Metagenome Analysis of Preterm Birth

Jaewoong Lee   Semin Lee

Department of Biomedical Engineering
Ulsan National Institute of Science and Technology

jwlee230@unist.ac.kr

2023-09-05

Overview

1. Introduction

Microbiome

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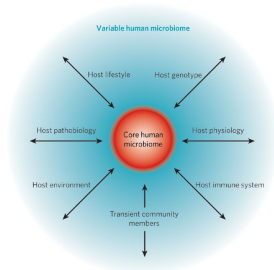

Figure: Concept of a core human microbiome (Turnbaugh et al., 2007)

rRNA

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

PTB:

1. PTB $<$ 37 GW (Gestational week)
2. 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 I

Data composition

59 pregnant women & 55 newborns

PTB

- Mother $\Rightarrow$ PTB: 30 & Normal: 29
- Newborn $\Rightarrow$ PTB: 25 & Normal: 30

Table: Clinical characteristics of mothers

Clinical	<37 GW (n=30)	≥37 GW (n=29)	p-value	Remarks
Cholesterol	289.6±72.6	269.6±50.6	0.496	
DBP	83.1±17.5	80.7±9.8	0.887	
Glucose	97.3±27.7	83.1±14.3	0.034	*
HDL	83.7±17.6	82.4±30.8	0.374	
Hb	11.6±1.4	12.2±1.6	0.095	
Hct	34.7±3.5	36.2±3.9	0.115	
LDL	151.3±32.6	157.3±26.4	0.779	
Mother Age	31.8±5.1	33.7±4.4	0.142	
SBP	142.0±28.0	129.3±15.7	0.079	
Weight gain	9.0±5.8	11.6±4.0	0.022	*
Advanced maternal age	8 (26.7%)	10 (34.5%)	0.580	
C-section	20 (66.7%)	24 (82.8%)	0.233	
Gestational Diabetes	0 (0.0%)	3 (10.3%)	0.112	
Hypertension	10 (33.3%)	5 (17.2%)	0.233	
Mother Antibiotics	13 (43.3%)	4 (13.8%)	0.020	*
Mother Steroid	18 (60.0%)	0 (0.0%)	0.000	*
Obesity	8 (26.7%)	7 (24.1%)	1.000	
PROM	12 (40.0%)	1 (3.4%)	0.001	*
Preterm Labor	13 (43.3%)	1 (3.4%)	0.000	*
Too much weight gain	4 (13.3%)	4 (13.8%)	1.000	

Data Composition III

Table: Clinical characteristics of newborns

Clinical	<37 GW (n=25)	≥37 GW (n=30)	p-value	Remarks
Apgar Score	8.4±1.2	9.7±0.7	0.000	*
Gestational Week	33.3±2.5	38.0±1.0	0.000	*
Hospitalized Day	20.6±19.5	8.3±6.2	0.002	*
Weight	2150.7±587.1	3267.9±397.8	0.000	*
CPAP	9 (36.0%)	2 (6.7%)	0.015	*
Dyspnea	11 (44.0%)	2 (6.7%)	0.002	*
Gender	12 (48.0%)	13 (43.3%)	0.790	
Neonate Antibiotics	7 (28.0%)	6 (20.0%)	0.537	
PROM	9 (36.0%)	1 (3.3%)	0.003	*
Respirator	6 (24.0%)	1 (3.3%)	0.039	*
Sepsis	6 (24.0%)	5 (16.7%)	0.521	

Statistical tests

- Continuous: M.W.W. test
- Categorical: Fisher exact test

3. Methods

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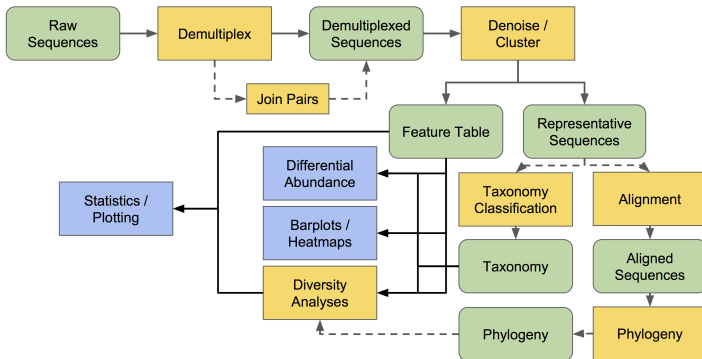
Figure: QIIME 2 workflow (Bolyen et al., 2019; Mandal et al., 2015; McDonald et al., 2012)

