

Metagenome Report for HMP2

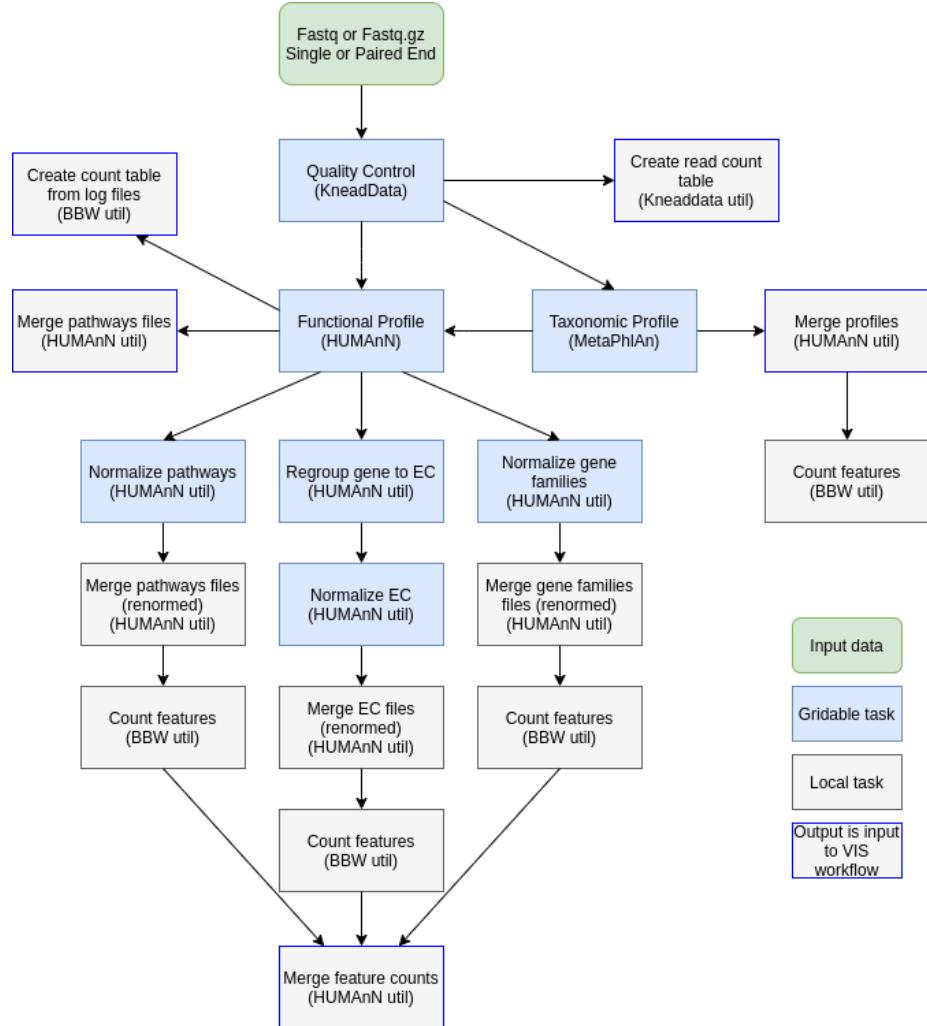
Author: LM

Date: 06/22/2021

Contents

Introduction	2
Quality Control	3
DNA Samples Quality Control	4
DNA Samples Tables of Filtered Reads	4
DNA Samples Plots of Filtered Reads	7
Taxonomic Profiling of Metagenomic Reads	9
Taxonomic Count Table	9
Alpha Diversity	10
Categorical variables	10
Continuous variables	12
Ordination	13
Species	13
Genera	20
Heatmaps	27
Species	27
Genera	29
Barplot	31
Species	31
Genera	45
Functional Profiling of Metagenomic Reads	59
Pathway Abundance	59
Data Processing Workflow Information	63
Software Versions	63
Tasks Run	63

Introduction



The data was run through the standard workflow for whole metagenome shotgun sequencing.

Quality Control

This report section contains information about the quality control processing for all 1298 paired-end fastq input files. These files were run through the [KneadData](#) QC pipeline. Reads were first trimmed then filtered against contaminate reference database hg38.

Data is organized by paired and orphan reads. When one read in a pair passes a filtering step and the other does not the surviving read is an orphan. The tables and plots are annotated as follows:

- raw : Untouched fastq reads.
- trim : Number of reads remaining after trimming bases with Phred score < 20. If the trimmed reads is < 50% of original length then it is removed altogether.
- hg38 : Number of reads after depleting against reference database hg38. The human genome database is used to remove reads originating from the host DNA.

DNA Samples Quality Control

DNA Samples Tables of Filtered Reads

DNA Paired end reads (partial table)

	Raw	Trim	hg38
CSM5FZ4M	17,662,709	14,203,545	14,191,802
CSM5MCUO	19,661,807	15,901,082	15,777,116
CSM5MCVL	15,712,972	10,621,253	10,614,698
CSM5MCVN	14,292,763	11,405,205	11,276,921
CSM5MCW6	11,519,145	9,105,913	9,067,660
CSM5MCWC	15,609,334	12,051,356	11,929,952
CSM5MCWE	16,869,140	13,095,559	12,793,095
CSM5MCWG	16,981,364	11,576,931	11,535,025
CSM5MCWQ	11,734,108	9,012,498	8,976,280
CSM5MCX3	18,755,959	15,035,352	13,416,035
CSM5MCXD	15,170,612	11,170,359	11,141,935
CSM5MCXH	15,408,027	12,289,380	12,264,273
CSM5MCXJ	13,903,250	11,145,332	11,132,334
CSM5MCXL	17,872,356	14,328,802	14,304,007
CSM5MCXN	14,743,546	11,874,108	11,866,008
CSM5MCXP	17,770,970	13,338,176	13,288,031
CSM5MCXR	13,150,550	9,880,262	9,809,632
CSM5MCXT	5,515,703	4,356,840	4,338,542
CSM5MCXV	15,963,867	10,695,779	10,671,042
CSM5MCY2	10,039,709	7,958,521	7,874,726

The table is too large to include the full table in this document. A partial table is shown which includes only 20 rows. Please see the data file for the full table: [qc_counts_pairs_table.tsv](#)

DNA Orphan reads (partial table)

	Trim orphan1	Trim orphan2	hg38 orphan1	hg38 orphan2
CSM5FZ4M	873,653	488,500	872,537	488,008
CSM5MCUO	732,340	524,751	727,316	520,450
CSM5MCVL	1,336,295	819,275	1,335,590	818,724
CSM5MCVN	625,572	355,717	619,482	351,357
CSM5MCW6	514,004	278,111	512,226	276,787
CSM5MCWC	999,170	363,909	992,209	360,137
CSM5MCWE	942,825	442,595	929,545	431,872
CSM5MCWG	1,145,471	819,127	1,142,536	815,953
CSM5MCWQ	854,455	300,054	851,872	298,860
CSM5MCX3	708,193	464,420	620,124	412,830
CSM5MCXD	576,588	1,008,378	575,270	1,005,967
CSM5MCXH	636,185	483,598	634,197	482,561
CSM5MCXJ	542,991	441,013	541,815	440,426
CSM5MCXL	812,060	433,203	810,721	432,372
CSM5MCXN	657,092	407,285	656,475	406,950
CSM5MCXP	796,785	1,062,469	793,633	1,058,624
CSM5MCXR	515,808	378,342	512,267	375,743
CSM5MCXT	240,127	152,895	239,277	152,134
CSM5MCXV	1,420,812	765,028	1,418,037	763,128
CSM5MCY2	313,064	310,260	309,924	307,068

The table is too large to include the full table in this document. A partial table is shown which includes only 20 rows. Please see the data file for the full table: [qc_counts_orphans_table.tsv](#)

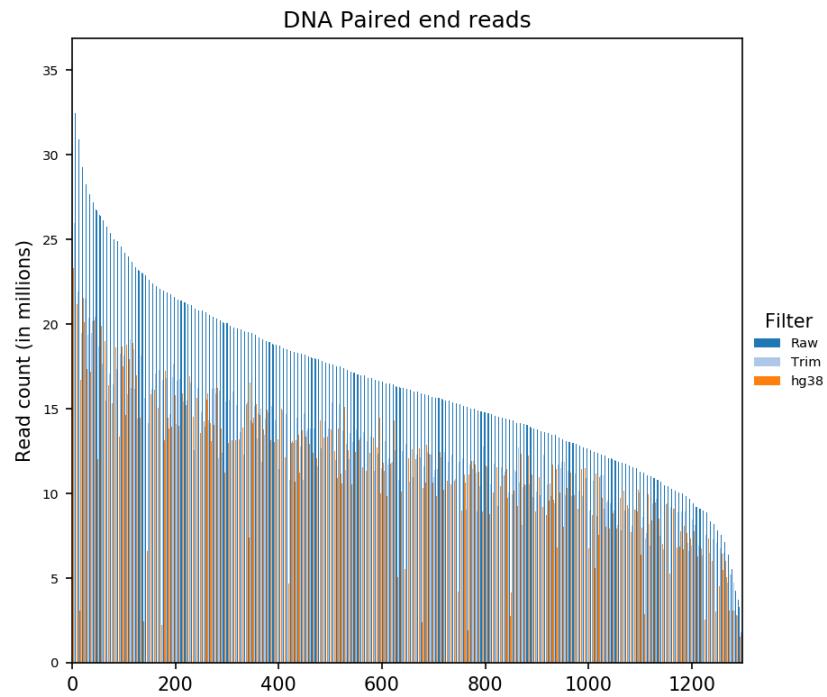
DNA microbial read proportion (partial table)

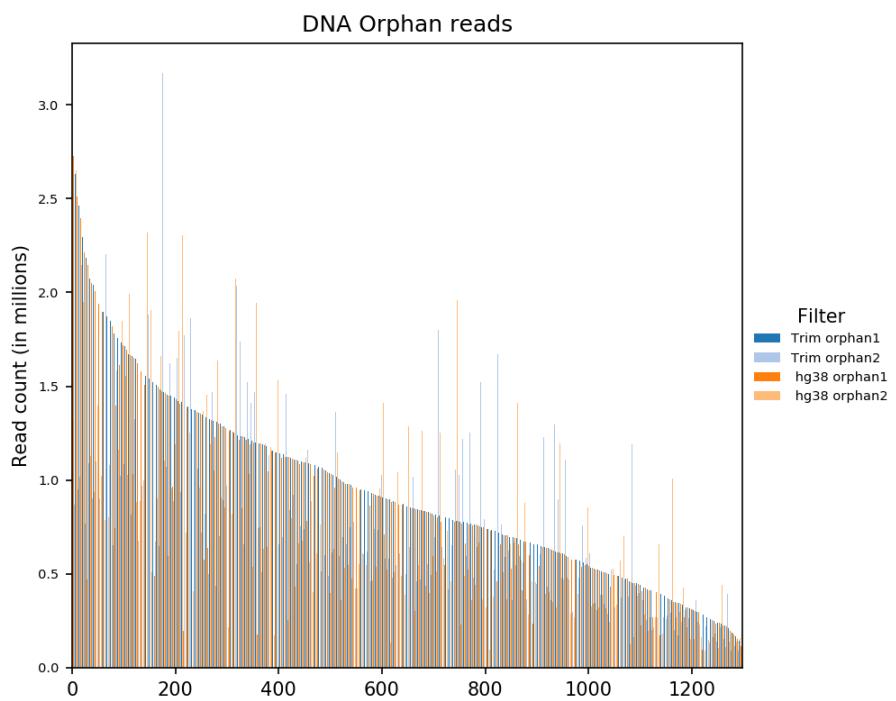
	hg38 / Trim	hg38 / Raw
CSM5FZ4M	0.99917	0.84202
CSM5MCUO	0.99235	0.83426
CSM5MCVL	0.99941	0.74411
CSM5MCVN	0.98896	0.82311
CSM5MCW6	0.99588	0.82149
CSM5MCWC	0.99019	0.80772
CSM5MCWE	0.97758	0.79904
CSM5MCWG	0.99655	0.73703
CSM5MCWQ	0.99609	0.81406
CSM5MCX3	0.89352	0.74421
CSM5MCXD	0.99757	0.78664
CSM5MCXH	0.99797	0.83224
CSM5MCXJ	0.99883	0.83605
CSM5MCXL	0.99830	0.83514
CSM5MCXN	0.99932	0.84090
CSM5MCXP	0.99637	0.79996
CSM5MCXR	0.99299	0.77981
CSM5MCXT	0.99589	0.82213
CSM5MCXV	0.99778	0.73682
CSM5MCY2	0.98968	0.81524

Proportion of reads remaining after removing host reads relative to the number of: i) quality-trimmed reads, and ii) raw unfiltered reads.

The table is too large to include the full table in this document. A partial table is shown which includes only 20 rows. Please see the data file for the full table: [microbial_counts_table.tsv](#)

DNA Samples Plots of Filtered Reads





Taxonomic Profiling of Metagenomic Reads

This report section contains information about the taxonomy for all DNA samples. These samples were run through [MetaPhlAn](#).

Taxonomic abundances are passed through a basic filter requiring each species or genus to have at least 0.01 % abundance in at least 10 % of all samples.

A total of 171 species and 16 genera were identified. After basic filtering 72 species and 15 genera remained.

Taxonomic Count Table

Total taxa per sample (partial table)

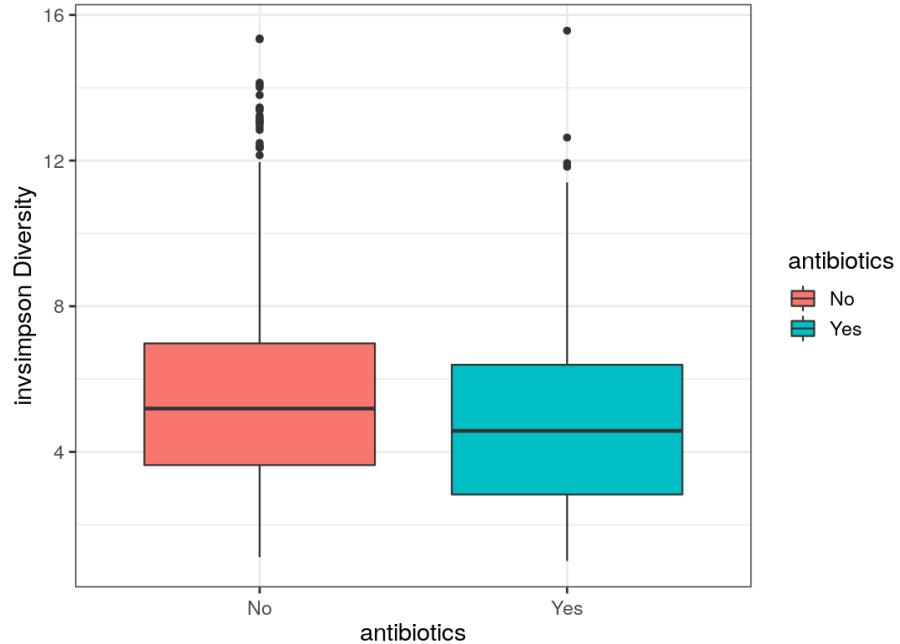
Species	Species filtered	Genera	Genera filtered
CSM5FZ3N_P	23	8	8
CSM5FZ3R_P	35	10	10
CSM5FZ3T_P	13	5	5
CSM5FZ3V_P	33	12	12
CSM5FZ3X_P	28	11	11
CSM5FZ3Z_P	18	8	8
CSM5FZ42_P	24	9	9
CSM5FZ44_P	22	9	9
CSM5FZ46_P	13	6	6
CSM5FZ4A_P	30	11	11
CSM5FZ4C_P	23	8	8
CSM5FZ4E_P	37	9	9
CSM5FZ4G_P	39	10	10
CSM5FZ4K_P	38	10	10
CSM5FZ4M	29	10	10
CSM5LLGB_P	40	14	14
CSM5MCTZ_P	49	10	10
CSM5MCU4_P	20	7	7
CSM5MCU8_P	46	12	12
CSM5MCUA_P	35	12	12

