WEB APPENDIX

Power of detecting associations with repeated microbiome measurements

For each subject, let X_i denote a metric for the “usual” microbiome at the specific body site and $y_i = 0/1$ denote the binary outcome. Suppose that $X_i \sim N(0,1)$ after normalization. We assume a logistic regression model to link the outcome with the usual microbiome: $\log(y_i = 1|X_i) = \alpha + \beta X_i$.

For each subject, we assume K repeated microbiome observations $x_{ij} = X_i + \varepsilon_{ij}$ with $\varepsilon_{ij} \sim N(0, \sigma^2)$. By definition, the intraclass correlation coefficient (ICC) for the microbiome metric is given by

$$ICC = \frac{Var(X_i)}{Var(X_i) + Var(\varepsilon_{ij})} = \frac{1}{1+\sigma^2}.$$

For each subject, we use $\bar{X}_i = \frac{\sum_{j=1}^{K} x_{ij}}{K}$ as the estimate for the “usual” microbiome metric. Then,

$$Var(\bar{X}_i) = 1 + \frac{\sigma^2}{K} = 1 + \frac{\frac{1}{ICC} - 1}{K}.$$

We assume N subjects in a case-control study with $N\phi$ cases and $N(1-\phi)$ controls. We test the null hypothesis $H_0: \beta = 0$ using the following score statistic based on $\{\bar{X}_i, y_i\}$ pairs:

$$Z = \frac{\frac{\sum_{i:y_i=1} \bar{X}_i}{N\phi} - \frac{\sum_{i:y_i=0} \bar{X}_i}{N(1-\phi)}}{\sqrt{\left(\frac{1}{N\phi} + \frac{1}{N(1-\phi)}\right) Var(\bar{X}_i)}}.$$

Let $\mu_+(\alpha, \beta) = E(X_i|y_i = 1)$ be expectation of microbiome metric in the case group and is a function of β. Similarly, $\mu_-(\alpha, \beta) = E(X_i|y_i = 0)$ be the expectation of microbiome metric in the control group. Both $\mu_+(\alpha, \beta)$ and $\mu_-(\alpha, \beta)$ can be numerically derived by Monte Carlo methods. Then, $E(\bar{X}_i|y_i = 1) = \mu_+(\alpha, \beta)$ and $E(\bar{X}_i|y_i = 0) = \mu_-(\alpha, \beta)$. The non-centrality parameter (NCP) of the score statistic is derived as

$$\xi = E_{\alpha,\beta}(Z) = \frac{\mu_+(\alpha,\beta) - \mu_-(\alpha,\beta)}{\sqrt{\frac{1}{N}\left(\frac{1}{\phi} + \frac{1}{1-\phi}\right)\left(1 + \frac{\frac{1}{ICC} - 1}{K}\right)}},$$

which leads to the formula for the required sample size:

$$N = \frac{\xi^2\left(\frac{1}{\phi} + \frac{1}{1-\phi}\right)\left(1 + \frac{\frac{1}{ICC} - 1}{K}\right)}{[\mu_+(\alpha,\beta) - \mu_-(\alpha,\beta)]^2}.$$

Derivation of the relationship between odds ratio and logistic regression coefficient β

Let $d_0 = \Phi^{-1}(25\%) = -0.6745$ and $d_1 = \Phi(75\%) = 0.6745$be the 25% and 75% quantiles for the standard normal distribution. We derive the odds ratio (OR) comparing the risk between the top 25% quantile and the bottom 25% quantile:

$$OR = \frac{P(y_i = 1|X_i > d_1)/P(y_i = 0|X_i > d_1)}{P(y_i = 1|X_i < d_0)/P(y_i = 0|X_i < d_0)}.$$

For rare diseases (e.g., colorectal cancer), OR is approximated by relative risk

$$RR = \frac{P(y_i = 1|X_i > d_1)}{P(y_i = 1|X_i < d_0)} = \frac{\int_{d_1}^{\infty} \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \phi(x)dx}{\int_{-\infty}^{d_0} \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \phi(x)dx}.$$

Note that for rare diseases, $\alpha \ll 0$, we have $1 + e^{\alpha+\beta x} \approx 1$. Thus,

$$OR \approx RR \approx \frac{\int_{d_1}^{\infty} e^{\beta x}\phi(x)dx}{\int_{-\infty}^{d_0} e^{\beta x}\phi(x)dx} = \frac{\Phi(\beta - d_1\,)}{\Phi(d_0 - \beta)}.$$

