

Metagenome Analysis of Preterm Birth

Jaewoong Lee Semin Lee

Department of Biomedical Engineering
Ulsan National Institute of Science and Technology

jwlee230@unist.ac.kr

2022-12-27

Overview

1 Introduction

2 Materials

3 Methods

4 Results

5 Discussion

6 References

1. Introduction

Microbiome

- Microbiota: the microorganisms which live inside & on humans (Turnbaugh et al., 2007)
- Microbiome: 10^{13} to 10^{14} microorganisms whose collective genome (Gill et al., 2006)



Figure: Concept of a core human microbiome (Turnbaugh et al., 2007)

rRNA

- Ribosomal RNA
- Well-known as a key to phylogeny (Olsen & Woese, 1993)

Preterm Birth (PTB)

PTB:

- ① PTB < 37 GW (Gestational week)
- ② Normal ≥ 37 GW

Detailed PTB:

- ① Early PTB < 34 GW
- ② 34 GW \leq Late PTB < 37 GW
- ③ Normal ≥ 37 GW

(J. Tucker & McGuire, 2004; Voronkov, Solonovych, Liashenko, & Revenko, 2018)

2. Materials

16S rRNA Sequencing

16S rRNA sequencing is the *reference method* for bacterial taxonomy & identification (Mignard & Flandrois, 2006)

Three main reasons (Janda & Abbott, 2007):

- 16S rRNA exists in almost all bacteria
- Functions of the 16S rRNA has not changed over evolution.
- 16S rRNA is large enough for bioinformatics

Data Composition

- JBNU/Helixco data
 - First data
 - Second data
 - Stool data

Table: Sample Information

Data	Participants	Samples	Remarks
First	24	107	-
Second	35	288	-
Third	10	106	-
Stool	63	126	Stool

3. Methods

Qiime 2 Workflow

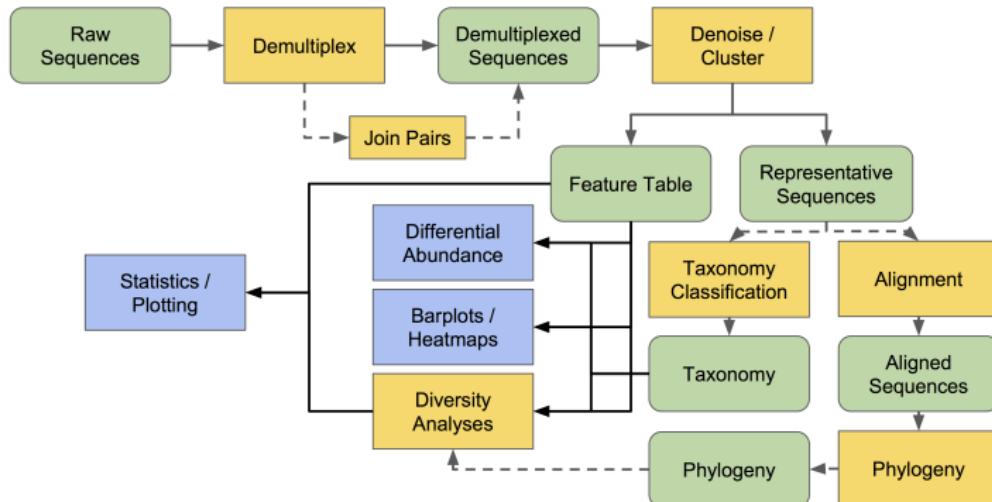


Figure: QIIME 2 workflow (Bolyen et al., 2019; Mandal et al., 2015; McDonald et al., 2012)

