Evaluation of Nanopore sequencing technology to differentiate *Salmonella* serotypes and serotype variants with the same or closely related antigenic formulae

Feng Xu[1,*], Chongtao Ge[1,*], Shaoting Li[2], Silin Tang[1], Xingwen Wu[1], Hao Luo[1], Xiangyu Deng[2], Guangtao Zhang[1], Abigail Stevenson[1], Robert C. Baker[1]

[1] Mars Global Food Safety Center, Beijing, 101407, China

[2] Center for Food Safety, University of Georgia, Griffin, GA, 30223, USA

*Corresponding authors: feng.y.xu@effem.com; chongtao.ge@effem.com

Highlights

- *Salmonella* serotypes or serotype variants with the same antigenic formula were differentiated by SNP typing.
- Nanopore sequencing followed by phage prediction identified the *Salmonella* serotype variants caused by phage conversion.
- The latest ONT technology is capable of high fidelity SNP typing of *Salmonella*.

ABSTRACT:

Our previous study demonstrated that whole genome sequencing (WGS) data generated by Oxford Nanopore Technologies (ONT) can be used for rapid and accurate prediction of *Salmnonella* serotypes. However, one limitation is that established methods for WGS-based serotype prediction cannot differentiate certain serotypes and serotype variants with the same or closely related antigenic formulae. This study aimed to evaluate Nanopore sequencing and corresponding data analysis for differentiation of these serotypes and serotype variants, thus overcoming this limitation. Five workflows that combined different flow cells, library construction methods and basecaller models were evaluated and compared. The workflow that consisted of the R9 flow cell, rapid sequencing library construction kit and guppy basecaller with base modified model performed best for Single Nucleotide Polymorphism (SNP) analysis. With this workflow, as high as 99.98% matched the identity of the assembled genomes and only less than five high quality SNPs (hqSNPs) between ONT and Illumina sequencing data were achieved. SNP typing allowed differentiation of Choleraesuis*sensu stricto*, Choleraesuis var. Kunzendorf, Choleraesuis var. Decatur, Paratyphi C, and Typhisuis that share the same antigenic formula 6,7:c:1,5. Prophage prediction further distinguished Orion var. 15^{+} and Orion var. 15^{+}, 34^{+}. Our study improves the readiness of ONT as a *Salmonella* subtyping and source tracking tool for food industry applications.

Keywords: *Salmonella*, Oxford Nanopore sequencing, phylogenetic analysis, SNP, variants

1. Introduction

Salmonella remains an important concern in the food industry since this foodborne pathogen continues to cause global public health and economic impact. Thus, it is imperative to reinforce *Salmonella* control measures in the food industry (GMA, 2009). Rapid methods for subtyping strains beyond the species level to serotype are essential to facilitate contamination incident investigations (Olaimat and Holley, 2012; Shi et al., 2015). There are more than 2,600 *Salmonella* serotypes currently described in the White-Kauffmann-Le Minor scheme (Grimont and Weill, 2007; Dieckmann and Malorny, 2011). Whole Genome Sequencing (WGS) based tools have now emerged as an alternative method for *Salmonella* serotyping (Zhang et al., 2015; Yoshida et al., 2016; Yachison et al., 2017; Ibrahim and Morin, 2018; Diep et al., 2019; Tang et al., 2019; Zhang et al., 2019; Uelze et al., 2020) and have been suggested to be the new gold standard (Banerji et al., 2020).

Oxford Nanopore Technologies (ONT) sequencing generates long-read sequences with rapid turnaround and has the potential for rapid identification of food-borne pathogens during routine monitoring and incident investigations in the food industry. Challenges in the basecalling accuracy of nanopore sequencing have been identified compared with short-read sequencing technologies (Amarasinghe et al., 2020). Errors caused by signal interpretation during basecalling are dominated by insertions and deletions (indels), especially in homopolymeric regions (Gargis et al, 2019). An average of 95% accuracy of raw read (Jain et al., 2017, 2018) and 1.06 errors per Kbp assembly (Lang et al., 2020) were recently reported.

In silico serotype prediction using ONT sequences was systematically evaluated in our previous study, and all the tested *Salmonella* isolates representing 34 serotypes were correctly predicted at the level of antigentic profile (Xu et al., 2020). Comparison with the conventional

serotyping method using antisera combined with biochemical tests, identified a limitation that the established WGS methods, with either Illumina or ONT data, were unable to differentiate serotypes that share the same or related antigenic formulae.

Serotypes with antigenic formula 6,7:c:1,5 including Paratyphi C (Vi + or Vi–), Typhisuis, Choleraesuis *sensu stricto*, Choleraesuis var. Kunzendorf and Choleraesuis var. Decatur, require additional biochemical tests to differentiate between them in conventional serotyping (Grimont and Weill, 2007). Choleraesuis is a serotype adapted to swine. This serotype has caused serious disease outbreaks in swine, including two in Denmark in 1999 - 2000 and 2012 - 2013 (Leekitcharoenphon et al., 2019). It also has a propensity to cause extraintestinal infections in humans (Chiu et al., 2004; Sirichote et al., 2010). Choleraesuis var. Kunzendorf is responsible for the majority of outbreaks among swine (Leekitcharoenphon et al., 2019). Serotype Orion has two closely related variants, Orion var. 15^{+} and Orion var. 15^{+}, 34^{+}. Phage conversion by $\square_{15}$ or $\square_{34}$ is needed to differentiate these two variants (Grimont and Weill, 2007). Orion has been isolated from a variety of hosts worldwide, including humans (Cabrera et al., 2006; Trafny et al., 2006), cattle (Alam et al., 2009), ducks (Rampersad et al., 2008), dogs (Gupta et al., 2016) and birds (Münch et al., 2012). Orion var.15^{+} and Orion var. 15^{+}, 34^{+} are more likely to be associated with poultry (Corry et al., 2002; McWhorter et al., 2015). Additional subtyping analysis is needed to differentiate these variants.

When the genetic determinant for a particular biotype or variant is known, it makes an effective target for futher differentiation. For example, SeqSero2 uses a Single Nucleotide Polymorphism (SNP) that inactivates tartrate fermentation to differentiate the *S.* Paratyphi B pathogen and a 7-bp deletion to identify O5- variants of *S.* Typhimurium (Zhang et al., 2019).

When such genetic determinants are not available, phylogenetic analysis, such as SNP typing and core/whole genome Multilocus Sequence Typing (MLST), can be used to differentiate biotypes and serotype. Nair et al. grouped the 6,7:c:1,5 biotypes of *S. enterica* subspecies *enterica* including Paratyphi C, Typhisuis, Choleraesuis *sensu stricto*, Cholerasuis var. Kunzendorf into four phyloclusters using the core SNP phylogeny (Nair et al., 2020). The authors also found that Choleraesuis var. Decatur genomes showed more SNP variation between them than was found across all of the remaining biotypes (Nair et al., 2020). With continuous improvement of the sequencing and basecalling accuracy of ONT, application of ONT sequences in SNP analysis has become more applicable (Greig et al., 2019; Taylor et al., 2019).

The goal of this study was to evaluate the combined use of latest improved ONT sequencing and modified subtyping analyses with primary serotype prediction in differentiating representative *Salmonella* serotypes or serotype variants that shared the same or related antigenic formulae.

2. Materials and methods

2.1. Bacterial strains

Two strains, *S. enterica* Choleraesuis var. Kunzendorf and *S. enterica* Orion var. 15^{+}, 34^{+}, were obtained from Dr. Martin Weidmann's laboratory (Cornell University, USA). Details of these strains can be found at www.foodmicrobetracker.com under the isolate IDs FSL R9-0095 and FSL R8-3858.

