CHAPTER TWELVE

Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis

Brandon M. Satinsky[*], **Scott M. Gifford**[†], **Byron C. Crump**[‡], **Mary Ann Moran**[§,1]

[*]Department of Microbiology, University of Georgia, Athens, Georgia, USA
[†]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[‡]College of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA
[§]Department of Marine Sciences, University of Georgia, Athens, Georgia, USA
[1]Corresponding author: e-mail address: mmoran@uga.edu

Contents

Abstract

Next generation sequencing-enabled metatranscriptomic and metagenomic datasets are providing unprecedented insights into the functional diversity of microbial communities, allowing detection of the genes present in a community as well as differentiation of those being actively transcribed. An emerging challenge of meta-omics approaches is how to quantitatively compare metagenomes and metatranscriptomes collected across spatial and temporal scales, or among treatments in experimental manipulations.

Methods in Enzymology, Volume 531
ISSN 0076-6879
http://dx.doi.org/10.1016/B978-0-12-407863-5.00012-5

Here, we describe the use of internal DNA and mRNA standards in meta-omics methodologies, and highlight how data collected in an absolute framework (per L or per cell) provides increased comparative power and insight into underlying causes of differences between samples.

1. INTRODUCTION

Metagenomic and metatranscriptomic methodologies have been used with great success in generating detailed information on community-level gene abundance and transcription patterns in marine, freshwater, soil, gut, and other natural microbial systems (Damon et al., 2011; Dinsdale et al., 2008; Gifford, Sharma, Rinta-Kanto, & Moran, 2011; Maurice, Haiser, & Turnbaugh, 2013; Ottesen et al., 2013; Poretsky et al., 2005; Vila-Costa, Sharma, Moran, & Casamayor, 2013). Most studies to date have collected meta-omics data in a relative framework, in which abundance of genes or messages is calculated as percent of the sequence library (Campbell, Yu, Heidelberg, & Kirchman, 2011; Hewson et al., 2009). However, a critical limitation of relative meta-omics data from complex natural communities is that they cannot provide information on the extent or directionality of changes in any particular gene or transcript molecule in comparative analyses. For instance, an observed decrease in the percent contribution of a transcript to the community metatranscriptome may be due to a decrease in the abundance of that transcript or to an increase in the abundance of an unrelated transcript (Fig. 12.1). In the application of meta-omics technologies to ecological and biogeochemical questions in complex microbial communities, the ability to recognize which genes and transcript molecules are changing in absolute abundance is crucial information, requiring datasets that are not influenced by the myriad nontarget processes and taxa changing simultaneously in a microbial cell or ecosystem.

To circumvent the limitations of relative metagenomic and metatranscriptomic datasets, internal genomic DNA or mRNA standards can be added at the initiation of sample processing (Gifford et al., 2011; Moran et al., 2013). Because these control molecules are mixed into and processed alongside the sample-derived nucleic acids, this allows quantification of losses throughout the preparation and analysis pipeline and, based on the number of standard molecules added at the beginning of sample processing and those recovered in the sequence library, calculation of the number of molecules of each gene or transcript in the original environment

