TECHNISCHE UNIVERSITÄT DRESDEN

Institut für Medizinische Mikrobiologie und Hygiene

Direktor: Herr Prof. Dr. med. Enno Jacobs

Array hybridization and whole genome sequencing as new typing tools for *Legionella pneumophila*

DISSERTATIONSSCHRIFT

zur Erlangung des akademischen Grades

Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt

der Medizinischen Fakultät Carl Gustav Carus

der Technischen Universität Dresden

von

Markus Petzold (M. Sc.)

aus Dresden

- Dresden 2017 -

1. Gutachter: Prof. Dr. med. Enno Jacobs
2. Gutachter: Prof. Dr. rer. nat. Michael Göttfert

Tag der Einreichung: 08.11.2017
Tag der mündlichen Prüfung (Verteidigungstermin): 14.02.2018

Die Dissertation wurde im Zeitraum vom September 2013 bis November 2017 im Institut für Medizinische Mikrobiologie und Hygiene der Technischen Universität Dresden Angefertigt.

gez.: _____
Vorsitzender der Promotionskommission
Prof. Dr. med. Andreas Seidler, MPH

Acknowledgement

My most sincere gratitude goes to you, Christian, for all inspiring, excessive, often funny and pointed, but always friendly discussions. I appreciated the cities we traveled, the wine we drank, and the input you drummed into my brain. You encouraged me to think outside the box – something really important in the life of a scientist. I always felt a valued member of the lab and I am really thankful for the responsibilities and the trust you gave me through the years. You showed me that first-class mentorship and kindness is possible, something I unfortunately mentioned too rarely. Thank you very much.

For kindly agreeing and allowing me to do my work in his institute I thank Prof. Dr. Enno Jacobs.

Many thank also to Prof. Dr. Michael Göttfert who was patient and understanding when things became difficult.

Many thanks go to Stefan Monecke, Massimo Mentasti, Jacob-Moran Gilad, Dag Harmsen, Karola Prior, Anika Wunderlich, Sophia David, Klaus Heuner, Barbara Schrammel, Roger Dumke, Alexander Thürmer and the Applied Maths team for accompany me along the scientific path.

My colleagues from the *Legionella*-lab, Kerstin L., Susi, Ines, Kerstin R., Sigrid and Edith, deserve special thanks for welcoming me in the family.

And what makes life bearable? Sweets (and coffee and beer)! I will miss the ‘Schweinsohrsession’, the birthday cakes, BBQs, the beer hours and the PhD-office crew. You brought the sweet and hoppy side of life into the dreary grey labs: Christian, Anne, Jenny, Tanja, Corinna, Steffi, Leo Lausemaus, Kati, Charlotte, Ellen, Katharina, Fabian, Anja, Sarah, Manu, Christin, Diana, Nicole, Catharina and the numerous Foamys.

Last but not least, I want to thank Dr. Hans-Udo Graf, Dr. Peter Klaren, Prof. Huub Op den Camp and Dr. Jan Keltjens for introducing me to the fascinating world of biology and microbiology.

- Für meine Familie -

Summary

To understand transmissible human diseases, disciplines such as epidemiology and the surveillance of affected cases are as essential as the knowledge about the pathogenesis and the course of a disease. Epidemiologists categorize and estimate factors for public health risks by taking metadata into account including geographic aspects, health and social states to study a disease transmission and prevent further cases. In addition, a focus on the causative agents itself is necessary in order to understand their ecology and hence their virulence traits. The causative agents for a severe pneumonia named Legionnaires' disease (LD) are bacteria of the genus *Legionella*. The putative sources of LD infection are any aerosol-generating natural or man-made fresh water systems. Due to this ubiquitous distribution of legionellae, it is difficult to find the source of infection. Therefore, it is necessary to isolate the bacterium from the suffering patients to further characterize it in the laboratory and to compare the clinical isolates with isolates obtained from probable environmental sources.

The predominant species isolated from LD patients is *Legionella pneumophila* serogroup (Sg) 1. Intensive genotyping of *L. pneumophila* Sg1 isolates by using the current gold standard method, the sequence-based typing scheme (SBT), revealed limitations in the discrimination of several sequence types (ST) which could not be compensated for by additional phenotypic typing scheme. In practical terms, this means that several clones or STs are disproportional frequently found in both, patients and water systems, and cannot be distinguished by current methods. Therefore, a distorted picture of endemic and globally-spread clones is generated and current typing methods cannot add substantial information during the identification of the infectious source. The aim of this thesis is to develop and implement new typing methods for *L. pneumophila* isolates with a higher resolution than the gold standard methods.

A DNA-DNA hybridization based microarray was designed and equipped with probes that target specifically *L. pneumophila* virulence factors and genes that are involved in the biosynthesis of lipopolysaccharide structures. Legionellae can be subgrouped on the basis of their lipopolysaccharide structures. Here, the usually phenotypic characterization of *L. pneumophila* Sg1 is successfully transmitted to a DNA-based genotypic method. Furthermore, the detailed validation of the DNA-microarray revealed a higher discriminatory power in comparison to the gold standard methods. It enables previously indistinguishable clones to be subdivided, providing valuable information about probable sources of infection.

The second new tool for typing of *L. pneumophila* is based on the core genome of the bacteria. An extended SBT-scheme was extracted from the core genome and accordingly named *core genome multilocus sequence typing* (cgMLST). This genome wide gene-by-gene typing approach allows a high genomic resolution of *L. pneumophila* isolates

by retaining epidemiological concordance. A major advantage of this genome-based method is the detection of large recombination events within the analysed genomes, which is, so far, reserved for whole genome sequencing. The population structure of legionellae is largely driven by recombination and horizontal gene transfer rather than by spontaneous mutations. Therefore, the detection of recombination events is essential for typing of *L. pneumophila* isolates. In addition, the cgMLST-scheme assigns a core genome sequence type to the analysed isolate and allows backwards compatibility with the current SBT-scheme.

Both methods proved to be fast, reliable and robust typing methods through their application during outbreak investigations. Furthermore, both systems are particularly suited as routine molecular typing tools for the surveillance of single cases. The raw data are verified and translated into uniform portable codes, which enables the easy transfer and comparison of results. The standardized and portable quality of the results of both methods enables the establishment of a curated global database. This qualifies both methods as potential new gold standard methods for the genotyping of *L. pneumophila* isolates.

Zusammenfassung

Zu den essentiellen Grundlagen, die notwendig sind um Infektionskrankheiten zu verstehen, gehören neben den Kenntnissen zur Pathogenese und zum Krankheitsverlauf auch die Epidemiologie. Sie fokussiert sich auf die Überwachung von Krankheitsfällen und kategorisiert gesundheitliche Risikofaktoren, unter Berücksichtigung geographischer, krankheitsspezifischer und sozialer Aspekte, um die Ausbreitung der Krankheit zu untersuchen und weitere Fälle zu verhindern. Ebenso wichtig bei der Ergründung von Infektionskrankheiten sind fundierte Kenntnisse über den Krankheitserreger selbst. Sie sind notwendig, um die Ökologie und dadurch die Virulenz der Pathogene zu verstehen. Das verursachende Pathogen der Legionärskrankheit, einer schweren Pneumonie, sind Bakterien der Gattung *Legionella*. Potentielle Infektionsquellen für die *Legionella*-Pneumonie sind jene natürlichen und künstlichen Frisch- und Süßwassersysteme, aus denen Aerosole entstehen können. Aufgrund dieser ubiquitären Verteilung von Legionellen ist es dringend erforderlich das Bakterium aus erkrankten Patienten zu isolieren und im Labor zu analysieren. Erst der Vergleich von Patientenisolaten mit Isolaten aus möglichen Infektionsquellen ermöglicht die Identifizierung von Infektionsquellen.

Patienten, die an einer *Legionella*-Pneumonie erkrankten, wurden größtenteils durch die Art *Legionella pneumophila* Serogruppe 1 (Sg1) infiziert. Das intensive Genotypisieren von *L. pneumophila* Sg1-Isolaten durch sequenzbasiertes Typisieren (SBT), dem aktuellen Goldstandard der Legionellentypisierung, zeigte eine starke Einschränkung in der Diskriminationsfähigkeit einiger Sequenztypen. Weiterführende phänotypische Typisierungsmethoden können diese Problem häufig nicht mehr kompensiert. Im Umkehrschluss bedeutet dies, dass einzelne Klone und Sequenztypen überproportional häufig in Patienten und Wassersystemen gefunden werden, da sie durch aktuelle Typisierungsmethoden nicht unterschieden werden können. So entsteht ein verzerrtes Bild von endemisch und teilweise global vorkommenden Sequenztypen. Die aktuellen Typisierungsmethoden sind folglich bei der Suche nach Infektionsquellen nur bedingt von Nutzen. Das Ziel dieser Doktorarbeit ist, neue Typisierungsmethoden mit höherem Auflösungspotential als die aktuellen Methoden für *L. pneumophila* Isolate zu entwickeln um diese in Routine- als auch Ausbruchsituationen einzusetzen.

Ein Microarray, basierend auf DNA-DNA Hybridisierung, wurde entworfen und mit Sonden ausgestattet, die spezifisch für Gene von Virulenzfaktoren und der Lipopolysaccharidbiosynthese von *L. pneumophila* sind. Legionellen können auf Basis ihrer Lipopolysaccharidstruktur in Subgruppen eingeteilt werden. Diese bislang phänotypische Charakterisierung von *L. pneumophila* Sg1 ist hier erstmals erfolgreich auf eine genotypsiche Methode übertragen und im Labor etabliert worden. Die detaillierte Validierung des DNA-Microarrays zeigt zudem eine höhere Diskriminationsfähigkeit als die

aktuellen Referenzmethoden, wodurch wertvolle Informationen über mögliche Infektionsquellen erhalten werden.

Eine weitere neue Methode für die Typisierung von *L. pneumophila* Sg1 basiert auf dem Kerngenom der Bakterien. Bei diesem Gen-für-Gen Typisierungsansatz, welcher Kerngenom Multilokus Sequenztypiserung genannt wird (*engl.: core genome multilocus sequence typing, cgMLST*), werden die Gene des Kerngenoms analysiert. Dies ermöglicht eine hohe genomische Auflösung von *L. pneumophila*-Isolaten ohne Verlust der epidemiologischen Konkordanz. Ein weiterer großer Nutzen dieser genombasierten Methode ist die Detektion von Rekombinationen größerer Genomabschnitte, etwas das bisher nur der Gesamtgenomanalyse vorbehalten ist. Da die Populationsstruktur von Legionellen größtenteils durch Rekombinationen und horizontalen Gentransfer geprägt ist, und weniger durch spontane Mutationen, stellt die Detektion von Rekombinationen einen essentiellen Teil bei der Analyse einzelner Isolate dar. Das neue cgMLST-Schema ermöglicht zusätzlich die Einteilung von *L. pneumophila*-Isolaten in sogenannte Kerngenomsequenztypen und ermöglicht eine Rückwärtskompatibilität zum aktuellen sequenzbasierten Typisierungsschema.

Beide Methoden überzeugen als schnelle, zuverlässige und robuste Typisierungsmethode während ihrer Anwendung bei Legionellenausbrüchen. Des Weiteren können sie als Routinemethoden für die Überwachung einzelner Krankheitsfälle eingesetzt werden. Für die Entwicklung neuer Methoden ist vor allem eine Vereinheitlichung der Rohdaten notwendig. Ein einheitliches Protokoll sowie Dateninterpretation ermöglichen den Vergleich und Austausch der ermittelten Ergebnisse. Die standardisierte Qualität der Ergebnisse beider Methoden ermöglicht den Aufbau und Pflege globaler Datenbanken. Dadurch sind beide Methoden potentielle neue Referenzmethoden für die Genotypisierung von *L. pneumophila*.

Table of contents

Part I - Introduction

Background

In June 1976, the annual convention of the American Legion was held in Philadelphia. Approximately one week after the convention, 182 people, mostly convention attendees, suffered from cough and fevers A total of 147 patients were hospitalized due to severe pneumonia, the cause of which was unknown, and 29 patients died (Fraser, et al., 1977). Several months later, the investigation team of the Centers for Disease Control and Prevention (CDC) succeeded in isolating the outbreak's etiological agent – a gram-negative non-spore forming bacillus ranging from small rods of 2-3 µm to longer filaments of 8-10 µm. This bacterium was initially called Legionnaires' disease bacterium due to its historical association with the American Legion Convention, and was later named *Legionella pneumophila*. Likewise, the pneumonia it causes was named Legionnaires' disease (LD) or legionellosis (McDade, et al., 1977; Brenner, et al., 1979; Horwitz and Silverstein, 1980). Retrospective examination determined *Legionella* spp. as the cause of several pneumonia cases dating back to 1947 (McDade, et al., 1977; Glick, et al., 1978; McDade, et al., 1979). Since then, 60 species and more than 70 serogroups (Sg) of *Legionella* have been described worldwide (Bajrai, et al., 2016; Ishizaki, et al., 2016).

Environment - Ecology

Legionellae are ubiquitously found in natural and anthropogenic freshwater environments such as cooling towers (CT), spas and drinking water distribution systems (van Heijnsbergen, et al., 2015). The optimum temperature for growth is from 35 °C to 37 °C (Feeley, et al., 1978; Kusnetsov, et al., 1996). However, legionellae have also been found at temperatures as low as 10 °C and as high as 70 °C (Fliermans, et al., 1981; Tison and Seidler, 1983; Lesnik, et al., 2016). In aqueous environments, legionellae are able to colonize persisting biofilms (Declerck, 2010). Their development is characterised by a biphasic cycle that comprises a replicative exponential phase form (EPF) and a transmissive stationary form (SPF). Certain environmental factors trigger these extracellular forms. This causes them to enter for instance the 'viable but non-culturable' (VBNC) state, which is characterised by a low metabolic activity, or to form long filaments (FF) (Faulkner and Garduno, 2002; Robertson, et al., 2014; Kirschner, 2016). In addition, legionellae are able to parasitize protozoa such as amoeba that graze on biofilms (Declerck, 2010). After internalisation by protozoa, legionellae evade lysosomal degradation and survive within vacuoles. From within these vacuoles, the pathogen secretes more than 300 effector

proteins into the host cell via a Dot/Icm type IV translocation system. These effectors intercept the endoplasmatic reticulum (ER)-Golgi traffic, recruit ER markers and remodel the host to form its own replicative niche named Legionella containing vacuole (LCV), which results in a highly infectious and stress resistant mature intracellular form (MIF) (Ensminger, 2016). After replication, the infected and apoptotic host cell lyses and a planktonic transmissive form of the bacterium is released into the environment, where it either infects other hosts or becomes sessile within biofilms communities. Recent studies have reported the Caenorhabditis nematodes as a natural host for *L. pneumophila*, which may become infected upon ingestion of a *Legionella*-containing protozoan (Brassinga, et al., 2010; Hellinga, et al., 2015; Rasch, et al., 2016). Furthermore, legionellae, especially the species *L. longbeachae*, are frequently isolated from soil or composted matter (Steele, et al., 1990; Den Boer, et al., 2007).

Infection of the respiratory tract

Another rather accidental host is the respiratory tract of mammals. Although several studies demonstrated that animals may become infected with legionellae, the main concern of the pathogen on human has focused on human infections (Fitzgeorge, et al., 1983; Boldur, et al., 1987; Fabbi, et al., 1993; Hägele, et al., 2000; Dorer, et al., 2006; Harding, et al., 2012). Legionellae can enter the human lung via aerosolised water droplets. These aerosols can be emitted by any water-bearing system in contact with air, such as CTs, spas, showers, fountains, humidifiers and waste water treatments plants, etc. (Figure 1) (Berendt, 1980; van Heijnsbergen, et al., 2015; Allegra, et al., 2016). Consequently, small *Legionella*-containing aerosol droplets can reach the lower respiratory tract where they encounter alveolar macrophages. These macrophages can subsequently be infected by the pathogen in a similar way to amoeba. Macrophages and amoeba are professional phagocytes that generally eliminate phagocytosed targets by means of lysosomal degradation. Studies on animals, human cell lines and amoeba showed a similar mode of infection by *L. pneumophila*. It has been argued that the long co-evolution of *L. pneumophila* with its natural host amoeba lead to the selection of an infection mechanism that allows its uptake and replication within other similar phagocytosing cells (Escoll, et al., 2013). Nevertheless, the infectious dose of aerosols lacks adequate evaluation. Several effects have been described, including the particle size, the packing of aerosols with infectious bacteria, the replicative form of infectious bacteria, the charge of the droplet and/or the charge of the bacterium's outer membrane and the device that's emits aerosols (Berendt, 1980; Dennis and Lee, 1988; Gaboriaud, et al., 2008; Gosselin, et al., 2011; Hines, et al., 2014). However, it is generally assumed that the human lung represents an evolutionary dead end for

legionellae, although the first putative person-to-person transmission was reported recently (Correia, et al., 2016).

Clinical symptoms

Legionellae cause two forms of respiratory infections: a non-pneumonic mild form named Pontiac fever and the LD mentioned above, with its clinical manifestation of an atypical pneumonia. Pontiac fever is a self-limiting illness with influenza-like symptoms (e.g. muscle pain, cough, fever, headache and chills) and usually lasts up to five days. The incubation time is 24 to 48 hours. So far, no deaths have been reported due to Pontiac fever. The more severe form, LD, features influenza-like symptoms as well as symptoms of severe pneumonia, including shortness of breath, nausea, vomiting, diarrhoea, chest pains, confusion and the coughing of blood. The incubation period is from two to ten days, and in some cases up to 20 days (WHO, 2007). In general, healthy people seldom become infected with legionellae. However, several host factors have been described that predispose one to becoming infected with LD. There is a clear increase of reported cases in people older than 50 years, and males are affected twice to four times as frequently as females. Only a few cases have been reported in children (Phin, et al., 2014). Smokers or former smokers, as well as people with chronic lung diseases (e.g. chronic obstructive respiratory disease [COPD]), are more susceptible. Furthermore, people who are immunosuppressed due to illnesses such as cancer, diabetes, and kidney failures, and

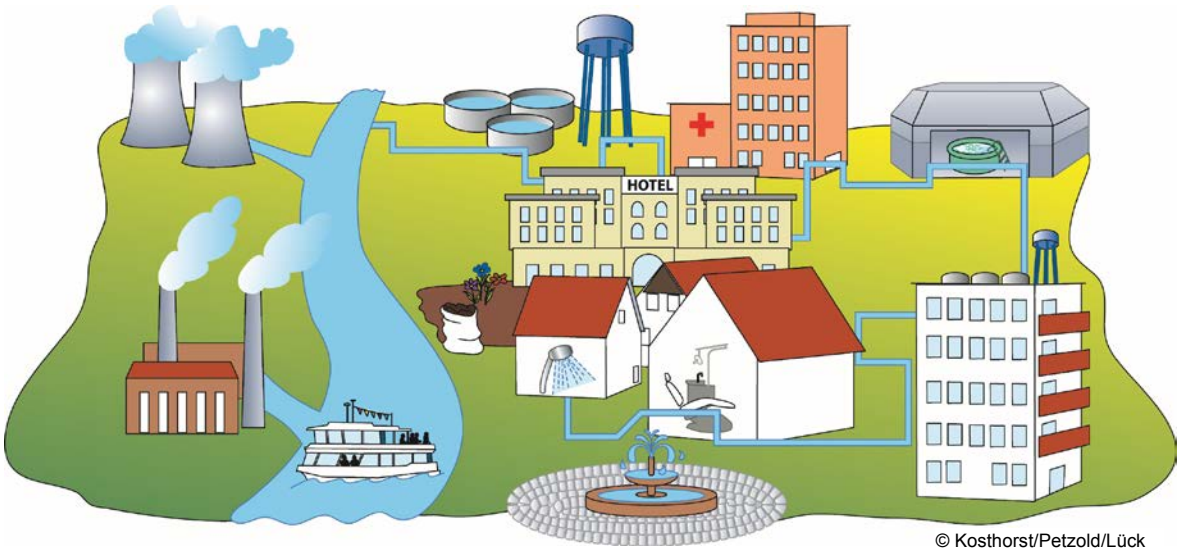


Figure 1: Schematic route of *Legionella* dissemination from natural and artificial reservoirs. *Legionella* can enter the fresh water system from environmental systems such as rivers or waste water treatment plants. Colonization of fresh water systems by *Legionella* can occur under optimal conditions that enable the amplification of legionellae. Aerosols can be distributed by e.g. private and public spas, cooling towers on buildings and plants, fountains, dental units and showers. Large apartment buildings, hotels, cruise ships and hospitals with branched and often unclear water networks possess more likely niches that are favorable for *Legionella*-amplification. Another, water-independent source of infection especially for the species *L. longbeachae* is (potting) soil.

people who take immunosuppressive or anti-inflammatory medications are prone to infection by legionellae.

Epidemiology

Legionnaires'' disease accounts for 2 to 8 % of community acquired pneumonia (CAP) (von Baum, et al., 2008; Bartlett, 2011). However, since surveillance systems for legionellae differ between countries, and are completely missing in some countries, the incidence of *Legionella*-pneumonia is generally thought to be higher. In several countries LD is a notifiable disease. Surveillance schemes have been implemented in the USA, Canada, New Zealand, Australia, Japan and Singapore. In Europe, surveillance is organized by the European Centre for Disease Prevention and Control (ECDC). The crude notification rate in Europe is 14 cases per million inhabitants. However, the rates vary from one in Poland to 51 per million inhabitants in Slovenia. In recent years, the number of reported LD cases in Europe has increased steadily, from 5,000 cases in 2011 to more than 7,000 cases in 2015. France, Spain, Italy and Germany report the most cases (68 % of cases reported in Europe) (ECDC, 2016). Similarly, the notification rate in the USA has increased from four in 2000 to eleven cases per million inhabitants in 2009, with a higher notification rate in the north-eastern states than in other states (CDC, 2011). A total of 5.000 cases were reported in the USA in 2015, with a fatality rate of 8 %, which is identical to that of Europe (CDC, 2015; ECDC, 2016).

The majority (70 %) of LD cases are sporadic single CAP cases. These patients acquired LD in any setting other than their domestic home, such as while travelling or during a health care stay. Around one fifth of the cases are travel associated (TALD), which means these patients had stayed at a public accommodation site (e.g. hotel) for two to ten days before the onset of their illness. The LD cases reported in health care settings (e.g. hospitals or nursing homes) account for approximately 8 %, while the remaining cases are attributed to clusters and outbreaks (Joseph, et al., 2010b; Beaute and ELDSNet, 2017). A cluster links two or more cases by space (e.g. accommodation site or work place) and has sufficient proximity in time, while outbreaks link cases in space and time (WHO, 2007; European Commission, 2012; Beaute and ELDSNet, 2017). In general, LD patients are infected by one single strain, although there have been cases reported with dual infections of *Legionella* strains (Wewalka, et al., 2014).

The European Legionnaires' Disease Surveillance Network (ELDSNet), which is coordinated by the ECDC, operates a platform that gathers a great deal of information on LD cases in Europe, with the help of national public health authorities. This means that TALDs, clusters and outbreaks can be detected rapidly. Its close contact with regional and national public health authorities and vice versa allows a prompt response when anomalies

are observed. It provides support materials that enable the public authorities to avoid the helplessness that they may face in the case of an LD outbreak. Several tools and guidelines assist them in identifying potential contaminated sources (e.g. GIS support, epidemiological surveys) and in assessing and minimizing the potential risks (prevention and control guidelines). This means that a network of local public health authorities, trained epidemiologists, microbiologists and risk assessors can called in order to identify the source of contamination and reduce the number of LD cases and fatalities (ECDC, 2012).

Laboratory analysis

The mere clinical picture of atypical pneumonia does not confirm legionellae as causative agent. Further laboratory analyses are urgently needed to confirm LD. There are several national and international guidelines that classify methods according to their potential to confirm the disease (CSTE, 2010; European Commission, 2012). The isolation of the bacteria by cultivation on selective media represents the most important method and is regarded as the 'gold standard'. Although legionellae have been isolated from blood and feces samples, the recommended sampling site for isolating living legionellae is the lower respiratory tract such as sputum, tracheal secretions, broncho-alveolar lavages (BAL), etc. (Edelstein, et al., 1979; Maiwald, et al., 1998; Rowbotham, 1998; Dunne, et al., 2017). Isolated colonies serve as a fundamental basis for further typing methods, which allows the detailed comparison of the clinical and environmental isolates. However, the isolation of legionellae by cultivation is time-consuming (up to ten days) and is subject to fluctuations related to the recovery rate. This is often due to patients' ongoing antibiotic therapies, which kill the antibiotic-sensitive legionellae prior to sampling, and which lowers the chance of isolating living bacteria. In addition, legionellae are slow-growing and are demanding bacteria with regard to nutrients. A special nutrient formulation that includes charcoal and essential supplements such as *L*-cysteine and ferrous sulfate is needed. Antibiotics are often needed to substantially diminish the growth of other respiratory microorganisms, hence the sensitivity range for cultivation of <10 to 80 % (Dunne, et al., 2017).

Therefore, clinicians prefer a rapid diagnosis that allows for adequate treatment. The detection of processed antigen (e.g. lipopolysaccharides) in urine samples was initially described for patients in Philadelphia the outbreak, 1976 (Berdal, et al., 1979; Sanford, 1979). The urinary antigen test (UAT) is a fast diagnostic method that allows the confirmation of LD within minutes (Dunne, et al., 2017). Since antigen can be even be excreted weeks after infection, a laboratory results must be examined in the context of the patient's clinical picture. Although this method is primarily limited to antigen of *L. pneumophila* Sg1, UATs are routinely used and confirm approximately 80 % of LD cases (ECDC, 2015b). Another fast and culture-independent method is the detection of legionellae

nucleic acid (DNA) by real-time polymerase chain reaction (PCR). This method has shown a higher sensitivity (83 %), together with a similar specificity of 90 % in comparison to the 100 % specificity using culture-based isolation (Cristovam, et al., 2017). Since legionellae do not colonize healthy people, a positive PCR-test is always indicative of an infection. Several partially multiplex systems have been introduced in recent years that target DNA of different *Legionella* species, but most prominently *L. pneumophila* and *L. pneumophila* Sg1 (Benitez and Winchell, 2013; Mentasti, et al., 2015; Benitez and Winchell, 2016). A recent study reported that PCR-based methods outcompetes UAT by detecting up to 30 % more LD cases (Avni, et al., 2016).

The three aforementioned methods rank first when laboratory confirmation of LD is required. Other methods, such as seroconversion, single high antibody titers against legionellae and direct fluorescent antibody tests (DFA) from respiratory samples, are inappropriate methods that lack sensitivity and specificity and should not be used to confirm LD cases (Dunne, et al., 2017).

The methods used to detect and/or isolate legionellae from environmental samples are similar. IN general, prior to detection, the capturing and/or the enrichment of the bacteria is necessary, for example by centrifugation or the filtration of water samples through membranes (ISO 11731, 1998; Villari, et al., 1998; ISO 11731-2, 2008; Reidt, et al., 2011; Wunderlich, et al., 2016). With regard to clinical samples, the primary goal for environmental samples is to isolate living bacteria on selective media. Furthermore, several PCR systems are available for the detection and enumeration of legionellae and are intensively compared to the cultivation method (Lee, et al., 2011; ISO 12869, 2012; Edagawa, et al., 2015; Collins, et al., 2017). However, intensive discussions that address the effect of the detection of free DNA or DNA from dead bacteria in samples are still ongoing and demand a viable-PCR (Schnetzinger, et al., 2013; Ditommaso, et al., 2015; Scaturro, et al., 2016).

Division and typing of legionellae

Of the 60 valid *Legionella* species described, half of them were isolated from clinical samples (Muder and Yu, 2002; Diederen, 2008; Bajrai, et al., 2016; Ishizaki, et al., 2016). However, the species most frequently isolated from respiratory material globally is *L. pneumophila* (95 %). Interestingly, in Australia and New Zealand another species is more prominent in clinical samples, where. *L. longbeachae* accounts for 50 to 70 % of clinical isolates and is commonly associated with potting soil (Slaon-Gardner, et al., 2011; Graham, et al., 2012).

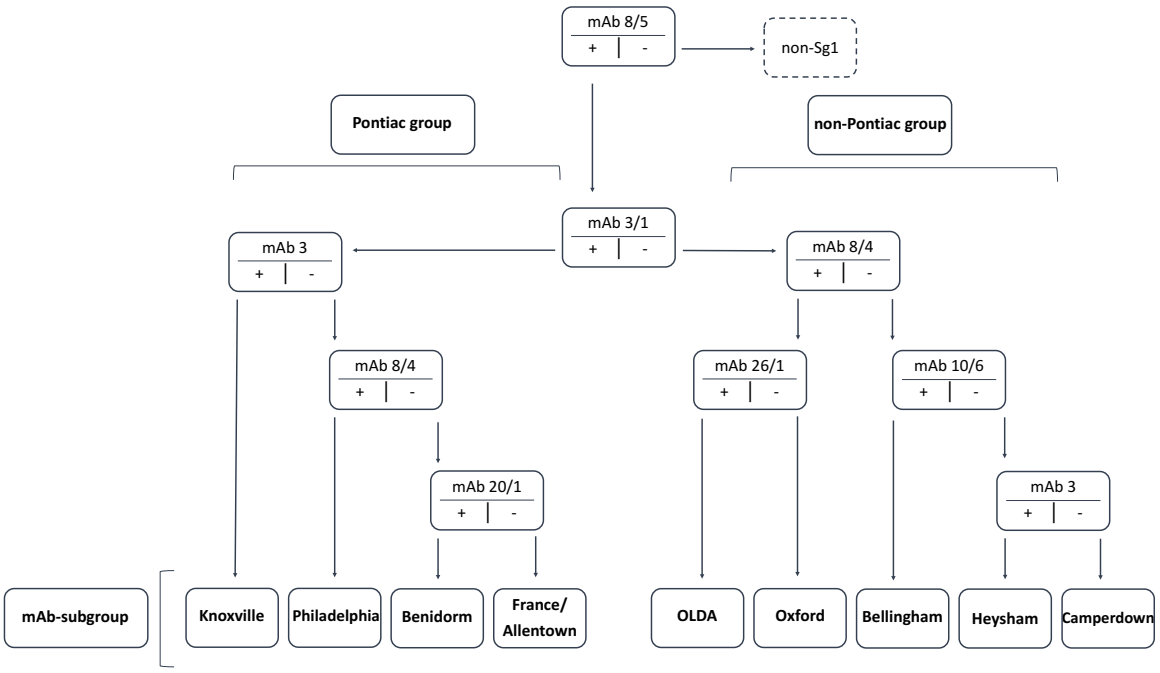


Figure 2 The mAb-based subgrouping scheme. The 'Dresden panel + mAb3' allows a hierarchical classification of *L. pneumophila* Sg1 isolates into mAb-subgroups (adapted from Helbig, et al., 2002).

The species *L. pneumophila* can further be subdivided into 15 serogroups (Sg) (Figure 2). This method is based on the structural differences of the lipopolysaccharide (LPS) on the bacteria's outer membrane which can be detected by specific monoclonal antibodies (mAb). Joly and colleagues described the potential of mAbs to divide *L. pneumophila* into Sgs that allows the differentiation of isolates from each other. A more sophisticated scheme named 'Dresden panel' (including mAb 3 of Joly's scheme) was set up by Helbig and colleagues and groups *L. pneumophila* isolates into 15 serogroups, and nine mAb-subgroups of Sg1 (Joly, et al., 1983; Joly, et al., 1986; Helbig, et al., 1997; Helbig, et al., 2002).

Two mAbs (mAb 8/5 and mAb 3/1) proved to be of particular value for epidemiological and clinical purposes as the most dominant and virulent strains could be identified. These mAbs specifically recognize the Sg1 strains that cause the majority of LD cases (mAb 8/5; 80 % of culture-confirmed LD cases) and strains of the Pontiac group (mAb 3/1), respectively. This represents 90 % of the Sg1 clinical isolates in CAP and TALD cases (Beaute, et al., 2013; Phin, et al., 2014; ECDC, 2016). Controversially, mAb $3/1^{+}$ isolates are seldom isolated from routine water samples (Lück, 2011).

