
Supplementary Information

The oral bacteriome and mycobiome across stages of oral carcinogenesis

Weiwei Heng[†], Wenmei Wang[†], Tingting Dai[†], Ping Jiang, Yong Lu, Ruowei Li, Miaomiao Zhang, Ruiqi Xie, Yifan Zhou, Maomao Zhao, Ning Duan, Zhengqin Ye*, Fuhua Yan*, Xiang Wang*.

Table of Contents

Supplementary Figure 1. Principal component analysis of the plaque bacteriome from HC, OPL, and OSCC individuals.

Supplementary Figure 2. Principal component analysis of the saliva bacteriome from HC, OPL, and OSCC individuals.

Supplementary Figure 3. Principal component analysis of the plaque mycobiome from HC, OPL, and OSCC individuals.

Supplementary Figure 4. Principal component analysis of the saliva mycobiome from HC, OPL, and OSCC individuals.

Supplementary Figure 5. Bubble plot of the plaque bacteriome and mycobiome from the HC, OPL, and OSCC individuals.

Supplementary Figure 6. Bubble plot of the saliva bacteriome and mycobiome from the HC, OPL, and OSCC individuals.

Supplementary Figure 7. Oral carcinogenesis-associated alterations in the abundance of the plaque bacteria examined using LEfSe.

Supplementary Figure 8. Oral carcinogenesis-associated alterations in the abundance of the saliva bacteria examined using LEfSe.

Supplementary Figure 9. Oral carcinogenesis-associated alterations in the abundance of the plaque fungi examined using LEfSe.

Supplementary Figure 10. Oral carcinogenesis-associated alterations in the abundance of the saliva fungi examined using LEfSe.

Supplementary Figure 11. Intra-kingdom correlation at the genera level in buccal mucosal samples.

Supplementary Figure 12. Intra-kingdom correlation at the genera level in plaque samples.

Supplementary Figure 13. Intra-kingdom correlation at the genera level in saliva samples.

Supplementary Figure 14. Inter-kingdom correlation at the genera level in buccal mucosal samples.

Supplementary Figure 15. Inter-kingdom correlation at the genera level in plaque samples.

Supplementary Figure 16. Inter-kingdom correlation at the genera level in saliva samples.

Supplementary Figure 17. Functional alterations in the plaque bacteriome.

Supplementary Figure 18. Functional alterations in the saliva bacteriome.

Supplementary Figure 19. Functional alterations in the plaque mycobiome.

Supplementary Figure 20. Functional alterations in the saliva mycobiome.

Supplementary Table 1. Baseline information of the HC, OPL and OSCC groups.

Supplementary Table 2. Differentially abundant bacterial phyla of buccal mucosal samples among three groups.

Supplementary Table 3. Differentially abundant bacterial phyla of plaque samples among three groups.

Supplementary Table 4. Differentially abundant bacterial phyla of saliva samples among three groups.

Supplementary Table 5. Differentially abundant fungal phyla of buccal mucosal samples among three groups.

Supplementary Table 6. Differentially abundant fungal phyla of plaque samples among three groups.

Supplementary Table 7. Differentially abundant fungal phyla of saliva samples among three groups.

Supplementary Table 8. Differentially abundant bacterial genera of buccal mucosal samples among three groups.

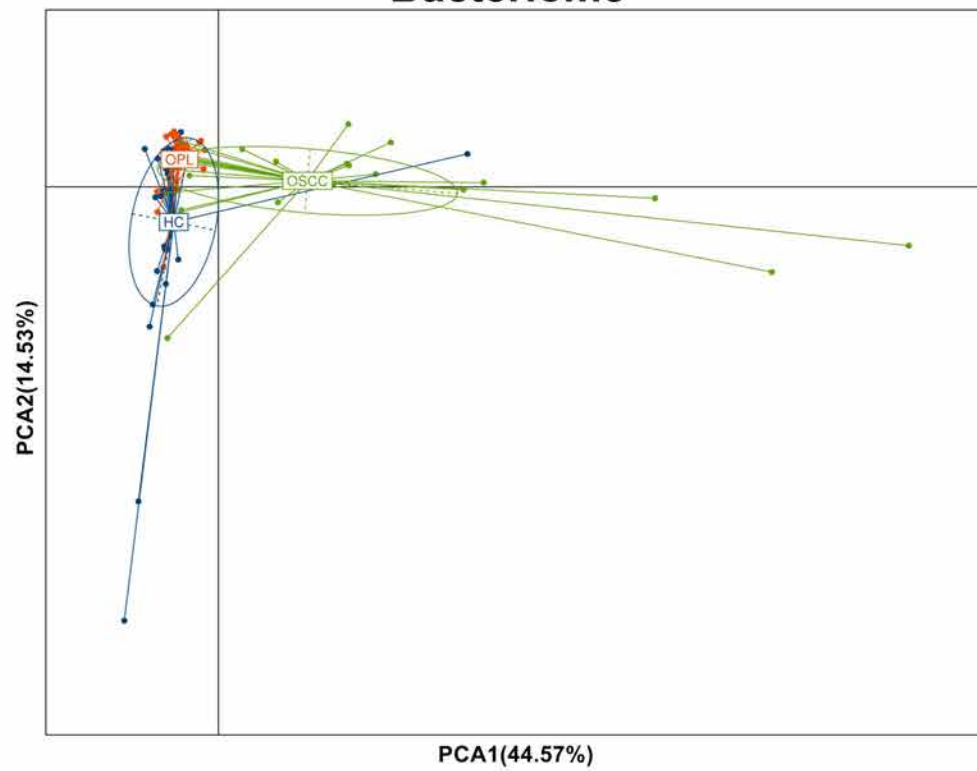
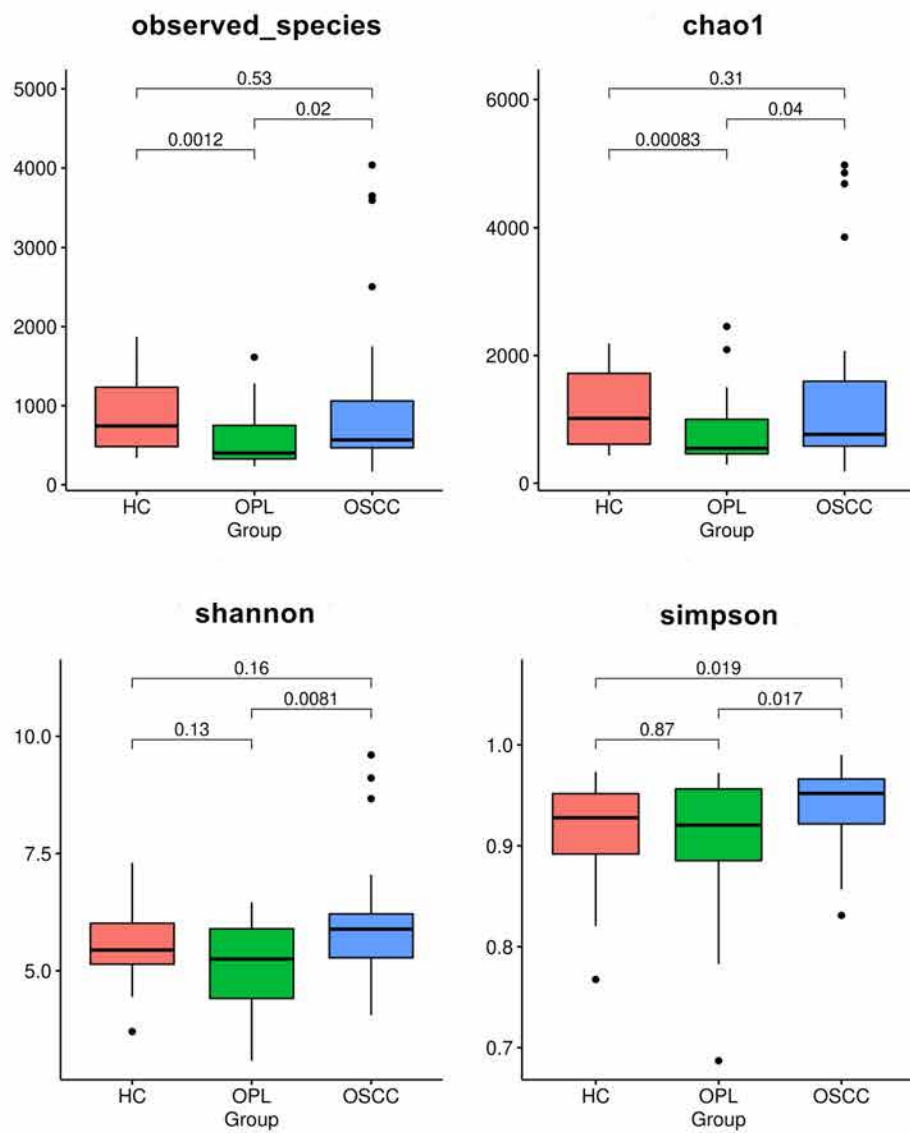
Supplementary Table 9. Differentially abundant bacterial genera of plaque samples among three groups.

Supplementary Table 10. Differentially abundant bacterial genera of saliva samples among three groups.

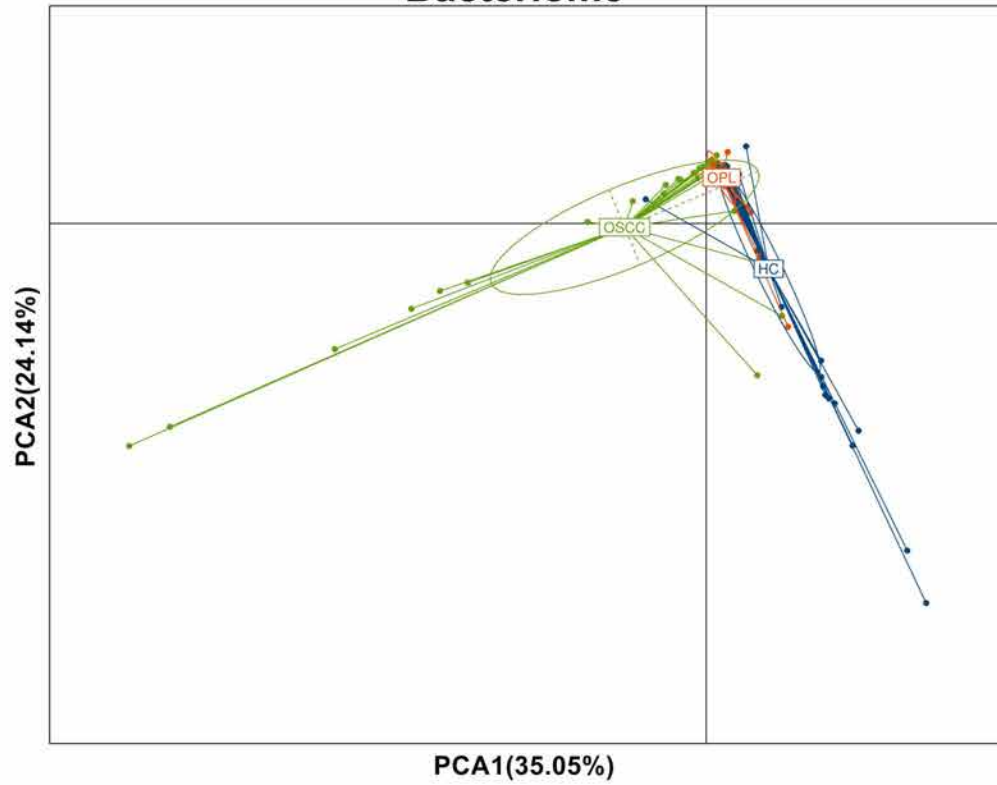
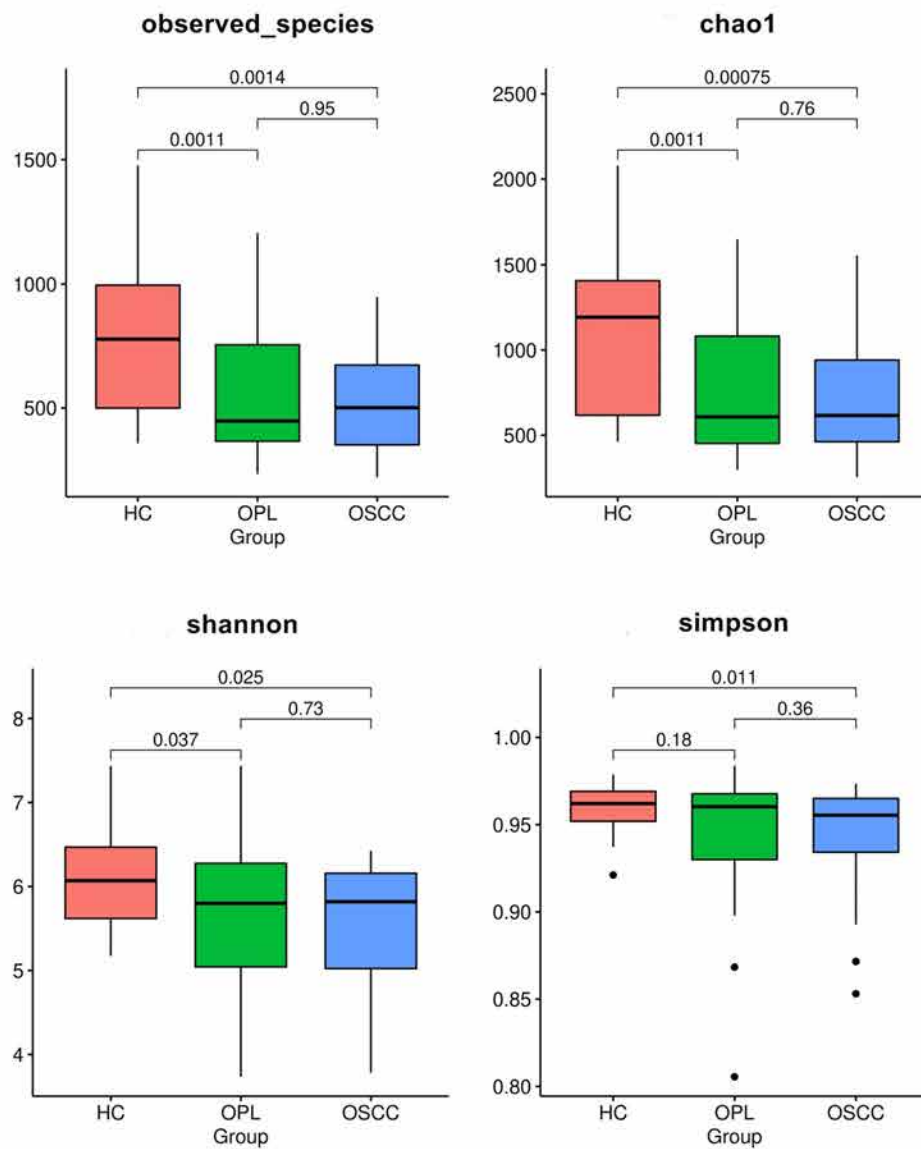
Supplementary Table 11. Differentially abundant fungal genera of buccal mucosal samples among three groups.

