Supplementary Tables

Table S1. Associations of probiotic taxa with smoking status stratified by race

Probiotic taxa	Race	Prevalence			*P* value [a]		
		Current-smokers (*N*=592)	Former-smokers (*N*=477)	Never-smokers (*N*=547)	Current- vs Never-smokers	Former- vs Never-smokers	Current- vs Non-smokers [b]
Phylum *Actinobacteria*							
Genus *Bifidobacterium*	African-American	85.02%	73.38%	67.19%	1.22×10^{-4}	0.44	2.83×10^{-4}
	European-American	87.08%	68.22%	67.47%	0.02	0.28	6.85×10^{-4}
Species *Bifidobacterium longum*	African-American	66.43%	46.77%	42.26%	1.58×10^{-6}	0.66	1.63×10^{-6}
	European-American	70.22%	42.99%	33.73%	8.18×10^{-6}	0.35	2.37×10^{-6}
Phylum *Firmicutes*							
Genus *Lactobacillus*	African-American	87.20%	73.76%	75.59%	5.70×10^{-3}	0.28	1.32×10^{-3}
	European-American	94.38%	72.90%	68.67%	1.07×10^{-4}	0.87	3.27×10^{-5}
Species *Lactobacillus crispatus*	African-American	62.08%	37.26%	34.65%	1.78×10^{-7}	0.67	2.55×10^{-8}
	European-American	58.99%	33.18%	33.73%	1.36×10^{-3}	0.16	1.09×10^{-5}
Species *Lactobacillus fermentum*	African-American	52.17%	41.44%	38.85%	0.02	0.70	0.03
	European-American	70.22%	37.85%	28.31%	1.01×10^{-5}	0.77	1.11×10^{-7}
Species *Lactobacillus gasseri*	African-American	69.57%	57.03%	55.64%	6.78×10^{-3}	0.68	5.04×10^{-3}
	European-American	79.21%	55.61%	46.99%	1.91×10^{-4}	0.88	1.30×10^{-5}
Species *Lactobacillus oris*	African-American	42.27%	29.66%	21.52%	7.08×10^{-5}	0.02	9.83×10^{-4}
	European-American	47.19%	23.36%	19.28%	1.13×10^{-3}	0.81	4.73×10^{-5}
Species *Lactobacillus panis*	African-American	41.55%	31.56%	28.08%	5.84×10^{-4}	0.36	5.11×10^{-3}
	European-American	43.82%	21.03%	19.28%	1.31×10^{-3}	0.34	8.56×10^{-6}
Species *Lactobacillus reuteri*	African-American	41.06%	22.81%	21.00%	6.89×10^{-6}	0.85	5.43×10^{-7}
	European-American	42.70%	23.83%	20.48%	0.02	0.73	4.05×10^{-3}
Species *Lactobacillus ultunensis*	African-American	19.32%	11.03%	7.35%	4.15×10^{-3}	0.09	0.01
	European-American	23.6%	8.88%	5.42%	7.55×10^{-3}	0.33	4.87×10^{-3}

[a] *P* values were calculated from logistic regression. Sequencing batch as well as other covariates (age, sex, BMI, alcohol consumption, oral health and disease status at the 1st follow-up and total energy intake) were adjusted.
[b] Non-smokers included former- and never-smokers.