However, the serotyping of *L. pneumophila* has limitations with regard to the potential for differentiation, initial studies reported (Figure 3). With the advent of genome-based methods of analysis, a larger spectrum of methods was applied to legionellae, basically by the restriction and the separation of DNA. The initial genome-based methods

for the typing of *L. pneumophila* Sg1 strains were evaluated in detail in 1999 by the European Working Group for Legionella Infections (EWGLI, which in 2012 became ESGLI, the affiliated study group of the European Society of Clinical Microbiology and Infectious Disease [ESCMID]) (Fry, et al., 1999). Several techniques were validated with regard to their discriminatory power and for their epidemiological concordance, using a European *L. pneumophila* strain collection (EUL). The amplified fragment-length polymorphism analysis (AFLP) (Valsangiacomo, et al., 1995) and the pulsed-field gel electrophoresis (PFGE) using the restriction nuclease SfiI (Lück, et al., 1991) proved to have the highest discrimination potential. However, since a discriminatory power value of 0.95 or higher is desirable, both methods required mAb-based subgrouping as a pre-screening method for acceptable discrimination (Fry, et al., 1999; Van Belkum, et al., 2007). While the molecular methods allowed a rapid screening of isolates within single laboratories, the interlaboratory exchange of results was unsatisfied, especially when gel-based methods were used. Feddersen and colleagues reported a gene-based typing scheme using *gyrB* for the first time, suggesting a higher discriminator potential than PFGE (Feddersen, et al., 2000). This scheme would enable the typing of *L. pneumophila* and simultaneous screening for quinolone-resistance determining regions. Nevertheless, it took three more years before the multilocus sequence typing (MLST) scheme was adapted for *L. pneumophila* Sg1, although it had already been implemented for highly infectious bacteria. It uses specific PCRs to amplify preselected parts of genes prior to sequencing (Maiden, et al., 1998). After quality control, the sequences are analyzed. Each variant of a specific gene, also referred to as

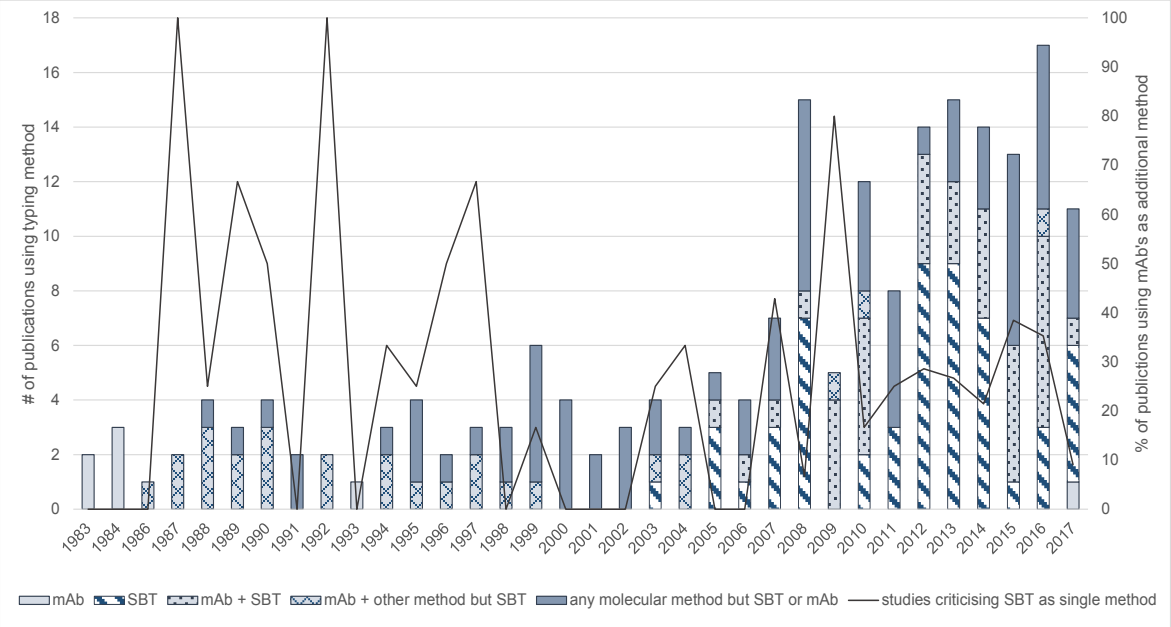


Figure 3: Literature overview of molecular typing of *L. pneumophila*. A Medline search for the typing methods of *L. pneumophila* was conducted (Suppl. Table 1). A total of 201 publications were included. Mab-based subgrouping was initially combined with all the kinds of molecular typing methods, mainly based on restriction techniques (1986 to 2007). The introduction of SBT as a typing method in 2003 (completed in 2007 with seven alleles) gradually replaced these techniques. The limitations of SBT were bypassed by adding another method (mainly mAb-subgrouping) or by developing new techniques. The solid line indicates to publications (by %) that reported concerns regarding SBT as a single typing method during the surveillance studies, case or outbreak reports.

allele, is assigned an allele number and each nucleotide polymorphism within an allele results in a particular allele number. In this way, a database of allele sequences and their corresponding allele numbers is generated, and an allele sequence can be compared to the set of known alleles within a database. The same allele number is only assigned to a sequence if a sequence matches a known allele sequence for 100 %. Any single polymorphism results in a new allele number. Since only one locus is less discriminatory, a set of alleles is used. Usually, MLST schemes comprise housekeeping genes that are not under selective pressure (Enright and Spratt, 1999; Spratt, 1999). Since the set for *L. pneumophila* contains genes that are virulence associated and are therefore probably under higher selective pressure than housekeeping genes, the scheme was named 'sequence-based typing' (SBT) to avoid misunderstandings. The initial set introduced by Gaia and colleagues comprised three genes (*flaA*, *proA* and *mompS*). The allele numbers were combined into a ternary-code revealing an index of discrimination (IOD) of 0.92, which was above that of the AFLP analysis. Once again, the IOD was 0.97 only by adding mAb-based subgrouping the IOD was 0.97 (Gaia, et al., 2003). The allelic scheme was extended to six alleles (IOD of 0.94) in 2005 and then to seven alleles (*flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*, and *neuA*) in 2007 to order to achieve an IOD of 0.96 (Gaia, et al., 2005; Ratzow, et al., 2007; Farhat, et al., 2011). The allele numbers were combined in strict order to an allelic profile or sequence type (ST). Since 2005, the SBT scheme has become the established method for *L. pneumophila* genotyping. A web-based platform was built and is curated by Public Health England (formerly called the Health Protection Agency), which facilitates a quality control for allele sequences, a growing database of allele sequences and STs, and further allows the retrieval of allelic profiles (Underwood, et al., 2006). For samples where no isolates can be obtained, a nested SBT (nSBT) approach was developed (Ginevra, et al., 2009a) in which alleles can be directly amplified in respiratory samples without the need for an isolated colony. Currently, 2,471 different STs are recorded in the ESGLI database (30th October 2017) (ESGLI, 2015).

Limitations of 'gold standard' typing methods

In general, the comparison of environmental and clinical isolates is either facilitated by using SBT as a single typing tool or, in two consecutive steps, by using mAb-subtyping as a pre-screening tool, followed by SBT. When isolates have the same mAb-subgroup, they are subsequently typed using SBT. Matching isolates for both methods means the identification of the source of infection and implies further actions that depend on the water system (e.g. decontamination, risk assessment, construction measures). The introduction of SBT as a globally recognized typing method for *L. pneumophila* was greeted by great expectation and surveillance laboratories extolled the robustness and comparability of the

method. Several studies were conducted to analyze ST variety in clinical or environmental isolates of specific regions or strain collections (Borchardt, et al., 2008; Chasqueira, et al., 2009; Harrison, et al., 2009; Amemura-Maekawa, et al., 2010; Tijet, et al., 2010). Simultaneously, LD surveillance was mainly based on SBT, and not only in Europe (Fry, et al., 2007; Joseph and Ricketts, 2010; CDC, 2011; Kozak-Muiznieks, et al., 2014; ECDC, 2015b). It soon became apparent that particular STs can exist as isolates in almost exclusively environmental samples, while others are mainly found in clinical samples. Similar tendencies were observed for globally spread STs, i.e. the ST1 clone and endemic clones (in Europe: ST23, ST37, ST47, ST62; in Germany, Berlin: ST182) were found in distinct areas (Cazalet, et al., 2008; Ginevra, et al., 2009b; Tijet, et al., 2010; Vergnes, et al., 2011; Krause, 2014). These results confirmed tentative doubts concerning this method as previous genome-based methods had also required an additional technique for higher resolution between isolates (Figure 3). Although the number of studies conducted using SBT as genetic typing tool increased, the major point of concern was the need for an additional typing method. As a result, mAb-subgrouping is still used as a pre-screening tool prior to SBT.

Major drawbacks concerning the identification of a potential source of infection were reported for endemic clones that could not be differentiated, even when both, the SBT and mAb-subgrouping methods were used. For example, isolates typed as ST1, mAb-subgroup OLDA or ST47, mAb-subgroup Allentown/France (especially in Western Europe) were frequently found in different kinds of environmental and clinical samples. A clear assignment of one environmental reservoir as a source of infection for a clinical case is unreliable. Therefore, several environmental sources could be regarded as possible sources of infection, even when no space temporal differences are obvious. Therefore, a further, third, method of differentiation was required. Laboratories facing these obstacles therefore developed other tools to further differentiate these clones. These tools used specific genetic elements as targets to differentiate clones. For example, the pathogenicity island (PA-I) was used as a target for variable element typing (VET) (Pannier, et al., 2010). The set of virulent genes on this PA-I differs between strains of the same clone. A further method uses the bacterial antiviral defense system of clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas), CRISPR-Cas. This genetic region is composed of spacers and repeats. By typing the number, order and sequences of theses spacers, a spoligotype is assigned to an isolate (Ginevra, et al., 2012). The multiple-locus variable number of repeats (VNTR) assay (MLVA) also showed the potential to differentiate between clones of the same ST and the mAb-subgroup. The number and size of tandem repeats spread throughout the genome is analyzed and, similarly to SBT, allows an allelic profile that can be compared between isolates (Pourcel, et al., 2007; Kahlisch, et al., 2010; Sobral, et al., 2011). Other laboratories developed microarrays with probes of the

L. pneumophila genome in order to verify the specific markers or marker genes found exclusively in clinical isolates, or to differentiate isolates of the same ST and mAb-subgroup by their clonal origin (Yzerman, et al., 2010; Den Boer, et al., 2013).

Despite these intense efforts to find a high-resolution typing method that is robust, reproducible and reliable, the majority of the tools were not adopted by other laboratories and appear to only have one single application. Reasons for this, are the lack of an international standardization and the absence of a curated database for the comparison and the exchange of results.

Unexpectedly, Underwood and colleagues revealed that SBT is in fact an optimal typing method for *L. pneumophila* on the basis of the whole genome sequences (WGS) of more than 800 different STs. They found that clustering based on WGS data was similar to ST-based clustering. They further identified, that *L. pneumophila* is spiked with recombined regions, which are mostly described for regions involved in phenotypical characteristics. This may explain the occurrence of different mAb-subtypes within a particular ST and vice versa (Underwood, et al., 2013). In the same year, Reuter and colleagues proved the feasibility of WGS for the analysis of isolates during an outbreak scenario (Reuter, et al., 2013).

Aim of this work

When this work started in 2013, next-generation sequencing (NGS) represented a major improvement in the molecular typing of bacteria and had triggered a wave of widespread use. However, the gold standard methods for *L. pneumophila* were SBT and mAb-based subgrouping, both of which demonstrated their limitations for the discrimination of several clones. Consequently, this work focus on new methods of molecular typing. It addresses issues of the limited molecular discrimination of the current gold standard methods and also the lack of standardized interpretation of WGS-data.

Although NGS is a promising tool that allows a widespread application, routine laboratories still cannot afford the second or third generation sequencing techniques. In addition, the analysis of NGS-raw data requires experienced staff, especially for the data interpretation. Therefore, this work introduces two new methods: a standardized NGS-based method named core genome MLST and a DNA-microarray that runs on a standard laboratory platform.

Part II reports the successive development of a DNA-DNA hybridization based microarray for the molecular typing of *L. pneumophila*. This microarray can be equipped with several hundred probes that hybridize to bacterial DNA. The microarray platform includes a reading device with automated data interpretation and is already reported for a variety of pathogens such as *Staphylococcus* spp., *Mycobacterium* spp.,

Salmonella enterica, *Mycoplasma* spp., *Brucella* spp., or influenza and is established in numerous laboratories (Monecke and Ehricht, 2005; Monecke, et al., 2008; Schmoock, et al., 2011; Braun, et al., 2012; Ruettger, et al., 2012; Schnee, et al., 2012). The most important aspect of this work concentrates on the identification of suitable targets that allow a higher discriminatory potential than current methods. An additional aspect is the simultaneous assignment of mAb-subgroups to isolates based on the genomic markers on the microarray.

A different aspect of current typing methods is addressed in Part III. Here, NGS data were processed and mapped against an extended scheme called cgMLST. However, the main concern with regard to NGS is the lack of standardized protocols for DNA extraction, sequencing and data interpretation. This problem was tackled by creating a standardized protocol for DNA treatment, sequencing and data processing using a fixed target scheme. The establishment of a genome-wide gene-by-gene typing approach for *L. pneumophila* was initially described by Moran-Gilad and colleagues (2015). However, this work simultaneously evaluates its detailed practical applicability during outbreak investigations in Germany.

Part II - Chapter I

A structural comparison of lipopolysaccharide biosynthesis loci of *Legionella pneumophila* serogroup 1 strains

Markus Petzold
Alexander Thürmer
Susan Menzel
Johan W. Mouton
Klaus Heuner
Christian Lück

BMC Microbiology

Petzold et al: A structural comparison of lipopolysaccharide biosynthesis loci of *Legionella pneumophila* serogroup 1 strains. BMC Microbiology 2013 **13**:198

Submitted 24 April 2013 | Accepted 31 August 2013 | Published 4 September 2013

Abstract

Background

The lipopolysaccharide (LPS) is the major immuno-dominant antigen of all *Legionella* species including *L. pneumophila*. Its diversity is the basis for the classification of *L. pneumophila* into serogroups and monoclonal subgroups and is thought to be involved in strain specific virulence. The understanding of the genetic basis of the LPS-antigen is incomplete. Thus, we analyzed the genetic locus involved in LPS-biosynthesis of *L. pneumophila* serogroup 1 (Sg1) strains with the focus on strain specific gene composition.

Results

The LPS-biosynthesis loci of 14 *L. pneumophila* Sg1 strains comprise two distinct regions: A 15 kb region containing LPS-biosynthesis genes that can be found in all *L. pneumophila* strains and a Sg1-specific 18 kb region. The 15 kb region is highly conserved among Sg1 strains as reflected by high homologies of single ORFs and by a consistent ORF arrangement. In contrast, the Sg1 specific 18 kb region is variable and partially disrupted by phage related genes. We propose that the region spanning from ORF 6 to ORF 11 of the Sg1-specific region is likely involved in late LPS-modification. Due to the high variability of this small region and various combinations of single ORFs within this region a strain specific LPS-structure could be synthesized including modifications of legionaminic acid derivates.

Conclusions

Our data clearly demonstrate that the gene structure of the LPS-biosynthesis locus of *L. pneumophila* Sg1 strains show significant interstrain variability. These data can be used for further functional analysis of the LPS synthesis to understand pathogenesis and reactivity with monoclonal antibodies. Moreover, variable but strain specific regions can serve as basis for the development of novel genotyping assays.

Background

Legionella pneumophila is one of 56 described species belonging to the genus *Legionella* of the family *Legionellaceae* (Pearce, et al., 2012). These gram-negative bacteria are ubiquitous inhabitants of natural and manmade aquatic environments where they survive parasitically in protozoa like amoeba (Rowbotham, 1980; Fields, et al., 2002) and in community structures such as biofilms (Declerck, 2010; Stewart, et al., 2012). Additionally, *Legionella* can infiltrate the human lung via inhaled aerosols (Fraser, 1980; Fields, et al., 2002) and subsequently infect alveolar macrophages (Isberg, et al., 2009)

which frequently cause a potential fatal pneumonia termed Legionnaires' disease (LD) (McDade, et al., 1977). *L. pneumophila* strains belonging to the serogroup 1 (Sg1) were predominantly reported in LD cases, especially in community acquired and travel-associated cases (Harrison, et al., 2009; Joseph, et al., 2010b).

Lipopolysaccharide (LPS) is the major immuno-dominant antigen of all *Legionella* species including *L. pneumophila* (Ciesielski, et al., 1986). It is the main component recognized by patient's sera and by diagnostic assays in urinary antigen detection (Helbig, et al., 2012). The LPS molecule possesses a high degree of diversity and thereby provides the basis for the classification of *L. pneumophila* into serogroups and subgroups by monoclonal antibodies (mAb) (Joly, et al., 1983; Helbig, et al., 1997; Helbig, et al., 2007). Sg1 strains are subdivided into nine mAb-subgroups using the Dresden monoclonal antibody panel (Table 1) (Helbig, et al., 2002).

The structural characterization of LPS of *L. pneumophila* identified several specific chemical attributes which differs it from the LPS molecules of other Gram-negative bacteria (reviewed in Knirel and Valvano, 2011). Particularly the *O*-antigen homopolymer structure consists of an unusual residue, 5-acetamidino-7-acetamido-8-*O*-acetyl-3, 5, 7, 9-tetradesoxy-*D*-glycero-*D*-galacto-nonulosonic acid (legionaminic acid) and its derivates (Knirel, et al., 1994; Zähringer, et al., 1995; Kooistra, et al., 2002a).

A central step in understanding the correlation of the LPS structure and pathogenesis of *L. pneumophila* was the description of the genetic background of LPS molecules by Lüneberg and colleagues (Lüneberg, et al., 2000). More precisely, a genetic locus composed of at least 28 open reading frames (ORF) is essential in LPS core oligosaccharide biosynthesis and LPS O-chain biosynthesis. The genes of this 31-36 kb cluster have characteristic functions required for the synthesis, transport, translocation and modification of LPS components. The *lag-1* gene of this biosynthesis locus encodes for an *O*-acetyltransferase which is responsible for the 8-*O*-acetylation of legionaminic acid (Zou, et al., 1999). Strains carrying a functional *lag-1* synthesize an LPS epitope that reacts with

Table 1: Monoclonal antibody-based subgrouping of *L. pneumophila* Sg1 strains using the Dresden panel

mAb-subgroup	8/5	3/1	3	8/4	10/6	20/1	26/1	30/4
Philadelphia	+	+	-	+	-	-	-	-
Allentown/France	+	+	-	-	-	-	-	+/-[a]
Benidorm	+	+	-	-	-	+	-	-
Knoxville	+	+	+	-	-	-	-	-
OLDA	+	-	-	+	-	+/-[a]	+	+/-[a]
Oxford	+	-	-	+	-	-	-	+/-[a]
Heysham	+	-	+	-	-	-	-	-
Camperdown	+	-	-	-	-	-	+	+
Bellingham	+	-	-	-	+	+	+	+

: Variable results with different strain

the mAb 3/1 (initially named mAb 2 (Joly, et al., 1986)) of the Dresden monoclonal antibody panel. This epitope is assumed to contribute to an increased virulence (Zou, et al., 1999) since mAb 3/1$^+$ strains represent the most prominent subgroup of clinical *Legionella* isolates. In contrast, strains lacking *lag-1* carry mainly deacetylated LPS molecules. These mAb 3/1$^-$ strains comprise only a small number of clinically identified *L. pneumophila* strains in immunocompetent patients (Harrison, et al., 2009; Joseph, et al., 2010a).

Besides the mAb 3/1-specific O-acetylation of the legionaminic acid epitope, to date it remains elusive how strain specific mAb-reactivities can be explained. Increased understanding of the genetic background and structural LPS properties of the different Sg1 strains could help to comprehend subgroup distributions among clinical and environmental isolates (Helbig, et al., 1995; Helbig, et al., 2002; Harrison, et al., 2006; Harrison, et al., 2009; Reimer, et al., 2010; Amemura-Maekawa, et al., 2012) and would deliver more insight in the role of LPS in the *L. pneumophila* life cycle.

To achieve this goal, we analyzed the LPS-biosynthesis loci of at least one member of each mAb-subgroup (excluding mAb-subgroup Oxford) of the *L. pneumophila* Sg1. In this study we focused on the genetically composition of the loci and putative genotype-phenotype correlations according to the Dresden panel of mAbs.

Methods

Phenotypic and genotypic characterization of *L. pneumophila* strains

Legionella pneumophila Sg1 strains Camperdown 1 (ATCC 43113), Heysham 1 (ATCC 43107) (Joly, et al., 1986), Uppsala 3 (Bernander, et al., 2003) and Görlitz 6543 (Thürmer, et al., 2009) were grown on buffered charcoal yeast extract (BCYE) agar plates (Oxoid, Wesel, Germany) for 48 h at 37 °C under a 5 % CO_2 atmosphere. Monoclonal subgrouping was accomplished using the Dresden panel of mAb as described elsewhere (Helbig, et al., 1997; Helbig, et al., 2002).

DNA extraction and sequence generation

DNA was extracted using the EZ1 DNA Tissue Kit (Qiagen, Hilden, Germany). Prior to sequencing DNA fragments of the LPS-biosynthesis locus were PCR-amplified using GoTaq polymerase (Promega, Madison, US-WI) and LPS-specific primers (Suppl. Table 2) which were designed based on published *L. pneumophila* genomes. Initial denaturation was carried out at 95 °C for 2 min followed by 30-35 cycles: 95 °C denaturation for 30 s, annealing at various temperatures for 1 min and elongation at 72 °C for 1 min/kb. Final elongation for 5 min at 72 °C completed the amplification protocol. The PCR result was

checked on 1.5 % agarose gel with 5 V/cm (LE Agarose, Biozym, Oldendorf, Germany) and purified (MSB Spin PCRapace, Stratec, Birkenfeld, Germany) for sequence reaction.

Sequencing reactions were accomplished by a cycle-sequencing procedure on an automated DNA sequencing machine (ABI Prism 377, Applied Biosystems, Waltham; US-MA).

The LPS-biosynthesis locus of the strain L10/23 was sequenced during a whole genome sequencing project. This strain was isolated during a cooling tower related outbreak in Ulm (Germany) in 2010 (von Baum, et al., 2010).

Sequence annotation and analysis

Obtained sequences of Camperdown 1, Heysham 1, Uppsala 3, Görlitz 6543 and L10/23 were assembled using SeqMan (Lasergene 8, DNASTAR, Madison, US-WI) and controlled against public databases using BLAST (Altschul, et al., 1997). ORF annotation of all analyzed strains was accomplished with GeneMark.hmm (Lukashin and Borodovsky, 1998) and Artemis (Rutherford, et al., 2000). The annotation and the numbering of ORFs was based on the initially described LPS-biosynthesis locus of strain RC1 (mAb-subgroup OLDA) (Lüneberg, et al., 2000) and if possible supplemented by further description of genes, gene products and their putative functions using BLAST, BLASTP (Altschul, et al., 1997; Altschul, et al., 2005), the LegionellaScope database of the MicroScope Microbial Genome Annotation Platform (Vallenet, et al., 2009) and the Conserved Domain Database (Marchler-Bauer, et al., 2011). Since Lüneberg et al. analyzed the strain RC1 which had 30 ORFs the numbering of ORFs in other *L. pneumophila* Sg1 strains with deviating ORF numbers is not continual (Lüneberg, et al., 2000). The genes *iraA* (ORF 29) and *iraB* (ORF 30) were not taken into account as part of the LPS-biosynthesis locus. Both formed a small 2-gene operon responsible for iron assimilation, infection and virulence (Viswanathan, et al., 2000).

The putative coding regions were compared to already known LPS-biosynthesis ORFs of published *L. pneumophila* strains using the SeqMan program. The LPS-biosynthesis clusters of the strains were deposited in the EMBL database under the number [EMBL: *HE980447*] for strain Camperdown 1 (mAb-subgroup Camperdown), [EMBL: *HE980446*] for strain Heysham 1 (mAb-subgroup Heysham), [EMBL: *HE980445*] for strain Uppsala 3 (mAb-subgroup Knoxville), [EMBL: *HF678227*] for strain Görlitz 6543 (mAb-subgroup Bellingham) and [EMBL: *HF545881*] for strain L10/23 (mAb-subgroup Knoxville) (Table 2).

Sequence homologies of single ORFs were calculated based on multiple alignments using BioNumerics 6.0 (Applied Maths NV, Sint-Martens-Latem, Belgium) and BLASTP

(Altschul, et al., 2005). Cluster analysis was performed using the UPGMA method of the BioNumerics 6.0 software package.

The sequences of other LPS-biosynthesis loci were obtained from complete genomes of the following strains: Paris (mAb-subgroup Philadelphia) (GenBank: *NC_006368.1*), Lens (mAb-subgroup Benidorm) (GenBank: *NC_006369.1*), Philadelphia 1 (mAb-subgroup Philadelphia) (GenBank: *NC_002942.5*), Alcoy 2300/99 (mAb-subgroup Knoxville) (GenBank: *NC_014125.1*), Corby (mAb-subgroup Knoxville) (GenBank: *NC_009494.2*), Lorraine (mAb-subgroup Allentown) (EMBL: *FQ958210*), HL 06041035 (mAb-subgroup Bellingham) (EMBL: *FQ958211*), RC1 (mAb-subgroup OLDA) (EMBL: *AJ277755*) and 130b (mAb-subgroup Benidorm) (EMBL: *FR687201.1*) (Table 2) (Lüneberg, et al., 2000; Chien, et al., 2004; Cazalet, et al., 2008; Glöckner, et al., 2008; D'Auria, et al., 2010; Schroeder, et al., 2010; Gomez-Valero, et al., 2011a). Since the genome of 130b is a draft version we closed a sequencing gap in scaffold 4 (position 918107 to 918206) using PCR and sequencing.

Table 2: LPS biosynthesis loci obtained from sequenced genomes of *L. pneumophila* Sg1 strains

Strain	mAb subgroup	Accession no.	Reference
Alcoy 2300/99	Knoxville	GenBank: NC_014125.1	(D'Auria, et al., 2010)
Corby	Knoxville	GenBank: NC_009494.2	(Glöckner, et al., 2008)
L10/23 (Ulm)*	Knoxville	EMBL: HF545881	this study
Uppsala 3*	Knoxville	EMBL: HE980445	this study
Paris	Philadelphia	GenBank: NC_006368.1	(Cazalet, et al., 2004)
Philadelphia 1	Philadelphia	GenBank: NC_002942.5	(Chien, et al., 2004)
HL 0604 1035	Bellingham	EMBL: FQ958211	(Gomez-Valero, et al., 2011b)
Görlitz 6543*	Bellingham	EMBL: HF678227	this study
Camperdown 1*	Camperdown	EMBL: HE980447	this study
Heysham 1*	Heysham	EMBL: HE980446	this study
130b (Wadsworth)	Benidorm	EMBL: FR687201 (FH93HPE01)	(Schroeder, et al., 2010)
Lens	Benidorm	GenBank: NC_006369.1	(Cazalet, et al., 2004)
Lorraine	Allentown	EMBL: FQ958210	(Gomez-Valero, et al., 2011b)
RC1*	OLDA	EMBL: AJ277755	(Lüneberg, et al., 2000)

* only LPS biosynthesis locus sequenced

Results and discussion

Two regions within the LPS-biosynthesis locus

To gain insight into the genetic composition and arrangement of the LPS biosynthesis locus we analyzed the loci of 14 *L. pneumophila* Sg1 strains. The strains represent members of all mAb–subgroups that can be distinguished by the Dresden monoclonal antibody panel (Table 1) besides the extremely rare mAb-subgroup Oxford. The

LPS biosynthesis loci of five strains were newly sequenced for this study. These were: Camperdown and Heysham 1 of the rarely found subgroups of the same name (Harrison, et al., 2009; Amemura-Maekawa, et al., 2012) and the strains Uppsala 3, Görlitz 6543 and L10/23. Eight LPS biosynthesis loci were obtained from complete genomes that have been published previously. Furthermore, for strain RC1 (mAb subgroup OLDA) the biosynthesis locus was available as well (Table 2).

The LPS-biosynthesis locus of each of the analyzed *L. pneumophila* Sg1 strains contained at least 28 ORFs and ranged in size from 30,644 bp (strain Lorraine) to 35,888 bp (strain 130b) with an average locus size of 33,398 bp respectively. The average ORF size within the locus was approximately 1 kb. The complete LPS-biosynthesis locus had a slightly lower % GC content (~35 %) than the adjacent regions (~38 %) and the total genome (~38.5 %), respectively (Suppl. Table 4).

Structural and comparative analysis of the loci confirmed a highly conserved 15 kb region from *wecA* (ORF 14) to lpg0748 (ORF 28) according to the Philadelphia1 genome as shown previously (Cazalet, et al., 2008). Additionally, all ORFs within this region were consistently orientated into the same direction (Figure 4).

A second region within the locus of 18 kb in size is spanning from lpg0779 (ORF 1) to lpg0764 (ORF 13). Here, the structural composition and the orientation of ORFs as well as the total amount of putative ORFs was less conserved (Table 3). Interestingly, many transposases and phage related genes were present in 8 strains (Figure 4).

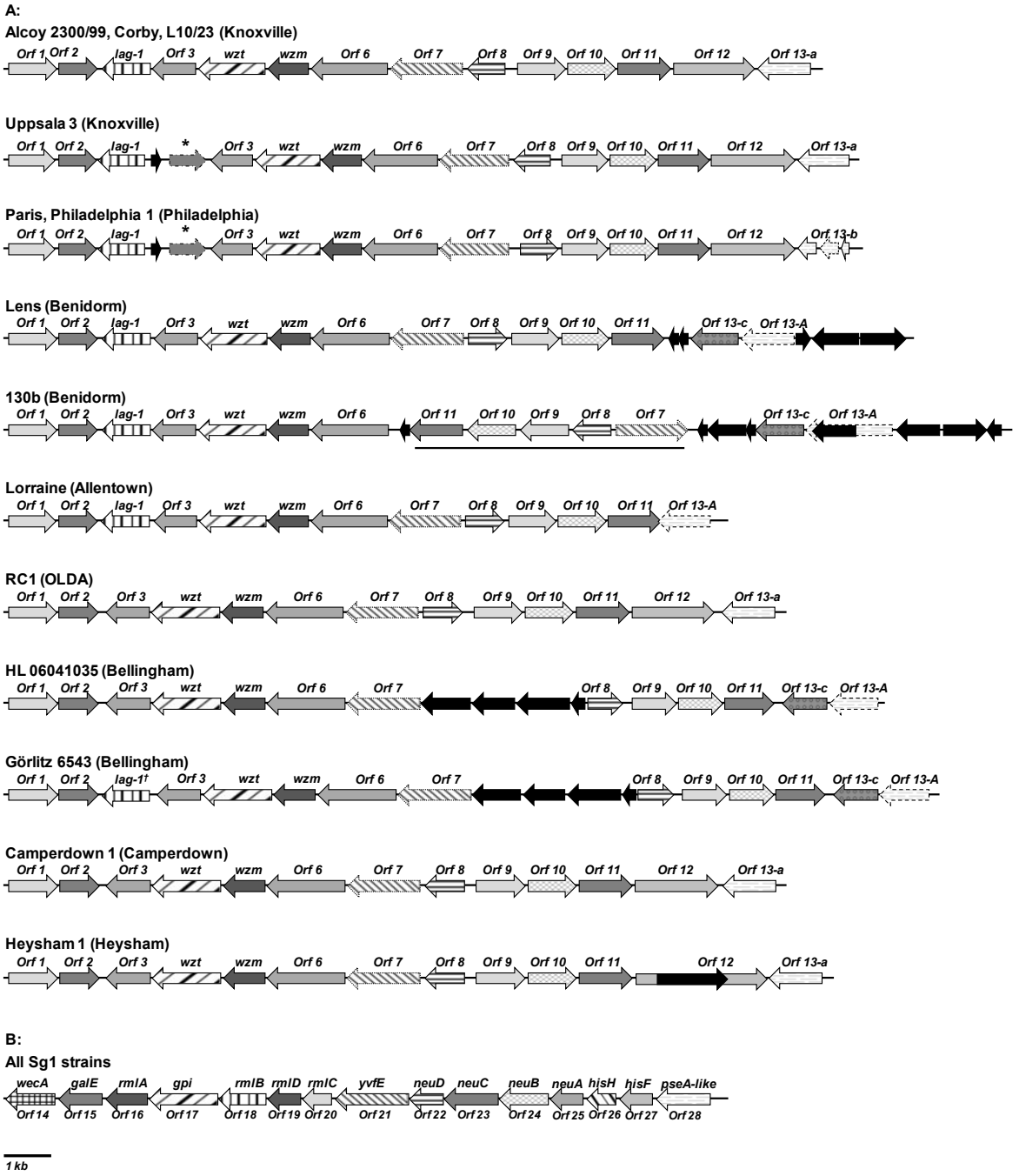


Figure 4: Structural representation of the LPS-biosynthesis locus. Shown are the LPS-biosynthesis loci of 14 *L. pneumophila* Sg1 strains and the corresponding monoclonal subgroup (in brackets). Strains Alcoy 2300/99, Corby and L10/23, and Paris and Philadelphia 1, respectively had the same genetic structure and monoclonal subtype and were therefore shown in one scheme. The numbering of ORFs was adopted by (Lüneberg, et al., 2000). **A**: shows the Sg1-specific 18 kb region (ORFs 1-13) and **B**: shows the 15 kb region (ORFs 14-28). The direction of transcription is indicated by arrowheads. The filled black arrows indicate transposases/phage-related proteins. Grey shades and hatched patters serve to distinguish ORFs. Asterisk in Uppsala 3, Philadelphia 1 and Paris represents a partial ORF 2 duplication (ORF 2 like) as described by (Bernander, et al., 2003). Underlined ORFs 7–11 in strain 130b represent an inversion. Görlitz 6543 carries a truncated *lag-1* marked with †.

