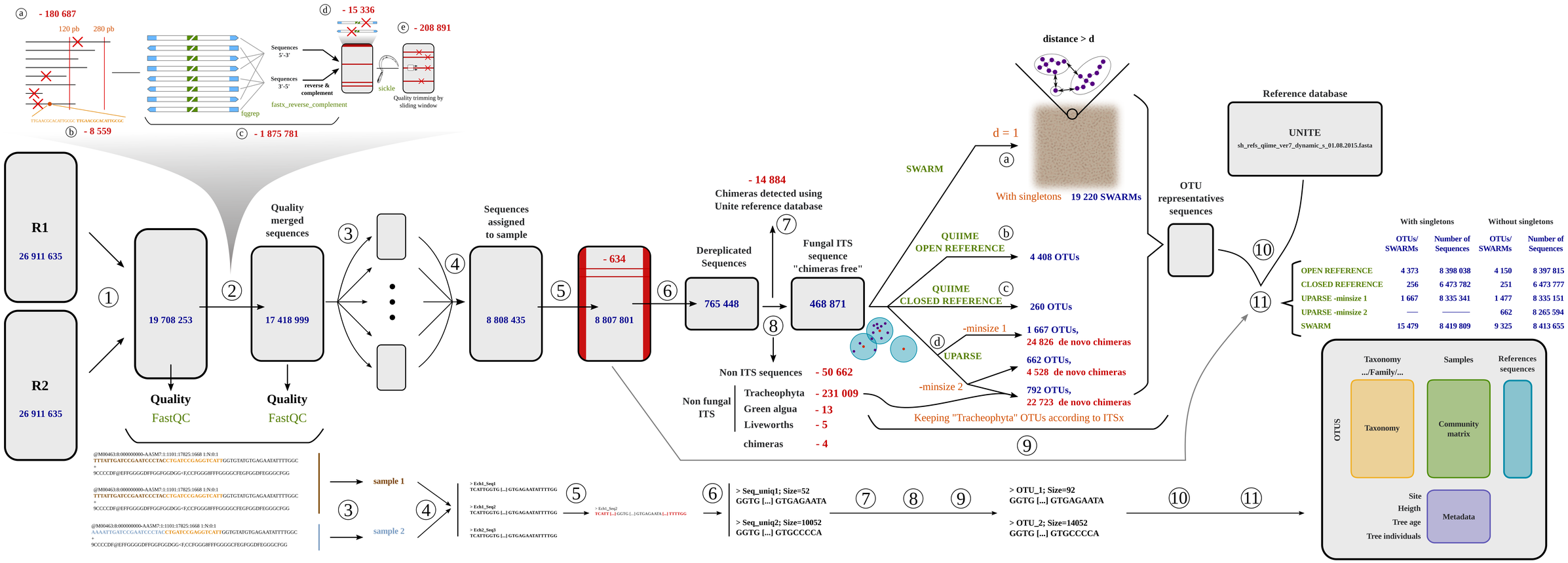



## Legend

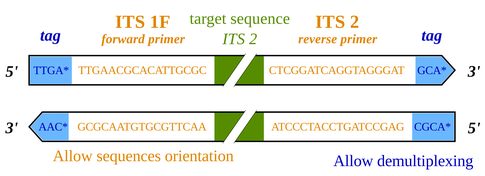
Software or script source

Commands

Option values when different from default

Number of sequences / clusters conserved

Number of sequences discarded




### 1 Merging sequences R1 and R2

| | | |
|---|---|---|
| qiime | join_paired_ends.py | 19 708 253 |

### 2 Quality filtering and orientation

**a Exclude sequences shorter than 180 bp and cut sequences to 280 bp.**

| | | |
|---|---|---|
| fastx | fastx_trimmer | |
| | -f 1 -l 281 | 19 708 253 |
| | -t 1 -m 180 | 19 527 566 |

**b Exclude sequences with primer artefacts**

*Exclude sequences with multiple primers*

| | | |
|---|---|---|
| Balint et al. 2014 | remove_multiprimer.py | 19 519 016 |

*Exclude sequences with multiple primers fw and rev*

| | | |
|---|---|---|
| Unix | grep | 19 519 007 |

**c Reorient sequences in 5'-3'**

| | | |
|---|---|---|
| fqgrep | fqgrep | 17 643 226 |
| fastx | fastx_reverse_complement | |

**d Exclude sequences with 3'-5' primers**

| | | |
|---|---|---|
| fqgrep | fqgrep | 17 627 890 |

**e Exclude bad quality sequences**

| | | |
|---|---|---|
| sickle | sickle | 17 418 999 |
| | -q 33 | |

### 3 Demultiplexing (81 samples)

| | | |
|---|---|---|
| Balint et al. 2014 | demultiplex.sh | |
| | *Tag forward* | 13 206 612 |
| | *Tag forward + reverse* | 8 808 435 |

*demultiplex.sh convert the sequences in fasta format*

### 4 Renamed each sequences with sample names and merge all sequences

| | | |
|---|---|---|
| Balint et al. 2014 | renamed.pl | 8 808 435 |

### 5 Supress primer and tags

| | | |
|---|---|---|
| fastx | fastx_trimmer | |
| | -f 32 | 8 808 435 |
| | -t 31 -m 120 | 8 807 801 |

*Delete 31 bases on each side and exclude sequences shorter than 120 bases*

### 6 Dereplicate sequences

| | | |
|---|---|---|
| SWARM | script bash | 765 448 |

### 7 Chimeras detection using database sequences

| | | |
|---|---|---|
| Usearch | uchime_ref | 750 564 |

### 8 Select fungal ITS sequences

| | | |
|---|---|---|
| ITSx | perl ITSx | 468 871 |

### 9 OTUs clustering

**a Clustering by network (swarm algorithm)**

| | | |
|---|---|---|
| SWARM | ./swarm | |
| | -f -t 4 -d 1 | 19 220 |

**b Clustering based on Qiime open Ref (97%)**

| | | |
|---|---|---|
| qiime | pick_open_otus | 4 408 |
| | -s 0.97 -m uclust | |

**c Clustering based on Qiime closed Ref (97%)**

| | | |
|---|---|---|
| qiime | pick_closed_otus | 260 |
| | -s 0.97 -m uclust | |

**d Clustering de novo using Uparse (97%)**

| | | |
|---|---|---|
| Usearch | sortbysize | |
| | -minsize 1 -minsize 2 | |
| Usearch | cluster_otus | 662 -minsize 2<br>1667 -minsize 1 |
| | --maxaccepts 0 | |
| | --maxrejects 0 | |

### 10 Assign taxonomy

| | | |
|---|---|---|
| a | qiime | assign_taxonomy.py |
| | RDP | -c 0.80 -m rdp |
| b | qiime | assign_taxonomy.py |
| | uclust | -similarity 0.90 -m uclust |
| c | qiime | assign_taxonomy.py |
| | mothur | -c 0.80 -m mothur |
| d | qiime | assign_taxonomy.py |
| | blast | -e 0.001 -m blast |

### 11 Format data for phyloseq package

*Build biom table using OTUs references sequences and the quality sequences obtained just before dereplication.*

| | |
|---|---|
| Usearch | -usearch_global |
| -strand plus -id 0.97 -top_hit_only | |

*Convert in biom format*

| | |
|---|---|
| Usearch | uc2otutab.py |
| biom | convert |