**SUPPLEMENTARY INFORMATION**

**Oral microbiome composition reflects prospective risk for esophageal cancers**

Brandilyn A. Peters,1 Jing Wu,1,2 Zhiheng Pei,2,3,4 Liying Yang,5 Mark P. Purdue,6 Neal D. Freedman,6 Eric J. Jacobs,7 Susan M. Gapstur,7 Richard B. Hayes,1,2 Jiyoung Ahn1,2,\*

1. Division of Epidemiology, Department of Population Health, NYU School of Medicine, New York, NY, USA

2. NYU Perlmutter Cancer Center, New York, NY, USA

3. Department of Pathology, NYU School of Medicine, New York, NY, USA

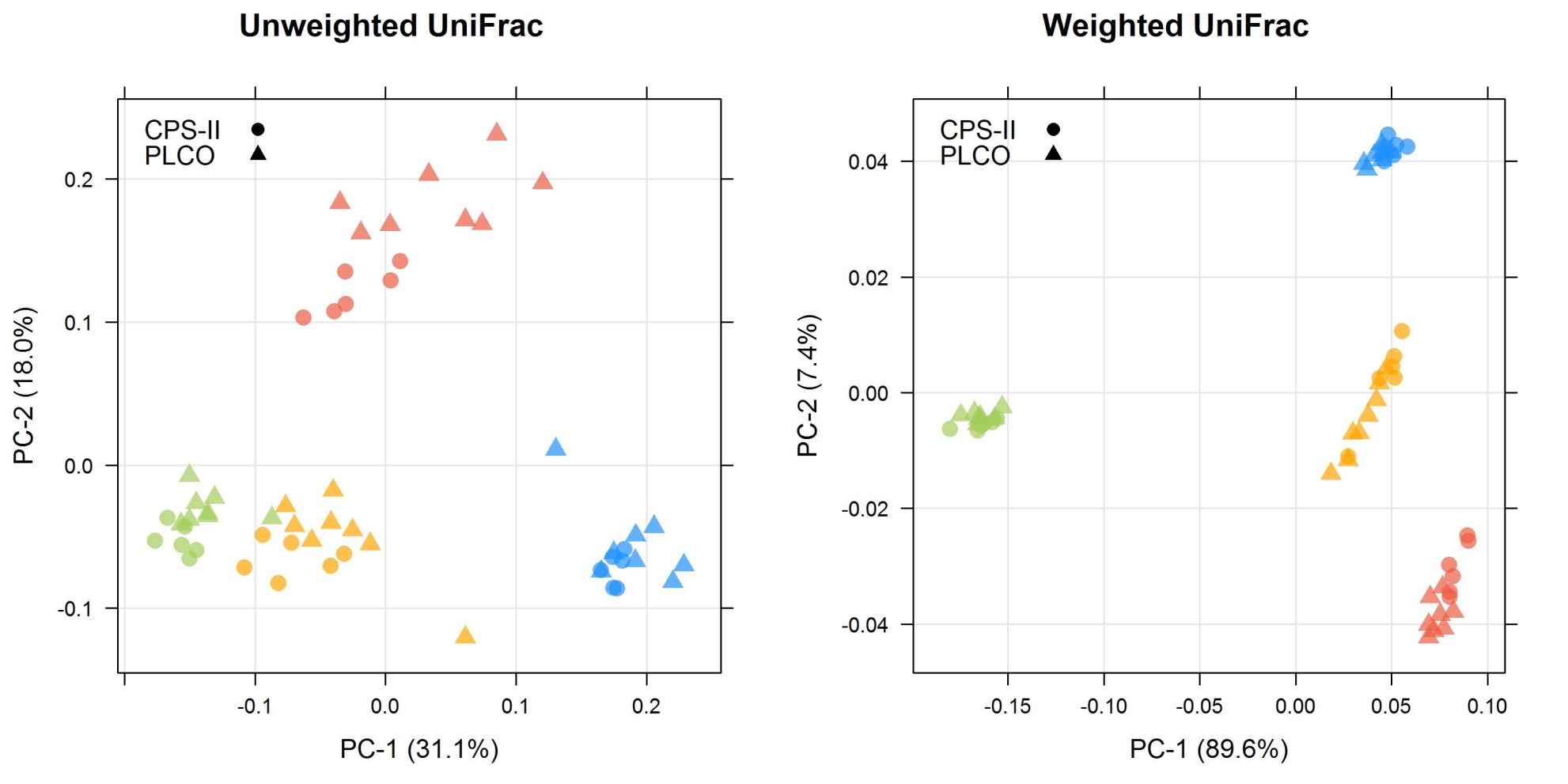
4. Department of Veterans Affairs New York Harbor Healthcare System, New York, NY, USA

