A Modified Low-quantity RNA-Seq Method for Microbial Community and Diversity Analysis Using Small Subunit Ribosomal RNA

Yong-Wei Yan, Ting Zhu, Bin Zou and Zhe-Xue Quan*

Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Biodiversity Science, School of Life Sciences, Fudan University, Shanghai, China
*For correspondence: quanzx@fudan.edu.cn

[Abstract] We propose a modified RNA-Seq method for small subunit ribosomal RNA (SSU rRNA)-based microbial community analysis that depends on the direct ligation of a 5' adaptor to RNA before reverse-transcription. The method requires only a low-input quantity of RNA (10-100 ng) and does not require a DNA removal step. Using this method, we could obtain more 16S rRNA sequences of the same regions (variable regions V1-V2) without the interference of DNA in order to analyze OTU (operational taxonomic unit)-based microbial communities and diversity. The generated SSU rRNA sequences are also suitable for the coverage evaluation for bacterial universal primer 8F (*Escherichia coli* position 8 to 27), which is commonly used for bacterial 16S rRNA gene amplification. The modified RNA-Seq method will be useful to determine potentially active microbial community structures and diversity for various environmental samples, and will also be useful for identifying novel microbial taxa.
Keywords: RNA-Seq, Low quantity, SSU rRNA, OTU, Microbial community

[Background] Ribosomal RNA (rRNA) accounts for more than 90% of the total microbial RNA, and is suitable for the analysis of microbial communities as an indicator of microbial physiological activity to synthesize proteins (Blazewicz *et al.*, 2013). The study of microbial community transcripts, including rRNA and mRNA, in a particular environment (Double RNA metatranscriptomics) has advantages in providing both functional and taxonomic information on microbes (Urich *et al.*, 2008), but has failed to perform OTU-based community comparisons. Although diversity indices could be calculated and compared using the V3 region of 16S rRNA sequences when gel-extracted SSU rRNA is analyzed, only a third of the resulting 16S rRNA sequences were found to be suitable for such analyses (Li *et al.*, 2014). Besides, such method usually requires high quantities of RNA (Li *et al.*, 2014). We recently developed a modified RNA-Seq method that uses an immediate adaptor ligation step at the 5' end of the RNA prior to reverse transcription. Consequently, we can obtain more 16S RNA reads which can be used for OTU-based community and diversity analysis especially regarding low-RNA-yield samples such as tap water, shower curtain and human skin (Yan *et al.*, 2017), and also mudflat sediment samples (Yan *et al.*, 2018).

Materials and Reagents

A. RNA extraction

1. Glass beads, acid-washed (≤ 106 μm) (Sigma-Aldrich, catalog number: G4649)

2. 2 ml bead beating tubes (QIAGEN, catalog number: 13118-400)

3. mirVANA™ miRNA Isolation Kit (Thermo Fisher Scientific, Invitrogen™, catalog number: AM1560)

B. RNA-Seq library preparation

1. PCR tubes (Corning, Axygen®, catalog number: PCR-02-L-C)

2. RNA-Seq Library Preparation Kit for Whole Transcriptome Discovery–Illumina Compatible (Gnomegen, catalog number: K02421-T)

3. DNase/RNase free water (Thermo Fisher Scientific, Invitrogen™, catalog number: AM9932)

4. Qubit™ RNA HS Assay Kit (Thermo Fisher Scientific, Invitrogen™, catalog number: Q32855)

5. RNaseOUT™ Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific, Invitrogen™, catalog number: 10777019)

6. Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, Invitrogen™, catalog number: Q32854)

7. Gnome Size Selector (Gnomegen, catalog number: R02424S)

C. Other materials

1. Pipette tips (10 μl, 200 μl, 1,000 μl) (Corning, Axygen®, catalog numbers: TF-300-R-S, TF-200-R-S, TF-1000-R-S)

2. 1.5 ml Eppendorf tubes (Eppendorf, catalog number: 022363204)

3. 100% ethanol

4. RNase-free 70% ethanol

5. Agarose (Biowest, catalog number: 111860)

6. 50x TAE Buffer (Sangon Biotech, catalog number: B548101-0500)

7. DL2,000 DNA Marker (Takara Bio, catalog number: 3427A)

Equipment

1. Pipettes (0-10 μl, 10-100 μl, 100-1,000 μl) (Eppendorf, catalog numbers: k03030, k03031, k03032)

2. Vortex-genie™ 2 (QIAGEN, model: Vortex-Genie® 2)

3. Fresco 17 centrifuge (Thermo Fisher Scientific, Thermo Scientific™, model: Heraeus™ Fresco™ 17)

4. Qubit 2.0 Fluorometer (Thermo Fisher Scientific, model: Qubit 2.0)

5. Automated thermal cycler TP600 (Takara Bio, model: TP600)

6. MagneSphere® Technology Magnetic Separation Stand (twelve-position) (Promega, catalog number: Z5342)

7. Electrophoresis apparatus (Bio-Rad Laboratories)

8. Gel imaging system

Software

1. Sickle v1.33 (Joshi and Fass, 2011)

2. Mothur v1.33.3 (Schloss *et al.*, 2009)

3. QIIME v1.8.0 (Caporaso *et al.*, 2010)

4. MIPE (Zou *et al.*, 2017)

Procedure

Note: Details of sample collection, storage and transport for low-biomass samples (tap water, shower curtain, mudflat water, leaf surface and human skin) please see the method described by Yan et al. (2017). Regarding mudflat sediment samples, details of sample collection and nucleic acid extraction are described in our recent work (Yan et al., 2018).

A. Nucleic acid extraction

1. Extract total RNA or total nucleic acids from low-biomass samples, which are unsuitable for DNase I digestion, using a mirVANA™ miRNA Isolation Kit. Five or more swabs are preferred to be pooled for nucleic acid extraction, while a whole piece of filter membrane is preferred to be cut into small pieces (~0.5 mm²) for extraction. To completely disrupt microbial cell walls, vortex such preprocessed samples with 0.7 g glass beads and 0.7 ml Lysis/Binding Buffer in a 2 ml bead beating tube at maximum speed prior to extraction. After centrifugation at 13,000 *x g* for 5 min at 4 °C, transfer the supernatant to a new 1.5 ml tube and extract total RNA strictly following the manufacturer's instructions.

2. Visualize the RNA samples in a 1% (w/v) agarose gel after electrophoresis to assess the RNA integrity (Figure 1). Store total RNA or total nucleic acids at -80 °C. Quantify total RNA using a Qubit™ RNA HS Assay Kit on Qubit 2.0 fluorometer before the preparation of RNA-Seq libraries.
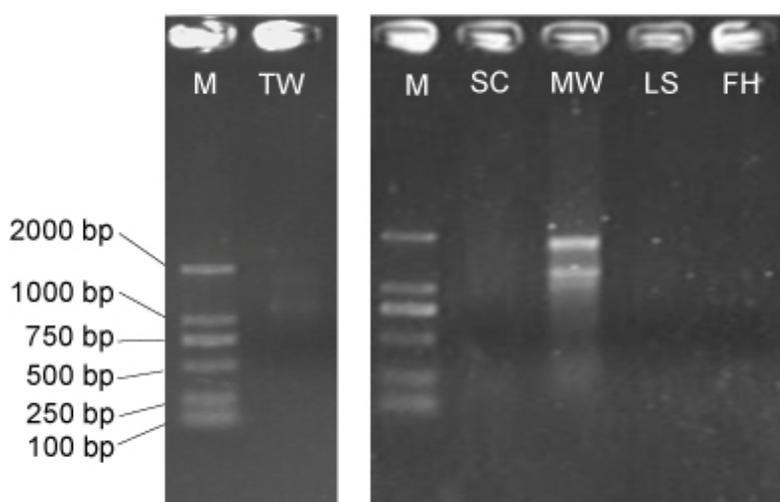
Figure 1. Total RNA electrophoresis for low-biomass samples. M: DL2,000 DNA marker; TW: tap water; SC; shower curtain; MW: mudflat surface water; LS: leaf surface; FH: forehead of human skin. Only TW and MW display integral 23S and 16S rRNA bands when 10 μl RNA is loaded, in contrast to samples that yield a very low concentration of RNA (< 2 ng/μl).

B. Preparation of RNA-Seq libraries

   This RNA-Seq library preparation protocol is modified from the Gnomegen RNA-Seq Library Preparation Kit protocol (catalog number: K02421-T). Prepare RNA-Seq libraries following the steps described below.

   1. Total RNA denaturation

      a. Add total RNA (10-100 ng) and DNase/RNase free water into a 0.2 ml PCR tube to make a 20-μl reaction volume.

      b. Mix well and incubate at 65 °C for 5 min using a thermal cycler program. Place the tube on ice immediately after the incubation and leave for at least 5 min.

         Note: We use this step to substitute the original fragmentation step as to keep the integrity of the RNA and directly ligate the RNA-Seq adaptor to the 5' end of the rRNA in the following step.

   2. Adaptor ligation to 5' end of RNA

      a. Add the following reagents sequentially to the denatured RNA, and make a reaction volume of 40.6 μl:

10x ligation buffer	3.6 μl (Final ~0.9x)
10 mM ATP	4 μl (Final ~1 mM)
Ligation enzyme mix	1 μl
Ligation enzyme supplement	1 μl
RNaseOUT (40 U/μl)	1 μl (Final ~1 U/μl)
Ligation enhancer mix	8 μl
RNA Seq 5' Adaptor B	2 μl

b. Mix well and incubate at 37 °C for 2 h using a thermal cycler program. Place the tube on ice after incubation.

Note: RNaseOUT is a separately-purchased reagent. Such specific ligation between single-stranded Adaptor B and RNA avoids the interference of DNA for the community analysis.

3. Purification of ligation product

Transfer 40 µl ligation product to a new 1.5 ml tube, and purify the product using a Gnome Size Selector, strictly following the manufacturer's instructions. A total volume of 10-µl purified product is obtained for the next step.

Note: Warm Gnome Size Selector to room temperature, and completely resuspend the Size Selector before use. Accurately use the volume that is recommended.

4. Synthesize the first strand of cDNA with a tagged random hexamer.

   a. Denaturation of ligation product

      i. Add 1 µl tagged RT primer and 1 µl dNTP (10 mM) to the purified ligation product in a 12-µl denaturation volume.

      ii. Mix well and incubate at 65 °C for 5 min using a thermal cycler program. Place the tube on ice immediately after the incubation and leave for at least 5 min.

   b. Reverse transcription

      i. Add the following reagents sequentially into the denatured ligation product, and make a final 20-µl reaction volume:

5x First Strand Buffer	4 µl
100 mM DTT	2 µl
RNaseOUT (40 U/µl)	1 µl
Reverse Transcriptase	1 µl

      ii. Mix well and use the following thermal cycler program to synthesize the cDNA:

         25 °C for 12 min;
         42 °C for 40 min;
         70 °C for 15 min;
         Hold at 4 °C.

5. cDNA purification

Purification is performed using a Gnome Size Selector strictly according to the manufacturer's instructions. 20 µl purified cDNA is obtained for the next step.

Note: Warm Gnome Size Selector to room temperature, and completely resuspend the Size Selector before use. Accurately use the volume that is recommended.

6. PCR amplification

The complete reagent mixture contains:

Purified single-stranded cDNA	20 µl
2x HiFi PCR Master Mix	25 µl
PCR Primer F	1.25 µl
Barcoded reverse primer bcX	1.25 µl

DNAse/RNase free water   2.5 µl

Note: Forward and reverse primers are designed based on the RNA Seq 5' Adaptor B and tag sequences, and are complementary to the standard Illumina forward and reverse primers. Therefore, the whole process avoids the utilization of microbe-related universal primers. The reverse primer also contains an 8-nucleotide (nt) indexing sequence to allow for multiplexing.

The PCR amplification conditions are as follows:

98 °C for 45 sec;

15 cycles of: 98 °C for 15 sec, 60 °C for 30 sec, 72 °C for 30 sec;

72 °C for 1 min;

Hold at 4 °C.

Note: Usually, 1-2 µl PCR products is enough for visualization on a 1% agarose gel. As a result, a wide range of clear and smear bands (100-2,000 bp) will be observed.