R-code for Power Calculations

```
# case.ratio: proportion of cases in all subjects, (0,1);
# alpha: critical value. You may choose 0.05, 0.001, ...
# n.repeats: 1,2,3 ...
# ICC: ICC values
# d0: bottom quantile
# d1: top quantile
# prevalence.code: prevalence in general population. 1--> 0.05, 2-->0.01, 3-->0.005 and 4--> 0.001
# OR is odds ratio. Can be a value or a vector
# Return number of cases to achieve 80% power.

Calculate.Sample.Size.Repeated.Samples <- function(case.ratio, alpha, ICC, d0, d1, prevalence.code, OR){

  # Please download a file "Beta.txt" from the link: https://www.dropbox.com/s/3og7p0yqjvwfazn/Beta.txt
  beta = read.table("Beta.txt");
  delta = beta[,prevalence.code+1];
  beta = beta[,1];
  threshold = qnorm(0.8)+qnorm(1-alpha/2);
  n.OR = length(OR);

  N = matrix(0, 5, n.OR);

  for(k in 1:n.OR){
      index = order(abs(pnorm(beta-d1)/pnorm(d0-beta)-OR[k]))[1];
      for(i in 1:5)
          N[i,k]=floor(case.ratio*threshold^2/delta[index]^2*(1/case.ratio+1/(1-case.ratio))*((1/ICC-1)/i+1));
  }
  return(N);
}
# Example for the sample size table in manuscript
ICC = c(0.44,0.37,0.42,0.18,0.29,0.58,0.44,0.43,0.47,0.81,0.48);
OR = c(1.5,2,2.5,3,3.5);
result = matrix(0,0,5);
for(k in 1:11){
   temp = Calculate.Sample.Size.Repeated.Samples(0.25, 0.05, ICC[k], qnorm(0.25), qnorm(0.75), 2, OR);
   result = rbind(result, temp[1:3,]);
}
write.table(result,"number.cases.1vs3.p05.txt",row.names=F,col.names=F)
```

Web Table 1. Intraclass correlation coefficients for five common phyla

		Body Site	***N***	**Actinobacteria**	**Bacteroidetes**	**Firmicutes**	**Fusobacteria**	**Proteobacteria**
CRC study	Gut	Stool	44	0.51	0.29	0.03	0.16	0.00
Costa Rica	Gut	Stool	121	0.44	0.37	0.42	0.33	0.61
	Oral	Saliva	42	0.34	0.48	0.38	0.56	0.47
HMP	Gut	Stool	107	0.00	0.34	0.39	0.19	0.21
	Oral	Saliva	94	0.16	0.41	0.29	0.42	0.48
	Oropharynx	Buccal mucosa	106	0.35	0.31	0.20	0.04	0.20
		Hard palate	101	0.00	0.44	0.47	0.48	0.24
		Keratinized gingiva	104	0.00	0.54	0.38	0.35	0.33
		Palatine tonsils	102	0.09	0.31	0.33	0.32	0.59
		Subgingival plaque	102	0.19	0.34	0.29	0.18	0.35
		Supragingival plaque	105	0.41	0.41	0.47	0.34	0.30
		Throat	103	0.07	0.30	0.22	0.14	0.50
		Tongue dorsum	106	0.44	0.39	0.51	0.51	0.50
		Oral average	103.6	0.19	0.38	0.36	0.29	0.38
	Airway	Anterior nares	78	0.24	0.00	0.10	0.02	0.07
	Skin	Left antecubital fossa	43	0.20	0.08	0.17	0.00	0.02
		Right antecubital fossa	49	0.10	0.00	0.29	0.00	0.00
		Left retroauricular crease	75	0.54	0.00	0.41	0.00	0.21
		Right retroauricular crease	83	0.61	0.00	0.47	0.10	0.46
		Skin average	62.5	0.36	0.02	0.34	0.02	0.17
	Vagina	Mid-vagina	37	0.37	0.62	0.58	0.18	0.01
		Posterior fornix	38	0.44	0.50	0.69	0.18	0.00
		Vaginal introitus	37	0.31	0.58	0.49	0.11	0.06
		Vagina average	37.3	0.37	0.57	0.59	0.15	0.02

Web Table 2. Intraclass correlation coefficients for four alpha-diversity measures

