FROGS Func

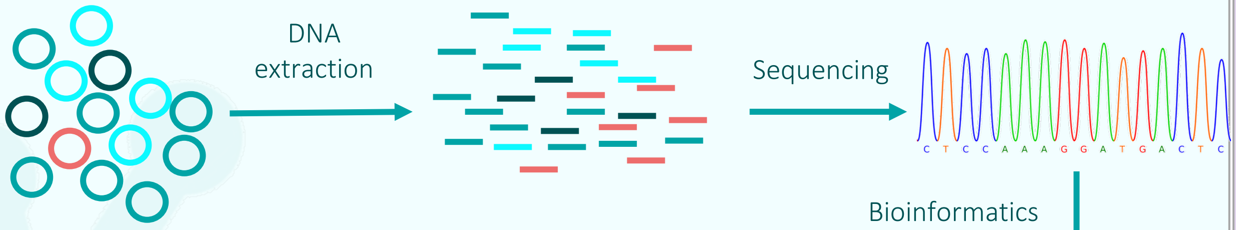
Which metabolic functions are present in the environment?

Lucas Auer, Gabryelle Agoutin,
Maria Bernard, Géraldine Pascal,
Maëlle Pomiès & Olivier Rué

Concepts

Metabarcoding principle

Determine the diversity of a environment by amplification and sequencing of a genetic marker.



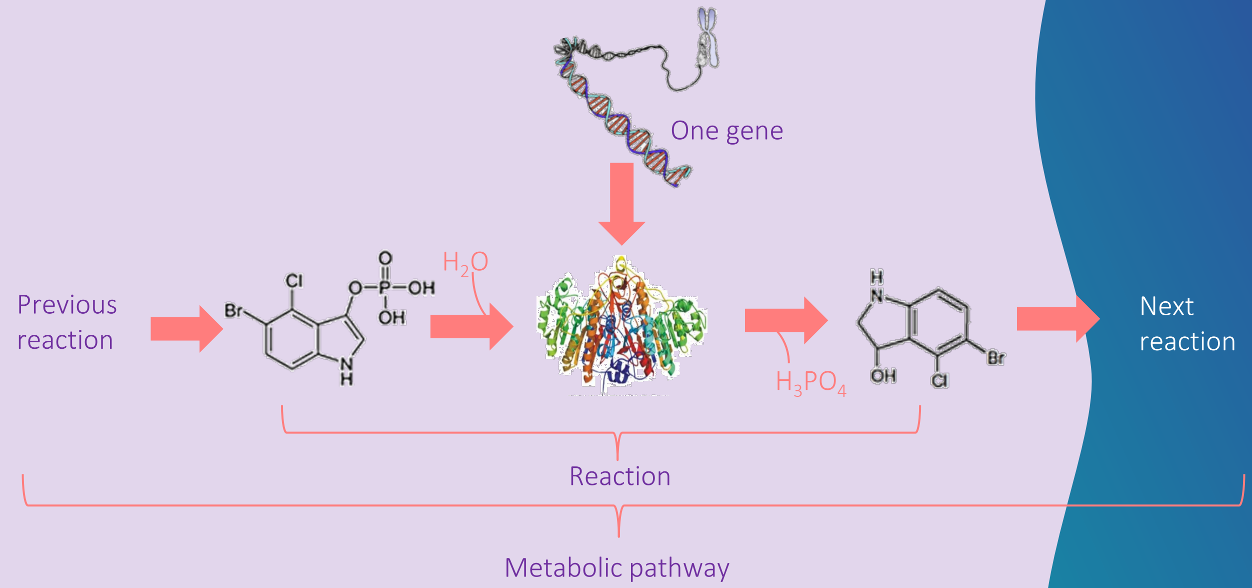
Bioinformatics analysis with FROGS

Abundancy table displaying microbial diversity per samples

	Taxonomic affiliation	Sample1	Sample2	Sample3
ASV1	Species A	3500	6300	210
ASV2	Species B	0	460	36
ASV3	Species C	400	700	500

Functional inference

Assuming that an organism has a metabolic function by the presence in the organism's genome of a known sequence having that function.



What are the metabolic functions present in a microbial community?

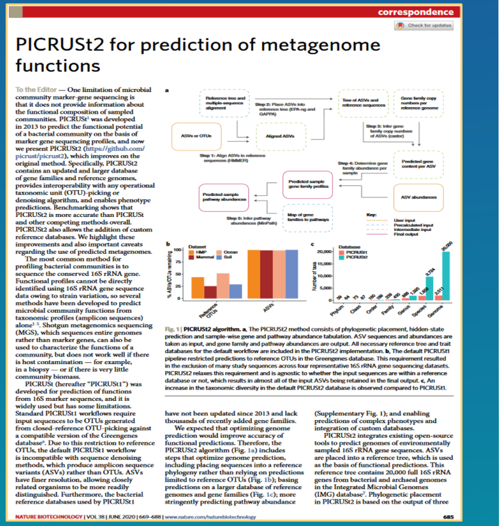
Based on PICRUSt2

PICRUSt (Phylogenetic investigation of communities by reconstruction of unobserved states) is an open-source tool.

It is a software that can predict functional abundances, based only on marker gene sequences.

PICRUSt2 is composed of 4 python applications.

No graphic interface exists to run PICRUSt2 for non-expert users.



correspondence

Check for updates

PICRUSt2 for prediction of metagenome functions

To the Editor — One limitation of microbial community marker-gene sequencing is that it does not provide information about the functional composition of sampled communities. PICRUSt[1] was developed in 2013 to predict the functional potential of a bacterial community on the basis of marker gene sequencing profiles, and now we present PICRUSt2 (https://github.com/picrust/picrust2), which improves on the original method. Specifically, PICRUSt2 contains an updated and larger database of gene families and reference genomes, provides interoperability with any operational taxonomic unit (OTU)-picking or denoising algorithm, and enables phenotype predictions. Benchmarking shows that PICRUSt2 is more accurate than PICRUSt and other competing methods overall. PICRUSt2 also allows the addition of custom reference databases. We highlight these improvements and also important caveats regarding the use of predicted metagenomes.

The most common method for profiling bacterial communities is to sequence the conserved 16S rRNA gene. Functional profiles cannot be directly identified using 16S rRNA gene sequence data owing to strain variation, so several methods have been developed to predict microbial community functions from taxonomic profiles (amplicon sequences) alone[1–5]. Shotgun metagenomics sequencing (MGS), which sequences entire genomes rather than marker genes, can also be used to characterize the functions of a community, but does not work well if there is host contamination — for example, in a biopsy — or if there is very little community biomass.

PICRUSt (hereafter "PICRUSt1") was developed for prediction of functions from 16S marker sequences, and it is widely used but has some limitations. Standard PICRUSt1 workflows require input sequences to be OTUs generated from closed-reference OTU-picking against a compatible version of the Greengenes database[6]. Due to this restriction to reference OTUs, the default PICRUSt1 workflow is incompatible with sequence denoising methods, which produce amplicon sequence variants (ASVs) rather than OTUs. ASVs have finer resolution, allowing closely related organisms to be more readily distinguished. Furthermore, the bacterial reference databases used by PICRUSt1 have not been updated since 2013 and lack thousands of recently added gene families.

We expected that optimizing genome prediction would improve accuracy of functional predictions. Therefore, the PICRUSt2 algorithm (Fig. 1a) includes steps that optimize genome prediction, including placing sequences into a reference phylogeny rather than relying on predictions limited to reference OTUs (Fig. 1b); basing predictions on a larger database of reference genomes and gene families (Fig. 1c); more stringently predicting pathway abundance (Supplementary Fig. 1); and enabling predictions of complex phenotypes and integration of custom databases.

PICRUSt2 integrates existing open-source tools to predict genomes of environmentally sampled 16S rRNA gene sequences. ASVs are placed into a reference tree, which is used as the basis of functional predictions. This reference tree contains 20,000 full 16S rRNA genes from bacterial and archaeal genomes in the Integrated Microbial Genomes (IMG) database[7]. Phylogenetic placement in PICRUSt2 is based on the output of three

Fig. 1 | **PICRUSt2 algorithm. a**, The PICRUSt2 method consists of phylogenetic placement, hidden-state prediction and sample-wise gene and pathway abundance tabulation. ASV sequences and abundances are taken as input, and gene family and pathway abundances are output. All necessary reference tree and trait databases for the default workflow are included in the PICRUSt2 implementation. **b**, The default PICRUSt1 pipeline restricted predictions to reference OTUs in the Greengenes database. This requirement resulted in the exclusion of many study sequences across four representative 16S rRNA gene sequencing datasets. PICRUSt2 relaxes this requirement and is agnostic to whether the input sequences are within a reference database or not, which results in almost all of the input ASVs being retained in the final output. **c**, An increase in the taxonomic diversity in the default PICRUSt2 database is observed compared to PICRUSt1.

NATURE BIOTECHNOLOGY | VOL 38 | JUNE 2020 | 669–688 | www.nature.com/naturebiotechnology 685

Douglas, G.M., Maffei, V.J., Zaneveld, J.R. *et al.* PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* 38, 685–688 (2020). https://doi.org/10.1038/s41587-020-0548-6

How does it work?

1. Places the ASVs into a reference phylogenetic tree and predicts the number of marker copies in each ASV.

FROGS Func - sequence placement & gene copy nb estimation

2. Estimates the number of copies of each function (gene) in each organism (ASV). Also calculates how many functions (genes) are present in each sample. And determines how many of ASVs there are using the number of copies of the marker.

FROGS Func - functional abundance estimation

3. Calculates how much of each pathway is present in each sample.

FROGS Func - pathways abundance estimation

It runs only on 16S, ITS or 18S

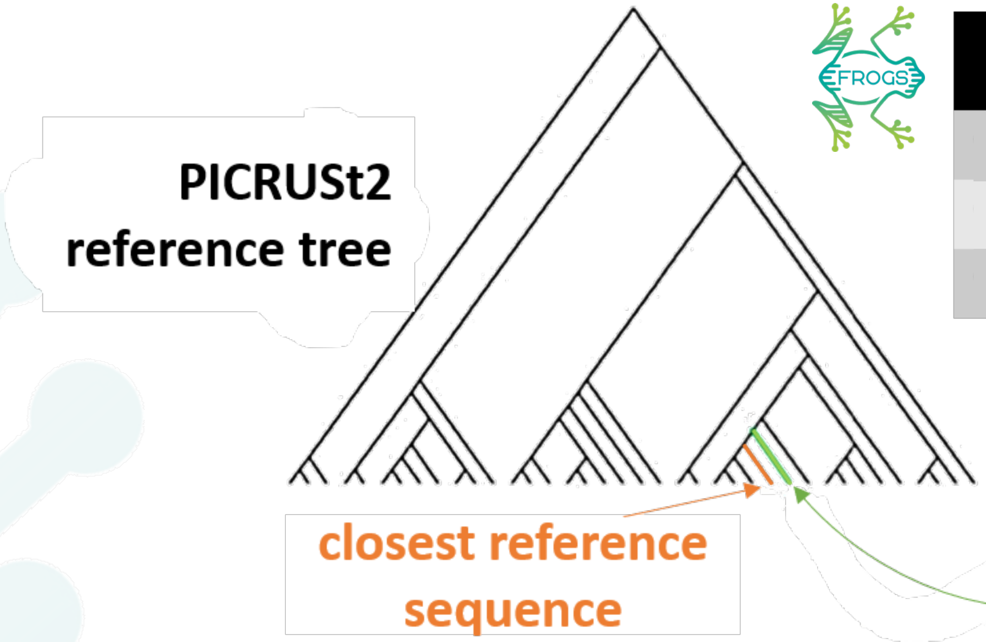
FROGS Func

FROGS Func - sequence placement & gene copy nb estimation

Overview

- FROGS Func - sequence placement & gene copy nb estimation is the first step of PICRUSt2.
- It inserts your studied sequences into a reference tree.
- By default, this reference tree is based on 20,000 16S sequences from genomes in the Integrated Microbial Genomes database.
- The numbers of marker gene copies (16S, ITS or 18S) are predicted in order to enable the normalization of the ASV abundance table.

Overview



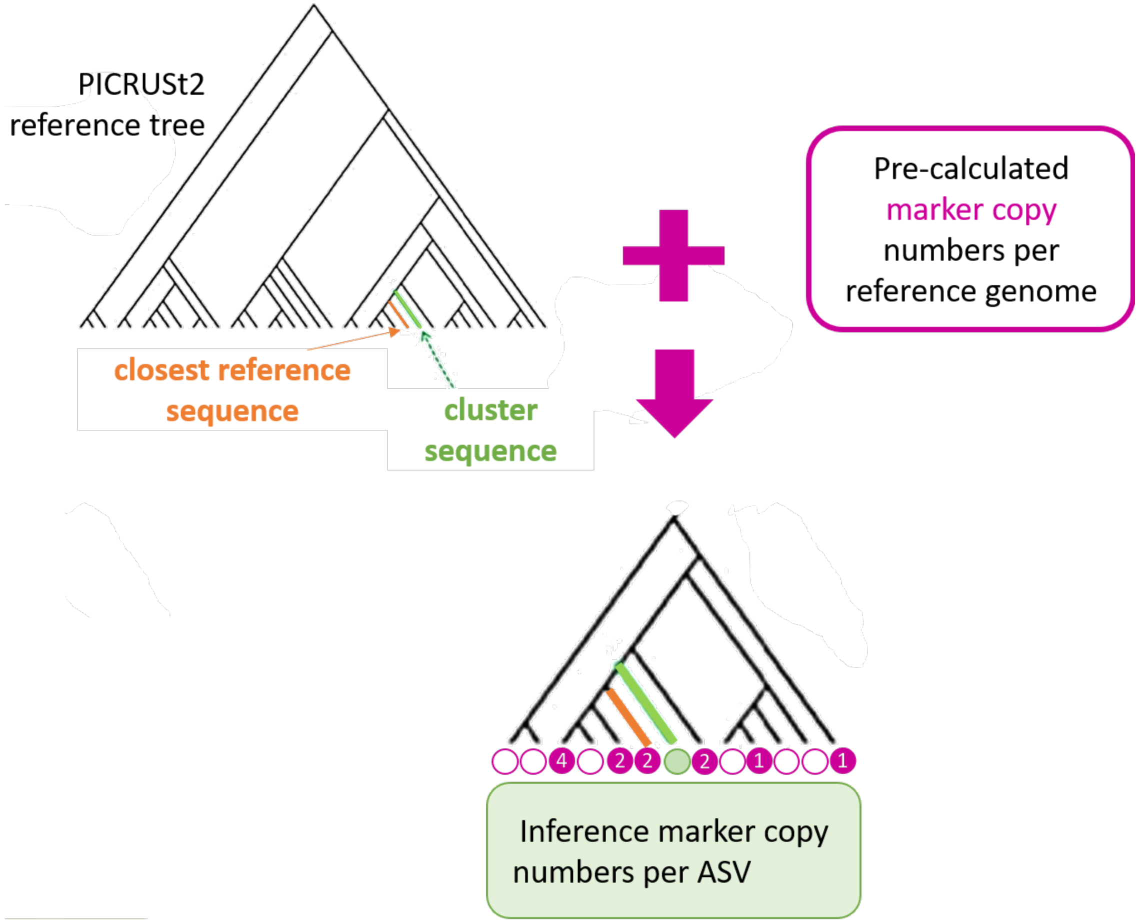
	taxonomic affiliation	fasta sequences	sample1	sample2	sample3
ASV1	species A	>AGGAG..	3500	6300	210
ASV2	species B	>ATGAG..	0	460	36
ASV3	species C	>AGGCT..	400	700	500

cluster sequence

2 input files are required for FROGS Func placeseqs & copynumber analysis:

- **fasta file** of ASV sequences (it can be from FROGS Core 1-Main cluster_filters or itsx step)
- **biom file** of ASV abundances with taxonomic affiliation information (it can be from FROGS Core 1-Main taxonomic affiliation step)

Overview



PICRUSt2 reference tree

closest reference sequence

cluster sequence

Pre-calculated marker copy numbers per reference genome

Inference marker copy numbers per ASV

Practice session

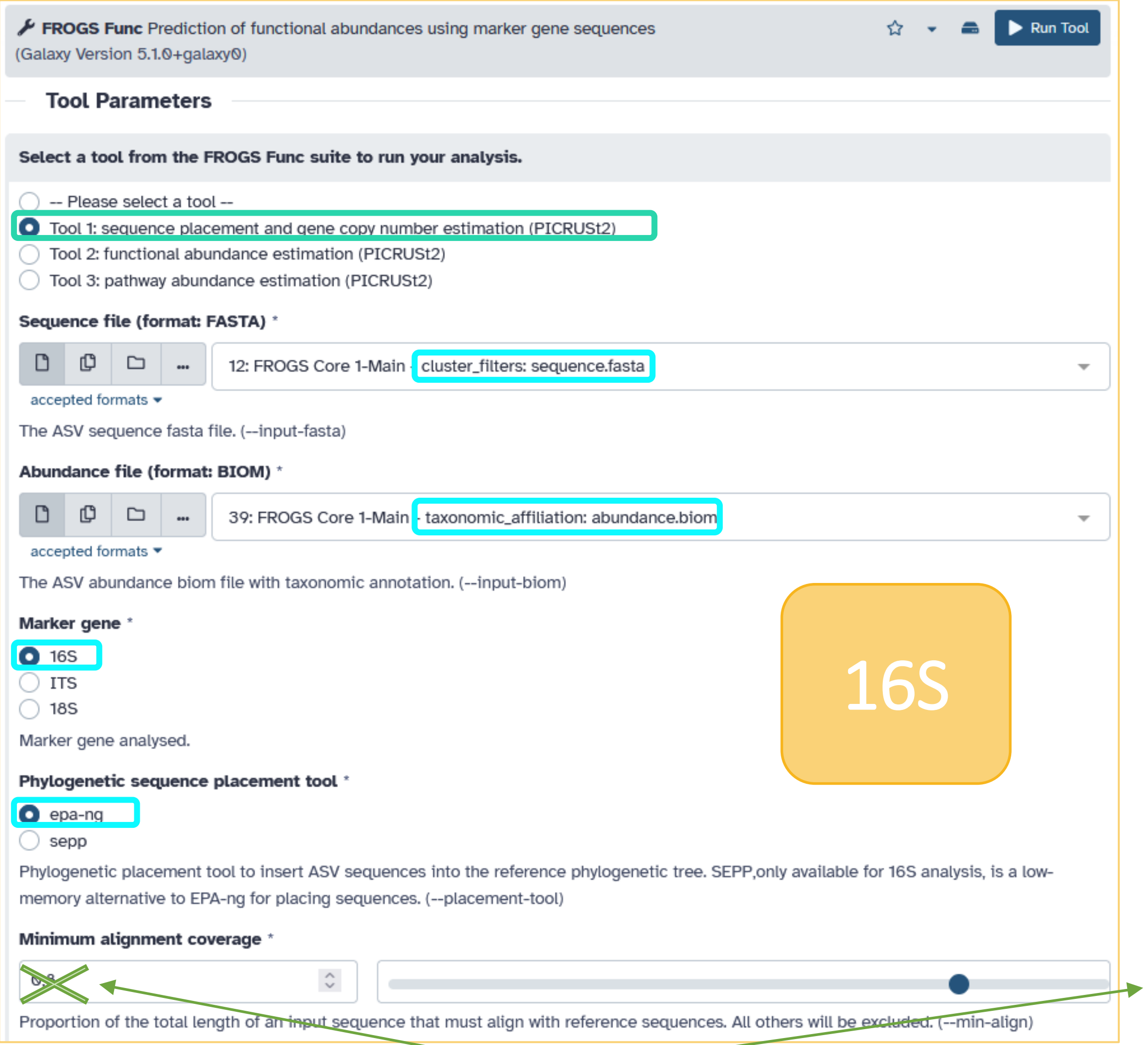
Please open the FROGS Func Tool 1 and familiarize yourself with the required parameters.

Please, enter all the information you have/understand.

- For Phylogenetic sequence placement tool, choose “epa-ng”.

Run the process !

Practice session



FROGS Func Prediction of functional abundances using marker gene sequences (Galaxy Version 5.1.0+galaxy0)

Tool Parameters

Select a tool from the FROGS Func suite to run your analysis.

○ -- Please select a tool --
● Tool 1: sequence placement and gene copy number estimation (PICRUSt2)
○ Tool 2: functional abundance estimation (PICRUSt2)
○ Tool 3: pathway abundance estimation (PICRUSt2)

Sequence file (format: FASTA) *

12: FROGS Core 1-Main - cluster_filters: sequence.fasta

accepted formats ▾

The ASV sequence fasta file. (--input-fasta)

Abundance file (format: BIOM) *

39: FROGS Core 1-Main - taxonomic_affiliation: abundance.biom

accepted formats ▾

The ASV abundance biom file with taxonomic annotation. (--input-biom)

Marker gene *

● 16S
○ ITS
○ 18S

Marker gene analysed.

Phylogenetic sequence placement tool *

● epa-ng
○ sepp

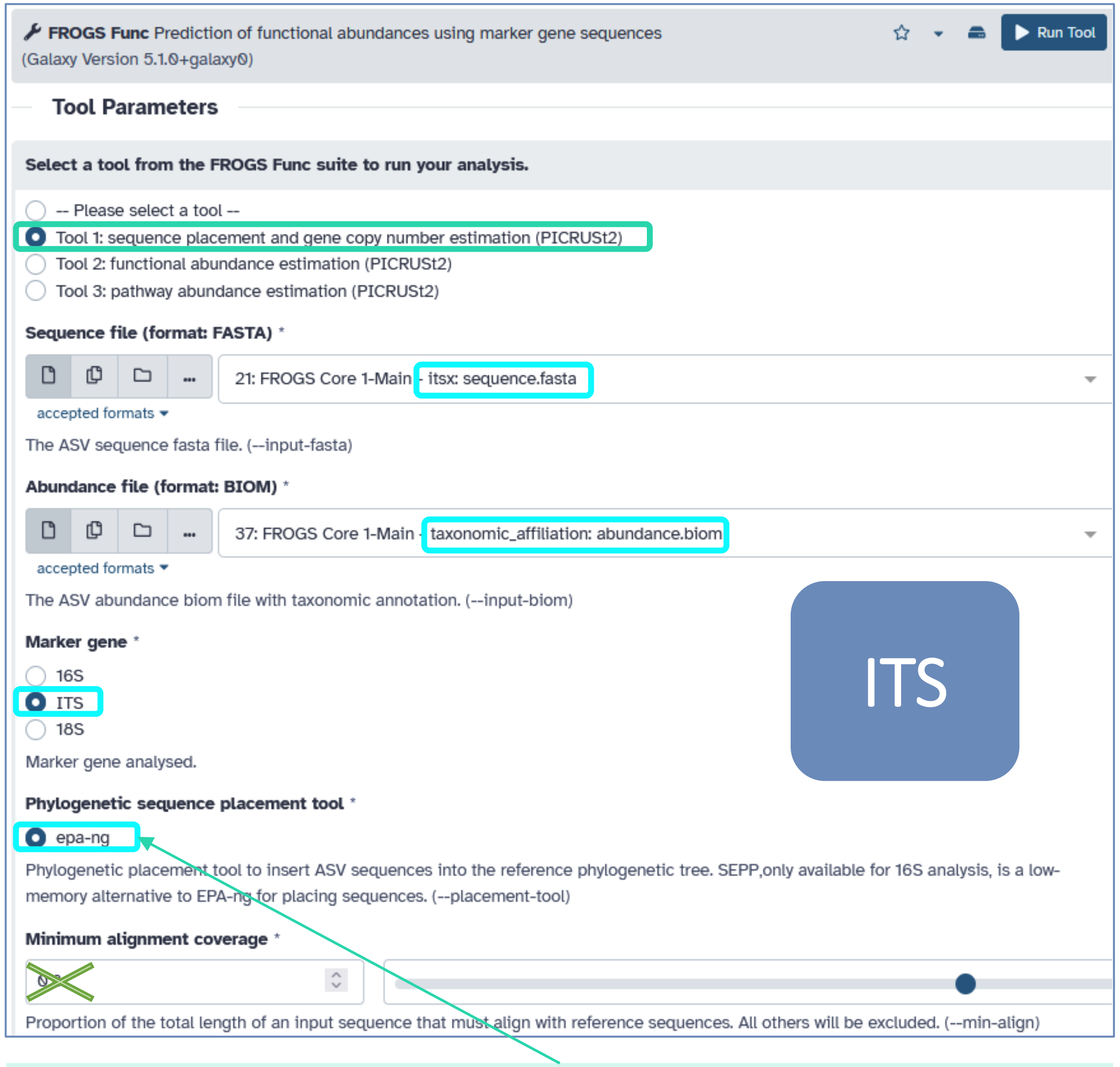
Phylogenetic placement tool to insert ASV sequences into the reference phylogenetic tree. SEPP,only available for 16S analysis, is a low-memory alternative to EPA-ng for placing sequences. (--placement-tool)

Minimum alignment coverage *

~~0.8~~

Proportion of the total length of an input sequence that must align with reference sequences. All others will be excluded. (--min-align)

16S



FROGS Func Prediction of functional abundances using marker gene sequences (Galaxy Version 5.1.0+galaxy0)

Tool Parameters

Select a tool from the FROGS Func suite to run your analysis.

○ -- Please select a tool --
● Tool 1: sequence placement and gene copy number estimation (PICRUSt2)
○ Tool 2: functional abundance estimation (PICRUSt2)
○ Tool 3: pathway abundance estimation (PICRUSt2)

Sequence file (format: FASTA) *

21: FROGS Core 1-Main - itsx: sequence.fasta

accepted formats ▾

The ASV sequence fasta file. (--input-fasta)

Abundance file (format: BIOM) *

37: FROGS Core 1-Main - taxonomic_affiliation: abundance.biom

accepted formats ▾

The ASV abundance biom file with taxonomic annotation. (--input-biom)

Marker gene *

○ 16S
● ITS
○ 18S

Marker gene analysed.

