EuroReference
journal of Reference
anses
French agency for food, environmental and occupational health & safety

Journal No. 12
Summer 2014

Point of view

Genomic epidemiological typing of pathogens: feedback on the IMMEM-10 Conference

Paul M. V. Martin[1] (paul.martin@ANSES.fr) & Sylvain Brisse[2]
1. ANSES, Laboratoire de Lyon, France
2. Institut Pasteur, Paris, France

The 10th International Meeting on Microbial Epidemiological Markers (IMMEM) was held at Institut Pasteur in Paris from 2 to 5 October 2013. The contents of the scientific communications presented at the meeting were described in detail in a recent publication (Brisse *et al.,* 2014). More than 400 participants from 40 different countries attended the meeting, which included 72 oral communications, 190 posters and, of course, multiple opportunities for discussion between sessions and during coffee breaks. The above show that this conference, the 10th since the first event held in Brussels in 1987, has been a successful scientific meeting; but is was not only that: the IMMEM-10 meeting will probably be considered a turning point in the development and use of epidemiological markers for pathogenic agents in public health. It has become clear that there will be a before and an after IMMEM-10. Here are the reasons why.

If we exclude the two welcome addresses, the general introduction on challenges in public health surveillance and the outstanding tribute paid to Mark Achtman and Brian Spratt, at least 40 communications out of 67 (60%) concerned at least one of the following key words: WGS (whole genome sequencing), NGS/HTS (next generation sequencing, high-throughput sequencing), pan-genome analysis, genome comparisons, microbial genome (or whole genome) analysis, or 'genome-wide'. Several other communications implicitly referred to the whole genome sequences of studied pathogens. The place given to the use of whole genomes in public health is a true landmark that we would like to highlight here.

A number of communications discussed the sequencing of dozens or hundreds of genomes of the same bacterial species: 957 genomes of *Clostridium difficile*, 237 genomes of Shiga toxin-producing *Escherichia coli* (STEC), 25 vancomycin-resistant *Enterococcus faecium*, 111 uropathogenic *Escherichia coli*, etc. Delegates from Public Health England (PHE) presented the first results on sequences of 1500 strains of *Salmonella*: 1000 S. Typhimurium, S. Typhi, and the most commonly found serovars in 2012, and 500 strains of other serovars. The conclusion for *Salmonella* was that there is no full congruence between the serovars, the current standard epidemiological biomarkers, and the results of WGS, confirming the results already obtained by MLST (multi-locus sequence typing) (Achtman *et al.,* 2012). It would therefore be necessary to entirely rethink the current epidemiological "classification" systems used in public health, by inventing new nomenclatures. As was put somewhat provocatively by Mark Achtman, in the near future we will have to "forget our gels" and genomic epidemiology will gradually replace "fingerprinting" methods.

These results, like those obtained for other bacterial species, pose a recurring question: should the systematic use of these new methods integrate data obtained over many decades with typing methods that have become "conventional" today, such as serotyping, MLST markers, PFGE (pulsed-field gel electrophoresis) or MLVA (multiple-locus VNTR analysis)? Although they now appear to be insufficient, traditional typing methods have proven their effectiveness as microbiological tools for use in public health.

It would not therefore be desirable, from a public health decision-making standpoint, and for methodological reasons, to loose the correspondence with molecular typing data accumulated over more than a quarter of a century, and the associated epidemiological knowledge on the spatial and temporal distribution of strains and their preferential association with various sources of infection. New data from WGS are rapidly proving their value in public health, on the basis of actual experience, during outbreaks or significant events that affect pathogen population dynamics (spread of a high pathogenicity clone, of a resistance plasmid, etc.). How can we reconcile changing practices made possible and desirable by high-throughput sequencing technologies without creating a rupture with former practices, which would be damaging for decision-making in public health? There are various solutions. Two communications at the meeting showed that classic typing data can still be integrated in the era of genomic epidemiology. F.-X. Weill from the Institut Pasteur in Paris presented the use of CRISPR markers (clustered regularly interspaced palindromic repeats) and their application in *Salmonella* epidemiology. This relatively new method can be used to perform simultaneous typing and subtyping of all *Salmonella* in real time (Fabre *et al.,* 2012). Characterisation of spacer variability of CRISPR markers is today a validated typing method for *Salmonella*. The study of 150 strains of serovar Typhimurium showed that the microevolution of spacers could be used to identify and individualise many subtypes of this major serotype. Sequences or presence/absence of these spacers, identifiable through conventional methods in molecular epidemiology (including the CRISPOL method using Luminex technology), are two characteristics of strains that can easily be extracted from the genome sequence.

In the same way MLST data, used as nomenclature reference for bacterial clones can easily be deduced from genomic sequences. Keith Jolley from the University of Oxford presented the concept of gene-by-gene genomic epidemiology and the bioinformatics tool associated with the Bacterial Isolates Genome Sequence Database (BIGSdb), which extends the

Point of view

concept of the MLST method to the entire genome (Maiden *et al.,* 2013).

The system makes it possible to develop a database of bacterial strains for each pathogenic species, in which the genomic sequences and metadata associated with each strain are stored. The BIGSdb system also contains a database that defines, one by one, all the genes of the species (pan-genome). It is then possible to define any combination of genes, called schemes, useful for strain genotyping. Genotyping schemes can include different numbers of genes, for example 7 genes like in MLST schemes, or several thousands. This flexibility enables the degree of discrimination of typing schemes to be modulated based on specific needs: for example, a few dozen genes may be sufficient to identify international clonal groups, while the pan-genome may be needed to decrypt transmission events during a localised outbreak. The BIGSdb system can also be used to define schemes corresponding to groups of genes of interest (virulence, resistance). Accessible via its web interface, this system is a simple and fast tool for extracting from genomic sequences, medically important data. Moreover, this tool and other equivalent systems under development are designed to enable each community of expert microbiologists on a given pathogen to define algorithms that can be used to establish the correspondence between the genome sequences and traditional epidemiological markers (Jolley and Maiden, 2010).

These two examples show that we have entered a transition period, rather than reaching a breaking point. This transition towards pan-genomic molecular epidemiology will help microbiologists working in the area of surveillance of pathogens and outbreak control, to continue fulfilling public health requirements. It can be expected that data from whole sequences of isolates obtained during outbreaks and other important events in public health will provide new knowledge on the circulation of pathogenic agents and the epidemiology of the diseases they cause.

References

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.