

# Training on Galaxy: Metabarcoding

June 2022 - Webinar

## FROGS Practice on function inference

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

---

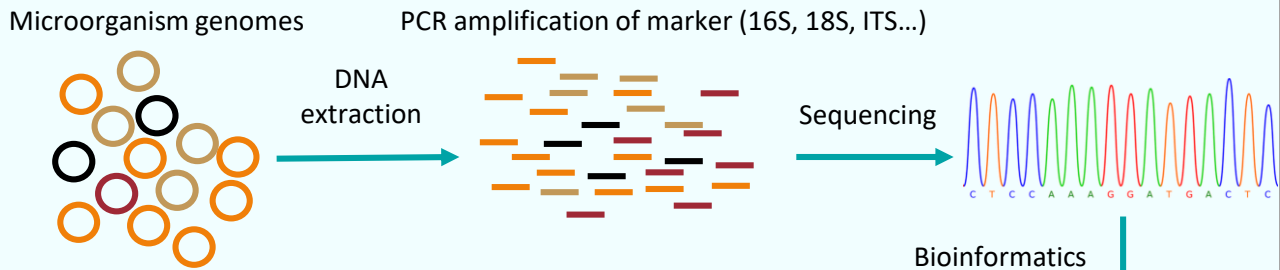
What metabolic functions are present in the environment?

---

# Concepts

## Metabarcoding principle

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

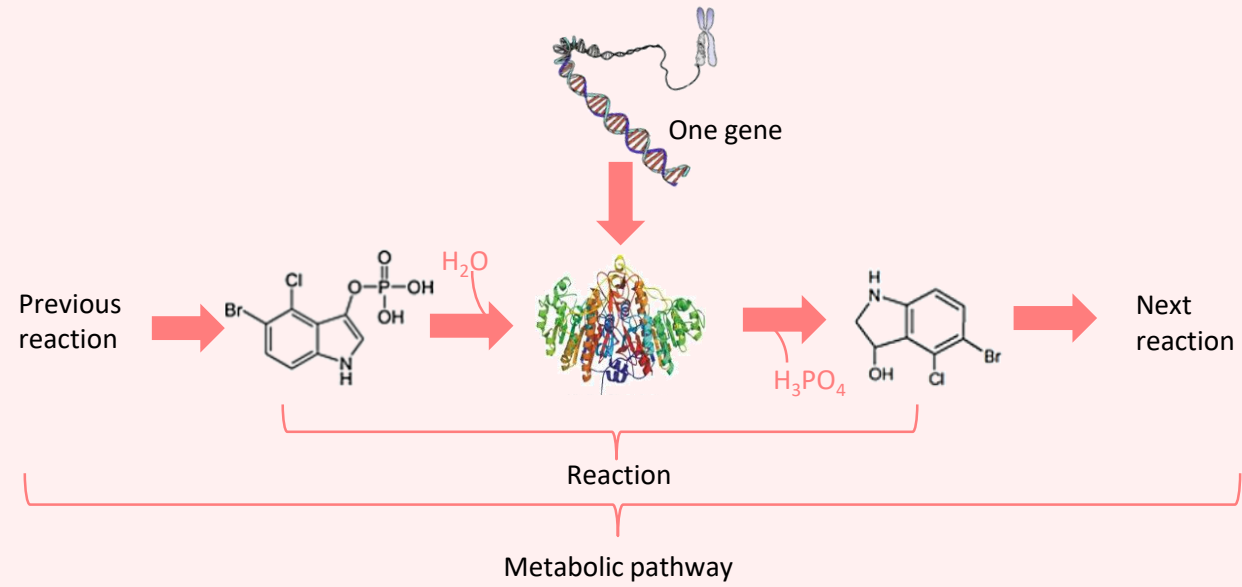


Abundance table displaying microbial diversity per samples

	Taxonomic affiliation	Sample1	Sample2	Sample3
OTU1	Species A	3500	6300	210
OTU2	Species B	0	460	36
OTU3	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 metabolic functions are presents in a microbial community ?

# Based on PICRUSt2

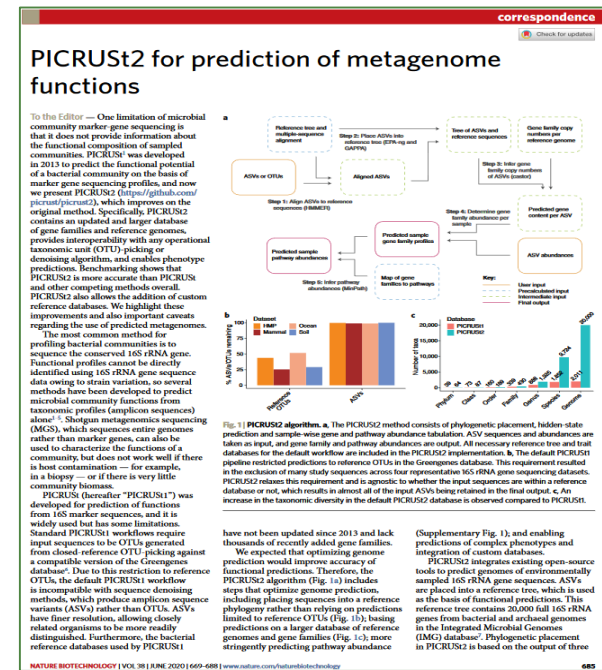
PICRUSt (Phylogenetic investigation of communities by reconstruction of unobserved states) is an open-source tool.

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 OTUs into a reference phylogenetic tree.

