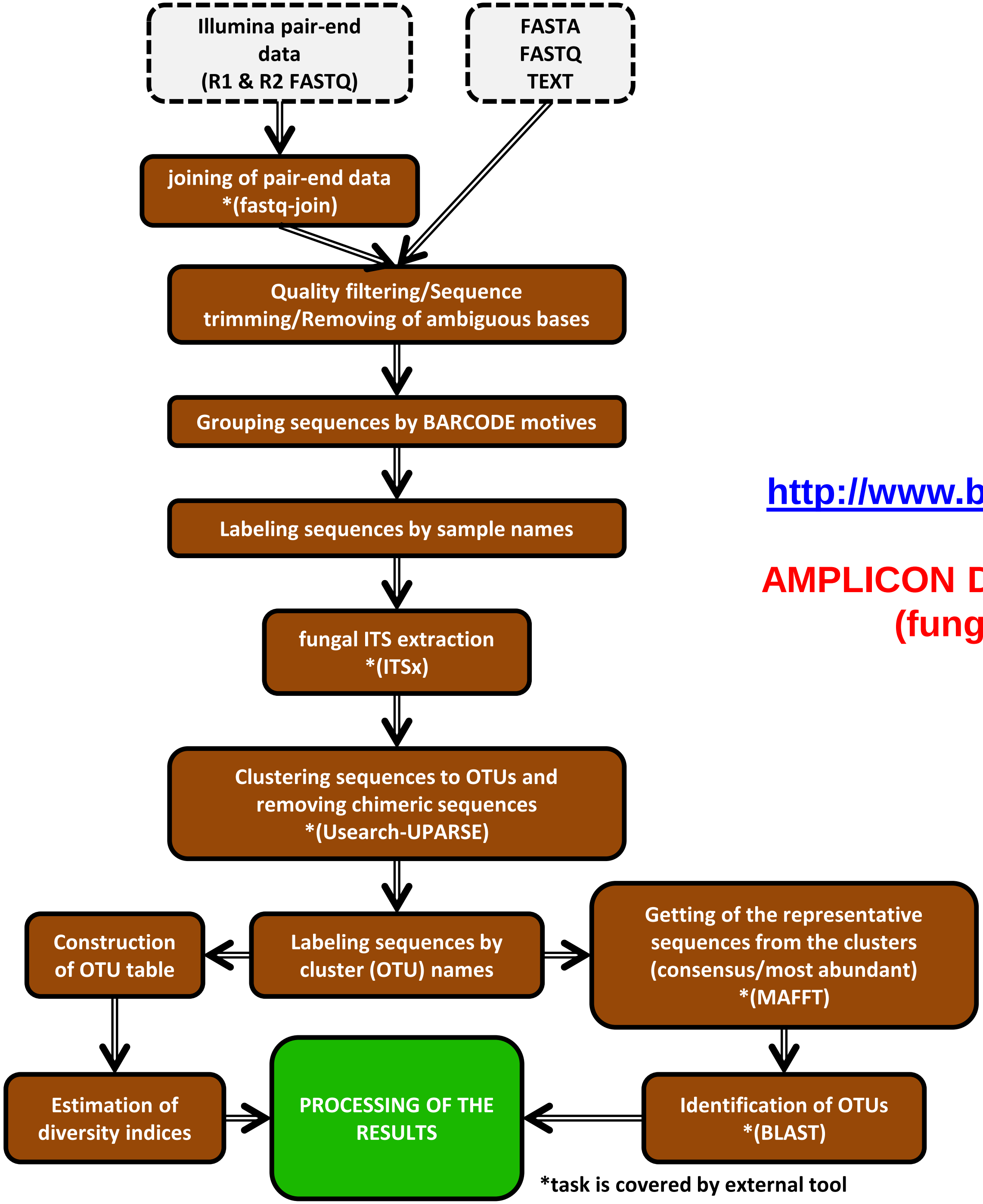


Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Labeling sequences by cluster (OTU) names

Construction of OTU table

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

Estimation of diversity indices

Identification of OTUs *(BLAST)

PROCESSING OF THE RESULTS

*task is covered by external tool

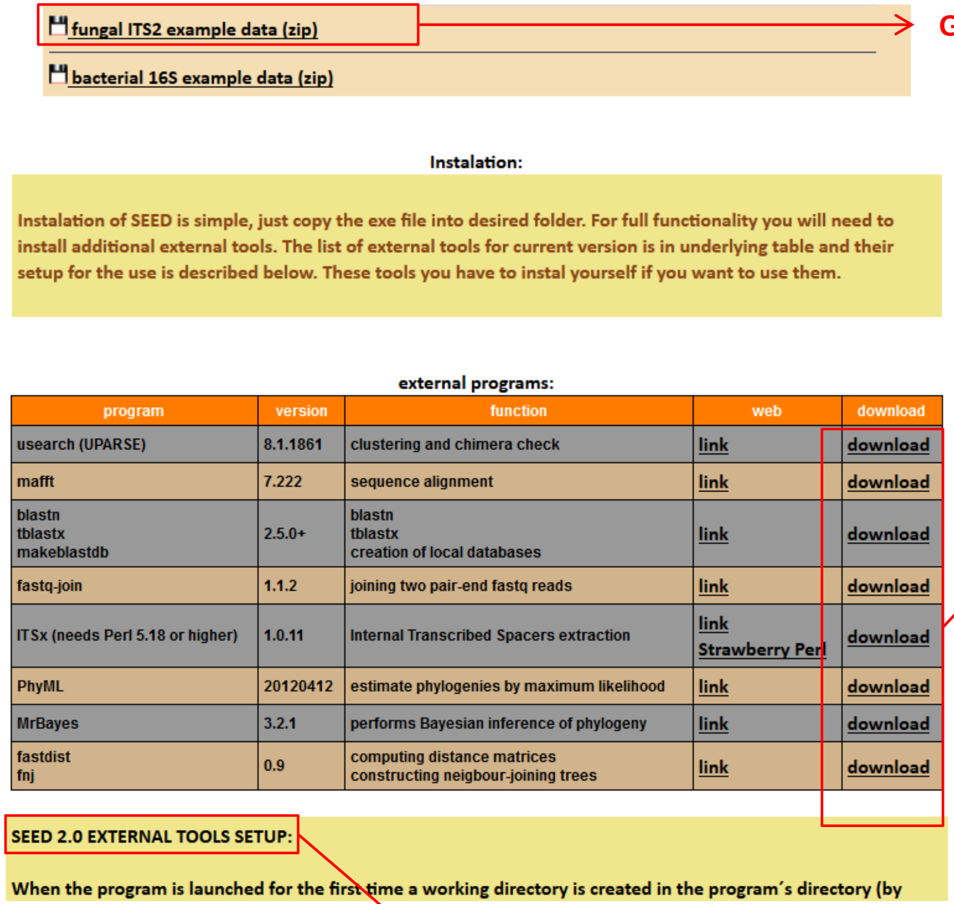
SEED v2.0

http://www.biomed.cas.cz/mbu/lbwrf/seed/

AMPLICON DATA PROCESSING TUTORIAL (fungal amplicons example)

Tomáš Větrovský
vetrovsky@biomed.cas.cz
Laboratory of Environmental Microbiology
Institute of Microbiology of the Academy of Sciences of the Czech Republic

http://www.biomed.cas.cz/mbu/lbwrf/seed/seed2.0.php



fungal ITS2 example data (zip)

bacterial 16S example data (zip)

Instalation:

Instalation of SEED is simple, just copy the exe file into desired folder. For full functionality you will need to install additional external tools. The list of external tools for current version is in underlying table and their setup for the use is described below. These tools you have to instal yourself if you want to use them.

external programs:

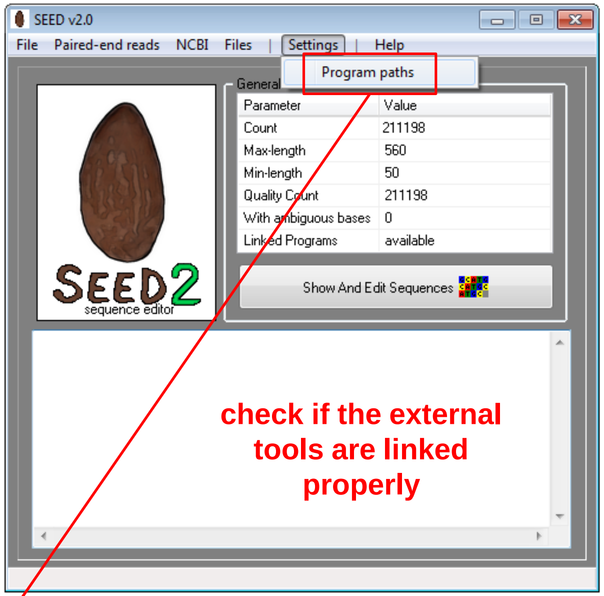
program	version	function	web	download
usearch (UPARSE)	8.1.1861	clustering and chimera check	link	download
mafft	7.222	sequence alignment	link	download
blastn tblastx makeblastdb	2.5.0+	blastn tblastx creation of local databases	link	download
fastq-join	1.1.2	joining two pair-end fastq reads	link	download
ITSx (needs Perl 5.18 or higher)	1.0.11	Internal Transcribed Spacers extraction	link Strawberry Perl	download
PhyML	20120412	estimate phylogenies by maximum likelihood	link	download
MrBayes	3.2.1	performs Bayesian inference of phylogeny	link	download
fastdist fnj	0.9	computing distance matrices constructing neigbour-joining trees	link	download

SEED 2.0 EXTERNAL TOOLS SETUP:

When the program is launched for the first time a working directory is created in the program´s directory (by

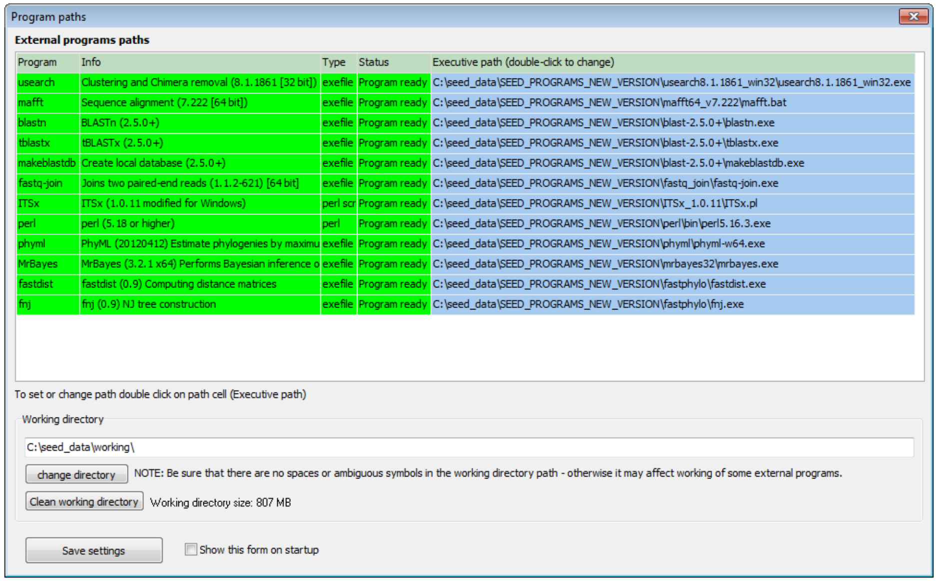
Get example data...

Download external tools



Parameter	Value
Count	211198
Max-length	560
Min-length	50
Quality Count	211198
With ambiguous bases	0
Linked Programs	available

check if the external tools are linked properly



External programs paths

Program	Info	Type	Status	Executive path (double-click to change)
usearch	Clustering and Chimera removal (8.1.1861 [32 bit])	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\usearch8.1.1861_win32\usearch8.1.1861_win32.exe
mafft	Sequence alignment (7.222 [64 bit])	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\mafft64_v7.222\mafft.bat
blastn	BLASTn (2.5.0+)	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\blast-2.5.0+\blastn.exe
tblastx	tBLASTx (2.5.0+)	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\blast-2.5.0+\tblastx.exe
makeblastdb	Create local database (2.5.0+)	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\blast-2.5.0+\makeblastdb.exe
fastq-join	Joins two paired-end reads (1.1.2-621) [64 bit]	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\fastq_join\fastq-join.exe
ITSx	ITSx (1.0.11 modified for Windows)	perl scr	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\ITSx_1.0.11\ITSx.pl
perl	perl (5.18 or higher)	perl	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\perl\bin\perl5.16.3.exe
phyml	PhyML (20120412) Estimate phylogenies by maximu	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\phyml\phyml-w64.exe
MrBayes	MrBayes (3.2.1 x64) Performs Bayesian inference o	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\mrbayes32\mrbayes.exe
fastdist	fastdist (0.9) Computing distance matrices	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\fastphylo\fastdist.exe
fnj	fnj (0.9) NJ tree construction	exefile	Program ready	C:\seed_data\SEED_PROGRAMS_NEW_VERSION\fastphylo\fnj.exe

To set or change path double click on path cell (Executive path)

Working directory

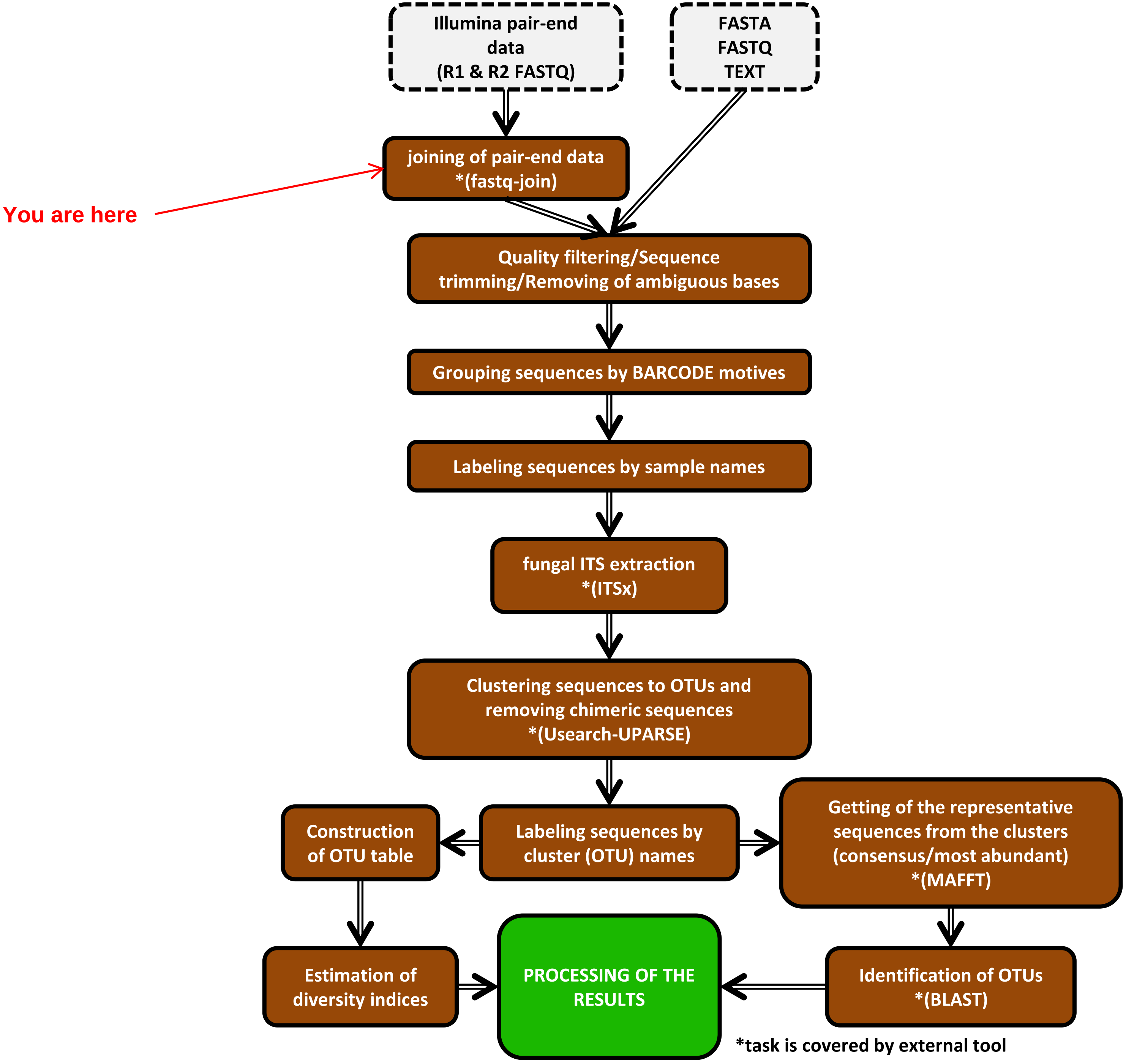
C:\seed_data\working\

NOTE: Be sure that there are no spaces or ambiguous symbols in the working directory path - otherwise it may affect working of some external programs.

Working directory size: 807 MB

Set external tools...

Note:
Windows 8 & Windows 10 - disable SmartScreen to avoid blocking of external tools...



Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

You are here

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

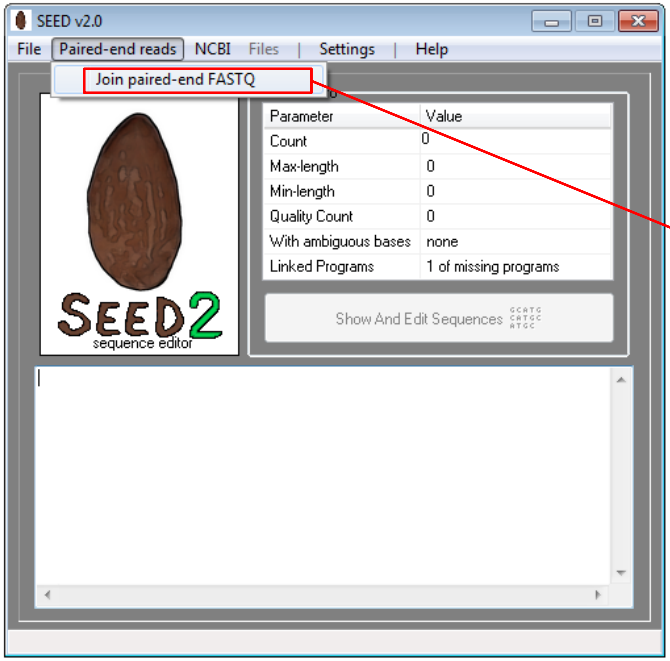
Estimation of diversity indices

PROCESSING OF THE RESULTS

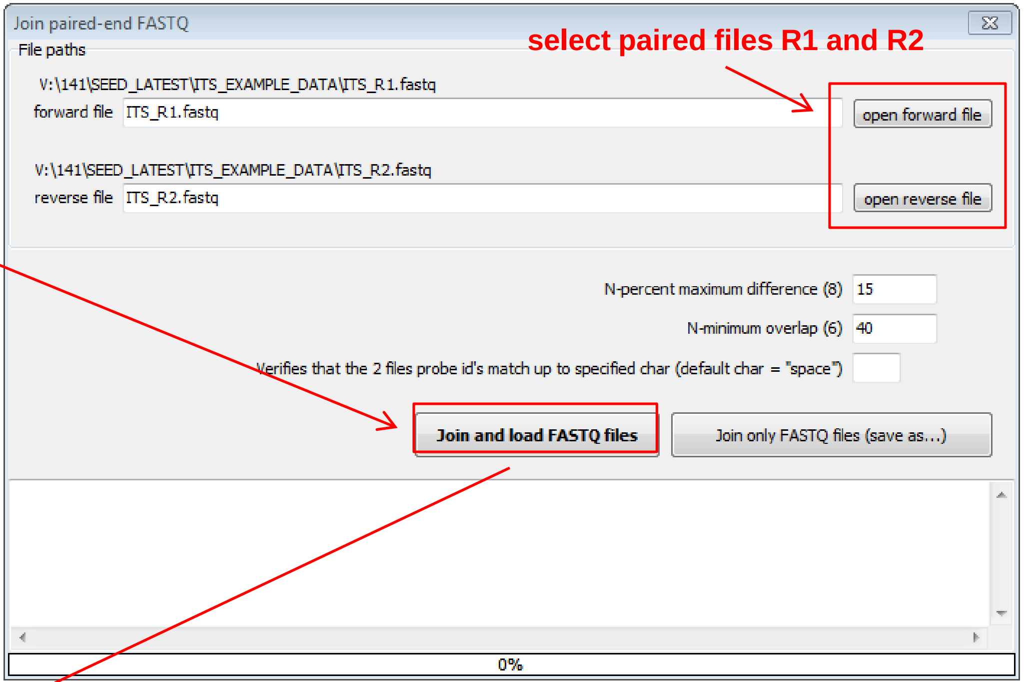
Identification of OTUs *(BLAST)

*task is covered by external tool

Join pared-end Illumina reads…



Parameter	Value
Count	0
Max-length	0
Min-length	0
Quality Count	0
With ambiguous bases	none
Linked Programs	1 of missing programs

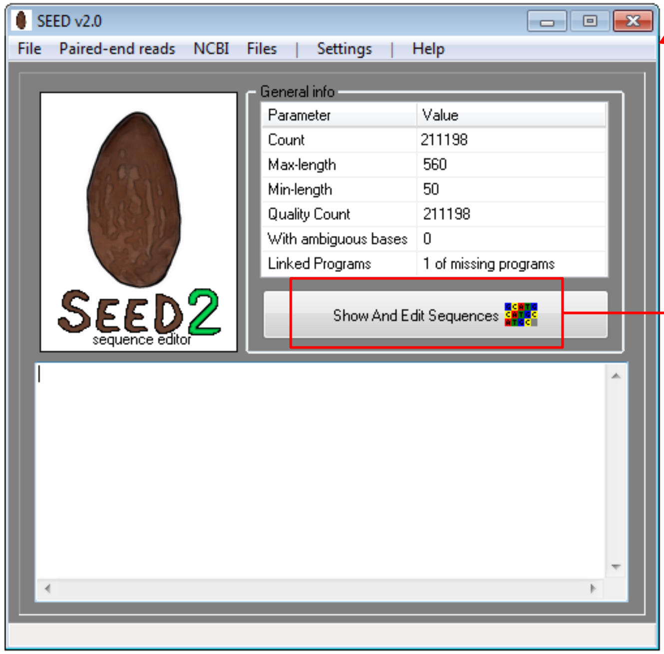


select paired files R1 and R2

V:\141\SEED_LATEST\ITS_EXAMPLE_DATA\ITS_R1.fastq
forward file ITS_R1.fastq

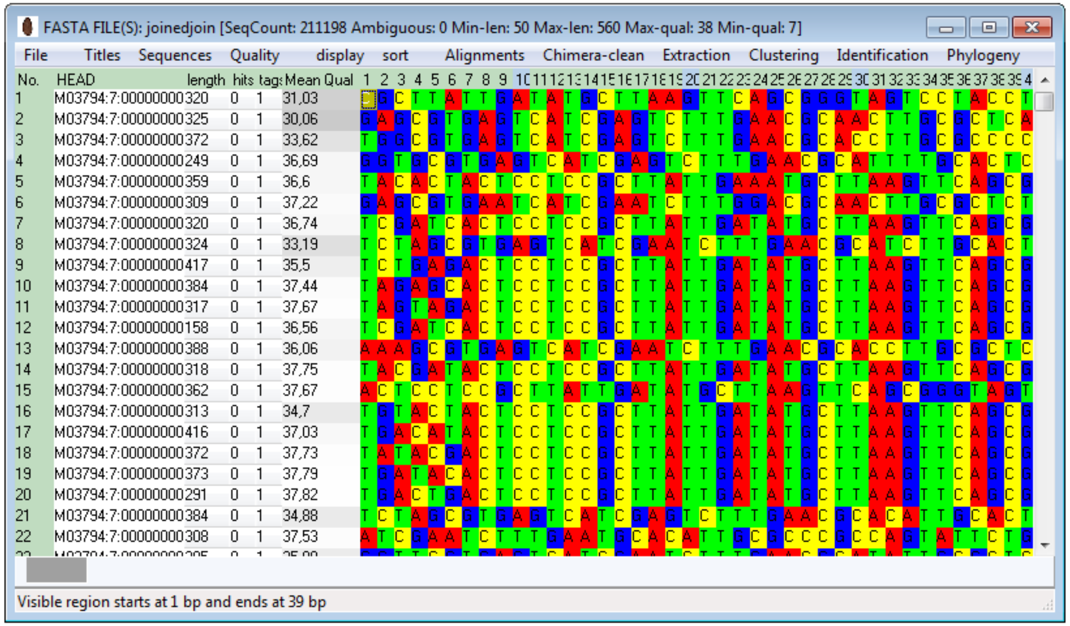
V:\141\SEED_LATEST\ITS_EXAMPLE_DATA\ITS_R2.fastq
reverse file ITS_R2.fastq

N-percent maximum difference (8) 15
N-minimum overlap (6) 40
Verifies that the 2 files probe id's match up to specified char (default char = "space")



General info

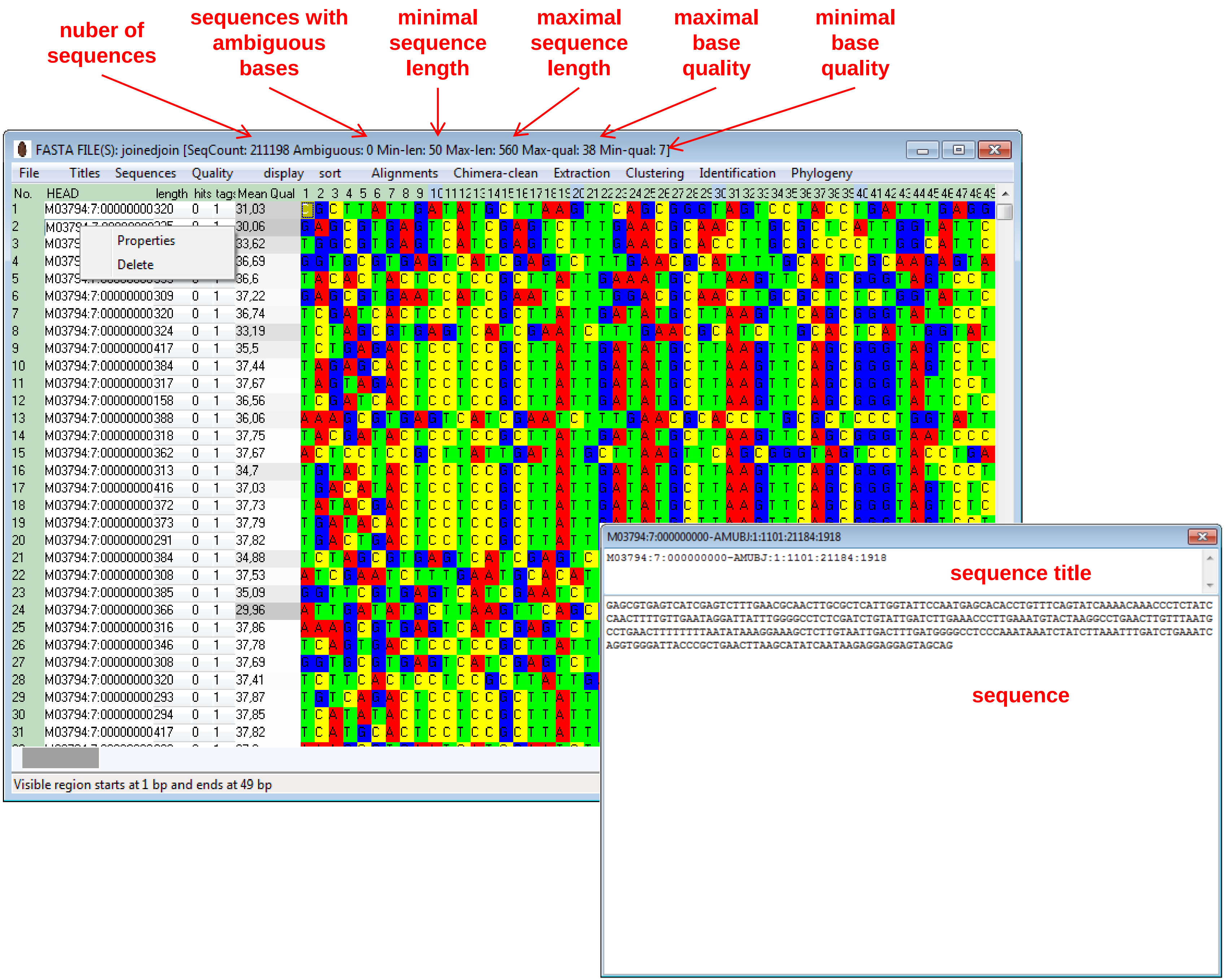
Parameter	Value
Count	211198
Max-length	560
Min-length	50
Quality Count	211198
With ambiguous bases	0
Linked Programs	1 of missing programs



FASTA FILE(S): joinedjoin [SeqCount: 211198 Ambiguous: 0 Min-len: 50 Max-len: 560 Max-qual: 38 Min-qual: 7]

No.	HEAD	length	hits	tags	Mean Qual
1	M03794:7:00000000	320	0	1	31,03
2	M03794:7:00000000	325	0	1	30,06
3	M03794:7:00000000	372	0	1	33,62
4	M03794:7:00000000	249	0	1	36,69
5	M03794:7:00000000	359	0	1	36,6
6	M03794:7:00000000	309	0	1	37,22
7	M03794:7:00000000	320	0	1	36,74
8	M03794:7:00000000	324	0	1	33,19
9	M03794:7:00000000	417	0	1	35,5
10	M03794:7:00000000	384	0	1	37,44
11	M03794:7:00000000	317	0	1	37,67
12	M03794:7:00000000	158	0	1	36,56
13	M03794:7:00000000	388	0	1	36,06
14	M03794:7:00000000	318	0	1	37,75
15	M03794:7:00000000	362	0	1	37,67
16	M03794:7:00000000	313	0	1	34,7
17	M03794:7:00000000	416	0	1	37,03
18	M03794:7:00000000	372	0	1	37,73
19	M03794:7:00000000	373	0	1	37,79
20	M03794:7:00000000	291	0	1	37,82
21	M03794:7:00000000	384	0	1	34,88
22	M03794:7:00000000	308	0	1	37,53

Visible region starts at 1 bp and ends at 39 bp



FASTA FILE(S): joinedjoin [SeqCount: 211198 Ambiguous: 0 Min-len: 50 Max-len: 560 Max-qual: 38 Min-qual: 7]

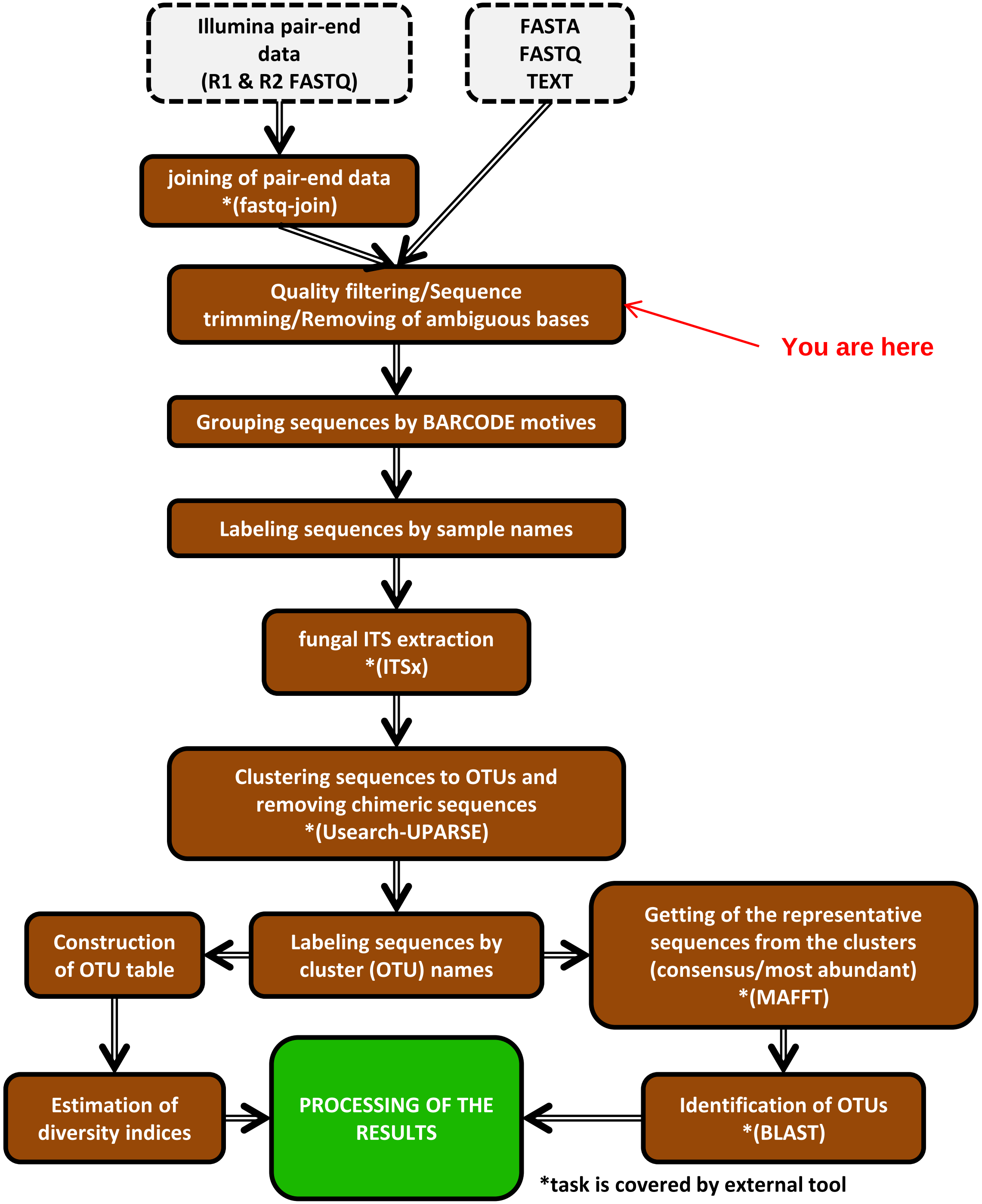
No.	HEAD	length	hits	tags	Mean Qual
1	M03794:7:00000000	320	0	1	31,03
2	M0379	[illegible]	[illegible]	[illegible]	30,06
3	M0379				33,62
4	M0379				36,69
5	M0379	[illegible]	[illegible]	[illegible]	36,6
6	M03794:7:00000000	309	0	1	37,22
7	M03794:7:00000000	320	0	1	36,74
8	M03794:7:00000000	324	0	1	33,19
9	M03794:7:00000000	417	0	1	35,5
10	M03794:7:00000000	384	0	1	37,44
11	M03794:7:00000000	317	0	1	37,67
12	M03794:7:00000000	158	0	1	36,56
13	M03794:7:00000000	388	0	1	36,06
14	M03794:7:00000000	318	0	1	37,75
15	M03794:7:00000000	362	0	1	37,67
16	M03794:7:00000000	313	0	1	34,7
17	M03794:7:00000000	416	0	1	37,03
18	M03794:7:00000000	372	0	1	37,73
19	M03794:7:00000000	373	0	1	37,79
20	M03794:7:00000000	291	0	1	37,82
21	M03794:7:00000000	384	0	1	34,88
22	M03794:7:00000000	308	0	1	37,53
23	M03794:7:00000000	385	0	1	35,09
24	M03794:7:00000000	366	0	1	29,96
25	M03794:7:00000000	316	0	1	37,86
26	M03794:7:00000000	346	0	1	37,78
27	M03794:7:00000000	308	0	1	37,69
28	M03794:7:00000000	320	0	1	37,41
29	M03794:7:00000000	293	0	1	37,87
30	M03794:7:00000000	294	0	1	37,85
31	M03794:7:00000000	417	0	1	37,82

Visible region starts at 1 bp and ends at 49 bp

M03794:7:000000000-AMUBJ:1:1101:21184:1918

```
M03794:7:000000000-AMUBJ:1:1101:21184:1918
```

```
GAGCGTGAGTCATCGAGTCTTTGAACGCAACTTGCGCTCATTGGTATTCCAATGAGCACACCTGTTTCAGTATCAAAACAAACCCTCTATC
CAACTTTTGTTGAATAGGATTATTTGGGGCCTCTCGATCTGTATTGATCTTGAAACCCTTGAAATGTACTAAGGCCTGAACTTGTTTAATG
CCTGAACTTTTTTTAATATAAAGGAAAGCTCTTGTAATTGACTTTGATGGGGCCTCCCAAATAAATCTATCTTAAATTTGATCTGAAATC
AGGTGGGATTACCCGCTGAACTTAAGCATATCAATAAGAGGAGGAGTAGCAG
```



Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

You are here

Grouping sequences by BARCODE motives

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

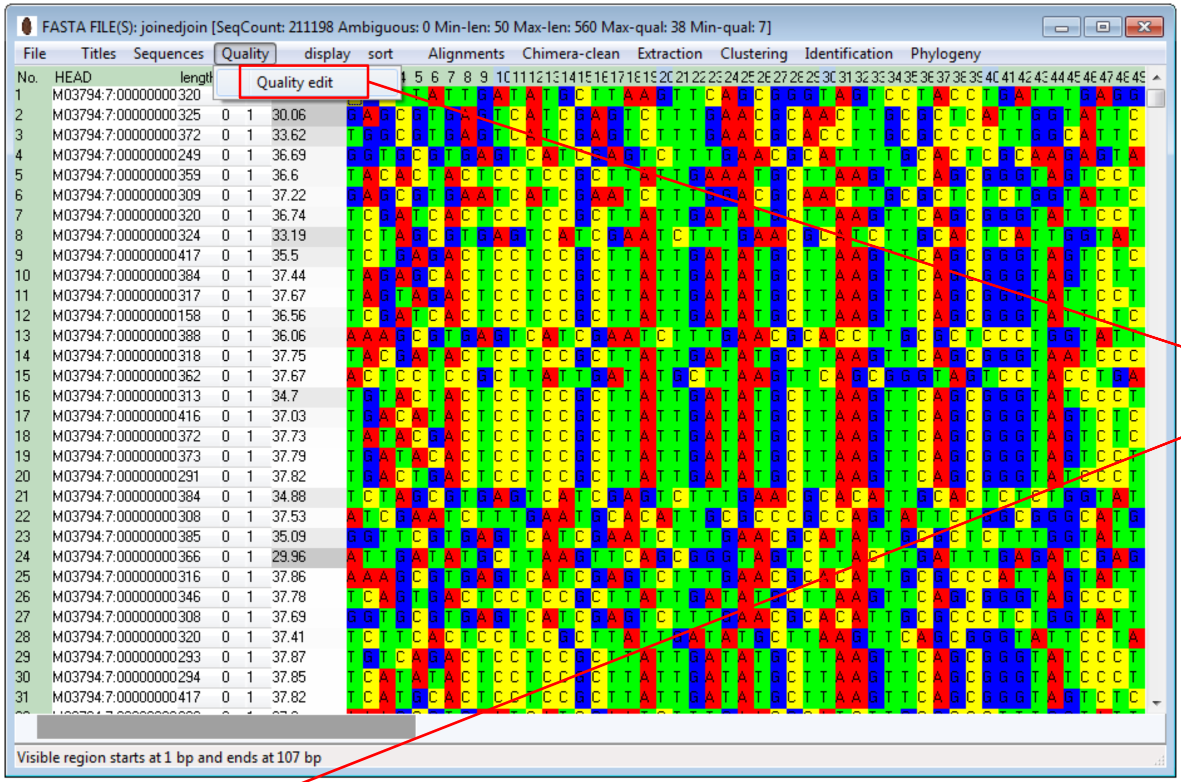
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool

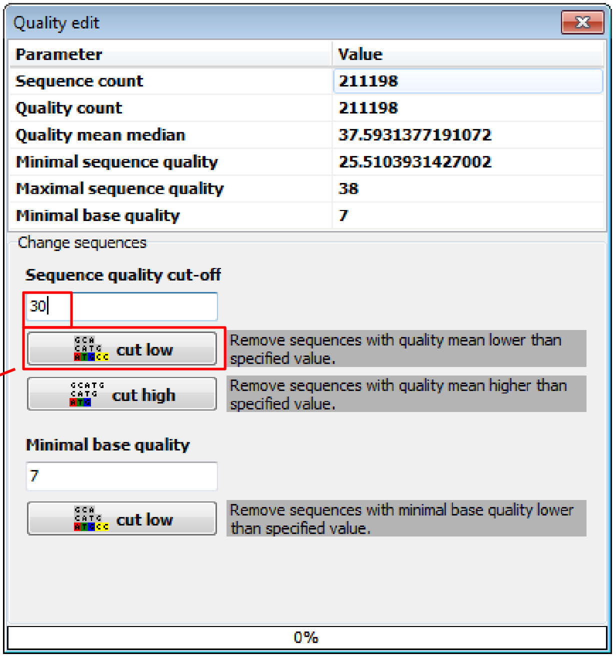
Filter sequences by their quality…



FASTA FILE(S): joinedjoin [SeqCount: 211198 Ambiguous: 0 Min-len: 50 Max-len: 560 Max-qual: 38 Min-qual: 7]

No.	HEAD	length			
1	M03794:7:00000000	320			
2	M03794:7:00000000	325	0	1	30.06
3	M03794:7:00000000	372	0	1	33.62
4	M03794:7:00000000	249	0	1	36.69
5	M03794:7:00000000	359	0	1	36.6
6	M03794:7:00000000	309	0	1	37.22
7	M03794:7:00000000	320	0	1	36.74
8	M03794:7:00000000	324	0	1	33.19
9	M03794:7:00000000	417	0	1	35.5
10	M03794:7:00000000	384	0	1	37.44
11	M03794:7:00000000	317	0	1	37.67
12	M03794:7:00000000	158	0	1	36.56
13	M03794:7:00000000	388	0	1	36.06
14	M03794:7:00000000	318	0	1	37.75
15	M03794:7:00000000	362	0	1	37.67
16	M03794:7:00000000	313	0	1	34.7
17	M03794:7:00000000	416	0	1	37.03
18	M03794:7:00000000	372	0	1	37.73
19	M03794:7:00000000	373	0	1	37.79
20	M03794:7:00000000	291	0	1	37.82
21	M03794:7:00000000	384	0	1	34.88
22	M03794:7:00000000	308	0	1	37.53
23	M03794:7:00000000	385	0	1	35.09
24	M03794:7:00000000	366	0	1	29.96
25	M03794:7:00000000	316	0	1	37.86
26	M03794:7:00000000	346	0	1	37.78
27	M03794:7:00000000	308	0	1	37.69
28	M03794:7:00000000	320	0	1	37.41
29	M03794:7:00000000	293	0	1	37.87
30	M03794:7:00000000	294	0	1	37.85
31	M03794:7:00000000	417	0	1	37.82

Visible region starts at 1 bp and ends at 107 bp



Quality edit

Parameter	Value
Sequence count	211198
Quality count	211198
Quality mean median	37.5931377191072
Minimal sequence quality	25.5103931427002
Maximal sequence quality	38
Minimal base quality	7

Change sequences

Sequence quality cut-off

30

Remove sequences with quality mean lower than specified value.

