



# Direct Metatranscriptome RNA-seq and Multiplex RT-PCR Amplicon Sequencing on Nanopore MinION – Promising Strategies for Multiplex Identification of Viable Pathogens in Food

## OPEN ACCESS

### Edited by:

Victor Ladero,  
Spanish National Research Council,  
Spain

### Reviewed by:

Yangchao Luo,  
University of Connecticut,  
United States  
Miten Jain,  
University of California, Santa Cruz,  
United States

### \*Correspondence:

Mingqun Xu  
xum@neb.com  
Boce Zhang  
boce\_zhang@uml.edu

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Food Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 21 November 2019

**Accepted:** 10 March 2020

**Published:** 09 April 2020

### Citation:

Yang M, Cousineau A, Liu X,  
Luo Y, Sun D, Li S, Gu T, Sun L,  
Dillow H, Lepine J, Xu M and Zhang B  
(2020) Direct Metatranscriptome  
RNA-seq and Multiplex RT-PCR  
Amplicon Sequencing on Nanopore  
MinION – Promising Strategies for  
Multiplex Identification of Viable  
Pathogens in Food.  
Front. Microbiol. 11:514.  
doi: 10.3389/fmicb.2020.00514

Manyun Yang<sup>1†</sup>, Alyssa Cousineau<sup>2†</sup>, Xiaobo Liu<sup>1</sup>, Yaguang Luo<sup>3</sup>, Daniel Sun<sup>2,4</sup>,  
Shaohua Li<sup>2,5</sup>, Tingting Gu<sup>1</sup>, Luo Sun<sup>2</sup>, Hayden Dillow<sup>1</sup>, Jack Lepine<sup>6</sup>, Mingqun Xu<sup>2\*</sup> and  
Boce Zhang<sup>1\*</sup>

<sup>1</sup> Department of Biomedical and Nutritional Sciences, University of Massachusetts, Lowell, MA, United States, <sup>2</sup> New England Biolabs, Inc., Ipswich, MA, United States, <sup>3</sup> Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD, United States, <sup>4</sup> Department of Chemistry, Brandeis University, Waltham, MA, United States, <sup>5</sup> U.S. Food and Drug Administration, Silver Spring, MD, United States, <sup>6</sup> Biomolecular Characterization Lab, University of Massachusetts Lowell, Lowell, MA, United States

Viable pathogenic bacteria are major biohazards that pose a significant threat to food safety. Despite the recent developments in detection platforms, multiplex identification of viable pathogens in food remains a major challenge. A novel strategy is developed through direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing on Nanopore MinION to achieve real-time multiplex identification of viable pathogens in food. Specifically, this study reports an optimized universal Nanopore sample extraction and library preparation protocol applicable to both Gram-positive and Gram-negative pathogenic bacteria, demonstrated using a cocktail culture of *E. coli* O157:H7, *Salmonella enteritidis*, and *Listeria monocytogenes*, which were selected based on their impact on economic loss or prevalence in recent outbreaks. Further evaluation and validation confirmed the accuracy of direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing using Sanger sequencing and selective media. The study also included a comparison of different bioinformatic pipelines for metatranscriptomic and amplicon genomic analysis. MEGAN without rRNA mapping showed the highest accuracy of multiplex identification using the metatranscriptomic data. EPI2ME also demonstrated high accuracy using multiplex RT-PCR amplicon sequencing. In addition, a systemic comparison was drawn between Nanopore sequencing of the direct metatranscriptome RNA-seq and RT-PCR amplicons. Both methods are comparable in accuracy and time. Nanopore sequencing of RT-PCR amplicons has higher sensitivity, but Nanopore metatranscriptome sequencing excels in read length and dealing with complex microbiome and non-bacterial transcriptome backgrounds.

**Keywords:** multiplex identification, viable pathogens, nanopore, metatranscriptome, multiplex RT-PCR

## INTRODUCTION

Biological threats, including bacteria, viruses, and parasites, remain as the top food safety challenge in the United States. According to the CDC surveillance for foodborne disease outbreaks most recent annual report, in 2016 outbreaks attributed to bacterial infection comprised 44% of the total 645 outbreaks and caused 76% of the 847 hospitalization cases (Centers for Disease Control and Prevention, 2018). More importantly, a recent report published by U.S. Department of Agriculture Economic Research Service (USDA ERS) stated that food safety challenges caused an annual loss of \$15.5 billion to the economy and the top 10 infectious bacteria alone contribute to \$10 billion in economic loss (Hoffmann et al., 2015). These statistics revealed that bacterial infection is the primary concern among all biological threats. To cope with the threat of bacterial infection to public health, the demand for a rapid and highly sensitive method to detect and identify bacterial pathogens in food is enormous and becoming more urgent, especially after the implementation of the Food Safety Modernization Act (FSMA) in 2011.

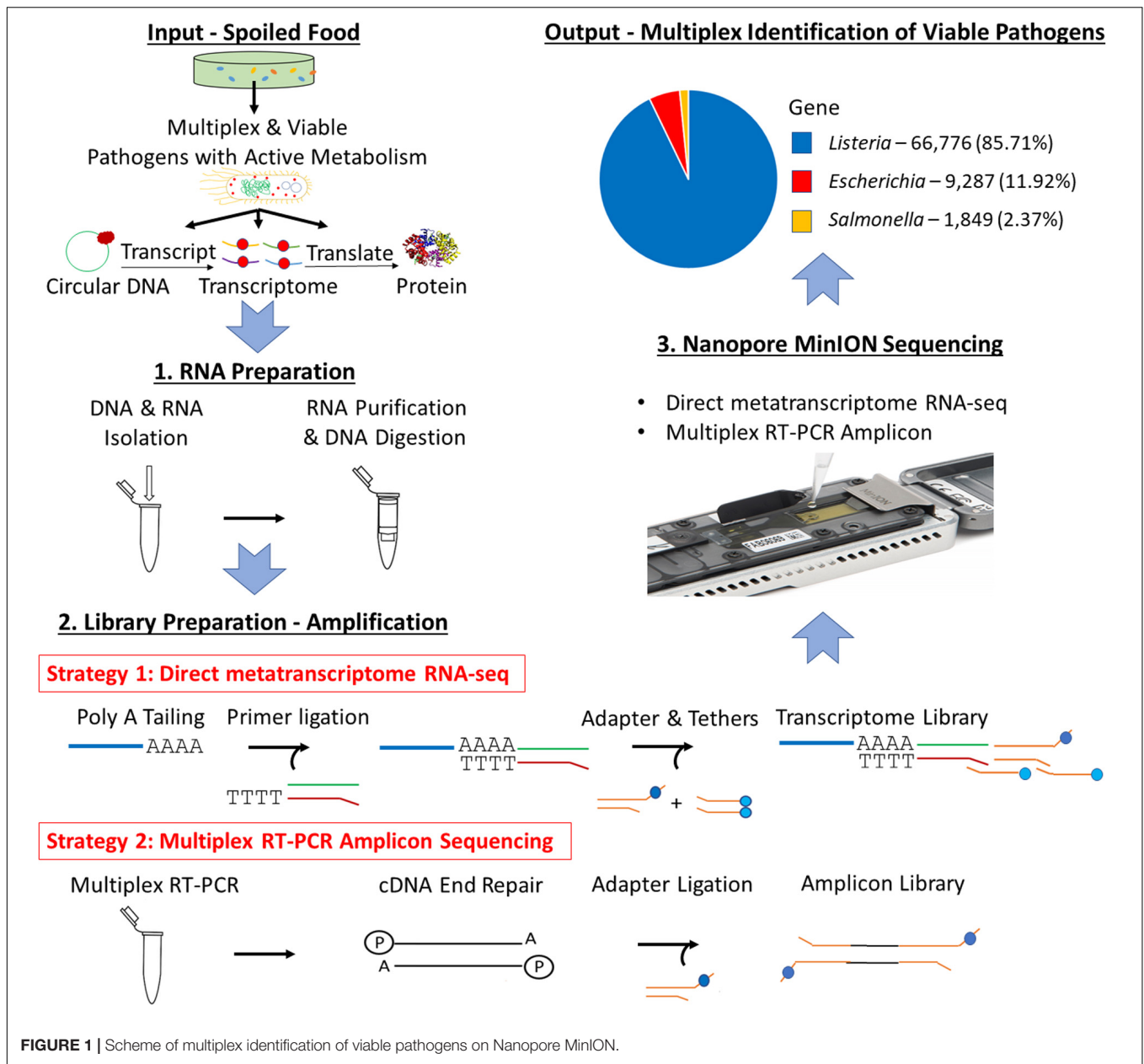
Commercial food safety testing methods include traditional plate counting methods, immunological techniques such as enzyme-linked immunosorbent assay (ELISA), lateral flow immunoassay chip, electrochemical biosensors, and chromatography, as well as nucleic acid-based approaches (Heiat et al., 2014). As summarized in **Supplementary Table 1**, the widely recognized cultivation methods and commercial rapid detection systems that are available for food defense applications have major limitations, such as large sample size, long turnaround time, and intensive labor demands. Rapid detection systems also failed to address unique changes to food safety and food defense, despite the recent success in medical and clinical diagnostics.

