The ENGAGE Proficiency Test Report 2016

THE ENGAGE PROFICIENCY TEST REPORT 2016

Oksana Lukjancenko, Susanne Karlsmose Pedersen, Rene S. Hendriksen

1. edition, March 2017.

Technical University of Denmark, National Food Institute, Research Group of Genomic Epidemiology, Kgs. Lyngby, Denmark

Contents

1. Introduction

The main objective of this proficiency test (PT) is to facilitate the production of reliable laboratory results of consistently good quality within the area of whole genome sequencing (WGS).

The PT evaluates the consistency and robustness of ENGAGE consortium members' ability to perform deoxyribonucleic acid (DNA) extraction, library preparation, the WGS, and assembly following different laboratory protocols, software tools, and sequence platforms for the reliability of submitted sequence data to the public repositories. This ensures harmonization and standardization in WGS and data analysis, with the aim to produce comparable data for the ENGAGE initiative. To meet these objectives, the laboratory work and analyses performed for this PT should be performed using the methods routinely employed in the individual laboratories.

The PT consists of a "Wet-lab" component targeting three common bacterial pathogens. The Wet-lab components assess the laboratories ability to perform DNA preparation, sequencing procedures and, if laboratories routinely do so, the analysis of epidemiological markers; Multi Locus Sequence Typing (MLST) and antimicrobial resistance (AMR) genes.

The individual laboratory data are confidential and only known by the participating laboratory and the PT organizers (DTU Food).

2. Materials and Methods

2.1 Participants

A pre-notification to announce the ENGAGE proficiency test was distributed on the 12th July 2016 by e-mail to the eight ENGAGE consortium partners. Seven of the eight partners signed up and participated in the PT. Only, the National Institute of Public Health – National Institute of Hygiene in Poland did not participate as they have not initiated in-house WGS. Some of the seven partners however, only took part in testing a subset of the target organisms after agreement with the PT organizers.

2.2 Strains

Two strains of *Campylobacter jejuni*, *Listeria monocytogenes*, and *Klebsiella pneumonia* were selected for the wet-lab in 2016. In a GMI end-user analysis of what species to target, *Campylobacter* and *Listeria* have been indicated being of interest (2). *Campylobacter* was selected for this PT due to its many repeats and rearrangements and *Listeria* due to it being part of many genomic pilot projects and it's genetically heterogeneous with limited repeats and rearrangement. One of the *Listeria* strains belonged to a less virulent MLST – ST-121, whereas the other strain was of to a known virulent type, ST-2. We also included *Klebsiella* due to its many resistance genes for evaluating if the detection of these as can be used to indicate the quality level of the sequencing.

Individual sets of the strains were lyophilized as KWIK STIKs by Microbiologics, St. Cloud, Minnesota, USA and the corresponding DNA were purified and pooled by DTU-Food prior to distribution in individual vials for each participant.

To better be able to assess the differences in the sequences generated by the participants, each of the six strains in the Wet-lab component were sequenced on the PacBio to get a closed reference genome. This was done by creating 10kb template libraries using “10kb DNA Template Prep Kit 1.0” from Pacific Biosciences, which were then sequenced using C2 chemistry on single-molecule real-time (SMRT) cells with a 180min collection protocol. The data was then de novo assembled using the Hierarchical Genome Assembly Process (HGAP) within the Pacific Biosciences SMRTAnalysis software package. Polishing and finishing the genome were performed with custom python scripts, Quiver and Gepard, a dot plot tool to identify overlapping regions.

2.3 Distribution

On 24th October 2016, bacterial strains in agar stab cultures together with the corresponding purified and dried DNA and a welcome letter were dispatched in double pack containers (class UN 6.2) to the participating laboratories according to the International Air Transport Association (IATA) regulations as UN3373, biological substances Category B.

2.4 Procedure

The protocol was made available on the website (http://www.globalmicrobialidentifier.org/Workgroups/About-the-GMI-Proficiency-Test-2016) allowing the PT participants access to all necessary information at any time. Additional relevant information was distributed by email directly to the participants.

The protocol presented instructions as to the handling of the received bacterial cultures and DNA.

Participants were requested to capture information in relation to the questions presented in the SurveyMonkey.

Deadline for submission of results was initially set for 14th December 2016 but was extended to 13th January 2017.After this date, participants, #93 and #104 who had not yet submitted results according to the level of their sign-up, were approached to confirm if they were planning on submitting results. By the beginning of February 2017, all relevant data was captured and the data analysis was instigated. This report summarizes the results and allows for ensures full anonymity for the participants, as only the PT-organizers has access to the individual results.

2.5.1 SurveyMonkey

Apart from three questions relating to the contact information of the participant, 40 questions were asked focused on the storage of bacterial cultures and DNA prior to analysis, the cultivation and DNA extraction procedure, the quality assurance parameters applied, details related to the sequencing and analysis of the obtained sequencing data.

2.5.2 Sequencing

The participants uploaded raw sequence files in fastq format. The reads were *de novo* assembled applying the standard assembly pipeline used by the web-services from Center for Genomic Epidemiology (CGE) https://cge.cbs.dtu.dk//services/all.php, except for the reads which were not trimmed prior to the assembly.

For the raw reads, the following QC metrics were calculated:

- Number of reads that map to reference chromosome
- Proportion of reads that map to reference chromosome out of all reads that map to total reference DNA
- Coverage, total reference DNA. The number of reads mapping to the total reference DNA multiplied with the average length of the reads divided by the total size of the reference genome

For the assemblies, the following QC parameters were calculated:

- Size of assembled genome
- Size of assembled genome per total size of DNA sequence
- Total number of contigs
- N50 (defined as the length of the shortest contig, in the set of largest contigs that represents at least 50% of the assembly)

In addition to the calculation of the above QC metrics and parameters, participants were requested to provide the identification of the strains corresponding MLST and AMR genes to support the assessment of the sequence quality. Participates identified the MLSTs and AMR genes using the software of their choice. To assess the proficiency of the participants, the PT organizers used a command line version of the CGE MLST-Finder v.1.7 (1) and ResFinder 2.1 (3) (Threshold for %ID = 98% and HSP/Query length = 60%) including the CGE standard assembly pipeline on the participant's raw reads to compare the results with those reported by the participants.

3. Results

3.1 Participation

Seven laboratories responded to the pre-notification and were enrolled in the ENGAGE PT. When the deadline for submitting results was reached, all seven laboratories had uploaded data. Seven partners, #104, #114, #115, #77, #82, #93, and #95 submitted raw reads of both the culture and the DNA for both *Campylobacter* strains. Only four partners, #104, #115, #77, and #82 submitted raw reads of both the culture and the DNA for both *Listeria* strains and five partners, #104, #114, #77, #82, and #93 submitted raw reads of the *Klebsiella* cultures and the DNA.

3.2 Method description

The bacterial cultures were stored at 4°C by 86% (n = 6) of the participants prior to the analysis. In addition, one participant, #104 (14%) stored the reference material at -20°C.

Four participants (57%) stored the DNA in the time between reception and processing at room temperature (5 days by #115, 12 days by #77, 14 days by #82 and 41 days by #114) whereas the remaining three participants, #93, #95 and #104 stored the DNA at 4°C.

All seven participants inoculated the bacterial cultures onto various types of blood agar. The *Listeria* and *Klebsiella* strains were incubated at 37°C between 16 to 24 hours in contrast to *Campylobacter* which were incubated at 42°C for 48 hours.

By five partners, the Genomic DNA was extracted from the Gram negative and positive using a number of different commercially available kits including, Easy-DNA and PureLink Genomic DNA Mini Kit (Gram negative) from Invitrogen, Minikit (Gram negative) and QIAamp DNA Mini kit from Qiagen, Charge Switch gDNA Mini Bacteria Kit (Gram positive) and Genomic Mini from A & A Biotechnology. Two of the participants have modified the used Gram positive protocols by lysostaphin treatment prior to extraction. Two partners used a commercially available automatically DNA purification instrument/ robot, the MagNA Pure LC / MagNA Pure LC DNA Isolation Kit III (Bacteria, Fungi) from Roche and the QIAsymphony/ DSP DNA Mini Kit from Qiagen.

DNA concentrations (ng/µl) of the bacterial cultures and DNA were determined prior to library preparation on a Qubit by four partners. In addition, one participant used the Nanodrop and another participant the GloMax® 96 Microplate Luminometer (QIAsymphony) and a third a quantifluor kit read on POLARstar Omega plate reader.

For the *Campylobacter* cultures, the DNA concentration ranged from 0.26 to 80 ng/µl and from 0.22 to 647.39 ng/µl for the provided DNA (Table 1). For the *Listeria* culture, the DNA concentration ranged from 0.28 to 92.05 ng/µl and from 0.24 to 33.52 ng/µl for the DNA. The DNA concentration ranged from 0.18 to 34.3 ng/µl and from 0.24 to 58.3 ng/µl for the *Klebsiella* bacterial culture and DNA, respectively (Table 1).