Remove sequences with quality mean higher than specified value.

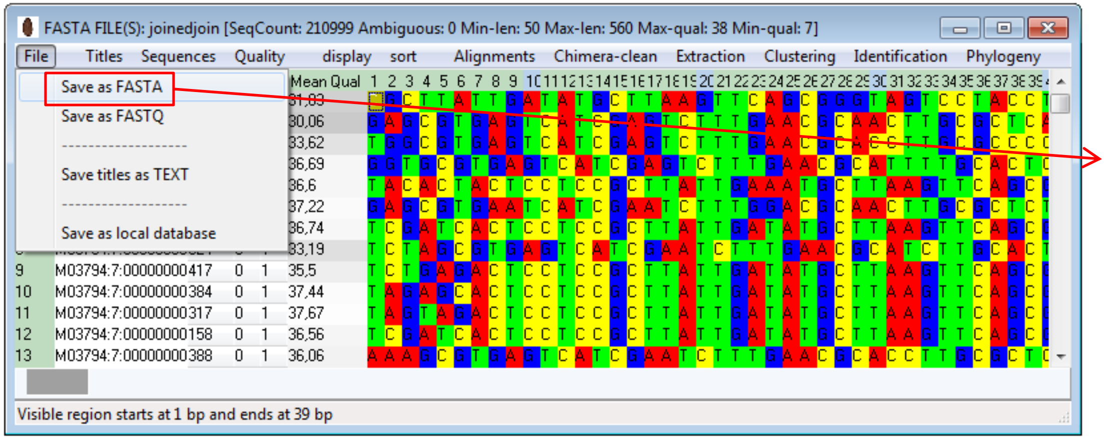
Minimal base quality

7

Remove sequences with minimal base quality lower than specified value.

0%

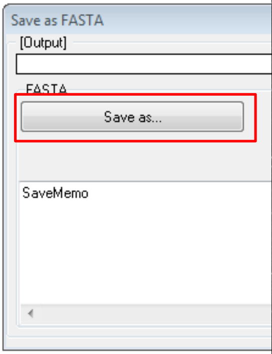
Save files as FASTA after each important step…



FASTA FILE(S): joinedjoin [SeqCount: 210999 Ambiguous: 0 Min-len: 50 Max-len: 560 Max-qual: 38 Min-qual: 7]

					Mean Qual
					31,03
					30,06
					33,62
					36,69
					36,6
					37,22
					36,74
					33,19
9	M03794:7:00000000	417	0	1	35,5
10	M03794:7:00000000	384	0	1	37,44
11	M03794:7:00000000	317	0	1	37,67
12	M03794:7:00000000	158	0	1	36,56
13	M03794:7:00000000	388	0	1	36,06

Visible region starts at 1 bp and ends at 39 bp

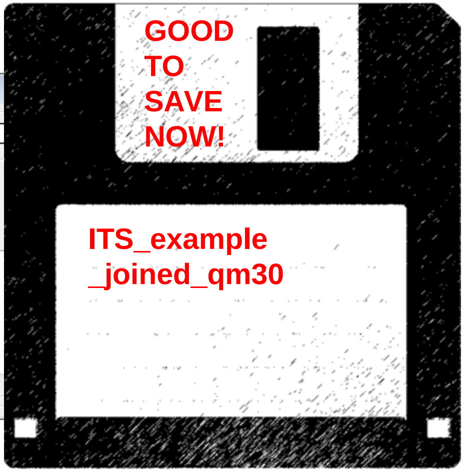


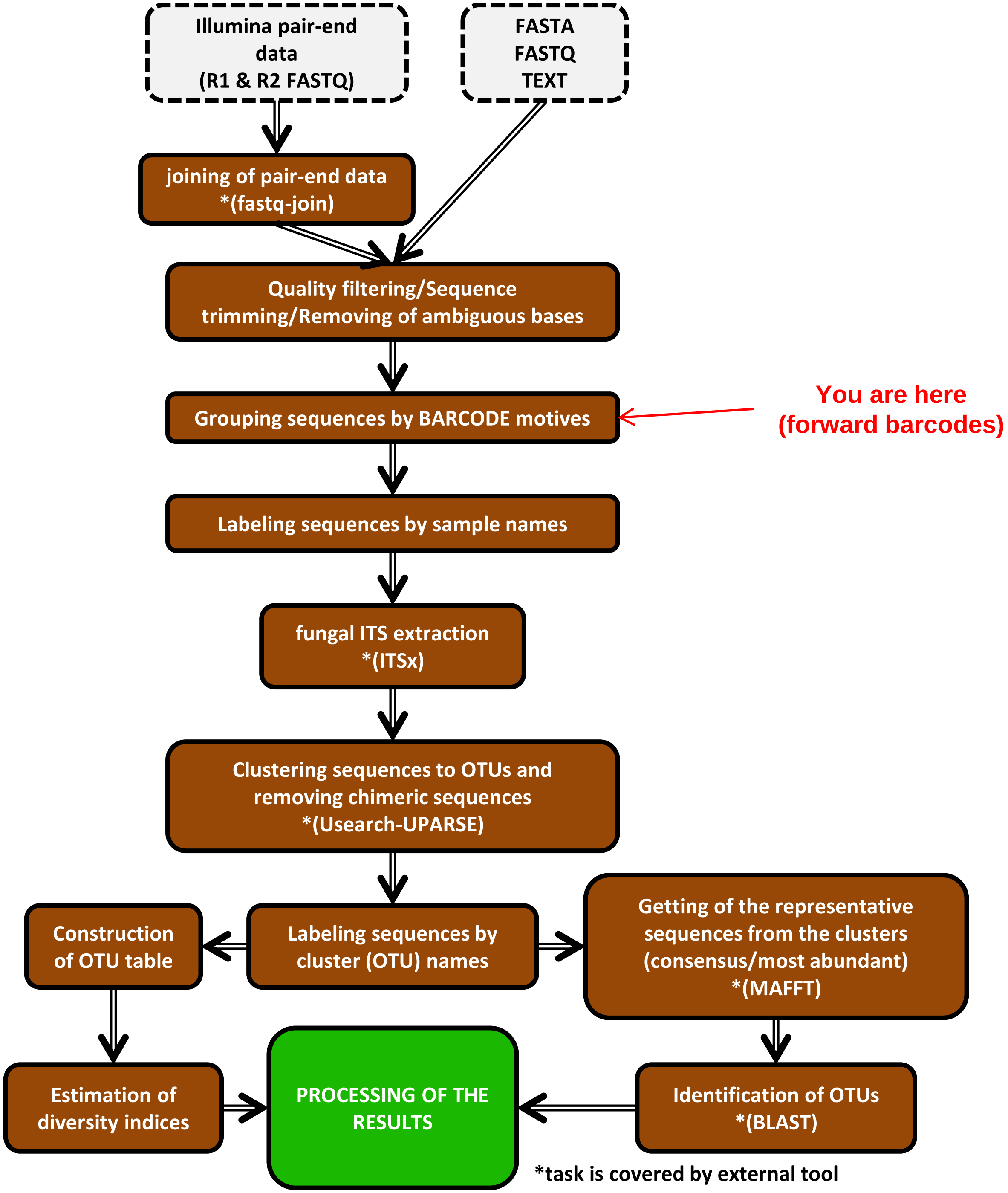
Save as FASTA

[Output]

FASTA

SaveMemo





Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

You are here (forward barcodes)

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool

Forward primer

gITS7	GTGARTCATCGARTCTTTG

TAG SPACER PRIMER

Tagged Forward primers

gITS7_T02	GAGCGTGARTCATCGARTCTTTG
gITS7_T03	TGGCGTGARTCATCGARTCTTTG
gITS7_T06	GGTGCGTGARTCATCGARTCTTTG
gITS7_T08	AAAGCGTGARTCATCGARTCTTTG
gITS7_T10	TCTAGCGTGARTCATCGARTCTTTG

Sequence motive and tag name to search (TAB delimited)

GAGCGTGA	gITS7_T02
TGGCGTGA	gITS7_T03
GGTGCGTGA	gITS7_T06
AAAGCGTGA	gITS7_T08
TCTAGCGTGA	gITS7_T10

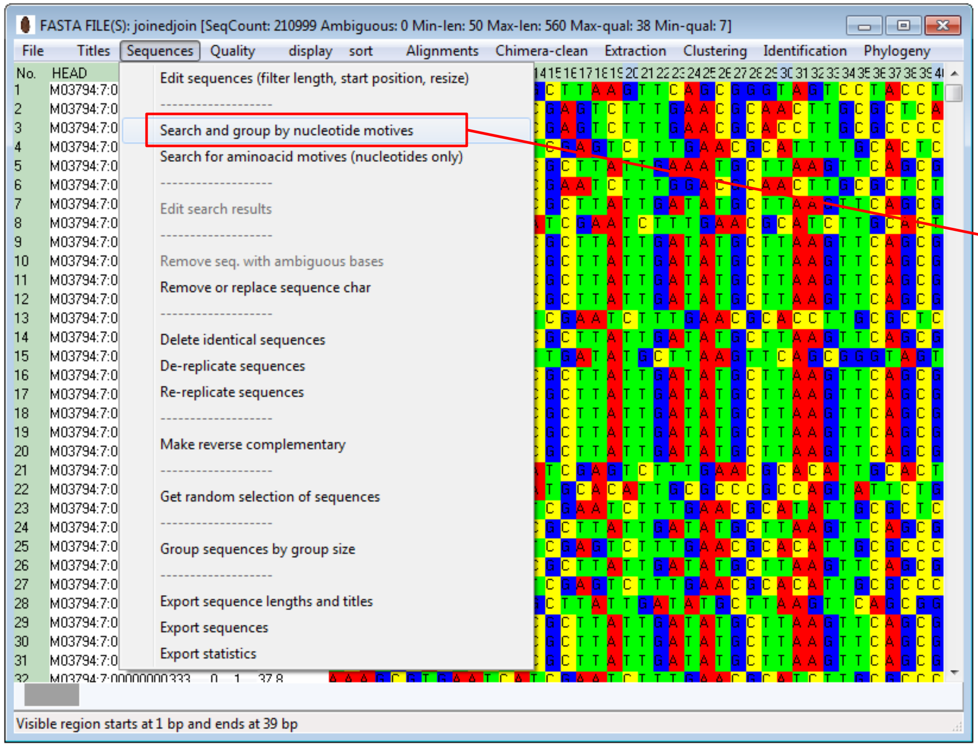
Reverse primer

ITS4	TCCTCCGCTTATTGATATGC

Tagged Reverse primers

ITS4_301	TACGTGACTCCTCCGCTTATTGATATGC
ITS4_302	TCTCGTACTCCTCCGCTTATTGATATGC
ITS4_303	TATACGACTCCTCCGCTTATTGATATGC
ITS4_304	TAGTAGACTCCTCCGCTTATTGATATGC
ITS4_305	TCTCACACTCCTCCGCTTATTGATATGC

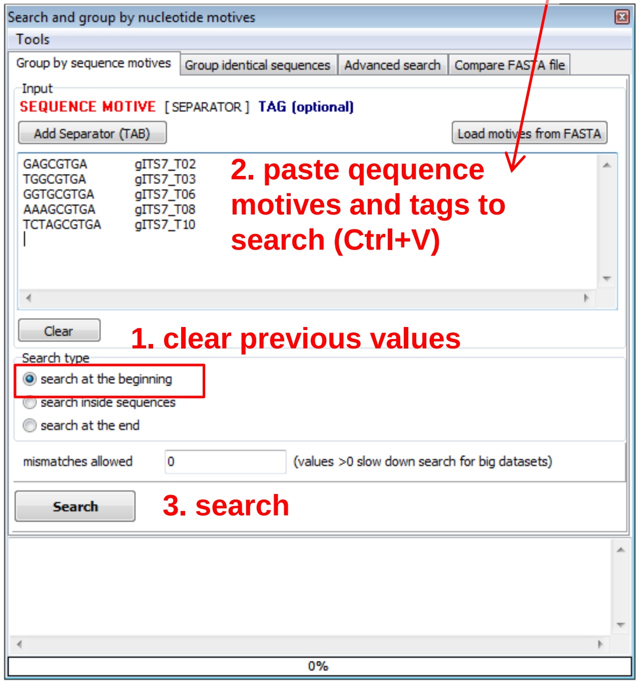
TACGTGACTCCT	ITS4_301
TCTCGTACTCCT	ITS4_302
TATACGACTCCT	ITS4_303
TAGTAGACTCCT	ITS4_304
TCTCACACTCCT	ITS4_305



FASTA FILE(S): joinedjoin [SeqCount: 210999 Ambiguous: 0 Min-len: 50 Max-len: 560 Max-qual: 38 Min-qual: 7]

Visible region starts at 1 bp and ends at 39 bp

Search for the forward tag motives...



Search and group by nucleotide motives

SEQUENCE MOTIVE [SEPARATOR] TAG (optional)

```
GAGCGTGA	gITS7_T02
TGGCGTGA	gITS7_T03
GGTGCGTGA	gITS7_T06
AAAGCGTGA	gITS7_T08
TCTAGCGTGA	gITS7_T10
```

Search type
● search at the beginning
○ search inside sequences
○ search at the end

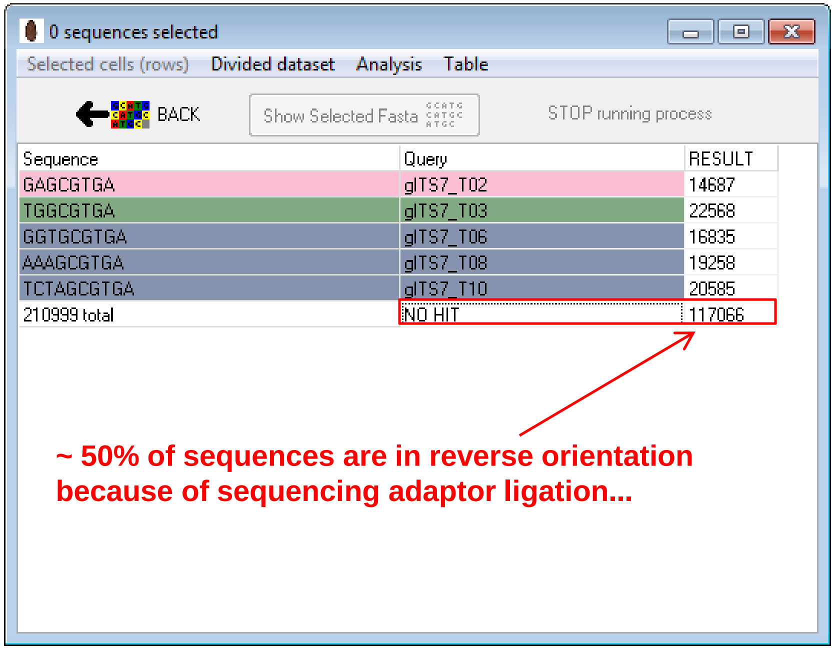
mismatches allowed: 0 (values >0 slow down search for big datasets)

0%

2. paste qequence motives and tags to search (Ctrl+V)

1. clear previous values

3. search

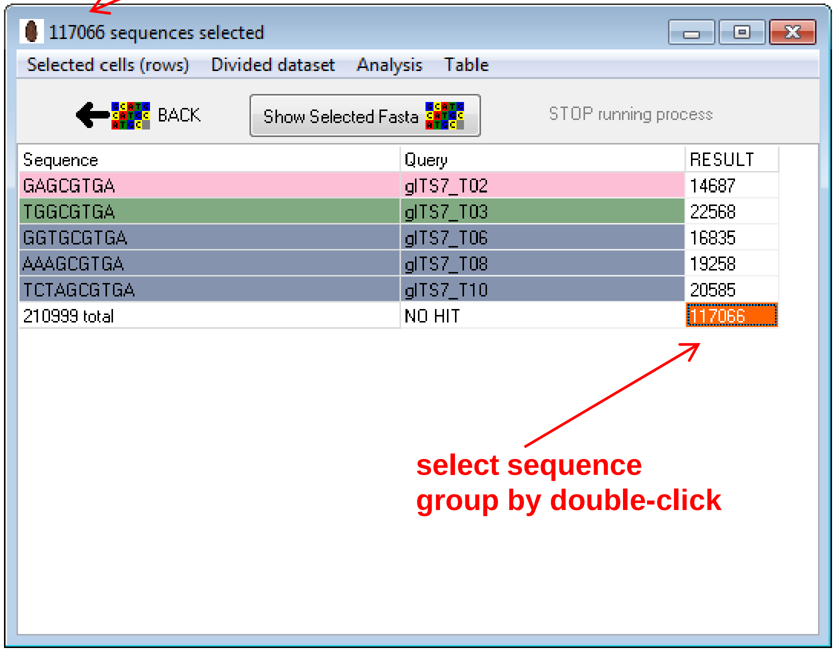


0 sequences selected

Sequence	Query	RESULT
GAGCGTGA	gITS7_T02	14687
TGGCGTGA	gITS7_T03	22568
GGTGCGTGA	gITS7_T06	16835
AAAGCGTGA	gITS7_T08	19258
TCTAGCGTGA	gITS7_T10	20585
210999 total	NO HIT	117066

~ 50% of sequences are in reverse orientation because of sequencing adaptor ligation...

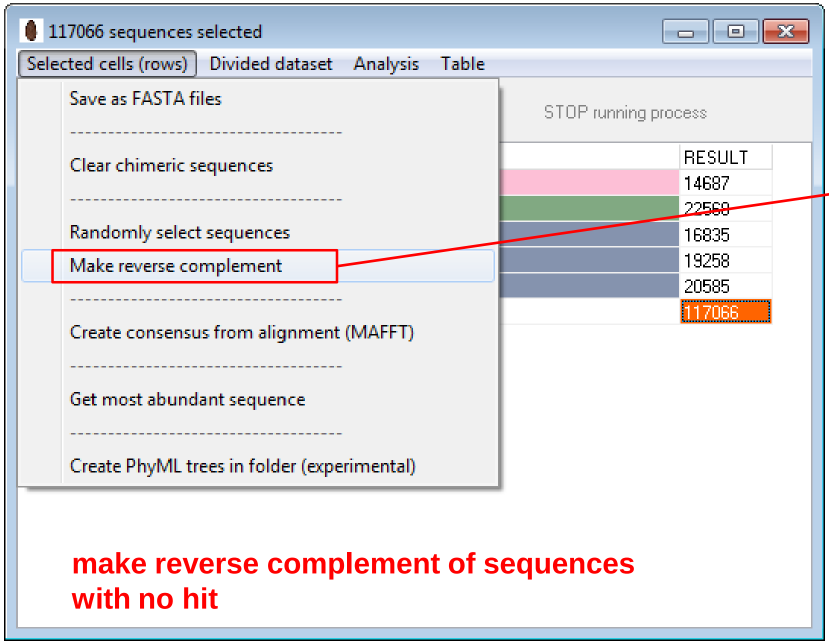
total nuber of selected sequences



117066 sequences selected

Sequence	Query	RESULT
GAGCGTGA	gITS7_T02	14687
TGGCGTGA	gITS7_T03	22568
GGTGCGTGA	gITS7_T06	16835
AAAGCGTGA	gITS7_T08	19258
TCTAGCGTGA	gITS7_T10	20585
210999 total	NO HIT	117066

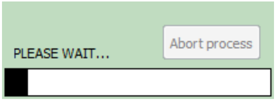
select sequence group by double-click



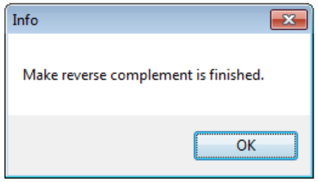
117066 sequences selected

make reverse complement of sequences with no hit

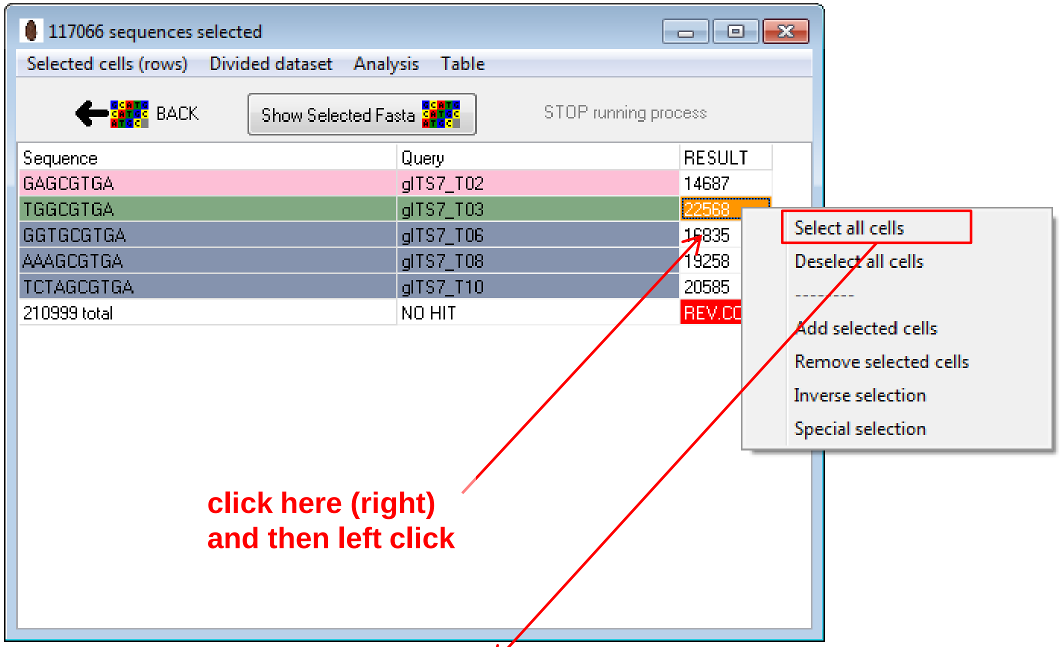
it may take a while...



...and then...

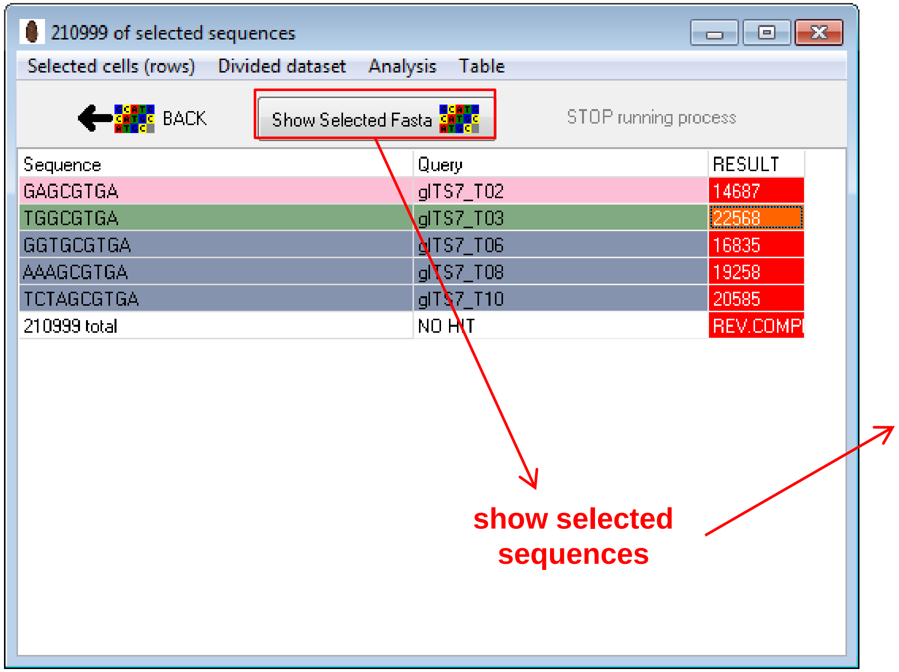


Make reverse complement is finished.



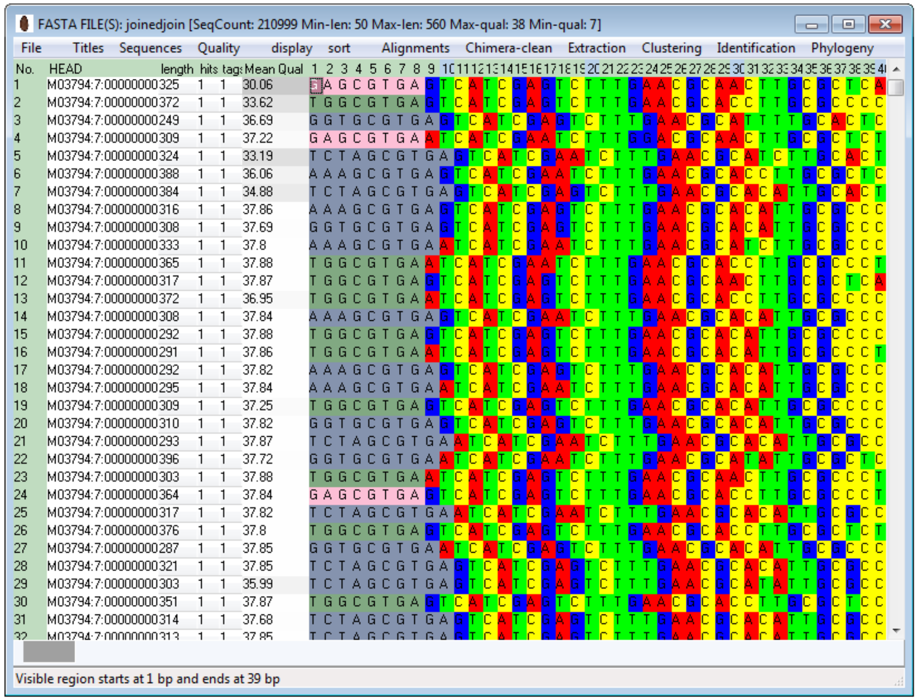
Sequence	Query	RESULT
GAGCGTGA	glTS7_T02	14687
TGGCGTGA	glTS7_T03	22568
GGTGCGTGA	glTS7_T06	16835
AAAGCGTGA	glTS7_T08	19258
TCTAGCGTGA	glTS7_T10	20585
210999 total	NO HIT	REV.CO

click here (right) and then left click



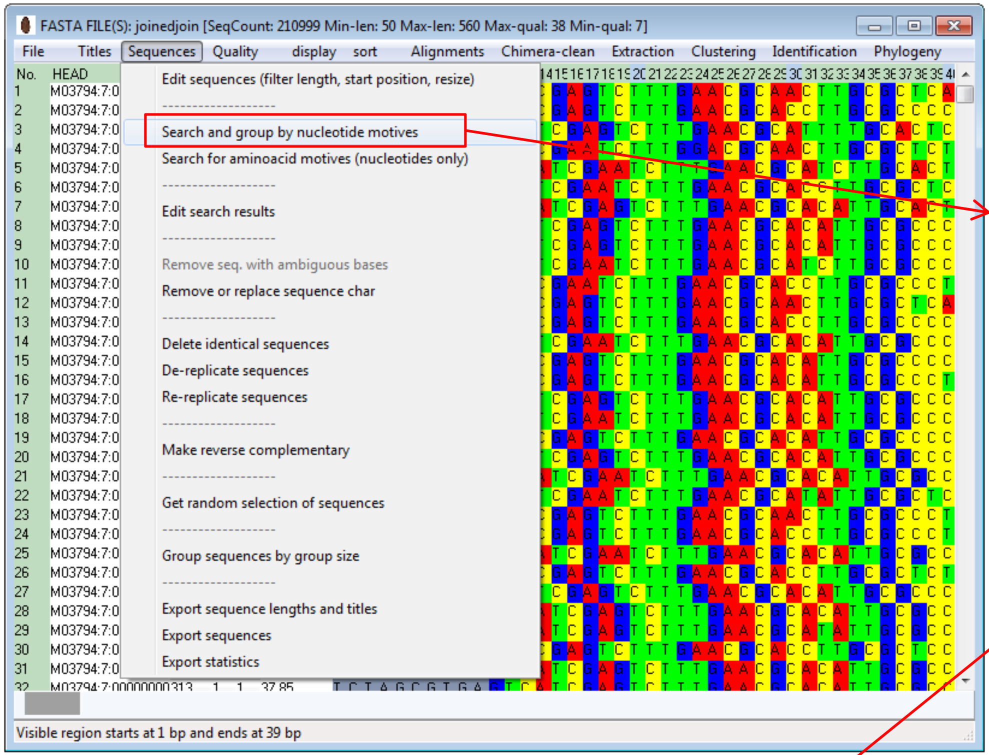
Sequence	Query	RESULT
GAGCGTGA	glTS7_T02	14687
TGGCGTGA	glTS7_T03	22568
GGTGCGTGA	glTS7_T06	16835
AAAGCGTGA	glTS7_T08	19258
TCTAGCGTGA	glTS7_T10	20585
210999 total	NO HIT	REV.COMP

show selected sequences



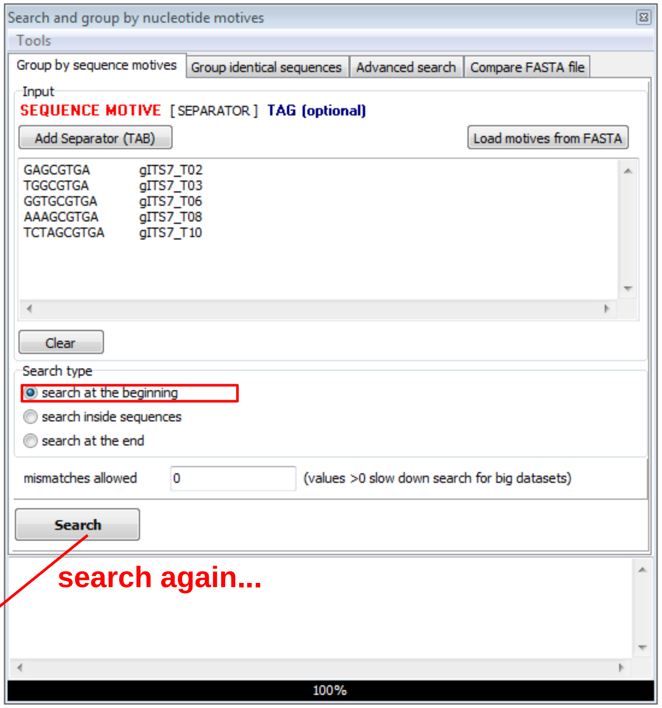
No.	HEAD	length	hits	tags	Mean Qual
1	M03794:7:0000000	325	1	1	30.06
2	M03794:7:0000000	372	1	1	33.62
3	M03794:7:0000000	249	1	1	36.69
4	M03794:7:0000000	309	1	1	37.22
5	M03794:7:0000000	324	1	1	33.19
6	M03794:7:0000000	388	1	1	36.06
7	M03794:7:0000000	384	1	1	34.88
8	M03794:7:0000000	316	1	1	37.86
9	M03794:7:0000000	308	1	1	37.69
10	M03794:7:0000000	333	1	1	37.8
11	M03794:7:0000000	365	1	1	37.88
12	M03794:7:0000000	317	1	1	37.87
13	M03794:7:0000000	372	1	1	36.95
14	M03794:7:0000000	308	1	1	37.84
15	M03794:7:0000000	292	1	1	37.88
16	M03794:7:0000000	291	1	1	37.86
17	M03794:7:0000000	292	1	1	37.82
18	M03794:7:0000000	295	1	1	37.84
19	M03794:7:0000000	309	1	1	37.25
20	M03794:7:0000000	310	1	1	37.82
21	M03794:7:0000000	293	1	1	37.87
22	M03794:7:0000000	396	1	1	37.72
23	M03794:7:0000000	303	1	1	37.88
24	M03794:7:0000000	364	1	1	37.84
25	M03794:7:0000000	317	1	1	37.82
26	M03794:7:0000000	376	1	1	37.8
27	M03794:7:0000000	287	1	1	37.85
28	M03794:7:0000000	321	1	1	37.85
29	M03794:7:0000000	303	1	1	35.99
30	M03794:7:0000000	351	1	1	37.87
31	M03794:7:0000000	314	1	1	37.68
32	M03794:7:0000000	313	1	1	37.85

Visible region starts at 1 bp and ends at 39 bp



FASTA FILE(S): joinedjoin [SeqCount: 210999 Min-len: 50 Max-len: 560 Max-qual: 38 Min-qual: 7]

Visible region starts at 1 bp and ends at 39 bp



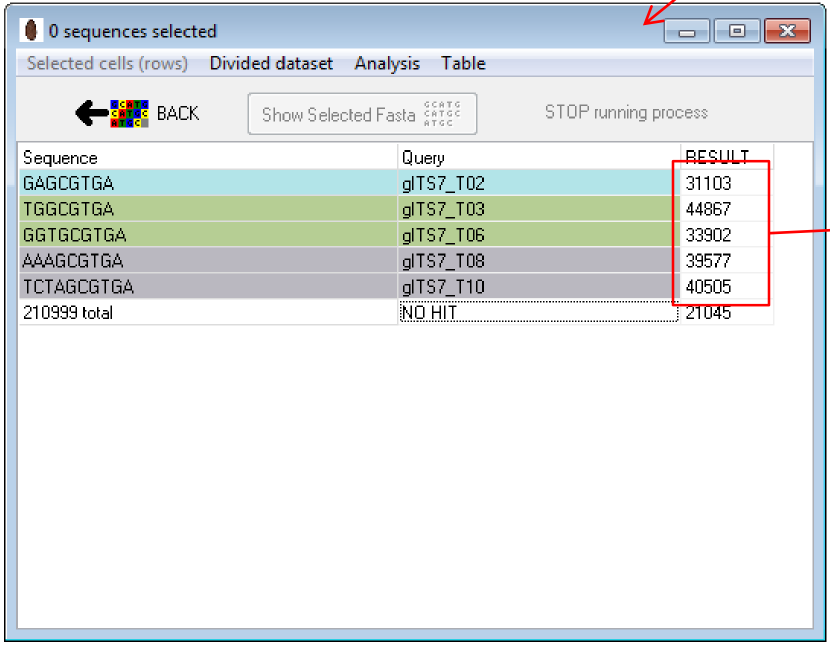
GAGCGTGA	gITS7_T02
TGGCGTGA	gITS7_T03
GGTGCGTGA	gITS7_T06
AAAGCGTGA	gITS7_T08
TCTAGCGTGA	gITS7_T10

● search at the beginning
○ search inside sequences
○ search at the end

mismatches allowed 0 (values >0 slow down search for big datasets)

100%

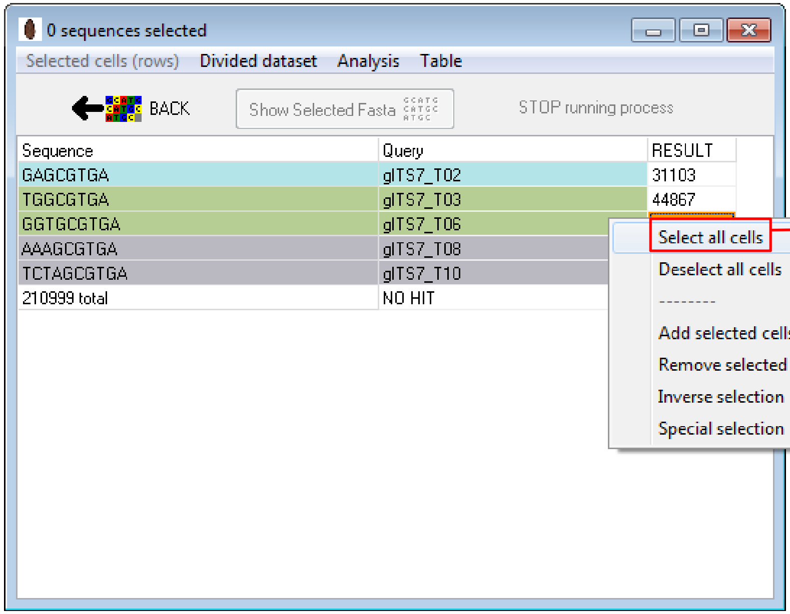
search again...



0 sequences selected

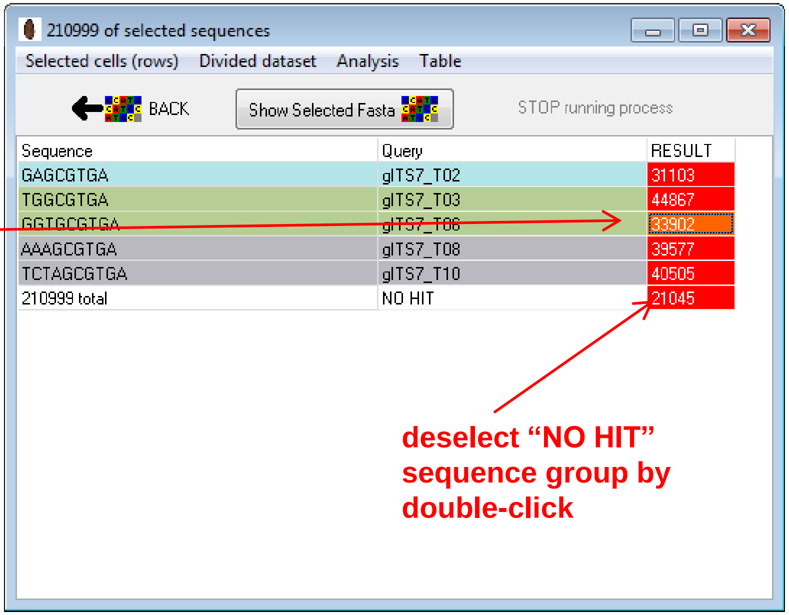
Sequence	Query	RESULT
GAGCGTGA	gITS7_T02	31103
TGGCGTGA	gITS7_T03	44867
GGTGCGTGA	gITS7_T06	33902
AAAGCGTGA	gITS7_T08	39577
TCTAGCGTGA	gITS7_T10	40505
210999 total	NO HIT	21045

now all sequences which contain the searched motives have the same orientation...



