Etat de l'art du projet de recherche

Stephane Robin

15 janvier 2020

Table des matières

1. Introduction

Le présent document cherche à définir les fondements sur lesquels va reposer le stage et en particulier à connaître l'état actuel des connaissances concernant l'analyse bio-informatique du locus CRISPR-Cas chez *Mycobacterium tuberculosis*.

L'objectif du stage est de comprendre les liens entre les sept lignées de *M. tuberculosis* et les alternances de gènes et espaceurs du locus CRISPR-Cas, au moyen d'outils développés au sein de l'équipe AND de l'Université de Franche Comté.

2. Evolution de la branche humaine de la tuberculose

2.1. Diversité génétique de la tuberculose

Une bactérie survit d'autant mieux qu'elle est capable de s'adapter à son environnement au travers de mutations génétiques. Le polymorphisme génétique est à l'origine de la diversité génétique et correspond, dans le cas de notre étude, à des variations de séquences d'ADN entre différentes souches de *M. tuberculosis*. Ces variations sont dues à des mutations successives au cours de l'évolution de la bactérie, et elles permettent l'analyse phylogénique de *M. tuberculosis*. Il existe plusieurs formes de polymorphisme, le polymorphisme chromosomique lié à un changement du nombre de chromosomes ou de leurs structures, le polymorphisme d'insertion, de délétion et d'inversion qui provoquent un

changement spécifique de certaines séquences du génome, et le polymorphisme nucléotidique SNP *Single Nucleotide Polymorphism* lié au changement d'une seule paire de bases [1] du génome de *M. tuberculosis*. Nous allons détailler plus particulièrement ce dernier cas.

Certaines mutations n'ont aucun impact évolutif sur *M. tuberculosis*. En revanche, des changements fonctionnels peuvent avoir lieu lorsque ces mutations entraînent des modifications d'acides aminés dans les régions codantes, cela peut-être le cas lors d'une adaptation à l'environnement ou lors d'une nouvelle forme de résistance aux antibiotiques. Les SNPs synonymes ne changent pas la séquence de protéine, ainsi la substitution d'un codon [2] par un autre codon peut engendrer le même acide aminé. Au contraire, les SNPs non-synonymes changent la séquence de protéine, et engendrent donc l'incorporation d'un acide aminé différent. Chez *M. tuberculosis*, les SNPs sont peu sujets à des phénomènes d'homoplasie [3](seuls 1,1 % des SNPs sont homoplasiques), ce qui suggère que la structure de *M. tuberculosis* favorise les clonages plutôt que les recombinaisons entre branches. Pour de tels organismes clonaux, l'identification de mutations homoplasiques est un excellent moyen de déterminer les différentes souches bactériennes, et ainsi de procéder à des études phylogéniques [4] et de classification.

2.2. Co-évolution de la tuberculose avec l'homme moderne

Le développement des maladies s'adapte à la densité de population concernée. En effet, auprès d'une foule dense, les infections se répandent plus largement et deviennent plus virulentes, alors qu'auprès d'une population moins importante, elles ont une croissance plus faible, laissant parfois place à des périodes où les infections restent latentes.

Une période charnière dans l'histoire de l'humanité est la transition démographique du Néolithique, qui a vu il y a 10 000 ans, suite à l'apparition de l'agriculture et de l'élevage, un accroissement de la population, favorisant la naissance de nombreuses maladies. Les maladies humaines plus anciennes se développaient auprès de populations moins denses et produisaient des phases chroniques de latence et de réactivation permettant aux populations infectées de survivre.

Nous allons voir que la tuberculose conjugue ces deux modèles de maladie.

L'étude phylogénique de Comas et al.[1] se base exclusivement sur l'étude du génome [5] complet de toutes les lignées connues de *M. tuberculosis* en utilisant les SNPs comme marqueurs pour construire les relations entre les différentes branches. Les résultats obtenus rejoignent de précédentes études effectuées à partir d'autres marqueurs, et confirment l'existence de sept principales lignées de tuberculose. On remarque en particulier que plusieurs branches d'origine animale se sont regroupées avec la lignée 6 d'Afrique de l'Ouest, et que les lignées modernes 2 d'Asie de l'Est, 3 d'Asie Centrale et 4 d'Europe ont des origines proches.

1. **Paire de bases** : appariement de 2 bases nucléiques situées sur 2 brins complémentaires d'ADN, reliées par des ponts d'hydrogène.
2. **Codon** : ensemble composé de trois nucléotides consécutifs spécifiant l'incorporation d'un acide aminé déterminé. Le code génétique est ainsi lu trois nucléotides par trois nucléotides.
3. **Homoplasie** : similitude de caractères chez différentes espèces, qui ne provient pas d'un ancêtre commun, mais peut par exemple provenir d'une adaptation à l'environnement. Diffère de l'homologie qui est une similitude de caractères observée chez deux espèces différentes, provenant de l'héritage d'un ancêtre commun.
4. **Phylogénie** : étude des liens entre espèces apparentées, permettant de retracer les principales étapes de l'évolution des organismes depuis un ancêtre commun.
5. **Génome** : ensemble de l'information génétique d'un organisme. Par extension, le génome se réfère aussi au support physique de cette information génétique, la macromolécule d'ADN. L'annotation des gènes est le processus permettant d'identifier l'emplacement des gènes dans l'ADN, de déterminer leurs fonctions et leurs possibles interactions.

L'analyse phylogénique de Comas et al. corrobore la conjecture selon laquelle la tuberculose est originaire d'Afrique. Par ailleurs, s'appuyant sur les origines africaines de l'espèce humaine, cette étude cherche également à déterminer l'ancienneté de l'association entre la tuberculose et son hôte humain. Pour cela, l'analyse des divergences des génomes de la tuberculose est comparée à celle d'une arborescence génétique déjà établie à partir de mitochondries [6] de l'être humain.

Les similitudes relevées montrent que la tuberculose a infecté les premiers hommes d'Afrique. Pour aller plus loin, l'étude de Comas et al. a tenu compte de trois dates importantes dans l'évolution biologique de l'être humain qui ont été reportées sur l'analyse phylogénique de *M. tuberculosis* des lignées 5 et 6 d'Afrique de l'Ouest : l'émergence de l'homo sapiens correspondant au MTBC-185 [7], l'émergence de l'haplogroupe [8] mitochondrial de la lignée 3 chez l'homme correspondant au MTBC-70, et le début de la transition démographique du Néolithique correspondant au MTBC-10.

La branche MTBC-185 suggère l'apparition de mutations à partir de lignées africaines il y a 174 000 ans, c'est à dire que la dispersion de la tuberculose précèderait celle de l'homo sapiens.

La branche MTBC-70 révèle des corrélations avec l'histoire de l'humanité telle qu'elle a pu être décrite par l'archéologie, en montrant l'apparition des sept différentes lignées de tuberculose :

- il y a 73 000 ans, apparition des lignées 5 et 6 correspondant à une première migration humaine importante vers l'Afrique de l'Ouest,
- il y a 67 000 ans, apparition de la lignée 1 correspondant à une migration humaine importante autour de l'Océan Indien,
- il y a 64 000 ans, apparition de la lignée 7 concernant une population qui est restée en Afrique ou est revenue en Afrique après une première migration,
- il y a 46 000 ans, apparition de la lignée 4 correspondant à une migration humaine importante vers l'Europe,
- il y a 42 000 ans, apparition des lignées 2 et 3 correspondant à une migration humaine importante vers l'Asie de l'Est et l'Asie Centrale.

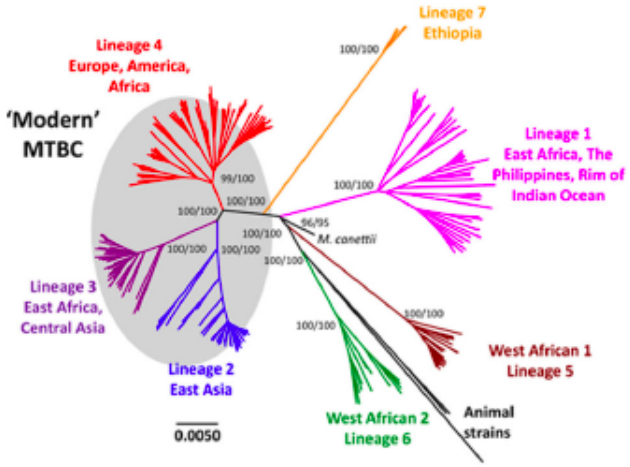


Figure 1 – Phylogénie du génome complet de MTBC, d'après *Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans*

Dans tous les cas, la tuberculose aurait infecté l'espèce humaine et évolué conjointement avec elle

6. **Mitochondrie** : centrale énergétique des cellules qui contribue à la production d'ATP.
7. **MTBC** : Mycobacterium Tuberculosis Complex.
8. **Haplogroupe** : groupe possédant les mêmes caractères génétiques et partageant un ancêtre commun suivant une mutation SNP.

depuis 70 000 ans, mais son apparition serait antérieure à la transition démographique du Néolithique.

