JCM
Journals.ASM.org

Whole-Genome Mapping as a Novel High-Resolution Typing Tool for *Legionella pneumophila*

Thijs Bosch,[a] Sjoerd M. Euser,[b] Fabian Landman,[a] Jacob P. Bruin,[b] Ed P. IJzerman,[b] Jeroen W. den Boer,[b] Leo M. Schouls[a]

National Institute for Public Health and the Environment, Bilthoven, the Netherlands[a]; Regional Public Health Laboratory Kennemerland, Haarlem, the Netherlands[b]

***Legionella* is the causative agent for Legionnaires' disease (LD) and is responsible for several large outbreaks in the world. More than 90% of LD cases are caused by *Legionella pneumophila*, and studies on the origin and transmission routes of this pathogen rely on adequate molecular characterization of isolates. Current typing of *L. pneumophila* mainly depends on sequence-based typing (SBT). However, studies have shown that in some outbreak situations, SBT does not have sufficient discriminatory power to distinguish between related and nonrelated *L. pneumophila* isolates. In this study, we used a novel high-resolution typing technique, called whole-genome mapping (WGM), to differentiate between epidemiologically related and nonrelated *L. pneumophila* isolates. Assessment of the method by various validation experiments showed highly reproducible results, and WGM was able to confirm two well-documented Dutch *L. pneumophila* outbreaks. Comparison of whole-genome maps of the two outbreaks together with WGMs of epidemiologically nonrelated *L. pneumophila* isolates showed major differences between the maps, and WGM yielded a higher discriminatory power than SBT. In conclusion, WGM can be a valuable alternative to perform outbreak investigations of *L. pneumophila* in real time since the turnaround time from culture to comparison of the *L. pneumophila* maps is less than 24 h.**

Legionella is a rod-shaped, Gram-negative bacterial pathogen that is ubiquitous in aquatic reservoirs. It is the causative agent for Legionnaires' disease (LD), an acute pneumonia, characterized by clinical symptoms such as cough, fever, and radiological signs of infiltration that do not differ from pneumonia caused by other pathogens. LD is thought to account for 2% to 15% of all community-acquired pneumonias (1–3) and proves fatal in about 6% of cases (4).

Several large outbreaks of LD have been reported worldwide. Examples include Murcia, Spain (449 confirmed cases); Barrow-in-Furness, United Kingdom (179 confirmed cases); and Quebec City, Canada (182 confirmed cases). These outbreaks often involved contaminated cooling towers that can infect hundreds of people within a short time period, until the source of infection is detected and appropriate control measures are taken (5–7). In the source investigation of such outbreaks, epidemiological analyses together with genotypic comparisons between clinical and environmental isolates are essential (8, 9).

More than 90% of LD cases are caused by *Legionella pneumophila*, and as this species is commonly found in the environment, adequate typing methods are needed to differentiate between isolates in order to confirm or reject a potential source of infection (10). Sequence-based typing (SBT), a variant of the classic multilocus sequence typing schemes (11), is an internationally recognized procedure for genotyping *L. pneumophila* isolates. It is a rapid, discriminatory, and reproducible seven-gene molecular typing method that is recommended as the method of choice by the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) study group for *Legionella* infections (ESGLI) (12).

However, recent studies have described situations in which SBT did not provide the discriminatory level that was needed to distinguish between outbreak-related *L. pneumophila* isolates and nonoutbreak isolates (13). Exploring novel techniques such as next-generation sequencing (NGS) has shown that NGS can provide comparable, if not better, discriminatory power within *L. pneumophila* isolates; however, it is still too laborious and time-consuming to implement in routine surveillance (10, 13–15).

Another novel molecular analysis method called whole-genome mapping (WGM) has recently been used as a high-resolution typing technique (16). WGM uses high-resolution, ordered, whole-genome restriction maps for comparative analysis. In a recent study, WGM successfully discriminated between isolates belonging to livestock-associated methicillin-resistant *Staphylococcus aureus* (LA-MRSA) and identified the transmission and persistence of this genetically homogeneous MRSA variant, where current typing techniques failed (17). Although WGM has been very successful for LA-MRSA, the number of reports in which WGM was applied for molecular typing of other bacterial pathogens is very limited (18, 19).