4. Results

4. Results

4.1. Taxonomy Overview

Proportion Distribution

Proportion

- Minimum: 0.0
- Mean: 0.00008
- Median: 0.0
- Maximum: 0.793

Proportion without Zero

- Minimum: 0.00002
- Mean: 0.00008
- Median: 0.00153
- Maximum: 0.793

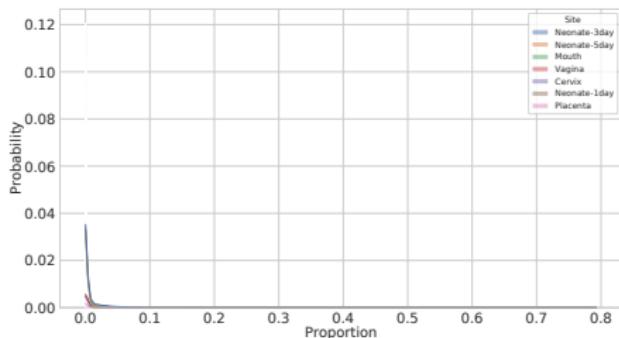


Figure: Proportion distribution

Microbial community with Proportion

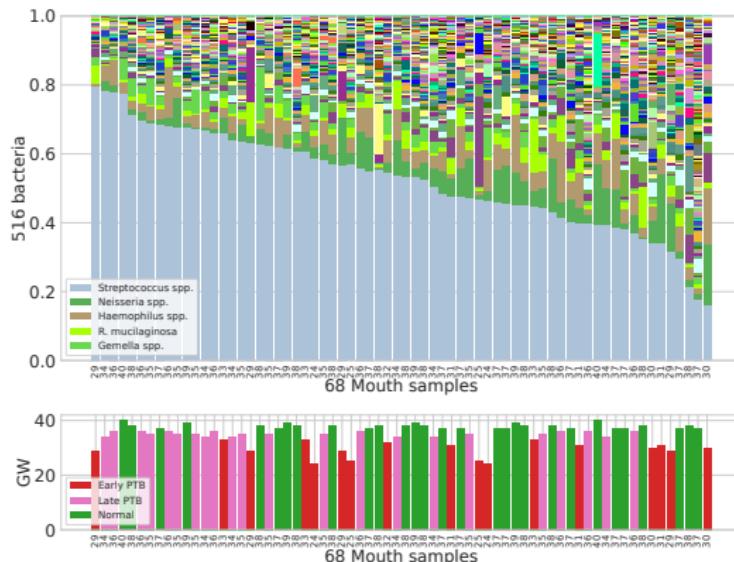
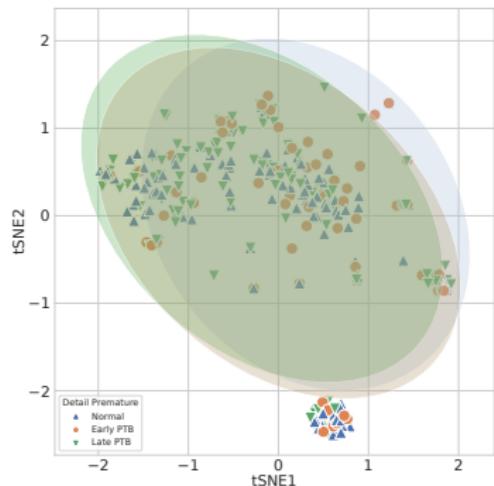
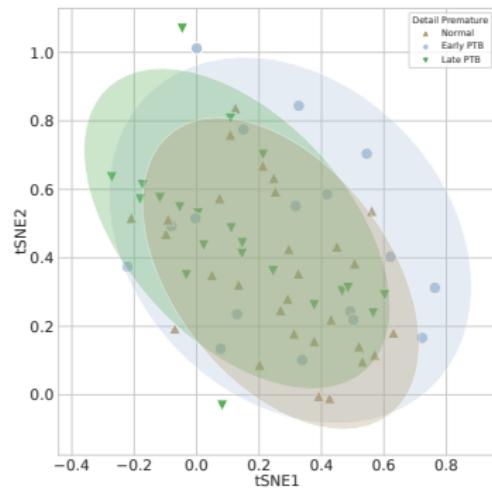


Figure: Microbial community with Proportion

t-SNE with Proportion



(a) All



(b) Mother Mouth

Figure: t-SNE plot of Proportion with PTB

Notable Taxa I

Streptococcus spp.

- *S. mutans*: pathogen of dental caries
- Membrane vesicles of Group B *Streptococcus* disrupt feto-maternal barrier (Surve et al., 2016).
∴ Leading to PTB

Neisseria spp.

- *N.* colonize the mucosal surfaces of many animals.
- Only two pathogens in 11 known species
 - *N. meningitidis*: Meningitis & Sepsis
 - *N. gonorrhoeae*: Gonorrhea (sexual transmitted disease)
- *N. gonorrhoeae* results adverse pregnancy outcomes (Heumann, Quilter, Eastment, Heffron, & Hawes, 2017).

Notable Taxa II

Haemophilus spp.

- *H.* inhabit the mucous membranes of upper respiratory tract, mouth, vagina, & intestinal tract.
- *H. influenzae*: Influenza
- PTB caused by *H. influenzae* (Hills et al., 2022) and *H. parainfluenzae* (Mendz, Petersen, Quinlivan, & Kaakoush, 2014).

R. mucilaginosa

- *Rhodotorula mucilaginosa*
- *R.* is a genus of pigmented yeasts.
- Catheter-related bloodstream infection due to *R. mucilaginosa* (Kitazawa, Ishigaki, Seo, Yoshino, & Ota, 2018).
- ∴ *Rhodotorula* bloodstream infections

Gemella spp.

- *G. bacteria* are primarily found in mucous membranes of human e.g. oral cavity & upper digestive tract.
- *G. haemolysans* causes endocarditis (Mosquera, Zabalza, Laniero, & Blanco, 2000).
- Women who have PTB also have significant lower level of *G.* (Li et al., 2021).

4. Results

4.2. Diversity Index

Diversity Index

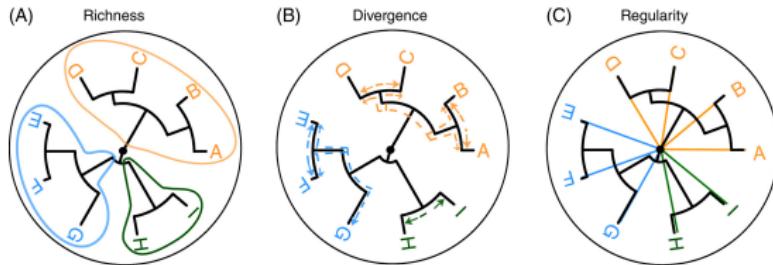


Figure: Three dimensions of phylogenetic information (C. M. Tucker et al., 2017)