2.2. Genomic DNA extraction and quality control

Strains were incubated for 20 hrs on Trypticase Soy Agar (TSA) at 37°C. Genomic DNA was

extracted using a QIAamp DNA mini kit (Qiagen, Hilden, Germany), following the instructions provided by the manufacture. The double-stranded DNA (dsDNA) was quantified using a Qubit 3.0 fluorimeter (Life Technologies, Paisley, UK). The quality of DNA was assessed using a NanoDrop 1000 instrument (Thermo Fisher Scientific, Delaware, USA).

2.3. Illumina sequencing and genome assembly

Illumina sequences were performed at the Beijing Genomics Institute (BGI, Shenzhen, China). The library for sequencing was constructed by BGI and sequenced using a HiSeq X Ten platform (Illumina, San Diego, CA, USA) to 2 × 150 cycles. Genomes were assembled by the SOAPdenovo v1.05 (URL: http://soap.genomics.org.cn).

2.4. Oxford Nanopore sequencing and genome assembly

Five workflows with combinations of different flow cells, library construction methods and basecaller models were used in this study for comparison purposes (Supplementary Table 1). The DNA library was prepared with library construction kits following the manufacturer's instructions as indicated in each workflow. For flow cell, compared with commercial available R9 (version 9.4.1), R10 (version 10.0, early access) is a new design of nanopore, with a longer barrel and dual reader head. Libraries were sequenced with qualified R9 and R10 flow cells (active pores number >= 800) on a GridION sequencer (Oxford Nanopore Technologies, Oxford, UK). Real-time basecalling was performed using Guppy version 3.2.6 with the corresponding basecalling model as indicated in each workflow, which was integrated in the MinKNOW software v3.5.4. Fastq data were obtained for further analysis. Adaptor sequences were trimmed with Porechop v0.2.3 (URL:

https://github.com/rrwick/Porechop), and the quality of trimmed data was assessed using NanoStat v1.1.2 (De Coster et al., 2018). Filtlong v0.2.0 (URL: https://github.com/rrwick/Filtlong) was applied to remove reads shorter than 1,000 bp. Genomes were *de novo* assembled by Wtdbg2 v2.4 (Ruan and Li, 2019) with default parameters. The assembled genomes were corrected using Racon v1.3.3 (Vaser et al. 2017). Consensus was obtained from one round of Medaka version 0.8.0 (https://github.com/nanoporetech/medaka). QUAST v5.0.2 was applied to assess the assembled contigs (Gurevich et al., 2013). DNAdiff v1.3 (MUMmer version, Kurtz et al., 2004) was used to evaluate the base level comparison between ONT and Illumina assemblies.

2.5. Phylogenetic analysis

2.5.1. kSNP tree

A k-mer based approach, kSNP3 (Gardner et al., 2015) was used to assess the clustering of isolates sharing the same antigenic formula with FSL R9-0095 and isolates sharing closely related antigenic formulae with FSL R8-3858. Kchooser was used to determine an optimum k-mer size (Carroll et al., 2017). A parsimony tree was generated by kSNP3 using the set of core SNPs identified. Clade robustness was assessed using a bootstrap analysis with 1,000 replicates (Felsenstein, 1985). Twenty-six draft genome sequences were downloaded from GenBank and further confirmed by SeqSero2 (Zhang et al., 2019) and SISTR (Yoshida et al., 2016) for the phylogenetic analysis of strain FSL R9-0095 (Supplementary Table 2). These draft genome sequences included six strains of Typhisuis, eight strains of Paratyphi C, two strains of Choleraesuis *sensu stricto*, three strains of Choleraesuis var. Decatur, and seven

142 strains of Choeraesuis var. Kunzendorf. The Illumina sequences of strain FSL R9-0095 and
143 the corresponding ONT sequences generated using all the sequencing workflows were
144 analyzed through kSNP analysis. Twenty-seven draft genomes were downloaded from
145 GenBank and further confirmed by SeqSero2 and SISTR for the phylogenetic analysis of
146 strain FSL R8-3858 (Supplementary Table 2). These draft genomes included 16 strains of
147 Orion, seven strains of Orion var. 15^{+}, and four strains of Orion var. 15^{+}, 34^{+}. The Illumina
148 sequences of strain FSL R8-3858 and the corresponding ONT sequences generated using all
149 the sequencing workflows were analyzed through kSNP analysis.

150 *2.5.2. CFSAN high quality SNP (hqSNP) pipeline*

151 The hqSNP pipeline developed by the Center for Food Safety and Applied Nutrition (CFSAN
152 SNP Pipeline v.1.0.0/FDA) was used for SNP calling of *Salmonella* strains with default
153 quality filters (Davis et al., 2015). Read mapping was performed using Burrows-Wheeler
154 Aligner (v0.7.17-r1188) (BWA-MEM) with the -x ont2d option (Li and Durbin, 2010; Hyeon
155 et al., 2018) for the ONT sequences. An SNP matrix and alignment concatenated SNP were
156 produced with customized Python scripts.

157 *2.5.3. Phylogenetic tree generated from CFSAN hqSNP analysis*

158 Maximum likelihood (ML) phylogenetic trees were built using PhyML v3.3 based on the
159 hqSNP matrices and visualized with Figtree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).
160 All the phylogenetic trees shown in this paper were midpoint rooted.

161 *2.6. Phage detection*

162 All the Illumina data of Orion and its varirants, as well as the assembled genomes of strain

FSL R8-3858 sequenced using either Illumina or ONT, were submitted to the Phage Search Tool - Enhanced Release (PHASTER API website, http://phaster.ca). Results were extracted from the files returned from the server (Zhou et al., 2011; Arndt et al., 2016; Worley et al., 2018). Sequences classified by PHASTER as questionable (score between 70 and 90), intact (score > 90) or incomplete (score < 70) were all recorded.

3. Results and discussion

3.1. Overview of ONT raw reads

The same ONT fast5 data were used for basecalling by the high accuracy (HAC) model and base modified model (workflow 1 vs workflow 2) (Supplementaryable 1). The data with quality score >= 12 were considered high quality data. The mean read length, mean read quality and percentage of high-quality data (high quality data/total bases) were similar between data generated by these two basecalling models (Table 1). The read length of sequences generated from workflow 3 was close to 4 Kbp (Table 1). One strand of dsDNA prepared by the 1D2 library preparation kit was sequenced followed by its complementary strand for workflow 4. This workflow generated sequences with the longest mean read length (Table 1). However, its mean read quality of the sequences was the lowest (Table 1).

Two equal aliquots of the same batch of extracted genomic DNA were sequenced with the R9 and R10 flow cells for workflows 1 and 5, respectively. The yield of the R9 flow cell was higher than that of the R10 flow cell (Table 1). The mean read length of sequences from the R9 flow cell were shorter than that from the R10 flow cell (Table 1). The mean read quality of sequences from the R9 flow cell were higher than that from the R10 flow cell (Table 1). The percentage of high quality data of sequences from the R9 flow cell were higher than that from

the R10 flow cell (Table 1).

According to the previous study sequences produced in two hrs from ONT were sufficient for *Salmonella* serotype prediction (Xu et al., 2020). Therefore, the raw ONT sequences from the first two hrs of sequencing time were extracted to investigate whether they were sufficient for additional subtyping analysis. About 100× genome coverage was achieved after two hrs sequencing, except with workflow 5 which used the R10 flow cell for FSL R9-0095 sequencing (58×, Table 1). This might have resulted from a dramatic decrease of active pores on the R10 from 1,223 to 155 after the library was loaded, although it increased after 1.5 hrs. Overall, comparisons of the sequencing data yield, mean read length, mean quality score and high-quality data percentage from different workflow using two hrs sequencing data were in line with those using the data of full runs (24, 48, 72 hrs, Table 1).