0 sequences selected

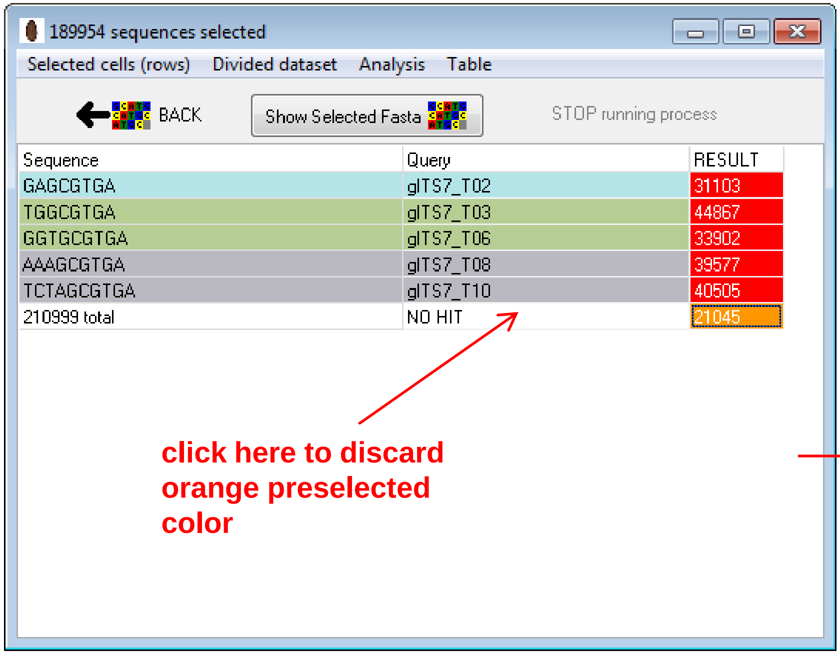
Sequence	Query	RESULT
GAGCGTGA	gITS7_T02	31103
TGGCGTGA	gITS7_T03	44867
GGTGCGTGA	gITS7_T06	
AAAGCGTGA	gITS7_T08	
TCTAGCGTGA	gITS7_T10	
210999 total	NO HIT	



210999 of selected sequences

Sequence	Query	RESULT
GAGCGTGA	gITS7_T02	31103
TGGCGTGA	gITS7_T03	44867
GGTGCGTGA	gITS7_T06	33902
AAAGCGTGA	gITS7_T08	39577
TCTAGCGTGA	gITS7_T10	40505
210999 total	NO HIT	21045

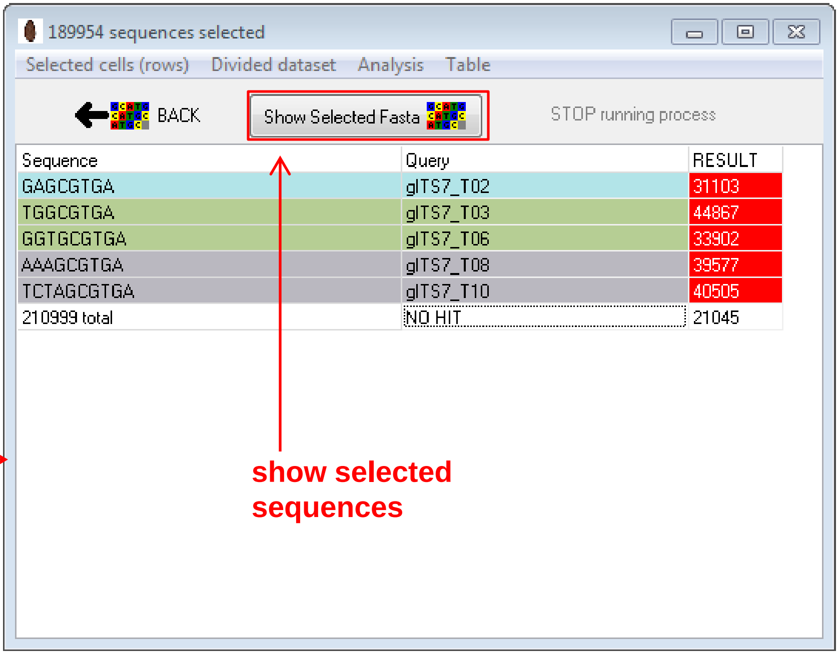
deselect “NO HIT” sequence group by double-click



189954 sequences selected

Sequence	Query	RESULT
GAGCGTGA	gITS7_T02	31103
TGGCGTGA	gITS7_T03	44867
GGTGCGTGA	gITS7_T06	33902
AAAGCGTGA	gITS7_T08	39577
TCTAGCGTGA	gITS7_T10	40505
210999 total	NO HIT	21045

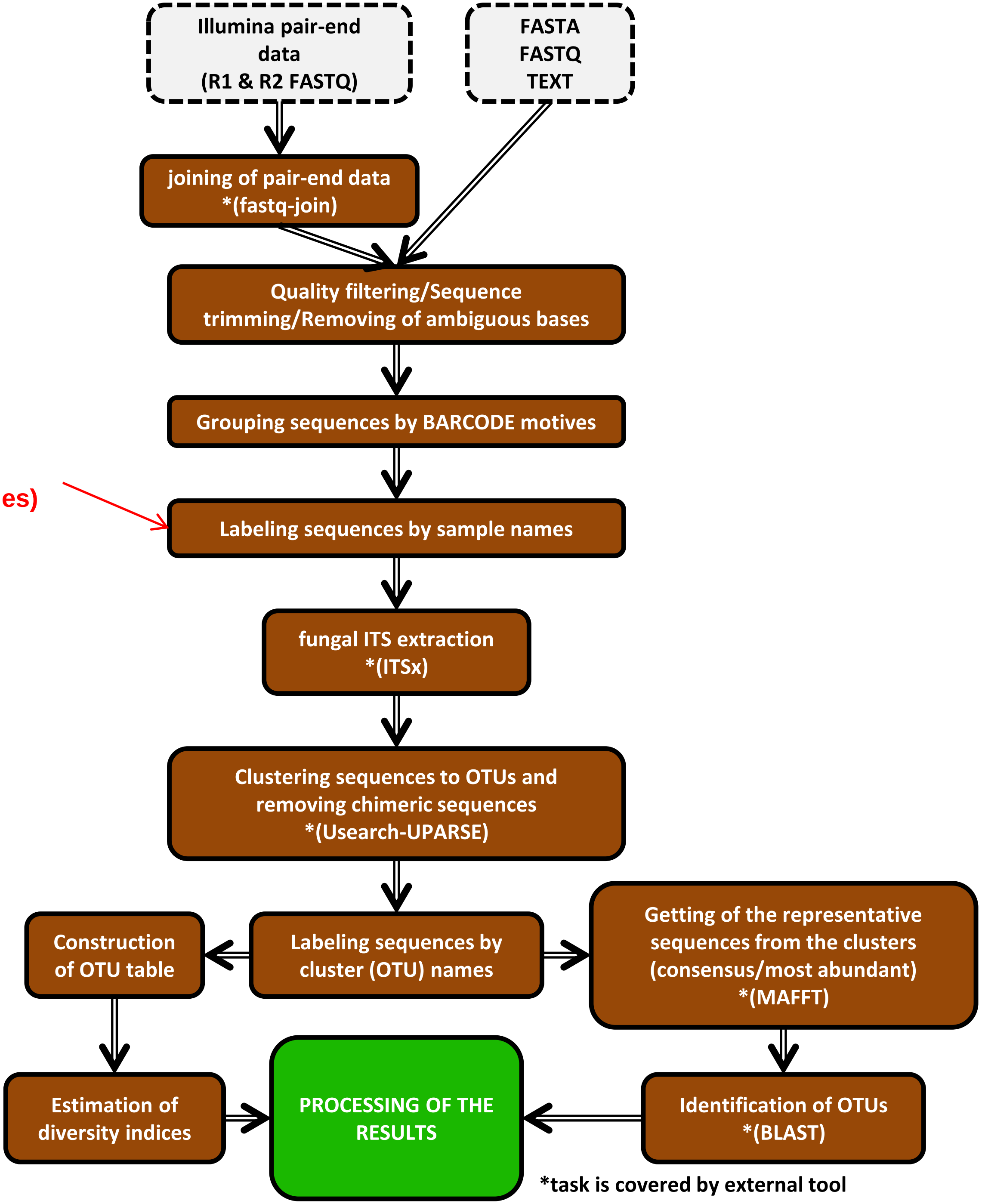
click here to discard orange preselected color



189954 sequences selected

Sequence	Query	RESULT
GAGCGTGA	gITS7_T02	31103
TGGCGTGA	gITS7_T03	44867
GGTGCGTGA	gITS7_T06	33902
AAAGCGTGA	gITS7_T08	39577
TCTAGCGTGA	gITS7_T10	40505
210999 total	NO HIT	21045

show selected sequences



Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

You are here
(forward barcodes)

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

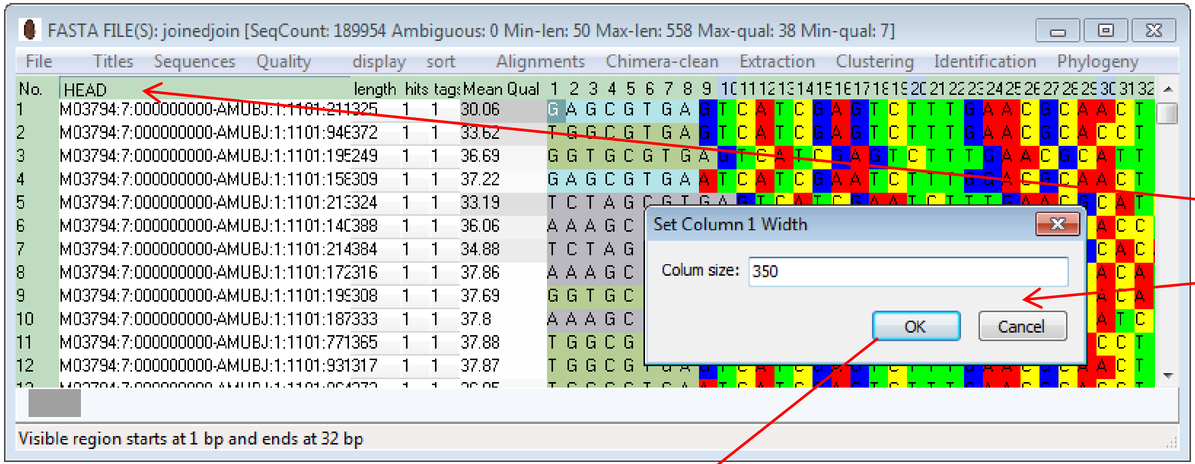
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool

Resize title column...



No.	HEAD	length	hits	tags	Mean Qual
1	M03794:7:000000000-AMUBJ:1:1101:211325	1	1	30.06
2	M03794:7:000000000-AMUBJ:1:1101:94E372	1	1	33.62
3	M03794:7:000000000-AMUBJ:1:1101:19E249	1	1	36.69
4	M03794:7:000000000-AMUBJ:1:1101:15E309	1	1	37.22
5	M03794:7:000000000-AMUBJ:1:1101:21324	1	1	33.19
6	M03794:7:000000000-AMUBJ:1:1101:14C388	1	1	36.06
7	M03794:7:000000000-AMUBJ:1:1101:214384	1	1	34.88
8	M03794:7:000000000-AMUBJ:1:1101:172316	1	1	37.86
9	M03794:7:000000000-AMUBJ:1:1101:19E308	1	1	37.69
10	M03794:7:000000000-AMUBJ:1:1101:187333	1	1	37.8
11	M03794:7:000000000-AMUBJ:1:1101:771365	1	1	37.88
12	M03794:7:000000000-AMUBJ:1:1101:931317	1	1	37.87

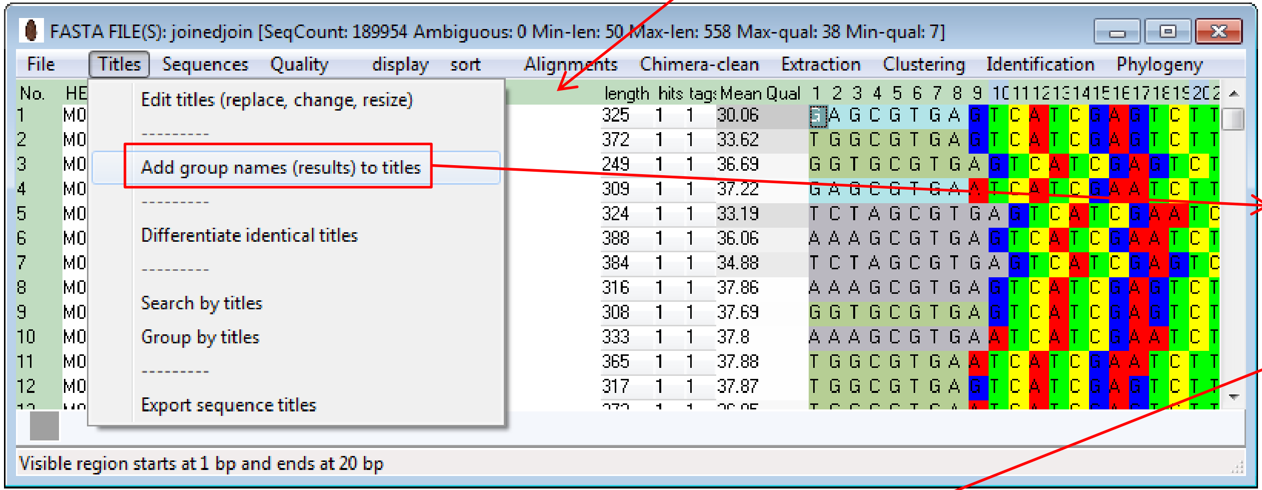
Set Column 1 Width

Colum size: 350

Visible region starts at 1 bp and ends at 32 bp

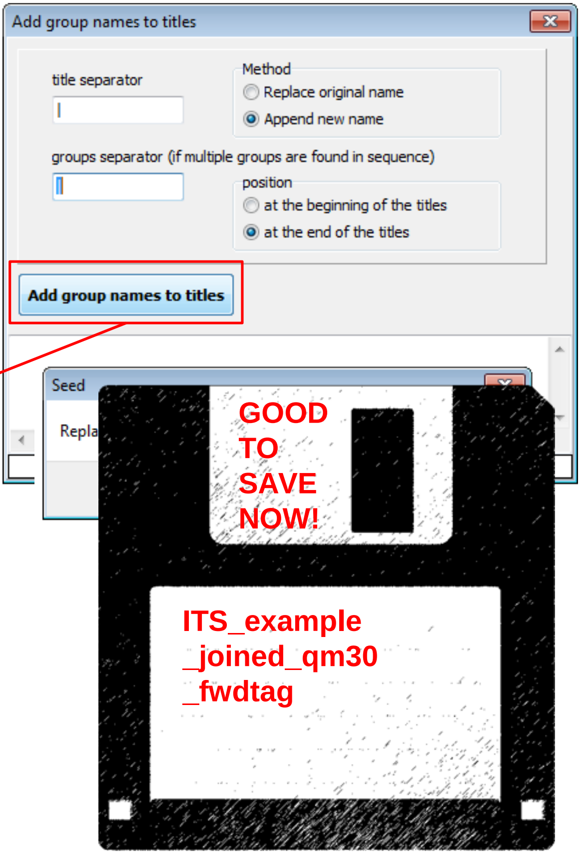
click here to set the width

Add group (tag) name to title...



No.	length	hits	tag	Mean Qual
1	325	1	1	30.06
2	372	1	1	33.62
3	249	1	1	36.69
4	309	1	1	37.22
5	324	1	1	33.19
6	388	1	1	36.06
7	384	1	1	34.88
8	316	1	1	37.86
9	308	1	1	37.69
10	333	1	1	37.8
11	365	1	1	37.88
12	317	1	1	37.87

Visible region starts at 1 bp and ends at 20 bp


Add group names to titles

title separator: |

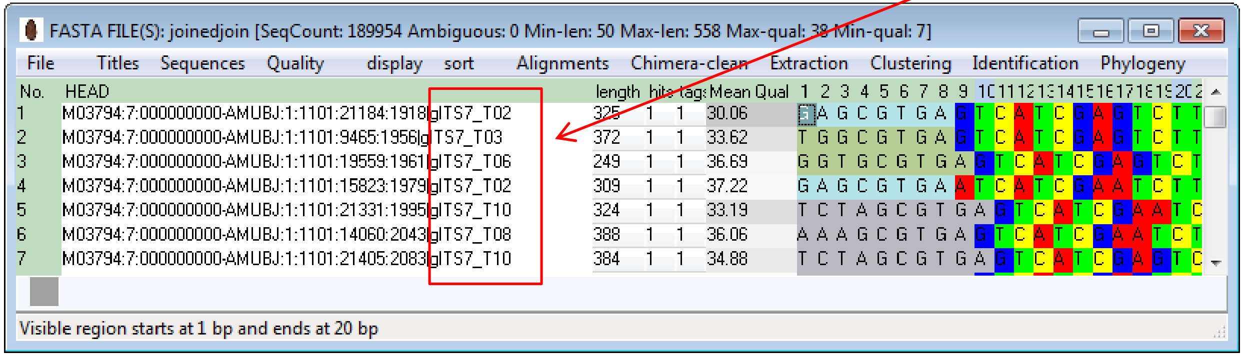
Method
○ Replace original name
● Append new name

groups separator (if multiple groups are found in sequence): |

position
○ at the beginning of the titles
● at the end of the titles

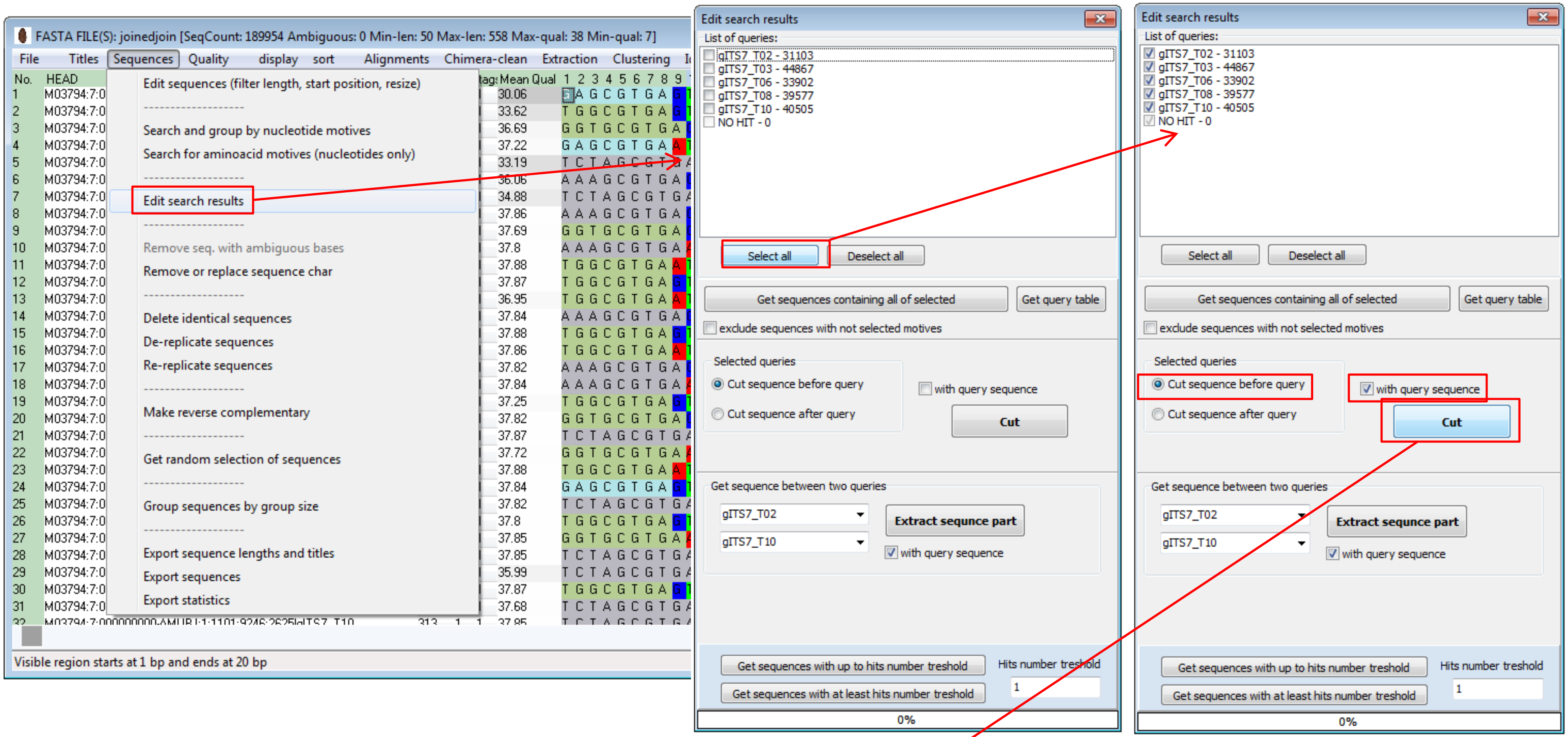
GOOD TO SAVE NOW!

ITS_example_joined_qm30_fwdtag



No.	HEAD	length	hits	tag	Mean Qual
1	M03794:7:000000000-AMUBJ:1:1101:21184:1918|gITS7_T02	325	1	1	30.06
2	M03794:7:000000000-AMUBJ:1:1101:9465:1956|gITS7_T03	372	1	1	33.62
3	M03794:7:000000000-AMUBJ:1:1101:19559:1961|gITS7_T06	249	1	1	36.69
4	M03794:7:000000000-AMUBJ:1:1101:15823:1979|gITS7_T02	309	1	1	37.22
5	M03794:7:000000000-AMUBJ:1:1101:21331:1995|gITS7_T10	324	1	1	33.19
6	M03794:7:000000000-AMUBJ:1:1101:14060:2043|gITS7_T08	388	1	1	36.06
7	M03794:7:000000000-AMUBJ:1:1101:21405:2083|gITS7_T10	384	1	1	34.88

Visible region starts at 1 bp and ends at 20 bp



FASTA FILE(S): joinedjoin [SeqCount: 189954 Ambiguous: 0 Min-len: 50 Max-len: 558 Max-qual: 38 Min-qual: 7]

No.	HEAD	Mean Qual	1	2	3	4	5	6	7	8	9
1	M03794:7:0	30.06	G	A	G	C	G	T	G	A	G
2	M03794:7:0	33.62	T	G	G	C	G	T	G	A	G
3	M03794:7:0	36.69	G	G	T	G	C	G	T	G	A
4	M03794:7:0	37.22	G	A	G	C	G	T	G	A	A
5	M03794:7:0	33.19	T	C	T	A	G	C	G	T	G
6	M03794:7:0	36.06	A	A	A	G	C	G	T	G	A
7	M03794:7:0	34.88	T	C	T	A	G	C	G	T	G
8	M03794:7:0	37.86	A	A	A	G	C	G	T	G	A
9	M03794:7:0	37.69	G	G	T	G	C	G	T	G	A
10	M03794:7:0	37.8	A	A	A	G	C	G	T	G	A
11	M03794:7:0	37.88	T	G	G	C	G	T	G	A	A
12	M03794:7:0	37.87	T	G	G	C	G	T	G	A	G
13	M03794:7:0	36.95	T	G	G	C	G	T	G	A	A
14	M03794:7:0	37.84	A	A	A	G	C	G	T	G	A
15	M03794:7:0	37.88	T	G	G	C	G	T	G	A	G
16	M03794:7:0	37.86	T	G	G	C	G	T	G	A	A
17	M03794:7:0	37.82	A	A	A	G	C	G	T	G	A
18	M03794:7:0	37.84	A	A	A	G	C	G	T	G	A
19	M03794:7:0	37.25	T	G	G	C	G	T	G	A	G
20	M03794:7:0	37.82	G	G	T	G	C	G	T	G	A
21	M03794:7:0	37.87	T	C	T	A	G	C	G	T	G
22	M03794:7:0	37.72	G	G	T	G	C	G	T	G	A
23	M03794:7:0	37.88	T	G	G	C	G	T	G	A	A
24	M03794:7:0	37.84	G	A	G	C	G	T	G	A	G
25	M03794:7:0	37.82	T	C	T	A	G	C	G	T	G
26	M03794:7:0	37.8	T	G	G	C	G	T	G	A	G
27	M03794:7:0	37.85	G	G	T	G	C	G	T	G	A
28	M03794:7:0	37.85	T	C	T	A	G	C	G	T	G
29	M03794:7:0	35.99	T	C	T	A	G	C	G	T	G
30	M03794:7:0	37.87	T	G	G	C	G	T	G	A	G
31	M03794:7:0	37.68	T	C	T	A	G	C	G	T	G

Visible region starts at 1 bp and ends at 20 bp

Edit search results

List of queries:

- ☐ gITS7_T02 - 31103
- ☐ gITS7_T03 - 44867
- ☐ gITS7_T06 - 33902
- ☐ gITS7_T08 - 39577
- ☐ gITS7_T10 - 40505
- ☐ NO HIT - 0

☐ exclude sequences with not selected motives

Selected queries

● Cut sequence before query ☐ with query sequence

○ Cut sequence after query

Get sequence between two queries

gITS7_T02

gITS7_T10

☑ with query sequence

Hits number treshold: 1

0%

Edit search results

List of queries:

- ☑ gITS7_T02 - 31103
- ☑ gITS7_T03 - 44867
- ☑ gITS7_T06 - 33902
- ☑ gITS7_T08 - 39577
- ☑ gITS7_T10 - 40505
- ☑ NO HIT - 0

☐ exclude sequences with not selected motives

Selected queries

● Cut sequence before query ☑ with query sequence

○ Cut sequence after query

Get sequence between two queries

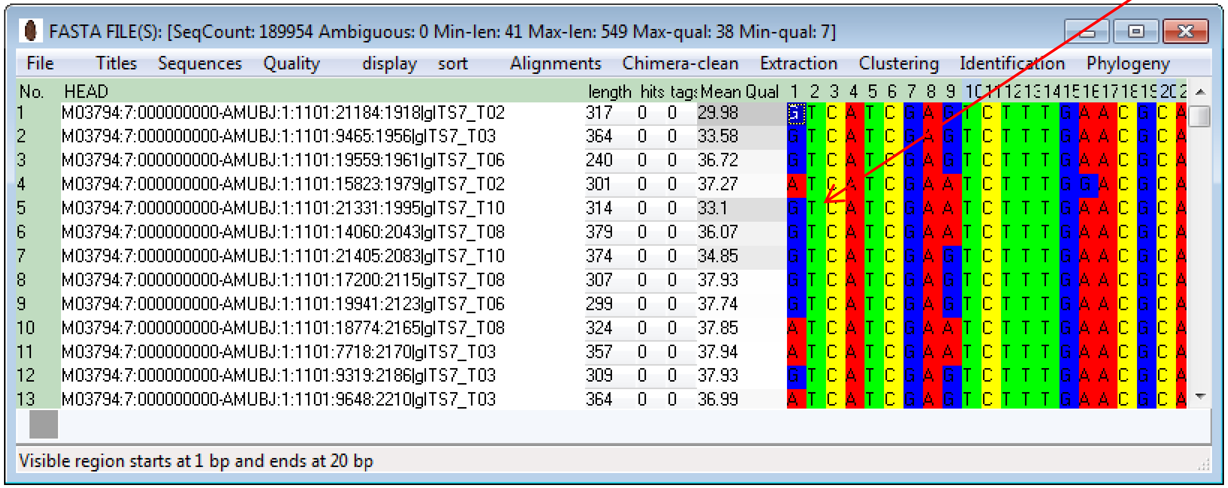
gITS7_T02

gITS7_T10

☑ with query sequence

Hits number treshold: 1

0%

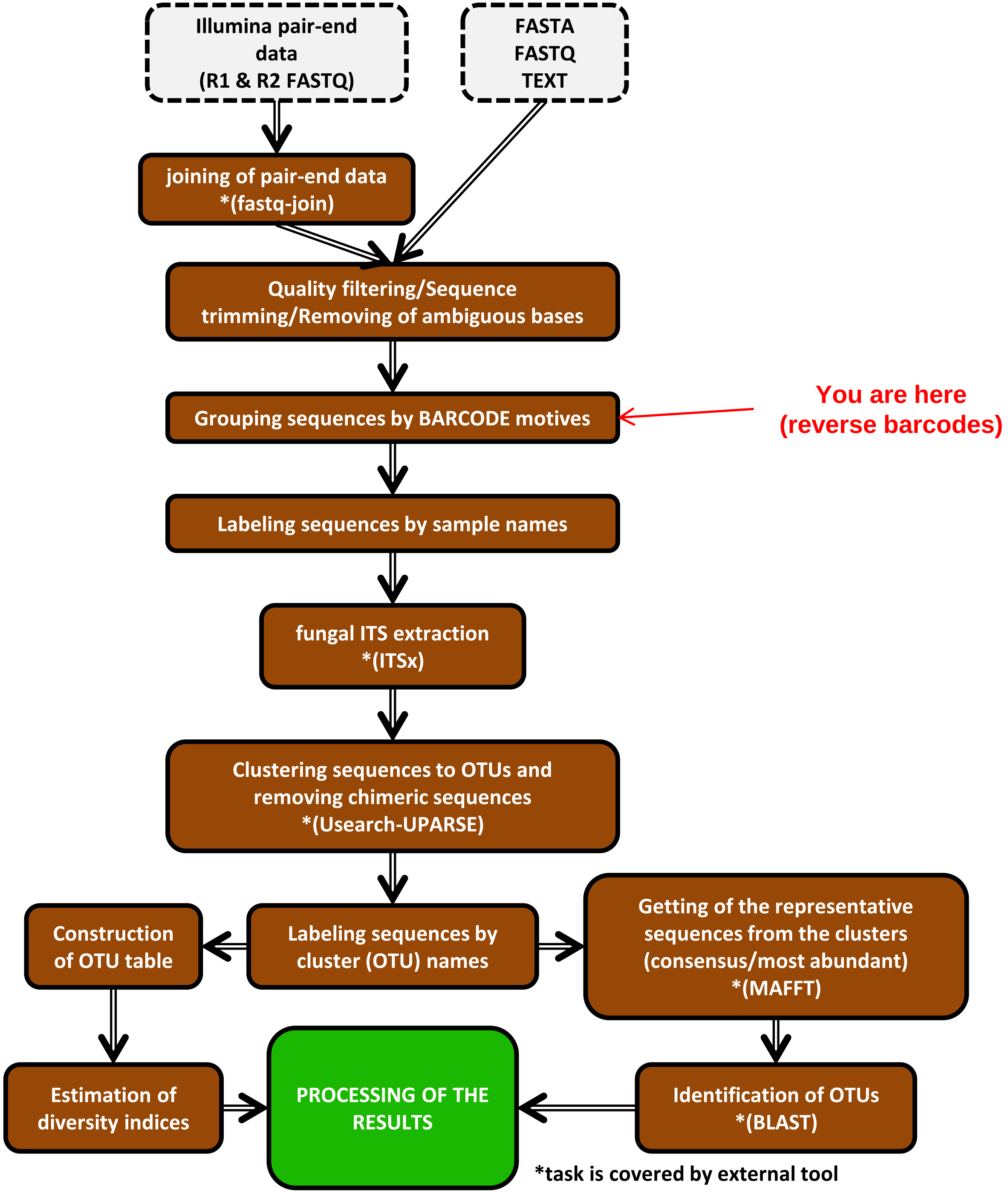


FASTA FILE(S): [SeqCount: 189954 Ambiguous: 0 Min-len: 41 Max-len: 549 Max-qual: 38 Min-qual: 7]

No.	HEAD	length	hits	tags	Mean Qual	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	2
1	M03794:7:000000000-AMUBJ:1:1101:21184:1918|gITS7_T02	317	0	0	29.98	G	T	C	A	T	C	G	A	G	T	C	T	T	T	G	A	A	C	G	C	A
2	M03794:7:000000000-AMUBJ:1:1101:9465:1956|gITS7_T03	364	0	0	33.58	G	T	C	A	T	C	G	A	G	T	C	T	T	T	G	A	A	C	G	C	A
3	M03794:7:000000000-AMUBJ:1:1101:19559:1961|gITS7_T06	240	0	0	36.72	G	T	C	A	T	C	G	A	G	T	C	T	T	T	G	A	A	C	G	C	A
4	M03794:7:000000000-AMUBJ:1:1101:15823:1979|gITS7_T02	301	0	0	37.27	A	T	C	A	T	C	G	A	A	T	C	T	T	T	G	G	A	C	G	C	A
5	M03794:7:000000000-AMUBJ:1:1101:21331:1995|gITS7_T10	314	0	0	33.1	G	T	C	A	T	C	G	A	A	T	C	T	T	T	G	A	A	C	G	C	A
6	M03794:7:000000000-AMUBJ:1:1101:14060:2043|gITS7_T08	379	0	0	36.07	G	T	C	A	T	C	G	A	A	T	C	T	T	T	G	A	A	C	G	C	A
7	M03794:7:000000000-AMUBJ:1:1101:21405:2083|gITS7_T10	374	0	0	34.85	G	T	C	A	T	C	G	A	G	T	C	T	T	T	G	A	A	C	G	C	A
8	M03794:7:000000000-AMUBJ:1:1101:17200:2115|gITS7_T08	307	0	0	37.93	G	T	C	A	T	C	G	A	G	T	C	T	T	T	G	A	A	C	G	C	A
9	M03794:7:000000000-AMUBJ:1:1101:19941:2123|gITS7_T06	299	0	0	37.74	G	T	C	A	T	C	G	A	G	T	C	T	T	T	G	A	A	C	G	C	A
10	M03794:7:000000000-AMUBJ:1:1101:18774:2165|gITS7_T08	324	0	0	37.85	A	T	C	A	T	C	G	A	A	T	C	T	T	T	G	A	A	C	G	C	A
11	M03794:7:000000000-AMUBJ:1:1101:7718:2170|gITS7_T03	357	0	0	37.94	A	T	C	A	T	C	G	A	A	T	C	T	T	T	G	A	A	C	G	C	A
12	M03794:7:000000000-AMUBJ:1:1101:9319:2186|gITS7_T03	309	0	0	37.93	G	T	C	A	T	C	G	A	G	T	C	T	T	T	G	A	A	C	G	C	A
13	M03794:7:000000000-AMUBJ:1:1101:9648:2210|gITS7_T03	364	0	0	36.99	A	T	C	A	T	C	G	A	G	T	C	T	T	T	G	A	A	C	G	C	A

Visible region starts at 1 bp and ends at 20 bp

Remove the tag motives from sequences...



Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

You are here (reverse barcodes)

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

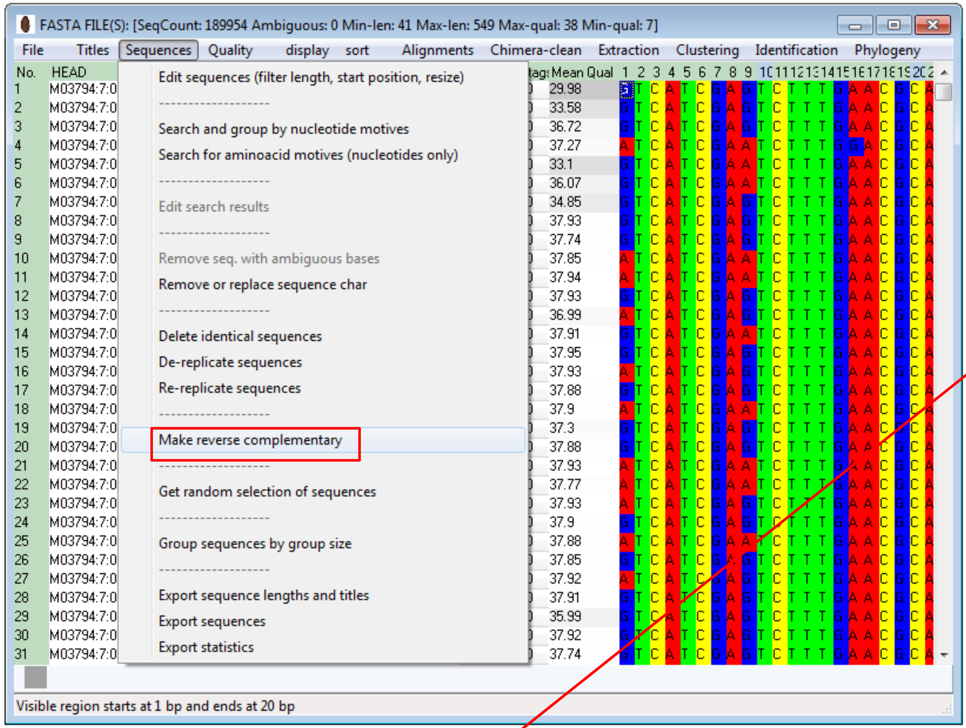
Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

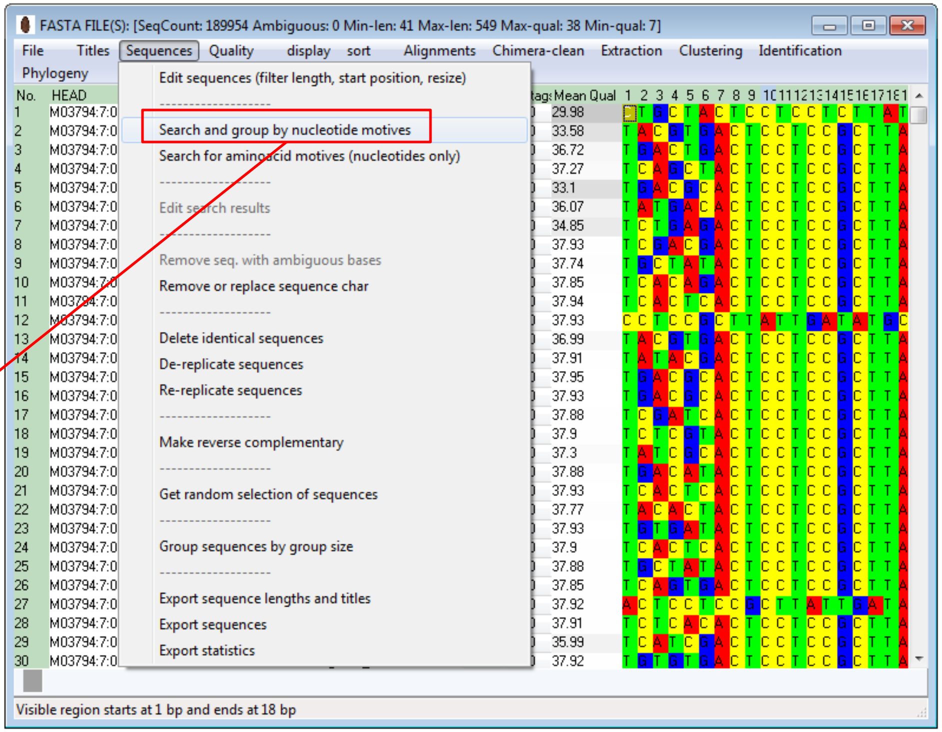
Estimation of diversity indices

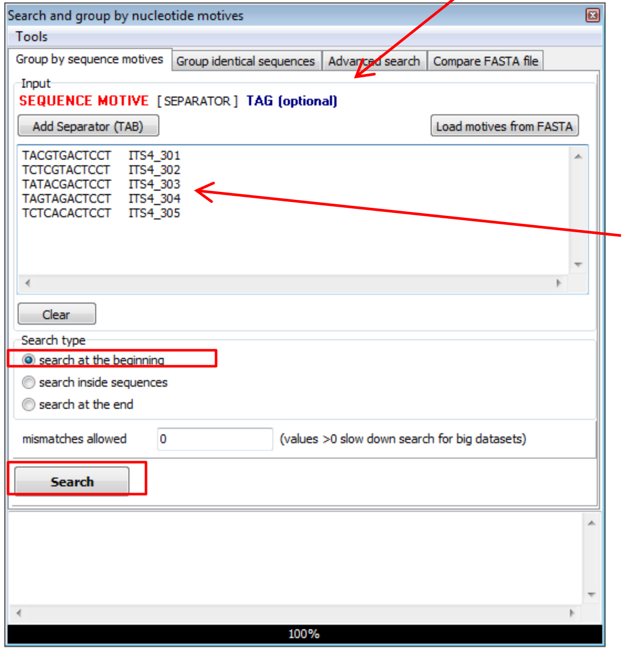
PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool







TACGTGACTCCT	ITS4_301
TCTCGTACTCCT	ITS4_302
TATACGACTCCT	ITS4_303
TAGTAGACTCCT	ITS4_304
TCTCACACTCCT	ITS4_305

● search at the beginning
○ search inside sequences
○ search at the end

mismatches allowed: 0 (values >0 slow down search for big datasets)

Search for the reverse tag motives...