There are two major unique challenges in food defense, especially in identifying biohazards in food. First, viable bacteria are the etiological agents of foodborne illnesses. Most rapid detection methods have limited discretionary power to identify bacterial viability (summarized in **Supplementary Table 1**). Current genome-based technology includes polymerase chain reaction (PCR), real-time PCR (qPCR), fluorescence *in situ* hybridization (FISH), nucleic acid sequence-based amplification (NASBA), and loop-mediated isothermal amplification (LAMP). PCR based methods are sensitive and specific but easily generate cross contamination between pre-PCR and post-PCR products. Insufficient permeability of cell walls and the inherent autofluorescence of the substrate will decrease the efficiency of FISH (Baschien et al., 2008). NASBA and LAMP do not require a thermocycler, however, NASBA shows a size range limitation of target RNA and LAMP requires complex primer design that cannot be adapted to multiplex amplification (Zanoli and Spoto, 2013). In addition, all these approaches suffer from the limited capability to identify viable pathogens (Malhotra et al., 2014). False positive results are a major issue for DNA-based approaches. This is due to the inability to differentiate DNA molecules in viable bacterial cells from the genomic background, which is comprised of stable DNA molecules

from the microbiota, the food matrices, and dead pathogens inactivated during food processing and storage (Sheridan et al., 1998; Hellyer et al., 1999). In contrast, transcriptome-based technologies which utilize RNA as alternative biomarkers for bacterial viability hold more promise, because RNA molecules tend to have a shorter half-life than DNA in the environment when cells are inactivated (Sheridan et al., 1998; Hellyer et al., 1999). Recent progress was made using reverse transcription PCR (RT-PCR), but false positives also plague RT-PCR approaches (Lu et al., 2016). This can be explained by non-specific amplification of RNA molecules from food matrices and microbiota (Ju et al., 2016; Takahashi et al., 2018). Subsequent sequencing of the RT-PCR amplicons has the potential to significantly improve the accuracy of the transcriptome-based approach by identifying the origin of the amplicons.

The second prominent challenge is multiplex identification without the need for assay customization to each individual microbial threat. Each food commodity often faces multiple, and sometimes random, threats from dozens of major etiological agents (Crowe et al., 2015). A monitoring and inspection system should entail capacities of multiplex identification. Nonetheless, conventional systems depend on the customization of recognition elements, like antibodies or enzymes, to achieve multiplex detection, which can be self-prohibitory economically (Faggioli et al., 2017; Wang et al., 2019). Therefore, a feasible strategy should enable multiplex identification without the need to customize for individual threats, which can be of great importance and benefit to food defense. Several multiplex RT-PCR methods were developed for *S. aureus*, *Salmonella* and *Listeria* using food models over the last decade (Bao et al., 2008; Kawasaki et al., 2010; Ruiz-Rueda et al., 2011; Garrido et al., 2013; Salihah et al., 2016; Ding et al., 2017). However, a recent validation study suggests that multiplex RT-PCR may also generate false positive results in real food samples, especially if rRNA is the target template (Ju et al., 2016). Very recently, Next Generation Sequencing (NGS) platforms, such as Illumina, have emerged as a new strategy for food defense (Diaz-Sanchez et al., 2013; Solieri et al., 2013; Mayo et al., 2014; Moran-Gilad, 2017; Taboada et al., 2017) but, its applications in food testing are very limited. NGS does not permit timely analysis, as these platforms generate sequence reads in parallel and not in series, so data analysis can add significant burden to total turnaround time. Additionally, NGS relies on non-portable and expensive equipment, which is also economically self-prohibitory for the food industry.

The novel Oxford Nanopore MinION sequencer has emerged as a promising method of food pathogen detection based on its rapid, cost effective, portable, and high-throughput RNA and DNA sequencing workflows (Pritchard et al., 2016; Walsh et al., 2017; Hyeon et al., 2018; Taylor et al., 2019). Nanopore sequencing is a third-generation sequencing platform that can produce long reads on DNA and RNA molecules and perform real-time metagenomic and metatranscriptomic sequence analysis on the pocket-sized Nanopore MinION device (Garalde et al., 2018). This technology can be used to identify viral pathogens, as well as microorganisms such bacteria and fungi (Greninger et al., 2015; Juul et al., 2015; Cheng et al., 2018).



A few studies have demonstrated Nanopore’s potential for food safety application using metagenomic sequencing in clinical and food samples (Quick et al., 2015), however, like other genomic approaches, stable DNA molecules can cause false-positive identification, and the studies did not include a validation of whether the nanopore metagenomic sequencing data only correlates with viable pathogens. Direct RNA sequencing on Nanopore was successfully developed in 2018 on the Nanopore MinION (Garalde et al., 2018).

Therefore, for the first time, RNA-enabled Nanopore sequencing is evaluated for its potential in achieving multiplex identification of viable pathogens in this study (Figure 1). Specifically, an optimized universal RNA extraction and DNA digestion method is developed to simplify and standardize the

RNA preparation for both Gram-positive and Gram-negative bacteria. In addition, the work also includes an accuracy evaluation of different bioinformatic pipelines using whole metatranscriptome datasets, which includes MEGAN as a tool to compare the impact of rRNA on the accuracy of taxonomic analysis. Direct metatranscriptomic RNA sequencing and multiplex RT-PCR amplicon sequencing were evaluated and compared using a cocktail culture of *Escherichia coli* O157:H7 (*E. coli* O157:H7), *Salmonella enteritidis* (*S. enteritidis*), and *Listeria monocytogenes* (*L. monocytogenes*) in both standard general-purpose media and a food model. The three bacteria were selected based on their impact on economic loss or prevalence in recent outbreaks.

## RESULTS

### Verification of mRNA as Biomarkers for Bacteria Viability

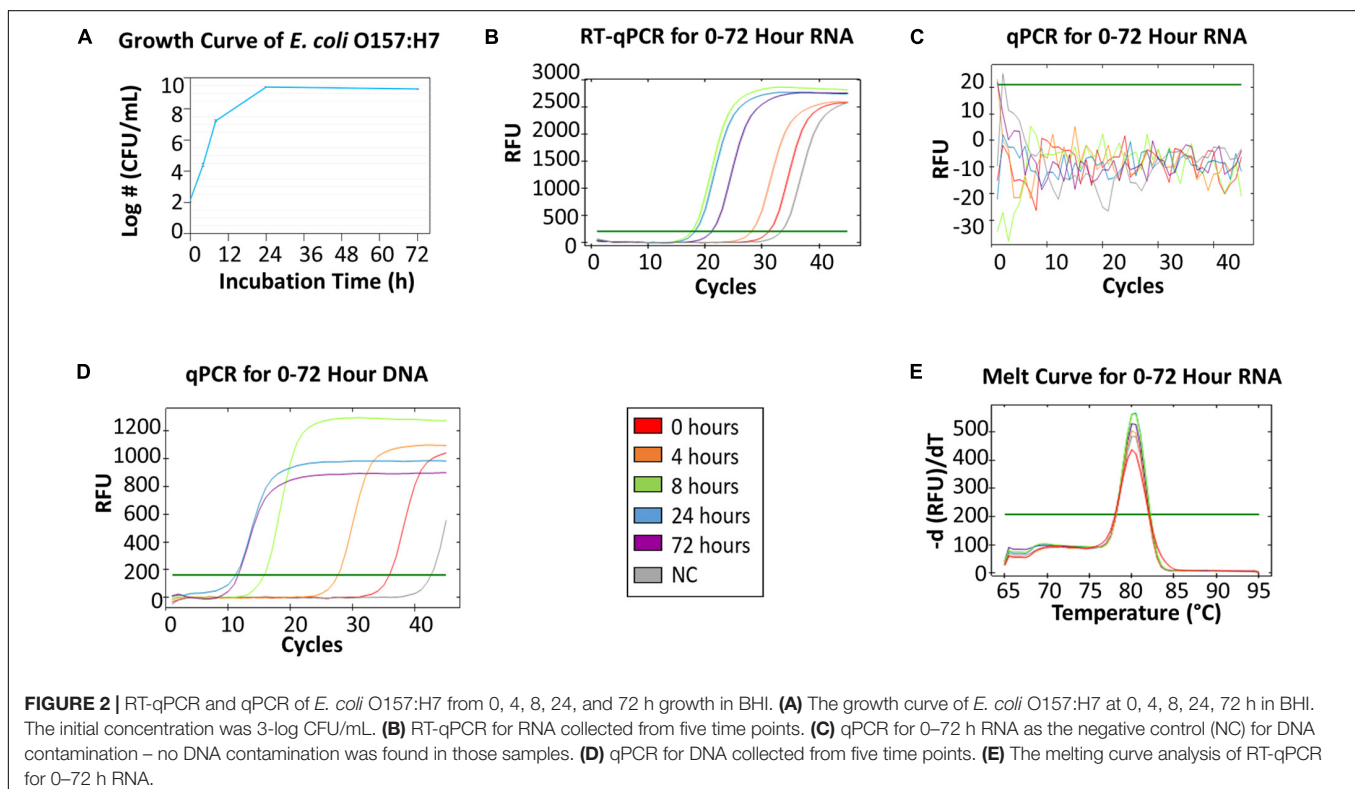
Figure 2 showed qPCR and RT-qPCR results of *E. coli* O157:H7 samples collected from 5 time points. *E. coli* O157:H7 growth curve (Figure 2A) resembles a typical microbial growth curve with an exponential phase from 0 to 24 h and a stationary phase from 24 to 72 h. Bacterial counts at 72 h showed a slight decrease from 24 h, which could indicate the start of death phase. The RT-qPCR of mRNA collected at different time points (Figure 2B) showed that the greatest amount of RNA was found in 8-h and 24-h samples, followed by a decline of mRNA concentration at 72 h. A high alignment between mRNA concentration and viable cell density can be established between Figures 2A,B. The results indicate mRNA has good correlation with viable bacterial count. The melt curve (Figure 2E) showed 5 peaks from the 5 different time points, which indicates that the same mRNA was amplified. In the negative control, no colony was identified on BHI agar and no amplicon was detected by gel electrophoresis.

