



Méthodes

Création d'un réseau de génotypage CRISPR en France et en Europe : expérience de cinq ans en matière de recherche, de formation et de santé publique.

Christophe Sola¹ (christophe.sola@u-psud.fr), Guislaine Refrégier¹ et Michel Gomgnimbou^{1,2}.

1. Institut de génétique et microbiologie, équipe IGEPE et services de génotypage « Beads4Med », CNRS-université Paris-Sud, UMR8621, Orsay, France.

2. Centre Muraz, Bobo-Dioulasso, Burkina Faso.

Nous avons lancé une initiative de partenariat et de mise en réseau de laboratoires entre une entreprise industrielle privée, Luminex Corporation (TX), et un laboratoire public, l'Institut de génétique et microbiologie (IGM, UMR8621), pour la mise au point de technologies d'analyse parallélisées de santé publique sur microbilles et à haut débit, et ce compte tenu de l'intérêt croissant suscité par les objets génétiques polymorphes, plus précisément les séquences CRISPR (*Clustered Regularly Interspersed Short Palindromic Repeats*). Nous savons que ces loci sont utiles pour la surveillance et le contrôle des maladies infectieuses bactériennes ; nous allons montrer comment nos études initiales sur une maladie respiratoire, la tuberculose, nous ont conduits à nous intéresser aux pathogènes de l'alimentation et présentons de nouvelles perspectives.

Mots-clés : CRISPR, hybridation sur microbilles multiplex, *Legionella pneumophila*, *Salmonella enterica*, *Mycobacterium tuberculosis*.

Contexte et historique de la recherche sur les CRISPR ; typage du complexe *Mycobacterium tuberculosis*

La première caractérisation d'un locus CRISPR a été réalisée sur *E. coli* par Ishino *et al.* en 1987. À cette époque, l'acronyme « CRISPR » n'existait pas ; il a été forgé par Jansen *et al.* en 2002. En 2004, un article des CDC (*Centers for Disease Control*, centres épidémiologiques américains) d'Atlanta décrivait le passage de la technologie des puces à ADN 2D à celle des puces 3D pour la technique de « spoligotypage » de la tuberculose, mise au point en 1997 par Kamerbeek *et al.* (Cowan *et al.* 2004). Cette technique de génotypage permet d'évaluer la diversité des loci CRISPR dans le complexe *M. tuberculosis* (MTBC) et demeure l'une des principales techniques de génotypage utilisées pour l'épidémiologie moléculaire de la tuberculose. Des projets de bases de données de spoligotypage (SpolIDB1 à SpolIDB4 et aujourd'hui SITVITWEB³) ont contribué à donner une grande visibilité internationale au spoligotypage et nous ont permis d'élargir notre connaissance de la structure phylogéographique des populations du MTBC ainsi que de l'histoire évolutive des loci CRISPR. À ce jour, près de 840 publications utilisent le terme « spoligotypage » comme mot-clé dans PubMed.

Le MTBC, agent de la tuberculose humaine et bovine, contient un locus CRISPR unique et « figé » (sans nouvelle acquisition d'espaces) comprenant une répétition directe (DR, *Direct Repeat*) de 36 paires de bases (pb) de séquence consensus : 5'-GTCGTCAGACCCAAAACCCCGAGAGGGGACGGAAAC-3'. Sur un total de 94 espaces publiés, seuls 68 se sont révélés spécifiques du MTBC et 26 ont été observés dans l'écotype *M. canettii*. Les résultats ne révèlent pratiquement aucune variation génétique au niveau du DR.