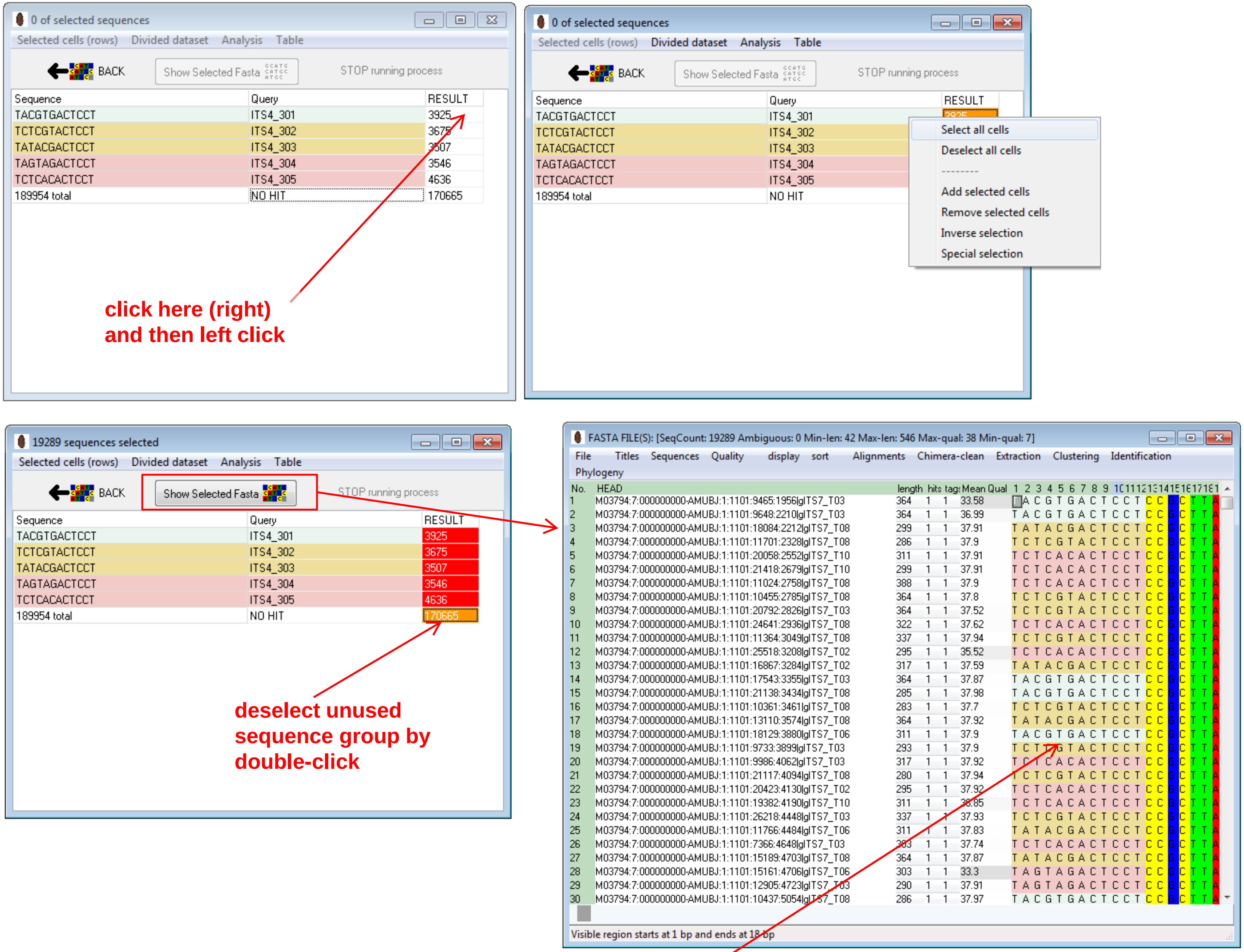
TACGTGACTCCT	ITS4_301
TCTCGTACTCCT	ITS4_302
TATACGACTCCT	ITS4_303
TAGTAGACTCCT	ITS4_304
TCTCACACTCCT	ITS4_305

reverse primer

ITS4	TCCTCCGCTTATTGATATGC

tagged reverse primers

ITS4_301	TACGTGACTCCTCCGCTTATTGATATGC
ITS4_302	TCTCGTACTCCTCCGCTTATTGATATGC
ITS4_303	TATACGACTCCTCCGCTTATTGATATGC
ITS4_304	TAGTAGACTCCTCCGCTTATTGATATGC
ITS4_305	TCTCACACTCCTCCGCTTATTGATATGC



0 of selected sequences

Sequence	Query	RESULT
TACGTGACTCCT	ITS4_301	3925
TCTCGTACTCCT	ITS4_302	3675
TATACGACTCCT	ITS4_303	3507
TAGTAGACTCCT	ITS4_304	3546
TCTCACACTCCT	ITS4_305	4636
189954 total	NO HIT	170665

click here (right)
and then left click

0 of selected sequences

Sequence	Query	RESULT
TACGTGACTCCT	ITS4_301	
TCTCGTACTCCT	ITS4_302	
TATACGACTCCT	ITS4_303	
TAGTAGACTCCT	ITS4_304	
TCTCACACTCCT	ITS4_305	
189954 total	NO HIT	

19289 sequences selected

Sequence	Query	RESULT
TACGTGACTCCT	ITS4_301	3925
TCTCGTACTCCT	ITS4_302	3675
TATACGACTCCT	ITS4_303	3507
TAGTAGACTCCT	ITS4_304	3546
TCTCACACTCCT	ITS4_305	4636
189954 total	NO HIT	170665

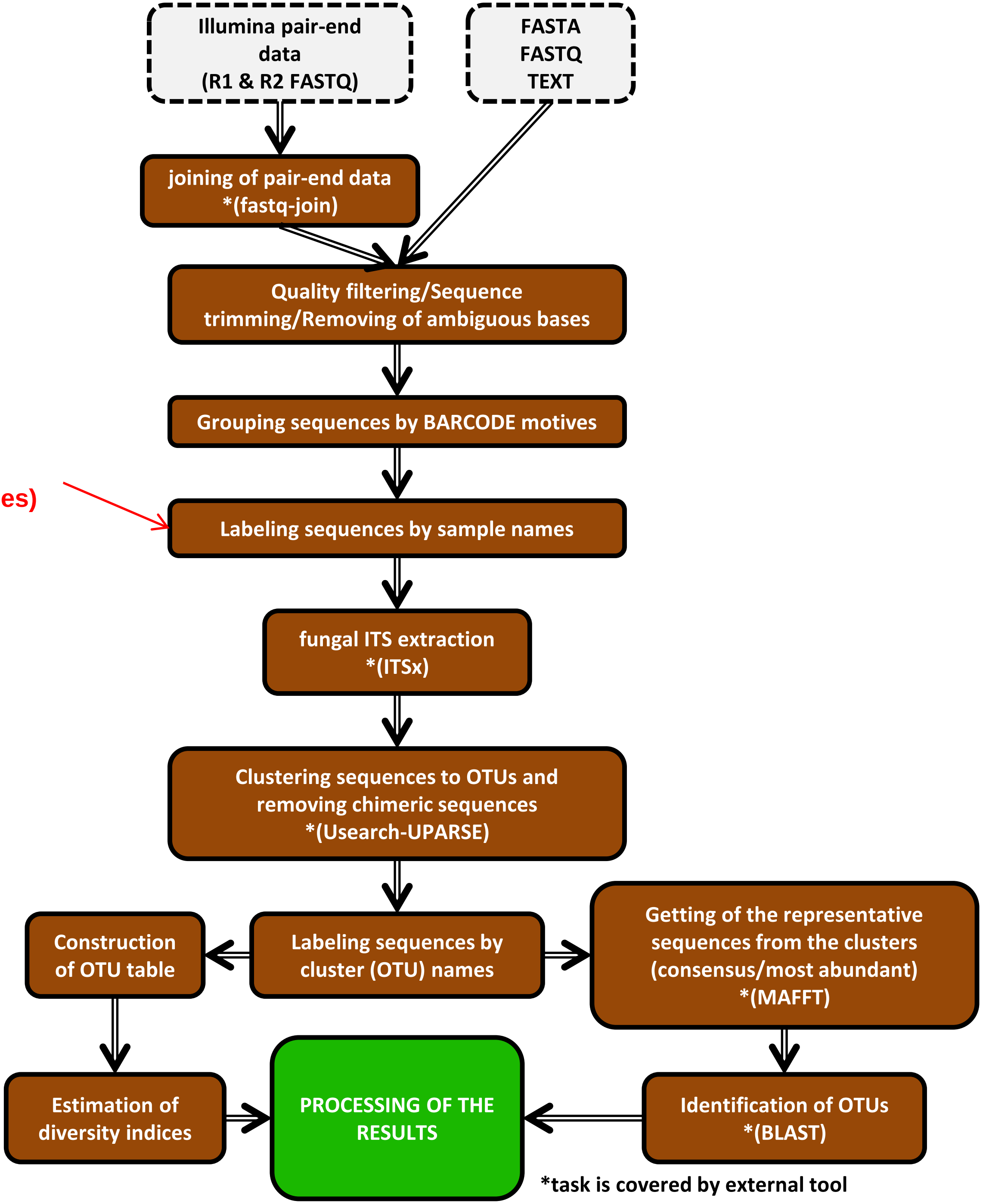
deselect unused sequence group by double-click

FASTA FILE(S): [SeqCount: 19289 Ambiguous: 0 Min-len: 42 Max-len: 546 Max-qual: 38 Min-qual: 7]

No.	HEAD	length	hits	tags	Mean Qual	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	1
1	M03794:7:000000000-AMUBJ:1:1101:9465:1956|gITS7_T03	364	1	1	33.58	T	A	C	G	T	G	A	C	T	C	C	T	C	C	G	C	T	T	A
2	M03794:7:000000000-AMUBJ:1:1101:9648:2210|gITS7_T03	364	1	1	36.99	T	A	C	G	T	G	A	C	T	C	C	T	C	C	G	C	T	T	A
3	M03794:7:000000000-AMUBJ:1:1101:18084:2212|gITS7_T08	299	1	1	37.91	T	A	T	A	C	G	A	C	T	C	C	T	C	C	G	C	T	T	A
4	M03794:7:000000000-AMUBJ:1:1101:11701:2328|gITS7_T08	286	1	1	37.9	T	C	T	C	G	T	A	C	T	C	C	T	C	C	G	C	T	T	A
5	M03794:7:000000000-AMUBJ:1:1101:20058:2552|gITS7_T10	311	1	1	37.91	T	C	T	C	A	C	A	C	T	C	C	T	C	C	G	C	T	T	A
6	M03794:7:000000000-AMUBJ:1:1101:21418:2679|gITS7_T10	299	1	1	37.91	T	C	T	C	A	C	A	C	T	C	C	T	C	C	G	C	T	T	A
7	M03794:7:000000000-AMUBJ:1:1101:11024:2758|gITS7_T08	388	1	1	37.9	T	C	T	C	A	C	A	C	T	C	C	T	C	C	G	C	T	T	A
8	M03794:7:000000000-AMUBJ:1:1101:10455:2785|gITS7_T08	364	1	1	37.8	T	C	T	C	G	T	A	C	T	C	C	T	C	C	G	C	T	T	A
9	M03794:7:000000000-AMUBJ:1:1101:20792:2826|gITS7_T03	364	1	1	37.52	T	C	T	C	G	T	A	C	T	C	C	T	C	C	G	C	T	T	A
10	M03794:7:000000000-AMUBJ:1:1101:24641:2936|gITS7_T08	322	1	1	37.62	T	C	T	C	A	C	A	C	T	C	C	T	C	C	G	C	T	T	A
11	M03794:7:000000000-AMUBJ:1:1101:11364:3049|gITS7_T08	337	1	1	37.94	T	C	T	C	G	T	A	C	T	C	C	T	C	C	G	C	T	T	A
12	M03794:7:000000000-AMUBJ:1:1101:25518:3208|gITS7_T02	295	1	1	35.52	T	C	T	C	A	C	A	C	T	C	C	T	C	C	G	C	T	T	A
13	M03794:7:000000000-AMUBJ:1:1101:16867:3284|gITS7_T02	317	1	1	37.59	T	A	T	A	C	G	A	C	T	C	C	T	C	C	G	C	T	T	A
14	M03794:7:000000000-AMUBJ:1:1101:17543:3355|gITS7_T03	364	1	1	37.87	T	A	C	G	T	G	A	C	T	C	C	T	C	C	G	C	T	T	A
15	M03794:7:000000000-AMUBJ:1:1101:21138:3434|gITS7_T08	285	1	1	37.98	T	A	C	G	T	G	A	C	T	C	C	T	C	C	G	C	T	T	A
16	M03794:7:000000000-AMUBJ:1:1101:10361:3461|gITS7_T08	283	1	1	37.7	T	C	T	C	G	T	A	C	T	C	C	T	C	C	G	C	T	T	A
17	M03794:7:000000000-AMUBJ:1:1101:13110:3574|gITS7_T08	364	1	1	37.92	T	A	T	A	C	G	A	C	T	C	C	T	C	C	G	C	T	T	A
18	M03794:7:000000000-AMUBJ:1:1101:18129:3880|gITS7_T06	311	1	1	37.9	T	A	C	G	T	G	A	C	T	C	C	T	C	C	G	C	T	T	A
19	M03794:7:000000000-AMUBJ:1:1101:9733:3899|gITS7_T03	293	1	1	37.9	T	C	T	C	G	T	A	C	T	C	C	T	C	C	G	C	T	T	A
20	M03794:7:000000000-AMUBJ:1:1101:9986:4062|gITS7_T03	317	1	1	37.92	T	C	T	C	A	C	A	C	T	C	C	T	C	C	G	C	T	T	A
21	M03794:7:000000000-AMUBJ:1:1101:21117:4094|gITS7_T08	280	1	1	37.94	T	C	T	C	G	T	A	C	T	C	C	T	C	C	G	C	T	T	A
22	M03794:7:000000000-AMUBJ:1:1101:20423:4130|gITS7_T02	295	1	1	37.92	T	C	T	C	A	C	A	C	T	C	C	T	C	C	G	C	T	T	A
23	M03794:7:000000000-AMUBJ:1:1101:19382:4190|gITS7_T10	311	1	1	36.85	T	C	T	C	A	C	A	C	T	C	C	T	C	C	G	C	T	T	A
24	M03794:7:000000000-AMUBJ:1:1101:26218:4448|gITS7_T03	337	1	1	37.93	T	C	T	C	G	T	A	C	T	C	C	T	C	C	G	C	T	T	A
25	M03794:7:000000000-AMUBJ:1:1101:11766:4484|gITS7_T06	311	1	1	37.83	T	A	T	A	C	G	A	C	T	C	C	T	C	C	G	C	T	T	A
26	M03794:7:000000000-AMUBJ:1:1101:7366:4648|gITS7_T03	303	1	1	37.74	T	C	T	C	A	C	A	C	T	C	C	T	C	C	G	C	T	T	A
27	M03794:7:000000000-AMUBJ:1:1101:15189:4703|gITS7_T08	364	1	1	37.87	T	A	T	A	C	G	A	C	T	C	C	T	C	C	G	C	T	T	A
28	M03794:7:000000000-AMUBJ:1:1101:15161:4706|gITS7_T06	303	1	1	33.3	T	A	G	T	A	G	A	C	T	C	C	T	C	C	G	C	T	T	A
29	M03794:7:000000000-AMUBJ:1:1101:12905:4723|gITS7_T03	290	1	1	37.91	T	A	G	T	A	G	A	C	T	C	C	T	C	C	G	C	T	T	A
30	M03794:7:000000000-AMUBJ:1:1101:10437:5054|gITS7_T08	286	1	1	37.97	T	A	C	G	T	G	A	C	T	C	C	T	C	C	G	C	T	T	A

Visible region starts at 1 bp and ends at 18 bp

searched sequence motives



Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

You are here
(reverse barcodes)

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

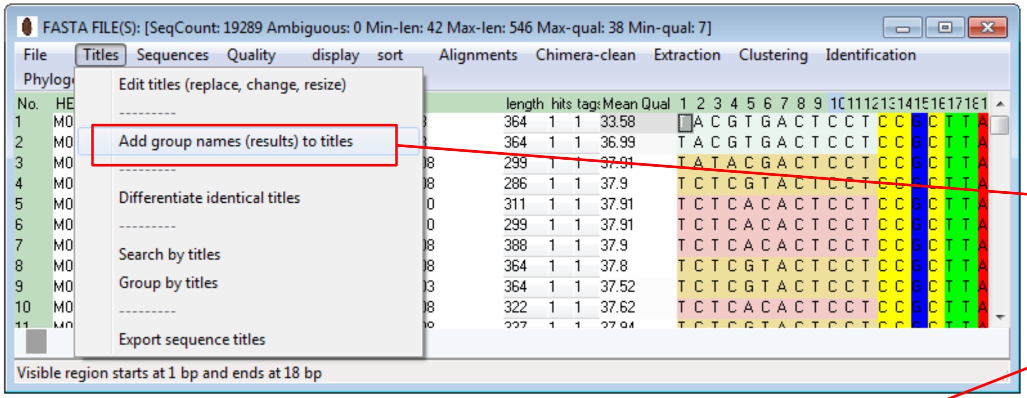
Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

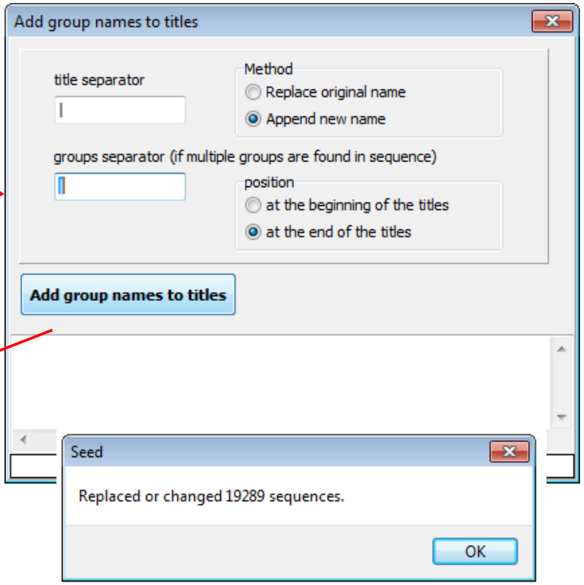
*task is covered by external tool



FASTA FILE(S): [SeqCount: 19289 Ambiguous: 0 Min-len: 42 Max-len: 546 Max-qual: 38 Min-qual: 7]

No.	HE	length	hits	tags	Mean Qual	1 2 3 4 5 6 7 8 9 1C11121314151E171E1
1	M0	364	1	1	33.58	T A C G T G A C T C C T C C G C T T A
2	M0	364	1	1	36.99	T A C G T G A C T C C T C C G C T T A
3	M0	299	1	1	37.91	T A T A C G A C T C C T C C G C T T A
4	M0	286	1	1	37.9	T C T C G T A C T C C T C C G C T T A
5	M0	311	1	1	37.91	T C T C A C A C T C C T C C G C T T A
6	M0	299	1	1	37.91	T C T C A C A C T C C T C C G C T T A
7	M0	388	1	1	37.9	T C T C A C A C T C C T C C G C T T A
8	M0	364	1	1	37.8	T C T C G T A C T C C T C C G C T T A
9	M0	364	1	1	37.52	T C T C G T A C T C C T C C G C T T A
10	M0	322	1	1	37.62	T C T C A C A C T C C T C C G C T T A

Visible region starts at 1 bp and ends at 18 bp



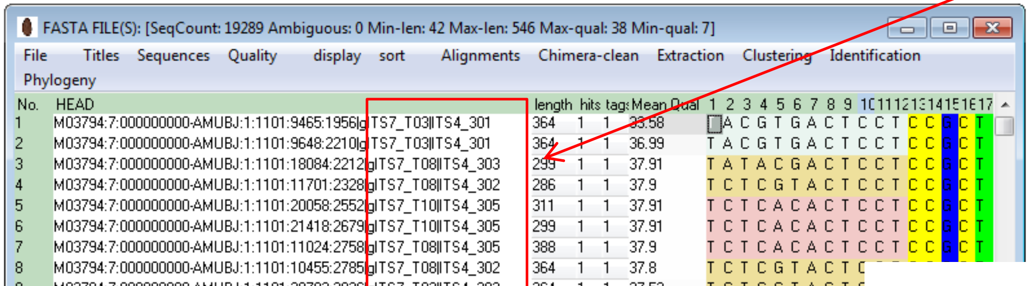
title separator: |

Method: ○ Replace original name ● Append new name

groups separator (if multiple groups are found in sequence): |

position: ○ at the beginning of the titles ● at the end of the titles

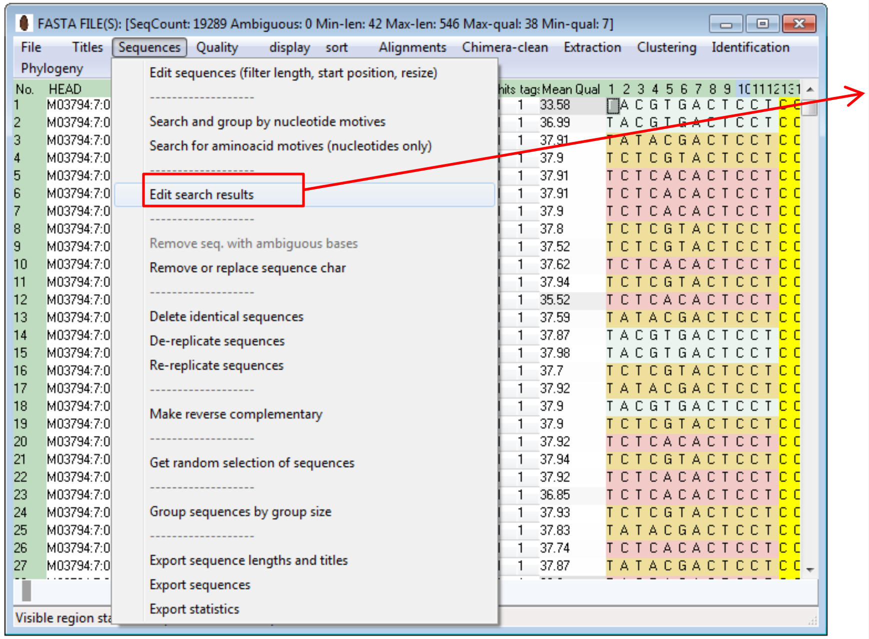
Replaced or changed 19289 sequences.



FASTA FILE(S): [SeqCount: 19289 Ambiguous: 0 Min-len: 42 Max-len: 546 Max-qual: 38 Min-qual: 7]

No.	HEAD	length	hits	tags	Mean Qual	1 2 3 4 5 6 7 8 9 1C11121314151E17
1	M03794:7:000000000-AMUBJ:1:1101:9465:1956|gITS7_T03|ITS4_301	364	1	1	33.58	T A C G T G A C T C C T C C G C T
2	M03794:7:000000000-AMUBJ:1:1101:9648:2210|gITS7_T03|ITS4_301	364	1	1	36.99	T A C G T G A C T C C T C C G C T
3	M03794:7:000000000-AMUBJ:1:1101:18084:2212|gITS7_T08|ITS4_303	299	1	1	37.91	T A T A C G A C T C C T C C G C T
4	M03794:7:000000000-AMUBJ:1:1101:11701:2328|gITS7_T08|ITS4_302	286	1	1	37.9	T C T C G T A C T C C T C C G C T
5	M03794:7:000000000-AMUBJ:1:1101:20058:2552|gITS7_T10|ITS4_305	311	1	1	37.91	T C T C A C A C T C C T C C G C T
6	M03794:7:000000000-AMUBJ:1:1101:21418:2679|gITS7_T10|ITS4_305	299	1	1	37.91	T C T C A C A C T C C T C C G C T
7	M03794:7:000000000-AMUBJ:1:1101:11024:2758|gITS7_T08|ITS4_305	388	1	1	37.9	T C T C A C A C T C C T C C G C T
8	M03794:7:000000000-AMUBJ:1:1101:10455:2785|gITS7_T08|ITS4_302	364	1	1	37.8	T C T C G T A C T C
9	M03794:7:000000000-AMUBJ:1:1101:20792:2826|gITS7_T03|ITS4_302	364	1	1	37.52	T C T C G T A C T C

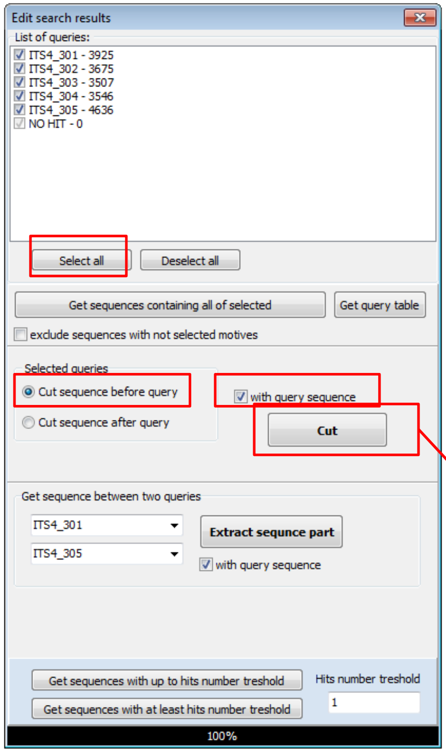
Visible region starts at 1 bp and ends at 17 bp



FASTA FILE(S): [SeqCount: 19289 Ambiguous: 0 Min-len: 42 Max-len: 546 Max-qual: 38 Min-qual: 7]

No.	HEAD	tags	Mean Qual	1 2 3 4 5 6 7 8 9 1C11121:1
1	M03794:7:0	1	33.58	T A C G T G A C T C C T C C
2	M03794:7:0	1	36.99	T A C G T G A C T C C T C C
3	M03794:7:0	1	37.91	T A T A C G A C T C C T C C
4	M03794:7:0	1	37.9	T C T C G T A C T C C T C C
5	M03794:7:0	1	37.91	T C T C A C A C T C C T C C
6	M03794:7:0	1	37.91	T C T C A C A C T C C T C C
7	M03794:7:0	1	37.9	T C T C A C A C T C C T C C
8	M03794:7:0	1	37.8	T C T C G T A C T C C T C C
9	M03794:7:0	1	37.52	T C T C G T A C T C C T C C
10	M03794:7:0	1	37.62	T C T C A C A C T C C T C C
11	M03794:7:0	1	37.94	T C T C G T A C T C C T C C
12	M03794:7:0	1	35.52	T C T C A C A C T C C T C C
13	M03794:7:0	1	37.59	T A T A C G A C T C C T C C
14	M03794:7:0	1	37.87	T A C G T G A C T C C T C C
15	M03794:7:0	1	37.98	T A C G T G A C T C C T C C
16	M03794:7:0	1	37.7	T C T C G T A C T C C T C C
17	M03794:7:0	1	37.92	T A T A C G A C T C C T C C
18	M03794:7:0	1	37.9	T A C G T G A C T C C T C C
19	M03794:7:0	1	37.9	T C T C G T A C T C C T C C
20	M03794:7:0	1	37.92	T C T C A C A C T C C T C C
21	M03794:7:0	1	37.94	T C T C G T A C T C C T C C
22	M03794:7:0	1	37.92	T C T C A C A C T C C T C C
23	M03794:7:0	1	36.85	T C T C A C A C T C C T C C
24	M03794:7:0	1	37.93	T C T C G T A C T C C T C C
25	M03794:7:0	1	37.83	T A T A C G A C T C C T C C
26	M03794:7:0	1	37.74	T C T C A C A C T C C T C C
27	M03794:7:0	1	37.87	T A T A C G A C T C C T C C

Visible region sta



List of queries:

- ☑ ITS4_301 - 3925
- ☑ ITS4_302 - 3675
- ☑ ITS4_303 - 3507
- ☑ ITS4_304 - 3546
- ☑ ITS4_305 - 4636
- ☑ NO HIT - 0

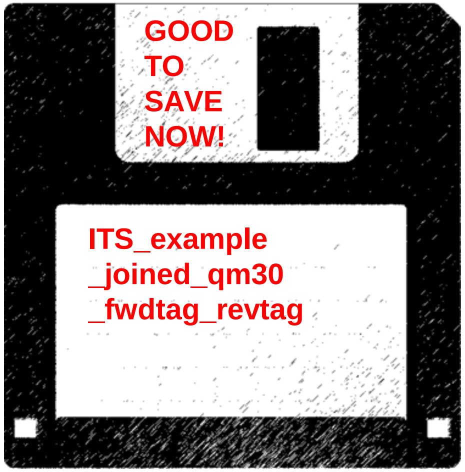
☐ exclude sequences with not selected motives

Selected queries: ● Cut sequence before query ○ Cut sequence after query ☑ with query sequence

Get sequence between two queries: ITS4_301, ITS4_305 ☑ with query sequence

Hits number treshold: 1

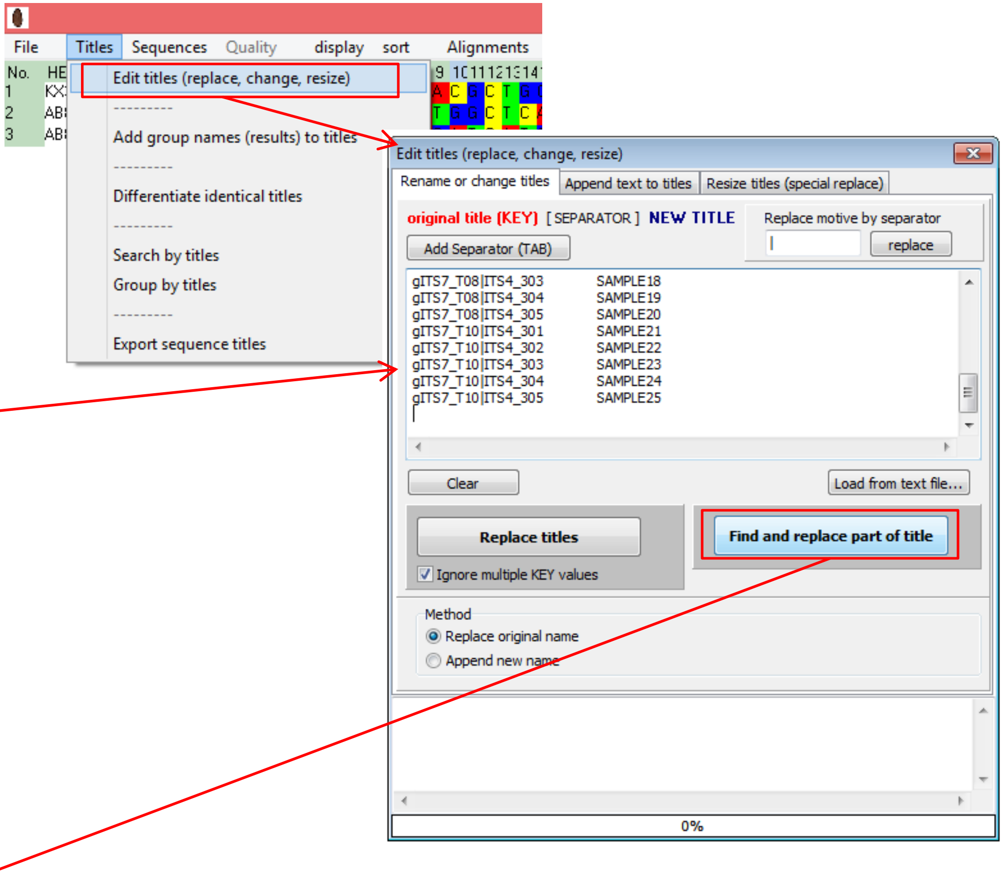
100%



Remove the tag motives from sequences...

Replace tag names by sample name…

	FWD Primer	REV Primer			
SAMPLE01	gITS7_T02	ITS4_301		gITS7_T02|ITS4_301	SAMPLE01
SAMPLE02	gITS7_T02	ITS4_302		gITS7_T02|ITS4_302	SAMPLE02
SAMPLE03	gITS7_T02	ITS4_303		gITS7_T02|ITS4_303	SAMPLE03
SAMPLE04	gITS7_T02	ITS4_304		gITS7_T02|ITS4_304	SAMPLE04
SAMPLE05	gITS7_T02	ITS4_305		gITS7_T02|ITS4_305	SAMPLE05
SAMPLE06	gITS7_T03	ITS4_301		gITS7_T03|ITS4_301	SAMPLE06
SAMPLE07	gITS7_T03	ITS4_302		gITS7_T03|ITS4_302	SAMPLE07
SAMPLE08	gITS7_T03	ITS4_303		gITS7_T03|ITS4_303	SAMPLE08
SAMPLE09	gITS7_T03	ITS4_304		gITS7_T03|ITS4_304	SAMPLE09
SAMPLE10	gITS7_T03	ITS4_305		gITS7_T03|ITS4_305	SAMPLE10
SAMPLE11	gITS7_T06	ITS4_301		gITS7_T06|ITS4_301	SAMPLE11
SAMPLE12	gITS7_T06	ITS4_302		gITS7_T06|ITS4_302	SAMPLE12
SAMPLE13	gITS7_T06	ITS4_303		gITS7_T06|ITS4_303	SAMPLE13
SAMPLE14	gITS7_T06	ITS4_304		gITS7_T06|ITS4_304	SAMPLE14
SAMPLE15	gITS7_T06	ITS4_305		gITS7_T06|ITS4_305	SAMPLE15
SAMPLE16	gITS7_T08	ITS4_301		gITS7_T08|ITS4_301	SAMPLE16
SAMPLE17	gITS7_T08	ITS4_302		gITS7_T08|ITS4_302	SAMPLE17
SAMPLE18	gITS7_T08	ITS4_303		gITS7_T08|ITS4_303	SAMPLE18
SAMPLE19	gITS7_T08	ITS4_304		gITS7_T08|ITS4_304	SAMPLE19
SAMPLE20	gITS7_T08	ITS4_305		gITS7_T08|ITS4_305	SAMPLE20
SAMPLE21	gITS7_T10	ITS4_301		gITS7_T10|ITS4_301	SAMPLE21
SAMPLE22	gITS7_T10	ITS4_302		gITS7_T10|ITS4_302	SAMPLE22
SAMPLE23	gITS7_T10	ITS4_303		gITS7_T10|ITS4_303	SAMPLE23
SAMPLE24	gITS7_T10	ITS4_304		gITS7_T10|ITS4_304	SAMPLE24
SAMPLE25	gITS7_T10	ITS4_305		gITS7_T10|ITS4_305	SAMPLE25

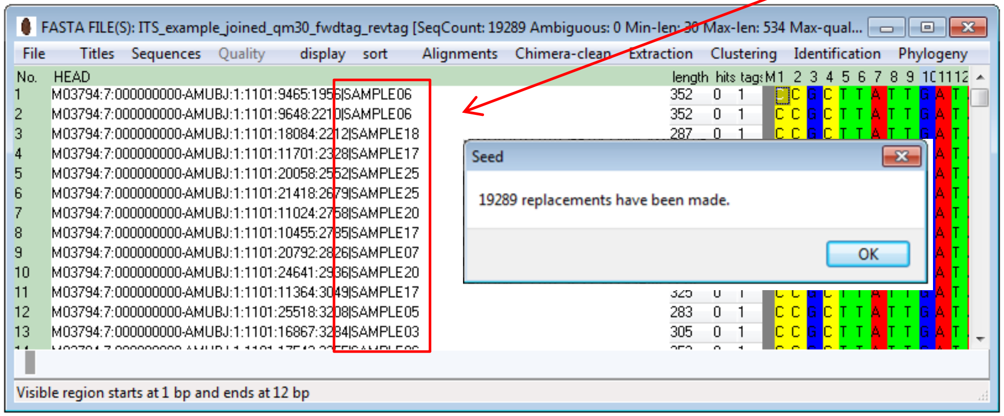


Replace motive by separator: |

```
gITS7_T08|ITS4_303	SAMPLE18
gITS7_T08|ITS4_304	SAMPLE19
gITS7_T08|ITS4_305	SAMPLE20
gITS7_T10|ITS4_301	SAMPLE21
gITS7_T10|ITS4_302	SAMPLE22
gITS7_T10|ITS4_303	SAMPLE23
gITS7_T10|ITS4_304	SAMPLE24
gITS7_T10|ITS4_305	SAMPLE25
```

☑ Ignore multiple KEY values

Method
● Replace original name
○ Append new name

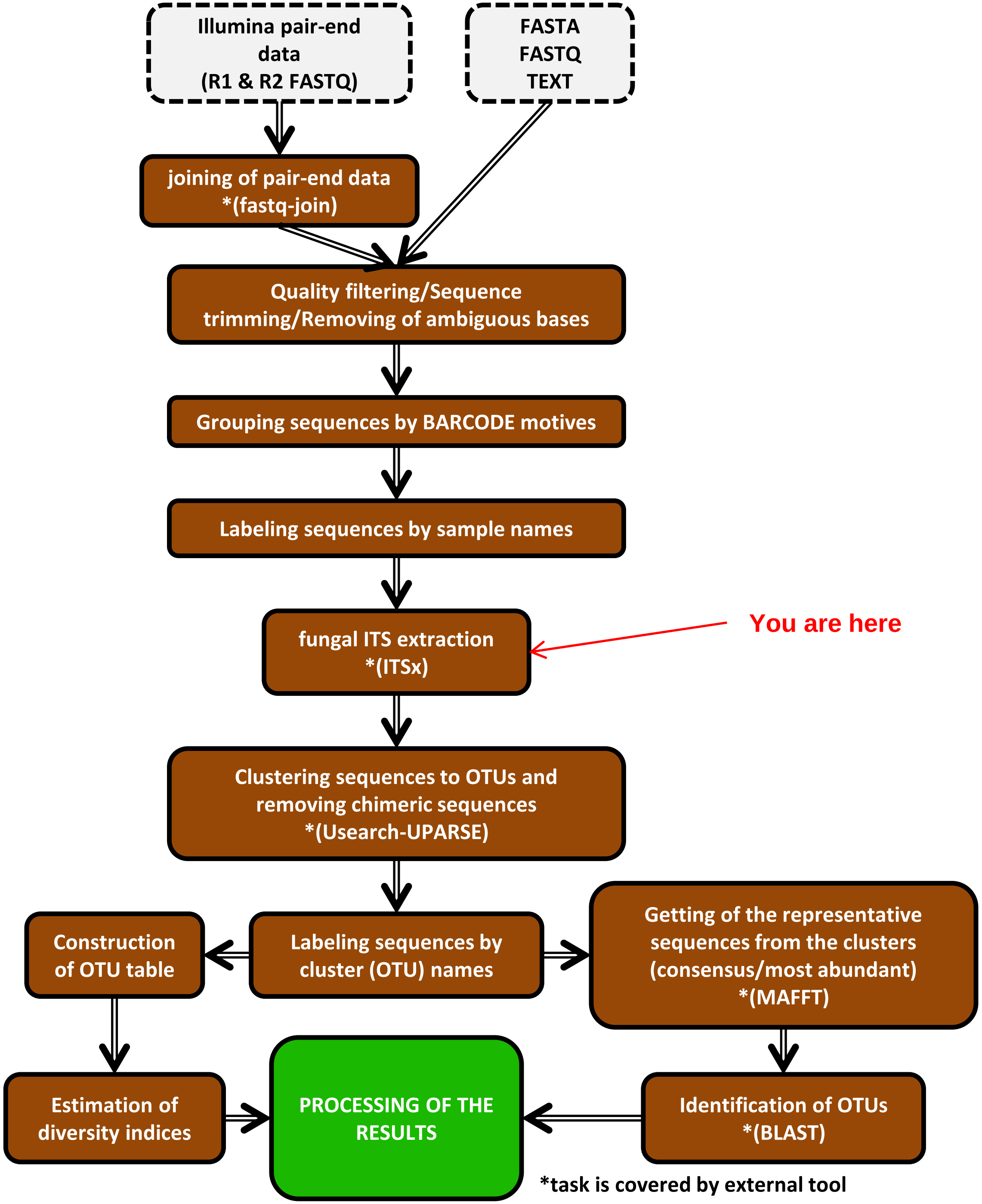


FASTA FILE(S): ITS_example_joined_qm30_fwdtag_revtag [SeqCount: 19289 Ambiguous: 0 Min-len: 30 Max-len: 534 Max-qual...

```
No. HEAD                                                length hits tag
1   M03794:7:000000000-AMUBJ:1:1101:9465:1956|SAMPLE06   352   0    1
2   M03794:7:000000000-AMUBJ:1:1101:9648:2210|SAMPLE06   352   0    1
3   M03794:7:000000000-AMUBJ:1:1101:18084:2212|SAMPLE18  287   0    1
4   M03794:7:000000000-AMUBJ:1:1101:11701:2328|SAMPLE17
5   M03794:7:000000000-AMUBJ:1:1101:20058:2552|SAMPLE25
6   M03794:7:000000000-AMUBJ:1:1101:21418:2679|SAMPLE25
7   M03794:7:000000000-AMUBJ:1:1101:11024:2758|SAMPLE20
8   M03794:7:000000000-AMUBJ:1:1101:10455:2785|SAMPLE17
9   M03794:7:000000000-AMUBJ:1:1101:20792:2826|SAMPLE07
10  M03794:7:000000000-AMUBJ:1:1101:24641:2936|SAMPLE20
11  M03794:7:000000000-AMUBJ:1:1101:11364:3049|SAMPLE17  325   0    1
12  M03794:7:000000000-AMUBJ:1:1101:25518:3208|SAMPLE05  283   0    1
13  M03794:7:000000000-AMUBJ:1:1101:16867:3234|SAMPLE03  305   0    1
```

19289 replacements have been made.