Supplementary Table 12. Differentially abundant fungal genera of plaque samples among three groups.

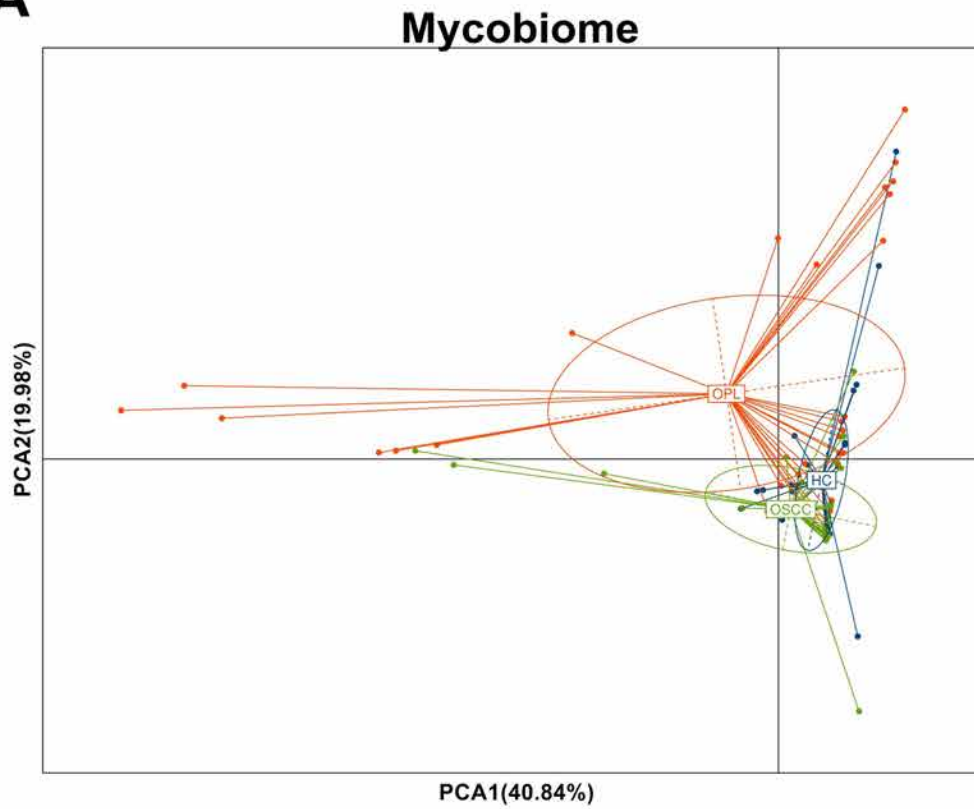
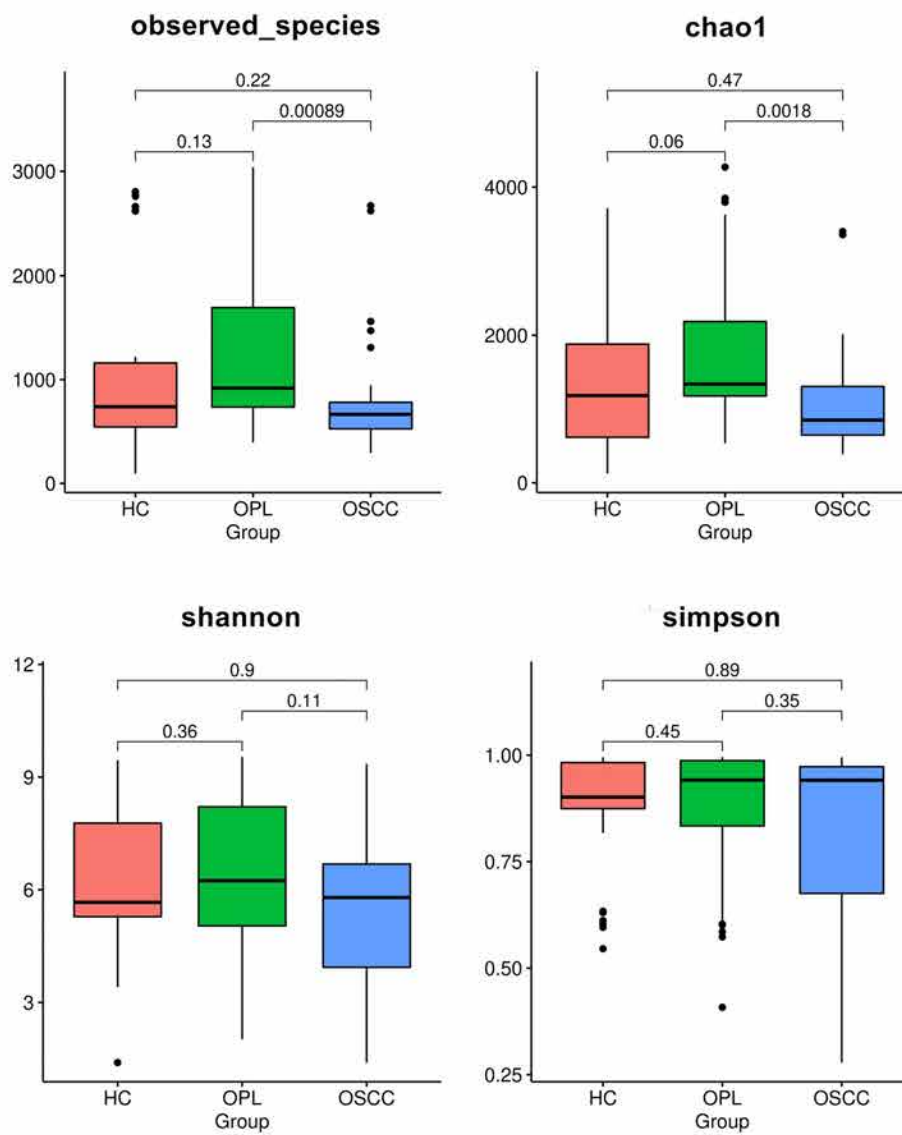
Supplementary Table 13. Differentially abundant fungal genera of saliva samples among three groups.

A**Bacteriome****B**

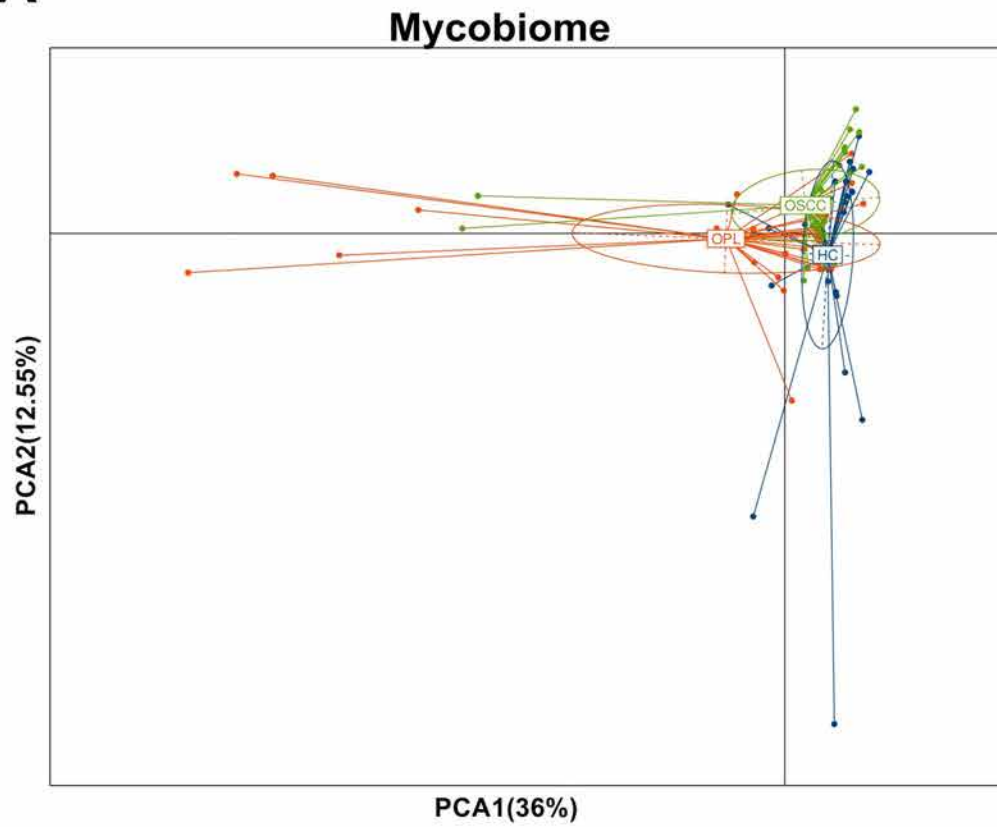
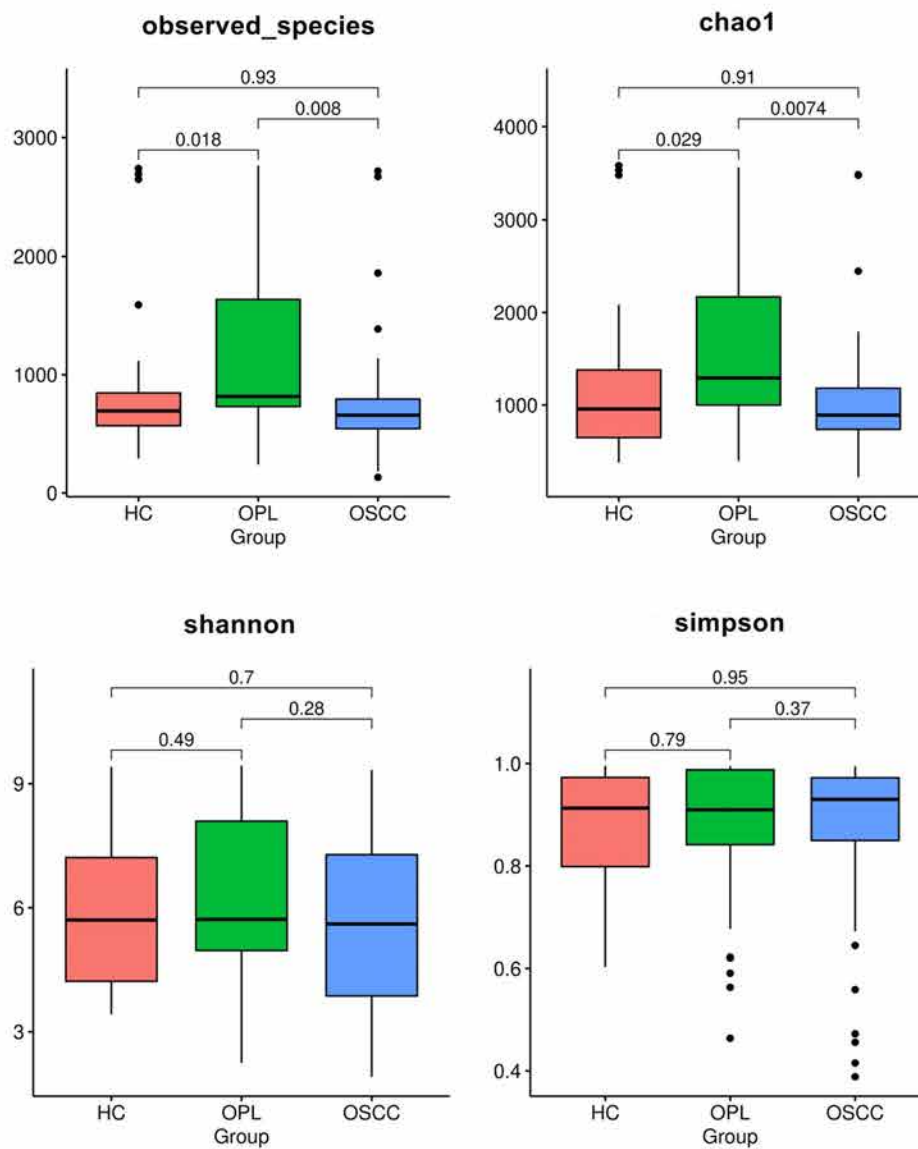
Supplementary Figure 1. Principal component analysis of the plaque bacteriome from HC, OPL, and OSCC individuals. (A) The plaque bacteriome of individuals with OSCC was statistically significantly different ($P < 0.05$) from HC and OPL individuals. (B) Box plots show the diversity and richness of the plaque bacteriome from the HP, OPL, and OSCC groups at the OTU level.

A**Bacteriome****B**

Supplementary Figure 2. Principal component analysis of the saliva bacteriome from HC, OPL, and OSCC individuals. (A) The plaque mycobiome of individuals with OSCC was statistically significantly different ($P < 0.05$) from HC and OPL individuals. (B) Box plots show the diversity and richness of the saliva bacteriome from the HP, OPL, and OSCC groups at the OTU level.

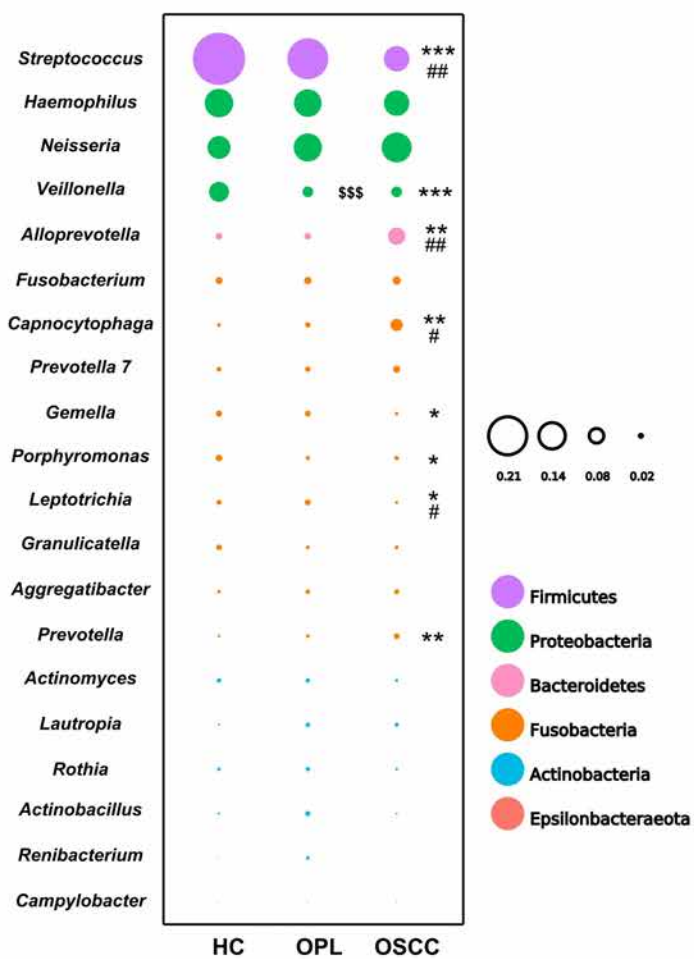
A**B**

Supplementary Figure 3. Principal component analysis of the plaque mycobiome from HC, OPL, and OSCC individuals. (A) The saliva bacteriome of individuals with OSCC was statistically significantly different ($P < 0.05$) from HC and OPL individuals. (B) Box plots show the diversity and richness of the plaque mycobiome from the HP, OPL, and OSCC groups at the OTU level.

