UNIVERSITE SIDI MOHAMED BEN ABDELLAH
FACULTE DES SCIENCES ET TECHNIQUES DE FES
Département de Biologie

Master: Biotechnologie Microbienne

PROJET DE FIN D'ETUDES

Analyse phylogénétique de trois complexes Mycobactériens: Le complexe *Mycobacterium tuberculosis*, le complexe *Mycobacterium avium* et le complexe *Mycobacterium abscessus*

Stage effectué à: Laboratoire de Biotechnologie Microbienne, FST- Fès.

Présenté par: EL-AMMARI Abdelghaffar.

Encadré par: Pr. IRAQUI HOUSSAINI Mohammed.

-Etablissement d'accueil: FST-Fès.

Soutenu le: le 25 juin 2012.

Jury:

-Pr. IRAQUI HOUSSAINI Mohammed;

-Pr. HAGGOUD Abdellatif;

-Pr. TAHIRI Ali;

-Pr. GUISSI Sanae.

UNIVERSITE SIDI MOHAMED BEN ABDELLAH
FACULTE DES SCIENCES ET TECHNIQUES DE FES
Département de Biologie

Année universitaire : 2011-2012

DEDICACES

Je dédie ce travail à mes très chers parents, à mes frères et mes sœurs, à mes amis et à tous ceux qui me sont chers.

REMERCIEMENTS

- **Pr. Mohammed IRAQUI HOUSSAINI:**

Je vous remercie pour tout ce que vous m'avez apporté, pour votre choix du sujet, vos conseils, votre disponibilité, votre encouragement, pour m'avoir fait confiance et m'avoir laissé la liberté nécessaire à l'accomplissement de mes travaux, tout en y gardant un œil critique et avisé.

Je garderai toujours les meilleurs souvenirs de votre immense savoir et de votre sens du devoir. Qu'il me soit permis de vous exprimer ici mon profond respect et mon intense admiration pour vos qualités humaines exceptionnelles.

- **Pr. Abdellatif HAGGOUD:**

J'aimerai vous adresser un remerciement particulier pour tous les efforts que vous avez subi pour l'accomplissement de notre formation avec une grande patience et pédagogie.

Je vous remercie également de votre disponibilité à participer à la critique de ce travail. Votre présence parmi les membres de jury m'honore.

- **Pr. Ali TAHIRI:**

Je vous remercie infiniment d'avoir accepté de juger ce travail;

Veuillez trouver ici l'expression de mon profond respect et de ma haute considération.

- **Pr. Sanae GUISSI:**

Je vous remercie, Mme, infiniment d'avoir accepté de juger ce travail

Qu'il me soit permis de vous exprimer mon estime et ma sincère reconnaissance.

TABLE DES MATIERES

LISTE DES ABREVIATIONS

- ADN: Acide Désoxyribonucléique;
- *ARN tm*: transfer messenger RNA;
- *ARNr 16S*: Acide Ribonucléique 16S;

- BAAR: Bacilles Acido-Alcoolo-Résistants;
- BLAST: Basic Local Alignment Search Tool;
- DDBJ: DNA DataBank of Japan;
- EMBL: European Molecular Biology Laboratory;
- *gyrB*: sous-unité β de l'enzyme ADN gyrase;
- HGT: Horizontal Gene Ttransfer;
- *Hsp65*: Heat-shock protein de 65 kDa;
- Indel: Insertion/ délétion;
- *ITS*: Intergénic Spacer;
- kDa: Kilo-Dalton;
- MAC: *Mycobacterium Avium* Complex;
- ML: Maximum Likelihood;
- MLSA: Multigene or multilocus Sequence Analysis;
- MP: Maximum de Parcimonie;
- MTBC: *Mycobacterium Tuberculosis* Complex;
- NJ: Neighbor Joining;
- OMS: Organisation Mondiale de la Santé;
- OTU: Operational Taxonomic Units;
- RGM: Rapidly Growing Mycobacteria;
- *rpoB:* Sous unité β de l'ARN polymérase;
- *rrn:* Ribosomal ribonucleotid;
- *rrs:* gène d'*ARNr 16S*;
- SGM: Slowly Growing Mycobacteria;
- SNP: Single-Nucleotide Polymorphism;
- *sodA*: superoxyde dismutase A;
- *tuf*: Facteur d'élongation Tu;
- UPGMA: Unweighted Pair Group Method withArithmetic Mean;

LISTE DES FIGURES

LISTE DES TABLEAUX

RESUME

La phylogénie moléculaire du genre *Mycobacterium* est compliquée parce qu'il y a une grande similarité génomique entre les espèces Mycobactériennes. Le complexe *Mycobacterium tuberculosis* par exemple comprend des espèces très apparentées entre elles, en outre, ce complexe pose un problème très sérieux de santé publique dans le monde entier. Ce constat concerne aussi d'autres complexes *Mycobactériens* comme le complexe *Mycobacterium avium* ou celui de *Mycobacterium abscessus.*

Inférer une phylogénie de ces complexes à partir d'un seul locus s'avère souvent moins robuste. Ainsi, une analyse phylogénétique des trois complexes précédents a été menée en utilisant les loci suivants: *ARNr 16S, hsp65, rpoB, tuf, sodA, ARNtm, gyrB* et *sceA1*. Dans le but de détecter une combinaison pertinente pour réaliser la phylogénie des Mycobactéries en question, ces loci sont analysés individuellement puis collectivement (MLSA: multilocus Sequence Analysis).

Les résultats obtenus montrent que les combinaisons de deux ou de trois faites à partir des gènes d'*ARNr 16S*, de *hsp65*, de *tuf* et de *gyrB* aboutissent à une phylogénie fiable.

Mots clés: phylogénie, *Mycobactéries*, loci, complexe *M. tuberculosis*, complexe *M. avium*, complexe *M. abscessus*, MLSA, combinaison de gènes.

INTRODUCTION

La Phylogénie (phylogenèse) est une reconstruction de l'histoire évolutive des êtres vivants en étudiant leurs liens de parentés. Elle permet de classer les êtres vivants, d'étudier leur biodiversité, de faire des études épidémiologiques, …etc. Cette discipline de l'évolution reposait au début sur des critères phénotypiques. Actuellement la phylogénie moléculaire, basée sur les séquences d'acides nucléiques ou d'acides aminés, a largement amélioré notre compréhension sur le changement et la biodiversité du monde vivant y compris le monde bactérien.

Le genre *Mycobacterium* est le seul représentant de la famille des *Mycobacteriaceae* de l'ordre des *Actinomycetales*. Les espèces de ce genre sont caractérisées par un ensemble de critères morphologiques, biochimiques et moléculaires; parmi ceux-ci leur capacité de résister aux acides et

aux alcools, de ce fait, les mycobactéries sont qualifiées de B.A.A.R. (bacilles acido-alcoolo-résistants) [5]. L'espèce type est *Mycobacterium tuberculosis* ou bacille de Koch (1896).

La tuberculose, la lèpre et les mycobactérioses sont des maladies sérieuses de santé publique. La tuberculose est causée par des espèces mycobactériennes strictement pathogènes appartenant au complexe *M. tuberculosis comme M. tuberculosis*. La lèpre est due à *Mycobacterium leprae*. Les mycobactéries responsables des mycobactérioses sont des bactéries pathogènes opportunistes comme *Mycobacterium avium-intracellulare, Mycobacterium kansasii, Mycobacterium xenopi, Mycobacterium ulcerans*…etc. [8]

La tuberculose touche un tiers de la population mondiale et cause environ deux millions de décès chaque année [47].

En se basant sur la vitesse de croissance, les mycobactéries peuvent être divisées en deux groupes: les espèces à croissance rapide (RGM: rapidly growing mycobacteria) et les espèces à croissance lente (SGM: slowly growing mycobacteria) [2], L'ajout du critère de pigmentation des colonies permet de différencier 4 groupes au sein du genre *Mycobacterium* (Cf. classification de Runyon, Tableau 2).

La majorité des mycobactéries pathogènes sont à croissance lente (SGM). La culture des mycobactéries, qui nécessite plusieurs semaines, reste encore la méthode de référence dans le diagnostic de la tuberculose et des mycobactérioses humaines. La lenteur de croissance des mycobactéries est responsable du délai important de rendu des résultats. En effet, un délai minimum de quatre semaines est requis pour que l'isolat mycobactérien puisse être identifié. Actuellement, les techniques moléculaires permettent un gain de temps précieux dans la prise en charge de la maladie, notamment avec les méthodes d'amplification génique, applicables directement aux échantillons cliniques [5]. Le séquençage des fragments d'ADN, amplifiés génétiquement, permet d'avoir les séquences qui seront analysées par des méthodes de phylogénie, ce qui permettra l'identification des mycobactéries.

La phylogénie des mycobactéries a beaucoup profité des avancées réalisées en terme de séquençage des génomes mycobactériens et de la disponibilité des séquences dans les banques de données (Genbank, EMBL, DDBJ…etc.) qui contiennent la majorité des séquences génomiques obtenues.

Il existe plusieurs travaux publiés sur la phylogénie du genre *Mycobacterium***,** la plupart d'entre eux fixent les objectifs suivants:

- Différencier entre les espèces appartenant au genre *Mycobacterium*, surtout celles qui sont proches;

- Identifier des souches isolées à partir des échantillons cliniques;
- Tester la pertinence de certains gènes de ménages comme marqueurs phylogénétiques fiables.

Les phylogénies moléculaires des mycobactéries, basées sur une seule cible moléculaire, sont jugées inexactes et ayant une discrimination insuffisante [13,6]. Aujourd'hui et depuis l'édition du rapport de la commission de réévaluation de la définition des espèces en bactériologie [42], l'approche de l'analyse multigénique (MLSA: multigene or multilocus sequence Analysis,) est devenue dominante. Ainsi plusieurs études ont été faites avec deux, trois, quatre, cinq, six, et même avec sept [26] gènes de ménage.

Les travaux, basés sur l'approche MLSA, ont signalé que les arbres phylogénétiques obtenus par cette approche sont plus robustes que ceux résultant de l'analyse d'un seul locus [26, 2, 13,…etc.].

La problématique de la phylogénie des mycobactéries n'est pas encore résolue car il existe des espèces phylogénétiquement très proches et sont difficilement discernables, parmi celles-ci, on note:

- Les membres du complexe *M. tuberculosis* (MTBC pour *Mycobacterium tuberculosis* complex: *M. tuberculosis, Mycobacterium bovis, Mycobacterium africanum, Mycobacterium microti, Mycobacterium caprae et Mycobacterium canetti*) [13, 19, 53, 25, 9, 23...etc.];
- Les membres du complexe *M. avium* (MAC pour *M. avium* complex) [9];
- Les membres de *Mycobacterium abscessus* sensu lato *(M. abscessus* sensu stricto*, Mycobacterium bolleti et Mycobacterium massiliense)* [2, 22, 52]*;*
- *Les membres de Mycobacterium fortiutum (M. fortuitum subsp. Acetamidolyticum, M. fortuitum subsp. Fortuitum)* [2, 9];
- Les membres de *M. abscessus* sensu lato vis-à-vis de *Mycobacterium immunogenum et Mycobacterium chelonae* [2, 53, 9, 22];
- *Mycobacterium marinum vis-à-vis de M. ulcerans* [13, 19, 53, 9];
- *M. kansasii vis-à-vis de Mycobacterium gastri* [13, 19, 3, 53, 9];
- *Mycobacterium murale vis-à-vis de Mycobacterium tokaiense* [13];
- *Mycobacterium malmoense* vis-à-vis de *Mycobacterium szulgai* [13, 19];
- *…etc.*

Cependant, le groupe qui pose un problème sérieux, aussi bien en santé publique qu'en analyse phylogénétique, est le complexe *M. tuberculosis* (MTBC). Ce complexe est facilement identifiable des autres espèces, mais il est très difficile de distinguer entre ses membres.

A part un seul gène *gyrB* qui code pour la sous-unité β de l'enzyme ADN gyrase, dont les travaux ont montré qu'il est pertinent dans ce sens [19, 10, 28,], aucun autre gène n'est capable de distinguer entre les membres du complexe MTBC.

En utilisant ce gène ainsi que 7 autres gènes de ménage reconnus comme étant utiles pour faire la phylogénie du genre *Mycobacterium* (*ARNr 16S, hsp65, rpoB, tuf, sodA, secA1* et *ARN tm*), notre objectif est de répondre à la question suivante:

Existe-t-il une combinaison de gènes qui pourra différencier entre les membres du complexe MTBC?

Ce manuscrit contient deux parties:

- *Une partie bibliographique:* met l'accent sur les principales méthodes utilisées pour faire l'analyse phylogénétique d'une part et sur les principales caractéristiques phénotypiques, génotypiques, pathogéniques et taxonomiques du genre *Mycobacterium* d'autre part;
- *Une partie pratique:* dans laquelle, une analyse phylogénétique sera réalisée à partir des espèces et des séquences choisies.

PARTIE BIBLIOGRAPHIQUE

I. LA PHYLOGENIE; DEFINITION, INTERETS ET METHODES:

Malgré la diversité des formes de vie observées, c'est l'unité qui se dégage du monde vivant:

- L'unité d'organistaion: Les êtres vivants sont constitués de cellules;
- L'unité de fonctionnement: le métabolisme fondamental est le même pour toutes les cellules, l'ADN est le support de l'information génétique, le génotype détermine toujours le phénotype selon les mêmes principes de synthèse protéique et en utilisant toujours les mêmes codes génétiques.

Tous les êtres vivants, ont donc entre eux des relations de parentés plus ou moins étroites que l'analyse phylogénétique tente d'étudier

Parmi entre autres rôles, La phylogénie permet:

- De classer les êtres vivants et d'étudier leur biodiversité;
- De faire des études épidémiologiques;
- De mesurer le taux de mutation;
- De déterminer l'origine géographique des espèces;
- De dater les événements évolutionnaires;
- …etc.

1. L'arbre phylogénétique:

Le support formel de la représentation, admis par tous [30], est l'arbre phylogénétique (fig.1).

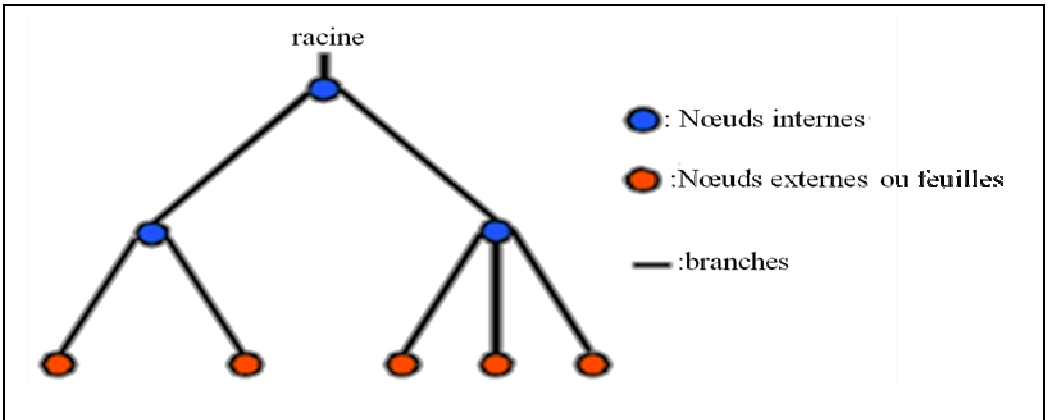

Figure 1: Un arbre phylogénétique montrant ses principaux constituants.

Les feuilles: représentent les taxons ou les unités évolutives (OTU: operational taxonomic units);

Les nœuds internes: représentent les ancêtres virtuels ayant divergé;

Les branches: définissent les relations entre les taxons en termes de descendance.

L'arbre enraciné possède une origine qui est un nœud unaire appelé racine ou ancêtre, il est orienté dans le temps, alors que l'arbre non enraciné est une représentation intemporelle des relations phylogénétiques (Fig.2).

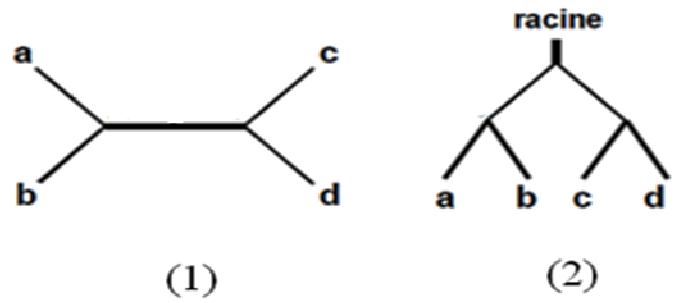

Figure 2: Arbre non enraciné (1) et arbre enraciné (2). les lettres ***a,b,c*** *et* ***d*** *représentent les taxons* [30].

1.1. Enraciner un arbre phylogénétique:

Il existe 2 principales méthodes pour enraciner un arbre [30]:

- *Méthode du groupe externe (outgroup):* Elle consiste à inclure un groupe connu a priori comme externe au groupe d'intérêt; la racine correspond alors à la branche qui relie le groupe externe aux autres séquences.
- *Méthode de mid-point*: La racine est le point de l'arbre équidistant de toutes les feuilles.

1.2. Nombre d'arbres possibles pour n taxons:

Le nombre d'arbres enracinés et non enracinés possibles pour un nombre de **n** taxons ou de **n** séquences est $\frac{(2n-3)!}{(2^{(n-2)})(n-2)!}$ et $\frac{(2n-5)!}{(2^{(n-3)})(n-3)!}$ respectivement [30]. Comme ces formules le montrent, le nombre de topologies croît exponentiellement en fonction du nombre de séquences ou de taxons (tableau 1).