Visible region starts at 1 bp and ends at 12 bp



Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

You are here

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

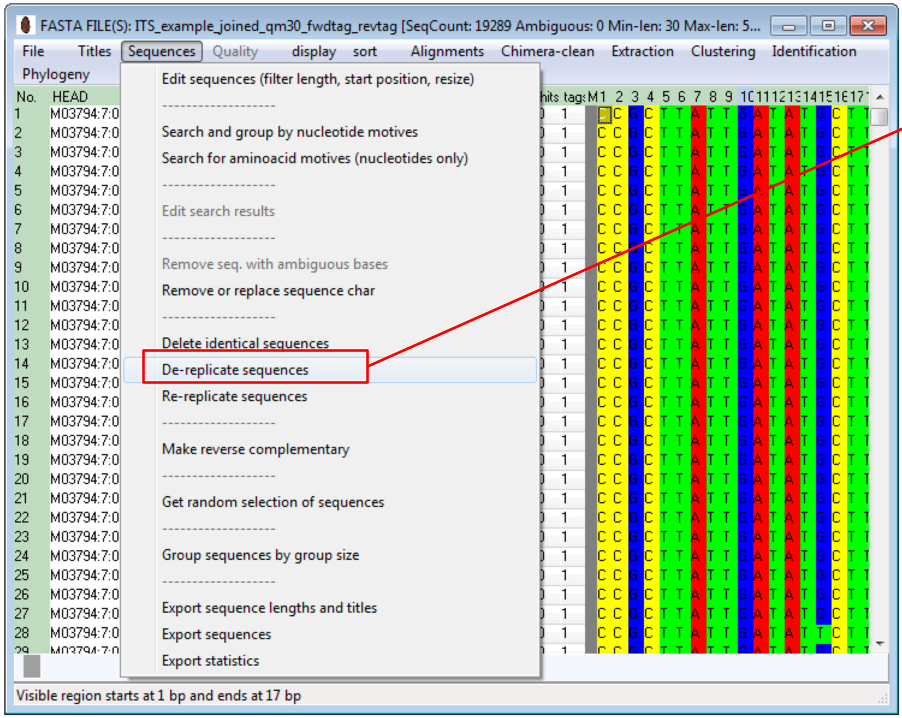
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

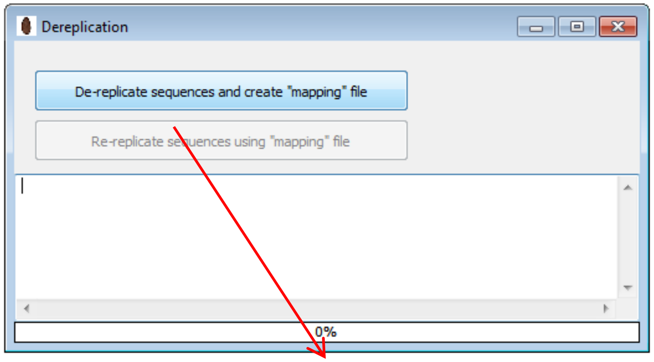
*task is covered by external tool

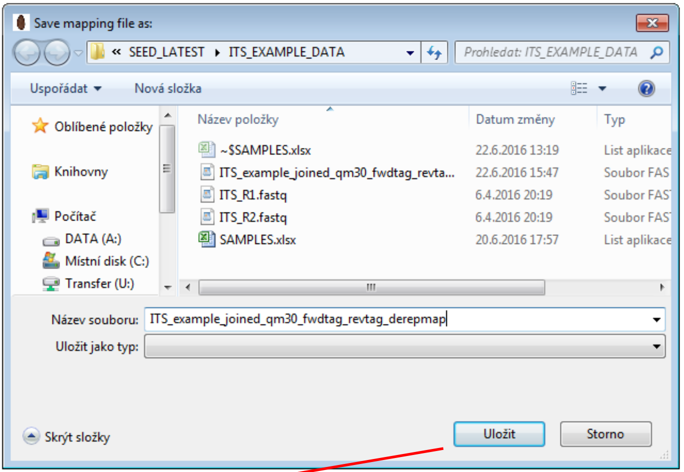
De-replicate sequences before ITS extraction…



FASTA FILE(S): ITS_example_joined_qm30_fwdtag_revtag [SeqCount: 19289 Ambiguous: 0 Min-len: 30 Max-len: 5...

Visible region starts at 1 bp and ends at 17 bp

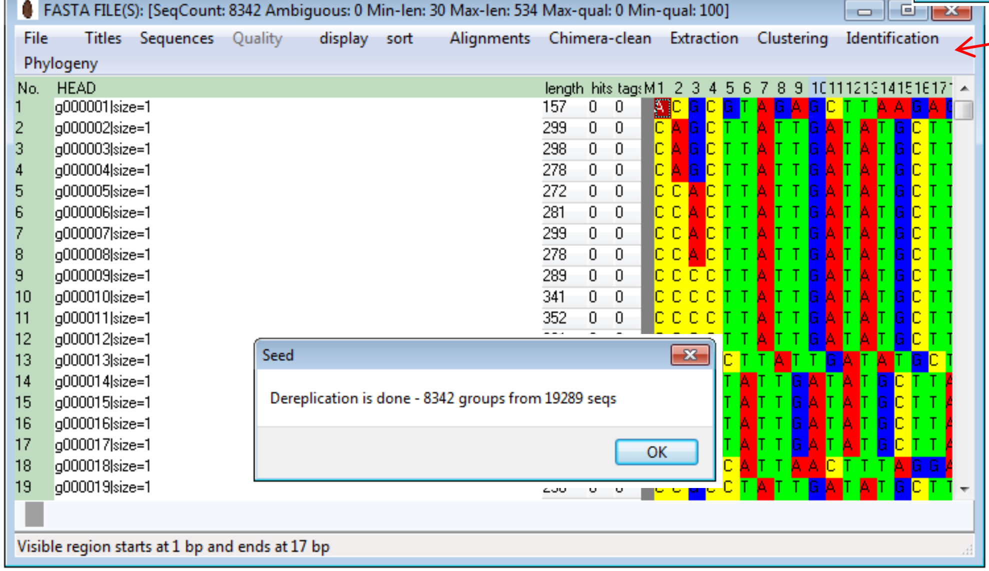




Název položky	Datum změny	Typ
~$SAMPLES.xlsx	22.6.2016 13:19	List aplikace
ITS_example_joined_qm30_fwdtag_revta...	22.6.2016 15:47	Soubor FAS
ITS_R1.fastq	6.4.2016 20:19	Soubor FAS
ITS_R2.fastq	6.4.2016 20:19	Soubor FAS
SAMPLES.xlsx	20.6.2016 17:57	List aplikace

Název souboru: ITS_example_joined_qm30_fwdtag_revtag_derepmap

Uložit jako typ:



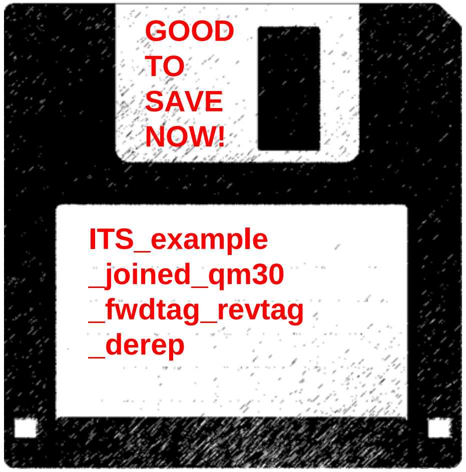
FASTA FILE(S): [SeqCount: 8342 Ambiguous: 0 Min-len: 30 Max-len: 534 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tags
1	g000001|size=1	157	0	0
2	g000002|size=1	299	0	0
3	g000003|size=1	298	0	0
4	g000004|size=1	278	0	0
5	g000005|size=1	272	0	0
6	g000006|size=1	281	0	0
7	g000007|size=1	299	0	0
8	g000008|size=1	278	0	0
9	g000009|size=1	289	0	0
10	g000010|size=1	341	0	0
11	g000011|size=1	352	0	0
12	g000012|size=1			
13	g000013|size=1			
14	g000014|size=1			
15	g000015|size=1			
16	g000016|size=1			
17	g000017|size=1			
18	g000018|size=1			
19	g000019|size=1			

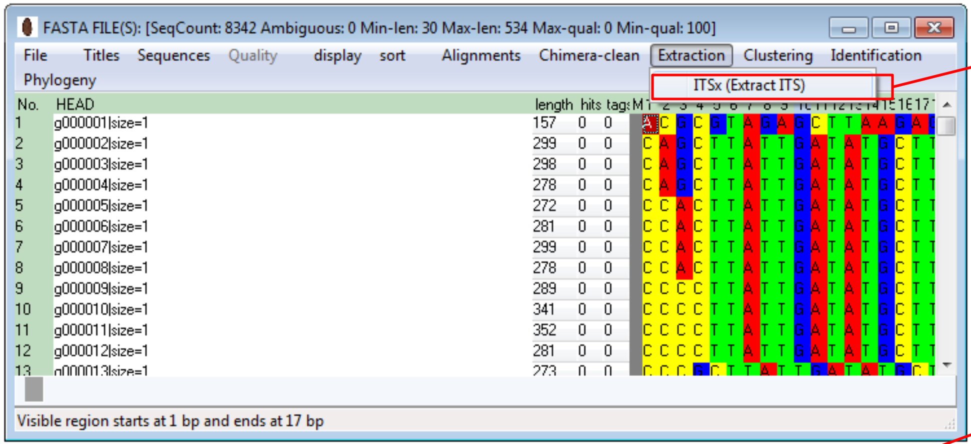
Seed

Dereplication is done - 8342 groups from 19289 seqs

Visible region starts at 1 bp and ends at 17 bp

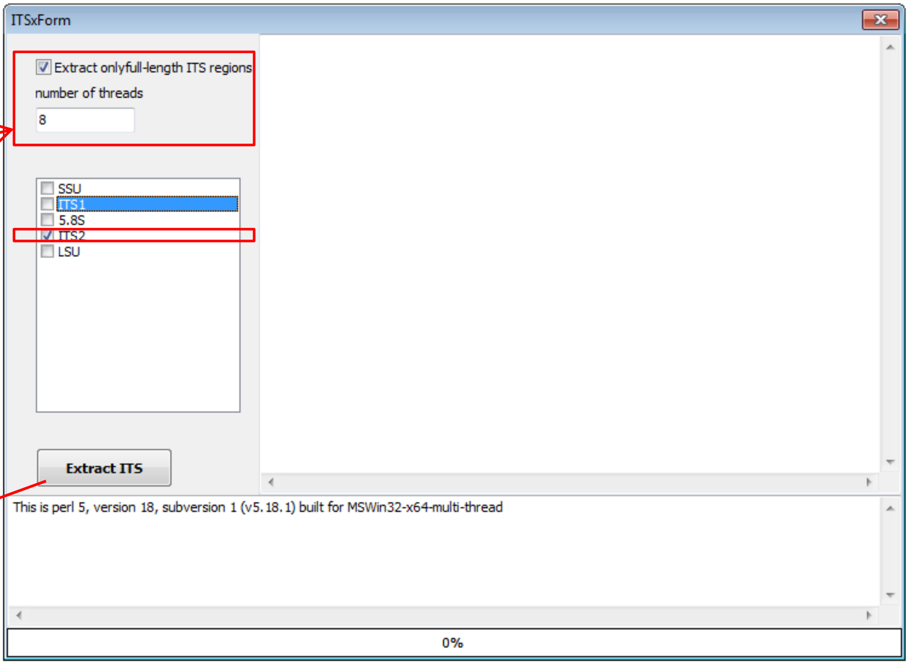


Extract fungal Internal Transcribed Spacers...



No.	HEAD	length	hits	tag:	M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
1	g000001|size=1	157	0	0	A C G C G T A G A G C T T A A G A
2	g000002|size=1	299	0	0	C A G C T T A T T G A T A T G C T
3	g000003|size=1	298	0	0	C A G C T T A T T G A T A T G C T
4	g000004|size=1	278	0	0	C A G C T T A T T G A T A T G C T
5	g000005|size=1	272	0	0	C C A C T T A T T G A T A T G C T
6	g000006|size=1	281	0	0	C C A C T T A T T G A T A T G C T
7	g000007|size=1	299	0	0	C C A C T T A T T G A T A T G C T
8	g000008|size=1	278	0	0	C C A C T T A T T G A T A T G C T
9	g000009|size=1	289	0	0	C C C C T T A T T G A T A T G C T
10	g000010|size=1	341	0	0	C C C C T T A T T G A T A T G C T
11	g000011|size=1	352	0	0	C C C C T T A T T G A T A T G C T
12	g000012|size=1	281	0	0	C C C C T T A T T G A T A T G C T
13	g000013|size=1	273	0	0	C C C G C T T A T T G A T A T G C

Visible region starts at 1 bp and ends at 17 bp



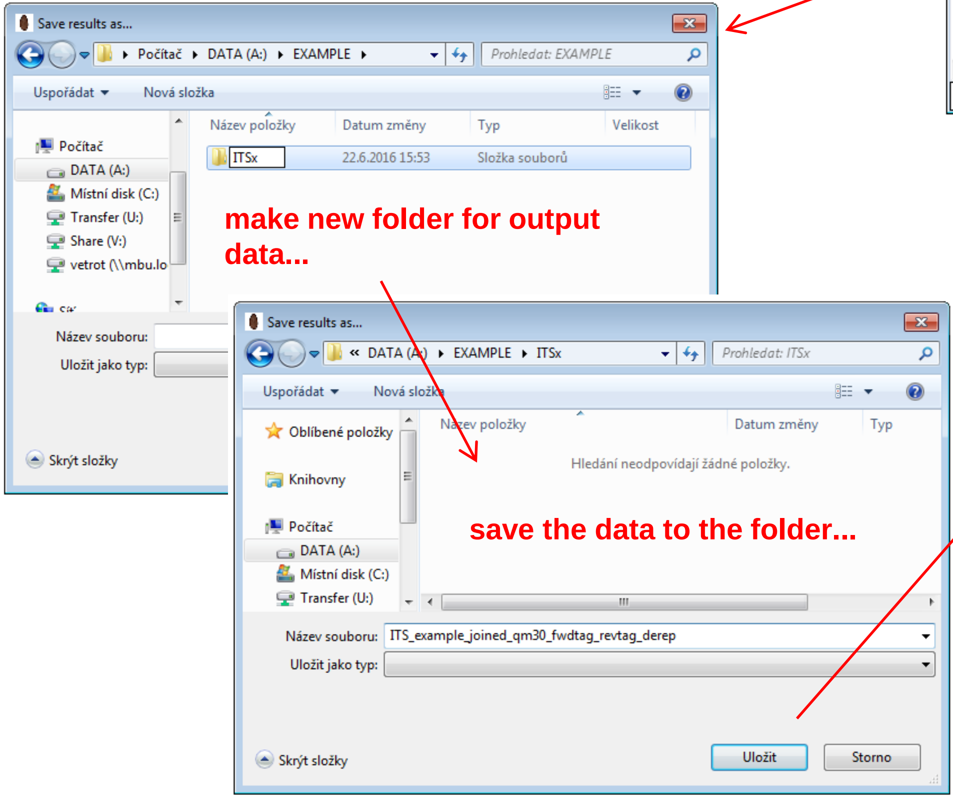
☑ Extract onlyfull-length ITS regions

number of threads: 8

☐ SSU
☐ ITS1
☐ 5.8S
☑ ITS2
☐ LSU

This is perl 5, version 18, subversion 1 (v5.18.1) built for MSWin32-x64-multi-thread

0%



Název položky	Datum změny	Typ	Velikost
ITSx	22.6.2016 15:53	Složka souborů	

Název souboru:

Uložit jako typ:

make new folder for output data...

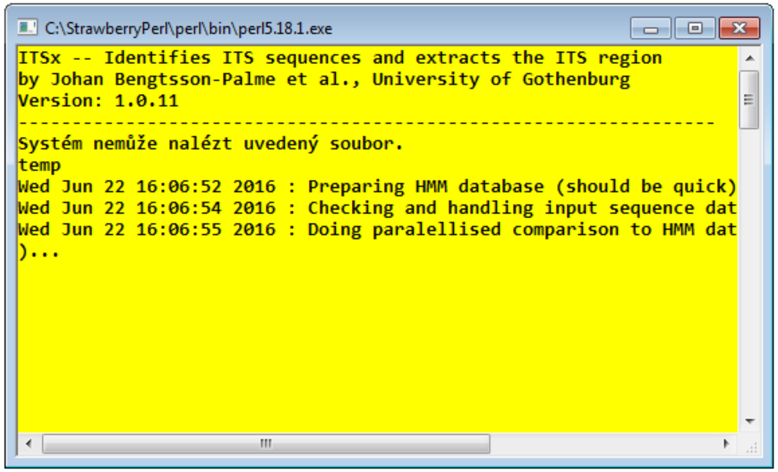
Název položky | Datum změny | Typ

Hledání neodpovídají žádné položky.

save the data to the folder...

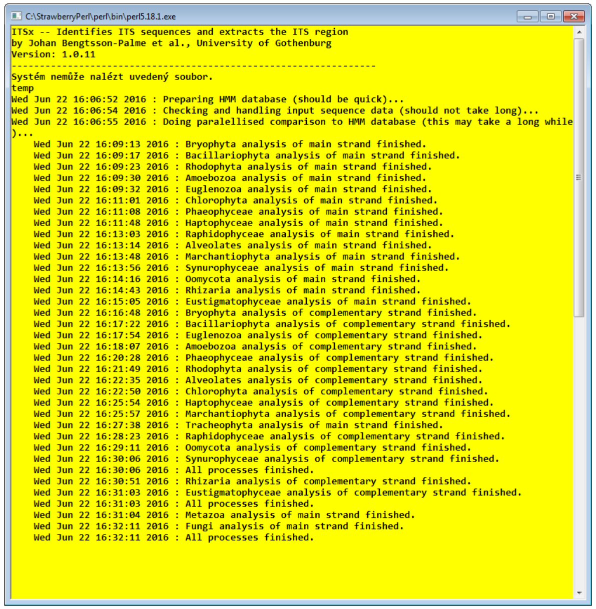
Název souboru: ITS_example_joined_qm30_fwdtag_revtag_derep

Uložit jako typ:

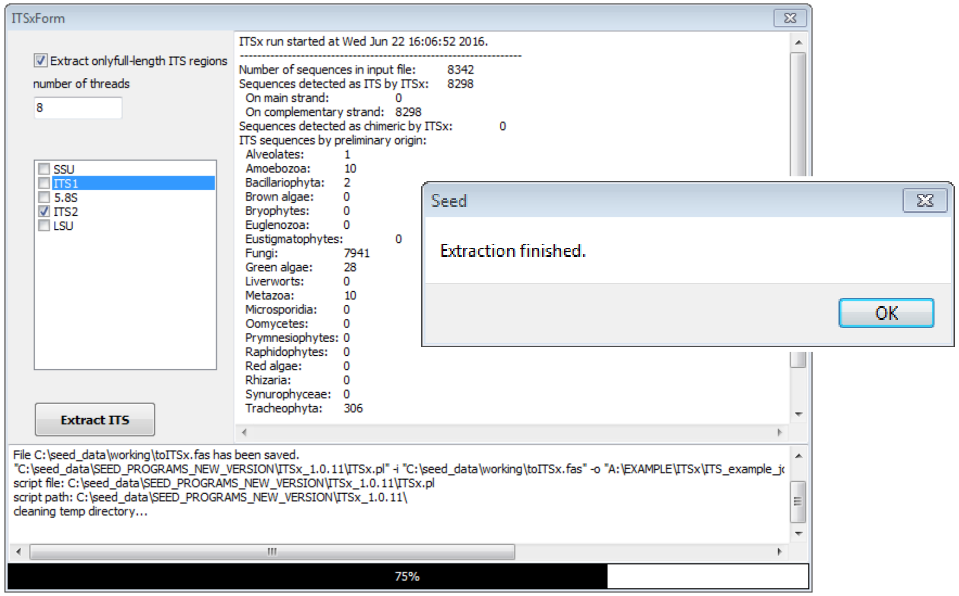


```
ITSx -- Identifies ITS sequences and extracts the ITS region
by Johan Bengtsson-Palme et al., University of Gothenburg
Version: 1.0.11
-----------------------------------------------------------------
Systém nemůže nalézt uvedený soubor.
temp
Wed Jun 22 16:06:52 2016 : Preparing HMM database (should be quick)
Wed Jun 22 16:06:54 2016 : Checking and handling input sequence dat
Wed Jun 22 16:06:55 2016 : Doing paralellised comparison to HMM dat
)...
```

extraction should starts immediately...



```
ITSx -- Identifies ITS sequences and extracts the ITS region
by Johan Bengtsson-Palme et al., University of Gothenburg
Version: 1.0.11
-----------------------------------------------------------------
Systém nemůže nalézt uvedený soubor.
temp
Wed Jun 22 16:06:52 2016 : Preparing HMM database (should be quick)...
Wed Jun 22 16:06:54 2016 : Checking and handling input sequence data (should not take long)...
Wed Jun 22 16:06:55 2016 : Doing paralellised comparison to HMM database (this may take a long while
)...
    Wed Jun 22 16:09:13 2016 : Bryophyta analysis of main strand finished.
    Wed Jun 22 16:09:17 2016 : Bacillariophyta analysis of main strand finished.
    Wed Jun 22 16:09:23 2016 : Rhodophyta analysis of main strand finished.
    Wed Jun 22 16:09:30 2016 : Amoebozoa analysis of main strand finished.
    Wed Jun 22 16:09:32 2016 : Euglenozoa analysis of main strand finished.
    Wed Jun 22 16:11:01 2016 : Chlorophyta analysis of main strand finished.
    Wed Jun 22 16:11:08 2016 : Phaeophyceae analysis of main strand finished.
    Wed Jun 22 16:11:48 2016 : Haptophyceae analysis of main strand finished.
    Wed Jun 22 16:13:03 2016 : Raphidophyceae analysis of main strand finished.
    Wed Jun 22 16:13:14 2016 : Alveolates analysis of main strand finished.
    Wed Jun 22 16:13:48 2016 : Marchantiophyta analysis of main strand finished.
    Wed Jun 22 16:13:56 2016 : Synurophyceae analysis of main strand finished.
    Wed Jun 22 16:14:16 2016 : Oomycota analysis of main strand finished.
    Wed Jun 22 16:14:43 2016 : Rhizaria analysis of main strand finished.
    Wed Jun 22 16:15:05 2016 : Eustigmatophyceae analysis of main strand finished.
    Wed Jun 22 16:16:48 2016 : Bryophyta analysis of complementary strand finished.
    Wed Jun 22 16:17:22 2016 : Bacillariophyta analysis of complementary strand finished.
    Wed Jun 22 16:17:54 2016 : Euglenozoa analysis of complementary strand finished.
    Wed Jun 22 16:18:07 2016 : Amoebozoa analysis of complementary strand finished.
    Wed Jun 22 16:20:28 2016 : Phaeophyceae analysis of complementary strand finished.
    Wed Jun 22 16:21:49 2016 : Rhodophyta analysis of complementary strand finished.
    Wed Jun 22 16:22:35 2016 : Alveolates analysis of complementary strand finished.
    Wed Jun 22 16:22:50 2016 : Chlorophyta analysis of complementary strand finished.
    Wed Jun 22 16:25:54 2016 : Haptophyceae analysis of complementary strand finished.
    Wed Jun 22 16:25:57 2016 : Marchantiophyta analysis of complementary strand finished.
    Wed Jun 22 16:27:38 2016 : Tracheophyta analysis of main strand finished.
    Wed Jun 22 16:28:23 2016 : Raphidophyceae analysis of complementary strand finished.
    Wed Jun 22 16:29:11 2016 : Oomycota analysis of complementary strand finished.
    Wed Jun 22 16:30:06 2016 : Synurophyceae analysis of complementary strand finished.
    Wed Jun 22 16:30:06 2016 : All processes finished.
    Wed Jun 22 16:30:51 2016 : Rhizaria analysis of complementary strand finished.
    Wed Jun 22 16:31:03 2016 : Eustigmatophyceae analysis of complementary strand finished.
    Wed Jun 22 16:31:03 2016 : All processes finished.
    Wed Jun 22 16:31:04 2016 : Metazoa analysis of main strand finished.
    Wed Jun 22 16:32:11 2016 : Fungi analysis of main strand finished.
    Wed Jun 22 16:32:11 2016 : All processes finished.
```

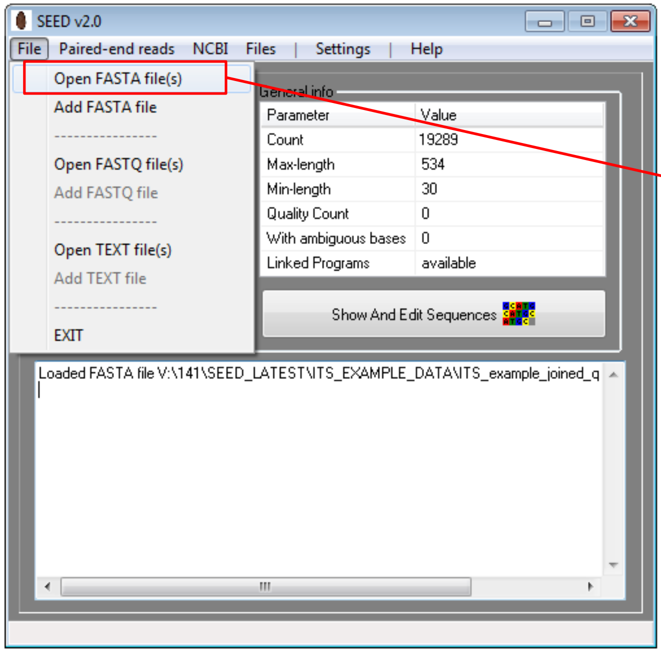


```
ITSx run started at Wed Jun 22 16:06:52 2016.
-----------------------------------------------------------------
Number of sequences in input file:      8342
Sequences detected as ITS by ITSx:      8298
  On main strand:                 0
  On complementary strand:   8298
Sequences detected as chimeric by ITSx:          0
ITS sequences by preliminary origin:
  Alveolates:          1
  Amoebozoa:           10
  Bacillariophyta:     2
  Brown algae:         0
  Bryophytes:          0
  Euglenozoa:          0
  Eustigmatophytes:            0
  Fungi:               7941
  Green algae:         28
  Liverworts:          0
  Metazoa:             10
  Microsporidia:       0
  Oomycetes:           0
  Prymnesiophytes: 0
  Raphidophytes:       0
  Red algae:           0
  Rhizaria:            0
  Synurophyceae:   0
  Tracheophyta:        306
```

```
File C:\seed_data\working\toITSx.fas has been saved.
"C:\seed_data\SEED_PROGRAMS_NEW_VERSION\ITSx_1.0.11\ITSx.pl" -i "C:\seed_data\working\toITSx.fas" -o "A:\EXAMPLE\ITSx\ITS_example_j
script file: C:\seed_data\SEED_PROGRAMS_NEW_VERSION\ITSx_1.0.11\ITSx.pl
script path: C:\seed_data\SEED_PROGRAMS_NEW_VERSION\ITSx_1.0.11\
cleaning temp directory...
```

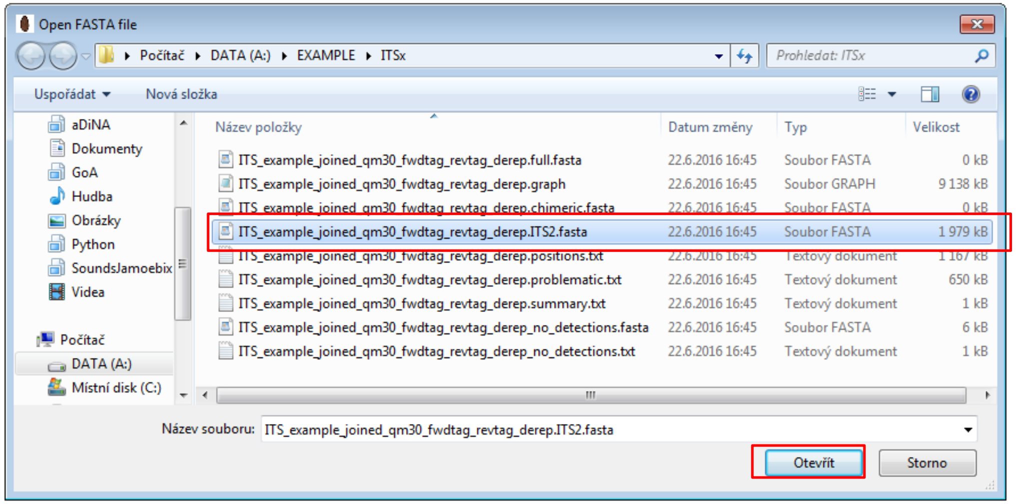
Extraction finished.

close all windows and open the result...



Parameter	Value
Count	19289
Max-length	534
Min-length	30
Quality Count	0
With ambiguous bases	0
Linked Programs	available

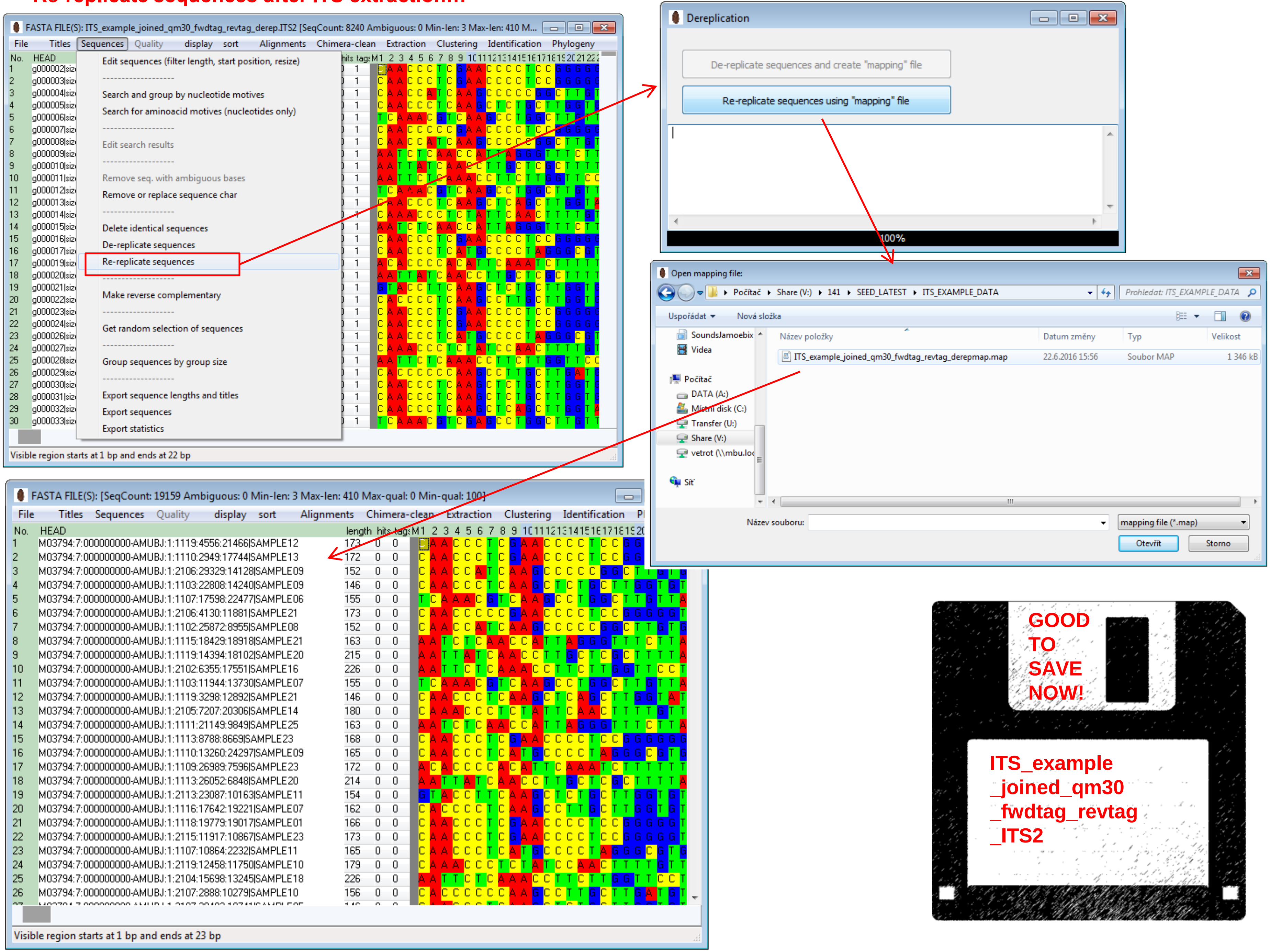
```
Loaded FASTA file V:\141\SEED_LATEST\ITS_EXAMPLE_DATA\ITS_example_joined_q
```



Název položky	Datum změny	Typ	Velikost
ITS_example_joined_qm30_fwdtag_revtag_derep.full.fasta	22.6.2016 16:45	Soubor FASTA	0 kB
ITS_example_joined_qm30_fwdtag_revtag_derep.graph	22.6.2016 16:45	Soubor GRAPH	9 138 kB
ITS_example_joined_qm30_fwdtag_revtag_derep.chimeric.fasta	22.6.2016 16:45	Soubor FASTA	0 kB
ITS_example_joined_qm30_fwdtag_revtag_derep.ITS2.fasta	22.6.2016 16:45	Soubor FASTA	1 979 kB
ITS_example_joined_qm30_fwdtag_revtag_derep.positions.txt	22.6.2016 16:45	Textový dokument	1 167 kB
ITS_example_joined_qm30_fwdtag_revtag_derep.problematic.txt	22.6.2016 16:45	Textový dokument	650 kB
ITS_example_joined_qm30_fwdtag_revtag_derep.summary.txt	22.6.2016 16:45	Textový dokument	1 kB
ITS_example_joined_qm30_fwdtag_revtag_derep_no_detections.fasta	22.6.2016 16:45	Soubor FASTA	6 kB
ITS_example_joined_qm30_fwdtag_revtag_derep_no_detections.txt	22.6.2016 16:45	Textový dokument	1 kB

Název souboru: ITS_example_joined_qm30_fwdtag_revtag_derep.ITS2.fasta

Re-replicate sequences after ITS extraction…



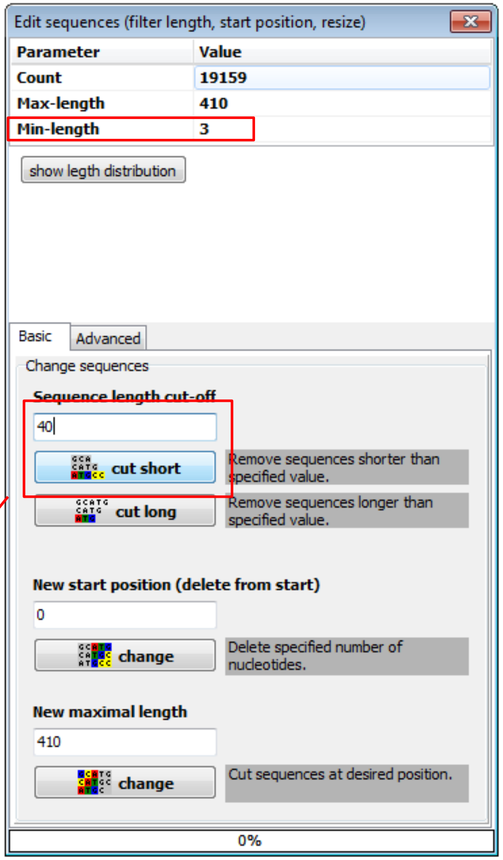
No.	HEAD	length	hits	tag
1	M03794:7:000000000-AMUBJ:1:1119:4556:21466|SAMPLE12	173	0	0
2	M03794:7:000000000-AMUBJ:1:1110:2949:17744|SAMPLE13	172	0	0
3	M03794:7:000000000-AMUBJ:1:2106:29329:14128|SAMPLE09	152	0	0
4	M03794:7:000000000-AMUBJ:1:1103:22808:14240|SAMPLE09	146	0	0
5	M03794:7:000000000-AMUBJ:1:1107:17598:22477|SAMPLE06	155	0	0
6	M03794:7:000000000-AMUBJ:1:2106:4130:11881|SAMPLE21	173	0	0
7	M03794:7:000000000-AMUBJ:1:1102:25872:8955|SAMPLE08	152	0	0
8	M03794:7:000000000-AMUBJ:1:1115:18429:18918|SAMPLE21	163	0	0
9	M03794:7:000000000-AMUBJ:1:1119:14394:18102|SAMPLE20	215	0	0
10	M03794:7:000000000-AMUBJ:1:2102:6355:17551|SAMPLE16	226	0	0
11	M03794:7:000000000-AMUBJ:1:1103:11944:13730|SAMPLE07	155	0	0
12	M03794:7:000000000-AMUBJ:1:1119:3298:12892|SAMPLE21	146	0	0
13	M03794:7:000000000-AMUBJ:1:2105:7207:20306|SAMPLE14	180	0	0
14	M03794:7:000000000-AMUBJ:1:1111:21149:9849|SAMPLE25	163	0	0
15	M03794:7:000000000-AMUBJ:1:1113:8788:8669|SAMPLE23	168	0	0
16	M03794:7:000000000-AMUBJ:1:1110:13260:24297|SAMPLE09	165	0	0
17	M03794:7:000000000-AMUBJ:1:1109:26989:7596|SAMPLE23	172	0	0
18	M03794:7:000000000-AMUBJ:1:1113:26052:6848|SAMPLE20	214	0	0
19	M03794:7:000000000-AMUBJ:1:2113:23087:10163|SAMPLE11	154	0	0
20	M03794:7:000000000-AMUBJ:1:1116:17642:19221|SAMPLE07	162	0	0
21	M03794:7:000000000-AMUBJ:1:1118:19779:19017|SAMPLE01	166	0	0
22	M03794:7:000000000-AMUBJ:1:2115:11917:10867|SAMPLE23	173	0	0
23	M03794:7:000000000-AMUBJ:1:1107:10864:2232|SAMPLE11	165	0	0
24	M03794:7:000000000-AMUBJ:1:2119:12458:11750|SAMPLE10	179	0	0
25	M03794:7:000000000-AMUBJ:1:2104:15698:13245|SAMPLE18	226	0	0
26	M03794:7:000000000-AMUBJ:1:2107:2888:10279|SAMPLE10	156	0	0

Visible region starts at 1 bp and ends at 23 bp

Remove short sequences...

FASTA FILE(S): [SeqCount: 19159 Ambiguous: 0 Min-len: 3 Max-len: 410 Max-qual: 0 Min-qual: 100]

Visible region starts at 1 bp and ends at 23 bp



Edit sequences (filter length, start position, resize)

Parameter	Value
Count	19159
Max-length	410
Min-length	3

Sequence length cut-off

40

Remove sequences shorter than specified value.

Remove sequences longer than specified value.

New start position (delete from start)

0

Delete specified number of nucleotides.