In this study, the capability of whole-genome mapping to differentiate *L. pneumophila* isolates was investigated using epidemiologically related and nonrelated *L. pneumophila* isolates.

MATERIALS AND METHODS

Strain selection and study design. We selected 53 *L. pneumophila* isolates to generate 57 different whole-genome maps (WGMs). For validation, two environmental *L. pneumophila* strains B9006 and D9010 were used for reproducibility and stability experiments. *L. pneumophila* strains

Received 21 May 2015 Returned for modification 24 June 2015
Accepted 18 July 2015
Accepted manuscript posted online 22 July 2015
Citation Bosch T, Euser SM, Landman F, Bruin JP, IJzerman EP, den Boer JW, Schouls LM. 2015. Whole-genome mapping as a novel high-resolution typing tool for *Legionella pneumophila*. J Clin Microbiol 53:3234–3238. doi:10.1128/JCM.01369-15.
Editor: D. J. Diekema
Address correspondence to Thijs Bosch, thijs.bosch@rivm.nl.

doi:10.1128/JCM.01369-15

TABLE 1 Bacterial strains used in this study

Experiment	Strains(s) (no. of isolates)	Sequence type(s) (no. of isolates)	WGMs created
Reproducibility of WGM	B9006	477	3
Stability of WGM	D9010	47	2
Comparison with *in silico* maps	A9001	45	1
	D9010	47	1
Legionella transmission events	Bovenkarspel (28)	62	28
	Amsterdam (7)	42	7
Discriminatory power of WGM	Unrelated Legionella isolates (15)	23, 37, 47 (8), 62, 444, 485, 493, 524	15
Total	53		57

A9001(derived from a patient) and D9010 were selected for comparison of whole-genome maps created in our laboratory with *in silico* maps based on their whole-genome sequences obtained using a hybrid assembly approach of Illumina and PacBio sequencing (BaseClear, Leiden, The Netherlands). The discriminatory power of whole-genome mapping for *L. pneumophila* and the capability to identify transmission events was studied using 35 isolates obtained during two well-documented outbreaks in the Netherlands of *L. pneumophila* (20, 21) and 15 epidemiologically unrelated isolates originating from different patients in the Netherlands (Table 1). These 15 isolates, as well as the three strains used for reproducibility and stability experiments (A9001, B9006, and D9010) were collected within the Dutch National *Legionella* Outbreak Detection Program between 2002 and 2012.

All isolates used in this study came from preexisting collections, and the isolates used to create WGMs were also characterized by SBT (12, 22).

Whole-genome mapping of *L. pneumophila*. The input of high-molecular-weight (HMW) DNA is required for whole-genome mapping. The isolation of HMW DNA was performed using the same protocol as for *S. aureus* with two modifications. First, no lysostaphine was added during the spheroplasting step, and second, the incubation time of the spheroplasting step was reduced to 1 h. The creation of whole-genome maps and the analyses of WGMs were performed as described previously (16). Based on the previous validation of WGM for *S. aureus*, a combination of a filtering setting that excluded fragments of <3,000 bp from the comparison and a relative tolerance of 15% and an absolute tolerance of 1,000 bp was used to compensate for the variation in sizing and the presence or absence of smaller fragments during the clustering and alignment of *Legionella* WGMs.

RESULTS

Assessing the optimal settings for whole-genome mapping of *L. pneumophila*. To assess the optimal settings for WGM of *L. pneumophila*, a series of validation experiments were conducted. First, the DNA sample obtained from *L. pneumophila* strain B9006 was used to create WGMs on three consecutive days. The obtained WGMs showed >99.5% similarity between WGMs created from the same DNA sample, while distinct maps were obtained from unrelated isolates under these conditions.

Second, the temporal stability of *L. pneumophila* genomes under laboratory conditions was determined by subculturing *L. pneumophila* strain D9010 for 30 days, and WGMs were created from DNA isolated from the day 1 and day 30 cultures. The resulting WGMs were indistinguishable, with a similarity of 99.3% between the maps.

