Journal of Biotechnology 235 (2016) 181–186

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Whole genome sequence-based serogrouping of *Listeria monocytogenes* isolates

Patrick Hyden[a], Ariane Pietzka[b,*], Anna Lennkh[b], Andrea Murer[b], Burkhard Springer[b], Marion Blaschitz[b], Alexander Indra[b], Steliana Huhulescu[b], Franz Allerberger[b], Werner Ruppitsch[b,c], Christoph W. Sensen[a]

[a] *Institute of Molecular Biotechnology, Graz University of Technology, Graz, Austria*
[b] *Institute of Medical Microbiology and Hygiene, Austrian Agency for Health and Food Safety, Graz, Austria*
[c] *Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria*

ARTICLE INFO

Article history:
Received 16 February 2016
Received in revised form 3 June 2016
Accepted 7 June 2016
Available online 8 June 2016

Keywords:
Listeria monocytogenes
Serotyping
Serogrouping
Serogroup prediction
Whole genome sequencing
Core genome MLST

ABSTRACT

Whole genome sequencing (WGS) is currently becoming the method of choice for characterization of *Listeria monocytogenes* isolates in national reference laboratories (NRLs). WGS is superior with regards to accuracy, resolution and analysis speed in comparison to several other methods including serotyping, PCR, pulsed field gel electrophoresis (PFGE), multilocus sequence typing (MLST), multilocus variable number tandem repeat analysis (MLVA), and multivirulence-locus sequence typing (MVLST), which have been used thus far for the characterization of bacterial isolates (and are still important tools in reference laboratories today) to control and prevent listeriosis, one of the major sources of foodborne diseases for humans. Backward compatibility of WGS to former methods can be maintained by extraction of the respective information from WGS data. Serotyping was the first subtyping method for *L. monocytogenes* capable of differentiating 12 serovars and national reference laboratories still perform serotyping and PCR-based serogrouping as a first level classification method for *Listeria monocytogenes* surveillance. Whole genome sequence based core genome MLST analysis of a *L. monocytogenes* collection comprising 172 isolates spanning all 12 serotypes was performed for serogroup determination. These isolates clustered according to their serotypes and it was possible to group them either into the IIa, IIc, IVb or IIb clusters, respectively, which were generated by minimum spanning tree (MST) and neighbor joining (NJ) tree data analysis, demonstrating the power of the new approach.

1. Introduction

Listeria monocytogenes is the causative agent of listeriosis, one of the major food-borne diseases affecting humans. It is a facultative intracellular pathogen of humans and animals and widespread in the environment (Pietzka et al., 2011; Allerberger et al., 2015). Listeriosis is characterized by symptoms like gastroenteritis, encephalitis, meningitis, and septicemia. Typically, pregnant woman, the elderly, and immuno-compromised people are affected. The high case-fatality rate of 20% to 30% makes *L. monocytogenes* a leading cause of food-borne related human mortality (Nyarko and Donnelly, 2015).

L. monocytogenes is a ubiquitous microorganism with the ability to survive in a variety of food sources (for example cheese), to grow at low temperatures (for example on cooled shelves in stores), to survive even freezing and high salt conditions and to withstand nitrite preservation methods, respectively. The ability to form biofilms on food contact surfaces (Allerberger and Wagner 2010; Jordan et al., 2008) can facilitate the persistence, dissemination and food contamination at several stages of food production (Allerberger et al., 2015). Nearly all sporadic and epidemic human listeriosis cases are linked to the consumption or use of contaminated food or feed (Schlech et al., 1983; Allerberger et al., 2015; Fretz et al., 2010).