3.2. Overview of ONT assembled contigs

The ONT sequencing data were assembled using Wtdbg2, and corrected by 2 rounds of Racon. Consensus were then obtained by one round of Medaka (Supplementary Figure 1). The overview of the assembled contigs through the reads obtained on GridION are summarized below (Table 2). The contigs generated from Illumina sequencing were used as a benchmark for comparisons. In general, higher matching identity (> 99.90%) and fewer SNPs by DNAdiff (< 1,000) were obtained from workflows 2, 3 and 5 compared with workflows 1 or 4 (< 99.90% of matching identity and >1,000 of SNPs by DNAdiff). DNA libraries prepared with the rapid kit (workflows 1, 2 and 5) and 1D2 kit (workflow 4) generated fewer contigs (< 5) and higher N50 length (> 4.6 Mbp) in general, except for the 48 hrs data obtained with workflow 5 for FSL R8-3858 sequencing (N50 = 3.51 Mbp). The DNA library

prepared using the PCR kit (workflow 3) generated more contigs and shorter N50 length (Table 2). The reason could be fragmentation of genomic DNA followed by a PCR amplification step during the library preparation. Taylor et al. (2019) reported that four hrs ONT sequences generated full-length genomes with an average identity of 99.87% for *S. Bareilly* and 99.89% for *Escherichia coli* in comparison to the respective Illumina references. With the accuracy improvement of ONT sequencing, especially for the basecaller and flow cell, the highest matching identity was improved to 99.99% as demonstrated using sequencing workflow 3 with the PCR kit in this study (Table 2).

Two sequencing workflows were tested to reduce the impact of methylation, as an average of 95% of the discrepant positions between the sequencing technologies (ONT *versus* Illumina) were caused by methylations (Greig et al., 2019). These were workflow 2 (R9 flow cell + rapid kit + base modified model basecaller) and workflow 3 (R9 flow cell + PCR kit + HAC model basecaller). The base modified model basecaller was trained with human and *E. coli* reads, which allowed for calling 5-methylcytosine (5mC) and N6-methyladenine (6mA) modified bases. In comparison with workflow 1, the total SNP loci generated from workflow 2 decreased from 4492 - 4854 to 29 - 38 for both tested strains (Table 2). Since workflow 3 included a PCR amplification step in library preparation, it could exclude methylations from the library (Liu et al., 2019). Compared with workflow 1, the total SNP loci decreased to 42 - 245 for both tested strains (Table 2). Workflow 4 generated similar SNP numbers to that of workflow 1, ranging from 4,400 to 6,000, which suggested that sequencing both DNA strands might not improve accuracy (Table 2). Despite the low data quantity and genome coverage, fewer SNPs were obtained by sequencing on R10 flow cells (workflow 5) compared to R9

flow cells (workflow 1). R10 flow cell had a longer barrel and dual reader head, enabling improved resolution of homopolymeric regions and improving the consensus accuracy of ONT sequencing data (https://nanoporetech.com/about-us/news/r103-newest-nanopore-high-accuracy-nanopore-sequencing-now-available-store). Current results indicated that flow cell R10 flow cell did improve sequencing accuracy.

3.3. Differentialtion of serotypes and serotype variants that share the same antigenic formula

3.3.1 SNP analysis

Isolates belonging to each of the three serotypes (Typhisuis, Paratyphi C and Choleraesuis) and two serotype variants (Chloeraesuis var. Decatur abd Chloeraesuis var. Kunzendorf) sharing the same antigenic formula 6,7:c:1,5 formed a distinct clade in the kSNP phylogeny with 100% bootstrap support for each clade (Figure 1). This suggested that kSNP analysis using assembled genomes could further differentiate serotypes and variants with this antigenic formula. Genomes of strain FSL R9-0095 sequenced with either the Illumina or ONT platforms fell into the Choleraesuis var. Kunzenforf clade (Figure 1), which indicated that ONT sequencing was equivalent to Illumina for kSNP analysis in identifying these serotypes and serotype variants. Since the SNP numbers generated through sequencing workflows 1 and 4 were higher than those through the other three workflows, the branches of these four corresponding genomes were longer than the other ONT genomes with 100% of bootstrap support (Figure 1). A shorter branch was observed in the subclade through use of workflow 5 (sequencing on the R10 flow cell).

Genomes of strain FSL R9-0095 sequenced on the Illumina platform were used as a

reference for read mapping in the analysis of CFSAN hqSNP. ONT sequences generated using the five different workflows and from different sequencing time were analyzed using the CFSAN hqSNP pipeline. The numbers of hqSNP are listed below (Table 2). Sequencing workflows 2 and 3 generated 0 hqSNP between Illumina data and corresponding ONT data for strain FSL R9-0095. Workflow 5 generated 1 hqSNP after two or 24 hrs of sequencing of the same strain. Overall, the 5 sequencing workflows yielded fewer than 10 hqSNPs compared with the corrsponding Illumina sequences (Table 2).

The phylogenetic trees generated by hqSNP using ONT and/or Illumina sequences for strain FSL R9-0095 are presented below (Figure 2a, 2b and 2c). Three phylogenetic trees were constructed. In order to compare ONT and Illumina sequences, the first tree was built using both ONT and Illumina sequences, including Illumina sequences from GenBank, Illumina sequence of the tested isolate, and ONT sequences generated by different workflows and different sequencing time (Figure 2a). In the second tree, one ONT-sequenced isolate was included to replace the corresponding Illumina-sequenced genome to assess hqSNP-based phylogenetic analysis using ONT sequencing (Figure 2b). The third tree was built using Illumina sequences of these isolates only. Strain FSL R9-0095 was correctly placed in the Cholerasuis var. Kunzendorf clade (Figures 2a, 2b and 2c). All the different serotypes or serotype variants under antigenic formula 6,7:c:1,5 were clearly separated into their respective clades, and all the clades were supported by 100% bootstrap. This indicated that the CFSAN hqSNP-based phylogenetic analysis with either ONT or Illumina data was able to differentiate Choleraesuis var. Kunzendorf isolates. In addition, less than 30 mins was needed for library preparation by the rapid library preparation kit, while about three and five hrs were

required using the 1D2 and PCR kits, respectively. In summary, workflow 2 (which used the rapid library preparation kit) was the more optimal procedure to deliver the desired phylogenetic analysis in terms of accuracy and tunaround time. The phylogenetic trees constructed from data sets of two sequencing times (two hrs and 24 hrs) using workflow 2 showed identical topology (data not shown), suggesting that ONT sequences collected within two hrs were sufficient for CFSAN hqSNP-based phylogenetic analysis to generate correct variant prediction.

Longo et al. (2019) showed that WGS followed by Short Read Sequence Typing for Bacterial Pathogens (SRST2) for MLST (Inouye et al., 2014) in *Salmonella* TypeFinder (https://cge.cbs.dtu.dk/services/SalmonellaTypeFinder/) could correctly identify all the analyzed *S.* Choleraesuis var. Kunzendorf isolates. However, *Salmonella* TypeFinder cannot currently accept raw ONT reads as inputs and few tools are available to identify Cholerasuis var. Kunzendorf using ONT sequences. In this study, phylogenetic analysis through kSNP or CFSAN hqSNP analysis using ONT sequences successfully identified the isolate FSL R9-0095 to the variant level, which suggests the feasibility of applying ONT sequencing for differentiation of the serotypes or serotype variants sharing the same formula 6,7:c:1,5. The phylogenetic result from this study was in accordance with the result from Nair et al. (2020) using Illumina sequences.

