Microbiology Discovery
ISSN 2052-6180

HOAJ
Herbert Open Access Journals

Special Section | Microbial Genomics | Methodology Open Access

Determining RNA quality for NextGen sequencing: some exceptions to the gold standard rule of 23S to 16S rRNA ratio§

Arvind A Bhagwat[1*], Z. Irene Ying[1], Jeff Karns[1] and Allen Smith[2]
*Correspondence: Arvind.bhagwat@ars.usda.gov
§Mention of brand or firm name does not constitute an endorsement by the U.S. Department of 13 Agriculture above others of a similar nature not mentioned.
[1]Environmental Microbial & Food Safety Laboratory and [2]Diet Genomics and Immunology Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, USDA, 10300 Baltimore Avenue, BARC-E, Beltsville, MD 20705-2350, USA.

Abstract

Background: Using next-generation-sequencing technology to assess entire transcriptomes requires high quality starting RNA. Traditionally, the ratios of 23S to 16S ribosomal RNA bands from agarose gel electrophoresis have been used to judge integrity of RNA. Currently, RNA quality is routinely judged using automated microfluidic gel electrophoresis platforms and associated algorithms.

Findings: Here we report that the two most popular automated platforms (i.e., BioAnalyzer and Experion systems of Agilent Technology and Bio-Rad Laboratories respectively) and their associate algorithms are based on limited data sets of model organisms. The systems perform data interpretation with presumption that prokaryotic rRNA molecules are eluted in two unique peaks corresponding to 23S and 16S molecules. However, certain microorganisms carry intervening sequences in their rRNA structural genes that are subsequently excised during ribosome formation. In such instances, the 23S and 16S rRNA components are eluted in multiple peaks. As a result, current algorithms used by microfluidic platforms read such samples as 'degraded' and assign them poor RNA quality scores. We observed RNA isolated from several *Citrobacter* and *Salmonella* isolates generated false quality scores and low 23S to 16S ribosomal RNA ratios.

Conclusions: For RNA-sequencing projects involving non-model organisms, relying solely on automated algorithms for 'quality control' of RNA could be misleading. Multiple peaks corresponding to 23S or 16S RNA could be due to occurrence of multiple intervening sequences in rRNA genes.

Keywords: RNA-Seq, RNA quality, *salmonella* transcriptome, agarose gel electrophoresis

Background

The decreasing cost and increasing availability of next-generation sequencing (NGS) technology will lead the way to its wide ranging applications across the fields of microbiology [9]. One powerful application of NGS technology is sequencing an organism's entire mRNA population, allowing for high-resolution gene expression studies. Obtaining high quality RNA preparations is extremely important in such experiments. Researchers have relied on guidelines such as a high ratio of A_{260}/A_{280} (>2.0) to measure RNA purity, and a high ratio of 23S rRNA to 16S rRNA (~1.5) to measure RNA integrity [14]. More recently, gel electrophoresis profiles from automated microfluidic platforms are used to determine RNA integrity [15,17]. However, interpretation of RNA quality by automated methods is limited by the associated algorithms, which are constructed from a limited number of biological samples and biased for eukaryotic RNA analyses. A number of prokaryotes do not possess intact 23S rRNA molecules [7], which is not accounted for by platforms using computer algorithms. Here we report several *Salmonella sp.* isolates that do not carry intact 23S rRNA, and as a result, their RNA preparations may be erroneously judged as degraded or having low RNA integrity.

Methods

Bacterial strains and growth conditions *Salmonella enterica* serovar strains paratyphi B, paratyphi A, Montevideo, and Panama were from Salmonella Genetic Stock Center (SARB Collection, University of Calgary, Canada). *Salmonella enterica* Serovar Typhimurium strain SL1344, *Escherichia coli* O157:H7 EDL933, *Escherichia albertii* USDA 181, and *Citrobacter rodentium* ATCC 51459 have been described earlier [2,3,8,11,16]. Bacterial cultures were streaked on Luria-Bertani (LB) agar plates from freezer stocks, and a single colony was inoculated in LB broth and grown at 37°C in a shaker incubator for 18-20 h.