La base de données de tuberculose étudiée de façon probabiliste par Comas et al.[1] montre que le Néolithique a fortement contribué à l'expansion de la maladie il y a 10 000 ans grâce à l'augmentation de la densité de population et à la probabilité de co-infection avec d'autres maladies également dépendantes de la densité de population. La possibilité pour la tuberculose de muter d'une variété animale vers une variété humaine n'est en revanche pas retenue par Comas et al. En effet, l'analyse phylogénique de la tuberculose montre que les branches humaines ont divergé des branches animales avant le Néolithique.

Le Néolithique n'était pas la seule période où l'augmentation de la population fut importante, toutefois la concentration de population qui s'en est suivie a permis l'apparition, auprès de la tuberculose, de caractères fortement dépendants de la densité de population qu'elle affecte. Le Néolithique a donc marqué un tournant dans l'histoire de la tuberculose, qui a alors commencé à conjuguer les deux principaux modèles de maladie, d'une part dépendant de la densité de population et d'autre part s'apparentant à une infection chronique. En effet, le mode de transmission aérosol de la tuberculose s'est parfaitement adapté aux foules, et elle a montré à travers les âges des périodes de latence et de réactivation.

Il faut donc considérer que la co-existence de la tuberculose avec l'espèce humaine depuis des milliers d'années a conduit la maladie à s'adapter aux changements du génome humain et inversement. Les prochaines études sur la tuberculose devraient donc se concentrer sur des génomes complets de la tuberculose et de l'être humain choisis en rapport à leurs associations.

En particulier, la tuberculose a dû s'adapter aux autres infections ayant touché l'espèce humaine, avec plus ou moins de succès. Dans cet ordre d'idée, une étude récente de Perry S. et al.[7, 8] suggère que l'infection d'un organisme par *Helicobacter Pylori* pourrait protéger de la tuberculose sous sa forme active. A contrario, nous ne savons pas si la tuberculose latente pourrait protéger contre les ulcères et les cancers de l'estomac causés par *Helicobacter Pylori*.

3. Le développement de souches résistantes aux antibiotiques

3.1. L'expansion de la lignée 4 de *M. tuberculosis*

La lignée 4 de *M. tuberculosis* est la plus répandue de par le monde et pour cette raison a fait l'objet de nombreuses publications. Brynildsrud O.B. et al.[3] utilisent des méthodes d'analyse discrète et une approche bayésienne [9] en phylogénie moléculaire pour obtenir de manière formelle l'évolution phylogéographique de la lignée 4 de *M. tuberculosis*. Ils estiment que le plus récent ancêtre commun de la lignée 4 est apparu en Europe en 1096 après JC. Si on considère l'Europe en tant que continent au sens large, cela ne contredit pas les résultats de O'Neil M.B. et al.[14] qui estiment l'origine de la lignée 4 autour de la méditerranée.

L'analyse phylogéographique de Brynildsrud O.B. suggère que les premières vagues de migration de la lignée 4 hors d'Europe se sont déroulées au début du 13ème siècle vers l'Asie du Sud-Est. Il est également possible d'établir une correspondance entre la structure des isolats [10] Vietnamiens et l'époque de l'expansion coloniale française en Indochine au 19ème siècle.

Les vagues suivantes de migration de la lignée 4 se sont dirigées vers l'Afrique de l'Ouest au 15ème siècle, puis vers Afrique de l'Est et du Sud au 17ème siècle. Les échanges continus avec le Portugal

9. **Approche bayésienne** : méthode probabiliste basée sur le calcul des probabilités postérieures des arbres phylogéniques par la combinaison d'une probabilité antérieure avec la fonction de vraisemblance

10. **Isolat** : fragment d'organisme qu'on a isolé à des fins d'examen histo-pathologiques ou pour être cultivé in vitro.

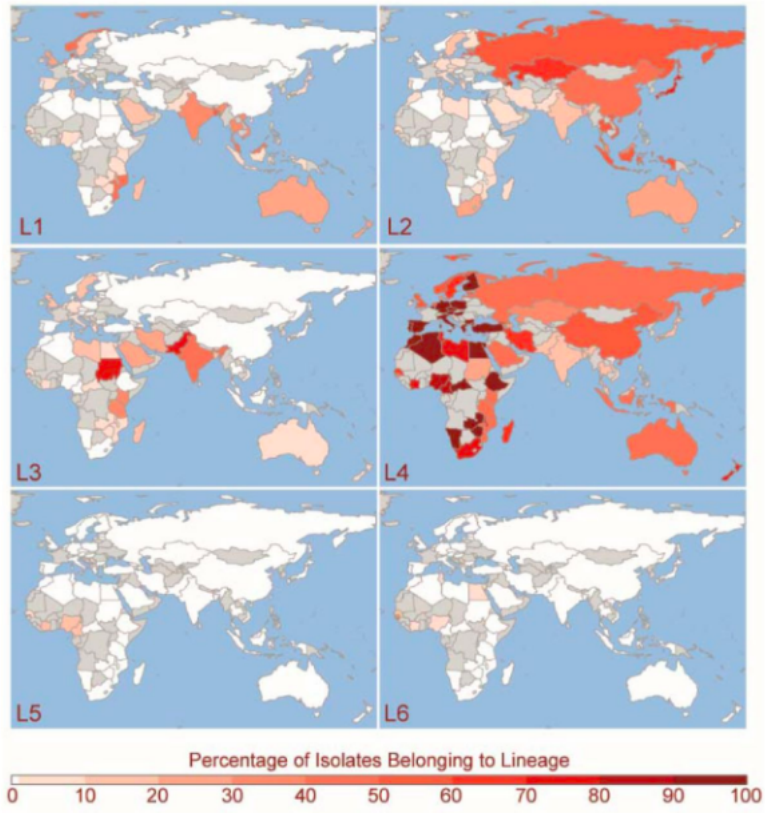


Figure 2 – Distribution géographique des lignées 1 à 6, d'après *Lineage specific histories of Mycobacterium tuberculosis dispersal in Africa and Eurasia*

dès le 15ème siècle ont favorisé la dispersion de la maladie, ce qui a été renforcé plus tard par la colonisation française de l'Afrique de l'Ouest. Ces échanges avec les populations européennes ont prévalu à une transmission locale de la tuberculose jusqu'au 19ème siècle.

La transmission de la maladie en Amérique date, elle aussi, du 15ème siècle avec la colonisation du continent, mais il faudra attendre le 17ème siècle pour voir l'explosion de la maladie en Amérique du Sud. Ce retard dans l'évolution de la maladie par rapport à la branche africaine peut s'expliquer par le taux de mortalité élevé des populations aborigènes au contact des européens.

La première migration interne de la maladie en Afrique date de l'Empire Zulu au 19ème siècle et se dirigeait vers le Nord et l'Est africain.

Ainsi, Brynildsrud O.B. et al. montrent que la dispersion de la lignée 4 est essentiellement liée à l'expansion coloniale européenne en Afrique et en Amérique entre le 17ème et le 19ème siècle.

3.2. L'adaptation de la lignée 4 pour devenir résistante aux antibiotiques

Nous avons déjà vu que la tuberculose a su s'adapter à l'évolution géographique de l'humanité en suivant les différentes migrations humaines pour créer de nouvelles lignées ou de nouvelles souches. Il apparaît que la tuberculose est également capable de suivre l'évolution médicale de l'humanité. L'étude de Brynildsrud O.B. et al.[3] constate chez *M. tuberculosis* l'émergence croissante d'une résistance à de multiples antibiotiques entre 1960 et 2000 au travers de la phylogénie de la lignée 4.

Des mutations spontanées dans le génome de la tuberculose peuvent altérer les protéines qui sont la cible des médicaments, ce qui rend les bactéries résistantes à ces médicaments. Prenons comme

exemple une mutation du gène rpoB de *M. tuberculosis*, qui code pour la sous-unité β[11] de l'ARN polymérase[12] de la bactérie. Dans la tuberculose non résistante, la rifampicine se lie à cette sous-unité β et perturbe l'élongation de la transcription de l'ARN. La mutation dans le gène rpoB modifie la séquence des acides aminés et donc de la sous-unité β. Dans ce cas, la rifampicine ne peut plus se lier à la sous-unité β de l'ARN et empêcher la transcription. La bactérie est devenue résistante. C'est bien le cas de la tuberculose, qui est considérée aujourd'hui comme une maladie résistante aux antibiotiques.

