



# Training on Galaxy: Metabarcoding

March 2024 - Webinar

## FROGS Practice on function inference

LUCAS AUER, MARIA BERNARD, LAURENT CAUQUIL, VINCENT DARBOT, MAHENDRA MARIADASSOU, GÉRALDINE PASCAL & OLIVIER RUÉ

---

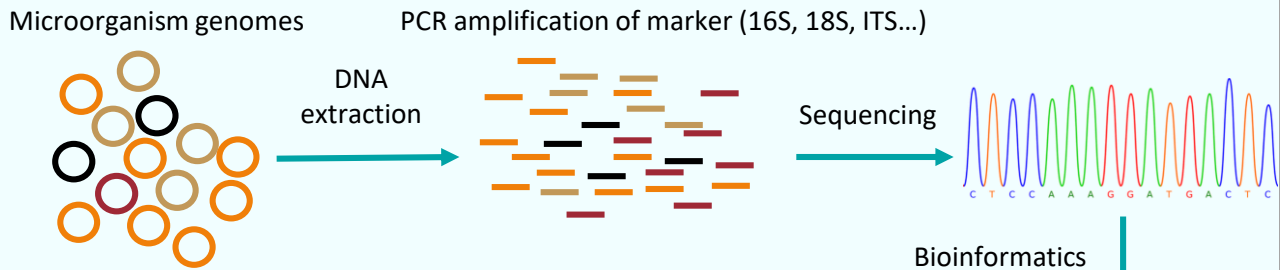
What metabolic functions are present in the environment?

---

# Concepts

## Metabarcoding principle

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



Abundancy table displaying microbial diversity per samples

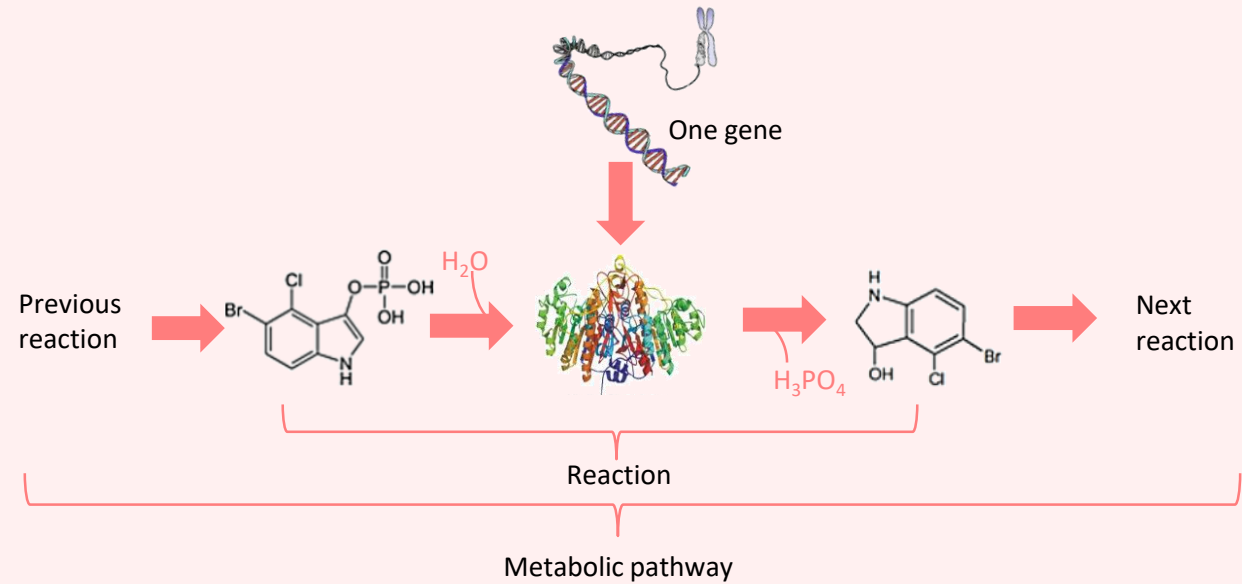
	Taxonomic affiliation	Sample1	Sample2	Sample3
ASV1	Species A	3500	6300	210
ASV2	Species B	0	460	36
ASV3	Species C	400	700	500

Bioinformatics analysis with FROGS



## 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 metabolic functions are presents in a microbial community ?

# Based on PICRUSt2

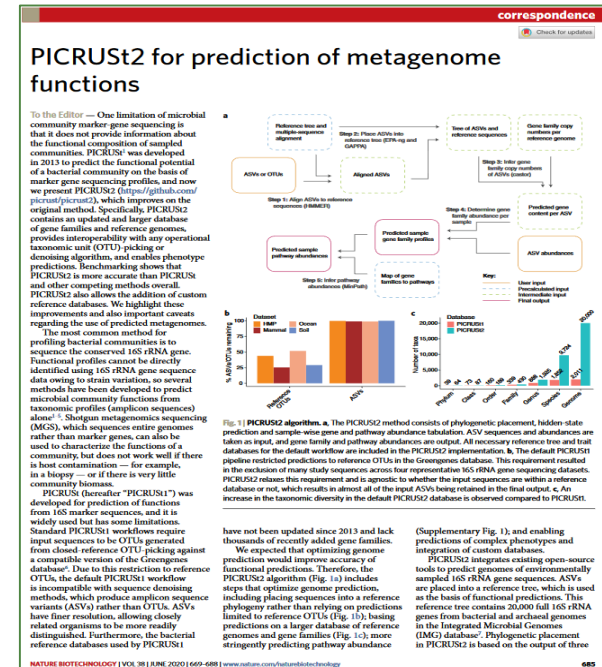
PICRUSt (Phylogenetic investigation of communities by reconstruction of unobserved states) is an open-source tool.

It is a software for predicting 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.

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 it works ?

---

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

FROGSFUNC\_1\_placeseqs\_copynumber

2. Predicts number of function copy number in each ASV and calculates functions abundances in each sample and ASV abundances according to marker copy number.

FROGSFUNC\_2\_functions

3. Calculates pathway abundances in each sample.

FROGSFUNC\_3\_pathways



It runs only on 16S, ITS or 18S

---

FROGSFUNC\_1\_placeseqs\_copynumber

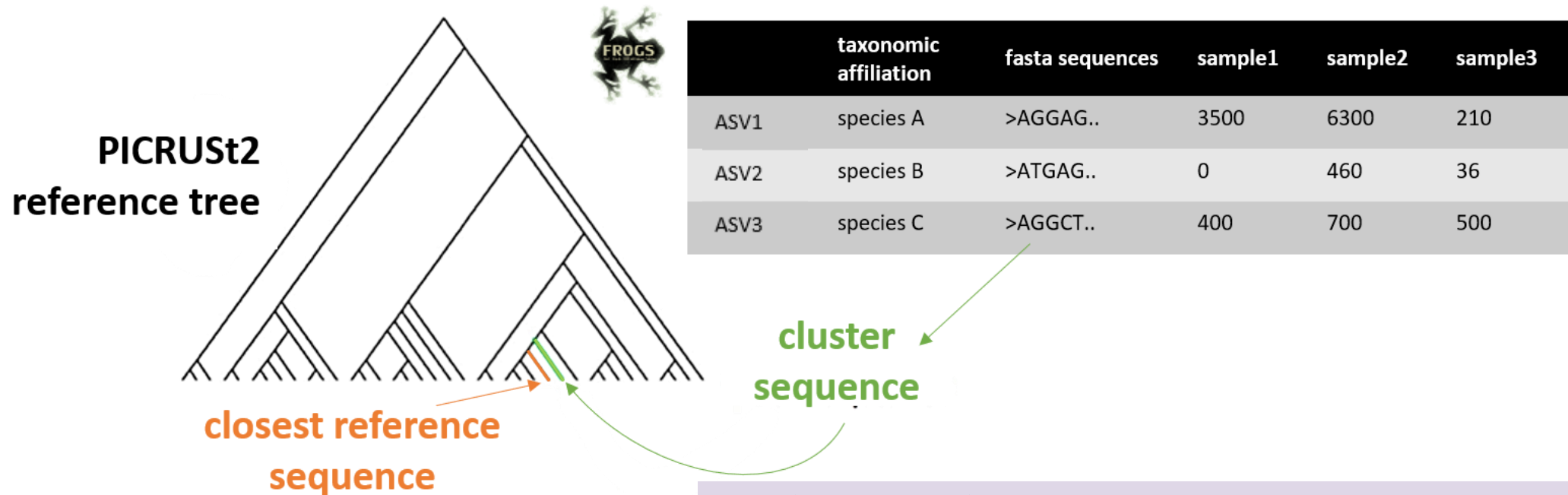
---

# FROGSFUNC\_1\_placeseqs\_copynumber

---

- **FROGSFUNC\_1\_placeseqs\_copynumber** 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](#).
- Prediction of the **copy numbers** of the marker gene (16S, ITS or 18S) in order to normalize the ASV abundances table thereafter.

# FROGSFUNC\_1\_placeseqs\_copynumber

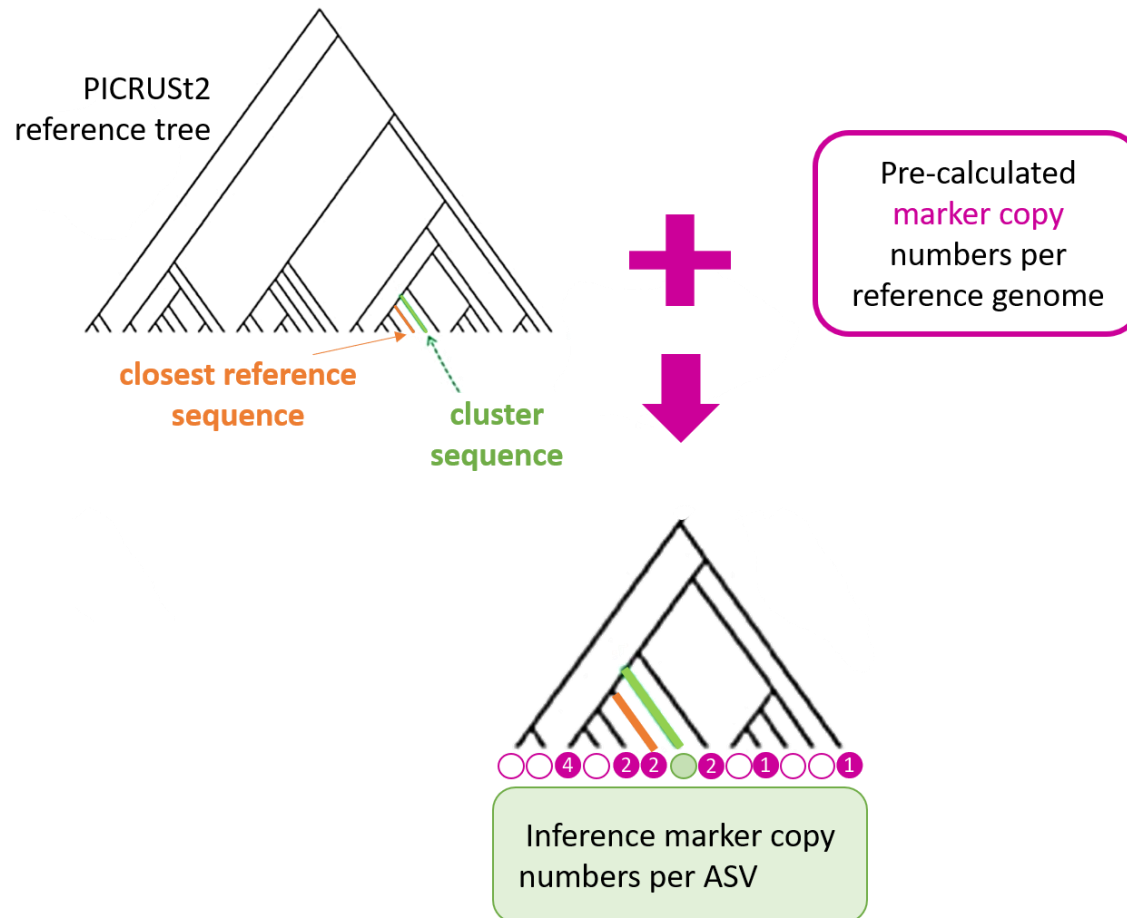


**2 input files** are required for FROGSFUNC\_1\_placeseqs\_copynumber analysis:

- **fasta file** of ASV sequences (it can be from FROGS\_4 cluster\_filters step)
- **biom file** of ASV abundances with taxonomic affiliation information (it can be from FROGS\_5\_taxonomic\_affiliation step)



# FROGSFUNC\_1\_placeseqs\_copynumber



# FROGSFUNC\_1\_placeseqs\_copynumber

**FROGSFUNC\_1\_placeseqs\_and\_copynumbers** Places ASVs into a reference phylogenetic tree. (Galaxy Version 4.1.0+galaxy1) ☆ Favorite 🔄 Versions ▼ Options

**Sequence file**  
15: FROGS\_4 Cluster filters: clusterFilters\_sequences.fasta

The sequence file to analyse (format: fasta). (--input-fasta)

**Biom file**  
25: FROGS\_5 Taxonomic affiliation: affiliation\_abundance.biom

The abundance file to analyse (format: biom). Taxonomic affiliations must be inside (FROGS taxonomic\_affiliation step). (--input-biom)

**Taxonomy marker**  
 16S  
 ITS  
 18S

Taxonomic marker of interest. (--ref-dir)

**Placement tool**  
 epa-ng  
 sepp

Placement tool for insertion of sequences into the reference tree. SEPP is a low-memory alternative to EPA-ng for placing sequences, and is only available for 16S analysis. (--placement-tool)

**Minimum alignment length**  
0.8

Proportion of the total length of an input sequence that must align with reference sequences. All others will be out. (default: 0.80) (--min-align)

**Email notification**  
 No

Send an email notification when the job completes.


epa-ng is the only choice for ITS and 18S  
Since epa-ng is greedy then it may force to reduce the dataset to run.

epa-ng is very memory and computing power intensive



# Input files

---

- **Sequences file:** The ASV fasta sequence file.
- **biom file:** The ASV [biom](#) file. Taxonomic affiliations must be done before (biom file form FROGS\_5\_taxonomic\_affiliation tool).
- **taxonomy marker:** 16S, ITS and 18S only available.  
 If your ASVs are based on another marker, you cannot use this tool.
- **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. **So, if the tool crashes with EPA-ng, try again with SEPP.**
- **minimum alignment length:** Proportion of the total length of an input sequence that must align with reference sequences. All other will be out.