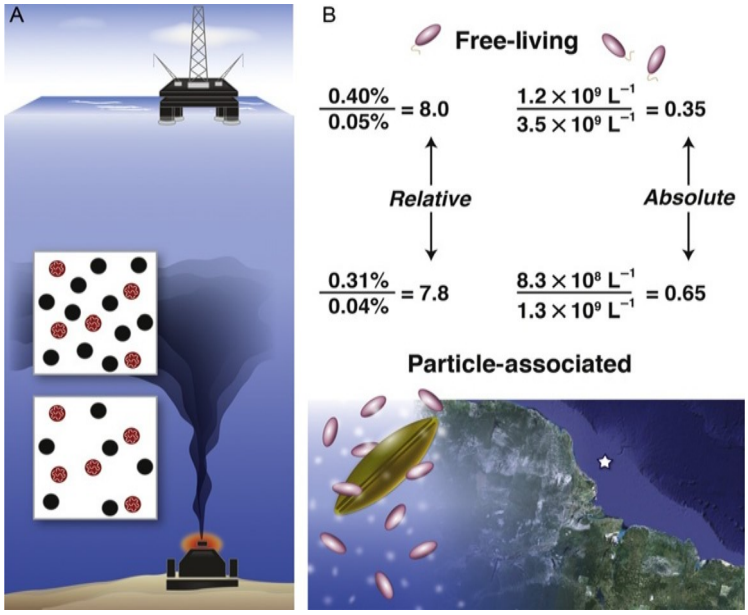

Figure 12.1 Examples of improved quantification of meta-omics data through the use of internal standards. (A) Transcripts binning to SAR11 member HTCC7211 accounted for a smaller fraction of the community metatranscriptome in the oil plume caused by the Deepwater Horizon accident compared to nonimpacted control samples below the plume, yet the absolute number of transcripts contributed by this taxon was not different. HTCC7211 is one of several bacteria taxa dominant in the prespill community that did not respond to the presence of oil, whereas other taxa greatly increased in number and activity in the hydrocarbon-impacted seawater (from Rivers et al., 2013). (B) Particle-associated bacteria in the Amazon River plume in June 2010 had twofold higher expression of proteorhodopsin genes than free-living bacteria, yet expression estimates calculated incorrectly from relative data would not have shown the differential regulation of this ecologically important gene (Satinsky B. et al., unpublished).

(e.g., gene copies per liter of water or average transcripts per microbial cell). Internal standards based on a known quantity of added control molecules are used routinely in quantitative PCR studies for calculating absolute gene and transcript abundance (Church, Short, Jenkins, Karl, & Zehr, 2005) and in microarray and RNA-seq analyses to normalize expression shifts in genes across different developmental stages or tissue types (Hannah, Redestig, Leisse, & Willmitzer, 2008; van de Peppel et al., 2003).

The benefits of quantitative meta-omics datasets can be illustrated by the following two examples. In the first, sequences binning to SAR11 member HTCC7211 in the bathypelagic waters of the Gulf of Mexico accounted for

0.84% of the bacterial metatranscriptome in natural seawater but only 0.07% in seawater exposed to oil and gas contamination from the Deepwater Horizon accident, indicating a 12-fold underrepresentation of HTCC7211 following the accident. Yet absolute transcript numbers for this taxon calculated based on internal standard normalization revealed that transcripts were present in equal numbers in impacted and nonimpacted seawater (2.8×10^{11} and 3.4×10^{11} transcripts/L; Fig. 12.1), and that the change in percent contribution of HTCC7211 populations was due to large increases in gammaproteobacteria groups that bloomed in response to hydrocarbon inputs (Fig. 12.1; Rivers et al., 2013). In a second example, expression ratios for proteorhodopsin genes in the near-shore Amazon River plume were nearly identical for the free-living and particle-associated bacteria when calculated on a relative basis (% of the metatranscriptome/% of the metagenome ≈ 8 for both free-living and particle-associated; Fig. 12.1). Yet on an absolute basis, the per-gene transcription level of proteorhodopsin was twofold higher for bacteria associated with particulate material compared to free-living cells in this ecosystem (Fig. 12.1). In these examples, normalization based on internal standard recovery provided insights into growth and regulation differences for bacteria in their natural environment, information that can be leveraged in comparative analyses across samples (e.g., within a time series, across a transect, or during a manipulative experiment) (Gifford et al., 2011; Moran et al., 2013; Fig. 12.1).

To generate quantitative-omics data, internal control sequences must be readily distinguished from natural microbial community sequences during bioinformatic analyses. For metatranscriptomes, artificial mRNAs produced by *in vitro* transcription from constructed DNA templates can be used as internal standards and preparation of mRNA standards with or without a poly(A) tail customizes them for bacterial/archaeal or eukaryotic studies. For metagenomes, genomic DNA obtained from a cultured microorganism not present in the studied environment can be added as an internal standard. In our marine and estuarine studies, DNA from the thermophilic bacterium *Thermus thermophilus* (ATCC) has served as the standard.

