

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Mäki, Anita; Tirola, Marja

Title: Directional high-throughput sequencing of RNAs without gene-specific primers

Year: 2018

Version: Published version

Copyright: © Authors, 2018

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Mäki, A., & Tirola, M. (2018). Directional high-throughput sequencing of RNAs without gene-specific primers. *BioTechniques*, 65(4), 219-223. <https://doi.org/10.2144/btn-2018-0082>

Directional high-throughput sequencing of RNAs without gene-specific primers

Anita Mäki^{*,1} & Marja Tiirola¹

¹Department of Biological and Environmental Science, Nanoscience Center, University of Jyväskylä, PO Box 35, FI-40014 Finland

BioTechniques 65: 219–223 (October 2018) 10.2144/btn-2018-0082

Keywords: algae • ligation • next-generation sequencing • phytoplankton • primer bias • random priming • ribosomal RNA

Ribosomal RNA analysis is a useful tool for characterization of microbial communities. However, the lack of broad-range primers has hampered the simultaneous analysis of eukaryotic and prokaryotic members by amplicon sequencing. We present a complete workflow for directional, primer-independent sequencing of size-selected small subunit ribosomal RNA fragments. The library preparation protocol includes gel extraction of the target RNA, ligation of an RNA oligo to the 5'-end of the target, and cDNA synthesis with a tailed random-hexamer primer and further barcoding. The sequencing results of a phytoplankton mock community showed a highly similar profile to the biomass indicators. This method has universal potential for microbiome studies, and is compatible for the 5'-end sequencing of other RNA types with minimum library preparation costs.

The phylogenetic characterization of microbial communities rests mainly on the small subunit ribosomal RNA (SSU rRNA) genes, 16S rRNA in prokaryotes and its counterpart 18S rRNA in eukaryotes; however, the lack of good universal primer sequences and incomplete reference databases complicates the analysis of eukaryotic species. Our previous phytoplankton mock community study showed that RNA-based community profiling correlated better with the biomass measures than DNA-based sequencing, in which the variation in the rRNA gene copy numbers per algae cell can be over 100-fold [1]. Moreover, profiling microbial communities using 16S/18S rRNA instead of rRNA genes especially brings out living organisms with active protein synthesis and is free of dissolved or relic DNA [2].

To amplify RNA templates without the natural poly(A)-tail, the use of gene-specific primers can be avoided by 3'- and 5'-end poly(A) tailing of RNA or by ligation of adaptor sequences as reviewed by van Dijk *et al.* [3]. When primer-independent methods have been applied, our knowledge on the rRNA

molecules and the diversity of life have been expanded [4–6]. Commercial kits have been fabricated for nontargeted amplification of RNA. In small RNA library construction of NEBNext products of New England Biolabs (MA, USA), the workflow is based on a ligation of 3' and 5' adaptors, whereas in mRNA library construction, cDNA synthesis is based on RNA fragmentation and random priming. Diagenode's (Seraing, Belgium) CATS RNA-seq kit is based on poly(A) tailing and the ligation-free 'Capture and Amplification by Tailing and Switching' method. In the common Illumina (CA, USA) library preparation workflow, random fragmentation of the cDNA sample is followed by 5' and 3' adaptor ligation. Machida and Lin [7] describe in their article other commercial library preparation kits, such as SMART cDNA library construction kit of Clontech and Exact START Eukaryotic mRNA 5'- & 3'- RACE kit of Epicenter (WI, USA), for which the procedures are easy to repeat. However, the kits available are either designed to cover the entire length of RNAs, and therefore not efficient for

rRNA sequencing where the full alignment of the fragments would be beneficial, or are rather expensive to use for environmental studies. When we searched a suitable method to perform rRNA 5'-end sequencing of phytoplankton samples, amplification of poly(A)-tailed cDNA products with an oligo(dT) primer preferentially resulted in internal poly(A) priming and truncated amplification fragments, as was warned by Nam *et al.* [8]. Therefore, we recognized a lack of a directional 5'-end sequencing method that could be readily used for preparing barcoded RNA sequencing libraries without the need for gene-specific primers or oligo(dT) priming.

