Effects of HAART and Time on the Beta Diversity of Breast Milk Microbiome in HIV-infected Postpartum Women

Ningxin Ma

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2021

Committee:
Barbra Richardson
James Hughes

Program Authorized to Offer Degree:

Biostatistics

University of Washington

Abstract

Effects of HAART and Time on the Beta Diversity of
Breast Milk Microbiome in HIV-infected Postpartum Women

Ningxin Ma

Chair of the Supervisory Committee:

Barbra Richardson
Department of Biostatistics

Introduction Highly active antiretroviral therapy (HAART) has been used in HIV-infected pregnant women to suppress viral replication and reduce perinatal transmission. However, the influence of HAART on the breast milk microbiome remains largely unknown. In addition, analysis of unbalanced longitudinal studies of β-diversity data is limited by a lack of appropriate statistical methods. The purpose of this study is to investigate the effects of HAART and time on β-diversity of breast milk microbiome from HIV-infected women during the first month postpartum using a novel statistical analysis method.

Methods HIV-infected pregnant women in Nairobi, Kenya in two separate studies were randomized to receive either HAART (treated group) during pregnancy to 6 months postpartum or short course zidovudine (control group) up to delivery. Breast milk samples were collected from 25 treated and 24 control women every week during the first month postpartum. These samples were subjected to 16S ribosomal sequencing. Microbial community analysis (PERMANOVA with restricted permutation for β-diversity) was used to determine the effects of HAART and time on the breast milk microbiome.

Results PERMANOVA analysis revealed statistically significant changes in breast

milk microbiome β-diversity when comparing postpartum week 1 to week 4 ($p < 0.01$). In contrast, no obvious difference was detected between the treated and control groups.

Conclusion PERMANOVA analysis with restricted permutation was used to evaluate the effects of time and treatment in mixed models on an unbalanced longitudinal dataset. During the first month postpartum, the β-diversity of the breast milk microbiome changed significantly in HIV-infected women from both arms of the trials. In contrast, HAART treatment had minimal effects on the β-diversity.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First of all, I would like to give my genuine thanks and gratitude to my advisor, Dr. Barbra Richardson. It has been a great pleasure to work with and learn from her. Her passion, knowledge and support have greatly inspired me and guided me through this endeavor. Moreover, this work wouldn't be complete without the insightful feedback from Dr. James Hughes. His suggestions have pushed this work to a higher level. I am extremely grateful for their tremendous input.

Next, I would like to thank the Department of Biostatistics and student cohorts for providing accommodations and the warmest support throughout this particularly difficult year. Graduate school during pandemic is not what I expected and working from home is not easy, but it turns out to be a great journey and we manage to build strong bonds together.

Last, my sincere thanks go to my friends and family. I wouldn't be able to make such a decision two years ago without your understanding and unconditional love. Thank you for bringing me joy and keeping me company all the time.

Chapter 1

INTRODUCTION

Human breast milk provides complete nutrition for infants in the first 6 months. In addition, it plays an important role in seeding the gut microbiome of infants, which is essential for protection against infection and immune system maturation [20, 13]. Thus, the maternal microbiota is considered a significant determinant of infant health [8]. Using culture-dependent techniques it has been shown that most bacteria from human breast milk are *Staphylococcus, Streptococcus, Lactobacillus* and *Bifidobacterium* spp [10]. The advancement of polymerase chain reaction (PCR) and next-generation sequencing suggest a more diverse and complex microbial community [15, 9]. Moreover, the microbiome composition undergoes distinct changes over lactation stages [14, 6, 10]. In a recent study in Guatemala, [11] identified a shift from *Staphylococcus* and *Streptococcus* to *Sphingobium* and *Pseudomonas* in the breast milk samples collected over the first 6 months postpartum.

HIV infection and the subsequent loss of CD4 cells can disrupt the balance between host and endogenous microbiota, leading to potential shifts in bacterial composition and diversity. A previous study of 121 breastfeeding women from a high-HIV prevalence area showed that HIV RNA-positive breast milk samples had an increased bacterial diversity and higher frequency of *Lactobacillus* spp [12]. In a more recent study in Kenya, no significant differences in composition or diversity were detected in breast milk samples from HIV-infected women, suggesting the microbiome resilience under immunosuppression [18].

Highly active antiretroviral therapy (HAART) suppresses viral replication and helps CD4 cells recover in HIV-infected patients. Consequently, the immune response to ART can induce microbiome changes. Previous studies on gut microbiome show that HIV-

infected individuals have decreased richness and diversity [25, 21]. These alterations can contribute to inflammation, pathogenesis, and the persistence of HIV [23, 16]. Moreover, HAART has not been consistently found to help restore gut microbiota, and may sometimes even enhance dysbiosis [3, 22]. In contrast, the changes in breast milk microbiome induced by HAART are not well understood.

In this study, we aim to compare the breast milk microbiome of HAART-treated HIV-infected women to those without HAART over time with a particular interest in β-diversity, which measures the difference in microbial composition between communities. Permutational multivariate analysis of variance (PERMANOVA) was developed for multivariate ANOVA in a chosen dissimilarity measure [2] and has been widely used in biodiversity studies. Some types of mixed effects models can be constructed in PERMANOVA using centroids [4, 5]. However, this method does not work on phylogenetic distance nor in unbalanced longitudinal analyses. To evaluate the effects of time and HAART accounting for the random effect of subject, we develop a method of using PERMANOVA with restricted permutations to characterize changes in microbiome β-diversity.

Chapter 2

METHODS

2.1 Study Design and Procedures

This study used data and specimens from two previous randomized clinical trials among HIV-infected pregnant women. Recruitment, description of study cohort, and sample collection have been published [19, 7]. In [19], women in the treatment arm received HAART (300 mg of zidovudine (ZDV), 150 mg of lamivudine, and 200 mg of nevirapine twice daily from 34 weeks gestation until 6 months postpartum. The control arm in [7] received the Thai-CDC regimen which was 300 mg of ZDV twice daily from 34 weeks gestation until the onset of labor, and every 3 hours from onset of labor until delivery.