4. Results
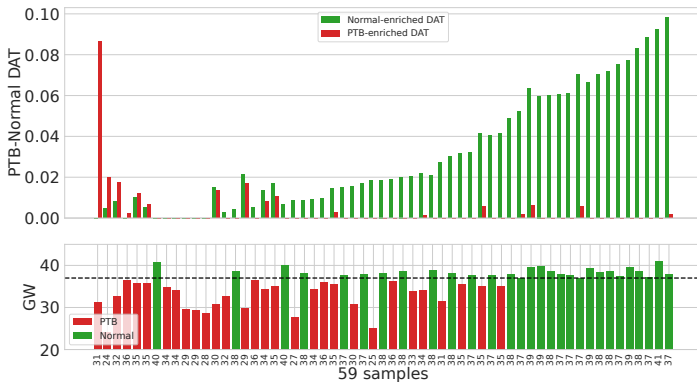
4. Results

4.1. Taxonomy Overview

Figure: Microbial community with Proportion
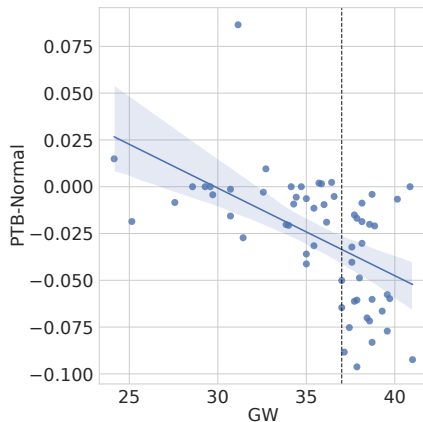
Microbial community with Proportion II

Figure: Microbial community with Proportion
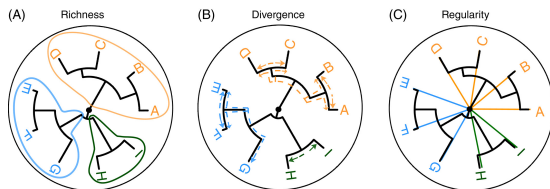
4. Results

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
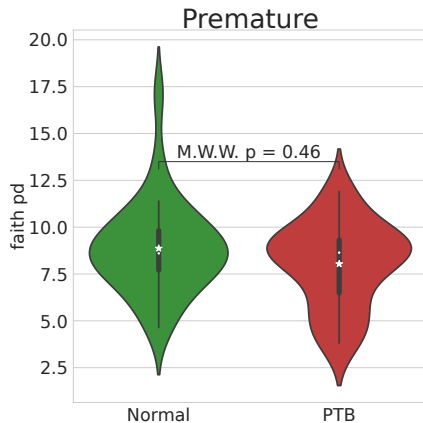
Figure: Premature & Faith's PD

4. Results

4.2. Diversity Index
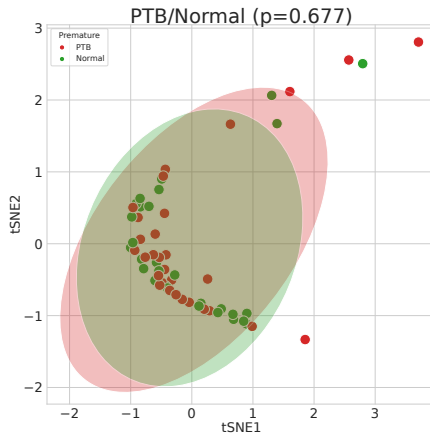
4.2.2. Beta-diversity

Figure: Hamming distance index t-SNE plot

4. Results

4.3. Taxonomy Analyses

4. Results

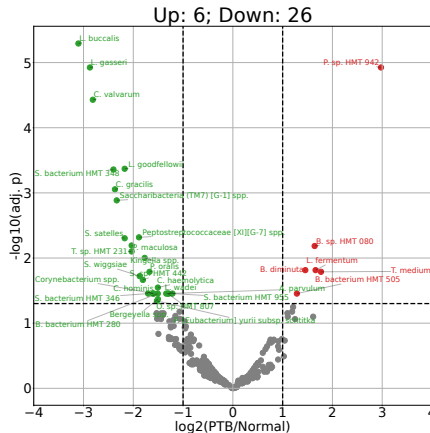## 4.3. Taxonomy Analyses

4.3.1. Differentially Abundant Taxa

Volcano plots

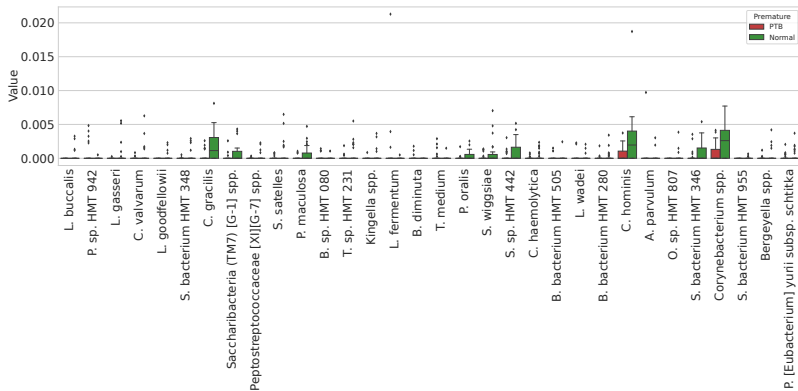Figure: DAT in Mouth

Box plots

Figure: DAT box plots

4. Results

4.4. Machine Learning

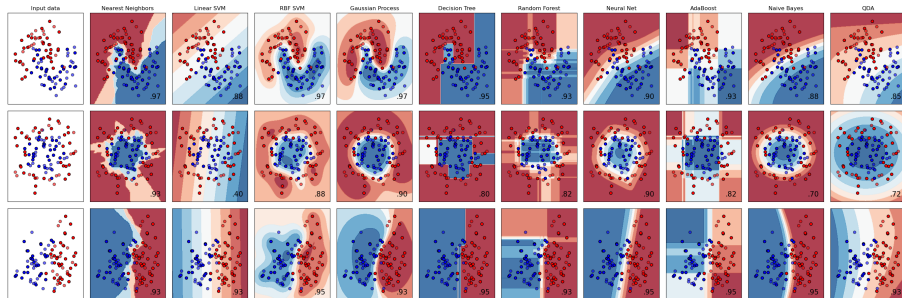Figure: Classification Comparison (Pedregosa et al., 2011)

4. Results

4.4. Machine Learning

4.4.1. Random Forest Classifier on Proportion
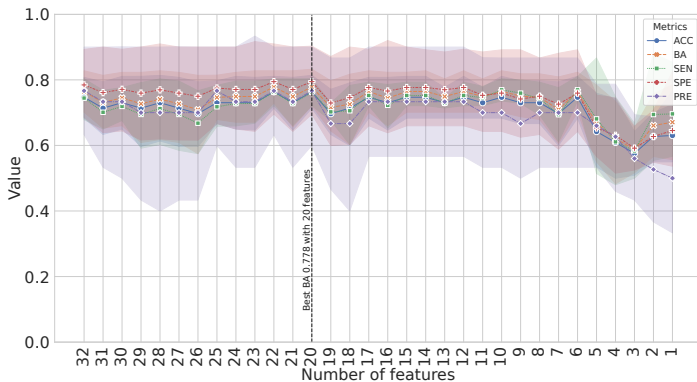
Figure: RF evaluations with feature counts
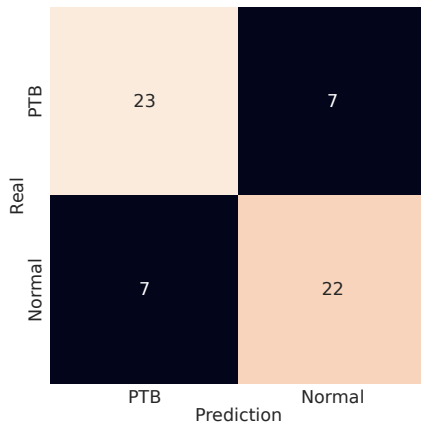
Figure: RF confusion matrix
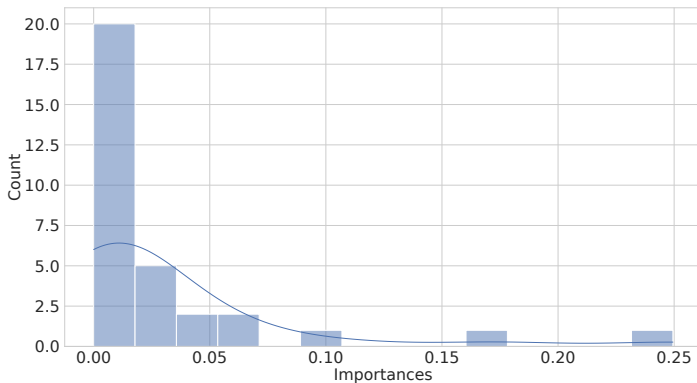
Figure: RF importances