During the investigation of a listeriosis outbreak, rapid and accurate subtyping methods are essential for the identification and subsequent elimination of the source of the contaminated food (Pichler et al., 2011). Serotyping of *L. monocytogenes* is based on somatic (O) and flagellar (H) antigens and was the first *Listeria monocytogenes* subtyping scheme. This approach was developed at

* Corresponding author at: Austrian Agency for Health and Food Safety, Institute of Medical Microbiology and Hygiene, National Reference Laboratory for Listeria, Beethovenstr. 6, A–8010 Graz, Austria.
E-mail address: ariane.pietzka@ages.at (A. Pietzka).

http://dx.doi.org/10.1016/j.jbiotec.2016.06.005

the end of the 1970's (Seeliger and Höhne 1979) and allows the differentiation of 12 serotypes. Phylogenetic investigations revealed that the species *L. monocytogenes* consists of four genetic lineages, Lineages I–IV, comprising specific serotypes (lineage I: serotypes 1/2b, 3b, 4b, 4d, 4e, and 7; lineage II: serotypes 1/2a, 1/2c, 3a, and 3c; lineage III: serotypes 4b, 1/2a, 4a and 4c; lineage IV: 4a, 4c serotypes (Haase et al., 2014). About 96% of all reported human listeriosis cases are caused by Lineages I and II (serotypes 4b, 1/2a, 1/2b) isolates (Kasper et al., 2009; Doumith et al., 2004; Seeliger and Höhne, 1979).

For practical and traditional purposes, serotyping is still the first level response subtyping method in public health laboratories, despite its limited value for tracking isolates (Nyarko and Donnelly, 2015; Doumith et al., 2004) due to the low discriminatory power of the method, the insufficient reproducibility and antigen sharing between serotypes, respectively (Schönberg et al., 1996). Testing with antisera sets is time-consuming and demanding. Diagnostic sera/antisera have to be checked on behalf of the internal quality assurance of the laboratory using the sera/antisera on a regular basis (Rili-BÄK, part B3; valid in Germany since 1.5.2015). Above all, the production of the sera requires the use of vertebrate animals and to the best of our knowledge there is currently only one manufacturer for *Listeria* sera/antisera worldwide (Denka Seiken, Japan). To circumvent these serotyping limitations, a five-plex PCR assay including genes *lmo0737*, *lmo1118*, *ORF2819*, *ORF2110*, and *prs* was developed for molecular separation of the four major serotypes (1/2a, 1/2b, 1/2c, and 4b) (Doumith et al., 2004).

With the recent evolution of whole-genome sequencing technologies, high-resolution typing schemes have been developed for the characterization of *L. monocytogenes* strains (Ruppitsch et al., 2015a; Pightling et al., 2015; Kwong et al., 2016; Maury et al., 2016). The developed core genome (cg) MLST scheme based typing of *L. monocytogenes* represents an expansion of the classical seven gene MLST scheme (Salcedo et al., 2003) and is in all aspects superior for the tracking and source identification, as compared to the current gold standard methods PFGE and fAFLP (Schmid et al., 2014; Ruppitsch et al., 2015b). Access to the genomic sequence allows not only strain characterization at a very high resolution, it also facilitates the rapid extraction of specific sequence data, thus making sequence-based serotyping quite promising as the new gold standard for the rapid and accurate characterization of *Listeria* strains (Kwong et al., 2016).

In our study, we applied next-generation sequencing (NGS) based core-genome (cg) MLST minimum-spanning tree (MST) analysis to a collection of 172 *L. monocytogenes* isolates from the Austrian National Reference Laboratory for Listeria (NRLL) including isolates from the Seeliger collection (Haase et al., 2011), as well as type strains for all known serotypes (Ruppitsch et al., 2015a), and in addition 45 isolates from the NRLL with previously uncharacterized serotypes (*i.e.* a total of 207 isolates), to determine the serogroup of all isolates *via* core genome MLST (cgMLST) profiles. Isolates with known serotypes and lineages were selected to cover the entire genomic diversity of the species as described previously (Ruppitsch et al., 2015a). In addition to the known multiplex-PCR targets (Doumith et al., 2004; Kwong et al., 2016) further serogroup-specific determinants were identified in this work to improve usability and robustness of the combined speciation and serotyping from WGS data in a single workflow.