5. Division of Translational Medicine, Department of Medicine, NYU School of Medicine, New York, NY, USA

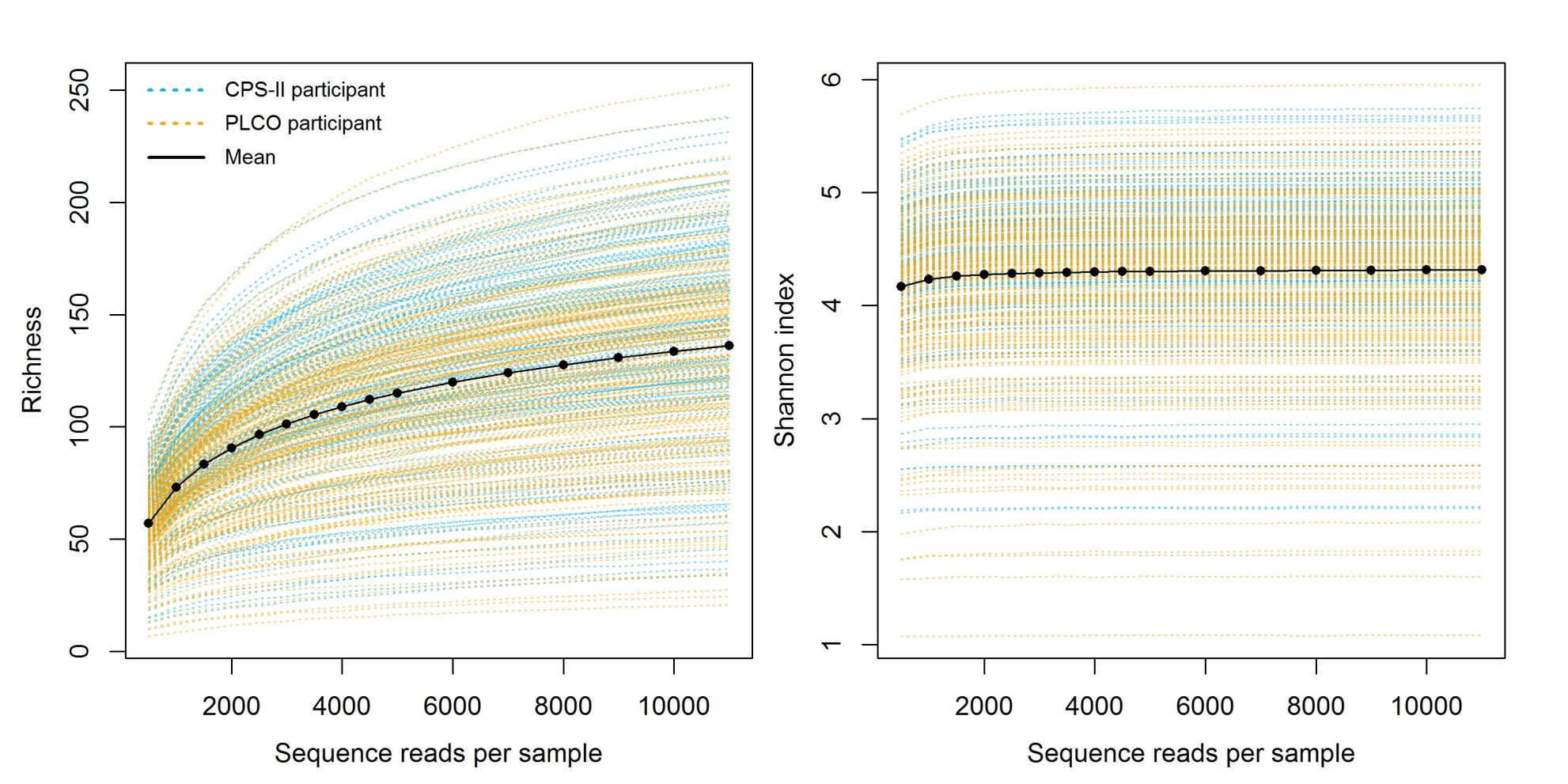
6. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA

7. Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA

\*Correspondence should be addressed to Jiyoung Ahn (650 1st Avenue, 5th Floor, New York, NY, 10016; phone: 212-263-3390; fax: 212-263-8570; email: Jiyoung.Ahn@nyumc.org)

