Implementation of molecular techniques in the diagnosis of legionnaires' disease and in the investigation of legionella outbreaks

Mentasti, M.

IMPLEMENTATION OF MOLECULAR TECHNIQUES IN THE DIAGNOSIS OF LEGIONNAIRES' DISEASE AND IN THE INVESTIGATION OF LEGIONELLA OUTBREAKS.

MASSIMO MENTASTI

A thesis submitted in partial fulfilment of the
requirement of the University of Westminster
for the degree of Doctor of Philosophy

The research programme was carried out
in collaboration with Public Health England

October 2016

"Rien ne se perd, rien ne se crée, tout se transforme"

(Antoine-Laurent de Lavoisier)

Abstract

Background. Legionnaires' Disease (LD) is a mild to severe, potentially lethal, respiratory syndrome caused by members of the Legionella genus, in particular L. *pneumophila* serogroup (sg) 1 alone causes about 95% of culture confirmed cases. The infection is usually acquired by inhalation of aerosols originating from contaminated fresh water sources, consequently typing of both clinical and environmental isolates is crucial to rapidly identify the possible source and prevent further cases. Legionellae are difficult to isolate by culture, moreover as respiratory samples are not available for up to 65% patients, alternative techniques are needed to diagnose LD and maximise the amount of typing data that can be obtained to aid investigations. Urinary antigen detection and serology provide very limited information regarding the infecting strain, while the advent of PCR and Sanger sequencing has allowed reliable diagnostic and typing methods to be introduced.

Objectives. The aim of this study was to improve existing diagnostic and typing molecular assays, and to develop new ones to further standardise diagnostic and typing procedures across members of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) Study Group for Legionella Infections (ESGLI). Utility of the assays was assessed in both routine and outbreak scenarios.

Methods. A wide range of both *in silico* and *in vitro* experiments were used to design and validate specific oligonucleotides to improve detection and typing of *L. pneumophila*. Genomic DNA was manually extracted and prepared for Whole Genome Sequencing (WGS) using Illumina platforms. A bioinformatic approach was used to design a WGS based typing scheme and decipher the evolution of *L. pneumophila* sg1 Sequence Type (ST) 47, a major disease-causing strain.

Results. A real-time PCR detecting *L. pneumophila* and sg1 specific targets was validated with international colleagues and made available to ESGLI members. Sequence based typing was improved and expanded, and specific typing guidelines produced. A 50 gene core-genome MLST was identified as the best approach to improve the current typing method. ST47 was shown to be a 'chimera' between ST109 and ST62, and a specific real-time PCR was designed to detect this strain.

Conclusions. The results of this study allowed researchers to obtain faster and more accurate diagnosis of LD, and *L. pneumophila* typing data from both isolates and primary samples. A metagenomics approach is presently under evaluation to obtain typing results by WGS directly from clinical and environmental samples.

List of contents

Introduction

The *Legionella* genus comprises Gram-negative bacteria widespread in both natural (e.g., streams, lakes, rivers, natural spas, water in hydrothermal areas, and sub-terrestrial groundwater sediments) and man-made (e.g., air-conditioning systems, potable water supplies, ornamental fountains, and plumbing fixtures and fittings in hospitals, shops and homes) freshwater habitats. Legionella strains have also been recovered from moist potting soils, mud, riverbanks, and rainwater run-off (Harrison, 2005). Legionellae can live as free-cells or infect and multiply within protozoan hosts such as *Acanthamoeba*, *Hartmanella*, *Naegleria*, *Valkampfia*, *Echinamoeba*, *Tetrahymena* and *Cyclidium*. Occasionally, they can also infect humans causing either a mild to severe form of pneumonia known as Legionnaires' disease (LD) or a 'flu-like syndrome known as Pontiac Fever. Man-made environments are the source in almost all cases, thus regular maintenance and appropriate disinfection measures are essential to prevent LD (Harrison, 2005).

LD can manifest both as sporadic cases and explosive outbreaks. The last reported UK incidence was 5.2 cases per million inhabitants (331 total cases), compared to an average of 11.4 (0.1 - 37.4) in Europe (ECDC, 2015). Usually patients require antibiotic treatment and hospitalisation, often in an intensive care unit, with the most critical cases needing extra-corporeal membrane oxygenation. The infection is typically acquired by inhalation of aerosols originated from colonised water systems (Phin *et al.*, 2014). So far person-to-person transmission has only been demonstrated in one instance involving two fatal cases in Portugal (Correia *et al.*, 2016). After the first documented incident involving 182 patients with 29 fatalities that occurred in 1976 during the 58th annual meeting of the American Legion (Philadelphia, USA), many other outbreaks have been identified worldwide (Phin *et al.*, 2014). LD is also a well-recognised nosocomial infection characterised by an increased mortality rate with respect to community acquired cases (ca. 30% *vs* 10%) due to the weakened immune response of the infected patients (Phin *et al.*, 2014). Often, LD is acquired during travel, consequently robust and reproducible typing methods are necessary to compare results obtained on strains isolated in different countries in order to identify the infecting source.

Out of 59 different species included in the *Legionella* genus (www.bacterio.net/legionella.html), *Legionella pneumophila* serogroup (sg) 1 strains are responsible for ca. 85% of culture confirmed LD cases in Europe (ECDC, 2015). Indeed, these strains are also the leading cause of LD outbreaks: to date only one documented outbreak has been shown to be caused by a non-sg1 strain of *L. pneumophila* (Sedgwick *et al.*, 2007). Diagnostic methods were originally designed to target sg1 strains (the first to be isolated from patients), thus they often lack sensitivity for other strains and species belonging to the *Legionella* genus. The role of such strains in the epidemiology of LD infections is consequently underestimated.

As first line antibiotics such as β-lactams are not effective against an intracellular pathogen such as *Legionella*, prompt and accurate diagnosis is crucial to provide patients with an appropriate treatment, i.e., macrolides (usually azithromycin) and/or fluoroquinolones (usually levofloxacin) (Pedro-Botet and Yu, 2006). Clinical manifestation and examination by chest X-ray do not allow physicians to clearly differentiate LD from other forms of pneumonia, hence targeted laboratory tests are necessary to obtain a correct diagnosis.

After the Philadelphia outbreak, culture and serology were the only laboratory methods available to diagnose LD. Legionellae require L-cysteine and iron salts to grow, consequently they are not isolated unless dedicated media such as Buffered Charcoal Yeast Extract (BCYE) are used (Harrison, 2005). Typically three to seven days are required to obtain *Legionella* colonies (with negative culture results issued only after 10 days of incubation) and *Legionella* strains are often overgrown by faster growing microorganisms such as *Candida* spp. and/or *Pseudomonas aeruginosa*. Serology requires serum samples to be collected >14 days after onset of symptoms, which may serve as a useful retrospective, but not primary diagnostic method. In the early 1980s, commercial kits for the detection of *L. pneumophila* antigen in urine started to become available, and whilst they proved valuable in making a prompt diagnosis of LD, they are most reliable only when sg1 strains are involved and they provide no typing data to aid outbreak investigation. Nevertheless, with up to 65% of LD patients failing to produce sputum (von Baum *et al.*, 2008), having

instead a so-called “dry cough”, detection of *L. pneumophila* sg1 antigen in urine still remains the main diagnostic method for LD. After the advent of PCR in the mid-late 1980s, infectious diseases were promptly and specifically diagnosed prior the isolation of the causative agent in culture or when organisms could not be isolated. It became then possible to quickly and reliably diagnose LD cases caused by *Legionella* strains other than *L. pneumophila* sg1 and to later design typing methods for *L. pneumophila* based on DNA sequence analysis.

