



## Point de vue

### Typage épidémiologique génomique des pathogènes : retour sur le Congrès IMMEM-10

Paul M. V. Martin<sup>1</sup> (paul.martin@anses.fr) & Sylvain Brisse<sup>2</sup>

1. Anses, Laboratoire de Lyon, France

2. Institut Pasteur, Paris, France

La 10<sup>e</sup> réunion IMMEM (International Meeting on Microbial Epidemiological Markers) s'est tenue à l'Institut Pasteur à Paris du 2 au 5 octobre 2013. Le contenu des communications scientifiques présentées au meeting est très longuement détaillé dans une récente publication (Brisse *et al.*, 2014). Elle a rassemblé 400 participants venus de 40 pays pour écouter 72 communications orales, voir 190 posters et, bien sur, échanger abondamment entre les sessions, pendant les pauses café, etc.... Jusqu'ici cette réunion, la 10<sup>e</sup> depuis celle de Bruxelles en 1987, ressemble à un congrès scientifique réussi mais classique. Nous allons voir qu'il n'en fut rien et que cette réunion restera probablement comme une marque durable, une *hallmark*, véritable étape charnière dans le développement et l'utilisation des marqueurs épidémiologiques des agents pathogènes en santé publique. Il est apparu qu'il y avait un avant et un après IMMEM-10. Pourquoi ? Si l'on enlève les 2 *welcome addresses*, l'introduction générale sur les défis de la surveillance en santé publique et le magnifique hommage à Mark Achtman et Brian Spratt, au moins 40 communications sur 67 (60%) faisaient référence à au moins un des mots-clés suivant : WGS (Whole Genome Sequencing), NGS/HTS (next generation sequencing, High throughput sequencing), pangenome analysis, genome comparisons, microbial genome (ou whole-genome) analysis, genome-wide. Et plusieurs autres communications faisaient implicitement référence aux séquences complètes des génomes des pathogènes étudiés. La place accordée à l'utilisation des génomes complets en santé publique constitue le véritable tournant que nous voulons souligner ici.

Plusieurs communications s'appuyaient sur le séquençage de dizaines ou de centaines de génomes de la même espèce bactérienne : 957 génomes de *Clostridium difficile*, 237 génomes de STEC (*Escherichia Coli productrices de Shiga toxine*), 25 vancomycine-résistant *Enterococcus faecium*, 111 uropathogenic *E. coli*, etc... Les délégués de Public Health England (PHE) présentaient les premiers résultats portant sur les séquences de 1500 souches de *Salmonella* : 1000 *S. Typhimurium*, *S. Typhi* et les sérovars les plus fréquemment retrouvés en 2012, et 500 souches des « autres sérovars ». La conclusion était, pour *Salmonella*, qu'il n'existait pas une congruence totale entre les bio marqueurs épidémiologiques que sont les sérovars, et les résultats du WGS, confirmant les résultats déjà obtenus par MLST (multi locus sequence typing) (Achtman *et al.*, 2012). Il sera donc nécessaire de repenser complètement les systèmes actuels de « classification » épidémiologique utilisés en santé publique, en inventant de nouvelles nomenclatures. Par ailleurs, Mark Achtman annonce très justement que dans un avenir proche nous devons « oublier nos gels » et que l'épidémiologie génomique remplacera progressivement les méthodes de « *fingerprinting* ». Ces résultats, comme d'autres obtenus sur d'autres espèces

bactériennes, posent une question récurrente: est-ce que l'utilisation systématique de ces nouvelles méthodes doit intégrer les données obtenues depuis de nombreuses décennies avec les méthodes de typage devenues aujourd'hui « traditionnelles » comme les sérotypes, ou les marqueurs MLST, PFGE (pulse field gel electrophoresis) ou MLVA (multiple loci VNTR analysis) ? Même si elles apparaissent aujourd'hui insuffisantes, les méthodes de typage « traditionnelles » ont fait la preuve de leur efficacité comme outils microbiologiques au service de la santé publique.

Il serait donc insatisfaisant, du point de vue de la prise de décision en santé publique, de perdre pour des raisons méthodologiques la correspondance avec les données de typage moléculaire accumulées depuis plus d'un quart de siècle, et les connaissances associées en épidémiologie sur la distribution temporo-spatiale des souches et leur association préférentielle à diverses sources d'infections. Les nouvelles données issues du WGS font rapidement leurs preuves en santé publique, à la lumière d'expériences réelles, lors d'épidémies ou d'événements importants touchant à la dynamique des populations de pathogènes (diffusion d'un clone plus pathogène, d'un plasmide de résistance aux antibiotiques, etc.). Comment concilier le changement de pratique rendu possible et souhaitable par les technologies de séquençage haut débit, sans créer une rupture avec les pratiques antérieures, dommageable pour l'aide à la décision en santé publique ? Il existe des solutions.