		Body site	***N***	**PD_tree†**	**Chao1**	**S_{obs}†**	**Shannon Index**
CRC study	Gut	Stool	34	0.34	0.37	0.33	0.48
Costa Rica	Gut	Stool	121	0.58	0.44	0.43	0.47
	Oral	Saliva	42	0.60	0.55	0.51	0.33
HMP	Gut	Stool	100	0.66	0.58	0.60	0.62
	Oral	Saliva	94	0.39	0.41	0.43	0.42
	Oropharynx	Buccal mucosa	101	0.26	0.29	0.24	0.06
		Hard palate	95	0.25	0.14	0.20	0.25
		Keratinized gingiva	102	0.28	0.19	0.22	0.27
		Palatine tonsils	100	0.39	0.20	0.27	0.26
		Subgingival plaque	100	0.39	0.33	0.32	0.37
		Supragingival plaque	104	0.40	0.38	0.46	0.48
		Throat	98	0.20	0.17	0.17	0.24
		Tongue dorsum	105	0.40	0.20	0.22	0.30
		Oral average	99	0.32	0.24	0.26	0.28
	Airway	Anterior nares	64	0.55	0.48	0.42	0.28
	Skin	Left antecubital fossa	14	0.49	0.30	0.40	0.44
		Right antecubital fossa	15	0.46	0.35	0.43	0.45
		Left retroauricular crease	63	0.61	0.65	0.63	0.61
		Right retroauricular crease	71	0.39	0.45	0.47	0.51
		Skin average	41	0.49	0.44	0.48	0.50
	Vagina	Mid-vagina	37	0.56	0.28	0.37	0.49
		Posterior fornix	38	0.54	0.34	0.31	0.54
		Vaginal introitus	37	0.54	0.27	0.47	0.49
		Vagina average	37	0.55	0.29	0.38	0.51

† PD_tree - phylogenetic distance whole tree; S_{obs} - the number of observed species or operational taxonomic units (OTU).

Web Table 3. Intraclass correlation coefficients for unweighted Unifrac distance*

		Body Site	*N*	PCoA1-5 Average	PCoA 1	PCoA2	PCoA3	PCoA4	PCoA5
CRC study	Gut	Stool	34	0.61	0.84	0.45	0.70	0.74	0.31
Costa Rica	Gut	Stool	121	0.68	0.81	0.72	0.58	0.63	0.65
	Oral	Saliva	42	0.73	0.76	0.67	0.79	0.72	0.72
HMP	Gut	Stool	100	0.50	0.76	0.47	0.35	0.45	0.47
	Oral	Saliva	86	0.26	0.54	0.34	0.11	0.10	0.21
	Oropharynx	Buccal mucosa	101	0.11	0.09	0.17	0.02	0.21	0.08
		Hard palate	95	0.31	0.44	0.56	0.15	0.21	0.18
		Keratinized gingiva	102	0.30	0.61	0.45	0.05	0.25	0.14
		Palatine tonsils	100	0.24	0.35	0.46	0.26	0.05	0.07
		Subgingival plaque	100	0.31	0.47	0.28	0.33	0.33	0.14
		Supragingival plaque	104	0.26	0.45	0.12	0.30	0.42	0.00
		Throat	98	0.23	0.38	0.29	0.19	0.00	0.29
		Tongue dorsum	105	0.26	0.58	0.39	0.07	0.06	0.19
		Oral average	93	0.25	0.42	0.34	0.17	0.19	0.14
	Airway	Anterior nares	64	0.39	0.28	0.67	0.09	0.39	0.52
	Skin	Left antecubital fossa	14	0.19	0.43	0.00	0.00	0.16	0.34
		Right antecubital fossa	15	0.09	0.18	0.24	0.03	0.00	0.00
		Left retroauricular crease	63	0.15	0.22	0.25	0.18	0.08	0.03
		Right retroauricular crease	71	0.21	0.16	0.32	0.20	0.21	0.14
		Skin average	41	0.16	0.25	0.20	0.10	0.11	0.13
	Vagina	Mid-vagina	37	0.46	0.60	0.54	0.42	0.28	0.49
		Posterior fornix	38	0.45	0.70	0.69	0.78	0.07	0.00
		Vaginal introitus	37	0.39	0.73	0.33	0.42	0.21	0.25
		Vagina average	37	0.43	0.68	0.52	0.54	0.18	0.25

* First five principal coordinates (PCoA1-5) of the distance matrices.

Web Table 4. Intraclass correlation coefficients for weighted Unifrac distance*

		Body Site	***N*[a]**	**PCoA1-5 Average**	**PCoA1**	**PCoA2**	**PCoA3**	**PCoA4**	**PCoA5**
CRC study	Gut	Stool	34	0.52	0.65	0.28	0.44	0.51	0.74
Costa Rica	Gut	Stool	121	0.52	0.45	0.55	0.59	0.48	0.51
	Oral	Saliva	42	0.40	0.57	0.50	0.40	0	0.52
HMP	Gut	Stool	100	0.41	0.44	0.54	0.16	0.38	0.50
	Oral	Saliva	86	0.42	0.68	0.32	0.41	0.29	0.38
	Oropharynx	Buccal mucosa	101	0.20	0.22	0.16	0.36	0.12	0.16
		Hard palate	95	0.33	0.39	0.41	0.35	0.20	0.32
		Keratinized gingiva	102	0.31	0.60	0.20	0.04	0.35	0.34
		Palatine tonsils	100	0.27	0.31	0.46	0.26	0.00	0.30
		Subgingival plaque	100	0.26	0.42	0.49	0.14	0.20	0.07
		Supragingival plaque	104	0.31	0.39	0.38	0.45	0.13	0.21
		Throat	98	0.29	0.31	0.44	0.23	0.22	0.23
		Tongue dorsum	105	0.33	0.59	0.45	0.20	0.38	0.01
		Oral average	93.2	0.29	0.40	0.37	0.26	0.20	0.21
	Airway	Anterior nares	64	0.21	0.34	0.00	0.10	0.15	0.43
	Skin	Left antecubital fossa	14	0.12	0.08	0.20	0.12	0.00	0.22
		Right antecubital fossa	15	0.16	0.34	0.11	0.09	0.27	0.00
		Left retroauricular crease	63	0.45	0.69	0.27	0.49	0.38	0.40
		Right retroauricular crease	71	0.38	0.68	0.00	0.41	0.24	0.57
		Skin average	40.8	0.28	0.45	0.14	0.28	0.22	0.30
	Vagina	Mid-vagina	37	0.48	0.55	0.68	0.32	0.22	0.61
		Posterior fornix	38	0.26	0.55	0.58	0.13	0.00	0.06
		Vaginal introitus	37	0.42	0.48	0.53	0.62	0.06	0.40
		Vagina average	37.3	0.39	0.53	0.60	0.36	0.09	0.36

* First five principal coordinates (PCoA1-5) of the distance matrices.

Web Table 5. Proportion of explained variance for unweighted Unifrac distance*

		PCoA1	**PCoA2**	**PCoA3**	**PCoA4**	**PCoA5**	**PCoA1-5**
CRC study	Stool	3.6%	2.5%	2.2%	2.1%	1.8%	12.2%
Costa Rica	Saliva	19.8%	5.8%	4.1%	3.2%	3.2%	36.0%
Costa Rica	Stool	11.0%	6.2%	4.0%	2.1%	2.0%	25.3%
HMP	Anterior_nares	14.4%	8.7%	5.9%	4.8%	4.6%	38.5%
HMP	Attached_Keratinized_gingiva	19.8%	10.1%	7.7%	5.8%	4.9%	48.3%
HMP	Buccal_mucosa	11.5%	7.1%	6.5%	4.4%	3.9%	33.4%
HMP	Hard_palate	9.3%	6.9%	5.6%	4.1%	3.6%	29.6%
HMP	Left_Antecubital_fossa	9.8%	5.5%	5.0%	3.9%	3.5%	27.7%
HMP	Left_Retroauricular_crease	18.7%	8.9%	7.0%	6.0%	4.9%	45.4%
HMP	Mid_vagina	27.0%	9.4%	7.5%	6.0%	5.6%	55.6%
HMP	Palatine_Tonesils	10.8%	8.6%	5.2%	4.6%	3.9%	33.1%
HMP	Posterior_fonix	29.6%	10.3%	8.4%	6.6%	5.0%	59.8%
HMP	Right_Antecubital_fossa	8.7%	5.4%	4.4%	3.9%	3.8%	26.2%
HMP	Right_Retroauricular_crease	19.1%	9.6%	7.0%	5.1%	4.7%	45.5%
HMP	Saliva	7.2%	6.8%	4.7%	3.9%	3.7%	26.5%
HMP	Stool	9.0%	4.3%	3.8%	3.5%	2.9%	23.5%
HMP	Subgingival_plague	14.4%	5.3%	5.1%	4.2%	3.7%	32.8%
HMP	Supragingival_plague	14.7%	5.1%	4.9%	4.5%	3.5%	32.7%
HMP	Throat	10.2%	6.9%	5.4%	4.1%	3.8%	30.5%
HMP	Tongue_dorsum	14.9%	8.4%	5.3%	4.6%	3.8%	37.0%
HMP	Vaginal_introitus	22.5%	8.3%	6.9%	6.1%	5.3%	49.1%

* First five principal coordinates (PCoA1-5) of the distance matrices.