Une souche de *M. tuberculosis* est appelée MDR-TB *Multi-Drug-Resistant Tuberculosis* si elle est résistante au deux anti-tuberculeux de première intention les plus puissants, l'isoniazide et le rifampicine. Dans ce cas, certaines régions du génome de *M. tuberculosis* sont impliquées dans la résistance à plus d'un médicament. La découverte de nouvelles cibles moléculaires s'avère essentielle pour lutter contre ce développement de la résistance chez *M. tuberculosis*. Une souche de *M. tuberculosis* est appelée XDR-TB si elle est de surcroît résistante aux anti-tuberculeux de seconde intention tels que le fluoroquinolone et l'aminoglycoside.

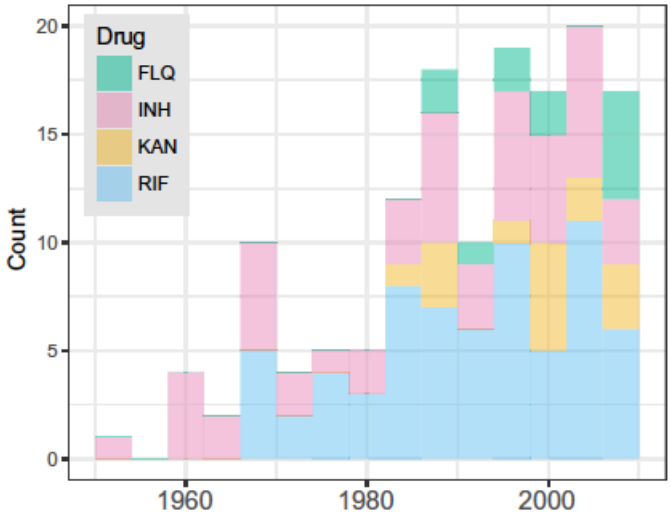


Figure 3 – Transmission de la résistance ces dernières années à l'échelle mondiale, d'après *Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation*. FLQ=fluoroquinolone, INH=isoniazide, KAN=aminoglycoside, RIF=rifampicine

Les causes de la résistance de *M. tuberculosis* aux antibiotiques sont multiples, mais il s'agit principalement de l'utilisation inappropriée ou incorrecte d'antibiotiques, et de l'interruption précoce des traitements. Dans ce cas, les souches résistantes se transmettent génétiquement de générations en générations. Toutefois, ces souches résistantes de *M. tuberculosis* peuvent aussi se transmettre directement à une personne saine, qui dans ce cas, se retrouve infectée avec une souche MDR-TB sans avoir pris de traitement inapproprié contre la tuberculose.

Brynildsrud O.B. et al. étudient également le gène lldD2 impliqué dans la réplication de *M. tuberculosis* au sein des macrophages[13] humains. Ils identifient au niveau des codons 3 et 253 la présence de nombreux promoteurs[14] et mutations non-synonymes qui ont évolué indépendamment.

Une recherche au sein d'une base de données recouvrant les lignées 1 à 6 a révélé que la mutation du codon 3 a émergé indépendamment dans les lignées 1, 2 et 4, alors que la mutation du codon 253 est

11. **Sous-unité** β : élément de l'ARN polymérase des bactéries qui est composé de la structure suivante $\alpha_2\beta\beta'\omega$
12. **ARN polymérase** : complexe enzymatique responsable de la synthèse de l'ARN à partir d'ADN.
13. **Macrophage** : cellule appartenant aux globules blancs qui infiltre les tissus et est capable de phagocytose.
14. **Promoteur** : région de l'ADN située à proximité d'un gène et indispensable à la transcription de l'ADN.

apparue à plusieurs reprises dans la lignée 4 et est présente dans pratiquement tous les isolats de la lignée 2. Brynildsrud O.B. et al. constatent que les mutations de lldD2 ont commencé à apparaître bien avant l'utilisation des antibiotiques sur tous les continents. Ceci suppose une adaptation locale de *M. tuberculosis* à de profonds changements chez l'hôte humain, qui s'est opérée en parallèle sur les différents continents. Par ailleurs, les souches hébergeant des mutations du promoteur lldD2 présentent un avantage significatif en terme de transmissibilité.

Il ne fait aucun doute que des souches MDR-TB peuvent traverser les frontières, comme cela a déjà été observé avec la lignée 2 entre l'Europe de l'Est et l'Europe de l'Ouest. Toutefois, le jeune âge relatif des souches résistantes pourrait expliquer le manque de migrations observées de ces souches. Brynildsrud O.B. et al. démontrent que, d'un point de vue mondial, la migration humaine a joué un rôle négligeable dans l'élaboration des modèles de résistance aux antimicrobiens. En effet, la migration des souches résistantes s'est avérée marginale. Il s'agit plutôt d'un phénomène local. La restriction géographique de souches résistantes suggère même de lutter contre ce type de mutation de *M. tuberculosis* de façon nationale plutôt que de recourir à une politique globale de traitement antibiotique.

4. Le locus CRISPR-Cas

4.1. Quelques caractéristiques du génome de *M. tuberculosis*

La souche H37Rv de *M. tuberculosis* est la souche de tuberculose la plus étudiée en laboratoire, depuis sa découverte en 1905. Elle sert aujourd'hui de référence pour le séquençage et l'annotation du génome de *M. tuberculosis*. Constitué d'environ 4 millions de paires de base et 3959 gènes, ce génome se caractérise par un taux élevé de guanine G et de cytosine C (65,6%), et un codon GTG qui sert de codon d'initiation dans 35% des gènes.

Parmi les marqueurs génétiques utilisés pour des études phylogéniques ou d'épidémiologie moléculaire, on retrouve les SNPs, les loci CRISPR, les MIRU [15], et les VNTR [16]. L'association des résultats obtenus par ces marqueurs génère un profil allélique [17] utile pour l'étude du complexe *M. tuberculosis*. La base de données mondiale de marqueurs moléculaires de la tuberculose SITVIT [18] présentée par Demay C. et al.[12] contient les génotypes de *M. tuberculosis* obtenus à partir des marqueurs moléculaires MIRU et VNTR.

4.2. Description du locus CRISPR-Cas

Le locus CRISPR *Clustered Regularly Interspaced Short Palindromic Repeats* est une famille de séquences répétées (DR pour *Direct Repeat*) dans l'ADN formant un palindrome, qui se trouve à l'état naturel chez 40% des bactéries (dont le *M. tuberculosis*) et la plupart des archées. CRISPR

15. **Marqueur MIRU Mycobacterial Interspaced Repetitive Units** : séquences nucléotidiques courtes répétitives en tandem entrecoupées de mycobactéries. La méthode MIRU actuellement utilisée sur *M. tuberculosis* est composée de 12 loci MIRU différents. Un mirutype est un modèle à 12 chiffres représentant le nombre de répétitions de chacun de ces 12 loci spécifiques.
16. **Marqueur VNTR Variable Number of Tandem DNA Repeats** : séquences nucléotidiques courtes en tandem à nombre variable. Cinq répétitions en tandem exactes (locus ETR) sont utilisées pour l'analyse VNTR du complexe *M. tuberculosis*.
17. **Allèle** : version variable d'un même gène.
18. **Base de données SITVIT** : base de données de l'Institut Pasteur de Guadeloupe consultable en ligne `http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/query`, permettant n'analyser des data liées au MTBC. Elle comprend les spoligotypes de *M. tuberculosis*, ainsi que les marqueurs utilisés pour les détecter MIRU12, VNTR, SIT, MIT, VIT, les différentes branches de MTBC, les pays d'origine et l'année de découverte.

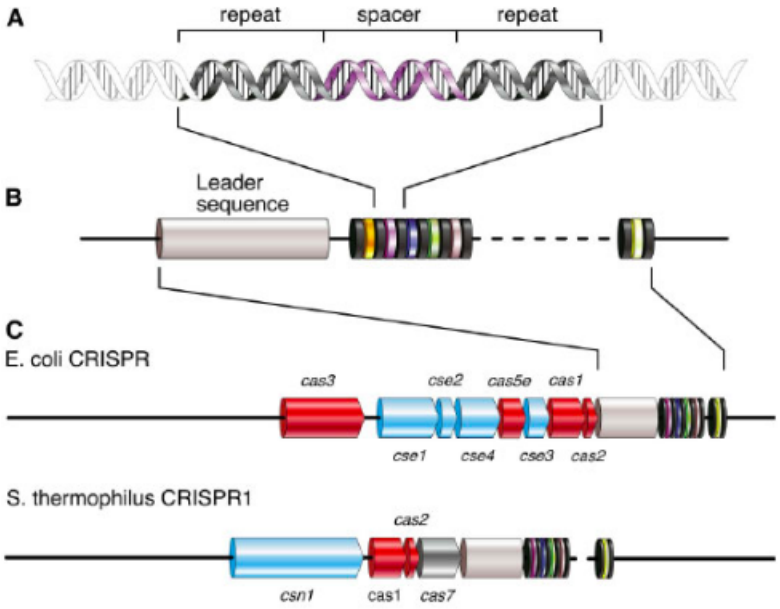


Figure 4 – Structure du locus CRISPR-Cas, d'après
https://www.sinobiological.com/crispr-locus.html

est héritable par transmission aux cellules filles et se conserve donc pour une même espèce. Chez *M. tuberculosis*, chaque série de répétitions contient 36 bp [19] ; les répétitions étant régulièrement espacées par des espaceurs de 34 à 41 bp. A l'heure actuelle, 104 espaceurs ont été identifiés dans toutes les souches de *M. tuberculosis*. Les loci CRISPR sont généralement adjacents aux gènes Cas, dont ils sont séparés par une séquence de 300 à 500 bp, appelée leader qui contrôle à la fois l'acquisition de l'ADN viral par les espaceurs et la fabrication des protéines. Les gènes Cas produisent des protéines aux fonctionnalités multiples et notamment les enzymes [20] capables de couper l'ADN en vue de sa réparation.