Calculations based on the internal standards assume that the natural nucleic acid (mRNA or genomic DNA) and the internal standards (artificial transcripts or exogenous genomic DNA) behave similarly throughout the sample and library preparation steps. However, the natural nucleic acid is enclosed in cell membranes at the initiation of processing while the internal standards are not, potentially resulting in underestimation of natural nucleic acid abundance due to incomplete cell lysis, or alternatively,

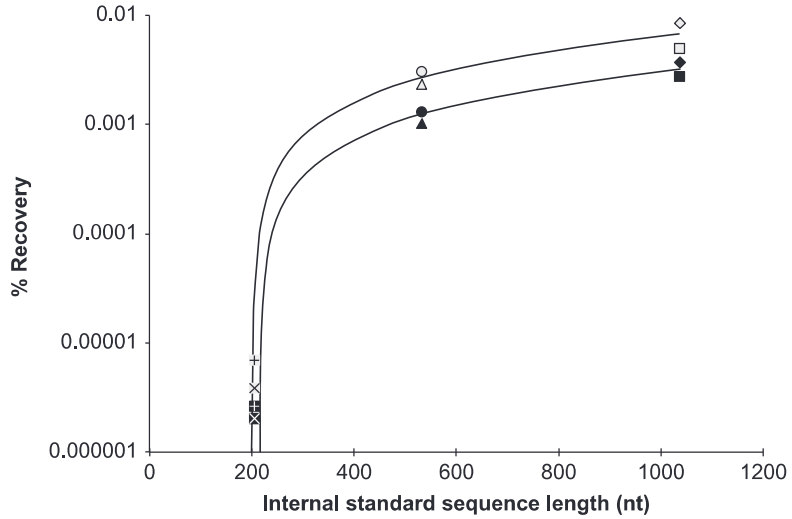

Figure 12.2 Recovery of internal mRNA standards as a function of standard length for two replicate metatranscriptome libraries (black series and gray series) from the Amazon River near Tapajos in June 2011. Two different internal standards of three different lengths (each represented as a different shaped symbol) were added to the samples at the initiation of nucleic acid extraction and percent recoveries were calculated as S_s (internal standard reads in the sequence library) $\times$ 100/S_a (internal standards added to the sample).

underestimation of standard abundance due to longer exposure time to mechanical shearing or RNAse degradation. In the case of mRNA processing, transcript length can affect recovery because of biases against small transcripts during solid-phase extraction methods and library preparation (Fig. 12.2), although this will affect both artificial and natural transcripts alike. Given an average bacterial and archaeal gene size of 924 bp (Xu et al., 2006), we use an internal standard of ~1000 nt to track recovery of mRNAs from typical prokaryotic genes. Internal standard length can be scaled down if small RNAs or short transcripts are the focus of the study; scaling up from 1000 nt does not appear to be necessary because of minimal effect on recovery for lengths >500 nt (Fig. 12.2).

2. METHOD OVERVIEW

The method given here describes the synthesis of internal mRNA standards and then the addition and quantification of mRNA and DNA internal standards for metatranscriptome and metagenome analysis.

mRNA standards are synthesized using custom templates or commercially available plasmids that are transcribed *in vitro* to RNA. A known number of standards are added to the sample of interest and metatranscriptome processing and sequencing proceeds according to the user's protocol. The number of internal standards recovered in the sequence library is quantified via BLAST homology searches. Data normalization is then based on the number of standards identified in a sequence library relative to the number of standards added. As a note of caution for working with RNA, care should be taken to avoid all contaminating nucleic acids and nucleases through the use of sterile technique and cleaning the working area with RNase*Zap*® or a similar reagent.

DNA standards can be prepared by purchasing or extracting DNA from a cultured microbe that is unrelated to microbes anticipated to be present in the system of interest and for which a complete genome sequence is available. A known number of genome copies are added to the sample, and metagenome processing and sequencing proceeds according to the user's protocol. The number of standard reads recovered in the sequence library is quantified via a two-step BLAST homology search and used for quantitative metagenomic analysis.

