

The oral microbiome and breast cancer and non-malignant breast disease, and its relationship with the fecal microbiome in the Ghana Breast Health Study

Zeni Wu¹, Doratha A. Byrd^{1,2}, Yunhu Wan¹, Daniel Ansong³, Joe-Nat Clegg-Lamptey⁴, Beatrice Wiafe-Addai⁵, Lawrence Edusei⁴, Ernest Adjei³, Nicholas Titiloye³, Florence Dedey⁴, Francis Aitpillah³, Joseph Oppong³, Verna Vanderpuye⁴, Ernest Osei-Bonsu³, Casey L. Dagnall^{1,6}, Kristine Jones^{1,6}, Amy Hutchinson^{1,6}, Belynda D. Hicks^{1,6}, Thomas U. Ahearn¹, Jianxin Shi¹, Rob Knight⁷, Richard Biritwum⁸, Joel Yarney⁴, Seth Wiafe⁹, Baffour Awuah³, Kofi Nyarko⁸, Jonine D. Figueroa^{1,10}, Rashmi Sinha¹, Montserrat Garcia-Closas¹, Louise A. Brinton¹, Emily Vogtmann¹

Table of Contents

Table S1 Associations of genus presence/absence with breast cancer and non-malignant breast disease in the Ghana Breast Health Study (provided as separate excel files)

Table S2 Associations of genus relative abundance with breast cancer and non-malignant breast disease in the Ghana Breast Health Study (provided as separate excel files)

Table S3 Sensitivity analyses of association between alpha-diversity with breast cancer in participants did not use antibiotics within the last 30 days

Table S4 Sensitivity analyses of association between beta-diversity with breast cancer in participants did not use antibiotics within the last 30 days

Figure S1 Rarefaction curve for Observed Amplicon Sequence Variants (ASVs) in oral samples.

Figure S2 Principal coordinate analysis (PCoA) of the first five principal components of artificial community, rodogut, and oral samples, based on Bray Curtis, Jaccard, and Unweighted UniFrac and Weighted UniFrac distances.

Figure S3 Principal coordinate analysis (PCoA) of the first five principal components of breast cancer (red), non-malignant breast disease, and controls, based on Bray Curtis, Jaccard, and Unweighted UniFrac and Weighted UniFrac distances.

Figure S4 Genus-level presence correlations between the oral and fecal microbiome within breast cancer cases, non-malignant cases, and controls.

Table S3 Sensitivity analyses of association between alpha-diversity with breast cancer in participants did not use antibiotics within the last 30 days

	N	Case	OR (95% CI)	P-value
Alpha diversity				
Observed ASVs				
Continuous (10 ASVs per unit)	572	327	0.87 (0.84-0.91)	1.40E-10
Q1	217	82	Reference	
Q2	123	79	0.38 (0.23-0.64)	0.0002
Q3	115	79	0.37 (0.21-0.62)	0.0002
Q4	117	87	0.26 (0.15-0.44)	1.18E-06
Faith's PD				
Continuous	572	327	0.76 (0.69-0.82)	5.78E-11
Q1	222	82	Reference	
Q2	118	79	0.37 (0.22-0.61)	0.0001
Q3	114	79	0.40 (0.23-0.68)	0.0007
Q4	118	87	0.27 (0.16-0.47)	3.02E-06
Shannon index				
Continuous	572	327	0.57 (0.44-0.74)	2.74E-05
Q1	196	82	Reference	
Q2	135	79	0.63 (0.38-1.04)	0.0696
Q3	110	79	0.39 (0.22-0.69)	0.0012
Q4	131	87	0.54 (0.32-0.90)	0.0181

Adjusted for age (continuous), study center (Komfo Anoyke Teaching Hospital, Korle Bu Teaching Hospital, or Peace and Love Hospital), body mass index (BMI, categorical), education (junior secondary school or lower, senior secondary school/some college or technical school or more, other or unknown), family history of cancer (yes, no, unknown), number of full-term pregnancies (0, 1-2, 3-4, 5+ pregnancies), alcohol drinking (yes, no, unknown), and current hormonal contraceptive use (yes, no).