RNA isolation procedure

RNA was isolated from stationary phase cells grown in LB broth for 18-20 h. Cells were stabilized with RNA stabilization reagent (RSR-3) [5] and RNA was isolated using Qiagen RNeasy kit as described before [1,5]. RNA was also isolated by suspending cells (after RNA stabilization) in hot Trizol reagent (65 °C) and purified using RNeasy silica columns and on-column DNase treatment (Qiagen). After elution from Qiagen column, RNA was treated with DNase (New England Biolabs, USA) and repurified with a RNeasy column. DNA contamination was checked by

doi: 10.7243/2052-6180-1-10

PCR using primers against *invA* gene of *Salmonella sp.* [10].

Microfluidic gel electrophoresis of RNA

RNA integrity was analyzed using chip-based microcapillary electrophoresis systems (Agilent 2100 BioAnalyzer, Agilent Technology, USA) or Experion Bio-Rad Laboratories, USA), RNA molecular weight ladder provided in kits was used as reference. Electrophoretic runs were performed following manufacturer's instructions and the molecular weight and integrity of bacterial RNA samples was determined using BioAnalyzer and Experion software programs.

Determination of Intervening sequences of Salmonella sp. 23S gene

PCR probes were designed for helix-25 rrlH-F1 (5'GCGCCGGTAAGGTGATATG 3') and rrlH-R1 (5' GCTATCTCCCGGTTTGATTG 3') and helix-45 rrlH-F2 (5'CCGATGCAAACTGCGAATAC 3') and rrlH-R2 (5' TTCTCTACCTGACCACCTG 3') intervening sequences [12]. These primers are located in the *rrlH* gene (NC_016810.1, at positions 291299 and 292227) of *S. enterica* serovar Typhimurium strain SL 1344. PCR was carried out using template DNA isolated from a single colony, and the thermal profile consisted of 30 cycles of 1 min of denaturation (94°C), 1 min of annealing (56°C), and 1 min of extension (72°C). A final extension step for 3 min at 72°C was performed. PCR products were electrophoresed on 1.5% agarose gel in 0.5xTris-borate EDTA buffer.

Results and discussion

Formaldehyde-containing agarose gel electrophoresis patterns have been historically used to determine RNA integrity. Samples having a 23S ribosomal band at twice the intensity of 16S ribosomal band (28 S and 18 S for eukaryotic RNA) were considered to have intact RNA molecules with high integrity [14]. However, with the advent of capillary electrophoresis and microfluidic chip based procedures, RNA samples can be separated into different integrity categories [15,17]. Currently two major microcaplillary electrophoretic RNA separation systems (BioAnalyzer, Agilent Technology, USA and Experion, Bio-Rad Laboratories, USA) are in wide use. Both methods automatically compare features from a single electrophoretic run and the data are analyzed with a proprietary regression models in order to estimate RNA quantity and integrity. The final output of the quality analysis is expressed on a 1 to 10 arbitrary scale as either RNA Integration Number or RIN or RNA Quality Indicator or RQI (BioAnalyzer and Experion platforms respectively).

