WHOLE GENOME SEQUENCING AS MOLECULAR TYPING METHOD FOR LEGIONELLA PNEUMOPHILA

F. Echahidi[a], A. Meghraoui[b], S. Park[c], O. Soetens[a],D. Piérard[a], S. Roisin[b], M. Montesinos[b] and I. Wybo[a].

National Reference Centre for *Legionella pneumophila*

[a] Vrije Universiteit Brussel (VUB), Universitair Ziekenhuis Brussel (UZ Brussel), Department of Microbiology and Infection control, Laarbeeklaan 101, 1090 Brussels, Belgium.

[b] Laboratoire Hospitalier Universitaire de Bruxelles - University Laboratory of Brussels (LHUB-ULB), Department of Microbiology, Route de Lennik 808, 1070 Brussels, Belgium.

[c]Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA 30329 and Association of Public Health Laboratories (APHL), Silver Spring, MD, 20910, USA

BACKGROUND

Legionella pneumophila (L.pn) is a gram-negative facultative intracellular pathogenic bacterium that can cause Pontiac fever and legionellose. Pontiac fever is a mild disease that can cure spontaneously while legionellose is pneumonia with mortality rate ranging from less than 1% and up to 80% depending on several factors mainly the immune system of the host. The bacterium is found in air conditioning systems, cooling towers, showers, fountains and other devices that make aerosols. People become infected by inhalation of aerosols contaminated by Legionella. Typing of L.pn is until now mainly performed using the gold standard Sequence based typing (SBT) method. A standardized and accurate epidemiological typing method is of high need when an outbreak occurs. The SBT is very discriminatory and reliable; however, when dealing with very frequent sequence types, it is often very difficult to conclude whether environment and patients isolates belonging to the same frequent ST are related. Other methods, like Pulsed-field Gel Electrophoresis (PFGE), are applied to further discriminate such isolates. But it has a limited value because its discriminatory power is also not very high. The recent advances in whole genome sequencing (WGS) for outbreak investigations, both for nosocomial as for community outbreaks of several public health important microorganisms like L.pn are currently making big revolution. In order to implement WGS, as alternative method for SBT in routine use, we evaluated the new available WGS tools for the typing of strains from old outbreaks, small clusters and unrelated isolates

MATERIALS & METHODS

A panel of 77 clinical and environmental L.pn isolates was selected for WGS typing evaluation. Out of these isolates, 7 and 12 (+4 as non-related controls) isolates belonging to 2 important old outbreaks respectively. The remaining isolates belonging to small clusters as well as non-related isolates of very diverse ST types. Genomic DNA was extracted and purified from pure cultures of the isolates using Qiagen dneasy blood & tissue kit (Qiagen). The proper amount of genomic DNA was then used for the library preparation using the Kappa HyperPlus Library Preparation Kit (Kapa Biosystems). Sequencing was performed on the MiSeq sequencing system (Illumina) using the MiSeq Reagent kit v2 (500 cycle): 2 x 250 bp read-length. Assembly of the reads was done with the SPAdes assembly software. Sequencing and assembly were performed at BRIGHTcore WGS platform of VUB-ULB. WGS data were further analysed using Core genome/whole genome MLST (cg/wgMLST) using the available schemes, namely the “Ridom SeqSphere+ software”, “Applied Maths-Bionumerics” software and the 50 loci cgMLST (CDC/ESGLI_ESCMID).

RESULTS

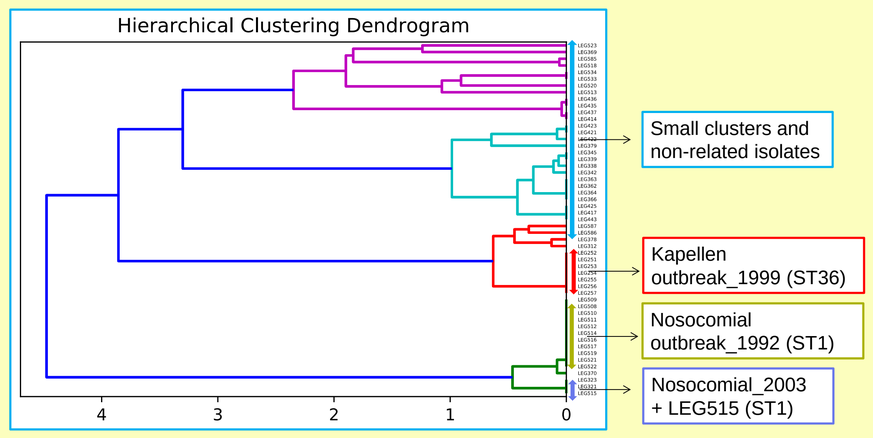

Figure1. Analysis of the available 53 WGS data using the **50 loci L. pn typing scheme**. This scheme is under validation and is based on the following publication: David, Sophia., *et al*. Journal of clinical microbiology, 2016.

