

Earth Hologenome Initiative final report

DMB0038

2023-10-29

Contents

Introduction	2
Data availability	2
1. General overview	3
1.1 Produced data	3
1.2 Sample metadata	4
2 MAG catalogue	6
2.1 Phylogenetic tree	6
2.2 Quality of MAGs	7
2.3 Functional attributes of MAGs	20
2.4 Functional ordination of MAGs	21
3. Sequencing depth assessment	22
3.1 DNA fractions	22
3.2 Estimated vs. mapped prokaryotic fraction	23
3.3 Extra sequencing effort required	24
4. Count data	26
4.1 Minimum genome-coverage filtering	26
4.2 Genome-size normalisation	26
4.3 Count table	27
5. Taxonomic composition	28
6. Diversity analyses	29
6.1 Alpha diversity	29
6.2 Relationship between sequencing and alpha diversities	35
6.3 Beta diversity	36

Introduction

This is the report automatically generated by the Earth Holohome Initiative bioinformatic pipeline. Here you will find basic information of the dataset, which aims to guide you in assessing whether data quality is good enough, whether more sequencing is needed, as well as formulating downstream statistical analyses. Note that the analyses and visualisations displayed here are not study-specific, so they are not intended to serve as conclusive analyses. The reporting is still in developmental phase.

Data availability

The data used for generating this report is available through the following links:

Count table

https://sid.erda.dk/share_redirect/BaMZodj9sA/DMB/DMB0038/DMB0038_counts.tsv.gz

It contains the read-count table specifying how many reads mapped to which MAG in each sample. This table is used for quantifying MAGs.

Coverage table

https://sid.erda.dk/share_redirect/BaMZodj9sA/DMB/DMB0038/DMB0038_coverage.tsv.gz

It contains the breadth of coverage values for each MAG in each sample. This table is used for filtering MAGs.

Sample metadata

https://sid.erda.dk/share_redirect/BaMZodj9sA/DMB/DMB0038/DMB0038_metadata.tsv.gz

It contains the basic metadata of each sample, including host species, location, sample type and similar.

MAG metadata

https://sid.erda.dk/share_redirect/BaMZodj9sA/DMB/DMB0038/DMB0038_mag_info.tsv.gz

It contains the structural, taxonomic and quality information of all MAGs. It is used for MAG filtering and visualisations.

MAG phylogenetic tree

https://sid.erda.dk/share_redirect/BaMZodj9sA/DMB/DMB0038/DMB0038.tree.gz

It contains the phylogenetic tree of all MAGs. It is used for phylogenetic diversity analyses and visualisations.

MAG function table

https://sid.erda.dk/share_redirect/BaMZodj9sA/DMB/DMB0038/DMB0038_merged_kegg.tsv.gz

It contains the distilled KEGG pathway fullness values for all MAGs. It is used for functional diversity analyses and visualisations.

1. General overview

1.1 Produced data

Number of samples: 38

Amount of total data: 146.59 GB

The amount of total DNA data sequenced in gigabases (GB, one billion nucleotide bases). As a reference, the amount of data processed in this batch is equivalent to 47.29 human genomes.

Amount of discarded data: 6.41 GB (4.37 % of the raw data)

The amount of data discarded due to low quality or lack of informativeness during data preprocessing. Discarding 5-15% of the produced data is within the expected range, due to formation of adaptor dimers, inclusion of adaptors in sequencing reads due to short insert sizes, low sequencing quality, etc.

Amount of host data: 7.05 GB (5.03 % of the quality-filtered data)

The amount of data mapped against the host genome. The percentage refers to the amount of data mapped to the host genome respect to quality-filtered data. Note that this value can be very variable depending on the biological features of the sample (e.g., anal swabs contain more host DNA than faeces) and the employed reference genome (e.g., the chances for mapping to the genome are lower as the distance between the study species and the employed reference genome differ).

Amount of metagenomic data: 133.13 GB (94.97 % of the quality-filtered data)

The amount of data not-mapped against the host genome. Note that this fraction does not only include the prokaryotic DNA that we intend to reconstruct into MAGs, but also dietary DNA, viruses, etc., which will not map against the MAG catalogue.

Estimated prokaryotic data: 88.9 GB (66.77 % of the metagenomic data)

The amount and proportion of data belonging to prokaryotic genomes respect to the total metagenomic fraction, as estimated from singleM analysis. Note that this is an estimation that relies on the genome sizes of genomes available in reference databases. If a given taxon is not properly represented, genome size estimations can be less accurate.

Number of MAGs: 581

Out of which 521 (89.7%) are new species

The number of metagenome-assembled genomes (MAG) or draft bacterial genomes reconstructed from the metagenomic data.

Total mapped sequencing depth: 621.1 million reads (88.82 GB)

The amount of reads (and nucleotide bases) that were mapped to the entire MAG catalogue. Note that the amount of bases is only an approximation estimated by multiplying the exact number of mapped reads by 250 bp.

Average mapped sequencing depth: 16.34±9.4 million reads (2.34±2.82 GB)

This is the average number of reads (and nucleotide bases) mapped to each sample. Note that the amount of bases is only an approximation estimated by multiplying the exact number of mapped reads by 250 bp.

1.2 Sample metadata

The samples processed in this batch were collected in the following locations:

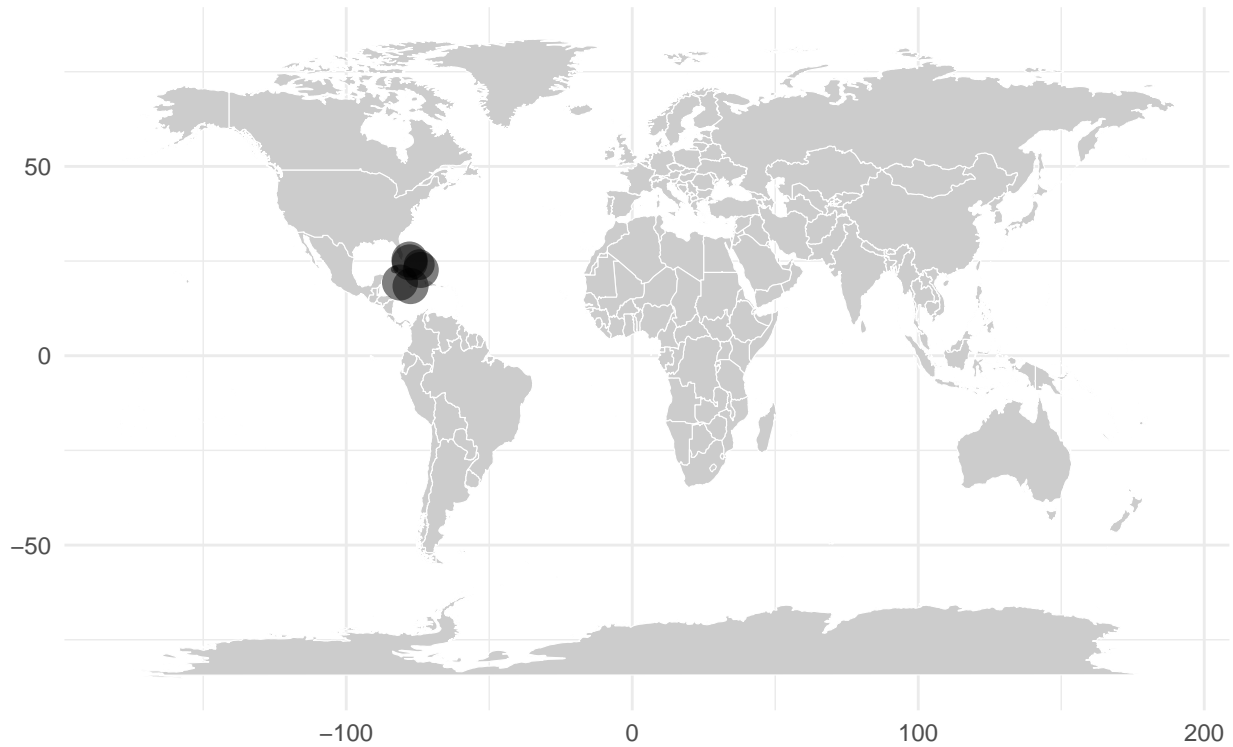


Figure 1: Geographic location of samples. Size indicates the number of samples.

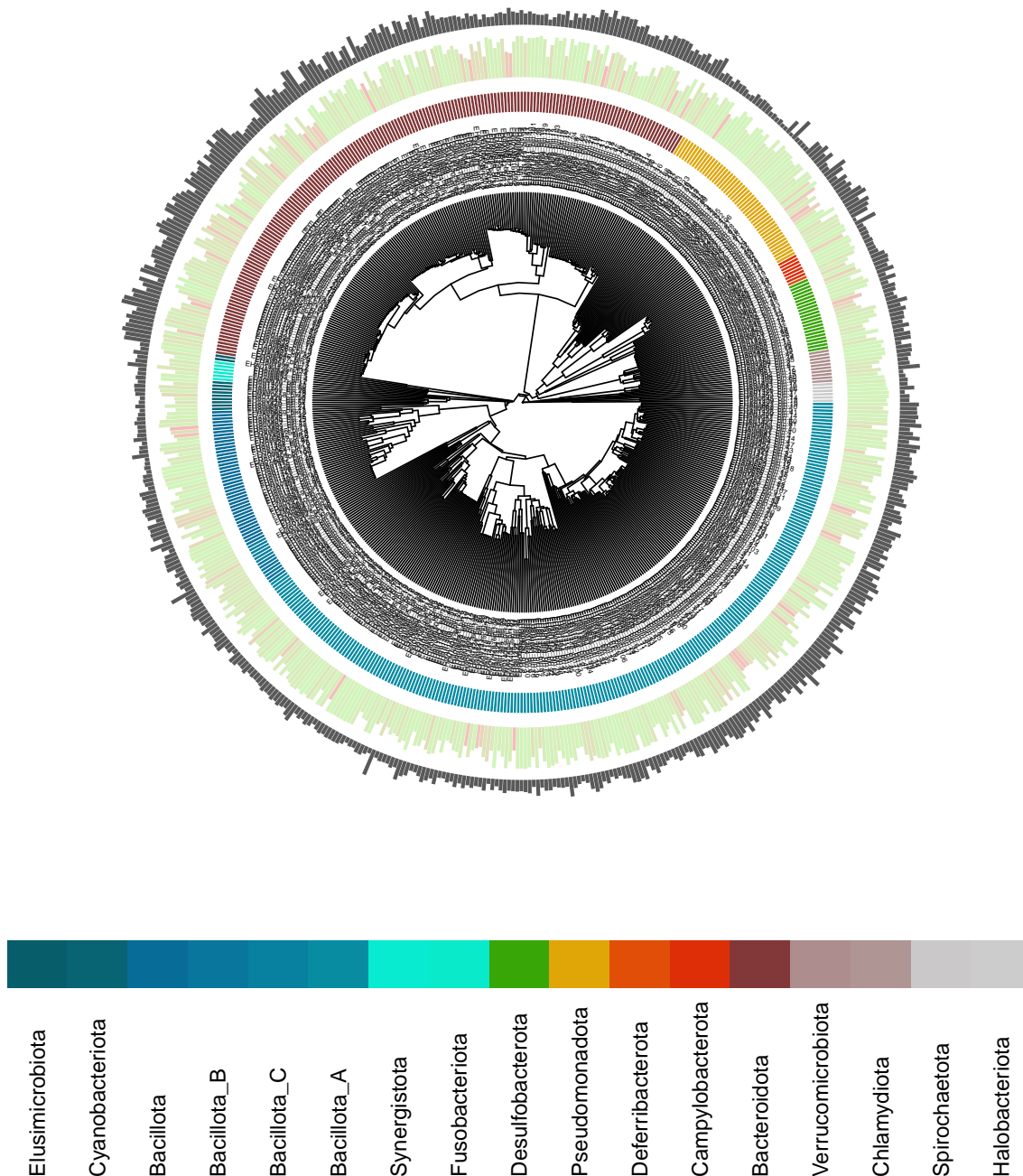
sample	sample_type	country	latitude	longitude
EH100678	Faecal	Bahamas	24.05862	-74.46739
EH100682	Faecal	Bahamas	25.40986	-77.87054
EH100743	Faecal	Jamaica	18.38248	-77.51726
EH100679	Faecal	Bahamas	25.40986	-77.87054
EH100674	Faecal	Jamaica	18.38248	-77.51726
EH100689	Faecal	Bahamas	24.66547	-77.80168
EH100688	Faecal	Bahamas	22.65397	-73.93493
EH100690	Faecal	Bahamas	24.66547	-77.80168
EH100692	Faecal	Bahamas	24.66547	-77.80168
EH100753	Faecal	Bahamas	25.40986	-77.87054
EH100696	Faecal	Cayman Islands	19.32504	-81.20698
EH100751	Faecal	Jamaica	18.38248	-77.51726
EH100680	Faecal	Bahamas	24.05862	-74.46739
EH100683	Faecal	Bahamas	24.05862	-74.46739
EH100732	Faecal	Cuba	22.77456	-83.03659
EH100695	Faecal	Cayman Islands	19.32504	-81.20698
EH100693	Faecal	Bahamas	24.66547	-77.80168
EH100731	Faecal	Bahamas	22.65397	-73.93493
EH100675	Faecal	Jamaica	18.38248	-77.51726
EH100686	Faecal	Bahamas	22.65397	-73.93493
EH100681	Faecal	Bahamas	24.05862	-74.46739

sample	sample_type	country	latitude	longitude
EH100758	Faecal	Cuba	22.77456	-83.03659
EH100685	Faecal	Bahamas	22.65397	-73.93493
EH100752	Faecal	Bahamas	25.40986	-77.87054
EH100697	Faecal	Cayman Islands	19.32504	-81.20698
EH100742	Faecal	Jamaica	18.38248	-77.51726
EH100698	Faecal	Cayman Islands	19.32504	-81.20698
EH100687	Faecal	Bahamas	22.65397	-73.93493
EH100684	Faecal	Bahamas	22.65397	-73.93493
EH100733	Faecal	Cuba	22.77456	-83.03659
EH100691	Faecal	Bahamas	24.66547	-77.80168
EH100757	Faecal	Bahamas	25.40986	-77.87054
EH100699	Faecal	Cayman Islands	19.32504	-81.20698
EH100677	Faecal	Bahamas	24.05862	-74.46739
EH100700	Faecal	Cayman Islands	19.32504	-81.20698
EH100730	Faecal	Bahamas	25.40986	-77.87054
EH100694	Faecal	Bahamas	24.66547	-77.80168
EH100676	Faecal	Jamaica	18.38248	-77.51726

2 MAG catalogue

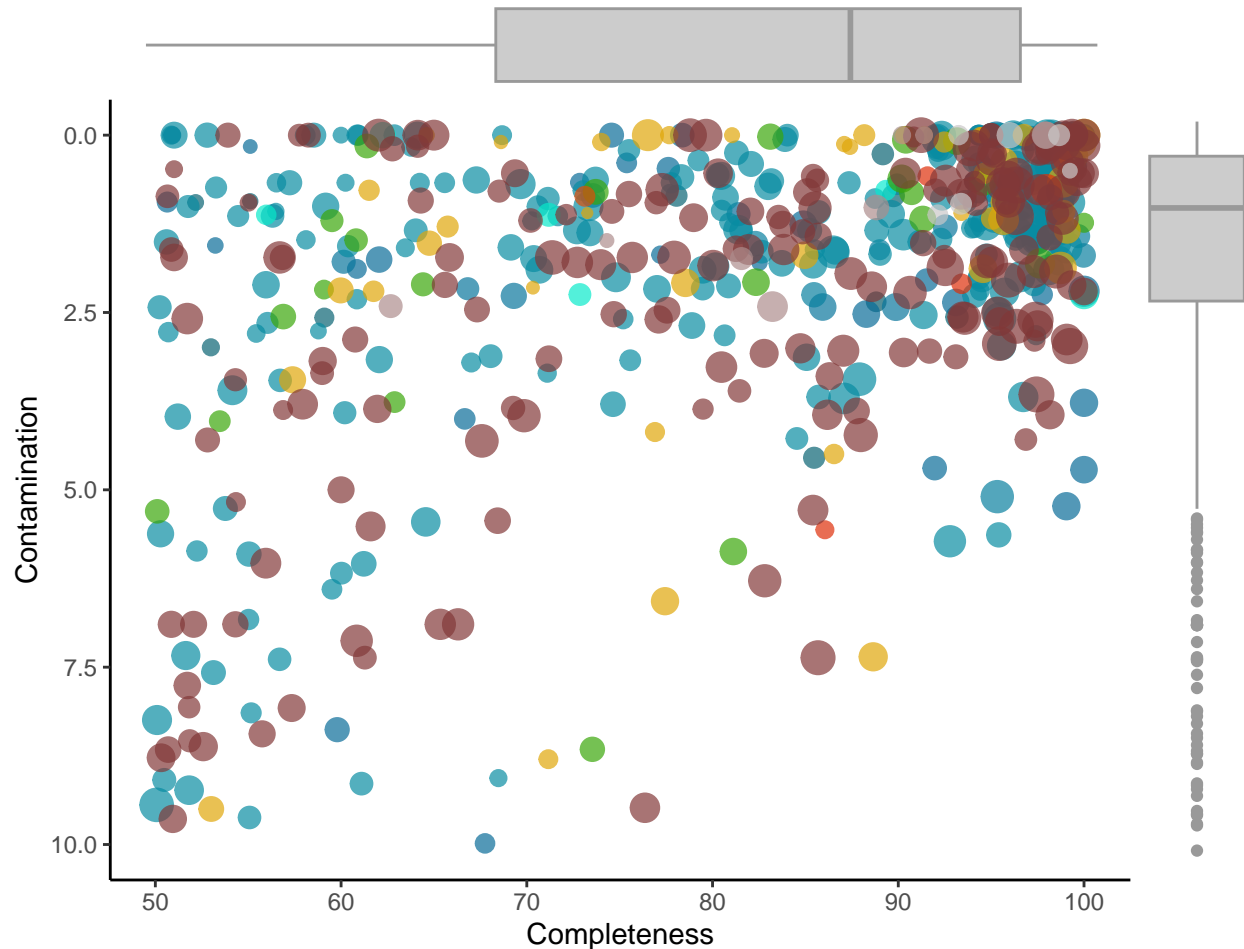
2.1 Phylogenetic tree

The phylogenetic tree is constructed by placing the MAG sequences within the reference archaeal and bacterial trees using GTDBTK, followed by merging both trees. The external grey ring indicates genome sizes of the MAGs, while the middle ring shows an overview of completeness (height) and contamination (green = low, red = high). The original tree can be downloaded here: **Phylogenetic tree file**.