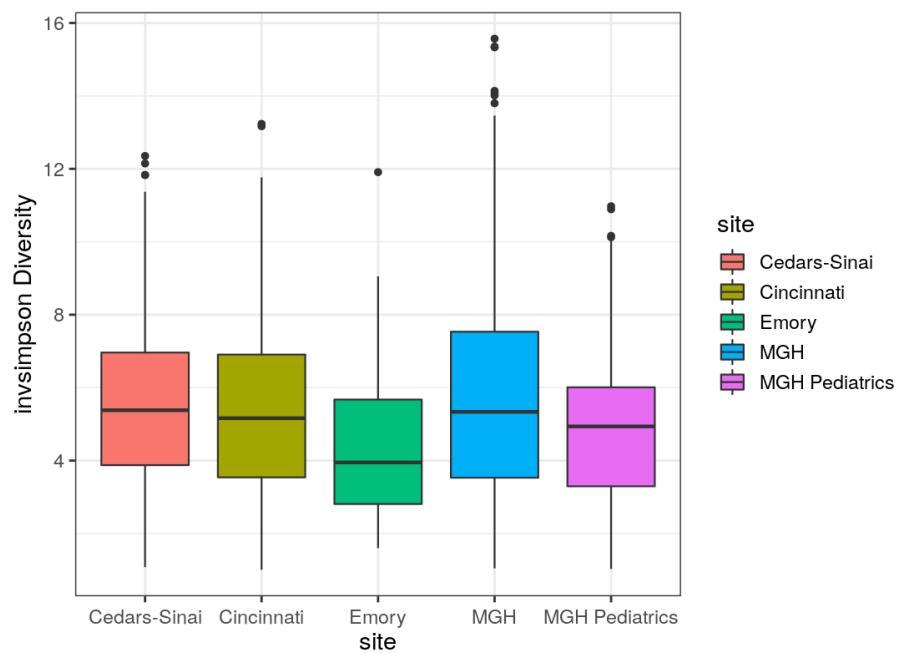
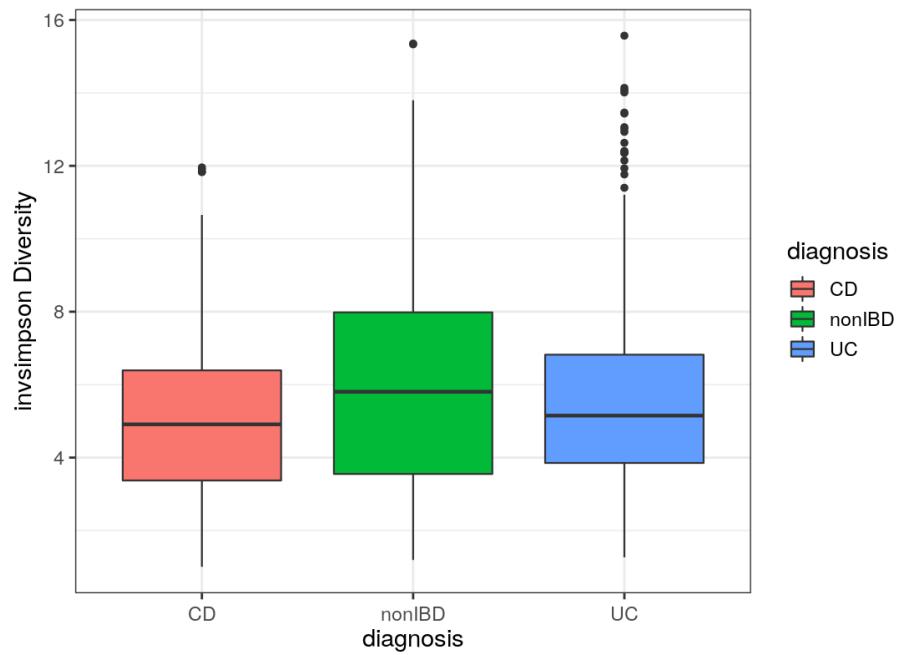
The table is too large to include the full table in this document. A partial table is shown which includes only 20 rows. Please see the data file for the full table: [taxa_counts_table.tsv](#)

Alpha Diversity

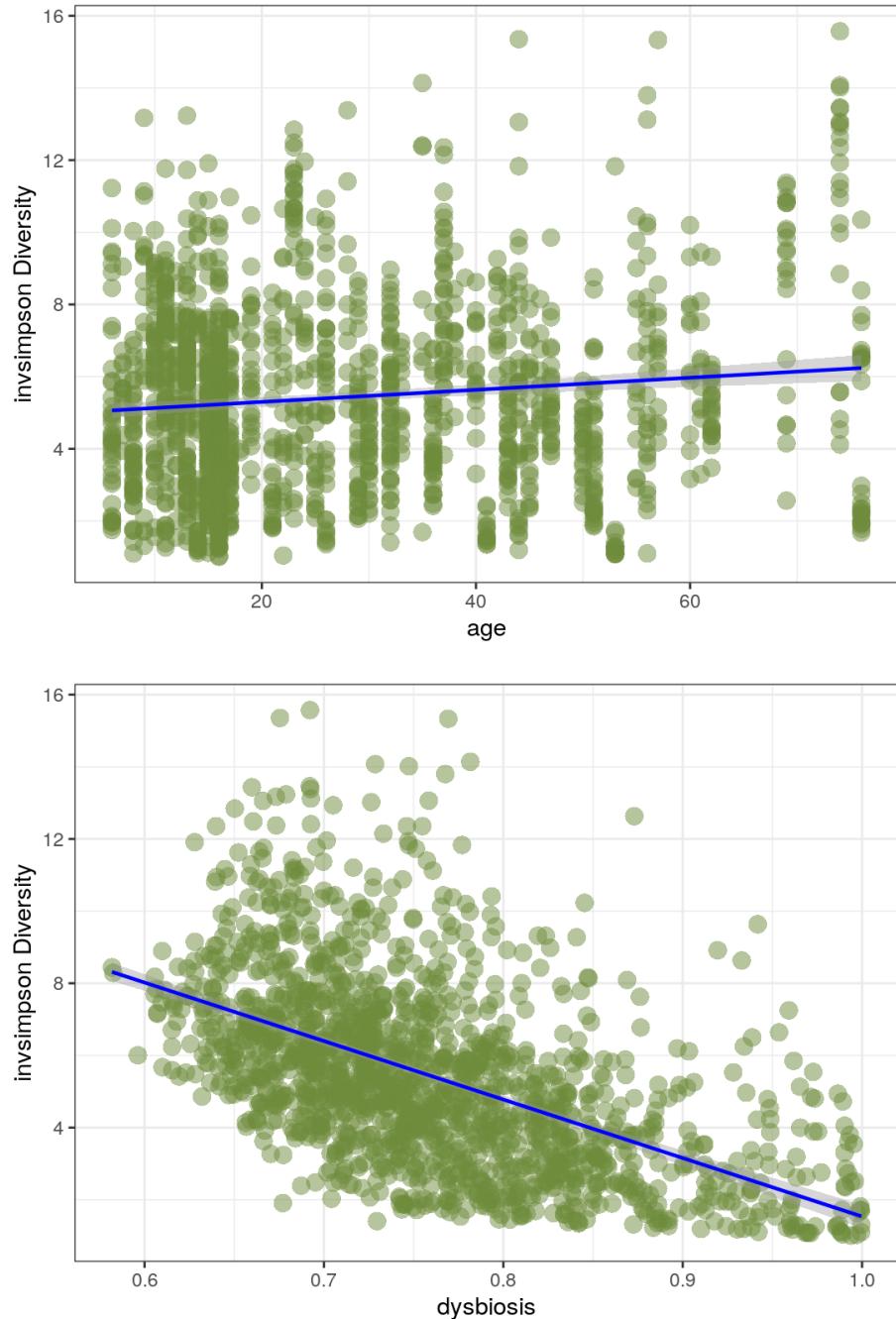
Alpha diversity, computed with the Inverse Simpson Index, were computed on un-filtered normalized data at the species level. Boxplots were generated for categorical/discrete variables while scatterplots were generated for continuous variables.

Categorical variables



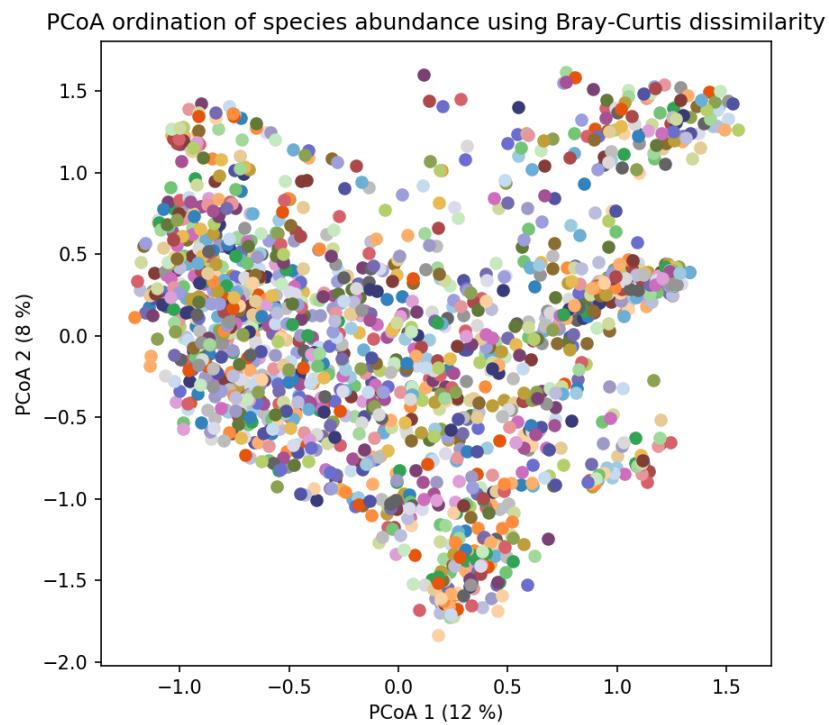


Continuous variables



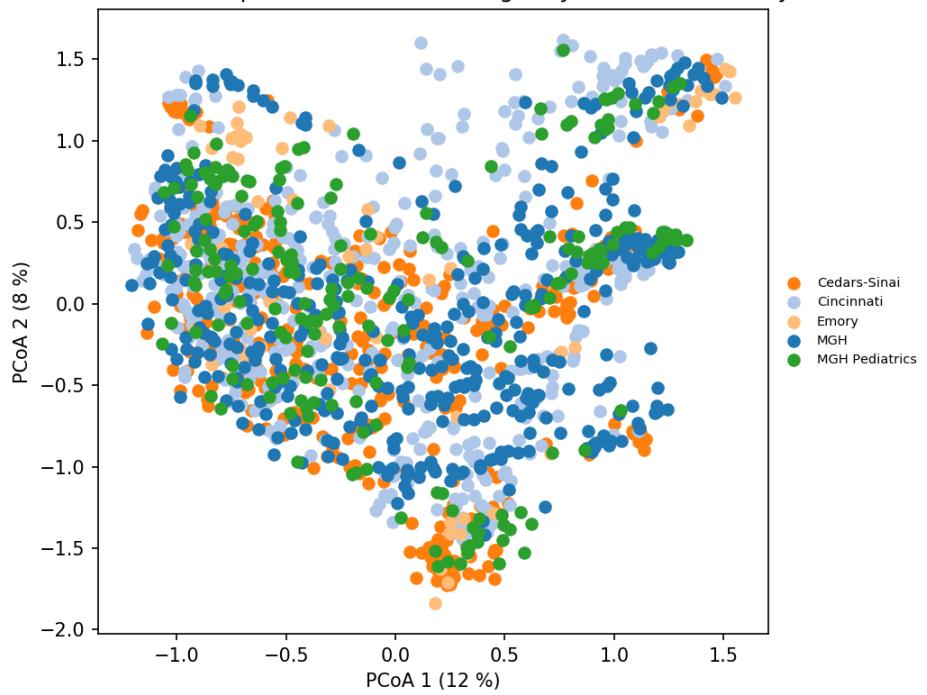
Ordination

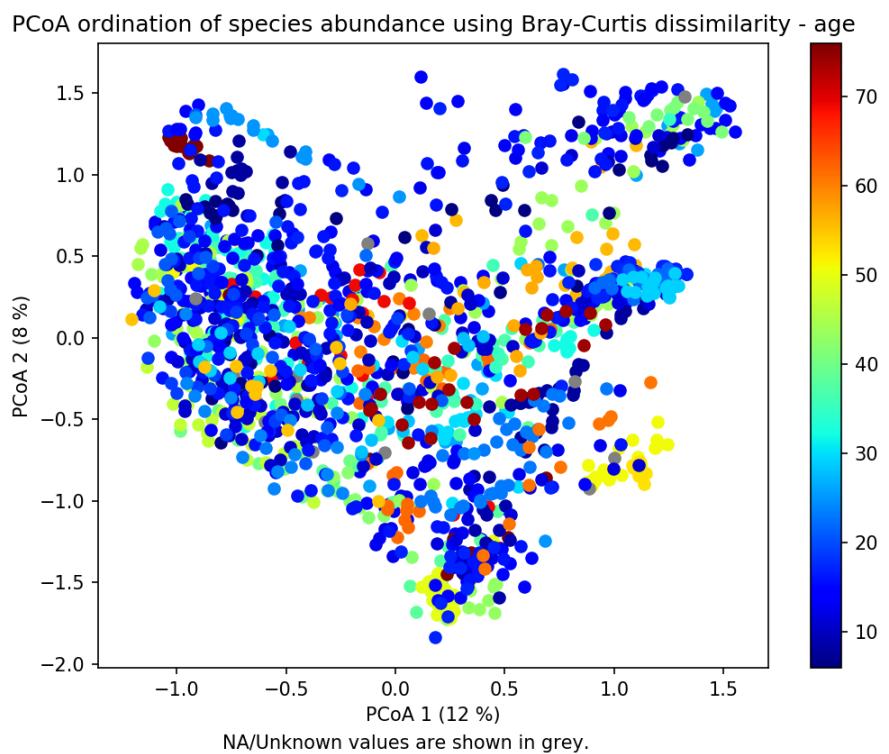
Species



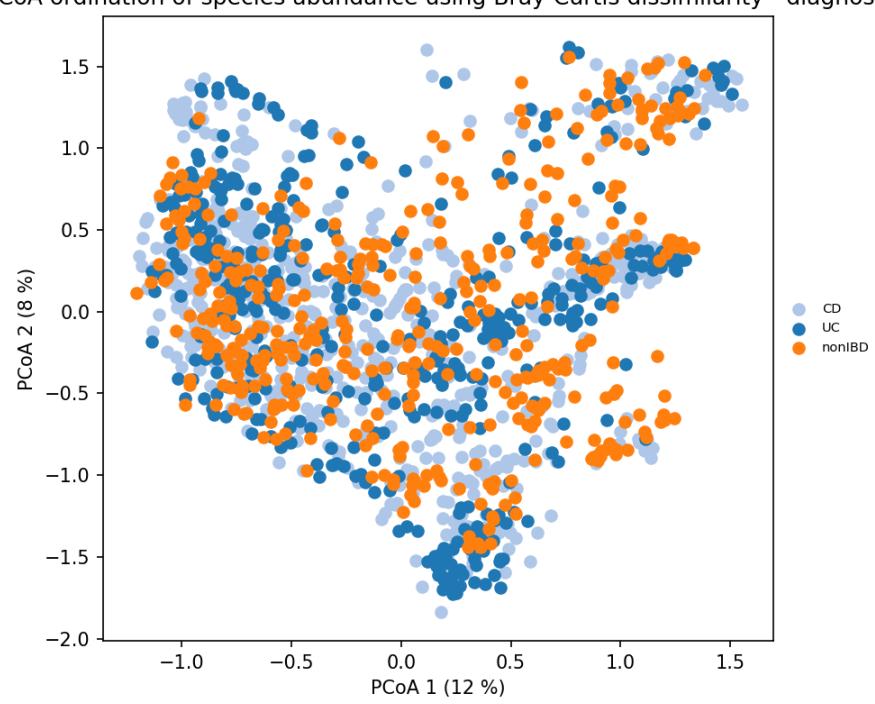
Principal coordinate analysis of variance among samples, based on Bray-Curtis dissimilarities between species profiles of samples. Numbers in parenthesis on each axis represent the amount of variance explained by that axis.

PCoA ordination of species abundance using Bray-Curtis dissimilarity - site

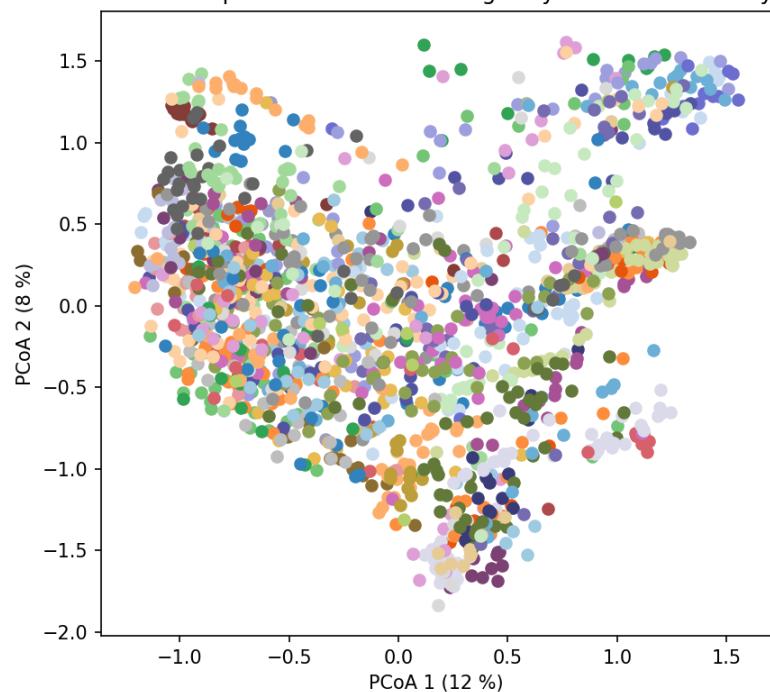




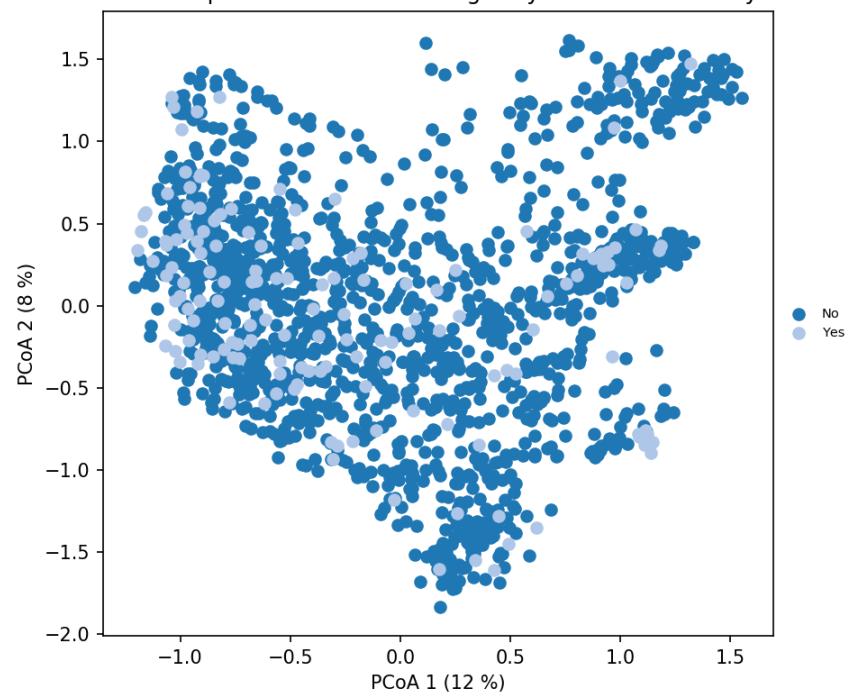
PCoA ordination of species abundance using Bray-Curtis dissimilarity - diagnosis