Besides reproducibility, a comparative analysis between WGMs from *L. pneumophila* isolates A9001 and D9010 created in our laboratory and their *in silico* counterparts generated in the BioNumerics software was performed. The *in silico* maps, based on the obtained whole-genome sequences, and the WGMs obtained in the laboratory showed high similarities of 99.2% for A9001 and 98.4% for D9010.

Whole-genome mapping confirms two well-known *L. pneumophila* outbreaks. To assess whether WGM is capable of identifying outbreaks of *L. pneumophila*, isolates of two well-known *L. pneumophila* outbreaks from the Netherlands were subjected to WGM.

The first set comprised seven isolates from an outbreak in the Amsterdam region (21). Six isolates were cultured from patients while a single isolate was obtained from the identified source of the outbreak (a cooling tower). All isolates belonged to serogroup 1, and characterization with SBT revealed they were of sequence type 42 (ST42). The WGMs of the isolates were indistinguishable, showing a similarity of >99.9% between the most distinct maps (Fig. 1).

The second outbreak comprised 28 isolates obtained from the large *L. pneumophila* outbreak at the Westfriese flower show in Bovenkarspel (20), of which 23 isolates originated from patients and 5 isolates came from different sampling points from the likely source, a whirlpool. Previous characterization showed that all isolates yielded serogroup 1 and ST62. The WGMs of the 23 patient isolates were indistinguishable, with a similarity between the maps of 99.4%. All WGMs of isolates sampled from the likely source, a whirlpool, were indistinguishable from the WGMs of the isolates from patients.

A comparison of the two outbreaks revealed major differences between the whole-genome maps, and the similarity between the WGMs of the isolates from the two outbreaks was only 56.5%. Based on reproducibility experiments, which yielded >98% similar profiles, and the result of the above-described outbreaks, we chose to set the cutoff value at 98% for indistinguishable profiles, while isolates with similarities between 95% and 98% were considered highly related.

Discriminatory power of WGM for *L. pneumophila*. To determine the discriminatory power of WGM, maps were created of *L. pneumophila* isolates originating from 15 epidemiologically unrelated patients. Previous characterization with SBT revealed eight different types, and eight isolates had the same ST, namely, ST47. Comparative analyses of the whole-genome maps of the unrelated isolates together with the WGMs from the two outbreaks yielded very distinctive maps with a similarity between the maps ranging from 43% to 90% (Fig. 2). Four clusters were found based on a similarity cutoff of 98% for indistinguishable WGMs. The first cluster (cl-01) contained all WGMs from the Amsterdam outbreak, while cl-02 consisted of 28 of the 29 Bovenkarspel isolates and a single epidemiologically unrelated isolate. Although this