2.2 Quality of MAGs

The two main parameters to assess the quality of reconstructed MAG as completeness and contamination (or redundancy). These parameters are calculated by CheckM based on the presence of single-copy core genes in the MAGs. Completeness values should ideally be as close to 100 as possible, because incompleteness introduces biases when performing functional analyses. Contamination should be as close to 0 as possible, because it indicates that the MAG probably contains DNA fragments that in reality belong to a different genome, and therefore can distort the results and introduce noise. In the below plot, MAGs are ordinated according to their completeness and contamination scores, while dot sizes indicate their genome-size. The following page(s) contain(s) an overview of these values for every MAG, while the complete MAG feature table can be downloaded here: **[MAG feature table file](#)**



genome	domain	phylum	comp	cont	size
EHA01404_bin.132	d__Bacteria	p__Elusimicrobiota	88.81	2.349	1.61
EHA01421_bin.6	d__Bacteria	p__Elusimicrobiota	96.62	0.000	1.79
EHA00671_bin.18	d__Bacteria	p__Cyanobacteriota	93.16	2.564	2.14
EHA01409_bin.28	d__Bacteria	p__Cyanobacteriota	59.10	2.572	1.49
EHA01426_bin.52	d__Bacteria	p__Cyanobacteriota	52.99	2.991	1.21
EHA00651_bin.24	d__Bacteria	p__Cyanobacteriota	52.17	0.949	1.05
EHA00660_bin.21	d__Bacteria	p__Cyanobacteriota	78.06	0.854	1.86
EHA00641_bin.12	d__Bacteria	p__Cyanobacteriota	55.11	0.932	0.99
EHA01426_bin.50	d__Bacteria	p__Cyanobacteriota	85.47	4.548	1.87
EHA00658_bin.2	d__Bacteria	p__Cyanobacteriota	89.74	0.854	2.03
EHA01421_bin.58	d__Bacteria	p__Cyanobacteriota	89.17	0.267	1.93
EHA00659_bin.8	d__Bacteria	p__Bacillota	75.38	0.391	1.61
EHA00660_bin.11	d__Bacteria	p__Bacillota	72.85	0.667	1.38
EHA00645_bin.15	d__Bacteria	p__Bacillota	94.11	0.653	1.60
EHA01409_bin.32	d__Bacteria	p__Bacillota	91.53	2.042	2.12
EHA00666_bin.3	d__Bacteria	p__Bacillota	94.30	0.000	1.60
EHA00645_bin.9	d__Bacteria	p__Bacillota	88.28	2.522	3.19
EHA01404_bin.24	d__Bacteria	p__Bacillota	97.33	1.444	1.74
EHA00667_bin.17	d__Bacteria	p__Bacillota	66.66	4.000	1.77
EHA00664_bin.17	d__Bacteria	p__Bacillota	53.23	1.555	1.02
EHA01409_bin.82	d__Bacteria	p__Bacillota	59.79	8.381	2.54
EHA01409_bin.25	d__Bacteria	p__Bacillota	67.75	9.984	1.58
EHA01409_bin.100	d__Bacteria	p__Bacillota	97.19	0.000	1.28
EHA01404_bin.82	d__Bacteria	p__Bacillota	97.75	2.247	1.74
EHA00672_bin.14	d__Bacteria	p__Bacillota	100.00	0.000	1.91
EHA00641_bin.9	d__Bacteria	p__Bacillota	100.00	0.000	1.77
EHA01421_bin.11	d__Bacteria	p__Bacillota	99.52	0.000	2.90
EHA00657_bin.4	d__Bacteria	p__Bacillota	100.00	3.773	3.38
EHA01421_bin.39	d__Bacteria	p__Bacillota	99.05	5.231	3.37
EHA01404_bin.45	d__Bacteria	p__Bacillota	100.00	4.716	3.16
EHA01426_bin.58	d__Bacteria	p__Bacillota	84.89	1.257	3.22
EHA00651_bin.42	d__Bacteria	p__Bacillota	100.00	2.201	4.30
EHA01426_bin.95	d__Bacteria	p__Bacillota	50.64	0.943	1.66
EHA00659_bin.2	d__Bacteria	p__Bacillota	78.37	0.000	2.41
EHA00645_bin.11	d__Bacteria	p__Bacillota	50.91	0.000	1.19
EHA01421_bin.14	d__Bacteria	p__Bacillota	91.96	4.694	2.40
EHA00650_bin.3	d__Bacteria	p__Bacillota	96.22	0.000	2.84
EHA00669_bin.20	d__Bacteria	p__Bacillota	98.11	0.943	2.90
EHA00641_bin.3	d__Bacteria	p__Bacillota	99.05	0.000	3.04
EHA00666_bin.36	d__Bacteria	p__Bacillota	94.11	0.653	1.59
EHA00674_bin.8	d__Bacteria	p__Bacillota	95.30	0.915	1.55
EHA00655_bin.5	d__Bacteria	p__Bacillota	77.62	0.471	2.12
EHA01409_bin.14	d__Bacteria	p__Bacillota	60.85	1.886	1.46
EHA01404_bin.67	d__Bacteria	p__Bacillota	74.56	0.000	2.51
EHA01415_bin.46	d__Bacteria	p__Bacillota	93.10	2.358	3.14
EHA01426_bin.89	d__Bacteria	p__Bacillota	98.07	1.121	3.43
EHA00657_bin.12	d__Bacteria	p__Bacillota	95.28	0.000	3.13
EHA01426_bin.88	d__Bacteria	p__Bacillota	81.40	1.114	2.27
EHA01415_bin.25	d__Bacteria	p__Bacillota	62.05	1.754	2.87
EHA01426_bin.78	d__Bacteria	p__Bacillota	100.00	0.000	0.96
EHA00652_bin.6	d__Bacteria	p__Bacillota	55.11	0.160	0.85
EHA01404_bin.87	d__Bacteria	p__Bacillota	77.19	1.685	1.24

genome	domain	phylum	comp	cont	size
EHA01426_bin.82	d__Bacteria	p__Bacillota	82.53	1.797	1.00
EHA01415_bin.37	d__Bacteria	p__Bacillota_B	60.14	1.785	2.28
EHA00658_bin.19	d__Bacteria	p__Bacillota_B	69.30	2.266	2.79
EHA01426_bin.71	d__Bacteria	p__Bacillota_B	85.93	2.423	3.17
EHA01421_bin.75	d__Bacteria	p__Bacillota_B	89.87	2.423	3.27
EHA01415_bin.55	d__Bacteria	p__Bacillota_B	73.55	0.850	2.42
EHA01426_bin.26	d__Bacteria	p__Bacillota_B	66.82	2.165	2.06
EHA00661_bin.28	d__Bacteria	p__Bacillota_B	92.53	1.292	2.47
EHA01404_bin.116	d__Bacteria	p__Bacillota_C	99.68	1.424	2.31
EHA01415_bin.80	d__Bacteria	p__Bacillota_C	82.31	1.529	2.61
EHA00667_bin.15	d__Bacteria	p__Bacillota_C	89.00	1.615	1.67
EHA00645_bin.3	d__Bacteria	p__Bacillota_C	95.33	5.096	5.04
EHA00666_bin.37	d__Bacteria	p__Bacillota_C	85.46	2.250	2.59
EHA01421_bin.29	d__Bacteria	p__Bacillota_C	99.98	1.696	2.39
EHA01409_bin.76	d__Bacteria	p__Bacillota_A	80.70	0.877	2.92
EHA01404_bin.20	d__Bacteria	p__Bacillota_A	85.72	1.085	2.25
EHA01426_bin.70	d__Bacteria	p__Bacillota_A	56.72	3.457	2.16
EHA01404_bin.74	d__Bacteria	p__Bacillota_A	79.47	2.163	2.20
EHA00646_bin.24	d__Bacteria	p__Bacillota_A	98.93	0.806	3.46
EHA01409_bin.13	d__Bacteria	p__Bacillota_A	60.92	0.000	1.66
EHA01409_bin.79	d__Bacteria	p__Bacillota_A	92.08	0.000	1.57
EHA01409_bin.70	d__Bacteria	p__Bacillota_A	99.10	1.342	2.97
EHA01409_bin.67	d__Bacteria	p__Bacillota_A	95.41	5.635	2.58
EHA01409_bin.16	d__Bacteria	p__Bacillota_A	77.68	0.806	1.06
EHA01426_bin.42	d__Bacteria	p__Bacillota_A	80.65	2.822	1.81
EHA01409_bin.24	d__Bacteria	p__Bacillota_A	97.31	2.104	2.20
EHA01404_bin.117	d__Bacteria	p__Bacillota_A	94.73	0.476	1.85
EHA00656_bin.1	d__Bacteria	p__Bacillota_A	81.29	0.249	2.44
EHA00645_bin.1	d__Bacteria	p__Bacillota_A	94.35	0.000	1.67
EHA01404_bin.47	d__Bacteria	p__Bacillota_A	59.51	6.402	1.60
EHA01404_bin.119	d__Bacteria	p__Bacillota_A	85.51	1.612	1.53
EHA01404_bin.78	d__Bacteria	p__Bacillota_A	92.06	0.201	1.45
EHA01409_bin.89	d__Bacteria	p__Bacillota_A	68.05	3.116	2.22
EHA01404_bin.41	d__Bacteria	p__Bacillota_A	94.32	0.806	1.88
EHA00664_bin.6	d__Bacteria	p__Bacillota_A	75.50	0.213	1.92
EHA01409_bin.44	d__Bacteria	p__Bacillota_A	91.15	1.505	2.33
EHA00645_bin.22	d__Bacteria	p__Bacillota_A	98.65	1.525	2.42
EHA00641_bin.14	d__Bacteria	p__Bacillota_A	94.35	1.209	2.01
EHA01409_bin.97	d__Bacteria	p__Bacillota_A	96.31	1.478	1.76
EHA00674_bin.4	d__Bacteria	p__Bacillota_A	97.76	1.118	2.08
EHA01409_bin.84	d__Bacteria	p__Bacillota_A	98.65	0.000	2.92
EHA00666_bin.32	d__Bacteria	p__Bacillota_A	92.34	0.000	3.52
EHA00670_bin.3	d__Bacteria	p__Bacillota_A	94.63	0.468	1.95
EHA01409_bin.69	d__Bacteria	p__Bacillota_A	97.42	2.820	1.51
EHA01415_bin.70	d__Bacteria	p__Bacillota_A	60.85	2.313	1.56
EHA00651_bin.4	d__Bacteria	p__Bacillota_A	95.96	0.000	2.23
EHA00671_bin.12	d__Bacteria	p__Bacillota_A	94.35	2.016	1.95
EHA00654_bin.13	d__Bacteria	p__Bacillota_A	68.47	9.064	1.21
EHA00665_bin.9	d__Bacteria	p__Bacillota_A	96.37	0.000	1.82
EHA00660_bin.12	d__Bacteria	p__Bacillota_A	95.96	0.000	2.12
EHA01421_bin.15	d__Bacteria	p__Bacillota_A	95.16	0.806	1.86
EHA00668_bin.6	d__Bacteria	p__Bacillota_A	95.56	0.806	2.12

genome	domain	phylum	comp	cont	size
EHA00657_bin.11	d__Bacteria	p__Bacillota_A	95.16	0.000	2.05
EHA00672_bin.6	d__Bacteria	p__Bacillota_A	95.56	0.000	2.07
EHA00655_bin.4	d__Bacteria	p__Bacillota_A	94.35	0.403	2.04
EHA00661_bin.34	d__Bacteria	p__Bacillota_A	93.95	1.309	1.68
EHA01404_bin.85	d__Bacteria	p__Bacillota_A	56.49	1.075	1.10
EHA00654_bin.19	d__Bacteria	p__Bacillota_A	68.67	0.000	1.47
EHA00650_bin.11	d__Bacteria	p__Bacillota_A	70.27	1.209	1.36
EHA00657_bin.15	d__Bacteria	p__Bacillota_A	94.75	0.026	1.99
EHA00663_bin.7	d__Bacteria	p__Bacillota_A	94.75	1.686	1.95
EHA01404_bin.19	d__Bacteria	p__Bacillota_A	92.14	0.223	1.84
EHA01426_bin.11	d__Bacteria	p__Bacillota_A	94.51	2.013	3.48
EHA00659_bin.11	d__Bacteria	p__Bacillota_A	51.76	1.006	2.05
EHA00666_bin.6	d__Bacteria	p__Bacillota_A	99.41	0.352	1.78
EHA00660_bin.32	d__Bacteria	p__Bacillota_A	64.02	1.342	2.23
EHA00660_bin.9	d__Bacteria	p__Bacillota_A	57.23	0.671	2.54
EHA00653_bin.4	d__Bacteria	p__Bacillota_A	96.08	0.671	3.32
EHA01404_bin.75	d__Bacteria	p__Bacillota_A	98.65	1.342	4.22
EHA00652_bin.16	d__Bacteria	p__Bacillota_A	94.74	1.260	4.46
EHA00660_bin.8	d__Bacteria	p__Bacillota_A	96.73	3.691	4.00
EHA00651_bin.26	d__Bacteria	p__Bacillota_A	92.78	5.725	4.66
EHA00655_bin.6	d__Bacteria	p__Bacillota_A	72.70	1.342	3.27
EHA00662_bin.5	d__Bacteria	p__Bacillota_A	64.56	5.452	3.74
EHA00659_bin.34	d__Bacteria	p__Bacillota_A	55.04	5.906	2.58
EHA00644_bin.4	d__Bacteria	p__Bacillota_A	98.65	0.978	3.93
EHA01415_bin.5	d__Bacteria	p__Bacillota_A	90.28	1.403	4.42
EHA00660_bin.35	d__Bacteria	p__Bacillota_A	81.37	1.331	3.73
EHA00669_bin.21	d__Bacteria	p__Bacillota_A	98.65	0.894	4.44
EHA01421_bin.33	d__Bacteria	p__Bacillota_A	96.62	1.006	2.78
EHA01415_bin.50	d__Bacteria	p__Bacillota_A	95.72	2.215	3.70
EHA01404_bin.130	d__Bacteria	p__Bacillota_A	69.14	1.582	2.88
EHA01404_bin.73	d__Bacteria	p__Bacillota_A	59.17	1.002	3.08
EHA01409_bin.64	d__Bacteria	p__Bacillota_A	95.00	0.000	2.25
EHA00661_bin.29	d__Bacteria	p__Bacillota_A	99.20	0.949	4.58
EHA01402_bin.23	d__Bacteria	p__Bacillota_A	70.68	1.898	3.21
EHA01415_bin.11	d__Bacteria	p__Bacillota_A	53.13	7.576	2.47
EHA00658_bin.17	d__Bacteria	p__Bacillota_A	51.64	7.336	3.62
EHA01421_bin.61	d__Bacteria	p__Bacillota_A	50.48	9.090	2.26
EHA01421_bin.54	d__Bacteria	p__Bacillota_A	95.46	2.589	4.96
EHA01415_bin.36	d__Bacteria	p__Bacillota_A	87.08	3.711	4.42
EHA01426_bin.57	d__Bacteria	p__Bacillota_A	62.06	3.164	3.07
EHA00656_bin.11	d__Bacteria	p__Bacillota_A	50.23	2.426	2.29
EHA01426_bin.83	d__Bacteria	p__Bacillota_A	73.40	1.376	3.00
EHA01415_bin.60	d__Bacteria	p__Bacillota_A	89.60	1.112	3.36
EHA00652_bin.2	d__Bacteria	p__Bacillota_A	54.15	3.596	3.85
EHA00651_bin.35	d__Bacteria	p__Bacillota_A	50.09	8.245	3.91
EHA01402_bin.22	d__Bacteria	p__Bacillota_A	74.80	2.004	3.85
EHA00657_bin.20	d__Bacteria	p__Bacillota_A	50.07	9.441	5.51
EHA01404_bin.54	d__Bacteria	p__Bacillota_A	55.96	2.109	3.13
EHA00660_bin.28	d__Bacteria	p__Bacillota_A	97.13	0.632	4.25
EHA00647_bin.5	d__Bacteria	p__Bacillota_A	70.37	1.740	2.99
EHA00659_bin.19	d__Bacteria	p__Bacillota_A	97.23	0.569	4.60
EHA01426_bin.1	d__Bacteria	p__Bacillota_A	77.01	2.167	3.20

genome	domain	phylum	comp	cont	size
EHA00654_bin.10	d__Bacteria	p__Bacillota_A	98.06	0.474	6.05
EHA00650_bin.10	d__Bacteria	p__Bacillota_A	69.66	0.690	3.95
EHA01402_bin.7	d__Bacteria	p__Bacillota_A	73.32	1.002	3.88
EHA00651_bin.38	d__Bacteria	p__Bacillota_A	51.21	3.968	2.85
EHA01426_bin.46	d__Bacteria	p__Bacillota_A	79.47	1.898	3.19
EHA01404_bin.114	d__Bacteria	p__Bacillota_A	53.27	0.738	1.86
EHA00652_bin.14	d__Bacteria	p__Bacillota_A	67.28	0.632	3.08
EHA00650_bin.5	d__Bacteria	p__Bacillota_A	60.02	6.175	2.01
EHA00654_bin.3	d__Bacteria	p__Bacillota_A	98.00	0.000	4.40
EHA00651_bin.21	d__Bacteria	p__Bacillota_A	98.10	1.898	4.21
EHA00647_bin.3	d__Bacteria	p__Bacillota_A	58.53	0.000	2.50
EHA01409_bin.103	d__Bacteria	p__Bacillota_A	61.22	6.043	2.65
EHA00671_bin.8	d__Bacteria	p__Bacillota_A	96.43	1.371	4.15
EHA00646_bin.10	d__Bacteria	p__Bacillota_A	99.36	0.019	3.89
EHA00654_bin.24	d__Bacteria	p__Bacillota_A	56.31	1.139	2.17
EHA00648_bin.12	d__Bacteria	p__Bacillota_A	56.04	2.647	1.85
EHA01421_bin.16	d__Bacteria	p__Bacillota_A	80.46	0.574	3.30
EHA01415_bin.63	d__Bacteria	p__Bacillota_A	80.16	1.848	3.23
EHA01421_bin.64	d__Bacteria	p__Bacillota_A	97.96	1.297	4.06
EHA01426_bin.23	d__Bacteria	p__Bacillota_A	95.31	2.490	3.79
EHA00660_bin.3	d__Bacteria	p__Bacillota_A	97.01	0.000	4.60
EHA00656_bin.15	d__Bacteria	p__Bacillota_A	97.70	1.393	3.84
EHA01404_bin.118	d__Bacteria	p__Bacillota_A	74.64	3.793	2.64
EHA00669_bin.25	d__Bacteria	p__Bacillota_A	82.99	0.722	3.12
EHA01404_bin.83	d__Bacteria	p__Bacillota_A	95.19	1.166	3.89
EHA00658_bin.22	d__Bacteria	p__Bacillota_A	95.80	0.646	3.60
EHA00666_bin.35	d__Bacteria	p__Bacillota_A	94.35	0.000	1.68
EHA00656_bin.4	d__Bacteria	p__Bacillota_A	64.11	0.000	3.08
EHA00661_bin.13	d__Bacteria	p__Bacillota_A	61.10	9.142	2.19
EHA00647_bin.4	d__Bacteria	p__Bacillota_A	96.55	1.997	4.23
EHA01404_bin.60	d__Bacteria	p__Bacillota_A	60.20	3.913	2.08
EHA01421_bin.81	d__Bacteria	p__Bacillota_A	98.25	0.771	2.74
EHA01402_bin.14	d__Bacteria	p__Bacillota_A	96.08	0.944	3.08
EHA01426_bin.41	d__Bacteria	p__Bacillota_A	96.30	0.420	3.56
EHA00664_bin.5	d__Bacteria	p__Bacillota_A	97.49	1.480	3.33
EHA01421_bin.19	d__Bacteria	p__Bacillota_A	96.35	0.752	3.63
EHA00668_bin.10	d__Bacteria	p__Bacillota_A	85.42	1.825	3.27
EHA01404_bin.57	d__Bacteria	p__Bacillota_A	98.27	0.574	3.16
EHA00667_bin.4	d__Bacteria	p__Bacillota_A	83.73	1.807	2.92
EHA00641_bin.4	d__Bacteria	p__Bacillota_A	98.27	0.574	3.73
EHA00667_bin.14	d__Bacteria	p__Bacillota_A	52.79	0.000	2.56
EHA00663_bin.4	d__Bacteria	p__Bacillota_A	98.27	0.000	4.31
EHA00655_bin.7	d__Bacteria	p__Bacillota_A	98.27	0.574	4.08
EHA01409_bin.40	d__Bacteria	p__Bacillota_A	89.02	1.677	2.13
EHA00668_bin.4	d__Bacteria	p__Bacillota_A	50.28	5.620	3.06
EHA01409_bin.107	d__Bacteria	p__Bacillota_A	55.02	6.828	1.72
EHA00651_bin.9	d__Bacteria	p__Bacillota_A	99.05	1.582	4.54
EHA01426_bin.109	d__Bacteria	p__Bacillota_A	50.60	1.505	2.59
EHA01402_bin.25	d__Bacteria	p__Bacillota_A	91.36	2.536	3.33
EHA00659_bin.33	d__Bacteria	p__Bacillota_A	95.65	1.207	3.61
EHA01404_bin.121	d__Bacteria	p__Bacillota_A	97.04	1.175	3.73
EHA01426_bin.29	d__Bacteria	p__Bacillota_A	86.61	1.630	3.22