# Output files

---

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_marker.tsv

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.biom

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_closests\_ref\_sequences.txt

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.fasta

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_excluded.tsv

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_tree.nwk



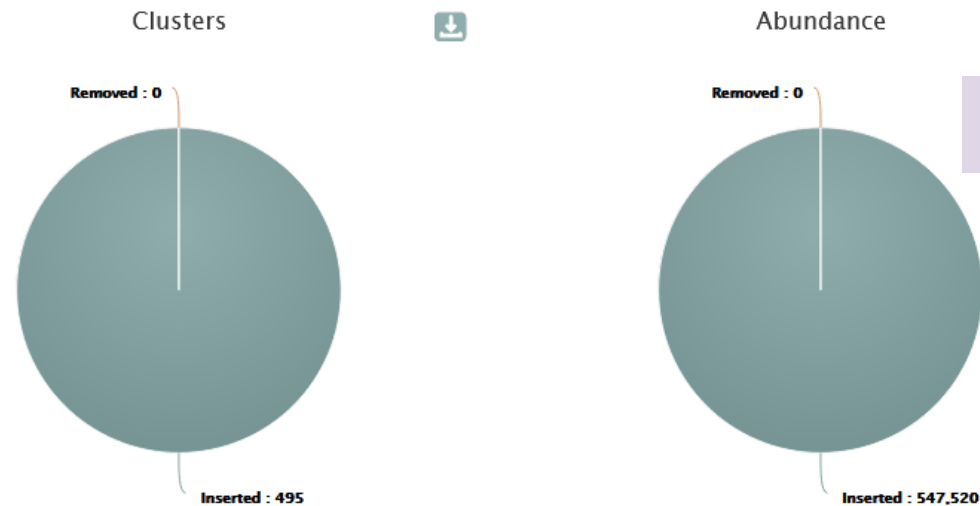
FROGSFUNC\_1\_placeseqs\_and\_copynumbers: report.html

# FROGSFUNC\_1: report.html

---

## Insertion in reference tree summary

---



Here, all our sequences have found a place inside the PICRUSt2 reference tree.

The html report file describes that ASVs are contained or not in the phylogenetic tree.

Note that PICRUSt2 uses its own reference tree to affiliate ASVs from reference sequences.

The report file indicates for each ASV which is the closest PICRUSt2 reference sequence, and compares it to the original FROGS taxonomy.

Clicking on the sequence ID gives you more information about it [JGI](#) database.

# Where are my ASVs inserted in the phylogenetic reference tree ?

<https://jgi.doe.gov/>

ASV	Nb sequences	FROGS Taxonomy	PICRUSt2 closest ID (JGI)	PICRUSt2 closest reference name
Cluster_1	84849	Bacteria;Firmicutes;Bacilli;Lactobacillales;Listeriaceae;Brochothrix;Brochothrix thermosphacta	2576861686	Brochothrix thermosphacta FSL F6-1036
Cluster_10	4188	Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Flavobacterium;Flavobacterium sp.	2724679776	Flavobacterium sp. 9
Cluster_100	696	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas;Stenotrophomonas sp.	2639762796	Stenotrophomonas maltophilia OC194
Cluster_101	752	Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Weeksellaceae;Chryseobacterium;Chryseobacterium sp.	2600255101	Chryseobacterium haifense DSM 19056

first part of the table

# Where are my ASVs inserted in the phylogenetic 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 <sup>††</sup>	PICRUSt2 closest taxonomy	<sup>††</sup> NSTI <sup>††</sup>	NSTI Confidence <sup>††</sup>	Lowest same taxonomic rank between FROGS and PICRUSt2 <sup>††</sup>	Comment <sup>††</sup>
Brochothrix thermosphacta FSL F6-1036	Bacteria;Firmicutes;Bacilli;Bacillales;Listeriaceae;Brochothrix;Brochothrix thermosphacta	0.062	Good	Up to Species	/
Flavobacterium sp. 9	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Flavobacteriaceae;Flavobacterium;Flavobacterium sp.	0.057	Good	Up to Species	/
Stenotrophomonas maltophilia OC194	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas;Stenotrophomonas maltophilia				
Chryseobacterium haifense DSM 19056	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Weeksellaceae;Kaistella;Chryseobacterium haifense	0.063	Good	Up to Family	/

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

second part of the table

# Where are my ASVs inserted in the phylogenetic reference tree ?

PICRUSt2 closest reference name <sup>†‡</sup>	PICRUSt2 closest taxonomy	<sup>†‡</sup> NSTI <sup>†‡</sup>	 NSTI Confidence <sup>†‡</sup>	Lowest same taxonomic rank between FROGS and PICRUSt2 <sup>†‡</sup>	Comment <sup>†‡</sup>
Brochothrix thermosphacta FSL F6-1036	Bacteria;Firmicutes;Bacilli;Bacillales;Listeriaceae;Brochothrix;Brochothrix thermosphacta	0.062	Good	Up to Species	/
Flavobacterium sp. 9	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Flavobacteriaceae;Flavobacterium;Flavobacterium sp.	0.057	Good	Up to Species	/
Stenotrophomonas maltophilia OC194	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas;Stenotrophomonas maltophilia	0.036	Good	Up to Genus	/
Chryseobacterium haifense DSM 19056	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Weeksellaceae;Kaistella;Chryseobacterium haifense	0.063	Good	Up to Family	/

second part of the table



## → 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:

- $0 < \text{Good} < 0.5$
- $0.5 \leq \text{Medium} < 1$
- $1 \leq \text{Bad} < 2$
- To exclude  $\geq 2$

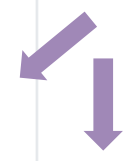
PICRUSt2 sets NSTI threshold to 2 per default. Some studies have shown that this threshold is permissive. Thus, it is important to see if the taxonomies between PICRUSt2 and FROGS are quite similar or not, in order to potentially choose a more stringent threshold afterwards.



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

# Where are my ASVs inserted in the phylogenetic reference tree ?

PICRUSt2 closest reference name <sup>†↓</sup>	PICRUSt2 closest taxonomy	<sup>†↓</sup> NSTI <sup>†↓</sup>	NSTI Confidence <sup>†↓</sup>	Lowest same taxonomic rank between FROGS and PICRUSt2 <sup>†↓</sup>	Comment <sup>†↓</sup>
Brochothrix thermosphacta FSL F6-1036	Bacteria;Firmicutes;Bacilli;Bacillales;Listeriaceae;Brochothrix;Brochothrix thermosphacta	0.062	Good	Up to Species	/
Flavobacterium sp. 9	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Flavobacteriaceae;Flavobacterium;Flavobacterium sp.	0.057	Good	Up to Species	/
Stenotrophomonas maltophilia OC194	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas;Stenotrophomonas maltophilia	0.036	Good	Up to Genus	/
Chryseobacterium haifense DSM 19056	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Weeksellaceae;Kaistella;Chryseobacterium haifense	0.063	Good	Up to Family	/



second part of the table

# Where are my ASVs inserted in the phylogenetic reference tree ?

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

## Comment :

- *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

PICRUSt2 closest reference name	PICRUSt2 closest taxonomy	NSTI	NSTI Confidence	Lowest same taxonomic rank between FROGS and PICRUSt2	Comment
Brochothrix thermosphacta FSL F6-1036	Bacteria;Firmicutes;Bacilli;Bacillales;Listeriaceae;Brochothrix;Brochothrix thermosphacta	0.062	Good	Up to Species	/
Flavobacterium sp. 9	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Flavobacteriaceae;Flavobacterium;Flavobacterium sp.	0.057	Good	Up to Species	/
Stenotrophomonas maltophilia OC194	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas;Stenotrophomonas maltophilia	0.036	Good	Up to Genus	/
Chryseobacterium haifense DSM 19056	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Weeksellaceae;Kaistella;Chryseobacterium haifense	0.063	Good	Up to Family	/

Search « Up to Species » for obtaining less ambiguous reference



# Where are my ASVs inserted in the phylogenetic reference tree ?

PICRUSt2 closest reference name	PICRUSt2 closest taxonomy	NSTI	NSTI Confidence	Lowest same taxonomic rank between FROGS and PICRUSt2	Comment
Brochothrix thermosphacta FSL F6-1036	Bacteria;Firmicutes;Bacilli;Bacillales;Listeriaceae;Brochothrix;Brochothrix thermosphacta	0.062	Good	Up to Species	/
Flavobacterium sp. 9	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Flavobacteriaceae;Flavobacterium;Flavobacterium sp.	0.057	Good	Up to Species	/
Stenotrophomonas maltophilia OC194	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas;Stenotrophomonas maltophilia	0.036	Good	Up to Genus	/
Chryseobacterium haifense DSM 19056	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Weeksellaceae;Kaistella;Chryseobacterium haifense	0.063	Good	Up to Family	/

PICRUSt2 reference tree is base on NCBI taxonomy. If you want more « Up to Sepices », i.e. more correspondances between FROGS affiliation and PICRUSt2 affiliation , think to use 16S REFseq databank in FROGS\_5 Taxonomic affiliation step

For this exemple, affiliation with 16S SILVA 138.1 gives 73 « Up to Species » and with 16S REFseq 20230726 gives 146 « Up to Species »

**FROGS\_5 Taxonomic affiliation** Taxonomic affiliation of each ASV's seed by RDPtools and BLAST

Using reference database 

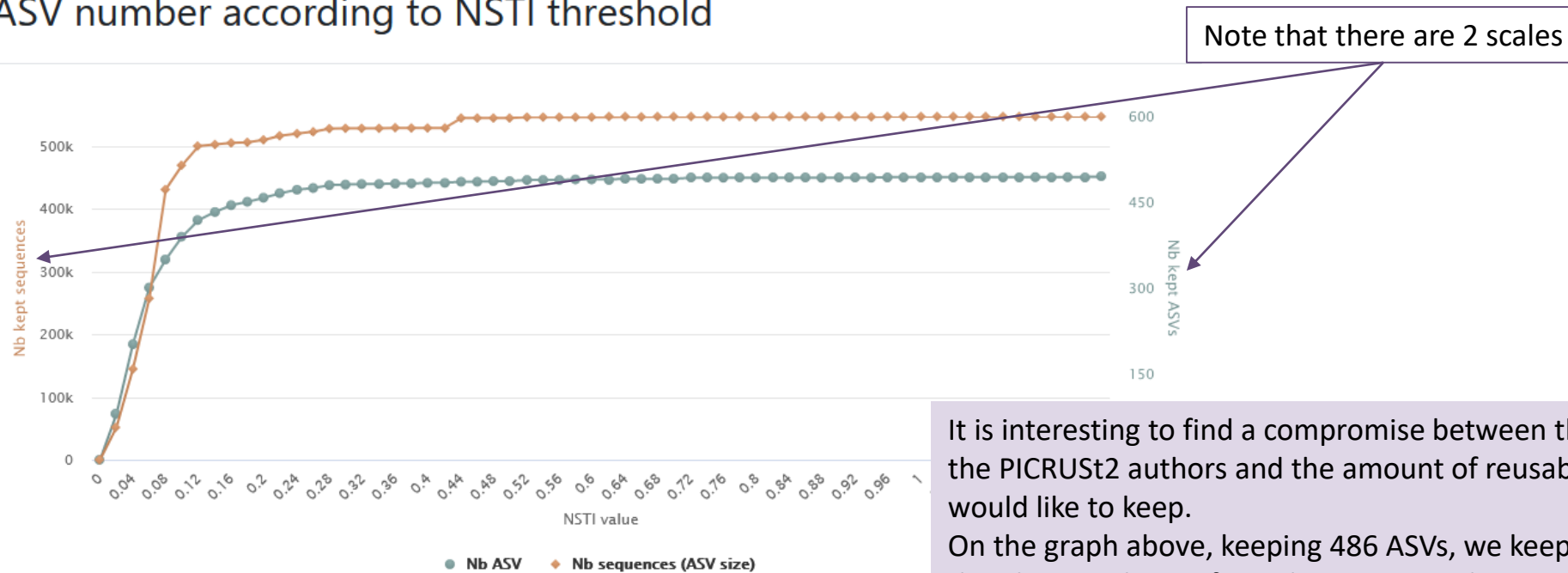
16S REFseq Bacteria 20230726

Select reference from the list

... - - - - -

# Another key for choosing NSTI threshold

## ASV number according to NSTI threshold



N.B.: Select area to zoom in.