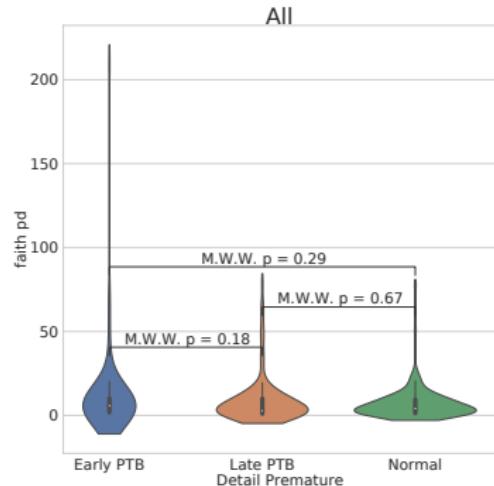
- A quantitative measure that shows richness, divergence, and regularity (C. M. Tucker et al., 2017)
- Alpha diversity: the richness of taxa **at a single community**
- Beta diversity: taxonomy differentiation **between communities**

4. Results

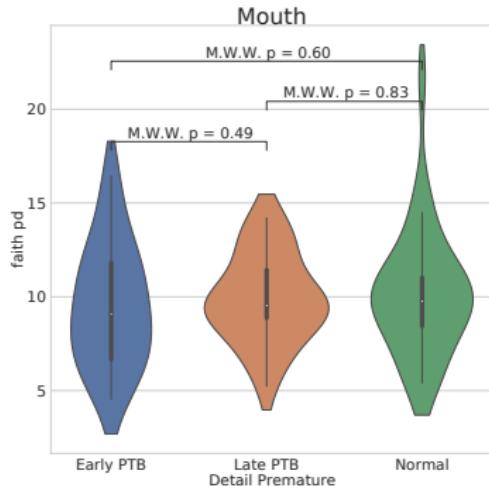
4.2. Diversity Index

4.2.1. Alpha-diversity

Violin Plot with Alpha-diversity I



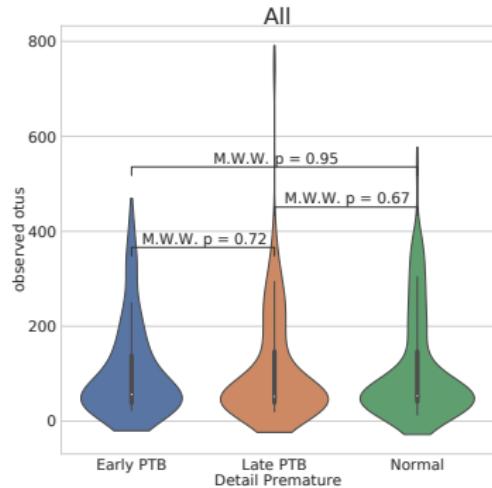
(a) All



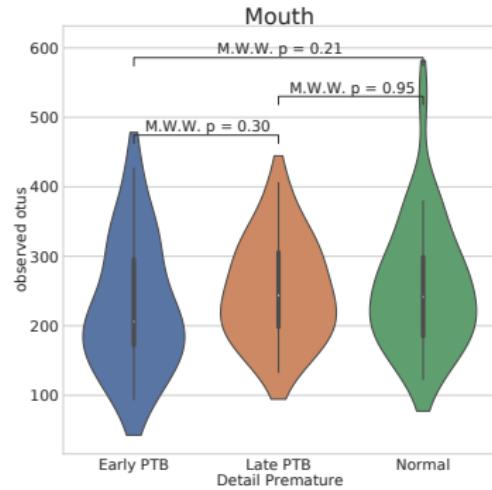
(b) Mother Mouth

Figure: Detail premature & Faith's PD