Tableau 1: Nombre d'arbres possibles en fonction du nombre de séquences [30].

nombres de séquences	nombres d'arbres non enracinés possibles	nombres d'arbres enracinés possibles
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10395
8	10395	135135
9	135135	34459425
10	34459425	2.13 E 15
15	2.13 E15	8E 21

2. La phylogénie moléculaire:

Pour préciser les relations phylogénétiques entre différents taxons, il faut se limiter aux caractères communs qui peuvent apporter une information. Ces caractères communs, correspondant aux ressemblances, sont appelés homologies.

Les caractères utilisés peuvent être de nature variée: morphologiques, anatomiques, embryologiques, chromosomiques, moléculaires...etc. Il faut identifier les caractères informatifs qui peuvent présenter différents états au sein de l'échantillon. Les résultats des observations ainsi réalisées sont relevés dans une matrice taxons/caractères, où est précisé, pour chaque caractère et pour chaque organisme l'état du caractère observé.

La phylogénie moléculaire étudie l'histoire évolutive des espèces étudiées à la base d'une portion de leur séquence moléculaire. Elle est due à la découverte de la variabilité des protéines et des acides nucléiques homologues des espèces. Ses principaux avantages sont l'universalité, la rapidité et l'objectivité [30].

- ***Hypothèse de l'horloge moléculaire:***

Cette hypothèse considère que le taux d'accumulation des mutations dans le génome d'organismes différents est le même dans des régions homologues, ainsi il est possible d'estimer le temps de divergences des espèces en comparant leur diversité moléculaire [30].

3. Méthodes de reconstructions phylogénétiques:

Lors d'une reconstruction d'arbres phylogénétiques, la première étape consiste à mettre en correspondance les sites des séquences de manière à pouvoir comparer ce qui est comparable. Cette étape est nommée alignement multiple. Les séquences utilisées pour la reconstruction peuvent être de l'ADN, de l'ARN ou des séquences protéiques.

3.1. Alignement multiple des séquences:

L'Alignement multiple porte sur plusieurs séquences à la fois et dans leur intégralité (Fig.3). Il permet de mettre en évidence les relations entre ces séquences que l'on ne peut pas visualiser en comparant les séquences 2 à 2.

Les principales applications de l'alignement multiple sont:

- Caractérisation des régions conservées des protéines (motifs et domaines conservés);
- Prédiction des structures 2D et 3D par comparaison avec des structures connues;
- Construction d'arbres phylogénétiques des séquences homologues;
- …etc. [7].

Dans le cas des séquences d'ADN, elles se composent des caractères discontinus qui peuvent avoir 5 états différents: soit une adénine, soit une guanine, soit une cytosine, soit une thymine soit une insertion ou une délétion. Les sites qui ont les mêmes états dans chaque séquence s'appellent des sites conservés *(match).* Un changement d'état dans un site s'appelle une substitution *(mismatch)* ou un indel (insertion-délétion ou *gap*). Les programmes d'alignement multiple les plus utilisés sont: MUSCLE, CLUSTALW et T-Coffee.

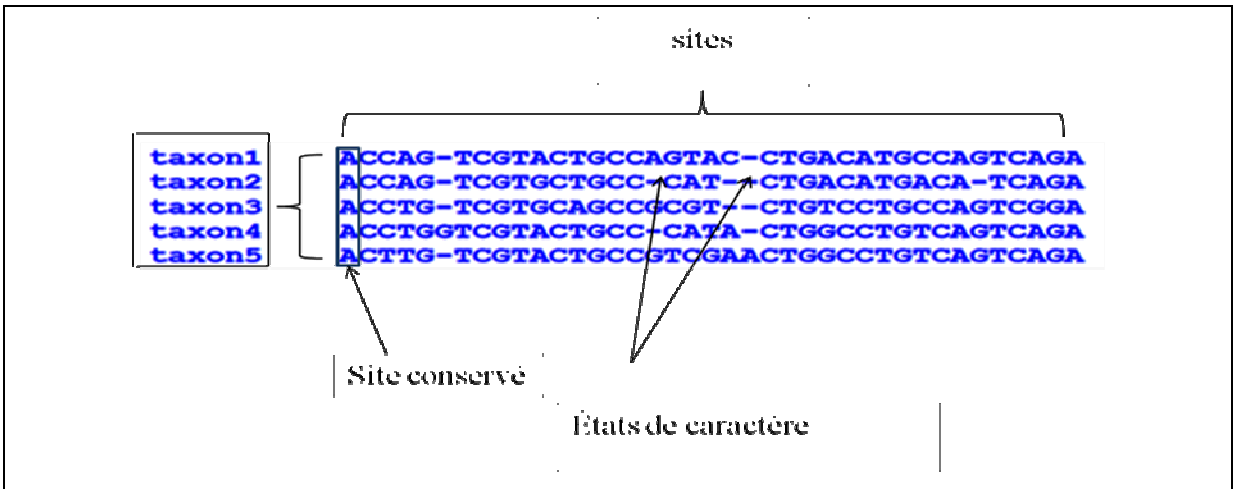

Figure 3: Alignement multiple de cinq séquences représentant cinq taxons.

Une fois les séquences alignées, une méthode de reconstruction d'arbres phylogénétiques peut être appliquée pour obtenir l'arbre qui reflète le mieux les données. Il existe actuellement au moins deux approches de reconstruction phylogénétiques: les méthodes basées sur les distances et celles basées sur les caractères [27]:

3.2. Méthodes des distances:

Ce sont des méthodes basées sur la mesure de distances entre séquences prises 2 à 2, c'est à dire le nombre de substitutions de nucléotides ou d'acides aminés entre ces 2 séquences.

Deux étapes sont nécessaires:

- Calcul des distances à partir d'un jeu de données (séquences alignées);
- Construction d'arbre phylogénétique en utilisant une méthode de distance.

3.2.1. Calcul des distances:

La méthode la plus simple de comparer deux séquences est d'évaluer leur similitude et leur différence. La similarité **(S)** entre deux séquences est égale au nombre de sites synonymes **(M)** divisé par la longueur de la séquence **(L)**. La distance observée **(D)** entre deux séquences est donnée par la formule: **D=1–S où S=M/L.** Le résultat est une matrice dite de distances [30].

- **Exemple:** le tableau 2 ci-dessous montre comment on calcule une matrice de distances à partir d'un jeu de données. Ce dernier est composé de 4 séquences nucléotidiques, chacune de celles ci est composée de 5 sites **(L=5)**.

Tableau 2: Un jeu de données composé de 4 séquences nucléotidiques (1) et la matrice de distances obtenue à partir de ce jeu de données (2). Les valeurs de la matrice sont à diviser par la longueur des séquences qui est égale à 5 [55].

Séquence 1: **AAAAA**
Séquence 2: **AAAAG**
Séquence 3: **AACGC**
Séquence 4: **ATGGT**

(1)

	1	2	3	4
1	0			
2	1	0		
3	3	3	0	
4	4	4	3	0

(2)

3.2.2. Construction d'arbre de distances:

Plusieurs méthodes ont été développées pour construire un arbre phylogénétique à partir d'une matrice de distance.

❖ ***UPGMA (Unweighted Pair Group Method withArithmetic Mean):***

C'est une méthode agglomérative qui regroupe progressivement les séquences de plus en plus proches. Elle impose que les distances soient ultra-métriques (fig.5), donc que les séquences évoluent à une vitesse constante (hypothèse d'horloge moléculaire).

A partir de la matrice de distances précédente (tableau 2), la paire la plus proche est identifiée: c'est 1 et 2, par la suite ces deux objets sont réunis en un seul (virtuel) nommé 5. La matrice de distances est recalculée ainsi sur **n-1** objets, selon la formule : **D (5, i)= [D (1, i) + D (2, i)] / 2** [30]. La nouvelle matrice de distance suivante est obtenue (tableau 3):

Tableau 3: *La nouvelle matrice de distances. Toutes les valeurs sont à diviser par 4 (L=4).*

	5	**3**	**4**
5	0		
3	2	0	
4	4	3	0

On recommence tant que la matrice contient plus de 3 lignes et colonnes. A la fin on a un arbre à trois branches qu'on redéveloppe en remontant les calculs (Fig.4).

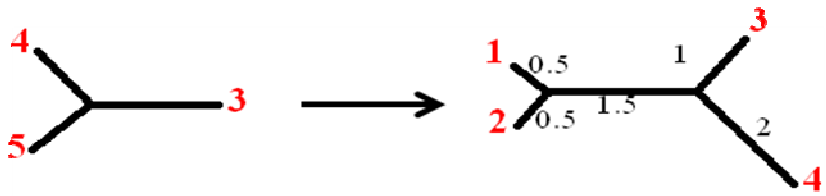

Figure 4: Arbre phylogénétique obtenu à partir du jeu de données précédent. Les numéros en gras aux sommets représentent les séquences utilisées, tandis que ceux à coté des branches représentent les distances entre ces séquences (la distance entre la séquences 1 et la séquence 3 par exemple est égale à 0.5+1.5 +1=3) [14].

L'inconvénient majeur de la méthode UPGMA est sa sensibilité à des taux de mutations différents sur les différentes branches car elle considère que toutes les branches ont le même taux de mutations [14].

❖ ***Neighbor Joining (NJ):***

Cette méthode est développée pour corriger la méthode UPGMA afin d'autoriser un taux de mutation différent sur les branches. Elle assume que les distances sont proches de

L'additivité, mais pas ultramétrique (Fig.5), donc elle n'implique pas l'hypothèse d'horloge moléculaire [37].

Figure 5: (a) Arbre "ultramétrique" obtenu par la méthode UPGMA; (b) le même arbre "additif" obtenu par la méthode de neighbor joining [37].

3.2.3. Avantages et désavantages des méthodes de distances:

❖ ***Avantages:***

Les méthodes de distances sont les seules disponibles pour analyser un certain type de données: distances immunologiques, distances d'hybridation d'acides nucléiques. Elles sont rapides et permettent d'analyser de grandes bases de données et de tester un grand nombre d'hypothèses alternatives. Elles permettent aussi d'intégrer des modèles de changements évolutifs qui ne sont pas intégrables dans d'autres méthodes [30, 14].

❖ ***Désavantages:***

Le fait de réduire la matrice de caractères à une matrice de distance induit la perte d'une certaine quantité d'information. En plus, les méthodes de distance ne permettent pas de combiner dans une même matrice des caractères de nature différente (par exemple caractères morphologiques et séquences d'ADN) [30, 14].

3.3. Méthodes de caractères:

L'approche basée sur les caractères est statistiquement plus robuste que les méthodes de distances. Mais elles sont en contre partie très lentes. Cette approche regroupe les méthodes de parcimonie (MP) et de maximum de vraisemblance (ML) [14, 30].

3.3.1. Méthode de parcimonie (MP):

Cette méthode postule que, pour un groupe d'espèces, la phylogénie la plus vraisemblable est celle qui nécessite le plus petit nombre de changements évolutifs. L'arbre phylogénétique est conçu de manière à impliquer le minimum d'événements évolutifs.

La méthode de parcimonie s'appuie sur deux hypothèses principales:

- Tous les sites changent indépendamment les uns des autres;
- La vitesse du changement est lente et constante à travers les lignées évolutives.

❖ ***Procédure d'analyse par la méthode de parcimonie:***

Les principales étapes sont:

- Identification des sites informatifs: Un site est informatif uniquement s'il y a au moins deux types de nucléotides présents dans ce site et représentés chacun dans au moins deux séquences.
- Génération de toutes les topologies d'arbres possibles pour les séquences données;
- Calcul du nombre minimum de substitutions pour chaque site informatif;
- Calcul de la somme de changements pour chaque arbre;
- Choix de la topologie de l'arbre qui nécessite le moins de changements c.à.d. l'arbre le plus court.

▪ **Exemple:** prenons le jeu de données suivant (tableau 4) qui comprend 4 séquences nucléotidiques alignées a, b, c et d:

Tableau 4: *Jeu de données composé de 4 séquences nucléotidiques. Les sites informatifs sont en gras [55].*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
a	A	A	G	A	**G**	T	**G**	C	**A**
b	A	G	C	C	**G**	T	**G**	C	**G**

c	A	G	A	T	**A**	T	**C**	C	**A**
d	A	G	A	G	**A**	T	**C**	C	**G**

Les sites informatifs sont: 5, 7 et 9. On peut donc réduire les séquences aux seuls sites informatifs (tableau 5):

Tableau 5: *Matrice réduite aux seuls sites informatifs du jeu de données précédent [55].*

	5	**7**	**9**
a	G	G	A
b	G	G	G
c	A	C	A
d	A	C	G

Pour les 4 séquences, il y a 3 arbres non enracinés possibles (Fig.6). Ces trois arbres sont analysés en recherchant la séquence ancestrale et en comptant le nombre de mutations pour chacun.

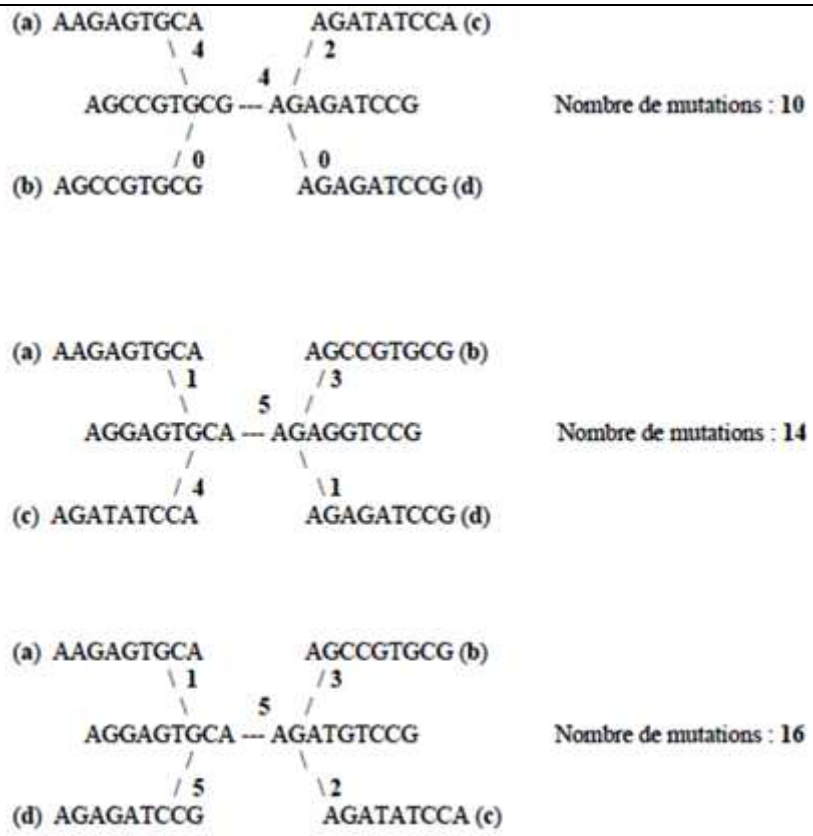

Figure 6: Les 3 arbres possibles pour le jeu de données du tableau 4 [55].

L'arbre en haut de ces trois arbres (Fig.6) est celui nécessitant le moins de mutations (10 mutations), c'est donc le plus parcimonieux.

❖ ***Avantages et inconvénients de la méthode de parcimonie:***

La parcimonie est une méthode de caractères qui fournit l'information sur les séquences ancestrales et qui permet l'évaluation des différents arbres. Cependant, seulement une partie d'information (sites informatifs) est utilisée. En plus, la méthode ne corrige pas les substitutions multiples et ne calcule pas les longueurs de branches et elle est aussi lente et inutilisable lorsque l'on a un grand nombre de séquences.

3.3.2. Méthode de maximum de vraisemblance (ML):

C'est une méthode probabiliste qui recherche l'arbre optimal en attribuant une probabilité à chaque changement dans les séquences.

La vraisemblance **L** est la probabilité d'observer les données **D** sachant que l'on considère l'hypothèse **H**, **L = Pr (D/H).**

La démarche consiste à rechercher la vraisemblance des données **D** sous différentes hypothèses évolutives **H** d'un modèle **M** et à retenir les hypothèses qui rendent cette vraisemblance maximale. Les données **D** sont des séquences comparées et l'hypothèse **H** est l'arbre phylogénétique. Le but est de trouver l'arbre dont la vraisemblance est maximale, en considérant un certain modèle d'évolution [30,11]

Deux hypothèses principales sont posées:

- les changements affectant la même séquence sont indépendants;
- les changements, affectant l'ensemble des séquences étudiées, sont indépendants.
  - **Exemple:** soit le jeu de données suivant (tableau 6) constitué de 4 séquences a, b, c et d.

Tableau 6: *Un jeu de données comprenant 4 séquences nucléotidiques. "j" est un site parmi N sites* [55].

	1					**j**
a)	C	G	A	G	A	**C**
b)	A	G	C	G	A	**C**
c)	A	G	A	T	T	**A**
d)	G	G	A	T	A	**G**

Plusieurs arbres sont possibles à partir de ce jeu de données. On veut estimer la probabilité que l'arbre suivant (Fig.7) soit le plus vraisemblable. La vraisemblance de l'arbre est en général indépendante de la position de la racine, il peut être donc enraciné d'une manière arbitraire (Fig.8).

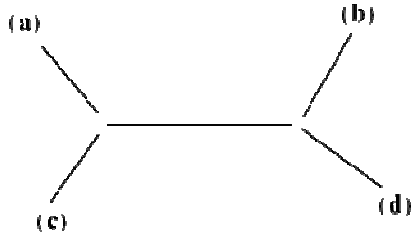

Figure 7: L'arbre dont on veut calculer la vraisemblance.

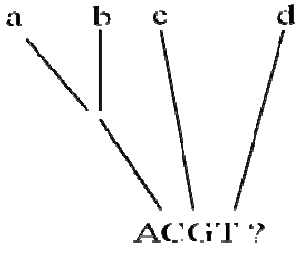

Figure 8: Une des possibilités d'enracinement de l'arbre.