Ces séquences CRISPR incorporent dans les espaceurs des fragments d'ADN de bactériophages qui ont déjà infecté la bactérie, et sont stockés pour détecter et détruire l'ADN de bactériophages similaires en cas de nouvelle infection. Par conséquent, CRISPR-Cas est un système immunitaire naturel utilisé par les bactéries pour se protéger des infections virales.

4.3. Fonctionnement du système CRISPR-Cas

Les systèmes CRISPR-Cas sont de trois types et utilisent les différents gènes Cas pour intégrer des fragments de gènes étrangers dans les espaceurs de CRISPR. Par exemple, dans le cas d'une bactérie qui détecte la présence d'ADN ou d'ARN d'un virus, elle produit une enzyme nucléase appelée Cas9 capable de couper l'ADN viral, puis une séquence d'ARN CRISPR notée crARN correspondant à celle de l'ADN du virus et servant de guide ARN, et finalement une séquence d'ARN traceur notée trARN. Lorsque trARN trouve sa cible parmi le génome du virus, Cas9 sectionne l'ADN viral puis en incorpore un fragment dans un espaceur du génome de la bactérie, conservant ainsi en mémoire une trace de ce virus en vue d'une éventuelle infection future. Les espaceurs servent donc de banque de mémoire en conservant l'ADN des virus qui ont attaqué la bactérie. Cette fonctionnalité va être exploitée de différentes manières par les biologistes.

19. **bp** : une paire de base.
20. **Enzyme de restriction** : protéine capable de couper un fragment d'ADN au niveau d'une séquence de nucléotides caractéristique appelée site de restriction. Chaque enzyme de restriction reconnaît ainsi un site spécifique.

La technologie CRISPR-Cas9, s'inspirant du système du même nom, a d'abord été utilisée pour typer les souches bactériennes, suivant une technique appelée spoligotypage. CRISPR-Cas9 est actuellement principalement employé comme ciseau moléculaire afin d'éditer le génome et d'introduire localement des modifications génétiques.

5. Le spoligotypage

La région DR du locus CRISPR-Cas présente un niveau de polymorphisme suffisant pour pouvoir classer phylogéographiquement les souches de *M. tuberculosis*. Le polymorphisme entre les différentes souches résulte des variations et de l'identité des espaceurs. C'est ce polymorphisme qui est exploité en 1997 par Kamerbeek et al. et expliqué dans [16] comme technique de génotypage spécifique de *M. tuberculosis*. Le *Spacer Oligonucleotide Typing*, repose sur la détection de séquences répétitives trouvées entre les gènes d'un agent infectieux au sein d'un locus CRISPR-Cas. Pour ce faire, la région DR d'un isolat à tester subit un traitement par amplification PCR [21] ou celui d'une puce à ADN [22], pour dévoiler un motif de taches correspondant aux espaceurs.

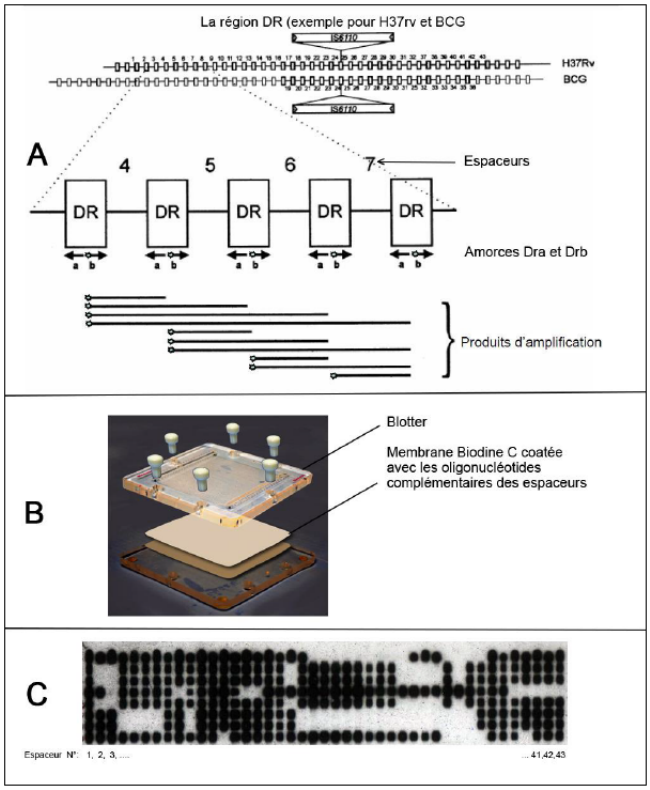


Figure 5 – Les différentes étapes du spoligotypage d'après *Etudes descriptive, épidémiologique, moléculaire et spatiale des souches Mycobacterium tuberculosis circulant à Antananarivo, Madagascar*

21. **PCR Polymerase Chain Reaction** : méthode de réaction en chaîne utilisant un polymère pour dupliquer en grand nombre une séquence d'ADN spécifique. La méthode PCR repose sur le cycle thermique, qui expose les séquences à des cycles répétés de chauffage et de refroidissement pour permettre différentes réactions dépendantes de la température comme la fusion de l'ADN et la réplication de l'ADN par les enzymes. La méthode PCR utilise deux agents principaux : les polymères d'ADN i.e. des macromolécules répétant un même motif structural d'ADN et les amorces de séquençage.

22. **Puce à ADN** : ensemble de molécules d'ADN fixées sur une petite surface solide permettant de mesurer le niveau d'expression d'un grand nombre de gènes simultanément, ou de déterminer le génotype de plusieurs régions d'un génome.

La comparaison de ces motifs permet la différentiation des souches. Quarante-trois espaceurs les plus polymorphes ont été utilisés pour le typage des mycobactéries suivant Kamerbeek et al. La classification classique de MTBC utilise donc un groupe de 43 bits représentant la présence ou l'absence d'espaceurs dans le locus CRISPR, qu'on appelle spoligotype. Des études pour augmenter le niveau de discrimination du spoligotypage ont été faites en 2010 utilisant 68 espaceurs. A l'heure actuelle l'équipe AND de l'université de Franche Comté utilise 98 espaceurs pour ce génotypage.

