asm microbe 2019
JUNE 20 – 24 ● SAN FRANCISCO

AES-1176
June 23, 2019

A Pilot Study Evaluating Nanopore Sequencing Technology for *Salmonella* Serotype Prediction

Silin Tang[1,*], Feng Xu[1,*], Chongtao Ge[1], Hao Luo[1], Guangtao Zhang[1], Bob Baker[1], Martin Wiedmann[2], Xiangyu Deng[3]

[1]Mars Global Food Safety Center, Beijing 101407, China, [2]Department of Food Science, Cornell University, Ithaca, NY 14850, USA, [3]Center for Food Safety, University of Georgia, Griffin, GA, 30223, USA

*Corresponding author: silin.tang@effem.com; feng.y.xu@effem.com

MARS Global Food Safety Center

Abstract

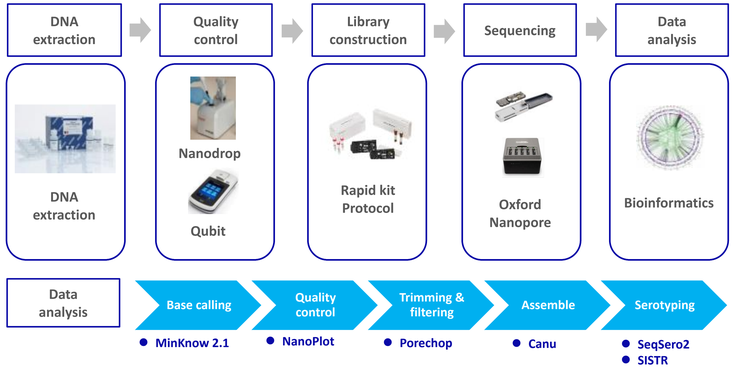
The sequencing platforms developed by Oxford Nanopore Technologies (ONT) using long-read, single-molecule nanopore sequencing technology have advantages in portability, affordability, real-time base calling, and simplicity compared with other sequencing technologies. They provide a potential whole genome sequencing (WGS) method to meet the needs of food industry for effective and efficient *Salmonella* identification. This pilot study sequenced 24 *Salmonella enterica* strains (17 serovars) to explore whether WGS data generated by the ONT sequencing system could accurately predict *Salmonella* serotypes. WGS data generated by Illumina Hiseq (200X) of the same isolates were used to predict serotype as the benchmark of accuracy. We found that the quality score and high-quality data percentage of the reads declined over the sequencing time, whilst the data generated within the first two hours were sufficient for serotype prediction. The size of data generated within the first two hours ranged from 201 to 1,072 MB (42-203X), and the mean read lengths ranged from 3,589 to 11,721 bp. Consensus serotype prediction results were obtained after assembly from SeqSero2 and SISTR for all 24 isolates using the ONT sequencing data. All predictions were identical to the corresponding results generated by Illumina Hiseq. This pilot study indicated that the WGS data generated by the ONT sequencing system had the potential to be used to predict *Salmonella* serotypes within two hours of sequencing time, and the accuracy was comparable with that of the Illumina sequencing system.

Introduction

Methods that can differentiate *Salmonella* beyond the species level (*e.g.* to serovars) are essential for controlling food safety control across the supply chain (Olaimat *et al.* 2012, Barco *et al.* 2013, Shi *et al.* 2015). WGS-based serotyping is gradually replacing conventional serological methods due to its higher resolution of genetic information, and larger serovar coverage of *in silico* serovar prediction tools (Zhang *et al.* 2015, Yohshida *et al.* 2016).

The Illumina sequencing platforms have emerged as the gold standard, however, limitations such as long turnaround time and complex sample preparation procedures are still obstacles for in-house deployment in the food industry. The ONT sequencing platforms are real-time and long-read sequencing portable devices that can generate data for identification and characterization of microbes. We explored the feasibility of using ONT sequencing platforms for *Salmonella* confirmation and identification in a pilot study using various *Salmonella* strains. In this study we compared the performance and corresponding bioinformatics pipelines for *Salmonella* serovar prediction of the ONT platforms WGS data generated using an Illumina sequencing platform.

Materials and Methods



Scheme 1. Workflow of *in silica Salmonella* serotype prediction

Results

Data generation

- The mean quality score of data generated during sequencing of each *Salmonella* decreased gradually over 48 hrs of sequencing time.

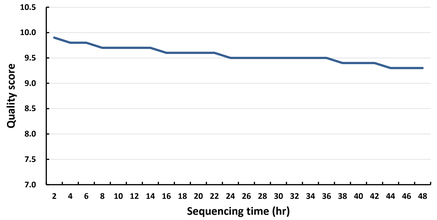
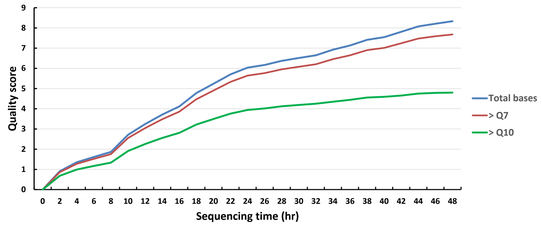


Figure 1. Example of change in mean quality score over 48 hrs sequencing (strain No. 1604).