Highest Importances

1. *Cardiobacterium hominis*
2. *Campylobacter gracilis*
3. *Corynebacterium* spp.
4. *Selenomonas* sp. *HMT 442*
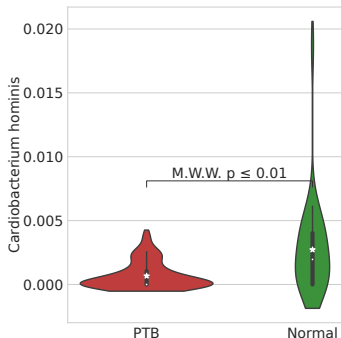
Figure: *C. hominis*

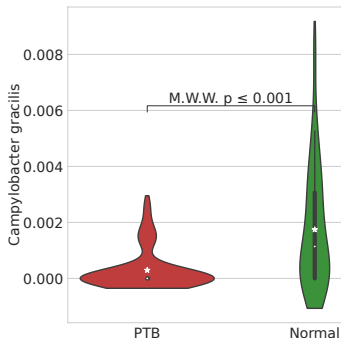**Cardiobacterium hominis**

- *C.hominis* causes endocarditis (Malani, Aronoff, Bradley, & Kauffman, 2006).

Figure: *C. gracilis*

Campylobacter gracilis

- Odontogenic brain abscess with *C. gracilis* (Jang, Elliott, Herman, & Booher, 2021).
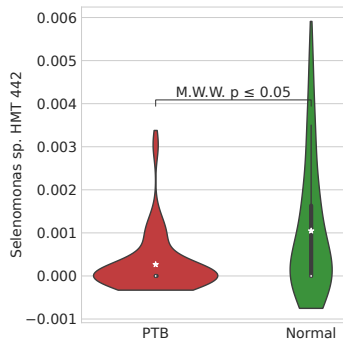
Figure: *Corynebacterium* spp.
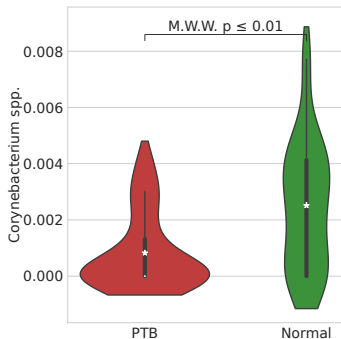
Corynebacterium spp.

- *C.* sp. HMSC078H07 associated with PTB (Huo, Jiang, & Zhao, 2022) enriched in PTB ?

Figure: *S.* sp. *HMT 442*

Selenomonas sp. *HMT 442*

- *S.* sp. *HMT 442* associated with dental caries (da Costa Rosa et al., 2021).

5. Discussion

6. References

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

da Costa Rosa, T., de Almeida Neves, A., Azcarate-Peril, M. A., Divaris, K., Wu, D., Cho, H., . . . others (2021). The bacterial microbiome and metabolome in caries progression and arrest. *Journal of Oral Microbiology*, *13*(1), 1886748.

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

Huo, Y., Jiang, Q., & Zhao, W. (2022). Meta-analysis of metagenomics reveals the signatures of vaginal microbiome in preterm birth. *Medicine in Microecology*, *14*, 100065.

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.

Jang, T. L., Elliott, B. P., Herman, D. G., & Booher, K. J. (2021). Odontogenic brain abscess with campylobacter gracilis and fusobacterium nucleatum complicated by rupture into the ventricle. *Infectious Diseases in Clinical Practice*, *29*(6), e437–e439.

Malani, A., Aronoff, D., Bradley, S., & Kauffman, C. (2006). Cardiobacterium hominis endocarditis: two cases and a review of the literature. *European Journal of Clinical Microbiology and Infectious Diseases*, *25*, 587–595.

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

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.

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.

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.

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.