Journal of Microbiological Methods 75 (2008) 172–176

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Isolation and analysis of mRNA from environmental microbial communities

Ken C. McGrath, Skye R. Thomas-Hall, Chu Ting Cheng, Lesina Leo, Andrei Alexa, Susanne Schmidt, Peer M. Schenk *

School of Integrative Biology, University of Queensland, St Lucia, QLD 4072, Australia

ARTICLE INFO

Article history:
Received 28 January 2008
Received in revised form 12 May 2008
Accepted 19 May 2008
Available online 25 June 2008

Keywords:
cDNA library
Environmental transcriptomics
Metatranscriptomics
Microbiome
mRNA isolation
Unculturable microorganisms

ABSTRACT

The advent of metagenomics has revealed that our planet harbors millions of previously undiscovered microbial species. However, functional insights into the activities of microbial communities cannot easily be obtained using metagenomics. Using transcriptional analyses to study microbial gene functions is currently problematic due to difficulties working with unstable microbial mRNA as a small fraction of total cellular RNA. Current techniques can be expensive and time consuming, and still result in significant levels of rRNA contamination. We have adapted techniques to rapidly isolate high high-quality RNA from environmental samples and developed a simple method for specific isolation of mRNA by size separation. This new technique was evaluated by constructing cDNA libraries directly from uncultured environmental microbial communities, including agricultural soil samples, aquatic flocculants, organic composts, mammalian oral and faecal samples, and wastewater sludge. The sequencing of a fraction of these cDNA clones revealed a high degree of novelty, demonstrating the potential of this approach to capture a large number of unique transcripts directly from the environment. To our knowledge, this is the first study that uses gel electrophoresis to isolate mRNA from microbial communities. We conclude that this method could be used to provide insights into the microbial 'metatranscriptome' of entire microbial communities. Coupled with high-throughput sequencing or the construction of cDNA microarrays, this approach will provide a useful tool to study the transcriptional activities of microorganisms, including those of entire microbial communities and of non-culturable microorganisms.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The extraction of total RNA from Environmental Microbial Communities (EMCs) has been a focus of several publications (Hurt et al., 2001; Zoetendal et al., 2006) and microbial mRNA has previously been analyzed using different approaches. These include the analysis of individual transcripts (Burgmann et al., 2003), differential display (Fleming et al., 1998; Brzostowicz et al., 2003), subtracted libraries (Poretsky et al., 2005) and fluorescent in-situ hybridization (Pernthaler and Amann, 2004). However, the functions of prokaryotic genes and their transcripts remain difficult to study due to technical problems related with the isolation of mRNA. The pool of total RNA consists predominantly of ribosomal RNA, with approximately 1–5% mRNA (Neidhardt and Umbarger, 1996). Additionally, prokaryotic transcripts usually have no poly(A) tail, and because of simultaneous transcription and translation, mRNA is usually fragmented and unstable (Nakazato et al., 1975). Because of these factors, the separation of mRNA from the total RNA pool is difficult, and cDNA libraries are dominated by rRNA clones (Botero et al., 2005). While methods are available that allow partial removal of 23S and 16S ribosomal RNA (MICROB*Express*™, Ambion; mRNA-ONLY, EpiCentre), these techniques are very limited in their species range and ability to remove all forms of rRNA (e.g. 5S rRNA) (Poretsky et al., 2005).

Here we report a new approach for isolating high-quality mRNA from diverse EMCs from terrestrial, commensal and aquatic sources, and the construction of cDNA libraries. Sequence analysis of a fraction of the clones revealed a high degree of diversity and novelty. This simple method will assist research in the field of metatranscriptomics, and our results emphasize the need for future large-scale analyses of microbial community transcriptomes.

2. Materials and methods

2.1. Total RNA isolation

2.1.1. Terrestrial soil samples

Several soil types, including organic compost and garden topsoil from a private household (Brisbane, Australia), and sugarcane soil (Jacob's Well, Queensland, Australia) were used for mRNA isolation. The initial stages up to freezing of the pellet were performed directly on site, to pre-process the samples and freeze them to minimize changes to the RNA profiles and degradation. For each sample, approximately 20 g of soil was placed in a 50 mL tube, and 20 mL of

* Corresponding author. Level 5, John Hines Building, School of Integrative Biology, University of Queensland, St Lucia, QLD 4072, Australia. Tel.: +61 7 33658817; fax: +61 7 33651699.

