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

Journal No. 12
Summer 2014

Methods

A CRISPR genotyping network in France and Europe: A 5-year experiment for Research, Training and Global Health.

Christophe Sola[1] (christophe.sola@u-psud.fr), Guislaine Refrégier[1] 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.

We created a partnership and laboratory networking initiative between a private industrial company, Luminex Corporation (TX), and a public laboratory, Institut de Génétique et Microbiologie (IGM, UMR8621), for the development of high-throughput microbead-based multiplexed technologies for public health, in relation to the growing interest in polymorphic genetic objects, i.e. CRISPRs (Clustered Regularly Interspersed Short Palindromic Repeats). These loci are known to be useful for surveillance and control of infectious bacterial diseases; we show how an initial focus on a respiratory disease, tuberculosis, led us to an interest in food-borne pathogens and to new perspectives.

Keywords: CRISPR, multiplex, microbead-based hybridization, *Legionella pneumophila, Salmonella enterica, Mycobacterium tuberculosis.*

Context and History of CRISPR research; typing the *Mycobacterium tuberculosiss*

The first characterization of a CRISPR locus was made in *E. coli* by Ishino *et al.* in 1987. At that time, the CRISPR acronym did not exist; it was created in 2002 by Jansen et al. In 2004, a paper by the CDC (*Centers for Disease Control*) in Atlanta described the switch from 2D to 3D DNA chips technology for the tuberculosis "*spoligotyping*" technique, developed in 1997 by Kamerbeek et al. (Cowan *et al.* 2004). This genotyping technique assays the diversity of the CRISPR locus in the *M. tuberculosis* complex (MTBC) and remains one of the first-line genotyping techniques used for molecular epidemiology of tuberculosis. International spoligotyping database projects (SpolDB1 to SpolDB4 and now SITVITWEB[3]) helped provide high visibility for spoligotyping and enabled us to increase our knowledge concerning both the phylogeographical population structure of the MTBC and the evolution of CRISPR loci. There are now around 840 published references with *spoligotyping* as a keyword in PubMed.

MTBC, the agent of human and bovine tuberculosis, harbours one unique and "frozen" CRISPR locus, with a Direct Repeat (DR) of 36 bp with the consensus sequence: 5'-GTCGTCAGACCCAAAACCCCGAGAGGGGACGGAAAC-3'. A total of 94 spacers have been published, although only 68 were found to be specific to the MTBC and 26 were only found in the *M. canettii* ecotype. Almost no genetic variation within the DR repeat was found.

Multiplex-bead-array assays (MBAA) are considered to be a recent innovation, although single-cell analysis can be found in the literature as far back as 1977. MBAA can be used both for immunological and molecular ligands (Dunbar *et al.* 2006). Multiplexing (defined as a quantitative assay of multiple analytes simultaneously in small volumes of material) of up to 500 coloured bead types, each with its own spectral signature, as well as lasers, microfluidics and bioinformatic innovations, provide robust devices and equally robust results. The potential cost- and time-savings when compared with single-plex methods provide a strong incentive for the routine use of these methods both in research and clinical laboratories. Indeed, multiplexing technologies are to genetics what optical fibres are to information technologies: a way to have multiple information channels in a single pipe and to produce and transmit a deluge of data. This amazing process should however of course be complemented by data storage and data analysis processes (e.g. by cloud computing and data-mining systems).

When creating the IGEPE team in 2007, the business development unit of Luminex in the Netherlands (Luminex BV, Oosterhout) helped us establish microbead-based flow cytometry in our research team. The original team actually consisted of only two people, but grew rapidly. No laboratory in Europe had tried to reproduce the transfer of spoligotyping in bead-based format; this first step had been done at the CDC as previously mentioned, and Zhang *et al.* extended in our laboratory the original 43-spacer spoligotyping format to a 68-spacer format, which was shown later to be more efficient for achieving better discrimination in M. bovis and South-East-Asian genotypes of MTBC clinical isolates (Zhang *et al.* 2010, 2011).

With funding from the "Région Ile-de-France" and technical assistance from Luminex BV, we started to provide genotyping services, training, and sales of microbead-based spoligotyping services, and were also able to develop innovative non-catalogue coupled microbead-based assays with the support of a few customers in Europe (the RIVM in the Netherlands) and especially in France (Inserm research team in Montpellier, Bichat Hospital and National Reference Tuberculosis Centre in Paris, ANSES teams). We trained dedicated staff within these teams and looked at new ways of applying microbead-based techniques to public health issues. **Figure 1** shows a brief chronology of method development over the last six years.