This secondary analysis included 25 treated (HAART) and 24 control (Thai-CDC) subjects. Breast milk sample collection occurred once a week during the first 4 weeks postpartum. A total of 173 samples were subjected to 16S ribosomal sequencing.

2.2 Microbial Community Characterization

Matrices of relative abundances at multiple taxonomic levels were generated. Taxa with relative abundances less than 3% in all samples were classified as "other". Stacked bar plots were created on phylum or order level to display the proportions of major taxa in each group and at each time point.

The unweighted UniFrac (qualitative), weighted (quantitative) UniFrac and Jaccard dissimilarity distances were used to analyze microbiota diversity between samples. Unweighted UniFrac and Jaccard are qualitative measures based on presence / absence of species, while weighted UniFrac further incorporates phylogenetic information. Each of these distance measurements were used to calculate beta diversity, which is defined as the difference in microbial composition among communities. Each distance measure

was visualized with principal coordinate analysis (PCoA) plots.

2.3 Statistical Analysis

To investigate the changes in microbiome composition between the treatment and control groups over time, PERMANOVA was conducted in the space of weighted UniFrac distance. The analyses were executed using the vegan and permute packages in R. Besides PERMANOVA, zero-inflated beta regression (ZIBR) and negative binomial mixed model (NBMM) can also be used for correlated microbiome data, but these methods test individual biomarkers. To compare microbiomes at the community-level, [24] proposed a correlated sequence kernel association test (cSKAT). Furthermore, [17] extended cSKAT into a generalized linear mixed model, which can handle other types of traits besides those following a Gaussian distribution (GLMM-MiRKAT). It was used as an alternative method to confirm the PERMANOVA results. The details of this analysis is included in the Appendix D.

Instead of deriving the null distribution from theory, PERMANOVA randomly shuffles the data to create a distribution of test F-statistics under the null hypothesis and then calculates the empirical p-value [1]. Our implementation of PERMANOVA performs permutations based on null hypotheses to obtain correct p-values. This ensures the null distribution generated through permutations matches the corresponding F-statistic. Therefore, to investigate the different effects of time and treatment on microbiome β-diversity, it was necessary to construct two models and perform two separate permutation schemes. The sample codes were shown in the Appendix B. A confirmatory analysis using centroids in PERMANOVA was performed and included in the Appendix C.

2.3.1 Time effect

The model contains two covariates: group and time point, and uses type III sums of squares (SS) to evaluate the marginal effect of time. Under the null hypothesis, the microbiomes at four time points should be equivalent in terms of their β-diversity index. However, completely random shuffling of the time points is inappropriate since obser-

vations from the same subject are correlated. Therefore, permutations were restricted within each subject.

In cases in which the PERMANOVA results show strong evidence against the null hypothesis, post-hoc tests were performed to identify which pair(s) of time points were different. P-values for multiple testing were adjusted using the Benjamini-Hochberg procedure. Detailed information is shown in the Appendix.

2.3.2 Treatment effect

The random effect of subject (id) needs to be taken into consideration when evaluating the effect of HAART. Therefore, the model contains three covariates: group, id and time point, and uses type I sequential sums of squares. However, the default PERMANOVA setting incorrectly calculates the F-statistic of treatment group F_g as the ratio of the mean squares (MS) of group to MS of residual, and uses random permutations.

The correct F_g should be the ratio of MS(group) to MS(id). Accordingly, the permutations are implemented to randomly assign group labels to each id. In each permuted scenario, F-statistic of treatment group is calculated to generate the null distribution. An empirical p-value is calculated as (1 + number of $F > F_g$) / (1 + total number of F). It is worth noting that this permutation scheme requires the number of observations within each subject to be equal. After removal of subjects with missing data, 17 subjects in the treated group and 15 in the control group, each with 4 time points, were retained.

Sensitivity analyses were done to examine the robustness of the results. This analysis was repeated on (i) subsets of subjects with three complete time points and (ii) subsets of a single time point, which resulted in slightly larger sample sizes. These outcomes were compared qualitatively to see if removal of missing data dramatically changed the overall conclusion.

Besides PERMANOVA, a distance-based kernel association test based on the generalized linear mixed model (GLMM-MiRKAT) was used as an alternative analysis method. The details of these analyses are included in the Appendix.

Chapter 3

RESULTS

3.1 Breast Milk Microbiome Composition

Overall, 24 control and 25 HAART-treated HIV-infected women were included in the study. Women had a median age of 25.5 and 26.0 in control and treated groups, respectively. Baseline CD4 count was lower in treated (260 cells/mm^3) than control (446 cells/mm^3) women (Table 3.1). A total of 173 breast milk samples were collected during the first month postpartum and subjected to 16S sequencing. Among all 49 participants, 32 had four samples, 13 had three samples, 2 had two samples and 2 had only one sample.

Stacked bar plots representing relative abundance of bacterial taxa on the order level are shown in Figure 3.1. The taxa with the largest abundance are compared between groups and across time points. In both groups, Bacillales and Lactobacillales had the highest abundance. They were 29.3% and 20.9% in the treated group, 23.6% and 25.7% in the control group, respectively (Figure 3.1A). This is consistent with previous findings that genus *Staphylococcus*, Bacillales and genus *Streptococcus*, Lactobacillales were predominant in the breast milk microbiome of HIV-infected women [12, 18]. Over the first 4 weeks, these two orders had a decreasing abundance, while the proportions of Actinomycetales and Clostridiales increased (Figure 3.1B).

To visualize the high-dimensional microbiome composition data, PCoA plots were generated based on weighted UniFrac metrics. The first two principal components account for 21.9% and 16.3% of variation, respectively. The inter-sample distances between groups or time points were compared. The 95% confidence interval ellipse of the HAART group largely covered that of the control group, although the HAART group had slightly larger dispersion (Figure 3.2A). In contrast, samples at four time points showed some

Table 3.1: Baseline characteristics of the subjects and sample sizes at each time point. Displayed in Number (%) or Median (IQR). Statistical significance assessed by Fisher's exact test, Mann-Whitney test.

