Community profiling - amplicon

Nils Peder Willassen
Centre for bioinformatics
UiT The Arctic University of Norway
ELIXIR Norway

Community profiling - terminology

Community

"Group or association of populations of two or more different species"

Richness

The species richness is how much species there are in a sample

Eveness

The species evenness is who equal the relative number of species are in a sample

Diversity = Richness + Evenness

Community profiling - terminology

Relative abundance

Number of each species relative to total number of all species in a sample

Clustering

Grouping sequences into clusters (or bins) bases on percent similarity (commonly 97%). Each bin/cluster is termed an OTU.

OTU: operational taxonomical unit

Most commonly used microbial diversity unit. Whilst sometimes used as a proxy for species it is a distinct entity.

Chimera

PCR artefact

Barcode/index

Short piece of DNA added to each read that is sample specific; allows for multiplexing of samples

Community profiling - terminology

Multiplexing/Demultiplexing

Pooling multiple DNA samples together before sequencing. During downstream analysis samples will be demultiplexed (separated by sample) based on the barcodes.

OTU table (Biome table)

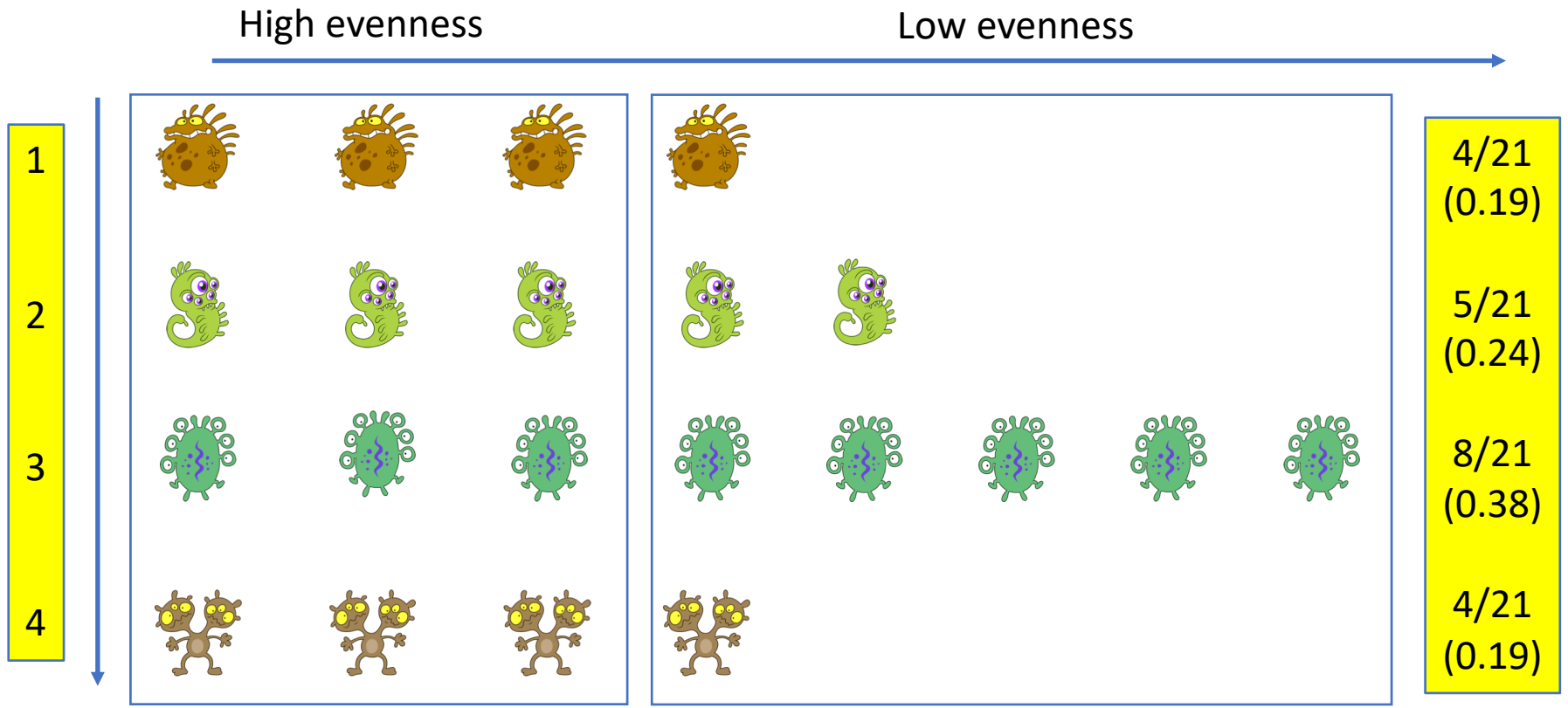
Matrix containing counts of OTUs and corresponding metadata (e.g. taxonomy)

Community profiling - diversity

Richness & Evenness

Eveness

The species ***evenness*** is who equal the relative number of species are



Richness

The species ***richness*** is how many species there are in a sample

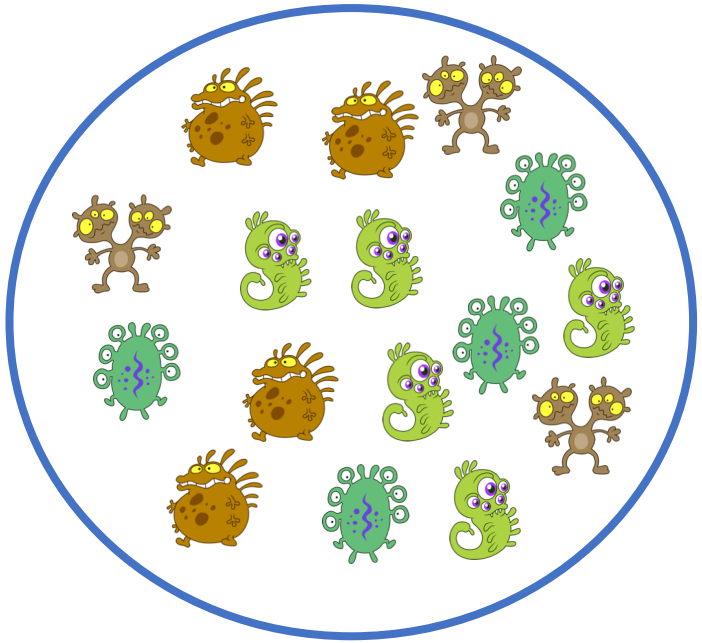
Relative abundance

Number of each species relative to total number of all species in a sample

Community profiling - diversity

Diversity = Richness + Evenness

Community 1



Richness = 4

Evenness (high) >

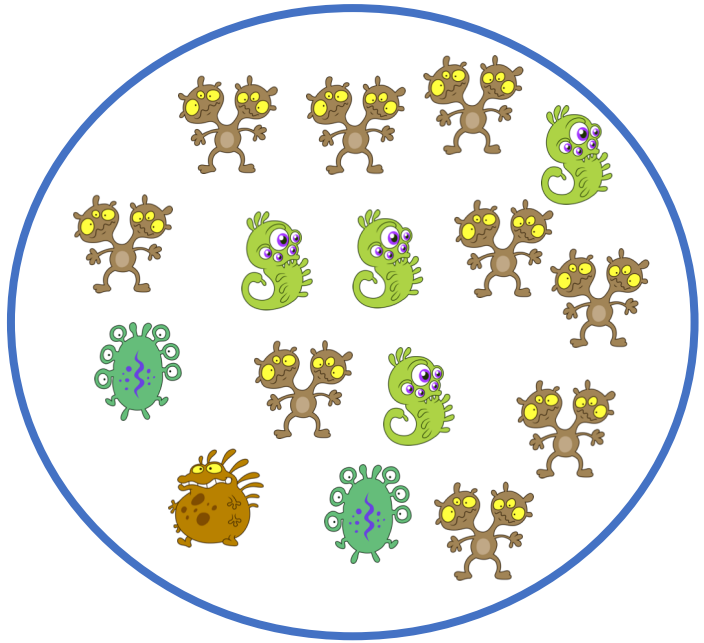
Diversity >

Richness = 4

Eveness (low)

Diversity

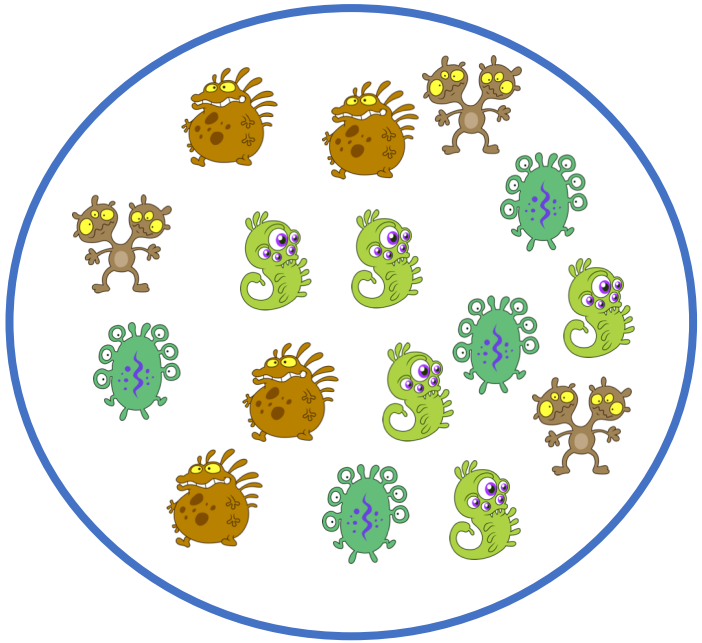
Community 2



Community profiling - diversity

Diversity = Richness + Evenness

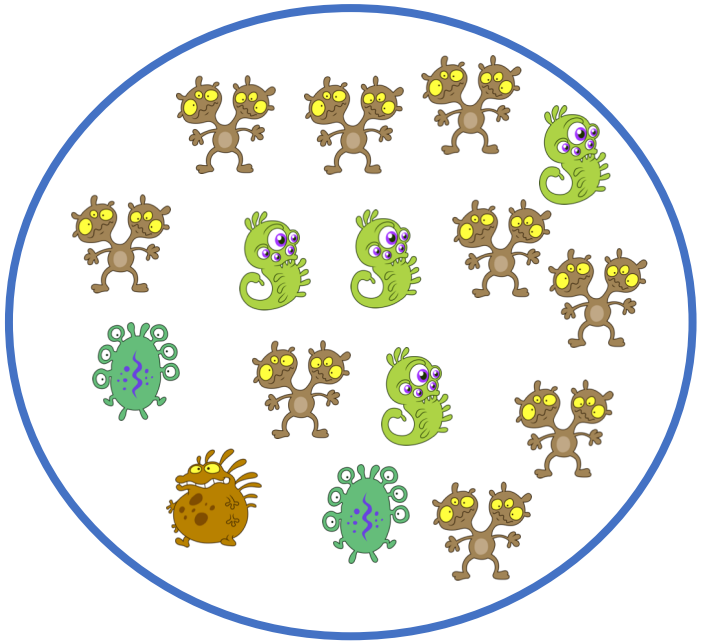
Community 1



OTU table

OTU ID	Community 1	Community 2	Taxon
OTU1	4	1	
OTU2	4	2	
OTU3	5	4	
OTU4	3	9	

Community 2



OTU = Operational taxonomic Unit

Community profiling - amplicon

Classification using markers

Marker gene (e.g. 16S)