Table 3: Amino acid similarity of the *L. pneumophila* Sg1-specific LPS biosynthesis region from lpg0769-lpg0761 (ORF 1-ORF 15) of strain Philadelphia 1 to other Sg1 strains

Gene name of *L. pneumophila* Sg1 strain Philadelphia 1		Amino acid similarity [%]*												
		Philadelphia[#]	Knoxville[#]				Benidorm[#]		Bellingham[#]		Allentown[#]	OLDA[#]	Camperdown[#]	Heysham[#]
		Paris	2300/99 Alcoy	Corby	Uppsala 3	Ulm	130b	Lens	HL 0604 1035	Görlitz 6543	Lorraine	RC1	Camperdown 1	Heysham 1
lpg0761 (galE)	ORF 15	100	100	100	100	100	97.1	96.0	99.8	99.8	99.8	100	100	100
lpg0762 (wecA)	ORF 14	100	99.5	99.5	99.5	99.5	93.4	93.1	93.7	93.7	93.4	98.8	99.5	99.5
lpl0803[A]	ORF 13	-	40.3	40.3	40.3	40.3	trans.[c]	100	98.2	98.2	96.6	41.8	40.3	40.3
lpg0765	ORF 12	100	98.6	98.7	98.6	98.6	-	-	-	-	-	98.7	98.6	trans.[c]
lpg0766	ORF 11	100	96.6	96.6	96.6	96.6	93.2	93.2	93.7	93.7	93.1	96.6	96.6	96.6
lpg0767	ORF 10	100	96.2	96.2	96.2	96.2	96.6	97.1	98.9	98.9	97	95.6	96.2	96.2
lpg0768	ORF 9	100	30.6	30.6	30.6	30.6	98.4	99	99	99	98.9	99.4	30.6	30.6
lpg0769	ORF 8	100	31	31	31	31	97.9	97.4	98.4	98.4	97.4	100	31	31
lpg0770	ORF 7	100	90.6	90.6	90.6	90.6	32	31.9	31.9	31.9	99.8	99.9	90.6	90.6
lpg0771	ORF 6	100	38.8	38.7	38.7	38.7	38.8	99.1	100	100	38.8	38.6	99.1	38.7
lpg0772 (wzm)	ORF 5	100	100	100	100	100	100	100	100	100	100	100	100	100
lpg0773 (wzt)	ORF 4	100	99	99.6	100	100	100	99.6	100	99.5	99	99.8	100	100
lpg0774	ORF 3	100	91.6	86.4	98.7	92.1	89	86.4	100	86.4	91.6	99.5	99.8	99.8
lpg0775[a]		100	-	-	100	-	-	-	-		-	-	-	-
lpg0776[b]		100	-	-	100	-	-	-	-		-	-	-	-
lpg0777 (lag-1)		100	96.8	94.9	100	96.8	94.9	94.9	-	94.7[T]	96.8	-	-	-
lpg0778	ORF 2	100	97.9	97.4	100	97.7	97.4	97.4	99.6	96.5	97.9	98.9	98.7	98.7
lpg0779	ORF 1	100	99.8	99.1	99.8	99.8	98.9	98.9	100	98.9	99.8	99.4	99.8	99.8

[#] Monoclonal antibody subgroup according to the 'Dresden' panel; * Determined by UPGMA clustering method based on multiple sequence alignment
[A] ORF 13 (lpg0764/lpg0764b/lpg0763) of Philadelphia 1 not displayed, ORF 13-A of strain Lens was used, [a] Partial duplication of ORF 2 (*lpg0778*), [b, c] Transposase
[†] *Lag-1* of Görlitz 6543 has no functional start codon
Underlined numbers indicate different clusters of corresponding ORF (see also Fig. 4)
The highly conserved 15 kb region (ORF14 – ORF 28) is not completely shown and only reflected by WecA and GalE

The heterogeneous nature of the 18 kb region and the extremely high conserved 15 kb region found in our study are largely in agreement with earlier results. These proposed to separate the locus into a Sg1-specific and a *L. pneumophila*-specific region (Cazalet, et al., 2008; Mérault, et al., 2011). Microarray analysis of Sg1 and non-Sg1 strains have identified a 13 kb region (ORF 16-28) which is present in all *L. pneumophila* strains and a 20 kb region (ORF 1-15) generally found in all Sg1 strains (Cazalet, et al., 2008). The two regions were defined based on the LPS-biosynthesis loci of the Sg1 strain Paris (Cazalet, et al., 2004).

To determine the putative breakpoint between both regions is difficult. However, based on our analysis of the structural composition we would rather separate the LPS biosynthesis locus between lpg0763 (ORF 13) and *wecA* (ORF 14). This is in agreement with recent data, since the genes *wecA* (ORF 14) and *galE* (ORF 15) were demonstrated to be present in non-Sg1 strains with lower amino acid similarities when compared to Sg1 strains (55-61%) (Mérault, et al., 2011).

The previously mentioned ORF 13 is located next to the breakpoint region. In total, four different types of ORFs were found in the analyzed region of Sg1 strains here named ORF 13-a, -b, -c and -A. In each of the strains Lens, 130b, HL 06041035 and Görlitz 6543 two ORFs were found. These strains carried a putative conserved protein of unknown function (here referred to as ORF 13-A). A transposase-disrupted ORF 13 A was present in strain 130b (Figure 4). Additionally, the strains carried an ORF which shared features of the radical *S*-adenosylmethionine (SAM) superfamily (CDD: cd01335) named ORF 13-c (Suppl. Table 3). Interestingly, all these strains lacked the ORF 12. However, even though the strain Lorraine lacked ORF 12 as well, it carried only a single ORF 13-A variant.

A distinct ORF of unknown function with amino acid similarity to ORF 13-A of only 38 %, here named ORF 13-a, was present in the remaining strains with the exceptions of a truncated form in strains RC1, Philadelphia 1 and Paris. Philadelphia 1 and Paris shared high similarities with ORF 13-a but a deletion led to a frame shift resulting into three smaller fragments (pooled as ORF 13-b).

A conserved region found in all serogroup 1 strains

Within the conserved region several genes were found which are proposed to be involved in the biosynthesis of the highly acetylated core region which is composed of mannose, *N*-acetyl-glucosamine (GlcNAc), *N*-acetyl-quinovosamine (QuiNAc) and rhamnose residues (Zähringer, et al., 1995). A vast number of ORFs, more specifically ORF 21 through 25 and 28, were recently reported to facilitate the biosynthesis of the repetitive legionaminic acid residues of the *O*-antigen (Knirel, et al., 1994; Glaze, et al., 2008). The pyrodoxal-phosphate dependent aminotransferase (ORF 21), the

acetyltransferase *neuD* (ORF 22) and a dehydratase (lpg0966) located outside of the locus are likely to synthesize the precursor molecule of legionaminic acid, UDP-*N*,*N*'-diacetylbacillosamine (UDP-Bac2Ac4Ac) (Schoenhofen, et al., 2006). Contradictory to our findings, functions of the *neuD* products are described highlighting that the acetyltransferase is involved in Lag-1-independent *O*-acetylation of few legionaminic acid residues close to the LPS-core of *L. pneumophila* (Lüneberg, et al., 2000; Knirel, et al., 2003; Lewis, et al., 2006). Furthermore, the adjacent genes encoding for NeuC (ORF 23), NeuB (ORF 24) and NeuA (ORF 25) were recently identified to be involved in the subsequent biosynthetic processes converting UDP-Bac2Ac4Ac into CMP-5,7-diacetamido-3,5,7,9-tetradeoxy-*D*-glycero-*D*-talo-nonulosonic acid (CMP-Leg5Ac7Ac) (Glaze, et al., 2008). Moreover, the ORF 28 is homologous to the *ptmG* gene of *Campylobacter jejuni* (Cj1324) which converts the CMP-Leg5Ac7Ac residue to CMP-5-acetamidino-7-acetamido-3,5,7,9-tetradeoxy-*D*-glycero-*D*-talo-nonulosonic acid (CMP-Leg5Am7Ac) (McNally, et al., 2007), the dominant residue of the *O*-antigen of non-Sg1 strains of *L. pneumophila* (Knirel, et al., 2001). A functional correlation of the ORFs of this region is supported by recent transcriptomic data of strain Paris in which the ORFs 21-17 and 28-22 were transcribed as operons (Sahr, et al., 2012). Since all analyzed Sg1 strains and a broad number of non-Sg1 strains carry ORF 28 it can be assumed that CMP-Leg5Am7Ac is a common residue of the *L. pneumophila* LPS molecule which might subsequently become modified in a mAb-subgroup or even strain specific manner (Ledesma, et al., 1995; Farhat, et al., 2011; Mérault, et al., 2011).

Three clusters of the *O*-acetyltransferase Lag-1

A well examined phenotype variation is linked to the presence and absence of the *lag-1* gene. *Lag-1* encodes for an *O*-acetyltransferase that conferred reactivity with mAb 3/1 and is exclusively found in Sg1 strains. Our results revealed three clusters of the *lag-1* genes, although without any detectable relation to the mAb-subgroup switch which supports recent findings (Kozak, et al., 2009) (Figure 5). *Lag-1* was previously reported to be involved in mAb-subgroup switches of different strains. However, this was generally due to gene deletion or loss-of-function mutations of *lag-1* (Lück, et al., 2001; Bernander, et al., 2003; Wagner, et al., 2007; Thürmer, et al., 2009). Complete and functional *lag-1* genes were present in all mAb $3/1^+$ strains and were absent in all mAb $3/1^-$ strains. Besides that, the Philadelphia subgroup strains (Philadelphia 1 and Paris) as well as the Knoxville-subgroup strain Uppsala 3 carried a transposase and a partial duplication of ORF 2 adjacent to *lag-1*. Bernander et al. reported the region from ORF 2 to ORF 3 as unstable (Bernander, et al., 2003). Looping-out of the intermediate located *lag-1* gene is assumed to be a potential consequence. Under *in vitro* conditions the deletion of the *lag-1* gene occurred at with

frequency of 10^{-6} to 10^{-7} (C. Lück, unpublished results). Detailed analysis of the region from ORF 2 to ORF 3 including *lag-1* of these strains revealed remarkably high similarities of Uppsala 3 to the Philadelphia-subgroup strains Philadelphia 1 and Paris (>98-100 %) whereas the remaining Knoxville-subgroup strains clustered in a different group (Table 3; Figure 5). The high similarity of this 4 kb region between strain Uppsala 3 and the strains Paris and Philadelphia 1 may indicate horizontal gene transfer of this region. However, this had no impact on the specific mAb reactivity for all other analyzed Knoxville-subgroup strains. Horizontal gene transfer between strain Paris and Philadelphia 1 was recently reported for a large genome fragment which also harbored the LPS biosynthesis locus (Gomez-Valero, et al., 2011b). These observations are confirmed by our results since the LPS biosynthesis loci of both strains are almost identical in composition, arrangement and amino acid similarity (Suppl. Table 3). A truncated *lag-1* gene was found in the strain Görlitz 6543 (mAb-subgroup Bellingham) as recently reported (Thürmer, et al., 2009). The whole gene is present but carries a mutated start codon. Since Görlitz 6543 showed no reactivity with mAb 3/1 it was assumed that the mutation significantly impairs the production of a functional *O*-acetyltransferase. Phylogenetic analysis showed 99.9 % amino acid similarity of Görlitz 6543 to Corby (mAb-subgroup Knoxville), 130b and Lens (both mAb-subgroup Benidorm) (Figure 5).

ABC-transporter genes *wzt* and *wzm* as Sg1-specfic marker region

Noticeable conserved genes within the heterogenic region were *wzt* (ORF 4) and *wzm* (ORF 5) which are almost identical among all analyzed Sg1 strains (Table 3). *Wzm* encodes for a protein containing a transmembrane domain while *wzt* encodes for a

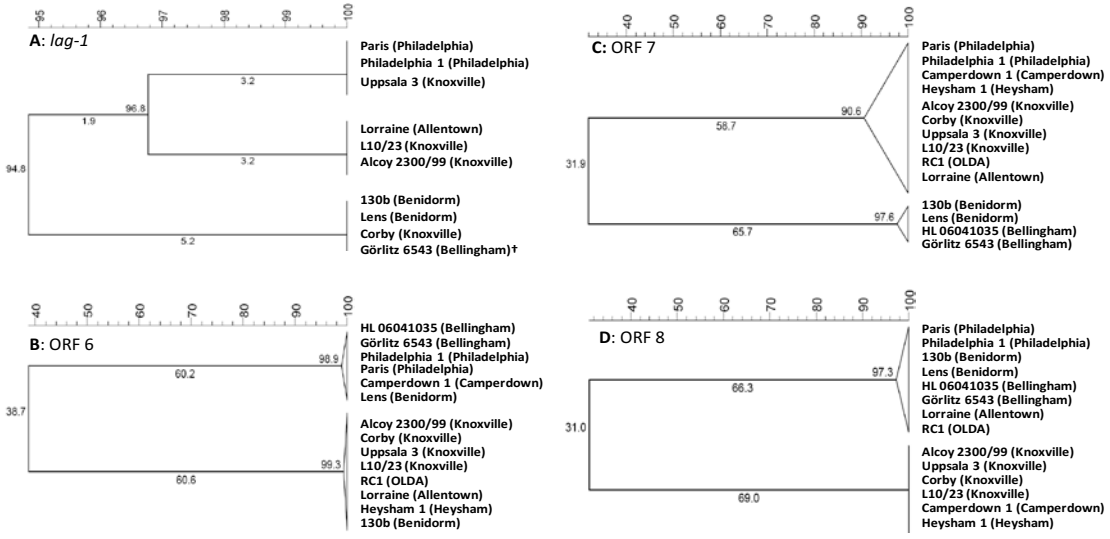


Figure 5: Dendrogram of variable ORFs. Multiple amino acid based cluster analysis using UPGMA (BioNumerics, Applied Maths NV, Belgium). The phylogenetic trees of gene *lag-1* and of the ORFs 6, 7 and 8 are shown. ORF 9 is identical to the phylogenetic tree of ORF 8 and is therefore not shown. Similarity values and branch distances were depicted in percentages [%]. The strain-specific mAb-subgroup is indicated in brackets. The mutated start codon of *lag-1* of Görlitz 6543 was neglected for similarity analysis and is indicated with †.

nucleotide binding domain of an ABC transporter system which mediates the *O*-antigen translocation across the inner membrane (Greenfield and Whitfield, 2012). Recently, both genes were evaluated as marker genes for PCR based discrimination between *L. pneumophila* Sg1 and non-Sg1 strains (Mérault, et al., 2011). The ABC transporter-dependent *O*-antigen pathway interacts with WecA (ORF 14), an UDP-GlcNAc-1-transferase which initiates *O*-chain biosynthesis at the cytoplasmic site of the inner membrane (Greenfield and Whitfield, 2012). The low amino acid similarity of WecA between Sg1 and non-Sg1 that was described recently combined with the absence of *wzm* and *wzt* in non-Sg1 genomes indicate a different *O*-chain biosynthesis mechanism for non-Sg1 strains than found in Sg1 strains (Mérault, et al., 2011).

ORF 6 through 11 involved in *O*-antigen modification

The most variable region within the Sg1-specific region in terms of low similarities on the amino acid level and the diverse arrangement of single ORFs was found from ORF 6 to ORF 11. The strains of mAb-subgroup Benidorm 130b and Lens were almost identical regarding the amino acid similarities of the single ORFs within the Sg1-specific region. Interestingly, strain 130b carried a large inverted fragment containing ORF 7 to ORF 11 (Figure 4). This region was surrounded by transposases suggesting their potential contribution to the inversion. Since the strain 130b showed no altered reactivity pattern using the Dresden panel compared to other Benidorm strains it could be stated that the inversion had no detectable effect on the LPS phenotype detected by monoclonal antibodies.

The adjacent ORF 6 showed a high degree of variability between *L. pneumophila* Sg1 strains represented by two clusters of low amino acid similarity (<39 %) (Table 4). Interestingly, the two analyzed strains of the mAb-subgroup Benidorm, 130b and Lens, cluster into two distinct groups. This either indicates that the product of ORF 6 has probably no effect on the LPS structure of strains of the same monoclonal subgroup or that it has the same function despite low similarity.

However, ORF 6 products might be involved in the establishment of a mAb-subgroup discriminating epitope. More precisely, only the mAb-subgroups Heysham and Knoxville react with mAb 3. This indicates a similar epitope which in turn could possibly be traced back to specific ORFs within the Sg1-specific region. However, strains of both mAb-subgroups were highly homologous regarding the whole LPS-biosynthesis with the exception of *lag-1* which is present in Knoxville strains (Table 3, Table 4). In addition, the strain Camperdown 1, not reacting with mAb 3, carried a very similar LPS-biosynthesis locus as Heysham 1 and the Knoxville strains. It is the single ORF 6 in which Camperdown 1 clusters differently to Heysham 1. It can be assumed that the combination

of ORF 6 to 9 which is exclusively found in Knoxville and Heysham strains leads to reactivity with mAb 3. Another ORF 6 as found in the genetically very similar strain Camperdown 1 could alter the LPS epitope and is thereby not recognized by mAb 3. Furthermore, the mAb 3 epitope was not influenced by *O*-acetylation of the legionaminic acid residue since the Knoxville strains were mAb 3/1$^+$ and carried the *lag-1* gene whereas the strain Heysham 1 is negative for both markers.

Modification of legionaminic acid in transposon mutants

Two additional ORFs, ORF 8 and ORF 9, within in the highly variable region from ORF 6 to ORF 11 are most likely involved in O-antigen modification. The genetic nature of the ORF 8 products displayed two different clusters which was comparable to the clustering of ORF 9. Both clusters share poor amino acid similarities of 31 % (ORF 8) and 30.7 % (ORF 9) (Table 3, Figure 5). These differences in amino acid similarity were also reflected by the ORF orientation. Both ORFs were orientated into opposite directions in strains of the mAb-subgroups Knoxville, Camperdown and Heysham which form a separate cluster in both ORFs (Figure 4). For the remaining mAb-subgroups (Philadelphia, Allentown, Benidorm, Bellingham and OLDA) the ORFs are oriented into identical directions. *In silico* analysis of these loci predicted a five-gene operon from ORF 8 to ORF 12 suggesting a coupled functional entity (Price, et al., 2005). These strains were also grouped into a single cluster. However, recent transcriptomic data obtained from strain Paris revealed a four-gene operon which lacks ORF 8 (Sahr, et al., 2012). For all strains regardless of the distance in the phylogenetic tree BLASTP predicted a methyltransferase function for ORF 8 and a siliac acid synthetase function (neuB family) for ORF 9 (Lüneberg, et al., 2000; Kooistra, et al., 2002b; Wagner, et al., 2007).

It is reported that the putative methyltransferase encoded by ORF 8 is responsible for *N*-methylation of the 5-acetimydyol amino group of the legionaminic acid (Kooistra, et al., 2002b; Wagner, et al., 2007). This is supported by studies on the legionaminic acid pathway of *Campylobacter*. The *ptmH* gene (Cj1325) of *C. jejuni* is a homologue of ORF 8 of the Knoxville, Camperdown and Heysham subgroup cluster (Figure 5) (McNally, et al., 2007). The *ptmH* product catalyzes the modification of CMP-Leg5Am7Ac to the *N*-methylated residue CMP-5-acetimidoyl (*N*-methyl) amino-7-acetamido-3,5,7,9-tetradeoxynon-2-ulosonic acid (CMP-Leg5AmNMe7Ac), the main residue of the Sg1 *O*-antigen. Disruption of ORF 8 in the Bellingham-subgroup strain Görlitz 6543 led to loss-of-reactivity with the Bellingham-subgroup specific mAb 10/6 and mAb 20/1 and resulted in a mAb-subgroup switch from subgroup Bellingham to Camperdown. In similar mutants of the mAb 3/1$^+$ strain 130b the reactivity with mAb 20/1 was also lost when ORF 8 or ORF 11 was disrupted leading to a switch from mAb-subgroup Benidorm to Allentown. The wild type

strains 130b and these mutants did not react with mAb 10/6. This supported the assumption that the mAb 3/1-specific epitope generated by the *O*-acetyltransferase Lag-1 masks the *N*-methyl group and hinders binding of mAb 10/6 (Wagner, et al., 2007). This is in agreement with earlier observations which reported a correlation between ORF 8 and *N*-methylated legionaminic acid residues for the mAb 3/1^{-} strain RC1 (Kooistra, et al., 2002b). However, the fact that mutants of both strains, 130b and Görlitz 6543, lost the reactivity with mAb 20/1, indicated that ORF 8 and/or ORF 11 are also involved in the generation or modification of another epitope which is not blocked by the *O*-acetyl group.

To find putative ORF candidates, next to ORF 8, that are responsible for synthesis or modification of the common epitope bound by mAb 20/1, we looked for similar but unique ORFs within the Sg1-specific region of Bellingham- and Benidorm-subgroup strains. Phylogenetic analyses identified ORF 7 as a putative subgroup discriminating gene since the mAb-subgroups Benidorm and Bellingham clustered in specific separate group when compared to the other mAb-subgroups (Figure 5). The presence of two different ORF 7 variants is in agreement with recent results obtained by subgroup specific PCR amplification (Thürmer, et al., 2009).

Conclusions

Characterization of the LPS-biosynthesis loci of *L. pneumophila* Sg1 strains revealed two mayor regions: A Sg1-specific region of 18 kb and a conserved 15 kb region containing genes found in Sg1 and non-Sg1 strains. The conserved region carries genes involved in outer core and *O*-chain biosynthesis of LPS molecules.

The variable and heterogeneous Sg1-specific region raised questions concerning the genetic basis for subgroup specific mAb-reactivity. Switches from one monoclonal subtype to another in transposon induced mutants gave a first indication for the function of different gene products. The most variable region from ORF 6 to ORF 11 is likely involved in *O*-antigen modifications and could deliver more insight in the strain specific LPS structures and more important, in strain specific virulence. The ORFs within this region could act in a pathway-like manner explaining the broad variability of the LPS molecule among the Sg1 strains. Furthermore, it is also not excluded that each ORF of this region has an own function in the late modification of legionaminic acid derivates which could be regulated in a life cycle or growth phase-depended way. Further studies using specific mutation in these ORFs, mRNA assays and chemical analysis are required in order to elucidate the role of different genes in the synthesis of the subgroup specific structures in different strains.

Authors' contribution

MP generated sequences of strains Camperdown 1 and Heysham 1, conducted comparative genetic and phylogenic studies, interpreted the results and drafted the manuscript. AT and SM generated sequences of strains Uppsala 3 and Görlitz 6543. KH generated the genome sequence of strain L10/23. JWM reviewed the manuscript. CL conceived and supervised the work, assisted with inspiring discussions and ideas, helped interpreting the results and reviewed the manuscript. All authors read and approved the manuscript.

Acknowledgement

We thank Sigrid Gäbler, Kerstin Lück and Ines Wolf for technical assistance. This work was partly supported by the Robert Koch-Institute grant 1369–364 to Christian Lück.

Dedicated to the memory of Dr. Jürgen Helbig, Dresden, Germany.

Part II - Chapter II

Rapid genotyping of *Legionella pneumophila* serogroup 1 strains by a novel DNA microarray-based assay during the outbreak investigation in Warstein, Germany 2013

Markus Petzold
Ralf Ehricht
Peter Slickers
Stefan Pleischl
Ansgar Brockmann
Martin Exner
Stefan Monecke
Christian Lück

International Journal of Hygiene and Environmental Health

Petzold et al: Rapid genotyping of *Legionella pneumophila* serogroup 1 strains by a novel DNA microarray-based assay during the outbreak investigation in Warstein, Germany 2013. IJHEH 2017 220(4):673-678

Submitted 11 December 2015 | Accepted 12 February 2016 | Published Online 22 February 2016

Abstract

Between 1 August and 6 September 2013, an outbreak of Legionnaires' disease (LD) with 78 cases confirmed by positive urinary antigen tests occurred in Warstein, North Rhine-Westphalia, Germany. *Legionella* (*L.*) *pneumophila*, serogroup (Sg) 1, monoclonal antibody (mAb) subgroup Knoxville, sequence type (ST) 345, was identified as the epidemic strain. This strain was isolated from seven patients. To detect the source of the infection, epidemiological typing of clinical and environmental strains was performed in two consecutive steps. First, strains were typed by monoclonal antibodies. Indistinguishable strains were further subtyped by sequence-based typing (SBT) which is the internationally recognized standard method for epidemiological genotyping of *L. pneumophila*. In an early stage of the outbreak investigation, many environmental isolates were found to belong to the mAb subgroup Knoxville, but to two different STs, namely to ST345, the epidemic strain, and to ST600. A majority of environmental isolates belonged to ST600 whereas the epidemic ST345 strain was less common in environmental samples. To rapidly distinguish both Knoxville strains, we applied a novel typing method based on DNA-hybridization on glass chips. The new assay can easily and rapidly discriminate *L. pneumophila* Sg1 strains. Thus, we were able to quickly identify the sources harboring the epidemic strain, i.e., two cooling towers of different companies, the waste water treatment plants (WWTP) of the city and one company as well as water samples of the river Wester and its branches.

Introduction

Legionellae are ubiquitous gram-negative bacteria, which occupy natural and manmade aquatic environments. They are the causative agents of Legionnaires' disease (LD), which occurs mostly as sporadic cases of pneumonia acquired by inhalation or aspiration of legionellae from different environmental sources, such as cooling towers, whirlpool spas and warm water supplies (Mercante and Winchell, 2015). Currently, the genus *Legionella* comprises 58 species and more than 70 serogroups with *Legionella* (*L.*) *pneumophila* serogroup (Sg) 1 causing the majority of human infections (Helbig, et al., 2002; Harrison, et al., 2009; Parte and Euzéby, 2017). In general, *Legionella* spp. are responsible for 2-20 % of community acquired pneumonia cases (CAP) (Torres, et al., 2014). Approximately 10 % of cases occur in outbreaks or clusters (European Centre for Disease Prevention and Control, 2015b). *Legionella* outbreaks may present a serious threat to public health; and management of outbreaks can be extremely difficult especially if the source of infection is unknown (McAdam, et al., 2014).

Epidemiological typing, i.e., comparison of clinical and environmental isolates of *L. pneumophila*, is mandatory to identify the source of an outbreak (von Baum, et al., 2010). Usually, this is performed in two consecutive steps. First, isolates are typed by monoclonal

antibodies (mAb) (Helbig, et al., 2002). Indistinguishable isolates are further genotyped by sequence-based typing (SBT) which is the internationally recognized standard method for epidemiological genotyping of *L. pneumophila*. This is performed by partial sequencing of seven genes and comparison of these alleles to a database to assign a sequence type (ST) (ESGLI, 2015).

Isolates reacting with the mAb 3/1 belong to the subgroups of *L. pneumophila* strains causing the majority of community acquired and travel associated cases (Lück, 2011) as well as all cooling tower associated outbreaks yet observed (Walser, et al., 2014).

The epidemic strain of the outbreak in Warstein was characterized as *L. pneumophila* Sg1, mAb subgroup Knoxville ST345, which was isolated from seven clinical samples. However, even knowing the epidemic strain, the identification of the source proved to be difficult since the majority of environmental samples yielded enormous *L. pneumophila* counts and multiple different *L. pneumophila* Sgs, mAb subgroups and STs. SBT was not practical and next-generation sequencing was not possible given the high number of samples and isolates, the limited resources and the urgency in investigating a large-scale and high-profile outbreak. Therefore, we used a *L. pneumophila* Sg1-specific DNA-microarray as a fast and discriminatory tool to investigate outbreak isolates.

Materials and Methods

Patient samples investigated during the outbreak

Between 1 August and 6 September 2013 an outbreak of LD with 78 laboratory confirmed cases occurred in the town of Warstein, North Rhine-Westphalia, Germany. All these patients had at least one positive urine sample tested with the Binax ELISA (Alere, Cologne, Germany). During this outbreak, respiratory specimens from ten patients were obtained and cultured on buffered charcoal yeast extract (BCYE) and BCYE agar plates supplemented with cefamandole (4 mg/l), polymyxin B (80.000 IU/l) and anisomyxin (80 mg/l; BMPA medium; Oxoid, Wesel, Germany) at 1:10 and 1:100 dilutions with and without heat pretreatment (3 min at 60 °C) for up to ten days at 37 °C at 5 % CO_2 atmosphere. Eight *Legionella* isolates could be obtained from seven patients. Isolates were initially serotyped using a latex agglutination test (Oxoid) and further subtyped using the ‘Dresden panel’ of mAbs (Helbig, et al., 2002). Genotyping was performed according to the SBT method (ESGLI, 2015). Respiratory samples from three culture-negative samples were investigated using the *L. pneumophila* serogroup 1-specific PCR (Mérault, et al., 2011) and the direct nested SBT scheme (Mentasti, et al., 2014).

Environmental isolates obtained during the outbreak

A variety of environmental samples were taken in order to identify the source of the outbreak. Those included a cooling tower (CT) from a company (source A), another CT and an industrial sewage pretreatment plant (source B), the municipal water supply system and the municipal waste water treatment plant (WWTP) (source C), as well as the river Wester and its branches (source D). In total, 880 environmental samples were initially cultured and screened for the presence of *Legionella* spp. by the Institute of Hygiene and Public Health at the University of Bonn according to ISO 11731/1998 (ISO 11731, 1998). *Legionella*-positive culture plates were immediately sent to the German national reference laboratory for *Legionella* in Dresden for further subtyping as described above.

L. pneumophila isolates tested to validate the newly developed DNA-based microarray

In total, 163 epidemiologically unrelated strains and isolates were tested (Suppl. Table 5) including 80 European *Legionella* (EUL) reference strains used for the validation of the SBT scheme (Gaia, et al., 2005; Ratzow, et al., 2007) and isolates from the strain collection of the German national reference laboratory for *Legionella* that were originally isolated from patients (n=57) and water systems (n=24). 23 of these strains were related to five confirmed case clusters or outbreak situations (Table 4). In addition, strains were included for which complete genome sequences are publicly available. These were the Sg1 strains Philadelphia-1 (GenBank *NC_002942.5*) (Chien, et al., 2004), Lens (GenBank *NC_006369.1*), Paris (GenBank *NC_006368.1*) (Cazalet, et al., 2004) and Corby (GenBank *NC_009494.2*) (Glöckner, et al., 2008).

Table 4: Set of epidemiological related strains isolated during outbreak or single case investigations for microarray validation

Set	number patients/ clinical isolates/ environmental strains	mAb subgroup	sequence type (ST)	allele formula (*flaA, pilE, asd, mip, mompS, proA, neuA*)	match clinical and environmental microarray typing >93%	pattern similar to other strains	reference
A	7(8)/31	Knoxville	345	6, 10, 19, 3, 19, 4, 11	yes	unique	(Maisa, et al., 2015)
B	6/0	Knoxville	182	3, 4, 1, 3, 35, 9, 11	yes	unique	(Weist, et al., 2006)
C	0/35	Knoxville	600	6, 10, 19, 28, 19, 14, 11	yes	unique	(Maisa, et al., 2015)
D	2/1	Allentown/France	82	5, 1, 22, 10, 6, 10, 6	yes	unique	this study
E	3/4	Knoxville	62	8, 10, 3, 15, 18, 1, 6	yes	unique	(von Baum, et al., 2010)
F	1/2	Philadelphia	332	7, 10, 17, 6, 14, 11,3	yes	unique	this study
G	2/2	Knoxville	35	3, 4, 1, 1, 1, 9, 1	yes	unique	(Feeley, et al., 1978; Beyrer, et al., 2007)

DNA Microarray Design

The primers and probes used in the newly developed DNA-microarray are listed in the supplemental table 6 and 7. These sequences were derived from published sequences using the Array Design software package (Alere Technologies, Jena, Germany). We used variants of 25 different target genes which are part of the LPS-biosynthesis cluster (Thürmer, et al., 2009) as well as variable genetic elements absent or present in the genomes of *L. pneumophila* Sg1 strains (Pannier, et al., 2010). Additionally, allelic variants of the *neuA* and the *pilE* genes were included that also are used in the SBT scheme (Ratzow, et al., 2007).

Based on published sequences of the specific target regions mentioned above and on the complete genome sequences of the Sg1 strains Philadelphia-1 (Chien, et al., 2004), Lens, Paris (Cazalet, et al., 2004) and Corby (Glöckner, et al., 2008), probes and primers were designed. In order to provide comparable binding efficiency, these sequences had to be specific for each target, to be free from self-hybridizing sequences, and to have similar lengths, GC-contents and melting temperatures.

Finally, these probe and primer sequences were again verified against the GenBank database using BLAST (Camacho, et al., 2009) to exclude false-positive reactions resulting from possible cross-hybridizations. The Array Design Software Package (Alere Technologies) was used to list all theoretically possible hybridizations of the probes to the *L. pneumophila* genomes, allowing up to five mismatches per probe, although measurable signals can only be produced with up to two mismatches. This facilitated a prediction of hybridization patterns for the sequenced strains. Finally, 97 probes and a biotin staining control were spotted in quadruplicate on arrays that were mounted onto ArrayStrips (Alere Technologies) (Suppl. Figure 1). The probes range in length from 22 to 33 nucleotides with a median length of 27 nucleotides.

Hybridization experiments were then carried out in order to optimize the protocol. The experimental procedures and data processing were initially evaluated by comparing DNA-microarray results of the sequenced *L. pneumophila* strains Philadelphia-1, Lens, Paris and Corby with the *in silico* predicted reactivities that were based on the published genomes of these strains. The assay protocol, especially with regard to hybridization and washing temperatures, was modified and optimized stepwise until the experimental results for these four strains fully matched the *in silico* predicted hybridization patterns. The resulting protocol and data processing step are described below.

DNA microarray typing

L. pneumophila isolates were subcultured and recloned on BCYE agar for 48 h. Chromosomal DNA was extracted using the automatic DNA extraction system EZ1

(QIAgen, Hilden, Germany) and the DNA concentration was measured for purity using the Implen P300 (Implen, Munich, Germany) and GeneQuant™ pro (GE Healthcare, Freiburg, Germany).

A multiplex linear primer elongation with 62 primers was used for targeted amplification and dUTP-linked biotin incorporation using the Alere HybridizationPlus Kit (Alere Technologies). In brief, 1 µg bacterial DNA was added to a master mix containing 3.9 µl labeling buffer (B1), 0.1 µl DNA polymerase (B2) and 1 µl of the primer mix (0.135 µmol/L of each primer). Linear amplification was performed after an initial denaturation step (5 min at 96 °C) in 45 cycles of 20 s at 60 °C, 30 s at 72 °C and 20 s at 96 °C.

ArrayStrips were read out within 20 min by a reading device (ArrayMate, Alere Technologies) according to the manufacturer's instructions.