3. DNA TEMPLATE AND VECTOR DESIGN FOR INTERNAL RNA STANDARDS

Two approaches are available for obtaining the DNA template for standard synthesis. One approach involves commercially available plasmids that contain an RNA polymerase binding site. These are advantageous because of ease of use and low cost (Gifford et al., 2011; Moran et al., 2013), although the vectors make transcript length customization more difficult and they often contain regions of homology to functional proteins or to sequences deposited mistakenly into databases as functional proteins. This homology can make the subsequent identification of reads derived from standards more challenging in a high-throughput bioinformatics pipeline. A second approach involves the synthesis of custom DNA fragments that are inserted into plasmids. These fragments can easily be designed without homology to protein encoding genes, and provide optimal control of both length and composition.

For both template approaches, the final plasmid should contain the following components (in order): a T7 RNA polymerase promoter sequence, the internal standard sequence, and a restriction site targeting a unique site in

the plasmid and preferably producing a blunt end (Fig. 12.3). For poly-A selective transcriptomes, a poly-A tail can be included in custom synthesized templates between the RNA polymerase promoter and the internal standard sequence. Whether using commercially available plasmids or custom synthesized internal standard templates, sequences should first be analyzed against relevant databases to identify regions of homology that could interfere with unambiguous identification of the standard in the sequence library.

Template size is also an important consideration because downstream processing steps during RNA processing and library preparation can lead to biases in the size of transcripts recovered. Based on addition of the six standards shown in Fig. 12.2 (representing two variations in base composition for each of three sizes: 200, 500, and 1000 nt), recovery efficiency in the sequence library was several orders of magnitude lower for the 200 nt mRNA standards compared to the others (Fig. 12.2). However, the duplicate standards at each size were recovered with nearly identical efficiencies, indicating that base composition is not an important factor in standard

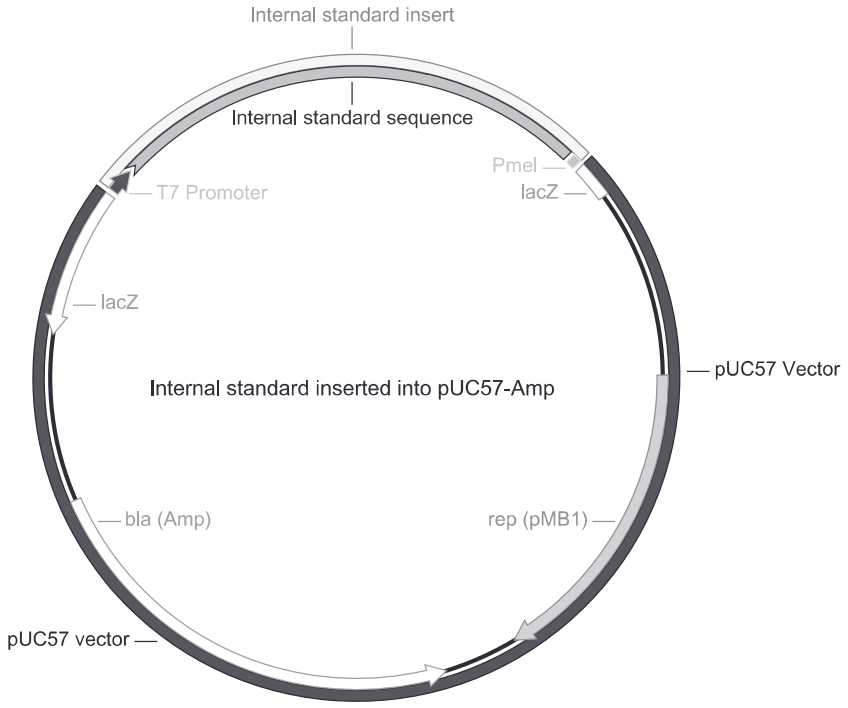

Figure 12.3 Genetic construct for *in vitro* transcription of a synthetic mRNA internal standard.

