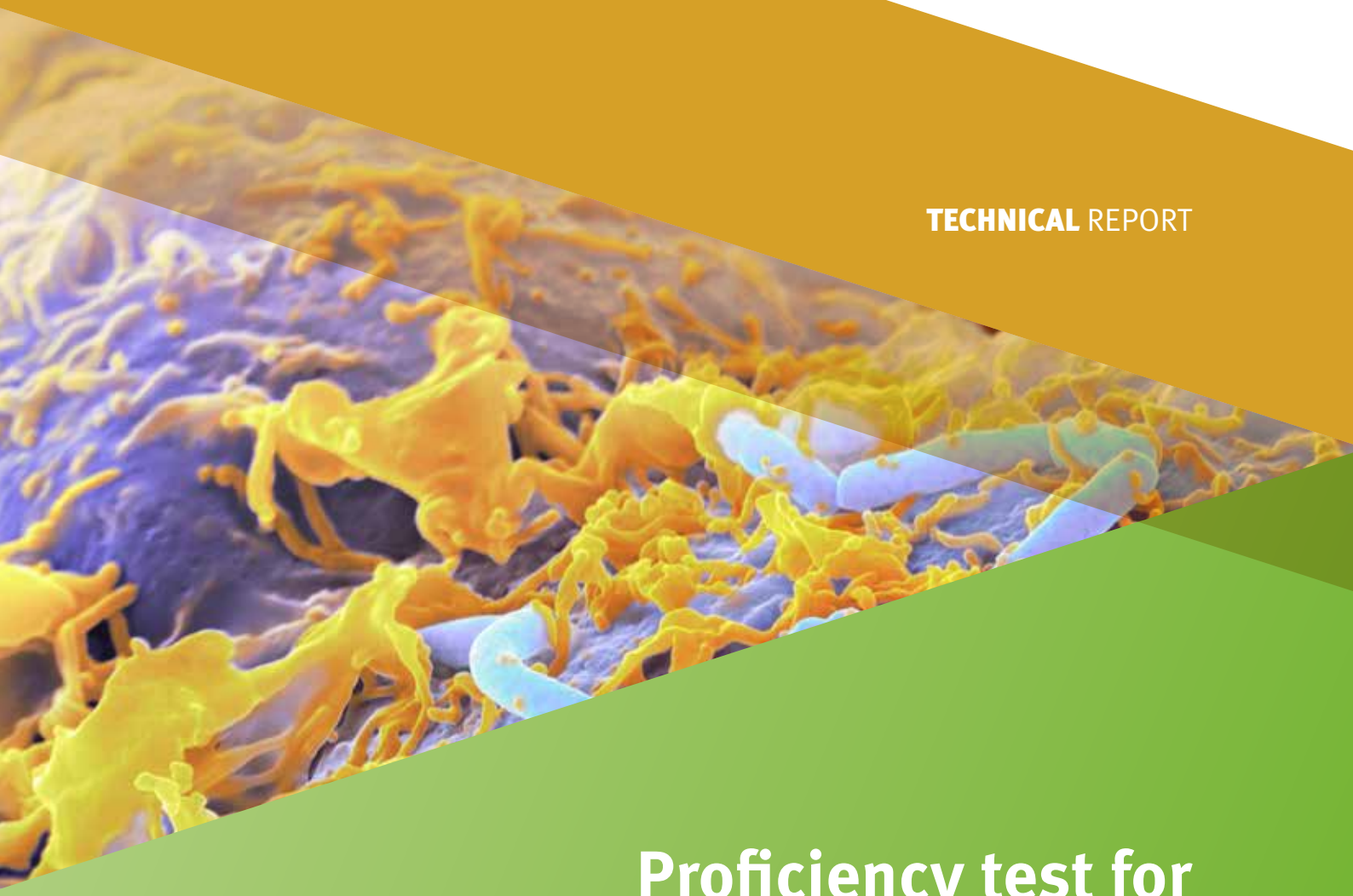
ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

TECHNICAL REPORT



Proficiency test for *Listeria monocytogenes* whole genome assembly

2018

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Proficiency test for *Listeria monocytogenes* whole genome assembly

2018

This report of the European Centre for Disease Prevention and Control (ECDC) was coordinated by Ivo Van Walle and supported by Erik Alm, Daniel Palm, Taina Niskanen, Marc Struelens and Johanna Takkinen.

Suggested citation: European Centre for Disease Prevention and Control. Proficiency test for *Listeria monocytogenes* whole genome assembly – 2018. Stockholm: ECDC; 2019.

Stockholm, May 2019

ISBN: 978-92-9498-339-8
DOI: 10.2900/188017
Catalogue number: TQ-03-19-325-EN-N

Contents

Figures

Tables

Annex

Abbreviations

AD	Allelic distance
BWA	Burrows-Wheeler Aligner
cgMLST	Core genome multilocus sequence typing
EFSA	European Food Safety Authority
QC	Quality control
WGS	Whole genome sequencing

Executive summary

Within its mission to operate EU surveillance networks, ECDC supports the integration of whole genome sequencing (WGS) data into surveillance and multi-country outbreak investigations of foodborne diseases including listeriosis as one of the priority diseases. To evaluate the inter-laboratory reproducibility and portability of *Listeria monocytogenes* genome assemblies, ECDC organised a proficiency test for national public health reference laboratories with WGS typing capabilities in the EU/EEA, as well as EFSA and the EU Reference Laboratory for *L. monocytogenes*.

This report presents the results of the proficiency test. Each participant received a total of 15 sets of raw sequence reads, which were to be assembled by one or more pipelines of their choice. The resulting assemblies were then compared to the reference assembly generated by ECDC on several quality metrics. There were 16 participants, submitting results for 29 pipelines.

