



Training on Galaxy: Metabarcoding

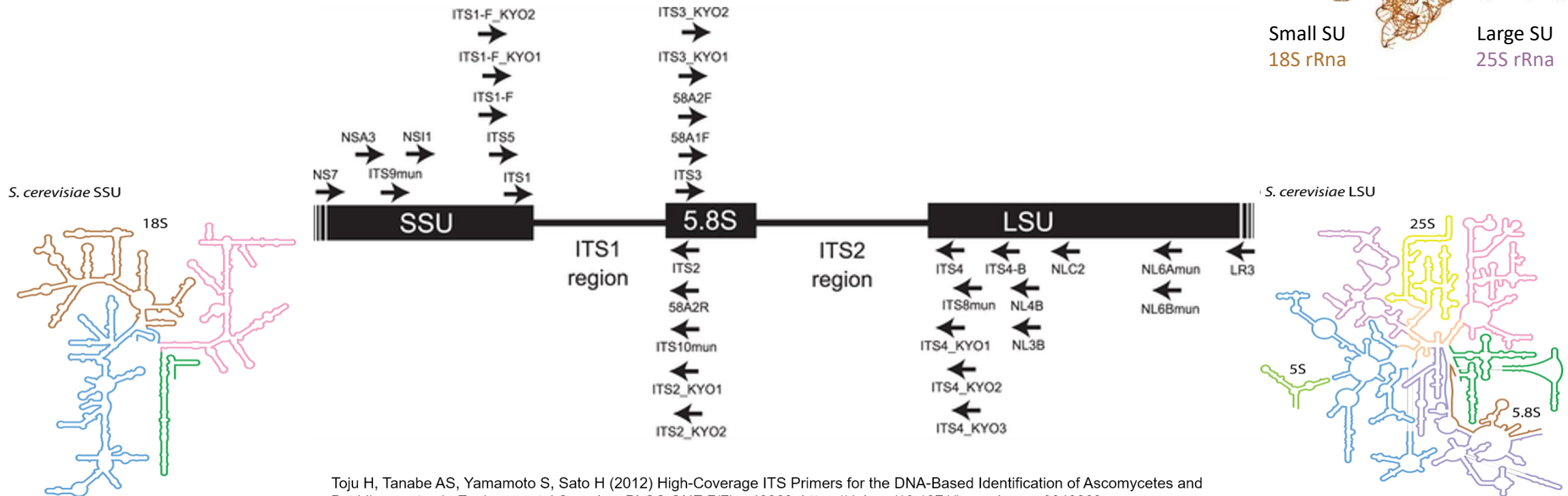
June 2022 - Webinar

FROGS Practice on ITS data

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

What is a ITS ?

Map of nuclear ribosomal RNA genes and their ITS regions.



Toju H, Tanabe AS, Yamamoto S, Sato H (2012) High-Coverage ITS Primers for the DNA-Based Identification of Ascomycetes and Basidiomycetes in Environmental Samples. PLOS ONE 7(7): e40863. <https://doi.org/10.1371/journal.pone.0040863>

Secondary Structures of rRNAs from All Three Domains of Life
Anton S. Petrov, Chad R. Bernier, Burak Gulen, Chris C. Waterbury, Eli Hershkovits, Chiaolong Hsiao, Stephen C. Harvey, Nicholas V. Hud, George E. Fox, Roger M. Wartell, Loren Dean Williams February 5, 2014 <https://doi.org/10.1371/journal.pone.0088222>

What is a ITS ?

- Size polymorphism of ITS (from 361 to 1475 bases in UNITE 7.1)
- Highly conserved regions of the neighboring of ITS1 and ITS2
- Lack of a generalist and abundant ITS databank (several small specialized databanks)
- Multiple copies* (14 to 1400 copies (mean at 113, median at 80))
- Do not target Glomeromycetes/Glomeromycota (→ alternative: 18S)



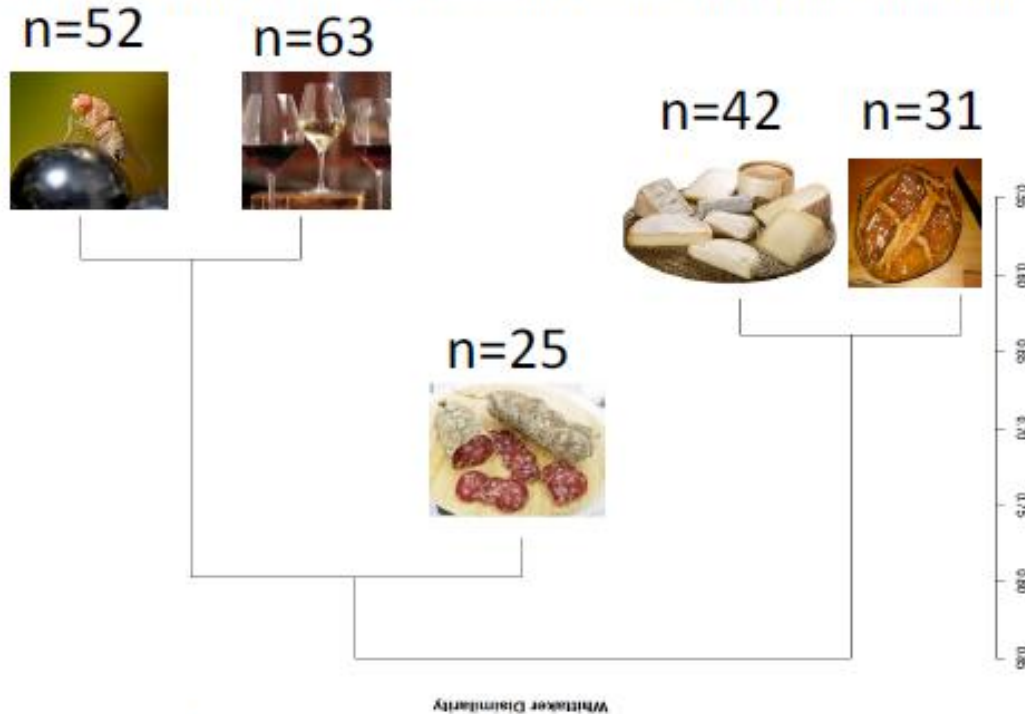
If your sequencing platform preprocesses your data, it has to keep short and long sequences

* <https://doi.org/10.1111/mec.14995>

ITS data form METABARFOOD project metaprogramme MEM

Yeast catalog in food ecosystem

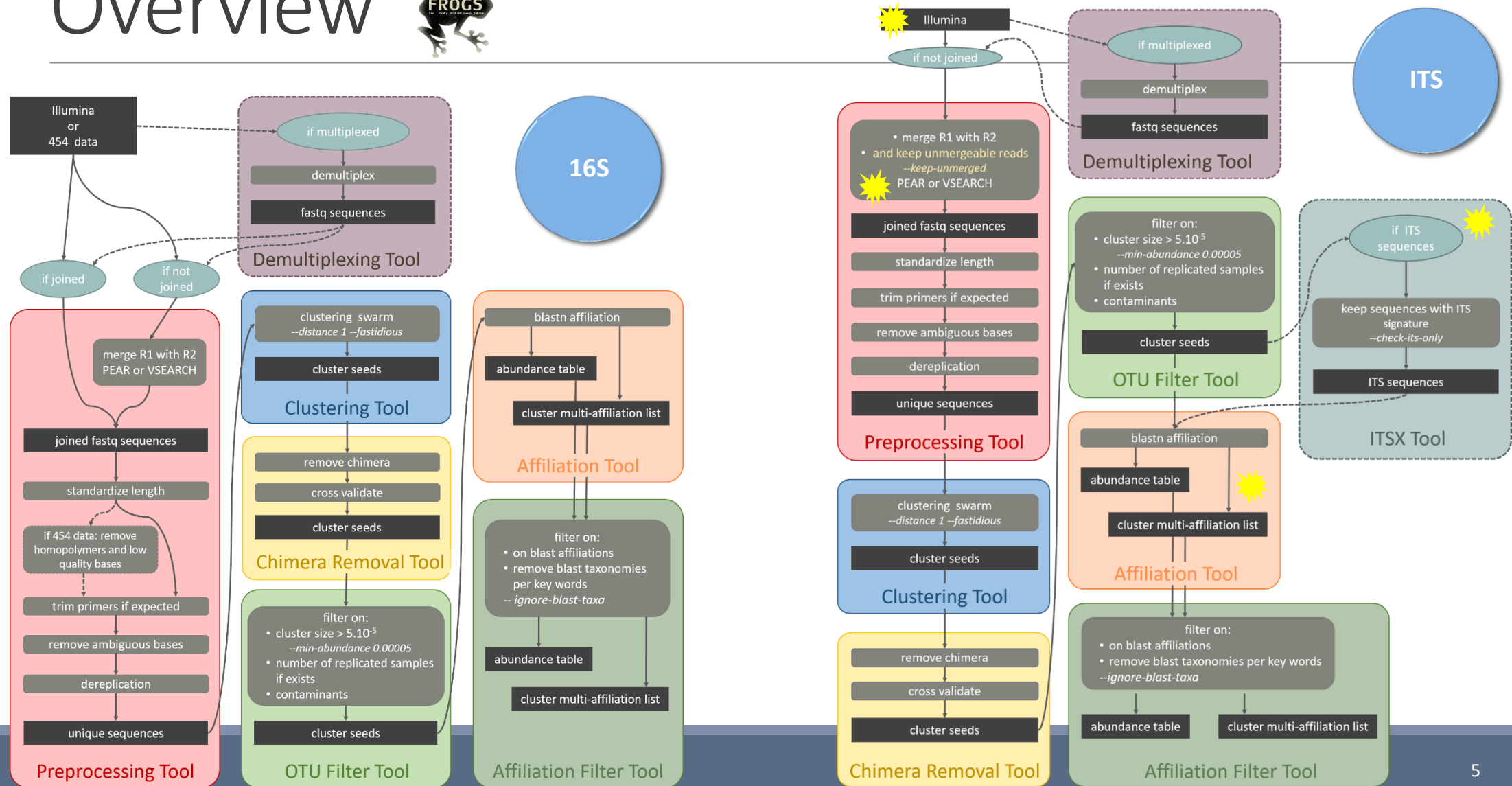
Number of yeast species reported at least twice in each ecosystem and their dissimilarity between ecosystems, as measured by the Whittaker distance



The universal fungal barcode, the Internal Transcribed Spacer (ITS) region, displays considerable size variation amongst yeasts and other micro-eukaryotes.

There are also several repeats leading to sequencing errors or termination.

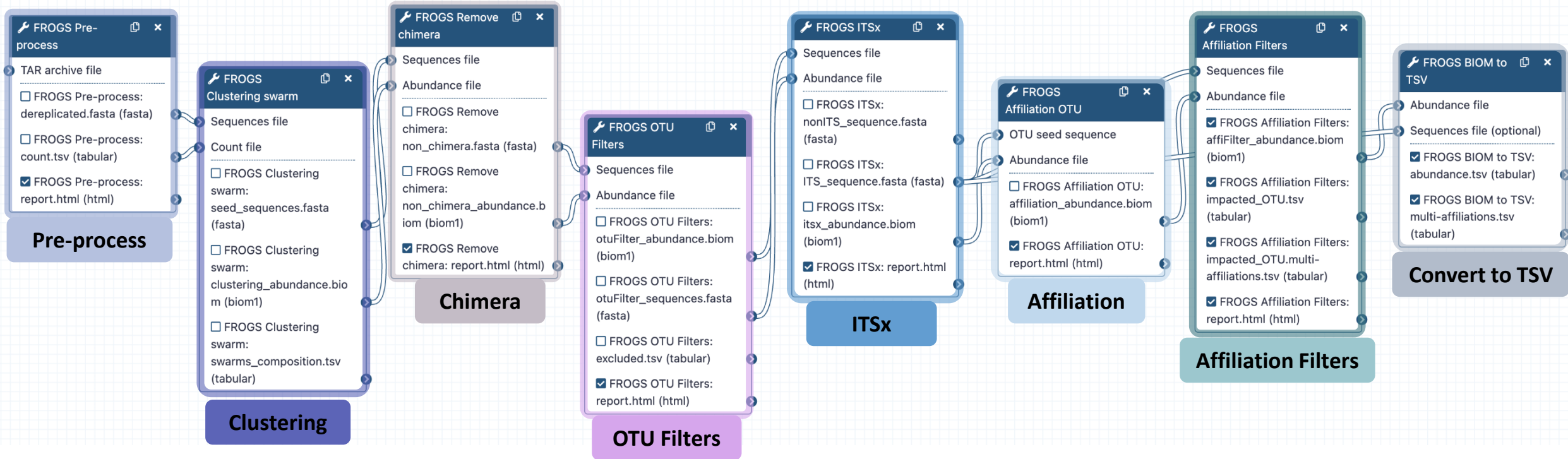
Overview

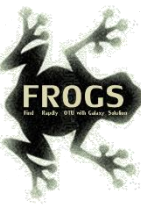




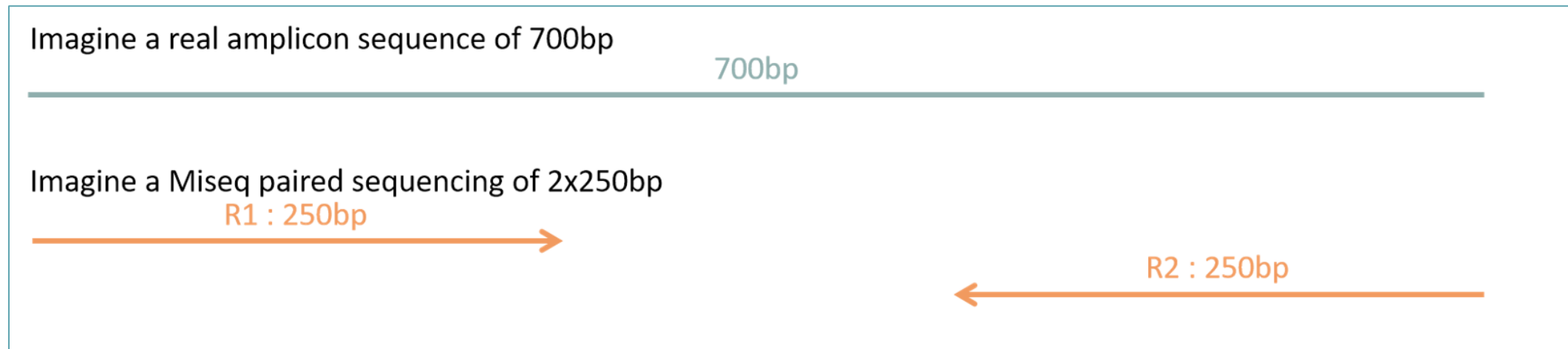
FROGS Pipeline

Minimal pipeline for ITS amplicon analyses

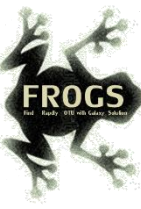




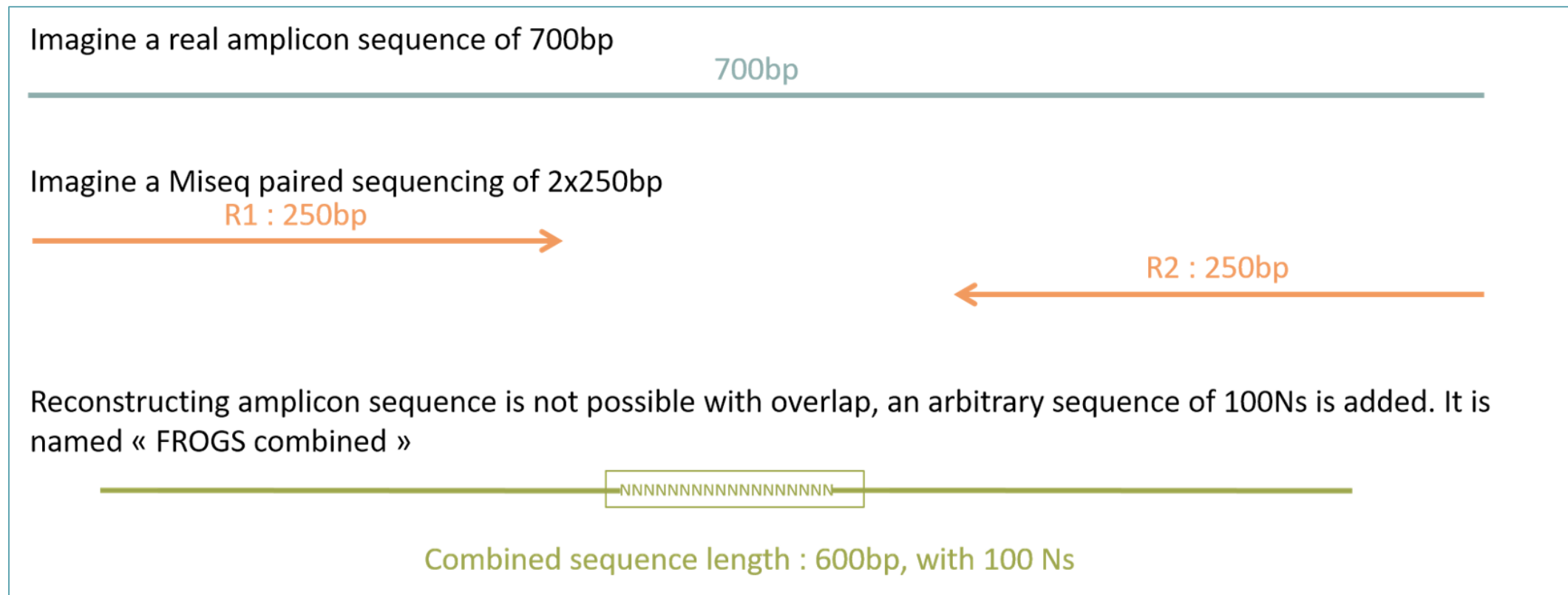
Problematic:
some ITS reads (Miseq sequencing) are non-overlapping
sequences



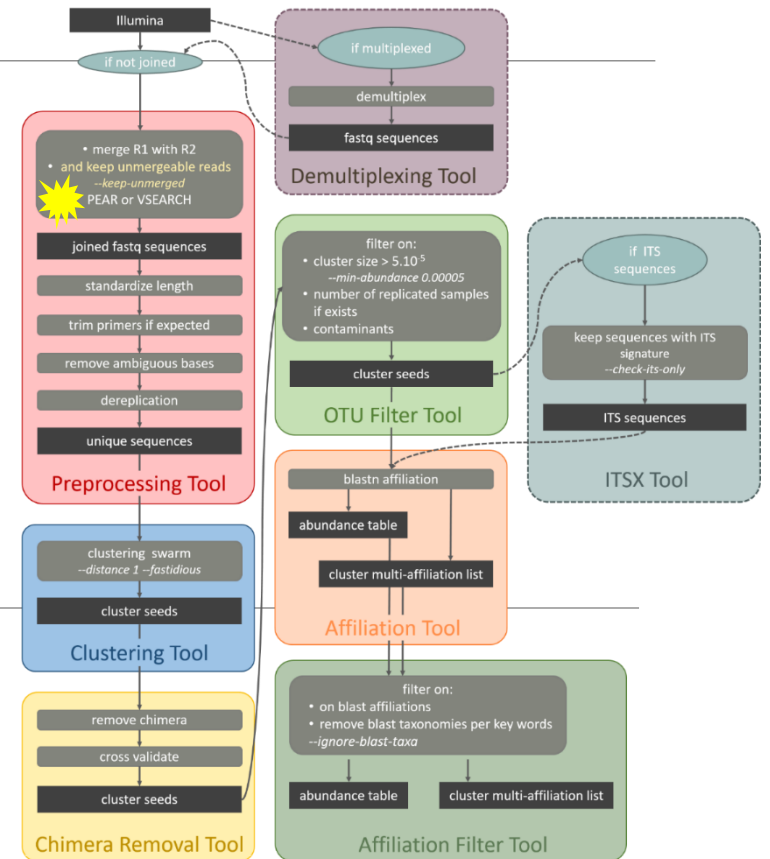
Consequence: during bioinformatics process, these reads are lost and underlying organisms will be never represented in the abundance table.



Solution: in preprocess step – creation of “FROGS combined” sequences



Pre-process tool





Sequencer

illumina

Select the sequencing technology used to produce the sequences.

Input type

TAR Archive

Samples files can be provided in a single TAR archive or sample by sample (with one or two files each).

TAR archive file

1: ITS1.tar.gz

The TAR file containing the sequences file(s) for each sample.

Are reads already merged ?

No

The archive contains 1 file by sample : R1 and R2 pair are already merged in one sequence.

Reads 1 size

250

The maximum read1 size.

Reads 2 size

250

The maximum read2 size.

Mismatch rate.

0.1

The maximum rate of mismatch in the overlap region

Merge software

Vsearch

Select the software to merge paired-end reads.

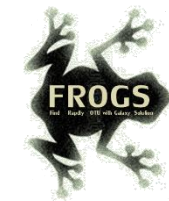
Would you like to keep unmerged reads?

Yes

No : Unmerged reads will be excluded; Yes : unmerged reads will be artificially combined with 100 N. (default No)

To keep FROGS combined sequences, choose YES



**Minimum amplicon size**

50

The minimum size for the amplicons (with primers).

Maximum amplicon size

490

The maximum size for the amplicons (with primers).

Sequencing protocol

Illumina standard

The protocol used for sequencing step: standard or custom with PCR primers as sequencing primers.

5' primer

CTTGGTCATTTAGAGGAAGTAA

The 5' primer sequence (wildcards are accepted). The orientation is detailed below in 'Primers parameters' help section.

3' primer

GCATCGATGAAGAACGCAGC

The 3' primer sequence (wildcards are accepted). The orientation is detailed below in 'Primers parameters' help section.

Email notification No

Send an email notification when the job completes.

 Execute

Primer 5': CTTGGTCATTTAGAGGAAGTAA
Primer 3': GCATCGATGAAGAACGCAGC

Exercise

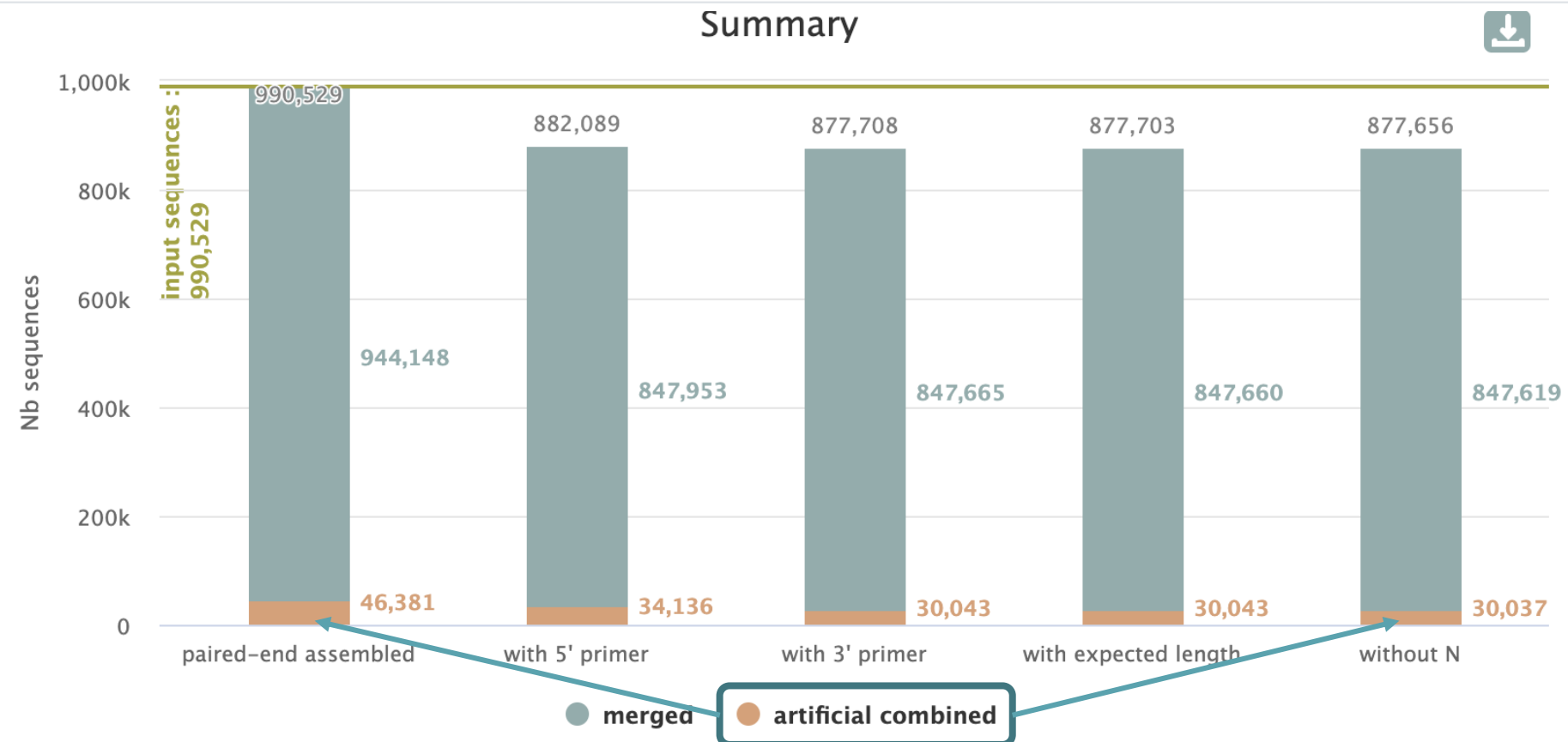
Go to « [ITS](#) » history

Launch the pre-process tool on this data set

→ objective: understand preprocess report and « FROGS combined sequences »

Explore Preprocess report.html

Preprocess summary





2 tables:

Details on merged sequences

Show entries Search: [CSV](#)

Samples	before process	% kept	paired-end assembled	with 5' primer	with 3' primer	with expected length	without N
complexe-ADN-1	56,284	85.79	52,757	48,305	48,292	48,292	48,288
echantillon1-1	32,883	81.01	31,137	26,653	26,640	26,640	26,639
echantillon1-2	56,166	91.15	53,736	51,225	51,200	51,200	51,195
echantillon1-3	84,102	71.56	80,060	60,224	60,188	60,187	60,181
echantillon2-1	53,405	86.65	51,004	46,282	46,276	46,275	46,275

Own tag for combined sequences

```
>Cluster_20410 1:N:0:ATATAA
AAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGAT
>Cluster_2881 1:N:0:ATATAA
AAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGAT
>Cluster_10465 1:N:0:ATTACA
AAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGAT
>Cluster_2714_FROGS_combined R1_desc:1:N:0:
AAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGAT
>Cluster_6993_FROGS_combined R1_desc:1:N:0:
AAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGAT
>Cluster_2580_FROGS_combined R1_desc:1:N:0:
AAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGAT
```

Details on artificial combined sequences

Show entries Search: [CSV](#)

Samples	before process	% kept	paired-end assembled	with 5' primer	with 3' primer	with expected length	without N
complexe-ADN-1	56,284	4.57	3,527	2,814	2,572	2,572	2,571
echantillon1-1	32,883	3.03	1,746	1,151	997	997	997
echantillon1-2	56,166	2.81	2,430	1,789	1,582	1,582	1,581
echantillon1-3	84,102	2.61	4,042	2,574	2,199	2,199	2,199
echantillon2-1	53,405	2.46	2,401	1,554	1,317	1,317	1,316

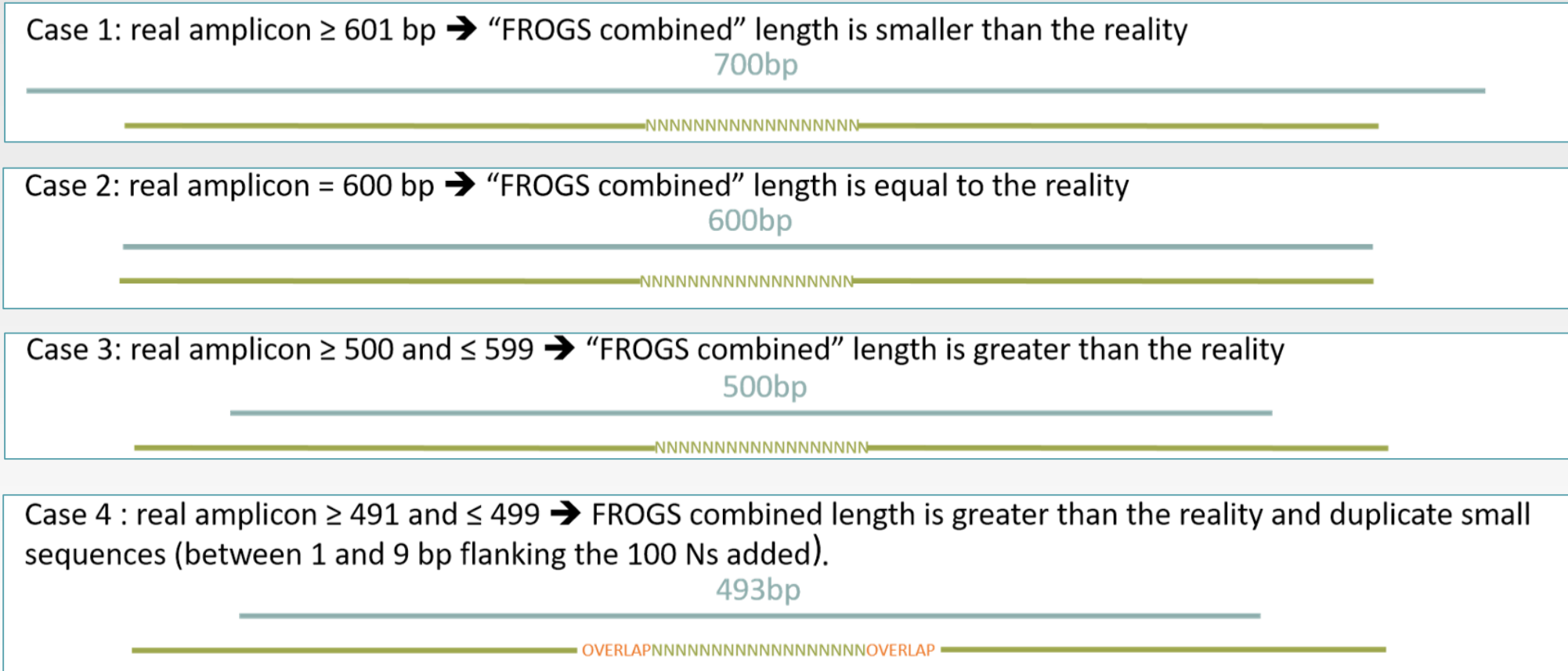
Filter only on minimum length for « combined ».

Minimum length = $R1 + 100N + R2 - \text{primers sizes}$

If the primers are very internal to the read, after trimming them, the combined sequence could be smaller than a read. FROGS rejects these cases.

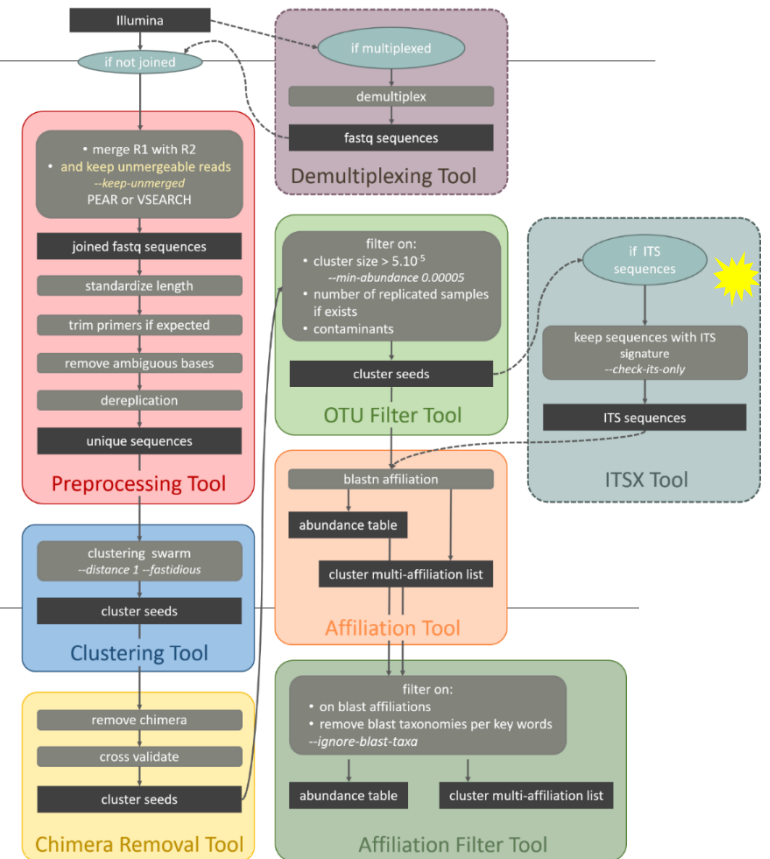
FROGS "combined" sequences are **artificial** and present particular features especially on size.

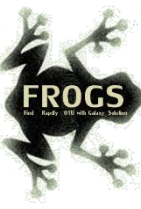
Imagine a MiSeq sequencing of 2x250pb with reads impossible to overlap. So FROGS "combined" length = 600 bp.





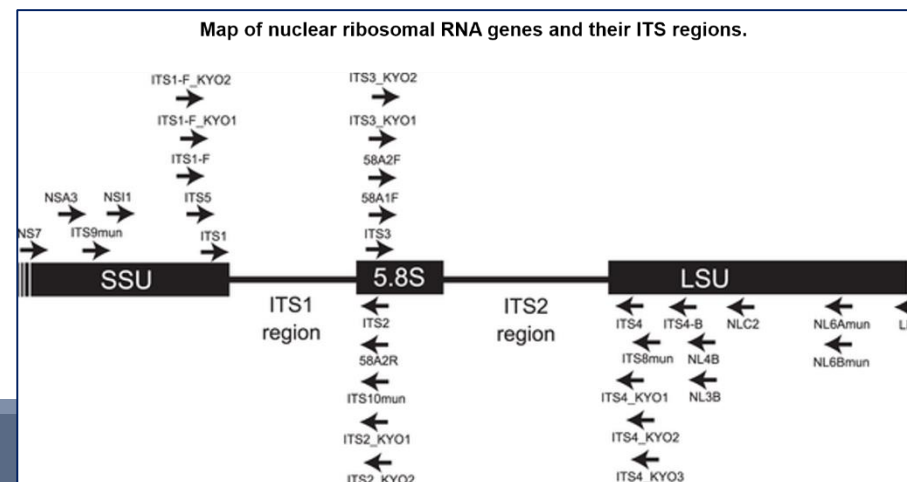
ITSx tools





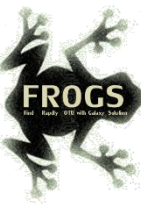
What is the purpose of the ITSx tool?

- ITSx is a tool to **filter** sequences.
- ITSx **identifies** and **trims** ITS regions in sequences.
- It **excludes** the highly conserved neighboring sequences **SSU**, **5S** and **LSU** rRNA.
- If the ITS1 or ITS2 region is not detected, the sequence is discarded.
- You can choose to check only if the sequence is detected as an ITS.
In this case, the sequence is not trimmed, only sequences not detected as ITS are rejected (*e.g.* contaminants).

