ASM Microbe 2023 | Poster P42738

Development and validation of an internal bioinformatic pipeline for 16S rRNA amplicon sequencing

Timothy B. Doyle, Jessica Ewald, Valerie Kantro, Mariana Castanheira *
JMI Laboratories, North Liberty, IA, USA

Introduction

- The gut microbiota is not a fixed composition of microbes, but rather a dynamic ecosystem influenced by patient age, diet, and antibiotic exposure.
- There is a growing body of evidence on the impact of the gut microbiome on a huge range of diseases and disorders, from inflammatory bowel disease (IBD) to autism to depression.
- The ability to provide accurate analysis of stool samples from each patient is paramount to better comprehend the relationship between gut microbiome and human health.
- We developed and validated an end-to-end 16S rRNA bioinformatics analysis pipeline for the taxonomic classification of bacteria from stool samples.

Materials and Methods

Samples tested

- Two separate ATCC mock communities that mimic mixed microbiome samples:
  - MSA-1000 is a 10-strain even mix (each at 10% relative abundance).
  - MSA-1002 is a 20-strain even mix (each at 5% relative abundance).
- A total of 17 fresh, unpreserved stool samples were collected in sterile vials and sent to JMI Laboratories for processing. Each collected sample was required to meet the following criteria:
  - Can be formed or unformed.
  - Minimum of 1 mL or 1 g.
  - Collected fresh, stored at 4 °C and delivered to JMI Laboratories within 24 hours.
- ZymoBIOMICS microbiome spike-in control:
  - Consists of equal cell numbers of two bacterial strains, *Imtechella halotolerans* and *Allobacillus halotolerans*, which are foreign to the human microbiome.

16S rRNA amplicon sequencing and analysis

- Total genomic DNA was extracted using the QIAamp PowerFecal DNA Kit. DNA was evenly split for sequencing at JMI Laboratories and CosmosID.
- Targeted V3-V4 amplicons from each sample were generated using custom PCR amplification primers.
- DNA libraries were prepared using the Nextera XT™ library index kit and sequenced on a MiSeq sequencer.
- Samples were analyzed on both the JMI analysis pipeline and the CosmosID pipeline as part of this validation.

JMI Analysis pipeline (Figure 1)

- All FASTQ files uploaded to the JMI Laboratories 16S rRNA analysis pipeline must pass initial quality control metrics including average read quality (≥Q26), total number of reads (≥50,000) and average read length (≥150 bp).
- Passing FASTQs are trimmed and filtered based on length and quality, and all adapter sequences removed.
- Sample classification is performed based on alignment of sample reads with Kraken 2 database, which utilizes the NCBI Refseq Targeted Loci Project's collection of 16S sequences.
- Relative abundances are calculated using Bracken and normalized to allow for comparison between samples with different total read counts.

Results

- The accuracy and specificity for both mock communities was 100% between the JMI analysis pipeline and ATCC specifications.
  - All genera were correctly identified above the analytical sensitivity limit (0.5%).
  - Relative abundance for 10-strain mix ranged from 3.77%–18.78% (Figure 2).
  - Relative abundance for 20-strain mix ranged from 0.84%–10.82% (Figure 3).
- The degree of accuracy for the 17 fresh stool samples compared to the CosmosID analysis ranged from 64.0%–87.8% with p-values >0.954 (Table 1).
  - Discordant results entirely driven by lower abundance genera and taxonomy classifications.
  - The accuracy for genera with a relative abundance ≥1.5% was 100%; precision rate was 92.6%.
- Inter- and intra-assay reproducibility showed a high agreement between replicates, with p-values ≥0.993 (Data not shown).
- Limit of detection, as determined using serial dilutions of the ZymoBIOMICs spike-in control, was established at 2x10^6 CFU/extracted sample (Data not shown).

Conclusions

- A 16s rRNA bioinformatics pipeline developed by JMI Laboratories was validated using ATCC mock communities and by comparing sequencing results against the CosmosID pipeline.
- The JMI pipeline showed 100% accuracy at the genus level for ATCC mock communities and 100% accuracy of unknown stool samples when compared to the CosmosID pipeline, for genera ≥1.5%.
  - Discordant results between JMI and CosmosID can largely be attributed to very low abundance genera (<1.5% relative abundance).
- The development and validation of a 16S rRNA pipeline gives JMI Laboratories the ability to ensure high quality, reproducible metagenomic results where all steps of the analysis are understood and well documented.
- JMI Laboratories has continued to validate additional specimen types and microbiome conditions using this pipeline.