E-mail address: p.schenk@uq.edu.au (P.M. Schenk).

0167-7012/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2008.05.019

sterile distilled water (Milli-Q) was added. The tube was shaken vigorously until the soil sample was completely suspended in the water (20–30 s). The tube was left to settle for 10 s to allow larger particles (i.e. stones) to fall out of suspension, and the supernatant was decanted into individual microcentrifuge tubes (2 mL), and centrifuged for 2 min (14,000 ×g) to pellet the microbial contents. The supernatant was discarded, and the microbial pellet was frozen and stored at −80 °C. Total RNA was isolated from the frozen pellet using the PowerSoil™ RNA extraction kit (MoBio, USA), with the first two buffers being added to the frozen pellet simultaneously.

2.1.2. Aquatic samples

Aquatic samples were collected from a highly eutrophic freshwater lake (University of Queensland, St Lucia Campus, Australia), as well as activated flocculant (Oxley waste water treatment plant, Brisbane, Australia). For each sample, 10 Falcon tubes with 50 mL of aquatic sample each were collected and centrifuged for 2 min at 14,000 ×g. The supernatant was discarded, and the unfrozen pellets were used directly for RNA extraction immediately using the SV Total RNA Isolation Kit (Promega).

2.1.3. Commensal samples

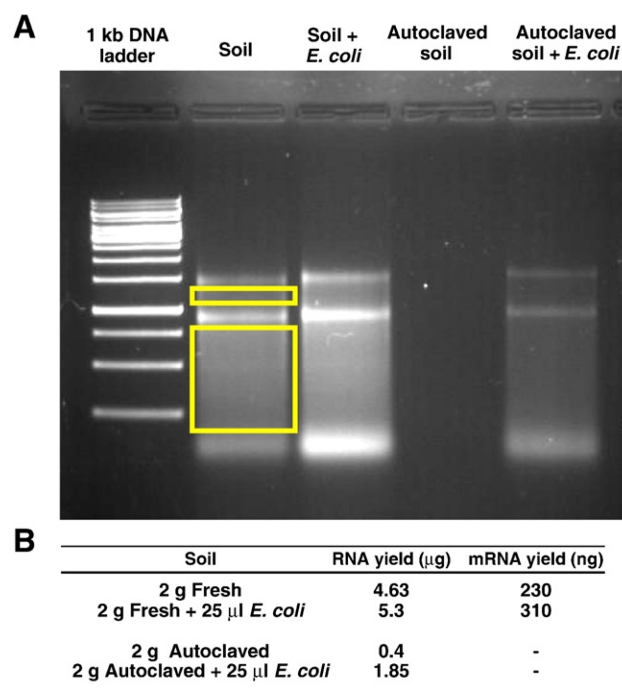
Commensal samples including bovine rumen material, human oral samples from teeth, tongue and inner cheek, and human faecal samples were used for extraction. Approximately 200–400 mg of samples were was added to a 2 mL microcentrifuge tube and snap-frozen at −80 °C. These samples were then processed using the SV Total RNA Isolation Kit (Promega).

2.2. mRNA isolation

This step was carried out in an RNase-free environment, with all surfaces coming in contact with the RNA sample cleaned with RNase away™ (Invitrogen, Australia). All solutions were made from RNase-free stocks and DEPC-treated water. All pipette tips were RNAse-free. Each total RNA sample was run on a 1.5% agarose gel containing ethidium bromide at 100 V for 45 min in 1×TAE buffer (40 mM Tris-base; 20 mM acetic acid; 1 mM ethylene diamine tetraacetic acid pH 8.0). The 23S, 16S and 5S rRNA bands were identified on the gel using UV illumination (Fig. 1A). The regions between the 23S and 16S, and the 16S and 5S rRNA bands were excised, and the RNA was extracted from the agarose using the Wizard SV Gel and PCR Clean-Up System (Promega).