3.4. Differentiation of serotype variants that differ by minor antigens

3.4.1 SNP analysis

Similar results were obtained from both kSNP and CFSAN hqSNP trees generated using either the ONT or the Illumina-sequenced genomes of FSL R8-3858. All the ONT sequences

were clustered with their corresponding Illumina sequences (Suplementary Figure 2 and Supplementary Figure 3a). Serotype Orion and its variants Orion var. 15^+ and Orion var. 15^+, 34^+ were indistinguishable in both SNP trees (Supplementary Figure 2 and Supplementary Figure 3), indicating that phylogenetic clustering is not able to differentiate the two Orion variants from each other and Orion.

3.4.2. Detection of prophage

Since the Orion variants are caused by phage conversion and phylogenetic analysis could not differentiate them, *in silico* prophage detection was used for their differention. Prophage prediction using Illumina data correctly identified Orion, Orion var. 15^+, or Orion var. 15^+, 34^+ in 27 of 42 isolates (Supplementary Table 1). Four sequences were untypable or questionable since ϕ_{15} and/or ϕ_{34} could not be identified (Supplementary Table 1). Moreover, 11 isolates had questionable prohage prediciton that was inconsistent with the original phage typing records (Supplementary Table 1). A 25.6 Kbp region was detected with Illumina data as phage ϕ_{15} with a score of 20 (marked as incomplete), and a 36.8 Kbp region was detected as phage ϕ_{34} with a score of 80 (marked as questionable). Sub-optimal assembly and identification of the two phages was likely due to frequent occurrence of sequencing gaps in the prophage regions caused by short Illumina reads.

Twenty-seven of the 42 Illumina sequences tested were identified to be in accordance with their original record through PHASTER (supplymentary Table 1). By contrast, both prophages for genomes assembled from ONT data were detected with higher scores (100 - 150 for ϕ_{15} and 80 - 90 for ϕ_{34}), except the genome assembled from 48 hr sequences using workflow 4 (Table 3). These results showed that the much longer ONT reads (compared with

Illumina reads) allowed the detection of more complete prophages for definitive differentiation of Orion variants.

5. Conclusion

In this study we demonstrated that long-read nanopore sequencing technology can be used as a subtyping tool to differentiate *Salmonella* serotypes or serotype variants which share the same or related antigenic formulae. The matching identity between ONT and Illumina sequences was improved to 99.98 – 99.99% using the workflow including the bioinformatics pipeline. SNP-based phylogenetic analysis successfully identified serotypes and serotype variants including Choleraesuis, Choleraesuis var. Kunzendorf, Choleraesuis var. Decatur, Paratyphi C, Typhisuis by using ONT sequences. ONT sequencing can also successfully differentiate variants of Orion, Orion var. 15^{+} and Orion var. 15^{+}, 34^{+} by addition of PHASTER prediction.

Acknowledgement

The authors would like to thank Dr. Martin Weidmann at Cornell University for providing strains used in this study. The authors would also like to thank Oxford Nanopore Technologies (Thomas Bray, Lin Jin, Xinran Yu, Xiang Chen, Iain MacLaren, Richard Compton, Stephen Rudd, David Dai) for supporting the establishment of ONT capability at the Mars Global Food Safety Center.

References

Alam, M.J., Renter, D.G., Ives, S.E., Thomson, D.U., Sanderson, M.W., Hollis, L.C., Nagaraja, T.G., 2009. Potential associations between fecal shedding of *Salmonella* in feedlot cattle treated for apparent respiratory disease and subsequent adverse health

outcomes. Vet. Res. 40, 2. http://dx.doi.org/10.1051/vetres:2008040.

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie M.E., Gouil ,Q., 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 21, 30. https://doi.org/10.1186/s13059-020-1935-5.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44, 16 - 21. https://doi.org/10.1093/nar/gkw387.

Banerji, S., Simon, S., Tille, A., Fruth, A., Flieger, A., 2020. Genome-based *Salmonella* serotyping as the new gold standard. Sci. Rep. 10 (1), 4333. Doi: 10.1038/s41598-020-61254-1.

Cabrera, R., Ruiz, J., Ramírez, M., Bravo, L., Fernández, A., Aladueña, A., Echeíta, A., Gascón, J., Alonso, P.L., Vila, J., 2006. Dissemination of *Salmonella enterica* serotype Agona and multidrug-resistant *Salmonella enterica* serotype Typhimurium in Cuba. Am. J. Trop. Med. Hyg. 74, 1049 - 1053.

Carroll, L.M., Wiedmann, M., den Bakker, H., Siler, J., Warchocki, S., Kent, D., Lyalina, S., Davis, M., Sischo, W., Besser, T., Warnick, L.D., Pereira, R.V., 2017. Whole-genome sequencing of drug resistant *Salmonella enterica* isolates from dairy cattle and humans in New York and Washington States reveals source and geographic associations. Appl. Environ. Microbiol. 83, e00140-17. https://doi.org/10.1128/AEM.00140-17.

Chiu, C.H., Su, L.H., Chu, C., 2004. *Salmonella enterica* serotype Choleraesuis: epidemiology, pathogenesis, clinical disease, and treatment. Clin. Microbiol. Rev. 17, 311 - 322. Doi: 10.1128/cmr.17.2.311-322.2004.

361 Corry, J.E.L., Allen, V.M., Hudson, W.R., Breslin, M.F., Davies, R.H., 2020. Sources of
362 *Salmonella* on broiler carcasses during transportation and processing: modes of
363 contamination and methods of control. J. Appl. Microbiol. 92(3): 424 – 432.
364 Davis, S., Pettengill, J.B., Luo, Y., Payne, J., Shpuntoff, A., Rand, H., Strain, E., 2015.
365 CFSAN SNP Pipeline: an automated method for constructing SNP matrices from
366 next-generation sequence data. PeerJ Computer Science 1, e20.
367 https://doi.org/10.7717/peerj-cs.20.
368 De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., Van Broeckhoven, C., 2018. NanoPack:
369 visualizing and processing long-read sequencing data. Bioinformatics 34 (15), 2666 -
370 2669. Doi: 10.1093/bioinformatics/bty149.
371 Dieckmann, R., Malorny, B., 2011. Rapid screening of epidemiologically important
372 *Salmonella enterica* subsp. *enterica* serovars by whole-cell matrix-assisted laser
373 desorption ionization - time of flight mass spectrometry. Appl. Environ. Microbiol.
374 77(12): 4136 - 4146.
375 Diep, B., Barretto, C., Portmann, A-C., Fournier, C., Karczmarek, A., Voets, G., Li, S., Deng,
376 X., Klijn, A., 2019. *Salmonella* serotyping: comparison of the traditional method to a
377 microarray-based method and an in silico platform using whole genome sequencing data.
378 Front. Microbiol. 10, 2554. Doi: 10.3389/fmicb.2019.02554.
379 Felsenstein, J., 1985. Confidence-delimitation on phylogenies - an approach using the
380 bootstrap. Evolution. 39, 783-791.