The qPCR of *E. coli* O157:H7 DNA (Figure 2D) showed that the amount of DNA in 72-h samples was greater than the amount in 24-h samples, which contradicted the data from the viable bacterial counts. This indicated DNA accumulation from non-viable cells was present in 72-h samples, which was consistent with other studies (del Mar Lleó et al., 2000; Delgado-Viscogliosi et al., 2009). Hence, DNA was not a great indicator of bacteria viability. Additionally, the same qPCR amplicon was detected by gel electrophoresis in the negative control of sodium hypochlorite

treated *E. coli* O157:H7. Therefore, the results demonstrate that the global transcriptome, especially mRNA, of bacteria could be a robust indicator of cell viability.

### Direct Metatranscriptome RNA-seq on Nanopore MinION and NGS iSeq 100

Tables 1, 2 showed the results of direct metatranscriptome RNA-seq of *E. coli* O157:H7, *S. enteritidis* and *L. monocytogenes* cocktail in BHI and LJE 24-h culture using different approaches of sequencing and bioinformatics pipelines. Both EPI2ME, MG-RAST, and MEGAN miss-identified the three pathogens as other species (Table 2). MEGAN with non-rRNA mapping successfully identified the three bacteria without miss-identification as *Listeria*, *E. coli* and *Salmonella* at 91.1, 5.4, and 3.6% in BHI and 67.5, 20, and 12.5% in LJE, respectively (Table 2). The results agreed with plate counting confirmation and the growth curve of cocktail culture that all three bacteria were present (Supplementary Figure 1 and Supplementary Table 2). The mean read length was close to 1,200 bp (Table 1 and Supplementary Figure 2), which agrees with the size of 16S RNA in bacteria. The average quality scores were 83.4% in BHI and 83.7 in LJE. There was no miss-identification of the bacteria using a quality score cut-off at 7.0 (80%) using MEGAN analysis with non-rRNA mapping. No false positive identification of any bacteria was found in the negative control of sodium hypochlorite treated cocktail culture (negative control). In addition, the three pathogens were identified with genus-level in NGS for positive control, without carrying out other species (Table 2). Therefore,



**TABLE 1** | Direct metatranscriptome RNA-seq and amplicon sequencing of cocktail bacterial culture on Nanopore platform and NGS iSeq 100.

Sequencing	Sample	Flow cell chemistry	Running time (h)	Total yield (Mbases)	Reads analyzed	Passed reads	Failed reads	Mean read-length (bp)	Mean quality score
Oxford Nanopore sequencing	BHI-24 h	FLO-MIN106 R9.4 Rev D	16	487.7	412,000	409,744	2,256	1,135	83.4%
Direct metatranscriptome RNA-seq	LJE-24 h	FLO-MIN106 R9.4 Rev D	19	223.1	180,962	180,001	961	1,233	83.7%
RT-PCR amplicon	BHI-4 h	FLO-MIN106 R9.4 Rev D	3	15.6	29,290	29,279	11	534	90.9%
	LJE-4 h	FLO-MIN106 R9.4 Rev D	24	200.2	442,647	442,325	322	452	89.3%
Illumina Next generation sequencing	BHI-24 h	BPC29611-1919	21.5	1590.0	15,376,000 (total reads)	10,450,360 (passing filter reads)	-	150	94.2% (Avg%Q30)
	LJE-24 h								

the results strongly support that direct metatranscriptome RNA-seq on Nanopore MinION can achieve multiplex identification of viable pathogens.

### Multiplex RT-PCR Amplicon Sequencing on Nanopore MinION

Similarly, multiplex RT-PCR amplicon sequencing also successfully identified the three bacteria in the 4-h cocktail culture sample (Figure 3). *E. coli* O157:H7, *S. enteritidis* and *L. monocytogenes* were observed in a real-time phylogenetic tree generated by EPI2ME in less than 15 min and the distribution was, respectively, 84.7, 13.7, and 1.7% in BHI sample, and 50.9, 43.4, and 5.7% in LJE sample (Figures 3A,B and Table 2). The average read quality was 90.9 and 89.3%. A total of 29,279 reads were analyzed in BHI culture with a 3-h running time (early termination due to high quality score) and 442,325 reads in LJE cocktail culture with a 24-h running time (Table 1). The average sequence length was 534 bp in BHI sample and 432 bp in LJE sample (Supplementary Table 3 and Supplementary Figure 2) which was consistent with multiplex RT-PCR products (520, 244, and 153 bp) (van der Velden et al., 2000; Vázquez-Boland et al., 2001; Xiao et al., 2012; Nguyen et al., 2016). No false positive identification of any bacteria was found in the negative control of sodium hypochlorite treated cocktail culture.

### Quality Control and Comparison of Bioinformatic Pipelines

In this study, the quality score of direct metatranscriptome RNA-seq was 83.4 and 83.7%, while 90.9 and 89.3% for multiplex RT-PCR amplicon sequencing from MinKNOW QC report (Table 1). Raw data was collected by nanopore real-time sequencing software MinKNOW and analyzed with different bioinformatic databases and pipelines.

In metatranscriptomic direct RNA-seq, MinKNOW misidentified *Bacillus* as the top genus in both the BHI and LJE samples (Table 2). This error could be caused by the similarity between *Listeria* and *Bacillus*, especially with their housekeeping genes and rRNA (Borezee et al., 2000; Ferreira et al., 2004). Although MG-RAST eliminated this misreading, other untargeted bacteria counted for a close proportion to *S. enteritidis* (1.5%) (Table 2). MEGAN was able to eliminate other untargeted bacteria except *Bacillus* (still misidentified as 63.6%), which again was likely caused by rRNA or other housekeeping genes. Therefore, MEGAN with non-rRNA mapping was performed and successfully identified all three primary bacteria of *Listeria*, *E. coli* O157:H7 and *Salmonella* without any miss-identification (Table 2).

The results of multiplex RT-PCR sequencing showed that three targeted bacteria were anchored accurately by MinKNOW (Table 2), and no miss-identification appeared in the results.

### Gel Electrophoresis of RT-PCR and PCR Amplicon

RT-PCR was used to verify the presence of all three bacteria in the cocktail culture, and PCR was used to verify the complete removal of DNA contamination using the protocol described above.

**TABLE 2 |** Results of MinION R9.4 Rev D direct metatranscriptome RNA-seq and RT-PCR amplicon sequencing, as well as NGS iSeq 100 i1 system for BHI and LJE samples collected from 4-hour and 24-hour culture with different initial growth concentration via different bioinformatic pipelines\*.

Sequencing	Sample	Bioinformatic pipeline										
		MinKNOW/EPI2ME		MG-RAST		MEGAN		MEGAN-rRNA excluded		Galaxy and NCBI BLAST		
		Taxon	Cumulative reads	Taxon	Cumulative reads	Taxon	Percentage	Taxon	Percentage	Taxon	Cumulative reads	
Oxford Nanopore sequencing	BHI-24h	<i>Bacillus</i>	65,746 (37.6%)	<i>Listeria</i>	75,562 (85.28%)	<i>Bacillus</i>	63.6%	<i>Listeria</i>	91.1%	–	–	
		<i>Listeria</i>	54,140 (31.0%)	<i>Escherichia</i>	4,560 (5.15%)	<i>Listeria</i>	29.5%	<i>Escherichia</i>	5.4%	–	–	
		<i>Escherichia</i>	29,312 (16.8%)	<i>Salmonella</i>	1,281 (1.45%)	<i>Escherichia</i>	4.5%	<i>Salmonella</i>	3.6%	–	–	
		<i>Lactobacillus</i>	10,718 (6.1%)	<i>Coptotermes</i>	1,237 (1.40%)	<i>Salmonella</i>	2.3%	–	–	–	–	
		<i>Staphylococcus</i>	8,772 (5.0%)	<i>Bacillus</i>	952 (1.07%)	–	–	–	–	–	–	
		<i>Salmonella</i>	6,155 (3.5%)	<i>Tetragenococcus</i>	435 (0.49%)	–	–	–	–	–	–	
	LJE-24h	<i>Bacillus</i>	25,434 (34.5%)	<i>Listeria</i>	66,776 (76.55%)	<i>Bacillus</i>	63.6%	<i>Listeria</i>	67.5%	–	–	
		<i>Escherichia</i>	23,084 (31.3%)	<i>Escherichia</i>	9,287 (10.65%)	<i>Listeria</i>	29.5%	<i>Escherichia</i>	20.0%	–	–	
		<i>Listeria</i>	16,751 (22.7%)	<i>Salmonella</i>	1,849 (2.12%)	<i>Escherichia</i>	4.5%	<i>Salmonella</i>	12.5%	–	–	
		<i>Salmonella</i>	4,943 (6.7%)	<i>Bacillus</i>	1,028 (1.18%)	<i>Salmonella</i>	2.3%	–	–	–	–	
		<i>Edwardsiella</i>	3,515 (4.8%)	<i>Lactobacillus</i>	728 (0.83%)	–	–	–	–	–	–	
		<i>Escherichia</i>	5,770 (84.7%)	<i>Escherichia</i>	1,342 (70.19%)	–	–	–	–	–	–	
	RT-PCR amplicon	BHI-4h	<i>Salmonella</i>	931 (13.7%)	<i>Salmonella</i>	474 (24.79%)	–	–	–	–	–	–
			<i>Listeria</i>	113 (1.7%)	<i>Enterobacter</i>	43 (2.25%)	–	–	–	–	–	–
			–	–	<i>Listeria</i>	20 (1.05%)	–	–	–	–	–	–
LJE-4h		<i>Escherichia</i>	77,437 (50.9%)	<i>Escherichia</i>	1,852 (88.11%)	–	–	–	–	–	–	
		<i>Salmonella</i>	66,098 (43.4%)	<i>Listeria</i>	179 (8.52%)	–	–	–	–	–	–	
		<i>Listeria</i>	8,711 (5.7%)	<i>Enterobacter</i>	45 (2.14%)	–	–	–	–	–	–	