FROGSFUNC\_step1\_placeseqs

2. Predicts number of marker and function copy number in each OTU.

FROGSFUNC\_step2\_copynumbers

3. Calculates functions abundances in each sample.

FROGSFUNC\_step3\_functions

4. Calculates pathway abundances in each sample.

FROGSFUNC\_step4\_pathways



It runs only on 16S, ITS or 18S

---

FROGSFUNC\_step1\_placeseqs

---

# FROGSFUNC\_step1\_placeseqs

---

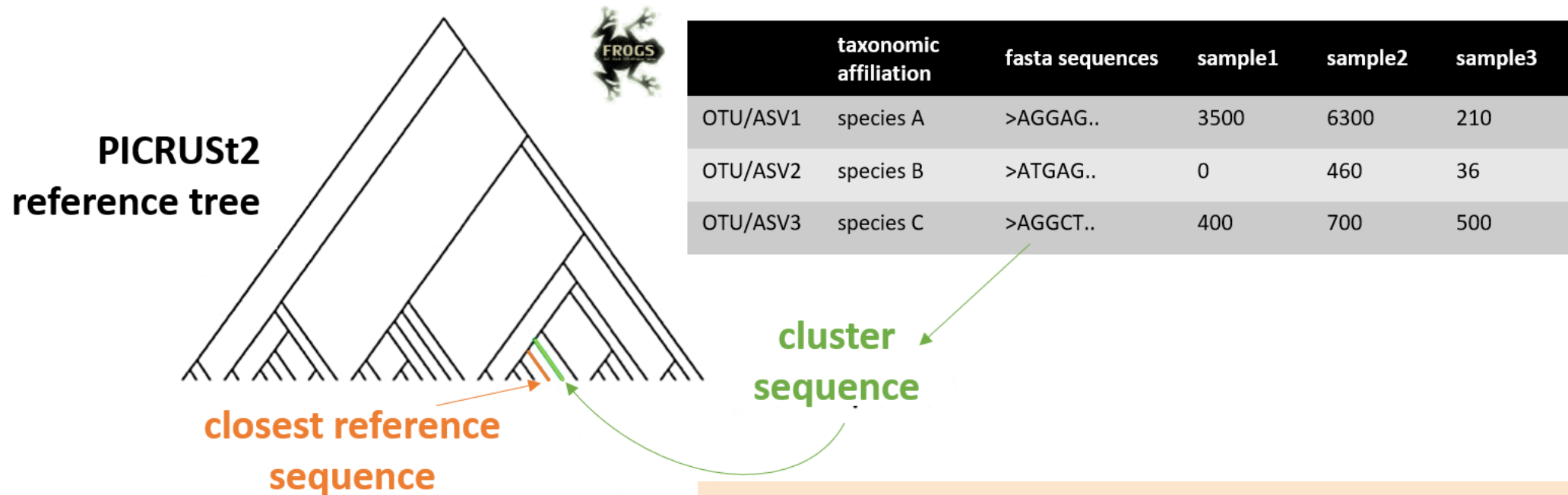
**FROGSFUNC\_step1\_placeseqs** is the first step of PICRUSt2.

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 script performs this step, which specifically:

- Aligns your study sequences with a multiple-sequence alignment of reference 16S, ITS or 18S sequences with [HMMER](#).
- Finds the most likely placements of your study sequences in the reference tree with [EPA NG](#) or [SEPP](#).
- Produces a treefile with the most likely placement for each sequence as the new tips with [GAPPA](#).

# FROGSFUNC\_step1\_placeseqs



**2 input files** are required for FROGSFUNC\_step1\_placeseqs analysis:

- **fasta file** of OTU/ASV sequences (it can be from FROGS Filters step)
- **biom file** of OTU/ASV abundances with taxonomic affiliation information (it can be from FROGS Affiliation OTU step)



# FROGSFUNC\_step1\_placeseqs

**FROGSFUNC\_step1\_placeseqs** Places the OTUs into a reference phylogenetic tree. (Galaxy Version 4.0.0+galaxy1) ☆ Favorite ▼ Options

**Sequence file**  
   29: FROGS OTU Filters: otuFilter\_sequences.fasta

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

**Biom file**  
   33: FROGS Affiliation OTU: Pintail100affiliation\_abundance.biom

The abundance file to analyse (format: biom). Taxonomic affiliations must be inside (FROGS Affiliation OTU 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. (--placement-tool)

**Minimum alignment length**

Proportion of the total length of an input sequence that must align with reference sequences. All other 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 OTU/ASV fasta sequence file.
- **biom file:** The OTU [biom](#) file. Taxonomic affiliations must be done before (biom file form FROGS Affiliation OTU tool).
- **taxonomy marker:** 16S, ITS and 18S only available.  
 If your OTUs/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.

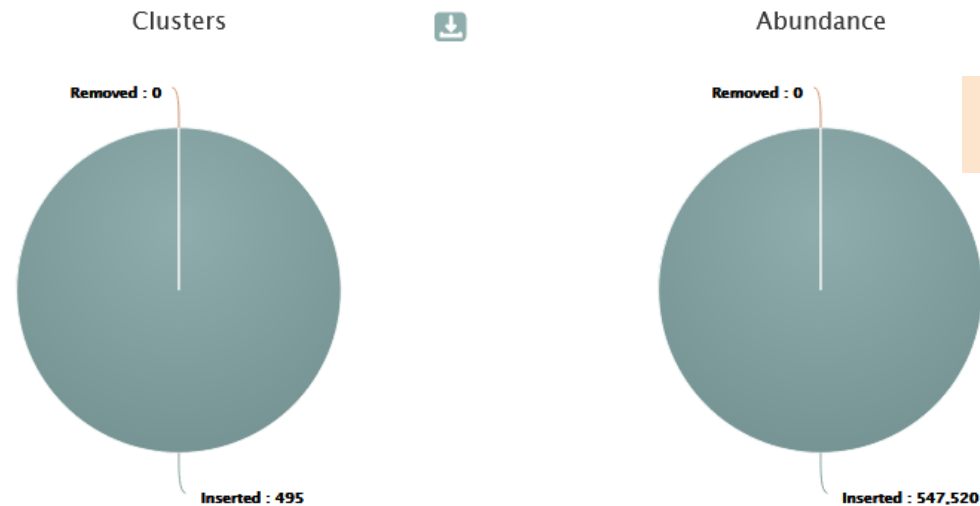
# Output files

---

FROGSFUNC_step1_placeseqs: frogsfunc_placeseqs.biom	  
FROGSFUNC_step1_placeseqs: frogsfunc_placeseqs_closests_ref_sequences.txt	  
FROGSFUNC_step1_placeseqs: frogsfunc_placeseqs.fasta	  
FROGSFUNC_step1_placeseqs: frogsfunc_placeseqs_excluded.tsv	  
FROGSFUNC_step1_placeseqs: frogsfunc_placeseqs_tree.nwk	  
FROGSFUNC_step1_placeseqs: report.html	  

# FROGSFUNC\_step1\_placeseqs: 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 which OTUs are contained or not in the phylogenetic tree.

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

The report file indicates for each OTU 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 OTUs inserted in the phylogenetic reference tree ?

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

Cluster <sup>↑↓</sup>	Nb sequences <sup>↑↓</sup>	FROGS Taxonomy <sup>↑↓</sup>	PICRUSt2 closest ID (JGI) <sup>↑↓</sup>	PICRUSt2 closest reference name <sup>↑↓</sup>
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.	2617271209	Flavobacterium hydatis DSM 2063
Cluster_100	696	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas;Multi-affiliation	2531839535	Stenotrophomonas maltophilia EPM1
Cluster_101	752	Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Weeksellaceae;Chryseobacterium;Chryseobacterium sp.	2639762901	Chryseobacterium jeonii DSM 17048
Cluster_102	524	Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Weeksellaceae;Chryseobacterium;Chryseobacterium antarcticum	2609460311	Chryseobacterium antarcticum LMG 24720

first part of the table

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

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

PICRUSt2 closest taxonomy	NSTI	NSTI Confidence	Lowest same taxonomic rank between FROGS and PICRUSt2	Comment
Bacteria;Firmicutes;Bacilli;Bacillales;Listeriaceae;Brochothrix;Brochothrix thermosphacta	0.0089	Good	Species	/
Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Flavobacteriaceae;Flavobacterium;Flavobacterium hydatis	0.021	Good		
Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas;Stenotrophomonas maltophilia	0.0073	Good	Species	identical taxonomy
Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Weeksellaceae;Kaistella;Chryseobacterium jeonii	0.12	Good	Family	/
Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Weeksellaceae;Kaistella;Chryseobacterium antarcticum	0.042	Good	Species	/

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

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.

