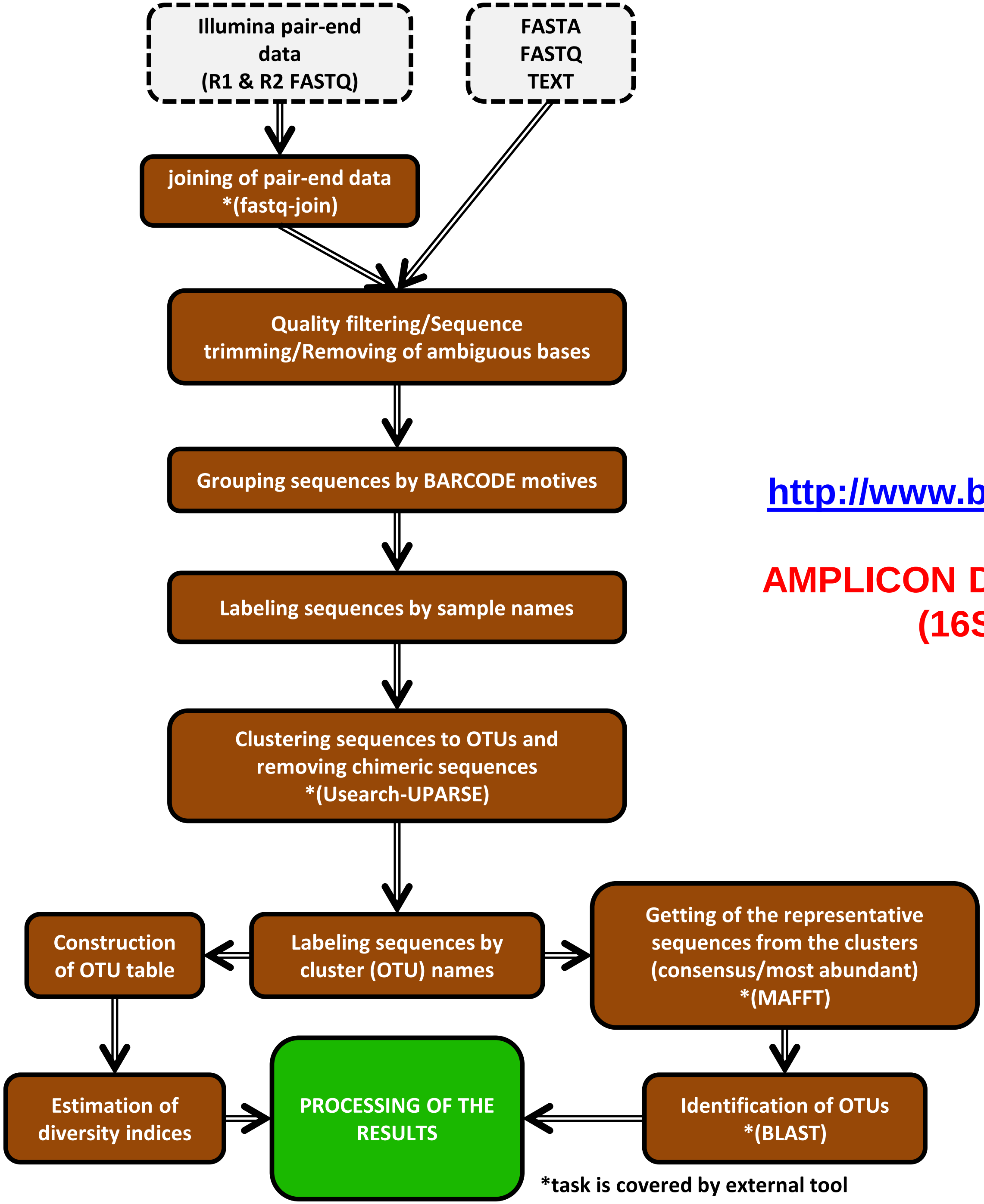


Illumina pair-end data (R1 & R2 FASTQ)

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

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

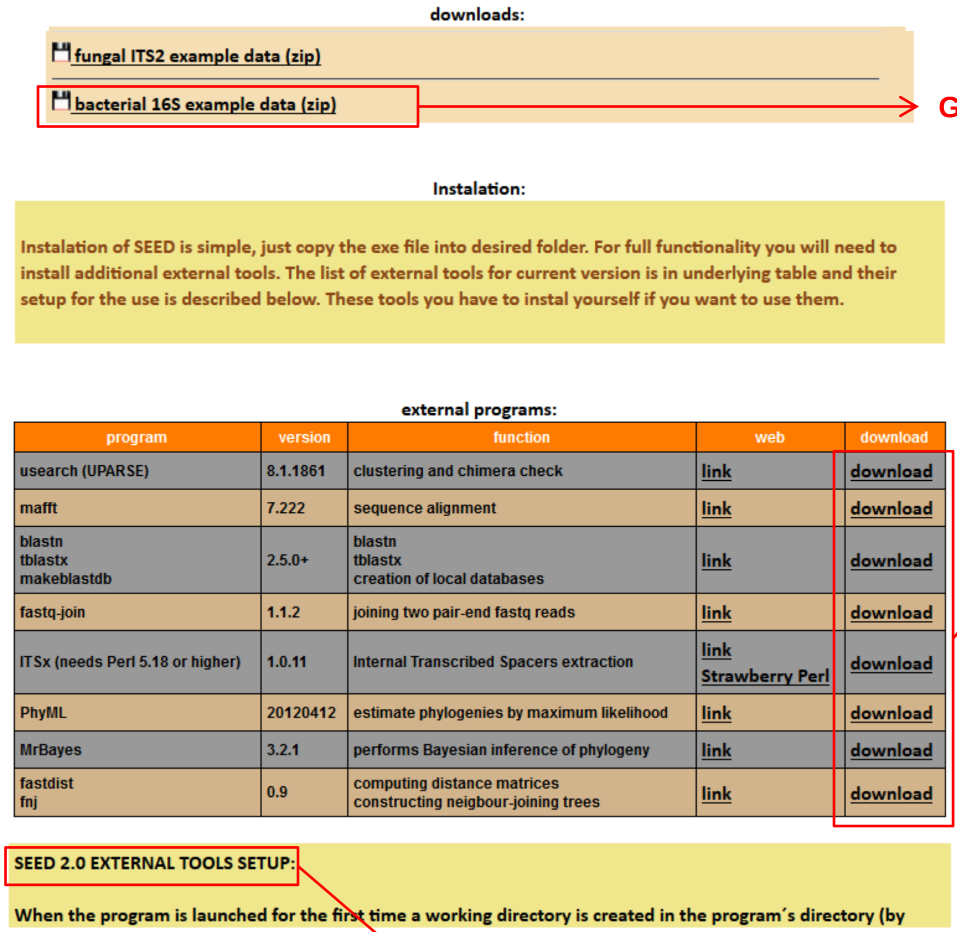
SEED v2.0

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

AMPLICON DATA PROCESSING TUTORIAL (16S 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



downloads:

- 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.