For the *Campylobacter* culture, the total amount of DNA ranged from 0.001 to 4.8 µg and from 0.001 to 3.12 µg for the provided DNA (Table 2). For the *Listeria* culture the total amount of DNA ranged from 0.001 to 4.88 µg and from 0.001 to 3.11 µg for the total amount of DNA. The total amount of DNA ranged from 0.001 to 1.43 µg and from 0.001 to 3.43 µg for the *Klebsiella* culture and DNA, respectively (Table 2). Laboratory #77, consistently reported the concentrations of 0.001.

All seven participants responded to the method applied to measure the DNA quality (e.g. RIN, 260/280 ratio and/or 260/230 ratio) prior to library preparation for bacterial cultures and DNA received. For bacterial cultures, two (29%) of the laboratories used the Nanodrop, one used the Qubit, another one the BioPhotometer plus (Eppendorf), and a third one quantifluor kit read on POLARstar Omega plate reader. In addition, two (29%) did not measure the DNA quality. For the DNA received, the laboratories used the same method to measure the DNA concentration except for one of the participants that reported not measuring the DNA of the cultures, and which used the Nanodrop.

Up to five of the laboratories depending on participation reported the measurement of the DNA quality (e.g. RIN or 260/280 ratio) for bacterial cultures and DNA received (Table 3). Among the five laboratories providing data of the DNA quality for the cultures, the level ranged from about 1.47 to 12.1 (Table 3).

Four participants reported the measurement of the DNA quality (260/230 ratio) for bacterial cultures and DNA received (Table 4). For the cultures and received DNA, the DNA quality ranged from 0.2 to 2.4 (Table 4).

Two out of the seven laboratories assessed the quality visually on an agarose gel.

Of the seven participants, five used the Illumina Nextera XT DNA sample preparation kit FC-131-1024 (n = 2) and FC-131-1096 (n = 2) and one indicated using the Illumina NEB Next Ultra DNA Library prep kit E6040L for the preparation of the sample library before sequencing. Two participants using the Illumina Nextera XT DNA kit FC-131-1024 or FC-131-1096, respectively indicated using this in combination with the Nextera XT Index kit FC-131-1001 or FC-121-1012. In addition, one participant did not indicate the cat no. but the lot no.

The genomic DNA was prepared for pair-end sequencing by all seven (100%) participants. The libraries were sequenced by five participants (71%) using an Illumina MiSeq platform whereas two used the HiSeq 2000 or the HiSeq 2500 platforms, respectively. The read length of the sequences was set between 100 (n = 1), 250 (n = 1), 251 (n = 3) and up to 300 bp (n = 1). The reads were trimmed before upload by one, #115 out of the seven participants using trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic). Five participants indicated that if assembled by themselves, three would have used SPAdes http://bioinf.spbau.ru/spades, one would have used the PATRIC provided tool (https://www.patricbrc.org) ; Assembly Strategy: FullSpades, output file "contigs.fa" and finally one have used Assembler 1.2: https://cge.cbs.dtu.dk/servicesAssembler available from CGE.

3.3 Sequencing, MLST, and antimicrobial resistance genes

For *Campylobacter* GMI16-001 the expected MLST was ST7426 which was found by all laboratories except for Laboratory #114 that had mixed up the two strains GMI16-001-BACT and GMI16-002-BACT as well as GMI16-001-DNA and GMI16-002-DNA, explaining the incorrect MLSTs. Two laboratories, #95 and #114 did not report MLST data (own tool) but these were provided by PT-organizer (CGE tool) and found correct (Table 5).

The *Campylobacter*, GMI16-001 was pan-susceptible why no resistance genes were expected. Laboratory #114 reported however, resistance data matching the profile of GMI16-002 due to the mix up (Table 6).

The MLST ST6238 was expected in *Campylobacter* strain; GMI16-002. This was reported by all participants except for Laboratory #114 due to the above reported mistakes (Table 5).

A very high degree of concordance was observed between the reported resistance genes detected by own tools and the CGE reference tool and between culture and DNA samples. Only four participants reported what own tool being used to identify the resistance genes, #77, #82, and #93 used the CGE ResFinder whereas #104 used Blastn. Some of the resistance genes, were determined “like” which indicate that the homology to the reference genes were less than 100% which is often seen due to minute sequencing errors. The gene *aph*(2'')-like was reported by a number of laboratories. In contrast, the CGE tool did not detect this specific gene which doesn't mean that it is not present. It merely indicate

that the commandline version of the CGE ResFinder tool did not pick up this gene most likely due to a higher threshold in homology than used by the laboratories. Laboratory #82 reported chromosomal point mutations which are not yet included the commandline version of the CGE ResFinder tool why this very well could be true. Running the commandline version of the CGE ResFinder tool for the genome of *Campylobacter* strain; GMI16-002 submitted by laboratory #114 showed resistance genes that do not match any of the expected profiles of the PT strains (Table 6).

Only four laboratories, #77, #82, #104, and #115 tested the two *Listeria* strains; GMI16-003 and GMI16-004. In all cases, the four laboratories managed to identify the correct and expected MLST ST-2 and ST-121, respectively (Table 5).

The two *Listeria* strains were both pan-susceptible and no resistance genes were reported nor identified using the commandline version of the CGE ResFinder tool.

Five laboratories, #77, #82, #93, #104, and #114 tested the *Klebsiella* strains, GMI16-005. The commandline version of the CGE MLSTFinder tool was used to test the submitted genome, GMI16-005-DNA laboratory #114 which didn't submit own data. In all cases, the laboratories managed to identify the correct and expected MLST ST-512 (Table 5).

The same laboratories were involved in testing the *Klebsiella* strains, GMI16-006. For this strain laboratories #114 didn't submit own data. All MLST profiles were correct, ST-15 (Table 5).

Both of the *Klebsiella* strains were multidrug resistant harbouring a number of resistance genes (Table 7-8). *Klebsiella* strains, GMI16-005 were found to contain the following genes, *aad*A2, *aac*(6')-Ib), bla_{TEM-1A}, bla_{KPC-3}, bla_{OXA-9}, bla_{SHV-11}, *oqx*A, *oqx*B, *aac*(6')Ib-cr, *fos*A, *mph*(A), *cat*A1, *sul*1, and *dfr*A12. Most of the genes were identified by both own and CGE tools indicated by a very high concordance. Several of the laboratories report the genes being with a lower homology than the reference gene indicated by being determined "like". The mutation, *aac*(6')Ib-cr was not identified by laboratory #82 using own tools for both the culture and DNA in contrast to the CGE tool. The laboratory however, identified the presence of the gene, *aac*(6')Ib as all did. This indicate that the laboratory might have used another tool not able to identify this mutation in the *aac*(6')Ib gene. Similarly, the CGE tool was not able to detect neither the *aac*(6')Ib nor the mutation *aac*(6')Ib-cr in GMI16-005-DNA for laboratory #104 indicating a potentially truncated gene. Almost all of the laboratories identified the gene, *fos*A in a "like" version. The commandline version of the CGE ResFinder tool did not pick up this gene most likely due to a higher threshold in homology than used by the laboratories (Table 7).

The *Klebsiella* strains, GMI16-006 contained the following genes, *aad*A1, *aac*(6')-Ib, *aac*(3)-Iid, *aph*(3')-Via, *str*A, *str*B, bla_{NDM-1}, bla_{OXA-9}, $bla_{CTX-M-15}$, bla_{SHV-1}/ bla_{SHV-28}, bla_{TEM-1b}/bla_{TEM-1a}, *qnr*S1, *oqx*B, *oqx*A, *aac*(6')Ib-cr, *sul*2, *tet*(D), *dfr*A14, and *fos*A. The concordance was very high between the laboratories testing the strain GMI16-006. In two incidences, the commandline version of the CGE ResFinder tool identified bla_{TEM-1a} whereas all "own" testing as well as the remaining testing by the CGE tool identified the gene, $bla_{TEM-1b.}$ The difference between the two genes is only a few SNPs why the error is often observed. Inconsistences in detection of the gene bla_{SHV} gene were observed. Some laboratories couldn't distinguish the type of bla_{SHV} and reported bla_{SHV-28} or bla_{SHV-28}, respectively. Consistency however, between "own" and CGE data was seen. The same explanation given for the bla_{TEM} gene also accounts the bla_{SHV} gene. Almost all of the laboratories identified the gene, *fos*A in a "like" version. The

commandline version of the CGE ResFinder tool did not pick up this gene most likely due to a higher threshold in homology than used by the laboratories. Laboratory #93 reported the detection of the *bla*$_{\text{LEN-12-Like}}$ gene not reported by others (Table 8).

