

Web Material

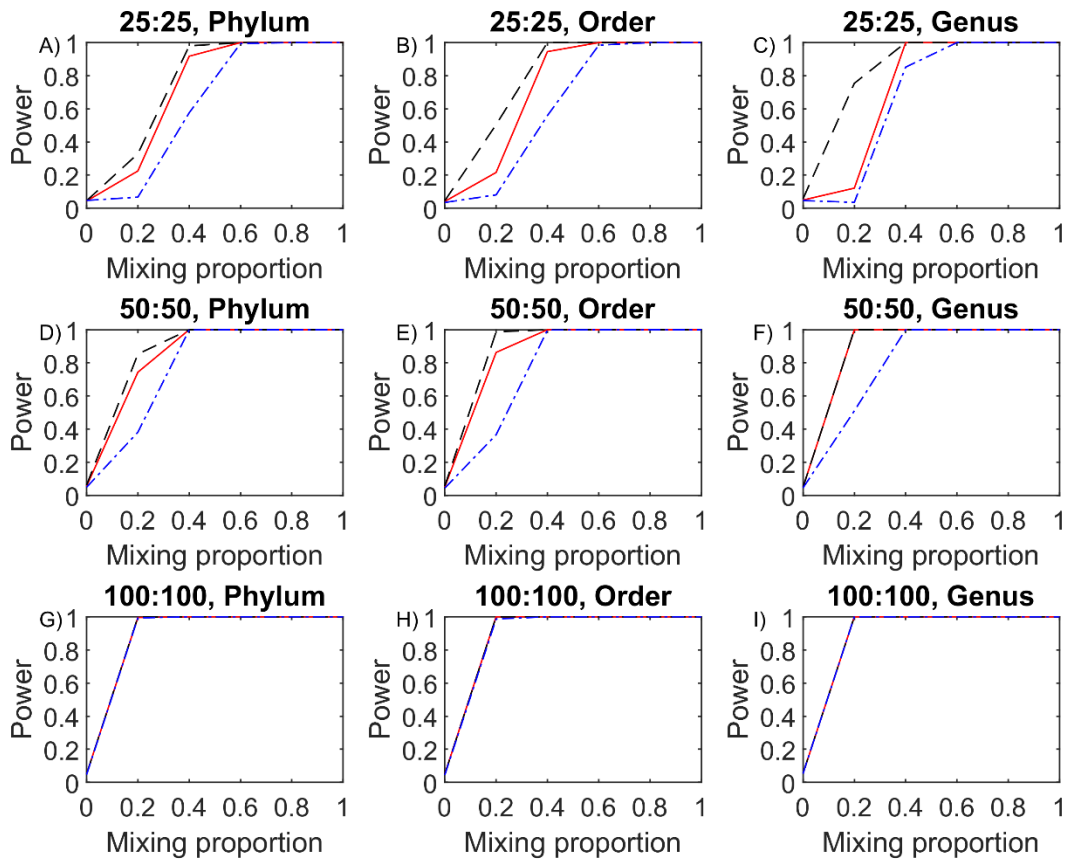
Power of Microbiome Beta-Diversity Analyses Based on Standard Reference Samples

Mitchell H. Gail, Yunhu Wan, and Jianxin Shi

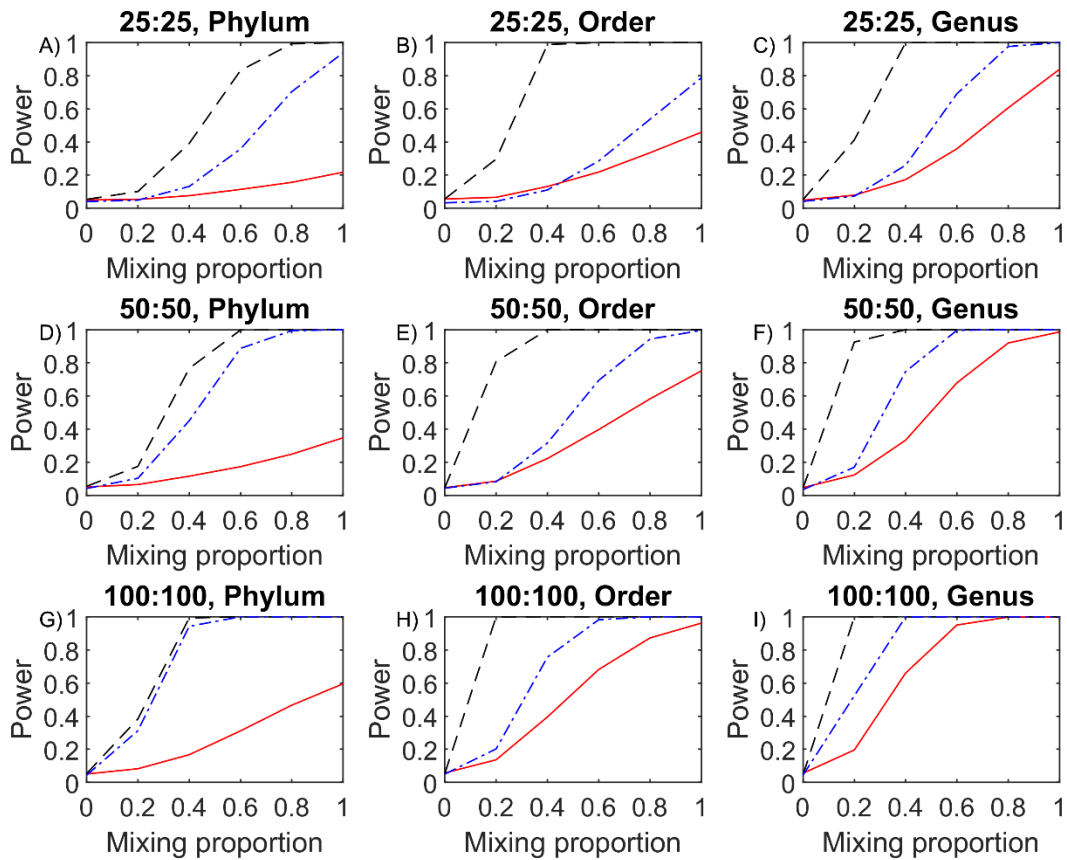
Contents

Web Figures 1–12

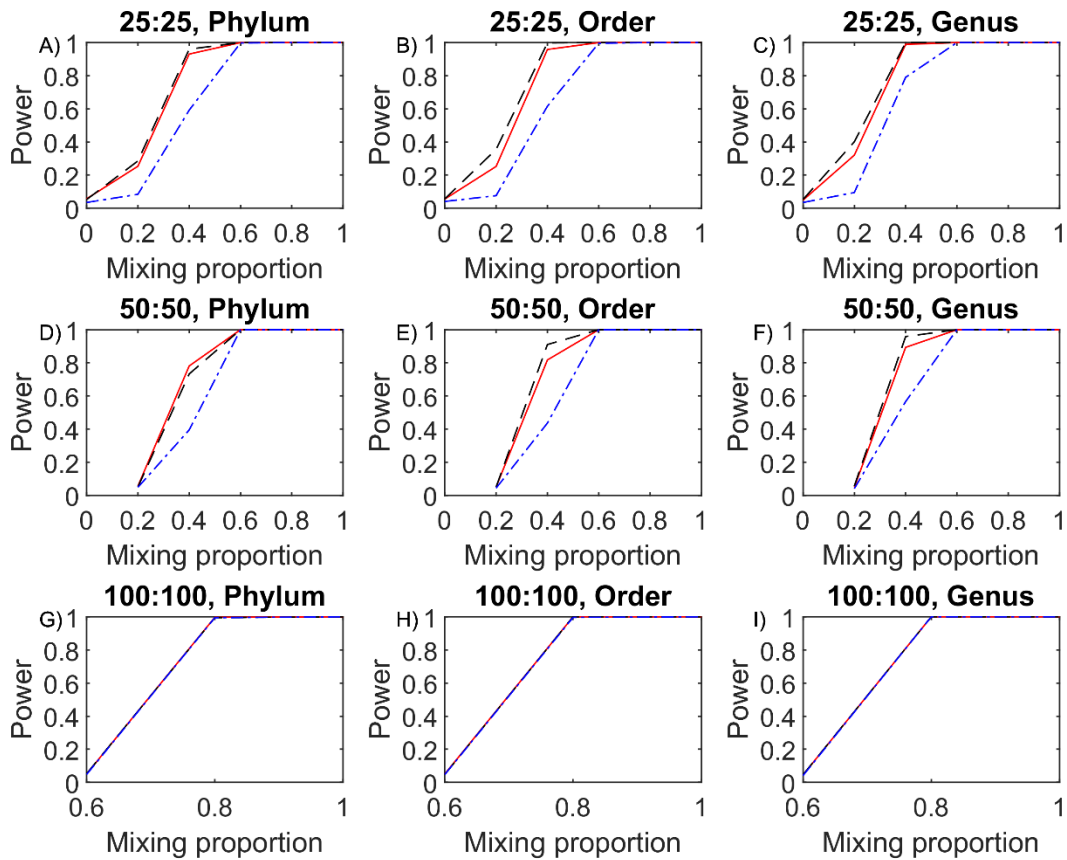
Web Tables 1–5



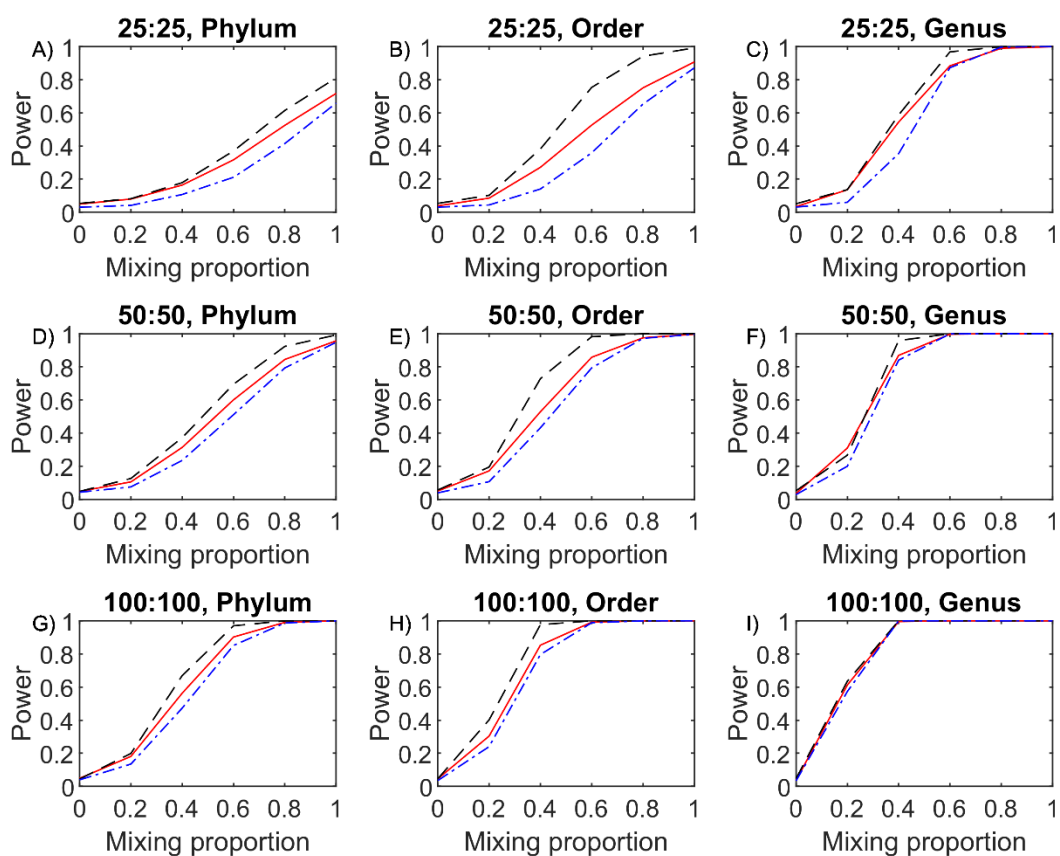
Web Figure 1. Power to discriminate skin from saliva samples using Bray-Curtis distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



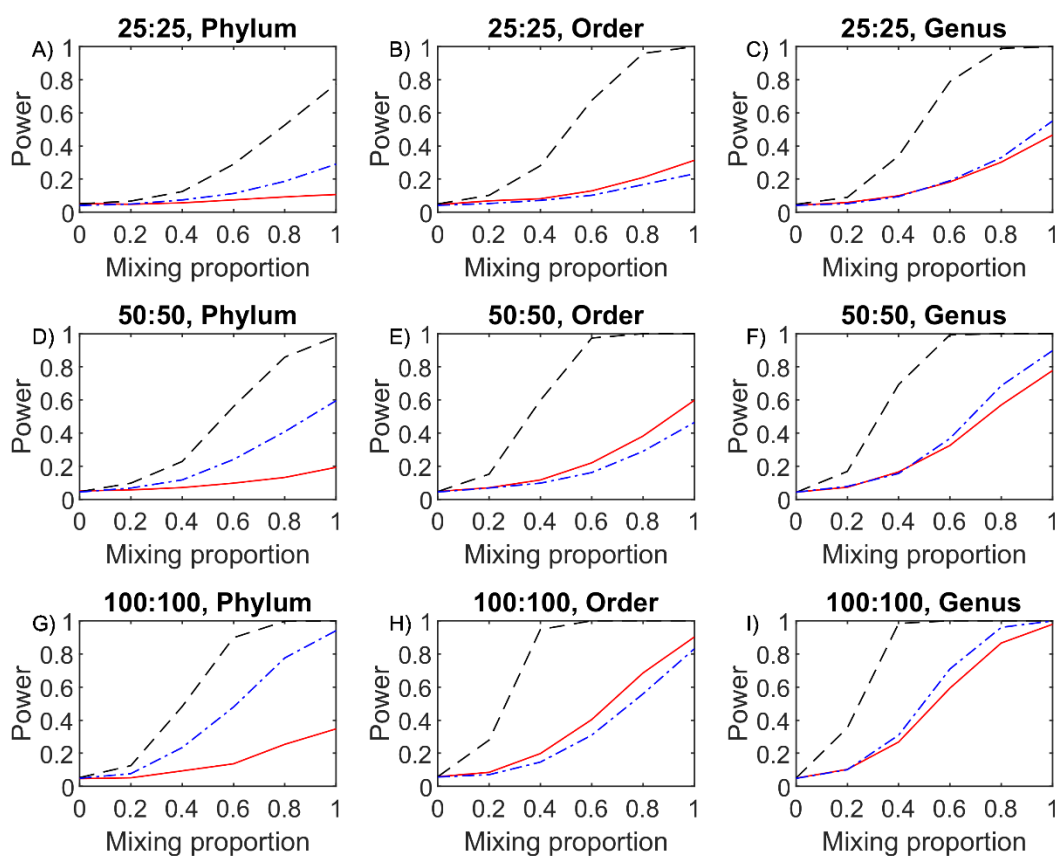
Web Figure 2. Power to discriminate skin from saliva samples using Unweighted Unifrac distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or $2df$ test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



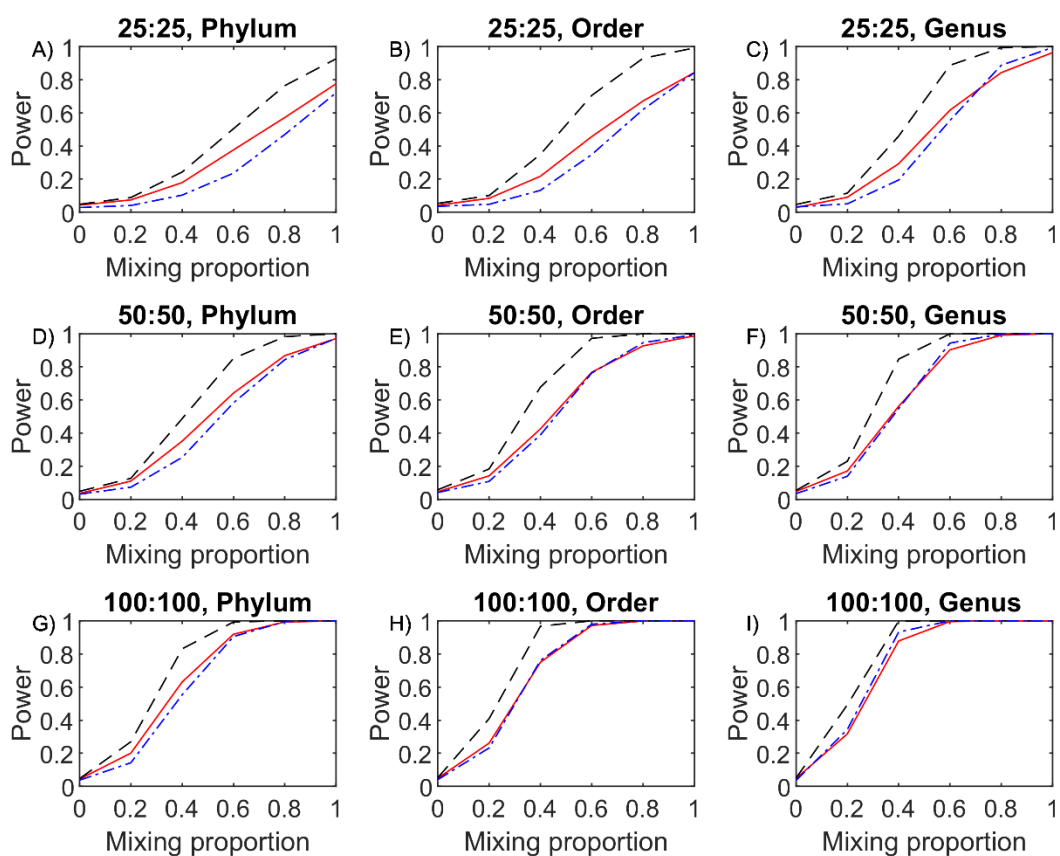
Web Figure 3. Power to discriminate skin from saliva samples using Weighted Unifrac distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