Violin Plot with Alpha-diversity II



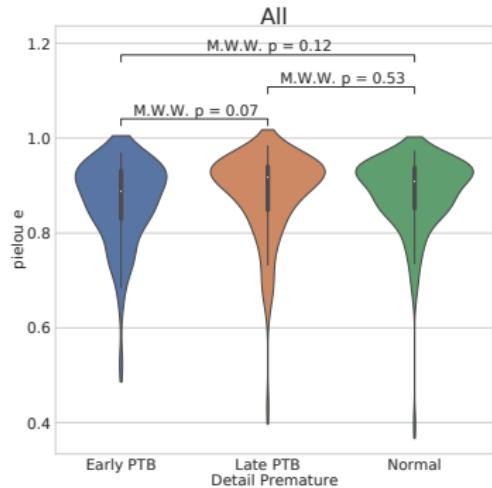
(a) All



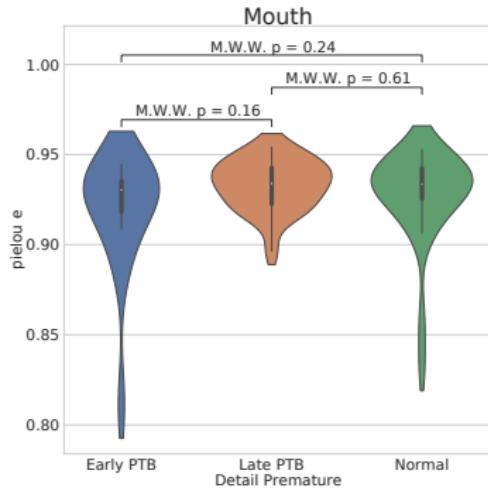
(b) Mother Mouth

Figure: Detail premature & Observed OTUs

Violin Plot with Alpha-diversity III



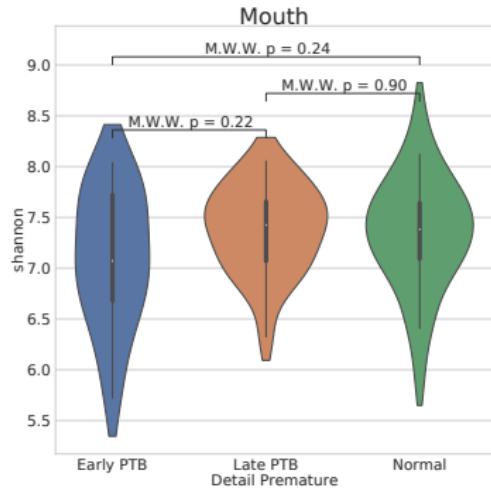
(a) All



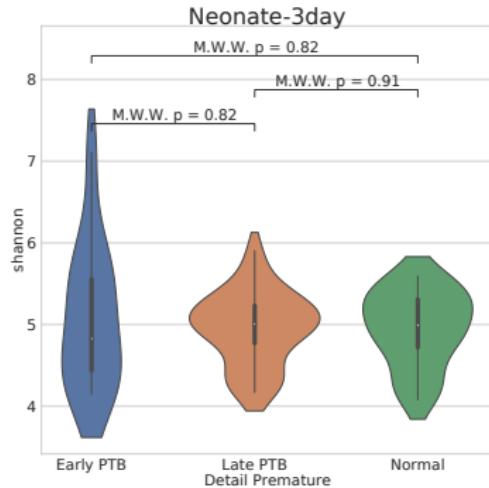
(b) Mother Mouth

Figure: Detail premature & Pielou Evenness

Violin Plot with Alpha-diversity IV



(a) All



(b) Mother Mouth

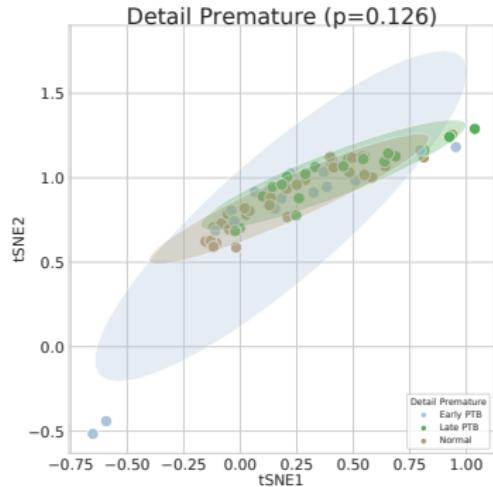
Figure: Detail premature & Shannon Entropy