3.3 Sequencing, Quality markers

All seven laboratories submitted sequencing data for the *Campylobacter* GMI16-001 and GMI16-002 related to the quality metrics and parameters from both the received bacterial culture and corresponding DNA. For Listeria GMI16-003 and GMI16-004, four laboratories participated, #77, #82, #104, and #115 both the received bacterial culture and corresponding DNA except for laboratory #77 which didn't submit data for the corresponding DNA of GMI16-004. In testing the *Klebsiella* strains, GMI16-005 and GMI16-006, the following laboratories participated #77, #82, #93, #104, and #114 submitting data for both the bacterial culture and corresponding.

The quality metrics and parameters of GMI16-001-BACT, GMI16-002-BACT, GMI16-001-DNA, and GMI16-002-DNA from laboratory #114 were excluded the analysis due to mixing up the two strains.

Initially, the quality markers were evaluated for potential contamination which revealed that all genomes were of only one species.

The medians of the number of reads mapped to the reference DNA sequences was somehow consistent between the three species with a tendency of having higher medians for the DNA than the BACT samples (Figure 1). The 25% upper and lower quartiles ranged largely for the *Listeria* and *Klebsiella* species compared to *Campylobacter*. In general, two laboratories, #93 and #115 were determined outliers with a high number of reads mapped to the reference DNA sequences for all species compared to the other laboratories (Figure 1). This can be explained by the used sequencing platform e.g. Hiseq 2000 and Hiseq 2500 in contrast to Miseq being used by all other laboratories. In addition, these platforms often provide more reads and of shorter length e.g. 100bp as indicated by laboratory #93. The lowest observed values were of laboratory #104 and #114 (Figure 1).

The proportion of reads produced which map directly to the closed genome of the same strain should not exceed more than 100% indicating an error e.g. contaminations. The medians of the proportion of reads produced which map directly to the closed genome were almost 100% for both the DNA and the culture of the two *Campylobacter* genomes. A very little range of the 25% upper and lower quartiles were observed. This indicated that all laboratories performed equally well with the exception of laboratory #115. For GMI16-001-BACT, laboratory #115 only had 56.2% of reads mapped to the reference DNA sequence (Figure 2). The median of the proportion of reads produced which map directly to the closed genome of the *Listeria* GMI16-003-BACT were as well almost 100% with a tight upper and lower quartile centered around. In contrast, for the GMI16-003-DNA, the upper and lower quartile were much larger but still close to 100% indicating that all reads produced map to the reference. Lower proportions of reads produced which map directly to the closed genome were observed for *Listeria* GMI16-004 with a median of close to 89% for BACT and 86% for DNA. The reason for this might be due to a high number of plamids. The medians of the proportions of reads produced which map directly to the closed genome for *Klebsiella* GMI16-005 and GMI16-005 were about 95% or greater, especially for GMI16-005 which indicates a nice fit of the reads to the reference. In general, the proportions of

reads produced which map directly to the closed genome were lower for the laboratories, #93, #104, and #115 than the others participating (Figure 2).

The total number of contigs assembled should ideally be less than 1000 indicating good quality – the lower the better. For *Campylobacter* GMI16-001 and *Campylobacter* GMI16-002, the medians are between 100 and 150 contigs with a tight 25% quartile fit except for GMI16-001-BACT and GMI16-002-DNA where laboratory #104 produced 459 and 2.131 contigs, respectively and being considered an outlier. The 25% quartiles are much broader for *Listeria* GMI16-003 and GMI16-004 with an overall median less than 250 contigs indicating some unexpected difficulties sequencing *Listeria*. Similarly, the *Klebsiella* genomes of GMI16-005 and GMI16-006 revealed medians below the same number of 200 contigs. For the DNA samples, the 25% quartiles are really tight compared to the BACT, indicating that the problems can be related to the DNA purification step of the bacteria. This seems to be a general observation (Figure 3).

The size of the assembled genomes was observed to match the expected size of the species with *Campylobacter* being around 2mb, *Listeria* about 3mb, and *Klebsiella* of around 5mb. For the *Campylobacter* genomes, laboratory #104 was considered an outlier with a size of 2.077.671bp (107.92%) for GMI16-001-BACT and 3.260.167bp (171.72%) for GMI16-002-DNA (Figure 4-5). For *Listeria*, the size the assembled genomes as well as the proportion of the size to the reference DNA sequences were much broader with larger 25% quartiles especially for the DNA samples. In contrast, the proportion of the size to the reference DNA sequences of *Klebsiella* GMI16-005 were in average 99.3% with an outlier of 111.67% (laboratory #104) (Table 4). Larger 25% quartiles were observed for *Klebsiella* GMI16-006 but still close to the expected size of the species genome and with an almost 100% in proportion to the reference DNA sequence (Table 3-4).

The N50 length is defined as the length for which the collection of all contigs of that length or longer contains at least half of the sum of the lengths of all contigs, and for which the collection of all contigs of that length or shorter also contains at least half of the sum of the lengths of all contigs. A N50 more than 15000 normally indicate good quality and were obtained by all laboratories for all of the genomes. The lowest N50 value observed was 83.000 and by Laboratory #104 (Figure 6).

The depth (bp) of the coverage is calculated based on the number of bps sequenced divided by the total size (both chromosome and plasmids) of the closed genome (same strain). This number can be rounded to the nearest integer. In essence this number describes the number of times the sequenced bps covers the reference DNA and is often ended with an “x” (e.g. 30x) which also serve as a good average number in depth. All of the laboratories for all genomes were observed to have an overall depth of between 50X to 100X which is ideal (Figure 7).

4. Discussion

The majority of the submitted MLST data were correct and in line with the expected value. The results of MLST analysis revealed a systematic error for participant #114 when submitting the data causing a mix up of the test genomes for the MLST and resistance genes prediction. However, the MLST was correct for all PT strains when re-analysed using the CGE reference method.

Most of the submitted AMR genes were in concordance with the expected results. Some deviations however were observed mostly due to the tools, own or CGE reference lower threshold setting ignoring genes with a lower homology.

One of the objectives for the ENGAGE PT was to assess a range of quality markers to evaluate the performance by the consortium partners. Overall, the PT test show that all laboratories perform satisfactory with the exception of laboratory #104 which in general produced a low number of reads, a lower percentage of mapping reads to the references, a high number of contigs, a high size of the assembly, and a high proportion in the size of the assembly per reference sequence. A few other laboratories could benefit from an assessment of own sequencing quality including laboratory #114. It is noteworthy to mention that the lower quality of the sequeces produced by laboratory #104 did not affect the prediction of MLSTs nor resistance genes.

Laboratory #104 have indicated that they tried to select standard parameters of sequencing (routinely used) with a depth oscillates of about 30x which can affect the results of laboratory #104. In addition, the laboratory submitted trimmed sequences as indicated in the protocol but failed to remove adapters which normally are removed by the platform itself. This might have affected the quality of the sequences as the PT organizers didn't enhance any of the submitted data. The PT organizers offered the laboratory #104 to re-submit data with removed adapters but this was not possible due to the timeline and deliverable of this report.

5. Conclusions

The pilot PT was a useful exercise as it allowed ENGAGE consortium partners to assess the quality of own data as well as to identify critical points for improvement. In general, all data were satisfactory but the PT organizer encourage especially laboratory #104 to upload data which has removed adapters as well as ensuring the genomes being matched to the identical reference genome to avoid wrong prediction of the MLST and resistance genes.

Reference List

1. **Larsen, M. V., S. Cosentino, S. Rasmussen, C. Friis, H. Hasman, R. L. Marvig, L. Jelsbak, T. Sicheritz-Ponten, D. W. Ussery, F. M. Aarestrup, and O. Lund**. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J.Clin.Microbiol. **50(4)**:1355-1361.

2. **Moran-Gilad, J., V. Sintchenko, S. K. Pedersen, W. J. Wolfgang, J. Pettengill, E. Strain, and R. S. Hendriksen**. 2015. Proficiency testing for bacterial whole genome sequencing: an end-user survey of current capabilities, requirements and priorities. BMC.Infect.Dis. **15**:174-0902.

3. **Zankari, E., H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F. M. Aarestrup, and M. V. Larsen**. 2012. Identification of acquired antimicrobial resistance genes. J.Antimicrob.Chemother. **67**:2640-2644.

THE ENGAGE PROFICIENCY TEST REPORT 2016
TABLES AND FIGURES

Table 1: The low and high range of DNA concentration (ng/µl) measured for both the bacterial cultures and DNA received.