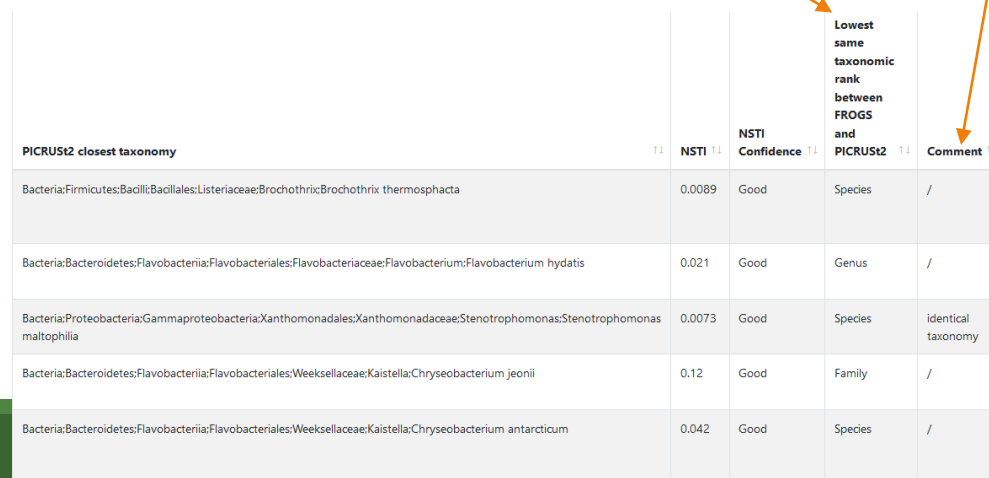
# FROGSFUNC\_step1\_placeseqs: report.html

**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 OTU/ASV sequence is strictly the same as the reference sequence.



PICRUST2 closest taxonomy	NSTI	NSTI Confidence	Lowest same taxonomic rank between FROGS and PICRUST2	Comment
Bacteria:Firmicutes:Bacilli:Bacillales>Listeriaceae:Brochothrix:Brochothrix thermosphacta	0.0089	Good	Species	/
Bacteria:Bacteroidetes:Flavobacteriia:Flavobacteriales:Flavobacteriaceae:Flavobacterium:Flavobacterium hydatis	0.021	Good	Genus	/
Bacteria:Proteobacteria:Gammaproteobacteria:Xanthomonadales:Xanthomonadaceae:Stenotrophomonas:Stenotrophomonas maltophilia	0.0073	Good	Species	identical taxonomy
Bacteria:Bacteroidetes:Flavobacteriia:Flavobacteriales:Weeksellaceae:Kaistella:Chryseobacterium jeonii	0.12	Good	Family	/
Bacteria:Bacteroidetes:Flavobacteriia:Flavobacteriales:Weeksellaceae:Kaistella:Chryseobacterium antarcticum	0.042	Good	Species	/



---

FROGSFUNC\_step2\_copynumbers

---

# FROGSFUNC\_step2\_copynumbers

---

It runs hidden-state prediction (hsp) to predict function abundances with **castor-R** of each OTUs placed in the PICRUSt2 reference phylogenetic tree.

**2 input files** are required for FROGSFUNC\_step2\_copynumbers analysis:

- **Tree file** (format newick nwk): The file contains the tree informations from FROGSFUNC\_step1\_placeseqs (FROGSFUNC\_step1\_placeseqs output : frogsfunc\_placeseqs\_tree.nwk)
- **OTUs biom file**: The abundance file to analyse *i.e.* FROGSFUNC\_step1\_placeseqs tool output file (format biom). (frogsfunc\_placeseqs.biom)

# FROGSFUNC\_step2\_copynumbers

---

FROGSFUNC\_step2\_copynumbers **runs hsp twice:**

- Prediction of the **copy numbers** of the marker gene (16S, ITS or 18S) in order to normalize the OTU abundances table thereafter.
- Prediction of the **functions abundances**, using different databases:

EC : <https://enzyme.expasy.org/>

KO : <https://www.genome.jp/kegg/ko.html>

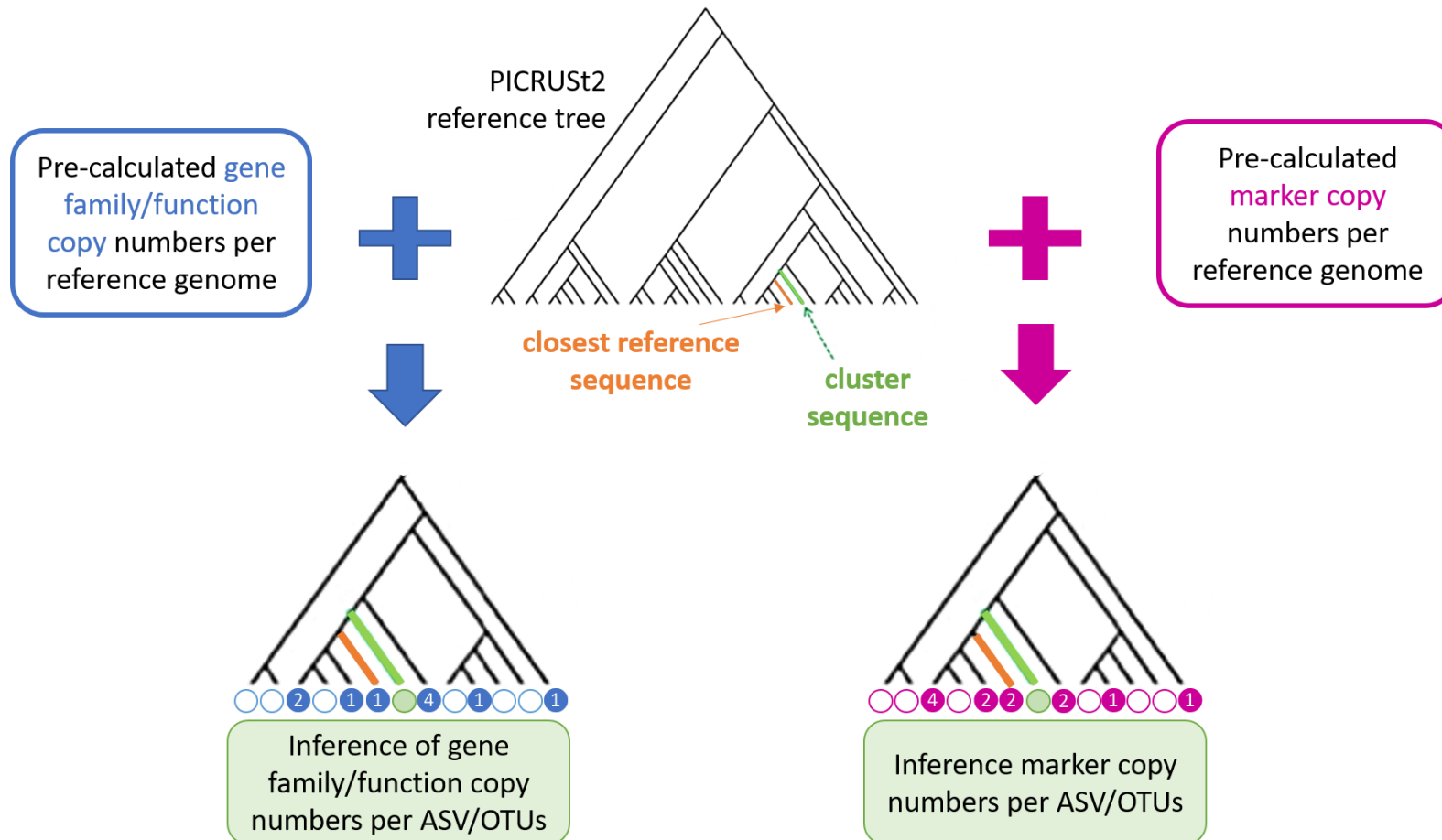
PFAM : <http://pfam.xfam.org/>

COG : <https://www.ncbi.nlm.nih.gov/research/cog-project/>

TIGRFAM : <https://tigrfams.jcvi.org/cgi-bin/index.cgi>

PHENO : <https://phenodb.org/>

# FROGSFUNC\_step2\_copynumbers



# FROGSFUNC\_step2\_copynumbers

**FROGSFUNC\_step2\_copynumbers** Predicts number of marker and function copy number in each OTU. (Galaxy Version 4.0.0+galaxy1) Favorite Options

**Biom file**  
70: FROGSFUNC\_step1\_placeseqs: frogsfunc\_placeseqs.biom

The abundance file to analyse i.e. FROGSFUNC\_step1\_placeseqs tool output file (frogsfunc\_placeseqs.biom). (--input-biom)

**Tree file**  
66: FROGSFUNC\_step1\_placeseqs: frogsfunc\_placeseqs\_tree.nwk

The file contains the tree information from FROGSFUNC\_step1\_placeseqs tool (frogsfunc\_placeseqs\_tree.nwk). (--tree)

**Taxonomic marker**  
 16S  
 ITS  
 18S

Taxonomic marker of interest.

**Function table**  
 Select/Unselect all

EC  KO

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


**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)

**Email notification**  
 No










Send an email notification when the job completes.

## 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 for next FROGSFUNC tools, the information from Metacyc (EC) or KEGG (KO) are required.
-  For ITS and 18S markers, 'EC' is only available.

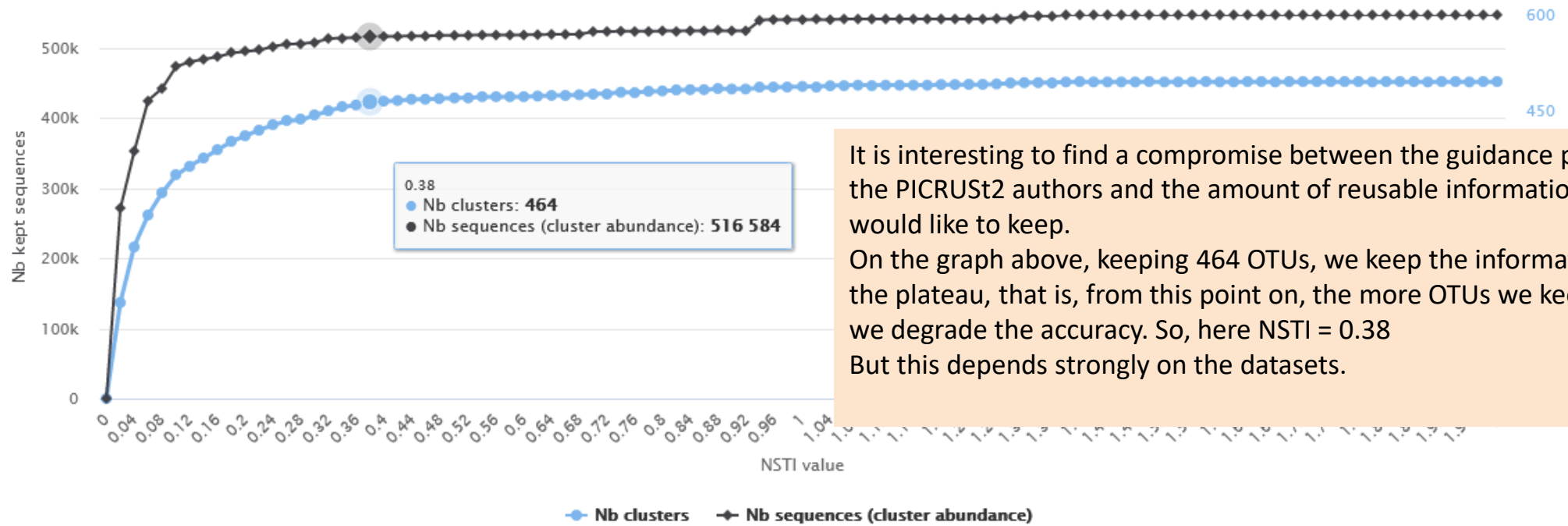
# Outputs

---

FROGSFUNC_step2_copynumbers: frogsfunc_copynumbers_marker.tsv	  
FROGSFUNC_step2_copynumbers: frogsfunc_copynumbers_predicted_functions.tsv	  
FROGSFUNC_step2_copynumbers: report.html	  

# FROGSFUNC\_step2\_copynumbers: report.html

## Cluster/sequence number according to NSTI threshold



It is interesting to find a compromise between the guidance provided by the PICRUSt2 authors and the amount of reusable information you would like to keep. On the graph above, keeping 464 OTUs, we keep the information before the plateau, that is, from this point on, the more OTUs we keep the more we degrade the accuracy. So, here NSTI = 0.38 But this depends strongly on the datasets.



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

# FROGSFUNC\_step2\_copynumbers: copynumber\_predicted\_functions.tsv

---

sequence	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
Cluster_1	2 0	2 0	0 0	0 0	0 0
Cluster_10	0 0	3 0	0 0	0 0	0 0
Cluster_100	3 0	8 0	0 1	0 0	0 0
Cluster_101	0 0	2 0	0 0	0 0	0 0
Cluster_102	0 0	3 0	0 0	1 0	0 0
Cluster_103	0 0	2 0	0 0	0 0	0 0
Cluster_104	0 0	2 0	0 0	0 0	0 0
Cluster_105	3 0	3 0	0 0	0 0	0 0
Cluster_106	1 0	4 0	0 0	0 0	0 0
Cluster_107	2 0	3 0	0 0	0 0	0 0
Cluster_108	3 0	7 0	0 0	0 0	1 0
Cluster_109	0 0	2 0	0 0	0 0	0 0
Cluster_11	5 0	5 0	0 1	0 0	0 0
Cluster_110	0 0	3 0	0 0	0 0	0 0
Cluster_111	9 0	10 0	0 0	0 0	0 0
Cluster_112	3 0	2 0	0 0	0 0	0 0
Cluster_113	4 0	4 0	0 0	0 0	0 0
Cluster_114	0 0	2 0	0 0	0 0	0 0



Required file for FROGSFUNC\_step3\_functions tool.