Web Table 6. Proportion of explained variance for weighted Unifrac distance*

		PCoA1	**PCoA2**	**PCoA3**	**PCoA4**	**PCoA5**	**PCoA1-5**
CRC study	Stool	15.3%	5.2%	5.8%	3.9%	3.5%	33.7%
Costa Rica	Saliva	24.1%	15.5%	11.0%	5.3%	4.6%	60.4%
Costa Rica	Stool	31.4%	8.7%	5.0%	4.3%	3.6%	53.1%
HMP	Anterior_nares	25.0%	13.6%	10.7%	7.6%	5.8%	62.7%
HMP	Attached_Keratinized_gingiva	39.6%	17.0%	6.8%	5.6%	2.8%	71.8%
HMP	Buccal_mucosa	31.4%	16.8%	7.2%	4.9%	3.3%	63.6%
HMP	Hard_palate	32.8%	16.8%	6.0%	4.4%	3.8%	63.7%
HMP	Left_Antecubital_fossa	32.4%	14.4%	7.4%	6.0%	4.3%	64.5%
HMP	Left_Retroauricular_crease	51.3%	11.0%	5.8%	5.1%	3.3%	76.4%
HMP	Mid_vagina	56.4%	13.8%	5.8%	4.4%	3.4%	83.8%
HMP	Palatine_Tonesils	30.0%	15.5%	8.5%	6.6%	3.2%	63.8%
HMP	Posterior_fonix	62.2%	15.4%	3.6%	3.5%	3.2%	88.0%
HMP	Right_Antecubital_fossa	34.5%	12.7%	7.0%	6.5%	4.5%	65.3%
HMP	Right_Retroauricular_crease	50.3%	10.4%	7.4%	5.9%	3.1%	77.0%
HMP	Saliva	25.6%	16.2%	6.8%	5.1%	4.4%	58.2%
HMP	Stool	46.9%	6.6%	5.9%	3.7%	3.0%	66.1%
HMP	Subgingival_plague	31.1%	9.7%	7.9%	5.8%	4.5%	59.0%
HMP	Supragingival_plague	26.8%	15.1%	8.8%	4.7%	4.4%	59.7%
HMP	Throat	28.0%	16.7%	7.8%	5.7%	4.0%	62.2%
HMP	Tongue_dorsum	28.6%	19.0%	6.7%	5.8%	4.5%	64.6%
HMP	Vaginal_introitus	54.7%	13.0%	6.1%	4.1%	3.5%	81.4%

* First five principal coordinates (PCoA1-5) of the distance matrices.

Web Table 7. Number of cases (assuming equal number of controls) required to detect an association with specified odds ratio (OR), at significance level=0.05, 80% power, based on 1, 2, or 3 fecal Specimens per Subject and the intraclass correlation coefficient (ICC) estimated from the fecal samples from the three populations.

		Estimated ICC[a]	**#Specimens per Subject**	**OR[b]**				
				1.5	**2**	**2.5**	**3**	**3.5**
			1	1496	498	273	190	151
Relative abundance of phylum-level taxa	Actinobacteria	0.44	2	1077	359	196	136	109
			3	938	312	171	119	94
			1	1780	593	325	226	180
	Bacteroidetes	0.37	2	1219	406	222	154	123
			3	1032	344	188	131	104
			1	1568	522	286	199	158
	Firmicutes	0.42	2	1113	371	203	141	112
			3	961	320	175	122	97
			1	3659	1219	668	464	370
	Fusobacteria	0.18	2	2158	719	394	274	218
			3	1658	552	302	210	167
			1	2271	757	414	288	229
	Proteobacteria	0.29	2	1464	488	267	186	148
			3	1196	398	218	151	121
Alpha diversity			1	1135	378	207	144	114
	PD_tree	0.58	2	897	299	163	113	90
			3	817	272	149	103	82
			1	1496	498	273	190	151
	Chao1	0.44	2	1077	359	196	136	109
			3	938	312	171	119	94
			1	1531	510	279	194	155
	# species	0.43	2	1095	365	199	139	110
			3	949	316	173	120	96
			1	1401	467	255	177	141
	Shannon index	0.47	2	1030	343	188	130	104
			3	906	302	165	115	91
Beta diversity			1	813	271	148	103	82
	Unweighted UniFrac PCoA1	0.81	2	735	245	134	93	74
			3	710	236	129	90	71
			1	1372	457	250	174	138
	Weighted UniFrac PCoA1	0.48	2	1015	338	185	128	102
			3	896	298	163	113	90

Web Table 8. Number of cases (assuming equal number of controls) required to detect an association with specified odds ratio (OR), at significance level=0.001, 80% power, based on 1, 2, or 3 fecal Specimens per Subject and the intraclass correlation coefficient (ICC) estimated from the fecal samples from the three populations.