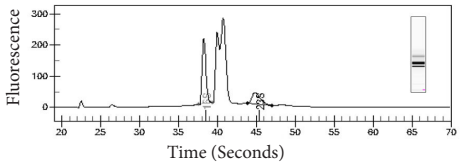
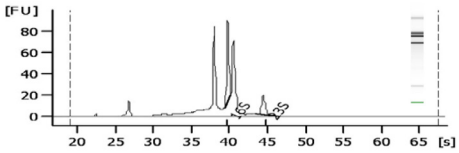
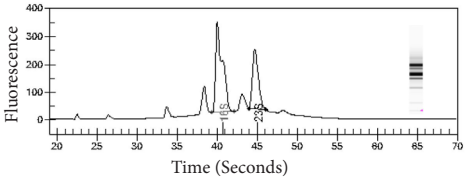
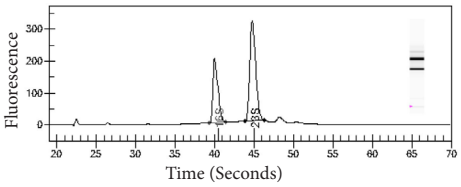
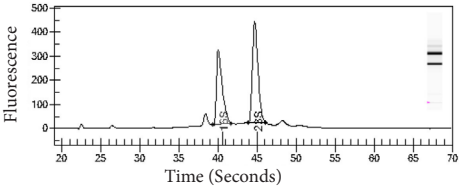
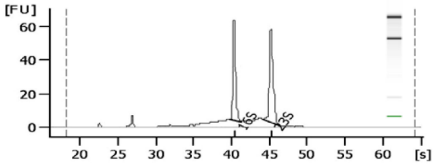
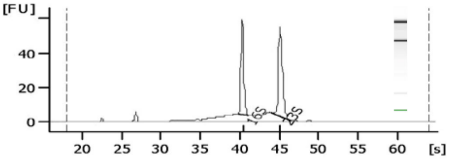
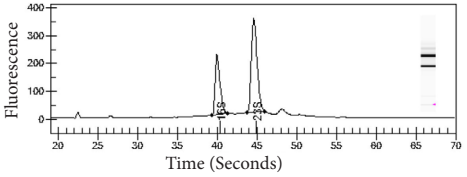
When we performed RNA extraction on different *Salmonella* isolates we observed that previously tried and trusted RNA isolation protocols established in this laboratory [1,4,5] generated either poor 23S/16S ratios (**Table 1**, sample 1) and/or poor 'Quality Scores' (**Table 1**, samples 2 and 3) on both microfluidic platforms. Additionally, the quality score generated on BioAnalyzer platform was not consistent with the conventional indicator such as 23S/16S ratio (i.e., **Table 1**, sample 1 vs. sample 4) (Experion platform does not generate quality score for prokaryotic RNA samples). Moreover, identification of 16S peak by Experion platform was inconsistent and varied between samples (eg., 38.5 s vs. 40.6 s for samples 1 and 3 respectively; **Figure 1**). To minimize any possibility of degradation, RNA was isolated from cells suspended directly in hot Trizol reagent. Both RNA isolation protocols gave RNA with poor 23S/16S ratios ranging from 0.3 to 0.5 (samples 1-3, **Table 1**). However, using the same RNA isolation protocol, RNA from *E. coli* K-12 and O157:H7strain EDL933 grown under identical conditions gave quality scores exceeding 8.0 with high ratio of 23S/16S rRNA (sample 4, data for *E. coli* K-12 not shown).

A literature search of prokaryotic ribosomal RNA patterns revealed the existence of fragmented 23S RNA in several genera of the family *Enterobacteriaceae* [13]. Fragmentation occurs immediately after transcription as a result of RNase III dependent excision of intervening sequences (IVSs) at two sites within the 23S molecule: helix-25 and helix-45. The fragmented 23S molecule is not religated thus interrupting its linear continuity [6]. The IVSs are approximately 90 bp intervening sequences and have not been reported in the model organism *E. coli*, providing a possible explanation for the apparent incongruity in microfluidic data for RNA integrity scores and 23S/16S ratios.

In some *Salmonella* strains, IVSs occur at helix-25 and helix-45 positions and may vary in numbers [12]. *Salmonella* and *E. coli* carry total seven ribosomal RNA synthesis genes (*rrl*) per chromosome and in each *Salmonella enterica* strain the occurrence of IVSs across all *rrl* genes within a chromosome is highly variable. In order to ascertain if IVSs are indeed the cause of poor RNA integrity scores and low 23S/16S ratios we designed primers specific to helix-25 and helix-45 of *Salmonella enterica* serovar Typhimurium strain SL1344 and *E. coli*. Using these PCR primers 731 and 733 bp amplicons for helix-25 and helix-45 positions, respectively, were generated in *E. coli* where IVSs are absent (data not shown). On the other hand, a higher molecular weight amplicon (~830 bp) was generated from those *rrl* loci which harbored IVSs (**Figure 1**). Our PCR data corroborates earlier reported IVSs in helix-25 and helix-45 positions [12]. As expected, RNA from *Salmonella* strains which do not carry IVSs in either of the helixes such as serovar Montevideo strain SARB-30, serovar Panama strain SARB-40, and serovar Paratyphi A strain SARB-42 (**Figure 1**, lanes 9-14) [12], gave high RNA integrity scores and high 23S/16S ratios on both microfluidic platforms (**Table 1**, samples 6-8). Conversely, strains that were shown by PCR to contain IVSs (**Figure 1**, lanes 1-8, 15 and 16) yielded RNA with lower RNA integrity scores (**Table 1**, samples 1-3, and 5). Also it may be noted that both platforms generated identical electropherogram patterns (**Table 1**, samples 6 and 7 vs. 8 where all three species lack IVSs) although quality scores for prokaryotic RNA were generated by only BioAnalyzer platform.