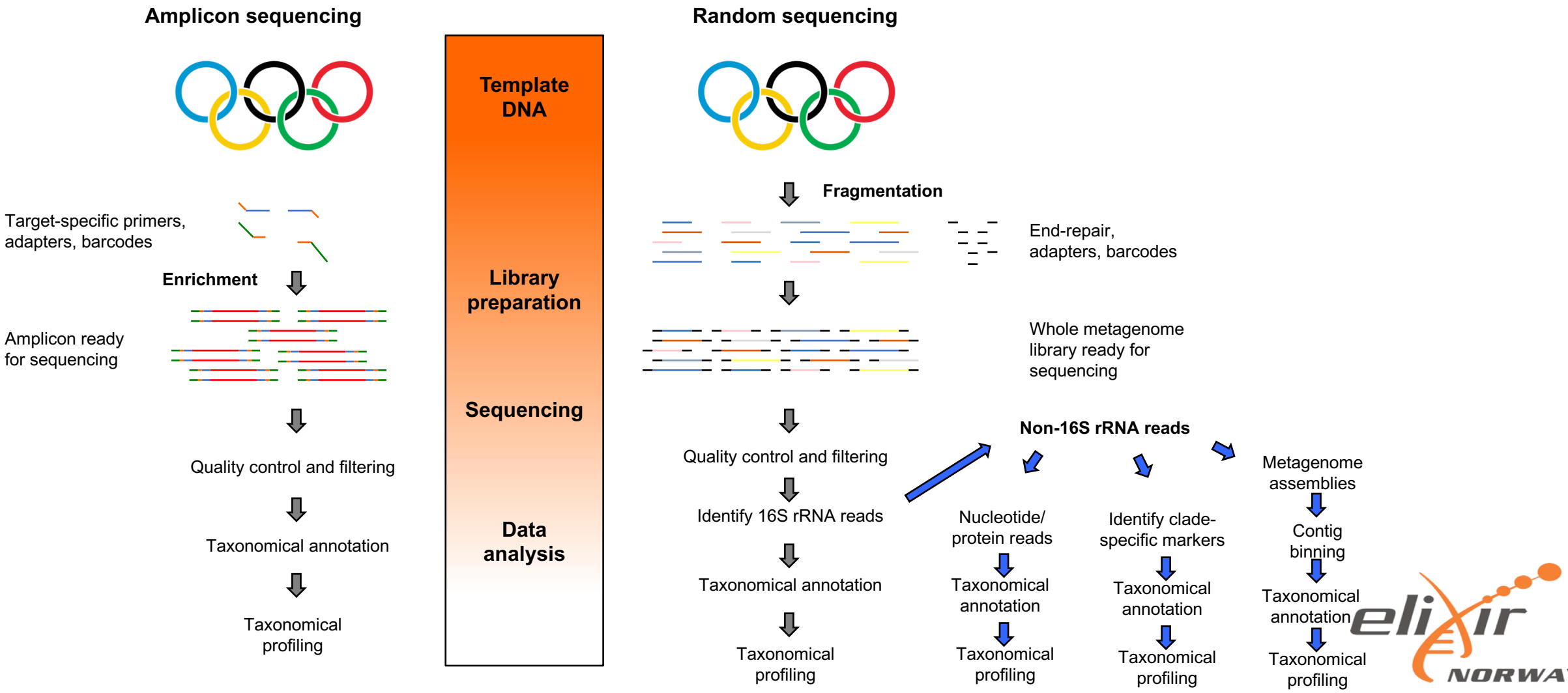
```
AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCTCAGA TAACTGGGCC
CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAAA
TGGATTTATC TGCTCTTCGC GTTGAAGAAG TACAAAATGT CATTAATGCT ATGCAGAAAA
TCTTAGAGTG TCCCATCTGT CTGGAGTTGA TCAAGGAACC TGTCTCCACA AAGTGTGACC

TCTTAGAGTG TCCCATCTGT CTGGAGTTGA TCAAGGAACC TGTCTCCACA AAGTGTGACC
ACATATTTTG CAAATTTTGC ATGCTGAAAC TTCTCAACCA GAAGAAAGGG CCTTCACAGT
GTCCTTTATG TAAGAATGAT ATAACCAAAA GGAGCCTACA AGAAAGTACG AGATTTAGTC
AACTTGTTGA AGAGCTATTG AAAATCATTT GTGCTTTTCA GCTTGACACA GGTTTGGAGT

GTCCTTTATG TAAGAATGAT ATAACCAAAA GGAGCCTACA AGAAAGTACG AGATTTAGTC
AACTTGTTGA AGAGCTATTG AAAATCATTT GTGCTTTTCA GCTTGACACA GGTTTGGAGT
ATGCAAACAG CTATAATTTT GCAAAAAAGG AAAATAACTC TCCTGAACAT CTAAAAGATG
AAGTTTCTAT CATCCAAAGT ATGGGCTACA GAAACCGTGC CAAAAGACTT CTACAGAGTG
AACCCGAAAA TCCTTCCTTG CAGGAAACCA GTCTCAGTGT CCAACTCTCT AACCTTGGAA

ATGCAAACAG CTATAATTTT GCAAAAAAGG AAAATAACTC TCCTGAACAT CTAAAAGATG
AAGTTTCTAT CATCCAAAGT ATGGGCTACA GAAACCGTGC CAAAAGACTT CTACAGAGTG
AACCCGAAAA TCCTTCCTTG CAGGAAACCA GTCTCAGTGT CCAACTCTCT AACCTTGGAA
CTGTGAGAAC TCTGAGGACA AAGCAGCGGA TACAACCTCA AAAGACGTCT GTCTACATTG

AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCTCAGA TAACTGGGCC
CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAAA
TGGATTTATC TGCTCTTCGC GTTGAAGAAG TACAAAATGT CATTAATGCT ATGCAGAAAA
TCTTAGAGTG TCCCATCTGT CTGGAGTTGA TCAAGGAACC TGTCTCCACA AAGTGTGACC
```

OUT table

Constructed from biom file

#OTU ID	A1	A2	B1	B2	C1	C2	D1	D2	ConsensusLineage
denovo0	1	0	0	0	0	0	0	0	k__Bacteria
denovo1	0	1	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__
denovo2	1	0	1	0	0	1	0	0	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides
denovo3	0	0	0	0	0	2	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Dialister; s__
denovo4	0	1	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus
denovo5	2	0	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__
denovo6	0	0	0	0	1	1	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae
denovo7	0	0	0	0	3	1	10	11	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__
denovo8	1	7	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia; s__
denovo9	0	0	0	1	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae
denovo10	1	0	0	2	0	1	1	0	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Desulfovibrionales; f__Desulfovibrionaceae; g__; s__
denovo11	0	0	0	0	0	0	0	3	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__Finegoldia; s__
denovo12	0	0	0	0	0	0	0	1	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales
denovo13	0	0	0	0	0	1	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae
denovo14	12	13	6	13	121	58	1	12	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Dialister; s__
denovo15	30	16	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae
denovo16	0	0	0	1	0	0	0	0	k__Bacteria; p__Firmicutes; c__Bacilli
denovo17	8	4	0	3	1	0	1	2	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales
denovo18	0	0	1	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales
denovo19	0	0	0	0	1	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales

Community profiling – amplicon vs random sequencing



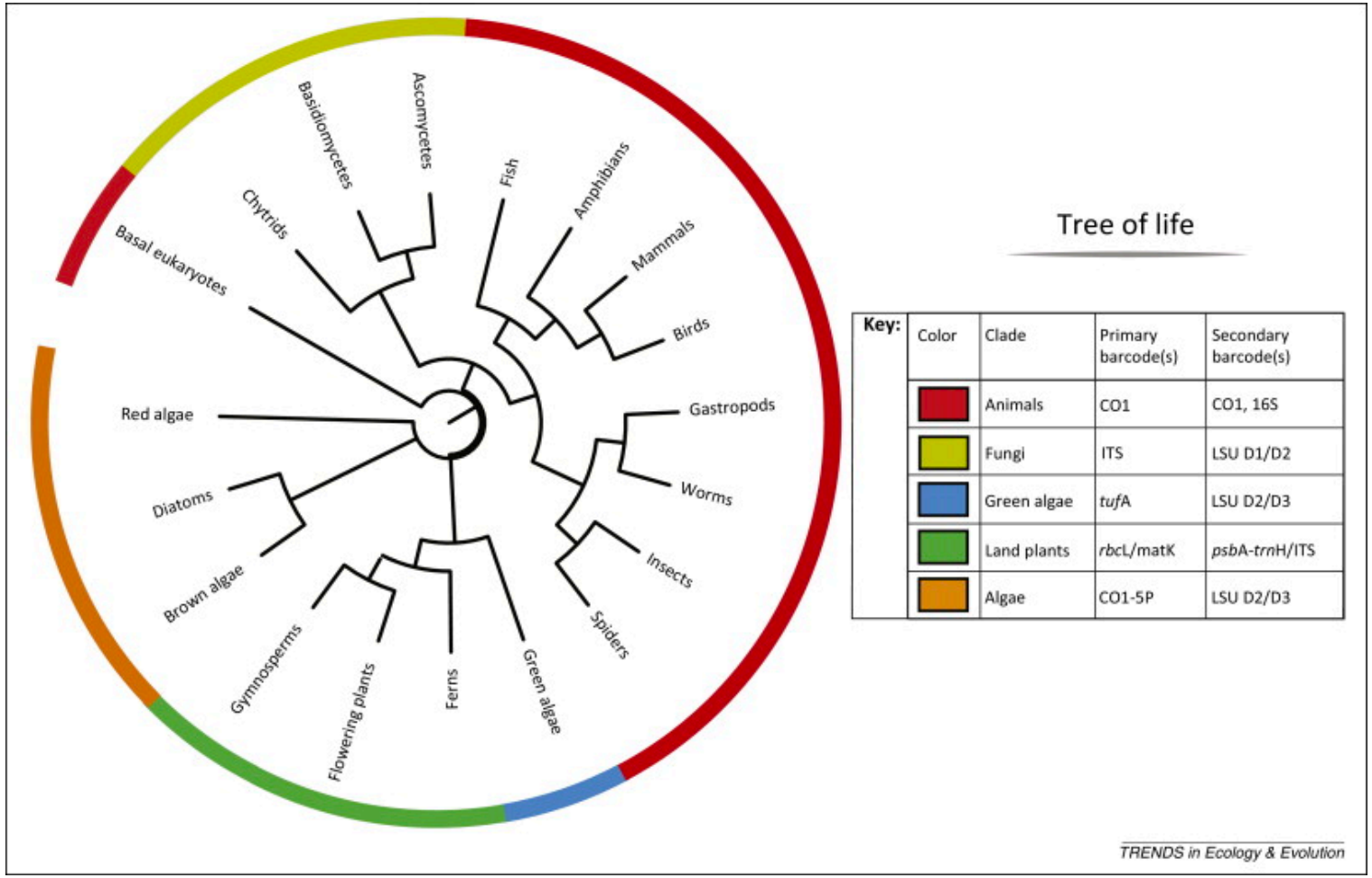
Community profiling - amplicon

Community profiling

Amplicon sequencing

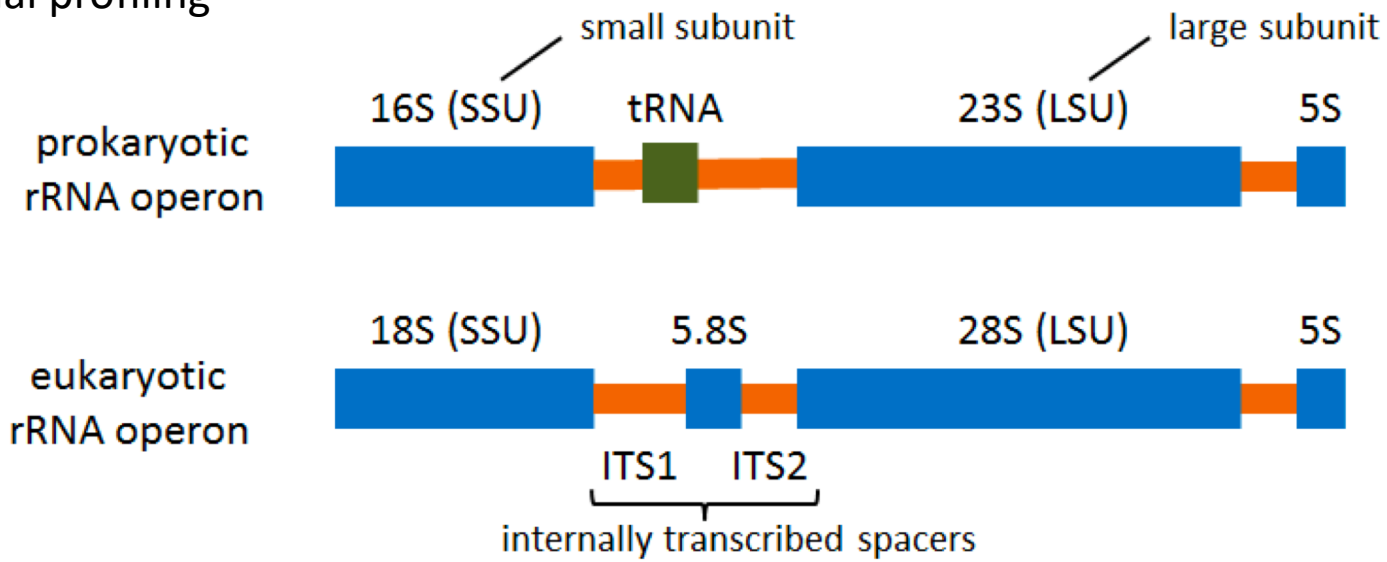
Targeted sequencing

Metabarcoding



Community profiling - amplicon

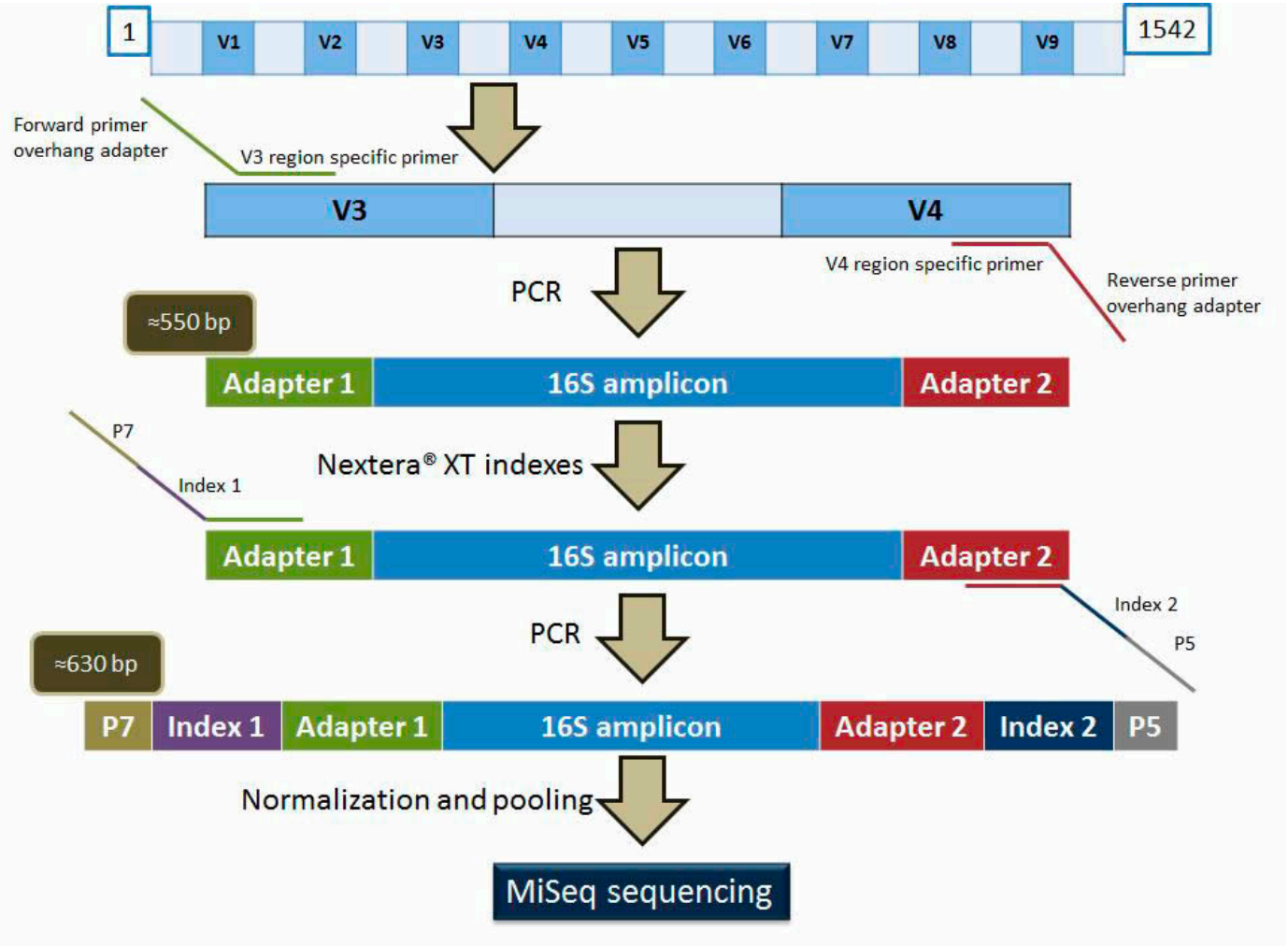
Markers for microbial profiling



Type	LSU	SSU
prokaryotic	5S - 120 bp 23S - 2906 bp	16S - 1542 bp
eukaryotic	5S - 121 bp 5.8S - 156 bp 28S - 5070 bp	18S - 1869 bp