Table S2. Associations of common taxa with smoking status stratified by race

Taxa	Race	Median relative abundance			*P* value [a]		
		Current-smokers (*N*=592)	Former-smokers (*N*=477)	Never-smokers (*N*=547)	Current- vs Never-smokers	Former- vs Never-smokers	Current- vs Non-smokers [b]
Phylum *Actinobacteria*	African-American	10.30%	10.39%	9.70%	1.10×10^{-4}	0.74	1.04×10^{-5}
	European-American	16.00%	9.91%	8.17%	6.36×10^{-9}	0.64	1.51×10^{-16}
Family *Actinomycetaceae*	African-American	2.88%	2.38%	2.37%	0.03	0.16	3.51×10^{-3}
	European-American	3.94%	2.74%	2.60%	2.26×10^{-3}	0.87	3.42×10^{-6}
Genus *Actinomyces*	African-American	2.86%	2.24%	2.29%	0.03	0.15	2.58×10^{-3}
	European-American	3.93%	2.65%	2.59%	2.02×10^{-3}	0.95	2.55×10^{-6}
Species *Actinomyces graevenitzii*	African-American	0.34%	0.17%	0.17%	6.20×10^{-6}	0.58	1.31×10^{-7}
	European-American	0.38%	0.12%	0.11%	2.38×10^{-4}	0.81	4.04×10^{-8}
Species *Actinomyces odontolyticus*	African-American	1.37%	0.87%	0.94%	1.48×10^{-3}	0.60	4.68×10^{-5}
	European-American	1.80%	1.10%	0.97%	1.13×10^{-3}	0.77	3.65×10^{-6}
Genus *Rothia*	African-American	5.32%	5.79%	5.38%	8.29×10^{-4}	0.45	2.22×10^{-4}
	European-American	10.08%	4.85%	3.81%	4.89×10^{-6}	0.65	3.54×10^{-11}
Species *Rothia mucilaginosa*	African-American	4.60%	5.01%	4.49%	8.51×10^{-4}	0.35	2.12×10^{-4}
	European-American	8.94%	4.06%	2.88%	4.24×10^{-7}	0.23	2.54×10^{-12}
Family *Coriobacteriaceae*	African-American	0.14%	0.11%	0.15%	1.18×10^{-3}	0.26	7.72×10^{-5}
	European-American	0.11%	0.07%	0.08%	0.03	0.40	4.51×10^{-5}
Genus *Atopobium*	African-American	0.14%	0.11%	0.15%	1.73×10^{-3}	0.25	1.18×10^{-4}
	European-American	0.10%	0.07%	0.07%	0.03	0.38	4.68×10^{-5}
Species *Atopobium parvulum*	African-American	0.13%	0.10%	0.13%	2.11×10^{-3}	0.21	1.52×10^{-4}
	European-American	0.09%	0.06%	0.06%	0.02	0.42	1.11×10^{-5}
Phylum *Proteobacteria*	African-American	2.68%	5.92%	7.26%	2.26×10^{-12}	0.57	4.67×10^{-11}
	European-American	1.60%	6.79%	7.06%	1.04×10^{-8}	0.77	6.11×10^{-11}
Family *Neisseriaceae*	African-American	0.10%	0.58%	0.95%	9.50×10^{-13}	0.32	2.54×10^{-10}
	European-American	0.04%	1.19%	1.35%	8.30×10^{-13}	0.24	1.17×10^{-15}
Genus *Neisseria*	African-American	0.09%	0.54%	0.88%	1.91×10^{-12}	0.38	3.40×10^{-10}
	European-American	0.03%	1.15%	1.28%	2.51×10^{-12}	0.27	3.24×10^{-15}
Species *Neisseria pharyngis*	African-American	0.01%	0.08%	0.10%	5.96×10^{-9}	0.68	2.34×10^{-7}
	European-American	0.01%	0.09%	0.11%	1.65×10^{-7}	0.50	7.59×10^{-7}
Species *Neisseria subflava*	African-American	0.03%	0.30%	0.54%	8.01×10^{-11}	0.33	1.20×10^{-8}
	European-American	0.01%	0.67%	0.76%	1.23×10^{-11}	0.40	2.50×10^{-14}
Family *Pasteurellaceae*	African-American	1.86%	3.82%	4.18%	5.82×10^{-9}	0.96	8.36×10^{-9}
	European-American	1.32%	3.86%	4.33%	5.93×10^{-6}	0.78	4.58×10^{-7}
Genus *Aggregatibacter*	African-American	0.10%	0.13%	0.19%	1.31×10^{-3}	0.22	0.02
	European-American	0.01%	0.05%	0.12%	5.17×10^{-3}	0.50	3.66×10^{-3}
Genus *Haemophilus*	African-American	1.52%	3.51%	3.69%	6.39×10^{-9}	0.83	3.78×10^{-9}
	European-American	1.13%	3.43%	4.10%	7.07×10^{-6}	0.70	6.46×10^{-7}
Species *Haemophilus parahaemolyticus*	African-American	0.16%	0.47%	0.48%	5.21×10^{-10}	0.43	2.35×10^{-10}
	European-American	0.16%	0.54%	0.43%	2.93×10^{-5}	0.61	3.48×10^{-6}
Species *Haemophilus paraphrohaemolyticus*	African-American	1.27%	2.83%	2.96%	6.59×10^{-8}	0.95	9.05×10^{-8}
	European-American	0.99%	2.68%	3.14%	2.94×10^{-5}	0.58	4.26×10^{-6}

Phylum *Bacteroidetes*							
Species *Prevotella sp. oral taxon 313*	African-American	4.74%	3.64%	3.98%	8.31×10^{-3}	0.73	2.86×10^{-3}
	European-American	3.41%	2.05%	1.73%	6.90×10^{-3}	0.29	1.64×10^{-4}
Family *Flavobacteriaceae*	African-American	0.05%	0.10%	0.10%	1.03×10^{-3}	0.94	2.42×10^{-3}
	European-American	0.03%	0.09%	0.07%	0.03	0.02	0.01
Phylum *Firmicutes*							
Genus *Gemella*	African-American	1.46%	1.85%	2.29%	7.02×10^{-7}	0.04	1.26×10^{-4}
	European-American	0.95%	1.99%	2.47%	8.41×10^{-9}	0.31	8.99×10^{-10}
Species *Streptococcus oligofermentans*	African-American	0.13%	0.35%	0.31%	9.93×10^{-13}	0.67	7.49×10^{-15}
	European-American	0.13%	0.61%	0.79%	6.23×10^{-8}	0.56	5.42×10^{-9}
Species *Streptococcus sp. oral taxon 057*	African-American	9.61%	8.72%	8.16%	7.79×10^{-4}	0.36	5.33×10^{-3}
	European-American	12.47%	9.52%	7.79%	7.73×10^{-9}	0.18	1.90×10^{-9}
Species *Streptococcus sp. oral taxon 070*	African-American	24.82%	21.80%	21.43%	0.31	0.73	0.47
	European-American	20.27%	25.33%	30.50%	4.31×10^{-5}	0.23	6.70×10^{-7}
Genus *Megasphaera*	African-American	0.26%	0.12%	0.14%	1.09×10^{-5}	0.37	4.86×10^{-7}
	European-American	0.25%	0.10%	0.09%	4.52×10^{-3}	0.79	1.63×10^{-6}
Species *Megasphaera micronuciformis*	African-American	0.24%	0.11%	0.13%	4.20×10^{-5}	0.38	2.23×10^{-6}
	European-American	0.25%	0.09%	0.09%	6.58×10^{-3}	0.91	5.84×10^{-6}

[a] *P* values were calculated from logistic regression. Sequencing batch as well as other covariates (age, sex, BMI, alcohol consumption, oral health and disease status at the 1st follow-up and total energy intake) were adjusted.
[b] Non-smokers included former- and never-smokers.