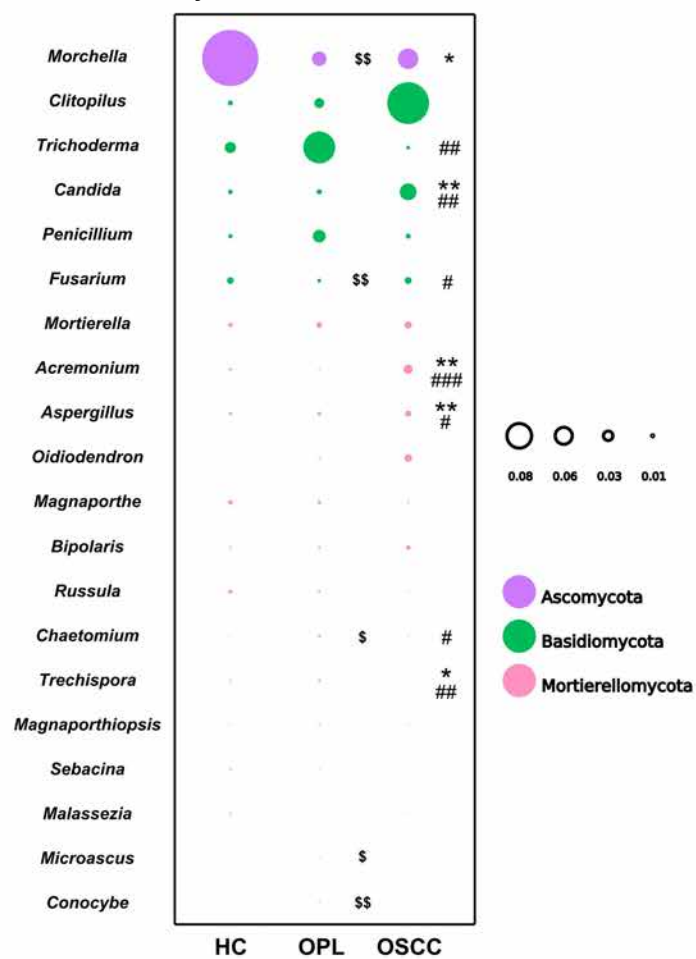
A**B**

Supplementary Figure 4. Principal component analysis of the saliva mycobiome from HC, OPL, and OSCC individuals. (A) The saliva mycobiome of individuals with OSCC was statistically significantly different ($P < 0.05$) from HC and OPL individuals. (B) Box plots show the diversity and richness of the saliva mycobiome from the HP, OPL, and OSCC groups at the OTU level.

A



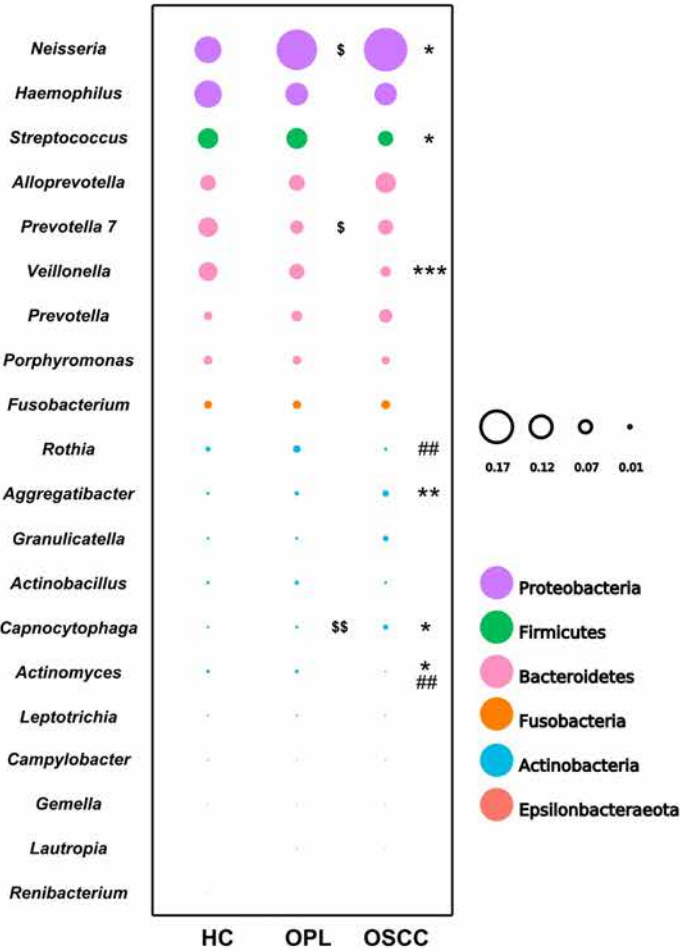
B



Supplementary Figure 5. Bubble plot of the plaque (A) bacteriome and (B) mycobiome from HC, OPL, and OSCC individuals. Each bubble represents one genus and the bubble size represents the relative abundance of each genus. For the OSCC group, *, ** and *** denote $P < 0.05$, $P < 0.01$ and $P < 0.001$ versus the HC group, respectively. For the OSCC group, #, ## and ### denote $P < 0.05$, $P < 0.01$ and $P < 0.001$ versus the OPL group, respectively. For the OPL group, \$, \$\$ and \$\$\$ denote $P < 0.05$, $P < 0.01$ and $P < 0.001$ versus the HC group, respectively.

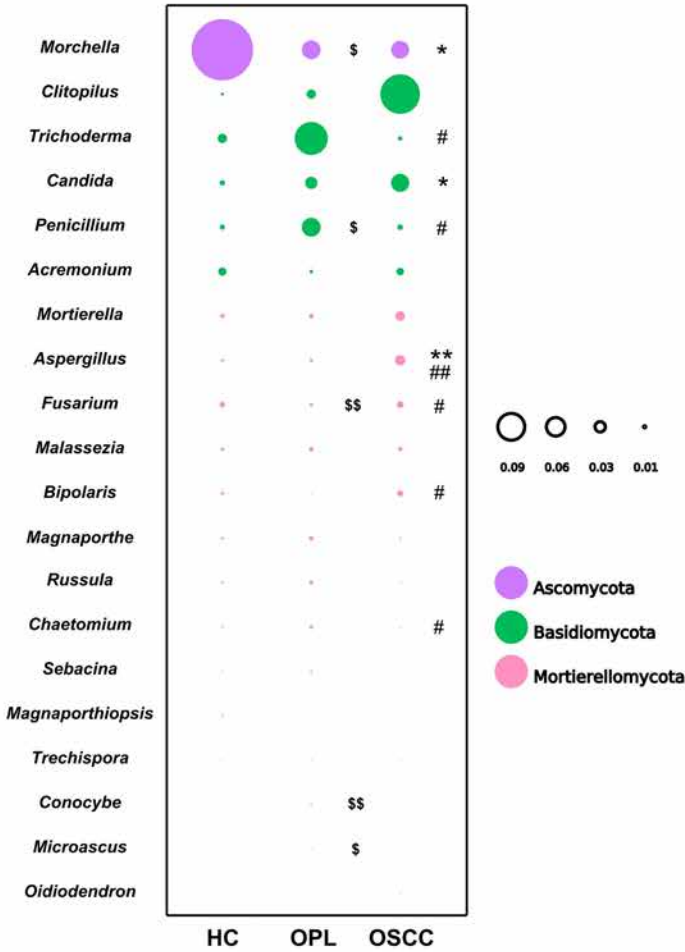
A

Bacteriome



B

Mycobiome

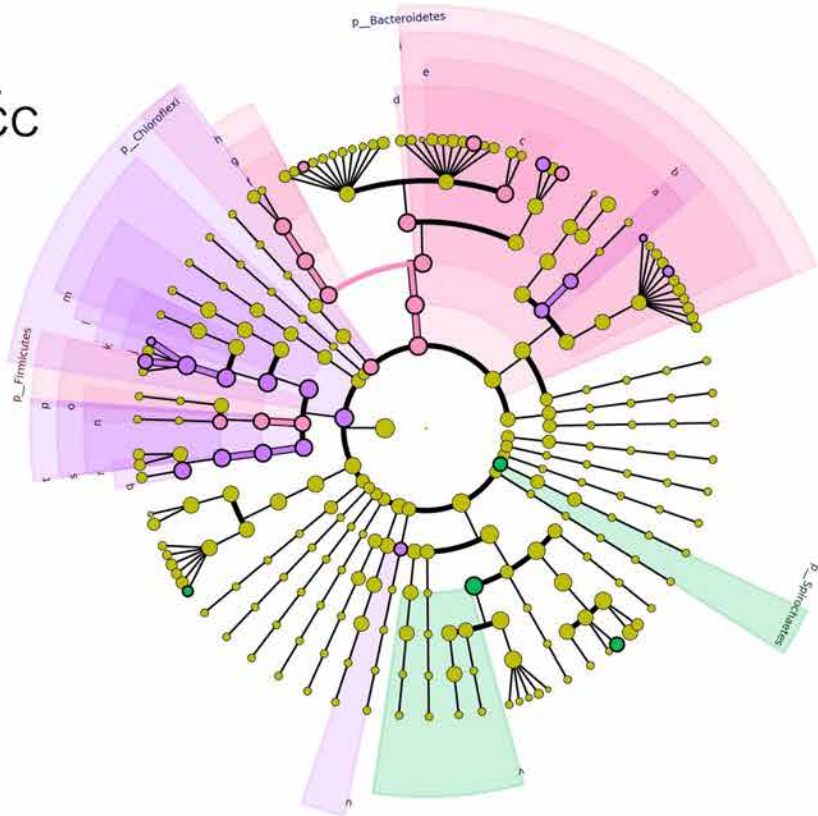


Supplementary Figure 6. Bubble plot of the saliva (A) bacteriome and (B) mycobiome from the HC, OPL, and OSCC individuals. Each bubble represents one genus and the bubble size represents the relative abundance of each genus. For the OSCC group, *, ** and *** denote $P < 0.05$, $P < 0.01$ and $P < 0.001$ versus the HC group, respectively. For the OSCC group, # and ## denote $P < 0.05$ and $P < 0.01$ versus the OPL group, respectively. For the OPL group, \$ and \$\$ denote $P < 0.05$ and $P < 0.01$ versus the HC group, respectively.

A

Cladogram

HC
OPL
OSCC

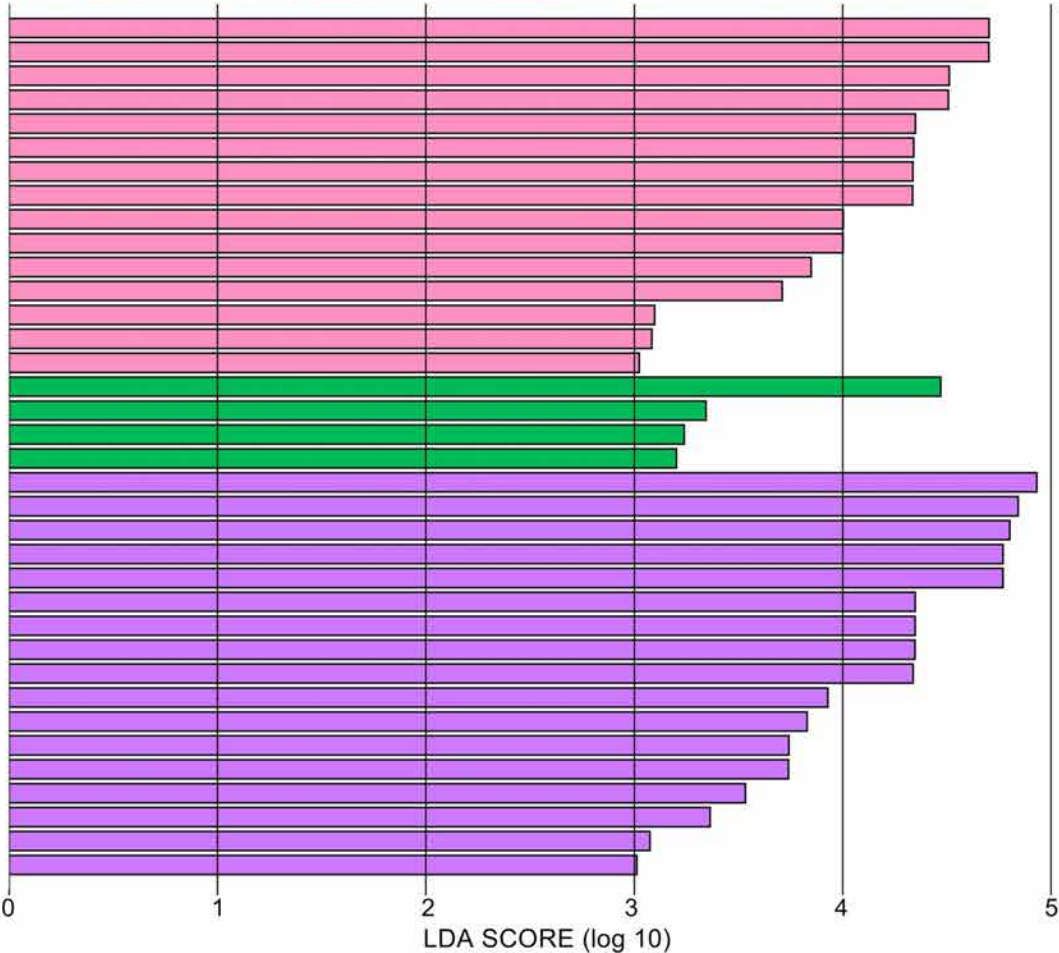


a: f_Corynebacteriaceae
b: o_Corynebacteriales
c: g_Alloprevotella
d: f_Prevotellaceae
e: o_Bacteroidales
f: g_Capnocytophaga
g: f_Flavobacteriaceae
h: o_Flavobacteriales
i: c_Bacteroidia
j: g_Streptococcus
k: f_Streptococcaceae
l: o_Lactobacillales
m: c_Bacilli
n: f_Ruminococcaceae
o: o_Clostridiales
p: c_Clostridia
q: g_Veillonella
r: f_Veillonellaceae
s: o_Selenomonadales
t: c_Negativicutes
u: c_Planctomycetacia
v: o_Betaproteobacteriales

B

HC OPL OSCC

c_Bacteroidia
p_Bacteroidetes
f_Prevotellaceae
o_Bacteroidales
g_Alloprevotella
o_Flavobacteriales
g_Capnocytophaga
f_Flavobacteriaceae
o_Clostridiales
c_Clostridia
p_Chloroflexi
s_Prevotellaintermedia
f_Ruminococcaceae
s_Prevotellasp_oralcloneFW035
s_Porphyrimonasendodontalis
o_Betaproteobacteriales
s_AggregatibacteraphrophilusATCC33389
s_Leptotrichiasp_oraltaxon847
p_Spirochaetes
p_Firmicutes
c_Bacilli
o_Lactobacillales
g_Streptococcus
f_Streptococcaceae
o_Selenomonadales
c_Negativicutes
f_Veillonellaceae
g_Veillonella
s_Actinomycesweissii
s_Scapharcabroughtonii
f_Corynebacteriaceae
o_Corynebacteriales
s_Streptococcusalivariussubsp_salivarius
s_Porphyrimonassp_oralcloneHF001
s_Actinomycessp_oralcloneGU009
c_Planctomycetacia

