

HOWTO : Du BIOM issu de FROGS à la prédiction métabolique et fonctionnelle Tax4Fun

Le BIOM sortant de l'étape 4 de la pipeline FROGS (Outil « FROGS Affiliation OTU ») contient la table d'abondance des OTUs et leurs taxonomies. A partir de ces informations, une prédiction métabolique et fonctionnelle peut être réalisée avec l'outil «Function Table for Tax4Fun matrix (Galaxy Version 1.0.0)».

Trois remarques :

- La base de données utilisée pour l'outil «Function Table for Tax4Fun matrix (Galaxy Version 1.0.0)» est Silva version 123.
- L'équipe FROGS recommande une affiliation taxonomique avec blastn, ce qui génère des multi-affiliations. Les OTUs dont la taxonomie contient un terme « multi-affiliation » ne trouveront pas de correspondance avec les organismes KEGG.
- L'outil «Function Table for Tax4Fun matrix (Galaxy Version 1.0.0)» prend en entrée un objet phyloseq de format rdata.

4 étapes intermédiaires doivent être effectuées avant l'utilisation de l'outil «Function Table for Tax4Fun matrix (Galaxy Version 1.0.0)», et permettent de :

- convertir le fichier de format biom en format tabular,
- remplacer les termes « multi-affiliation » par des chaînes vides,
- convertir le fichier de format tabular en format biom,
- construire un objet phyloseq de format rdata à partir du fichier de format biom.

Afin de réaliser ces étapes rapidement à la suite du pipeline FROGS, le workflow « Function Table for Tax4Fun matrix » est disponible au format .ga



Hélène BILLARD, UMR 1280 PhAN, Inra / Université, Nantes.

auprès du support Galaxy Sigenae / Bioinfo Genotoul :
support.sigenae@inra.fr.

Vous pourrez l'importer dans votre espace Galaxy, vérifier les paramètres des outils détaillés ci-dessous, et l'exécuter sur vos données.

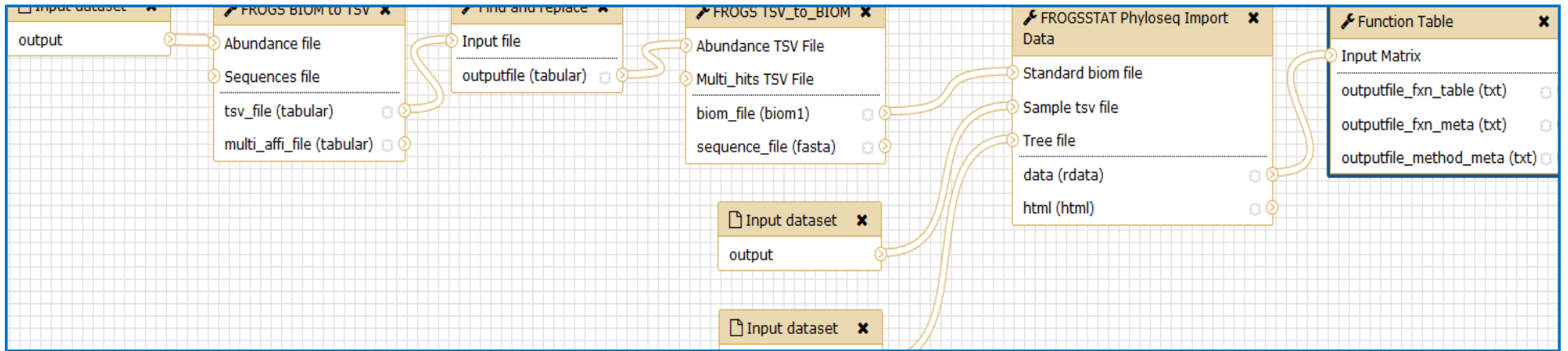


Figure 1: Workflow "Function Table for Tax4Fun matrix".

1^{ère} étape : convertir le fichier de format biom en format tabular.

FROGS BIOM to TSV Converts a BIOM file in TSV file. Options
(Galaxy Version 2.1.0)

Abundance file
[File icon] [Copy icon] [Folder icon] 76: FROGS Affiliation OTU: affiliation.biom
The BIOM file to convert (format: BIOM).

Sequences file
[File icon] [Copy icon] [Folder icon] Nothing selected
The sequences file (format: fasta). If you use this option the sequences will be add in TSV.

Extract multi-alignments
 Yes No
If you have used FROGS affiliation on your data, you can extract information about multiple alignements in a second TSV.

Figure 2: Formulaire de l'outil "FROGS BIOM to TSV (Galaxy Version 2.1.0)".