Abbreviations: ASVs, amplicon sequence variants; CI, confidence interval; OR, odds ratio; PD, Phylogenetic Diversity

Table S4 Sensitivity analyses of association between beta-diversity with breast cancer in participants did not use antibiotics within the last 30 days

	Breast cancer cases vs. controls
Bray-Curtis	<0.0001
Jaccard	<0.0001
Unweighted UniFrac	<0.0001
Weighted UniFrac	0.0002

Adjusted for age (continuous), study center (Komfo Anoyke Teaching Hospital, Korle Bu Teaching Hospital, or Peace and Love Hospital), body mass index (BMI, categorical), education (junior secondary school or lower, senior secondary school/some college or technical school or more, other or unknown), family history of cancer (yes, no, unknown), number of full-term pregnancies (0, 1-2, 3-4, 5+ pregnancies), alcohol drinking (yes, no, unknown), and current hormonal contraceptive use (yes, no).

Bonferroni adjusted p-value significance threshold for beta-diversity comparisons was $p < 0.05/4 = 0.0125$

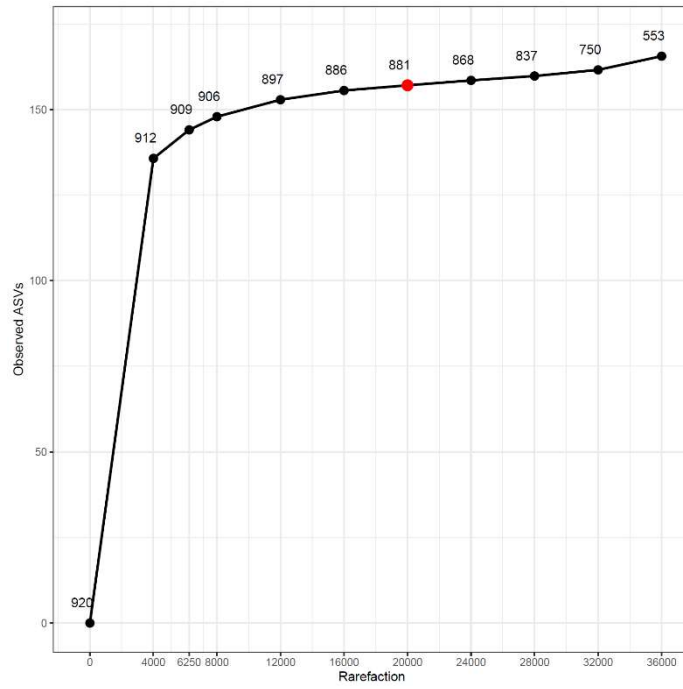


Figure S1 Rarefaction curve for Observed Amplicon Sequence Variants (ASVs) in oral samples. The red dot is the selected rarefaction level. Each number indicates the number of samples at that specific rarefaction cutpoint.