Numéro	Numéro espaceur en spoligotypage	Numéro espaceur dans le génome	séquence des oligonucléotides (5'→3')	Numéro	Numéro espaceur en spoligotypage	Numéro espaceur dans le génome	séquence des oligonucléotides (5'→3')
1	1	2	CGCTCCCCTAGTCGT	53	-	17	TCTTGAGCAACGCCATC
2	2	3	TGGGCGACAGCTTTTGA	54	-	45	AAGTTGGCGCTGGGG
3	3	4	CTTCCAGTGATCGCCTT	55	-	48	AACCGTCCCACCTG
4	4	12	TCATACGCCGACCAATC	56	-	49	AACACTTTTTTTGAGCGTGG
5	5	13	TTCTGACCACTTGTGCG	57	-	50	CGGAAACGCAGCACC
6	6	14	TCATTTCCGGCTT	58	-	54	CGATCATGAGAGTTGCG
7	7	15	TGAGGAGAGCGAGTACT	59	-	55	TTTTCGCTGTTGTGGTTCT
8	8	18	TGAAACCGCCCCCAG	60	-	56e	AGCACCTCCCTTGACAA
9	9	19	ACTCGGAATCCCATGTG	61	-	57	TGCTGACTTCGCCTGTA
10	10	20	CTCTAGTTGACTTCCGG	62	-	58	CGAGCAGCGGCATAC
11	11	21	CAGGTGAGCAACGGC	63	-	59	GCATCCACTCGTCGC
12	12	22	ATGGGATATCTGCTGCC	64	-	60	TGGTAATTGCGTCACGG
13	13	23	ATTGCCATTCCCTCTCC	65	-	61	ACCATCCGACGCAGG
14	14	24	TTTCGGTGTGATGCGGA	66	-	66	CCACGCTACTGCTCC
15	15	25	TGAATAACGCGCAGTGAAT	67	-	67	CACCGCCGATGACAG
16	16	26	TCGCACGAGTTCCCG	68	-	68	GTGTTTCGGCCGTGC
17	17	27	CCGGCAACAATCGCG	69	-	69	GTTGCATTCGTCGACTG
18	18	28	TGCAGATGGTCCGGG	70	-	70	GGCGGCGCCGAGAA
19	19	29	TTGCGCTAACTGGCTTG	71	-	71	TTCCATGACTTGACGCC
20	20	30	ATTTCCTTGACCTCGCC	72	-	72	CGATGCGGCCACTAG
21	21	31	CGATGTCGATGTCCCAA	73	-	73	GCTGACCCCATGGATG
22	22	32	ACGGCACGATTGAGACA	74	-	74	CAACAAGGTCTACGCGT
23	23	33	GTCCAGCTCGTCCGT	75	-	75	GATCAGGCGAAGGCG
24	24	34	GCCTGCTGGGTGAGA	76	-	76	ATTGCAGCGACGGGC
25	25	35	GGAGCCGATCAGCGA	77	-	77	CAACGACGCTGTATTGG
26	26	36	CTTCAGCACCACCATCA	78	-	78	AGCAGCATGGACGGTTT
27	27	37	TTCGTGATCTCTTCCCG	79	-	79	GCGGATGTGGTGGTC
28	28	38	GATCACAACACCAACTAATG	80	-	80	GTACATAGCGAGCTG
29	29	39	GAAATACAGGCTCCACG	81	-	81	GCCGCGGGTTTCGTT
30	30	40	TCTTGACGATGCGGTTG	82	-	82	GGGGCGTGTGTTCGT
31	31	41	TTCGCGTCAGACAGGTT	83	-	83	CTGGTGTGCTTATGCCT
32	32	42	ACTCCCGACCAAATAGG	84	-	84	CAAATGTTTGGACTGTGATC
33	33	43	TCGACACCGACATGAC	85	-	85	TTGTCGCGCGCCTTTTT
34	34	44	GAAGTCACCTCGCCC	86	-	86	GTTTCAGTTTTCTTGTCCC
35	35	46	AGTCCGTACGCTCGAAA	87	-	87	CTGGTTGTTGCCCGG
36	36	47	CGAAATCCAGCACCACA	88	-	88	TGTTCGGTGTTCTCCTG
37	37	51	TTTGAGCGCGAACTCGT	89	-	89	TCATGACGAGCCCGCA
38	38	52	TGGATGGCGGATGCG	90	-	90	ACACGGCCTGATCGGT
39	39	53	AAATCGGCGTGGGTAAC	91	-	91	CGGATTGTCTGGCCC
40	40	62	TCATACAGGTCCAGTGC	92	-	92	TAAGCACGCGTCTGTCA
41	41	63	GCTTTCCGGCTTCTATC	93	-	93	GACCACCGAATCACCAT
42	42	64	GACATGGACGAGCGC	94	-	94	TCTGGTAGTGGGCTTCT
43	43	65	CAGAATCGCACCGGG	95	-	11	ACATGCCGTGGCTCA
44	-	1	CAACCCGGAATTCTTGC	96	-	16	CACGACGTTAGGGCA
45	-	5	CAGGCGTGGCTAGG	97	-	5	CGGCAGGCGTGGCTA
46	-	6	GTCGCCGTAAGTGCC	98	-	6	CCGTCGCCGTAAGTG
47	-	7	GTTGACCACGAATTTTCAG	99	-	17	GAGCAACGCCATCAT
48	-	8	GCTGGCGCGCATCAT	100	-	11	TGAGCCACGGCATGT*
49	-	9	CCATATCGGGGACGG	101	-	16	ATGCCCTAACGTCGTG*
50	-	10	GCGTCGTGCCATCAG	102	-	5	TAGCCACGCCTGCCG*
51	-	11	CCGTGCACATGCCGT	103	-	6	CACTTACGGCGACGG*
52	-	16	ACGTTAGGGCATGCAG	104	-	17	GATGATGGCGTTGCT*

Figure 6 – Espaceurs connus chez MTBC, d'après *Etudes descriptive, épidémiologique, moléculaire et spatiale des souches Mycobacterium tuberculosis circulant à Antananarivo, Madagascar*

La technique ne nécessite pas une importante quantité d'ADN car elle est basée sur une amplification de la région DR par PCR. Les spoligotypes ainsi obtenus peuvent être partagées entre laboratoires et corroborent les résultats recueillis à partir d'autres marqueurs génétiques. Ces données numériques permettent de bien différencier les souches de *M. tuberculosis* et sont de moindre coût comparativement à d'autres méthodes. Cependant, le spoligotypage éprouve des difficultés à bien différencier les

souches au sein de grandes familles de *M. tuberculosis* telles que la lignée 2 par exemple.

Jusqu'à présent, le spoligotypage a permis de fournir une image globale de la diversité des souches de *M. tuberculosis*.

Une nouvelle technologie permettant de combiner le spoligotypage avec des tests moléculaires de sensibilité aux anti-tuberculeux, appelée spoligoriftypage, a été développée pour aboutir à la version TB-SPRINT qui a été décrite en 2013 par Gomgnimbou et al. dans leur article [18]. Elle consiste au typage tuberculose-spoligo-rifampicine-isoniazide fonctionnant sur des systèmes à base de microbilles, à partir notamment de 43 espaceurs, 11 SNPs présents sur rpoB aux positions 516, 526 et 531. Cette nouvelle génération de spoligotypage fournit donc, en plus des données classiques de génotypage, une prédiction basée sur la mutation des profils de résistance aux médicaments.

5.1. Vers une normalisation des spoligotypes

Au début du spoligotypage, il n'existait pas de norme pour décrire les motifs formés par les espaceurs ou simplement les numéroter. Chaque laboratoire utilisait son propre système de numérotation accompagné d'un schéma descriptif du motif. Ce manque de normalisation entravait les possibilités de comparaison des résultats obtenus et le développement d'une vision mondiale de l'évolution de *M. tuberculosis*. Une méthode standardisée de description des spoligotypes a été proposée en 2001 par Dale JW dans son article [10].

Tout d'abord, une base de données centralisée regroupant tous les motifs connus et de leurs numérotations associées existe au RIVM Rijksinstituut voor Volksgezondheid en Milieuhygiene, Bilthoven, Netherlands. Elle peut être consultée au `http://www.caontb.rivm.nl`. A partir de 2001, les nouveaux motifs doivent prendre un unique format de numérotation pour être répertoriés dans cette base de données. Toutefois, cela nécessite l'interrogation systématique de la base de données et la comparaison avec les éléments déjà existants pour chaque nouveau spoligotype. Pour éviter cette perte de temps, de nombreux laboratoires utilisent des systèmes rationnels avec des codes descriptifs assignés à chaque isolat.

Dale JW et al.[10] proposent d'utiliser exclusivement un système rationnel octal ou hexadécimal, sachant qu'il est aisé de passer de l'un à l'autre et qu'il est également facile de retrouver l'état initial binaire. Ainsi, les motifs de spoligotype comprenant 43 bits seraient réduits dans le système octal en 14 groupes de 3 bits auquel s'ajouterait un unique bit, ce qui donnerait finalement un ensemble de 15 chiffres en écriture octale. En ce qui concerne le système hexadécimal, les motifs de 43 bits seraient réduits en 6 groupes de 8 bits avec un dernier groupe ne comprenant que 3 bits, soit 6 groupes de 2 chiffres hexadécimaux. Notons qu'un bit symbolise dans ce cas la présence ou l'absence d'une espaceur dans le locus étudié.