Definitive typing of *Legionella pneumophila* is essential to support or refute an epidemiology link between clinical and environmental isolates; this can play a pivotal role in the decision to launch further environmental investigations in order to determine the possible infection source with the aim of preventing further cases. Typing using monoclonal antibodies (mAb) was one of the first phenotypic methods to be introduced (Helbig *et al.*, 2002), but it offers limited discriminatory power only for *L. pneumophila* sg1 strains. This approach splits sg1 isolates into mAb 3/1 positive (i.e., Philadelphia, Allentown, France, Knoxville and Benidorm subgroups) that are likely to cause community acquired cases and outbreaks, while mAb 3/1negative strains (i.e. Oxford, OLDA, Bellingham, Camperdown and Heysham subgroups) are usually related to hospital acquired cases and are found in immunocompromised patients. Presently Sequence-Based Typing (SBT) is the gold standard for typing of *L. pneumophila* isolates and the pillar of LD outbreak investigation (Gaia *et al.*, 2005; Ratzow *et al.*, 2007). Briefly, the method is a modification of the Multi-Locus Sequence Typing, MLST (Maiden *et al.*, 1998) principle where 4 of the 7 targets (namely, *flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA* and *neuA*) are not housekeeping genes, hence the SBT name instead of MLST. The seven targets are amplified by PCR using specific sets of primers and PCR products are sequenced using the Sanger method. Tracefiles are then analysed using a specific Sequence Quality Tool (SQT) available online on the *L. pneumophila* SBT database (www.hpa-bioinformatics.org.uk/cgi-bin/legionella/sbt/seq_assemble_legionella1.cgi): consensus sequences obtained from forward and reverse primers are trimmed and compared to previously assigned allele numbers and then set into the above pre-determined order to generate a 7-digit allelic profile (e.g. 1,4,3,1,1,1,1) and a Sequence

Type (ST) represented by a number (e.g., ST1). The database of *L. pneumophila* Sequence Types (STs) is hosted by Public Health England (PHE) and contains data submitted by colleagues from worldwide locations. As of 15 July 2016, it holds 10,908 entries comprising 2191 STs from 61 countries (bioinformatics.phe.org.uk/legionella/legionella_sbt/php/sbt_homepage.php). A nested modification (nSBT) increasing the sensitivity of the SBT method is also available to obtain typing data directly from those primary samples where *L. pneumophila* DNA was detected but the organism was not isolated by culture (Ginevra *et al.*, 2008; Mentasti and Fry, 2012). SBT and nSBT have been extensively and successfully used to analyse both LD clusters and outbreaks, however in some instances full allelic profiles cannot be obtained (and consequently a ST cannot be designated), due to strains carrying mutations in the primer binding sites that prevent the amplification of one or more targets, or presence of insufficient target DNA. Also, as common sequence types such as ST1, ST23, ST37, ST42, ST47 and ST62 cause ca. 50% of LD cases, some investigations remain unresolved (David *et al.*, 2016).

The main aim of this study was to improve laboratory methods for analysing samples related to LD cases by designing, validating and implementing state of the art molecular techniques able to rapidly and effectively detect and type *L. pneumophila*. The ability and utility of these techniques to rapidly identify and tackle possible LD outbreaks was assessed.

Development of *L. pneumophila* 3plex qPCR by the European Society of Clinical Microbiology and Infectious Diseases Study Group for Legionella Infections

Although PCR has proven decisive in making diagnosis of many infectious diseases, detection of *L. pneumophila* DNA by this method is still not considered by the European Centre for Disease Control (ECDC) as proven evidence of LD (http://ecdc.europa.eu/en/healthtopics/legionnaires_disease/surveillance/Pages/EU-case-definition.aspx), due to lack of reliable, internationally standardised methods. Following the formation of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Study Group for Legionella Infections (ESGLI)

in 2012, the design and validation of a specific real-time PCR was identified as a priority to standardise molecular diagnosis of *L. pneumophila* infections across Europe and then approach the ECDC to consider *L. pneumophila* PCR positive patients as proven cases.

A survey of ESGLI laboratories in November 2012 identified ABI 7500 (ThermoFisher Scientific), LightCycler 480II (Roche) and RotorGene-Q (Qiagen) as the three real-time PCR platforms mostly used by members (23 out of 30), and it was subsequently decided to validate an in-house assay detecting both *L. pneumophila* and sg1 markers that could be widely and easily implemented across Europe. The National Reference Laboratory at PHE Colindale was given the task to develop the assay. A subgroup comprising members of five countries was formed, led and coordinated by the PhD candidate.

A qPCR method targeting the macrophage infectivity potentiator gene (*mip*) of *L. pneumophila* was already routinely used at PHE Colindale on the LightCycler 2.0 (Roche) since 2008 (Mentasti *et al.*, 2012), however this method was based on dual hybridisation probes which are less commonly used than other types, such as hydrolysis probes, due to the fact that they are less readily available, more complicated to design and more expensive; furthermore, the LightCycler 2.0 only allows simultaneous detection of two targets, thus limiting the options for inclusion of multiple targets and an internal control. In 2011, French colleagues at the National Legionella Reference Laboratory (Lyon) and the Pasteur Institute (Paris), used a bioinformatic approach to identify the O-antigen ABC transporter permease (*wzm*) as the molecular marker for sg1 strains of *L. pneumophila*. Subsequently they designed and validated a real-time PCR using hydrolysis probes to reliably detect such strains (Mérault *et al.*, 2011). It was therefore decided to adopt a rational approach to design a new *mip* assay based on hydrolysis probes and to combine this with the *wzm* assay. An internal control assay targeting the green fluorescent protein gene (*gfp*) (Murphy *et al.*, 2007) was also added to check for inhibition where *mip* and *wzm* amplification were negative.

A 402 bp fragment internal to the *mip* gene is the fourth target of the *L. pneumophila* SBT protocol, consequently sequence data for this fragment are available for all the strains present in the *L. pneumophila* SBT database; 9423

sequences for a total of 59 *mip* alleles were available at the beginning of the study (October, 2012) and had risen to 74 by the end of it (February, 2015). Allele sequences were downloaded from the database and aligned using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2). Oligonucleotides were designed using Primer3 (http://primer3plus.com/web_3.0.0/primer3web_input.htm) in conserved regions allowing a maximum of one mismatch in either the primer or probe binding sites. A validation protocol following criteria published by Saunders *et al.* (2013) was written by PHE Colindale and distributed to ESGLI colleagues to perform a multicentre validation. A large panel of reference strains (251), external quality assessment (EQA) samples (100) and clinical samples (132) was included to complete the validation. A full validation report was later published (Mentasti *et al.*, 2015) under the auspices of ESGLI and a detailed protocol for each of the three platforms was made available to ESGLI members with the intent to promote the use of this assay.