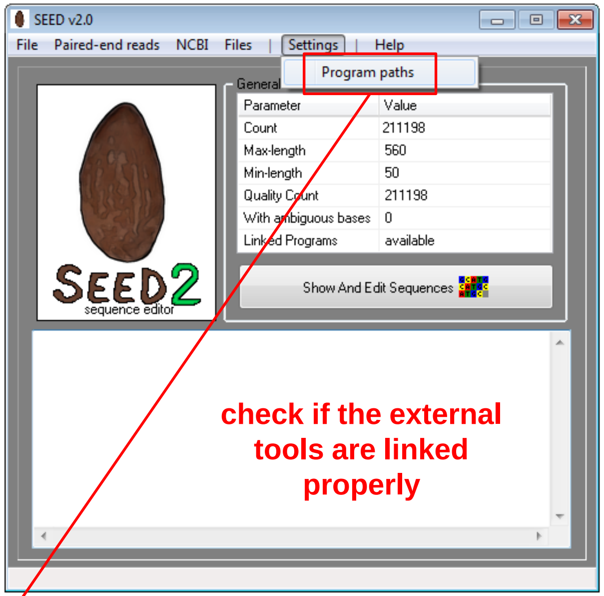
external programs:

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

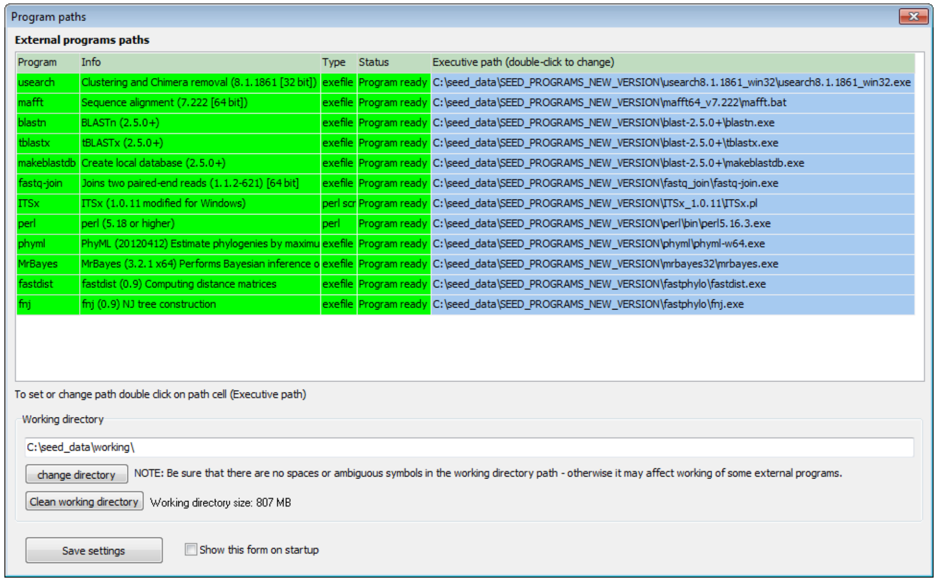
Get example data...



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

Download external tools

check if the external tools are linked properly



Program paths

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\

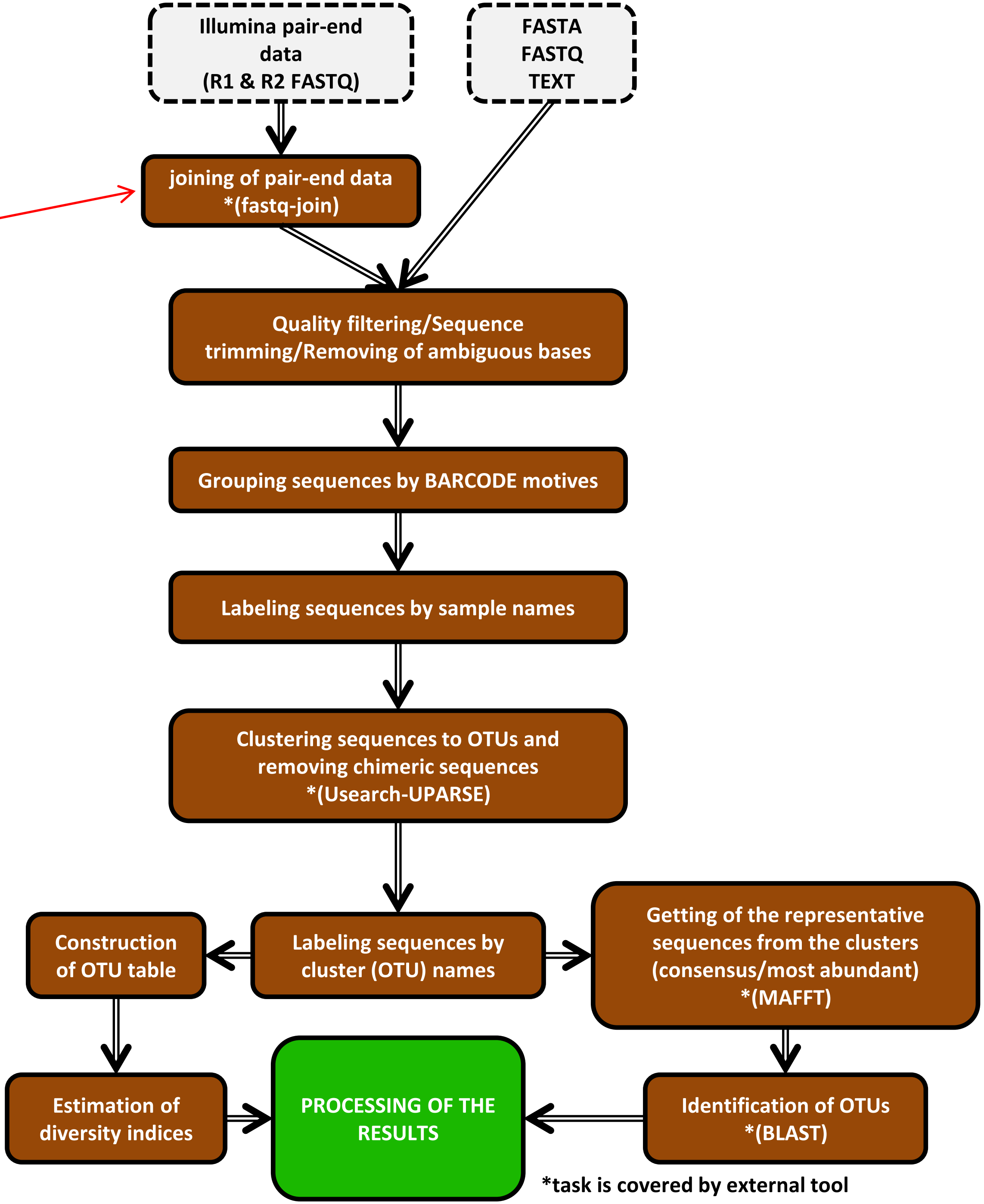
NOTE: Be sure that there are no spaces or ambiguous symbols in the working directory path - otherwise it may affect working of some external programs.

