

# Metagenome Analysis of Preterm Birth

Jaewoong Lee    Semin Lee

Department of Biomedical Engineering  
Ulsan National Institute of Science and Technology

*jwlee230@unist.ac.kr*

2022-12-28

# 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  $\geq 37$  GW

Detailed PTB:

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

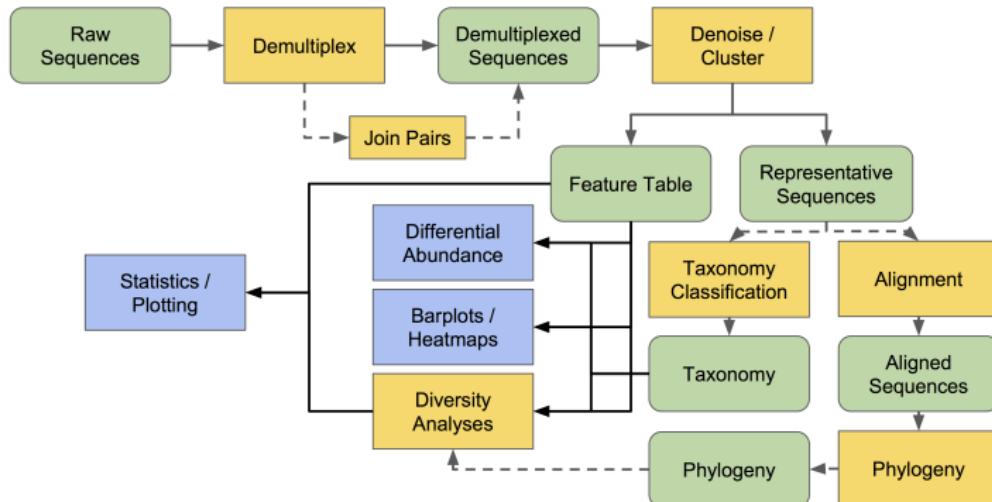
- JBNU/Helixco data
  - First data
  - Second data
  - Stool data

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

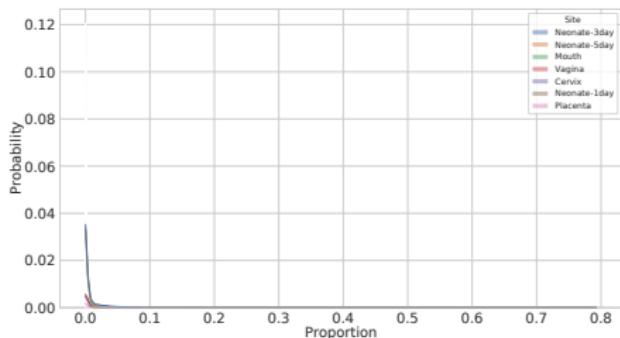


Figure: Proportion distribution

# Microbial community with Proportion

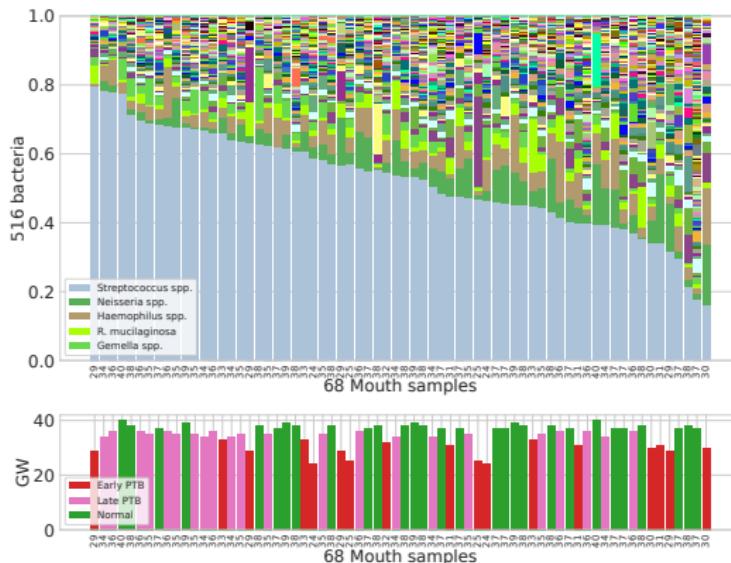
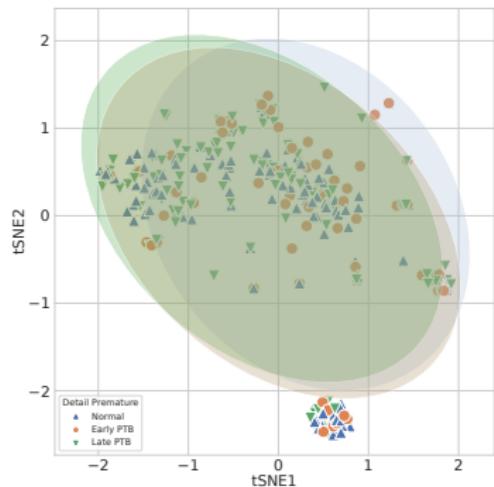
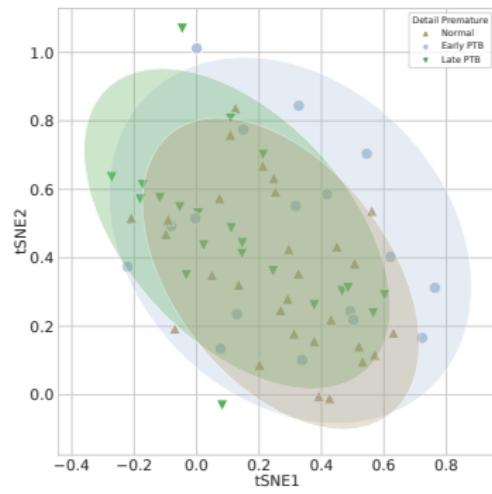


