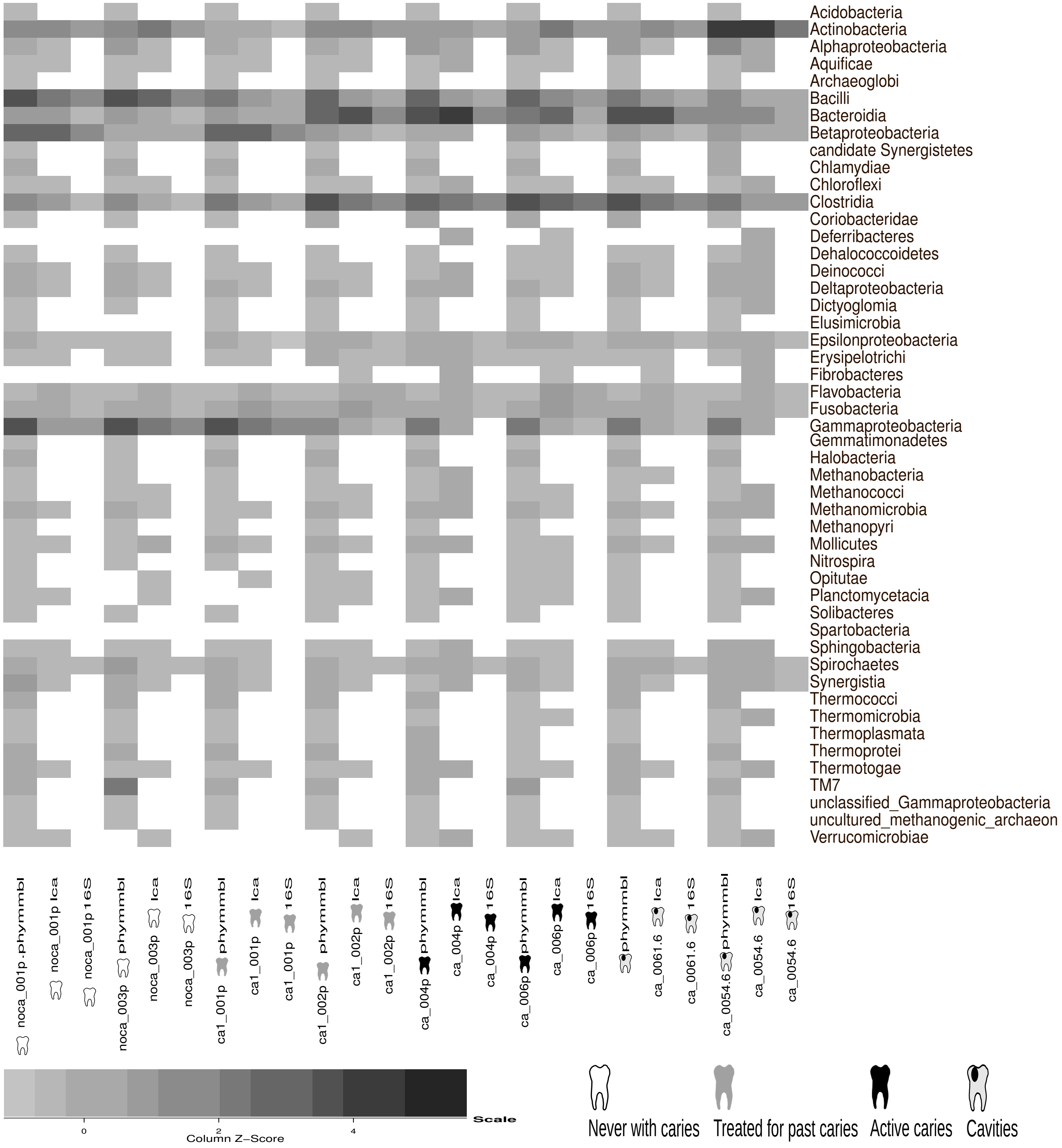
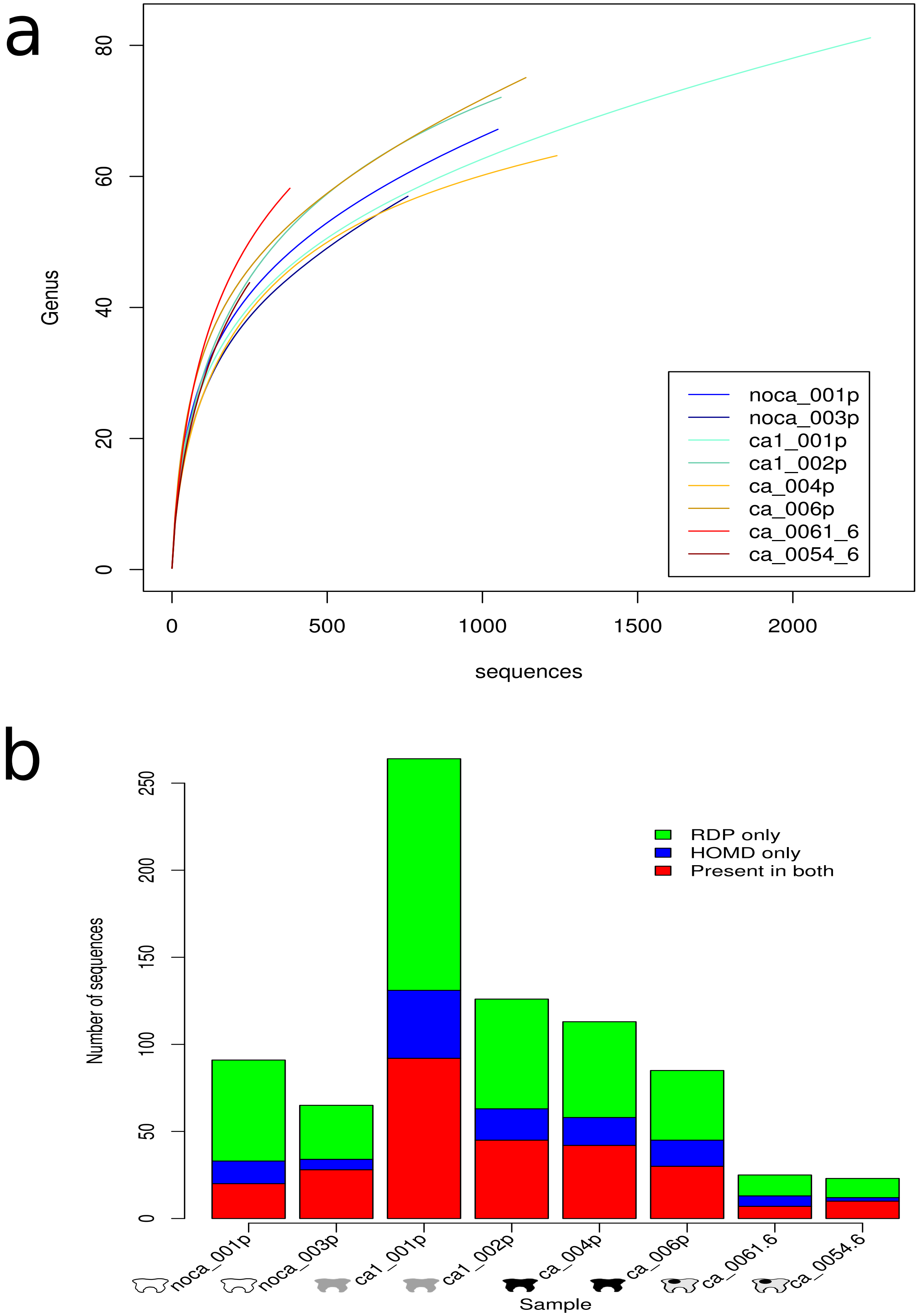
Belda-Ferre *et al.* Supplementary Fig 1