genome	domain	phylum	comp	cont	size
EHA00646_bin.4	d__Bacteria	p__Bacillota_A	51.82	9.234	3.76
EHA01421_bin.43	d__Bacteria	p__Bacillota_A	92.07	0.057	4.19
EHA01404_bin.30	d__Bacteria	p__Bacillota_A	97.42	1.476	3.95
EHA01426_bin.12	d__Bacteria	p__Bacillota_A	73.70	0.632	2.93
EHA00666_bin.7	d__Bacteria	p__Bacillota_A	100.00	0.000	0.91
EHA01421_bin.93	d__Bacteria	p__Bacillota_A	98.10	0.210	3.42
EHA00667_bin.20	d__Bacteria	p__Bacillota_A	53.77	5.263	2.50
EHA01426_bin.79	d__Bacteria	p__Bacillota_A	96.51	1.265	3.78
EHA01421_bin.70	d__Bacteria	p__Bacillota_A	86.49	1.647	4.93
EHA01404_bin.131	d__Bacteria	p__Bacillota_A	94.15	0.343	3.23
EHA00657_bin.16	d__Bacteria	p__Bacillota_A	97.17	1.481	4.65
EHA01404_bin.77	d__Bacteria	p__Bacillota_A	92.39	0.292	3.53
EHA00645_bin.24	d__Bacteria	p__Bacillota_A	78.32	0.591	2.71
EHA01409_bin.11	d__Bacteria	p__Bacillota_A	93.62	0.936	2.11
EHA00656_bin.14	d__Bacteria	p__Bacillota_A	83.90	0.061	2.51
EHA00651_bin.37	d__Bacteria	p__Bacillota_A	98.65	0.518	3.71
EHA01421_bin.27	d__Bacteria	p__Bacillota_A	98.65	0.671	3.30
EHA00656_bin.3	d__Bacteria	p__Bacillota_A	98.65	1.769	3.73
EHA01415_bin.9	d__Bacteria	p__Bacillota_A	96.67	1.006	3.33
EHA01415_bin.47	d__Bacteria	p__Bacillota_A	56.52	0.671	1.41
EHA00652_bin.8	d__Bacteria	p__Bacillota_A	60.87	0.000	1.73
EHA01404_bin.106	d__Bacteria	p__Bacillota_A	94.11	2.263	2.42
EHA01426_bin.6	d__Bacteria	p__Bacillota_A	92.36	0.894	2.30
EHA01415_bin.24	d__Bacteria	p__Bacillota_A	96.64	0.223	2.43
EHA00655_bin.12	d__Bacteria	p__Bacillota_A	63.77	0.167	1.41
EHA00651_bin.25	d__Bacteria	p__Bacillota_A	98.65	0.305	3.60
EHA01404_bin.71	d__Bacteria	p__Bacillota_A	75.19	2.593	1.58
EHA00650_bin.2	d__Bacteria	p__Bacillota_A	71.67	1.140	1.58
EHA00659_bin.14	d__Bacteria	p__Bacillota_A	77.78	0.671	1.69
EHA01404_bin.96	d__Bacteria	p__Bacillota_A	63.46	1.588	1.33
EHA01409_bin.2	d__Bacteria	p__Bacillota_A	87.91	3.440	5.13
EHA01409_bin.33	d__Bacteria	p__Bacillota_A	60.00	0.000	1.05
EHA01409_bin.106	d__Bacteria	p__Bacillota_A	67.01	3.205	1.49
EHA01409_bin.52	d__Bacteria	p__Bacillota_A	94.11	1.307	1.59
EHA01426_bin.8	d__Bacteria	p__Bacillota_A	98.65	0.000	2.17
EHA01409_bin.71	d__Bacteria	p__Bacillota_A	89.93	0.692	1.77
EHA01404_bin.64	d__Bacteria	p__Bacillota_A	79.24	0.356	1.81
EHA00660_bin.22	d__Bacteria	p__Bacillota_A	60.38	1.565	1.91
EHA00660_bin.16	d__Bacteria	p__Bacillota_A	80.27	0.559	2.23
EHA01415_bin.79	d__Bacteria	p__Bacillota_A	96.19	0.000	3.76
EHA01404_bin.9	d__Bacteria	p__Bacillota_A	95.41	0.000	2.49
EHA01404_bin.53	d__Bacteria	p__Bacillota_A	92.30	1.466	1.98
EHA01404_bin.124	d__Bacteria	p__Bacillota_A	84.03	0.000	1.98
EHA00661_bin.33	d__Bacteria	p__Bacillota_A	97.98	0.000	2.37
EHA01404_bin.99	d__Bacteria	p__Bacillota_A	65.58	0.671	1.61
EHA01409_bin.47	d__Bacteria	p__Bacillota_A	94.40	2.197	1.42
EHA01404_bin.27	d__Bacteria	p__Bacillota_A	60.24	0.671	1.19
EHA01426_bin.91	d__Bacteria	p__Bacillota_A	88.75	0.894	1.60
EHA01421_bin.31	d__Bacteria	p__Bacillota_A	58.07	0.000	1.62
EHA01415_bin.85	d__Bacteria	p__Bacillota_A	78.89	2.684	2.93
EHA01415_bin.100	d__Bacteria	p__Bacillota_A	85.06	3.131	3.13
EHA01415_bin.1	d__Bacteria	p__Bacillota_A	95.56	2.969	3.65

genome	domain	phylum	comp	cont	size
EHA01409_bin.6	d__Bacteria	p__Bacillota_A	71.10	3.355	1.37
EHA01404_bin.10	d__Bacteria	p__Bacillota_A	95.08	0.000	2.56
EHA00641_bin.16	d__Bacteria	p__Bacillota_A	97.27	0.865	3.10
EHA01409_bin.17	d__Bacteria	p__Bacillota_A	82.65	1.643	2.20
EHA01402_bin.39	d__Bacteria	p__Bacillota_A	50.70	2.777	1.57
EHA01421_bin.95	d__Bacteria	p__Bacillota_A	55.07	9.619	2.17
EHA01404_bin.11	d__Bacteria	p__Bacillota_A	81.53	1.342	2.10
EHA00666_bin.16	d__Bacteria	p__Bacillota_A	97.98	0.671	2.24
EHA00650_bin.9	d__Bacteria	p__Bacillota_A	98.65	0.335	2.32
EHA00669_bin.6	d__Bacteria	p__Bacillota_A	97.98	0.447	2.15
EHA01402_bin.27	d__Bacteria	p__Bacillota_A	99.32	0.000	2.16
EHA01409_bin.20	d__Bacteria	p__Bacillota_A	58.78	2.762	1.04
EHA01404_bin.103	d__Bacteria	p__Bacillota_A	97.97	1.006	2.58
EHA01404_bin.81	d__Bacteria	p__Bacillota_A	71.14	0.894	2.13
EHA01404_bin.105	d__Bacteria	p__Bacillota_A	58.11	1.476	1.39
EHA01409_bin.75	d__Bacteria	p__Bacillota_A	83.17	0.645	2.58
EHA01409_bin.21	d__Bacteria	p__Bacillota_A	80.89	1.230	1.42
EHA01404_bin.126	d__Bacteria	p__Bacillota_A	64.24	0.671	1.48
EHA01426_bin.3	d__Bacteria	p__Bacillota_A	84.54	4.276	2.00
EHA00658_bin.10	d__Bacteria	p__Bacillota_A	92.38	0.167	2.95
EHA00659_bin.17	d__Bacteria	p__Bacillota_A	88.84	1.342	2.43
EHA00656_bin.12	d__Bacteria	p__Bacillota_A	82.07	0.406	2.75
EHA01415_bin.74	d__Bacteria	p__Bacillota_A	54.46	1.140	1.85
EHA01404_bin.98	d__Bacteria	p__Bacillota_A	50.86	0.000	1.35
EHA01415_bin.97	d__Bacteria	p__Bacillota_A	62.87	0.000	1.78
EHA01415_bin.19	d__Bacteria	p__Bacillota_A	56.69	7.389	2.16
EHA01426_bin.106	d__Bacteria	p__Bacillota_A	52.24	5.862	1.72
EHA01421_bin.65	d__Bacteria	p__Bacillota_A	85.71	3.691	2.40
EHA01404_bin.13	d__Bacteria	p__Bacillota_A	75.57	3.173	1.79
EHA01415_bin.77	d__Bacteria	p__Bacillota_A	80.86	2.125	2.46
EHA01404_bin.137	d__Bacteria	p__Bacillota_A	55.43	2.796	1.29
EHA01415_bin.3	d__Bacteria	p__Bacillota_A	55.16	8.144	1.67
EHA00643_bin.20	d__Bacteria	p__Bacillota_A	85.23	1.342	2.50
EHA01421_bin.60	d__Bacteria	p__Bacillota_A	73.02	0.782	1.60
EHA01426_bin.60	d__Bacteria	p__Bacillota_A	70.59	1.006	1.91
EHA01409_bin.27	d__Bacteria	p__Bacillota_A	51.01	0.000	2.98
EHA01409_bin.57	d__Bacteria	p__Bacillota_A	99.36	0.421	3.92
EHA01404_bin.48	d__Bacteria	p__Bacillota_A	80.49	1.062	1.75
EHA00642_bin.23	d__Bacteria	p__Bacillota_A	87.35	0.671	2.09
EHA01409_bin.50	d__Bacteria	p__Synergistota	89.33	0.777	1.63
EHA00667_bin.5	d__Bacteria	p__Fusobacteriota	72.86	2.247	2.01
EHA00652_bin.4	d__Bacteria	p__Fusobacteriota	71.18	1.123	2.25
EHA00670_bin.11	d__Bacteria	p__Fusobacteriota	55.97	1.123	1.57
EHA00672_bin.2	d__Bacteria	p__Fusobacteriota	100.00	2.247	3.67
EHA00643_bin.1	d__Bacteria	p__Fusobacteriota	62.24	0.000	2.85
EHA01426_bin.85	d__Bacteria	p__Desulfobacterota	90.72	0.810	2.49
EHA01421_bin.48	d__Bacteria	p__Desulfobacterota	97.63	1.808	3.99
EHA00659_bin.25	d__Bacteria	p__Desulfobacterota	64.41	2.103	2.28
EHA00675_bin.3	d__Bacteria	p__Desulfobacterota	53.47	4.032	1.80
EHA00663_bin.12	d__Bacteria	p__Desulfobacterota	50.10	5.303	2.39
EHA01415_bin.12	d__Bacteria	p__Desulfobacterota	73.53	8.658	2.58
EHA00645_bin.23	d__Bacteria	p__Desulfobacterota	59.08	2.177	1.46

genome	domain	phylum	comp	cont	size
EHA00646_bin.15	d__Bacteria	p__Desulfobacterota	93.16	0.084	3.18
EHA01409_bin.51	d__Bacteria	p__Desulfobacterota	100.00	1.235	1.50
EHA00652_bin.1	d__Bacteria	p__Desulfobacterota	90.18	0.613	3.35
EHA01402_bin.33	d__Bacteria	p__Desulfobacterota	61.39	0.153	2.43
EHA00654_bin.8	d__Bacteria	p__Desulfobacterota	91.32	1.183	2.95
EHA00666_bin.18	d__Bacteria	p__Desulfobacterota	81.12	5.870	3.14
EHA01404_bin.46	d__Bacteria	p__Desulfobacterota	62.88	3.763	1.80
EHA00651_bin.20	d__Bacteria	p__Desulfobacterota	83.11	0.021	2.81
EHA00657_bin.14	d__Bacteria	p__Desulfobacterota	90.41	0.065	3.21
EHA01426_bin.96	d__Bacteria	p__Desulfobacterota	60.81	1.479	2.15
EHA00660_bin.10	d__Bacteria	p__Desulfobacterota	56.89	2.558	2.64
EHA00669_bin.26	d__Bacteria	p__Desulfobacterota	94.55	0.657	3.32
EHA00658_bin.11	d__Bacteria	p__Desulfobacterota	73.71	0.799	2.70
EHA01426_bin.39	d__Bacteria	p__Desulfobacterota	59.51	1.205	2.05
EHA01421_bin.45	d__Bacteria	p__Desulfobacterota	82.34	2.073	3.17
EHA00641_bin.11	d__Bacteria	p__Pseudomonadota	100.00	0.000	0.63
EHA01421_bin.92	d__Bacteria	p__Pseudomonadota	76.52	0.000	4.55
EHA00651_bin.31	d__Bacteria	p__Pseudomonadota	99.80	0.093	5.05
EHA01426_bin.53	d__Bacteria	p__Pseudomonadota	59.98	2.189	2.75
EHA00646_bin.22	d__Bacteria	p__Pseudomonadota	53.01	9.499	2.70
EHA01415_bin.45	d__Bacteria	p__Pseudomonadota	77.44	6.570	3.36
EHA00666_bin.14	d__Bacteria	p__Pseudomonadota	95.67	1.191	4.11
EHA01415_bin.76	d__Bacteria	p__Pseudomonadota	57.39	3.448	3.01
EHA00660_bin.31	d__Bacteria	p__Pseudomonadota	98.74	1.872	4.70
EHA01415_bin.56	d__Bacteria	p__Pseudomonadota	94.56	1.883	3.48
EHA00658_bin.28	d__Bacteria	p__Pseudomonadota	95.78	0.949	3.93
EHA01426_bin.31	d__Bacteria	p__Pseudomonadota	84.96	1.693	3.16
EHA00645_bin.8	d__Bacteria	p__Pseudomonadota	95.70	0.369	3.04
EHA01409_bin.88	d__Bacteria	p__Pseudomonadota	99.23	0.572	1.79
EHA00653_bin.6	d__Bacteria	p__Pseudomonadota	64.75	1.523	2.57
EHA00667_bin.18	d__Bacteria	p__Pseudomonadota	99.43	0.293	3.73
EHA00658_bin.29	d__Bacteria	p__Pseudomonadota	61.75	2.201	1.76
EHA00648_bin.9	d__Bacteria	p__Pseudomonadota	86.54	4.495	1.56
EHA00675_bin.11	d__Bacteria	p__Pseudomonadota	100.00	0.000	4.84
EHA00667_bin.11	d__Bacteria	p__Pseudomonadota	88.65	7.354	3.65
EHA00670_bin.17	d__Bacteria	p__Pseudomonadota	78.54	2.082	3.43
EHA01415_bin.106	d__Bacteria	p__Pseudomonadota	87.40	0.164	1.00
EHA00658_bin.4	d__Bacteria	p__Pseudomonadota	94.50	0.000	1.00
EHA00644_bin.22	d__Bacteria	p__Pseudomonadota	94.40	0.000	1.01
EHA00656_bin.6	d__Bacteria	p__Pseudomonadota	93.40	1.098	0.95
EHA00663_bin.2	d__Bacteria	p__Pseudomonadota	73.26	1.098	0.68
EHA01409_bin.45	d__Bacteria	p__Pseudomonadota	98.88	0.656	4.50
EHA00648_bin.4	d__Bacteria	p__Pseudomonadota	68.61	0.097	0.81
EHA00666_bin.8	d__Bacteria	p__Pseudomonadota	99.10	1.342	2.96
EHA00672_bin.3	d__Bacteria	p__Pseudomonadota	70.33	2.150	0.78
EHA00641_bin.7	d__Bacteria	p__Pseudomonadota	91.39	0.000	1.17
EHA01404_bin.89	d__Bacteria	p__Pseudomonadota	81.05	0.000	0.96
EHA01415_bin.41	d__Bacteria	p__Pseudomonadota	87.09	0.126	1.00
EHA01404_bin.28	d__Bacteria	p__Pseudomonadota	90.32	0.000	1.19
EHA01421_bin.62	d__Bacteria	p__Pseudomonadota	65.74	1.289	1.86
EHA01421_bin.47	d__Bacteria	p__Pseudomonadota	61.51	0.777	1.71
EHA01421_bin.42	d__Bacteria	p__Pseudomonadota	98.92	1.075	2.02

genome	domain	phylum	comp	cont	size
EHA00665_bin.7	d__Bacteria	p__Pseudomonadota	76.90	4.185	1.50
EHA01404_bin.88	d__Bacteria	p__Pseudomonadota	95.06	1.173	1.98
EHA00643_bin.29	d__Bacteria	p__Pseudomonadota	88.17	0.000	1.84
EHA00643_bin.11	d__Bacteria	p__Pseudomonadota	74.01	0.097	1.24
EHA01404_bin.6	d__Bacteria	p__Pseudomonadota	71.16	8.797	1.47
EHA00647_bin.6	d__Bacteria	p__Pseudomonadota	94.62	0.537	1.68
EHA01421_bin.53	d__Bacteria	p__Pseudomonadota	98.92	0.000	1.76
EHA01404_bin.123	d__Bacteria	p__Pseudomonadota	64.61	0.000	0.98
EHA01404_bin.16	d__Bacteria	p__Pseudomonadota	77.67	0.000	1.23
EHA01421_bin.59	d__Bacteria	p__Pseudomonadota	98.92	0.000	2.06
EHA01426_bin.74	d__Bacteria	p__Pseudomonadota	92.47	0.097	1.70
EHA00652_bin.12	d__Bacteria	p__Pseudomonadota	96.77	0.000	1.96
EHA01404_bin.95	d__Bacteria	p__Deferribacterota	73.13	0.862	1.64
EHA01426_bin.59	d__Bacteria	p__Campylobacterota	98.87	0.352	1.56
EHA00662_bin.23	d__Bacteria	p__Campylobacterota	98.34	0.000	1.54
EHA00645_bin.19	d__Bacteria	p__Campylobacterota	97.97	0.768	4.31
EHA00667_bin.3	d__Bacteria	p__Campylobacterota	86.05	5.563	1.33
EHA00668_bin.5	d__Bacteria	p__Campylobacterota	93.41	2.095	1.58
EHA00669_bin.11	d__Bacteria	p__Campylobacterota	91.57	0.578	1.44
EHA00659_bin.30	d__Bacteria	p__Campylobacterota	99.41	0.000	1.70
EHA01415_bin.91	d__Bacteria	p__Bacteroidota	83.78	1.257	3.03
EHA00643_bin.26	d__Bacteria	p__Bacteroidota	98.86	0.000	3.73
EHA01426_bin.19	d__Bacteria	p__Bacteroidota	75.50	0.830	2.84
EHA01409_bin.38	d__Bacteria	p__Bacteroidota	100.00	0.146	4.80
EHA01404_bin.58	d__Bacteria	p__Bacteroidota	95.45	2.628	3.53
EHA00667_bin.6	d__Bacteria	p__Bacteroidota	54.31	3.448	2.09
EHA01415_bin.26	d__Bacteria	p__Bacteroidota	96.35	0.551	4.94
EHA01409_bin.77	d__Bacteria	p__Bacteroidota	97.50	2.688	4.22
EHA00670_bin.2	d__Bacteria	p__Bacteroidota	99.33	2.115	3.79
EHA00659_bin.35	d__Bacteria	p__Bacteroidota	95.76	0.705	4.40
EHA01426_bin.68	d__Bacteria	p__Bacteroidota	62.76	0.192	2.57
EHA00669_bin.16	d__Bacteria	p__Bacteroidota	98.24	0.128	4.78
EHA00650_bin.14	d__Bacteria	p__Bacteroidota	52.81	4.297	2.49
EHA00671_bin.5	d__Bacteria	p__Bacteroidota	93.39	0.833	4.17
EHA00657_bin.1	d__Bacteria	p__Bacteroidota	99.10	0.384	5.01
EHA01402_bin.18	d__Bacteria	p__Bacteroidota	82.78	3.076	3.57
EHA00666_bin.15	d__Bacteria	p__Bacteroidota	67.32	2.458	2.70
EHA01409_bin.80	d__Bacteria	p__Bacteroidota	50.86	1.612	2.40
EHA00648_bin.5	d__Bacteria	p__Bacteroidota	71.19	3.152	2.92
EHA00643_bin.19	d__Bacteria	p__Bacteroidota	98.50	1.923	5.26
EHA00645_bin.5	d__Bacteria	p__Bacteroidota	55.02	0.952	1.17
EHA00669_bin.5	d__Bacteria	p__Bacteroidota	76.36	9.482	4.05
EHA01409_bin.58	d__Bacteria	p__Bacteroidota	98.66	0.111	1.97
EHA01402_bin.38	d__Bacteria	p__Bacteroidota	65.84	1.724	3.50
EHA00654_bin.9	d__Bacteria	p__Bacteroidota	98.84	1.176	4.75
EHA01404_bin.70	d__Bacteria	p__Bacteroidota	95.93	0.576	4.07
EHA01415_bin.102	d__Bacteria	p__Bacteroidota	97.58	0.769	4.27
EHA01409_bin.91	d__Bacteria	p__Bacteroidota	90.83	0.026	1.67
EHA01421_bin.80	d__Bacteria	p__Bacteroidota	73.98	1.826	4.22
EHA01404_bin.35	d__Bacteria	p__Bacteroidota	92.56	0.641	4.53
EHA00657_bin.9	d__Bacteria	p__Bacteroidota	99.61	0.641	4.17
EHA01426_bin.45	d__Bacteria	p__Bacteroidota	86.30	3.397	3.27

genome	domain	phylum	comp	cont	size
EHA00657_bin.8	d__Bacteria	p__Bacteroidota	99.35	0.000	4.81
EHA00644_bin.20	d__Bacteria	p__Bacteroidota	94.79	1.775	5.51
EHA00671_bin.21	d__Bacteria	p__Bacteroidota	97.28	1.923	4.91
EHA01409_bin.53	d__Bacteria	p__Bacteroidota	64.12	0.156	2.10
EHA01409_bin.36	d__Bacteria	p__Bacteroidota	68.51	0.789	2.07
EHA00660_bin.23	d__Bacteria	p__Bacteroidota	91.23	0.000	5.95
EHA00651_bin.14	d__Bacteria	p__Bacteroidota	93.87	0.192	5.68
EHA00667_bin.19	d__Bacteria	p__Bacteroidota	55.95	6.034	4.15
EHA01409_bin.9	d__Bacteria	p__Bacteroidota	50.86	6.896	3.06
EHA00674_bin.1	d__Bacteria	p__Bacteroidota	60.00	5.000	3.02
EHA00661_bin.7	d__Bacteria	p__Bacteroidota	93.48	2.564	4.27
EHA00671_bin.25	d__Bacteria	p__Bacteroidota	50.31	8.777	3.51
EHA01402_bin.20	d__Bacteria	p__Bacteroidota	57.93	3.793	4.12
EHA00651_bin.17	d__Bacteria	p__Bacteroidota	96.34	0.705	6.52
EHA00641_bin.6	d__Bacteria	p__Bacteroidota	85.52	1.025	4.95
EHA01426_bin.14	d__Bacteria	p__Bacteroidota	82.81	6.282	5.01
EHA00647_bin.7	d__Bacteria	p__Bacteroidota	77.29	0.769	5.05
EHA01421_bin.67	d__Bacteria	p__Bacteroidota	72.72	1.794	4.37
EHA01409_bin.86	d__Bacteria	p__Bacteroidota	96.87	4.291	1.97
EHA01404_bin.135	d__Bacteria	p__Bacteroidota	87.04	3.039	4.56
EHA01404_bin.101	d__Bacteria	p__Bacteroidota	65.00	0.000	4.18
EHA01409_bin.90	d__Bacteria	p__Bacteroidota	65.59	2.111	2.81
EHA01426_bin.75	d__Bacteria	p__Bacteroidota	79.65	0.000	4.80
EHA01421_bin.22	d__Bacteria	p__Bacteroidota	64.13	0.000	4.10
EHA01426_bin.54	d__Bacteria	p__Bacteroidota	62.02	0.000	4.87
EHA01409_bin.10	d__Bacteria	p__Bacteroidota	65.34	6.896	4.33
EHA01404_bin.49	d__Bacteria	p__Bacteroidota	77.93	1.724	5.13
EHA00645_bin.20	d__Bacteria	p__Bacteroidota	95.00	0.000	2.34
EHA00642_bin.19	d__Bacteria	p__Bacteroidota	52.58	8.620	3.67
EHA00660_bin.26	d__Bacteria	p__Bacteroidota	80.48	3.270	4.40
EHA00650_bin.16	d__Bacteria	p__Bacteroidota	90.63	2.219	5.07
EHA00669_bin.23	d__Bacteria	p__Bacteroidota	87.98	4.228	5.35
EHA00654_bin.15	d__Bacteria	p__Bacteroidota	80.03	1.839	4.39
EHA00672_bin.18	d__Bacteria	p__Bacteroidota	85.68	7.367	5.66
EHA00670_bin.16	d__Bacteria	p__Bacteroidota	95.44	2.939	5.54
EHA01421_bin.83	d__Bacteria	p__Bacteroidota	75.68	1.724	4.32
EHA00671_bin.4	d__Bacteria	p__Bacteroidota	59.01	3.184	3.34
EHA01415_bin.39	d__Bacteria	p__Bacteroidota	51.72	2.586	4.40
EHA01415_bin.93	d__Bacteria	p__Bacteroidota	67.58	4.310	5.14
EHA00648_bin.13	d__Bacteria	p__Bacteroidota	96.77	1.617	3.73
EHA00669_bin.3	d__Bacteria	p__Bacteroidota	99.07	2.881	4.44
EHA00657_bin.7	d__Bacteria	p__Bacteroidota	96.37	2.695	5.64
EHA00650_bin.7	d__Bacteria	p__Bacteroidota	95.94	0.929	4.99
EHA00671_bin.13	d__Bacteria	p__Bacteroidota	93.58	2.583	4.87
EHA00651_bin.32	d__Bacteria	p__Bacteroidota	98.88	0.793	5.37
EHA00646_bin.3	d__Bacteria	p__Bacteroidota	95.62	0.691	5.14
EHA01409_bin.56	d__Bacteria	p__Bacteroidota	66.30	6.896	4.65
EHA01409_bin.87	d__Bacteria	p__Bacteroidota	97.58	0.000	1.38
EHA01404_bin.52	d__Bacteria	p__Bacteroidota	87.45	1.951	4.83
EHA01404_bin.90	d__Bacteria	p__Bacteroidota	69.85	3.955	5.03
EHA00652_bin.18	d__Bacteria	p__Bacteroidota	96.28	1.140	6.86
EHA00646_bin.11	d__Bacteria	p__Bacteroidota	92.53	1.865	6.61

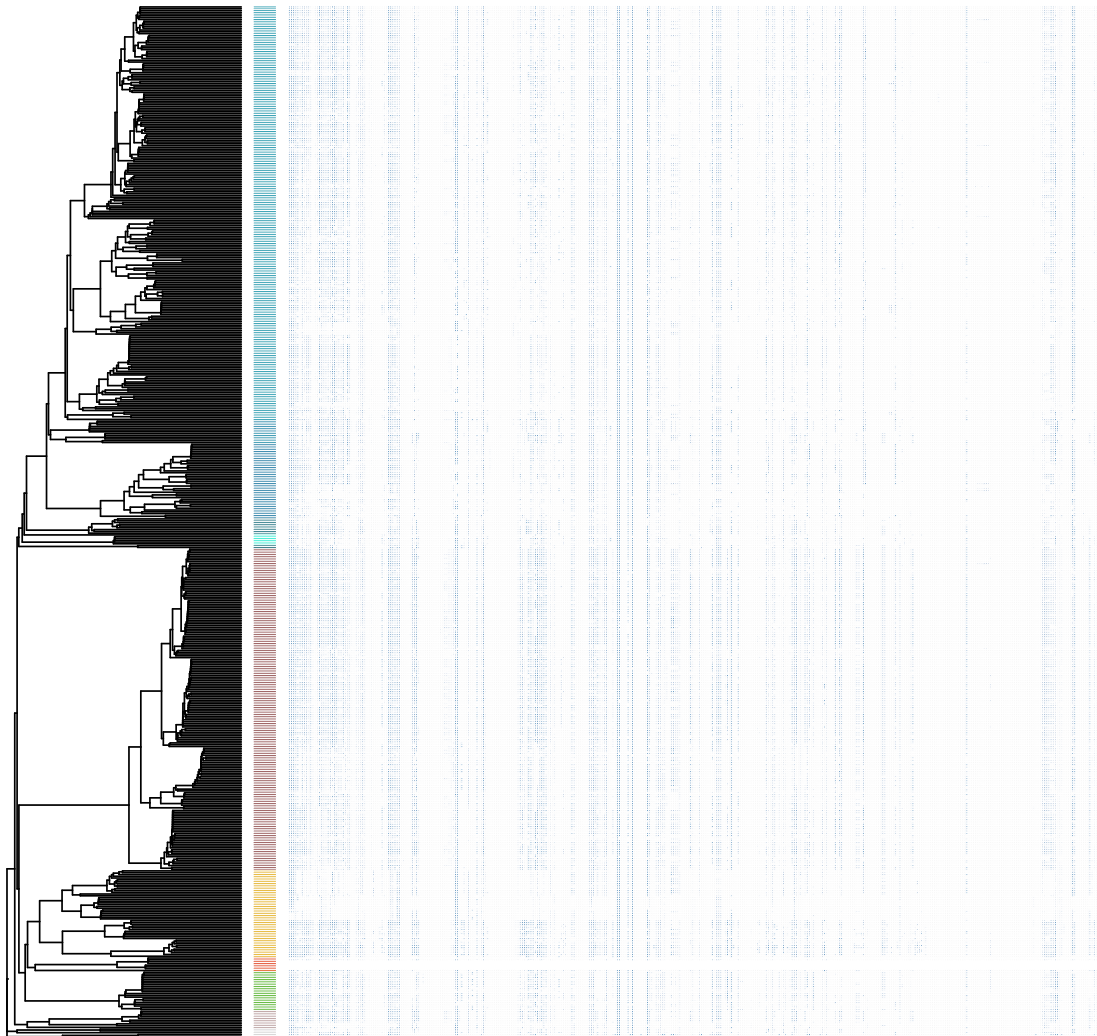
genome	domain	phylum	comp	cont	size
EHA01426_bin.32	d__Bacteria	p__Bacteroidota	61.59	5.516	3.86
EHA01421_bin.79	d__Bacteria	p__Bacteroidota	71.37	1.724	5.48
EHA01415_bin.83	d__Bacteria	p__Bacteroidota	50.69	8.662	2.89
EHA00655_bin.1	d__Bacteria	p__Bacteroidota	85.69	1.425	3.06
EHA00657_bin.18	d__Bacteria	p__Bacteroidota	88.57	2.137	4.10
EHA00668_bin.8	d__Bacteria	p__Bacteroidota	95.07	1.790	4.31
EHA00651_bin.29	d__Bacteria	p__Bacteroidota	93.95	0.805	4.17
EHA01415_bin.35	d__Bacteria	p__Bacteroidota	85.09	0.805	3.28
EHA01421_bin.10	d__Bacteria	p__Bacteroidota	74.68	2.524	2.82
EHA01415_bin.40	d__Bacteria	p__Bacteroidota	78.79	0.000	5.15
EHA01404_bin.37	d__Bacteria	p__Bacteroidota	97.66	0.185	4.69
EHA00659_bin.42	d__Bacteria	p__Bacteroidota	60.84	7.128	4.69
EHA01415_bin.51	d__Bacteria	p__Bacteroidota	51.01	1.724	3.17
EHA01426_bin.102	d__Bacteria	p__Bacteroidota	60.76	2.881	2.76
EHA00646_bin.9	d__Bacteria	p__Bacteroidota	95.93	2.044	5.63
EHA01404_bin.100	d__Bacteria	p__Bacteroidota	95.35	2.509	4.92
EHA01415_bin.44	d__Bacteria	p__Bacteroidota	98.18	0.877	2.71
EHA00643_bin.24	d__Bacteria	p__Bacteroidota	76.97	0.952	2.13
EHA01426_bin.25	d__Bacteria	p__Bacteroidota	83.55	1.154	2.35
EHA01421_bin.24	d__Bacteria	p__Bacteroidota	70.20	1.203	2.39
EHA01409_bin.105	d__Bacteria	p__Bacteroidota	61.95	3.859	3.37
EHA01415_bin.104	d__Bacteria	p__Bacteroidota	51.72	7.758	3.15
EHA00655_bin.10	d__Bacteria	p__Bacteroidota	97.58	0.185	5.68
EHA00659_bin.23	d__Bacteria	p__Bacteroidota	97.44	3.655	6.04
EHA00670_bin.12	d__Bacteria	p__Bacteroidota	97.39	2.602	6.10
EHA00652_bin.13	d__Bacteria	p__Bacteroidota	99.25	2.973	6.14
EHA01426_bin.108	d__Bacteria	p__Bacteroidota	56.69	1.724	4.71
EHA01409_bin.26	d__Bacteria	p__Bacteroidota	97.33	2.898	1.85
EHA00651_bin.41	d__Bacteria	p__Bacteroidota	94.15	0.961	2.56
EHA01415_bin.95	d__Bacteria	p__Bacteroidota	81.45	3.605	2.17
EHA00672_bin.17	d__Bacteria	p__Bacteroidota	98.15	1.442	2.80
EHA00657_bin.6	d__Bacteria	p__Bacteroidota	84.85	1.362	2.55
EHA00664_bin.18	d__Bacteria	p__Bacteroidota	98.55	1.095	2.30
EHA00667_bin.8	d__Bacteria	p__Bacteroidota	93.10	3.125	2.50
EHA00643_bin.23	d__Bacteria	p__Bacteroidota	85.41	0.576	2.33
EHA00647_bin.1	d__Bacteria	p__Bacteroidota	72.12	1.121	1.75
EHA00648_bin.7	d__Bacteria	p__Bacteroidota	94.11	0.653	1.73
EHA00654_bin.18	d__Bacteria	p__Bacteroidota	99.51	0.480	2.36
EHA00646_bin.7	d__Bacteria	p__Bacteroidota	54.34	5.172	1.48
EHA00659_bin.7	d__Bacteria	p__Bacteroidota	91.68	3.044	2.58
EHA01421_bin.69	d__Bacteria	p__Bacteroidota	87.75	3.888	2.94
EHA00658_bin.30	d__Bacteria	p__Bacteroidota	69.26	3.846	2.33
EHA00664_bin.8	d__Bacteria	p__Bacteroidota	51.85	8.538	2.09
EHA01426_bin.104	d__Bacteria	p__Bacteroidota	56.84	1.724	2.54
EHA01409_bin.8	d__Bacteria	p__Bacteroidota	68.43	5.435	2.89
EHA01404_bin.91	d__Bacteria	p__Bacteroidota	91.34	1.522	2.82
EHA00667_bin.12	d__Bacteria	p__Bacteroidota	81.26	1.602	2.87
EHA00641_bin.5	d__Bacteria	p__Bacteroidota	61.28	7.366	2.21
EHA01409_bin.92	d__Bacteria	p__Bacteroidota	57.34	8.076	3.28
EHA00651_bin.40	d__Bacteria	p__Bacteroidota	93.75	0.480	2.17
EHA01426_bin.101	d__Bacteria	p__Bacteroidota	95.83	0.961	1.98
EHA01421_bin.78	d__Bacteria	p__Bacteroidota	99.51	0.000	1.98

genome	domain	phylum	comp	cont	size
EHA00651_bin.39	d__Bacteria	p__Bacteroidota	92.48	1.714	2.21
EHA00645_bin.17	d__Bacteria	p__Bacteroidota	92.00	0.782	2.33
EHA00662_bin.13	d__Bacteria	p__Bacteroidota	98.07	0.240	2.43
EHA00666_bin.10	d__Bacteria	p__Bacteroidota	51.01	0.480	1.18
EHA01404_bin.129	d__Bacteria	p__Bacteroidota	79.49	3.862	1.70
EHA00660_bin.19	d__Bacteria	p__Bacteroidota	85.64	0.637	2.17
EHA00659_bin.10	d__Bacteria	p__Bacteroidota	99.28	0.526	2.88
EHA00656_bin.13	d__Bacteria	p__Bacteroidota	99.04	0.159	3.13
EHA01426_bin.103	d__Bacteria	p__Bacteroidota	56.88	3.875	1.48
EHA00670_bin.10	d__Bacteria	p__Bacteroidota	95.00	0.238	3.05
EHA00666_bin.12	d__Bacteria	p__Bacteroidota	95.00	0.000	2.34
EHA00668_bin.11	d__Bacteria	p__Bacteroidota	51.82	8.064	1.94
EHA00659_bin.36	d__Bacteria	p__Bacteroidota	99.46	0.000	3.45
EHA01415_bin.114	d__Bacteria	p__Bacteroidota	58.24	0.000	2.46
EHA01421_bin.84	d__Bacteria	p__Bacteroidota	77.55	2.468	2.95
EHA00660_bin.4	d__Bacteria	p__Bacteroidota	100.00	2.204	3.16
EHA01415_bin.81	d__Bacteria	p__Bacteroidota	55.75	8.440	3.04
EHA01421_bin.18	d__Bacteria	p__Bacteroidota	99.80	0.537	3.61
EHA00653_bin.2	d__Bacteria	p__Bacteroidota	69.36	0.537	3.26
EHA00672_bin.11	d__Bacteria	p__Bacteroidota	52.06	6.896	3.01
EHA00642_bin.21	d__Bacteria	p__Bacteroidota	81.94	1.612	4.17
EHA00659_bin.18	d__Bacteria	p__Bacteroidota	86.18	3.942	4.01
EHA00654_bin.16	d__Bacteria	p__Bacteroidota	90.39	0.537	4.22
EHA01426_bin.21	d__Bacteria	p__Bacteroidota	85.41	5.286	4.13
EHA00651_bin.23	d__Bacteria	p__Bacteroidota	50.94	9.639	3.38
EHA00655_bin.2	d__Bacteria	p__Bacteroidota	100.00	0.000	4.81
EHA01421_bin.66	d__Bacteria	p__Bacteroidota	77.10	2.598	3.38
EHA00668_bin.7	d__Bacteria	p__Bacteroidota	84.73	3.005	3.96
EHA00669_bin.8	d__Bacteria	p__Bacteroidota	94.08	0.146	4.42
EHA00658_bin.7	d__Bacteria	p__Bacteroidota	95.12	0.268	3.03
EHA00660_bin.24	d__Bacteria	p__Bacteroidota	54.31	6.896	2.81
EHA01404_bin.42	d__Bacteria	p__Bacteroidota	90.29	3.064	3.77
EHA01426_bin.77	d__Bacteria	p__Bacteroidota	64.29	0.922	2.78
EHA01415_bin.42	d__Bacteria	p__Bacteroidota	82.04	1.075	3.20
EHA01421_bin.32	d__Bacteria	p__Bacteroidota	50.65	0.862	2.18
EHA00659_bin.13	d__Bacteria	p__Bacteroidota	80.30	0.537	3.76
EHA01409_bin.78	d__Bacteria	p__Bacteroidota	58.98	3.355	2.21
EHA01426_bin.4	d__Bacteria	p__Bacteroidota	53.91	0.000	2.52
EHA00667_bin.16	d__Bacteria	p__Bacteroidota	57.75	0.000	2.00
EHA00659_bin.31	d__Bacteria	p__Bacteroidota	74.58	1.075	2.46
EHA00660_bin.33	d__Bacteria	p__Bacteroidota	98.17	3.942	3.62
EHA00661_bin.20	d__Bacteria	p__Bacteroidota	100.00	0.537	3.83
EHA00643_bin.25	d__Bacteria	p__Bacteroidota	94.98	0.358	3.78
EHA00651_bin.16	d__Bacteria	p__Bacteroidota	78.98	1.164	3.51
EHA00662_bin.7	d__Bacteria	p__Bacteroidota	83.83	1.612	4.32
EHA01409_bin.60	d__Bacteria	p__Verrucomicrobiota	83.23	2.417	4.11
EHA01404_bin.50	d__Bacteria	p__Verrucomicrobiota	74.31	1.486	0.87
EHA01402_bin.19	d__Bacteria	p__Verrucomicrobiota	81.56	1.740	2.05
EHA01409_bin.15	d__Bacteria	p__Verrucomicrobiota	62.67	2.410	2.25
EHA00651_bin.8	d__Bacteria	p__Verrucomicrobiota	97.95	0.000	2.95
EHA00667_bin.2	d__Bacteria	p__Verrucomicrobiota	97.91	0.000	2.99
EHA00660_bin.34	d__Bacteria	p__Verrucomicrobiota	97.95	0.000	3.11

genome	domain	phylum	comp	cont	size
EHA01409_bin.35	d__Bacteria	p__Verrucomicrobiota	91.39	0.000	1.17
EHA00661_bin.17	d__Bacteria	p__Verrucomicrobiota	88.82	1.020	2.64
EHA00641_bin.17	d__Bacteria	p__Chlamydiota	90.20	0.000	1.23
EHA01421_bin.17	d__Bacteria	p__Spirochaetota	96.00	0.000	2.79
EHA00662_bin.10	d__Bacteria	p__Spirochaetota	92.13	1.123	1.62
EHA01426_bin.66	d__Bacteria	p__Spirochaetota	93.25	0.000	1.55
EHA01409_bin.18	d__Archaea	p__Halobacteriota	93.38	0.967	1.81
EHA01409_bin.4	d__Archaea	p__Halobacteriota	99.24	0.501	0.94
EHA00645_bin.14	d__Archaea	p__Halobacteriota	98.66	0.000	1.84