Phylogenetic sequence placement tool *

● epa-ng

Phylogenetic placement tool to insert ASV sequences into the reference phylogenetic tree. SEPP,only available for 16S analysis, is a low-memory alternative to EPA-ng for placing sequences. (--placement-tool)

Minimum alignment coverage *

~~0.8~~

Proportion of the total length of an input sequence that must align with reference sequences. All others will be excluded. (--min-align)

ITS

Don’t worry about this setting; it’s an internal setting in Picrust and will be removed in the next release of FROGS.

epa-ng option is the only one available for ITS and 18S markers. Because epa-ng is greedy, it may force to reduce the dataset to run.

Settings

- **placement tool**: epa-ng or sepp are placement tools for insertion of sequences into the PICRUSt2 reference tree.
  sepp is a low-memory alternative to epa-ng for placing sequences.

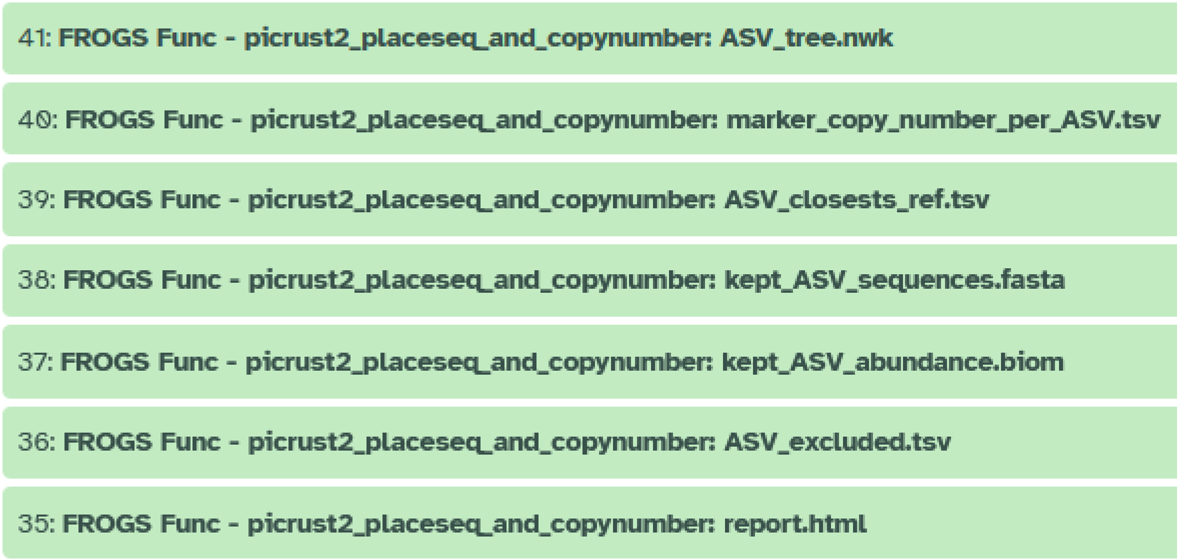
- If the tool crashes when you're using epa-ng , try using sepp instead.

Practice session

What are the outputs ?

Practice session

What are the outputs ?



41: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_tree.nwk**

40: **FROGS Func - picrust2_placeseq_and_copynumber: marker_copy_number_per_ASV.tsv**

39: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_closests_ref.tsv**

38: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_sequences.fasta**

37: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_abundance.biom**

36: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_excluded.tsv**

35: **FROGS Func - picrust2_placeseq_and_copynumber: report.html**

Practice session

How many ASVs are kept in the dataset?

How many sequences does that amount to?

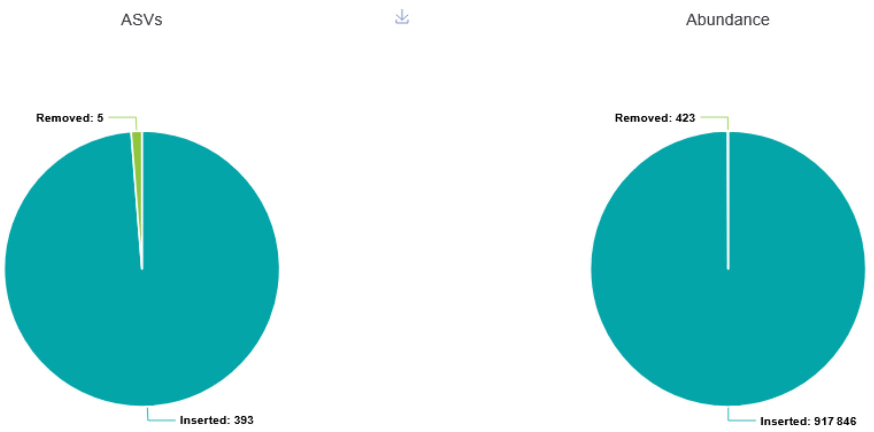
What impact does this have on the subsequent functional prediction of your samples?

Practice session

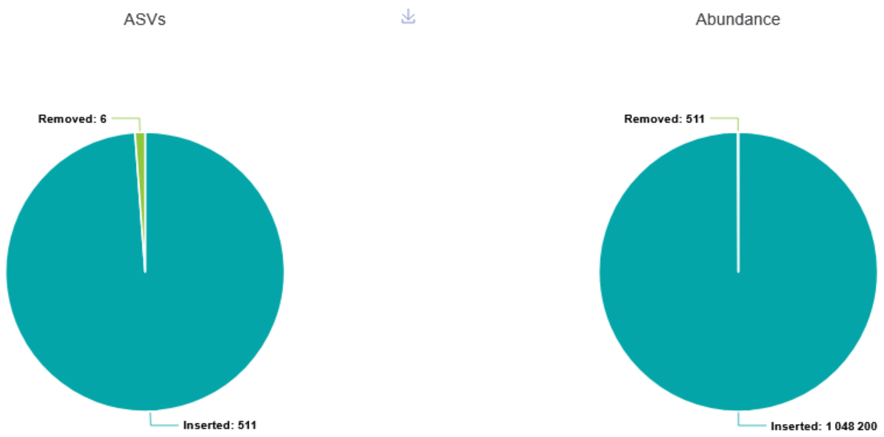
How many ASVs are kept in the dataset?
How many sequences does that amount to?

What impact does this have on the subsequent functional prediction of your samples?

16S
swarm



16S
DADA2

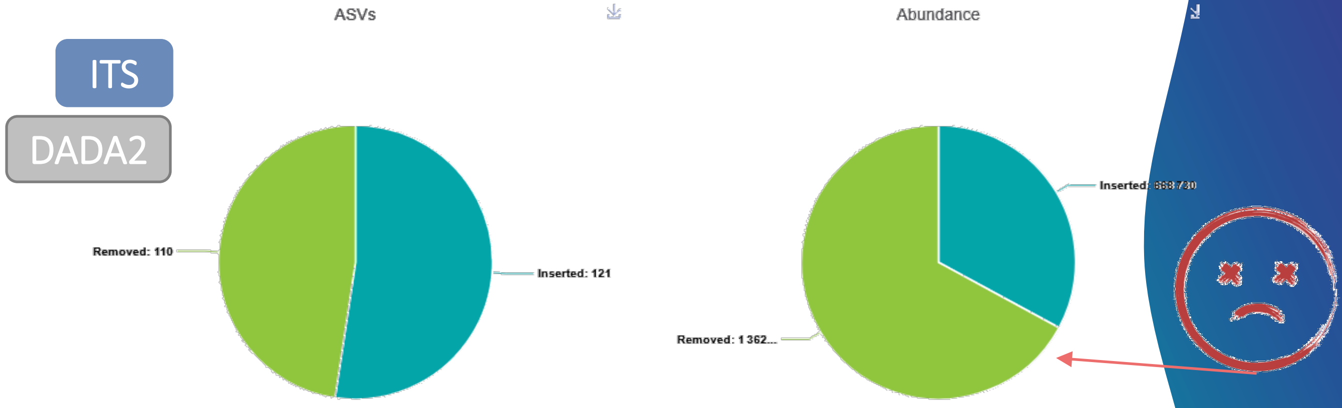


- Only a few ASVs are not placed in the reference tree.
- And these ASVs represent a small number of sequences.
- The datasets from this 1st step is a good representation of the original information.
- We can continue the functional inference analysis.

Practice session

How many ASVs are kept in the dataset?
How many sequences does that amount to?

What impact does this have on the subsequent functional prediction of your samples?



- A lot of ASVs are not placed in the reference tree.
- And these ASVs represent a large number of sequences.
- The datasets from this 1st step is not a good representation of the original information.
- It is better not to continue with the functional inference analysis.

Picrust2 is not well adapted to ITS data.
The FROGS group is currently working on a new functional inference strategy for yeasts and fungi.

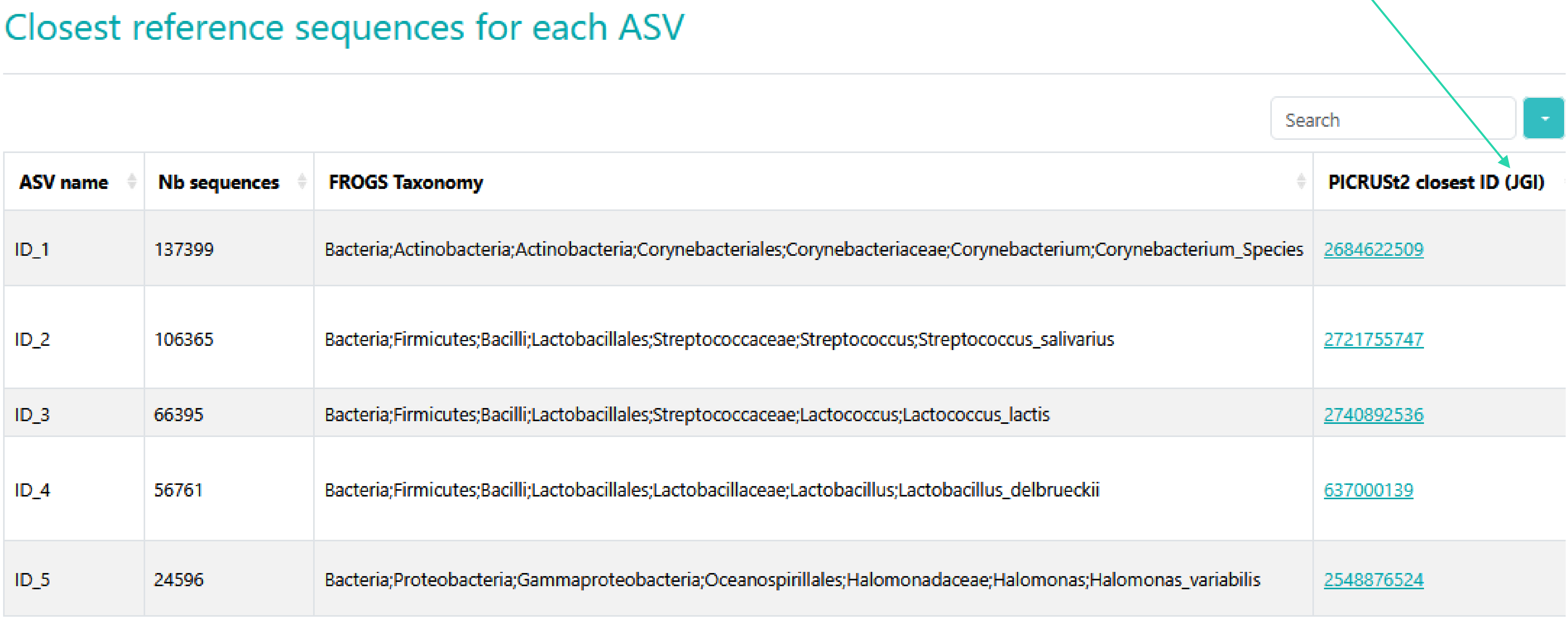
Practice session

Please, analyse the 'Closest reference sequences for each ASV' table in the report.html file.

Practice session

Please, analyse the 'Closest reference sequences for each ASV' table in the report.html file.

https://jgi.doe.gov/



Closest reference sequences for each ASV

Search

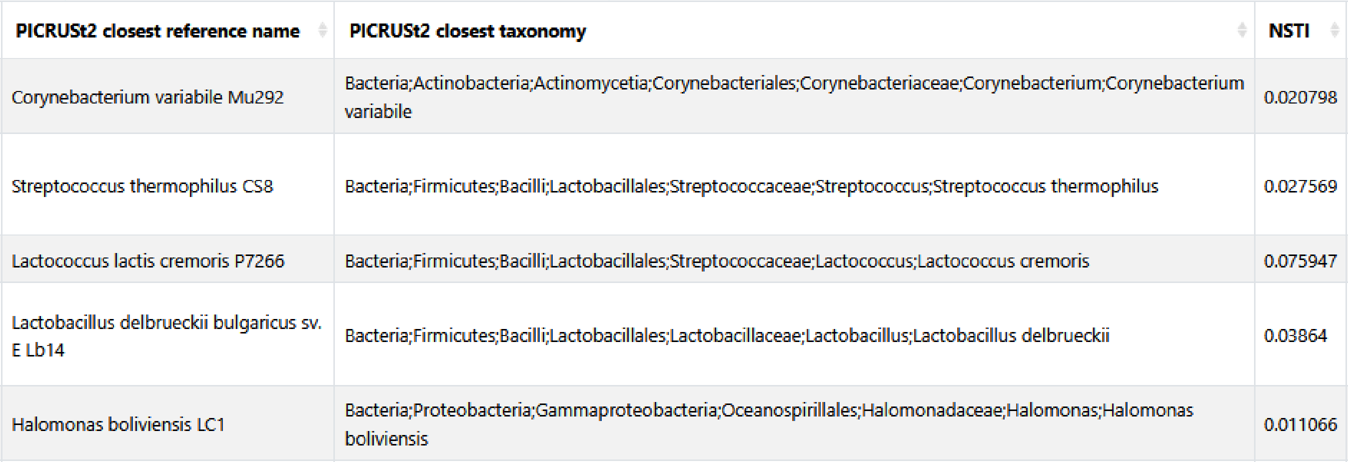
ASV name	Nb sequences	FROGS Taxonomy	PICRUSt2 closest ID (JGI)
ID_1	137399	Bacteria;Actinobacteria;Actinobacteria;Corynebacteriales;Corynebacteriaceae;Corynebacterium;Corynebacterium_Species	2684622509
ID_2	106365	Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus;Streptococcus_salivarius	2721755747
ID_3	66395	Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Lactococcus;Lactococcus_lactis	2740892536
ID_4	56761	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus;Lactobacillus_delbrueckii	637000139
ID_5	24596	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Halomonadaceae;Halomonas;Halomonas_variabilis	2548876524

Practice session

Please, analyse the 'Closest reference sequences for each ASV' table in the report.html file.

Sequence name in the Picrust reference tree

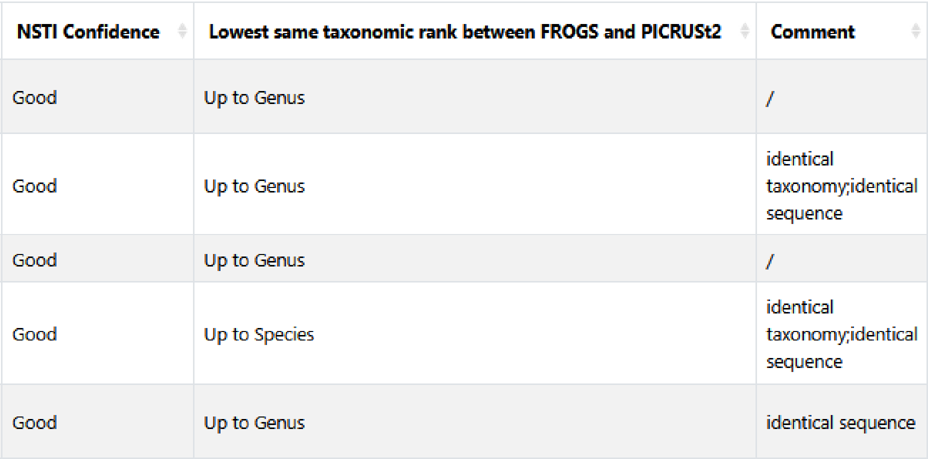
Taxonomy (JGI) of the closest reference sequence from the ASV inserted in the reference tree under the following format: Kingdom;Phylum;Class;Order;Family;Genus;Species



PICRUSt2 closest reference name	PICRUSt2 closest taxonomy	NSTI
Corynebacterium variabile Mu292	Bacteria;Actinobacteria;Actinomycetia;Corynebacteriales;Corynebacteriaceae;Corynebacterium;Corynebacterium variabile	0.020798
Streptococcus thermophilus CS8	Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus;Streptococcus thermophilus	0.027569
Lactococcus lactis cremoris P7266	Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Lactococcus;Lactococcus cremoris	0.075947
Lactobacillus delbrueckii bulgaricus sv. E Lb14	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus;Lactobacillus delbrueckii	0.03864
Halomonas boliviensis LC1	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Halomonadaceae;Halomonas;Halomonas boliviensis	0.011066

Nearest Sequenced Taxon Index (NSTI) is the phylogenetic distance between the ASV and the nearest sequenced reference genome.

Confidence tag of NSTI



NSTI Confidence	Lowest same taxonomic rank between FROGS and PICRUSt2	Comment
Good	Up to Genus	/
Good	Up to Genus	identical taxonomy;identical sequence
Good	Up to Genus	/
Good	Up to Species	identical taxonomy;identical sequence
Good	Up to Genus	identical sequence

NSTI confidence

According to the NSTI score, we guide you in the confidence you can bring to the issue affiliation of PICRUSt2. Four levels are given:

- 0 < Good < 0.5
- 0.5 <= Medium < 1
- 1 <= Bad < 2
- To exclude >= 2

PICRUSt2 sets NSTI threshold to 2 per default. Some studies have shown that this threshold is too permissive.

Thus, it is important to see if the taxonomies between PICRUSt2 and FROGS are quite similar or not, in order to potentially choose a more stringent threshold afterwards.

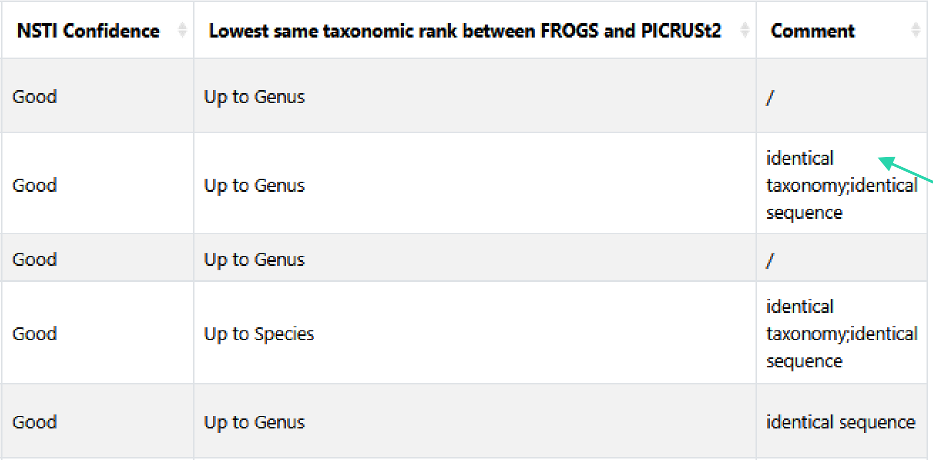
For example, a NSTI lower than 0.5, with “species” as lowest common taxonomic rank between FROGS and PICRUSt2 will product a good prediction.

Practice session

Please, analyse the 'Closest reference sequences for each ASV' table in the report.html file.

PICRUSt2 closest reference name	PICRUSt2 closest taxonomy	NSTI
Corynebacterium variabile Mu292	Bacteria;Actinobacteria;Actinomycetia;Corynebacteriales;Corynebacteriaceae;Corynebacterium;Corynebacterium variabile	0.020798
Streptococcus thermophilus CS8	Bacteria;Firmicutes;Ba	
Lactococcus lactis cremoris P7266	Bacteria;Firmicutes;Bac	
Lactobacillus delbrueckii bulgaricus sv. E Lb14	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus;Lactobacillus delbrueckii	0.03864
Halomonas boliviensis LC1	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Halomonadaceae;Halomonas;Halomonas boliviensis	0.011066

Lowest same taxonomic rank between FROGS and PICRUSt2 :
Lowest common taxonomic rank between FROGS and PICRUSt2 affiliations.



NSTI Confidence	Lowest same taxonomic rank between FROGS and PICRUSt2	Comment
Good	Up to Genus	/
Good	Up to Genus	identical taxonomy;identical sequence
Good	Up to Genus	/
Good	Up to Species	identical taxonomy;identical sequence
Good	Up to Genus	identical sequence

- *identical taxonomy*: if the FROGS and PICRUSt2 taxonomic affiliations are identical.
- *identical sequence*: if the ASV sequence is strictly the same as the reference sequence.

it is a mark of unambiguity

Practice session

Please, the NSTI graphics

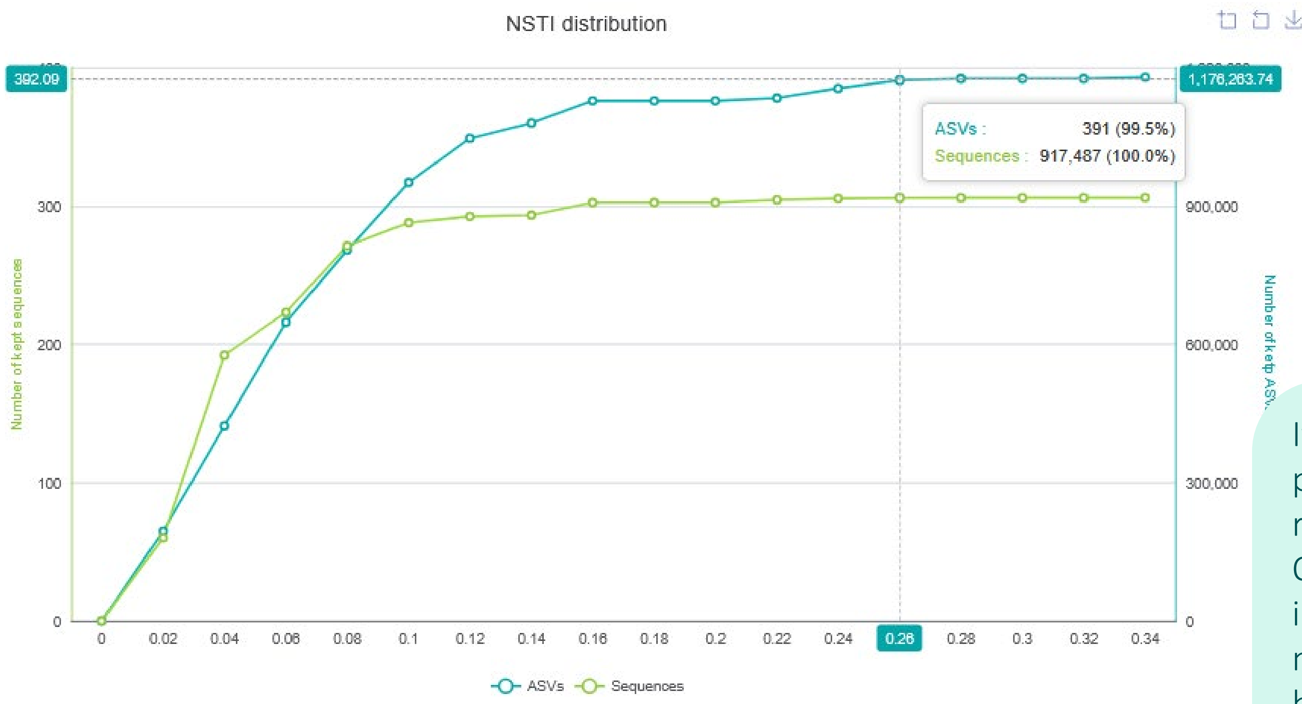
Practice session

Please, the NSTI graphics

16S

swarm

Number of ASVs according to the NSTI threshold



ASVs : 391 (99.5%)
Sequences : 917,487 (100.0%)

It is interesting to find a compromise between the guidance provided by the PICRUSt2 authors and the amount of reusable information that you would like to keep.
On the graph above, keeping 391 ASVs, we keep the information before the plateau, that is, from this point on, the more ASVs we keep the more we degrade the accuracy. So, here NSTI = 0.26
But this depends strongly on the datasets.

This graph allows you to set the “NSTI cut-off” parameter of the next tool

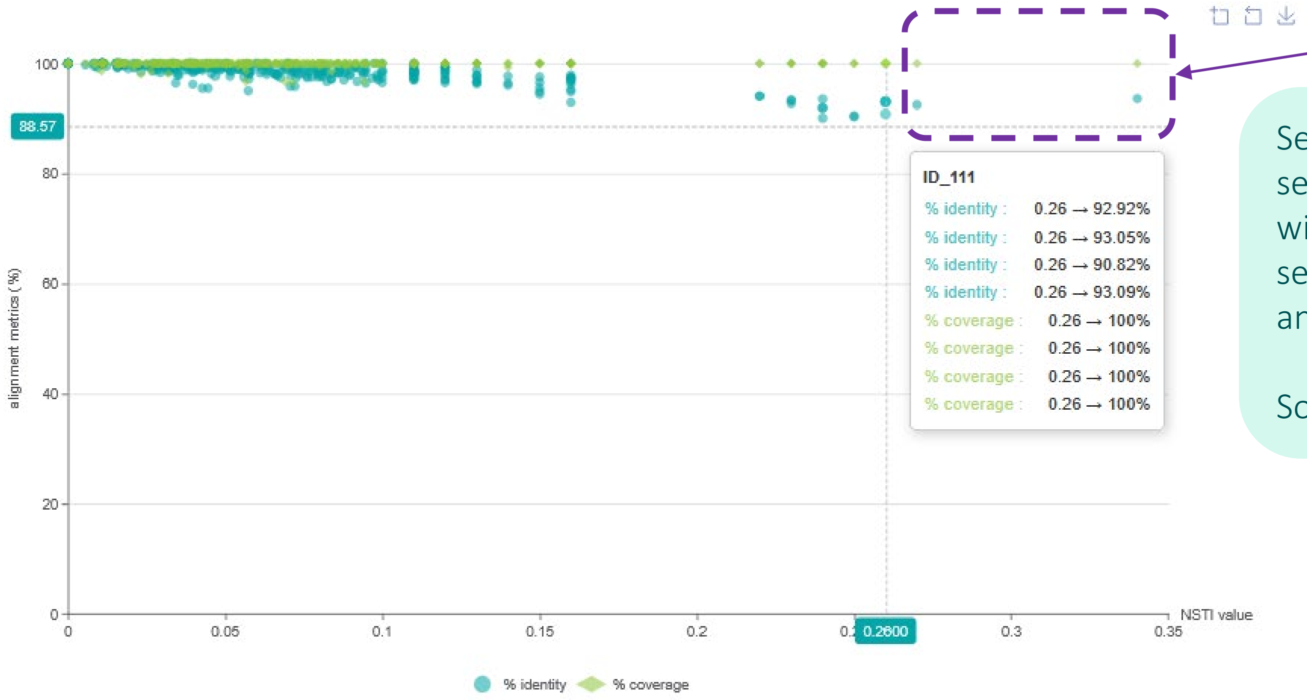
Practice session

Please, the NSTI graphics

16S

swarm

NSTI vs %identity and %coverage between kept ASVs and their closest PICRUSt2 reference sequence



ID_111

% identity :	0.26 → 92.92%
% identity :	0.26 → 93.05%
% identity :	0.26 → 90.82%
% identity :	0.26 → 93.09%
% coverage :	0.26 → 100%
% coverage :	0.26 → 100%
% coverage :	0.26 → 100%
% coverage :	0.26 → 100%

In settings of FROGS Func 2, at the NSTI cut-off, if you choose to fix it at 0.26, these ASV will be removed.