Maternal Characteristics	**Control (n=24)**	**HAART (n=25)**	**P-value**
maternal age	25.5 (21.9, 29.3)	26.0 (24.0, 29.0)	0.989
parity	2 (0.5, 2)	1 (1, 2)	0.542
primiparous	18 (75%)	20 (80%)	0.742
Antenatal Characteristics			
premature labor	4 (16.7%)	1 (4%)	0.190
baseline CD4 count	446 (268, 660)	280 (248, 421)	0.049
plasma HIV RNA	4.6 (4.4, 5.4)	4.9 (4.4, 5.0)	0.984
Postpartum Characteristics			
mastitis	4 (16.7%)	3 (12%)	0.702
breast abscess	1 (4.2%)	2 (8%)	>0.999
Supplements			
iron	7 (29.2%)	17 (68%)	0.01
folate	8 (33.3%)	16 (64%)	0.047
vitamin A	3 (12.5%)	1 (4%)	0.349
multi vitamins	1 (4.2%)	1 (4.2%)	>0.999
any antibiotic	10 (41.7%)	5 (20%)	0.128
antimalarials	2 (8.3%)	0 (0%)	0.235
Breast Milk Sample Size			
week 1	21	24	
week 2	21	22	
week 3	22	23	
week 4	18	22	

differences in both the centroid and dispersion (Figure 3.2B). Finally, the samples were stratified by group to assess the time effect. Figure 3.3 shows that the ellipses of 4 weekly clusters largely overlapped in the control group, while samples were more widely spread and had slight separation in the HAART group.

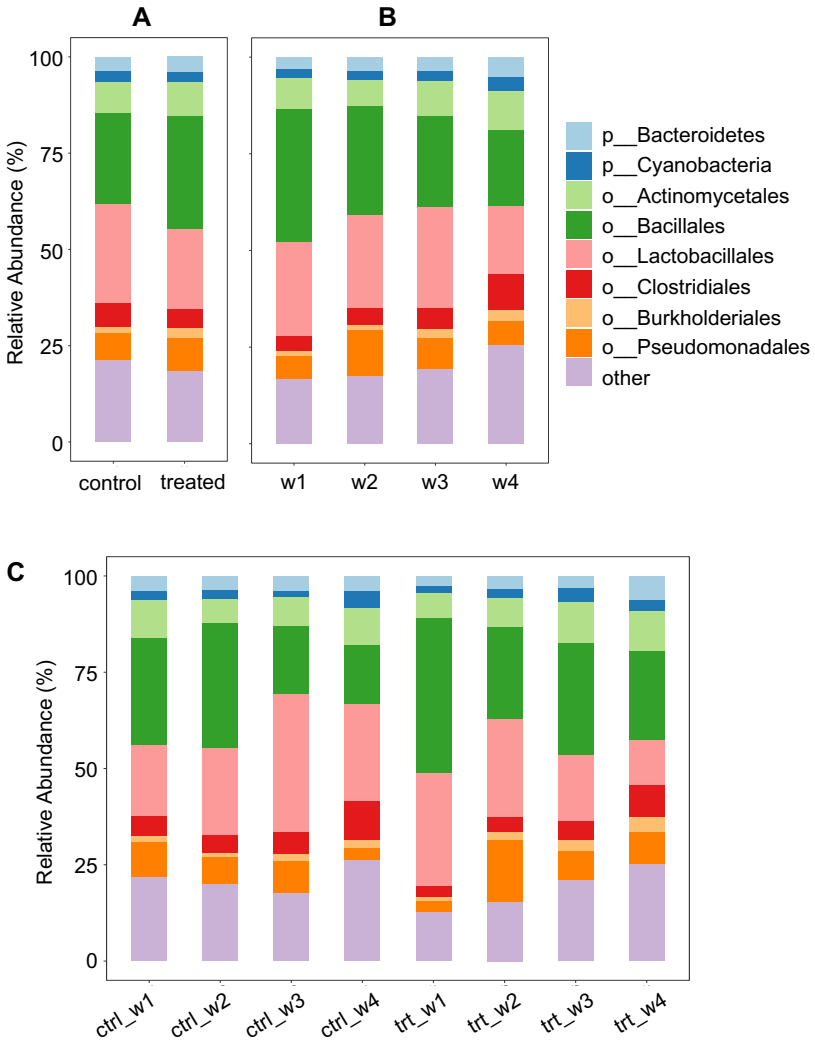

Figure 3.1: Stacked bar charts of the mean relative abundance of the 8 most abundant taxa across individuals in (A) groups (ctrl vs. trt), (B) time points (week 1/2/3/4), (C) groups at each week.

Similar analyses were performed on unweighted UniFrac and Jaccard distance measures. However, the first two principal components only explained a small amount of variance and did not capture the dissimilarities among samples (15.5% for UniFrac, 4.4% for Jaccard. Figure A.1). Weighted UniFrac distance reduces the effects of low abundance species and thus was used for analyses.