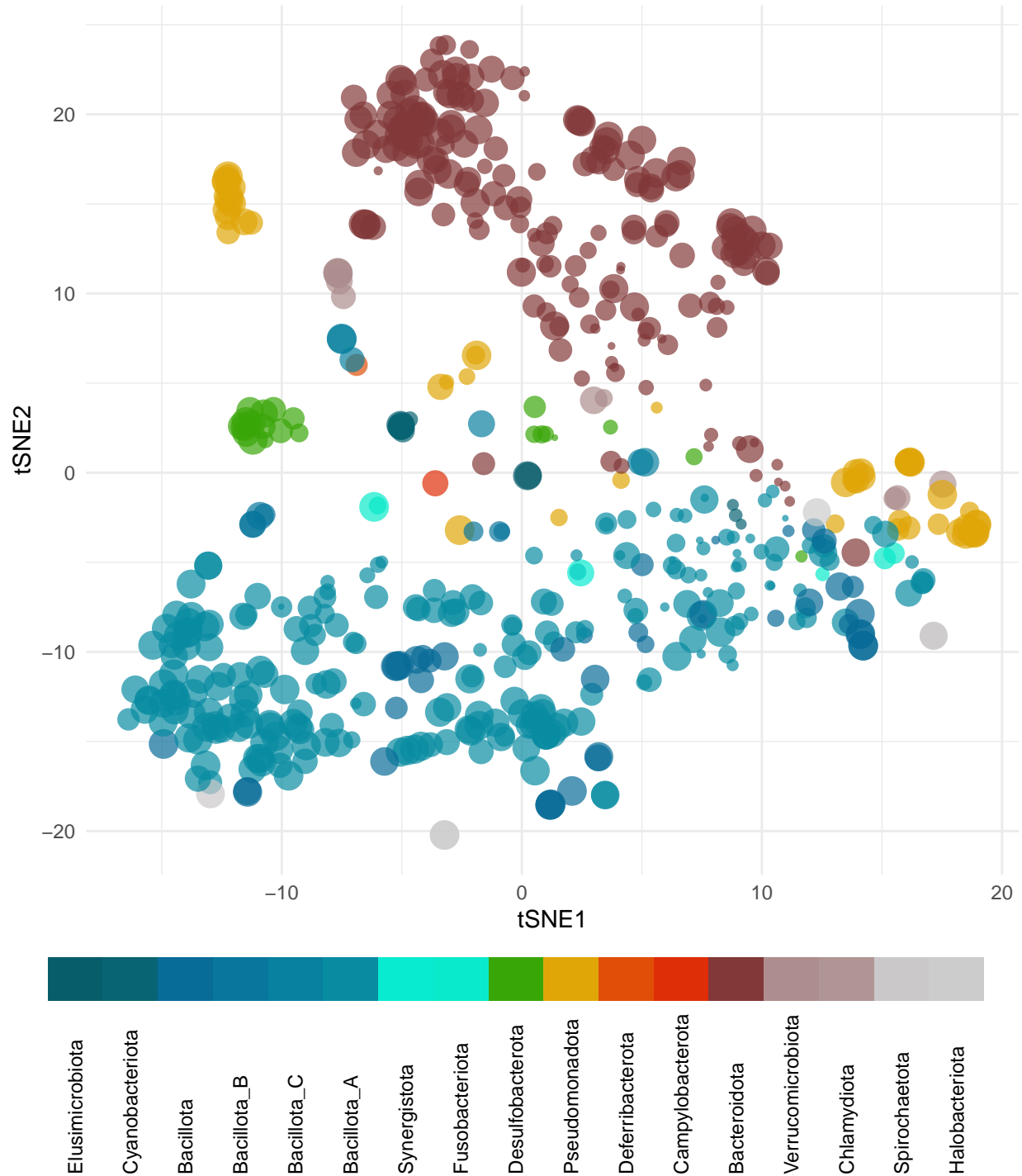
2.3 Functional attributes of MAGs

Prokaryotic genomes carry different sets of genes that confer their organisms with different functional capabilities. By annotating the microbial genes against the KEGG catalogue, enables calculating how full each of the hundreds of metabolic functions is in each genome, and in this way infer functional capabilities of each strain. The below figure shows the MAG phylogenetic tree associated with fullness (between 0-1) levels for each metabolic pathway.



2.4 Functional ordination of MAGs

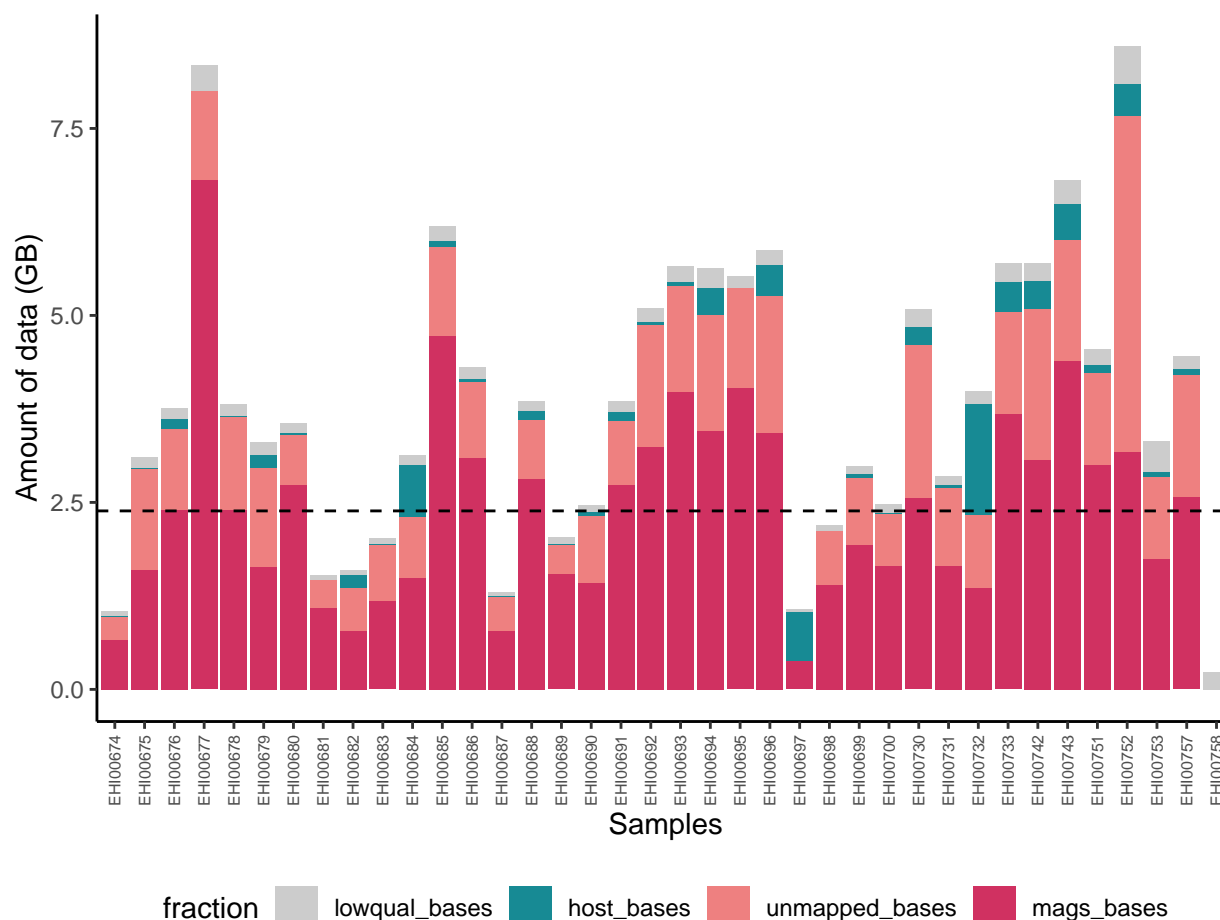
Using the functional annotation, it is possible to ordinate prokaryotic genomes on a bidimensional space. In doing so, one can assess how close any group of bacteria are in functional terms, or how functionally diverse the members of a given phylum can be.



3. Sequencing depth assessment

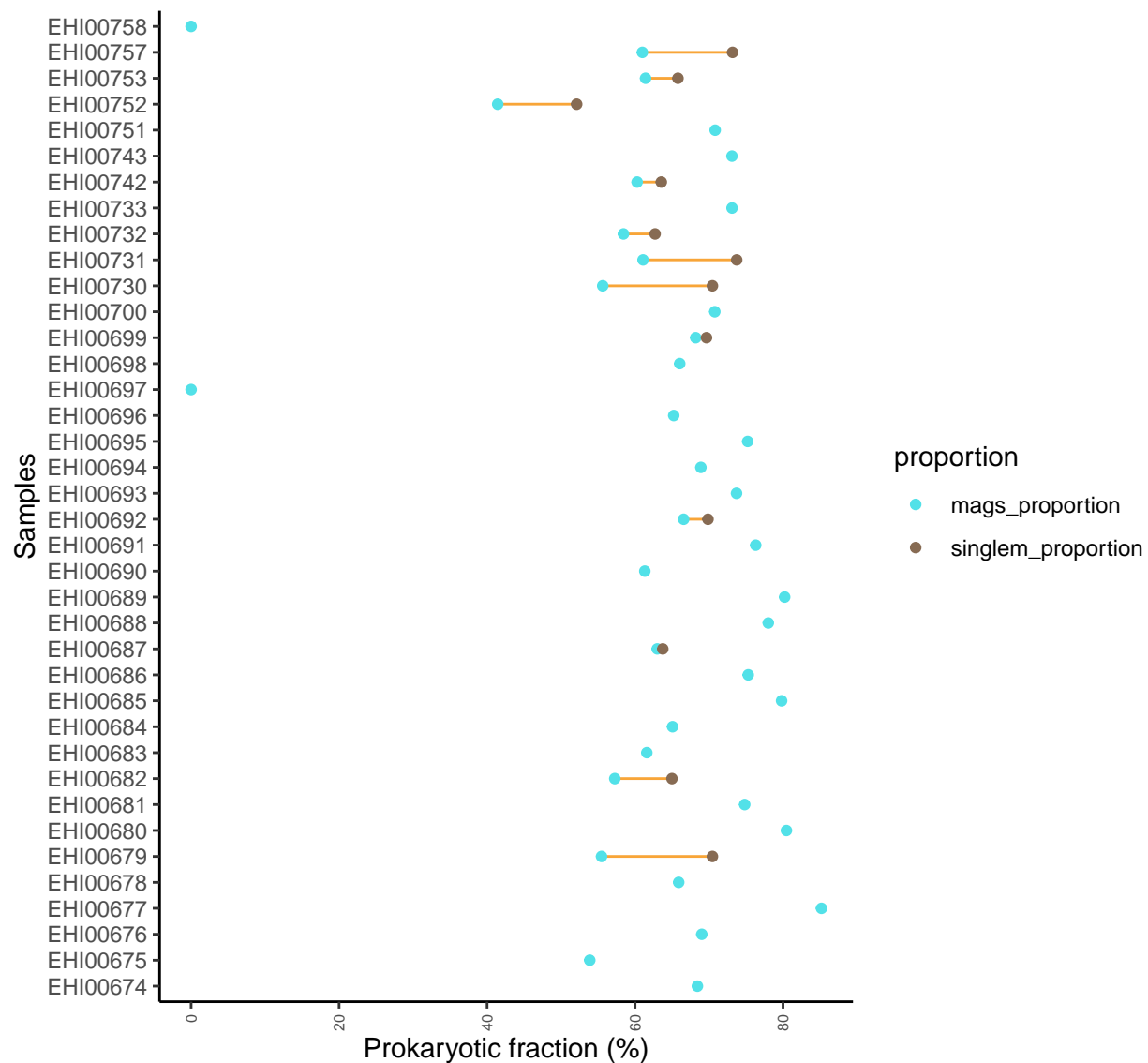
3.1 DNA fractions

When performing genome-resolved metagenomic analyses on host-associated microbial communities, the data usually contains a mixture of origins. One fraction is low-quality data that is discarded in the bioinformatic preprocessing due to lack of informativeness. These data include low-quality bases, adaptors, low-complexity reads and alike, which do not contribute to the study. Another fraction belongs to the host genome against which the data are mapped. The host fraction can be very variable depending on the species and the sample type, and while it is not informative for metagenomic analyses, it can be used for genomic analyses. The rest is what we call the metagenomic fraction. Part of the metagenomic fraction is built into draft bacterial genomes or MAGs, against which metagenomic reads are mapped later on to quantify relative representation of genomes. The fraction that is not built into MAGs is what is also unmapped against the MAG catalogue. This last fraction includes DNA dietary items, viruses and other organisms, but can values_to include prokaryotic DNA of bacteria and archaea that were unable to be reconstructed. In order to have representative results, the number of reads mapped to the MAG catalogue should be similar across samples. However, multiple reasons can create large imbalances, including uneven sequencing depth, different microbiome complexity across samples, different amount of host or non-microbial reads in the dataset, etc. The following plot shows the distribution of reads across samples.



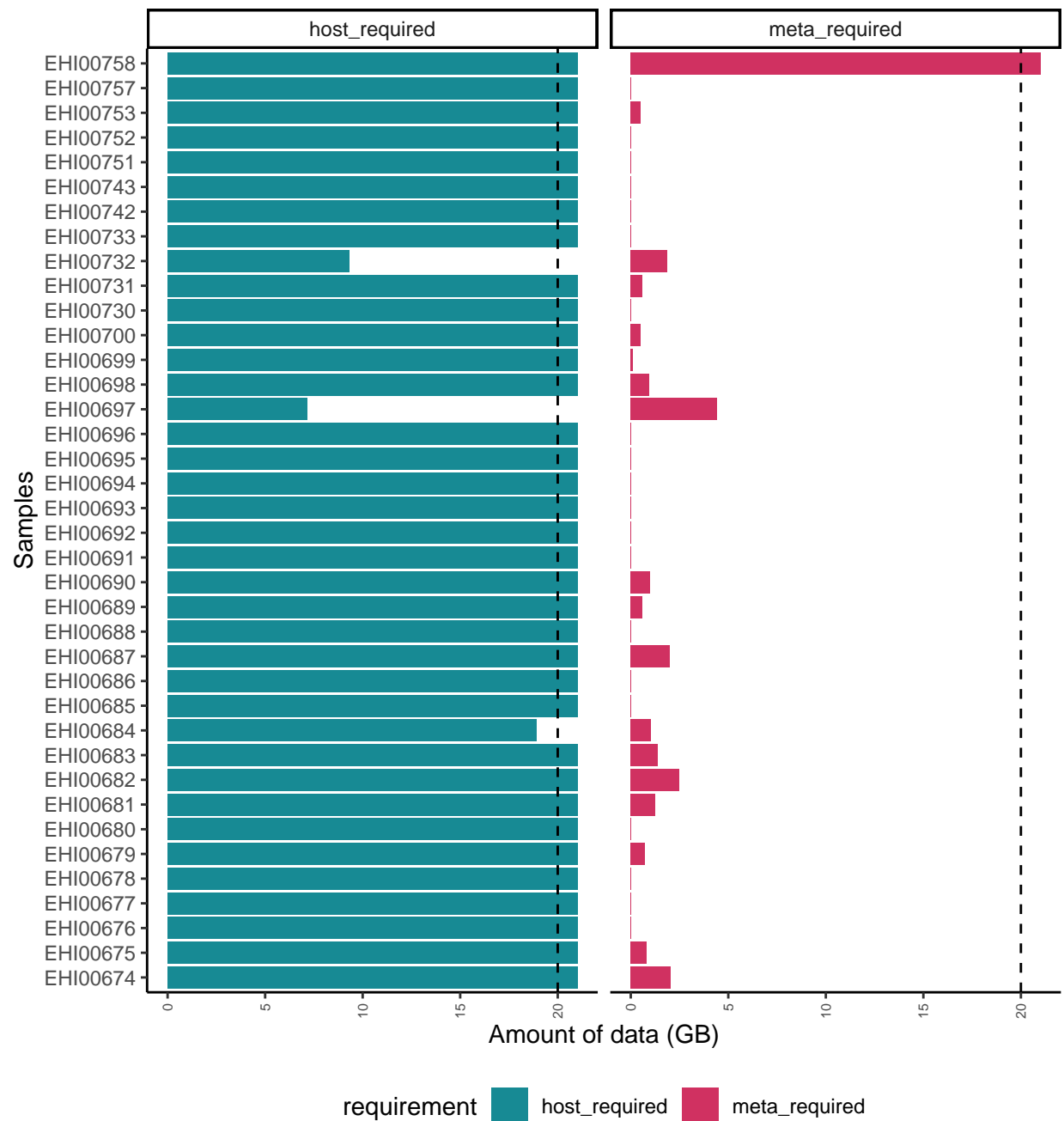
3.2 Estimated vs. mapped prokaryotic fraction

The main metric employed to assess the representativeness of a prokaryotic community using genome-resolved metagenomics is to measure the percentage of reads mapped to the MAG catalogue. This metric, however can be slippery when prokaryotic DNA is not dominant in the metagenomic mixture. Host DNA, dietary DNA, viral DNA and other non-prokaryotic DNA sequences can drastically underestimate representativeness in such cases. In the EHI we also employ another metric which estimated the proportion of prokaryotic sequences in a sample relying on coverage values on marker genes and extrapolation of genome sizes of the closest relatives. Combining both approaches, we can better estimate whether a prokaryotic community has been properly represented, or whether further sequencing is required. In the below chart, a longer orange line indicates that the mapping rate is further away from the expected fraction of prokaryotic DNA. The absence of lines and brown dots indicate that the mapping rate and the estimated fraction match, so the representativeness is correct. When the mapping value is at zero, it indicates that not enough metagenomic data was generated for this type of analyses, usually due to excessive host DNA.



3.3 Extra sequencing effort required

Estimating the amount of host and metagenomic data required for robust hologenomic analyses is very complex, because these values are dependent on many system and study-specific factors. Here you will find the minimum data requirements that work for most systems, which take as reference a desired amount of 5GB of host DNA and 2GB of prokaryotic DNA. The calculations of extra sequencing required take into account the different fractions of sequencing data, including low-quality, host, MAG catalogue-mapped and other metagenomic reads. The bars exceeding the vertical dashed bar set at 20GB of data indicate that more than 20GBs are required to reach the desired amount of host or prokaryotic data. You can check the table for the actual value, but bear in mind that generating more than 20GB of data is economically prohibitive in most cases.



sample	mags_bases	unmapped_bases	host_bases	lowqual_bases	meta_required	host_required
EHI00674	0.67	0.31	0.00	0.06	2.06	Inf
EHI00675	1.59	1.36	0.01	0.14	0.80	1546.90
EHI00676	2.40	1.08	0.14	0.14	0.00	130.53
EHI00677	6.82	1.18	0.00	0.34	0.00	Inf
EHI00678	2.40	1.24	0.01	0.15	0.00	1896.20
EHI00679	1.64	1.32	0.17	0.17	0.72	93.76
EHI00680	2.74	0.66	0.03	0.13	0.00	589.77
EHI00681	1.09	0.37	0.00	0.06	1.27	Inf
EHI00682	0.78	0.58	0.18	0.06	2.50	42.84
EHI00683	1.19	0.74	0.02	0.07	1.37	502.98
EHI00684	1.50	0.80	0.71	0.12	1.04	18.91
EHI00685	4.73	1.20	0.07	0.19	0.00	435.95
EHI00686	3.09	1.02	0.04	0.16	0.00	534.44
EHI00687	0.78	0.46	0.00	0.05	2.02	Inf
EHI00688	2.81	0.79	0.12	0.12	0.00	156.16
EHI00689	1.55	0.38	0.01	0.08	0.59	1007.98
EHI00690	1.42	0.90	0.05	0.08	1.00	242.55
EHI00691	2.74	0.85	0.12	0.14	0.00	156.57
EHI00692	3.24	1.63	0.04	0.19	0.00	632.40
EHI00693	3.98	1.42	0.05	0.21	0.00	560.34
EHI00694	3.45	1.56	0.37	0.26	0.00	70.58
EHI00695	4.04	1.33	0.01	0.15	0.00	2759.47
EHI00696	3.43	1.83	0.42	0.19	0.00	64.01
EHI00697	0.39	0.00	0.65	0.03	4.42	7.16
EHI00698	1.40	0.72	0.01	0.08	0.95	1102.79
EHI00699	1.93	0.90	0.04	0.11	0.11	369.52
EHI00700	1.66	0.69	0.02	0.12	0.51	620.01
EHI00730	2.56	2.05	0.24	0.23	0.00	100.75
EHI00731	1.65	1.05	0.05	0.11	0.61	283.14
EHI00732	1.36	0.97	1.50	0.16	1.88	9.31
EHI00733	3.69	1.36	0.41	0.25	0.00	63.92
EHI00742	3.07	2.02	0.37	0.23	0.00	71.20
EHI00743	4.39	1.62	0.48	0.31	0.00	64.03
EHI00751	3.00	1.24	0.11	0.21	0.00	202.71
EHI00752	3.18	4.49	0.43	0.50	0.00	91.40
EHI00753	1.74	1.09	0.08	0.41	0.50	204.18
EHI00757	2.57	1.64	0.08	0.17	0.00	274.29
EHI00758	0.00	0.00	0.00	0.24	Inf	Inf