Twelve participants, including 10 of the 14 participating public health reference laboratories, had at least one concordant pipeline for Illumina reads. The other participants were provided with individual feedback on possibilities to improve their pipeline(s). Participants with a concordant pipeline are recommended to use that for their own analyses as well as for any sharing of assemblies with other organisations including ECDC. For EU-level surveillance purposes ECDC will only accept assemblies generated with a concordant pipeline. Any new pipelines or updates to existing pipelines should go through the same proficiency testing before being used for sharing data with ECDC. For outbreak investigation purposes when more detailed analysis can be needed, raw sequence reads are proposed to be shared instead of or in addition to assemblies for isolates included in the cluster.

For Ion Torrent reads, it was not possible to establish concordance. ECDC suggests that any countries producing these reads share not only the reads with other organisations but also the extracted allele sequences for at least the core genome in the form of a fasta file. This was shown to produce acceptable results and allows other organisations, including ECDC, to perform their allele calling as with any regular assembly.

It was also found that the assembly process can be used to remove low-level contamination. Conversely, low-level contamination can give rise to much longer assembly lengths than the expected length due to the presence of a large number of very small contigs with very low quality. It is recommended that assembly pipelines include removal of such small and unreliable contigs, ideally in a way that still alerts the user to the likely presence of low-level contamination.

EU laboratories that have installed a new or updated pipeline are welcome to have its concordance assessed by ECDC at any time.

For further questions and comments, contact fwd@ecdc.europa.eu.

1 Introduction

The European Centre for Disease Prevention and Control (ECDC) is an independent European Union (EU) agency with a mandate to operate the dedicated surveillance networks. Its mission is to identify, assess and communicate current and emerging threats to human health from communicable diseases. ECDC fosters the development of sufficient capacity for diagnosis, detection, identification and characterisation of infectious agents that may threaten public health. It maintains and extends such cooperation and supports the implementation of quality assurance schemes [1,2].

ECDC supports the integration of whole genome sequence (WGS) data to enhance EU surveillance and multi-country outbreak investigations of communicable diseases including listeriosis [3]. Monitoring of national capacities for WGS-enhanced national surveillance of listeriosis indicated that 14 EU/EEA countries had capabilities covering approximately half of the EU's notified cases in 2017 [4]. These laboratories reported using a diversity of next-generation sequencing platforms and bioinformatic analysis pipelines. To evaluate the inter-laboratory reproducibility and portability of *Listeria monocytogenes (Lm)* genome assemblies, ECDC organised a proficiency test exercise for national public health reference laboratories with WGS typing capabilities in the EU/EEA, as well as EFSA and the EU Reference Laboratory for *L. monocytogenes* for feed and food safety.

The aim of the proficiency test is to support national public health reference laboratories performing WGS-based typing in generating good quality and comparable genome assemblies for *Lm*. Assemblies can be used for many analyses, including whole and core genome multilocus sequence typing (wg/cgMLST) data. Good quality assemblies are therefore necessary to produce wg/cgMLST data that are comparable between laboratories. In addition, as part of ECDC's initiative to strengthen Lm surveillance at the EU level, countries can also choose to submit assemblies to the European Surveillance System (TESSy).

The main objectives of the proficiency test are:

- evaluating individual laboratory and pipeline performance
- identifying and justifying problem areas; and
- providing continuing education and technical recommendations for achieving the required proficiency level for participation in EU surveillance.

2 Study design and methods

2.1 Selection of sequences

A set of 15 sequences were selected primarily to reflect:

- good intrinsic quality of the reads to avoid effects on quality not related to the assembly process. The average coverage of the genome was restricted between 55x–100x to avoid effects on quality from both too low and very high coverage. The reads were also all classified as 'Accepted' according to the quality controls (QCs) described in Van Walle et al. [5].
- the diversity in sequencing protocols and read lengths used in the EU/EEA to ensure that laboratories are able to process reads generated by other laboratories. Included were results generated with Illumina sequencers (MiSeq 2x150, MiSeq 2x250, MiSeq 2x300, NextSeq 2x150, HiSeq 2x100) and Ion Torrent.
- genetic similarity between sequences. One pair of sequences had 0 allelic differences (AD) on the core genome according to the reference methodology used.

It should be noted that an assembly that passes the QCs of point 1 above can still have low quality due to e.g. individual bases being wrongly called by the assembler. In particular, this can be the case if no post-assembly optimisation is performed by mapping the reads back onto the assembly and selecting the consensus base for each position (further referred to as consensus calling). This proficiency test is designed to assess primarily quality issues due to assembly since both the input sequence reads and resulting reference assemblies are selected not to have such quality issues. Table 1 gives an overview of the test sequences.

Table 1. Test sequences

Code	Sequencing protocol	Average genome coverage (N)	Comment
101	Illumina MiSeq 2X300	61	
102	Illumina MiSeq 2X250	65	
103	Illumina MiSeq 2X150	67	
104	Illumina NextSeq 2X150	93	Matches 105 with AD=0
105	Illumina HiSeq 2X100	72	Matches 104 with AD=0
106	Ion Torrent	75	
107	Illumina MiSeq 2X250	73	
108	Illumina MiSeq 2X150	84	
109	Illumina MiSeq 2X150	80	
110	Illumina MiSeq 2X150	61	Low-level contamination with *Citrobacter*[1]
111	Illumina HiSeq 2X100	78	
112	Illumina MiSeq 2X150	100	
113	Illumina MiSeq 2X150	99	
114	Illumina MiSeq 2X150	62	
115	Illumina MiSeq 2X150	88	Low-level contamination with *Methylobacterium*, *Sphingobium* and *Sphingomonas*[1]

[1]: detected by one of the participants.

2.2 Testing

Participants were provided with the de-identified raw reads as FASTQ files for each sequence. They were asked to provide the corresponding whole genome assemblies as FASTA files. Assemblies generated by up to three different pipelines per participant were accepted and a detailed description of each pipeline had to be provided as well in order to determine the cause of potential issues.