	Low range (ng/µl)	High range (ng/µl)
GMI16-001-BACT *(Campylobacter)*	0.26	64.8
GMI16-002-BACT *(Campylobacter)*	0.27	80
GMI16-001-DNA *(Campylobacter)*	0.27	647.39
GMI16-002-DNA *(Campylobacter)*	0.22	90.89
GMI16-003-BACT *(Listeria)*	0.28	92.05
GMI16-004-BACT *(Listeria)*	0.28	68.7
GMI16-003-DNA *(Listeria)*	0.25	33.52
GMI16-004-DNA *(Listeria)*	0.24	23.77
GMI16-005-BACT *(Klebsiella)*	0.18	34.3
GMI16-006-BACT *(Klebsiella)*	2.2	28.6
GMI16-005-DNA *(Klebsiella)*	0.25	52.18
GMI16-006-DNA *(Klebsiella)*	0.24	58.3

Table 2: The low and high range of total DNA amount (µg) measured for both the bacterial cultures and DNA received.

	Low range (µg)	High range (µg)
GMI16-001-BACT *(Campylobacter)*	0.001*	3.88
GMI16-002-BACT *(Campylobacter)*	0.001*	4.8
GMI16-001-DNA *(Campylobacter)*	0.001*	2.28
GMI16-002-DNA *(Campylobacter)*	0.001*	3.12
GMI16-003-BACT *(Listeria)*	0.001*	4.52
GMI16-004-BACT *(Listeria)*	0.001*	4.88
GMI16-003-DNA *(Listeria)*	0.001*	3.11
GMI16-004-DNA *(Listeria)*	0.001*	1.79
GMI16-005-BACT *(Klebsiella)*	0.001*	1
GMI16-006-BACT *(Klebsiella)*	0.001*	1.43
GMI16-005-DNA *(Klebsiella)*	0.001*	3.43
GMI16-006-DNA *(Klebsiella)*	0.001*	2.45

* All values from #77

Table 3: The low and high range of the measured DNA quality (e.g. RIN or 260/280 ratio) for both the bacterial cultures and DNA received.

	Low range	High range
GMI16-001-BACT *(Campylobacter)*	1.47	3.28
GMI16-002-BACT *(Campylobacter)*	1.84	2.54
GMI16-001-DNA *(Campylobacter)*	2	3.55
GMI16-002-DNA *(Campylobacter)*	1.75	2.26
GMI16-003-BACT *(Listeria)*	1.7	1.93
GMI16-004-BACT *(Listeria)*	1.77	1.91
GMI16-003-DNA *(Listeria)*	1.82	1.92
GMI16-004-DNA *(Listeria)*	1.78	1.86
GMI16-005-BACT *(Klebsiella)*	1.85	11.2
GMI16-006-BACT *(Klebsiella)*	1.73	10.2
GMI16-005-DNA *(Klebsiella)*	1.72	12.1
GMI16-006-DNA *(Klebsiella)*	1.78	13

Table 4: The low and high range of the measured DNA quality (260/230 ratio) for both the bacterial cultures and DNA received.

	Low range	High range
GMI16-001-BACT *(Campylobacter)*	0.79	2.4
GMI16-002-BACT *(Campylobacter)*	1.71	2.24
GMI16-001-DNA *(Campylobacter)*	0.54	2.3
GMI16-002-DNA *(Campylobacter)*	0.21	1.62
GMI16-003-BACT *(Listeria)*	1.31	1.73
GMI16-004-BACT *(Listeria)*	1.35	1.95
GMI16-003-DNA *(Listeria)*	0.32	1.75
GMI16-004-DNA *(Listeria)*	0.2	1.83
GMI16-005-BACT *(Klebsiella)*	1.36	1.92
GMI16-006-BACT *(Klebsiella)*	1.22	1.76
GMI16-005-DNA *(Klebsiella)*	0.24	1.76
GMI16-006-DNA *(Klebsiella)*	0.32	1.6

Table 5: Determined MLST for both the bacterial culture and DNA received.

			GMI16-001		GMI16-002		GMI16-003		GMI16-004		GMI16-005		GMI16-006	
	Participant		Expected MLST	Obtained MLST	Expected MLST	Obtained MLST	Expected MLST	Obtained MLST	Expected MLST	Obtained MLST	Expected MLST	Obtained MLST	Expected MLST	Obtained MLST
BACT	#77	Own tool	ST-7426	7426	ST-6238	6238	ST-2	2	ST-121	121	ST-512	512	ST-15	15
		CGE tool		7426		6238		2		121		512		15
	#82	Own tool		7426		6238		2		121		512		15
		CGE tool		7426		6238		2		121		512		15
	#93	Own tool				6238						512		15
		CGE tool		7426		6238						512		15
	#95	Own tool												
		CGE tool		7426		6238								
	#104	Own tool		7426		6238		2		121		512		15
		CGE tool		7426		6238		2		121		512		15
	#114*	Own tool												
		CGE tool		**6238**		**7426**						512		
	#115	Own tool		7426		6238		2		121				
		CGE tool		7426		6238		2		121				
DNA	#77	Own tool	ST 7426	7426	ST-6238	6238	ST-2	2	ST-121	121	ST-512	512	ST-15	15
		CGE tool		7426		6238		2		121		512		15
	#82	Own tool		7426		6238		2		121		512		15
		CGE tool		7426		6238		2		121		512		15
	#93	Own tool				6238						512		15
		CGE tool		7426		6238						512		15
	#95	Own tool												
		CGE tool		7426		6238								
	#104	Own tool		7426		6238		2		121		512		
		CGE tool		7426		6238		2		121		512		15
	#114*	Own tool												
		CGE tool		**Unk**		**Unk**								15
	#115	Own tool		7426		6238		2		121				
		CGE tool		7426		6238		2		121				

* Laboratory #114 mixed up the two strains GMI16-001-BACT and GMI16-002-BACT as well as GMI16-001-DNA and GMI16-002-DNA why the incorrect MLSTs. Some laboratories did not report MLST data (own tool) but these were provided by PT-organizer (CGE tool) marked in light gray. Deviating results indicated in bold.

Table 6: Determined antimicrobial resistance genes in *Campylobacter* GMI16-002 for both the bacterial culture and DNA received

	Participant		GMI16-002																				
BACT	#77	Own tool	aadE	aph(3')-III	aph(2'')-like	tet(O)-like																	
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#82	Own tool	aadE	aph(3')-III	aph(2'')-like	tet(O)-like	gyrA T86I (Quinolone)	23S A2075G (Macrolide)															
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#93	Own tool	aadE	aph(3')-III	aph(2'')-like	tet(O)-like																	
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#95	Own tool																					
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#104	Own tool	aadE	aph(3')-III		tet(O)-like																	
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#114	Own tool																					
		CGE tool																					
	#115	Own tool																					
		CGE tool	aadE	aph(3')-III		tet(O)																	
DNA	#77	Own tool	aadE	aph(3')-III	aph(2'')	tet(O)																	
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#82	Own tool	aadE	aph(3')-III	aph(2'')-If-like	tet(O)-like	gyrA T86I (Quinolone)	23S A2075G (Macrolide)															
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#93	Own tool	aadE	aph(3')-III	aph(2'')-like	tet(O)-like																	
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#95	Own tool																					
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#104	Own tool	aadE	aph(3')-III		tet(O)-like																	
		CGE tool	aadE	aph(3')-III		tet(O)																	
	#114*	Own tool																					
		CGE tool					**aac(6')-IIc**	**blaSHV-12**	**blaCTX-M-15**	**blaTEM-1B**	**dfrA18**	**strB**	**strA**	**ere(A)**	**tet(A)**	**tet(D)**	**sul2**	**sul1**	**QnrB49**	**floR**			
	#115	Own tool																					
		CGE tool	aadE	aph(3')-III		tet(O)																	

* Laboratory #114 mixed up the two strains GMI16-001-BACT and GMI16-002-BACT why the incorrect AMR profile. The expected AMR profile for GMI16-002-BACT was reported for the pan-susceptible GMI16-001-BACT. Similarly, the laboratory #114 mixed up the GMI16-001-DNA and GMI16-002-DNA why the incorrect AMR profile for GMI16-002-DNA. Some laboratories did not report AMR data (own tool) but these were provided by PT-organizer (CGE tool) marked in light gray. Deviating results indicated in bold.