# FROGSFUNC\_step2\_copynumbers: copynumber\_predicted\_marker.tsv

---

sequence	16S_rRNA_Count	metadata_NSTI
Cluster_1	1	0.00891
Cluster_10	1	0.021849
Cluster_100	1	0.007370999999999999
Cluster_101	1	0.11353599999999998
Cluster_102	1	0.042736
Cluster_103	1	0.01323
Cluster_104	1	0.007503
Cluster_105	1	0.006364
Cluster_106	1	0.027795999999999998
Cluster_107	1	0.63742
Cluster_108	5	0.012044



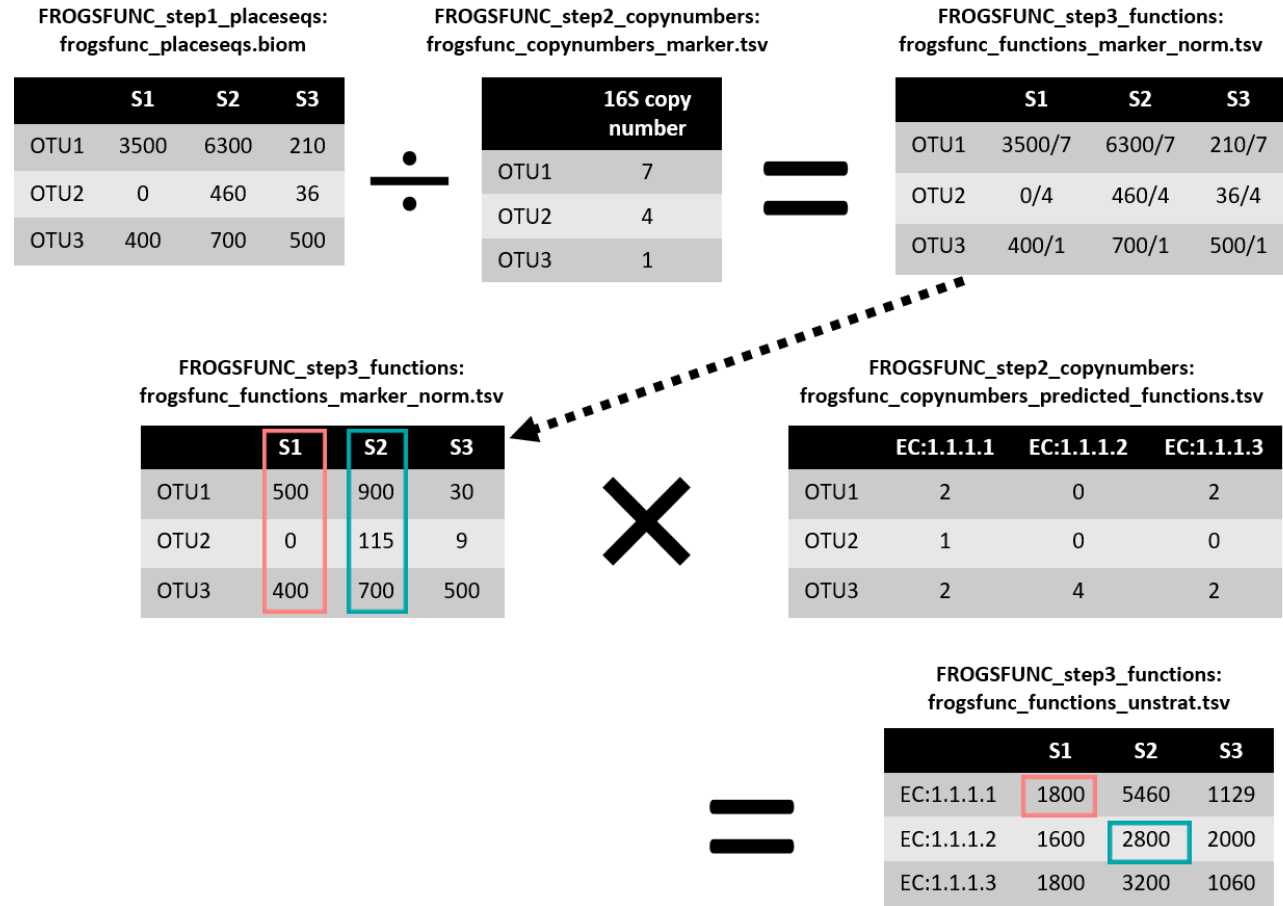
Required file for FROGSFUNC\_step3\_functions tool.

---

# FROGSFUNC\_step3\_functions

---

# FROGSFUNC\_step3\_functions



# FROGSFUNC\_step3\_functions

**FROGSFUNC\_step3\_functions** Calculates functions abundances in each sample. (Galaxy Version 4.0.0+galaxy1) ☆ Favorite ▼ Options

**Biom file**  
70: FROGSFUNC\_step1\_placeseqs: frogsfunc\_placeseqs.biom

The abundance file i.e. FROGSFUNC\_step1\_placeseqs tool output file (frogsfunc\_placeseqs.biom). (--input-biom)

**Function file**  
72: FROGSFUNC\_step2\_copynumbers: frogsfunc\_copynumbers\_predicted\_functions.tsv

Copy number table of functions present in the predicted genome for each OTU i.e. FROGSFUNC\_step2\_copynumbers tool output file (frogsfunc\_copynumbers\_predicted\_functions.tsv). (--function)

**Marker file**  
73: FROGSFUNC\_step2\_copynumbers: frogsfunc\_copynumbers\_marker.tsv

Table of predicted marker copy number i.e. FROGSFUNC\_step2\_copynumbers output (frogsfunc\_copynumbers\_marker.tsv). (--marker)

**NSTI cut-off**  
0,38

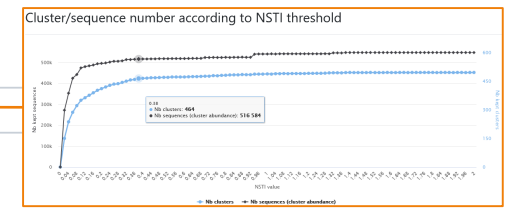
Any sequence with an NSTI above this threshold will be out. (default: 2) (--max-nsti)

**Email notification**  
 No

Send an email notification when the job completes.


















Please note that here Galaxy puts the same tsv file in both entries. Change the 1<sup>st</sup>.