2.3 Analysis

The returned sequence assemblies were first imported into a BioNumerics database (BioNumerics 7.6.3, Applied Maths). Core genome MLST was called using the scheme of Moura et al. [6]. The QCs described in Van Walle et al. were subsequently applied to each sequence [5]. These included:

- three QCs for contamination with another species based on genome length and contamination based on alignment of the assembly to reference genomes of common contaminants, as well as of the expected species. These QCs are further referred to as 'QC genome length', 'QC common contaminants' and

'QC expected species'. The 'QC expected species' was not further considered because the sequences were selected to be *Lm*.
- one QC for sequencing of a non-pure culture, i.e. two or more clones of the expected species, based on the number of core genome loci with more than one allele detected (further referred to as 'multiple calls'. Allele calling was performed on assembly only. This QC is further referred to as 'QC core genome multiple calls' or 'QC CGM'.
- one QC for read and assembly quality based on the proportion of core genome loci detected. Allele calling was performed on assembly only. This QCs is further referred to as 'QC core genome coverage' or 'QC CGC'.

The result of each QC either passed ('PASS'), passed with a warning ('WARN') or failed ('FAIL'). After applying the QCs, the allelic distances (AD) between all assemblies were determined, not counting missing loci in either sequence as counting towards the distance.

The reference assemblies were generated with BioNumerics 7.6.3, using SPAdes 3.7.1 and consensus calling performed by BioNumerics, based on mapping the reads onto the assembly using Burrows-Wheeler Aligner (BWA) and selecting the consensus base per position [7,8]. Contigs with length <300 and <1 000 bp were removed from Illumina and Ion Torrent reads-based assemblies. Aside from the study of Van Walle et al. [5], it is a priori not guaranteed whether assemblies generated by this pipeline are in fact a good reference for these particular sequences. Two analyses were performed to assess this:

- distance to other pipelines. The proportion of assemblies generated by all the other pipelines with an allelic distance (AD) of 0, 1, 2, 3, 4, 5, 6 or ≥7 to the corresponding sequence from the selected pipeline using the core genome scheme of Moura et al. [6]. This expresses how similar this pipeline is in general to all the other pipelines. Sequence pairs for which one or both sequences failed or passed with a warning the sequence read/assembly quality QC were excluded from this analysis to avoid including artificially low AD values.
- distance between similar isolates. There was one pair of isolates in the dataset that had AD=0 based on the reference pipeline. Other pipelines may be near-identical or even identical to each other, e.g. based on the same assembler, and will as a result produce near-zero or zero differences between them for the same isolate. However, this may be the result of introducing the same mistakes in the assembly, and if these mistakes are introduced at different locations for the two similar isolates, the distance between the two similar isolates can be greater than zero.

The assemblies from each test pipeline where then compared to those of the reference pipeline, separately for each of the two platforms (Illumina: n=14, Ion Torrent: n=1). This was treated as an additional sixth QC, further referred to as 'QC core genome allele sequences' or 'QC CGS'. Each isolate was classified as PASS, WARN or FAIL on this metric when the AD between the test and reference pipeline is respectively 0-1, 2-3 or ≥4 alleles. This choice was made as in Van Walle et al. (2018). The cut-off of 4 is explicitly used, but as a cluster cut-off [5]. It would make sense then that isolates with more than 4 differences due to errors are classified as failing. PASS and WARN are equally distributed below that (0–1 and 2–3 differences respectively).

Finally, pipelines were classified per sequencing platform as concordant, near-concordant or discordant based on the following criteria:

- concordant: maximum 10% of isolates, rounded up, have QC core genome coverage, QC core genome multiple calls and/or QC allele sequences equal to WARN and the rest of the isolates have all three of these QCs equal to PASS. One FAIL in any of the three QCs for any of the isolates therefore excludes a pipeline from being concordant. This seemingly strict choice was made because of the relatively low number of isolates in the test and even one FAIL may therefore correspond in reality to a substantial fraction of isolates in routine application not having an assembly of sufficient quality.
- near-concordant: maximum 20% of isolates, rounded up, have QC core genome coverage, QC core genome multiple calls and/or QC allele sequences equal to WARN and the rest of the isolates have all three of these QCs equal to PASS. As for the concordant category, one FAIL in any of the three QCs for any of the isolates therefore excludes a pipeline from being near-concordant. This seemingly strict choice was made because of the relatively low number of isolates in the test and even one FAIL may therefore correspond in reality to a substantial fraction of isolates in routine application not having an assembly of sufficient quality.
- discordant: not concordant or near-concordant.

Additional quality issues related to the other three QCs on contamination were assessed qualitatively to determine if this is an issue or not.