Figure: Microbial community with Proportion

# 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

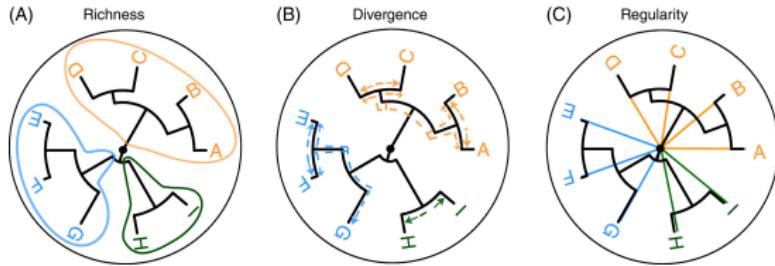


Figure: Three dimensions of phylogenetic information (C. M. Tucker et al., 2017)

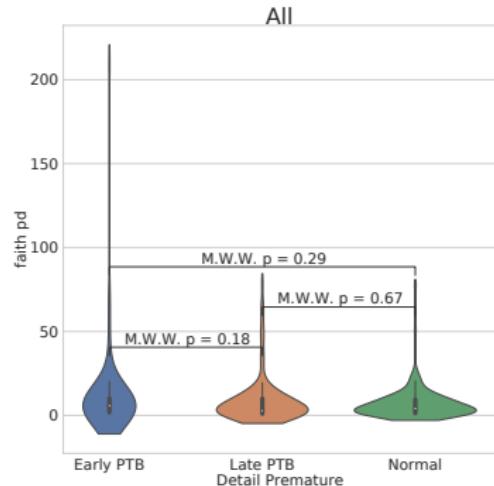
- A quantitative measure that shows richness, divergence, and regularity (C. M. Tucker et al., 2017)
- Alpha diversity: the richness of taxa **at a single community**
- Beta diversity: taxonomy differentiation **between communities**