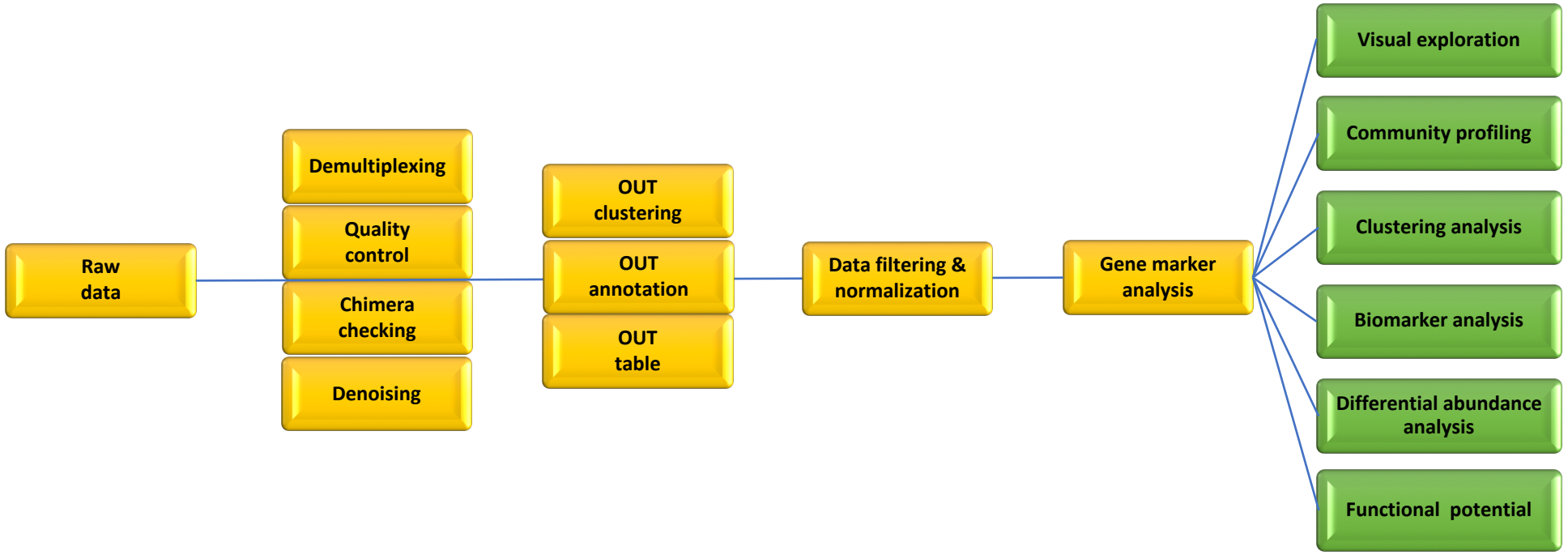
Community profiling - amplicon

Illumina amplicon sequencing



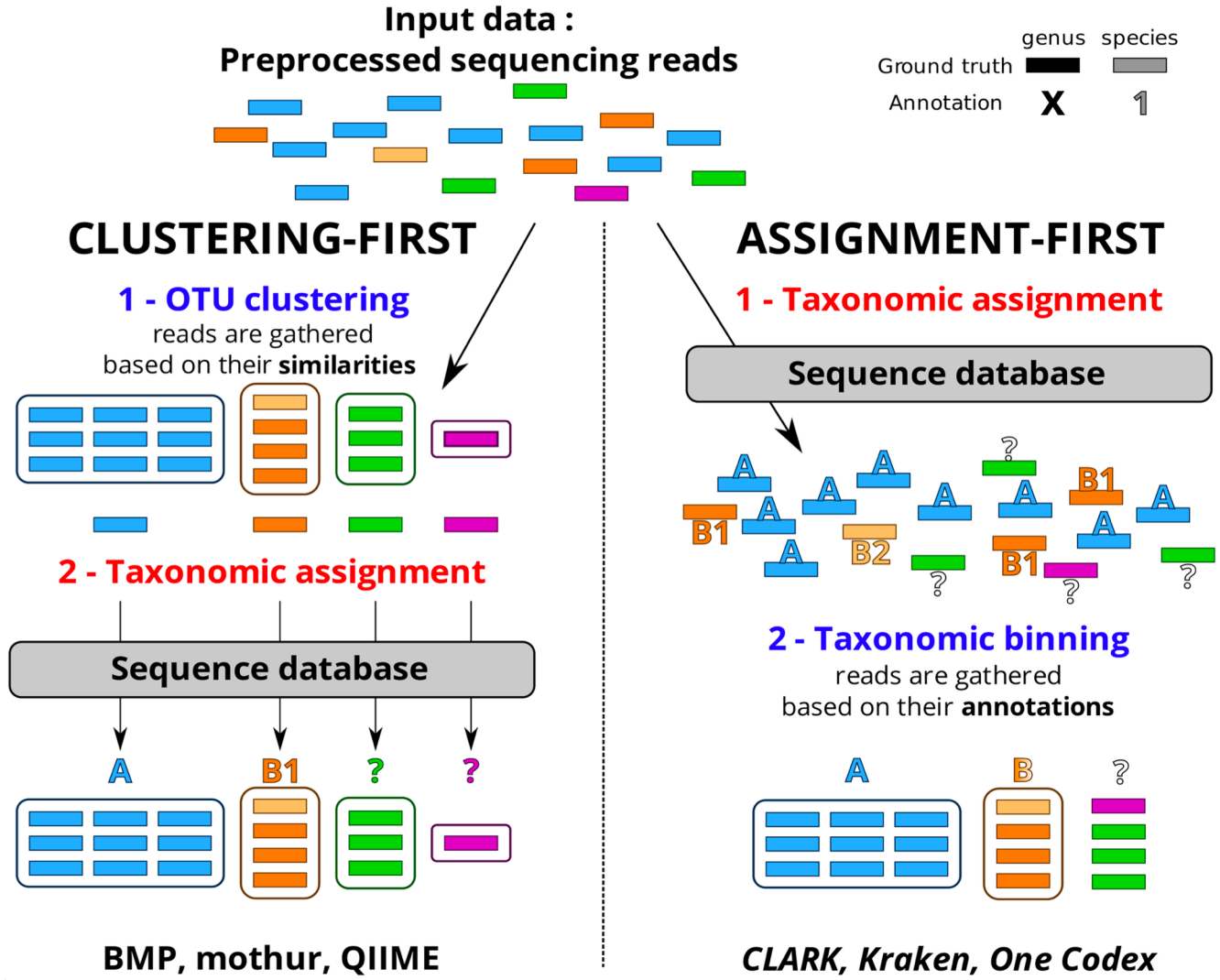
Community profiling - amplicon

Amplicon sequencing workflow



Community profiling - amplicon

OUT clustering

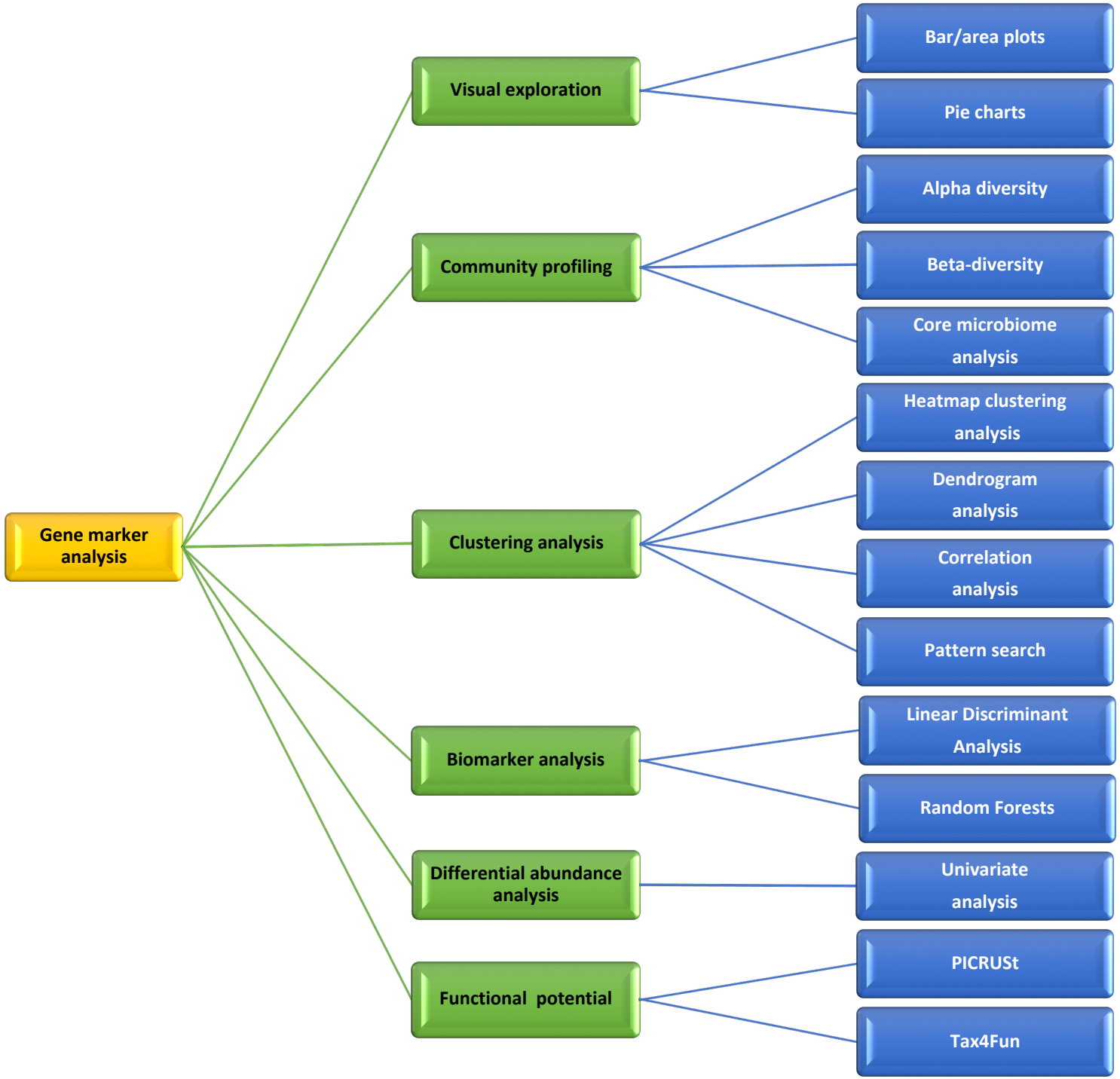


Community profiling - amplicon

OUT table

OUT # Samples Taxonomic linkage

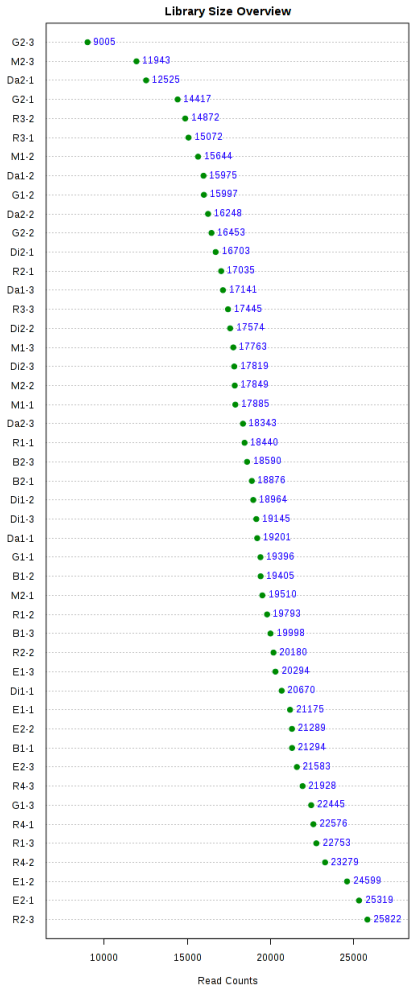
# Constructed from biom file									
#OTU ID	A1	A2	B1	B2	C1	C2	D1	D2	ConsensusLineage
denovo0	1	0	0	0	0	0	0	0	k__Bacteria
denovo1	0	1	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__
denovo2	1	0	1	0	0	1	0	0	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides
denovo3	0	0	0	0	0	2	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Dialister; s__
denovo4	0	1	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus
denovo5	2	0	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira; s__
denovo6	0	0	0	0	1	1	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae
denovo7	0	0	0	0	3	1	10	11	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__; s__
denovo8	1	7	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia; s__
denovo9	0	0	0	1	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae
denovo10	1	0	0	2	0	1	1	0	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Desulfovibrionales; f__Desulfovibrionaceae; g__; s__
denovo11	0	0	0	0	0	0	0	3	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__Finegoldia; s__
denovo12	0	0	0	0	0	0	0	1	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales
denovo13	0	0	0	0	0	1	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae
denovo14	12	13	6	13	121	58	1	12	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Dialister; s__
denovo15	30	16	0	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae
denovo16	0	0	0	1	0	0	0	0	k__Bacteria; p__Firmicutes; c__Bacilli
denovo17	8	4	0	3	1	0	1	2	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales
denovo18	0	0	1	0	0	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales
denovo19	0	0	0	0	1	0	0	0	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales

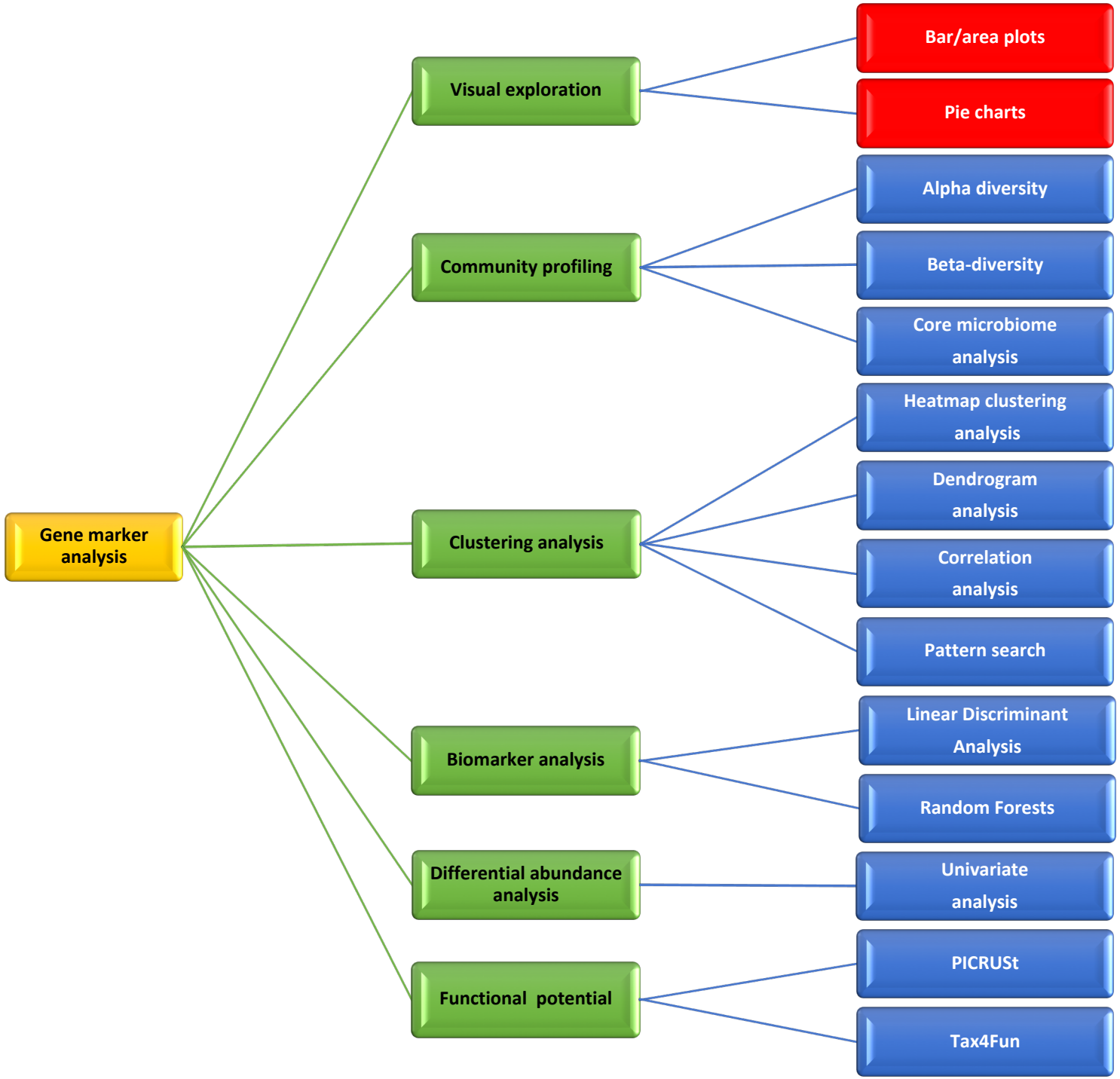

Gene marker analysis
Visual exploration
Community profiling
Clustering analysis
Biomarker analysis
Differential abundance analysis
Functional potential
Bar/area plots
Pie charts
Alpha diversity
Beta-diversity
Core microbiome analysis
Heatmap clustering analysis
Dendrogram analysis
Correlation analysis
Pattern search
Linear Discriminant Analysis
Random Forests
Univariate analysis
PICRUSt
Tax4Fun

Community profiling – analysis and visualization

Input OUT table + metadata

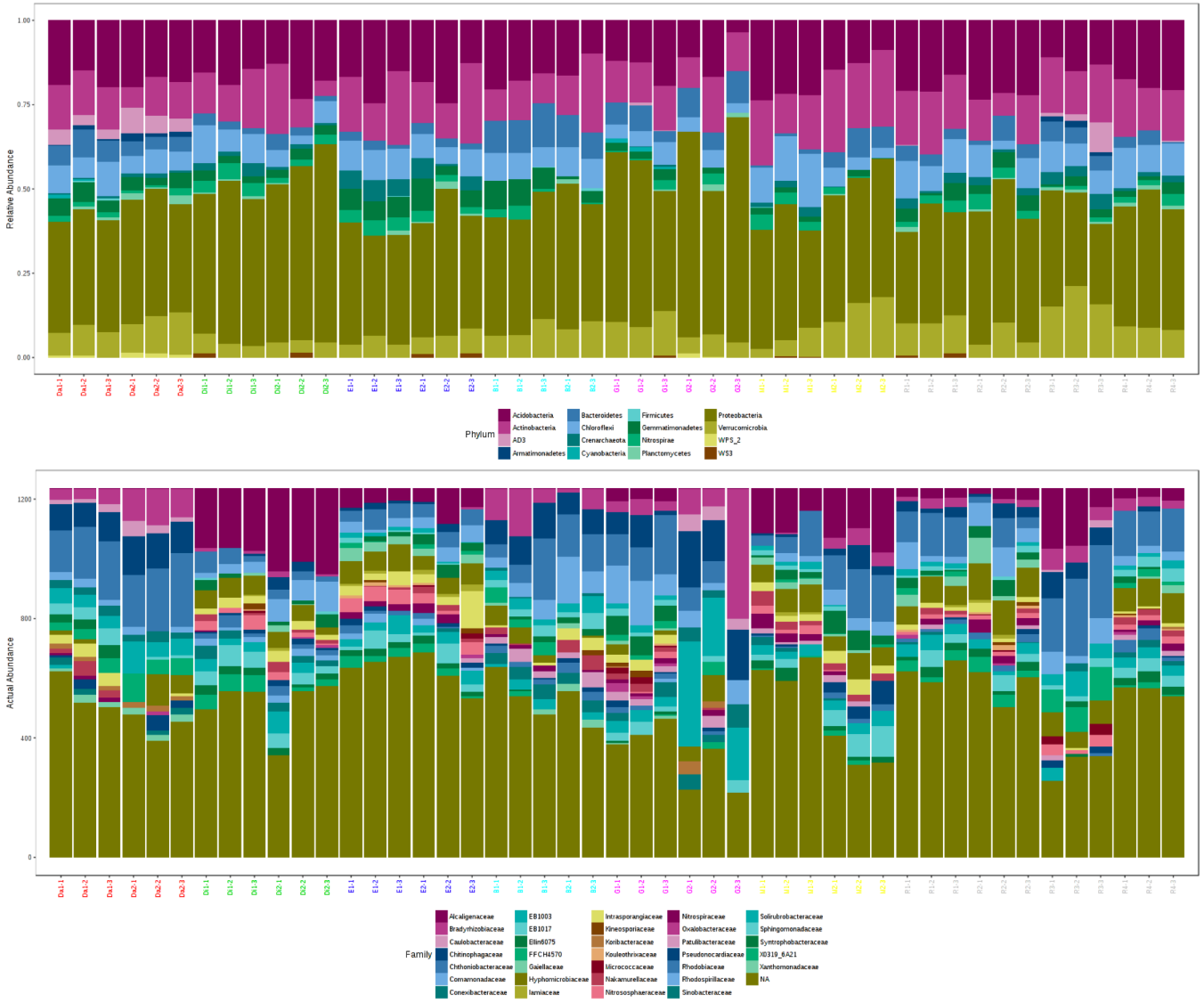
Data type:	OTU abundance table
File format:	biom
OTU annotation:	Greengenes
OTU number:	13931
OTU with ≥ 2 counts:	2942
Sample number:	47
Number of experimental factors:	6
Total read counts:	880232
Average counts per sample:	18728
Maximum counts per sample:	25822
Minimum counts per sample:	9005

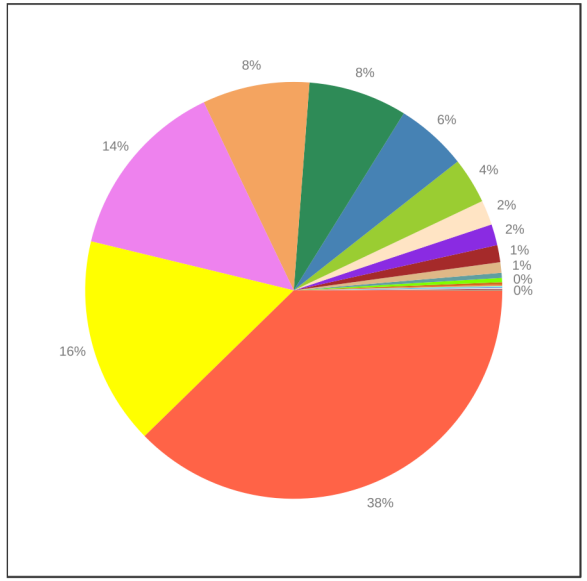


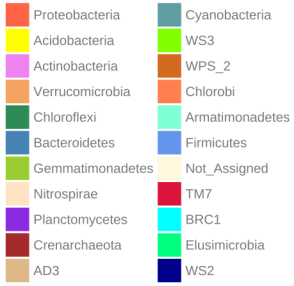

Gene marker analysis
Visual exploration
Community profiling
Clustering analysis
Biomarker analysis
Differential abundance analysis
Functional potential
Bar/area plots
Pie charts
Alpha diversity
Beta-diversity
Core microbiome analysis
Heatmap clustering analysis
Dendrogram analysis
Correlation analysis
Pattern search
Linear Discriminant Analysis
Random Forests
Univariate analysis
PICRUSt
Tax4Fun

Community profiling – analysis and visualization

Plots and charts



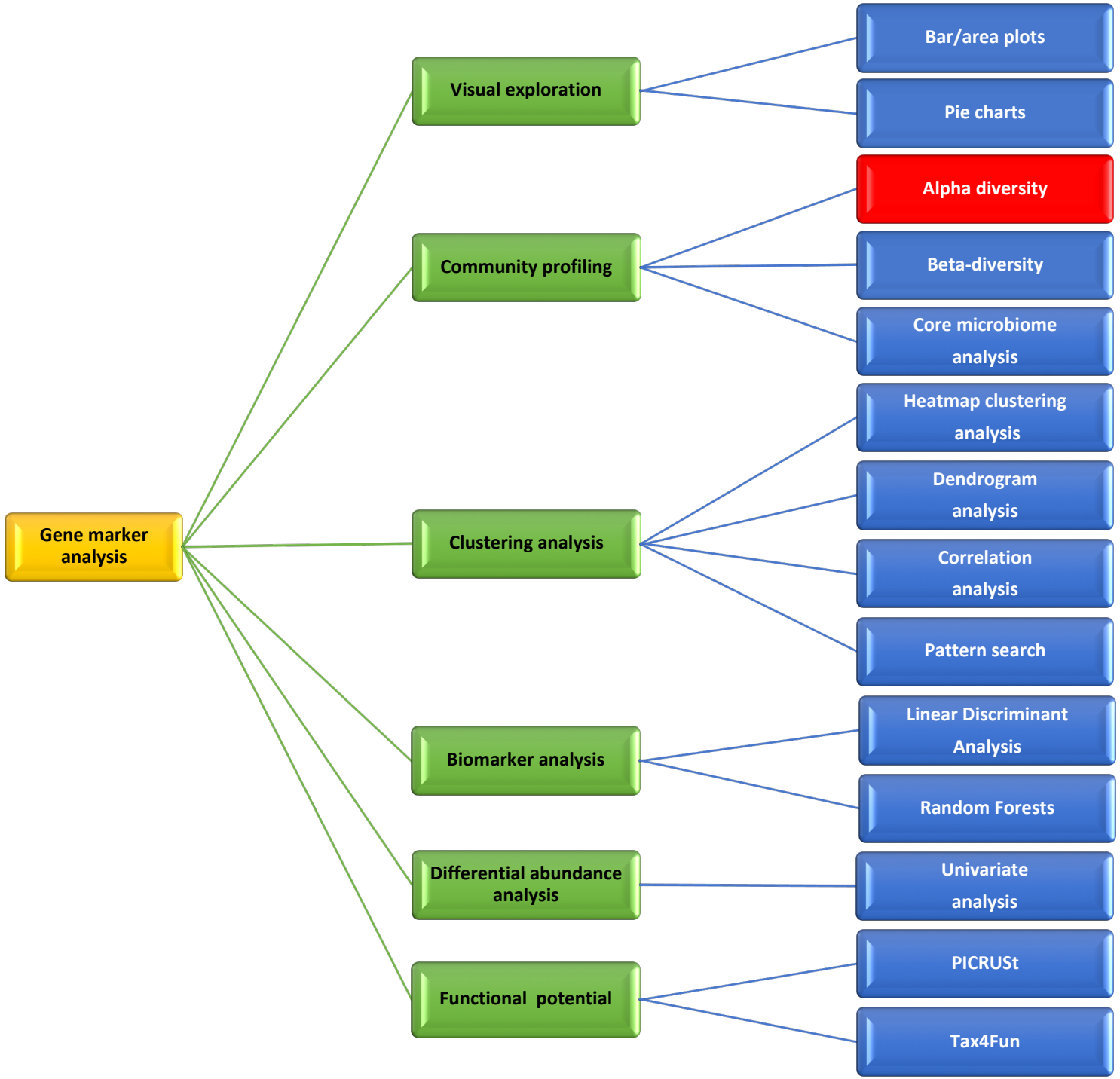




Community profiling – analysis and visualization

Alpha, beta and gamma diversity (introduced by R.H. Whittaker in 1972)