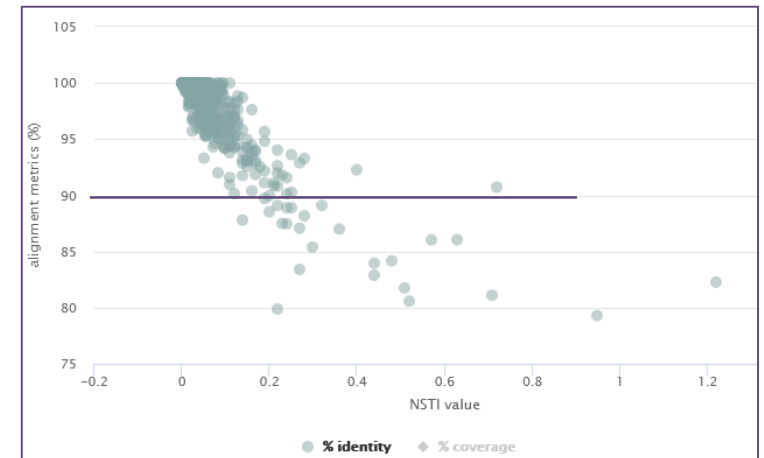
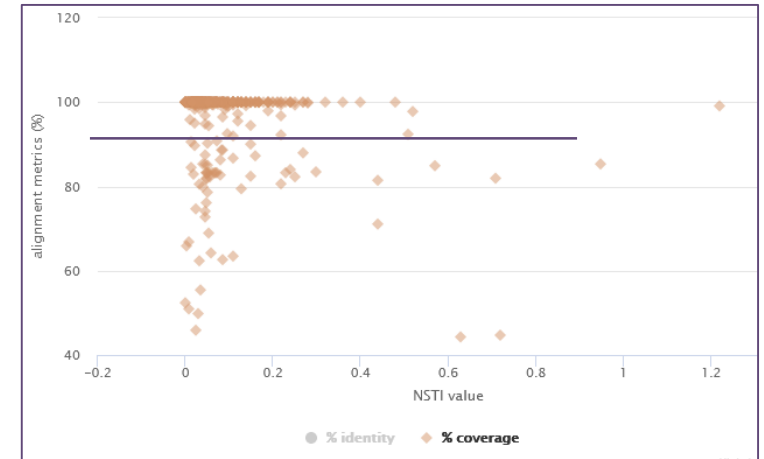
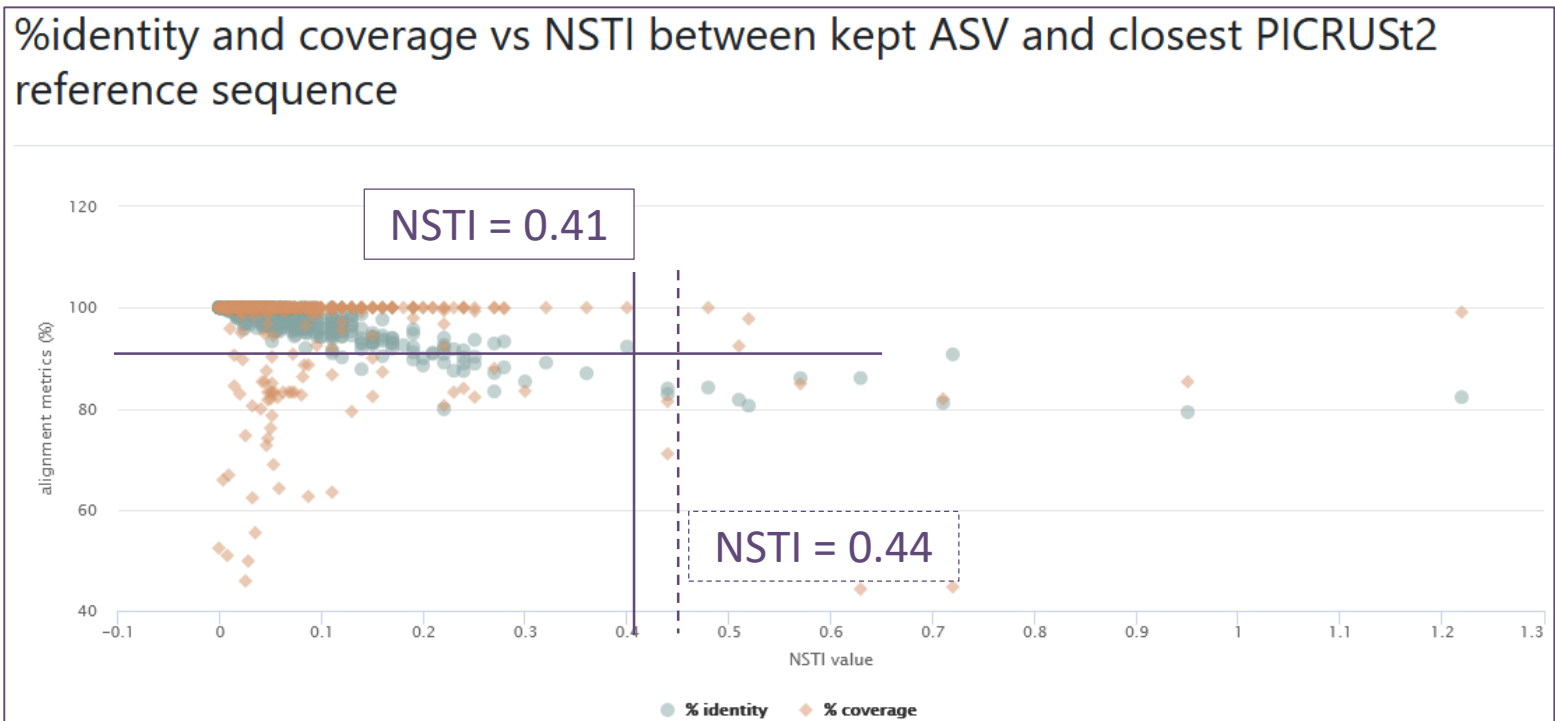
It is interesting to find a compromise between the guidance provided by the PICRUSt2 authors and the amount of reusable information that you would like to keep. On the graph above, keeping 486 ASVs, we keep the information before the plateau, that is, from this point on, the more ASVs we keep the more we degrade the accuracy. So, here NSTI = 0.44 But this depends strongly on the datasets.



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

# Another key for choosing NSTI threshold

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



With %id and %cov cutoff at 90%, we can choose a NSTI cutoff at 0.41 or 0.44, we will select same data. So we can choose 0.41.

# Output files

---

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_marker.tsv

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.biom

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_closests\_ref\_sequences.txt

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.fasta

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_excluded.tsv



FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_tree.nwk

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: report.html

# FROGSFUNC\_1: tree.nwk

---

```
``
(2609460310:0.0708,2713896746:0.079553):0.020861):0.018755):0.049721,2667528167:0.021242):0.05
(2634166173:0.020087,((2630968917:0.042097,(2695420983:0.0414,((641380439:0.006668,(2751185746
(2630968293:0.07529,(2634166307:0.010503,2545824761:0.03788):0.019526):0.024717):0.015629):0.0
((2630968881:0.054077,2654587584:0.013434):0.046038,2585427602:0.014665):0.024954):0.022354):0
((2526164557:0.215838,((2524023070:0.051755,(2627853707:0.116778,2509276063:0.03521):0.033363)
((2627853601:0.074019,
((640963037:0.107024,2675903215:0.024681):0.038088,2728369219:0.027685):0.020008):0.023448,274
0.021443):0.036439,(((2600255390:0.00005,Cluster 103:0.022437):0.088344,(2617271337:0.049512,
(2585427837:0.000001,2693429891:0.001563):0.000001):0.00005,Cluster 29:0.031725):0.013777):0.0
((((2700988729:0.065717,2695420957:0.031748):0.022352,(2698536752:0.009476,((((2695421021:0.
((2643221790:0.0157,2728368985:0.008527):0.014802,(2695420960:0.010196,Cluster 188:0.040768):0.0
(2738541279-cluster:0.000001,((2636415472:0.00924,(2648501293:0.00005,Cluster 202:0.029009):0.
((((((2734482184:0.025815,2648501732:0.010421):0.005141,
(2513020052:0.021475,2695420959:0.057031):0.010852):0.005258,2754412712:0.020212):0.011771,265
((2700988712:0.032914,2695420925-cluster:0.003371):0.007089,((2619618807:0.048997,(2523533607
(2693429909:0.000001,Cluster 192:0.14387):0):0.017851,((2684622654:0.032648,(2551306066:0.0026
(2693429883:0.000001,Cluster 321:0.05075):0):0.019173):0.006793):0.012944):0.039508):0.029699,
(2523533550:0.002444,Cluster 37:0.02644):0.010937):0.020635):0.038474):0.000001,2643221667:0.0
((((2739367857-cluster:0.001694,2585427836:0.026929):0.008262,((((2619618994:0.089002,((25246
```



# Output files

---

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_marker.tsv

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.biom

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_closests\_ref\_sequences.txt

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.fasta

➔ FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_excluded.tsv

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_tree.nwk

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: report.html

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

# Output files

---

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_marker.tsv  
FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.biom  
FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_closests\_ref\_sequences.txt  
→ FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.fasta  
FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_excluded.tsv  
FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_tree.nwk  
FROGSFUNC\_1\_placeseqs\_and\_copynumbers: report.html

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

# Output files

---



FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_marker.tsv
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs.biom
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs_closests_ref_sequences.txt
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs.fasta
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs_excluded.tsv
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs_tree.nwk
FROGSFUNC_1_placeseqs_and_copynumbers: report.html

In this table, you will find all the details of the correspondences between your sequences and those of PICRUSt2.

# FROGSFUNC\_1: closest\_ref\_sequences.txt

#Cluster	Nb sequences	FROGS Taxonomy	PICRUSt2 closest ID
Cluster_1	84849	Bacteria;Firmicutes;Bacilli;Lactobacillales;Listeriaceae;Brochothrix;Brochothrix thermosphacta	2576861686
Cluster_2	31333	Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Vibrionaceae;Photobacterium;unknown species	2724679053
Cluster_3	40711	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Latilactobacillus;Lactobacillus sakei	2728369693
Cluster_4	22275	Bacteria;Actinobacteriota;Actinobacteria;Propionibacteriales;Propionibacteriaceae;Cutibacterium;unknown species	2537562124
Cluster_5	29355	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Leuconostoc;Leuconostoc inhae KCTC 3774	641522636

PICRUSt2 closest reference name	PICRUSt2 closest taxonomy
Brochothrix thermosphacta FSL F6-1036	Bacteria;Firmicutes;Bacilli;Bacillales;Listeriaceae;Brochothrix;Brochothrix thermosphacta
Photobacterium kishitanii 201212X	Bacteria;Proteobacteria;Gammaproteobacteria;Vibrionales;Vibrionaceae;Photobacterium;Photobacterium kishitanii
Lactobacillus curvatus JCM 1096, DSM 20019	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Latilactobacillus;Latilactobacillus curvatus
Cutibacterium acnes SK182	Bacteria;Actinobacteria;Actinomycetia;Propionibacteriales;Propionibacteriaceae;Cutibacterium;Cutibacterium acnes
Leuconostoc citreum KM20	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Leuconostoc;Leuconostoc citreum

NSTI	NSTI Confidence	FROGS and PICRUSt2 lowest same taxonomic rank	Comment	Cluster sequence	PICRUSt2 closest reference sequence	%id	%cov	score
0.062	Good	Up to Species	/	GACGAACGCTGGCGGCGTC	TTAACGAGAGTTTGATCCTGGCTCAGGACGAACGCTG	100.0	100.0	493.5
0.05	Good	Up to Genus	/	ATTGAACGCTGGCGGCAGG	GAGTAATGCCTGGGAATATACCCTGATGTGGGGGATAA	99.75	81.91	400.0
0.068	Good	Up to Genus	/	GACGAACGCTGGCGGCGTC	TTTTAATCGAGAGTTTGATCCTGGCTCAGGACGAACG	98.84	83.27	412.0
0.039	Good	Up to Genus	identical sequence	GACGAACGCTGGCGGCGTC	TTCCATTGGAGAGTTTGATCCTGGCTCAGGACGAACG	100.0	100.0	468.0
0.072	Good	Up to Genus	/	GATGAACGCTGGCGGCGTC	GAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGG	97.17	100.0	444.5

# Output files

---



FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_marker.tsv
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs.biom
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs_closests_ref_sequences.txt
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs.fasta
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs_excluded.tsv
FROGSFUNC_1_placeseqs_and_copynumbers: frogsfunc_placeseqs_tree.nwk
FROGSFUNC_1_placeseqs_and_copynumbers: report.html

Abundance table without those that may be excluded

# Output files

---



FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_marker.tsv

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.biom

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_closests\_ref\_sequences.txt

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.fasta

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_excluded.tsv

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_tree.nwk

FROGSFUNC\_1\_placeseqs\_and\_copynumbers: report.html

# FROGSFUNC\_1: marker.tsv

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

sequence	16S_rRNA_Count	metadata_NSTI
Cluster_1	1	0.062403
Cluster_10	1	0.057594000000000006
Cluster_100	1	0.036314
Cluster_101	1	0.063955
Cluster_102	1	0.019805
Cluster_103	1	0.022487
Cluster_104	1	0.074696
Cluster_105	7	0.0042
Cluster_106	1	0.048352
Cluster_107	1	0.295495
Cluster_108	1	0.018517
Cluster_109	1	0.017555
Cluster_11	1	0.017435
Cluster_110	1	0.068808000000000001
Cluster_111	1	0.031835
Cluster_112	1	0.110719000000000001
Cluster_113	1	0.009720999999999999
Cluster_114	4	0.043952
Cluster_115	1	0.105228
Cluster_116	1	0.238852
Cluster_117	1	0.006305
Cluster_118	1	0.035066
Cluster_119	1	0.053155999999999995
Cluster_12	1	0.060149
Cluster_120	1	0.194218
Cluster_121	1	0.111322
Cluster_122	1	0.031979
Cluster_123	2	0.037387000000000004
Cluster_124	1	0.26648499999999997
Cluster_125	2	0.091262
Cluster_126	1	0.051895000000000004
Cluster_127	6	0.019323
Cluster_128	4	0.022711000000000002

---

# FROGSFUNC\_2\_functions

---



# How it works ?

---

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

FROGSFUNC\_1\_placeseqs\_copynumber

2. Predicts number of function copy number in each ASV and calculates functions abundances in each sample and ASV abundances according to marker copy number.

FROGSFUNC\_2\_functions

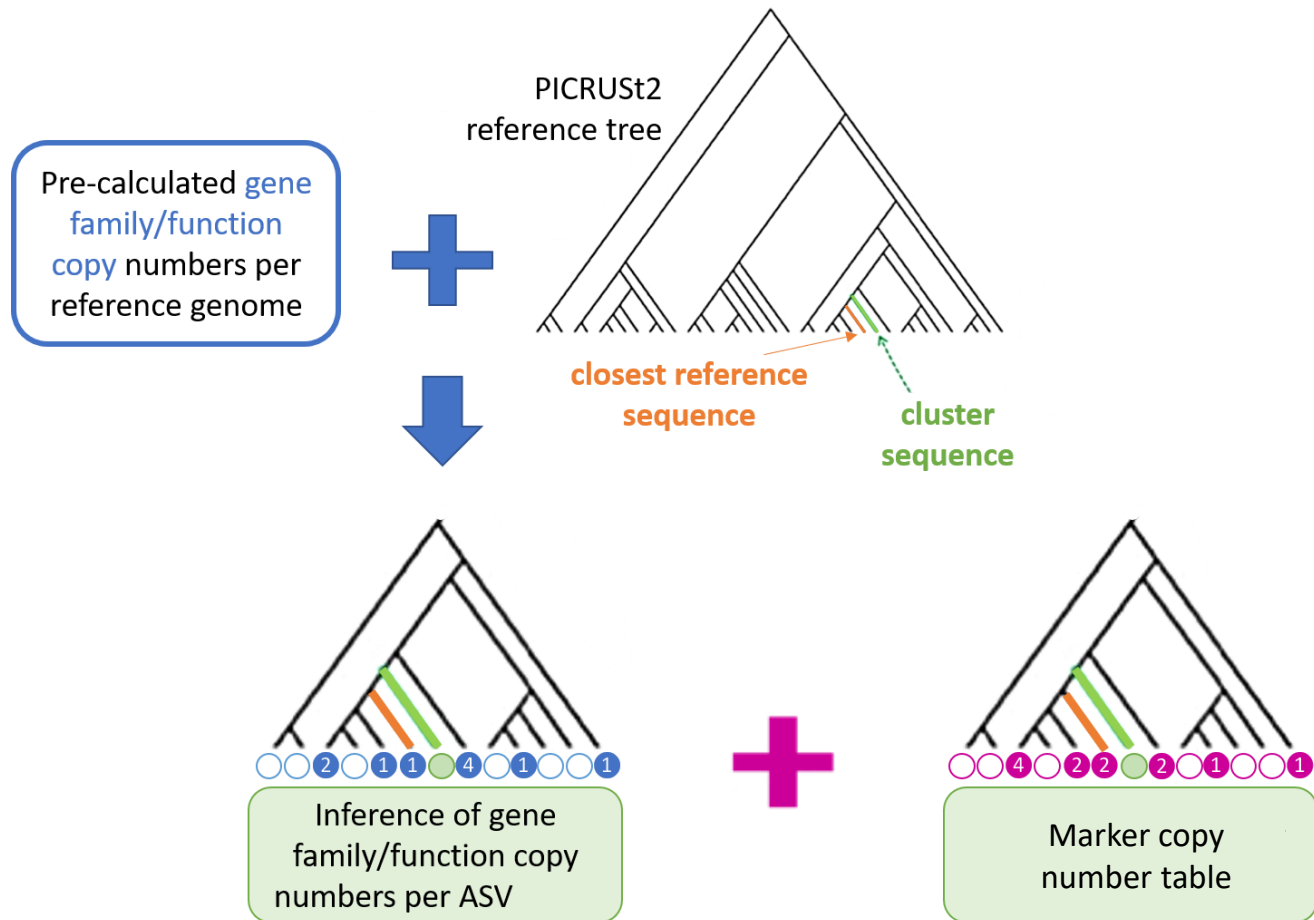
3. Calculates pathway abundances in each sample.