Les dosages parallèles sur réseaux optiques de microbilles (MBAA, *Multiplex-Bead-Array Assays*) sont une innovation récente. Les MBAA peuvent être utilisés pour des ligands immunologiques et moléculaires (Dunbar *et al.* 2006). Le parallélisme ou multiplexage (défini par le dosage quantitatif de plusieurs analytes simultanément dans de petits volumes

de matière) permet de doser jusqu'à 500 analytes grâce à auran de types de billes colorées, présentant chacune leur propre signature spectrale. La miniaturisation des lasers, la microfluidique et les innovations bio-informatiques permettent de disposer de systèmes robustes et de résultats tout aussi robustes. Les économies possibles par rapport aux méthodes en analyses uniques et non simultanées, tant en termes de temps que d'argent, font le succès croissant de ces méthodes dans les laboratoires de recherche et les laboratoires cliniques. En effet, les technologies de parallélisme sont à la génétique ce que la fibre optique est à l'informatique : un moyen de disposer de nombreux canaux d'informations dans un seul tuyau et de produire et transmettre une avalanche de données. Cependant, ces procédés extraordinaires doivent évidemment être associés à des processus de stockage et d'analyse de données (par exemple, avec des systèmes de *cloud computing* et de *data mining* domaines qui constituent aujourd'hui les *big data*).

Lorsque nous avons fondé l'équipe IGEPE en 2007, la filiale hollandaise de Luminex Corp. Luminex (Luminex BV, Oosterhout) nous a aidés à mettre en place la cytométrie en flux sur microbilles au sein de notre équipe de recherche. L'équipe initiale, composée de deux personnes, s'est rapidement étoffée. Aucun laboratoire européen n'avait essayé de reproduire le transfert du spoligotypage sur un format « microbilles » ; cette première étape avait été réalisée au CDC, comme expliqué plus haut. Dans notre équipe, Zhang *et al.* ont fait évoluer, le format de spoligotypage de 43 espaces initial vers un format sur 68 espaces, plus efficace pour discriminer les souches de *M. bovis* et des génotypes d'Asie du Sud-Est sur des isolats cliniques de MTBC (Zhang *et al.* 2010, 2011).

Grâce au soutien financier de la Région Île-de-France et à l'assistance technique de Luminex BV, nous avons commencé à fournir des services de génotypage, de formation et de commercialisation de services de typage sur microbilles ; nous avons également pu mettre au point de nouveaux dosages personnalisés sur microbilles couplés (hors catalogue) avec l'aide de quelques clients ou partenaires situés en Europe

3. Spoligo-International Type, VNTR-International Type. http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/.



Méthodes

(comme le RIVM ou le KIT aux Pays-Bas), et notamment en France (grâce à différentes équipes de recherche comme l'Inserm de Montpellier, l'hôpital Bichat, le centre national de référence pour la tuberculose à Paris ou des équipes de l'Anses). Nous avons formé des collaborateurs spécialisés au sein de ces équipes et avons recherché de nouveaux moyens d'utiliser les techniques sur microbilles pour les problèmes de santé publique. La **figure 1** est une mini-chronologie de l'évolution des méthodes développées au cours des six dernières années.

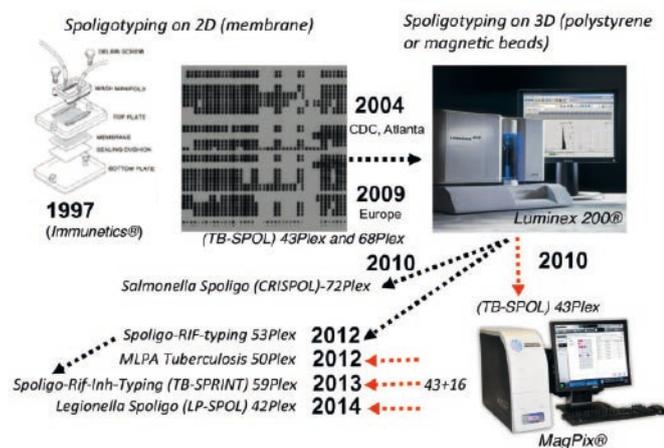


Figure 1. Mini-chronologie de la mise au point des techniques à microbilles sur plates-formes Luminex ; depuis 2010, toutes les techniques sont conçues sur les systèmes Luminex 200® et MagPix®. Flèches noires : dosages conçus sur Luminex 200® ; flèches rouges : dosages conçus sur MagPix®.