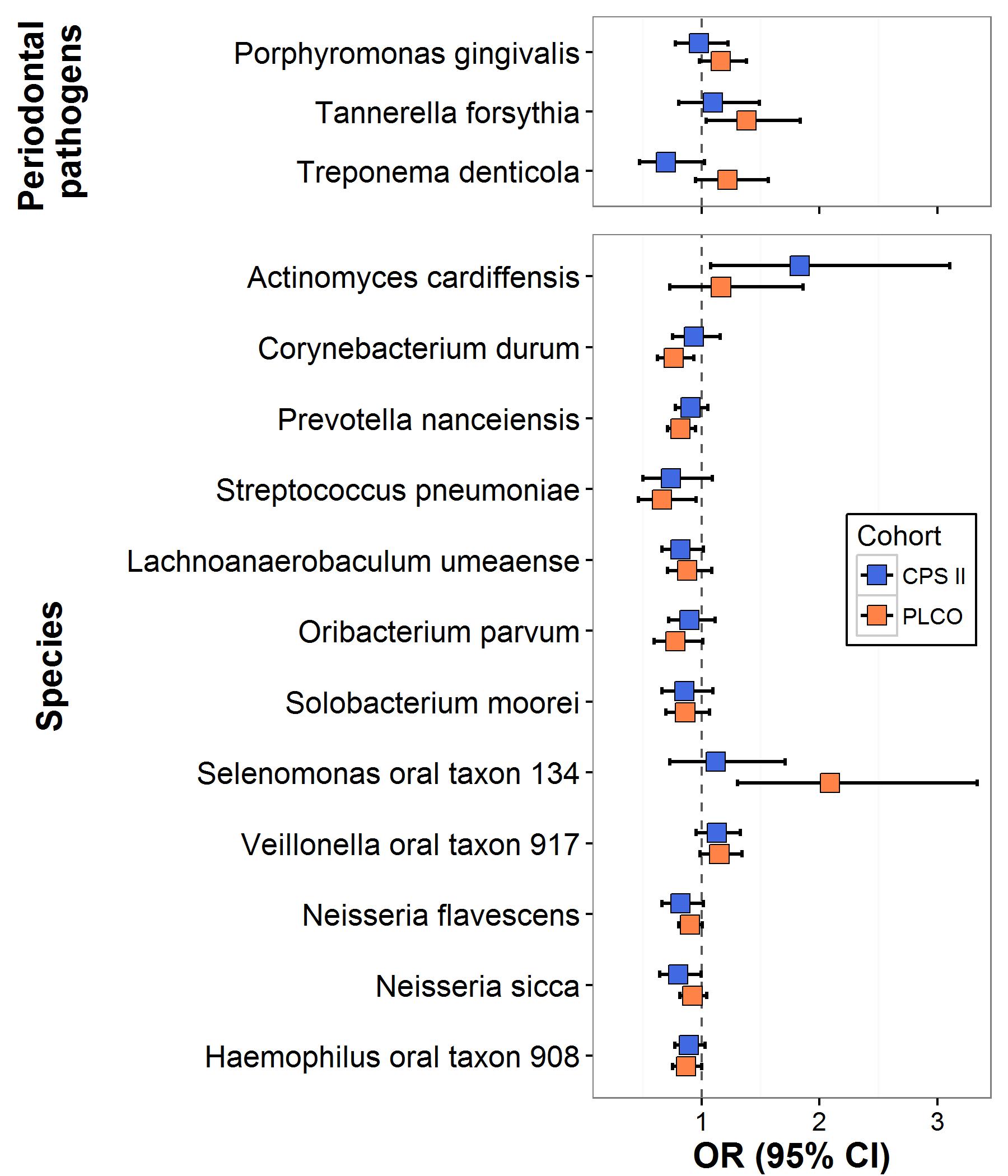
****

**Supplementary Figure 1.** Principal coordinate analysis (PCoA) of the unweighted and weighted UniFrac distances for quality control oral wash samples provided by 4 volunteers. The 4 volunteers are differentiated by color, with triangles representing the replicates sequenced in the PLCO sequencing batch, and circles representing the replicates sequenced in the CPS-II sequencing batch.

****

**Supplementary Figure 2.** Rarefaction curves of richness and the Shannon index. These indices were calculated for 100 iterations of rarefied OTU tables, and the average over the iterations was taken for each participant. n=356 participants with library sizes ≥ 11,000 reads are included in the plot.

**Supplementary Figure 3. Forest plot of odds ratios (OR) and 95% confidence intervals (95% CI) for associations of clr-transformed species abundance with EAC risk in the CPS-II and PLCO cohorts.** See Supplementary Table 4 for numeric display of the OR (95% CI) estimates.



**Supplementary Table 1.** Quality control sample coefficients of variation (CV) and intra‐class correlation coefficients (ICC) for the Shannon diversity index (from 100 iterations of rarefied OTU tables of 4500 reads/sample) and relative abundance of major oral phyla. Six replicates from four volunteers were included in the CPS-II sequencing batch, and eight replicates from each of the same four volunteers were included in the PLCO batch, for a total of 56 quality control samples.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | CV | | | | ICC |
|  | Subject 1 | Subject 2 | Subject 3 | Subject 4 |  |
| Shannon index | 0.86 | 3.89 | 3.09 | 3.05 | 0.99 |
| Firmicutes | 2.84 | 2.5 | 1.34 | 1.39 | 0.99 |
| Bacteroidetes | 4.96 | 10.37 | 7.19 | 7.16 | 0.99 |
| Proteobacteria | 3.91 | 9.71 | 6.49 | 6.56 | 0.98 |
| Actinobacteria | 9.82 | 12.6 | 18.69 | 12.33 | 0.94 |
| Fusobacteria | 6.8 | 9.87 | 8.55 | 9.47 | 0.99 |

**Supplementary Table 2.** α- and β-diversity in relation to esophageal adenocarcinoma and squamous cell carcinoma.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Adenocarcinomaa** | | **Squamous Cell Carcinomaa** | |
| **α-diversityb,c** |  |  |  |  |
|  | **OR (95% CI)** | **p** | **OR (95% CI)** | **p** |
| **Richness** | 1.00 (0.99, 1.01) | 0.97 | 1.00 (0.98, 1.02) | 0.86 |
| **Shannon index** | 0.97 (0.65, 1.45) | 0.89 | 1.43 (0.53, 3.88) | 0.48 |
| **β-diversityd,c** |  |  |  |  |
|  | **R2 (%)e** | **p** | **R2 (%)e** | **p** |
| **Unweighted UniFrac** | 0.39 | 0.34 | 0.78 | 0.85 |
| **Weighted UniFrac** | 0.43 | 0.33 | 1.12 | 0.39 |

aAdenocarcinoma includes 81 EAC cases 160 matched controls, and Squamous Cell Carcinoma includes 25 ESCC cases and 50 matched controls.

bRichness and Shannon index were estimated by taking the mean for each participant over 100 iterations of rarefied OTU tables of 4,500 sequence reads per sample. Odds ratios are from conditional logistic regression with case/control status as the binary outcome and richness or Shannon index as the predictor of interest.

cAll models used matched sets as strata and adjusted for smoking status, BMI category, and alcohol drinking level. Missing values in BMI category and drinking level were filled in by multiple imputation (mice package, R); model results from 10 imputations were pooled to yield final model results.

dPERMANOVA of the unweighted or weighted UniFrac distance using 'adonis' function (vegan package, R); all adjustment factors were included in the model first before case/control status.

ePercent variation in overall microbiome composition explained by case/control status.