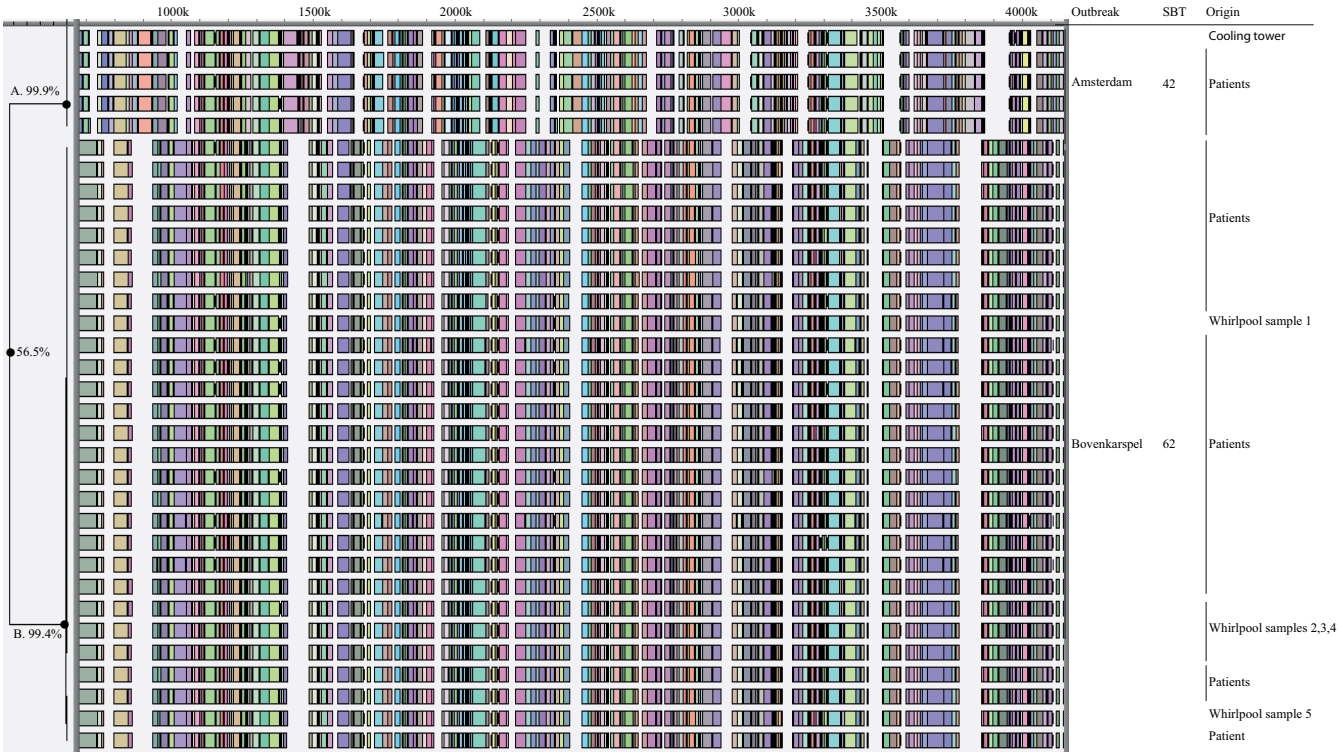

FIG 1 Whole-genome mapping of *L. pneumophila* of two well-documented outbreaks. The complete WGMs are displayed. The dendrogram on the left denotes the clusters A and B and their similarities. On the right hand side, outbreak, SBT, and origin of the isolates are indicated.

unrelated isolate was separated from the Bovenkarspel isolates in the spanning tree, it had a similarity with the other maps of 98.8% and yielded the same SBT as the Bovenkarspel isolates. The other two clusters (cl-03 and cl-04) were comprised of seven unrelated isolates (all typed as ST47). One cluster (cl-03) comprised four maps that yielded a similarity between the maps of 99.7%, while the other cluster (cl-04) consisting of three maps and showed a similarity between the most distinct maps of 99.2%. The similarity between clusters cl-03 and cl-04 was only 43% despite the fact that all isolates in these groups yielded ST47. The isolates that were not part of a cluster belonged to eight epidemiologically unrelated isolates and to a single isolate of the Bovenkarspel outbreak.

DISCUSSION

In this study, we used whole-genome mapping (WGM) as a high-resolution typing method for *L. pneumophila*. We were able to reveal a considerable degree of variation among different *L. pneumophila* strains, and WGM seems to be a useful typing tool to identify outbreaks.

WGM was able to confirm the two *L. pneumophila* outbreaks and link the likely source of each outbreak with isolates obtained from patients (20, 21).

The discriminatory power of WGM for *L. pneumophila* was best illustrated by the high degree of variation between the WGMs of two well-documented Dutch *L. pneumophila* outbreaks and the epidemiologically unrelated isolates, where the most distinct maps only showed a similarity of 43%. In comparison, WGM of LA-MRSA in a recent study still revealed an 84% similarity between the most distinct LA-MRSA isolates (17).