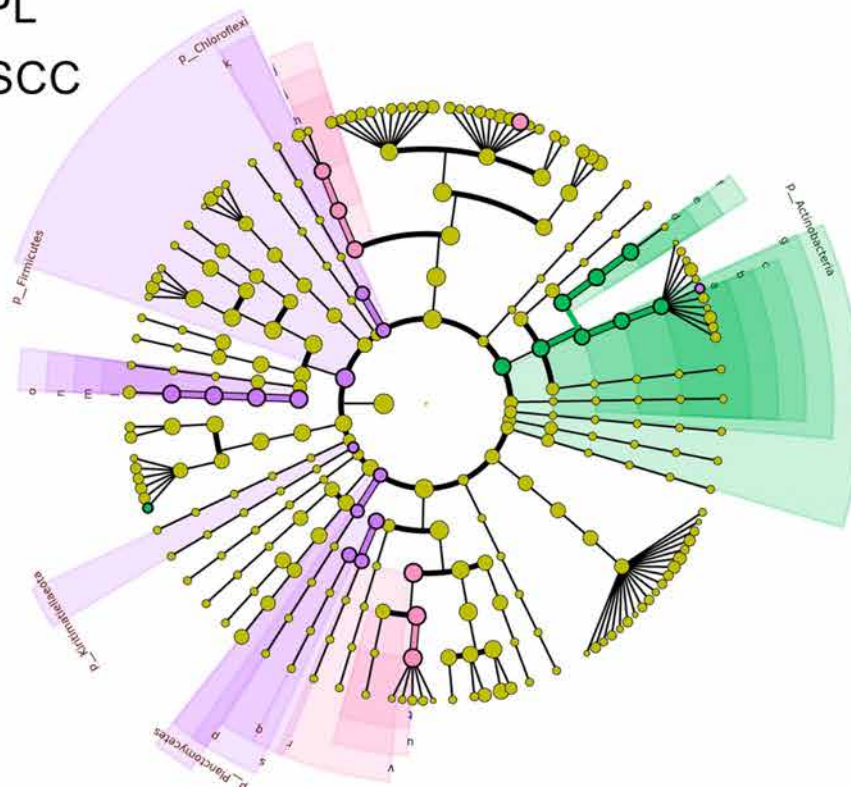


Supplementary Figure 7. Oral carcinogenesis-associated alterations in the abundance of the plaque bacteria examined using LEfSe. (A) Cladogram indicates the phylogenetic distribution of active bacteria that were remarkably enriched. (B) Bar plots at the species level with significant differences in abundance based on LEfSe. The color bars show the LDA scores of species that were enriched in the indicated condition: purple bar (HC), green bar (OPL), and pink bar (OSCC).

A

Cladogram

HC
OPL
OSCC

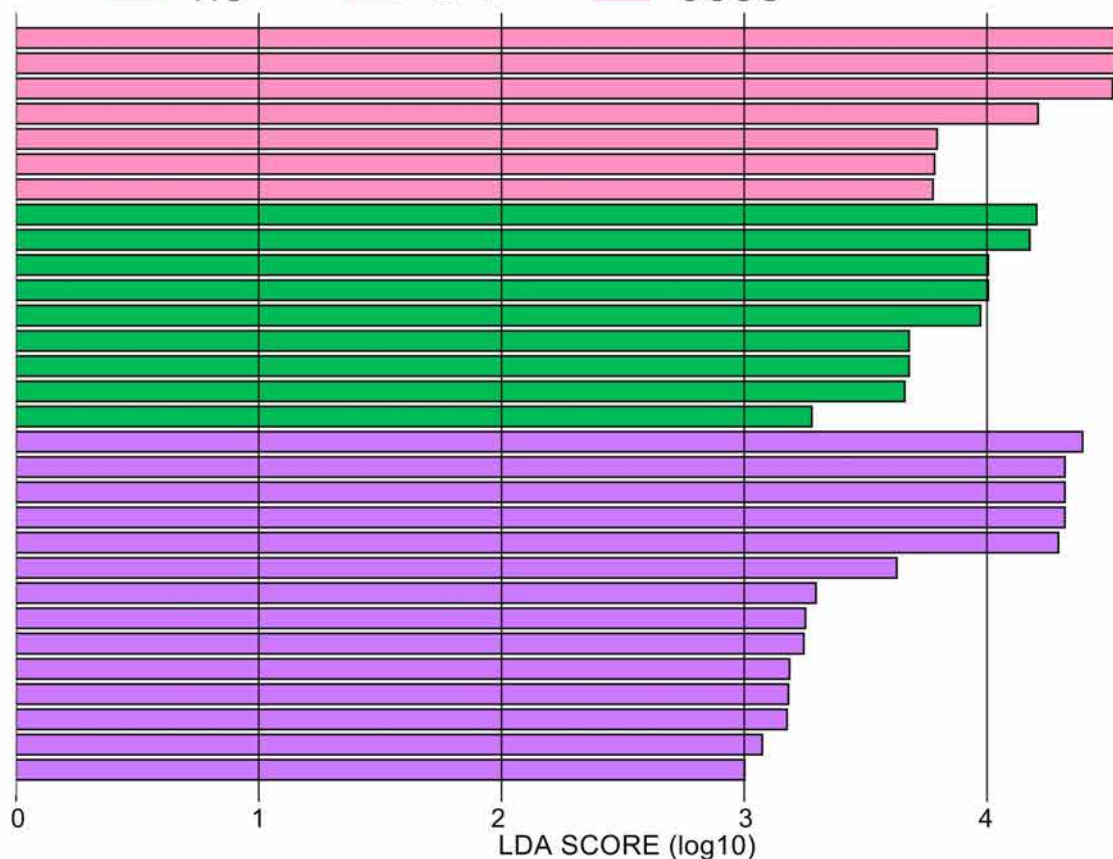


a: g_Actinomyces
b: f_Actinomycetaceae
c: o_Actinomycetales
d: g_Rothia
e: f_Micrococcaceae
f: o_Micrococcales
g: c_Actinobacteria
h: g_Capnocytophaga
i: f_Flavobacteriaceae
j: o_Flavobacteriales
k: c_Annaerolineae
l: g_Veillonella
m: f_Veillonellaceae
n: o_Selenomonadales
o: c_Negativicutes
p: c_Plantomycetacia
q: o_Rhizobiales
r: o_Sphingomonadales
s: c_Alphaproteobacteria
t: g_Neisseria
u: f_Neisseriaceae
v: o_Betaproteobacteriales

B

HC OPL OSCC

o_Betaproteobacteriales
f_Neisseriaceae
g_Neisseria
s_Prevotellaintermedia
g_Capnocytophaga
o_Flavobacteriales
f_Flavobacteriaceae
p_Actinobacteria
c_Actinobacteria
o_Micrococcales
f_Micrococcaceae
g_Rothia
o_Actinomycetales
f_Actinomycetaceae
g_Actinomyces
s_Leptotrichiasp_oraltaxon847
p_Firmicutes
f_Veillonellaceae
o_Selenomonadales
c_Negativicutes
g_Veillonella
c_Alphaproteobacteria
o_Rhizobiales
p_Chloroflexi
o_Sphingomonadales
c_Annaerolineae
p_Kiritimatiellaeota
s_Actinomycesoris
p_Plantomycetes
c_Plantomycetacia

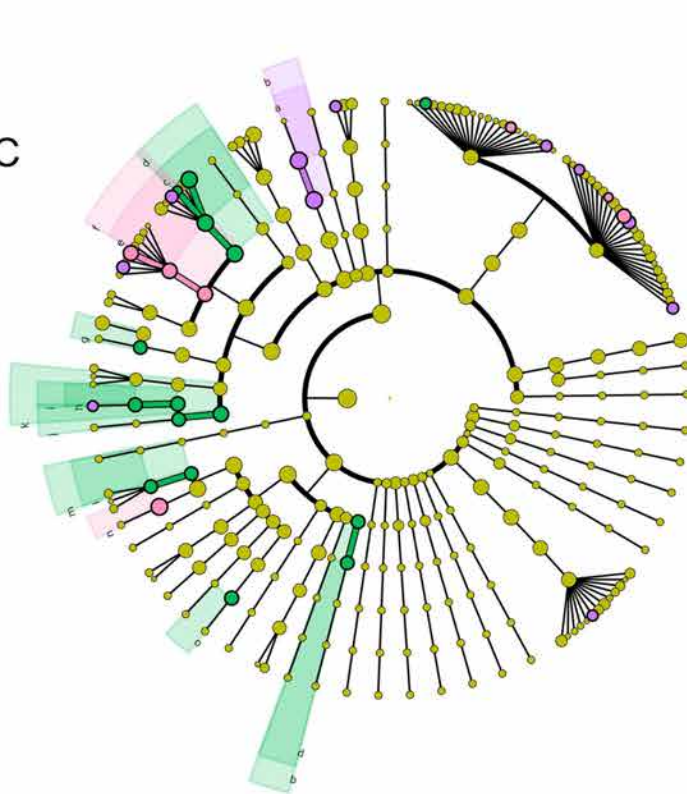


Supplementary Figure 8. Oral carcinogenesis-associated alterations in the abundance of the saliva bacteria examined using LEfSe. (A) Cladogram indicates the phylogenetic distribution of active bacteria that were remarkably enriched. (B) Bar plots at the species level with significant differences in abundance based on LEfSe. The color bars show the LDA scores of species that were enriched in the indicated condition: purple bar (HC), green bar (OPL), and pink bar (OSCC).

A

Cladogram

HC
OPL
OSCC

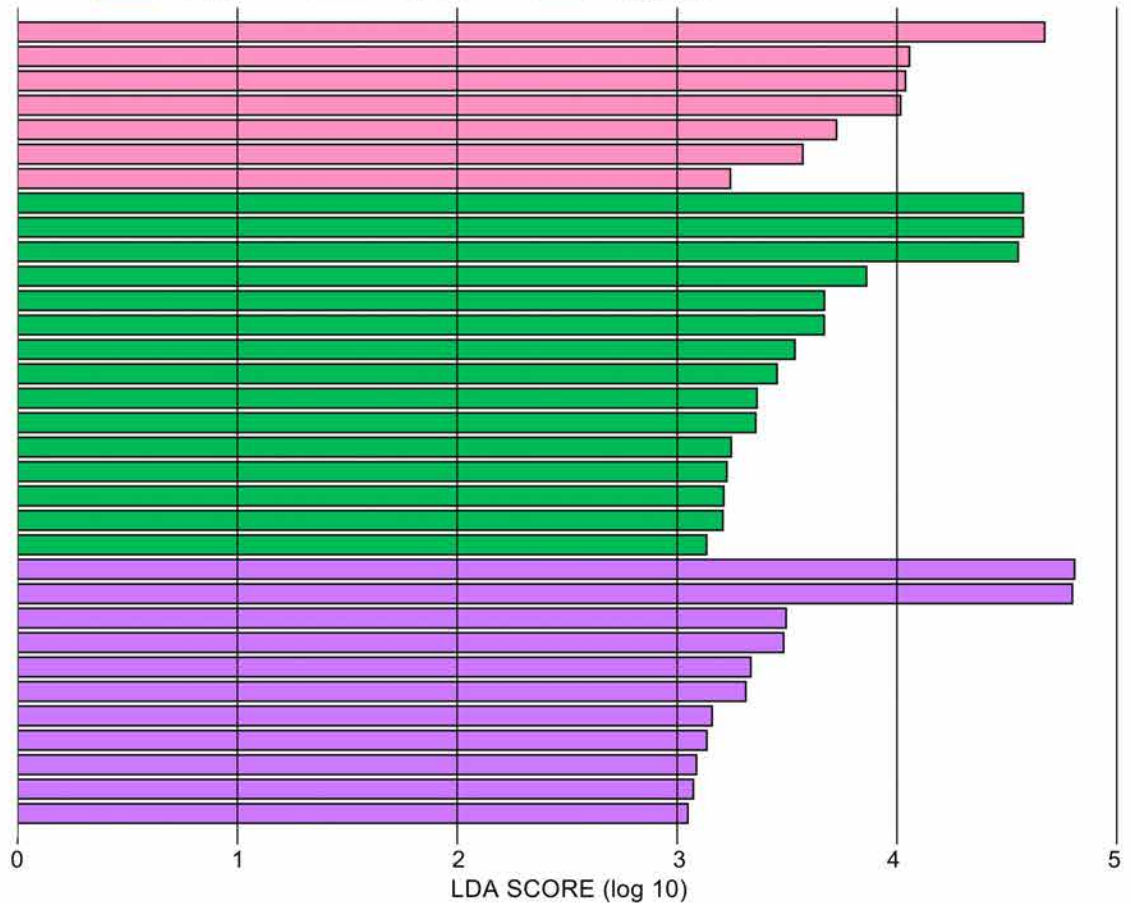


a: g_Morchella
b: f_Morchellaceae
c: g_Trichoderma
d: f_Hypocreaceae
e: g_Acremonium
f: f_Hypocreales_fam_Incertae_sedis
g: g_Magnaporthiopsis
h: g_Chaetomium
i: f_Chaetomiaceae
j: f_Lasiosphaeriaceae
k: o_Sordariales
l: g_Conocybe
m: f_Bolbitiaceae
n: g_Clitopilus
o: g_Trechispora
p: o_Tremellales
q: c_Tremellomycetes