4. Count data

4.1 Minimum genome-coverage filtering

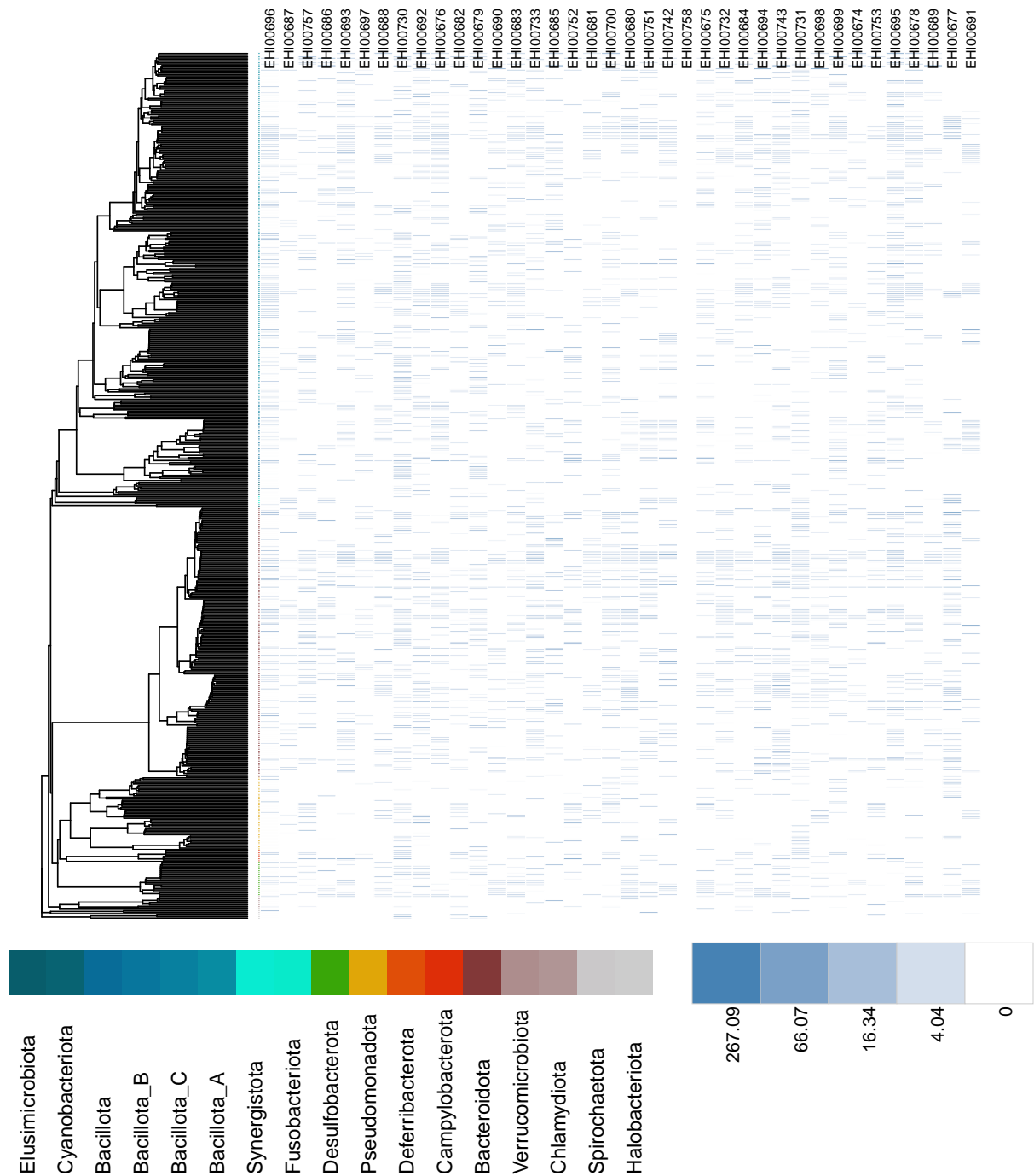
Mapping of sequencing reads against the reference genome catalogue is not perfect, and in consequence, all MAGs tend to get a few reads assigned. Implementing a minimum genome coverage filter aims at minimising artificial inflation of diversity due to this artifact of genome-resolved metagenomic analysis. The EHI pipeline implements a default coverage threshold of 0.3 (%30) of a genome. However, if the sequencing depth is low and uneven across samples, this filtering can also introduce distortion, by introducing false positives.

4.2 Genome-size normalisation

Bacterial genomes can vary between 1 and 8 MB, which make relative representation of each genome dependent on its size. To account for genome size biases, read-counts can be normalised by applying a normalisation factor that modifies the read numbers according to the size of each genome compared to the average genome size in the dataset.

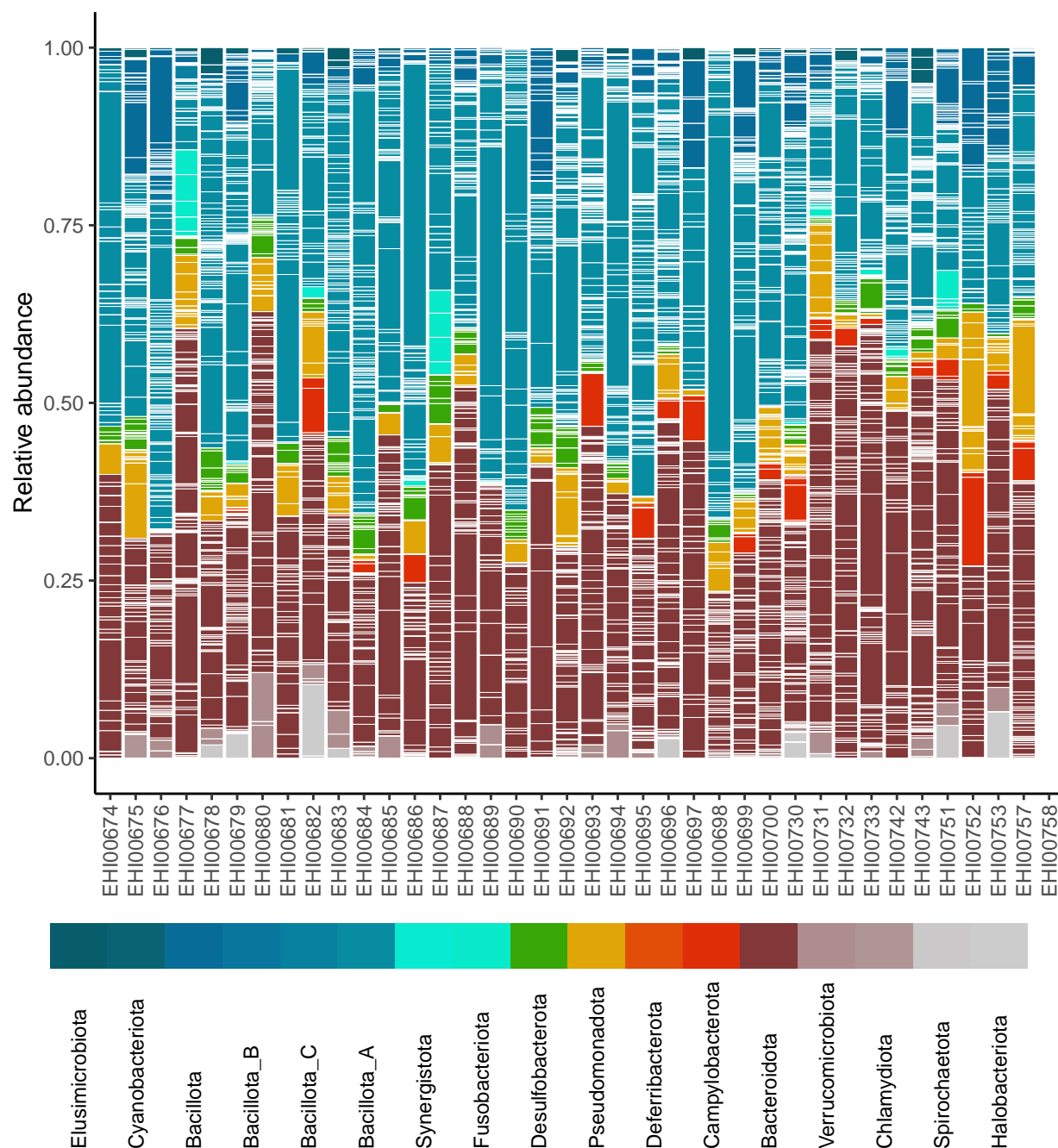
4.3 Count table

Once low-coverage genome counts have been filtered out, and the read counts have been normalised into genome counts, we can visualise the relative MAG abundances per sample. Note that the count scale is log-transformed.



5. Taxonomic composition

Taxonomic composition of each sample shown after total sum scaling (TSS) normalisation. Note that TSS normalisation simply divides each count value for the total count for the sample, thus transforming the data to 0-1 scale. TSS or similar procedures that normalise data to the 0-1 scale are required for some downstream analyses, such as Hill number computation. However, it is important to bear in mind that TSS normalisation is agnostic to sequencing depth, thus if depth is very uneven across samples, the results can be largely distorted. Each tile in the figure corresponds to a MAG, while colors indicate their phylum-level taxonomic assignment.



6. Diversity analyses

6.1 Alpha diversity

Diversity estimations for each sample.

Richness: Number of MAGs per sample (after applying coverage filter).

Neutral diversity: Hill number of $q=1$ (Shannon diversity), a diversity metric that accounts for richness and evenness (relative abundances) of the MAGs.

Phylogenetic (phylo) diversity: Phylogenetic Hill number of $q=1$, a diversity metric that accounts for richness and evenness (relative abundances), as well as phylogenetic relationships among MAGs.

Functional (func) diversity: Functional Hill number of $q=1$, a diversity metric that accounts for richness and evenness (relative abundances), as well as functional dissimilarities among MAGs.

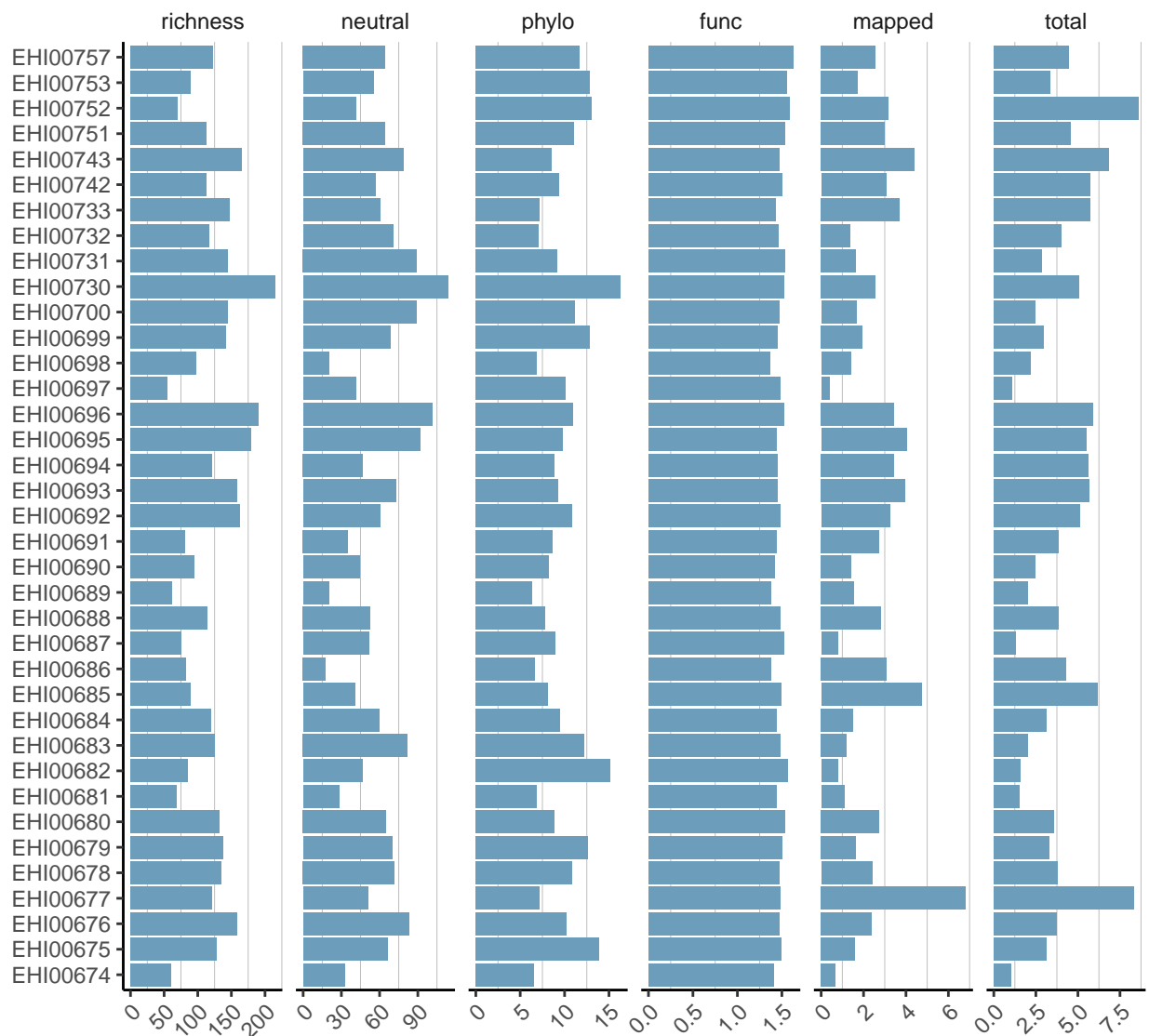
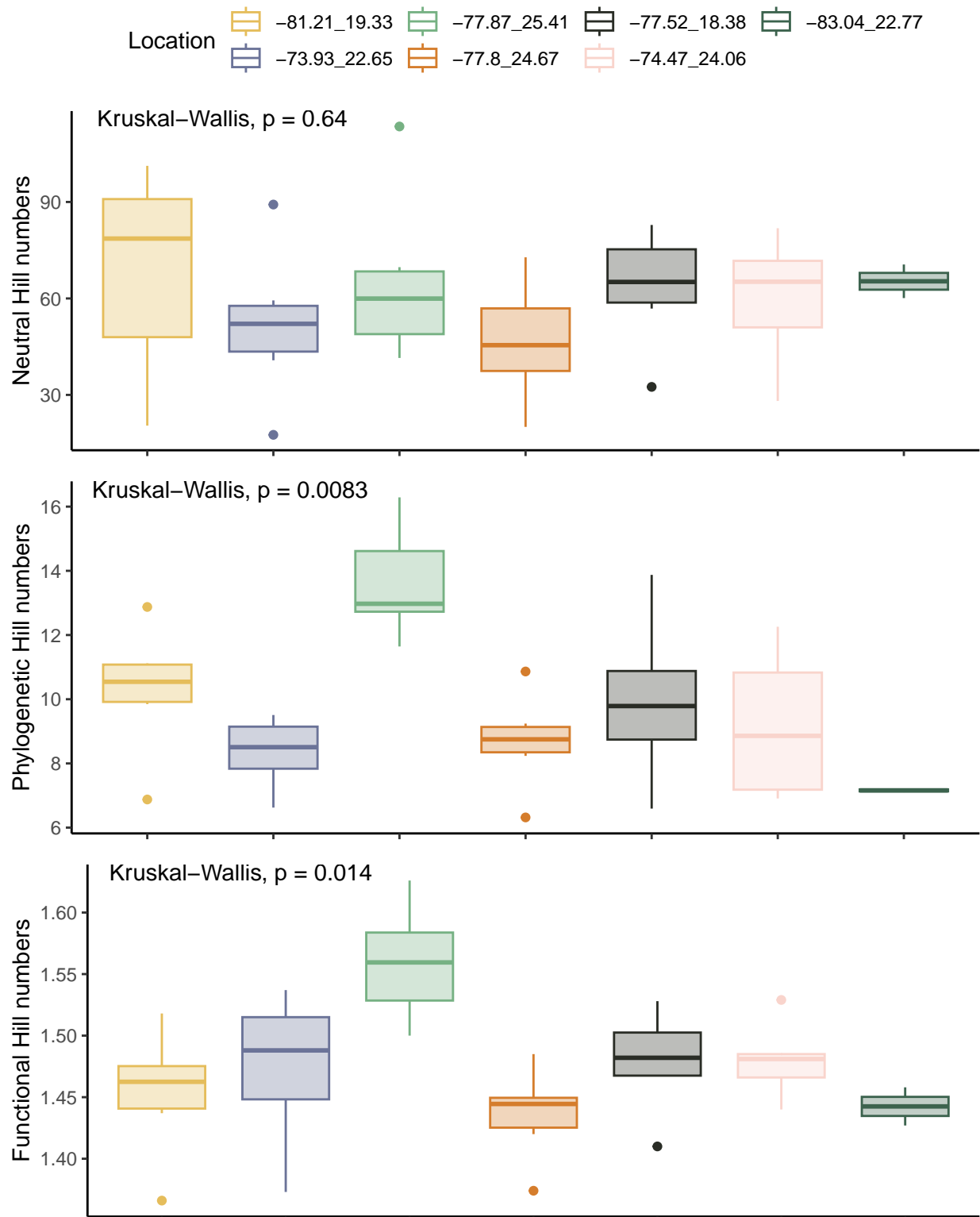