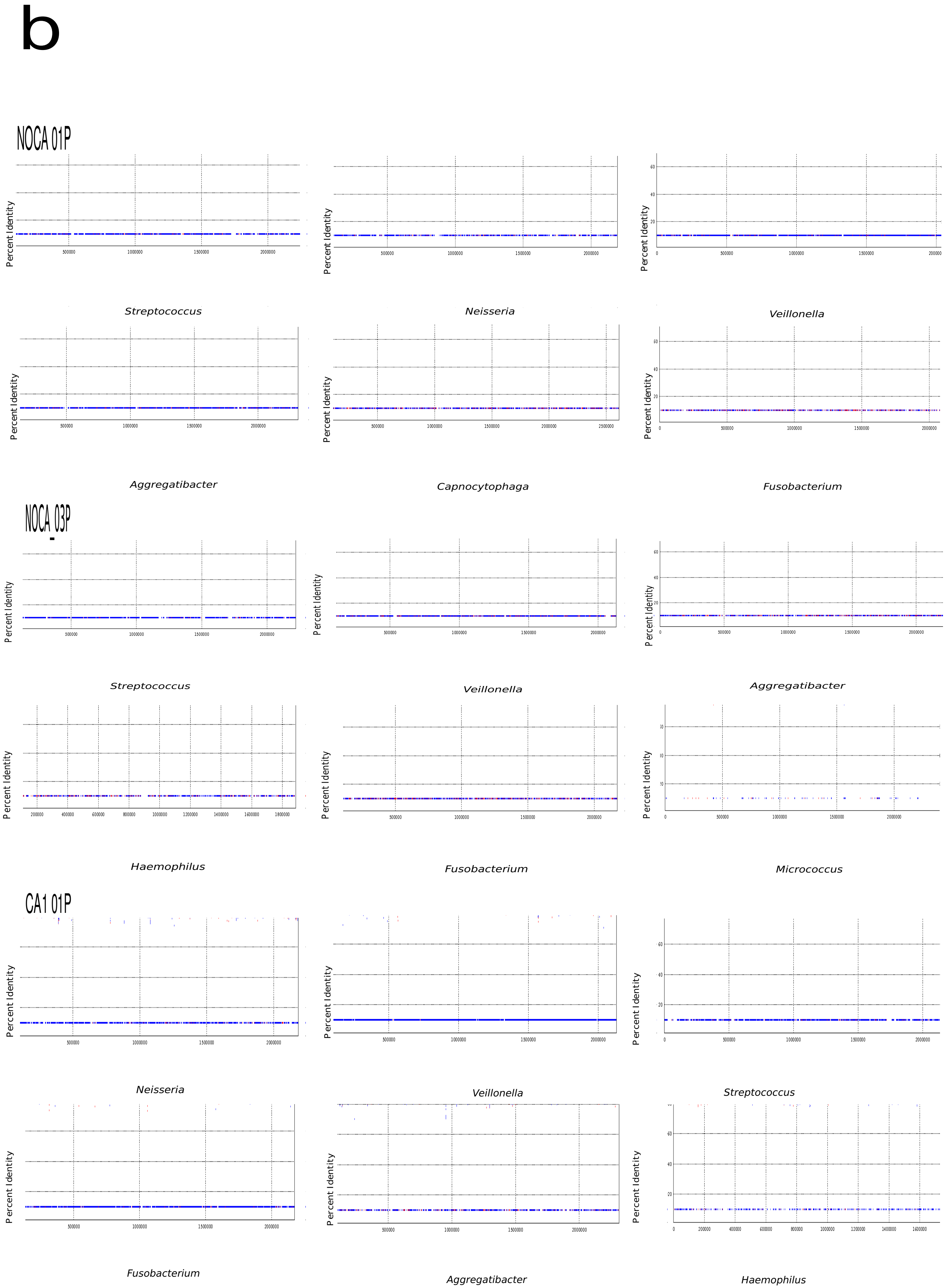


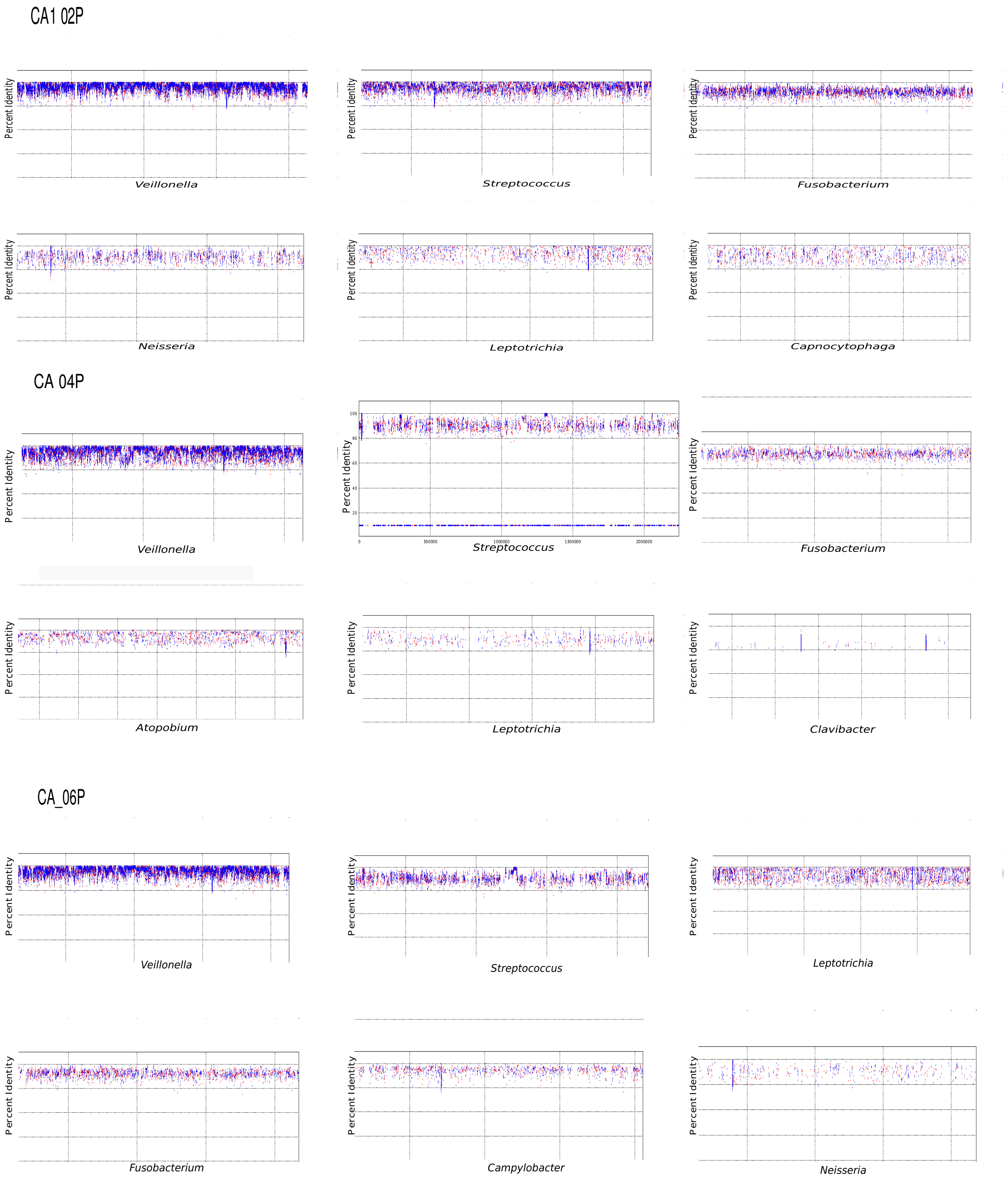
Supplementary Fig 1 | Bacterial diversity in the metagenome of the oral cavity by 16S rRNA, LCA and phymmBL approaches. Similar results are obtained for the most common microbial groups, but 16S rRNA reads detect a smaller fraction of taxa**.**



Supplementary Figure 2. (a) Rarefaction curves for the 8 metagenomic samples studied. Curves were estimated based on the 16S rRNA reads extracted from the metagenomes. No clear differences in diversity are observed between the health status of the samples. (b) New Operational Taxonomic Units (OTUs) in the oral cavity. Sequences corresponding to 16S rRNA genes were BLASTed against the Ribosomal Database Project (RDP) and the Human Oral Microbiome Database (HOMD) using reads which were over 400 bp and which aligned over 90% of the length. A read was considered to represent a new OTU if the sequence identity against its top hit was between 80-95%. The graphs show the number of new OTUs with hits in the RDP, the HOMD or both. A complete list is available from the authors upon request.