To assess whether WGM is capable of identifying *L. pneumophila* transmission events, maps of epidemiologically unrelated isolates were compared to maps obtained from outbreaks. Analysis showed that the two outbreaks form different clusters, while most unrelated isolates cluster as singletons. However, in the cluster containing the Bovenkarspel isolates, an additional nonrelated isolate yielding the same ST62 was present. Although the map of this isolate showed some differences with the Bovenkarspel isolates, it still groups within this cluster based on our criterion for indistinguishable isolates. Whether this strain represents the same strain as obtained during the outbreak remains unclear. However, ST62 is (after ST47) the second most frequently found ST among clinical *L. pneumophila* serogroup 1 isolates in the Netherlands, which makes it more likely to find an epidemiologically nonrelated ST62 strain with a corresponding map compared to those of strains with uncommon STs. Another group of isolates for which we were not able to make a clear distinction with WGM was the group of isolates yielding ST47, resulting in two additional clusters. It may be that these two clusters contain isolates representing the same strain, with a previously unknown epidemiological link, but it is more likely that WGM has difficulties separating strains within this ST. However, WGM was able to split these strains into two distinct clusters with very limited similarity between the clusters, indicating that WGM has a higher discriminatory power and is more capable than SBT to assess whether isolates belong to a transmission event.

The criterion for indistinguishable WGMs of ≥98% was based on replicates of *L. pneumophila* strains created in this study. In addition, the possible effect of mobile genetic elements made us determine that maps with similarities between 95% and 98% may represent the same strain, while maps with a similarities of <95%

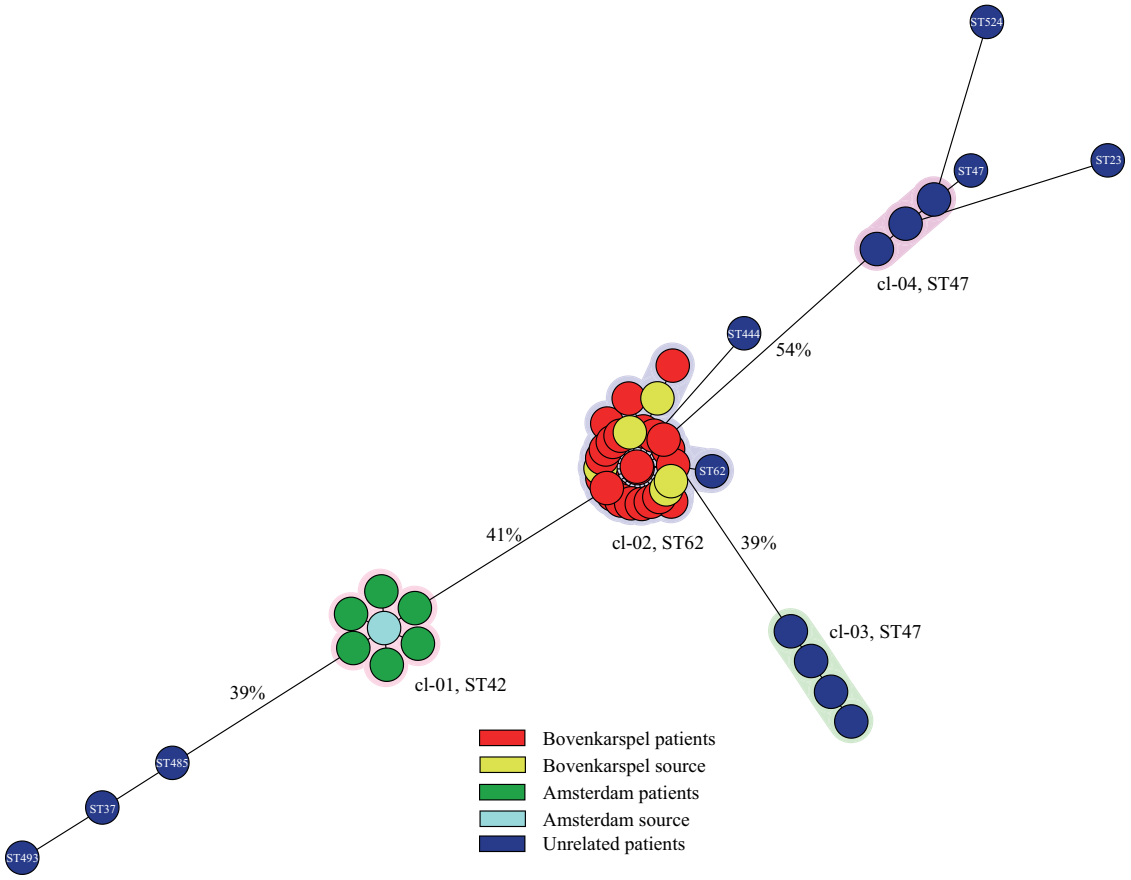

FIG 2 Minimum spanning tree depicting the genotypic diversity of *L. pneumophila* isolates ($n = 50$). Each node represents the WGM of a single *L. pneumophila* isolate. The halos represent clusters based on a similarity cutoff of ≥98%.