FROGSFUNC\_3\_pathways





It runs only on 16S, ITS or 18S

# FROGSFUNC\_2\_functions



# FROGSFUNC\_2\_functions

FROGSFUNC_1 abundance ASV						FROGSFUNC_1 marker.tsv 				FROGSFUNC_2 marker_norm.tsv			
	S1	S2	S3	16S copy number			S1	S2	S3		S1	S2	S3
ASV1	3500	6300	210	7		ASV1	3500/7	6300/7	210/7				
ASV2	0	460	36	4		ASV2	0/4	460/4	36/4				
ASV3	400	700	500	1		ASV3	400/1	700/1	500/1				

FROGSFUNC_2 marker_norm.tsv						FROGSFUNC_2 EC_copynumber_predicted.tsv 				FROGSFUNC_2 functions_unstrat_EC.tsv			
	S1	S2	S3	EC:1.1.1.1		EC:1.1.1.2		EC:1.1.1.3			S1	S2	S3
ASV1	500	900	30	2	0	2				EC:1.1.1.1	1800	5460	1129
ASV2	0	115	9	1	0	0				EC:1.1.1.2	1600	2800	2000
ASV3	400	700	500	2	4	2				EC:1.1.1.3	1800	3200	1060

# FROGSFUNC\_2\_functions

**FROGSFUNC\_2\_functions** Calculates functions abundances in each sample. (Galaxy Version 4.1.0+galaxy1)

☆ Favorite

🔄 Versions

▼ Options

## Biom file

51: FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.biom

The abundance file i.e. FROGSFUNC\_1\_placeseqs\_copynumber tool output file (frogsfunc\_placeseqs.biom). (--input-biom)

## Sequence file

49: FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs.fasta

The fasta file i.e. from FROGSFUNC\_1\_placeseqs\_copynumber tool output file (frogsfunc\_placeseqs.fasta). (--input-fasta)

## Tree file

47: FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_placeseqs\_tree.nwk

The file contains the tree information from FROGSFUNC\_1\_placeseqs\_copynumber tool (frogsfunc\_placeseqs\_tree.nwk). (--input-tree)

## Marker file

52: FROGSFUNC\_1\_placeseqs\_and\_copynumbers: frogsfunc\_marker.tsv

Table of predicted marker copy number i.e. FROGSFUNC\_1\_placeseqs\_copynumber output (frogsfunc\_marker.tsv). (--input-marker)

## Taxonomic marker

- 16S
- ITS
- 18S

Taxonomic marker of interest.

Thanks to the previous prediction of the **copy numbers of the marker gene** (16S, ITS or 18S) in FROGSFUNC\_1, FROGSFUNC\_2 can normalize the ASV abundances table.

# FROGSFUNC\_2\_functions

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

Select/Unselect all

EC

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 optional. ITS and 18S : 'EC' only available. (--functions)

## NSTI cut-off

0.41

Any sequence with an NSTI above this threshold will be out. (default: 2) (--max-nsti)

## Identity alignment cut-off

0.9

Percentage identity of the alignment between the input sequence and the PICRUSt2 reference sequence. Below this threshold, all sequences will be discarded. (default: None) (--min-blast-ident)

## Coverage alignment cut-off

0.9

Coverage identity of the alignment between the input sequence and the PICRUSt2 reference sequence. Below this threshold, all sequences will be discarded. (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 phylogenetic independent contrast (pic), continuous traits reconstruction using squared-change parsimony (scp) (default: mp). (--hsp-method)

# FROGSFUNC\_2\_functions



## 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.

### Target function database

Select/Unselect all

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 optional. ITS and 18S : 'EC' only available. (--functions)

### NSTI cut-off

Any sequence with an NSTI above this threshold will be out. (default: 2) (--max-nsti)

### Identity alignment cut-off

Percentage identity of the alignment between the input sequence and the PICRUSt2 reference sequence. Below this threshold, all sequences will be discarded. (default: None) (--min-blast-ident)

### Coverage alignment cut-off

Coverage identity of the alignment between the input sequence and the PICRUSt2 reference sequence. Below this threshold, all sequences will be discarded. (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 phylogenetic independent contrast (pic), continuous traits reconstruction using squared-change parsimony (scp) (default: mp). (--hsp-method)

# FROGSFUNC\_2\_functions

## Target function database

Select/Unselect all

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

Any sequence with an NSTI above this threshold will be out. (default: 2) (--max-nsti)

## Identity alignment cut-off

Percentage identity of the alignment between the input sequence and the PICRUSt2 reference sequence. Below this threshold, all sequences will be discarded. (default: None) (--min-blast-ident)

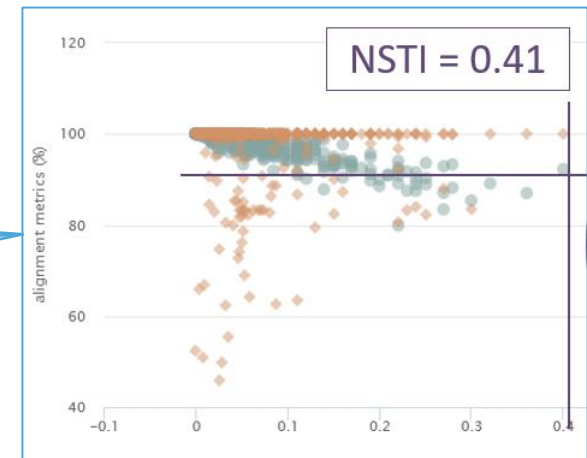
## Coverage alignment cut-off

Coverage identity of the alignment between the input sequence and the PICRUSt2 reference sequence. Below this threshold, all sequences will be discarded. (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 phylogenetic independent contrast (pic), continuous traits reconstruction using squared-change parsimony (scp) (default: mp). (--hsp-method)



# FROGSFUNC\_2\_functions

## Target function database

Select/Unselect all

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

Any sequence with an NSTI above this threshold will be out. (default: 2) (--max-nsti)

## Identity alignment cut-off

Percentage identity of the alignment between the input sequence and the PICRUSt2 reference sequence. Below this threshold, all sequences will be discarded. (default: None) (--min-blast-ident)

## Coverage alignment cut-off

Coverage identity of the alignment between the input sequence and the PICRUSt2 reference sequence. Below this threshold, all sequences will be discarded. (default: None) (--min-blast-cov)

## HSP method

mp  
 emp\_prob  
 pic  
 scp  
 subtree\_average

Method used  
for prediction.

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 phylogenetic independent contrast (pic), continuous traits reconstruction using squared-change parsimony (scp) (default: mp). (--hsp-method)



# Outputs

---

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_unstrat\_EC.tsv**

**FROGSFUNC\_2\_functions: EC\_copynumbers\_predicted.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_excluded.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_weighted\_nsti.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_marker\_norm.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.fasta**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.biom**



**FROGSFUNC\_2\_functions: report.html**

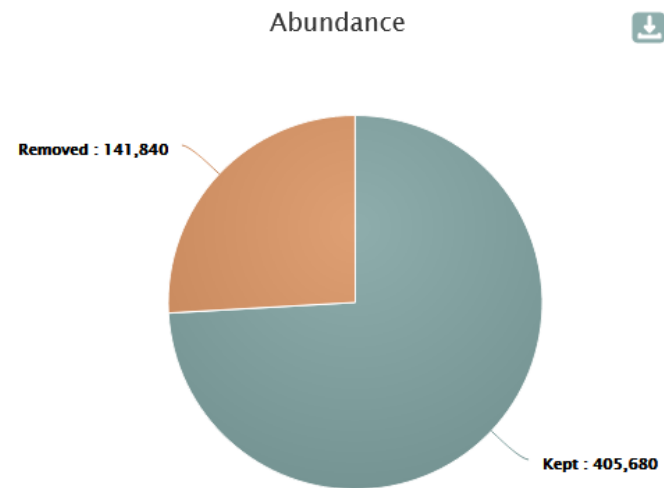
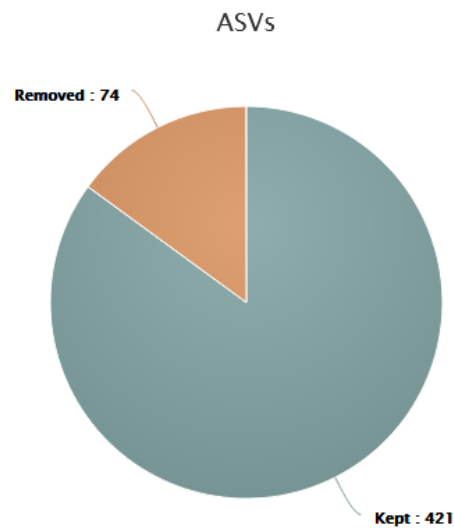
# FROGSFUNC\_2 : report.html

---

ASVs are excluded if the associated NSTI is above the threshold, or if the alignment values are below the thresholds.

## Metagenome functional profile summary

---



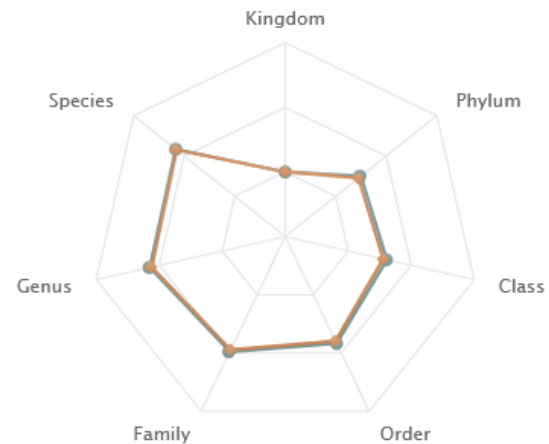
# FROGSFUNC\_2 : report.html

---

Number of different taxonomic ranks before (green) and after (orange) application of the filters.

Remaining diversity after filtering for functional inference

---



*Number of different taxonomic observations per rank before and after applying NSTI or alignments thresholds*

- Before
- ◆ After

# FROGSFUNC\_2 : report.html

Gene families/function from KEGG or Metacyc databases are classified according to 4 hierarchy levels. The graph shows the proportion of each level within the selected samples.

Function abundances per sample

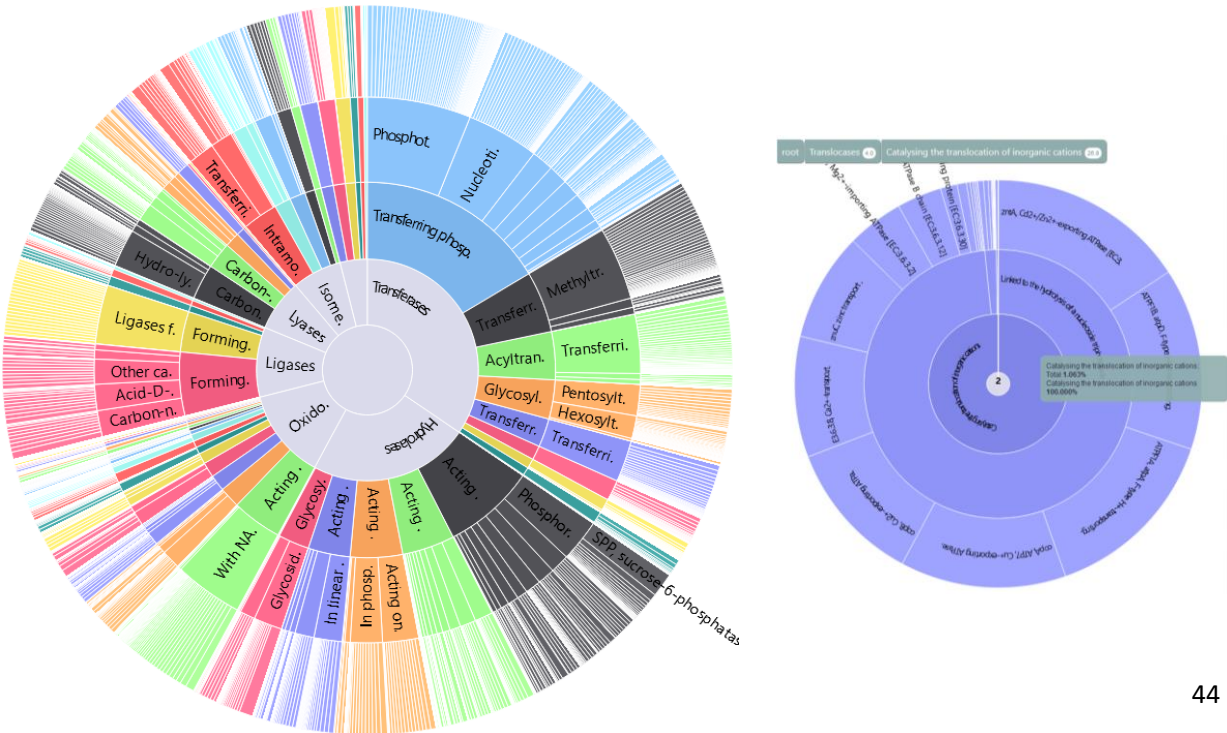
Display global distribution

[↓ CSV](#)

Show  entries



<input type="checkbox"/>	Samples	↑↓ Weighted NSTI	↑↓ Nb function_id retrieved
<input checked="" type="checkbox"/>	BHT0.LOT01	0.081	2,627
<input checked="" type="checkbox"/>	BHT0.LOT03	0.075	2,701
<input checked="" type="checkbox"/>	BHT0.LOT04	0.066	2,909
<input checked="" type="checkbox"/>	BHT0.LOT05	0.081	2,836
<input checked="" type="checkbox"/>	BHT0.LOT06	0.085	2,815
<input checked="" type="checkbox"/>	BHT0.LOT07	0.082	2,787
<input checked="" type="checkbox"/>	BHT0.LOT08	0.082	2,895
<input type="checkbox"/>	BHT0.LOT10	0.074	2,764
<input type="checkbox"/>	CDT0.LOT02	0.073	2,553
<input type="checkbox"/>	CDT0.LOT04	0.053	2,693

Display distribution



# Outputs

---

	<b>FROGSFUNC_2_functions: frogsfunc_functions_unstrat_EC.tsv</b>
	<b>FROGSFUNC_2_functions: EC_copynumbers_predicted.tsv</b>
	<b>FROGSFUNC_2_functions: frogsfunc_functions_excluded.tsv</b>
	<b>FROGSFUNC_2_functions: frogsfunc_functions_weighted_nsti.tsv</b>
	<b>FROGSFUNC_2_functions: frogsfunc_functions_marker_norm.tsv</b>
	<b>FROGSFUNC_2_functions: frogsfunc_functions.fasta</b>
	<b>FROGSFUNC_2_functions: frogsfunc_functions.biom</b>
	<b>FROGSFUNC_2_functions: report.html</b>

ASV abundance table without excluded ASVs (NSTI, %identity or %coverage thresholds alignment).

ASV Sequence file without excluded ASVs (NSTI, blast %identity or blast %coverage thresholds).

# Outputs

---

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_unstrat\_EC.tsv**

**FROGSFUNC\_2\_functions: EC\_copynumbers\_predicted.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_excluded.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_weighted\_nsti.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_marker\_norm.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.fasta**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.biom**

**FROGSFUNC\_2\_functions: report.html**



ASV normalized abundance table

# FROGSFUNC\_2: marker\_norm.tsv

Table with normalized abundances per marker copy number from FROGSFUNC\_1 step.

Ex: cluster\_123 have two 16S copies and its abundance before this normalization in BHT0.LOT01 was 35

Cluster_124	461	0	6	6	0
Cluster_123	640	35	6	55	11
Cluster_126	511	73	7	10	33

normalized	BHT0.LOT01	BHT0.LOT03	BHT0.LOT04
Cluster_1	791.0	402.0	433.0
Cluster_10	0.0	0.0	0.0
Cluster_100	0.0	0.0	0.0
Cluster_101	0.0	0.0	0.0
Cluster_102	0.0	0.0	0.0
Cluster_103	0.0	0.0	0.0
Cluster_104	0.0	0.0	0.0
Cluster_106	0.0	0.0	0.0
Cluster_108	0.0	0.0	0.0
Cluster_109	0.0	0.0	0.0
Cluster_11	3.0	0.0	4.0
Cluster_110	0.0	0.0	0.0
Cluster_111	2.0	8.0	59.0
Cluster_112	0.0	0.0	0.0
Cluster_113	0.0	4.0	5.0
Cluster_114	0.0	0.0	0.5
Cluster_115	0.0	6.0	7.0
Cluster_117	0.0	0.0	0.0
Cluster_118	0.0	46.0	6.0
Cluster_119	0.0	0.0	0.0
Cluster_12	1333.0	188.0	127.0
Cluster_120	0.0	0.0	0.0
Cluster_121	0.0	0.0	0.0
Cluster_122	0.0	0.0	0.0
Cluster_123	17.5	3.0	27.5
Cluster_125	45.5	6.0	22.5

# Outputs

---

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_unstrat\_EC.tsv**

**FROGSFUNC\_2\_functions: EC\_copynumbers\_predicted.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_excluded.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_weighted\_nsti.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_marker\_norm.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.fasta**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.biom**

**FROGSFUNC\_2\_functions: report.html**



sample	weighted_NSTI
BHT0.LOT01	0.0814507179687713
BHT0.LOT03	0.07523644621312382
BHT0.LOT04	0.06550232405467385
BHT0.LOT05	0.08141930786656948
BHT0.LOT06	0.08495448189855995
BHT0.LOT07	0.08161575516954905
BHT0.LOT08	0.08233567661364216

the mean of NSTI value per sample.



# Outputs

---



**FROGSFUNC\_2\_functions: frogsfunc\_functions\_unstrat\_EC.tsv**

**FROGSFUNC\_2\_functions: EC\_copynumbers\_predicted.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_excluded.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_weighted\_nsti.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_marker\_norm.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.fasta**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.biom**

**FROGSFUNC\_2\_functions: report.html**

# FROGSFUNC\_2: excluded.tsv

#Cluster	FROGS_taxonomy	PICRUSt2_taxonomy
Cluster_2	Bacteria;Proteobacteria;Gammaproteobacteria;Enterobac	Bacteria;Proteobacteria;Gammaproteobacter
Cluster_3	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;l	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lact
Cluster_7	Bacteria;Firmicutes;Bacilli;Erysipelotrichales;Erysipelotricha	Bacteria;Tenericutes;Mollicutes;Acholeplasma
Cluster_9	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;l	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lact
Cluster_14	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;l	Bacteria;Firmicutes;Bacilli;Lactobacillales;Lact
Cluster_17	Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae	Bacteria;Firmicutes;Bacilli;Lactobacillales;Strep

exclusion_paramater	value_parameter
min_blast_coverage	coverage = 0.8190999999999999
min_blast_coverage	coverage = 0.8327
min_blast_identity,min_blast_coverage,max_nsti	identity = 0.8289,coverage = 0.711,nsti = 0.43916499999999997
min_blast_coverage	coverage = 0.8632
min_blast_coverage	coverage = 0.8327
min_blast_coverage	coverage = 0.66870000000000001

Information (FROGS taxonomy, PICRUSt2 taxonomy, exclusion\_parameter, value\_parameter) on deleted ASV that are out of the cut-off values selected in this step.

# Outputs

---



**FROGSFUNC\_2\_functions: frogsfunc\_functions\_unstrat\_EC.tsv**

**FROGSFUNC\_2\_functions: EC\_copynumbers\_predicted.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_excluded.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_weighted\_nsti.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_marker\_norm.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.fasta**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.biom**

**FROGSFUNC\_2\_functions: report.html**

# FROGSFUNC\_2: EC\_copynumber\_predicted.tsv

ASV	EC:1.1.1.1	EC:1.1.1.10	EC:1.1.1.100	EC:1.1.1.101	EC:1.1.1.102	EC:1.1.1.103	EC:1.1.1.105
Cluster_1	2	0	2	0	0	0	0
Cluster_10	1	0	4	0	0	0	0
Cluster_100	2	0	8	0	0	1	0
Cluster_101	0	0	2	0	0	0	1
Cluster_102	0	0	3	0	0	0	1
Cluster_103	0	0	2	0	0	0	0
Cluster_104	0	0	2	0	0	0	0
Cluster_105	1	0	3	0	0	0	0
Cluster_106	1	0	4	0	0	0	0
Cluster_107	3	0	4	0	0	0	0
Cluster_108	3	0	10	0	0	0	0
Cluster_109	0	0	2	0	0	0	0
Cluster_11	5	0	5	0	0	1	0
Cluster_110	0	0	2	0	0	0	0
Cluster_111	5	0	5	0	0	1	0
Cluster_112	2	0	1	0	0	0	0
Cluster_113	4	0	3	0	0	0	0
Cluster_114	2	0	3	0	0	0	0

Output table of predicted function copy numbers per ASV.

One per chosen target function database (EC, KO, PFAM, COG, TIGRFAM, PHENO).

# Outputs

---



**FROGSFUNC\_2\_functions: frogsfunc\_functions\_unstrat\_EC.tsv**

**FROGSFUNC\_2\_functions: EC\_copynumbers\_predicted.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_excluded.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_weighted\_nsti.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions\_marker\_norm.tsv**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.fasta**

**FROGSFUNC\_2\_functions: frogsfunc\_functions.biom**

**FROGSFUNC\_2\_functions: report.html**

# FROGSFUNC\_2: unstrat\_EC.tsv

---

classification	db_link	observation_name	BHT0.LOT01	BHT0.LOT03	BHT0.LOT04
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.1	<a href="https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.1">https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.1</a>	EC:1.1.1.1	13205	15364	13783
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.100	<a href="https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.100">https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.100</a>	EC:1.1.1.100	20029	21513	23461
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.103	<a href="https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.103">https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.103</a>	EC:1.1.1.103	177	2011	334
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.105	<a href="https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.105">https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.105</a>	EC:1.1.1.105	0	0	0
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.108	<a href="https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.108">https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.108</a>	EC:1.1.1.108	142	194	127
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.11	<a href="https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.11">https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.11</a>	EC:1.1.1.11	2	10	2
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.122	<a href="https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.122">https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.122</a>	EC:1.1.1.122	16	109	54
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.125	<a href="https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.125">https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.125</a>	EC:1.1.1.125	2218	505	1656



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

---

# FROGSFUNC\_3\_pathways

---

# How it works ?

---

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

FROGSFUNC\_1\_placeseqs\_copynumber

2. Predicts number of function copy number in each ASV and calculates functions abundances in each sample and ASV abundances according to marker copy number.

FROGSFUNC\_2\_functions

3. Calculates pathway abundances in each sample.

FROGSFUNC\_3\_pathways



It runs only on 16S, ITS or 18S



# FROGSFUNC\_3

**FROGSFUNC\_3\_pathways** 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.

FROGSFUNC\_2 functions\_unstrat\_EC.tsv

	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



FROGSFUNC\_3 pathways\_unstrat per sample and per reference

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

# FROGSFUNC\_3

**FROGSFUNC\_3\_pathways** Calculates pathway abundances in each sample.  
(Galaxy Version 4.1.0+galaxy1)

☆ Favorite

🔄 Versions

▼ Options

## Function abundance file

61: FROGSFUNC\_2\_functions: frogsfunc\_functions\_unstrat\_EC.tsv

TSV function abundances table from FROGSFUNC\_2\_functions tool, FROGSFUNC\_2\_functions\_unstrat\_EC.tsv for Metacyc database or FROGSFUNC\_2\_functions\_unstrat\_KO.tsv for Kegg database (unstratified table). (--input-file)

## Taxonomic marker

- 16S
- ITS
- 18S

Taxonomic marker of interest.

## Pathway reference

- Metacyc
- Kegg

Metacyc is the only choice for ITS and 18S

For 16S marker, choose Metacyc or KEGG in accordance with your choice in the FROGSFUNC\_1\_placeseqs\_copynumbers tool.  
For 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)

## Email notification

No

Send an email notification when the job completes.

✓ Execute



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.

# Outputs

---



# FROGSFUNC\_3: report.html

## Pathway abundances per sample

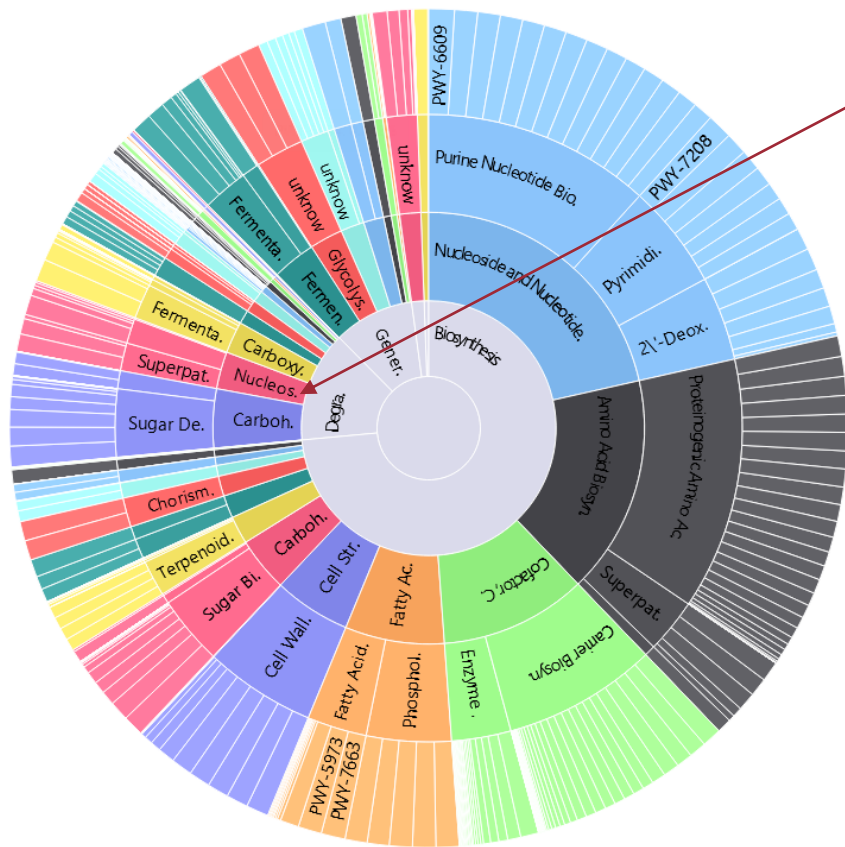
Display global distribution ← Displays the pathway distribution on all samples

Show 10 entries Search:  CSV

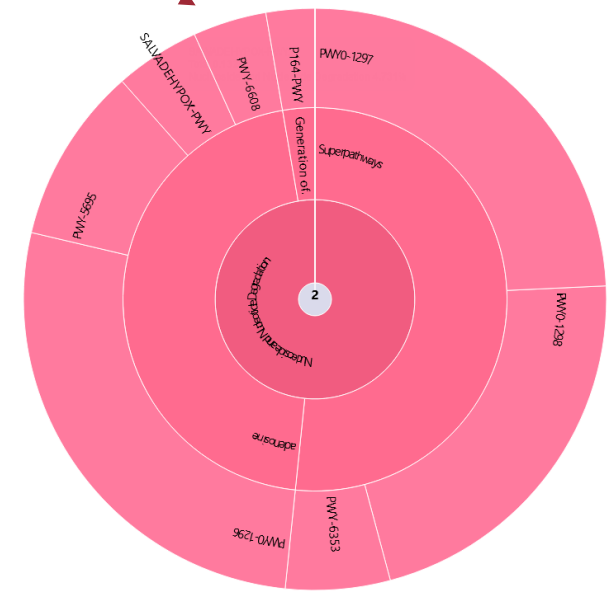
<input type="checkbox"/>	Samples	Nb pathway retrieved
<input checked="" type="checkbox"/>	BHT0.LOT01	350
<input checked="" type="checkbox"/>	BHT0.LOT03	362
<input checked="" type="checkbox"/>	BHT0.LOT04	369
<input checked="" type="checkbox"/>	BHT0.LOT05	364
<input checked="" type="checkbox"/>	BHT0.LOT06	360
<input checked="" type="checkbox"/>	BHT0.LOT07	357
<input checked="" type="checkbox"/>	BHT0.LOT08	368
<input checked="" type="checkbox"/>	BHT0.LOT10	367
<input type="checkbox"/>	CDT0.LOT02	333
<input type="checkbox"/>	CDT0.LOT04	344

Display distribution ← Displays the pathway distribution on selected samples

# What is the distribution of pathway abundances in the samples ?



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



# Outputs

---



**FROGSFUNC\_3\_pathways: frogsfunc\_pathways\_unstrat.tsv**

**FROGSFUNC\_3\_pathways: report.html**

pathways abundance predictions of metagenome per sample.