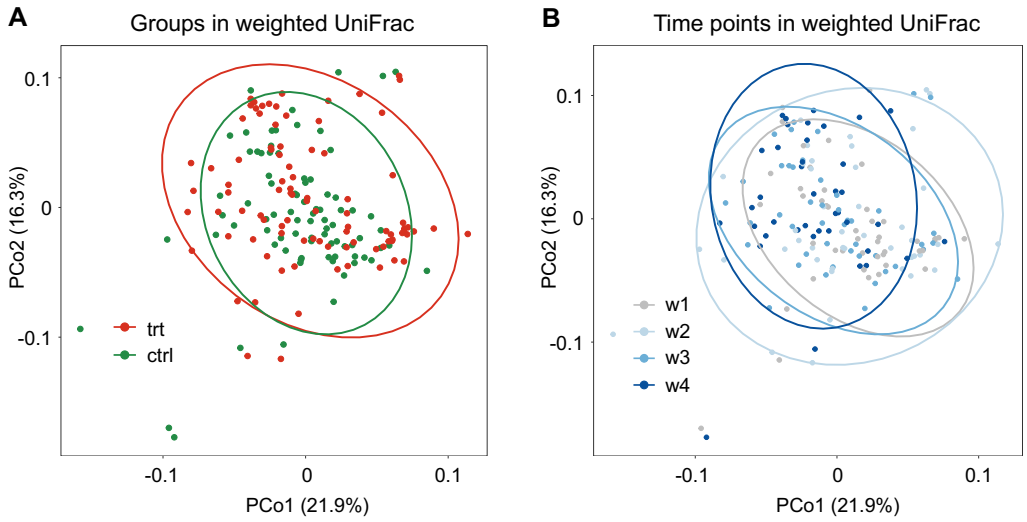

Figure 3.2: Beta diversity between group (A) and across time points (B). PCoA plots of β-diversity by weighted UniFrac distance matrix showing the sample points with 95% confidence interval ellipses.

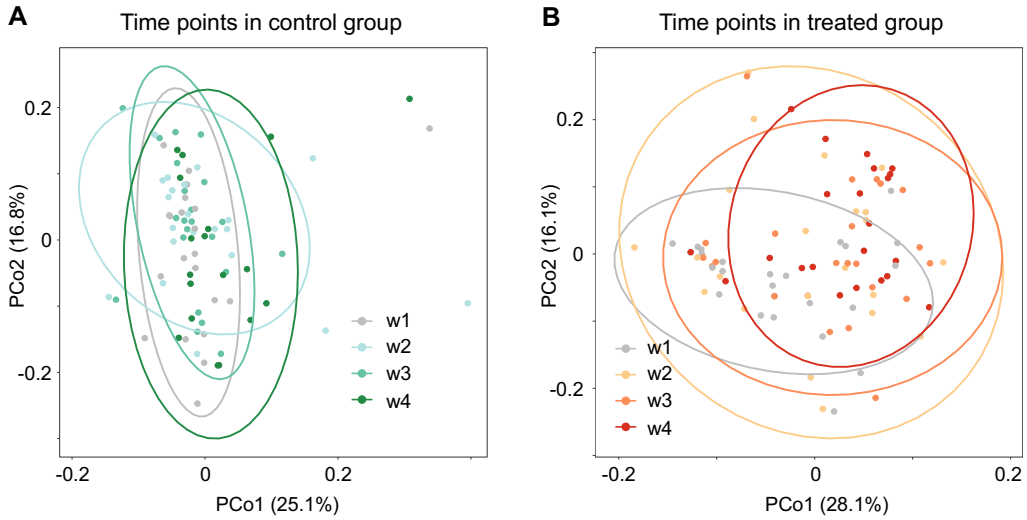

Figure 3.3: Beta diversity comparison between time points, separated by group. PCoA plots of β-diversity by weighted UniFrac distance showing the sample points with 95% confidence interval ellipses. The 4 colors represent 4 time points within (A) control group and (B) treated group.

3.2 Significant Changes in Beta Diversity Over Time

This analysis assessed changes over time in β-diversity during the first month postpartum. Instead of including id in the model, the random effect of subject was accounted for through random shuffling of observations within each subject, and generating an empirical null distribution to estimate a p-value. Based on PERMANOVA results in weighted UniFrac metrics, the centers of sample clusters by time points are statistically significantly different (p = 0.005; Table 3.2).

Table 3.2: Sample dissimilarity in weighted UniFrac distance was associated with time.

	DF	Sums of Sqs	Mean Sqs	F	Pr (F)
Group	1	0.128	0.128	1.71	
Time Point	3	0.394	0.131	1.76	0.005 **
Residual	168	12.541	0.075		
Total	172	13.063			

Table 3.3: Pairwise comparisons indicated that β-diversity in weighted UniFrac distance at week 1 was significantly different from the last time point week 4.

Pairwise comparison	P-value adjusted
W1 - W2	0.676
W1 - W3	0.051
W1 - W4	0.006 **
W2 - W3	0.641
W2 - W4	0.105
W3 - W4	0.105

Furthermore, to identify where the difference came from, pairwise post-hoc tests with Benjamini-Hochberg p-value adjustment was done. Only the comparison for week 1 versus week 4 (p = 0.006) was statistically significant (Table 3.3). This difference in microbiome composition was apparent in the visualization in the PCoA plots, with the least overlap between week 1 and week 4 in the treated group (Figure 3.3B).

3.3 Microbiome diversity not altered by HAART

An additional question of interest was the effect of HAART on β-diversity. PERMANOVA analysis was conducted with permutation of group labels for each subject. Such a permutation scheme requires an equal number of observations per cluster, therefore only 15 control and 17 treated subjects who had 4 time-point samples were included. The analysis suggested that in the first postpartum month, HAART regimen did not significantly change the breast milk microbiome diversity in HIV-infected women (p = 0.308; Table 3.4).

Sensitivity analyses assessed the robustness of results to missing data. To obtain a larger sample size, various combinations of time points were chosen and the analysis was repeated. At week 3, samples came from 22 control and 23 treated subjects and were not statistically significantly different (p = 0.107; Table A.1). Analysis on data from 3 time-points (week 1/2/3 and week 1/3/4) combinations also confirmed the absence of any treatment effect on β-diversity (p = 0.452 and 0.366, respectively; Table A.2, A.3). Overall, these outcomes were consistent, indicating that removal of missing data did not qualitatively change the result, and there was no detectable alteration on β-diversity by HAART.

Table 3.4: HAART did not significantly alter microbiome β-diversity in weighted UniFrac distance in PERMANOVA analysis with 32 subjects (n_{ctrl} = 15, n_{trt} = 17) who had samples collected all 4 weeks.