The *L. pneumophila* 3plex qPCR has been routinely used by PHE Colindale to diagnose and confirm LD cases since 2013. This method has proven effective and specific in outbreak scenarios where rapid detection and differentiation of *L. pneumophila* sg1 DNA from both clinical and environmental samples is imperative to provide timely results to the Incident Management Team (IMT). Since 2012, this method has been used in at least seven instances where urgent results were specifically requested by the IMT to aid their actions. In all cases the assay proved reliable and informative (Mentasti *et al.*, 2016).

Improvement of the SBT protocol

As LD is often travel-associated (i.e., clinical and environmental strains isolated in different countries), when typing *L. pneumophila*, reproducible and comparable results must be obtained. As the ECDC official collaborating centre for legionellosis from 1998 to 2013, PHE Colindale standardised the *L. pneumophila* SBT method among Legionella National Reference Laboratories in Europe (and beyond) by providing training and guidelines (Mentasti and Fry, 2012) based on previously published papers, and improving them with *ad hoc* experiments. Primer sequences were modified to obtain improved sequencing

results. Primers were re-designed to bind SBT targets externally to the originally published primers to improve the sequencing results at beginning and end of the target, and obtain fully double-stranded sequences. M13 forward (5' - GCGTATTGCTCAAAATACTG - 3') and M13 reverse (5' - GGTATCACCTGCGGTTCCA - 3') tags were attached to the amplification primers so that fewer primers were required when sequencing the 7 PCR products (Mentasti *et al.*, 2012). SBT results of Legionella reference laboratories in Europe and beyond were analysed yearly using a *L. pneumophila* SBT EQA scheme run by PHE under the auspices of ECDC. A total of 10 SBT EQA panels were distributed between 2003 and 2014 with participants raising from 16 to 38. A brief summary of results in detailed in Table 1.

After the SBT scheme was completed, it became evident that the amplification of *neuA*, the seventh SBT target was not possible in some strains of *L. pneumophila*, consequently a full 7-allele profile could not be obtained or a Sequence Type designated. This issue had already been described for other pathogens such as *Haemophilus influenzae* where the deletion of an operon encoding a *fuculokinase*, which is part of the seven-locus sequence typing scheme (Riddenberg *et al.*, 2010), caused some strains to be untypable. A study in collaboration with the German Legionella Reference laboratory in Dresden (Farhat *et al.*, 2011) showed that in those *L. pneumophila* strains a variant (designated *neuAh*) with about 70% sequence homology to *neuA* is present, thus *neuA* primers are not able to bind it and the target is not amplified. New primers (*neuA*_up and *neuA*_do2) were designed with the same binding sites of the original primers to amplify both *neuA* and *neuAh*, however due to sequence variation, they contained several ambiguities, the resulting sequencing results were of suboptimal quality and did not pass the strict criteria of the *L. pneumophila* SQT. New primers (*neuAh*_L and *neuAh*_R) were later designed to specifically amplify and sequence *neuAh* with QV20+ results (i.e., the total number of bases in the entire trace that have basecaller quality value of ≥20) of ca. 800 when *neuAh*_L and *neuAh*_R were used, whilst values of ca. 400 and 250, respectively were obtained, when *neuA*_up and *neuA*_do2 were used. Considering that 354 bp is the length of the fragment used to determine the allele type for the *neuAh* target, a QV20+ result of ca. 250 for the reverse

sequence is well below what is necessary to obtain a good quality double-stranded consensus sequence across the entire allelic region. The new *neuAh* primer sequences were distributed to all international laboratories taking part in the *L. pneumophila* SBT EQA. Results obtained worldwide on a panel of five selected strains showed that the new primers were reliable in both amplifying and characterising the *neuAh* target. Design and international validation of these primers together with a full characterisation of the *neuAh* target were later published (Mentasti *et al.*, 2013). A legionella outbreak caused by one such strain occurred in 2005 on a cruise ship in northern Europe. As part of the validation of the new primers it was possible to retrospectively complete the 7 allele profile of the infecting strain which was characterised as *L. pneumophila* sg5 ST1327 (11,14,16,31,15,13,210). Legionella Reference laboratories worldwide adopted the *neuAh* primers and where possible completed the allelic profile of those *L. pneumophila* strains previously uploaded to the SBT database without the seventh target. This allowed the database curators to later assign an ST number to them. By mid-2013 of the about 1600 or so extant STs, 100 (about 6%) had been allocated thanks to the fact that the *neuA/neuAh* issue was resolved. A total of 29 *neuAh* alleles distributed in 224 different STs (478 entries), has been so far submitted to the SBT database (accessed on 15/07/2016).

The above SBT modifications and guidelines were adopted to investigate the molecular epidemiology of *L. pneumophila* in Israel and published in collaboration with colleagues at Jerusalem University Hospital and Israeli Ministry of Health (Moran-Gilad *et al.*, 2014). This study was the first of its kind published from a Middle Eastern country. Results showed many similarities to the epidemiological trends found in European countries, especially in relation to incidence, seasonality and methods used for laboratory diagnosis. Differences were also seen, including a larger proportion of nosocomial cases than that reported in many countries. Other unusual features included a surprisingly high percentage (i.e., 43%) of *L. pneumophila* sg1 ‘OLDA/Oxford’ ST1 (1,4,3,1,1,1,1) strains compared to other counties (e.g., 4.8% in the UK) resulting in reduced clinical strain diversity; a notable number of clinical cases involving *L. pneumophila* sg1 ‘Allentown/France ST40 (3,6,1,14,14,9,11) and the presence of the rare *L. pneumophila* sg3 ST338 (2,10,15,28,9,4,13) strain among

environmental isolates. A substantial percentage of typed strains yielded novel STs which were added to the *L. pneumophila* SBT database.

A fatal LD case in an Israeli infant was also analysed with the above protocols and published in in collaboration with Israeli Public Health colleagues (Moran-Gilad *et al*., 2012). Although this was not the first described case of LD in an infant (Unit for Surveillance and Control of Communicable Diseases, 2009) it was the first to fully characterise a clinical and epidemiologically linked environmental strain (*L. pneumophila* sg1 Oxford/Olda ST1) in a paediatric LD case, and to demonstrate humidifiers (filled with non-sterile tap water) as a source of LD. Further to these results, the Israeli Public Health Authorities approached the Standards Institution of Israel to set up new regulations mandating hazard labelling of cold water humidifiers and package inserts were included in user manuals with the aim to prevent further LD cases associated with the use of tap water in humidifiers.