Working directory size: 807 MB

☐ Show this form on startup

Set external tools...

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



You are here

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

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)

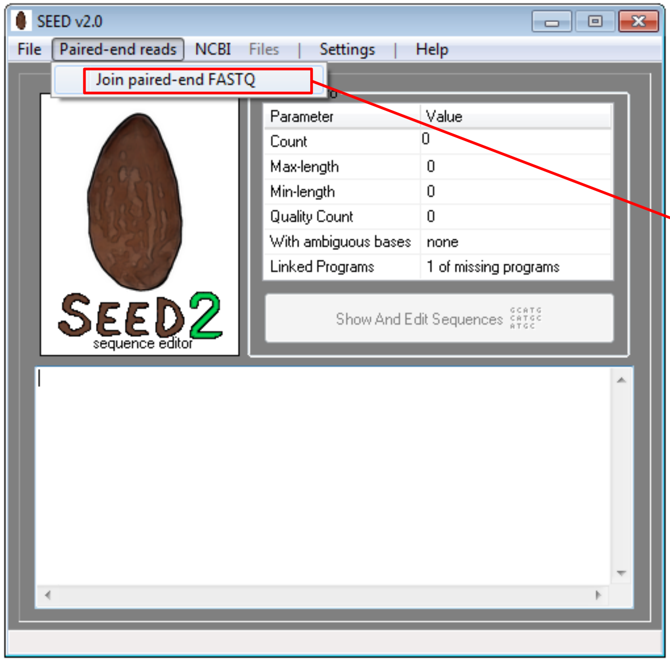
Estimation of diversity indices

PROCESSING OF THE RESULTS

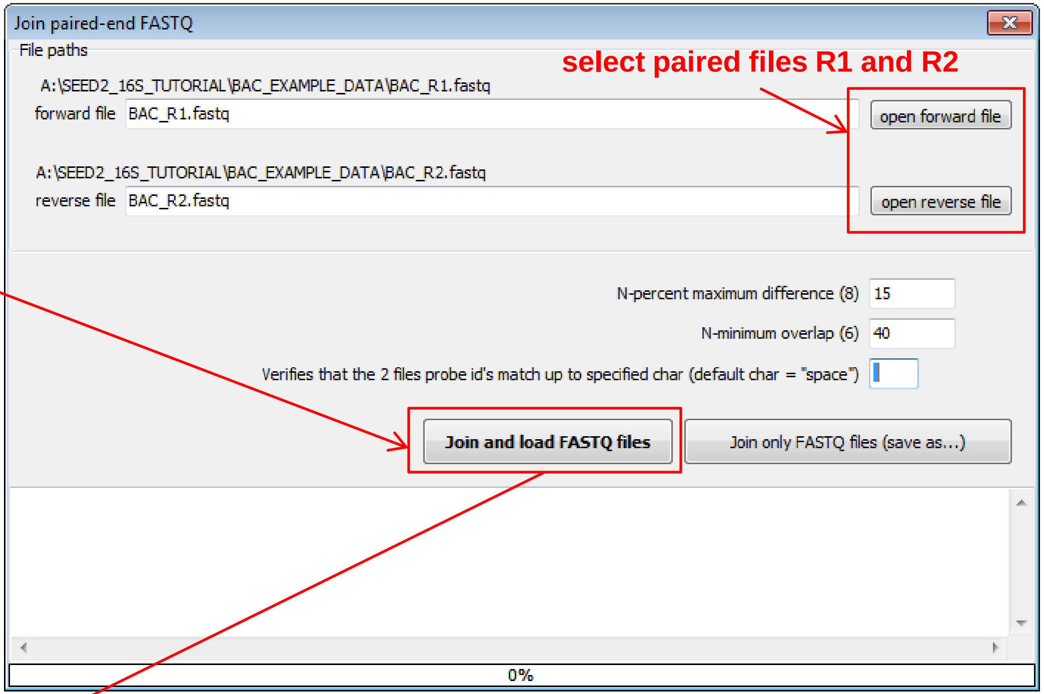
Identification of OTUs *(BLAST)