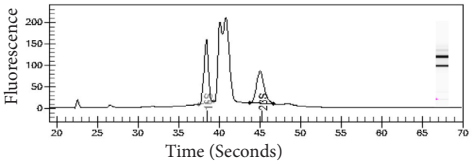
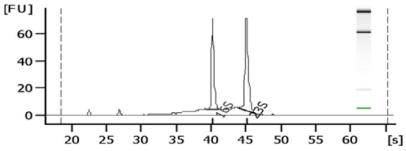
 doi: 10.7243/2052-6180-1-10

Table 1. Electrophorograms of RNA isolated from different *Salmonella* and *E. coli* strains on microfluidic platform[a].

	Strain	Intervening Sequences (IVSs)[b]		23S/16S ratio[c]	RIN[d]	Microfluidic electrophorograms (Inset: Instrument software generated 'virtual gel' pattern)
		Helix-25	Helix-45			
1	***Salmonella enterica* serovar Paratyphi B (SARB 43)**	-	+	0.31	8.1	
2	***Salmonella enterica* serovar Typhimurium SL1344 NalR**	-	+	0.4	6.0	
3	***Salmonella enterica* serovar Paratyphi B (SARB-45)**	-	+	0.52	7.7	
4	***Escherichia coli* O157:H7 EDL933**	-	-	1.5	8.5	
5	***Salmonella enterica* serovar Paratyphi B (SARB-46)**	-	+	1.19	9.5	
6	***Salmonella enterica* serovar Montevideo (SARB-30)**	-	-	1.3	8.5	
7	***Salmonella enterica* serovar Panama (SARB 40)**	-	-	1.3	8.3	
8	***Salmonella enterica* serovar Paratyphi A (SARB-42)**	-	-	1.69	9.8	

doi: 10.7243/2052-6180-1-10

Continuation of Table 1.

	Strain	Intervening Sequences (IVSs)[b]		23S/16S ratio[c]	RIN[d]	Microfluidic electrophorograms (Inset: Instrument software generated 'virtual gel' pattern)
9	***Citrobacter rodentium*** **ATCC 51459**	+	+	0.7	6.1	
10	***Escherichia albertii*** **USDA-181**	-	-	1.3	8.8	

a: All samples were analyzed on both platforms and generated identical electrophoretic pattern. Electrophorogram of samples 2, 6, 7, and 1 10 are from BioAnalyzer platform, and the rest of the electrophorograms are from Experion platform.
b: Data for the occurrence of Intervening Sequences (IVSs) is based on PCR analysis (**Figure 1**).
c: 23S/16S ratios are reported from respective microfluidic electrophorograms.
d: Quality Score for RNA as computed by BioAnalyzer platform (RNA Integration Number or RIN) (Experion platform generates quality scores only for eukaryotic RNA samples).

