

# Microbiome Premature

Jaewoong Lee

Ulsan National Institute of Science and Technology

*jwlee230@unist.ac.kr*

2020-09-07

# Overview

1 Introduction

2 Materials

3 Methods

4 Results

5 Proceedings

References

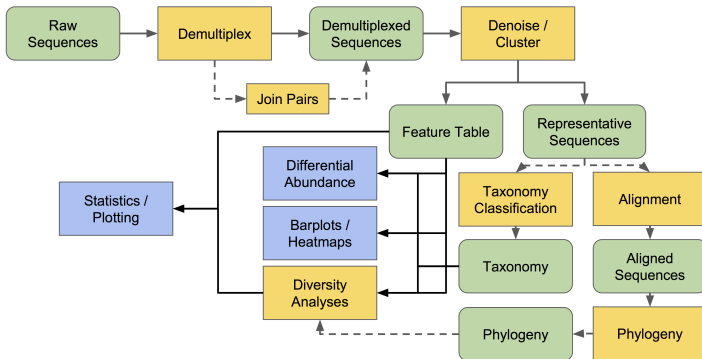


# Ribosomal RNA

# Premature

# 16s rRNA Sequencing

# Qiime 2



**Figure:** QIIME 2 workflow (Bolyen et al., 2019; Mandal et al., 2015; McDonald et al., 2012)

# Denoising Techniques

- DADA2 (Callahan et al., 2016)
- Deblur (Amir et al., 2017)



# Taxonomy Classification

- Greengenes (DeSantis et al., 2006)
- SILVA (Pruesse et al., 2007; Quast et al., 2012)

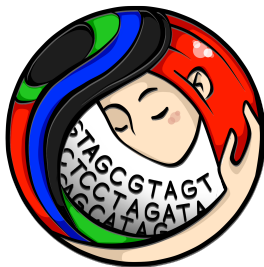


Figure: Mothur

Note: Still in progress

# t-distributed Stochastic Neighbor Embedding

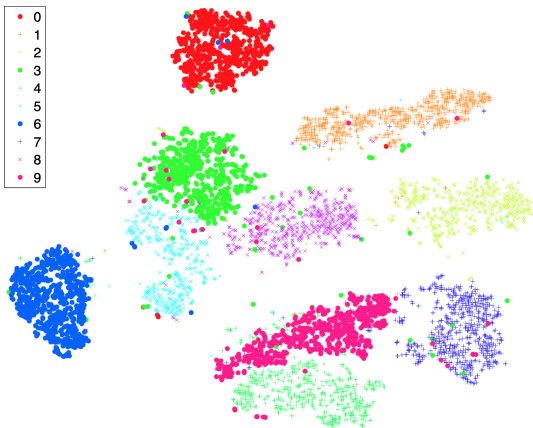
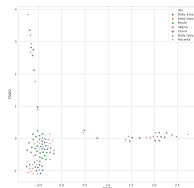
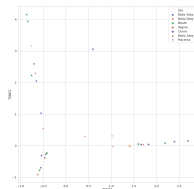
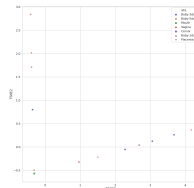


Figure: t-SNE with handwritten data (Maaten & Hinton, 2008)

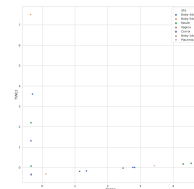
# t-SNE with Site



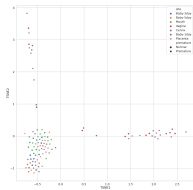
(a) DADA2 + GG (b) DADA2 + SILVA



(a) Deblur + GG (b) Deblur + SILVA

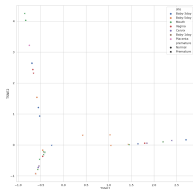


## t-SNE with Site + Premature



(a) DADA2 + GG

(b) DADA2 + SILVA



(a) Deblur + GG

(b) Deblur + SILVA

# Yields I

# Requirements I

# Expectations I



- Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Xu, Z. Z., ... others (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems*, 2(2).
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using qiime 2. *Nature Biotechnology*, 37(8), 852-857. Retrieved from <https://doi.org/10.1038/s41587-019-0209-9> doi: 10.1038/s41587-019-0209-9
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). Dada2: high-resolution sample inference from illumina amplicon data. *Nature methods*, 13(7), 581-583.

- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... Andersen, G. L. (2006). Greengenes, a chimera-checked 16s rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7), 5069–5072.
- Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov), 2579–2605.
- Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease*, 26(1), 27663. doi: 10.3402/mehd.v26.27663

# References III

- McDonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D., ... Caporaso, J. G. (2012). The biological observation matrix (biom) format or: how i learned to stop worrying and love the ome-ome. *GigaScience*, 1(1), 7. doi: 10.1186/2047-217X-1-7
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). Silva: a comprehensive online resource for quality checked and aligned ribosomal rna sequence data compatible with arb. *Nucleic acids research*, 35(21), 7188–7196.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glöckner, F. O. (2012). The silva ribosomal rna gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41(D1), D590–D596.