Setting the NSTI to 0.26, we can see that the sequences included in the final dataset – which will be used for functional inference – consist of sequences with a minimum coverage of 100% and a minimum identity of 93.09%.

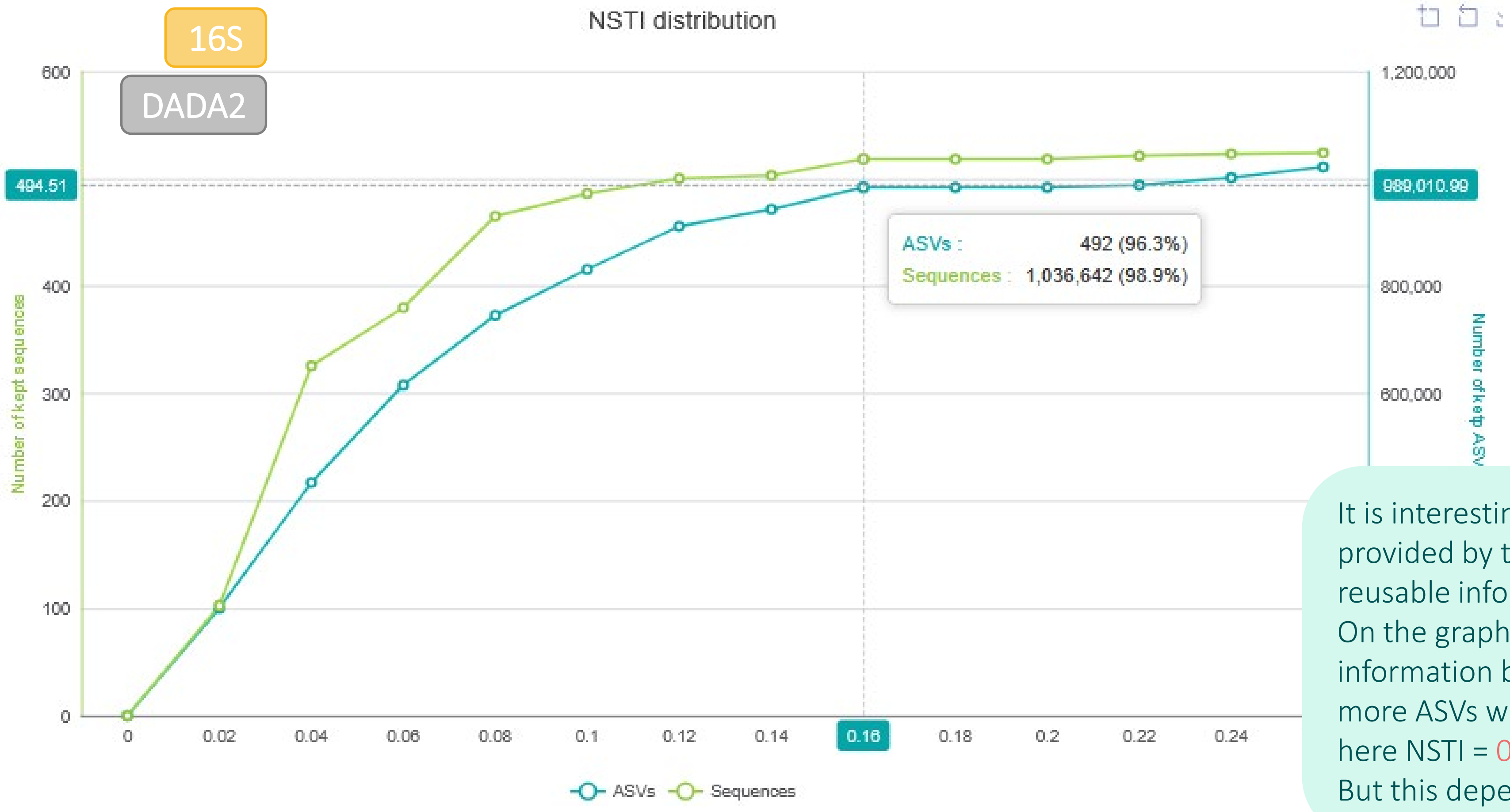
So in this case, we are fairly confident.

Practice session

Please, the NSTI graphics

16S

DADA2



It is interesting to find a compromise between the guidance provided by the PICRUSt2 authors and the amount of reusable information that you would like to keep.
On the graph above, keeping 492 ASVs, we keep the information before the plateau, that is, from this point on, the more ASVs we keep the more we degrade the accuracy. So, here NSTI = 0.16
But this depends strongly on the datasets.

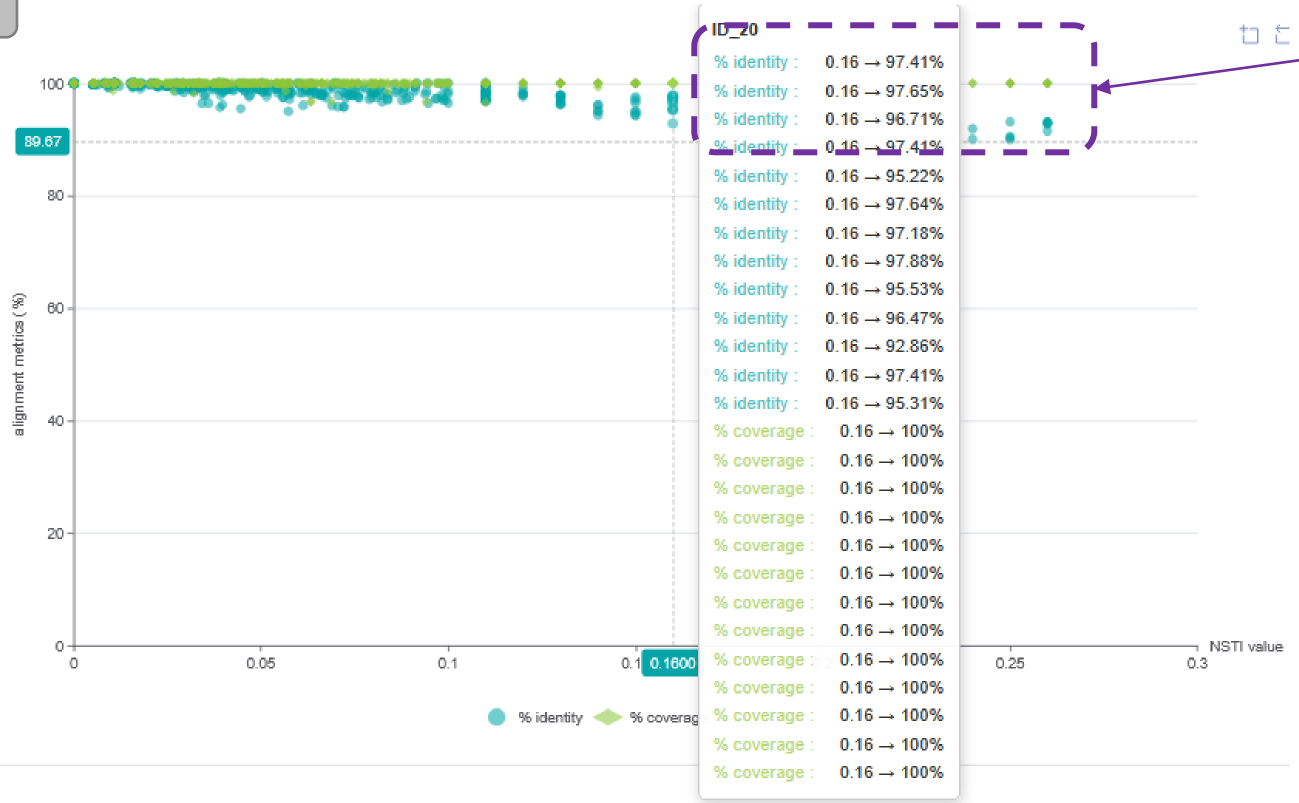
This graph allows you to set the “NSTI cut-off” parameter of the next tool

Practice session

Please, the NSTI graphics

16S

DADA2



ID_20

% identity : 0.16 → 97.41%
% identity : 0.16 → 97.65%
% identity : 0.16 → 96.71%
% identity : 0.16 → 97.41%
% identity : 0.16 → 95.22%
% identity : 0.16 → 97.64%
% identity : 0.16 → 97.18%
% identity : 0.16 → 97.88%
% identity : 0.16 → 95.53%
% identity : 0.16 → 96.47%
% identity : 0.16 → 92.86%
% identity : 0.16 → 97.41%
% identity : 0.16 → 95.31%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%
% coverage : 0.16 → 100%

In settings of FROGS Func 2, at the NSTI cut-off, if you choose to fix it at 0.16, these ASV will be removed.

Setting the NSTI to 0.16, we can see that the sequences included in the final dataset – which will be used for functional inference – consist of sequences with a minimum coverage of 100% and a minimum identity of 95.22%.

So in this case, we are fairly confident.

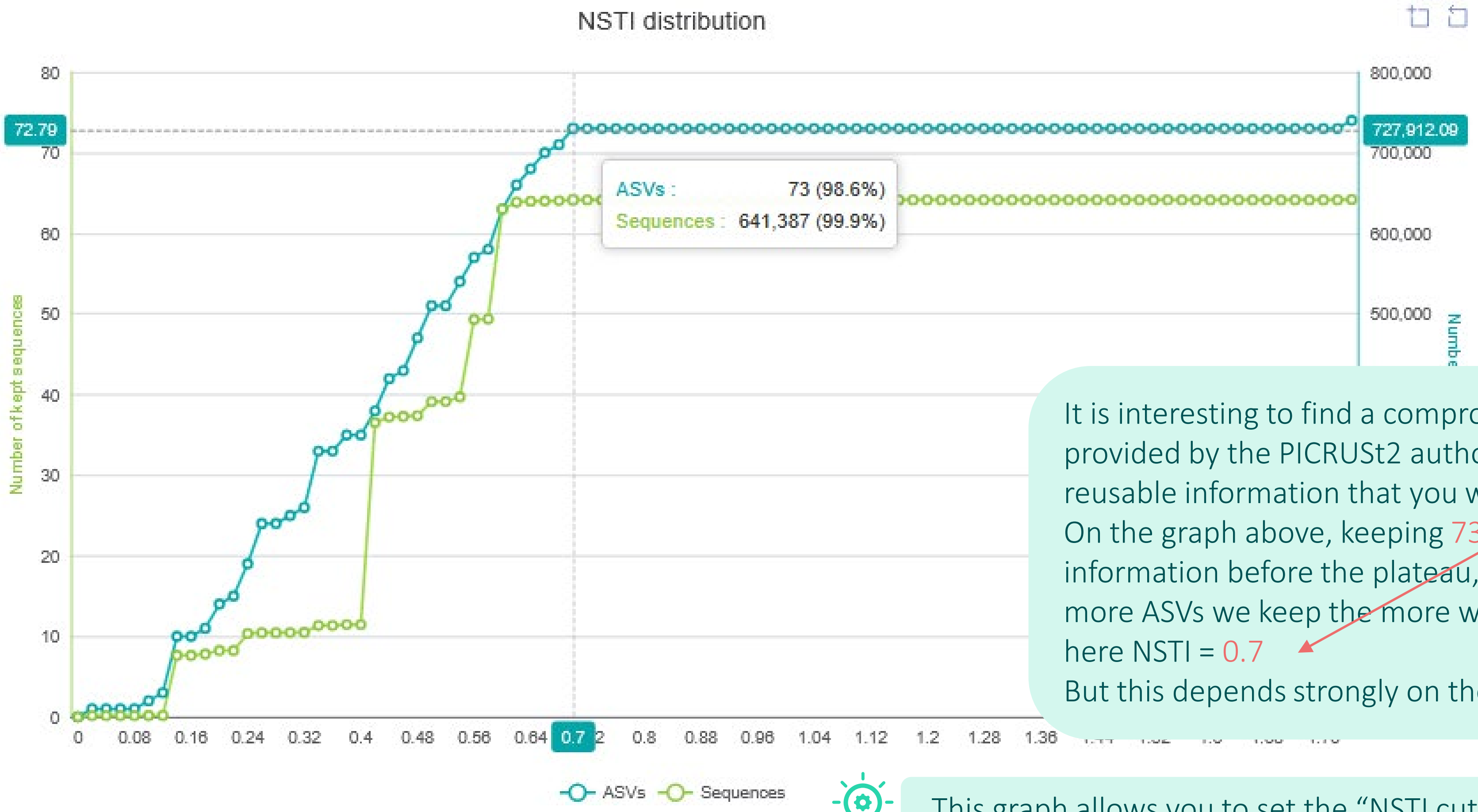
Practice session

Please, the NSTI graphics

Number of ASVs according to the NSTI threshold

ITS

swarm



It is interesting to find a compromise between the guidance provided by the PICRUSt2 authors and the amount of reusable information that you would like to keep.
On the graph above, keeping 73 ASVs, we keep the information before the plateau, that is, from this point on, the more ASVs we keep the more we degrade the accuracy. So, here NSTI = 0.7
But this depends strongly on the datasets.

This graph allows you to set the “NSTI cut-off” parameter of the next tool

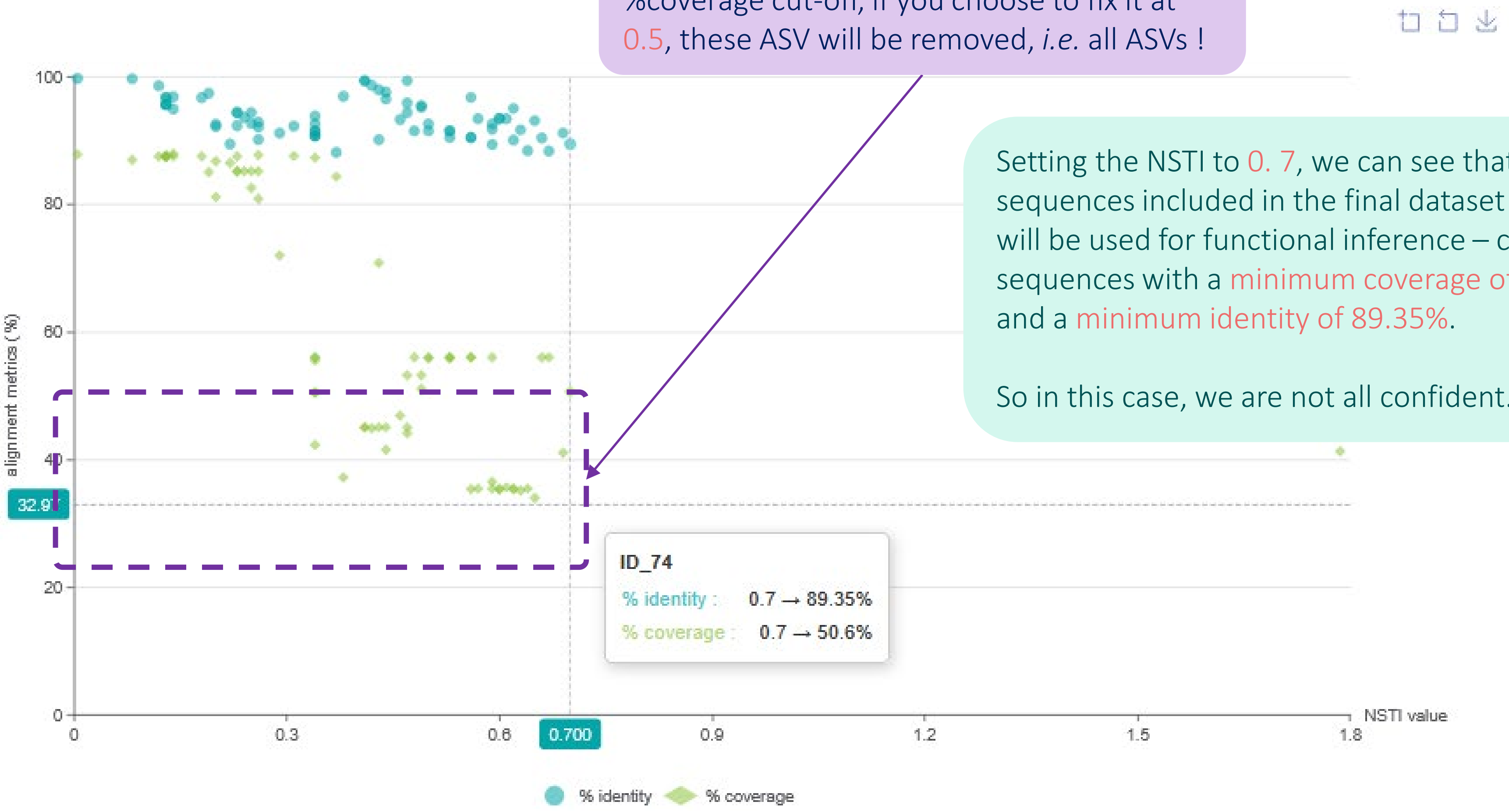
Practice session

Please, the NSTI graphics

ITS

swarm

NSTI vs %identity and %coverage between kept ASVs and their closest PICRUSt2 reference sequence



In settings of FROGS Func 2, at the %coverage cut-off, if you choose to fix it at 0.5, these ASV will be removed, *i.e.* all ASVs !

Setting the NSTI to 0. 7, we can see that the sequences included in the final dataset – which will be used for functional inference – consist of sequences with a minimum coverage of 50.6% and a minimum identity of 89.35%.

So in this case, we are not all confident.

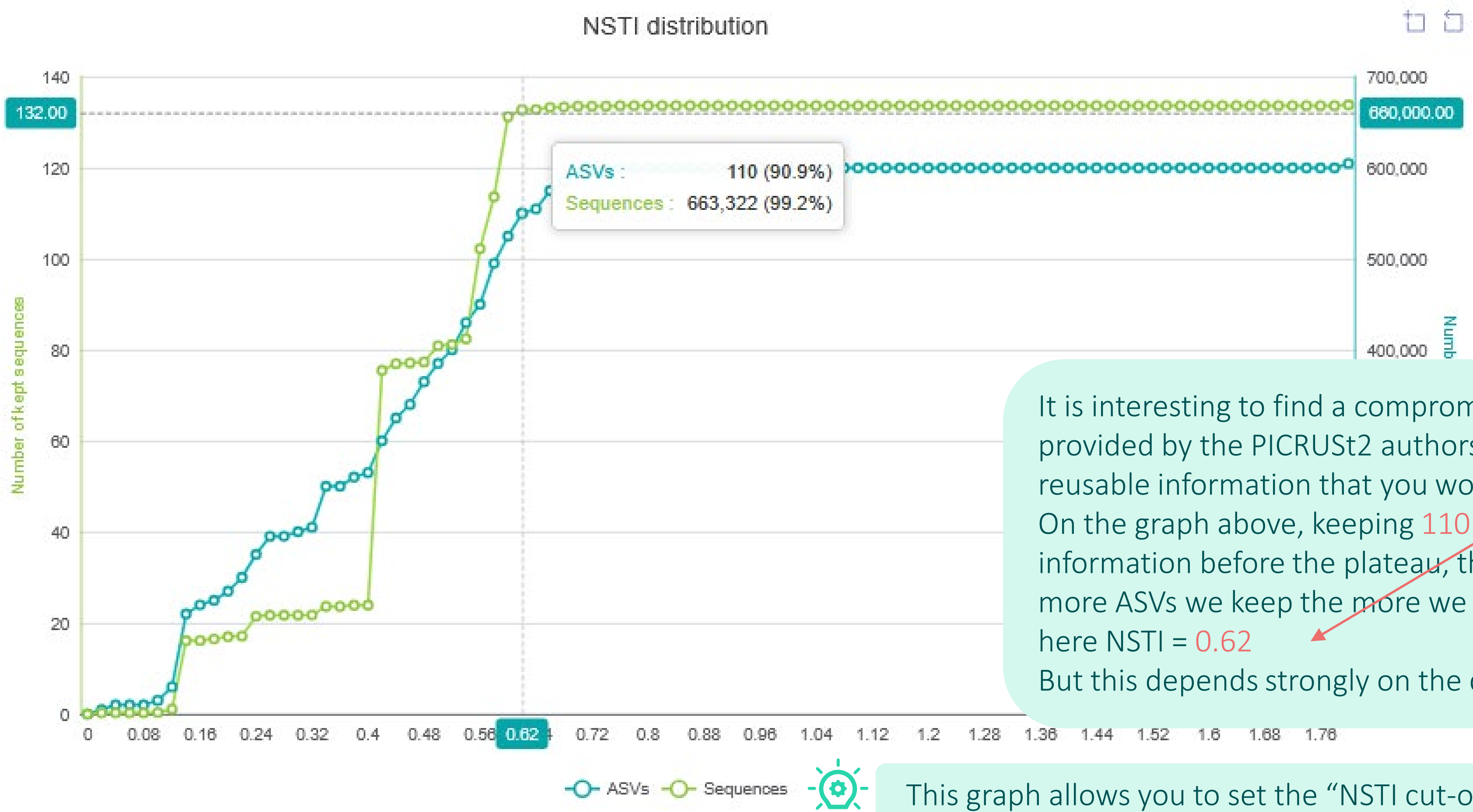
Practice session

Please, the NSTI graphics

Number of ASVs according to the NSTI threshold

ITS

DADA2



It is interesting to find a compromise between the guidance provided by the PICRUSt2 authors and the amount of reusable information that you would like to keep.
On the graph above, keeping 110 ASVs, we keep the information before the plateau, that is, from this point on, the more ASVs we keep the more we degrade the accuracy. So, here NSTI = 0.62
But this depends strongly on the datasets.

This graph allows you to set the “NSTI cut-off” parameter of the next tool

Practice session

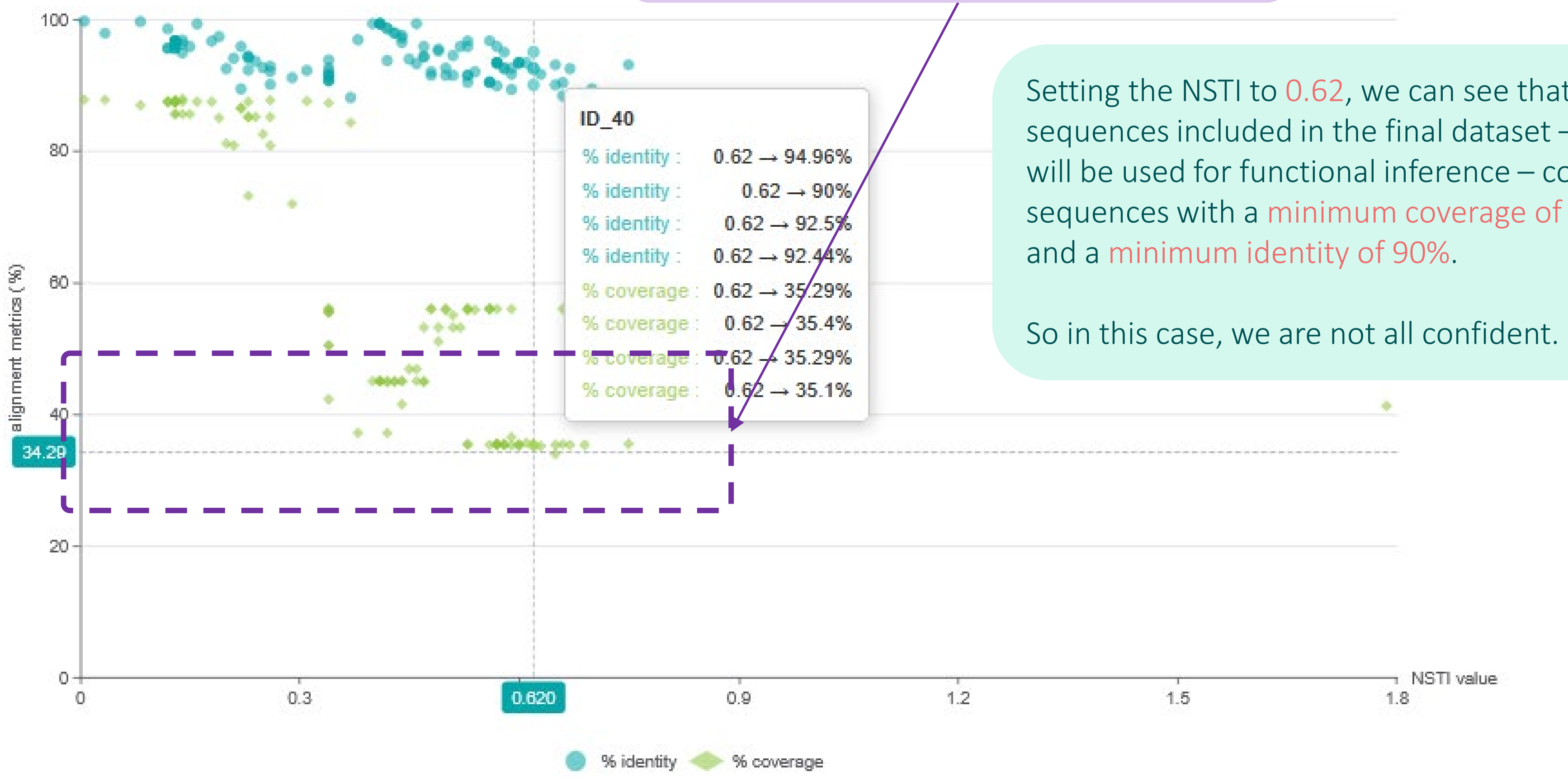
Please, the NSTI graphics

NSTI vs %identity and %coverage between kept ASVs and their closest PICRUSt2 reference sequence

ITS

DADA2

In settings of FROGS Func 2, at the %coverage cut-off, if you choose to fix it at 0.5, these ASV will be removed, *i.e.* all ASVs !