Technique CRISPOL : application de la technique CRISPR à *Salmonella enterica*

Parmi les entérobactéries, *Salmonella enterica* est un pathogène alimentaire fréquent et revêt un grand intérêt économique. Bien que coûteuse et fastidieuse, l'identification de milliers de sérovars demeure à ce jour la technique de référence en surveillance épidémiologique. Le tout premier système de typage de sérovars date de 1934, mais il a été régulièrement perfectionné depuis, les dernières améliorations datant de 2007. Bien que les méthodes (PFGE², MLST³, MLVA⁴), les bases de données et les réseaux (PulseNet) de génotypage moléculaire soient bien en place, il est nécessaire de disposer de techniques plus économiques. L'épidémie d'*E. coli* O104:H4 qui a récemment touché l'Allemagne et la France a également mis l'accent sur la nécessité d'utiliser davantage les loci plus discriminants, comme les loci CRISPR, dans la prise en charge des crises sanitaires.

C'est grâce à une énorme opération de séquençage de loci CRISPR, sur les collections historiques du Laboratoire des entérobactéries de l'Institut Pasteur, menée grâce à l'engagement sans faille de son directeur actuel, le Docteur F.X. Weill, et au soutien des services de génomique et de santé publique, que ces équipes ont pu analyser de manière exhaustive la diversité génétique des loci CRISPR chez *Salmonella*.

Il existe deux loci CRISPR chez *Salmonella* : CRISPR1 et CRISPR2. Le locus CRISPR1 se situe en aval du gène *iap*, tandis que le locus CRISPR2 se trouve en amont du gène *ycgF*. Les DR de ces deux loci font 29 paires de base de longueur et la séquence consensus est la suivante : 5'-CGGTTTATCCCCGCTGGCGCGGGGAACAC-3'. Une analyse des CRISPR par PCR et séquençage de 783 souches appartenant à 130 sérotypes a révélé la présence de 3 800 espaceurs mesurant en moyenne 32 pb (Fabre *et al.* 2012). Une corrélation a été observée entre la composition en espaceurs, le sérotype et le type MLST. De plus, la microévolution des espaceurs (duplication, triplification, perte ou gain d'espaceurs, présence de variantes d'espaceurs selon la technique SNP ou VNTR) a permis de différencier des sous-types au sein des sérotypes prédominants comme Typhimurium (STM), sérotype le plus répandu dans le monde. Dans huit génomes et 150 souches du sérotype Typhimurium et de son variant monophasique 1,4,[5],12:i:-, 57 allèles CRISPR1, 62 allèles CRISPR2 et 83 allèles CRISPR1-CRISPR2 combinés ont été identifiés. Quarante espaceurs uniques (dont quatre avec variantes, par exemple des variantes SNP ou VNTR) ont été identifiés dans les allèles CRISPR1. Trente-neuf espaceurs uniques (dont deux avec variante SNP) ont été identifiés dans les allèles CRISPR2. Certaines populations bien caractérisées, comme les isolats DT104 multirésistants aux médicaments (MDR, *Multidrug-Resistant*), les isolats ST313 MDR d'Afrique et les isolats DT2 de pigeons, présentaient chacune des allèles CRISPR caractéristiques. Cet important polymorphisme de la composition en espaceurs a permis de mettre au point un dosage d'hybridation parallèle en phase liquide sur microbilles, le CRISPOL (*CRISPR Polymorphism*). Ce spoligotypage cible actuellement 72 espaceurs sélectionnés parmi les 3800 caractérisés.

La méthode CRISPOL a été brevetée par l'Institut Pasteur et il s'agit maintenant d'une technique de routine au sein de l'Institut Pasteur et d'un laboratoire de l'Anses (Fabre *et al.* 2012). Il permet l'identification, presque en temps réel, des épidémies émergentes de *Salmonella enterica* sérovar Typhimurium, selon une nouvelle nomenclature de clones CT (*CRISPOL Type*). Notre équipe a participé à la conception, à la mise en œuvre et à l'optimisation de cette méthode, ainsi qu'à l'optimisation du processus de production des réactifs, et elle est en mesure de fournir aujourd'hui des microbilles couplées ayant fait l'objet d'un contrôle de qualité pour les laboratoires de santé publique. La **figure 2** illustre et décrit le principe de la technique CRISPOL