381 Gardner, S.N., Hall, B.G., 2013. When whole-genome alignments just won't work: kSNP v2
382 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial
383 genomes. PLoS One 8, e81760. https://doi.org/10.1371/journal.pone.0081760.
384 Gargis, A.S., Cherney, B., Conley, A.B., McLaughlin, H.P., Sue, D., 2019. Rapid detection of
385 genetic engineering, structural variation, and antimicrobial resistance markers in bacterial
386 biothreat pathogens by nanopore sequencing. Sci. Rep. 9(1), 13501. Doi:
387 10.1038/s41598-019-49700-1.
388 GMA, 2009. The association of food, beverage and consumer products companies (GMA):
389 Control of *Salmonella* in low-moisture foods.
390 Greig, D.R., Jenkins, C., Gharbia, S., Dallman, T.J., 2019. Comparison of single-nucleotide
391 variants identified by Illumina and Oxford Nanopore Technologies in the context of a
392 potential outbreak of Shiga toxin-producing *Escherichia coli*. GigaScience 8, 1 - 12.
393 https://doi.org/10.1093/gigascience/giz104.
394 Grimont, P., Weill, F., 2007. Antigenic formulae of the *Salmonella* serovars, (9th ed.) Paris:
395 WHO Collaborating Centre for Reference and Research on *Salmonella*.
396 Gupta, S.K., McMillan, E.A., Jackson, C.R., Desai, P.T., Porwollik, S., McClelland, M., Hiott,
397 L.M., Humayoun, S.B., Frye, J.G., 2016. Draft genome sequence of *Salmonella enterica*
398 subsp. *enterica* serovar Orion strain CRJJGF 00093 (phylum Gammaproteo bacteria).
399 Genome Announc. 4(5), e01063-16. Doi:10.1128/genomeA.01063-16.
400 Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for
401 genome assemblies. Bioinformatics. 29 (8), 1072 - 1075.
402 Doi:10.1093/bioinformatics/btt086.

403 Hyeon, J-Y., Li, S., Mann, D.A., Zhang, S., Li, Z., Chen, Y., Deng, X., 2018.
404 Quasimetagenomics based and real-time-sequencing-aided detection and subtyping of
405 *Salmonella enterica* from food samples. Appl. Environ. Microbiol. 84, e02340-17.
406 https://doi.org/10.1128/AEM.02340-17.
407 Ibrahim, G.M., Morin, P.M., 2018. *Salmonella* serotyping using whole genome sequencing.
408 Front. Microbiol. 9, 2993. https://doi.org/10.3389/fmicb.2018.02993.
409 Inouye, M., Dashnow, H., Raven, L., Schultz, M., Pope, B.J., Tomita, T., Zobel, J., Holt, K.E,
410 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology
411 labs. Genome Med. 6(11): 1.46.
412 Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D.,
413 Dilthey, A.T., Fiddes, I.T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H.E.,
414 Pedersen, B.S., Rhie, A., Richardson, H., Quinlan, A.R., Snutch, T.P., Tee, L., Paten, B.,
415 Phillippy, A.M., Simpson, J.T., Loman, N.J., Loose, M., 2018. Nanopore sequencing and
416 assembly of a human genome with ultra-long reads. Nat. Biotechnol. 36, 338.
417 https://doi.org/10.1038/nbt.4060.
418 Jain, M., Tyson, J.R., Loose, M., L C Ip, C., Eccles, D., O'Grady, J., Malla, S., Leggett, R.M.,
419 Wallerman, O., Jansen, H., Zalunin, V., Birney, E., Brown, B., Snutch, T., Olsen, H.,
420 2017. MinION analysis and reference consortium: phase 2 data release and analysis of
421 R9.0 chemistry, vol. 6.
422 Kurtz, S,. Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg,
423 S,L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5,
424 R12.

425 Lang, D., Zhang, S., Ren, P., Liang, F., Sun, Z., Meng, G., Tan, Y., Hu, J., Li, X., Lai, Q.,
426 Han, L., Wang, D., Hu, F., Wang, W., Liu, S., 2020. Comparison of the two up-to-date
427 sequencing technologies for genome assembly: HiFi reads of Pacbio Sequel II system
428 and ultralong reads of Oxford Nanopore. bioRxiv. Doi:
429 https://doi.org/10.1101/2020.02.13.948489.
430 Leekitcharoenphon, P., Sørensen, G., Löfström, C., Battisti, A., Szabo, I., Wasyl, D., Slowey,
431 R., Zhao, S., Brisabois, A., Kornschober, C., Kärssin, A., Szilárd, J., Cerný, T., Svendsen,
432 C.A., Pedersen, K., Aarestrup, F.M., Hendriksen, R.S., 2019. Cross-border transmission
433 of *Salmonella* Choleraesuis var. Kunzendorf in European pigs and wild boar: infection,
434 genetics, and evolution. Front. Microbiol. 10, 179. Doi: 10.3389/fmicb.2019.00179.
435 Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler
436 transform. Bioinformatics 26, 589 - 595.
437 iu, Q., Georgieva, D.C., Egli, D. and Wang, K., 2019. NanoMod: a computational tool to
438 detect DNA modifications using Nanopore long-readsequencing data. BMC Genomics
439 20, 78. https://doi.org/10.1186/s12864-018-5372-8.
440 Longo, A., Petrin, S., Mastrorilli, E., Tiengo, A., Lettini, A.A., Barco, L., Ricci, A., Losasso,
441 C., Cibin, V., 2019. Characterizing *Salmonella enterica* serovar Choleraesuis, var.
442 Kunzendorf: a comparative case study. Front. Vet. Sci. 6, 316.
443 Doi:10.3389/fvets.2019.00316.
444 McWhorter, A.R., Davos, D., Chousalkar, K.K., 2015. Pathogenicity of Salmonella strains
445 isolated from egg shells and the layer farm environment in Australia. Appl. Environ.
446 Microbiol. 81: 405–414. doi:10.1128/AEM.02931-14.

447 Mottawea, W., Duceppe, M-O., Dupras, A.A., Usongo, V., Jeukens, J., Freschi, L.,
448 Emond-Rheault, J-G., Hamel, J., Kukavica-Ibrulj, I., Boyle, B., Gill, A., Burnett, E.,
449 Franz, E., Arya, G., Weadge, J.T., Gruenheid, S., Wiedmann, M., Huang, H., Daigle, F.,
450 Moineau, S., Bekal, S., Levesque, R.C., Goodridge, L.D. Ogunremi, D., 2018.
451 *Salmonella enterica* prophage sequence profiles reflect genome diversity and can be used
452 for high discrimination subtyping. Front. Microbiol. 9, 836. Doi:
453 10.3389/fmicb.2018.00836.
454 Münch, S., Braun, P., Wernery, U., Kinne, J., Pees, M., Flieger, A., Tietze, E., Rabsch, W.,
455 2012. Prevalence, serovars, phage types, and antibiotic susceptibilities of *Salmonella*
456 strains isolated from animals in the United Arab Emirates from 1996 to 2009. Trop.
457 Anim. Health Prod. 44, 1725 - 1738. http://dx.doi.org/10.1007/s11250-012-0130-4.
458 Nair, S., Fookes, M.,Corton, C., Thomson, N,R., Wain, J., Langridge, G.C., 2020. Genetic
459 Markers in S. Paratyphi C Reveal Primary Adaptation to Pigs. Microorganisms, 8, 657.
460 doi:10.3390/microorganisms8050657.
461 Olaimat, A.N., Holley, R.A., 2012. Factors influencing the microbial safety of fresh produce:
462 a review. Food Microbiol. 32 (1), 1 - 19. https://doi.org/10.1016/j.fm.2012.04.016.
463 Rampersad, J., Johnson, J., Brown, G., Samlal, M., Ammons, D., 2008. Comparison of
464 olymerase chain reaction and bacterial culture for *Salmonella* detection in the Muscovy
465 duck in Trinidad and Tobago. Rev. Panam. Salud. Publica. 23, 264 - 267.
466 http://dx.doi.org/10.1590/S1020-49892008000400006.
467 Ruan, J., Li, H., 2019. Fast and accurate long-read assembly with Wtdbg2. 10.1101/530972.