Setting the NSTI to 0.62, we can see that the sequences included in the final dataset – which will be used for functional inference – consist of sequences with a minimum coverage of 35.1% and a minimum identity of 90%.

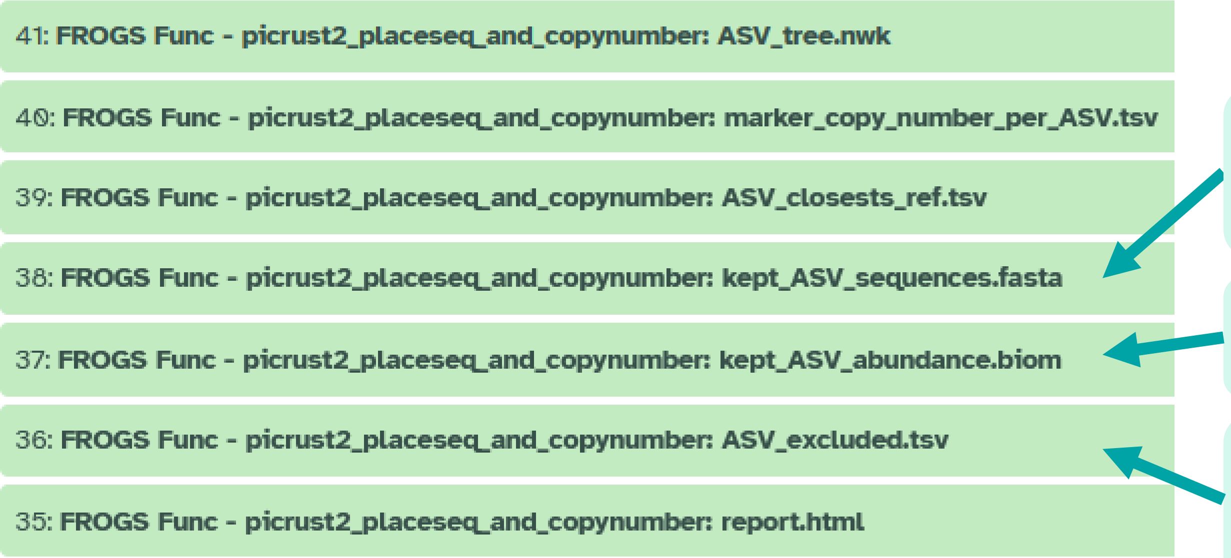
So in this case, we are not all confident.

Practice session

Please, check the other output files.

Practice session

Please, check the other output files.



41: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_tree.nwk**

40: **FROGS Func - picrust2_placeseq_and_copynumber: marker_copy_number_per_ASV.tsv**

39: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_closests_ref.tsv**

38: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_sequences.fasta**

37: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_abundance.biom**

36: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_excluded.tsv**

35: **FROGS Func - picrust2_placeseq_and_copynumber: report.html**

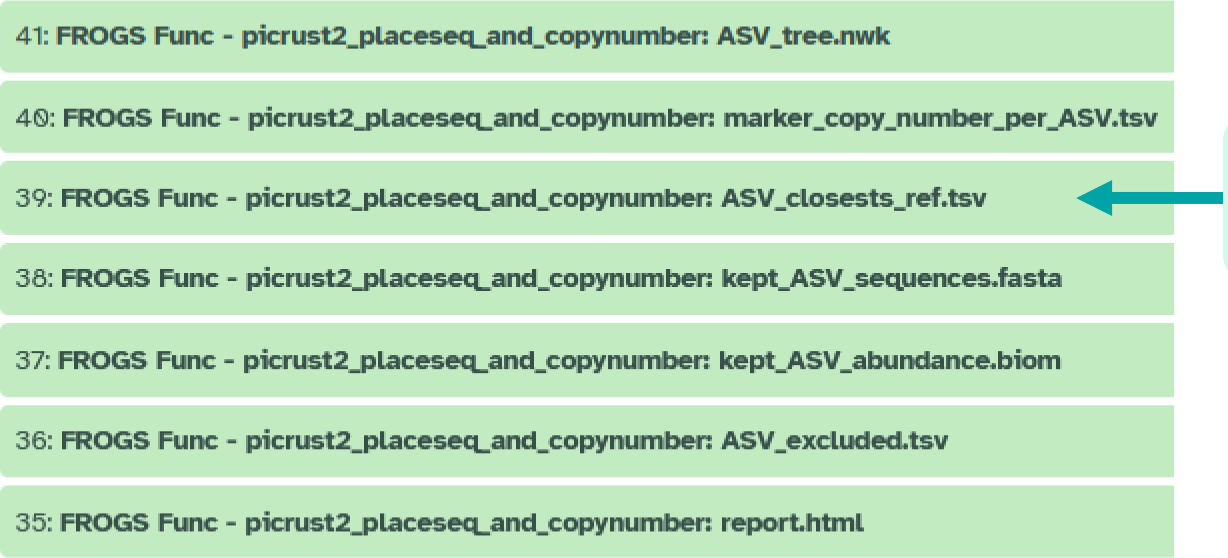
This file contains all the ASV fasta sequences (without those that may be excluded)

Abundance table without those that may be excluded

This file contains all the ASVs that could not be placed in the PICRUSt2 reference tree.

Practice session

Please, check the other output files.



41: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_tree.nwk**

40: **FROGS Func - picrust2_placeseq_and_copynumber: marker_copy_number_per_ASV.tsv**

39: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_closests_ref.tsv**

38: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_sequences.fasta**

37: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_abundance.biom**

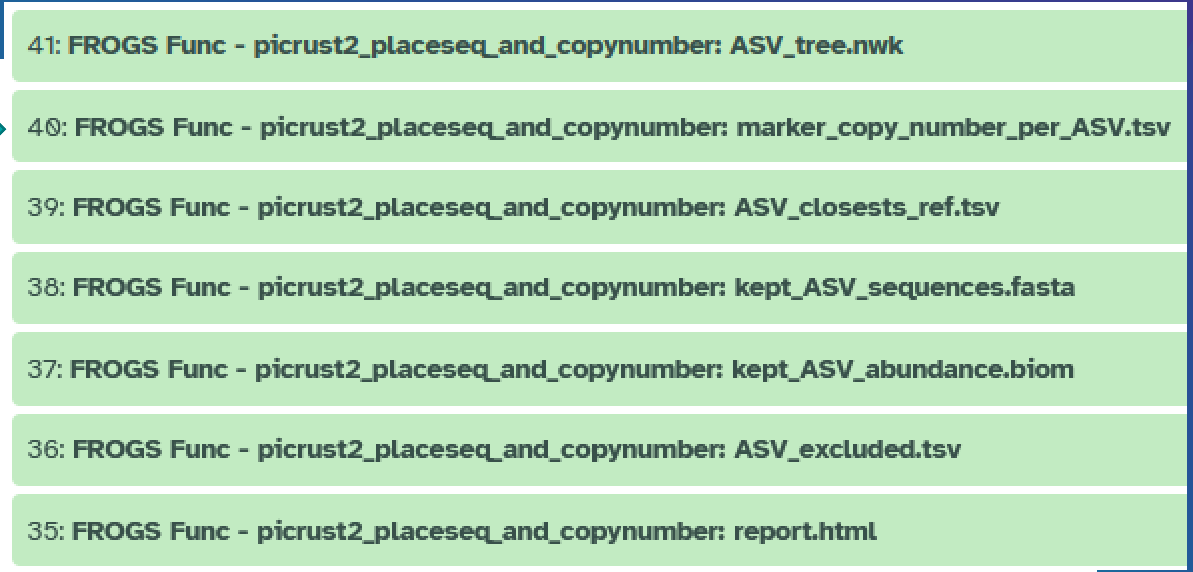
36: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_excluded.tsv**

35: **FROGS Func - picrust2_placeseq_and_copynumber: report.html**

The table contained in the report.html file is available in TSV format.

Practice session

Please, check the other output files.



41: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_tree.nwk**

40: **FROGS Func - picrust2_placeseq_and_copynumber: marker_copy_number_per_ASV.tsv**

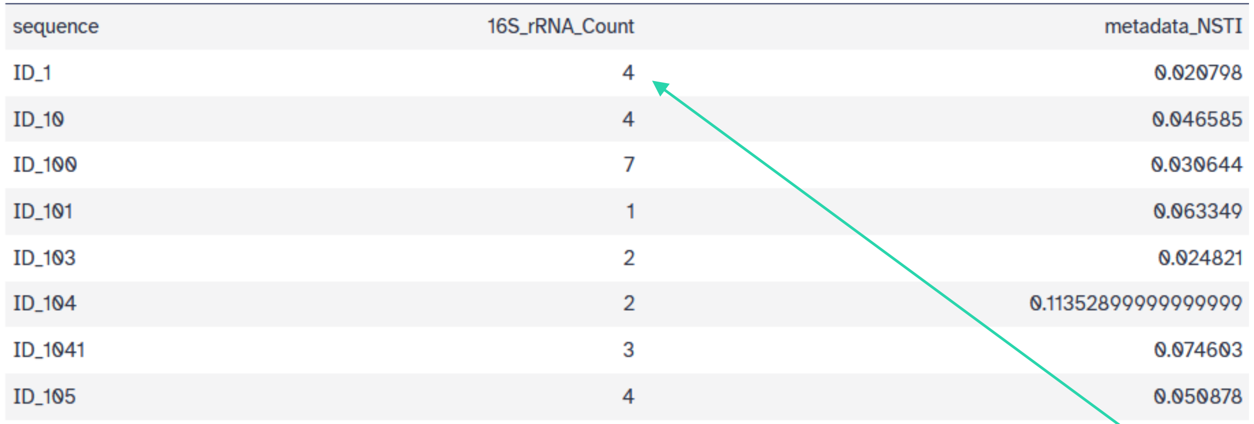
39: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_closests_ref.tsv**

38: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_sequences.fasta**

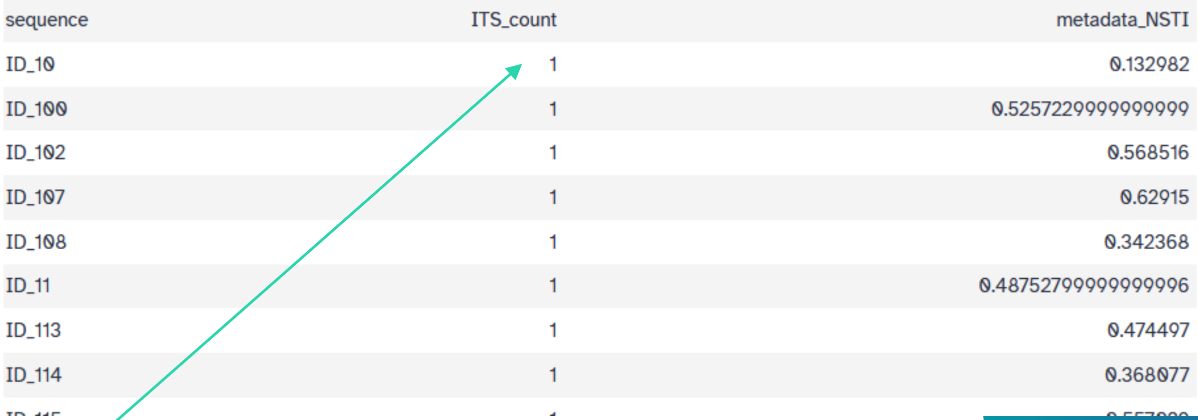
37: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_abundance.biom**

36: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_excluded.tsv**

35: **FROGS Func - picrust2_placeseq_and_copynumber: report.html**



sequence	16S_rRNA_Count	metadata_NSTI
ID_1	4	0.020798
ID_10	4	0.046585
ID_100	7	0.030644
ID_101	1	0.063349
ID_103	2	0.024821
ID_104	2	0.11352899999999999
ID_1041	3	0.074603
ID_105	4	0.050878

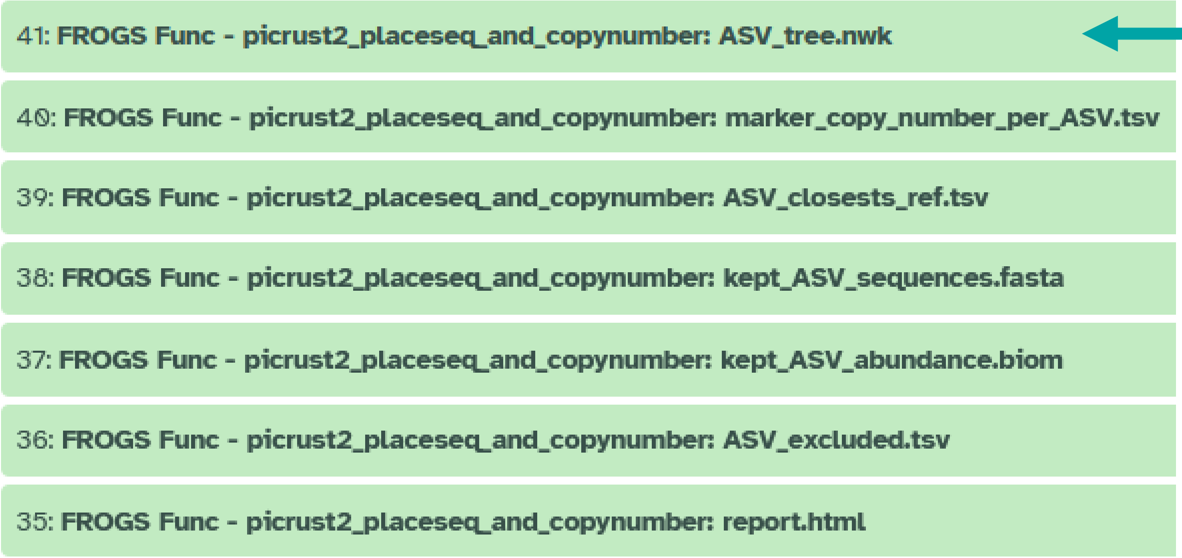


sequence	ITS_count	metadata_NSTI
ID_10	1	0.132982
ID_100	1	0.5257229999999999
ID_102	1	0.568516
ID_107	1	0.62915
ID_108	1	0.342368
ID_11	1	0.48752799999999996
ID_113	1	0.474497
ID_114	1	0.368077

Prediction by PICRUSt2 of 16S/ITS copy number for each ASV (placed in the reference tree).

Practice session

Please, check the other output files.



41: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_tree.nwk**

40: **FROGS Func - picrust2_placeseq_and_copynumber: marker_copy_number_per_ASV.tsv**

39: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_closests_ref.tsv**

38: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_sequences.fasta**

37: **FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_abundance.biom**

36: **FROGS Func - picrust2_placeseq_and_copynumber: ASV_excluded.tsv**

35: **FROGS Func - picrust2_placeseq_and_copynumber: report.html**

This file contains the phylogenetic tree with only kept ASVs.

FROGS Func

FROGS Func - functional abundance estimation

How does it work?

1. Places the ASVs into a reference phylogenetic tree and predicts the number of marker copies in each ASV.

FROGS Func - sequence placement & gene copy nb estimation

2. Estimates the number of copies of each function (gene) in each organism (ASV). Also calculates how many functions (genes) are present in each sample. And determines how many of ASVs there are using the number of copies of the marker.

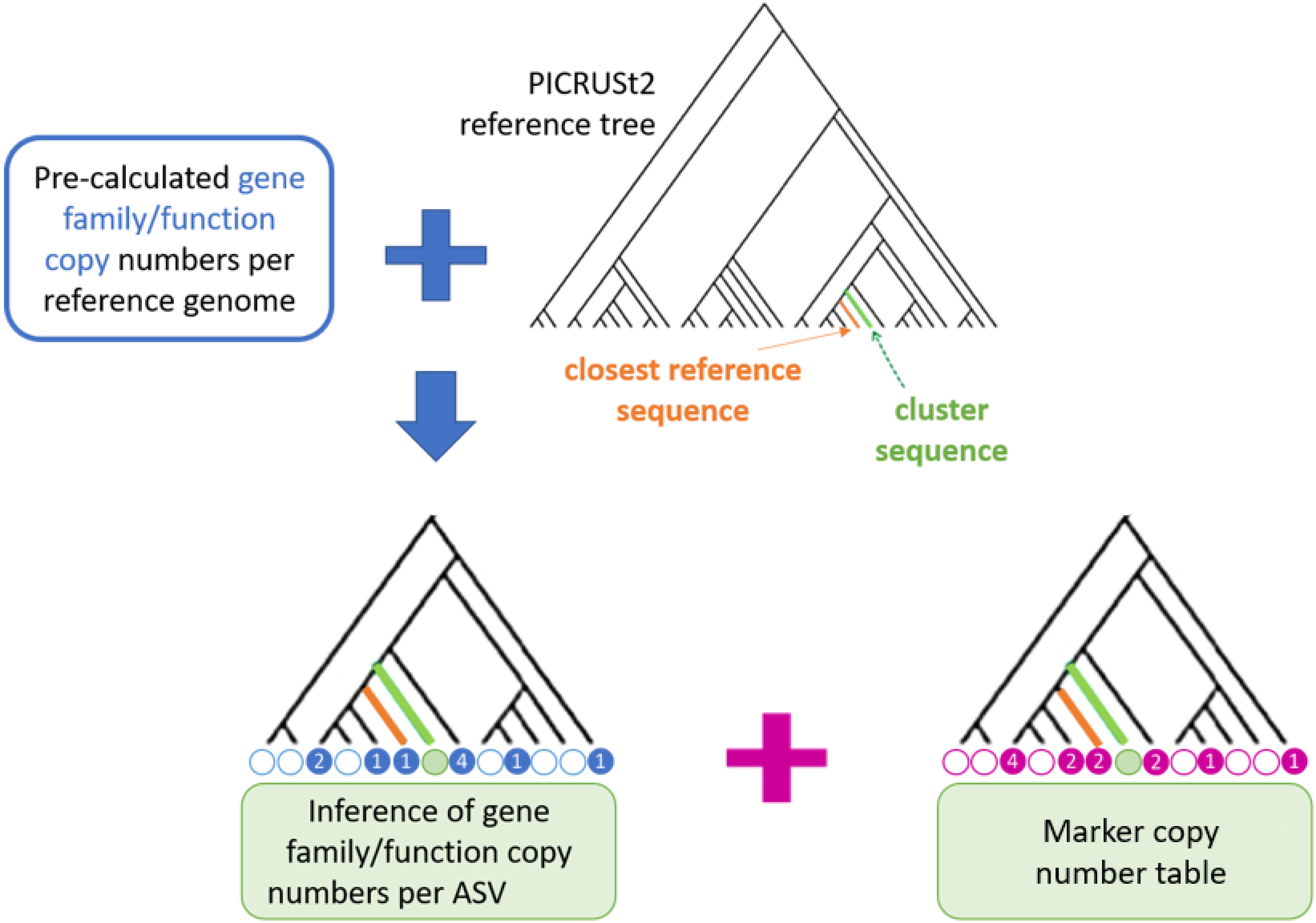
FROGS Func - functional abundance estimation

3. Calculates how much of each pathway is present in each sample.

FROGS Func - pathways abundance estimation

It runs only on 16S, ITS or 18S

Overview



Pre-calculated gene family/function copy numbers per reference genome

PICRUSt2 reference tree

closest reference sequence

cluster sequence

Inference of gene family/function copy numbers per ASV

Marker copy number table

Overview

FROGS Func ‘1’ ASV abundance table

	S1	S2	S3
ASV1	3500	6300	210
ASV2	0	460	36
ASV3	400	700	500

÷

FROGS Func ‘1’ marker copy nb table

	16S copy number
ASV1	7
ASV2	4
ASV3	1

=

FROGS Func ‘2’ ASV abundance normalized table

	S1	S2	S3
ASV1	3500/7	6300/7	210/7
ASV2	0/4	460/4	36/4
ASV3	400/1	700/1	500/1

FROGS Func ‘2’ ASV abundance normalized table

	S1	S2	S3
ASV1	500	900	30
ASV2	0	115	9
ASV3	400	700	500

x

FROGS Func ‘2’ EC predicted copy nb table

	EC:1.1.1.1	EC:1.1.1.2	EC:1.1.1.3
ASV1	2	0	2
ASV2	1	0	0
ASV3	2	4	2

=

FROGS Func ‘2’ EC unstrat abundance table

	S1	S2	S3
EC:1.1.1.1	1800	5460	1129
EC:1.1.1.2	1600	2800	2000
EC:1.1.1.3	1800	3200	1060

Practice session

Please open the FROGS Func Tool 2 and familiarize yourself with the required parameters.

Please, enter all the information you have/understand.

- For Target function database, choose “EC”.
- For NSTI cut-off, choose the previous NSTI chosen.
- For 16S: identity and coverage cut-offs choose 90% (0.9).
- For ITS: identity cut-off choose 85% (0.85) and coverage cut-off choose 50% (0.5).
- For HSP method, choose « mp ».

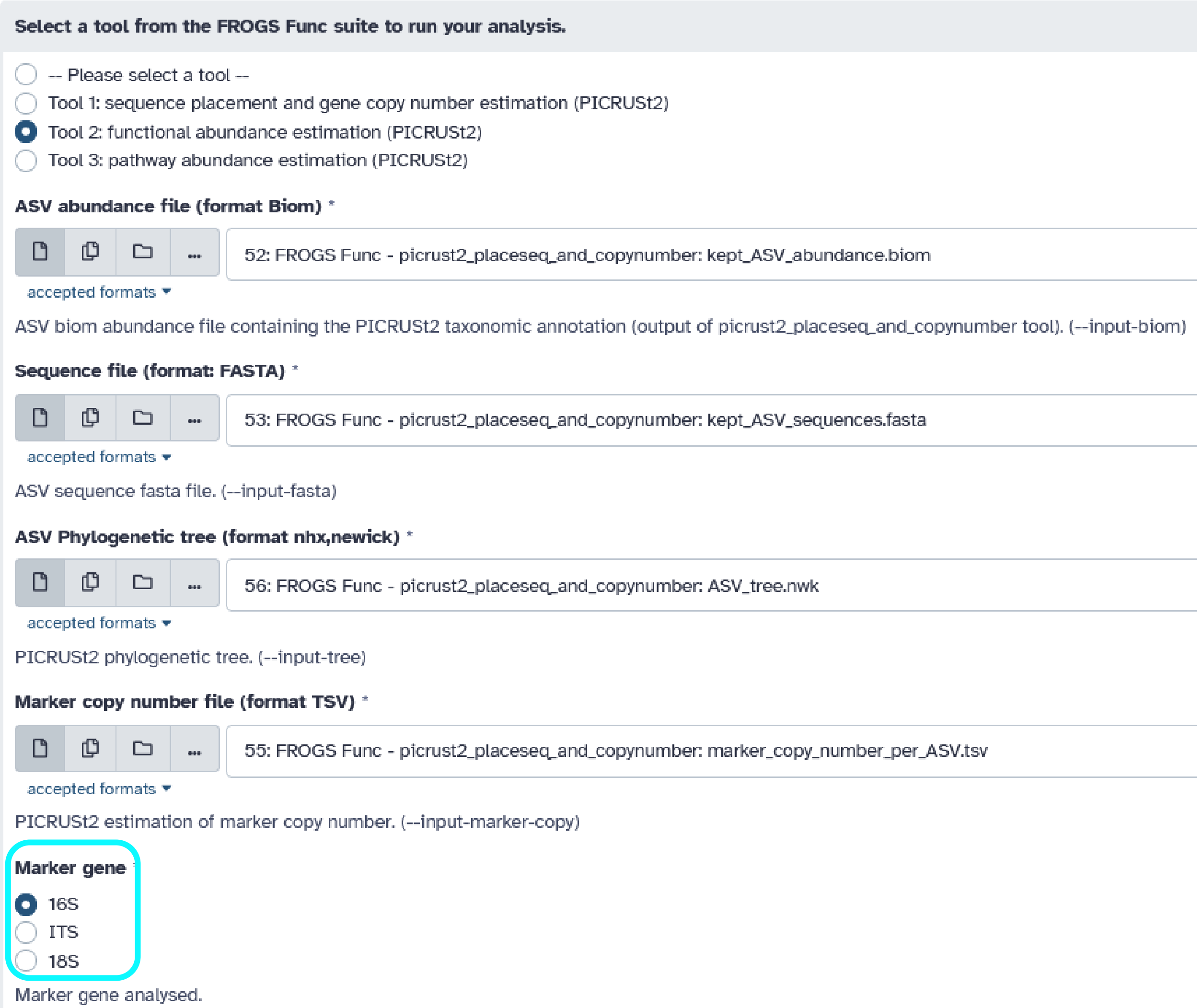
Perhaps, it would be better not to continue using the ITS dataset.

Run the process !

Practice session

16S

swarm



Select a tool from the FROGS Func suite to run your analysis.