Table 7: Determined antimicrobial resistance genes in *Klebsiella* GMI16-005 for both the bacterial culture and DNA received

	Participant		GMI16-005													
BACT	#77	Own tool	aadA2	aac(6')-Ib)	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr	fosA	mph(A)	catA1	sul1	dfrA12
		CGE tool	aadA2	aac(6')-Ib	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr		mph(A)	catA1	sul1	dfrA12
	#82	Own tool	aadA2-like	aac(6')-Ib	blaTEM-1A-like	blaKPC-3	blaOXA-9-like	blaSHV-11	oqxA	oqxB		fosA-like	mph(A)	catA1-like	sul1	dfrA12
		CGE tool	aadA2	aac(6')-Ib	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr		mph(A)	catA1	sul1	dfrA12
	#93	Own tool	aadA2	aac(6')-Ib	blaTEM-1A-like	blaKPC-3	blaOXA-9-like	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr-like	fosA-like	mph(A)	catA1-like	sul1	dfrA12
		CGE tool	aadA2	aac(6')-Ib	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr		mph(A)	catA1	sul1	dfrA12
	#95	Own tool														
		CGE tool														
	#104	Own tool	aadA2-like	aac(6')-Ib	blaTEM-1A-like	blaKPC-3	blaOXA-9-like	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr-like	fosA-like	mph(A)	catA1-like	sul1	dfrA12
		CGE tool	aadA2	aac(6')-Ib	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr		mph(A)	catA1	sul1	dfrA12
	#114	Own tool														
		CGE tool	aadA2	aac(6')-Ib	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr		mph(A)	catA1	sul1	dfrA12
	#115	Own tool														
		CGE tool														
DNA	#77	Own tool	aadA2	aac(6')-Ib)	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr	fosA	mph(A)	catA1	sul1	dfrA12
		CGE tool	aadA2	aac(6')-Ib	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr		mph(A)	catA1	sul1	dfrA12
	#82	Own tool	aadA2-like	aac(6')-Ib	blaTEM-1A-like	blaKPC-3	blaOXA-9-like	blaSHV-11	oqxA	oqxB		fosA-like	mph(A)	catA1-like	sul1	dfrA12
		CGE tool	aadA2	aac(6')-Ib	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr		mph(A)	catA1	sul1	dfrA12
	#93	Own tool	aadA2	aac(6')-Ib	blaTEM-1A-like	blaKPC-3	blaOXA-9-like	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr-like	fosA-like	mph(A)	catA1-like	sul1	dfrA12
		CGE tool	aadA2	aac(6')-Ib	blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr		mph(A)	catA1	sul1	dfrA12
	#95	Own tool														
		CGE tool														
	#104	Own tool	aadA2-like	aac(6')-Ib	blaTEM-1A-like	blaKPC-3	blaOXA-9-like	blaSHV-11	oqxA	oqxB	aac(6')Ib-cr-like	fosA-like	mph(A)	catA1-like	sul1	dfrA12
		CGE tool	aadA2		blaTEM-1A	blaKPC-3	blaOXA-9	blaSHV-11	oqxA	oqxB			mph(A)	catA1	sul1	dfrA12
	#114	Own tool														
		CGE tool														
	#115	Own tool														
		CGE tool														

Data for the CGE tool were provided by PT-organizer and marked in light gray. Deviating results indicated in bold.

Table 8: Determined antimicrobial resistance genes in *Klebsiella* GMI16-006 for both the bacterial culture and DNA received

	Participant		GMI16-006																					
BACT	#77	Own tool	aadA1	aac(6')-Ib	aac(3)-Iid	aph(3')-Via	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTem-1b		QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14	fosA		blaSHV-1	
		CGE tool	aadA1	aac(6')-Ib	aac(3)-IId	aph(3')-VIa	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTEM-1B		QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14			blaSHV-1	
	#82	Own tool	aadA1	aac(6')-Ib	aac(3)-IId-like	aph(3')-VIa-like	strA-like	strB-like	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTEM-1B-like		QnrS1	oqxA-like	oqxB-like		sul2	tet(D)	dfrA14-like	fosA-like	blaSHV-28		
		CGE tool	aadA1	aac(6')-Ib	aac(3)-IId	aph(3')-VIa	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15		blaTEM-1A	QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14		blaSHV-28		
	#93	Own tool	aadA1	aac(6')-Ib		aph(3')VIa-like	strA-like	strB-like	blaNDM-1	blaOXA-9-like	blaCTX-M-15	blaTEM-1B		QnrS1	oqxA-like	oqxB-like	aac(6')Ib-cr-like	sul2	tet(D)	dfrA14-like	fosA-like		blaSHV-1	
		CGE tool	aadA1	aac(6')-Ib	aac(3)-IId	aph(3')-VIa	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTEM-1B		QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14				
	#95	Own tool																						
		CGE tool																						
	#104	Own tool	aadA1	aac(6')-Ib	aac(3)-IId-like	aph(3')-VIa-like	strA-like	strB-like	blaNDM-1	blaOXA-9-like	blaCTX-M-15	blaTEM-1B		QnrS1	oqxB-like	oqxA-like	aac(6')Ib-cr-like	sul2	tet(D)	dfrA14-like,	fosA	blaSHV-28		
		CGE tool	aadA1	aac(6')-Ib	aac(3)-IId	aph(3')-VIa	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15		blaTEM-1A	QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14				
	#114	Own tool																						
		CGE tool																						
	#115	Own tool																						
		CGE tool																						
DNA	#77	Own tool	aadA1	aac(6')-Ib	aac(3)-Iid	aph(3')-Via	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTem-1b		QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14	fosA		blaSHV-1	
		CGE tool	aadA1	aac(6')-Ib	aac(3)-IId	aph(3')-VIa	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTEM-1B		QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14				
	#82	Own tool	aadA1	aac(6')-Ib	aac(3)-IId-like	aph(3')-VIa-like	strA-like	strB-like	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTEM-1B-like		QnrS1	oqxA-like	oqxB-like		sul2	tet(D)	dfrA14-like	fosA-like	blaSHV-28		
		CGE tool	aadA1	aac(6')-Ib	aac(3)-IId	aph(3')-VIa	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTEM-1B		QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14		blaSHV-28		
	#93	Own tool	aadA1	aac(6')-Ib	aac(3)-IId-like	aph(3')VIa-like	strA-like	strB-like	blaNDM-1	blaOXA-9-like	blaCTX-M-15	blaTEM-1B		QnrS1	oqxA-like	oqxB-like	aac(6')Ib-cr-like	sul2	tet(D)	dfrA14-like	fosA-like		blaSHV-1	blaLEN12-like
		CGE tool	aadA1	aac(6')-Ib	aac(3)-IId	aph(3')-VIa	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTEM-1B		QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14				
	#95	Own tool																						
		CGE tool																						
	#104	Own tool	aadA1	aac(6')-Ib	aac(3)-IId-like	aph(3')-VIa-like	strA-like	strB-like	blaNDM-1	blaOXA-9-like	blaCTX-M-15	blaTEM-1B		QnrS1	oqxB-like	oqxA-like	aac(6')Ib-cr-like	sul2	tet(D)	dfrA14-like,	fosA	blaSHV-28		
		CGE tool	aadA1	aac(6')-Ib	aac(3)-IId-like	aph(3')-VIa-like	strA-like	strB-like	blaNDM-1	blaOXA-9-like	blaCTX-M-15	blaTEM-1B		QnrS1	oqxB-like	oqxA-like	aac(6')Ib-cr-like	sul2	tet(D)	dfrA14-like,	fosA	blaSHV-28		
	#114	Own tool																						
		CGE tool	aadA1	aac(6')-Ib	aac(3)-IId	aph(3')-VIa	strA	strB	blaNDM-1	blaOXA-9	blaCTX-M-15	blaTEM-1B		QnrS1	oqxB	oqxA	aac(6')Ib-cr	sul2	tet(D)	dfrA14				
	#115	Own tool																						
		CGE tool																						

Data for the CGE tool were provided by PT-organizer and marked in light gray. Deviating results indicated in bold.