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

Methods

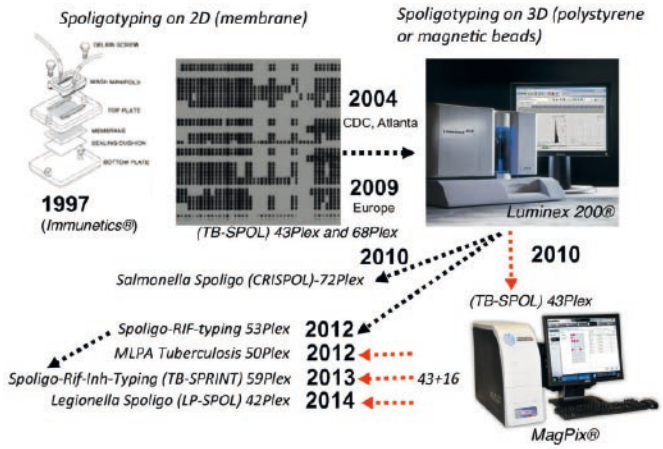

Figure 1. Brief chronology of the development of microbead-based methods on Luminex platforms; since 2010, all methods have been developed on both the Luminex 200® and on the MagPix®. Black arrows: assays developed on the Luminex 200®; red arrows: assays developed on the MagPix®.

The CRISPOL technique: Application of CRISPR technique to *Salmonella enterica*

Among enterobacteriae, *Salmonella* enterica is a major food-borne pathogen and is consequently of great economic interest. For the time being, the identification of thousands of serovars remains the gold standard though it is a tedious and costly method. The initial serovar typing scheme goes back to 1934, but has been regularly updated, most recently in 2007. Even if molecular genotyping methods (PFGE, MLST, MLVA), databases and networks (PulseNet) are well implemented, there is a need for more cost-effective methods. The recent *E. coli* 0104:H4 outbreak in Germany and France also pointed up the need for an increased use of very discriminatory loci such as CRISPR loci when responding to health crises.

An exhaustive analysis of the genetic diversity of CRISPR loci in *Salmonella* was performed thanks to a huge CRISPR loci sequencing operation on the historical collections of the Enterobacteria laboratory at the Pasteur Institute, with the firm commitment of its current director, Dr F.X. Weill, and the support of the genomic and public health departments.

In *Salmonella*, there are two CRISPR loci, CRISPR1 and CRISPR2. The CRISPR1 locus is located downstream from the *iap* gene, whereas CRISPR2 is located upstream from the *ygcF* gene. The DRs of both CRISPR loci are 29 bp long and have the consensus sequence 5-CGGTTTATCCCCGCTGGCGCGGGGAACAC-3. CRISPR analysis by PCR and sequencing of 783 strains belonging to 130 serotypes revealed the presence of 3,800 spacers with a mean size of 32 bp (Fabre *et al.* 2012). The spacer content was found to be correlated with both serotype and multilocus sequence type (MLST). Furthermore, spacer microevolution (duplication, triplication, loss or gain of spacers, presence of SNP variant spacers or VNTR variant spacers) discriminated between subtypes within prevalent serotypes such as Typhimurium (STM), the most prevalent serotype worldwide. In eight genomes and 150 strains of serotype Typhimurium and its monophasic 1,4,[5],12:i:- variant, 57 CRISPR1, 62 CRISPR2 alleles and 83 CRISPR1-CRISPR2 combined alleles were found. Forty unique spacers (including four with variants, such as SNP or VNTR variants) were identified in CRISPR1. Thirty-nine unique spacers (including two with a SNP variant) were identified in CRISPR2. Particular well-characterized populations, such as multidrug-resistant DT104 isolates, African MDR ST313 isolates, and DT2 isolates from pigeons, each had typical CRISPR alleles. Based on this high polymorphism of the spacer contents, a microbead-based liquid hybridization assay, CRISPOL (for CRISPR polymorphism) has been developed. This assay targets 72 spacers identified previously.

The CRISPOL assay was patented by the Institut Pasteur, and implemented in routine both at the Pasteur Institute and at the ANSES laboratory (Fabre *et al.* 2012). It enables the identification of emerging *Salmonella enterica* serovar Typhimurium outbreaks almost in real time, using a new CT (CRISPOL type) clone nomenclature. Our team contributed to the development, implementation and optimization of this assay, and to the optimization of the reagent production process, and is now a supplier of quality-controlled coupled-microbeads to these public health laboratories. **Figure 2** describes the principles and schematizes the CRISPOL technique.