○ -- Please select a tool --
○ Tool 1: sequence placement and gene copy number estimation (PICRUSt2)
● Tool 2: functional abundance estimation (PICRUSt2)
○ Tool 3: pathway abundance estimation (PICRUSt2)

ASV abundance file (format Biom) *

52: FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_abundance.biom

accepted formats

ASV biom abundance file containing the PICRUSt2 taxonomic annotation (output of picrust2_placeseq_and_copynumber tool). (--input-biom)

Sequence file (format: FASTA) *

53: FROGS Func - picrust2_placeseq_and_copynumber: kept_ASV_sequences.fasta

accepted formats

ASV sequence fasta file. (--input-fasta)

ASV Phylogenetic tree (format nhx,newick) *

56: FROGS Func - picrust2_placeseq_and_copynumber: ASV_tree.nwk

accepted formats

PICRUSt2 phylogenetic tree. (--input-tree)

Marker copy number file (format TSV) *

55: FROGS Func - picrust2_placeseq_and_copynumber: marker_copy_number_per_ASV.tsv

accepted formats

PICRUSt2 estimation of marker copy number. (--input-marker-copy)

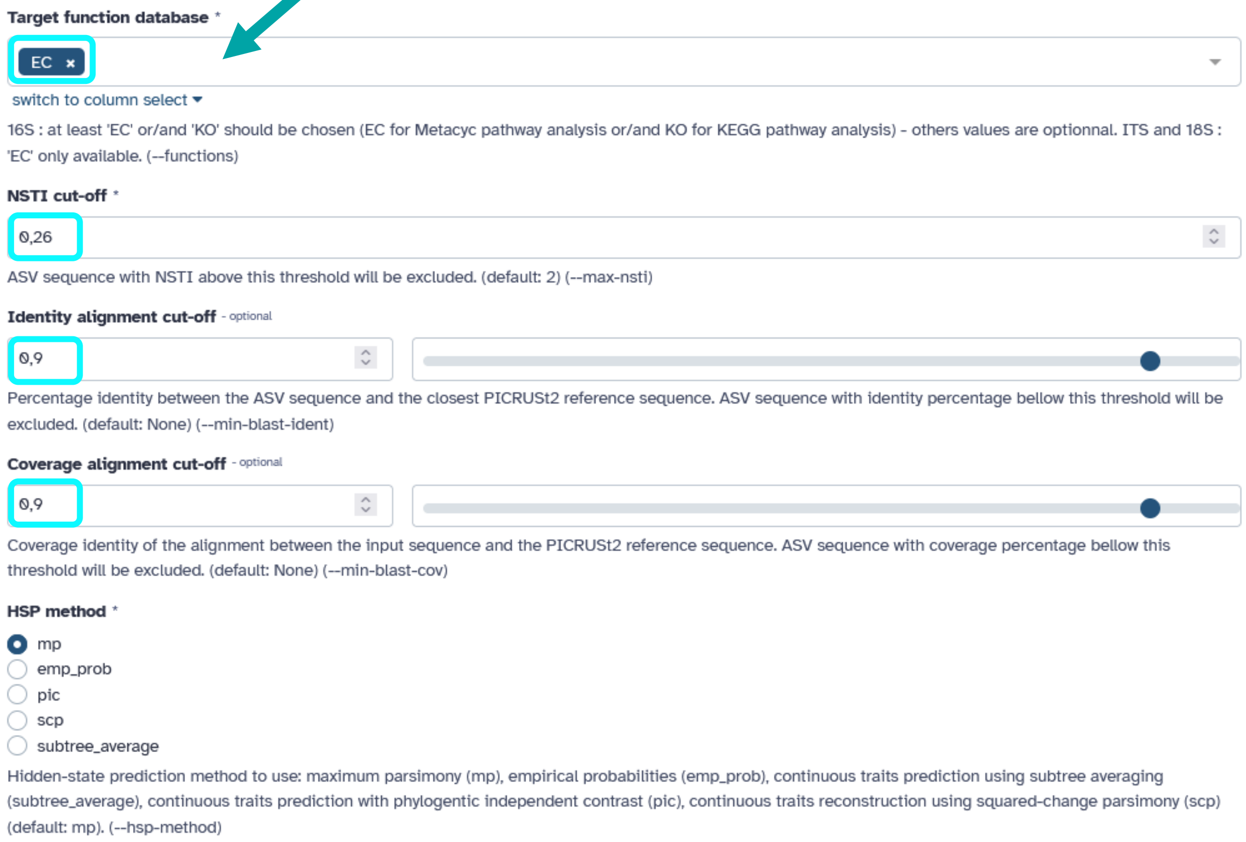
Marker gene *

● 16S
○ ITS
○ 18S

Marker gene analysed.

Practice session

Prediction of the **functions abundances**, using different databases:
EC : https://enzyme.expasy.org/
KO : https://www.genome.jp/kegg/ko.html
PFAM : http://pfam.xfam.org/
COG : https://www.ncbi.nlm.nih.gov/research/cog-project/
TIGRFAM : https://tigrfams.jcvi.org/cgi-bin/index.cgi
PHENO : https://phenodb.org/



Target function database *

EC

switch to column select

16S : at least 'EC' or/and 'KO' should be chosen (EC for Metacyc pathway analysis or/and KO for KEGG pathway analysis) - others values are optionnal. ITS and 18S : 'EC' only available. (--functions)

NSTI cut-off *

0,26

ASV sequence with NSTI above this threshold will be excluded. (default: 2) (--max-nsti)

Identity alignment cut-off - optional

0,9

Percentage identity between the ASV sequence and the closest PICRUSt2 reference sequence. ASV sequence with identity percentage bellow this threshold will be excluded. (default: None) (--min-blast-ident)

Coverage alignment cut-off - optional

0,9

Coverage identity of the alignment between the input sequence and the PICRUSt2 reference sequence. ASV sequence with coverage percentage bellow this threshold will be excluded. (default: None) (--min-blast-cov)

HSP method *

● mp
○ emp_prob
○ pic
○ scp
○ subtree_average

Hidden-state prediction method to use: maximum parsimony (mp), empirical probabilities (emp_prob), continuous traits prediction using subtree averaging (subtree_average), continuous traits prediction with phylogentic independent contrast (pic), continuous traits reconstruction using squared-change parsimony (scp) (default: mp). (--hsp-method)

16S
swarm
NSTI cut-off = 0.26

16S
DADA2
NSTI cut-off = 0.16

ITS
swarm
NSTI cut-off = 0.7
%id = 0.85
%cov = 0.5

ITS
DADA2
NSTI cut-off = 0.62
%id = 0.85
%cov = 0.5

Perhaps it would be better not to continue using the ITS dataset.

Function table choice: Which default pre-calculated count table to use ?

- For 16S rRNA gene you can choose between: ‘EC’, ‘KO’, ‘PFAM’, ‘COG’, ‘TIGRFAM’, and/or ‘PHENO’.
  You must select at least ‘EC’ or ‘KO’ because the information from Metacyc (EC) or KEGG (KO) are required.
- For **ITS** and **18S** markers, **‘EC’ is only** available.

Practice session

Explore the report.html output.

How many ASVs are kept in the dataset?

Has the diversity changed before and after the application of the NSTI/%coverage/%identity cut-off ?

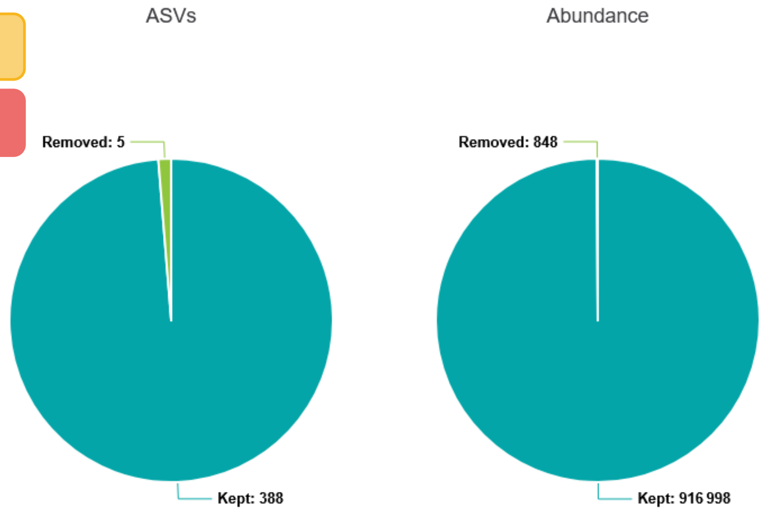
Display the distribution of predicted functions in AOP1_PPC_S* abundances.

Practice session

How many ASVs are kept in the dataset?

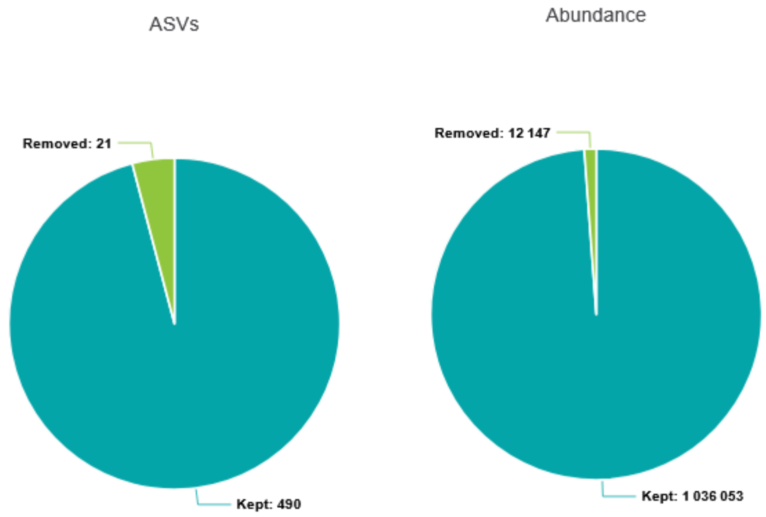
16S

swarm



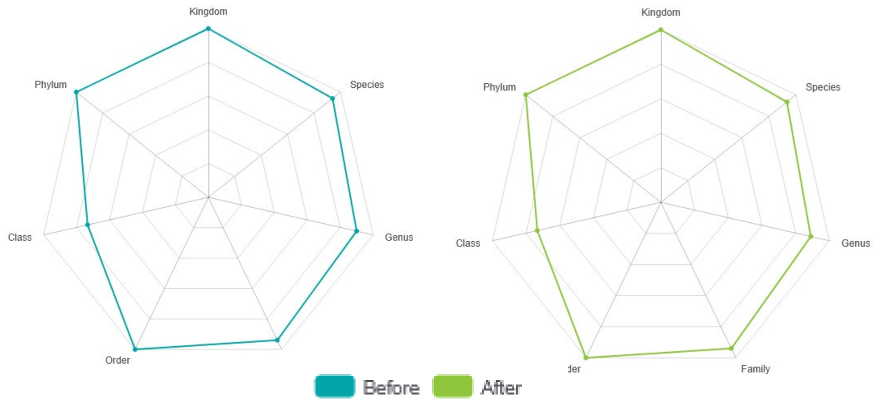
16S

DADA2

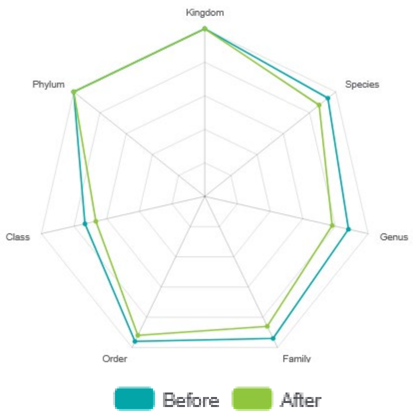


Has the diversity changed before and after the application of the NSTI/%coverage/%identity cut-off ?

Remaining diversity after filtering



Remaining diversity after filtering

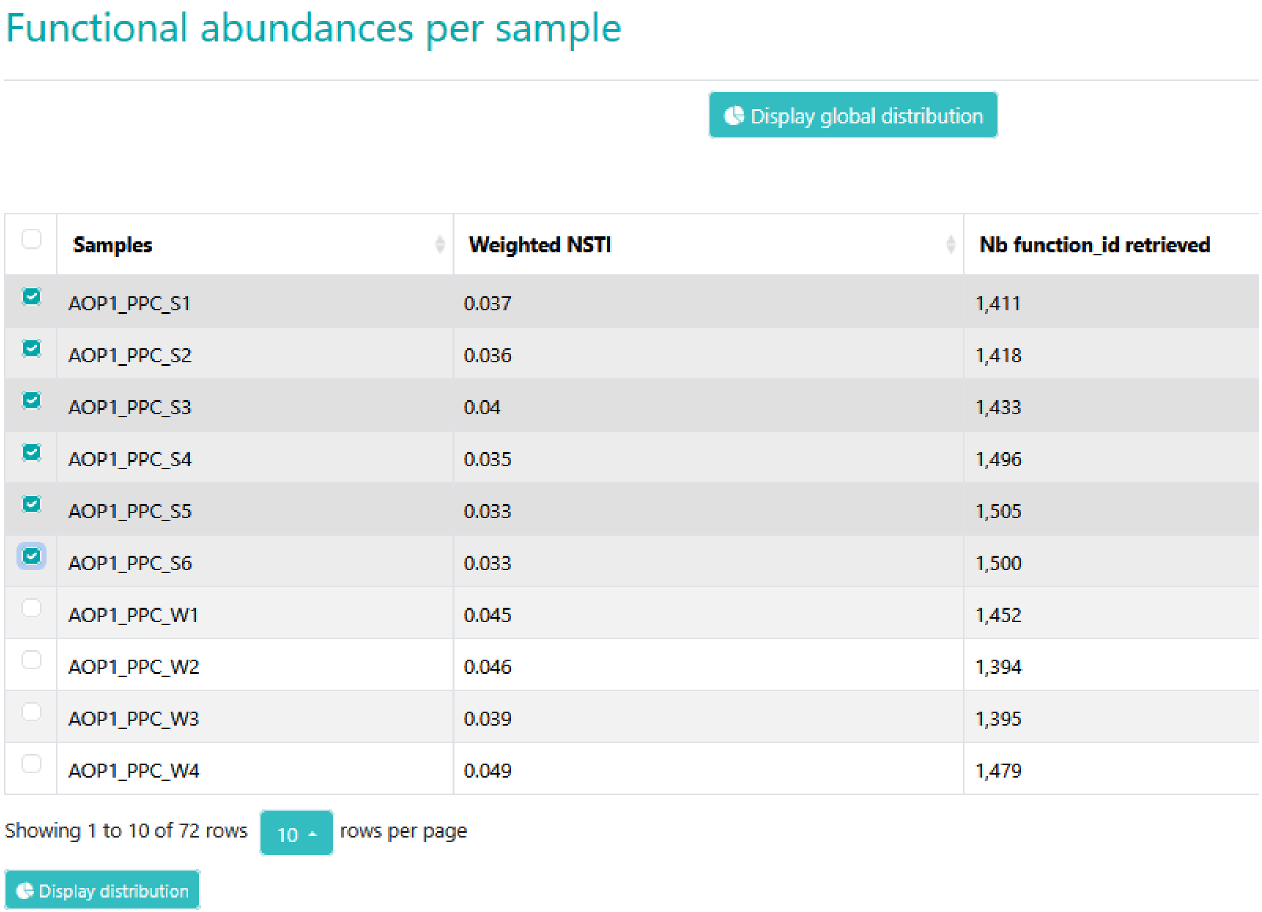


Practice session

Display the distribution of predicted functions in AOP1_PPC_S* abundances.

16S

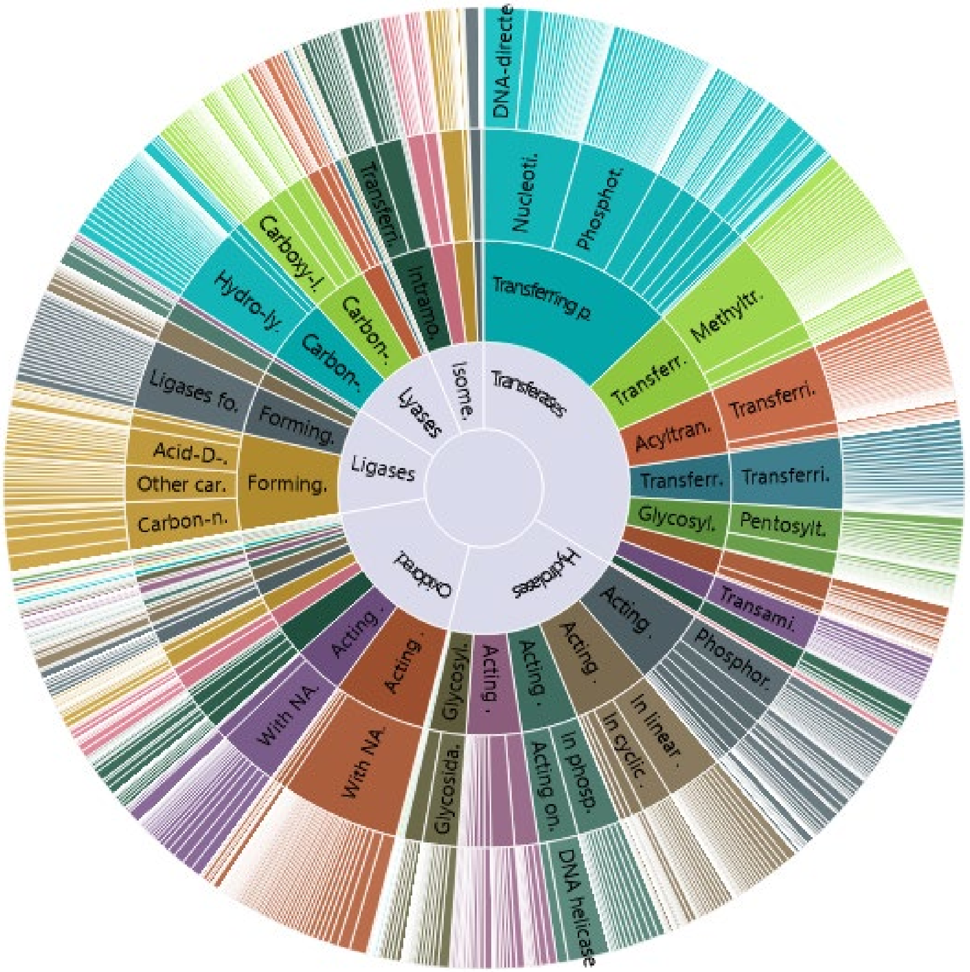
swarm



Functional abundances per sample

☐	Samples	Weighted NSTI	Nb function_id retrieved
☑	AOP1_PPC_S1	0.037	1,411
☑	AOP1_PPC_S2	0.036	1,418
☑	AOP1_PPC_S3	0.04	1,433
☑	AOP1_PPC_S4	0.035	1,496
☑	AOP1_PPC_S5	0.033	1,505
☑	AOP1_PPC_S6	0.033	1,500
☐	AOP1_PPC_W1	0.045	1,452
☐	AOP1_PPC_W2	0.046	1,394
☐	AOP1_PPC_W3	0.039	1,395
☐	AOP1_PPC_W4	0.049	1,479

Showing 1 to 10 of 72 rows 10 rows per page

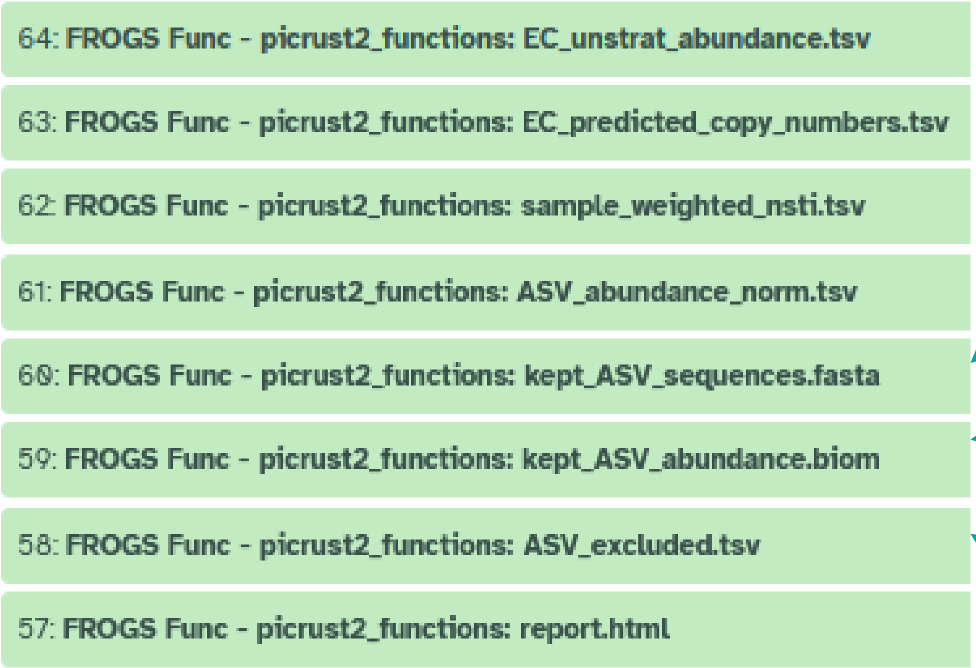


Practice session

Please, check the other output files.

Practice session

Please, check the other output files.



64: FROGS Func - picrust2_functions: EC_unstrat_abundance.tsv

63: FROGS Func - picrust2_functions: EC_predicted_copy_numbers.tsv

62: FROGS Func - picrust2_functions: sample_weighted_nsti.tsv

61: FROGS Func - picrust2_functions: ASV_abundance_norm.tsv

60: FROGS Func - picrust2_functions: kept_ASV_sequences.fasta

59: FROGS Func - picrust2_functions: kept_ASV_abundance.biom

58: FROGS Func - picrust2_functions: ASV_excluded.tsv

57: FROGS Func - picrust2_functions: report.html

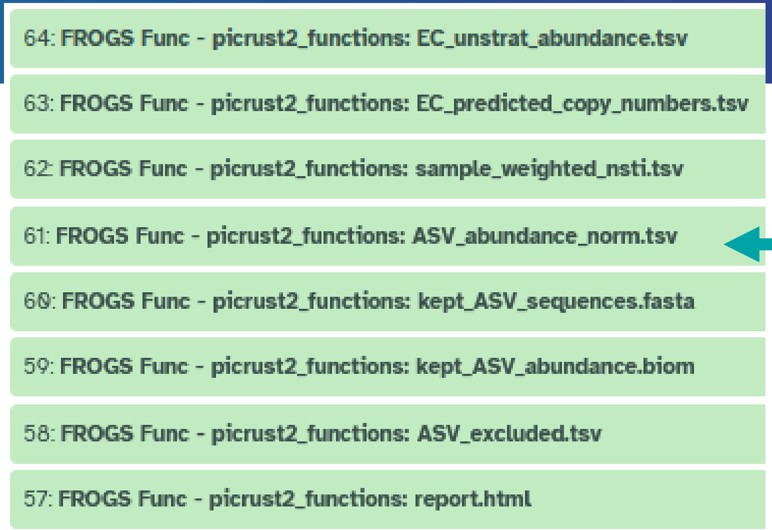
This file contains all the ASV fasta sequences (without those that may be excluded)

Abundance table without those that may be excluded

Information regarding the ASVs that have been removed because they do not meet the thresholds selected at this stage *e.g.* FROGS taxonomy, PICRUSt2 taxonomy, exclusion_parameter, value_parameter

Practice session

Please, check the other output files.



64: FROGS Func - picrust2_functions: EC_unstrat_abundance.tsv

63: FROGS Func - picrust2_functions: EC_predicted_copy_numbers.tsv

62: FROGS Func - picrust2_functions: sample_weighted_nsti.tsv

61: FROGS Func - picrust2_functions: ASV_abundance_norm.tsv

60: FROGS Func - picrust2_functions: kept_ASV_sequences.fasta

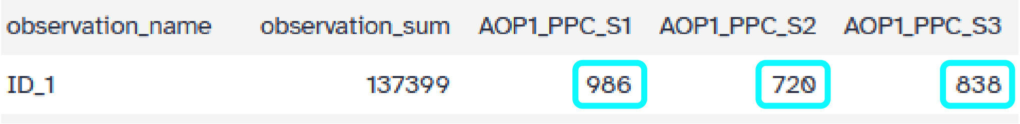
59: FROGS Func - picrust2_functions: kept_ASV_abundance.biom

58: FROGS Func - picrust2_functions: ASV_excluded.tsv

57: FROGS Func - picrust2_functions: report.html

ASV normalized abundance table.

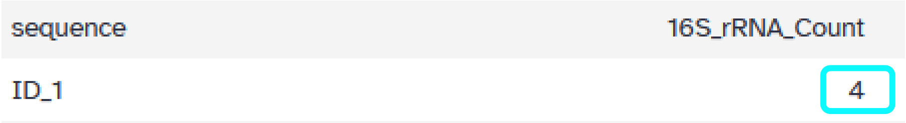
FROGS Core 2 - Companion - biom_to_tsv: abundance.tsv



observation_name	observation_sum	AOP1_PPC_S1	AOP1_PPC_S2	AOP1_PPC_S3
ID_1	137399	986	720	838

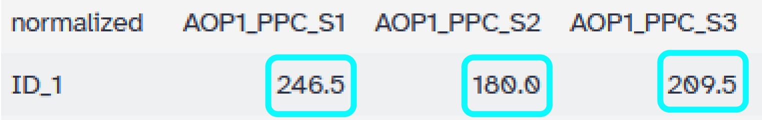
/

FROGS Func '1' - picrust2_**placeseq_and_copynumber**: marker_copy_number_per_ASV.tsv



sequence	16S_rRNA_Count
ID_1	4

FROGS Func '2' - picrust2_**functions**: ASV_abundance_norm.tsv



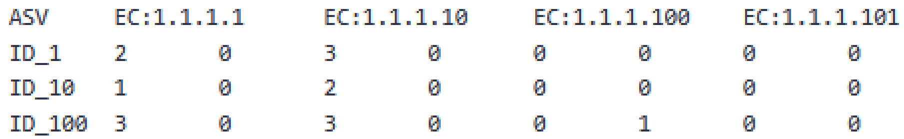
normalized	AOP1_PPC_S1	AOP1_PPC_S2	AOP1_PPC_S3
ID_1	246.5	180.0	209.5

Practice session

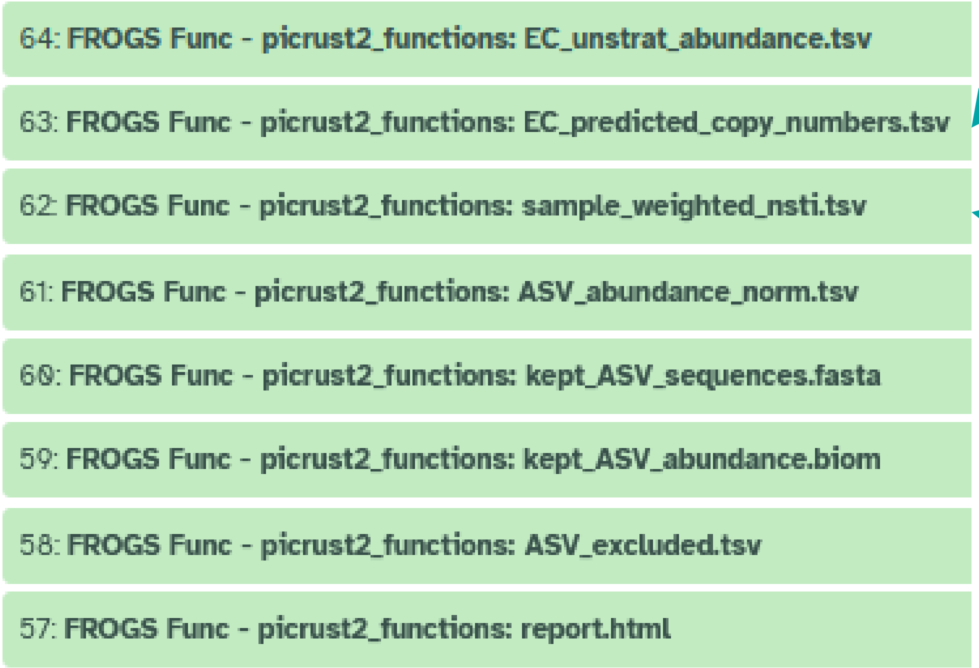
Please, check the other output files.