Table S3. Associations of rare bacterial taxa with smoking status stratified by race

Taxa	Race	Prevalence			*P* value [a]		
		Current-smokers (*N*=592)	Former-smokers (*N*=477)	Never-smokers (*N*=547)	Current- vs Never-smokers	Former- vs Never-smokers	Current- vs Non-smokers [b]
Phylum *Actinobacteria*							
Family *Bifidobacteriaceae*	African-American	95.17%	88.59%	84.25%	6.15×10^{-4}	0.37	3.95×10^{-3}
	European-American	96.07%	91.59%	85.54%	0.03	0.15	0.02
Species *Actinomyces lingnae_[NVP]*	African-American	92.75%	88.59%	87.40%	3.20×10^{-5}	0.59	1.32×10^{-5}
	European-American	91.57%	87.38%	84.94%	0.06	0.34	0.01
Phylum *Proteobacteria*							
Family *Burkholderiaceae*	African-American	45.41%	61.60%	66.67%	9.58×10^{-10}	0.11	2.95×10^{-9}
	European-American	36.52%	65.42%	77.11%	1.76×10^{-7}	0.63	6.61×10^{-9}
Genus *Lautropia*	African-American	43.72%	61.22%	66.14%	3.11×10^{-11}	0.12	5.25×10^{-11}
	European-American	36.52%	64.49%	77.11%	1.76×10^{-7}	0.53	1.12×10^{-8}
Genus *Kingella*	African-American	70.29%	77.95%	82.94%	1.62×10^{-3}	0.16	2.51×10^{-3}
	European-American	60.11%	82.24%	90.36%	5.52×10^{-6}	0.53	4.05×10^{-7}
Species *Kingella denitrificans*	African-American	27.29%	41.44%	44.36%	7.09×10^{-5}	0.53	1.19×10^{-4}
	European-American	19.10%	50.00%	51.81%	3.24×10^{-5}	0.68	3.86×10^{-7}
Species *Kingella elongata*	African-American	65.22%	74.14%	78.22%	1.95×10^{-3}	0.26	2.31×10^{-3}
	European-American	54.49%	80.37%	86.14%	1.79×10^{-5}	0.67	2.38×10^{-8}
Species *Neisseria oralis*	African-American	21.98%	38.02%	50.92%	1.69×10^{-12}	0.00	1.94×10^{-11}
	European-American	15.17%	49.53%	58.43%	4.96×10^{-9}	0.60	7.20×10^{-11}
Genus *Cardiobacterium*	African-American	35.99%	52.85%	54.59%	1.38×10^{-6}	0.89	1.73×10^{-5}
	European-American	30.90%	61.68%	70.48%	8.44×10^{-6}	0.48	1.20×10^{-7}
Phylum *Bacteroidetes*							
Species *Prevotella nanceiensis*	African-American	85.99%	86.69%	87.14%	0.26	0.95	0.39
	European-American	68.54%	85.51%	94.58%	5.88×10^{-6}	0.07	5.99×10^{-6}
Species *Capnocytophaga sputigena*	African-American	42.75%	59.32%	60.89%	2.55×10^{-6}	0.94	5.15×10^{-6}
	European-American	43.82%	62.62%	66.27%	6.77×10^{-3}	0.74	2.48×10^{-3}
Phylum *Firmicutes*							
Family *Lactobacillaceae*	African-American	88.65%	77.19%	78.22%	0.01	0.41	3.69×10^{-3}
	European-American	94.38%	76.17%	74.70%	1.68×10^{-3}	0.78	5.29×10^{-4}
Genus *Enterococcus*	African-American	65.46%	66.54%	72.44%	3.15×10^{-3}	0.10	0.02
	European-American	57.87%	77.10%	79.52%	5.71×10^{-4}	0.43	1.92×10^{-4}
Genus *Lachnospiraceae_[G-2]*	African-American	57.25%	54.75%	60.63%	3.67×10^{-3}	0.01	0.01
	European-American	38.76%	50.47%	63.86%	0.03	0.58	0.05
Species *Lachnoanaerobaculum umeaense*	African-American	76.09%	82.13%	81.63%	0.02	0.92	9.38×10^{-3}
	European-American	65.73%	81.31%	85.54%	0.02	0.49	6.84×10^{-3}
Species *Eubacterium infirmum*	African-American	66.18%	71.10%	75.59%	0.06	0.07	0.12
	European-American	52.81%	72.43%	82.53%	5.43×10^{-5}	0.45	1.54×10^{-4}
Phylum *Spirochaetes*							
Species *Treponema denticola*	African-American	77.78%	56.27%	56.96%	4.00×10^{-4}	0.60	3.01×10^{-5}
	European-American	51.69%	40.19%	48.19%	0.54	0.37	0.19

[a] *P* values were calculated from logistic regression. Sequencing batch as well as other covariates (age, sex, BMI, alcohol consumption, oral health and disease status at the 1st follow-up and total energy intake) were adjusted.

[b] Non-smokers included former- and never-smokers.