- Alpha diversity
  - species richness (number of taxa) within a single microbial ecosystem.
  - *How many different microbial species can be detected in one sample?*
- Beta diversity
  - diversity in microbial community between different environments (difference in taxonomic abundance profiles from different samples).
  - *How different is the microbial composition in one environment compared to another?*
- Gamma diversity
  - is a measure of the overall number of species (sum) within a landscape.
  - *How many different microbial species can be detected in the region of interest?*


Gene marker analysis
Visual exploration
Community profiling
Clustering analysis
Biomarker analysis
Differential abundance analysis
Functional potential
Bar/area plots
Pie charts
Alpha diversity
Beta-diversity
Core microbiome analysis
Heatmap clustering analysis
Dendrogram analysis
Correlation analysis
Pattern search
Linear Discriminant Analysis
Random Forests
Univariate analysis
PICRUSt
Tax4Fun

Community profiling – analysis and visualization

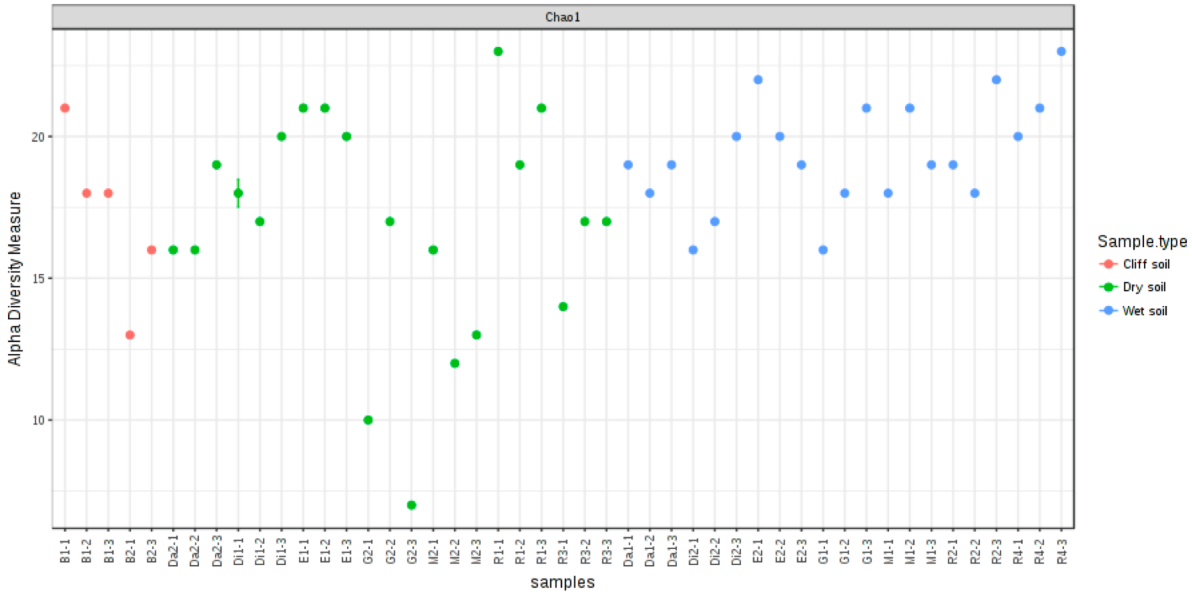
Alpha diversity

- Observed richness and abundance
  - the number of observed unique OUTs in a sample
- Chao1 and CHA estimators
  - Accounts for observed OTUs in addition to low abundance species (singletons and doubletons)
  - Estimate richness – not diversity
- Shannon, Simpson, Fischer index
  - Take into account both richness, abundance and evenness
  - Estimate diversity

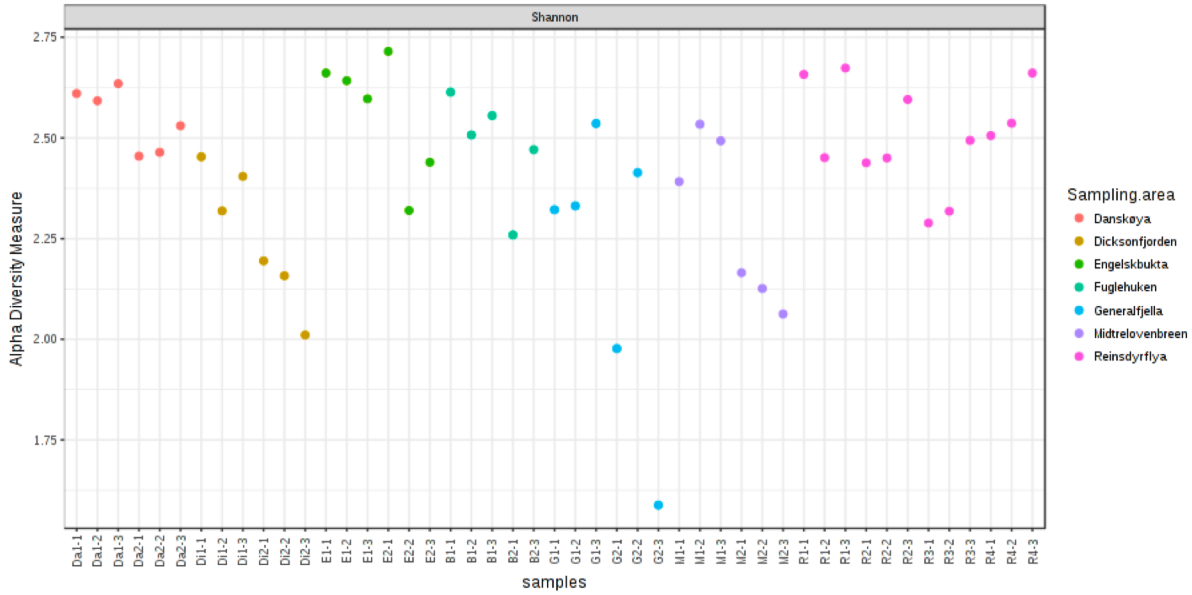
Community profiling – analysis and visualization

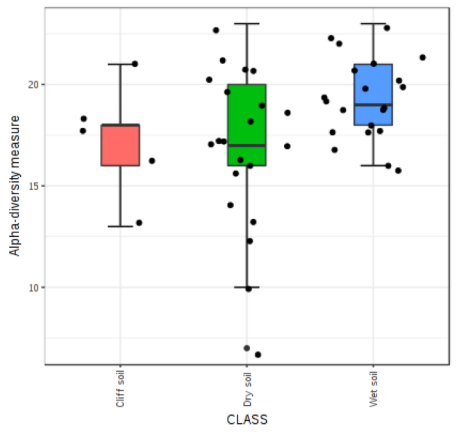
Community profiling - Alphadiversity

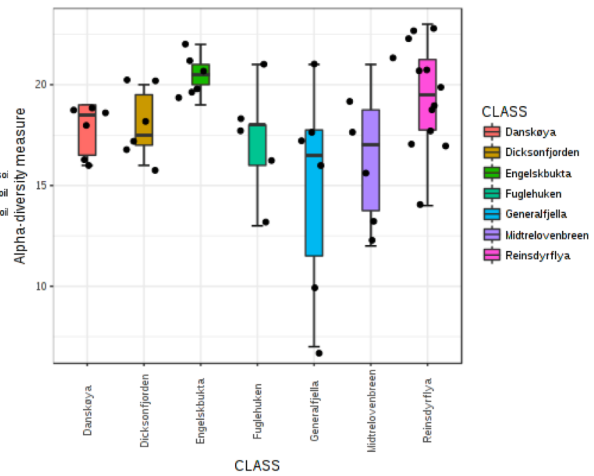
Chao1

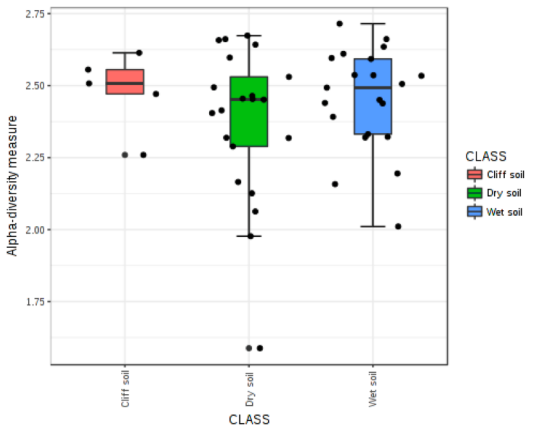


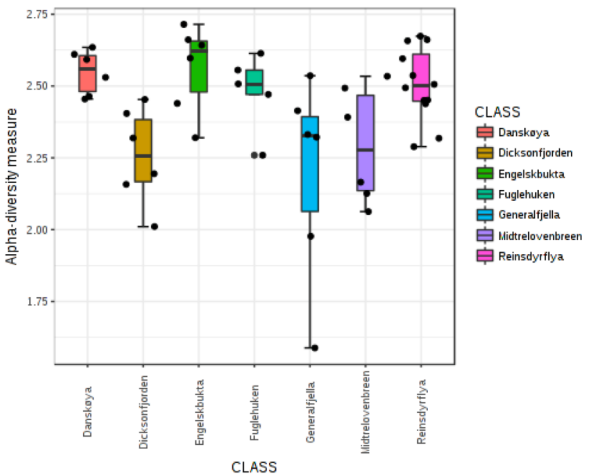
Shannon

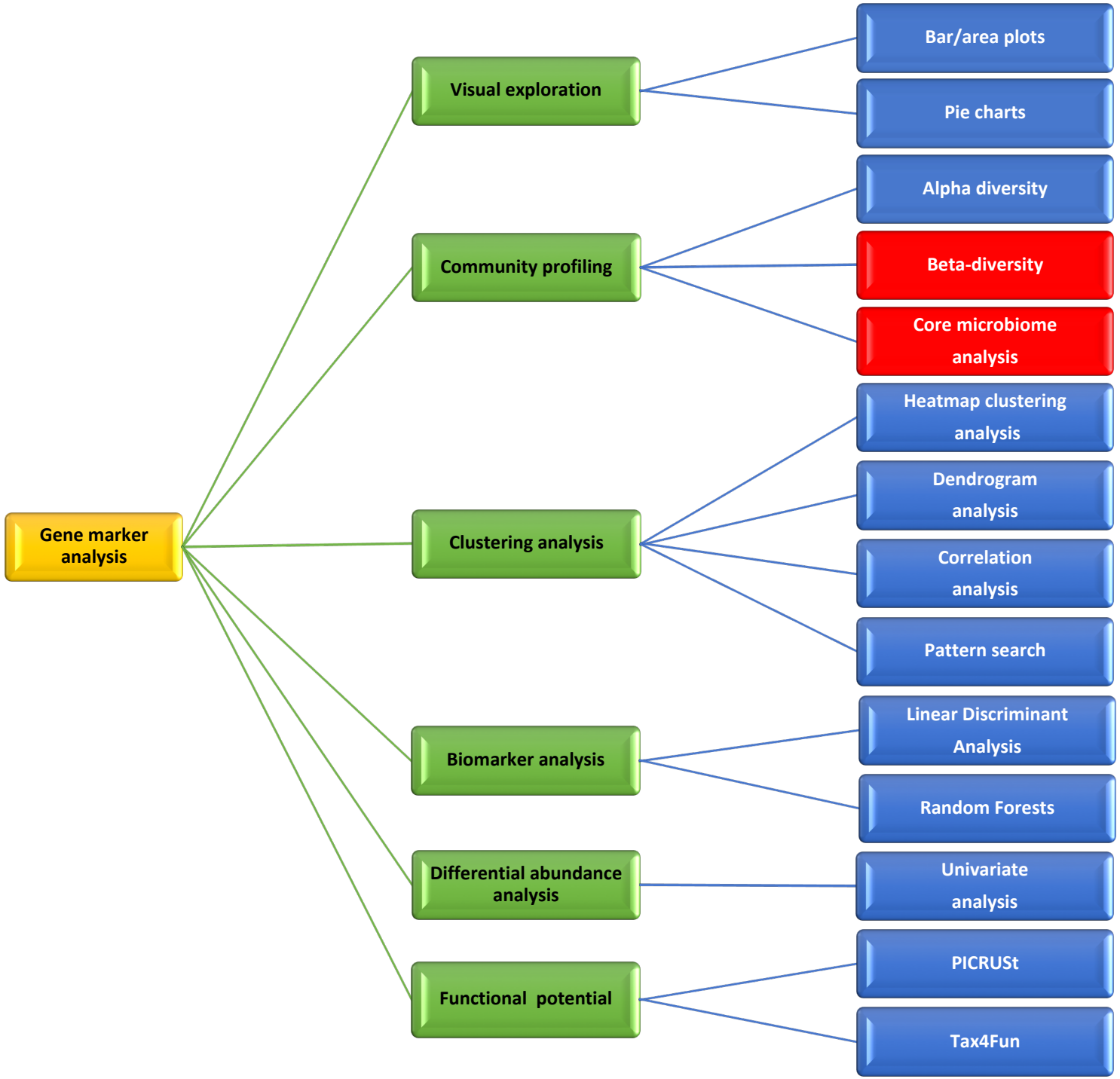











Gene marker analysis
Visual exploration
Community profiling
Clustering analysis
Biomarker analysis
Differential abundance analysis
Functional potential
Bar/area plots
Pie charts
Alpha diversity
Beta-diversity
Core microbiome analysis
Heatmap clustering analysis
Dendrogram analysis
Correlation analysis
Pattern search
Linear Discriminant Analysis
Random Forests
Univariate analysis
PICRUSt
Tax4Fun

Community profiling – analysis and visualization

Beta diversity

- Diversity between communities
- Expressed as the ecological distance between two communities

- Bray-Curtis dissimilarity
  - a statistic used to quantify the compositional dissimilarity between two different sites, based on counts at each site.