Output table of predicted function copy numbers per ASV.

One table per chosen target function database
e.g. EC, KO, PFAM, COG, TIGRFAM,PHENO.



```
ASV     EC:1.1.1.1      EC:1.1.1.10     EC:1.1.1.100    EC:1.1.1.101
ID_1    2       0       3       0       0       0       0       0
ID_10   1       0       2       0       0       0       0       0
ID_100  3       0       3       0       0       1       0       0
```



64: **FROGS Func - picrust2_functions: EC_unstrat_abundance.tsv**

63: **FROGS Func - picrust2_functions: EC_predicted_copy_numbers.tsv**

62: **FROGS Func - picrust2_functions: sample_weighted_nsti.tsv**

61: **FROGS Func - picrust2_functions: ASV_abundance_norm.tsv**

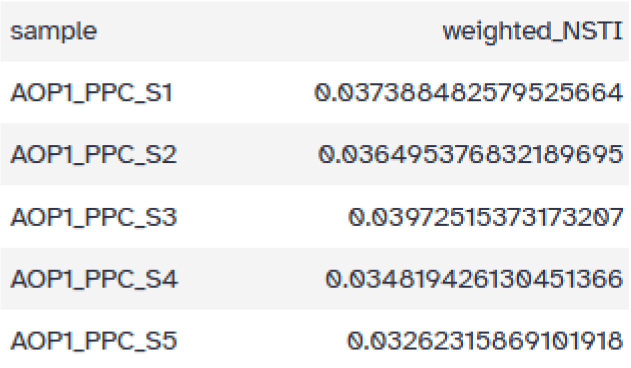
60: **FROGS Func - picrust2_functions: kept_ASV_sequences.fasta**

59: **FROGS Func - picrust2_functions: kept_ASV_abundance.biom**

58: **FROGS Func - picrust2_functions: ASV_excluded.tsv**

57: **FROGS Func - picrust2_functions: report.html**

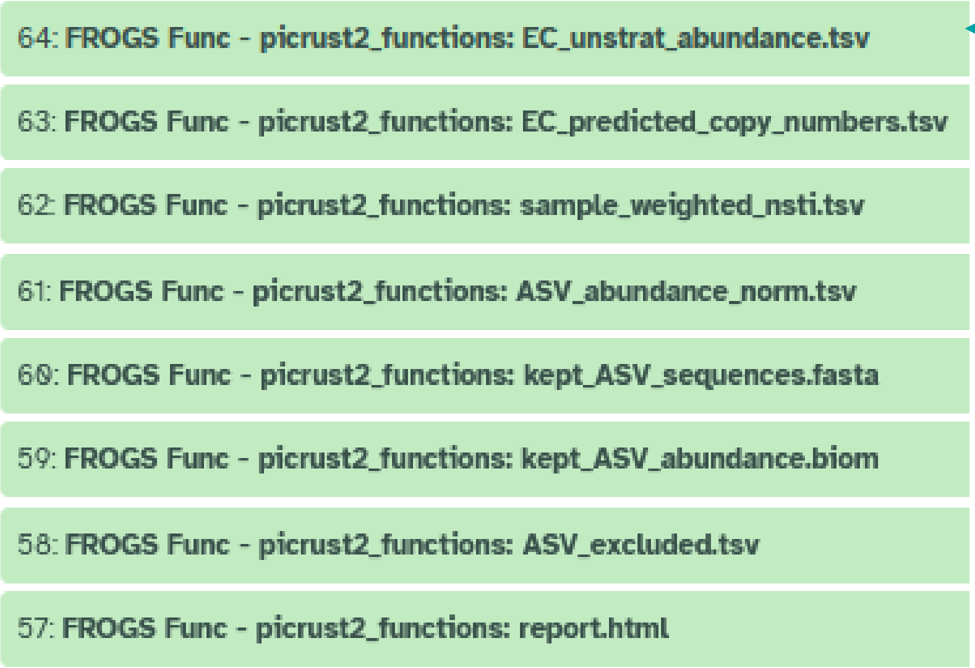
the mean of NSTI value per sample.



sample	weighted_NSTI
AOP1_PPC_S1	0.037388482579525664
AOP1_PPC_S2	0.03649537683218969‌5
AOP1_PPC_S3	0.03972515373173207
AOP1_PPC_S4	0.034819426130451366
AOP1_PPC_S5	0.03262315869101918

Practice session

Please, check the other output files.



64: FROGS Func - picrust2_functions: EC_unstrat_abundance.tsv

63: FROGS Func - picrust2_functions: EC_predicted_copy_numbers.tsv

62: FROGS Func - picrust2_functions: sample_weighted_nsti.tsv

61: FROGS Func - picrust2_functions: ASV_abundance_norm.tsv

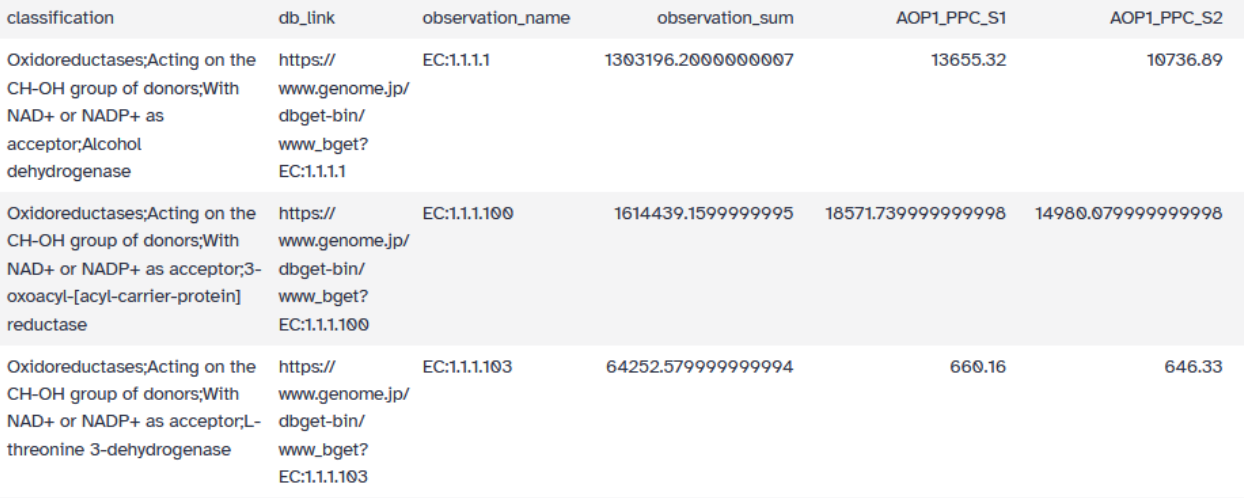
60: FROGS Func - picrust2_functions: kept_ASV_sequences.fasta

59: FROGS Func - picrust2_functions: kept_ASV_abundance.biom

58: FROGS Func - picrust2_functions: ASV_excluded.tsv

57: FROGS Func - picrust2_functions: report.html

Abundance table per function and per sample



classification	db_link	observation_name	observation_sum	AOP1_PPC_S1	AOP1_PPC_S2
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;Alcohol dehydrogenase	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.1	EC:1.1.1.1	1303196.2000000007	13655.32	10736.89
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;3-oxoacyl-[acyl-carrier-protein] reductase	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.100	EC:1.1.1.100	1614439.1599999995	18571.739999999998	14980.079999999998
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;L-threonine 3-dehydrogenase	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.103	EC:1.1.1.103	64252.579999999994	660.16	646.33

You can do statistical analyses of the information in this table.

FROGS Func

FROGS Func - pathway abundance estimation

How does it work?

1. Places the ASVs into a reference phylogenetic tree and predicts the number of marker copies in each ASV.

FROGS Func - sequence placement & gene copy nb estimation

2. Estimates the number of copies of each function (gene) in each organism (ASV). Also calculates how many functions (genes) are present in each sample. And determines how many of ASVs there are using the number of copies of the marker.

FROGS Func - functional abundance estimation

3. Calculates how much of each pathway is present in each sample.

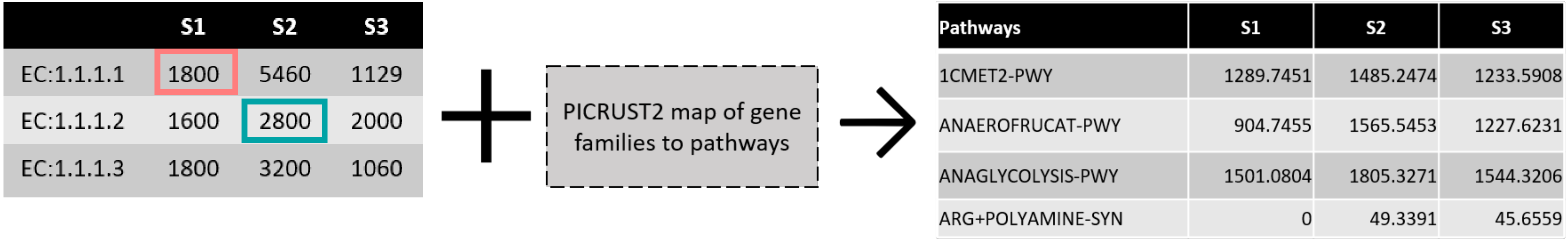
FROGS Func - pathways abundance estimation

It runs only on 16S, ITS or 18S

Overview

FROGS Func - pathway abundance estimation tool infers MetaCyc/KEGG pathway abundances based on EC/KO number abundances.

- Regroups EC/KO numbers to MetaCyc/KEGG reactions.
- Infers which MetaCyc/KEGG pathways are present based on these reactions with MinPath.
- Calculates and returns the abundance of pathways identified as present.



FROGS Func ‘2’ EC unstrat abundance table

	S1	S2	S3
EC:1.1.1.1	1800	5460	1129
EC:1.1.1.2	1600	2800	2000
EC:1.1.1.3	1800	3200	1060

PICRUST2 map of gene families to pathways

FROGS Func ‘3’pathways_unstrat per sample and per reference table

Pathways	S1	S2	S3
1CMET2-PWY	1289.7451	1485.2474	1233.5908
ANAEROFRUCAT-PWY	904.7455	1565.5453	1227.6231
ANAGLYCOLYSIS-PWY	1501.0804	1805.3271	1544.3206
ARG+POLYAMINE-SYN	0	49.3391	45.6559

Practice session

Please open the FROGS Func Tool 3 and familiarize yourself with the required parameters.

Please, enter all the information you have/understand.

- 16S/ITS
- Metacyc
- No normalization

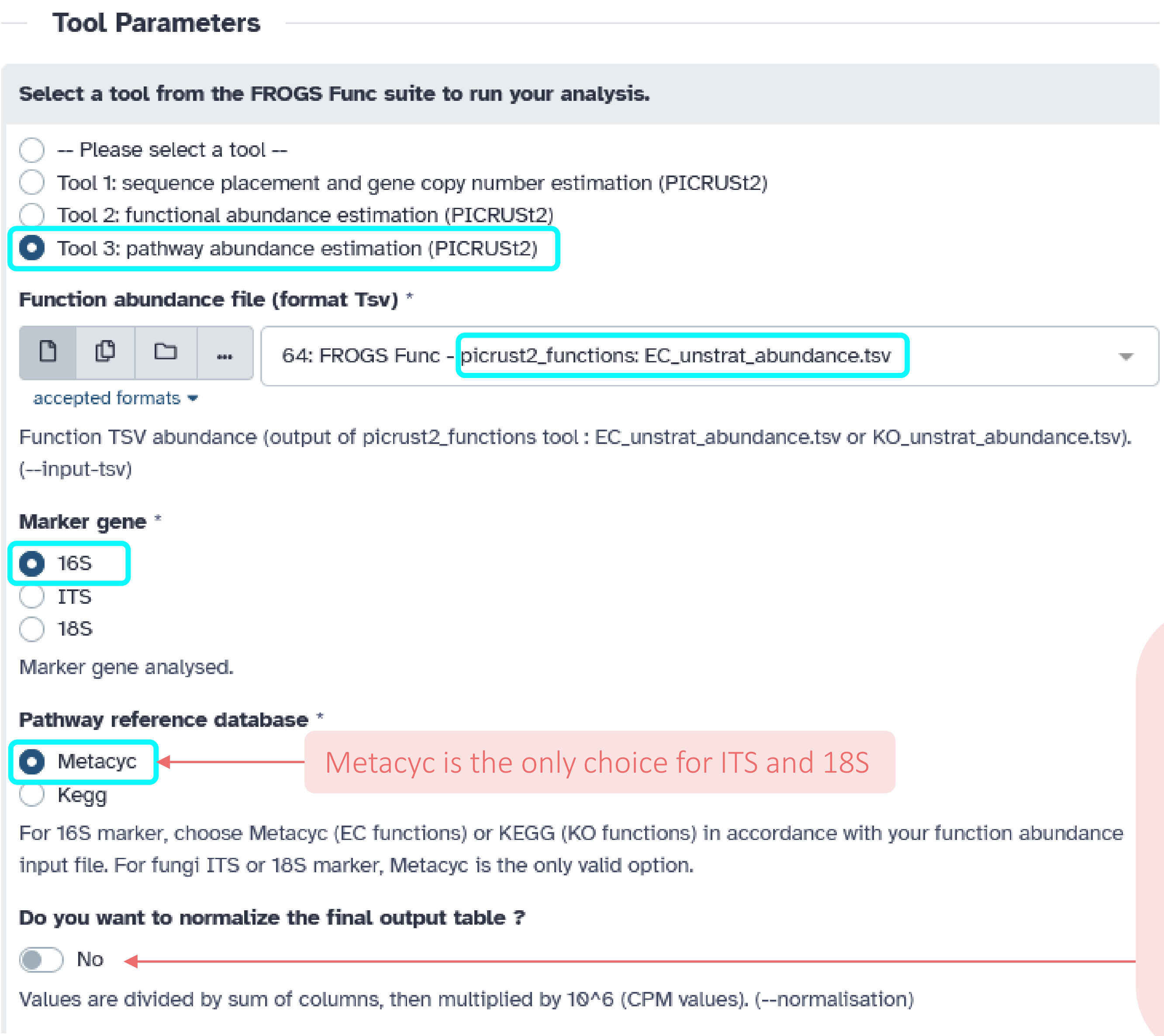
Perhaps, it would be better not to continue using the ITS dataset.

Run the process !

Practice session

16S

swarm



Tool Parameters

Select a tool from the FROGS Func suite to run your analysis.

○ -- Please select a tool --
○ Tool 1: sequence placement and gene copy number estimation (PICRUSt2)
○ Tool 2: functional abundance estimation (PICRUSt2)
● Tool 3: pathway abundance estimation (PICRUSt2)

Function abundance file (format Tsv) *

64: FROGS Func - picrust2_functions: EC_unstrat_abundance.tsv

accepted formats

Function TSV abundance (output of picrust2_functions tool : EC_unstrat_abundance.tsv or KO_unstrat_abundance.tsv). (--input-tsv)

Marker gene *

● 16S
○ ITS
○ 18S

Marker gene analysed.

Pathway reference database *

● Metacyc
○ Kegg

Metacyc is the only choice for ITS and 18S

For 16S marker, choose Metacyc (EC functions) or KEGG (KO functions) in accordance with your function abundance input file. For fungi ITS or 18S marker, Metacyc is the only valid option.

Do you want to normalize the final output table ?

No

Values are divided by sum of columns, then multiplied by 10^6 (CPM values). (--normalisation)

normalization: values are divided by sum of columns, then multiplied by 10^6 (Count Per Million values).

This normalization allows to compare the samples between them. But to perform more precise statistical analysis, some tools as DESeq2 need the non-normalized abundance table to perform the normalization by themselves.

So be careful which table to use for further analysis.

Practice session

Explore the report.html output.

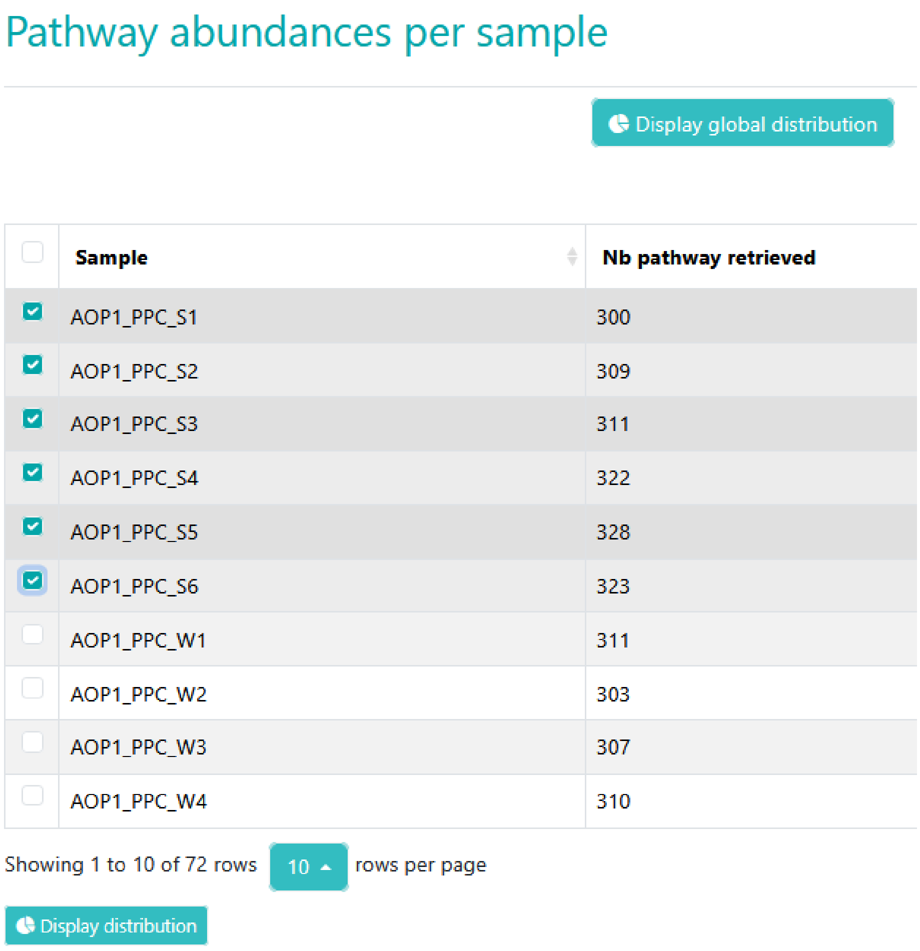
Display the distribution of predicted pathways in AOP1_PPC_S* abundances.

Explore the ‘Metacyc_unstrat_abundance.tsv’ output.

Practice session

Display the distribution of predicted pathways in AOP1_PPC_S* abundances.

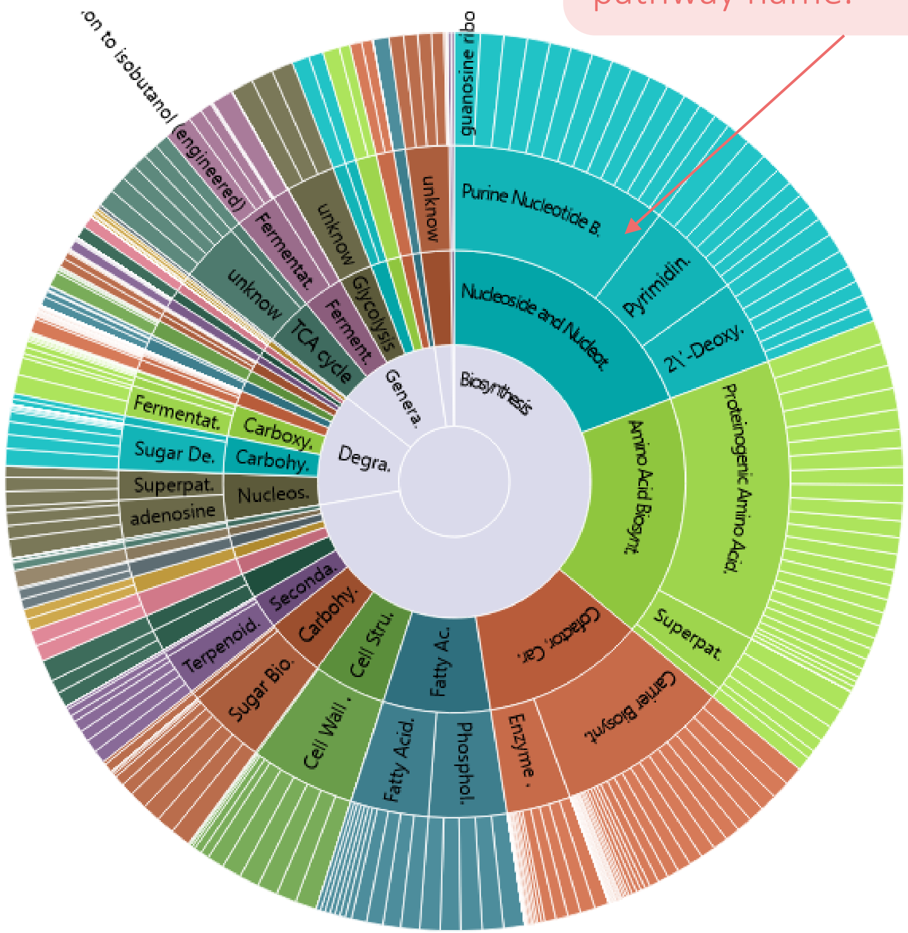
For more pathway details, double-click on a the interest pathway name.



Pathway abundances per sample

☐	Sample	Nb pathway retrieved
☑	AOP1_PPC_S1	300
☑	AOP1_PPC_S2	309
☑	AOP1_PPC_S3	311
☑	AOP1_PPC_S4	322
☑	AOP1_PPC_S5	328
☑	AOP1_PPC_S6	323
☐	AOP1_PPC_W1	311
☐	AOP1_PPC_W2	303
☐	AOP1_PPC_W3	307
☐	AOP1_PPC_W4	310

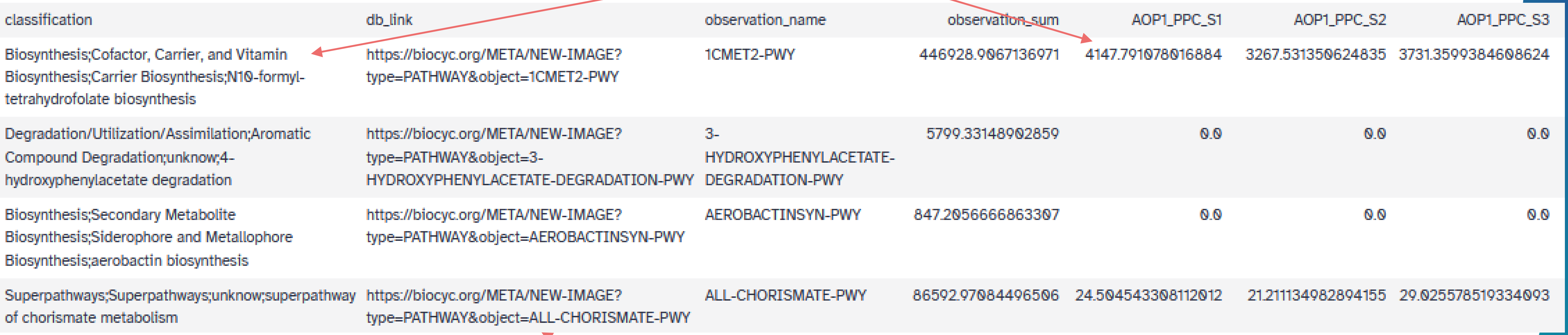
Showing 1 to 10 of 72 rows 10 rows per page



Practice session

Explore the ‘Metacyc_unstrat_abundance.tsv’ output.

Abundance of each pathway in each sample.