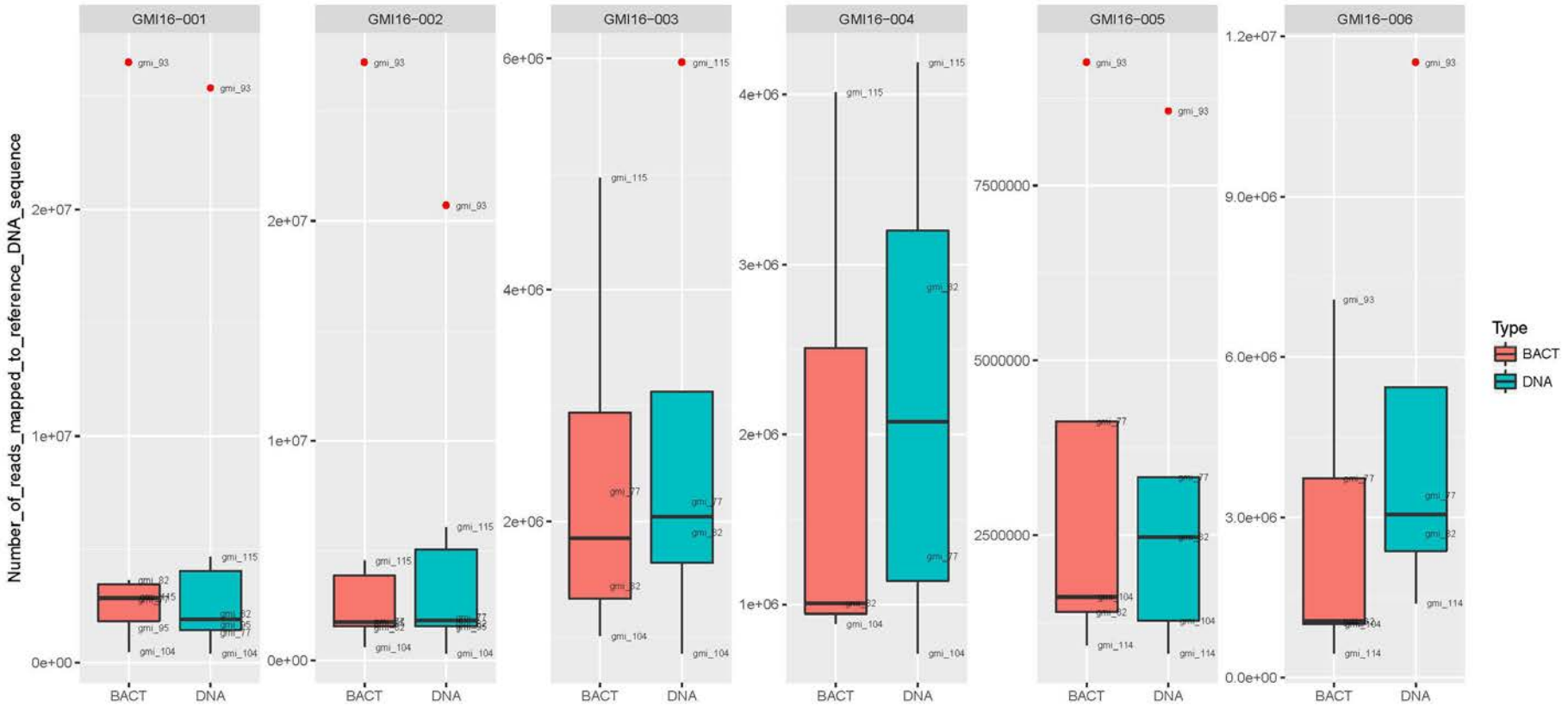

Figure 1: Number (No) of reads mapped to the reference sequence.

The black line inserted the box represent the median which indicates 50% of the data being greater than this value. The top and botton of the box indicate the upper and lower quartiles which is 25% of the data being greater or lower than this value. The end of the whiskers indicates the maximum (greatest value) or minimum (lowest value) excluding outliers. The red dots represent values considered outliers.

Results for participant 2 omitted for both sample types of strain 1, 4, 5, and 6. The whiskers represent minimum and maximum values (range) and the box represent the Q1, Median, and Q3, respectively.

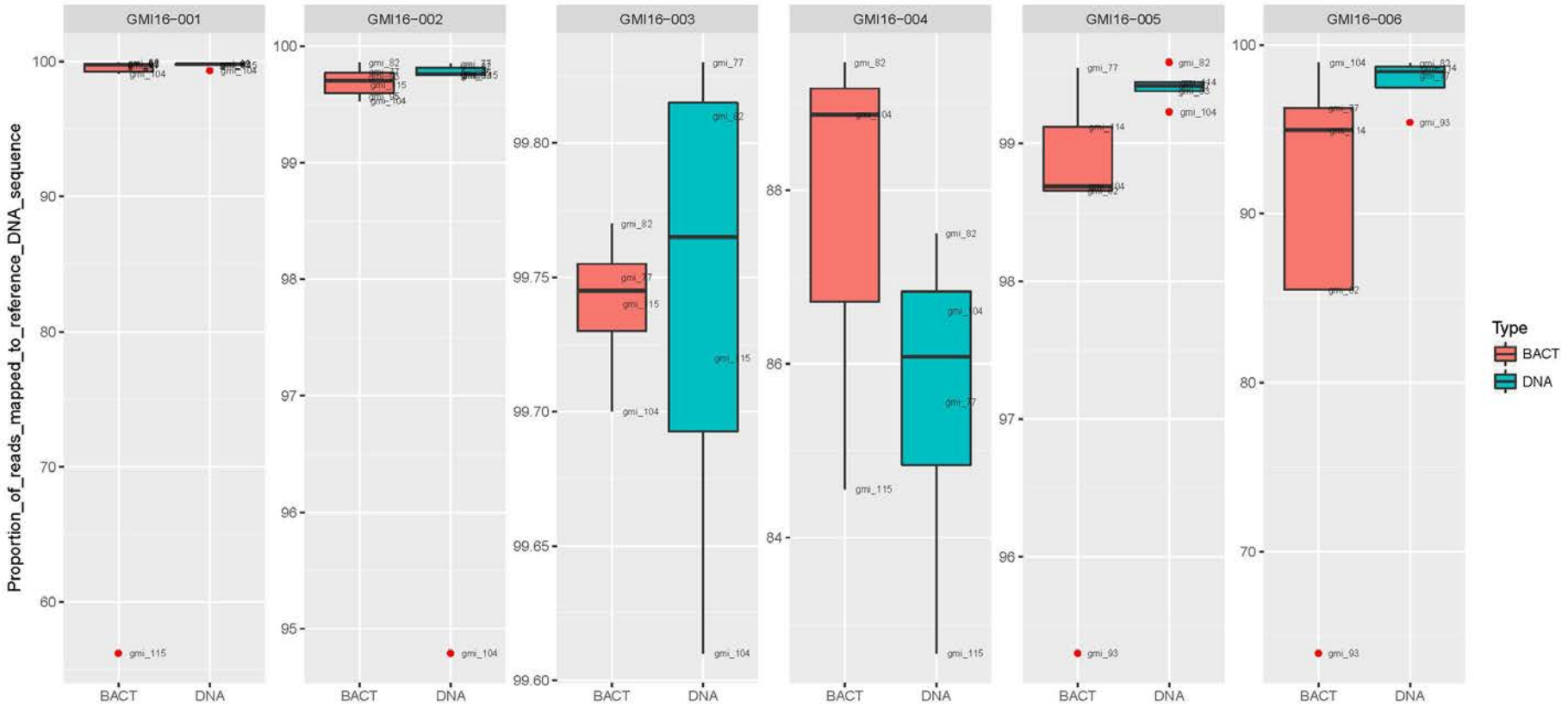

Figure 2: Proportion (%) of reads mapped to the reference DNA sequence.

The black line inserted the box represent the median which indicates 50% of the data being greater than this value. The top and botton of the box indicate the upper and lower quartiles which is 25% of the data being greater or lower than this value. The end of the whiskers indicates the maximum (greatest value) or minimum (lowest value) excluding outliers. The red dots represent values considered outliers.