2.3. cDNA synthesis, cloning and sequence analysis

The isolated mRNA (ranging between 200–500 ng for each sample) was vacuum-concentrated to a volume of 15 μl μL and the first cDNA strand was synthesized using random hexamers (0.3 μg) and SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacuer's instructions. Double-stranded cDNA was synthesized by adding the following to the initial cDNA reaction: 13 μL of 10× *E. coli* Ligase Buffer, 1 μL of 10 U/μL *E. coli* Ligase, 4 μL of 10 U/μL Klenow Fragment, 1 μL of 1 U/μL RNase H, 3 μL of 10 mM dNTP's and 106 μL of Milli Milli-Q water to a final volume of 150 μL. The reaction was incubated at 16 °C for 2 h. A total of 1.5 μL of T4 DNA polymerase (7.9 U/μL) was added to the reaction, followed by incubation for further 5 min to ensure the presence of blunt ends. The double-stranded cDNA was purified using the Wizard SV Gel and PCR Clean-Up System (Promega), and vacuum-concentrated to a final volume of 5 μL in preparation for cloning. The cDNA was then cloned and transformed into *E. coli* Top10 cells using the pCR-Blunt system (Invitrogen). Each cDNA library yielded 1000–4000 colonies. A total of 13,056 individual white colonies were cultured in 2 mL of LB broth, and the plasmid DNA of 100 clones was extracted and sequenced by the Australian Genome Research Facility. Both forward and reverse primers (m13-F and -R) were used to generate the sequence data, and vector sequence was removed. cDNA clone sequences were submitted to the NCBI database under the accession numbers ES544490–ES544589. All sequences were analyzed using blastn and blastx homology alignments to Genbank using the default settings (database date July, 2007) and the details of the lowest *E*-value return from either search were recorded (Supplementary Table 1).



Soil	RNA yield (μg)	mRNA yield (ng)
2 g Fresh	4.63	230
2 g Fresh + 25 μl *E. coli*	5.3	310
2 g Autoclaved	0.4	-
2 g Autoclaved + 25 μl *E. coli*	1.85	-

Fig. 1. A) Agarose gel image showing total RNA separation from 4 separate extractions from fresh or autoclaved soil with or without the addition of *E. coli*. Highlighted regions, containing mRNA, were used for extraction and cDNA synthesis. Note that the first lane includes a 1 kb DNA GeneRuler™ ladder (Fermentas) and hence does not correspond to RNA sizes. B) Summary of extraction efficiency, comparing yields of total RNA and mRNA.

3. Results

3.1. Total RNA extraction from diverse environmental microbial communities

Methods for total RNA extraction were performed for a diverse selection of microbial communities. Initial attempts to isolate total RNA from these EMCs had varied success, and several RNA extraction protocols were tested before a successful method for each sample matrix was developed. Three methods were developed for three general microbial community descriptions: terrestrial, aquatic and commensal. To obtain high-quality RNA from soil samples it was best not to freeze complete soil immediately after sampling, but to employ a quick pre-processing step on site to extract a microbial suspension from soil particles (see Materials and methods). For commensal samples (bovine rumen, human oral, and human intestinal), total RNA could be isolated with minimal sample processing when the samples were frozen with a large surface area (as a smear on the inside of a tube) and then rapidly thawed while resuspending the sample in the presence of lysis buffer. RNA from aquatic samples (waste water and pond water) could be extracted using the same method as commensal samples, after the aqueous samples were centrifuged to pellet their microbial contents. To investigate the RNA yields achieved using this method, an extraction from soil was carried out, using either fresh or autoclaved soil. To compare this method to RNA extractions from culturable bacteria, 25 μl of log-phase *E. coli* culture was added as a