Figure 1. 16S rRNA analysis pipeline developed and validated by JMI Laboratories

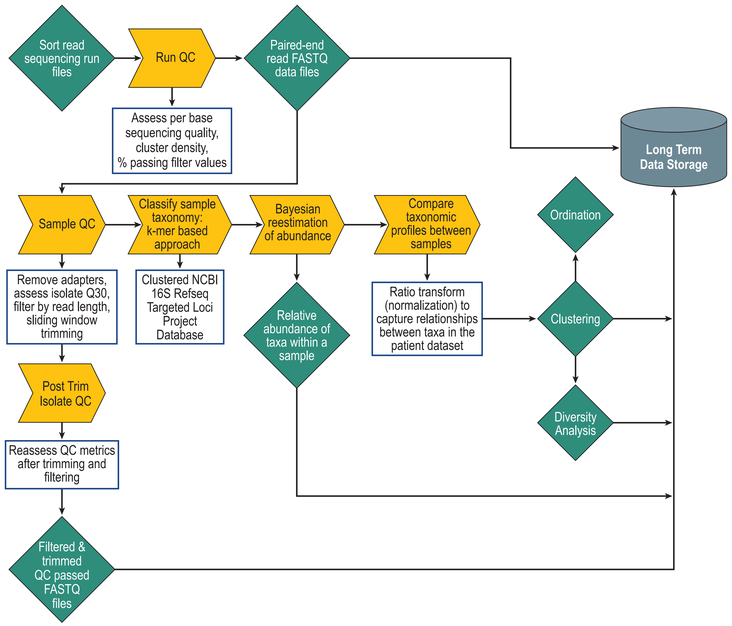


Figure 2. Sequencing results of known control ATCC MSA-1000 performed and analyzed at JMI Laboratories

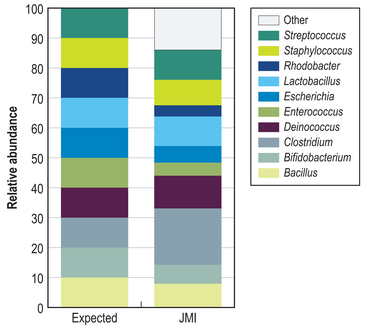
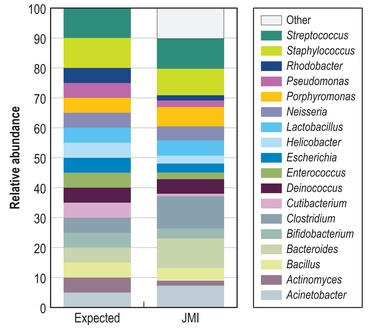


Figure 3. Sequencing results of known control ATCC MSA-1002 performed and analyzed at JMI Laboratories



Staphylococcus and *Streptococcus* genera are each at 10% in product

Table 1. The number of genera identified, accuracy and p-value analysis for fresh stool samples analyzed at CosmosID and JMI Laboratories

Sample number	Genera count	Accuracy	p-value
JMI 1	36	86.1	0.975
JMI 2	18	83.3	0.994
JMI 3	29	82.8	0.984
JMI 4	39	79.5	0.981
JMI 5	16	81.3	0.997
JMI 6	29	79.3	0.962
JMI 7	36	83.3	1.000
JMI 8	34	82.4	0.990
JMI 9	19	84.2	0.963
JMI 10	35	74.3	0.986
JMI 14	29	75.9	0.987
JMI 15	41	87.8	0.969
JMI 16	26	80.8	0.995
JMI 17	22	68.2	0.977
JMI 18	25	64.0	0.992
JMI 19	24	79.2	0.979
JMI 20	33	78.8	0.954

Genera count = number of genera with <0.5% relative abundance

References

21 CFR 58; Good Laboratory Practices for Non-Clinical Laboratory Studies

Leber, A. Clinical Microbiology Procedures Handbook. Current Edition.

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*. **41:** e1.

Srinivasan R, Karaoz U, Volegova M, MacKichan J, Kato-Maeda M, Miller S, Nadarajan R, Brodie EL, Lynch SV. 2015. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. PLoS One. 10(2):e0117617.

www.cosmosid.com

Contact

Mariana Castanheira, PhD
JMI Laboratories
345 Beaver Kreek Centre, Suite A
North Liberty, Iowa 52317
Phone: (319) 665-3370
Fax: (319) 665-3371
Email: mariana-castanheira@jmilabs.com

SCAN ME

To obtain a PDF of this poster:
Scan the QR code or visit
https://www.jmilabs.com/data/posters/ASMMicrobe2023_16sMicrobiome.pdf