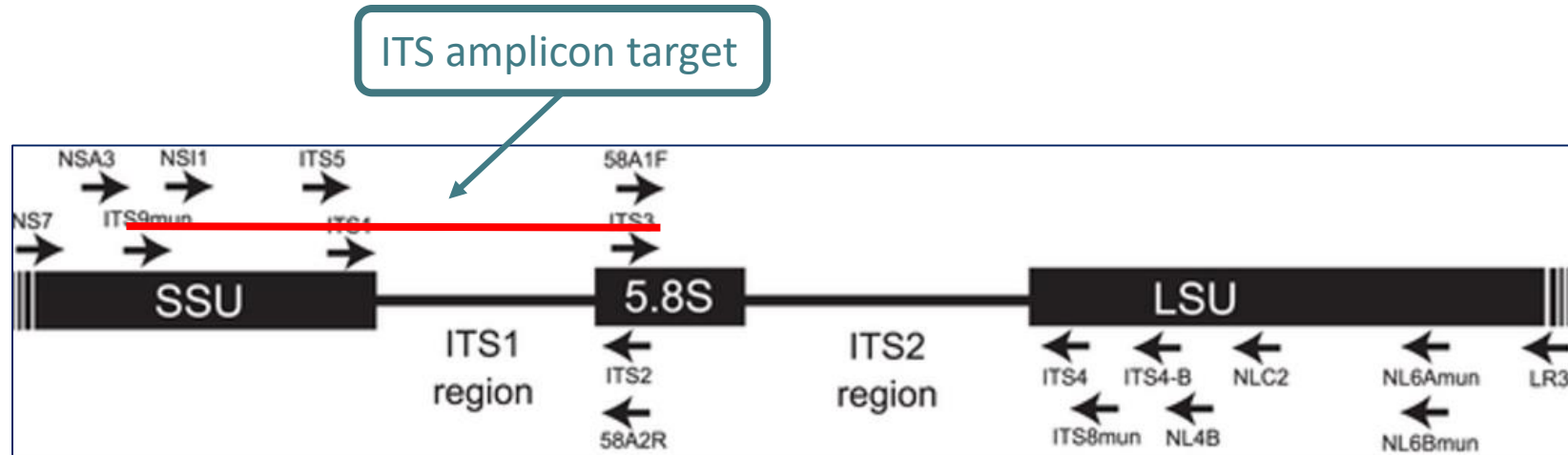



Bengtsson-Palme, J., et al. (2013), Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol Evol*, 4: 914-919.


<https://doi.org/10.1111/2041-210X.12073>

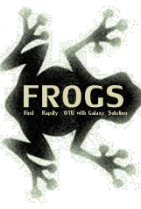


What is the purpose of the ITSx tool?



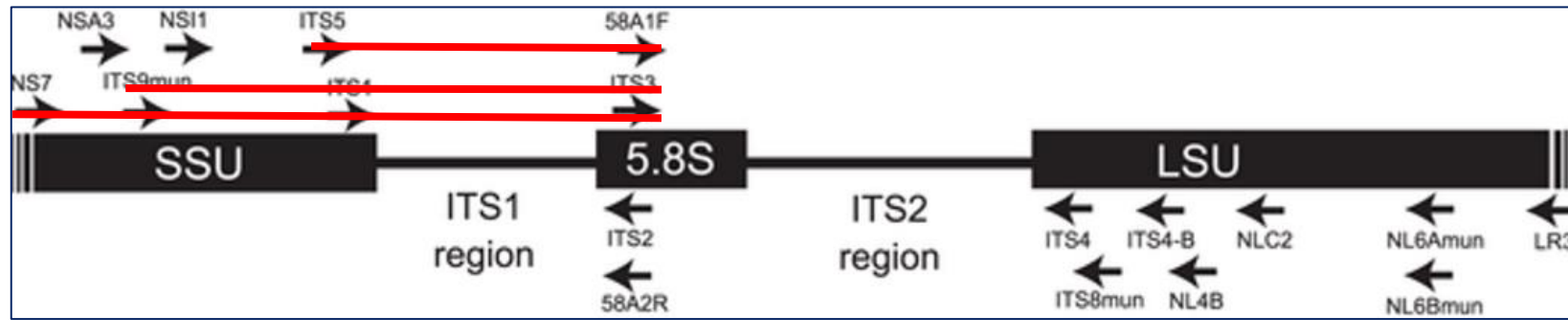
1st case: choose to trim
ITS1 is well detected
SSU part and 5.8S part are trimmed
Result: 

2nd case: choose to check only
ITS1 is well detected
SSU part and 5.8S part are not trimmed
Result: 



Check only if sequence is detected as ITS? Yes or not?

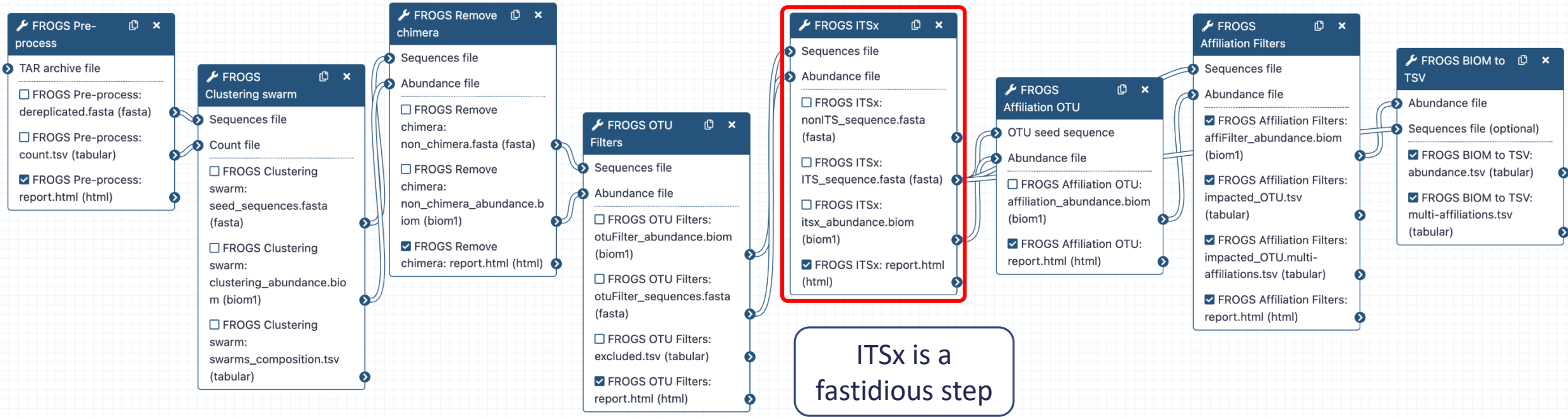
- If not, only ITS1 or ITS2 part will be conserved
- This is interesting to keep only the ITS parts without the flanking sequences in case of :
 - comparison of sequenced amplicons with different primers targeting the same region to be amplified.
 - using a database with only ITS part

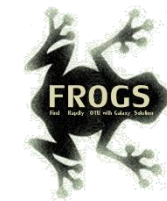




When should we use ITSx ?

After filtering !





Sequences file

13: FROGS OTU Filters: otuFilter_sequences.fasta

The sequence file to filter (format: fasta).

Abundance file

12: FROGS OTU Filters: otuFilter_abundance.biom

The abundance file to filter (format: BIOM).

ITS region

ITS1 ▾

Which fungal ITS region is targeted: either ITS1 or ITS2

Trim conserved sequence (SSU, 5.8S, LSU) ?

Yes

If Yes, only part of the sequences with ITS signature will be kept, SSU, LSU or 5.8S regions will be trimmed. (default : No)

Choose pertinent organisms to scan?

Select/Unselect all

- Fungi
- Alveolata
- Bryophyta
- Bacillariophyta
- Amoebozoa
- Euglenozoa
- Chlorophyta
- Rhodophyta
- Phaeophyceae
- Marchantiophyta
- Metazoa
- Oomycota
- Haptophyceae
- Raphidophyceae
- Rhizaria
- Synurophyceae
- Tracheophyta
- Eustigmatophyceae

Save a lot of time by checking pertinent organism group model to scan.

Email notification

No

Send an email notification when the job completes.

By default, only part of the sequence with ITS signature is kept (SSU, LSU or 5.8S are trimmed)

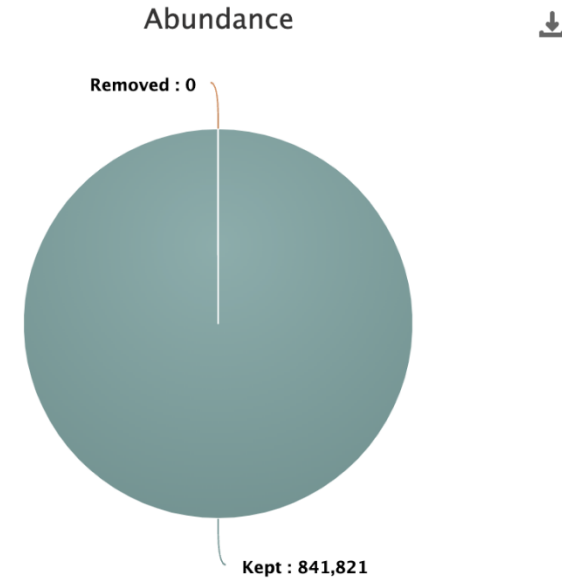
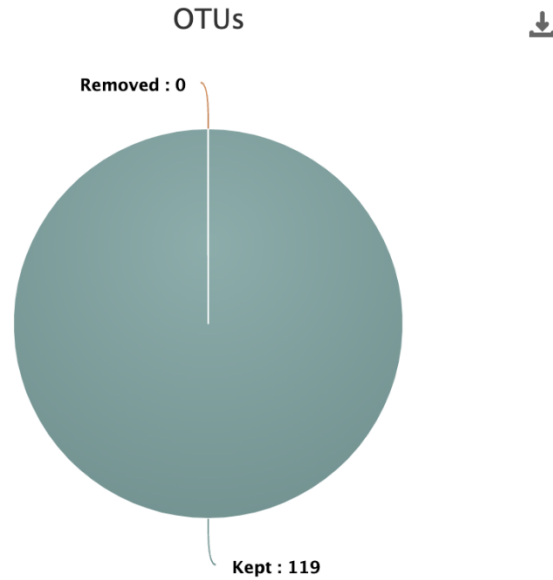
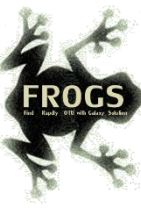
By default, sequences are considered as FUNGI sequences. Change it, if it is not the case.

Careful !



- The ITSx step is time consuming and has to be done on clusters. We advise our users to apply ITSx in 5th step:
 1. Preprocess step,
 2. Clustering step,
 3. Chimera removing step,
 4. Filter on OTUs abundances and replicats step,
 5. ITSx if Fungi ITS amplicons.

Filters (ITSx) summary



Filters (ITSx) by samples

↓ CSV

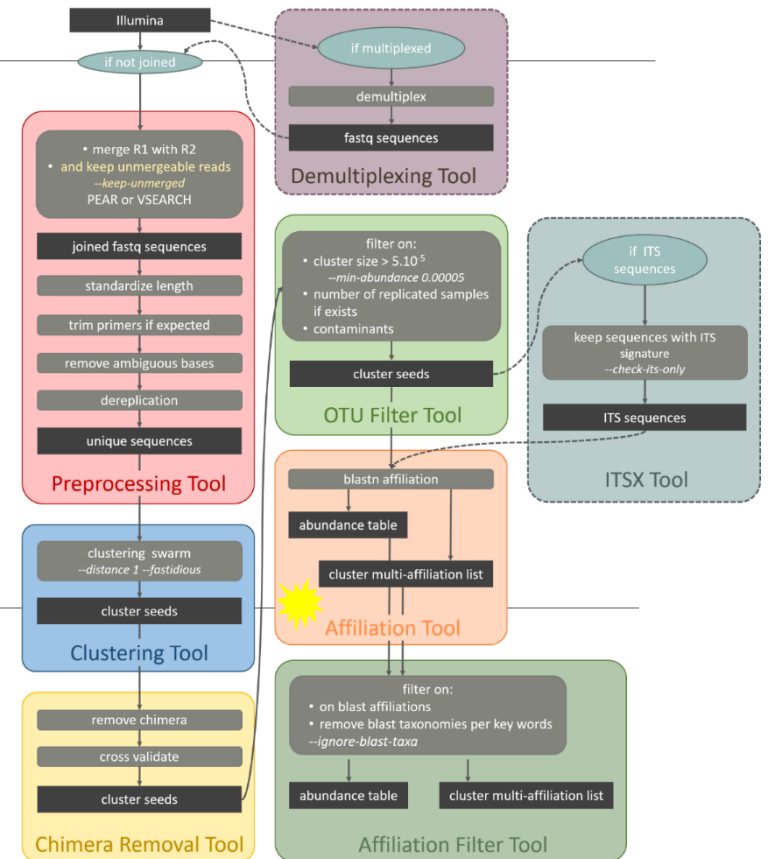
Show entries

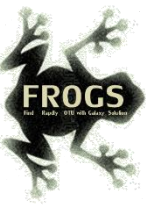
Search:

OTUs removed by sample

Sample name	↕ Initial	↕ Kept	↕ Initial abundance	↕ Kept abundance
complexe-ADN-1	92	92	46,438	46,438
echantillon1-1	71	71	26,497	26,497
echantillon1-2	71	71	50,940	50,940

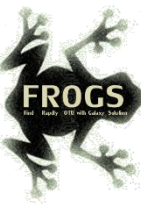
ITS Affiliation





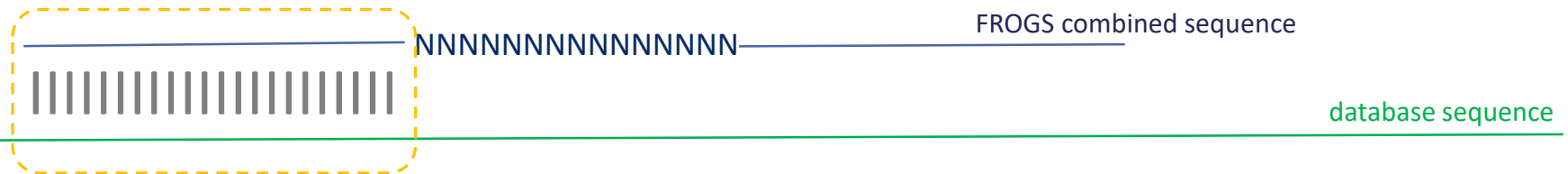
What is special about the affiliation of ITS/FROGS combined sequences?

- 2 alignment tools - blastn+ or needleall - are used to find alignments between each OTU and the database.
- Only the best hits with the same score are reported.
- blastn+ is used for classical **merged read pair**, and blastn+ then needleall are used for **artificially combined sequence**.
- For each alignment, several metrics are computed: %identity, %coverage and alignment length.



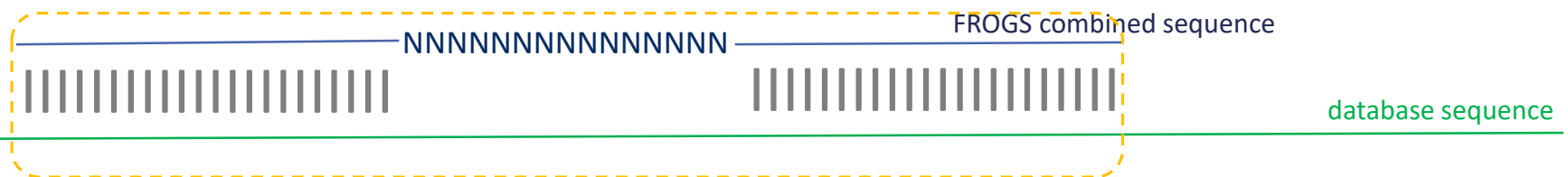
What is special about the affiliation of ITS/FROGS combined sequences?

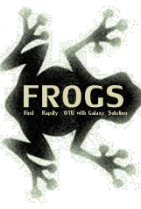
- blastn+ *i.e.* a local aligner, is not usable for "**combined**" sequences



Between combined and the database sequence, alignment is perfect until N stretch with blastn+. Information about the 2nd part of sequence are not explored !

- It is necessary to use a global aligner *i.e.* Needleall (the sequence must be aligned in its entirety), but it is computationally too hard.

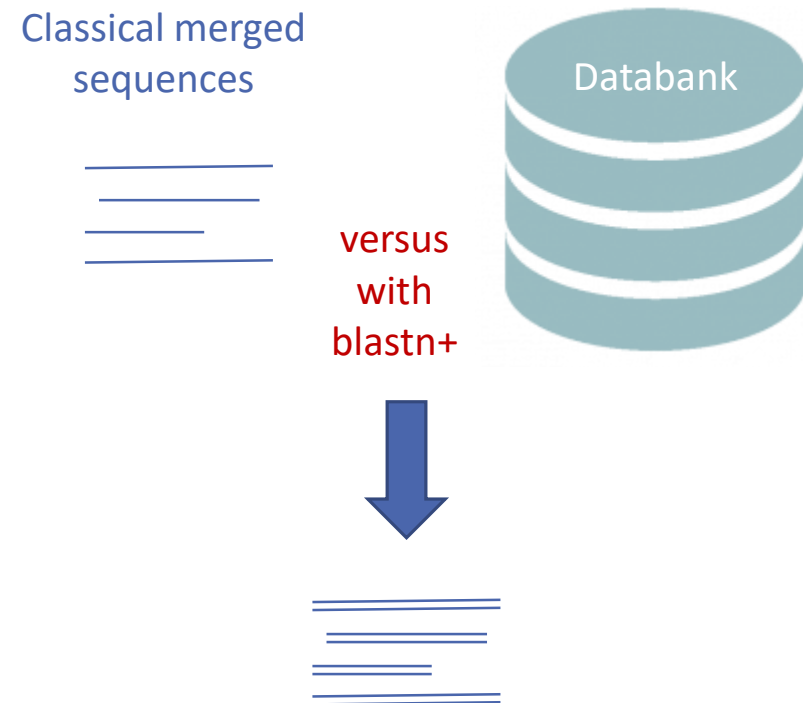




What is special about the affiliation of ITS/FROGS combined sequences?

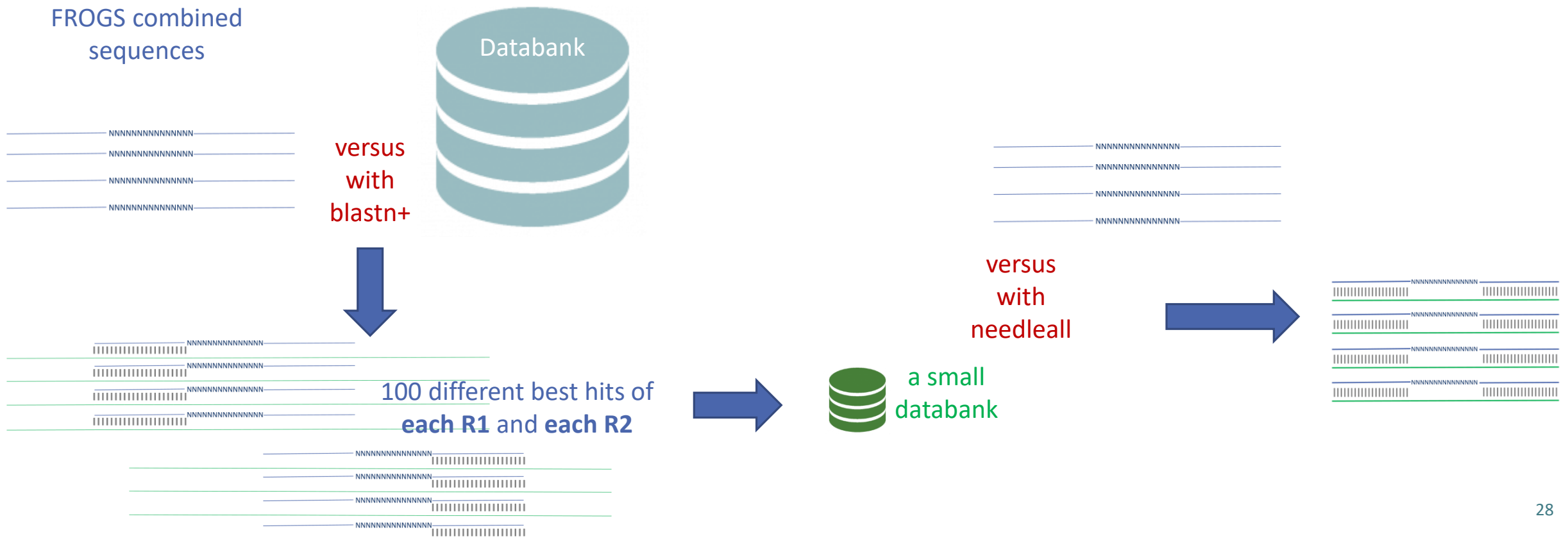
Solution:

- 1st step treat classical merged sequences with blastn+



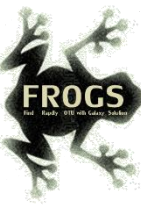
What is special about the affiliation of ITS/FROGS combined sequences?

- 2nd step for FROGS combined sequences: use blastn+ to create a small databank and align with needleall this small databank versus FROGS combined sequences





Careful, with "combined" sequences, we introduced some modification on identity percentage



Case 1: a sequencing of overlapping sequences *i.e.* 16S V3-V4 amplicon MiSeq sequencing

Imagine a real amplicon sequence of 400bp



Reconstructing amplicon sequence is a merged sequence (length : 400bp, with 100bp overlap)

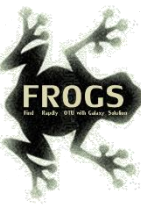


Affiliation is notably made by a local alignment with NCBI Blast+



Imagine a perfect sequencing without error:

classical %id = number of matches / alignment length = 400 matches / 400 positions = 100% identity



Case2: a sequencing of non-overlapping sequences case of ITS1 amplicon MiSeq sequencing

Imagine a real amplicon sequence of 700bp

700bp

Reconstructing a FROGS combined sequence (length : 600bp, with 100Ns)

NNNNNNNNNNNNNNNNNN

Affiliation could not be made by a local alignment but with a global alignment with Emboss needleall

NNNNNNNNNNNNNNNNNN

Imagine a perfect sequencing without error:

classical %id = number of matches / alignment length = (250+250 matches) / 700 positions = 71%



Case2: a sequencing of non-overlapping sequences case of ITS1 amplicon MiSeq sequencing

Filtering on %id will systematically removed “FROGS combined” OTUs.

So, we replaced the classical %id by a %id computed on the sequenced bases only.

% sequenced bases identity = number of matches / (seed length – artificial added N)

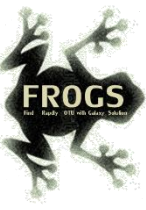
Case 1 : 16S V3V4 → overlapped sequence

% sequenced bases identity = 400 matches / 400 bp = **100 %**

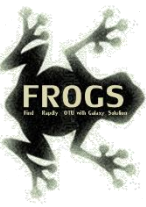
Case 2 : very large ITS1 → “FROGS combined” shorter than the real sequence

% sequenced bases identity = (250 + 250) / (600 - 100) = **100%**

This calculation allows the 100% identity score to be returned on FROGS "combined" shorter or longer than reality in case of perfect sequencing. And returns a lower percentage of identity in the case of repeated small overlaps kept in the FROGS "combined".



Affiliation Post-process



What is the purpose of the *Affiliation post-process* tool ?

This tool allows **grouping OTUs together** in accordance with the %id and %cov chosen by the user and according to the following criteria:

1. They must have the same affiliation

Or

2. If they have "multi-affiliation" tag in FROGS taxonomy, they must have in common in their list of possible affiliations at least one identical affiliation.



What is the purpose of the *Affiliation post-process* tool ?

In consequence:

The different affiliations involved in multi-affiliation are merged.

The abundances are added together.

It is the most abundant OTU seed that is kept.



FROGS Affiliation postprocess Optionnal step to resolve inclusive amplicon ambiguities and to aggregate OTUs based on alignment metrics (Galaxy Version 3.2.3+galaxy2)

☆ Favorite

▼ Options

Sequences file

17: FROGS ITSx: ITS_sequence.fasta

The sequence file to filter (format: fasta).

Abundance file

20: FROGS Affiliation OTU: affiliation_abundance.biom

The abundance file to filter (format: BIOM).

Is this an amplicon hyper variable in length?

Yes

Yes, we have combined sequences

Multi-affiliation tag may be resolved by selecting the shortest amplicon reference. For this you need the reference fasta file of your targetted amplicon.

Using reference database

ITS UNITE Fungi 8.3

Select reference from the list

minimum identity for aggregation

100

OTUs will be aggregated if they share the same taxonomy with at least X% identity.

minimum coverage for aggregation

100

OTUs will be aggregated if they share the same taxonomy with at least X% alignment coverage.

Email notification

No

Send an email notification when the job completes.

Here, we wanted to aggregate OTUs only if they are identical

✓ Execute



Example

After Preprocessing + Clustering + OTU Filter:

>Cluster_3

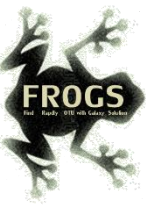
```
AAGTCGTAAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAGTGATTGCCTTTATAGGCTTATAACTATATCCACTTACACCTGTGAACTGTTCTACTACTTGACGCAAGTCGAGT
ATTTTTACAAACAATGTGTAATGAACGTCGTTTTATTATAACAAAATAAACTTTCAACAACGGATCTCTTGGCTCTC
```

>Cluster_54

```
AAGTCGTAAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAGTGATTGCCTTTATAGGCTTATAACTATATCCACTTACACCTGTGAACTGTTCTACTACTTGACGCAAGTCGAGT
ATTTTTACAAACAATGTGTAATGAACGTCGTTTTATTATAACAAAATAAACTTTCAACAACGGATCTCTTGGTTCCG
```

>Cluster_414_FROGS_combined

```
AAGTCGTAAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAGTGATTGCCTTTATAGGCTTATAACTATATCCACTTACACCTGTGAACTGTTCTACTACTTGACGCAAGTCGAGT
ATTTTTACAAACAATGTGTAATGAACGTCGTTTTATTATAACAAAATAAACTTTCAACAACGGATCTCTTGGCTCTCGCATCGATGAAGAACGCAGCAGATCGGAANNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
CATTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAGTGATTGCCTTTATAGGCTTATAACTATATCCACTTACACCTGTGAACTGTTCTACTAC
TTGACGCAAGTCGAGTATTTTTACAAACAATGTGTAATGAACGTCGTTTTATTATAACAAAATAAACTTTCAACAACGGATCTCTTGGCTCTC
```



Exemple

After Preprocessing + Clustering + OTU Filter + **ITSX** :

>Cluster_3

GTGATTGCCTTTATAGGCTTATAACTATATCCACTTACACCTGTGAACTGTTCTACTACTTGACGCAAGTCGAGTATTTTTACAAACAATGTGTAATGAACGTCGTTTTATTATAACAAAATA

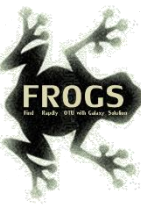
>Cluster_54

GTGATTGCCTTTATAGGCTTATAACTATATCCACTTACACCTGTGAACTGTTCTACTACTTGACGCAAGTCGAGTATTTTTACAAACAATGTGTAATGAACGTCGTTTTATTATAACAAAATA

>Cluster_414_FROGS_combined

GTGATTGCCTTTATAGGCTTATAACTATATCCACTTACACCTGTGAACTGTTCTACTACTTGACGCAAGTCGAGTATTTTTACAAACAATGTGTAATGAACGTCGTTTTATTATAACAAAATA

These 3 sequences have become **strictly identical** !



Example

After Preprocessing + Clustering + OTU Filter + ITSX + **Affiliation Post-process** :

Cluster_3, Cluster_54 and Cluster_414_FROGS_combined are **aggregated** in a same OTU

FROGS Affiliation postprocess: aggregation_composition.txt

Cluster_1 Cluster_244 Cluster_448_FROGS_combined Cluster_471_FROGS_combined

Cluster_2 Cluster_320 Cluster_357 Cluster_435 Cluster_468 Cluster_312 Cluster_364 Cluster_477 Cluster_466 Cluster_480

Cluster_3 Cluster_54 Cluster_414_FROGS_combined

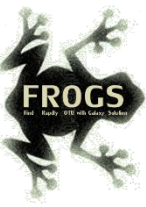
Cluster_4 Cluster_15 Cluster_27 Cluster_42 Cluster_67 Cluster_77 Cluster_137 Cluster_209 Cluster_422

Cluster_5 Cluster_5171

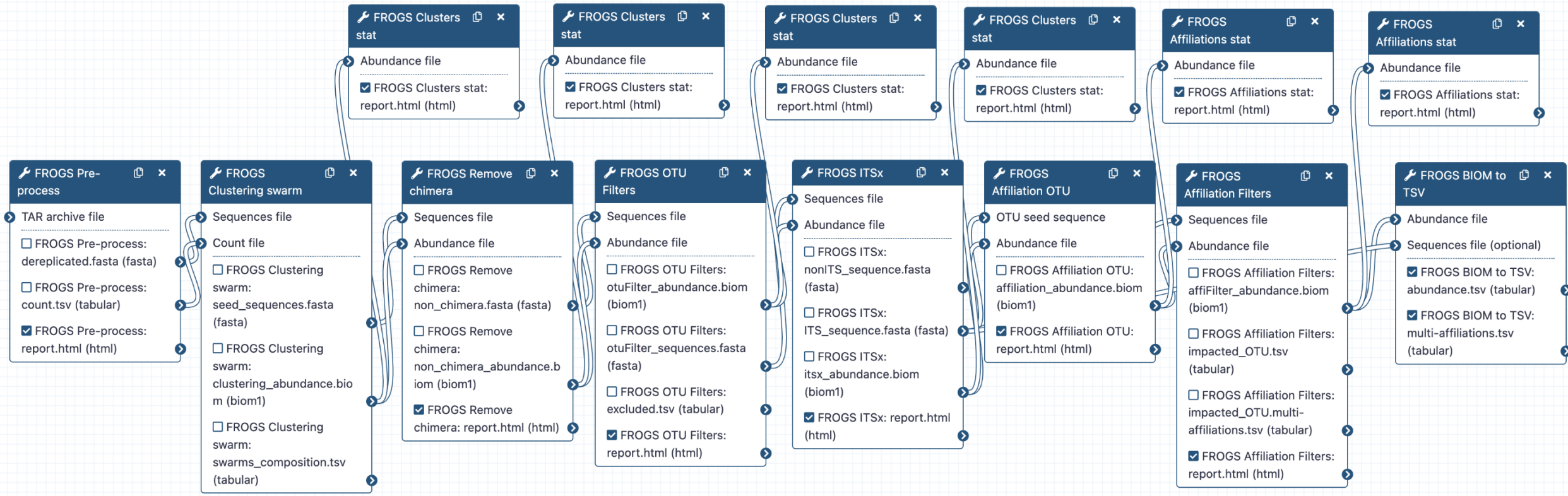
Cluster_6 Cluster_53

Cluster_9 Cluster_71

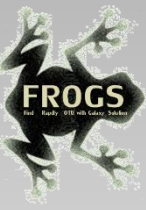
Cluster_7



Workflow creation

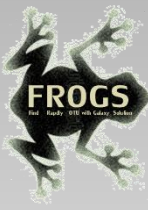


A workflow links FROGS steps together and when it is launched, all the steps run automatically.



Practice

CREATE YOUR OWN WORKFLOW !



Exercise



! New Galaxy server, needed tools/databanks are added on demand

Tools

search tools

Upload Data

BASIC TOOLS

- Monitoring
- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort

Search Workflows

+ Create Import

Name	Tags	Updated	Sharing	Bookmarked
16S		a few seconds ago		<input type="checkbox"/>
ITS		2 minutes ago		<input type="checkbox"/>

History

Rechercher des données

ITS

21 shown

252.24 MB

- 21: FROGS Affiliation O TU: report.html
- 20: FROGS Affiliation O TU: affiliation_abundance.biom
- 19: FROGS ITSx: report.html
- 18: FROGS ITSx: itsx_abundance.biom

Exercise

Create Workflow

Name

3

ITS_formation

Annotation

A description of the workflow; annotation is shown alongside shared or published workflows.







✕ Cancel

✓ Create


4








Exercise

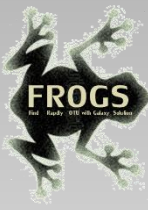
Search Workflows + Create Import

Name	Tags	Updated	Sharing	Bookmarked	
▼ ITS_formation		a few seconds ago		<input type="checkbox"/>	
▼ 16S		3 minutes ago		<input type="checkbox"/>	
▼ ITS		5 minutes ago		<input type="checkbox"/>	

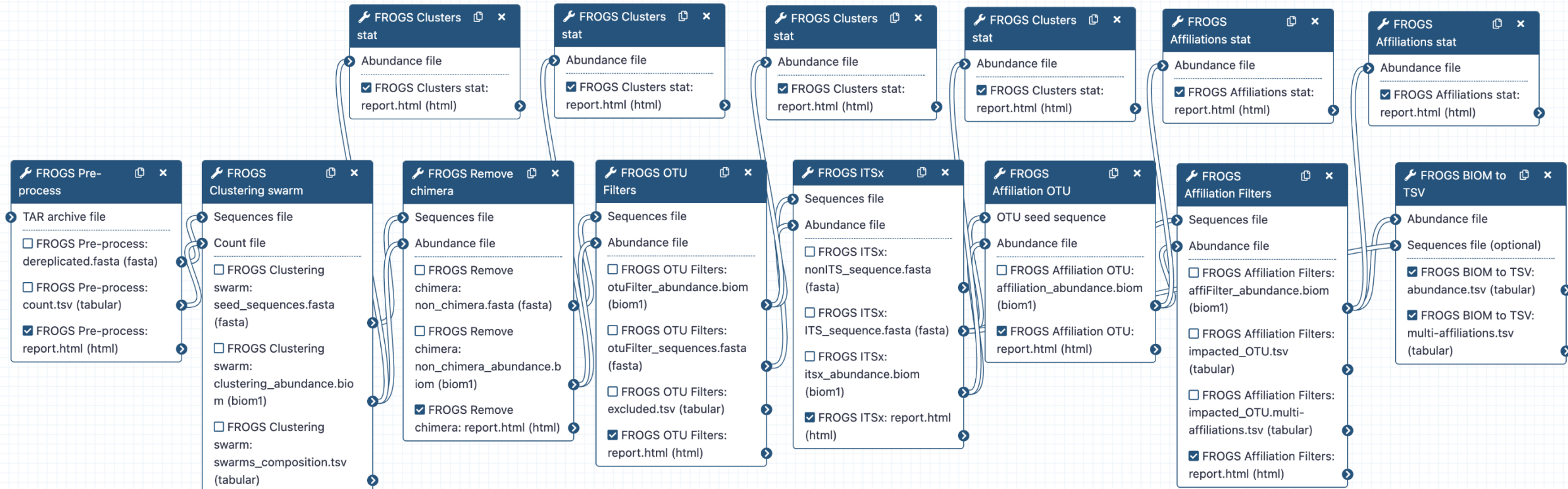
Name Tags

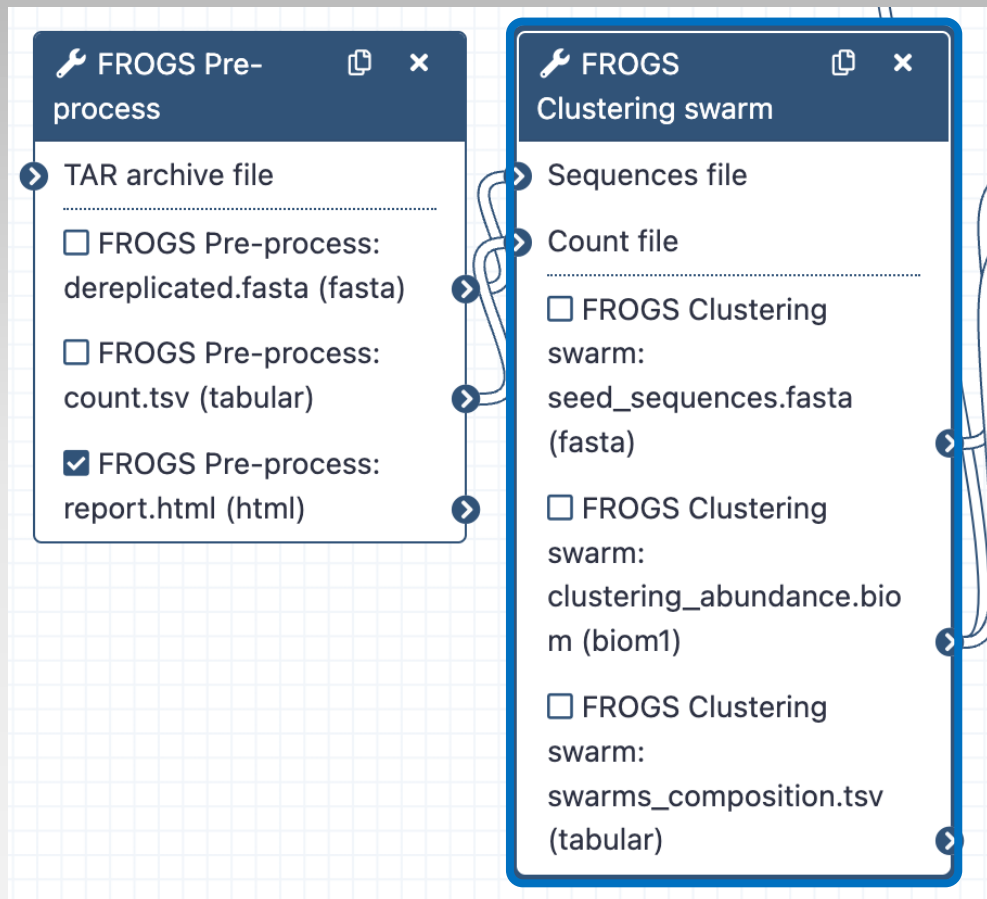
▼ ITS_formation 

-  Edit
-  Copy
-  Download
-  Rename
-  Share
-  View
-  Delete



Solution of exercise:





FROGS Clustering swarm Single-linkage clustering on sequences (Galaxy Version 3.2.3+galaxy2)

Label

Add a step label.

Step Annotation

Add an annotation or notes to this step. Annotations are available when a workflow is viewed.

Sequences file
Data input 'sequence_file' (fasta)
The dereplicated sequences file (format: fasta).

Count file
Data input 'count_file' (tabular)
It contains the count by sample for each sequence (format: TSV).

FROGS guidelines version
New guidelines from version 3.2

Denoising step prior to a d3 clustering is no more recommended since FROGS 3.2, but you can still choose it.

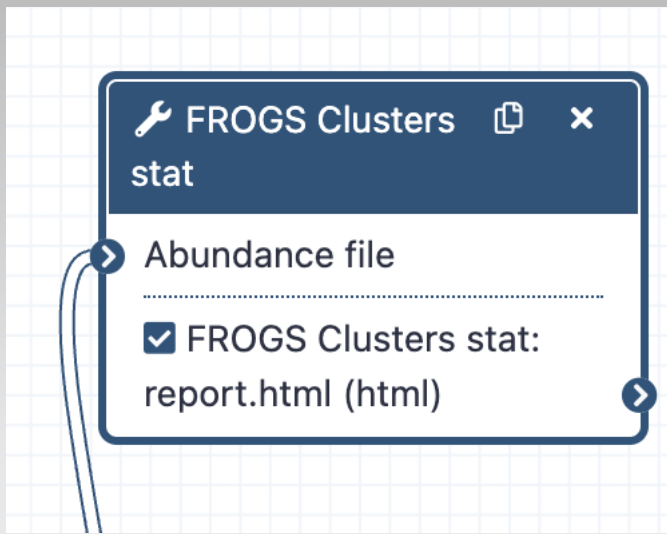
↔ Aggregation distance clustering
1

Maximum number of differences between sequences in each aggregation swarm step. (recommended d=1)

↔ Refine OTU clustering
 Yes

Clustering will be performed with the swarm --fastidious option, which is recommended and only usable in association with a distance of 1 (default and recommended: Yes)

For each tool, think to:
1. Set parameters



Configure Output: 'FROGS Clusters stat: report.html'

Label

This will provide a short name to describe the output - this must be unique across workflows.

Rename dataset

This action will rename the output dataset. Click here for more information. Valid input variables are:

- **biom** (Abundance file)

Change datatype

Leave unchanged

This action will change the datatype of the output to the indicated datatype.

Add Tags

This action will set tags for the dataset.

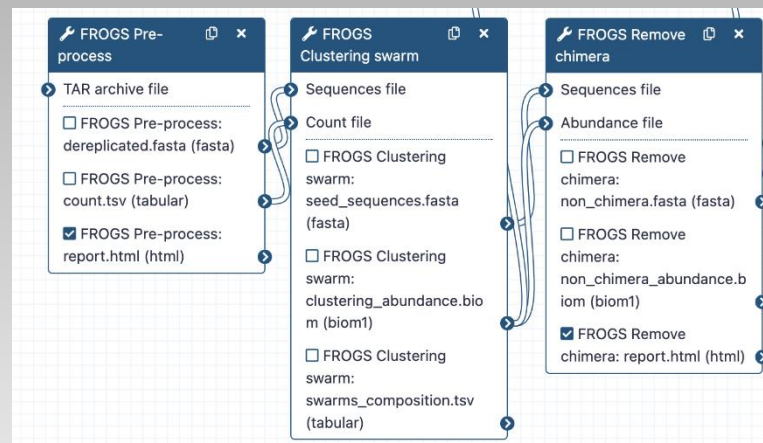
Remove Tags

This action will remove tags for the dataset.

[Assign columns](#)

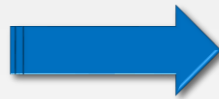
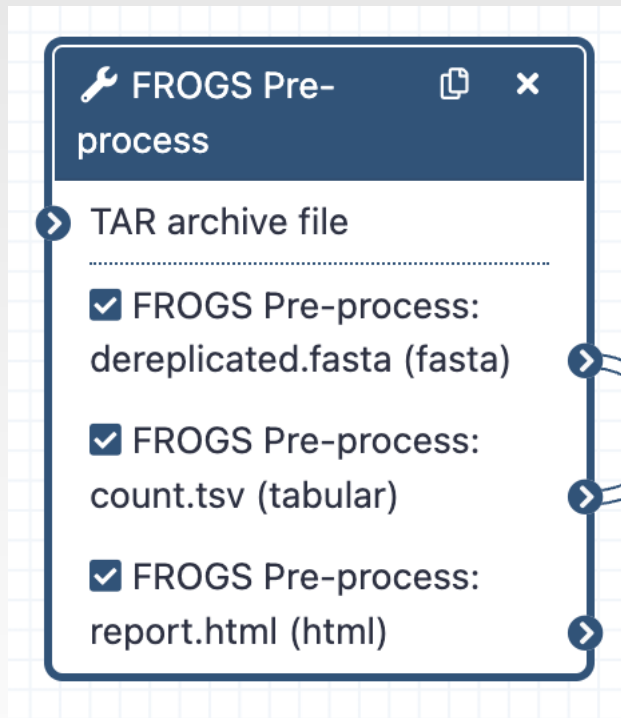
For each tool, think to:

1. Set parameters
2. Rename output files

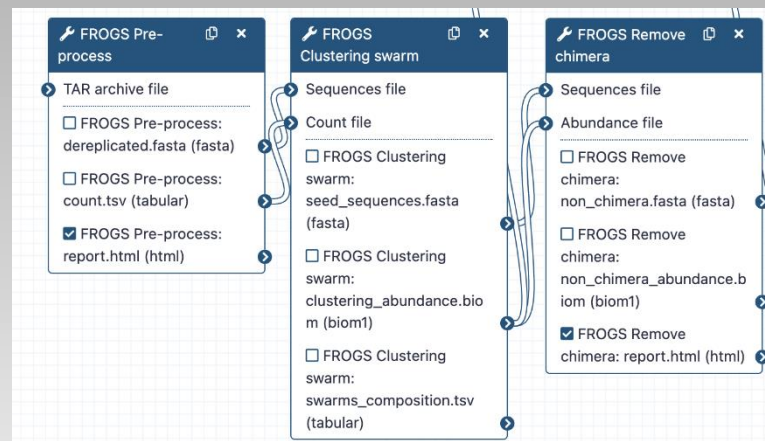


For each tool, think to:

1. Set parameters
2. Rename output files
3. Hide intermediate files

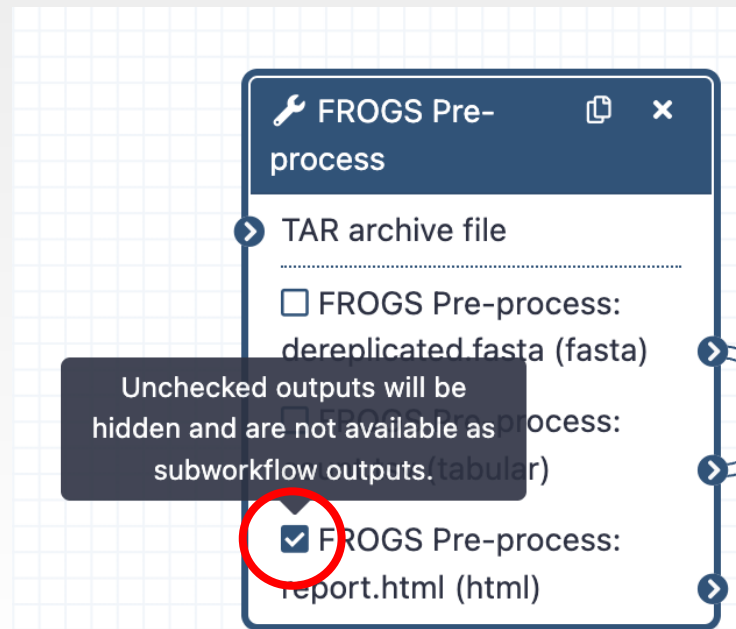


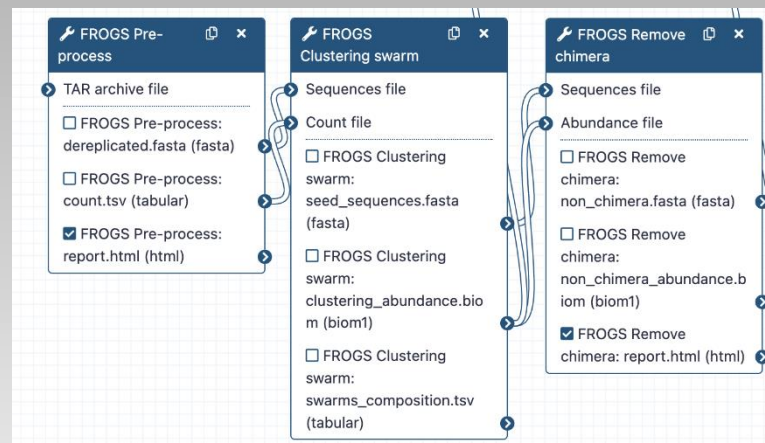
<u>11: FROGS Remove chimera: report.html</u>	👁️ ✎️ ✕
<u>10: FROGS Remove chimera: non chimera abundance.biom</u>	👁️ ✎️ ✕
<u>9: FROGS Remove chimera: non chimera.fasta</u>	👁️ ✎️ ✕



For each tool, think to:

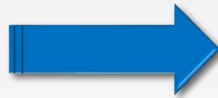
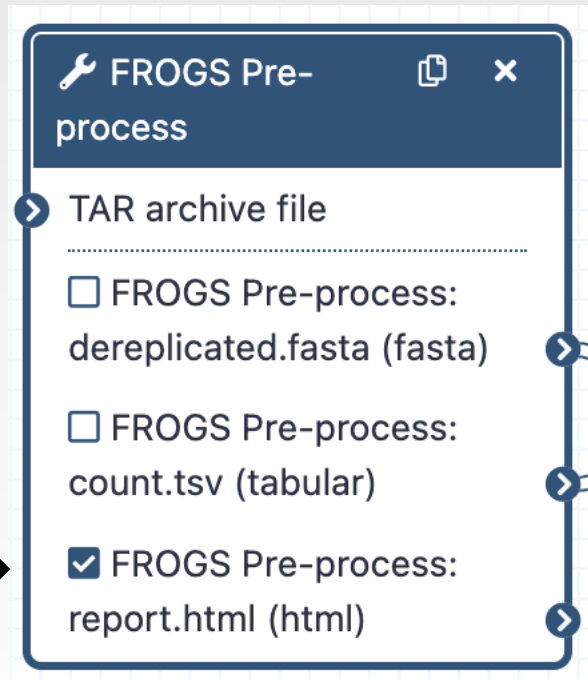
1. Set parameters
2. Rename output files
3. Hide intermediate files



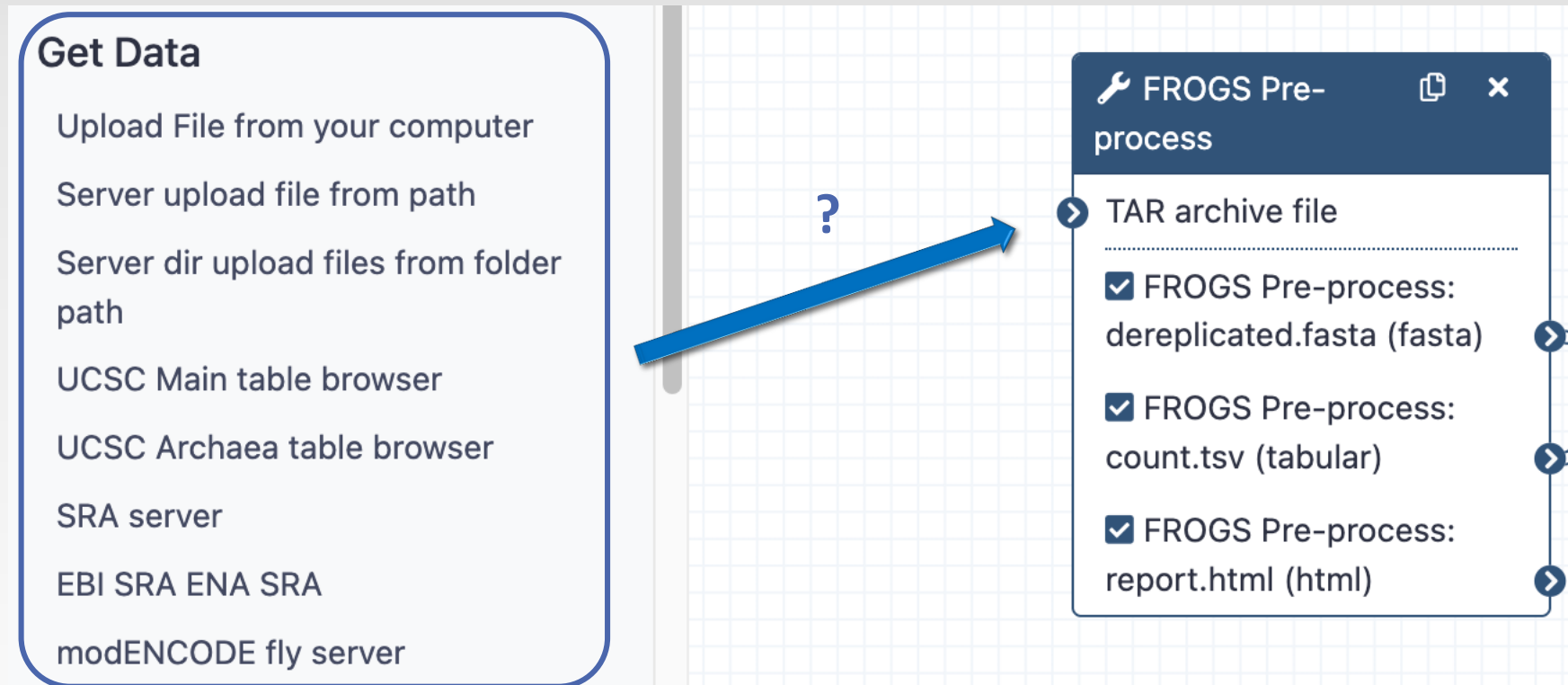


For each tool, think to:

1. Set parameters
2. Rename output files
3. Hide intermediate files

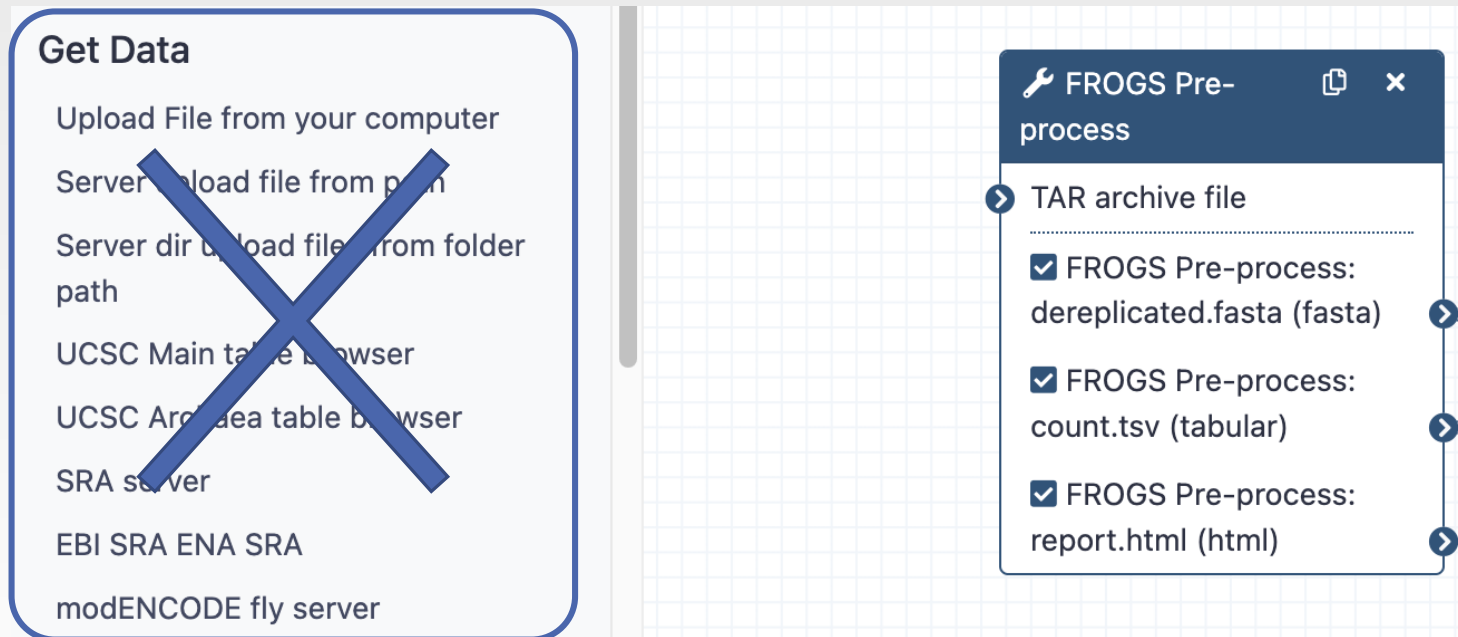


Could you integrate « upload file » in the workflow ?

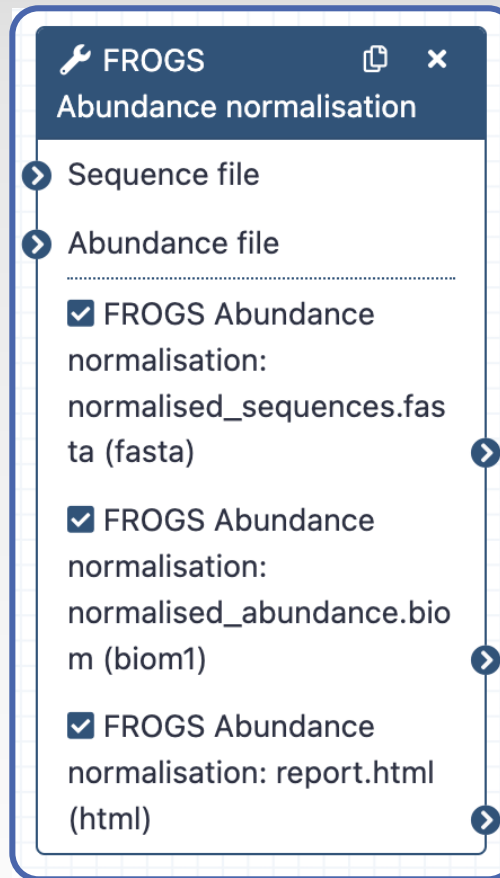


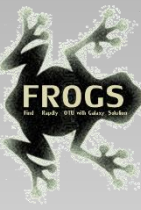
Could you integrate « upload file » in the workflow ?

Upload file cannot be automatized because the workflow, at each run, will be processed with different input data

The image shows a screenshot of a workflow editor interface. On the left, a panel titled 'Get Data' contains a list of data sources: 'Upload File from your computer', 'Server upload file from path', 'Server dir upload files from folder path', 'UCSC Main table browser', 'UCSC Archive table browser', 'SRA server', 'EBI SRA ENA SRA', and 'modENCODE fly server'. A large blue 'X' is drawn over this list. On the right, a workflow step titled 'FROGS Pre-process' is shown. It has a 'TAR archive file' input and three checked checkboxes for outputs: 'FROGS Pre-process: dereplicated.fasta (fasta)', 'FROGS Pre-process: count.tsv (tabular)', and 'FROGS Pre-process: report.html (html)'. The workflow step is connected to the next step in the sequence by a blue arrow.

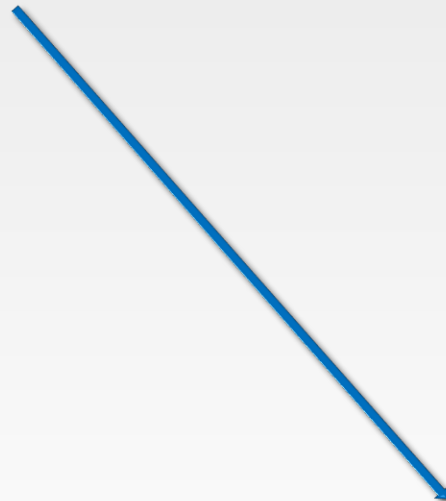
Could you integrate « Normalisation tool » in the workflow ?





Could you integrate « Normalisation tool » in the workflow ?

Yes but only if you select
« sampling by the number of
sequence of the smallest sample »



FROGS Abundance normalisation Normalise OTU abundance. (Galaxy Version 4.0.0+ galaxy1) ☆ ▼

Label

Add a step label.

Step Annotation

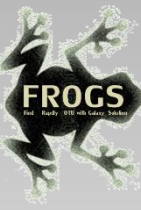
Add an annotation or notes to this step. Annotations are available when a workflow is viewed.

Sequence file
Data input 'input_fasta' (fasta).
Sequence file to normalise (format: fasta).

Abundance file
Data input 'input_biom' (biom1).
Abundance file to normalise (format: BIOM).

Sampling method
 Sampling by the number of sequences of the smallest sample
 Select a number of sequences

Sampling by the number of sequences of the smallest sample, or select a number manually



Exercise

When your workflow is built

1. Run your own workflow with ITS data with :

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/ITS1.tar.gz

2. Import metadata for statistics analyses

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/Webinar_data/metadata ITS.tsv

3. Run FROGS_stat tools