*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

A:\SEED2_16S_TUTORIAL\BAC_EXAMPLE_DATA\BAC_R1.fastq

forward file BAC_R1.fastq

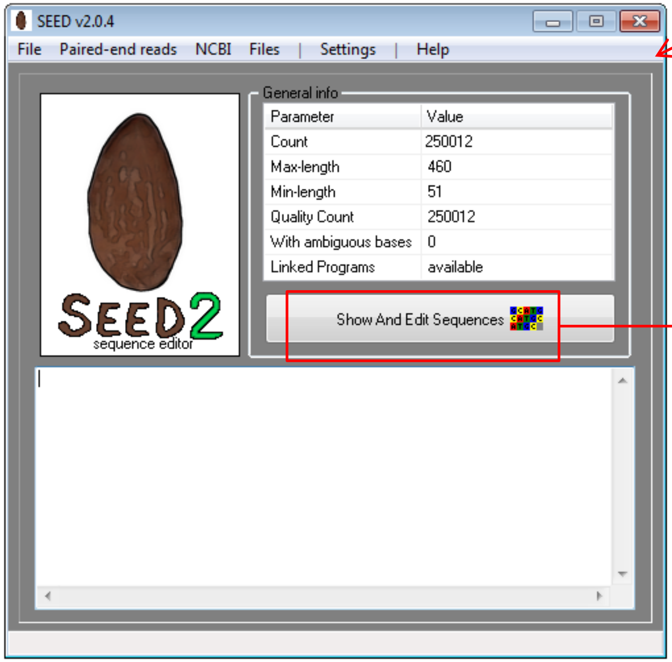
A:\SEED2_16S_TUTORIAL\BAC_EXAMPLE_DATA\BAC_R2.fastq

reverse file BAC_R2.fastq

N-percent maximum difference (8) 15

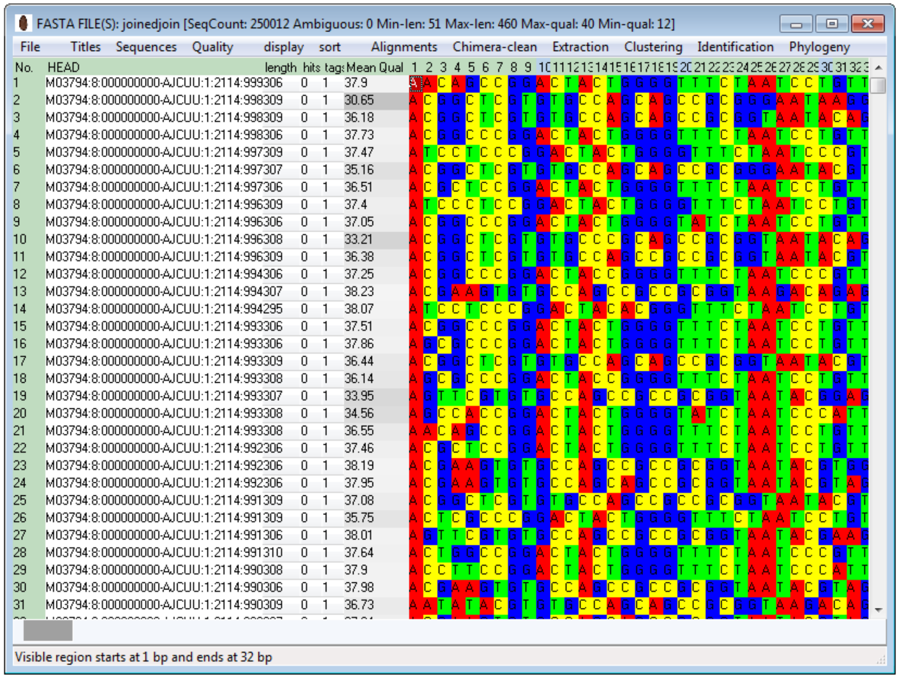
N-minimum overlap (6) 40

Verifies that the 2 files probe id's match up to specified char (default char = "space")



General info

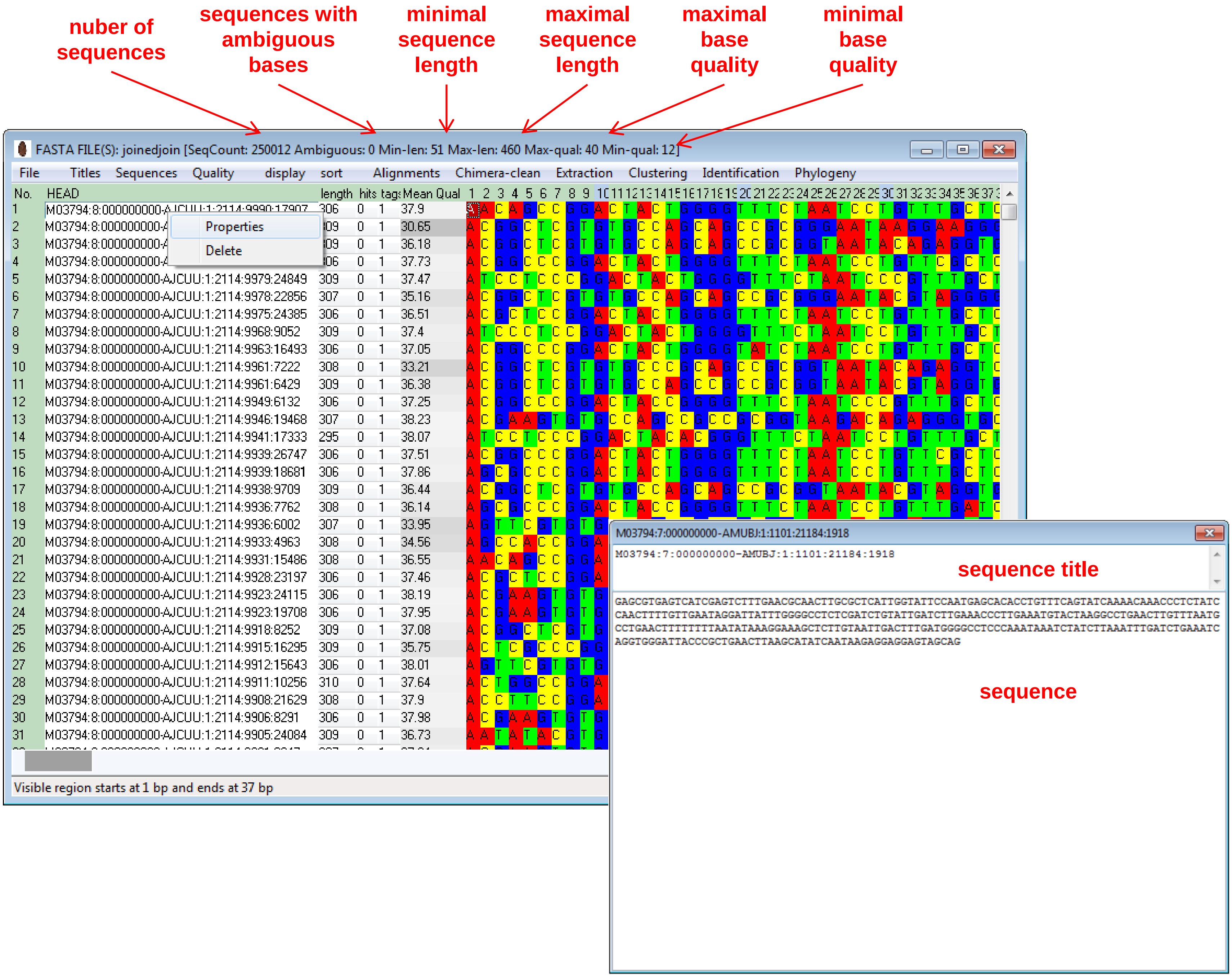
Parameter	Value
Count	250012
Max-length	460
Min-length	51
Quality Count	250012
With ambiguous bases	0
Linked Programs	available