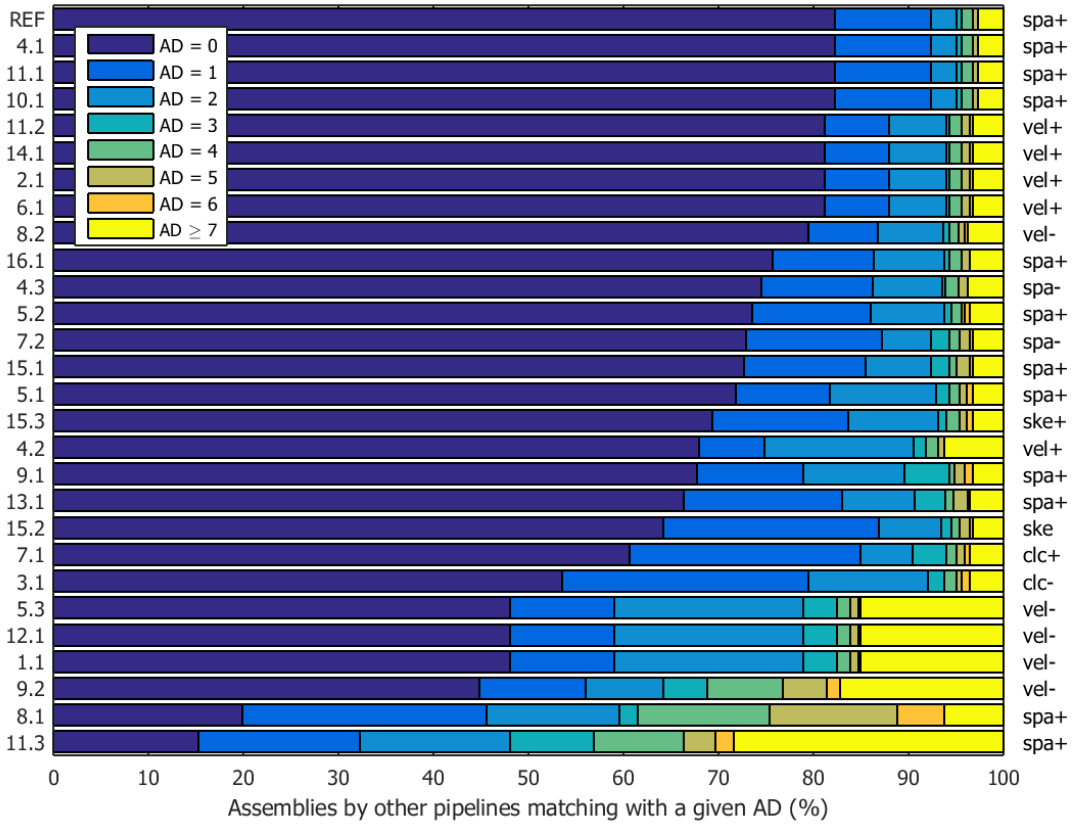
3 Results

3.1 Illumina reads

Figure 1 shows the distance histogram of each pipeline to all the others. Together with two other pipelines, the reference pipeline has the largest proportion of AD=0 distances, as well as AD<2 distances, to all the other pipelines, and by this criterion is the most 'central' assembly. Figure 2 shows the distance between the two isolates (104 and 105) that have AD=0 according to the reference pipeline, for each pair of pipelines. There are a number of pipelines that have a distance of AD≥4, and often substantially above this threshold, versus the same pipeline for the other isolate. Therefore they are not a good reference even if between subsets of similar pipelines they may produce comparable results for individual isolates. In addition, the distance between two isolates of a given pipeline is almost invariably the sum of the distances versus the reference, i.e. the first column and row sum up to the diagonal. This indicates that the differences in the pipeline occur at different parts of the genome for the two isolates, consistent with a random error. Assemblers used were SPAdes, Velvet, SKESA and CLC Assembly Cell (Qiagen) [7,9–10].

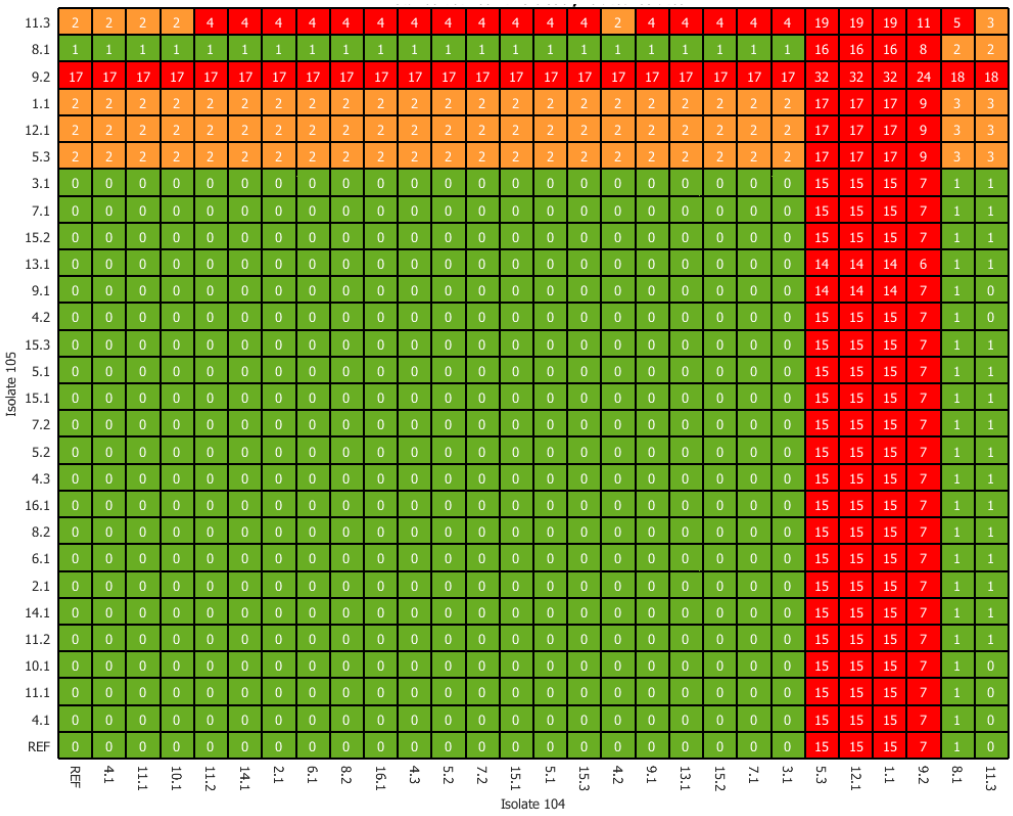
An overview of the core genome allele sequence quality comparison is given in Figures 3–4, which show respectively the median (rounded) and maximum allelic distance between each pipeline for all of the isolates sequenced on an Illumina platform. Only pairs of sequences that both pass the QC on core genome coverage are included. The final classification of each pipeline in terms of concordance with the reference is given in Table 2.

Figure 1. Distance histogram between assemblies from one pipeline and corresponding assemblies of all other pipelines, Illumina reads



Right axis indicates the assembler used (spa: SPAdes, vel: Velvet, ske: SKESA, clc: CLC Assembly Cell) and whether a form of consensus calling is performed (+: yes, -: no).