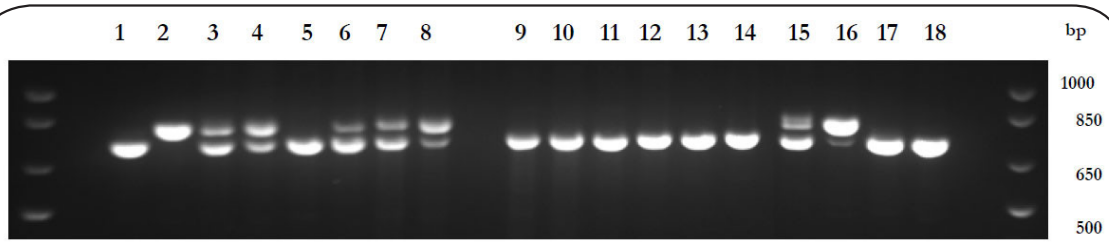


Figure 1. Detection of Intervening sequences (IVSs) in rrn loci by PCR.
PCR products which were obtained using various bacterial strains were electrophoresed using 1.5% agarose gel in 0.5x Tris-borate EDTA buffer. The *rrl* gene may include IVSs at helix 25 (odd numbered lanes) and/or at helix-45 (even numbered lanes). Lower band in each lane (~730 bp) represents PCR product from *rrl* gene which has no IVSs, while slower migrating (~833 bp) band indicates presence of IVSs. Two flanking lanes contain DNA molecular weight ladder with indicated bp sizes). DNA templates isolated from various strains, SARB-43 (lanes 1, 2), SARB-45 (lanes 3, 4), SARB-46 (lanes 5, 6), SARB-47 (lanes 7, 8), SARB-30 (lanes 9, 10), SARB-40 (lanes 11, 12), SARB-42 (lanes 13, 14), serovar Typhimurium SL1344 (lanes 15, 16), and E. *albertii* USDA 181 (lanes 17, 18).

We also examined RNA patterns in *C. rodentium,* which is a mouse pathogen and excellent model to study entero-hemorrhagic infections in animals [16], and the newly emerging food borne pathogen *E. albertii* [11]. Based on RNA electrophoretic patterns, IVSs appear to be present in *C. rodentium* but absent in *E. albertii* (**Figure 1**, lanes 17 and 18; data not shown for *C. rodentium*).

Both microfluidic platforms generated identical electrophoretic patterns for all RNA samples. However, we observed a slight discrepancy in the quantification algorithms of bacterial RNA among the two microfluidic platforms. In comparison with BioAnalyzer data, the Experion system consistently produced higher RNA estimations by a factor of 1.54 $\pm$ 0.6 ($P < 0.001$) and also appeared to tilt the ratio of 23S to 16S in favor of 23S by a factor of 1.11 $\pm$ 0.2 ($P = 0.012$).

Conclusions

While automated microfluidic platforms have many advantages such as user-independent, automated, and reproducible data generation for RNA integrity and quantification, they have certain limitations. For RNASeq experiments, most NextGen sequence providers require microfluidic electropherogram profiles to accompany samples to ensure RNA integrity. Thus, care must be taken when evaluating microbial RNA integrity and quantity data from microfluidic platforms to take into account non-traditional banding patterns that may arise

from IVSs. Specifically, in microbial RNASeq experiments of non-model organisms, relying completely on microfluidic platforms and on quality scores generated by the associated algorithms could possibly mislead researchers as otherwise perfect RNA preparations would be judged as 'degraded' or as having poor 23S to 16S ratios.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Authors' contributions	AAB	ZIY	JK	AS
Research concept and design	✓	--	--	--
Collection and/or assembly of data	✓	✓	--	✓
Data analysis and interpretation	✓	✓	✓	✓
Writing the article	✓	--	--	✓
Critical revision of the article	✓	✓	✓	✓
Final approval of article	✓	--	✓	✓
Statistical analysis	✓	--	--	✓

Acknowledgement

AAB, ZIY, JK and AS would like to thank Lumei Cheung and Sukla Lakshman for their excellent technical assistance.

Publication history

Editor: Janet Donaldson, Mississippi State University, USA.
EIC: Dr. Todd R. Callaway, U.S. Department of Agriculture, USA.
Received: 29-Jul-2013 Accepted: 24-Sep- 2013
Published: 01-Oct-2013