This parameter can be set using FROGSFUNC\_step2\_copynumbers: report.html

# Outputs

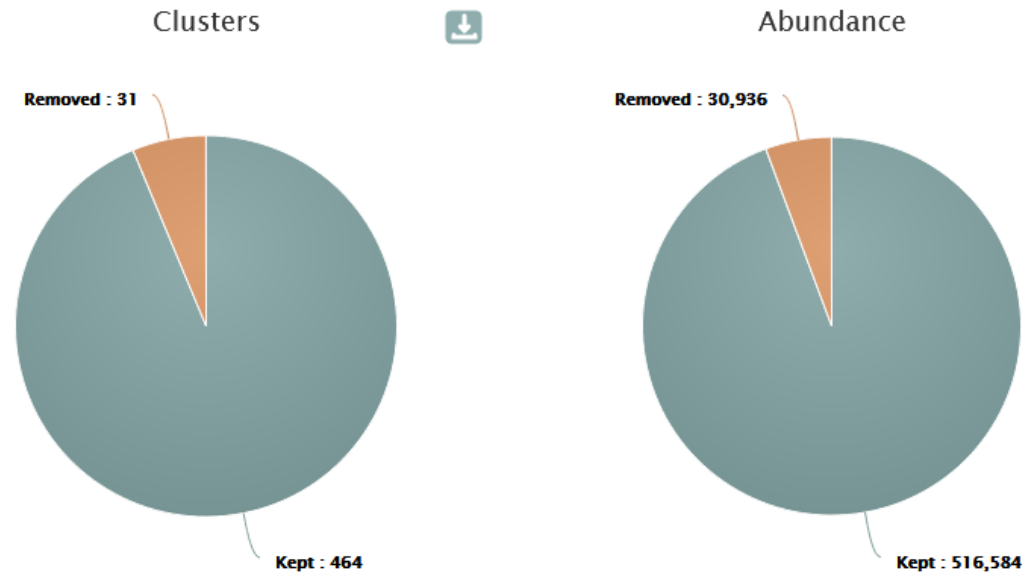
---

78: FROGSFUNC_step3_functions: frogsfunc_functions_unstrat.tsv	  
77: FROGSFUNC_step3_functions: frogsfunc_functions_excluded.tsv	  
76: FROGSFUNC_step3_functions: frogsfunc_functions_weighted_nsti.tsv	  
75: FROGSFUNC_step3_functions: frogsfunc_functions_marker_norm.tsv	  
74: FROGSFUNC_step3_functions: report.html	  

# FROGSFUNC\_step3\_functions: report.html

How many OTUs/sequences are kept after the process?

## Metagenome functional profile summary



OTUs are out if the NSTI associated is above the threshold. In this example all OTUs with a NSTI > 0.38. Here, only 464 OTUs are kept in the functional inference process.

# What is the distribution of gene/function abundances in the samples ?

Display global distribution ← Displays the distribution on all samples

Show 10 entries Search:

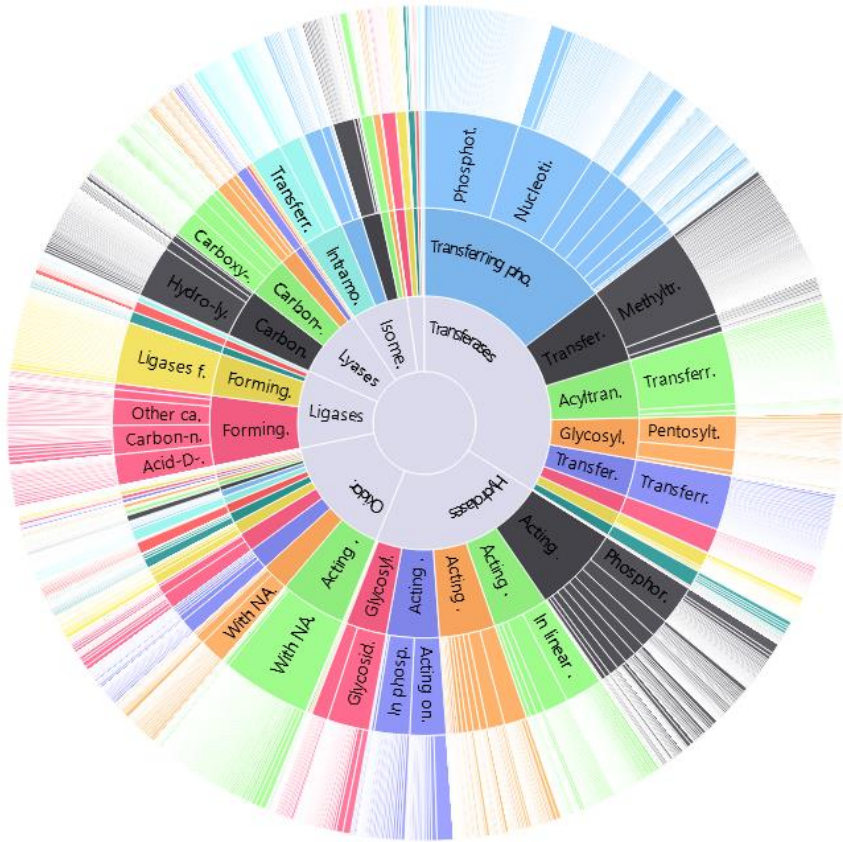
CSV

<input type="checkbox"/>	Samples	↑ Weighted NSTI	↑ Nb gene retrieved
<input checked="" type="checkbox"/>	BHT0.LOT01	0.043	4,490
<input checked="" type="checkbox"/>	BHT0.LOT03	0.06	4,633
<input checked="" type="checkbox"/>	BHT0.LOT04	0.042	4,830
<input checked="" type="checkbox"/>	BHT0.LOT05	0.049	4,740
<input checked="" type="checkbox"/>	BHT0.LOT06	0.064	4,675
<input checked="" type="checkbox"/>	BHT0.LOT07	0.048	4,662
<input checked="" type="checkbox"/>	BHT0.LOT08	0.06	4,830
<input checked="" type="checkbox"/>	BHT0.LOT10	0.039	4,676
<input type="checkbox"/>	CDT0.LOT02	0.032	4,293
<input type="checkbox"/>	CDT0.LOT04	0.044	4,457

Display distribution ← Displays the distribution on selected samples

# What is the distribution of gene/function abundances in the samples ?

---



The innermost circle represents the highest hierarchical level of gene families according to Metacyc or Kegg databases.

The more we go outwards, the more the hierarchical level becomes precise until indicating the identifier of the gene family.



# Output

## Function abundance table - unstratified.



Required file for FROGSFUNC\_step4\_pathways 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	<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 on it to understand the information.

# Output Function excluded

Cluster	FROGS_taxonomy	PICRUSt2_taxonomy	NSTI
Cluster_107	Bacteria;Firmicutes;Bacilli;Mycoplasmatales;Mycoplasmataceae;Candidatus Bacilloplasma;unknown species	Bacteria;Tenericutes;Mollicutes;Haloplasmatales;Haloplasmataceae;Haloplasma;Haloplasma_contractile	0.63742
Cluster_122	Bacteria;Firmicutes;Bacilli;Erysipelotrichales;Erysipelotrichaceae;Faecalitalea;Multi-affiliation	Bacteria;Firmicutes;Bacilli;Lactobacillales;Enterococcaceae;Enterococcus;Enterococcus_pallens	0.679824
Cluster_124	Bacteria;Firmicutes;Clostridia;Christensenellales;Christensenellaceae;Christensenellaceae R-7 group;unknown species	Bacteria;Firmicutes;Clostridia;Eubacteriales;Christensenellaceae;Christensenella;Christensenella_minuta	0.420828
Cluster_144	Bacteria;Firmicutes;Clostridia;Clostridia UCG-014;unknown family;unknown genus;gut metagenome	Bacteria;Firmicutes;Clostridia;Eubacteriales;Clostridiaceae;Butyrivibrio;Butyrivibrio_pullicaeorum	0.726233
Cluster_152	Bacteria;Spirochaetota;Spirochaetia;Spirochaetales;Spirochaetaceae;unknown genus;unknown species	Bacteria;Spirochaetes;Spirochaetia;Spirochaetales;Spirochaetaceae;Treponema;Treponema_caldarium	0.53047
Cluster_16	Bacteria;Fusobacteriota;Fusobacteriia;Fusobacteriales;Leptotrichiaceae;Hyponocyclicus;unknown species	Bacteria;Firmicutes;Clostridia;Eubacteriales;unclassified;unclassified;Clostridiales_bacterium_Firm_08	1.3093270000000001
Cluster_166	Bacteria;Bacteroidota;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenellaceae RC9 gut group;unknown species	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Odoribacteraceae;Butyrivibrio;Butyrivibrio_virosa	0.959354
Cluster_172	Bacteria;Actinobacteriota;Actinobacteria;Actinomycetales;Actinomycetaceae;Actinomyces;Actinomyces viscosus	Bacteria;Actinobacteria;Actinomycetia;Micromonosporales;Micromonosporaceae;Asanoa;Asanoa_ishikariensis	0.48663500000000004
Cluster_178	Bacteria;Patescibacteria;Saccharimonadia;Saccharimonadales;Saccharimonadaceae;Candidatus Saccharimonas;unknown species	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Acidovorax;Acidovorax_citrullii	0.79263500000000001
Cluster_189	Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Weeksellaceae;unknown genus;unknown species	Bacteria;Bacteroidetes;Flavobacteriia;Flavobacteriales;Weeksellaceae;Apibacter;Apibacter_mensalis	0.381465
Cluster_204	Bacteria;Fusobacteriota;Fusobacteriia;Fusobacteriales;Fusobacteriaceae;Fusobacterium;Multi-affiliation	Bacteria;Firmicutes;Negativicutes;Veillonellales;Veillonellaceae;Veillonella;Veillonella_magna	1.281097
Cluster_234	Bacteria;Bacteroidota;Bacteroidia;Bacteroidales;Prevotellaceae;Alloprevotella;unknown species	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Alloprevotella;Alloprevotella_rava	0.6157819999999999
Cluster_25	Bacteria;Firmicutes;Bacilli;Mycoplasmatales;Mycoplasmataceae;Candidatus Bacilloplasma;unknown species	Bacteria;Firmicutes;Clostridia;Eubacteriales;unclassified;unclassified;Clostridiales_bacterium_VE202-01	0.694566
Cluster_252	Bacteria;Actinobacteriota;Coriobacteriia;Coriobacteriales;Atopobiaceae;Olsenella;Olsenella umbonata	Bacteria;Actinobacteria;Actinomycetia;Nakamurellales;Nakamurellaceae;Nakamurella;Nakamurella_multipartita	0.7311479999999999
Cluster_266	Bacteria;Patescibacteria;Gracilibacteria;Candidatus Peribacteria;unknown family;unknown genus;metagenome	Bacteria;Proteobacteria;Epsilonproteobacteria;Nautiliales;Nitratiruptoraceae;Nitratiruptor;Nitratiruptor_sp.	1.027285
Cluster_355	Bacteria;Actinobacteriota;Acidimicrobiia;Microtrichales;Ilumatobacteraceae;unknown genus;unknown species	Bacteria;Firmicutes;Clostridia;Eubacteriales;Syntrophomonadaceae;Dethiobacter;Dethiobacter_alkaliphilus	0.876819
Cluster_358	Bacteria;Actinobacteriota;Coriobacteriia;Coriobacteriales;Atopobiaceae;Olsenella;unknown species	Bacteria;Actinobacteria;Actinomycetia;Streptomycetales;Streptomycetaceae;Streptomyces;Streptomyces_sp.	0.835499
Cluster_373	Bacteria;Actinobacteriota;Actinobacteria;Actinomycetales;Actinomycetaceae;Actinomyces;Actinomyces sp.	Bacteria;Actinobacteria;Actinomycetia;Propionibacteriales;Propionibacteriaceae;Cutibacterium;Propionibacterium_humerusii	0.539983
Cluster_39			
Cluster_40			
Cluster_41			
Cluster_42			
Cluster_43			
Cluster_438	Bacteria;Actinobacteriota;Actinobacteria;Micrococcales;Microbacteriaceae;Microbacterium;Microbacterium phyllosphaerae	Bacteria;Actinobacteria;Actinomycetia;Propionibacteriales;Propionibacteriaceae;Micrococcus;Micrococcus_sagamiensis	0.41153599999999996
Cluster_453	Bacteria;Firmicutes;Bacilli;Mycoplasmatales;Mycoplasmataceae;unknown genus;unknown species	Bacteria;Firmicutes;Clostridia;Eubacteriales;unclassified;unclassified;Clostridiales_bacterium_VE202-01	0.713615
Cluster_46	Bacteria;Fusobacteriota;Fusobacteriia;Fusobacteriales;Fusobacteriaceae;Cetobacterium;unknown species	Bacteria;Firmicutes;Clostridia;Eubacteriales;unclassified;unclassified;Clostridiales_bacterium_Firm_08	1.377319
Cluster_47	Bacteria;Bacteroidota;Bacteroidia;Bacteroidales;Rikenellaceae;Rikenellaceae RC9 gut group;unknown species	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes;Alistipes_inops	1.044139
Cluster_496	Bacteria;Firmicutes;Clostridia;Peptostreptococcales-Tissierellales;Peptostreptococcaceae;Proteocatella;Proteocatella sphenisci	Bacteria;Firmicutes;Clostridia;Eubacteriales;Peptostreptococcaceae;Paenoclostridium;Paenoclostridium_sordellii	0.432676
Cluster_7	Bacteria;Firmicutes;Bacilli;Erysipelotrichales;Erysipelotrichaceae;ZOR0006;unknown species	Bacteria;Firmicutes;Clostridia;Eubacteriales;Christensenellaceae;Christensenella;Christensenella_minuta	0.925652
Cluster_77	Bacteria;Actinobacteriota;Actinobacteria;Actinomycetales;Actinomycetaceae;Trueperella;Multi-affiliation	Bacteria;Actinobacteria;Actinomycetia;Streptomycetales;Streptomycetaceae;Streptomyces;Candidatus_Streptomyces_massiliensis	0.467055
Cluster_97	Bacteria;Fusobacteriota;Fusobacteriia;Fusobacteriales;Fusobacteriaceae;Cetobacterium;unknown species	Bacteria;Firmicutes;Clostridia;Eubacteriales;unclassified;unclassified;Clostridiales_bacterium_Firm_08	1.273517

Information (FROGS taxonomy, PICRUSt2 taxonomy, NSTI) about removed sequences that have a NSTI value above the NSTI threshold chosen in this step.