		Estimated ICC[a]	**#Specimens per Subject**	**OR**[b]				
				1.5	**2**	**2.5**	**3**	**3.5**
			1	3256	1085	594	413	329
Relative abundance of phylum-level taxa	Actinobacteria	0.44	2	2344	781	428	297	237
			3	2040	680	372	259	206
			1	3872	1290	707	491	392
	Bacteroidetes	0.37	2	2652	884	484	336	268
			3	2246	748	410	285	227
			1	3411	1137	622	433	345
	Firmicutes	0.42	2	2422	807	442	307	245
			3	2092	697	382	265	211
			1	7960	2653	1453	1010	805
	Fusobacteria	0.18	2	4696	1565	857	596	475
			3	3608	1202	658	458	365
			1	4940	1646	902	627	500
	Proteobacteria	0.29	2	3186	1062	581	404	322
			3	2602	867	475	330	263
Alpha diversity			1	2470	823	451	313	250
	PD_tree	0.58	2	1951	650	356	247	197
			3	1778	592	324	225	180
			1	3256	1085	594	413	329
	Chao1	0.44	2	2344	781	428	297	237
			3	2040	680	372	259	206
			1	3332	1110	608	423	337
	# species	0.43	2	2382	794	435	302	241
			3	2065	688	377	262	209
			1	3048	1016	556	387	308
	Shannon index	0.47	2	2240	746	409	284	226
			3	1971	657	360	250	199
Beta diversity			1	1768	589	323	224	179
	Unweighted UniFrac PCoA1	0.81	2	1600	533	292	203	162
			3	1544	514	282	196	156
			1	2985	995	545	379	302
	Weighted UniFrac PCoA1	0.48	2	2208	736	403	280	223
			3	1950	650	356	247	197

Web Table 9. Number of cases (assuming equal number of controls) required to detect an association with specified odds ratio (OR), at significance level=0.0001, 80% power, based on 1, 2, or 3 fecal Specimens per Subject and the intraclass correlation coefficient (ICC) estimated from the fecal samples from the three populations.

		Estimated ICC[a]	**#Specimens per Subject**	**OR[b]**				
				1.5	**2**	**2.5**	**3**	**3.5**
			1	4270	1423	779	542	432
Relative abundance of phylum-level taxa	Actinobacteria	0.44	2	3075	1024	561	390	311
			3	2676	892	488	339	270
			1	5078	1692	927	644	514
	Bacteroidetes	0.37	2	3479	1159	635	441	352
			3	2945	981	537	374	298
			1	4474	1491	817	568	452
	Firmicutes	0.42	2	3176	1058	580	403	321
			3	2744	914	501	348	277
			1	10439	3479	1906	1325	1056
	Fusobacteria	0.18	2	6159	2053	1124	782	623
			3	4732	1577	864	600	479
			1	6479	2159	1183	822	655
	Proteobacteria	0.29	2	4179	1393	763	530	423
			3	3412	1137	623	433	345
Alpha diversity			1	3239	1079	591	411	327
	PD_tree	0.58	2	2559	853	467	324	259
			3	2332	777	425	296	236
			1	4270	1423	779	542	432
	Chao1	0.44	2	3075	1024	561	390	311
			3	2676	892	488	339	270
			1	4370	1456	798	554	442
	# species	0.43	2	3124	1041	570	396	316
			3	2709	903	494	344	274
			1	3998	1332	730	507	404
	Shannon index	0.47	2	2938	979	536	373	297
			3	2585	861	472	328	261
Beta diversity			1	2319	773	423	294	234
	Unweighted UniFrac PCoA1	0.81	2	2099	699	383	266	212
			3	2026	675	369	257	205
			1	3914	1304	714	497	396
	Weighted UniFrac PCoA1	0.48	2	2897	965	529	367	293
			3	2557	852	467	324	258

Web Table 10. Number of cases (assuming equal number of controls) required to detect an association with specified odds ratio (OR), at significance level=0.00001, 80% power, based on 1, 2, or 3 fecal Specimens per Subject and the intraclass correlation coefficient (ICC) estimated from the fecal samples from the three populations.

