

How to create a custom reference database for BLAST in SEED2

https://www.mothur.org/wiki/Silva_reference_files

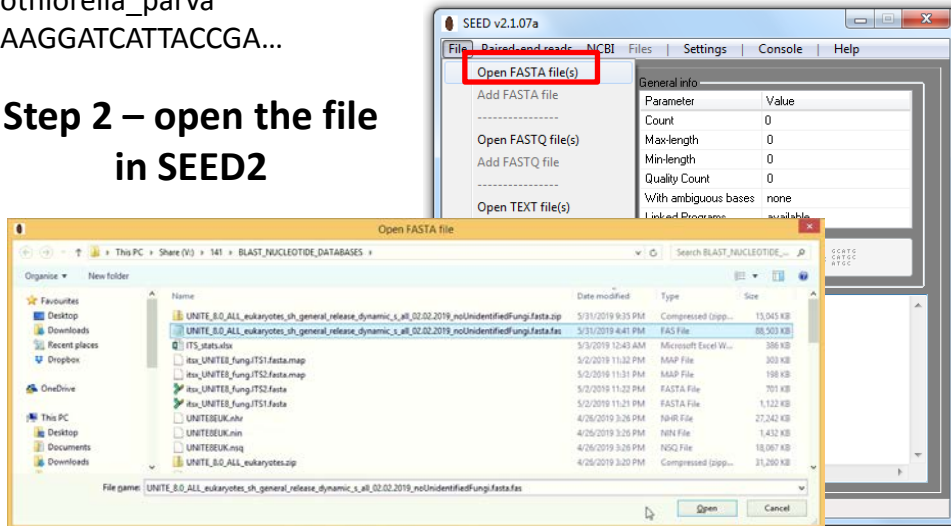
Step 1 – download reference sequences as FASTA or create FASTA file from your edited sequences and put desired information into the titles of the sequences, e.g.:

<Species name> | <ACCESSION#> | <full taxonomy>

Here is example of FASTA:

```
>Thelephora_sp|UDB014120|k__Fungi;p__Basidiomycota;c__Agaricomycetes;o__Thelephorales;f__Thelephoraceae;g__Thelephora;s__Thelephora_sp
GGAAGGATCATTACT...
>Dothiorella_parva|KC898234|k__Fungi;p__Ascomycota;c__Dothideomycetes;o__Botryosphaeriales;f__Botryosphaeriaceae;g__Dothiorella;s__Dothiorella_parva
AAGGATCATTACCGA...
```

Step 2 – open the file in SEED2



Main page
Recent changes
Random page
Help
Tools
What links here
Related changes
Special pages
Printable version
Permanent link
Page information

Page Discussion Read View source View history Search

We will be offering an R workshop December 18-20, 2019. [Learn more.](#)

Silva reference files

If you use the SILVA reference files you should be aware of their [dual-use license](#).

Contents [hide]

- 1 Release 132
- 2 Release 128
- 3 Release 123
- 4 Release 119
- 5 Release 102

Release 132

The SILVA alignment is... sequences as well as arc... reference alignment out... previous version of the S... sequences available for... database that is used for... document where you can read about the process that we used to generate these references.

- Full length sequences and taxonomy references (188247 bacteria, 4626 archaea, and 20246 eukarya sequences). This reference could be customized for alignments, but could also be used for classification. The uncompressed version is ~9.9 GB and the compressed version is 348 MB.

<https://unite.ut.ee/repository.php>

QIIME release (download)

Three sets of QIIME files are released, corresponding to the SHs resulting from clustering at the 97% and 99% threshold levels. The third set of files is the result of a dynamic use of clustering thresholds, such that some SHs are delimited at the 97% level, some at the 97.5% level, some at the 98% level, and so on; these choices were made manually by experts of those particular lineages of fungi. The syntax is the same throughout the three sets of files.

Each SH is given a stable name of the accession number type, here shown in the FASTA file of the dynamic set:

```
>SH099456.05FU_FJ357315_refs
CACAAATATGAGGGCGGGCTGGCACTCCTTGAGAGGACCGGC...
```

SH099456 = accession number of the SH
05FU = global key release 5, organism group Fungi
FJ357315 = GenBank/UNITE accession number of sequence chosen to refs = this is a manually designated Refs
(reps = this is an automatically chosen Reps)

Step 4 – use the reference database for BLAST identification

FASTA FILE(S): ITS_OTUs [SeqCount: 5191 Ambiguous: 0 Min-len: 40 Max-len: 395 Max-qual: 0 Min-qual: 13]

No.	HEAD	Sequences	Quality	display	sort	Alignments	Chimera-clean	Extraction	Clustering	Identification	Phylogeny
1	Otu1	170	0	1		C C A A C C C T C A G G C C C C T T C T G G G C				NCBI BLAST	
2	Otu2	191	0	1		T A T T C T C A A T T C T C T C A G T T T T G C				Taxonomy by name	
3	Otu3	192	0	1		A A T T C T C A A C C T A T A A A C C T T G T				Taxonomy by accession or taxID	
4	Otu4	153	0	1		C A A C C A T C A A G C C C T A G G C T T G T					
5	Otu5	227	0	1		A A T T C T C A A A C C T G T A A C T G G T T T A A C A G C T G G T T T T G G T A T T G					
6	Otu6	210	0	1		A T A A A T C T C A A C C G C A C C C T G C T A T C A T T A G C A A G G T C G T C G G					

select "NCBI BLAST" in "Identification" menu

NCBI BLAST (ITS_OTUs)

Run BLAST (settings)

Load BLAST results

Save BLAST results

Load PluotF BLAST results

STATUS	Sel Result	Accession	Description

BLAST settings

MYREFERENCE
 UNITEBALLEUK

DATABASE SOURCE

remote (NCBI)
 local

Select local database file(s)

number of threads: 1

F:\LAB\FM_DATABASE_TEST\
-Specification
 Restrict search with the given Entrez query
NOT (environmental samples[organism] OR metagenomes[organism] OR unidentified[organism])

Choose a BLAST program to run: blastn

Parameters
Type: megablast
Tasks at the same time: 10
Results per sequence: 10
E-value threshold: 0.00001

Restart FAILED and UNDONE RUN (removes previous results)

select the database folder and the reference database

...run BLAST