	DF	Sums of Sqs	Mean Sqs	F	Pr (>F) [1]
Group	1	0.118	0.118	1.12	0.308
Id	30	3.169	0.106		
Time Point	3	0.323	0.108		
Residual	93	6.546	0.070		
Total	127	10.157			

[1] F-statistic and p-value were obtained as described in Methods.

Chapter 4

DISCUSSION

This study analyzed the β-diversity of breast milk microbiome among HIV-infected women during the first month postpartum. Women who had been receiving HAART from 34 weeks gestation showed minimal difference from controls. However, there were statistically significant alterations in β-diversity when breast milk samples from week 1 compared to week 4 postpartum. Restricted permutation techniques were used to obtain empirical p-values accounting for random effects. This method allows us to partition variation in the space of weighted UniFrac distance with no multivariate normality assumption.

β-diversity at week 1 was found to be significantly different from week 4 postpartum. An alteration in temporal β-diversity is expected, as previous studies have shown microbiome dynamics over lactation stage [14, 6, 10]. However, despite prolonged exposure to HAART, no obvious difference in β-diversity was detected between the HAART group and the control group. These conclusions are further confirmed by GLMM-MiRKAT, a distance-based kernel association test for correlated microbiome samples. Since all the samples were collected over a relatively short period of time, HAART impact might have been overshadowed by the natural modulations in microbiome composition. Overall, differences in breast milk microbiome β-diversity is not associated with prolonged HAART but rather with time after delivery in HIV-infected women.

To handle the mixed effects model on the unbalanced longitudinal data, our analysis adopted PERMANOVA with restricted permutation. This methodology has several advantages. First, it can be used for any distance measure, such as Bray-Curtis dissimilarity, and UniFrac distance. The mixed model shown in the Appendix C only works with distance that does not need phylogenetic information (e.g., Euclidean distance). Therefore,

our method expands the scope of applications of mixed model PERMANOVA in microbiome studies. Second, strata is used to account for the random effect, and restricted permutations are implemented to obtain correct p-values. When analyzing the crossed factor (e.g. time), a balanced design is not required for permutation within strata. In our case, all subjects with at least two samples were included in the analysis of the effect of time. This enables us to take full advantage of the sample size and increase statistical power.

This proposed methodology also has several limitations. First, the permutation of strata itself requires equal sample size per strata. For example, in the analysis of the treatment effect, each person must have an equal number of observations, therefore 26% of the samples were excluded from that analysis. Although sensitivity analyses confirmed that excluded samples did not qualitatively change the results, a loss of statistical power is inevitable for such unbalanced designs. Moreover, the dataset only covers 4 time points during the first month postpartum. Our failure to detect any HAART effect could be the result of temporal changes overshadowing any effects of HAART. Further research is needed to establish any long-term effect of HAART on breast milk microbiome.

BIBLIOGRAPHY

[1] Marti J. Anderson. Permutation tests for univariate or multivariate analysis of variance and regression. *Can. J. Fish. Aquat. Sci.*, 58(3):626–639, 2001.

[2] Marti J. Anderson. Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley StatsRef Stat. Ref. Online*, pages 1–15, 2017.

[3] Alessandra Bandera, Ilaria De Benedetto, Giorgio Bozzi, and Andrea Gori. Altered gut microbiome composition in HIV infection: Causes, effects and potential intervention. *Curr. Opin. HIV AIDS*, 13(1):73–80, 2018.

[4] Steve Brewer. Doing repeated measure manova w adonis, 2014.

[5] Steve Brewer. Nested Permanova with repeated measures, 2015.

[6] Raul Cabrera-Rubio, M. Carmen Collado, Kirsi Laitinen, Seppo Salminen, Erika Isolauri, and Alex Mira. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. *Am. J. Clin. Nutr.*, 2012.

[7] Michael H. Chung, James N. Kiarie, Barbra A. Richardson, Dara A. Lehman, Julie Overbaugh, and Grace C. John-Stewart. Breast milk HIV-1 suppression and decreased transmission: A randomized trial comparing HIVNET 012 nevirapine versus short-course zidovudine. *AIDS*, 2005.

[8] Mercedes Gomez De Agüero, Stephanie C. Ganal-Vonarburg, Tobias Fuhrer, Sandra Rupp, Yasuhiro Uchimura, Hai Li, Anna Steinert, Mathias Heikenwalder, Siegfried Hapfelmeier, Uwe Sauer, Kathy D. McCoy, and Andrew J. Macpherson. The maternal microbiota drives early postnatal innate immune development. *Science (80-.).*, 2016.

[9] Leónides Fernández, Susana Langa, Virginia Martín, Antonio Maldonado, Esther Jiménez, Rocío Martín, and Juan M. Rodríguez. The human milk microbiota: Origin and potential roles in health and disease. *Pharmacol. Res.*, 2013.

[10] Carlos Gomez-Gallego, Izaskun Garcia-Mantrana, Seppo Salminen, and María Carmen Collado. The human milk microbiome and factors influencing its composition and activity. *Semin. Fetal Neonatal Med.*, 2016.

[11] Emmanuel Gonzalez, Nicholas J.B. Brereton, Chen Li, Lilian Lopez Leyva, Noel W. Solomons, Luis B. Agellon, Marilyn E. Scott, and Kristine G. Koski. Distinct Changes Occur in the Human Breast Milk Microbiome Between Early and Established Lactation in Breastfeeding Guatemalan Mothers. *Front. Microbiol.*, 2021.

[12] Raquel González, Inácio Mandomando, Victoria Fumadó, Charfudin Sacoor, Eusébio Macete, Pedro L. Alonso, and Clara Menendez. Breast milk and gut microbiota in African mothers and infants from an area of high HIV prevalence. *PLoS One*, 8(11), 2013.

[13] Bethany M. Henrick, Xiao Dan Yao, Laila Nasser, Ava Roozrogousheh, and Kenneth L. Rosenthal. Breastfeeding behaviors and the innate immune system of human milk: Working together to protect infants against inflammation, HIV-1, and other infections. *Front. Immunol.*, 2017.

