Cheatography

Validation Metagenomics Workflows Cheat Sheet

by [deleted] via cheatography.com/2754/cs/18375/

Introduction

Shotgun metagenomics is a powerful platform to characterize human microbiomes. However, to translate such survey data into consumer--relevant products or services, it is critical to have a robust metage-nom-ics workfl ow. We present a tool – spike-in DNA – to assess performance of metagenomics workfl ows. The spike-in is DNA from two organisms – Alivibrio fi scheri and Rhodopseudomonas palustris , in a ratio of 4:1 added to samples before DNA extraction. With a valid workfl ow, the output ratio of relative abundances of these organisms should be close to 4. This expectation was tested in samples of varying diversities (n = 110), and the mean ratio was 4.73 (99% CI [4.0, 5.24]). We anticipate this tool to be a relevant community resource for assessing the quality of shotgun metagenomics workfl ows and thereby enable robust characterization of microbiomes..

Source: https://www.future-science.com/doi/pdf/10.2144/btn-2018-0089

Stage 1

1. . Break workflow into discrete modules, e.g., DNAextraction and library preparation.
2. Add spike-in genomic DNA to the sample of interest at the first step of the module. F or instance, spike-inbefore tagmentation if library preparation is the module.
3. Module with the maximum deviation from expected ratio is identified and iterated upon for improvement.
4. Put modules together and use spike-in to validate the entire workflow (Stage 2).

Stage 2

1. Assess variance in the spike-in ratio using the experimental design outlined (Figure 1).
2. Spike-in ratio should be closest to the expected value when the spike-in genomic DNA is added to sterile saline and processed through the workflow.
3. Test the spike-in performance in samples of varying complexities. Ensure that these samples include microbiomes of interest. Defining the acceptable variance is left to the operator's discretion. Based upon all the samples described here, we defined the acceptable range to be between 4 and 5.4(99% confidence intervals).

Figure 1

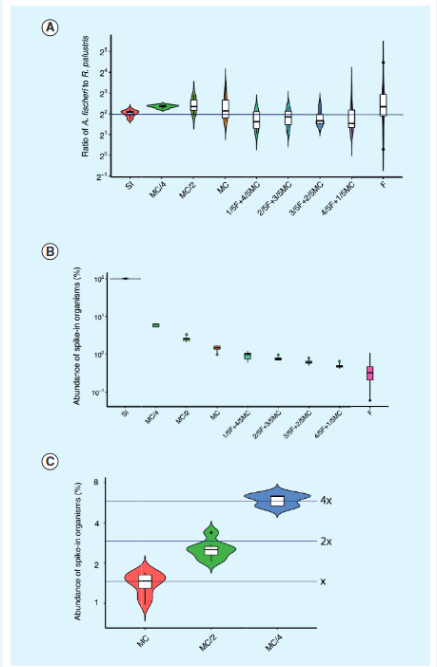

Figure 1. Spike-in genomic DNA can be used to validate metagenomics workflows. (A) Overall, the output ratio of the spike-in matches with the input ratio in samples of varying diversities. The solid line is the expected ratio - 4. **(B and C)** Spike-in reads are inversely and linearly related to the microbial load of the samples. Data shown in **(C)** is a subset of the data in **(B)**. The expectation regarding the linear relationship of spike-in abundances with microbial load is better tested in dilutions of the defined mock community. The horizontal intercepts denote the median relative abundances. **(A and C)** Colored areas represent violin plots and whiskerplots represent boxplots. **(B)** Colored areas represent whiskerplots. The figures are all generated using the R package ggplot2 as part of ggpubR.
F: Fecal material; MC: Undiluted mock community; MC/4: Mock community diluted fourfold; MC/2: Mock community diluted twofold; SI: Spike-in added to sterile saline.

Stage 3

Spike-in can be used for per run QC as follows:

1. Include triplicates of just the spike-in added to sterile saline as positive control.
2. A pooled sample can be created by mixing the samples of interest. The spike-in genomic DNA can be added to this pool in triplicate and processed through the workflow. Spike-in performance is calculated as outlined in Stage 2.