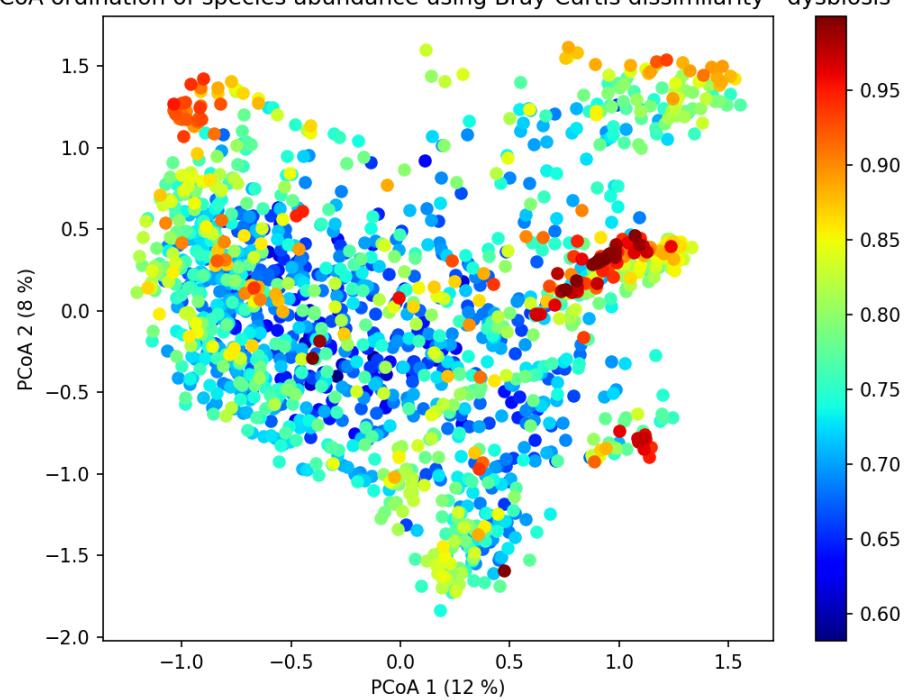
PCoA ordination of species abundance using Bray-Curtis dissimilarity - subject



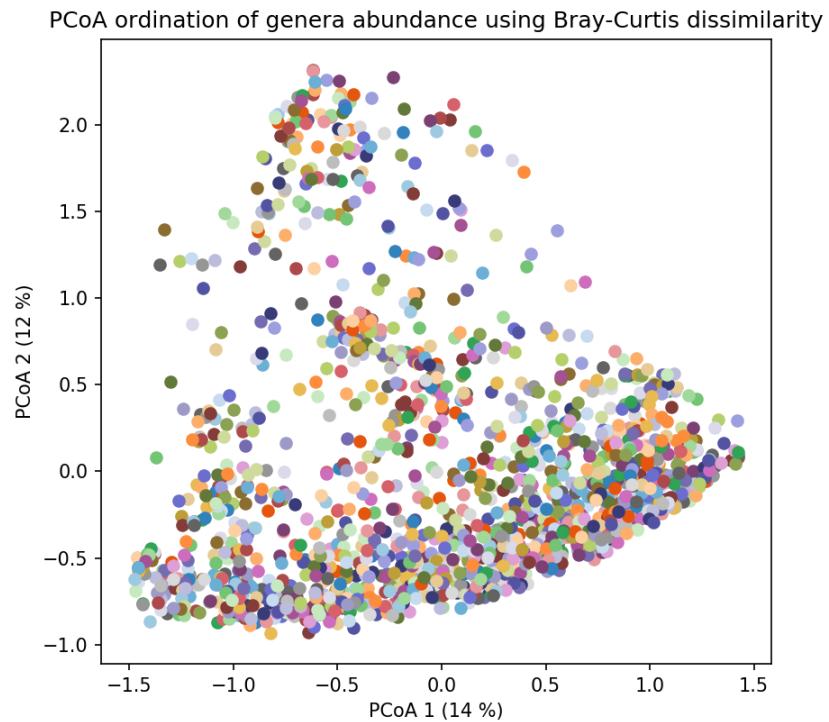
PCoA ordination of species abundance using Bray-Curtis dissimilarity - antibiotics



PCoA ordination of species abundance using Bray-Curtis dissimilarity - dysbiosis

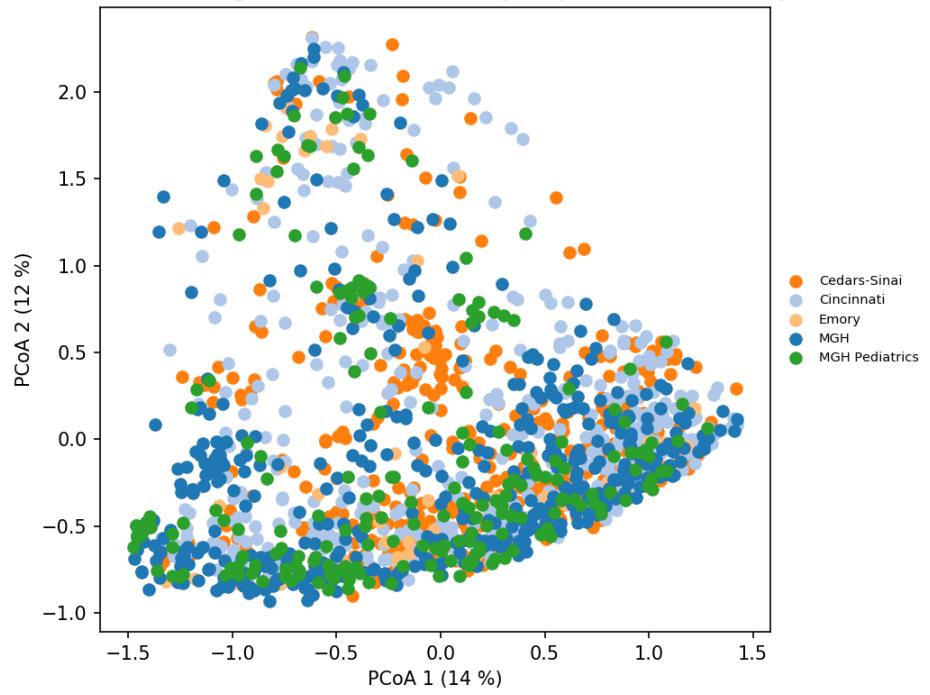


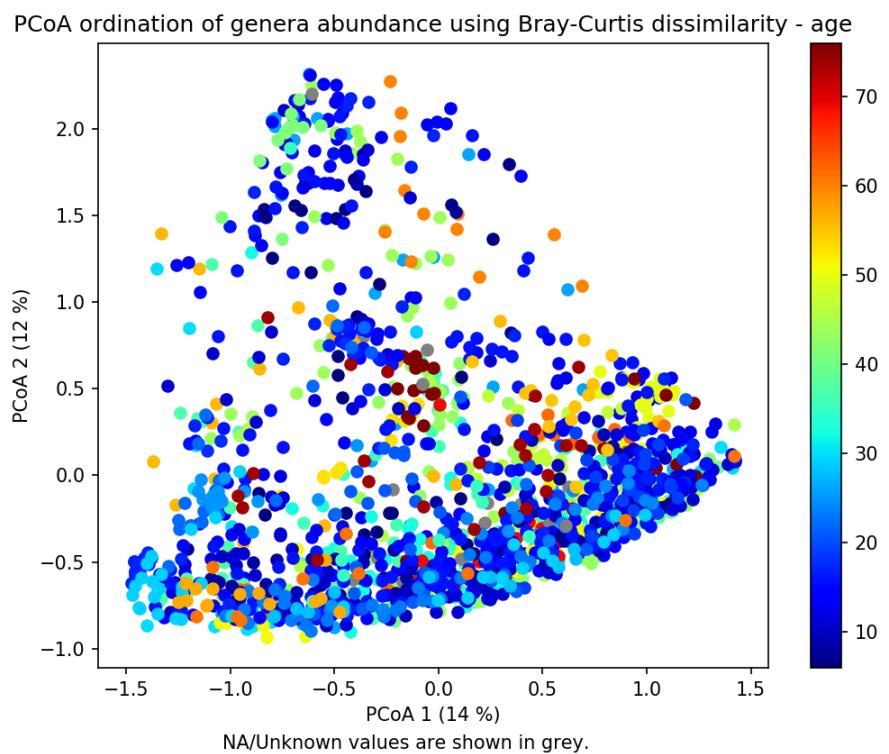
Genera



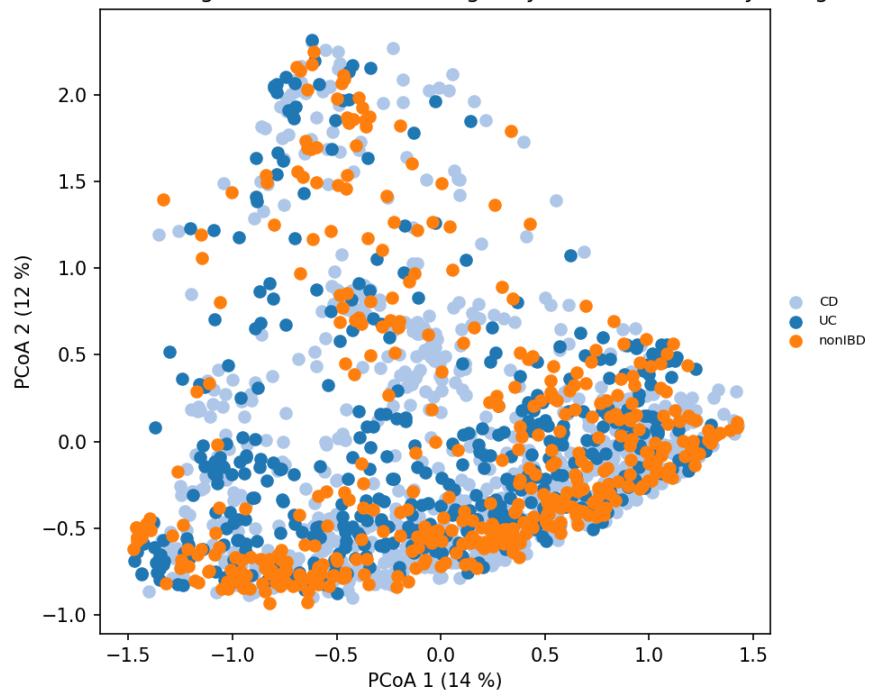
Principal coordinate analysis of variance among samples, based on Bray-Curtis dissimilarities between genera profiles of samples. Numbers in parenthesis on each axis represent the amount of variance explained by that axis.

PCoA ordination of genera abundance using Bray-Curtis dissimilarity - site

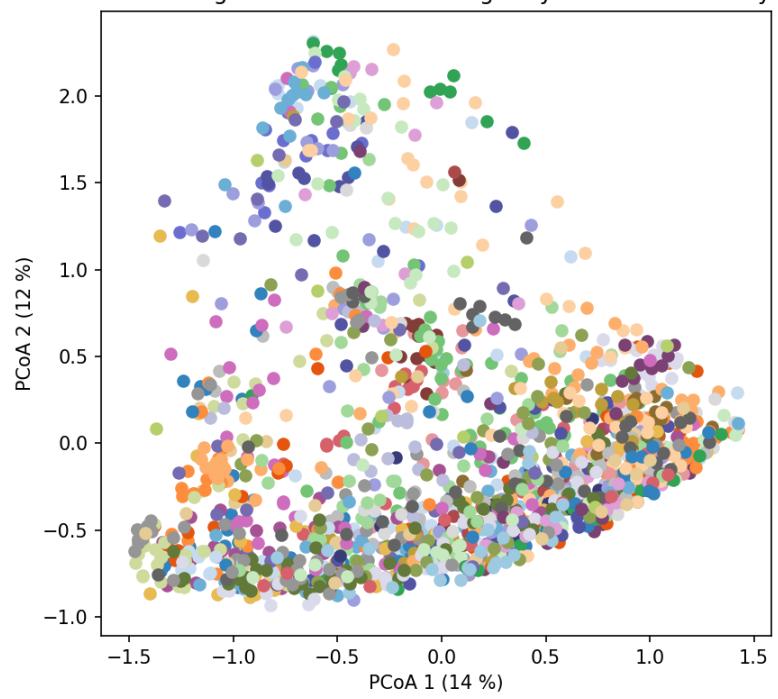




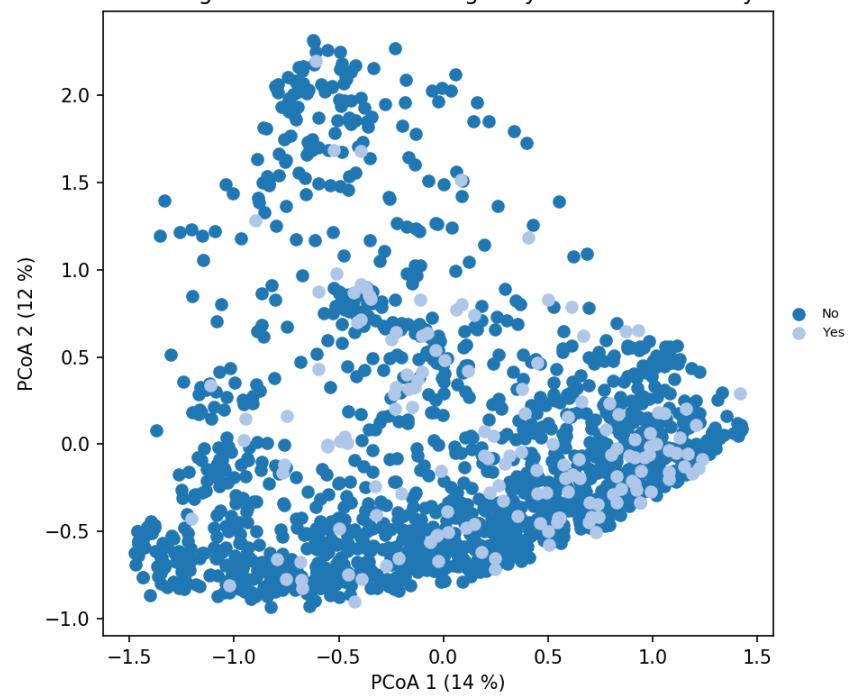
PCoA ordination of genera abundance using Bray-Curtis dissimilarity - diagnosis



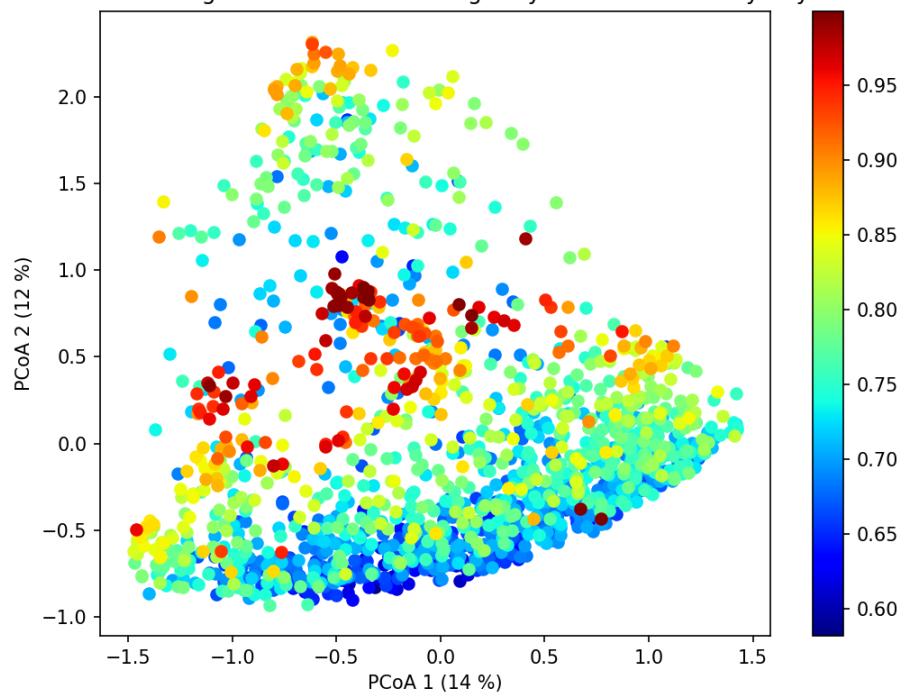
PCoA ordination of genera abundance using Bray-Curtis dissimilarity - subject



PCoA ordination of genera abundance using Bray-Curtis dissimilarity - antibiotics