recovery. For the sequence data represented in Fig. 12.2, steps in the RNA isolation, purification, and amplification relied on solid-phase extraction, while Illumina library preparation included cDNA shearing and size selection (225 bp target size), all of which could lead to size bias for both artificial and natural mRNAs. Other extraction and library preparation methods may result in different size biases, but it is not straightforward to correct for size biases as transcript length depends on operon structure rather than individual gene length. Nonetheless, an internal standard can be selected that approximates the average size of the natural nucleic acid molecules being targeted (i.e., genomic DNA standards for metagenomes and artificial mRNAs of typical gene length for metatranscriptomes).

4. mRNA STANDARD PREPARATION

4.1. Required materials

- *Equipment*: 4 °C microcentrifuge, 10-, 20-, 200-, and 1000-μL pipettes, water bath, 37 °C shaking incubator, thermocycler, gel electrophoresis equipment and reagents, microfluidic electrophoresis instrument or fluorometry-based instrument for measuring nucleic acid concentration.
- *Media*: LB agar, LB agar + ampicillin (100 μg/mL final concentration), LB medium, LB medium + ampicillin (100 μg/mL final concentration), SOC medium (2% tryptone, 0.5% yeast extract, 10 m*M* sodium chloride, 2.5 m*M* potassium chloride, 10 m*M* magnesium chloride, 10 m*M* magnesium sulfate, 20 m*M* glucose).
- *Bacterial cell line*: One Shot® Top10 Chemically Competent *Escherichia coli* (Life Technologies, Grand Island, NY).
- *Template DNA*: Custom synthesized DNA template (T7 RNA polymerase promoter, internal standard sequence, unique restriction site) inserted into a plasmid.
- *Restriction digest and end repair*: Restriction enzyme matching unique restriction site and corresponding buffers, mung bean nuclease for end repair on digests that do not produce blunt ends.
- *Commercially available kits*: Ambion MEGAscript® T7 Kit (Life Technologies), Quant-iT™ RiboGreen® RNA Assay Kit (Life Technologies), miniPrep plasmid extraction kit.
- *Other reagents*: phenol:chloroform:isoamyl alcohol (24:24:1, pH ~7), citrate-saturated phenol:chloroform:isoamyl alcohol (24:24:1, pH 4.7),

sterile 2-propanol, ice cold 70% ethanol, nuclease-free 3 *M* sodium acetate, nuclease-free TE buffer, sterilized 100% glycerol; 1% agarose gel, nuclease-free water, RNase*Zap*® (Life Technologies).
- *Disposables*: nuclease-free 10-, 20-, 200-, and 1000-μL filter tips, nuclease-free PCR tubes, nuclease-free microcentrifuge tubes, gloves.

4.2. Plasmid amplification and stock preparation

4.2.1 Resuspension of plasmid DNA

If beginning with lyophilized plasmid DNA, spin briefly to ensure the contents are at the bottom of the tube. Resuspend the plasmid DNA in a volume of TE buffer to produce a stock concentration of 0.1 μg/μL. To prepare a working solution, add 1 μL of the stock solution to 99 μL of nuclease-free water to produce a final concentration of 1 ng/μL. The resuspended plasmid DNA can be stored at −20 °C.