4. Results

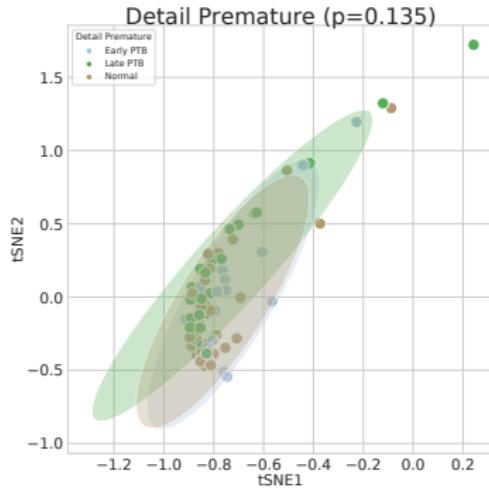
4.2. Diversity Index

4.2.2. Beta-diversity

Beta-diversity t-SNE plots I



(a) Bray-Curtis



(b) Euclidean

Figure: Beta-diversity t-SNE plots

4. Results

4.3. Taxonomy Analyses

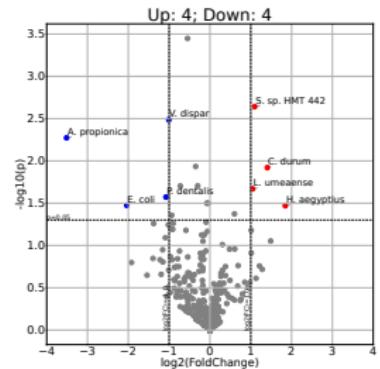
4. Results

4.3. Taxonomy Analyses

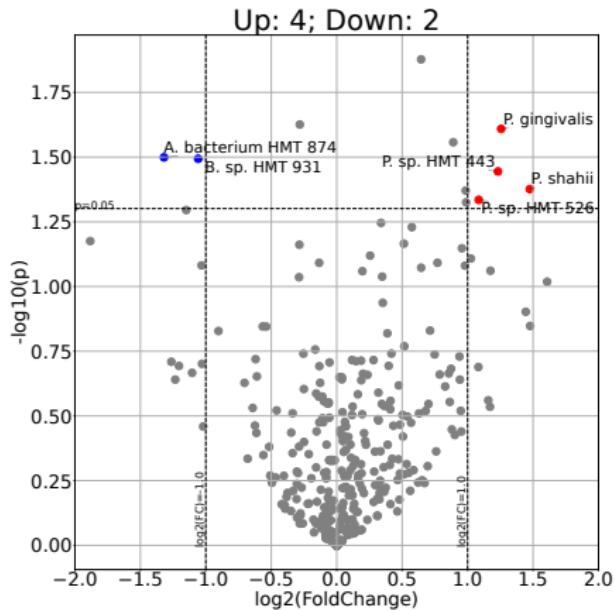
4.3.1. Differentially Abundant Taxa

Volcano plots I

Volcano plots II



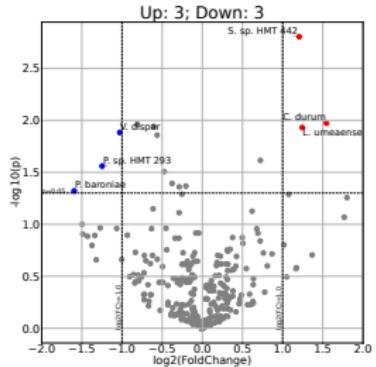
(a) Early PTB vs. other



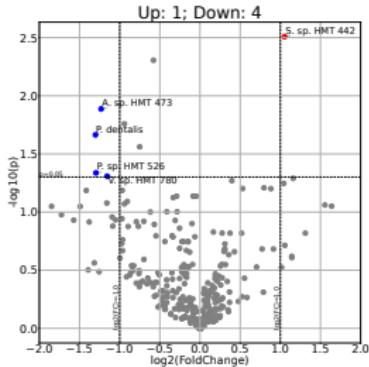
(b) PTB vs. Normal

Figure: DAT in Mouth

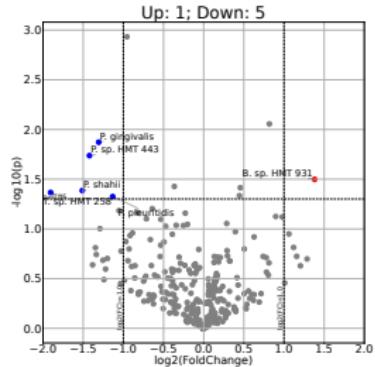
Volcano plots III



(a) Early vs. Late



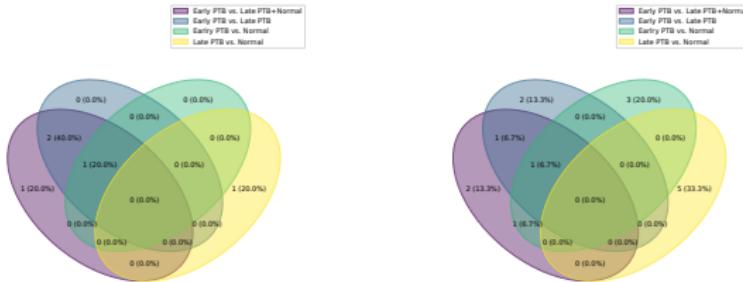
(b) Early vs. Normal



(c) Late vs. Normal

Figure: DAT in Mouth

Venn Diagrams



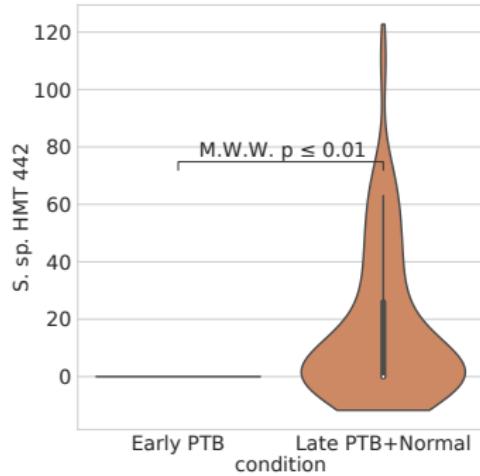
(a) Up-regulated DAT (b) Down-regulated DAT

Figure: DAT Venn diagram

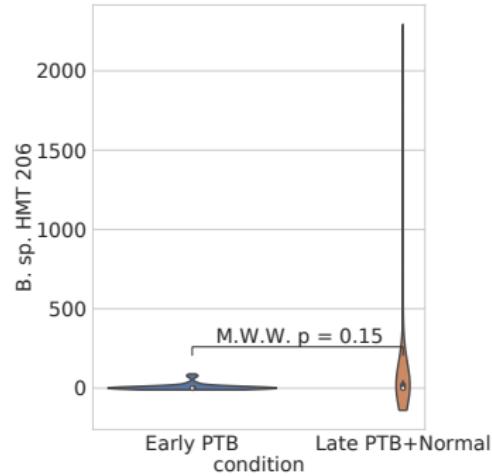
Union of DAT

A total of 19 taxa selected as DAT.