[14] Katherine M. Hunt, James A. Foster, Larry J. Forney, Ursel M.E. Schütte, Daniel L. Beck, Zaid Abdo, Lawrence K. Fox, Janet E. Williams, Michelle K. McGuire, and Mark A. McGuire. Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS One*, 2011.

[15] Ted Jost, Christophe Lacroix, Christian Braegger, and Christophe Chassard. Assessment of bacterial diversity in breast milk using culture-dependent and culture-independent approaches. *Br. J. Nutr.*, 2013.

[16] Wei Li A. Koay, Lilly V. Siems, and Deborah Persaud. The microbiome and HIV persistence: Implications for viral remission and cure. *Curr. Opin. HIV AIDS*, 2018.

[17] Hyunwook Koh, Yutong Li, Xiang Zhan, Jun Chen, and Ni Zhao. A distance-based kernel association test based on the generalized linear mixed model for correlated microbiome studies. *Front. Genet.*, 10(MAY):1–14, 2019.

[18] Rabia Maqsood, Joshua B. Reus, Lily I. Wu, LaRinda A. Holland, Ruth Nduati, Dorothy Mbori-Ngacha, Elizabeth Maleche-Obimbo, Emily R. Begnel, Soren Gantt, Ednah Ojee, Dalton Wamalwa, Grace John-Stewart, Jennifer Slyker, Dara A. Lehman, and Efrem S. Lim. Breast Milk Virome and Bacterial Microbiome Resilience in Kenyan Women Living with HIV. *mSystems*, 6(2):1–13, 2021.

[19] Chung M.H., Kiarie J.N., Richardson B.A., Lehman D.A., Overbaugh J., Kinuthia J., Njiri F., and John-Stewart G.C. Highly active antiretroviral therapy versus zidovudine/nevirapine effects on early breast milk HIV type-1 RNA: A phase II randomized clinical trial. *Antivir. Ther.*, 2008.

[20] Pia S. Pannaraj, Fan Li, Chiara Cerini, Jeffrey M. Bender, Shangxin Yang, Adrienne Rollie, Helty Adisetiyo, Sara Zabih, Pamela J. Lincez, Kyle Bittinger, Aubrey Bailey,

Frederic D. Bushman, John W. Sleasman, and Grace M. Aldrovandi. Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. *JAMA Pediatr.*, 2017.

[21] Yang Sun, Yingfei Ma, Ping Lin, Yi Wei Tang, Liying Yang, Yinzhong Shen, Renfan Zhang, Li Liu, Jun Cheng, Jiashen Shao, Tangkai Qi, Yan Tang, Rentian Cai, Liqian Guan, Bin Luo, Meiyan Sun, Ben Li, Zhiheng Pei, and Hongzhou Lu. Fecal bacterial microbiome diversity in chronic HIV-infected patients in China. *Emerg. Microbes Infect.*, 2016.

[22] Yirui Xie, Jia Sun, Li Wei, Haiyin Jiang, Caiqin Hu, Jiezuan Yang, Ying Huang, Bing Ruan, and Biao Zhu. Altered gut microbiota correlate with different immune responses to HAART in HIV-infected individuals. *BMC Microbiol.*, 2021.

[23] Alexander S. Zevin, Lyle McKinnon, Adam Burgener, and Nichole R. Klatt. Microbial translocation and microbiome dysbiosis in HIV-associated immune activation. *Curr. Opin. HIV AIDS*, 2016.

[24] Xiang Zhan, Lingzhou Xue, Haotian Zheng, Anna Plantinga, Michael C. Wu, Daniel J. Schaid, Ni Zhao, and Jun Chen. A small-sample kernel association test for correlated data with application to microbiome association studies. *Genet. Epidemiol.*, 42(8):772–782, 2018.

[25] Gili Zilberman-Schapira, Niv Zmora, Shlomik Itav, Stavros Bashiardes, Hila Elinav, and Eran Elinav. The gut microbiome in human immunodeficiency virus infection. *BMC Med.*, 2016.

Appendix A

SUPPLEMENTARY FIGURES AND TABLES

Table A.1: Cross-sectional PERMANOVA at week 3 indicated no significant effect of HAART on β-diversity.

	DF	Sums of Sqs	Mean Sqs	F	Pr (>F)
Group	1	0.108	0.118	1.56	0.118
Residual	43	2.963	0.069		
Total	44	3.071			

Table A.2: No effect of HAART in PERMANOVA analysis with 38 subjects (n_{ctrl} = 18, n_{trt} = 20) who had samples collected for weeks 1/2/3.

	DF	Sums of Sqs	Mean Sqs	F	Pr (>F) [1]
Group	1	0.092	0.092	0.91	0.452
Id	36	3.611	0.100		
Time Point	2	0.117	0.059		
Residual	74	4.819	0.065		
Total	113	8.639			

Table A.3: No effect of HAART in PERMANOVA analysis with 36 subjects (n_{ctrl} = 17, n_{trt} = 19) who had samples collected for weeks 1/3/4.

	DF	Sums of Sqs	Mean Sqs	F	Pr (>F) [1]
Group	1	0.098	0.098	1.07	0.366
Id	34	3.131	0.092		
Time Point	2	0.295	0.148		
Residual	70	4.326	0.062		
Total	107	8.639			

[1] F-statistic and p-value were obtained as described in Methods.