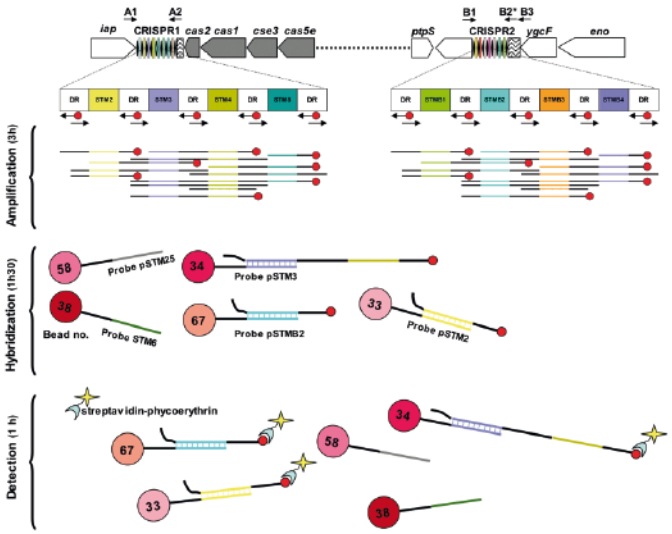

Figure 2. Description of all steps of the CRISPOL technique. Contrary to MTBC, *Salmonella* has two CRISPR loci. A total of 72 spacer probes (including 4 SNP variants) are assessed during this four-five hour method. The technique was implemented in routine analysis and can detect outbreaks in the early stages for a few Euros per assay. (Reproduced from Fabre *et al.* 2012 with authorization.)

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

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

Journal No. 12
Summer 2014

Methods

Legionella pneumophila ST1/pulsotype Paris and the LP-SPOL technique

Legionella pneumophila is a gram-negative facultative intracellular pathogenic bacterium that was identified as the infectious agent of Legionnaire's Disease (LD) or Legionellosis in 1977. Several species of the *Legionella* genus can cause LD, but *L. pneumophila* is responsible for the major outbreaks of LD and for more than ~90% of all identified clinical cases, and *L. pneumophila* serogroup 1 accounts for ~85%. The organism is quite ubiquitous in water environment, whether natural or artificial water networks. Its pathogenicity is related to pulmonary infections, mainly acute pneumonia, which can be fatal in quite a large number of cases (29 fatal in 182 cases in the first reported outbreak).

The reference method for studying the molecular epidemiology of *L. pneumophila* is restriction enzyme analysis using pulse-field gel electrophoresis (PFGE). This method, although laborious, remains a reference even though sequence-based typing (SBT), monoclonal antibody typing and more recently MLVA (Multiple-locus Variable Number of Tandem Repeat Analysis) have been developed. However such methods have limitations in molecular epidemiological investigations due to their lack of power of discrimination of certain strains and they are tedious and slow.

In a previous study, Ginevra *et al.* showed that in certain cases, some indistinguishable strains either by the STB or PFGE methods, particularly within the *L. pneumophila* ST1/Paris pulsotype, could be studied more efficiently using CRISPR, and they developed a membrane-based spoligotyping method. The purpose of our recent collaboration with the National Reference Centre on *Legionella* was to transfer the previously-developed membrane-based spoligotyping technique to the microbead-based format. The *Legionella* direct repeat is a 37 bp repeat with the following sequence:
5'-CCAATAATCCCTCATCTAAAAATCCAACCACTGAAAC-3'.
We recently successfully transferred the membrane-based method to microbeads, both on Luminex 200® and on a Magpix® device and are planning to launch an international multicentric study. The method appears promising for tracking certain specific clones of *L. pneumophila* (Gomgnimbou *et al.* 2014).