# FROGSFUNC\_3: Pathway abundance tables

Abundance of each pathway inside each sample.

classification	db_link	observation_name	BHT0.LOT01	BHT0.LOT03	BHT0.LOT04	BHT0.LOT05
Biosynthesis;Cofactor, Carrier, and Vitamin Biosynthesis;Carrier Biosynthesis;1CMET2-PWY	<a href="https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=1CMET2-PWY">https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=1CMET2-PWY</a>	1CMET2-PWY	5927	5375	5114	5923
Degradation/Utilization/Assimilation;Aromatic Compound Degradation;unknown;3-HYDROXYPHENYLACETATE-DEGRADATION-PWY	<a href="https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=3-HYDROXYPHENYLACETATE-DEGRADATION-PWY">https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=3-HYDROXYPHENYLACETATE-DEGRADATION-PWY</a>	3-HYDROXYPHENYLACETATE-DEGRADATION-PWY	40	50	147	65
Biosynthesis;Secondary Metabolite Biosynthesis;Siderophore and Metallophore Biosynthesis;AEROBACTINSYN-PWY	<a href="https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=AEROBACTINSYN-PWY">https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=AEROBACTINSYN-PWY</a>	AEROBACTINSYN-PWY	8	0	5	7
Superpathways;Superpathways;unknown;ALL-CHORISMATE-PWY	<a href="https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ALL-CHORISMATE-PWY">https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ALL-CHORISMATE-PWY</a>	ALL-CHORISMATE-PWY	81	215	444	211
Generation of Precursor Metabolites and Energy;Fermentation;Fermentation to Short-Chain Fatty Acids;ANAEROFrucAT-PWY	<a href="https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ANAEROFrucAT-PWY">https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ANAEROFrucAT-PWY</a>	ANAEROFrucAT-PWY	7998	7020	8141	7347
Generation of Precursor Metabolites and Energy;Glycolysis;unknown;ANAGLYCOLYSIS-PWY	<a href="https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ANAGLYCOLYSIS-PWY">https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ANAGLYCOLYSIS-PWY</a>	ANAGLYCOLYSIS-PWY	7791	7449	8141	7353
Degradation/Utilization/Assimilation;Amino Acid Degradation;Proteinogenic Amino Acid Degradation;ARGDEG-PWY	<a href="https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ARGDEG-PWY">https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ARGDEG-PWY</a>	ARGDEG-PWY	3	0	47	39
Biosynthesis;Amino Acid Biosynthesis;Proteinogenic Amino Acid Biosynthesis;ARGSYN-PWY	<a href="https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ARGSYN-PWY">https://biocyc.org/META/NEW-IMAGE?type=PATHWAY&amp;object=ARGSYN-PWY</a>	ARGSYN-PWY	6527	5811	6431	6372

External link on the pathway



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

---

# Statistics

---

DIFFERENTIAL ANALYSIS



---

# FROGSSTAT DESeq2 preprocess

---

**FROGSSTAT DESeq2 Preprocess** import a Phyloseq object and prepare it for DESeq2 differential abundance analysis (Galaxy Version 4.1.0+galaxy1)

☆ Favorite

🔄 Versions

▼ Options

### Type of analysis

ASV

FUNCTION

Type of data to perform the differential analysis. ASV: DESeq2 is run on the ASV abundance table. FUNCTION: predicted function abundance table from FROGSFUNC\_2\_function tool.

classification	db_link	observation_name	BHT0.LOT01	BHT0.LOT03	BHT0.LOT04
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.1	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.1	EC:1.1.1.1	13205	15364	13783
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.100	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.100	EC:1.1.1.100	20029	21513	23461
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.103	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.103	EC:1.1.1.103	177	2011	334
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.105	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.105	EC:1.1.1.105	0	0	0
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.108	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.108	EC:1.1.1.108	142	194	127
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.11	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.11	EC:1.1.1.11	2	10	2
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.122	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.122	EC:1.1.1.122	16	109	54
Oxidoreductases;Acting on the CH-OH group of donors;With NAD+ or NADP+ as acceptor;EC:1.1.1.125	https://www.genome.jp/dbget-bin/www_bget?EC:1.1.1.125	EC:1.1.1.125	2218	505	1656

### Function abundances file

61: FROGSFUNC\_2\_functions: frogsfunc\_functions\_unstrat\_EC.tsv

Input file of predicted function abundances (frogsfunc\_functions\_unstrat.tsv from FROGSFUNC\_2\_function tool).

### Metadata associated to samples (format: TSV)

2: metadata\_chaillou.tsv

The file must contain the metadata that characterise each sample. (--samplefile)

	EnvType	Description	FoodType
BHT0.LOT01	BoeufHache	LOT1	Meat
BHT0.LOT03	BoeufHache	LOT3	Meat
BHT0.LOT04	BoeufHache	LOT4	Meat

### Experimental variable

EnvType

The factor that could have an effect on ASV/FUNCTION abundances. Ex: Treatment, etc.

### Do you want to correct a confounding factor?

False

If yes, specify the counfouding factor

### Email notification

No

Send an email notification when the job completes.

✓ Execute

## Outputs

FROGSSTAT DESeq2 Preprocess: function\_data.Rdata

FROGSSTAT DESeq2 Preprocess: function\_dds.Rdata

---

# FROGSSTAT DESeq2 visualisation

---

Type of analysis

- ASV
- FUNCTION

Type of data to perform the differential analysis. ASV: DESeq2 is run on the ASV abundance table. FUNCTION: DESeq2 is run on predicted function abundance table from FROGSFUNC\_2\_function tool.

Data object (format: data.RData)

74: FROGSSTAT DESeq2 Preprocess: function\_data.Rdata **data.Rdata**

For ASV: asv\_data.Rdata from FROGSSTAT\_Phyloseq\_Import\_Data tool - For FUNCTION: function\_data.Rdata from FROGSSTAT\_DESeq2\_Preprocess tool. (--abundanceData)

DESeq2 object (format: dds.RData)

73: FROGSSTAT DESeq2 Preprocess: function\_dds.Rdata **dds.Rdata**

This is the result of FROGSSTAT\_DESeq2\_Preprocess tool asv\_dds.Rdata or function\_dds.Rdata (--dds)

Experimental variable

EnvType

The factor that could have an effect on ASV/FUNCTION abundances. Ex : Treatment (var)

The experimental variable is it quantitative or qualitative?

Qualitative

If qualitative, choose 2 conditions to compare

Condition 1 considered as reference

BoeufHache

One condition of the experimental variable (e.g. with) (--mod2)

Condition 2 to be compared to the reference

VeauHache

Another condition of the experimental variable (e.g. without) (--mod1)

	EnvType	Description	FoodType
BHT0.LOT01	BoeufHache	LOT1	Meat
BHT0.LOT03	BoeufHache	LOT3	Meat
BHT0.LOT04	BoeufHache	LOT4	Meat
VHT0.LOT07	VeauHache	LOT7	Meat
VHT0.LOT08	VeauHache	LOT8	Meat
VHT0.LOT10	VeauHache	LOT10	Meat
SFT0.LOT01	SaumonFume	LOT1	Seafood
SFT0.LOT02	SaumonFume	LOT2	Seafood
SFT0.LOT03	SaumonFume	LOT3	Seafood

Adjusted p-value threshold

0.05 **For the moment, we keep default p-value**

Threshold used for statistical significance of the differentially abundant ASV/FUNCTION analysis (--padj)

# Outputs

---

FROGSSTAT DESeq2 Visualisation: ipath\_under.tsv

FROGSSTAT DESeq2 Visualisation: ipath\_over.tsv



FROGSSTAT DESeq2 Visualisation: report.nb.html

# DESeq2 visualisation: report.html

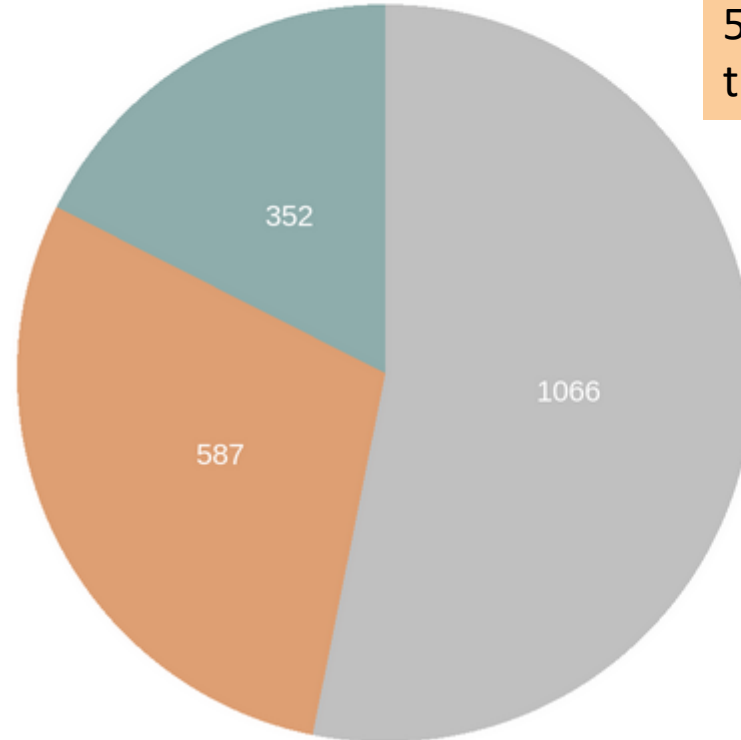
---

	ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Level_4
	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text"/>	<input type="text"/>	<input type="text" value="A"/>	<input type="text" value=","/>	<input type="text" value="All"/>
413	EC:3.4.24.30	1.96849	4.53492	1.30759	3.46814	0.000524070	0.00249661	Hydrolases
691	EC:6.3.2.43	2.57016	-6.11515	2.29021	-2.67013	0.00758220	0.0215825	Ligases

You chose to compare VeauHache to the reference modality BoeufHache. This implies that a positive log2FoldChange means more abundant in VeauHache than in BoeufHache.

# DESeq2 visualisation: report.html

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

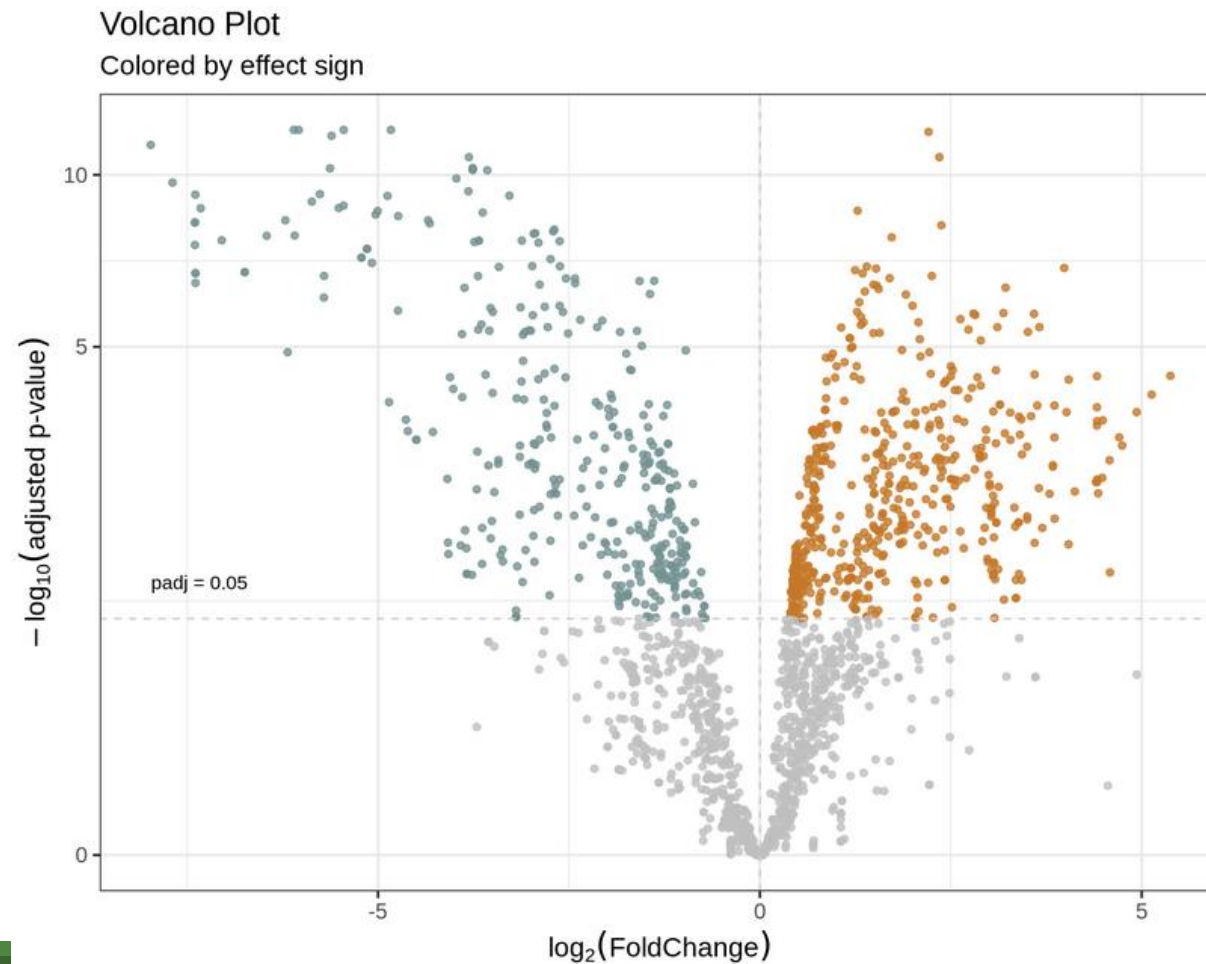


587 functions are more abundant in VeauHache than in BoeufHache

352 functions are more abundant in BoeufHache than in VeauHache

- Differentially Abundant (log-fold change < 0)
- Differentially Abundant (log-fold change > 0)
- Not Differentially Abundant