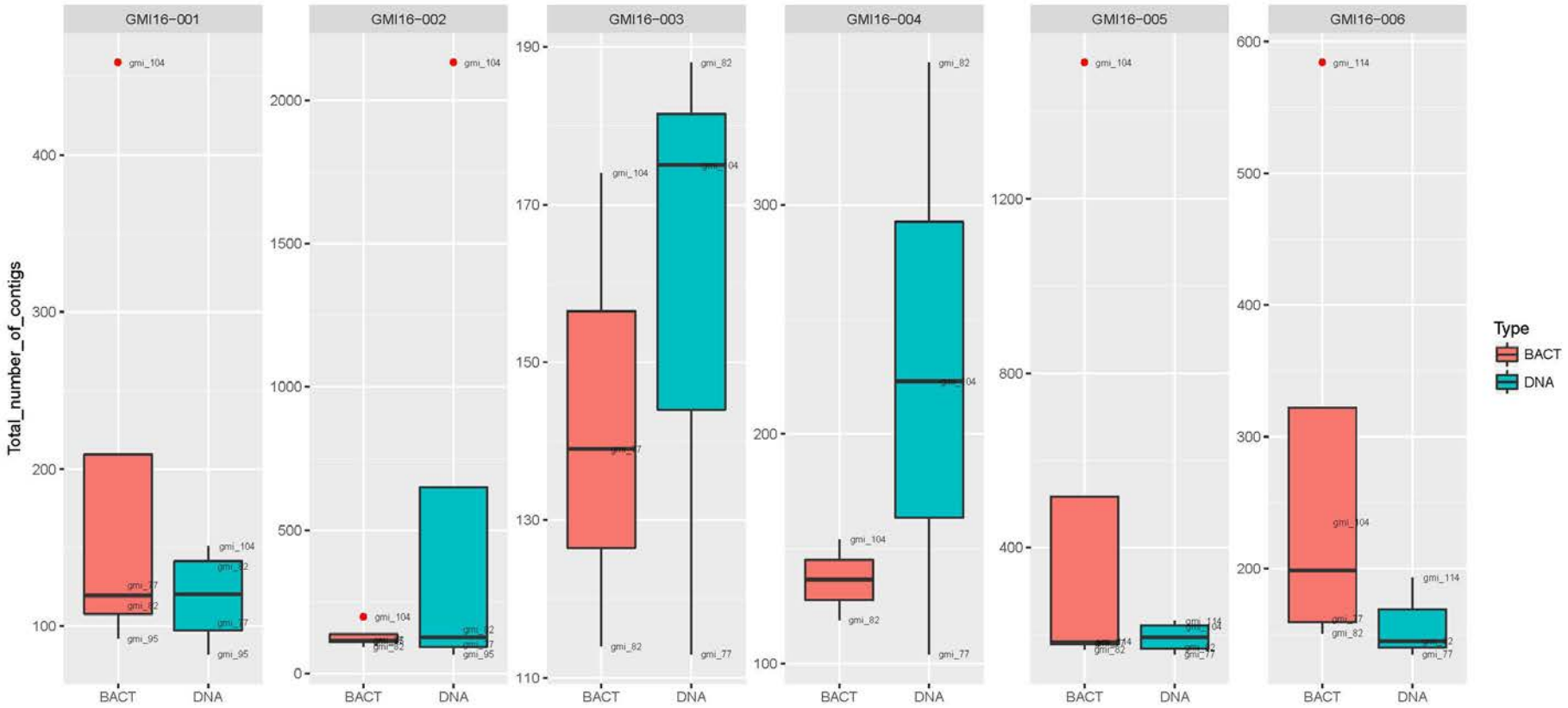

Figure 3: Total number (No) of contigs

The black line inserted the box represent the median which indicates 50% of the data being greater than this value. The top and botton of the box indicate the upper and lower quartiles which is 25% of the data being greater or lower than this value. The end of the whiskers indicates the maximum (greatest value) or minimum (lowest value) excluding outliers. The red dots represent values considered outliers.

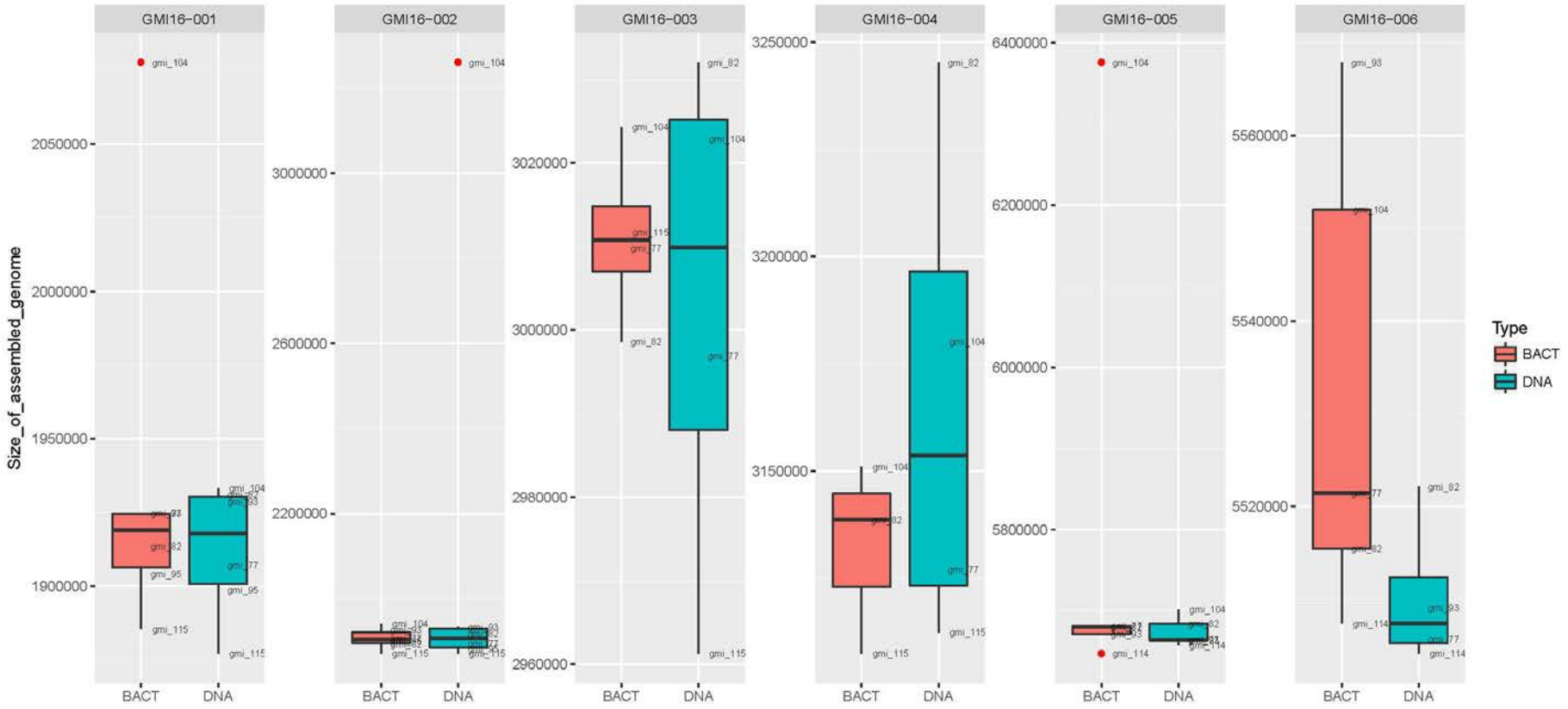

Figure 4: Size (bp) of the assembled genome

The black line inserted the box represent the median which indicates 50% of the data being greater than this value. The top and botton of the box indicate the upper and lower quartiles which is 25% of the data being greater or lower than this value. The end of the whiskers indicates the maximum (greatest value) or minimum (lowest value) excluding outliers. The red dots represent values considered outliers.

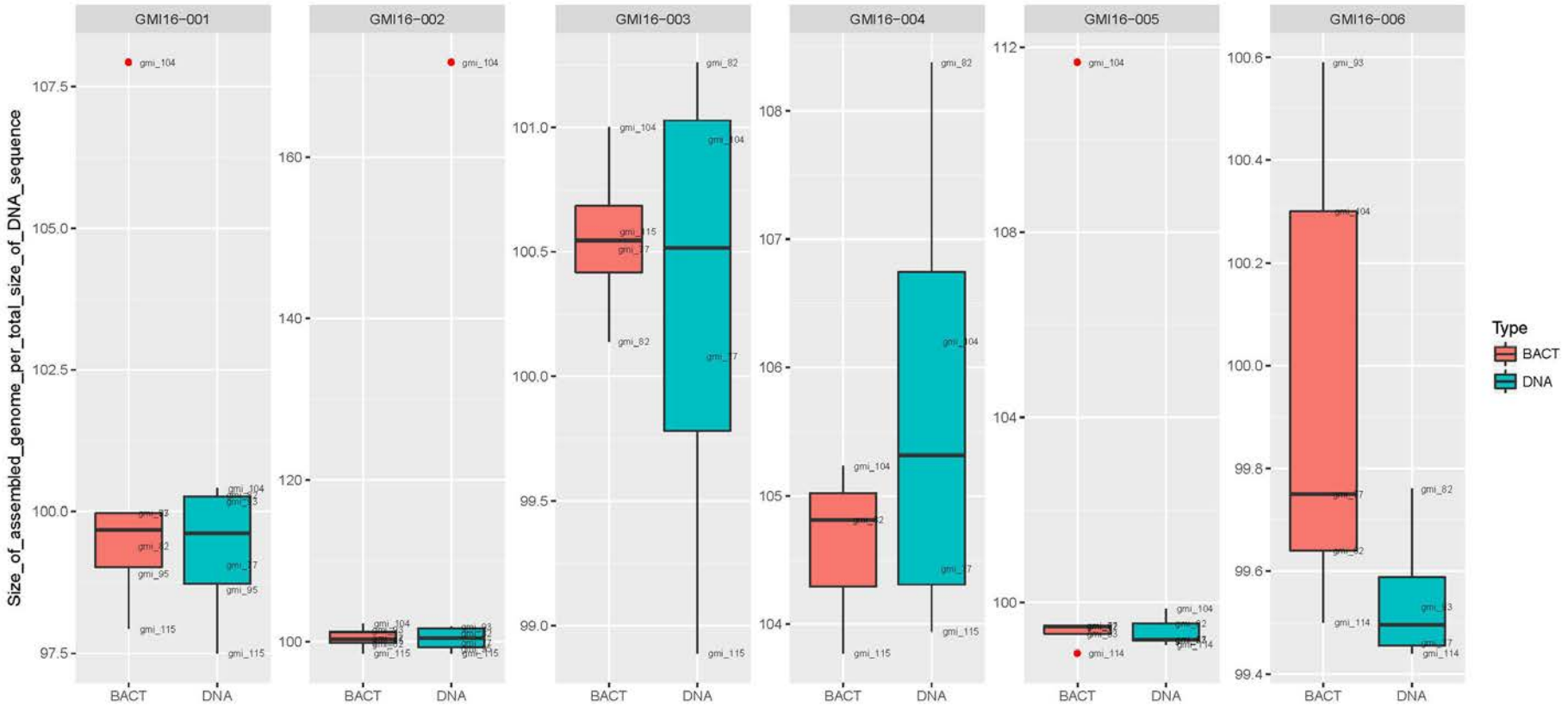

Figure 5: Proportion (%) of the assembled genome per reference DNA sequence

The black line inserted the box represent the median which indicates 50% of the data being greater than this value. The top and botton of the box indicate the upper and lower quartiles which is 25% of the data being greater or lower than this value. The end of the whiskers indicates the maximum (greatest value) or minimum (lowest value) excluding outliers. The red dots represent values considered outliers.