1. Pulsed Field Gel Electrophoresis
2. Multi-Locus sequence Typing
3. Multi-Locus VNTR Analysis



Méthodes

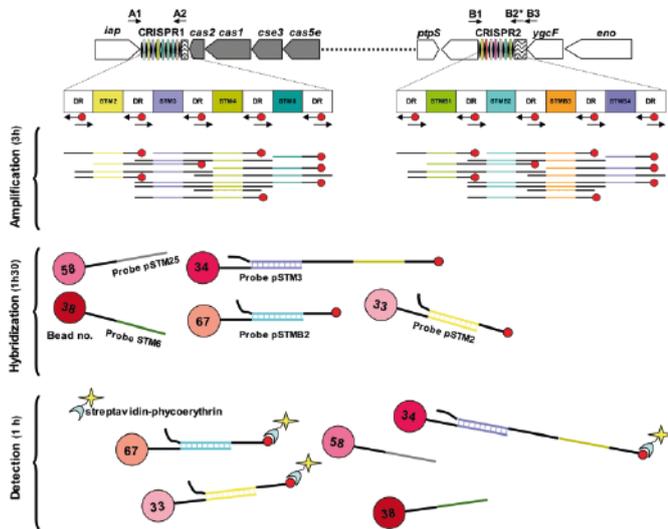


Figure 2. Description de toutes les étapes de la technique CRISPOL. Contrairement au MTBC, *Salmonella* possède deux loci CRISPR. Au total, cette technique de quatre à cinq heures permet d'identifier 72 sondes espaceurs (dont 4 variantes SNP). Elle est utilisée comme technique analytique de routine et permet de détecter les épidémies dès leurs premiers stades, pour quelques euros par dosage. (Reproduit d'après Fabre *et al.* 2012, avec autorisation.)

Legionella pneumophila ST1/pulsotype Paris et technique LP-SPOL

Legionella pneumophila est une bactérie pathogène intracellulaire Gram-négative facultative qui a été identifiée comme étant l'agent infectieux de la maladie des légionnaires ou légionellose en 1977. Plusieurs espèces du genre *Legionella* peuvent provoquer la légionellose, mais *L. pneumophila* est responsable des principales épidémies de légionellose et de plus de ~90 % de l'ensemble des cas cliniques identifiés. Le sérotype 1 de *L. pneumophila* est à lui seul responsable de ~85 % de ces cas. Il s'agit d'un organisme relativement répandu dans les milieux aquatiques, qu'il s'agisse des circuits d'eau naturels ou artificiels. Son pouvoir pathogène est lié à des infections pulmonaires, principalement la pneumonie aiguë, qui peuvent être fatales dans un nombre relativement élevé de cas (29 décès sur 182 cas lors de la première épidémie signalée). La technique de référence pour l'épidémiologie moléculaire de *L. pneumophila* est l'analyse par enzymes de restriction et électrophorèse en champ pulsé (PFGE, *Pulse-Field Gel Electrophoresis*). Bien que lourde, cette technique demeure une référence, et ce, malgré la mise au point du typage séquentiel (SBT, *Sequence-Based Typing*), du typage par anticorps monoclonaux et, plus récemment, de l'analyse multilocus du nombre variable de répétitions en tandem (MLVA), déjà citée plus haut. Ces méthodes présentent toutefois des limites en matière d'épidémiologie moléculaire : elles ne présentent pas de pouvoir discriminant pour certaines souches et se révèlent lentes et fastidieuses.

Dans une étude précédente, Ginevra *et al.* ont démontré que, dans certains cas, des souches indifférenciables par

l'une ou l'autre des méthodes STB et PFGE, notamment au sein de *L. pneumophila* ST1/pulsotype Paris, pouvaient être étudiées de manière plus efficace avec la technique CRISPR, et ils ont mis au point une méthode de spoligotypage sur membrane. L'objectif de notre récente collaboration avec le Centre national de référence des légionelles était de transférer la technique de spoligotypage de la membrane précédemment mise au point vers un format « microbilles ». La DR de *Legionella* est une répétition de 37 pb de séquence suivante : 5'-CCAATAATCCCTCATCTAAAATCCAACCACTGAAAC-3'. Nous avons récemment transféré la technique membranaire sur des microbilles, sur les appareils Luminex 200® et MagPix®, et envisageons de lancer une étude multicentrique internationale. Cette méthode semble prometteuse pour dépister certains clones de *L. pneumophila* (Gomgnimbou *et al.* 2014).

Des méthodes de diagnostic et de surveillance intégrées encore plus perfectionnées peuvent être mises au point

En 2009-2010, notre équipe est devenue site pilote, en France, pour l'essai d'une nouvelle technologie d'imagerie par fluorescence (MagPix®) qui met fin à la plupart des systèmes de microfluidique ainsi qu'à la technologie laser utilisée en cytométrie de flux. Cette technologie utilise un aimant qui attire des billes magnétiques étalées sur une surface, une technologie mettant aussi des LED à la place des lasers, et une caméra CCD à la place des photomultiplicateurs. Le dispositif est miniaturisé et portable, de paillasse ou de terrain, puisqu'il ne nécessite pas de régulation stricte de la température ambiante. Il supporte un format de dosage à 50 paramètres maximum « 50-plex », soit réduit de moitié par rapport au format du Luminex 200 (100 paramètres) et avec un temps de lecture par échantillon légèrement plus long. Cette plate-forme plus robuste facilite son implémentation dans des milieux non strictement dédiés à la recherche, en particulier en milieu hospitalier ou laboratoire d'analyses médicales. En collaboration avec d'autres équipes (le *Royal Tropical Institute* ou KIT à Amsterdam) et grâce à l'expertise de nos membres en interne, nous avons conçu des dosages plus complexes qui utilisent des SNP et le typage par délétion pour fonctionner sur les plate-formes MagPix et Luminex. Nous avons également développé plusieurs stratégies de biologie moléculaire pour optimiser ces méthodes multiplexées. Une première approche qui permet de faciliter le multiplexage des PCR consiste à utiliser le principe de l'amplification multiplex de sonde nucléique dépendant des ligatures (MLPA, *Multiple Ligation-dependent Probe Amplification*). La MLPA fonctionne très bien en génétique humaine ; elle a donc été utilisée pour concevoir un dosage fonctionnant sur l'appareil MagPix® capable d'identifier simultanément (i) les espèces du MTBC (MTBC contre mycobactéries non tuberculeuses), (ii) les génotypes de résistance aux médicaments et (iii) les génotypes au niveau de la sous-espèce, voire du clade, sur la base des caractéristiques des SNP et des signatures de délétion (Bergval *et al.* 2012). Une seconde solution, pour réaliser ce type de dosages, consiste à utiliser le principe des oligonucléotides à double amorçage (DPO, *Dual-Priming Oligonucleotides*) pour réaliser simultanément (i) des spoligotypages et obtenir un profil de sensibilité génétique à la rifampine sur les mutations de pharmacorésistance les plus fréquentes, ou pour réaliser simultanément (ii) des spoligotypages et obtenir un profil de



Méthodes

sensibilité génétique à la rifampine et à l'isoniazide sur les mutations les plus fréquentes, avec 90 % de sensibilité et 100 % de spécificité (Gomgnimbou *et al.* 2012, 2013). La première méthode utilise une évaluation directe sur les positions 516, 526 et 531 du point chaud de résistance à la rifampicine de 81 pb du gène *rpoB* et une évaluation indirecte sur les autres positions du point chaud à l'aide de sondes moléculaires « flottantes ». La seconde méthode ajoute la détection des mutations de la position 315 du gène *katG* et des positions -7 et -15 du gène *inhA*, qui sont toutes responsables de la résistance à l'isoniazide. Ce dosage permet donc de détecter et surveiller simultanément la transmission de bacilles tuberculeux multirésistants. Ce dernier dosage (TB SPRINT) fait actuellement l'objet d'une évaluation directe sur des échantillons biologiques. La diffusion d'un tel dosage, dans des laboratoires de référence dont le personnel aura été correctement formé, permettra également de modifier la façon dont les études de résistance aux médicaments antituberculeux sont régulièrement réalisées dans les pays en développement où la prévalence de la TB est élevée, et avec une plus grande efficacité et à plus faible coût.