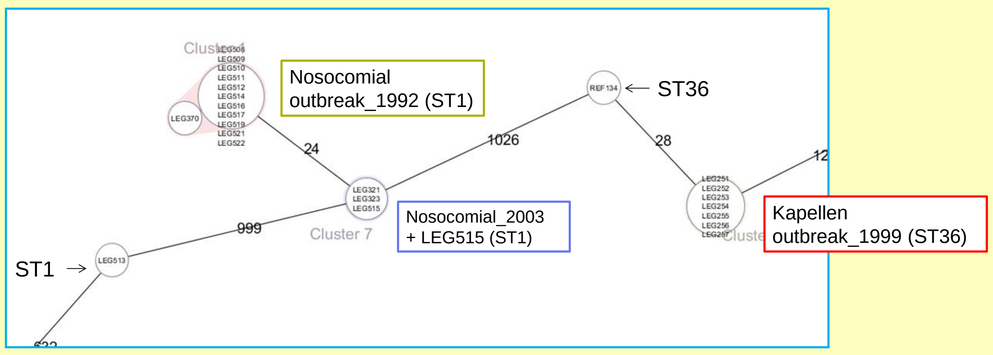

Figure2. Analysis by minimum spanning tree of the available 53 WGS data using **the RIDOM SeqSphere cgMLST L. pn typing scheme** (**1521 core genes).** (Only the part of the MST showing the 2 outbreaks is shown in this figure). This typing scheme is based on the following publication: Moran-Gilad J. *et al*. Surveillance and outbreak reports, 2015.

Main conclusions based on the 2 typing schemes above for the 53 analysed WGS data:

- The known outbreaks and clusters could be detected by both schemes
- One isolate (LEG370) clustered near the nosocomial outbreak_1992 clone, this isolate belongs to ST1, same as for the outbreak. This confirms the compatibility of the new typing scheme with SBT typing.
- One isolate (LEG515) from the outbreak_1992 clustered within another small nosocomial cluster in another hospital but in the same town. This isolate gave originally discrepant results: by an old genotyping method, it was typed within the outbreak clone while MAbs typing classified this isolate outside of the clone. This latter result is now confirmed by the new WGS typing methods.

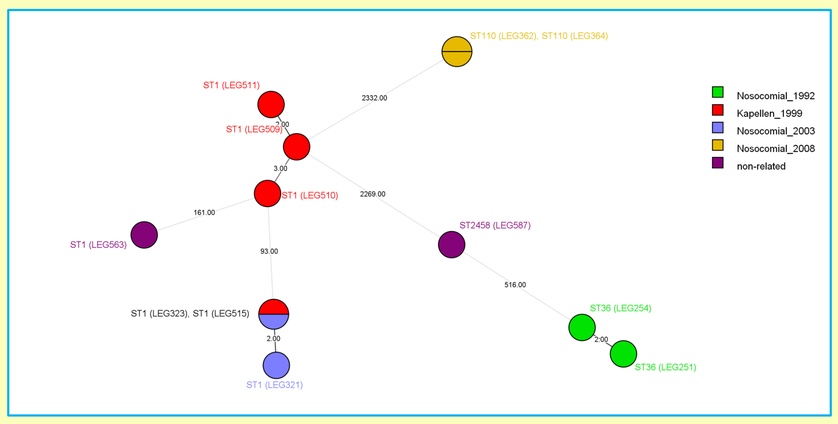

Figure3. Analysis by minimum spanning tree of 12 WGS data by BioNumerics (Applied Maths) wgMLST L. pn typing scheme (**1521 core genes and 4249 accessory genes**) (analysis of remaining data still ongoing).
This typing scheme is also based on the following publication: Moran-Gilad J. et al. Surveillance and outbreak reports, 2015.

Main conclusions based on the BioNumerics typing scheme for the 12 analysed WGS data:

- Also by using this scheme, known outbreaks and clusters could be detected
- Same result obtained for LEG515 as by the other 2 typing schemes (LEG370 not yet analysed)
- Compatibility with SBT maintained and the higher resolution allows for clear discrimination of non-related ST1 isolates (except for LEG515 also here classified by another clone)

CONCLUSIONS

- Based on quality control results at each of the different steps of WGS, optimal results were obtained by using the above mentioned protocol.
- Based on the data interpretation schemes evaluated, concordant results were obtained by Legionella pneumophila cg/wg MLST in comparison to SBT results for the tested isolates.
- Comparable results were also obtained by the different tested interpretations typing tools.
- The availability of public WGS interpretation tools allows for standardisation between different laboratories.
- The higher discriminatory power of WGS in comparison to previously described typing methods will allow for more accurate results and conclusions, especially for environment investigations in outbreak situations.
- The higher throughput of WGS is also very valuable for identifying a potential outbreak source, as several samples are generally taken from a big number of suspected areas.

Part of this work was performed in the frame of Belgian National Reference Centre for *Legionella pneumophila*, supported by the Ministry of Social Affairs through a fund within the National Health Insurance System. We thank our microbiologist colleagues for sending their strains to the NRC.