Deux communications du meeting ont montré que l'intégration des données de typage « traditionnel » classique peut être réalisée dans l'ère de l'épidémiologie génomique. F.-X. Weill de l'Institut Pasteur à Paris a présenté l'utilisation des marqueurs CRISPR (clustered regularly interspaced palindromic repeats) et leur application à l'épidémiologie des *Salmonella*. Cette méthode relativement nouvelle permet de réaliser en temps réel typage et sous-typage simultané de toutes les *Salmonella* (Fabre *et al.*, 2012). La caractérisation de la variabilité des « spacers » (espaceurs) des CRISPR constitue aujourd'hui une méthode de typage validée pour *Salmonella*. L'étude de 150 souches de *S. Typhimurium* a montré que la microévolution des spacers permettait d'identifier et d'individualiser de nombreux sous-types de ce sérotype majeur. La séquence ou la présence/absence de ces spacers, identifiable par des méthodes classiques d'épidémiologie moléculaire (y compris la méthode CRISPOL utilisant une technologie Luminex) sont deux caractéristiques des souches qui peuvent être aisément extraites de la séquence génomique après WGS.

De la même façon, les données du MLST, base de la nomenclature de référence pour les clones bactériens, peuvent être facilement déduite des données du WGS. Keith Jolley (Univ. Oxford) a présenté le concept d'épidémiologie génomique gène-par-gène et l'outil bioinformatique associé



## Point de vue

BIGSdb, qui étend le concept de la méthode MLST à l'échelle du génome (Maiden *et al*, 2013). Le système permet de créer une base de données de souches bactériennes, pour chaque espèce pathogène, dans laquelle sont stockées les séquences génomiques et les métadonnées associées à chaque souche. Le système BIGSdb comprend également une base de données où sont définis un à un l'ensemble des gènes de l'espèce (pan-génome). Il est alors possible de définir toute combinaison de gènes (appelée schéma) utile au génotypage des souches. Les schémas de génotypage peuvent inclure différents nombres de gènes, par exemple sept comme dans le cas des schémas MLST, ou plusieurs milliers. Cette flexibilité permet de faire varier le degré de discrimination des schémas de typage en fonction des besoins : par exemple, quelques dizaines de gènes peuvent suffire à l'identification de groupes clonaux internationaux, tandis que le pangénome peut être nécessaire pour déchiffrer les événements de transmission durant une épidémie localisée. Le système BIGSdb permet également de définir des schémas correspondant à des groupes de gènes d'intérêt (virulence, résistance). Accessible par son interface web, ce système représente donc un outil simple et rapide pour extraire des séquences génomiques, les informations médicalement importantes. De plus, cet outil et d'autres systèmes équivalents en développement sont conçus pour permettre à chaque communauté de microbiologistes experts d'un pathogène particulier de définir des algorithmes permettant d'établir la correspondance entre les séquences génomiques et les marqueurs épidémiologiques « traditionnels » (Jolley and Maiden, 2010) .

Ces deux exemples montrent que nous sommes aujourd'hui dans une phase de transition et non pas de rupture. Cette

transition permettra d'accompagner les microbiologistes impliqués dans la surveillance des pathogènes et le contrôle des épidémies, dans le changement vers l'épidémiologie moléculaire pangénomique tout en permettant de répondre aux contraintes de santé publique. Ces données nouvelles, issues de séquences complètes d'isolats obtenus lors d'épidémies ou d'autres événements important du point de vue de la santé publique, ouvriront les portes à des connaissances nouvelles sur la circulation des agents pathogènes et l'épidémiologie des maladies qu'ils entraînent.

### Bibliographie

Brisse S, Brehony C, Conceicao T, Cubero M, Glasner C, Le Gouil M, Renvoise A, Sheppard S, Weinert LA. 2014. Microbial molecular markers and epidemiological surveillance in the era of high throughput sequencing: an update from the IMMEM-10 conference. *Res Microbiol* **165**:140-153.

Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, *et al*. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. *PloS One* 2012; 7 :e36995.

Achtman M, Wain J, Weill F-X, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Uesbeck A, Dougan G, Harrison LH, Brisse S, and the *S. enterica* working Group. 2012. Multilocus Sequence Typing as a Replacement for Serotyping in *Salmonella enterica*. *PLoS Pathog* **8**:e1002776.

Maiden MC, van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* **11**:728-736.

Jolley KA, Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**:595.