La vraisemblance de cet arbre est égale à la somme des logs de vraisemblance de chaque site, ce qui revient à calculer la vraisemblance de chaque site.

$$\ln L = \ln L1 + lnL2 + lnL3 + \cdots + lnLN = \sum_{j=1}^{N} lnLj$$

- Ln L= log de la vraisemblance de l'arbre;
- ln Lj= log de la vraisemblance d'un site j tel que "j=1 à N" [11].

La vraisemblance du site "j" (Fig.9) par exemple est la somme des probabilités de toutes les possibilités de reconstruction de l'état ancestral sous le modèle choisi.

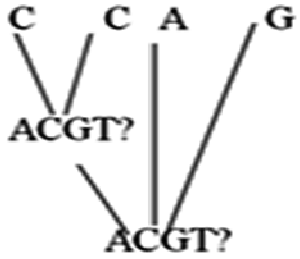

Figure 9:La vraisemblance au site "j".

Cette procédure est répétée pour tous les arbres possibles et l'arbre ayant la plus haute vraisemblance est choisi [55].

❖ ***Avantages et inconvénients de la méthode ML:***

La méthode de **ML** est considérée comme la plus fiable de toutes les méthodes phylogénétiques, c'est celle qui conduit au résultat le plus proche de l'arbre évolutif réel. En plus, elle permet d'appliquer les différents modèles d'évolution (p.ex. le modèle de Kimura qui tient compte de différences entre transitions et transversions) et d'estimer la longueur des branches en fonction de changement évolutif. Par contre, c'est la méthode qui demande la plus grosse puissance de calcul et prend le plus de temps.

La méthode du maximum de vraisemblance a ainsi bénéficié de nombreuses améliorations successives visant à décrire au mieux les processus d'évolution des séquences biologiques. Cette méthode est particulièrement performante et ses propriétés statistiques séduisantes en font la

méthode la plus utilisée à l'heure actuelle dans le domaine de la reconstruction phylogénétique [30, 11].

4. Fiabilité des arbres phylogénétiques:

Le "Bootstraping" est la méthode la plus souvent utilisée pour tester la fiabilité des branches internes. Le bootstrap consiste à effectuer un tirage des sites au hasard avec remise. Chaque tirage produit un nouvel alignement "artificiel" qui est utilisé pour construire un arbre "artificiel". Pour chaque branche interne, le pourcentage des arbres "artificiels" contenant cette branche est calculé [30].

Prenons le jeu de données suivant: 4 séquences avec 9 sites, c'est l'alignement réel (tableau 7).

Tableau 7: Alignement réel de 9 séquences [55].

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
a	A	A	G	A	G	T	G	C	A
b	A	G	C	C	G	T	G	C	G
c	A	G	A	T	A	T	C	C	A
d	A	G	A	G	A	T	C	C	G

Pour calculer le bootstrap, au moins 100 nouveaux alignements "artificiels" sont réalisés, en prenant arbitrairement les 9 colonnes. Une colonne peut être présente plus d'une fois alors qu'une autre peut être totalement absente dans un alignement artificiel donné. Ainsi, 100 arbres (artificiels) sont obtenus. Ensuite pour chaque branche interne, le nombre de fois où elle est présente dans les 100 arbres est compté. Cette fréquence avec laquelle est retrouvée une branche interne est la valeur de bootstrap (plus elle est élevée plus la fiabilité de la branche est importante).

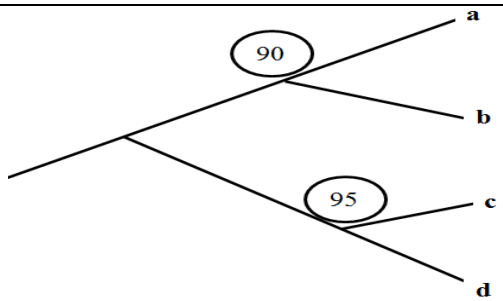

Figure 10: L'arbre phylogénétique obtenu, les valeurs 90 et 95 sont les valeurs de bootstrap [55].

Les valeurs de 90 et 95 dans l'arbre de la figure 19 indiquent que ces 2 branches sont retrouvées respectivement dans 90 et 95 des arbres artificiels. Les branches définies par une valeur de bootstrap supérieure à 95% sont considérées comme des branches fiables.

II. LES MYCOBACTERIES:

1. Généralités:

Le genre *Mycobactérium* est le seul représentant de la famille des *Mycobacteriaceae*. Depuis la découverte de *M. tuberculosis* par R. Koch, et jusqu'à maintenant, les scientifiques continuent de découvrir de nouvelles espèces (Fig.11). Aujourd'hui plus de 160 espèces mycobactériennes ont été décrites (*www.bacterio.cict.fr/index.html*) et cela Grâce à l'avènement des techniques d'identification moléculaire [46].

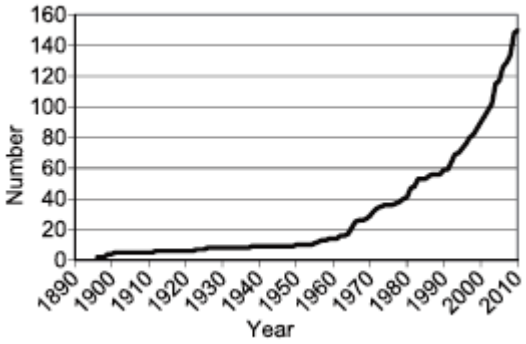

Figure 11: Nombre d'espèces/sous-espèces approuvées de 1896 à 2010 [8].

Ce genre comprend des bactéries dont le génome a une teneur élevée en G et C (de 62 à 72%) [2] et dont la paroi contient des acides mycoliques, qui n'existent que chez quelques genres voisins (*Corynebacterium*, *Nocardia*, *Rhodococcus*, *Tsukamurella*, *Dietzia* et *Gordonia*).

Les mycobactéries sont des bacilles aérobies stricts, immobiles, asporulés et acapsulés. La présence d'acides mycoliques pariétaux est responsable d'une propriété tinctoriale particulière [5]: ce sont des bacilles acido-alcoolo-résistants (BAAR). Pour les colorer, il faut donc utiliser la technique de Ziehl-Neelsen ou la coloration à l'auramine. La croissance *in vitro* des mycobactéries est souvent très lente

(de 1 à 8 semaines) et s'obtient classiquement sur des milieux spécifiques comme ceux de Löwenstein-Jensen et de Coletsos [5].

2. Tuberculose et mycobactérioses humaines:

En bactériologie médicale, classiquement trois groupes de mycobactéries sont à distinguer:

- les espèces du complexe tuberculosis, responsables de la tuberculose;
- les mycobactéries atypiques ou non tuberculeuses, responsables des mycobactérioses humaines;
- *Mycobacterium leprae* ou bacille de Hansen, agent de la lèpre, un deuxième fléau de la santé publique, après la tuberculose, qu'occupe l'OMS.

La tuberculose est une infection due le plus souvent à *M. tuberculosis*, le bacille de Koch (BK), et plus rarement à *Mycobacterium bovis* (agent de la tuberculose des bovins) ou à *M. africanum*. Dans le monde, la tuberculose est responsable de plus de 2 millions de décès par an et son incidence annuelle est de 8 à 9 millions de nouveaux cas. L'atteinte respiratoire représente 75 à 80 % des tuberculoses et les tuberculoses extrapulmonaires comprennent notamment les formes miliaires, ganglionnaires, osseuses et la méningite tuberculeuse [5].

3. Phylogénie et classification:

La taxinomie des mycobactéries est assez compliquée, la différenciation des mycobactéries en deux groupes, pathogènes et atypiques, est actuellement dominante [40]. Selon la classification de Runyon (tableau 8), basée sur le taux de croissance et la pigmentation des colonies, les mycobactéries sont divisées ainsi:

- **GROUPE O:** Les bacilles tuberculeux;
- **GROUPE I:** Mycobactéries à croissance lente, photochromogènes;
- **GROUPE II:** Mycobactéries à croissance lente, scotochromogènes;
- **GROUPE III:** Mycobactéries à croissance lente, non chromogènes;
- **GROUPE IV:** Mycobactéries à croissance rapide, pigmentées ou non.

Tableau 8: Classification des mycobactéries selon la vitesse de croissance et la pigmentation des colonies (Kanai 1990).

Croissance	Groupe	Pathogénicité pour l'homme	
		Oui	Non
Croissance lente	Bacilles tuberculeux	*M. tuberculosis* *M. bovis* *M. africanum*	*M. microti*
	I	*M. kansasii* *M. marinum* *M. simiae* *M. asiaticum*	
	II	*M. scrofulaceum* *M. szulgai*	*M. gordonae* *M. farcinogenes*
	III	*M. avium* *M. intracellulare* *M. xenopi* *M. malmoense* *M. haemophilum* *M. ulcerans*	*M. gastri* *M.nonchromogenicum* *M. terrae* *M. triviale* *M. paratuberculosis* *M. lepraemurium*
Croissance rapide	IV	*M. fortuitum* *M. chelonei subsp. chelonei* *M. chelonei subsp. abscessus*	*M. smegmatis* *M. phlei* *M. chitae* *M. flavescens* *M. parafortuitum* *M. thermoresistible* *M. aurum* *M. duvalii* *M. neoaurum* *M. glivum* *M. vaccae* *M. komossense* *M. senegalense*

Les sociétés "American Thoracic Society and the Infectious Diseases Society of America" exigent que les mycobactéries doivent être identifiées jusqu'au niveau espèces, chose qui aide beaucoup à déterminer leur signification clinique ainsi que les options d'un traitement consécutif [17].

Les méthodes d'identification basées sur les caractéristiques phénotypiques et biochimiques sont lentes et peu précises pour déterminer les espèces mycobactériennes. Aujourd'hui les méthodes moléculaires sont devenues de plus en plus utilisées. Les principaux loci largement utilisés sont: le gène *d'ARNr 16S* (*rrs*) [21], l'espace intergénique *ARNr 16S-23S (ITS1)* [1, 38, 35], et *hsp65* [33, 45]. D'autres travaux ont utilisé d'autres loci comme *gyrB* [19], *rpoB* [16, 20], *dnaJ1* [18,19], *recA* [3], *sodA* [54], *secA1* [53], *tuf* [25], *ssrA* [25], *smpB* [26], et le gène d'une protéine de 32-kDa [41].

Le problème c'est que ces loci ne sont pas amplifiés chez toutes les espèces ou ils ne permettent pas une discrimination suffisante entre les espèces fortement apparentées [58], par exemple:

Le gène d'*ARNr 16S* ne différencie pas entre *M. kansassi* pathogène de *M. gastri* non pathogène [34] ou entre *M. fortuitum subsp. Acetamidolyticum et M. fortuitum subsp. Fortuitum* qui sont différenciables par les diagnostiques cliniques traditionnelles. Ce gène ne permet pas de différencier les espèces à l'intérieur du complexe de M. *tuberculosis* MTBC ou à l'intérieur du complexe de *M. avium* (MAC) [49].

Les premiers 500 pb du gène d'*ARNr 16S* sont identiques entre *M. marinum* et *M. ulcerans* [49], entre *M. abscessus* et *M. chelonae* et entre *M. vaccae* et *M. vanbaalenii* [13]. *M. marinum* et *M. ulcerans* ont aussi l'espace intergénique ARNr 16S-23S identique [36].

Le locus *gyrB* a été testé seulement chez les mycobactéries à croissance lente (SGM) et a besoin d'une étude approfondie chez les mycobactéries à croissance rapide (RGM) [9].

hsp65 est le deuxième locus le plus conservé après le gène *ARNr 16S*, cependant il est incapable de différencier entre les membres de MTBC (à l'exception de *M. africanum*) [13, 5], entre ceux du complexe MAC ou entre les espèces de *M. fortuitum* [24].

Le locus *rpoB* est également utilisé pour identifier les espèces mycobactériennes [20], mais il n'arrive pas à différencier entre les membres de MTBC, un arbre phylogénétique basé sur le locus *rpoB* est moins robuste par rapport à un arbre basé sur *hsp65*, sur *sodA* ou sur l'*ARNr 16S* [9].

Les analyses phylogénétiques reposant sur une seule cible pour classer les bactéries au niveau genre ou espèces sont avérées peu discriminantes [6]. L'utilisation combinée de plusieurs gènes pour ce but a vu le jour depuis le rapport de la "ad hoc committee" pour la réévaluation de la définition des espèces bactériennes [42] et elle est connue sous le nom de l'analyse multigénique des séquences (MLSA pour multigene or multilocus sequence analysis). L'avantage de cette approche est sa bonne résolution étant donné que les gènes sélectionnés sont omniprésents dans le genre, présents en une seule copie et ne subissent pas de transfert horizontal (HGT pour horizontal gene transfer).

L'approche MLSA n'est pas encore exploré dans l'ensemble du genre Mycobacterium, [2, 13].

PARTIE PRATIQUE

I. MATERIEL ET METHODES:

1. Souches bactériennes utilisées:

La présente étude est focalisée essentiellement sur les membres du complexe *M. tuberculosis*: *M. tuberculosis, M. bovis, M. africanum, M. caprae, M. microti* et *M. canetti,* car, comme il s'est mentionné dans l'introduction de ce manuscrit, ce sont ceux qui posent des grands problèmes sanitaires et phylogénétiques, ainsi leur étude a été l'objectif de notre travail.

Cependant, létude a été élargie sur d'autres complexes qui posent les mêmes problèmes que le complexe MTBC mais à moindre degré, ce sont:

- Le complexe *Mycobacterium abscessus sensu lato: M. abscessus* sensu stricto*, M. bolleti* et *M. massiliense;*
- Et le complexe *Mycobacterium avium (MAC): M. avium subsp.avium, M. avium subsp.paratuberculosis* et *M. avium subsp.silvaticum.*

En compatibilité avec ce qui est recommandé [42], seulement les souches types ont été utilisées, sauf dans le cas où la séquence recherchée n'était pas disponible pour la souche type de l'espèce (tableau 16; annexes).

Pour l'étude de la variabilité intraspécifique, des souches normales ont été utilisées avec la souche type de l'espèce (tableaux 17, 18 et 19; annexes).

2. Séquences génétiques:

Comme a été signalé avant, 8 gènes de ménages ont été choisis: *ARNr 16S, hsp65, rpoB, tuf, sodA, ARN tm, gyrB* et *secA1*. Ce choix est justifié par 2 raisons:

- Ce sont des gènes jugés être pertinents pour faire la phylogénie du genre Mycobacterium [2, 26, 19, 53];
- Les séquences de ces gènes sont disponibles dans la base de données Genbank (*http://www.ncbi.nlm.nih.gov/genbank/*).

Le gène d'*ARNr 16S* est le premier gène utilisé pour l'identification bactérienne [29]. Il est généralement admis comme étant la meilleure cible pour étudier les relations phylogénétiques au niveau de l'espèce, car il est présent chez toutes les bactéries [15]. En plus, il est caractérisé par un taux d'évolution qui conduit à une variabilité intra spécifique faible et à une variabilité interspécifique significative pour plusieurs espèces [15]. En outre, il est très disponible dans les bases de données publiques (*http://www.ncbi.nlm.nih.gov/genbank/*).

Le gène *hsp65*, qui est présent dans toutes les mycobactéries, appartient à la famille des gènes des protéines de choc thermique (Hsp pour heat-shock protein). Ce gène est très conservé [13].

Le gène *rpoB* code pour la sous-unité β de l'ARN polymérase, une enzyme oligomérique responsable de la synthèse de l'ARN [13].

Le gène *tuf* code pour le facteur d'élongation Tu (EF-Tu) [39].

Le gène *ARNtm* code pour un "Transfer messenger RNA ou TmRNA" qui peut jouer à la fois le rôle d'ARNt ou d'ARNm, il contribue aussi à la dégradation des peptides incomplètement synthétisés et au recyclage des ribosomes bloqués [39].

Le gène *sodA* code pour une métallo-enzyme, une superoxyde dismutase, qui constitue l'un des principaux mécanismes de défense contre le stress oxydatif [13].

Le gène *gyrB* code pour la sous unité β de l'ADN gyrase, une enzyme universellement distribuée et essentielle pour la réplication de l'ADN [19].

Finalement, le gène *secA1* code pour une protéine SecA1 qui est l'homologue Mycobactérien de la protéine SecA d'Escherichia coli. Il est impliqué dans l'exportation (sec pour sécrétion) des protéines à travers la membrane plasmique [53.].

Les séquences de chaque gène ont été téléchargées à partir de Genbank (*http://www.ncbi.nlm.nih.gov/genbank/*), une base de données publique, puis elles sont organisées dans un classeur Excel pour faciliter leur manipulation. Par la suite, ces séquences sont analysées en effectuant un alignement multiple avec le logiciel BLAST (Basic Local Alignment Search Tool:*http://blast.ncbi.nlm.nih.gov*). Cette analyse a pour but d'uniformiser leurs tailles et de ne garder que les domaines conservés.

Le programme BLAST est basé sur un algorithme qui trouve des régions similaires entre les séquences. Le programme compare les séquences de nucléotides ou de protéines entre elles ou avec les séquences contenues dans les bases de données. BLAST peut être utilisé pour déduire les relations fonctionnelles et évolutives entre les séquences, ainsi que pour identifier les membres de familles de gènes (*http://blast.ncbi.nlm.nih.gov*).

3. Analyse phylogénétique des séquences:

Les séquences ainsi triées manuellement, sont soumises à une analyse phylogénétique grâce à l'interface web (*http://phylogeny.lirmm.fr/phylo_cgi/simple_phylogeny.cgi*) (Fig.12).