Original pattern															
Binary code	1 0 0	1 1 1	1 1 0	1 1 1	1 1 1	0 1 1	1 1 0	1 1 1	1 1 1	1 1 1	0 0 0	0 1 1	0 1 0	0 0 0	0
Octal blocks	100	111	110	111	111	011	110	111	111	111	000	011	010	000	0
Octal code	4	7	6	7	7	3	6	7	7	7	0	3	2	0	0
Written as	476773677703200														

Original pattern						
Binary code	1 0 0 1 1 1 1	1 0 1 1 1 1 1	1 0 1 1 1 1 0	1 1 1 1 1 1 1	1 1 0 0 0 0 1 1	0 1 0 0 0 0 0
Hex blocks	1001111	1011111	1011110	1111111	11000011	0100000
Hex code	4F	5F	5E	7F	C3	20
Written as	4F-5F-5E-7F-C3-20					

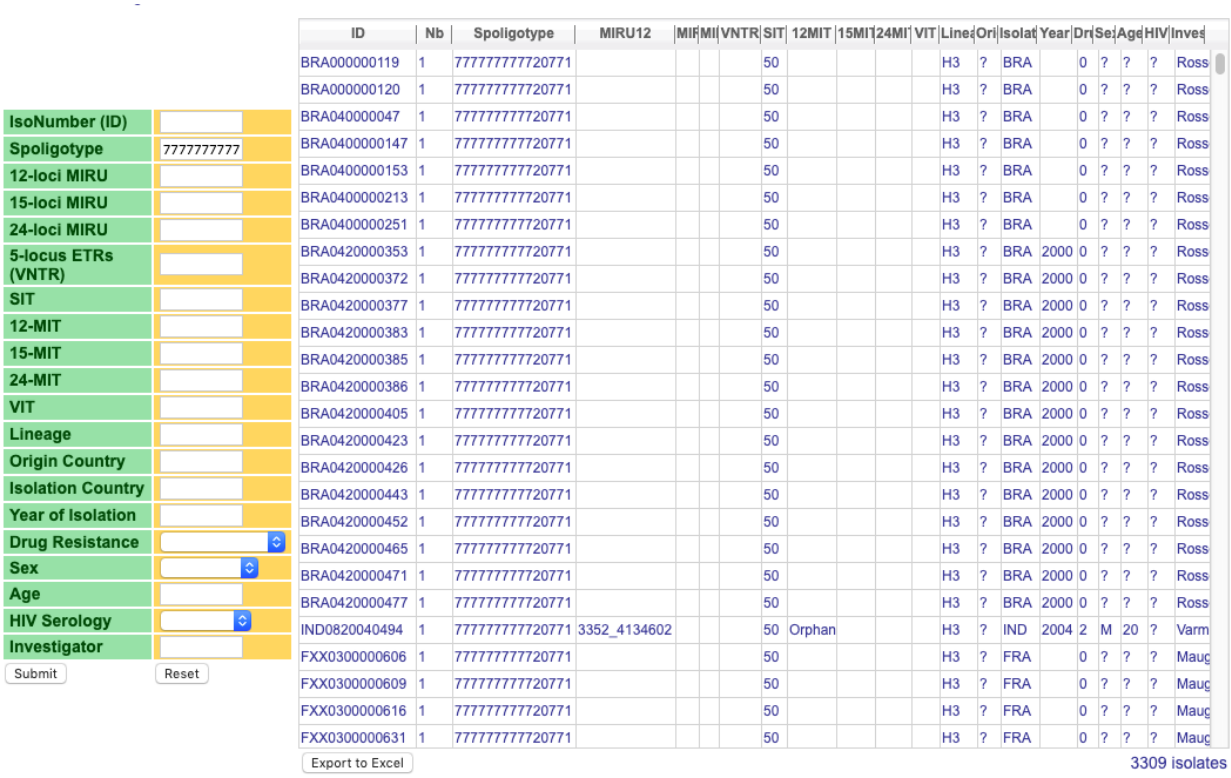
Figure 7 – Exemple de système rationnel octal et hexadécimal, d'après `https://www.mbovis.org/spoligotype-nomenclature.php`

Le site Mbovis.org, `https://www.mbovis.org/database.php`, bien que dédié aux souches animales de MTBC, fournit une application pratique permettant de transformer rapidement les spoligotypes binaires en système octal ou hexadécimal.

5.2. Quel outil informatique pour le spoligotypage ?

Les technologies PCR de génotypage utilisent toujours différents marqueurs tels que les SNPs pour obtenir en laboratoire des résultats fiables. Des logiciels informatiques de prédiction du génotype ont également fait leur apparition pour optimiser les coûts et le gain de temps. Ils offrent un outil de comparaison des résultats obtenus expérimentalement et in silico.

SpolPred est un logiciel de prédiction rapide et précis des spoligotypes de *M. tuberculosis* à partir de séquences génomiques courtes appelées reads [23]. Cet outil développé par Coll F. et al. fonctionne efficacement avec des reads provenant de plateformes telles que Illumina GAII ou HiSeq. SpolPred utilise des fichiers de séquences de reads simples ou par paires au format FASTQ, afin de produire une prédiction de spoligotype au format octal, qui est ensuite comparée au spoligotype correspondant dans la base SITVIT.



ID	Nb	Spoligotype	MIRU12	SIT	12MIT	15MIT	24MIT	VIT	Lineage	Origin	Isolation	Year	Drug	Sex	Age	HIV	Investigator
BRA000000119	1	777777777720771		50					H3	?	BRA		0	?	?	?	Ross
BRA000000120	1	777777777720771		50					H3	?	BRA		0	?	?	?	Ross
BRA040000047	1	777777777720771		50					H3	?	BRA		0	?	?	?	Ross
BRA0400000147	1	777777777720771		50					H3	?	BRA		0	?	?	?	Ross
BRA0400000153	1	777777777720771		50					H3	?	BRA		0	?	?	?	Ross
BRA0400000213	1	777777777720771		50					H3	?	BRA		0	?	?	?	Ross
BRA0400000251	1	777777777720771		50					H3	?	BRA		0	?	?	?	Ross
BRA0420000353	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000372	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000377	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000383	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000385	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000386	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000405	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000423	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000426	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000443	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000452	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000465	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000471	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
BRA0420000477	1	777777777720771		50					H3	?	BRA	2000	0	?	?	?	Ross
IND0820040494	1	777777777720771	3352_4134602	50	Orphan				H3	?	IND	2004	2	M	20	?	Varm
FXX0300000606	1	777777777720771		50					H3	?	FRA		0	?	?	?	Maug
FXX0300000609	1	777777777720771		50					H3	?	FRA		0	?	?	?	Maug
FXX0300000616	1	777777777720771		50					H3	?	FRA		0	?	?	?	Maug
FXX0300000631	1	777777777720771		50					H3	?	FRA		0	?	?	?	Maug

Figure 8 – Exemple de recherche effectuée sur SITVIT2 à partir du spoligotype 777777777720771 dont le résultat est exploitable au format Excel

Dans leur étude [2], Coll F. et al. montrent en 2012 l'utilité de SpolPred en comparant les spoligotypes obtenus par le logiciel avec les résultats de laboratoire. Ils dévoilent ainsi les limites de la méthode expérimentale qui a répertorié cinq faux spoligotypes, alors que SpolPred a su éviter ces erreurs de classification du génotype. Par ailleurs, il apparaît que SpolPred offre plus de rapidité avec des résultats pratiquement identiques à ceux obtenus avec la méthode bio-informatique par assemblage.

23. **Reads** : mélange de courtes séquences oligonocléotidiques de 20 à 200 bp générées par des séquenceurs

Cette dernière, développée en 2008 à l'aide du logiciel Velvet, consiste à fusionner des fragments d'ADN issus d'une plus longue séquence afin d'en reconstruire la séquence originale.

Toutefois, d'après l'étude de Xia et al.[9], la précision de SpolPred est fortement réduite lorsque les reads n'ont pas une taille uniforme, comme par exemple lorsqu'ils proviennent de séquençages Ion Torrent ou de la plateforme de diagnostique clinique Illumina MiSeq. Ainsi, lorsque les reads ne sont pas uniformes, la précision des résultats dépend fortement de leurs tailles et donc du choix initial fait par l'opérateur. Par ailleurs, SpolPred demande à l'utilisateur de spécifier la direction de lecture des reads, et le logiciel n'utilise donc qu'une partie des informations fournies par les reads.

Une problématique de SpolPred en 2020 est que le logiciel n'est plus disponible au téléchargement en ligne. En effet, une visite sur le site officiel `http://www.pathogenseq.org/spolpred`, fourni comme référence dans le document [2] de Coll F. et al., montre que le nom du domaine est à vendre. Preston M., qui a fait partie de l'équipe de recherche de Coll F. pour le développement de SpolPred, a bien créé un site proposant le téléchargement du logiciel `https://www.mybiosoftware.com/spolpred-predict-the-spoligotype-from-raw-sequence-reads.html`, mais le lien est inactif en janvier 2020.

Une alternative à SpolPred est SpoTyping présenté en 2016 dans l'article [9] de Xia et al. comme étant 20 à 40 fois plus rapide que SpolPred pour prédire avec précision des spoligotypes de *M. tuberculosis* à partir de reads de taille uniforme ou variable. Par ailleurs, SpoTyping lit chaque read dans les deux directions en exploitant complètement les informations fournies. SpoTyping réduit la durée des recherches en intégrant l'algorithme BLAST [24] dans ses calculs. Il compare les isolats testés avec ceux ayant le même spoligotype dans la base de données mondiale SITVIT, qui regroupe les données épidémiologiques [25] associées à des isolats de même spoligotype.