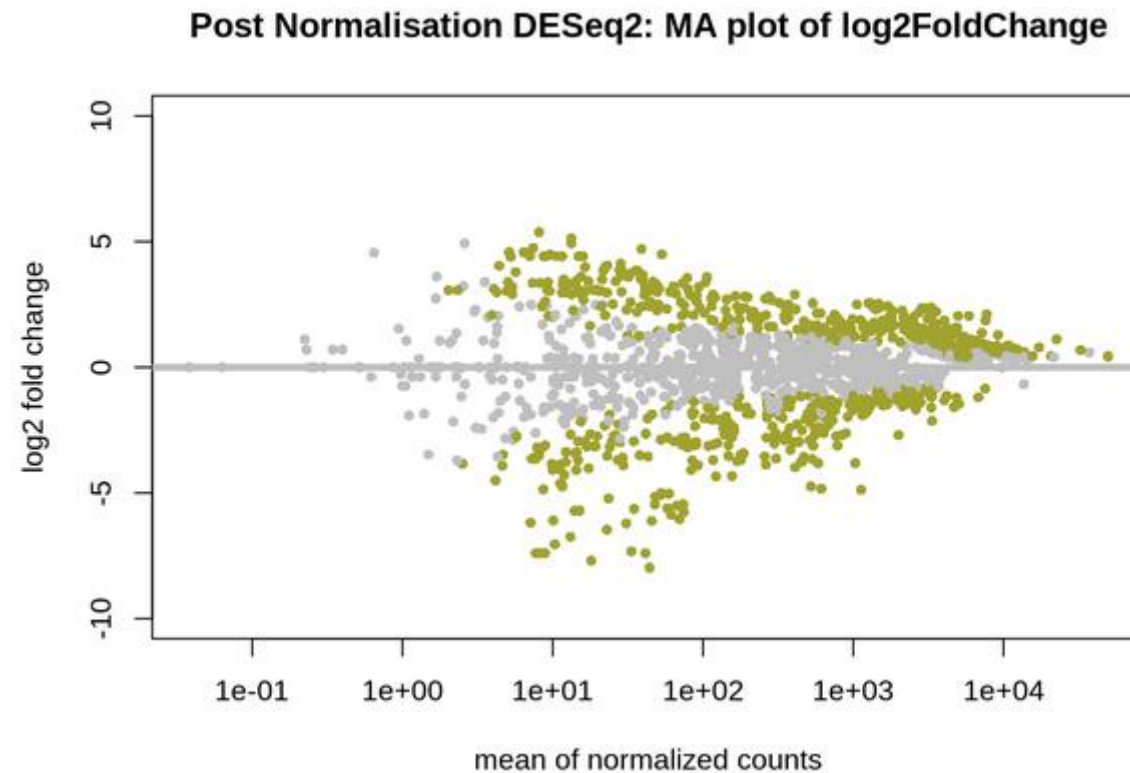
# DESeq2 visualisation: report.html



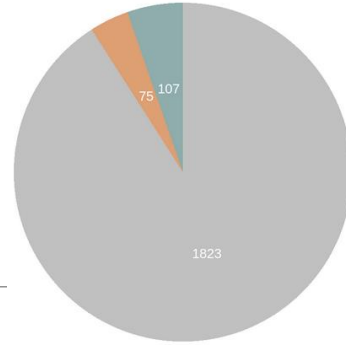


# DESeq2 visualisation: report.html

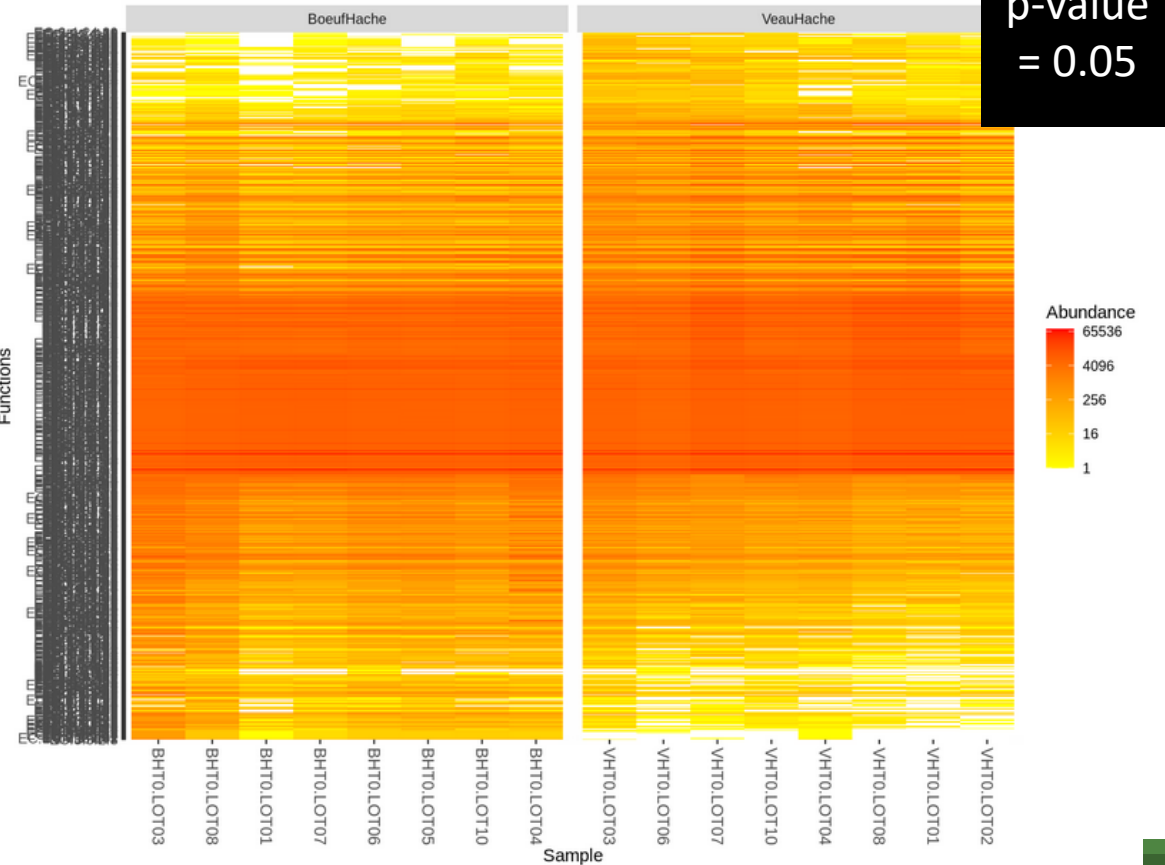
---



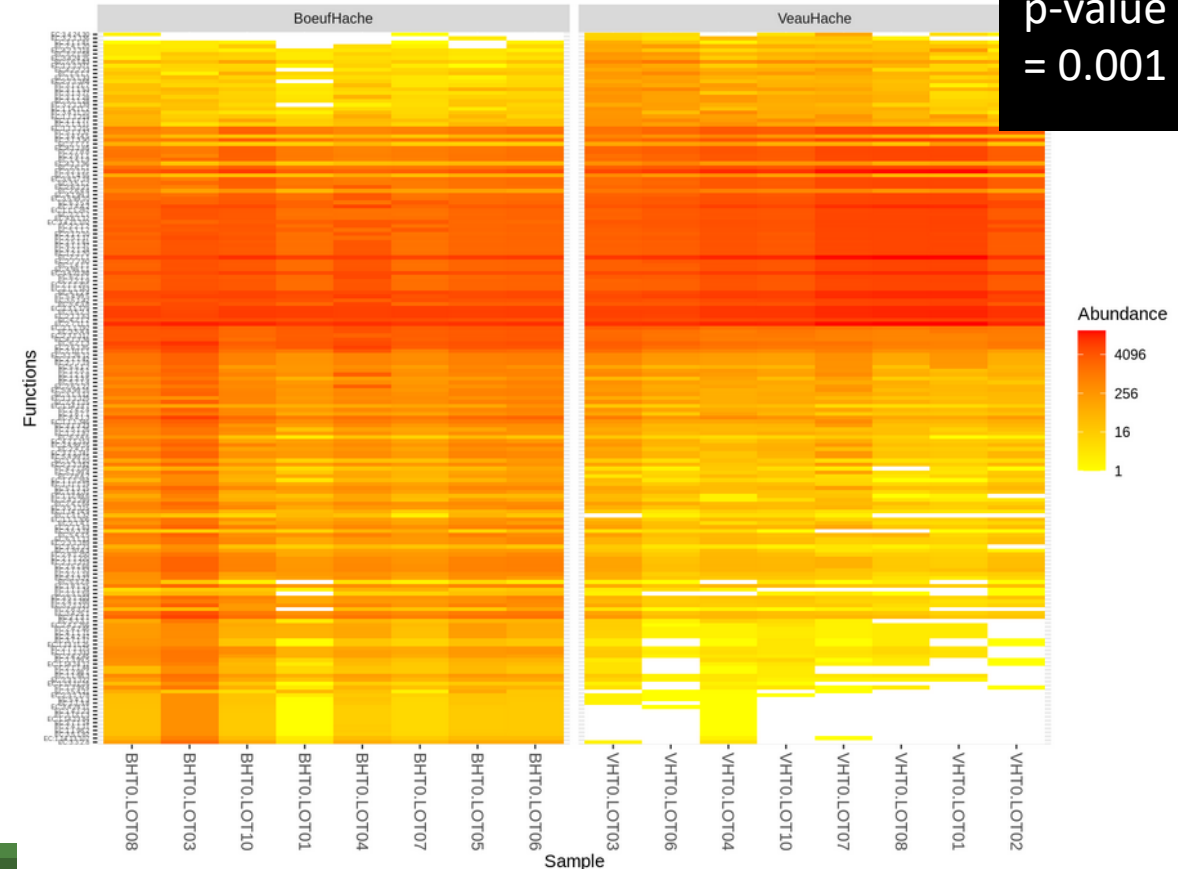
# DESeq2 visualisation: report.html



Heatmap plot of DA asv or functions, between 2 conditions  
EnvType\_VeauHache\_vs\_BoeufHache



Heatmap plot of DA asv or functions, between 2 conditions  
EnvType\_VeauHache\_vs\_BoeufHache

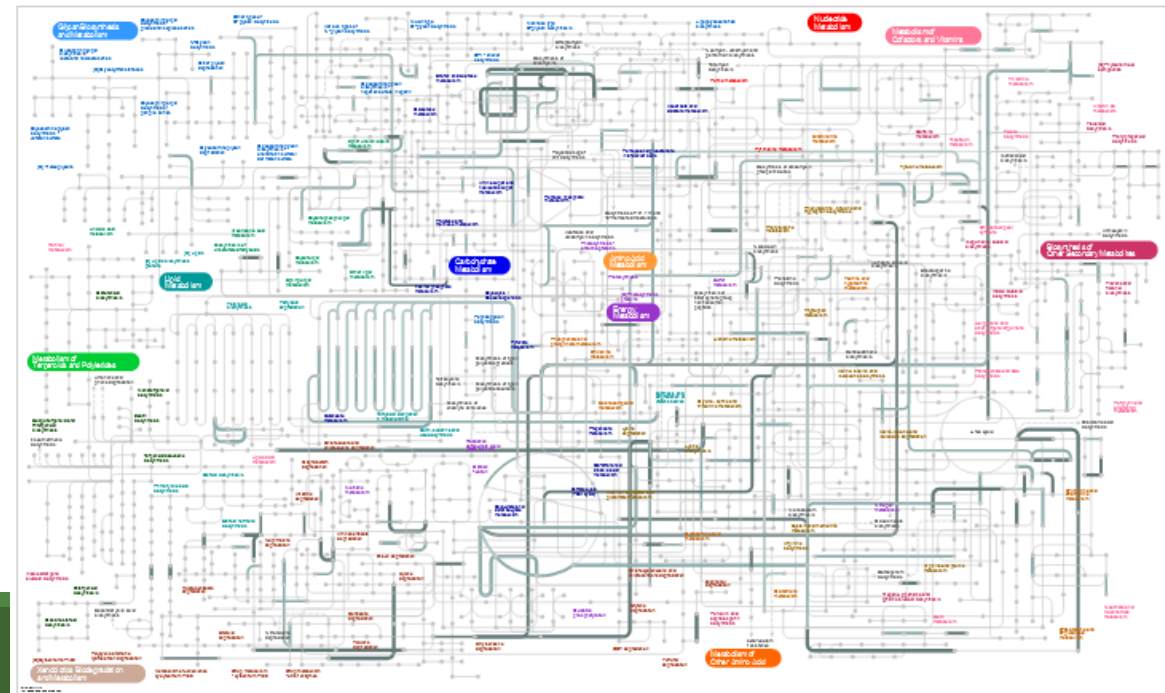
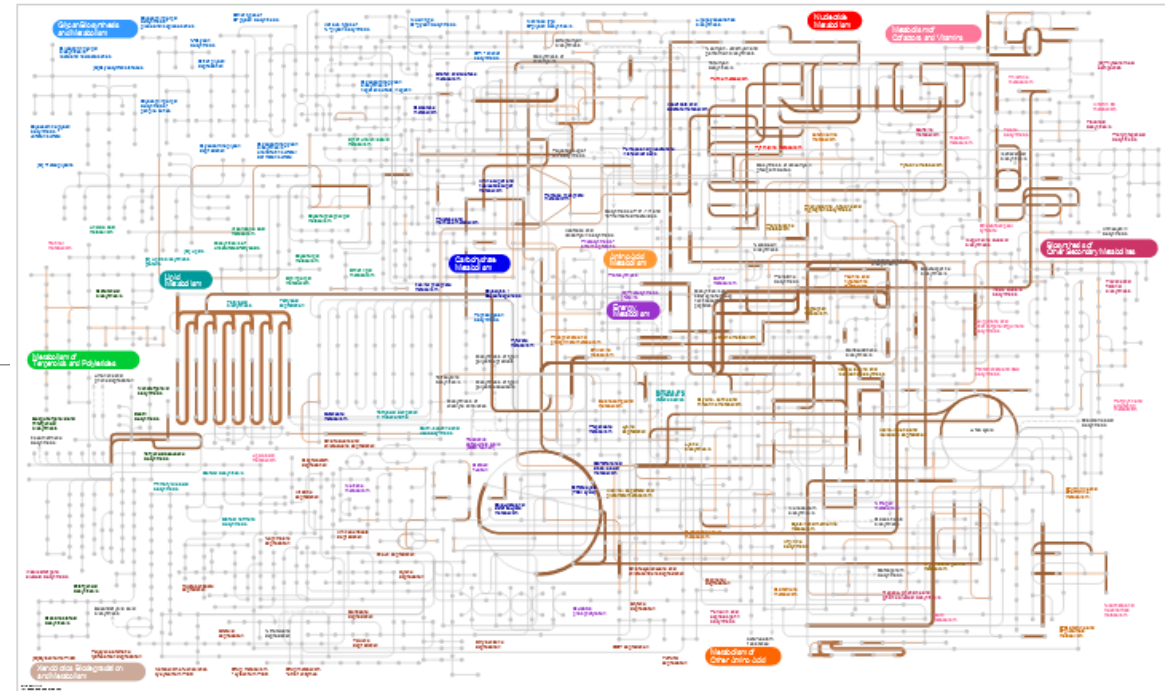