Image analysis

The raw data of the signal intensity of each spot was normalized as described elsewhere (Monecke, et al., 2007; Monecke, et al., 2008). Briefly, the normalized intensity (NI) was first calculated, in which NI = 1 – (M / BG), where M is the mean intensity of a spot, and BG is the intensity of the local background, giving results between 0 (weak signal, i.e., negative) and 1.0 (strong signal, i.e., positive). For further analyses, median values for all four spots of one probe were used.

For the interpretation of normalized data, flexible breakpoints were introduced that depended on overall signal quality and staining intensity. An experiment was regarded as valid based on the NI median of the biotin staining control and the positive probes that recognize the alleles of the gene *rnpB* (ribonuclease P RNA component). For the calculation of the 'average *rnpB* & staining control value' only NI ≥ 0.2 were considered as positive. A valid experiment yielded an 'average *rnpB* & staining control value' above 0.666 and has to include at least one positive *rnpB* probe.

Raw data of each valid experiment were translated in a 97-ternary marker hybridization profile (HP) that was composed of 'positive' (pos), 'ambiguous' (amb) and 'negative' (neg). In valid experiments, the average NIs of a probe that were above 2/3 of the 'average *rnpB* & staining control value' were regarded positive, all that were above 1/3 but below 2/3 of this value were regarded as ambiguous. Signals that were below 1/3 of this value were considered negative (Suppl. Table 5).

Evaluation of microarray performance

To evaluate the array for reproducibility and stability, 103 strains were tested at least twice in independent experiments. Furthermore, sets of related strains were analyzed in

order to define a threshold that could assign related isolates to each other based on a HP-similarity score. Therefore, HPs were compared based on the mapping of categorical data. A predefined similarity scoring matrix was used where pos:pos and neg:neg matches were assigned as 1 (i.e. 100 % similarity) for self-matches. Matches amb:neg, neg:amb, amb:pos, pos:amb were regarded as 0.5, thus having 50 % similarity. Mismatches with pos:neg or neg:pos were counted as 0, having no similarity. Based on this matrix a dendrogram was calculated using the Unweighted Pair Group Method with Arithmetic mean algorithm (UPGMA) of the BioNumerics 7.5 software (Applied Maths, Sint-Martens-Latem, Belgium).

Results

The detection of the epidemic strain, typed as *L. pneumophila* Sg1, mAb subgroup Knoxville ST345, proved to be difficult due to the high number of other *Legionella* spp. that were present in the environmental samples. For instance, among the first randomly picked 34 colonies from environmental samples, only four were characterized as the epidemic strain using the standardized mAb subgrouping and SBT (Figure 6). A major obstacle for the investigation was a simultaneous and abundant presence of a second *L. pneumophila* isolate with the same mAb subgroup Knoxville but with another sequence type - ST600, a double locus variant of ST345. When processing the following samples, we put more attention on the colony morphology in order to increase our detection rate but colonies of

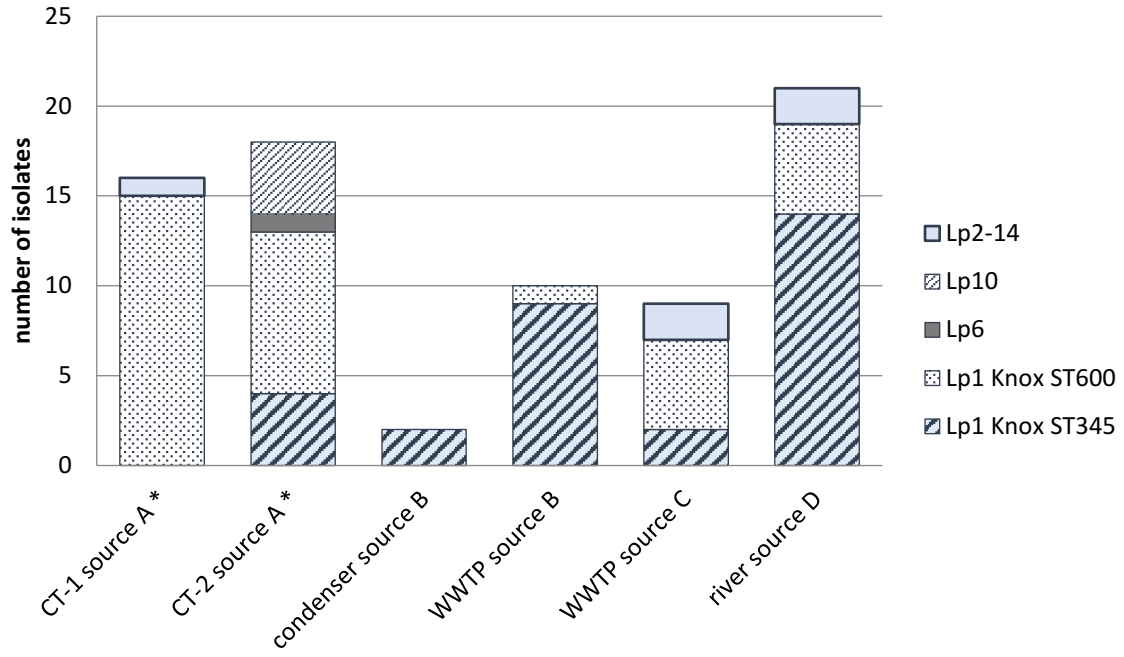


Figure 6: Different *L. pneumophila* Sgs isolated in different sources. After identification of small numbers of the epidemic strain only among randomly picked colonies from sources indicated with asterisk (*), we tried to focus on the colony shape and increased the detection rate. Source A: pipeline production site; source B: private brewery; source C: municipal WWTP; source D: the river Wester with its branches.

the prominent ST600 strain could not be distinguished visually with certainty from the epidemic strain ST345.

By using the DNA microarray it was possible to discern the epidemic ST345 isolates from the background Knoxville ST600 isolates. Hybridization profiles of the first fully SBT genotyped isolates showed that the epidemic strains and the background isolates clearly differed from each. Both STs grouped into separate clusters with a HP-similarity score of around 90 % (Figure 7). Concurrently, we observed intra-cluster HP-similarity scores > 96 % for ST345 isolates and > 94 % for ST600 isolates, respectively (Figure 8). With this fast and discriminatory method in hand, it was possible to characterize all *L. pneumophila* Sg1 isolates found during the ongoing Warstein outbreak investigation. In total, 66 environmental isolates were analyzed. Thirty-one of these formed a separate cluster and matched with HP-similarity scores > 96 % to the clinical isolates of the epidemic strain. Further 35 isolates could be assigned to the ST600 cluster with a HP-similarity score > 94 % (Figure 8). The similarity between the two clusters remained at 90 % after all isolates were analyzed. Subsequently, isolates were confirmed as the epidemic strain or ST600 by SBT.

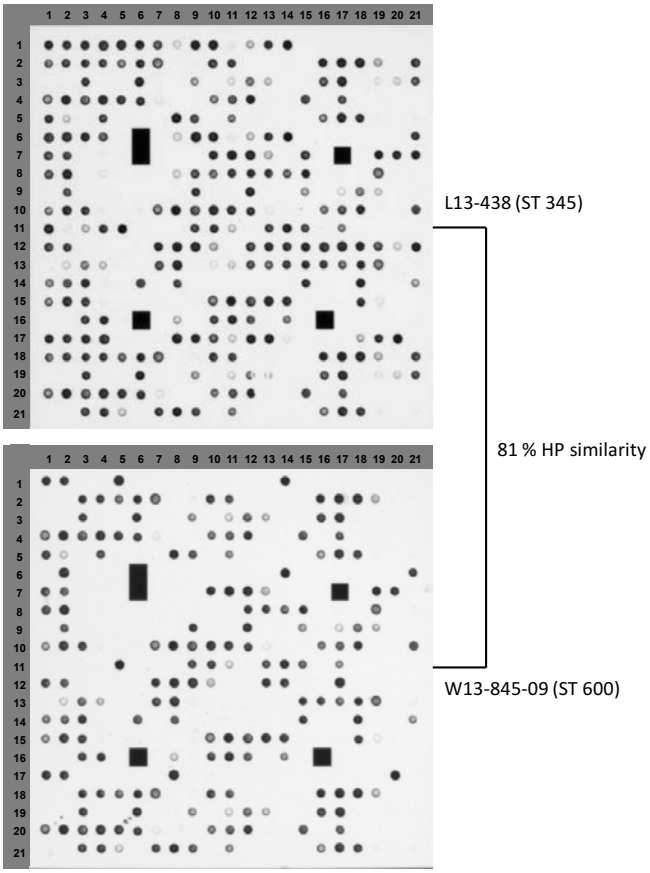
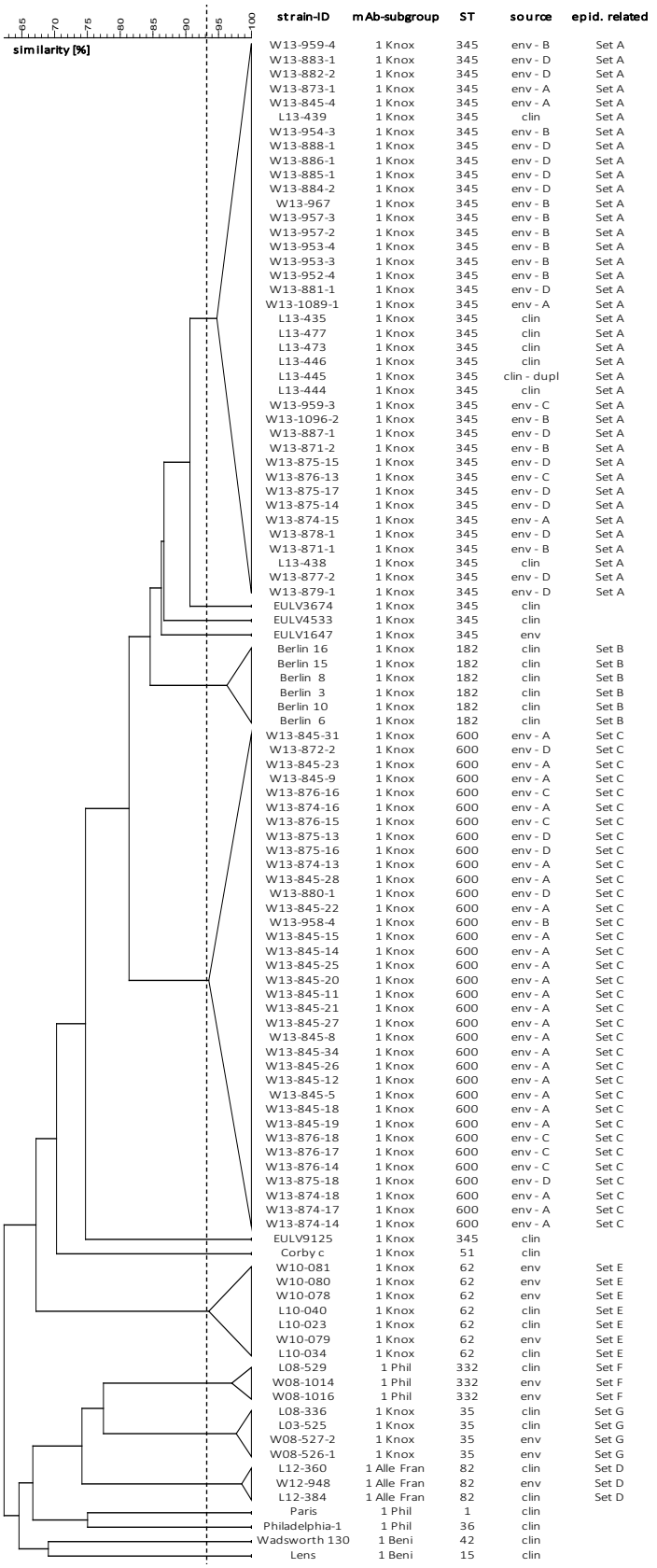


Figure 7: Two DNA microarray pictures. Positive hybridization is indicated by grey/black spots. The upper array shows a representative isolate of the epidemic strain ST345 (L13-438). The array below represents a ST600 background isolate (here W13-845-09). Similarity is based on normalized hybridization profiles (HP) and calculated using the Unweighted Pair Group Method with Arithmetic mean algorithm (UPGMA). Position of probes can be compared with the s scheme of the array (Suppl. Figure 1).



strain-ID	mAb-subgroup	ST	source	epid. related
W13-959-4	1 Knox	345	env - B	Set A
W13-883-1	1 Knox	345	env - D	Set A
W13-882-2	1 Knox	345	env - D	Set A
W13-873-1	1 Knox	345	env - A	Set A
W13-845-4	1 Knox	345	env - A	Set A
L13-439	1 Knox	345	clin	Set A
W13-954-3	1 Knox	345	env - B	Set A
W13-888-1	1 Knox	345	env - D	Set A
W13-886-1	1 Knox	345	env - D	Set A
W13-885-1	1 Knox	345	env - D	Set A
W13-884-2	1 Knox	345	env - D	Set A
W13-967	1 Knox	345	env - B	Set A
W13-957-3	1 Knox	345	env - B	Set A
W13-957-2	1 Knox	345	env - B	Set A
W13-953-4	1 Knox	345	env - B	Set A
W13-953-3	1 Knox	345	env - B	Set A
W13-952-4	1 Knox	345	env - B	Set A
W13-881-1	1 Knox	345	env - D	Set A
W13-1089-1	1 Knox	345	env - A	Set A
L13-435	1 Knox	345	clin	Set A
L13-477	1 Knox	345	clin	Set A
L13-473	1 Knox	345	clin	Set A
L13-446	1 Knox	345	clin	Set A
L13-445	1 Knox	345	clin - dupl	Set A
L13-444	1 Knox	345	clin	Set A
W13-959-3	1 Knox	345	env - C	Set A
W13-1096-2	1 Knox	345	env - B	Set A
W13-887-1	1 Knox	345	env - D	Set A
W13-871-2	1 Knox	345	env - B	Set A
W13-875-15	1 Knox	345	env - D	Set A
W13-876-13	1 Knox	345	env - C	Set A
W13-875-17	1 Knox	345	env - D	Set A
W13-875-14	1 Knox	345	env - D	Set A
W13-874-15	1 Knox	345	env - A	Set A
W13-878-1	1 Knox	345	env - D	Set A
W13-871-1	1 Knox	345	env - B	Set A
L13-438	1 Knox	345	clin	Set A
W13-877-2	1 Knox	345	env - D	Set A
W13-879-1	1 Knox	345	env - D	Set A
EULV3674	1 Knox	345	clin	
EULV4533	1 Knox	345	clin	
EULV1647	1 Knox	345	env	
Berlin 16	1 Knox	182	clin	Set B
Berlin 15	1 Knox	182	clin	Set B
Berlin 8	1 Knox	182	clin	Set B
Berlin 3	1 Knox	182	clin	Set B
Berlin 10	1 Knox	182	clin	Set B
Berlin 6	1 Knox	182	clin	Set B
W13-845-31	1 Knox	600	env - A	Set C
W13-872-2	1 Knox	600	env - D	Set C
W13-845-23	1 Knox	600	env - A	Set C
W13-845-9	1 Knox	600	env - A	Set C
W13-876-16	1 Knox	600	env - C	Set C
W13-874-16	1 Knox	600	env - A	Set C
W13-876-15	1 Knox	600	env - C	Set C
W13-875-13	1 Knox	600	env - D	Set C
W13-875-16	1 Knox	600	env - D	Set C
W13-874-13	1 Knox	600	env - A	Set C
W13-845-28	1 Knox	600	env - A	Set C
W13-880-1	1 Knox	600	env - D	Set C
W13-845-22	1 Knox	600	env - A	Set C
W13-958-4	1 Knox	600	env - B	Set C
W13-845-15	1 Knox	600	env - A	Set C
W13-845-14	1 Knox	600	env - A	Set C
W13-845-25	1 Knox	600	env - A	Set C
W13-845-20	1 Knox	600	env - A	Set C
W13-845-11	1 Knox	600	env - A	Set C
W13-845-21	1 Knox	600	env - A	Set C
W13-845-27	1 Knox	600	env - A	Set C
W13-845-8	1 Knox	600	env - A	Set C
W13-845-34	1 Knox	600	env - A	Set C
W13-845-26	1 Knox	600	env - A	Set C
W13-845-12	1 Knox	600	env - A	Set C
W13-845-5	1 Knox	600	env - A	Set C
W13-845-18	1 Knox	600	env - A	Set C
W13-845-19	1 Knox	600	env - A	Set C
W13-876-18	1 Knox	600	env - C	Set C
W13-876-17	1 Knox	600	env - C	Set C
W13-876-14	1 Knox	600	env - C	Set C
W13-875-18	1 Knox	600	env - D	Set C
W13-874-18	1 Knox	600	env - A	Set C
W13-874-17	1 Knox	600	env - A	Set C
W13-874-14	1 Knox	600	env - A	Set C
EULV9125	1 Knox	345	clin	
Corby c	1 Knox	51	clin	
W10-081	1 Knox	62	env	Set E
W10-080	1 Knox	62	env	Set E
W10-078	1 Knox	62	env	Set E
L10-040	1 Knox	62	clin	Set E
L10-023	1 Knox	62	clin	Set E
W10-079	1 Knox	62	env	Set E
L10-034	1 Knox	62	clin	Set E
L08-529	1 Phil	332	clin	Set F
W08-1014	1 Phil	332	env	Set F
W08-1016	1 Phil	332	env	Set F
L08-336	1 Knox	35	clin	Set G
L03-525	1 Knox	35	clin	Set G
W08-527-2	1 Knox	35	env	Set G
W08-526-1	1 Knox	35	env	Set G
L12-360	1 Alle Fran	82	clin	Set D
W12-948	1 Alle Fran	82	env	Set D
L12-384	1 Alle Fran	82	clin	Set D
Paris	1 Phil	1	clin	
Philadelphia-1	1 Phil	36	clin	
Wadsworth 130	1 Beni	42	clin	
Lens	1 Beni	15	clin	

Figure 8: Dendrogram of hybridization profiles (HP). In total 106 selected *L. pneumophila* Sg1 isolates were included. Sequence type and mAb subgroup are indicated. Epidemiological related strains are summarized in seven sets (Set A-G) (Table 4). Unrelated clinical (clin) and environmental (env) isolates are designated including the sample sites during the outbreak (source A-D). The dashed line indicates a cut-off value at the HP-similarity score of 94 %. The dendrogram is calculated using the Unweighted Pair Group Method with Arithmetic mean algorithm (UPGMA).

The epidemic strain was isolated from samples of two cooling towers of different companies (source A and B), the sewage pretreatment plant (source B), the municipal water supply system and WWTP (source C), as well as the water samples of the river Wester and its branches (source D) (Figure 6 and Figure 8).

To verify HP-similarity scores we validated the DNA-microarray by using five sets of related isolates and several unrelated isolates. Each set of epidemiological related isolates formed clusters with a HP-similarity score ≥ 94 %, as seen for the epidemic strain ST345 and the ST600 outgroup isolates of the Warstein outbreak. The reference value of a HP-similarity score ≥ 94 % was also seen for 103 unrelated strains that were tested at least twice. Valid experiments of repeatedly tested strains never showed HP-similarity scores below this threshold. We therefore regarded an HP-similarity score of 94 % and higher as good indication to define isolates as related to each when typed during an outbreak scenario, thus putatively belonging, i.e., to the same clone.

Whether the cluster formed by the epidemic strain was unique or a ST345 specific cluster was further verified by testing unrelated *L. pneumophila* ST345 isolates of the mAb subgroup Knoxville from Belgium (EULV4533), France (EULV9125) and the Netherlands (EULV1647, EULV3674). The three Belgian and Dutch isolates were the closest isolates to the ST345 isolates with 90 % (EULV4533), 92 % (EULV1647) and 94 % similarity, respectively. One French isolate (EUL4533) had similarities of 85 % (Figure 8) to both clusters.

Discussion

In August 2013, the town of Warstein faced the largest outbreak of LD ever reported in Germany with 78 laboratory-confirmed cases, including two fatalities (Maisa, et al., 2015). The significant number of confirmed cases during the outbreak required an investigation for possible environmental sources by detailed and intensive sampling of all kinds of water reservoirs, including cooling towers, municipal and industrial sewage and WWTPs, municipal tap water and river water.

There are two major tasks for the field investigation team. First, confirm that the patients were infected with the same strain, the outbreak strain. Accidental clusters caused by different strains have occasionally been detected (Pereira, et al., 2006). The choice of diagnostic tools plays thereby an important role to avoid pseudoepidemics due to false-positive results (Rota, et al., 2014). On the other side long-lasting outbreaks could be detected by appropriate diagnostic assays (Scaturro, et al., 2015). Second, the highest resolution possible to detect the source of the infection by comparing clinical isolates with environmental isolates should be used.

Clinical as well as environmental isolates of *L. pneumophila* Sg1 were analyzed using the novel DNA microarray. We found an excellent reproducibility of the hybridization signals obtained with the same strains from the ESGLI strain collection (EUL strains) in independent experiments. Furthermore, the experimental results for strains with published genome sequences showed the expected results.

The results obtained by the DNA-microarray could be confirmed by established typing methods, e.g. mAb subgrouping and SBT (Helbig, et al., 2002; Gaia, et al., 2005; Ratzow, et al., 2007). Thus, the microarray succeeded in assigning isolates to the epidemic strain of the Warstein outbreak. This supports the application of the microarray during the outbreak investigation as a fast and reliable genotyping tool. Using the DNA microarray enabled us to rapidly genotype several environmental isolates and hence to identify potential sources.

Due to the fact that several sources harbored the epidemic strain we cannot say with certainty whether one or more sources were the origin of infection during the outbreak. On the other side it was extremely helpful to implement control methods for the reservoirs where we found the epidemic strain.

The DNA-microarray could be used as surveillance and subtyping tool for *L. pneumophila* Sg1 isolates and it proved to be valuable in the current situation of a simultaneous presence of epidemic and environmental strains that yielded the same mAb subgroup. However, prior to routine use (as opposed to an emergency use as in the discussed outbreak), its discriminatory power must be evaluated based on a larger and more diverse panel.

Conclusion

Legionella is a frequent cause of bacterial pneumonia in the world. In the event of an outbreak a quick and appropriate response is necessary to prevent further cases. Microbiological diagnostic tools and subtyping techniques to detect the source of infection in a timely manner are indispensable. The newly developed microarray is a promising tool for epidemiological investigations comprising the ability to discriminate successfully between *L. pneumophila* Sg1 strains in a rapid and robust manner.

Authors' contribution

MP generated putative targets for the array, conducted all typing experiments, performed, analyzed and interpreted the DNA-microarray results, and drafted the manuscript. RE developed the DNA-microarray platform and the reading devices. PS finalized the probe and primer sequences. SP, AB and ME supported the investigation as on-site investigator. SM and CL initialized the project of the DNA-microarray, supervised the work, helped interpreting the results and reviewed the manuscript. All authors read and approved the manuscript.

Acknowledgements

The authors acknowledge the contribution of all those individuals involved in the outbreak: Local, Regional and National Health Agency staff, the Local Authority, Environmental Health officers, hospital staff and microbiologists – for their dedication and professionalism in the face of such a large outbreak. Furthermore, we thank Kerstin Lück, Susan Menzel (TU Dresden) and Elke Müller (Alere technologies) for excellent laboratory work.

Thanks also to Fedoua Echaidi of the University Hospital of Brussels in Belgium, Sophie Jarraud of the National Reference Centre of *Legionella* in France and Jacob de Bruin of the National Institute for Public Health and the Environment in The Netherlands for providing us unrelated ST345 isolates. Many thanks to Dr. Norman K. Fry (PHE, London) for curating the European SBT database. This study was in part supported by the Robert Koch-Institute/German Federal Ministry of Health grant 1369-351.

Part II - Comment

The work summarized in Part II focuses on the establishment and validation of a novel DNA-DNA hybridization based microarray for the genotyping of *L. pneumophila* Sg1 strains. Therefore, a variety of genetic targets were extracted from published *L. pneumophila* genomes by the colleagues from Alere™. Furthermore, the LPS biosynthesis loci was analysed in detail to define DNA-probes that distinguish *L. pneumophila* isolates into clonal groups and/or clonal complexes. The DNA-microarray runs on a robust and user-friendly platform that was invented by the team at Alere™. It eliminates excessive hands-on time and has a higher level of discrimination in comparison to the current gold standard methods, such as SBT and mAb-subgrouping.

Ideally, this DNA-microarray would also be able implement current typing methods and assigns e.g. it would simultaneously implement a ST (without sequencing) and a mAb-subgroup (without using mAbs) to an analysed strain. However, it is evident that transferring the complete set of alleles of more than 2,400 STs (SBT Database, acc: 08/2017) to the microarray platform is not feasible. In anticipation of the preliminary results, it appears that the clustering of clonal groups based on hybridization patterns does not support ST-specific clustering. By contrast, the implementation of DNA-probes that differentiate strains into mAb-subgroups appears realistic as only nine mAb-subgroups for *L. pneumophila* Sg1 strains are described. However, this requires detailed information on the genetic background of the differentiation of *L. pneumophila* isolates into Sgs and subsequently into Sg1-specific mAb-subgroups – a research topic that is largely unknown.

Several studies have suggested the LPS-biosynthesis locus as the main genomic region within the genome that is responsible for Sg-specific LPS modifications (Lüneberg, et al., 2000). A highly conserved part within this locus, spanning ORF 14 to ORF 28, can be found within the genomes of all Sgs. This is thought to be involved in early steps of the biosynthesis of legionaminic acid, the main component of the distal *O*-specific chain (McNally, et al., 2006; Schoenhofen, et al., 2006; Glaze, et al., 2008). By contrast, a heterogenetic region (ORF 1-ORF 13) is exclusively found in Sg1 strains and is putatively transferred via genomic recombination and/or homologous gene transfer. Furthermore, it has been demonstrated that differences within this region have an impact on the LPS structure, leading to a loss-of-reactivity with LPS-specific mAbs and subsequently resulting in mAb-subgroup switches (Zou, et al., 1999; Bernander, et al., 2003; Wagner, et al., 2007). Therefore, it can be assumed that this heterogenic region is involved in specific modifications of the LPS structure, to be more precisely the legionaminic acid of the *O*-chain, of *L. pneumophila* Sg1 strains (Helbig, et al., 1995; Zähringer, et al., 1995; Knirel, et al., 1996; Kooistra, et al., 2001; Lüneberg, et al., 2001; Kooistra, et al., 2002a; Kooistra, et al., 2002b; Mérault, et al., 2011).

The detailed analysis of the LPS-biosynthesis locus of all the *L. pneumophila* Sg1 mAb-subgroups in this work revealed several genes/ORFs that may be involved in mAb-subgroup specific LPS modifications. These candidates exhibited conserved regions within a certain mAb-subgroup and variabilities to all other mAb-subgroups. A total of five genes with 37 variants were included in the DNA-microarray's set of genetic targets, which putatively divides *L. pneumophila* Sg1 isolates into mAb-subgroups according to a genotypic approach.

The overall discriminatory potential of the DNA-microarray was evaluated using 550 clinical and environmental isolates (Suppl. Table 8). All the strains were typed by using the DNA-microarray, SBT and mAb-subgrouping. The hybridization profiles (HP) generated by the DNA-microarray were combined to HP-patterns. Hybridization patterns with similarities of 90 % and more were bundled into one HP-pattern. In this way, the strains were distinguished into 153 HP-patterns. The index of discriminatory power was 0.975 (95 % confidence interval [CI]: 0.971-0.981), which was higher than the mAb-subgrouping (IOD: 0.829; CI 95 % 0.818-0.839), SBT (IOD: 0.947; CI 95 % 0.935-0.958) and both methods combined (IOD 0.971; CI 95 % 0.965-0.977) (Hunter and Gaston, 1988; Gaia, et al., 2005; Ratzow, et al., 2007).

Moreover, by extracting the hybridisation results of the selected probes that specifically hybridise to LPS-biosynthesis genes, it was possible to construct a typing scheme similar to the mAb-subgrouping 'Dresden panel + mAb3' scheme (Figure 9). The *lag-1* gene can achieve a first essential discrimination between isolates. The positive hybridization results separate the isolates into the 'Pontiac group' and the 'non-Pontiac

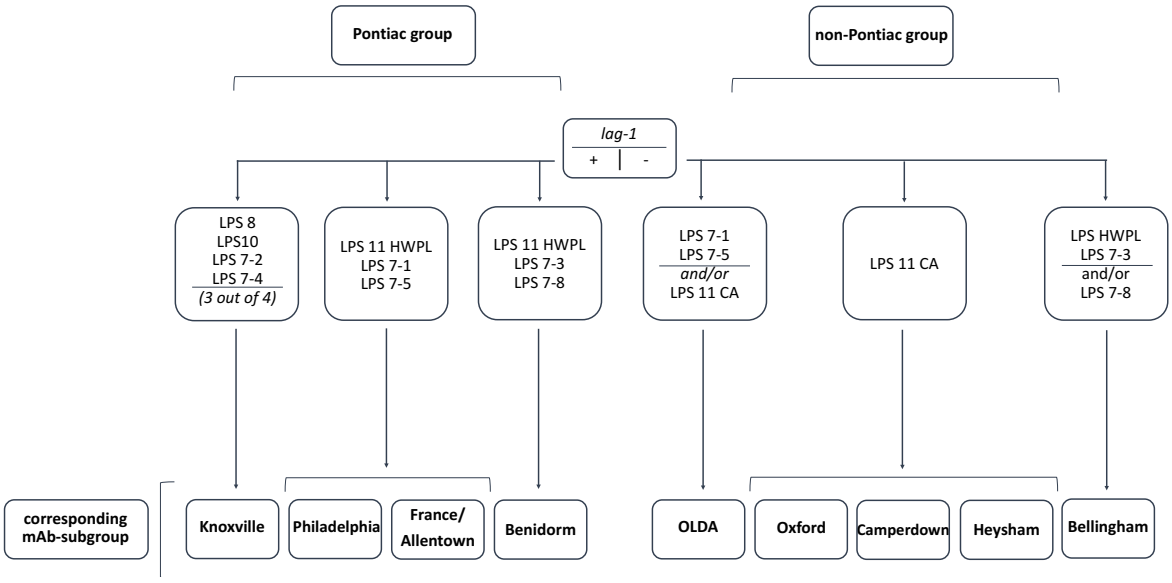


Figure 9: DNA-microarray based mAb-subgrouping. A classification into the corresponding mAb-subgroup is possible on the basis of on the reactivity with LPS biosynthesis genes on the DNA-microarray. The *lag-1* probes divide isolates into the 'Pontiac group' and the 'non-Pontiac group'.

group', in a similar way to the mAb 3/1 [compare Figure 2]. Furthermore, the reactivity of the isolate's DNA with certain mAb-subgroup specific probes assigns a strain according to the phenotypic mAb-subgroup. In this way, the mAb-subgroups Knoxville, Benidorm, OLDA and Bellingham could be classified. Limitations were observed for the mAb-subgroups Philadelphia and France/Allentown and for Oxford, Camperdown and Heysham that could not be distinguished.

The application of the DNA-microarray becomes essential when the isolates of certain endemic clones cannot be distinguished by other methods. These clones are disproportionally found in specific regions and/or sample types. It I is impossible for SBT and mAb-subgrouping to distinguish the isolates of these clones from each other. A prominent example is represented by the *L. pneumophila* isolates, which are typed as mAb-subgroup OLDA, ST1. Isolates with these characteristics can be found in several environmental samples, making it difficult to rely on the results when environmental and clinical isolates are compared. However, this dilemma is partially solved by the DNA-microarray. The group of *L. pneumophila* mAb-subgroup OLDA ST1 isolates can be further divided into smaller clusters. This enabled for instance the identification of the infectious source of contamination in a fatal LD case of a new-born in which the environmental and clinical isolates shared more that 98 % HP-similarity (Figure 10).

This indicates that the DNA-microarray allows a fast and reliable genotyping of *L. pneumophila* Sg1 isolates. It proved to be an important tool during the LD outbreak in Warstein in 2013, when sample volumes increased enormously and typing results were required immediately. A central advantage of the DNA-microarray over the SBT and the mAb-subgrouping is its greater discriminatory power, which is based on the HP-patterns. It enables endemic clones to be distinguished when current methods fail. Importantly, the robust HP-patterns allow a global comparison of isolates, independent of the user, due to a transferrable platform. The detailed analysis of the LPS-biosynthesis loci revealed specific gene targets that allow the transfer of the mAb-based subgrouping scheme to a genotypic approach. Whether certain STs can be subdivided by the DNA-microarray requires further evaluation. However, an initial analysis has already revealed that a division into ST-specific patterns has proved to be difficult due to the strong heterogeneity within several STs.