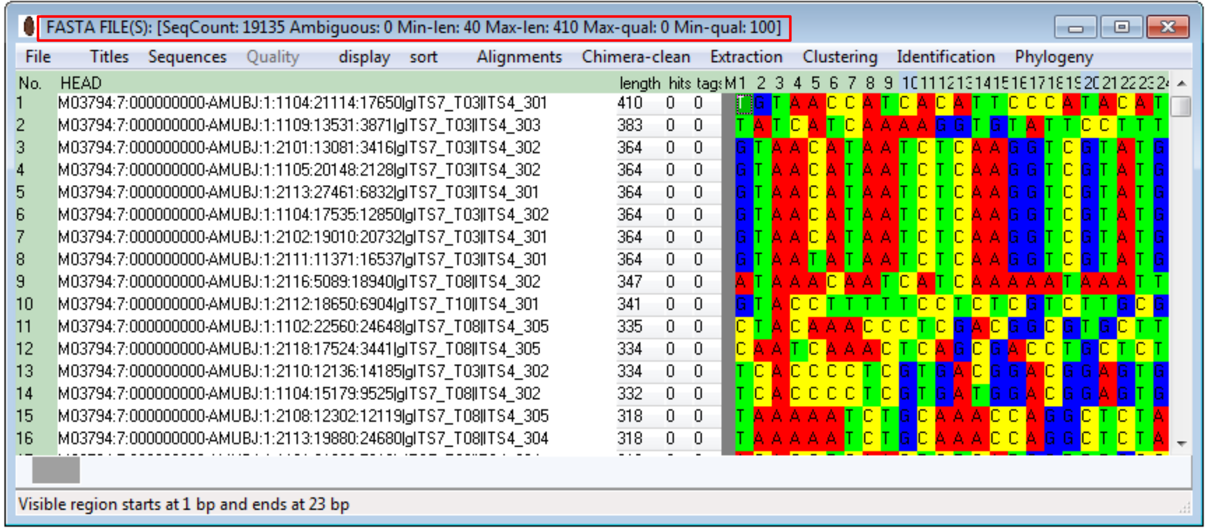
New maximal length

410

Cut sequences at desired position.

0%

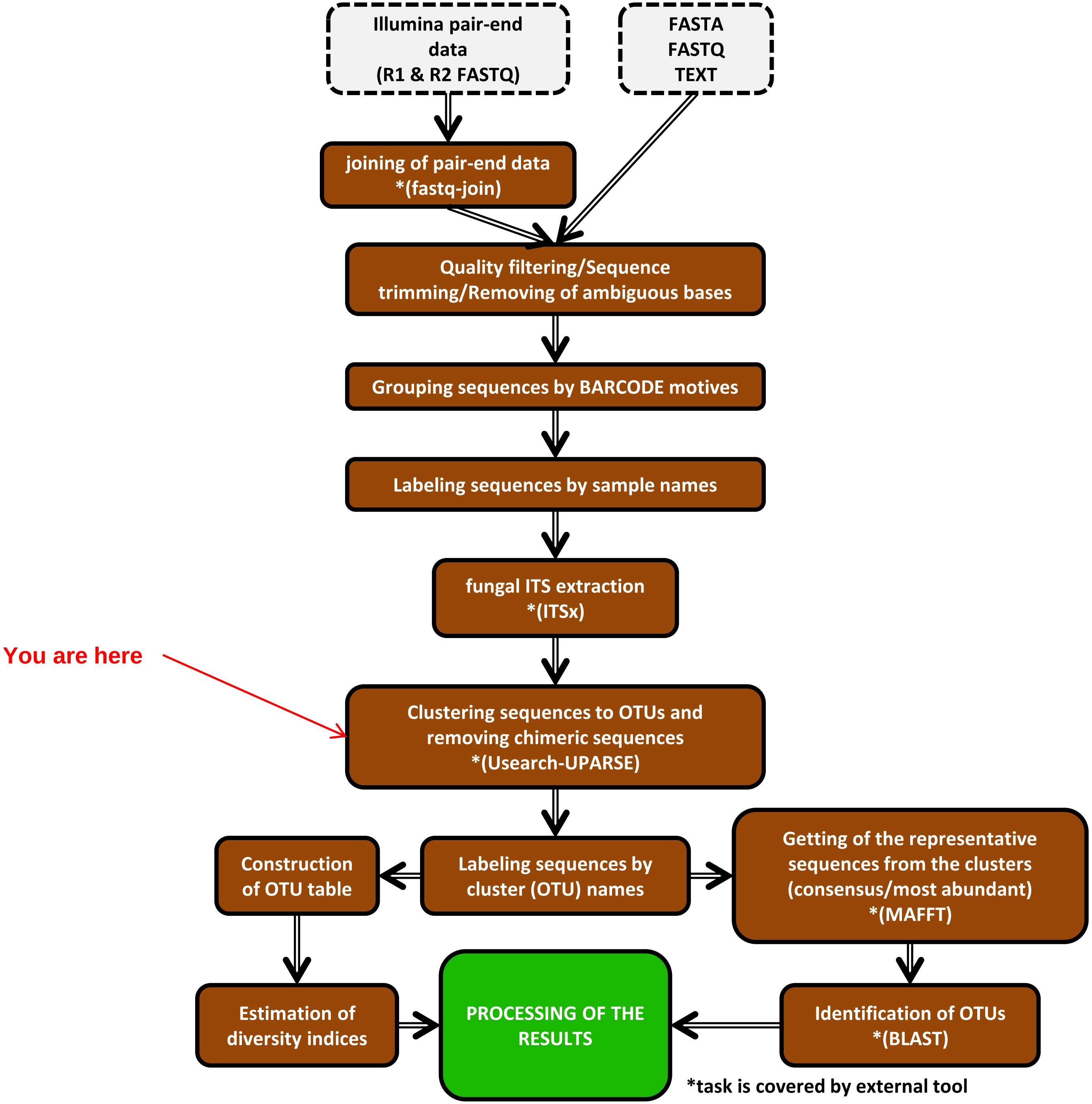
set sequence length cutoff threshold...



FASTA FILE(S): [SeqCount: 19135 Ambiguous: 0 Min-len: 40 Max-len: 410 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tags
1	M03794:7:000000000-AMUBJ:1:1104:21114:17650|gITS7_T03|ITS4_301	410	0	0
2	M03794:7:000000000-AMUBJ:1:1109:13531:3871|gITS7_T03|ITS4_303	383	0	0
3	M03794:7:000000000-AMUBJ:1:2101:13081:3416|gITS7_T03|ITS4_302	364	0	0
4	M03794:7:000000000-AMUBJ:1:1105:20148:2128|gITS7_T03|ITS4_302	364	0	0
5	M03794:7:000000000-AMUBJ:1:2113:27461:6832|gITS7_T03|ITS4_301	364	0	0
6	M03794:7:000000000-AMUBJ:1:1104:17535:12850|gITS7_T03|ITS4_302	364	0	0
7	M03794:7:000000000-AMUBJ:1:2102:19010:20732|gITS7_T03|ITS4_301	364	0	0
8	M03794:7:000000000-AMUBJ:1:2111:11371:16537|gITS7_T03|ITS4_301	364	0	0
9	M03794:7:000000000-AMUBJ:1:2116:5089:18940|gITS7_T08|ITS4_302	347	0	0
10	M03794:7:000000000-AMUBJ:1:2112:18650:6904|gITS7_T10|ITS4_301	341	0	0
11	M03794:7:000000000-AMUBJ:1:1102:22560:24648|gITS7_T08|ITS4_305	335	0	0
12	M03794:7:000000000-AMUBJ:1:2118:17524:3441|gITS7_T08|ITS4_305	334	0	0
13	M03794:7:000000000-AMUBJ:1:2110:12136:14185|gITS7_T03|ITS4_302	334	0	0
14	M03794:7:000000000-AMUBJ:1:1104:15179:9525|gITS7_T08|ITS4_302	332	0	0
15	M03794:7:000000000-AMUBJ:1:2108:12302:12119|gITS7_T08|ITS4_305	318	0	0
16	M03794:7:000000000-AMUBJ:1:2113:19880:24680|gITS7_T08|ITS4_304	318	0	0

Visible region starts at 1 bp and ends at 23 bp



Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

You are here

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

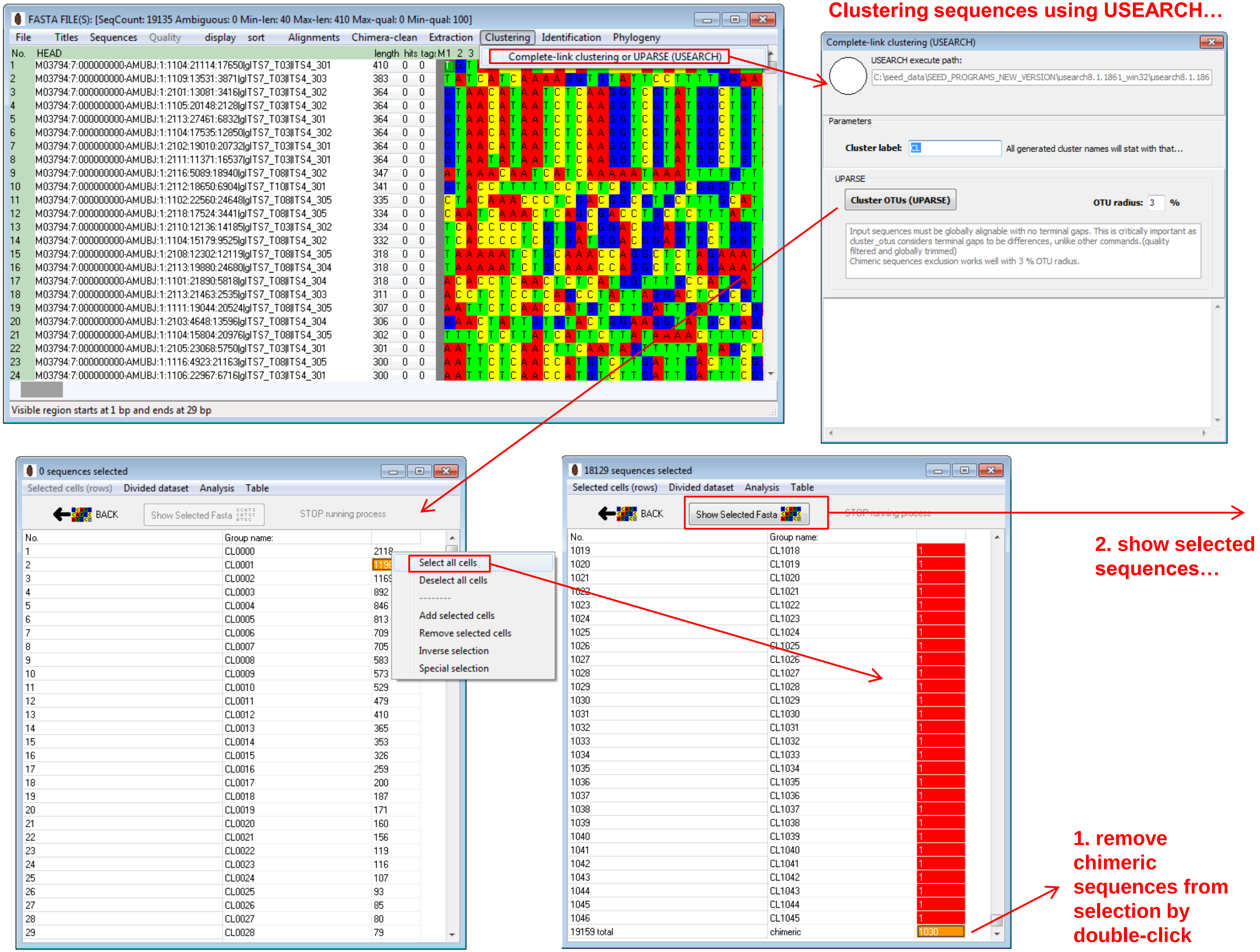
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool

Clustering sequences using USEARCH…



FASTA FILE(S): [SeqCount: 19135 Ambiguous: 0 Min-len: 40 Max-len: 410 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tags
1	M03794:7:000000000-AMUBJ:1:1104:21114:17650|gITS7_T03|ITS4_301	410	0	0
2	M03794:7:000000000-AMUBJ:1:1109:13531:3871|gITS7_T03|ITS4_303	383	0	0
3	M03794:7:000000000-AMUBJ:1:2101:13081:3416|gITS7_T03|ITS4_302	364	0	0
4	M03794:7:000000000-AMUBJ:1:1105:20148:2128|gITS7_T03|ITS4_302	364	0	0
5	M03794:7:000000000-AMUBJ:1:2113:27461:6832|gITS7_T03|ITS4_301	364	0	0
6	M03794:7:000000000-AMUBJ:1:1104:17535:12850|gITS7_T03|ITS4_302	364	0	0
7	M03794:7:000000000-AMUBJ:1:2102:19010:20732|gITS7_T03|ITS4_301	364	0	0
8	M03794:7:000000000-AMUBJ:1:2111:11371:16537|gITS7_T03|ITS4_301	364	0	0
9	M03794:7:000000000-AMUBJ:1:2116:5089:18940|gITS7_T08|ITS4_302	347	0	0
10	M03794:7:000000000-AMUBJ:1:2112:18650:6904|gITS7_T10|ITS4_301	341	0	0
11	M03794:7:000000000-AMUBJ:1:1102:22560:24648|gITS7_T08|ITS4_305	335	0	0
12	M03794:7:000000000-AMUBJ:1:2118:17524:3441|gITS7_T08|ITS4_305	334	0	0
13	M03794:7:000000000-AMUBJ:1:2110:12136:14185|gITS7_T03|ITS4_302	334	0	0
14	M03794:7:000000000-AMUBJ:1:1104:15179:9525|gITS7_T08|ITS4_302	332	0	0
15	M03794:7:000000000-AMUBJ:1:2108:12302:12119|gITS7_T08|ITS4_305	318	0	0
16	M03794:7:000000000-AMUBJ:1:2113:19880:24680|gITS7_T08|ITS4_304	318	0	0
17	M03794:7:000000000-AMUBJ:1:1101:21890:5818|gITS7_T08|ITS4_304	318	0	0
18	M03794:7:000000000-AMUBJ:1:2113:21463:2535|gITS7_T08|ITS4_303	311	0	0
19	M03794:7:000000000-AMUBJ:1:1111:19044:20524|gITS7_T08|ITS4_305	307	0	0
20	M03794:7:000000000-AMUBJ:1:2103:4648:13596|gITS7_T08|ITS4_304	306	0	0
21	M03794:7:000000000-AMUBJ:1:1104:15804:20976|gITS7_T08|ITS4_305	302	0	0
22	M03794:7:000000000-AMUBJ:1:2105:23068:5750|gITS7_T03|ITS4_301	301	0	0
23	M03794:7:000000000-AMUBJ:1:1116:4923:21163|gITS7_T08|ITS4_305	300	0	0
24	M03794:7:000000000-AMUBJ:1:1106:22967:6716|gITS7_T03|ITS4_301	300	0	0

Visible region starts at 1 bp and ends at 29 bp

Complete-link clustering (USEARCH)

USEARCH execute path: C:\seed_data\SEED_PROGRAMS_NEW_VERSION\usearch8.1.1861_win32\usearch8.1.186

Parameters

Cluster label: CL All generated cluster names will stat with that…

UPARSE

OTU radius: 3 %

Input sequences must be globally alignable with no terminal gaps. This is critically important as cluster_otus considers terminal gaps to be differences, unlike other commands.(quality filtered and globally trimmed)
Chimeric sequences exclusion works well with 3 % OTU radius.

0 sequences selected

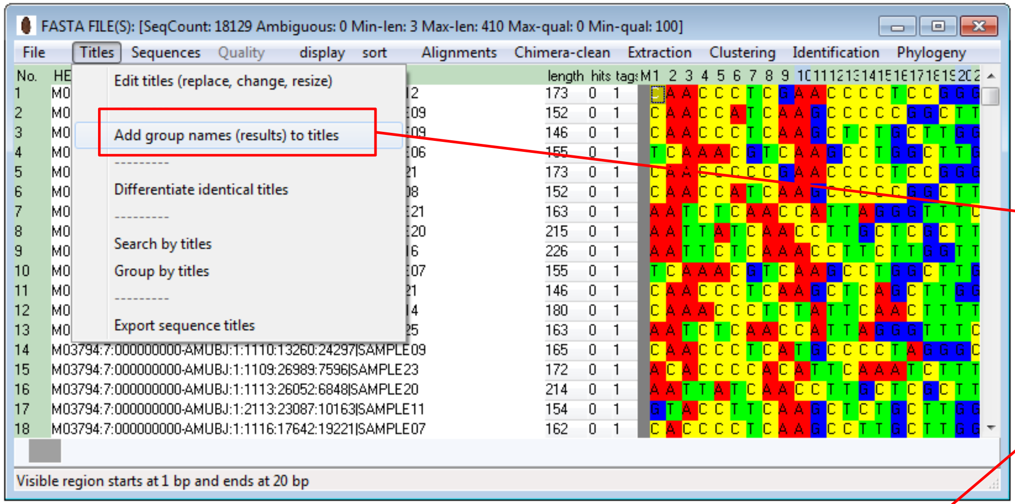
No.	Group name:	
1	CL0000	2118
2	CL0001	119
3	CL0002	116
4	CL0003	892
5	CL0004	846
6	CL0005	813
7	CL0006	709
8	CL0007	705
9	CL0008	583
10	CL0009	573
11	CL0010	529
12	CL0011	479
13	CL0012	410
14	CL0013	365
15	CL0014	353
16	CL0015	326
17	CL0016	259
18	CL0017	200
19	CL0018	187
20	CL0019	171
21	CL0020	160
22	CL0021	156
23	CL0022	119
24	CL0023	116
25	CL0024	107
26	CL0025	93
27	CL0026	85
28	CL0027	80
29	CL0028	79

18129 sequences selected

No.	Group name:	
1019	CL1018	1
1020	CL1019	1
1021	CL1020	1
1022	CL1021	1
1023	CL1022	1
1024	CL1023	1
1025	CL1024	1
1026	CL1025	1
1027	CL1026	1
1028	CL1027	1
1029	CL1028	1
1030	CL1029	1
1031	CL1030	1
1032	CL1031	1
1033	CL1032	1
1034	CL1033	1
1035	CL1034	1
1036	CL1035	1
1037	CL1036	1
1038	CL1037	1
1039	CL1038	1
1040	CL1039	1
1041	CL1040	1
1042	CL1041	1
1043	CL1042	1
1044	CL1043	1
1045	CL1044	1
1046	CL1045	1
19159 total	chimeric	1030

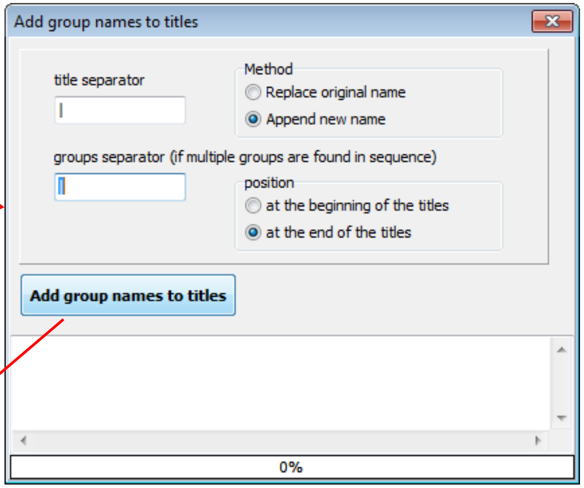
2. show selected sequences…

1. remove chimeric sequences from selection by double-click



No.	HE	length	hits	tags	M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	2
1	M0 12	173	0	1		C	A	A	C	C	C	T	C	G	A	A	C	C	C	C	T	C	C	G	G	G
2	M0 09	152	0	1		C	A	A	C	C	A	T	C	A	A	G	C	C	C	C	C	G	G	C	T	T
3	M0 09	146	0	1		C	A	A	C	C	C	T	C	A	A	G	C	T	C	T	G	C	T	T	G	G
4	M0 06	155	0	1		T	C	A	A	A	C	G	T	C	A	A	G	C	C	T	G	G	C	T	T	G
5	M0 21	173	0	1		C	A	A	C	C	C	C	C	G	A	A	C	C	C	C	T	C	C	G	G	G
6	M0 08	152	0	1		C	A	A	C	C	A	T	C	A	A	G	C	C	C	C	C	G	G	C	T	T
7	M0 21	163	0	1		A	A	T	C	T	C	A	A	C	C	A	T	T	A	G	G	G	T	T	T	C
8	M0 20	215	0	1		A	A	T	T	A	T	C	A	A	C	C	T	T	G	C	T	C	G	C	T	T
9	M0 16	226	0	1		A	A	T	T	C	T	C	A	A	A	C	C	T	T	C	T	T	G	G	T	T
10	M0 07	155	0	1		T	C	A	A	A	C	G	T	C	A	A	G	C	C	T	G	G	C	T	T	G
11	M0 21	146	0	1		C	A	A	C	C	C	T	C	A	A	G	C	T	C	A	G	C	T	T	G	G
12	M0 14	180	0	1		C	A	A	A	C	C	C	T	C	T	A	T	T	C	A	A	C	T	T	T	T
13	M0 25	163	0	1		A	A	T	C	T	C	A	A	C	C	A	T	T	A	G	G	G	T	T	T	C
14	M03794:7:000000000-AMUBJ:1:1110:13260:24297|SAMPLE09	165	0	1		C	A	A	C	C	C	T	C	A	T	G	C	C	C	C	T	A	G	G	G	C
15	M03794:7:000000000-AMUBJ:1:1109:26989:7596|SAMPLE23	172	0	1		A	C	A	C	C	C	C	A	C	A	T	T	C	A	A	A	T	C	T	T	T
16	M03794:7:000000000-AMUBJ:1:1113:26052:6848|SAMPLE20	214	0	1		A	A	T	T	A	T	C	A	A	C	C	T	T	G	C	T	C	G	C	T	T
17	M03794:7:000000000-AMUBJ:1:2113:23087:10163|SAMPLE11	154	0	1		G	T	A	C	C	T	T	C	A	A	G	C	T	C	T	G	C	T	T	G	G
18	M03794:7:000000000-AMUBJ:1:1116:17642:19221|SAMPLE07	162	0	1		C	A	C	C	C	C	T	C	A	A	G	C	C	T	T	G	C	T	T	G	G

Visible region starts at 1 bp and ends at 20 bp



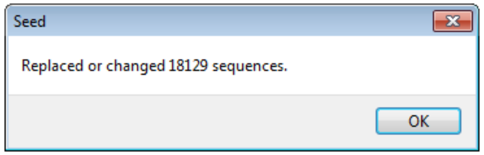
title separator: |

Method
○ Replace original name
● Append new name

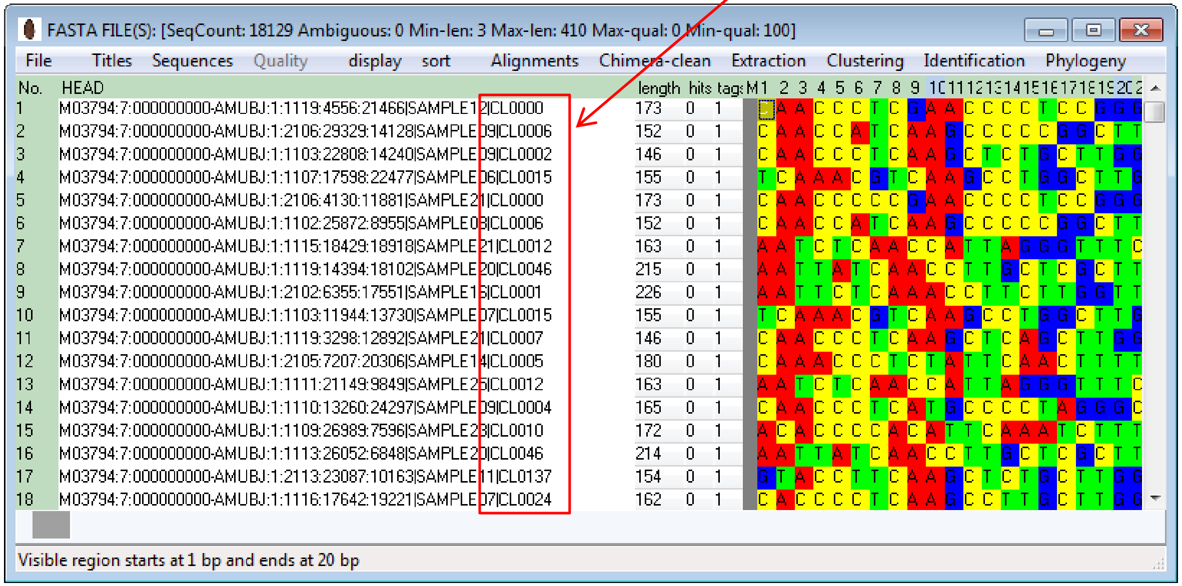
groups separator (if multiple groups are found in sequence): |

position
○ at the beginning of the titles
● at the end of the titles

Add cluster names to titles…



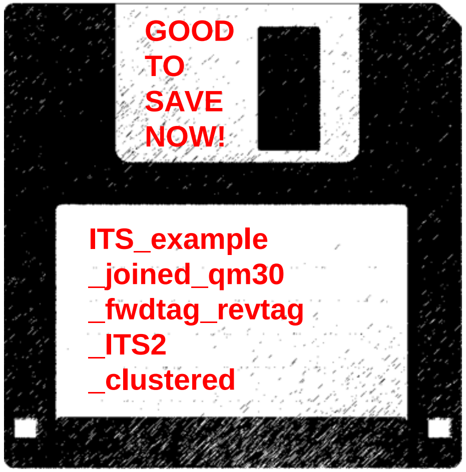
Replaced or changed 18129 sequences.

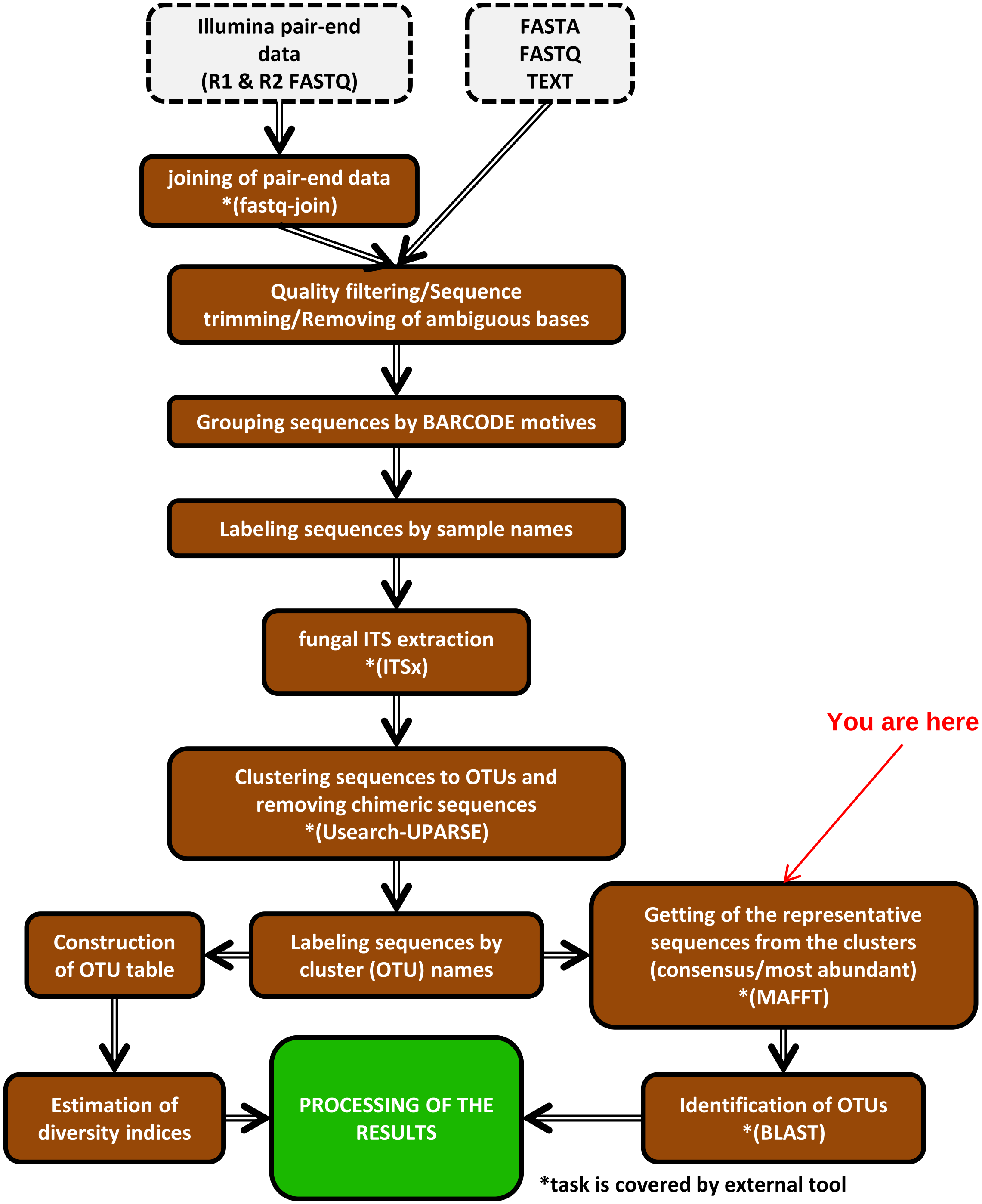


No.	HEAD	length	hits	tags	M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	2
1	M03794:7:000000000-AMUBJ:1:1119:4556:21466|SAMPLE12|CL0000	173	0	1		C	A	A	C	C	C	T	C	G	A	A	C	C	C	C	T	C	C	G	G	G
2	M03794:7:000000000-AMUBJ:1:2106:29329:14128|SAMPLE09|CL0006	152	0	1		C	A	A	C	C	A	T	C	A	A	G	C	C	C	C	C	G	G	C	T	T
3	M03794:7:000000000-AMUBJ:1:1103:22808:14240|SAMPLE09|CL0002	146	0	1		C	A	A	C	C	C	T	C	A	A	G	C	T	C	T	G	C	T	T	G	G
4	M03794:7:000000000-AMUBJ:1:1107:17598:22477|SAMPLE06|CL0015	155	0	1		T	C	A	A	A	C	G	T	C	A	A	G	C	C	T	G	G	C	T	T	G
5	M03794:7:000000000-AMUBJ:1:2106:4130:11881|SAMPLE21|CL0000	173	0	1		C	A	A	C	C	C	C	C	G	A	A	C	C	C	C	T	C	C	G	G	G
6	M03794:7:000000000-AMUBJ:1:1102:25872:8955|SAMPLE08|CL0006	152	0	1		C	A	A	C	C	A	T	C	A	A	G	C	C	C	C	C	G	G	C	T	T
7	M03794:7:000000000-AMUBJ:1:1115:18429:18918|SAMPLE21|CL0012	163	0	1		A	A	T	C	T	C	A	A	C	C	A	T	T	A	G	G	G	T	T	T	C
8	M03794:7:000000000-AMUBJ:1:1119:14394:18102|SAMPLE20|CL0046	215	0	1		A	A	T	T	A	T	C	A	A	C	C	T	T	G	C	T	C	G	C	T	T
9	M03794:7:000000000-AMUBJ:1:2102:6355:17551|SAMPLE16|CL0001	226	0	1		A	A	T	T	C	T	C	A	A	A	C	C	T	T	C	T	T	G	G	T	T
10	M03794:7:000000000-AMUBJ:1:1103:11944:13730|SAMPLE07|CL0015	155	0	1		T	C	A	A	A	C	G	T	C	A	A	G	C	C	T	G	G	C	T	T	G
11	M03794:7:000000000-AMUBJ:1:1119:3298:12892|SAMPLE21|CL0007	146	0	1		C	A	A	C	C	C	T	C	A	A	G	C	T	C	A	G	C	T	T	G	G
12	M03794:7:000000000-AMUBJ:1:2105:7207:20306|SAMPLE14|CL0005	180	0	1		C	A	A	A	C	C	C	T	C	T	A	T	T	C	A	A	C	T	T	T	T
13	M03794:7:000000000-AMUBJ:1:1111:21149:9849|SAMPLE25|CL0012	163	0	1		A	A	T	C	T	C	A	A	C	C	A	T	T	A	G	G	G	T	T	T	C
14	M03794:7:000000000-AMUBJ:1:1110:13260:24297|SAMPLE09|CL0004	165	0	1		C	A	A	C	C	C	T	C	A	T	G	C	C	C	C	T	A	G	G	G	C
15	M03794:7:000000000-AMUBJ:1:1109:26989:7596|SAMPLE23|CL0010	172	0	1		A	C	A	C	C	C	C	A	C	A	T	T	C	A	A	A	T	C	T	T	T
16	M03794:7:000000000-AMUBJ:1:1113:26052:6848|SAMPLE20|CL0046	214	0	1		A	A	T	T	A	T	C	A	A	C	C	T	T	G	C	T	C	G	C	T	T
17	M03794:7:000000000-AMUBJ:1:2113:23087:10163|SAMPLE11|CL0137	154	0	1		G	T	A	C	C	T	T	C	A	A	G	C	T	C	T	G	C	T	T	G	G
18	M03794:7:000000000-AMUBJ:1:1116:17642:19221|SAMPLE07|CL0024	162	0	1		C	A	C	C	C	C	T	C	A	A	G	C	C	T	T	G	C	T	T	G	G

Visible region starts at 1 bp and ends at 20 bp

This file will be used for OUT table construction…





Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

You are here

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

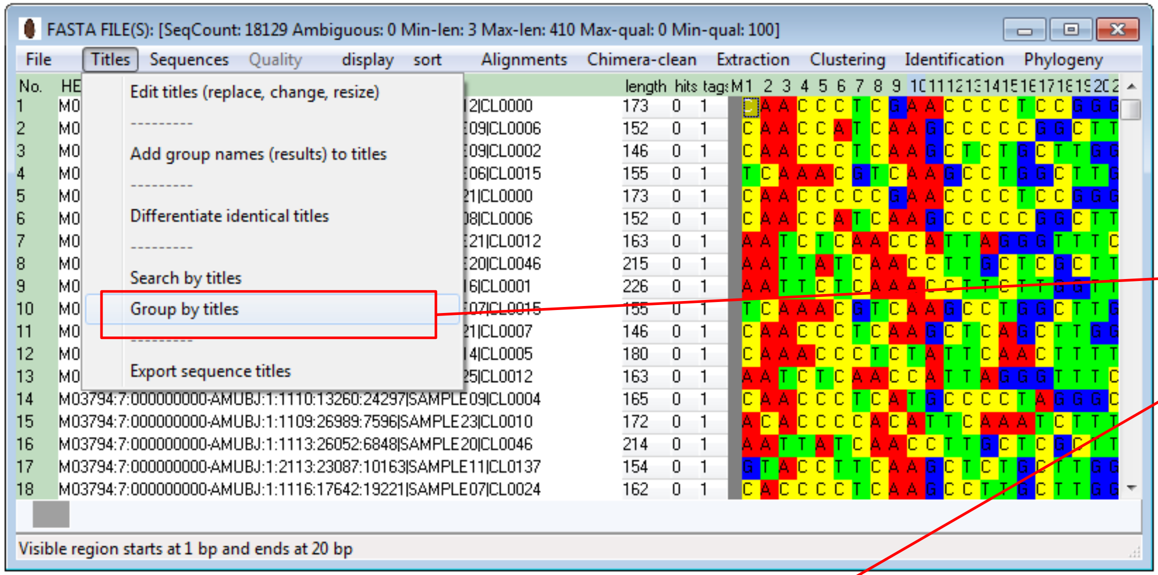
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool

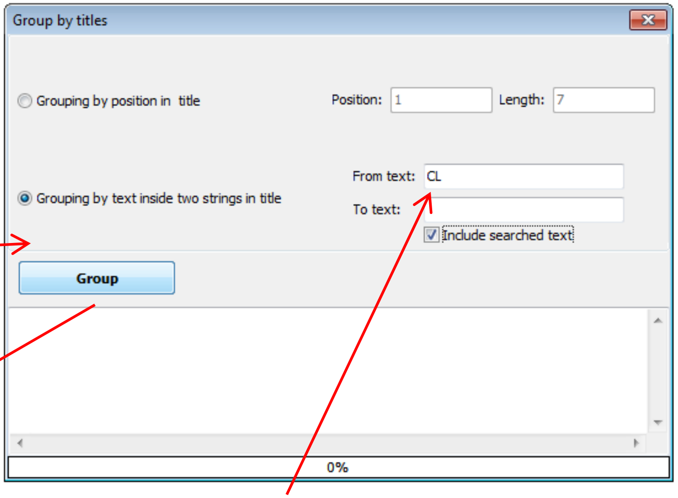
Get clusters (OTUs) representative sequences…



FASTA FILE(S): [SeqCount: 18129 Ambiguous: 0 Min-len: 3 Max-len: 410 Max-qual: 0 Min-qual: 100]

No.	HE		length	hits	tags	M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	2
1	M0	2|CL0000	173	0	1		C	A	A	C	C	C	T	C	G	A	A	C	C	C	C	T	C	C	G	G	G
2	M0	09|CL0006	152	0	1		C	A	A	C	C	A	T	C	A	A	G	C	C	C	C	C	G	G	C	T	T
3	M0	09|CL0002	146	0	1		C	A	A	C	C	C	T	C	A	A	G	C	T	C	T	G	C	T	T	G	G
4	M0	06|CL0015	155	0	1		T	C	A	A	A	C	G	T	C	A	A	G	C	C	T	G	G	C	T	T	G
5	M0	21|CL0000	173	0	1		C	A	A	C	C	C	C	C	G	A	A	C	C	C	C	T	C	C	G	G	G
6	M0	08|CL0006	152	0	1		C	A	A	C	C	A	T	C	A	A	G	C	C	C	C	C	G	G	C	T	T
7	M0	21|CL0012	163	0	1		A	A	T	C	T	C	A	A	C	C	A	T	T	A	G	G	G	T	T	T	C
8	M0	20|CL0046	215	0	1		A	A	T	T	A	T	C	A	A	C	C	T	T	G	C	T	C	G	C	T	T
9	M0	6|CL0001	226	0	1		A	A	T	T	C	T	C	A	A	A	C	C	T	T	C	T	T	G	G	T	T
10	M0	07|CL0015	155	0	1		T	C	A	A	A	C	G	T	C	A	A	G	C	C	T	G	G	C	T	T	G
11	M0	21|CL0007	146	0	1		C	A	A	C	C	C	T	C	A	A	G	C	T	C	A	G	C	T	T	G	G
12	M0	4|CL0005	180	0	1		C	A	A	A	C	C	C	T	C	T	A	T	T	C	A	A	C	T	T	T	T
13	M0	25|CL0012	163	0	1		A	A	T	C	T	C	A	A	C	C	A	T	T	A	G	G	G	T	T	T	C
14	M03794:7:000000000-AMUBJ:1:1110:13260:24297|SAMPLE09|CL0004		165	0	1		C	A	A	C	C	C	T	C	A	T	G	C	C	C	C	T	A	G	G	G	C
15	M03794:7:000000000-AMUBJ:1:1109:26989:7596|SAMPLE23|CL0010		172	0	1		A	C	A	C	C	C	C	A	C	A	T	T	C	A	A	A	T	C	T	T	T
16	M03794:7:000000000-AMUBJ:1:1113:26052:6848|SAMPLE20|CL0046		214	0	1		A	A	T	T	A	T	C	A	A	C	C	T	T	G	C	T	C	G	C	T	T
17	M03794:7:000000000-AMUBJ:1:2113:23087:10163|SAMPLE11|CL0137		154	0	1		G	T	A	C	C	T	T	C	A	A	G	C	T	C	T	G	C	T	T	G	G
18	M03794:7:000000000-AMUBJ:1:1116:17642:19221|SAMPLE07|CL0024		162	0	1		C	A	C	C	C	C	T	C	A	A	G	C	C	T	T	G	C	T	T	G	G