Rapid investigation of LD

Using traditional culture methods, at least six days are necessary to isolate and fully type *L. pneumophila* strains. In some cases *L. pneumophila* cannot be isolated due to overgrowth of contaminant flora or due to the use of appropriate antibiotic therapy (clinical sample) or disinfectant (environmental samples) prior to specimen collection. Whilst these limitations might be tolerated for sporadic LD cases bearing little risk to the wider population, when a possible outbreak is suspected the use of rapid molecular tests could be decisive in identifying the possible source and so enabling prompt public health measures to prevent more LD cases from occurring.

In preparation for the 2012 Olympics held in London (UK), legionella together with gastrointestinal pathogens and ‘flu were declared as a top public health threats. PHE Colindale was requested to devise a rapid response strategy (RRS) should an LD outbreak have occurred during the Olympics. The adopted approach was the rapid detection of *L. pneumophila* and sg1 DNA on both clinical and environmental samples by using the ESGLI 3plex PCR (Mentasti *et*

al., 2015) followed by nested SBT (Ginevra *et al.*, 2009; Mentasti and Fry, 2012) applied to DNA extracts PCR-positive for *L. pneumophila* DNA. The efficacy and reliability of molecular methods in detecting and typing *L. pneumophila* was previously retrospectively demonstrated (Mentasti *et al.*, 2012). Briefly, in a total of 227 respiratory samples obtained from legionella urinary antigen positive patients, *L. pneumophila* was successfully cultured from 147 (64.8%), while *L. pneumophila* DNA was detected from 194 (85.5%); moreover, partial to full typing results were obtained on 43/47 (91%) PCR-positive culture-negative samples. Increased/improved typing results can be obtained retrospectively on DNA extracts following culture-negative results, indeed such results could instead be rapidly obtained (if required) while waiting for culture results.

The RRS was first used to investigate an outbreak that occurred in Stoke-on-Trent (Coetzee *et al.*, 2012) in summer 2012. The initial typing results, obtained in just 17.5 hours (samples received at 17:00 on 24th July and results reported to the IMT at 10:30 on 25th July), were sufficient to identify the LD cluster as a definite point-source outbreak. Where *L. pneumophila* was cultured from clinical samples the preliminary results were all confirmed and the infecting strain characterised as *L. pneumophila* sg1 'Benidorm' ST1268 (2,9,2,5,3,17,15). PCR screening of environmental samples, identified three specimens positive for *L. pneumophila* and sg1 DNA. Nested SBT yielded a high-quality unambiguous ST1268 profile in a swab from a hot tub present in retail premises, whilst partial profiles, namely 0,0,0,15,18,1,6 and 8,0,0,15,18,1,6, consistent with ST62 (8,10,3,15,18,1,6) were obtained from water concentrates collected from cooling towers located in industrial premises. The hot tub was identified as the outbreak source, however it was not possible to isolate *L. pneumophila* from the swab sample due to high counts of *Pseudomonas aeruginosa*. *Legionella pneumophila* sg1 'Allentown/France' ST62 was instead isolated from the water concentrates. Significantly, ST62 is the fourth main cause of LD cases, however due to the availability of these rapid typing results the Health and Safety Executive visited the premises where the cooling towers were located and implemented urgent disinfection measures. No cases due to ST62 strains were identified in the area.

Due to the success during the Stoke-on-Trent investigation, the RRS has since been used by PHE Colindale to investigate urgent LD cases and possible outbreaks. So far it has rapidly delivered crucial epidemiological data in seven significant LD incidents: initial typing results were available in 1.58 day (SD 1.01) while results were obtained in 9.53 (SD 3.73) days when standard culture and typing protocols were used (Mentasti *et al.*, 2016).

Rapid detection and evolution of ST47

For reasons yet to be understood, *L. pneumophila* sg1 ST47 (5,10,22,15,6,2,6) is the leading cause of LD in Belgium (Vekens *et al.*, 2012), England & Wales (Harrison *et al.*, 2009), France (Ginevra *et al.*, 2008) and The Netherlands (Den Boer *et al.*, 2008). Despite more than 2000 different *L. pneumophila* STs present in the SBT database, ST47 alone causes more than 20-25% of community acquired LD cases, but it is rarely isolated from the environment when investigations following LD incidents are carried out.

With the aim of designing a specific test to improve detection of ST47 in the environment and tracing the evolution of ST47, the genome of a total of 129 ST47 isolates was sequenced and analysed in collaboration with colleagues at the Wellcome Trust Sanger Institute (Cambridge, UK), the French Legionella Reference Laboratory in Lyon (France) and the Pasteur Institute (Paris, France). WGS analysis revealed that the ST47 genomes (3.47Mb) contain a plasmid of about 150kb, interestingly no recombination was found and a maximum of 33 SNPs were identified between any pair of strains. Comparison of these genomes to that of the 32 STs representative of the main *L. pneumophila* clades (Underwood *et al.*, 2013), revealed 64 possible ST47 specific Open Reading Frames (ORFs). Among these, the ones present in the ST47 plasmid and the ones that generated positive hits when compared using BLASTN against the NCBI database where excluded. Primers were designed to amplify fragments from the remaining 21 ORFs, analysed by BLASTN against the NCBI database and then tested on a subset of reference, clinical and environmental strains present in the collections of the Legionella reference laboratories of England & Wales and France. LPO_1073 was identified as the target with highest specificity. PFAM (http://pfam.xfam.org) was not able to

assign the translated sequence to any so far described protein family and the only positive match on the NCBI database is the “Lorraine” strain of *L. pneumophila*, which is the reference ST47 strain. A hydrolysis probe was designed and added to the primers to set up and validate a real time PCR assay. Interestingly, the only positive results apart from ST47 isolates, were 11 ST109 (5,1,22,15,6,10,6) isolates from France, while 8 ST109 isolates from England & Wales gave no reaction. When the remaining 25 ST109 isolates present on the *L. pneumophila* SBT database were tested, 26 isolates in total (out of 44) resulted in positive results. WGS analysis revealed that “France” ST109 had acquired a 16,266 bp mobile element, probably a phage, (absent in the genome of “England & Wales” ST109) carrying LPO_1073. When sequenced genomes of closely related STs, namely ST74 (5,1,22,30,6,10,6) and ST82 (5,1,22,10,6,10,6) were added to the WGS analysis, it became evident that a gradual acquisition by homologous recombination of two large fragments, 170,636 bp and 9,326 bp respectively, of the ST62 genome led to the formation of ST47 (Fig. 1) indicating that these strains shared a common niche allowing recombination to occur. Significantly, only 44 ST109 isolates are present in the SBT database in contrast to 612 ST47 isolates: the acquisition by “France” ST109 of a large portion (13.5 %) of ST62 genome, a well-known cause of LD, undoubtedly generated a strain with increased ability to cause disease with respect to its ancestor. Analysis of the recombined regions revealed that they contain core genes (including 4 of the 7 SBT targets) thus a clear explanation of this increased incidence was not found (Mentasti *et al.*, submitted).