**Supplementary Table 3.** Oral taxaa associated with EAC, stratified by years to case diagnosis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **≤3 years**  **(n=52/103 cases/controls)** | | **>3 years**  **(n=29/57 cases/controls)** | | **p-int.c** |
| **Taxon (Class; Order; Family; Genus; Species)** | **OR (95% CI)b** | **pb** | **OR (95% CI)b** | **pb** |  |
| **Actinobacteria** |  |  |  |  |  |
| Actinobacteria; Actinobacteria; Actinomycetales (order) | 1.46 (1.01, 2.12) | 0.04 | 1.19 (0.72, 1.96) | 0.49 | 0.36 |
| Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces; cardiffensis (species) | 1.25 (0.83, 1.9) | 0.28 | 1.89 (1.04, 3.44) | 0.04 | 0.41 |
| Actinobacteria; Actinobacteria; Corynebacteriales; Corynebacteriaceae; Corynebacterium; durum (species) | 0.83 (0.69, 0.99) | 0.04 | 0.92 (0.73, 1.16) | 0.47 | 0.63 |
| **Bacteroidetes** |  |  |  |  |  |
| Bacteroidetes; Bacteroidetes C-1; Bacteroidetes O-1; Bacteroidetes F-1; Bacteroidetes G-3 (genus) | 1.28 (0.77, 2.12) | 0.34 | 2.32 (1.14, 4.73) | 0.02 | 0.13 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Porphyromonas; gingivalisd (species) | 1.05 (0.89, 1.23) | 0.59 | 1.04 (0.84, 1.3) | 0.71 | 0.83 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Tannerella; forsythiad (species) | 1.16 (0.92, 1.47) | 0.21 | 1.29 (0.92, 1.82) | 0.14 | 0.48 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Alloprevotella (genus) | 0.87 (0.74, 1.02) | 0.09 | 0.89 (0.72, 1.1) | 0.29 | 0.88 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella; nanceiensis (species) | 0.82 (0.71, 0.94) | 0.01 | 0.89 (0.74, 1.06) | 0.18 | 0.41 |
| **Firmicutes** |  |  |  |  |  |
| Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus; pneumoniae (species) | 0.86 (0.64, 1.16) | 0.33 | 0.61 (0.4, 0.95) | 0.03 | 0.19 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Lachnoanaerobaculum; umeaense (species) | 0.78 (0.63, 0.97) | 0.03 | 0.92 (0.73, 1.15) | 0.46 | 0.28 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium (genus) | 0.8 (0.65, 0.99) | 0.04 | 0.72 (0.52, 1.01) | 0.06 | 0.69 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium; parvum (species) | 0.85 (0.69, 1.05) | 0.12 | 0.82 (0.61, 1.09) | 0.17 | 0.87 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium (genus) | 0.77 (0.61, 0.96) | 0.02 | 0.97 (0.7, 1.32) | 0.83 | 0.3 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium; moorei (species) | 0.78 (0.62, 0.96) | 0.02 | 0.98 (0.72, 1.32) | 0.88 | 0.27 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Selenomonas; oral taxon 134 (species) | 1.59 (1.06, 2.37) | 0.02 | 1.16 (0.75, 1.8) | 0.50 | 0.43 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella; oral taxon 917 (species) | 1.16 (1.01, 1.33) | 0.03 | 1.11 (0.92, 1.34) | 0.27 | 0.76 |
| **Proteobacteria** |  |  |  |  |  |
| Proteobacteria; Betaproteobacteria (class) | 0.89 (0.78, 1.03) | 0.17 | 0.85 (0.68, 1.07) | 0.17 | 0.7 |
| Proteobacteria; Betaproteobacteria; Neisseriales (order) | 0.9 (0.79, 1.03) | 0.13 | 0.86 (0.69, 1.07) | 0.18 | 0.67 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae (family) | 0.9 (0.79, 1.03) | 0.12 | 0.87 (0.7, 1.08) | 0.21 | 0.75 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria (genus) | 0.89 (0.79, 1) | 0.06 | 0.88 (0.73, 1.07) | 0.21 | 0.9 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; flavescens (species) | 0.91 (0.81, 1.01) | 0.07 | 0.9 (0.75, 1.07) | 0.23 | 0.9 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; sicca (species) | 0.9 (0.79, 1.03) | 0.13 | 0.91 (0.77, 1.08) | 0.29 | 0.96 |
| Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus; oral taxon 908 (species) | 0.88 (0.78, 0.99) | 0.04 | 0.95 (0.8, 1.13) | 0.57 | 0.45 |
| **Spirochaetes** |  |  |  |  |  |
| Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Treponema; denticolad (species) | 0.89 (0.7, 1.12) | 0.32 | 1.14 (0.85, 1.52) | 0.38 | 0.17 |

aTaxon raw counts were normalized with the clr transformation and used as predictors in conditional logistic regression models; all taxa with p<0.05 in the unstratified analysis for EAC (main text Table 3) are included in this table.

bOdds ratios from conditional logistic regression using matched sets as strata and adjusting for smoking status, BMI category, and alcohol drinking level. Model parameters and p-values were pooled over 10 models from 10 imputed data sets (missing values in BMI category and alcohol drinking level were imputed) using 'mice' package, R.

c"p-int" is the p-value for the interaction term (time category [≤3 years or >3 years] \* taxon).

dPre-selected "red complex" periodontal pathogens (*P. gingivalis*, *T. forsythia*, *T. denticola*) are additionally included in the table.