PCoA ordination of genera abundance using Bray-Curtis dissimilarity - dysbiosis

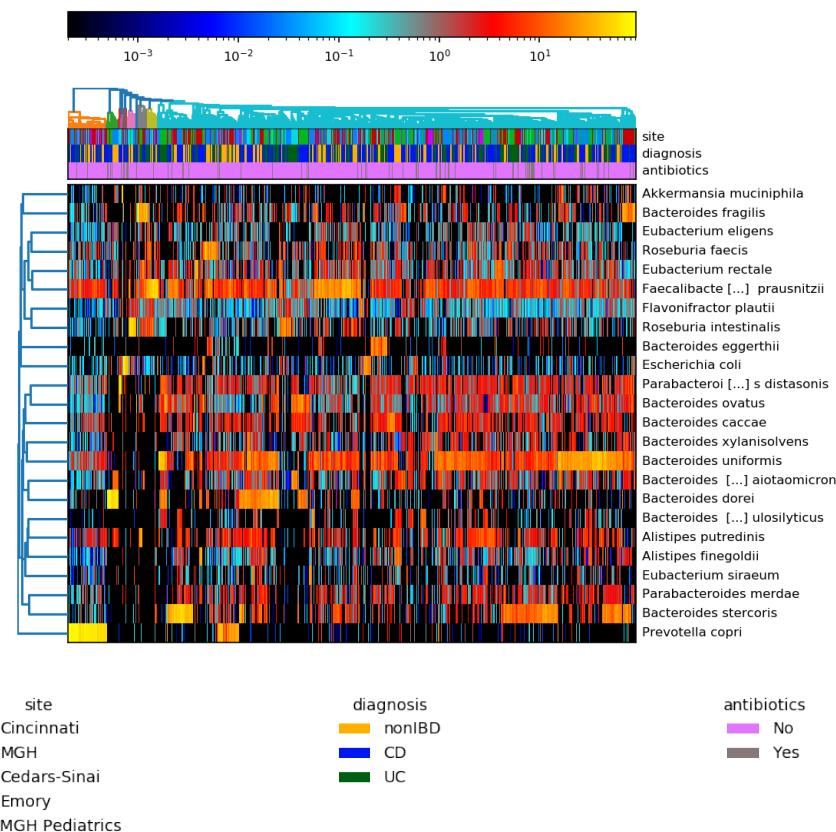


Heatmaps

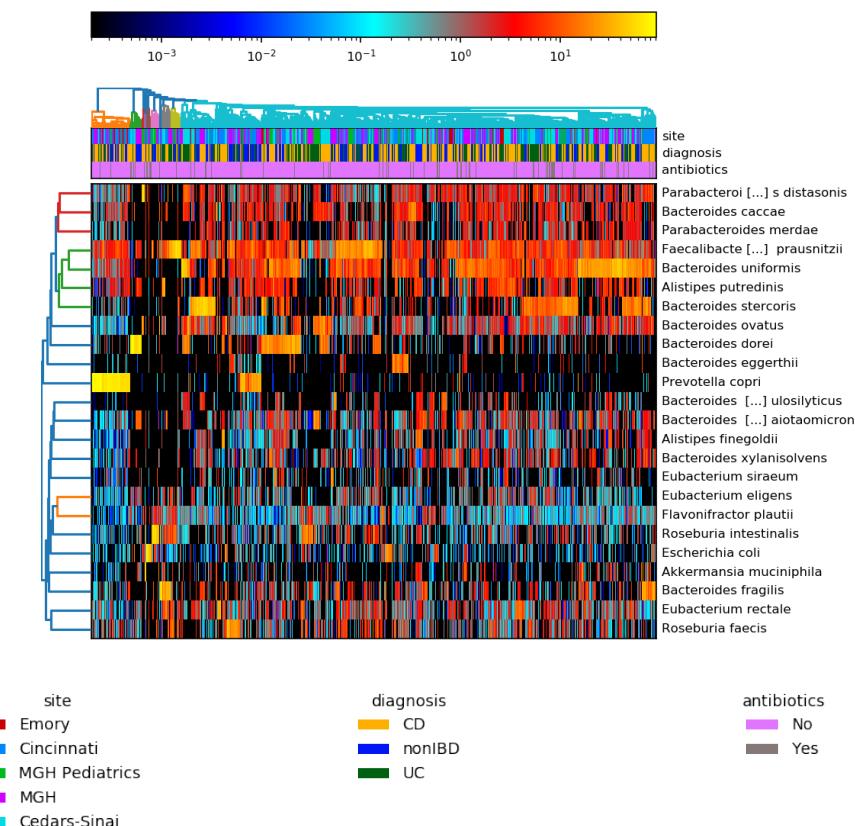
Hierarchical clustering of samples and species and genera, using top 25 species and genera with highest mean relative abundance among samples. The ‘average linkage’ clustering on the Euclidean distance metric was used to cluster samples. The species and genera dendrogram is based on pairwise (Spearman and Bray-Curtis) correlation between taxa. Samples are columns and species and genera are rows. The heatmaps were generated with [Hclust2](#).

Species

Top 25 species by average abundance (Spearman)

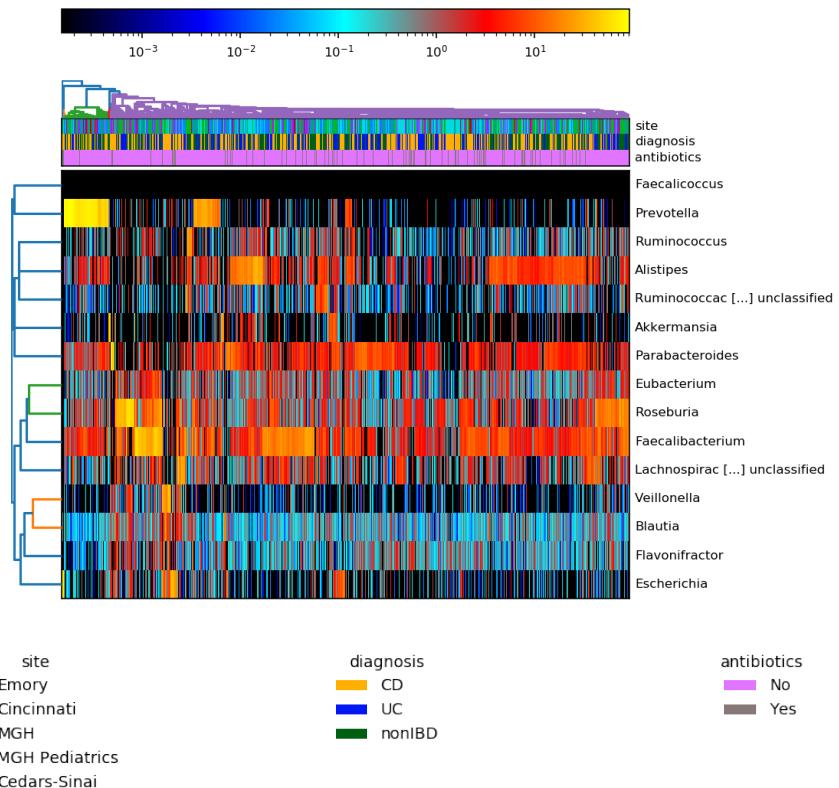


Top 25 species by average abundance (Bray-Curtis)

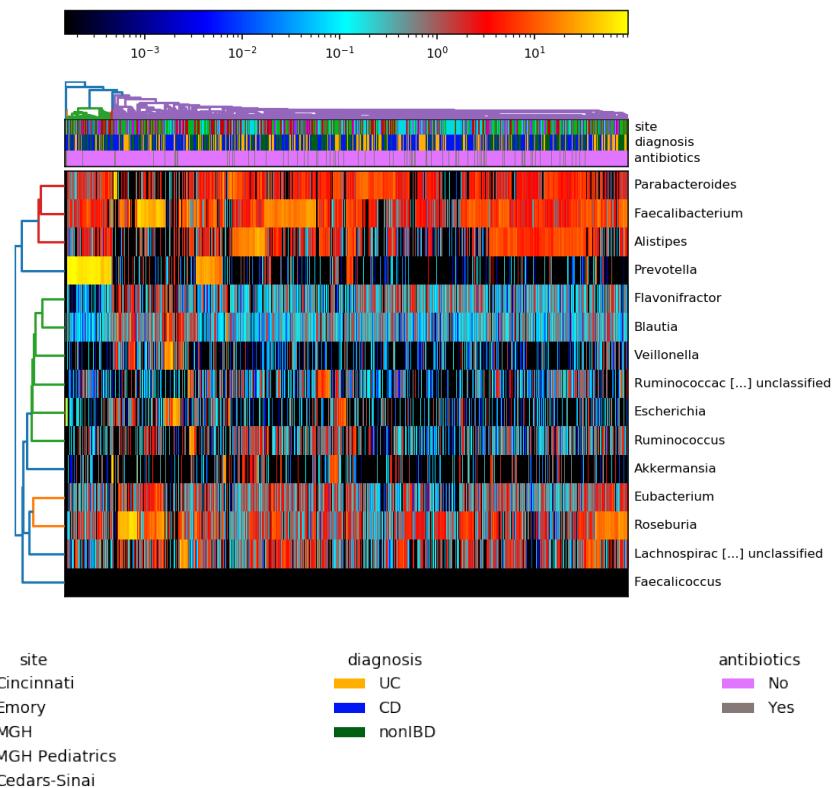


Genera

Top 25 genera by average abundance (Spearman)

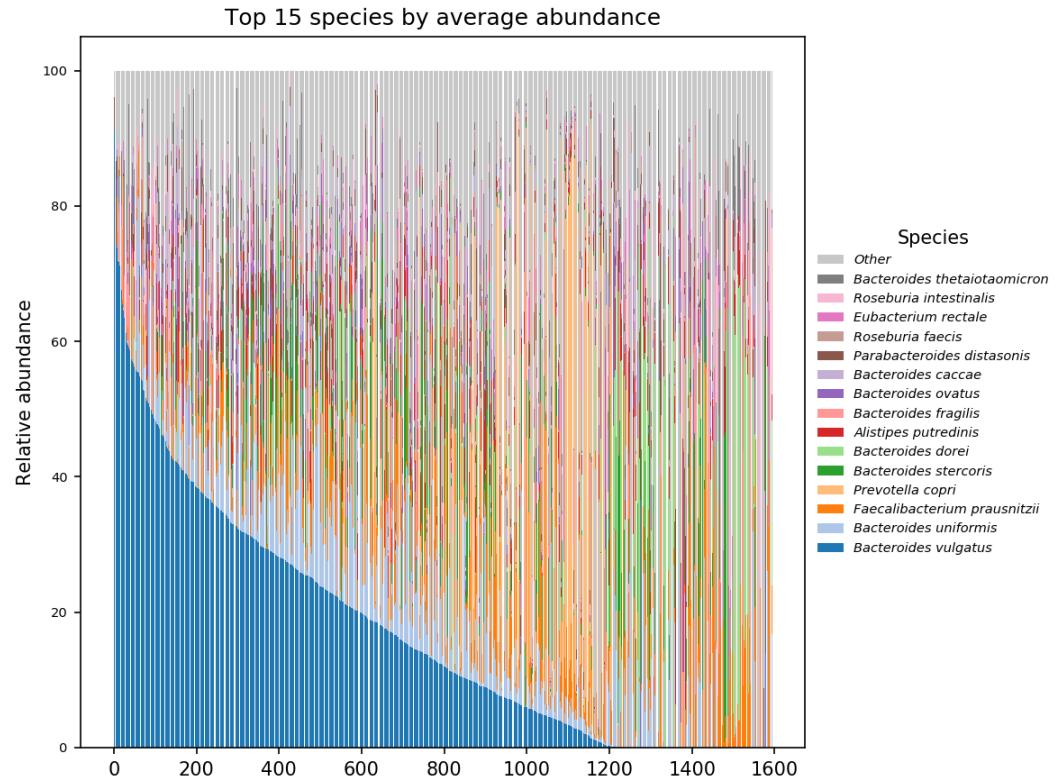


Top 25 genera by average abundance (Bray-Curtis)



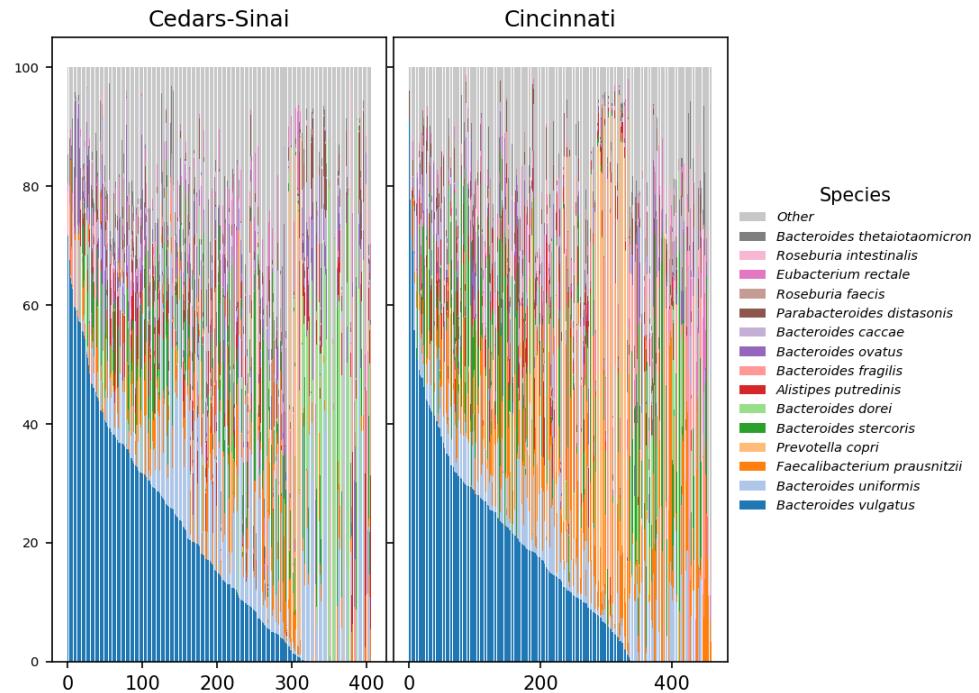
Barplot

Species

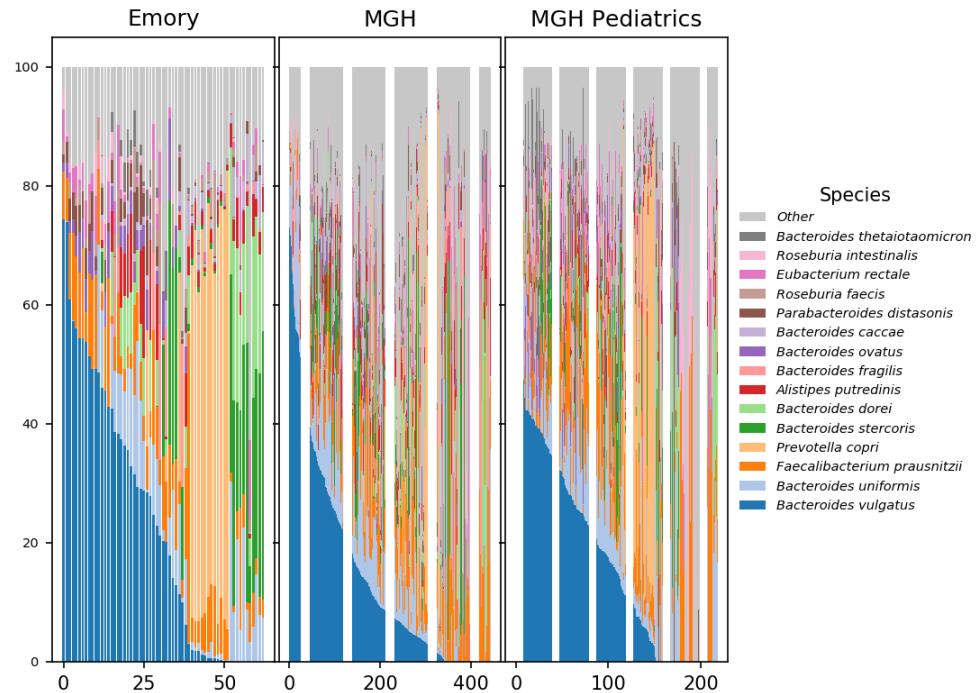


Stacked barplot of 15 most abundant species among samples. Samples in the plot were sorted on the species with the highest mean abundances among samples, in decreasing order.