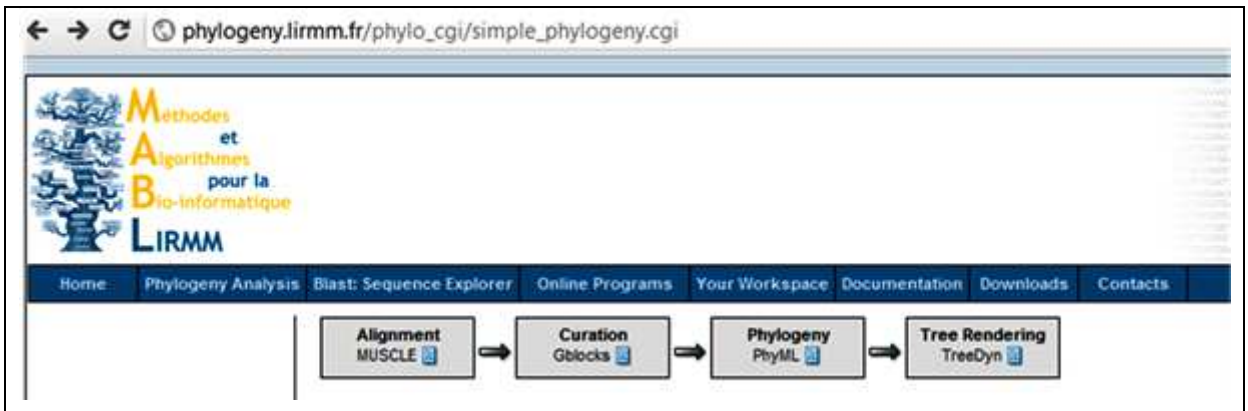

Figure 12: L'interface web utilisée pour obtenir des arbres phylogénétiques. Noter les principales étapes de l'analyse: alignement, curation, phylogénie et finalement dessin de l'arbre.

Cette interface comprend un ensemble de programmes informatiques qui génèrent des arbres phylogénétiques en passant par les étapes suivantes:

- **1ère étape: Alignement multiple**; le programme MUSCLE réalise un alignement multiple des séquences nucléotidiques ou peptidiques (tableau 9). Ce programme est plus précis et plus rapide que d'autres programmes comme CLUSTAW ou T-Coffee [15];

Tableau 9: Alignement multiple par MUSCLE en format CLUSTAL; les astérisques indiquent que les sites sont identiques; gène de l'ARNr 16S.

```
Nocardia_f      GTGCTTAACACATGCAAGTCGAGCGGTAAGGCC-CTTCGG-GGTACACGAGCGGCGAACG
M._abscess      GTGCTTAACACATGCAAGTCGAACGGAAAGGCC-CTTCGG-GGTACTCGAGTGGCGAACG
M._bolleti      GTGCTTAACACATGCAAGTCGAACGGAAAGGCC-CTTCGG-GGTACTCGAGTGGCGAACG
M._massili      GTGCTTAACACATGCAAGTCGAACGGAAAGGCC-CTTCGG-GGTACTCGAGTGGCGAACG
M._tubercu      GTGCTTAACACATGCAAGTCGAACGGAAAGGTCTCTTCGGAGATACTCGAGTGGCGAACG
M._african      GTGCTTAACACATGCAAGTCGAACGGAAAGGTCTCTTCGGAGATACTCGAGTGGCGAACG
M._caprae       GTGCTTAACACATGCAAGTCGAACGGAAAGGTCTCTTCGGAGATACTCGAGTGGCGAACG
M._microti      GTGCTTAACACATGCAAGTCGAACGGAAAGGTCTCTTCGGAGATACTCGAGTGGCGAACG
M._canetti      GTGCTTAACACATGCAAGTCGAACGGAAAGGTCTCTTCGGAGATACTCGAGTGGCGAACG
M._bovis_C      GTGCTTAACACATGCAAGTCGAACGGAAAGGTCTCTTCGGAGATACTCGAGTGGCGAACG
M._avium_p      GTGCTTAACACATGCAAGTCGAACGGAAAGGCCTCTTCGGAGGTACTCGAGTGGCGAACG
M._avium_u      GTGCTTAACACATGCAAGTCGAACGGAAAGGCCTCTTCGGAGGTACTCGAGTGGCGAACG
M._avium_b      GTGCTTAACACATGCAAGTCGAACGGAAAGGCCTCTTCGGAGGTACTCGAGTGGCGAACG
                ********************** *** **** * ****** * *** **** ********
```

- **2ème étape: curation;** le programme Gblocks [4] élimine les positions mal alignées ainsi que les régions divergentes, pour ne laisser que les parties conservées et bien alignées sans insertions/délétions (Fig.13).

Figure 13: Curation des séquences alignées; les régions soulignées en bleu sont les seules qui seront prises en compte dans l'analyse phylogénétique; gène de l'ARNr 16S.

- **3ème étape: phylogénie;** PhyML est un logiciel dédié à l'analyse phylogénétique [18] basé sur le principe du maximum de vraisemblance (ML), comme s'est mentionné précédemment, cette méthode probabiliste génère, parmi plusieurs arbres phylogénétiques possibles, celui qui est le plus vraisemblable (Fig.14).

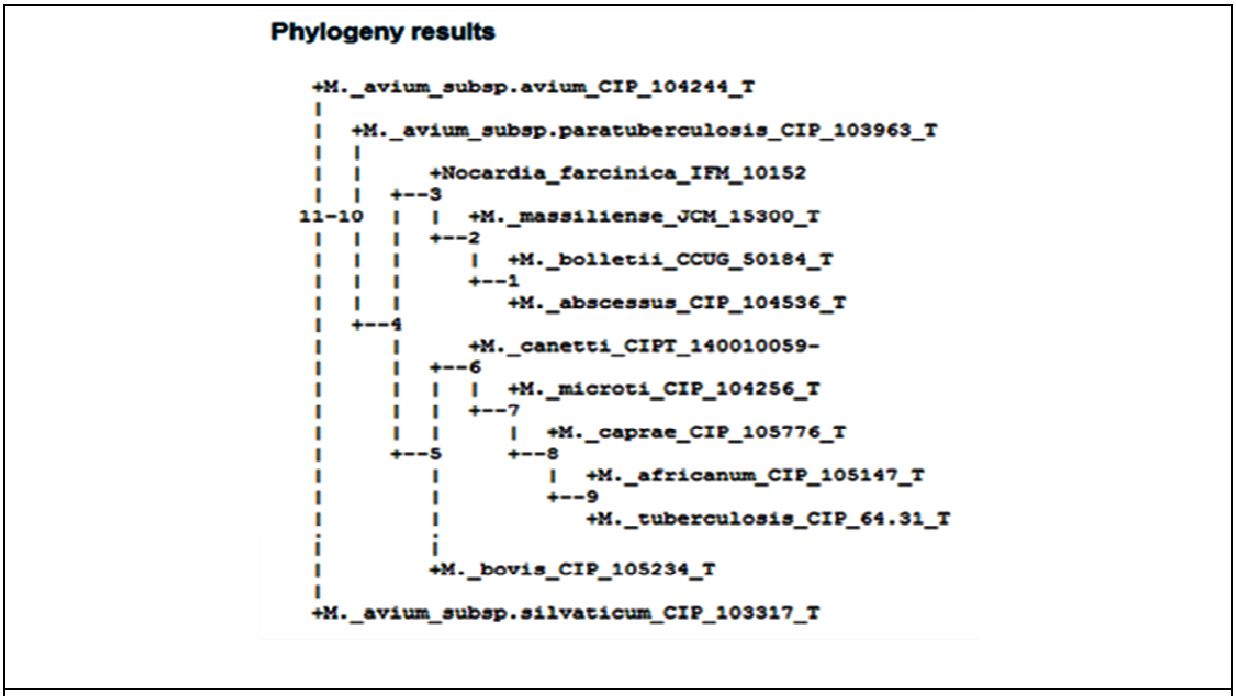

Figure 14: Résultat obtenu après analyse phylogénétique; gène de l'ARNr 16S.

- **4ème étape: dessin de l'arbre phylogénétique;**

Finalement les arbres (fig.15) sont dessinés et visualisés par le logiciel TreeDyn [12].

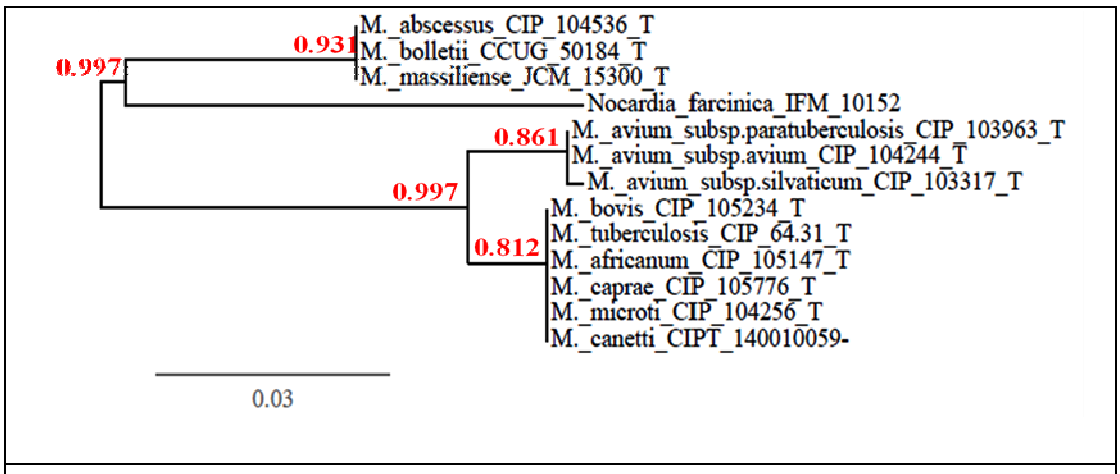

Figure 15: Arbre phylogénétique généré par TreeDyn. Les numéros en gras sont les valeurs de bootstrap qui renseignent sur le degré des soutiens des branches; la longueur des branches est proportionnelle au nombre de substitutions par site; cas

du gène de l'ARNr 16S.

La procédure de "bootstraping" est utilisée pour évaluer le degré de soutien des branches et ainsi la robustesse des arbres.

Dans un premier temps, l'analyse phylogénétique a été effectuée avec une seule séquence, les arbres ainsi obtenus sont basés sur une seule séquence.

Dans un deuxième temps, les arbres sont obtenus non pas à partir d'une seule séquence, mais à partir de plusieurs séquences concaténées. Pour concaténer les séquences, une fonction de concaténation Excel a été utilisée. Le but est de détecter la combinaison de gènes qui permet d'avoir la phylogénie la plus robuste.

4. Etude de la variabilité intraspécifique:

L'étude de la variabilité intraspécifique pour un gène donné est un aspect très important; cette dernière doit être nulle sinon très faible car, si elle est grande, des souches appartenant à la même espèce seront classées dans des espèces différentes.

Parmi les 8 loci, l'étude de la variabilité intraspécifique a été réalisée pour 3 loci (*ARNr 16S*, *hsp65*, *gyrB* et *tuf*) pour les membres du complexe MTBC. En plus de la souche type de l'espèce considérée, d'autres souches appartenant à la même espèce ont été incluses (tableaux 17, 18 et 19; annexes). Le nombre des souches incluses a été déterminé par la disponibilité des séquences dans la base de données, c'est pourquoi, pour une séquence donnée, certaines espèces ne sont représentées que par une seule souche qui est la souche type. En utilisant BLAST (http://blast.ncbi.nlm.nih.gov), la mesure de la variabilité intraspécifique est effectuée en déterminant le pourcentage d'identité entre les séquences prises 2 à 2.

Tableau 10:Nombre de souches utilisées pour chaque gène.

gène / espèce	***hsp65***	***tuf***	***gyrB***
M. tuberculosis	14	14	15
M. bovis	3	2	4
M. africanum	3	2	2

M. microti	2	1	1
M. caprae	1	1	1
M. canetti	3	1	4
nombre total	**26**	**21**	**27**

II. RESULTATS ET DISCUSSION:

Cet axe comprend les sous-axes suivants:

- Les tailles des séquences triées par le programme BLAST;
- Les arbres phylogénétiques obtenus par l'analyse individuelle des différents loci;
- Les résultats de l'étude de la variabilité intraspécifique;
- Les arbres phylogénétiques générés par l'analyse combinée des 4 loci: *ARNr 16S, hsp65, tuf* et *gyrB*.

1. Taille des séquences:

Les séquences récupérées à partir de Genbank *(http://www.ncbi.nlm.nih.gov/genbank/)*, sont ensuite analysées par le programme BLAST *(http://blast.ncbi.nlm.nih.gov)* dans le but d'uniformiser leurs tailles, ces dernières sont devenues ainsi:

- 525 à 541 pb pour les séquences d'*ARNr 16S*;
- 372 pb pour les séquences de *hsp65*;
- 359 pb pour les séquences de *rpoB*;
- 629 à 632 pb pour les séquences de *tuf*;
- 414 pb pour les séquences de *sodA*;
- 287 à 289 pb pour les séquences de *l'ARNtm*;
- 943 pb pour les séquences de *gyrB*;
- 463 pb pour les séquences de *secA1*.

Il est à signaler que la différence de tailles à l'intérieur des gènes *d'ARNr 16S, d'ARNtm,* et de *tuf* est due à des insertions/délétions.

2. La phylogénie avec un seul gène:

8 arbres sont obtenus chacun à partir de l'un des 8 gènes utilisés (*ARNr 16S*, *hsp65*, *rpoB*, *tuf*, *sodA*, *ARNtm*, *gyrB* et *secA*).

Pour chacun des 5 premiers gènes, 12 souches mycobactériennes ont été incluses dans l'analyse. A l'exception de *M. canetti*, ces souches sont toutes des souches types.

Nocardia abscessus, *Nocardia farcinica* et *Rhodococcus corynebacterioides* sont utilisées comme groupes externes "outgroup" pour enraciner les arbres.

2.1. Arbre phylogénétique obtenu avec le gène *ARNr 16S*:

Les séquences des membres des complexes MTBC et MAC ont une taille de 541 pb, celles du complexe *M. abscessus* ont une taille de 527 pb. La séquence de *Nocardia abscessus* IFM 10152 "outgroup", a une taille de 525 pb. Comme il est signalé ci-dessus, cette différence de taille est due à des insertions/délétions.

L'Arbre phylogénétique généré par le gène de *l'ARNr 16S* (Fig.16) montre que Les 3 complexes étudiés ont été séparés l'un de l'autre avec des valeurs de "bootstrap" plus au moins importantes: 0.785, 0.809 et 0.923 pour le complexe MTBC, le complexe MAC et le complexe *M. abscessus*, respectivement. Les 2 complexes MTBC et MAC ont été considérés comme un seul groupe soutenu par un nœud ayant une valeur de "bootstrap" égal à 0.971.

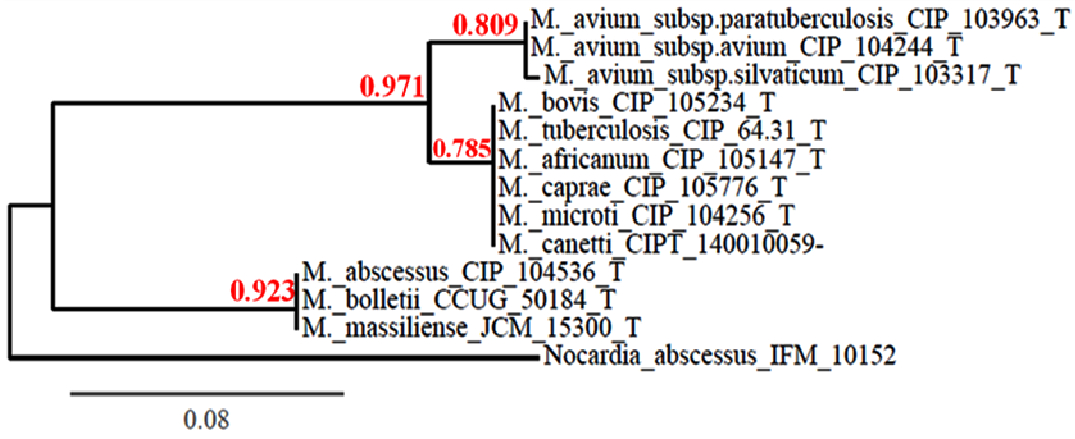

Figure 16: Arbre phylogénétique obtenu avec la séquence du gène d'ARNr 16S. Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. La longueur de la barre en dessous de l'arbre indique 0.08 substitutions par site. T: souche type.

A l'exception de la souche *M. avium subsp. Silvaticum* qui a été séparée des autres membres du complexe MAC, aucune autre souche n'a été identifiée jusqu'au niveau espèce. Le nombre de substitutions par site est égal à 0.08, une valeur qui est inférieure aux valeurs trouvées pour les autres loci. Aucun des nœuds des 3 complexes n'a été soutenu par un "bootstrap" supérieur à 0.95**.** De ce fait, et bien que largement utilisé dans la taxonomie et l'identification [13], ce gène n'est pas approprié pour différencier entre les espèces très proches comme celles qui forment le complexe MTBC.

2.2. Arbre phylogénétique obtenu avec le gène *hsp65*:

Comme pour le gène de *l'ARNr 16S*, l'utilisation du gène *hsp65* (Fig.17) permet de différencier entre les 3 complexes mais avec des valeurs de "bootstrap" sensiblement supérieures: 0.985, 0.979 et 0.852 au lieu de 0.785, 0.809 et 0.923 (Fig.16), respectivement pour les complexes MTBC, MAC et *M. abscessus*.