Figure 2. Core genome distance between one pair of similar isolates for all combinations of pipelines, Illumina reads



In accordance with the thresholds used for the QC on core genome allele sequences, distances of 0–1 (PASS) are coloured green, 2–3 (WARN) orange and ≥4 (FAIL) red.

Figure 3. Median core genome distance between the respective assemblies generated by each pair of pipelines, Illumina reads

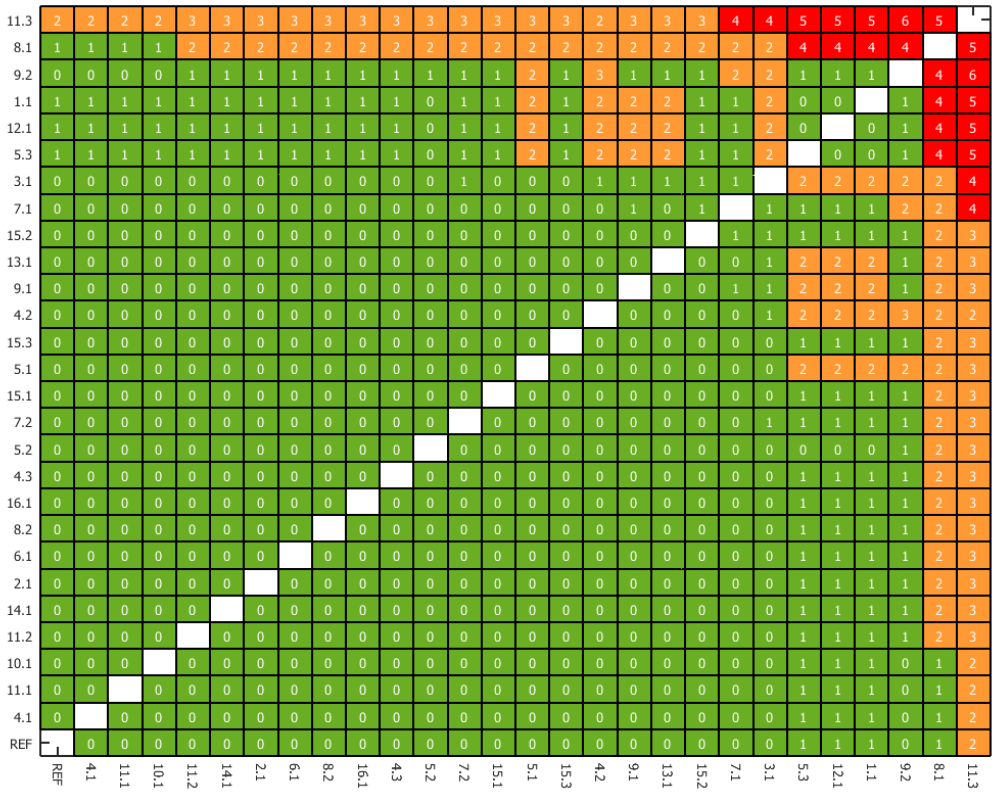
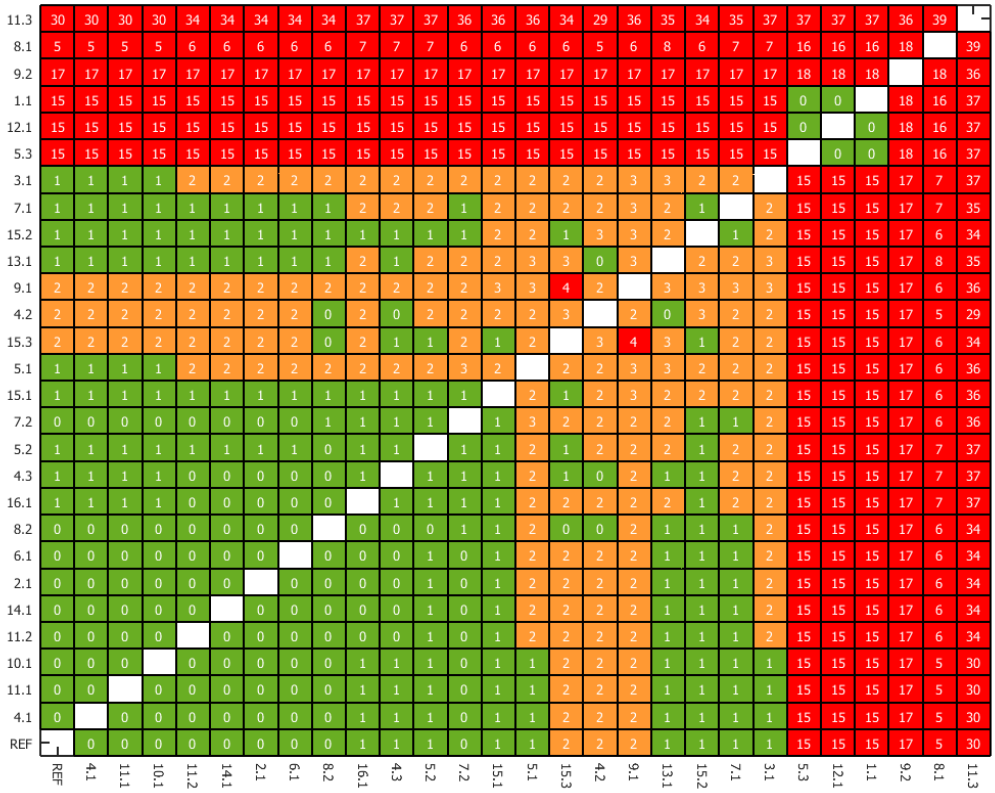


Figure 4. Maximum core genome distance between the respective assemblies generated by each pair of pipelines, Illumina reads