468 Shi, C., Singh, P., Ranieri, M.L., Wiedmann, M., Moreno Switt, A.I., 2015. Molecular
469 methods for serovar determination of *Salmonella*. Crit. Rev. Microbiol. 41 (3), 309 - 325.
470 https://doi.org/10.3109/1040841X.2013.837862.
471 Sirichote, P., Hasman, H., Pulsrikarn, C., Schønheyder, H. C., Samulioniené, J.,
472 Pornruangmong, S., Bangtrakulnonth, A., Aarestrup, F.M., Hendriksen, R.S., 2010.
473 Molecular characterization of extended-spectrum cephalosporinase-producing
474 *Salmonella enterica* serovar Choleraesuis isolates from patients in Thailand and
475 Denmark. J. Clin. Microbiol. 48, 883 - 888. Doi: 10.1128/JCM.01792-09.
476 Tang, S. Orsi, R.S., Luo, H., Ge, C., Zhang, G., Baker, R.C., Stevenson A., Wiedmann, M.,
477 2019. Assessment and comparison of molecular subtyping and characterization methods
478 for *Salmonella*. Front Microbiol. 10, 1591.
479 Taylor, T.L., Volkening, J.D., DeJesus, E., Simmons, M., Dimitrov, K.M., Tillman, G.E.,
480 Suarez, D.L., Afonso, C.L., 2019. Rapid, multiplexed, whole genome and plasmid
481 sequencing of foodborne pathogens using long-read nanopore technology. Sci. Rep. 9 (1),
482 16350. https://doi.org/10.1038/s41598-019-52424-x.
483 Trafny, E.A., Kozłowska, K., Szpakowska, M., 2006. A novel multiplex PCR assay for the
484 detection of *Salmonella enterica* serovar Enteritidis in human faeces. Lett. Appl.
485 Microbiol. 43, 673 - 679. http://dx.doi.org/10.1111/j.1472-765X.2006.02007.x.
486 Vaser, R., I. Sovic, N. Nagarajan, and M. Sikic. 2017. Fast and accurate *de novo* genome
487 assembly from long uncorrected reads. Genome Res. 27 (5), 737 - 746. Doi:
488 10.1101/gr.214270.116.

Uelze, L., Borowiak, M, Deneke, C., Szabó, I., Fischer, J., Tausch, S.H., Malorny, B., 2020. Performance and accuracy of four open-source tools for in silico serotyping of *Salmonella* spp. based on whole-genome short-read sequencing data. Appl. Environ. Microbiol. 86, e02265-19. https://doi.org/10.1128/AEM.02265-19.

Worley, J., Meng, J., Allard, M.W., Brown, E.W., Timme, R.E., 2018. *Salmonella enterica* phylogeny based on whole-genome sequencing reveals two new clades and novel patterns of horizontally acquired genetic elements. mBio 9, e02303-18. https://doi.org/10.1128/mBio.02303-18.

Xu, F., Ge, C., Luo, H., Li, S., Wiedmann, M., Deng, X., Zhang, G., Stevenson, A., Baker, R.C., Tang, S., 2020. Evaluation of real-time nanopore sequencing for *Salmonella* serotype prediction. Food Microbiol. 89, 103452. Doi: 10.1016/j.fm.2020.103452.

Yachison, C.A., Yoshida, C., Robertson, J., Nash, J.H.E., Kruczkiewicz, P., Taboada, E.N., Walker, M., Reimer, A., Christianson, S., Nichani, A., PulseNet Canada Steering, C., Nadon, C., 2017. The validation and implications of using whole genome sequencing as a replacement for traditional serotyping for a national *Salmonella* reference laboratory. Front. Microbiol. 8, 1044. https://doi.org/10.3389/fmicb.2017.01044.

Yoshida, C.E., Kruczkiewicz, P., Laing, C.R., Lingohr, E.J., Gannon, V.P., Nash, J.H., Taboada, E.N., 2016. The *Salmonella* in silico typing Resource (SISTR): an open webaccessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. PloS One 11 (1), e0147101. https://doi.org/10.1371/journal.pone.0147101.

509 Zhang, S., Yin, Y., Jones, M.B., Zhang, Z., Deatherage Kaiser, B.L., Dinsmore, B.A.,
510 Fitzgerald, C., Fields, P.I., Deng, X., 2015. *Salmonella* serotype determination utilizing
511 high-throughput genome sequencing data. J. Clin. Microbiol. 2015, 53(5):1685-1692.
512 Zhang, S., Den-Bakker, H.C., Li, S., Chen, J., Dinsmore, B.A., Lane, C., Lauer, A.C., Fields,
513 P.I., Deng, X., 2019. SeqSero2: rapid and improved *Salmonella* serotype determination
514 using whole genome sequencing data. Appl. Environ. Microbiol. 85(23). Pii, e01746-19.
515 Doi: 10.1128/AEM.01746-19.
516 Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S., 2011. PHAST: a fast phage
517 search tool. Nucleic Acids Res. 39, 347 - 352.

Table 1. Overview of ONT raw reads generated using different sequencing workflows.

Strain	Sequencing workflow No.	Flowcell	Library preparation kit	Basecalling model	Sequencing time	Mean read length	Mean read quality	Number of reads	Reads length N50	Total bases (Mb)	Coverage	> Q7 (Mb)	> Q12 (Mb)	Percentage of high-quality data
FSL R9-0095	1	R9	Rapid	HAC	2hr	5,799.7	12.1	86,580	9,119	502.1	105	502.1	295.6	58.87%
					24hr	5,816.1	12.1	555,409	8,926	3,230.3	673	3,230.3	1,819.9	56.34%
	2	R9	Rapid	Base modified	2hr	4,450.2	11.7	132,018	8,607	587.5	122	586.3	319.0	54.41%
					24hr	4,441.2	11.7	812,214	8,580	3,607.1	751	3,600.5	1,942.6	53.95%
	3	R9	PCR kit	HAC	2hr	4,431.7	12.5	100,147	5,930	443.8	92	443.8	298.4	67.24%
					72hr	3,945.2	10.7	1,846,360	5,745	7,284.2	1,518	7,283.4	2,213.0	30.38%
	4	R9	1D2	HAC	2hr	12,570.7	7.3	84,118	25,490	813.6	170	642.4	38.3	5.96%
					48hr	12,489.7	7.2	800,235	25,417	7,649.3	1,594	5,921.2	346.5	5.85%
	5	R10	Rapid	HAC	2hr	9,101.6	9.5	30,805	14,162	280.4	58	280.4	0.0	0%
					24hr	9,237.1	9.3	214,847	14,158	1,983.6	413	1,983.6	0.0	0%
FSL R8-3858	1	R9	Rapid	HAC	2hr	7,026.3	12.0	128,056	12,443	899.7	187	899.6	536.0	59.58%
					48hr	7,081.5	11.5	2,091,542	12,969	14,811.0	3,086	14,809.9	6,799.6	45.91%
	2	R9	Rapid	Base modified	2hr	7,056.8	11.8	140,039	12,840	988.2	206	988.0	527.0	53.34%
					48hr	7,081.5	11.5	2,091,542	12,969	14,811.0	3,086	14,809.9	6,799.6	45.91%
	3	R9	PCR kit	HAC	2hr	4,516.6	12.4	104,149	5,781	470.4	98	470.4	319.1	67.84%
					72hr	4,484.1	11.4	4,424,664	5,746	19,840.7	4,133	19,839.9	8,596.0	43.33%
	4	R9	1D2	HAC	2hr	16,224.2	7.4	84,081	30,317	1,098.7	229	884.5	32.6	3.69%
					48hr	15,497.2	7.4	643,208	29,007	8,045.2	1,676	6,429.2	298.8	4.65%
	5	R10	Rapid	HAC	2hr	7,894.2	9.2	76,008	13,929	600.0	125	600.0	0.0	0%
					48hr	9,092.8	9.3	471,007	14,312	4,282.8	892	4,282.8	0.0	0%

Table 2. Overview of ONT assembled contigs through data analysis pipeline 1.