FASTA FILE(S): joinedjoin [SeqCount: 250012 Ambiguous: 0 Min-len: 51 Max-len: 460 Max-qual: 40 Min-qual: 12]

No.	HEAD	length	hits	tag	Mean Qual	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	32
1	M03794:8:000000000-AJCUU:1:2114:999	306	0	1	37.9	A	A	C	A	G	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T
2	M03794:8:000000000-AJCUU:1:2114:998	309	0	1	30.65	A	C	G	G	C	T	C	G	T	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	G	A	A	T	A	A	G
3	M03794:8:000000000-AJCUU:1:2114:998	309	0	1	36.18	A	C	G	G	C	T	C	G	T	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	A
4	M03794:8:000000000-AJCUU:1:2114:998	306	0	1	37.73	A	C	G	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T
5	M03794:8:000000000-AJCUU:1:2114:997	309	0	1	37.47	A	T	C	C	T	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	C	G
6	M03794:8:000000000-AJCUU:1:2114:997	307	0	1	35.16	A	C	G	G	C	T	C	G	T	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	G	A	A	T	A	C	G
7	M03794:8:000000000-AJCUU:1:2114:997	306	0	1	36.51	A	C	G	C	T	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T
8	M03794:8:000000000-AJCUU:1:2114:996	309	0	1	37.4	A	T	C	C	C	T	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G
9	M03794:8:000000000-AJCUU:1:2114:996	306	0	1	37.05	A	C	G	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	A	T	C	T	A	A	T	C	C	T	G	T
10	M03794:8:000000000-AJCUU:1:2114:996	308	0	1	33.21	A	C	G	G	C	T	C	G	T	G	T	G	C	C	C	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	A
11	M03794:8:000000000-AJCUU:1:2114:996	309	0	1	36.38	A	C	G	G	C	T	C	G	T	G	T	G	C	C	A	G	C	C	G	C	C	G	C	G	G	T	A	A	T	A	C	G
12	M03794:8:000000000-AJCUU:1:2114:994	306	0	1	37.25	A	C	G	G	C	C	C	G	G	A	C	T	A	C	C	G	G	G	G	T	T	T	C	T	A	A	T	C	C	C	G	T
13	M03794:8:000000000-AJCUU:1:2114:994	307	0	1	38.23	A	C	G	A	A	G	T	G	T	G	C	C	A	G	C	C	G	C	C	G	C	G	G	T	A	A	G	A	C	A	G	A
14	M03794:8:000000000-AJCUU:1:2114:994	295	0	1	38.07	A	T	C	C	T	C	C	C	G	G	A	C	T	A	C	A	C	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G
15	M03794:8:000000000-AJCUU:1:2114:993	306	0	1	37.51	A	C	G	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T
16	M03794:8:000000000-AJCUU:1:2114:993	306	0	1	37.86	A	G	C	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T
17	M03794:8:000000000-AJCUU:1:2114:993	309	0	1	36.44	A	C	G	G	C	T	C	G	T	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	G
18	M03794:8:000000000-AJCUU:1:2114:993	308	0	1	36.14	A	G	C	G	C	C	C	G	G	A	C	T	A	C	C	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T
19	M03794:8:000000000-AJCUU:1:2114:993	307	0	1	33.95	A	G	T	T	C	G	T	G	T	G	C	C	A	G	C	C	G	C	C	G	C	G	G	T	A	A	T	A	C	G	G	A
20	M03794:8:000000000-AJCUU:1:2114:993	308	0	1	34.56	A	G	C	C	A	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	A	T	C	T	A	A	T	C	C	C	A	T
21	M03794:8:000000000-AJCUU:1:2114:993	308	0	1	36.55	A	A	C	A	G	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T
22	M03794:8:000000000-AJCUU:1:2114:992	306	0	1	37.46	A	C	G	C	T	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T
23	M03794:8:000000000-AJCUU:1:2114:992	306	0	1	38.19	A	C	G	A	A	G	T	G	T	G	C	C	A	G	C	C	G	C	C	G	C	G	G	T	A	A	T	A	C	G	T	G
24	M03794:8:000000000-AJCUU:1:2114:992	306	0	1	37.95	A	C	G	A	A	G	T	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	T	A	C	G	T	A
25	M03794:8:000000000-AJCUU:1:2114:991	309	0	1	37.08	A	C	G	G	C	T	C	G	T	G	T	G	C	C	A	G	C	C	G	C	C	G	C	G	G	T	A	A	T	A	C	G
26	M03794:8:000000000-AJCUU:1:2114:991	309	0	1	35.75	A	C	T	C	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G
27	M03794:8:000000000-AJCUU:1:2114:991	306	0	1	38.01	A	G	T	T	C	G	T	G	T	G	C	C	A	G	C	C	G	C	C	G	C	G	G	T	A	A	T	A	C	G	A	A
28	M03794:8:000000000-AJCUU:1:2114:991	310	0	1	37.64	A	C	T	G	G	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	C	G	T
29	M03794:8:000000000-AJCUU:1:2114:990	308	0	1	37.9	A	C	C	T	T	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	C	A	T
30	M03794:8:000000000-AJCUU:1:2114:990	306	0	1	37.98	A	C	G	A	A	G	T	G	T	G	C	C	A	G	C	C	G	C	C	G	C	G	G	T	A	A	T	A	C	G	T	A
31	M03794:8:000000000-AJCUU:1:2114:990	309	0	1	36.73	A	A	T	A	T	A	C	G	T	G	T	G	C	C	A	G	C	A	G	C	C	G	C	G	G	T	A	A	G	A	C	A