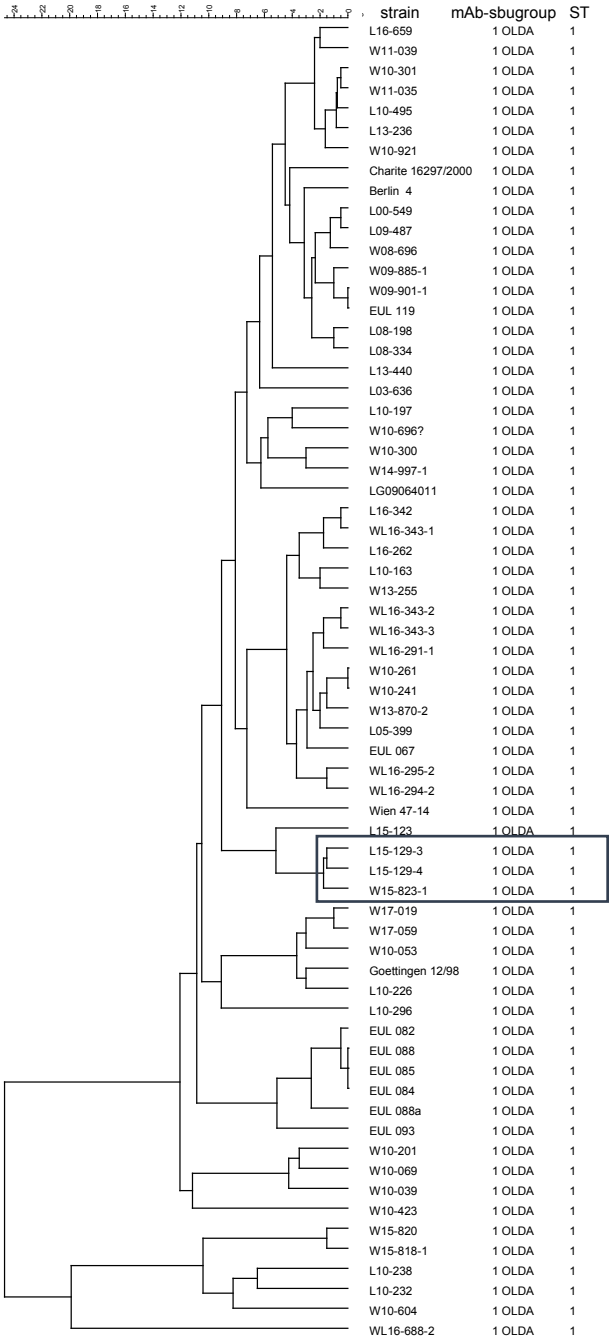


Figure 10: Dendrogram of hybridization profiles of ST1, mAb-subgroup OLDA isolates. Highlighted isolates represent a single case of two clinical isolates L15-129-3, -4 and an environmental isolate W15-823-. 1 of the household water system. The dendrogram is calculated using the Unweighted Pair Group Method with Arithmetic mean algorithm (UPGMA). [results presented at GNPI/DGPI in Dresden 2017; see Publication record].

Part III - Chapter I

Confirming Legionnaires' Disease Outbreak by Genome-Based Method, Germany, 2012

Florian Burckhardt
Andre Brion
Jeanette Lahm
Heinz-Ulrich Koch
Karola Prior
Markus Petzold
Dag Harmsen
Christian Lück

Emerging Infectious Diseases

Bruckhardt et al: Confirming Legionnaires' Disease Outbreak by Genome-Based Method, Germany 2012. EID 2016;22(7):1303-1304

Submitted 20 October 2015 | Accepted 07 April 2016 | Published 22 December 2016

To the Editor

We report an outbreak of Legionnaires' disease in southwestern Germany. On July 31, 2012, the State Health Agency of Rhineland-Palatinate was informed by the local health department of the city of Zweibrücken that ten patients tested positive for *Legionella pneumophila*, the bacterium that causes Legionnaires' disease. The onset of disease for all case-patients was from June 26 through July 25, which exceeded the yearly average of one to four patients a month. By August 23, we had received notifications of 19 patients with pneumonia and notification of 1 patient who did not exhibit pneumonia. We set three parameters for reporting a patient as a Legionnaires' disease case-patient. First, the patient had to either live in or have been visiting the city of Zweibrücken in June 2012 before onset of disease. Second, the respiratory samples from the patient had to contain *L. pneumophila* or the results of patient's serogroup 1 urinary antigen test had to be positive for the bacterium (ECDC, 2015a). Finally, clinical or radiologic confirmation of the disease was required. Of 20 patients who fit the case definition, 14 were male and six were female. Nine smoked and 2 were immunocompromised; none died.

All case-patients were positive for *L. pneumophila* serogroup 1 urinary antigen. From clinical samples of two patients, legionellae were cultured, and the infecting strain was confirmed as *L. pneumophila* serogroup 1, monoclonal subgroup Allentown-France, sequence type (ST) 82 (Moran-Gilad, et al., 2015; ESGLI, 2015). Currently, 118 strains of this ST are found in the European database for sequence-based typing of *L. pneumophila* (ESGLI, 2015). Most ST82 strains were isolated from clinical samples; thus, this ST appears more likely than other strains to infect humans. Further, three respiratory samples from case-patients were positive in a PCR for *L. pneumophila* serogroup 1 but were negative by culture (Mérault, et al., 2011). These samples were investigated with the nested sequence-based typing protocol, which allows typing data to be obtained directly from clinical samples (ESGLI, 2015). Of the three samples, two were confirmed as ST82.

The local health authority did not initially identify likely sources of transmission such as cooling towers, public spas, or warm water supply systems in the vicinity of the patients (Mercante and Winchell, 2015). Environmental samples were taken from the homes of 15 of the 20 patients; all samples tested negative for *Legionella* (ISO 11731, 1998).

To find the source of the outbreak, we plotted 20 home and seven work addresses of patients using Quantum-GIS software and found that 18 addresses were within a 2-km radius of each other, including two patients who had limited mobility and had not left their homes during their incubation period (Figure 11) (QGIS, 2012). We conducted a site visit on August 22 to inspect a sewage plant and 2 large manufacturing plants (A and B) that were within the same 2-km radius. Neither the sewage plant nor plant A had a potential *Legionella* source. Plant B had a cooling tower mounted on a rooftop that was described by

the company as a closed circuit cooling system, indicating that no aerosols would be released, and thus was missed by the initial local health department inquiry. However, closed circuit referred only to the primary cooling circuitry, whereas excess heat was exchanged through wet surface cooling, allowing release of aerosols into the atmosphere. The local health department immediately shut down the cooling tower, and plant B used shot-dose chlorine to disinfect it. Before disinfection, we obtained three swab specimens and 250-mL samples of water from the reservoir and plated them in dilutions with and without acid wash (ISO 11731, 1998; ISO 11731-2, 2008). Samples without acid wash were completely overgrown, whereas a single 1-mL sample with acid wash showed 20 *Legionella* colonies after seven days. Three colonies were typed and found to belong to the epidemic strain. Of the 27 work and home addresses, six were within a 1-km radius of the cooling tower, and 18 were within a 4-km radius (Figure 11). No further cases occurred within the incubation period (up to 14 days after closure of the cooling tower).

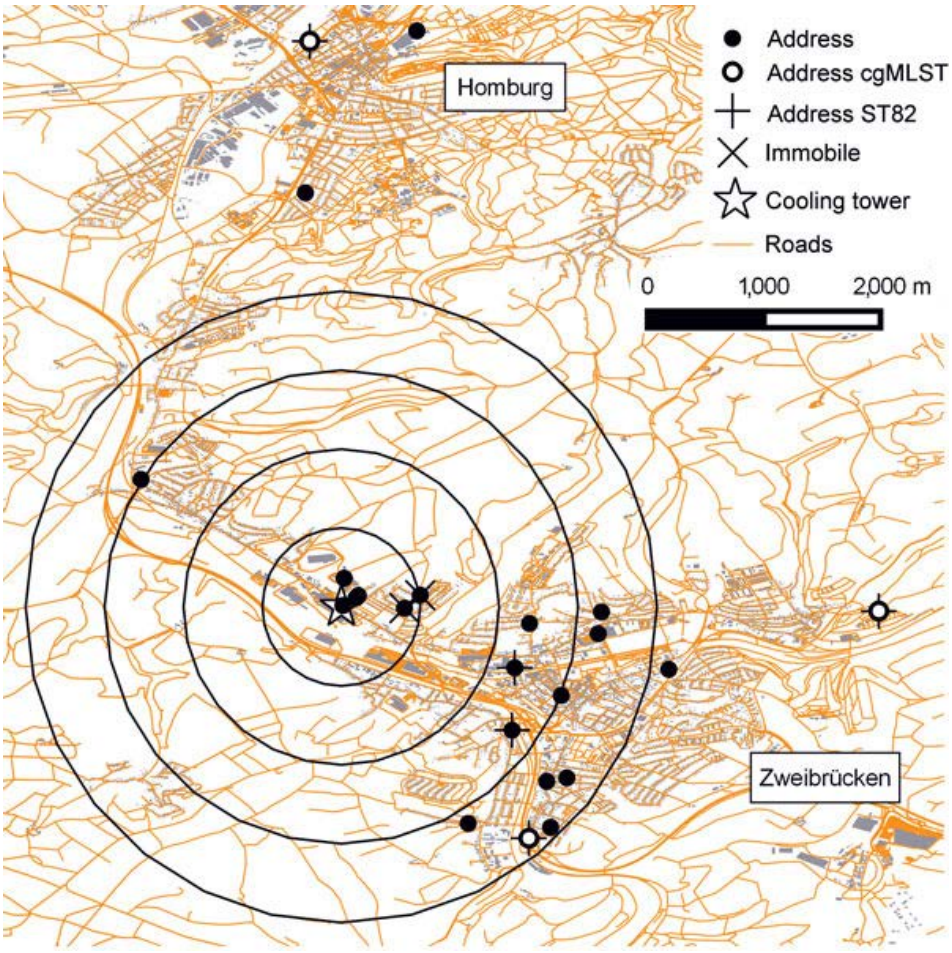


Figure 11: Geographic distribution of cooling tower and home and work addresses (n = 23) of patients. One patient may be represented twice with home and work address, because place of infection is unknown. The addresses marked “immobile” belong to 2 patients who had not left their homes. Two samples had undergone core genome multilocus sequence typing (cgMLST), and sequence type (ST) 82 was typed (represented by 2 home addresses and 1 work address). For 2 samples, only ST82 was typed. Two dots in the 1-km radius are overlapping each other. Four addresses (9 km, 10 km, 19 km, and 26 km from the cooling tower) are outside the scale of the map. Circle radii are from 1 km to 4 km, centered on the cooling tower. Shapefiles for mapping by OpenStreetMap contributors.

To further confirm this cooling tower as the source of the outbreak, we applied core genome multilocus sequence typing (cgMLST) (Moran-Gilad, et al., 2015). We analyzed allelic differences of 1,521 gene targets of the core genome of *L. pneumophila* using the pairwise ignore missing values option in SeqSphere^{+} software (Ridom GmbH, Münster, Germany). Results showed that the strains from 2 patients with culture-positive test results and the three environmental ST82 strains were identical in their cgMLST profile, which covers 47 % of the Philadelphia-1 reference genome. Currently, no German law requires a registry for cooling towers; such a registry would accelerate identification of potential *L. pneumophila* emission during outbreaks (Maisa, et al., 2015). In January 2015, a code of conduct for maintenance of cooling towers went into effect (VDI, 2015). Modern typing methods such as cgMLST can serve as supporting tools in confirming infection origin. However, this method must be validated on a larger scale, and its discriminatory power compared with that of current typing methods. Further cgMLST studies with other ST82 strains are underway.

Authors' contribution

FB, AB, JL, HUK conducted the local microbiological, geographical and epidemiological investigations. FB and CL drafted the manuscript and supervised the outbreak investigation. KP and DH performed the cgMLST analysis. MP performed general typing and laboratory analysis of the isolates and assisted during cgMLST analysis.

Acknowledgments

We acknowledge the support of Jürgen Blanz of the Regional Trade Control, Rhineland Palatinate, and thank the visited companies for their full and unconditional cooperation. We also acknowledge the contribution of all those individuals involved in the outbreak—local, regional and national Health Protection Agency staff, the Local Authority, the Health and Safety Executive, Environmental Health Officers, hospital staff and microbiologists —for their dedication and professionalism in the face of the outbreak.

Part III - Chapter II

Epidemiological information is key when interpreting whole genome sequence data – Lessons learned from a large L*egionella pneumophila* outbreak in Warstein, Germany, 2013

Markus Petzold*
Karola Prior*
Jacob Moran-Gilad
Dag Harmsen
Christian Lück

* Authors contributed equally

Eurosurveillance

Petzold et al: Epidemiological information is key when interpreting whole genome sequence data – Lessons learned from a large *Legionella pneumophila* outbreak in Warstein, Germany, 2013. Euro Surveill. 2017

Submitted 21 February 2017 | Accepted 18 July 2017

Abstract

Introduction

Whole genome sequencing (WGS) is increasingly used in Legionnaires' disease (LD) outbreak investigations, owing to its higher resolution than sequence-based typing, the gold standard typing method for *Legionella pneumophila,* in the analysis of endemic strains. Recently, a gene-by-gene typing approach based on 1,521 core genes called core genome multilocus sequence typing (cgMLST) was described that enables a robust and standardised typing of *L. pneumophila*.

Methods

We applied this cgMLST scheme to isolates obtained during the largest outbreak of LD reported so far in Germany. In this outbreak, the epidemic clone ST345 had been isolated from patients and four different environmental sources. In total 42 clinical and environmental isolates were retrospectively typed.

Results

Epidemiologically unrelated ST345 isolates were clearly distinguishable from the epidemic clone. Remarkably, epidemic isolates split up into two distinct clusters, ST345-A and ST345-B, each respectively containing a mix of clinical and epidemiologically-related environmental samples.

Discussion/conclusion

The outbreak was therefore likely caused by both variants of the single sequence type, which pre-existed in the environmental reservoirs. The two clusters differed by 40 alleles located in two neighbouring genomic regions of ca 42 and 26 kb. Additional analysis supported horizontal gene transfer of the two regions as responsible for the difference between the variants. Both regions comprise virulence genes and have previously been reported to be involved in recombination events. This corroborates the notion that genomic outbreak investigations should always take epidemiological information into consideration when making inferences. Overall, cgMLST proved helpful in disentangling the complex genomic epidemiology of the outbreak.

Introduction

Legionella spp. is the causative agent of Legionnaires' disease (LD) named after its first occurrence during a convention of the American Legion in 1976 (Fraser, et al., 1977; McDade, et al., 1977). These rod shaped Gram-negative bacteria inhabit all kinds of natural

and man-made fresh water reservoirs including cooling towers (CT), spas and drinking water systems. Inhalation of legionellae-containing aerosols originating from contaminated environmental reservoirs is the main route of infection. However, a case of a person-to-person transmission of *Legionella* under special circumstances was recently reported (Correia, et al., 2016).

LD accounts for 2–20% of community-acquired pneumonia (CAP) and the number of cases in Europe reached almost 6,000 in 2014 (Torres, et al., 2014). In Germany ca 1,000 cases are reported annually, representing an incidence of ca 11 cases per million population (RKI, 2016). In Europe ca 10 % of the cases are related to clusters or outbreaks. So far, 60 species and more than 70 serogroups (Sg) of the genus *Legionella* were reported from which around half were implicated in human infections (Bajrai, et al., 2016). The vast majority of LD cases is caused by *L. pneumophila* serogroup Sg1 isolates, especially monoclonal antibody (mAb) 2/3–1 positive strains (ECDC, 2015b). Hitherto, all CT-related outbreaks reported worldwide were caused by these subtypes (Walser, et al., 2014).

Molecular and serological typing methods are predominately applied to the species *L. pneumophila*. The two well-established epidemiological typing methods for comparison of clinical and environmental isolates consist of the subgrouping scheme based on mAbs and the sequence based typing (SBT) method, an adapted multilocus sequence typing (MLST) variant that defines sequence types (ST) (Helbig, et al., 2002; ESGLI, 2015). Other methods have been described but lack uniform interpretation of results in inter-laboratory comparison studies (Van Belkum, et al., 2007; Sabat, et al., 2013).

Currently, SBT is the gold standard to genotype *L. pneumophila* isolates. The allelic profile of seven genes enables the assignment of an ST to the corresponding isolate. A database, curated by Public Health England (PHE), London, United Kingdom, in cooperation with the European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden, facilitates the exchange of typing data and can be queried for surveillance and epidemiological studies of *L. pneumophila* (ESGLI, 2015). Currently, the database consists of more than 11,000 reported isolates with 2,298 different STs (status as of 05 January 2017). Despite the index of discrimination of the SBT scheme being around 0.92, typing of frequently circulating STs with this method, e.g. ST1, ST47 and ST23, proves less informative to further differentiate strains within these rather big clonal groups (David, et al., 2016b). An additional typing step is thus needed, but modalities attempted thus far such as spoligotyping were of limited value (Ginevra, et al., 2012).

Due to a higher level of discrimination compared with gold standard typing methods of different bacteria, including *L. pneumophila,* whole genome sequencing (WGS) has become a frequently applied tool in outbreak investigations (Reuter, et al., 2013; Graham, et al., 2014; Lévesque, et al., 2014; McAdam, et al., 2014; Moran-Gilad, et al., 2015; Raphael, et al., 2016). While use of this tool has mostly relied on analysis of single nt

polymorphisms (SNPs), a few studies are based on a genome-wide gene-by-gene typing approach for *L. pneumophila* Sg1 strains. These extended MLST schemes enable a detailed comparison of two or more isolates by either considering all genes of a species (pangenome) in what is called whole genome MLST (wgMLST), or alternatively, a set of conserved genes of a species, namely core genome MLST (cgMLST) (Maiden, et al., 2013). Analysis of several related strains and strains that were involved in small outbreaks using these cgMLST or wgMLST produced results that were in agreement with current standard typing methods, indicating the suitability of these methods as typing tools for *L. pneumophila* Sg1 isolates (Moran-Gilad, et al., 2015; Burckhardt, et al., 2016; Raphael, et al., 2016).

Here, we report in detail the retrospective application of a previously described cgMLST scheme consisting of 1,521 genes (Moran-Gilad, et al., 2015), to the largest outbreak of LD reported so far in Germany, in order to validate this scheme on a large and homogenous set of isolates. The outbreak occurred in the summer of 2013 in the city of Warstein. In total, 78 confirmed LD cases were reported and multiple potential environmental sources of infection carrying the outbreak strain were implicated. These included several CTs, municipal and private waste water treatment plants (WWTPs) and the river Wester, which runs through the city of Warstein. The outbreak strain was characterised as *L. pneumophila* Sg1, mAb-subgroup Knoxville, ST345 (Maisa, et al., 2015).

Methods

Cultivation of *L. pneumophila* isolates and DNA extraction

Respiratory samples (bronchoalveolar lavages, BAL) from outbreak patients, with and without heat treatment at 50 °C for 30 min were plated on non-selective buffered charcoal-yeast extract (BCYE) agar and a selective agar containing cefamandole, polymyxin B, and anisomycin (BMPA) and incubated at 36 °C in humidified atmosphere supplemented with 5% CO_2. Isolated strains were initially serotyped by using a latex agglutination test (Oxoid, Wesel, Germany) and confirmed by using the Dresden panel of mAbs as described elsewhere (Helbig, et al., 2002). The environmental isolates were isolated according to ISO11731/1998 (Essig, et al., 2016; Petzold, et al., 2016) and typed in a similar way. Additionally, all samples were typed according the *L. pneumophila* SBT protocol (ESGLI, 2015).

DNA from respiratory samples was extracted using the EZ1 DNA tissue kit (Qiagen, Hilden, Germany) according to manufacturer instructions. Clinical samples were tested with a *L. pneumophila* specific PCR (DUPLICα RealTime Legionella pneumophila Kit,

Euroclone, Milan, Italy) and a *L. pneumophila* Sg1 specific PCR (Mérault, et al., 2011). Furthermore, direct genotyping from three culture negative PCR-positive clinical samples was attempted using the nested SBT (nSBT) protocol (Mentasti, et al., 2012).

Whole genome sequencing and assembly

Deep frozen clinical and environmental isolates collected during the outbreak (stored in 15% glycerol at − 80 °C) were thawed, sub-cultured on BCYE-agar plates (Oxoid, Wesel, Germany), and incubated for another 48 hour as described above. We additionally included unrelated isolates of ST345 as well as two strains of *L. pneumophila* Sg1 mAb-subgroup Knoxville ST600, a double locus variant of ST345 frequently isolated during the outbreak (Table 5). Colonies were harvested and resuspended in sterile distilled water for subsequent DNA extraction using the purification protocol for Gram-negative bacteria of the MagAttract HMW DNA Kit (Qiagen, Hilden, Germany).

Table 5: *Legionella pneumophila* serogroup 1 samples and isolates used for the retrospective analysis of a 2013 Legionnaires' disease outbreak by core genome multilocus sequence typing, Germany (n = 46)

Sample ID / source (C/E)[a]	Epidemiological context to outbreak	Culture	Monoclonal subgroup[b]	ST[c]	Allelic profile (*flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*, *neuA*)	Outbreak clone[d]
L13–435 (C)	Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
L13–438 (C)	Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
L13–439 (C)	Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
L13–444 (C)	Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
L13–445/-446 (C)[e]	Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
L13–473 (C)	Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
L13–477 (C)	Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–845–1 (E)	Cooling tower, source A, Warstein outbreak	+	Knoxville	600	6, 10, 19, 28, 19, 4, 11	NA[f]
W13–845–4 (E)	Cooling tower, source A, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–845–8 (E)	Cooling tower, source A, Warstein outbreak	+	Knoxville	600	6, 10, 19, 28, 19, 4, 11	NA[f]
W13–871–1 (E)	Condenser, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–873–1 (E)	Pump shaft, source A, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–874–15 (E)	Pump shaft, source A, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–875–15 (E)	River inlet, source A, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–875–17 (E)	River inlet, source A, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–876–13 (E)	Aeration basin, source C, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–878–1 (E)	River water, source D, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–879–1 (E)	River water, source D, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–952–4 (E)	Pre-sedimentation basin, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–953–3 (E)	Pre-sedimentation basin, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
W13–953–4 (E)	Pre-sedimentation basin, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
W13–954–3 (E)	Pre-sedimentation basin, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
W13–957–2 (E)	Outlet, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–959–3 (E)	Aeration basin, source C, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
W13–959–4 (E)	Inlet from source B, source C, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
W13–1093 (E)	Cooling tower, source A, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W13–1096–2 (E)	Cooling tower, source A, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W14–178 (E)	Pre-sedimentation basin, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W14–472 (E)	Pre-sedimentation basin, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W14–474 (E)	Pre-sedimentation basin, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
W14–476 (E)	Pre-sedimentation basin, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-B
W14–489 (E)	Pre-sedimentation basin, source B, Warstein outbreak	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	ST345-A
P13–308 (C)	Clinical sample, Warstein outbreak	-	ND[f]	345	6, 10, 19, 3, 19, 4, 11	NA[g]
P13–402 (C)	Clinical sample, Warstein outbreak	-	ND[f]	ND	6, 10, 0, 3, 0, 4, 11	NA[g]
P13–733 (C)	Clinical sample, Warstein outbreak	-	ND[f]	345	6, 10, 19, 3, 19, 4, 11	NA[g]
EULV1461 (C)	Unrelated isolate (France)	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	NA[g]
EULV1647 (E)	Unrelated isolate (the Netherlands)	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	NA[g]
EULV1654 (C)	Unrelated isolate (the Netherlands)	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	NA[g]
EULV3674 (C)	Unrelated isolate (the Netherlands)	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	NA[g]
EULV5358 (C)	Unrelated isolate (France)	+	OLDA	345	6, 10, 19, 3, 19, 4, 11	NA[g]
EULV6345 (C)	Unrelated isolate (France)	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	NA[g]
EULV9125 (C)	Unrelated isolate (France)	+	Knoxville	345	6, 10, 19, 3, 19, 4, 11	NA[g]
Corby (C)	Unrelated isolate (United Kingdom)	+	Knoxville	51	6, 10, 15, 28, 9, 14, 6	NA[g]
Alcoy 2300/99 (C)	Unrelated isolate (Spain)	+	Knoxville	578	6, 10, 15, 13, 9, 14, 6	NA[g]
Philadelphia-1	Unrelated isolate (United States)	+	Philadelphia	36	3, 4, 1, 1, 14, 9, 1	NA[g]

C: clinical sample, E: environmental sample, ID: identity, NA: not applicable, ND: not determined, ST: sequence type
[a]: source of sample: C corresponds to clinical isolates; E corresponds to environmental isolates
[b]: Monoclonal subgrouping as described by Helbig and coworkers (Helbig, et al., 2002)
[c]: Sequence based typing by using Sanger sequencing according to the ESCMID Study Group for Legionella Infections (ESGLI) protocol (ESGLI, 2015)
[d]: Outbreak ST345 strain clone assignment by core genome multiclocus sequences typing
[e]: Isolate L13-445/-446 were isolated from the same patient.
[f]: The molecular antibody-subgroup and/or sequence type were/was not determined because no isolate was obtained and/or because of incomplete direct sequence-based typing.
[g]: Clone assignment within the ST345 outbreak strain was not applicable when the isolate was from a chosen non-outbreak reference strain, when the isolate had another ST than ST345, or when no isolate could be obtained.

Sequencing libraries were prepared using the Nextera XT library prep kit (Illumina GmbH, Munich, Germany) for a 250 bp paired-end sequencing run on an Illumina MiSeq sequencer. Samples were sequenced to aim for a minimum 100-fold coverage using Illumina's recommended standard protocols with dual-index barcoding and rotation of barcodes over time. Sequencing run quality (Q30 and output) had to fulfil the manufacturer's minimum specifications. The resulting FASTQ files were quality trimmed and assembled de novo using the Velvet assembler that is integrated in Ridom SeqSphere $^+$ v.3.0 software (Ridom GmbH, Münster, Germany) (Jünemann, et al., 2013). Here, reads were trimmed at their 5'- and 3'-ends until an average base quality of 30 was reached in a window of 20 bases, and the assembly was performed with Velvet version 1.1.04 (Zerbino and Birney, 2008) using optimised k-mer size and coverage cut-off values based on the average length of contigs with > 1,000 bp.

Core genome multilocus sequence typing cgMLST analysis

A cgMLST was performed using SeqSphere $^+$ with the *L. pneumophila* typing scheme described by Moran-Gilad et al. (2015). This scheme includes 1,521 core genome genes and the basic local alignment search tool (BLAST)-based allele calling procedure details have been described previously (Moran-Gilad, et al., 2015). The percentage of good cgMLST targets determined the overall sequence quality of every sample such that samples containing at least 95% of extracted cgMLST targets were considered typeable. Alleles for each gene were assigned automatically by the SeqSphere $^+$ software to ensure a unique nomenclature. The combination of all alleles in each strain formed an allelic profile that was used to generate minimum spanning trees (MST). Targets with missing values in one of the strains compared were omitted during distance calculation. In order to maintain backwards compatibility with *L. pneumophila* SBT, sequences of the seven genes comprising the allelic profile of the SBT schemes were separately extracted from finished genomes and WGS data and then queried against the SBT database (ESGLI, 2015) in order to assign classic STs in silico.

Detection of recombined regions

The de novo assembly FASTA contig files of four samples were chosen (L13–435, L13–473, W13–879–1, and W13–952–4) in order to analyse their genomes for putative recombined regions. Mauve (version 20150226 build 10, default parameters) (Darling, et al., 2010) was used to calculate a multiple alignment of the four genomes. A SeqSphere $^+$ function was used to convert the Mauve alignment file from XMFA format into a FASTA and thereby concatenating the alignments for each of the four sample sequences and replacing all ambiguous bases against 'N'. Gubbins (version 1.4.5, default parameters) (Croucher, et

al., 2015) was used for recombination prediction based on the Mauve alignment. Predicted recombined regions (> 7,500 bp) were subsequently scanned with SeqSphere $^{+}$ against all cgMLST targets (using BLAST with thresholds 66% identity and 50% overlap) to reveal corresponding targets within the recombined region.

Data availability

All raw reads generated were submitted to the European Nt Archive (http://www.ebi.ac.uk/ena/) of European molecular biology laboratory (EMBL) European Bioinformatics Institute (EBI) under the study accession number PRJEB12633. The cgMLST targets as well as the allelic profiles of each isolate were deposited at the cgMLST.org nomenclature server (http://www.cgmlst.org).

Results

The outbreak occurred in the city of Warstein, Germany, in 2013 and involved 78 laboratory-confirmed LD cases (Maisa, et al., 2015). Respiratory samples from 10 patients tested positive using the *L. pneumophila* and the Sg1-specific PCRs. From seven of these patients, eight clinical isolates were recovered. All these isolates were characterised as *L. pneumophila* Sg1, mAb-subgroup Knoxville, ST345 and regarded as being of the particular outbreak strain. This outbreak strain was also isolated from an industrial CT (source A; source designation as in (Maisa, et al., 2015), two WWTPs (source B and C) and the river Wester (source D) running through the town of Warstein.

In total, 42 strains were used in cgMLST analysis. These strains comprised the eight clinical isolates identified by conventional means as outbreak strain and 25 epidemiologically related environmental isolates. Furthermore, seven unrelated ST345 strains (six environmental and one clinical) as well as two reference genomes of the strains Corby and Alcoy 2300/99 were also included (Table 5). All seven unrelated ST345 isolates that were deposited in the European SBT database for *L. pneumophila* (ESGLI, 2015) until the outbreak occurred were part of those strains. Furthermore, two environmental *L. pneumophila* Sg1 isolates of the same mAb-subgroup Knoxville as the outbreak strain but of a different but close genotype, ST600, were analysed. These ST600 isolates were found in higher numbers than the epidemic strain in all environmental samples taken during the outbreak but were not recovered from any of the clinical samples (Table 5) (Petzold, et al., 2016).

Although the original scheme as described by Moran-Gilad et al. (2015) consists of 1,521 targets, some targets can happen to be absent in some strains. Therefore, a minimum spanning tree (MST) was constructed based on 1,475 targets of the cgMLST scheme that were present in all analysed genomes. Remarkably, the MST identified two clearly

distinguishable clusters of the ST345 isolates obtained during the outbreak, hereafter referred to as ST345-A and ST345-B (Figure 12).

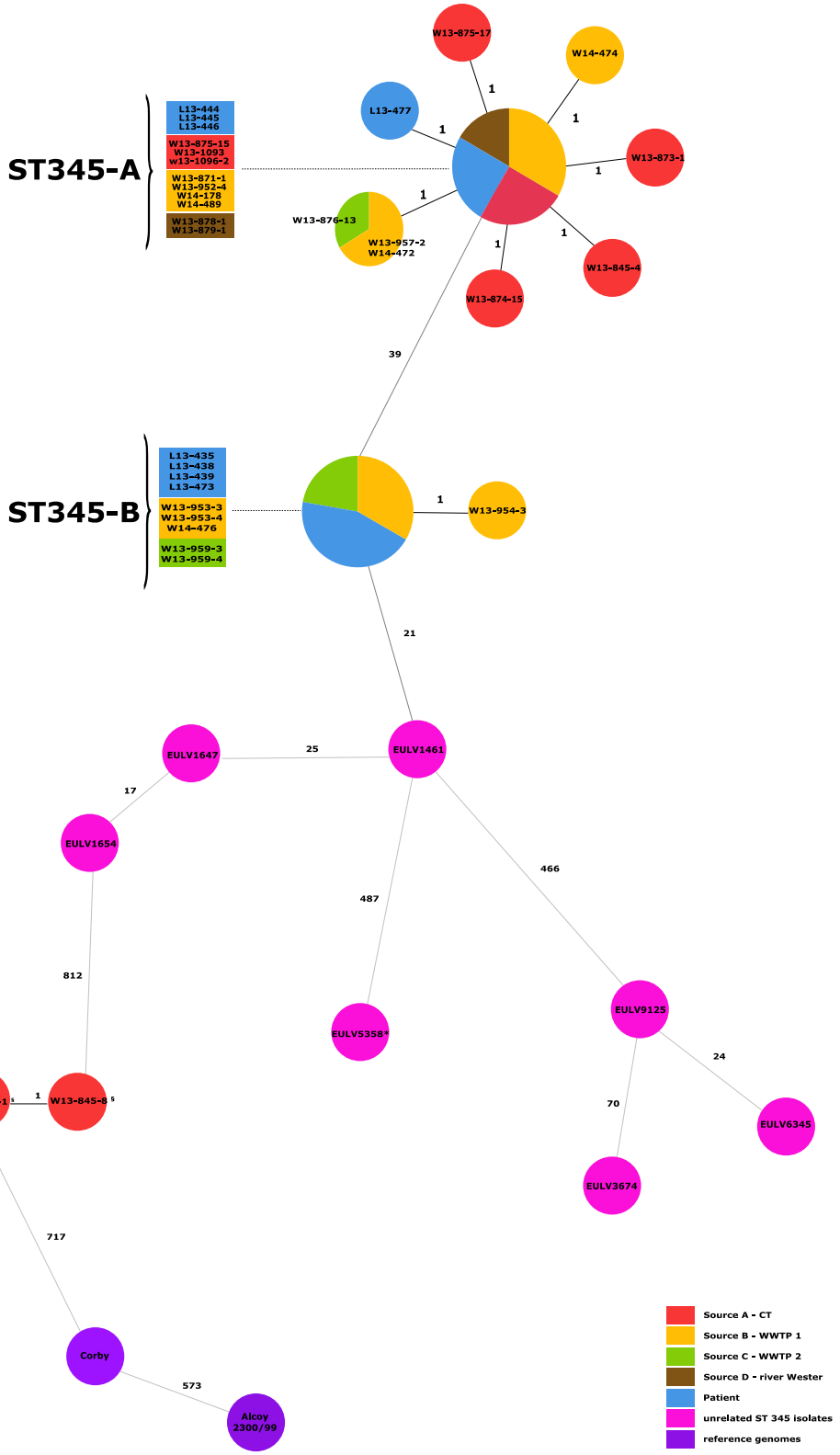


Figure 12: Analysis by minimum spanning tree based on 1,475 core genome multilocus sequence typing targets of isolates recovered in 2013 during a Legionnaires' disease outbreak, Germany (n = 42 strains). Numbers on lines represent the allelic differences between the isolates. All but two isolates in the Figure are sequence type (ST) 345. The two ST600 isolates are indicated with (§). All ST345 isolates which are unrelated to the outbreak and which serve as outgroup belong to the monoclonal antibody (mAb)-subgroup Knoxville, except EULV5358, which is mAb subgroup OLDA and which is indicated with (*).CT: cooling tower; WWTP: waste water treatment plant.