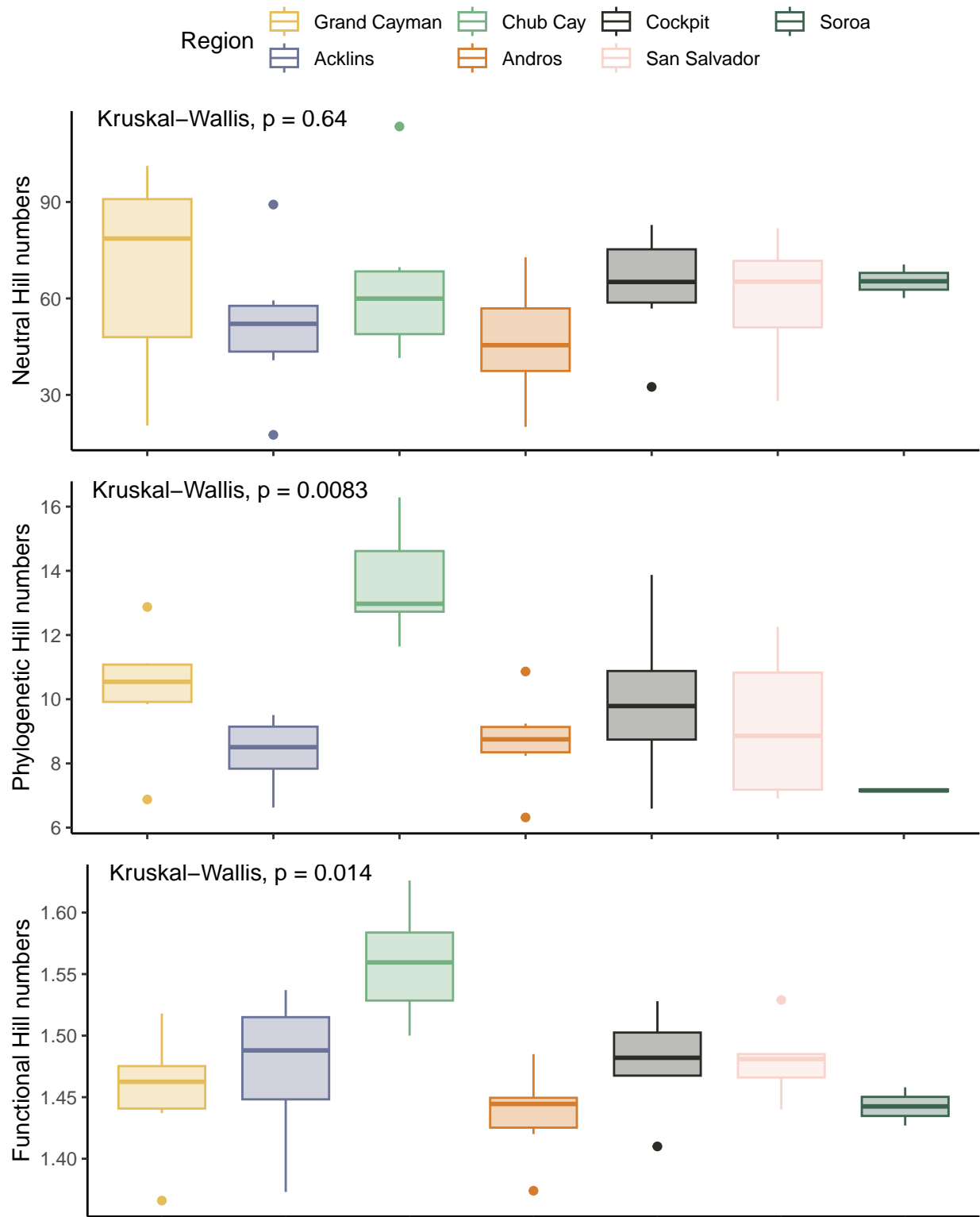
Table 4: Mapped and Total amount of data are shown in GB (giga-bases)

sample	richness	neutral	phylo	func	mapped	total
EHI00696	190	101.237	10.960	1.518	3.430	5.868
EHI00687	75	51.579	8.942	1.523	0.783	1.297
EHI00757	123	64.415	11.643	1.626	2.567	4.455
EHI00686	82	17.581	6.624	1.373	3.094	4.307
EHI00693	159	72.791	9.241	1.450	3.975	5.656
EHI00697	54	41.198	10.126	1.476	0.386	1.067
EHI00688	114	52.637	7.758	1.485	2.812	3.850
EHI00730	215	113.479	16.288	1.521	2.565	5.078
EHI00692	162	60.243	10.864	1.485	3.244	5.099
EHI00676	159	82.850	10.203	1.469	2.401	3.758
EHI00682	85	46.715	15.128	1.568	0.780	1.596
EHI00679	137	69.732	12.678	1.500	1.645	3.306
EHI00690	95	44.051	8.235	1.420	1.424	2.459
EHI00683	125	81.839	12.258	1.481	1.189	2.022
EHI00733	147	60.117	7.219	1.427	3.688	5.706
EHI00685	89	40.760	8.071	1.491	4.730	6.188
EHI00752	70	41.494	13.077	1.589	3.176	8.595
EHI00681	69	28.127	6.909	1.440	1.093	1.528
EHI00700	144	88.972	11.118	1.473	1.660	2.480
EHI00680	132	65.208	8.859	1.529	2.736	3.559
EHI00751	113	64.296	11.104	1.528	2.999	4.552
EHI00742	113	56.852	9.372	1.505	3.071	5.697
EHI00675	128	65.964	13.873	1.495	1.592	3.102
EHI00732	117	70.557	7.101	1.458	1.360	3.986
EHI00684	120	59.392	9.508	1.436	1.498	3.132
EHI00694	121	46.891	8.820	1.448	3.452	5.636
EHI00743	165	78.383	8.533	1.467	4.395	6.805
EHI00731	144	89.203	9.214	1.537	1.648	2.853
EHI00698	97	20.448	6.877	1.366	1.399	2.199
EHI00699	142	68.270	12.874	1.452	1.931	2.983
EHI00674	60	32.495	6.593	1.410	0.669	1.039
EHI00753	89	55.487	12.870	1.551	1.742	3.324
EHI00695	179	91.539	9.847	1.437	4.036	5.528
EHI00678	135	71.700	10.831	1.466	2.403	3.815
EHI00689	62	20.056	6.318	1.374	1.551	2.029
EHI00677	121	50.998	7.182	1.485	6.818	8.344
EHI00691	81	35.247	8.681	1.441	2.738	3.846

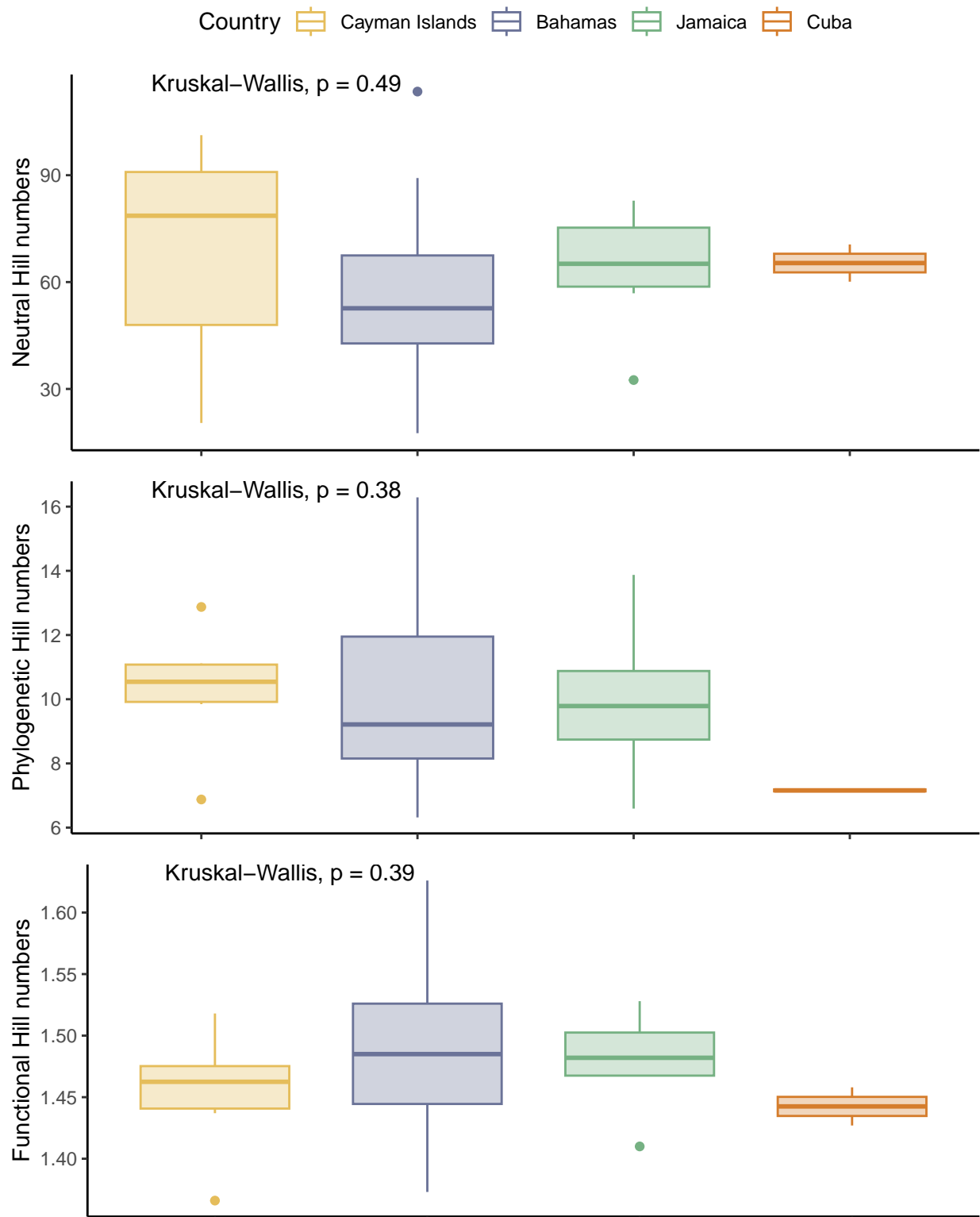
Alpha diversity variation across locations.



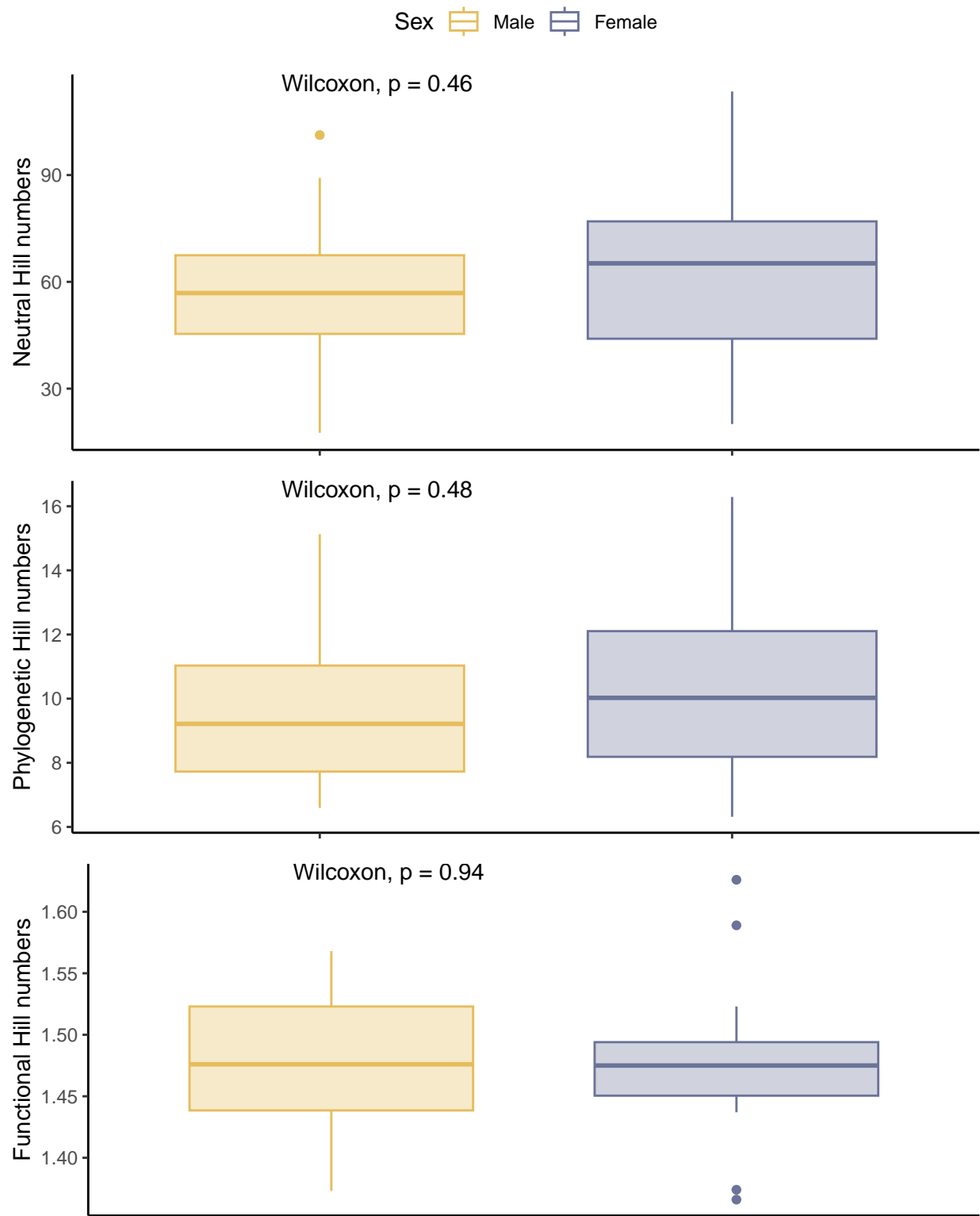
Alpha diversity variation across regions.



Alpha diversity variation across countries.

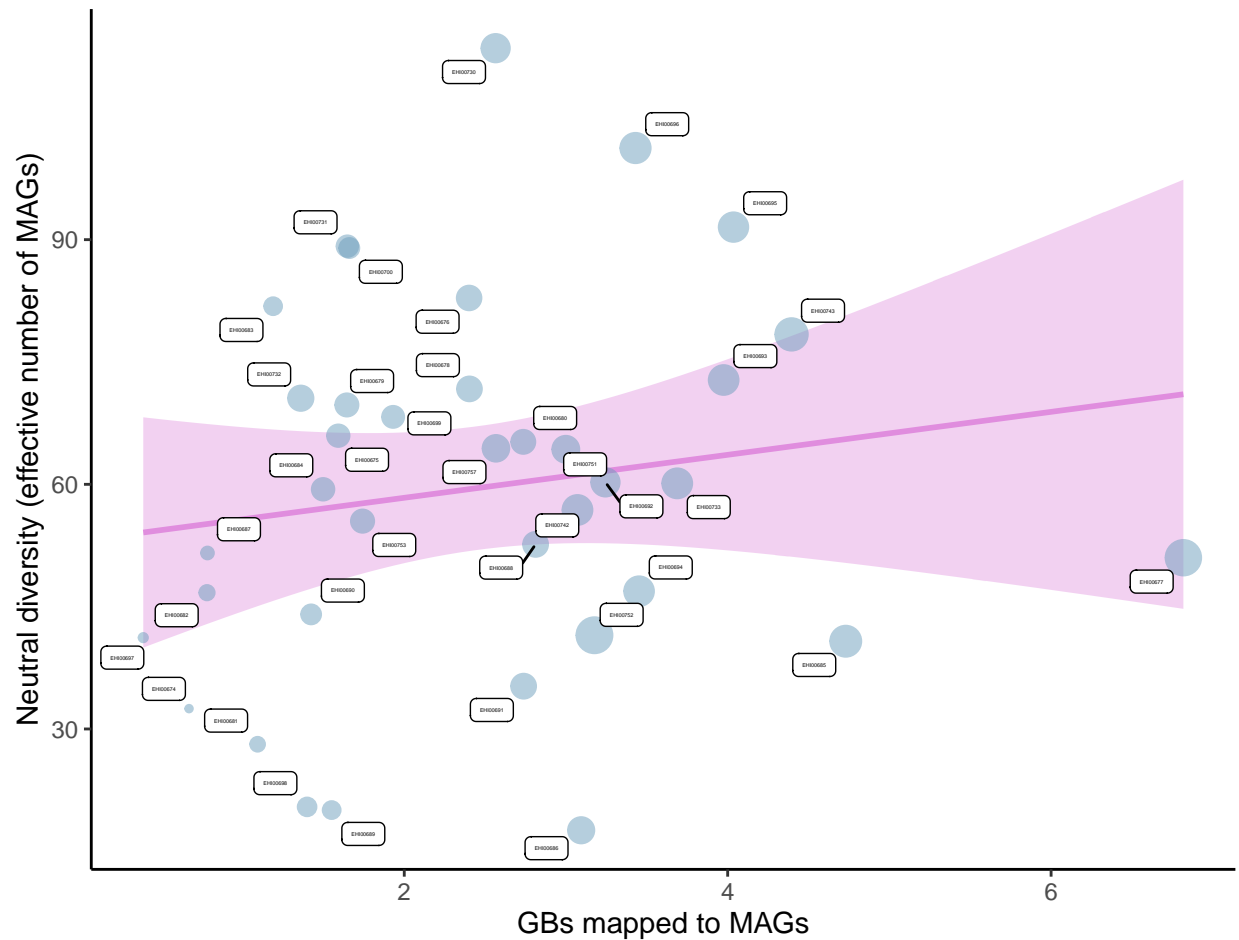


Alpha diversity variation between sexes.



6.2 Relationship between sequencing and alpha diversities

The microbial diversity in a sample is often correlated with sequencing depth, mainly when the sequencing effort is not large enough to properly capture the entirety of the diversity in the sample. The following plot shows the relationship between neutral diversity (Hill number of $q=1$, Shannon diversity) and the GBs mapped to the MAG catalogue. The size of the dots indicates the total sequencing effort performed for that sample (including host DNA, non-mapped reads, etc.). The flatter the regression line the lower is the correlation, and thus lower is the effect of sequencing depth in the results. However, it must be noted that this analysis only compares the sequencing depth in relation to the MAG catalogue, which might not be complete or representative of the system. Hence, the results shown here must be interpreted cautiously.