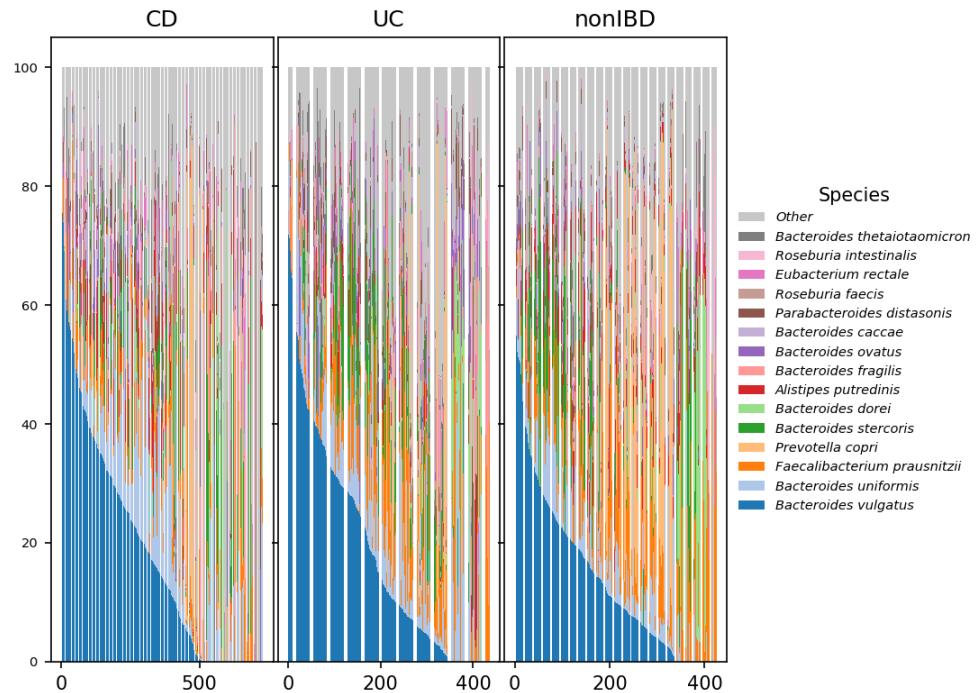
Top 15 species by average abundance - site Cedars-Sinai to Cincinnati



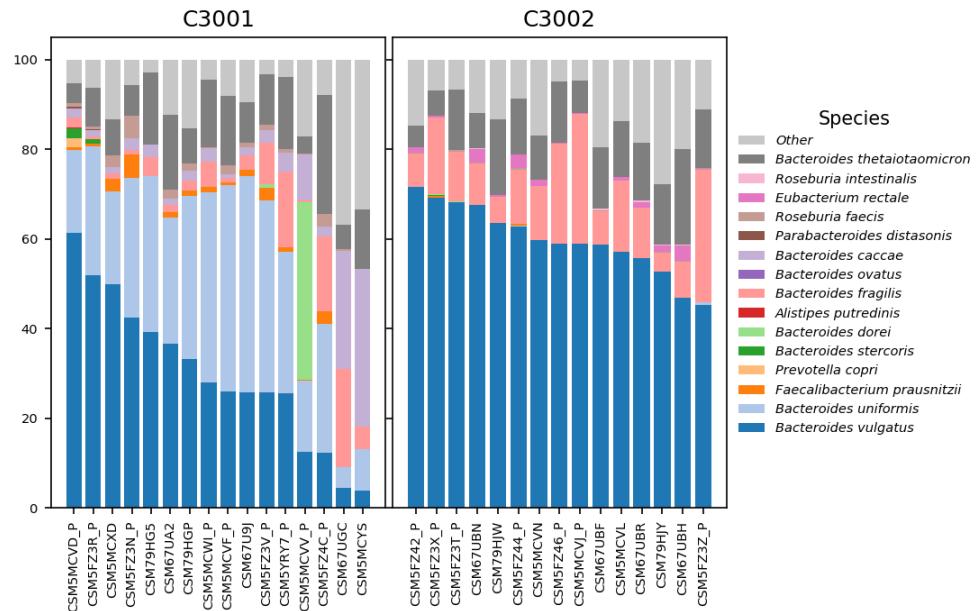
Top 15 species by average abundance - site Emory to MGH Pediatrics



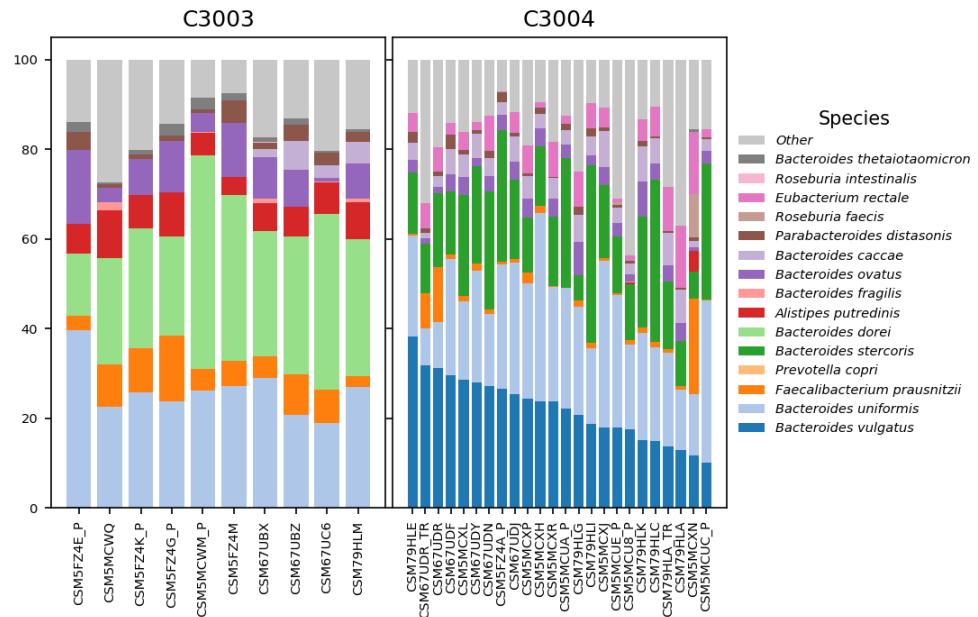
Top 15 species by average abundance - diagnosis CD to nonIBD



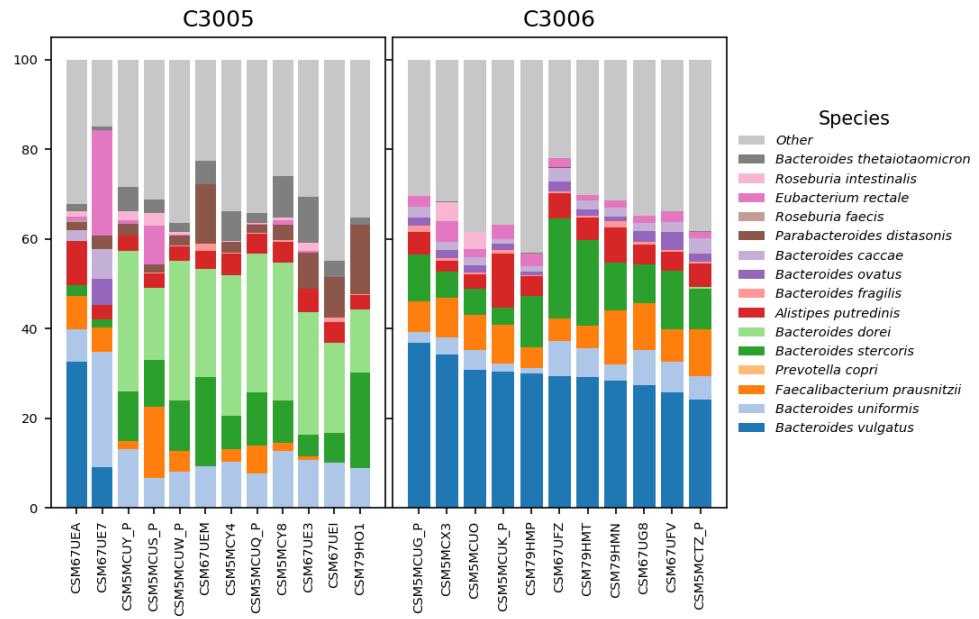
Top 15 species by average abundance - subject C3001 to C3002



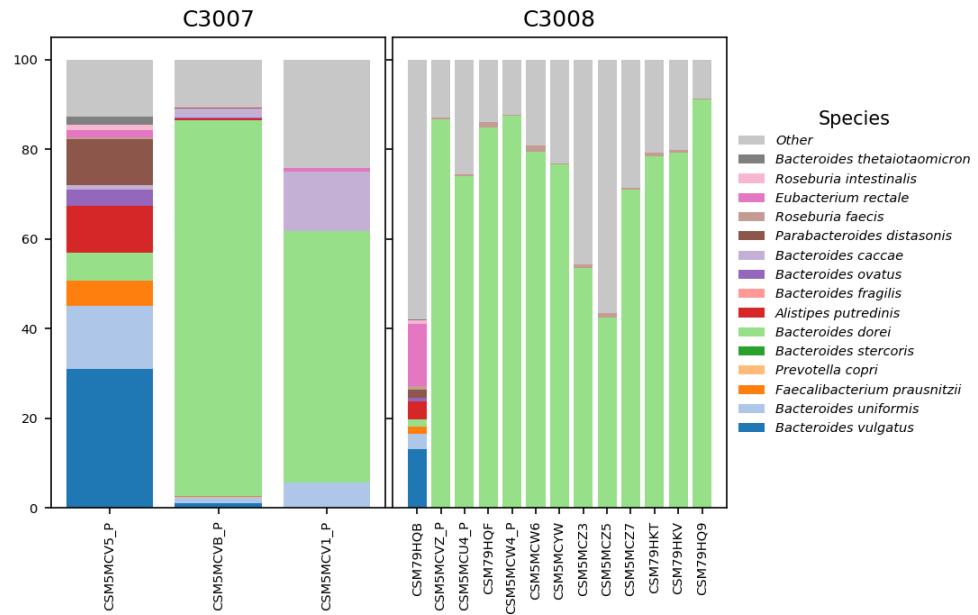
Top 15 species by average abundance - subject C3003 to C3004



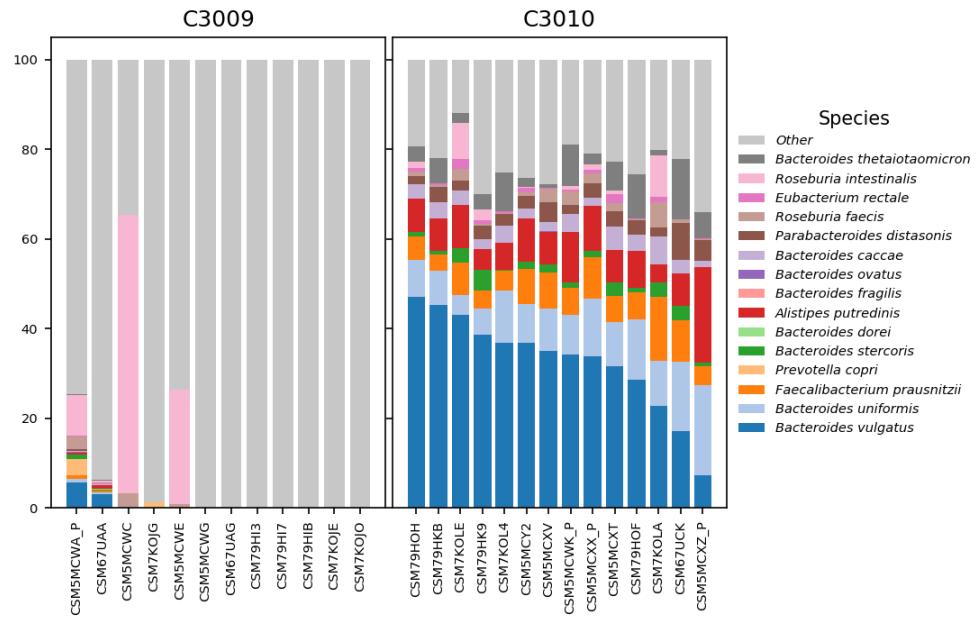
Top 15 species by average abundance - subject C3005 to C3006



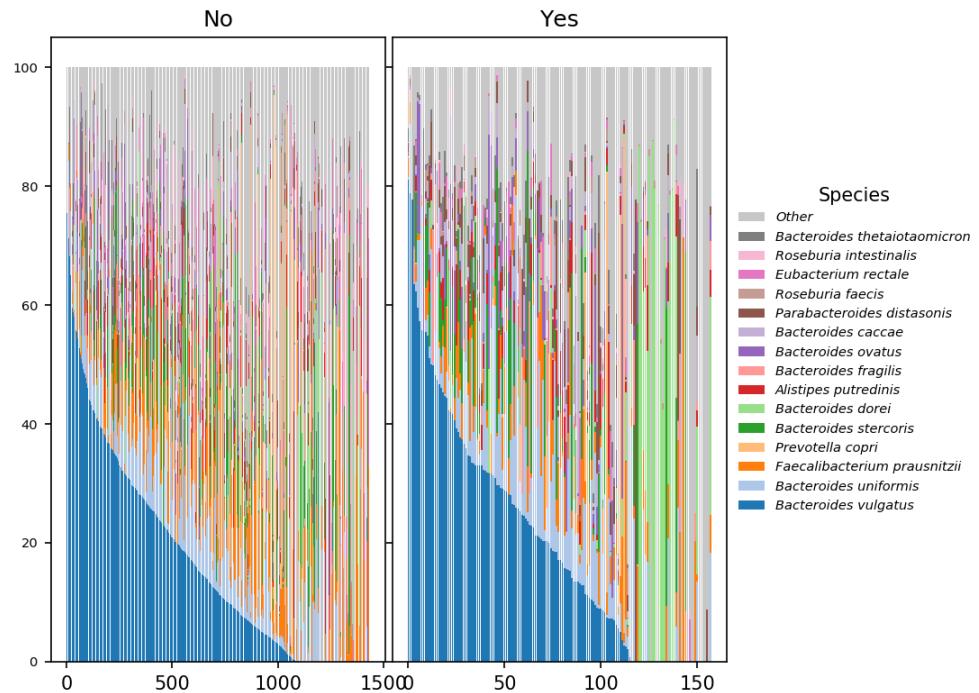
Top 15 species by average abundance - subject C3007 to C3008

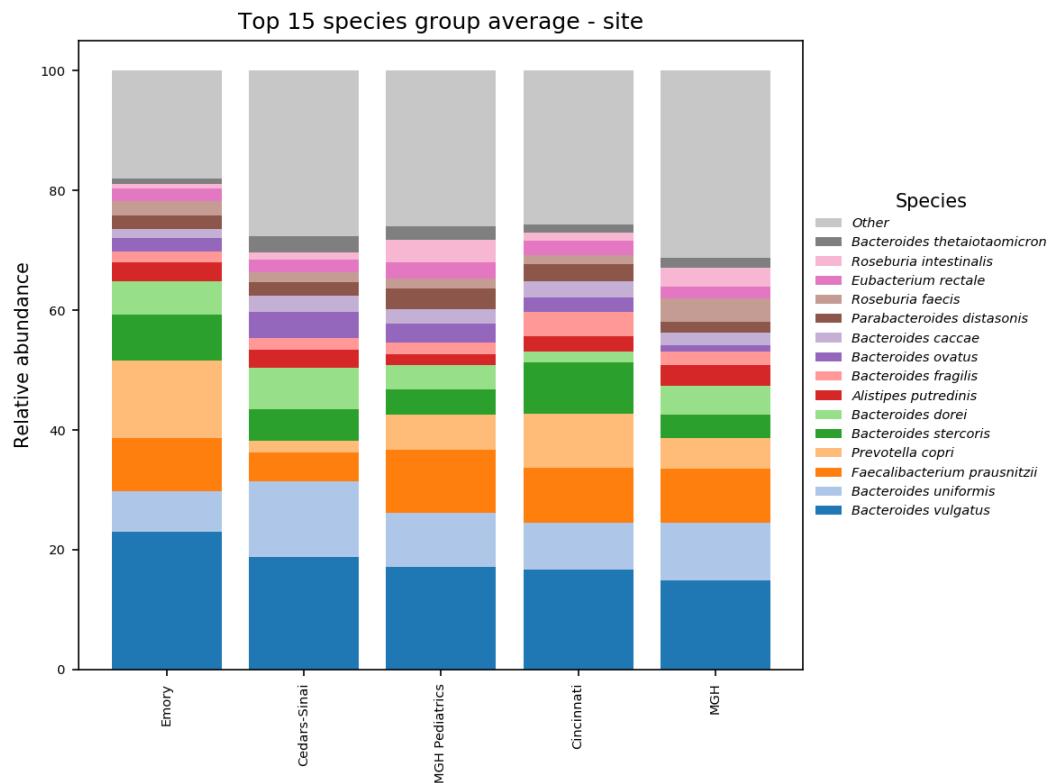


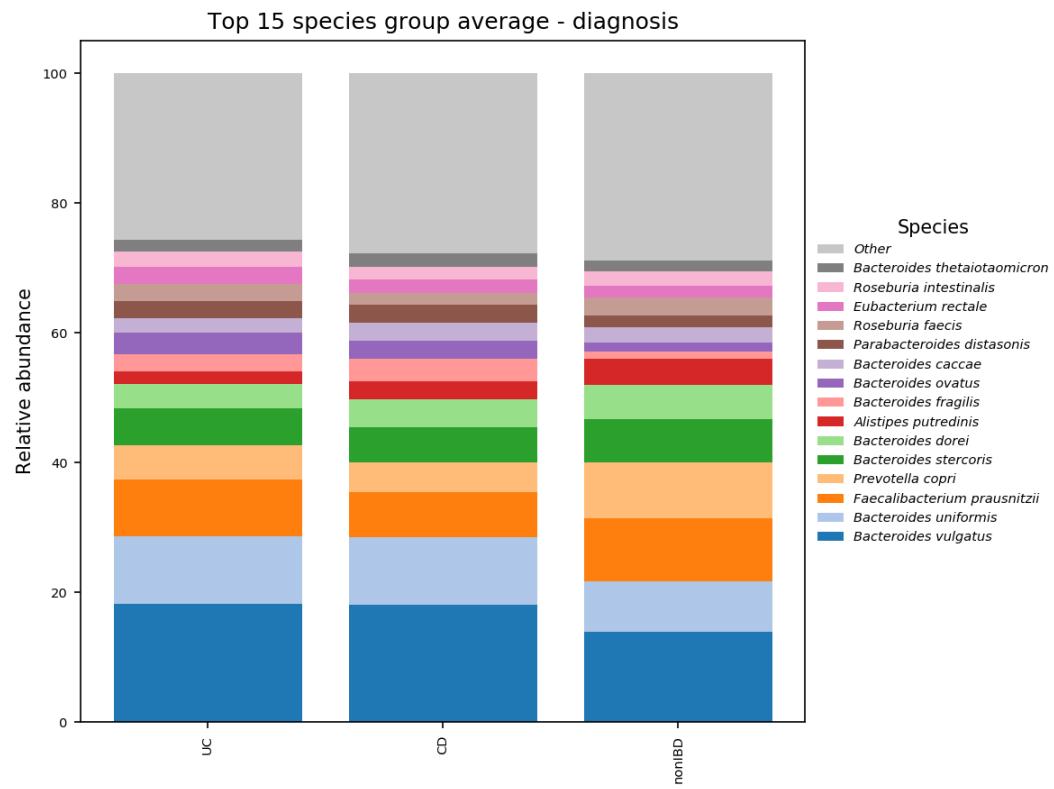
Top 15 species by average abundance - subject C3009 to C3010