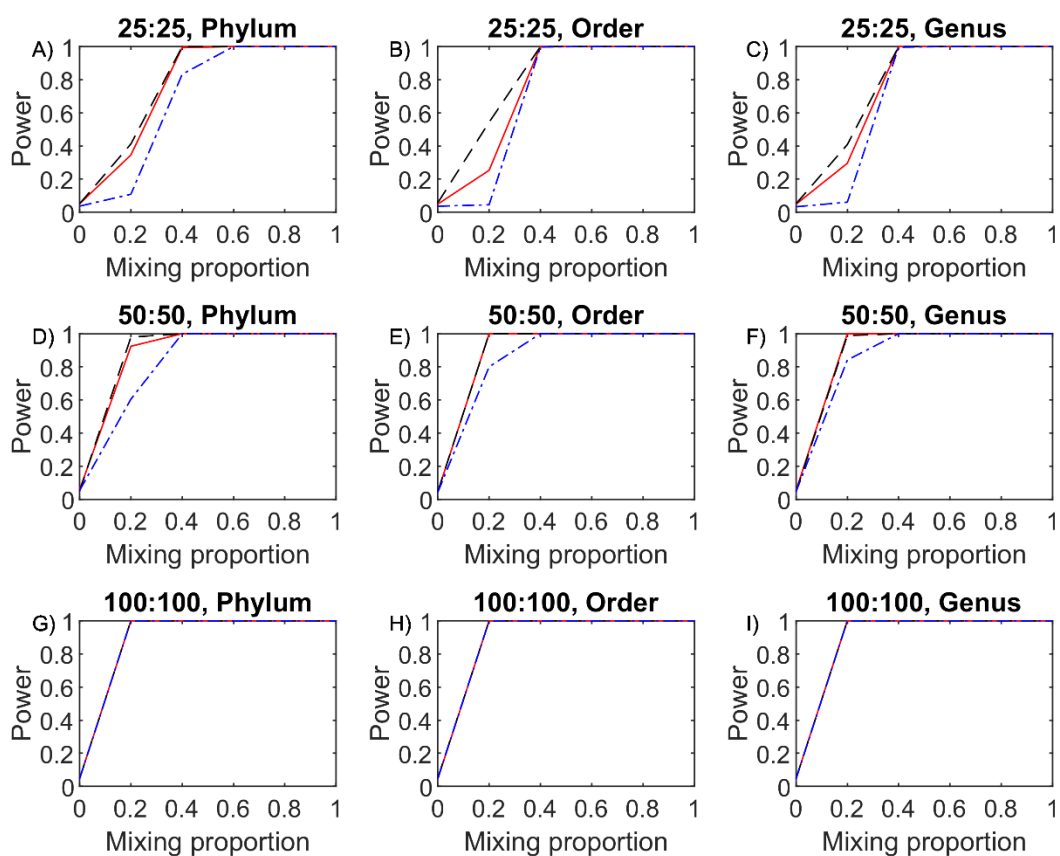
Web Figure 4. Power to discriminate skin from nasal samples using Bray-Curtis distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



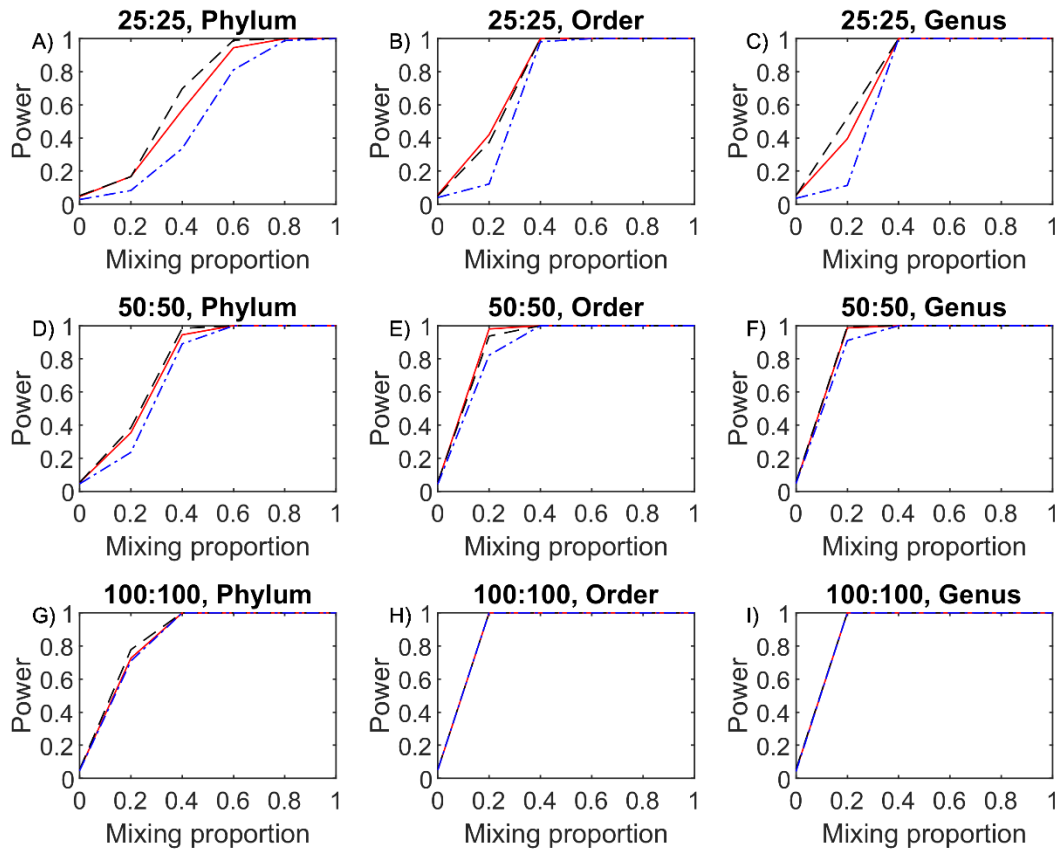
Web Figure 5. Power to discriminate skin from nasal samples using Unweighted Unifrac distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