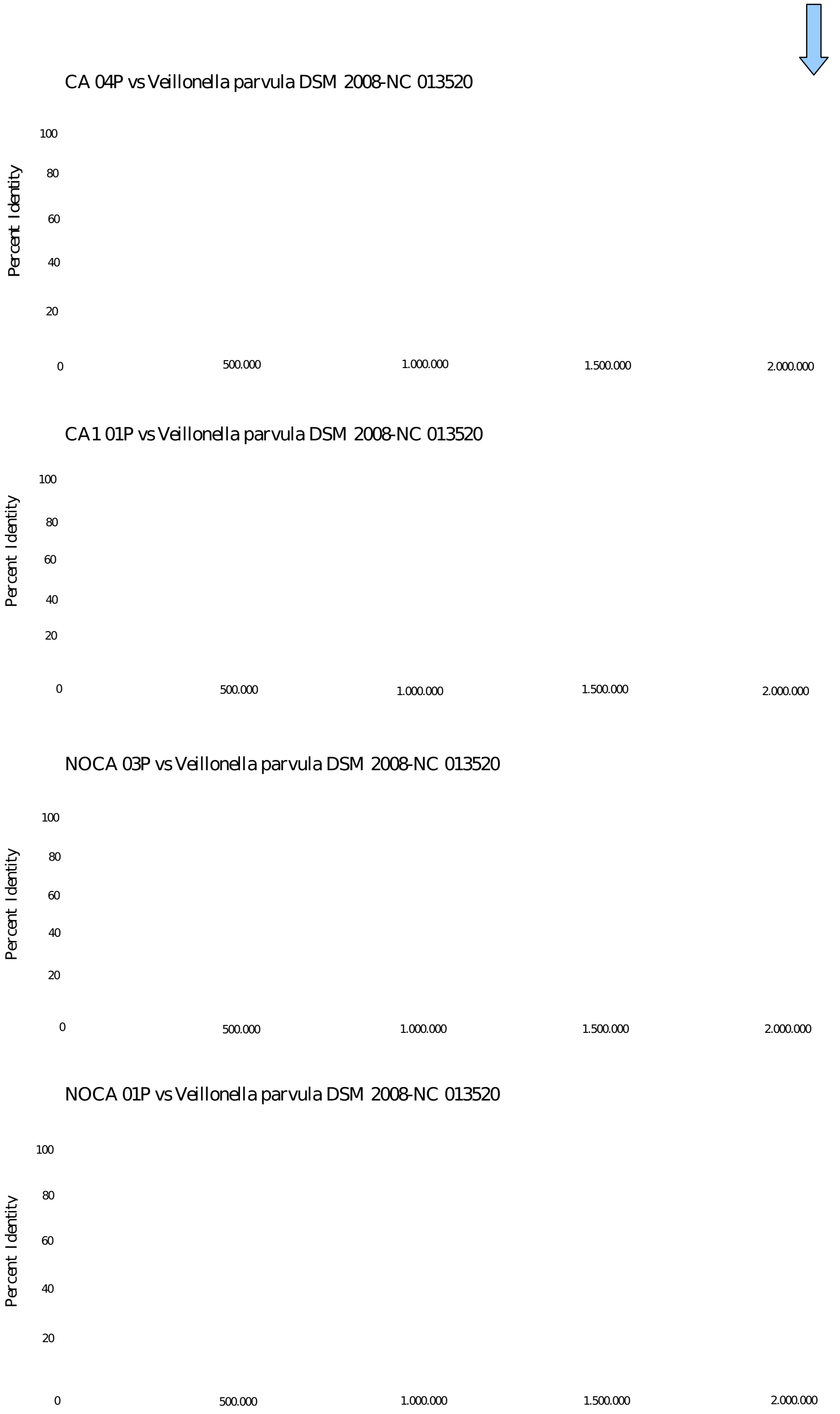
Belda-Ferre et al. Supplementary Fig.3


b
NOCA 01P
Percent Identity
500000
1000000
1500000
2000000
2500000
0
20
40
60
Streptococcus
Neisseria
Veillonella
Aggregatibacter
Capnocytophaga
Fusobacterium
NOCA_03P
Streptococcus
Veillonella
Aggregatibacter
200000
400000
600000
800000
1000000
1200000
1400000
1600000
1800000
Haemophilus
Fusobacterium
Micrococcus
CA1 01P
Neisseria
Veillonella
Streptococcus
Fusobacterium
Aggregatibacter
Haemophilus



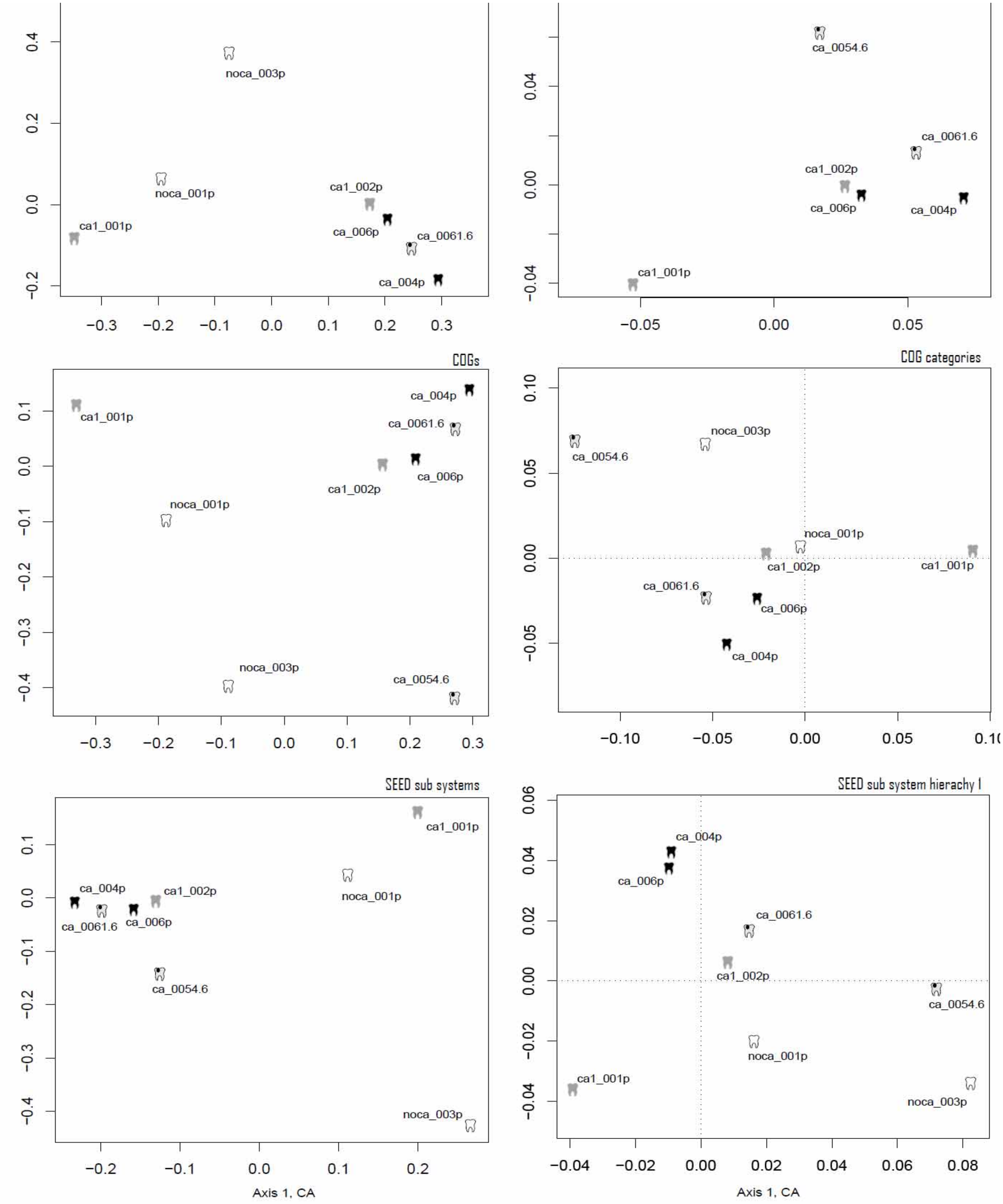
Supplementary Fig. 3 | Metagenomic recruitment of the 4 most common species per sample (**a**), and the 6 most common genera (**b**) against six metagenomes corresponding to supragingival dental plaque samples. A species was assumed to be present in the metagenome if the mode of the frequency distribution of percent identity was above 94%, following Konstantinidis and Tiedje[31]. A genus was assumed to be present in the metagenome if the mode of the frequency distribution of percent identity was above 90%. For genera recruitment plots, the graph corresponding to the most common species for each genus was selected as a representative