(Continued)

TABLE 2 | Continued

Sequencing	Sample	Bioinformatic pipeline									
		MinKNOW/EPI2ME		MG-RAST		MEGAN		MEGAN-rRNA excluded		Galaxy and NCBI BLAST	
		Taxon	Cumulative reads	Taxon	Cumulative reads	Taxon	Percentage	Taxon	Percentage	Taxon	Cumulative reads
Illumina Next generation sequencing Direct metatranscriptome RNA-seq	BHI-24h	-	-	<i>Listeria</i>	163,627 (91.8%)	-	-	-	-	<b><i>Escherichia</i></b>	<b>19,207 (52.8%)</b>
		-	-	<i>Escherichia</i>	13,706 (7.7%)	-	-	-	-	<i>Listeria</i>	<b>9,817 (27.0%)</b>
		-	-	<i>Salmonella</i>	899 (0.5%)	-	-	-	-	<i>Salmonella</i>	<b>7,339 (20.2%)</b>
	LJE-24h	-	-	<i>Listeria</i>	153,815 (92.9%)	-	-	-	-	<i>Escherichia</i>	<b>18,387 (54.4%)</b>
		-	-	<i>Escherichia</i>	9,988 (6.0%)	-	-	-	-	<i>Listeria</i>	<b>9,136 (27.0%)</b>
		-	-	<i>Salmonella</i>	1,732 (1.0%)	-	-	-	-	<i>Salmonella</i>	<b>6,279 (18.6%)</b>

\*Bold indicates the optimal bioinformatics pipeline without miss-identification.

**Supplementary Figure 3A.** shows the verification of 24-h LJE cocktail culture, which was used in metatranscriptomic direct RNA-seq. The results showed the RT-PCR product of three expected bands for *stx*, *invA* and *inlA* in the 24-h cocktail culture, which confirms the presence of all three target pathogens. No bands appeared on negative controls using only PCR without the RT step, which indicates the absence of DNA contamination.

Further validation was performed for the 4-h LJE cocktail culture sample, which was used in the multiplex RT-PCR amplicon sequencing. **Supplementary Figure 3B.** shows multiplex RT-PCR amplicon with three bands. The sizes of the amplicons are consistent with previous reports of 520 (*stx*), 244 (*invA*) and 153 (*inlA*) bp (**Supplementary Figure 3B** line 2). No RT-PCR products were detected in the negative control (line 3, 4, and 5) using only PCR without the RT step, which indicates that there was no DNA contamination in the sample.

## DISCUSSION

### Viability and Multiplex Identification

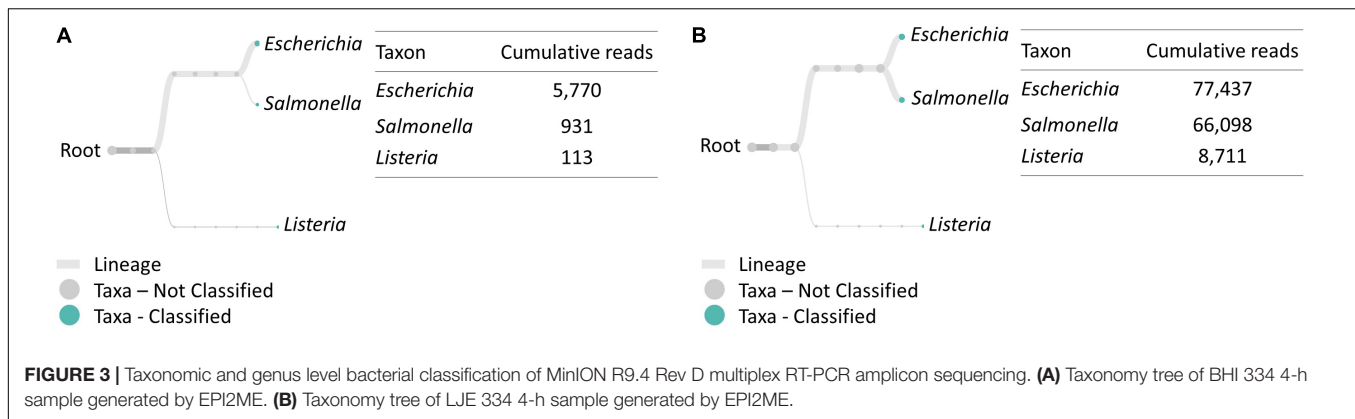
In this study, RNA-enabled Nanopore sequencing is evaluated, for the first time, for its potential in achieving multiplex identification of viable pathogens. The optimized universal RNA extraction and DNA digestion method was developed to simplify and standardize the RNA preparation for both Gram-positive and Gram-negative bacteria. Direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing were evaluated and compared using a cocktail culture of *E. coli* O157:H7, *S. enteritidis*, and *L. monocytogenes* in both standard general-purpose media and a food model.

False positives are a major issue for DNA-based approaches. This is due to the inability to differentiate DNA molecules in viable bacterial cells from the genomic background, which is comprised of stable DNA molecules from the microbiota, the food matrices, and dead pathogens inactivated during food processing and storage. Both approaches developed in this study only utilize RNA, especially mRNA, as the ultimate sequencing target, which eliminated false positive identification typically caused by DNA contamination.

Random and unknown threats from multiple infectious bacteria poses a significant threat to the safety and security of food supplies worldwide. A feasible strategy should enable multiplex identification without the need to customize for an individual threat. Therefore, the developed universal protocol is applicable to both Gram-positive and Gram-negative bacteria. RNA from multiple pathogens in one food sample can be collected from one extraction and library preparation step, followed by the universal sequencing protocol.

### Comparison of Direct Metatranscriptome RNA-seq and Multiplex RT-PCR Amplicon Sequencing

The developed method successfully identified all three bacteria from cocktail culture in BHI and LJE by direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon



sequencing. Nonetheless, the two sequencing approaches entail different capacities and challenges. Direct metatranscriptome RNA-seq does not require assay customization for an individual biohazard if the bioinformatic database includes the target microbiota. Multiplex RT-PCR amplicons comprise the target gene copies, and they can be easily captured by the motor membrane protein when passing through the nanopores. As a result, it shows higher accuracy, greater quality score, better quality control, and less turnaround time.

The two strategies result in different read length. In this study, we extracted total bacterial RNA that is comprised of a majority of rRNA and a small number of mRNA and tRNA for nanopore sequencing. The bioanalyzer results showed that the majority of RNA from *E. coli* O157:H7, *S. enteritidis* and *L. monocytogenes* cocktail culture was 16S RNA with 1250–2100 nucleotides, and 23S RNA with 2250–3950 nucleotides, respectively. Direct metatranscriptome RNA-seq sequencing successfully identified all three bacteria. The RNA read length ranged from 0 to 3000 nucleotides, with the most abundant read length between 400–1600 nucleotides. The read length from direct metatranscriptome RNA-seq is approximately the full length of RNA (Cho et al., 2014; Byrne et al., 2017). Thus, this method can provide approximal full-length RNA sequence. However, in multiplex RT-PCR amplicon sequencing, the amplicons for each strain have different expected sizes, which were observed in the sequencing read length results. In the real-time analysis of multiplex RT-PCR amplicon sequencing, all three bacteria were identified within 15 min and the resulting read lengths were 510 bp (*stx*), 244 bp (*invA*) and 153 bp (*inlA*), respectively. Some reports suggest that Nanopore excels in long RNA reads up to thousands of nucleotides, and sequencing of short reads tends to be more challenging due to their higher and non-uniform error profiles, which might result in a large fraction of reads remaining unmapped or unused (Grabherr et al., 2011; Madoui et al., 2015; de Lannoy et al., 2017). However, amplicon sequencing showed less error than metatranscriptomic direct RNA-seq. Multiplex RT-PCR amplicon sequencing successfully identified all three target bacteria using MinKNOW (Table 2) in real time. Direct metatranscriptome RNA-seq could experience issues of miss-identification when the assay is plagued by using

a less efficient library preparation or choosing inappropriate bioinformatic pipelines. Therefore, more work is warranted to improve library preparation and bioinformatic pipelines for practical applications.

Both methods are comparable in their total turnaround time. Direct metatranscriptome RNA-seq does not include an additional RT-PCR step, but the library preparation, bioinformatic analysis, and mapping could easily offset the time difference. The total turnaround time for direct metatranscriptome RNA-seq is approximately 6.5 h, which includes RNA purification (3.5 h), library preparation (1.5 h), Nanopore sequencing (1 h), and bioinformatic analysis (0.5 h). Multiplex amplicon sequencing takes approximately 6 h, which includes RNA purification (3.5 h), RT-PCR (2 h), library preparation (0.5 h), Nanopore sequencing (15 min), and bioinformatics analysis (0.5 h).