Table S4. Associations of probiotic taxa with smoking status stratified by sequencing batch

Probiotic taxa	Race	Prevalence			*P* value [a]		
		Current-smokers (*N*=592)	Former-smokers (*N*=477)	Never-smokers (*N*=547)	Current- vs Never-smokers	Former- vs Never-smokers	Current- vs Non-smokers [b]
Phylum *Actinobacteria*							
Genus *Bifidobacterium*	Frist batch	84.24%	75.09%	70.18%	0.02	0.60	4.70×10^{-3}
	Second batch	88.71%	66.04%	64.12%	1.77×10^{-5}	0.64	6.64×10^{-6}
Species *Bifidobacterium longum*	Frist batch	69.95%	50.57%	43.16%	1.90×10^{-5}	0.66	2.62×10^{-6}
	Second batch	62.37%	38.21%	35.88%	7.77×10^{-7}	0.39	6.73×10^{-7}
Phylum *Firmicutes*							
Genus *Lactobacillus*	Frist batch	90.89%	77.36%	76.49%	4.70×10^{-3}	0.34	2.27×10^{-4}
	Second batch	86.02%	68.40%	70.23%	7.06×10^{-4}	0.91	5.48×10^{-4}
Species *Lactobacillus crispatus*	Frist batch	60.59%	37.36%	37.19%	1.84×10^{-5}	0.24	8.21×10^{-7}
	Second batch	62.37%	33.02%	31.30%	1.28×10^{-6}	0.74	7.33×10^{-8}
Species *Lactobacillus fermentum*	Frist batch	61.33%	42.64%	34.04%	4.45×10^{-6}	0.31	7.56×10^{-6}
	Second batch	49.46%	36.32%	37.40%	0.10	0.59	0.03
Species *Lactobacillus gasseri*	Frist batch	73.65%	63.02%	56.14%	0.04	0.78	5.28×10^{-3}
	Second batch	69.89%	48.11%	49.62%	1.98×10^{-4}	0.96	1.07×10^{-4}
Species *Lactobacillus oris*	Frist batch	44.58%	29.06%	20.35%	2.19×10^{-6}	0.08	2.30×10^{-5}
	Second batch	41.94%	24.06%	21.37%	6.55×10^{-3}	0.40	3.49×10^{-3}
Species *Lactobacillus panis*	Frist batch	36.45%	24.53%	19.65%	3.49×10^{-3}	0.98	3.19×10^{-3}
	Second batch	54.84%	29.72%	31.68%	1.21×10^{-4}	0.90	6.00×10^{-5}
Species *Lactobacillus reuteri*	Frist batch	43.10%	27.55%	21.75%	4.38×10^{-5}	0.97	2.74×10^{-5}
	Second batch	38.17%	17.92%	19.85%	8.25×10^{-4}	0.68	3.28×10^{-5}

[a] *P* values were calculated from logistic regression. Covariates including age, sex, race, BMI, alcohol consumption, oral health and disease status at the 1st follow-up and total energy intake were adjusted.

[b] Non-smokers included former- and never-smokers.

Table S5. Associations of common taxa with smoking status stratified by sequencing batch

Taxa	Race	Median relative abundance			*P* value [a]		
		Current-smokers (*N*=592)	Former-smokers (*N*=477)	Never-smokers (*N*=547)	Current- vs Never-smokers	Former- vs Never-smokers	Current- vs Non-smokers [b]
Phylum *Actinobacteria*	First batch	11.90%	7.89%	6.55%	6.52×10^{-13}	0.43	2.52×10^{-17}
	Second batch	13.03%	12.13%	11.40%	0.06	0.88	0.02
Family *Actinomycetaceae*	First batch	3.36%	2.67%	2.56%	8.03×10^{-4}	0.53	3.53×10^{-6}
	Second batch	2.49%	2.29%	2.36%	0.06	0.57	0.05
Genus *Actinomyces*	First batch	3.34%	2.60%	2.56%	6.65×10^{-4}	0.47	2.01×10^{-6}
	Second batch	2.48%	2.23%	2.25%	0.05	0.56	0.04
Species *Actinomyces graevenitzii*	First batch	0.30%	0.11%	0.11%	1.27×10^{-8}	0.39	8.28×10^{-15}
	Second batch	0.45%	0.22%	0.28%	7.97×10^{-3}	0.97	5.57×10^{-3}
Species *Actinomyces odontolyticus*	First batch	1.81%	1.12%	1.11%	3.13×10^{-4}	0.72	2.30×10^{-7}
	Second batch	1.04%	0.82%	0.78%	1.43×10^{-3}	0.52	4.62×10^{-3}
Genus *Rothia*	First batch	6.24%	3.54%	2.68%	1.50×10^{-11}	0.19	2.56×10^{-14}
	Second batch	7.61%	7.53%	7.28%	0.39	0.85	0.17
Species *Rothia mucilaginosa*	First batch	5.42%	3.01%	2.16%	6.82×10^{-12}	0.26	1.93×10^{-15}
	Second batch	6.36%	6.38%	6.14%	0.23	0.49	0.15
Family *Coriobacteriaceae*	First batch	0.08%	0.05%	0.05%	0.01	0.04	1.27×10^{-5}
	Second batch	0.36%	0.20%	0.22%	8.91×10^{-4}	0.89	6.50×10^{-4}
Genus *Atopobium*	First batch	0.08%	0.05%	0.05%	0.02	0.03	2.21×10^{-5}
	Second batch	0.36%	0.20%	0.22%	8.97×10^{-4}	0.88	6.46×10^{-4}
Species *Atopobium parvulum*	First batch	0.07%	0.04%	0.04%	0.02	0.04	2.09×10^{-5}
	Second batch	0.26%	0.17%	0.18%	8.16×10^{-4}	0.85	4.43×10^{-4}
Phylum *Proteobacteria*	First batch	1.81%	4.36%	5.07%	2.12×10^{-10}	0.41	4.94×10^{-9}
	Second batch	3.51%	9.00%	9.19%	6.80×10^{-13}	0.76	1.04×10^{-14}
Family *Neisseriaceae*	First batch	0.04%	0.47%	0.72%	1.62×10^{-10}	0.72	4.09×10^{-10}
	Second batch	0.13%	1.56%	1.82%	3.30×10^{-16}	0.97	3.78×10^{-16}
Genus *Neisseria*	First batch	0.04%	0.44%	0.67%	2.80×10^{-10}	0.78	5.22×10^{-10}
	Second batch	0.11%	1.50%	1.59%	1.05×10^{-15}	0.95	1.01×10^{-15}
Species *Neisseria pharyngis*	First batch	0.01%	0.03%	0.06%	1.22×10^{-5}	0.32	2.71×10^{-4}
	Second batch	0.03%	0.29%	0.27%	9.03×10^{-12}	0.45	3.96×10^{-11}
Species *Neisseria subflava*	First batch	0.02%	0.24%	0.41%	7.15×10^{-10}	0.80	2.20×10^{-9}
	Second batch	0.04%	0.77%	1.02%	1.40×10^{-13}	0.67	1.16×10^{-13}
Family *Pasteurellaceae*	First batch	1.48%	2.97%	3.59%	4.40×10^{-8}	0.39	7.81×10^{-7}
	Second batch	2.21%	5.38%	5.52%	8.34×10^{-8}	0.61	7.51×10^{-10}
Genus *Aggregatibacter*	First batch	0.05%	0.06%	0.15%	1.79×10^{-4}	0.15	4.07×10^{-3}
	Second batch	0.10%	0.20%	0.18%	0.06	0.89	0.05
Genus *Haemophilus*	First batch	1.14%	2.54%	3.08%	7.68×10^{-8}	0.48	8.00×10^{-7}
	Second batch	1.78%	4.65%	4.58%	5.96×10^{-8}	0.58	4.45×10^{-10}
Species *Haemophilus parahaemolyticus*	First batch	0.14%	0.37%	0.41%	1.27×10^{-7}	0.93	6.92×10^{-7}
	Second batch	0.19%	0.62%	0.60%	1.61×10^{-9}	0.18	4.90×10^{-11}
Species *Haemophilus paraphrohaemolyticus*	First batch	0.99%	2.07%	2.61%	4.50×10^{-7}	0.36	7.67×10^{-6}
	Second batch	1.47%	3.72%	3.59%	9.53×10^{-7}	0.74	1.20×10^{-8}