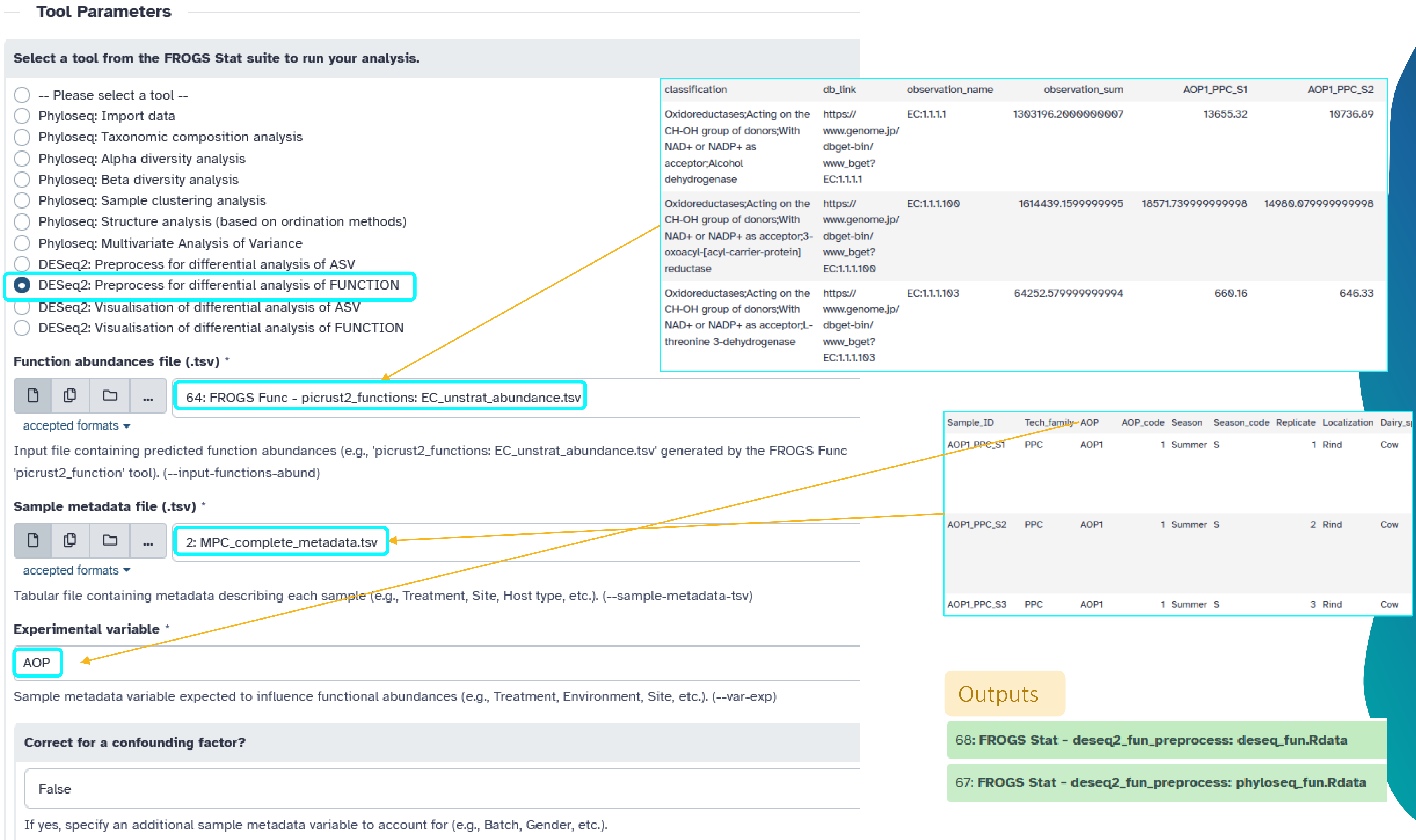
classification	db_link	observation_name	observation_sum	AOP1_PPC_S1	AOP1_PPC_S2	AOP1_PPC_S3
Biosynthesis;Cofactor, Carrier, and Vitamin Biosynthesis;Carrier Biosynthesis;N10-formyl-tetrahydrofolate biosynthesis	https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&object=1CMET2-PWY	1CMET2-PWY	446928.9067136971	4147.791078016884	3267.531350624835	3731.3599384608624
Degradation/Utilization/Assimilation;Aromatic Compound Degradation;unknow;4-hydroxyphenylacetate degradation	https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&object=3-HYDROXYPHENYLACETATE-DEGRADATION-PWY	3-HYDROXYPHENYLACETATE-DEGRADATION-PWY	5799.33148902859	0.0	0.0	0.0
Biosynthesis;Secondary Metabolite Biosynthesis;Siderophore and Metallophore Biosynthesis;aerobactin biosynthesis	https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&object=AEROBACTINSYN-PWY	AEROBACTINSYN-PWY	847.2056666863307	0.0	0.0	0.0
Superpathways;Superpathways;unknow;superpathway of chorismate metabolism	https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&object=ALL-CHORISMATE-PWY	ALL-CHORISMATE-PWY	86592.97084496506	24.504543308112012	21.211134982894155	29.025578519334093

Here is a link that you can copy and paste into your web browser.

From this table of abundance it is quite possible to make statistical analyses on it to understand the information.

Statistics on functional data

Preprocess for differential analysis of FUNCTION



Tool Parameters

Select a tool from the FROGS Stat suite to run your analysis.

- ○ -- Please select a tool --
- ○ Phyloseq: Import data
- ○ Phyloseq: Taxonomic composition analysis
- ○ Phyloseq: Alpha diversity analysis
- ○ Phyloseq: Beta diversity analysis
- ○ Phyloseq: Sample clustering analysis
- ○ Phyloseq: Structure analysis (based on ordination methods)
- ○ Phyloseq: Multivariate Analysis of Variance
- ○ DESeq2: Preprocess for differential analysis of ASV
- ● DESeq2: Preprocess for differential analysis of FUNCTION
- ○ DESeq2: Visualisation of differential analysis of ASV
- ○ DESeq2: Visualisation of differential analysis of FUNCTION

classification	db_link	observation_name	observation_sum	AOP1_PPC_S1	AOP1_PPC_S2
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;Alcohol dehydrogenase	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.1	EC:1.1.1.1	1303196.2000000007	13655.32	10736.89
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;3-oxoacyl-[acyl-carrier-protein] reductase	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.100	EC:1.1.1.100	1614439.1599999995	18571.739999999998	14980.079999999998
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;L-threonine 3-dehydrogenase	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.103	EC:1.1.1.103	64252.579999999994	660.16	646.33

Function abundances file (.tsv) *

64: FROGS Func - picrust2_functions: EC_unstrat_abundance.tsv

accepted formats ▾

Input file containing predicted function abundances (e.g., 'picrust2_functions: EC_unstrat_abundance.tsv' generated by the FROGS Func 'picrust2_function' tool). (--input-functions-abund)

Sample metadata file (.tsv) *

2: MPC_complete_metadata.tsv

accepted formats ▾

Tabular file containing metadata describing each sample (e.g., Treatment, Site, Host type, etc.). (--sample-metadata-tsv)

Sample_ID	Tech_family	AOP	AOP_code	Season	Season_code	Replicate	Localization	Dairy_s
AOP1_PPC_S1	PPC	AOP1	1	Summer	S	1	Rind	Cow
AOP1_PPC_S2	PPC	AOP1	1	Summer	S	2	Rind	Cow
AOP1_PPC_S3	PPC	AOP1	1	Summer	S	3	Rind	Cow

Experimental variable *

AOP

Sample metadata variable expected to influence functional abundances (e.g., Treatment, Environment, Site, etc.). (--var-exp)

Correct for a confounding factor?

False

If yes, specify an additional sample metadata variable to account for (e.g., Batch, Gender, etc.).

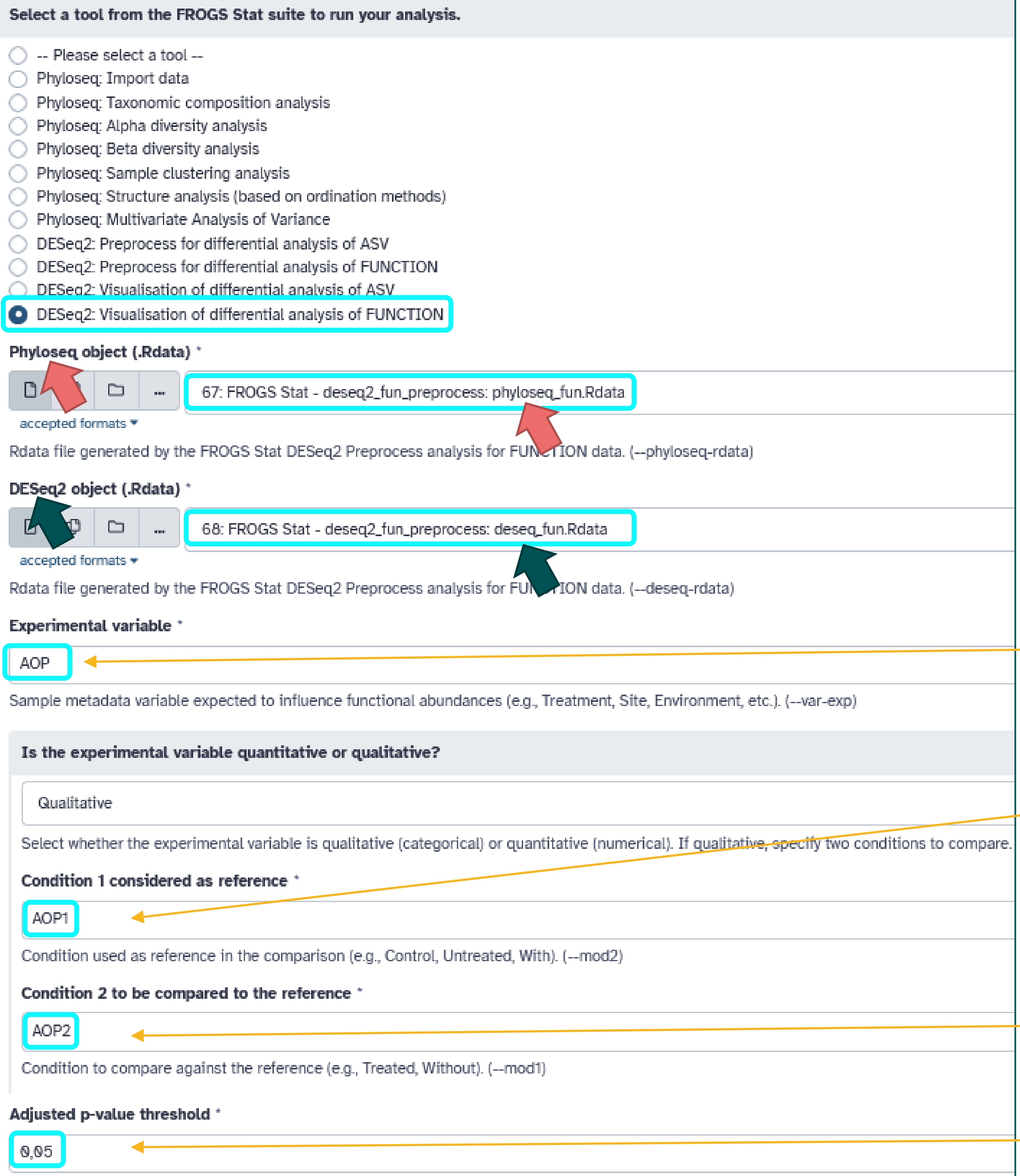
Outputs

68: **FROGS Stat - deseq2_fun_preprocess: deseq_fun.Rdata**

67: **FROGS Stat - deseq2_fun_preprocess: phyloseq_fun.Rdata**

Statistics on functional data

DESeq2: Visualisation of differential analysis of FUNCTION



Select a tool from the FROGS Stat suite to run your analysis.

- ○ -- Please select a tool --
- ○ Phyloseq: Import data
- ○ Phyloseq: Taxonomic composition analysis
- ○ Phyloseq: Alpha diversity analysis
- ○ Phyloseq: Beta diversity analysis
- ○ Phyloseq: Sample clustering analysis
- ○ Phyloseq: Structure analysis (based on ordination methods)
- ○ Phyloseq: Multivariate Analysis of Variance
- ○ DESeq2: Preprocess for differential analysis of ASV
- ○ DESeq2: Preprocess for differential analysis of FUNCTION
- ○ DESeq2: Visualisation of differential analysis of ASV
- ● DESeq2: Visualisation of differential analysis of FUNCTION

Phyloseq object (.Rdata) *

67: FROGS Stat - deseq2_fun_preprocess: phyloseq_fun.Rdata

Rdata file generated by the FROGS Stat DESeq2 Preprocess analysis for FUNCTION data. (--phyloseq-rdata)

DESeq2 object (.Rdata) *

68: FROGS Stat - deseq2_fun_preprocess: deseq_fun.Rdata

Rdata file generated by the FROGS Stat DESeq2 Preprocess analysis for FUNCTION data. (--deseq-rdata)

Experimental variable *

AOP

Sample metadata variable expected to influence functional abundances (e.g., Treatment, Site, Environment, etc.). (--var-exp)

Is the experimental variable quantitative or qualitative?

Qualitative

Select whether the experimental variable is qualitative (categorical) or quantitative (numerical). If qualitative, specify two conditions to compare.

Condition 1 considered as reference *

AOP1

Condition used as reference in the comparison (e.g., Control, Untreated, With). (--mod2)

Condition 2 to be compared to the reference *

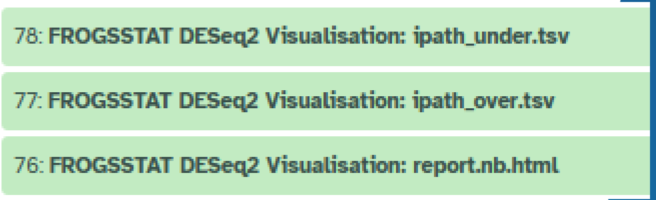
AOP2

Condition to compare against the reference (e.g., Treated, Without). (--mod1)

Adjusted p-value threshold *

0,05

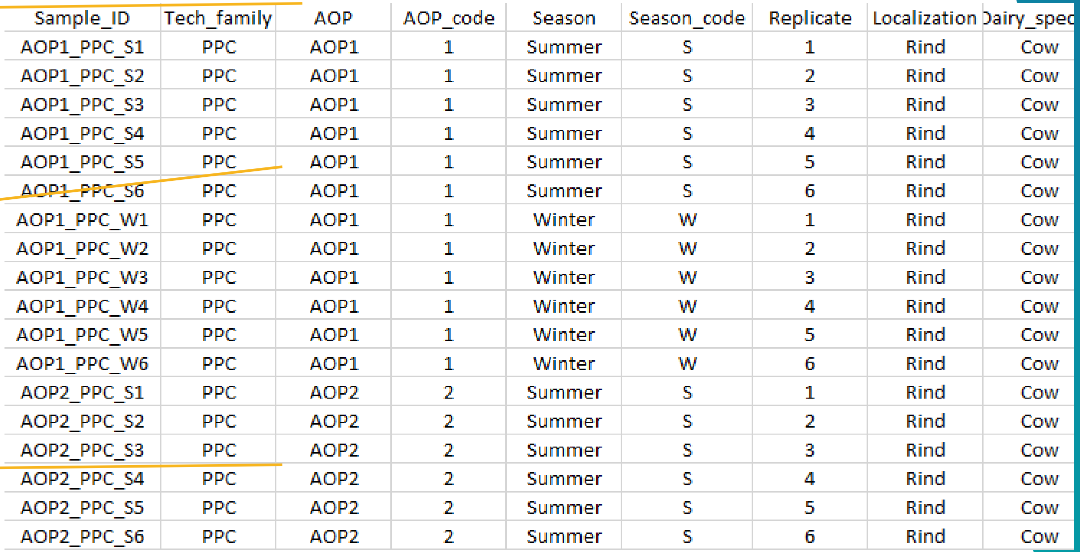
Outputs



78: **FROGSSTAT DESeq2 Visualisation: ipath_under.tsv**

77: **FROGSSTAT DESeq2 Visualisation: ipath_over.tsv**

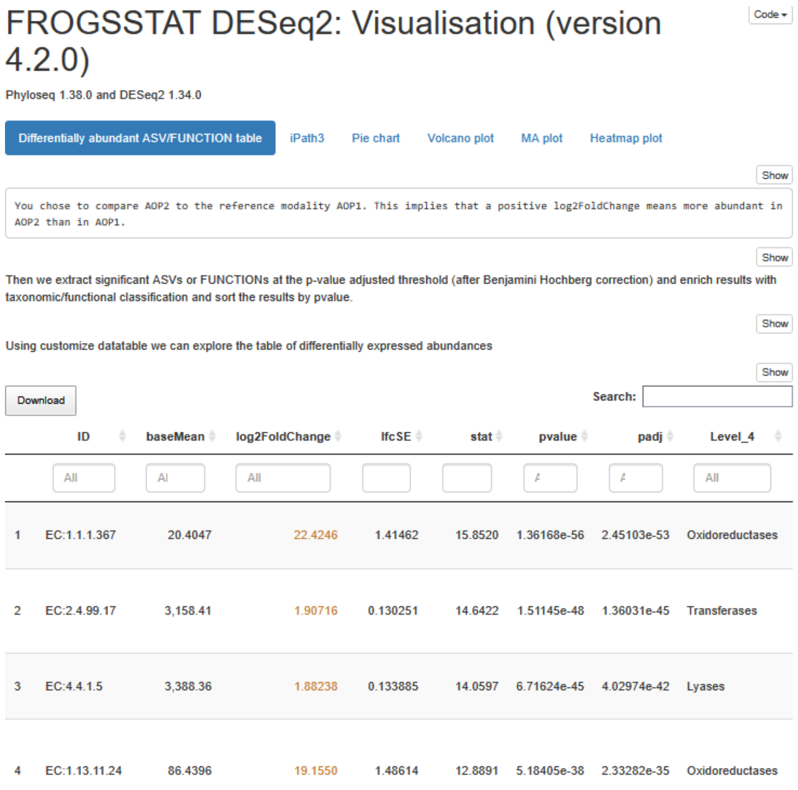
76: **FROGSSTAT DESeq2 Visualisation: report.nb.html**



Sample_ID	Tech_family	AOP	AOP_code	Season	Season_code	Replicate	Localization	Dairy_spec
AOP1_PPC_S1	PPC	AOP1	1	Summer	S	1	Rind	Cow
AOP1_PPC_S2	PPC	AOP1	1	Summer	S	2	Rind	Cow
AOP1_PPC_S3	PPC	AOP1	1	Summer	S	3	Rind	Cow
AOP1_PPC_S4	PPC	AOP1	1	Summer	S	4	Rind	Cow
AOP1_PPC_S5	PPC	AOP1	1	Summer	S	5	Rind	Cow
AOP1_PPC_S6	PPC	AOP1	1	Summer	S	6	Rind	Cow
AOP1_PPC_W1	PPC	AOP1	1	Winter	W	1	Rind	Cow
AOP1_PPC_W2	PPC	AOP1	1	Winter	W	2	Rind	Cow
AOP1_PPC_W3	PPC	AOP1	1	Winter	W	3	Rind	Cow
AOP1_PPC_W4	PPC	AOP1	1	Winter	W	4	Rind	Cow
AOP1_PPC_W5	PPC	AOP1	1	Winter	W	5	Rind	Cow
AOP1_PPC_W6	PPC	AOP1	1	Winter	W	6	Rind	Cow
AOP2_PPC_S1	PPC	AOP2	2	Summer	S	1	Rind	Cow
AOP2_PPC_S2	PPC	AOP2	2	Summer	S	2	Rind	Cow
AOP2_PPC_S3	PPC	AOP2	2	Summer	S	3	Rind	Cow
AOP2_PPC_S4	PPC	AOP2	2	Summer	S	4	Rind	Cow
AOP2_PPC_S5	PPC	AOP2	2	Summer	S	5	Rind	Cow
AOP2_PPC_S6	PPC	AOP2	2	Summer	S	6	Rind	Cow

To start with, we'll set the p-value to 0.05.

FROGS Stat - deseq2_fun_visualisation: report.nb.html



FROGSSTAT DESeq2: Visualisation (version 4.2.0)

Phyloseq 1.38.0 and DESeq2 1.34.0

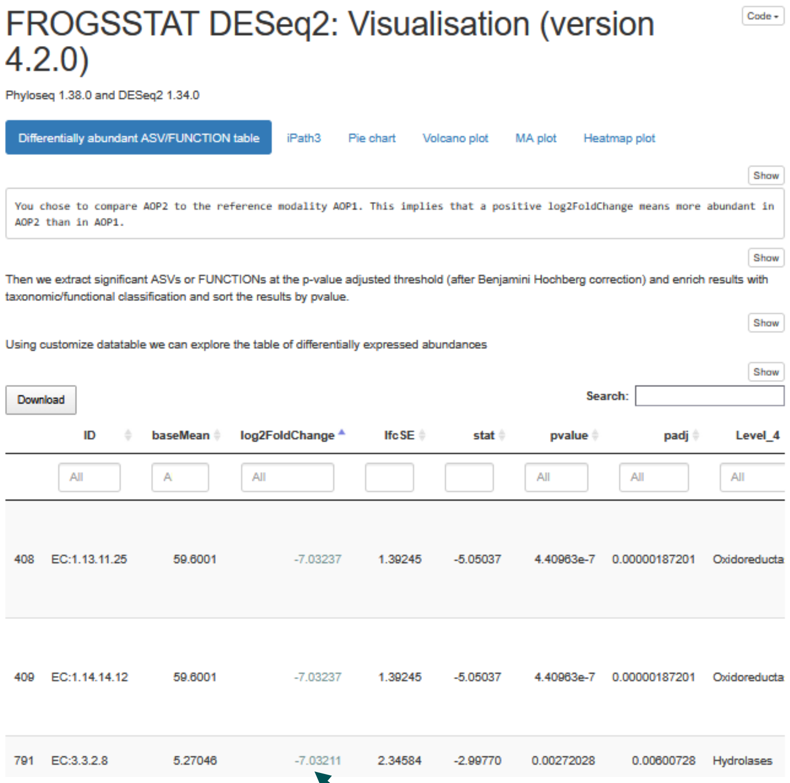
```
You chose to compare AOP2 to the reference modality AOP1. This implies that a positive log2FoldChange means more abundant in
AOP2 than in AOP1.
```

Then we extract significant ASVs or FUNCTIONs at the p-value adjusted threshold (after Benjamini Hochberg correction) and enrich results with taxonomic/functional classification and sort the results by pvalue.

Using customize datatable we can explore the table of differentially expressed abundances

	ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Level_4
1	EC:1.1.1.367	20.4047	22.4246	1.41462	15.8520	1.36168e-56	2.45103e-53	Oxidoreductases
2	EC:2.4.99.17	3,158.41	1.90716	0.130251	14.6422	1.51145e-48	1.36031e-45	Transferases
3	EC:4.4.1.5	3,388.36	1.88238	0.133885	14.0597	6.71624e-45	4.02974e-42	Lyases
4	EC:1.13.11.24	86.4396	19.1550	1.48614	12.8891	5.18405e-38	2.33282e-35	Oxidoreductases

We chose to compare AOP2 to the reference modality AOP1. This implies that a positive log2FoldChange means more abundant in AOP2 than in AOP1.



FROGSSTAT DESeq2: Visualisation (version 4.2.0)

Phyloseq 1.38.0 and DESeq2 1.34.0

```
You chose to compare AOP2 to the reference modality AOP1. This implies that a positive log2FoldChange means more abundant in
AOP2 than in AOP1.
```

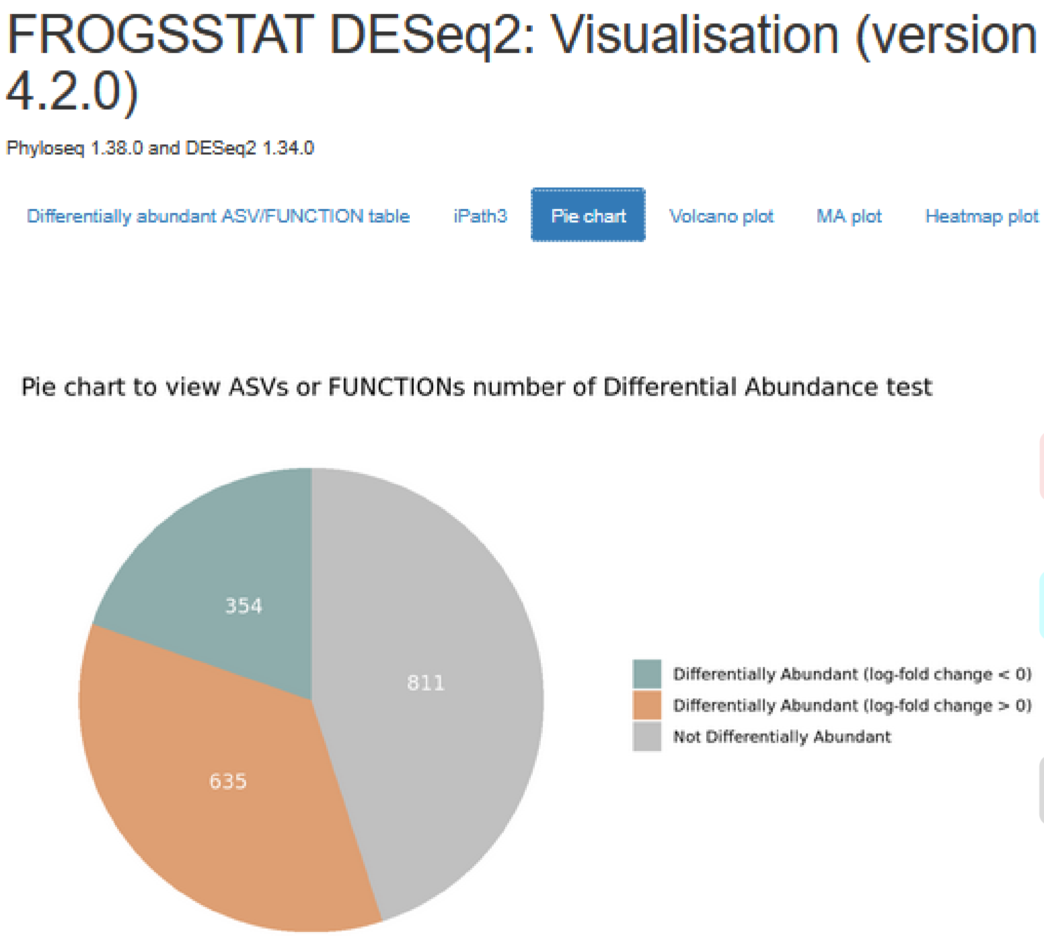
Then we extract significant ASVs or FUNCTIONs at the p-value adjusted threshold (after Benjamini Hochberg correction) and enrich results with taxonomic/functional classification and sort the results by pvalue.

