



HAL
open science

FROGSFUNC: Smart integration of PICRUSt2 software into FROGS pipeline

Vincent Darbot, Moussa Samb, Maria Bernard, Olivier Rué, Géraldine Pascal

► **To cite this version:**

Vincent Darbot, Moussa Samb, Maria Bernard, Olivier Rué, Géraldine Pascal. FROGSFUNC: Smart integration of PICRUSt2 software into FROGS pipeline. JOBIM 2022, Jul 2022, Rennes, France. . hal-03806133

HAL Id: hal-03806133

<https://hal.inrae.fr/hal-03806133>

Submitted on 7 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



FROGSFUNC:

Smart integration of PICRUSt2 software into FROGS pipeline

Vincent DARBOT¹, Moussa SAMB¹, Maria BERNARD², Olivier RUÉ³ and Géraldine PASCAL¹

¹GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet Tolosan, France. ²Univ. Paris-Saclay, INRAE, AgroParisTech, GABI, SIGENAE, F-78352, Jouy-en-Josas, France. ³Univ. Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, Jouy-en-Josas, France
Corresponding Author: vincent.darbot@inrae.fr



Background

Metabarcoding is the large-scale taxonomic identification of complex environmental samples via analysis of DNA reads of one marker gene (16S, ITS, 18S, COI, etc.).

The aim of metabarcoding analysis is to provide a table of abundance of OTUs/ASVs, as close as possible to the species, per sample as well as a descriptive statistical analysis of the composition of the targeted microbial population of the samples.

The various tools developed within FROGS [1,2] offers such results. They allow users to process their data in command lines or *via* in a user-friendly Galaxy [3] interface and to obtain different graphical and descriptive outputs.