Violin plots I



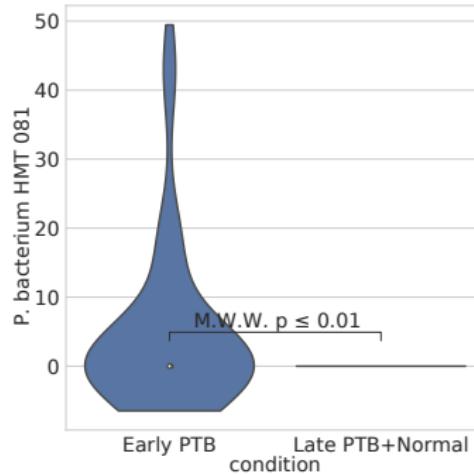
(a) *Selenomonas* sp. HMT 442



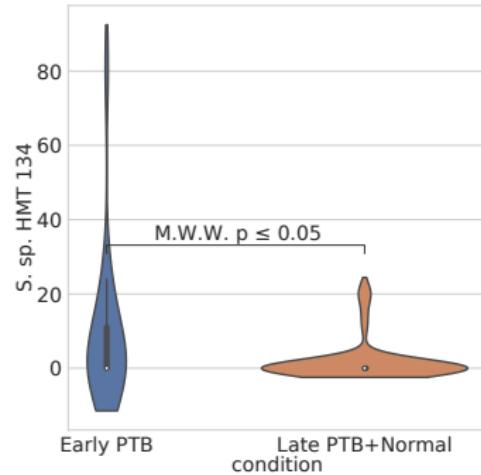
(b) *Bergeyella* sp. HMT 206

Figure: Up-regulated taxa

Violin plots II



(a) *P. bacterium* HMT 081



(b) *Selenomonas* sp. HMT 134

Figure: Down-regulated taxa

4. Results

4.4. Machine Learning

ML algorithm comparison

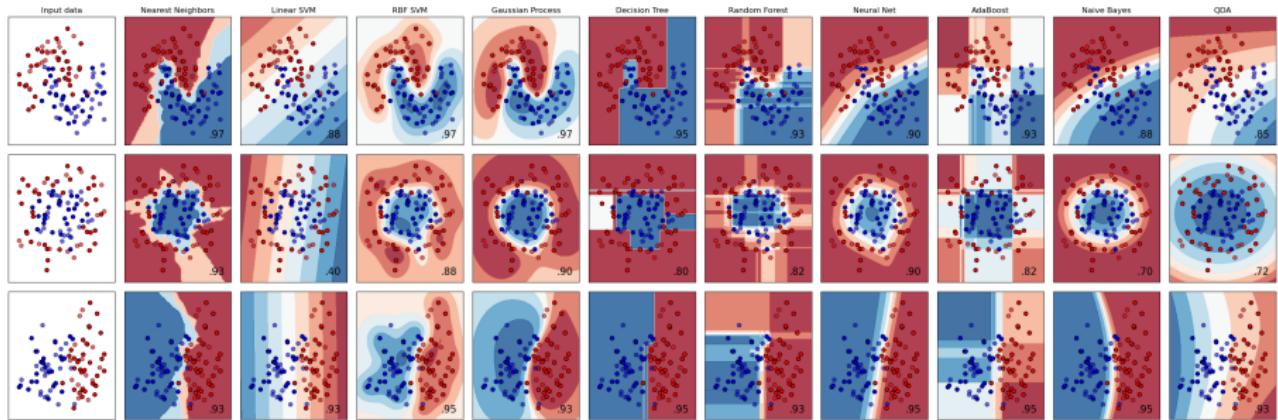


Figure: Classification Comparison (Pedregosa et al., 2011)

Oversampling

SMOTE (Chawla, Bowyer, Hall, & Kegelmeyer, 2002)

- Synthetic Minority Oversampling Technique
- An algorithm that makes pseudo-sample
- Using K -Nearest Neighbor algorithm