are considered different strains. The same cutoff values were previously established for LA-MRSA and were also used in two other studies using WGM for MRSA with the USA300 genotype and *Pseudomonas* (16, 18, 23). It therefore seems that these cutoff values can be utilized for WGM in general regardless of the microorganism used.

Current molecular characterization of *L. pneumophila* isolates is mainly based on SBT (12). Based on this study, WGM can be a suitable high-resolution alternative to type *L. pneumophila* isolates and perform transmission investigations due to its higher discriminatory power. However, with current developments regarding next-generation sequencing (NGS), we believe it is only a matter of time before NGS will become the ultimate typing tool for *L. pneumophila* and virtually all microorganisms. Yet, for the time being, WGM can be a valuable alternative to perform outbreak investigations in real time since the turnaround time from culture to comparison of the *L. pneumophila* maps is less than 24 h.

REFERENCES

1. **Braun JJ, de Graaff CS, de Goey J, Zwinderman AH, Petit PL.** 2004. Community-acquired pneumonia: pathogens and course in patients admitted to a general hospital. Ned Tijdschr Geneeskd **148:**836–840. (In Dutch.)
2. **Fields BS, Benson RF, Besser RE.** 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. Clin Microbiol Rev **15:**506–526. http://dx.doi.org/10.1128/CMR.15.3.506-526.2002.
3. **Sopena N, Sabria M, Pedro-Botet ML, Manterola JM, Matas L, Dominguez J, Modol JM, Tudela P, Ausina V, Foz M.** 1999. Prospective study of community-acquired pneumonia of bacterial etiology in adults. Eur J Clin Microbiol Infect Dis **18:**852–858. http://dx.doi.org/10.1007/s100960050419.
4. **Joseph CA, Ricketts KD, European Working Group for *Legionella* Infections.** 2010. Legionnaires disease in Europe 2007–2008. Euro Surveill **15:**19493.
5. **Bennett E, Ashton M, Calvert N, Chaloner J, Cheesbrough J, Egan J, Farrell I, Hall I, Harrison TG, Naik FC, Partridge S, Syed Q, Gent RN.** 2014. Barrow-in-Furness: a large community legionellosis outbreak in the UK. Epidemiol Infect **142:**1763–1777. http://dx.doi.org/10.1017/S0950268813002483.
6. **Garcia-Fulgueiras A, Navarro C, Fenoll D, Garcia J, Gonzalez-Diego P, Jimenez-Bunuales T, Rodriguez M, Lopez R, Pacheco F, Ruiz J, Segovia M, Balandron B, Pelaz C.** 2003. Legionnaires' disease outbreak in Murcia, Spain. Emerg Infect Dis **9:**915–921. http://dx.doi.org/10.3201/eid0908.030337.
7. **Levesque 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:**e103852. http://dx.doi.org/10.1371/journal.pone.0103852.
8. **Chiarini A, Bonura C, Ferraro D, Barbaro R, Cala C, Distefano S, Casuccio N, Belfiore S, Giammanco A.** 2008. Genotyping of *Legionella pneumophila* serogroup 1 strains isolated in Northern Sicily, Italy. New Microbiol **31:**217–228.
9. **Fry NK, Bangsborg JM, Bernander S, Etienne J, Forsblom B, Gaia V, Hasenberger P, Lindsay D, Papoutsi A, Pelaz C, Struelens M, Uldum SA, Visca P, Harrison TG.** 2000. Assessment of intercentre reproducibility and epidemiological concordance of *Legionella pneumophila* serogroup

1 genotyping by amplified fragment length polymorphism analysis. Eur J Clin Microbiol Infect Dis **19:**773–780. http://dx.doi.org/10.1007/s100960000359.