Although the original scheme as described by Moran-Gilad and colleagues consists of 1,521 targets, it happens that targets are absent in some strains. Therefore, a minimum spanning tree (MST) was constructed based on 1,475 targets of the cgMLST scheme that were present in all analysed genomes. Remarkably, the MST identified two clearly distinguishable clusters of the ST345 isolates obtained during the outbreak, hereafter referred to as ST345-A and ST345-B.

Cluster ST345-A consisted of four clinical isolates (including two isolates from the same patient L13–445/-446) and 17 isolates recovered from all four putative environmental sources A–D. From these isolates, twelve showed no allelic difference and nine isolates differed each in a single allele from this central node. The remaining 10 ST345 outbreak isolates grouped as a separate cluster ST345-B, which differed in 39 alleles from ST345-A. Nine of the 10 ST345-B isolates showed an identical cgMLST profile and were isolated from four clinical and five environmental samples of sources B and C (both WWTPs). One environmental sample from source B showed one allele difference. Direct comparison of both clusters using all 1,521 cgMLST targets revealed in fact 40 alleles difference. The seven unrelated ST345 isolates were quite diverse revealing from 17 allelic differences between EULV1647 and EULV1654 up to 1,023 differences between EULV5358 and EULV3674. The unrelated ST345 isolate EULV1461 had only 21 allelic differences to the epidemic clone ST345-B. The two ST600 isolates showing only one allele difference from each other as well as the genomes of Corby and Alcoy 2300/99 clearly differ from the ST345 clones in more than 800 alleles (Figure 12).

We further investigated the differences between the two ST345 clusters in more detail. The aforementioned 40 different alleles are apparently located on two distinct neighbouring genomic regions including respectively 27 targets (recombination region 1; corresponding genes of reference strain Philadelphia-1 lpg2604–2636) and 13 targets (recombination region 2; lpg2666–2687) (Table 6).

The affected genes comprise virulence factors including genes of the Dot/Icm Type IV secretion system and genes involved in the muramyl synthesis (Ensminger and Isberg, 2009; Liu, et al., 2012). In addition, a SNP analysis for these regions revealed several SNPs ranging from two to > 118 SNPs per gene with an average of two SNPs per 100 bp (Table 6). To investigate if the two regions resulted from potential large recombination events and juxtaposition of the involved genes and regions, the genomes of four isolates were chosen, two from each cluster (ST345-A: W13–879–1/W13–952–4; ST345-B: L13–435/L13–473), the genomes were multiple aligned with Mauve, and searched for evidence of ‘import’ of divergent sequences from a distantly related source using Gubbins. Two large recombination regions of ca 42 and 26 kb size were predicted and subsequently screened for all cgMLST targets. The scanning procedure resulted exactly in the 27 and 13 cgMLST targets that were already detected as potentially recombinatory by SeqSphere $^{+}$. Finally, the

40 targets were compared against published genomes, thereby revealing that the ST345-A cluster differed from the ST36 strain Philadelphia-1 only in three of these (Table 6).

Table 6: Core genome multilocus sequence typing targets differing between the two ST345 outbreak variants identified in a 2013 Legionnaires' disease outbreak, Germany

Target[a]	Begin	End	Gene name	GenBank protein_ID	cgMLST allele number ST345-A	cgMLST allele number ST345-B	SNPs per target
Differing cgMLST targets in 42 kb recombination region (27 targets)							
lpg2604	2938631	2939434	NA[b]	YP_096609.1	1	4	11
lpg2606	2940021	2940887	NA[b]	YP_096611.1	1	4	5
lpg2607	2941026	2943062	*pepO*	YP_096612.1	1	4	21
lpg2608	2943206	2944120	*lpxC*	YP_096613.1	1	2	9
lpg2609	2944368	2945564	*ftsZ*	YP_096614.1	1	4	12
lpg2610	2945759	2947021	*ftsA*	YP_096615.1	1	4	26
lpg2612	2947732	2948838	*ddl*	YP_096617.1	1	4	16
lpg2614	2949745	2951154	*murC*	YP_096619.1	1	4	17
lpg2615	2951164	2952348	*ftsW*	YP_096620.1	3	5	14
lpg2616	2952345	2953688	*murD*	YP_096621.1	1	4	9
lpg2617	2953702	2954820	*mraY*	YP_096622.1	1	4	7
lpg2618	2954902	2956287	*murF*	YP_096623.1	1	4	13
lpg2619	2956480	2957259	NA[b]	YP_096624.1	1	4	7
lpg2620	2957264	2960758	NA[b]	YP_096625.1	1	5	63
lpg2621	2960933	2961613	NA[b]	YP_096626.1	1	4	9
lpg2622	2961715	2962776	NA[b]	YP_096627.1	1	4	13
lpg2623	2963086	2963898	NA[b]	YP_096628.1	1	4	12
lpg2624	2963973	2964455	*greA*	YP_096629.1	1	4	10
lpg2625	2964464	2967667	*carB*	YP_096630.1	1	4	77
lpg2626	2967794	2968066	NA[b]	YP_096631.1	1	4	22
lpg2627	2968179	2969360	NA[b]	YP_096632.1	1	4	85
lpg2629	2970147	2971217	NA[b]	YP_096634.1	1	4	13
lpg2630	2971214	2972215	NA[b]	YP_096635.1	1	4	15
lpg2631	2972463	2973947	*pepA*	YP_096636.1	1	4	25
lpg2633	2974501	2974818	NA[b]	YP_096638.1	3	5	5
lpg2635	2976210	2977781	*mviN*	YP_096640.1	1	4	26
lpg2636	2978139	2978405	*rpsT*	YP_096641.1	1	2	2
Differing cgMLST targets of 26 kb recombination region (13 targets)							
lpg2666	3013221	3014102	NA[b]	YP_096671.1	1	4	9
lpg2667	3014236	3015114	*rpoH*	YP_096672.1	1	4	5
lpg2668	3015387	3016316	*ftsX*	YP_096673.1	1	4	15
lpg2671	3018077	3019402	NA[b]	YP_096676.1	1	4	17
lpg2672	3019399	3020703	NA[b]	YP_096677.1	1	4	25
lpg2673	3020700	3021245	NA[b]	YP_096678.1	1	4	10
lpg2676	3022673	3023836	*dotB*	YP_096681.1	1	4	21
lpg2678	3025688	3026485	NA[b]	YP_096683.1	1	4	4
lpg2679	3026515	3027459	NA[b]	YP_096684.1	1	4	17
lpg2680	3027690	3028718	*murE3*	YP_096685.1	1	4	27
lpg2683	3030713	3032560	NA[b]	YP_096688.1	1	4	118
lpg2684	3032560	3033420	NA[b]	YP_096689.1	3	5	35
lpg2687	3037576	3038031	*icmV*	YP_096692.1	1	4	34

cgMLST: core genome multilocus sequence typing, ID: identity, SNP: single nt polymorphism, NA: not applicable
[a]: Reference genome *L. pneumophila* strain Philadelphia-1; GenBank accession number NC_002942.5; cgMLST allele number for each target: "1"
[b]: There is currently no assigned name for this gene

Discussion

Here we present the results of the analysis of *L. pneumophila* ST345 strains isolated during the outbreak of Warstein 2013 (Maisa, et al., 2015) using a recently published cgMLST scheme. While less than 10% of reported LD cases occur in clusters and outbreaks (Mercante and Winchell, 2015), each outbreak must be regarded as a serious threat for public health since LD is a potentially life-threatening disease with case fatality rate of ca 10% (Phin, et al., 2014). Since the clinical picture of LD is not specific, the diagnosis always requires laboratory investigation. Of 78 epidemiologically and laboratory-confirmed cases of this large outbreak we were able to isolate the epidemic strain from seven patients. In

three additional clinical samples that were culture-negative, nSBT allowed the complete ST in two (P13–308 and P13–733, Table 5), and a nearly complete allelic profile in the third (P13–402). Thus, we could detect the epidemic strain in samples of 10 patients (Table 5). The rate of complete or almost complete identified STs (10/78; 12%) is in the range as reported from other outbreaks (Lévesque, et al., 2014; McAdam, et al., 2014; Sánchez-Busó, et al., 2016). However, there is a need to improve the recovery of clinical isolates in general in order to assign patients properly to an outbreak.

In the last 5 years, Legionnaires' disease outbreak investigations have increasingly included WGS (Reuter, et al., 2013; Graham, et al., 2014; Lévesque, et al., 2014; McAdam, et al., 2014; Raphael, et al., 2016). The main approach has been SNP-based, by mapping reads of clinical and environmental strains against a known reference genome. Although this enables precise differentiation between outbreak and non-outbreak isolates, the use of different reference genomes and mapping approaches makes SNP-based typing difficult to standardise. With the standardised generation, analysis and interpretation of WGS data and the establishment of a comprehensive bioinformatics pipeline and nomenclature, cgMLST allows to overcome this obstacle (Moran-Gilad, et al., 2015; Raphael, et al., 2016).

In this study, the application of cgMLST to a Legionnaires' disease outbreak revealed two distinct clusters of the epidemic *L. pneumophila* clone, namely ST345-A and ST345-B, differing in 40 alleles. This difference clearly exceeds the preliminary threshold for a WGS cluster of four alleles difference, as shown previously (Moran-Gilad, et al., 2015; Burckhardt, et al., 2016). Both clusters were indistinguishable by common gold standard methods and other typing methods (Petzold, et al., 2016). Since strains in both clusters of the epidemic clone were almost equally distributed among clinical samples and epidemiologically linked to environmental strains by place and time of occurrence, we assume that this outbreak was caused by a single epidemic ST with two variants, which were already present in the environmental reservoirs before the outbreak occurred.

Since the WGS analysis demonstrated a notable distance between the outbreak clusters, we closely examined the arrangement of the differing alleles. This analysis suggested that two major recombination events, most probably by horizontal gene transfer (HGT), may explain the differences between the two variants. Interestingly, the regions involved (42 and 26 kb) have already been reported as involved in a recombination event in a Spanish endemic clone of the same mAb-subgroup Knoxville (Sánchez-Busó, et al., 2014). The results of our investigation should serve as a note of caution for the use of WGS in outbreak investigations. Although gene-by-gene allele calling procedures like cgMLST inherently mitigate, in contrast to SNP calling procedures, against the effects of smaller recombination events, the method is prone to effects of large recombination events. Therefore, epidemiological information and/or compensation for recombination with methods as implemented in Gubbins or BratNextGen (Marttinen, et al., 2012) are strongly

recommended and, ideally, could be implemented in WGS-based typing and cgMLST standardised workflows.

An intriguing aspect would be the identification of the potential donor of the HGT regions. Comparison of the recombined regions with published genomes revealed a high similarity of the ST345-A variant to the *L. pneumophila* strain Philadelphia-1 (ST36), which was the causative agent of the first described outbreak in Philadelphia, 1976 (Fraser, et al., 1977; Mercante, et al., 2016). Additional 19 isolates of the same ST36 described by Mercante et al. were identical to the Philadelphia-1 strain for these 40 targets (data not shown) (Mercante, et al., 2016). Furthermore, three unrelated ST345 isolates (EULV1461; EULV1647 and EULV1654) are highly similar to the second cluster, ST345-B. We therefore assume that the ST345-B variant is the ancestral strain and the isolates of the ST345-A cluster evolved most probably by uptake of two large fragments from a donor strain in water systems. During the outbreak in Warstein environmental isolates of different sero- and mAb-subgroups were screened, but not tested in more depth and unfortunately not stored for later analysis which makes it impossible to identify the donor of the recombined regions.

Both ST345 clusters, ST345-A and ST345-B, which were identified during this outbreak, were identified in clinical and environmental sources. Several distinct potential environmental sources were confirmed that all harboured the epidemic clone including a CT, WWTPs and the river Wester that runs through the city of Warstein (Maisa, et al., 2015). ST345-A isolates were detected in all four environmental reservoirs (sources A–D) while ST345-B isolates were found in two of them (sources B and C). However, all sources are located close to each other and are connected to or use the water of the river Wester (source D). The extensive epidemiological investigations indicated that this outbreak must be regarded as a multifactorial event with more than one sole source of contamination. It cannot be excluded that ST345-B inhabited the remaining two sources as well but might not have been sampled or isolated during the outbreak. The final clarification regarding which source might have contaminated other sources or whether one source is the main source of infection may never be solved in detail.

The suitability as well as the usability of a cgMLST to become a new standard typing method for *L. pneumophila* Sg1 isolates was recently discussed and requires further evaluation and refinement (Moran-Gilad, et al., 2015; David, et al., 2016a). Core genome MLST combines a high discriminatory level with a standardised workflow and nomenclature which enables a global comparability of isolates. The latter is an important keynote for the surveillance and epidemiological investigation of LD wherein travel-associated infections play a significant role (European Centre for Disease Prevention and Control, 2015b). Thus, having the same typing tool is crucial. Combining knowledge on international level to define a robust scheme, a comprehensible workflow and uniform interpretation of data is mandatory. This is currently being mitigated by an international working group set up by the

European Study Group for *Legionella* Infections (ESGLI) to ensure that cgMLST is globally implemented in a fit-for-purpose manner while maintaining backwards compatibility (Gordon, et al., 2017).

Conclusion

Application of the cgMLST scheme for *L. pneumophila* demonstrated its usability during outbreak investigations. Core genome MLST showed a superior discriminatory power when compared with current gold standard typing methods, allowing for a higher resolution which resulted in finding that the epidemic strain split up into two variants. Furthermore, cgMLST indicated horizontal gene transfer as potential reason for the difference between both variants. This was confirmed by additional bioinformatics analyses. The value of classical epidemiological data was reinforced during the outbreak investigation, as such data anchored the isolates in time and space. These epidemiological data supported the findings that the large outbreak of LD in Warstein was caused by two variants, ST345-A and ST345-B, of the same ST345 clone. In the WGS era, cgMLST allows for a standardised workflow and nomenclature with high resolution and can even identify recombination events when allelic differences are clustered. However, the establishment of a globally uniform scheme needs to be well communicated and orchestrated in order to be cost-efficient and fit-for-purpose.

Authors' contribution

Initiated the study: CL, DH and JM-G; analysed epidemiological data: MP, CL; interpreted epidemiological findings: MP, CL; performed laboratory analyses and standard genotyping: MP; performed WGS and cgMLST analysis: KP, DH, MP; wrote the manuscript: MP, CL; helped to draft the manuscript: DH, JMG, KP. All authors critically revised the manuscript and approved the final version.

Acknowledgements

The authors acknowledge the contribution of all those individuals involved in the outbreak – Local, Regional and National Health Agency staff, the Local Authority, Environmental Health officers, hospital staff and microbiologists – for their dedication and professionalism in the face of such a large outbreak. Furthermore, we thank Kerstin Lück, Susan Menzel (TU Dresden) for their devotional commitment during the outbreak. Many thanks to the National Reference laboratories for Legionella in Belgium, France and the Netherlands for providing us unrelated ST345 isolates. Part of this work was presented at the 67th Annual Meeting of the German Society for Hygiene and Microbiology in Münster, Germany, 27 – 30 September 2015. This work received financial support from the Robert Koch-Institute (BMG/RKI 1369-464).

Part III - Comment

The application of cgMLST to surveillance and outbreak investigations brings the molecular typing of *L. pneumophila* to the next (second and third) generation's sequencing level. The cgMLST scheme used for *L. pneumophila* was initially described by Moran-Gilad and colleagues (2015). They tested the scheme using a few isolates of humidifier-related LD cases. The main focus of this work was to evaluate the method used during outbreak situations. I was able to use collected WGS data analyzed by the colleagues from Munster.

In addition to the outbreaks described in Zweibrücken (2012) and Warstein (2013), two further outbreaks were analyzed in retrospective and support the benefit of the application.

During the winter of 2002/2003 seven nosocomial LD cases were reported in a psychiatric ward in Germany (Ellsäßer and Hiller, 2004). All the patients resided in one section of a two-part building. Construction work on that section of the building had been completed two months before the outbreak occurred. Laboratory confirmation of the LD cases was limited to positive UATs on clinical samples. Isolates obtained from samples of the hot water drinking system were typed as *L. pneumophila* Sg1, mAb-subgroup Knoxville (ST182) and Philadelphia (ST1). Six months later, in June 2003, five LD cases were reported. This time all the cases were accommodated in the other part of the same building, which had once more been completed just two month before. This time respiratory specimens and drinking water samples were collected. Clinical isolates were typed as *L. pneumophila* Sg1, Knoxville, ST182. The water isolates showed the same characteristics. The second wave of infections was therefore caused by the ST182 strain, which was probably the causative strain of the first wave. For the cgMLST analysis, several environmental isolates were available from the first and second waves of infection, as well as from a routine follow up control in 2009 (Figure 13). Remarkably, the isolates showed a maximum of two different alleles. There was no significant difference between the strains of the two outbreak waves and the control isolate obtained six years later.

Another interesting cluster occurred on a cruise ship in 2003 in which eight patients suffered from LD, one of them including died. Clinical isolates were obtained and the epidemic strain was typed as *L. pneumophila* Sg1, mAb-subgroup Knoxville, ST35. The outbreak strain was found in the spa-pool and in the shower hose of the ship's hair-dresser (Beyrer, et al., 2007). Five years later, another LD patient was reported who was a passenger on a cruise ship during the incubation period (Hering, 2009). Despite the new name of the ship, the responsible public health authorities and the consultant laboratory for *Legionella* identified the cruise ship as the same ship. Laboratory investigations of the clinical and environmental samples confirmed the causative strain as ST35, mAb-subgroup

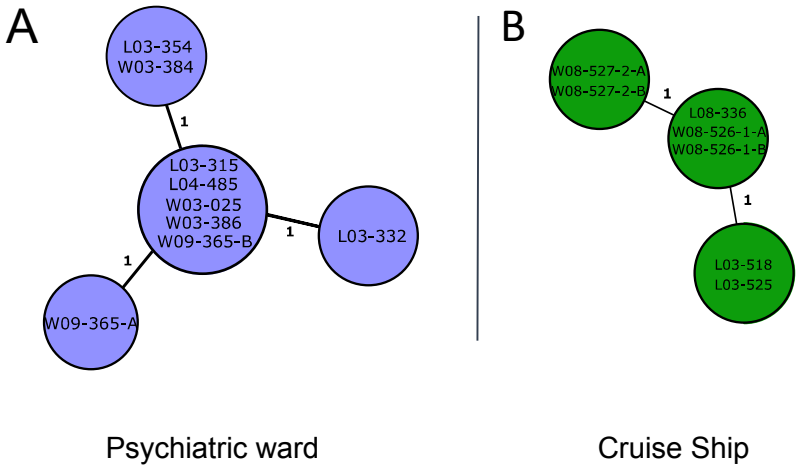


Figure 13: Core genome MLST analysis of two clusters of *L. pneumophila* Sg1 isolates. A) Isolates obtained during a biphasic outbreak in a psychiatric ward in 2003 in Frankfurt/Oder (Germany) characterized as ST182, mAb-subgroup Knoxville. The environmental isolate W03-025 was obtained during the first cluster of LD cases. The isolates W09-365-A and –B were collected in 2009 during follow up control. B) Isolates were obtained in two waves of LD cases on a cruise ship in 2003 and 2008 and characterized as ST35, mAb-subgroup Knoxville. 'L' represents clinical isolates, 'W' represents water samples. 1,400 cgMLST targets of 1,521 targets were analyzed.

Knoxville, which was identical to that of the strain isolated five years before. The cgMLST indicates that the isolates obtained five years later did not differ from the initial clinical isolates (Figure 13).

These scenarios demonstrated how cgMLST can be applied to outbreak investigations. Thus, the new method confirms current the gold standard methods. Furthermore, cgMLST represents a robust system. Isolates sampled several years apart cluster together, proving the stability of the cgMLST scheme's targets. The analysis of alleles is not representative of a complete genome, and finding no allelic difference between two isolates does not necessarily exclude other mutations, deletions or insertions within the genome, irrespective of whether the seven alleles of SBT are used or 1,521 alleles of cgMLST. However, the chance of identifying mutations using the cgMLST scheme is significantly higher due to the number of targets. Therefore, it is remarkable that the isolates of the outbreaks in the psychiatric ward and the cruise ship do not differ, which indicates a rather slow mutation rate within the core genome. The rate of mutation for selected *L. pneumophila* strains was calculated as 0.49-071 SNPs per genome and year, thus three to five mutations per genome of the aforementioned outbreak strains. These SNPs might be not detected by cgMLST as the targets are well selected coding sequences and SNPs mainly occur in non-coding regions (Sánchez-Busó, et al., 2014; David, et al., 2016b). Furthermore, Coscollá and colleagues calculated that the rate of recombination is four times higher than the rate of mutation (Coscollá, et al., 2011). The cgMLST proved its suitability as an analysis tool during the outbreak scenarios and confirmed the current typing methods.

In contrast to these outbreaks, the outbreak in Warstein revealed an unexpected feature of the cgMLST scheme. Core genome MLST distinguished the epidemic strain of the Warstein outbreak ST345 into two distinct clusters. Both clusters differed from each other by 40 alleles. Neither the current typing methods nor the newly developed DNA-microarray were able to unravel the differences between the two clusters. A detailed analysis identified recombination events as the cause of the difference. The reason for the detection of the recombined regions was the core genome scheme itself. The close proximity of the selected 1,521 coding regions (average 2 genes between alleles; range 1 to 128 genes) enabled the detection of large connected recombined regions. It should be noted here, that the arrangement of genes was based on the reference genome used and may vary among genomes. However, it appears unlikely that the genes that were arranged in close proximity in one strain show a completely different arrangement in another genome and that these genes also show allelic differences.

In addition, several genomes of *L. pneumophila* were sequenced and analyzed in detail. The species *L. pneumophila* contains several molecular markers of horizontal gene transfer (HGT) or recombination. Furthermore, a large arsenal of genes of eukaryotic origin can be found within the genome, which may have been acquired by the HGT or by their convergent evolution from their hosts (Cazalet, et al., 2004; Lurie-Weinberger, et al., 2010; Gomez-Valero, et al., 2011a). It is generally assumed that recombination is the driving force that shapes the adaptation and virulence of *L. pneumophila*. Large recombination fragments have been described, which were up to 141 kb in size (Sánchez-Busó, et al., 2014). Recently, researchers from the Wellcome Trust Sanger Institute and the Institute Pasteur identified several hot-spots for homologous recombination. They point to multi-fragment recombination of *L. pneumophila*, which implies the incorporation of several DNA segments from the same donor DNA into the genome (David and Sánchez-Busó, et al., 2017). This is certainly a conceivable process since *L. pneumophila* is naturally capable of external DNA acquisition (Stone and Kwaik, 1999).

Other cgMLST schemes of *L. pneumophila* have recently been reported (David, et al., 2016a; Raphael, et al., 2016). David and colleagues tested various target sets accurately, such as ribosomal genes and different sets of the core genome, including the set used in this work. Raphael and colleagues used a whole genome MLST pipeline of Bionumerics®, which is similar to the concept of cgMLST but uses many more targets. The conclusion drawn by all who applied a genome wide gene-by-gene typing approach was that a MLST-based method allows the assignment of traceable types and transferable results, which is important to the final user, usually public health authorities and reference laboratories. As consequence, a whole genome SNP-based method is not favorable as surveillance tool for legionellae (Maiden, et al., 2013).

Most importantly, a new typing scheme should have a higher resolution than the current methods and must fulfil the validation guidelines for molecular typing methods (Van Belkum, et al., 2007). Furthermore, epidemiological concordance is inversely correlated to resolution, which means that SNP-based whole genome analysis will (probably) always show differences between clinical and environmental isolates, and even between two isolates obtained from the same sample on the same day (Ensminger, et al., 2012). Thus, the trade-off between epidemiological concordance and resolution should be fathomed to allow surveillance and typing. Moreover, bioinformatic pipelines should be built-up and maintained. Similarly to the current SBT database, an online tool is preferred that assigns a core genome sequence type after the submission of WGS raw data, and provides further supportive information, for example, concerning putative HGT events or similarities to other strains (Schjørring, et al., 2017). In addition, during times of big data, the storage of huge amounts of WGS data needs be included in the bioinformatics pipeline (central vs. local storage; backup). A new method should maintain backward compatibility in order to obtain STs of SBT. Likewise, the multicopy dilemma of *mompS* needs to be solved bioinformatically (Moran-Gilad, et al., 2015; Gordon, et al., 2017).

An international group of *Legionella*-epidemiologists and molecular analysts has been formed to establish a new gene-by-gene typing approach for *L. pneumophila* and *L. longbeachae*, and to deal with the above issues.

Part IV - Conclusion

Understanding the ecology, and hence the distribution, of pathogenic bacteria is fundamental if epidemiologists are to keep pace with the rapidly evolving organisms. Many aspects need to be taken into account. Models have been developed of the infection and amplification of the bacteria, their physiological response to environmental stressors, their evolution and the co-evolution of their natural hosts, and the global spread of bacteria. Since the discovery of the genetic code, these topics have been accompanied and supported by molecular tools. In recent years, the molecular analysis of infectious diseases has undergone a remarkable transformation from phenotypic methods to gene or genome based methods. These tools currently deliver an immense data amount of information that needs to be analysed, understood and interpreted until a complete picture of a pathogen's lifestyle can be drawn, which will probably never occur.

An important aspect of surveillance and epidemiological investigations is the collection of as much information as possible concerning the patients who suffers from the disease. Classical epidemiology uses parameters such as time, place and person to identify epidemiological links. However, it is not only the affected person who is screened. The triggering pathogen itself and its environmental habitat also needs to be analysed in much detail as possible. Clinical microbiology allows the identification of the species and molecular typing methods are useful for characterizing and categorizing the pathogen.

Legionellae is a versatile pathogen in terms of possible study fields, which range from intracellular replication to biofilm formation, and from secreting virulence factors to co-evolution with its natural hosts. Within four decades, since the initial description of legionellae as a public health threat, knowledge concerning this pathogen has grown from day to day. However, the reported cases of LD have also increased each year, which is largely due to a higher awareness of this bacterium in health care institutions.

The isolation of legionellae by means of cultivation still represents a serious obstacle. An isolate forms the basis for further typing methods, which allow a comparison between clinical and environmental samples. Culture independent typing methods are limited to the detection of soluble antigen (e.g. urinary antigen tests) and DNA via PCR. However, these methods usually only provide information about presence or absence of the bacteria in the sample analysed. For legionellae, several typing techniques based on the isolated bacteria were evaluated and a gold standard method was finally validated and determined. For several years, SBT and mAb-subgrouping served as the globally recognized epidemiological typing scheme. Neither system runs perfectly, but they are acceptable in the absence of any better tool. The major shortcoming of current methods, namely their limited discriminatory resolution, has been frequently discussed in recent years. They do not deliver satisfactory typing results, particularly in the case of endemic clones such as the

ST48, ST37 and ST62 or globally spread clones such as ST1 SBT, even in combination with phenotypic serogrouping (Ginevra, et al., 2012; Lück, et al., 2015; Mentasti, et al., 2017). Several in-house methods have been developed by local laboratories that deal with undistinguishable clones. ThThe validation and introduction of a new molecular typing tool needs to achieve several standard parameters, including a satisfactory discriminatory power and the reproducibility and transferability of the results between laboratories (Van Belkum, et al., 2007).

The results presented in this work focus on the current limitations of *L. pneumophila* typing, such as the non-distinctiveness of globally spread STs and dominant endemic clones. In this respect, it is of utmost importance to maintain a high level of discrimination without losing epidemiological concordance, a common problem when molecular resolution increases. Furthermore, attention should be paid to a new method's working capability in routine laboratories and its performance during labour-intensive investigations such as outbreaks. It is necessary to validate and compare a newly implemented tool using current methods. The known deficits should be compensated without the introduction of new shortcomings or bottlenecks.

Both methods use whole genome DNA and both methods 'map' the DNA to a predefined template. The template used on the microarray is a single-stranded DNA-probe and represents a rather physiochemical-based mapping. By contrast, the cgMLST template is an *in silico* sequence that is compared to the sequence of the screened isolate. Both 'mapping patterns' can be translated into transferable codes (hybridization pattern for the microarray or cgMLST sequence type for the cgMLST-scheme).

The development of a DNA-hybridization based microarray was validated against current gold standard methods and passed all the demands for molecular typing: the index of discrimination is higher than that of the SBT and mAb-based subgrouping, the hands on time is comparable to these methods, the results are transferable and reproducible, and their output user-friendly (Gaia, et al., 2005; Krause, 2014; David, et al., 2016a). The study sought to implement DNA-based subgrouping of the *L. pneumophila* Sg1 isolates, which is usually done phenotypically using mAbs. The assignment of isolates to the virulent 'Pontiac group' on molecular basis may be an essential step in the screening procedures of environmental and clinical isolates. In addition, four of the nine *L. pneumophila* Sg1 subgroups can clearly be distinguished by specific hybridization patterns. The subgroups Philadelphia and Allentown/France, as well as the very rare isolated subgroups Heysham, Camperdown and Oxford were indistinguishable by the DNA-microarray. Similarly, the assignment of mAb-subgroups also allows the microarray to partially assign STs to an isolate. However, the classification is limited to STs that are clearly distinct from each other. By contrast, heterogeneous STs, such as the ST1, are separated in subgroups. This enables differentiation between frequently found isolates that are characterized as ST1,

mAb-subgroup OLDA. The DNA-microarray combines several abilities, such as the subgrouping of isolates, a rough ST assignment and the higher resolution of large heterogeneous ST, which were not distinguishable by the current methods. Thus, the DNA-microarray represents an optimized screening tool for larger sample volumes, particularly when the time window is narrow.

The initial idea to rather transfer the phenotypic mAb-based subgrouping to a molecular level succeeded partially by implementing specific probes onto the DNA-microarray. In addition, the focus was to identify the molecular traits of specific ST and mAb-subgroup combinations, and to find the molecular correlation between a particular ST and its LPS phenotype. The hierarchical analysis of hybridization patterns allows the division of isolates of the same ST into different subgroups. Multiple clusters of the same ST and subgroup that differed from each other were still observed, and however, some ST/subgroup combinations were very similar to each other. This indicates that mAb-subgroups, and hence the LPS-biosynthesis loci, do not correlate with STs. David and coworkers recently described the LPS-biosynthesis region as a hotspot for homologous recombination. This explanation supports the incongruence between the STs and mAb-subgroups of the *L. pneumophila* Sg1 isolates (David, et al., 2017).

The evaluation of the cgMLST during outbreak investigations proved the applicability of this tool. The selection of the cgMLST scheme's 1,521 genes was based on Underwood and co-workers who determined the core genome of *L. pneumophila* Sg1 (Underwood, et al., 2013). The analysis of the population structure based on the core genome largely agreed with the population analysis based on SBT. A detailed validation of the scheme was performed by David and colleagues, who found that the index of discrimination was higher than that of current gold standard methods (IOD of cgMLST: 0.999). However, they further reported that only 40 % of the analysed strains were fully typeable, which means that only a proportion of genomes carry all the 1,521 genes. This may be due to a larger set of accessory genes within the scheme that are difficult to sequence and/or assemble, or these genes may generally be absent in the analysed genomes (David, et al., 2016a). A group of experts in the field of molecular typing and *Legionella* is currently being established to verify the perfect set of core genes. Remarkably, the scheme used in this work reflects the genome of *L. pneumophila* in extraordinary depth. In this way, the identification of large genome arrangements such as recombination events becomes visible. It had generally been assumed that the expansion of *L. pneumophila* follows a rather clonal pattern with little diversification due to the low extent of recombination. However, recent studies support the more complex genome plasticity of the *L. pneumophila* genomes, with several exchangeable regions of other serogroups and even eukaryotic hosts (David, et al., 2017).

The WGS data that form the basis for the cgMLST allows, albeit bioinformatically challenging, backward compatibility for the determination of allele numbers of the classic

SBT. With the introduction of WGS, a new flood of information concerning the ecology and spread of *L. pneumophila* became visible which lead to paradigm shifts. In comparison to the current methods and the DNA-microarray presented here, the cgMLST allows the assignment of both, the classic ST and a core genome ST. However, it remains depending on the implementation of further tools that are able to detect recombination events. Moreover, the generation of genome data requires huge storage capacities, an issue that has not yet been adequately tackled.

In general, both methods fulfil the guidelines for molecular typing methods. They represent current laboratory standards and have a balanced trade-off between molecular resolution and epidemiological concordance.

With regard to the laboratory-analytic chain of the analysis of one isolate, starting from sample processing to the analysis of final results, both methods can be performed within one workday. The use of current sequencing machines for cgMLST means that the sequencing coverage per nucleotide allows a robust and reliable interpretation of the results. By contrast, the DNA-microarray requires physically repeated experiments to increase the reliability of the results. Repeated experiments on the same genome show differences between hybridization patterns up to 4 %, which must be taken into account during data interpretation. A fully automated interpretation of the results can be implemented into the analysis-pipeline for both methods, leading to the assignment, for example, of clonal groups (DNA-microarray) or a core genome ST. Additional information can be extracted from the raw results, such as a DNA-based subgroup (DNA-microarray), the classical ST or an indication of the recombination events (cgMLST).

It is very likely, that future molecular typing methods rely on WGS data. The complete sequenced genomes provide information that are not only useful during the diagnosis of LD patients, but provide also information that are required for the identification of the source of infection and the epidemiological tracing of single strains or outbreaks. It is important to establish a typing method which is globally accepted. Whole genome sequencing supports a robust scheme for the legionella typing, and allows hierarchical levels that provide the users and epidemiologist a deeper insight into certain isolates. However, WGS is currently performed by specialized reference centers. For laboratories that are not equipped with NGS-experienced staff or that do not have the money to use these machines routinely, the DNA-microarray represents an alternative to NGS and is certainly an advantage to current typing methods with regard to a fast and reliable molecular analysis during outbreak investigations and routine analyses.