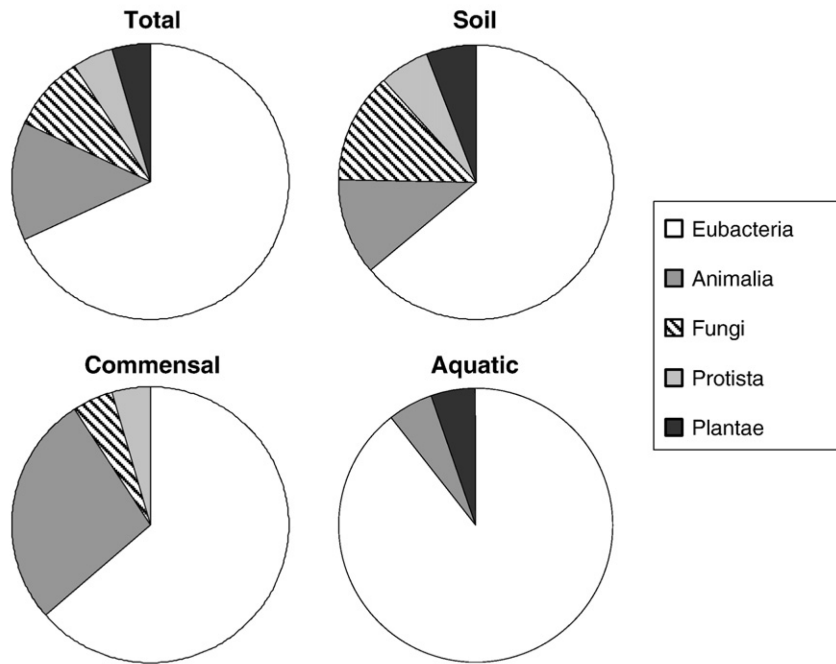


Fig. 2. Distribution of cDNA sequences obtained from different environmental microbial communities into taxonomic domains based on the highest BLAST match to existing sequences in Genbank.

known mRNA source. High High-quality RNA was indicated by minimal rRNA degradation, visible as distinct rRNA bands and the absence of fragmentation (smearing) after gel electrophoresis (Fig. 1A). A yield of 2.3 µg of total RNA per g soil (115 ng mRNA/g soil) was obtained when using fresh soil, while log-phase *E. coli* culture resulted in approximately 60 µg total RNA per ml. As expected, autoclaving the soil resulted in no visible RNA being isolated from the soil (Lane 4, Fig. 1A) although a low OD^{260} reading indicated the presence of a small residual amount (Fig. 1B). Furthermore, the addition of 25 µl *E. coli* culture to 2 g of fresh soil increased the total yield to 2.65 µg of total RNA per g soil (155 ng mRNA/g soil), although not by the same amount as when added to autoclaved soil.

3.2. mRNA isolation from total RNA

The pool of total RNA consists predominantly of ribosomal RNA, with approximately 1–5% mRNA (Neidhardt and Umbarger, 1996). We investigated an existing method of removing rRNA from total RNA (MICROB*Express*™, Ambion), but found that rRNA bands were still visible after gel electrophoresis, indicating that this method was not successful for removing all rRNA (data not shown). This residual rRNA interferes with any analysis of the mRNA transcripts present in the sample. Using size fractionation of total RNA by agarose gel electrophoresis, 23S, 16S and 5S rRNA bands were clearly distinct and could be efficiently removed by excision of the agarose between ribosomal bands (highlighted in Fig. 1A). The mRNA was subsequently purified from the agarose, producing an mRNA mixture containing transcripts of various sizes. This method was able to yield mRNA concentrations that were equivalent to approximately 5% of the total RNA concentration (Fig. 1B), although the mRNA may still contain trace levels of rRNA that migrates outside of the 3 major bands.

3.3. Sequence analysis of mRNA transcripts from EMCs

The mRNA isolated from terrestrial, aquatic and commensal microbial communities was used to construct plasmid cDNA libraries in *E. coli*. Out of the 13,056 clones produced, a random sample of 100 cDNA clones was selected and sequenced, representing transcripts

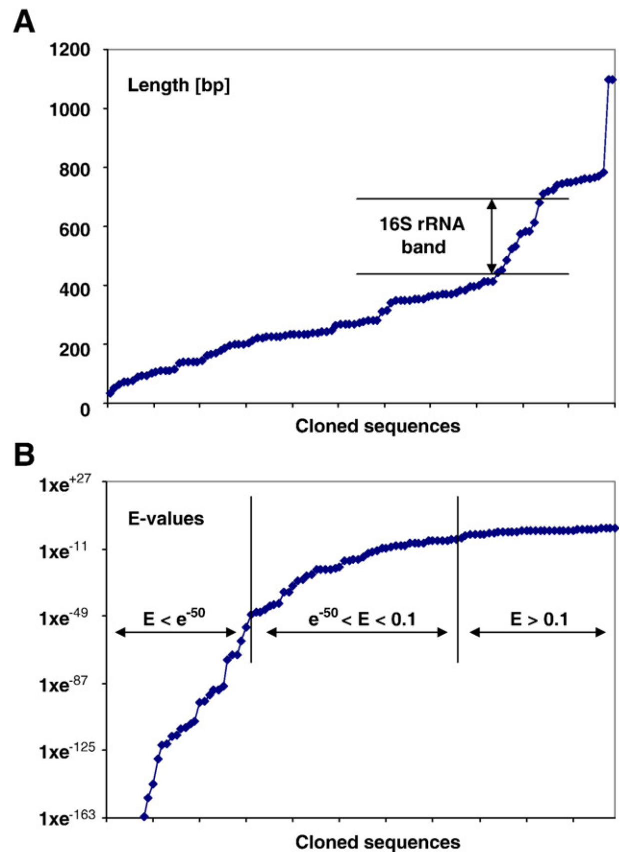


Fig. 3. Distributions of sizes (A) and *E*-values (B) of cDNA transcripts isolated from environmental microbial communities. *E*-values were obtained from BLAST searches of sequenced cDNA. Note the low frequency of sizes around the region occupied by the 16S rRNA band (A). The cutoffs used to discuss novelty of the obtained sequences ($E > e^{-50}$ and $E > 0.1$) are indicated by vertical lines (B).

from each of the EMCs (19 from aquatic, 22 from commensal, and 59 from terrestrial soil). Microbial cDNA sequences were compared to Genbank using the BLAST algorithm for nucleotides (blastn) and proteins (blastx). All accession numbers, as well as an analysis of BLAST results, are listed in Supplementary Table 1. Each cDNA sequence was assigned to a group of organisms, based on the lowest *E*-value returned during blastn and blastx analyses (Fig. 2). The majority of sequences returned matches to bacteria (68%), followed by animal (13%), fungi (9%), protista and plants (4.5% each; Fig. 2). Notably, soil samples showed a higher content of fungal sequences (13%) while commensal microbial communities contained more sequences with homology to animals (27%, mostly from the host species) and none from plants. None of the sequences showed a close match to an existing Archaea sequence entry.

Eleven of the 100 sequences matched to rRNA, but most of these rRNA sequences showed the closest homology to eukaryotic organisms, and hence may not have been accurately excluded during the mRNA isolation recovery by gel electrophoresis, given that we excised between the regions of prokaryotic rRNA (Fig. 1A). It is likely that the other rRNA clones from prokaryotic sources originate from partial rRNA products migrating with mRNA or the limitation of size separation by gel electrophoresis. The size of the analyzed cDNA sequences reflected the general distribution of mRNA transcripts captured by this method (Fig. 3A). Most of the sequences were either less than 450 bp, or between 700 and 800 bp in size. This corresponds to the two excised regions used to isolate the mRNA from total RNA. It should be noted that mRNA transcripts that co-migrate with the rRNA bands are not isolated with this method, although partial products of such transcripts may still be represented.

The levels of homology of the cDNA sequences to existing database entries revealed that this approach recovered mostly new sequences from EMCs. Only seven of the analyzed transcripts returned an *E*-value of zero (Supplementary Table 1), indicating a perfect match to existing genetic data. The distribution of *E*-values (Fig. 3B) shows that most sequences (71%) had *E*-values greater than 10^{-50}, and 19% of sequences had an *E*-value greater than 0.1, reflecting the ability of this method to capture novel mRNA transcripts directly from microbial communities.

4. Discussion

The isolation and analysis of mRNA transcripts from environmental microbial samples is an important step to increase our understanding of the complex processes of microbial ecology. Environmental transcriptomics or metatranscriptomics target functional gene expression within microbial communities without bias towards known sequences, and provide a new approach for identifying and analysing analyzing community-specific variants of key functional genes (Poretsky et al., 2005). Recently, Rohwer (2007) envisaged that “massive sequencing of RNA populations will become routine and replace the current array technologies”. We have employed a novel yet simple technique using size separation of mRNA from high-quality total RNA that provides opportunities to study the gene expression patterns of entire environmental microbial communities that contain many currently unculturable organisms. This approach was suitable for extracting viable mRNA from a range of EMCs, and from various and distinct environmental samples. These included terrestrial (organic compost soil, garden topsoil, sugarcane field soils), aquatic (eutrophic fresh-water lake, activated flocculant of communal waste water) and commensal (cow rumen, human oral; human faecal) samples. Variations from this protocol (e.g. extraction from fresh or frozen soil without pre-processing step) and many other techniques (e.g. other commercial kits; methods using CTAB or LiCl) gave a smear of rRNA when viewed on a gel. The method presented here can be performed using commonly found laboratory equipment, and results in considerably less rRNA sequences in libraries than other methods (Poretsky et al., 2005). Limitations include a possible bias towards microbes that are more stable, easier to pellet and lyse, and towards fragments that do not co-migrate with rRNA during gel electrophoresis. Also, small fragments are easier to convert into cDNA and to clone into libraries.