Visible region starts at 1 bp and ends at 20 bp

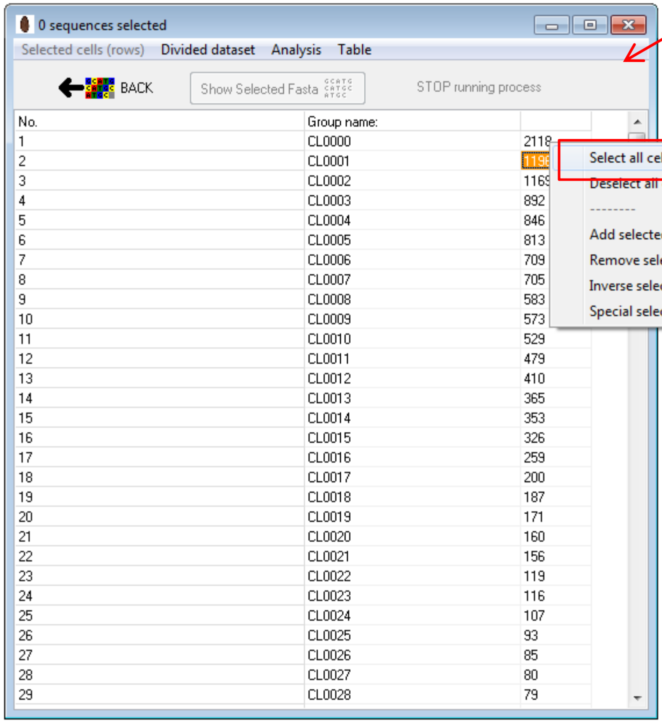


○ Grouping by position in title — Position: 1 Length: 7

● Grouping by text inside two strings in title — From text: CL To text:

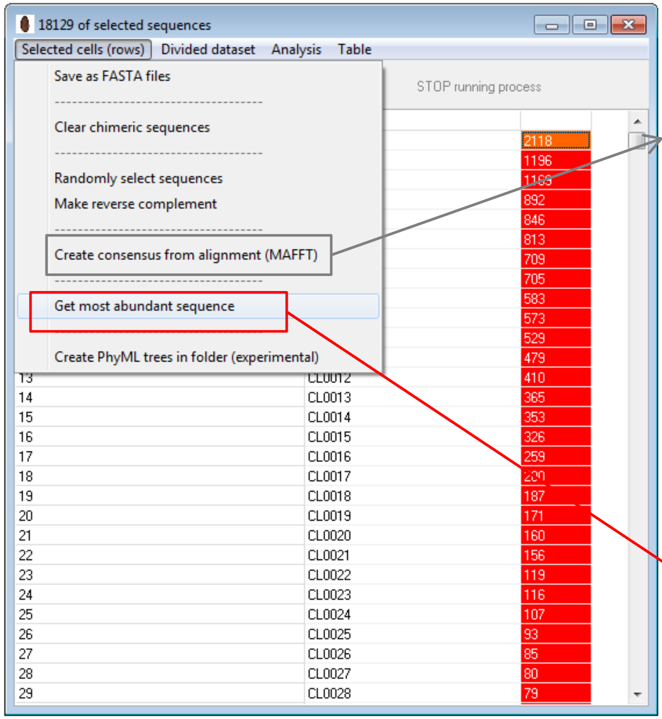
☑ Include searched text

…be careful to choose the unique identifier of desired text motive…(e.g.: "|CL")



0 sequences selected

No.	Group name:	
1	CL0000	2118
2	CL0001	1196
3	CL0002	1169
4	CL0003	892
5	CL0004	846
6	CL0005	813
7	CL0006	709
8	CL0007	705
9	CL0008	583
10	CL0009	573
11	CL0010	529
12	CL0011	479
13	CL0012	410
14	CL0013	365
15	CL0014	353
16	CL0015	326
17	CL0016	259
18	CL0017	200
19	CL0018	187
20	CL0019	171
21	CL0020	160
22	CL0021	156
23	CL0022	119
24	CL0023	116
25	CL0024	107
26	CL0025	93
27	CL0026	85
28	CL0027	80
29	CL0028	79

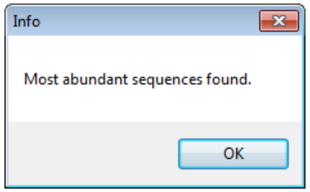


18129 of selected sequences

		2118
		1196
		1169
		892
		846
		813
		709
		705
		583
		573
		529
		479
13	CL0012	410
14	CL0013	365
15	CL0014	353
16	CL0015	326
17	CL0016	259
18	CL0017	200
19	CL0018	187
20	CL0019	171
21	CL0020	160
22	CL0021	156
23	CL0022	119
24	CL0023	116
25	CL0024	107
26	CL0025	93
27	CL0026	85
28	CL0027	80
29	CL0028	79

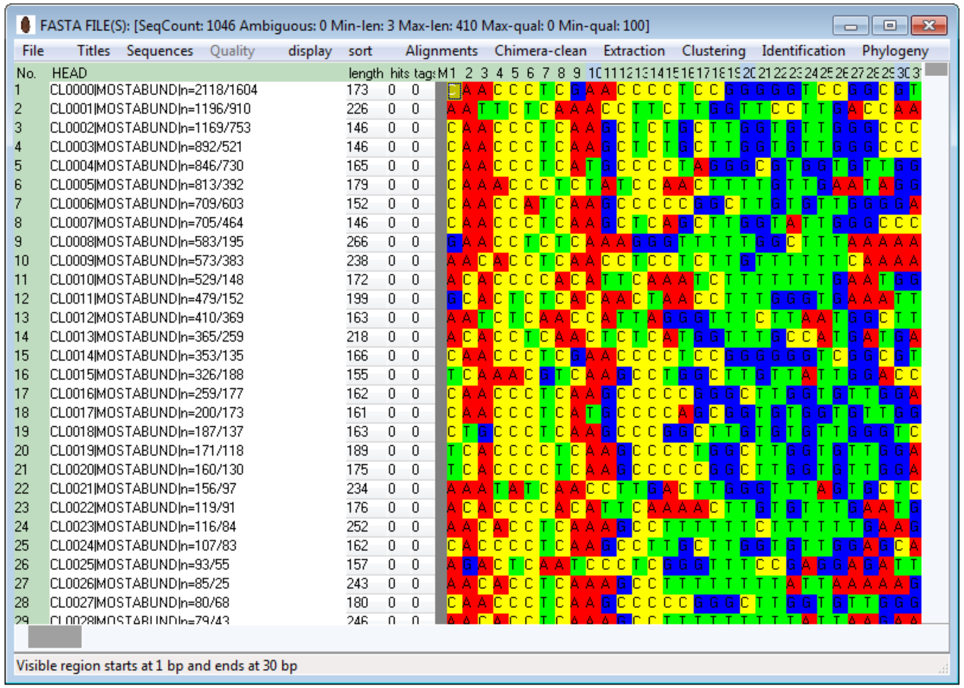
(alternative) compute a consensus from aligned sequences using „mafft„ aligner (it may take a long while)

get the most abundant sequence from each cluster (fast)



Most abundant sequences found.

OUTs representative sequences - the most abundant sequences…


FASTA FILE(S): [SeqCount: 1046 Ambiguous: 0 Min-len: 3 Max-len: 410 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tags	M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	CL0000|MOSTABUND|n=2118/1604	173	0	0		C	A	A	C	C	C	T	C	G	A	A	C	C	C	C	T	C	C	G	G	G	G	G	T	C	C	G	G	C	G	T
2	CL0001|MOSTABUND|n=1196/910	226	0	0		A	A	T	T	C	T	C	A	A	A	C	C	T	T	C	T	T	G	G	T	T	C	C	T	T	G	A	C	C	A	A
3	CL0002|MOSTABUND|n=1169/753	146	0	0		C	A	A	C	C	C	T	C	A	A	G	C	T	C	T	G	C	T	T	G	G	T	G	T	T	G	G	G	C	C	C
4	CL0003|MOSTABUND|n=892/521	146	0	0		C	A	A	C	C	C	T	C	A	A	G	C	T	C	T	G	C	T	T	G	G	T	G	T	T	G	G	G	C	C	C
5	CL0004|MOSTABUND|n=846/730	165	0	0		C	A	A	C	C	C	T	C	A	T	G	C	C	C	C	T	A	G	G	G	C	G	T	G	G	T	G	T	T	G	G
6	CL0005|MOSTABUND|n=813/392	179	0	0		C	A	A	A	C	C	C	T	C	T	A	T	C	C	A	A	C	T	T	T	T	G	T	T	G	A	A	T	A	G	G
7	CL0006|MOSTABUND|n=709/603	152	0	0		C	A	A	C	C	A	T	C	A	A	G	C	C	C	C	C	G	G	C	T	T	G	T	G	T	T	G	G	G	G	A
8	CL0007|MOSTABUND|n=705/464	146	0	0		C	A	A	C	C	C	T	C	A	A	G	C	T	C	A	G	C	T	T	G	G	T	A	T	T	G	G	G	C	C	C
9	CL0008|MOSTABUND|n=583/195	266	0	0		G	A	A	C	C	T	C	T	C	A	A	A	G	G	G	T	T	T	T	T	G	G	C	T	T	T	A	A	A	A	A
10	CL0009|MOSTABUND|n=573/383	238	0	0		A	A	C	A	C	C	T	C	A	A	C	C	T	C	C	T	C	T	T	G	T	T	T	T	T	T	C	A	A	A	A
11	CL0010|MOSTABUND|n=529/148	172	0	0		A	C	A	C	C	C	C	A	C	A	T	T	C	A	A	A	T	C	T	T	T	T	T	T	T	G	A	A	T	G	G
12	CL0011|MOSTABUND|n=479/152	199	0	0		G	C	A	C	T	C	T	C	A	C	A	A	C	T	A	A	C	C	T	T	T	G	G	G	T	G	A	A	A	T	T
13	CL0012|MOSTABUND|n=410/369	163	0	0		A	A	T	C	T	C	A	A	C	C	A	T	T	A	G	G	G	T	T	T	C	T	T	A	A	T	G	G	C	T	T
14	CL0013|MOSTABUND|n=365/259	218	0	0		A	C	A	C	C	T	C	A	A	C	T	C	T	C	A	T	G	G	T	T	T	G	C	C	A	T	G	A	T	G	A
15	CL0014|MOSTABUND|n=353/135	166	0	0		C	A	A	C	C	C	T	C	G	A	A	C	C	C	C	T	C	C	G	G	G	G	G	G	T	C	G	G	C	G	T
16	CL0015|MOSTABUND|n=326/188	155	0	0		T	C	A	A	A	C	G	T	C	A	A	G	C	C	T	G	G	C	T	T	G	T	T	A	T	T	G	G	A	C	C
17	CL0016|MOSTABUND|n=259/177	162	0	0		C	A	A	C	C	C	T	C	A	A	G	C	C	C	C	C	G	G	G	C	T	T	G	G	T	G	T	T	G	G	A
18	CL0017|MOSTABUND|n=200/173	161	0	0		C	A	A	C	C	C	T	C	A	T	G	C	C	C	C	A	G	C	G	G	T	G	T	G	G	T	G	T	T	G	G
19	CL0018|MOSTABUND|n=187/137	163	0	0		C	T	G	C	C	C	T	C	A	A	G	C	C	C	G	G	C	T	T	G	T	G	T	G	T	T	G	G	G	T	C
20	CL0019|MOSTABUND|n=171/118	189	0	0		T	C	A	C	C	C	C	T	C	A	A	G	C	C	C	C	T	G	G	C	T	T	G	G	T	G	T	T	G	G	A
21	CL0020|MOSTABUND|n=160/130	175	0	0		T	C	A	C	C	C	C	T	C	A	A	G	C	C	C	C	C	G	G	C	T	T	G	G	T	G	T	T	G	G	A
22	CL0021|MOSTABUND|n=156/97	234	0	0		A	A	A	T	A	T	C	A	A	C	C	T	T	G	A	C	T	T	G	G	G	T	T	T	A	G	T	G	C	T	C
23	CL0022|MOSTABUND|n=119/91	176	0	0		A	C	A	C	C	C	C	A	C	A	T	T	C	A	A	A	A	C	T	T	G	T	G	T	T	T	G	A	A	T	G
24	CL0023|MOSTABUND|n=116/84	252	0	0		A	A	C	A	C	C	T	C	A	A	A	G	C	C	T	T	T	T	T	T	C	T	T	T	T	T	T	G	A	A	G
25	CL0024|MOSTABUND|n=107/83	162	0	0		C	A	C	C	C	C	T	C	A	A	G	C	C	T	T	G	C	T	T	G	G	T	G	T	T	G	G	A	G	C	A
26	CL0025|MOSTABUND|n=93/55	157	0	0		A	G	A	C	T	C	A	A	T	C	C	C	T	C	G	G	G	T	T	T	C	C	G	A	G	G	A	G	A	T	T
27	CL0026|MOSTABUND|n=85/25	243	0	0		A	A	C	A	C	C	T	C	A	A	A	G	C	C	T	T	T	T	T	T	T	T	A	T	T	A	A	A	A	A	G
28	CL0027|MOSTABUND|n=80/68	180	0	0		C	A	A	C	C	C	T	C	A	A	G	C	C	C	C	C	G	G	G	C	T	T	G	G	T	G	T	T	G	G	G
29	CL0028|MOSTABUND|n=79/43	246	0	0		A	A	C	A	C	C	T	C	A	A	A	G	C	C	T	T	T	T	T	T	T	T	T	A	T	T	A	A	G	A	A

Visible region starts at 1 bp and ends at 30 bp

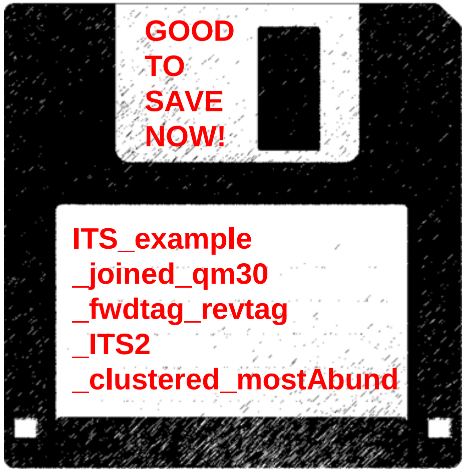
CL0000|MOSTABUND|n=2118/1604

cluster name

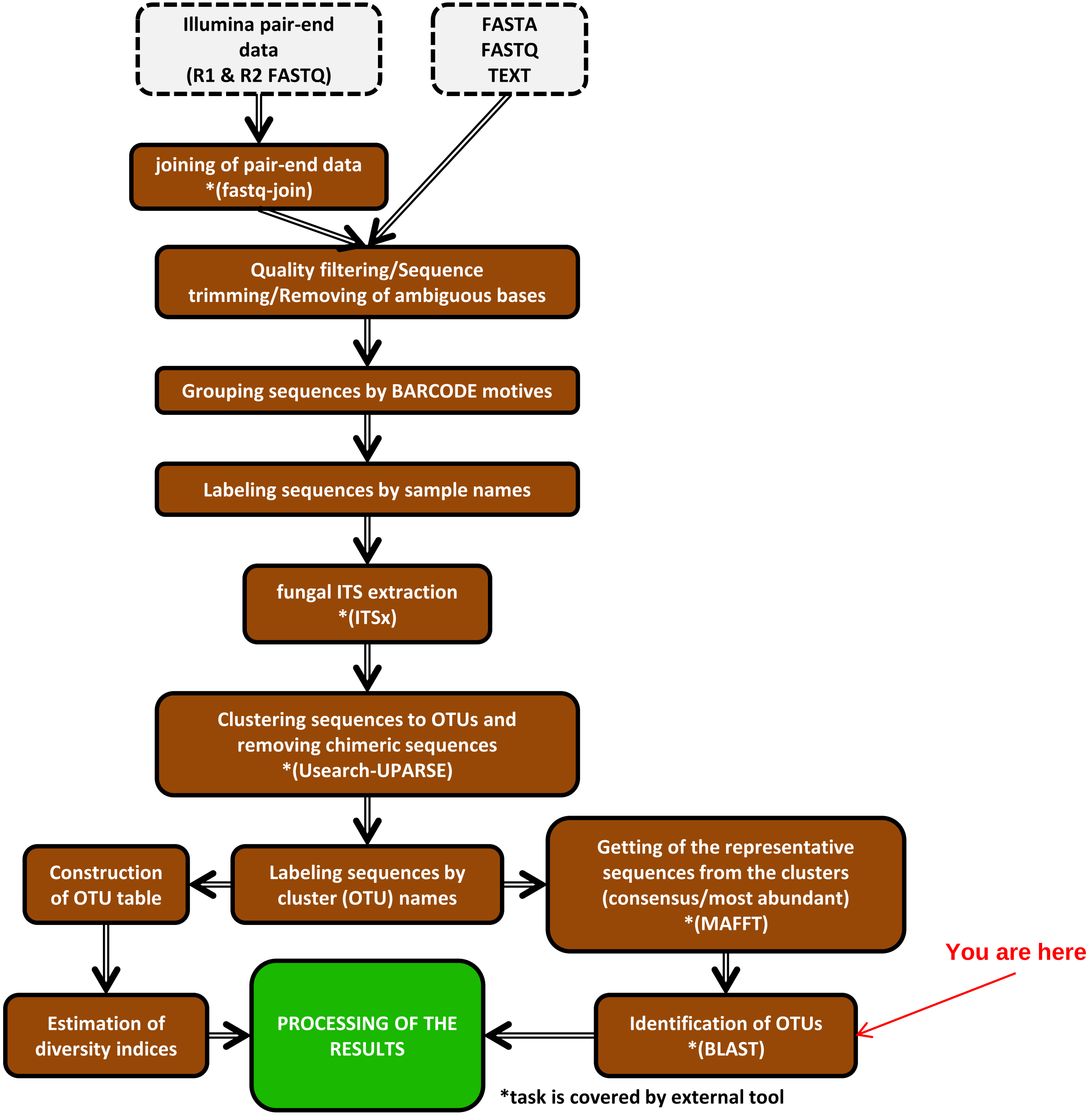
type of selection

number of sequences in group

number of most abundant identical sequences in group


GOOD TO SAVE NOW!

ITS_example_joined_qm30_fwdtag_revtag_ITS2_clustered_mostAbund



Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

Construction of OTU table

Labeling sequences by cluster (OTU) names

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

Estimation of diversity indices

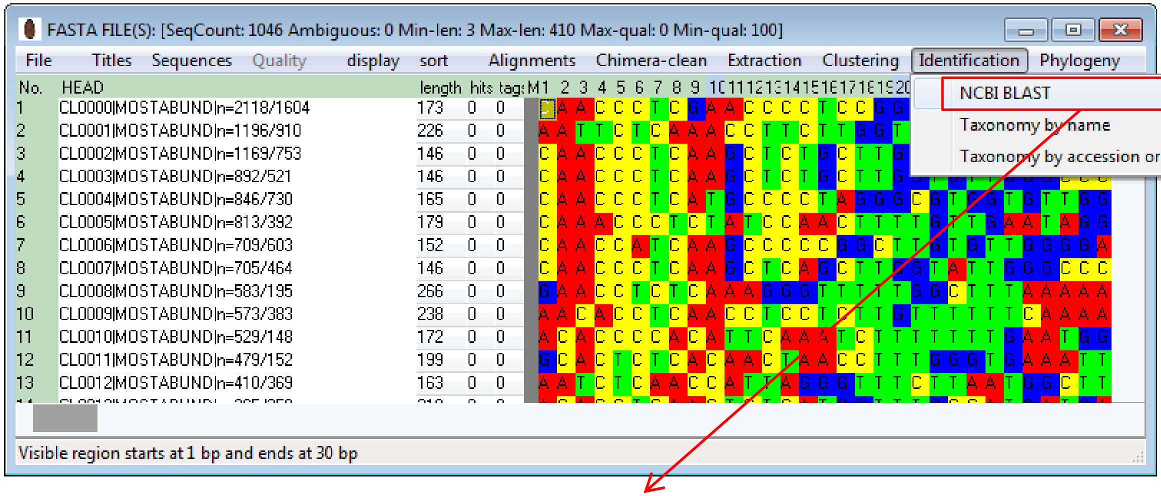
PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

You are here

*task is covered by external tool

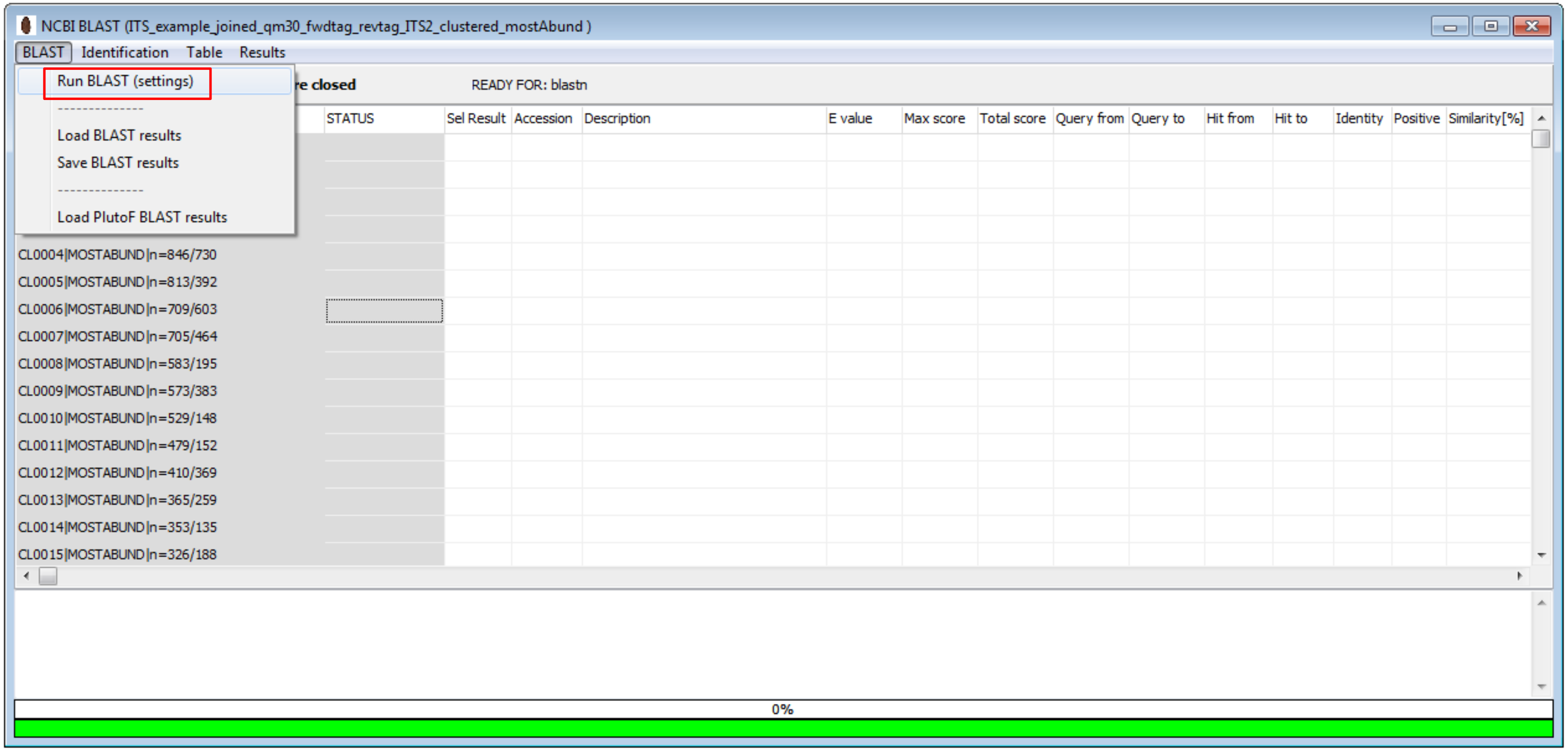
Identification of representative sequences…



FASTA FILE(S): [SeqCount: 1046 Ambiguous: 0 Min-len: 3 Max-len: 410 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tags	M1 …
1	CL0000|MOSTABUND|n=2118/1604	173	0	0	CAACCCTCGAACCCCTCCGG
2	CL0001|MOSTABUND|n=1196/910	226	0	0	AATTCTCAAACCTTCTTGGT
3	CL0002|MOSTABUND|n=1169/753	146	0	0	CAACCCTCAAGCTCTGCTTG
4	CL0003|MOSTABUND|n=892/521	146	0	0	CAACCCTCAAGCTCTGCTTG
5	CL0004|MOSTABUND|n=846/730	165	0	0	CAACCCTCATGCCCCTAGGGCGTGGTGTTGG
6	CL0005|MOSTABUND|n=813/392	179	0	0	CAAACCCTCTATCCAACTTTTGTTGAATAGG
7	CL0006|MOSTABUND|n=709/603	152	0	0	CAACCATCAAGCCCCCGGCTTGTGTTGGGGA
8	CL0007|MOSTABUND|n=705/464	146	0	0	CAACCCTCAAGCTCAGCTTGGTATTGGGCCC
9	CL0008|MOSTABUND|n=583/195	266	0	0	GAACCTCTCAAAGGGTTTTTGGCTTTAAAAA
10	CL0009|MOSTABUND|n=573/383	238	0	0	AACACCTCAACCTCCTCTTGTTTTTTCAAAA
11	CL0010|MOSTABUND|n=529/148	172	0	0	ACACCCCACATTCAAATCTTTTTTTGAATGG
12	CL0011|MOSTABUND|n=479/152	199	0	0	GCACTCTCACAACTAACCTTTGGGTGAAATT
13	CL0012|MOSTABUND|n=410/369	163	0	0	AATCTCAACCATTAGGGTTTCTTAATGGCTT

Visible region starts at 1 bp and ends at 30 bp



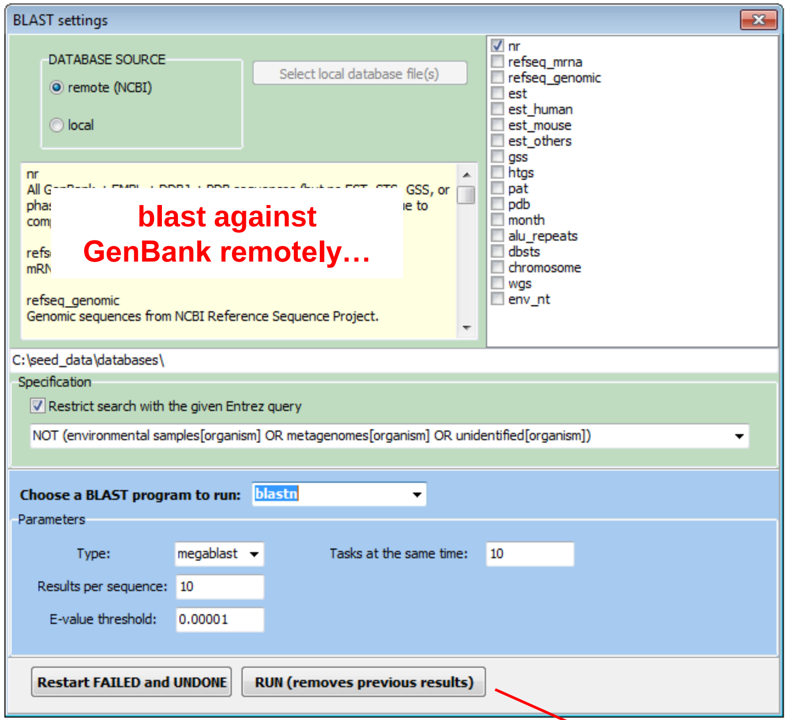
NCBI BLAST (ITS_example_joined_qm30_fwdtag_revtag_ITS2_clustered_mostAbund)

re closed

READY FOR: blastn

	STATUS	Sel Result	Accession	Description	E value	Max score	Total score	Query from	Query to	Hit from	Hit to	Identity	Positive	Similarity[%]
CL0004|MOSTABUND|n=846/730														
CL0005|MOSTABUND|n=813/392														
CL0006|MOSTABUND|n=709/603														
CL0007|MOSTABUND|n=705/464														
CL0008|MOSTABUND|n=583/195														
CL0009|MOSTABUND|n=573/383														
CL0010|MOSTABUND|n=529/148														
CL0011|MOSTABUND|n=479/152														
CL0012|MOSTABUND|n=410/369														
CL0013|MOSTABUND|n=365/259														
CL0014|MOSTABUND|n=353/135														
CL0015|MOSTABUND|n=326/188														

0%



DATABASE SOURCE
● remote (NCBI)
○ local

nr
All G… GSS, or
phas… ie to
com
refs
mRN

refseq_genomic
Genomic sequences from NCBI Reference Sequence Project.

blast against GenBank remotely…

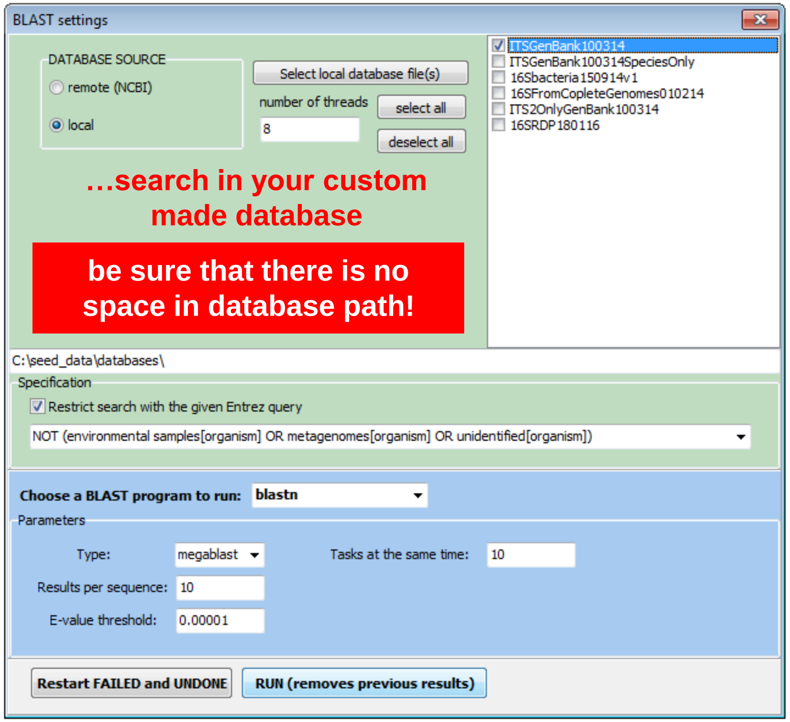
☑ nr
☐ refseq_mrna
☐ refseq_genomic
☐ est
☐ est_human
☐ est_mouse
☐ est_others
☐ gss
☐ htgs
☐ pat
☐ pdb
☐ month
☐ alu_repeats
☐ dbsts
☐ chromosome
☐ wgs
☐ env_nt

C:\seed_data\databases\

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

or



DATABASE SOURCE
○ remote (NCBI)
● local

number of threads: 8

…search in your custom made database

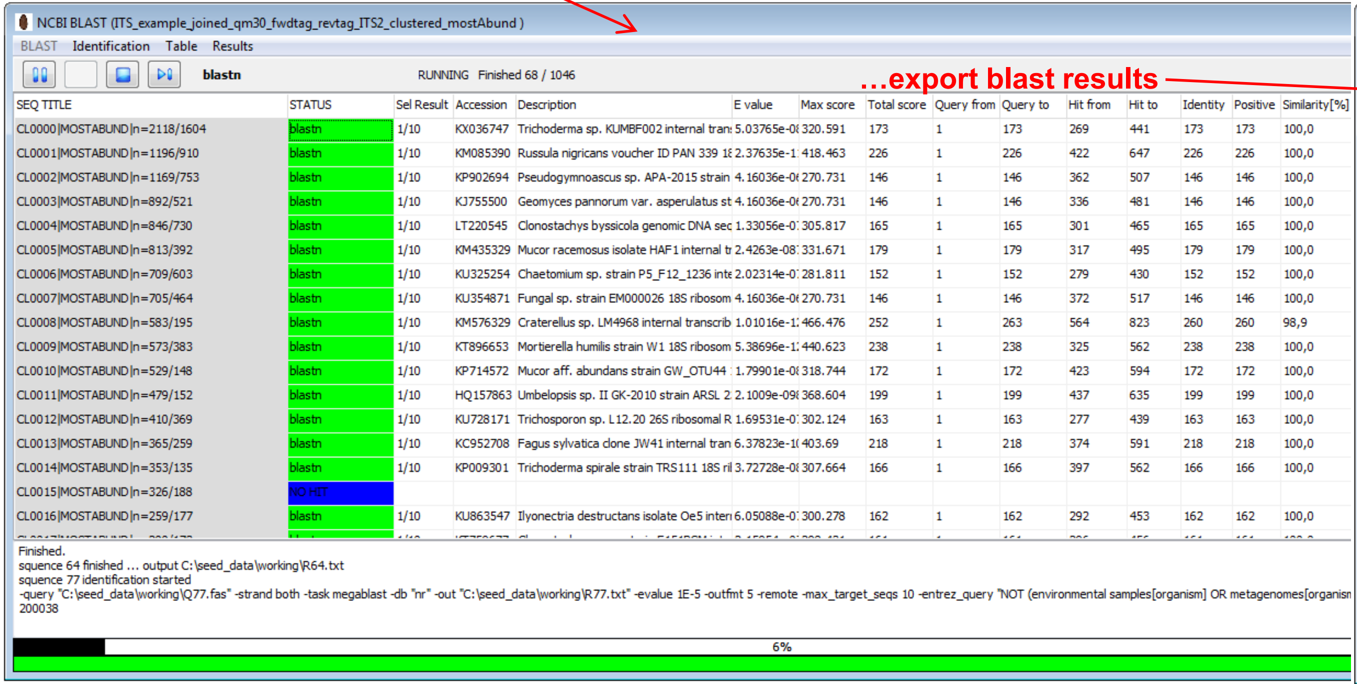
be sure that there is no space in database path!

☑ ITSGenBank100314
☐ ITSGenBank100314SpeciesOnly
☐ 16Sbacteria150914v1
☐ 16SFromCopleteGenomes010214
☐ ITS2OnlyGenBank100314
☐ 16SRDP180116

C:\seed_data\databases\

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



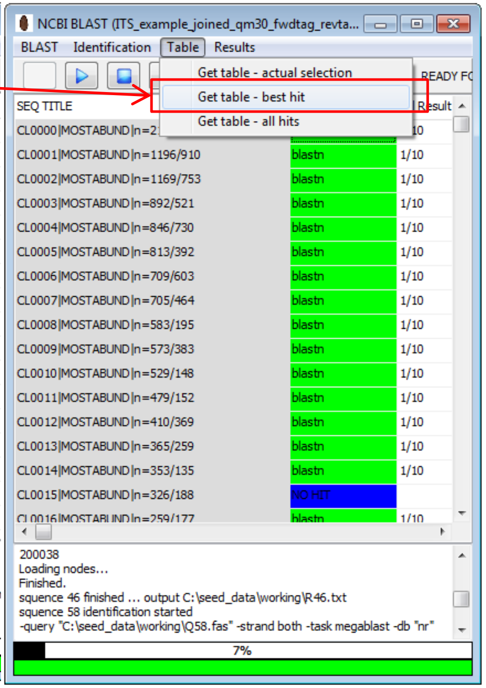
RUNNING Finished 68 / 1046

SEQ TITLE	STATUS	Sel Result	Accession	Description	E value	Max score	Total score	Query from	Query to	Hit from	Hit to	Identity	Positive	Similarity[%]
CL0000|MOSTABUND|n=2118/1604	blastn	1/10	KX036747	Trichoderma sp. KUMBF002 internal tran	5.03765e-0	320.591	173	1	173	269	441	173	173	100,0
CL0001|MOSTABUND|n=1196/910	blastn	1/10	KM085390	Russula nigricans voucher ID PAN 339 1	2.37635e-1	418.463	226	1	226	422	647	226	226	100,0
CL0002|MOSTABUND|n=1169/753	blastn	1/10	KP902694	Pseudogymnoascus sp. APA-2015 strain	4.16036e-0	270.731	146	1	146	362	507	146	146	100,0
CL0003|MOSTABUND|n=892/521	blastn	1/10	KJ755500	Geomyces pannorum var. asperulatus st	4.16036e-0	270.731	146	1	146	336	481	146	146	100,0
CL0004|MOSTABUND|n=846/730	blastn	1/10	LT220545	Clonostachys byssicola genomic DNA sec	1.33056e-0	305.817	165	1	165	301	465	165	165	100,0
CL0005|MOSTABUND|n=813/392	blastn	1/10	KM435329	Mucor racemosus isolate HAF1 internal t	2.4263e-08	331.671	179	1	179	317	495	179	179	100,0
CL0006|MOSTABUND|n=709/603	blastn	1/10	KU325254	Chaetomium sp. strain P5_F12_1236 inte	2.02314e-0	281.811	152	1	152	279	430	152	152	100,0
CL0007|MOSTABUND|n=705/464	blastn	1/10	KU354871	Fungal sp. strain EM000026 18S ribosom	4.16036e-0	270.731	146	1	146	372	517	146	146	100,0
CL0008|MOSTABUND|n=583/195	blastn	1/10	KM576329	Craterellus sp. LM4968 internal transcrib	1.01016e-1	466.476	252	1	263	564	823	260	260	98,9
CL0009|MOSTABUND|n=573/383	blastn	1/10	KT896653	Mortierella humilis strain W1 18S ribosom	5.38696e-1	440.623	238	1	238	325	562	238	238	100,0
CL0010|MOSTABUND|n=529/148	blastn	1/10	KP714572	Mucor aff. abundans strain GW_OTU44	1.79901e-0	318.744	172	1	172	423	594	172	172	100,0
CL0011|MOSTABUND|n=479/152	blastn	1/10	HQ157863	Umbelopsis sp. II GK-2010 strain ARSL 2	2.1009e-09	368.604	199	1	199	437	635	199	199	100,0
CL0012|MOSTABUND|n=410/369	blastn	1/10	KU728171	Trichosporon sp. L12.20 26S ribosomal R	1.69531e-0	302.124	163	1	163	277	439	163	163	100,0
CL0013|MOSTABUND|n=365/259	blastn	1/10	KC952708	Fagus sylvatica clone JW41 internal tran	6.37823e-1	403.69	218	1	218	374	591	218	218	100,0
CL0014|MOSTABUND|n=353/135	blastn	1/10	KP009301	Trichoderma spirale strain TRS111 18S ri	3.72728e-0	307.664	166	1	166	397	562	166	166	100,0
CL0015|MOSTABUND|n=326/188	NO HIT													
CL0016|MOSTABUND|n=259/177	blastn	1/10	KU863547	Ilyonectria destructans isolate Oe5 inter	6.05088e-0	300.278	162	1	162	292	453	162	162	100,0

```
Finished.
squence 64 finished … output C:\seed_data\working\R64.txt
squence 77 identification started
-query "C:\seed_data\working\Q77.fas" -strand both -task megablast -db "nr" -out "C:\seed_data\working\R77.txt" -evalue 1E-5 -outfmt 5 -remote -max_target_seqs 10 -entrez_query "NOT (environmental samples[organism] OR metagenomes[organism
200038
```