Multiplex RT-PCR amplicon sequencing requires substantially less RNA input, which could translate into less microbial input. The method only requires 36.5 ng RNA input for multiplex RT-PCR, and 33.8 ng amplicon for library preparation and sequencing (Supplementary Table 2). The amplicon sequencing method is more sensitive and could be applicable for food commodities with low bacterial loading around  $10^1$ – $10^4$  CFU/g. 500 ng RNA input on Nanopore MinION is recommended by the supplier for metatranscriptomic direct RNA-seq. However, significant RNA loss was observed during the library preparation due to the three purification steps. The initial purified RNA concentration before library preparation was 3490 ng and 1338 ng in BHI and LJE, respectively, and only 744 and 130 ng were yielded for Nanopore sequencing. RNA loss can be as high as 80–90%, which significantly restricted sensitivity of the assay. Redesign of the library preparation protocol to minimize RNA loss can have profound significance for assay sensitivity and feasibility for clinical applications.

The two strategies pose different levels of complexities. Direct metatranscriptome RNA-seq may be applicable in foods with a complex microbiome (e.g. cultured food). Direct metatranscriptome RNA-seq does not require assay customization for an individual biohazard, if the bioinformatic



database includes the target microbiota. The multiplex RT-PCR amplicon sequencing requires complex primer design and validation. Not all RT-PCR primers work in multiplex RT-PCR, due to potential primer interaction, non-specific amplification, and amplification bias. The amplicon sequencing may be more suitable for high-throughput and continuous monitoring of foodborne pathogens with high risk factors.

## Comparison Between Different Bioinformatic Pipelines

Bioinformatic analysis has significant impact to the accuracy of Nanopore sequencing. Different computational pipelines of the same nanopore data may lead to different results. Normally, MinION pipeline contains primer trimming, alignment, variant calling and consensus generation (Loman and Quinlan, 2014; Wood and Salzberg, 2014; Menzel et al., 2016; Kerkhof et al., 2017), and EPI2ME conducts real-time surveillance of nanopore sequencing. First, reads containing raw data are base called by MinKNOW, and then extracted into a FASTQ file for mapping to reference transcriptome or genome (Li and Durbin, 2010; Mitsuhashi et al., 2017; Rang et al., 2018), aligned to sequence via primer trimming and coverage normalization. During this process, low quality or low coverage reads (read hit) are filtered out to generate final sequence for BLAST in NCBI<sup>1</sup>. MinION chemistry provides a simplified and rapid report of nanopore running, including read number, read length, cumulative read, taxonomy tree and quality control. MEGAN and MG-RAST are popular software or service for metagenome or metatranscriptome analysis. The similarity between them is that they perform computational analysis of multiple datasets for taxonomic content based on family and genus level. In contrast, MEGAN is able to perform taxonomical, functional and interactive analyses, which is the comparison of taxonomic and functional contents based on the SEED hierarchy and KEGG pathways (Borthong et al., 2018; Huson et al., 2018). In this study, MG-RAST was used in taxonomic analysis, and MEGAN was selected for mRNA analysis by removing rRNA from the whole metatranscriptome datasets. Taxonomical classification using MEGAN on non-rRNA data showed higher fidelity to types of pathogens inoculated in the sample and was consistent with our separate plate count and NGS validation. Moreover, both MG-RAST and MEGAN showed higher accuracy compared with EPI2ME for taxonomical classification using the whole metatranscriptome data. The results agree with previous reports on bioinformatic analysis using non-rRNA, because different bacteria can have very similar housekeeping genes in forms of rRNA, which leads to errors in taxonomical classification (Klappenbach et al., 2000; Schmieder et al., 2011). Multiplex RT-PCR amplicon sequencing obtained a rapid and accurate taxonomic content because this nanopore sequencing method poses a high sensitivity. In addition, adequate and complete BLAST database may further improve the accuracy, rapidness, and quality for the multiplex identification of viable pathogens in food.

<sup>1</sup> Available at: <https://www.ncbi.nlm.nih.gov> (accessed January 6, 2019).

## CONCLUSION

Novel strategies are developed through direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing on Nanopore MinION to achieve real-time multiplex identification of viable pathogens in food. This study reports an optimized universal Nanopore sample extraction and library preparation protocol applicable to both Gram-positive and Gram-negative bacteria. Further evaluation and validation confirmed the accuracy of direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing using Sanger sequencing and selective media. The result of NGS (positive control) from this study confirmed the identification of the three pathogens performed by Nanopore direct metatranscriptome RNA-seq, even though NGS had time-consuming and cost prohibitive limitations. The study also included a comparison of different bioinformatic pipelines for metatranscriptomic and amplicon genomic analysis. In addition, direct metatranscriptome RNA-seq and RT-PCR amplicon sequencing were compared for their respective advantages in sample inputs, accuracy, sensitivity, and time effectiveness for potential applications.

Both direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing require further development to address some pressing challenges. (A) Optimization of direct metatranscriptome RNA-seq sequencing may include minimizing RNA loss in the library preparation step; comparison of bioinformatic pipelines to eliminate miss-identified and unclassified targets; cross-domain identification of prokaryotes, eukaryotes, and viruses. This is especially important in complex mixtures. For example, less RNA loss during library preparation often leads to increased RNA yield and input on Nanopore device, which improves data quality and accuracy of the taxonomical identification. (B) Multiplex RT-PCR amplicon sequencing can benefit from: multiplex primer development; inclusivity/exclusivity evaluations; reduced amplification bias.

To the best of our knowledge, this is the first report of metatranscriptome sequencing of cocktail microbial RNAs on the emerging Nanopore platform. Direct RNA-seq and RT-PCR amplicon sequencing of the metatranscriptome enable the direct identification of nucleotide analogs in RNAs, which is highly informative for determining microbial identities while detecting ecologically relevant processes. The information pertained in this study could be important for future revelatory research including predicting antibiotic resistance, elucidating host-pathogen interactions, prognosing disease progression, and investigating microbial ecology, etc.

## MATERIALS AND METHODS

### Bacterial Strains and Culturing

*Escherichia coli* O157:H7 (ATCC 43895), *S. enteritidis* (ATCC 13076), and *L. monocytogenes* (ATCC 19115) were acquired from ATCC (Manassas, VA). The three bacteria were cultured using Brain Heart Infusion (BHI) broth and agar (BD, Franklin Lakes, NJ, United States) at 37°C for 24 h either in separate individual cultures or in cocktail cultures. Romaine lettuce (*Lactuca sativa*

*L. var. longifolia*) juice extract (LJE) was used as a food model in this study. Romaine LJE was prepared according to our previous publications (Shen et al., 2012). Briefly, 250 g fresh Romaine lettuce heart (Fresh Express) and 200 ml DI Water was blended in a Waring 7011G Commercial Blender for 1 min. The blended mixture was then filtered through Büchner funnel with P5 filter paper. The filtrate was centrifuged at  $2300 \times g$  for 10 min (low speed centrifugation), the supernatant from low speed centrifugation was then centrifuged at  $3200 \times g$  for 30 min (high speed centrifugation). High speed centrifugation supernatant was then filtered through 0.2 micron filter membrane (vacuum filter 0.2-micron, Thermo Fisher Scientific) and diluted to 4% using sterilized DI water (COD = 800 ppm) to grow bacteria.

Cocktail culture of *E. coli* O157:H7, *S. enteritidis*, and *L. monocytogenes* in BHI or LJE were obtained by inoculating appropriate volume of 24-h stock culture of the individual bacteria to achieve the initial concentration shown in **Supplementary Table 2**. The concentration for each bacteria was determined by plate counting methods using selective agars. Oxford *Listeria* selective agar base (Oxford formulation) with Oxford modified *Listeria* selective supplement was used for the selective quantification of *L. monocytogenes*. MacConkey agar (BD, Franklin Lakes, NJ, United States) was used to differentiate and quantify *E. coli* O157:H7 and *S. enteritidis*. Both cultures were incubated at 37°C for 24 h before quantification using an automated plate counter (Scan 300, Interscience Laboratories Inc., Woburn, MA, United States).

## Verification of mRNA as Biomarkers for Bacteria Viability

*Escherichia coli* O157:H7 was selected as a model organism to demonstrate that mRNA is a valid indicator for bacteria viability. Aliquot of overnight *E. coli* O157:H7 culture was inoculated at 3-log CFU/mL in BHI broth and incubated at 37°C. Culture was sampled at 0, 4, 8, 24, and 72 h, fractions from the sample culture were taken and plated on BHI agar to determine viable bacteria counts. Remaining fractions were used for DNA and RNA purification using DNeasy blood and tissue kit (Qiagen, Germantown, MD, United States) and Monarch Total RNA Miniprep Kit (New England Biolabs, Ipswich, MA, United States) following the supplier protocols, respectively. In RNA preparation, to lyse the cells, the cell pellet obtained from initial centrifugation was incubated at 37°C for 1 h with 300 rpm mixing in 250  $\mu$ L 3 mg/mL lysozyme (Alfa Aesar, Haverhill, MA, United States) in Tris-EDTA buffer (Sigma-Aldrich, St. Louis, MO, United States). Purified DNA and RNA from these four time points were quantified using Qubit dsDNA HS Assay Kit and Qubit RNA HS Assay Kit (Invitrogen, Carlsbad, CA, United States), and also quantified using NEB Luna Universal qPCR Master Mix and Luna Universal One-Step RT-qPCR Kit (New England Biolabs, Ipswich, MA, United States) following supplier protocols. A Biorad CFX-96 Touch real time PCR detection system was used for qPCR and RT-qPCR testing. The primer pairs used in this test were designated as Stx1A and sequence was listed in **Supplementary Table 3**. *E. coli* O157:H7 inactivated with 13.4 mmol/L of sodium hypochlorite was used

as the negative control to test whether mRNA and/or DNA can be used as viability biomarkers (Skinner et al., 2018).