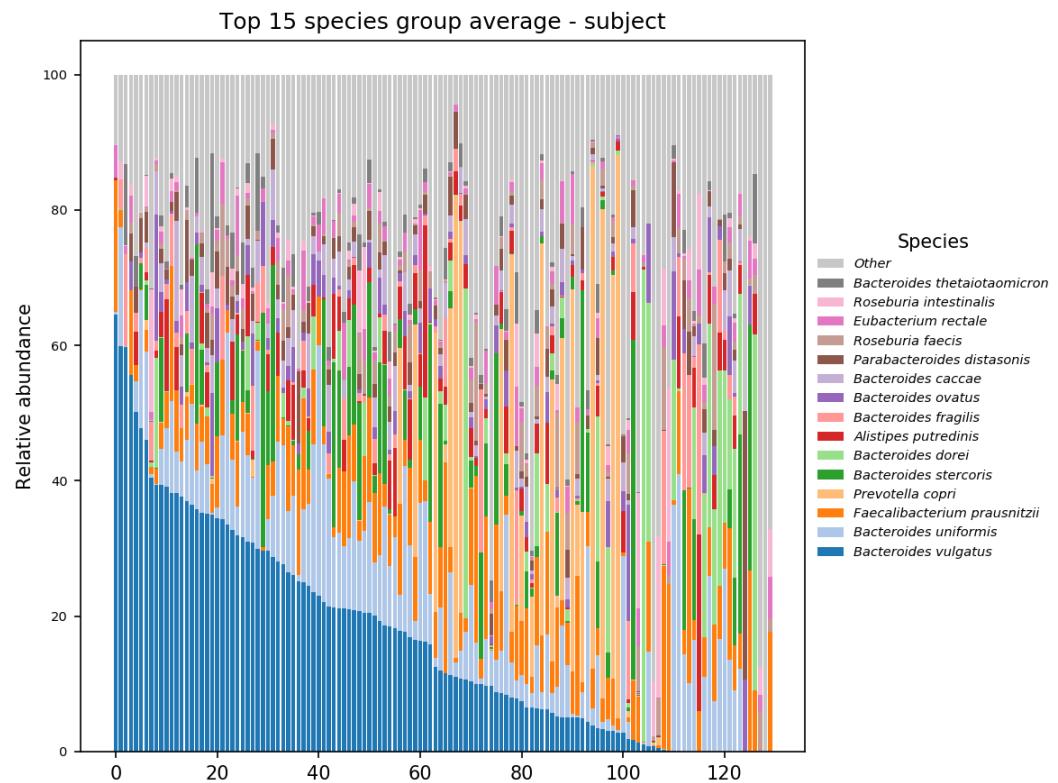


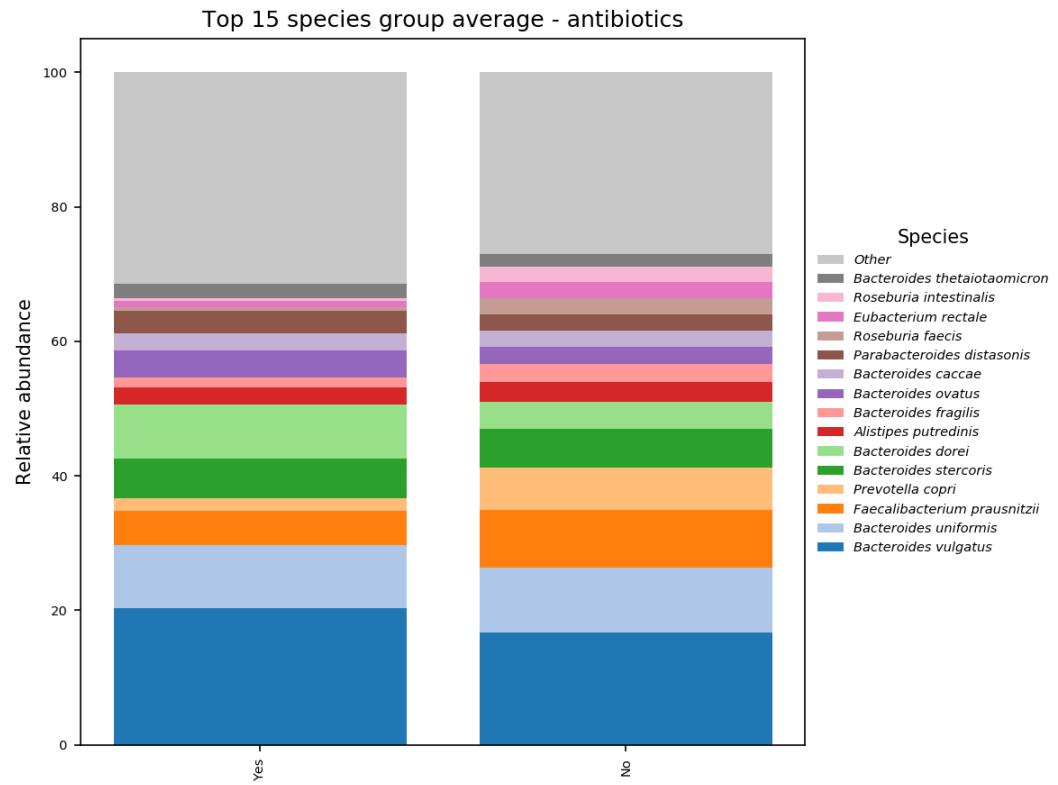
Top 15 species by average abundance - antibiotics





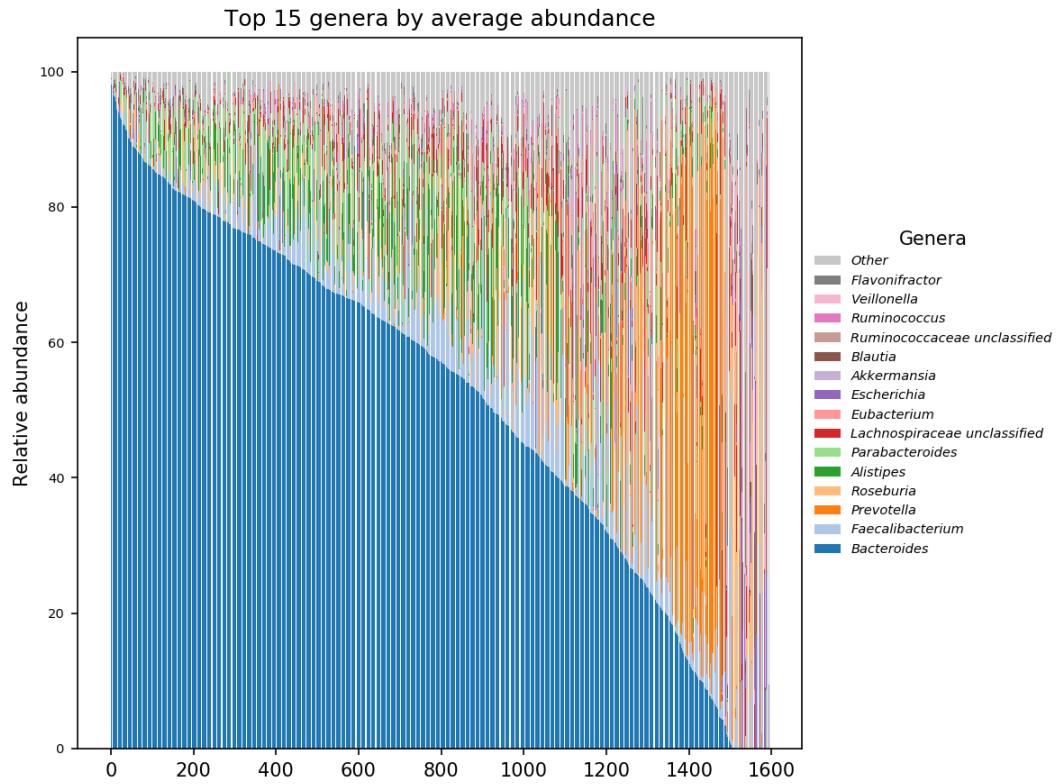






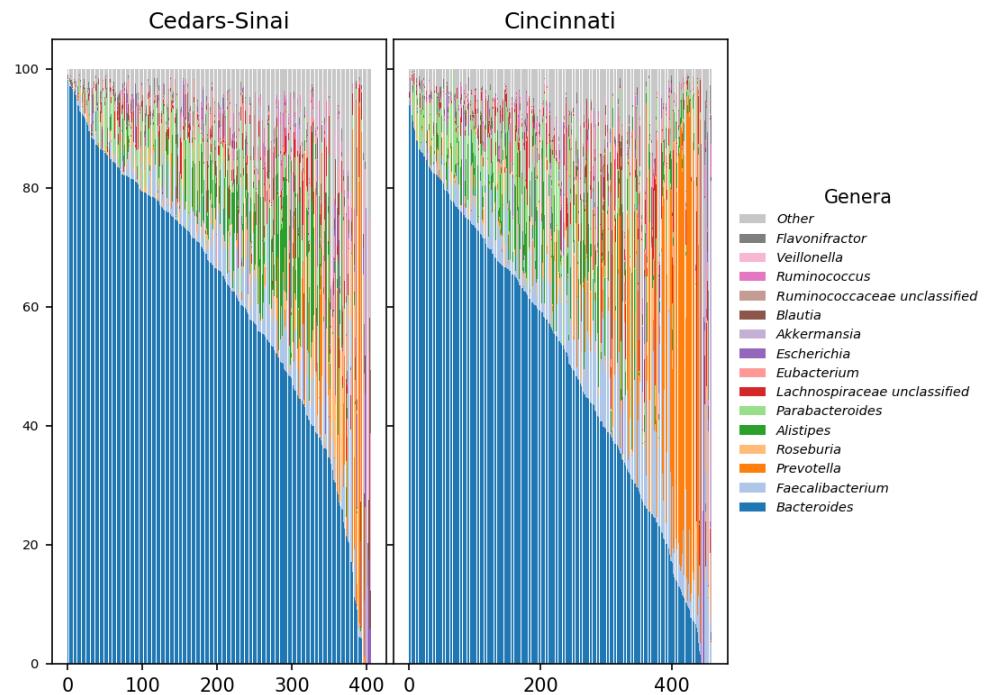
Stacked barplot of species abundance grouped by metadata and species average abundance grouped by metadata. For cases where a metadata variable has many levels only the first 5 grouped barplots will be shown.

Genera

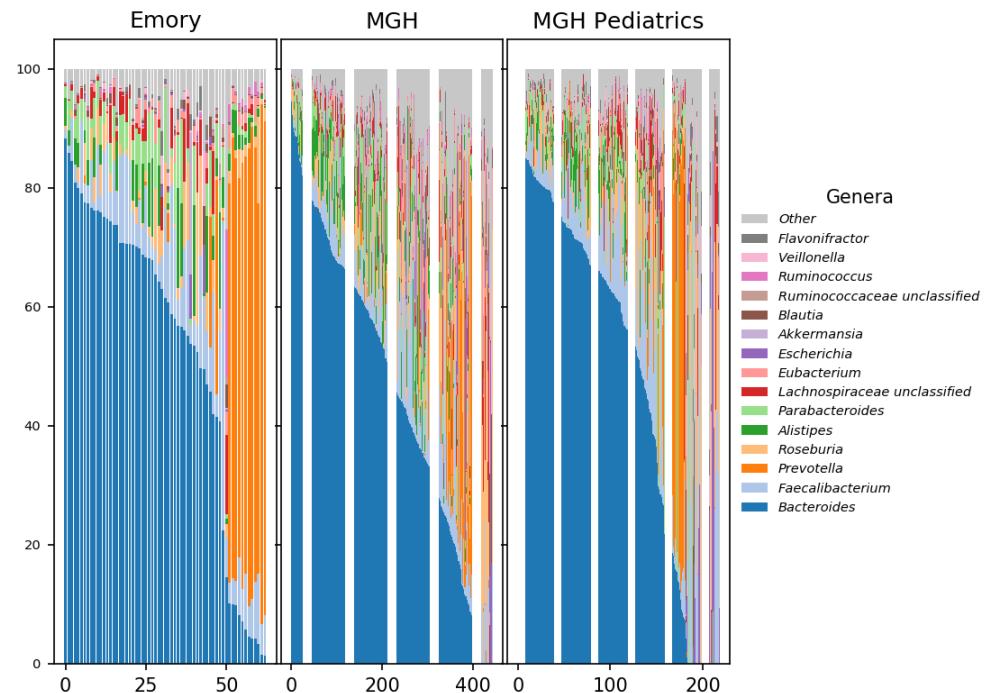


Stacked barplot of 15 most abundant genera among samples. Samples in the plot were sorted on the genera with the highest mean abundances among samples, in decreasing order.

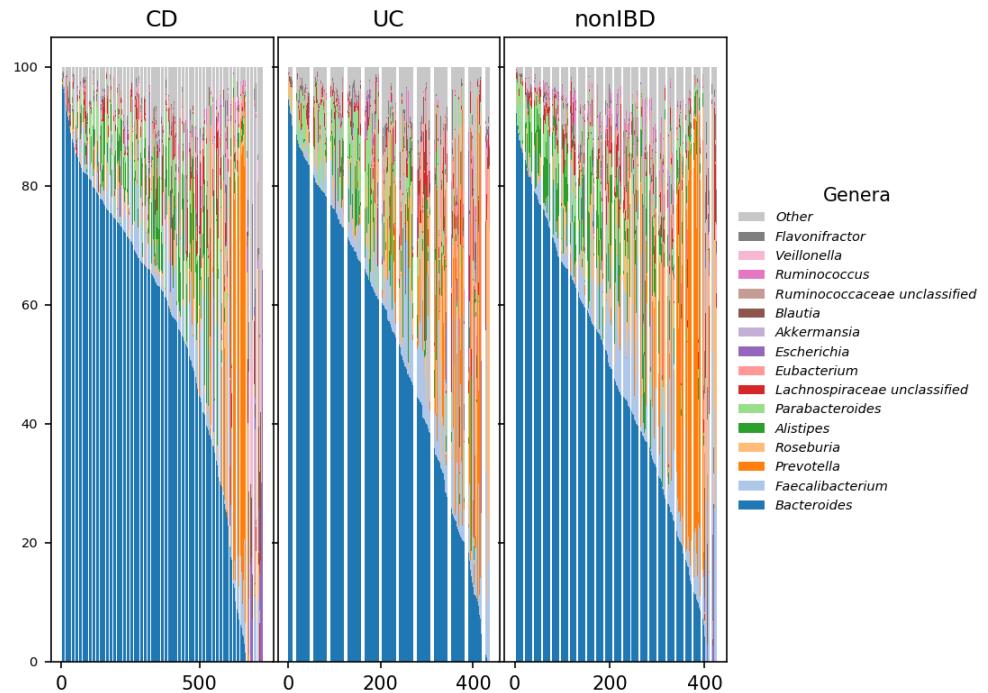
Top 15 genera by average abundance - site Cedars-Sinai to Cincinnati



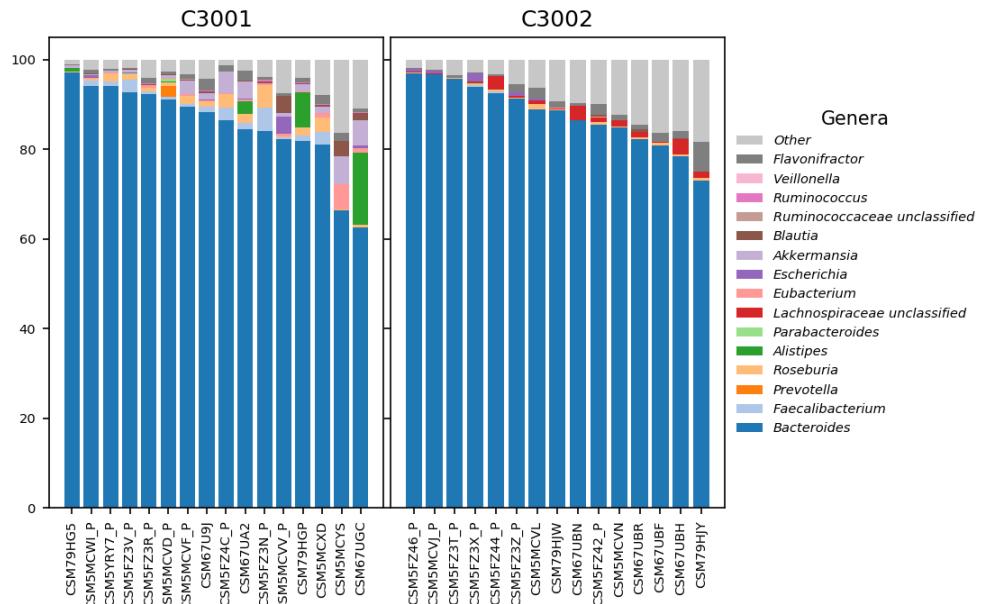
Top 15 genera by average abundance - site Emory to MGH Pediatrics



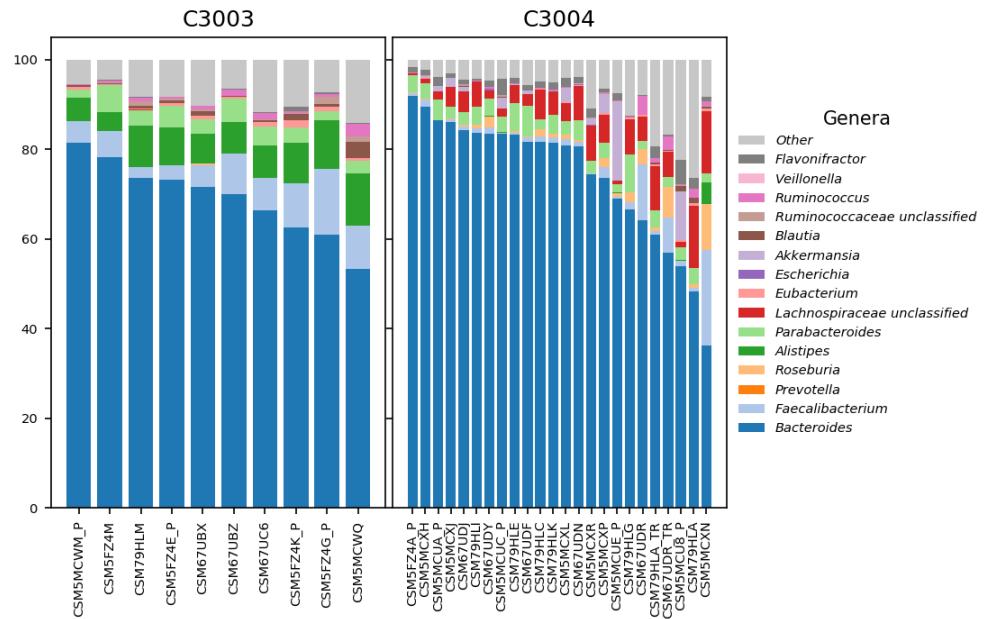
Top 15 genera by average abundance - diagnosis CD to nonIBD