Belda-Ferre *et al.* Supplementary Fig 4



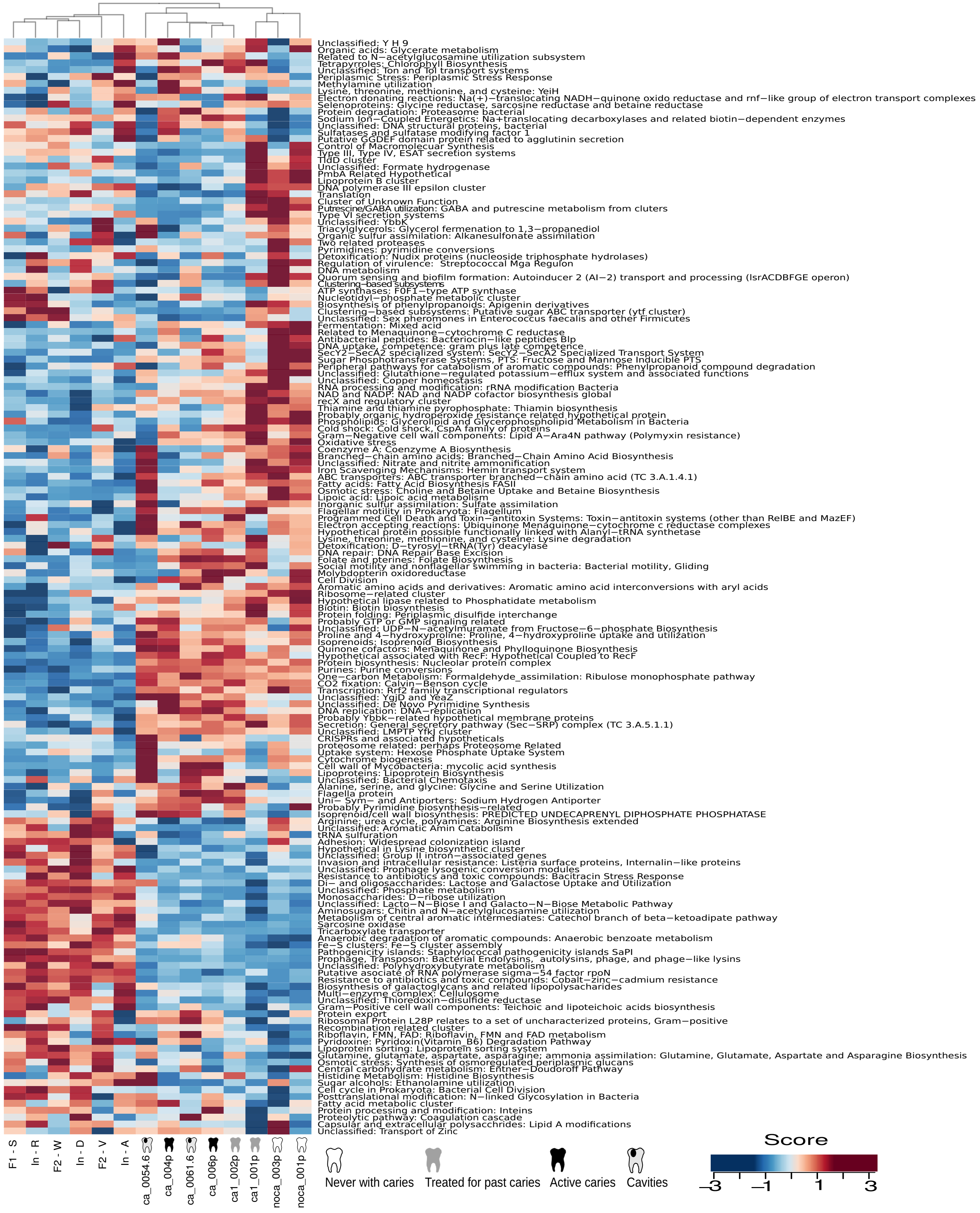
Supplementary Fig. 4 | Metagenomic recruitment of the species *Veillonella parvula* at the protein level in four supragingival dental plaque samples from the oral microbiome. Plots were done using Promer (see supplementary methods). The arrow indicates a genomic island that is absent in the two individuals without caries, showing differences in gene content between strains of this species inhabiting different individuals. Genes in the island were still present in samples CA_04P and CA1_01P under lower levels of coverage equivalent to those of samples NOCA_01P and NOCA_03P.

Belda-Ferre *et al.* Supplementary Fig 5



Supplementary Fig. 5 | Correspondence Analyses (CoAs) of the functional annotation of oral samples. This figure depicts the plot of the first 2 axes of a CoA performed on a table of frequencies of functional classes per sample. Six independent analyses were done over different annotations systems: TigrFam, COG and SEED Subsystems in the right column; and their respective integration into higher hierarchies in the left column: TigrFam main roles, COG categories, and Subsystem Hierarchy 1. Similar clustering of the samples were obtained with the different methods

Belda-Ferre *et al.*, Supplementary Fig. 6



Supplementary Fig. 6 | Functional profiles from oral and gut metagenomic samples. Classification was based on SEED Subsystem Hierarchy 2. Counts were first normalized to the total number of reads per sample and then normalized by function. Blue to red gradient indicates levels of under/ over-representation. Large blocks of gene categories are over-represented in each of the two microbiotas, indicating that the gut and the oral cavity are two functionally distinct ecosystems. Within the oral samples, some categories are over-represented in individuals without dental caries.

Supplementary Table 1. Features of oral samples and their metagenomes.