Strain	Sequencing workflow No.	Flowcell	Library Preparation kit	Basecalling model	Sequencing time	Total length (bp)	Contigs Numbers	N50 Length (bp)	SNPs	InDels	Matching identity (%)	High quality SNPs
FSL R9-0095	1	R9	Rapid	HAC	2hr	4,792,171	3	4,736,332	4,778	3,116	99.83	4
					24hr	4,789,608	2	4,741,208	4,492	2,534	99.85	2
	2	R9	Rapid	Base modified	2hr	4,825,067	4	4,740,589	31	901	99.98	0
					24hr	4,741,424	1	4,741,424	29	652	99.98	0
	3	R9	PCR kit	HAC	2hr	4,667,484	49	183,328	42	493	99.99	0
					72hr	4,561,879	48	174,019	245	1,001	99.97	0
	4	R9	1D2	HAC	2hr	4,742,094	1	4,742,094	5,043	5,840	99.77	1
					48hr	4,790,286	2	4,741,889	5,200	4,911	99.78	4
	5	R10	Rapid	HAC	2hr	4,790,793	2	4,742,687	619	1,672	99.95	1
					24hr	4,783,849	2	4,735,698	811	2,179	99.93	1
FSL R8-3858	1	R9	Rapid	HAC	2hr	4,665,249	1	4,665,249	4,831	2,617	99.84	100
					48hr	4,665,134	1	4,665,134	4,854	2,705	99.83	73
	2	R9	Rapid	Base modified	2hr	4,666,028	1	4,666,028	33	677	99.98	1
					48hr	4,666,136	1	4,666,136	38	666	99.98	1
	3	R9	PCR kit	HAC	2hr	4,583,190	31	273,714	56	483	99.98	1
					72hr	4,500,549	33	180,233	59	407	99.99	1
	4	R9	1D2	HAC	2hr	4,666,639	1	4,666,639	5,605	6,524	99.74	69
					48hr	4,698,102	2	4,665,652	5,964	6,724	99.72	61
	5	R10	Rapid	HAC	2hr	4,666,923	1	4,666,923	691	1,919	99.94	4
					48hr	4,651,530	2	3,508,306	608	1,873	99.95	3

Table 3. Result of predicted prophage for strain FSL R8-3858.

Sequencing technology	Sequencing workflow No.	Flowcell	Library preparation kit	Basecalling model	Sequencing time	Prophage □15			Prophage □34		
						Length (kb)	Completeness	Score	Length (bp)	Completeness	Score
Illumina						25.6	Incomplete	20	36.8	Questionable	80
ONT	1	R9	Rapid	HAC	2hr	45.9	Intact	130	63.7	Questionable	90
					48hr	45.9	Intact	120	47.1	Questionable	80
	2	R9	Rapid	Base modified	2hr	44.3	Intact	100	43.1	Questionable	90
					48hr	46	Intact	100	43.7	Questionable	80
	3	R9	PCR kit	HAC	2hr	44.3	Intact	100	44.1	Questionable	90
					72hr	44.3	Intact	100	45.1	Questionable	90
	4	R9	1D2	HAC	2hr	45.7	Intact	150	43.7	Questionable	90
					48hr	45.6	Intact	150	33.2	Incomplete	30
	5	R10	Rapid	HAC	2hr	46	Intact	110	60.1	Questionable	90
					48hr	44.3	Intact	130	47.1	Questionable	90

Note: When the score was above 90, the completeness was recorded as intact; when the score was between 70 and 90, the completeness was recorded as questionable; when the score was below 70, the completeness was recorded as incomplete. More information about the result interpretation for scoring the prophage region can be found on PASTER API (http://phaster.ca).

Figure legends

Figure 1. Maximum parsimony tree of the serotype formula 6,7:c:1,5 based on k-mer-based SNP analysis. The tree was built using kSNP3 with the core SNPs identified among both Illumina and ONT sequences of the strain FSL R9-0095 and 26 sequences downloaded from the NCBI. Clades of Typhisuis, Paratyphi C, Choleraesuis, Choleraesuis var. Kunzendorf, Choleraesuis var. Decatur were annotated separately. Only high support in the analysis (bootstrap >= 90%) are labeled in this tree. The tree is midpoint rooted and the scale axis is provided below the tree.

Figure 2. Maximum likelihood phylogenetic tree of the serotype formula 6,7:c:1,5 using the CFSAN hqSNP pipeline. The tree was constructed with PhyML using hqSNPs identified among Illumina and ONT sequences of the strain FSL R9-0095 and 26 sequences downloaded from the NCBI. Illumina sequences of FSL R9-0095 was used as the reference genome. (a) The tree includes both the Illumina and the ONT sequencing data of strain FSL R9-0095. Clades of Typhisuis, Paratyphi C, Choleraesuis, Choleraesuis var. Kunzendorf, Choleraesuis var. Decatur are annotated separatel. (b) The data of strain FSL R9-0095 replaced with SNPs from two hrs ONT sequencing data using sequencing workflow 2. (c) The tree only includes Illumina sequencing data of strain FSL R9-0095, which were used as a benchmark phylogenetic tree. Only high support in the analysis (bootstrap >= 90%) are labeled in this tree. The tree is midpoint rooted and the scale axis is provided below the tree.