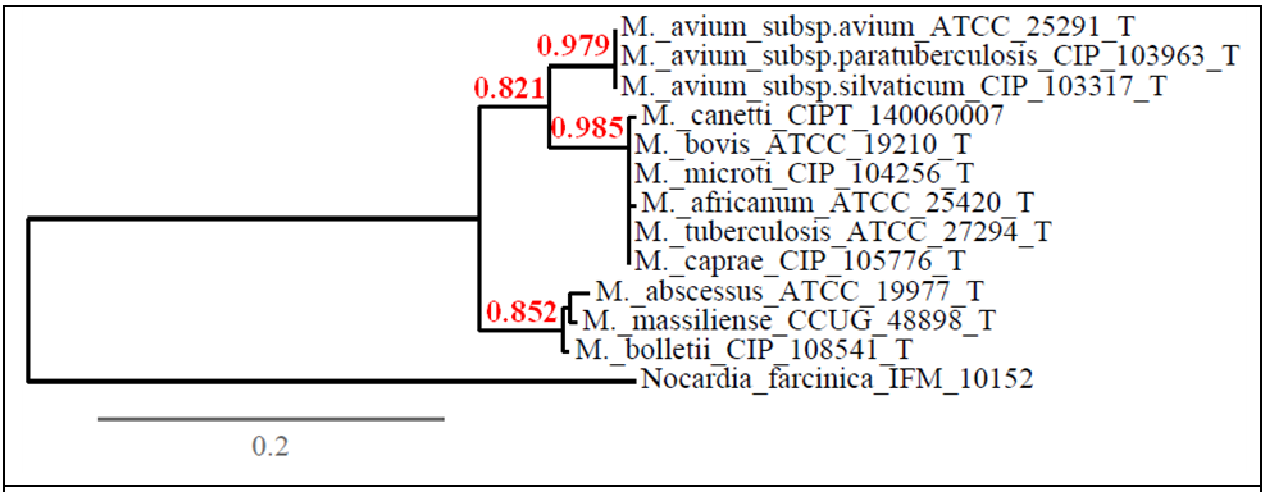

Figure 17: Arbre phylogénétique obtenu avec le gène hsp65. Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. La longueur de la barre en dessous de l'arbre correspond à 0.2 substitution par site. T: souche type.

A la différence du gène de *l'ARNr 16S*, le gène *hsp65* permet de distinguer entre les 3 membres du complexe *M. abscessus.* Il permet aussi de distinguer 2 espèces parmi les 6 espèces du complexe MTBC, ce sont *M. canetti* et *M. africanum*, ces dernières diffèrent de leurs voisins MTBC par 2 SNP (single-nucleotide polymorphism) au niveau des positions 207 et 348 respectivement (Fig.18): *M.*

canetti a une thymine au lieu d'une cytosine dans la position 207 et *M. africanum* possède une guanine au lieu d'une cytosine au niveau de la position 348.

Le nombre des substitutions par site (0.2) est supérieur à celui *d'ARNr 16S* (0.08) ce qui confirme que le gène *hsp65* est plus variable que le gène *ARNr 16S* [32]. 2 nœuds parmi les 3 ont des valeurs de "bootstrap" supérieures à 0.95.

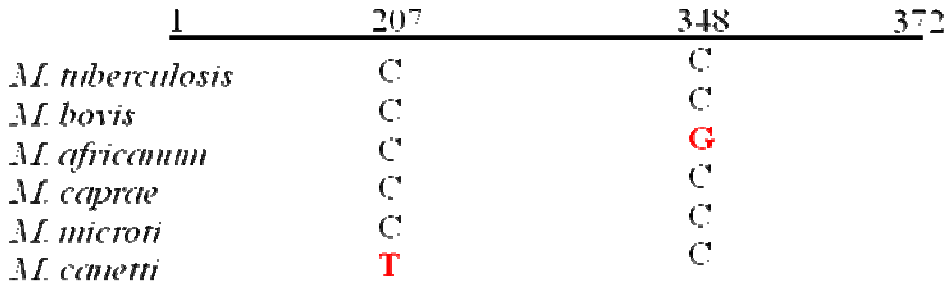

Figure 18: Carte physique de la séquence hsp65. Les deux SNP (en gras) permettent de différencier M. canetti et M. africanum des autres membres du complexe MTBC.1: début de la séquence; 372: fin de la séquence.

Par rapport à l'arbre *d'ARNr 16S* (Fig.16), Cet arbre montre donc un bon pouvoir de discrimination ainsi qu'une relative robustesse Un constat déjà rapporté par Devulder et al. [13]. Cependant, les membres du complexe avium ne sont pas différenciés ainsi que 4 espèces du complexe MTBC (Fig.17).

2.3. Arbre phylogénétique obtenu avec le gène *rpoB*:

Comme les 2 gènes précédents (*ARNr 16S* et *hsp65*), l'arbre du locus *rpoB* (Fig.19) permet de différencier entre les 3 complexes en question. Le nombre de substitution par site est 0.2, une valeur égale à celle du locus *hsp65* ce qui signifie que *rpoB* a la même variabilité que *hsp65*.

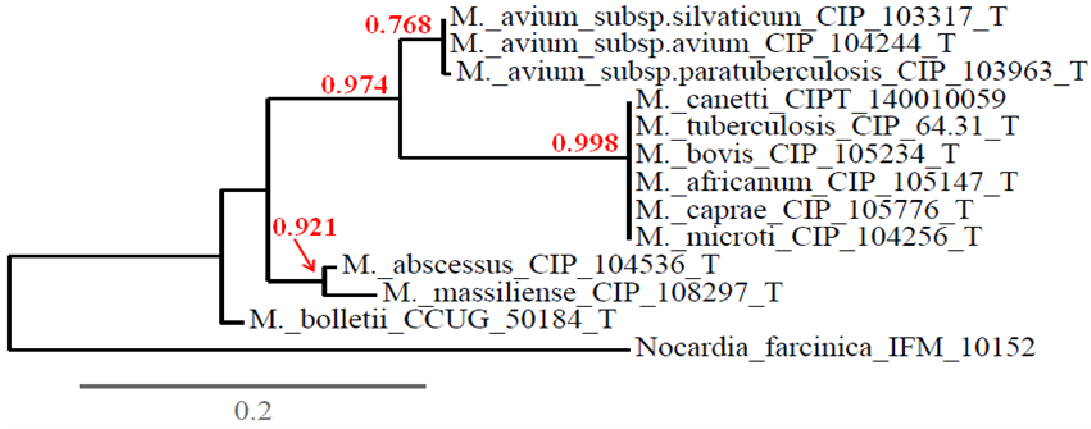

Figure 19: Arbre phylogénétique généré par le gène rpoB. Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. La longueur de la barre en dessous de l'arbre indique 0.2 substitutions par site. T: souche type.

Aucun membre du complexe MTBC n'a été différencié de ces voisins. Un seul membre du complexe MAC (*M. avium subsp.paratuberculosis*) a été distingué des 3 autres. Les 3 membres du complexe *M. abscessus* ne sont pas considérés comme un seul groupe parce que *M. bolleti* a été placée avant le nœud qui soutient *M. abscessus* et *M. massiliense.* Il y a seulement un seul nœud, parmi les 3, qui a une valeur de "bootstrap" supérieure à 0.95.

Aucun membre du complexe MTBC n'a été différencié jusqu'au niveau espèce. Le locus *rpoB* a été recommandé, comme un gène additionnel du locus *ARNr 16S*, pour étudier la phylogénie de des mycobactéries [20].

2.4. Arbre phylogénétique obtenu avec le gène *tuf*:

En se basant sur l'arbre inféré (Fig.20), Le gène *tuf* discerne bien entre les 3 complexes étudiés. Il est aussi capable de distinguer les espèces suivantes:

- *M. canetti* et *M. microti* des autres membres du complexe MTBC;
- Tous les membres du complexe *M. abscessus*;
- *M. avium subsp. avium* des autres membres du complexe MAC.

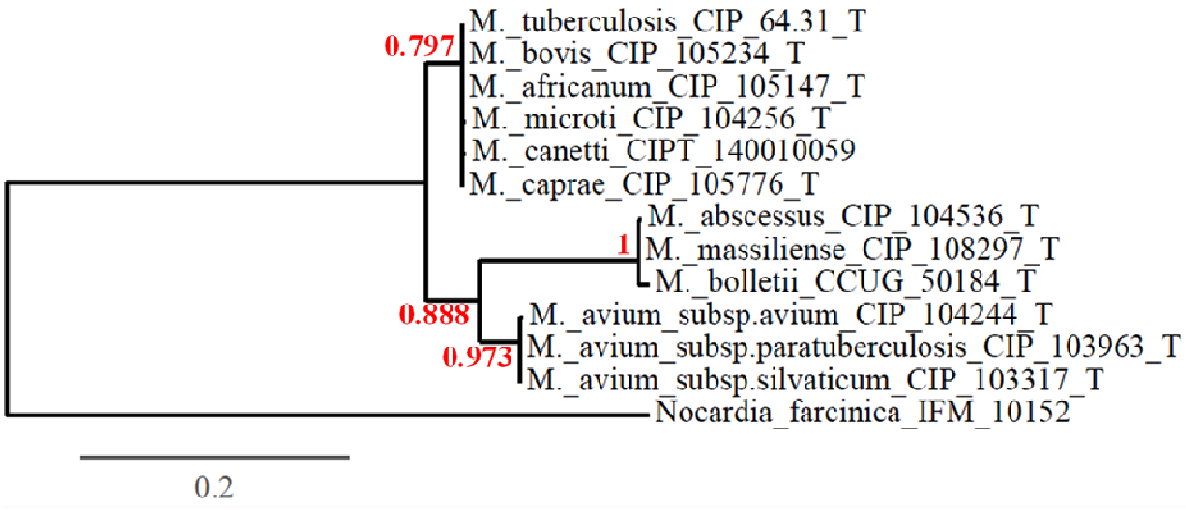

Figure 20: Arbre phylogénétique obtenu par le gène tuf. Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. La longueur de la barre en dessous de l'arbre indique 0.2 substitutions par site. T: souche type.

M. canetti et *M. microti* sont différenciées de leurs voisins par 2 SNP au niveau des positions 176 et 623 respectivement (Fig.21): *M. canetti* possède une guanine au lieu d'une cytosine au niveau de la position 176 alors que *M. microti* a une adénine au lieu d'une cytosine dans la position 623.

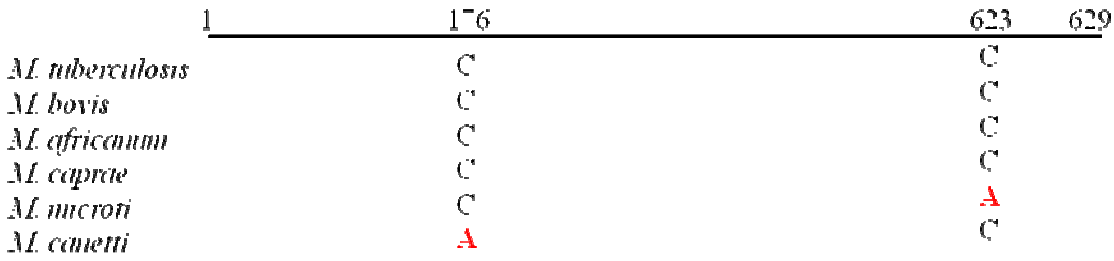

Figure 21: Carte physique de la séquence tuf. Les deux SNP (en gras) par lesquels diffèrent M. canetti et M. microti des autres membres du complexe MTBC.1: début de la séquence; 629: fin de la séquence.

Parmi les 3 nœuds (Fig.20), ceux qui soutiennent le complexe MAC et le complexe *M. abscessus*, ont des valeurs de "bootstrap" supérieures à 0.95. En harmonie avec ce qui est signalé par d'autres travaux [25], Le gène tuf possède une variabilité similaire à celles de *hsp65* et *rpoB* (0.2 substitutions par site).

Le gène *tuf* possède ainsi un pouvoir résolutif et une robustesse proches de *hsp65*, de ce fait il est suggéré comme un candidat fiable pour faire la phylogénie des mycobactéries [25]. Toutefois, il n'arrive pas à séparer toutes les espèces du complexe MTBC ou du complexe MAC.

2.5. Arbre phylogénétique du gène *sodA*:

D'après l'arbre obtenu (Fig.22), le locus *sodA* est capable de:

- Séparer entre les 3 complexes;
- Séparer les membres du complexe MAC;
- Séparer les membres du complexe *M. abscessus*.

Aucun membre du complexe MTBC n'a été discerné des autres. 2 des 3 nœuds ont une valeur de "bootstrap" supérieure à 0.95 (Fig.22). Le taux de substitutions par site (0.2) équivaut à ceux de *tuf*, *rpoB* et *hsp65*.

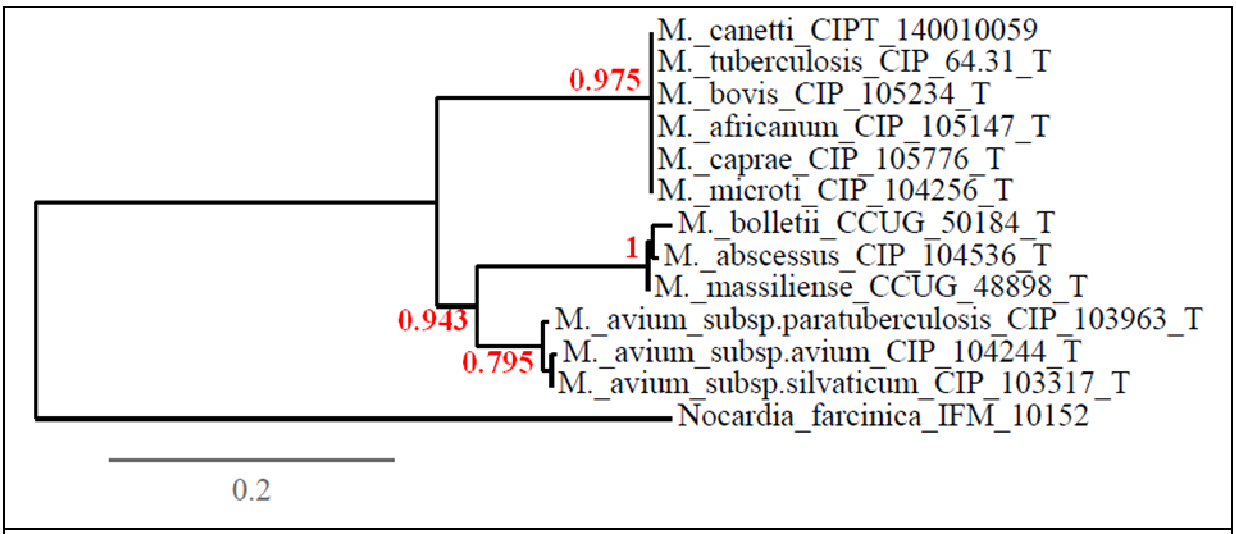

Figure 22: Arbre phylogénétique obtenu par le gène sodA. Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. La longueur de la barre en dessous de l'arbre indique 0.2 substitutions par site. T: souche type.

Vue sa variabilité élevée [13] et son taux d'évolution rapide [26], il semble que *sodA* pourrait être très utile pour faire une phylogénie avec d'autres gènes de ménages ayant une faible variabilité comme le gène de *l'ARNr 16S*. Mais, comme les loci précédents, il est incapable de séparer les membres du complexe MTBC.

2.6. Arbre phylogénétique obtenu avec le gène *d'ARNtm*:

En plus du groupe externe *Nocardia farcinica* IFM 10152 utilisée pour l'enracinement de l'arbre, 11 souches seulement et pas 12 ont été incluses dans cette analyse car la séquence de *l'ARNtm* de *M. canetti* est indisponible dans la base de données Genbank.

Ce qui est important dans l'arbre (Fig.23) généré par de ce locus est que:

- Les 3 complexes sont différenciables;
- Aucun membre des 3 complexes n'a été différencié des autres;
- 2 des 3 nœuds ont un "bootstrap" supérieur à 0.95.

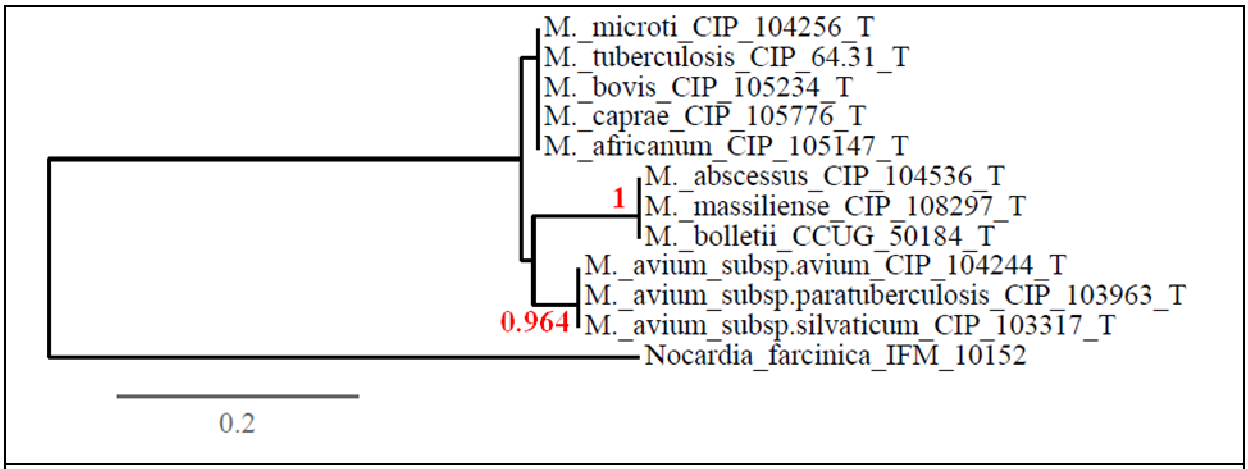

Figure 23: Arbre phylogénétique inféré par le gène d'ARNtm. Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. La longueur de la barre en dessous de l'arbre indique 0.2 substitutions par site. T: souche type.