Visible region starts at 1 bp and ends at 32 bp



FASTA FILE(S): joinedjoin [SeqCount: 250012 Ambiguous: 0 Min-len: 51 Max-len: 460 Max-qual: 40 Min-qual: 12]

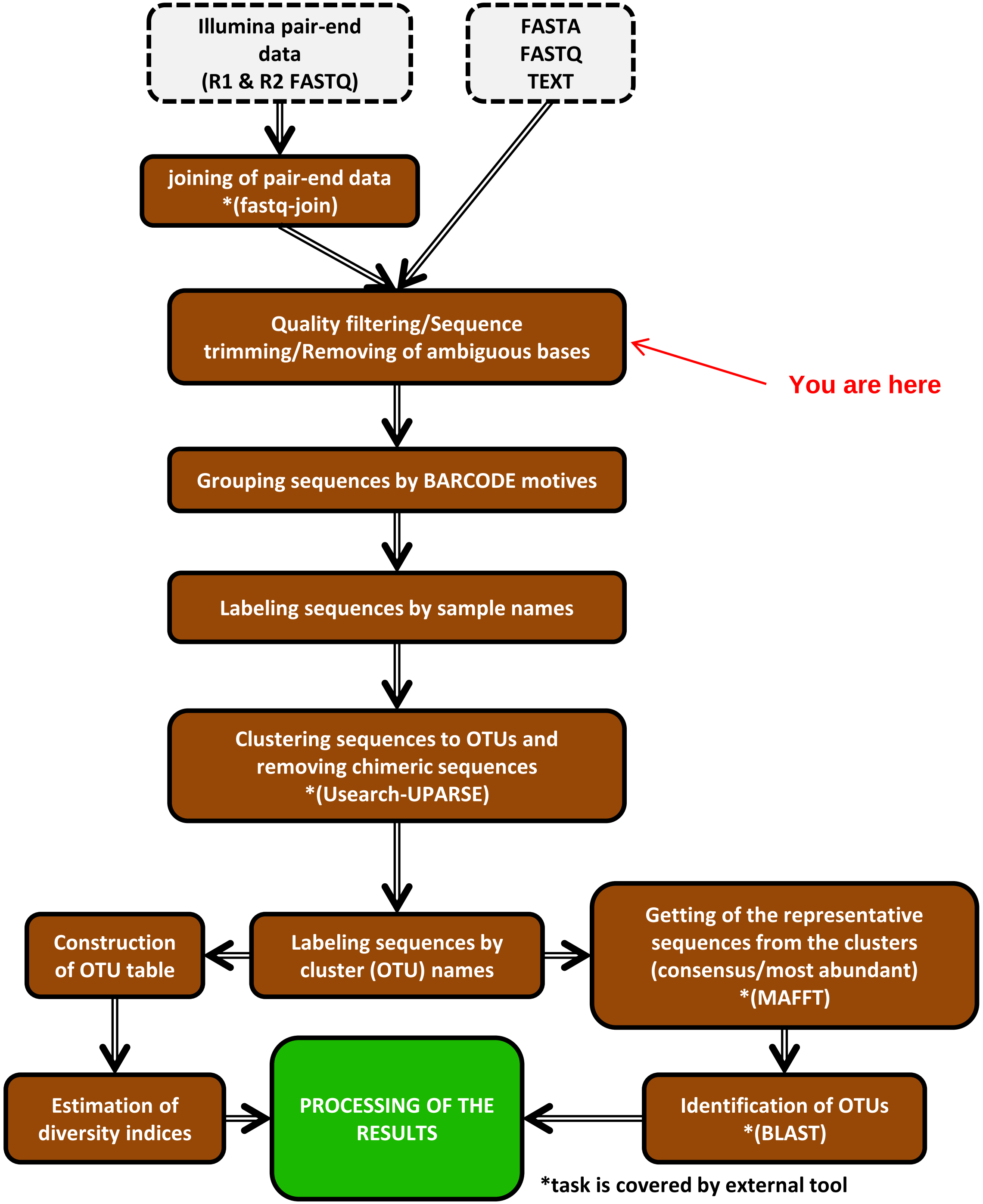
No.	HEAD	length	hits	tags	Mean Qual
1	M03794:8:000000000-AJCUU:1:2114:9990:17907	306	0	1	37.9
2	M03794:8:000000000-A	09	0	1	30.65
3	M03794:8:000000000-A	09	0	1	36.18
4	M03794:8:000000000-A	06	0	1	37.73
5	M03794:8:000000000-AJCUU:1:2114:9979:24849	309	0	1	37.47
6	M03794:8:000000000-AJCUU:1:2114:9978:22856	307	0	1	35.16
7	M03794:8:000000000-AJCUU:1:2114:9975:24385	306	0	1	36.51
8	M03794:8:000000000-AJCUU:1:2114:9968:9052	309	0	1	37.4
9	M03794:8:000000000-AJCUU:1:2114:9963:16493	306	0	1	37.05
10	M03794:8:000000000-AJCUU:1:2114:9961:7222	308	0	1	33.21
11	M03794:8:000000000-AJCUU:1:2114:9961:6429	309	0	1	36.38
12	M03794:8:000000000-AJCUU:1:2114:9949:6132	306	0	1	37.25
13	M03794:8:000000000-AJCUU:1:2114:9946:19468	307	0	1	38.23
14	M03794:8:000000000-AJCUU:1:2114:9941:17333	295	0	1	38.07
15	M03794:8:000000000-AJCUU:1:2114:9939:26747	306	0	1	37.51
16	M03794:8:000000000-AJCUU:1:2114:9939:18681	306	0	1	37.86
17	M03794:8:000000000-AJCUU:1:2114:9938:9709	309	0	1	36.44
18	M03794:8:000000000-AJCUU:1:2114:9936:7762	308	0	1	36.14
19	M03794:8:000000000-AJCUU:1:2114:9936:6002	307	0	1	33.95
20	M03794:8:000000000-AJCUU:1:2114:9933:4963	308	0	1	34.56
21	M03794:8:000000000-AJCUU:1:2114:9931:15486	308	0	1	36.55
22	M03794:8:000000000-AJCUU:1:2114:9928:23197	306	0	1	37.46
23	M03794:8:000000000-AJCUU:1:2114:9923:24115	306	0	1	38.19
24	M03794:8:000000000-AJCUU:1:2114:9923:19708	306	0	1	37.95
25	M03794:8:000000000-AJCUU:1:2114:9918:8252	309	0	1	37.08
26	M03794:8:000000000-AJCUU:1:2114:9915:16295	309	0	1	35.75
27	M03794:8:000000000-AJCUU:1:2114:9912:15643	306	0	1	38.01
28	M03794:8:000000000-AJCUU:1:2114:9911:10256	310	0	1	37.64
29	M03794:8:000000000-AJCUU:1:2114:9908:21629	308	0	1	37.9
30	M03794:8:000000000-AJCUU:1:2114:9906:8291	306	0	1	37.98
31	M03794:8:000000000-AJCUU:1:2114:9905:24084	309	0	1	36.73

