



HAL
open science

FROGS: Find Rapidly OTUs with Galaxy Solution

Frédéric Escudie, Lucas Auer, Maria Bernard, Laurent Cauquil, Katia Vidal,
Sarah Maman, Mahendra Mariadassou, Guillermina Hernandez Raquet,
Géraldine Pascal

► **To cite this version:**

Frédéric Escudie, Lucas Auer, Maria Bernard, Laurent Cauquil, Katia Vidal, et al.. FROGS: Find Rapidly OTUs with Galaxy Solution. Colloque d'Écologie Microbienne organisé par l'Association Française d'Écologie Microbienne (AFEM), Nov 2015, Anglet, France. Oxford University Press, 34, 7 p., 2015, Bioinformatics. hal-01886389

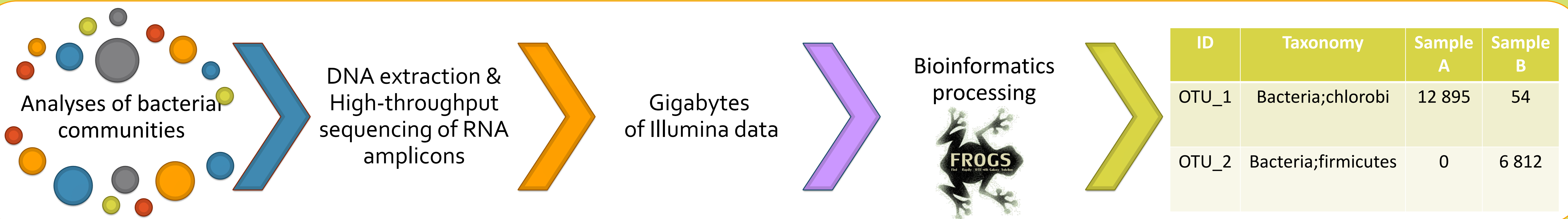
HAL Id: hal-01886389

<https://hal.science/hal-01886389v1>

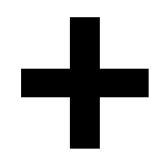
Submitted on 3 Jun 2020

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.

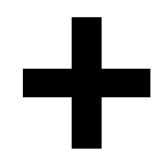
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