# DESeq2 visualisation: report.html

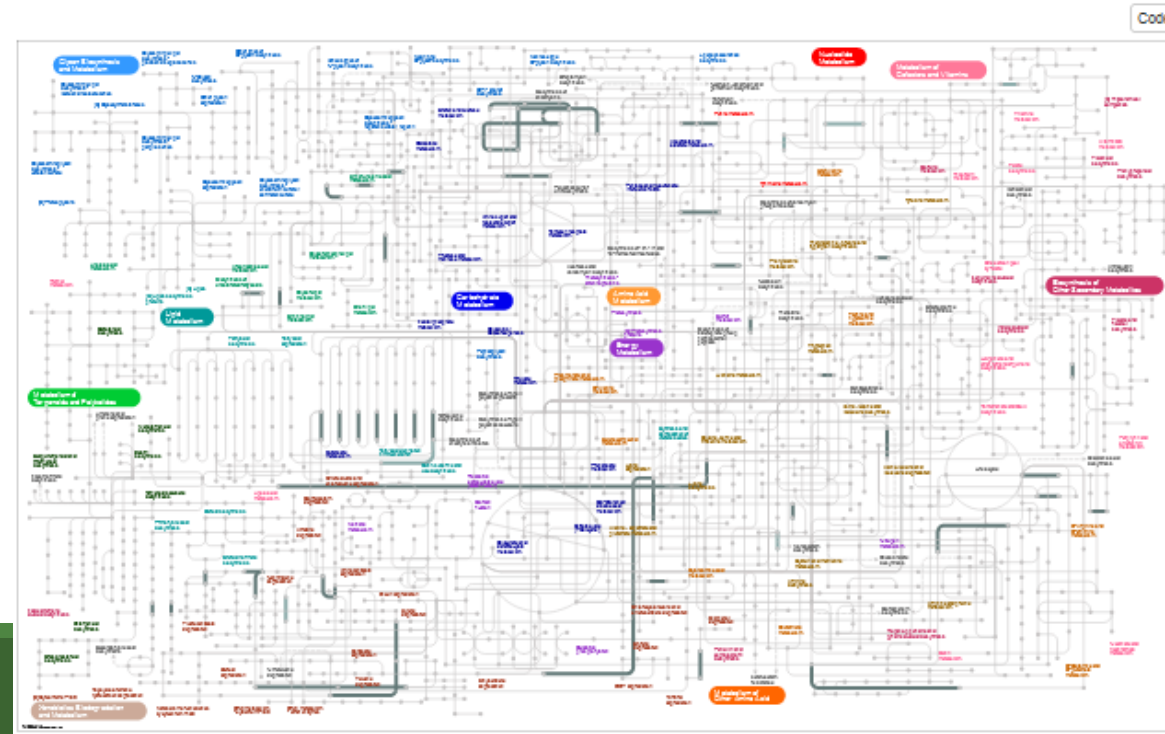
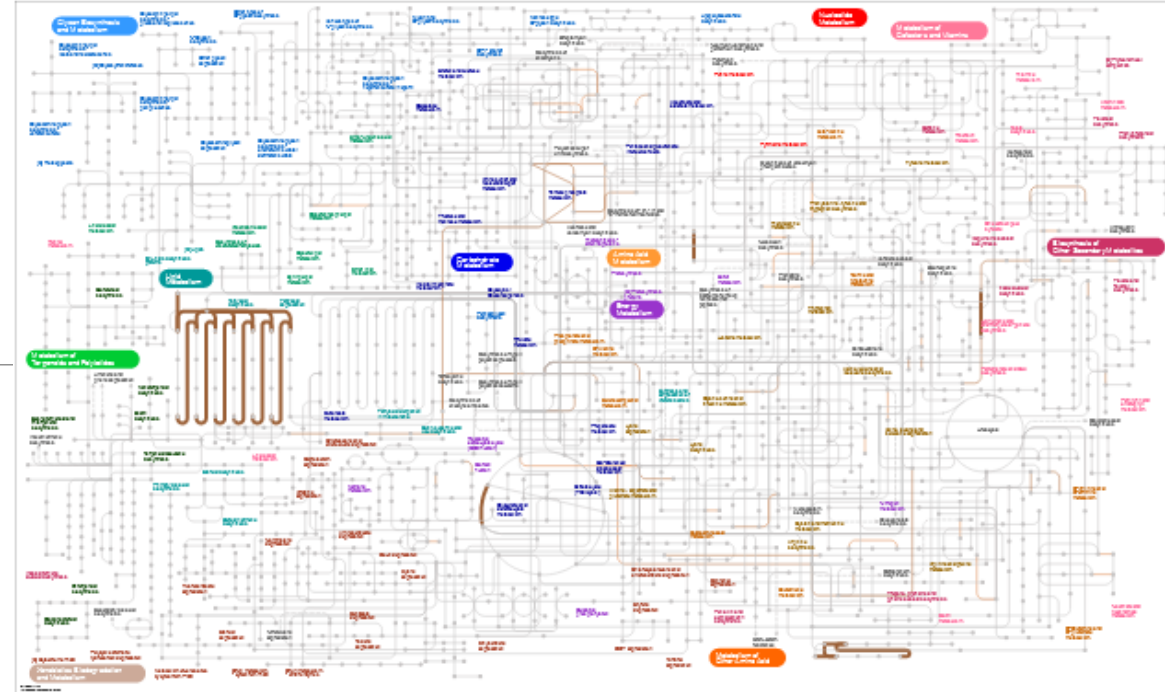
You chose to compare VeauHache to the reference modality BoeufHache. This implies that the over-abundant pathways (first image) and under-abundant pathways (second image) are involved in VeauHache condition.

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

p-value  
= 0.05



# DESeq2 visualisation: report.html



p-value  
= 0.001

# Outputs

---



FROGSSTAT DESeq2 Visualisation: ipath\_under.tsv



FROGSSTAT DESeq2 Visualisation: ipath\_over.tsv

FROGSSTAT DESeq2 Visualisation: report.nb.html

To visualise and explore metabolic pathways with [IPATH3 website](#)

# Go to IPATH3

ID Color Width

EC2.6.1.1 #637978 W12

EC2.7.8.8 #637978 W12

EC2.2.1.1 #8EADAC W12

EC4.1.1.65 #637978 W12

EC4.1.99.3 #8EADAC W12

EC2.2.1.2 #8EADAC W12

EC4.2.1.118 #637978 W12

EC2.2.1.7 #8EADAC W12

EC4.99.1.1 #8EADAC W12

EC2.5.1.17 #8EADAC W12

ipath\_over.tsv

ID Color Width

EC1.3.99.4 #DE894E W12

EC2.7.1.175 #DE894E W12

EC3.4.25.1 #DE894E W12

EC1.3.99.5 #DE894E W12

EC3.3.2.8 #DE894E W12

EC2.7.7.53 #DE894E W12

EC2.5.1.68 #DE894E W12

EC1.14.14.12 #DE894E W12

EC2.1.1.219 #DE894E W12

EC2.1.1.220 #DE894E W12

ipath\_under.tsv

<https://pathways.embl.de/>

iPATH3 Pathway maps Tools Share Help

## 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 capabilities. 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

Display



#### Secondary metabolites

An overview of pathways involved in biosynthesis of secondary metabolites

Display



#### Antibiotics

An overview of pathways involved in the biosynthesis of various antibiotics

Display



#### Microbial metabolism

Covers microbial metabolism in diverse environments

Display

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

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

# Go to IPATH3

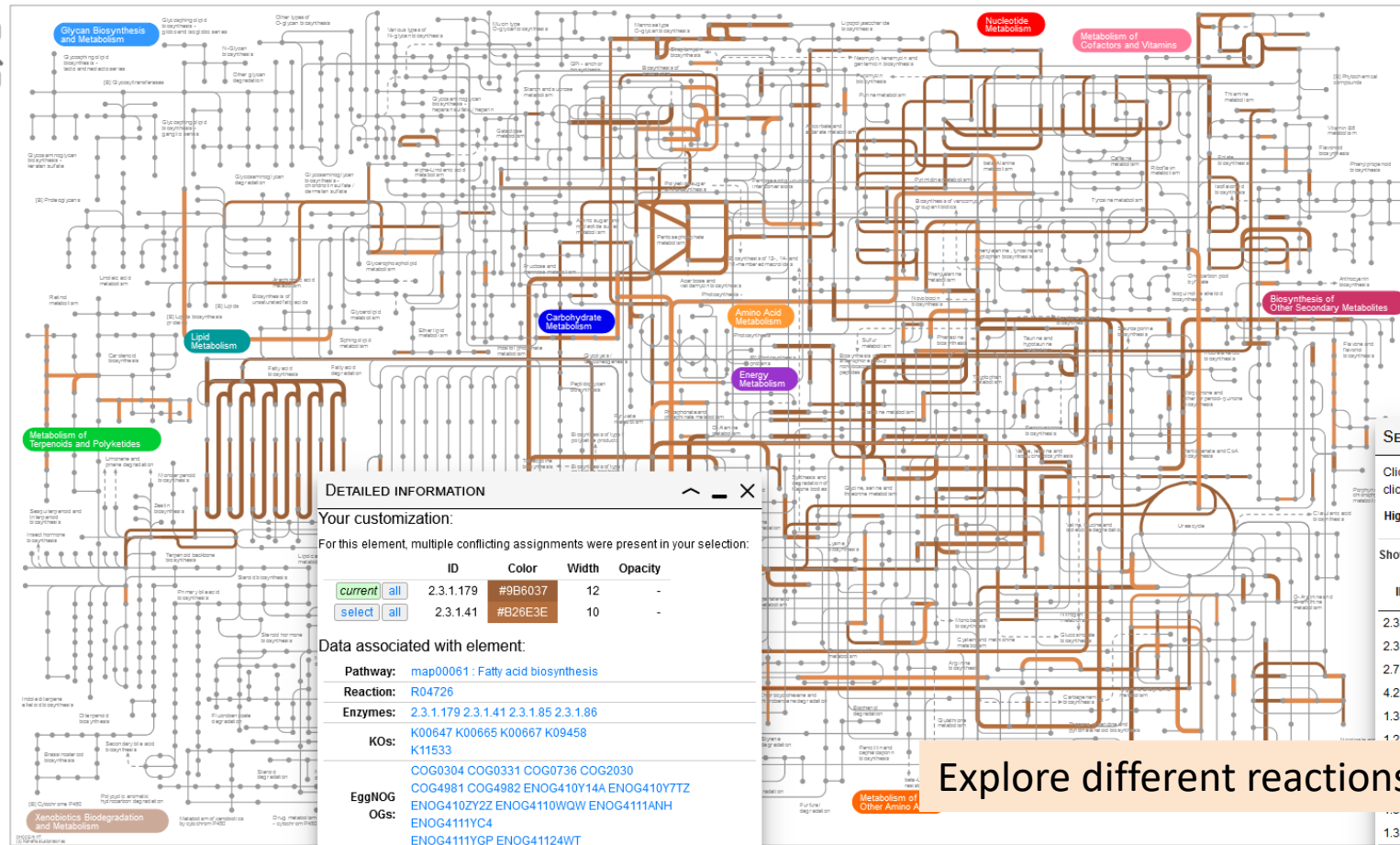
Help ▾

- Help pages
- Video tutorials
- About & Contact



Pathway maps ▾ Tools Share My account Help

Search current map Search geraldine ▾



**CONTROLS**

Customize Export

Element selection:

- EC3.6.4.13 #9B6037 WB
- EC4.2.2.7 #DE894E WB
- EC5.4.99.23 #9B6037 WB
- EC2.3.1.12 #9B6037 WB

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: [COG0304](#) [COG0331](#) [COG0736](#) [COG2030](#) [COG4981](#) [COG4982](#) [ENOG410Y14A](#) [ENOG410Y7TZ](#) [ENOG410ZYZZ](#) [ENOG4110WQW](#) [ENOG4111ANH](#) [ENOG4111Y4](#) [ENOG4111YGP](#) [ENOG41124WT](#)

Other: 9164 KEGG genes, 1305 NCBI genes, 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
1.3.1.9	7	0	7

Page: 1 / 59 (588 IDs) << < 1 2 3 4 5 ... 59 >>

Explore different reactions customize