More sophisticated and integrated methods for diagnosis and surveillance can be developed

In 2009-2010, our team became the French beta-tester site of a new fluorescence imaging technology (MagPix®) that eliminates most of the microfluidics systems as well as the Laser technology used in flow cytometry. This technology uses a magnet that attracts spreaded paramagnetic beads on a surface, LED technology and a CCD camera, thus providing a portable, bench or field device, with a reduced 50-plex format and a slightly longer read-out time. In collaboration with other teams (the Royal Tropical Institute in Amsterdam or KIT) or by using in-house expertise, we developed more complex assays that use SNPs and deletion typing to work on this new platform, as well as other molecular biology strategies to optimize the multiplexed assays. One first solution to facilitate PCR multiplexing is to use the MLPA (Multiple Ligation-dependent Probe Amplification) principle. MLPA is used very successfully in human genetics and was applied to develop an assay on the MagPix, which simultaneously (i) identifies MTBC species (MTBC versus non-tuberculosis mycobacteria), (ii) identifies drug-resistance genotypes, and (iii) allows sub-species and even clade-level identification of genotypes based on characteristics-SNPs and signatures (Bergval *et al.* 2012). A second solution for performing such assays uses the DPO principle (dual-priming oligonucleotides) to simultaneously perform (i) spoligotyping and obtain a rifampin genetic susceptibility profile of the most frequent drug-resistance mutations or to simultaneously perform (ii) spoligotyping, obtain a rifampin and isoniazid genetic susceptibility profile on the most frequent mutations, with 90% sensitivity and 100% specificity (Gomgnimbou *et al.* 2012, 2013). The first method uses the direct assessment on positions 516, 526 and 531 of the rpoB 81 bp Hot-Spot and indirect assessment on the other positions of the rpoB gene due to "sloppy" molecular probes. The second method adds the detection of katG 315 and inhA -7 and -15 mutations, all mutations responsible for isoniazid resistance. With this assay we can thus simultaneously detect and track multi-drug-resistant tuberculosis bacilli transmission. This last assay (TB-SPRINT) is currently being evaluated directly on biological samples. Such an assay, if disseminated to well-trained reference laboratories, could also change the way drug-resistance surveys are regularly performed in developing countries with high TB prevalence, with greater efficiency and lower costs.

Outlook for CRISPR typing on new models or new technologies

Spoligotyping could soon become a generic method, since polymorphic CRISPR regions are described at an increased pace and for more and more pathogens such as *Corynebacterium diphteriae, Lactobacilli* and *Streptococci*. Spoligotyping is widely accepted internationally as a first-line method for genotyping MTBC for molecular epidemiological studies and for enhancing our understanding of the global phylogeographical structure of the *M. tuberculosis* bacilli population. When associated with drug-resistance gene mutations, it now provides a unique laboratory method for tracking the spread of MDR-TB, easier to analyze than Whole-Genome Sequencing and Next-Generation Sequencing (WGS/NGS). The possibility of running spoligotyping on flow cytometers or fluorescence imaging devices offers a number of advantages. These microbead-based systems are high-throughput, thus allowing faster flows and a better standardization of assays, making it easier to develop and implement these techniques. These systems are cost-effective, and well suited for use in routine analysis and for the surveillance and control of infectious diseases. Such typing systems, which can detect variations in CRISPR content, are dependent on the rate of spacer turnover and/or acquisition, a process that is strongly dependent on phage challenges and hence on environmental conditions.

It is too early to predict whether CRISPR-based methods will gain wide acceptance at a time when WGS/NGS and Mass Spectrometry are tending to become the preferences of clinical and public health microbiologists. It is certain that the price *per sample*, the possibility of fully automating technological platforms, and the user-friendliness of data processing and results management software applications will play a key role in the success or failure of CRISPR-based methods. Quality

Methods

controlled, niche assays (non-Luminex catalogue), that may only be run once or twice a year for very specific purposes on these devices, are unlikely to find a market unless an adequate ecosystem of many multiplexed techniques is specifically developed. This is the mission and vision of the IGEPE team and of "Beads4Med" genotyping services, to serve public health laboratories by expanding our activities through the sustained, partner-based, realistic but ambitious development of our products and services portfolio for European and world-wide public health issues.

Acknowledgments

We are grateful to the Centre national de la recherche scientifique (CNRS), université Paris-Sud, fondation Mérieux, and the région Ile-de-France for financial support, and to the Luminex BV company (M. F. Topin and M. J. van Gils) for technical support. This work also required the scientific collaboration/partnerships of several research teams, among which the following are specifically acknowledged: Dr. F.X. Weill and Dr. S. Le Hello (institut Pasteur, Paris), Dr. C. Ginevra and Dr. S. Jarraud (French National Reference Centre for *Salmonella*, Lyon), Dr. M.L. Boschiroli and Dr. A. Brisabois (ANSES, Maisons-Alfort, France). We also thank all our faithful international partners, customers and collaborators.

References

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, 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, Bachiyska E, Sola C. 2012. "Spoligoriftyping" 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, Bachiyska 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, Boschiroli ML, Guillard B, Andremont 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.