PHE “firepower” against LD

A major aim of the PHE Legionella reference laboratory is to generate typing data to help describe the epidemiology of LD in England & Wales, and to rapidly respond to clusters of LD cases to determine whether a point-source outbreak is possibly occurring. Environmental investigations are then triggered when required to identify the infection source and prevent further cases.

The LPO_1073 real-time PCR was added to the rapid LD investigation approach using *L. pneumophila* 3plex PCR and direct nSBT. The added value

of this assay derives from data gathered from LD epidemiology: as ST47 is the leading cause of sporadic LD, molecular detection of ST47 when investigating clusters can be informative of increased LD incidence and exclude outbreaks.

The RRS plus LPO_1073 PCR was used for the first time to rapidly investigate a cluster of 4 LD cases that occurred during the same week in the summer of 2013 in South Wales. As LD is a relatively uncommon type of infection, identification of four cases in a short period of time in patients with a possible epidemiological link was highly indicative of a point-source outbreak, consequently the PHE full molecular approach was deployed (Table 2). The same day (Day 0) that respiratory specimens were received, *L. pneumophila* sg1 DNA was detected by qPCR in all 4 patients and 2 were also positive with the LPO_1073 assay. This result indicated that a point-source outbreak was highly unlikely as patients were likely to have been infected by different strains of *L. pneumophila* sg1. Direct nested SBT results were available on Day 1 and showed the 2 LPO_1073 PCR positive patients respectively were infected by an ST47 strain and by a strain with a partial allelic profile (5,0,22,15,6,2,6) consistent with ST47. The other two patients were infected by strains with allelic profiles distinct from each other and from ST47. *Legionella pneumophila* was isolated from 3 of the 4 patients and SBT results available on Day 8 confirmed the LPO_1073 and direct nSBT preliminary results (Table 2).

Whole Genome Sequencing (WGS) typing approach

Despite SBT being the current gold-standard for typing of *L. pneumophila*, a few STs cause the majority of infections, consequently some investigations remain unresolved. Furthermore, some strains carry mutations in the SBT primer binding sites; thus they cannot be fully typed.

The development of Next Generation Sequencing (NGS) platforms has progressively improved access to Whole Genome Sequencing (WGS) and allowed a wider use of this technique. The application of four WGS-based approaches, namely Single Nucleotide polymorphism (SNP) analysis, extended Multi-Locus Sequence Typing (MLST), gene presence/absence and a Kmer-based method comparing the kmer content (18-mers) between all *de novo*

assemblies in order to determine their similarity to each other (https://github.com/phe-bioinformatics/kmerid), was evaluated for the epidemiological typing of *L. pneumophila* using criteria published by the ESCMID Study Group for Epidemiological Markers (ESGEM) (Van Belkum *et al.*, 2007) and compared to results obtained by SBT. Briefly, the criteria are based on the assessment of typability (proportion of isolates that can be assigned to a type using a particular method), reproducibility (proportion of pairs of sequencing replicates assigned to the same type), epidemiological concordance (proportion of epidemiologically related set of isolates assigned to the same type), discriminatory power (proportion of epidemiologically unrelated set of isolates assigned to different types) and stability (ability to assign a particular strain to a same type after different culture procedures and conditions).

Paired-end libraries were prepared on DNA samples extracted at PHE, while sequencing was performed with Illumina HiSeq at Wellcome Trust Sanger Institute (UK) on paired-end reads of 100 bases. Where a closely related genome (KmerID > 90%) was already available, genomes were assembled using an in-house mapping pipeline, otherwise *de novo* assembly was applied (David *et al.*, 2016).

In a total of 370 genomes (45 distinct STs), all methods produced a higher index of discrimination (from D = 0.972 using ribosomal MLST to D = 0.999 using SNP analysis) with respect to SBT (D = 0.940). All methods were able to differentiate epidemiologically unrelated isolates of the same ST into further types, while SNP analysis separated almost all isolates into different types. Highly discriminatory methods (such as SNP analysis) produce poor epidemiological concordance, generating virtually a different type for each analysed isolate. The best compromise was identified in a core genome (cg) MLST approach using 50 genes (D = 0.991): this method has a lower discriminatory power than SNP analysis but maintains good epidemiological concordance.

Unfortunately, the short reads on which Illumina technology is based produce results that do not allow result compatibility with SBT: *mompS*, the fifth SBT target, is present in multiple copies (usually two) in the genome of *L.*

pneumophila. In those strains where the *mompS* copies are not identical, the allele type of the SBT copy cannot be determined by Illumina WGS analysis, consequently Sanger sequencing will still be necessary in some instances. NGS platforms generating longer reads (e.g. PacBio or MinION) are available, however they do not allow high-throughput analysis, thus costs are much higher and not suitable for routine use in a public health environment.

A WGS working group was established within ESGLI during the meeting held in London in September 2015 and is presently discussing the results of this comparative study to set up a new gold-standard method for typing *of L. pneumophila* isolates.

Future perspectives

As discussed, direct nSBT is presently the fastest way to obtain typing results when investigating outbreaks, however the level of data that can be obtained is limited to the seven SBT targets. Hence, no more than a ST (if full allele profiles are obtained) can be resolved. Studies are presently addressed at exploiting the full potential of WGS using metagenomics; this approach applies a suite of genomic technologies and bioinformatics tools to directly access the genetic content of entire communities of (micro-)organisms without the need for cultivation.

PHE and Imperial College have recently set up a collaborative study to design and validate a metagenomics protocol to be used during LD outbreak investigations. WGS would be applied directly to DNA extracted from clinical samples to allow scientists to obtain a greater amount of data about the infecting strain than just the ST, with the aim of promptly identifying strain-specific PCR target(s) and designing specific primers. Environmental specimens could be then screened for presence of the specific target(s) by SYBR-green real-time PCR to quickly identify the infection source and apply appropriate disinfection measures to prevent further LD cases.

Acknowledgments

The candidate would like to thank Dr Pamela Greenwell for six years of unconditional guidance and support through his MSc and PhD journeys.

The candidate would also like to acknowledge internationally recognised Legionella experts working at Regional and National Reference Laboratories in Europe and beyond. A particular “thank you” to Dr Tim Harrison; Dr Norman Fry and Dr Vicki Chalker, Public Health England, UK; Dr Christian Lück, Dresden University, Germany; Dr Natalia A. Kozak-Muiznieks, CDC Atlanta, USA; Prof Darja Kese, Ljubljana University, Slovenia; Dr Jakub Mrazek, National Public Health Laboratory (Czech Republic); Prof Sophie Jarraud, Lyon University, France; Prof Jacob Moran-Gilad, Ministry of Health (Israel); Dr Fedoua Echahidi, Universitair Ziekenhuis Brussel (Belgium); Dr Soren Uldum, Staten Serum Institute (Denmark); Ms Sophia David and Prof Julian Parkhill, Wellcome Trust Sanger Institute (UK).