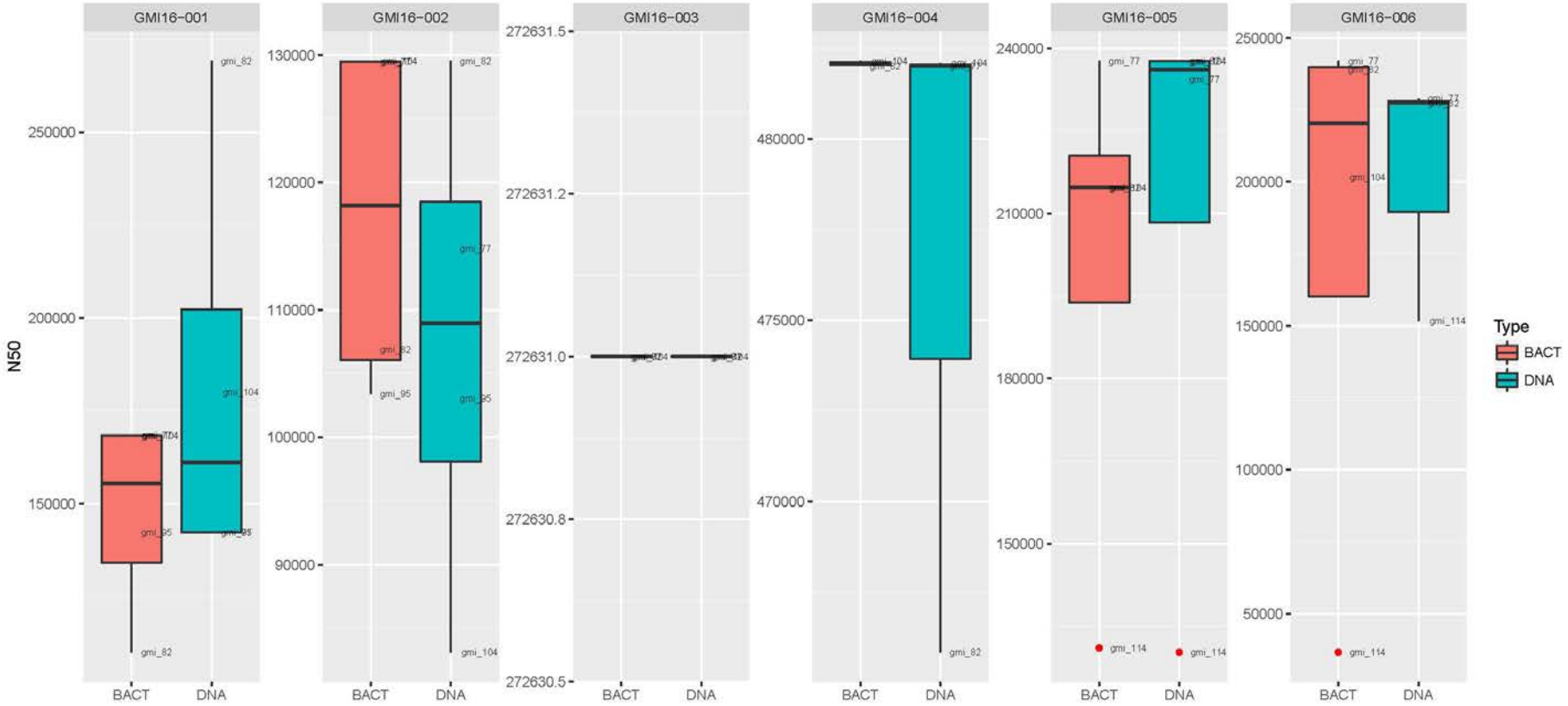

Figure 6: N50 - average length (bp) of sequences

The black line inserted the box represent the median which indicates 50% of the data being greater than this value. The top and botton of the box indicate the upper and lower quartiles which is 25% of the data being greater or lower than this value. The end of the whiskers indicates the maximum (greatest value) or minimum (lowest value) excluding outliers. The red dots represent values considered outliers.

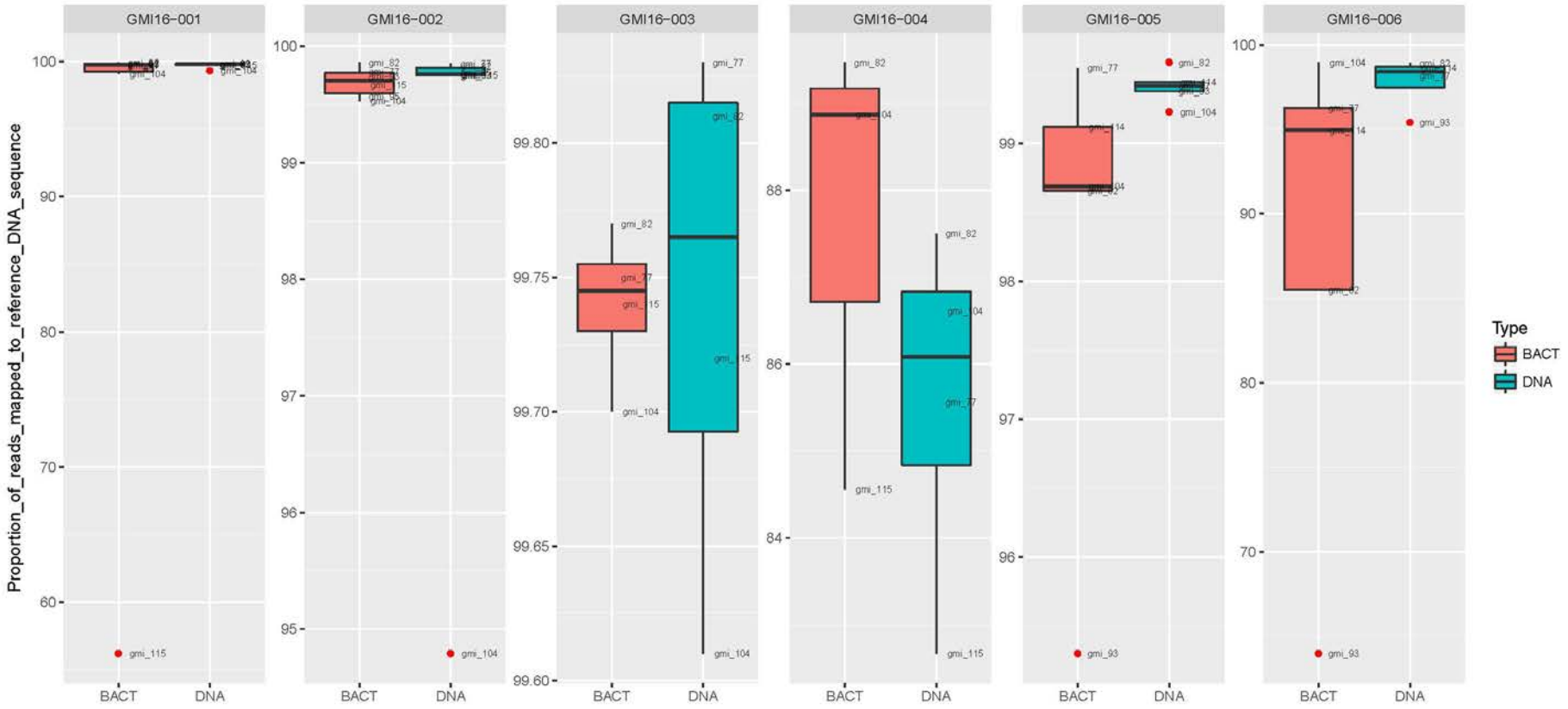

Figure 7: Depth of coverage (X) of the sequences

The black line inserted the box represent the median which indicates 50% of the data being greater than this value. The top and botton of the box indicate the upper and lower quartiles which is 25% of the data being greater or lower than this value. The end of the whiskers indicates the maximum (greatest value) or minimum (lowest value) excluding outliers. The red dots represent values considered outliers.