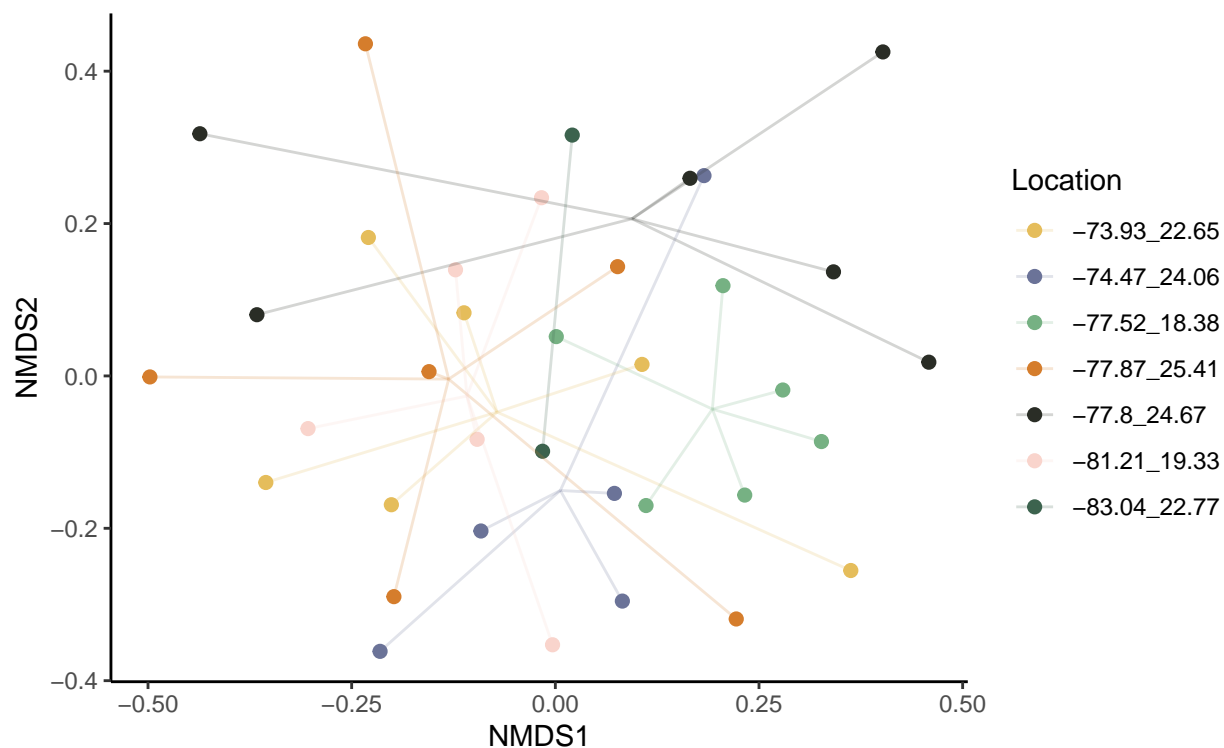
6.3 Beta diversity

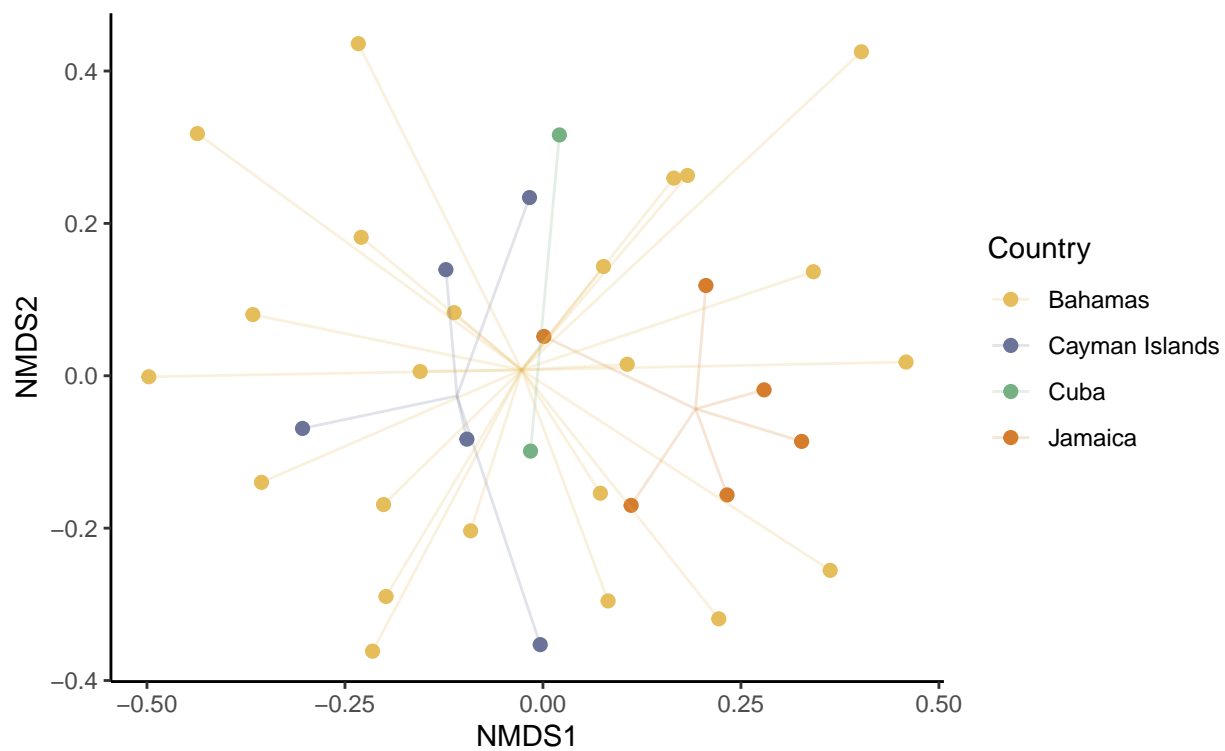
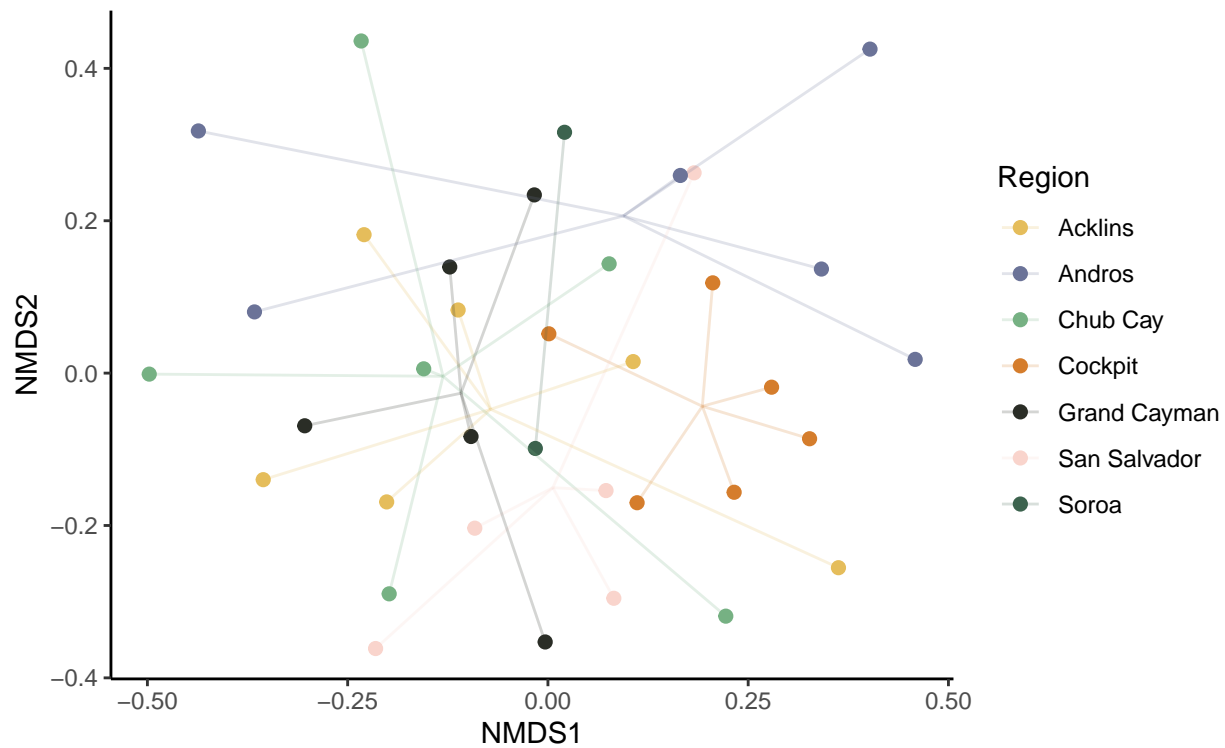
Pairwise beta diversities quantify the dissimilarity between samples. Within the Hill numbers framework, beta diversities can be calculated for neutral, phylogenetic and functional diversities, at the desired order of diversity (q -value). Often times, beta diversities differ depending on the components of the diversity. For instance, phylogenetic and functional diversities often display lower beta values than neutral diversities due to phylogenetic and functional redundancy of the microbiota (i.e. different yet closely related and functionally similar MAGs replace each other). Pairwise beta diversities can be displayed in a 2-dimensional ordination using NMDS, and can be used to test for compositional differences between variables of interest. In the following pages, PERMANOVAs are conducted against all metadata variables with >1 different values, and NMDS ordinations are displayed for all metadata features containing between 2 and 8 different values. Note that outliers (often technical failures) can distort the ordination considerably.

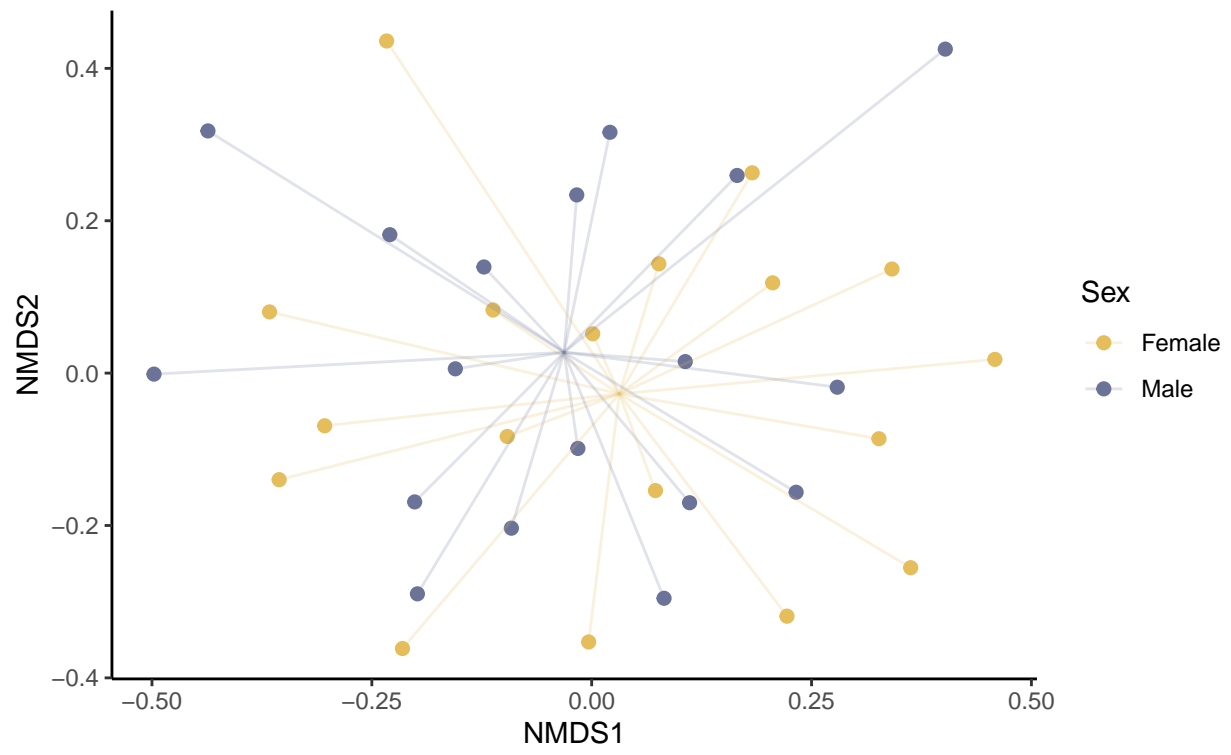
6.3.1 Neutral beta diversity

The PERMANOVA analysis based on neutral beta diversities indicates whether microbiomes vary between metadata variables.

	Df	SumOfSqs	R2	F	Pr(>F)
location	6	1.3247170	0.1191033	0.6509051	0.994
sex	1	0.3001424	0.0269853	0.8848571	0.650
Residual	28	9.4975634	0.8539114	NA	NA
Total	35	11.1224228	1.0000000	NA	NA



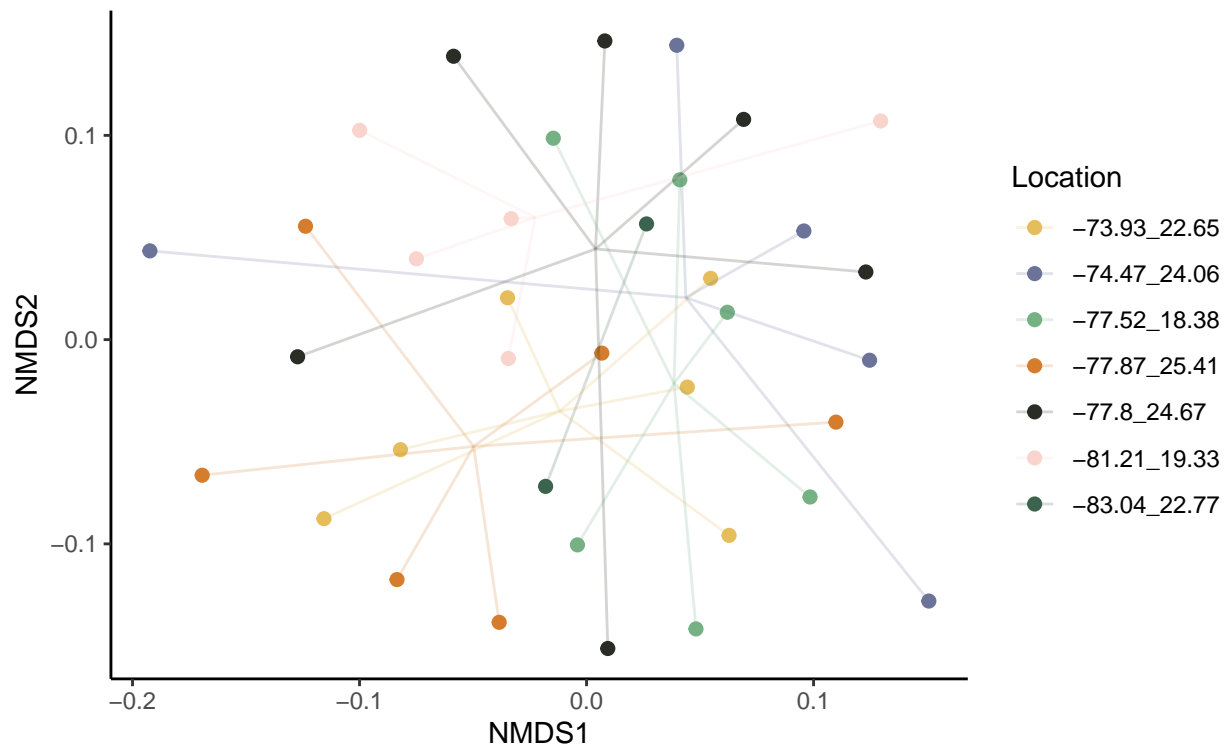


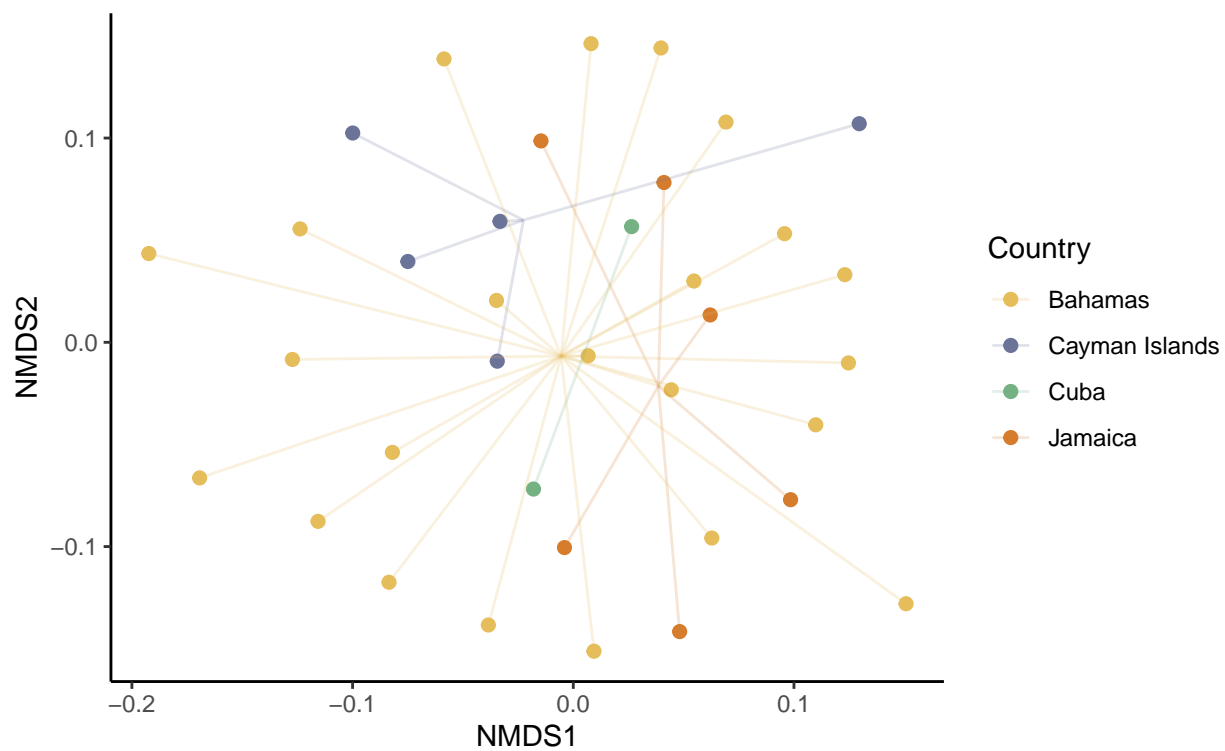
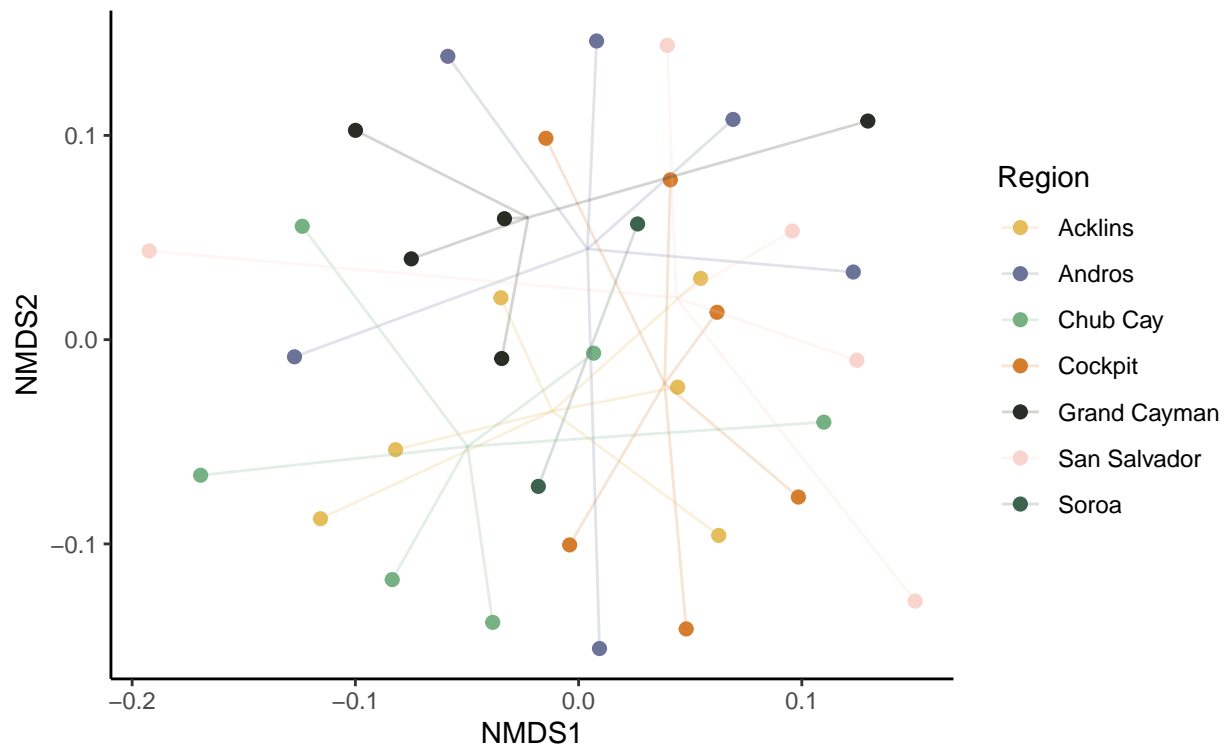


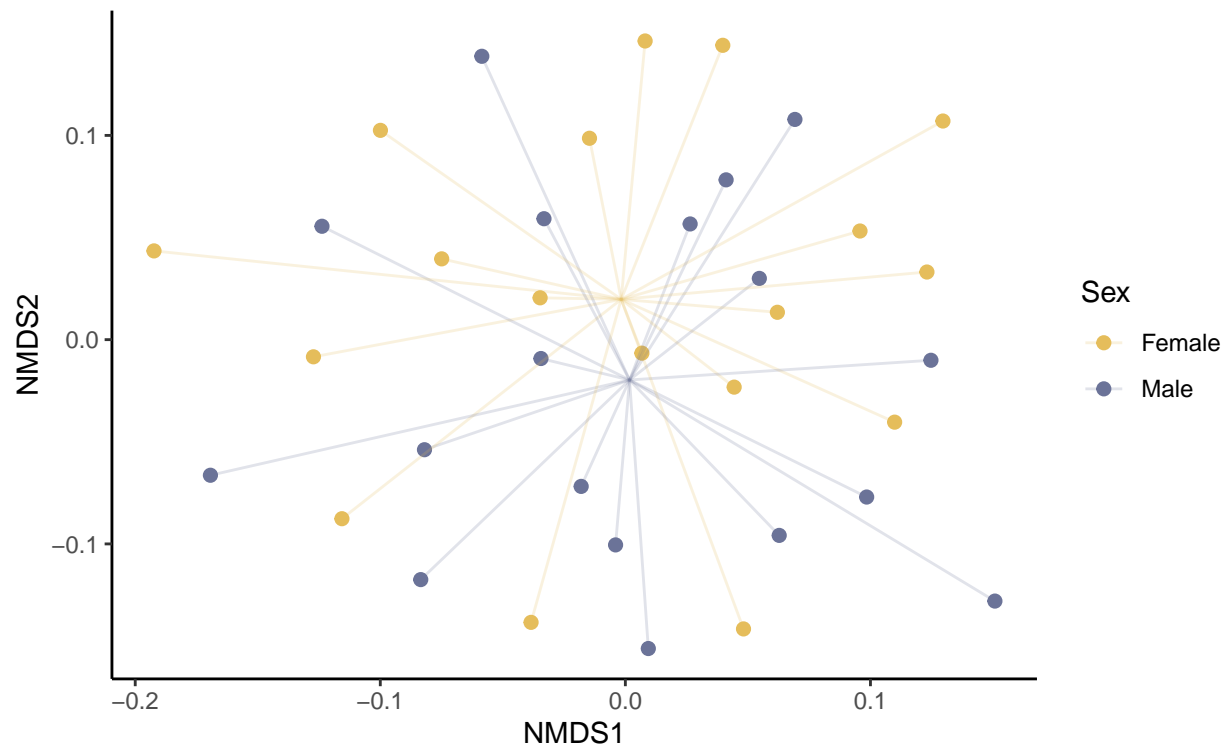
6.3.2 Phylogenetic beta diversity

The PERMANOVA analysis based on phylogenetic beta diversities indicates whether phylogenetic attributes of microbiomes vary between metadata variables.

	Df	SumOfSqs	R2	F	Pr(>F)
location	6	0.0719329	0.0944958	0.5008412	0.959
sex	1	0.0190495	0.0250247	0.7958078	0.644
Residual	28	0.6702459	0.8804795	NA	NA
Total	35	0.7612283	1.0000000	NA	NA







6.3.3 Functional beta diversity

The PERMANOVA analysis based on functional beta diversities indicates whether functional attributes of microbiomes vary between metadata variables.

	Df	SumOfSqs	R2	F	Pr(>F)
location	6	0.160799	0.2675132	1.6968056	0.287
sex	1	-0.001951	-0.0032458	-0.1235272	0.714
Residual	28	0.442240	0.7357326	NA	NA
Total	35	0.601088	1.0000000	NA	NA