Vu l'objectif fixé pour ce travail, ce locus ne jouera pas un rôle important.

2.7. Arbre phylogénétique obtenu avec le gène gyrB:

11 souches au lieu de 12 sont analysées. La séquence *gyrB* pour *M. avium subsp. silvaticum* est absente dans la base de données Genbank. *Rhodococcus corynebacterioides* IFO 14404 a servi du groupe externe.

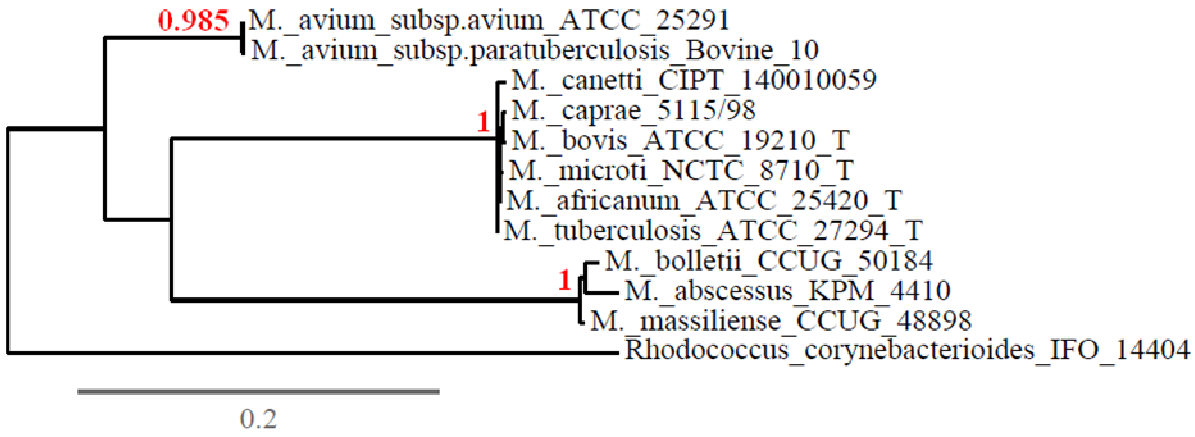

Figure 24: Arbre phylogénétique obtenu par le gène gyrB. Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. La longueur de la barre en dessous de l'arbre correspond à 0.2 substitution par site. T: souche type.

L'arbre obtenu (Fig.24) montre qu'il ya une distinction entre:

- Les 3 complexes analysés;
- Les membres du complexe MTBC;
- Les membres du complexe *M. abscessus*;

Les 3 nœuds ont une valeur de "bootstrap" supérieure à 0.95.

Les membres du complexe MTBC sont différenciés entre eux par des SNP présents au niveau des sites 54, 114, 135, 390, 690,789, 829 et 837 (Fig.25).

	54	114	135	390	690	789	829	837
M. Tuberculosis (9 10)	C	C	G	C	T	C	**G**	C
M. Bovis (4 4)	C	C	**A**	C	T	**T**	T	C
M. Africanum (2 2)	C	C	G	C	T	C	T	C
M. Caprae (2 2)	C	C	**A**	C	**G**	C	T	C
M. Microti (1 1)	**T**	C	G	C	T	C	T	C
M. canetti CIPT 140010059	C	**T**	G	**T**	T	C	**G**	**T**
M. canetti CIPT 140060005	C	C	G	C	T	C	**G**	**T**
M. canetti CIPT 140060007	C	C	G	C	T	C	**G**	**T**
M. canetti CIPT 140060001	C	**T**	G	**T**	T	C	**G**	**T**

Figure 25: Carte physique de la séquence gyrB. Les 8 SNP (en gras et en rouge) qui différencient entre les membres du complexe MTBC. les 4 positions 54, 135, 789 et 829 sont décrites par Kasai et al. [19], la position 690 a été rapportée par NIEMANN et al. [28], les positions 114, 390 et 837 sont

détectées dans cette étude. Noter que 2 sur 4 seulement des souches de M. canetti qui ont ces 3 nouveaux SNP. Les rapports renseignent sur le nombre de souches ayant le SNP en question sur le nombre total des souches incluses. 1: début de la séquence; 943: fin de la séquence.

Il est à signaler que les positions 54, 135, 789 et 829 sont décrites par Kasai et al. [19], la position 690 a été rapportée par NIEMANN et al. [28] alors que les positions 114, 390 et 837 sont détectées dans cette étude. Ces nouvelles positions permettent de distinguer *M. canetti* des autres membres. Cependant seulement une position (position 837), de ces 3 nouvelles positions, est conservée chez toutes les souches de *M. canetti* étudiées, on peut dire donc que c'est cette position qui permet réellement de différencier *M. canetti* des autres MTBC.

Il apparait donc que le locus *gyrB* est doué d'un pouvoir résolutif très important permettant de différencier entre les espèces les plus proches comme celles du complexe MTBC. Du fait que les 3 nœuds ont une valeur de "bootstrap" supérieure à 0.95, l'arbre phylogénétique fourni par *gyrB* est robuste. Ainsi, et comme il est rapporté auparavant [19, 28], *gyrB* est très utile pour identifier ou faire la phylogénie de Mycobactéries.

Toutefois, les membres du complexes MAC n'ont pas été séparés.

2.8. Arbre phylogénétique obtenu avec le gène *secA1*:

Les séquences *secA1* de 2 souches manquent car elles n'existent pas dans la base de données Genbank. Pour cela l'analyse a porté seulement sur 10 souches, même le groupe externe n'a pas été introduit du fait qu'une souche qui pourrait remplir efficacement ce rôle n'a pas été trouvée, l'arbre est alors enraciné par la méthode alternative de "mid-point".

Selon l'arbre obtenu (Fig.26), secA1 permet de différencier:

- Les membres du complexe *M. abscessus*;
- Les membres du complexe MAC;
- *M. canetti* des autres MTBC.

Un seul nœud qui a une valeur de "bootstrap" supérieur à 0.95. En outre, Ce locus est incapable de différencier 5 espèces du complexe MTBC.

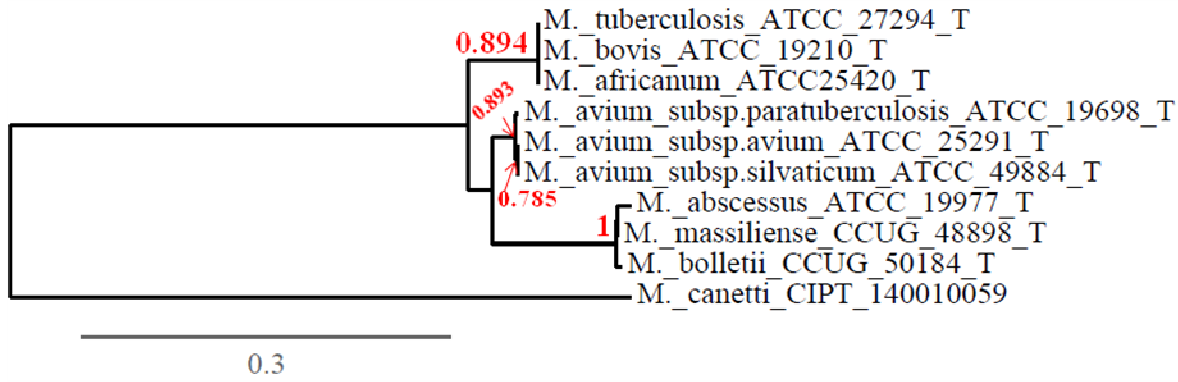

Figure 26: Arbre phylogénétique obtenu à partir du gène secA1. Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. La longueur de la barre en dessous de l'arbre indique 0.2 substitutions par site. T: souche type.

2.9. Synthèse des résultats de l'analyse phylogénétique avec un seul locus:

Les 2 tableaux 11 et 12 représentés ci-dessous résument les principaux résultats de l'analyse phylogénétique précédente, ils permettent de faire une comparaison rapide entre ces résultats.

Tableau 11: Capacité des différents loci à distinguer entre les membres de chacun des trois complexes étudiés. Pour chaque locus, les rapports renseignent sur le nombre des espèces différenciables d'un complexe sur le nombre total des membres qui forment celui-ci.

Locus / Propriété		***16S ARNr***	***hsp65***	***rpoB***	***tuf***	***sodA***	***ARN tm***	***gyrB***	***secA1***
proportion des membres différenciables	complexe MTBC	0/6	2/6	0/6	2/6	0/6	0/6	6/6	1/6
	complexe MAC	1/3	0/6	1/3	1/3	3/3	0/3	0/3	3/3
	complexe *M. abscessus*	0/3	3/3	3/3	3/3	3/3	0/3	3/3	3/3

Tableau 12: Résultats obtenus pour: les valeurs du "bootstrap" supportant les nœuds des 3 complexes, la proportion des nœuds ayant un "bootstrap" > 0.95 et le nombre de substitutions par site. Le signe – indique que le nœud est absent.

Locus / propriété		***ARNr 16S***	***hsp65***	***rpoB***	***tuf***	***sodA***	***ARN tm***	***gyrB***	***secA1***
"bootstrap"	MTBC	0.785	0.985	0.998	0.797	0.975	0.65	1	0.894
	MAC	0.809	0.979	0.768	0.973	0.795	0.964	0.985	0.893
	Complexe *M. Abscessus*	0.923	0.852	-	1	1	1	1	1
proportion des nœuds ayant un "bootstrap" > 0,95		0/3	2/3	1/3	2/3	2/3	2/3	3/3	1/3
nombre de substitutions par site		0.08	0.2	0.2	0.2	0.2	0.2	0.2	0.3

L'objectif fixé pour ce travail est de trouver une combinaison de loci qui donnera une phylogénie robuste des membres du complexe *M. tuberculosis* en premier lieu et des complexes *M. avium* et *M. tuberculosis* en deuxième lieu.

D'après ces résultats, les loci qui semblent être pertinents sont:

- **Le locus *hsp65*:** car il différencie 2 des 6 espèces du complexe MTBC, ce sont *M. africanum* et *M. canetti;*
- **Le locus *tuf*:** parce qu'il permet de distinguer *M. microti* et *M. canetti* des autres membres du complexe;

- **Le locus *gyrB*:** Les 8 SNP qui se trouvent dans ce gène permettent de différencier entre tous les membres du complexe MTBC. Il doit être donc inclus dans n'importe quelle combinaison suggérée.
- **Le locus *ARNr 16S*:** même s'il ne différencie aucun membre du complexe MTBC, ce gène est choisi car qu'il est largement utilisé en phylogénie [13].

3. Variabilité intraspécifique:

La variabilité intraspécifique est une notion très importante en analyse phylogénétique. En effet, un locus qui présente une grande variabilité intraspécifique ne doit pas être utilisé, car cela peut conduire à la différenciation entre des souches appartenant à la même espèce.

Ainsi, la variabilité intraspécifique est étudiée pour 3 loci jugés être intéressants dans ce travail (*hsp65*, *tuf* et *gyrB*). Seuls les membres du complexe MTBC sont inclus. Le gène *d'ARNr 16S* n'est pas concerné par cette analyse parce qu'il très conservé [13, 21] et même la variabilité interspécifique est nulle entre les espèces du MTBC.

Les résultats sont présentés sous forme de tableaux. Pour ne pas encombrer le texte, 2 tableaux sont inclus avec le texte (tableaux 13 et 14) tandis qu'un troisième est rattaché aux annexes (tableau 20). Ces tableaux apportent des informations aussi bien sur la variabilité intraspécifique que sur la variabilité interspécifique.

*Tableau 13: Matrice montrant la variation intraspécifique et interspécifique du gène **hsp65** mesurées chez 26 souches du complexe MTBC; les valeurs du tableau indiquent le % d'identité calculé par*

	N°	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
M. tuberculosis ATCC 27294T	1																									
M.tuberculosis Erdman	2	100																								
M.tuberculosis KZN 4207	3	100	100																							
M.tuberculosis KIT 77710	4	100	100	100																						
M.tuberculosis CCDC5180	5	100	100	100	100																					
M.tuberculosis KIT 77009	6	100	100	100	100	100																				
M.tuberculosis RGTB423	7	100	100	100	100	100	100																			
M.tuberculosis CTRI-2	8	100	100	100	100	100	100	100																		
M.tuberculosis CCDC5079	9	100	100	100	100	100	100	100	100																	
M.tuberculosis RGTB327	10	100	100	100	100	100	100	100	100	100																
M.tuberculosis CDC1551	11	100	100	100	100	100	100	100	100	100	100															
M.tuberculosis KZN 1435	12	100	100	100	100	100	100	100	100	100	100	100														
M.tuberculosis F11	13	100	100	100	100	100	100	100	100	100	100	100	100													
M.tuberculosis H37Ra	14	100	100	100	100	100	100	100	100	100	100	100	100	100												
M bovis ATCC 19210T	15	100	100	100	100	100	100	100	100	100	100	100	100	100	100											
M bovis CIP 105050	16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100										
M bovis AF2122/97	17	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100									
M africanum ATCC 25420T	18	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73								
M africanum CIP 105147	19	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	100							
M.africanum2 KIT 77710	20	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99.73	99.73						
M microti CIP 104256T	21	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99.73	99.73	100					
M microti ATCC 19422	22	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99.73	99.73	100	100				
M caprae CIP 105776T	23	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99.73	99.73	100	100	100			
M canetti CIPT 140060007	24	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.46	99.46	99.73	99.73	99.73	99.73		
M canetti CIPT 140060001	25	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.46	99.46	99.73	99.73	99.73	99.73	100	
M canetti CIPT 140060005	26	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.73	99.46	99.46	99.73	99.73	99.73	99.73	100	100

BLAST.

*Tableau 14: Matrice montrant la variation intraspécifique et interspécifique du gène **gyrB** mesurées*

	N°	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
M. tuberculosis T021T	1																										
M. tuberculosis KPMKT590	2	100																									
M. tuberculosis KPMKT031	3	100	100																								
M. tuberculosis KPMKT643	4	100	100	100																							
M. tuberculosis KPMKT673	5	99.89	99.89	99.89	99.89																						
M. tuberculosis KPMKT677	6	100	100	100	100	99.89																					
M. tuberculosis KPMKT741	7	100	100	100	100	99.89	100																				
M. tuberculosis KPMKT713	8	100	100	100	100	99.89	100	100																			
M. tuberculosis KPMKT699	9	100	100	100	100	99.89	100	100	100																		
M. tuberculosis KPMKT697	10	100	100	100	100	99.89	100	100	100	100																	
M. tuberculosis H37Rv	11	100	100	100	100	99.89	100	100	100	100	100																
M. tuberculosis CTRI-2	12	100	100	100	100	99.89	100	100	100	100	100	100															
M. tuberculosis RGTB423	13	99.89	99.89	99.89	99.89	99.78	99.89	99.89	99.89	99.89	99.89	99.89	99.89														
M. tuberculosis CCDC5079	14	100	100	100	100	99.89	100	100	100	100	100	100	100	99													
M. tuberculosis RGTB327	15	100	100	100	100	99.89	100	100	100	100	100	100	100	99	100												
M. bovis T704T	16	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.57	99.68	99.68											
M. bovis IKEDA	17	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.57	99.68	99.68	100										
M. bovis T702	18	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.57	99.68	99.68	100	100									
M. bovis AF2122/97	19	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.57	99.68	99.68	100	100	100								
M. africanum T801T	20	99.89	99.89	99.89	99.89	99.78	99.89	99.89	99.89	99.89	99.89	99.89	99.89	99.89	99.89	99.89	99.78	99.78	99.78	99.78	100						
M. africanum GM041182	21	99.89	99.89	99.89	99.89	99.78	99.89	99.89	99.89	99.89	99.89	99.89	99.89	99.89	99.89	99.89	99.78	99.78	99.78	99.78	99.78	99.78					
M. caprae 5115/98	22	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.78	99.78	99.78	99.78	99.89	99.89	99.68				
M. microti NCTC 8710T	23	99.78	99.78	99.78	99.78	99.68	99.78	99.78	99.78	99.78	99.78	99.78	99.78	99.68	99.78	99.78	99.68	99.68	99.68	99.68	99.89	99.89		99.47			
M. canettii CIPT140010059	24	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.68	99.78	99.78	99.78	99.57	99.57	99.36	99.04	99.15		
M. canetti2 CIPT140060005	25	99.25	99.25	99.25	99.25	99.15	99.25	99.25	99.25	99.25	99.25	99.25	99.25	99.15	99.25	99.25	98.94	98.94	98.94	98.94	99.15	99.15	98.94	99.04	99.15	100	
M. canetti3 CIPT140060007	26	99.25	99.25	99.25	99.25	99.15	99.25	99.25	99.25	99.25	99.25	99.25	99.25	99.15	99.25	99.25	98.94	98.94	98.94	98.94	99.15	99.15	98.94	99.04	99.15	100	
M. canetti4 CIPT140060001	27	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.57	99.68	99.68	99.36	99.36	99.36	99.36	99.57	99.57	99.36	99.47	100	99.15	99.15

chez 27 membres du complexe MTBC; les valeurs du tableau indiquent le % d'identité calculé par BLAST.

Les résultats présentés dans les tableaux 13 et 14 ci-dessus et le tableau 20 (annexes) montrent que la variabilité intraspécifique de l'ensemble des gènes (*hsp65*, *tuf* et *gyrB*) est très faible; en effet le pourcentage de variabilité dans tous les cas est toujours inférieur à 0.85%. Ce sont donc des bonnes cibles moléculaires pour faire la phylogénie des *Mycobactéries.*

4. La phylogénie avec une combinaison de gène:

A partir des 4 loci choisis, les 5 combinaisons suivantes ont été effectuées:

- Une première combinaison avec *l'ARNr 16S* et *gyrB*;
- Une deuxième combinaison avec *hsp65* et *gyrB*;
- Une troisième combinaison avec *tuf* et *gyrB*;
- Une quatrième combinaison avec *hsp65*, *tuf* et *gyrB*;
- Une cinquième combinaison avec *l'ARNr 16S*, *hsp65* et *gyrB.*

Par la suite 5 arbres phylogénétiques ont été obtenus chacun à partir de l'une de ces 5 combinaisons. Pour ne pas alourdir le texte, 2 de ces 5 arbres seulement sont intégrés avec le texte, ce sont ceux obtenus à partir de la combinaison "*hsp65* et *gyrB*" (Fig.27) et à partir de la combinaison "*hsp65*, *tuf* et *gyrB*" (Fig.28), les 3 arbres qui restent sont attachés aux annexes.