Preliminary results of this project were presented at Legionella European conferences in Paris (2009), Copenhagen (2010), Vienna (2011), Dresden (2012), Athens (2013) and London (2015), and at the International Legionella Conference in Melbourne (2013).

Author’s declaration

I declare that all the material contained in this thesis is my own work.

REFERENCES

Coetzee, N., Duggal, H., Hawker, J., Ibbotson, S., Harrison, T.G., Phin, N., Laza-Stanca, V., Johnston, R., Iqbal, Z., Rehman, Y., Knapper, E., Robinson, S. & Aigbogun, N. (2012). An outbreak of Legionnaires' disease associated with a display spa pool in retail premises, Stoke-on-Trent, United Kingdom, July 2012. *EuroSurveill* **17**, pii=20271. Available from: www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20271

Correia, A.M., Ferreira, J.S., Borges, V., Nunes, A., Gomes, B., Capucho, R., Gonçalves, J., Antunes, D.M., Almeida, S., Mendes, A., Guerreiro, M., Sampaio, D.A., Vieira, L., Machado, J., Simões, M.J., Gonçalves, P., Gomes, J.P. (2016). Probable person-to-person transmission of Legionnaires' Disease. *N Engl J Med* **374**, 497-498.

David, S., Mentasti, M., Tewolde, R., Aslett, M., Harris, S.R., Afshar, B., Underwood, A., Fry, N.K., Parkhill, J., Harrison, T.G. (2016). Evaluation of an optimal epidemiologic typing scheme for *Legionella pneumophila* with whole genome sequence data using validation guidelines. *J Clin Microbiol* pii: JCM.00432-16. [Epub ahead of print]

Den Boer, J.W., Bruin, J.P., Verhoef, L.P.B., Van der Zwaluw, K., Jansen, R., Yzerman, E.P.F. (2008). Genotypic comparison of clinical *Legionella* isolates and patient-related environmental isolates in The Netherlands, 2002-2006. *Clin Microbiol Infect* **14**, 459–466.

European Centre for Disease Prevention and Control. (2015). Legionnaires' disease in Europe, 2013 Surveillance Report. Available online at: http://ecdc.europa.eu/en/publications/Publications/legionnaires-disease-2015.pdf.

Farhat C, Mentasti M, Jacobs E, Fry NK, Lück C. (2011). The *N*-acylneuraminate cytidyltransferase gene, *neuA*, is heterogenous in *Legionella pneumophila* strains but can be used as a marker for epidemiological typing in

the consensus sequence-based typing scheme. *J Clin Microbiol* **49**, 4052 - 4058.

Gaia, V., Fry, N.K., Afshar, B., Lück, P.C., Meugnier, H., Etienne, J., Peduzzi, R. & Harrison, T.G. (2005). Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. *J Clin Microbiol* **43**, 2047-2052.

Ginevra, C., Lopez, M., Forey, F., Reyrolle, M., Meugnier, H., Vandenesch, F., Etienne, J., Jarraud, S. & Molmeret, M. (2009). Evaluation of a nested-PCR-derived sequence-based typing method applied directly to respiratory samples from patients with Legionnaires' disease. *J Clin Microbiol* **47**, 981-987.

Ginevra, C., Forey, F., Campèse, C., Reyrolle, M., Che, D., Etienne, J., Jarraud, S. (2008). Lorraine strain of *Legionella pneumophila* serogroup 1, France. *Emerg Infect Dis* **14**, 673–675.

Harrison, T.G. (2005). Legionella. In: Borriello SP, Murrray PR, Funke G, eds. Topley & Wilson's Microbiology and Microbial Infections. London: Hodder Arnold, 1761-1785.

Helbig, J.H., Bernander, S., Castellani Pastoris, M., Etienne, J., Gaia, V., Lauwers, S., Lindsay, D., Lück, P.C., Marques, T., Mentula, S., Peeters, M.F., Pelaz, C., Struelens, M., Uldum, S.A., Wewalka, G. & Harrison, T.G. (2002). Pan-European study on culture-proven Legionnaires' disease: distribution of *Legionella pneumophila* serogroups and monoclonal subgroups. *Eur J Clin Microbiol Infect Dis* **21**, 710–716.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A,. Feavers, I.M., Achtman, M., Spratt, B.G. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci* **95**, 3140 - 3145.

Mentasti, M. and Fry N.K. (2012). Nested Sequence-Based Typing (NSBT) protocol for epidemiological typing of *Legionella pneumophila* directly from clinical samples, v. 2.0. Available from: http://bioinformatics.phe.org.uk/legionella/legionella_sbt/php/sbt_homepage.php

Mentasti, M., Fry, N.K., Afshar, B., Palepou-Foxley, C., Naik, F.C. & Harrison, T.G. (2012). Application of *Legionella pneumophila*-specific quantitative real-time PCR combined with direct amplification and sequence-based typing in the diagnosis and epidemiological investigation of Legionnaires' disease. *Eur J Clin Microbiol Infect Dis* **31**, 2017-2028.

Mentasti, M., Underwood, A., Lück, C., Kozak-Muiznieks, N.A., Harrison, T.G. & Fry, N.K. (2014). Extension of the *Legionella pneumophila* sequence-based typing scheme to include strains carrying a variant of the N-acylneuraminate cytidylyltransferase gene. *Clin Microbiol Infect* **20**, 435-441.

Mentasti, M., Kese, D., Echahidi, F., Uldum, S.A., Afshar, B., David, S., Mrazek, J., De Mendonça, R., Harrison, T.G., Chalker, V.J. (2015). Design and validation of a qPCR assay for accurate detection and initial serogrouping of *Legionella pneumophila* in clinical specimens by the ESCMID Study Group for Legionella Infections (ESGLI). *Eur J Clin Microbiol Infect Dis* **34**, 1387-1393.

Mentasti, M., Afshar, B., Collins, S., Walker, J., Harrison, T.G., Chalker, V. (2016). Rapid investigation of cases and clusters of Legionnaires' disease in England and Wales using direct molecular typing. *J Med Microbiol* **65**, 484-493.

Mentasti, M., Cassier, P., David, S., Ginevra, C., Gomez-Valero, L., Underwood, A., Afshar, B., Etienne, J., C. Buchrieser, C., Chalker, V.C., Harrison, T.G., Jarraud, S. Rapid detection and evolutionary analysis of *Legionella pneumophila* serogroup 1 ST47. Accepted.

Mérault N, Rusniok C, Jarraud S, Gomez-Valero L, Cazalet C, Marin M, Brachet E, Aegerter P, Gaillard JL, Etienne J, Herrmann JL; DELPH-I Study Group, Lawrence C, Buchrieser C. (2011). Specific real-time PCR for

simultaneous detection and identification of *Legionella pneumophila* serogroup 1 in water and clinical samples. *Appl Environ Microbiol* **77**, 1708-1717.

Moran-Gilad, J., Lazarovitch, T., Mentasti, M., Harrison, T., Weinberger, M., Mordish, Y., Mor, Z., Stocki, T., Anis, E., Sadik, C., Amitai, Z., Grotto I. (2012). Humidifier-associated paediatric Legionnaires' disease, Israel, February 2012. *Euro Surveill* **17**. Available from: www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20293.