Age	Sex	Sample[1]	CAO's Index[2]	Number of reads	% Replic[3]	% Human DNA	Total Mbp	Contigs >5kbp	Largest contig	N50 Contig Size	16S reads[4]	Simpson Index[5]	Shannon Index[5]	Chao1 Index[6]	ACE Index[7]
23	Male	NOCA_01P	0	347927	1.2	40.59	77.54	13	12856	898	543	0.93	3.19	100 ± 24.6	94.6 ± 4.9
39	Male	NOCA_03P	0	330073	3.95	22.76	100.13	49	43857	1083	374	0.91	2.94	92 ± 28.4	83.7 ± 4.8
36	Male	CA1_01P	8 (1)	494659	3.93	2.23	203.71	657	46856	2230	1160	0.94	3.21	120 ± 24.8	120.4 ±5.8
29	Male	CA1_02P	6 (4)	315892	3.95	2.74	129.85	154	15919	1071	575	0.92	3.11	85.2 ± 9	89.9 ± 4.7
36	Fem	CA_04P	25 (15)	402049	4.54	11.54	142.37	181	19835	939	663	0.89	2.89	74.4 ± 9.9	73.5 ± 4.2
49	Male	CA_06P	11 (8)	354192	2.88	10.83	123.27	47	51033	872	615	0.95	3.38	129.2 ± 41	115.9 ±5.8
49	Male	CA_06_1.6	11 (8)	305820	3.25	66.97	37.52	0	3376	667	194	0.92	3.21	77 ± 13.3	77.1 ± 4.3
42	Male	CA_05_4.6	10 (7)	291162	3.19	74.99	27.67	2	29784	661	130	0.88	2.82	55.3 ± 8.3	66.3 ± 4.6

[1] Samples marked with "P" indicate supranginvival dental plaque samples. Samples with a number code indicate the tooth from which the cavity sample was taken, following the international WHO nomenclature

[2] Number of caried, absent and obstructed teeth (wisdom teeth were excluded). Number between brackets indicate the number of exposed caries

[3] Proportion of filtered artificial replicates during pyrosequencing

[4] Number of 16S rRNA sequences detected in the metagenome and assigned by the RDP classifier.

[5] Diversity indexes were calculated at the genus level based on 16S rRNA sequences extracted from the metagenomes

[6] Data indicate Chao1 richness index (number of expected genera in the sample) and its corresponding standard errors

[7] Data indicate ACE richness index (number of expected genera in the sample) and its corresponding standard errors

Supplementary Table 2. Level of funcional assignment of the metagenomic sequences.

ID	Health status	Total reads	cd (n) [a]	cd (%)	cog (n) [b]	cog (%)	Tfam (n) [c]	Tfam (%)	seed (n) [d]	seed(%)
NOCA_01P	○	204218	126729	62	108929	53	82457	40	111497	50
NOCA_03P	○	244881	116575	48	95327	39	74356	30	93391	38
CA1_01P	▨	464594	321997	69	280652	60	214050	46	271868	59
CA1_02P	▨	295072	182091	62	150966	51	118716	40	146161	55
CA_04P	●	339503	192003	57	161384	48	126281	37	158887	48
CA_06P	●	306740	182349	59	151524	49	119477	39	146032	47
CA_05_4.6	◉	70503	40999	58	31864	45	26245	37	29625	42
CA_06_1.6	◉	97722	54305	56	45440	46	35395	36	44552	46

(n): absolute count
(%): percentage of the total reads in sample
[a] cdd : conserved domains of NCBI Conserved Domains Database
[b] cog: cluster of orthologous groups
[c] Tfam: Tigr Fams
[d] seed : Seed / MG-RAST sub systems

Supplementary Table 3 | Potential new Operational Taxonomic Units (OTUs) in the oral cavity. Sequences corresponding to 16S rRNA genes were extracted from the 8 metagenomes. BLASTN was performed against the Ribosomal Database Project and the Human Oral Microbiome Database (HOMD) using reads which were over 400 bp and which aligned over 90% of the length. Only OTUs with hits between 80-95% nucleotide identity were considered.

CLOSEST SPECIES	**Number of Reads**	**%ID**
uncultured bacterium	71	0.808-0.946
Acetobacter pasteurianus IFO 3283-01-42C	61	0.801-0.886
Streptococcus pyogenes MGAS10750	33	0.812-0.949
Mycoplasma arthritidis 158L3-1	23	0.801-0.864
Lactococcus lactis	20	0.801-0.945
Helicobacter mustelae 12198	18	0.802-0.86
Bacteroides vulgatus	17	0.803-0.898
Dickeya chrysanthemi	17	0.828-0.932
Actinobacillus pleuropneumoniae L20	16	0.808-0.929
Neisseria meningitidis 8013	16	0.808-0.943
Dechloromonas sp. HZ	16	0.812-0.896
Streptococcus sp.	15	0.805-0.949
Prevotella sp.	12	0.883-0.947
Francisella noatunensis subsp. noatunensis	12	0.803-0.915
Bacteroides helcogenes	12	0.832-0.935
Kinetoplastibacterium blastocrithidii	11	0.840-0.939
Veillonella parvula DSM 2008	11	0.815-0.903
Pseudomonas filiscindens	11	0.824-0.883
psychrophilic marine bacterium PS32	10	0.808-0.949
Bacillus cereus	9	0.824-0.896
Prevotella melaninogenica (T)	9	0.841-0.944
Parabacteroides goldsteinii	8	0.804-0.910
Bacteroides massiliensis	8	0.828-0.924
uncultured *Moraxellaceae bacterium*	8	0.834-0.948
Lactobacillus rhamnosus	8	0.811-0.907
Capnocytophaga sp.	7	0.911-0.946
Porphyromonas sp. oral clone EP003	7	0.857-0.923
uncultured *Veillonella sp.*	7	0.801-0.939
Bacillus pumilus	7	0.816-0.909
uncultured *Porphyromonas sp.*	6	0.917-0.929
Haemophilus parasuis SH0165	6	0.808-0.935
uncultured *Tsukamurella sp.*	6	0.810-0.923
Actinomyces naeslundii	6	0.841-0.948
Veillonella parvula	6	0.872-0.948
Fusobacterium nucleatum (T)	6	0.817-0.934
Neisseria sicca	6	0.912-0.949