6%

…export blast results

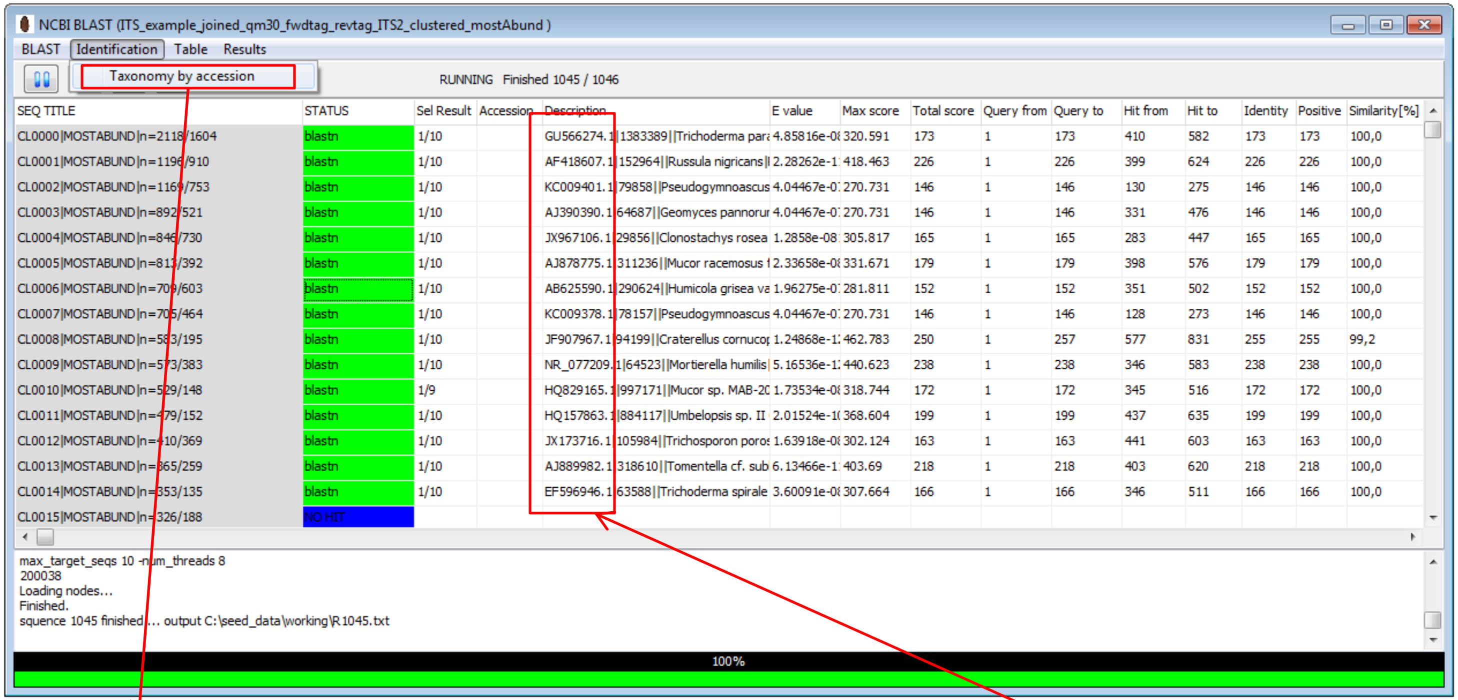


SEQ TITLE		Result
CL0000|MOSTABUND|n=2		10
CL0001|MOSTABUND|n=1196/910	blastn	1/10
CL0002|MOSTABUND|n=1169/753	blastn	1/10
CL0003|MOSTABUND|n=892/521	blastn	1/10
CL0004|MOSTABUND|n=846/730	blastn	1/10
CL0005|MOSTABUND|n=813/392	blastn	1/10
CL0006|MOSTABUND|n=709/603	blastn	1/10
CL0007|MOSTABUND|n=705/464	blastn	1/10
CL0008|MOSTABUND|n=583/195	blastn	1/10
CL0009|MOSTABUND|n=573/383	blastn	1/10
CL0010|MOSTABUND|n=529/148	blastn	1/10
CL0011|MOSTABUND|n=479/152	blastn	1/10
CL0012|MOSTABUND|n=410/369	blastn	1/10
CL0013|MOSTABUND|n=365/259	blastn	1/10
CL0014|MOSTABUND|n=353/135	blastn	1/10
CL0015|MOSTABUND|n=326/188	NO HIT	

```
200038
Loading nodes…
Finished.
squence 46 finished … output C:\seed_data\working\R46.txt
squence 58 identification started
-query "C:\seed_data\working\Q58.fas" -strand both -task megablast -db "nr"
```

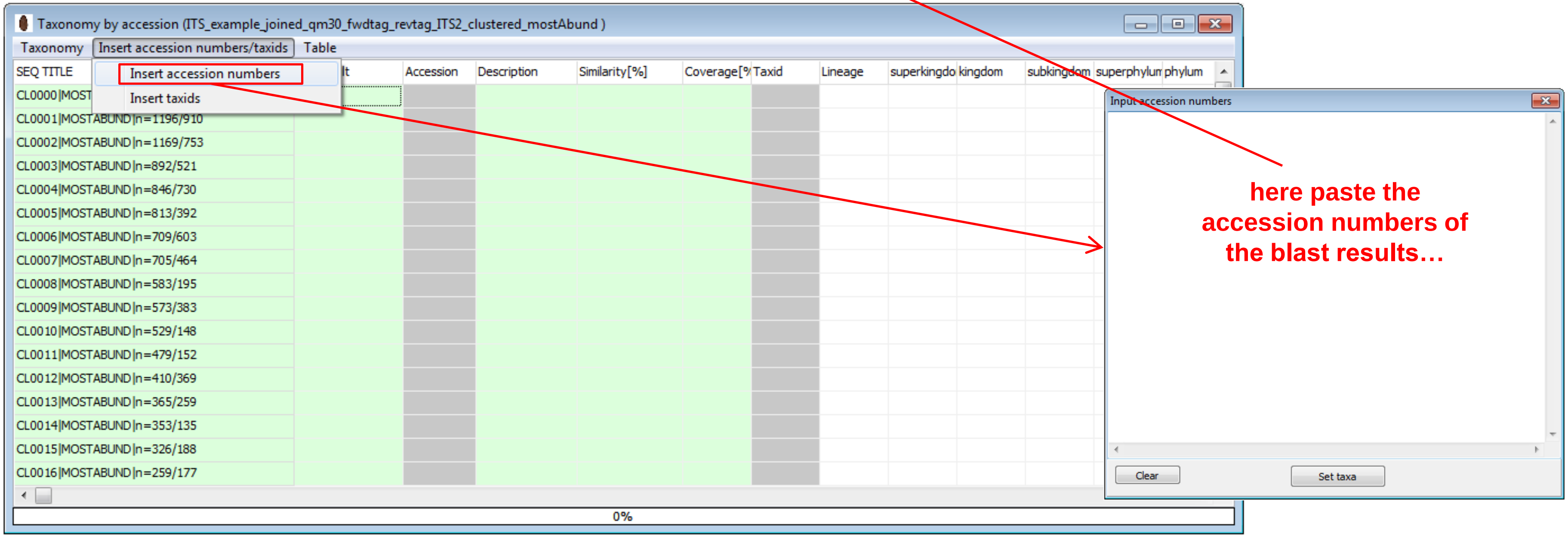
7%

Get taxonomic classification…



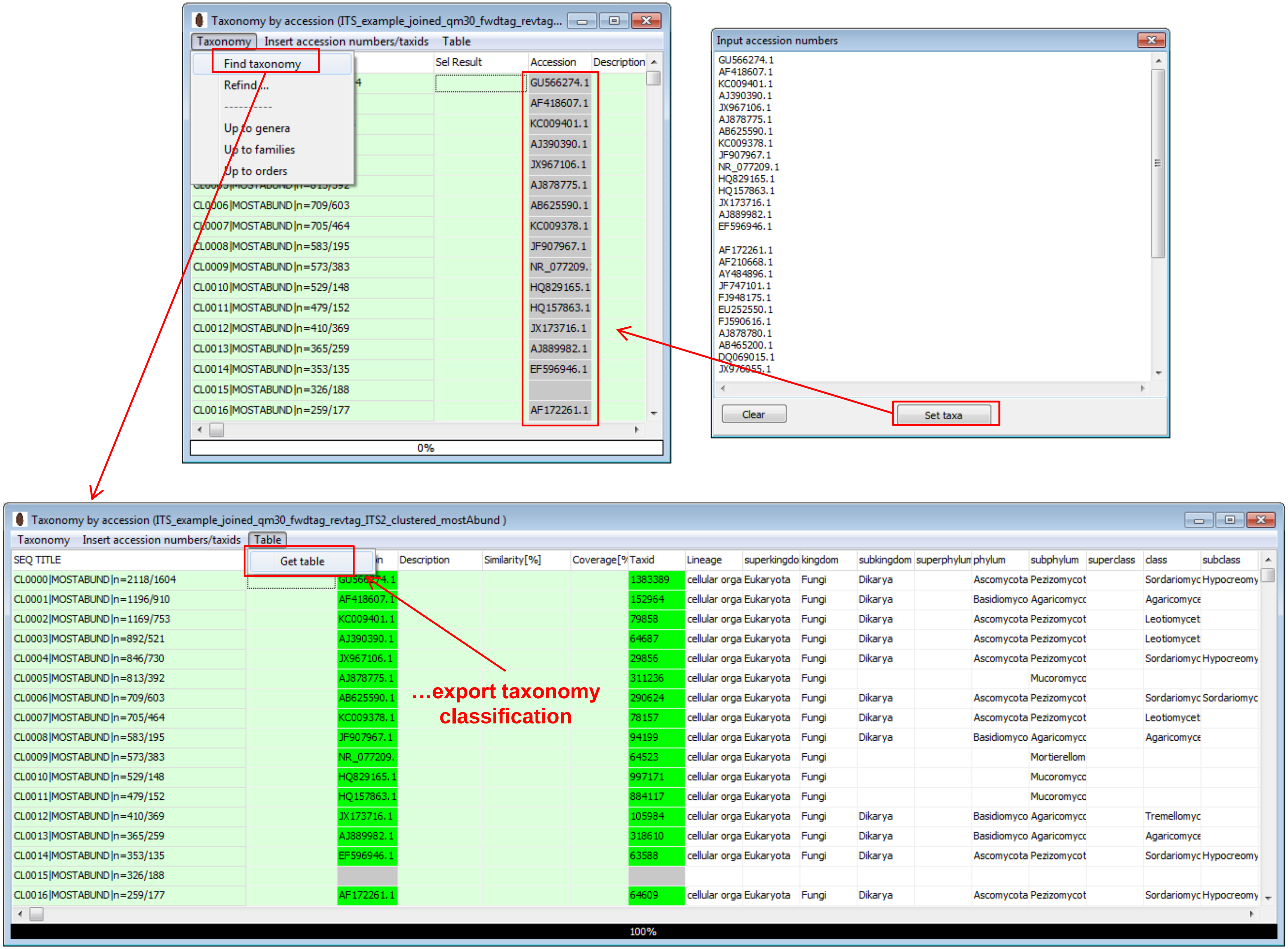
SEQ TITLE	STATUS	Sel Result	Accession	Description	E value	Max score	Total score	Query from	Query to	Hit from	Hit to	Identity	Positive	Similarity[%]
CL0000|MOSTABUND|n=2118/1604	blastn	1/10		GU566274.1|1383389||Trichoderma para	4.85816e-0	320.591	173	1	173	410	582	173	173	100,0
CL0001|MOSTABUND|n=1196/910	blastn	1/10		AF418607.1|152964||Russula nigricans|	2.28262e-1	418.463	226	1	226	399	624	226	226	100,0
CL0002|MOSTABUND|n=1169/753	blastn	1/10		KC009401.1|79858||Pseudogymnoascus	4.04467e-0	270.731	146	1	146	130	275	146	146	100,0
CL0003|MOSTABUND|n=892/521	blastn	1/10		AJ390390.1|64687||Geomyces pannorur	4.04467e-0	270.731	146	1	146	331	476	146	146	100,0
CL0004|MOSTABUND|n=846/730	blastn	1/10		JX967106.1|29856||Clonostachys rosea	1.2858e-08	305.817	165	1	165	283	447	165	165	100,0
CL0005|MOSTABUND|n=813/392	blastn	1/10		AJ878775.1|311236||Mucor racemosus f	2.33658e-0	331.671	179	1	179	398	576	179	179	100,0
CL0006|MOSTABUND|n=709/603	blastn	1/10		AB625590.1|290624||Humicola grisea va	1.96275e-0	281.811	152	1	152	351	502	152	152	100,0
CL0007|MOSTABUND|n=705/464	blastn	1/10		KC009378.1|78157||Pseudogymnoascus	4.04467e-0	270.731	146	1	146	128	273	146	146	100,0
CL0008|MOSTABUND|n=583/195	blastn	1/10		JF907967.1|94199||Craterellus cornucop	1.24868e-1	462.783	250	1	257	577	831	255	255	99,2
CL0009|MOSTABUND|n=573/383	blastn	1/10		NR_077209.1|64523||Mortierella humilis|	5.16536e-1	440.623	238	1	238	346	583	238	238	100,0
CL0010|MOSTABUND|n=529/148	blastn	1/9		HQ829165.1|997171||Mucor sp. MAB-20	1.73534e-0	318.744	172	1	172	345	516	172	172	100,0
CL0011|MOSTABUND|n=479/152	blastn	1/10		HQ157863.1|884117||Umbelopsis sp. II	2.01524e-1	368.604	199	1	199	437	635	199	199	100,0
CL0012|MOSTABUND|n=410/369	blastn	1/10		JX173716.1|105984||Trichosporon poros	1.63918e-0	302.124	163	1	163	441	603	163	163	100,0
CL0013|MOSTABUND|n=365/259	blastn	1/10		AJ889982.1|318610||Tomentella cf. sub	6.13466e-1	403.69	218	1	218	403	620	218	218	100,0
CL0014|MOSTABUND|n=353/135	blastn	1/10		EF596946.1|63588||Trichoderma spirale	3.60091e-0	307.664	166	1	166	346	511	166	166	100,0
CL0015|MOSTABUND|n=326/188	NO HIT													

```
max_target_seqs 10 -num_threads 8
200038
Loading nodes...
Finished.
squence 1045 finished... output C:\seed_data\working\R1045.txt
```



SEQ TITLE	lt	Accession	Description	Similarity[%]	Coverage[%	Taxid	Lineage	superkingdo	kingdom	subkingdom	superphylum	phylum
CL0000|MOST												
CL0001|MOSTABUND|n=1196/910												
CL0002|MOSTABUND|n=1169/753												
CL0003|MOSTABUND|n=892/521												
CL0004|MOSTABUND|n=846/730												
CL0005|MOSTABUND|n=813/392												
CL0006|MOSTABUND|n=709/603												
CL0007|MOSTABUND|n=705/464												
CL0008|MOSTABUND|n=583/195												
CL0009|MOSTABUND|n=573/383												
CL0010|MOSTABUND|n=529/148												
CL0011|MOSTABUND|n=479/152												
CL0012|MOSTABUND|n=410/369												
CL0013|MOSTABUND|n=365/259												
CL0014|MOSTABUND|n=353/135												
CL0015|MOSTABUND|n=326/188												
CL0016|MOSTABUND|n=259/177												

here paste the accession numbers of the blast results…



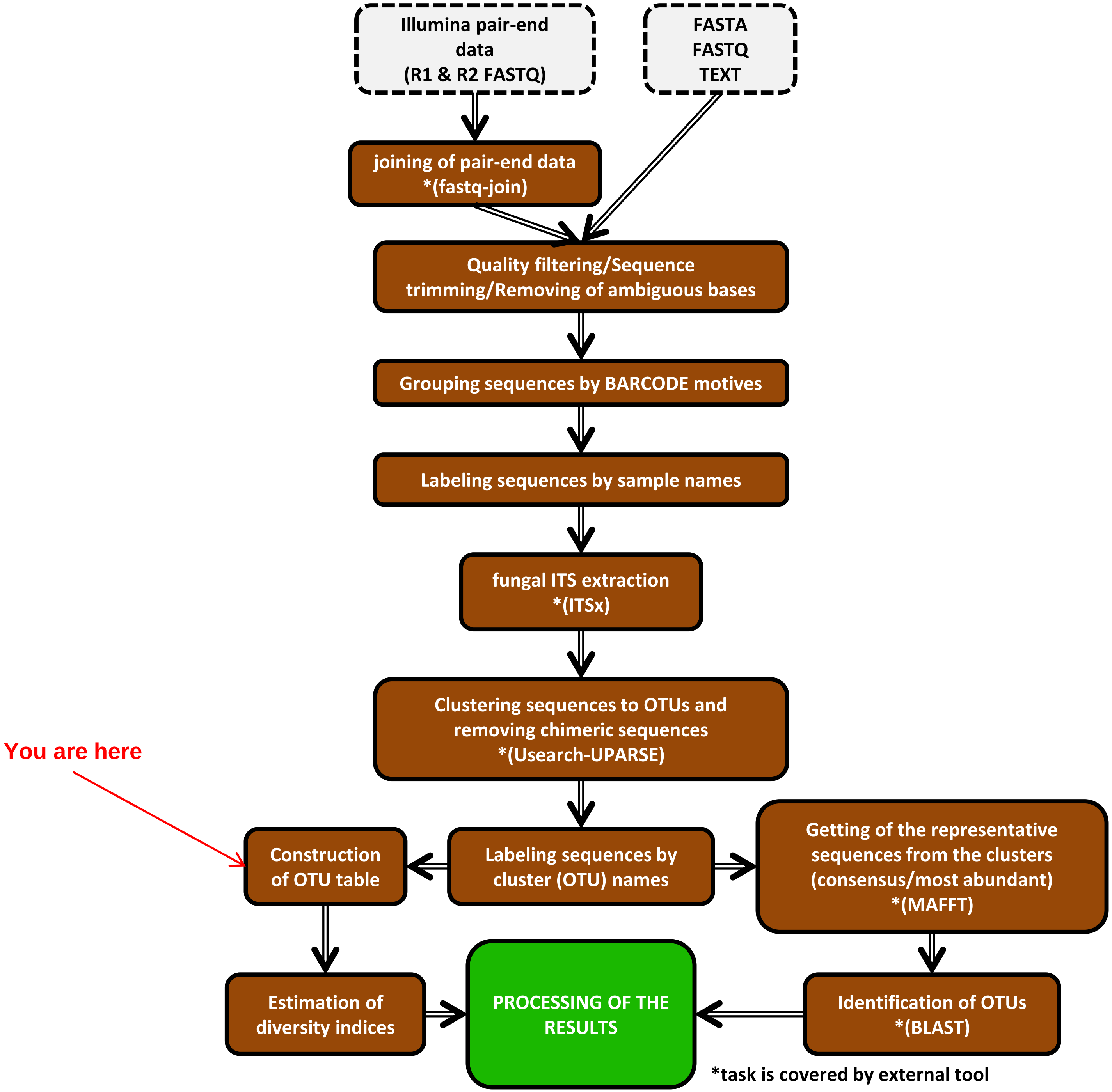
	Sel Result	Accession	Description
4		GU566274.1	
		AF418607.1	
		KC009401.1	
		AJ390390.1	
		JX967106.1	
CL0005|MOSTABUND|n=813/392		AJ878775.1	
CL0006|MOSTABUND|n=709/603		AB625590.1	
CL0007|MOSTABUND|n=705/464		KC009378.1	
CL0008|MOSTABUND|n=583/195		JF907967.1	
CL0009|MOSTABUND|n=573/383		NR_077209.	
CL0010|MOSTABUND|n=529/148		HQ829165.1	
CL0011|MOSTABUND|n=479/152		HQ157863.1	
CL0012|MOSTABUND|n=410/369		JX173716.1	
CL0013|MOSTABUND|n=365/259		AJ889982.1	
CL0014|MOSTABUND|n=353/135		EF596946.1	
CL0015|MOSTABUND|n=326/188			
CL0016|MOSTABUND|n=259/177		AF172261.1	

GU566274.1
AF418607.1
KC009401.1
AJ390390.1
JX967106.1
AJ878775.1
AB625590.1
KC009378.1
JF907967.1
NR_077209.1
HQ829165.1
HQ157863.1
JX173716.1
AJ889982.1
EF596946.1

AF172261.1
AF210668.1
AY484896.1
JF747101.1
FJ948175.1
EU252550.1
FJ590616.1
AJ878780.1
AB465200.1
DQ069015.1
JX976055.1

SEQ TITLE		n	Description	Similarity[%]	Coverage[%	Taxid	Lineage	superkingdo	kingdom	subkingdom	superphylum	phylum	subphylum	superclass	class	subclass
CL0000|MOSTABUND|n=2118/1604		GU566274.1				1383389	cellular orga	Eukaryota	Fungi	Dikarya		Ascomycota	Pezizomycot		Sordariomyc	Hypocreomy
CL0001|MOSTABUND|n=1196/910		AF418607.1				152964	cellular orga	Eukaryota	Fungi	Dikarya		Basidiomyco	Agaricomyc		Agaricomyce	
CL0002|MOSTABUND|n=1169/753		KC009401.1				79858	cellular orga	Eukaryota	Fungi	Dikarya		Ascomycota	Pezizomycot		Leotiomycet	
CL0003|MOSTABUND|n=892/521		AJ390390.1				64687	cellular orga	Eukaryota	Fungi	Dikarya		Ascomycota	Pezizomycot		Leotiomycet	
CL0004|MOSTABUND|n=846/730		JX967106.1				29856	cellular orga	Eukaryota	Fungi	Dikarya		Ascomycota	Pezizomycot		Sordariomyc	Hypocreomy
CL0005|MOSTABUND|n=813/392		AJ878775.1				311236	cellular orga	Eukaryota	Fungi				Mucoromyc			
CL0006|MOSTABUND|n=709/603		AB625590.1				290624	cellular orga	Eukaryota	Fungi	Dikarya		Ascomycota	Pezizomycot		Sordariomyc	Sordariomyc
CL0007|MOSTABUND|n=705/464		KC009378.1				78157	cellular orga	Eukaryota	Fungi	Dikarya		Ascomycota	Pezizomycot		Leotiomycet	
CL0008|MOSTABUND|n=583/195		JF907967.1				94199	cellular orga	Eukaryota	Fungi	Dikarya		Basidiomyco	Agaricomyc		Agaricomyce	
CL0009|MOSTABUND|n=573/383		NR_077209.				64523	cellular orga	Eukaryota	Fungi				Mortierellom			
CL0010|MOSTABUND|n=529/148		HQ829165.1				997171	cellular orga	Eukaryota	Fungi				Mucoromyc			
CL0011|MOSTABUND|n=479/152		HQ157863.1				884117	cellular orga	Eukaryota	Fungi				Mucoromyc			
CL0012|MOSTABUND|n=410/369		JX173716.1				105984	cellular orga	Eukaryota	Fungi	Dikarya		Basidiomyco	Agaricomyc		Tremellomyc	
CL0013|MOSTABUND|n=365/259		AJ889982.1				318610	cellular orga	Eukaryota	Fungi	Dikarya		Basidiomyco	Agaricomyc		Agaricomyce	
CL0014|MOSTABUND|n=353/135		EF596946.1				63588	cellular orga	Eukaryota	Fungi	Dikarya		Ascomycota	Pezizomycot		Sordariomyc	Hypocreomy
CL0015|MOSTABUND|n=326/188																
CL0016|MOSTABUND|n=259/177		AF172261.1				64609	cellular orga	Eukaryota	Fungi	Dikarya		Ascomycota	Pezizomycot		Sordariomyc	Hypocreomy

…export taxonomy classification



Illumina pair-end data (R1 & R2 FASTQ)

FASTA
FASTQ
TEXT

joining of pair-end data *(fastq-join)

Quality filtering/Sequence trimming/Removing of ambiguous bases

Grouping sequences by BARCODE motives

Labeling sequences by sample names

fungal ITS extraction *(ITSx)

Clustering sequences to OTUs and removing chimeric sequences *(Usearch-UPARSE)

You are here

Construction of OTU table

Labeling sequences by cluster (OTU) names

Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

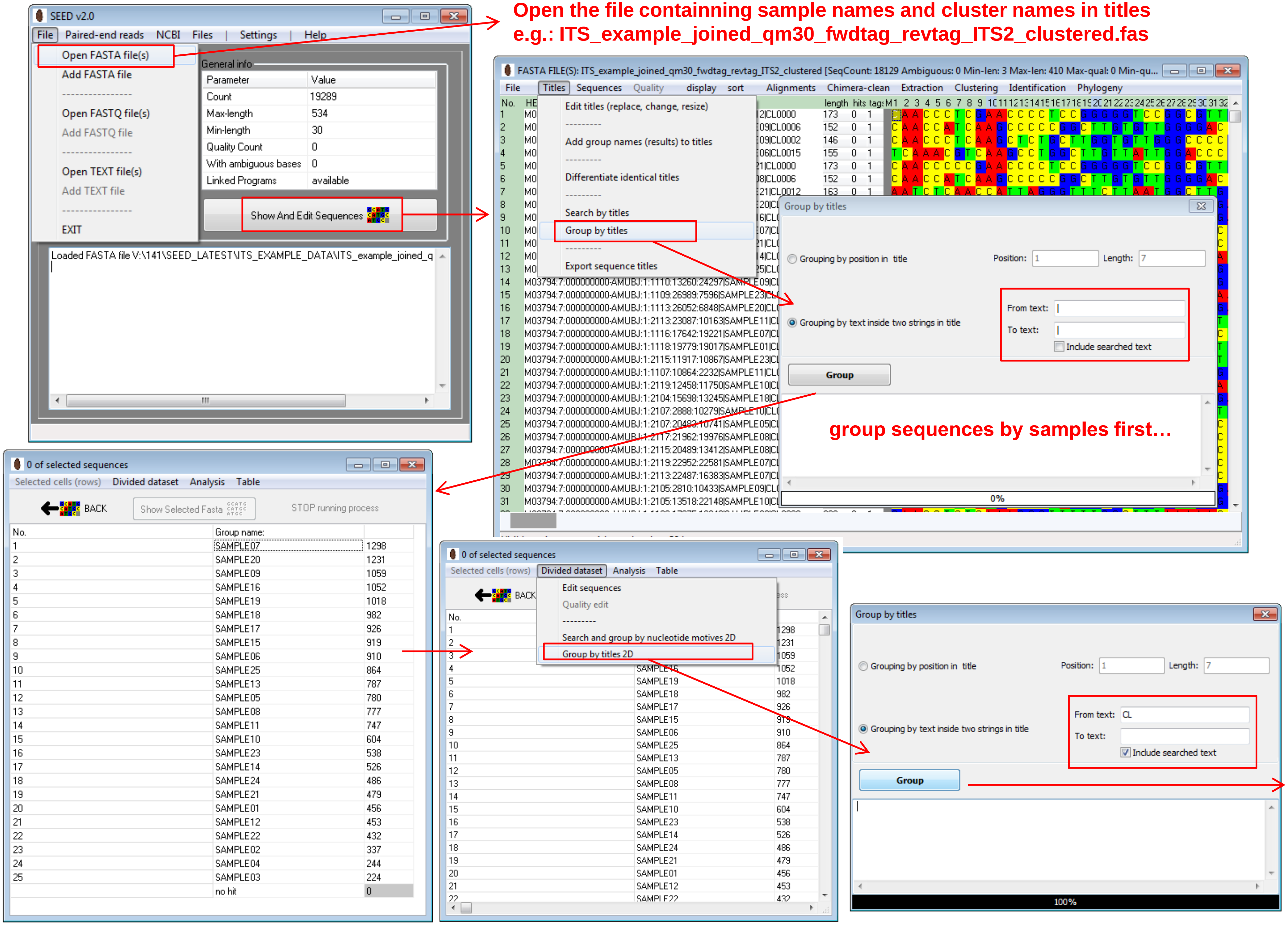
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool

OTU table construction…



Parameter	Value
Count	19289
Max-length	534
Min-length	30
Quality Count	0
With ambiguous bases	0
Linked Programs	available

Loaded FASTA file V:\141\SEED_LATEST\ITS_EXAMPLE_DATA\ITS_example_joined_q

Open the file containning sample names and cluster names in titles
e.g.: ITS_example_joined_qm30_fwdtag_revtag_ITS2_clustered.fas

FASTA FILE(S): ITS_example_joined_qm30_fwdtag_revtag_ITS2_clustered [SeqCount: 18129 Ambiguous: 0 Min-len: 3 Max-len: 410 Max-qual: 0 Min-qu...

```
No. HE                                   length hits tags
1   M0          2|CL0000                 173    0    1
2   M0          09|CL0006                152    0    1
3   M0          09|CL0002                146    0    1
4   M0          06|CL0015                155    0    1
5   M0          21|CL0000                173    0    1
6   M0          08|CL0006                152    0    1
7   M0          21|CL0012                163    0    1
8   M0          20|CL
9   M0          16|CL
10  M0          07|CL
11  M0          21|CL
12  M0          14|CL
13  M0          25|CL
14  M03794:7:000000000-AMUBJ:1:1110:13260:24297|SAMPLE09|CL
15  M03794:7:000000000-AMUBJ:1:1109:26989:7596|SAMPLE23|CL
16  M03794:7:000000000-AMUBJ:1:1113:26052:6848|SAMPLE20|CL
17  M03794:7:000000000-AMUBJ:1:2113:23087:10163|SAMPLE11|CL
18  M03794:7:000000000-AMUBJ:1:1116:17642:19221|SAMPLE07|CL
19  M03794:7:000000000-AMUBJ:1:1118:19779:19017|SAMPLE01|CL
20  M03794:7:000000000-AMUBJ:1:2115:11917:10867|SAMPLE23|CL
21  M03794:7:000000000-AMUBJ:1:1107:10864:2232|SAMPLE11|CL
22  M03794:7:000000000-AMUBJ:1:2119:12458:11750|SAMPLE10|CL
23  M03794:7:000000000-AMUBJ:1:2104:15698:13245|SAMPLE18|CL
24  M03794:7:000000000-AMUBJ:1:2107:2888:10279|SAMPLE10|CL
25  M03794:7:000000000-AMUBJ:1:2107:20483:10741|SAMPLE05|CL
26  M03794:7:000000000-AMUBJ:1:2117:21962:19976|SAMPLE08|CL
27  M03794:7:000000000-AMUBJ:1:2115:20489:13412|SAMPLE08|CL
28  M03794:7:000000000-AMUBJ:1:2119:22952:22581|SAMPLE07|CL
29  M03794:7:000000000-AMUBJ:1:2113:22487:16383|SAMPLE07|CL
30  M03794:7:000000000-AMUBJ:1:2105:2810:10433|SAMPLE09|CL
31  M03794:7:000000000-AMUBJ:1:2105:13518:22148|SAMPLE10|CL
```

Group by titles

○ Grouping by position in title — Position: 1 Length: 7

● Grouping by text inside two strings in title — From text: To text: ☐ Include searched text

group sequences by samples first…

0 of selected sequences

No.	Group name:	
1	SAMPLE07	1298
2	SAMPLE20	1231
3	SAMPLE09	1059
4	SAMPLE16	1052
5	SAMPLE19	1018
6	SAMPLE18	982
7	SAMPLE17	926
8	SAMPLE15	919
9	SAMPLE06	910
10	SAMPLE25	864
11	SAMPLE13	787
12	SAMPLE05	780
13	SAMPLE08	777
14	SAMPLE11	747
15	SAMPLE10	604
16	SAMPLE23	538
17	SAMPLE14	526
18	SAMPLE24	486
19	SAMPLE21	479
20	SAMPLE01	456
21	SAMPLE12	453
22	SAMPLE22	432
23	SAMPLE02	337
24	SAMPLE04	244
25	SAMPLE03	224
	no hit	0

0 of selected sequences

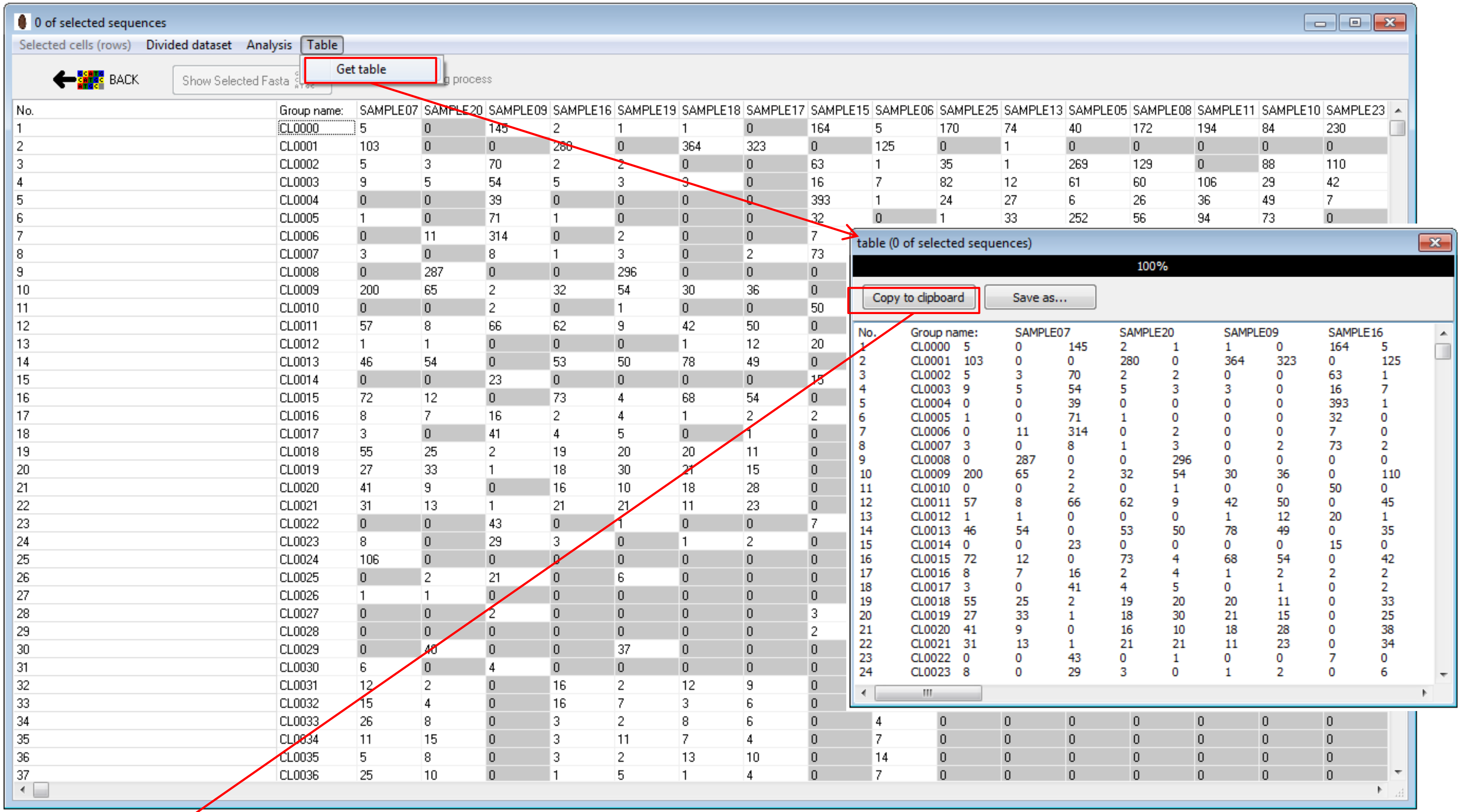
No.		
1		1298
2		1231
3		1059
4	SAMPLE16	1052
5	SAMPLE19	1018
6	SAMPLE18	982
7	SAMPLE17	926
8	SAMPLE15	919
9	SAMPLE06	910
10	SAMPLE25	864
11	SAMPLE13	787
12	SAMPLE05	780
13	SAMPLE08	777
14	SAMPLE11	747
15	SAMPLE10	604
16	SAMPLE23	538
17	SAMPLE14	526
18	SAMPLE24	486
19	SAMPLE21	479
20	SAMPLE01	456
21	SAMPLE12	453
22	SAMPLE22	432

Group by titles

○ Grouping by position in title — Position: 1 Length: 7

● Grouping by text inside two strings in title — From text: CL To text: ☑ Include searched text

OTU table is done…



No.	Group name:	SAMPLE07	SAMPLE20	SAMPLE09	SAMPLE16	SAMPLE19	SAMPLE18	SAMPLE17	SAMPLE15	SAMPLE06	SAMPLE25	SAMPLE13	SAMPLE05	SAMPLE08	SAMPLE11	SAMPLE10	SAMPLE23
1	CL0000	5	0	145	2	1	1	0	164	5	170	74	40	172	194	84	230
2	CL0001	103	0	0	280	0	364	323	0	125	0	1	0	0	0	0	0
3	CL0002	5	3	70	2	2	0	0	63	1	35	1	269	129	0	88	110
4	CL0003	9	5	54	5	3	3	0	16	7	82	12	61	60	106	29	42
5	CL0004	0	0	39	0	0	0	0	393	1	24	27	6	26	36	49	7
6	CL0005	1	0	71	1	0	0	0	32	0	1	33	252	56	94	73	0
7	CL0006	0	11	314	0	2	0	0	7								
8	CL0007	3	0	8	1	3	0	2	73								
9	CL0008	0	287	0	0	296	0	0	0								
10	CL0009	200	65	2	32	54	30	36	0								
11	CL0010	0	0	2	0	1	0	0	50								
12	CL0011	57	8	66	62	9	42	50	0								
13	CL0012	1	1	0	0	0	1	12	20								
14	CL0013	46	54	0	53	50	78	49	0								
15	CL0014	0	0	23	0	0	0	0	15								
16	CL0015	72	12	0	73	4	68	54	0								
17	CL0016	8	7	16	2	4	1	2	2								
18	CL0017	3	0	41	4	5	0	1	0								
19	CL0018	55	25	2	19	20	20	11	0								
20	CL0019	27	33	1	18	30	21	15	0								
21	CL0020	41	9	0	16	10	18	28	0								
22	CL0021	31	13	1	21	21	11	23	0								
23	CL0022	0	0	43	0	1	0	0	7								
24	CL0023	8	0	29	3	0	1	2	0								
25	CL0024	106	0	0	0	0	0	0	0								
26	CL0025	0	2	21	0	6	0	0	0								
27	CL0026	1	1	0	0	0	0	0	0								
28	CL0027	0	0	2	0	0	0	0	3								
29	CL0028	0	0	0	0	0	0	0	2								
30	CL0029	0	40	0	0	37	0	0	0								
31	CL0030	6	0	4	0	0	0	0	0								
32	CL0031	12	2	0	16	2	12	9	0								
33	CL0032	15	4	0	16	7	3	6	0								
34	CL0033	26	8	0	3	2	8	6	0	4	0	0	0	0	0	0	0
35	CL0034	11	15	0	3	11	7	4	0	7	0	0	0	0	0	0	0
36	CL0035	5	8	0	3	2	13	10	0	14	0	0	0	0	0	0	0
37	CL0036	25	10	0	1	5	1	4	0	7	0	0	0	0	0	0	0

```
No.     Group name:     SAMPLE07        SAMPLE20        SAMPLE09        SAMPLE16
1       CL0000  5       0       145     2       1       1       0       164     5
2       CL0001  103     0       0       280     0       364     323     0       125
3       CL0002  5       3       70      2       2       0       0       63      1
4       CL0003  9       5       54      5       3       3       0       16      7
5       CL0004  0       0       39      0       0       0       0       393     1
6       CL0005  1       0       71      1       0       0       0       32      0
7       CL0006  0       11      314     0       2       0       0       7       0
8       CL0007  3       0       8       1       3       0       2       73      2
9       CL0008  0       287     0       0       296     0       0       0       0
10      CL0009  200     65      2       32      54      30      36      0       110
11      CL0010  0       0       2       0       1       0       0       50      0
12      CL0011  57      8       66      62      9       42      50      0       45
13      CL0012  1       1       0       0       0       1       12      20      1
14      CL0013  46      54      0       53      50      78      49      0       35
15      CL0014  0       0       23      0       0       0       0       15      0
16      CL0015  72      12      0       73      4       68      54      0       42
17      CL0016  8       7       16      2       4       1       2       2       2
18      CL0017  3       0       41      4       5       0       1       0       2
19      CL0018  55      25      2       19      20      20      11      0       33
20      CL0019  27      33      1       18      30      21      15      0       25
21      CL0020  41      9       0       16      10      18      28      0       38
22      CL0021  31      13      1       21      21      11      23      0       34
23      CL0022  0       0       43      0       1       0       0       7       0
24      CL0023  8       0       29      3       0       1       2       0       6
```

…paste the table to excel…

Combine the obtained information…

blast identification

Taxonomy classification

OTU table

Group name:	Organism	subphylum	SAMPLE07	SAMPLE20	SAMPLE09	SAMPLE16	SAMPLE19	SAMPLE18	SAMPLE17	SAMPLE15	SAMPLE06	SAMPLE25	SAMPLE13
CL0000	Trichoderma paraviridescens	Pezizomycotina	5	0	145	2	1	1	0	164	5	170	74
CL0001	Russula nigricans	Agaricomycotina	103	0	0	280	0	364	323	0	125	0	1
CL0002	Pseudogymnoascus pannorum	Pezizomycotina	5	3	70	2	2	0	0	63	1	35	1
CL0003	Geomyces pannorum var. asperulatus	Pezizomycotina	9	5	54	5	3	3	0	16	7	82	12
CL0004	Clonostachys rosea	Pezizomycotina	0	0	39	0	0	0	0	393	1	24	27
CL0005	Mucor racemosus f. sphaerosporus	Mucoromycotina	1	0	71	1	0	0	0	32	0	1	33
CL0006	Humicola grisea var. grisea	Pezizomycotina	0	11	314	0	2	0	0	7	0	8	5
CL0007	Pseudogymnoascus roseus	Pezizomycotina	3	0	8	1	3	0	2	73	2	112	37
CL0008	Craterellus cornucopioides	Agaricomycotina	0	287	0	0	296	0	0	0	0	0	0
CL0009	Mortierella humilis	Mortierellomycotina	200	65	2	32	54	30	36	0	110	10	4
CL0010	Mucor sp. MAB-2010h	Mucoromycotina	0	0	2	0	1	0	0	50	0	57	194

+ additional metadata

Process the results…