		Estimated ICC[a]	**#Specimens per Subject**	**OR[b]**				
				1.5	**2**	**2.5**	**3**	**3.5**
			1	5274	1758	963	669	533
Relative abundance of phylum-level taxa	Actinobacteria	0.44	2	3797	1265	693	482	384
			3	3305	1101	603	419	334
			1	6272	2090	1145	796	634
	Bacteroidetes	0.37	2	4296	1432	784	545	434
			3	3637	1212	664	461	368
			1	5525	1841	1009	701	559
	Firmicutes	0.42	2	3923	1307	716	498	397
			3	3388	1129	618	430	343
			1	12892	4297	2354	1636	1305
	Fusobacteria	0.18	2	7606	2535	1389	965	770
			3	5844	1948	1067	742	591
			1	8002	2667	1461	1016	810
	Proteobacteria	0.29	2	5161	1720	942	655	522
			3	4214	1404	769	535	426
Alpha diversity			1	4001	1333	730	508	405
	PD_tree	0.58	2	3160	1053	577	401	319
			3	2880	960	526	365	291
			1	5274	1758	963	669	533
	Chao1	0.44	2	3797	1265	693	482	384
			3	3305	1101	603	419	334
			1	5396	1798	985	685	546
	# species	0.43	2	3858	1286	704	489	390
			3	3346	1115	611	424	338
			1	4937	1645	901	626	499
	Shannon index	0.47	2	3629	1209	662	460	367
			3	3192	1064	583	405	323
Beta diversity			1	2865	954	523	363	290
	Unweighted UniFrac PCoA1	0.81	2	2592	864	473	329	262
			3	2502	834	456	317	253
			1	4834	1611	882	613	489
	Weighted UniFrac PCoA1	0.48	2	3577	1192	653	454	362
			3	3158	1052	576	401	319

Web Table 11. Number of cases (assuming cases:controls=1:3) required to detect an association with specified odds ratio (OR), at significance level=0.05, 80% power, based on 1, 2, or 3 fecal Specimens per Subject and the intraclass correlation coefficient (ICC) estimated from the fecal samples from the three populations.

		Estimated ICC[a]	**#Specimens per Subject**	**OR[b]**				
				1.5	**2**	**2.5**	**3**	**3.5**
			1	997	332	182	126	101
Relative abundance of phylum-level taxa	Actinobacteria	0.44	2	718	239	131	91	72
			3	625	208	114	79	63
			1	1186	395	216	150	120
	Bacteroidetes	0.37	2	812	270	148	103	82
			3	688	229	125	87	69
			1	1045	348	190	132	105
	Firmicutes	0.42	2	742	247	135	94	75
			3	641	213	117	81	64
			1	2439	813	445	309	246
	Fusobacteria	0.18	2	1439	479	262	182	145
			3	1105	368	201	140	111
			1	1514	504	276	192	153
	Proteobacteria	0.29	2	976	325	178	124	98
			3	797	265	145	101	80
Alpha diversity			1	757	252	138	96	76
	PD_tree	0.58	2	598	199	109	75	60
			3	545	181	99	69	55
			1	997	332	182	126	101
	Chao1	0.44	2	718	239	131	91	72
			3	625	208	114	79	63
			1	1021	340	186	129	103
	# species	0.43	2	730	243	133	92	73
			3	633	211	115	80	64
			1	934	311	170	118	94
	Shannon index	0.47	2	686	228	125	87	69
			3	604	201	110	76	61
Beta diversity			1	542	180	98	68	54
	Unweighted UniFrac PCoA1	0.81	2	490	163	89	62	49
			3	473	157	86	60	47
			1	914	304	167	116	92
	Weighted UniFrac PCoA1	0.48	2	676	225	123	85	68
			3	597	199	109	75	60

Web Table 12. Number of cases (assuming cases: controls =1:3) required to detect an association with specified odds ratio (OR), at significance level=0.001, 80% power, based on 1, 2, or 3 fecal Specimens per Subject and the intraclass correlation coefficient (ICC) estimated from the fecal samples from the three populations.

		Estimated ICC^a	**#Specimens per Subject**	**OR^b**				
				1.5	**2**	**2.5**	**3**	**3.5**
			1	2170	723	396	275	219
Relative abundance of phylum-level taxa	Actinobacteria	0.44	2	1563	521	285	198	158
			3	1360	453	248	172	137
			1	2581	860	471	327	261
	Bacteroidetes	0.37	2	1768	589	322	224	179
			3	1497	499	273	190	151
			1	2274	758	415	288	230
	Firmicutes	0.42	2	1614	538	294	205	163
			3	1394	464	254	177	141
			1	5306	1768	969	673	537
	Fusobacteria	0.18	2	3130	1043	571	397	316
			3	2405	801	439	305	243
			1	3293	1097	601	418	333
	Proteobacteria	0.29	2	2124	708	387	269	215
			3	1734	578	316	220	175
Alpha diversity			1	1646	548	300	209	166
	PD_tree	0.58	2	1301	433	237	165	131
			3	1185	395	216	150	120
			1	2170	723	396	275	219
	Chao1	0.44	2	1563	521	285	198	158
			3	1360	453	248	172	137
			1	2221	740	405	282	224
	# species	0.43	2	1588	529	290	201	160
			3	1377	459	251	174	139
			1	2032	677	371	258	205
	Shannon index	0.47	2	1493	497	272	189	151
			3	1314	438	240	166	133
Beta diversity			1	1179	393	215	149	119
	Unweighted UniFrac PCoA1	0.81	2	1067	355	194	135	108
			3	1029	343	188	130	104
			1	1990	663	363	252	201
	Weighted UniFrac PCoA1	0.48	2	1472	490	268	186	149
			3	1300	433	237	165	131