Our target was to construct a library preparation workflow to study the 5'-end of the SSU rRNAs of all organisms without the bias caused by gene-specific primers or poly(A) tailing. We validated the workflow (Figure 1) by exploiting a eukaryotic cell pool consisting of six phytoplankton species: *Diatoma tenuis*, *Melosira arctica*, *Apocalathium malmogiense*, *Kryptoperidinium foliaceum*, *Monoraphidium*

METHOD SUMMARY

The primer-independent RNA library construction workflow includes ligation of the RNA oligo (M13) to the 5'-end of the gel-extracted small subunit ribosomal RNA, and subsequent reverse transcription of the template using a random hexamer primer with a sequencing adapter overhang. Following amplification of the RNA fragments with barcoded Ion Torrent primers, the amplicons are size-selected with magnetic bead purification.

sp. and *Chlorella pyrenoidosa*. The biomass, dry mass, carbon content and cell number values of each species of the pool were determined earlier [1], but symbiotic prokaryotic cells were unidentified. Here, total RNA was extracted from the mock cells (a detailed protocol and list of the reagents is available in the Supplementary Protocol) using a Direct-Zol RNA MicroPrep isolation kit (Zymo Research, CA, USA). 16S/18S rRNA fragments were cut from a precast 1% agarose E-Gel EX gel (Invitrogen, MA, USA) and purified using a Zymoclean Gel RNA Recovery Kit (Zymo Research). The universal M13-RNA forward adapter sequence (5'-UGUAAAACGACGGCCAGU-3') was ligated to the purified 16S/18S rRNA fragments with T4 RNA ligase (Promega, WI, USA) applying the manufacturer's instructions. After purification of the ligation product, Ion Torrent sequencing adapter P1 was exploited as an overhang

in the random primed (5'-CCTCTCTATG GGCAGTCGGTGATNNNNNN-3') cDNA synthesis. Purified cDNA was amplified with a barcoded Ion Torrent sequencing adapter (IonA) with M13-sequence in the 3'-end and P1 as the reverse primer. For one-step size-selection and purification of amplicons, the dual size selection procedure of the ProNex Size-Selective Purification System (Promega) was applied, targeting the selection between 300 and 550 bp. Sequencing was performed with the Ion Torrent Personal Genome Machine (PGM) as described by Mäki *et al.* [1]. The data was analyzed using CLC Genomics Workbench 11 software (www.qiagenbioinformatics.com) using a phytoplankton reference database comprising mock community sequences [1] or Silva v128 16S reference rRNA gene database for prokaryotic analysis and using OTU 97% identity clustering level (Supplementary Figure 2).

Relative abundances of strains were used to calculate Bray-Curtis similarity matrix, based on which nonmetric multidimensional scaling (NMDS) was calculated with 1000 repeats in PRIMER v. 6.1.12 and PERMANOVA+ v.1.0.2 (PRIMER-E/Quest Research Limited, Albany, New Zealand).

Two steps were critical for the workflow. First, when the RNA template is ligated to the M13 RNA oligo, self-ligation of fragments may occur. To avoid this, we tested blocking the free OH-group of the original RNA with ddATP using the TdT enzyme, but this caused extra steps in the process. The other method was to block self-ligation by introducing an excess of M13-RNA adapters during the ligation reaction. When the concentration of the M13-RNA adapter in the ligation reaction was 10,000 times greater than the rRNA concentration, self-ligation of 16S/18S rRNAs was efficiently prevented (Supplementary Figure 1), and the latter procedure was selected.