None of the sequences in our study showed a close match to an existing Archaea sequence entry. This may be due to a low abundance of transcriptionally-active Archaea in the analyzed samples, or limited entries for non-extremophile Archaea in the public databases. Most BLAST searches did not yield a close match to any existing Genbank entry and the attributed kingdoms may be biased towards the number of available sequences. A low (2%) degree of redundancy was observed in the cDNA libraries, and only two sequences (matching to coding sequence of *Escherichia coli* aquaporin Z and a putative helicase from *Geobacter sulfurreducens*) appeared twice in the 100 sequenced clones (Supplementary Table 1). However, the fact that these sequences were identical and had the same length suggests that these may stem from *E. coli* cell divisions during the 1 h incubation step after transformation with cDNA ligation products rather than a true representation of redundancy in these microbial communities. Large-scale sequencing of microbial cDNA clones should be carried out to more accurately reveal the diversity of expressed genes in the EMCs described here.

Identifying key functional genes and profiling the transcriptomes of microbial communities offer a broad range of biotechnological and medical applications. The use of this technique to isolate mRNA from diverse microbial communities will most certainly aid researchers to reveal expressed genes with important functions and to establish microbial transcriptional activity profiles, including those present in complex EMCs with unculturable species. For example, cDNA clones can be used to construct custom microarrays to analyze the expression of thousands of microbial genes from EMCs in parallel and our preliminary experiments using cDNA microarrays containing all 13,056 anonymous clones showed reproducible expression patterns for independent biological replicates (unpublished data). It has also been pointed out that genes from highly-diverse EMCs may encode many novel biologically active peptides (Blake, 2004) and expression libraries generated using this method may provide natural compound collections suited for future biodiscovery projects.

Acknowledgements

We wish to thank Prof. Linda Blackall for helpful discussions and acknowledge the Australian Greenhouse Office (to S.S. and P.M.S) for funding of this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mimet.2008.05.019.

References

Blake, D., 2004. Biodiscovery — from reef to outback. Nature 429, 15–17.

Botero, L.M., D'Imperio, S., Burr, M., McDermott, T.R., Young, M., Hassett, D.J., 2005. Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. Appl. Environ. Microbiol. 71, 1267–1275.

Brzostowicz, P.C., Walters, D.M., Thomas, S.M., Nagarajan, V., Rouviere, P.E., 2003. mRNA differential display in a microbial enrichment culture: simultaneous identification of three cyclohexanone monooxygenases from three species. Appl. Environ. Microbiol. 69, 334–342.

Burgmann, H., Widmer, F., Sigler, W.V., Zeyer, J., 2003. mRNA extraction and reverse transcription-PCR protocol for detection of nifH gene expression by *Azotobacter vinelandii* in soil. Appl. Environ. Microbiol. 69, 1928–1935.

Fleming, J.T., Yao, W.H., Sayler, G.S., 1998. Optimization of differential display of prokaryotic mRNA: application to pure culture and soil microcosms. Appl. Environ. Microbiol. 69, 3698–3706.

Hurt, R.A., Qiu, X., Wu, L., Roh, Y., Palumbo, A.V., Tiedje, J.M., Zhou, J., 2001. Simultaneous recovery of RNA and DNA from soils and sediments. Appl. Environ. Microbiol. 67, 4495–4503.

Nakazato, H., Venkatesan, S., Edmonds, M., 1975. Polyadenylic acid sequences in *E. coli* messenger RNA. Nature 256, 144–146.

Neidhardt, F.C., Umbarger, H.E., 1996. In: Neidhardt, F.C. (Ed.), *Escherichia coli and Salmonella*: Cellular and Molecular Biology, 1. ASM Press, Washington, DC, pp. 13–16.
Pernthaler, A., Amann, R., 2004. Simultaneous fluorescence in situ hybridization of mRNA and rRNA in environmental bacteria. Appl. Environ. Microbiol. 70, 5426–5433.
Poretsky, R.S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W.M., Moran, M.A., Hollibaugh, T.J., 2005. Analysis of microbial gene transcripts in environmental samples. Appl. Environ. Microbiol. 71, 4121–4126.
Rohwer, F., 2007. Real-time microbial ecology. Environ. Microbiol. 9, 10.
Zoetendal, E.G., Booijink, C.C., Klaassens, E.S., Heilig, H.G., Kleerebezem, M., Smidt, H., de Vos, W.M., 2006. Isolation of RNA from bacterial samples of the human gastrointestinal tract. Nat. Protoc. 1, 954–959.