B

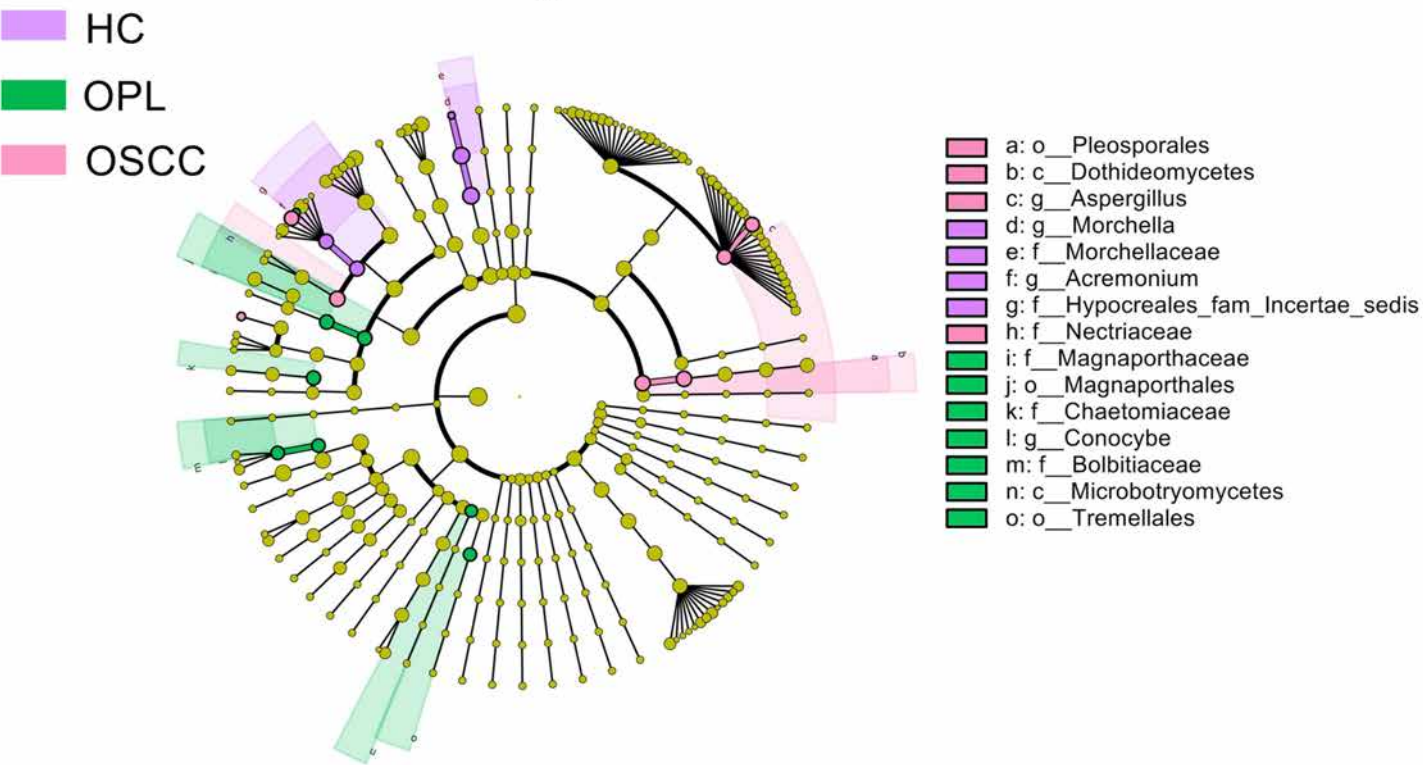
HC OPL OSCC

g_Clitopilus
g_Acremonium
s_Acremonium_exuvium
f_Hypocreales_fam_Incertae_sedis
s_AspERGILLUS_ochraceopetaliformis
s_AspERGILLUS_fumigatus
s_Penicillium_cryptum
f_Hypocreaceae
g_Trichoderma
s_Trichoderma_asperellum
o_Sordariales
g_Conocybe
f_Bolbitiaceae
f_Chaetomiaceae
s_Trichoderma_koningiopsis
g_Chaetomium
g_Trechispora
s_Penicillium_simplicissimum
c_Tremellomycetes
o_Tremellales
g_Magnaporthiopsis
f_Lasiosphaeriaceae
f_Morchellaceae
g_Morchella
s_Penicillium_bialowiezense
s_AspERGILLUS_aculeatus
s_Chaetomium_atrobrunneum
s_Acremonium_persicinum
s_Oidiodendron_rhodogenum
s_AspERGILLUS_tamarai
s_Mortierella_echinula
s_AspERGILLUS_flavus
s_Trichoderma_longibrachiatum



Supplementary Figure 9. Oral carcinogenesis-associated alterations in the abundance of the plaque fungi examined using LEfSe. (A) Cladogram indicates the phylogenetic distribution of active fungi that were remarkably enriched. (B) Bar plots at the species level with significant differences in abundance based on LEfSe. The color bars show the LDA scores of species that were enriched in the indicated condition: purple bar (HC), green bar (OPL), and pink bar (OSCC).

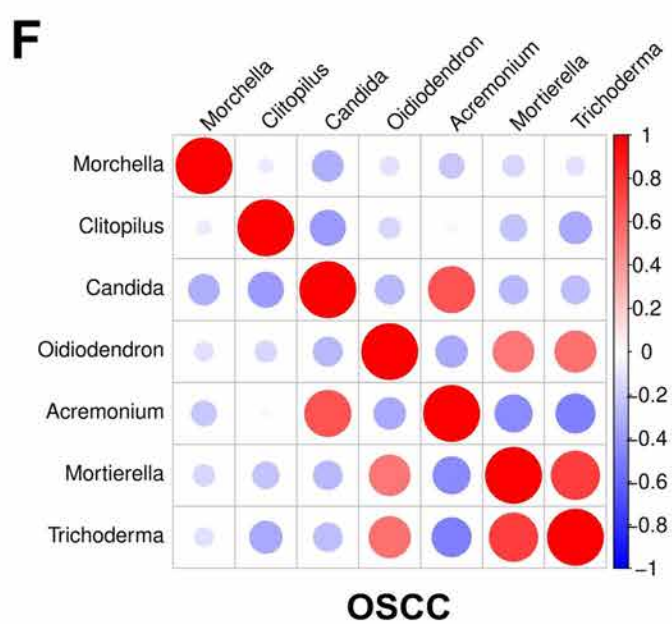
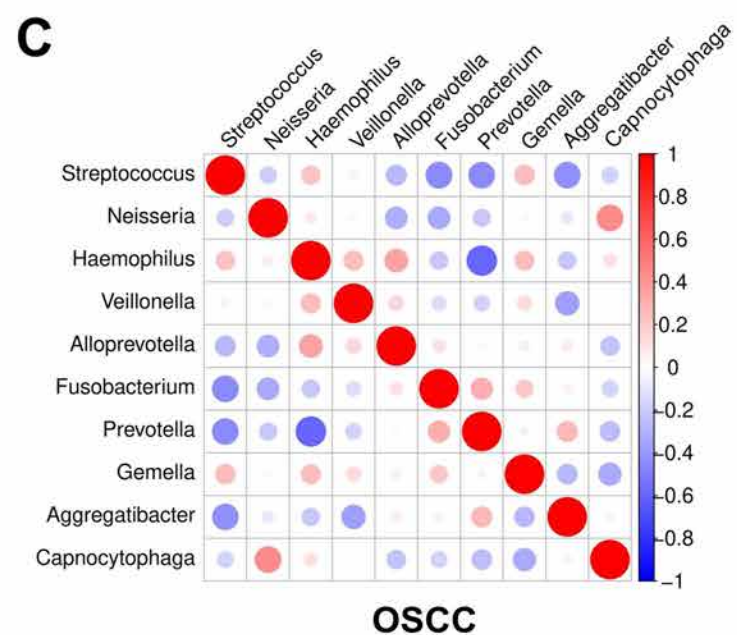
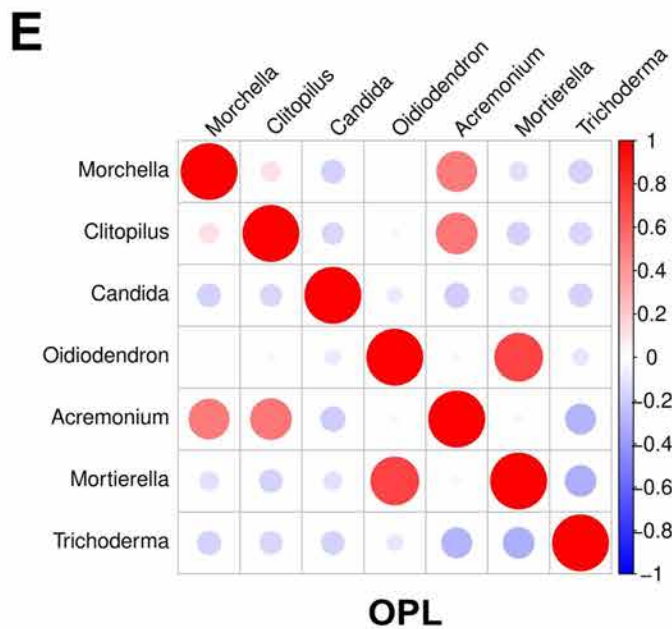
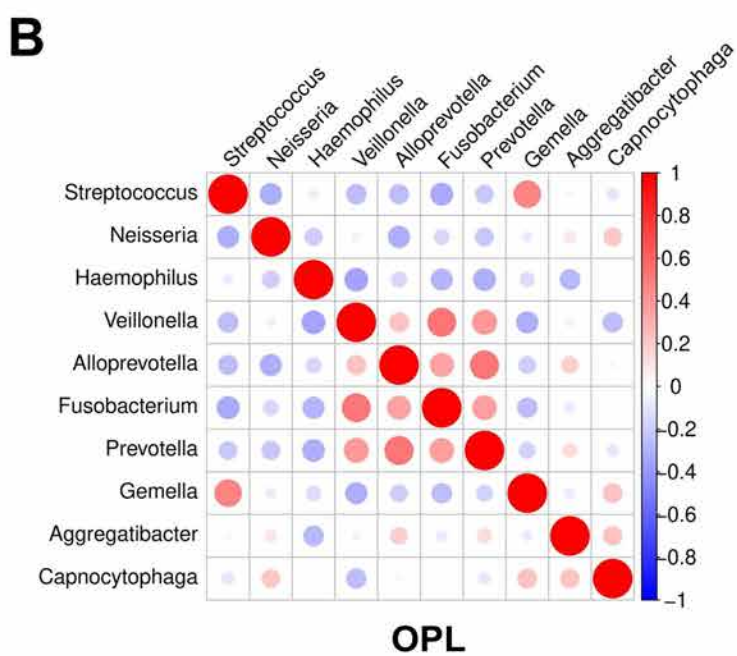
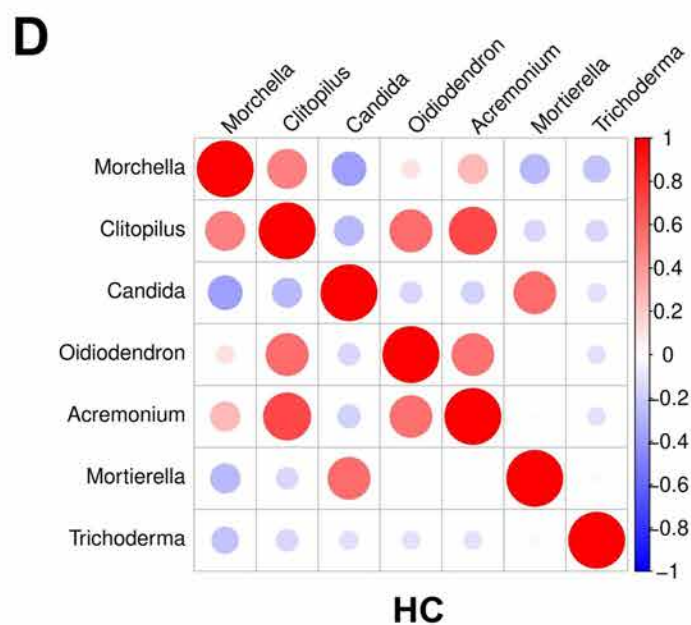
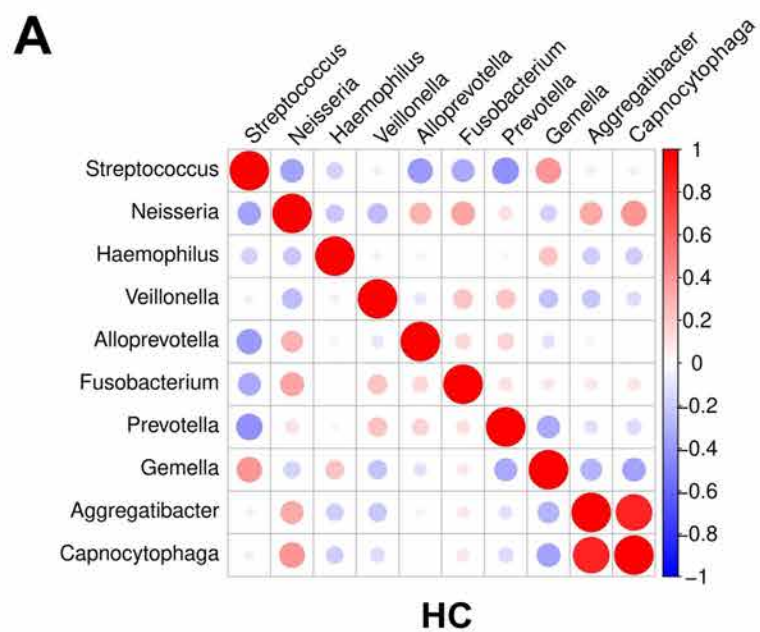
A Cladogram