Perspectives du typage CRISPR pour de nouveaux modèles ou de nouvelles technologies

Le spoligotypage pourrait se généraliser rapidement à de nombreux pathogènes, car la description des régions CRISPR polymorphes devient de plus en plus courante et concerne un nombre croissant d'agents infectieux tels que *Corynebacterium diphtheriae*, les lactobacilles ou les streptocoques, ou encore des pathogènes de plantes. Le spoligotypage est largement reconnu comme une technique très utile en épidémiologie moléculaire du MTBC. Il a permis de découvrir et d'approfondir nos connaissances de la diversité génétique et phylogéographique mondiale de la population de *M. tuberculosis*. Associée aux mutations des gènes de résistance aux médicaments, le spoligotypage constitue aujourd'hui une technique de laboratoire unique, pour la surveillance de la propagation de la TB MDR, plus facile à réaliser que le séquençage du génome entier et le séquençage de nouvelle génération (WGS/NGS, *Whole-Genome Sequencing/Next-Generation Sequencing*), et en complémentarité de la technique MLVA. La possibilité de réaliser le spoligotypage sur des cytomètres en flux ou sur des appareils d'imagerie par fluorescence présente un certain nombre d'avantages. Ces systèmes à microbilles sont à haut débit et permettent donc un flux plus rapide de résultats, une meilleure standardisation des dosages, une grande facilité de conception et de mise en œuvre de ces techniques, et pour lesquelles de nouvelles adaptations pour des applications locales sont faciles à mettre en place. Ces systèmes sont économiques et adaptés à une utilisation dans des analyses de routine ou pour la surveillance et le contrôle des maladies infectieuses. De tels systèmes de typage capables de détecter des variations dans la composition en CRISPR dépendent du taux de remplacement et/ou d'acquisition des espaceurs, processus fortement dépendant de l'exposition aux phages, et donc, des conditions environnementales.

Il est trop tôt pour dire si les méthodes fondées sur les CRISPR seront ou non acceptées à grande échelle, à l'heure où le WGS/NGS et la spectrométrie de masse ont tendance à avoir la préférence des microbiologistes, tant dans les laboratoires cliniques que dans les laboratoires de santé publique. Il est

certain que le prix *par échantillon*, la possibilité de disposer de plates-formes technologiques entièrement automatisées et la convivialité des logiciels de traitement des données et de gestion des résultats vont jouer un rôle clé dans la réussite ou l'échec des méthodes fondées sur les CRISPR. Il est néanmoins assez peu probable que des dosages de niche, même ayant bénéficié d'un contrôle qualité, et qui ne doivent être réalisés qu'une ou deux fois par an et dans certaines situations très spécifiques sur ces appareils, débouchent sur des marchés importants, à moins qu'un écosystème adapté, constitué de nombreuses techniques multiplexées soit spécialement mis au point. Au sein de l'équipe IGEPE et des services de génotypage « Beads4Med », nous nous sommes fixés pour mission et pour cap d'assister les laboratoires de santé publique en élargissant progressivement notre palette d'activités et de services, par un développement constant, collaboratif, réaliste, mais ambitieux, de notre offre de produits et de services pour les questions de santé publique à l'échelle européenne et mondiale.

Remerciements

Nous remercions le Centre national de la recherche scientifique (CNRS), l'université Paris-Sud, la fondation Mérieux et la Région Île-de-France pour leur soutien financier, ainsi que l'entreprise Luminex BV (M.F. Topin et M.J. van Gils) pour son assistance technique. Ce travail a également nécessité la collaboration scientifique de plusieurs équipes de recherche, parmi lesquelles nous remercions particulièrement : les Docteurs F.X. Weill et S. Le Hello (Institut Pasteur, Paris), les Docteurs C. Ginevra et S. Jarraud (Centre national de référence des salmonelles, Lyon) et les Docteurs M.L. Boschirolini et A. Brisabois (Anses, Maisons-Alfort, France). Nous adressons également nos remerciements à tous nos partenaires, clients et collaborateurs internationaux pour leur confiance.