The second critical step was related to random priming, which was not fully continuous throughout the sequences, but the specific priming site was affected by the primary or secondary structure of the RNA in the first strand cDNA synthesis, as already recognized in previous studies [9,10]. Nonuniform random priming caused noncontinuous length distribution in the final sequencing products. When the final sequencing data were sorted to 150–200 bp, 200–300 bp, 300–500 bp and 150–500 bp size fractions, it was found that the selection of size fraction affected the outcome, the relative abundance of species (Supplementary Figure 2). Strain-specific distribution of sequence lengths was also noticeable when random primed cDNAs of *A. malmogiense*, *Monoraphidium* sp. and *M. arctica* were amplified with the eukaryotic rRNA-specific forward primer Euk1A (5'-CTGGTTGATCCTGCCAG-3') (Supplementary Figure 3A). Adjusting the number on degenerate nucleotides in the random hexamer did not improve the uneven priming pattern, since five, six and seven degenerate bases (N) resulted in a similar patterning, but eight N-bases generated even stronger peak formation, when the RNA of *Monoraphidium* sp. was studied (Supplementary Figure 3B).

To make the random priming more continuous, we tested addition of the organic solvent dimethyl sulfoxide (DMSO), which should improve the efficiency of cDNA synthesis by decreasing the secondary

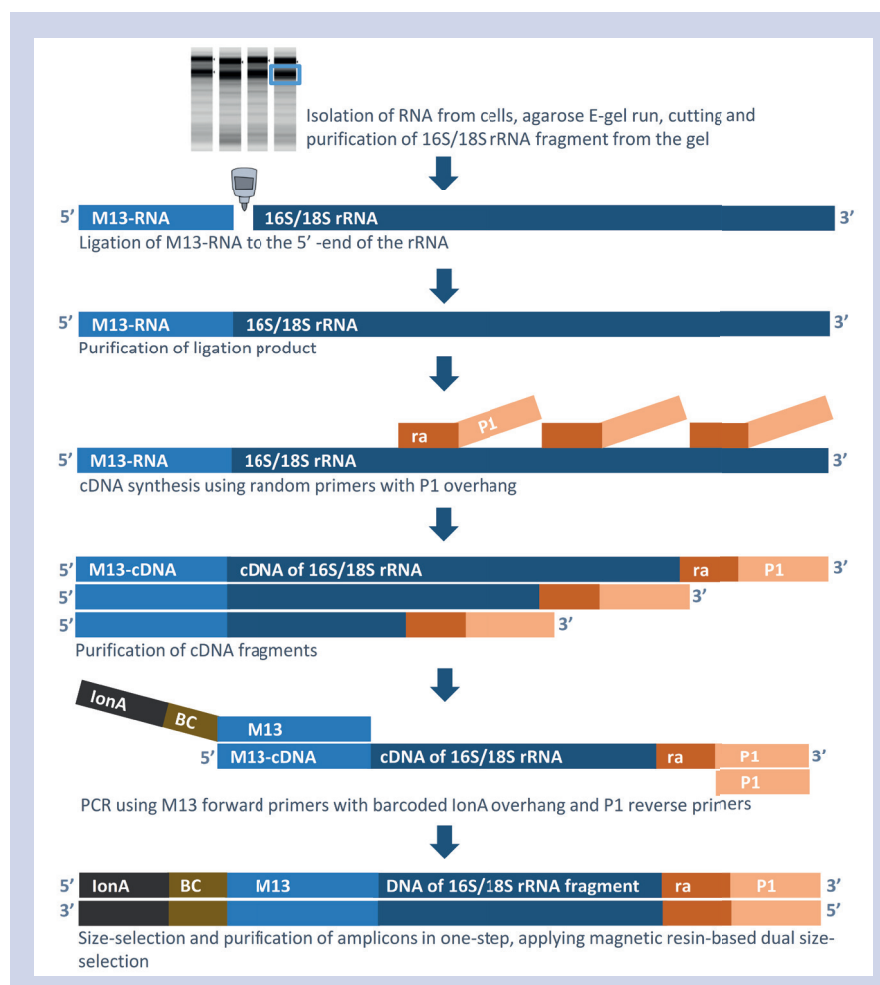


Figure 1. Construction of rRNA sequencing library without gene-specific primers. Ribosomal 16S/18S RNA fragments are extracted from total RNA, M13 RNA oligo is ligated to the purified product, and cDNA synthesis is primed with a random primer containing P1-adapter overhang for the Ion Torrent sequencing. Finally, barcoding and amplification of the construct is performed using the M13 and P1 sites to create the sequencing library.