4.2.2 Chemical transformation of plasmid into Top10 E. coli cells

Prior to beginning the transformation, ensure that all required media are prepared and sterilized. Place frozen competent cells and a prelabeled tube on ice. Prewarm a hot water bath to 42 °C. To a tube on ice, add 2 μL of (∼2 ng) the plasmid working solution to 100 μL of thawed competent cells and flick the tube gently to mix. Incubate the mixture for 30 min on ice, then heat shock in the 42 °C hot water bath for 45 s. Immediately place the tube on ice for 2 min and then add 500 μL of SOC or LB liquid medium to the tube and incubate at 37 °C for 1 h with shaking (∼225 rpm). During this time prewarm LB-Amp agar plates in a 37 °C incubator. From the tube, pipet and spread 10, 100, and 200 μL on three separate LB-Amp agar plates. Place the plates upside down in a 37 °C incubator for 12–24 h. Following incubation, inoculate a single, well-isolated colony from one of the plates and place into 5-mL LB-Amp media. Grow the liquid culture at 37 °C for ∼8 h with vigorous shaking (∼300 rpm). Remove 850 μL of the starter culture and place into a 2-mL freezer vial with 150 μL of sterilized 100% glycerol, mix thoroughly, and store at −80 °C. To work from the frozen stocks, place a loopful of stock into 10 mL of LB-Amp liquid medium and grow at 37 °C with vigorous shaking (∼300 rpm) for 12–16 h. Harvest cells by centrifugation at 6000 × *g* for 15 min at 4 °C. Discard the supernatant and recover the plasmid DNA using a commercially available plasmid mini-prep kit.

4.3. Plasmid linearization and *in vitro* transcription

4.3.1 Linearization of plasmid template

Digest 2 μg of plasmid with restriction enzyme targeting the site at the end of the template sequence according to the restriction enzyme protocol. Sticky ends created by nonblunt-end cutting enzymes should be removed using mung bean nuclease. After digestion and end repair, bring the reaction to 100 μL by adding TE buffer, add 100 μL of phenol:chloroform:isoamyl alcohol (25:24:1, pH ~7), and mix by vortexing. Spin the mixture for 5 min at 12,000 × *g* in a microcentrifuge. Transfer the aqueous phase to a new tube and add 0.1 volumes (~10 μL) of 3 *M* sodium acetate and 0.7 volumes (~70 μL) of isopropanol to the tube. Mix thoroughly and incubate for 10 min at room temperature, and centrifuge for 30 min at 12,000 × *g* at 4 °C. Discard supernatant and wash pellet with 200 μL of ice cold 70% ethanol. Centrifuge for 5 min and discard the supernatant, being careful not to disturb the pellet. Air-dry the pellet to remove residual ethanol before resuspending the pellet in 5 μL of nuclease-free water. Transfer 2 μL of linearized plasmid into a new tube and add 2 μL of nuclease-free water. Use 1 μL of the diluted sample to check the concentration and analyze the remaining 3 μL on a 1% agarose gel to check for complete digestion and the presence of a single-sized product. Retain the 3 μL of undiluted DNA template for subsequent steps.

4.3.2 Synthesis and purification of mRNA internal standard

Synthesis of the internal standards from a template containing a T7 promoter is completed through the use of an *in vitro* transcription reaction using the Ambion MEGAscript® High Yield T7 Kit. In a 0.2-mL tube at room temperature, combine 2 μL of ATP solution, 2 μL of CTP solution, 2 μL of GTP solution, 2 μL of UTP solution, 2 μL of 10× reaction buffer, 1 μg of linearized template DNA (up to 8 μl), and 2 μL of enzyme mix, and bring the total reaction volume to 20 μL with nuclease-free water. Mix thoroughly by flicking and incubate the mixture at 37 °C in a thermocycler with a heated lid for 16 h. Degrade the plasmid DNA by adding 1 μL of Turbo DNAse to the reaction tube and incubating for 15 min at 37 °C. Add 20 μL of citrate-saturated (pH 4.7) phenol:chloroform:isoamyl alcohol (25:24:1) to the tube. Vortex the mixture for 1 min and centrifuge for 2 min at 12,000 × *g* to separate the phases. Transfer the upper aqueous phase to a fresh tube and add 1 volume of chloroform:isoamyl alcohol (24:1). Vortex the mixture for 1 min and centrifuge for 2 min at 12,000 × *g*. Transfer the upper aqueous phase to a fresh tube and add 0.1 volumes of 3 *M* sodium acetate and 0.7 volumes of isopropanol. Mix by vortexing and incubate for 10 min at room temperature and

then centrifuge for 30 min at 4 °C. Carefully discard the supernatant and wash the pellet with 200 μL of ice cold 70% ethanol. Centrifuge for 5 min and carefully remove the supernatant without disturbing the pellet. Air-dry the pellet until no residual ethanol remains and resuspend the dried pellet in 50 μL of nuclease-free water. Quantify the RNA fluorometrically using a Quant-iT™ RiboGreen® RNA Assay Kit and check the transcript size using a microfluidic electrophoresis instrument (e.g., Experion Automated Electrophoresis System, Agilent 2100 Bioanalyzer, or Agilent 2200 TapeStation). Store the mRNA internal standard stock at −80 °C.