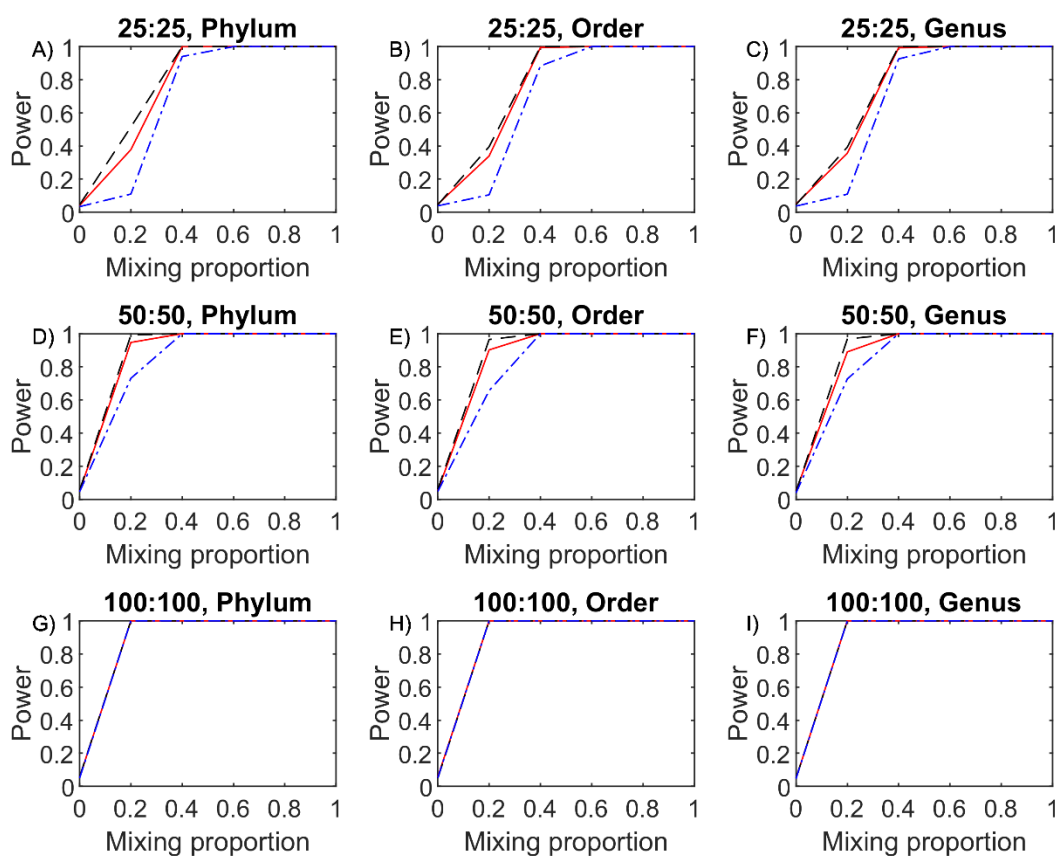
Web Figure 6. Power to discriminate skin from nasal samples using Weighted Unifrac distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



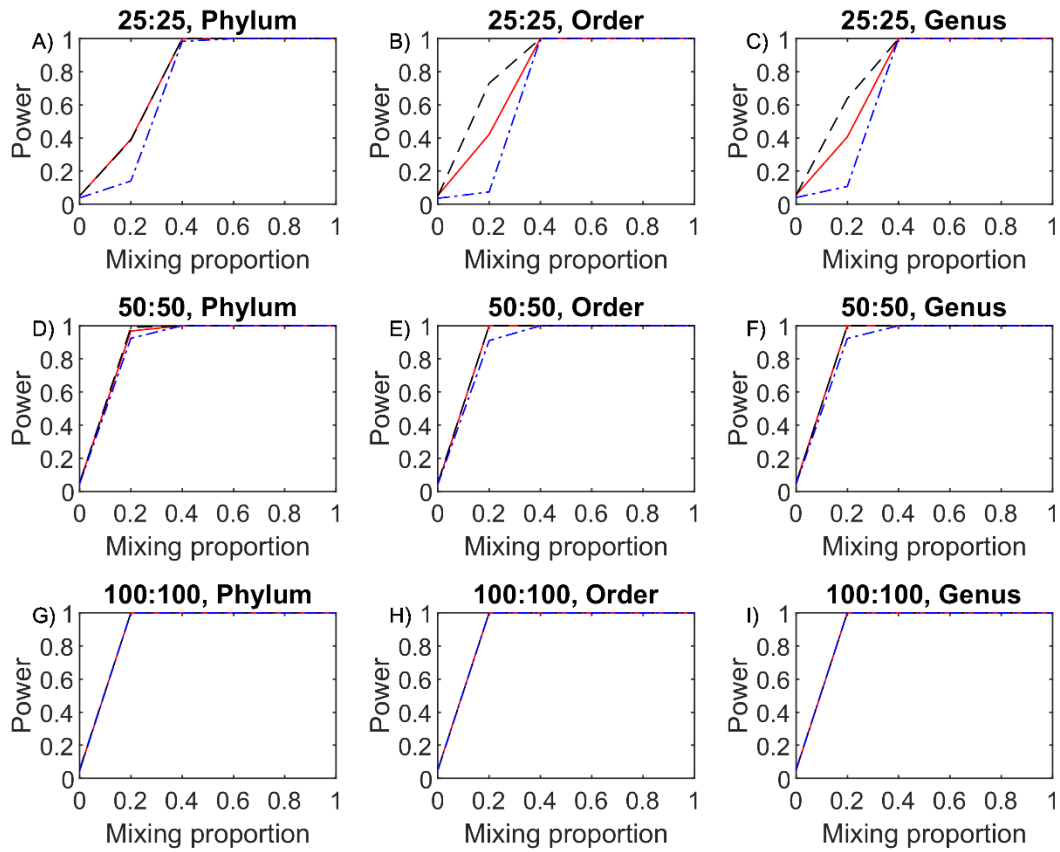
Web Figure 7. Power to discriminate skin from stool samples using Bray-Curtis distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