# Output functions\_marker\_norm

---

normalized	BHT0.LOT01	BHT0.LOT03	BHT0.LOT04	BHT0.LOT05	BHT0.LOT06	BHT0.LOT07
Cluster_1	791.0	402.0	433.0	911.0	1232.0	653.0
Cluster_10	0.0	0.0	0.0	16.0	6.0	0.0
Cluster_100	0.0	0.0	0.0	0.0	0.0	0.0
Cluster_101	0.0	0.0	0.0	0.0	0.0	0.0
Cluster_102	0.0	0.0	0.0	0.0	0.0	2.0
Cluster_103	0.0	0.0	0.0	0.0	0.0	0.0
Cluster_104	0.0	0.0	0.0	0.0	1.0	0.0
Cluster_105	0.0	0.0	0.0	0.0	1.0	3.0
Cluster_106	0.0	0.0	0.0	0.0	0.0	0.0
Cluster_108	0.0	0.0	0.0	1.4	0.2	0.4
Cluster_109	0.0	0.0	0.0	0.0	0.0	0.0
Cluster_11	3.0	0.0	4.0	11.0	5.0	0.0

16S/ITS/18S abundance table normalized by the number of markers per genome.

---

# FROGSFUNC\_step4\_pathways

---

# FROGSFUNC\_step4\_pathways

**FROGSFUNC\_step4\_pathways** is the fourth step of PICRUSt2.

It infers MetaCyc/KEGG pathway abundances based on EC/KO number abundances.

There are 3 steps performed at this stage:

- 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\_step3\_functions:  
frogsfunc\_functions\_unstrat.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\_step4\_pathways: frogsfunc\_pathways\_unstrat.tsv

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\_step4\_pathways

FROGSFUNC\_step4\_pathways Calculates pathway abundances in each sample. (Galaxy Version 4.0.0+galaxy1) ☆ Favorite ▼ Options

Function abundance file

78: FROGSFUNC\_step3\_functions: frogsfunc\_functions\_unstrat.tsv 📁

TSV function abundances table from FROGSFUNC\_step3\_function tool, frogsfunc\_functions\_unstrat.tsv (unstratified table). (--input-file)

Taxonomic marker

16S  
 ITS  
 18S

Taxonomic marker of interest.

Pathway reference

Metacyc ←  
 Kegg

For 16S marker, choose Metacyc or KEGG in accordance with your choice in the FROGSFUNC\_step2\_copynumbers tool. For ITS or 18S marker, Metacyc is the on

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



If this option is set, the pathway abundances file (frogsfunc\_functions\_unstrat.tsv) is normalized: 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

---

80: FROGSFUNC\_step4\_pathways: frogsfunc\_pathways\_unstrat.tsv



79: FROGSFUNC\_step4\_pathways: report.html



# FROGSFUNC\_step4\_pathways: report.html

Pathway abundances per sample

← Displays the pathway distribution on all samples

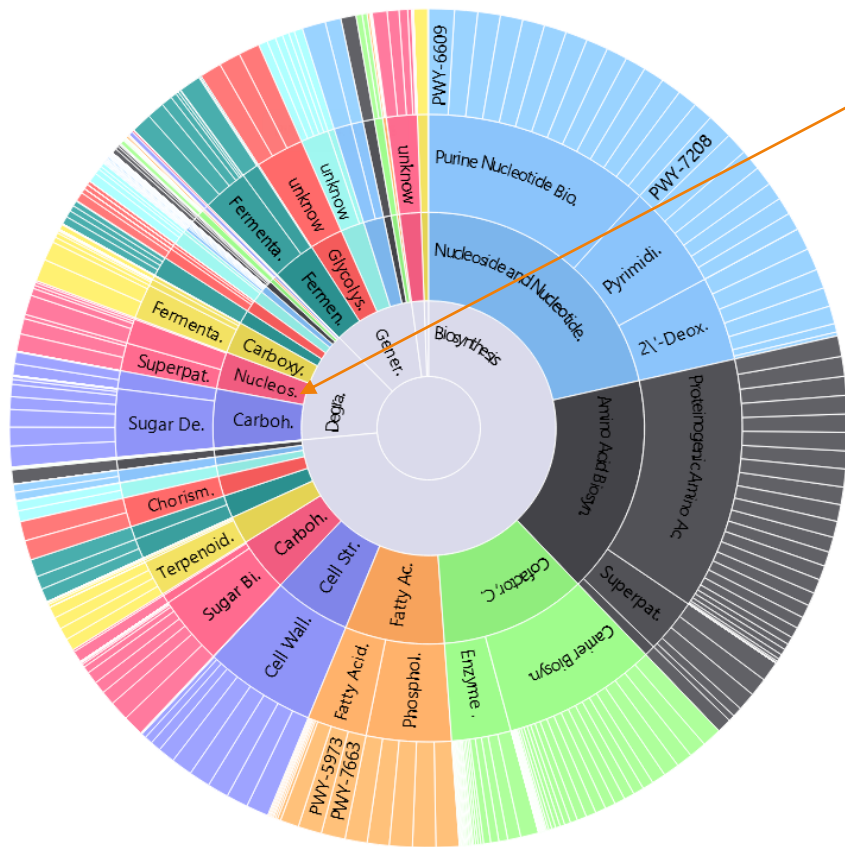
Show  entries  Search:

<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

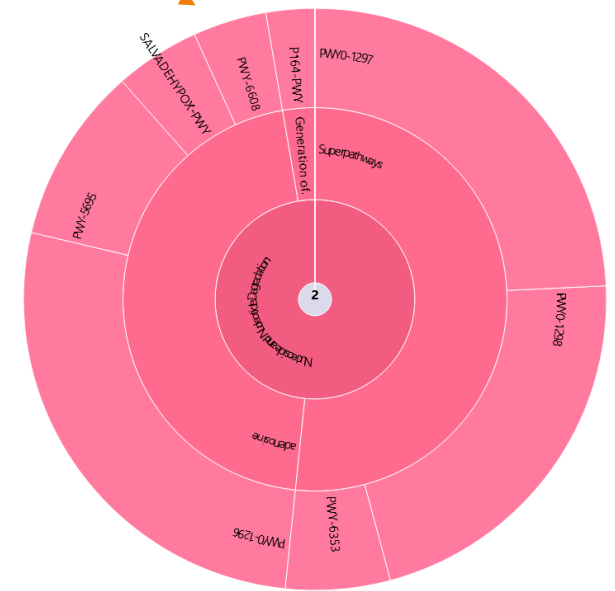
← Displays the pathway distribution on selected samples



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



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



# Output 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.