**Supplementary Table 4.** Oral taxaa associated with EAC, stratified by cohort.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **CPS-II**  **(n=30/59 cases/controls)** | | **PLCO**  **(n=51/101 cases/controls)** | | **p-int.c** |
| **Taxon (Class; Order; Family; Genus; Species)** | **OR (95% CI)b** | **pb** | **OR (95% CI)b** | **pb** |  |
| **Actinobacteria** |  |  |  |  |  |
| Actinobacteria; Actinobacteria; Actinomycetales (order) | 1.49 (0.86, 2.57) | 0.16 | 1.28 (0.9, 1.83) | 0.17 | 0.67 |
| Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces; cardiffensis (species) | 1.83 (1.08, 3.1) | 0.03 | 1.16 (0.73, 1.86) | 0.52 | 0.24 |
| Actinobacteria; Actinobacteria; Corynebacteriales; Corynebacteriaceae; Corynebacterium; durum (species) | 0.94 (0.76, 1.16) | 0.54 | 0.76 (0.62, 0.94) | 0.01 | 0.13 |
| **Bacteroidetes** |  |  |  |  |  |
| Bacteroidetes; Bacteroidetes C-1; Bacteroidetes O-1; Bacteroidetes F-1; Bacteroidetes G-3 (genus) | 2.01 (0.86, 4.67) | 0.11 | 1.59 (1.03, 2.46) | 0.04 | 0.93 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Porphyromonas; gingivalisd (species) | 0.98 (0.78, 1.22) | 0.83 | 1.16 (0.98, 1.38) | 0.08 | 0.11 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Tannerella; forsythiad (species) | 1.09 (0.8, 1.49) | 0.57 | 1.38 (1.04, 1.84) | 0.03 | 0.35 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Alloprevotella (genus) | 0.93 (0.75, 1.15) | 0.50 | 0.83 (0.71, 0.97) | 0.02 | 0.16 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella; nanceiensis (species) | 0.91 (0.78, 1.05) | 0.20 | 0.82 (0.71, 0.95) | 0.01 | 0.47 |
| **Firmicutes** |  |  |  |  |  |
| Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus; pneumoniae (species) | 0.74 (0.5, 1.09) | 0.13 | 0.66 (0.46, 0.95) | 0.03 | 0.80 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Lachnoanaerobaculum; umeaense (species) | 0.82 (0.66, 1.02) | 0.07 | 0.88 (0.71, 1.08) | 0.23 | 0.56 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium (genus) | 0.74 (0.53, 1.03) | 0.07 | 0.84 (0.67, 1.05) | 0.13 | 0.69 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium; parvum (species) | 0.9 (0.72, 1.11) | 0.32 | 0.78 (0.6, 1.01) | 0.06 | 0.37 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium (genus) | 0.86 (0.66, 1.11) | 0.25 | 0.85 (0.68, 1.06) | 0.14 | 0.74 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium; moorei (species) | 0.85 (0.66, 1.1) | 0.21 | 0.86 (0.69, 1.07) | 0.17 | 0.84 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Selenomonas; oral taxon 134 (species) | 1.12 (0.73, 1.71) | 0.61 | 2.09 (1.31, 3.34) | 0.002 | 0.02 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella; oral taxon 917 (species) | 1.13 (0.96, 1.33) | 0.16 | 1.15 (0.98, 1.34) | 0.08 | 0.92 |
| **Proteobacteria** |  |  |  |  |  |
| Proteobacteria; Betaproteobacteria (class) | 0.66 (0.46, 0.93) | 0.02 | 0.91 (0.8, 1.04) | 0.18 | 0.15 |
| Proteobacteria; Betaproteobacteria; Neisseriales (order) | 0.73 (0.55, 0.97) | 0.03 | 0.91 (0.79, 1.04) | 0.15 | 0.33 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae (family) | 0.72 (0.54, 0.97) | 0.03 | 0.91 (0.8, 1.04) | 0.16 | 0.31 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria (genus) | 0.72 (0.54, 0.96) | 0.02 | 0.91 (0.81, 1.02) | 0.12 | 0.28 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; flavescens (species) | 0.82 (0.66, 1.02) | 0.07 | 0.9 (0.81, 1) | 0.06 | 0.77 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; sicca (species) | 0.8 (0.64, 0.99) | 0.04 | 0.92 (0.82, 1.04) | 0.20 | 0.51 |
| Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus; oral taxon 908 (species) | 0.89 (0.77, 1.03) | 0.12 | 0.87 (0.75, 1) | 0.05 | 0.47 |
| **Spirochaetes** |  |  |  |  |  |
| Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Treponema; denticolad (species) | 0.7 (0.47, 1.03) | 0.07 | 1.22 (0.95, 1.57) | 0.12 | 0.04 |

aTaxon raw counts were normalized with the clr transformation and used as predictors in conditional logistic regression models; all taxa with p<0.05 in the unstratified analysis for EAC (main text Table 3) are included in this table.

bOdds ratios from conditional logistic regression using matched sets as strata and adjusting for smoking status, BMI category, and alcohol drinking level. Model parameters and p-values were pooled over 10 models from 10 imputed data sets (missing values in BMI category and alcohol drinking level were imputed) using 'mice' package, R.

c"p-int" is the p-value for the interaction term (cohort [CPS-II or PLCO] \* taxon).

dPre-selected "red complex" periodontal pathogens (*P. gingivalis*, *T. forsythia*, *T. denticola*) are additionally included in the table.

**Supplementary Table 5.** Oral taxaa associated with EAC, stratified by smoking status.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Never smokers**  **(n=21/59 cases/controls)** | | **Ever smokers**  **(n=60/101 cases/controls)** | | **p-int.c** |
| **Taxon (Class; Order; Family; Genus; Species)** | **OR (95% CI)b** | **pb** | **OR (95% CI)b** | **pb** |  |
| **Actinobacteria** |  |  |  |  |  |
| Actinobacteria; Actinobacteria; Actinomycetales (order) | 0.58 (0.26, 1.3) | 0.18 | 1.56 (1.11, 2.17) | 0.01 | 0.26 |
| Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces; cardiffensis (species) | 0.93 (0.4, 2.15) | 0.86 | 1.67 (1.09, 2.58) | 0.02 | 0.31 |
| Actinobacteria; Actinobacteria; Corynebacteriales; Corynebacteriaceae; Corynebacterium; durum (species) | 0.9 (0.68, 1.19) | 0.46 | 0.82 (0.7, 0.97) | 0.02 | 0.61 |
| **Bacteroidetes** |  |  |  |  |  |
| Bacteroidetes; Bacteroidetes C-1; Bacteroidetes O-1; Bacteroidetes F-1; Bacteroidetes G-3 (genus) | 1.76 (0.78, 3.95) | 0.17 | 1.84 (1.1, 3.08) | 0.02 | 0.89 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Porphyromonas; gingivalisd (species) | 1.13 (0.85, 1.5) | 0.41 | 1.06 (0.89, 1.28) | 0.51 | 0.79 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Tannerella; forsythiad (species) | 1.38 (0.88, 2.16) | 0.16 | 1.28 (1.01, 1.62) | 0.04 | 0.66 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Alloprevotella (genus) | 1.03 (0.74, 1.43) | 0.86 | 0.84 (0.72, 0.97) | 0.02 | 0.87 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella; nanceiensis (species) | 0.95 (0.75, 1.21) | 0.68 | 0.82 (0.73, 0.93) | 0.001 | 0.58 |
| **Firmicutes** |  |  |  |  |  |
| Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus; pneumoniae (species) | 0.93 (0.56, 1.54) | 0.76 | 0.67 (0.49, 0.91) | 0.01 | 0.19 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Lachnoanaerobaculum; umeaense (species) | 1.4 (0.93, 2.11) | 0.11 | 0.7 (0.57, 0.86) | 0.001 | 0.02 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium (genus) | 0.73 (0.46, 1.17) | 0.19 | 0.77 (0.63, 0.95) | 0.01 | 0.94 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium; parvum (species) | 0.91 (0.66, 1.25) | 0.55 | 0.84 (0.68, 1.05) | 0.12 | 0.77 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium (genus) | 0.91 (0.61, 1.36) | 0.65 | 0.79 (0.65, 0.96) | 0.02 | 0.82 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium; moorei (species) | 0.93 (0.63, 1.38) | 0.72 | 0.8 (0.66, 0.96) | 0.02 | 0.85 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Selenomonas; oral taxon 134 (species) | 1.87 (0.99, 3.51) | 0.05 | 1.4 (0.97, 2.02) | 0.07 | 0.76 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella; oral taxon 917 (species) | 1.22 (0.94, 1.58) | 0.13 | 1.11 (0.97, 1.27) | 0.14 | 0.26 |
| **Proteobacteria** |  |  |  |  |  |
| Proteobacteria; Betaproteobacteria (class) | 0.96 (0.71, 1.29) | 0.77 | 0.83 (0.73, 0.95) | 0.01 | 0.87 |
| Proteobacteria; Betaproteobacteria; Neisseriales (order) | 0.95 (0.72, 1.26) | 0.72 | 0.84 (0.74, 0.96) | 0.01 | 0.85 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae (family) | 0.95 (0.72, 1.25) | 0.70 | 0.84 (0.74, 0.96) | 0.01 | 0.86 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria (genus) | 0.97 (0.75, 1.25) | 0.81 | 0.84 (0.75, 0.94) | 0.003 | 0.68 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; flavescens (species) | 1 (0.81, 1.23) | 0.99 | 0.86 (0.77, 0.95) | 0.003 | 0.43 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; sicca (species) | 1.03 (0.79, 1.35) | 0.80 | 0.87 (0.77, 0.99) | 0.03 | 0.82 |
| Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus; oral taxon 908 (species) | 0.99 (0.81, 1.22) | 0.93 | 0.84 (0.75, 0.95) | 0.004 | 0.37 |
| **Spirochaetes** |  |  |  |  |  |
| Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Treponema; denticolad (species) | 1.02 (0.71, 1.47) | 0.92 | 1.05 (0.84, 1.3) | 0.68 | 0.22 |

aTaxon raw counts were normalized with the clr transformation and used as predictors in traditional logistic regression models; all taxa with p<0.05 in the unstratified analysis for EAC (main text Table 3) are included in this table.

bOdds ratios from traditional logistic regression adjusted for matching factors (age, sex, race, cohort, time to case diagnosis), BMI category, and alcohol drinking level. Model parameters and p-values were pooled over 10 models from 10 imputed data sets (missing values in BMI category and alcohol drinking level were imputed) using 'mice' package, R.

c"p-int" is the p-value for the interaction term (smoking status [never or ever] \* taxon).

dPre-selected "red complex" periodontal pathogens (*P. gingivalis*, *T. forsythia*, *T. denticola*) are additionally included in the table.

**Supplementary Table 6.** Oral taxaa associated with EAC, stratified by obesity.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Non-obese (BMI<30 kg/m2)**  **(n=59/124 cases/controls)** | | **Obese (BMI≥30 kg/m2)**  **(n=17/22 cases/controls)** | | **p-int.c** |
| **Taxon (Class; Order; Family; Genus; Species)** | **OR (95% CI)b** | **pb** | **OR (95% CI)b** | **pb** |  |
| **Actinobacteria** |  |  |  |  |  |
| Actinobacteria; Actinobacteria; Actinomycetales (order) | 1.28 (0.91, 1.8) | 0.15 | 2.32 (0.66, 8.2) | 0.18 | 0.52 |
| Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces; cardiffensis (species) | 1.89 (1.25, 2.87) | 0.003 | 0.69 (0.1, 4.87) | 0.69 | 0.02 |
| Actinobacteria; Actinobacteria; Corynebacteriales; Corynebacteriaceae; Corynebacterium; durum (species) | 0.82 (0.7, 0.96) | 0.01 | 1.07 (0.63, 1.8) | 0.80 | 0.18 |
| **Bacteroidetes** |  |  |  |  |  |
| Bacteroidetes; Bacteroidetes C-1; Bacteroidetes O-1; Bacteroidetes F-1; Bacteroidetes G-3 (genus) | 1.65 (1.08, 2.51) | 0.02 | 0.8 (0.23, 2.79) | 0.71 | 0.99 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Porphyromonas; gingivalisd (species) | 1.06 (0.91, 1.24) | 0.45 | 1.5 (0.76, 2.95) | 0.23 | 0.49 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Tannerella; forsythiad (species) | 1.18 (0.95, 1.47) | 0.12 | 0.73 (0.31, 1.71) | 0.45 | 0.71 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Alloprevotella (genus) | 0.84 (0.72, 0.99) | 0.04 | 0.79 (0.48, 1.3) | 0.34 | 0.30 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella; nanceiensis (species) | 0.82 (0.72, 0.93) | 0.002 | 0.88 (0.61, 1.27) | 0.48 | 0.6 |
| **Firmicutes** |  |  |  |  |  |
| Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus; pneumoniae (species) | 0.78 (0.6, 1.01) | 0.06 | 0.18 (0.01, 2.48) | 0.19 | 0.62 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Lachnoanaerobaculum; umeaense (species) | 0.83 (0.69, 0.99) | 0.04 | 0.81 (0.45, 1.44) | 0.44 | 0.51 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium (genus) | 0.84 (0.68, 1.04) | 0.11 | 0.32 (0.06, 1.8) | 0.18 | 0.11 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium; parvum (species) | 0.85 (0.7, 1.04) | 0.11 | 1.27 (0.63, 2.54) | 0.48 | 0.85 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium (genus) | 0.84 (0.69, 1.03) | 0.10 | 0.7 (0.38, 1.29) | 0.24 | 0.67 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium; moorei (species) | 0.85 (0.7, 1.03) | 0.09 | 0.7 (0.38, 1.28) | 0.23 | 0.65 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Selenomonas; oral taxon 134 (species) | 1.37 (0.97, 1.93) | 0.07 | 5.1 (0.55, 46.93) | 0.14 | 0.69 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella; oral taxon 917 (species) | 1.14 (1.01, 1.3) | 0.04 | 1.08 (0.57, 2.05) | 0.79 | 0.91 |
| **Proteobacteria** |  |  |  |  |  |
| Proteobacteria; Betaproteobacteria (class) | 0.86 (0.75, 0.99) | 0.03 | 0.61 (0.27, 1.38) | 0.22 | 0.98 |
| Proteobacteria; Betaproteobacteria; Neisseriales (order) | 0.87 (0.77, 0.99) | 0.04 | 0.6 (0.24, 1.47) | 0.24 | 0.98 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae (family) | 0.87 (0.77, 0.99) | 0.04 | 0.57 (0.21, 1.58) | 0.25 | 0.96 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria (genus) | 0.87 (0.78, 0.97) | 0.02 | 0.58 (0.21, 1.56) | 0.25 | 0.87 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; flavescens (species) | 0.88 (0.8, 0.97) | 0.01 | 0.65 (0.21, 1.99) | 0.39 | 0.82 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; sicca (species) | 0.92 (0.82, 1.04) | 0.19 | 0.94 (0.63, 1.39) | 0.73 | 0.70 |
| Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus; oral taxon 908 (species) | 0.92 (0.82, 1.03) | 0.13 | 0.84 (0.58, 1.2) | 0.32 | 0.57 |
| **Spirochaetes** |  |  |  |  |  |
| Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Treponema; denticolad (species) | 1 (0.82, 1.21) | 0.96 | 0.64 (0.3, 1.37) | 0.24 | 0.55 |

aTaxon raw counts were normalized with the clr transformation and used as predictors in traditional logistic regression models; all taxa with p<0.05 in the unstratified analysis for EAC (main text Table 3) are included in this table.

bOdds ratios from traditional logistic regression adjusted for matching factors (age, sex, race, cohort, time to case diagnosis), smoking status, and alcohol drinking level. Model parameters and p-values were pooled over 10 models from 10 imputed data sets (missing values in alcohol drinking level were imputed) using 'mice' package, R. Participants missing BMI were omitted from the analysis.

c"p-int" is the p-value for the interaction term (obesity [non-obese or obese] \* taxon).

dPre-selected "red complex" periodontal pathogens (*P. gingivalis*, *T. forsythia*, *T. denticola*) are additionally included in the table.