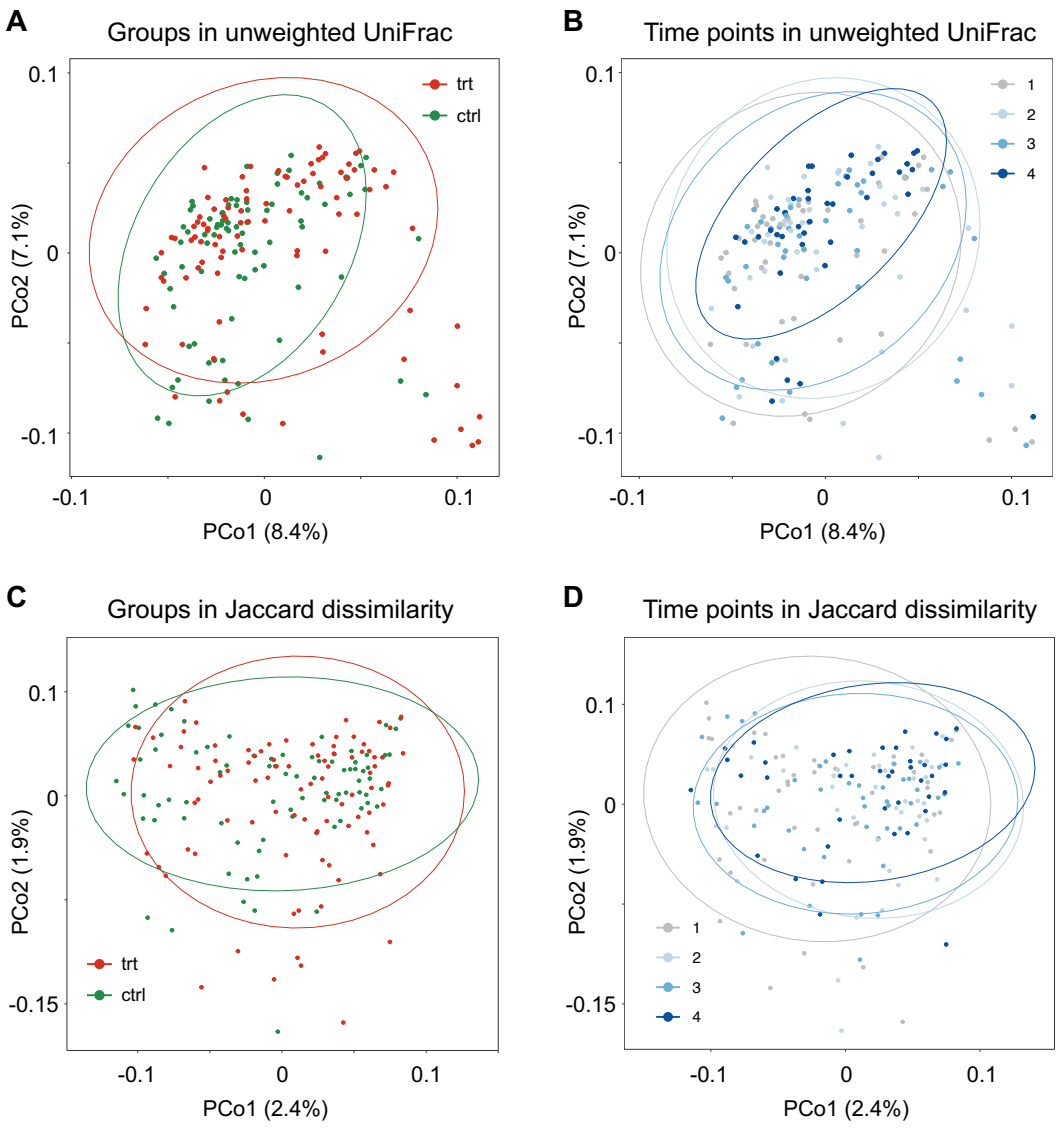

Figure A.1: β-diversity comparison between groups, time points measured by unweighted UniFrac (A, B) and Jaccard dissimilarity (C, D). PCoA plots showing the sample points with 95% confidence interval ellipses.

Appendix B

SAMPLE CODE OF PERMANOVA

B.1 Time effect

```
# dat is the meta dataset for 173 samples
# it contains id (subject id), grp (group label), tp (time point)
# WUDM is a matrix representing the weighted unifrac distances
# wudm is a dist object (lower triangular matrix), wudm = as.dist(WUDM)

perm <- how(nperm = 999, within = Within(type = "free"),
        plots = Plots(strata = dat$id))
adonis2(wudm ~ grp + tp, data = dat, permutations = perm, by = "margin")

# post-hoc
pair_dat <- function(dist_mat, info, t1, t2) {
  ind <- info$tp == t1 | info$tp == t2
  pair_info <- info[ind, ]
  pair_mat <- as.dist(dist_mat[ind, ind])
  return(list(dm = pair_mat, pi = pair_info))
}

# pairwise time points
pair_tp <- combn(unique(dat$tp), 2)
pval <- c()
for (i in 1:ncol(pair_tp)) {
  x <- pair_dat(WUDM, dat, pair_tp[1,i], pair_tp[2,i])
  perm <- how(nperm = 999, within = Within(type = "free"),
          plots = Plots(strata = x$pi$id))
  res <- adonis2(x$dm ~ grp + tp, data = x$pi, permutations = perm, by = "margin")
  p <- res$`Pr(>F)`[2]
  pval <- c(pval, p)
}

# p-value adjustment
padj <- p.adjust(pval, method = "BH")
rbind(pair_tp, padj)
```

B.2 Treatment effect

```
# select subjects with 4 time point samples: dat_1234, wudm_1234
# SS table
x <- adonis(wudm_1234 ~ grp + id + tp, data = dat_1234, by = "terms",
       permutations = 2)
# calculate the correct F-statistic for group
f.grp <- x$aov.tab$F.Model[1] / x$aov.tab$F.Model[2]

# manually permute
n <- 999; f <- c(); dt <- dat_1234
perm <- how(within = Within(type = "none"),
       plots = Plots(strata = dat_1234$id, type = "free"))

for (i in 1:n) {
  dt$grp <- shuffle(dt$grp, control = perm)
  x <- adonis(wudm_1234 ~ grp + id + tp, data = dt, by = "terms",
         permutations = 2)
  fp <- x$aov.tab$F.Model[1] / x$aov.tab$F.Model[2]
  f <- c(f, fp)
}

# empirical p-value
sum(f>f.grp, 1)/(n+1)
```

Appendix C

COMPARISON OF MIXED MODEL AND RESTRICTED PERMUTATIONS

To validate the results of our restricted permutation methodology, we compared it to a well-established method of analyzing mixed effects models in PERMANOVA. This method is appropriate for balanced designs and distance measures that do not rely on phylogenetic information. Therefore, a subset of 15 control and 17 treated subjects with all 4 time point sampled was used. Group and time point were modeled as fixed effects, and subject id as a random effect. Euclidean distance instead of a β-diversity measure was used as the response.