Phylum *Bacteroidetes*							
Species *Prevotella sp. oral taxon 313*	First batch	3.35%	1.74%	1.75%	4.00×10^{-3}	0.98	1.87×10^{-4}
	Second batch	5.92%	4.37%	4.38%	0.02	0.19	0.02
Family *Flavobacteriaceae*	First batch	0.03%	0.04%	0.04%	0.02	0.21	0.01
	Second batch	0.14%	0.23%	0.20%	1.44×10^{-3}	0.58	1.42×10^{-3}
Phylum *Firmicutes*							
Genus *Gemella*	First batch	1.16%	1.70%	2.14%	5.42×10^{-7}	0.19	2.61×10^{-5}
	Second batch	1.42%	2.29%	2.48%	4.79×10^{-9}	0.07	8.01×10^{-8}
Species *Streptococcus oligofermentans*	First batch	0.12%	0.55%	0.62%	1.90×10^{-14}	0.73	3.83×10^{-16}
	Second batch	0.15%	0.36%	0.29%	1.21×10^{-6}	0.49	1.98×10^{-6}
Species *Streptococcus sp. oral taxon 057*	First batch	11.15%	10.16%	8.51%	1.34×10^{-6}	0.11	7.28×10^{-5}
	Second batch	9.02%	8.00%	7.56%	3.56×10^{-6}	0.55	1.25×10^{-6}
Species *Streptococcus sp. oral taxon 070*	First batch	26.64%	29.70%	32.86%	5.03×10^{-4}	0.37	9.16×10^{-4}
	Second batch	16.75%	18.87%	18.05%	0.22	0.60	0.13
Genus *Megasphaera*	First batch	0.26%	0.08%	0.10%	1.16×10^{-4}	0.21	1.01×10^{-7}
	Second batch	0.25%	0.13%	0.15%	6.28×10^{-5}	0.71	5.75×10^{-5}
Species *Megasphaera micronuciformis*	First batch	0.25%	0.08%	0.09%	2.71×10^{-4}	0.18	3.71×10^{-7}
	Second batch	0.21%	0.12%	0.14%	2.82×10^{-4}	0.72	3.06×10^{-4}

[a] *P* values were calculated from logistic regression. Covariates including age, sex, race, BMI, alcohol consumption, oral health and disease status at the 1st follow-up and total energy intake were adjusted.

[b] Non-smokers included former- and never-smokers.

Table S6. Associations of rare bacterial taxa with smoking status sequencing batch

Taxa	Race	Prevalence			*P* value [a]		
		Current-smokers (*N*=592)	Former-smokers (*N*=477)	Never-smokers (*N*=547)	Current- vs Never-smokers	Former- vs Never-smokers	Current- vs Non-smokers [b]
Phylum *Actinobacteria*							
Family *Bifidobacteriaceae*	First batch	95.07%	90.94%	83.51%	4.43×10^{-3}	0.12	5.68×10^{-3}
	Second batch	96.24%	88.68%	85.88%	0.01	0.72	0.02
Species *Actinomyces lingnae_[NVP]*	First batch	90.39%	83.02%	80.35%	2.85×10^{-4}	0.83	3.09×10^{-5}
	Second batch	96.77%	94.34%	93.51%	9.55×10^{-3}	0.28	0.02
Phylum *Proteobacteria*							
Family *Burkholderiaceae*	First batch	41.38%	64.15%	72.28%	7.48×10^{-13}	0.08	4.52×10^{-12}
	Second batch	45.70%	62.26%	67.18%	7.77×10^{-6}	0.40	7.30×10^{-7}
Genus *Lautropia*	First batch	40.64%	63.40%	72.28%	1.38×10^{-13}	0.06	1.08×10^{-12}
	Second batch	43.55%	61.79%	66.41%	1.39×10^{-6}	0.42	7.13×10^{-8}
Genus *Kingella*	First batch	64.53%	78.11%	84.21%	3.12×10^{-5}	0.23	4.79×10^{-6}
	Second batch	73.12%	82.08%	86.26%	1.06×10^{-3}	0.31	1.56×10^{-3}
Species *Kingella denitrificans*	First batch	23.65%	39.62%	42.11%	2.29×10^{-4}	0.48	3.20×10^{-5}
	Second batch	27.42%	52.36%	51.53%	4.36×10^{-7}	0.96	5.76×10^{-8}
Species *Kingella elongata*	First batch	59.61%	76.60%	82.11%	6.88×10^{-6}	0.26	1.27×10^{-7}
	Second batch	67.20%	77.36%	79.01%	0.01	0.92	7.89×10^{-3}
Species *Neisseria oralis*	First batch	20.69%	45.28%	57.19%	1.17×10^{-13}	5.63×10^{-3}	7.76×10^{-14}
	Second batch	18.28%	40.57%	48.85%	7.21×10^{-9}	0.08	4.59×10^{-9}
Genus *Cardiobacterium*	First batch	30.54%	51.32%	56.14%	3.69×10^{-6}	0.22	1.06×10^{-6}
	Second batch	43.01%	63.68%	62.98%	7.23×10^{-6}	0.82	2.32×10^{-6}
Phylum *Bacteroidetes*							
Species *Prevotella nanceiensis*	First batch	82.27%	85.66%	92.63%	7.59×10^{-5}	0.08	3.74×10^{-3}
	Second batch	77.42%	86.79%	85.88%	0.08	0.78	4.75×10^{-3}
Species *Capnocytophaga sputigena*	First batch	42.36%	55.85%	60.35%	4.29×10^{-4}	0.38	4.58×10^{-4}
	Second batch	44.62%	66.98%	64.89%	1.19×10^{-5}	0.44	2.04×10^{-6}
Phylum *Firmicutes*							
Family *Lactobacillaceae*	First batch	91.63%	80.00%	80.35%	0.04	0.22	2.93×10^{-3}
	Second batch	87.63%	72.64%	73.66%	1.70×10^{-3}	0.97	1.31×10^{-3}
Genus *Enterococcus*	First batch	65.27%	74.34%	80.35%	2.98×10^{-5}	0.16	2.47×10^{-4}
	Second batch	58.60%	67.45%	68.32%	0.03	0.79	0.01
Genus *Lachnospiraceae_[G-2]*	First batch	51.72%	55.47%	67.72%	7.53×10^{-4}	5.44×10^{-3}	8.13×10^{-3}
	Second batch	51.61%	49.53%	54.96%	0.16	0.40	0.11
Species *Lachnoanaerobaculum umeaense*	First batch	72.66%	78.49%	84.21%	3.71×10^{-3}	0.32	7.71×10^{-3}
	Second batch	73.66%	85.85%	81.30%	0.09	0.14	2.97×10^{-3}
Species *Eubacterium infirmum*	First batch	58.37%	67.55%	75.44%	3.69×10^{-5}	0.02	2.85×10^{-4}
	Second batch	70.43%	76.89%	80.15%	0.13	0.64	0.11
Phylum *Spirochaetes*							
Species *Treponema denticola*	First batch	67.24%	47.17%	52.63%	0.01	0.18	3.22×10^{-3}
	Second batch	75.81%	51.42%	56.11%	8.03×10^{-3}	0.74	1.59×10^{-3}

[a] *P* values were calculated from logistic regression. Covariates including age, sex, race, BMI, alcohol consumption, oral health and disease status at the 1st follow-up and total energy intake were adjusted.

[b] Non-smokers included former- and never-smokers.

Supplementary Figures

Figure S1. Difference in relative abundance of common taxa between current-smokers and never-smokers. This figure shows the difference of relative abundance of all common taxa (with a relative abundance among never-smokers of >0.10%) between current- and never-smokers. The red color represents that the taxon was more abundant among current-smokers and the blue color represent that the taxon was more abundant among never-smokers. The grey color represents that the taxon didn't showed a differential abundance. The pentagram represents that the difference of relative abundance reached a Bonferroni-corrected threshold.

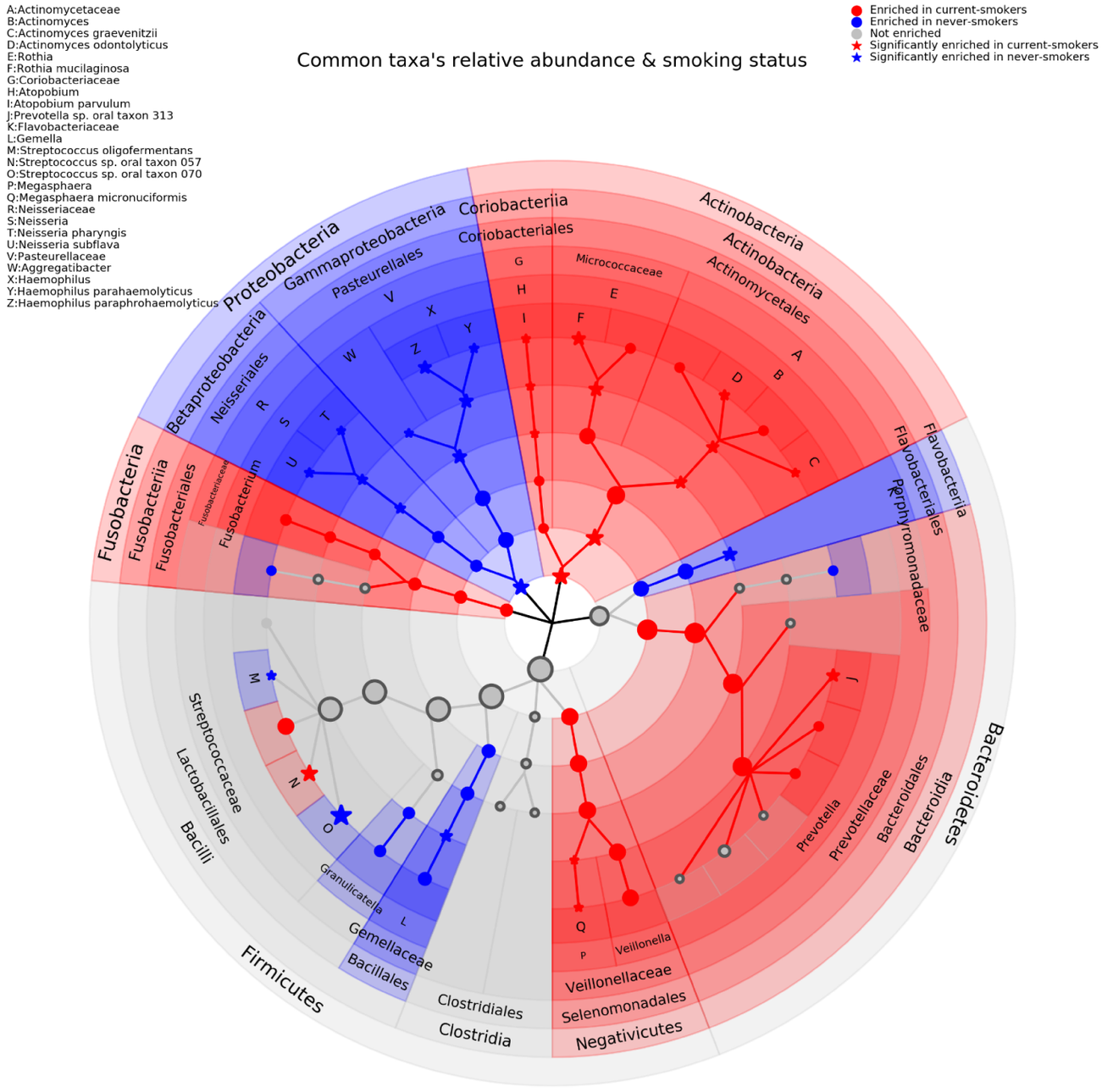
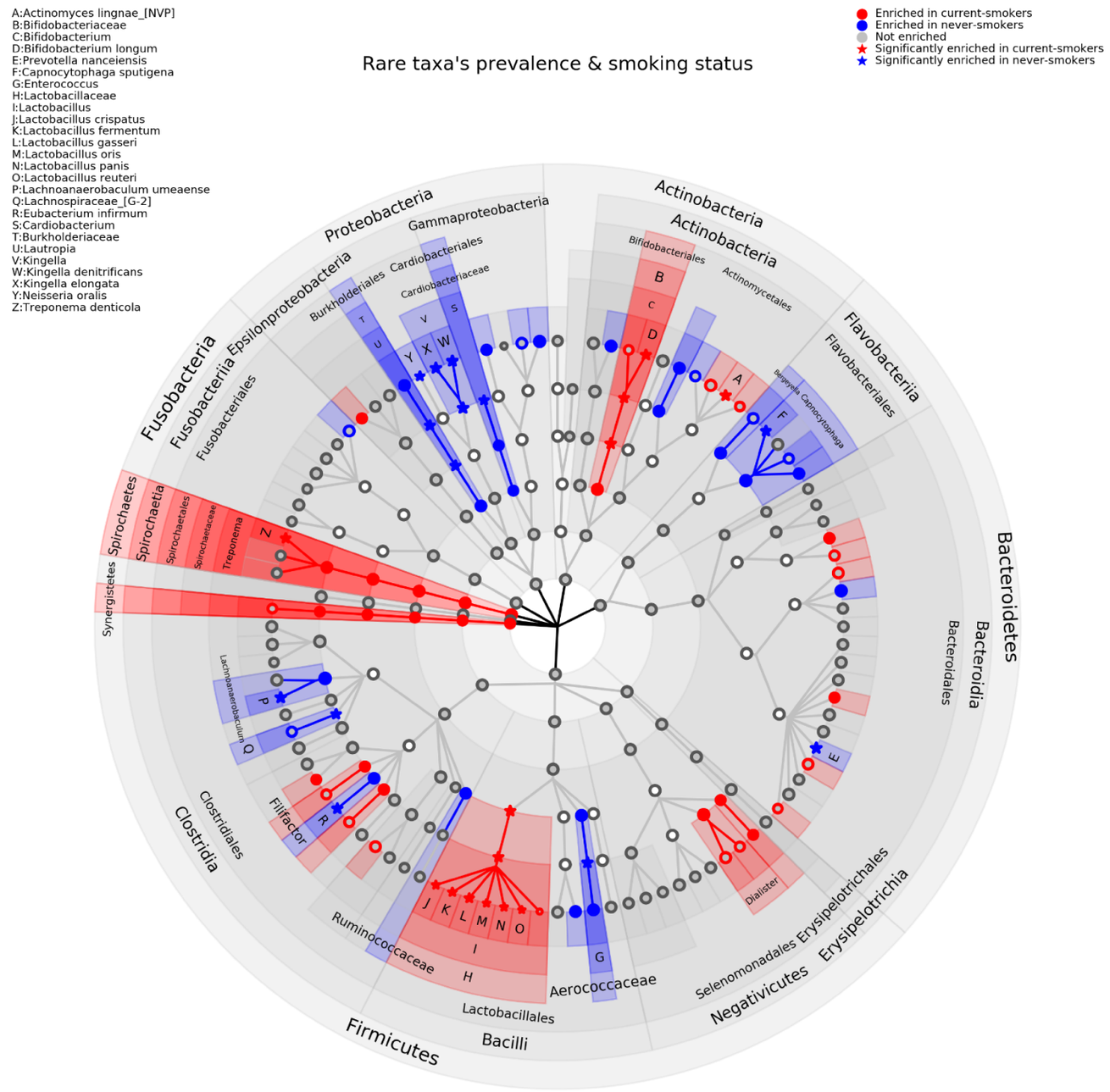