- Data was generated faster before ~24 hrs. Data size with higher quality (Q score >10) became relatively flat after 24 hrs (Figure 2).
- The quality score of most sequence data was above 7, which passes the internal QC of MinKNOW. Serotype prediction using data of Q > 7 and Q > 10 produced identical results. Therefore, data that passed the default QC (Q score > 7) was used for the following serotype prediction.

Figure 2. Data generation and percentage of data with different quality score (strain No. 1604).

Serotype prediction

- Data generated in different sequencing times were analyzed using the same bioinformatics pipeline. Two hours' data were found to be sufficient to predict correct *Salmonella* serotypes.
- Consensus serotype prediction results were obtained using both ONT and Illumina sequencing (Table 1).

Table 1. Data generated within first 2 hrs and *in-silico* serotype prediction.

Strain No.	Mean read length (bp)	Data size (MB)	ONT		Illumina Hiseq	
			SeqSero2	SISTR	SeqSero2	SISTR
1602	3,589	201	Tennessee	Tennessee	Tennessee	Tennessee
1604	8,129.2	910	Tennessee	Tennessee	Tennessee	Tennessee
1605	8,244.4	857	Tennessee	Tennessee	Tennessee	Tennessee
1609	8,264.3	859	Tennessee	Tennessee	Tennessee	Tennessee
1610	10,758.2	860	Typhimurium	Typhimurium	Typhimurium	Typhimurium
1701	9,317.5	783	Saintpaul	Saintpaul	Saintpaul	Saintpaul
1702	7,479.6	957	Typhimurium	Typhimurium	Typhimurium	Typhimurium
1704	8,041	708	Typhimurium	Typhimurium	Typhimurium	Typhimurium
1705	4,664	243	Thompson	Thompson	Thompson	Thompson
1801	7,927.2	254	Bareilly	Bareilly	Bareilly	Bareilly
1802	8,383.4	939	Newport	Newport	Newport	Newport
1803	11,787.8	566	Anatum	Anatum	Anatum	Anatum
1804	7,466.2	956	Mbandaka	Mbandaka	Mbandaka	Mbandaka
1805	8,086	647	Virchow	Virchow	Virchow	Virchow
1806	11,721.8	469	Mississippi	Mississippi	Mississippi	Mississippi
1807	7,730.8	1,072	Cerro	Cerro	Cerro	Cerro
1808	8,788.5	987	Derby	Derby	Derby	Derby
1809	9,627.4	942	Derby	Derby	Derby	Derby
1810	6,094.6	878	Meleagridis	Meleagridis	Meleagridis	Meleagridis
1811	11,022.4	812	Kentucky	Kentucky	Kentucky	Kentucky
1812	9,083.3	581	Javiana	Javiana	Javiana	Javiana
1813	6,340.1	456	Weltevreden	Weltevreden	Weltevreden	Weltevreden
1814	8,375	1,004	Derby	Derby	Derby	Derby
1815	7,898.1	884	Blockley	Blockley	Blockley	Blockley

Protocol finalization

- A protocol was finalized based on the results of this study (Table 2).
- Further studies will use the same experimental steps and bioinformatics pipeline.

Table 2. Procedure for *Salmonella* serotype prediction using ONT sequencing.

	Step	Duration (hr)
1	DNA extraction from isolates on non-selective agar	3
2	Quality control	0.5
3	Library construction (rapid kits)	1
4	Sequencing	2
5	Data analysis (open source software)	6
	Total*	12.5

* Total duration for 5 strains; time required will increase/decrease with sample number increase/decrease.

Conclusion

- An *in silico Salmonella* serotype prediction protocol was established. The entire process from culture of pure isolates on agar to prediction can be completed within one day.
- Consensus serotype prediction was obtained from ONT and Illumina WGS sequencing for all 24 *Salmonella* strains tested.
- A comprehensive study will be designed as the next step to conduct further systematic evaluation of the application of ONT sequencing for *Salmonella* confirmation and serotype prediction.

Acknowledgement

We would like to thank our collaborator, Oxford Nanopore Technologies (Dr. Thomas Bray, Dr. Lin Jin, Xinran Yu, Xiang Chen, Dr. Iain MacLaren, Richard Compton, Stephen Rudd, Dr. David Dai) for supporting the establishment of ONT capability at the Mars Global Food Safety Center.

Oxford NANOPORE Technologies

References

- **Olaimat, A. N., Holley, R. A. (2012).** Factors influencing the microbial safety of fresh produce: A review. Food Microbiology 32(1): 1-19.
- **Barco, L., Barrucci, F., Olsen, J. E., Ricci, A. (2013).** *Salmonella* source attribution based on microbial subtyping. International Journal of Food Microbiology 163(2-3): 193-203.
- **Shi, C., Singh, P., Ranieri, M. L., Wiedmann, M., Switt, A. I. M. (2015).** Molecular methods for serovar determination of *Salmonella*. Critical Reviews in Microbiology 41(3): 309.
- **Zhang, S., Yin, Y., Jones, M. B., Zhang, Z., Deatherage Kaiser, B. L., Dinsmore, B. A., Fitzgerald, C., Fields, P. I., Deng, X. (2015).** *Salmonella* serotype determination utilizing high-throughput genome sequencing data. Journal of Clinical Microbiology 53(5): 1685-1692.
- **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 web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. PLoS ONE 11(1): e0147101.