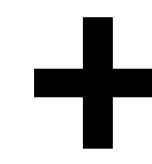
User-friendly



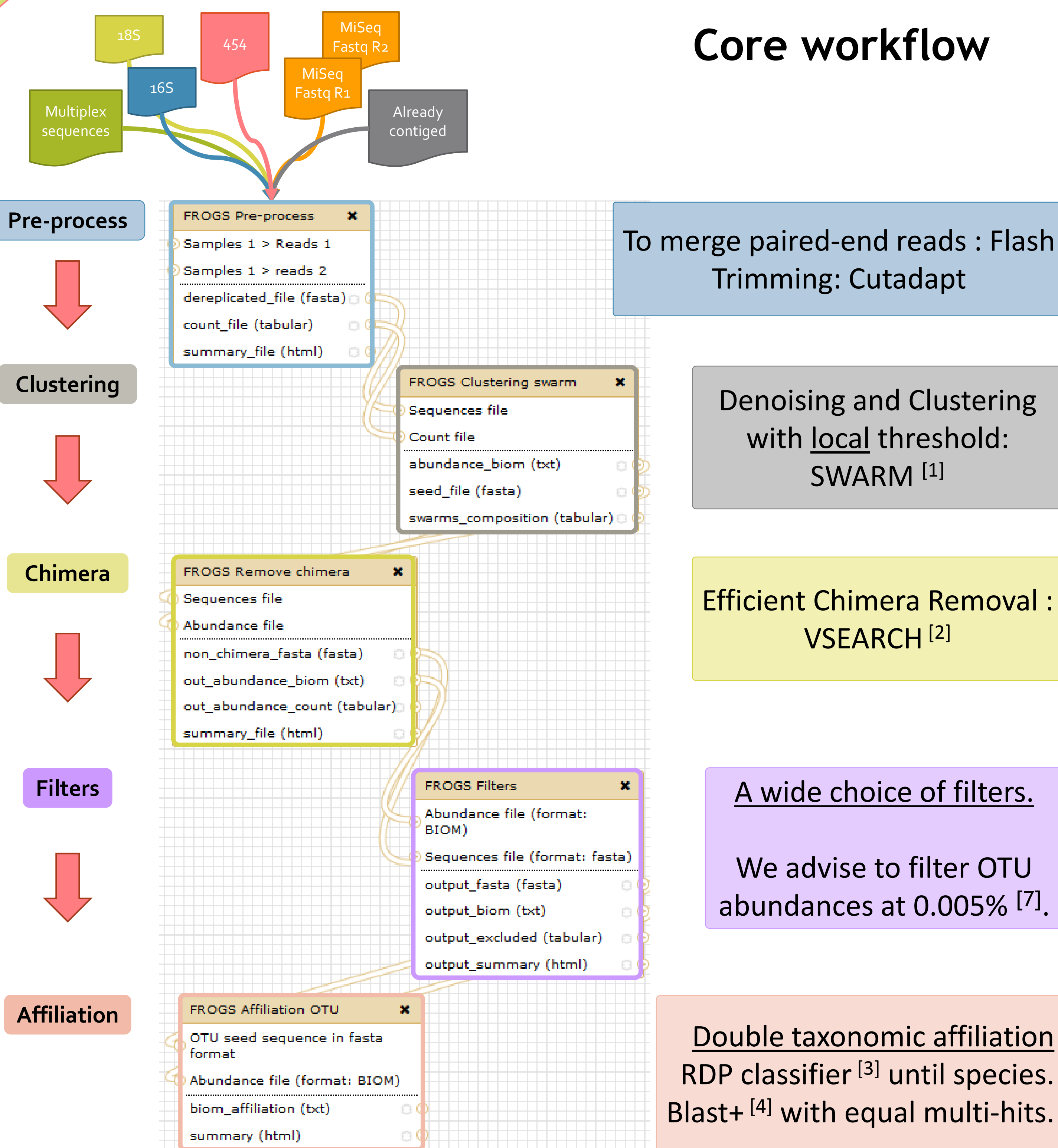
Accuracy



Speed



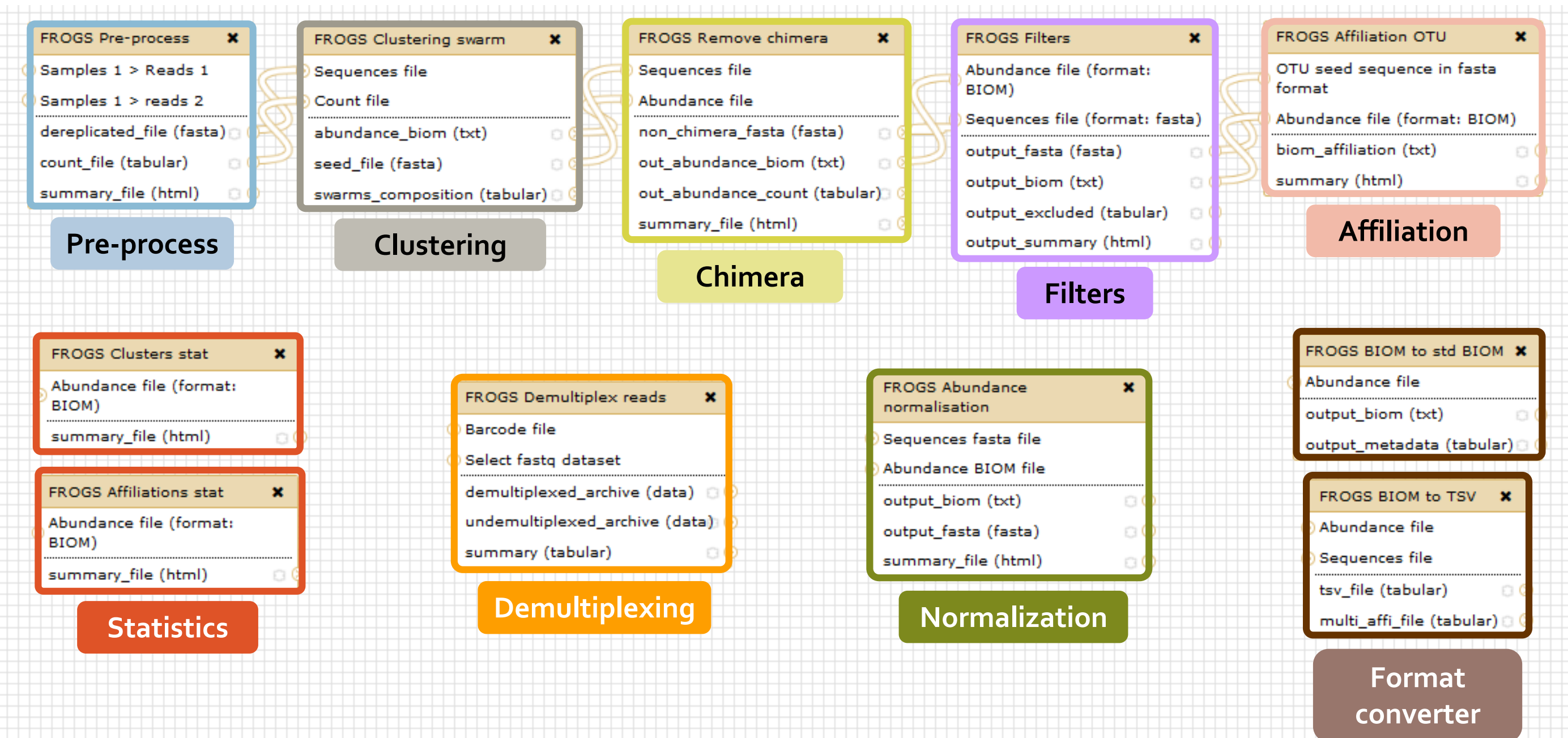
Scalability



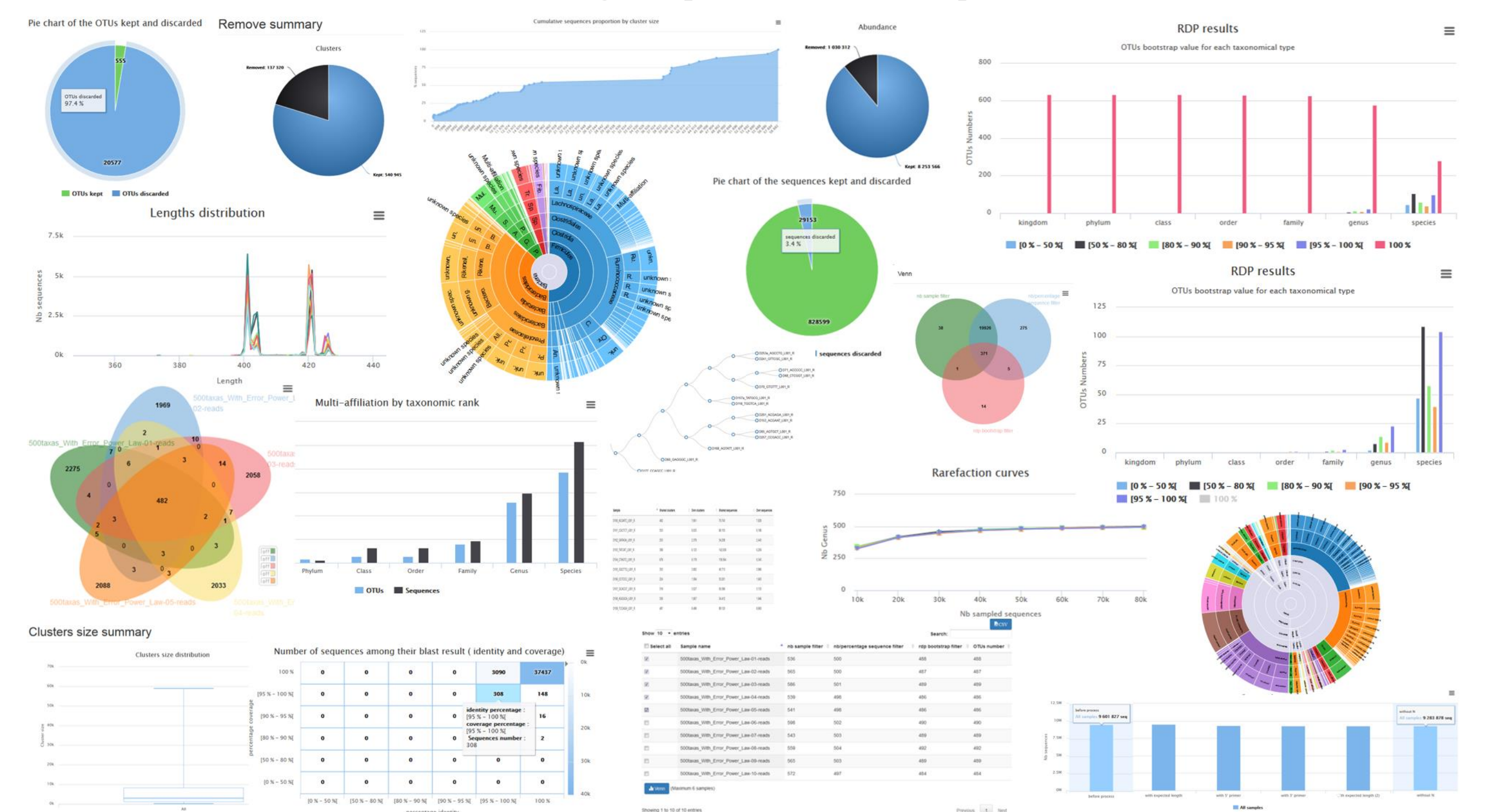
FROGS = complete analyze