Part VI - Appendices

List of abbreviations

abbreviation	explanation
°C	degrees Celsius
µg	microgram
µl	microliter
µmol	micromole
AFLP	amplified fragment length polymorphism
amb	ambiguous
B	background
BAL	broncho-alveolar lavage
BCYE	buffered charcoal yeast extract
BLAST (BLASTP)	Basic Local Alignment Search Tool (Protein BLAST)
BMPA	BCYE with cefamandole, plymyxin B and anisomycin
CAP	community acquired pneumonia
CDC	Centers for Disease Control and Prevention
cgMLST	core genome multilocus sequence typing
CI	confidence interval
clin.	clinical
cm	centimeter
CMP-5,7-diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*talo*-nonulosonic acid	CMP-Leg5Ac7Ac
CMP-5-acetamidino-7-acetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*talo*-nonulosonic acid	CMP-Leg5Am7Ac
CMP-5-acetimidoyl (*N*-methyl) amino-7-acetamido-3,5,7,9-tetradeoxynon-2-ulosonic acid	CMP-Leg5AmNMe7Ac
CO_2	carbon dioxide
cooling tower	CT
COPD	chronic obstructive respiratory disease
CRISPR	clustered regularly interspaced short palindromic repeats
defect of organelle trafficing/intra cellular multipliction	Dot/icm
DFA	direct fluorescent antibody test
DNA	deoxyribonucleic acid
Dot/Icm	defect in the organelle trafficking/intracellular multiplication
dUTP	deoxyuridine triphosphate

abbreviation	explanation
ELDSNet	European Legionnaires' disease network
ECDC	European Center for Disease Prevention and Control
ELISA	Enzyme-linke immunosorbend assay
EMBL	European Molecular Biology Laboratory
env.	environmental
ER	endoplasmatic reticulum
ESGLI	ESCMID (European Society of Clinical Microbiology and Infectious Disease) Study Group for Legionella Infections
et al.	*et alii* (and others)
etc.	*et cetera* (and the rest)
EUL	European Legionella reference strains
e.g.	*exempli gratia* (for example)
FF	filamentous form
GC	guanine/cytosine
GIS	geographic information system
GlcNAc	N-acetyl-glucosamine
h	hour
HGT	horizontal gene transfer
HP	hybridization profile
ID	identity
i.e.	*id est* (that is)
IOD	index of discriminatory power
ISO	International Organization for Standardization
kb	kilobase
L	liter
L (in e.g. L13-438)	*Legionella* (clinical legionellae isolate)
LCV	Legionella containing vacuole
LD	Legionnaires' Disease
Lp	*Legionella pneumophila*
LPS	lipopolysaccharide
M	mean intensity
mAb	monoclonal antibody
MIF	Mature intracellular form
min	minute
MLST	multilocus sequence typing
mRNA	messenger ribonucleic acid
MST	minimum spanning tree
n	number
NA	Not applicable
ND	Not detected
neg	negative
NGS	next generation sequencing
NI	normalized intensity
nSBT	nested sequence based typing
ORF	open reading frame

abbreviation	explanation
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
PHE	Public Health England
pos	positive
QuiNAc	*N*-acetyl-quinovosamine
s	second
SAM	S-adenosylmethionine
SBT	sequence-based typing
Sg	serogroup
SNP	single nucleotide polymorphism
spp.	*species pluralis*
ST	sequence type
UAT	urinary antigen test
UDP-*N,N'*-diacetylbacillosamine	UDP-Bac2Ac4Ac
UPGMA	Unweighted Pair Group Method with Arithmetric mean
V	Volt
VBNC	viable but non-culturable
W (in e.g. W13-845)	water (environmental legionellae isolate)
WGS	whole genome sequencing
WWTP	waste water treatment plant

List of Figures

List of Tables

Supplementary Data

The supplementary data and materials are stored at the research data repository Zenodo. Zenodo supports open science and is funded by the European Organization for Nuclear Research (CERN), the research program OpenAIRE and the European Commission.

The supplementary data can be retrieved under
https://doi.org/10.5281/zenodo.1037181

Publication record

Petzold M., Prior K., Moran-Gilad J., Harmsen D. and Lück C.:
Epidemiologic information is key when interpreting whole genome sequence data -lessons learned from the genomic analysis of the largest German *Legionella pneumophila* outbreak (Warstein, 2013). *Eurosurveillance*, 2017, 22(45)

Burckhardt F., Brion A., Lahm J., Koch H.-U., Prior K., Petzold M., Harmsen D. and Lück C.:
Confirming Legionnaires' Disease Outbreak by Genome-Based Method, Germany, 2012. *Emerging Infectious Disease* 2016, 22(7)

Petzold M., Ehricht R., Slickers P., Pleischl S., Brockmann A., Exner M., Monecke S. and Lück C.:
Rapid genotyping of *Legionella pneumophila* serogroup 1 strains by a novel DNA microarray-based assay during the outbreak investigation in Warstein, Germany 2013. *International Journal of Hygiene and Environmental Health* 2017, 220(4):673-678

Petzold M., Thürmer A., Menzel S., Mouton J.W., Heuner K. and Lück C.:
A structural comparison of lipopolysaccharide biosynthesis loci of *Legionella pneumophila* serogroup 1 strains. *BMC Microbiology* 2013, 13:198

Table 7: Overview of impact factors and ranks based on InCites™ Journal Citation Report® 2016 (Clarivate Analytics, 2017)

Journal	Impact Factor	Rank in category
BMC Microbiology	2,644	# 59/123; Microbiology
International Journal of Hygiene and Environmental Health	4,643	# 15/176; Public, Environmental and Occupational Health # 13/84; Infectious Diseases
Emerging Infectious Diseases	8,222	# 3/84; Infectious Diseases
Eurosurveillance	7,202	# 5/84; Infectious Diseases

Lück C., Petzold M. and Bock-Hensley O.:
Mikrobiologische, hygienische und epidemiologische Untersuchungen bei einem Todesfall eines Neugeborenen durch Legionellen. Poster presentation at the 43rd/25th annual meeting of the Society for Neonatalogy and Paediatric Intensive care (GNPI)/German Society for Paediatric Infectiology (DGPI) in Dresden, 8 – 10 June 2017

Petzold M., Jarraud S., Ehricht R., Jacotin N., Meyer T., Slickers P., Ziegler A., Monecke S. and Lück C.:
LegioType AS-1: Rapid microarray-based genotyping of *Legionella pneumophila* SG1 isolates. Poster presentation at the 4th ECCMID Study Group for Legionella Infections (ESGLI) congress in Amsterdam 22 – 23 September 2016

Petzold M., Prior K., Moran-Gilad J., Harmsen D. and Lück C.:
Retrospective investigation of the largest German *Legionella pneumophila* outbreak to date in Warstein (2013) by core genome MLST revealed two different clones as causative agents. Oral presentation at the 67th annual meeting of the German Society for Hygiene and Microbiology (DGHM) in Münster 27 – 30 September 2015

Petzold M., Prior K., Moran-Gilad J., Harmsen D. and Lück C.:
Retrospective investigation of the largest German *Legionella pneumophila* outbreak to date in Warstein (2013) by core genome MLST revealed two different clones as causative agents. Oral presentation at the 3rd ECCMID Study Group for Legionella Infections (ESGLI) congress in London 16 – 17 September 2015

Lück C., Petzold M., Lück, K., Brockmann A., Pleischl S. and Exner M.: **Microbiological Investigation during an Outbreak of Legionellosis in Warstein, Germany, August, 2013**. Oral presentation at the 4th Joint Conference of the German Society for Hygiene and Microbiology (DGHM) and the Association for General and Applied Microbiology (VAAM) in Dresden 5 – 8 October 2014

Lück C., Jarraud R., Ehricht R., Jacotin N., Meyer T., Petzold M., Slickers P., Ziegler A. and Monecke S.:

Microarray-based strain assignment of *Legionella pneumophila* isolates. Poster presentation at the 24th European Congress of Clinical Microbiology and Infectious Disease (ECCMID) in Barcelona 10 – 13 May 2014

Petzold M., Thürmer A., Menzel S., Mouton J.W., Heuner K. and Lück C.:

The lipopolysaccharide biosynthesis loci of *Legionella pneumophila* serogroup 1 strains: A structural comparison. Poster presentation at the 8th International Conference on Legionella in Melbourne 29 October – 1 November 2013

Petzold M., Thürmer A. and Lück C.:

A structural comparison of lipopolysaccharide biosynthesis loci of *Legionella pneumophila* Sg1. Poster presentation at the 1st ECCMID Study Group for Legionella Infections (ESGLI) congress in Dresden 5 – 7 September 2012

References

Allegra S, Leclerc L, Massard PA, Girardot F, Riffard S, Pourchez J. 2016. Characterization of aerosols containing *Legionella* generated upon nebulization. Sci. Rep. 6:33998.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389-3402.

Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA, Yu YK. 2005. Protein database searches using compositionally adjusted substitution matrices. FEBS J 272(20):5101-9.

Amemura-Maekawa J, Kura F, Helbig JH, Chang B, Kaneko A, Watanabe Y, Isobe J, Nukina M, Nakajima H, Kawano K, Tada Y, Watanabe H, Working Group for Legionella in J. 2010. Characterization of *Legionella pneumophila* isolates from patients in Japan according to serogroups, monoclonal antibody subgroups and sequence types. J. Med. Microbiol. 59(Pt 6):653-659.

Amemura-Maekawa J, Kikukawa K, Helbig J, Kaneko S, Suzuki-Hashimoto A, Furuhata K, Chang B, Murai M, Ichinose M, Ohnishi M, Kura F, Japan atWGfLi. 2012. Distribution of monoclonal antibody subgroups and sequence-based types among *Legionella pneumophila* serogroup 1 isolates derived from cooling tower water, bathwater, and soil in Japan. Appl. Environ. Microbiol. 78(12):4263-4270.

Avni T, Bieber A, Green H, Steinmetz T, Leibovici L, Paul M. 2016. Diagnostic accuracy of PCR alone and compared to urinary antigen testing for detection of *Legionella* spp.: A systematic review. J. Clin. Microbiol. 54(2):401-411.

Bajrai LH, Azhar EI, Yasir M, Jardot P, Barrassi L, Raoult D, La Scola B, Pagnier I. 2016. *Legionella saoudiensis* sp. nov., isolated from a sewage water sample. Int. J. Syst. Evol. Microbiol. 66(11):4367-4371.

Bartlett JG. 2011. Diagnostic tests for agents of community-acquired pneumonia. Clin. Infect. Dis. 52 Suppl 4:S296-304.

Beaute J, Zucs P, De Jong B. 2013. Legionnaires' disease in Europe, 2009-2010. Euro Surveill. 18(10):20417.

Beaute J, ELDSNet. 2017. Legionnaires' disease in Europe, 2011 to 2015. Euro Surveill. 22(27).

Benitez AJ, Winchell JM. 2013. Clinical application of a multiplex real-time PCR assay for simultaneous detection of *Legionella* species, *Legionella pneumophila*, and *Legionella pneumophila* serogroup 1. J. Clin. Microbiol. 51(1):348-51.

Benitez AJ, Winchell JM. 2016. Rapid detection and typing of pathogenic nonpneumophila *Legionella* spp. isolates using a multiplex real-time PCR assay. Diagn. Microbiol. Infect. Dis. 84(4):298-303.

Berdal BP, Farshy CE, Feeley JC. 1979. Detection of *Legionella pneumonophila* antigen in urine by enzyme-linked immunospecific assay. J. Clin. Microbiol. 9(5):575-578.

Berendt RF. 1980. Survival of *Legionella pneumophila* in aerosols: effect of relative humidity. J. Infect. Dis. 141(5):689-689.

Bernander S, Jacobson K, Helbig JH, Lück PC, Lundholm M. 2003. A hospital-associated outbreak of Legionnaires disease caused by *Legionella pneumophila* serogroup 1 is characterized by stable genetic fingerprinting but variable monoclonal antibody patterns. J. Clin. Microbiol. 41(6):2503-2508.

Beyrer K, Lai S, Dreesman J, Lee JV, Joseph C, Harrison T, Surman-Lee S, Lück C, Brodhun B, Buchholz U, Windorfer A. 2007. Legionnaires disease outbreak associated with a cruise liner, August 2003: epidemiological and microbiological findings. Epidemiol. Infect. 135(5):802-810.

Boldur I, Cohen A, Tamarin-Landau R, Sompolinsky D. 1987. Isolation of *Legionella pneumophila* from calves and the prevalence of antibodies in cattle, sheep, horses, antelopes, buffaloes and rabbits. Vet. Microbiol. 13(4):313-20.

Borchardt J, Helbig JH, Lück PC. 2008. Occurrence and distribution of sequence types among *Legionella pneumophila* strains isolated from patients in Germany: common features and differences to other regions of the world. Eur. J. Clin. Microbiol. Infect. Dis. 27(1):29-36.

Brassinga AK, Kinchen JM, Cupp ME, Day SR, Hoffman PS, Sifri CD. 2010. *Caenorhabditis* is a metazoan host for *Legionella*. Cell. Microbiol. 12(3):343-61.

Braun SD, Ziegler A, Methner U, Slickers P, Keiling S, Monecke S, Ehricht R. 2012. Fast DNA serotyping and antimicrobial resistance gene determination of *Salmonella enterica* with an oligonucleotide microarray-based assay. PLoS One 7(10):e46489.

Brenner DJ, Steigerwalt AG, McDade JE. 1979. Classification of the Legionnaires' disease bacterium: *Legionella pneumophila*, genus novum, species nova, of the family Legionellaceae, familia nova. Ann. Intern. Med. 90(4):656-658.

Burckhardt F, Brion A, Lahm A, Koch HU, Prior K, Petzold M, Harmsen D, Lück C. 2016. Confirming Legionnaires' disease outbreak by genome-based method, Germany, 2012. Emerg. Infect. Dis. 22(7).

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC.Bioinformatics. 10:421.

Cazalet C, Rusniok C, Brüggemann H, Zidane N, Magnier A, Ma L, Tichit M, Jarraud S, Bouchier C, Vandenesch F, Kunst F, Etienne J, Glaser P, Buchrieser C. 2004. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. Nat. Genet. 36(11):1165-1173.

Cazalet C, Jarraud S, Ghavi-Helm Y, Kunst F, Glaser P, Etienne J, Buchrieser C. 2008. Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. Genome Res. 18(3):431-441.

Centers for Disease Control and Prevention. 2011. Legionellosis - United States, 2000-2009. MMWR Morb Mortal Weekly Report - Centers for Disease Control and Prevention 60(32):1083-1086.

Centers for Disease Control and Prevention. 2015. Summary of Notifiable Infectious Diseases and Conditions - United States, 2013. MMWR Morb Mortal Weekly Report - Centers for Disease Control and Prevention 62(53).

Chasqueira MJ, Rodrigues L, Nascimento M, Marques T. 2009. Sequence-based and monoclonal antibody typing of *Legionella pneumophila* isolated from patients in Portugal during 1987-2008. Eurosurveillance 14(28).

Chien M, Morozova I, Shi S, Sheng H, Chen J, Gomez SM, Asamani G, Hill K, Nuara J, Feder M, Rineer J, Greenberg JJ, Steshenko V, Park SH, Zhao B, Teplitskaya E, Edwards JR, Pampou S, Georghiou A, Chou IC, Iannuccilli W, Ulz ME, Kim DH, Geringer-Sameth A, Goldsberry C, Morozov P, Fischer SG, Segal G, Qu X, Rzhetsky A, Zhang P, Cayanis E, De Jong PJ, Ju J, Kalachikov S, Shuman HA, Russo JJ. 2004. The genomic sequence of the accidental pathogen *Legionella pneumophila*. Science 305(5692):1966-1968.

Ciesielski CA, Blaser MJ, Wang WL. 1986. Serogroup specificity of *Legionella pneumophila* is related to lipopolysaccharide characteristics. Infect. Immun. 51(2):397-404.

Clarivate Analytics. 2017. 2016 Journal Citation Report®. [upd: 14.06.2017.] Acc: 16.08.2017. https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action.

Collins SL, Stevenson D, Mentasti M, Shaw A, Johnson A, Crossley L, Willis C. 2017. High prevalence of *Legionella* in non-passenger merchant vessels. Epidemiol. Infect. 145(4):647-655.

Correia AM, Ferreira JS, Borges V, Nunes A, Gomes B, Capucho R, Goncalves J, Antunes DM, Almeida S, Mendes A, Guerreiro M, Sampaio DA, Vieira L, Machado J, Simoes MJ, Goncalves P, Gomes JP. 2016. Probable person-to-person transmission of Legionnaires' Disease. N. Engl. J. Med. 374(5):497-498.

Coscollá M, Comas I, Gonzlez-Candelas F. 2011. Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. Mol. Biol. Evol. 28(2):985-1001.

Council of State and Territorial Epidemiologists. 2010. Public Health Reporting and National Notification for Legionellosis (Position Paper 09-ID-45). CST.

Cristovam E, Almeida D, Caldeira D, Ferreira JJ, Marques T. 2017. Accuracy of diagnostic tests for Legionnaires' disease: a systematic review. J. Med. Microbiol. 66(4):485-489.

Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 43(3):e15.

D'Auria G, Jimnez-Hernndez N, Peris-Bondia F, Moya A, Latorre A. 2010. *Legionella pneumophila* pangenome reveals strain-specific virulence factors. BMC Genomics 11(1):181.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5(6):e11147.

David S, Mentasti M, Tewolde R, Aslett M, Harris SR, Afshar B, Underwood A, Fry NK, Parkhill J, Harrison TG. 2016a. Evaluation of an optimal epidemiologic typing scheme for *Legionella pneumophila* with whole genome sequence data using validation guidelines. J. Clin. Microbiol. 54(8):2135-2148.

David S, Rusniok C, Mentasti M, Gomez-Valero L, Harris SR, Lechat P, Lees J, Ginevra C, Glaser P, Ma L, Bouchier C, Underwood A, Jarraud S, Harrison TG, Parkhill J, Buchrieser C. 2016b. Multiple major disease-associated clones of *Legionella pneumophila* have emerged recently and independently. Genome Res. 26(11):1555-1564.

David S, Sánchez-Busó L, Harris SR, Marttinen P, Rusniok C, Buchrieser C, Harrison TG, Parkhill J. 2017. Dynamics and impact of homologous recombination on the evolution of *Legionella pneumophila*. PLoS Genet 13(6):e1006855.

Declerck P. 2010. Biofilms: the environmental playground of *Legionella pneumophila*. Environ. Microbiol. 12(3):557-566.

Den Boer JW, Yzerman EP, Jansen R, Bruin JP, Verhoef LP, Neve G, van der Zwaluw K. 2007. Legionnaires' disease and gardening. Clin. Microbiol. Infect. 13(1):88-91.

Den Boer JW, Euser SM, Nagelkerke NJ, Schuren F, Jarraud S, Etienne J. 2013. Prediction of the origin of French *Legionella pneumophila* strains using a mixed-genome microarray. BMC Genomics 14:435.

Dennis PJ, Lee JV. 1988. Differences in aerosol survival between pathogenic and non-pathogenic strains of *Legionella pneumophila* serogroup 1. J. Appl. Bacteriol. 65(2):135-141.

Diederen BM. 2008. *Legionella* spp. and Legionnaires disease. J. Infect. 56(1):1-12.

Ditommaso S, Ricciardi E, Giacomuzzi M, Arauco Rivera SR, Zotti CM. 2015. *Legionella* in water samples: how can you interpret the results obtained by quantitative PCR? Mol. Cell. Probes 29(1):7-12.

Dorer MS, Kirton D, Bader JS, Isberg RR. 2006. RNA interference analysis of *Legionella* in *Drosophila* cells: exploitation of early secretory apparatus dynamics. PLoS Pathog. 2(4):e34.

Dunne WM, Jr., Picot N, van Belkum A. 2017. Laboratory Tests for Legionnaire's Disease. Infect. Dis. Clin. North Am. 31(1):167-178.

Edagawa A, Kimura A, Kawabuchi-Kurata T, Adachi S, Furuhata K, Miyamoto H. 2015. Investigation of *Legionella* contamination in bath water samples by culture, amoebic co-culture, and real-time quantitative PCR methods. Int. J. Environ. Res. Public Health 12(10):13118-30.

Edelstein PH, Meyer RD, Finegold SM. 1979. Isolation of *Legionella pneumophila* from blood. Lancet 1(8119):750-751.

Ellsäßer G, Hiller I. 2004. [Zu zwei nosokomialen Legionellose-Ausbrüchen in einem Klinikum im Land Brandenburg]. Epidemiologisches Bulletin 11:89-91.

Enright MC, Spratt BG. 1999. Multilocus sequence typing. Trends Microbiol. 7(12):482-7.

Ensminger AW, Isberg RR. 2009. Legionella pneumophila Dot/Icm translocated substrates: a sum of parts. Curr. Opin. Microbiol.

Ensminger AW, Yassin Y, Miron A, Isberg RR. 2012. Experimental evolution of *Legionella pneumophila* in mouse macrophages leads to strains with altered determinants of environmental survival. PLoS Pathog. 8(5):e1002731.

Ensminger AW. 2016. *Legionella pneumophila*, armed to the hilt: justifying the largest arsenal of effectors in the bacterial world. Curr. Opin. Microbiol. 29:74-80.

Escoll P, Rolando M, Gomez-Valero L, Buchrieser C. 2013. From amoeba to macrophages: exploring the molecular mechanisms of *Legionella pneumophila* infection in both hosts. Curr. Top. Microbiol. Immunol. 376:1-34.

Essig A, von Baum H, Gonser T, Haerter G, Lück C. 2016. Microbiological diagnosis and molecular typing of *Legionella* strains during an outbreak of legionellosis in Southern Germany. Int. J. Med. Microbiol. 306(2):109-114.

European Centre for Disease Prevention and Control. 2012. European Legionnaires' Disease Surveillance Network (ELDSNet): Operating procedures. Stockholm: European Centre for Disease Control and Prevention.

European Centre for Disease Prevention and Control. 2015a. Case definitions for Legionnaires' disease. Stockholm: European Centre for Disease Prevention and Control.

European Centre for Disease Prevention and Control. 2015b. Legionnaires' disease in Europe, 2013. Stockholm: European Centre for Disease Prevention and Control.

European Centre for Disease Prevention and Control. 2016. Annual Epidemiological Report 2014 - Legionnaires' Disease. [upd: 30.12.2016.] Acc: 01.08.2017.

European Commission. 2012. Legislation on Communicable diseases: amending Decision 2002/253/EC laying down case definitions for reporting communicable diseases to the Community network under Decision No 2119/98/EC of the European Parliament and of the Council. Official Journal of the European Union. p 17-18.

Fabbi M, Magnino S, Scanziani E, Pastoris MC. 1993. *Legionella* pneumonia in a calf. J. Infect. 27(2):215-216.

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(12):4052-4058.

Faulkner G, Garduno RA. 2002. Ultrastructural analysis of differentiation in *Legionella pneumophila*. J. Bacteriol. 184(24):7025-7041.

Feddersen A, Meyer HG, Matthes P, Bhakdi S, Husmann M. 2000. GyrA sequence-based typing of *Legionella*. Med Microbiol Immunol (Berl) 189(1):7-11.

Feeley JC, Gorman GW, Weaver RE, Mackel DC, Smith HW. 1978. Primary isolation media for Legionnaires' disease bacterium. J. Clin. Microbiol. 8(3):320-5.

Fields BS, Benson RF, Besser RE. 2002. Legionella and Legionnaires disease: 25 years of investigation. Clin. Microbiol. Rev. 15(3):506-526.

Fitzgeorge RB, Baskerville A, Broster M, Hambleton P, Dennis PJ. 1983. Aerosol infection of animals with strains of *Legionella pneumophila* of different virulence: comparison with intraperitoneal and intranasal routes of infection. J. Hyg. (Lond.) 90(1):81-89.

Fliermans CB, Cherry WB, Orrison LH, Smith SJ, Tison DL, Pope DH. 1981. Ecological distribution of *Legionella pneumophila*. Appl. Environ. Microbiol. 41(1):9-16.

Fraser DW, Tsai TR, Orenstein W, Parkin WE, Beecham HJ, Sharrar RG, Harris J, Mallison GF, Martin SM, McDade JE, Shepard CC, Brachman PS. 1977. Legionnaires' disease: description of an epidemic of pneumonia. N. Engl. J. Med. 297(22):1189-97.

Fraser DW. 1980. Legionellosis: evidence of airborne transmission. Ann. N. Y. Acad. Sci. 353:61-6.

Fry NK, Alexiou-Daniel S, Bangsborg JM, Bernander S, Castellani PM, Etienne J, Forsblom B, Gaia V, Helbig JH, Lindsay D, Lück C, Pelaz C, Uldum SA, Harrison TG. 1999. A multicenter evaluation of genotypic methods for the epidemiologic typing of *Legionella pneumophila* serogroup 1: results of a pan-European study. Clin. Microbiol. Infect. 5(8):462-477.

Fry NK, Afshar B, Bellamy W, Underwood AP, Ratcliff RM, Harrison TG, members of the European Working Group for Legionella I. 2007. Identification of *Legionella* spp. by 19 European reference laboratories: results of the European Working Group for Legionella Infections external quality assessment scheme using DNA sequencing of the macrophage infectivity potentiator gene and dedicated online tools. Clin. Microbiol. Infect. 13(11):1119-1124.

Gaboriaud F, Gee ML, Strugnell R, Duval JF. 2008. Coupled electrostatic, hydrodynamic, and mechanical properties of bacterial interfaces in aqueous media. Langmuir 24(19):10988-95.

Gaia V, Fry NK, Harrison TG, Peduzzi R. 2003. Sequence-based typing of *Legionella pneumophila* serogroup 1 offers the potential for true portability in legionellosis outbreak investigation. J. Clin. Microbiol. 41(7):2932-2939.

Gaia V, Fry NK, Afshar B, Lück PC, Meugnier H, Etienne J, Peduzzi R, Harrison TG. 2005. Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. J. Clin. Microbiol. 43(5):2047-2052.

Ginevra C, Lopez M, Forey F, Reyrolle M, Meugnier H, Vandenesch F, Etienne J, Jarraud S, Molmeret M. 2009a. Evaluation of a nested-PCR-derived sequence-based typing method applied directly to respiratory samples from patients with Legionnaires' Disease. J. Clin. Microbiol. 47(4):981-987.

Ginevra C, Duclos A, Vanhems P, Campse C, Forey F, Lina G, Che D, Etienne J, Jarraud S. 2009b. Host-related risk factors and clinical features of community-acquired Legionnaires' disease due to the Paris and Lorraine endemic strains, 1998-2007, France. Clin. Infect. Dis.

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(3):696-701.

Glaze PA, Watson DC, Young NM, Tanner ME. 2008. Biosynthesis of CMP-N,N -diacetyllegionaminic acid from UDP-N,N -diacetylbacillosamine in *Legionella pneumophila*. Biochemistry 47(10):3272-3282.

Glick TH, Gregg MB, Berman B, Mallison G, Rhodes WW, Jr., Kassanoff I. 1978. Pontiac fever. An epidemic of unknown etiology in a health department: I. Clinical and epidemiologic aspects. Am. J. Epidemiol. 107(2):149-60.

Glöckner G, Albert-Weissenberger C, Weinmann E, Jacobi S, Schunder E, Steinert M, Hacker J, Heuner K. 2008. Identification and characterization of a new conjugation/type IVA secretion system (*trb/tra*) of *Legionella pneumophila* Corby localized on two mobile genomic islands. Int. J. Med. Microbiol. 298(5-6):411-428.

Gomez-Valero L, Rusniok C, Cazalet C, Buchrieser C. 2011a. Comparative and functional genomics of *Legionella* identified eucaryotic like proteins as key players in host-pathogen interactions. Front. Microbiol. 12(2):1-20.

Gomez-Valero L, Rusniok C, Jarraud S, Vacherie B, Rouy Z, Barbe V, Medigue C, Etienne J, Buchrieser C. 2011b. Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. BMC Genomics 12:536.

Gordon M, Yakunin E, Valinsky L, Chalifa-Caspi V, Moran-Gilad J, Infections ESGfL. 2017. A bioinformatics tool for ensuring the backwards compatibility of *Legionella pneumophila* typing in the genomic era. Clin. Microbiol. Infect. 23(5):306-310.

Gosselin F, Duval JF, Simonet J, Ginevra C, Gaboriaud F, Jarraud S, Mathieu L. 2011. Impact of the virulence-associated MAb3/1 epitope on the physicochemical surface properties of *Legionella pneumophila* sg1: An issue to explain infection potential? Colloids Surf B Biointerfaces 82(2):283-290.

Graham FF, White PS, Harte DJ, Kingham SP. 2012. Changing epidemiological trends of legionellosis in New Zealand, 1979-2009. Epidemiol. Infect. 140(8):1481-96.

Graham RM, Doyle CJ, Jennison AV. 2014. Real-time investigation of a *Legionella pneumophila* outbreak using whole genome sequencing. Epidemiol. Infect. 142(11):2347-2351.

Greenfield LK, Whitfield C. 2012. Synthesis of lipopolysaccharide *O*-antigens by ABC transporter-dependent pathways. Carbohydr. Res. 356:12-24.

Hägele S, Köhler R, Merkert H, Schleicher M, Hacker J, Steinert M. 2000. *Dictyostelium discoideum*: a new host model system for intracellular pathogens of the genus *Legionella*. Cell. Microbiol. 2(2):165-71.

Harding CR, Schroeder GN, Reynolds S, Kosta A, Collins JW, Mousnier A, Frankel G. 2012. *Legionella pneumophila* pathogenesis in the *Galleria mellonella* infection model. Infect. Immun. 80(8):2780-90.

Harrison TG, Doshi N, Fry NK, Joseph CA. 2006. Comparison of clinical and environmental isolates of *Legionella pneumophila* obtained in the UK over 19 years. Clin. Microbiol. Infect. 13(1):78-85.

Harrison TG, Afshar B, Doshi N, Fry NK, Lee JV. 2009. Distribution of *Legionella pneumophila* serogroups, monoclonal antibody subgroups and DNA sequence types in recent clinical and environmental isolates from England and Wales (2000-2008). Eur. J. Clin. Microbiol. Infect. Dis. 28(7):781-791.

Helbig JH, Lück PC, Knirel YA, Witzleb W, Zähringer U. 1995. Molecular characterization of a virulence-associated epitope on the lipopolysaccharide of *Legionella pneumophila* serogroup 1. Epidemiol. Infect. 115(1):71-8.

Helbig JH, Kurtz JB, Pastoris MC, Pelaz C, Lück PC. 1997. Antigenic lipopolysaccharide components of *Legionella pneumophila* recognized by monoclonal antibodies: possibilities and limitations for division of the species into serogroups. J. Clin. Microbiol. 35(11):2841-2845.

Helbig JH, Bernander S, Castellani PM, Etienne J, Gaia V, Lauwers S, Lindsay D, Lück PC, Marques T, Mentula S, Peeters MF, Pelaz C, Struelens M, Uldum SA, Wewalka G, Harrison TG. 2002. Pan-European study on culture-proven Legionnaires disease: distribution of *Legionella pneumophila* serogroups and monoclonal subgroups. Eur. J. Clin. Microbiol. Infect. Dis. 21(10):710-716.

Helbig JH, Benson RF, Pelaz C, Jacobs E, Lück PC. 2007. Identification and serotyping of atypical *Legionella pneumophila* strains isolated from human and environmental sources. J. Appl. Microbiol. 102(1):100-105.

Helbig JH, Jacobs E, Lück C. 2012. *Legionella pneumophila* urinary antigen subtyping using monoclonal antibodies as a tool for epidemiological investigations. Eur. J. Clin. Microbiol. Infect. Dis. 31:1673-1677.

Hellinga JR, Garduno RA, Kormish JD, Tanner JR, Khan D, Buchko K, Jimenez C, Pinette MM, Brassinga AK. 2015. Identification of vacuoles containing extraintestinal differentiated forms of *Legionella pneumophila* in colonized *Caenorhabditis elegans* soil nematodes. Microbiologyopen.

Hering A. 2009. Legionellen-Erkrankungen auf einem Kreuzfahrtschiff. Epidemiologisches Bulletin 10:85-87.

Hines SA, Chappie DJ, Lordo RA, Miller BD, Janke RJ, Lindquist HA, Fox KR, Ernst HS, Taft SC. 2014. Assessment of relative potential for *Legionella* species or surrogates inhalation exposure from common water uses. Water Res 56:203-13.

Horwitz MA, Silverstein SC. 1980. Legionnaires' disease bacterium (*Legionella pneumophila*) multiples intracellularly in human monocytes. J. Clin. Invest. 66(3):441-450.

Hunter PR, Gaston MA. 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. 26(11):2465-6.

International Organization for Standardization. 1998. ISO 11731:1998: Water Quality - Detection and enumeration of *Legionella*. International Organization for Standardization, Genevra, Switzerland.

International Organization for Standardization. 2008. EN ISO 11731-2. Water Quality-Detection and Enumeration of *Legionella* Part2: Direct membrane filtration method for waters with low bacterial counts. International Organization for Standardization, Genevra, Switzerland.

International Organization for Standardization. 2012. ISO/TS 12869:2012: Water Quality - Detection and quantification of *Legionella* spp. and /or *Legionella pneumophila* by concentration and genic amplification by quantitative polymerase chain reaction (qPCR). International Organization for Standardization, Genevra, Switzerland.

Isberg RR, O' Connor TJ, Heidtman M. 2009. The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. Nat. Rev. Microbiol. 7(1):13-24.