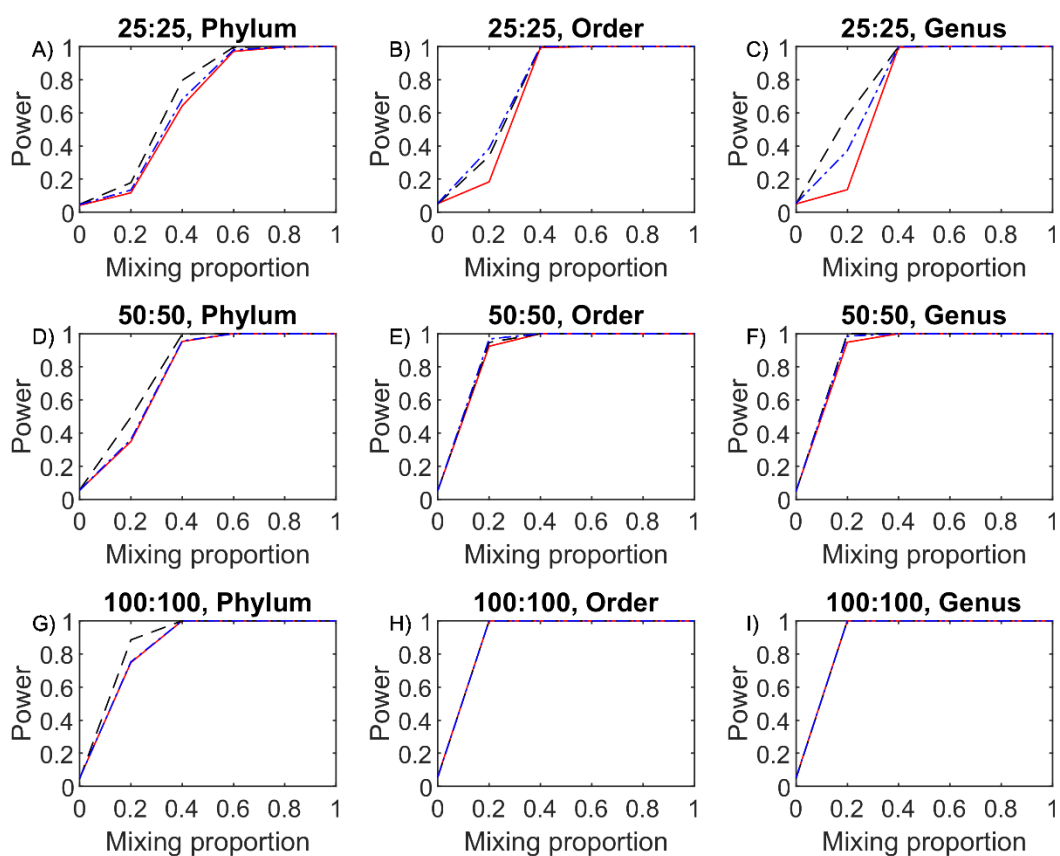
Web Figure 8. Power to discriminate skin from stool samples using Unweighted Unifrac distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