structures of RNA [11]. However, 10% DMSO did not reduce the patterning in the random priming (Supplementary Figure 3C). We also tested whether the non-continuous random priming was only related to rRNA, which has strong secondary structures, or universal. We used our random priming procedure for the protein-coding RNA of the firefly luciferase (*Fluc*) gene. The RNA was transcribed from control template DNA of the HiScribe T7 Quick High Yield RNA Synthesis kit (New England Biolabs Inc.). The size distribution of amplified cDNA fragments of *Fluc* was not better than that of rRNA (Supplementary Figure 3D), showing that the secondary structures are not the only reason for the noncontinuous random priming pattern of the rRNA. When the tailed random primer P1–6N was manufactured using hand-mixing of the degenerate bases (Integrated DNA Technologies, Freising, Germany) to guarantee that all four bases will be equally represented, size distribution of the PCR-amplified cDNA fragments was more even and spread out to more peaks (Supplementary Figure 3E), suggesting that special handmixing of (tailed) random oligos is recommended and would solve part of the noncontinuous nature of the random priming.

In rRNA analysis every sequence matters, and species-specific differences in sequence lengths may affect the outcome of the analysis if quality trimming selects the data. The effect of the noncontinuous random priming was further avoided by maintaining a large size distribution during the data trimming and analysis. This resulted in a realistic relationship between the mock community species, when sequencing results were compared with defined biomass indicators (Figure 2A & B). In the NMDS ordination of data, our workflow closely resembled dry mass proportions of the cells of the mock pool (Figure 2B), and also prokaryotic partners of the mock community were analyzed at the same sequencing run (Figure 2C).

As even half of the microbial diversity may remain unrevealed using gene-specific primers [12], efficient primer-free techniques are needed to study environmental microbiomes. We predict that the method here described may provide an affordable alternative for commercial kits to study the real diversity of the rRNA world. However, special care has to be given for ensuring that noncontinuous random priming does not affect the results.