**Supplementary Table 7.** Oral taxaa associated with EAC, stratified by fruit and vegetable intake.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Low intake**  **(n=36/68 cases/controls)** | | **High intake**  **(n=35/73 cases/controls)** | | **p-int.c** |
| **Taxon (Class; Order; Family; Genus; Species)** | **OR (95% CI)b** | **pb** | **OR (95% CI)b** | **pb** |  |
| **Actinobacteria** |  |  |  |  |  |
| Actinobacteria; Actinobacteria; Actinomycetales (order) | 0.98 (0.6, 1.6) | 0.93 | 2.39 (1.36, 4.19) | 0.003 | 0.05 |
| Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Actinomyces; cardiffensis (species) | 1.83 (0.97, 3.46) | 0.06 | 1.67 (0.92, 3.02) | 0.09 | 0.27 |
| Actinobacteria; Actinobacteria; Corynebacteriales; Corynebacteriaceae; Corynebacterium; durum (species) | 0.8 (0.64, 1) | 0.05 | 0.83 (0.67, 1.04) | 0.10 | 0.49 |
| **Bacteroidetes** |  |  |  |  |  |
| Bacteroidetes; Bacteroidetes C-1; Bacteroidetes O-1; Bacteroidetes F-1; Bacteroidetes G-3 (genus) | 1.73 (0.92, 3.23) | 0.09 | 2.1 (1.15, 3.82) | 0.02 | 0.62 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Porphyromonas; gingivalisd (species) | 1.04 (0.85, 1.27) | 0.71 | 1.13 (0.87, 1.47) | 0.36 | 0.93 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Tannerella; forsythiad (species) | 1.39 (1.03, 1.86) | 0.03 | 1.03 (0.68, 1.56) | 0.88 | 0.25 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Alloprevotella (genus) | 0.88 (0.69, 1.12) | 0.30 | 0.77 (0.63, 0.95) | 0.01 | 0.41 |
| Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella; nanceiensis (species) | 0.81 (0.68, 0.97) | 0.02 | 0.81 (0.69, 0.95) | 0.01 | 0.91 |
| **Firmicutes** |  |  |  |  |  |
| Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus; pneumoniae (species) | 0.65 (0.42, 1) | 0.05 | 0.69 (0.46, 1.04) | 0.08 | 0.81 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Lachnoanaerobaculum; umeaense (species) | 0.8 (0.61, 1.04) | 0.09 | 0.76 (0.59, 0.98) | 0.03 | 0.72 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium (genus) | 0.79 (0.56, 1.11) | 0.17 | 0.67 (0.5, 0.9) | 0.01 | 0.38 |
| Firmicutes; Clostridia; Clostridiales; Lachnospiraceae XIV; Oribacterium; parvum (species) | 0.91 (0.69, 1.2) | 0.50 | 0.82 (0.62, 1.07) | 0.15 | 0.39 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium (genus) | 0.67 (0.48, 0.95) | 0.03 | 0.88 (0.68, 1.13) | 0.30 | 0.41 |
| Firmicutes; Erysipelotrichia; Erysipelotrichales; Erysipelotrichaceae; Solobacterium; moorei (species) | 0.69 (0.5, 0.97) | 0.03 | 0.88 (0.69, 1.12) | 0.28 | 0.48 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Selenomonas; oral taxon 134 (species) | 1.78 (1.06, 2.97) | 0.03 | 1.54 (0.9, 2.61) | 0.11 | 0.95 |
| Firmicutes; Negativicutes; Selenomonadales; Veillonellaceae; Veillonella; oral taxon 917 (species) | 1.2 (0.99, 1.46) | 0.06 | 1.09 (0.91, 1.3) | 0.36 | 0.90 |
| **Proteobacteria** |  |  |  |  |  |
| Proteobacteria; Betaproteobacteria (class) | 0.82 (0.67, 1) | 0.06 | 0.85 (0.7, 1.01) | 0.07 | 0.86 |
| Proteobacteria; Betaproteobacteria; Neisseriales (order) | 0.82 (0.67, 1) | 0.05 | 0.87 (0.73, 1.03) | 0.09 | 0.72 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae (family) | 0.82 (0.68, 1) | 0.05 | 0.86 (0.73, 1.02) | 0.09 | 0.73 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria (genus) | 0.84 (0.71, 1) | 0.05 | 0.84 (0.72, 0.99) | 0.04 | 0.96 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; flavescens (species) | 0.88 (0.75, 1.02) | 0.10 | 0.88 (0.77, 1.01) | 0.06 | 0.97 |
| Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria; sicca (species) | 0.87 (0.73, 1.03) | 0.10 | 0.83 (0.69, 1) | 0.05 | 0.74 |
| Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus; oral taxon 908 (species) | 0.92 (0.77, 1.1) | 0.37 | 0.85 (0.72, 0.99) | 0.03 | 0.86 |
| **Spirochaetes** |  |  |  |  |  |
| Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Treponema; denticolad (species) | 1.18 (0.88, 1.58) | 0.27 | 0.84 (0.6, 1.17) | 0.29 | 0.18 |

aTaxon raw counts were normalized with the clr transformation and used as predictors in traditional logistic regression models; all taxa with p<0.05 in the unstratified analysis for EAC (main text Table 3) are included in this table.

bOdds ratios from traditional logistic regression adjusted for matching factors (age, sex, race, cohort, time to case diagnosis), smoking status, BMI category, and alcohol drinking level. Model parameters and p-values were pooled over 10 models from 10 imputed data sets (missing values in BMI category and alcohol drinking level were imputed) using 'mice' package, R. Participants missing fruit/vegetable intake were omitted from the analysis.

c"p-int" is the p-value for the interaction term (fruit/vegetable intake [low or high] \* taxon).

dPre-selected "red complex" periodontal pathogens (*P. gingivalis*, *T. forsythia*, *T. denticola*) are additionally included in the table.