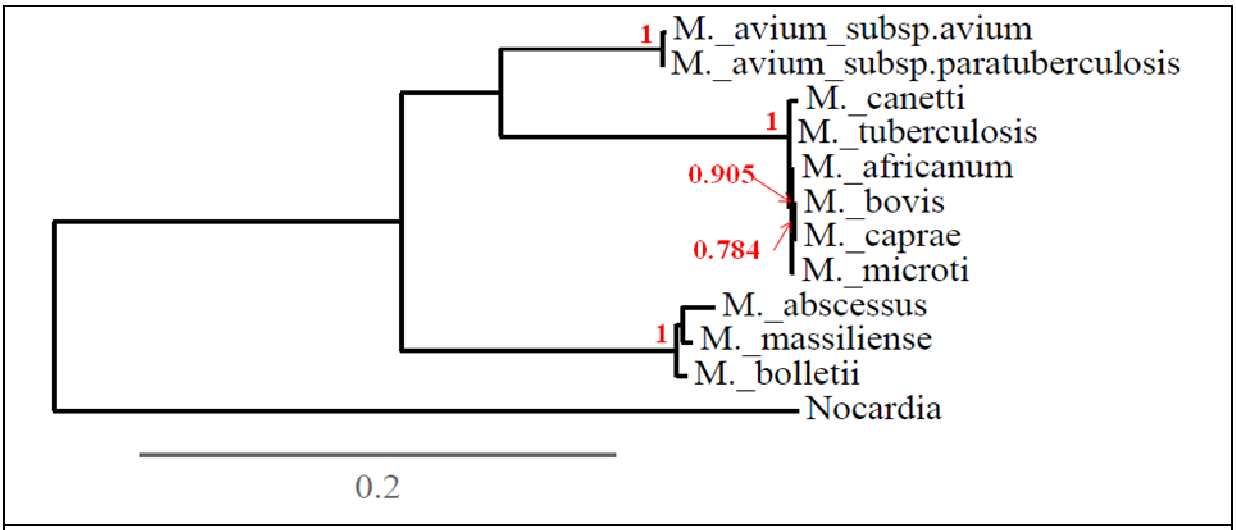

Figure 27: Arbre calculé à partir de la combinaison (hsp65 + gyrB). Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. Le nombre de substitutions par site est égal à 0.2.

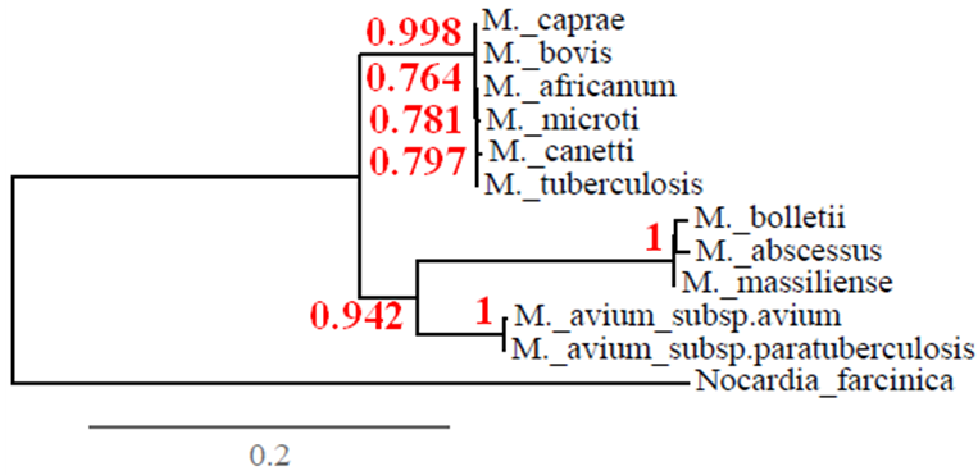

Figure 28: Arbre obtenu à partir de la combinaison (hsp65 + tuf + gyrB). Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. Le nombre de substitutions par site est égal à 0.2.

Le tableau 15 ci-dessous propose une comparaison entre les résultats obtenus à partir de l'analyse individuelle et combinée de ces gènes.

Tableau 15: Comparaison des principaux résultats obtenus avec l'analyse individuelle et combinée des 4 loci ARNr 16S, hsp65, tuf et gyrB. Le signe + indique que la différenciation a eu lieu alors que le signe – indique le contraire.

	gènes individuels				**gènes conbinés**				
	ARNr 16S	**hsp65**	**tuf**	**gyrB**	**ARNr 16S et gyrB**	**hsp65 et gyrB**	**Tuf et gyrB**	**hsp65, tuf et gyrB**	**ARNr 16S, hsp65 et yrB**
Taille de la séquence en pb	525-541	372	629-632	943	1468-1484	1315	1572	2201-2207	1842-1856
Différenciation entre les membres du complexe MTBC	-	-	-	+	+	+	+	+	+
Différenciation entre les membres du complexe MAC	-	-	-	-	+	+	+	+	+
Différenciation entre les membres du complexe *M. abscessus*	+	+	+	+	+	+	+	+	+
Nombre de substitutions par site	0.08	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.1
nombre totale des nœuds ayant une valeur de "bootstrap" >0.95	1	2	2	3	4	4	3	4	4

nœuds avec "bootstrap">0.75 à l'intérieur du MTBC	0	0	0	2	2	2	3	3	2

On constate généralement que:

- La robustesse et fiabilité des arbres obtenus à partir des gènes concaténés sont supérieures à celles des arbres inférés à partir des gènes pris individuellement car les valeurs de "bootstrap" et le nombre des complexes différenciables augmentent;
- Toutes les combinaisons permettent de différencier entre tous les membres des 3 complexes;
- Les membres du complexe *M. abscessus* sont facilement différenciables par rapport aux membres des 2 autres complexes (MTBC et MAC);
- Les membres du complexe MTBC sont mieux différenciables avec une combinaison de 2 ou de 3 gènes;
- La meilleure combinaison est celle qui comprend les gènes *gyrB, hsp65* et *tuf*.

Enfin, Les combinaisons qui comprennent 2 ou 3 semblent très intéressantes, non pas seulement pour faire, la phylogénie de ces 3 complexes mycobactériens mais aussi pour leur identification.
Actuellement, il n'existe pas un kit commercial qui pourrait différencier entre les membres du complexe *M. tuberculosis*. Ainsi, Un kit basé sur la meilleure combinaison détectée (*hsp65, tuf et gyrB,*) s'avère capable de remplir efficacement cette fonction.

CONCLUSION

Le genre *Mycobacterium* comprend des espèces phylogénétiquement très apparentées, Plusieurs d'entre elles sont pathogènes pour l'Homme ou pour l'Animal, Parmi les mycobactéries qui posent des problèmes aussi bien en santé publique qu'en phylogénie, on trouve le complexe *M. tuberculosis*, le complexe *M. abscessus* et le complexe *M. avium.*

Parmi les trois complexes étudiés, c'est surtout le complexe *M. tuberculosis* qui embarrasse les phylogénéticiens car il regroupe des espèces ayant une variabilité interspécifique faible, ainsi la majorité des loci n'arrive pas à séparer entre ses membres. L'aptitude de 8 loci (*ARNr 16S, hsp65, rpoB, tuf, sodA, ARNtm, gyrB* et *secA1*) à générer une phylogénie fiable a été évaluée pour ce complexe ainsi que pour les deux autres.

Les résultats montrent que quatre (*ARNr 16S, hsp65, tuf* et *gyrB*) de ces huit loci donnent des phylogénies robustes et résolutives si on en fait une combinaison de deux ou de trois. Il semble très intéressant d'inclure la meilleure combinaison détectée (*gyrB, hsp65 et tuf*) dans un kit commercial qui sera destiné à l'identification des espèces Mycobactériennes très proches comme celles qui ont fait l'objet de ce travail.

PERSPECTIVES

Dans une future étude, Il est souhaitable de tester la fiabilité d'autres loci autres que ceux inclus dans cette étude, ainsi que d'augmenter le nombre de souches étudiées.

REFERENCES BIBLIOGRAPHIQUES

1) **Adekambi T., and al.** 2004. Amoebal coculture of "Mycobacterium massiliense" sp. nov. from the sputum of a patient with hemoptoic pneumonia. J. Clin. Microbiol. **42:**5493–5501.
2) **Adekambi T., and M. Drancourt.** 2004. 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. 54:2095–2105.
3) **Blackwood K. S., C. He, J. Gunton, C. Y. Turenne, J. Wolfe, and A. M. Kabani**. 2000. Evaluation of recA sequences for identification of Mycobacterium species. J. Clin. Microbiol. 38:2846–2852.
4) **Castresana J.,** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution. 17: 540-552.
5) **Cattoir V.,** 2004. Identification moléculaire des mycobactéries et détection de la résistance aux antibiotiques, Ann Biol Cli. 62: 405-13.
6) **Charles L., Carbone, I., Davies, K. G., Bird, D., Burke, M., Kerry, B. R. & Opperman C. H.** 2005. Phylogenetic analysis of Pasteuria penetrans by use of multiple genetic loci. J Bacteriol 187, 5700–5708.
7) **Coutouly G., E. Klein, E. Barbieri and M. Kriat**. 2006. Travaux dirigés de biochimie, biologie moléculaire et bioinformatique. Chapitre 3. Groupe liaison SA.
8) **Dai J., Y. Chen and M. Lauzardo.** 2011. Web-Accessible Database of hsp65 Sequences from Mycobacterium Reference Strains. J. Clin. Microbiol. 49: 2296-2303.
9) **Dai J., Y. Chen, S. Dean, J. G. Morris, M. Salfinger and J. A. Johnson.** 2011. Multiple-Genome Comparison Reveals New Loci for Mycobacterium Species Identification. J. Clin. Microbiol. 49: 144-153.
10) **Dauendorffer J. N., I. Guillemin, A. Aubry, C. T. Pernot and others.** 2003. Identification of Mycobacterial Species by PCR Sequencing of Quinolone Resistance Determining Regions of DNA Gyrase Genes. J. Clin. Microbiol. 4: 1311-1315.
11) **Delsuc F., Et E. J. P. Douzery.** 2004. Les méthodes probabilistes en phylogénie moléculaire; Les modèles d'évolution des séquences et le maximum de vraisemblance. Biosystema. 22:59-74.
12) **Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., DufayardJ.F., Guindon S., Lefort V., Lescot M., Claverie J.M., Gascuel O.** 2008. robust phylogenetic analysis for the non-specialist.Nucleic Acids Res. 36: 465-469**.**

13) **Devulder G., M. Perouse de Montclos, and J. P. Flandrois.** 2005. A multigene approach to phylogenetic analysis using the genus Mycobacterium as a model. Int. J. Syst. Evol. Microbiol. 55:293–302.

14) **Durbin R., S. Eddy, A. Krogh and G. Mitchison.** 2002. Biological sequence analysis, Probabilistic models of proteins and nucleic acids. Chapitre 6 et 7. Cambridge university press.

15) **Edgar Robert C.,** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32(5), 1792-97.

16) **Gingeras T. R., G. Ghandour, E. Wang, A. Berno, P. M. Small, F. Drobniewski, D. Alland, E. Desmond, M. Holodniy, and J. Drenkow.** 1998. Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic Mycobacterium DNA arrays. Genome Res. 8:435–448.

17) Griffith D. E. and al., 2007. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. Am. J. Respir. Crit. Care Med. 175:367–416.

18) **Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O.,** 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology, 59(3):307-21.

19) **Kasai H., T. Ezaki, and S. Harayama**. 2000. Differentiation of phylogenetically related slowly growing mycobacteria by their gyrB sequences. J. Clin. Microbiol. 38:301–308.

20) **Kim B. J., S. H. Lee, M. A. Lyu, S. J. Kim, G. H. Bai, G. T. Chae, E. C. Kim, C. Y. Cha, and Y. H. Kook. 1999.** Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (rpoB). J. Clin. Microbiol. 37:1714–1720.

21) **Kirschner P. and E. C. Bottger**. 1998. Species identification of mycobacteria using rDNA sequencing. Methods Mol. Biol. 101:349–361.

22) **Macheras E., A. Roux, F. Ripoll, V. S. Tardy, C. Gutierrez, J. L. Gaillard and B. Heym**. 2009. Inaccuracy of Single-Target Sequencing for Discriminating Species of the Mycobacterium abscessus Group. J. Clin. Microbiol. 4: 2596-2600.

23) **Marmiesse M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. Cole and R. Brosch.** 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex. Microbiology. 150: 483–496.

24) **McNabb A., D. Eisler, K. Adie, M. Amos, M. Rodrigues, G. Stephens, W. A. Black, and J. Isaac-Renton.** 2004. Assessment of partial sequencing of the 65-kilodalton heat shock protein gene (hsp65) for routine identification of Mycobacterium species isolated from clinical sources. J. Clin. Microbiol. 42:3000–3011.

25) **Mignard S., and J. P. Flandrois**. 2007. Identification of Mycobacterium using the EF-Tu encoding (tuf) gene and the tmRNA encoding (ssrA) gene. J. Med. Microbiol. 56:1033–1041.

26) **Mignard S., and J. P. Flandrois.** 2008. A seven-gene, multilocus, genus-wide approach to the phylogeny of mycobacteria using supertrees. Int. J. Syst. Evol. Microbiol. 58:1432–1441.

27) **Narayanan P.,** 2005. Bioinformatics a primer. Chapitre 12. New age international publishers.

28) **Niemann S., D. Harmsen, S. R. Gerdes And E. Richter**. 2000. Differentiation of Clinical Mycobacterium tuberculosis Complex Isolates by gyrB DNA Sequence Polymorphism Analysis. J. Clin. Microbiol. 38: 3231-3234.

29) Ninet B., M. Monod, S. Emler, J. Pawlowski, C. Metral, P. Rohner, R. Auckenthaler, and B. Hirschel. 1996. Two different 16S rRNA genes in a mycobacterial strain. J. Clin. Microbiol. 34:2531–2536.

30) **Ranwez V.,** 2002. Méthodes efficaces pour reconstruire de grandes phylogénies suivant le principe du maximum de vraisemblance. Chapitre 1 et 2.

31) **Reischl U., K. Feldmann, L. Naumann, B. J. Gaugler, B. Ninet, B. Hirschel, and S. Emler.** 1998. 16S rRNA sequence diversity in Mycobacterium celatum strains caused by presence of two different copies of 16S rRNA gene. J. Clin. Microbiol. 36:1761–1764.

32) **Ringuet, H., Akoua-Koffi, C., Honore, S., Varnerot, A., Vincent, V., Berche, P., Gaillard, J. L. & Pierre-Audigier, C.,** 1999. hsp65 sequencing for identification of rapidly growing mycobacteria. J., Clin Microbiol. 37: 852–857.

33) Ringuet, H., C. Akoua-Koffi, S. Honore, A. Varnerot, V. Vincent, P. Berche, J. L. Gaillard, and C. Pierre-Audigier. 1999. hsp65 sequencing for identification of rapidly growing mycobacteria. J. Clin. Microbiol. 37:852–857.

34) **Rogall T., T. Flohr, and E. C. Bottger.** 1990. Differentiation of Mycobacterium species by direct sequencing of amplified DNA. J. Gen. Microbiol. 136:1915–1920.

35) Roth A., M. Fischer, M. E. Hamid and others. 1998. Differentiation of Phylogenetically Related Slowly Growing Mycobacteria Based on 16S-23S rRNA Gene Internal Transcribed Spacer Sequences . J. Clin. Microbiol. 36: 139-147.

36) **Roth A., M. Fischer, M. E. Hamid, S. Michalke, W. Ludwig, and H. Mauch.** 1998. Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences. J. Clin. Microbiol. 36:139–147.

37) **Saitou N. and M. Nei.**1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. Mol. Biol. Evol.4: 406-425.

38) **Sansila R., P. Hongmanee and others.** 1998. Differentiation between Mycobacterium tuberculosis and Mycobacterium avium by Amplification of the 16S-23S Ribosomal DNA Spacer. J. Clin. Microbiol. 36: 2399-2403.

39) **Shin J. Hyun, E. J. Cho, J. Y. Lee and others.** 2009. Novel Diagnostic Algorithm Using tuf Gene Amplification and Restriction Fragment Length Polymorphism is Promising Tool for Identification of Nontuberculous Mycobacteria. J. Microbiol. Biotechnol. 19: 323-330.

40) **Skrypnyk A.V.,** 2011. Molecular Phylogeny of Representatives of the Genus Mycobacterium Based on Structural Analysis of the Hypervariable Region A of the 16S Ribosomal RNA Gene. Cytology and Genetics. 44: 134-139.

41) **Soini H., E. C. Bottger, and M. K. Viljanen.** 1994. Identification of mycobacteria by PCR-based sequence determination of the 32-kilodalton protein gene. J. Clin. Microbiol. 32:2944–2947.

42) **Stackebrandt E., Frederiksen, W., Garrity, G.M., Grimont, P. A., Ka¨ mpfer, P., Maiden, M. C., Nesme, X., Rossello´ -Mora, R., Swings, J. & other authors** .2002. Report of the ad hoc committee for the reevaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 52, 1043–1047.