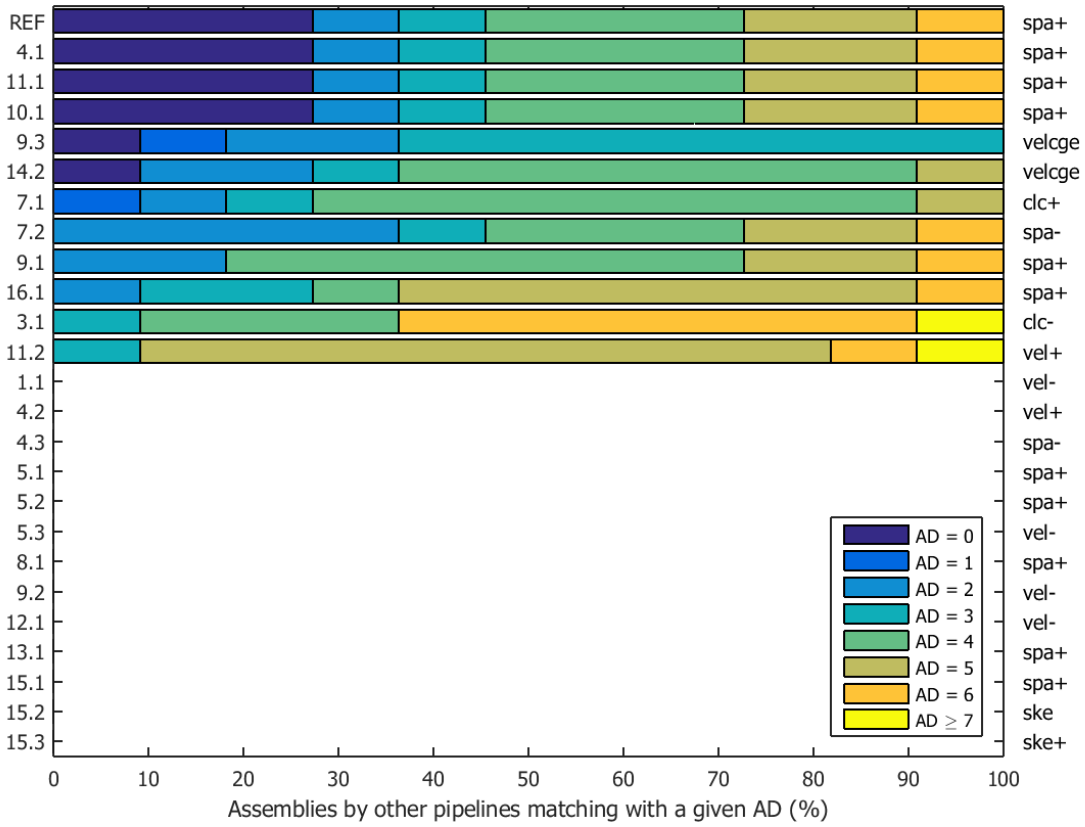
3.2 Ion Torrent reads

Figure 5 shows the distance histogram of each pipeline to all the others. Together with three other pipelines, the reference pipeline has the largest proportion of AD=0 distances, as well as AD<2 distances, to all the other pipelines, and by this criterion is the most 'central' assembly. However, the proportion is very low compared to that for Illumina reads. Figure 6 shows the distance between each pipeline for the single Ion Torrent isolate. Along the same line, these distances are much larger than for the Illumina reads and more than half of the pipelines do not have a distance calculated between them because the QC on core genome coverage for the one isolate is WARN or FAIL for either or both pipelines.

3.3 Low-level contamination

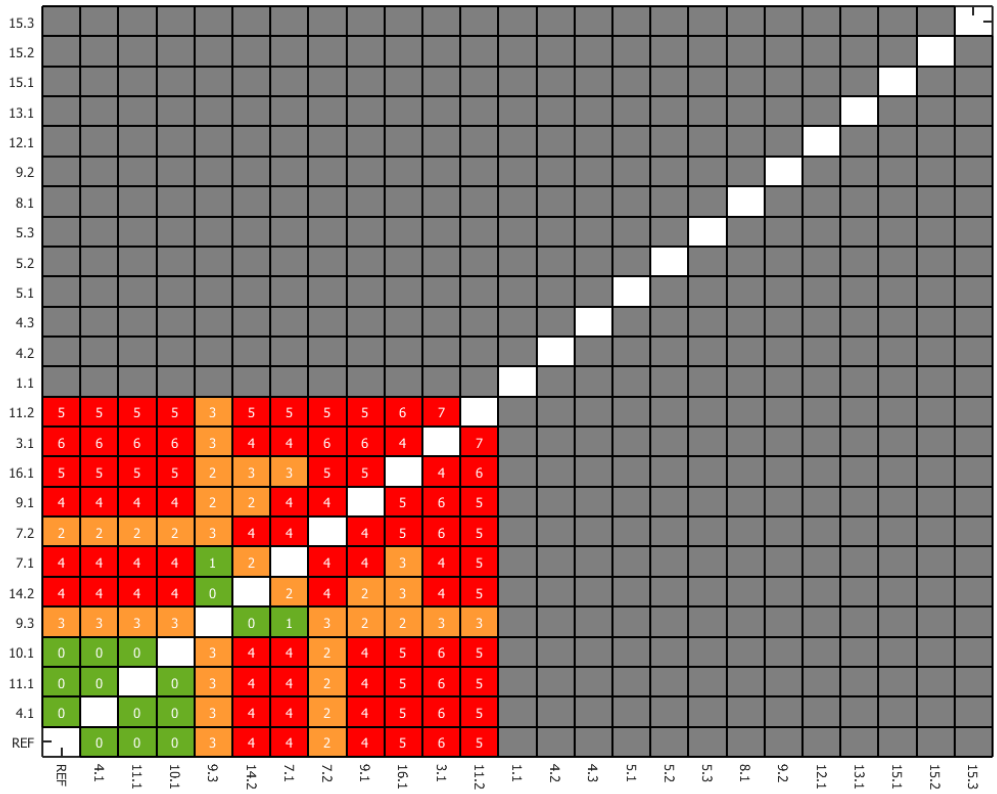
Two isolates (110 and 115) were found by one of the participants to have low-level contamination. Upon further inspection, this low-level contamination was found to give rise to substantially longer assembly lengths for certain pipelines almost exclusively due to very small (<1 kb) length contigs. In the reference assembly, these very small contigs were filtered out (Chapter 2).