Ishizaki N, Sogawa K, Inoue H, Agata K, Edagawa A, Miyamoto H, Fukuyama M, Furuhata K. 2016. *Legionella thermalis* sp. nov., isolated from hot spring water in Tokyo, Japan. Microbiol. Immunol. 60(3):203-208.

Joly JR, Chen YY, Ramsay D. 1983. Serogrouping and subtyping of *Legionella pneumophila* with monoclonal antibodies. J. Clin. Microbiol. 18(5):1040-1046.

Joly JR, McKinney RM, Tobin JO, Bibb WF, Watkins ID, Ramsay D. 1986. Development of a standardized subgrouping scheme for *Legionella pneumophila* serogroup 1 using monoclonal antibodies. J. Clin. Microbiol. 23(4):768-771.

Joseph C, Ricketts K, Yadav R, Patel S, Patel S. 2010a. Travel-associated Legionnaires' disease in Europe in 2009. Euro Surveill. 15(41).

Joseph CA, Ricketts KD. 2010. Legionnaires disease in Europe 2007-2008. Euro Surveill. 15(8):19493.

Joseph CA, Ricketts KD, Yadav R, Patel S, Legionella EWG. 2010b. Travel-associated Legionnaires' disease in Europe in 2009. Eurosurveillance 15(41):5-11.

Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinoeski J, Mellmann A, Goesmann A, von Haeseler A, Stoye J, Harmsen D. 2013. Updating benchtop sequencing performance comparison. Nat. Biotechnol. 31(4):294-296.

Kahlisch L, Henne K, Draheim J, Brettar I, Höfle MG. 2010. High-resolution *in situ* genotyping of *Legionella pneumophila* populations in drinking water by multiple-locus variable-number tandem-repeat analysis using environmental DNA. Appl. Environ. Microbiol. 76(18):6186-6195.

Kirschner AKT. 2016. Determination of viable legionellae in engineered water systems: Do we find what we are looking for? Water Research 93:276-288.

Knirel YA, Rietschel ET, Marre R, Zähringer U. 1994. The structure of the O-specific chain of *Legionella pneumophila* serogroup 1 lipopolysaccharide. Eur. J. Biochem. 221(1):239-245.

Knirel YA, Moll H, Zähringer U. 1996. Structural study of a highly O-acetylated core of *Legionella pneumophila* serogroup 1 lipopolysaccharide. Carbohydr. Res. 293(2):223-234.

Knirel YA, Senchenkova SN, Kocharova NA, Shashkov AS, Helbig JH, Zähringer U. 2001. Identification of a homopolymer of 5-acetamidino-7-acetamido-3,5,7,9-tetradeoxy-D-glycero-D-talo-nonulosonic acid in the lipopolysaccharides of *Legionella pneumophila* non-1 serogroups. Biochemistry-Moscow 66(9):1035-1041.

Knirel YA, Shashkov AS, Tsvetkov YE, Jansson PE, Zähringer U. 2003. 5,7-diamino-3,5,7,9-tetradeoxynon-2-ulosonic acids in bacterial glycopolymers: chemistry and biochemistry. Adv. Carbohydr. Chem. Biochem. 58:371-417.

Knirel YA, Valvano MA. 2011. Bacterial Lipopolysaccharides: Structure, Chemical Synthesis, Biogenisis and Interaction with Host Cells. Vienna: Springer.

Kooistra O, Lüneberg E, Lindner B, Knirel YA, Frosch M, Zähringer U. 2001. Complex O-acetylation in *Legionella pneumophila* serogroup 1 lipopolysaccharide. Evidence for two genes involved in 8-O-acetylation of legionaminic acid. Biochemistry 40(25):7630-7640.

Kooistra O, Herfurth L, Lüneberg E, Frosch M, Peters T, Zähringer U. 2002a. Epitope mapping of the O-chain polysaccharide of *Legionella pneumophila* serogroup 1 lipopolysaccharide by saturation-transfer-difference NMR spectroscopy. Eur. J. Biochem. 269(2):573-582.

Kooistra O, Lüneberg E, Knirel YA, Frosch M, Zähringer U. 2002b. N-Methylation in polylegionaminic acid is associated with the phase-variable epitope of *Legionella pneumophila* serogroup 1 lipopolysaccharide. Identification of 5-(N,N-dimethylacetimidoyl)amino and 5-acetimidoyl(N-methyl)amino-7-acetamido-3,5,7,9-tetradeoxynon-2-ulosonic acid in the O-chain polysaccharide. Eur. J. Biochem. 269(2):560-572.

Kozak-Muiznieks NA, Lucas CE, Brown E, Pondo T, Taylor TH, Frace M, Miskowski D, Winchell JM. 2014. Prevalence of sequence types among clinical and environmental isolates of *Legionella pneumophila* serogroup 1 in the United States from 1982 to 2012. J. Clin. Microbiol. 52(1):201-211.

Kozak NA, Benson RF, Brown E, Alexander NT, Taylor TH, Shelton BG, Fields BS. 2009. Distribution of *lag-1* alleles and sequence-based types among *Legionella pneumophila* serogroup 1 clinical and environmental isolates in the United States. J. Clin. Microbiol. 47(8):2525-2535.

Krause C. 2014. Vorkommen und Verteilung von Spezies, Serogruppen, monoklonalen Subtypen und Sequenztypen von *Legionella* spp. unter klinischen Isolaten in Deutschland und weltweit. Dresden: Medical Faculty Universtiy of Technology Dresden. 104 p.

Kusnetsov JM, Ottoila E, Martikainen PJ. 1996. Growth, respiration and survival of *Legionella pneumophila* at high temperatures. J. Appl. Bacteriol. 81(4):341-347.

Ledesma E, Camaró ML, Carbonell E, Sacristn T, Mart A, Pellicer S, Llorca J, Herrero P, Dasi MA. 1995. Subtyping of *Legionella pneumophila* isolates by arbitrarily primed polymerase chain reaction. Can. J. Microbiol. 41(9):846-848.

Lee JV, Lai S, Exner M, Lenz J, Gaia V, Casati S, Hartemann P, Lück C, Pangon B, Ricci ML, Scaturro M, Fontana S, Sabria M, Sanchez I, Assaf S, Surman-Lee S. 2011. An international trial of quantitative PCR for monitoring *Legionella* in artificial water systems. J. Appl. Microbiol. 110(4):1032-44.

Lesnik R, Brettar I, Höfle MG. 2016. *Legionella* species diversity and dynamics from surface reservoir to tap water: from cold adaptation to thermophily. ISME J 10(5):1064-80.

Lévesque S, Plante PL, Mendis N, Cantin P, Marchand G, Charest H, Raymond F, Huot C, Goupil-Sormany I, Desbiens F, Faucher SP, Corbeil J, Tremblay C. 2014. Genomic characterization of a large outbreak of *Legionella pneumophila* serogroup 1 strains in Quebec City, 2012. PLoS One 9(8):e103852.

Lewis AL, Hensler ME, Varki A, Nizet V. 2006. The group B streptococcal sialic acid *O*-acetyltransferase is encoded by *neuD*, a conserved component of bacterial sialic acid biosynthetic gene clusters. J. Biol. Chem. 281(16):11186-92.

Liu M, Haenssler E, Uehara T, Losick VP, Park JT, Isberg RR. 2012. The *Legionella pneumophila* EnhC protein interferes with immunostimulatory muramyl peptide production to evade innate immunity. Cell Host Microbe 12(2):166-76.

Lück C, Bender L, Ott M, Helbig JH, Hacker J. 1991. Analysis of *Legionella pneumophila* serogroup 6 strains isolated from a hospital warm water supply over a three-year period by using genomic long-range mapping techniques and monoclonal antibodies. Appl. Environ. Microbiol. 57(11):3226-3231.

Lück C, Freier T, Steudel C, Knirel YA, Lüneberg E, Zähringer U, Helbig JH. 2001. A point mutation in the active site of *Legionella pneumophila O*-acetyltransferase results in modified lipopolysaccharide but does not influence virulence. Int. J. Med. Microbiol. 291(5):345-352.

Lück C. 2011. *Legionella pneumophila*. Genetic diversity of patients and environmental isolates (in German). Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz 54(6):693-698.

Lück C, Brzuszkiewicz E, Rydzewski K, Koshkolda T, Sarnow K, Essig A, Heuner K. 2015. Subtyping of the *Legionella pneumophila* "Ulm" outbreak strain using the CRISPR-Cas system. Int. J. Med. Microbiol. 305(8):828-37.

Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26(4):1107-15.

Lüneberg E, Zetzmann N, Alber D, Knirel YA, Kooistra O, Zähringer U, Frosch M. 2000. Cloning and functional characterization of a 30 kb gene locus required for lipopolysaccharide biosynthesis in *Legionella pneumophila*. Int. J. Med. Microbiol. 290(1):37-49.

Lüneberg E, Mayer B, Daryab N, Kooistra O, Zähringer U, Rohde M, Swanson J, Frosch M. 2001. Chromosomal insertion and excision of a 30 kb unstable genetic element is responsible for phase variation of lipopolysaccharide and other virulence determinants in *Legionella pneumophila*. Mol. Microbiol. 39(5):1259-1271.

Lurie-Weinberger MN, Gomez-Valero L, Merault N, Glöckner G, Buchrieser C, Gophna U. 2010. The origins of eukaryotic-like proteins in *Legionella pneumophila*. Int. J. Med. Microbiol. 300(7):470-481.

Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95(6):3140-5.

Maiden MC, Jansen 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(10):728-36.

Maisa A, Brockmann A, Renken F, Lück C, Pleischl S, Exner M, Daniels-Haardt I, Jurke A. 2015. Epidemiological investigation and case-control study: a Legionnaires' disease outbreak associated with cooling towers in Warstein, Germany, August-September 2013. Euro Surveill. 20(46).

Maiwald M, Helbig JH, Lück PC. 1998. Laboratory methods for the diagnosis of *Legionella* infections. J. Microbiol. Methods 33(1):59-79.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39(Database issue):D225-9.

Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, Corander J. 2012. Detection of recombination events in bacterial genomes from large population samples. Nucleic Acids Res. 40(1):e6.

McAdam PR, Vander Broek CW, Lindsay DS, Ward MJ, Hanson MF, Gillies M, Watson M, Stevens JM, Edwards GF, Fitzgerald JR. 2014. Gene flow in environmental *Legionella pneumophila* leads to genetic and pathogenic heterogeneity within a Legionnaires' disease outbreak. Genome Biol. 15(11):504.

McDade JE, Shepard CC, Fraser DW, Tsai TR, Redus MA, Dowdle WR. 1977. Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. N. Engl. J. Med. 297(22):1197-203.

McDade JE, Brenner DJ, Bozeman FM. 1979. Legionnaires' disease bacterium isolated in 1947. Ann. Intern. Med. 90(4):659-661.

McNally DJ, Schoenhofen IC, Mulrooney EF, Whitfield DM, Vinogradov E, Lam JS, Logan SM, Brisson JR. 2006. Identification of labile UDP-ketosugars in *Helicobacter pylori*, *Campylobacter jejuni* and *Pseudomonas aeruginosa*: key metabolites used to make glycan virulence factors. Chembiochem 7(12):1865-8.

McNally DJ, Aubry AJ, Hui JP, Khieu NH, Whitfield D, Ewing CP, Guerry P, Brisson JR, Logan SM, Soo EC. 2007. Targeted metabolomics analysis of *Campylobacter coli* VC167 reveals legionaminic acid derivatives as novel flagellar glycans. J. Biol. Chem. 282(19):14463-75.

Mentasti M, Fry N, Afshar B, Palepou-Foxley C, Naik F, Harrison T. 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(8):2017-2028.

Mentasti M, Underwood A, Lück C, Kozak-Muiznieks NA, Harrison TG, Fry NK. 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(7):O435-O441.

Mentasti M, Kese D, Echahidi F, Uldum SA, Afshar B, David S, Mrazek J, R. DM, Harrison TG, Chalker VJ. 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(7):1387-1393.

Mentasti M, Cassier P, David S, Ginevra C, Gomez-Valero L, Underwood A, Afshar B, Etienne J, Parkhill J, Chalker V, Buchrieser C, Harrison TG, Jarraud S, Infections ESGfL. 2017. Rapid detection and evolutionary analysis of *Legionella pneumophila* serogroup 1 sequence type 47. Clin. Microbiol. Infect. 23(4):264 e1-264 e9.

Mérault N, Rusniok C, Jarraud S, Gomez-Valero L, Cazalet C, Marin M, Brachet E, Aegerter P, Gaillard JL, Etienne J, Herrmann JL, Group tD-IS, 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(5):1708-1717.

Mercante JW, Winchell JM. 2015. Current and emerging *Legionella* diagnostics for laboratory and outbreak investigations. Clin. Microbiol. Rev. 28(1):95-133.

Mercante JW, Morrison SS, Desai HP, Raphael BH, Winchell JM. 2016. Genomic analysis reveals novel diversity among the 1976 Philadelphia Legionnaires' Disease outbreak isolates and additional ST36 strains. PLoS One 11(9):e0164074.

Monecke S, Ehricht R. 2005. Rapid genotyping of methicillin-resistant *Staphylococcus aureus* (MRSA) isolates using miniaturised oligonucleotide arrays. Clin. Microbiol. Infect. 11(10):825-833.

Monecke S, Berger-Bachi B, Coombs G, Holmes A, Kay I, Kearns A, Linde HJ, O'Brien F, Slickers P, Ehricht R. 2007. Comparative genomics and DNA array-based genotyping of pandemic Staphylococcus aureus strains encoding Panton-Valentine leukocidin. Clin. Microbiol. Infect. 13(3):236-49.

Monecke S, Jatzwauk L, Weber S, Slickers P, Ehricht R. 2008. DNA microarray-based genotyping of methicillin-resistant *Staphylococcus aureus* strains from Eastern Saxony. Clin. Microbiol. Infect. 14(6):534-545.

Moran-Gilad J, Prior K, Yakunin E, Harrison TG, Underwood A, Lazarovitch T, Valinsky L, Lück C, Krux F, Agmon V, Grotto I, Harmsen D. 2015. Design and application of a core genome multilocus sequence typing scheme for investigation of Legionnaires' disease incidents. Euro Surveill. 20(28).

Muder RR, Yu VL. 2002. Infection due to *Legionella* species other than *L. pneumophila*. Clin. Infect. Dis. 35(8):990-998.

Pannier K, Heuner K, Lück C. 2010. Variable genetic element typing: a quick method for epidemiological subtyping of *Legionella pneumophila*. Eur. J. Clin. Microbiol. Infect. Dis. 29(4):481-487.

Parte AC, Euzéby JP. 2017. List of Prokaryotic names with Standing in Nomenclature. [upd: 09.07.2017.] Acc: 10.08.2017. http://www.bacterio.net/.

Pearce MM, Theodoropoulos N, Mandel MJ, Brown E, Reed KD, Cianciotto NP. 2012. *Legionella cardiaca* sp. nov., isolated from a case of native valve endocarditis in a human heart. IJSB 62(Pt 12):2946-2954.

Pereira AJ, Broadbent J, Mahgoub H, Morgan O, Bracebridge S, Reacher M, Ibbotson S, Lee JV, Harrison TG, Nair P. 2006. Legionnaires disease: when an outbreak is not an outbreak. Euro Surveill. 11(11).

Petzold M, Ehricht R, Slickers P, Pleischl S, Brockmann A, Exner M, Monecke S, Lück C. 2016. Rapid genotyping of *Legionella pneumophila* serogroup 1 strains by a novel DNA microarray-based assay during the outbreak investigation in Warstein, Germany 2013. Int J Hyg Environ Health doi:10.1016/j.ijheh.2016.02.004.

Phin N, Parry-Ford F, Harrison T, Stagg HR, Zhang N, Kumar K, Lortholary O, Zumla A, Abubakar I. 2014. Epidemiology and clinical management of Legionnaires' disease. Lancet Infect. Dis. 14(10):1011-1021.

Pourcel C, Visca P, Afshar B, Arezzo D, Vergnaud G, Fry NK. 2007. Identification of variable-number tandem-repeat (VNTR) sequences in *Legionella pneumophila* and development of an optimized multiple-locus VNTR analysis typing scheme. J. Clin. Microbiol. 45(4):1190-1199.

Price MN, Huang KH, Alm EJ, Arkin AP. 2005. A novel method for accurate operon predictions in all sequenced prokaryotes. Nucleic Acids Res. 33(3):880-92.

QGIS Development Team. 2012. Quantum Geographic Information System - QGIS. [upd: 10.08.2017.] Acc: http://qgis.osgeo.org.

Raphael BH, Baker DJ, Nazarian E, Lapierre P, Bopp D, Kozak-Muiznieks NA, Morrison SS, Lucas CE, Mercante JW, Musser KA, Winchell JM. 2016. Genomic resolution of outbreak-associated *Legionella pneumophila* serogroup 1 isolates from New York State. Appl. Environ. Microbiol. 82(12):3582-3590.

Rasch J, Krüger S, Fontvieille D, Ünal C, Michel R, Labrosse A, Steinert M. 2016. Legionella-protozoa-nematode interactions in aquatic biofilms and influence of Mip on *Caenorhabditis elegans* colonization. IJMM 306(6):443-451.

Ratzow S, Gaia V, Helbig JH, Fry NK, Lück PC. 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(6):1965-1968.

Reidt U, Geisberger B, Heller C, Friedberger A. 2011. Automated immunomagnetic processing and separation of *Legionella pneumophila* with manual detection by sandwich ELISA and PCR amplification of the ompS gene. J Lab Autom 16(2):157-64.

Reimer AR, Au S, Schindle S, Bernard KA. 2010. *Legionella pneumophila* monoclonal antibody subgroups and DNA sequence types isolated in Canada between 1981 and 2009: Laboratory Component of National Surveillance. Eur. J. Clin. Microbiol. Infect. Dis. 29(2):191-205.

Reuter S, Harrison TG, Koser CU, Ellington MJ, Smith GP, Parkhill J, Peacock SJ, Bentley SD, Torok ME. 2013. A pilot study of rapid whole-genome sequencing for the investigation of a *Legionella* outbreak. BMJ Open 3(1).

Robert Koch-Institut. 2016. Infektionsepidemiologisches Jahrbuch meldepflichtiger Krankheiten für 2015.: Robert-Koch-Institut.

Robertson P, Abdelhady H, Garduno RA. 2014. The many forms of a pleomorphic bacterial pathogen-the developmental network of *Legionella pneumophila*. Front. Microbiol. 5:670.

Rota MC, Fontana S, Montano-Remacha C, Scaturro M, Caporali MG, Vullo V, Scorzolini L, Ercole A, Ricci ML. 2014. Legionnaires' disease pseudoepidemic due to falsely positive urine antigen test results. J. Clin. Microbiol. 52(6):2279-2280.

Rowbotham TJ. 1980. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. J. Clin. Pathol. 33(12):1179-1183.

Rowbotham TJ. 1998. Isolation of *Legionella pneumophila* serogroup 1 from human feces with use of amebic cocultures. Clin. Infect. Dis. 26(2):502-503.

Ruettger A, Nieter J, Skrypnyk A, Engelmann I, Ziegler A, Moser I, Monecke S, Ehricht R, Sachse K. 2012. Rapid spoligotyping of *Mycobacterium tuberculosis* complex bacteria by use of a microarray system with automatic data processing and assignment. J. Clin. Microbiol. 50(7):2492-5.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16(10):944-5.

Sabat AJ, Budimir A, Nashev D, Sá-Leao R, van Dijl JM, Laurent F, Grundmann H, Friedrich AW. 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill. 18(4):20380.

Sahr T, Rusniok C, Dervins-Ravault D, Sismeiro O, Coppee JY, Buchrieser C. 2012. Deep sequencing defines the transcriptional map of *L. pneumophila* and identifies growth phase-dependent regulated ncRNAs implicated in virulence. RNA Biol. 9(4):503-19.

Sánchez-Busó L, Comas I, Jorques G, Gonzalez-Candelas F. 2014. Recombination drives genome evolution in outbreak-related *Legionella pneumophila* isolates. Nat. Genet. 46(11):1205-1211.

Sánchez-Busó L, Guiral S, Crespi S, Moya V, Camaro ML, Olmos MP, Adrian F, Morera V, Gonzalez-Moran F, Vanaclocha H, Gonzalez-Candelas F. 2016. Genomic Investigation of a Legionellosis Outbreak in a Persistently Colonized Hotel. Front. Microbiol. 6:1556.

Sanford JP. 1979. Legionnaires' disease: one person's perspective. Ann. Intern. Med. 90(4):699-703.

Scaturro M, Fontana S, Crippa S, Caporali MG, Seyler T, Veschetti E, Villa G, Rota MC, Ricci ML. 2015. An unusually long-lasting outbreak of community-acquired Legionnaires' disease, 2005-2008, Italy. Epidemiol.Infect 143(11):2416-2425.

Scaturro M, Fontana S, Dell'eva I, Helfer F, Marchio M, Stefanetti MV, Cavallaro M, Miglietta M, Montagna MT, De Giglio O, Cuna T, Chetti L, Sabattini MA, Carlotti M, Viggiani M, Stenico A, Romanin E, Bonanni E, Ottaviano C, Franzin L, Avanzini C, Demarie V, Corbella M, Cambieri P, Marone P, Rota MC, Bella A, Ricci ML. 2016. A multicenter study of viable PCR using propidium monoazide to detect *Legionella* in water samples. Diagn. Microbiol. Infect. Dis. 85(3):283-8.

Schjørring S, Stegger M, Kjelso C, Lilje B, Bangsborg JM, Petersen RF, David S, Uldum SA, Infections ESGfL. 2017. Genomic investigation of a suspected outbreak of *Legionella pneumophila* ST82 reveals undetected heterogeneity by the present gold-standard methods, Denmark, July to November 2014. Euro Surveill. 22(25).

Schmoock G, Ehricht R, Melzer F, Elschner M, Tomaso H, Neubauer H, Al Dahouk S. 2011. Development of a diagnostic multiplex polymerase chain reaction microarray assay to detect and differentiate *Brucella* spp. Diagn. Microbiol. Infect. Dis. 71(4):341-53.

Schnee C, Schulsse S, Hotzel H, Ayling RD, Nicholas RA, Schubert E, Heller M, Ehricht R, Sachse K. 2012. A novel rapid DNA microarray assay enables identification of 37 *Mycoplasma* species and highlights multiple *Mycoplasma* infections. PLoS One 7(3):e33237.

Schnetzinger F, Pan Y, Nocker A. 2013. Use of propidium monoazide and increased amplicon length reduce false-positive signals in quantitative PCR for bioburden analysis. Appl. Microbiol. Biotechnol. 97(5):2153-62.

Schoenhofen IC, McNally DJ, Vinogradov E, Whitfield D, Young NM, Dick S, Wakarchuk WW, Brisson JR, Logan SM. 2006. Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*: enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. J. Biol. Chem. 281(2):723-32.

Schroeder GN, Petty NK, Mousnier A, Harding CR, Vogrin AJ, Wee B, Fry NK, Harrison TG, Newton HJ, Thomson NR, Beatson SA, Dougan G, Hartland EL, Frankel G. 2010. *Legionella pneumophila* strain 130b possesses a unique combination of type IV secretion systems and novel Dot/Icm secretion system effector proteins. J. Bacteriol. 192(22):6001-6016.

Slaon-Gardner T, Stirzaker S, Knuckey D, Pennington K, Knope K, Fitzsimmons G, Martin N, Siripol S, Richards K, Witheridge J, Wright P, Barry C, Ormond J, Group NARW. 2011. Australia's notifiable disease status, 2009: annual report of the National Notifiable Diseases Surveillance System. Commun Dis Intell Q Rep 35(2):61-131.

Sobral D, Le Cann P, Gerard A, Jarraud S, Lebeau B, Loisy-Hamon F, Vergnaud G, Pourcel C. 2011. High-throughput typing method to identify a non-outbreak-involved *Legionella pneumophila* strain colonizing the entire water supply system in the town of Rennes, France. Appl. Environ. Microbiol. 77(19):6899-907.

Spratt BG. 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. Curr. Opin. Microbiol. 2(3):312-6.

Steele TW, Lanser J, Sangster N. 1990. Isolation of *Legionella longbeachae* serogroup 1 from potting mixes. Appl. Environ. Microbiol. 56(1):49-53.

Stewart CR, Muthye V, Cianciotto NP. 2012. *Legionella pneumophila* persists within biofilms formed by *Klebsiella pneumoniae*, *Flavobacterium* sp., and *Pseudomonas fluorescens* under dynamic flow conditions. PLoS One 7(11):e50560.

Stone BJ, Kwaik YA. 1999. Natural competence for DNA transformation by *Legionella pneumophila* and its association with expression of type IV pili. J. Bacteriol. 181(5):1395-1402.

The ESCMID Study for *Legionella* Infections. 2015. The ESGLI SBT database for the typing of *Legionella pneumophila*. Afshar B. [upd: 11.08.2017.] Acc: 11.08.2017. http://www.hpa-bioinformatics.org.uk/legionella/legionella_sbt/php/sbt_homepage.php.

Thürmer A, Helbig JH, Jacobs E, Lück PC. 2009. PCR-based serotyping of *Legionella pneumophila*. J. Med. Microbiol. 58(Pt 5):588-595.

Tijet N, Tang P, Romilowych M, Duncan C, Ng V, Fisman DN, Jamieson F, Low DE, Guyard C. 2010. New endemic *Legionella pneumophila* serogroup I clones, Ontario, Canada. Emerg. Infect. Dis. 16(3):447-454.

Tison DL, Seidler RJ. 1983. *Legionella* incidence and density in potable drinking water supplies. Appl. Environ. Microbiol. 45(1):337-339.

Torres A, Blasi F, Peetermans WE, Viegi G, Welte T. 2014. The aetiology and antibiotic management of community-acquired pneumonia in adults in Europe: a literature review. Eur. J. Clin. Microbiol. Infect. Dis. 33(7):1065-1079.

Underwood AP, Bellamy W, Afshar B, Fry NK, Harrison TG. 2006. Chapter 44: Development of an online tool for European working group for *Legionella* infections sequence based typing, including automatic quality assessment and data submission. In: Cianciotto NP, Abu-Kwaik Y, Edelstein PH, Fields BS, Geary DF, Harrison TG, Joseph CA, Ratcliff RM, Stout JE, Swanson MS, editors. Legionella. Washington DC: ASM Press. p 163-166.

Underwood AP, Jones G, Mentasti M, Fry NK, Harrison TG. 2013. Comparison of the *Legionella pneumophila* population structure as determined by sequence-based typing and whole genome sequencing. BMC Microbiol. 13:302.

Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, Rouy Z, Roche D, Salvignol G, Scarpelli C, Medigue C. 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. Database (Oxford) 2009:bap021.

Valsangiacomo C, Baggi F, Gaia V, Balmelli T, Peduzzi R, Piffaretti JC. 1995. Use of amplified fragment length polymorphism in molecular typing of *Legionella pneumophila* and application to epidemiological studies. J. Clin. Microbiol. 33(7):1716-1719.

Van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S, Struelens M. 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin. Microbiol. Infect. 13:1-46.

van Heijnsbergen E, Schalk JA, Euser SM, Brandsema PS, Den Boer JW, de Roda Husman AM. 2015. Confirmed and potential sources of legionella reviewed. Environ. Sci. Technol. 49(8):4797-4815.

VDI. 2015. VDI 2047-2: Securing hygienically sound operation of evaporative cooling systems (VDI Cooling Tower Code of Practice): Beuth Verlag, Berlin, 2015.

Vergnes M, Ginevra C, Kay E, Normand P, Thioulouse J, Jarraud S, Maurin M, Schneider D. 2011. Insertion sequences as highly resolutive genomic markers for sequence type 1 *Legionella pneumophila* Paris. J. Clin. Microbiol. 49(1):315-324.

Villari P, Motti E, Farullo C, Torre I. 1998. Comparison of conventional culture and PCR methods for the detection of *Legionella pneumophila* in water. Lett. Appl. Microbiol. 27(2):106-110.

Viswanathan VK, Edelstein PH, Pope CD, Cianciotto NP. 2000. The *Legionella pneumophila iraAB* locus is required for iron assimilation, intracellular infection, and virulence. Infect. Immun. 68(3):1069-1079.

von Baum H, Ewig S, Marre R, Suttorp N, Gonschior S, Welte T, Lück C, Competence Network for Community Acquired Pneumonia Study G. 2008. Community-acquired *Legionella* pneumonia: new insights from the German competence network for community acquired pneumonia. Clin. Infect. Dis. 46(9):1356-1364.

von Baum H, Härter G, Essig A, Lück C, Gonser T, Embacher A, Brockmann S. 2010. Preliminary report: outbreak of Legionnaires' disease in the cities of Ulm and Neu-Ulm in Germany, December 2009 - January 2010. Euro Surveill. 15(4):19472.

Wagner C, Krönert C, Lück PC, Jacobs E, Cianciotto NP, Helbig JH. 2007. Random mutagenesis of *Legionella pneumophila* reveals genes associated with lipopolysaccharide synthesis and recognition by typing monoclonal antibodies. J. Appl. Microbiol. 103(5):1975-1982.

Walser SM, Gerstner DG, Brenner B, Höller C, Liebl B, Herr CE. 2014. Assessing the environmental health relevance of cooling towers-a systematic review of legionellosis outbreaks. Int. J. Hyg. Environ. Health 217(2-3):145-154.

Weist K, Brandt C, Lück PC, Wagner J, Eckmanns T, Rüden H. 2006. 10 years of Legionella surveillance: Change of *Legionella* subtype preceded epidemic of nosocomial Legionnaire's disease. In: Cianciotto NP, Abu-Kwaik Y, Edelstein PH, Fields BS, Geary DF, Harrison TG, Joseph CA, Ratcliff RM, Stout JE, Swanson MS, editors. *Legionella*: ASM Press, Washington DC. p 128-131.

Wewalka G, Schmid D, Harrison TG, Uldum SA, Lück C. 2014. Dual infections with different *Legionella* strains. Clin. Microbiol. Infect. 20(1):O13-O19.

World Health Organization. 2007. *Legionella* and the prevention of legionellosis. Geneva: World Health Organization.

Wunderlich A, Torggler C, Elsasser D, Lück C, Niessner R, Seidel M. 2016. Rapid quantification method for *Legionella pneumophila* in surface water. Anal. Bioanal. Chem. 408(9):2203-13.

Yzerman E, den Boer JW, Caspers M, Almal A, Worzel B, van der Meer W, Montijn R, Schuren F. 2010. Comparative genome analysis of a large Dutch *Legionella pneumophila* strain collection identifies five markers highly correlated with clinical strains. BMC Genomics 11:433.

Zähringer U, Knirel YA, Lindner B, Helbig JH, Sonesson A, Marre R, Rietschel ET. 1995. The lipopolysaccharide of *Legionella pneumophila* serogroup 1 (strain Philadelphia 1): chemical structure and biological significance. Prog. Clin. Biol. Res. 392:113-139.

Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18(5):821-829.

Zou CH, Knirel YA, Helbig JH, Zähringer U, Mintz CS. 1999. Molecular cloning and characterization of a locus responsible for O acetylation of the O polysaccharide of *Legionella pneumophila* serogroup 1 lipopolysaccharide. J. Bacteriol. 181(13):4137-4141.

Statement of Authorship

Technische Universität Dresden
Medizinische Fakultät Carl Gustav Carus
Promotionsordnung 24.07.2011

Erklärung zur Eröffnung des Promotionsverfahrens

1. Hiermit versichere ich, dass ich die vorliegende Arbeit ohne zulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

2. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Unterstützung von folgenden Personen erhalten: *Dr. med. Christian Lück. Einzelne Abschnitte wurden mit den genannten Co-Autoren der jeweiligen Publikationen gemeinsam bearbeitet. Dies ist angegeben bei den jeweiligen Publikationen angegeben.*

3. Weitere Personen waren an der geistigen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mit weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

4. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

5. Die Inhalte der Dissertation wurden in folgender Form veröffentlicht: *Die Veröffentlichungen in Form von Fachartikeln wurden deutlich in der Arbeit angegeben. Des Weiteren sind Poster- oder Präsentationsbeiträge auf nationalen oder internationalen Tagungen gehalten worden (siehe Publikationsliste).*

6. Ich bestätige, dass es keine zurückliegenden erfolglosen Promotionsverfahren gab.

7. Ich bestätige, dass ich die Promotionsordnung der Medizinischen Fakultät der Technischen Universität Dresden anerkenne.

8. Ich habe die Zitierrichtlinien für Dissertationen an der Medizinischen Fakultät der Technischen Universität Dresden zur Kenntnis genommen und befolgt.

Dresden, 01.11.2017

Markus Petzold

Declaration of Compliance

Technische Universität Dresden
Medizinische Fakultät Carl Gustav Carus
Promotionsordnung 24.10.2014

Hiermit bestätige ich die Einhaltung der folgenden aktuellen gesetzlichen Vorgaben im Rahmen meiner Dissertation

□ Das zustimmende Votum der Ethikkommission bei klinischen Studien, epidemiologischen Untersuchungen mit Personenbezug oder Sachverhalten, die das Medizinproduktegesetzt betreffen
Aktenzeichen der zuständigen Ethikkommission…….

□ Die Einhaltung der Bestimmung des Tierschutzgesetzes
Aktenzeichen der Genehmigungsbehörde zum Vorhaben/zur Mitwirkung…….....

□ Die Einhaltung des Gentechnikgesetzes
Projektnummer……

■ Die Einhaltung der Datenschutzbestimmung der Medizinischen Fakultät und des Universitätsklinikums Carl Gustav Carus

Dresden, 01.11.2017

Markus Petzold