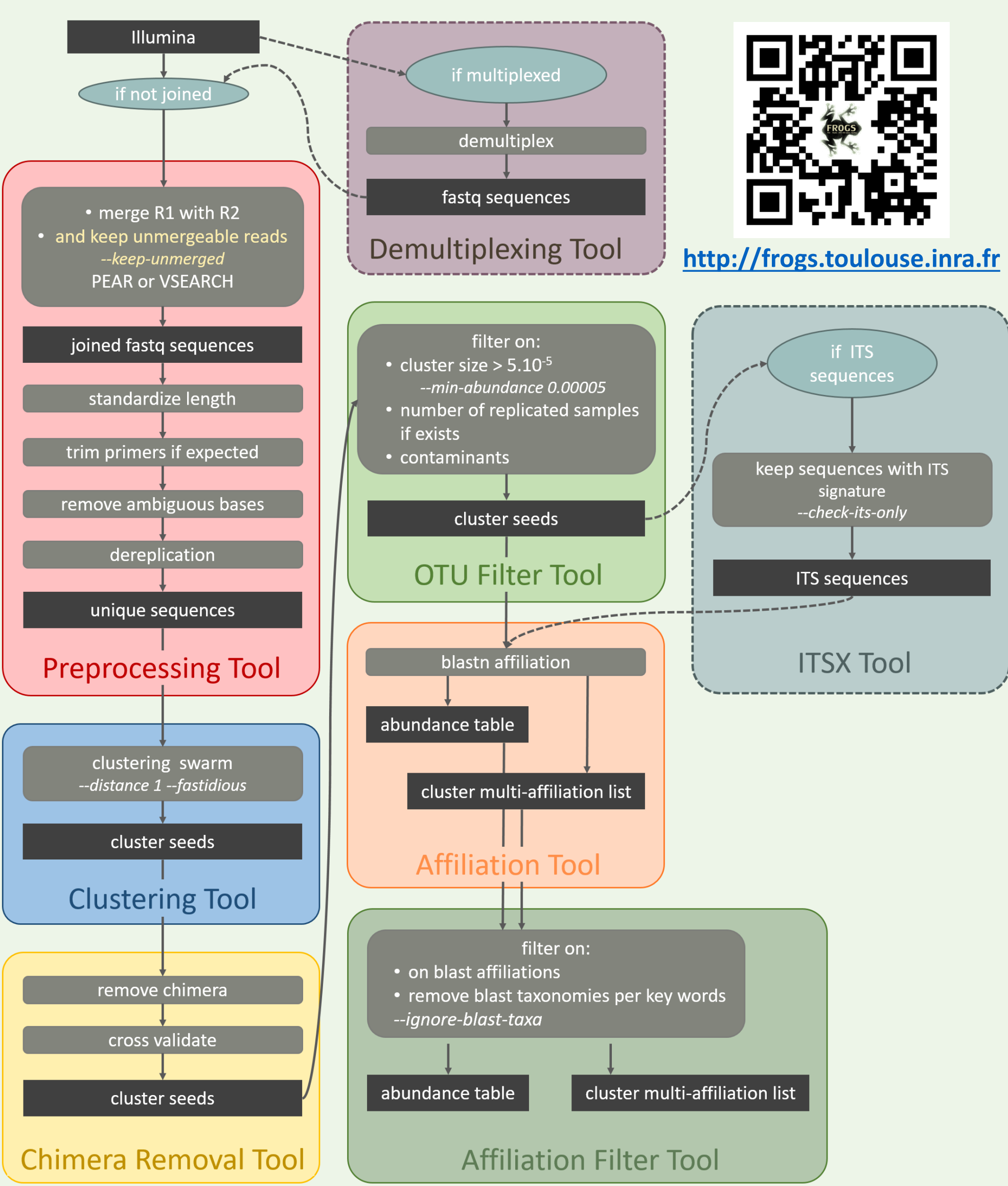
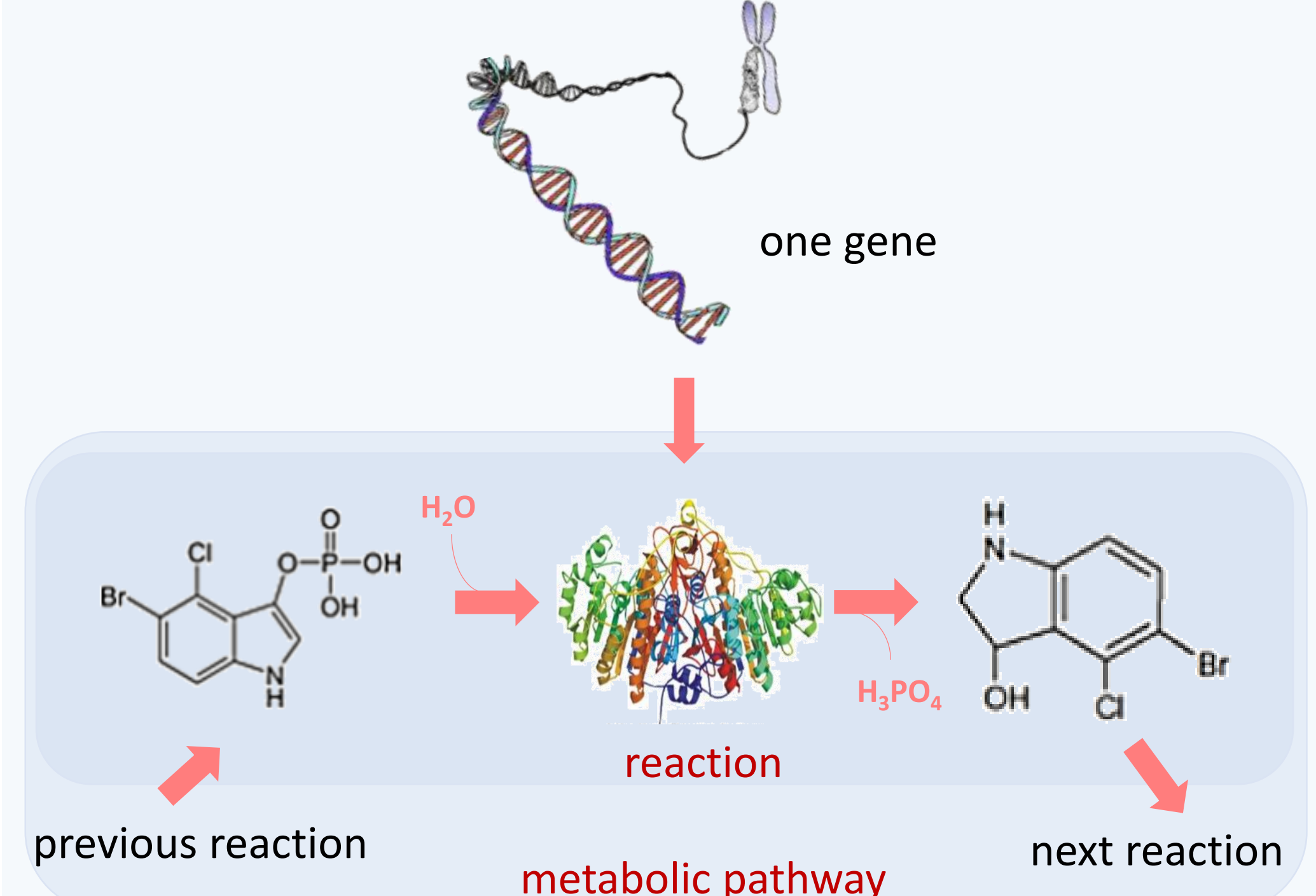
Unlike metagenomics [4], metabarcoding does not provide these functional profiles of a microbial population, by being restricted to one marker gene.