First, a full model was fit, with group, id and time as fixed effects (Table C.1). The F-statistic for time was correctly calculated as the ratio of MS(time) to MS(residual), and so was the permuted p-value. Time was not found to be statistically significant ($F = 1.229, p = 0.187$). However, the F-statistic of group should be calculated as the ratio of MS(group) and MS(id) to account for the random effect. Permutations also needed to be performed accordingly. Therefore, 4 observations for each subject were averaged into one centroid, and the distance between centroids was then fit into a second model with group only (Table C.2). No statistically significant effect of treatment was detected ($F = 1.427,\ p = 0.183$).

Next, the analyses of dissimilarity in Euclidean distance were repeated using our novel method with restricted permutations. When testing for the effect of time, a model with group and time was fit, and the null distribution of F(time) was generated by permuting observations within each subject id (Table C.3). Similar to the previous method, time was not found to be statistically significant ($F = 1.056,\ p = 0.182$). A second model with group, id, and time was fit using sequential sums of squares. The F-statistic of

group was calculated as the ratio of MS group to MS id, and the p-value was obtained through permutation of group labels among ids (Table C.4. $F = 1.427,\ p = 0.19$).

When comparing the mixed model with the restricted permutations, both approaches indicate the absence of time and treatment effect in Euclidean distance, and the empirical p-values are fairly similar. Overall, these results confirm the validity of the restricted permutation method.

Table C.1: Evaluation of time effect using a full model with group and subject id in Euclidean distance. Group and time are fixed effects.

	DF	Sums of Sqs	Mean Sqs	F	Pr (>F)
Group	1	2999	2999		
Id	30	63035	2101		
Time Point	3	4636	1545	1.229	0.187
Residual	93	116943	1258		
Total	127	187613			

Table C.2: To account for the random effect of subject, a second model was fit with time-averaged observations of each person.

	DF	Sums of Sqs	Mean Sqs	F	Pr (>F)
Group	1	750	750	1.427	0.183
Residual	30	15759	525		
Total	31	16509			

Table C.3: Repeat PERMANOVA analysis in Euclidean distance with restricted permutations. For time effect, a model with group and time was fit, and p-value was obtained through permutation of time.

	DF	Sums of Sqs	Mean Sqs	F	Pr (>F)
Group	1	2999	2999		
Time Point	3	4636	1545	1.056	0.182
Residual	123	179978	1463		
Total	127	187613			

Table C.4: To evaluate the HAART effect, a model with group, id, and time was fit. F statistic was calculated as the ratio of MS group to MS id, and p-value was obtained through permutation of group labels between ids.

	DF	Sums of Sqs	Mean Sqs	F	Pr (>F)
Group	1	2999	2999	2999/2101 = 1.427	0.19
Id	30	63035	2101		
Time Point	3	4636	1545		
Residual	93	116943	1258		
Total	127	187613			

Appendix D

GLMM-MIRKAT

To address correlated microbiomes in family-based or longitudinal studies, Koh et al. proposed a distance-based kernel association test based on a generalized linear mixed model (GLMM-MiRKAT) [17]. Compared to previous methods, GLMM-MiRKAT is a community-level association test and can handle non-Gaussian traits. Because GLMM-MiRKAT can be sensitive to the choice of distance measure, [17] adopted a data-driven method (aGLMM-MiRKAT) and considered multiple distance metrics (Jaccard dissimilarity, Bray-Curtis dissimilarity, UniFrac distances). With the optimal distance measure, aGLMM-MiRKAT can achieve robustly high power based on simulation experiments.

As specified below, a GLMM-MiRKAT model with random intercepts for the group was fit first, adjusting for time point (Table D.1). All the five distance measures suggested no obvious effect of HAART on the breast milk microbiome composition (Table D.2). This matches with the visual inspection of the overlapping ellipses between control and treated groups in PCoA plot and the PERMANOVA with restricted permutation results (Figure 3.2A; Figure A.1).

To evaluate the temporal effect, a random-intercept model was fit for time points in continuous scale, adjusting for group. Jaccard dissimilarity ($p = 0.0004$), Bray-Curtis dissimilarity ($p = 0.0002$), generalized UniFrac distance ($p = 0.0002$), and weighted UniFrac distance ($p = 0.0004$) all indicate a strong association between microbial composition and time point, while unweighted UniFrac distance did not (Table D.2). Again, these GLMM-MiRKAT results are consistent with our previous conclusion based on PERMANOVA with restricted permutation that time postpartum was associated with statistically significant differences in breast milk microbiome composition in HIV-infected women.

Table D.1: Model details for assessing the effects of treatment and time, respectively. Column y refers to the trait of interest, and covariate includes traits for adjustment. These two models assumed random intercepts.

	y	covariate	time.pt	mod	random slope
Model 1	Group	-	Time point	Binomial	F
Model 2	Time point	Group	-	Gaussian	F

Table D.2: Summary of the results using different distance measures. aGLMM-MiRKAT column represents the p-value obtained from the optimal choice of metric.

	Jaccard	Bray-Curtis	Unweighted UniFrac	Generalized UniFrac	Weighted UniFrac	aGLMM-MiRKAT
Model 1	0.18	0.31	0.14	0.08	0.24	0.23
Model 2	0.0004	0.0002	0.2795	0.0002	0.0004	0.0002