Access : <http://galaxy-workbench.toulouse.inra.fr>

Interconnectable with BIOM format

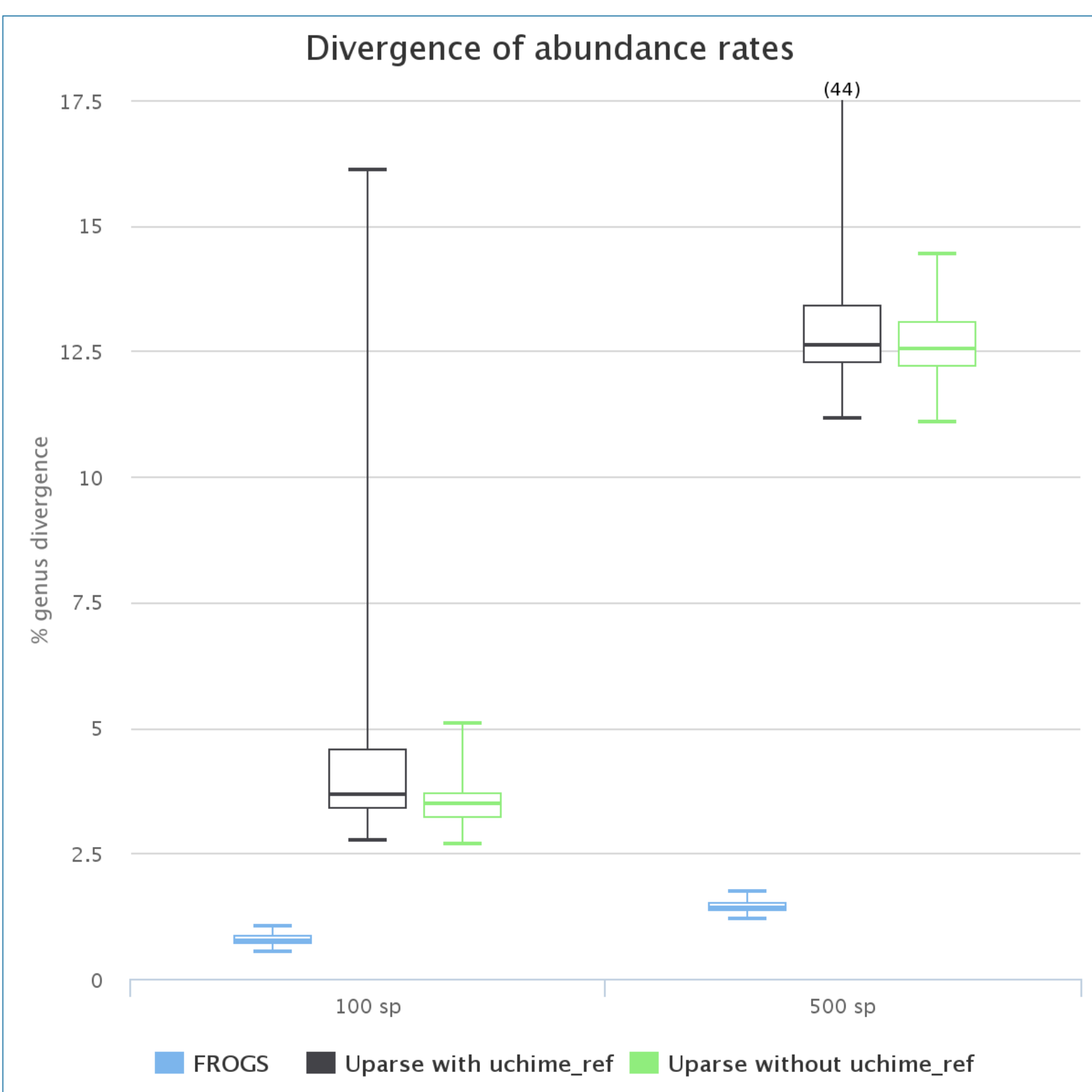


Great graphics outputs



Accuracy on simulated data set

For the tests we use two sets of species manually extracted from LTP 115. These species were chosen to obtain near and distant species in the same dataset. For each set of species we generate 5 artificial runs of 10 samples. In each sample the abundance of species following a power law. The ranking of species abundance is randomly determined in each sample. Then, an Illumina error profile is applied on reads and 20% of chimera are added. All the runs are processed after FROGS preprocess by FROGS and UPARSE [6].