L'intérêt d'un outil tel que SpolPred ou Spotyping est qu'il est capable de combiner le spoligotypage avec d'autres méthodes telles que MIRU (unités répétitives entrecoupées de mycobactéries) et VNTR nombre variable de répétitions d'ADN en tandem) en utilisant la base de données SITVIT.

SpoTyping utilise des fichiers de séquences de reads simples ou par paires au format FASTQ et des fichiers de séquences complètes de génomes ou de contigs [26] assemblés au format FASTA. Les séquences de reads sont regroupées en une unique séquence continue au format FASTA pour être ensuite soumise à l'algorithme BLAST qui détecte les régions similaires. Finalement la base de données SITVIT permet d'identifier les isolats ayant le même spoligotype. SpoTyping est limité à une lecture de 250 Mbp au sein des séquences de reads testées, lors de l'utilisation du swift mode qui accélère le temps de traitement.

SpoTyping propose un rapport statistique permettant de résumer le rapprochement avec les spoligotypes trouvés dans la base de données SITVIT, ainsi qu'une estimation du nombre de correspondances positives pour chaque espaceur.

D'après le repository `https://github.com/xiaeryu/SpoTyping-v2.0/blob/master/SpoTyping-v2.0-commandLine/SpoTyping-README.pdf`, les spécifications techniques de SpoTyping sont les suivantes :
- SpoTyping peut s'exécuter sur les principaux systèmes d'exploitation, contrairement à SpolPred

24. **BLAST Basic Local Alignment Search Tool** : logiciel basé sur l'algorithme du même nom qui détecte des régions similaires entre plusieurs séquences biologiques. Le programme compare les séquences de nucléotides aux séquences contenues dans la base de données BLAST pour fournir des résultats statistiquement significatifs.

25. **Epidémiologie** : discipline scientifique qui étudie les problèmes de santé dans les populations humaines, leur fréquence, leur géographie ainsi que les facteurs influents.

26. **Contig** : séquence génomique continue et ordonnée, générée par l'assemblage des clones d'une bibliothèque génomique qui se chevauchent.

Figure 9 – Exemple de fonctionnement de SpoTyping, d'après *SpoTyping : fast and accurate in silico Mycobacterium spoligotyping from sequence reads*

qui utilise exclusivement Linux. Il se présente à la fois sous forme de script et sous forme d'application avec une interface graphique.
- SpoTyping est un logiciel open-source qui peut se télécharger gratuitement à l'adresse https://github.com/xiaeryu/SpoTyping-v2.0. SpoTyping nécessite l'utilisation de Python2.7 et BLAST.
- il est recommandé d'utiliser le swift mode paramétré par défaut si le débit de séquençage[27] est inférieur à 135 Mbp. Pour les débits de séquençage inférieurs à 135 Mbp ou supérieurs à 1260 Mbp, les seuils doivent être réglés entre 0.018 et 0.1486 fois la profondeur de lecture estimée pour les hits sans erreur, et entre 0.018 et 0.1488 fois la profondeur de lecture estimée pour les hits tolérant une erreur. Notons que la profondeur de lecture est définie par le débit de séquençage divisé par 4 500 0000 qui correspond à l'estimation de la longueur d'un génome de *M. tuberculosis*.

Le fichier obtenu propose une prédiction de spoligotype au format de code binaire et octal. Le fichier log obtenu contient le nombre de correspondances positives des résultats de BLAST pour chaque séquence d'espaceurs. Le fichier xls Excel obtenu fournit le résultat de la recherche de spoligotype auprès de la base de données SITVIT2[28].

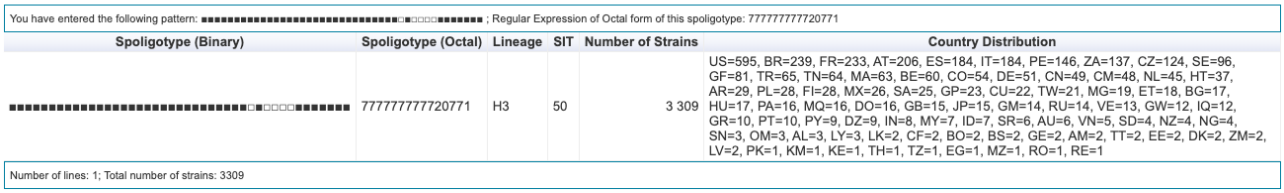
L'étude de Iwai H. et al.[11] envisage une autre possibilité de travail et montre l'intérêt d'une analyse de *M. tuberculosis* à l'aide de serveurs, appelée CASTB, et notamment le spoligotypage. Le Webserver fournit une vue complète des données, mais les performances de chaque outil utilisé ne sont pas décrites dans l'article. Il est probable que le spoligotypage prenne plus de temps en passant par un serveur suite au problème de disponibilité des données et aux lenteurs de téléchargement de ces données. Il semblerait que SpoTyping, de par sa configuration locale, puisse fournir un résultat en une minute.

27. **Séquençage du génome** : consiste, par des méthodes chimiques ou de biologie moléculaire, à déterminer l'ordre des nucléotides de l'ADN.
28. **Base de données SITVIT2** : mise à jour de la base de données SITVIT, consultable en ligne http://www.pasteur-guadeloupe.fr:8081/SITVIT2/index.jsp

5.3. Comparaison de spoligotypes

Une fois les spoligotypes de différentes lignées obtenus, il est nécessaire de les comparer pour chercher à faire ressortir les points communs ou certains traits pouvant être liés à une mutation particulière. Il existe à l'heure actuelle un premier outil en ligne de comparaison du nom de SpolSimilaritySearch, accessible à l'adresse `http://www.pasteur-guadeloupe.fr:8081/SpolSimilaritySearch/index.jsp`, et présenté par Couvin D. et al.[19]. SpolSimilaritySearch incorpore un algorithme de recherche de similitudes entre spoligotypes dans la base de données SITVIT2. Cet outil permet d'analyser les modèles de propagation et d'évolution de *M. tuberculosis* en comparant des modèles de spoligotypes similaires, de distinguer les modèles répandus, confinés ou spécifiques, d'identifier les modèles ayant de grands blocs supprimés ou encore de fournir les modèles de distribution par pays pour chaque spoligotype interrogé.

Par exemple, si on sélectionne le spoligotype 777777777720771 appartenant à la lignée H3, et qu'on interroge la base SpolSimilaritySearch, on obtient les rapprochements suivants :


You have entered the following pattern: ■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■■■■■■ ; Regular Expression of Octal form of this spoligotype: 777777777720771

Spoligotype (Binary)	Spoligotype (Octal)	Lineage	SIT	Number of Strains	Country Distribution
■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■■■■■■	777777777720771	H3	50	3 309	US=595, BR=239, FR=233, AT=206, ES=184, IT=184, PE=146, ZA=137, CZ=124, SE=96, GF=81, TR=65, TN=64, MA=63, BE=60, CO=54, DE=51, CN=49, CM=48, NL=45, HT=37, AR=29, PL=28, FI=28, MX=26, SA=25, GP=23, CU=22, TW=21, MG=19, ET=18, BG=17, HU=17, PA=16, MQ=16, DO=16, GB=15, JP=15, GM=14, RU=14, VE=13, GW=12, IQ=12, GR=10, PT=10, PY=9, DZ=9, IN=8, MY=7, ID=7, SR=6, AU=6, VN=5, SD=4, NZ=4, NG=4, SN=3, OM=3, AL=3, LY=3, LK=2, CF=2, BO=2, BS=2, GE=2, AM=2, TT=2, EE=2, DK=2, ZM=2, LV=2, PK=1, KM=1, KE=1, TH=1, TZ=1, EG=1, MZ=1, RO=1, RE=1

