 Legionellosis, or Legionnaires’ Disease, is a pathology characterized by an acute pneumonia and is fatal in 10 to 15 % of cases in France. The infection is caused by a bacterium from the Legionella genus, usually from the Legionella pneumophila species. Epidemiological data suggest that all strains do not have the same pathogenic potential.

The aim of the project is to address this question: **What are the molecular determinants of the pathogenic potential in legionellosis?**

### Whole genome sequencing

- 16 clinical isolates from deceased patients
- 16 clinical isolates from patients that promptly recovered from the infection

MiSeq (illumina), paired-end reads of 301 bp

Quality control: removal of remaining adaptors and of low quality bases (FastQC, fastq-mcf2)

### Sequence Based Typing

Typing is usually performed by sequencing seven highly conserved genes with the Sanger technology

Need to adapt the method to NGS data

Published tools can not be used here: either they require a whole alleles' database for L. pneumophila that is unavailable for download (SRST2, BigSdb) or it is dedicated to bacterial species not including L. pneumophila (MLST)

We propose a small and fast pipeline based on FASTA that can easily be turned into routine practice in clinical laboratories

### Comparison of assemblers

As genome assembly is still not a simple and fully resolved task, we have tested several software for our data. QUAST and REAPR are then used to evaluate the accuracy of the assemblies, respectively with and without a close reference genome.

#### Results vary upon tools and parameters.

There is no clear winner. The best approach for a strain is not necessary the best one for another strain, but all assemblers give similar results for most strains.

Velvet and Newbler consistently give good results across all strains.

#### Next steps

- Ongoing work: tests of Medusa17 and Pilon18 to fill gaps and reduce the number of scaffolds -> preliminary results look promising
- Annotation of genomes
- Characterization of genetic differences between strains, with respect to the severity of the associated legionellosis -> comparative genomics, SNP

---

References

1. 1) [http://www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
2. 2) [http://code.google.com/p/seq-utils/](http://code.google.com/p/seq-utils/)

Acknowledgements

This work was supported by the Fondation pour la Recherche Médicale, grant number DB20131228568 to Sj, and was performed using the computing facilities of the CC LBBE/PRABI.

**Analysis of genomes from 32 strains of Legionella pneumophila of variable severity to identify determinants of pathogenicity**

Amandine CAMPAN-FOURNIER1,2, Christine OGER2, Christophe GINEVRA3,4, Vincent NAVRATIL2, Anne-Gaëlle RANC3,4, Guy PERRIERE1,2 and Sophie JARRAUD1,4

1 Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, UMR CNRS 5558, 43 bd du 11 novembre 1918, 69622 Villeurbanne cedex, France
2 Pôle Rhône-Alpes de Bio-Informatique, Université Lyon 1, 43 bd du 11 novembre 1918, 69622 Villeurbanne cedex, France
3 Hospices Civils de Lyon, Centre National de Référence des Legionelles, 59 bd Pinel, 69500 Bron, France
4 Centre International de Recherche en Infectiologie, Université Lyon 1, UMR CNRS 5308, Inserm U1111, Ecole Normale Supérieure de Lyon, 46 allée d’Italie, 69364 Lyon cedex 07, France

**http://www.cdc.gov/legionella/index.html**