43) **Stadthagen-Gomez, G., A. C. Helguera-Repetto, J. F. Cerna-Cortes, R. A. Goldstein, R. A. Cox, and J. A. Gonzalez-y-Merchand.** 2008. The organization of two rRNA (rrn) operons of the slow-growing pathogen Mycobacterium celatum provides key insights into mycobacterial evolution. FEMS Microbiol. Lett. 280:102–112.

44) **Takewaki S., K. Okuzumi, H. Ishiko, K. Nakahara, A. Ohkubo, and R. Nagai**. 1993. Genus-specific polymerase chain reaction for the mycobacterial dnaJ gene and species specific oligonucleotide probes. J. Clin. Microbiol. 31:446–450.

45) **Telenti A., F. Marchesi, M. Balz, F. Bally, E. C. Bottger, and T. Bodmer.** 1993. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. J. Clin. Microbiol. 31:175–178.

46) **Tortoli E.** 2003. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. Clin Microbiol Rev. 16: 319-54.

47) **Troesch A., H. Nguyen, C. G. Miyada, S. Desvarenne, T. R. Gingeras, P. M. Kaplan, P. Cros, and C. Mabilat**. 1999. *Mycobacterium* species identification and rifampin resistance testing with high-density DNA probe arrays. J. Clin. Microbiol. 37:49–55.

48) **Tsukamura M., I. Yano, and T. Imaeda.** 1986. Mycobacterium fortuitum subspecies acetamidolyticum, a new subspecies of Mycobacterium fortuitum. Microbiol. Immunol. 30:97–110.

49) **Turenne C. Y., L. Tschetter, J. Wolfe, and A. Kabani.** 2001. Necessity of quality-controlled 16S rRNA gene sequence databases: identifying nontuberculous Mycobacterium species. J. Clin. Microbiol. 39:3637–3648.

50) **Yamada-Noda, M., K. Ohkusu, H. Hata, M. M. Shah, P. H. Nhung, X. S. Sun, M. Hayashi, and T. Ezaki.** 2007. Mycobacterium species identification a new approach via dnaJ gene sequencing. Syst. Appl. Microbiol. 30:453–462.

51) Zeigler D. R., 2003. Gene sequences useful for predicting relatedness of whole genomes in bacteria. Int J Syst Evol Microbiol 53, 1893–1900.

52) **Zelazny A. M., J. M. Root , Y. R. Shea and others. 2009.** Cohort Study of Molecular Identification and Typing of Mycobacterium abscessus, Mycobacterium massiliense, and Mycobacterium bolleti. J. Clin. Microbiol. 4: 1985-1995.

53) **Zelazny A. M., L. B. Calhoun, L. Li, Y. R. Shea, and S. H. Fischer.** 2005. Identification of Mycobacterium species by secA1 sequences. J. Clin. Microbiol. 43:1051–1058.

54) **Zolg J. W., and S. Philippi-Schulz.** 1994. The superoxide dismutase gene, a target for detection and identification of mycobacteria by PCR. J. Clin. Microbiol. 32:2801–2812.

55) http://www.infobiogen.fr/doc/tutoriel/PHYLO/phylogenie.html.

UNIVERSITE SIDI MOHAMED BEN ABDELLAH
FACULTE DES SCIENCES ET TECHNIQUES DE FES
Département de Biologie

ANNEXES

Tableau 16: Pour les 9 espèces, ce tableau montre les souches et les numéros d'accession. T: souche type.

UNIVERSITE SIDI MOHAMED BEN ABDELLAH
FACULTE DES SCIENCES ET TECHNIQUES DE FES
Département de Biologie

Espèces	Souches utilisées avec les numéros d'accession pour chaque gène							
	16S RNA		***hsp65***		***rpoB***		***tuf***	
	Souche	**N° d'accession**	**Souche**	**N° d'accession**	**Souche**	**N° d'accession**	**Souche**	**N° d'accession**
M. tuberculosis	CIP 64.31 T	AF547977	ATCC 27294 T	JF491311	CIP 64.31 T	AY544974	CIP 64.31 T	DQ986023
M. bovis	CIP 105234 T	AF547903	ATCC 19210 T	JF491332	CIP 105234 T	AY544894	CIP 105234 T	DQ986022
M. africanum	CIP 105147 T	AF547893	ATCC 25420 T	JF491313	CIP 105147 T	AY544880	CIP 105147 T	DQ986020
M. caprae	CIP 105776 T	AF547975	CIP 105776 T	AF547884	CIP 105776 T	AY544972	CIP 105776 T	DQ986024
M. microti	CIP 104256 T	AF547947	CIP 104256 T	AF547856	CIP 104256 T	AY544944	CIP 104256 T	DQ986019
M. canetti	CIPT 140010059-	NC_015848	CIPT 140060007	AJ749924	CIPT 140010059	NC_015848	CIPT 140010059	NC_015848
M. avium subsp. avium	CIP 104244 T	AF547898	ATCC 25291 T	JF491291	CIP 104244 T	AY544887	CIP 104244 T	DQ986050
M. avium subsp. paratuberculosis	CIP 103963 T	AF547899	CIP 103963 T	AF547809	CIP 103963 T	AY544888	CIP 103963 T	DQ986049
M. avium subsp. silvaticum	CIP 103317 T	AF547900	CIP 103317 T	AF547810	CIP 103317 T	AY544889	CIP 103317 T	DQ986051
M. abscessus	CIP 104536 T	AF547892	ATCC 19977 T	JF491290	CIP 104536 T	AY544879	CIP 104536 T	DQ986060
M. bolletii	CCUG 50184 T	DQ987740	CIP 108541 T	AY859675	CCUG 50184 T	DQ987717	CCUG 50184 T	DQ986061
M. massiliense	JCM 15300 T	AB548602	CCUG 48898 T	AY596465	CIP 108297 T	EU191921	CIP 108297 T	EU191941

Suite du tableau 16; le signe – indique que la séquence est absente

UNIVERSITE SIDI MOHAMED BEN ABDELLAH
FACULTE DES SCIENCES ET TECHNIQUES DE FES
Département de Biologie

Espèces	Souches utilisées avec les N° d'accession pour chaque gène							
	sodA		**tmRNA**		**gyrB**		**secA1**	
	souche	**N° d'accession**	**souche**	**N° d'accession**	**souche**	**N° d'accession**	**souche**	**N° d'accession**
M. tuberculosis	CIP 64.31 T	AY544875	CIP 64.31 T	DQ480269	ATCC 27294 T	AB014194	ATCC 27294 T	AY724732
M. bovis	CIP 105234 T	AY544809	CIP 105234 T	DQ480270	ATCC 19210 T	AB014184	ATCC 19210 T	AY724705
M. africanum	CIP 105147 T	AY544796	CIP 105147 T	DQ480267	ATCC 25420 T	AB014192	ATCC25420 T	AY724702
M. caprae	CIP 105776 T	AY544873	CIP 105776 T	DQ480271	5115/98-	AJ276122	-	-
M. microti	CIP 104256 T	AY544846	CIP 104256 T	DQ480268	NCTC 8710 T	AB014205	-	-
M. canetti	CIPT 140010059	NC_015848	-	-	CIPT 140010059	NC_015848	CIPT 140010059	NC_015848
M. avium subsp. avium	CIP 104244 T	AY544802	CIP 104244 T	DQ480276	ATCC 25291 -	NZ_ACFI01000001	ATCC 25291 T	AY724704
M. avium subsp. paratuberculosis	CIP 103963 T	AY544803	CIP 103963 T	DQ480277	Bovine 10	AB014317	ATCC 19698 T	FJ442856
M. avium subsp. silvaticum	CIP 103317 T	AY544804	CIP 103317 T	DQ480275	-	-	ATCC 49884 T	FJ442857
M. abscessus	CIP 104536 T	AY544795	CIP 104536 T	DQ480289	KPM 4410-	AB014186	ATCC 19977 T	AY724701
M. bolletii	CCUG 50184 T	DQ987732	CCUG 50184 T	DQ985998	CCUG 50184	GU143882	CCUG 50184 T	FJ442839
M. massiliense	CCUG 48898 T	AY593975	CIP 108297 T	EU191937	CCUG 48898	GU143885	CCUG 48898 T	FJ442844

*Tableau 17: Souches et numéros d'accession utilisés pour étudier la variabilité intraspécifique du gène **hsp65**. T: souche type.*

Espèce	***hsp65***	
	souche	N° d'accession
M. tuberculosis	ATCC 27294 T	JF491311
M.tuberculosis	Erdman	AP012340
M.tuberculosis	KZN 4207	NC_016768
M.tuberculosis	KIT 77710	AY299175
M.tuberculosis	CCDC5180	NC_017522
M.tuberculosis	KIT 77009	AY299174
M.tuberculosis	RGTB423	NC_017528
M.tuberculosis	CTRI-2	NC_017524
M.tuberculosis	CCDC5079	NC_017523
M.tuberculosis	RGTB327	NC_017026
M.tuberculosis	CDC1551	NC_002755
M.tuberculosis	KZN 1435	NC_012943
M.tuberculosis	F11	NC_009565
M.tuberculosis	H37Ra	NC_009525
M. bovis	ATCC 19210 T	JF491332
M. bovis	CIP 105050	AF547814
M. bovis	AF2122/97	NC_002945
M. africanum	ATCC 25420 T	JF491313
M.africanum	CIP 105147	AF547803
M.africanum2	KIT 77710	AY299176
M. microti	CIP 104256 T	AF547856
M. microti	ATCC 19422	AY299135
M. caprae	CIP 105776 T	AF547884
M. canetti	CIPT 140060007	AJ749924
M. canetti	CIPT 140060001	AJ749922
M. canetti	CIPT 140060005	AJ749923

*Tableau 18: Souches et numéros d'accession utilisés pour étudier la variabilité intraspécifique du gène **gyrB**. T: souche type.*

Espèce	***gyrB***	
	Souche	N° d'accession
M. tuberculosis	T 021 T	AB014194
M. tuberculosis	KPM KY590	AB014209
M. tuberculosis	KPM KY631	AB014210
M. tuberculosis	KPM KY643	AB014211
M. tuberculosis	KPM KY673	AB014212
M. tuberculosis	KPM KY677	AB014213
M. tuberculosis	KPM KY741	AB014242
M. tuberculosis	KPM KY713	AB014231
M. tuberculosis	KPM KY699	AB014220
M. tuberculosis	KPM KY697	AB014218
M. tuberculosis	H37Rv	AL123456
M. tuberculosis	CTRI-2	NC_017524
M. tuberculosis	RGTB423	NC_017528
M. tuberculosis	CCDC5079	NC_017523
M. tuberculosis	RGTB327	NC_017026
M. tuberculosis	CDC1551	NC_002755
M. tuberculosis	KZN 1435	NC_012943
M. tuberculosis	F11	NC_009565
M. tuberculosis	H37Ra	NC_009525
M. tuberculosis	UT205	HE608151
M. bovis	T 704 T	AB014184
M. bovis	IKEDA	AB018554
M. bovis	T702	AB014193
M. bovis	AF2122/97	NC_002945
M. africanum	T 801 T	AB014192
M. africanum	GM041182	NC_015758
M. caprae	5115/98-	AJ276122
M. microti	NCTC 8710 T	AB014205
M. canetti1	CIPT 140010059-	NC_015848
M.canetti2	CIPT 140060005	AJ749914
M. canetti3	CIPT 140060007	AJ749915

M. canetti4	CIPT 140060001	AJ749913

*Tableau 19: Souches et numéros d'accession utilisés pour étudier la variabilité intraspécifique du gène **tuf**. T: souche type.*

Espèce	***tuf***	
	Souche	N° d'accession
M. tuberculosis	CIP 64.31 T	DQ986023
M. tuberculosis	Erdmann	X63539
M. tuberculosis	KZN 4207	NC_016768
M. tuberculosis	RGTB423	NC_017528
M. tuberculosis	CCDC5079	NC_017523
M. tuberculosis	CTRI-2	NC_017524
M. tuberculosis	UT205	HE608151
M. tuberculosis	H37Rv	AL123456
M. tuberculosis	RGTB327	NC_017026
M. tuberculosis	CDC1551	NC_002755
M. tuberculosis	KZN 1435	NC_012943
M. tuberculosis	F11	NC_009565
M. tuberculosis	H37Ra	NC_009525
M. tuberculosis	CCDC5180	NC_017522
M. bovis	CIP 105234 T	DQ986022
M. bovis	AF2122/97	NC_002945
M. africanum	CIP 105147 T	DQ986020
M. africanum	GM041182	NC_015758
M. caprae	CIP 105776 T	DQ986024
M. microti	CIP 104256 T	DQ986019
M. canetti	CIPT 140010059-	NC_015848

*Tableau 20: Une matrice montrant la variation intraspécifique et interspécifique du gène **tuf** mesurées chez 21 membres du complexe MTBC;les valeurs du tableau indiquent le % d'identité calculé par BLAST.*

	N°	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
M. tuberculosis CIP 64.31 T	1																				
M. tuberculosis Erdmann	2	100																			
M. tuberculosis KZN 4207	3	100	100																		
M. tuberculosis RGTB423	4	100	100	100																	
M. tuberculosis CCDC5079	5	100	100	100	100																
M. tuberculosis CTRI-2	6	100	100	100	100	100															
M. tuberculosis UT205	7	100	100	100	100	100	100														
M. tuberculosis H37Rv	8	100	100	100	100	100	100	100													
M. tuberculosis RGTB327	9	100	100	100	100	100	100	100	100												
M. tuberculosis CDC1551	10	100	100	100	100	100	100	100	100	100											
M. tuberculosis KZN 1435	11	100	100	100	100	100	100	100	100	100	100										
M. tuberculosis F11	12	100	100	100	100	100	100	100	100	100	100	100									
M. tuberculosis H37Ra	13	100	100	100	100	100	100	100	100	100	100	100	100								
M. tuberculosis CCDC5180	14	100	100	100	100	100	100	100	100	100	100	100	100	100							
M. bovis CIP 105234 T	15	100	100	100	100	100	100	100	100	100	100	100	100	100	100						
M. bovis AF2122/97	16	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100					
M. africanum CIP 105147 T	17	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100				
M. africanum GM041182	18	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100			
M. caprae CIP 105776 T	19	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100		
M. microti CIP 104256 T	20	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	99.84	
M. canetti CIPT 140010059-	21	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.68	99.53

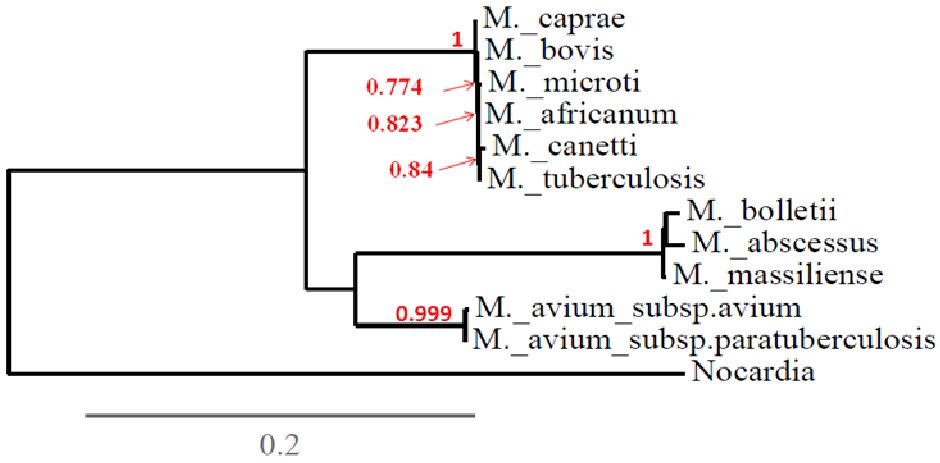

Figure 29: Arbre obtenu à partir de la combinaison (tuf + gyrB). Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. Le nombre de substitutions par site est égal à 0.2.

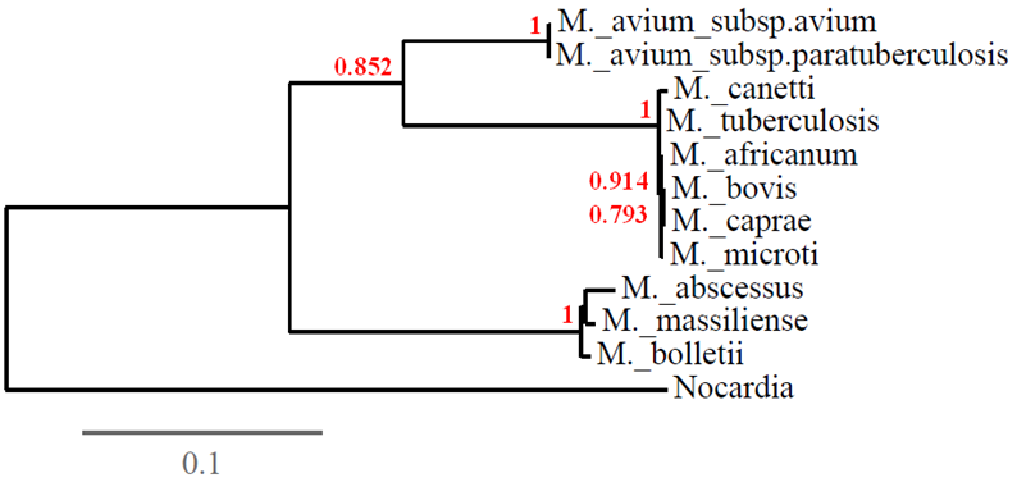

Figure 30: Arbre obtenu à partir de la combinaison (ARNr 16S + hsp65 + gyrB). Les numéros en gras sont les valeurs de "bootstrap", seules les valeurs supérieures à 0.75 sont indiquées. Le nombre de substitutions par site est égal à 0.1.