7. Purification of PCR product

   Size select the 400-600 base pair (bp) PCR product using a Gnome Size Selector strictly according to the manufacturer's instructions.

   Note: Warm Gnome Size Selector to room temperature, and completely resuspend the Size Selector before use. Accurately use the volume that is recommended. Other size ranges of PCR product can also be obtained using Gnome Size Selector according to the manufacturer's instructions.

C. High-throughput sequencing

   Quantify barcoded PCR product from different samples using a Qubit™ dsDNA HS Assay Kit on Qubit 2.0 fluorometer. Mix together PCR product from different samples with equal quantity (at least 10 ng for each sample) and sequence on an Illumina MiSeq platform using the 2 x 300 paired end protocol. For each sample, paired-end reads formatted as two FASTQ files will be obtained.

Data analysis

1. Pre-process paired-end reads with Sickle software v1.33 using the command pe to trim and filter reads with a Phred quality score below 20.

2. *De novo* merge paired-end reads using the command join_paired_end.py with default parameters in QIIME v1.8.0.

3. Transform FASTQ files into fasta format using the command fastq.info in mothur v1.33.3.

4. Remove merged sequences containing ambiguous nucleotides and homopolymer lengths longer than eight nucleotides using the command screen.seqs in mothur v1.33.3.

5. Remove merged sequences shorter than 250 bp using the command screen.seqs in mothur v1.33.3.

6. Classify the reads against the SILVA SSU v119 database in MIPE (Zou *et al.*, 2017) with a bootstrap cut-off of 80% and remove SSU rRNA reads identified as chloroplast, mitochondria or

human. Mismatches of the primer 8F (5'-AGAGTTTGAT (C/T) (A/C) TGGCTCAG-3') (Mao *et al.*, 2012) in all the bacterial 16S rRNA sequences are also identified with this software (Table 1).

Table 1. Non-coverage rates of the bacterial universal primer 8F

	Non-coverage rates (%)[a]				
	TW[b]	**SC**	**MW**	**LS**	**FH**
Domain					
Bacteria	1.70	0.57	1.74	3.10	0.72
Phylum					
Acidobacteria	0.79				
Actinobacteria		0.34			0.36
Cyanobacteria			0.32		
Proteobacteria	0.50	0.33	0.46	0.72	0.52
Unclassified	2.15	25.69	10.90	38.69	8.70

[a]*The non-coverage rates were calculated by dividing non-coverage sequences (at least one mismatch within primer 8F) of taxa with their relative total sequences. The phyla with less than 10 non-coveraged sequences in the datasets are not shown.*

[b]*TW: tap water; SC; shower curtain; MW: mudflat surface water; LS: leaf surface; FH: forehead of human skin.*

7. For each sample, extract all the identified prokaryotic 16S rRNA reads using get.seqs command and deposit the reads in a fasta file. Merge all the fasta files of all the samples using cat command in Linux, and use count.seqs command in mothur v1.33.3 to differentiate sequences according to samples.

8. Align and trim sequences to leave the 8F-V1-V2 regions (*E. coli* position 8 to 242) using align.seqs and pcr.seqs commands in mothur v1.33.3 respectively.

9. Analyze OTU based community structures (Figure 2) and calculate diversity indices following the MiSeq SOP pipeline (https://www.mothur.org/wiki/MiSeq_SOP) as described in Kozich *et al.* (2013). Representative OTU sequences at a cut-off of 0.03 are taxonomically classified against the SILVA SSU v119 database with a bootstrap cut-off of 50%. OTUs belonging to chloroplasts or mitochondria are also removed from the analysis.
   Note: All scripts for computational analysis are available in our supplementary material "Supplementary.docx".