10. **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. http://dx.doi.org/10.1186/1471-2180-13-302.
11. **Brehony C, Jolley KA, Maiden MC.** 2007. Multilocus sequence typing for global surveillance of meningococcal disease. FEMS Microbiol Rev **31:**15–26. http://dx.doi.org/10.1111/j.1574-6976.2006.00056.x.
12. **Gaia V, Fry NK, Afshar B, Luck 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:**2047–2052. http://dx.doi.org/10.1128/JCM.43.5.2047-2052.2005.
13. **Graham RM, Doyle CJ, Jennison AV.** 2014. Real-time investigation of a *Legionella pneumophila* outbreak using whole genome sequencing. Epidemiol Infect **142:**2347–2351. http://dx.doi.org/10.1017/S0950268814000375.
14. **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:**e002175.
15. **Sabat AJ, Budimir A, Nashev D, Sa-Leao R, van Dijl J, Laurent F, Grundmann H, Friedrich AW, ESCMID Study Group of Epidemiological Markers (ESGEM).** 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill **18:**20380.
16. **Bosch T, Verkade E, van Luit M, Pot B, Vauterin P, Burggrave R, Savelkoul P, Kluytmans J, Schouls L.** 2013. High resolution typing by whole genome mapping enables discrimination of LA-MRSA (CC398) strains and identification of transmission events. PLoS One **8:**e66493. http://dx.doi.org/10.1371/journal.pone.0066493.
17. **Bosch T, Verkade E, van Luit M, Landman F, Kluytmans J, Schouls LM.** 2015. Transmission and persistence of livestock-associated methicillin-resistant *Staphylococcus aureus* among veterinarians and their household members. Appl Environ Microbiol **81:**124–129. http://dx.doi.org/10.1128/AEM.02803-14.
18. **Boers SA, Burggrave R, van Westreenen M, Goessens WH, Hays JP.** 2014. Whole-genome mapping for high-resolution genotyping of *Pseudomonas aeruginosa*. J Microbiol Methods **106:**19–22. http://dx.doi.org/10.1016/j.mimet.2014.07.020.
19. **Kotewicz ML, Jackson SA, LeClerc JE, Cebula TA.** 2007. Optical maps distinguish individual strains of *Escherichia coli* O157: H7. Microbiology **153:**1720–1733. http://dx.doi.org/10.1099/mic.0.2006/004507-0.
20. **Den Boer JW, Yzerman EP, Schellekens J, Lettinga KD, Boshuizen HC, Van Steenbergen JE, Bosman A, Van den Hof S, Van Vliet HA, Peeters MF, Van Ketel RJ, Speelman P, Kool JL, Conyn-Van Spaendonck MA.** 2002. A large outbreak of Legionnaires' disease at a flower show, the Netherlands, 1999. Emerg Infect Dis **8:**37–43. http://dx.doi.org/10.3201/eid0801.010176.
21. **Sonder GJ, van den Hoek JA, Bovee LP, Aanhane FE, Worp J, Du Ry van Beest Holle M, van Steenbergen JE, den Boer JW, Ijzerman EP, Coutinho RA.** 2008. Changes in prevention and outbreak management of Legionnaires disease in the Netherlands between two large outbreaks in 1999 and 2006. Euro Surveill **13:**pii=18983.
22. **Ratzow S, Gaia V, Helbig JH, Fry NK, Luck 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:**1965–1968. http://dx.doi.org/10.1128/JCM.00261-07.
23. **Shukla SK, Pantrangi M, Stahl B, Briska AM, Stemper ME, Wagner TK, Zentz EB, Callister SM, Lovrich SD, Henkhaus JK, Dykes CW.** 2012. Comparative whole-genome mapping to determine *Staphylococcus aureus* genome size, virulence motifs, and clonality. J Clin Microbiol **50:**3526–3533. http://dx.doi.org/10.1128/JCM.01168-12.