Streptococcus sanguinis SK36	6	0.816-0.931
Actinomyces sp.	5	0.885-0.939
Bacteroides stercoris ATCC 43183	5	0.833-0.895
Prevotella melaninogenica	5	0.853-0.947
Fusobacterium sp. oral clone ASCF06	5	0.83-0.933
TM7 [G-1] sp.	5	0.839-0.947
uncultured *Actinomyces sp.*	5	0.817-0.923
uncultured *Prevotella sp.*	5	0.830-0.910
Parabacteroides distasonis	4	0.838-0.922
Actinomyces odontolyticus	4	0.844-0.941
uncultured *Corynebacterium sp.*	4	0.933-0.944
Campylobacter gracilis	4	0.803-0.914
uncultured *Neisseria sp.*	4	0.813-0.876
Moraxella catarrhalis	4	0.872-0.917
Burkholderia glathei (T)	4	0.855-0.865
Corynebacterium sp.	4	0.815-0.942
Terrahaemophilus aromaticivorans	4	0.929-0.942
Actinomyces sp. oral strain Hal-1065	3	0.836-0.844
Streptococcus mitis	3	0.816-0.948
Bacteroides tectus	3	0.871-0.921
Selenomonas sp.	3	0.916-0.931
Fusobacterium nucleatum ss. nucleatum	3	0.886-0.949
Aggregatibacter aphrophilus	3	0.863-0.897
Actinomyces sp. oral clone IP073	3	0.818-0.843
Leptotrichia sp. oral clone HE012	3	0.8-0.8897
Parabacteroides merdae	3	0.880-0.918
Bacteroides coprocola	3	0.867-0.886
Moraxella bovoculi (T)	3	0.871-0.934
Bacteroides thetaiotaomicron	3	0.841-0.909
uncultured candidate division TM7	2	0.882-0.889
Haemophilus parainfluenzae	2	0.939-0.946
Eikenella corrodens	2	0.867-0.876
uncultured *Capnocytophaga sp.*	2	0.932-0.944
Bacteroides intestinalis	2	0.805-0.904
Bifidobacterium adolescentis (T)	2	0.830-0.836
Bacteroides salyersiae	2	0.839-0.887
Prevotella denticola	2	0.910-0.948
Rehmannia glutinosa var. purpurea' phytoplasma	2	0.801-0.830
Moraxella sp.	2	0.916-0.933
Acetobacter pasteurianus IFO 3283-07	2	0.805-0.807
Streptococcus anginosus	2	0.876-0.944
uncultured *Abiotrophia sp.*	2	0.849-0.912
Actinomyces israelii	2	0.942-0.944
uncultured *Megasphaera sp.*	2	0.830-0.874
Xanthomonas translucens pv. poae	2	0.822-0.834
Terrahaemophilus sp.	2	0.914-0.915

Bacteroides caccae	2	0.919-0.937
Aggregatibacter sp.	2	0.874-0.934
Prevotella salivae (T)	2	0.830-0.929
Actinomyces sp. oral taxon 180	2	0.830-0.944
Selenomonas sputigena	2	0.817-0.854
Veillonella dispar	2	0.947-0.949
uncultured *Selenomonas sp.*	2	0.803-0.891
Actinomyces oris	2	0.880-0.889
Fusobacterium nucleatum ss. vincentii	2	0.896-0.942
Clostridium difficile 630	2	0.803-0.845
Pectobacterium atrosepticum	2	0.887-0.91
Prevotella denticola (T)	2	0.831-0.947
Selenomonas sp. oral clone GT010	2	0.931-0.939
Kingella oralis (T)	2	0.808-0.934
Parascardovia denticolens	2	0.834-0.910
Neisseria elongata (T)	2	0.807-0.903
Selenomonas infelix (T)	2	0.895-0.897
Rothia dentiocariosa	2	0.895-0.900
Actinomyces sp. oral clone IO076	2	0.907-0.918
Bacteroides uniformis	1	0.81
Capnocytophaga granulosa	1	0.8
Rothia sp. oral taxon 188	1	0.95
Prevotella sp. oral clone GI032	1	0.82
Haemophilus sp.	1	0.95
Actinomyces israelii (T)	1	0.95
uncultured *Gemella sp.*	1	0.81
uncultured *Pseudanabaena sp.*	1	0.9
Gemella sp. oral clone ASCE02	1	0.85
Prevotella tannerae	1	0.92
Bacteroides acidifaciens	1	0.92
Campylobacter concisus	1	0.85
Bergeyella sp.	1	0.94
Lactobacillus paracasei subsp. paracasei	1	0.82
Streptococcus sp. oral clone ASCG04	1	0.86
Streptococcus sp. oral clone ASCA03	1	0.83
Capnocytophaga sp. oral clone BR085	1	0.82
Lachnospiraceae [G-4] sp.	1	0.94
Streptococcus suis BM407	1	0.9
Neisseria gonorrhoeae	1	0.95
Micrococcus luteus NCTC 2665	1	0.95
Olsenella uli	1	0.94
Pseudomonas aeruginosa	1	0.86
Atopobium parvulum DSM 20469	1	0.85
Prevotella sp. oral clone BE073	1	0.8
Capnocytophaga sp. AHN9756	1	0.94
Prevotella intermedia	1	0.91

Leptotrichia sp. oral clone IK040	1	0.93
Aster yellows phytoplasma B	1	0.82
Kytococcus sedentarius	1	0.81
Kordia algicida (T)	1	0.9
Tannerella forsythensis	1	0.92
Rothia sp.	1	0.81
Xanthomonas axonopodis pv. citrumelo	1	0.81
Capnocytophaga sp. AHN9687	1	0.94
Neisseria pharyngis	1	0.92
Streptococcus gordonii	1	0.93
Eubacterium sp. oral clone DO016	1	0.91
Porphyromonas sp. oral clone DP023	1	0.95
Streptococcus sp. Culture clone SRC DSC22	1	0.89
Actinomyces oricola	1	0.93
Chryseobacterium sp. IMMIB L-1519	1	0.85
Xanthomonas axonopodis pv. syngonii	1	0.85
Streptococcus mitis B6	1	0.93
Prevotella oralis	1	0.81
Burkholderia sp. m35b	1	0.94
TM7 [G-3] sp.	1	0.87
Rothia dentocariosa (T)	1	0.92
Bacteroides-like sp. oral clone AU126	1	0.86
Neisseria sp.	1	0.86
Lautropia mirabilis	1	0.89
Fusobacterium nucleatum ss. animalis	1	0.89
Lachnospiraceae bacterium 'Oral Taxon 107'	1	0.9
uncultured *eubacterium*	1	0.92
uncultured candidate division *SR1 bacterium*	1	0.85
Clostridiales bacterium CD3:22	1	0.84
Eubacterium [XIVa][G-1] saburreum	1	0.94
Treponema pectinovorum	1	0.86
Fusobacterium nucleatum ss. polymorphum	1	0.87
uncultured *gamma proteobacterium*	1	0.81
Bacteroides pyogenes	1	0.89
Prevotella sp. oral clone GU027	1	0.9
Leptotrichia sp.	1	0.8
Bacteroides fragilis	1	0.87
Neisseria flava	1	0.94
Solobacterium moorei	1	0.92
Capnocytophaga gingivalis	1	0.83
uncultured *Haemophilus sp.*	1	0.95
Kingella oralis	1	0.93
Fusobacterium sp. oral clone CZ006	1	0.93
Veillonellaceae bacterium oral taxon 155	1	0.83
Granulicatella sp. oral clone ASCG05	1	0.85
Leptotrichia sp. oral clone DR011	1	0.89