Visible region starts at 1 bp and ends at 37 bp

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

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

```
GAGCGTGAGTCATCGAGTCTTTGAACGCAACTTGCGCTCATTGGTATTCCAATGAGCACACCTGTTTCAGTATCAAAACAAACCCTCTATC
CAACTTTTGTTGAATAGGATTATTTGGGGCCTCTCGATCTGTATTGATCTTGAAACCCTTGAAATGTACTAAGGCCTGAACTTGTTTAATG
CCTGAACTTTTTTTTAATATAAAGGAAAGCTCTTGTAATTGACTTTGATGGGGCCTCCCAAATAAATCTATCTTAAATTTGATCTGAAATC
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

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)

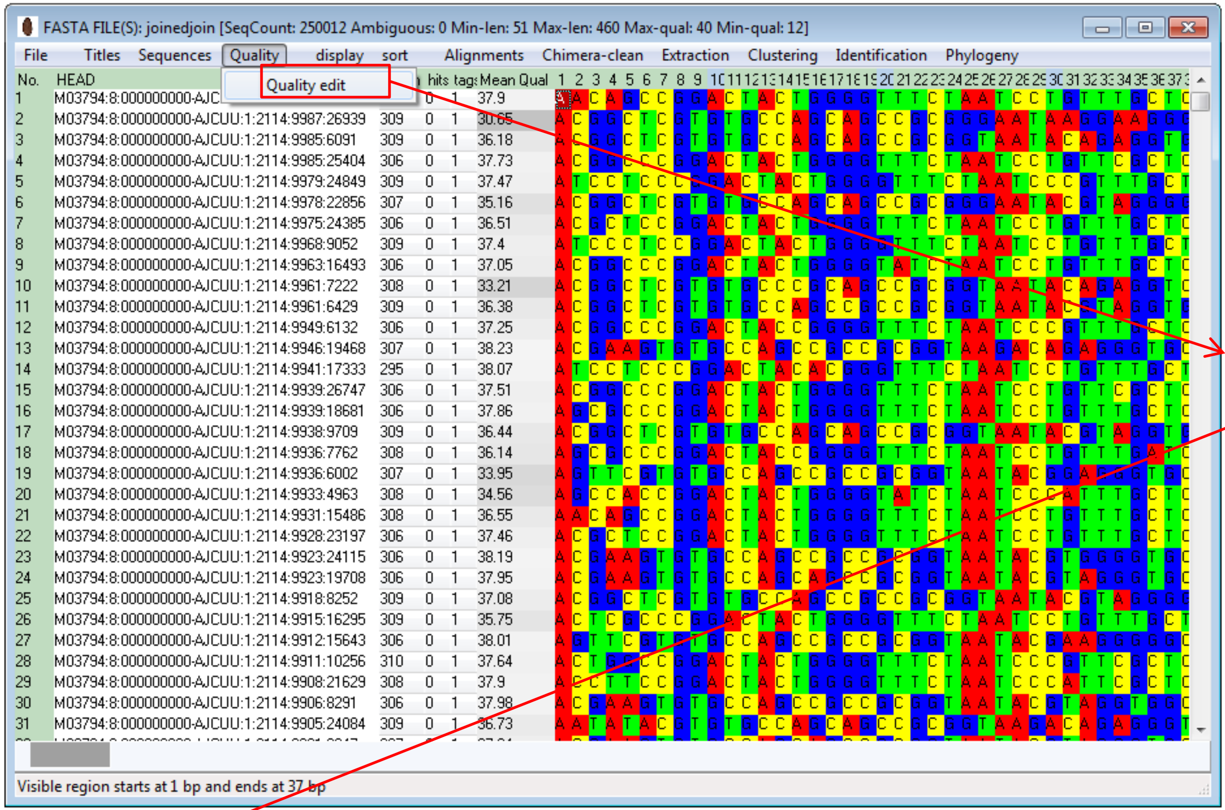
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs
*(BLAST)

*task is covered by external tool

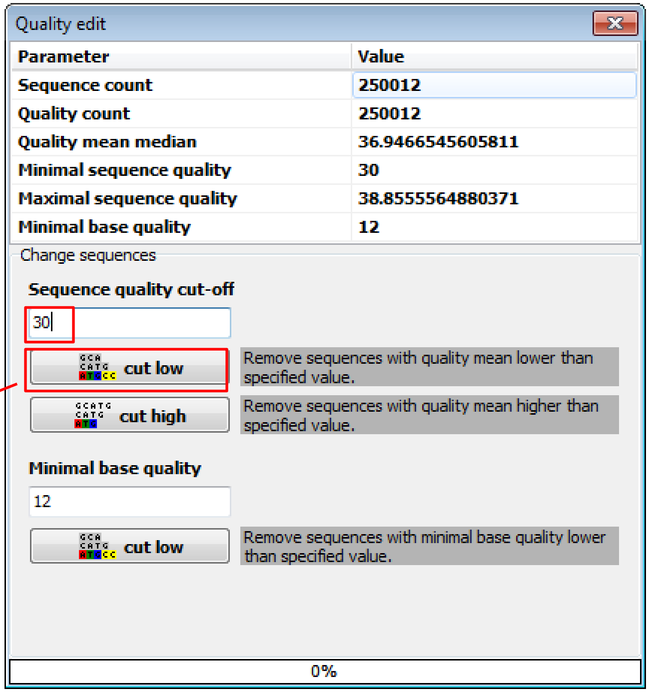
Filter sequences by their quality…



FASTA FILE(S): joinedjoin [SeqCount: 250012 Ambiguous: 0 Min-len: 51 Max-len: 460 Max-qual: 40 Min-qual: 12]

No.	HEAD		hits	tag	Mean Qual
1	M03794:8:000000000-AJC		0	1	37.9
2	M03794:8:000000000-AJCUU:1:2114:9987:26939	309	0	1	30.65
3	M03794:8:000000000-AJCUU:1:2114:9985:6091	309	0	1	36.18
4	M03794:8:000000000-AJCUU:1:2114:9985:25404	306	0	1	37.73
5	M03794:8:000000000-AJCUU:1:2114:9979:24849	309	0	1	37.47
6	M03794:8:000000000-AJCUU:1:2114:9978:22856	307	0	1	35.16
7	M03794:8:000000000-AJCUU:1:2114:9975:24385	306	0	1	36.51
8	M03794:8:000000000-AJCUU:1:2114:9968:9052	309	0	1	37.4
9	M03794:8:000000000-AJCUU:1:2114:9963:16493	306	0	1	37.05
10	M03794:8:000000000-AJCUU:1:2114:9961:7222	308	0	1	33.21
11	M03794:8:000000000-AJCUU:1:2114:9961:6429	309	0	1	36.38
12	M03794:8:000000000-AJCUU:1:2114:9949:6132	306	0	1	37.25
13	M03794:8:000000000-AJCUU:1:2114:9946:19468	307	0	1	38.23
14	M03794:8:000000000-AJCUU:1:2114:9941:17333	295	0	1	38.07
15	M03794:8:000000000-AJCUU:1:2114:9939:26747	306	0	1	37.51
16	M03794:8:000000000-AJCUU:1:2114:9939:18681	306	0	1	37.86
17	M03794:8:000000000-AJCUU:1:2114:9938:9709	309	0	1	36.44