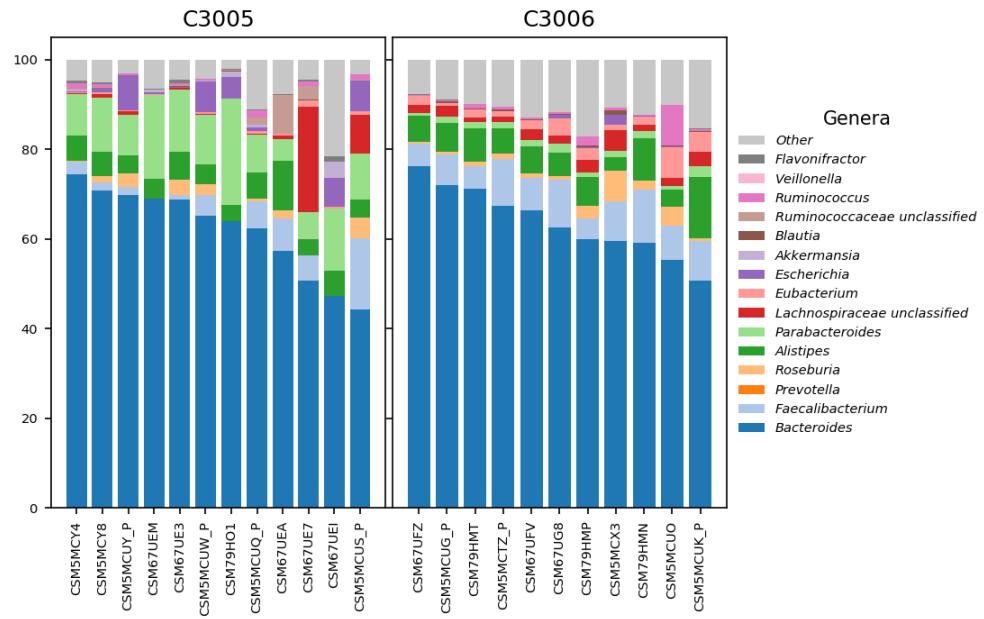
Top 15 genera by average abundance - subject C3001 to C3002



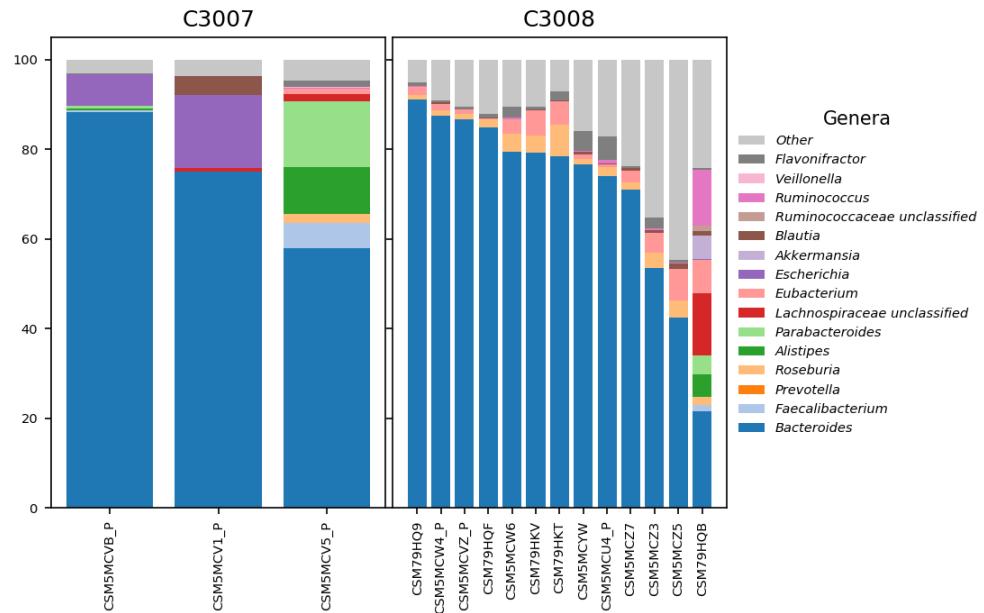
Top 15 genera by average abundance - subject C3003 to C3004



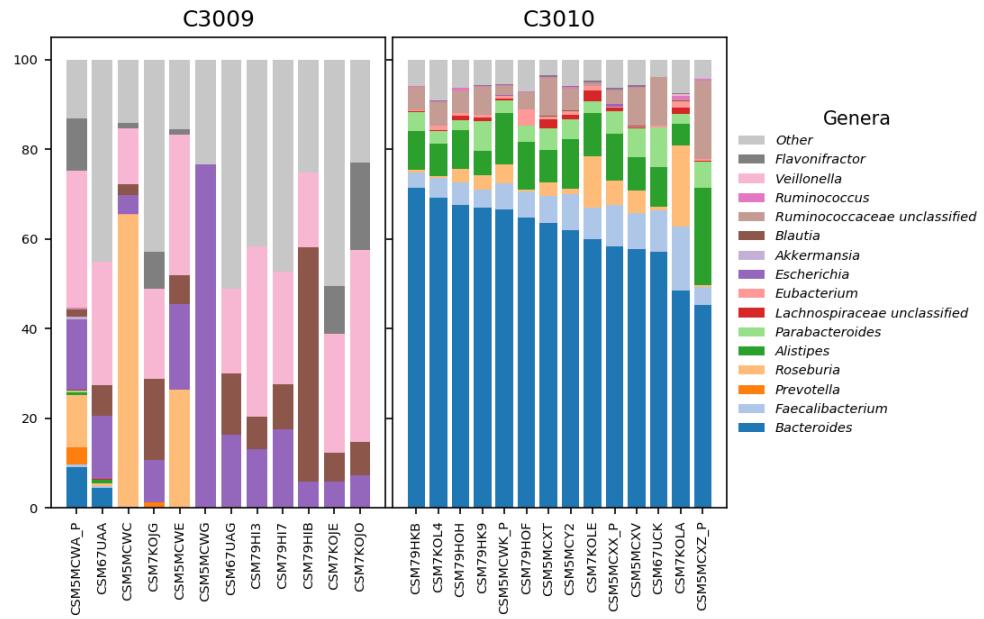
Top 15 genera by average abundance - subject C3005 to C3006



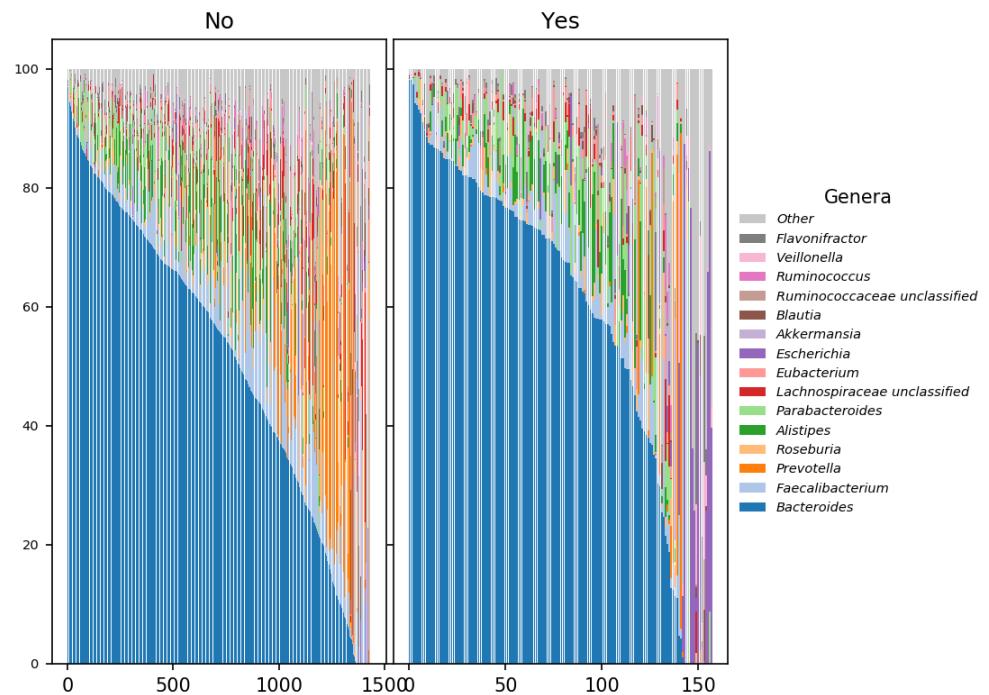
Top 15 genera by average abundance - subject C3007 to C3008

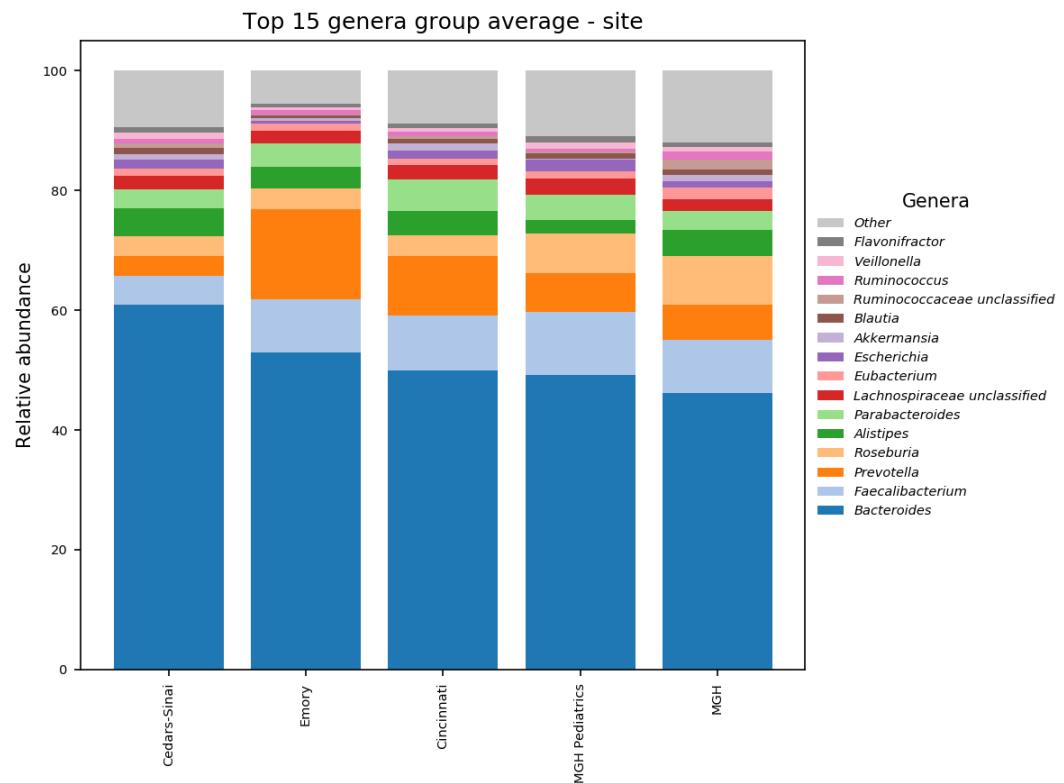


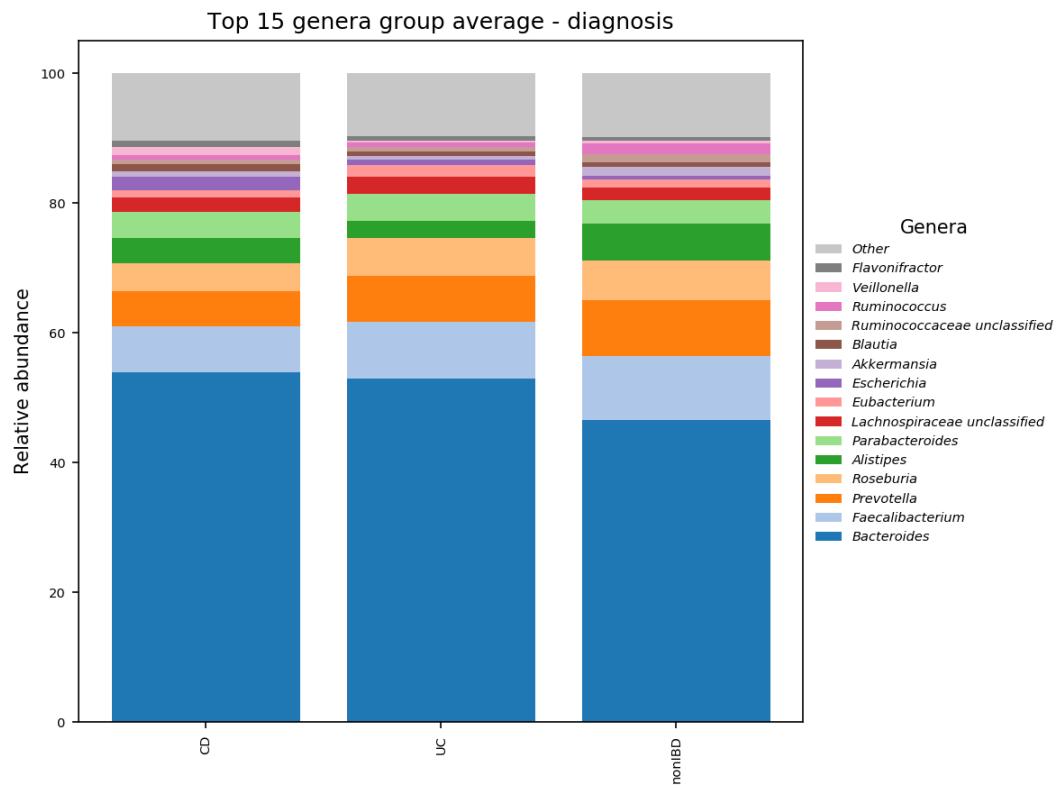
Top 15 genera by average abundance - subject C3009 to C3010

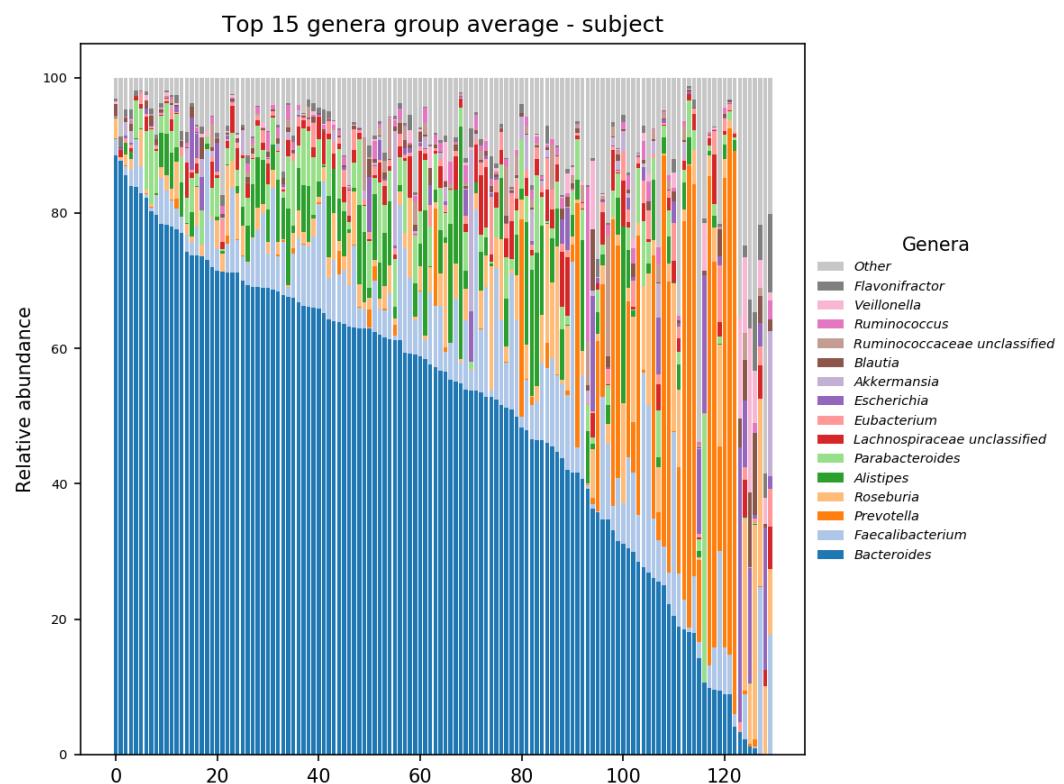


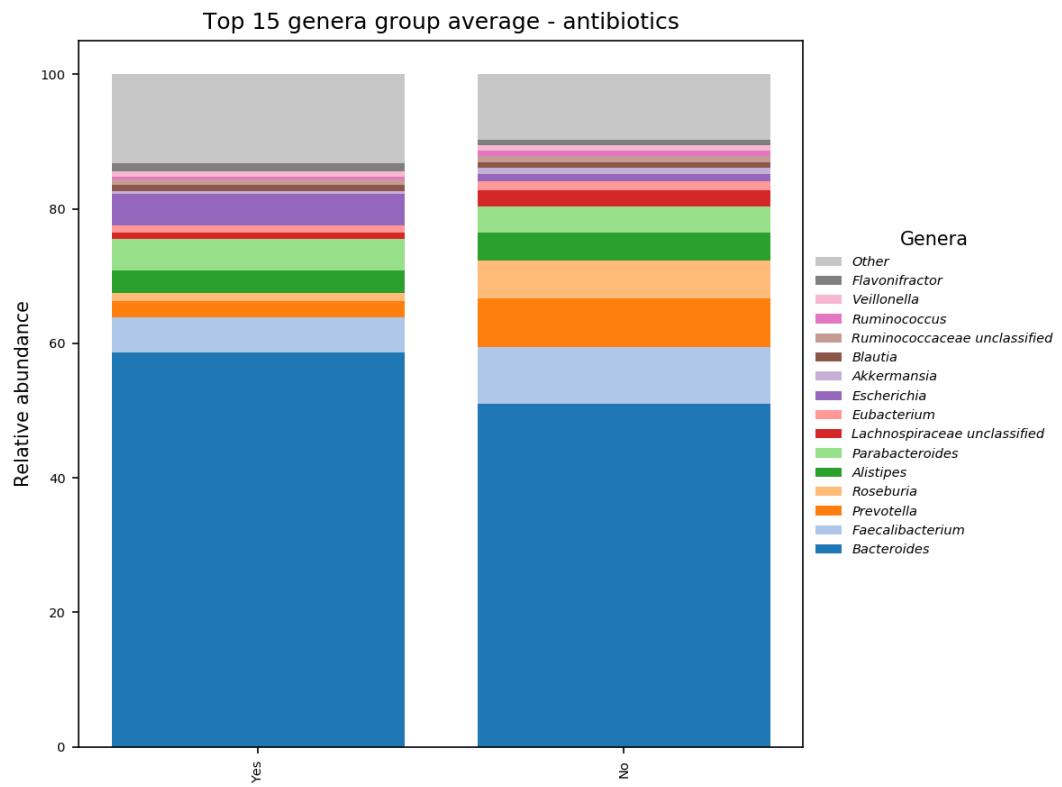
Top 15 genera by average abundance - antibiotics











Stacked barplot of genera abundance grouped by metadata and genera average abundance grouped by metadata. For cases where a metadata variable has many levels only the first 5 grouped barplots will be shown.