## Direct Metatranscriptome RNA-seq on Nanopore MinION

One dimensional direct metatranscriptome RNA-seq was performed using 24-h cocktail culture of *E. coli* O157:H7, *S. enteritidis*, and *L. monocytogenes* in BHI or LJE. RNA was extracted from the cocktail culture by Monarch Total RNA Miniprep Kit including DNase I (NEB, Ipswich, MA, United States) which was confirmed by multiplex RT-PCR and gel electrophoresis. DNA was completely digested using DNase I (NEB # T2010S, working concentration: 0.1 U/ $\mu$ l), which was verified by multiplex PCR and gel electrophoresis. The primers stx, invA and LisA2 were selected for *E. coli* O157:H7, *S. enteritidis*, and *L. monocytogenes*, respectively. RT-PCR products were analyzed using gel electrophoresis with 1.2% agarose gel. The One Taq One-Step RT-PCR Kit (NEB # E5310S) was used for nucleic acid amplification. The thermal cycler condition: reverse transcription at 48°C for 15 min; initial denaturation at 94°C for 1 min; denaturation at 94°C for 15 s, annealing at 53°C for 30 s, extension at 68°C for 40 s with 40 cycles; final extension at 68°C for 5 min.

The prepared RNA samples were further modified with poly(A) tailing and library preparation by following suppliers' protocols. Direct metatranscriptome RNA-seq was developed based on supplier's direct RNA-seq protocol (RNA Kit SQK-RNA001, Oxford Nanopore Technologies, Oxford, United Kingdom). The MinION flow cell was primed using a priming mix, and then 75  $\mu$ l of sample was loaded to the SpotON sample port dropwise to avoid bubbles. After adding the sample, MinKNOW software was initiated to start a sequencing run. Cocktail culture inactivated with 13.4 mmol/L of sodium hypochlorite was used as the negative control to test whether metatranscriptome sequencing can eliminate false positive identification.

## Direct Metatranscriptome RNA-seq on NGS iSeq 100 Sequencing System for Control

RNA extracts of 24-h cocktail culture of *E. coli* O157:H7, *S. enteritidis*, and *L. monocytogenes* in BHI or LJE was obtained by NEB Monarch Total RNA Miniprep Kit including DNase I as we did for direct metatranscriptome RNA-seq on a MinION sequencer (Oxford Nanopore). The quantification and quality control of samples were tested by Qubit RNA HS Assay (Thermo Fisher Scientific) and 2100 Bioanalyzer system (RNA 6000 Pico Kit/High Sensitivity DNA Kit and 2100 Expert Software, Agilent, Santa Clara, CA, United States). NEBNext Ultra II Directional RNA Library Prep Kit (NEB # E7760), NEBNext rRNA Depletion Kit (NEB # E6310) and NEBNext Multiplex Oligos for Illumina (Set 2, NEB # E7500) were used for NGS library prep.

RNA samples from BHI (quadruplicate) were developed by following the NEB Protocol for use with FFPE RNA, NEBNext rRNA Depletion Kit (Human/Mouse/Rat) (E6310) and NEBNext

Ultra II Directional RNA Library Prep Kit for Illumina (E7760, E7765) as their RNA Integrity Number (RIN) was equal to 1 to 2, while samples from LJE (quadruplicate) were treated by following NEB Protocol for use with NEBNext rRNA Depletion Kit (Human/Mouse/Rat) (NEB # 6310) as their RIN was above 7. Briefly, 100 ng of FFPE RNA was diluted in 12  $\mu$ l of Nuclease-free Water for RNA probe hybridization. Twelve microliters of total RNA was mixed with 1  $\mu$ l of NEBNext rRNA Depletion Solution and 2  $\mu$ l of Probe Hybridization Buffer by following the thermal condition: heated lid at 105°C; 2 min at 95°C; ramp down to 22°C at 0.1°C/s; 5 min hold at 22°C. After that, RNase H Digestion and DNase I Digestion were performed. First, 15  $\mu$ l of sample were mixed with 5  $\mu$ l of RNase H master mix (2  $\mu$ l of NEBNext RNase H, 2  $\mu$ l of NEBNext RNase H Reaction Buffer, 1  $\mu$ l of Nuclease-free Water) and incubated for 30 min at 37°C (heated lid at 40°C), and then 20  $\mu$ l of sample was added to 30  $\mu$ l of DNase I digestion master mix (5  $\mu$ l of DNase I Reaction Buffer, 2.5  $\mu$ l of DNase I, 22.5  $\mu$ l of Nuclease-free Water) and incubated for another 30 min with the same condition. After RNA purification by beads (NEBNext RNA Sample Purification Beads, NEB # E7104S, 2.2X beads), 5  $\mu$ l of sample with Nuclease-free Water was reacted with 1  $\mu$ l of random primers for 5 min at 65°C (heated lid at 105°C, and hold at 4°C) for priming. The 6  $\mu$ l of primed RNA was mixed with 4  $\mu$ l of NEBNext First Strand Synthesis Reaction Buffer, 8  $\mu$ l of NEBNext Strand Specificity Reagent and 2  $\mu$ l of NEBNext First Strand Synthesis Enzyme Mix to form the first strand cDNA (thermal condition: heated lid at 80°C, 10 min at 25°C, 15 min at 42°C, 15 min at 70°C and hold at 4°C), followed by second strand cDNA synthesis by reacting with 8  $\mu$ l of NEBNext Second Strand Synthesis Reaction Buffer with dUTP, 4  $\mu$ l of NEBNext Second Strand Synthesis Enzyme Mix and 48  $\mu$ l of Nuclease-free Water for 1 h at 16°C. After cDNA synthesis, a bead purification (1.8X) was performed as before and 50  $\mu$ l of purified cDNA was eluted. End Prep of cDNA was carried out in a reaction of 50  $\mu$ l of Second Strand Synthesis Product, 7  $\mu$ l of NEBNext Ultra II End Prep Reaction Buffer, 3  $\mu$ l of NEBNext Ultra II End Prep Enzyme Mix (thermal condition: heated lid at 75°C, 30 min at 20°C, 30 min at 65°C and hold at 4°C), and Adaptor Ligation in a reaction of 60  $\mu$ l of End Prepped DNA, 2.5  $\mu$ l of 25-fold Diluted Adapter, 1  $\mu$ l of NEBNext Ligation Enhancer and 30  $\mu$ l of NEBNext Ultra II Ligation Master Mix with incubation for 15 min at 20°C, followed by another 15 min incubation at 37°C by adding 3  $\mu$ l of USER Enzyme. DNA libraries were purified with 0.9X DNA purification beads. Finally, 15  $\mu$ l of adaptor ligated DNA, 25  $\mu$ l of NEBNext Ultra II Q5 Master Mix, 5  $\mu$ l of Universal PCR Primer/i5 Primer, 5  $\mu$ l of Index (X) Primer/i7 Primer (NEBNext Multiplex Oligos for Illumina Set 2, NEB # E7500) were mixed together for PCR Enrichment (thermal condition: initial denaturation at 98°C for 30 s; denaturation at 98°C for 10 s, annealing/extension at 65°C for 75 s with 16 cycles; final extension at 65°C for 5 min with 16 cycles). After PCR enrichment, 0.9X beads were used to purify DNA and 0.1X TE was used to elute DNA for Assess Library Quality on Agilent 2100 Bioanalyzer. RNA samples from LJE and BHI samples were treated the same, however, the LJE samples had

an additional step of RNA Fragmentation via mixing 5  $\mu$ l of sample with 4  $\mu$ l of NEBNext First Strand Synthesis Reaction Buffer and 1  $\mu$ l of random primers, following the incubation at 94°C for 15 min.

Library quantification was done on CFX-96 Touch system (Bio-Rad Laboratories, Hercules, CA, United States) using NEBNext Library Quant Kit (NEB # E7630) by following NEBNext Library Quant Kit Protocol. Libraries were diluted at 1:1,000, 1:10,000 and 1:100,000 and 4  $\mu$ l of template was used in each 20  $\mu$ l total reaction volume. The thermal condition: initial denaturation at 95°C for 1 min; denaturation at 95°C for 15 s with 35 cycles; extension at 63°C for 45 s. The calculation of quantitative library DNA was achieved by using NEBioCalculator v1.10.0.

An iSeq 100 system (Illumina, San Diego, CA, United States) was used to sequence 24-h cocktail cultures of BHI-derived and LJE-derived *E. coli* O157:H7, *L. monocytogenes*, and *S. enteritidis* as a control dataset. All eight libraries were normalized to a 1 nM concentration and 5  $\mu$ l of library pool was diluted to a 50 pM loading concentration. PhiX was spiked-in to the library pool at 10%. This value was used in order to increase base diversity. Paired-end reads were generated at 151 cycles each with a 6-nucleotide indexing read.

## Multiplex RT-PCR Amplicon Sequencing on Nanopore MinION

Multiplex RT-PCR amplicon sequencing was conducted using 4-h cocktail culture of *E. coli* O157:H7, *S. enteritidis*, and *L. monocytogenes* in BHI or LJE. RNA extraction and DNA digestion were performed and verified using the same protocols listed above. Major virulent genes selected in this study include: *stx* and *stx1A* localized to the lambdoid prophages H19B and H19J in *E. coli* O157:H7 (Zhang et al., 2000; Toma et al., 2003; Nguyen et al., 2016); *invA*, a critical component of the *Salmonella* pathogenicity island 1 (SPI1) in *S. enteritidis* (van der Velden et al., 2000; Nguyen et al., 2016); and *inlA*, encoded internalins genes in the *inlAB* operon outside the *Listeria* pathogenicity island 1 (LIPI-1) in *L. monocytogenes* (Vázquez-Boland et al., 2001; Xiao et al., 2012; Nguyen et al., 2016).