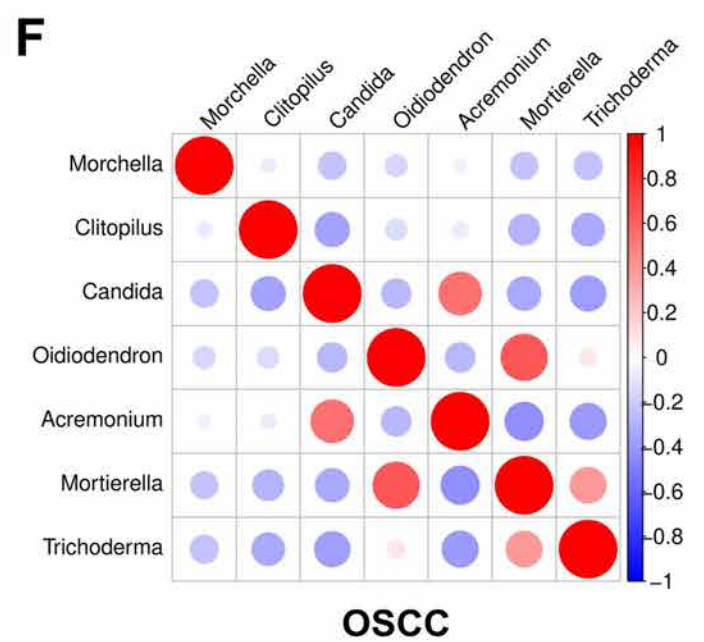
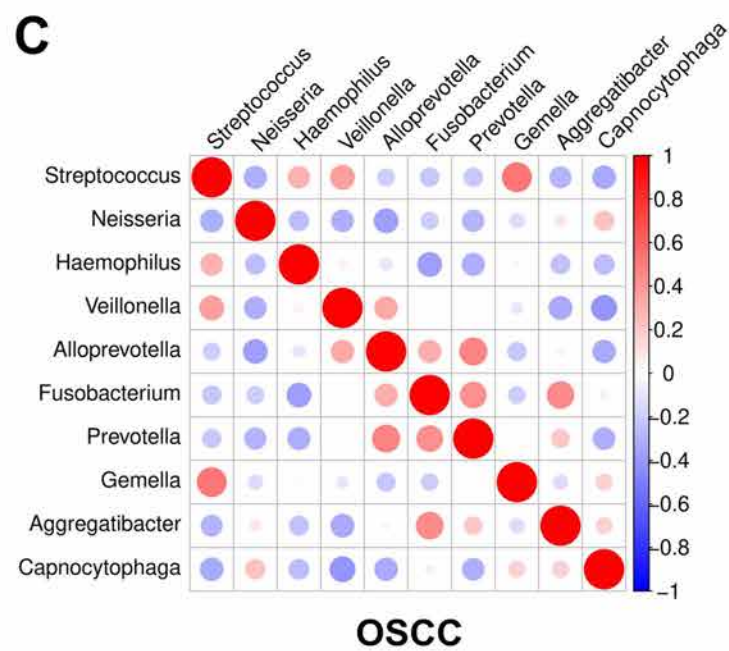
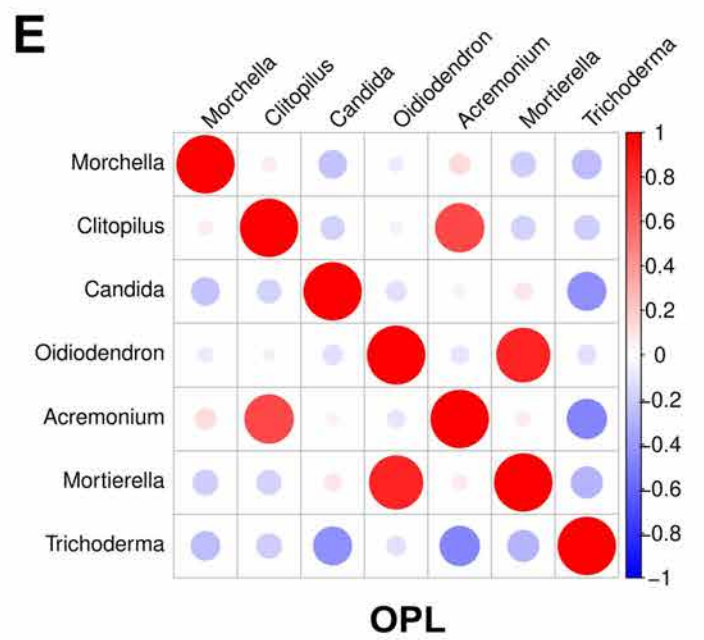
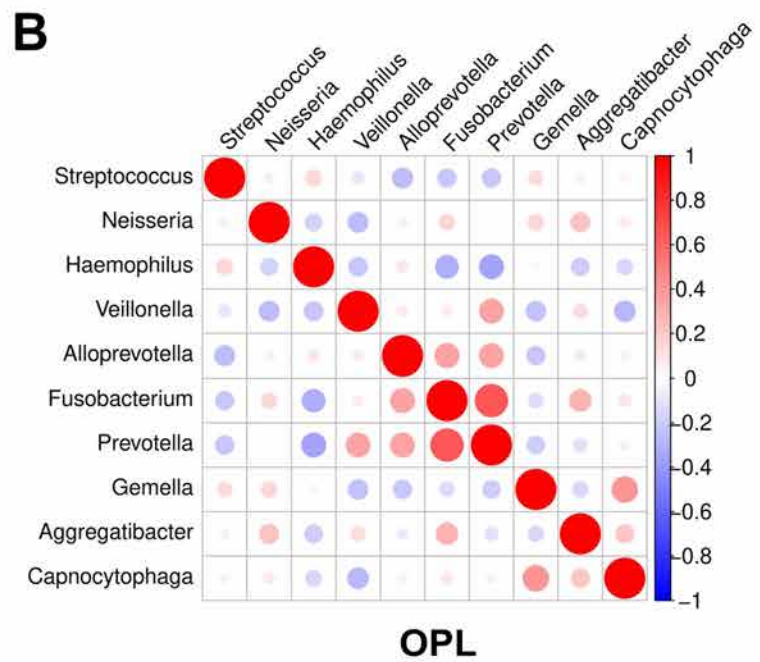
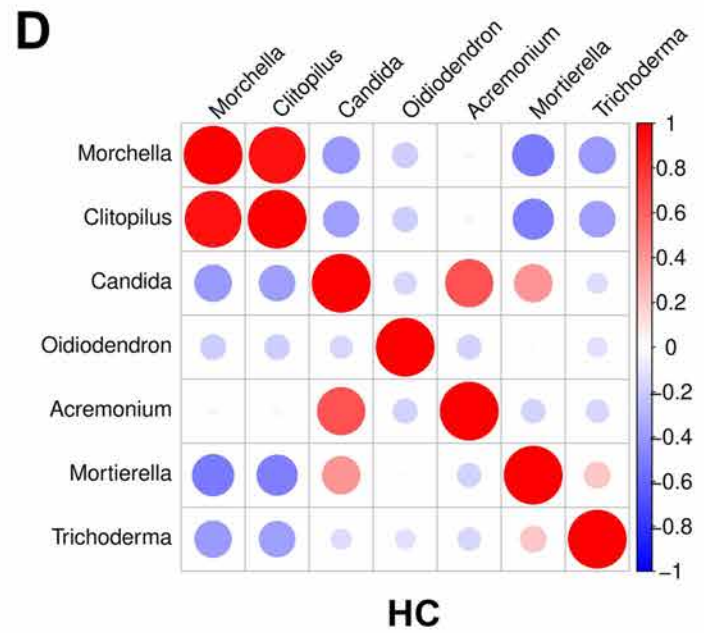
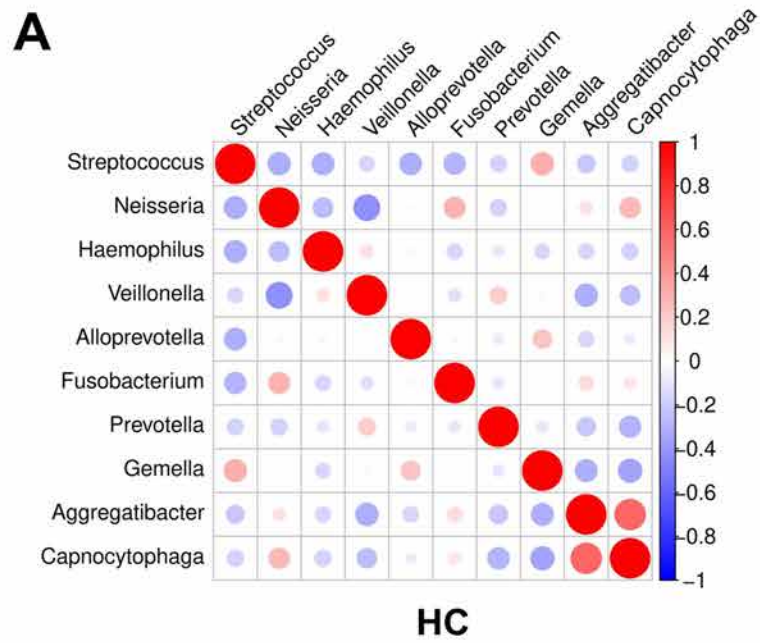
B HC OPL OSCC



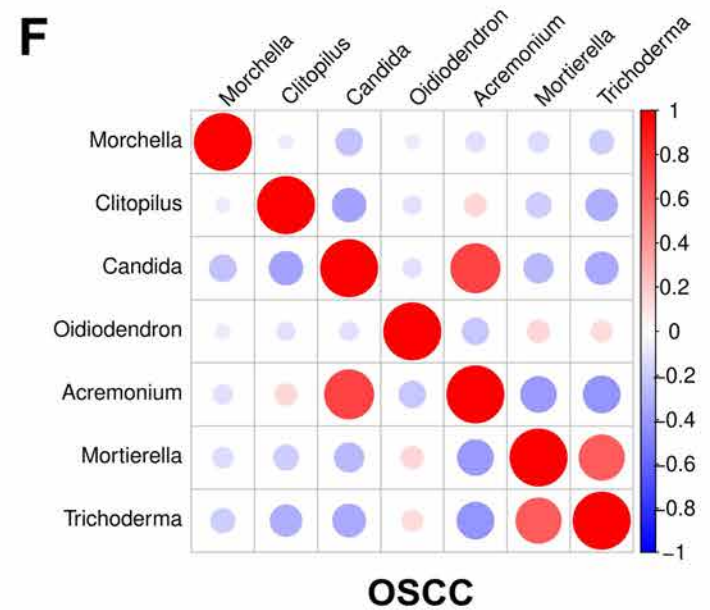
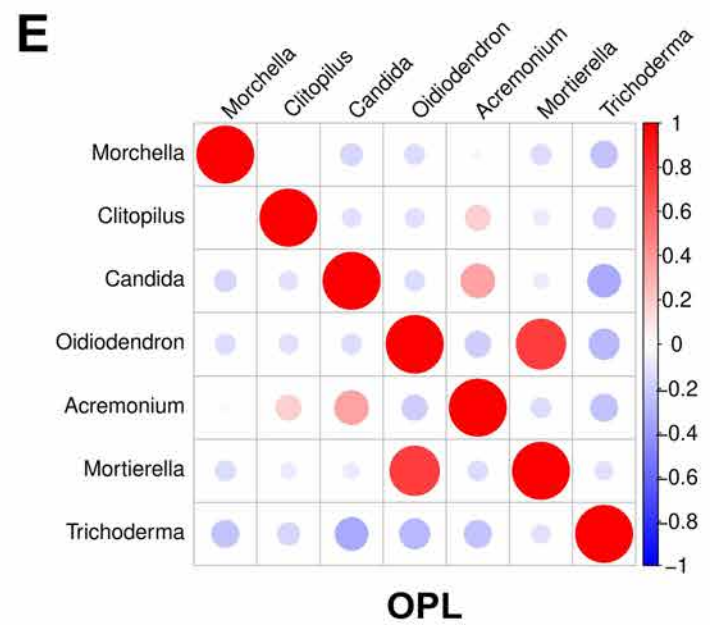
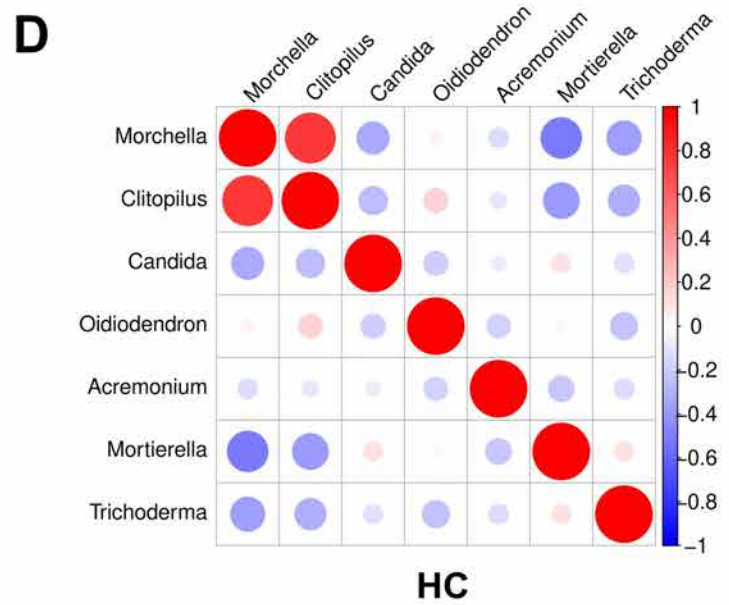
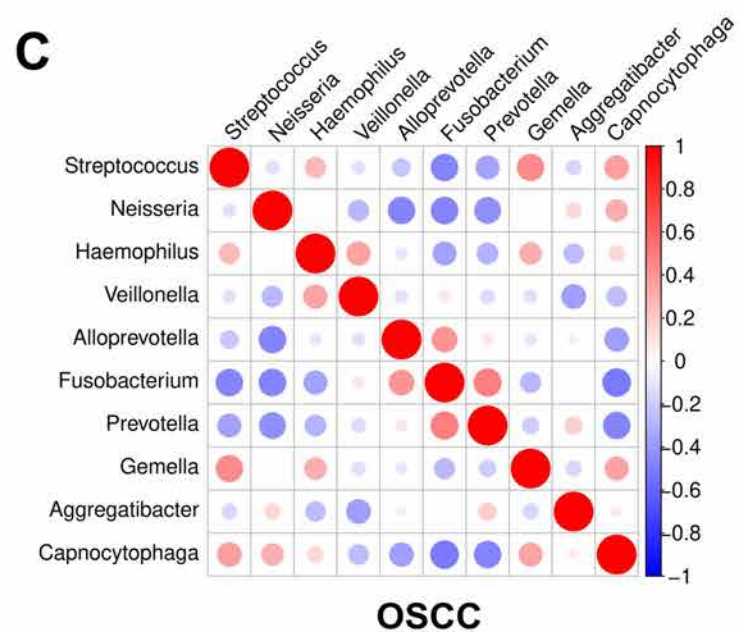
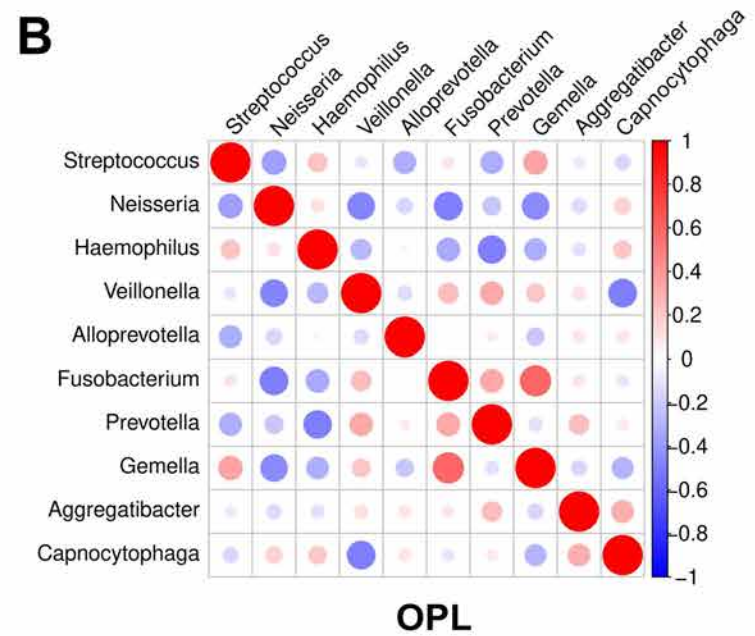
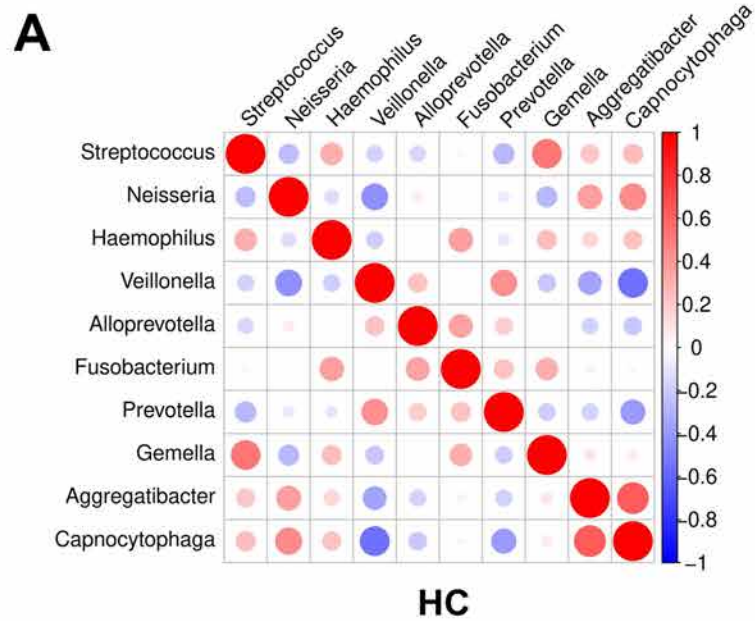
Supplementary Figure 10. Oral carcinogenesis-associated alterations in the abundance of the saliva fungi examined using LEfSe. (A) Cladogram indicates the phylogenetic distribution of active fungi that were remarkably enriched. (B) Bar plots at the species level with significant differences in abundance based on LEfSe. The color bars show the LDA scores of species that were enriched in the indicated condition: purple bar (HC), green bar (OPL), and pink bar (OSCC).