## 4. Results

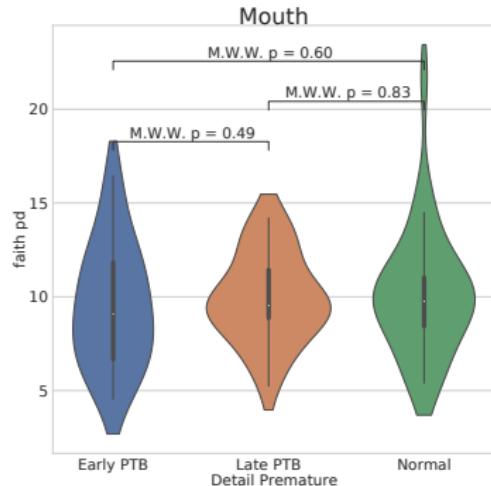
### 4.2. Diversity Index

#### 4.2.1. Alpha-diversity

# Violin Plot with Alpha-diversity I



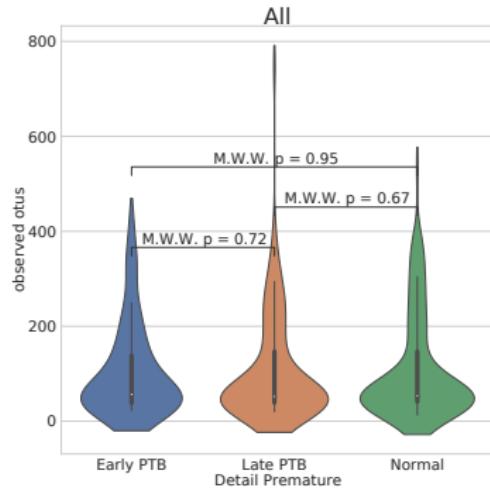
(a) All



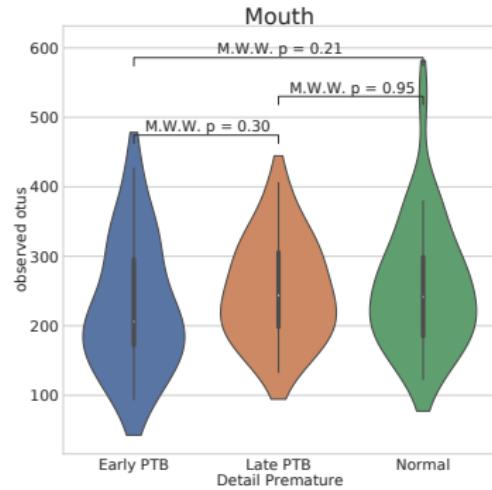
(b) Mother Mouth

Figure: Detail premature & Faith's PD

# Violin Plot with Alpha-diversity II



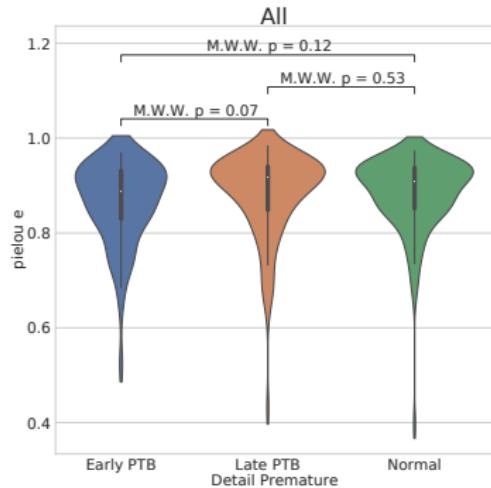
(a) All



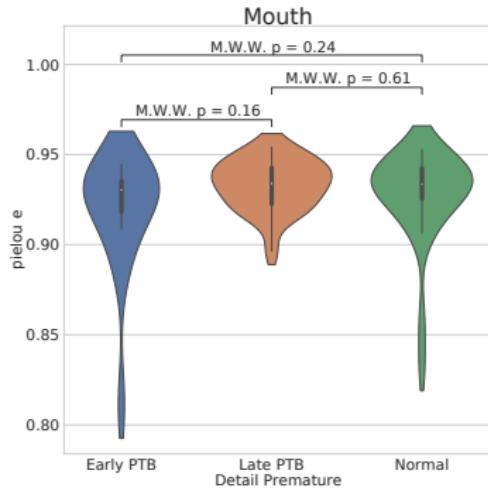
(b) Mother Mouth

Figure: Detail premature & Observed OTUs

# Violin Plot with Alpha-diversity III



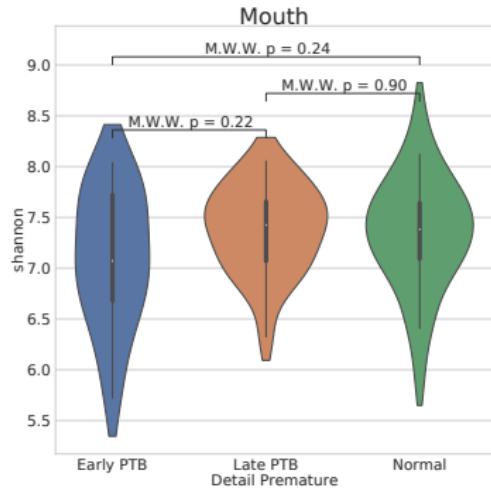
(a) All



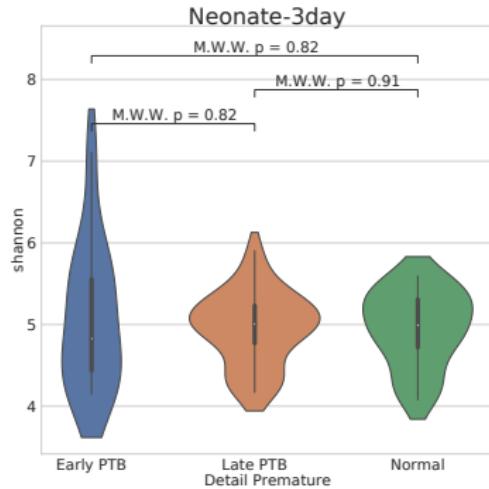
(b) Mother Mouth

Figure: Detail premature & Pielou Evenness

# Violin Plot with Alpha-diversity IV



(a) All



(b) Mother Mouth

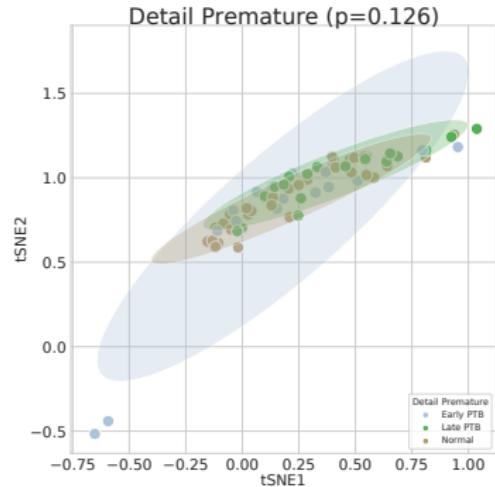
Figure: Detail premature & Shannon Entropy