2^{ème} étape : remplacer les termes « multi-affiliation » par des chaînes vides.

Find and replace (sed command) (Galaxy Version 1.0.0) Options

Input file
[File icon] [Copy icon] [Folder icon] 2: FROGS BIOM to TSV: abundance.tsv

Item to find
Multi-affiliation

Replace item

Figure 3: Formulaire de l'outil "Find and replace (Galaxy Version 1.0.0)".

3^{ème} étape : convertir le fichier de format tabular en format biom.

FROGS TSV_to_BIOM Converts a TSV file in a BIOM file. (Galaxy Version 2.0.0)

Abundance TSV File
7: Replacedfile
Your FROGS abundance TSV file. Take care to keep original column names.

Multi_hits TSV File
Nothing selected
TSV file describing multi_hit blast results.

Extract seeds in FASTA file
Yes No
If there is a 'seed_sequence' column in your TSV table, you can extract seed sequences in a separated FASTA file.

Execute

Figure 4: Formulaire de l'outil "FROGS TSV_to_BIOM (Galaxy Version 2.0.0)".

4^{ème} étape : construire un objet phyloseq de format rdata à partir du fichier de format biom.

FROGSSTAT Phyloseq Import Data from 3 files: biomfile, samplefile, treefile (Galaxy Version 1.0.1)

Standard biom file
96: FROGS TSV_to_BIOM: abundance.biom
The file contains the OTU informations (format: biom1).

Sample tsv file
108: samples.tabular
The file contains the samples informations (format: tabular).

Tree file
102: FROGS_Tree.nhx
The file contains the tree informations (format: Newick - nhx or nwk).

Names of taxonomics levels
Kingdom Phylum Class Order Family Genus Species
The ordered taxonomic levels stored in BIOM. Each level is separated by one space.

Do you want to normalise your data ?
Yes No
To normalise data before analysis.

Execute

Figure 5: Formulaire de l'outil "FROGSSTAT Phyloseq Import Data (Galaxy Version 1.0.1)".

A l'issue de ces 4 étapes, l'outil «Function Table for Tax4Fun matrix (Galaxy Version 1.0.0)» peut être utilisé :

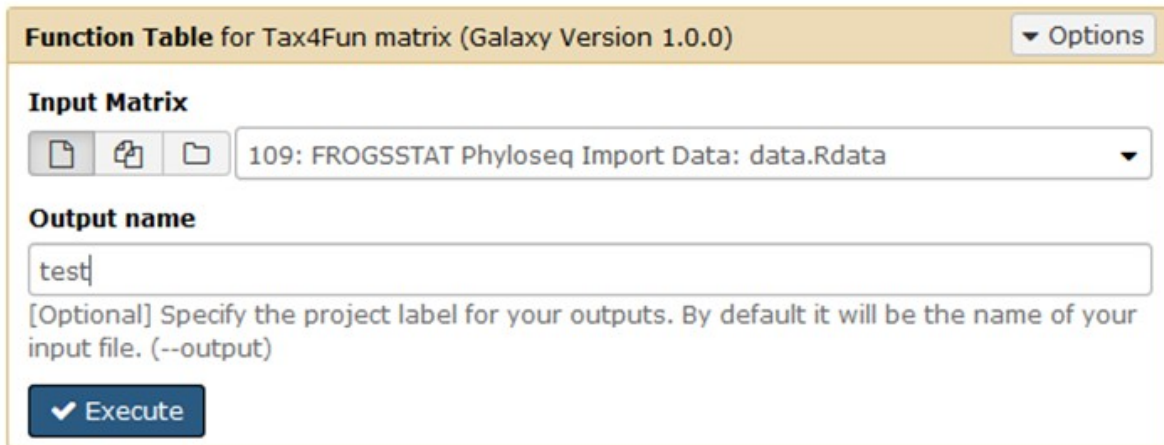


Figure 6: Formulaire de l'outil "Function Table (Galaxy Version 1.0.0)".

Une description des fichiers en entrée et en sortie, ainsi que les références avec des liens vers le package R « themetagenomics » et le manuel sont disponibles sur le formulaire de l'outil.

Les arguments, décrits dans le manuel, sont les suivants:

```
t4f(otu_table=ABUND,rows_are_taxa=FALSE,tax_table=TAX,reference_path=tmp,type='uproc',short=TRUE,cn_normalize=TRUE,sample_normalize=TRUE,drop=TRUE)
```

Les en-têtes des colonnes des tables en sortie de l'outil sont décalés d'une colonne vers la gauche.