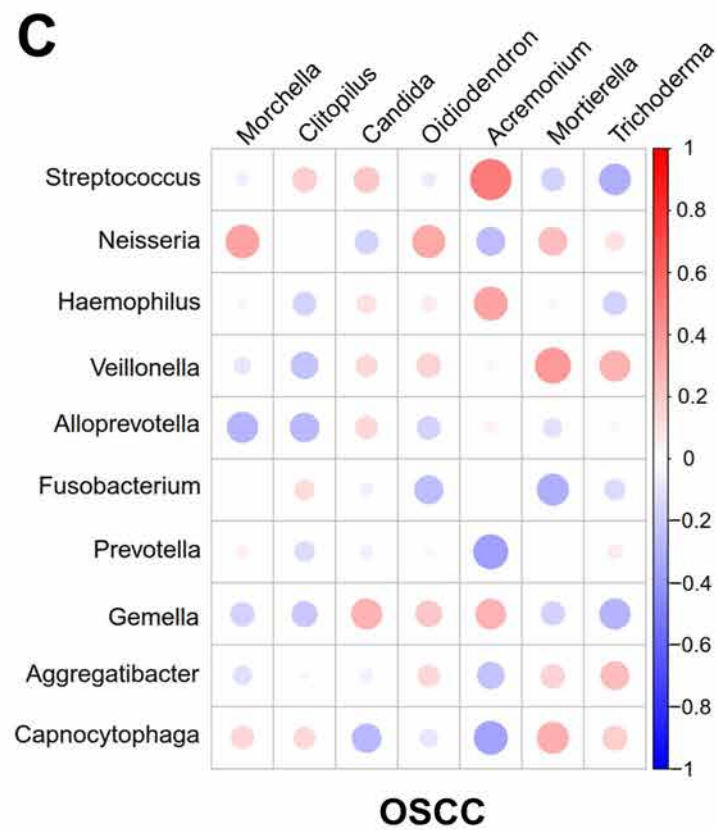
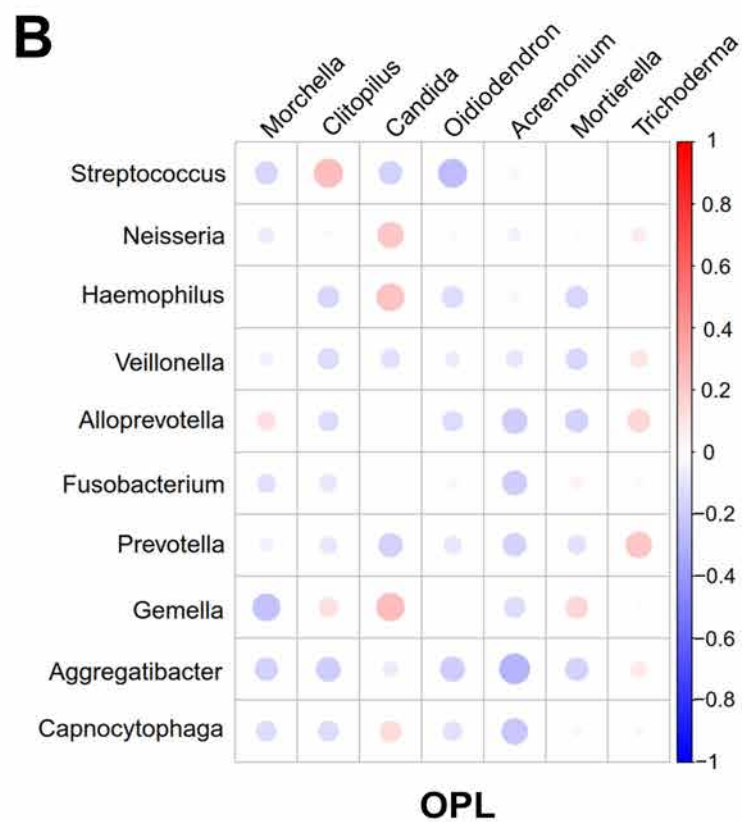
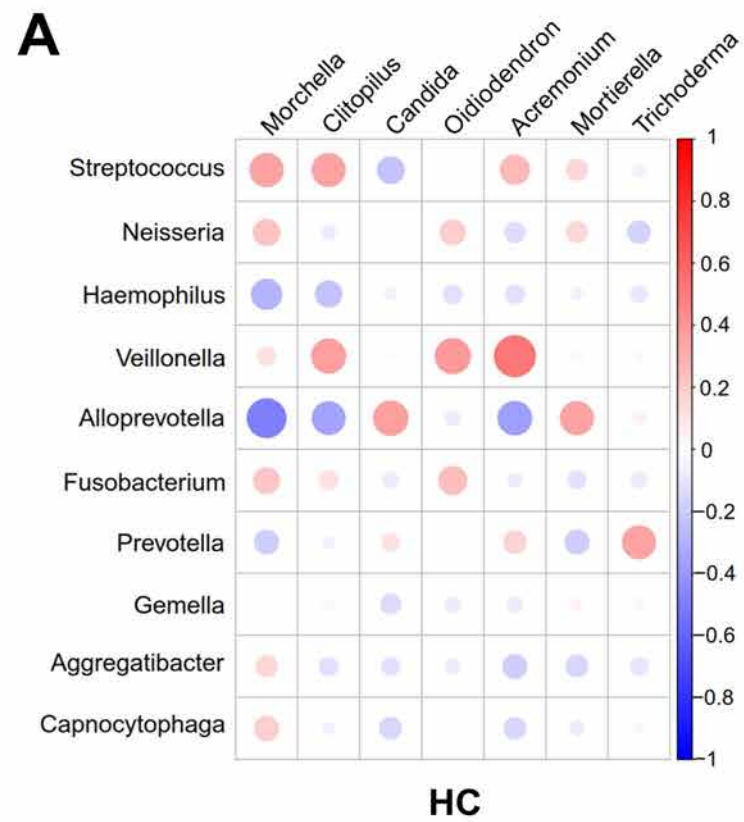
Supplementary Figure 11. Intra-kingdom correlation at the genera level in buccal mucosal samples. Intra-kingdom correlations are shown between (A–C) bacterial and (D–F) fungal genera in the buccal mucosal samples from the HC, OPL, or OSCC groups. Red: positive correlation, blue: negative correlation. Circle size and color shading indicate the value of the correlation coefficient, with bigger circles with darker coloring representing higher coefficient values (maximum = 1) and smaller circles with lighter coloring representing lower coefficient values (minimum = 0).



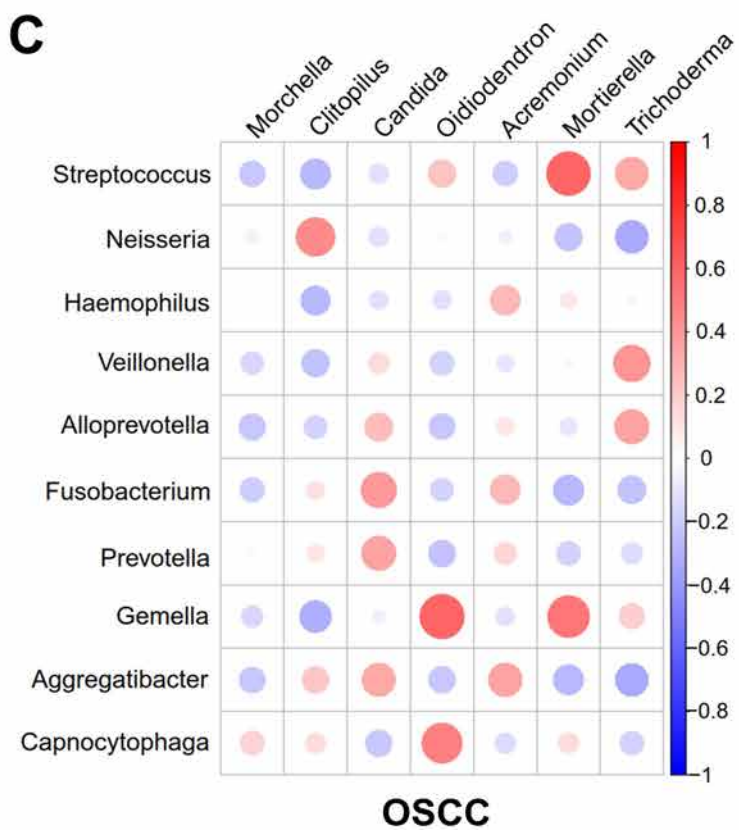
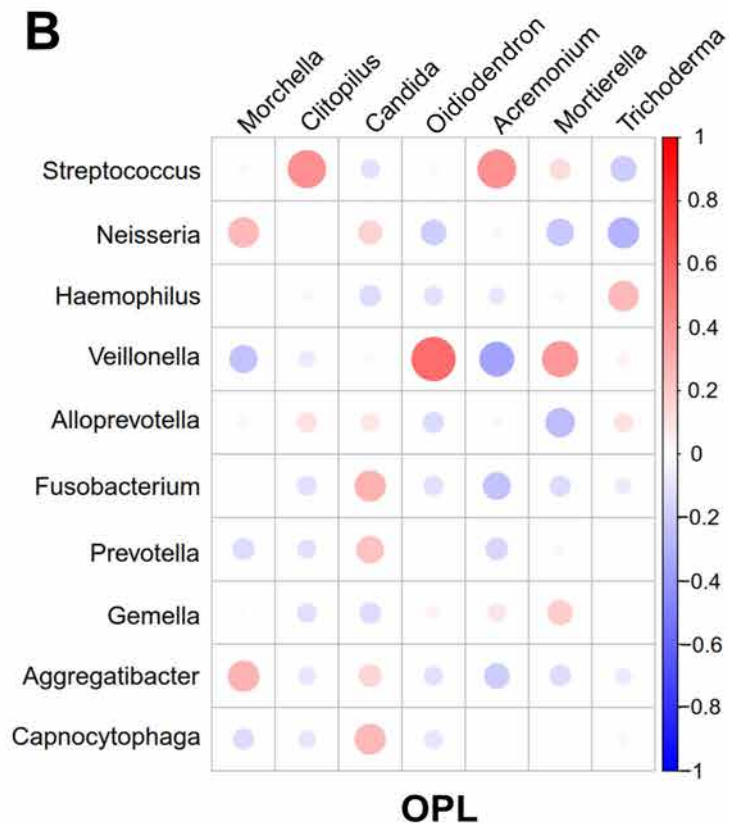
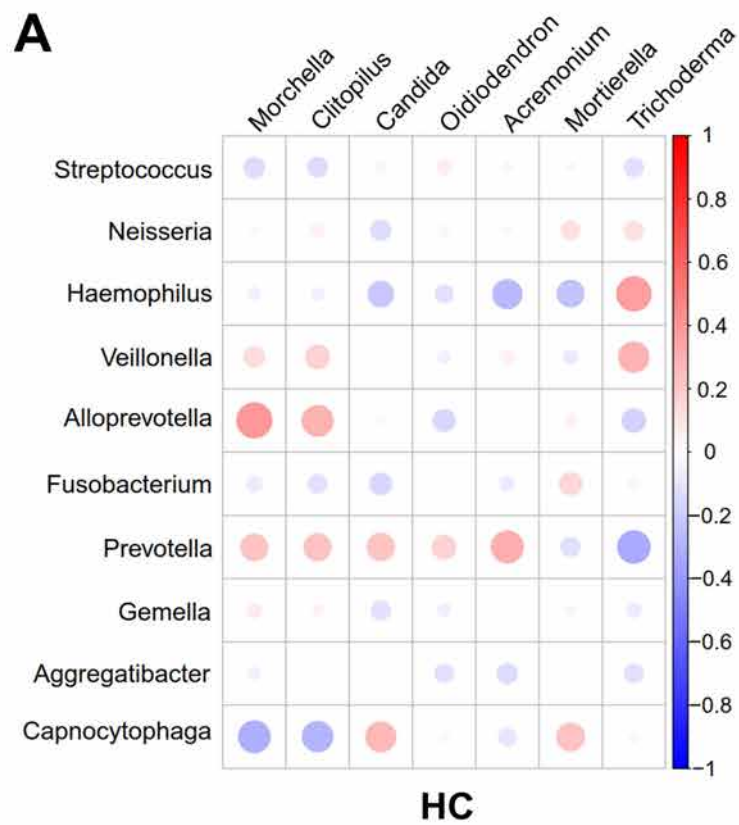
Supplementary Figure 12. Intra-kingdom correlation at the genera level in plaque samples. Intra-kingdom correlations are shown between (A–C) bacterial and (D–F) fungal genera in the plaque samples from the HC, OPL, or OSCC groups. Red: positive correlation, blue: negative correlation. Circle size and color shading indicate the value of the correlation coefficient, with bigger circles with darker coloring representing higher coefficient values (maximum = 1) and smaller circles with lighter coloring representing lower coefficient values (minimum = 0).



Supplementary Figure 13. Intra-kingdom correlation at the genera level in saliva samples. Intra-kingdom correlations are shown between (A–C) bacterial and (D–F) fungal genera in the saliva samples from the HC, OPL, or OSCC groups. Red: positive correlation, blue: negative correlation. Circle size and color shading indicate the value of the correlation coefficient, with bigger circles with darker coloring representing higher coefficient values (maximum = 1) and smaller circles with lighter coloring representing lower coefficient values (minimum = 0).

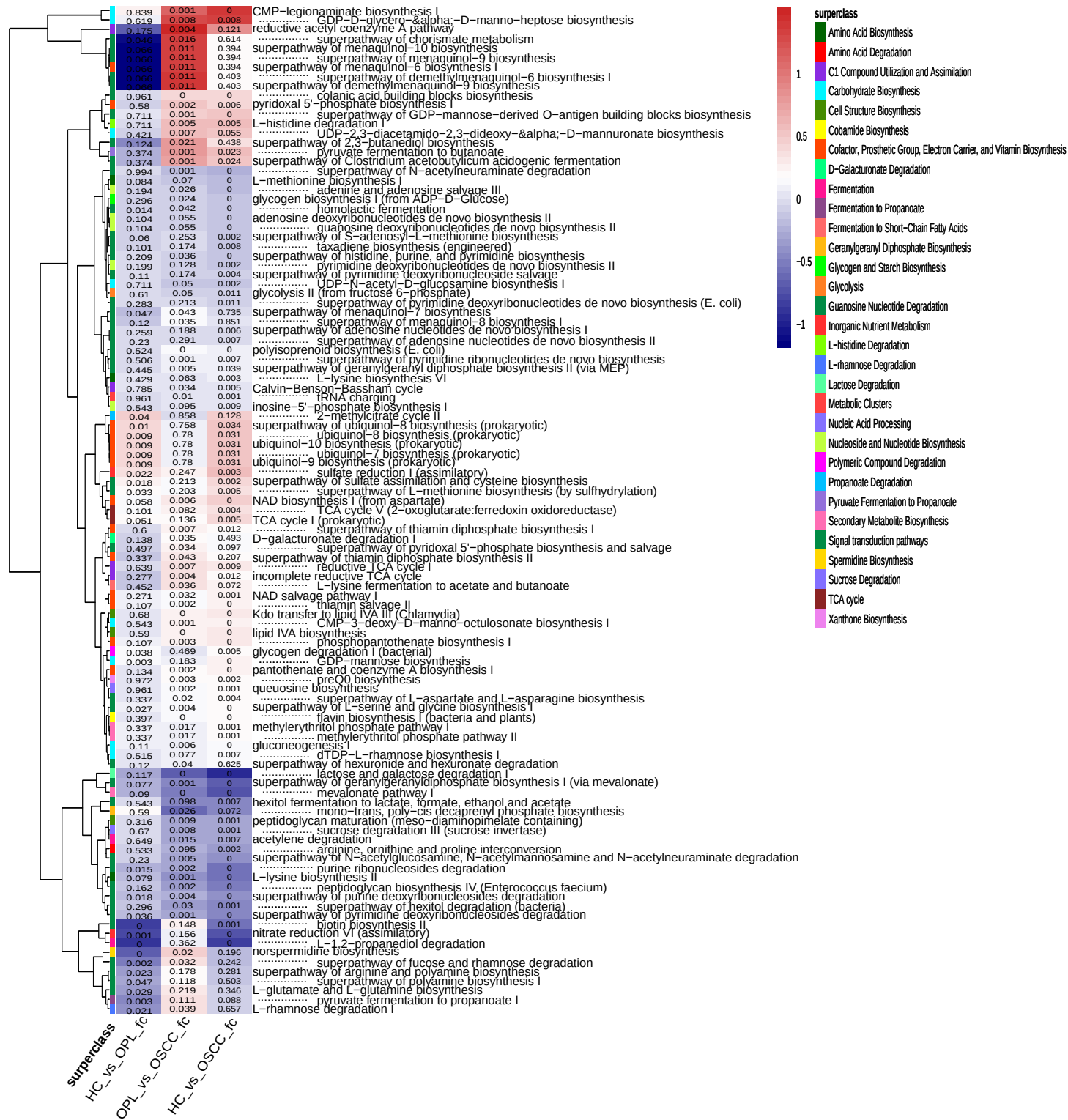


Supplementary Figure 14. Inter-kingdom correlation at the genera level in buccal mucosal samples. Inter-kingdom correlations were observed between bacterial and fungal genera in the buccal mucosal samples from the (A) HC, (B) OPL, or (C) OSCC groups. Red: positive correlation, blue: negative correlation. Circle size and color shading indicate the value of the correlation coefficient, with bigger circles with darker coloring representing higher coefficient values (maximum = 1) and smaller circles with lighter coloring representing lower coefficient values (minimum = 0).

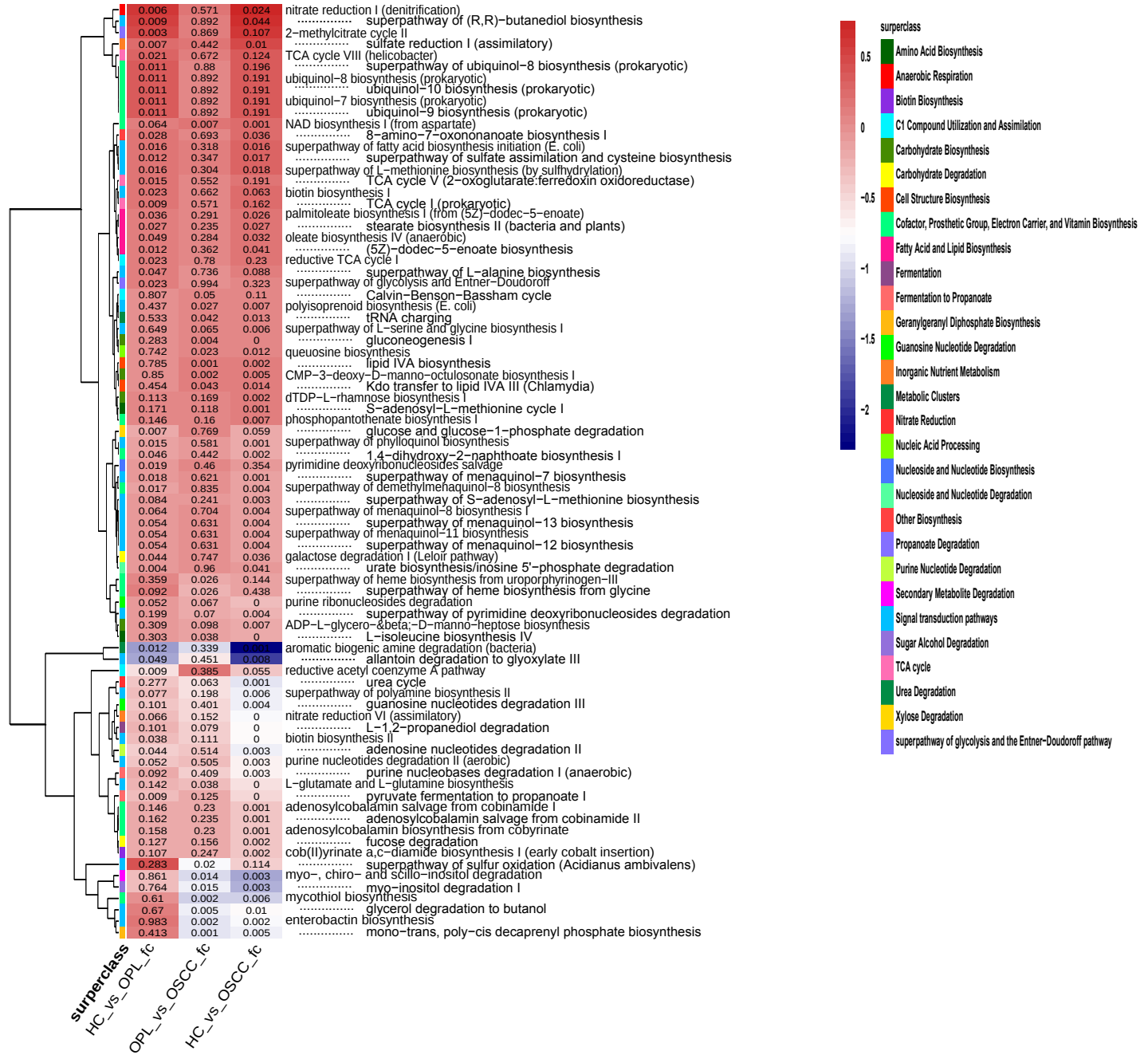


Supplementary Figure 15. Inter-kingdom correlation at the genera level in plaque samples. Inter-kingdom correlations were observed between bacterial and fungal genera in the plaque samples from the (A) HC, (B) OPL, or (C) OSCC groups. Red: positive correlation, blue: negative correlation. Circle size and color shading indicate the value of the correlation coefficient, with bigger circles with darker coloring representing higher coefficient values (maximum = 1) and smaller circles with lighter coloring representing lower coefficient values (minimum = 0).

Supplementary Figure 16. Inter-kingdom correlation at the genera level in saliva samples. Inter-kingdom correlations were observed between bacterial and fungal genera in the saliva samples from the (A) HC, (B) OPL, or (C) OSCC groups. Red: positive correlation, blue: negative correlation. Circle size and color shading indicate the value of the correlation coefficient, with bigger circles with darker coloring representing higher coefficient values (maximum = 1) and smaller circles with lighter coloring representing lower coefficient values (minimum = 0).



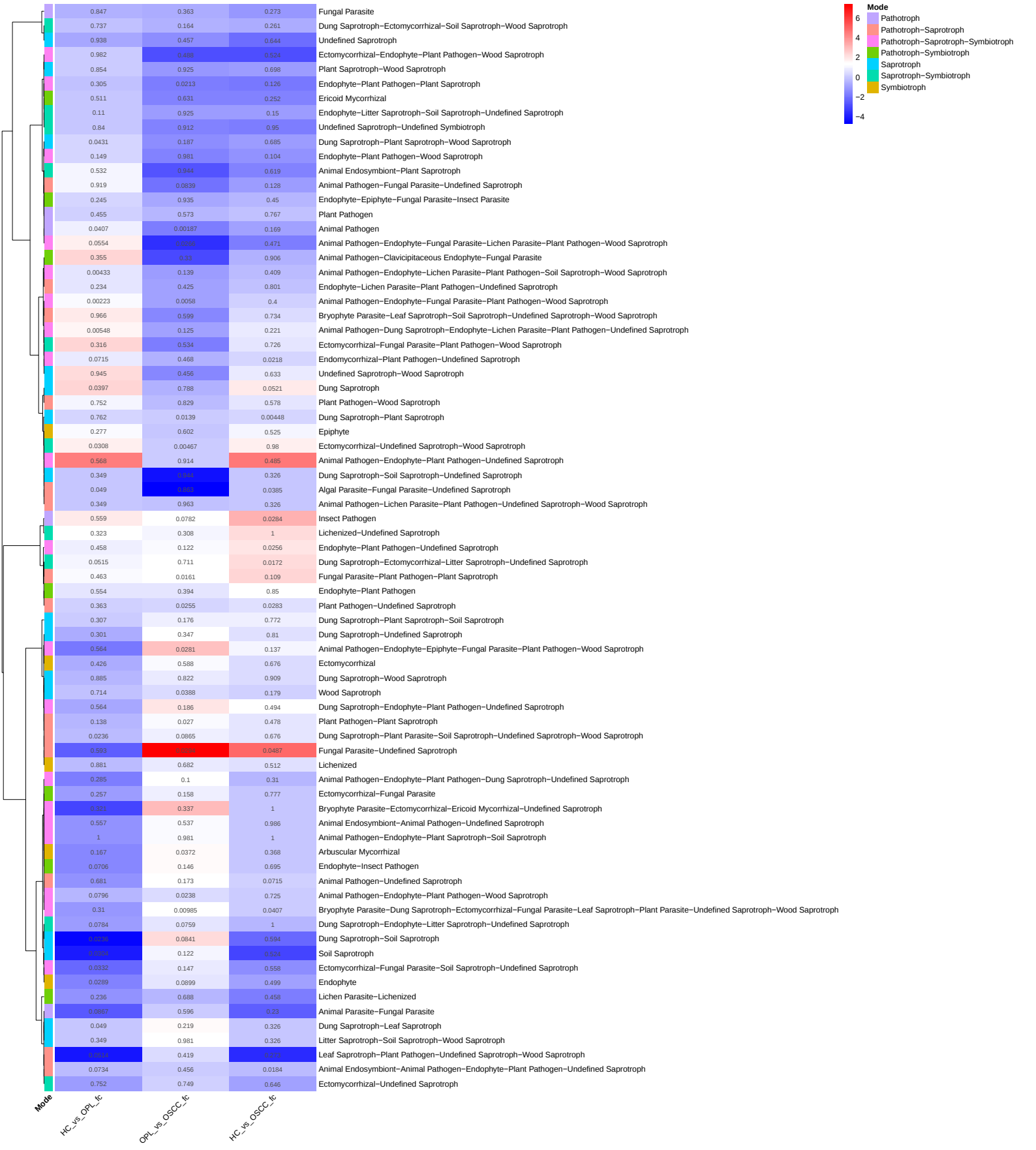
Supplementary Figure 17. Functional alterations in the plaque bacteriome. The relative abundance of functional pathways was compared among HC, OPL, and OSCC individuals. Differentially abundant pathways were plotted, and the exact *P* values are presented in the heatmap. Generalized fold change is indicated by color gradients. The generalized fold change > 0 : enriched in the latter; generalized fold change < 0 : enriched in the former.



Supplementary Figure 18. Functional alterations in the saliva bacteriome. The relative abundance of functional pathways was compared among HC, OPL, and OSCC individuals. Differentially abundant pathways were plotted, and the exact *P* values are presented in the heatmap. Generalized fold change is indicated by color gradients. The generalized fold change > 0 : enriched in the latter; generalized fold change < 0 : enriched in the former.



Supplementary Figure 19. Functional alterations in the plaque mycobiome. The relative abundance of functional pathways was compared among HC, OPL, and OSCC individuals. Differentially abundant pathways were plotted, and the exact *P* values are presented in the heatmap. Generalized fold change is indicated by color gradients. The generalized fold change > 0 : enriched in the latter; generalized fold change < 0 : enriched in the former.



Supplementary Figure 20. Functional alterations in the saliva mycobiome. The relative abundance of functional pathways was compared among HC, OPL, and OSCC individuals. Differentially abundant pathways were plotted, and the exact *P* values are presented in the heatmap. Generalized fold change is indicated by color gradients. The generalized fold change > 0 : enriched in the latter; generalized fold change < 0 : enriched in the former.

Supplementary Table 1 Baseline information of the HC, OPL and OSCC groups

Characteristic	HC (n=30)	OPL (n=32)	OSCC (n=29)	<i>F</i> / χ^2	<i>P</i> value
Age, y	56.63±11.12	56.00±13.67	61.97±10.11	2.295	0.107
Gender, No. (%)				1.338	0.512
Female	15(50.00)	19(59.37)	13(44.83)		
Male	15(50.00)	13(40.63)	16(55.17)		
Smoking, No. (%)				0.018	0.991
Yes	14(46.67)	15(46.88)	14(48.28)		
No	16(53.33)	17(53.12)	15(51.72)		
Drinking, No. (%)				0.169	0.919
Yes	16(53.33)	16(50.00)	16(55.17)		
No	14(46.67)	16(50.00)	13(44.83)		
Dysplasia, No. (%)					
No		13(40.63%)			
Low grade		13(40.63%)			
High grade		6(18.75%)			
Differentiation,					
No. (%)					
Middle grade			14(48.28%)		
High grade			15(51.72%)		
pT stage, No. (%)					
T1			7(24.14%)		
T2			10(34.48%)		
T3			6(20.69%)		
T4			6(20.69%)		
pN stage, No. (%)					
N0			19(65.52%)		
N1			6(20.69%)		
N2			4(13.79%)		

HC, healthy control. OPL, oral premalignant lesion. OSCC, oral squamous cell carcinoma. pT stage, pathological T stage. pN stage, pathological N stage. Quantitative data are presented as mean ± standard deviation; categorical data are presented as frequency and percentage.