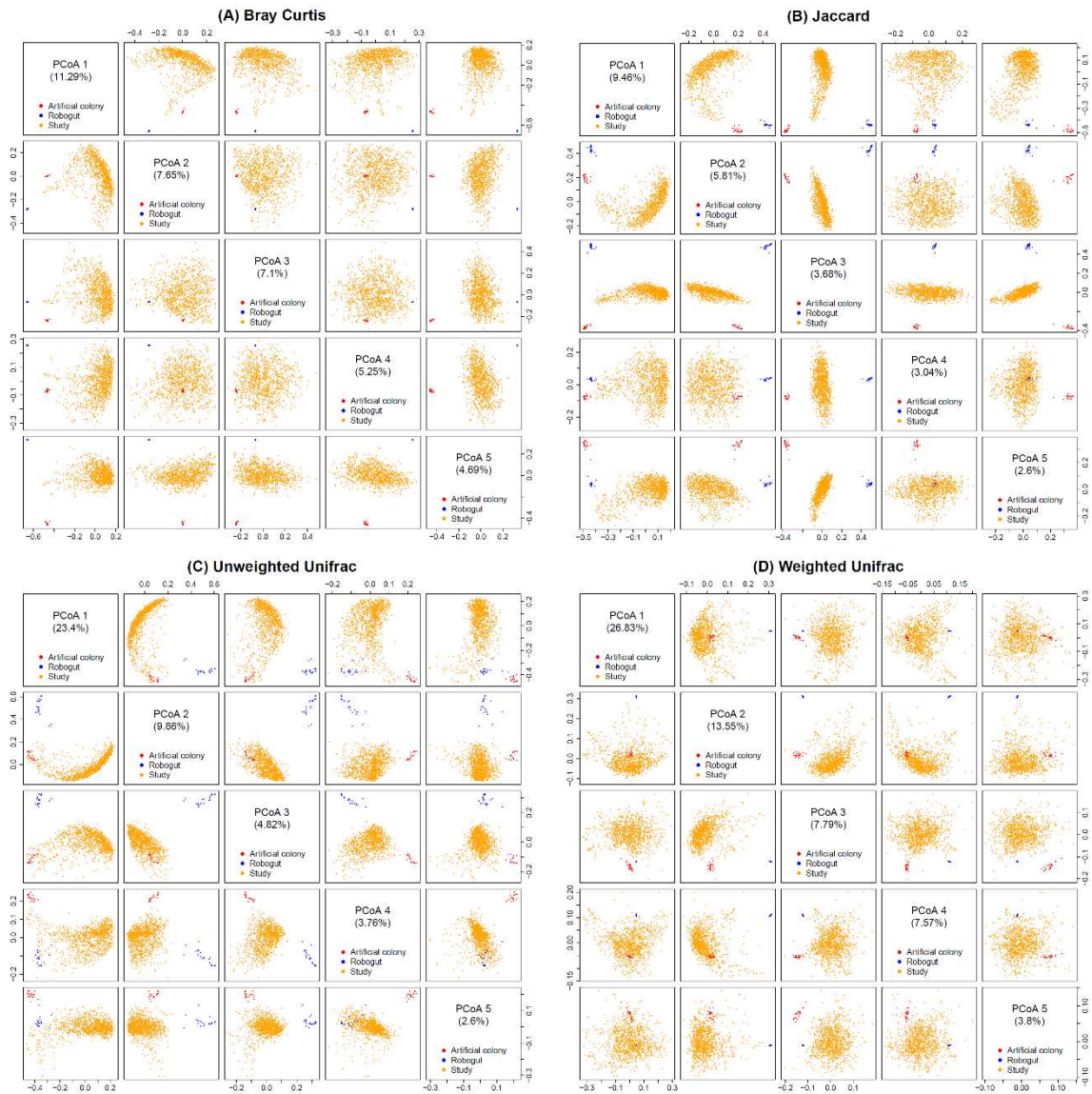


Figure S2 Principal coordinate analysis (PCoA) of the first five principal components of artificial community (red), robogut (blue), and oral samples (orange), based on Bray Curtis (A), Jaccard (B), and Unweighted UniFrac (C) and Weighted UniFrac distances (D).