Moran-Gilad, J., Mentasti, M., Lazarovitch, T., Huberman, Z., Stocki, T., Sadik, C., Shahar, T., Anis, E., Valinsky, L., Harrison, T.G., Grotto, I.; ESCMID Study Group for Legionella Infections (ESGLI). (2014). Molecular epidemiology of Legionnaires' disease in Israel. *Clin Microbiol Infect* **20**, 690-696.

Murphy, N.M., McLauchlin, J., Ohai, C., Grant, K.A. (2007). Construction and evaluation of a microbiological positive process internal control for PCR-based examination of food samples for *Listeria monocytogenes* and *Salmonella enterica* *Int J Food Microbiol* **120**, 110-119.

Pedro-Botet, L. and Yu, V.L. (2006). Legionella: macrolides or quinolones? *Clin Microbiol Infect* **12**, 25-30.

Phin, N., Parry-Ford, F., Harrison, T., Stagg, H.R., Zhang, N., Kumar, K., Lortholary, O., Zumla, A., Abubakar, I. (2014). Epidemiology and clinical management of Legionnaires' disease*. Lancet Infect Dis* **14**, 1011–1021.

Ratzow, S., Gaia, V., Helbig, J.H., Fry, N.K, Lück, P.C. (2007). Addition of *neuA*, the gene encoding *N*-acylneuraminate cytidylyl transferase, increases the discriminatory ability of the consensus sequence-based scheme for typing *Legionella pneumophila* serogroup 1 strains. *J Clin Microbiol* **45**, 1965-1968.

Reuter, S., Harrison, T.G., Köser, C.U., Ellington, M.J., Smith, G.P., Parkhill, J., Peacock, S.J., Bentley, S.D., Török, M.E. (2013). A pilot study of

rapid whole-genome sequencing for the investigation of a Legionella outbreak. BMJ Open doi: 10.1136/bmjopen-2012-002175.

Ridderberg, W., M. G. Fenger, and N. Nrskov-Lauritsen. (2010). *Haemophilus influenzae* may be untypable by the multilocus sequence typing scheme due to a complete deletion of the fucose operon. *J. Med. Microbiol* **59**, 740–742

Saunders, N., Zambon, M., Sharp, I., Siddiqui, R., Bermingham, A., Ellis, J., Vipond, B., Sails, A., Moran-Gilad, J., Marsh, P., Guiver, M.; HPA Microbiology Services Division. (2013). Guidance on the development and validation of diagnostic tests that depend on nucleic acid amplification and detection. *J Clin Virol* **56**, 260-270.

Sedgwick, J., Joseph, C., Chandrakumar, M., Harrison, T., Lee, J., de Jong B. (2007). Outbreak of respiratory infection on a cruise ship. *Euro Surveill* **12**. Available from: www.eurosurveillance.org/ViewArticle.aspx?ArticleId=3246.

Underwood, A.P., Jones, G., Mentasti, M., Fry, N.K., Harrison, T.G. (2013). Comparison of the *Legionella pneumophila* population structure as determined by sequence-based typing and whole genome sequencing. *BMC Microbiol* doi: 10.1186/1471-2180-13-302.

van Belkum, A., Tassios, P.T., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N.K., Fussing, V., Green, J., Feil, E., Gerner-Smidt, P., Brisse, S., Struelens, M.; European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Study Group on Epidemiological Markers (ESGEM). (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin Microbiol Infect* **13**, 1-46.

Unit for Surveillance and Control of Communicable Diseases. (2009). Legionnaires disease in a neonatal unit of a private hospital, Cyprus, December 2008: preliminary outbreak report. *Euro Surveill* **15**, 14(2).

Vekens, E., Soetens, O., De Mendonça, R., Echahidi, F., Roisin, S., Deplano, A., Eeckhout, L., Achtergael, W., Piérard, D., Denis, O., Wybo, I. (2012). Sequence-based typing of *Legionella pneumophila* serogroup 1 clinical isolates from Belgium between 2000 and 2010. *Euro Surveill* **17**, 20302.

von Baum, H., Ewig, S., Marre, R., Suttorp, N., Gonschior, S., Welte, T., Lück, C.; Competence Network for Community Acquired Pneumonia Study Group. (2008). Community-acquired *Legionella* pneumonia: new insights from the German competence network for community acquired pneumonia. *Clin Infect Dis* **46**, 1356-1364.

Appendix I: figure and tables

Figure 1. Circular representation of “Lorraine” ST47 genome and comparison to the genome of closely related strains (i.e., ST74, ST82, “England & Wales” ST109, “France” ST109) and ST62 obtained using CiVi (http://civi.cmbi.ru.nl).