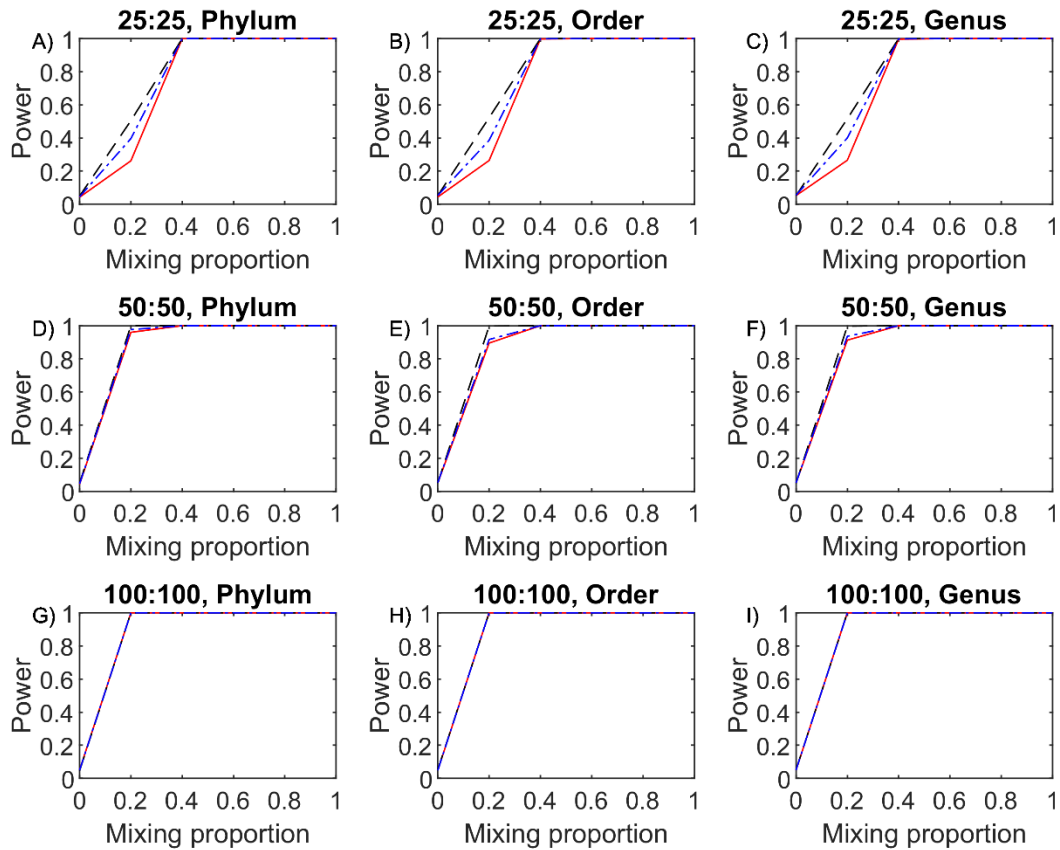
Web Figure 9. Power to discriminate skin from stool samples using Weighted Unifrac distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



Web Figure 10. Power to discriminate stool from nasal samples using Bray-Curtis distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



Web Figure 11. Power to discriminate stool from nasal samples using Unweighted Unifrac distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.



Web Figure 12. Power to discriminate stool from nasal samples using Weighted Unifrac distance at the phylum, order and genus levels with sample sizes 25, 50 or 100 in each group. The solid (red) locus refers to the standard reference method with the Hotelling T^2 test (or 2df test); the dashed-dot (blue) locus refers to the 5df standard reference test; the dashed (black) locus refers to MiRKAT.

Web Table 1. Phyla read counts from 18 HMP non-reference skin samples with at least 1000 total reads. Only phyla with a least one read in these 18 samples are shown.

Other	Acido- bacteria	Actino- bacteria	Armati- monadetes	Bacteroi- detes	Chloroflexi	Cyano- bacteria	Euryar- chaeota	Firmicutes	Fuso- bacteria	Gemmati- monadetes	Plancto- mycetes	Proteo- bacteria	Spiro- chaeetes	Synergis- tetes	Teneri- cutes	Thermi	WPS-2	Total reads
0	4	356	0	137	23	31	0	612	4	0	0	1169	0	0	0	0	0	2336
67	0	995	31	150	0	0	30	309	128	32	0	3331	0	0	0	43	0	5116
0	0	1429	0	14	0	42	0	263	4	0	0	112	0	0	0	11	0	1875
0	0	2361	0	39	0	0	0	192	0	0	0	90	0	0	1	0	0	2683
0	0	149	0	100	0	111	0	461	95	0	0	322	0	0	0	0	0	1238
0	0	951	0	215	0	7	0	909	0	0	0	533	0	0	0	0	0	2615
0	0	2806	0	80	0	2	0	664	1	0	0	97	0	0	0	0	0	3650
15	0	2290	0	581	0	0	0	998	8	0	0	880	0	0	32	0	0	4804
0	0	540	0	120	0	6	0	1691	2	0	0	327	0	0	0	0	0	2686
0	0	984	0	67	0	19	0	394	24	0	12	503	0	0	0	0	0	2003
0	0	476	0	67	0	0	0	524	10	0	0	325	0	0	0	0	0	1402
0	0	746	0	20	0	0	0	153	0	0	0	76	0	0	5	0	0	1000
0	22	592	0	22	0	32	0	472	11	0	0	164	0	0	15	0	0	1330
0	0	4159	0	1	0	0	0	63	0	0	0	58	0	0	0	0	0	4281
0	0	1756	0	7	0	2	0	394	5	0	0	637	0	0	0	0	0	2801
0	0	5336	0	20	0	35	0	166	1	0	0	59	0	0	8	0	0	5625
2	0	1090	0	1688	0	4	0	2288	301	0	0	405	15	3	8	0	0	5804
2	0	995	0	157	0	6	0	504	19	0	0	455	0	0	1	0	2	2141

Maximum likelihood parameter estimates from the corresponding Dirichlet multinomial distribution (1) from the CRAN program ‘dirmult’ follow. For the estimated Dirichlet parameters $\hat{\gamma}_k$, the values are 0.041066615, 0.019646245, 3.330978815, 0.009685926, 0.521687171, 0.009657488, 0.162877879, 0.009682801, 1.596910805, 0.196879877, 0.009688952, 0.009595361, 1.191197447, 0.009616721, 0.009457715, 0.075094277, 0.019994238, and 0.009413413. In an alternative parameterization, the estimated multinomial probabilities $\hat{\pi}_k$ are 0.005677570, 0.002716146, 0.460516818, 0.001339105, 0.072124660, 0.001335174, 0.022518307, 0.001338673, 0.220777232, 0.027219175, 0.001339524, 0.001326585, 0.164686264, 0.001329538, 0.001307555, 0.010381987, 0.002764257, and 0.001301430. The estimated overdispersion parameter is $\hat{\theta} = 0.1214605$. The overdispersion factor $\{1 + \theta(n_k - 1)\}$ multiplies the multinomial covariance matrix for $\hat{\pi}_k$, where n_k is the number of reads in the k^{th} sample. The 5 randomly phyla whose parameters were modified under the alternative were Bacteroidetes, Cyanobacteria, Fusobacteria, Spirochaetes and Actinobacteria. The first 3 had their γ_k parameters multiplied by ψ and the last two by $1/\psi$.