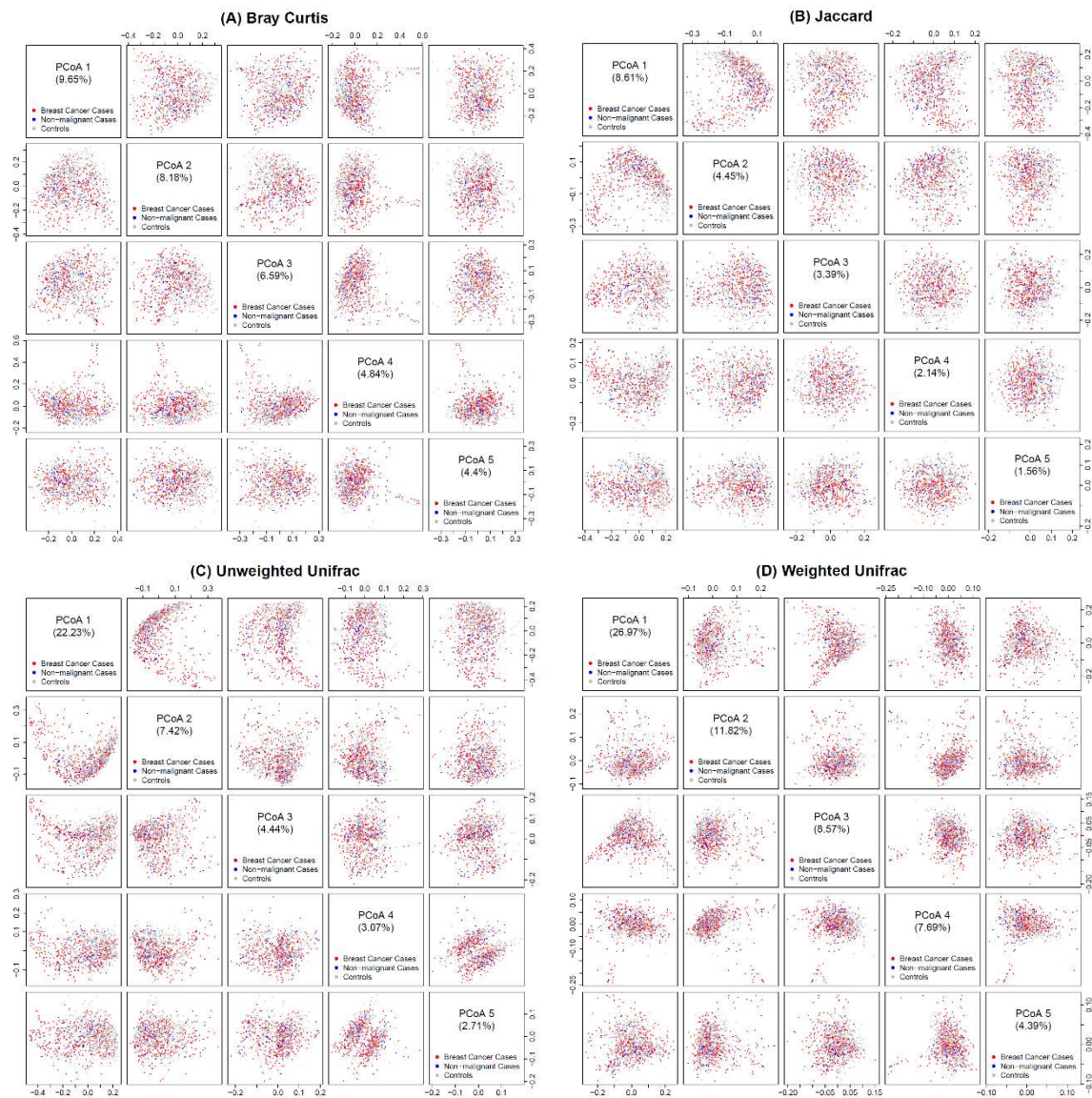


Figure S3 Principal coordinate analysis (PCoA) of the first five principal components of breast cancer (red), non-malignant breast disease (blue), and controls (grey), based on Bray Curtis (A), Jaccard (B), and Unweighted UniFrac (C) and Weighted UniFrac distances (D).