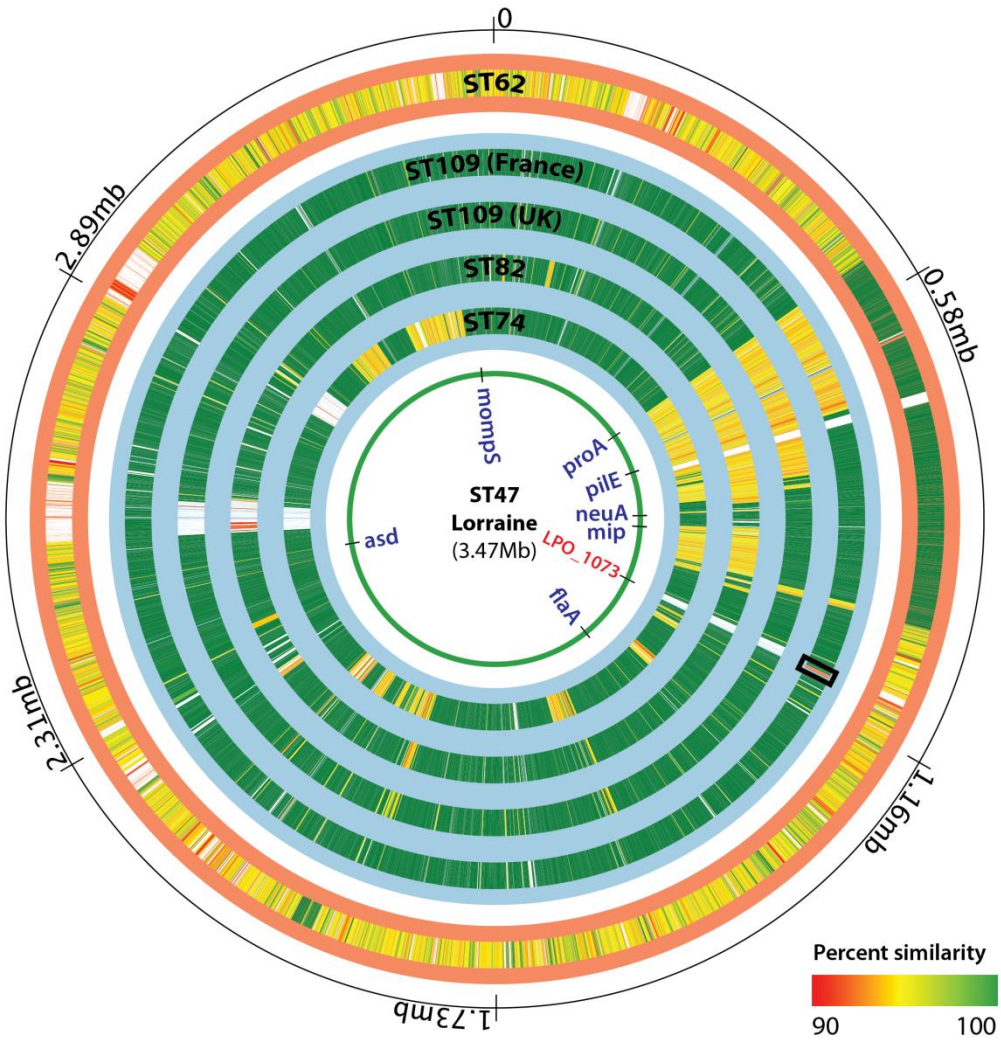

Table 1. Summary of SBT EQA results

Distribution	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Year	2003	2004	2006	2007	2008	2010	2010	2011	2012	2014
Participant	16	16	28	23	32	38	38	37	37	37
Expected Results	38.0%	71.2%	92.0%	37.4%	90.3%	94.4%	81.3%	80.0%	89.7%	95.7%

Table 2. Summary of results obtained during investigation of LD cluster occurred in South Wales in 2013.

Location	Sample	PCR			Nested SBT		*L. pneumophila* sg1		SBT	
		Lpn	sg1	ST47	Allelic Profile	ST	Culture	Sub-group	Allelic Profile	ST
South Wales (2013)	Sputum	+	+	-	2,3,0,10,2,1,6	0	+	Allentown/France	2,3,9,10,2,1,6	**23**
	Sputum	+	+	-	3,0,1,1,0,9,1	0	-	NA	NA	**NA**
	Sputum	+	+	+	5,0,22,15,6,2,6	0	+	Allentown/France	5,10,22,15,6,2,6	**47**
	Sputum	+	+	+	5,10,22,15,6,2,6	47	+	Allentown/France	5,10,22,15,6,2,6	**47**
4 LD cases in 1 week		**Day 0**			**Day 1**		**Day 5-7**		**Day 8**	

Appendix II: papers

PEER-REVIEWED PUBLICATIONS

1. **Farhat, C., Mentasti, M., Jacobs, E., Fry, N.K., Lück, C. (2011).** The *N*-acylneuraminate cytidyltransferase gene, *neuA*, is heterogenous in *Legionella pneumophila* strains but can be used as a marker for epidemiological typing in the consensus sequence-based typing scheme. *J Clin Microbiol* **49**, 4052 - 4058.
2. **Mentasti, M., Fry, N.K., Afshar, B., Palepou-Foxley, C., Naik, F.C. & Harrison, T.G. (2012).** Application of *Legionella pneumophila*-specific quantitative real-time PCR combined with direct amplification and sequence-based typing in the diagnosis and epidemiological investigation of Legionnaires' disease. *Eur J Clin Microbiol Infect Dis* **31**, 2017-2028.
3. **Moran-Gilad, J., Lazarovitch, T., Mentasti, M., Harrison, T., Weinberger, M., Mordish, Y., Mor, Z., Stocki, T., Anis, E., Sadik, C., Amitai, Z., Grotto I. (2012).** Humidifier-associated paediatric Legionnaires' disease, Israel, February 2012. *Euro Surveill* **17**.
4. **Mentasti, M., Underwood, A., Lück, C., Kozak-Muiznieks, N.A., Harrison, T.G. & Fry, N.K. (2014).** Extension of the *Legionella pneumophila* sequence-based typing scheme to include strains carrying a variant of the N-acylneuraminate cytidylyltransferase gene. *Clin Microbiol Infect* **20**, 435-441.
5. **Moran-Gilad, J., Mentasti, M., Lazarovitch, T., Huberman, Z., Stocki, T., Sadik, C., Shahar, T., Anis, E., Valinsky, L., Harrison, T.G., Grotto, I.; ESCMID Study Group for Legionella Infections (ESGLI). (2014).** Molecular epidemiology of Legionnaires' disease in Israel. *Clin Microbiol Infect* **20**, 690-696.
6. **Mentasti, M., Kese, D., Echahidi, F., Uldum, S.A., Afshar, B., David, S., Mrazek, J., De Mendonça, R., Harrison, T.G., Chalker, V.J. (2015).** Design and validation of a qPCR assay for accurate detection and initial serogrouping of *Legionella pneumophila* in clinical specimens by the ESCMID Study Group for Legionella Infections (ESGLI). *Eur J Clin Microbiol Infect Dis* **34**, 1387-1393.
7. **David, S., Mentasti, M., Tewolde, R., Aslett, M., Harris, S.R., Afshar, B., Underwood, A., Fry, N.K., Parkhill, J., Harrison, T.G. (2016).** Evaluation of an optimal epidemiologic typing scheme for *Legionella pneumophila* with whole genome sequence data using validation guidelines. *J Clin Microbiol* pii: JCM.00432-16. [Epub ahead of print]
8. **Mentasti, M., Afshar, B., Collins, S., Walker, J., Harrison, T.G., Chalker, V. (2016).** Rapid investigation of cases and clusters of Legionnaires' disease in England and Wales using direct molecular typing. *J Med Microbiol* **65**, 484-493.
9. **Mentasti, M., Cassier, P., David, S., Ginevra, C., Gomez-Valero, L., Underwood, A., Afshar, B., Etienne, J., C. Buchrieser, C., Chalker, V.C., Harrison, T.G., Jarraud, S.** Rapid detection and evolutionary analysis of *Legionella pneumophila* serogroup 1 ST47. Accepted.

Summary of PhD candidate's contribution to publications

#	PUBLICATION	Original idea	Study design	Method development	Experimental work	Data analysis	Data interpretation	Writing	Overall contribution
1	Farhat *et al.* (2011)				x	x	x	x	Minor contribution
2	Mentasti *et al.* (2012)	x	x		x	x	x	x	Major contribution
3	Moran-Gilad *et al.* (2012)		x	N/A	x	x	x	x	Major contribution
4	Mentasti *et al.* (2014)	x	x	x	x	x	x	x	Main contributor
5	Moran-Gilad *et al.* (2014)			N/A	x	x	x	x	Minor contribution
6	Mentasti *et al.* (2015)	x	x	x	x	x	x	x	Main contributor
7	David *et al.* (2016)		x		x		x	x	Minor contribution
8	Mentasti *et al.* (2016)	x	x	x	x	x	x	x	Main contributor
9	Mentasti *et al.* (accepted)	x	x	x	x	x	x	x	Main contributor

Appendix III: supporting letters

Appendix IV: guidelines