5. DNA STANDARD PREPARATION

5.1. Required materials

- *Equipment*: Refrigerator, small tube rocker, 65 °C water bath or oven, fluorometry-based instrument for measuring nucleic acid concentration, 10-, 20-, 200-, and 1000-μL pipettes.
- *Materials*: Genomic DNA from a cultured, sequenced microbe unlikely to be closely related to microbes in the natural community, for example, *T. thermophilus* DSM7039 [HB27] genomic DNA (American Type Culture Collection (ATCC), Manassas, VA).
- *Commercially available kit*: Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies).
- *Disposables*: Sterile 10-, 20-, 200-, and 1000-μL filter tips, nuclease-free microcentrifuge tubes, gloves.

5.2. Genomic standard stock preparation

- Resuspend the genomic DNA in a volume of nuclease-free water to produce a stock concentration of 0.1 μg/μL following procedures recommended by ATCC. After rehydration incubate overnight at 4 °C while rocking, and then incubate for 1 h at 65 °C. To prepare a working solution, add 1 μL of the stock solution to 99 μL of nuclease-free water to produce a final concentration of 1 ng/μL. Check the DNA concentration of stocks fluorometrically using Quant-iT™ PicoGreen® dsDNA Assay Kit. The genomic DNA can be stored at −20 °C.

6. INTERNAL STANDARD ADDITION

Internal standards should be incorporated into the sample in a known amount just prior to RNA/DNA extraction. Prepare a tube with the desired lysis solution and add a known number of internal standard copies/genomes

to the prepared lysis tube prior to the addition of the sample. The goal is to add an amount of internal standard sufficient for effective quantification in the sequence dataset, but not so high as to dominate the reads. This amount can be estimated from expected recovery of nucleic acids based on previous experience with the sample type. For example, if 5 μg of total RNA is expected from an extraction, the addition of 25 ng of internal standard (0.5% of the total RNA pool by weight) should be sufficient for a standard ~1000 nt in length. In our experience, a targeted ~0.5% addition has resulted in standards accounting for 0.1–5% of reads, depending on accuracy of our predicted RNA yield. When working with multiple standards, each standard should be added to the lysis tube independently in order to control for pipetting error.

7. INTERNAL STANDARD RECOVERIES AND QUANTIFICATION

Following sequencing, the number of mRNA internal standards recovered should be quantified by a BLASTn homology search for the template sequence using a bit score cutoff of 50, equivalent to an average percent identity of 98% in our analyses. The number of genomic internal standards should be quantified by first using a BLASTn homology search against the reference genome sequence to identify all potential standard reads, and subsequently taking any hits from the initial BLASTn homology search and performing a BLASTx search against the RefSeq Protein database to identify all protein encoding reads derived from the reference genome with a bit score cutoff of 40. The second annotation step against the RefSeq Protein database is necessary for identification of the standard reads due to a high number of false positives recruited by the BLASTn homology search. Following quantification, the internal standards should be removed from the dataset before further processing.