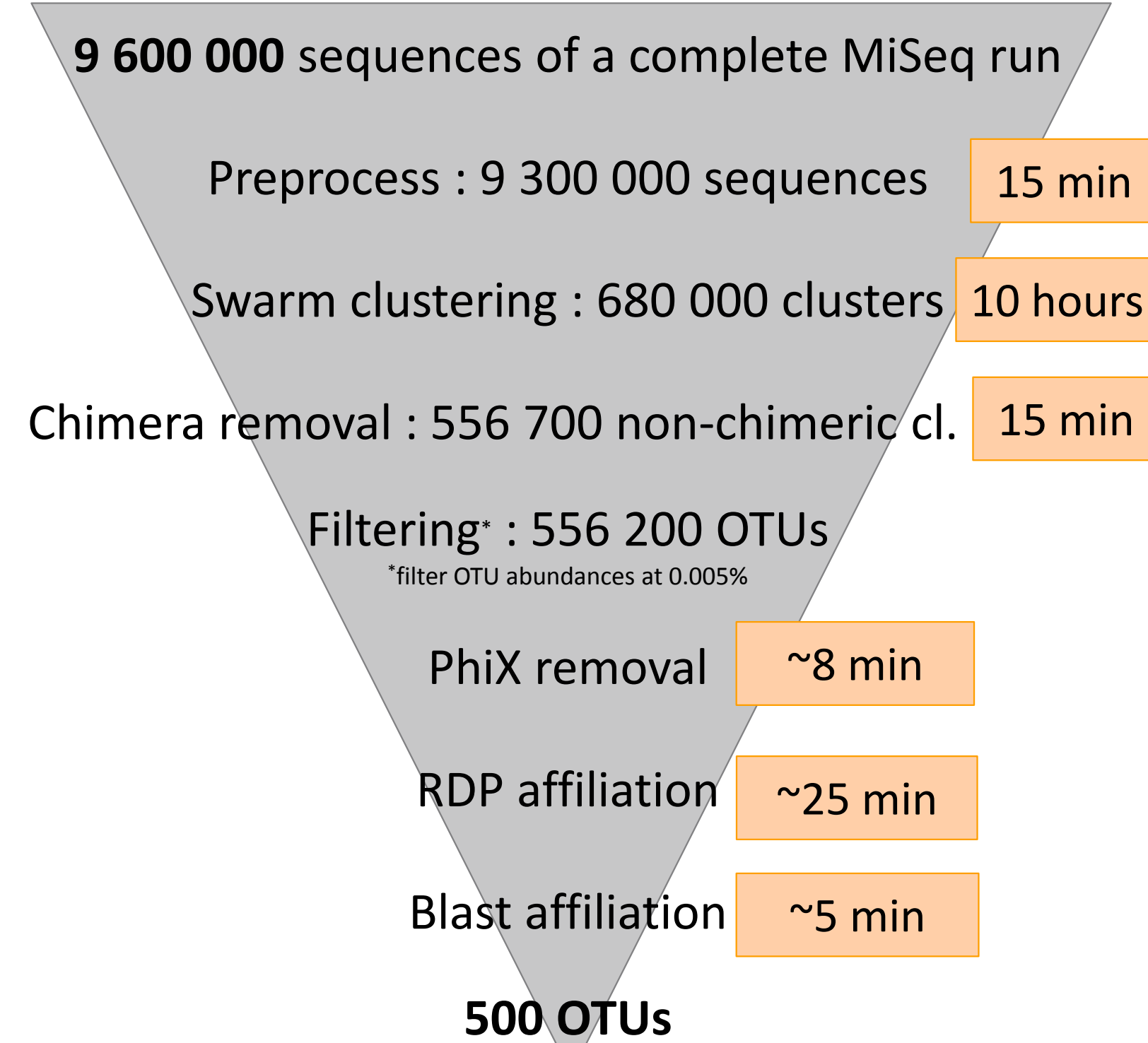
Divergence on the composition of microbial communities at the different taxonomic ranks

	FROGS guidelines*	UPARSE*
Kingdom	0.00	0.00
Phylum	0.38	7.73
Class	0.57	10.56
Order	0.81	11.48
Family	1.08	12.21
Genus	1.43	12.63
Species	1.53	12.67

	FROGS guidelines		UPARSE	
OTUs	Observed	Missing	Observed	Missing
496**	505#	0+	521#	3+
100**	105#	0+	111#	1+

**species/sample #OTU/sample + in OTU number

Speed on real datasets



Announcements

- Evaluate FROGS on others metrics and datasets (mock community, real already known community).
- Add other databases for affiliation (greengenes, ITS).
- Add FROGS in the toolshed.
- Github repository available <https://github.com/geraldinepascal/FROGS.git>
- Training sessions: Toulouse (December and April) & Jouy-en-Josas (January and in spring)

References

- Mahé & co. (2014) Swarm: robust and fast clustering method for amplicon-based studies.
- VSEARCH GitHub repository, doi:10.5281/zenodo.15524.
- Wang & co. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.
- Altschul & co. (1990) Basic local alignment search tool.
- Goecks & co. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.
- Edgar. (2013) UPARSE: Highly accurate OTU sequences from microbial amplicon reads.
- Bokulich (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing