UNIVERSITE AIX-MARSEILLE
FACULTE DE MEDECINE DE MARSEILLE
ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

THESE DE DOCTORAT

Présentée par

Mr. Mohamed SASSI

LA DIVERSITE DES ESPECES DU GROUPE *MYCOBACTERIUM ABSCESSUS* ET LEURS MYCOBACTÉRIOPHAGES

Soutenance le 25 Septembre 2013

En vue de l'obtention du grade de DOCTEUR de l'UNIVERSITE d'AIX-MARSEILLE
Spécialité : Pathologie Humaine-Maladies Infectieuses

Membres du jury de la Thèse :

Mr. le Professeur Jean-Louis HERRMANN	**Rapporteur**
Mr. le Docteur Paulo TAVARES	**Rapporteur**
Mr. le Docteur Christian CAMBILLAU	**Co-Directeur de Thèse**
Mr. le Professeur Michel DRANCOURT	**Directeur de Thèse**

Laboratoires d'accueil
Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes « URMITE », UM63, CNRS 7278, IRD 198, Inserm 1095, Faculté de Médecine
Architecture et Fonction des Macromolécules Biologiques, UMR CNRS 7257, Aix-Marseille-Université, Marseille, France

AVANT PROPOS

Le format de présentation de cette thèse correspond à une recommandation de la spécialité Maladies Infectieuses et Microbiologie, à l'intérieur du Master des Sciences de la Vie et de la Santé qui dépend de l'Ecole Doctorale des Sciences de la Vie de Marseille.

Le candidat est amené à respecter des règles qui lui sont composées et qui comportent un format de thèse utilisé dans le Nord de l'Europe et qui permet un meilleur rangement que les thèses traditionnelles ; par ailleurs, la partie introduction et bibliographie est remplacée par une revue envoyée dans un journal afin de permettre une évaluation extérieure de la qualité de la revue et de permettre à l'étudiant de commencer le plus tôt possible une bibliographie exhaustive sur le domaine de cette thèse. Par ailleurs, la thèse est présentée sur article publié, accepté ou soumis, associé d'un bref commentaire donnant le sens général du travail. Cette forme de présentation a paru plus en adéquation avec les exigences de la compétition internationale et permet de se concentrer sur des travaux qui bénéficieront d'une diffusion internationale.

Prof. Didier Raoult

ACKNOWLEDGEMENTS

Soyons reconnaissants aux personnes qui nous donnent
du bonheur ; elles sont les charmants jardiniers
par qui nos âmes sont fleuries.
Marcel Proust

Je tiens à exprimer en premier lieu toute ma reconnaissance au **Professeur Didier Raoult** pour m'avoir chaleureusement accueilli au sein de l'unité des Rickettsies et pour m'avoir permis l'octroi d'une bourse « Fondation Infectiopôle Sud ».

J'adresse mes plus chaleureux remerciements à mon Directeur de thèse, Monsieur le **Professeur Michel Drancourt**. Il n'a pas simplement accepté de diriger ma thèse, il m'a permis de travailler dans son équipe, m'a impliqué dans de si nombreux projets tous aussi excitants les uns que les autres, m'a transmis la passion de la recherche scientifique, et n'a eu de cesse de m'encourager et de me soutenir durant ces trois années. Je suis aussi très reconnaissant de la confiance qu'il m'a accordée et de la liberté qu'il m'a donnée dans mon travail. J'ai pu apprécier non seulement sa dimension scientifique, mais aussi sa non moins importante dimension humaine. J'en profite pour lui exprimer ici ma plus profonde gratitude.

Je souhaite remercier très chaleureusement mon co-directeur de thèse, le **Docteur Christian Cambillau** qui m'a beaucoup conseillé et aidé tant pour mes travaux sur les phages que sur le plan personnel. Il est difficile de trouver des mots assez forts pour souligner sa gentillesse, son humilité et sa patience à prodiguer des conseils pertinents.

Je remercie sincèrement le **Professeur Jean-Louis Herrmann** et le **Docteur Paolo Tavares**, d'avoir accepté d'être les rapporteurs de ma thèse.

Mes remerciements s'adressent également au **Professeur Pierre Pontarotti** et toute son équipe (Olivier, Philippe et Laurent) qui m'ont beaucoup aidé et appris surtout dans l'analyse phylogénétique et bio-informatique.

Je remercie tous les membres et ex-membres de l'unité des Rickettsies et AFMB pour le climat sympathique dans lequel ils m'ont permis de travailler. Les nombreuses discussions que j'ai pu avoir avec chacun d'entre eux m'ont beaucoup apporté.

Durant ces trois années de thèse, j'ai eu la chance de côtoyer de nombreuses personnes attachantes : que toutes soient remerciées pour les bons moments partagés. Je pense notamment (pardon à ceux que j'oublie) à Felix, Tahar, Majda, Samad, Fabrice, Micheline, Ginette, Amina, Aurélia, Adil, Nathalia, Fanny, Amira, Abdou, Alpha, Richard, Shady, Rita mais aussi de l'AFMB Badr, Cecilia, Silvia, Stéphanie, Eric, David, Johnny, Miguel, Mariella...

Enfin, les mots les plus simples étant les plus forts, je dédie ce travail à mes parents Mabrouka et Abdallah à qui j'adresse toute mon affection. Malgré mon éloignement depuis de nombreuses années, leur confiance, leur tendresse et leur amour me portent et me guident tous les jours. Merci pour avoir fait de moi ce que je suis aujourd'hui. Je dédie également cette thèse à mes parrains Dalila et Jean-Louis, qui étaient toujours des parents pour moi, leur amour, aides et soutient m'ont permis d'être la personne que je suis aujourd'hui, aussi mes sœurs Myriam et Linda, mon frère Moeen, ma future fiancée Asma avec qui la vie est plus belle et aussi sa famille et tous mes amis de longue dates et ceux que j'ai pu connaître ces dernières années (Emine, Aymen, Slim, Adam, Julien, Mathew, Jonathan, Tahar, Martin, Safia, Safa, Aurora, Claudia V, Claudia L, Nicolos, Marissa, Corina, Eva, Ana, Ana Maria, Iulia, Athena, ...).

J'espère que vous serez fiers de moi.

SOMMAIRE

RESUME

Mycobacterium abscessus sensu lato est la troisième espèce de mycobactéries la plus fréquemment documentée au cours des infections communautaires et nosocomiales dans les pays développés. Actuellement, l'espèce *Mycobacterium abscessus sensu lato* comprend deux sous espèces *M. abscessus* subsp. *abscessus* et *M. abscessus* subsp. *bolletii,* cette dernière regroupe des mycobactéries précédemment identifiées comme *M. bolletii* et *M. massiliense*. Ces changements taxonomiques montrent que la diversité de ces agents pathogènes opportunistes est mal décrite.

Au cours de notre travail de thèse, nous avons analysé 14 génomes publiés de *M. abscessus sensu lato* par approche bioinformatique, analyses fonctionnelles par BlastP et analyses phylogénétiques. Ce travail a montré que *M. abscessus sensu lato* comprend au moins cinq taxons différents spécifiés par des caractéristiques microbiologiques d'intérêt médical. En particulier chaque taxon est caractérisé par un répertoire spécifique de facteurs d'interaction avec l'hôte (y compris PE, PPE, LpqH, MCE, Yrbe et le système de sécrétion de type VII ESX3 et ESX4), de systèmes d'efflux de molécules, de multiples résistances et de prophages. Au cours d'un deuxième travail, nous avons développé une technique Multispacer Sequence Typing d'identification et de génotypage de *M. abscessus sensu lato* basée sur le séquençage de huit régions intergéniques. Cette technique distingue sans ambiguïté *M. massiliense* de *M. bolletii* et

M. abscessus et offre un outil d'identification pouvant être utilisée en routine au laboratoire.

Les mycobactéries *M. abscessus sensu lato* sont caractérisées par la présence de mycobactériophages. De façon surprenante, aucune étude structurale n'avait été réalisée pour aucun des ~ 3500 mycobactériophages décrits. Nous avons réalisé une étude génomique et structurale du bactériophage de *M. bolletii* que nous avons nommé Araucaria du fait de sa ressemblance à la plante Araucaria. Le mycobactériophage Araucaria appartient à la famille des *Siphoviridae* et possède un génome de 64 kb contenant 89 ORF, dont 27 sont annotés.

La résolution de cette première structure 3D d'un mycobactériophage a montré une capside et un connecteur similaires à ceux de plusieurs bactériophages de bactéries à Gram négatif et positif; et une queue hélicoïdale décorée par des pointes radiales (ressemblant à la plante Araucaria) susceptibles d'accueillir des dispositifs d'adhésion à son hôte et caractéristiques d'Araucaria. La partie basale (baseplate) du phage Araucaria présente des caractéristiques observées dans les phages qui se lient à des récepteurs de protéines. Ces résultats d'analyses structurales suggèrent un mécanisme d'adhésion d'Araucaria à son hôte en deux temps, par liaison de la queue aux saccharides de l'hôte dans un premier temps puis liaison de la baseplate aux protéines de la paroi cellulaire dans un deuxième temps. Ces hypothèses doivent être confirmées expérimentalement.

Enfin, nous avons analysé la présence de séquence de phages dans 48 génomes disponibles de *M. abscessus sensu lato*. Des séquences de prophage ont été détectées dans

47/48 génomes de *M. abscessus*, codant de 0 à 8 régions de prophages spécifiant quatre situations : i) prophages complets codant des protéines de structure, les protéines de lyse, des protéines d'intégration et des protéines nécessaires à la réplication et la recombinaison, ii) prophages incomplets dont certains gènes fonctionnels ne sont pas annotés avec certitude, iii) prophages incomplets ayant perdu une partie des gènes fonctionnels, iiii) éléments prophage-like codant seulement quelques protéines virales. Nous avons annoté 20304 protéines en 853 groupes: 30% des protéines annotées sont des protéines de mycobactériophages, 48% proviennent d'autres bactériophages infectant des bactéries à Gram négatif et positif et 23% sont des protéines virales. L'analyse fonctionnelle a montré 44% des protéines annotées impliquées dans la réplication de l'ADN et le métabolisme et 37% de protéines structurales. Par ailleurs, les espèces *M. abscessus* codent des protéines annotées comme répresseur de phage et dont le rôle reste indéfini.

Notre analyse phylogénétique suggère que les espèces de *M. abscessus sensu lato* ont été infectées par différents mycobactériophages et ont une histoire évolutive différente de celle des hôtes mycobactériens et contiennent aussi des protéines acquises par transfert horizontal dont la plupart sont des protéines de bactériophages et des protéines de fonctions inconnues.

Enfin, nous avons séquencé et analysé deux mycobactéries non-tuberculeuses responsables d'infections opportunistes, *Mycobacterium simiae* et *Mycobacterium septicum*. Le génome de *M. simiae* DSM 44165^{T} contient

5 782 968 pb et un taux de GC égal à 65,15%. Il code 5727 gènes dont 81,6% codent pour une fonction putative connue. Le génome de *M. septicum* DSM44393 ^{T} est 6 879 294 pb de longueur, avec un taux de GC égal à 66,73%. *M. septicum* code 6692 gènes dont 71,01% codent pour une fonction putative connue.

Nos travaux de recherche sur *M. abscessus sensu lato* et leurs mycobactériophages ont permis de résoudre la première structure d'un mycobactériophage, Araucaria, suggérant un mécanisme original d'infection en deux étapes. Egalement, ces travaux ont donné de nouveaux outils d'analyses au laboratoire de ces mycobactéries pathogènes opportunistes et constituent un socle pour de futures analyses fonctionnelles concernant le rôle de ces mycobactériophages dans la pathogénicité de leurs hôtes.

ABSTRACT

Mycobacterium abscessus sensu lato is the third group of mycobacteria most frequently encountered in community- and health care-associated infections in developed countries. *M. abscessus sensu lato* comprises of *M. abscessus* subsp. *abscessus* and *M. abscessus* subsp. *bolletii*. Later taxon accommodates isolates previously referred to as *Mycobacterium bolletii* and *Mycobacterium massiliense*. Current phenotypic and genetic markers poorly describe the diversity of this group of opportunistic pathogens. This high phenotypic and genotypic similarity leads to a confusing situation and some authors have questioned the current taxonomic classification.

In a first step, using bioinformatic approaches and functional analysis using BlastP, we reviewed the published genomes of 14 *M. abscessus* strains showing that *M. abscessus sensu lato* comprises of five different taxons specified by particular characteristics of microbiological and medical interests. Mycobacteria of the five clusters encode host-interaction factors (including PE, PPE, LpqH, MCE, Yrbe and type VII secretion system ESX3 and ESX4), drug-efflux systems, multiple resistances and prophages in agreement with their role as opportunistic pathogens. In a second step, based on sequencing of eight intergenic spaces, we developed a Multispacer Sequence Typing technique (MST) for *M. abscessus* group sub-species identification and strain genotyping. MST clearly differentiates formerly "*M. massiliense*" organisms from other *M. abscessus* subsp.

bolletii organisms. This technique could be a tool routinely used in laboratory.

These mycobacteria host bacteriophages which have not been explored so far. No structure of any mycobacteriophage is available, although more than 3,500 mycobacteriophages have been described to date. To fill this gap, we embarked in a genomic and structural study of a bacteriophage from *M. bolletii*. We named this mycobacteriophage Araucaria, it belongs to *Siphoviridae* family and possesses a 64-kb genome containing 89 ORFs, among which 27 could be annotated with certainty. We resolved Araucaria 3D structure, the first 3D structure ever resolved for a mycobacteriophage. Its capsid and connector share close similarity with several phages from $Gram^-$ or $Gram^+$ bacteria. The helical tail decorated by radial spikes, possibly host adhesion devices, according to which the phage name "Araucaria" was chosen being its most distinctive characteristics. Its host adsorption device, at the tail tip, assembles features observed in phages binding to protein receptors. All together, these results suggest that Araucaria may infect its mycobacterial host using a mechanism involving adhesion to cell wall saccharides and protein, a feature that remains to be further explored.

Beside *M. bolletii* mycobacteriophage, we analysed 48 *M. abscessus* sequenced genomes for encoding prophages. A total of 47/48 genomes encode at least one-prophage regions. Prophage regions were detected in 47/48 *M. abscessus* genomes. *M. abscessus* genomes harbor 0 to 8 prophage regions and these regions could be separated into four types: i) intact prophages who encodes structural

proteins, lysis proteins, integration proteins and sites and proteins necessary for the replication and the recombination, ii) questionable prophages which some functional genes could not be annotated with certainty, iii) incomplete prophage regions who lost some of the functional genes, iiii) prophage-like elements which encode few viral proteins. We annotated 20,304 viral and phage proteins clustered into 853 groups: 30% of the annotated proteins are mycobacteriophages proteins, 48% are other phages infecting Gram positive and negative bacteria and 23% are viral proteins. Moreover functional analysis showed that 44% of annotated proteins are implicated in DNA replication and metabolism and 37% are bacteriophages' proteins including structural proteins. Our phylogenetic analyses suggested that *M. abscessus* species were infected by different mycobacteriophages and have a different evolutionary history than the bacterial hosts and some proteins that are acquired by horizontal gene transfer mostly mycobacteriophages' proteins and hypothetical proteins.

Finally, we sequenced and analyzed two non-tuberculosis mycobacterium causing human infections, *Mycobacterium simiaie* and *Mycobacterium septicum.* The draft genome sequence of *M. simiae* DSM 44165^{T} is 5,782,968-pb long with 65.15 % GC content (one chromosome, no plasmid) containing 5,727 Open Reading Frames (ORFs) which 81.6% were assigned a putative function. The draft genome sequence of *M. septicum* DSM44393 ^{T} is 6,879,294-bp in length, with an overall GC content of the chromosome of 66.73% containing 6,692 protein-coding genes which 71.01 % were

assigned a putative function.

Our works on *M. abscessus sensu lato* and mycobacteriophages have solved the first structure of a mycobacteriophage, Araucaria, suggesting a novel mechanism of infection in two stages. Also, these studies have provided new tools for laboratory analysis of opportunistic pathogenic mycobacteria and provide a foundation for future functional analyses on the role of mycobacteriophages in the pathogenicity of their hosts.

I- INTRODUCTION: *MYCOBACTERIUM ABSCESSUS* ET MYCOBACTÉRIOPHAGES

Commentaire

Les mycobactéries sont phylogénétiquement classées parmi les bactéries à Gram positif à haut GC% caractérisées par une paroi cellulaire qui forme une véritable enveloppe cireuse et protectrice du fait de sa propriété d'acido-alcoolo-résistance et sa richesse exceptionnelle en acides gras et lipides. Il s'agit d'un genre bactérien extrêmement polymorphe qui comporte des mycobactéries à croissance rapide (moins de 7 jours pour produire des colonies visibles), des mycobactéries à croissance lente (7-60 jours pour produire des colonies visibles) et une espèce non-cultivée en dehors des animaux, *Mycobacterium leprae*, agent de la lèpre. La lèpre, la tuberculose et l'ulcère de Buruli causé par *Mycobacterium ulcerans* sont les trois principales infections à mycobactéries dans le monde tandis que les autres espèces dont *Mycobacterium abscessus* qui est l'objet de notre travail de thèse, constituent des pathogènes opportunistes et le troisième groupe de mycobactéries les plus fréquemment rencontrées dans les laboratoires de microbiologie clinique dans les pays développés [1, 2].

M. abscessus sensu lato regroupe des mycobactéries non tuberculeuses longtemps confondues avec *Mycobacterium chelonae* [3]. *M. abscessus sensu lato* est composé de *M. abscessus* subsp. *abscessus* et *M. abscessus* subsp. *bolletii* [4] qui a été précédemment identifié comme «*Mycobacterium massiliense* » ou « *Mycobacterium bolletii*» [5,6]. La discrimination et l'identification de *M. abscessus*

sensu lato restent difficiles. Les autres espèces étroitement apparentées comprennent *M. salmoniphilum* [7], *M. immunogenum* [8], *M. massiliense* [5], *M. bolletii* [6] et *Mycobacterium franklinii* [9] formant ensemble le complexe *M. chelonae-abscessus*. Sur le plan médical, *M. abscessus* est souvent isolé à partir d'échantillons des voies respiratoires recueillis chez des patients présentant une maladie pulmonaire chronique telle que la mucoviscidose [10]. Dans cette situation, il a été montré qu'un membre de *M. abscessus sensu lato*, *M. abscessus* subsp. *massiliense* est transmissible de patient à patient [11]. Ces mycobactéries sont également responsables d'infections de la peau et des tissus mous à la suite d'interventions chirurgicales et cosmétiques [1, 2]. Les infections à *M. abscessus* sont particulièrement critiques pour les patients immunodéprimés chez qui elles peuvent être fatales [12]. Les chats [13,14] et les dauphins [15,16] sont également infectés par *M. abscessus* alors que les poissons sont particulièrement infectés par *M. salmoniphilum* [17,7]. Les infections à *M. abscessus* sont également difficiles à traiter, en raison de la résistance naturelle à large spectre de ces mycobactéries qui peuvent également acquérir des facteurs de résistance [1, 2, 11].

Quelques publications ont rapporté des séquences de phages dans le génome des souches de *M. abscessus* [18, 19]. Cette particularité reste peu analysée. Actuellement, 3607 mycobactériophages ont été isolés essentiellement à partir d'échantillons environnementaux à l'aide de *M. smegmatis* comme hôte [20]. Ces mycobactériophages sont de l'ordre des Caudovirales,

la plupart de la famille des *Siphoviridae*. Ils sont caractérisés par la présence d'une queue non-contractile flexible relativement longue ; neuf mycobactériophages sont classés dans la famille des *Myoviridae*, caractérisés par la présence d'une queue contractile [20]. Bien que les mycobactériophages *Myoviridae* soient morphologiquement identiques, les siphovirus montrent des morphologies variables avec une longueur de queue variant de ~ 105 à ~ 300 nm et des structures à l'extrémité de la queue visiblement différentes selon les phages [20-24]. Pour la plupart, les capsides sont isométriques, bien que certains mycobactériophages comme Corndog, Che9c et Brujita aient des capsides aplaties. La taille de la capside isométrique varie de ~ 48 nm de diamètre (BP et Halo) à ~ 85 nm de diamètre (Bxz1). En général, la taille de la capside corrèle avec la taille du génome suggérant une densité d'emballage de l'ADN relativement constante [20-24]. Certains phages (par exemple D29) ont un large spectre d'hôtes et infectent aussi bien des mycobactéries à croissance rapide (par exemple *M. smegmatis)* que des mycobactéries à croissance lente (par exemple *M. tuberculosis*) [20], tandis que d'autres mycobactériophages (par exemple Barnyard et DS6) infectent un seul hôte [25]. Par ailleurs, plusieurs phages peuvent discriminer entre les souches d'une espèce particulière, par exemple le phage 33D permet de distinguer entre les souches *M. bovis* dont les souches vaccinales BCG, et plusieurs phages ont des préférences pour certaines souches de *M. smegmatis*. Les barrières moléculaires et génétiques opposées par l'hôte et le mécanisme d'infection des mycobactériophages ne

sont pas connus.

Au total, 531 mycobactériophages sont séquencés [Mycobactériophage Database] révélant leur grande diversité et le fait qu'ils véhiculent des gènes majoritairement de fonction inconnue. Les génomes des mycobactériophages sont regroupés en différents clusters et des singletons [20-24]. Une caractéristique inhabituelle est la taille moyenne du génome d'environ 70 kb avec des variations importantes entre 41 kb (Angel) et 150 kb (Bxz1, Catera) [24]. Cette taille moyenne est deux fois plus grande que celle des phages qui infectent les bactéries commensales [24]. Les principales différences dans la taille correspondent à la présence de gènes non structuraux.

Les génomes des mycobactériophages présentent un mosaïcisme plus accentué que celui des autres bactériophages [20-24]. Cela corrèle avec différentes histoires évolutives des différents segments de génome compliquant leurs études phylogénétiques [24]. Les mycobactériophages peuvent être utilisés comme outils de manipulation génétique des mycobactéries pathogènes. La construction de plasmide de réplication sur la base de phages TM4 et D29 a permis de construire des mutants knock-out [26-28]. En outre, les mycobactériophages sont utilisés comme marqueurs de diagnostic afin d'améliorer et d'accélérer la reconnaissance des mycobactéries pathogènes dans les échantillons de patients, ainsi que la résistance aux médicaments de ces pathogènes [29, 30]. En outre, des systèmes rapporteurs de la luciférase basés sur les phages TM4 et D29 ont été construits dans le même but [31, 32]. Aussi il a été démontré que les

mycobactériophages peuvent être utilisés comme agents anti-mycobactériens contre les infections causées par *M. avium* et *M. tuberculosis* [33].

La disponibilité du génome de 39 *M. abscessus*, 13 *M. massiliense* et deux *M. bolletii* dans les bases de données du National Center for Bioinformatics ouvre de nouvelles perspectives permettant d'évaluer la diversité de ces taxons et leurs mycobactériophages.

La première partie de nos travaux soumis à **BMC Genomics** est une revue de génomique comparative de 14 génomes publiés de *M. abscessus* afin d'analyser en profondeur leur diversité. Dans une deuxième partie, nous avons développé une approche moléculaire pour l'identification et le génotypage de *M. abscessus*. Ces résultats on été publiés dans **BMC Microbiology**. Dans la troisième partie, nous avons réalisé par microscopie électronique la première reconstruction 3D d'un mycobactériophage et une analyse bioinformatique du mycobactériophage de *M. bolletii* que nous avons nommé Araucaria. Ces résultats ont fait l'objet d'une publication dans **Journal of Virology**. Ensuite, nous avons investigué la diversité de prophages dans 48 génomes de *M. abscessus* et étudié l'histoire évolutive et fonctionnelle des prophages codés dans 48 génomes séquencés de *M. abscessus sensu lato*.

En parallèle, dans le but de comprendre la pathogénicité des mycobactéries non tuberculeuses nous avons séquencé le génome de deux mycobactéries pathogènes opportunistes; ces travaux sont

présentés dans la partie « annexes ». Ces deux génomes ont respectivement fait l'objet d'une publication dans **Genome Annoucement** et **Standards in Genomic Sciences**.

Article N.1:

Sassi M and Drancourt M. Reviewing post-genomic diversity of *Mycobacterium abscessus* sensu lato. BMC Genomics (Soumis)

Reviewing post-genomic diversity of *Mycobacterium abscessus*

Mohamed Sassi[1] and Michel Drancourt[1]*

[1] Aix Marseille Université, URMITE, UMR63, CNRS 7278, IRD 198, Inserm 1095, Marseille, France

*Corresponding author

E-mail : michel.drancourt@medecine.univ-mrs.fr

Word count abstract: 339/350

Word count text: 2881/3000

N° of tables: 4

N° of figures: 5

Abstract

Background *Mycobacterium abscessus sensu lato*, the third group of mycobacteria most frequently encountered in community- and health care-associated infections in developed countries, comprises of *M. abscessus* subsp. *abscessus* and *M. abscessus* subsp. *bolletii*. Later taxon accommodates isolates previously referred as *Mycobacterium bolletii* and *Mycobacterium massiliense*. The diversity of this group of opportunistic pathogens is poorly described.

Results Herein reviewing 14 published *M. abscessus sensu lato* genomes found a pangenome of 6,153 proteins and core genome of 3,947 proteins. Combining genome length (from 4.8-Mb to 5.51-Mb), average nucleotide identity (from 94.19% to 98.58%), phylogenomic analysis and the number of prophage regions (from 0 to 7) discriminates five clusters: cluster n° 1 comprises of *M. abscessus sensu stricto* mycobacteria, clusters n° 2A and n° 2B comprise of *M. massiliense* mycobacteria and clusters n° 3A and n° 3B comprise of *M. bolletii* mycobacteria. The mean number of prophage regions is 2 in cluster n° 1; 1.33 in cluster n° 2A; 3.5 in cluster n° 2B and five in clusters n° 3A and n° 3B. A total of 36 genes are uniquely present in cluster n° 1, 15 in clusters n° 2A and n° 2B and 15 in clusters n° 3A and n° 3B. These genes could be used in the detection and identification of *M. abscessus* organisms. Further, the mean number of host-interaction factors (including PE, PPE, LpqH, MCE, Yrbe and type VII secretion system ESX3 and ESX4) varies from 70 in cluster n° 1, 80 in cluster n° 2A, 74 in cluster n° 2B and 93 in clusters n° 3A and n° 3B. As for antibiotic and biocide resistance, mycobacteria of the five clusters encode drug-efflux systems and multiple resistances. They encode both penicillin-binding proteins targeted by β-lactam antibiotics and an Ambler class A β-lactamase for which inhibitors exist.

Conclusions Altogether, these data indicate that *M. abscessus sensu lato* comprises of five different taxons specified by particular characteristics of microbiological and medical interests. Genomics data here reported could base the development of laboratory tools for the description of *M. abscessus* diversity.

Keywords: *Mycobacterium abscessus*, *Mycobacterium bolletii*, *Mycobacterium massiliense*, Araucaria, mycobacteriophage, prophage.

Background

The non-tuberculous mycobacterium *Mycobacterium abscessus* was long confused with *Mycobacterium chelonae* [1]. Other closely related species include *Mycobacterium salmoniphilum* [2], *Mycobacterium immunogenum* [3], *Mycobacterium massiliense* [4], *Mycobacterium bolletii* [5] and *Mycobacterium franklinii* [6] altogether forming the *Mycobacterium chelonae-abscessus* complex. This complex is the third group of mycobacteria most frequently encountered in clinical microbiology laboratories in developed countries besides the *Mycobacterium tuberculosis* and *Mycobacterium avium* complexes [7, 8]. Bibliometrics retrieving over 1,700 publications in the Medline database illustrates the emerging status of this complex which members are causing both sporadic cases and outbreaks of community-acquired and health-care associated infections [9] (Figure 1). Not only humans but also cats [10,11] and dolphins [12-14] are infected while fishes are uniquely infected by *M. salmoniphilum* [2,15].

The species *M. abscessus sensu lato* is comprising of *M. abscessus* subsp. *abscessus* and *M. abscessus* subsp. *bolletii*. Later taxon accommodates isolates previously referred as *M. bolletii* or *M. massiliense* [16]. The availability of 39 *M. abscessus*, 13 *M. massiliense* and two *M. bolletii* genomes in National Center for BioInformatics (NCBI) genome database provides new opportunities to assess the diversity of this species. Here, we review 14 complete published *M. abscessus* genomes and compare them with the re-annotated *M. tuberculosis* H37Rv genome (Table 1) in order to in-depth analyse the diversity of *M. abscessus*.

Results and discussion

Mycobacterium abscessus diversity

Genetic analyses poorly describe the diversity of *M. abscessus sensu lato* mycobacteria. For example, the 16S rRNA gene yields an identical sequence for *M. abscessus* and *M. bolletii*, which shares 99% sequence identity with *M. massiliense, M. chelonae, M. immnogenum* and *M. salmoniphilum*. *Rpo*B gene sequencing founded the description of recent species [17-19] but yielded further conflicting results [20-22]. Multilocus sequencing analysis [23], multispacer sequence typing [24] and whole genome sequencing [8] differentiated *M. massiliense* from *M. bolletii*.

M. abscessus mycobacteria comprise one circular chromosome. In addition, *M. abscessus* ATCC 19977 contains one 23-kb plasmid identical to the *Mycobacterium marinum* pMM23 plasmid, encoding mer operon and mercury reductase protein, which may confer resistance to organomercury compounds [25]. The mean percentage of nucleotide sequence identity of core proteins [26] indicate that *M. abscessus sensu lato* mycobacteria form five clusters; cluster n°1 includes *M. abscessus* type strain and strains M93, 94, M152 and Go06; cluster n°2A includes *M. massiliense* type strain and strains M154 and M18; cluster n°2B includes strains 47J26, M115, M172 and M139; cluster n°3A includes *M. bolletii* type strain and cluster 3B includes *M. bolletii* strain M24 (Table 2). Interestingly, group 2B isolates have been made only in Malaysia, Asia, suggesting a limited regional diffusion of this cluster. *M. abscessus sensu lato* proteomes were aligned using Mauve software [27] to infer phylogeny using the Neighbor-Net algorithm in the package SplitsTree4 [28]. The phylogenomic network indicates that *M. abscessus* genomes form three clusters as defined above (Figure 2A). Also, a phylogenomic tree based on

gene content (i.e., the presence or absence of orthologs) (Figure 2B) differently organized *M. abscessus* genomes from the whole genome concatenated tree (Figure 2A) or even the phylogenetic tree based on *rpo*B gene sequences (Figure 2C). Interestingly, phylogenomic analysis showed that the *M. abscessus* gene repertoires have different evolutionary histories and suggested that differential gene loss and lateral gene acquisition are playing important roles in the evolution of some *M. abscessus* strains. Notably, the situation of strain Go06 is confusing, as it presents 98.4% nucleotide sequence identity with *M. abscessus* type strain in cluster n°1 (Figure 2A) whereas its *rpo*B gene sequence and single nucleotide polymorphisms analysis were indicative of *M. massiliense* [29 and 8]. Our analyses indicate that strain Go06 have an ambiguous classification as a chimera between *M. abscessus* and *M. massiliense* and is the only example compatible with a lateral transfer of *rpo*B gene within *M. abscessus*. *M. abscessus* median GC% content is 64.2%, ranging from 62.7% (*M. abscessus* ATCC 19977) to 64.2% (strain Go 06). The GC% is not characteristic of the clusters as the median GC% content of cluster n°1, 2A and 3 is 64.2%, close to the median 64.1% GC% content in cluster n°2B. However, there is a significant 14.7% variation in the genome length from 4.8-Mb (*M. abscessus* M154) to 5.51-Mb (*M. abscessus* M24) with a median of 5.07-Mb. In cluster n°1, median of genome size is 5.07-Mb, 4.89-Mb in cluster n°2A, 5.01-Mb in cluster n°2B and 5.28-Mb in cluster n°3.

Altogether, genomics analyses revealed a more heterogeneous structure of *M. abscessus sensu lato* than the one currently suggested by the nomenclature, which recognizes only two subspecies within *M.* abs*cessus* [16]. Recent whole genome sequencing analyses of clinical isolates in Great Britain also clearly

distinguished the three clusters here reported [8]. These data plead for revaluating the taxonomy of *M. abscessus*, which comprises of five taxons, three species corresponding to the clusters n° 1, 2 and 3 with *M. massiliense* and *M. bolletii* each comprising of two sub-species.

Mycobacterium abscessus pan- and core-genome

In order to normalize the predicted proteins and to minimize differences of presence/absence of genes and length, the coding sequences were predicted using prodigal software [30]. In total, we identified 70,309 protein-coding sequences. The number of protein-coding genes in each genome varies from 4,651 to 5,079 (Table 3). We first performed clustering using orthoMCL [31] with a conservative parameter value of 50% sequence identity this allows to find coding genes that are unique to the different clusters. In total, we identified 3,947 (64%) groups of proteins composing the *M. abscessus* core genome. They are containing a total number of 57,172 protein sequences. Then we annotated the sequences using Clusters of Orthologous Groups database (COG) [32] and BLASTP. We identified 36 genes unique to cluster n° 1 coding hypothetical proteins, proteins implicated in transcription, energy production and transport, carbohydrate metabolism and transport, lipid metabolism nucleotide metabolism and transport, amino acid metabolism and transport, post-translational modification and inorganic ion transport and metabolism (Table 3, Figure 3). Eleven unique genes were identified in M93, coding proteins implicated in amino acid metabolism and transport and transcription. Ten unique genes were identified in M94, coding lipid metabolism proteins. Within cluster n° 2, we identified 15 unique genes, coding hypothetical proteins, proteins implicated in transcription and lipid metabolism. Three unique genes were identified in *M. massiliense* type strain, coding lipid

metabolism proteins. Eight unique genes were identified in M18, coding lipid metabolism proteins. Four unique genes were identified in M115 and M139, coding respectively hypothetical proteins and lipid metabolism proteins. Twenty unique genes were identified in M172, coding proteins implicated in replication and repair, signal transduction and transcription. Within cluster n° 3, we identified 15 unique genes, coding hypothetical proteins, proteins implicated in amino acid metabolism and transport and translation. Nine unique genes were identified in *M. bolletii* type strain, coding hypothetical proteins and proteins implicated in transcription. Twenty-three unique genes were identified in M24, coding hypothetical proteins, proteins implicated in nucleotide metabolism and transport, amino-acid metabolism and transport, post-translational modifications and inorganic ion transport and metabolism. These unique genes

could base a refined identification of the taxon here defined as forming *M. abscessus*. However, we could not exclude that these unique genes could be due to a coding sequence, which arose denovo, to HGT or gene loss for the other subspecies. In the case of absence of a gene, this could also be due to a real loss or to anassembly artefact.

Mycobacterium abscessus **prophagome**

Differences in the genome size of *M. abscessus* mycobacteria correlate with the number of prophage regions which are detected in 13/14 *M. abscessus* genomes (Figure 4): *M. abscessus* M154 (*M. massiliense* cluster n° 2A) has the smallest genome encoding no prophage whereas *M. bolletii* M24 (cluster n° 3) has the largest genome encoding seven prophage regions (Figure 4). Prophage regions comprise up to 5% of the genome lenght in *M. abscessus* M172. The number of prophage regions in other

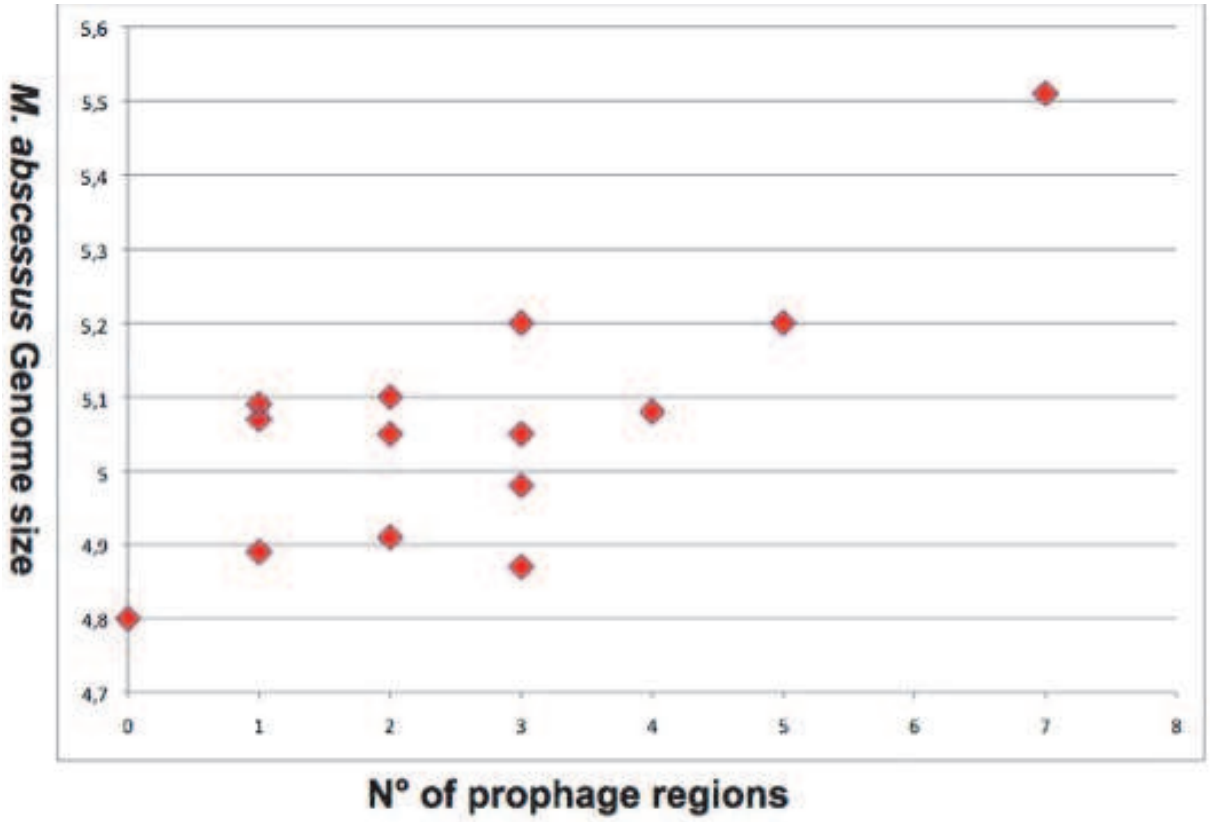


Figure 4. Correlation between *Mycobacterium abscessus* genomes size (y axis) and the number of prophages (x axis).

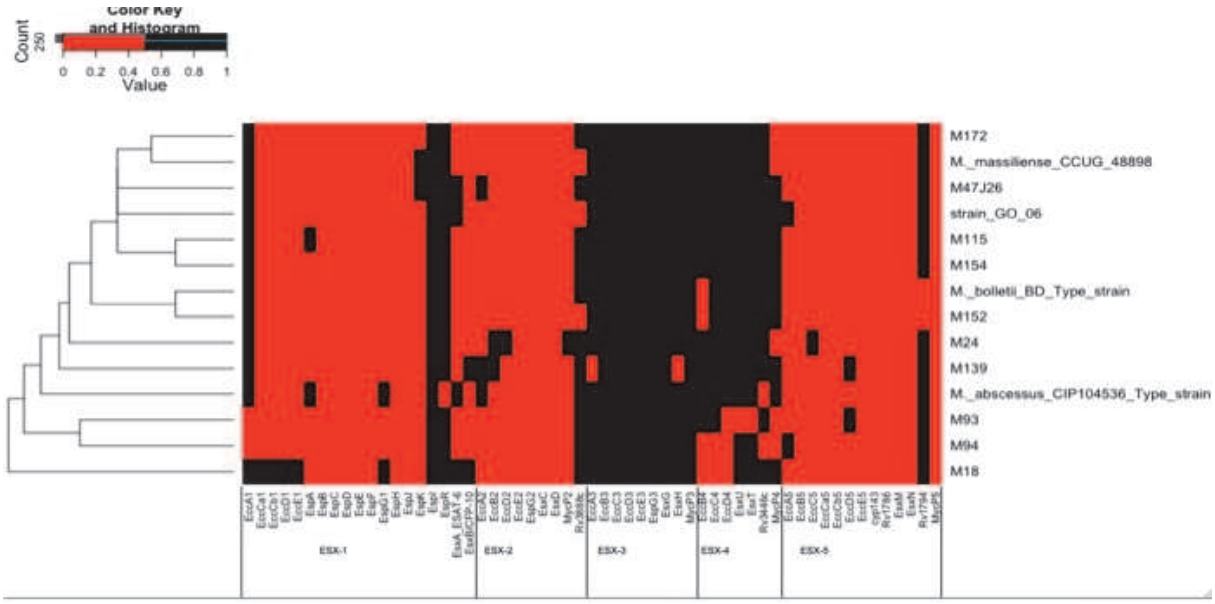


Figure 5. Heatmap clusterisation of *Mycobacterium abscessus* type VII secretion system compared to *Mycobacterium tuberculosis* H37Rv. M. abscessus strains are listed on the left side of the map.

genomes is diverse, ranging from one to six regions encoding putative genes in the subsystem of phages, prophages, transposable elements and plasmids, which might contribute to species diversity [33]. The mean number of prophage regions is 2 in cluster n° 1, 1.33 in cluster n° 2A, 3.5 in cluster n° 2B and 5 in cluster n° 3. This observation confirms the particularity of cluster 3: *M. bolletii* CIP108541 ^{T} contains a 13-kb and a 63-kb prophages whereas *M. bolletii* strain M24 contains seven prophage regions including one 17-kb region homologous to the *M. bolletii* CIP108541 ^{T} 13-kb region and a 27-kb region homologous to the *M. massiliense* CCUG 48898 50-kb region [34,35] (Table 4). *M. abscessus* genomes encode putative phage-related genes necessary for phage replication as well as phage-tail protein, phage endolysin, capsid proteins (major protein and scaffold proteins) and phage tape measure protein. Both ends of this region encode putative phage integrases. *M. abscessus* genomes encode small prophage-like regions. However, only *M. bolletii* mycobacteria have been reported to produce a mycobacteriophage that we named Araucaria after we recently resolved its electron microscopy 3D structure [36]. *M. abscessus* M94 genome harbours one particular pseudo-tRNA spanning the region 51,150-57,394 in contig 33, which is not observed in the other *M. abscessus* genomes [37].

Mycobacterium abscessus **resistome**

As all mycobacteria, *M. abscessus sensu lato* mycobacteria are embedded into a hydrophobic cell wall barrier to hydrophilic antibiotics. Accordingly, *M. abscessus* mycobacteria are multidrug resistant organisms exhibiting different drug resistance [38-40]. *M. abscessus sensu lato* genomes encode many proteins potentially involved in drug-efflux systems, including members of the major facilitator family, ABC transporters and MmpL proteins; Small

Multidrug Resistance-family, a family of lipophilic drug efflux proteins [41]; and a multidrug resistance stp protein similar to *M. tuberculosis* involved in spectinomycin and tetracycline resistance [42]. *M. abscessus*, *M. bolletii* and *M. massiliense* were reported to be *in-vitro* susceptible to amikacin; however, comparison with the *M. tuberculosis* H37Rv resistome and the antibiotic resistance databases indicate that *M. abscessus* mycobacteria encode an aminoglycoside 29-N-acetyltransferase and aminoglycoside phosphotransferases involved in resistance to aminoglycosides. Also, genetic analyses disclosed 16S rRNA gene mutations conferring aminoglycoside resistance [4,5,43]. Indeed, the presence of a single rRNA operon in all of the *M. abscessus* genomes favors the occurrence of dominant mutations conferring resistance to aminoglycosides and macrolides. *M. abscessus* genomes encode a rifampin ADP-ribosyl transferase and monooxygenases potentially involved in resistance to rifampin and tetracyclines. Moreover, *M. abscessus* genomes encode three *tet*(M) genes conferring resistance to tetracyclyine and doxycycline; the number of *tet*(M) genes was correlated to the resistance to cyclines in *Escherichia coli* [44]. However, *M. massiliense* was reported to be susceptible and *M. abscessus* and *M. bolletii* to be resistant to doxycycline [45]. *M. abscessus* mycobacteria encode resistance to fusidic acid, glycopeptides, MLS (Macrolide-Lincosamide-StreptograminB), phenicols, rifampicin, sulphonamide and trimethoprim. Also, *M. abscessus* genomes encode FolP homologs conferring resistance to cotrimoxazole, homolog of UDP-N- acetylglucosamine 1-carboxyvinyltransferase, a MurA protein conferring resistance to fosfomycin and homologs of 23S rRNA methylases conferring resistance to macrolides. Also, *M. abscessus* genome encodes an erm(41) gene which mutations were reported to confer clarithromycin resistance [46]. *In-vitro* tests showed that

M. massiliense clinical isolates could be distinguished from *M. abscessus* isolates for their susceptibility to ciprofloxacin [47] whereas *M. bolletii* isolates were reported to be resistant to all quinolones [48]. A mutation at codon 90 in gyrA gene was reported in clinical isolates of *M. abscessus* exhibiting high resistance to ciprofloxacin [47]. This observation contrasts with our genome analysis, which found no such mutations, suggesting that other mechanisms of resistance may be involved in high-level resistance to quinolones [48]. Accordingly, we found that *M. abscessus* mycobacteria encode qepA2, a plasmidic gene conferring quinolone resistance in gram-negative bacteria [49]. *M. abscessus* mycobacteria were reported to be *in-vitro* resistant to penicillin, amoxicillin, cefoxitin, ceftriaxone, cefotaxime and imipenen [4,5]. This contrasts with the fact that they encode Penicillin-binding proteins (PBPs), targets for β-lactam antibiotics (except for tabtoxinine-β-lactam, which inhibits glutamine synthetase), which are essential for peptidoglycan synthesis [50,51]. *M. abscessus* genomes encode an Ambler class A β-lactamase homologous to β-lactamases in gram-negative bacteria and to two β-lactamases in *M. tuberculosis*. β-lactamases inhibitors have not been evaluated against *M. abscessus sensu lato* mycobacteria.

Post-genomic analysis of host-interactions

M. abscessus mycobacteria are ubiquitous environmental organisms in soil and water [9] where they may have to cope with amoeba: *M. chelonae, M. abscessus, M. massiliense* and *M. immunogenum* were reported to survive within *Acanthamoeba polyphaga* tropohozoites and cysts [5]. Accordingly, our analyses indicate that *M. abscessus* mycobacteria encode factors implicated in host interactions. The mean number of genes encoding PE, PPE, LpqH, MCE, Yrbe and type VII secretion system is of 70 in

M. abscessus cluster n° 1, 80 in *M. massiliense* cluster n° 2A, 74 in *M. massiliense* cluster n° 2B and 93 in *M. bolletii* cluster n° 3. In *M. abscessus*, rough colonies lack *mmp*L4 (a gene required for glycopeptidolipid biosynthesis) and lost surface colonization, replication into human macrophages and stimulation of innate immune response; these observations suggested that glycopeptidolipid was a virulence factor [52-54]. Accordingly, glycopeptidolipids are required for sliding motility [55] and biofilm formation [56]. Glycopeptidolipids have also been suspected to inhibit phagocytosis of *M. avium* subsp*. avium* [57]. *M. abscessus* genomes encode Mammalian Cell Entry (MCE) proteins similar to *M. tuberculosis* H37Rv. MCE operon promotes internalization of *M. tuberculosis* by mammalian cells [58] and initiates rapid induction of transcription of genes involved in substrate trafficking [59]. The number of *mce* operons which correlated with pathogenicity [60], varies from six in cluster n° 2B mycobacteria to 13 in cluster n° 3 mycobacteria. In parallel, *M. abscessus* genomes encode 12 (cluster n° 1) to 21 (cluster n° 3A and 3B) copies of Yrbe proteins. As for secretion systems, recent evidences showed that mycobacteria evolved specialized type VII secretion systems to transport extracellular proteins across the cell wall [61]. Type VII secretion systems ESX-1 and ESX-5 are involved in cell-to-cell migration of *M. tuberculosis* [61,62]. In *M. abscessus*, our analyses indicate that ESX-3 and ESX-4 systems are conserved (Figure 5). However, *M. abscessus* M139 (cluster n° 2B) lacks two proteins of the ESX-3 system and *M. abscessus* M93 (cluster n° 1) lacks ESAT-6 like and CFP-10-like proteins secreted by the ESX-4 system. Interestingly, *M. abscessus* M18 (cluster n° 2A) encodes ESAT-6 and CFP-10 proteins secreted by ESX-1 system. In addition, there are two or three PE (proline-glutamate) and six (*M. massiliense*, *M. abscessus* M115 or *M. abscessus* 47J26) to 12 (*M. bolletii* M24) PPE (proline-proline-

glutamate) proteins, which are reported to be involved in the virulence of *M. tuberculosis* [63]. Our analyses further indicated that proteins related to phenazine biosynthesis, homogentisate catabolism, phenylacetic acid degradation and DNA degradation might have been transferred from *Actinobacteria* (e.g. *Rhodococcus* spp., *Streptomyces* sp.p) and pseudomonas (*Pseudomonas aeruginosa* and *Burkholderia cepacia*). Although distantly related, these bacteria share the same ecosystem as *M. abscessus* mycobacteria within cystic fibrosis microbiota.

Conclusions

M. abscessus mycobacteria emerge as opportunistic pathogens. Data here presented indicate that these mycobacteria are more diverse than previously suspected. We here define five clusters which, in addition to unsequenced species (*M. salmoniphilum*, *M. franklinii*, *M. chelonae* and *M. immunogenum*), may specify at least nine different taxons within the *M. chelonae-abscessus* complex. A very interesting feature feature is the small, "closed" pangenome, less than twice the size of the core genome. Accordingly, we identified 66 genes uniquely present in each cluster; these genes could be used in refined detection and identification of *M. abscessus* organisms. We indeed observed that each taxon has specificities; in particular, members of clusters n° 3A and n° 3B exhibit larger genomes encoding mycobacteriophages and more host-interaction factors than the other clusters. These particular capacities are relevant not only to microbiology but also to medicine as they specify different prognosis of infection and treatment. Therefore, genomics data here reported could serve to develop laboratory tools useful for the routine diagnosis of *M. abscessus* diversity.

Methods

Genome dataset

The whole genomes of 14 *M. abscessus* strains were downloaded from Genbank (Table 1). The genomic sequence, either contigs or finished genomes were concatenated to one pseudogenome per genome.

Prophage detection and genome annotation

Protein sequences were predicted using prodigal software [30] to generate normalized files containing the combined protein sequences of all 14 genomes. Prophage regions were detected using PHAST software (Table 4). Predicted proteins were annotated using BLASTp against the National Center for Biotechnology Information (NCBI) non-redundant (NR) database, UNIPROT (http://www.uniprot.org/), the Clusters of Orthologous Groups (COG) [33] and a home-made antibiotic resistance gene database.

Genome clustering and calculation of core genomes

Proteome sequences were compared using by BlastP and pairwise alignments using ClustalW and the ANI was determined by the mean percentage of nucleotide sequence identity of core proteins [26]. We clustered the *M. abscessus* homologous genes using orthoMCL [31] on the translated protein sequences of all predicted genes with a conservative parameter value of 50% sequence identity. The determination of the different unique core genomes was based on the homology clusters found by orthoMCL.

Phylogentic analysis

M. abscessus proteomes were aligned using Mauve software [27] to infer phylogeny using the Neighbor-Net algorithm in the package SplitsTree4 [28]. The orthologous group data found by orthoMCL were used to construct a whole-genome phylogenetic tree based on gene content. We generated a matrix of binary discrete characters ("0" and "1" for absence and presence, respectively) [64]. Using this matrix, we constructed a phylogenetic tree implementing the neighbor-joining (NJ) method within SplitsTree4 [28].

Competing interests

The authors declare that they have no competing interests.

Authors' contribution

MS carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript.

MD conceived of the study, and participated in its design and coordination and drafted the manuscript.

All authors read and approved the final manuscript.

Acknowledgments

MS is financially supported by the Infectiopôle Sud Foundation.

416 **Table 1 :** List of *Mycobacterium abscessus* genomes here studied.

Organism	Isolated from	Geography	BioProject
***M. abscessus* CIP104536 Type strain**	human knee infection	unknown	PRJNA61613, PRJNA15691
***M. abscessus* 47J26**	isputum sample from a patient with CF	England	PRJNA179981, PRJNA73255
***M. abscessus* M93**	sputum sample from a Malaysian patient presenting with a prolonged productive cough suggestive of a bacterial lower respiratory tract infection	Malysia	PRJNA180393, PRJNA84203
***M. abscessus* M94**	sputum sample of a Malaysian patient with a persistent cough and fever and consolidation in the chest radiograph	Malysia	PRJNA180394, PRJNA88149
***M. abscessus* M152**	acid-fast bacillus (AFB)-positive sputum of a Malaysian man	Malysia	PRJNA159789
***M. bolletii* BD Type strain**	respiratory tract specimen collected in woman with hemoptoic pneumonia	Marseille	PRJNA180015, PRJNA73695
***M. abscessus* M24**	the bronchoalveolar lavage fluid of a Malaysian patient	Malysia	PRJNA89595
***M. massiliense* CCUG 48898**	sputum specimen from hemoptoic pneumonia	Marseille	PRJNA180742, PRJNA65215
***M. massiliense* M18**	lymph node biopsy specimen from a Malaysian patient suspected of having tuberculous cervical lymphadenitis	Malysia	PRJNA89593
***M. massiliense* strain GO 06**	undergone knee joint surgery	Brazil	PRJNA170732, PRJNA168263
***M. abscessus* M154**	*M. massiliense* strain M154	malysia	PRJNA89603
***M. abscessus* M115**	sputum from a Malaysian patient presenting with persistent cough and loss of body weight suggestive of pulmonary tuberculosis	Malysia	PRJNA89601
***M. abscessus* M139**	sputum sample of a 26-year-old Nepalese male presenting with hemoptysis	Nepal	PRJNA159701
***M. abscessus* M172**	putum isolate from a Malaysian patient	Malysia	PRJNA89599

417 **Table 2 :** Average nucleodite identity and characteristics of *Mycobacterium abscessus* genomes.

Groups		Strain	Genome lenght Mb	Genome GC%	ANI Vs abscessus T	ANI Vs bolletii T	ANI Vs massiliense T
		M. abscessus CIP104536T	5,09	62,7	1,00	95,56	94,74
		M93	5,08	64,2	97,30	95,35	94,76
		M94	5,1	64,2	97,56	95,67	94,79
		M152	4,91	64,1	98,59	96,33	95,73
Group 1	***M. abscessus***	**strain GO 06**	5,07	64,2	98,35	95,23	95,64
		***M. massiliense* T**	5,2	64,2	95,56	96,13	1,00
		M18	4,89	64,2	96,66	96,09	97,57
	M. massiliense	**M154**	4,8	64,1	96,14	95,81	97,26
		M115	4,98	64,1	96,16	95,36	96,92
		M172	5,2	64,2	95,30	94,93	96,17
	M. massiliense	**M47J26**	4,87	64,1	96,23	95,74	96,93
Group 2	**genomespecies**	**M139**	5,05	64,1	95,94	95,64	96,88
		***M. bolletii* T**	5,05	64,2	94,51	1,00	95,33
Group 3	***M. bolletii***	**M24**	5,51	64,2	94,91	96,47	94,20

418

Table 3: *Mycobacterium abscessus* core genome and unique genes.

Organism	CDS	Unique genome	core Total genes
M. abscessus	**-**	**36**	**180**
M. abscessus Type strain	4954	-	-
strain GO 06	4944	-	-
M. abscessus M93	4733	11	11
M. abscessus M94	4841	10	10
M. abscessus M152	4762	-	-
M. massiliense	**-**	**15**	**107**
M. massiliense Type strain	4962	3	3
M. massiliense M18	4663	8	8
M. abscessus 47J26	4766	-	-
M. abscessus M154	4651	-	-
M. abscessus M115	4802	4	4
M. abscessus M139	4754	4	4
M. abscessus M172	5079	20	20
M. bolletii	**-**	**15**	**30**
M. bolletii Type strain	4733	9	9
M. abscessus M24	4960	23	23
***M. abscessus* core genome**	**-**	**3,947**	**57,172**

421 **Table 4:** *Mycobacterium abscessus* prophages.

Groups		Strain	REGION	REGION_LENGTH	CDS	REGION_POSITION
Group 1	*M. abscessus*	***M. abscessus ATCC19977.CIP104536T***	1	81Kb	110	1754551-1835095
		M93	1	16.4Kb	33	197463-213867
			2	38Kb	51	232006-270072
			3	53Kb	70	1762720-1815780
			4	20.2Kb	26	1820768-1841058
		M94	1	58.3Kb	84	1039523-1097850
			2	79.4Kb	99	4959719-5039151
		M152	1	48.9Kb	53	1897722-1946683
			2	34.9Kb	44	4784847-4819818
		Go06	1	65Kb	44	1768158-1833157
Group 2	*M. massiliense*	***M. massiliense BD***	1	12.5Kb	21	1600973-1613514
			2	31.3Kb	33	1620002-1651385
			3	50.4Kb	69	3907205-3957680
		M18	1	62.8Kb	67	4702725-4765592
		M154	0	0	0	0
	***M. massiliense* genomespecies**	***M115***	1	11.6Kb	10	1416841-1428481
			2	77.1Kb	102	1624644-1701770
			3	55.3Kb	79	3356346-3411651
		M172	1	55.1Kb	74	502478-557677
			2	50.7Kb	50	546109-596832
			3	59Kb	67	1934186-1993225
			4	31.1Kb	33	2050376-2081567
			5	39.4Kb	45	3711805-3751246
			6	19.6Kb	40	3753466-3773078
		M47J26	1	39.9Kb	48	1066714-1106668
			2	12.4Kb	16	3596408-3608873
			3	41.4Kb	42	3823414-3864899
		M139	1	35.9Kb	43	2906235-2942215
			2	12.5Kb	17	5033731-5046263

Groups		Strain	REGION	REGION_LENGTH	CDS	REGION_POSITION
Group 3	***M. bolletii***	***M. bolletii BD***	1	41.6Kb	47	1684736-1726377
			2	20.9Kb	38	1727918-1748849
			3	12.4Kb	16	3641720-3654182
		M24	1	37.1Kb	51	560940-598047
			2	37Kb	37	1680197-1717263
			3	17Kb	21	3830340-3847343
			4	18.1Kb	34	5051771-5069955
			5	26Kb	35	5155113-5181190
			6	19.2Kb	26	5213195-5232444
			7	26.5Kb	33	5312024-5338593

422

Figures legends:

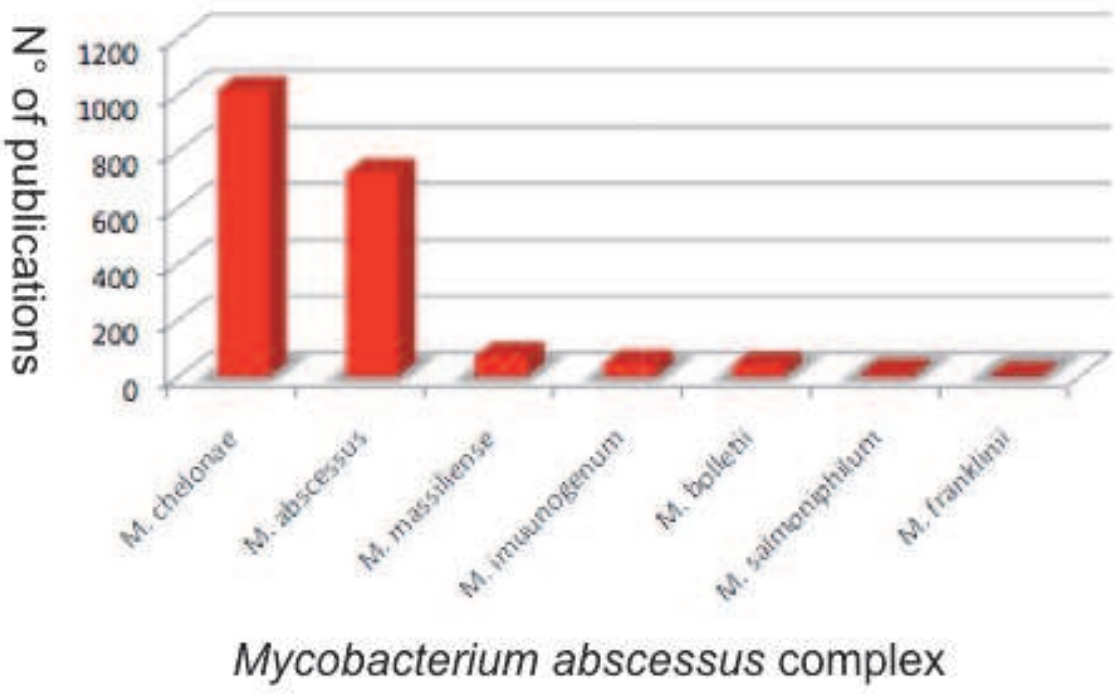


Figure 1. Number of publications retrieved in PubMed for *Mycobacterium abscessus* complex (March 2013).

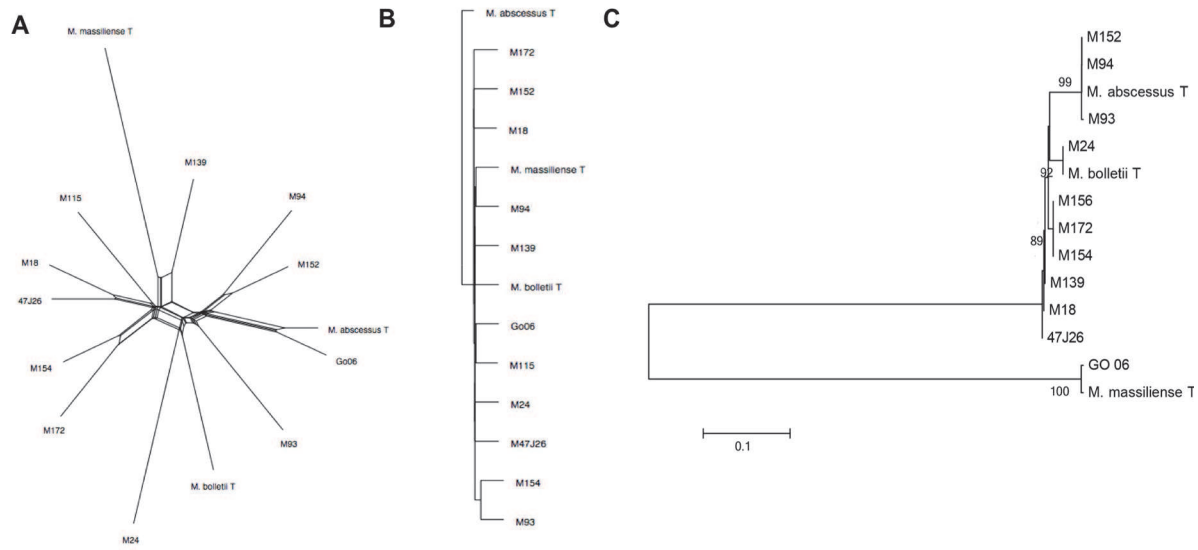


Figure 2. Phylogenomic analysis of *M. abscessus*. A. Aligned whole genomes phylogenetic network. B. Gene content phylogeny constructed from the matrix of discrete characters using the neighbor-joining method. C. RpoB gene based phylogentic tree using neighbor-joining method.

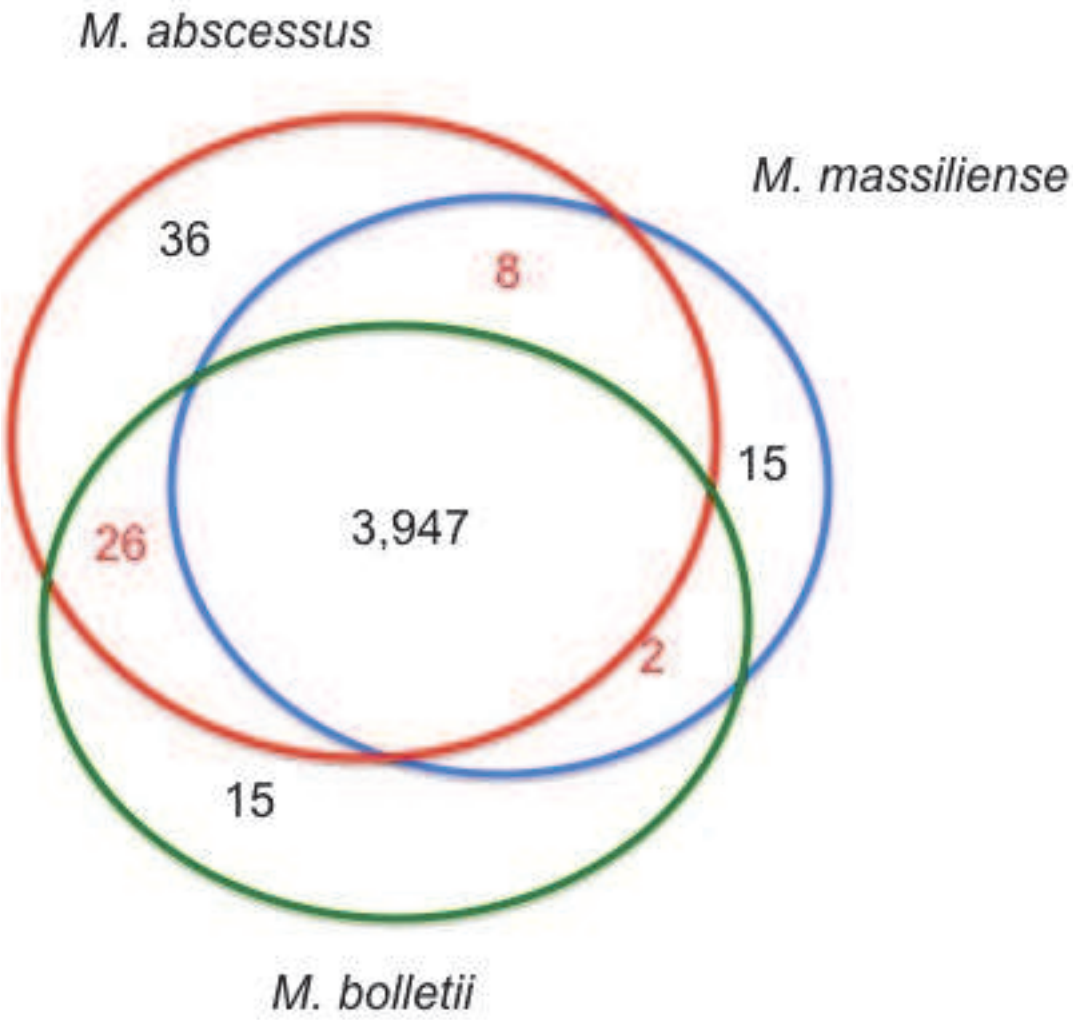


Figure 3. Core genomes in *M. abscessus* clusters.

REFERENCES

1. Kubica GP, Baess I, Gordon RE, Jenkins PA, Kwapinski JB, McDurmont C, Pattyn SR, Saito H, Silcox V, Stanford JL, Takeya K, Tsukamura M: **A co-operative numerical analysis of rapidly growing mycobacteria**. *J Gen Microbiol* 1972, **73:**55–70.

2. Ross AJ: ***Mycobacterium salmoniphilum*** **sp. nov. from salmonid fishes.** *Am Rev Respir Dis* 1960, **81:**241–250.

3. Wilson RW, Steingrube VA, Böttger EC, Springer B, Brown-Elliott BA, Vincent V, Jost KC Jr, Zhang Y, Garcia MJ, Chiu SH, Onyi GO, Rossmoore H, Nash DR, Wallace RJ Jr: ***Mycobacterium immunogenum*** **sp. nov., a novel species related to** ***Mycobacterium abscessus*** **and associated with clinical disease, pseudo-outbreaks, and contaminated metalworking fluids: an international cooperative study on mycobacterial taxonomy.** *Int J Syst Evol Microbiol* 2001, **51:**1751–1764.

4. Adékambi T, Reynaud-Gaubert M, Greub G, Gevaudan MJ, La Scola B, Raoult D, Drancourt M: **Amoebal coculture of "*Mycobacterium massiliense*" sp. nov. from the sputum of a patient with hemoptoic pneumonia**. *J Clin Microbiol* 2004, **42:**5493-5501.

5. Adékambi T, Berger P, Raoult D, Drancourt M: ***rpoB*** **gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of** ***Mycobacterium bolletii*** **sp. nov.,** ***Mycobacterium phocaicum*** **sp. nov. and** ***Mycobacterium aubagnense*** **sp. nov.** *Int J Syst Evol Microbiol* 2006, **56:**133-143.

6. Simmon KE, Brown-Elliott BA, Ridge PG, Durtschi JD, Mann LB, Slechta ES, Steigerwalt AG, Moser BD, Whitney AM, Brown JM, Voelkerding KV, McGowan KL, Reilly AF, Kirn TJ, Butler WR, Edelstein PH, Wallace RJ Jr, Petti CA: ***Mycobacterium chelonae-abscessus*** **complex associated with sinopulmonary disease, Northeastern USA**. *Emerg Infect Dis* 2011, **9:**1692-1700.

7. Petrini B: ***Mycobacterium abscessus*: an emerging rapid-growing potential pathogen**. *APMIS* 2006, **5:**319-328.

8. Bryant JM, Grogono DM, Greaves D, Foweraker Juliet, Roddick Iain, Inns T, Reacher M, Aworth CSH, Curran MD, Harris SR, Peacock SJ, Parkhill J, Floto RA: **Whole-genome sequencing to identify transmission of Mycobacterium abscessus between patients with cystic fibrosis: a retrospective cohort study.** *The Lancet* 2013, 381:1551-1560.

9. Medjahed H, Gaillard JL, Reyrat JM.: ***Mycobacterium abscessus*: a new player in the mycobacterial field.** *Trends Microbiol* 2010, **3:**117-123.

10. Jassies-van der Lee A, Houwers DJ, Meertens N, van der Zanden AG, Willemse T: **Localised pyogranulomatous dermatitis due to *Mycobacterium abscessus* in a cat: a case report.** *Vet J* 2009, **2:**304-306.

11. Albini S, Mueller S, Bornand V, Gutzwiller ME, Burnand C, Hüssy D, Abril C, Reitt K, Korczak BM, Miserez R: **Cutaneous atypical mycobacteriosis due to *Mycobacterium massiliense* in a cat.** (Article in German) *Schweiz Arch Tierheilkd* 2007, **12:**553-558.

12. Clayton LA, Stamper MA, Whitaker BR, Hadfield CA, Simons B, Mankowski JL: ***Mycobacterium abscessus* pneumonia in an Atlantic bottlenose dolphin (*Tursiops truncatus*).** *J Zoo Wildl Med*. 2012, **43**:961-965.

13. Wiinschmann A, Armien A, Harris NB, Brown-Elliott BA, Wallace RJ Jr, Rasmussen J, Willette M, Wolf T: **Disseminated panniculitis in a bottlenose dolphin (*Tursiops truncatus*) due to *Mycobacterium chelonae* infection**. *J Zoo Wildl Med* 2008, **39:**412-440.

14. Clayton LA, Stamper MA, Whitaker BR, Hadfield CA, Simons B, Mankowski JL: ***Mycobacterium abscessus* pneumonia in an Atlantic bottlenose dolphin (*Tursiops truncatus*).** *J Zoo Wildl Med* 2012, **43:**961-965.

15. Zerihun MA, Nilsen H, Hodneland S, Colquhoun DJ: ***Mycobacterium salmoniphilum*** **infection in farmed Atlantic salmon, Salmo salar L.** *J Fish Dis* 2011, **34:**769-781.

16. Leao SC, Tortoli E, Euzéby JP, Garcia MJ: **Proposal that *Mycobacterium massiliense* and *Mycobacterium bolletii* be united and reclassified as *Mycobacterium abscessus* subsp. *bolletii* comb. nov., designation of *Mycobacterium abscessus* subsp. *abscessus* subsp. nov. and emended description of *Mycobacterium abscessus***. *Int J Syst Evol Microbiol* 2011, **61:**2311-2313.

17. Adékambi T, Colson P, Drancourt M. ***rpoB*-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria**. *J Clin Microbiol* 2003, **41:**5699-5708.

18. Viana-Niero C, Lima KV, Lopes ML, Rabello MC, Marsola LR, Brilhante VC, Durham AM, Leao SC: **Molecular characterization of *Mycobacterium massiliense* and *Mycobacterium bolletii* in isolates collected from outbreaks of infections after laparoscopic surgeries and cosmetic procedures.** *J Clin Microbiol* 2008, **46:**850-855.

19. Arnold C, Barrett A, Cross L, Magee JG: **The use of *rpoB* sequence analysis in the differentiation of *Mycobacterium abscessus* and *Mycobacterium chelonae*: a critical judgement in cystic fibrosis?** *Clin Microbiol Infect* 2012, **18:**E131-133.

20. Leao SC, Tortoli E, Viana-Niero C, Ueki SY, Lima KV, Lopes ML, Yubero J, Menendez MC, Garcia MJ: **Characterization of mycobacteria from a major Brazilian outbreak suggests that revision of the taxonomic status of members of the *Mycobacterium chelonae* - *M. abscessus* group is needed.** *J Clin Microbiol* 2009, **47:**2691-2698.

21. Macheras E, Roux AL, Ripoll F, Sivadon-Tardy V, Gutierrez C, Gaillard JL, Heym B: **Inaccuracy of single-target sequencing for discriminating species of the *Mycobacterium abscessus* group.** *J Clin Microbiol* 2009, **47:**2596-2600.

22. Macheras E, Roux AL, Bastian S, Leão SC, Palaci M, Sivadon-Tardy V, Gutierrez C, Richter E, Rusch-Gerdes S, Pfyffer G, Bodmer T, Cambau E, Gaillard JL, Heym B: **Multilocus sequence analysis and *rpoB* sequencing of *Mycobacterium abscessus* (sensu lato) strains.** *J Clin Microbiol* 2011, **49:**491-499.

23. Zelazny AM, Root JM, Shea YR, Colombo RE, Shamputa IC, Stock F, Conlan S, McNulty S, Brown-Elliott BA, Wallace RJ Jr, Olivier KN, Holland SM, Sampaio EP: **Cohort study of molecular identification and typing of *Mycobacterium abscessus*, *Mycobacterium massiliense*, and *Mycobacterium bolletii*.** *J Clin Microbiol* 2009, **47:**1985-1995.

24. Sassi M, Ben Kahla I, Drancourt M: ***Mycobacterium abscessus* multispacer sequence typing.** *BMC Microbiol* 2013, **13:**1-3.

25. Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, Rottman M, Macheras E, Heym B, Herrmann JL, Daffé M, Brosch R, Risler JL, Gaillard JL: **Non-mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*.** *PLoS One* 2009, **4:**e5660.

26. Sentausa E, Fournier PE: **Advantages and limitations of genomics in prokaryotic taxonomy.** *Clin Microbiol Infect* 2013, Mar 13.

27. Darling AE, Mau B, Perna NT: **progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement**. *PLoS One* 2010, **5:**e11147.

28. Bryant D, Moulton V: **Neighbor-net: an agglomerative method for the construction of phylogenetic networks.** *Mol Biol Evol* 2004, **21:**255-265.

29. Raiol T, Ribeiro GM, Maranhão AQ, Bocca AL, Silva-Pereira I, Junqueira-Kipnis AP, Brigido Mde M, Kipnis A**: Complete genome sequence of *Mycobacterium massiliense*.** *J Bacteriol* 2012, **194:**5455.

30. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and translation initiation site identification**. *BMC Bioinformatics* 2010, **11:**119.

31. Li L, Stoeckert CJ Jr, Roos DS: **OrthoMCL: Identification of ortholog groups for eukaryotic genomes.** *Genome Res* 2003, **13:**2178-2189.

32. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4:**41.

33. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H: **Prophage genomics**. Microbiol Mol Biol Rev. 2003 Jun; **67**:238-276.

34. Choi GE, Cho YJ, Koh WJ, Chun J, Cho SN, Shin SJ: **Draft genome sequence of *Mycobacterium abscessus* subsp. *bolletii* BD (T).** *J Bacteriol* 2012, **194:**2756-2757.

35. Wong YL, Choo SW, Tan JL, Ong CS, Ng KP, Ngeow YF: **Draft genome sequence of *Mycobacterium bolletii* strain M24, a rapidly growing mycobacterium of contentious taxonomic status.** *J Bacteriol* 2012, **194:**4475.

36. Sassi M, Bebeacua C, Drancourt M, Cambillau C: **The first structure of a mycobacteriophage, Araucaria**. J Virol. 2013 [Epub ahead of print].

37. Choo SW, Wong YL, Leong ML, Heydari H, Ong CS, Ng KP, Ngeow YF: **Analysis of the genome of *Mycobacterium abscessus* strain M94 reveals an uncommon cluster of tRNAs.** *J Bacteriol* 2012, **194:**5724.

38. Adékambi T, Drancourt M: ***Mycobacterium bolletii* respiratory infections.** *Emerg Infect Dis* 2009, **15:**302-305.

39. Kim HY, Kim BJ, Kook Y, Yun YJ, Shin JH, Kim BJ, Kook YH: ***Mycobacterium massiliense*** **is differentiated from** ***Mycobacterium abscessus*** **and** ***Mycobacterium bolletii*** **by erythromycin ribosome methyltransferase gene (erm) and clarithromycin susceptibility patterns.** *Microbiol Immunol* 2010, **54:**347-353.

40. Koh WJ, Jeon K, Lee NY, Kim BJ, Kook YH, Lee SH, Park YK, Kim CK, Shin SJ, Huitt GA, Daley CL, Kwon OJ: **Clinical significance of differentiation of** ***Mycobacterium massiliense*** **from** ***Mycobacterium abscessus*****.** *Am J Respir Crit Care Med* 2011, **183:**405-410.

41. Paulsen IT, Skurray RA, Tam R, Saier MH Jr, Turner RJ, Weiner JH, Goldberg EB, Grinius LL: **The SMR family: a novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs**. *Mol Microbiol* 1996, **19:**1167-1175.

42. Ramón-García S, Martín C, De Rossi E, Aínsa JA: **Contribution of the Rv2333c efflux pump (the Stp protein) from** ***Mycobacterium tuberculosis*** **to intrinsic antibiotic resistance in** ***Mycobacterium bovis*** **BCG**. *J Antimicrob Chemother* 2007, **59:**544-547.

43. Nessar R, Cambau E, Reyrat JM, Murray A, Gicquel B: ***Mycobacterium abscessus*****: a new antibiotic nightmare.** *J Antimicrob Chemother* 2012, **67:**810-818.

44. Hu GZ, Pan YS, Wu H, Hu H, Xu R, Yuan L, Liu JH, Feng JK: **Prevalence of tetracycline resistance genes and identification of tet(M) in clinical isolates of** ***E. coli*** **from sick ducks in China.** *J Med Microbiol* 2013, Mar 8.

45. Duarte RS, Lourenço MC, Fonseca Lde S, Leão SC, Amorim Ede L, Rocha IL, Coelho FS, Viana-Niero C, Gomes KM, da Silva MG, Lorena NS, Pitombo MB, Ferreira RM, Garcia MH, de Oliveira GP, Lupi O, Vilaça BR, Serradas LR, Chebabo A, Marques EA, Teixeira LM, Dalcolmo M, Senna SG, Sampaio JL: **Epidemic of postsurgical infections caused by** ***Mycobacterium massiliense*****.** *J Clin Microbiol* 2009, **47:**2149-2155.

46. Maurer FP, Rüegger V, Ritter C, Bloemberg GV, Böttger EC: **Acquisition of clarithromycin resistance mutations in the 23S rRNA gene of *Mycobacterium abscessus* in the presence of inducible erm(41).** *J Antimicrob Chemother* 2012, **67:**2606-2611.

47. Monego F, Duarte RS, Nakatani SM, Araújo WN, Riediger IN, Brockelt S, Souza V, Cataldo JI, Dias RC, Biondo AW: **Molecular identification and typing of *Mycobacterium massiliense* isolated from postsurgical infections in Brazil.** *Braz J Infect Dis* 2011, **15:**436-441.

48. de Moura VC, da Silva MG, Gomes KM, Coelho FS, Sampaio JL, Mello FC, Lourenço MC, Amorim Ede L, Duarte RS: **Phenotypic and molecular characterization of quinolone resistance in *Mycobacterium abscessus* subsp. *bolletii* recovered from postsurgical infections.** *J Med Microbiol* 2012, **61:**115-125.

49. Cattoir V, Poirel L, Nordmann P: **Plasmid-mediated quinolone resistance pump QepA2 in an *Escherichia coli* isolate from France.** *Antimicrob Agents Chemother* 2008, **52:**3801-3804.

50. Basu J, Chattopadhyay R, Kundu M, Chakrabarti P: **Purification and partial characterization of a penicillin-binding protein from *Mycobacterium smegmatis***. *J Bacteriol* 1992, **174:**4829-4832.

51. Cayrou C, Henrissat B, Gouret P, Pontarotti P, Drancourt M: **Peptidoglycan: a post-genomic analysis.** *BMC Microbiol* 2012, **18:**294.

52. Catherinot E, Clarissou J, Etienne G, Ripoll F, Emile JF, Daffé M, Perronne C, Soudais C, Gaillard JL, Rottman M: **Hypervirulence of a rough variant of the *Mycobacterium abscessus* type strain.** *Infect Immun* 2007, **75:**1055–1058.

53. Nessar R, Reyrat JM, Davidson LB, Byrd TF: **Deletion of the mmpL4b gene in the *Mycobacterium abscessus* glycopeptidolipid biosynthetic pathway results in loss of surface colonization capability, but enhanced ability to replicate in human macrophages and stimulate their innate immune response.** *Microbiology* 2011, **157:**1187-1195.

54. Davidson LB, Nessar R, Kempaiah P, Perkins DJ, Byrd TF: ***Mycobacterium abscessus* glycopeptidolipid prevents respiratory epithelial TLR2 signaling as measured by HβD2 gene expression and IL-8 release**. *PLoS One* 2011, **6:**e29148.

55. Recht J, Martínez A, Torello S, Kolter R: **Genetic analysis of sliding motility in *Mycobacterium smegmatis*.** *J Bacteriol* 2000, **182:**4348–4351.

56. Recht J, Kolter R: **Glycopeptidolipid acetylation affects sliding motility and biofilm formation in *Mycobacterium smegmatis***. *J Bacteriol* 2001, **183:**5718–5724.

57. Villeneuve C, Etienne G, Abadie V, Montrozier H, Bordier C, Laval F, Daffe M, Maridonneau-Parini I, Astarie-Dequeker C: **Surface-exposed glycopeptidolipids of *Mycobacterium smegmatis* specifically inhibit the phagocytosis of mycobacteria by human macrophages. Identification of a novel family of glycopeptidolipids.** *J Biol Chem* 2003, **278:**5129–5130.

58. El-Shazly S, Ahmad S, Mustafa AS, Al-Attiyah R, Krajci D: **Internalization by HeLa cells of latex beads coated with mammalian cell entry (Mce) proteins encoded by the mce3 operon of *Mycobacterium tuberculosis*.** *J Med Microbiol* 2007, **56:**1145-1151.

59. Stavrum R, Valvatne H, Stavrum AK, Riley LW, Ulvestad E, Jonassen I, Doherty TM, Grewal HM: ***Mycobacterium tuberculosis* Mce1 protein complex initiates rapid induction of transcription of genes involved in substrate trafficking.** *Genes Immun* 2012, **13:**496-502.

60. Ishikawa J, Yamashita A, Mikami Y, Hoshino Y, Kurita H, Hotta K, Shiba T, Hattori M: **The complete genomic sequence of *Nocardia farcinica* IFM 10152.** *Proc Natl Acad Sci U S A* 2004, **101:**14925–14930.

61. Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, Appelmelk BJ, Bitter W: **Type VII secretion-mycobacteria show the way.** *Nat Rev Microbiol* 2007, **5:**88-91.

62. Simeone R, Bottai D, Brosch R: **ESX/type VII secretion systems and their role in host-pathogen interaction.** *Curr Opin Microbiol* 2009, **12:**4-10.

63. Mukhopadhyay S, Balaji KN: **The PE and PPE proteins of *Mycobacterium tuberculosis*.** *Tuberculosis (Edinb)* 2011, **91:**441-447.

64. Snel B, Bork P, Huynen MA: **Genome phylogeny based on gene content**. Nat Genet. 1999, **21**:108–110.

II- L'IDENTIFICATION DE *MYCOBACTERIUM ABSCESSUS*

Commentaire

M. abscessus sensu lato comprend des mycobactéries actuellement classées en *M. abscessus* subsp. *abscessus* et *M. abscessus* subsp. *bolletii* [4]. Ce dernier a été précédemment identifié comme « *M. bolletii*» ou «*M. massiliense* » [5, 6]. Cependant, ces organismes ne sont pas distingués en routine à l'aide des tests phénotypiques, y compris l'analyse de la composition en acides mycoliques et ils partagent 100% de similarité de séquence du gène de l'ARNr 16S. Ces organismes ont été initialement différenciés sur la base de > 3% de divergence de la séquence partielle du gène *rpo*B et de profils différents de sensibilité aux agents anti-microbiens [34]. Ensuite, des résultats contradictoires basés sur le séquençage *rpo*B ont été rapportés [35] et le séquençage combiné des gènes *rpo*B, *hsp*65 et *sec*A a été préconisé pour l'identification de *M. abscessus* [4,36, 37]. Dans la première partie de ce travail, nous avons montré que *M. abscessus sensu lato* est composé d'au moins cinq taxons différents spécifiés par des caractéristiques microbiologiques d'intérêt médical. Dans cette partie notre objectif a été de développer une méthode d'identification et génotypage permettant la discrimination rapide et efficace au laboratoire de *M. abscessus sensu lato*. En utilisant la souche type *M. abscessus* subsp *abscessus* CIP104536^T comme référence, nous avons séquencé huit régions intergéniques de 21 isolats cliniques et 48 génomes disponibles. Nous avons comparé notre méthode Multiple Spacer sequence Typing (MST) avec une méthode basée sur l'analyse des séquences des gènes de ménages

(MLSA) *arg*H (arginino- succinate lyase), *cya* (adenylate cyclase), *mur*C (UDP N-acetylmuramate-L-Ala ligase, *pta* (phosphate acetyl-transferase) and *pur*H (phoshoribosylminoimiazolcarboxy- lase ATPase subunit) [38]. Les résultats des deux méthodes MST et MLSA montrent que 37 organismes de *M. abscessus* sont regroupés en 12 et neuf types, respectivement, quatre *M. bolletii* et *M. abscessus* M139 sont regroupés en trois et quatre types, respectivement, et 27 *M. massiliense* sont regroupés en neuf et cinq types, respectivement. L'indice Hunter-Gaston était de 0,912 pour MST et de 0,903 pour MLSA. L'arbre basé sur les huit séquences concaténées de régions intergéniques est similaire à celui basé sur MLSA et le séquençage du gène *rpo*B. Nous avons observé trois groupes principaux qui comprennent chacun la souche de type des sous-espèces de *M. abscessus* respectifs. Deux isolats présentaient une discordance entre l'arbre *rpo*B et l'arbre basé sur MLSA, un isolat présentait une discordance entre MST et *rpo*B et un isolat présentait une discordance entre MST et MLSA. Le seul séquençage de la région intergénique n° 2 a permis l'identification précise des différents isolats au niveau des sous-espèces. La méthode MST est une nouvelle approche basée sur le séquençage permettant à la fois l'identification et le génotypage de *M. abscessus* et différenciant clairement *M. massiliense* et *M. bolletii*. Ce travail a fait l'objet d'un article de notre premier chapitre publié dans **BMC microbiology** ayant un impact factor de 3,10.

Sassi *et al. BMC Microbiology* 2013, **13**:3
http://www.biomedcentral.com/1471-2180/13/3

BMC Microbiology

RESEARCH ARTICLE Open Access

Mycobacterium abscessus multispacer sequence typing

Mohamed Sassi[1], Imen Ben Kahla[2] and Michel Drancourt[1*]

Article N.2 : **Sassi M,** Ben Kahla I and Drancourt M.

Mycobacterium abscessus multispacer sequence typing.

BMC Microbiol. 2013, 13:3.

Sassi et al. BMC Microbiology 2013, 13:3
http://www.biomedcentral.com/1471-2180/13/3

BMC Microbiology

RESEARCH ARTICLE Open Access

Mycobacterium abscessus multispacer sequence typing

Mohamed Sassi[1], Imen Ben Kahla[2] and Michel Drancourt[1*]

Abstract

Background: *Mycobacterium abscessus* group includes antibiotic-resistant, opportunistic mycobacteria that are responsible for sporadic cases and outbreaks of cutaneous, pulmonary and disseminated infections. However, because of their close genetic relationships, accurate discrimination between the various strains of these mycobacteria remains difficult. In this report, we describe the development of a multispacer sequence typing (MST) analysis for the simultaneous identification and typing of *M. abscessus* mycobacteria. We also compared MST with the reference multilocus sequence analysis (MLSA) typing method.

Results: Based on the *M. abscessus* CIP104536^T genome, eight intergenic spacers were selected, PCR amplified and sequenced in 21 *M. abscessus* isolates and analysed in 48 available *M. abscessus* genomes. MST and MLSA grouped 37 *M. abscessus* organisms into 12 and nine types, respectively; four formerly "*M. bolletii*" organisms and *M. abscessus* M139 into three and four types, respectively; and 27 formerly "*M. massiliense*" organisms grouped into nine and five types, respectively. The Hunter-Gaston index was off 0.912 for MST and of 0.903 for MLSA. The MST-derived tree was similar to that based on MLSA and *rpoB* gene sequencing and yielded three main clusters comprising each the type strain of the respective *M. abscessus* sub-species. Two isolates exhibited discordant MLSA- and *rpoB* gene sequence-derived position, one isolate exhibited discordant MST- and *rpoB* gene sequence-derived position and one isolate exhibited discordant MST- and MLSA-derived position. MST spacer n°2 sequencing alone allowed for the accurate identification of the different isolates at the sub-species level.

Conclusions: MST is a new sequencing-based approach for both identifying and genotyping *M. abscessus* mycobacteria that clearly differentiates formerly "*M. massiliense*" organisms from other *M. abscessus* subsp. *bolletii* organisms.

Keywords: *Mycobacterium*, *Mycobacterium abscessus*, *Mycobacterium massiliense*, *Mycobacterium bolletii*, Multispacer sequence typing, Genotyping

Background

Mycobacterium abscessus mycobacteria are increasingly being cultured from respiratory tract specimens collected from patients with chronic pulmonary diseases, including cystic fibrosis [1-9]. These mycobacteria are also responsible for skin and soft-tissue infections following surgical and cosmetic practices [10-12] and catheter-related bacteremia [13,14]. These infections are particularly critical for immune-compromised patients and may be fatal [15]. Water is suspected as a source of infection, as *M. abscessus* mycobacteria have been isolated from tap water [16]. Moreover, *M. abscessus* mycobacteria have been shown to be resistant to water-borne free-living amoebae [17,18]. *M. abscessus* infections are also associated with treatment failure owing, due to the natural broad-spectrum resistance to antibiotics in addition to acquired resistance, with subtle differences in the antibiotic susceptibility pattern being observed among isolates [19].

Indeed, *M. abscessus* is comprised of a heterogeneous group of mycobacteria currently classified into *M. abscessus* subsp. *abscessus* and *M. abscessus* subsp. *bolletii* [20,21], with the later subspecies accommodating mycobacteria previously identified as "*Mycobacterium bolletii*" or "*Mycobacterium massiliense*" [18,22]. However, these organisms are nearly indistinguishable using phenotypic tests including the mycolic acid pattern

* Correspondence: Michel. Drancourt@univmed.fr
[1]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UMR CNRS 7278, IRD 198, INSERM 1095. Faculté de médecine, 27, Boulevard Jean Moulin-Cedex 5, Marseille, France
Full list of author information is available at the end of the article

analysis and share 100% 16S rRNA gene sequence similarity [20]. They were initially differentiated on the basis of >3% *rpoB* gene sequence divergence and different antimicrobial susceptibility patterns [23,24]. Nevertheless, confusing results based on *rpo*B sequencing have been reported [21], and combining sequencing of the *rpo*B, *hsp65* and *secA* genes has been advocated for the optimal identification of the *M. abscessus* mycobacteria [25].

To further decrypt the diversity and genetic relationships among *M. abscessus* organisms, we investigated a collection of reference, sequenced genomes and clinical *M. abscessus* isolates using multispacer sequence typing (MST), which is a sequencing-based approach previously used for the species identification and genotyping of Mycobacteria, including *Mycobacterium avium* [26] and *Mycobacterium tuberculosis* [27] and non-mycobacterial pathogens, such as *Yersinia pestis* [28], *Rickettsia prowazekii* [29] and *Bartonella quintana* [30]. This approach was here compared with multilocus sequence analaysis which relies the sequencing of 5–8 genes (21, 25), and *rpoB* genes sequencing (23, 24).

Methods

Bacterial isolates

Reference *M. abscessus* CIP104536^T, *M. abscessus* DSMZ44567 (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany), *M. abscessus* subsp. *bolletii* CIP108541^T (herein referred as "*M. bolletii*") and *M. abscessus* subsp. *bolletii* CIP108297^T (herein referred as "*M. massiliense*" [23]) were used in this study. In addition, a collection of 17 *M. abscessus* clinical isolates from the mycobacteria reference laboratory of the Méditerranée Infection Institute, Marseille, France were also studied (Table 1). All of the mycobacteria were grown in 7H9 broth (Difco, Bordeaux, France) enriched with 10% OADC (oleic acid, bovine serum albumin, dextrose and catalase) at 37°C. As for the identification, DNA extraction and *rpo*B partial sequence-based identification were performed using the primers MYCOF and MYCOR2 (Table 1) as previously described [24]. In addition, the *rpoB* gene sequence retrieved from 48 *M. abscessus* sequenced genomes was also analysed (Additional file 1) (http://www.ncbi.nlm.nih.gov/).

Reference MLSA typing

Fragments from five housekeeping genes *argH* (argininosuccinate lyase), *cya* (adenylate cyclase), *murC* (UDP N-acetylmuramate-L-Ala ligase, *pta* (phosphate acetyltransferase) and *purH* (phoshoribosylminoimiazolcarboxylase ATPase subunit) were amplified using the sets of primers as previously described (21). The sequences of each one of these five housekeeping genes retrieved from 48 *M. abscessus* sequenced genomes, were also included in the MLSA analysis (Additional file 1).

MST analysis

Sequences of the whole intergenic spacers were extracted from the reference *M. abscessus* CIP104536^T (ATCC19977) genome (GenBank accession CU458896.1) using the perl script software and a total of 8 spacers with a 200-700-bp sequence size were further used in analysis. For each of these 8 spacers, specific PCR primers were designed using Primer3 software v 0.4.0 (http://frodo.wi.mit.edu/primer3) and tested *in silico* for specificity using BLAST software (http://www.ncbi.nlm.nih.gov). The PCR conditions were first optimized using DNA extracted from the reference *M. abscessus*, *"M. bolletii"* and *"M. massiliense"* isolates before analysis of DNA extracted from the 17 clinical isolates (Table 1). The PCR amplifications were performed in a 50 μl PCR mixture containing 5 μl 10x buffer (Qiagen, Courtaboeuf, France), 200 mM each dNTP, 1.5 mM $MgCl_2$, 1.25 U HotStarTaq polymerase (Qiagen), 1 μl each primer (10 pM), 33 μl nuclease-free water and 5 μl DNA template. The amplification program consisted of an initial 15 min denaturation step at 95°C followed by 40 cycles of 30 s at 95°C, 30 s at 60°C and 1 min at 72°C; the amplification was completed by a final 5-min elongation step at 72°C. Negative controls consisting of PCR mixture without DNA template were included in each PCR run. The products were visualized by gel electrophoresis, purified using a MultiScreen PCR filter plate (Millipore, Molsheim, France) and sequenced in both directions using the BigDye Terminator sequencing kit (Applied Biosystems, Villebon-sur-Yvette, France), as previously described [27]. The sequences were edited using the ChromasPro software (version 1.42; Technelysium Pty Ltd), aligned using Clustal W (MEGA 5 software) and compared with the reference *M. abscessus* ATCC 19977 sequences (GenBank accession CU458896.1). For MST and MLSA discrimination power was calculated using the Hunter-Gaston Index [31]:

$$DI = 1 - \left[\frac{1}{N(N-1)}\sum_{j=1}^{8} n_j(n_j-1)\right]$$

where D is the numerical index of discrimination, N is the total number of isolates in the sample population, s is the total number of different types and nj is the number of isolates belonging to the jth type.

Phylogenetic analysis

Phylogenetic trees were constructed based on *rpoB* gene, concatenated MLSA genes, concatenated spacers and MST spacer n°2 sequences using the neighbor–joining method with Kimura's two-parameter (K2P) distance correction model with 1000 bootstrap replications in the MEGA version 5 software package [32]. The *rpo*B gene sequence-based tree was rooted using *M. chelonae* strain

Table 1 Spacers characteristics used in this study

Name	Genome position*	Framing genes*	PCR primers	PCR product size (bp)
Spacer 1	106145-106396	MAB_0104:enoyl-CoA hydratase/isomerise	F : GGGATGCGCAGATGACGGGG	506
		MAB_0105c:oxidoreductase	R : GCTACCCCGAATGGGGCACG	
Spacer 2	173727-173985	MAB_0176:antigen 85-A precursor	F : TCGAGTTTCCTCCGGGCGGT	438
		MAB_0177:antigen 85-A/B/C precursor	R: AATCCAGGCAGAACGGCCGC	
Spacer 3	422777-423027	MAB_0423c:hypothetical protein	F: GCCATTGCTGTCCGTGCGGT	344
		MAB_0424:putative protease	R : GCCGCGAACAGGCCAAACAG	
Spacer 4	494411-494670	MAB_0495c:hypothetical protein	F: CGCCCTTGCGCAGGAGTGAT	528
		MAB_0496c:hypothetical protein	R: GCCTGGTTCGGACGGTGACG	
Spacer 5	761805-762060	MAB_0761c:putative 3-hydroxyacyl-CoA dehydrogenase	F : ACCACATCGGCGAGCGTGTG	545
		MAB_0762:hypothetical protein	R : CCAACACCGGGTCGCGGTAC	
Spacer 6	771170-771436	MAB_0772c:hypothetical protein	F : CGTCGGTCTTGCCGACCGTC	600
		MAB_0773:hypothetical protein	R : GGCGCCGACGATCTAGCACC	
Spacer 7	880381-880639	MAB_0887c:hypothetical protein	F: CGGCAGTGCAAGGTGCGTTG	519
		MAB_0888c:putative fumarylacetoacetase	R : GCACCGTGTCCGGTCCTCAG	
Spacer 8	959422-959678	MAB_0950c:putative amino acid permease family protein	F: GGGGCGTATGCGCCGTTACC	474
		MAB_0951:putative aminoglycoside phosphotransferase	R : CGAACGCGCTGTGATTCGGC	
Spacer 9	1002935-1003200	MAB_0995:hypothetical protein	F : GGCCGCGACAAGCTGATCGT	684
		MAB_0997c:hypothetical protein	R: ATGCAGGGCACCGTGCGTAG	
Spacer 10	1216613-1216879	MAB_1201c:transcription elongation factor GreA	F: CGTTCTCGCGCAGGTCTCCC	517
		MAB_1202c:hypothetical protein	R: CCGAACGATCCGTGCCGGTC	
Spacer 11	1818877-1819188	MAB_1818:hypothetical protein	F: AGCCAACTGCCATGGCGCTT	495
		MAB_1819c:hypothetical protein	R : ACCGAGACGTCATGCACCGC	

* With reference to *M. abscessus* ATCC 19977 genome.

CIP 104535^T and *M. immunogenum* strain CIP 106684^T *rpo*B gene sequences. A heatmap was constructed using the R statistical software based on the spacer profile as a distance matrix.

Results and discussion

rpoB identification and rpoB tree

The identification of *M. abscessus* CIP104536^T, *M. abscessus* DSMZ44567, *M. bolletii* CIP108541^T and *M. massiliense* CIP108297^T was confirmed by partial *rpo*B sequencing. The sequences were deposited in the GenBank database (GenBank accession: KC352778 - KC352795). Isolates P1, P2.1, P2.2, P2.3, P2.4, P2.5, P3.1, P3.2, P4, P5, P6, P7 and P8 exhibited 99% *rpo*B sequence similarity with *M. abscessus* ATCC19977^T and were identified as *M. abscessus*. Isolates P9 and P10 exhibited 99% *rpo*B sequence similarity with *"M. bolletii"* CIP108541^T and were identified as *"M. bolletii"* whereas isolate P11 exhibited 99% *rpo*B sequence similarity with *"M. massiliense"* CIP108297^T and was identified as *"M. massiliense"*. A total of 23 *M. abscessus* sequenced genomes were identified as *M. abscessus* since they exhibited 98 to 100% similarity with the *M. abscessus* type strain *rpoB* partial gene sequence. *M. abscessus* M24 shared 99% similarity with the *M. bolletii* type strain partial *rpoB* gene sequence. A total of 26 *M. abscessus* and *"M. massiliense"* sequenced genomes shared 99% to 100% similarity with *"M. massiliense"* partial *rpoB* gene sequence. The tree built from 69 partial *rpoB* gene sequences showed three distinct groups, each comprising the type strain (Figure 1a).

Reference MLSA analysis

Fragments for the expected size were amplified and sequenced for the five MLSA genes. The sequences were deposited in the GenBank database (GenBank accession: KC352742 - KC352759, KC352760 - KC352777, KC352796 - KC352813, KC352814 - KC352831, KC352832 - KC352849). Concatenation of the five sequences yielded a total of 19 different types, including 9 types for 37 *M. abscessus* organisms, four types for 4 *"M. bolletii"* organisms and *M. abscessus* M139 and five types for 27 *"M. massiliense"* organisms. The Hunter-Gaston Index for MLSA was of 0.903. The MLSA tree based on the five gene concatened sequences showed three principal clusters, i.e. a *M. abscessus* cluster, a *"M. bolletii"* cluster and a *"M. massiliense"* cluster (Figure 1b). Latter cluster comprised of five sub-clusters with *"M. massiliense"* type strain and P11 strain sub-clustering together close to *M. abscessus* 5S strain. Also,

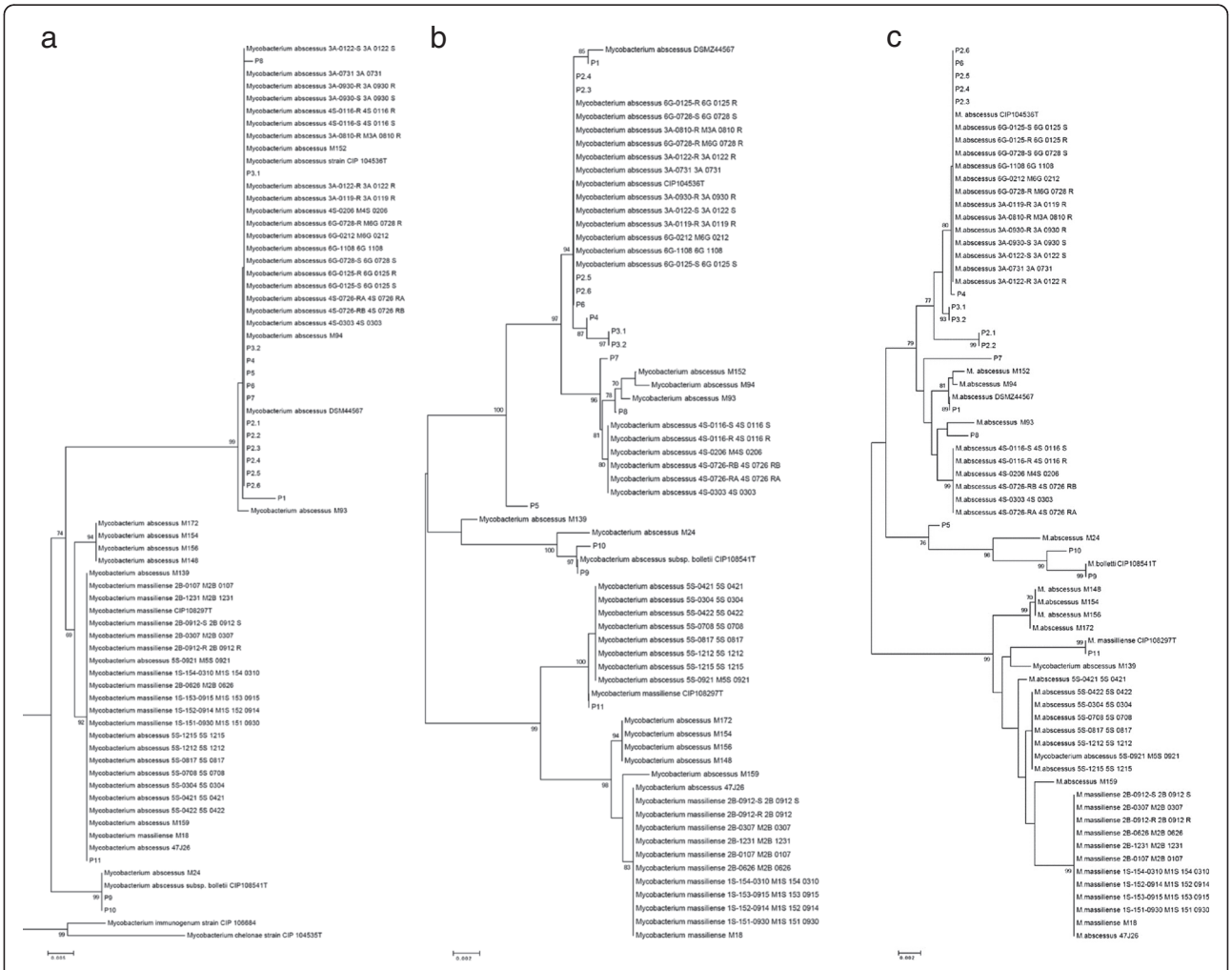


Figure 1 Phylogenetic tree based on *rpoB* gene sequence (a); based on the concatenated five MLSA gene sequences (b); and based on the concatenated eight polymorphic spacers (c).

MLSA-derived tree clustered *M. abscessus* M139 strain and P5 strain respectively identified as "*M. massiliense*", close to the "*M. bolletii*" whereas both strains clustered with *M. abscessus* in the *rpoB* gene sequence-derived tree.

MST analysis

Analysis of the reference *M. abscessus* ATCC 19977 complete genome sequence yielded 3538 intergenic spacers with > 300 spacers were 200–700 bp in length. Successful PCR sequencing was achieved for 8 spacers in all the isolates studied; the sequences were deposited in the GenBank database (GenBank accession: KC352850 - KC352890). In *M. abscessus* isolates, including the 37 sequenced genomes, the spacer sequence variability was generated by one to 12 single nucleotide polymorphisms (SNPs) (spacers n°1 and n°8), one to 18 SNPs and one to two nucleotide deletions (spacer n°2), one to two SNPs (spacers n°3 and n°7) and nucleotide insertion (spacers n°2 and n°5). In "*M. bolletii*" isolates, the spacer sequence polymorphisms were generated by one SNP for spacer n°1, two SNPs and one deletion for spacer n°2, two SNPs for spacer n°3 and nine SNPs for spacer n°7. In "*M. massiliense*" isolates, including 28 sequenced genomes, the spacer sequence polymorphism were generated by nine SNPs and one insertion (spacer n°1), one insertion (spacer n°3), five SNPs and two insertions (spacer n°4), one SNP (spacer n°5) and two SNPs (spacer n°7). Concatenation of the eight spacer sequences yielded a total of 24 types, with the 37 *M. abscessus* organisms grouped into 12 spacer types, four formerly "*M. bolletii*" organisms grouped into three spacer types and 28 formerly "*M. massiliense*" organisms grouped into nine spacer types. This yielded a Hunger-Gaston Index of 0.912. Spacer n°5 was found to be the most variable of the eight spacers under study, exhibiting 13 different alleles (Table 2). When combining the eight spacer sequences, a unique MST profile for each reference isolate was obtained, i.e., MST1 and MST2 for

Table 2 Spacers allelic polymorphism and MST [a] genotypes of *M. abscessus*, "*M. bolletii*" and "*M. massiliense*" isolates

Isolates	Spacer1	Spacer2	Spacer3	Spacer4	Spacer5	Spacer6	Spacer7	Spacer8	Genotype
***M.abscessus*_ATCC19977_CIP104536T**	1	1	1	1	1	1	1	1	**1**
***M.abscessus*_DSMZ44567**	2	1	2	2	2	1	2	1	**2**
P1	2	1	2	2	2	1	2	1	**2**
P2.1	1	2	1	3	1	1	2	2	**3**
P2.2	1	2	1	3	1	1	2	2	**3**
P2.3	1	1	1	1	1	1	1	1	**1**
P2.4	1	1	1	1	1	1	1	1	**1**
P2.5	1	1	1	1	1	1	1	1	**1**
P2.6	1	1	1	1	1	1	1	1	**1**
P3.1	3	1	2	1	1	1	2	1	**4**
P3.2	3	1	2	1	1	1	2	1	**4**
P4	1	1	1	1	1	1	1	2	**5**
P5	1	1	1	1	3	1	2	1	**6**
P6	1	1	1	1	1	1	1	1	**1**
P7	4	1	2	4	4	1	2	1	**7**
P8	4	1	2	4	4	1	3	1	**8**
M.abscessus_3A-0930-R_3A_0930_R	1	1	1	1	1	1	1	1	**1**
M.abscessus_3A-0930-S_3A_0930_S	1	1	1	1	1	1	1	1	**1**
M.abscessus_3A-0122-S_3A_0122_S	1	1	1	1	1	1	1	1	**1**
M.abscessus_3A-0731_3A_0731	1	1	1	1	1	1	1	1	**1**
M.abscessus_3A-0122-R_3A_0122_R	1	1	1	1	1	1	1	1	**1**
M.abscessus_3A-0119-R_3A_0119_R	1	1	1	1	1	1	1	1	**1**
M.abscessus_6G-0728-R_M6G_0728_R	1	1	1	1	1	1	1	1	**1**
M.abscessus_6G-0212_M6G_0212	1	1	1	1	1	1	1	1	**1**
M.abscessus _6G-1108_6G_1108	1	1	1	1	1	1	1	1	**1**
M.abscessus _6G-0728-S_6G_0728_S	1	1	1	1	1	1	1	1	**1**
M.abscessus_6G-0125-R_6G_0125_R	1	1	1	1	1	1	1	1	**1**
M.abscessus _6G-0125-S_6G_0125_S	1	1	1	1	1	1	1	1	**1**
M.abscessus_4S-0116-S_4S_0116_S	5	1	2	5	5	2	2	2	**9**
M.abscessus_4S-0116-R_4S_0116_R	5	1	2	5	5	2	2	2	**9**
M.abscessus_4S-0206_M4S_0206	5	1	2	5	5	2	2	2	**9**
M.abscessus_4S-0726-RB_4S_0726_RB	5	1	2	5	5	2	2	2	**9**
M.abscessus_4S-0303_4S_0303	5	1	2	5	5	2	2	2	**9**
M.abscessus_4S-0726-RA_4S_0726_RA	5	1	2	5	5	2	2	2	**9**
M.abscessus_M93	3	1	2	6	6	1	2	3	**10**
M.abscessus_M94	2	1	2	2	7	1	4	2	**11**
M.abscessus_M152	2	1	2	7	7	1	2	3	**12**
***M.bolletti*_CIP108541T**	6	3	3	3	8	1	5	2	**13**
P9	6	3	3	3	8	1	5	2	**13**
P10	7	4	1	3	8	1	2	2	**14**
M.abscessus_M24	8	3	4	8	8	1	2	2	**15**
***M.massilliense*_CIP108297T**	5	5	5	9	9	1	6	3	**16**
P11	5	5	5	9	9	1	6	3	**16**
M.massiliense _2B-0912-S_2B_0912_S	9	5	6	10	10	2	7	3	**17**
M.massiliense_2B-030_ M2B_0307	9	5	6	10	10	2	7	3	**17**
M.massiliense_2B-0912-R_2B_0912_R	9	5	6	10	10	2	7	3	**17**
M.massiliense_2B-0626_M2B_0626	9	5	6	10	10	2	7	3	**17**

Table 2 Spacers allelic polymorphism and MST [a] genotypes of *M. abscessus*, "*M. bolletii*" and "*M. massiliense*" isolates *(Continued)*

M.massiliense_2B-1231_M2B_1231	9	5	6	10	10	2	7	3	**17**
M.massiliense_2B-0107_M2B_0107	9	5	6	10	10	2	7	3	**17**
M.massiliense _1S-154-0310_M1S_154_0310	9	5	6	10	10	2	7	3	**17**
M.massiliense_1S-152-0914_M1S_152_0914	9	5	6	10	10	2	7	3	**17**
M.massiliense_1S-153-0915_M1S_153_0915	9	5	6	10	10	2	7	3	**17**
M.massiliense_1S-151-0930_M1S_151_0930	9	5	6	10	10	2	7	3	**17**
M.massiliense _M18	9	5	6	10	10	2	7	3	**17**
M.abscessus_M159	9	6	6	9	10	3	7	4	**18**
M.abscessus_47J26	9	5	6	6	11	4	7	3	**19**
M.abscessus_M172	10	7	2	9	12	3	8	5	**20**
M.abscessus_M154	10	7	2	9	12	3	8	5	**20**
M.abscessus_5S-1215_5S_1215	11	5	2	6	13	2	6	2	**21**
M.abscessus_5S-1212_5S_1212	11	5	2	6	13	2	6	2	**21**
M.abscessus_5S-0817_5S_0817	11	5	2	6	13	2	6	2	**21**
M.abscessus_5S-0708_5S_0708	11	5	2	6	13	2	6	2	**21**
M.abscessus_5S-0422_5S_0422	11	5	2	6	13	2	6	2	**21**
M.abscessus_5S-0304_5S_0304	11	5	2	6	13	2	6	2	**21**
M.abscessus_5S-0421_5S_0421	11	5	2	6	13	2	6	2	**21**
M.abscessus_M156	10	7	2	11	12	3	9	5	**22**
M.abscessus_M148	10	7	2	11	12	3	9	5	**23**
M.abscessus_M139	10	5	2	11	14	3	10	3	**24**
DI	**0.8295**	**0.6228**	**0.6969**	**0.8001**	**0.8371**	**0.6038**	**0.8084**	**0.7158**	**0.912**

[a] *MST* = Multispacer Sequence Typing. [b] isolates were listed with reference to their corresponding patient, for example *P1* = isolate 1 from patient 1, *P2.1* = isolate 1 from patient 2, etc. [c] *DI* = Discrimination index.

M. abscessus CIP104536^T and *M. abscessus* DSMZ44567 respectively, MST13 for "*M. bolletii*" CIP108541^T and MST16 for "*M. massiliense*" CIP108297^T. At the sequence level, we found that MST1 and MST2 genotypes differ by at most nine SNPs, whereas MST1 differed from MST13 by up to 18 SNPs, one insertion and two deletions and from MST16 by 14 SNPs, 11 deletions and two insertions (supplementary material). The 17 clinical *M. abscessus* isolates were grouped into eight MST types, named MST1 to MST8, with five *M. abscessus* isolates exhibiting the *M. abscessus* CIP104536^T MST1 genotype and one isolate (P1 strain) exhibiting the *M. abscessus* DSMZ44567 MST2 genotype. The P9 "*M. bolletii*" clinical isolate yielded the MST13 genotype in common with the reference "*M. bolletii*" CIP108541^T, whereas the P10 "*M. bolletii*" clinical isolate yielded a unique MST14 genotype that differ from MST13 by two SNPs in spacer n°1. *M. abscessus* M24 yielded the MST15 and differed from MST13 by four polymorphic spacers. In "*M. massiliense*" nine different profiles were generated MST 16 to MST24. The P11 "*M. massiliense*" clinical isolate shared the MST16 genotype with the reference "*M. massiliense*" CIP108297^T. "*M. massiliense*" 2B isolate, "*M. massiliense*" 1S isolate and "*M. massiliense*" M18 isolate shared the same MST profile (MST17). *M. abscessus* 5S isolate exhibited the MST21 profile.

MST based tree and comparaison with rpoB identification and MLSA analysis

The MST-phylogenetic tree clustered isolates from patients P1 to P8 with *M. abscessus* reference strain, isolates from P9 and P10 with "*M. bolletii*" and isolate from P11 with "*M. massiliense*", in agreement with their *rpo*B sequence-based identification and MLSA analysis (Figure 1c). The MST, MLSA and *rpo*B phylogenetic trees separated the *M. abscessus* isolates into three principal clusters depicted by *M. abscessus,* "*M. bolletii*" and "*M. massiliense*" isolates (Figure 1a, b and c). However, MST resolved "*M. bolletii*" cluster into two sub-clusters formed by isolate P5 and all of the other *M. bolletii* isolates with a 76% bootstrap value, wich is discordant with MLSA and *rpoB* based tree. Each cluster or sub-cluster of the *M. abscessus* isolates corresponded to different genotypes. The "*M. massiliense*" cluster was more disperse and divided into six sub-clusters with isolate P11 and "*M. massiliense*" type strain sub-clustering alone. The results of this analysis were consistent for 67 isolates and inconsistent for two isolates P5 and *M.*

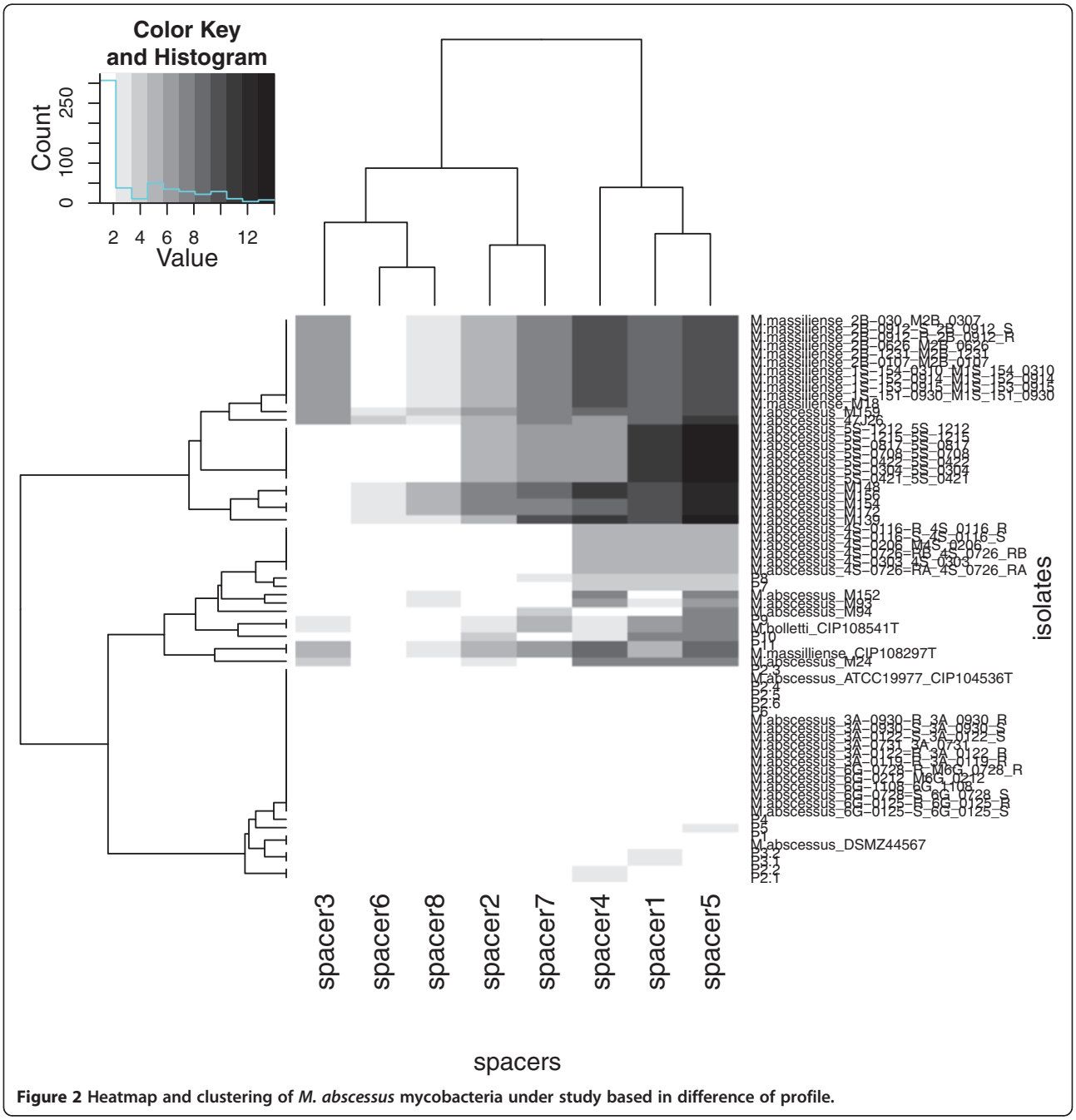


Figure 2 Heatmap and clustering of *M. abscessus* mycobacteria under study based in difference of profile.

abscessus M139. A heatmap incorporating all spacer patterns into a matrix further demonstrated that spacer n°2 was the most discriminating spacer (Figure 2). Hence, the tree based on the spacer n°2 sequence also discriminated the three *M. abscessus*, "*M. bolletii*" and "*M. massiliense*" clusters (Figure 3). This discrimination potential makes spacer n°2 a useful new tool for the accurate identification of *M. abscessus* subspecies. Furthermore, these data indicated that it was readily possible to discriminate isolates that would have been identified as "*M. bolletii*" [26] or "*M. massiliense*" [23] using a previous taxonomy proposal and are now grouped as *M. abscessus* subsp. *bolletii* according to a recent taxonomy proposal [20,21].

Conclusion

We herein developed a sequencing-based MST genotyping technique that allows the accurate identification and discrimination of *M. abscessus* mycobacteria. Therefore, MST could be added to the panel of molecular methods currently available for genotyping *M. abscessus* mycobacteria, with the advantages that MST is a PCR and

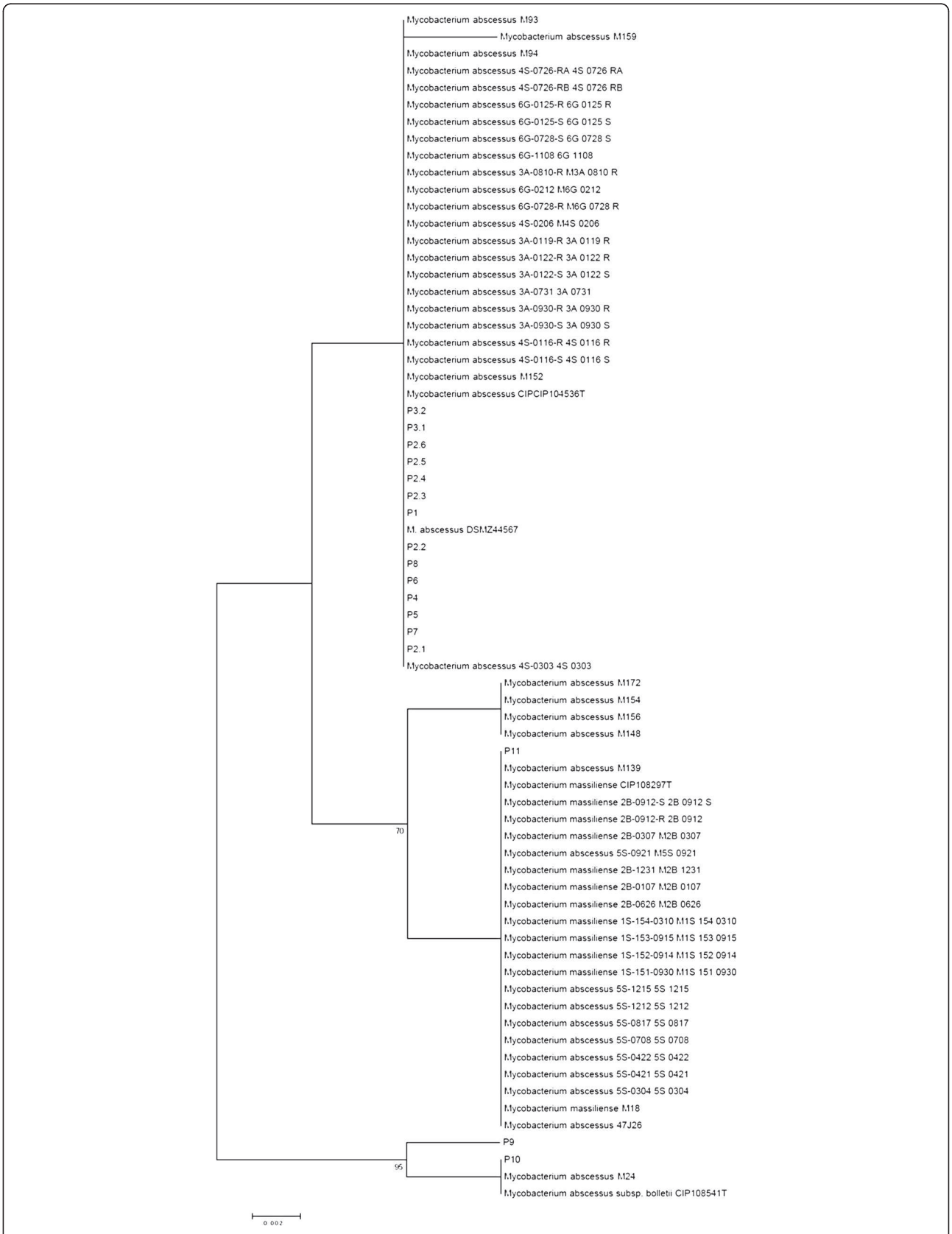


Figure 3 Phylogenetic tree based on MST spacer n°2 sequence.

sequencing-based technique, thereby providing a robust and accurate result without requiring a high DNA concentration and purity, as is the case for pulsed-field gel electrophoresis (PFGE) [5] and randomly amplified polymorphic DNA (RAPD) [33]. Furthermore, MST targets intergenic spacers, which undergo less evolutionary pressure and are thus more variable than the housekeeping genes targeted in multilocus sequence typing [21]. Also, MST incorporating sequencing is an open approach to described new genotypes more versatile than counting the number of tandem repeats [34]. We propose that MST could be incorporated into a polyphasic molecular approach to resolve the phylogenetic relationships of difficult-to-identify *M. abscessus* isolates [35]. Combining MST data with phylogenetic analyses clearly indicated that *M. abscessus* heterogeneity spans beyond the current two *M. abscessus* subspecies, as two "*M. massiliense*" isolates were readily discriminated from the other "*M. bolletii*" isolates [21]. These data, therefore, question the current nomenclature of *M. abscessus* mycobacteria, which incorporates mycobacteria previously recognized as "*M. bolletii*" and "*M. massiliense*" as "*M. abscessus* subsp. *bolletii*". The data presented here indicate that this nomenclature masks the underlying diversity of *M. abscessus* mycobacteria, potentially hampering the recognition of microbiological, epidemiological and clinical particularities that are linked to each subspecies. The elevation of "*M. massiliense*" as a new *M. abscessus* subspecies would accommodate the data produced in the present study [24].

Additional file

Additional file 1: *rpoB* and MLSA genes accession Number of 49 sequenced genomes.

Competing interests

The authors declare that they have no competing interest.

Authors' contributions

MS and IBK performed molecular analyses. MD designed the study. IBK, MS and MD interpreted data and wrote the draft. All authors read and approved the final manuscript.

Acknowledgments

IBK was financially supported by the Oeuvre Antituberculeuse des Bouches du Rhône. MS was financially supported by Infectiopole Sud Foundation.

Author details

[1]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UMR CNRS 7278, IRD 198, INSERM 1095. Faculté de médecine, 27, Boulevard Jean Moulin-Cedex 5, Marseille, France. [2]Laboratoire de Microbiologie et d'Immunologie, UR02/SP13, CHU Farhat Hached Sousse, Tunisie, France.

Received: 7 March 2012 Accepted: 20 December 2012
Published: 7 January 2013

References

1. Griffith DE, Girard WM, Wallace RJ Jr: **Clinical features of pulmonary disease caused by rapidly growing mycobacteria. An analysis of 154 patients.** *Am Rev Respir Dis* 1993, **147:**1271–1278.
2. Pierre-Audigier C, Ferroni A, Sermet-Gaudelus I, Le Bourgeois M, Offredo C, Vu-Thien H, Fauroux B, Mariani P, Munck A, Bingen E, Guillemot D, Quesne G, Vincent V, Berche P, Gaillard JL: **Age-related prevalence and distribution of nontuberculous mycobacterial species among patients with cystic fibrosis.** *J Clin Microbiol* 2005, **43:**3467–3470.
3. Olivier KN, Weber DJ, Wallace RJ Jr, Faiz AR, Lee JH, Zhang Y, Brown-Elliot BA, Handler A, Wilson RW, Schechter MS, Edwards LJ, Chakraborti S, Knowles MR, *et al*: **Nontuberculous mycobacteria. I: multicenter prevalence study in cystic fibrosis.** *Am J Respir Crit Care Med* 2003, **167:**828–834.
4. Chalermskulrat W, Sood N, Neuringer IP, Hecker TM, Chang L, Rivera MP, Paradowski LJ, Aris RM: **Non-tuberculous mycobacteria in end stage cystic fibrosis: implications for lung transplantation.** *Thorax* 2006, **61:**507–513.
5. Jönsson BE, Gilljam M, Lindblad A, Ridell M, Wold AE, Welinder-Olsson C: **Molecular epidemiology of *Mycobacterium abscessus*, with focus on cystic fibrosis.** *J Clin Microbiol* 2007, **45:**1497–1504.
6. Levy I, Grisaru-Soen G, Lerner-Geva L, Kerem E, Blau H, Bentur L, Aviram M, Rivlin J, Picard E, Lavy A, Yahav Y, Rahav G: **Multicenter cross-sectional study of nontuberculous mycobacterial infections among cystic fibrosis patients.** *Israel Emerg Infect Dis* 2008, **14:**378–384.
7. Griffith DE: **Emergence of nontuberculous mycobacteria as pathogens in cystic fibrosis.** *Am J Respir Crit Care Med* 2003, **167:**810–812.
8. Roux AL, Catherinot E, Ripoll F, Soismier N, Macheras E, Ravilly S, Bellis G, Vibet MA, Le Roux E, Lemonnier L, Gutierrez C, Vincent V, Fauroux B, Rottman M, Guillemot D, Gaillard JL, Jean-Louis Herrmann for the OMA Group: **Multicenter study of prevalence of nontuberculous mycobacteria in patients with cystic fibrosis in france.** *J Clin Microbiol* 2009, **47:**4124–4128.
9. Uyan ZS, Ersu R, Oktem S, Cakir E, Koksalan OK, Karadag B, Karakoc F, Dagli E: ***Mycobacterium abscessus* infection in a cystic fibrosis patient: a difficult to treat infection.** *Int J Tuberc Lung Dis* 2010, **14:**250–251.
10. Furuya EY, Paez A, Srinivasan A, Cooksey R, Augenbraun M, Baron M, Brudney K, Della-Latta P, Estivariz C, Fischer S, Flood M, Kellner P, Roman C, Yakrus M, Weiss D, Granowitz EV: **Outbreak of *mycobacterium abscessus* wound infections among "lipotourists" from the United States who underwent abdominoplasty in the Dominican Republic.** *Clin Infect Dis* 2008, **46:**1181–1188.
11. Koh SJ, Song T, Kang YA, Choi JW, Chang KJ, Chu CS, Jeong JG, Lee JY, Song MK, Sung HY, Kang YH, Yim JJ: **An outbreak of skin and soft tissue infection caused by *Mycobacterium abscessus* following acupuncture.** *Clin Microbiol Infect* 2010, **16:**895–901.
12. Viana-Niero C, Lima KV, Lopes ML, Rabello MC, Marsola LR, Brilhante VC, Durham AM, Leão SC: **Molecular characterization of *Mycobacterium massiliense* and *Mycobacterium bolletii* in isolates collected from outbreaks of infections after laparoscopic surgeries and cosmetic procedures.** *J Clin Microbiol* 2008, **46:**850–855.
13. Petrini B: ***Mycobacterium abscessus*: an emerging rapid-growing potential pathogen.** *APMIS* 2006, **114:**319–328.
14. Hayes D Jr: ***Mycobacterium abscessus* and other nontuberculous mycobacteria: evolving respiratory pathogens in cystic fibrosis: a case report and review.** *Southern Med J* 2005, **98:**657–661.
15. Sanguinetti M, Ardito F, Fiscarelli E, La Sorda M, D'argenio P, Ricciotti G, Fadda G: **Fatal pulmonary infection due to multidrug-resistant *Mycobacterium abscessus* in a patient with cystic fibrosis.** *J Clin Microbiol* 2001, **39:**816–819.
16. Shin JH, Lee HK, Cho EJ, Yu JY, Kang YH: **Targeting the *rpoB* gene using nested PCR-restriction fragment length polymorphism for identification of nontuberculous mycobacteria in hospital tap water.** *J Microbiol* 2008, **46:**608–614.
17. Huang WC, Chiou CS, Chen JH, Shen GH: **Molecular epidemiology of *Mycobacterium abscessus* infections in a subtropical chronic ventilatory setting.** *J Med Microbiol* 2010, **59:**1203–1211.
18. Adékambi T, Ben Salah I, Khlif M, Raoult D, Drancourt M: **Survival of environmental mycobacteria in *Acanthamoeba polyphaga*.** *Appl Environ Microbiol* 2006, **72:**5974–5981.
19. Koh WJ, Jeon K, Lee NY, Kim BJ, Kook YH, Lee SH, Park YK, Kim CK, Shin SJ, Huitt GA, Daley CL, Kwon OJ: **Clinical significance of differentiation of *Mycobacterium massiliense* from *Mycobacterium abscessus*.** *Am J Respir Crit Care Med* 2011, **183:**405–410.

20. Leao SC, Tortoli E, Viana-Niero C, Ueki SY, Lima KV, Lopes ML, Yubero J, Menendez MC, Garcia MJ: **Characterization of mycobacteria from a major Brazilian outbreak suggests that revision of the taxonomic status of members of the *Mycobacterium chelonae-M. abscessus* group is needed.** *J Clin Microbiol* 2009, **47:**2691–2698.
21. Macheras E, Roux AL, Bastian S, Leão SC, Palaci M, Sivadon-Tardy V, Gutierrez C, Richter E, Rüsch-Gerdes S, Pfyffer G, Bodmer T, Cambau E, Gaillard JL, Heym B: **Multilocus sequence analysis and *rpoB* sequencing of *Mycobacterium abscessus* (*sensu* lato) strains.** *J Clin Microbiol* 2011, **49:**491–499.
22. Adékambi T, Reynaud-Gaubert M, Greub G, Gevaudan MJ, La Scola B, Raoult D, Drancourt M: **Amoebal coculture of "*mycobacterium massiliense*" sp. nov. From the sputum of a patient with hemoptoic pneumonia.** *J Clin Microbiol* 2004, **42:**5493–5501.
23. Adékambi T, Drancourt M: **Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, hsp65, sodA, recA and *rpoB* gene sequencing.** *Int J Syst Evol Microbiol* 2004, **54:**2095–2105.
24. Adékambi T, Berger P, Raoult D, Drancourt M: ***rpoB* gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov.** *Int J Syst Evol Microbiol* 2006, **56:**133–143.
25. Macheras E, Roux AL, Ripoll F, Sivadon-Tardy V, Gutierrez C, Gaillard JL, Heym B: **Inaccuracy of single-target sequencing for discriminating species of the *Mycobacterium abscessus* group.** *J Clin Microbiol* 2009, **47:**2596–2600.
26. Cayrou C, Turenne C, Behr MA, Drancourt M: **Genotyping of *Mycobacterium avium* complex organisms using multispacer sequence typing.** *Microbiol* 2010, **156:**687–694.
27. Djelouadji Z, Arnold C, Gharbia S, Raoult D, Drancourt M: **Multispacer sequence typing for *Mycobacterium tuberculosis* genotyping.** *PLoS One* 2008, **3:**e2433.
28. Drancourt M, Roux V, Dang LV, Tran-Hung L, Castex D, Chenal-Francisque V, Ogata H, Fournier PE, Crubézy E, Raoult D: **Genotyping, Orientalis-like *Yersinia pestis*, and Plague Pandemics.** *Emer Infect Dis* 2004, **10:**1585–1592.
29. Wenjun LI, Mouffok N, Rovery C, Parola P, Raoult D: **Genotyping *Rickettsia conorii* detected in patients with Mediterranean spotted fever in Algeria using multispacer typing (MST).** *Clin Microbiol Inf* 2009, **15:**281–283.
30. Foucault C, La Scola B, Lindroos H, Andersson SGE, Raoult D: **Multispacer typing technique for sequence-based typing of *Bartonella Quintana*.** *J Clin Microbiol* 2005, **43:**41–48.
31. Hunter PR, Gaston MA: **Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity.** *J Clin Microbiol* 1988, **26:**2465–2466.
32. Kumar S, Tamura K, Jakobsen IB, Nei M: **MEGA2: molecular evolutionary genetics analysis software.** *Bioinformatics* 2001, **17:**1244–1245.
33. Zhang Y, Rajagopalan M, Brown BA, Wallace RJ Jr: **Randomly amplified polymorphic DNA PCR for comparison of *Mycobacterium abscessus* strains from nosocomial outbreaks.** *J Clin Microbiol* 1997, **35:**3132–3139.
34. Choi GE, Chulhun LC, Whang J, Kim HJ, Kwon OJ, Koh WJ, Shin SJ: **Efficient differentiation of *mycobacterium abscessus* complex isolates to the species level by a novel PCR-based variable-number tandem-repeat assay.** *J Clin Microbiol* 2011, **49:**1107–1109.
35. Zelazny AM, Root JM, Shea YR, Colombo RE, Shamputa IC, Stock F, Conlan S, McNulty S, Brown-Elliott BA, Wallace RJ Jr, Olivier KN, Holland SM, Sampaio EP: **Cohort study of molecular identification and typing of *Mycobacterium abscessus, Mycobacterium massiliense,* and *Mycobacterium bolletii.*** *J Clin Microbiol* 2009, **47:**1985–1995.

doi:10.1186/1471-2180-13-3
Cite this article as: Sassi *et al.*: *Mycobacterium abscessus* multispacer sequence typing. *BMC Microbiology* 2013 **13**:3.

III- LES MYCOBACTÉRIOPHAGES DE *MYCOBACTERIUM ABSCESSUS*

Commentaire

Dans la première partie de cette thèse, nous avons montré que les génomes de *M. abscessus sensu lato* codent pour des prophages. Au sein de *M. abscessus sensu lato*, *M. bolletii* se singularise par la production spontanée d'un mycobactériophage (Adékambi T, Drancourt M, données non publiées). Un total de 3427 mycobactériophages a été isolé et 448 génomes ont été séquencés et assemblés en 20 groupes (notés groupe A à groupe T) et sept singletons [24] (http://phagesdb.org/). Ce grand ensemble de génomes complets révèle une diversité considérable [20-25] et peu d'homologues détectables dans les bases de données [25]. Outre l'intérêt de phages infectant une bactérie pathogène, les études de mycobactériophages peuvent révéler des caractéristiques spécifiques des interactions mycobactéries / phages en raison des caractéristiques de la paroi cellulaire particulières de ces hôtes [27]. Bien qu'un grand nombre de mycobactériophages ait été isolé et séquencé [20-24], leurs mécanismes d'infection et de reconnaissance de l'hôte sont encore mal compris.

Tous les mycobactériophages caractérisés sont des phages ADN double brin de l'ordre des Caudovirales et appartiennent à la famille des *Siphoviridae* ou *Myoviridae* [24]. Nos objectifs ont été dans cette 3ème partie d'analyser en profondeur la particularité de *M. bolletii* à exprimer un mycobactériophage en essayant de comprendre son mécanisme d'infection par la résolution pour la première fois de la structure 3D d'un mycobactériophage par microscopie électronique;

puis d'enrichir le répertoire et la diversité des souches de *M. abscessus sensu lato* en mycobactériophages et de comprendre leurs histoires évolutives.

Mycobactériophage *de M. bolletii*

Nous rapportons ici l'annotation du génome et la caractérisation du prophage de *M. bolletii* que nous avons nommé Araucaria, ainsi que la première reconstruction 3D en microscopie électronique de particules isolées de l'ensemble du virion. Araucaria appartient à la famille des *Siphoviridae* et possède un génome de 64 kb contenant 89 protéines dont 27 sont annotées. La capside et le connecteur de Araucaria présentent des similitudes avec plusieurs phages de bactéries à Gram négatif ou Gram positif, sa caractéristique la plus distinctive et selon laquelle le nom du phage a été choisi est la queue hélicoïdale décorée par des pointes radiales, pouvant éventuellement servir à l'adhésion à l'hôte. Son dispositif d'adsorption, à l'extrémité de la queue, comprend des caractéristiques observées chez des phages liant à des récepteurs de protéines tels que les bactériophages SPP1. Ces résultats suggèrent qu'Araucaria peut infecter la mycobactérie hôte par un mécanisme impliquant l'adhésion aux saccharides de la paroi cellulaire et des protéines, une caractéristique qui doit encore être explorée. Ce travail a fait l'objet d'un 3ème article de thèse et a été publié dans J**ournal of Virology** dont l'impact factor est de 5,076.

Prophages de *M. abscessus sensu lato*

Notre objectif dans cette partie est d'analyser en détails le répertoire des prophages de *M. abscessus sensu lato* et de comprendre leur histoire évolutive. Nous avons analysé 48 génomes de *M. abscessus sensu lato* pour la présence de séquence de phages dans leurs génomes. Des séquences de prophage ont été détectées dans 47/48 génomes de *M. abscessus*, codant de 0 à 8 régions de prophages spécifiant quatre situations : i) prophages complets codant des protéines de structure, les protéines de lyse, des protéines d'intégration et des protéines nécessaires à la réplication et la recombinaison, ii) prophages incomplets dont certains gènes fonctionnels ne sont pas annotés avec certitude, iii) prophages incomplets ayant perdu une partie des gènes fonctionnels, iiii) éléments prophage-like codant seulement quelques protéines virales. Nous avons ainsi annoté 20304 protéines virales dans 853 groupes comprenant 30% qui sont des protéines de mycobactériophages, 48% sont d'autres phages qui infectent les bactéries à Gram négatif et positif et 23% sont des protéines virales. Par ailleurs l'analyse fonctionnelle a montré que 44% des protéines annotées sont impliquées dans la réplication de l'ADN et le métabolisme et 37% sont des protéines de bactériophages dont les protéines structurales. Par ailleurs, les espèces *M. abscessus* codent des protéines annotées comme répresseur de phage et dont le rôle reste indéfini. Notre analyse phylogénétique suggère que les espèces de *M. abscessus sensu lato* ont été infectées par différents mycobactériophages et ont une histoire évolutive différente de celle

des hôtes mycobactériens et contiennent aussi des protéines acquises par transfert horizontal dont la plupart sont des protéines de bactériophages et des protéines de fonctions inconnues. Ce travail est en cours de rédaction pour soumission, nous avons intégré un premier draft comme notre **4ème article** de thèse dans notre 3ème chapitre.

JVI
Journals.ASM.org

The First Structure of a Mycobacteriophage, the *Mycobacterium abscessus* subsp. *bolletii* Phage Araucaria

Mohamed Sassi,[a,b] Cecilia Bebeacua,[a] Michel Drancourt,[b] Christian Cambillau[a]

Article N.3 : **Sassi M,** Bebeacua C,

Drancourt M and Cambillau C.

The first structure of a mycobactériophage,

the *Mycobacterium abscessus* subsp.

bolletii Phage Araucaria.

J Virol. 2013 14:8099-8109.

JVI
Journals.ASM.org

The First Structure of a Mycobacteriophage, the *Mycobacterium abscessus* subsp. *bolletii* Phage Araucaria

Mohamed Sassi,[a,b] Cecilia Bebeacua,[a] Michel Drancourt,[b] Christian Cambillau[a]

Architecture et Fonction des Macromolécules Biologiques, UMR CNRS 7257, Aix-Marseille-Université, Marseille, France[a]; Aix-Marseille-Université, URMITE, UMR 63 CNRS 7278, IRD 198, INSERM 1095, Marseille, France[b]

The unique characteristics of the waxy mycobacterial cell wall raise questions about specific structural features of their bacteriophages. No structure of any mycobacteriophage is available, although ~3,500 have been described to date. To fill this gap, we embarked in a genomic and structural study of a bacteriophage from *Mycobacterium abscessus* subsp. *bolletii*, a member of the *Mycobacterium abscessus* group. This opportunistic pathogen is responsible for respiratory tract infections in patients with lung disorders, particularly cystic fibrosis. *M. abscessus* subsp. *bolletii* was isolated from respiratory tract specimens, and bacteriophages were observed in the cultures. We report here the genome annotation and characterization of the *M. abscessus* subsp. *bolletii* prophage Araucaria, as well as the first single-particle electron microscopy reconstruction of the whole virion. Araucaria belongs to *Siphoviridae* and possesses a 64-kb genome containing 89 open reading frames (ORFs), among which 27 could be annotated with certainty. Although its capsid and connector share close similarity with those of several phages from Gram-negative (Gram$^-$) or Gram$^+$ bacteria, its most distinctive characteristic is the helical tail decorated by radial spikes, possibly host adhesion devices, according to which the phage name was chosen. Its host adsorption device, at the tail tip, assembles features observed in phages binding to protein receptors, such as phage SPP1. All together, these results suggest that Araucaria may infect its mycobacterial host using a mechanism involving adhesion to cell wall saccharides and protein, a feature that remains to be further explored.

Mycobacterium abscessus subsp. *bolletii* is a member of the *Mycobacterium abscessus* complex of opportunistic pathogens responsible for outbreaks of skin and soft tissue infections following surgical and cosmetic practices (1). These mycobacteria are also responsible for a wide range of other infections, including catheter-related bacteremia (1) and respiratory tract infections in patients with lung disorders, particularly cystic fibrosis (1–5). *M. abscessus* mycobacteria are broadly resistant to antibiotics, so their infections may require curative surgery (6). *M. abscessus* subsp. *bolletii* was first isolated from a respiratory tract specimen collected from a woman presenting with hemoptoic pneumonia (7).

To date, 3,427 mycobacteriophages have been isolated and 448 mycobacteriophage genomes have been sequenced and assembled in 20 clusters (A through T) and seven sequenced singletons (8) (http://phagesdb.org/). This large set of complete genomes reveals a considerable diversity (8–11). Conversely, the large genetic repertoire of the predicted mycobacteriophage protein-coding genes possesses a low number of detectable homologues in the public databases (9, 10). Besides the interest in phages infecting a bacterial pathogen, studies of mycobacteriophages may reveal specific features of mycobacterium-phage interactions due to the peculiar cell wall features of these hosts (12). Despite the wealth of mycobacteriophages isolated and sequenced (8–11, 13), their infection and host recognition mechanisms are still poorly understood.

All mycobacteriophages characterized to date are double-stranded DNA (dsDNA)-tailed phages of the *Caudovirales* order and belong to the *Siphoviridae* or *Myoviridae* (9, 10). We report here the electron microscopic (EM) reconstruction of the *M. abscessus* subsp. *bolletii* siphophage Araucaria and its integrated prophage. To date, no other EM structure of any mycobacteriophage has been described. Due to the highly flexible nature of *Siphoviridae* tails, which limits the structural characterization, we used a methodology specially implemented to characterize their tail (14). Based on our EM reconstructions and bioinformatics analysis, we could achieve a pseudoatomic model for some parts of this siphoviridal virion and reveal the striking features of Araucaria's tail and host adsorption device (HAD).

MATERIALS AND METHODS

Mycobacterial host. *M. abscessus* subsp. *bolletii* CIP108541^T was grown in 7H9 broth (Difco, Bordeaux, France) enriched with 10% OADC (oleic acid, bovine serum albumin, dextrose, and catalase) in 8-ml tubes at 37°C.

Araucaria mycobacteriophage production and purification. Two liters of *M. abscessus* subsp. *bolletii* in 7H9 broth complemented with 100 mM $CaCl_2$ was cultured. *M. abscessus* subsp. *bolletii* culture was then centrifuged at 3,000 × *g* for 10 min and filtrated using 0.45-μm-pore-size filters (Minisart filters). The 2 liters of supernatant was ultracentrifuged (Beckman Coulter ultracentrifuge) at 22,000 rpm for 2 h. The pellet was resuspended in 2 ml of phosphate-buffered saline (PBS) and purified using preparative Superdex 200 (GE Healthcare) gel filtration.

***M. abscessus* subsp. *bolletii* prophage genome analysis and annotation.** The *M. abscessus* subsp. *bolletii* CIP108541^T genome was made available (15). The prophage region was predicted using PHAST (PHAge Search Tool) (16). Open reading frames (ORFs) were predicted using the software program Prodigal (http://prodigal.ornl.gov/) (17) with default parameters. The predicted protein-coding sequences were searched against the National Center for Biotechnology Information (NCBI) nonredundant (NR) database and UNIPROT (http://www.uniprot.org/) and against the Clusters of Orthologous Groups (COG) (18) using BLASTP. The ARAGORN software tool (19) was used to find tRNA genes,

Received 4 May 2013 Accepted 6 May 2013
Published ahead of print 15 May 2013
Address correspondence to Christian Cambillau, cambillau@afmb.univ-mrs.fr.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JVI.01209-13.

doi:10.1128/JVI.01209-13

TABLE 1 Summary of the data processing strategies employed for *M. abscessus* subsp. *bolletii* mycobacteriophage reconstruction

Structure	Symmetry	Resolution (Å)	No. of particles
Capsid	Icosahedral	30	7,431
Connector	C12	28	6,740
Tail	Helicoidal	24	2,565
Baseplate	C6	26	6,460

and a BLASTn search was conducted against the NR database. Proteins were also checked for domain using a hidden Markov model (HMM) search against the Pfam database (20). The predicted protein-coding sequences were also searched against the ACLAME (A CLAssification of Mobile genetic Elements) database. Tandem Repeat Finder was used for repetitive DNA prediction (21) and against Mycobacteriophage database (http://phagesdb.org/). CRISPRs were searched in the *M. abscessus* subsp. *bolletii* genome using the CRISPERfinder software program (http://crispr.u-psud.fr/Server/).

Electron microscopy. (i) Specimen preparation. Approximately 3 μl of sample was applied onto glow-discharged carbon-coated grids and incubated for 1 min. The grids were blotted, 10 μl of a 2% uranyl acetate solution was added, and they were incubated for 30 s. Stain excess was blotted, and the grids were transferred to the microscope for imaging.

(ii) Data collection. Approximately 1,500 charge-coupled device (CCD) images were collected using a Tecnai Spirit operated at 120 kV and a 2Kx2K CCD camera at a magnification of ×48,500, resulting in a pixel size of 4.95 Å/pixel.

(iii) Image processing. Particles (2,500 [full phage], 7,300 [capsid], 5,900 [connector], 2,000 [tail], or 6,400 [baseplate]) were manually selected using the program boxer from the EMAN2 package (22); extracted into boxes of 500 by 500 pixels (full phage, coarsened by 2), 200 by 200 pixels (capsid), 100 by 100 pixels (connector), 80 by 80 pixels (tail), or 100 by 100 pixels (baseplate); and combined into the five different data sets (Table 1). The data sets were pretreated using the SPIDER package (23) and submitted to maximum likelihood (ML) classification and alignment (24) using the Xmipp package (25). The initial models were built to form a visually selected class average representing a side view imposing the corresponding symmetry (C6 for the full phage, tail, and baseplate; C12 for the connector; and icosahedral for the capsid). The initial models were then refined by three-dimensional ML refinement first and further refined with SPIDER with a sampling rate of 5°. After refinements, final models were obtained at resolutions of approximately 30 Å (capsid), 28 Å (connector), 24 Å (tail), and 26 Å (baseplate) as estimated by Fourier shell correlation (FSC) and the ½-bit threshold criterion (26) (Fig. 1).

(iv) Tail helical processing. The tail particles pretreated as described above were submitted to helical processing. The rotational symmetry used was C6 and, as the particles were already aligned, the maximum allowed in-plane rotational angle was set to 10°. The initial helical parameters were determined using the Brandeis Helical Package (27) to calculate the Bessel orders of the basic layer lines (6 and −6). These were later refined by IHRSR++ (28) to a helical rise of 38 Å and a rotation between subunits of 26°.

(v) Structure visualization. Molecular graphics and analyses were performed with the UCSF Chimera package (Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco). The model/EM map or EM map/EM map fitting was performed by the option "fit in map" of the "volume" register. We then chose a mode of fitting in which a map of the atomic model was calculated at a given resolution (here 30 Å), and a correlation coefficient was calculated between the calculated and observed maps. A fit was considered good when a correlation coefficient better than 0.8 was obtained.

RESULTS

Genome characteristics of mycobacteriophage Araucaria. The mycobacteriophage Araucaria was isolated by concentrating a large volume of *M. abscessus* subsp. *bolletii* CIP108541^T culture and was not induced by usual methods. The Araucaria prophage genome sequence was predicted and retrieved from the sequenced *M. abscessus* subsp. *bolletii* CIP108541^T genome (15; M. Drancour, unpublished data). The Araucaria prophage genome is 64,129 bp long, within the average size of mycobacteriophage genomes (41,441 bp to 164,602 bp) (http://phagesdb.org/) (Table 2). The GC content was calculated to be 64.41%, a value similar to that of its host, *M. abscessus* subsp. *bolletii* (64.0%) (15), and within the average range of mycobacteriophage genomes (50.3% to 68.4%) (Tables 2 and 3). It has been proposed that two mycobacterial phage genomes displaying nucleotide sequence similarity above 50% should be included within the same cluster (13). Using this criterion, BLASTN and Dotter analysis of the Araucaria genome, mycobacteriophage clusters (A through T and singletons), and *Mycobacterium* prophages revealed that Araucaria has no discernible overall DNA sequence similarity to *Mycobacterium* prophages (Fig. 2). However, close to the above-mentioned cutoff criteria, Araucaria shares 52% sequence similarity to mycobacteriophage Dori and 47% sequence similarity to members of cluster B, together with even weaker sequence similarities to clusters A, G, I, and K (8, 11) (http://phagesdb.org/). Consequently, according to the above-described analysis and phylogenetic comparison of the Araucaria genome with mycobacteriophages (Dori, Bzx1, L5, Rosebush, Phaedrus, Halo, Brujita, and TM4) representing clusters A, G, I, and K, the Araucaria genome can be assigned to a Dori-like prophage (Fig. 3).

The genomic analysis of the Araucaria prophage genome revealed four main clusters dedicated to host integration, DNA replication, DNA packaging and lysis, and virion assembly (Fig. 4). Analysis of the Araucaria prophage genome identified 89 putative open reading frames (ORFs) encoding proteins but did not reveal the presence of tRNA or other small RNA coding sequences (Fig. 4; see also Table S1 in the supplemental material). Overall, based on BLAST analysis and genomic organization, a function could be proposed for only 27 of the 89 ORFs (30% of total) (see Table S1). This analysis showed also that 55 ORFs of the Araucaria

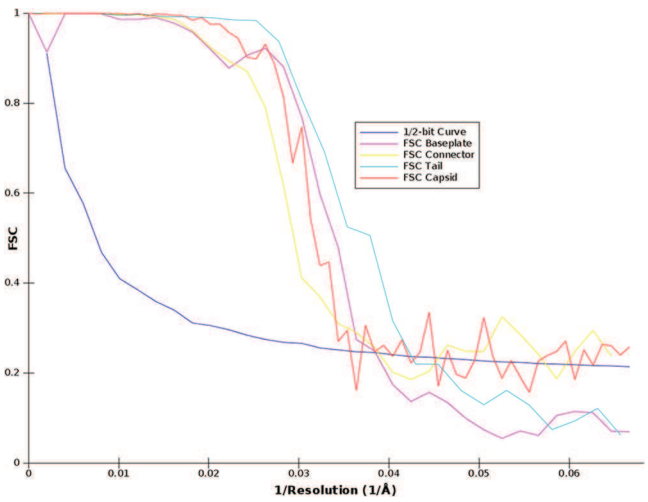


FIG 1 Graphs of the FSC curves of the final three-dimensional reconstructions. Graphs were obtained by correlation of two different three-dimensional reconstructions created by splitting the particles set into two subsets. The resolution was estimated by the ½-bit cutoff threshold criterion.

TABLE 2 Genometrics of Araucaria prophage, 4 mycobacterium prophages, and mycobacteriophage cluster genomes

Phage	Cluster	No. of members	No. of subclusters	Avg genome size (bp)	Avg GC content (%)	Avg no. of genes	Avg no. of tRNAs
Araucaria	Dori-like			64,129	64.2	89	0
M. abscessus type strain	Prophage			80,545	59.5	110	1
Mycobacterium tuberculosis H37Rv PhiRv1	Prophage			10,643	66.0	64	0
M. tuberculosis H37Rv PhiRv2	Prophage			8,759	66.8	45	0
Mycobacterium ulcerans Agy99 phiMUO1	Prophage			18,424	62.3	67	0
Dori	Singleton			64,613	66.0	93	0
DS6A	Singleton			60,588	68.4	97	0
Patience	Singleton			70,506	50.3	108	1
Wildcat	Singleton			78,296	57.2	148	24
A	A	10	242	51,484	63.3	89.3	0.9
B	B	5	116	68,667	67.1	98.1	0.0
C	C	2	55	155,578	64.7	228.7	33.4
D	D		14	6,471	59.7	86.9	0
E	E		40	75,398	63.1	143	1.8
F	F	3	76	5,741	61.5	105.2	0
G	G		21	41,837	66.6	62.3	0
H	H	2	5	69,953	57.1	98.7	0
I	I	2	6	49,954	66.5	78	0
J	J		12	109,821	60.9	78	0
K	K	5	35	59,689	67.0	233	1.5
L	L	2	14	74,978	58.9	118	8
M	M		4	81,636	61.3	142.5	20
N	N		6	42,756	66.2	65.5	0
O	O		4	70,759	65.4	113.5	0
P	P		8	47,376	67.1	82	0
Q	Q		4	53,757	67.4	78	0
R	R		2	71,102	56.0	96	0
S	S		2	65,172	63.4	107	0
T	T		3	42,833	66.2	None	None

genome (61%) share sequence similarity with other mycobacteriophage genes, among which only a few had known function and seven were found to be unique (see Table S1).

(i) Integration cassette. The ORF2 protein showed 88% and 58% sequence similarity with the integrase protein of *Mycobacterium massiliense* and mycobacteriophage Spartacus (cluster F1), respectively. Using the Pfam database, an integrase domain was detected toward the C terminus of this ORF protein, between amino acid positions 67 and 241 (Pfam: PF0058). Temperate mycobacteriophages typically encode either a serine or tyrosine integrase that mediates prophage integration (8–11, 13). In the Araucaria prophage genome, the ORF2 protein is perfectly identified as a tyrosine integrase and is located 22 kb distant from the cluster of structural proteins (Fig. 3).

(ii) DNA replication, recombination, and modification genes. Many genes associated with DNA replication, recombination, and modification could be identified in the Araucaria genome. The ORF16 protein shares 79% sequence similarity with an ORF protein from mycobacteriophage Dori and belongs to the YqaJ recombinase family. Such recombinases were shown to form oligomers and to function as processive alkaline exonucleases that digest linear double-stranded DNA in an Mg^{2+}-dependent fashion (29). The ORF17 protein shares 67% sequence similarity with the RecT recombinase from mycobacteriophage Brujita (cluster I1). The ORF19 protein sequence is 56% similar to the RusA endodeoxyribonuclease from mycobacteriophage MacNcheese (cluster K3). Finally, the ORF22, ORF24, and ORF46 proteins are putative methyltransferases, since they are homologous to methyltransferases from mycobacteriaphages Ramsey (cluster F1), Dlane (cluster F1), and *M. massiliense*.

(iii) DNA packaging and lysis genes. The ORF47 protein had several characteristics in common with phage terminase large subunit: it shares 74% identity with the large terminase subunit of mycobacteriophage Dori, and it contains a terminase domain between amino acids 61 and 454. Phage terminases are generally composed of a large and a small subunit encoded by side-by-side genes. Despite this, we could not identify sequence-wise a gene coding for the small terminase subunit in the Araucaria prophage genome. Despite this, the ORF48 protein could be assigned to the small terminase subunit with respect to its position and size.

The ORF84 protein possesses an *N*-acetylmuramoyl-L-alanine amidase domain. *N*-acetylmuramoyl-L-alanine amidases have been described to cleave the bond between *N*-acetylmuramoyl moieties and L-amino acid residues in cell wall glycopeptides. BLAST comparison of the ORF84 protein revealed homology to lysin A (64%) from mycobacteriophage TM4 (cluster K2 [30]). The ORF85 protein shares 82% sequence identity to *M. massiliense* cutinase, as well as with ORF proteins belonging to mycobacteriophage Marvin (cluster S). Cutinase that degrades cutin, a polyester protecting leaves, is also present in mycobacteria, for which it might be involved in cell wall wax degradation (31, 32). Finally, the ORF30 protein shares 38% sequence similarity with the holin (gp31) of the mycobacteriophage Fruitloop (cluster F1 [9, 10]), a small protein forming membrane pores through which the folded endolysin reaches its substrate. Two transmembrane

TABLE 3 Prophage and mycobacteriophage genomes used for this study

Phage	Cluster	Genome size (bp)	GC content (%)	No. of genes	No. of tRNAs
Araucaria	Dori-like	64,129	64.2	89	0
M. abscessus type strain	Prophage	80,545	59.5	110	1
M. tuberculosis H37Rv PhiRv1	Prophage	10,643	66.0	64	0
M. tuberculosis H37Rv PhiRv2	Prophage	8,759	66.8	45	0
M. ulcerans Agy99 phiMUO1	Prophage	18,424	62.3	67	0
Dori	Singleton	64,613	66.0	93	0
DS6A	Singleton	60,588	68.4	97	0
Patience	Singleton	70,506	50.3	108	1
Wildcat	Singleton	78,296	57.2	148	24
Bxb1	A1	50,550	63.6	86	0
L5	A2	52,297	62.3	85	3
Bxz2	A3	50,913	64.2	86	3
Peaches	A4	51,377	63.9	86	0
Airmid	A5	51,083	59.8	90	1
DaVinci	A6	51,547	61.5	97	1
Timshel	A7	53,278	63.1	85	1
Saintus	A8	49,228	61.2	93	1
Alma	A9	51,339	62.6	91	1
Twister	A10	51,094	65.0	87	1
Colbert	B1	67,774	66.5	100	0
Rosebush	B2	67,480	68.9	90	0
Phaedrus	B3	68,090	67.6	98	0
Nigel	B4	69,904	68.3	94	1
Acadian	B5	69,864	68.4	96	0
Bxz1	C1	156,102	64.8	225	35
Myrna	C2	164,602	65.4	229	41
Adjutor	D	64,511	59.9	86	0
244	E	74,483	63.4	142	2
che8	F1	59,471	61.3	112	0
Che9d	F2	56,276	60.9	111	0
Squirty	F3	60,285	62.4	None	None
Halo	G	42,289	66.7	64	0
Konstantine	H1	68,952	57.4	95	0
Barnyard	H2	70,797	57.5	109	0
Brujita	I1	47,057	66.8	74	0
Che9c	I2	57,050	65.4	84	0
Omega	J	110,865	61.4	237	2
Adephagia	K1	59,646	66.6	94	1
TM4	K2	52,797	68.1	89	0
MacnCheese	K3	61,567	67.3	99	0
Fionnbharth	K4	58,076	68.0	94	1
Larva	K5	62,991	65.3	96	1
LeBron	L1	73,453	58.8	120	9
Faith1	L2	75,960	58.9	129	12
Bongo	M	80,228	61.6	132	19
Redi	N	42,594	66.1	68	0
Corndog	O	69,777	65.4	99	0
BigNuz	P	48,984	66.7	82	0
Giles	Q	53,746	67.3	78	0
Send513	R	71,547	56.0	96	0
Marvin	S	65,100	63.4	107	0
Bernal13	T	42,392	66.2	None	None

helices have been detected in the ORF30 protein using TMHMM, encompassing residues 13 to 40 and 45 to 72.

(iv) Virion assembly cassette. Identification of the genes involved in virion formation and assembly is facilitated by their conserved order. The virion structure and assembly cassette likely spans ORF52 to ORF73 (Fig. 3). This ~25-kb segment is among the smallest virion structure and assembly cassettes observed in mycobacteriophages, along with those from mycobacteriophage BP (24 kb) and mycobacteriophage Marvin (20 kb) (33, 34). A putative function was assigned to the ORF55, -57, and -62 proteins on the basis of their similarity to proteins of mycobacteriophage Dori and members of cluster B. The ORF55 protein shares 70% sequence similarity with mycobacteriophage Dori gp8, annotated as the portal protein, and its sequence contains a phage Mu protein F-like domain, required for viral head morphogenesis, suggesting that the ORF55 protein could be the minor head

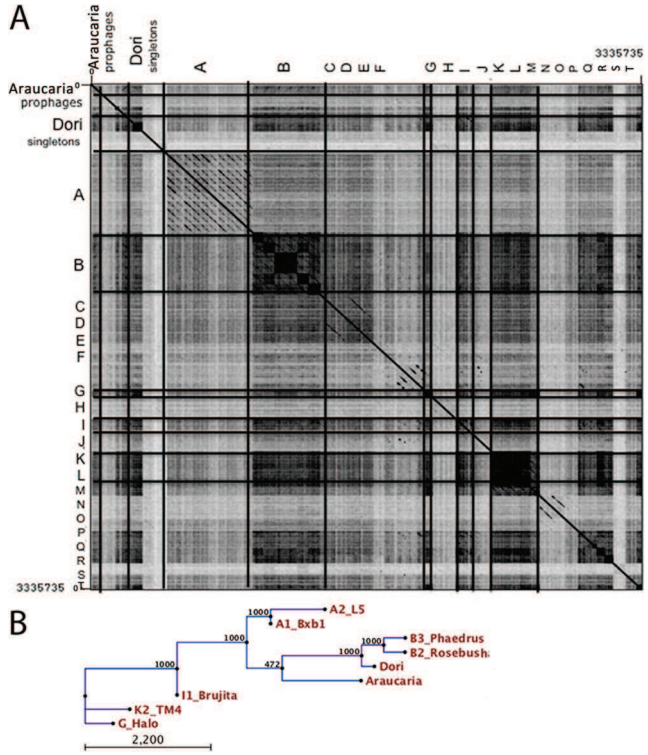


FIG 2 Mycobacteriophage Araucaria clusterization. (A) Dot plot comparison of Araucaria genome with 4 mycobacterium prophages and 47 mycobacteriophages representing mycobacteriophage clusters and listed in Table S1 in the supplemental material using Gepard (75). (B) Phylogenetic tree based on the nucleotide genome of Araucaria and its 8 closest mycobacteriophages using the neighbor joining method.

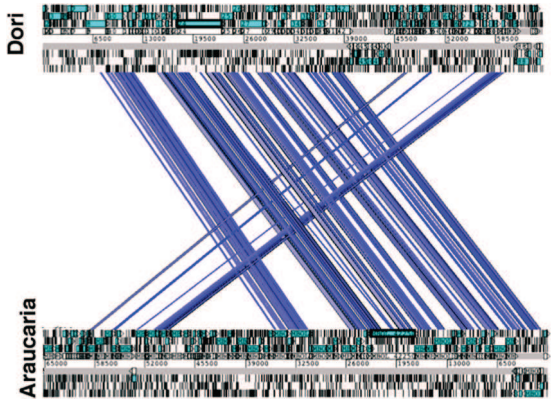


FIG 3 Overall alignment of the genome of Araucaria prophage with that of mycobacteriophage Dori.

protein. The ORF57 protein shares 55% sequence similarity with mycobacteriophage Dori gp10, annotated as the major capsid protein (MCP). The ORF62 protein is similar to major tail subunit proteins (MTP, Pham2299) (http://phagesdb.org/). We identified Araucaria ORF69 as encoding the tape measure protein (TMP) thanks to its genomic position as well as its very large size (5,379 bp). Analysis of its predicted secondary structure reveals a high propensity for alpha-helical and coiled-coil structures, which are hallmarks of TMPs. Finally, based on their position, we expected ORF70 to -73 to code for the host adsorption device (HAD).

Araucaria virion structure. (i) Capsid and head-to-tail connecting region. Bacteriophage capsids contain and protect the viral genome densely packed under high pressure (35). We computed a reconstruction of the Araucaria mycobacteriophage capsid at a 30-Å resolution using ~7,431 particle images and applying icosahedral symmetry (Fig. 5A and B). The mature capsid is ~600 Å wide along its 5-fold axes and is made of 60 hexamers and 11 pentamers of the major capsid protein (MCP; ORF57 protein), organized with a T=7 symmetry, as well as a dodecamer of the portal protein occupying a unique vertex (Fig. 5B and C). The large number of MCP structures reported to date has established the conservation of the HK97 MCP fold (so-called "Johnson fold") among tailed phages, herpesviruses, and some archaeal viruses (36–39). The HK97 MCP hexamers and pentamers could be fitted readily in the capsid EM structure (Fig. 5A).

The connector serves to attach the phage capsid to its tail and is composed of three different components organized as stacked rings: the portal protein and two head completion proteins (connector and stopper). It is located at a unique capsid vertex, where it replaces a penton motif (Fig. 5A to C). The 60-fold averaging procedure of the capsid reconstruction process averaged out the portal density. We therefore solved independently the structure of the connector region using ~6,740 particles and applying 12-fold symmetry along the connector channel axis (Fig. 5D to F). The reconstructed connector was then plugged at a vertex position using molecular fitting with Chimera (40) (Fig. 5B and C).

The portal, a dodecameric protein, discloses a conserved fold in tailed phages and herpesviruses (14, 36). It is involved in DNA packaging during assembly and in DNA release when phage infection occurs. The Araucaria portal (ORF52) exhibits weak sequence similarity with the HK97 portal sequence using the FFAS03 server (41) and HHpred (42) (Table 4). We used instead the structure of the described SPP1 dodecameric portal (43) to fit into the proximal region of the connector reconstruction. The two proteins contain similar numbers of amino acids (617 for Araucaria and 503 for SPP1). We observed a good agreement at this resolution between the SPP1 dodecameric atomic model and the EM map (correlation coefficient [cc] > 0.8) (Fig. 5D and E). This result further supports the suggested structural similarity between the Araucaria and HK97/SPP1 portal proteins (Table 4). The remaining region of the connector reconstruction reported here was assumed to account for the two rings of head completion proteins. It was modeled using SPP1 gp15 connecting module (PDB 2KBZ) and SPP1 gp16, the stopper (PDB 2KCA) (44) that closes the channel to prevent DNA release in the absence of an infection event (45) (Fig. 2E). An SPP1 dodecameric model of gp15 was found to fit well with the dimensions of the corresponding Araucaria connector map, while the SPP1 gp16 dodecamer was found to be slightly larger for the corresponding region in Araucaria (Fig. 5E and F). Dodecameric models of gp15 and gp16 were those used for fitting in the SPP1 map in the work of Lhuillier et al. (44).

(ii) Araucaria tail. We investigated the Araucaria tail structure. Initially, following our previously applied procedure (14), we produced a 6-fold averaged, low-resolution reconstruction of the whole Araucaria phage from selected straight-tailed virions. This reconstruction indicated clearly that the tail tube was composed of 34 stacked hexamers, including the tail terminator hexamer, at the interface with the connector, and the 33 MTP (ORF62) hexamers form the rest of the tube. In a subsequent step, we boxed short tail

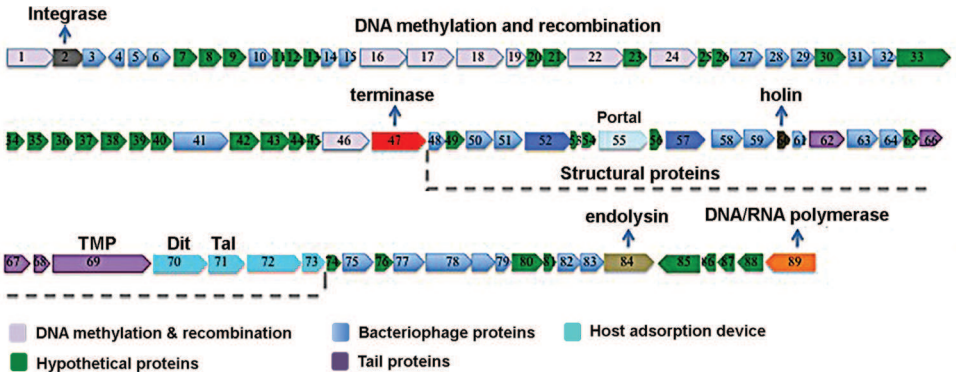


FIG 4 Map of the mycobacteriophage Araucaria genome.

segments of ~9 MTP rings and combined them in one data set processed with the appropriate helical symmetry.

The Araucaria tail extends over 1,100 Å between the connector and the baseplate (Fig. 6A). It exhibits a very peculiar decoration of the MTP rings, making the tail resemble a branch of the Araucaria tree (hence its given name). The diameter of the tail, including the decorations, is ~160 Å. The diameter of the core of the MTP rings is ~100 Å, at its largest extension, and 70 Å at the intersections between the rings (Fig. 6B and C). The MTP hexameric rings are rotated by 26° between each other, and the interhexamer distance is ~38 Å (Fig. 6B and C). The tail tube displays a 40-Å-wide central channel between the connector and the baseplate, which forms the DNA ejection passage (14, 45) (Fig. 6C). This channel is filled in our reconstruction, and we attributed the density to the tape measure protein (TMP; ORF69 protein), the molecular ruler controlling the tail length (46). The oligomeric TMP probably forms a long helical hexameric region anchored at both extremities of the tail. In SPP1, a similar density has been observed in the virion before infection but was absent after infection and DNA ejection through the tail channel (45). Although of lower resolution, the overall dimensions of the Araucaria tail components are in agreement with those of the SPP1 (45) or TP901-1 tail (14), not taking into account Araucaria's tail decorations.

(iii) HAD. The *Caudovirales* tail tip attaches the host adsorption device (HAD). In Araucaria, the HAD has a funnel shape of ~140 Å by 180 Å. The largest part of the funnel is located ~30 Å behind the tail junction and carries six bulbs of the dimensions 25 Å by 25 Å by 50 Å following a 6-fold symmetry (Fig. 7). Above the

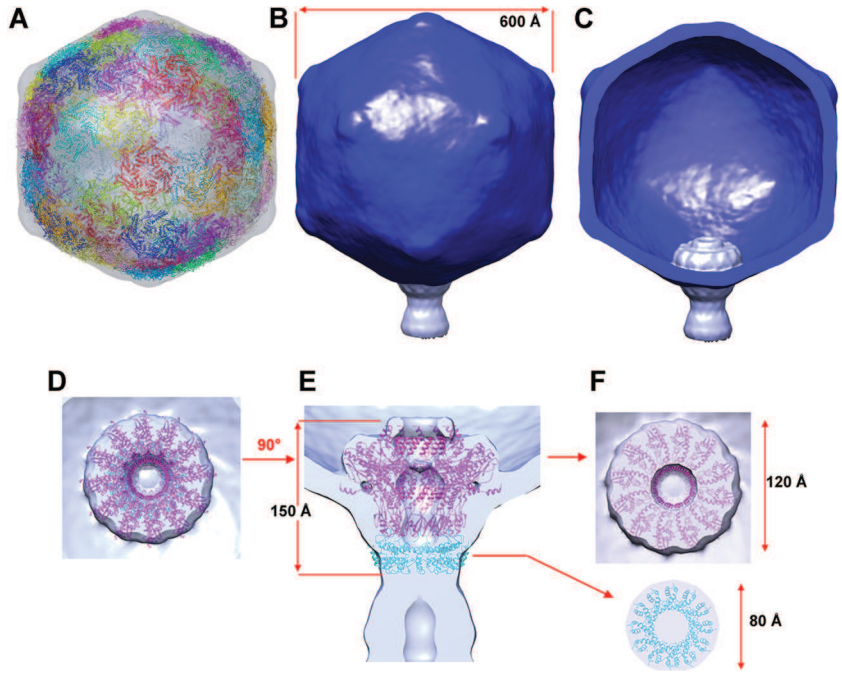


FIG 5 Reconstruction of Araucaria mycobacteriophage capsid and connector. (A) Surface rendering and pseudoatomic model of the Araucaria mature capsid icosahedral reconstruction viewed along an icosahedral 2-fold axis. The MCP was fitted into the capsid reconstruction using Chimera. (B) Capsid reconstruction and fitted reconstruction of the connector at a unique penton apex. (C) Cross-section of the capsid and connector, in the same orientation as in panel B. (D to F) EM reconstruction of the connector. (D) View of the connector's 12-fold averaged reconstruction from the capsid interior with the SPP1 portal fitted. (E) Side view of the connector reconstruction with the SPP1 portal and the first head completion protein dodecamer (SPP1 gp15) fitted. (F) Sliced views of the connector at the level of the upper portal (up) and of the first head completion protein (down).

TABLE 4 Sequence analysis of *M. abscessus* subsp. *bolletii* mycobacteriophage structural proteins

ORF protein	Annotation	FFASO3[a]		Hhpred			Pham
		Score	% identity	Probability	E value	*P* value	
ORF52	HK97 family phage portal protein	−6.44	10	88.8	2.9	9E−05	Pham 346
ORG55	Phage_Mu_F (minor head protein)						MHP
ORF57	MCP						Pham 3637
ORF62	MTP						Pham2299
ORF69	TMP						Pham 4154
ORF70	Dit *Lactococcus* phage tp901-1			71.63	7.8	0.00027	Pham 362
ORF71	Tal *Neisseria meningitidis* MC58	−12.3	15	98.66	9.4E−06	3.2E−10	Pham 363
ORF72	Putative HAD						Pham 3208
ORF73	Putative HAD						Pham 1371

[a] FFAS03 scores lower than −9.5 are considered significant.

largest part, the diameter of the HAD decreases to reach ~100 Å, the diameter of the MTP ring. Below it, the diameter regularly diminishes to reach 30 Å at the HAD lower part (Fig. 7A and B). HHpred searches revealed that the ORF protein following the TMP, ORF70, shares with ~80% probability structural similarity with the corresponding protein, Dit, in phage SPP1 and lactococcal phage TP901-1. The X-ray structures of Dit from phages SPP1 (gp19.1) (47), TP901-1 (ORF46 protein) (48), and p2 (ORF15 protein) (49) have been reported, and they were found to share close structural similarities. HHpred also predicted that the ORF71 protein shares structural similarity (98% similarity) with a prophage MUSO2 43-kDa tail protein (PDB 3CDD). This protein is structurally similar to phage T4 gp27 (50), to lactococcal phage p2 ORF16 protein (49), to SPP1 gp21 (51), and to TP901-1 Tal (ORF47 protein) (52). According to these predicted similarities, we then fit the phage p2 Dit-Tal complex (ORF15-ORF16) (49), a 1/1 assembly of Dit hexamer and Tal trimer, in the EM density of the Araucaria HAD (Fig. 7B and C). The fit was found to be excellent in the funnel upper part (cc > 0.90). Noteworthy, the p2 Dit (ORF15) arm and hand extension (49) was found to fit remarkably well with the positions of the HAD lateral bulbs (Fig. 7B and C). However, the Araucaria bulbs are larger than the p2 Dit "hands," which is in agreement with the longer sequence of the Araucaria Dit protein compared to the p2 one (478 versus 298 amino acids). The Tal proteins of Araucaria (ORF71 protein) and p2 (ORF15 protein) phages are of similar sizes (345 and 375 amino acids); however, Araucaria's Tal does not fill completely the lower part of the HAD funnel, leaving a large volume empty (~80 Å in length) (Fig. 7B). This volume might be accounted for by the ORF72 and/or ORF73 protein; both of these are in a position compatible with a HAD structural role (Fig. 7B).

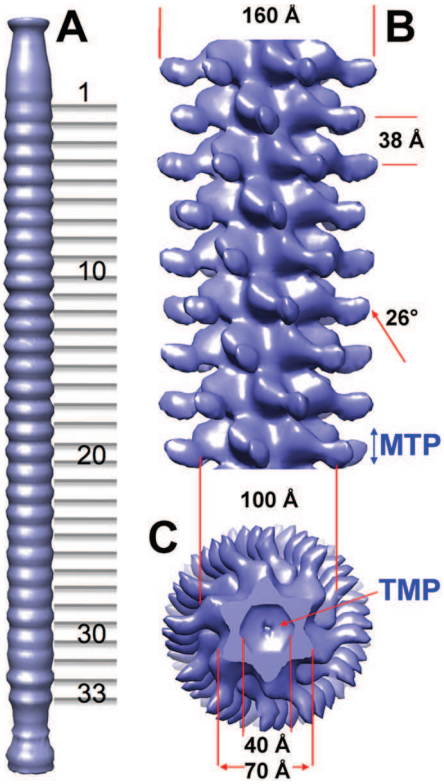


FIG 6 Reconstruction of the Araucaria mycobacteriophage tail. (A) Sixfold averaged reconstruction of the Araucaria phage tail from a few selected virions exhibiting an almost straight tail, making it possible to count the MTP rings. (B) Detailed view of the reconstruction of a segment of the tail (9 MTP rings) using helical symmetry. The helical parameters of the tail are shown. (C) Cross-section of the tail segment orthogonal to its long axis.

DISCUSSION

Genome organization. We have described here a temperate mycobacteriophage, Araucaria, isolated from *M. abscessus* subsp. *bolletii*, a multidrug-resistant emerging opportunistic pathogen that reveals a number of original insights into the diversity and evolution of bacteriophages. This suggests that the current collection of mycobacteriophages is far from being completely representative of the population of integrated mycobacteriophages.

The Araucaria prophage genome contains 17 genes (19%) that have bacterial homologues, notably among the *M. abscessus* group, which have no known function. These genes were therefore possibly acquired through horizontal transfer. The Araucaria genome contains all the genes required for lysogeny: those for an YqaJ recombinase (ORF16), a RecT recombinase (ORF17), and a RusA endodeoxyribonuclease (ORF19) similar to those of other mycobacteriophages. A notable feature of Araucaria is that its integrase gene is located far from the structural protein cluster.

In most siphophages, two proteins are required for cell lysis: holin, which form holes in the cell membrane, and lysin, which hydrolyzes the cell wall. ORF84 and ORF60 were identified as coding for lysin and holin, respectively, and the holin gene was found to be inserted among the structural genes. Araucaria's holin has two transmembrane helices and thus belongs to holin class II (53). The lysin gene is located on the right arm of the structural

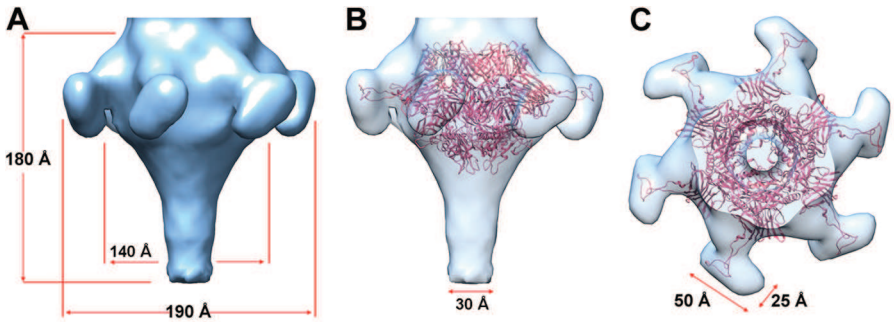


FIG 7 Reconstruction of the Araucaria host recognition device (HAD). (A) Sixfold averaged EM reconstruction of the HAD showing its funnel shape and the 6 bulbs of the protrusions. (B) Same view as in panel A, but with a Dit-Tal complex (ORF15-ORF16) from the phage p2 baseplate (red, 6Dit + 3Tal [49]) fitted in the EM map. (C) Same view as in panel B, rotated by 90°.

cassette. It is the only protein that shows sequence similarity with most of the mycobacteriophages sequenced so far, suggesting that it is highly conserved among them.

The terminase gene, transporting DNA into the proheads prior to attachment of the tail (54–56), is located close to the structural gene operon and far from the physical end of the genome, more than 20 kbp. This is also observed in the genomes of cluster A myobacteriophages (13, 30, 56–58).

The organization of the structural operon, encoding the virion proteins, is very similar to that of most other phages. The closest relatives of the Araucaria virion proteins are from mycobacteriophage Dori and cluster B mycobacteriophages, suggesting substantial lateral gene exchange among them. A few bacterial genes and a holin gene were observed to be inserted within the structural operon. Worth noticing, several examples of gene insertions within the structural gene operon are observed in the Wildcat or Corndog myobacteriophage (59). There are also examples of interruptions within the head genes of siphoviral phages, such as a large insertion between the head accessory protein and capsid protease genes in *Vibrio* phage SIO-2 (60). Little is known about the location of specific signals for gene expression in Araucaria or how these are regulated. However, Araucaria, like other mycobacteriophages such as TM4, encodes a WhiB-like protein (ORF32 protein) that may participate in Araucaria gene expression (61).

Overall structure of the Araucaria phage. We could assemble the complete EM structure of mycobacteriophage Araucaria from its four constituting modules, capsid, connector, tail, and host adsorption device, whose structures were solved independently (Fig. 8). The capsid and head-to-tail connector complex exhibit structures and overall dimensions very similar to those of other *Siphoviridae*, especially phages SPP1 (38, 45) and TP901-1 (14). The only difference between Araucaria and TP901-1 is the size of the stopper, much larger in the latter case. In contrast, the structure of the MTP rings is very peculiar. Noteworthy, the MTP (ORF62 protein) of Araucaria is much larger than the MTP of TP901-1 (352 versus 165 amino acids). The MTP rings of phage TP901-1 are about 110 Å in diameter, a value comparable to that of the inner core component of the Araucaria tail. It is likely that the exceeding ~190 MTP residues form Araucaria's tail protrusions. Such protrusions were already reported for phage λ and phage SPP1 (62, 63). In phage λ, the MTP (gpV) counts 246 amino acids. The N-terminal domain (gpV_N 1-159) forms the MTP rings (64), while residues 160 to 246 (gpV_C) belong to an Ig-like domain, probably involved with host cell wall saccharide interactions (63). In SPP1, the MTP-coding gene (*orf17.1*) is subject to a frameshift alternative reading, leading either to an MTP of 159 residues, the core, or to a longer form of 266 residues (62). Here again, the additional domain has a fold of an adhesin, probably involved in saccharide binding. However, this frameshift occurs in only 5 to 10% of the expressed MTPs, and the tail has not the hairy aspect of that of phage Araucaria. The Araucaria tail displays protrusions, which might play an accessory role, yet to be proved, in the interaction of the virion with the host cell wall through binding to its surface saccharides. Noteworthy, both phage λ and phage SPP1 attach to protein-specific receptors, LamB (65) and YueB (66), respectively. Hence, their binding to cell wall saccharides putatively provided by the tail protrusions might be only a first reversible step that maintains and orientates transitorily the phage before specific receptor recognition. Such a host cell

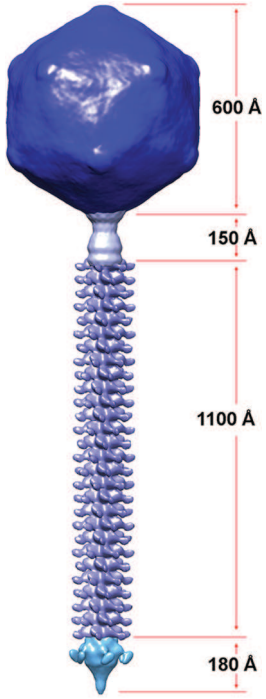


FIG 8 Assembled complete structure of Araucaria mycobacteriophage. The complete phage was assembled by fitting the individually refined reconstructions into the map obtained for the full phage.

wall scouting mechanism had a remarkable illustration with podophage T7 (67).

Most phages tail tips harbor a large macromolecular device, the head adsorption device (HAD). This macromolecular assembly displays different shapes in *Siphoviridae* according to the mechanism of attachment to the host. Phages, which attach to host's protein often, display a long, straight, element called "tail fiber" (45). In Gram-negative (Gram$^-$) bacteria (e.g., *Escherichia coli*), this is observed with phages binding to porins, such as phage T5 (68, 69) or phage λ. In Gram$^+$ bacteria, this tail fiber is observed in phage SPP1 (45, 66) and lactococcal phage c2 (70), which bind to extracellular components of the type 7 secretion system (T7SS), called, respectively, YueB (66) and PIP (70). In contrast, lactococcal phages of the 936 and P335 families seem to attach exclusively to polysaccharides and bear at their tail end a large macromolecular entity called the baseplate. The baseplate dimensions of phages p2 and TP901-1 are quite large, since these phages harbor a large number (18 to 54) of receptor binding proteins (RBPs), the specific proteins recognizing and attaching to the host (48, 49). These baseplates have been found, to date, to be composed of a central axial core formed by Dit and Tal, shared with the straight-tail tip phages (e.g., SPP1), and of a peripheral component formed by the RBPs as in phage p2 (49, 71) and eventually other proteins as in phages TP901-1 (48, 72) and Tuc2009 (73, 74).

Considering the above-described analysis, Araucaria should belong to the first class of phages, those with a straight tail tip. The Araucaria HAD is formed of Dit and Tal, plus other axial components, but is devoid of RBP-like structures as found in lactococcal phages p2 and TP901-1. Compared to SPP1, however, Araucaria possesses a very long Dit, 220 amino acids longer, with an extra domain appearing as a bulb in the EM density map. Araucaria's Dit is also longer, by 180 residues, than phage p2 Dit, which possesses a protrusion at the same position as Araucaria's bulb. However, neither BLAST nor HHpred returned hits related to such domains.

Araucaria's Tal possesses 345 amino acids, making it one of the shortest Tal sequences. While it covers most of the phage p2 Tal length, it corresponds only to the N-terminal domains of SPP1 and TP901-1 Tal, which are much longer, with 1,110 and 946 amino acids, respectively. The C-terminal domain of SPP1 has been found to be responsible for host YueB attachment, while in TP901-1, it is responsible for host cell wall hydrolysis. In Araucaria, the missing density could be assigned to ORF72 or -73. We suggest that Araucaria should bind to a protein receptor, since no evidence of a saccharide binding RBP presence is provided by the tail's tip shape or its sequence. It is tempting to speculate further, since mycobacteria possess a T7SS, that a component similar to T7SS YueB or PIP might serve as a receptor for Araucaria.

In spite of the great diversity of genetic mycobacteriophages and the mosaic nature of Araucaria, its homology with the mycobacteriophage Dori and those from cluster B suggests that Araucaria and these phages could share a similar structure and a similar mechanism of host infection. All together, our results provide evidence that the Araucaria mycobacteriophage shares building blocks similar to those of phages infecting Gram$^+$ or Gram$^-$ bacteria. They suggest also that Araucaria may infect its mycobacterial host by a two-step mechanism, first using reversible cell wall saccharide binding and then using irreversible binding to a cell wall emerging protein domain. The existence of such a mechanism, postulated for phages λ and SPP1, remains to be demonstrated.

ACKNOWLEDGMENTS

This work was supported by grants from the Agence Nationale de la Recherche (grants ANR-11-BSV8-004-01 "Lactophages") and by the Infectiopole Sud Foundation. The Chimera program, distributed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, was supported by NIGMS P41-GM103311.

REFERENCES

1. **Medjahed H, Gaillard JL, Reyrat JM.** 2010. Mycobacterium abscessus: a new player in the mycobacterial field. Trends Microbiol. **18:**117–123.
2. **Griffith DE.** 2011. The talking Mycobacterium abscessus blues. Clin. Infect. Dis. **52:**572–574.
3. **Haverkamp MH, van Wengen A, de Visser AW, van Kralingen KW, van Dissel JT, van de Vosse E.** 2012. Pulmonary Mycobacterium abscessus: a canary in the cystic fibrosis coalmine. J. Infect. **64:**609–612.
4. **Hongfei D, Xuerui H, Jing W, Naihui C.** 2012. Mycobacterium abscessus lung disease in a patient with previous pulmonary tuberculosis. Southeast Asian J. Trop. Med. Public Health **43:**959–963.
5. **Verregghen M, Heijerman HG, Reijers M, van Ingen J, van der Ent CK.** 2012. Risk factors for Mycobacterium abscessus infection in cystic fibrosis patients; a case-control study. J. Cyst. Fibros. **11:**340–343.
6. **Nessar R, Cambau E, Reyrat JM, Murray A, Gicquel B.** 2012. Mycobacterium abscessus: a new antibiotic nightmare. J. Antimicrob. Chemother. **67:**810–818.
7. **Adekambi T, Berger P, Raoult D, Drancourt M.** 2006. rpoB gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of Mycobacterium bolletii sp. nov., Mycobacterium phocaicum sp. nov. and Mycobacterium aubagnense sp. nov. Int. J. Syst. Evol. Microbiol. **56:**133–143.
8. **Hatfull GF.** 2010. Mycobacteriophages: genes and genomes. Annu. Rev. Microbiol. **64:**331–356.
9. **Hatfull GF.** 2012. Complete genome sequences of 138 mycobacteriophages. J. Virol. **86:**2382–2384.
10. **Hatfull GF.** 2012. The secret lives of mycobacteriophages. Adv. Virus Res. **82:**179–288.
11. **Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW.** 2010. Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. J. Mol. Biol. **397:**119–143.
12. **Brennan PJ.** 2003. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. Tuberculosis (Edinb.) **83:**91–97.
13. **Hatfull GF, Cresawn SG, Hendrix RW.** 2008. Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. Res. Microbiol. **159:**332–339.
14. **Bebeacua C, Lai L, Vegge CS, Brondsted L, van Heel M, Veesler D, Cambillau C.** 2013. Visualizing a complete Siphoviridae member by single-particle electron microscopy: the structure of lactococcal phage TP901-1. J. Virol. **87:**1061–1068.
15. **Choi GE, Cho YJ, Koh WJ, Chun J, Cho SN, Shin SJ.** 2012. Draft genome sequence of Mycobacterium abscessus subsp. bolletii BD(T). J. Bacteriol. **194:**2756–2757.
16. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: a fast phage search tool. Nucleic Acids Res. **39:**W347–W352.
17. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119.
18. **Tatusov RL, Galperin MY, Natale DA, Koonin EV.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28:**33–36.
19. **Laslett D, Canback B.** 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32:**11–16.
20. **Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD.** 2012. The Pfam protein families database. Nucleic Acids Res. **40:**D290–D301.
21. **Leplae R, Hebrant A, Wodak SJ, Toussaint A.** 2004. ACLAME: a CLAssification of Mobile genetic Elements. Nucleic Acids Res. **32:**D45–D49.
22. **Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ.**

2007. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. **157:**38–46.
23. **Shaikh TR, Gao H, Baxter WT, Asturias FJ, Boisset N, Leith A, Frank J.** 2008. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat. Protoc. **3:**1941–1974.
24. **Scheres SH.** 2010. Classification of structural heterogeneity by maximum-likelihood methods. Methods Enzymol. **482:**295–320.
25. **Scheres SH, Nunez-Ramirez R, Sorzano CO, Carazo JM, Marabini R.** 2008. Image processing for electron microscopy single-particle analysis using XMIPP. Nat. Protoc. **3:**977–990.
26. **van Heel M, Schatz M.** 2005. Fourier shell correlation threshold criteria. J. Struct. Biol. **151:**250–262.
27. **Owen CH, Morgan DG, DeRosier DJ.** 1996. Image analysis of helical objects: the Brandeis Helical Package. J. Struct. Biol. **116:**167–175.
28. **Egelman EH.** 2007. The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. J. Struct. Biol. **157:**83–94.
29. **Vellani TS, Myers RS.** 2003. Bacteriophage SPP1 Chu is an alkaline exonuclease in the SynExo family of viral two-component recombinases. J. Bacteriol. **185:**2465–2474.
30. **Ford ME, Stenstrom C, Hendrix RW, Hatfull GF.** 1998. Mycobacteriophage TM4: genome structure and gene expression. Tuber. Lung Dis. **79:**63–73.
31. **Martinez C, De Geus P, Lauwereys M, Matthyssens G, Cambillau C.** 1992. Fusarium solani cutinase is a lipolytic enzyme with a catalytic serine accessible to solvent. Nature **356:**615–618.
32. **Schué M, Maurin D, Dhouib R, N′Goma JCB, Delorme V, Lambeau G, Carriere F, Canaan S.** 2010. Two cutinase-like proteins secreted by Mycobacterium tuberculosis show very different lipolytic activities reflecting their physiological function. FASEB J. **24:**1893–1903.
33. **Mageeney C, Pope WH, Harrison M, Moran D, Cross T, Jacobs-Sera D, Hendrix RW, Dunbar D, Hatfull GF.** 2012. Mycobacteriophage Marvin: a new singleton phage with an unusual genome organization. J. Virol. **86:**4762–4775.
34. **Sampson T, Broussard GW, Marinelli LJ, Jacobs-Sera D, Ray M, Ko CC, Russell D, Hendrix RW, Hatfull GF.** 2009. Mycobacteriophages BPs, Angel and Halo: comparative genomics reveals a novel class of ultra-small mobile genetic elements. Microbiology **155:**2962–2977.
35. **Veesler D, Johnson JE.** 2012. Virus maturation. Annu. Rev. Biophys. **41:**473–496.
36. **Veesler D, Cambillau C.** 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol. Mol. Biol. Rev. **75:**423–433.
37. **Veesler D, Quispe J, Grigorieff N, Potter CS, Carragher B, Johnson JE.** 2012. Maturation in action: CryoEM study of a viral capsid caught during expansion. Structure **20:**1384–1390.
38. **White HE, Sherman MB, Brasiles S, Jacquet E, Seavers P, Tavares P, Orlova EV.** 2012. Capsid structure and its stability at the late stages of bacteriophage SPP1 assembly. J. Virol. **86:**6768–6777.
39. **Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE.** 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. Science **289:**2129–2133.
40. **Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE.** 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. **25:**1605–1612.
41. **Jaroszewski L, Li Z, Cai XH, Weber C, Godzik A.** 2011. FFAS server: novel features and applications. Nucleic Acids Res. **39:**W38–W44.
42. **Söding J, Biegert A, Lupas AN.** 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. **33:**W244–W248.
43. **Lebedev AA, Krause MH, Isidro AL, Vagin AA, Orlova EV, Turner J, Dodson EJ, Tavares P, Antson AA.** 2007. Structural framework for DNA translocation via the viral portal protein. EMBO J. **26:**1984–1994.
44. **Lhuillier S, Gallopin M, Gilquin B, Brasiles S, Lancelot N, Letellier G, Gilles M, Dethan G, Orlova EV, Couprie J, Tavares P, Zinn-Justin S.** 2009. Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. Proc. Natl. Acad. Sci. U. S. A. **106:**8507–8512.
45. **Plisson C, White HE, Auzat I, Zafarani A, Sao-Jose C, Lhuillier S, Tavares P, Orlova EV.** 2007. Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO J. **26:**3720–3728.
46. **Pedersen M, Ostergaard S, Bresciani J, Vogensen FK.** 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. Virology **276:**315–328.
47. **Veesler D, Robin G, Lichiere J, Auzat I, Tavares P, Bron P, Campanacci V, Cambillau C.** 2010. Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): a baseplate hub paradigm in gram-positive infecting phages. J. Biol. Chem. **285:**36666–36673.
48. **Veesler D, Spinelli S, Mahony J, Lichiere J, Blangy S, Bricogne G, Legrand P, Ortiz-Lombardia M, Campanacci V, van Sinderen D, Cambillau C.** 2012. Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. Proc. Natl. Acad. Sci. U. S. A. **109:**8954–8958.
49. **Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichiere J, van Heel M, Campanacci V, Moineau S, Cambillau C.** 2010. Structure of lactococcal phage p2 baseplate and its mechanism of activation. Proc. Natl. Acad. Sci. U. S. A. **107:**6852–6857.
50. **Kanamaru S, Leiman PG, Kostyuchenko VA, Chipman PR, Mesyanzhinov VV, Arisaka F, Rossmann MG.** 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature **415:**553–557.
51. **Goulet A, Lai-Kee-Him J, Veesler D, Auzat I, Robin G, Shepherd DA, Ashcroft AE, Richard E, Lichiere J, Tavares P, Cambillau C, Bron P.** 2011. The opening of the SPP1 bacteriophage tail, a prevalent mechanism in Gram-positive-infecting siphophages. J. Biol. Chem. **286:**25397–25405.
52. **Bebeacua C, Bron P, Lai L, Vegge CS, Brondsted L, Spinelli S, Campanacci V, Veesler D, van Heel M, Cambillau C.** 2010. Structure and molecular assignment of lactococcal phage TP901-1 baseplate. J. Biol. Chem. **285:**39079–39086.
53. **Barenboim M, Chang CY, Fdib Hajj Young R.** 1999. Characterization of the dual start motif of a class II holin gene. Mol. Microbiol. **32:**715–727.
54. **Black LW.** 1989. DNA packaging in dsDNA bacteriophages. Annu. Rev. Microbiol. **43:**267–292.
55. **Catalano CE, Cue D, Feiss M.** 1995. Virus DNA packaging: the strategy used by phage lambda. Mol. Microbiol. **16:**1075–1086.
56. **Gomathi NS, Sameer H, Kumar V, Balaji S, Dustackeer VN, Narayanan PR.** 2007. In silico analysis of mycobacteriophage Che12 genome: characterization of genes required to lysogenise Mycobacterium tuberculosis. Comput. Biol. Chem. **31:**82–91.
57. **Ford ME, Sarkis GJ, Belanger AE, Hendrix RW, Hatfull GF.** 1998. Genome structure of mycobacteriophage D29: implications for phage evolution. J. Mol. Biol. **279:**143–164.
58. **Sarkis GJ, Hatfull GF.** 1998. Mycobacteriophages. Methods Mol. Biol. **101:**145–173.
59. **Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WR, Jr, Hendrix RW, Hatfull GF.** 2003. Origins of highly mosaic mycobacteriophage genomes. Cell **113:**171–182.
60. **Baudoux AC, Hendrix RW, Lander GC, Bailly X, Podell S, Paillard C, Johnson JE, Potter CS, Carragher B, Azam F.** 2012. Genomic and functional analysis of Vibrio phage SIO-2 reveals novel insights into ecology and evolution of marine siphoviruses. Environ. Microbiol. **14:**2071–2086.
61. **Kormanec J, Homerova D.** 1993. Streptomyces aureofaciens whiB gene encoding putative transcription factor essential for differentiation. Nucleic Acids Res. **21:**2512.
62. **Auzat I, Droge A, Weise F, Lurz R, Tavares P.** 2008. Origin and function of the two major tail proteins of bacteriophage SPP1. Mol. Microbiol. **70:**557–569.
63. **Pell LG, Gasmi-Seabrook GM, Morais M, Neudecker P, Kanelis V, Bona D, Donaldson LW, Edwards AM, Howell PL, Davidson AR, Maxwell KL.** 2010. The solution structure of the C-terminal Ig-like domain of the bacteriophage lambda tail tube protein. J. Mol. Biol. **403:**468–479.
64. **Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR.** 2009. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. Proc. Natl. Acad. Sci. U. S. A. **106:**4160–4165.
65. **Charbit A, Werts C, Michel V, Klebba PE, Quillardet P, Hofnung M.** 1994. A role for residue 151 of LamB in bacteriophage lambda adsorption: possible steric effect of amino acid substitutions. J. Bacteriol. **176:**3204–3209.
66. **São-José C, Lhuillier S, Lurz R, Melki R, Lepault J, Santos MA, Tavares P.** 2006. The ectodomain of the viral receptor YueB forms a fiber that

triggers ejection of bacteriophage SPP1 DNA. J. Biol. Chem. **281:**11464–11470.
67. **Hu B, Margolin W, Molineux IJ, Liu J.** 2013. The bacteriophage t7 virion undergoes extensive structural remodeling during infection. Science **339:**576–579.
68. **Boulanger P, Jacquot P, Plancon L, Chami M, Engel A, Parquet C, Herbeuval C, Letellier L.** 2008. Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. J. Biol. Chem. **283:**13556–13564.
69. **Flayhan A, Wien F, Paternostre M, Boulanger P, Breyton C.** 2012. New insights into pb5, the receptor binding protein of bacteriophage T5, and its interaction with its Escherichia coli receptor FhuA. Biochimie **94:**1982–1989.
70. **Babu KS, Spence WS, Monteville MR, Geller BL.** 1995. Characterization of a cloned gene (pip) from Lactococcus lactis required for phage infection. Dev. Biol. Stand. **85:**569–575.
71. **Spinelli S, Desmyter A, Verrips CT, de Haard HJ, Moineau S, Cambillau C.** 2006. Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat. Struct. Mol. Biol. **13:**85–89.
72. **Spinelli S, Campanacci V, Blangy S, Moineau S, Tegoni M, Cambillau C.** 2006. Modular structure of the receptor binding proteins of Lactococcus lactis phages. The RBP structure of the temperate phage TP901-1. J. Biol. Chem. **281:**14256–14262.
73. **Mc Grath S, Neve H, Seegers JF, Eijlander R, Vegge CS, Brondsted L, Heller KJ, Fitzgerald GF, Vogensen FK, van Sinderen D.** 2006. Anatomy of a lactococcal phage tail. J. Bacteriol. **188:**3972–3982.
74. **Sciara G, Blangy S, Siponen M, Mc Grath S, van Sinderen D, Tegoni M, Cambillau C, Campanacci V.** 2008. A topological model of the baseplate of lactococcal phage Tuc2009. J. Biol. Chem. **283:**2716–2723.
75. **Krumsiek J, Arnold R, Rattei T.** 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics **23:**1026–1028.

Table S1: The functional details of predicted ORFs of Araucaria prophage

ORFs	Type	ORFs size bp	ORFs size aa	Predicted function	Homology with	E. value	% aa identity	pfam	cog family
1	CDS	1686	562	DNA/RNA non-specific endonuclease	hypothetical protein MA5S0422_3689 [Mycobacterium abscessus 5S-0422]	0.0	94%	gnl|CDD|150314	
2	CDS	762	254	integrase [Mycobacterium massiliense 1S-151-0930]	gp41 [Mycobacterium phage Spartacus]	2E-101	58%	gnl|CDD|144254	[L]
3	CDS	651	217	hypothetical protein	hypothetical protein MM1S1510930_3516 [Mycobacterium massiliense 1S-151-0930]	2E-146	94%		
4	CDS	435	145	hypothetical protein	gp57 [Mycobacterium phage Yoshi	2E_05	31%	gnl|CDD|144828	
5	CDS	231	77	hypothetical protein	gp58 [Mycobacterium phage Yoshi]	0,00000002	42%		[K]
6	CDS	288	96	hypothetical protein	gp41 [Mycobacterium phage Ardmore]	0,00000004	40%		
7	CDS	375	125	hypothetical protein	gp54 [Mycobacterium phage Dori]	0,0000005	43%		
8	CDS	327	109	hypothetical protein	hypothetical protein MM1S1510930_3511 [Mycobacterium massiliense 1S-151-0930]	9E-15	42%		
9	CDS	282	94	hypothetical protein	hypothetical protein MM1S1510930_3510 [Mycobacterium massiliense 1S-151-0930]	1E-55	93%		
10	CDS	372	124	hypothetical protein	gp41 [Mycobacterium phage Giles]	3E-36	62%		
11	CDS	192	64	hypothetical protein	hypothetical protein MM1S1510930_3508 [Mycobacterium massiliense 1S-151-0930]	7E-19	52%		
12	CDS	294	98	hypothetical protein	gp58 [Mycobacterium phage CrimD]	0,0000003	44%		
13	CDS	270	90	hypothetical protein	gp48 [Mycobacterium phage Redi]	3E-20	54%		
14	CDS	183	61	hypothetical protein	gp50 [Mycobacterium phage Redi]	7E-13	61%		
15	CDS	150	50	hypothetical protein	gp46 [Mycobacterium phage Babsiella]	0,000000001	70%		
16	CDS	1005	335	YqaJ recombinase domain	gp63 [Mycobacterium phage Dori]	0.0	79%	gnl|CDD|150301	[L]
17	CDS	1062	354	Recombinational DNA repair protein RecT (prophage associated)	gp49 [Mycobacterium phage Brujita]	6E-170	67%	gnl|CDD|146460	[L]

18	CDS	1068	356	cell division protein FtsK	gp50 [Mycobacterium phage Babsiella]	3E-121	54%	gnl|CDD|150161	[L]
19	CDS	429	143	putative resolvase/ Endodeoxyribonuclease RusA	gp73 [Mycobacteriophage MacnCheese]	5E-40	56%	gnl|CDD|147817	[L]
20	CDS	240	80	hypothetical protein	gp68 [Mycobacterium phage Dori]	3E-23	59%		
21	CDS	327	109	hypothetical protein	gp70 [Mycobacterium phage Dori]	2E-10	72%		
22	CDS	1518	506	DNA methylase [DNA replication, recombination, and repair]	gp69 [Mycobacterium phage Ramsey]	2E-170	58%		
23	CDS	402	134	hypothetical protein	hypothetical protein [Tuber melanosporum Mel28]	5.6	29%		
24	CDS	1119	373	DNA methylase N-4/N-6	gp61 [Mycobacterium phage DLane]	3E-114	52%		
25	CDS	129	43	hypothetical protein	hypothetical protein MM1S1510930_3496 [Mycobacterium massiliense 1S-151-0930]	1E-22	98%	gnl|CDD|144957	[L]
26	CDS	261	87	hypothetical protein	gp71 [Mycobacterium phage Baka]	7E-15	65%		
27	CDS	1011	337	hypothetical protein	hypothetical protein MM3A0810R_3184 [Mycobacterium abscessus 3A-0810-R]	0.0	84%		
28	CDS	327	109	hypothetical protein	gp8 [Mycobacterium phage Bxz1]	2E-22	46%		
29	CDS	366	122	hypothetical protein	gp74 [Mycobacterium phage Dori]	3E-57	79%		
30	CDS	591	197	hypothetical protein	hypothetical protein MA5S1215_0625 [Mycobacterium abscessus 5S-1215]	2E-75	92%		
31	CDS	354	118	hypothetical protein	gp57 [Mycobacterium phage Charlie]	3E-26	50%		
32	CDS	288	96	whiB-family transcriptional regulator	whiB-family transcriptional regulator [Mycobacterium massiliense 1S-151-0930]	4E-53	90%		
33	CDS	1701	567	hypothetical protein	hypothetical protein MM1S1510930_3487 [Mycobacterium massiliense 1S-151-0930]	0.0	95%		
34	CDS	264	88	hypothetical protein	hypothetical protein MM1S1520914_3692 [Mycobacterium massiliense 1S-152-0914]	2E-31	98%		
35	CDS	342	114	hypothetical protein	gp84 [Mycobacterium phage Dori]	7E-21	57%		
36	CDS	267	89	hypothetical protein	signal transduction histidine kinase [Sinorhizobium fredii NGR234]	0.41	43%		
37	CDS	225	75	hypothetical protein	hypothetical protein MM1S1510930_3483 [Mycobacterium massiliense 1S-151-0930]	0,000000001	56%		

38	CDS	363	121	hypothetical protein	gp85 [Mycobacterium phage Dori]	0%	58%		
39	CDS	216	72	hypothetical protein	hypothetical protein MA6G0125S_0396 [Mycobacterium abscessus 6G-0125-S]	0%	67%		
40	CDS	204	68	hypothetical protein	lipoprotein, releasing system, transmembrane protein, LolC/E family [Acinetobacter sp. WC-743	4.3	33%		
41	CDS	1863	621	resolvase domain-containing protein	hypothetical protein OUW_20551 [Mycobacterium abscessus M93]	0%	71%		
42	CDS	507	169	hypothetical protein/hypothetical aminoglycoside/multidrug efflux system	hypothetical protein OUW_20546 [Mycobacterium abscessus M93]	0%	78%		
43	CDS	546	182	PROBABLE LIPOPROTEIN LPPU	putative liporotein LppU [Mycobacterium massiliense 1S-151-0930]	0%	62%	gnl|CDD|143990	[L]
44	CDS	174	58	hypothetical protein	putative membrane protein [Mycobacterium massiliense 1S-151-0930]	5E-30	95%		
45	CDS	126	42	hypothetical protein	band 7 family membrane protein [Corynebacterium glucuronolyticum ATCC 51866]	7.8	35%		
46	CDS	996	332	putative dna modification methylase protein	putative dna modification methylase protein [Mycobacterium massiliense 1S-151-0930]	0.0	96%		
47	CDS	1527	509	Phage terminase, large subunit	gp3 [Mycobacterium phage Dori]	0.0	74%	gnl|CDD|144957	[L]
48	CDS	252	84	hypothetical protein	gp61 [Mycobacterium phage Phaedrus]	1E-18	56%	gnl|CDD|146059	
49	CDS	267	89	hypothetical protein	hydrolase [Pseudonocardia sp. P1]	2.8	33%		[S]
50	CDS	552	184	hypothetical protein	gp5 [Mycobacterium phage Dori]	3E-72	64%		
51	CDS	423	141	hypothetical protein	gp6 [Mycobacterium phage Dori]	5E-49	60%		
52	CDS	1851	617	portal protein	gp7 [Mycobacterium phage Dori]	0.0	59%		
53	CDS	162	54	hypothetical protein	gp65 [Mycobacterium phage Zemanar]	0.027	50%		
54	CDS	390	130	hypothetical protein	gp96 [Mycobacterium phage Spartacus]	1E-53	67%		
55	CDS	2355	785	minor head protein	gp8 [Mycobacterium phage Dori]	0.0	70%	gnl|CDD|152186	
56	CDS	177	59	hypothetical protein	gp9 [Mycobacterium phage Dori]	2E-14	55%	gnl|CDD|113023	
57	CDS	1782	594	Major Capsid protein	gp10 [Mycobacterium phage Dori]	0.0	55%		
58	CDS	771	257	hypothetical protein	gp16 [Mycobacterium phage Gadjet]	2E-77	49%		
59	CDS	534	178	hypothetical protein	gp12 [Mycobacterium phage Dori]	3E-31	43%		
60	CDS	261	87	holin	gp31 [Mycobacterium phage Fruitloop]	0,0003	38%		

61	CDS	249	83	hypothetical protein	hypothetical protein MM1S1510930_3457 [Mycobacterium massiliense 1S-151-0930]	3E-39	80%		
62	CDS	1056	352	Major tail subunit	gp17 [Mycobacterium phage Dori]	2E-111	66%		
63	CDS	1056	247	hypothetical protein	gp19 [Mycobacterium phage Dori]	3E-107	64%		
64	CDS	741	180	hypothetical protein	gp20 [Mycobacterium phage Dori]	7E-89	74%		
65	CDS	540	115	hypothetical protein	gp21 [Mycobacterium phage Dori]	0,00000005	58%		
66	CDS	345	108	Tail Assembly Chaperone	gp22 [Mycobacterium phage Dori]	1E-50	76%		
67	CDS	324	168	Tail Assembly Chaperone	gp27 [Mycobacterium phage Pipefish]	2E-32	51%		[R]
68	CDS	504	77	tape measure protein (phage rosebush)	gp28 [Mycobacterium phage Kamiyu]	6E-10	42%		
69	CDS	5379	1793	Phage tail length tape-measure protein	gp24 [Mycobacterium phage Dori]	0.0	55%	gnl|CDD|146000	
70	CDS	1434	478	Baseplate	gp25 [Mycobacterium phage Dori]	0.0	69%		
71	CDS	1035	345	Baseplate	gp26 [Mycobacterium phage Dori]	6E-131	57%		
72	CDS	1650	550	Baseplate	gp22 [Mycobacterium phage Halo]	4E-124	48%		
73	CDS	309	103	Baseplate	hypothetical protein OUW_20341 [Mycobacterium abscessus M93]	4E-62	89%		
74	CDS	222	74	hypothetical protein	hypothetical protein OUW_20346 [Mycobacterium abscessus M93]	5E-42	93%		
75	CDS	639	213	hypothetical protein	gp32 [Mycobacterium phage Dori]	8E-102	67%		
76	CDS	597	199	hypothetical protein	gp33 [Mycobacterium phage ChrisnMich]	5E-98	70%		
77	CDS	1278	426	hypothetical protein	gp34 [Mycobacterium phage Dori]	0.0	65%		
78	CDS	354	118	hypothetical protein	gp35 [Mycobacterium phage Dori]	2E-32	53%		
79	CDS	288	96	hypothetical protein	gp35 [Mycobacterium phage Dori]	1E-24	54%		
80	CDS	177	59	hypothetical protein	gp41 [Mycobacterium phage Harvey]	4E-11	56%		
81	CDS	696	232	hypothetical protein	gp37 [Mycobacterium phage Dori]	3E-61	53%		
82	CDS	156	52	hypothetical protein	hypothetical protein MM1S1510930_3437	8E-29	94%		
83	CDS	285	95	hypothetical protein	hypothetical protein MM1S1510930_3436 [Mycobacterium massiliense 1S-151-0930]	1E-53	94%		
84	CDS	1536	512	The gene TM4_gp29 (gp29) encodes a 58.6-kDa protein with the Zn-binding and amidase catalytic domains.mureinolytic enzyme	hypothetical protein TM4_gp29 [Mycobacterium phage TM4]	1E-116	46%	gnl|CDD|144925	
85	CDS	1164	388	hypothetical protein	gp53 [Mycobacterium phage Marvin]	5E-170	63%		

86	CDS	156	52	hypothetical protein	hypothetical protein MELLADRAFT_113625 [Melampsora larici-populina 98AG31]	0.65	43%		
87	CDS	303	101	hypothetical protein	gp44 [Mycobacterium phage Dori]	2E-22	55%		
88	CDS	717	239	hypothetical protein	hypothetical protein MA3A0119R_4120 [Mycobacterium abscessus 3A-0119-R	5E-94	62%		
89	CDS	1683	561	endonuclease	HNH endonuclease family protein [Mycobacterium abscessus 4S-0303]	3E-50	39%	gnl|CDD|148655	

Article N.4 **: Sassi M,** Gouret P, Chabrol O, Pantarotti P and Drancourt M**.**

Mycobacteriophage drived diversification: the *Mycobacterium abscessus* paradigm.

En cours de préparation.

Mycobacteriophage drived diversification :

The *Mycobacterium abscessus* paradigm

Mohamed Sassi[1], Philippe Gouret[2], Olivier Chabrol[2],

Pierre Pontarotti[2] and Michel Drancourt[1]*

Running title: *Mycobacterium abscessus* phages

[1] Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes UMR CNRS 6236 IRD198, IFR48, Institut Méditerranée Infection, Aix-Marseille-Université, Marseille, France

[2] Evolutionary biology and modeling, LATP UMR-CNRS 7353, Aix-Marseille Université, 13331, Marseille, France

*Corresponding author

E-mail: michel.drancourt@medecine.univ-mrs.fr

Abstract word count : **167**

Text word count: **2326**

Number of tables: **4**

Number of figures: **5**

Abstract

The analysis of *Mycobacterium abscessus* genomes including *M. abscessus*, *Mycobacterium bolletii* and *Mycobacterium massiliense* type strains found 1-8 prophage regions in 47/48 genomes ranging from small prophage-like elements to complete prophages. A total of 20,304 viral and phage proteins clustered into 853 orthologous groups comprised of 30% mycobacteriophages proteins, 48% other phages infecting gram positive and negative bacteria and 23% other viral proteins. Moreover, 44% of annotated proteins are implicated in DNA replication and metabolism and 37% are specific bacteriophage proteins including structural proteins. Phylogenomic and phylogentic analyses indicated that *M. abscessus* phages have a different evolutionary history than their mycobacterial hosts, which had been infected by at least ten different mycobacteriophages. In particular, 33 proteins occurred by horizontal transfer in the life of *M. abscessus*. These proteins are an integrase, specific mycobacteriophage proteins, hypothetical proteins and DNA replicaion and metabolism proteins. Gene exchanges, loss and gain were occurred in the specie *M. abscessus* by several mycobacteriophages. This first ever analysis of phage-mycobacterium co-evolution suggests that mycobactériophages play a role in *M. abscessus* diversifications.

Key words: *Mycobacterium abscessus, Mycobacterium bolletii, Mycobacterium massiliense*, prophages, mycobacteriophages

Introduction

Mycobacterium abscessus is an opportunistic pathogens (Ripoll et al., 2009) responsible for sporadic cases and outbreaks of skin and soft-tissue infections following surgical and cosmetic practices [Furuya et al. 2008, Koh et al. 2010, Viana-Niero, 2008]; catheter-related bacteremia [Petrini, 2006, Don Hayes, 2005]; and respiratory tract infections in patients with underlying lung disorders, particularly cystic fibrosis [Griffith DE et al., 1993; Pierre-Audrigier C et al., 2005, Oliver KN et al., 2003, Chalermstakulrat W et al., 2006, Jonsson eBE et al., 2007, Levy I et al., 2008 and Griffith DE et al., 2003]. These mycobacteria are broadly resistant to antibiotics and cure may require surgery (Griffith, D.E. et al., 2007).

There are a few evidences for mycobacteriophages in *M. abscessus* including a 81-kb prophage in the reference *M. abscessus* genome [Ripoll et al., 2009]. Also, a *M. abscessus* mycobacteriophage was isolated and its electron microscopy 3D structure was reselved [Sassi2013]. However, the repertoire of *M. abscessus* phages and their evolutionary history within this bacterium is unknown and no systematic exploration for prophages and mycobacteriophages has been performed among additionally available sequenced *M. abscessus* genomes, leaving undetermined whether these initial observation were unique or representative of the *M. abscessus*.

Here, exploiting available genome sequences by original bio-informatic analyses, we explored the repertoire of *M. abscessus* mycobacteriophages to gain insights into their evolution history among *M. abscessus.*

Materials and Methods

Establishing *M. abscessus* phage repertoire

The genomes of 48 *M. abscessus* strains were downloaded from Genbank (Table 1). As for 47 unfinished genomes, the contigs were reoriented based on *M. abscessus* type strain genome (GenBank GCF_000069185.1) as reference using MAUVE software [Darling AE et al., 2010]. The prophage regions were detected using PHAST software [Zhou Y et al., 2011]. Protein sequences were predicted using prodigal software [Hyatt D et al., 2010] in order to normalize prediction. *M. abscessus* pan-proteome was annotated using BlastP search with a cutoff E-value less than 0.001, percentage similarity more than 30% and an alignment length more than 50 amino acids against a home made database (including PHAST database, mimivirus, marseillevirus and additional mycobacteriophages proteins).

Determining *M. abscessus* phage phylogenies

M. abscessus proteomes were aligned using Mauve software [Darling AE et al., 2010]. Then *M. abscessus* tree was constructed using Neighbor-Net algorithm in the package SplitsTree4 [Huson DH et al., 2006]. Likewise, the annotated viral and phage proteins were concatenated for each genome and used for the construction of *M. abscessus* phage tree. The annotated viral and phage proteins were classified using OrthoMCL software [Li L et al., 2003]. Only protein sequences longer than 50-amino acid residues were considered for further analysis. Homologous sequences were selected using the all-against-all BlastP algorithm [Altschul F et al., 1997] with an E value of $<10^{-5}$. Then, clustering of the orthologous sequences was analyzed using the Markov Cluster algorithm [Van Dongen 2003]. The inflation index of 1.5 was used to regulate cluster tightness (granularity). The resulting orthologous groups were used to construct a whole-genome phylogenetic tree based on gene content. We generated a matrix of binary discrete characters ("0" and "1" for absence and presence, respectively) [Kalliopi G et al., 2011]. Using this matrix, we constructed a tree using the Neighbor-Net algorithm and a heatmap clusterization using R package [http://www.r-project.org/].

Detecting gene transfer events

The ortholog groups identified by OrthoMCL were submitted to PhyloPattern [Gouret P et al., 2009], a software library based on the Prolog language [Warren et al., 1997] for the automated analysis and manipulation of phylogenetic trees (within the DAGOBAH framework) Phuong Thi Le, et al., 2012]. The *M. abscessus* tree was used as a reference to infer topologies in order to detect gene gain and loss as previously described [Phuong Thi Le, et al., 2012]. The results were submitted to FIGENIX [Gouret P et al., 2005] for phylogenetic reconstruction within the DAGOBAH framework as described previously [Thi Le, et al., 2012]. These Horizontal Gene Transfer (HGT) events were detected using an in-house-built transfer filter called HGT agent [Phuong Thi Le, et al., 2012]. This filter uses PhyloPattern to annotate each internal duplication node of the tree with three tags, including the recipient species, the donor species and external species [Gouret P et al., 2009]. Then, This filter applies a an unique phyletic pattern to maximize the gene tree to find recipient species that are closer to donor species than to other external species that would otherwise be placed between the recipient and donor species in the species tree. As a result, a "donor" subtree only contains species of a specific group exclusively from the "recipient" group and vice versa. Using HG agent, one can specify the names of the donor species and the recipient species according to their usage.

Results

M. abscessus phage repertoire

Using PHAST software for prophage prediction, we found that *M. abscessus* M154 encodes no prophage regions whereas 47 genomes harbor one to eight prophage regions. A total of 171 predicted prophage regions could be separated into four types i) intact prophages encoding structural, lysis proteins, integration proteins and proteins necessary for replication and recombination ii) questionable prophages iii) incomplete prophage regions iiii) small prophage-like elements (Table 2).

Based on *M. abscessus* prophage regions homology, *M. abscessus* species could be separated into three clusters: *M. abscessus*, *M. massiliense* and *M. bolletii* (Figure 1). Few exceptions were observed; strain M139 and 1S_51_0915 strains showed prophage regions homology with *M. abscessus* cluster while *M. bolletii* M24 showed prophage regions homology with *M. massiliense* cluster (Figure 1). Interstingly, a 12-kb small prophage like element is conserved within the *M. abscessus* and *M. massiliense* cluster. We analyzed 242,067 proteins from 48 *M. abscessus* proteomes and we found a total of 20,304 (8.4%) proteins to be homologous to viral or phage proteins. these 20,304 proteins were grouped into 853 groups of orthologous proteins. All the species are represented in 239 groups (28.02%). Only three *M. abscessus* have

unique genes, two viral proteins in *M. abscessus* M94, four mycobacteriophage proteins in *M. abscessus* M159 and two viral proteins in *M. abscessus* M172. We found 30% of proteins to be homologous to mycobacteriophages proteins, 22% proteins to be homologous to proteins from phage infecting Gram-negative bacteria, 21% proteins to be homologous to proteins from phage infecting Gram-positive bacteria, 23% proteins to be homologous to viral proteins notably mimivirus and marseillevirus, 0.3 % proteins to be homologous to proteins from phage infecting archae (Figure 2). The annotation of the prophage found 44% to be implicated in DNA replication and bacterial or/and phage metabolism, 37% were annotated as bacteriophages proteins (including structural, integration and terminase) and 14% proteins remained with no functional annotation. Interstingly, 289 proteins were annotated as holin and 75 as lysin protein. Twenty-five *M. abscessus* genomes including type strains *M. abscessus*, *M. bolletii* and *M. massiliense* encode lysin proteins which belong to endolysin A and endolysin B families. Ta total of 37,3 % are homologous to lysin from mycobacteriophages. The remaining lysin proteins are homologous to lysin from phage infecting gram-positive bacteria (*bacillus* phages). Moreover, 156 proteins are repressor and anti-repressor proteins of the lambda repressor CI/C2 family (*Lactobacillus* phage and Staphylococcus phage), immunity repressor (*Bacillus* phage and *Geobacillus* phage) and Phage antirepressor protein KilAC domain (*Rhodococcus* phage).

M. bolletii genome encodes only one CI/C2 repressor homologous to the CI repressor from *Bacillus* phage and one putative repressors located out of Araucaria genome. While all other *M. abscessus* strains encode three to nine repressors. Only *M. abscessus* 6G genomes and *M. abscessus* type strain encode antirepressors.

Phylogenomic and phylogenetic analyses

The tree based on whole-genome content tree or *M. abscessus* treewas showed three clusters representing *M. abssessus, M. massiliense* and *M. bolletii.* However, contrary to what reported and in contrast to what would normally be expected M18 is not placed within the *M. massiliense* group. A phylogenomic prophage tree constructed based on gene content (i.e., the presence or absence of orthologous proteins) showed a species organization that was different from the *M. abscessus* tree (Figure 3). Rather, *M. massiliense* type strain clusters with *M. abscessus* group, while *M. abscessus* 4S species clusters with the *M. massiliense* 2B species and M159. This phylogenomic analysis shows that the *M. abscessus* viral and phage gene repertoires have different evolutionary histories. The phylogenomic prophage tree suggests that differential gene gain and loss and lateral gene acquisition may have played important roles in the evolution of some *M. abscessus* species. Also, the heatmap cluserization showed different species organization than *M. abscessus* species tree

suggesting that *M. abscessus* species may have been infected by several phages during their lifes (Figure 4). Like- wise, using the tree based on whole-genome content, individual phylogenetic analysis for the different orthologous proteins groups revealed many topologies that differ from that of the *M. abscessus* species. These results suggest that gene loss and HGT are relevant for all gene functions.

In order to estimate the number of phages infecting *M. abscessus*, a phylogenetic tree was constructed based on Tape measure proteins (TMP) (Figure 5). We could annotate of 135 TMP was found, and clustered using orthoMCL into ten groups. Interestingly, Araucaria TMP clusters with *M. massiliense* 1S and M172, suggesting that mycobacteriophages infecting this species may have similar features. The TMP based phylogenetic tree confirmed the clusterization of the heatmap and suggest that several mycobacteriophages infected *M. abscessus* species.

Detection of HGT cases

Among the 853 orthologous groups, phylogenetic tree were successfully reconstructed for 214 (25%) of the cases, out of which 33 (15.4%) cases were associated with strong boostrap support for HGT. Out of the 33 cases of HGT, 17 (51%) proteins were annotated as mycobacteriophage proteins, nine (27%) as implicated in DNA replicaion and metabolism, one (3%) integrase protein and six (18%) hypothetical proteins (Table 4). We found

15 proteins to be specific mycobacteriophage proteins (45 %), 16 to be phages infecting gram positive and negative bacteria (48 %) and two proteins to be mimivirus proteins (6%).

We identified 13 examples where *M. abscessus* species were probably gene donors in HGT events and 4 examples where groups of *M. abscessus* species where probably gene donors. *M. bolletii* type strain contributed only to one gene coding an ATP Binding Cassette transporter protein where's M115 strain contributed Five genes coding three bacteriophage proteins and two hypothetical proteins.

Discussion

Here updating PHAST-Database allowed depicting the co-evolution of mycobacteriophages within *M. abscessus*. A first aim of this study was to explore the repertoire of *M. abscessus* mycobacteriophages. Accordingly 171 prophage regions were observed in 47 genomes of *M. abscessus* being mosaics and sharing some homology between each other. Moreover, the annotation of *M. abscessus* viral and phage proteins showed homology to other mycobacteriophage proteins and phages infecting gram-negative and positive bacteria. The functional analyses of *M. abscessus* viral and phage proteins revealed the presence of proteins implicated on DNA replication and bacterial

and viral metabolism including ABC transporter. Also phage repressor proteins were identifed, as no CRISPRs found in *M. abscessus* genomes, *M. abscessus* phage repressors could serve to maintain their respective prophages in a quiescent state [Waldor MK and Friedman DI 2005]. Interestingly, *M. bolletii* encode two repressor proteins and has been reported to produce a mycobacteriophage [Sassi M et al., 2013]. Despite the presence of repressors, the mechanism of *M. bolletii* releasing mycobacteriophage unlike the other *M. abscessus* strains still unclear.

The phylogentic analyses of *M. abscessus* species tree shwoed the caracterization of three clusters *M. abscessus, M. bolletii* and *M. massiliense*. The *M. abscessus* phage trees showed that *M. abscessus* mycobacteriophages have a different evolutionary history than their hosts. These were confirmed by TMP protein sequence analyses which where reported to aid in mycobacteriophage classification [Smith KC et al., 2013]. Moreover, we could classify *M. abscessus* mycobacteriophages into at least ten clusters supporting the fact that *M. abscessus* were infected by at least ten different mycobacteriophages.

Interestingly, Araucaria prophage genome, the only reprted mycobacteriophage isolated from *M. bolleti* type strain gives no homology with other *M. abscessus* phage proteins [Sassi M et al., 2013]. Only *M. abscessus* 1S strains have few protein homologies

with Araucaria proteins. These were confirmed by phylogenetic studies using whole viral and phage proteins clusterization and TMP protein sequence based tree.

Finally, phages are reported to be implicated in lateral gene transfer [Canchaya C et al., 2003], and their production can be induced by antibiotics, as demonstrated in multidrug-resistant *Staphylococcus aureus* [Rolain JM et al, 2011]. HGT event were identified In *M. abscessus* and were occurred recently in the life of *M. abscessus*. The functional annotation of these HGT showed them to be integrase protein, mycobacteriophage proteins, hypothetical proteins and DNA replicaion and metabolism.

We previously observed that the diversity of *M. abscessus*, while are emerging pathogens, was indeed supported by the diversity of their mycobacteriophages [Sassi M et Drancourt M., 2013]. Interestingly, based on phylogentic analysis and prophage region homology we can separate *M. abscesus sensu lato* to three clusters *M. abscessus*, *M. massiliense* and *M. bolletii* with a clearly participation of myocbacteriophage in the evolution and diversity of thes mycobacteria.

Authors' contributions

MS, PG and OC performed the analyses. MS and MD designed the study. MS, PP and MD interpreted data and wrote the draft. All authors read and approved the final manuscript.

Acknowledgments

MS was financially supported by Infectiopole Sud Foundation.

Author details

1 Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UMR CNRS 7278, IRD 198, INSERM 1095. Faculté de médecine, 27, Boulevard Jean Moulin-Cedex 5, Marseille, France.

2 Evoolutionary biology and modeling, LATP UMR-CNRS 7353, Aix-Marseille University, 13331, Marseille, France

References

1. Adékambi T, Reynaud-Gaubert M, Greub G, Gevaudan MJ, La Scola B, et al: **Amoebal coculture of "*Mycobacterium massiliense*" sp. nov. from the sputum of a patient with hemoptoic pneumonia**. *J Clin Microbiol* 2004, **42**:5493–5501.

2. Adékambi T, Berger P, Raoult D, Drancourt M: ***rpo*B gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov.** *Int J Syst Evol Microbiol* 2006, **56**: 133–143.

3. Simmon KE, Brown-Elliott BA, Ridge PE, Durtschi JD, Bridge Mann L, Slechta ES, Steigerwalt AG, Moser BD, Whitney AM, Brown JM, Voelkerding KV, McGowan KL, Reilly AF, Kirn TJ, Butler WR, Edelstein PH, Wallace RJ Jr and Petti CA. ***Mycobacterium chelonae-abscessus* complex associated with sinopulmonary disease, northeastern USA.** *Emerging Infect*. Dis 2011, 17:1692-1700.

4. Leao SC, Tortoli E, Euzéby JP, Garcia MJ: **Proposal that *Mycobacterium massiliense* and *Mycobacterium bolletii* be united and reclassified as *Mycobacterium abscessus* subsp. *bolletii* comb. nov., designation of *Mycobacterium abscessus* subsp. *abscessus* subsp. nov. and emended description of *Mycobacterium abscessus*.** *Int J Syst Evol Microbiol* 2011, **61**:2311-3.

5. Furuya EY, Paez A, Srinivasan A, Cooksey R, Augenbraun M, Baron M, Brudney K, Della-Latta P, Estivariz C, Fischer S, Flood M, Kellner P, Roman C, Yakrus M, Weiss D, Granowitz EV. **Outbreak of *Mycobacterium abscessus* wound infections among "lipotourists" from the United States who underwent abdominoplasty in the Dominican Republic**. *Clin Infect Dis*. 2008 Apr 15;46(8):1181-8

6. Koh WJ, Jeon K, Lee NY, Kim BJ, Kook YH, Lee SH, Park YK, Kim CK, Shin SJ, Huitt GA, Daley CL, Kwon OJ. **Clinical significance of differentiation of *Mycobacterium massiliense* from *Mycobacterium abscessus*.** Am J Respir Crit Care Med. 2011 Feb 1;183(3):405-10. Epub 2010 Sep 10.

7. Hayes D Jr. ***Mycobacterium abscessus* and other nontuberculous mycobacteria: evolving respiratory pathogens in cystic fibrosis: a case report and review.** *South Med J*. 2005 Jun;98(6):657-61

8. Petrini B. ***Mycobacterium abscessus:*** **an emerging rapid-growing potential pathogen**. APMIS. 2006 May;114(5):319-28. Review.

9. Viana-Niero C, Lima KV, Lopes ML, Rabello MC, Marsola LR, Brilhante VC, Durham AM, Leão SC. **Molecular characterization of Mycobacterium massiliense and Mycobacterium bolletii in isolates collected from outbreaks of infections after laparoscopic surgeries and cosmetic procedures**. J Clin Microbiol. 2008 Mar;46(3):850-5. Epub 2008 Jan 3.

10. Griffith DE, Girard WM, Wallace RJ Jr: **Clinical features of pulmonary disease caused by rapidly growing mycobacteria. An analysis of 154 patients.** *Am Rev Respir Dis* 1993, **147**:1271–1278.

11. Pierre-Audigier C. et al: **Age-related prevalence and distribution of nontuberculous mycobacterial species among patients with cystic fibrosis.** *J Clin Microbiol.* 2005, **43**:3467–3470

12. Olivier KN et al: **Nontuberculous mycobacteria. I: multicenter prevalence study in cystic fibrosis.** *Am J Respir Crit Care Med* 2003, **167**:828–834.

13. Chalermskulrat W et al: **Non-tuberculous mycobacteria in end stage cystic fibrosis: implications for lung transplantation**. *Thorax* 2006, **61**:507–513.

14. Jonsson BE et al: **Molecular epidemiology of *Mycobacterium abscessus*, with focus on cystic fibrosis.** *J Clin Microbiol* 2007, **45**:1497– 1504.

15. Levy I et al: **Multicenter cross-sectional study of nontuberculous mycobacterial infections among cystic fibrosis patients, Israel.** *Emerg. Infect. Dis* 2008,**14**:378–384.

16. Griffith DE: **Emergence of nontuberculous mycobacteria as pathogens in cystic fibrosis.** *Am. J. Respir*. Crit. Care Med 2003, 167:810– 812.

17. Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, Rottman M, Macheras E, Heym B, Herrmann JL, Daffé M, Brosch R, Risler JL, Gaillard JL: **Non mycobacterial virulence genes in the genome of the emerging pathogen Mycobacterium abscessus**. *PLoS One* 2009, **19;4**(6):e5660.

18. Sassi M, Bebeacua C, Drancourt M, Cambillau C. **The First Structure of a Mycobacteriophage, the Mycobacterium abscessus subsp.**

bolletii Phage Araucaria. *J Virol*. 2013 Jul;87(14):8099-109.

19. Darling AE, Mau B, Perna NT. **progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement**. *PLoS One*. 2010 Jun 25;5(6):e11147.

20. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and translation initiation site identification**. *BMC Bioinformatics* 2010, **11:**119.

21. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. **PHAST: a fast phage search tool**. *Nucleic Acids Res*. 2011 Jul;39(Web Server issue):W347-52.

22. Li L, Stoeckert J, Christian J, Roos DS. OrthoMCL: **identification of ortholog groups for eukaryotic genomes**. *Genome Res*. 2003;13(9):2178–2189.

23. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res*. 1997;25(17):3389–3402.

24. Van Dongen S. **Graph clustering by flow simulation. PhD thesis**. University of Utrecht, The Netherlands 2000.

25. Huson D. H. and Bryant D., **Application of Phylogenetic Networks in Evolutionary Studies**, *Mol. Biol. Evol*., 23(2):254-267, 2006.

26. Georgiades K, Merhej M, El Karkouri K, Raoult D, Pontarotti P. **Gene gain and loss events in Rickettsia and Orientia species**. *Biology Direct*. 2011, 6:6

27. Gouret P, Thompson JD, Pontarotti P. **PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees.** *BMC Bioinf*. 2009;10:298.

28. Warren D, Pereira L, Pereira F. **Symposium on Artifical Intelligence and Programming Languages**. New York USA: SIGPLAN Notices; 1997. Prolog: the language and its implementation compared with Lisp; pp. 109–15.

29. Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EG. **FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform.** *BMC Bioinformatics*. 2005 Aug 5;6:198.

30. Thi Le P, Ramulu HG, Guijarro L, Paganini J, Gouret P, Chabrol O, Raoult D, Pontarotti P.**An automated approach for the identification of horizontal gene transfers from complete genomes reveals the rhizome of Rickettsiales**. *BMC Evol Biol*. 2012; 12: 243.

31. Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM, Petrova ZO, Dedrick RM, Pope WH; Science Education Alliance Phage Hunters Advancing Genomics And Evolutionary Science Sea-Phages Program, Modlin RL, Hendrix RW, Hatfull GF. **On the nature of mycobacteriophage diversity and host preference**. *Virology.* 2012 Dec 20;434(2):187-201.

32. Rolain JM, François P, Hernandez D, Bittar F, Richet H, Fournous G, Mattenberger Y, Bosdure E, Stremler N, Dubus JC, Sarles J, Reynaud-Gaubert M, Boniface S, Schrenzel J, Raoult D: **Genomic analysis of an emerging multiresistant *Staphylococcus aureus* strain rapidly spreading in cystic fibrosis patients revealed the presence of an antibiotic inducible bacteriophage.** *Biol Direct* 2009, **4**:1.

33. Loś JM, Loś M, Węgrzyn G: **Bacteriophages carrying Shiga toxin genes: genomic variations, detection and potential treatment of pathogenic bacteria.** *Future Microbiol* 2011, **6**(8):909-24.

34. Smith KC, Castro-Nallar E, Fisher JN, Breakwell DP, Grose JH, Burnett SH. **Phage cluster relationships identified through single gene analysis**. BMC Genomics. 2013 Jun 19;14:410.

35. Canchaya C, Proux C, Fournous G, Bruttin A, and Brussow H**. Prophage Genomics**. *Microbiology and molecular biology reviews*. June 2003, p. 238–276

36. Waldor MK, Friedman DI. **Phage regulatory circuits and virulence gene expression**. *Curr Opin Microbiol*. 2005 Aug;8(4):459-65.

Tables

Table 1: *M. abscessus* genomes used for this study

Table 2: *M. abscessus* species prophage regions

Table 3: *M. abscessus* HGT cases

Figures

Figure 1: Interacting Map based on *M. abscessus* prophage region homology. Blue: *M. abscessus* species. Red: *M. massiliense* species. Green: *M. bolletii* species.

Figure 2: *M. abscessus* viral and phage proteins annotation.

Figure 3: *M. abscessus* phylogeny. **A-** *M. abscessus* species split network tree. **B-** *M. abscessus* phage split network tree based on protein presence absence matrix.

Figure 4: heatmap clusterization based on protein presence absence matrix.

Figure 5: Phylogentic tree based on annotated tape measure protein sequences using neighbour joining method.

490 **Table 1:** *M. abscessus* genomes used for this study

Strain	Genome lenght Mb	Genome GC%
M154	4,8	64,1
2B-0107 M2B_0107	4,81	64,2
2B-0307 M2B_0307	4,81	64,2
2B-0912-R 2B_0912_R	4,81	64,2
2B-0912-S 2B_0912_S	4,81	64,2
2B-0626 M2B_0626	4,81	64,2
B-1231 M2B_1231	4,81	64,2
4S-0116-R 4S_0116_R	4,84	64
4S-0116-S 4S_0116_S	4,84	64
4S-0726-RA 4S_0726_RA	4,84	64
4S-0206 M4S_0206	4,86	64
4S-0303 4S_0303	4,86	64
4S-0726-RA 4S_0726_RB	4,86	64
M47J26	4,87	64,1
M18	4,89	64,2
1S_51_0915	4,89	64,2
1S-152-0930	4,9	64,2
1S_152_0914	4,9	64,2
M159	4,94	64,2
M115	4,98	64,1
M. bolletii BD	5,05	64,2
M93	5,08	64,2
M. abscessus.CIP104536T	5,09	62,7
M94	5,1	64,2
6G-0125-R 6G_0125_R	5,14	64,1
6G-0212 M6G_0212	5,14	64,1

Strain	Genome lenght Mb	Genome GC%
M. massiliense BD	5,2	64,2
M172	5,2	64,2
5S-1215 5S_1215	5,21	64,1
3A-0122-R 3A_0122_R	5,23	63,9
3A-0122-S 3A_0122_S	5,23	63,9
5S-0421 5S_0421	5,24	64,1
5S-1212 5S_1212	5,24	64,1
3A-0930-R 3A_0930_S	5,25	64
5S-0304 5S_0304	5,25	64,1
5S-0708 5S_0708	5,25	64,1
5S-0817 5S_0817	5,25	64,1
5S-0921 M5S_0921	5,25	64,1
3A-0930-R 3A_0930_R	5,27	64
3A-0119-R 3A_0119_R	5,28	63,8
3A-0810-R M3A_0810_R	5,29	64
5S-0422 5S_0422	5,32	64,1
6G-0728-S 6G_0728_S	5,32	64,1
6G-0125-S 6G_0125_S	5,33	64,1
6G-0728-R M6G_0728_R	5,34	64,1
6G-1108 6G_1108	5,34	64,1
3A-0731 3A_0731	5,39	64
M24	5,51	64,2

491

492 **Table 2:** *M. abscessus* species prophage regions

Strain	Genome size Mb	GC%	REGION	REGION_LENGTH Kb	COMPLETENESS	CDS	REGION_POSITION
M. abscessus CIP104536T	5.09	64.1	1	81	intact	110	1754551-1835095
M. bolletii BD	5.05	64.1	1	41.6	incomplete	47	1684736-1726377
			2	20.9	incomplete	38	1727918-1748849
			3	12.4	incomplete	16	3641720-3654182
M. massiliense BD	5.2	64.1	1	12.5	incomplete	21	1600973-1613514
			2	31.3	incomplete	33	1620002-1651385
			3	50.4	questionable	69	3907205-3957680
M24	5.51	63.8	1	37.1	questionable	51	560940-598047
			2	37	incomplete	37	1680197-1717263
			3	17	incomplete	21	3830340-3847343
			4	18.1	incomplete	34	5051771-5069955
			5	26	incomplete	35	5155113-5181190
			6	19.2	incomplete	26	5213195-5232444
			7	26.5	incomplete	33	5312024-5338593
M93	5.08	64.1	1	16.4	incomplete	33	197463-213867
			2	38	questionable	51	232006-270072
			3	53	incomplete	70	1762720-1815780
			4	20.2	incomplete	26	1820768-1841058
M94	5.1	64	1	58.3	questionable	84	1039523-1097850
			2	79.4	incomplete	99	4959719-5039151
M115	4.98	64.1	1	11.6	incomplete	10	1416841-1428481
			2	77.1	incomplete	102	1624644-1701770
			3	55.3	incomplete	79	3356346-3411651
M154	4.8	64.2	0	0	0	0	0
M159	4.94	64.2	1	16	incomplete	19	4919233-4935240
M172	5.2	63.9	1	55.1	questionable	74	502478-557677
			2	50.7	questionable	50	546109-596832

			3	59	questionable	67	1934186-1993225
			4	31.1	incomplete	33	2050376-2081567
			5	39.4	incomplete	45	3711805-3751246
			6	19.6	incomplete	40	3753466-3773078
M47J26	4.87	64.2	1	39.9	incomplete	48	1066714-1106668
			2	12.4	incomplete	16	3596408-3608873
			3	41.4	questionable	42	3823414-3864899
M18	4.89	64.2	1	62.8	questionable	67	4702725-4765592
3A-0119-R 3A_0119_R	5.28	64.1	1	41.1	intact	62	123714-164885
			2	43.6	questionable	56	408465-452128
			3	35.1	questionable	44	795576-830731
			4	20.1	incomplete	39	2834875-2855053
			5	38.2	questionable	45	2858522-2896821
			6	31.3	incomplete	9	2963183-2994488
			7	47.7	questionable	58	4372078-4419782
3A-0122-R 3A_0122_R	5.23	64.1	1	39.5	questionable	51	486333-525855
			2	31.3	incomplete	9	1122007-1153312
			3	20	incomplete	41	3430927-3450992
			4	38.1	questionable	47	3454461-3492645
			5	85	questionable	120	3968888-4053927
3A-0122-S 3A_0122_S	5.23	64.1	1	28.4	incomplete	29	727171-755621
			2	31.3	incomplete	9	855998-887303
			3	43.3	questionable	44	2261155-2304543
			4	17	incomplete	29	3132192-3149272
			5	40	questionable	58	3430676-3470735
			6	35.1	questionable	42	3775615-3810771
3A-0731 3A_0731	5.39	64.1	1	27.8	questionable	43	125845-153654
			2	11.8	incomplete	17	154064-165913
			3	34.9	questionable	50	1079517-1114497
			4	20.1	incomplete	40	1423251-1443430

			5	38.1	questionable	42	1446899-1485083
			6	31.3	incomplete	9	1551445-1582750
			7	34.8	questionable	43	3156535-3191432
			8	56	questionable	63	4018466-4074522
3A-0810-R M3A_0810_R	5.29	64.1	1	27.8	questionable	43	125781-153590
			2	11.8	incomplete	19	154000-165848
			3	37.9	questionable	50	408940-446855
			4	20.2	incomplete	40	1509208-1529448
			5	37.6	questionable	45	1532917-1570580
			6	31.1	incomplete	9	1636942-1668089
			7	45.5	questionable	63	3045844-3091362
			8	57	questionable	81	3918488-3975555
3A-0930-R 3A_0930_R	5.27	64.1	1	27.8	questionable	43	125821-153630
			2	11.8	incomplete	19	154040-165888
			3	39.9	questionable	51	409151-449145
			4	38.1	questionable	44	1529492-1567676
			5	31.3	incomplete	9	1634038-1665343
			6	22	incomplete	22	3042230-3064276
			7	34.1	questionable	43	3897212-3931331
			8	25.9	questionable	37	5246047-5271962
3A-0930-R 3A_0930_S	5.25	64.1	1	27.8	questionable	43	125833-153642
			2	11.8	incomplete	18	154052-165900
			3	20.1	incomplete	42	1162377-1182556
			4	38.1	questionable	44	1186025-1224209
			5	31.3	incomplete	9	1290571-1321876
			6	39.9	questionable	51	2252115-2292109
			7	43.4	questionable	58	3879712-3923155
			8	12.9	questionable	23	5234663-5247585
4S-0116-R 4S_0116_R	4.84	64.2	1	30.2	questionable	42	3246978-3277250
4S-0116-S 4S_0116_S	4.84	64.2	1	53.4	questionable	75	2294181-2347641

4S-0206 M4S_0206	4.86	64.2	1	31.1	incomplete	35	3235550-3266670
			2	30.2	questionable	42	3257124-3287396
4S-0303 4S_0303	4.86	64.2	1	31.1	incomplete	35	3218556-3249676
			2	30.2	questionable	42	3240130-3270402
4S-0726-RA 4S_0726_RA	4.84	64.2	1	53.4	questionable	76	3382701-3436159
4S-0726-RA 4S_0726_RB	4.86	64.2	1	53.4	intact	76	2746347-2799807
5S-0304 5S_0304	5.25	64	1	23.3	incomplete	16	1945826-1969168
			2	31.1	incomplete	33	2611298-2642403
			3	32.7	questionable	45	2630693-2663407
			4	28.2	incomplete	24	2683669-2711929
			5	20	incomplete	36	3151707-3171755
			6	46.4	questionable	41	3159765-3206257
5S-0421 5S_0421	5.24	64	1	27.2	incomplete	23	2193302-2220530
			2	41.8	questionable	59	2884649-2926459
			3	28.2	incomplete	22	2946721-2974981
			4	20	incomplete	36	3414759-3434783
			5	46.5	questionable	41	3422817-3469360
5S-0422 5S_0422	5.32	64	1	23.3	incomplete	16	2870816-2894158
			2	31.1	incomplete	33	3775349-3806454
			3	32.7	questionable	45	3794744-3827458
			4	28.2	incomplete	22	3847720-3875980
			5	20	incomplete	36	4315696-4335744
			6	46.5	questionable	40	4323754-4370297
5S-0708 5S_0708	5.25	64	1	31.1	incomplete	33	1676102-1707207
			2	32.7	questionable	46	1695497-1728211
			3	28.2	incomplete	25	1748473-1776733
			4	20	incomplete	36	2216449-2236497
			5	46.4	questionable	40	2224507-2270999
			6	27.2	incomplete	24	3919017-3946245
5S-0817 5S_0817	5.25	64	1	31.1	incomplete	33	2225104-2256209

			2	32.7	questionable	45	2244499-2277213
			3	28.2	incomplete	23	2297475-2325735
			4	20	incomplete	36	2765451-2785499
			5	46.5	questionable	42	2773509-2820052
			6	27.1	incomplete	23	4855708-4882907
5S-0921 M5S_0921	5.25	64	1	23.3	incomplete	14	1949994-1973336
			2	28.2	incomplete	30	3559204-3587488
			3	35.9	questionable	49	3575385-3611313
			4	28.2	incomplete	23	3629712-3657972
			5	20	incomplete	36	4097750-4117774
			6	46.4	questionable	41	4105808-4152300
5S-1212 5S_1212	5.24	64	1	31.1	incomplete	33	2383494-2414599
			2	32.7	questionable	45	2402889-2435603
			3	28.2	incomplete	21	2455865-2484125
			4	20	incomplete	36	2923840-2943888
			5	46.5	questionable	40	2931898-2978441
			6	19.4	incomplete	10	4254031-4273475
5S-1215 5S_1215	5.21	64	1	53.8	questionable	45	536941-590818
			2	13	incomplete	24	582131-595169
			3	31.1	incomplete	33	2146803-2177908
			4	32.7	questionable	45	2166198-2198912
			5	28.2	incomplete	22	2217311-2245571
			6	23.3	incomplete	16	4349482-4372824
6G-0125-R 6G_0125_R	5.14	64.1	1	54.2	questionable	69	66052-120303
			2	46.2	incomplete	35	798744-844965
6G-0125-S 6G_0125_S	5.33	64.1	1	54.2	questionable	69	1003141-1057392
			2	46.2	incomplete	36	1736529-1782750
6G-0212 M6G_0212	5.14	64.1	1	54.2	questionable	68	1070515-1124766
			2	46.2	incomplete	35	1803589-1849810
6G-0728-R M6G_0728_R	5.34	64.1	1	54.2	questionable	69	1004783-1059034

			2	64.9	questionable	44	1723225-1788224
6G-0728-S 6G_0728_S	5.32	64.1	1	64.9	questionable	44	716276-781275
			2	54.2	questionable	69	1354114-1408365
6G-1108 6G_1108	5.34	64.1	1	54.2	questionable	69	1004037-1058288
			2	46.2	incomplete	35	1737313-1783534
1S-152-0930	4.9	64.2	1	41.6	incomplete	47	2537516-2579135
			2	15.8	incomplete	32	2581230-2597119
			3	12.4	incomplete	16	2867809-2880274
1S_152_0914	4.9	64.2	1	41.6	incomplete	47	3559402-3601021
			2	15.8	incomplete	32	3603116-3619005
			3	12.4	incomplete	16	3889695-3902160
1S_51_0915	4.89	64.2	1	41.6	incomplete	47	2896041-2937660
			2	15.8	incomplete	31	2939755-2955644
			3	14	incomplete	9	3846091-3860097
2B-0107 M2B_0107	4.81	64.2	1	40	incomplete	48	325006-365014
			2	12.4	incomplete	16	2868969-2881434
2B-0307 M2B_0307	4.81	64.2	1	40	incomplete	49	324795-364803
			2	12.4	incomplete	16	2866959-2879424
2B-0912-R 2B_0912_R	4.81	64.2	1	39.9	incomplete	50	985086-1025040
			2	12.4	incomplete	16	3527642-3540107
2B-0912-S 2B_0912_S	4.81	64.2	1	39.9	incomplete	50	1403310-1443264
2B-0626 M2B_0626	4.81	64.2	1	39.9	incomplete	51	984096-1024050
			2	12.4	incomplete	16	3525970-3538435
2B-1231 M2B_1231	5.2	63.9	1	39.9	incomplete	48	1046358-1086336
			2	12.4	incomplete	16	3588716-3601181

193

194

495 **Table 3:** *M. abscessus* HGT cases

Putatve HGT	Homology	Parent Species	Recipient Species	Donnor Species	Nb Duplications before parent
1	PHAGE_Mycoba_Peaches-gi|282598664|ref|YP_003358761.1|gp58[Mycobacterium_phage_Peaches]	[M93]	[6G-0125-R]	[M93]	5
2	PHAGE_Plankt_PaV_LD-gi|371496158|ref|YP_004957306.1|ABCtransporter[PlanktothrixphagePaV-LD]	[PSEUDO~-5S-0421]	[M94]	[M93 M115]	4
3	PHAGE_Mycoba_LeBron-gi|304360967|ref|YP_003857149.1|gp18[Mycobacterium_phage_LeBron]	[M115]	[M172]	[M115]	2
4	PHAGE_Mycoba_Giles-gi|160700672|ref|YP_001552352.1|gp23[Mycobacterium_phage_Giles]	[M94]	[6G-0728-R]	[M94]	5
5	PHAGE_Mycoba_Che9c-gi|29566118|ref|NP_817687.1|gp10[Mycobacterium_phage_Che9c]	[M18]	[3A-0122_S]	[M18]	2
6	PHAGE_Mycoba_Pukovnik-gi|192824238|ref|YP_001994879.1|gp62[Mycobacterium_phage_Pukovnik]	[M154]	[M115]	[M154]	5
7	PHAGE_Tricho_2c-gi|116326757|ref|YP_803294.1|hypotheticalproteinTNAV2c_gp071[Trichoplusia_ni_ascovirus_2c]	[5S-0921]	[4S-0726]	[5S-0921]	4
8	PHAGE_Salmon_PVP_SE1-gi|363539742|ref|YP_004894027.1|hypotheticalprotein[SalmonellaphagePVP-SE1]	[M. massiliense T]	[M24]	[M. massiliense T]	5

9	PHAGE_Rhodoc_REQ3-gi|372449972|ref|YP_005087193.1|phageintegrase[RhodococcusphageREQ3]	[M172]	[3A_0930_S]	[M172]	3
10	PHAGE_Salmon_PVP_SE1-gi|363539618|ref|YP_004893903.1|phosphoribosylpyrophosphatesynthetase[SalmonellaphagePVP-SE1]	[3A-0122_S1]	[M. massiliense T]	[3A-0122_S1]	5
11	PHAGE_Mycoba_Myrna-gi|203454746|ref|YP_002225062.1|gp183[Mycobacterium_phage_Myrna]	[M172]	[M94]	[M172]	5
12	PHAGE_Mycoba_Omega-gi|29566822|ref|NP_818386.1|gp85[Mycobacterium_phage_Omega]	[M115]	[M24]	[M115]	5
13	PHAGE_Mycoba_Pacc40-gi|206600097|ref|YP_002241602.1|gp18[Mycobacterium_phage_Pacc40]	[M24]	[M94]	[M24]	6
14	PHAGE_Mycoba_Pacc40-gi|206600097|ref|YP_002241602.1|gp18[Mycobacterium_phage_Pacc40]	[PSEUDO~-M159]	[M24]	[M172 3A-0122_S6 3A-0122_S2 3A-0122_S4 47J26]	1
15	PHAGE_Acanth_mimivirus-gi|311977570|ref|YP_003986690.1|DNAtopoisomerase1b[Acanthamoebapolyphagamimivirus]	[3A-0122_S7]	[3A-0122_S5]	[3A-0122_S7]	2
16	PHAGE_Rhodoc_RER2-gi|372449922|ref|YP_005087145.1|hypotheticalprotein[RhodococcusphageRER2]	[M24]	[3A-0731]	[M24]	3
17	PHAGE_Aeromo_31-gi|66391812|ref|YP_238737.1|hypotheticalproteinPHG31p8[Aeromonas_phage_31]	[4S-0726]	[4S-0303]	[4S-0726]	6
18	PHAGE_Lactoc_P087-gi|229605000|ref|YP_002875699.1|putativecysteinesynthase[Lactococcus_phage_P087]	[47J26]	[5S-0708]	[47J26]	6

19	PHAGE_Mycoba_Myrna-gi|203454746|ref|YP_002225062.1|gp183[Mycobacterium_phage_Myrna]	[M172]	[M94]	[M172]	5
20	PHAGE_Plankt_PaV_LD-gi|371496158|ref|YP_004957306.1|ABCtransporter[PlanktothrixphagePaV-LD]	[3A-0122_S5]	[M94]	[3A-0122_S5]	2
21	PHAGE_Acanth_mimivirus-gi|311977513|ref|YP_003986633.1|putativedTDP-D-glucose4,6-dehydratase[Acanthamoebapolyphagamimivirus]	[3A-0122_S4]	[4S-0726-RA]	[3A-0122_S4]	4
22	PHAGE_Bacill_36-gi|156564011|ref|YP_001429750.1|PcrAhelicase[Bacillus_phage_0305phi8_36]	[3A-0122_S7]	[3A-0119-R]	[3A-0122_S7]	3
23	PHAGE_Mycoba_Che9c-gi|29566174|ref|NP_817745.1|gp68[Mycobacterium_phage_Che9c]	[M18]	[M94]	[M18]	3
24	PHAGE_Mycoba_Che8-gi|29565783|ref|NP_817355.1|gp17[Mycobacterium_phage_Che8]	[M24]	[M172 3A-0122_S4 3A-0122_S2 47J26]	[M24]	2
25	PHAGE_Tricho_2c-gi|116326757|ref|YP_803294.1|hypotheticalproteinTNAV2c_gp071[Trichoplusia_ni_ascovirus_2c]	[M115]	[4S-0116_S]	[M115]	3
26	PHAGE_Microm_MpV1-gi|313768434|ref|YP_004062114.1|hypotheticalprotein[Micromonassp.RCC1109virusMpV1]	[M115]	[M94]	[M115]	3
27	PHAGE_Mycoba_Pipefish-gi|109521870|ref|YP_655307.1|gp30[Mycobacterium_phage_Pipefish]	[M18]	[M172]	[M18]	1
28	PHAGE_Plankt_PaV_LD-gi|371496158|ref|YP_004957306.1|ABCtransporter[PlanktothrixphagePaV-LD]	[M. bolletii T]	[4S-0726-RA]	[M. bolletii T]	7

29	PHAGE_Lactoc_P087-gi|229605000|ref|YP_002875699.1|putativecysteinesynthase[Lactococcus_phage_P087]	[M24]	[M. bolletii T]	[M24]	7
30	PHAGE_Mycoba_Omega-gi|29566768|ref|NP_818332.1|gp31[Mycobacterium_phage_Omega]	[M94]	[6G-1108]	[M94]	3
31	PHAGE_Burkho_phi1026b-gi|38707948|ref|NP_945089.1|gp58[Burkholderia_phage_phi1026b]	[PSEUDO~-M154]	[M. bolletii T]	[M. massiliense T M154 M172 M159]	6
32	PHAGE_Mycoba_Cjw1-gi|29565933|ref|NP_817504.1|gp55[Mycobacterium_phage_Cjw1]	[PSEUDO~-M159]	[M. bolletii T]	[M172 3A-0122_S0 1S-152-0930 3A-0122_S1]	2
33	PHAGE_Brocho_BL3-gi|327409421|ref|YP_004301563.1|gp29[Brochothrixphage BL3]	[M115]	[M24]	[M115]	5

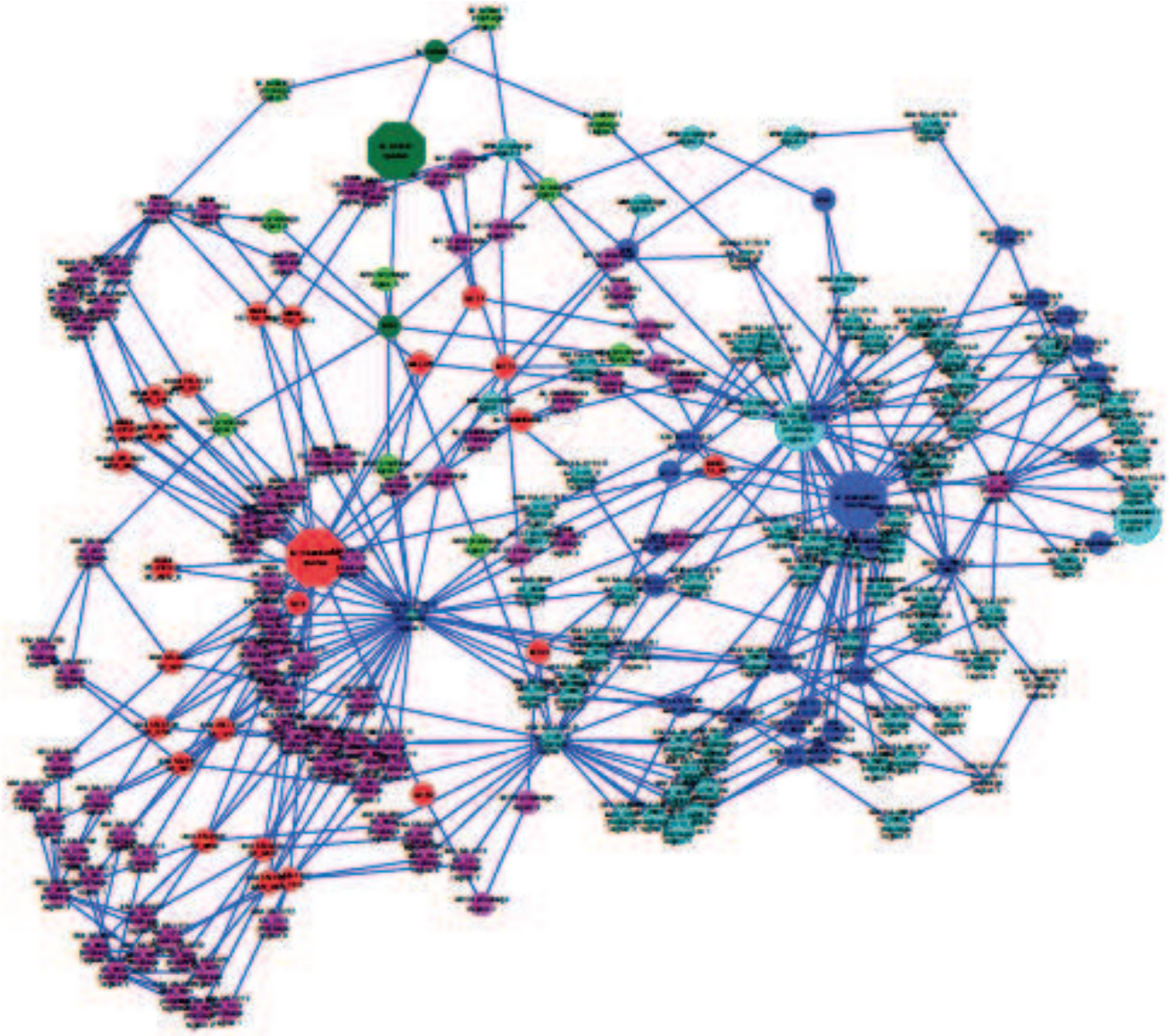


Figure 1: Interacting Map based on *M. abscessus* prophage region homology. Blue: *M. abscessus* species. Red: *M. massiliense* species. Green: *M. bolletii* species.

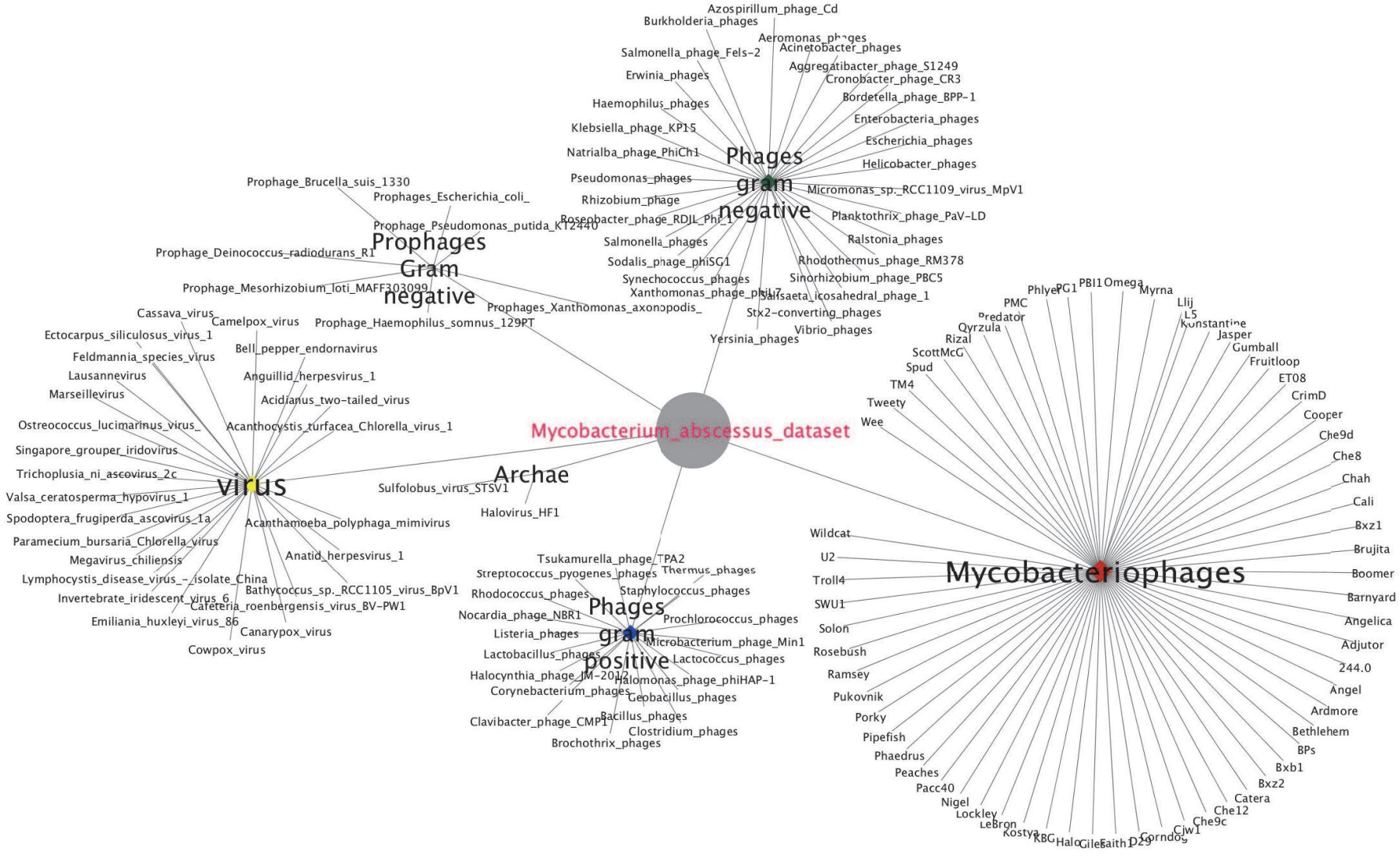
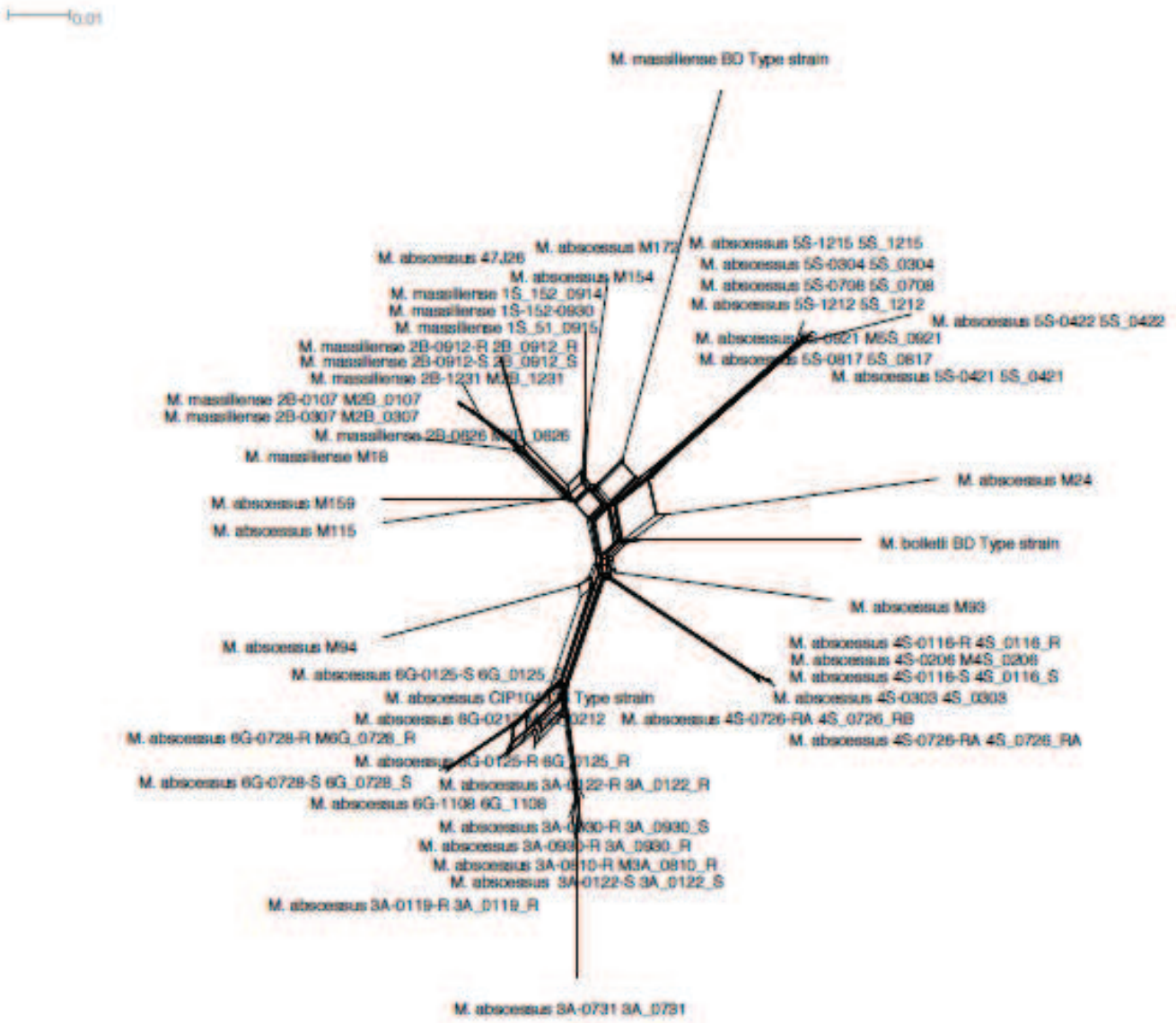
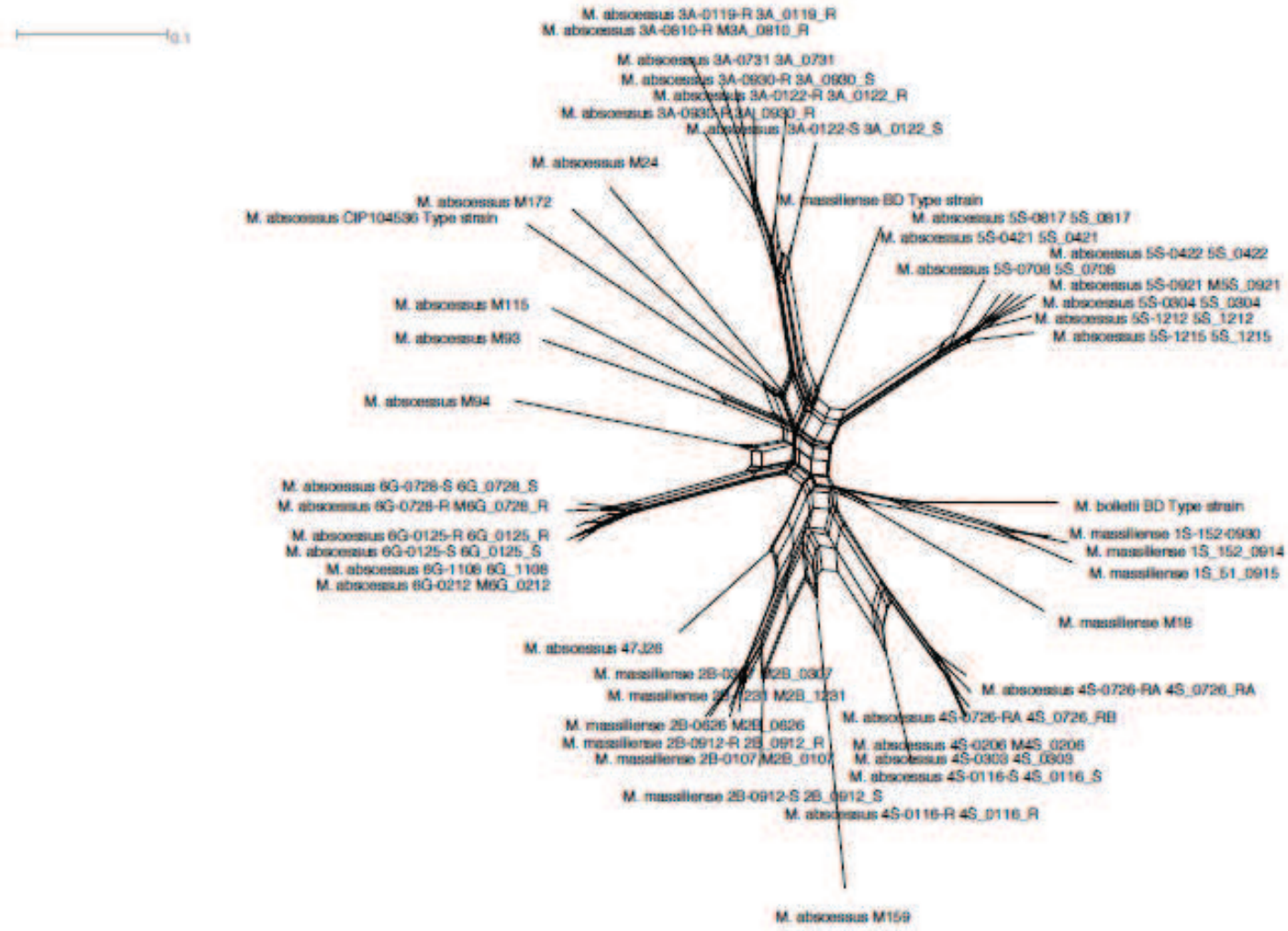


Figure 2: *M. abscessus* viral and phage proteins annotation.


0.01
M. massiliense BD Type strain
M. abscessus 47J26
M. abscessus M172
M. abscessus 5S-1215 5S_1215
M. abscessus M154
M. abscessus 5S-0304 5S_0304
M. massiliense 1S_152_0914
M. abscessus 5S-0708 5S_0708
M. massiliense 1S-152-0930
M. abscessus 5S-1212 5S_1212
M. massiliense 1S_51_0915
M. abscessus 5S-0422 5S_0422
M. massiliense 2B-0912-R 2B_0912_R
M. massiliense 2B-0912-S 2B_0912_S
M. massiliense 2B-1231 M2B_1231
M. massiliense 2B-0107 M2B_0107
M. massiliense 2B-0307 M2B_0307
M. massiliense M18
M. abscessus M24
M. abscessus M159
M. abscessus M115
M. bolletii BD Type strain
M. abscessus M93
M. abscessus M94
M. abscessus 4S-0116-R 4S_0116_R
M. abscessus 4S-0206 M4S_0206
M. abscessus 4S-0116-S 4S_0116_S
M. abscessus 4S-0303 4S_0303
M. abscessus 4S-0726-RA 4S_0726_RB
M. abscessus 4S-0726-RA 4S_0726_RA
M. abscessus 6G-0728-R M6G_0726_R
M. abscessus 6G-0728-S 6G_0728_S
M. abscessus 3A-0122-R 3A_0122_R
M. abscessus 6G-1108 6G_1108
M. abscessus 3A-0930-R 3A_0930_S
M. abscessus 3A-0930-R 3A_0930_R
M. abscessus 3A-0810-R M3A_0810_R
M. abscessus 3A-0122-S 3A_0122_S
M. abscessus 3A-0119-R 3A_0119_R
M. abscessus 3A-0731 3A_0731

A-



B-

Figure 3: *M. abscessus* phylogeny. **A-** *M. abscessus* species split network tree. **B-** *M. abscessus* phage split network tree based on protein presence absence matrix.

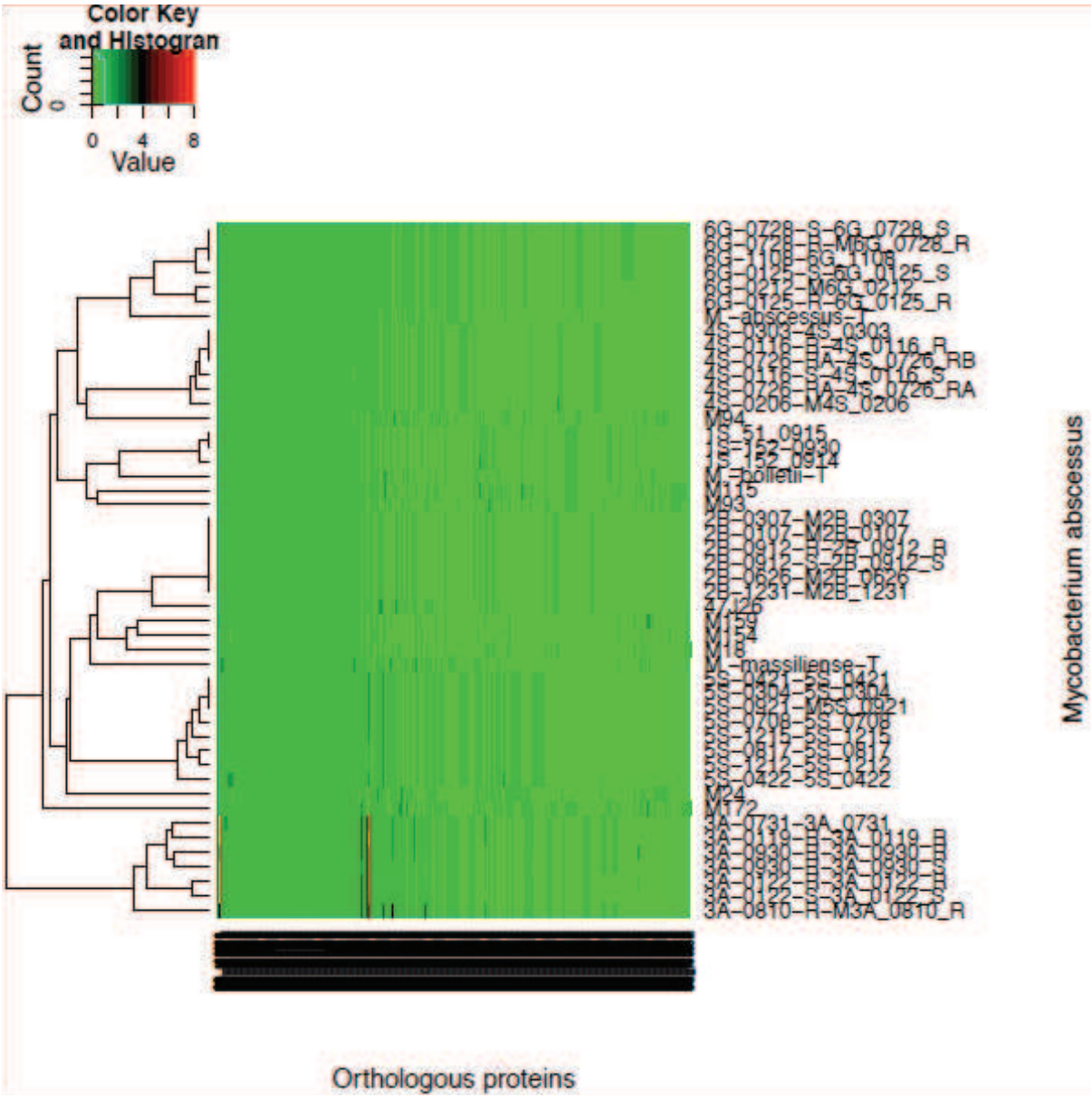


Figure 4: heatmap clusterization based on protein presence absence matrix.

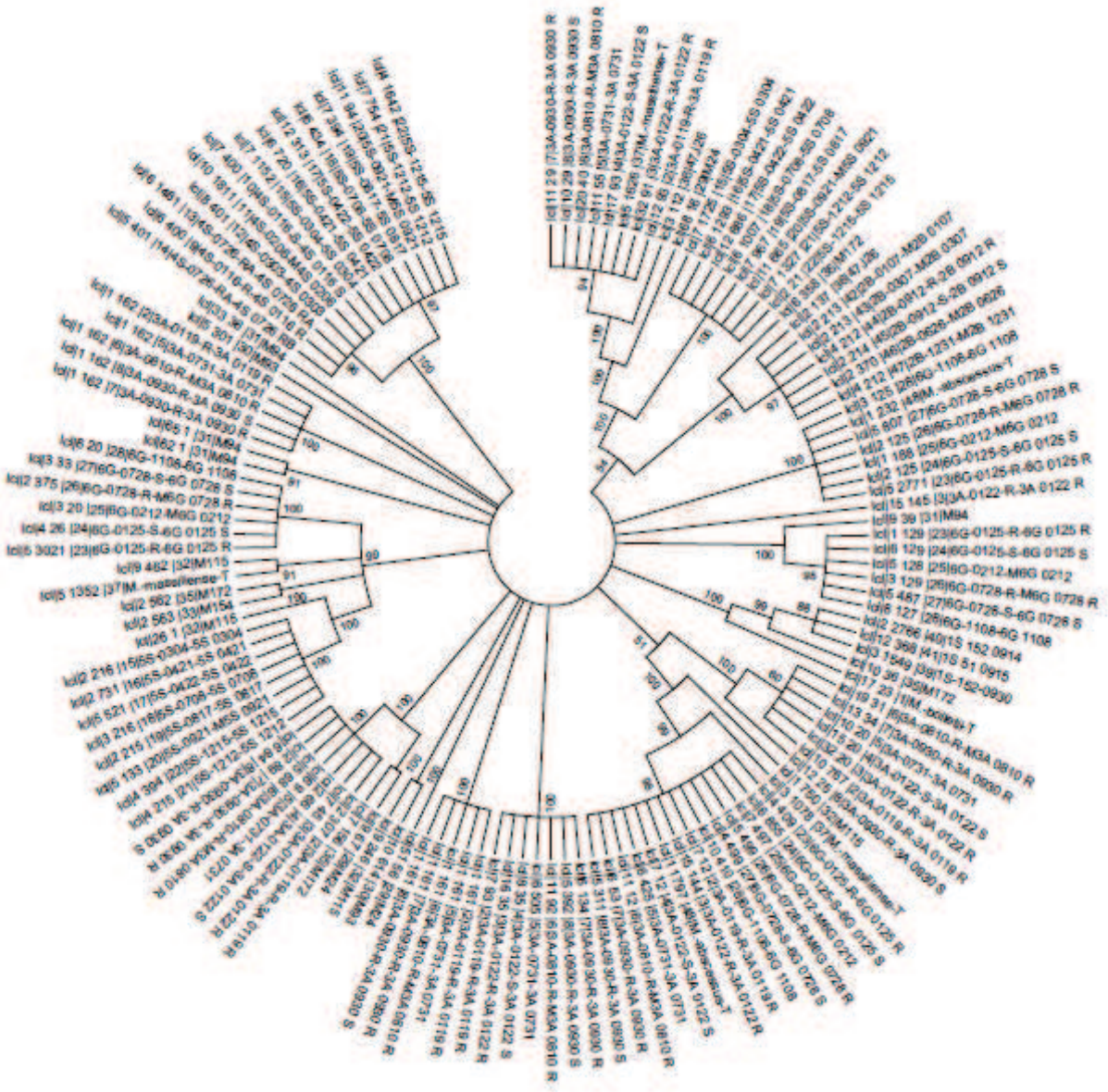


Figure 5: Phylogentic tree based on annotated tape measure protein sequences using neighbour joining method.

IV- DISCUSSION

Les données présentées dans cette thèse indiquent que les souches de *M. abscessus* sont plus diversifiées que ce que laisse entendre la nomenclature actuelle. Nous avons montré l'existence de cinq clusters qui, en plus des espèces non séquencées (*M. salmoniphilum*, *M. franklinii*, *M. chelonae* et *M. immunogenum*) indiquent au moins neuf taxons différents au sein du complexe *M. chelonae-abscessus*. Nous avons en effet observé que chaque taxon a des spécificités, en particulier, les membres des groupes n ° 3A et n ° 3B possèdent de grands génomes codant des mycobactériophages et plusieurs facteurs d'interaction avec l'hôte. Notre analyse des génomes a identifié les supports génétiques de la résistance aux antibiotiques et du caractère opportuniste de leur pathogénicité [1,2, 10]. Nous avons identifié 66 gènes uniques présents dans chaque groupe; ces gènes pourraient être utilisés dans la détection et l'identification des organismes *M. abscessus*. Par conséquent, les données génomiques rapportées par le présent travail pourraient servir à développer des outils de laboratoire utiles pour le diagnostic de routine de *M. abscessus*.

En particulier, nous avons ici développé une approche moléculaire basée sur le séquençage des séquences intergéniques MST. Cette technique a permis l'identification précise et la discrimination entre les souches de *M. abscessus* (Article 2). La technique de MST peut être ajoutée au panel des méthodes moléculaires actuellement disponibles [34-39] pour le génotypage de *M. abscessus* présentant comme principal avantage le fait d'être basée sur le séquençage fournissant ainsi des données robustes.

Une particularité remarquée grâce au séquençage des génomes de *M. abscessus* est la présence de séquences de prophages [18, 19]. La première observation d'un mycobactériophage en microscopie électronique à partir d'une culture de *M. bolletii* a été réalisée au laboratoire par Adekambi Toidi [données non publiées]. Ainsi nous avons analysé la séquence du prophage de *M. bolletii* et résolu par microscopie électronique la structure 3D de son mycobactériophage que nous avons nommé Araucaria. Ceci est une première étude structurale d'un mycobactériophage (Article 3). Nous avons aussi réalisé une étude de co-évolution de 48 génomes de *M. abscessus* par approches génomique et phylogénétique (Article 4). Les analyses génomiques des prophages de *M. bolletii* et 47 autres génomes de *M. abscessus* ont révélé une grande diversité génétique des prophages et une propriété mosaïque de leurs génomes. Ces génomes codent un grand nombre de gènes de fonction inconnue. Plusieurs études ont montré que les génomes des mycobactériophages possèdent une mine d'informations sur la diversité des phages qui infectent un hôte bactérien commun [26]. Ces études ont révélé la nature omniprésente d'une architecture mosaïque des génomes des mycobactériophages [25]. Les analyses génomiques et phylogénétiques ont indiqué que les mycobactériophages de *M. abscessus* ont une histoire évolutive différente de celle de leurs hôtes. En particulier, 33 protéines codées par des mycobactériophages indiquent des transferts horizontaux entre des souches de *M. abscessus.* Il a été démontré que le génome mosaïque des mycobactériophages est composé de différents segments

avec une histoire évolutive différente les unes des autres [20-25]. Toutefois, les mycobactériophages fournissent une vaste communauté de gènes échangeables qui se transmettent à leurs hôtes et peuvent conduire à leur diversification [20-25]. Effectivement, l'intégration de séquences de mycobactériophages confère de nouvelles informations à leurs hôtes. Les événements d'échanges génétiques entre les mycobactériophages et leurs hôtes peuvent être expliqués par la présence de plusieurs classes distinctes de gènes impliqués dans l'intégration et la recombinaison des génomes de mycobactériophages [25]. La présence d'une telle diversité de protéines suggère différents types d'intégration et de recombinaison notamment la recombinaison illégitime. Ces recombinaisons jouent un rôle clé dans la création de génome mosaïque et la co-évolution des mycobactériophages avec leur hôte.

Nous avons assemblé les quatre constituants : capside, connecteur, queue et dispositif d'adsorption à l'hôte ou baseplate pour déterminer la structure du phage Araucaria (Article 4). Cette toute première analyse structurale permet d'émettre des hypothèses quant aux mécanismes d'infection des mycobactéries par leurs phages. Par analogie avec les mécanismes connus d'infections par les *Siphoviridae*, la première étape du cycle de vie d'un phage est la reconnaissance d'un hôte bactérien, par un récepteur/anti-récepteur de l'interaction, suivie de l'injection de son génome dans la cellule. Les structures de la capside et du connecteur sont très similaires à celles des autres *Siphoviridae*, surtout les phages SPP1 [40, 41] et TP901 -1

[42]. Une étude récente par reconstruction cryo-EM de la capside de mycobactériophage BAKA appartenant au Cluster J a révélé une structure icosaédrique caractéristique des *Siphoviridae* [43]. En revanche, la structure de la queue du mycobactériophage Araucaria est très particulière. C'est une queue hélicoïdale décorée par des pointes radiales déjà signalées pour les phages λ et SPP1 [44, 41]. Par analogie avec les phages SPP1 et λ, la queue Araucaria pourrait jouer un rôle dans l'interaction du phage avec la paroi de la cellule de l'hôte grâce à la liaison à ses saccharides de surface. Il convient de noter, qu'aussi bien le phage λ que le phage SPP1 se fixent aux récepteurs spécifiques aux protéines respectivement LamB [45] et YueB [46]. Par conséquent, leur liaison aux saccharides de la paroi cellulaire peut être une première étape réversible qui maintient et oriente transitoirement le phage avant la reconnaissance du récepteur spécifique. Un tel mécanisme est comparable à celui du podophage T7 [47]. Grâce à la partie basale de la queue ou baseplate certains phages se fixent à une protéine de l'hôte [48-50] notamment ces phages se lient aux composants extracellulaires du système de type de sécrétion de type 7 (T7SS) YueB [45] et PIP [50]. En revanche, des phages semblent attacher exclusivement à des polysaccharides comme les phages P2 et TP901–1 grâce à une Receptor-Binding Protein (RBP). La baseplate d'Araucaria est formée de composantes axiales DIT et Tal, ainsi que d'autres, mais elle est dépourvue de protéines RBP comme on en trouve dans les phages p2 et TP901 -1. Par analogie, nous suggérons qu'Araucaria devrait se lier à un récepteur protéique,

car aucune preuve de la présence de la protéine RBP n’est fournie par la forme de la baseplate ou des séquences protéiques la formant. Il est tentant de spéculer, que comme les mycobactéries possèdent un T7SS, qu'un composant similaire à YueB ou PIP pourrait servir de récepteur pour Araucaria. En dépit de la grande diversité génétique des mycobactériophages et la nature du génome mosaïque d'Araucaria, son homologie avec le mycobactériophage Dori et celle des mycobactériophages du cluster B [Mycobactériophage Database] suggère qu’Araucaria et ces phages pourraient partager une structure similaire et le même mécanisme d'infection à l'hôte. Araucaria pourrait infecter son hôte en deux étapes, par liaison réversible aux saccharides dans une première étape, puis de façon irréversible à une protéine de la paroi cellulaire de l’hôte notamment un composant du T7SS dans une seconde étape. L'existence d'un tel mécanisme chez les mycobactériophages reste à démontrer expérimentalement.

IV- CONCLUSIONS GENERALES ET PERSPECTIVES

Les résultats de nos différentes études sur l'analyse génomique de *M. abscessus sensu lato* et leurs mycobactériophages ont apporté des données nouvelles permettant de mieux comprendre leur diversité et leur statut comme opportunistes pathogènes. Dans la première partie de notre thèse, nous avons réalisé une première étude comparative de génomes de *M. abscessus* et pu montrer qu'il existe au moins cinq taxons au sein de ce groupe. Dans une deuxième partie, nous avons développé une méthode d'identification et de génotypage pouvant être facilement et efficacement réalisé au laboratoire afin de distinguer entre les différents membres de ce groupe.

Pa ailleurs, dans notre troisième partie, nous avons résolu pour la première fois par microscopie électronique la structure d'un mycobactériophage et nous avons suggéré un mécanisme d'infection en deux étapes, une adhésion aux saccharides ensuite une liaison aux protéines. Dans cette même partie nous avons décrit le répertoire en prophages et en protéines virales de 48 génomes de *M. abscessus sensu lato*. Lors de notre analyse génomique de *M. abscessus* nous avons montré l'existence de trois taxons, *M. abscessus*, *M. massiliense* et *M. bolletii*. Nous avons remarqué une diversité plus grande de *M. abscessus* et *M. massiliense* et une distinction particulière de *M. bolletii* pour coder et secréter des mycobactériophages. Vu le nombre restreint des souches séquencées de *M. bolletii* ceci pourrait être confirmé en séquençant d'autres souches de ce taxon. Nous avons montré que les mycobactériophages de *M. abscessus* ont une histoire évolutive différente de celle de leur hôte et ils jouent un rôle dans la

diversification de cette espèce.

Aussi, dans un futur proche, il serait également intéressant d'étudier le spectre d'infection du phage Araucaria et de résoudre la structure de ses protéines structurales afin de confirmer et de mieux comprendre son mode d'infection. Il sera également important de comprendre le mode d'intégration et recombinaison de ces mycobactériophages dans leur hôte. Le rôle des répresseurs dans la biologie des mycobactériophages des espèces de *M. abscessus* reste encore pas connu. Une étude structurale et fonctionnelle de ces protéines pourrait aider à mieux à comprendre ce mécanisme.

Enfin une investigation plus approfondie sur l'expression par des souches de *M. abscessus* des mycobactériophages pouvant infecter ce groupe et autres pathogènes permettrait de développer de nouvaux outils de traitement.

Bien que la diversité génétique et la nouveauté de ces phages soit pleine d'intrigues, le nombre de génomes séquencés continue à croître et leurs mystères continuent à épaissir. Les études réalisées lors de cette thèse contribuent à une meilleure compréhension des mycobactériophages de *M. abscessus* et la diversité de ce taxon, et nous amènent à se poser de nombreuses questions concernant ces génomes, notamment l'existence des séquences de prophages, leurs rôles, leurs expressions et régulations, et leur éventuelle influence sur la physiologie de la bactérie hôte.

REFERENCES

1. Griffith DE. **The talking *Mycobacterium abscessus* blues**. Clin Infect Dis. 2011, **5**:572-4.

2. Petrini B: ***Mycobacterium abscessus*: an emerging rapid-growing potential pathogen**. *APMIS.* 2006, **5:**319-328.

3. Kubica GP, Baess I, Gordon RE, Jenkins PA, Kwapinski JB, McDurmont C, Pattyn SR, Saito H, Silcox V, Stanford JL, Takeya K, Tsukamura M: **A co-operative numerical analysis of rapidly growing mycobacteria**. *J Gen Microbiol* 1972, **73:**55–70.

4. Leao SC, Tortoli E, Euzéby JP, Garcia MJ: **Proposal that *Mycobacterium massiliense* and *Mycobacterium bolletii* be united and reclassified as *Mycobacterium abscessus* subsp. *bolletii* comb. nov., designation of *Mycobacterium abscessus* subsp. *abscessus* subsp. nov. and emended description of *Mycobacterium abscessus***. *Int J Syst Evol Microbiol* 2011, **61:**2311-2313.

5. Adékambi T, Reynaud-Gaubert M, Greub G, Gevaudan MJ, La Scola B, Raoult D, Drancourt M: **Amoebal coculture of "*Mycobacterium massiliense*" sp. nov. from the sputum of a patient with hemoptoic pneumonia**. *J Clin Microbiol* 2004, **42:**5493-5501.

6. Adékambi T, Berger P, Raoult D, Drancourt M: ***rpoB* gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov.** *Int J Syst Evol Microbiol* 2006, **56:**133-143.

7. Ross AJ: ***Mycobacterium salmoniphilum* sp. nov. from**

8. Wilson RW, Steingrube VA, Böttger EC, Springer B, Brown-Elliott BA, Vincent V, Jost KC Jr, Zhang Y, Garcia MJ, Chiu SH, Onyi GO, Rossmoore H, Nash DR, Wallace RJ Jr:

Mycobacterium immunogenum **sp. nov., a novel species related to** ***Mycobacterium abscessus*** **and associated with clinical disease, pseudo-outbreaks, and contaminated metalworking fluids: an international cooperative study on mycobacterial taxonomy.** *Int J Syst Evol Microbiol* 2001, **51:**1751–1764.

9. Simmon KE, Brown-Elliott BA, Ridge PG, Durtschi JD, Mann LB, Slechta ES, Steigerwalt AG, Moser BD, Whitney AM, Brown JM, Voelkerding KV, McGowan KL, Reilly AF, Kirn TJ, Butler WR, Edelstein PH, Wallace RJ Jr, Petti CA: ***Mycobacterium chelonae-abscessus*** **complex associated with sinopulmonary disease, Northeastern USA**. *Emerg Infect Dis* 2011, **9:**1692-1700.

10. Medjahed H, Gaillard JL, Reyrat JM.: ***Mycobacterium abscessus*****: a new player in the mycobacterial field.** *Trends Microbiol* 2010, **3:**117-123.

11. Bryant JM, Grogono DM, Greaves D, Foweraker Juliet, Roddick Iain, Inns T, Reacher M, Aworth CSH, Curran MD, Harris SR, Peacock SJ, Parkhill J, Floto RA: **Whole-genome sequencing to identify transmission of** ***Mycobacterium abscessus*** **between patients with cystic fibrosis: a retrospective cohort study.** *The Lancet* 2013, **381**:1551-1560.

12. Besada E. **Rapid growing mycobacteria and TNF-α blockers: case report of a fatal lung infection with** ***Mycobacterium abscessus*** **in a patient treated with infliximab, and literature review**. *Clin Exp Rheumatol*. 2011, **29**:705-707.

13. Jassies-van der Lee A, Houwers DJ, Meertens N, van der Zanden AG, Willemse T: **Localised pyogranulomatous dermatitis due to** ***Mycobacterium abscessus*** **in a cat: a case report.** *Vet J* 2009, **2:**304-306.

14. Albini S, Mueller S, Bornand V, Gutzwiller ME, Burnand C, Hüssy D, Abril C, Reitt K, Korczak BM, Miserez R: **Cutaneous atypical mycobacteriosis due to** ***Mycobacterium massiliense*** **in a cat.** (Article in German) *Schweiz Arch Tierheilkd* 2007, **12:**553-558.

15. Clayton LA, Stamper MA, Whitaker BR, Hadfield CA, Simons B, Mankowski JL: ***Mycobacterium abscessus*** **pneumonia in an Atlantic bottlenose dolphin (*Tursiops truncatus*).** *J Zoo Wildl Med*. 2012, **43**:961-965.

16. Wiinschmann A, Armien A, Harris NB, Brown-Elliott BA, Wallace RJ Jr, Rasmussen J, Willette M, Wolf T: **Disseminated panniculitis in a bottlenose dolphin (*Tursiops truncatus*) due to *Mycobacterium chelonae* infection**. *J Zoo Wildl Med* 2008, **39:**412-440.

17. Zerihun MA, Nilsen H, Hodneland S, Colquhoun DJ: ***Mycobacterium salmoniphilum*** **infection in farmed Atlantic salmon, Salmo salar L.** *J Fish Dis* 2011, **34:**769-781.

18. Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, Rottman M, Macheras E, Heym B, Herrmann JL, Daffé M, Brosch R, Risler JL, Gaillard JL: **Non-mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus***. *PLoS One* 2009, **4:**e5660.

19. Choo SW, Yusoff AM, Wong YL, Wee WY, Ong CS, Ng KP, Ngeow YF**: Genome analysis of *Mycobacterium massiliense* strain M172, which contains a putative mycobactériophage**. J Bacteriol 2012, **18**:5128.

20. Hatfull, G. F: **The secret lives of mycobactériophages**. Adv Virus Res 2012, **82**:179- 288.

21. Hatfull, G. F. 2012. **Complete genome sequences of 138 mycobactériophages**. J Virol 2012, **86**:2382-2384.

22. Hatfull, G. F., S. G. Cresawn, and R. W. Hendrix: **Comparative genomics of the mycobactériophages: insights into bacteriophage evolution**. Res Microbiol 2008, **159**:332- 339.

23. Hatfull, G. F., D. Jacobs-Sera, J. G. Lawrence, W. H. Pope, D. A. Russell, C. C. Ko, R. J. Weber, M. C. Patel, K. L. Germane, R. H. Edgar, N. N. Hoyte, C. A. Bowman, A. T. Tantoco, E. C. Paladin, M. S. Myers, A. L. Smith, M. S. Grace, T. T. Pham, M. B. O'Brien, A. M. Vogelsberger, A. J. Hryckowian, J. L.

Wynalek, H. Donis-Keller, M. W. Bogel, C. L. Peebles, S. G. Cresawn, and R. W. Hendrix: **Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size**. J Mol Biol 2010, **397**:119-143.

24. Hatfull GF. **Mycobactériophages: genes and genomes**. Annu Rev Microbiol 2010, **64**:331-56.

25. Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM, Petrova ZO, Dedrick RM, Pope WH; Science Education Alliance Phage Hunters Advancing Genomics And Evolutionary Science Sea-Phages Program, Modlin RL, Hendrix RW, Hatfull GF. **On the nature of mycobacteriophage diversity and host preference.** Virology 2012, **2**:187-201.

26. Bardarov, S., Kriakov, J., Carriere, C., Yu, S., Vaamonde, C., McAdam, R. A., Bloom, B. R., Hatfull, G. F. & Jacobs, W. R., Jr : **Conditionally replicating mycobacteriophages: a system for transposon delivery to *Mycobacterium tuberculosis***. Proc Natl Acad Sci USA 1997, **94**:10961–10966.

27. Harris, N. B., Feng, Z., Liu, X., Cirillo, S. L., Cirillo, J. D. & Barletta, R. G : **Development of a transposon mutagenesis system for *Mycobacterium avium* subsp. *paratuberculosis*.** FEMS Microbiol Lett 1999, **175** :21–26.

28. Rybniker, J., Wolke, M., Haefs, C. & Plum, G : **Transposition of Tn5367 in *Mycobacterium marinum*, using a conditionally recombinant mycobacteriophage**. J Bacteriol 2003, **185** :1745–1748.

29. Alcaide, F., Gali, N., Dominguez, J., Berlanga, P., Blanco, S., Orus, P. & Martin, R : **Usefulness of a new mycobacteriophage-based technique for rapid diagnosis of pulmonary tuberculosis**. J Clin Microbiol 2003, **41**:2867–2871.

30. Muzaffar, R., Batool, S., Aziz, F., Naqvi, A. & Rizvi, A : **Evaluation of the FASTPlaqueTB assay for direct detection of *Mycobacterium tuberculosis* in sputum specimens.** Int J Tuberc Lung Dis 2002, **6**:635–640.

31. Jacobs, W. R., Jr, Barletta, R. G., Udani, R. & 7 other authors : **Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages.** Science 1993, **260**:819–822.

32. Pearson, R. E., Jurgensen, S., Sarkis, G. J., Hatfull, G. F. & Jacobs, W. R., Jr : **Construction of D29 shuttle phasmids and luciferase reporter phages for detection of mycobacteria.** Gene 1996, **183**:129–136.

33. Broxmeyer, L., Sosnowska, D., Miltner, E., Chacon, O., Wagner, D., McGarvey, J., Barletta, R. G. & Bermudez, L. E : **Killing of *Mycobacterium avium* and *Mycobacterium tuberculosis* by a myco- bacteriophage delivered by a nonvirulent mycobacterium: a model for phage therapy of intracellular bacterial pathogens.** J Infect Dis 2002, **186**:1155–1160.

34. Adékambi T, Colson P, Drancourt M. ***rpoB*-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria**. *J Clin Microbiol* 2003, **41:**5699-5708.

35. Arnold C, Barrett A, Cross L, Magee JG: **The use of *rpoB* sequence analysis in the differentiation of *Mycobacterium abscessus* and *Mycobacterium chelonae*: a critical judgement in cystic fibrosis?** *Clin Microbiol Infect* 2012, **18:**E131-133.

36. Zelazny AM, Root JM, Shea YR, Colombo RE, Shamputa IC, Stock F, Conlan S, McNulty S, Brown-Elliott BA, Wallace RJ Jr, Olivier KN, Holland SM, Sampaio EP: **Cohort study of molecular identification and typing of *Mycobacterium abscessus*, *Mycobacterium massiliense*, and *Mycobacterium bolletii*.** *J Clin Microbiol* 2009, **47:**1985-1995.

37. Macheras E, Roux AL, Ripoll F, Sivadon-Tardy V, Gutierrez C, Gaillard JL, Heym B: **Inaccuracy of single-target sequencing for discriminating species of the *Mycobacterium abscessus* group.** *J Clin Microbiol* 2009, **47:**2596-2600.

38. Macheras E, Roux AL, Bastian S, Leão SC, Palaci M, Sivadon-Tardy V, Gutierrez C, Richter E, Rusch-Gerdes S, Pfyffer G, Bodmer T, Cambau E, Gaillard JL, Heym B: **Multilocus sequence analysis and *rpoB* sequencing of *Mycobacterium abscessus* (sensu lato) strains.** *J Clin Microbiol* 2011, **49:**491-499.

39. Kim SY, Kang YA, Bae IK, Yim JJ, Park MS, Kim YS, Kim SK, Chang J, Jeong SH: **Standardization of multilocus sequence typing scheme for *Mycobacterium abscessus* and *Mycobacterium massiliense***. Diagn Microbiol Infect Dis 2013.

40. Plisson C, White HE, Auzat I, Zafarani A, Sao-Jose C, Lhuillier S, Tavares P, Orlova EV: **Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection**. EMBO J 2007, **26**:3720–3728.

41. Auzat I, Droge A, Weise F, Lurz R, Tavares P: **Origin and function of the two major tail proteins of bacteriophage SPP1**. Mol. Microbiol 2008, **70**:557–569.

42. Bebeacua C, Lai L, Vegge CS, Brondsted L, van Heel M, Veesler D, Cambillau C: **Visualizing a complete Siphoviridae member by single- particle electron microscopy: the structure of lactococcal phage TP901-1.** J Virol 2013, **87**:1061–1068.

43. Pell LG, Gasmi-Seabrook GM, Morais M, Neudecker P, Kanelis V, Bona D, Donaldson LW, Edwards AM, Howell PL, Davidson AR, Maxwell K: **The solution structure of the C-terminal Ig-like domain of the bacteriophage lambda tail tube protein.** J Mol Biol 2010, **403**:468–479.

44. Pope WH, Jacobs-Sera D, Best AA, Broussard GW, Connerly PL, Dedrick RM, Kremer TA, Offner S, Ogiefo AH, Pizzorno MC, Rockenbach K, Russell DA, Stowe EL, Stukey J, Thibault SA, Conway JF, Hendrix RW, Hatfull GF: **Cluster j mycobacteriophages: intron splicing in capsid and tail genes**. PLoS One. 2013, **9**:8(7).

45. Charbit A, Werts C, Michel V, Klebba PE, Quillardet P, Hofnung M: **A role for residue 151 of LamB in bacteriophage lambda adsorption: possible steric effect of amino acid substitutions.** J Bacteriol 1994, **176**:3204–3209.

46. São-José C, Lhuillier S, Lurz R, Melki R, Lepault J, Santos MA, Tavares P: **The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA**. J Biol Chem 2006, **281**:11464–11470.

47. Hu B, Margolin W, Molineux IJ, Liu J: **The bacteriophage t7 virion undergoes extensive structural remodeling during infection**. Science 2013, **339**:576–579.

48. Boulanger P, Jacquot P, Plancon L, Chami M, Engel A, Parquet C, Herbeuval C, Letellier L: **Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities.** J Biol Chem 2008, **283**:13556–13564.

49. Flayhan A, Wien F, Paternostre M, Boulanger P, Breyton C: **New insights into pb5, the receptor binding protein of bacteriophage T5, and its interaction with its *Escherichia coli* receptor FhuA**. Biochimie 2012, **94**:1982–1989.

50. Babu KS, Spence WS, Monteville MR, Geller BL: **Characterization of a cloned gene (*pip*) from *Lactococcus lactis* required for phage infection.** Dev Biol Stand 1995, **85**:569–575.

ANNEXES

Standards in Genomic Sciences (2013) 8:306-317 DOI:10.4056/sigs.3707349

Non-contiguous genome sequence of *Mycobacterium simiae* strain DSM 44165^T

Mohamed Sassi[1], Catherine Robert[1], Didier Raoult[1] and Michel Drancourt[1*]

Article N°5 : **Sassi M,** Robert C, Drancourt M and Raoult D. Non-contiguous genome sequence of *Mycobacterium simiae* strain DSM 44165^T. **Stand Genomic Sci. 2013, 2: 306-317.**

Non-contiguous genome sequence of *Mycobacterium simiae* strain DSM 44165^T

Mohamed Sassi[1], Catherine Robert[1], Didier Raoult[1] and Michel Drancourt[1*]

[1]Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), Faculté de médecine, Marseille, France

*Corresponding author: Professor Michel Drancourt (Michel.Drancourt@univmed.fr)

Keywords: *Mycobacterium simiae* draft genome, non-tuberculous mycobacteria, SOLiD

Mycobacterium simiae is a non-tuberculosis mycobacterium causing pulmonary infections in both immunocompetent and imunocompromized patients. We announce the draft genome sequence of *M. simiae* DSM 44165^T. The 5,782,968-bp long genome with 65.15% GC content (one chromosome, no plasmid) contains 5,727 open reading frames (33% with unknown function and 11 ORFs sizing more than 5000 -bp), three rRNA operons, 52 tRNA, one 66-bp tmRNA matching with tmRNA tags from *Mycobacterium avium*, *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium microti*, *Mycobacterium marinum*, and *Mycobacterium africanum* and 389 DNA repetitive sequences. Comparing ORFs and size distribution between *M. simiae* and five other *Mycobacterium* species *M. simiae* clustered with *M. abscessus* and *M. smegmatis*. A 40-kb prophage was predicted in addition to two prophage-like elements, 7-kb and 18-kb in size, but no mycobacteriophage was seen after the observation of 10^6 *M. simiae* cells. Fifteen putative CRISPRs were found. Three genes were predicted to encode resistance to aminoglycosides, betalactams and macrolide-lincosamide-streptogramin B. A total of 163 CAZYmes were annotated. *M. simiae* contains ESX-1 to ESX-5 genes encoding for a type-VII secretion system. Availability of the genome sequence may help depict the unique properties of this environmental, opportunistic pathogen.

Introduction

Mycobacterium simiae is the type species for *M. simiae*, and is phylogenetically related to *Mycobacterium triplex* [1], *Mycobacterium genavense* [2], *Mycobacterium heidelbergense* [3], *Mycobacterium lentiflavum* [4], *Mycobacterium sherrisii* [5], *Mycobacterium parmense* [6], *Mycobacterium shigaense* [7], *Mycobacterium stomatepiae* [8] and *Mycobacterium florentinum* [9]. *M. simiae* is slow growing and photochromogenic, appearing rust-colored after exposure to light and is the only non-tuberculous mycobacterium that, is niacin positive, like *Mycobacterium tuberculosis* [10]. *M. simiae* was isolated initially from rhesus macaques in 1965 [11]. In immunocompetent patients, *M. simiae* is responsible for lymphadenitis [12,13], bone infection [14], respiratory tract infection [15] and skin infection [16]. *M. simiae* also causes infection in immunocompromized HIV-infected patients [17,18], including patients with immune reconstruction [19]. Tap water has proven to be a source of *M. simiae* infection in both community and hospital-acquired infection [20,21]. To understand the genetics of *M. simiae* in detail, we sequenced and annotated a draft genome of the type strain of *M. simiae* (DSM 44165^T).

Classification and features

M. simiae strain DSM 44165 ^{T} is the only genome sequenced strain within the *M. simiae* complex (Table 1).

The 16*S* rRNA gene sequence, derived from the *M. simiae* strain DSM 44165 ^{T} genome sequence showed 100% sequence similarity to that of *M. simiae* type strain DSM 44165 ^{T} /ATCC 25275 ^{T} previously deposited in GenBank (GenBank accession: GQ153280.1) and 99% sequence similarity with *M. sherrisii* (GenBank accession: AY353699.1). The *rpoB* gene sequence of *M. simiae* showed 98% similarity with *M. sherrisii* (GenBank accession: GQ166762.1), the closest mycobacterial species. The *rpoB* gene sequence-based phylogenetic tree (Figure 1) illustrates that *M. simiae* DSM 44165 ^{T} is phylogentically closest to *M. sherrisii*, *M. genavense*, *M. triplex*, *M. stomatepiae* and *M. florentinum*, which are all species constituting the *M. simiae* complex.

Table 1. Classification and general features of *Mycobacterium simiae* DSM44165^T [22].

MIGS ID	Property	Term	Evidence code
	Current classification	Domain *Bacteria*	TAS [23]
		Phylum *Actinobacteria*	TAS [24]
		Class *Actinobacteria*	TAS [25]
		Subclass *Actinobacteridae*	TAS [25,26]
		Order *Actinomycetales*	TAS [25-28]
		Suborder *Corynebacterineae*	TAS [25,26]
		Family *Mycobacteriaceae*	TAS [25-27,29]
		Genus *Mycobacterium*	TAS [27,30,31]
		Species *Mycobacterium simiae*	TAS [11,27]
	Gram stain	Weakly positive	TAS [11]
	Motility	Non motile	TAS [11]
	Sporulation	nonsporulating	NAS
	Temperature range	mesophile	TAS [11]
	Optimum temperature	37°C	TAS [11]
	Salinity	normal	TAS [11]
MIGS-22	Oxygen requirement	aerobic	TAS [11]
MIGS-6	Habitat	Soil	TAS [11]
MIGS-15	Biotic relationship	Free-living	NAS
MIGS-14	Pathogenicity	none	NAS
	Biosafety level	2	NAS
	Isolation	Macacus rhesus	TAS [11]
MIGS-4	Geographic location	Country India	TAS [11]
MIGS-5	Sample collection time	1965	TAS [11]
MIGS-4.1	Latitude	20.593684	NAS [11]
MIGS-4.2	Longitude	78.96288	NAS [11]
MIGS-4.3	Depth	Not reported	TAS [11]
MIGS-4.4	Altitude	Not reported	TAS [11]

Evidence codes - 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 [32].

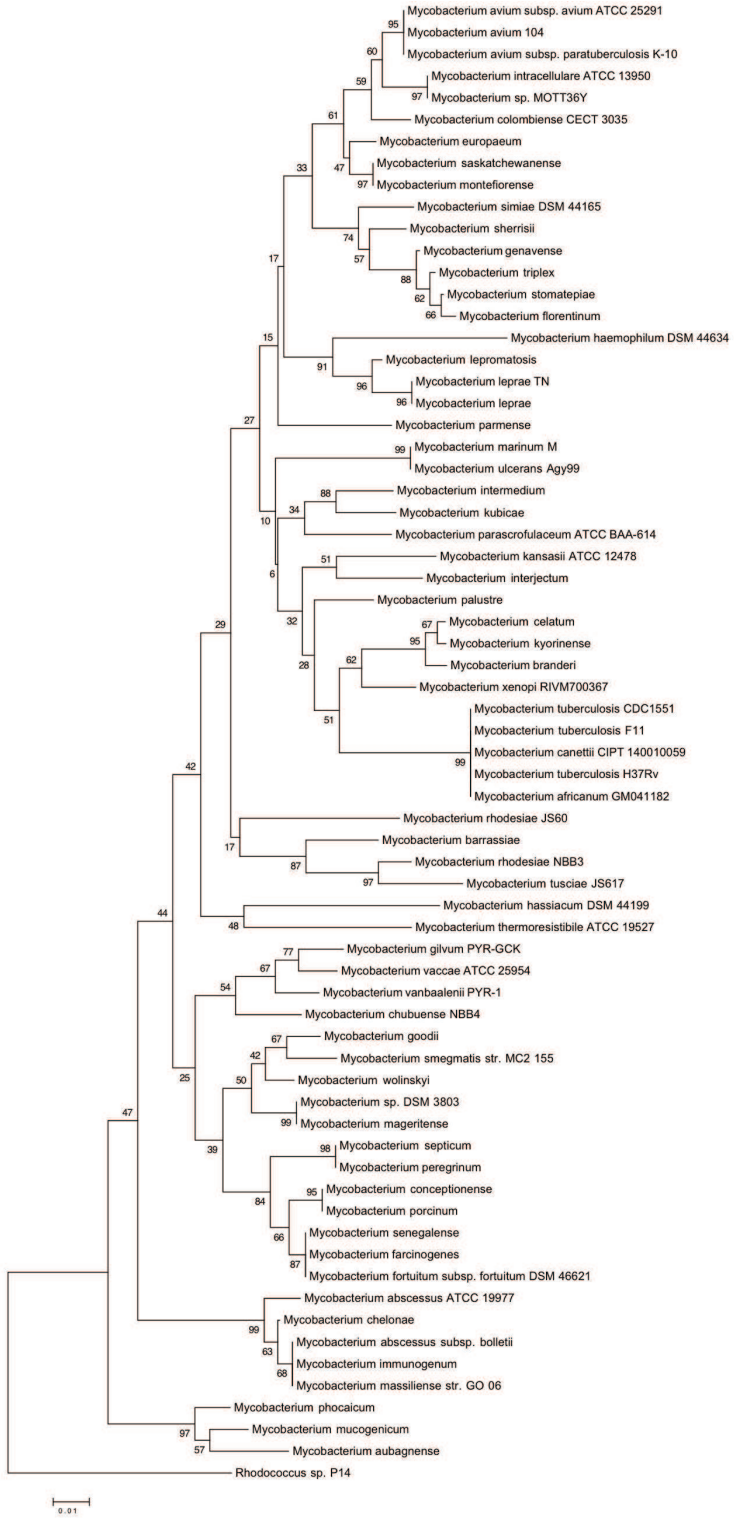


Figure 1. *rpo*B gene sequence based phylogenetic tree highlighting the position of *Mycobacterium simiae* DSM 44165 relative to other type strains within the *Mycobacterium* genus. Phylogenetic inferences obtained using the neighbor-joining method within MEGA. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. *Rhodococcus sp* P14 was used as an outgroup.

The *M. simiae* genome shares, 87%, 83%, 79% and 76% nucleotide similarity with the closest sequenced genomes of the species *Mycobacterium* sp: MOTT36Y (CP003491.1*)*, *M. intracellulare* ATCC 13950 (ABIN00000000*)*, *M. indicus pranii* MTCC 9506 (CP002275.1) and *M. avium* 104 (CP000479.1), respectively.

In order to complement the phenotypic traits previously reported for *M. simiae* [10], we observed 10^6 *M. simiae* cells by electron microscopy as previously described [33]. Briefly, *M. simiae* cells were deposited on carbon-reinforced Formvar-coated grids and negatively stained with 1.5 (w:v) phosphotungstic acid (ph 7.0). The grids were examined using a Hitachi HU-12 electron microscope (FEI, Lyon, France) at 89× magnification. No phage was observed in *M. simiae* DSM 44165 ^{T} cultures. *M. simiae* cells measured 1,226 nm in length and 594 nm in width of (Figure 2).

Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out as previously described [34]. The *M. simiae* spectra were imported into the MALDI Bio Typer software (version 2.0, Bruker, Wissembourg, France) and analyzed by standard pattern matching (with default parameter settings) against the main spectra of 3,769 bacteria, including spectra from 79 validly named mycobacterial species used as reference data, in the Bio Typer database (updated March 15th, 2012). The method of identification includes the m/z from 3,000 to 15,000 Da. For every spectrum, 100 peaks at most were taken into account and compared with the spectra in the database. For *M. simiae* DSM 44165 ^{T}, the score obtained was 1.7, matching that of *M. simiae* 423-B-I-2007-BSI thus suggesting that our isolate was a member of a *M. simiae* species. We incremented our database with the spectrum from *M. simiae* DSM 44165 ^{T} (Figure 3).

Genome sequencing and annotation

Genome project history

M. simiae is the first member of the *M. simiae* species complex for which a genome sequence has been completed. This organism was selected to gain understanding in the genetics of *M. simiae* complex in detail (Table 2).

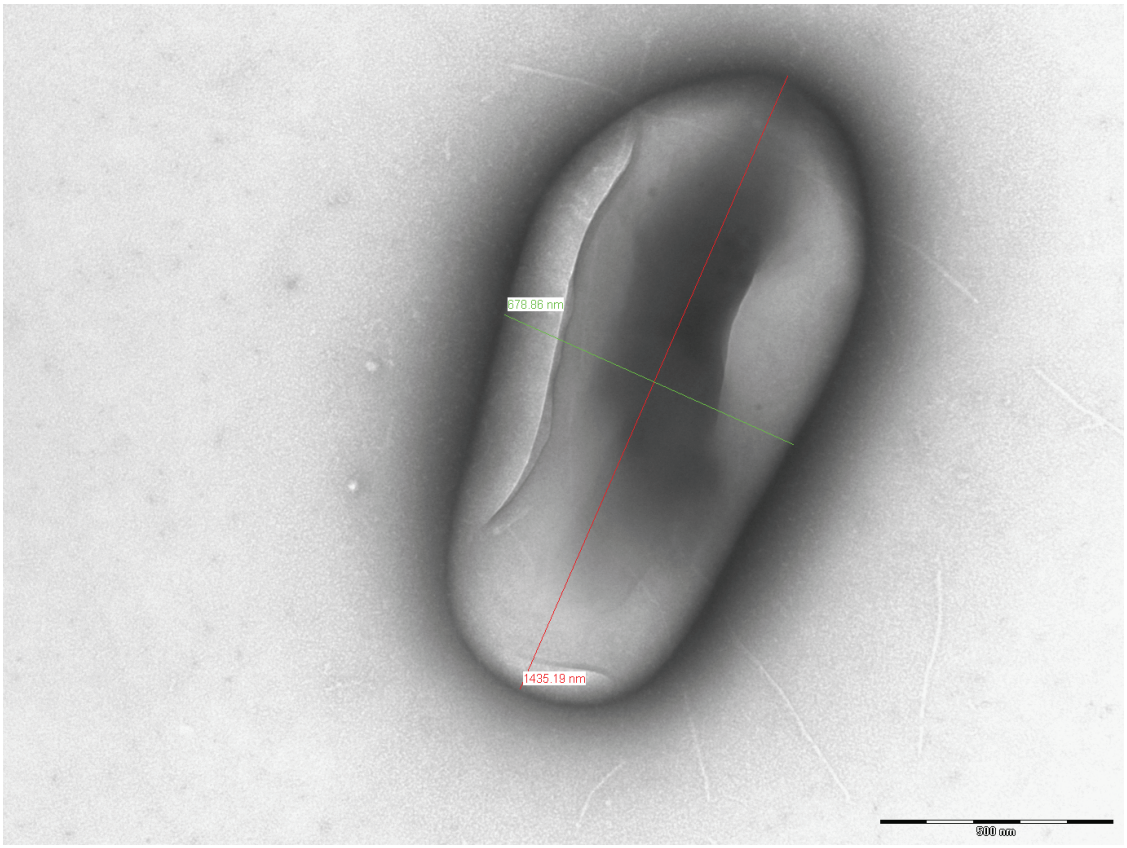


Figure 2. Electron microscopy graph of *M. simiae* DSM 44165^T

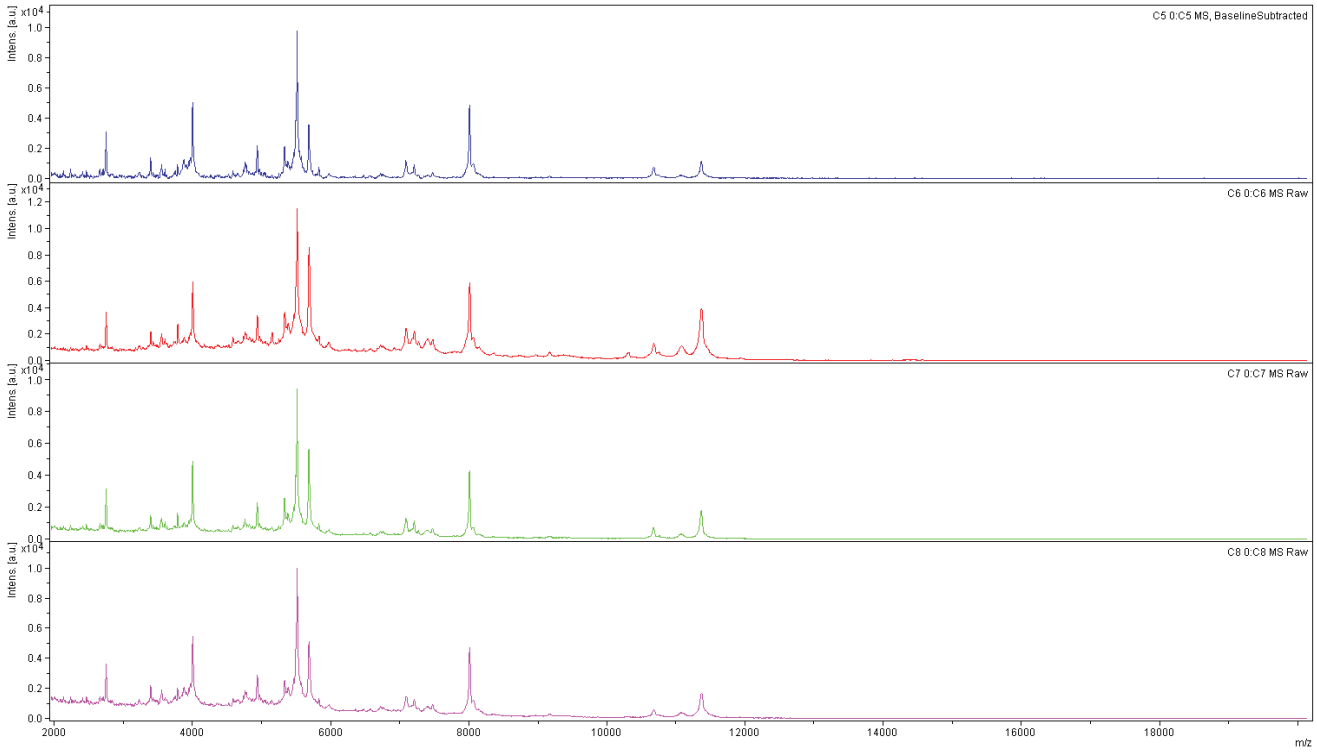


Figure 3. Reference mass spectrum from *M. simiae* strain DSM 44165. Spectra from 5 individual colonies were compared and a reference spectrum was generated.

Table 2. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	One 454 paired end 3-kb library
MIGS-29	Sequencing platforms	454 GS FLX Titanium
MIGS-31.2	Fold coverage	15.33
MIGS-30	Assemblers	Newbler version 2.5.3
MIGS-32	Gene calling method	Prodigal
	EMBL-EBI/NCBI project ID	PRJEB1560
	EMBL-EBI/Genbank ID	CBMJ020000001-CBMJ020000359
	EMBL-EBI Date of Release	June 27, 20113
MIGS-13	Source material identifier	DSM 44165^T
	Project relevance	Pangenome of opportunistic mycobacteria

Growth conditions and DNA isolation

M. simiae strain DSM 44165 ^{T} was grown in 7H9 broth (Difco, Bordeaux, France) enriched with 10% OADC (oleic acid, bovine serum albumin, dextrose and catalase) in 8-mL tubes at 37°C. The culture was centrifuged at 8,000 g for 10 min, the pellet was resuspended in 250 µL of phosphate buffered saline (PBS) and inactivated by heating at 95°C for one h. The sample was then transferred into a sterile screw-cap Eppendorf tube containing 0.3 g of acid-washed glass beads (Sigma, Saint-Quentin Fallavier, France) and shaken using a Bio 101 Fast Prep instrument (Qbiogene, Strasbourg, France) at level 6.5 (full speed) for 45 s. The supernatant was incubated overnight at 56°C with 25 µL proteinase K (20 mg/ml) and 180 µL T1 buffer from the Nucleospin Tissue Mini kit (Macherey-Nagel, Hoerdt, France). After a second mechanical lysis and a 15 min incubation at 70°C, total DNA was extracted using the NucleoSpin Tissue Mini kit (Macherey-Nagel, Hoerdt, France). The extracted DNA was eluted into 100 µL of elution buffer and stored at –20°C until used.

Genome sequencing and assembly

The concentration of the DNA was measured using a Quant-it Picogreen kit (Invitrogen) on the Genios Tecan fluorometer at 79.36 ng/µl. A 5 µg quantity of DNA was mechanically fragmented on the Covaris device (KBioScience-LGC Genomics, Teddington, UK) through miniTUBE-Red 5Kb. The DNA fragmentation was visualized in an Agilent 2100 BioAnalyzer on a DNA labchip 7500 with an optimal size of 3.57kb. The library was constructed according to the 454 Titanium paired end protocol (Roche, Boulogne-Billancourt, France). Circularization and nebulization were performed to generate a pattern with an optimum at 415 bp. After PCR amplification through 17 cycles followed by double size selection, the single stranded paired end library was quantified on the Quant-it Ribogreen kit (Invitrogen) on the Genios_Tecan fluorometer at 865pg/µL. The library concentration equivalence was calculated as 1.91E+09 molecules/µL. The library was stocked at -20°C until used. The library was clonally amplified with 0.5 cpb in 2 emPCR reactions with the GS Titanium SV emPCR Kit (Lib-L) v2 (Roche, Boulogne-Billancourt, France). The yield of the emPCR was 20.2%, which is somewhat high compared to the range of 5 to 20% from the Roche procedure. A total of 790,000 beads were loaded on the GS Titanium PicoTiterPlate PTP Kit 70x75 and sequenced with a GS Titanium Sequencing Kit XLR70 (Roche, Boulogne-Billancourt, France). The run was done overnight and analyzed on the cluster through the gsRunBrowser and gsAssembler_Roche. A total of 241,405 passed filter wells were obtained and generated 88.64Mb with an average 367 bp length. The passed filter sequences were assembled on the gsAssembler (Roche, Boulogne-Billancourt, France), with 90% identity and 40 bp as overlap, yielding one scaffold and 338 large contigs (>1,500 bp), generating a genome size of 5.78 Mb, which corresponds to a coverage of 15.33 × genome equivalents.

Genome annotation

Open reading frames (ORFs) were predicted using Prodigal [35,36] with default parameters. The predicted bacterial protein sequences were searched against the NCBI NR database, UNIPROT [37] and against COGs [38] using BLASTP. The ARAGORN software tool [39] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [40] and BLASTn against the NR database. Proteins were also checked for domain using a hidden Markov model (HMM) search against the PFAM database [41]. The Tandem Repeat Finder was used for repetitive DNA prediction [42]. The prophage region prediction was completed using PHAST (PHAge Search Tool) [43]. CRISPRs were found using the CRISPER finder [44].

The antibiotic resistance genes were annotated using. The CAZYmes, which are enzymes involved in the synthesis, metabolism, and transport of carbohydrates were annotated using CAZYmes Analysis Toolkit (CAT) (mothra.ornl.gov/cgi-bin/cat.cgi?tab=CAZymes)

Genome properties

M. simiae strain DSM 44165 ^{T} genome consists of a 5,782,968-pb long (65.15% GC content) chromosome without plasmids (Figure 4). Table 3 presents the nucleotide content and gene count levels of the genome and the distribution of genes into COGs functional categories is presented in Table 4.

The genome contains three rRNA (5S rRNA, 23S rRNA and 16S rRNA), 52 tRNA genes with one transfer-messenger RNA (tmRNA) and 5,727 ORFs with 4,673 ORFs (81.6%) having at least one PFAM domain. The properties and the statistics of the genome are summarized in Table 3. Of the coding sequences, 66% could be assigned to COG families (Table 4).

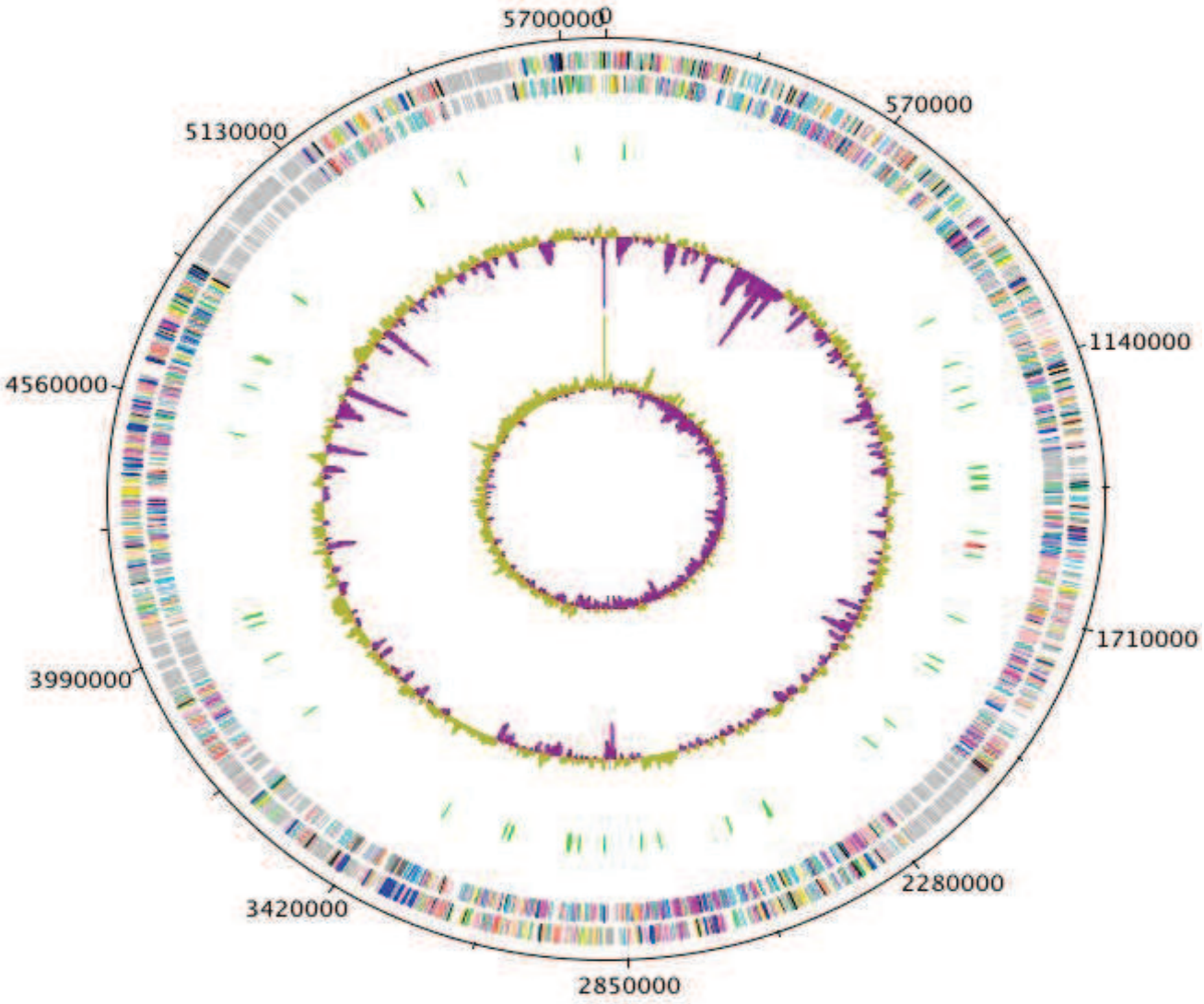


Figure 4. Graphical circular map of the chromosome. From outside to the center: Genes on the forward strand (colored by COG categories), genes on the reverse strand (colored by COG categories), RNA genes (tRNAs green, rRNAs red), GC content, and GC skew.

Table 3. Nucleotide content and gene count levels of the genome

Attribute	Value	% of total[a]
Genome size (bp)	5,782,968	100
DNA coding region (bp)	5,072,379	87.71
DNA G+C content (bp)	3,767,609	65.15
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	5,782	100
RNA genes	55	0.95
Protein-coding genes	5,727	99.04
Genes with function prediction	4,673	81.6
Genes assigned to COGs	4,105	71,67
Genes with peptide signals	377	6.58
Genes with transmembrane helices	1,144	19.97
CRISPR repeats	15	

a) The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome

Table 4: Number of genes associated with the 25 general COG functional categories

Code	Value	% age[a]	Description
J	157	2.74	Translation
A	1	0.02	RNA processing and modification
K	410	7,16	Transcription
L	171	2.99	Replication, recombination and repair
B	2	0.03	Chromatin structure and dynamics
D	34	0.59	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	41	0.72	Defense mechanisms
T	169	2.95	Signal transduction mechanisms
M	162	2.83	Cell wall/membrane biogenesis
N	48	0.84	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	23	0.40	Intracellular trafficking and secretion
O	132	2.30	Posttranslational modification, protein turnover, chaperones
C	400	6.98	Energy production and conversion
G	212	3.70	Carbohydrate transport and metabolism
E	151	2.64	Amino acid transport and metabolism
F	11	0.19	Nucleotide transport and metabolism
H	158	2,76	Coenzyme transport and metabolism
I	418	7,30	Lipid transport and metabolism
P	192	3,35	Inorganic ion transport and metabolism
Q	433	7,56	Secondary metabolites biosynthesis, transport and catabolism
R	656	11,45	General function prediction only
S	291	5,08	Function unknown
	1622	1,25	Not in COGs

a) The total is based on the total number of protein coding genes in the annotated genome.

The draft *M. simiae* genome has 389 DNA repetitive sequences and contains a 40-kb prophage like region with attachment sites. Two prophage like elements sized 7 kb and 8 kb containing six and 12 phage-like proteins respectively. A total of 15 questionable CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) were found and three genes encoding resistance to aminoglycosides, betalactamines and Macrolide-Lincosamide-StreptograminB (Table 3) were annotated. *M. simiae* DSM 44165^T showed the presence of 163 Carbohydrate-Active Enzymes genes belonging to 36 CAZy family (supplementary data S1).

Analysis of the distribution of *M. simiae* ORF size revealed 11 ORFs > 5,000-pb, including two ORFs > 10,000-pb: a 12,942-bp ORF showed 77% similarity with a *M. avium* 104 gene encoding a linear gramidicin synthase subunit D; a 14,415-bp ORF showed no similarity with NR database. We verified the open reading frames of the two ORFs using ORFs finder online software [45] and found that these ORFs encode 4,313 and 4,804 amino acids proteins respectively. A heatmap based on the distribution of ORFs sizes in *M. simiae* and five other genomes was done in R [46], which clusters *M. simiae* with *M. abscessus* and *M. smegmatis*, indicating that the three genomes have similar ORFs size distribution (Figure 5).

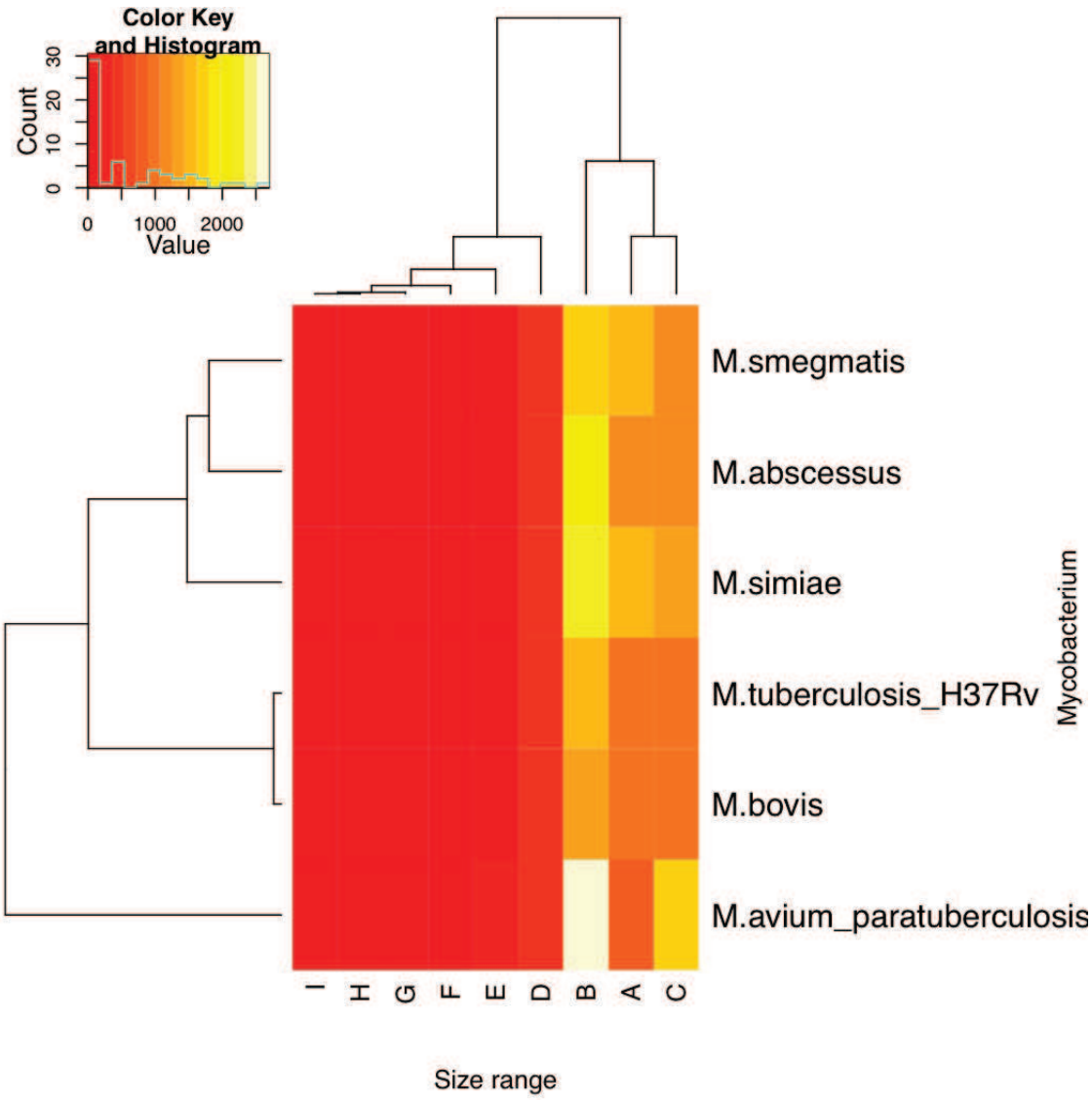


Figure 5. Heatmap of the ORFs size distribution of *M. simiae* compared with 5 other *Mycobacterium* genomes.

Recent evidence shows that mycobacteria have developed novel and specialized secretion systems for the transport of extracellular proteins across their hydrophobic, highly impermeable, cell wall [47]. *M. tuberculosis* genomes encode up to five of these transport systems, and ESX-1 and ESX-5 systems are involved in virulence [47]. In comparison with *M. tuberculosis* H37Rv type VII clusters using Blastp, a total of 77 proteins encoding a type VII secretion system were annotated in *M. simiae* (supplementary data II). ESX-5 seems to be a conserved cluster between *M. tuberculosis* and *M. simiae*, in agreement with opportunistic pathogenicity of *M. simiae*.

Acknowledgements

This study was supported by Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UMR CNRS 7278, IRD 198, INSERM 1095, Faculté de Médecine, Marseille, France

References

1. Floyd MM, Guthertz LS, Silcox VA, Duffey PS, Jang Y, Desmond EP, Crawford JT, Butler WR. Characterization of an SAV organism and proposal of *Mycobacterium triplex* sp. nov. *J Clin Microbiol* 1996; **34**:2963-2967. PubMed
2. Böttger EC, Hirschel B, Coyle MB. *Mycobacterium genavense* sp. nov. *Int J Syst Bacteriol* 1993; **43**:841-843. PubMed http://dx.doi.org/10.1099/00207713-43-4-841
3. Pfyffer GE, Weder W, Strässle A, Russi EW. *Mycobacterium heidelbergense* species nov. infection mimicking a lung tumor. *Clin Infect Dis* 1998; **27**:649-650. PubMed http://dx.doi.org/10.1086/517142
4. Springer B, Wu WK, Bodmer T, Haase G, Pfyffer GE, Kroppenstedt RM, Schröder KH, Emler S, Kilburn JO, Kirschner P, *et al*. Isolation and characterization of a unique group of slowly growing mycobacteria: description of *Mycobacterium lentiflavum* sp. nov. *J Clin Microbiol* 1996; **34**:1100-1107. PubMed
5. Selvarangan R, Wu WK, Nguyen TT, Carlson LD, Wallis CK, Stiglich SK, Chen YC, Jost KC, Jr., Prentice JL, Wallace RJ, Jr., *et al*. Characterization of a novel group of mycobacteria and proposal of *Mycobacterium sherrisii* sp. nov. *J Clin Microbiol* 2004; **42**:52-59. PubMed http://dx.doi.org/10.1128/JCM.42.1.52-59.2004
6. Fanti F, Tortoli E, Hall L, Roberts GD, Kroppenstedt RM, Dodi I, Conti S, Polonelli L, Chezzi C. *Mycobacterium parmense* sp. nov. *Int J Syst Evol Microbiol* 2004; **54**:1123-1127. PubMed http://dx.doi.org/10.1099/ijs.0.02760-0
7. Nakanaga K, Hoshino Y, Wakabayashi M, Fujimoto N, Tortoli E, Makino M, Tanaka T, Ishii NJ. *Mycobacterium shigaense* sp. nov., a novel slowly growing scotochromogenic mycobacterium that produced nodules in an erythroderma patient with severe cellular immunodeficiency and a history of Hodgkin's disease. *J Dermatol* 2012; **39**:389-396. PubMed http://dx.doi.org/10.1111/j.1346-8138.2011.01355.x
8. Pourahmad F, Cervellione F, Thompson KD, Taggart JB, Adams A, Richards RH. *Mycobacterium stomatepiae* sp. nov., a slowly growing, non-chromogenic species isolated from fish. *Int J Syst Evol Microbiol* 2008; **58**:2821-2827. PubMed http://dx.doi.org/10.1099/ijs.0.2008/001164-0
9. Tortoli E, Rindi L, Goh KS, Katila ML, Mariottini A, Mattei R, Mazzarelli G, Suomalainen S, Torkko P, Rastogi N. *Mycobacterium florentinum* sp. nov., isolated from humans. *Int J Syst Evol Microbiol* 2005; **55**:1101-1106. PubMed http://dx.doi.org/10.1099/ijs.0.63485-0
10. Rynkiewicz DL, Cage GD, Butler WR, Ampel NM. Clinical and microbiological assessment of *Mycobacterium simiae* isolates from a single laboratory in southern Arizona. *Clin Infect Dis* 1998; **26**:625-630. PubMed http://dx.doi.org/10.1086/514573
11. Karassova V, Weissfeiler J, Krasznay E. Occurrence of atypical mycobacteria in Macacus rhesus. *Acta Microbiol Acad Sci Hung* 1965; **12**:275-282. PubMed
12. Patel NC, Minifee PK, Dishop MK, Munoz FM. *Mycobacterium simiae* cervical lymphadenitis. *Pediatr Infect Dis J* 2007; **26**:362-363. PubMed http://dx.doi.org/10.1097/01.inf.0000258614.98241.4e
13. Cruz AT, Goytia VK, Starke J. *Mycobacterium simiae* complex infection in an immunocompetent child. *J Clin Microbiol* 2007; **45**:2745-2746. PubMed http://dx.doi.org/10.1128/JCM.00359-07
14. Bartanusz V, Savage JG. Destructive *Mycobacterium simiae* infection of the lumbar spine and retroperitoneum in an immunocompetent adult.

Spine J 2012; **12**:534-535. PubMed http://dx.doi.org/10.1016/j.spinee.2012.05.015

15. Baghaei P, Tabarsi P, Farnia P, Marjani M, Sheikholeslami FM, Chitsaz M, Gorji Bayani P, Shamaei M, Mansouri D, Masjedi MR, *et al*. Pulmonary disease caused by *Mycobacterium simiae* in Iran's national referral center for tuberculosis. *J Infect Dev Ctries* 2012; **6**:23-28. PubMed http://dx.doi.org/10.3855/jidc.1297

16. Piquero J, Casals VP, Higuera EL, Yakrus M, Sikes D, de Waard JH. Iatrogenic *Mycobacterium simiae* skin infection in an immunocompetent patient. *Emerg Infect Dis* 2004; **10**:969-970. PubMed http://dx.doi.org/10.3201/eid1005.030681

17. Al-Abdely HM, Revankar SG, Graybill JR. Disseminated *Mycobacterium simiae* infection in patients with AIDS. *J Infect* 2000; **41**:143-147. PubMed http://dx.doi.org/10.1053/jinf.2000.0700

18. Alcalá L, Ruiz-Serrano MJ, Cosín J, García-Garrote F, Ortega A, Bouza E. Disseminated infection due to *Mycobacterium simiae* in an AIDS patient: case report and review. *Clin Microbiol Infect* 1999; **5**:294-296. PubMed http://dx.doi.org/10.1111/j.1469-0691.1999.tb00146.x

19. Vitoria MA, González-Domínguez M, Salvo S, Crusells MJ, Letona S, Samper S, Sanjoaquín I. *Mycobacterium simiae* pulmonary infection unmasked during immune reconstitution in an HIV patient. *Diagn Microbiol Infect Dis* 2013; **75**:101-103. PubMed http://dx.doi.org/10.1016/j.diagmicrobio.2012.09.004

20. Conger NG, O'Connell RJ, Laurel VL, Olivier KN, Graviss EA, Williams-Bouyer N, Zhang Y, Brown-Elliott BA, Wallace RJ, Jr. *Mycobacterium simiae* outbreak associated with a hospital water supply. *Infect Control Hosp Epidemiol* 2004; **25**:1050-1055. PubMed http://dx.doi.org/10.1086/502342

21. El Sahly HM, Septimus E, Soini H, Septimus J, Wallace RJ, Pan X, Williams-Bouyer N, Musser JM, Graviss EA. *Mycobacterium simiae* pseudo-outbreak resulting from a contaminated hospital water supply in Houston, Texas. *Clin Infect Dis* 2002; **35**:802-807. PubMed http://dx.doi.org/10.1086/342331

22. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed http://dx.doi.org/10.1038/nbt1360

23. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576

24. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.

25. Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int J Syst Bacteriol* 1997; **47**:479-491. http://dx.doi.org/10.1099/00207713-47-2-479

26. 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. PubMed http://dx.doi.org/10.1099/ijs.0.65780-0

27. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. http://dx.doi.org/10.1099/00207713-30-1-225

28. Buchanan RE. Studies in the nomenclature and classification of bacteria. II. The primary subdivisions of the *Schizomycetes*. *J Bacteriol* 1917; **2**:155-164. PubMed

29. Chester FD. Report of mycologist: bacteriological work. *Delaware Agricultural Experiment Station Bulletin* 1897; **9**:38-145.

30. Lehmann KB, Neumann R. Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik, First Edition, J.F. Lehmann, München, 1896, p. 1-448.

31. Runyon EH, Wayne LG, Kubica GP. Genus I. *Mycobacterium* Lehmann and Neumann 1896, 363. *In*: Buchanan RE, Gibbons NE (eds), Bergey's Manual of Determinative Bacteriology, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 682-701.

32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. PubMed http://dx.doi.org/10.1038/75556

33. Brenner S, Horne RW. A negative staining method for high resolution electron microscopy of vi-

ruses. *Biochim Biophys Acta* 1959; **34**:103-110. PubMed http://dx.doi.org/10.1016/0006-3002(59)90237-9

34. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis* 2009; **49**:543-551. PubMed http://dx.doi.org/10.1086/600885

35. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. PubMed http://dx.doi.org/10.1186/1471-2105-11-119

36. Prodigal. http://prodigal.ornl.gov/

37. UNIPROT. http://www.uniprot.org/

38. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, *et al*. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 2003; **4**:41. PubMed http://dx.doi.org/10.1186/1471-2105-4-41

39. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 2004; **32**:11-16. PubMed http://dx.doi.org/10.1093/nar/gkh152

40. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. PubMed http://dx.doi.org/10.1093/nar/gkm160

41. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, *et al*. The Pfam protein families database. *Nucleic Acids Res* 2012; **40**:D290-D301. PubMed http://dx.doi.org/10.1093/nar/gkr1065

42. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999; **27**:573-580. PubMed http://dx.doi.org/10.1093/nar/27.2.573

43. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res* 2011; **39**:W347-W352. PubMed http://dx.doi.org/10.1093/nar/gkr485

44. CRISPER. http://crispr.u-psud.fr/server/

45. ORFs. http://www.ncbi.nlm.nih.gov/projects/gorf/

46. Heiberger RM, Holland B. 2004. *Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R, and SAS*. Springer Texts in Statistics. Springer. ISBN 0-387-40270-5.

47. Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, Appelmelk BJ, Bitter W. Type VII secretion--mycobacteria show the way. *Nat Rev Microbiol* 2007; **5**:883-891. PubMed http://dx.doi.org/10.1038/nrmicro1773

Article N°6 : **Sassi M,** Robert C, Drancourt M and Raoult D. Draft Genome Sequence of *Mycobacterium septicum* strain DSM44393^{T} . **Genome Announc. 2013, 15:1.**

genomeA
Journals.ASM.org

Noncontiguous Genome Sequence of *Mycobacterium septicum* Strain DSM 44393^T

Mohamed Sassi, Catherine Robert, Didier Raoult, Michel Drancourt

Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UMR CNRS 7278, IRD 198, INSERM 1095, Faculté des Médicine, Marseille, France

The rapidly growing *Mycobacterium septicum* rarely causes pulmonary infections. We report here the draft genome sequence of *M. septicum* strain DSM 44393^T, isolated from catheter-related bacteremia and initially identified as a member of *Mycobacterium fortuitum*.

Received 1 July 2013 **Accepted** 17 July 2013 **Published** 15 August 2013

Citation Sassi M, Robert C, Raoult D, Drancourt M. 2013. Noncontiguous genome sequence of *Mycobacterium septicum* strain DSM 44393^T. Genome Announc. 1(4):e00574-13. doi:10.1128/genomeA.00574-13.

Address correspondence to Michel Drancourt, Michel.Drancourt@univmed.fr.

The rapidly growing *Mycobacterium septicum* is found in water environments (1, 2). *M. septicum* rarely causes pulmonary infections (3, 4) but the name was coined after its initial isolation from the blood of a patient with catheter-related bacteremia, when it was identified as a member of the *Mycobacterium fortuitum* group (5). In order to get further knowledge on this group of mycobacteria, we report here the first draft genome sequence of *M. septicum* strain DSM 44393^T.

The *M. septicum* genome was sequenced using the 454 GS FLX Titanium pyrosequencing system (Roche, Boulogne-Billancourt, France). The 454 sequencing generated 371,276 reads (142,575,954 bp), assembled into contigs and scaffolds using Newbler version 2.6 (Roche) and checked using CLC Genomics Workbench v 4.7.2 (CLC bio, Aarhus, Denmark). Functional annotation was achieved using Prodigal (6) and BLASTp searches against the National Center for Biotechnology Information (NCBI) nonredundant (NR), UniProt (http://www.uniprot.org/), and COG databases (7). tRNA and rRNA genes were predicted using Aragorn and RNAmmer, respectively (8, 9).

M. septicum strain DSM 44393^T comprises 173 contigs (including 159 contigs of >1,500 bp) in 7 scaffolds. The draft genome size is 6,879,294 bp, with 66.73% G+C content. There are 56 RNAs. Of the 6,748 predicted genes, 6,692 (92.63%) are protein-coding genes, including 4,752 (71.01%) genes assigned to a putative function, 1,381 genes (20.64%) annotated as hypothetical proteins, and 199 (2.97%) genes identified as ORFans. We found 178 genes encoding resistance to aminoglycosides, β-lactamases, fosfomycin, fucidic acid, fluoroquinolones, macrolide-lincosamide-streptogramin B, phenicol, rifampin, tetracycline, trimethoprim, and glycopeptides. A total of 119 type VII secretion system proteins were annotated by comparison with reannotated *Mycobacterium tuberculosis* strain H37Rv. The *M. septicum* genome encodes 487 prophages and phage proteins and 11 clustered regularly interspaced short palindromic repeats (CRISPRs), according to the fast phage search tool (Phast) (10) and the CRISPERfinder online software program (http://crispr.u-psud.fr/Server/).

M. fortuitum members were determined to be the closest species to *M. septicum* based on the 16S rRNA gene sequence similarity. Accordingly, *M. septicum* presents 83.32% nucleotide sequence identity of core proteins with *M. fortuitum* subsp. *fortuitum*, confirming that these are two distinct species within the same group of mycobacteria (11). We identified 5,977 orthologous genes shared between *M. fortuitum* subsp. *fortuitum* and *M. septicum*. Also, 715 genes present in *M. septicum* are not found in *M. fortuitum*. These data could serve to set up new molecular tools for the refined identification of *M. septicum* in both environmental and clinical specimens to identify precise epidemiological and clinical features associated with this emerging opportunistic pathogen.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. CBMO00000000.1. The version described in this paper is the first version.

ACKNOWLEDGMENT

This study was supported by Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), Marseille, France.

REFERENCES

1. **Castillo-Rodal AI, Mazari-Hiriart M, Lloret-Sánchez LT, Sachman-Ruiz B, Vinuesa P, López-Vidal Y.** 2012. Potentially pathogenic nontuberculous mycobacteria found in aquatic systems. Analysis from a reclaimed water and water distribution system in Mexico City. Eur. J. Clin. Microbiol. Infect. Dis. **31**:683–694.
2. **Van Ingen J, Blaak H, de Beer J, de Roda Husman AM, van Soolingen D.** 2010. Rapidly growing nontuberculous mycobacteria cultured from home tap and shower water. Appl. Environ. Microbiol. **76**:6017–6019.
3. **Adékambi T, Drancourt M.** 2006. Isolation of *Mycobacterium septicum* from the sputum of a patient suffering from hemoptoic pneumonia. Res. Microbiol. **157**:466–470.
4. **Lian L, Deng J, Zhao X, Dong H, Zhang J, Li G, Xiao T, Wu Y, Li Q, Wan K.** 2013. The first case of pulmonary disease caused by *Mycobacterium septicum* in China. Int. J. Infect. Dis. **17**:e352–e354. doi:10.1016/j.ijid.2012.12.011.
5. **Schinsky MF, McNeil MM, Whitney AM, Steigerwalt AG, Lasker BA, Floyd MM, Hogg GG, Brenner DJ, Brown JM.** 2000. *Mycobacterium septicum* sp. nov., a new rapidly growing species associated with catheter-related bacteraemia. Int. J. Syst. Evol. Microbiol. **50**(Pt 2):575–581.

6. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. doi:10.1186/1471-2105-11-119.
7. **Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41. doi:10.1186/1471-2105-4-41.
8. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108.
9. **Laslett D, Canback B.** 2004. Aragorn, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32:**11–16.
10. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: a fast phage search tool. Nucleic Acids Res. **39:**W347–W352. doi:10.1093/nar/gkr485.
11. **Sentausa E, Fournier PE.** Advantages and limitations of genomics in prokaryotic taxonomy. Clin. Microbiol. Infect., in press.

« ***Non, la science n'est pas une illusion. Mais ce serait une illusion de croire que nous puissions trouver ailleurs ce qu'elle ne peut pas nous donner*** »

Sigmund Freud