MGTST-Outline

Nate Olson

2016-10-26

Objectives

- Provide a detailed description of the dataset and qa/qc methods used to validate assumptions regarding the sample composition
- Demonstrate how the dataset is used to evaluate pipeline performance and differential abundance methods

- Provide an R package to facilitate using the dataset for evaluating normalization and differential abundance detection methods

Abstract

Background

Methods

Experimental design

- two sample titration
- sample selection
- PCR, library prep, multiple laboratories

Sequencing

- PCR region
- barcode
- sequencing
- library prep qc

Mixture QC

- 16S qPCR
- ERCC qPCR

Sequence processing

- Pipelines
  - Mothur
  - QIIME
  - DADA2
  - POP

- Pipeline summaries
  - dataset characterizations

Data analysis

- Normalization Methods
- Differential Abundance Methods

Evalution Metrics

- Normalization bais and variance
  - Variance - negative binomial
  - Bias - difference expected value based on unmixed samples and observed value
- Differential abundance bias and variance
  - Variance - estimated by differential abundance methods
  - Bias - using pre and post specific OTUs

Results

Seq results summary table

- Seq dataset characterization - number of reads per sample ect.
- Seq Pipeline summary - otus, alpha and beta diversity of unmixed samples
  - provides a general characterization of unmixed samples
- Figure - MA plot color Pre and post specific OTUs, grey other OTUs, facet by pipeline
  - Pre and Post unique OTUs

Titration Validation

- Use of ERCC spike-ins to validate the pre and post samples were mixed as expected
- Bacterial abundance qPCR used to validate that bacterial DNA concentration is equivalent between unmixed pre and post samples

Sequence processing summary

- Table summarizing - Total and per sample OTUs and sequences
- Figure or Table? - Summary of Pre vs. Post specific OTUs - used in unmixed sample abundance free logFC ratio estimates
  * abundance and taxonomy
  * only OTUs not present in all four PCR replicates

Normalization

- section objective - used PCR replicate variance values to validate normalization methods
- Technical replicate variance distributions for different pipelines and normalization methods
- Bias for different pipelines and normalization methods

Differential Abundance

- section objective - demonstrate how the dataset can be used to evaluate the limit of differential abundance detection
- Differential abundance dectection between unmixed and tritrated samples
  - Pre and post specific OTUs
  - All OTUs

Discussion

- Validation of two sample titrations using qPCR
- Differences between pipelines
  - General statements
- Normalization methods
- Differential abundance methods

Acknowladgements

References

Supplemental

Sample Selection

wetlab QC

- sample concentration results summary

Seq data QA

- number of reads
- read length distributions
- PhiX error rate analysis
- base quality summary

Pipeline

- Seq budget - summarize fate of sequences, number successful merged read pairs, chimera filtered, alignment?, ect.
  - Table - pipeline sequence budget
    * number of reads filtered due to low quality
    * number of reads merged
    * number of chimeras

Characterizing Sources of variability

Response Variance

- Experimental replicate variance - how is the variance correlated with different types of technical replicates, how does the variance differ for pipelines
  - section objective - characterize count variance between PCR replicates
    - is the variance correlated with experimental values e.g. biological sample, PCR plate, well, sequencing depth, or observed count value
  - Figure - relationship between count and PCR replicate variance

Response linearity

- section objective - demonstrate how the dataset is used to characterize relative abundance estimates and identify potential sources of bias
- Figure observed vs expected plots
- Figure representative OTUs showing different types of response linearity
- Differentiating between high and low linearity OTUs