Capnocytophaga ochracea	1	0.91
uncultured *Bacteroidetes bacterium*	1	0.92
Eubacterium saburreum	1	0.94
Micrococcus luteus	1	0.83
Catonella morbi	1	0.83
Streptococcus gordonii str. Challis substr. CH1	1	0.82
Actinomyces orihominis	1	0.95
Veillonella atypica	1	0.95
Haemophilus influenzae	1	0.9
Lactobacillus vaginalis	1	0.83
Neisseria sp. oral clone AP015	1	0.93
Capnocytophaga sp. AHN8725	1	0.91

Supplementary Table 4 | Bacterial diversity in cavities from samples CA06-1.6 (a) and CA05-4.6 (b). Data show the number of contigs >500 bp giving a significant hit in the NR database (score >100).

a

BLAST hit against NRdb	Class	Contigs >500 bp
Veillonella parvula	Clostridia	166
Streptococcus pneumoniae	Bacilli	59
Streptococcus mitis	Bacilli	35
Capnocytophaga ochracea	Flavobacteria	20
Prevotella ruminicola	Bacteroidia	15
Porphyromonas gingivalis	Bacteroidia	10
Streptococcus gordonii	Bacilli	9
Corynebacterium aurimucosum	Actinobacteria	7
Corynebacterium efficiens	Actinobacteria	7
Fusobacterium nucleatum	Fusobacterium	7
Bacteroides fragilis	Bacteroides	6
Alistipes shahii	Bacteroides	5
Corynebacterium diphtheriae	Actinobacteria	5
Leptotrichia buccalis	Fusobacteria	5
Xylanimonas cellulosilytica	Actinobacteria	5
Aggregatibacter aphrophilus	γ-proteobacteria	4
Bacteroides thetaiotaomicron	Bacteroides	4
Corynebacterium glutamicum	Actinobacterium	4
Slackia heliotrinireducens	Actinobacteria	4
Bacteroides fragilis	Bacteroides	3
Campylobacter concisus	ε-proteobacteria	3
Corynebacterium kroppenstedtii	Actinobacteria	3
Streptococcus oralis	Bacilli	3
Bacteriophage Dp-1	Virus	2
Bacteroides vulgatus	Bacteroides	2
Beutenbergia cavernae	Actinobacteria	2
Campylobacter hominis	ε-proteobacteria	2
Conexibacter woesei	Actinobacteria	2
Corynebacterium jeikeium	Actinobacteria	2
Corynebacterium urealyticum	Actinobacteria	2
Kribbella flavida	Actinobacteria	2
Micrococcus luteus	Actinobacteria	2
Mycobacterium gilvum	Actinobacteria	2
Prevotella intermedia	Bacteroidia	2
Rhodopseudomonas palustris	α-proteobacteria	2

b

BLAST hit against NRdb	Class	Contigs >500 bp
Veillonella parvula	Clostridia	135
Xylanimonas cellulosilytica	Actinobacteria	46
Sanguibacter keddieii	Actinobacteria	30
Kineococcus radiotolerans	Actinobacteria	20
Porphyromonas gingivalis	Bacteroidia	19
Beutenbergia cavernae	Actinobacteria	17
Treponema denticola	Spirochaetes	17
Brachybacterium faecium	Actinobacteria	16
Kytococcus sedentarius	Actinobacteria	14
Actinomyces naeslundii	Actinobacteria	11
Kocuria rhizophila	Actinobacteria	10
Catenulispora acidiphila	Actinobacteria	9
Geodermatophilus obscurus	Actinobacteria	9
Micrococcus luteus	Actinobacteria	9
Nocardioides sp.	Actinobacteria	9
Thermomonospora curvata	Actinobacteria	9
Rothia mucilaginosa	Actinobacteria	8
Neisseria meningitidis	β-proteobacteria	7
Streptomyces coelicolor	Actinobacteria	6
Actinosynnema mirum	Actinobacteria	5
Arthrobacter chlorophenolicus	Actinobacteria	5
Clavibacter michiganensis	Actinobacteria	5
Corynebacterium efficiens	Actinobacteria	5
Nakamurella multipartita	Actinobacteria	5
Streptococcus sanguinis	Bacilli	5
Streptomyces avermitilis	Actinobacteria	5
Streptomyces griseus	Actinobacteria	5
Bifidobacterium longum	Actinobacteria	4
Nocardia farcinica	Actinobacteria	4
Propionibacterium acnes	Actinobacteria	4
Actinomyces oris	Actinobacteria	3
Bifidobacterium adolescentis	Actinobacteria	3
Gordonia bronchialis	Actinobacteria	3
Saccharopolyspora erythraea	Actinobacteria	3
Streptococcus gordonii	Bacilli	3
Streptomyces scabiei	Actinobacteria	3
Alistipes shahii	Bacteroides	2
Arthrobacter sp.	Actinobacteria	2
Corynebacterium aurimucosum	Actinobacteria	2
Corynebacterium diphtheriae	Actinobacteria	2
Corynebacterium urealyticum	Actinobacteria	2
Eggerthella lenta	Actinobacteria	2
Leifsonia xyli	Actinobacteria	2
Leptotrichia buccalis	Fusobacteria	2
Mycobacterium avium	Actinobacteria	2
Rhodococcus opacus	Actinobacteria	2
Stackebrandtia nassauensis	Actinobacteria	2
Streptosporangium roseum	Actinobacteria	2
Thermobifida fusca	Actinobacteria	2