PICRUSt2 [5] bypass this restriction and obtain a prediction of the functional potential of a sample, at low cost.

Firstly, PICRUSt2 placed the marker gene (16S, ITS or 18S) sequences of interest into a its reference tree, that is used as the basis of functional predictions. After, it predicts number of marker and function copy number in each OTU. Then, for each sample, it calculates functions abundances and finally, pathway abundances are inferred, based on functional profile.

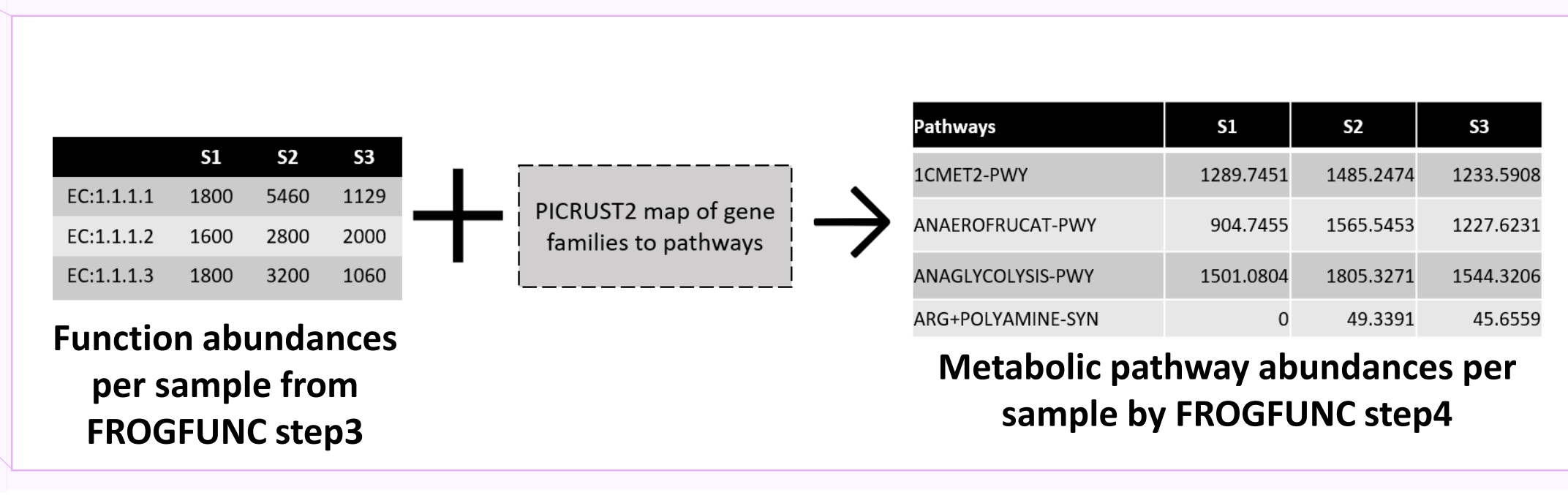
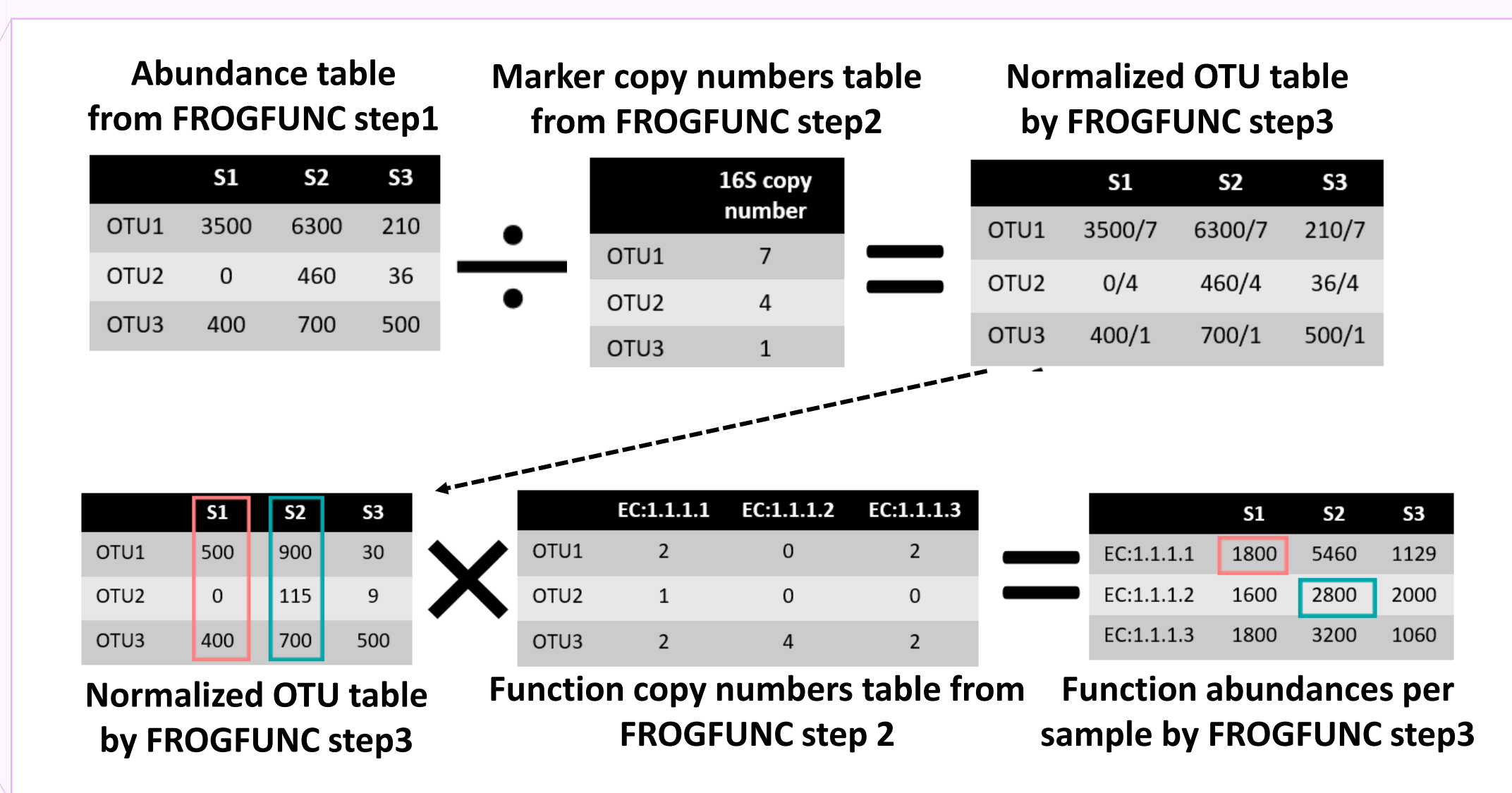
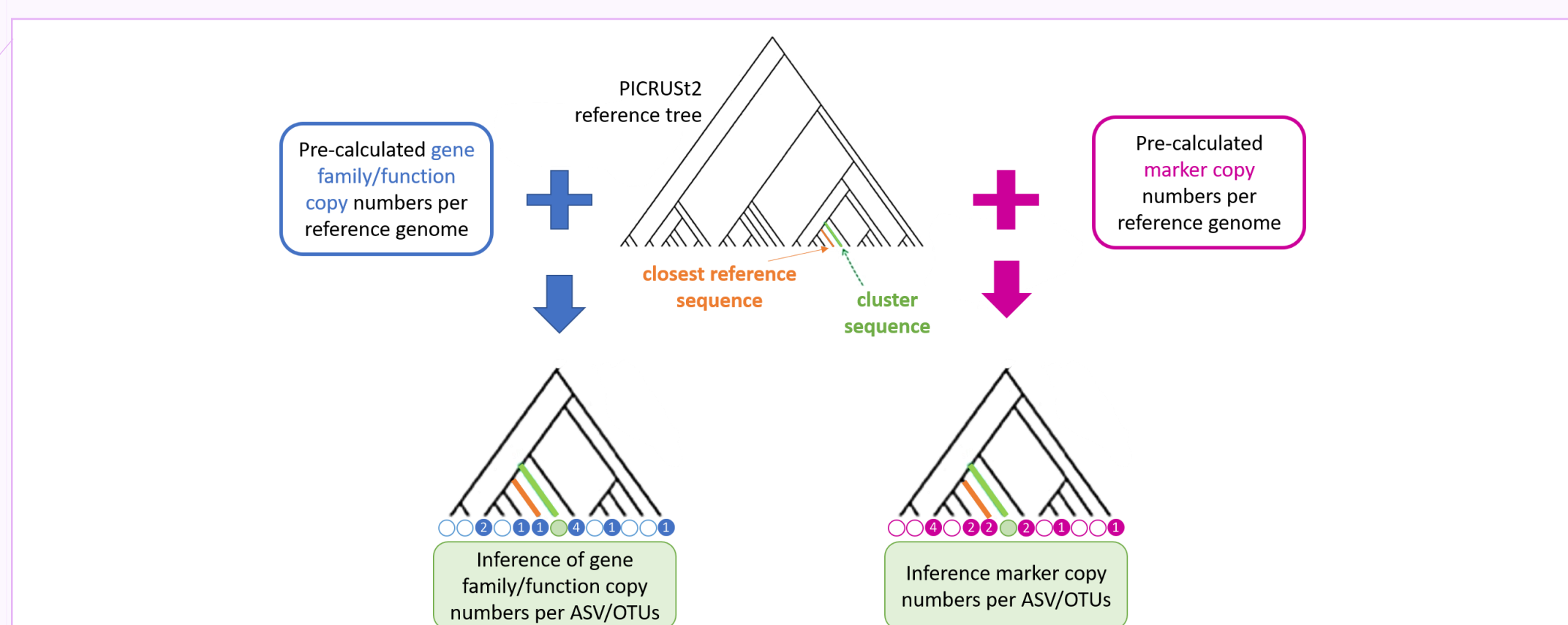
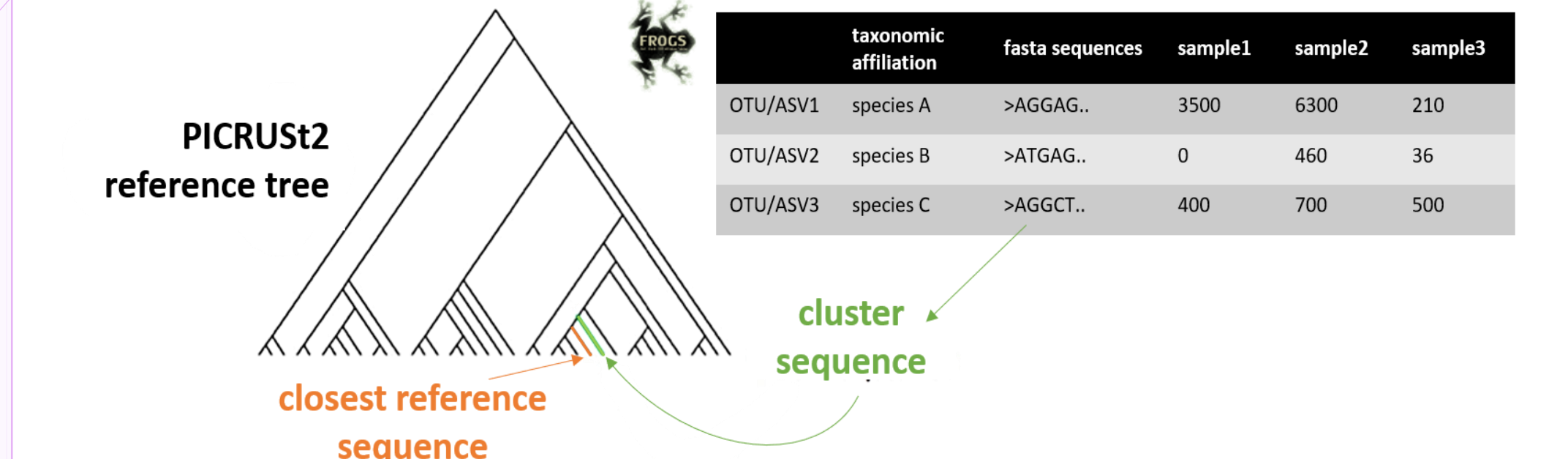
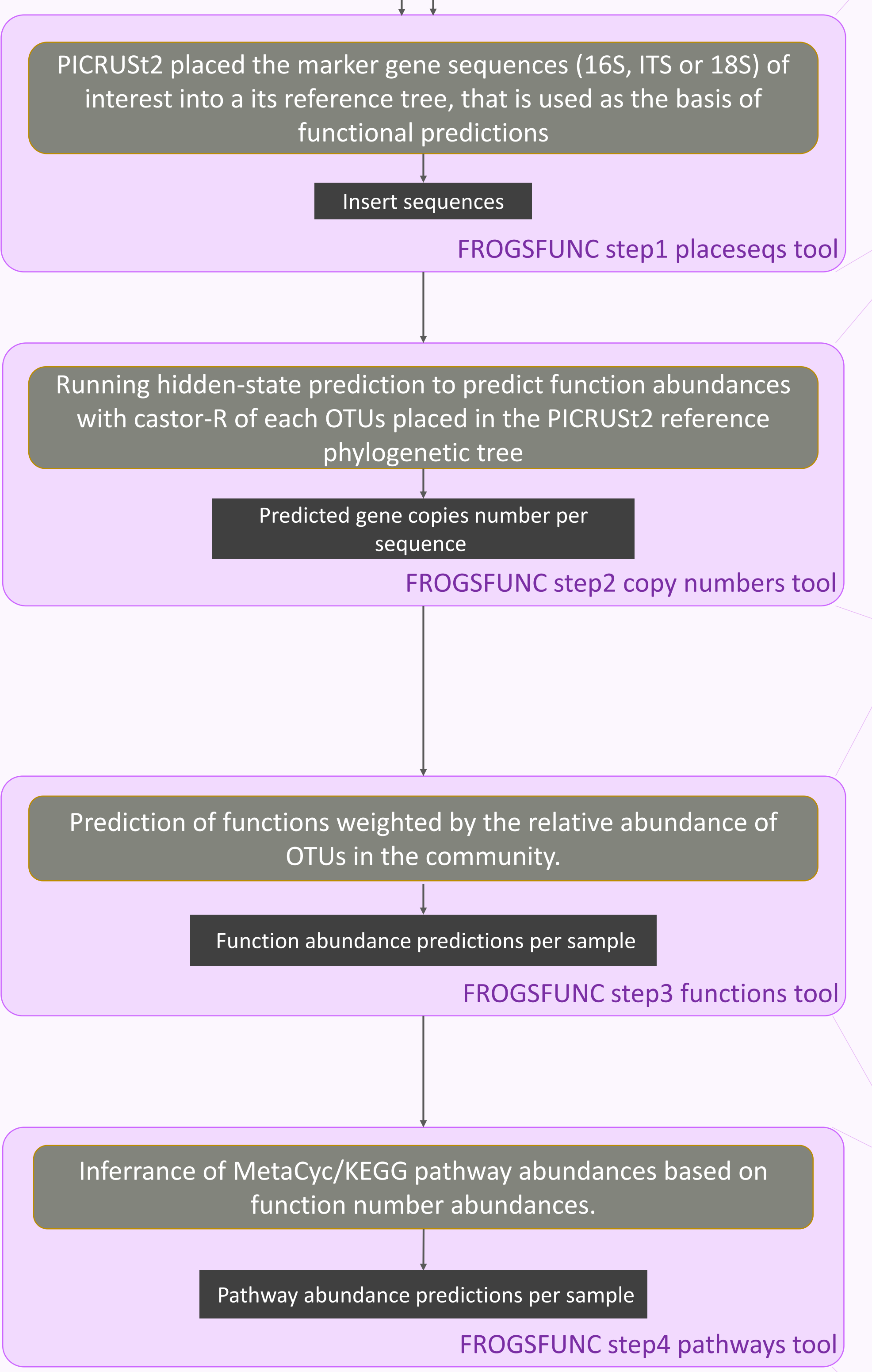
What metabolic functions are presents in a microbial community ?