Using customize datatable we can explore the table of differentially expressed abundances

	ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Level_4
408	EC:1.13.11.25	59.6001	-7.03237	1.39245	-5.05037	4.40963e-7	0.00000187201	Oxidoreducta
409	EC:1.14.14.12	59.6001	-7.03237	1.39245	-5.05037	4.40963e-7	0.00000187201	Oxidoreducta
791	EC:3.3.2.8	5.27046	-7.03211	2.34584	-2.99770	0.00272028	0.00600728	Hydrolases

We chose to compare AOP2 to the reference modality AOP1. This implies that a negative log2FoldChange means more abundant in AOP1 than in AOP2.

FROGS Stat - deseq2_fun_visualisation: report.nb.html



FROGSSTAT DESeq2: Visualisation (version 4.2.0)

Phyloseq 1.38.0 and DESeq2 1.34.0

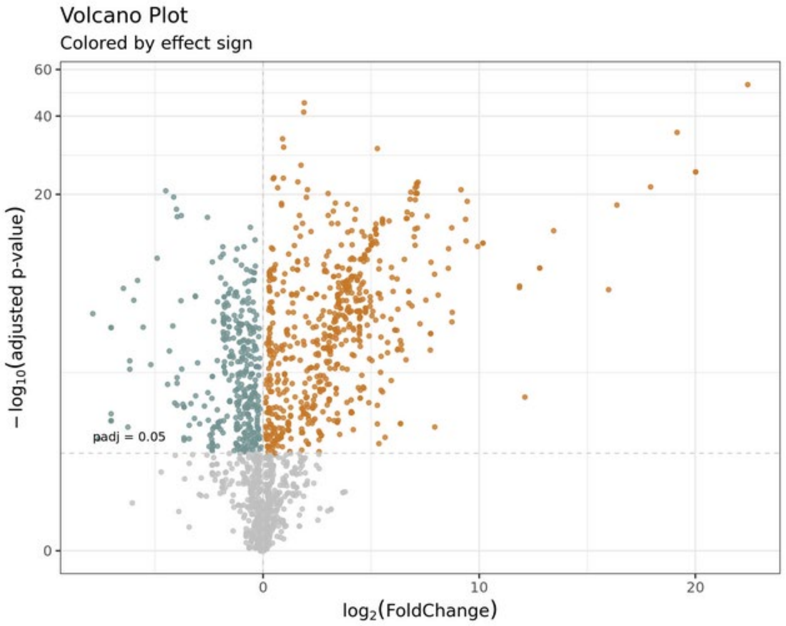
Pie chart to view ASVs or FUNCTIONs number of Differential Abundance test

635 functions are more abundant in AOP2 than in AOP1

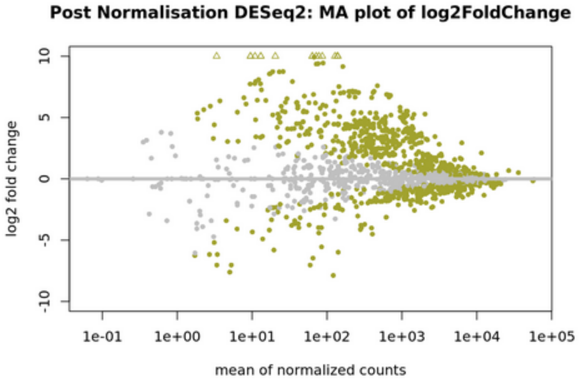
354 functions are more abundant in AOP1 than in AOP2

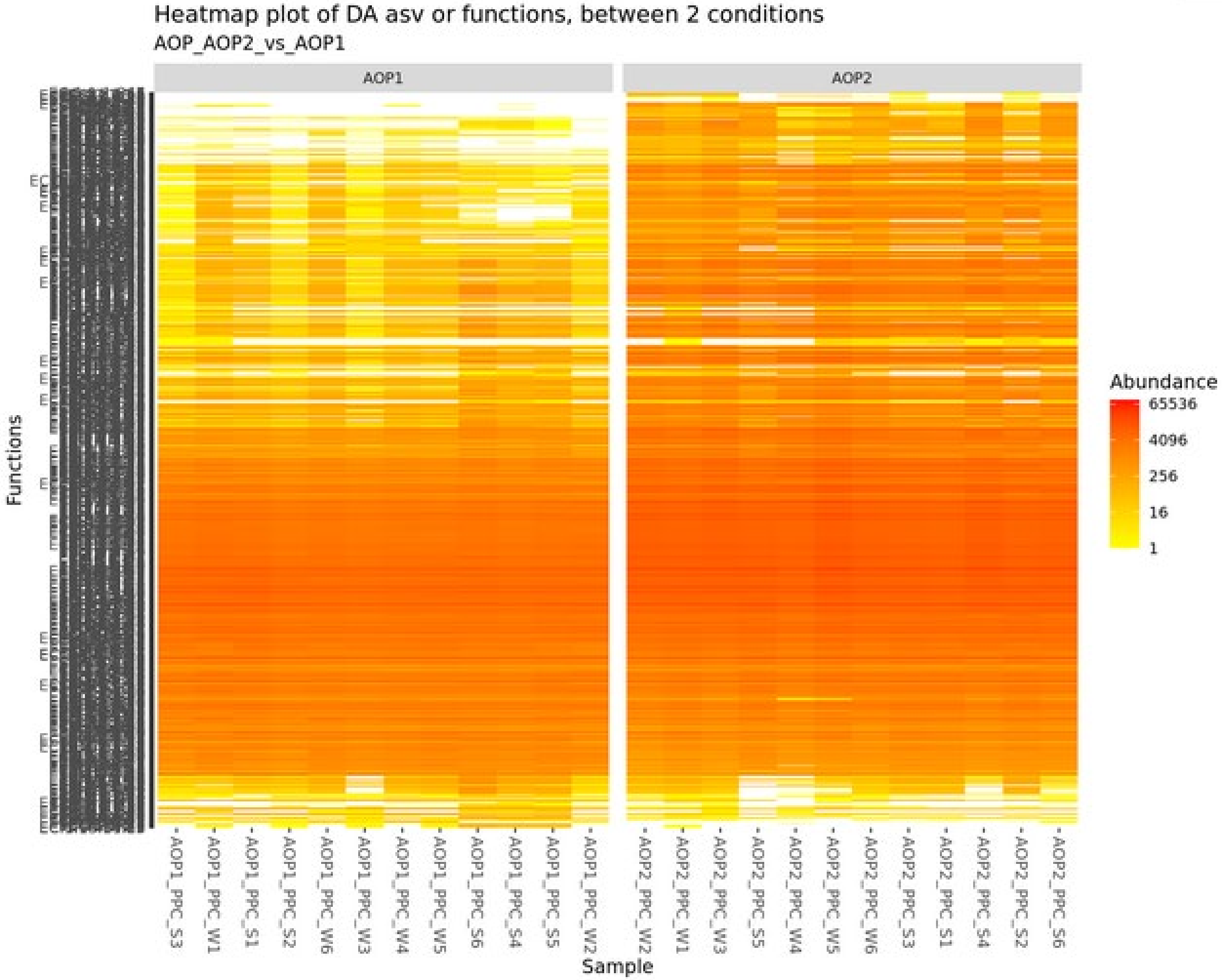
811 functions are not differentially abundant in AOP1 or in AOP2

FROGS Stat - deseq2_fun_visualisation: report.nb.html

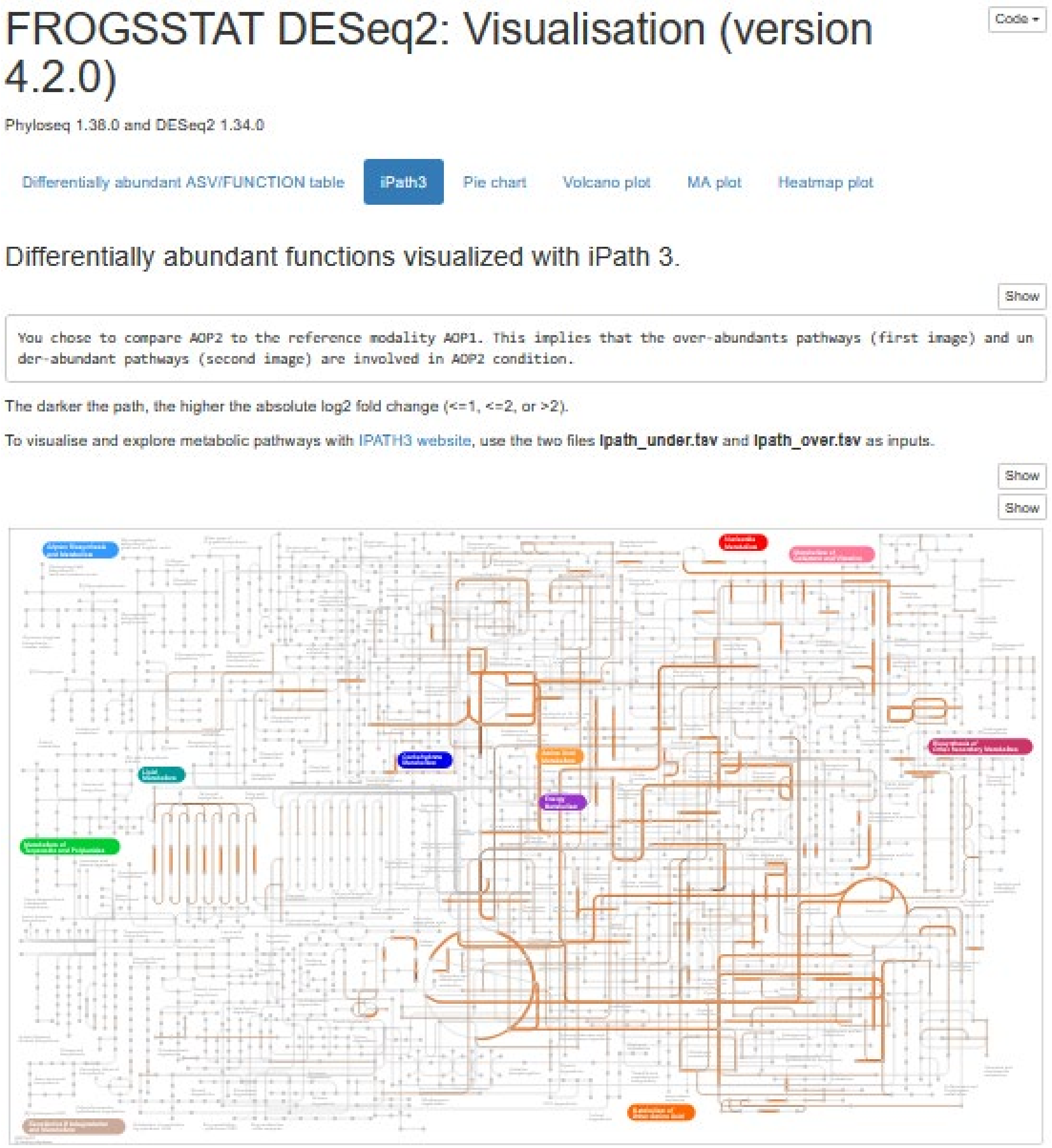

Volcano Plot

Colored by effect sign


Post Normalisation DESeq2: MA plot of log2FoldChange


Heatmap plot of DA asv or functions, between 2 conditions

AOP_AOP2_vs_AOP1



FROGSSTAT DESeq2: Visualisation (version 4.2.0)

Phyloseq 1.38.0 and DESeq2 1.34.0

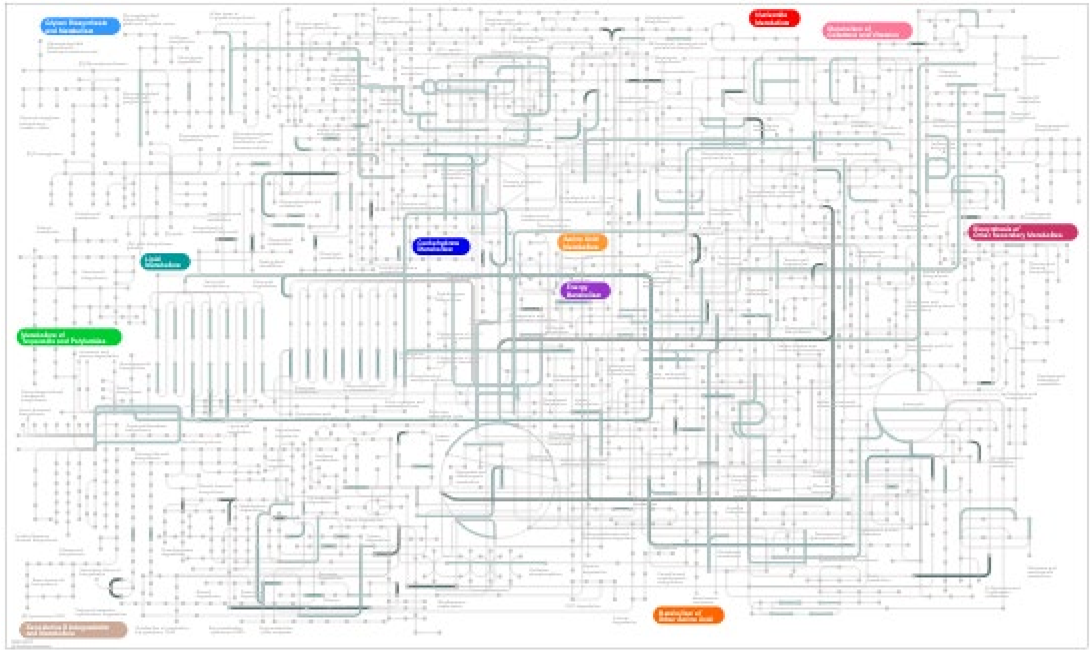
Differentially abundant functions visualized with iPath 3.

```
You chose to compare AOP2 to the reference modality AOP1. This implies that the over-abundants pathways (first image) and un
der-abundant pathways (second image) are involved in AOP2 condition.
```

The darker the path, the higher the absolute log2 fold change (<=1, <=2, or >2).

To visualise and explore metabolic pathways with IPATH3 website, use the two files **Ipath_under.tsv** and **Ipath_over.tsv** as inputs.

We chose to compare AOP2 to the reference modality AOP1. This implies that the over-abundants pathways (first image) and under-abundant pathways (second image) are involved in AOP2 condition.



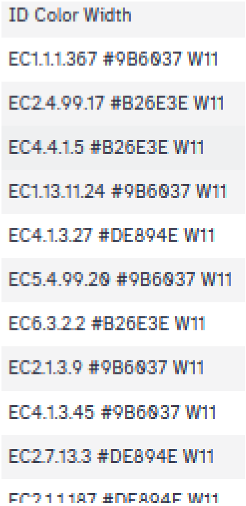
The darker the path, the higher the log2 fold change (<1, <2, or >2).

IPATH3

To visualise and explore metabolic pathways with IPATH3 website, use the two files ipath_under.tsv and ipath_over.tsv as inputs.

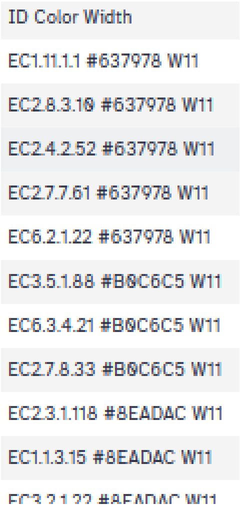
81: FROGSSTAT DESeq2 Visualisation: ipath_under.tsv

80: FROGSSTAT DESeq2 Visualisation: ipath_over.tsv



```
ID Color Width
EC1.1.1.367 #9B6037 W11
EC2.4.99.17 #B26E3E W11
EC4.4.1.5 #B26E3E W11
EC1.13.11.24 #9B6037 W11
EC4.1.3.27 #DE894E W11
EC5.4.99.20 #9B6037 W11
EC6.3.2.2 #B26E3E W11
EC2.1.3.9 #9B6037 W11
EC4.1.3.45 #9B6037 W11
EC2.7.13.3 #DE894E W11
EC2.1.1.187 #DE894E W11
```

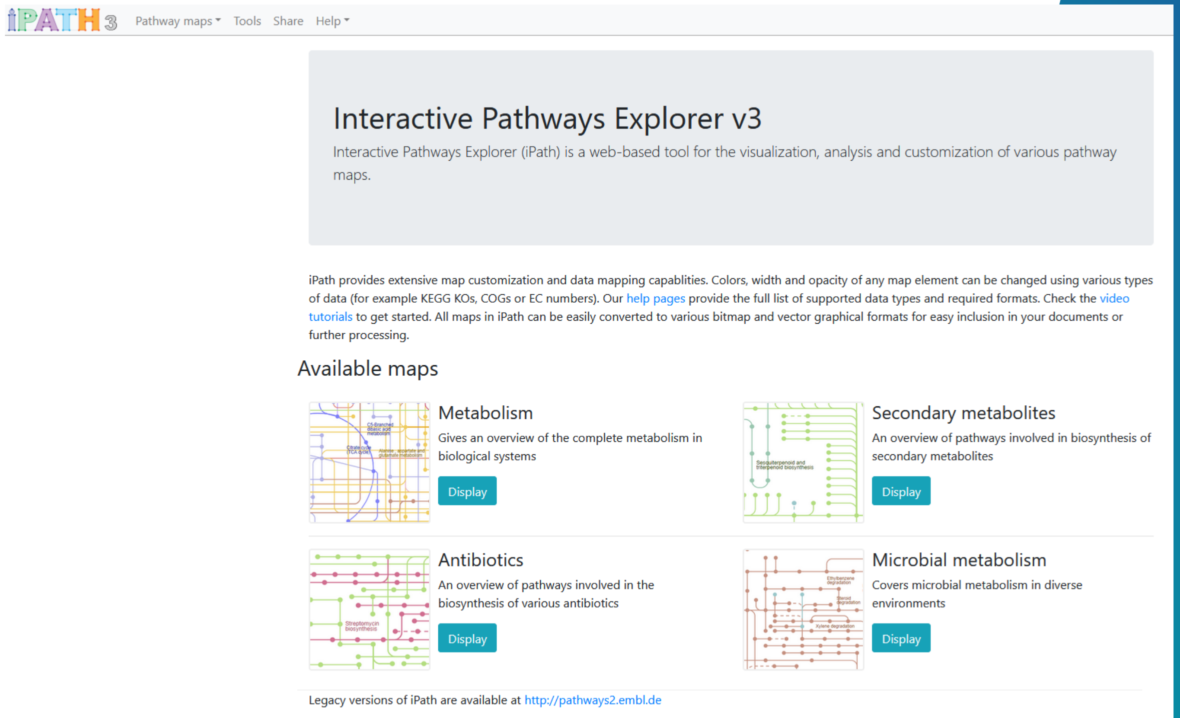
ipath_over.tsv



```
ID Color Width
EC1.11.1.1 #637978 W11
EC2.8.3.10 #637978 W11
EC2.4.2.52 #637978 W11
EC2.7.7.61 #637978 W11
EC6.2.1.22 #637978 W11
EC3.5.1.88 #B0C6C5 W11
EC6.3.4.21 #B0C6C5 W11
EC2.7.8.33 #B0C6C5 W11
EC2.3.1.118 #8EADAC W11
EC1.1.3.15 #8EADAC W11
EC3.2.1.22 #8EADAC W11
```

ipath_under.tsv

https://pathways.embl.de/



Interactive Pathways Explorer v3

Interactive Pathways Explorer (iPath) is a web-based tool for the visualization, analysis and customization of various pathway maps.

iPath provides extensive map customization and data mapping capablities. Colors, width and opacity of any map element can be changed using various types of data (for example KEGG KOs, COGs or EC numbers). Our help pages provide the full list of supported data types and required formats. Check the video tutorials to get started. All maps in iPath can be easily converted to various bitmap and vector graphical formats for easy inclusion in your documents or further processing.

Available maps

Metabolism
Gives an overview of the complete metabolism in biological systems

Secondary metabolites
An overview of pathways involved in biosynthesis of secondary metabolites

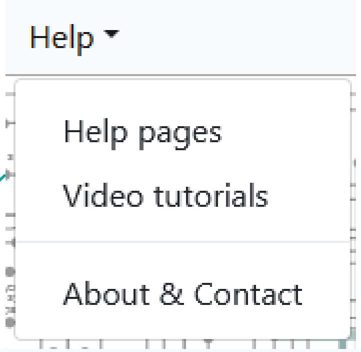
Antibiotics
An overview of pathways involved in the biosynthesis of various antibiotics

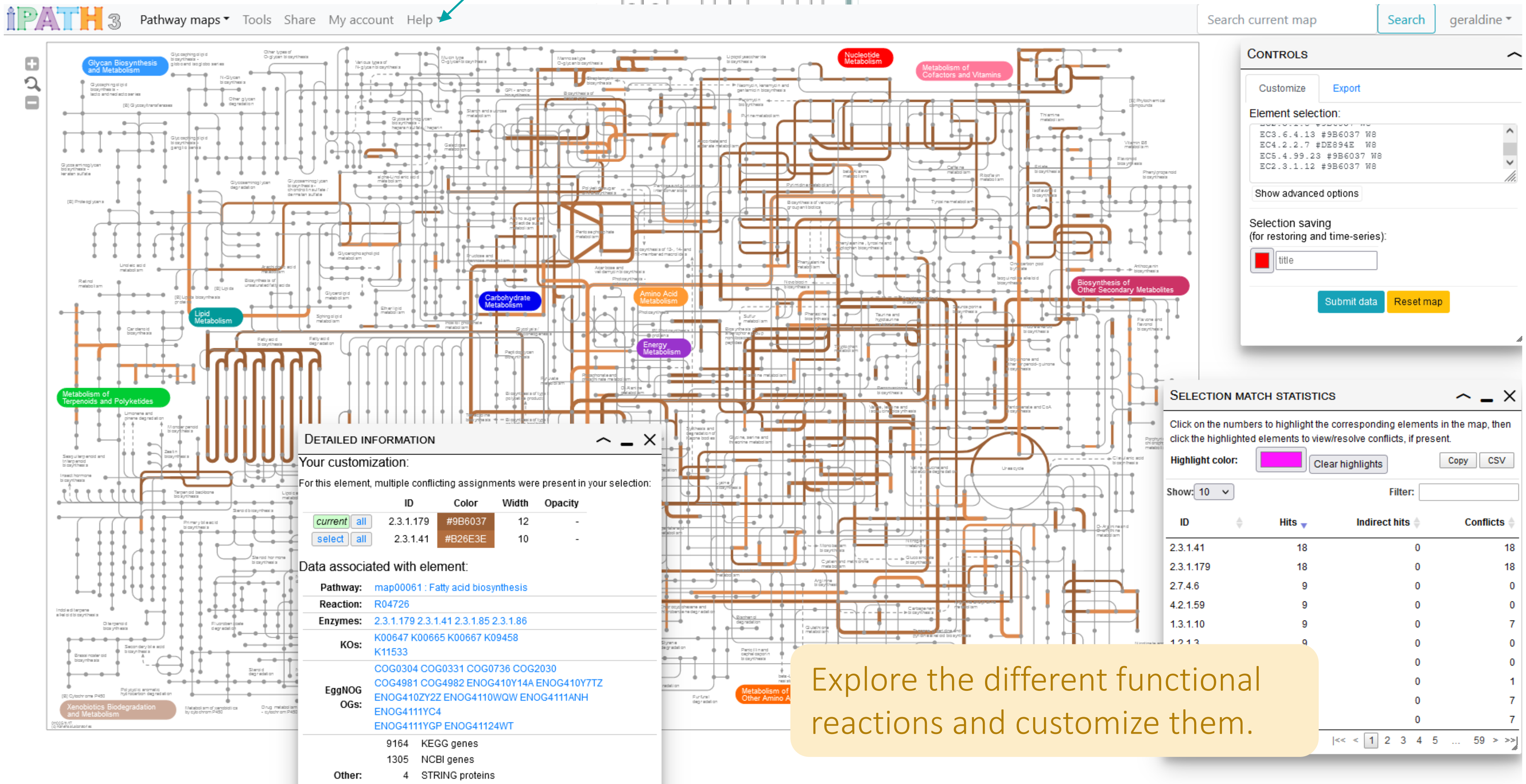
Microbial metabolism
Covers microbial metabolism in diverse environments

Legacy versions of iPath are available at http://pathways2.embl.de

Create a free account and paste “over” or “under” data

IPATH3





CONTROLS

Customize | Export

Element selection:

```
EC3.6.4.13 #9B6037 W8
EC4.2.2.7 #DE894E  W8
EC5.4.99.23 #9B6037 W8
EC2.3.1.12 #9B6037 W8
```

Show advanced options

Selection saving
(for restoring and time-series):

🟥 title

Submit data | Reset map

DETAILED INFORMATION

Your customization:

For this element, multiple conflicting assignments were present in your selection:

	ID	Color	Width	Opacity
current all	2.3.1.179	#9B6037	12	-
select all	2.3.1.41	#B26E3E	10	-

Data associated with element:

Pathway:	map00061 : Fatty acid biosynthesis	
Reaction:	R04726	
Enzymes:	2.3.1.179 2.3.1.41 2.3.1.85 2.3.1.86	
KOs:	K00647 K00665 K00667 K09458 K11533	
EggNOG OGs:	COG0304 COG0331 COG0736 COG2030 COG4981 COG4982 ENOG410Y14A ENOG410Y7TZ ENOG410ZY2Z ENOG4110WQW ENOG4111ANH ENOG4111YC4 ENOG4111YGP ENOG41124WT	
	9164	KEGG genes
	1305	NCBI genes
Other:	4	STRING proteins

SELECTION MATCH STATISTICS

Click on the numbers to highlight the corresponding elements in the map, then click the highlighted elements to view/resolve conflicts, if present.

Highlight color: 🟪 Clear highlights | Copy | CSV

Show: 10 | Filter:

ID	Hits	Indirect hits	Conflicts
2.3.1.41	18	0	18
2.3.1.179	18	0	18
2.7.4.6	9	0	0
4.2.1.59	9	0	0
1.3.1.10	9	0	7
1.2.1.3	9	0	0
		0	0
		0	1
		0	7
		0	7

|<< < 1 2 3 4 5 … 59 > >>|

Explore the different functional reactions and customize them.