A:Actinomycetaceae
B:Actinomyces
C:Actinomyces graevenitzii
D:Actinomyces odontolyticus
E:Rothia
F:Rothia mucilaginosa
G:Coriobacteriaceae
H:Atopobium
I:Atopobium parvulum
J:Prevotella sp. oral taxon 313
K:Flavobacteriaceae
L:Gemella
M:Streptococcus oligofermentans
N:Streptococcus sp. oral taxon 057
O:Streptococcus sp. oral taxon 070
P:Megasphaera
Q:Megasphaera micronuciformis
R:Neisseriaceae
S:Neisseria
T:Neisseria pharyngis
U:Neisseria subflava
V:Pasteurellaceae
W:Aggregatibacter
X:Haemophilus
Y:Haemophilus parahaemolyticus
Z:Haemophilus paraphrohaemolyticus
Common taxa's relative abundance & smoking status
Enriched in current-smokers
Enriched in never-smokers
Not enriched
Significantly enriched in current-smokers
Significantly enriched in never-smokers
Actinobacteria
Actinobacteria
Actinomycetales
Micrococcaceae
Coriobacteriia
Coriobacteriales
Proteobacteria
Gammaproteobacteria
Pasteurellales
Betaproteobacteria
Neisseriales
Fusobacteria
Fusobacteriia
Fusobacteriales
Fusobacteriaceae
Fusobacterium
Bacteroidetes
Flavobacteriia
Flavobacteriales
Porphyromonadaceae
Prevotellaceae
Prevotella
Bacteroidales
Bacteroidia
Firmicutes
Bacilli
Lactobacillales
Streptococcaceae
Granulicatella
Gemellaceae
Bacillales
Clostridia
Clostridiales
Negativicutes
Selenomonadales
Veillonellaceae
Veillonella

Figure S2. Difference in prevalence of probiotic taxa and rare taxa between current-smokers and never-smokers. This figure shows the difference of relative abundance of probiotic taxa and rare taxa (with a relative abundance among never-smokers of ≤0.10% and a prevalence among never-smokers of >30%) between current- and never-smokers. The red color represents that the taxon was more prevalent among current-smokers and the blue color represent that the taxon was more prevalent among never-smokers. The grey color represents that the taxon didn't showed a differential prevalence. The pentagram represents that the difference of prevalence reached a Bonferroni-corrected threshold.


A:Actinomyces lingnae_[NVP]
B:Bifidobacteriaceae
C:Bifidobacterium
D:Bifidobacterium longum
E:Prevotella nanceiensis
F:Capnocytophaga sputigena
G:Enterococcus
H:Lactobacillaceae
I:Lactobacillus
J:Lactobacillus crispatus
K:Lactobacillus fermentum
L:Lactobacillus gasseri
M:Lactobacillus oris
N:Lactobacillus panis
O:Lactobacillus reuteri
P:Lachnoanaerobaculum umeaense
Q:Lachnospiraceae_[G-2]
R:Eubacterium infirmum
S:Cardiobacterium
T:Burkholderiaceae
U:Lautropia
V:Kingella
W:Kingella denitrificans
X:Kingella elongata
Y:Neisseria oralis
Z:Treponema denticola
Rare taxa's prevalence & smoking status
Enriched in current-smokers
Enriched in never-smokers
Not enriched
Significantly enriched in current-smokers
Significantly enriched in never-smokers
Proteobacteria
Gammaproteobacteria
Actinobacteria
Actinobacteria
Flavobacteriia
Flavobacteriales
Epsilonproteobacteria
Fusobacteria
Fusobacteriia
Fusobacteriales
Bacteroidetes
Bacteroidia
Bacteroidales
Spirochaetes
Spirochaetia
Spirochaetales
Synergistetes
Firmicutes
Clostridia
Clostridiales
Bacilli
Lactobacillales
Negativicutes
Selenomonadales
Erysipelotrichia
Erysipelotrichales
Burkholderiales
Cardiobacteriales
Cardiobacteriaceae
Bifidobacteriales
Actinomycetales
Capnocytophaga
Spirochaetaceae
Treponema
Lachnoanaerobaculum
Filifactor
Ruminococcaceae
Aerococcaceae
Dialister