2. Material and methods

2.1. Bacterial isolates and DNA purification

Complete genomes of four isolates were downloaded from NCBI (Table 1). The 168 isolates (Austria n = 92, Germany n = 48, USA n = 6, Unknown n = 6, Canada n = 5, France n = 4, United Kingdom n = 3, Denmark n = 3, New Zealand n = 1) including 15 reference strains comprising the twelve serotypes (1/2a n = 30, 1/2b n = 21, 1/2c n = 4, 3a n = 1, 3b n = 2, 3c n = 1, 4a n = 1, 4b n = 42, 4c n = 2, 4d n = 1, 4e n = 1, 7 n = 1, 61 unknown), which were used for the assignment of *L. monocytogenes* isolates to serogroups by cgMLST, using whole genome sequencing (WGS) are listed in Supplement Table 1. In addition, a set of 45 isolates (Supplement Table 2) from NRLL, each without previous serogroup characterization, was used to evaluate the strategies to assign serogroups by cgMLST and WGS (test set). All strains were cultured overnight at 37 °C on RAPIDíL.Mono agar (Bio-Rad, Vienna, Austria) for species confirmation and sub-cultured on Columbia blood agar plates (BioMérieux, Marcy l'Etoile, France) prior to DNA extraction using the MagAttract HMW DNA Kit, according to the instructions of the manufacturer (Qiagen, Hilden, Germany). Isolates of Seeliger's historical "Special Listeria Culture Collection" (Haase et al., 2011) deposited at the German-Austrian binational consiliary laboratory for Listeria (AGES Graz, Austria) were reconstituted from the original agar slant, including also the oldest available *Listeria* strain (SLCC208) isolated from a soldier during the first world war (Hyden et al., 2015; A. Leclercq, personal communication), by adding Trypticase Soy Broth (BioMérieux, Marcy l'Etoile, France). Subsequently, the strains were subcultured on Columbia blood agar plates (BioMérieux, Marcy l'Etoile, France) and plated on RAPID'L.Mono plates (Bio-Rad, Vienna, Austria) for species verification, and finally grown overnight on Columbia blood agar plates (BioMérieux, Marcy l'Etoile, France) for isolation of genomic DNA, using the MagAttract HMW DNA Kit (Qiagen, Hilden, Germany), according to the instructions of the manufacturer.

Molecular serotyping was performed for the confirmation of serogroups, as described previously (Doumith et al., 2004).

Table 1
Complete genomes downloaded from NCBI GenBank and used to identify serogroup specific genes. MLST clonal complexes (CC) and lineages in accordance to http://bigsdb.web.pasteur.fr/listeria/listeria.html and cgMLST cluster types in accordance to the core genome defined by Ruppitsch et al. (2015a).

Sample ID	Serotype	MLST CC	Lineage	cgMLST CT	GenBank Accession No.
SLCC2540	3b	617	Lineage I	31	NC_018586
EGD-e	1/2a	35	Lineage II	1	NC_003210
EGD	1/2a	12	Lineage II	3	HG421741
10403S	1/2a	85	Lineage II	27	NC_017544

2.2. Whole genome sequencing, assembly and data analysis

Sequencing libraries were prepared using Nextera XT chemistry (Illumina Inc., San Diego, CA, USA) for a 2 × 300 bp sequencing run on an Illumina MiSeq sequencer. Samples were sequenced with an aim of a minimum coverage of 70-fold by preparing a library of 72 *L. monocytogenes* genomes each time. The resulting FASTQ files were first quality-trimmed and then *de-novo* assembled using the Velvet assembler version 1.1.08 (Zerbino and Birney, 2008) and subsequently integrated into the Ridom SeqSphere$^+$ software (Ruppitsch et al., 2015a) version 3.1 (Ridom GmbH, Münster, Germany). Sequence reads were trimmed at their 5′- and 3′-ends until an average PHRED value of 30 was reached in a window of 20 bases. The assembly was performed with the Velvet assembler, with the k-mer values and coverage cutoffs automatically optimized for each genome, based on the average length of contigs with >1000 bp. Contigs with an overall length less than 200 bp or an average coverage below five were discarded.