8. DATASET NORMALIZATION USING INTERNAL STANDARDS

8.1. Metatranscriptome normalization

Following identification of internal transcript standards, total transcript pool size and individual transcript abundances can be calculated as follows:

$$P_a = \frac{P_s \times S_a}{S_s}, \quad T_a = \frac{T_s \times P_a}{P_s}$$

P_a = total transcripts in the sample
P_s = protein encoding reads in the transcriptome library
S_a = molecules of internal standard added to the sample
S_s = internal standard reads in the sequence library
T_a = total molecules of any particular transcript type in the sample. This value can be divided by the mass or volume of sample collected to calculate the transcript abundance per volume or weight
T_s = number of transcripts of interest in the sequence library

8.2. Metagenome normalization

Following identification of internal genome standards, community gene pool size and individual gene abundances can be calculated as follows:

$$S_r = \frac{S_S}{S_P}$$

$$P_g = \frac{P_s \times S_a}{S_r}, \quad G_a = \frac{G_s \times P_g}{P_s}$$

S_r = no. of molecules of internal standard genome recovered from sequencing
S_S = no. of protein encoding internal standard reads in the sequence library
S_P = no. of protein encoding genes in the internal standard reference genome
P_g = total no. of protein encoding genes in the sample
P_s = no. of protein encoding sequences in the metagenome library
S_a = no. of molecules of internal standard genome added to the sample
G_a = no. of molecules of any particular gene category in the sample. This can then be divided by the mass or volume of sample collected to calculate the transcript abundance per volume or weight

ACKNOWLEDGMENTS

This research was funded by grants from the Gordon and Betty Moore Foundation (Marine Microbiology Investigator and River-Ocean Continuum of the Amazon) and the National Science Foundation (MCB-0702125).

REFERENCES

Campbell, B. J., Yu, L., Heidelberg, J. F., & Kirchman, D. L. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(31), 12776–12781.

Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M., & Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (nifH) expression in the oligotrophic North Pacific Ocean. *Applied and Environmental Microbiology*, *71*(9), 5362–5370.

Damon, C., Vallon, L., Zimmermann, S., Haider, M. Z., Galeote, V., Dequin, S., et al. (2011). A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes. *ISME Journal*, *5*(12), 1871–1880.

Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., et al. (2008). Functional metagenomic profiling of nine biomes. *Nature*, *452*(7187), 629–632.

Gifford, S. M., Sharma, S., Rinta-Kanto, J. M., & Moran, M. A. (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME Journal*, *5*(3), 461–472.

Hannah, M. A., Redestig, H., Leisse, A., & Willmitzer, L. (2008). Global mRNA changes in microarray experiments. *Nature Biotechnology*, *26*(7), 741–742.

Hewson, I., Poretsky, R. S., Beinart, R. A., White, A. E., Shi, T., Bench, S. R., et al. (2009). In situ transcriptomic analysis of the globally important keystone N2-fixing taxon Crocosphaera watsonii. *ISME Journal*, *3*(5), 618–631.

Maurice, C. F., Haiser, H. J., & Turnbaugh, P. J. (2013). Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell*, *152*(1–2), 39–50.

Moran, M. A., Satinsky, B., Gifford, S. M., Luo, H., Rivers, A., Chan, L. K., et al. (2013). Sizing up metatranscriptomics. *ISME Journal*, *7*(2), 237–243.

Ottesen, E. A., Young, C. R., Eppley, J. M., Ryan, J. P., Chavez, F. P., Scholin, C. A., et al. (2013). Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(6), E488–E497.

Poretsky, R. S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., et al. (2005). Analysis of microbial gene transcripts in environmental samples. *Applied and Environmental Microbiology*, *71*(7), 4121–4126.

Rivers, A., Sharma, S., Tringe, S. G., Martin, J., Joye, S. B., & Moran, M. A. (2013). Transcriptional response of bathypelagic marine bacterioplankton to the deepwater horizon oil spill. *ISME Journal*, http://dx.doi.org/10.1038/ismej.2013.129.

van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D., & Holstege, F. C. (2003). Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO Reports*, *4*(4), 387–393.

Vila-Costa, M., Sharma, S., Moran, M. A., & Casamayor, E. O. (2013). Diel gene expression profiles of a phosphorus limited mountain lake using metatranscriptomics. *Environmental Microbiology*, *15*(4), 1190–1203.

Xu, L., Chen, H., Hu, X., Zhang, R., Zhang, Z., & Luo, Z. W. (2006). Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Molecular Biology and Evolution*, *23*, 1107–1108.