Functional inference: we assume that an organism has a metabolic function by the presence in the organism's genome of a known sequence having that function.

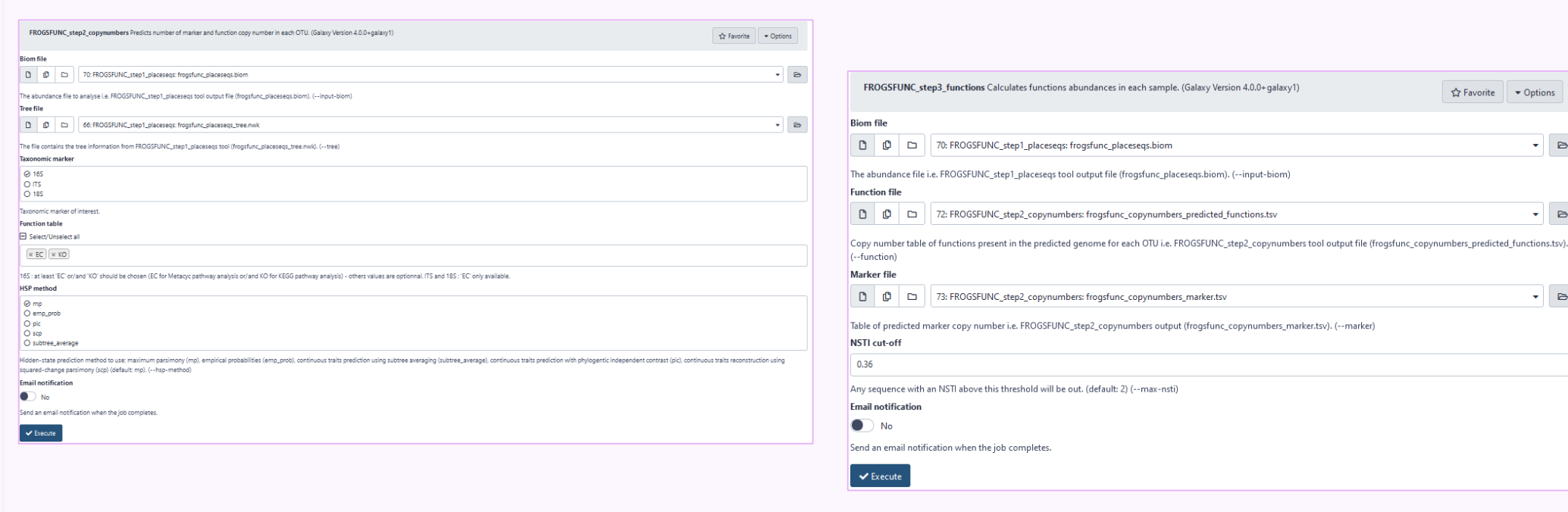


Functional profiling within FROGS

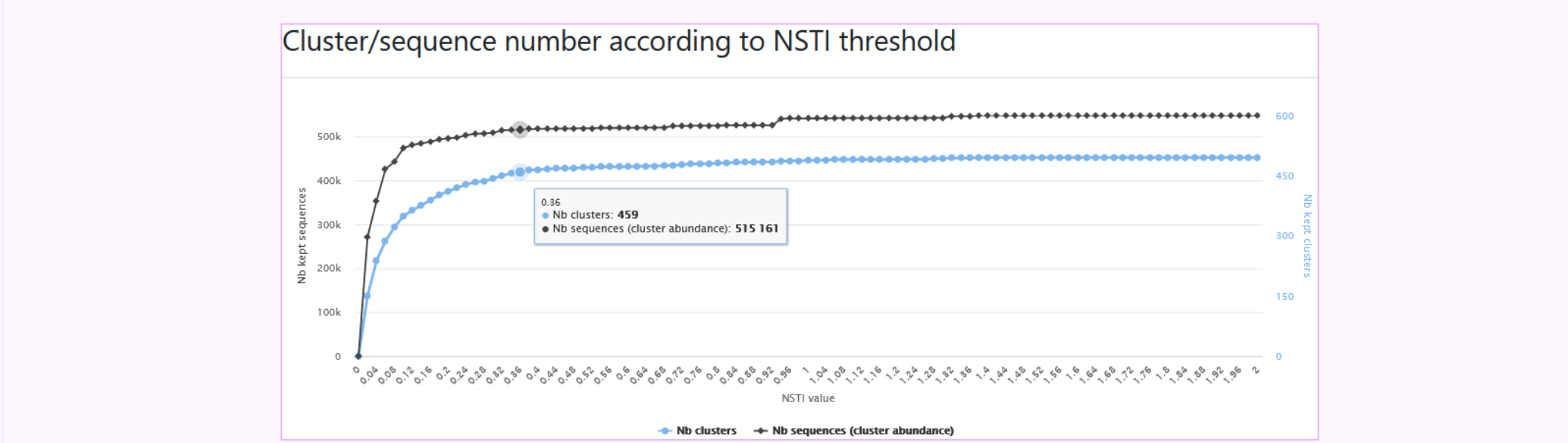
Workflow



Ability to use Galaxy interfaces



Various tables and graphical outputs are displayed to make the experience intuitive



FROGS highlights important information that are very useful for data interpretation. Some "hidden" PICRUSt2 outputs are exploited: reporting incongruence between taxonomic affiliations, Nearest Sequenced Taxon Index (NSTI) threshold confidence indicator, decision support graphic to help choosing the NSTI threshold



References

[1] Escudé F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet G, Combes S, Pascal G. FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics*. 2018 Apr 15;34(8):1287-1294. doi: 10.1093/bioinformatics/btx791. PMID: 29228191.
[2] Bernard M, Rué, O, Mariadassou M, and Pascal G; FROGS: a powerful tool to analyse the diversity of fungi with special management of internal transcribed spacers. *Briefings in Bioinformatics*. 2021, Nov. doi : 10.1093/bib/bbab318
[3] Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltmann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*. 2018 Jul 2;46(W1):W537-W544. doi: 10.1093/nar/gky379. PMID: 29790989; PMCID: PMC6030816.
[4] Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp*. 2012 Feb 9;2(1):3. doi: 10.1186/2042-5783-2-3. PMID: 22587947; PMCID: PMC3351745.
[5] Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020 Jun;38(6):685-688. doi: 10.1038/s41587-020-0548-6. PMID: 32483366; PMCID: PMC7365738.