## 4. Results

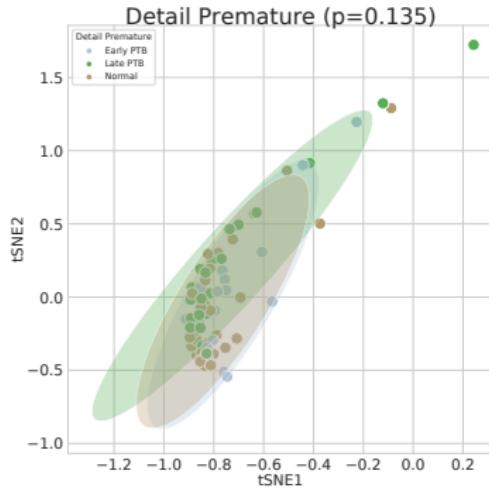
### 4.2. Diversity Index

#### 4.2.2. Beta-diversity

# Beta-diversity t-SNE plots I



(a) Bray-Curtis



(b) Euclidean

Figure: Beta-diversity t-SNE plots

## 4. Results

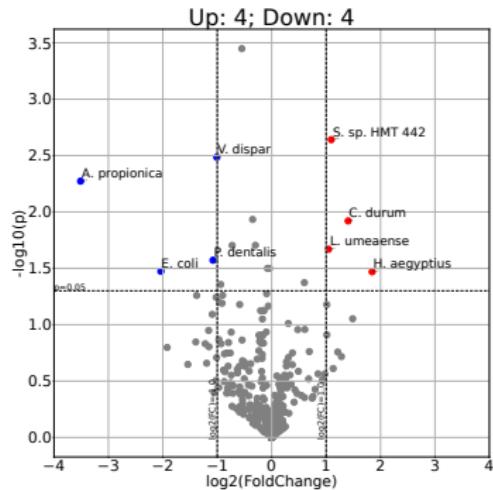
### 4.3. Taxonomy Analyses

## 4. Results

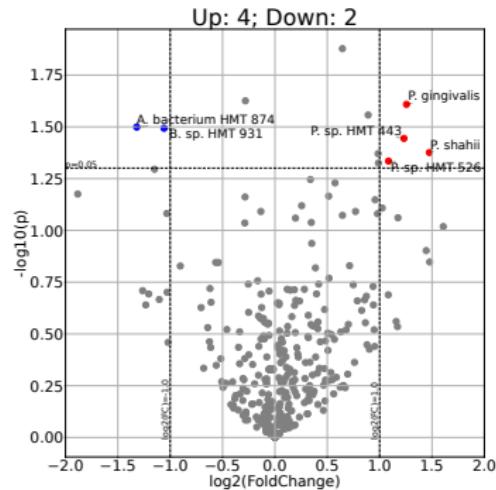
### 4.3. Taxonomy Analyses

#### 4.3.1. Differentially Abundant Taxa

# Volcano plots I



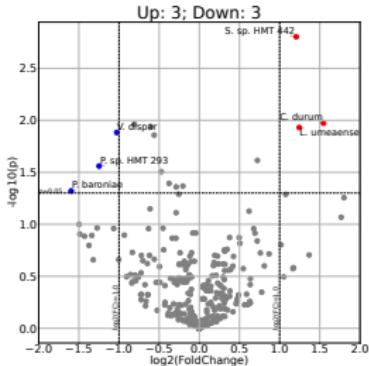
(a) Early PTB vs. other



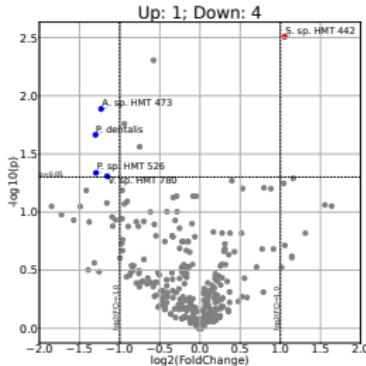
(b) PTB vs. Normal

Figure: DAT in Mouth

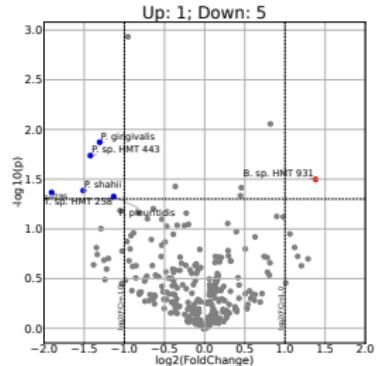
## Volcano plots II



(a) Early vs. Late



(b) Early vs. Normal



(c) Late vs. Normal

## Figure: DAT in Mouth

# Venn Diagrams

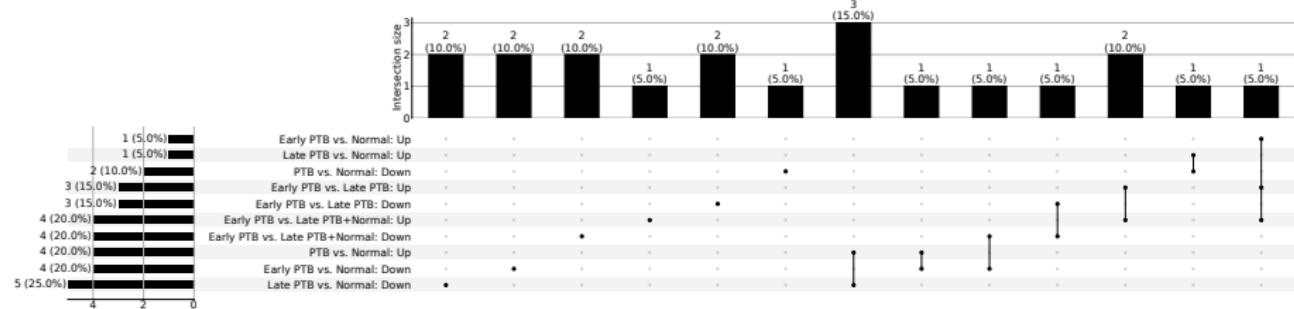
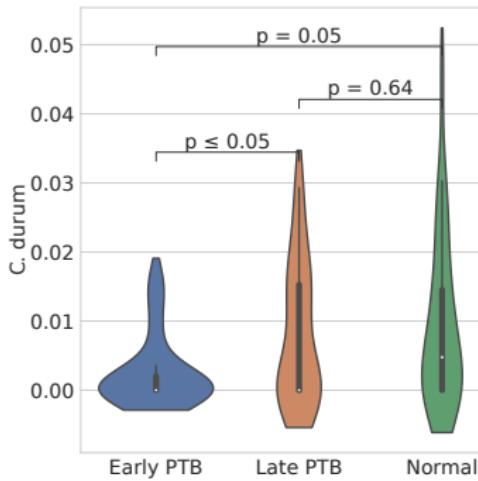


Figure: DAT Upset plot

## Union of DAT

A total of 20 taxa selected as DAT.

# Violin Plots I



## *C. durum*

- *Corynebacterium durum*
- Poly-microbial interactions  
⇒ Oral mucosal & Gingival cells (Redanz et al., 2021).
- Synergism between *C.* & *Streptococcus* in oral commensals (Treerat et al., 2020).

Figure: *C. durum*

# Violin Plots II

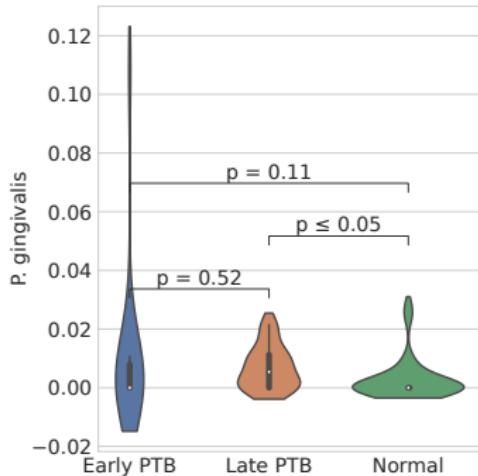


Figure: *P. gingivalis*

## *P. gingivalis*

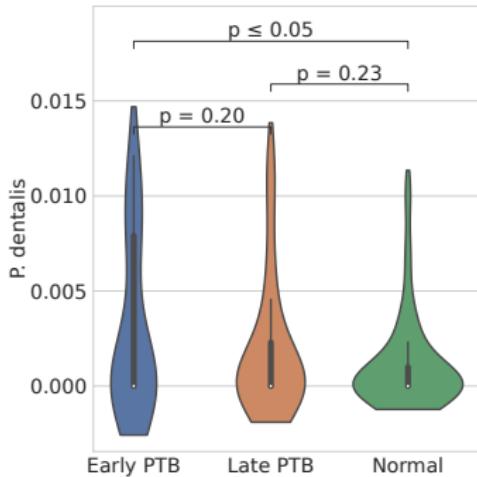
- *Porphyromonas gingivalis*
- *P. gingivalis* is well-known periodontitis pathogen.
- *P. gingivalis* is associated with bacterial vaginosis (Africa, Nel, & Stemmet, 2014).
- *P. gingivalis* ⇒ Adverse pregnancy outcomes (Miyauchi et al., 2018).