Functional Profiling of Metagenomic Reads

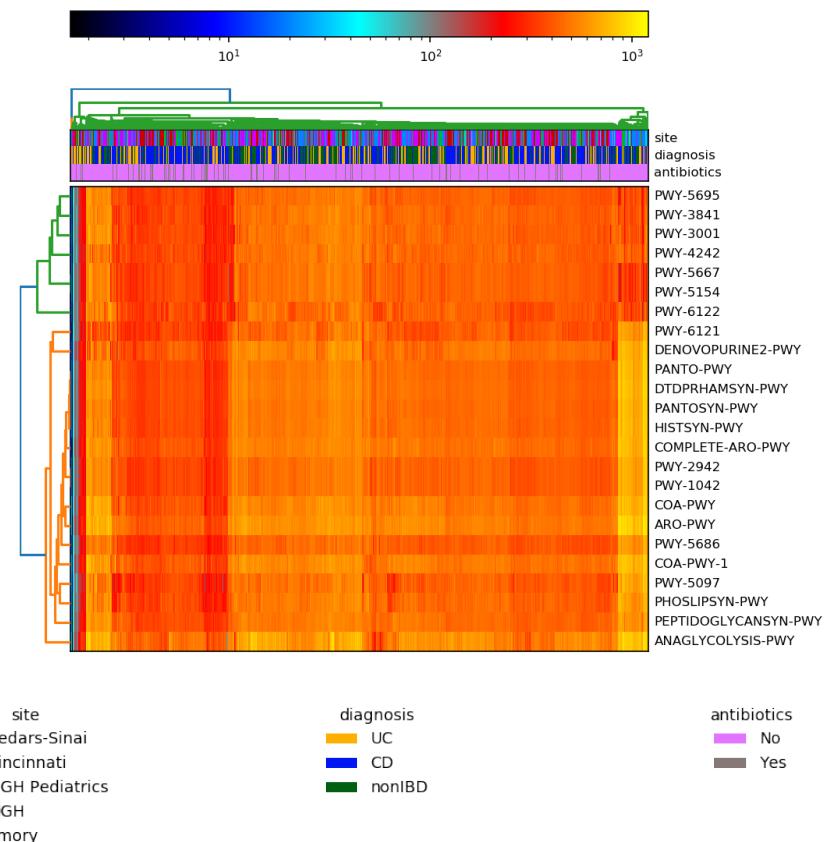
This report section contains preliminary exploratory figures that summarize HUMAnN functional profiling of all samples. HUMAnN performs species-specific and species-agnostic quantification of gene families, EC enzyme modules, and pathways, using the UniRef and MetaCyc databases. For more information on functional profiling and the databases used, see websites for [HUMAnN](#), [UniRef](#), and [MetaCyc](#).

Abundances are passed through a basic filter requiring each pathway to have at least 0.01 % abundance in at least 10 % of all samples. A total of 51 pathways were identified. After basic filtering 50 pathways remained.

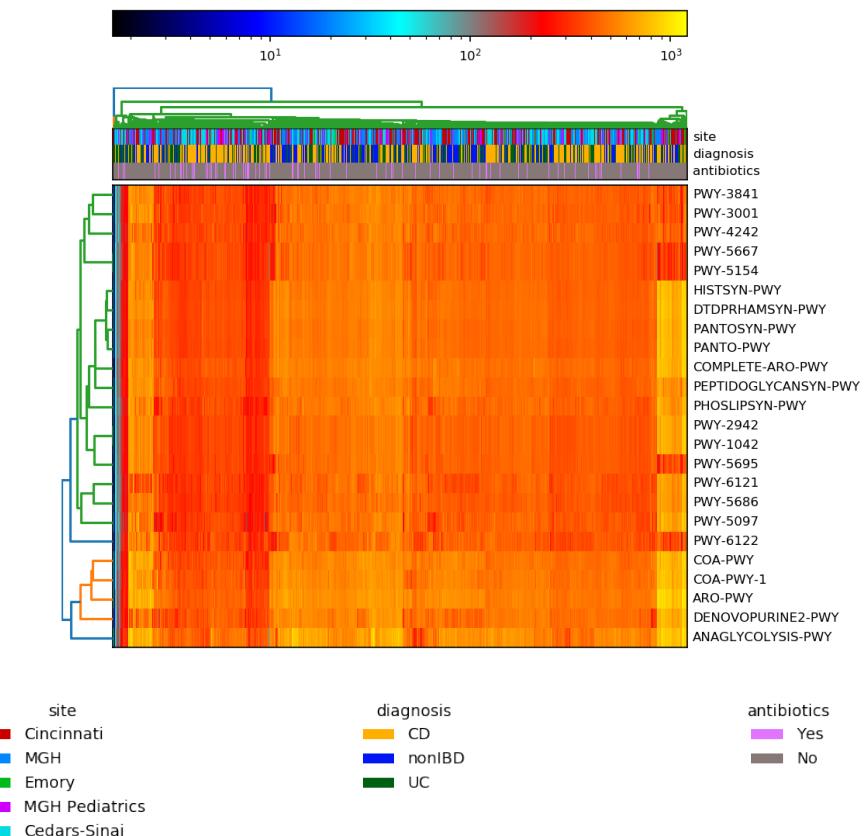
Pathway Abundance

Hierarchical clustering of samples and pathways, using top 25 pathways with highest mean relative abundance among samples. The ‘average linkage’ clustering on the Euclidean distance metric was used to cluster samples. The pathways dendrogram is based on pairwise (Spearman and Bray-Curtis) correlation between pathways. Samples are columns and pathways are rows. The heatmaps were generated with [Hclust2](#).

Top 25 pathways by average abundance (Spearman)



Top 25 pathways by average abundance (Bray-Curtis)



Top 25 pathways by average abundance (partial table)

	Average	Variance
DTDPRHAMSYN-PWY: dTDP-L-rhamnose biosynthesis I	6.93e+02	6.45e+04
PWY-1042: glycolysis IV (plant cytosol)	5.67e+02	2.33e+04
PWY-7219: adenosine ribonucleotides de novo biosynthesis	5.45e+02	1.86e+04
PWY-6109: adenine and adenosine salvage III	5.1e+02	1.52e+04
PWY-5680: UMP biosynthesis	5.05e+02	1.57e+04
PWY-7221: guanosine ribonucleotides de novo biosynthesis	4.88e+02	1.29e+04
PWY-5693: urate biosynthesis/inosine 5'-phosphate degradation	4.85e+02	1.35e+04
PWY-6387: UDP-N-acetylglucosaminyl-pentapeptide biosynthesis I (meso-diaminopimelate containing)	4.74e+02	1.44e+04
PWY-6386: UDP-N-acetylglucosaminyl-pentapeptide biosynthesis II (lysine-containing)	4.65e+02	1.34e+04
PEPTIDOGLYCAN5YN-PWY: peptidoglycan biosynthesis I (meso-diaminopimelate containing)	4.63e+02	1.36e+04
PWY-6385: peptidoglycan biosynthesis III (mycobacteria)	4.62e+02	1.26e+04
tRNA-CHARGING-PWY: tRNA charging	4.61e+02	8.7e+03
PWY-6700: queuosine biosynthesis	4.51e+02	1.22e+04
PWY-6122: 5-aminoimidazole ribonucleotide biosynthesis II	4.41e+02	1.15e+04
PWY-6277: superpathway of 5-aminoimidazole ribonucleotide biosynthesis	4.41e+02	1.15e+04
PWY-0163: chorismate biosynthesis from 3-dehydroquinate	4.33e+02	9.32e+03
COA-PWY-1: coenzyme A biosynthesis II (mammalian)	4.33e+02	8.73e+03
PWY-3841: folate transformations II	4.31e+02	7.91e+03
RIBOSYN2-PWY: flavin biosynthesis I (bacteria and plants)	4.29e+02	1.47e+04
PWY0-1319: CDP-diacylglycerol biosynthesis II	4.17e+02	8.66e+03

The table is too large to include the full table in this document. A partial table is shown which includes only 20 rows. Please see the data file for the full table: [top_average_pathways_names.tsv](#)

Detailed functions of the top 3 pathways can be found on the following MetaCyc pages:

- [DTDPRHAMSYN-PWY: dTDP-L-rhamnose biosynthesis I](#)
- [PWY-1042: glycolysis IV \(plant cytosol\)](#)
- [PWY-7219: adenosine ribonucleotides de novo biosynthesis](#)

To learn more about other pathways, search for the pathway by name on the MetaCyc website.

Data Processing Workflow Information

Software Versions

```
* kneaddata v0.6.1  
* humann v3  
* MetaPhlAn version 3
```

Tasks Run

```
* count_features.py --input metaphlan_taxonomic_profiles.tsv --output  
metaphlan_species_counts_table.tsv --include s__ --filter t__  
--reduce-sample-name  
  
* count_features.py --input genefamilies_relab.tsv --output  
humann_genealogies_relab_counts.tsv --reduce-sample-name --ignore-un-  
features --ignore-stratification  
  
* count_features.py --input pathabundance_relab.tsv --output  
humann_pathabundance_relab_counts.tsv --reduce-sample-name --ignore-  
un-features --ignore-stratification  
  
* count_features.py --input ecs_relab.tsv --output  
humann_ecs_relab_counts.tsv --reduce-sample-name --ignore-un-features  
--ignore-stratification  
  
* get_counts_from_humann_logs.py --input main --output  
humann_read_and_species_count_table.tsv  
  
* humann --input CSM5FZ4M.fastq --output main --o-log CSM5FZ4M.log  
--threads 1 --taxonomic-profile CSM5FZ4M_taxonomic_profile.tsv  
  
* humann --input CSM5MCU0.fastq --output main --o-log CSM5MCU0.log  
--threads 1 --taxonomic-profile CSM5MCU0_taxonomic_profile.tsv  
  
* humann --input CSM5MCVL.fastq --output main --o-log CSM5MCVL.log  
--threads 1 --taxonomic-profile CSM5MCVL_taxonomic_profile.tsv  
  
* humann --input CSM5MCVN.fastq --output main --o-log CSM5MCVN.log  
--threads 1 --taxonomic-profile CSM5MCVN_taxonomic_profile.tsv  
  
* humann --input CSM5MCW6.fastq --output main --o-log CSM5MCW6.log  
--threads 1 --taxonomic-profile CSM5MCW6_taxonomic_profile.tsv
```

```

* humann --input CSM5MCWC.fastq --output main --o-log CSM5MCWC.log
--threads 1 --taxonomic-profile CSM5MCWC_taxonomic_profile.tsv

* humann_join_tables --input main --output
metaphlan_taxonomic_profiles.tsv --file_name taxonomic_profile

* humann_join_tables --input main --output genefamilies.tsv
--file_name genefamilies

* humann_join_tables --input main --output pathabundance.tsv
--file_name pathabundance

* humann_join_tables --input genes --output genefamilies_relab.tsv

* humann_join_tables --input pathways --output pathabundance_relab.tsv

* humann_join_tables --input ecs --output ecs_relab.tsv

* humann_join_tables --input counts --output humann_feature_counts.tsv
--file_name _relab_counts.tsv

* humann_join_tables --input regrouped --output ecs.tsv --file_name
ecs

* humann_regroup_table --input CSM5FZ4M_genefamilies.tsv --output
CSM5FZ4M_ecs.tsv --groups uniref90_level4ec

* humann_regroup_table --input CSM5MCU0_genefamilies.tsv --output
CSM5MCU0_ecs.tsv --groups uniref90_level4ec

* humann_regroup_table --input CSM5MCVL_genefamilies.tsv --output
CSM5MCVL_ecs.tsv --groups uniref90_level4ec

* humann_regroup_table --input CSM5MCVN_genefamilies.tsv --output
CSM5MCVN_ecs.tsv --groups uniref90_level4ec

* humann_regroup_table --input CSM5MCW6_genefamilies.tsv --output
CSM5MCW6_ecs.tsv --groups uniref90_level4ec

* humann_regroup_table --input CSM5MCWC_genefamilies.tsv --output
CSM5MCWC_ecs.tsv --groups uniref90_level4ec

* humann_renorm_table --input CSM5FZ4M_ecs.tsv --output
CSM5FZ4M_ecs_relab.tsv --units relab --special n

```

```

* humann_renorm_table --input CSM5FZ4M_genefamilies.tsv --output
CSM5FZ4M_genefamilies_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5FZ4M_pathabundance.tsv --output
CSM5FZ4M_pathabundance_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCU0_ecs.tsv --output
CSM5MCU0_ecs_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCU0_genefamilies.tsv --output
CSM5MCU0_genefamilies_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCU0_pathabundance.tsv --output
CSM5MCU0_pathabundance_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCVL_ecs.tsv --output
CSM5MCVL_ecs_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCVL_genefamilies.tsv --output
CSM5MCVL_genefamilies_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCVL_pathabundance.tsv --output
CSM5MCVL_pathabundance_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCVN_ecs.tsv --output
CSM5MCVN_ecs_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCVN_genefamilies.tsv --output
CSM5MCVN_genefamilies_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCVN_pathabundance.tsv --output
CSM5MCVN_pathabundance_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCW6_ecs.tsv --output
CSM5MCW6_ecs_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCW6_genefamilies.tsv --output
CSM5MCW6_genefamilies_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCW6_pathabundance.tsv --output
CSM5MCW6_pathabundance_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCWC_genefamilies.tsv --output
CSM5MCWC_genefamilies_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCWC_pathabundance.tsv --output

```

```

CSM5MCWC_pathabundance_relab.tsv --units relab --special n

* humann_renorm_table --input CSM5MCWC_ecs.tsv --output
CSM5MCWC_ecs_relab.tsv --units relab --special n

* kneaddata --input CSM5FZ4M.fastq.gz --output main --threads 1
--output-prefix CSM5FZ4M --reference-db Homo_sapiens_demo --serial

* kneaddata --input CSM5MCU0.fastq.gz --output main --threads 1
--output-prefix CSM5MCU0 --reference-db Homo_sapiens_demo --serial

* kneaddata --input CSM5MCVL.fastq.gz --output main --threads 1
--output-prefix CSM5MCVL --reference-db Homo_sapiens_demo --serial

* kneaddata --input CSM5MCVN.fastq.gz --output main --threads 1
--output-prefix CSM5MCVN --reference-db Homo_sapiens_demo --serial

* kneaddata --input CSM5MCW6.fastq.gz --output main --threads 1
--output-prefix CSM5MCW6 --reference-db Homo_sapiens_demo --serial

* kneaddata --input CSM5MCWC.fastq.gz --output main --threads 1
--output-prefix CSM5MCWC --reference-db Homo_sapiens_demo --serial

* kneaddata_read_count_table --input main --output
kneaddata_read_count_table.tsv

* metaphlan CSM5FZ4M.fastq --input_type fastq --output_file
CSM5FZ4M_taxonomic_profile.tsv --samout CSM5FZ4M_bowtie2.sam --nproc 1
--no_map --tmp_dir main

* metaphlan CSM5MCU0.fastq --input_type fastq --output_file
CSM5MCU0_taxonomic_profile.tsv --samout CSM5MCU0_bowtie2.sam --nproc 1
--no_map --tmp_dir main

* metaphlan CSM5MCVL.fastq --input_type fastq --output_file
CSM5MCVL_taxonomic_profile.tsv --samout CSM5MCVL_bowtie2.sam --nproc 1
--no_map --tmp_dir main

* metaphlan CSM5MCVN.fastq --input_type fastq --output_file
CSM5MCVN_taxonomic_profile.tsv --samout CSM5MCVN_bowtie2.sam --nproc 1
--no_map --tmp_dir main

* metaphlan CSM5MCWC.fastq --input_type fastq --output_file
CSM5MCWC_taxonomic_profile.tsv --samout CSM5MCWC_bowtie2.sam --nproc 1
--no_map --tmp_dir main

```

```
* metaphlan.py CSM5MCW6.fastq --input_type fastq --output_file  
CSM5MCW6_taxonomic_profile.tsv --samout CSM5MCW6_bowtie2.sam --nproc 1  
--no_map --tmp_dir main
```