Internal Genomic DNA Standard for Quantitative Metagenome Analysis

Brandon M. Satinsky, Scott M. Gifford, Byron C. Crump, Christa Smith, Mary Ann Moran, Moran Lab

Abstract

Satinsky, Brandon M., et al. 'Use of internal standards for quantitative metatranscriptome and metagenome analysis.' *Methods in enzymology* 531 (2012): 237-250.

Citation: Brandon M. Satinsky, Scott M. Gifford, Byron C. Crump, Christa Smith, Mary Ann Moran, Moran Lab Internal Genomic DNA Standard for Quantitative Metagenome Analysis. **protocols.io**

dx.doi.org/10.17504/protocols.io.fgbbjsn

Published: 08 Sep 2016

Guidelines

DNA standards can be prepared by purchasing or extracting genomic 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.

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 standard from a cultured, sequenced microbe unrelated to microbes anticipated to be present in the system of interest. *Thermus thermophilus* DSM7039 [HB27] genomic DNA was purchased from the American Type Culture Collection (ATCC) for Satinsky et al. 2012.

•*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.

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.

Internal standard addition:

A known number of internal standard copies/genomes should be added to a sample just prior to DNA extraction. Depending on the sequencing depth, add genomic standard to reach ~0.1-1.0% of the total reads (less is needed as sequencing depth increases).

Internal standard recoveries and quantification:

Following sequencing, the number of genomic internal standards should be quantified, first by BLASTn homology search against the internal standard genome to identify all potential standard reads (bit score cutoff of 50), and second by BLASTx against the RefSeq Protein database to identify all protein encoding reads derived from the internal standard genome (bit score cutoff of 40). Following quantification internal standards should be removed from the dataset.

Dataset normalization using internal standards

Community gene pool size and individual gene abundances can be calculated as follows:

$$S_r = S_S / S_P$$

$$P_g = (P_s * S_a)/S_r;\ G_a = (G_s * P_g)/P_s$$

S_r = number of molecules of internal standard genome recovered from sequencing

S_s = number of protein encoding internal standard reads in the sequence library

S_P = number of protein encoding genes in the internal standard reference genome

$\mathbf{P_g}$ = total number of protein encoding genes in the sample

$\mathbf{P_s}$ = number of protein encoding sequences in the metagenome library

$\mathbf{S_a}$ = number of molecules of internal standard genome added to the sample

$\mathbf{G_a}$ = number 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

$\mathbf{G_s}$ = number of genes of interest in the sequence library

Protocol

Genomic standard stock preparation

Step 1.

Suspend genomic DNA in a volume of nuclease-free water to produce a stock concentration of 0.1 µg/µL.

Genomic standard stock preparation

Step 2.

Incubate rehydrated DNA overnight at 4°C while rocking, and then incubate for 1 h at 65°C.

Genomic standard stock preparation

Step 3.

Prepare working solution by adding 1 µL of the stock solution to 99 µL of nuclease-free water to produce a final concentration of 1 ng/µL.

Genomic standard stock preparation

Step 4.

Check the DNA concentration of stocks fluorometrically using Quant-iT™ PicoGreen® dsDNA Assay Kit

Genomic standard stock preparation

Step 5.

Genomic DNA can be stored at − 20 °C.

Internal standard addition

Step 6.

Just prior to sample DNA extraction, add enough standard DNA to each sample to reach 0.1-1.0% of

expected total reads.

Internal standard recoveries and quantification

Step 7.

Following sequencing, quantify the number of genomic standard reads (steps 8 and 9).

Internal standard recoveries and quantification

Step 8.

Using a bit score cutoff of 50, identify standard reads by BLASTn homology search aganst the internal standard genome.

Internal standard recoveries and quantification

Step 9.

Using the results from step 8 and a bit score cutoff of 40, perform a BLASTx against the RefSeq Protein database to identify all protein encoding reads derived from the internal standard genome.

Internal standard recoveries and quantification

Step 10.

Quantify recovered standard DNA reads and remove from dataset.

Dataset normlization using internal standards

Step 11.

Calculate the number of molecules of internal standard recovered from sequencing:

- $S_r=S_S/S_P$

$\mathbf{S_r}$ = number of molecules of internal standard genome recovered from sequencing

$\mathbf{S_s}$ = number of protein encoding internal standard reads in the sequence library

$\mathbf{S_P}$ = number of protein encoding genes in the internal standard reference genome

Dataset normlization using internal standards

Step 12.

Calculate the community gene pool size:

- $P_g=(P_s*S_a)/S_r$

$\mathbf{P_g}$ = total number of protein encoding genes in the sample

$\mathbf{P_s}$ = number of protein encoding sequences in the metagenome library

$\mathbf{S_a}$ = number of molecules of internal standard genome added to the sample

$\mathbf{S_r}$ = number of molecules of internal standard genome recovered from sequencing

Dataset normlization using internal standards

Step 13.

Calculate individual gene abundances for genes of interest:

- $G_a=(G_s*P_g)/P_s$

$\mathbf{G_a}$ = number 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

$\mathbf{G_s}$ = number of genes of interest in the sequence library

$\mathbf{P_g}$ = total number of protein encoding genes in the sample

$\mathbf{P_s}$ = number of protein encoding sequences in the metagenome library