Reference

1. Bhagwat AA and Bhagwat M. **Comparative analysis of transcriptional regulatory elements of glutamate-dependent acid-resistance systems of Shigella flexneri and Escherichia coli O157:H7**. *FEMS Microbiol Lett.* 2004; **234**:139-47. | Article | PubMed
2. Bhagwat AA, Chan L, Han R, Tan J, Kothary M, Jean-Gilles J and Tall BD. **Characterization of enterohemorrhagic Escherichia coli strains based on acid resistance phenotypes**. *Infect Immun.* 2005; **73**:4993-5003. | Article | PubMed Abstract | PubMed Full Text
3. Bhagwat AA, Jun W, Liu L, Kannan P, Dharne M, Pheh B, Tall BD, Kothary MH, Gross KC, Angle S, Meng J and Smith A. **Osmoregulated periplasmic glucans of Salmonella enterica serovar Typhimurium are required for optimal virulence in mice**. *Microbiology.* 2009; **155**:229-37. | Article | PubMed
4. Bhagwat AA and Keister DL. **Identification and cloning of Bradyrhizobium japonicum genes expressed strain selectively in soil and rhizosphere**. *Appl Environ Microbiol.* 1992; **58**:1490-5. | Article | PubMed Abstract | PubMed Full Text
5. Bhagwat AA, Phadke RP, Wheeler D, Kalantre S, Gudipati M and Bhagwat M. **Computational methods and evaluation of RNA stabilization reagents for genome-wide expression studies**. *J Microbiol Methods.* 2003; **55**:399-409. | Article | PubMed
6. Burgin AB, Parodos K, Lane DJ and Pace NR. **The excision of intervening sequences from Salmonella 23S ribosomal RNA**. *Cell.* 1990; **60**:405-14. | Article | PubMed
7. Evguenieva-Hackenberg E. **Bacterial ribosomal RNA in pieces**. *Mol Microbiol.* 2005; **57**:318-25. | Article | PubMed
8. Gawande PV and Bhagwat AA. **Inoculation onto solid surfaces protects Salmonella spp. during acid challenge: a model study using polyethersulfone membranes**. *Appl Environ Microbiol.* 2002; **68**:86-92. | Article | PubMed Abstract | PubMed Full Text
9. Guell M, Yus E, Lluch-Senar M and Serrano L. **Bacterial transcriptomics: what is beyond the RNA horiz-ome?** *Nat Rev Microbiol.* 2011; **9**:658-69. | Article | PubMed
10. Liming SH and Bhagwat AA. **Application of a molecular beacon-real-time PCR technology to detect Salmonella species contaminating fruits and vegetables**. *Int J Food Microbiol.* 2004; **95**:177-87. | Article | PubMed
11. Ling J, Sharma M and Bhagwat AA. **Role of RNA polymerase sigma-factor (RpoS) in induction of glutamate-dependent acid-resistance of Escherichia albertii under anaerobic conditions**. *FEMS Microbiol Lett.* 2008; **283**:75-82. | Article | PubMed
12. Pabbaraju K, Miller WL and Sanderson KE. **Distribution of intervening sequences in the genes for 23S rRNA and rRNA fragmentation among strains of the Salmonella reference collection B (SARB) and SARC sets**. *J Bacteriol.* 2000; **182**:1923-9. | Article | PubMed Abstract | PubMed Full Text
13. Pronk LM and Sanderson KE. **Intervening sequences in rrl genes and fragmentation of 23S rRNA in genera of the family Enterobacteriaceae**. *J Bacteriol.* 2001; **183**:5782-7. | Article | PubMed Abstract | PubMed Full Text
14. Sambrook J, Fritsch EF and Maniatis T. **Molecular Cloning: A Laboratory Manual**. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
15. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M and Ragg T. **The RIN: an RNA integrity number for assigning integrity values to RNA measurements**. *BMC Mol Biol.* 2006; **7**:3. | Article | PubMed Abstract | PubMed Full Text
16. Smith A and Bhagwat AA. **Hypervirulent-host-associated Citrobacter rodentium cells have poor acid tolerance**. *Curr Microbiol.* 2013; **66**:522-6. | Article | PubMed
17. Wilkes TM, Devonshire AS, Ellison SL and Foy CA. **Evaluation of a novel approach for the measurement of RNA quality**. *BMC Res Notes.* 2010; **3**:89. | Article | PubMed Abstract | PubMed Full Text

Citation:
Bhagwat AA, Ying ZI, Karns J and Smith A. **Determining RNA quality for NextGen sequencing: some exceptions to the gold standard rule of 23S to 16S rRNA ratio**[§]. *Microbiol Discov*. 2013; **1**:10.
http://dx.doi.org/10.7243/2052-6180-1-10