Prior to RT-PCR amplicon sequencing, an end repair/A tailing step (NEBNext End Repair/dA-tailing Module) was carried out for the RT-PCR products, followed by a ligation step using NEB Ultra II ligation master mix. Oxford Nanopore Ligation Sequencing Kit SQK-LSK108 and Library Loading Bead Kit EXP-LLB001 were used for the library preparation of RT-PCR amplicon. To validate Nanopore DNA sequencing results, the RT-PCR products were also sequenced using an Applied Biosystems 3130xl genetic analyzer by following a protocol at NEB. The DNA sequencing data collected by the sequencer was analyzed using EPI2ME and MG-RAST to confirm the identities of the three bacteria. Cocktail culture inactivated with 13.4 mmol/L of sodium hypochlorite was used as the negative control to test whether multiplex RT-PCR amplicon sequencing can eliminate false positive identification. Detailed protocols for RT-PCR are provided in the **Supplementary Material**.

## Data Analysis and Bioinformatics for Oxford Nanopore Sequencing Run

Sequencing reads were base-called via the local base-calling algorithm with MinKNOW software (v. 1.4.3). All FASTQ files of passed base-called reads were collected and combined to one file for analysis. EPI2ME, MG-RAST and MEGAN were used for metagenomics and taxonomic analysis.

## Data Analysis and Bioinformatics for iSeq100 Control Run

Sequencing reads were demultiplexed by unique six nucleotide barcodes using the bcl2fastq program<sup>2</sup>. Sequencing reads were uploaded to Galaxy (Galaxy) and adapter sequences were trimmed using Cutadapt (Galaxy) with a quality cutoff of 20. The program trimmed the adapter sequence in both the 5' and 3' orientations. Forward and Reverse reads were joined using FASTQjoiner (Galaxy). FASTQ files were converted to FASTA files using FASTQ to FASTA program<sup>3</sup> (GalaxyGalaxy). FASTA files were then aligned and mapped on NCBI BLAST by its refseq\_rna and blastn programs. The Common Taxonomic tree and Alignment Summary Viewer tools were used to analyze alignments per bacterial species and to generate phylogenetic trees.

## AUTHOR'S NOTE

This manuscript has been released as a Pre-Print at bioRxiv (Yang et al., 2019).

<sup>2</sup> Illumina Bcl2fastq Conversion Software. Available at: [https://support.illumina.com/sequencing/sequencing\\_software/bcl2fastq-conversion-software.html](https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html) (accessed July 21, 2019).

<sup>3</sup> Available at: <https://usegalaxy.org/> (accessed July 25, 2019).

## REFERENCES

- Bao, H.-F., Li, D., Guo, J.-H., Lu, Z.-J., Chen, Y.-L., Liu, Z.-X., et al. (2008). A highly sensitive and specific multiplex RT-PCR to detect foot-and-mouth disease virus in tissue and food samples. *Arch. Virol.* 153, 205–209. doi: 10.1007/s00705-007-1082-y
- Baschien, C., Manz, W., Neu, T. R., Marvanova, L., and Szwedzyk, U. (2008). *In situ* detection of freshwater fungi in an alpine stream by new taxon-specific fluorescence in situ hybridization probes. *Appl. Environ. Microbiol.* 74, 6427–6436. doi: 10.1128/AEM.00815-08
- Borezee, E., Msadek, T., Durant, L., and Berche, P. (2000). Identification in *Listeria monocytogenes* of MecA, a homologue of the *Bacillus subtilis* competence regulatory protein. *J. Bacteriol.* 182, 5931–5934. doi: 10.1128/jb.182.20.5931-5934.2000
- Borthong, J., Omori, R., Sugimoto, C., Suthienkul, O., Nakao, R., and Ito, K. (2018). Comparison of database search methods for the detection of *Legionella pneumophila* in water samples using metagenomic analysis. *Front. Microbiol.* 9:1272. doi: 10.3389/fmicb.2018.01272
- Byrne, A., Beaudin, A. E., Olsen, H. E., Jain, M., Cole, C., Palmer, T., et al. (2017). Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* 8:16027. doi: 10.1038/ncomms16027
- Centers for Disease Control and Prevention (2018). *Surveillance for foodborne disease outbreaks, United States, 2016, Annual Report*. Atlanta: Centers for Disease Control and Prevention.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found here: <https://www.ncbi.nlm.nih.gov/bioproject/609733>, Accession number: PRJNA609733.

## AUTHOR CONTRIBUTIONS

MY, AC, XL, DS, SL, TG, and LS provided assistance and conducted the experiments throughout the project. MY, HD, and JL aided in NGS validation. YL, MX, and BZ provided guidance and wrote the manuscript.

## FUNDING

The project is partially supported by the U.S. Department of Agriculture (S51600000035794).

## ACKNOWLEDGMENTS

Special appreciation goes to LiTing Chiu and Tong Wu for their support in this project.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.00514/full#supplementary-material>

- Cheng, J., Hu, H., Kang, Y., Chen, W., Fang, W., Wang, K., et al. (2018). Identification of pathogens in culture-negative infective endocarditis cases by metagenomic analysis. *Ann. Clin. Microbiol. Antimicrob.* 17:43. doi: 10.1186/s12941-018-0294-5
- Cho, H., Davis, J., Li, X., Smith, K. S., Battle, A., and Montgomery, S. B. (2014). High-resolution transcriptome analysis with long-read RNA sequencing. *PLoS One* 9:e108095. doi: 10.1371/journal.pone.0108095
- Crowe, S. J., Mahon, B. E., Vieira, A. R., and Gould, L. H. (2015). Vital signs: multistate foodborne outbreaks - United States, 2010–2014. *Morb. Mortal. Wkly Rep.* 64, 1221–1225. doi: 10.15585/mmwr.mm6443a4
- de Lannoy, C., de Ridder, D., and Risse, J. (2017). The long reads ahead: de novo genome assembly using the MinION. *F1000Res* 6:1083. doi: 10.12688/f1000research.12012.2
- del Mar Lleó, M., Pierobon, S., Tafi, M. C., Signoretto, C., and Canepari, P. (2000). mRNA detection by reverse transcription-PCR for monitoring viability over time in an *Enterococcus faecalis* viable but nonculturable population maintained in a laboratory microcosm. *Appl. Environ. Microbiol.* 66, 4564–4567. doi: 10.1128/aem.66.10.4564-4567.2000
- Delgado-Viscogliosi, P., Solignac, L., and Delattre, J. M. (2009). Viability PCR, a culture-independent method for rapid and selective quantification of viable *Legionella pneumophila* cells in environmental water samples. *Appl. Environ. Microbiol.* 75, 3502–3512. doi: 10.1128/aem.02878-08
- Diaz-Sanchez, S., Hanning, I., Pendleton, S., and D'Souza, D. (2013). Next-generation sequencing: the future of molecular genetics in poultry production and food safety. *Poultry Sci.* 92, 562–572. doi: 10.3382/ps.2012-02741