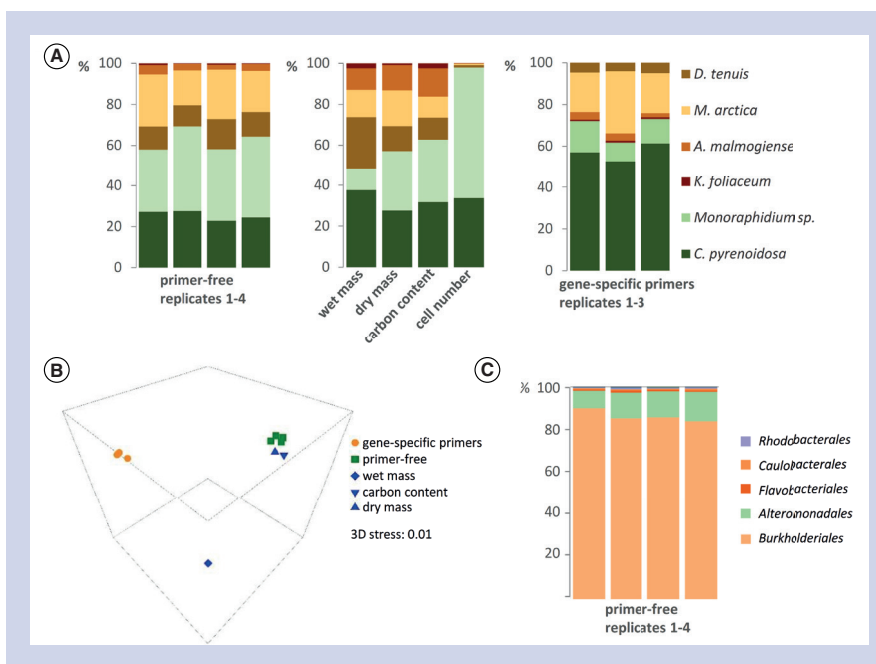


Figure 2. Comparison of the primer-free sequencing of rRNA and biomass indicators of the phytoplankton mock community. (A) Primer-free and gene-specific [1] sequencing of rRNA yielded different proportional relationships between species, where the results of the former better resembled the dry biomass and carbon content of the pool. Wet biomass [1] was determined microscopically. (B) Nonmetric multidimensional scaling ordination showed a close relationship between the primer-free sequencing results and biomass relationships. (C) Primer-free sequencing revealed simultaneously prokaryotic rRNAs of organisms that were grown in symbiosis with the algae cultures (main orders shown).

Acknowledgments

We thank Elina Virtanen for performing the Ion Torrent sequencing.

Author contributions

AM performed the molecular studies and data analysis and wrote the manuscript. MT reviewed and edited the manuscript and offered instructions and intellectual discussion during the study. AM and MT contributed to the study design.

Financial & competing interests disclosure

AM was supported by the European Research Council (ERC) through a grant awarded to MT under the European Union's Seventh Framework Programme (FP/2007-2013, grant 615146). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Supplementary data

To view the supplementary material that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2018-0082

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

- Mäki A, Salmi P, Mikkonen A, Kremp A, Tirola M. Sample preservation, DNA or RNA extraction and data analysis for high-throughput phytoplankton community sequencing. *Front. Microbiol.* 8, 1848 (2017).
- Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat. Microbiol.* 2(3), 16242 (2016).
- van Dijk EL, Jaszczyszyn Y, Thermes C. Library preparation methods for next-generation sequencing: tone down the bias. *Exp. Cell Res.* 322(1), 12–20 (2014).
- Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, Albertsen M. Retrieval of a million high-quality, full-length microbial 16S and



Premium labware for research and discovery

Cell and Tissue Culture

- Three different color coded growth surfaces
- Optimized, user-friendly geometries
- Labeling of all products with LOT number and expiration date



www.sarstedt.com

customerservice@sarstedt.us

18S rRNA gene sequences without primer bias. *Nat. Biotechnol.* 36(2), 190–195 (2018).

- Hoshino T, Inagaki F. A comparative study of microbial diversity and community structure in marine sediments using poly(A) tailing and reverse transcription-PCR. *Front. Microbiol.* 4, 160 (2013).
- Turchinovich A, Surowy H, Serva A, Zapatka M, Lichter P, Burwinkel B. Capture and amplification by tailing and switching (CATS). An ultrasensitive ligation-independent method for generation of DNA libraries for deep sequencing from picogram amounts of DNA and RNA. *RNA Biol.* 11(7), 817–828 (2014).
- Machida RJ, Lin Y. Four methods of preparing mRNA 5' end libraries using the Illumina sequencing platform. *PLoS One* 9(7), e101812 (2014).
- Nam DK, Lee S, Zhou G *et al.* Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc. Natl Acad. Sci. USA* 99(9), 6152–6156 (2002).
- Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* 38(12), e131 (2010).

- van Gorp TP, McIntyre LM, Verhoeven KJ. Consistent errors in first strand cDNA due to random hexamer mispriming. *PLoS One* 8(12), e85583 (2013).
- Yasukawa K, Konishi A, Inouye K. Effects of organic solvents on the reverse transcription reaction catalyzed by reverse transcriptases from avian myeloblastosis virus and Moloney murine leukemia virus. *Biosci. Biotechnol. Biochem.* 74(9), 1925–1930 (2010).
- Hong S, Bunge J, Leslin C, Jeon S, Epstein SS. Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J.* 3(12), 1365–1373 (2009).

First draft submitted: 13 June 2018; Accepted for publication: 6 August 2018

Address correspondence to: Anita Mäki; Department of Biological and Environmental Science, Nanoscience Center, PO Box 35, FI-40014 University of Jyväskylä, Finland; E-mail: anita.maki@jyu.fi

To purchase reprints of this article contact: s.cavana@future-science.com