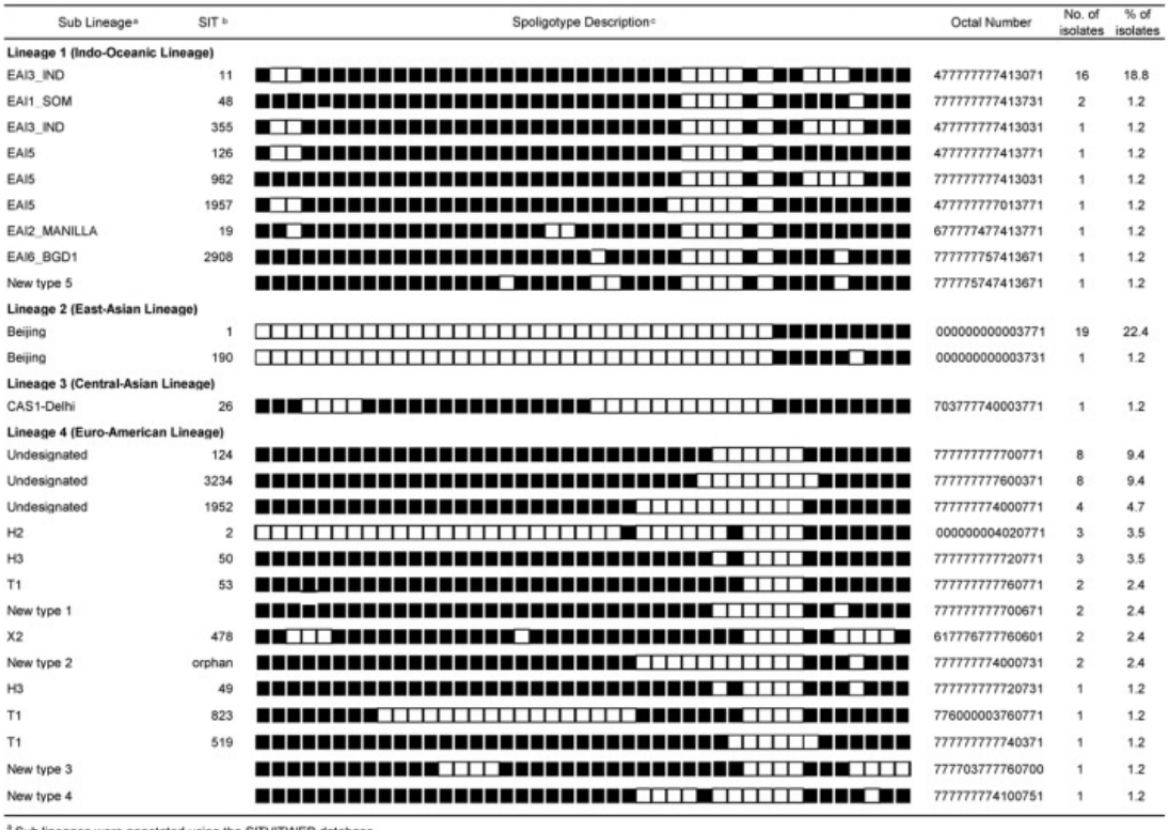
Number of lines: 1; Total number of strains: 3309

Figure 10 – Recherche effectuée sur le site de SpolSimilaritySearch

Cet outil pourrait donc s'avérer utile pour commencer à chercher des liens entre les sept lignées de *M. tuberculosis* et les spoligotypes de différentes souches. Un tableau comparatif de différents spoligotypes en fonction des lignées de *M. tuberculosis* pourrait ressembler à celui de la figure 11.

6. Conclusion

Nous avons brièvement vu que la tuberculose a su évoluer avec l'humanité tout en profitant des changements importants qui l'ont affectée. Sept lignées de *M. tuberculosis* ont pu se développer suite à des mutations, différentes infections ont été mémorisées dans l'ADN de la bactérie et des souches résistantes aux antibiotiques sont récemment apparues. Etudier les liens entre les mutations successives de la bactérie et les diverses lignées géographiques, chercher à mieux comprendre les mécanismes d'évolution de son génome lui permettant de développer des résistances aux antibiotiques, sont autant de questions ouvertes qui permettraient de faire avancer la lutte contre cette maladie. Pour cela, plusieurs marqueurs ont été utilisés jusqu'à présent permettant le génotypage de *M. tuberculosis*, que se soit en laboratoire ou in silico. La structure du locus CRISPR-Cas constitue un marqueur privilégié, permettant le spoligotypage de la tuberculose. Son étude prometteuse devrait permettre de mieux interpréter les sept lignées de *M. tuberculosis*.



Sub Lineage[a]	SIT [b]	Spoligotype Description[c]	Octal Number	No. of isolates	% of isolates
Lineage 1 (Indo-Oceanic Lineage)					
EAI3_IND	11		477777777413071	16	18.8
EAI1_SOM	48		777777777413731	2	1.2
EAI3_IND	355		477777777413031	1	1.2
EAI5	126		477777777413771	1	1.2
EAI5	962		777777777413031	1	1.2
EAI5	1957		477777777013771	1	1.2
EAI2_MANILLA	19		677777477413771	1	1.2
EAI6_BGD1	2908		777777757413671	1	1.2
New type 5			777775747413671	1	1.2
Lineage 2 (East-Asian Lineage)					
Beijing	1		000000000003771	19	22.4
Beijing	190		000000000003731	1	1.2
Lineage 3 (Central-Asian Lineage)					
CAS1-Delhi	26		703777740003771	1	1.2
Lineage 4 (Euro-American Lineage)					
Undesignated	124		777777777700771	8	9.4
Undesignated	3234		777777777600371	8	9.4
Undesignated	1952		777777774000771	4	4.7
H2	2		000000004020771	3	3.5
H3	50		777777777720771	3	3.5
T1	53		777777777760771	2	2.4
New type 1			777777777700671	2	2.4
X2	478		617776777760601	2	2.4
New type 2	orphan		777777774000731	2	2.4
H3	49		777777777720731	1	1.2
T1	823		776000003760771	1	1.2
T1	519		777777777740371	1	1.2
New type 3			777703777760700	1	1.2
New type 4			777777774100751	1	1.2

Figure 11 – Exemple comparatif de 26 spoligotypes correspondant à 4 lignées différentes de *M. tuberculosis*, d'après *Insight into genetic diversity of Mycobacterium tuberculosis in Kandy, Sri Lanka reveals predominance of the Euro-American lineage*

Références

[1] Comas I. et al. *Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans*, Nat Genet. 45(10) : 1176–1182. doi :10.1038/ng.2744

[2] Coll F. et al., *SpolPred : rapid and accurate prediction of Mycobacterium tuberculosis spoligotypes from short genomic sequences*, Bioinformatics. 28(22) :2991–3

[3] Brynildsrud O.B. et al., *Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation*, 4(10) : eaat5869. doi : 10.1126/sciadv.aat5869

[4] Driscoll J. R., *Spoligotyping for molecular epidemiology of the Mycobacterium tuberculosis complex*, 551 :117-28. doi : 10.1007/978-1-60327-999-4 10

[5] Jinek M. et al, *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*, 337(6096) :816-21. doi : 10.1126/science.1225829

[6] Gori A. et al, *Spoligotyping and Mycobacterium tuberculosis*, 11(8) : 1242–1248. doi : 10.3201/1108.040982

[7] Perry S. et al., *Infection with Helicobacter pylori is associated with protection against tuberculosis*, 5(1) :e8804. doi : 10.1371/journal.pone.0008804

[8] Perry S. et al, *The immune response to tuberculosis infection in the setting of Helicobacter pylori and helminth infections*, 141(6) : 1232–1243. doi : 10.1017/S0950268812001823

[9] Xia E. et al., *SpoTyping : fast and accurate in silico Mycobacterium spoligotyping from sequence reads*, 8 :19. doi 10.1186/s13073-016-0270-7

[10] Dale JW. et al., *Spacer oligonucleotide typing of bacteria of the Mycobacterium tuberculosis complex : recommendations for standardised nomenclature.*, 5(3) :216–9

[11] Iwai H et al., *CASTB (the comprehensive analysis server for the Mycobacterium tuberculosis complex) : A publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. Tuberculosis.*, 95(6) :843–4

[12] Demay C. et al., *SITVITWEB - A publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology.*, Infect Genet Evol. 12 :755–66

[13] McGovern Institute Channel, *Genome Editing with CRISPR-Cas9*, https://www.youtube.com/watch?v=2pp17E4E-O8

[14] O'Neil M.B. et al., *Lineage specific histories of Mycobacterium tuberculosis dispersal in Africa and Eurasia*, bioRxiv. 10.1101/210161

[15] Ratovonirina N. H., *Etudes descriptive, épidémiologique, moléculaire et spatiale des souches Mycobacterium tuberculosis circulant à Antananarivo, Madagascar*, Thèse de Doctorat de l'Université Paris-Saclay

[16] Kamerbeek et al., *Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology*, 35(4) : 907–914

[17] Mendis C. et al., *Insight into genetic diversity of Mycobacterium tuberculosis in Kandy, Sri Lanka reveals predominance of the Euro-American lineage*, International Journal of Infectious Diseases 87 84-91

[18] ,Gomgnimbou M. K. et al., *Tuberculosis-Spoligo-Rifampin-Isoniazid Typing : an All-in-One Assay Technique for Surveillance and Control of Multidrug-Resistant Tuberculosis on Luminex Devices*, 51(11) :3527-34. doi : 10.1128/JCM.01523-13

[19] Couvin D. et al., *SpolSimilaritySearch - A web tool to compare and search similarities between spoligotypes of Mycobacterium tuberculosis complex*, 105 :49-52. doi : 10.1016/j.tube.2017.04.007