- Ding, T., Suo, Y., Zhang, Z., Liu, D., Ye, X., Chen, S., et al. (2017). A multiplex RT-PCR assay for *S. aureus*, *L. monocytogenes*, and *Salmonella* spp. detection in raw milk with pre-enrichment. *Front. Microbiol.* 8:989. doi: 10.3389/fmicb.2017.00989
- Faggioli, F., Luigi, M., and Boubourakas, I. N. (2017). "Viroid amplification methods: RT-PCR, real-time RT-PCR, and RT-LAMP," in *Viroids and Satellites*, eds V. A. Hadidi, R. Flores, J. W. Randles, and P. Palukaitis (London: Academic Press), 381–391.
- Ferreira, A., Gray, M., Wiedmann, M., and Boor, K. J. (2004). Comparative genomic analysis of the sigB operon in *Listeria monocytogenes* and in other gram-positive bacteria. *Curr. Microbiol.* 48, 39–46. doi: 10.1007/s00284-003-4020-x
- Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15, 201. doi: 10.1038/nmeth.4577
- Garrido, A., Chapela, M.-J., Román, B., Fajardo, P., Lago, J., Vieites, J. M., et al. (2013). A new multiplex real-time PCR developed method for *Salmonella* spp. and *Listeria monocytogenes* detection in food and environmental samples. *Food Control* 30, 76–85. doi: 10.1016/j.foodcont.2012.06.029
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644. doi: 10.1038/nbt.1883
- Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., et al. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* 7:99. doi: 10.1186/s13073-015-0220-9
- Heiat, M., Ranjbar, R., and Alavian, S. M. (2014). Classical and modern approaches used for viral hepatitis diagnosis. *Hepat Mon* 14:e17632. doi: 10.5812/hepatmon.17632
- Hellyer, T. J., Desjardin, L. E., Hehman, G. L., Cave, M. D., and Eisenach, K. D. (1999). Quantitative analysis of mRNA as a marker for viability of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 37, 290–295. doi: 10.1128/jcm.37.2.290-295.1999
- Hoffmann, S., Maculloch, B., and Batz, M. (2015). *Economic burden of major foodborne illnesses acquired in the United States* (Washington, DC: Economic Research Service).
- Huson, D. H., Albrecht, B., Bačar, C., Bessarab, I., Górska, A., Jolic, D., et al. (2018). MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biol. Direct.* 13:6. doi: 10.1186/s13062-018-0208-7
- Hyeon, J.-Y., Li, S., Mann, D. A., Zhang, S., Li, Z., Chen, Y., et al. (2018). Quasimetagenomics-based and real-time-sequencing-aided detection and subtyping of *Salmonella enterica* from food samples. *Appl. Environ. Microbiol.* 84, e02340-17.
- Ju, W., Moyne, A. L., and Marco, M. L. (2016). RNA-based detection does not accurately enumerate living *Escherichia coli* O157:H7 cells on plants. *Front. Microbiol.* 7:223. doi: 10.3389/fmicb.2016.00223
- Juul, S., Izquierdo, F., Hurst, A., Dai, X., Wright, A., Kulesha, E., et al. (2015). What's in my pot? Real-time species identification on the MinION™. *bioRxiv* [preprint]. doi: 10.1101/030742
- Kawasaki, S., Fratamico, P. M., Horikoshi, N., Okada, Y., Takeshita, K., Sameshima, T., et al. (2010). Multiplex real-time polymerase chain reaction assay for simultaneous detection and quantification of *Salmonella* species, *Listeria monocytogenes*, and *Escherichia coli* O157:H7 in ground pork samples. *Foodborne Pathog. Dis.* 7, 549–554. doi: 10.1089/fpd.2009.0465
- Kerkhof, L. J., Dillon, K. P., Häggblom, M. M., and McGuinness, L. R. (2017). Profiling bacterial communities by MinION sequencing of ribosomal operons. *Microbiome* 5:116. doi: 10.1186/s40168-017-0336-9
- Klappenbach, J. A., Dunbar, J. M., and Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* 66:1328. doi: 10.1128/AEM.66.4.1328-1333.2000
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Loman, N. J., and Quinlan, A. R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* 30, 3399–3401. doi: 10.1093/bioinformatics/btu555
- Lu, H., Tang, Y., Lin, L., and Wolfgang, D. R. (2016). Next-generation sequencing confirmation of real-time RT-PCR false positive influenza-A virus detection in waterfowl and swine swab samples. *J. Next Gen. Seq. Appl.* 3. doi: 10.4172/2469-9853.1000134
- Madoui, M. A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., et al. (2015). Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics* 16:327. doi: 10.1186/s12864-015-1519-z
- Malhotra, S., Sharma, S., Bhatia, N., Kumar, P., and Hans, C. (2014). Molecular methods in microbiology and their clinical application. *J. Mol. Genet. Med.* 8:142. doi: 10.4172/1747-0862.1000142
- Mayo, B., Rachid, C., Alegria, A., Leite, A., Peixoto, R. S., and Delgado, S. (2014). Impact of next generation sequencing techniques in food microbiology. *Curr. Genomics* 15, 293–309. doi: 10.2174/1389202915666140616233211
- Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7:11257. doi: 10.1038/ncomms11257
- Mitsushashi, S., Kryukov, K., Nakagawa, S., Takeuchi, J. S., Shiraishi, Y., Asano, K., et al. (2017). A portable system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer. *Sci. Rep.* 7:5657. doi: 10.1038/s41598-017-05772-5
- Moran-Gilad, J. (2017). Whole genome sequencing (WGS) for food-borne pathogen surveillance and control – taking the pulse. *Eurosurveillance* 22:30547. doi: 10.2807/1560-7917.ES.2017.22.23.30547
- Nguyen, T. T., Van Giau, V., and Vo, T. K. (2016). Multiplex PCR for simultaneous identification of *E. coli* O157:H7, *Salmonella* spp. and *L. monocytogenes* in food. *3 Biotech* 6:205. doi: 10.1007/s13205-016-0523-6
- Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/c5ay02550h
- Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., et al. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* 16:114. doi: 10.1186/s13059-015-0677-2
- Rang, F. J., Kloosterman, W. P., and de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* 19, 90. doi: 10.1186/s13059-018-1462-9
- Ruiz-Rueda, O., Soler, M., Calvo, L., and García-Gil, J. L. (2011). Multiplex real-time PCR for the simultaneous detection of *Salmonella* spp. and *Listeria monocytogenes* in food samples. *Food Anal. Methods* 4, 131–138. doi: 10.1007/s12161-010-9163-3
- Salihah, N. T., Hossain, M. M., Lubis, H., and Ahmed, M. U. (2016). Trends and advances in food analysis by real-time polymerase chain reaction. *J. Food Sci. Technol.* 53, 2196–2209. doi: 10.1007/s13197-016-2205-0
- Schmieder, R., Lim, Y. W., and Edwards, R. (2011). Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* 28, 433–435. doi: 10.1093/bioinformatics/btr669
- Shen, C., Luo, Y., Nou, X., Bauchan, G., Zhou, B., Wang, Q., et al. (2012). Enhanced inactivation of *Salmonella* and *Pseudomonas* biofilms on stainless steel by use of T-128, a fresh-produce washing aid, in chlorinated wash solutions. *Appl. Environ. Microbiol.* 78, 6789–6798. doi: 10.1128/AEM.01094-12
- Sheridan, G. E., Masters, C. I., Shallcross, J. A., and MacKey, B. M. (1998). Detection of mRNA by reverse transcription-PCR as an indicator of viability in *Escherichia coli* cells. *Appl. Environ. Microbiol.* 64, 1313–1318. doi: 10.1128/aem.64.4.1313-1318.1998
- Skinner, B. M., Rogers, A. T., and Jacob, M. E. (2018). Susceptibility of *Escherichia coli* O157:H7 to disinfectants *in vitro* and in simulated footbaths amended with manure. *Foodborne Pathog. Dis.* 15, 718–725. doi: 10.1089/fpd.2018.2457
- Solieri, L., Dakal, T. C., and Giudici, P. (2013). Next-generation sequencing and its potential impact on food microbial genomics. *Ann. Microbiol.* 63, 21–37. doi: 10.1007/s13213-012-0478-8
- Taboada, E. N., Graham, M. R., Carriço, J. A., and Van Domselaar, G. (2017). Food safety in the age of next generation sequencing, bioinformatics, and open data access. *Front. Microbiol.* 8:909. doi: 10.3389/fmicb.2017.00909
- Takahashi, H., Ohkawachi, M., Horio, K., Kobori, T., Aki, T., Matsumura, Y., et al. (2018). RNase H-assisted RNA-primed rolling circle amplification for

- targeted RNA sequence detection. *Sci. Rep.* 8:7770. doi: 10.1038/s41598-018-26132-x
- Taylor, T. L., Volkening, J. D., DeJesus, E., Simmons, M., Dimitrov, K. M., Tillman, G. E., et al. (2019). Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology. *Sci. Rep.* 9:1635. doi: 10.1038/s41598-019-52424-x
- Toma, C., Lu, Y., Higa, N., Nakasone, N., Chinen, I., Baschkier, A., et al. (2003). Multiplex PCR assay for identification of human diarrheagenic *Escherichia coli*. *J. Clin. Microbiol.* 41, 2669–2671. doi: 10.1128/jcm.41.6.2669-2671.2003
- van der Velden, A. W., Lindgren, S. W., Worley, M. J., and Heffron, F. (2000). *Salmonella* pathogenicity island 1-independent induction of apoptosis in infected macrophages by *Salmonella enterica* serotype typhimurium. *Infect. Immun.* 68, 5702–5709. doi: 10.1128/iai.68.10.5702-5709.2000
- Vázquez-Boland, J. A., Kuhn, M., Berche, P., Chakraborty, T., Dominguez-Bernal, G., Goebel, W., et al. (2001). *Listeria* pathogenesis and molecular virulence determinants. *Clin. Microbiol. Rev.* 14, 584–640. doi: 10.1128/cmr.14.3.584-640.2001
- Walsh, A. M., Crispie, F., Claesson, M. J., and Cotter, P. D. (2017). Translating omics to food microbiology. *Annu. Rev. Food Sci. Technol.* 8, 113–134. doi: 10.1146/annurev-food-030216-025729
- Wang, D., Hinkley, T., Chen, J., Talbert, J. N., and Nugen, S. R. (2019). Phage based electrochemical detection of *Escherichia coli* in drinking water using affinity reporter probes. *Analyst* 144, 1345–1352. doi: 10.1039/C8AN01850B
- Wood, D. E., and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15:R46. doi: 10.1186/gb-2014-15-3-r46
- Xiao, L., Zhang, L., and Wang, H. H. (2012). Critical issues in detecting viable *Listeria monocytogenes* cells by real-time reverse transcriptase PCR. *J. Food Protec.* 75, 512–517. doi: 10.4315/0362-028x.jfp-11-346
- Yang, M., Cousineau, A., Liu, X., Sun, D., Li, S., Gu, T., et al. (2019). Direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing on Nanopore MinION – promising strategies for multiplex identification of viable pathogens in food. *Biorxiv* [Preprint]. doi: 10.1101/700674
- Zanoli, L. M., and Spoto, G. (2013). Isothermal amplification methods for the detection of nucleic acids in microfluidic devices. *Biosensors* 3, 18–43. doi: 10.3390/bios3010018
- Zhang, W.-L., Bielaszewska, M., Liesegang, A., Tschäpe, H., Schmidt, H., Bitzan, M., et al. (2000). Molecular characteristics and epidemiological significance of shiga toxin-producing *Escherichia coli* O26 strains. *J. Clin. Microbiol.* 38, 2134–2140.

**Conflict of Interest:** AC, DS, SL, LS, and MX were employed by the New England Biolabs, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yang, Cousineau, Liu, Luo, Sun, Li, Gu, Sun, Dillow, Lepine, Xu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.