www.bio-protocol.org/e2828
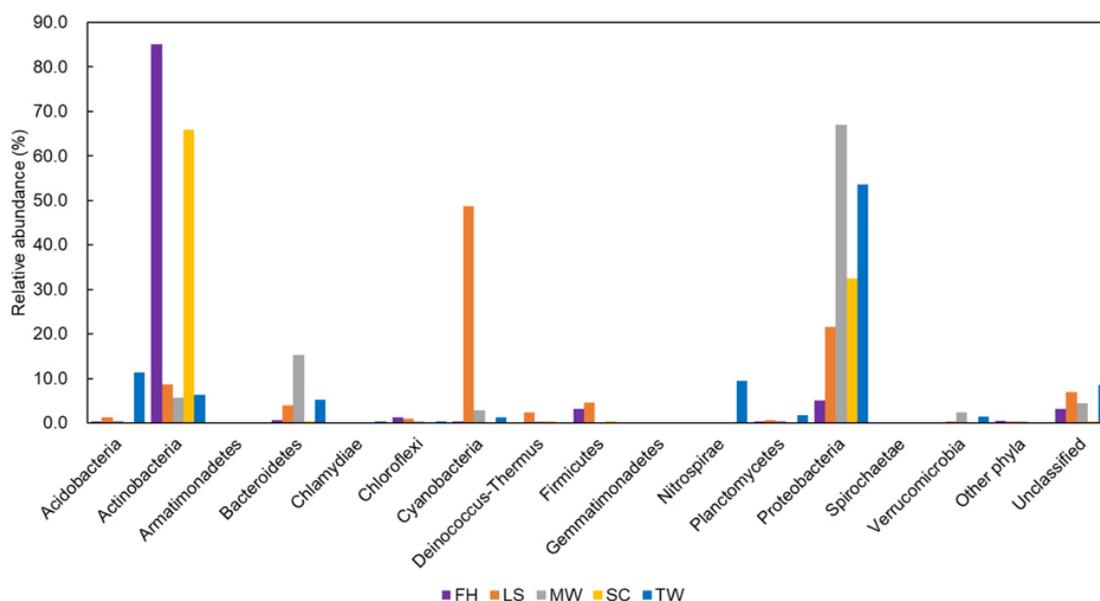
Figure 2. Comparisons of OTU-based bacterial communities in different samples at the phylum level. 'Other phyla' includes taxa that made up of small fractions (< 1%). 'Unclassified' includes sequences under a bootstrap cut-off value of 50% for bacteria. TW: tap water; SC; shower curtain; MW: mudflat surface water; LS: leaf surface; FH: forehead of human skin.

Notes

1. All the RNA samples and reaction reagents should be thawed and operated on ice. Once the reaction mix is prepared, perform the thermal cycler program as soon as possible.
2. For different samples, the optimum cycles used in the PCR amplification step can be determined using a 10-µl preliminary reaction system.

Acknowledgments

References

1. Blazewicz, S. J., Barnard, R. L., Daly, R. A. and Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J* 7(11): 2061-2068.
2. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D.,

Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5): 335-336.

3. Joshi, N. A., and Fass, J. N. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33).

4. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79(17): 5112-5120.

5. Li, X. R., Lv, Y., Meng, H., Gu, J. D. and Quan, Z. X. (2014). Analysis of microbial diversity by pyrosequencing the small-subunit ribosomal RNA without PCR amplification. *Appl Microbiol Biotechnol* 98(8): 3777-3789.

6. Mao, D. P., Zhou, Q., Chen, C. Y. and Quan, Z. X. (2012). Coverage evaluation of universal bacterial primers using the metagenomic datasets. *BMC Microbiol* 12: 66.

7. Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. and Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75(23): 7537-7541.

8. Urich, T., Lanzen, A., Qi, J., Huson, D. H., Schleper, C. and Schuster, S. C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* 3(6): e2527.

9. Yan, Y. W., Jiang, Q. Y., Wang, J. G., Zhu, T., Zou, B., Qiu, Q. F. and Quan, Z. X. (2018). Microbial communities and diversities in mudflat sediments analyzed using a modified metatranscriptomic method. *Front Microbiol* 9: 93.

10. Yan, Y. W., Zou, B., Zhu, T., Hozzein, W. N. and Quan, Z. X. (2017). Modified RNA-seq method for microbial community and diversity analysis using rRNA in different types of environmental samples. *PLoS One* 12(10): e0186161.

11. Zou, B., Li, J., Zhou, Q. and Quan, Z. X. (2017). MIPE: A metagenome-based community structure explorer and SSU primer evaluation tool. *PLoS One* 12(3): e0174609.