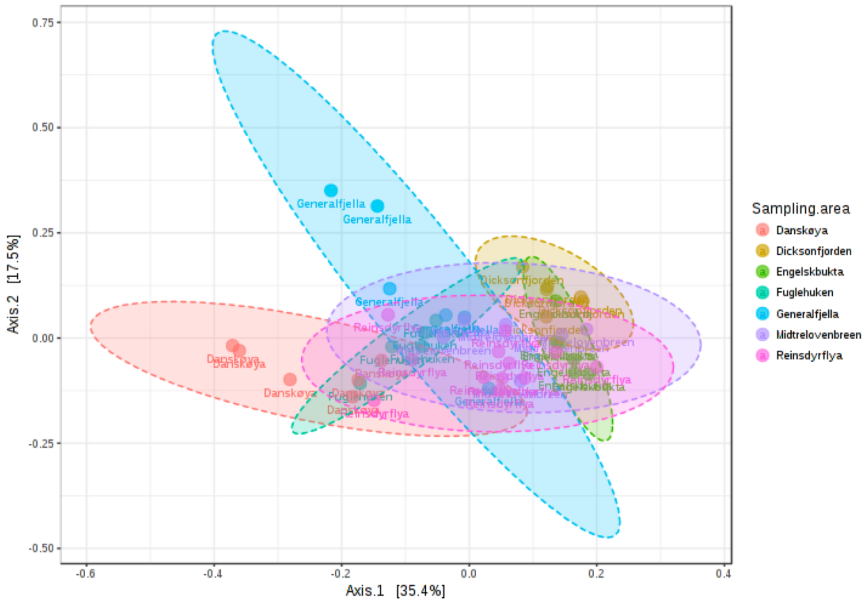
- Jaccard dissimilarity
  - a statistic used for comparing the similarity and diversity of sample sets.

- Unifrac (weighted/unweighted)
  - differs from dissimilarity measures such as Bray-Curtis dissimilarity in that it incorporates information on the relative relatedness of community members by incorporating phylogenetic distances between observed organisms in the computation. Both weighted (quantitative) and unweighted (qualitative) variants of UniFrac are widely used in microbial ecology, where the former accounts for abundance of observed organisms, while the latter only considers their presence or absence.

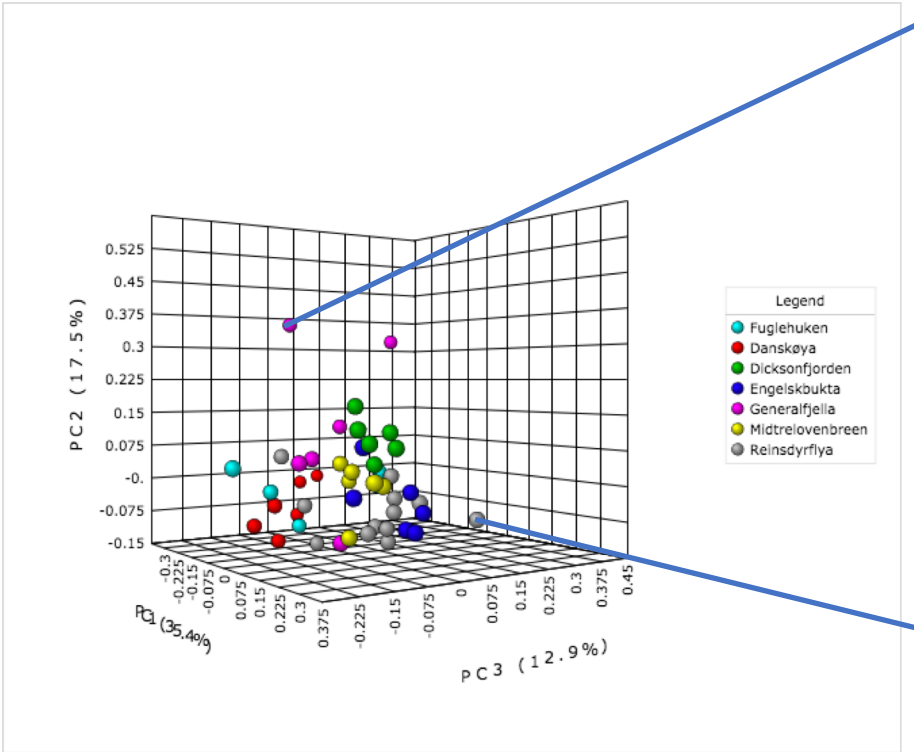
Community profiling – analysis and visualization

Community profiling - Betadiversity

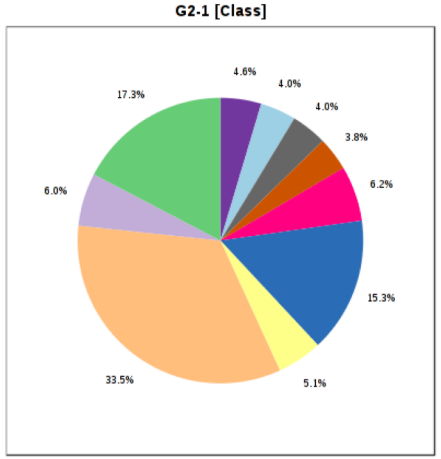
Principal Coordinates Analysis (PCoA)
Bray-Curtis index
Permutational MASNOVA statistics

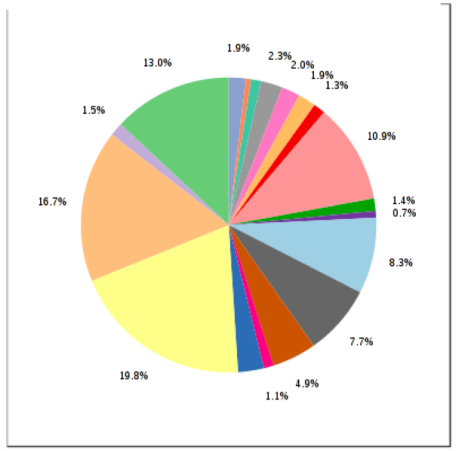


2D PCoA



3D PCoA

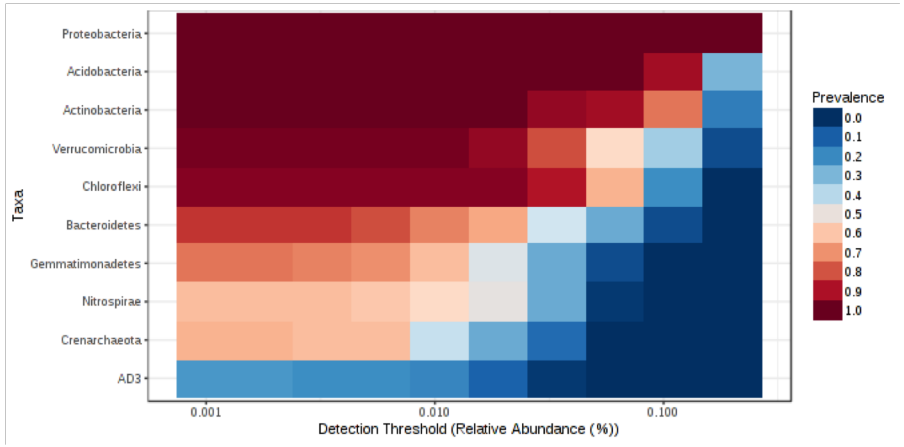




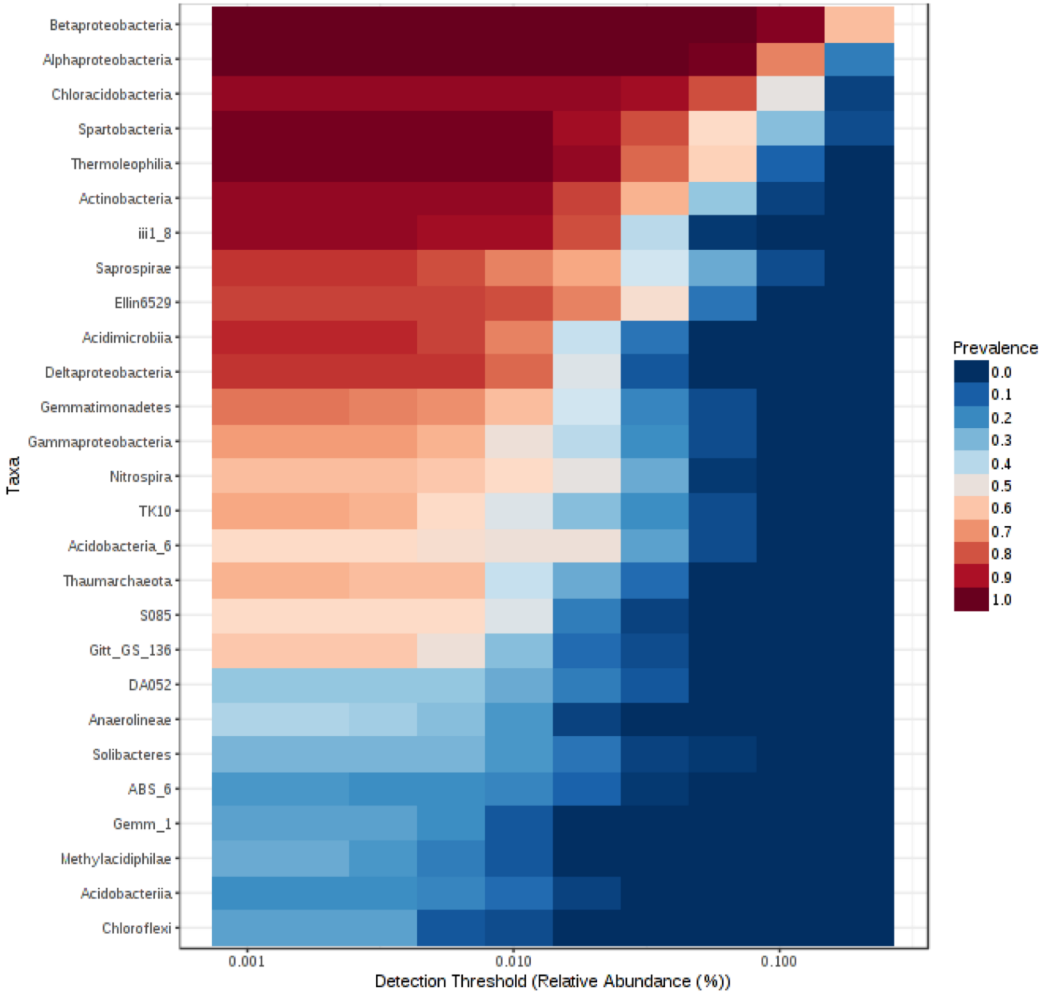
Community profiling – analysis and visualization

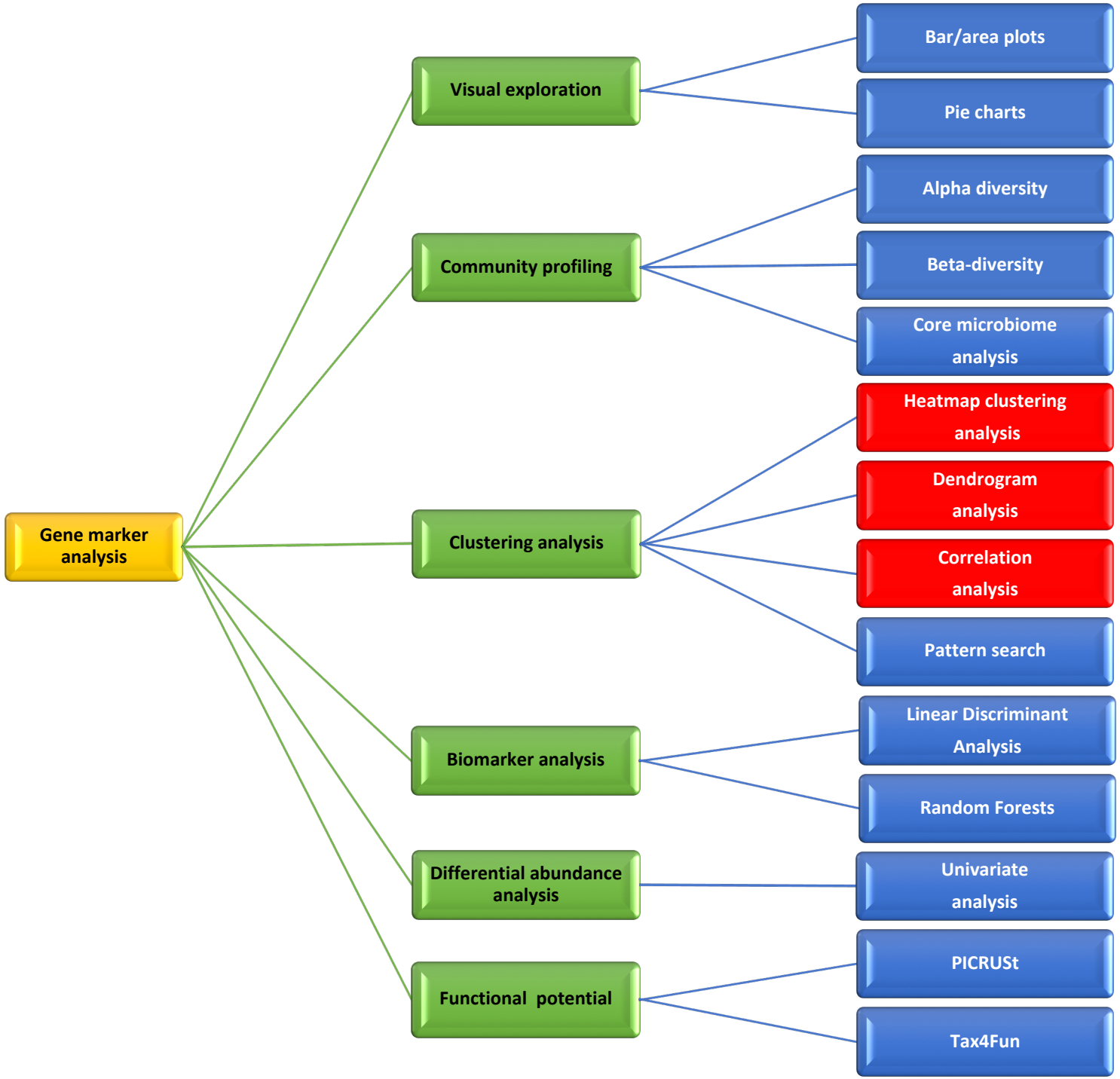
Core microbiome

Phylum



Class

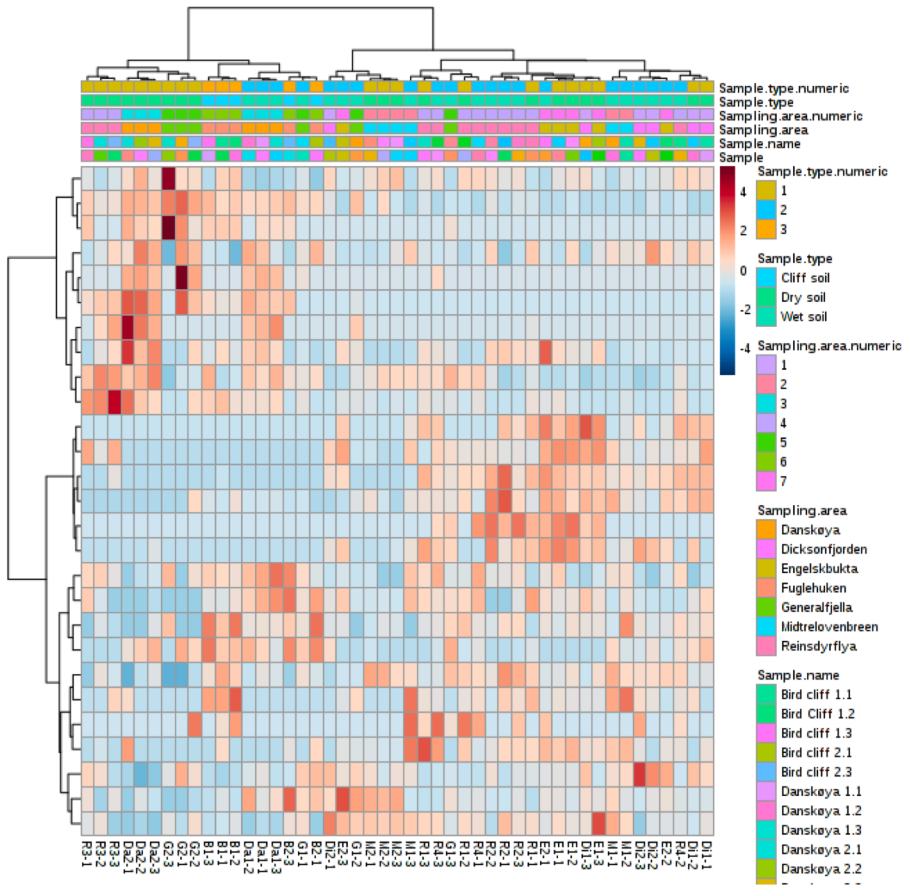



Gene marker analysis
Visual exploration
Community profiling
Clustering analysis
Biomarker analysis
Differential abundance analysis
Functional potential
Bar/area plots
Pie charts
Alpha diversity
Beta-diversity
Core microbiome analysis
Heatmap clustering analysis
Dendrogram analysis
Correlation analysis
Pattern search
Linear Discriminant Analysis
Random Forests
Univariate analysis
PICRUSt
Tax4Fun

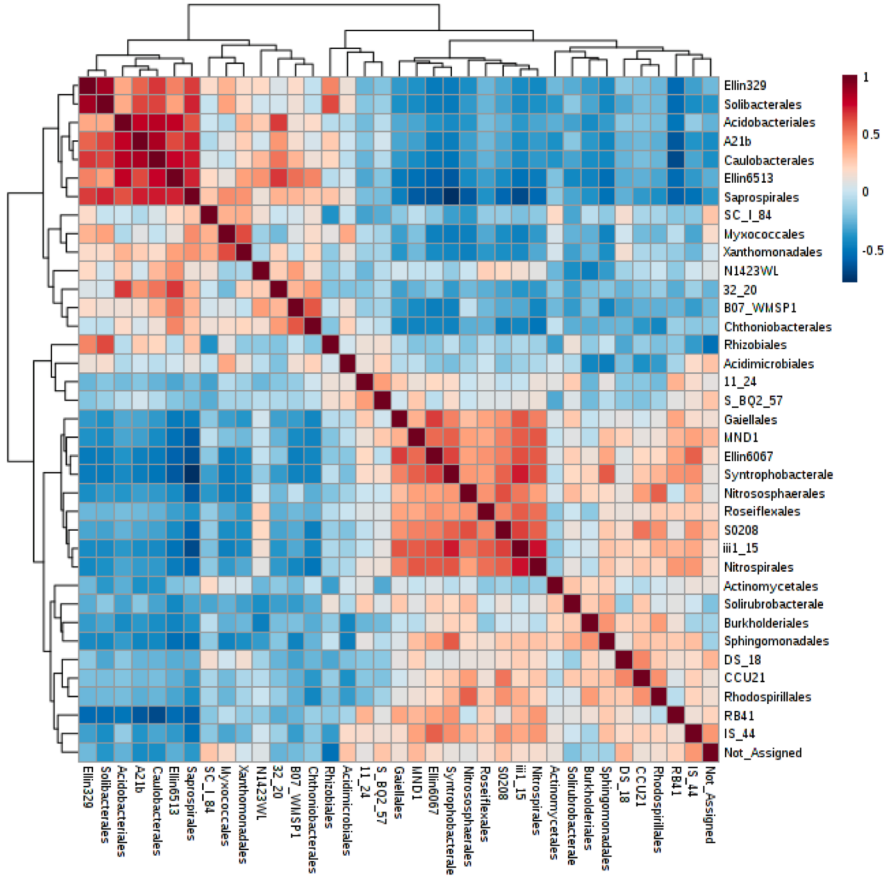
Community profiling – analysis and visualization

Clustering analysis

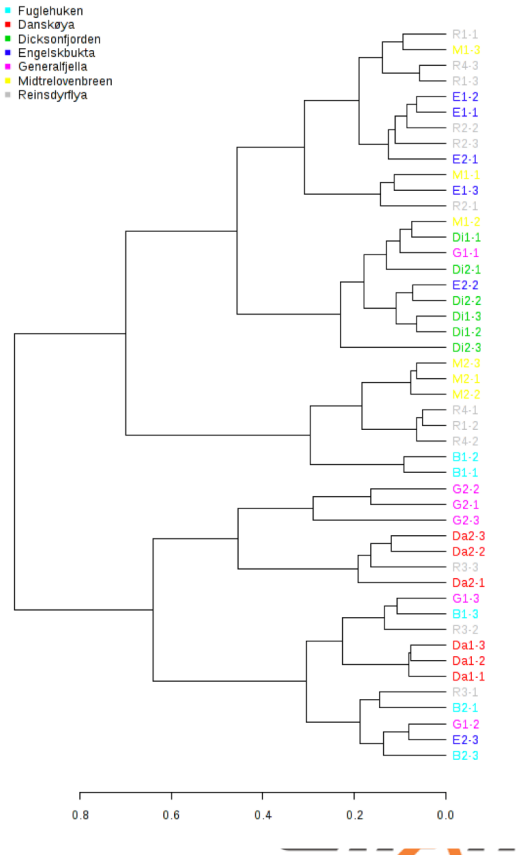
Heatmap

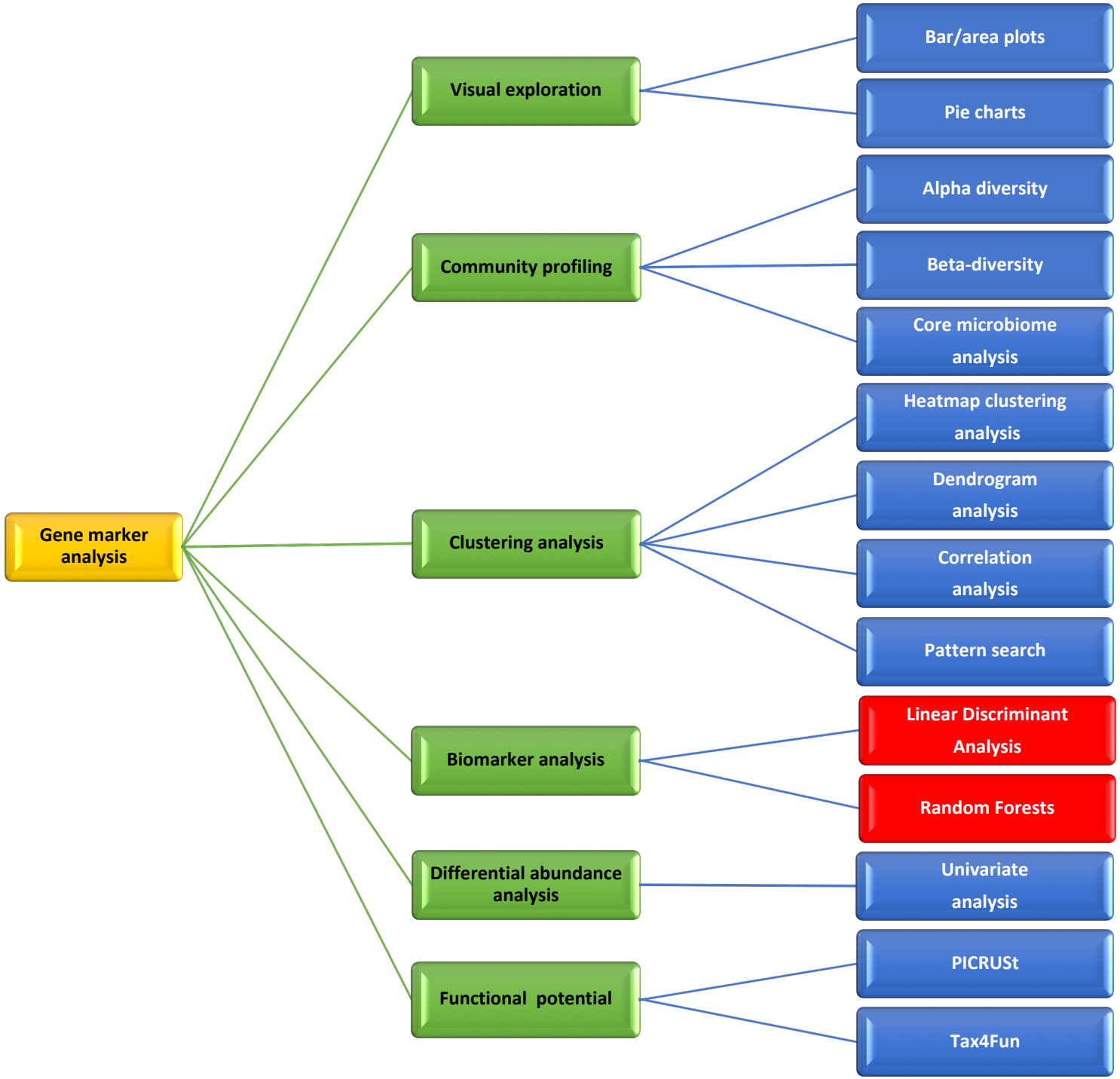


Correlation analysis



Dendrogram

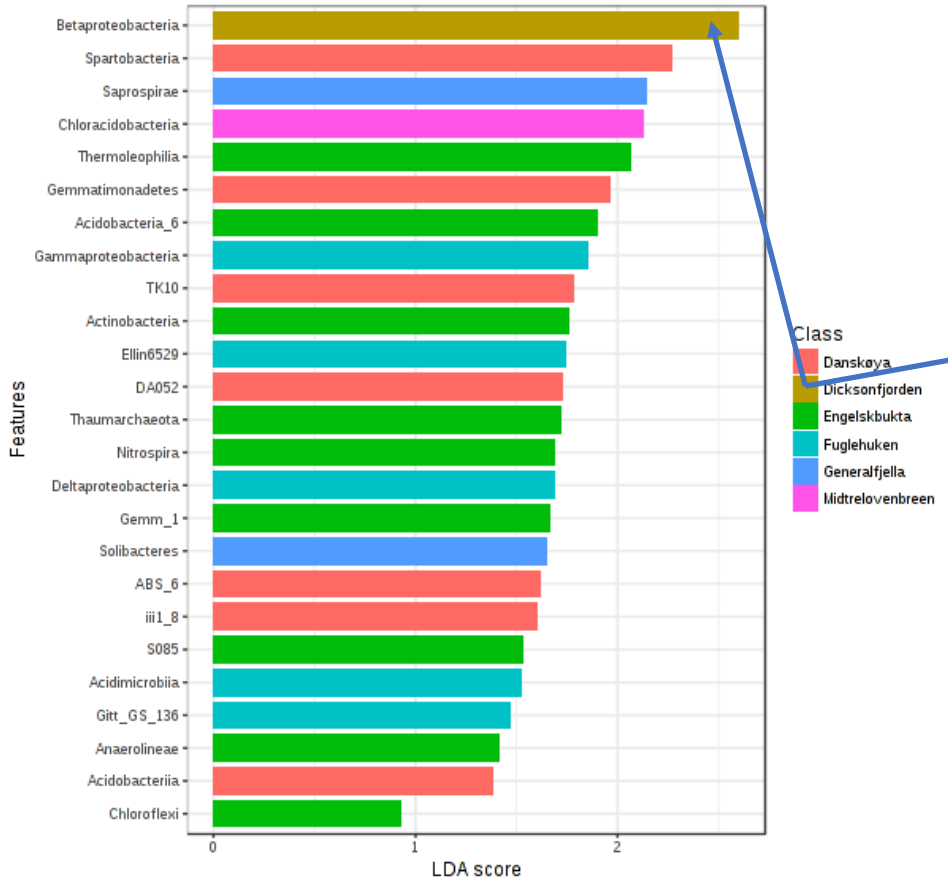



Gene marker analysis
Visual exploration
Community profiling
Clustering analysis
Biomarker analysis
Differential abundance analysis
Functional potential
Bar/area plots
Pie charts
Alpha diversity
Beta-diversity
Core microbiome analysis
Heatmap clustering analysis
Dendrogram analysis
Correlation analysis
Pattern search
Linear Discriminant Analysis
Random Forests
Univariate analysis
PICRUSt
Tax4Fun

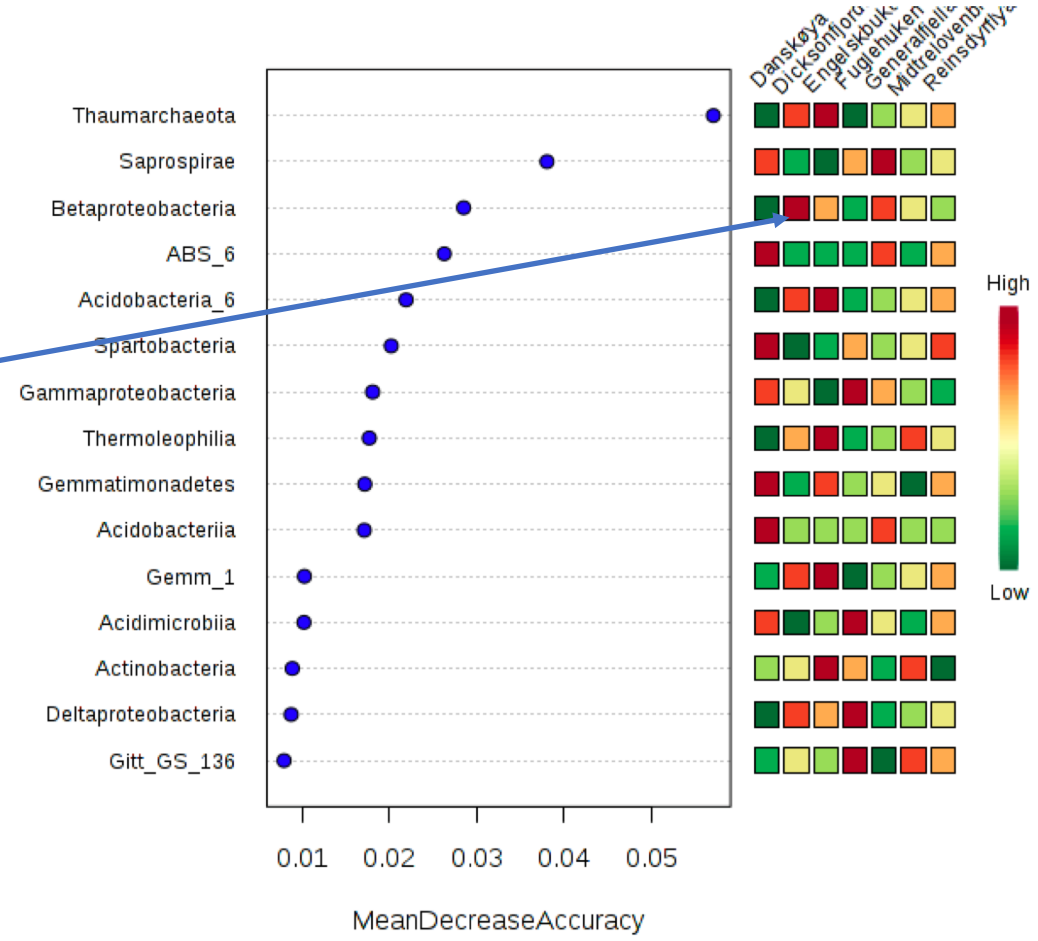
Community profiling – analysis and visualization

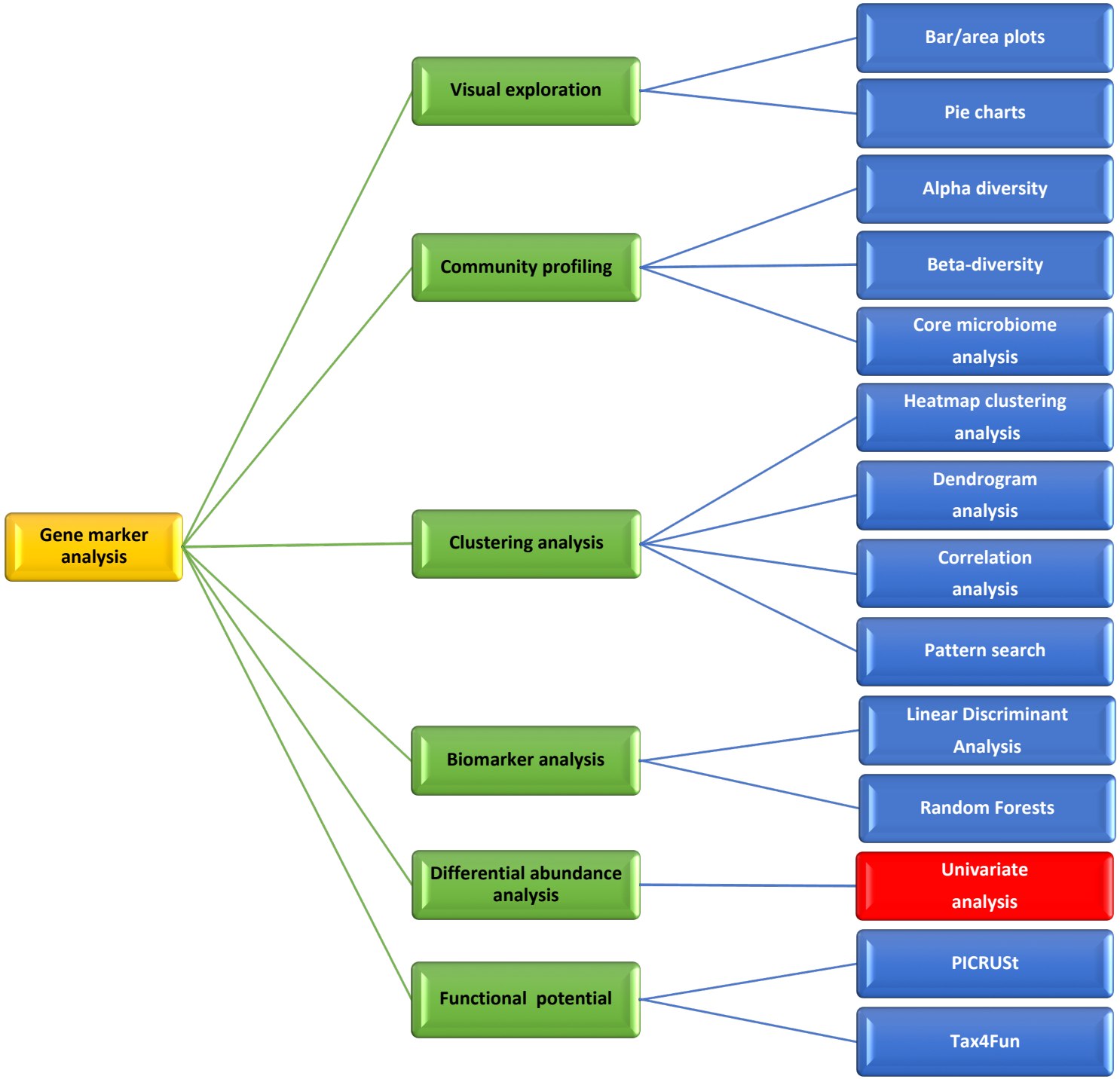
Biomarker analysis

Linear Discriminant Analysis (LDA) Effect Size (LEfSe)



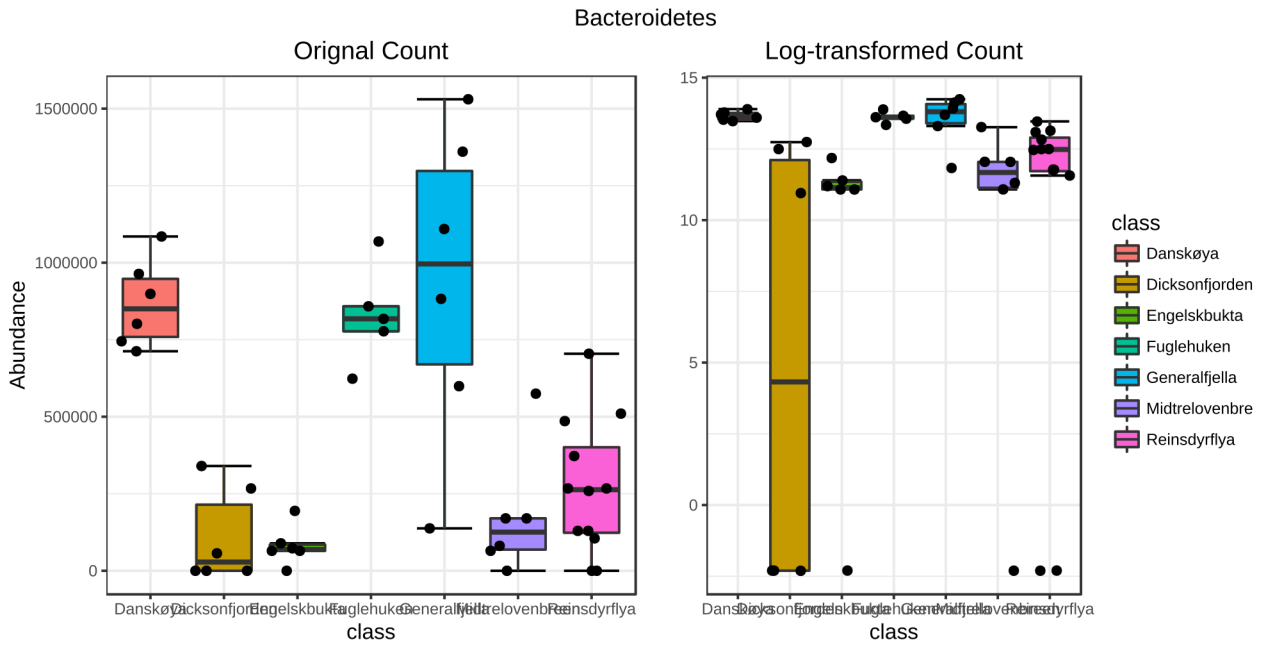
Random Forests

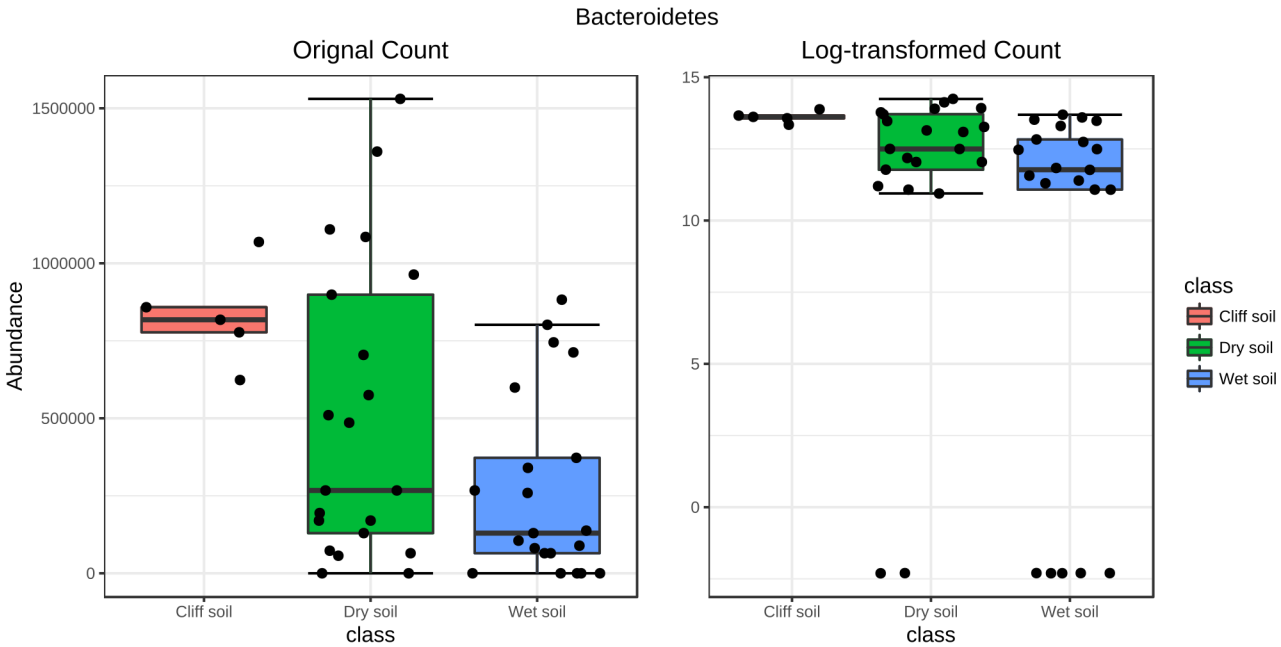



Gene marker analysis
Visual exploration
Community profiling
Clustering analysis
Biomarker analysis
Differential abundance analysis
Functional potential
Bar/area plots
Pie charts
Alpha diversity
Beta-diversity
Core microbiome analysis
Heatmap clustering analysis
Dendrogram analysis
Correlation analysis
Pattern search
Linear Discriminant Analysis
Random Forests
Univariate analysis
PICRUSt
Tax4Fun

Community profiling – analysis and visualization

Differential abundance analysis





Community profiling – amplicon vs shotgun

	16S amplicon	Shotgun
Analysis of large number of samples	pro	con
Depth - resolution	pro	con
Computational resources (and skills)	pro	con
Price	pro	con
PCR amplification bias	con	pro
Discovery of new bacterial genes and genomes	con	pro
Phylogenetic information	con	pro
Functional information	con	pro
Simultaneous study of several domains	con	pro

elixir NORWAY