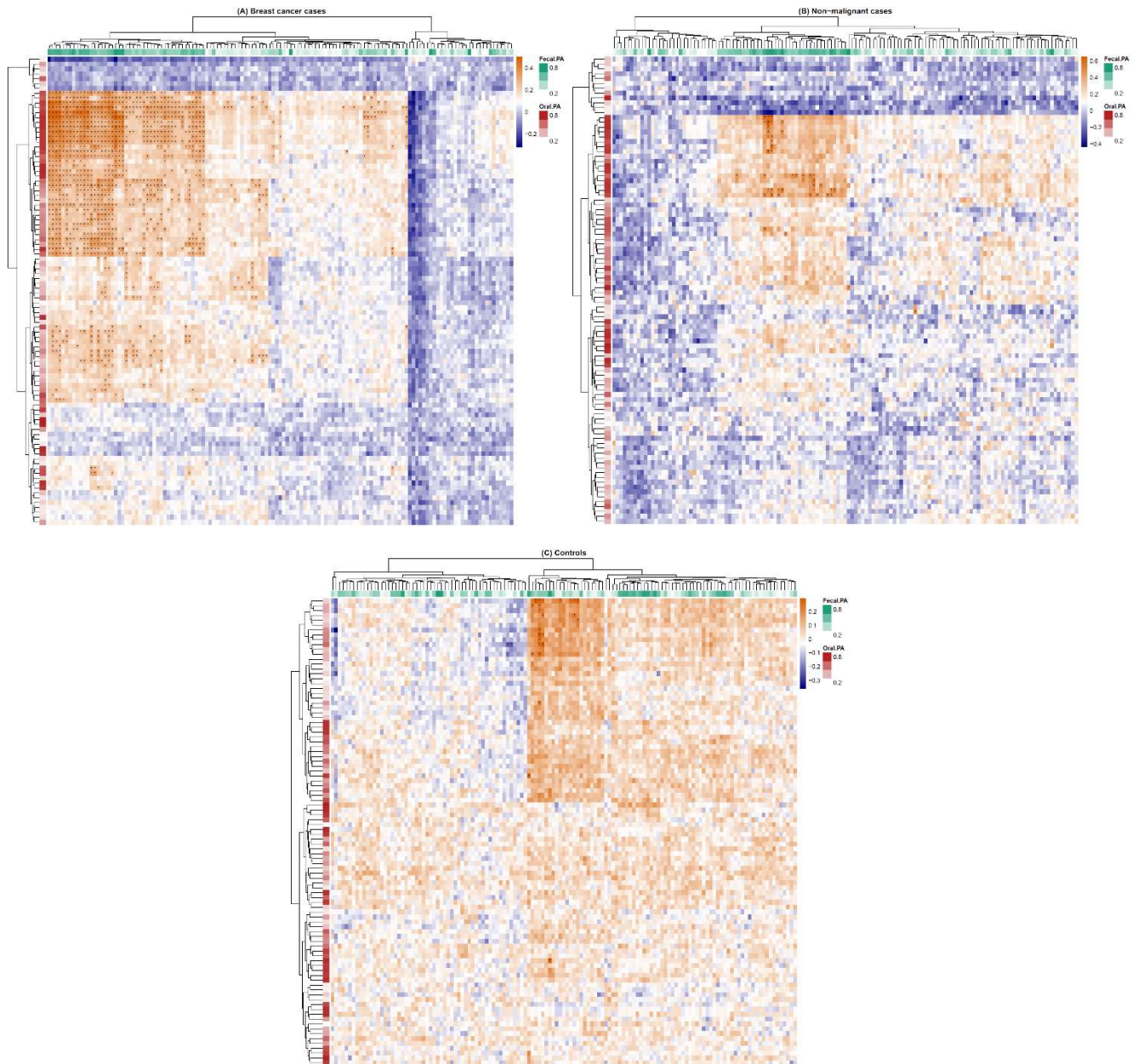


Figure S4 Genus-level presence correlations between the oral and fecal microbiome within breast cancer cases (A), non-malignant cases (B), and controls (C). Taxa were present in at least 10% of each group. Color of the cells (orange to blue) indicates the scale of correlation coefficients, and the star (*) in the cell indicates that the corresponding p-value was below the Bonferroni adjusted threshold $p < 0.05/(133 \times 96) = 23.9E-06$. Fecal.PA, presence of fecal taxa, Oral.PA, presence of oral taxa.