Web Table 13. Number of cases (assuming cases: controls =1:3) required to detect an association with specified odds ratio (OR), at significance level=0.0001, 80% power, based on 1, 2, or 3 fecal Specimens per Subject and the intraclass correlation coefficient (ICC) estimated from the fecal samples from the three populations.

		Estimated ICC[a]	**#Specimens per Subject**	**OR[b]**				
				1.5	**2**	**2.5**	**3**	**3.5**
			1	2847	949	519	361	288
Relative abundance of phylum-level taxa	Actinobacteria	0.44	2	2050	683	374	260	207
			3	1784	594	325	226	180
			1	3385	1128	618	429	342
	Bacteroidetes	0.37	2	2319	773	423	294	234
			3	1963	654	358	249	198
			1	2982	994	544	378	301
	Firmicutes	0.42	2	2117	705	386	268	214
			3	1829	609	334	232	185
			1	6959	2319	1270	883	704
	Fusobacteria	0.18	2	4106	1368	749	521	415
			3	3155	1051	576	400	319
			1	4319	1439	788	548	437
	Proteobacteria	0.29	2	2786	928	508	353	282
			3	2275	758	415	288	230
Alpha diversity			1	2159	719	394	274	218
	PD_tree	0.58	2	1706	568	311	216	172
			3	1555	518	283	197	157
			1	2847	949	519	361	288
	Chao1	0.44	2	2050	683	374	260	207
			3	1784	594	325	226	180
			1	2913	971	532	369	294
	# species	0.43	2	2083	694	380	264	210
			3	1806	602	329	229	182
			1	2665	888	486	338	269
	Shannon index	0.47	2	1959	653	357	248	198
			3	1723	574	314	218	174
Beta diversity			1	1546	515	282	196	156
	Unweighted UniFrac PCoA1	0.81	2	1399	466	255	177	141
			3	1350	450	246	171	136
			1	2609	869	476	331	264
	Weighted UniFrac PCoA1	0.48	2	1931	643	352	245	195
			3	1705	568	311	216	172

Web Table 14. Number of cases (assuming cases: controls =1:3) required to detect an association with specified odds ratio (OR), at significance level=0.00001, 80% power, based on 1, 2, or 3 fecal Specimens per Subject and the intraclass correlation coefficient (ICC) estimated from the fecal samples from the three populations.

		Estimated ICC[a]	**#Specimens per Subject**	OR[b]				
				1.5	**2**	**2.5**	**3**	**3.5**
			1	3516	1172	642	446	355
Relative abundance of phylum-level taxa	Actinobacteria	0.44	2	2531	843	462	321	256
			3	2203	734	402	279	223
			1	4181	1393	763	530	423
	Bacteroidetes	0.37	2	2864	954	523	363	289
			3	2425	808	442	307	245
			1	3683	1227	672	467	372
	Firmicutes	0.42	2	2615	871	477	332	264
			3	2259	753	412	286	228
			1	8595	2864	1569	1091	870
	Fusobacteria	0.18	2	5071	1690	926	643	513
			3	3896	1298	711	494	394
			1	5334	1778	974	677	540
	Proteobacteria	0.29	2	3440	1146	628	436	348
			3	2809	936	513	356	284
Alpha diversity			1	2667	889	487	338	270
	PD_tree	0.58	2	2107	702	384	267	213
			3	1920	640	350	243	194
			1	3516	1172	642	446	355
	Chao1	0.44	2	2531	843	462	321	256
			3	2203	734	402	279	223
			1	3597	1199	657	456	364
	# species	0.43	2	2572	857	469	326	260
			3	2230	743	407	283	225
			1	3291	1097	601	417	333
	Shannon index	0.47	2	2419	806	441	307	244
			3	2128	709	388	270	215
Beta diversity			1	1910	636	348	242	193
	Unweighted UniFrac PCoA1	0.81	2	1728	576	315	219	174
			3	1668	556	304	211	168
			1	3223	1074	588	409	326
	Weighted UniFrac PCoA1	0.48	2	2385	795	435	302	241
			3	2105	701	384	267	213