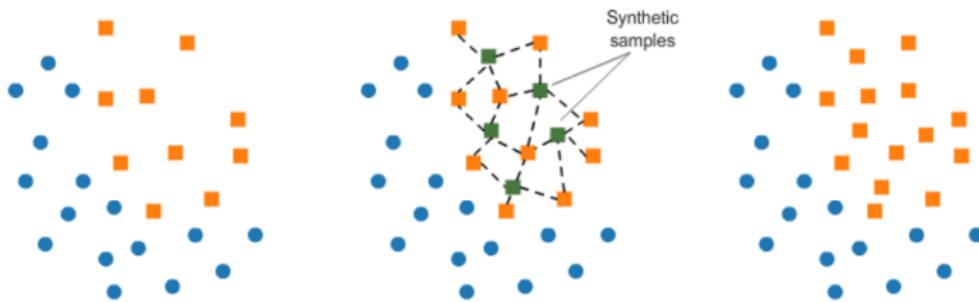


Figure: Workflow of SMOTE

4. Results

4.4. Machine Learning

4.4.1. Random Forest Classifier on Proportion

Random Forest with (Early vs. Late vs. Full) I

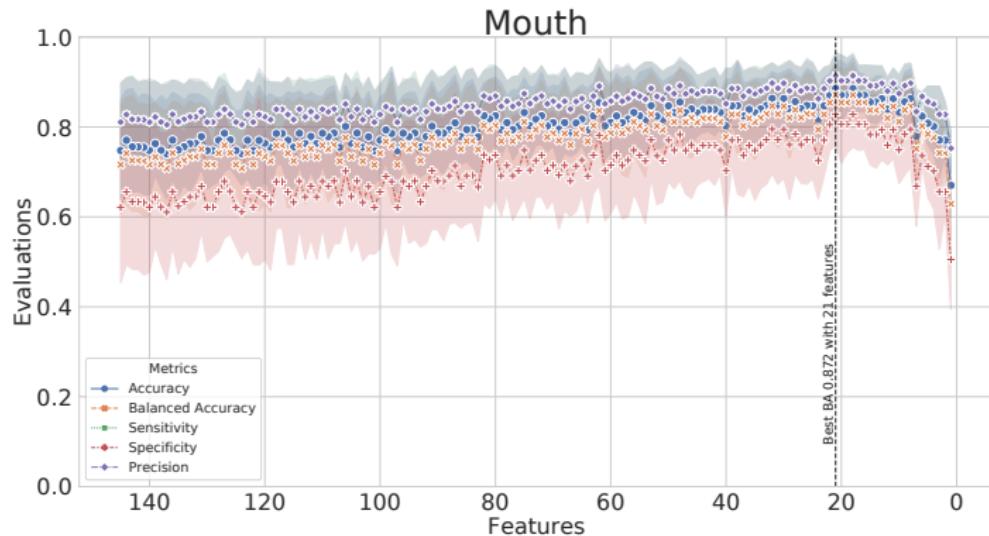


Figure: RF evaluations with feature counts

Random Forest with (Early vs. Late vs. Full) II

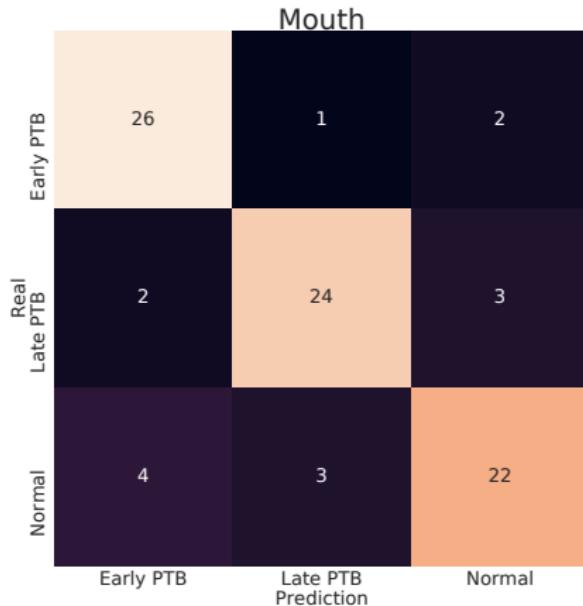


Figure: RF confusion matrix

Random Forest with (Early vs. Late vs. Full) III

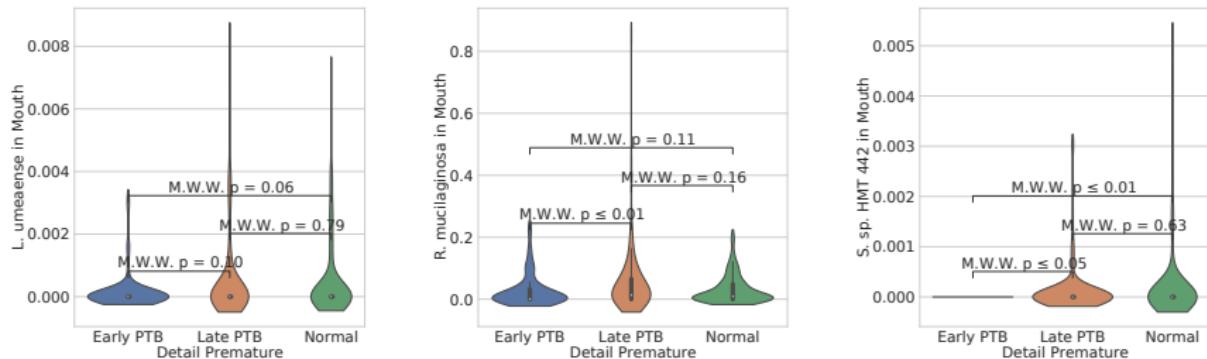


Figure: RF most important taxa

Random Forest with (Early vs. Late + Full) I

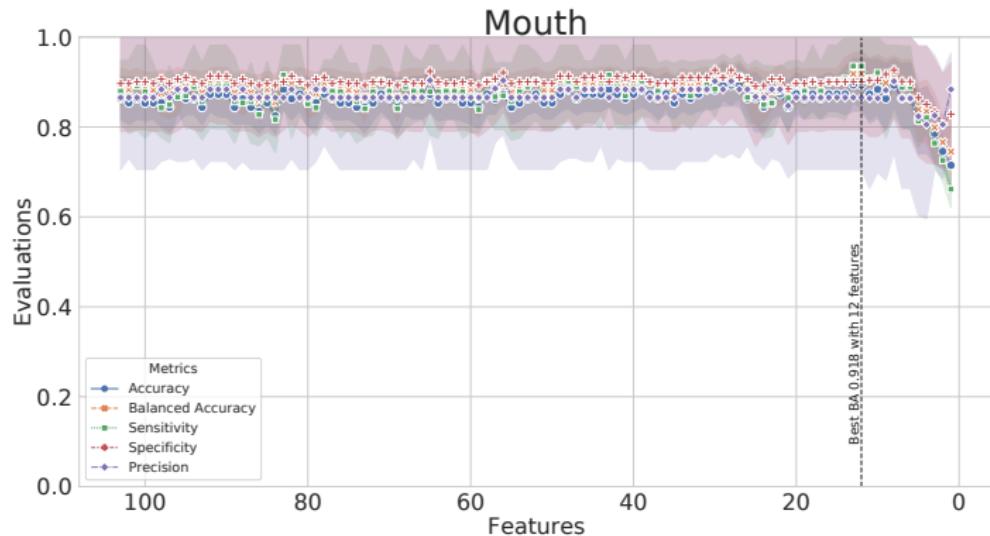


Figure: RF evaluations with feature counts

Random Forest with (Early vs. Late + Full) II

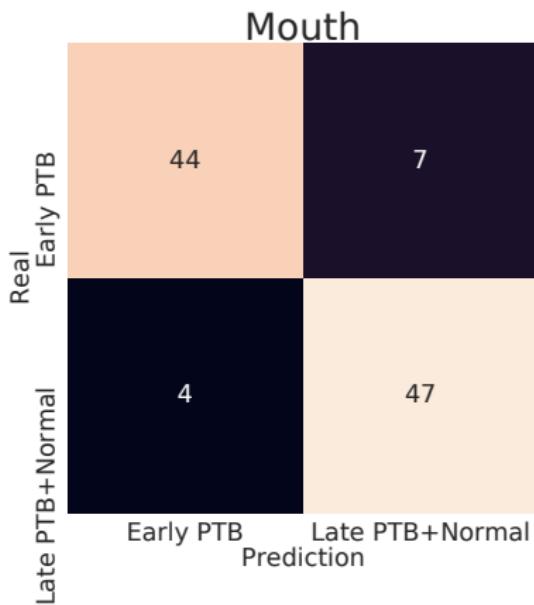


Figure: RF confusion matrix

Random Forest with (Early vs. Late + Full) III

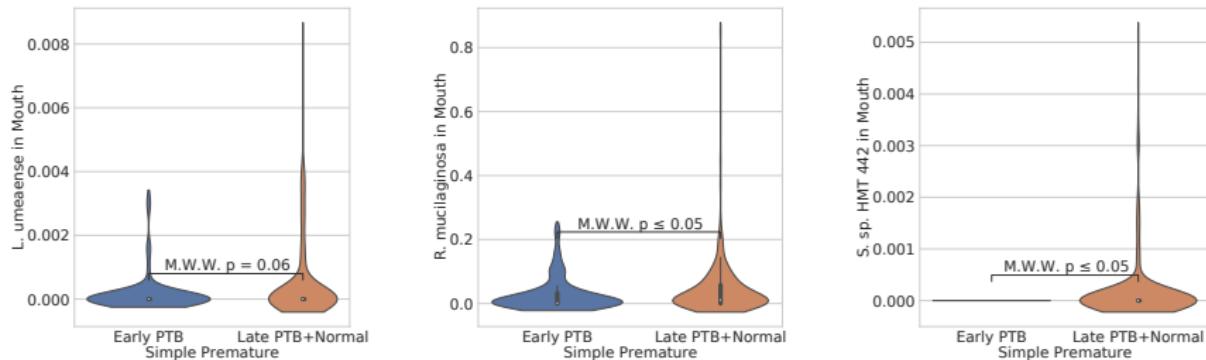


Figure: RF most important taxa

5. Discussion

6. References

References I

- 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
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321–357.
- Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., ... Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *science*, 312(5778), 1355–1359.

References II

- Heumann, C. L., Quilter, L. A., Eastment, M. C., Heffron, R., & Hawes, S. E. (2017). Adverse birth outcomes and maternal neisseria gonorrhoeae infection: a population-based cohort study in washington state. *Sexually transmitted diseases*, 44(5), 266.
- Hills, T., Sharpe, C., Wong, T., Cutfield, T., Lee, A., McBride, S., ... others (2022). Fetal loss and preterm birth caused by intraamniotic haemophilus influenzae infection, new zealand. *Emerging Infectious Diseases*, 28(9), 1747–1754.
- Janda, J. M., & Abbott, S. L. (2007). 16s rrna gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of clinical microbiology*, 45(9), 2761–2764.
- Kitazawa, T., Ishigaki, S., Seo, K., Yoshino, Y., & Ota, Y. (2018). Catheter-related bloodstream infection due to rhodotorula mucilaginosa with normal serum (1→ 3)- β -d-glucan level. *Journal de Mycologie Médicale*, 28(2), 393–395.