# Violin Plots III

## Prevotella spp.

- *P.* interactions with human health (Tett, Pasolli, Masetti, Ercolini, & Segata, 2021).
- *P.* spp. might have different role in PTB (Freitas, Bocking, Hill, & Money, 2018).
- *P.* spp. are associated with human infections (Sharma, Garg, Hasan, & Shirodkar, 2022).

# Violin Plots IV



## *P. dentalis*

- *Prevotella dentalis*
- *P. dentalis* is observed as colorectal cancer-promoting bacterium (Shi et al., 2022).

Figure: *P. dentalis*

# Violin Plots V

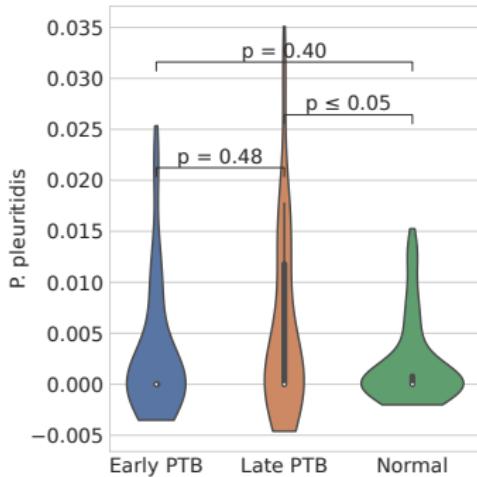


Figure: *P. pleuritidis*

## *P. pleuritidis*

- *Prevotella pleuritidis*
- *P. pleuritidis* causes lung abscess (Asif, Roy, & Ahmad, 2020).
- *P. pleuritidis* is abundant in oral microbiome of smokers (Al Bataineh et al., 2020).
- *P. pleuritidis* associated with gastric cancer (Yang et al., 2022).

## 4. Results

### 4.4. Machine Learning

# ML algorithm comparison

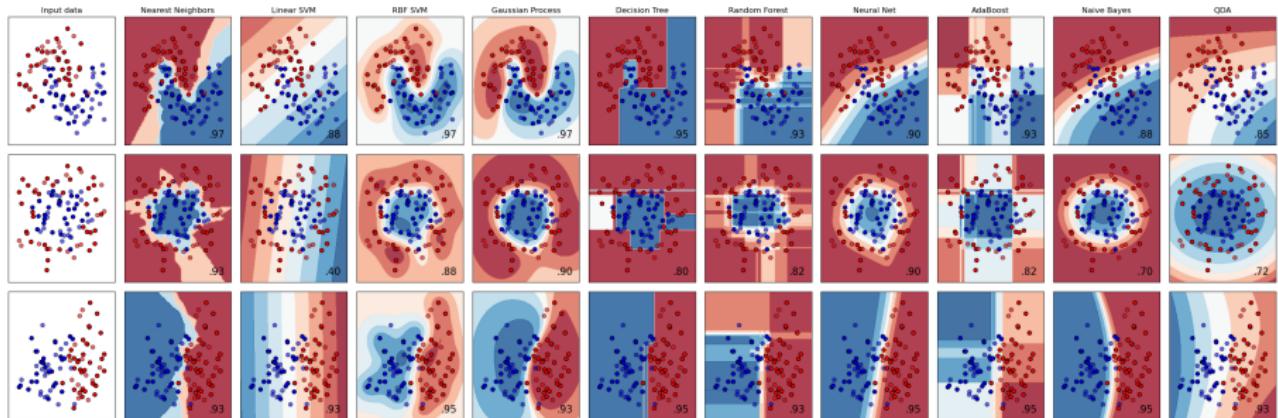


Figure: Classification Comparison (Pedregosa et al., 2011)

# Oversampling

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

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

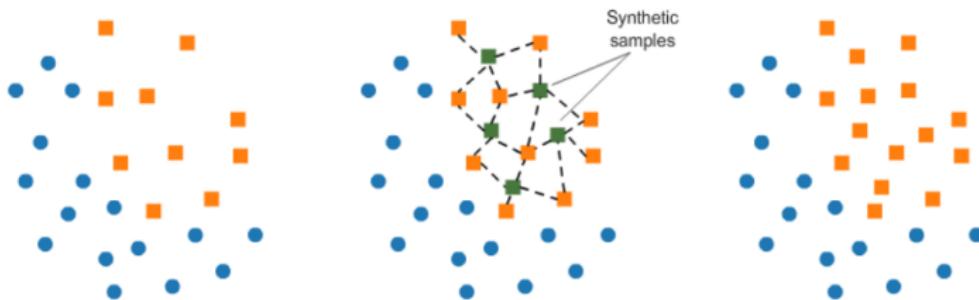


Figure: Workflow of SMOTE

## 4. Results

### 4.4. Machine Learning

#### 4.4.1. Random Forest Classifier on Proportion

# Random Forest with (Early vs. Late vs. Normal) I

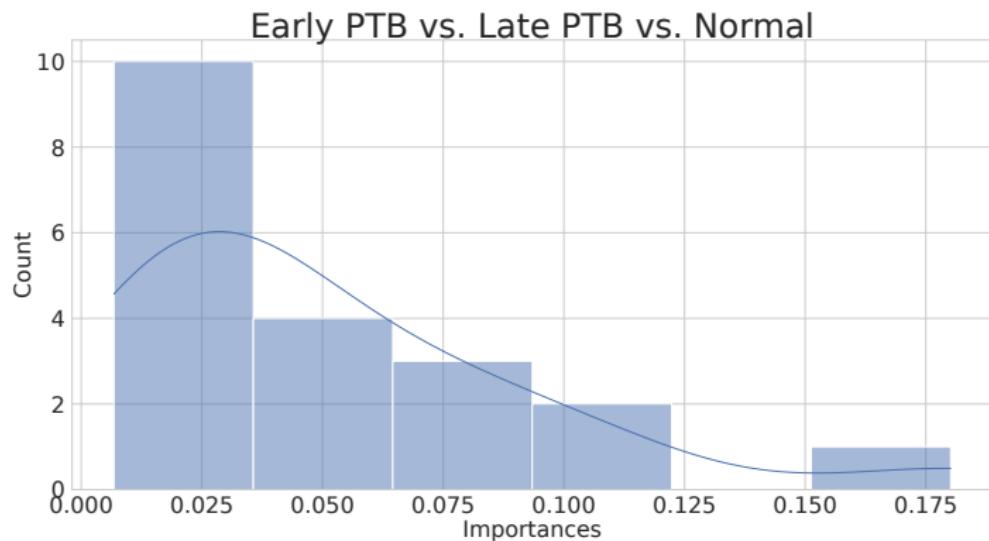


Figure: RF importances

# Random Forest with (Early vs. Late vs. Normal) II

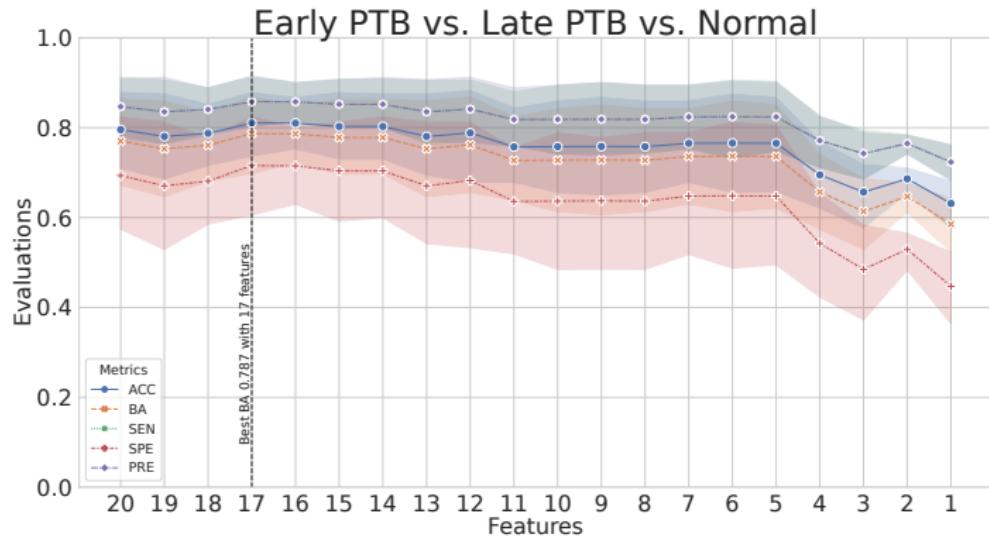


Figure: RF evaluations with feature counts

# Random Forest with (Early vs. Late vs. Normal) III

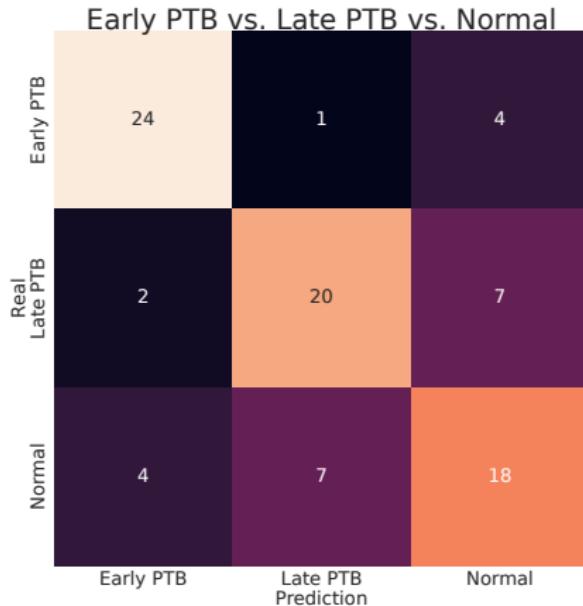


Figure: RF confusion matrix

# Random Forest with (Early vs. Late + Normal) I

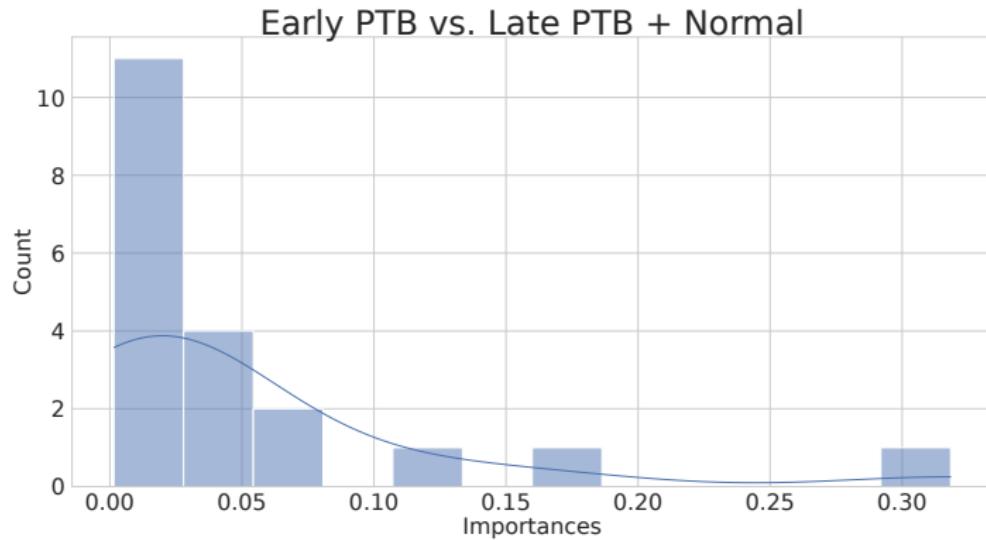


Figure: RF importances

# Random Forest with (Early vs. Late + Normal) II

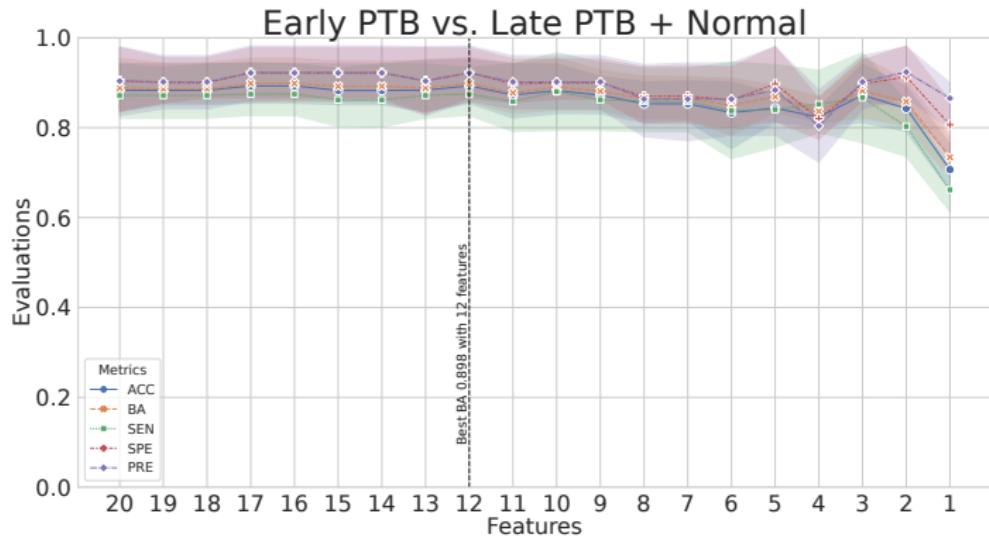


Figure: RF evaluations with feature counts

# Random Forest with (Early vs. Late + Normal) III

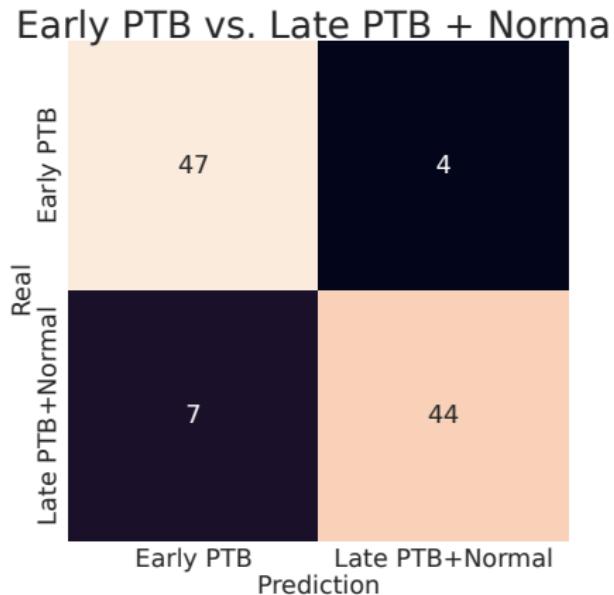


Figure: RF confusion matrix

# Random Forest with (PTB vs. Normal) I

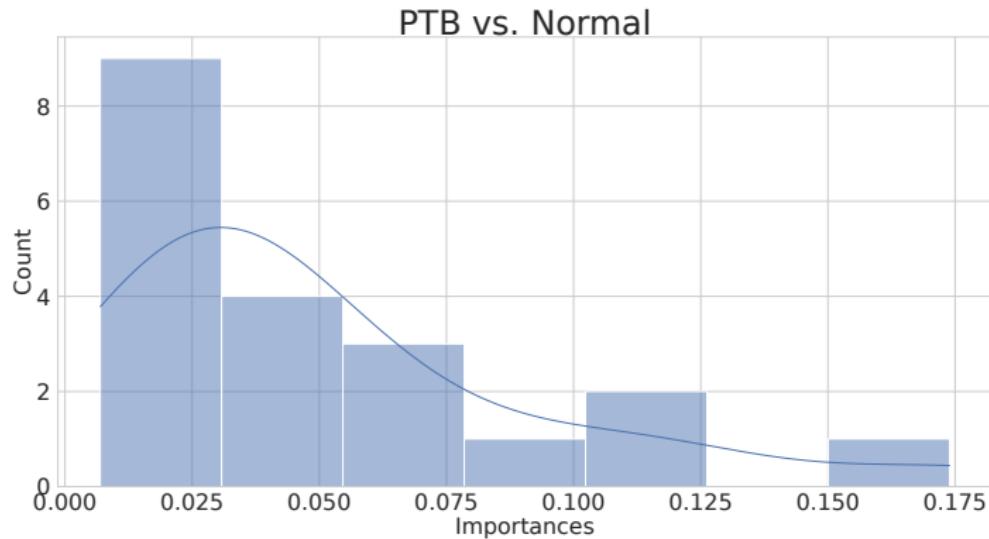


Figure: RF importances

# Random Forest with (PTB vs. Normal) II

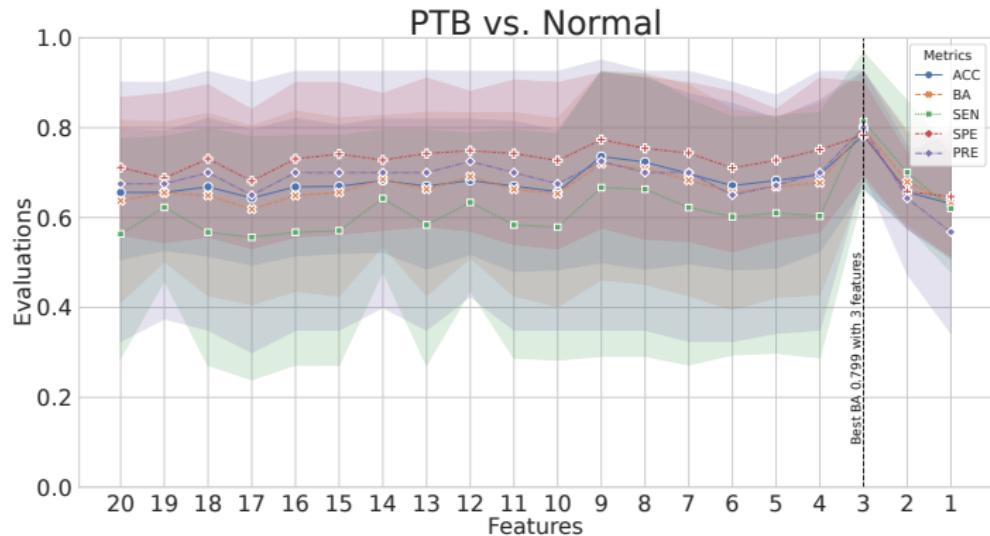


Figure: RF evaluations with feature counts

# Random Forest with (PTB vs. Normal) III

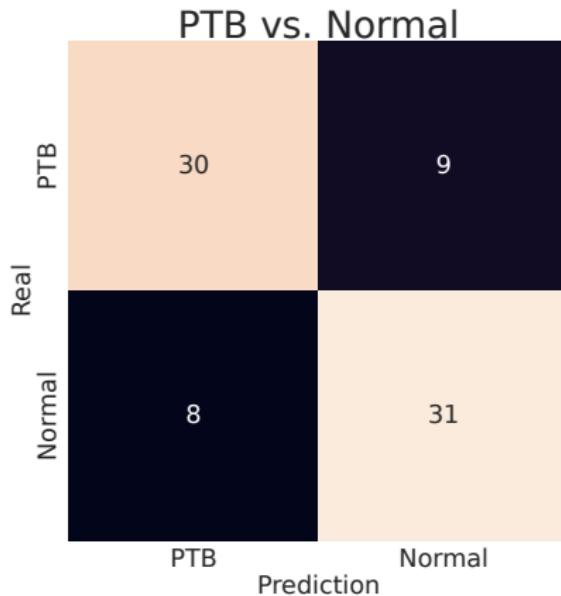


Figure: RF confusion matrix

## 5. Discussion

## 6. References

## References I

- Africa, C. W., Nel, J., & Stemmet, M. (2014). Anaerobes and bacterial vaginosis in pregnancy: virulence factors contributing to vaginal colonisation. *International journal of environmental research and public health*, 11(7), 6979–7000.
- Al Bataineh, M. T., Dash, N. R., Elkhazendar, M., Alnusairat, D. M. H., Darwish, I. M. I., Al-Hajjaj, M. S., & Hamid, Q. (2020). Revealing oral microbiota composition and functionality associated with heavy cigarette smoking. *Journal of translational medicine*, 18(1), 1–10.
- Asif, A. A., Roy, M., & Ahmad, S. (2020). Rare case of prevotella pleuritidis lung abscess. *BMJ Case Reports CP*, 13(9), e235960.

## References II

- 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.
- Freitas, A. C., Bocking, A., Hill, J. E., & Money, D. M. (2018). Increased richness and diversity of the vaginal microbiota and spontaneous preterm birth. *Microbiome*, 6(1), 1–15.
- 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 III

- 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)- $\beta$ -d-glucan level. *Journal de Mycologie Médicale*, 28(2), 393–395.

## References IV

- 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 V

- 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.
- Miyauchi, M., Ao, M., Furusho, H., Chea, C., Nagasaki, A., Sakamoto, S., ... Takata, T. (2018). Galectin-3 plays an important role in preterm birth caused by dental infection of porphyromonas gingivalis. *Scientific reports, 8*(1), 1–12.
- 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.

## References VI

- 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.
- Rao, B.-C., Zhang, G.-Z., Zou, Y.-W., Ren, T., Ren, H.-Y., Liu, C., ... Ren, Z.-G. (2022). Alterations in the human oral microbiome in cholangiocarcinoma. *Military Medical Research*, 9(1), 1–4.
- Redanz, U., Redanz, S., Treerat, P., Prakasam, S., Lin, L.-J., Merritt, J., & Kreth, J. (2021). Differential response of oral mucosal and gingival cells to corynebacterium durum, streptococcus sanguinis, and porphyromonas gingivalis multispecies biofilms. *Frontiers in Cellular and Infection Microbiology*, 11, 686479.

## References VII

- Sharma, G., Garg, N., Hasan, S., & Shirodkar, S. (2022). Prevotella: An insight into its characteristics and associated virulence factors. *Microbial Pathogenesis*, 105673.
- Shi, F., Liu, G., Lin, Y., Han, J., Chu, E. S., Shi, C., ... others (2022). Altered gut microbiome composition by appendectomy contributes to colorectal cancer. *Oncogene*, 1–11.
- 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.
- Tett, A., Pasolli, E., Masetti, G., Ercolini, D., & Segata, N. (2021). Prevotella diversity, niches and interactions with the human host. *Nature Reviews Microbiology*, 19(9), 585–599.

## References VIII

- Treerat, P., Redanz, U., Redanz, S., Giacaman, R. A., Merritt, J., & Kreth, J. (2020). Synergism between *corynebacterium* and *streptococcus sanguinis* reveals new interactions between oral commensals. *The ISME journal*, 14(5), 1154–1169.
- 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 IX

- 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.
- Yang, Y., Long, J., Wang, C., Blot, W. J., Pei, Z., Shu, X., . . . others (2022). Prospective study of oral microbiome and gastric cancer risk among asian, african american and european american populations. *International journal of cancer*, 150(6), 916–927.