Figure 5. Distance histogram between assemblies from one pipeline and corresponding assemblies of all other pipelines, Ion Torrent reads



Right axis indicates the assembler used (spa: SPAdes, vel: Velvet, ske: SKESA, clc: CLC Assembly Cell) and whether a form of consensus calling is performed (+: yes, -: no). Pipelines with no distances calculated to any other pipeline do not have a histogram shown.

Figure 6. Core genome distance between the respective assembly generated by each pair of pipelines, Ion Torrent reads



In accordance with the thresholds used for the QC on core genome allele sequences, distances of 0–1 (PASS) are coloured green, 2–3 (WARN) orange and ≥4 (FAIL) red. Distances not calculated are in grey.

Table 2. Concordance results for all pipelines

Pipeline	Assembler[1]	Concordance Illumina	Low-level contamination
REF	SPAdes+	Reference	None detected due to removal of small contigs
1.1	Velvet-	Discordant	Not detected, (mostly) filtered out by assembly pipeline
2.1	Velvet+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
3.1	CLC Assembly Cell-	Concordant	Detected in both isolates
4.1	SPAdes+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
4.2	Velvet+	Discordant	Not detected, (mostly) filtered out by assembly pipeline
4.3	SPAdes-	Discordant	Detected in both isolates
5.1	SPAdes+	Concordant	Detected in both isolates
5.2	SPAdes+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
5.3	Velvet-	Discordant	Not detected, (mostly) filtered out by assembly pipeline
6.1	Velvet+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
7.1	CLC Assembly Cell+	Concordant	Detected in one of the two isolates
7.2	SPAdes-	Concordant	Detected in both isolates
8.1	SPAdes+	Discordant	Detected in both isolates
8.2	Velvet-	Discordant	Not detected, (mostly) filtered out by assembly pipeline
9.1	SPAdes+	Discordant	Detected in both isolates
9.2	Velvet-	Discordant	Not detected, (mostly) filtered out by assembly pipeline
9.3	Velvet(unk)	n/a	Not detected, (mostly) filtered out by assembly pipeline
10.1	SPAdes+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
11.1	SPAdes+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
11.2	Velvet+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
11.3	SPAdes+	Discordant	Detected in both isolates
12.1	Velvet-	Discordant	Not detected, (mostly) filtered out by assembly pipeline
13.1	SPAdes+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
14.1	Velvet+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
14.2	Velvet(unk)	n/a	Not detected, (mostly) filtered out by assembly pipeline
15.1	SPAdes+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
15.2	SKESA-	Concordant	Not detected, (mostly) filtered out by assembly pipeline
15.3	SKESA+	Concordant	Not detected, (mostly) filtered out by assembly pipeline
16.1	SPAdes+	Concordant	Not detected, (mostly) filtered out by assembly pipeline

+: consensus calling performed (reads mapped back onto assembly and consensus base per position is chosen)
-: consensus calling not performed.
n/a.: not applicable (pipeline not used for Illumina reads).

3.4 Participant survey

After the proficiency test, an anonymous survey was performed on all 14 participating public health national reference laboratories, to which six participants replied. Five where satisfied with the received individual test report and one suggested using a newer SPAdes version for the reference assemblies. Regarding the question whether results would be used as documentation for accreditation, one replied yes, three no (of which one specified that there are no plans for accreditation) and two not applicable. Two participants suggested for the future to also include allele calling in the proficiency test.

4 Discussion and conclusion

A total of 16 organisations, including 14 EU/EEA public health national reference laboratories, EFSA and the EURL for *L. monocytogenes,* participated in the proficiency test, providing results for 29 assembly pipelines. For the Illumina reads, the reference assemblies were found to be appropriate and 10 of 14 public health national reference laboratories had at least one pipeline that was concordant with the reference. For these participants, sharing assemblies generated using (one of) their concordant pipelines for core genome comparison, either among each other or with ECDC, can be considered acceptable with respect to quality. The remaining four laboratories were each provided with individual feedback regarding possible causes for the issues found further assisted individually. Two of the laboratories have since updated their pipeline and generated concordant results. For the third laboratory, the cause is known and dependent on a software upgrade and this is possibly the same case for the fourth laboratory. Laboratories not having a concordant assembly pipeline should consider updating their pipeline as soon as possible to avoid issues with outbreak detection and in the meantime share raw reads with ECDC.

In terms of assemblers, concordant results were generated with SPAdes, Velvet, SKESA and CLC Assembly Cell. SPAdes and Velvet were the most widely used and gave rise to both concordant and discordant results, most likely due to the application or non-application of well performing consensus calling. For SPAdes, the most recent versions (3.11 and above) are expected to have a well performing built-in consensus calling and when switched on, they all delivered concordant results. Velvet does not have its own consensus calling and adding it would be helpful. CLC Assembly and SKESA were only used by two and one participants respectively. The impact of consensus calling on these cannot be assessed at this time, but adding an (additional) consensus calling step will likely only improve the quality. Among the commercial packages, BioNumerics/SPAdes and CLC Assembly were concordant and SeqSphere/Velvet was concordant provided that consensus calling was switched on.

For Ion Torrent reads, the appropriateness of the reference could not be well established and consequently concordance of a pipeline was not considered as a reliable quality indicator. Given that only very few EU/EEA countries generate Ion Torrent sequences for *Lm*, it is recommended that these countries share not only reads with other countries and/or ECDC, but also themselves extract the allele sequences as described in Van Walle et al. and share them since this method is shown to give reliable results [5]. Countries receiving these allele sequences in the form of a FASTA file can then perform allele calling on the data just like on a regular assembly.

The presence of low-level contamination caused some pipelines to generate assemblies with a very large amount of very small contigs and often with keyword coverage (in the case of SPAdes) below 1x. As a result, the assembly length was often more than one megabase longer than the expected length of 2.8–3.3 megabases. This interfered with the QC on genome length and also adds unnecessary clutter to the assembly. It is therefore recommended to include in the assembly pipeline both a minimum contig length (300 bp was the minimum used among those pipelines that applied this) and a minimum (keyword) coverage for each contig. In this way, the assembly process is also used to filter out low-level contamination. If it is important to pick up such low-level contamination events, either the reads can be used to detect them or the assembly before the removal of the small contigs. However, substantial contamination will not be removed by this assembly post-processing step as both the contaminant contig sizes and coverage will be larger.

5 Recommendations

For European surveillance of listeriosis, ECDC will only accept from laboratories assemblies that have been generated by a concordant pipeline as verified according to the methodology described in this report. Reporting laboratories should ensure their assembly pipeline is concordant, primarily to avoid issues with comparability and thus potentially affecting outbreak detection/verification. In addition, sharing of assemblies for comparison with other laboratories or ECDC can then also be done confidently.

Future *in silico* proficiency testing exercises may include allele calling concordance testing for certifying pipeline compatibility and comparability of WGS data.

References

1. European Parliament and Council of the European Union. Regulation (EC) No 851/2004 of the European Parliament and of the Council of 21 April 2004 establishing a European centre for disease prevention and control – Article 5.3. Strasbourg: European Parliament; 2004. Available from: http://ecdc.europa.eu/en/aboutus/Key%20Documents/0404_KD_Regulation_establishing_ECDC.pdf
2. European Parliament and Council of the European Union. Decision No 1082/2013/EU of the European Parliament and the Council of 22 October 2013 on serious cross-border threats to health and repealing Decision No 2119/98/EC (Text with EEA relevance). Strasbourg: European Parliament; 2013. Available from: http://ec.europa.eu/health/preparedness_response/docs/decision_serious_crossborder_threats_22102013_en.pdf
3. European Centre for Disease Prevention and Control. ECDC roadmap for integration of molecular and genomic typing into European-level surveillance and epidemic preparedness – Version 2.1, 2016-2019. Stockholm: ECDC; 2016. Available from: http://ecdc.europa.eu/publications-data/ecdc-roadmap-integration-molecular-typing-and-genomic-typing-european-level
4. European Centre for Disease Prevention and Control. Monitoring use of whole-genome sequencing for infectious diseases surveillance in Europe – 2015-2017. Stockholm: ECDC; 2018. Available from: http://ecdc.europa.eu/publications-data/monitoring-use-whole-genome-sequencing-infectious-disease-surveillance-europe
5. Van Walle I, Björkman JT, Cormican M, Dallman T, Mossong J, Moura A, et al. Retrospective validation of whole genome sequencing-enhanced surveillance of listeriosis in Europe, 2010 to 2015. Euro Surveill. 2018 Aug;23(33). Available from: http://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2018.23.33.1700798
6. Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, et al. Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. Nat Microbiol. 2016 Oct 10;2:16185.
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012 May;19(5):455-77.
8. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010 Mar 1;26(5):589-95.
9. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008 May;18(5):821-9.
10. Souvorov A, Agarwala R, Lipman DJ. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol. 2018 Oct 4;19(1):153.

Annex. List of participants

Country	Laboratory	Institute
Austria	NRC Listeria	Institute for Medical Microbiology and Hygiene Graz, AGES
Belgium	National Reference Centre Listeria	Scientific Institute Public Health
Czech Republic	National Reference Laboratory for Listeria	The National Instituteof Public Health
Denmark	Foodborne Infections	Statens Serum Institut
Spain	Neisseria, Listeria and Bordetella Unit	National Centre for Microbiology, Instituto de Salud Carlos III
Finland	Expert Microbiology	National Institute for Health and Welfare
Hungary	Department of phage-typing and molecular epidemiology	National Public Health Institute
Ireland	National Salmonella, Shigella and Listeria Reference Laboratory	University Hospital Galway
Italy	Department of Infectious diseases	Istituto Superiore di Sanità
Luxembourg	Epidémiologie et Génomique Microbienne	Laboratoire National de Santé
The Netherlands	Centre for Infectious Research, Diagnostics and Screening	National Institute for Public Health and the Environment
Norway	National Reference Laboratory for Enteropathogenic Bacteria	Norwegian Institute of Public Health
Portugal	Departamento de Doenças Infeciosas	Instituto Nacional de Saúde Doutor Ricardo Jorge
Sweden	MI-LB	Folkhälsomyndigheten
n/a	n/a	European Food Safety Authority
n/a	EU Reference Laboratory for *Listeria monocytogenes*	Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail

European Centre for Disease Prevention and Control (ECDC)

Gustav III:s Boulevard 40, 16973 Solna, Sweden

Tel. +46 858601000
Fax +46 858601001
www.ecdc.europa.eu

An agency of the European Union
www.europa.eu

Subscribe to our publications
www.ecdc.europa.eu/en/publications

Contact us
publications@ecdc.europa.eu

Follow us on Twitter
@ECDC_EU

Like our Facebook page
www.facebook.com/ECDC.EU

ECDC is committed to ensuring the transparency and independence of its work

In accordance with the *Staff Regulations for Officials and Conditions of Employment of Other Servants of the European Union* and the *ECDC Independence Policy*, ECDC staff members shall not, in the performance of their duties, deal with matters in which they may, directly or indirectly, have a personal interest that could impair their independence. Declarations of interest must be received from any prospective contractor before a contract can be awarded.
www.ecdc.europa.eu/en/aboutus/transparency

Publications Office of the European Union