References III

- Li, D., Huang, Y., Sadykova, A., Zheng, W., Lin, L., Jin, C., ... Pan, S. (2021). Composition of the microbial communities at different body sites in women with preterm birth and their newborns. *Medicine in Microecology*, 9, 100046.
- 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
- 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

References IV

- Mendz, G. L., Petersen, R., Quinlivan, J. A., & Kaakoush, N. O. (2014). Potential involvement of campylobacter curvus and haemophilus parainfluenzae in preterm birth. *Case Reports, 2014*, bcr2014205282.
- Mignard, S., & Flandrois, J.-P. (2006). 16s rrna sequencing in routine bacterial identification: a 30-month experiment. *Journal of microbiological methods, 67*(3), 574–581.
- Mosquera, J., Zabalza, M., Laniero, M., & Blanco, J. (2000). Endocarditis due to gemella haemolysans in a patient with hemochromatosis. *Clinical microbiology and infection, 6*(10), 566–568.
- Olsen, G. J., & Woese, C. R. (1993). Ribosomal rna: a key to phylogeny. *The FASEB journal, 7*(1), 113–123.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research, 12*, 2825–2830.

References V

- Surve, M. V., Anil, A., Kamath, K. G., Bhutda, S., Sthanam, L. K., Pradhan, A., ... others (2016). Membrane vesicles of group b streptococcus disrupt feto-maternal barrier leading to preterm birth. *PLoS pathogens*, 12(9), e1005816.
- Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., ... others (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, 92(2), 698–715.
- Tucker, J., & McGuire, W. (2004). Epidemiology of preterm birth. *Bmj*, 329(7467), 675–678.
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, 449(7164), 804–810.

References VI

Voronkov, L., Solonovych, A., Liashenko, A., & Revenko, I. (2018). Prognostic value of cognitive tests and their combination in patients with chronic heart failure and reduced left ventricular ejection fraction. *Eureka: health sciences*(6), 36–45.