Bibliographie

- Bergval I, Sengstake S, Brankova N, Levterova V, Abadia E, Tadumaze N, Bablishvili N, Akhalaia M, Tuin K, Schuitema A, Panaiotov S, Bachiyska E, Kantardjiev T, de Zwaan R, Schurch A, van Soolingen D, van't Hoog A, Cobelens F, Aspindzelashvili R, Sola C, Klatser P, Anthony R. 2012. Combined Species Identification, Genotyping, and Drug Resistance Detection of *Mycobacterium tuberculosis* Cultures by MLPA on a Bead-Based Array. *PLoS One*, 7:e43240.
- Cowan LS, Diem L, Brake MC, Crawford JT. 2004. Transfer of a *Mycobacterium tuberculosis* genotyping method, Spoligotyping, from a reverse line-blot hybridization, membrane-based assay to the Luminex multianalyte profiling system. *J Clin Microbiol*, 42:474-477.
- Dunbar SA. 2006. Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. *Clin Chim Acta*, 363:71-82.
- Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, de Romans S, Lim C, Roux C, Passet V, Diancourt L, Guibourdenche M, Issenhuth-Jeanjean S, Achtman M, Brisse S, Sola C, Weill FX. 2012. CRISPR Typing and Subtyping for Improved Laboratory Surveillance of *Salmonella* Infections. *PLoS One*, 7:e36995.
- Ginevra C, Jacotin N, Diancourt L, Guigon G, Arquilliere R, Meugnier H, Descours G, Vandenesch F, Etienne J, Lina G,



Méthodes

Caro V, Jarraud S. 2012. *Legionella pneumophila* sequence type 1/Paris pulsotype subtyping by spoligotyping. *J Clin Microbiol*, 50:696-701.

Gomgnimbou MK, Abadia E, Zhang J, Refregier G, Panaiotov S, Bachiyka E, Sola C. 2012. "Spoligorifotyping" a DPO-based direct-hybridization assay for TB control on a multianalyte microbead-based hybridization system. *J Clin Microbiol*, 50:3172-3179.

Gomgnimbou MK, Hernandez-Neuta I, Panaiotov S, Bachiyka E, Palomino JC, Martin A, Del Portillo P, Refrégier G, Sola C. 2013. "TB-SPRINT: TUBERCULOSIS-SPOLIGO-RIFAMPIN-ISONIAZID TYPING"; an All-in-One assay technique for surveillance and control of multi-drug resistant tuberculosis on Luminex® devices. *J Clin Microbiol*, 51:3527-3534.

Gomgnimbou MK, Ginevra C, Peron-Cane C, Versapuech M, Refrégier G, Jacotin N, Sola C, Jarraud S. 2014. Validation of a microbead-based format for spoligotyping of *Legionella pneumophila*. *J Clin Microbiol*. Apr 23. [Epub ahead of print].

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and

identification of the gene product. *J Bacteriol*, 169:5429-5433.

Jansen R, van Embden JD, Gaastra W, Schouls LM. 2002. Identification of a novel family of sequence repeats among prokaryotes. *Genomics*, 6:23-33.

Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol*, 35:907-914.

Zhang J, Abadia E, Refregier G, Tafaj S, Boschioli ML, Guillard B, Andreumont A, Ruimy R, Sola C. 2010. *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of 'spoligotyping' with new spacers and a microbead-based hybridization assay. *J Med Microbiol*, 59:285-294.

Zhang J, Heng S, Le Moullec S, Refregier G, Gicquel B, Sola C, Guillard B. 2011. A first assessment of the genetic diversity of *Mycobacterium tuberculosis* complex in Cambodia. *BMC Infect Dis*, 11:42.