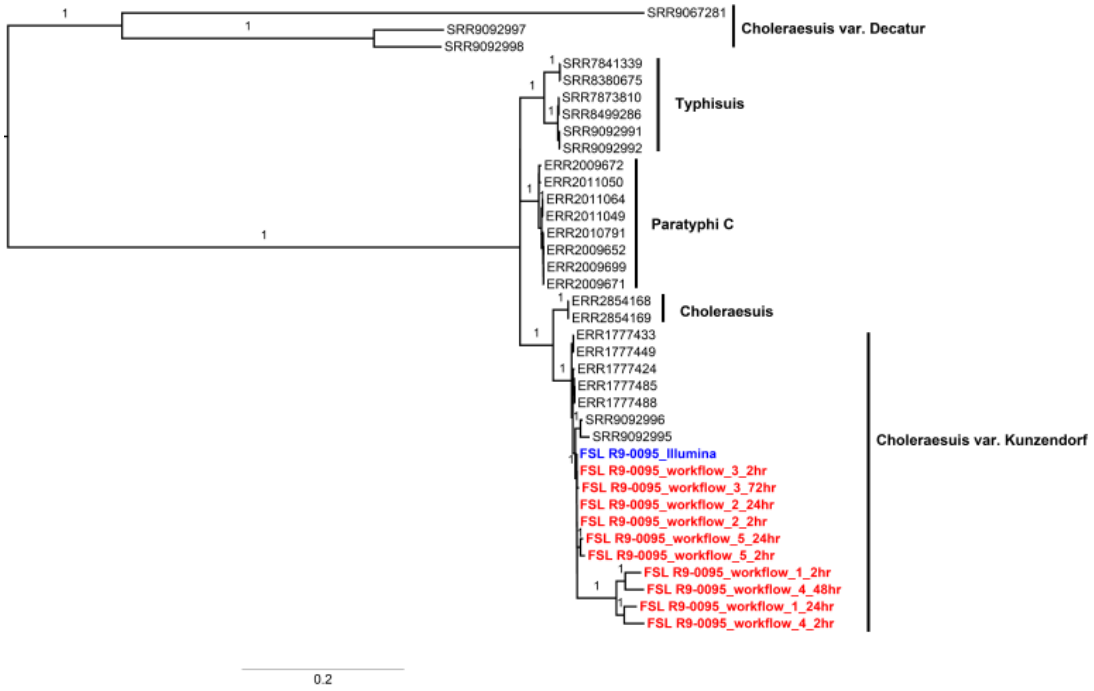

Figure 1

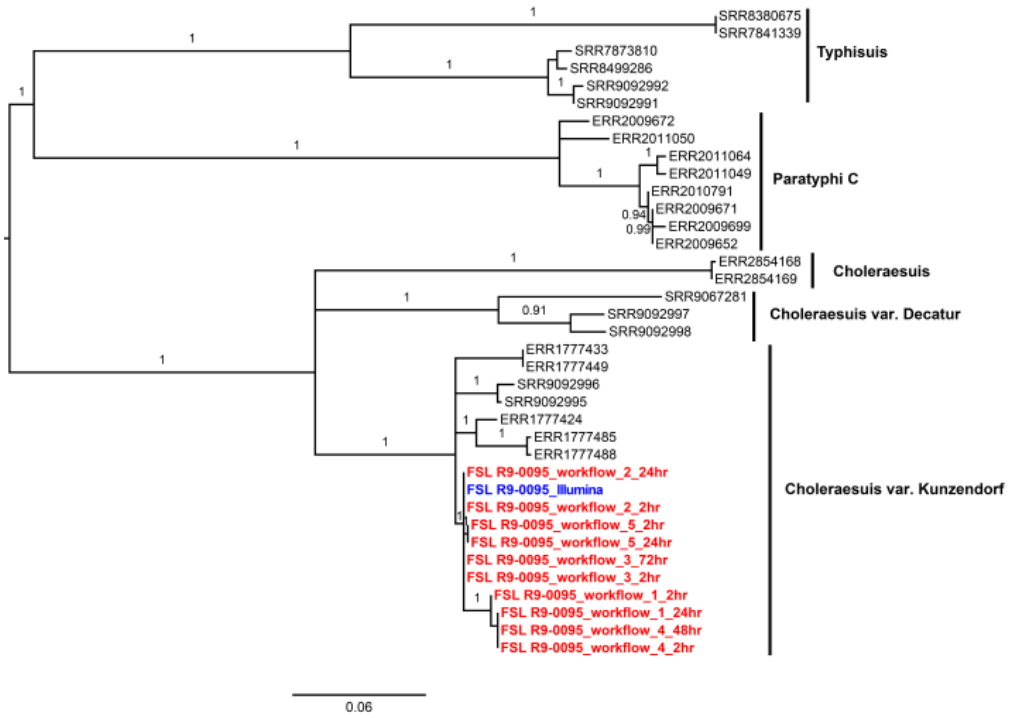

Figure 2a

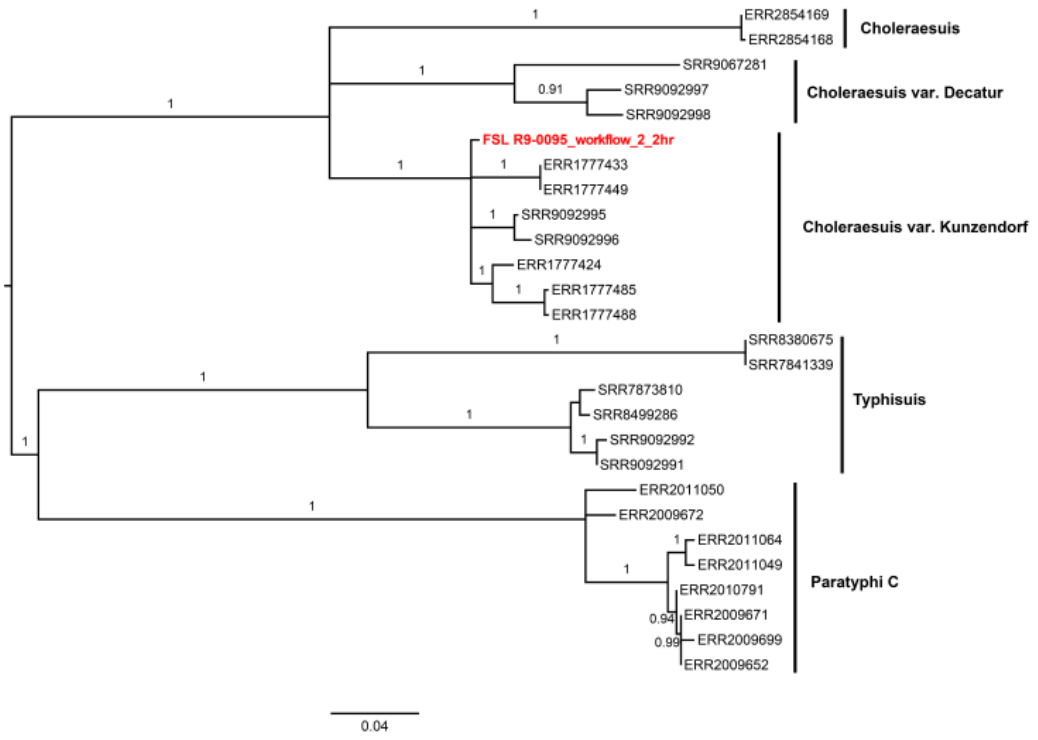

Figure 2b

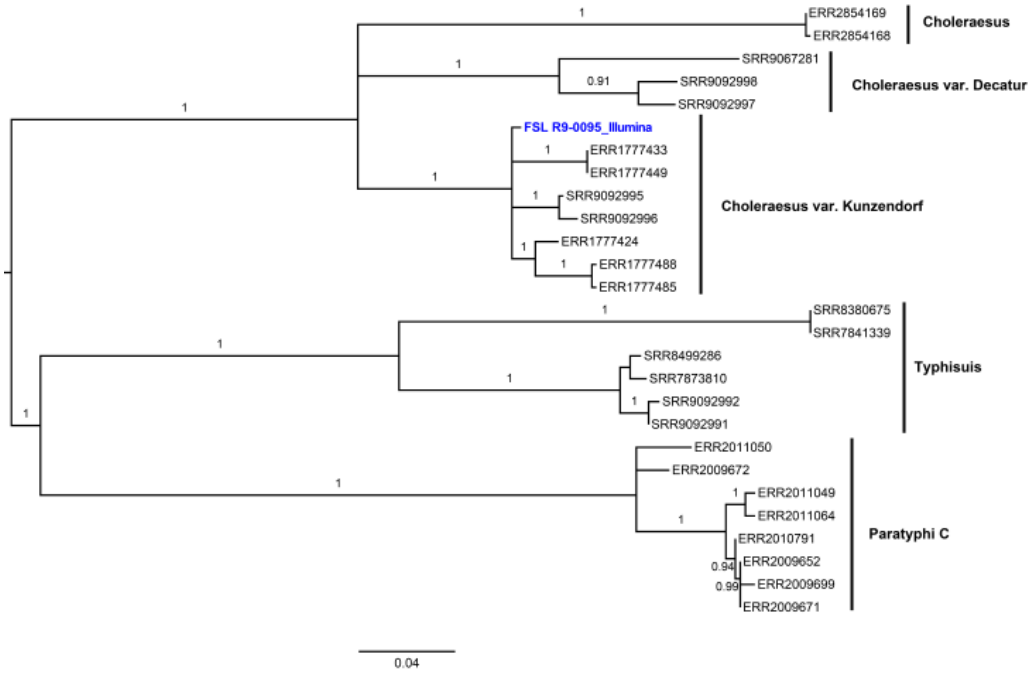

Figure 3c