Reference

1. Tvedebrink T. Overdispersion in allelic counts and theta-correction in forensic genetics. *Theor Popul Biol.* 2010;78(3):200-10.

Web Table 2. Power to discriminate nasal samples from a mixture of 40% skin and 60% nasal samples^a

Phylogenetic level	Sample size in each group	Bray-Curtis			Unweighted Unifrac			Weighted Unifrac		
		2df ^b	5df ^b	MiRKAT	2df	5df	MiRKAT	2df	5df	MiRKAT
Phylum	25	.16	.11	.18	.06	.07	.12	.18	.10	.24
	50	.31	.23	.37	.07	.12	.23	.35	.25	.49
	100	.56	.47	.67	.09	.23	.48	.63	.55	.83
Order	25	.27	.14	.38	.08	.07	.28	.22	.13	.35
	50	.53	.43	.73	.12	.10	.60	.42	.39	.68
	100	.85	.80	.98	.20	.15	.95	.75	.76	.97
Genus	25	.54	.35	.59	.10	.09	.34	.29	.19	.46
	50	.87	.84	.96	.17	.16	.69	.56	.55	.85
	100	1.00	1.00	1.00	.27	.31	.98	.88	.93	1.00

^aBased on 3000 simulations and rounded to two places. Bold numbers indicate the test with highest power.

^bThe 2df-test is equivalent to a Hotelling T² test and is computed by testing for no main associations in a linear regression of group indicator on the mean distances to the two reference samples. The 5df-test tests that there are no main associations and no associations with the two squared mean distances and with their product. See Methods for details.

Web Table 3. Power to discriminate stool samples from a mixture of 20% skin and 80% stool samples^a

Phylogenetic level	Sample size in each group	Bray-Curtis			Unweighted Unifrac			Weighted Unifrac		
		2df ^b	5df ^b	MiRKAT	2df	5df	MiRKAT	2df	5df	MiRKAT
Phylum	25	.35	.11	.41	.17	.08	.17	.38	.11	.51
	50	.92	.60	.98	.35	.24	.38	.95	.73	.99
	100	1.00	1.00	1.00	.73	.71	.78	1.00	1.00	1.00
Order	25	.25	.04	.54	.42	.12	.37	.34	.11	.40
	50	1.00	.80	1.00	.98	.82	.94	.90	.66	.96
	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Genus	25	.30	.06	.41	.39	.11	.52	.36	.11	.39
	50	1.00	.84	.99	1.00	1.00	1.00	.89	.73	.97
	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

^aBased on 3000 simulations and rounded to two places. Bold numbers indicate the test with highest power.

^bThe 2df-test is equivalent to a Hotelling T² test and is computed by testing for no main associations in a linear regression of group indicator on the mean distances to the two reference samples. The 5df-test tests that there are no main associations and no associations with the two squared mean distances and with their product. See Methods for details.

Web Table 4. Power to discriminate stool samples from a mixture of 20% stool and 80% nasal samples^a

Phylogenetic level	Sample size in each group	Bray-Curtis			Unweighted Unifrac			Weighted Unifrac		
		2df ^b	5df ^b	MiRKAT	2df	5df	MiRKAT	2df	5df	MiRKAT
Phylum	25	.39	.14	.39	.12	.13	.18	.26	.40	.50
	50	.97	.92	.99	.35	.36	.50	.96	.98	1.00
	100	1.00	1.00	1.00	.75	.75	.88	1.00	1.00	1.00
Order	25	.42	.07	.73	.18	.38	.33	.26	.39	.52
	50	1.00	.91	1.00	.92	.97	.95	.90	.92	1.00
	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Genus	25	.41	.11	.64	.14	.37	.58	.27	.40	.52
	50	1.00	.92	1.00	.95	.98	1.00	.92	.94	1.00
	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

^aBased on 3000 simulations and rounded to two places. Bold numbers indicate the test with highest power.

^bThe 2df-test is equivalent to a Hotelling T² test and is computed by testing for no main associations in a linear regression of group indicator on the mean distances to the two reference samples. The 5df-test tests that there are no main associations and no associations with the two squared mean distances and with their product. See Methods for details.

Web Table 5. P-values for testing association between BMI and beta diversity in the American Gut Project stool samples

Taxonomic Level	Statistical Test	Dissimilarity Measure		
		Bray-Curtis	Unweighted Unifrac	Weighted Unifrac
6 (Genus) 1186 taxa ^a	5 df standard reference	0.0012^b	0.0117	2.03x10⁻⁵
	2 df standard reference	0.0406	0.6696	0.1924
	MiRKAT	0.0102	0.0020	0.0347
5(Family) 479 taxa	5 df standard reference	0.0054	0.0079	2.52x10⁻⁵
	2 df standard reference	0.1197	0.7962	0.2147
	MiRKAT	0.0067	0.0073	0.0354
4(Order) 257 taxa	5 df standard reference	0.0017	7.059x10⁻⁴	4.79x10⁻⁴
	2 df standard reference	0.1540	0.4592	0.079
	MiRKAT	0.0387	0.0385	0.04468
3(Class) 124 taxa	5 df standard reference	0.0019	0.6044	5.59x10⁻⁴
	2 df standard reference	0.1707	0.6387	0.0871
	MiRKAT	0.0573	0.1299	0.0528
2(Phylum) 44 taxa	5 df standard reference	2.14x10⁻⁴	0.0257	2.36x10⁻⁴
	2 df standard reference	0.1031	0.6064	0.0846
	MiRKAT	0.0570	0.1600	0.0474

^aThis is the number of taxa identified in our data with the closed reference library Greengenes 13.8.

^bSmallest p-value is in bold.