18	M03794:8:000000000-AJCUU:1:2114:9936:7762	308	0	1	36.14
19	M03794:8:000000000-AJCUU:1:2114:9936:6002	307	0	1	33.95
20	M03794:8:000000000-AJCUU:1:2114:9933:4963	308	0	1	34.56
21	M03794:8:000000000-AJCUU:1:2114:9931:15486	308	0	1	36.55
22	M03794:8:000000000-AJCUU:1:2114:9928:23197	306	0	1	37.46
23	M03794:8:000000000-AJCUU:1:2114:9923:24115	306	0	1	38.19
24	M03794:8:000000000-AJCUU:1:2114:9923:19708	306	0	1	37.95
25	M03794:8:000000000-AJCUU:1:2114:9918:8252	309	0	1	37.08
26	M03794:8:000000000-AJCUU:1:2114:9915:16295	309	0	1	35.75
27	M03794:8:000000000-AJCUU:1:2114:9912:15643	306	0	1	38.01
28	M03794:8:000000000-AJCUU:1:2114:9911:10256	310	0	1	37.64
29	M03794:8:000000000-AJCUU:1:2114:9908:21629	308	0	1	37.9
30	M03794:8:000000000-AJCUU:1:2114:9906:8291	306	0	1	37.98
31	M03794:8:000000000-AJCUU:1:2114:9905:24084	309	0	1	36.73

Visible region starts at 1 bp and ends at 37 bp



Quality edit

Parameter	Value
Sequence count	250012
Quality count	250012
Quality mean median	36.9466545605811
Minimal sequence quality	30
Maximal sequence quality	38.8555564880371
Minimal base quality	12

Change sequences

Sequence quality cut-off: 30

Remove sequences with quality mean lower than specified value.

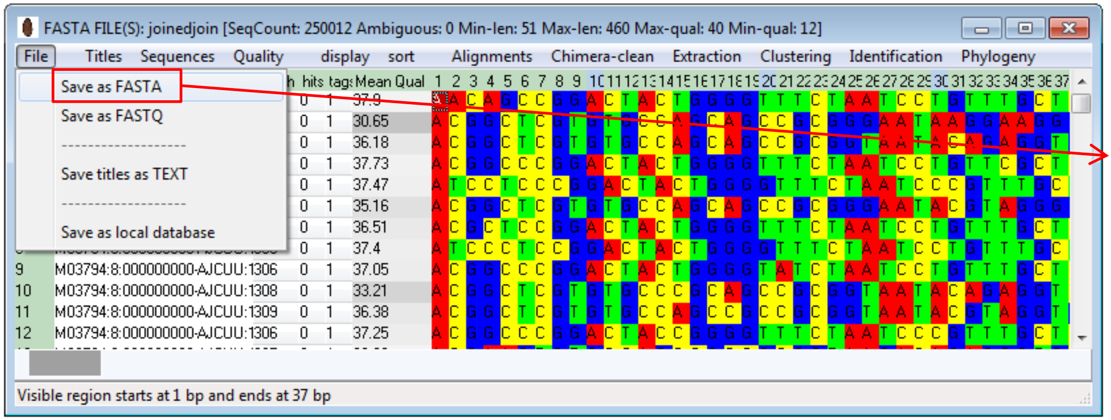
Remove sequences with quality mean higher than specified value.

Minimal base quality: 12

Remove sequences with minimal base quality lower than specified value.

0%

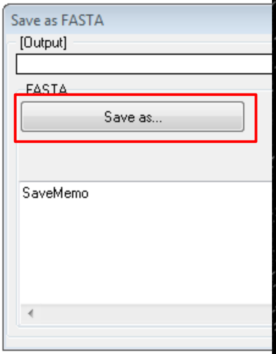
Save files as FASTA after each important step…



FASTA FILE(S): joinedjoin [SeqCount: 250012 Ambiguous: 0 Min-len: 51 Max-len: 460 Max-qual: 40 Min-qual: 12]

No.	HEAD	hits	tag	Mean Qual
9	M03794:8:000000000-AJCUU:1306	0	1	37.05
10	M03794:8:000000000-AJCUU:1308	0	1	33.21
11	M03794:8:000000000-AJCUU:1309	0	1	36.38
12	M03794:8:000000000-AJCUU:1306	0	1	37.25

Visible region starts at 1 bp and ends at 37 bp

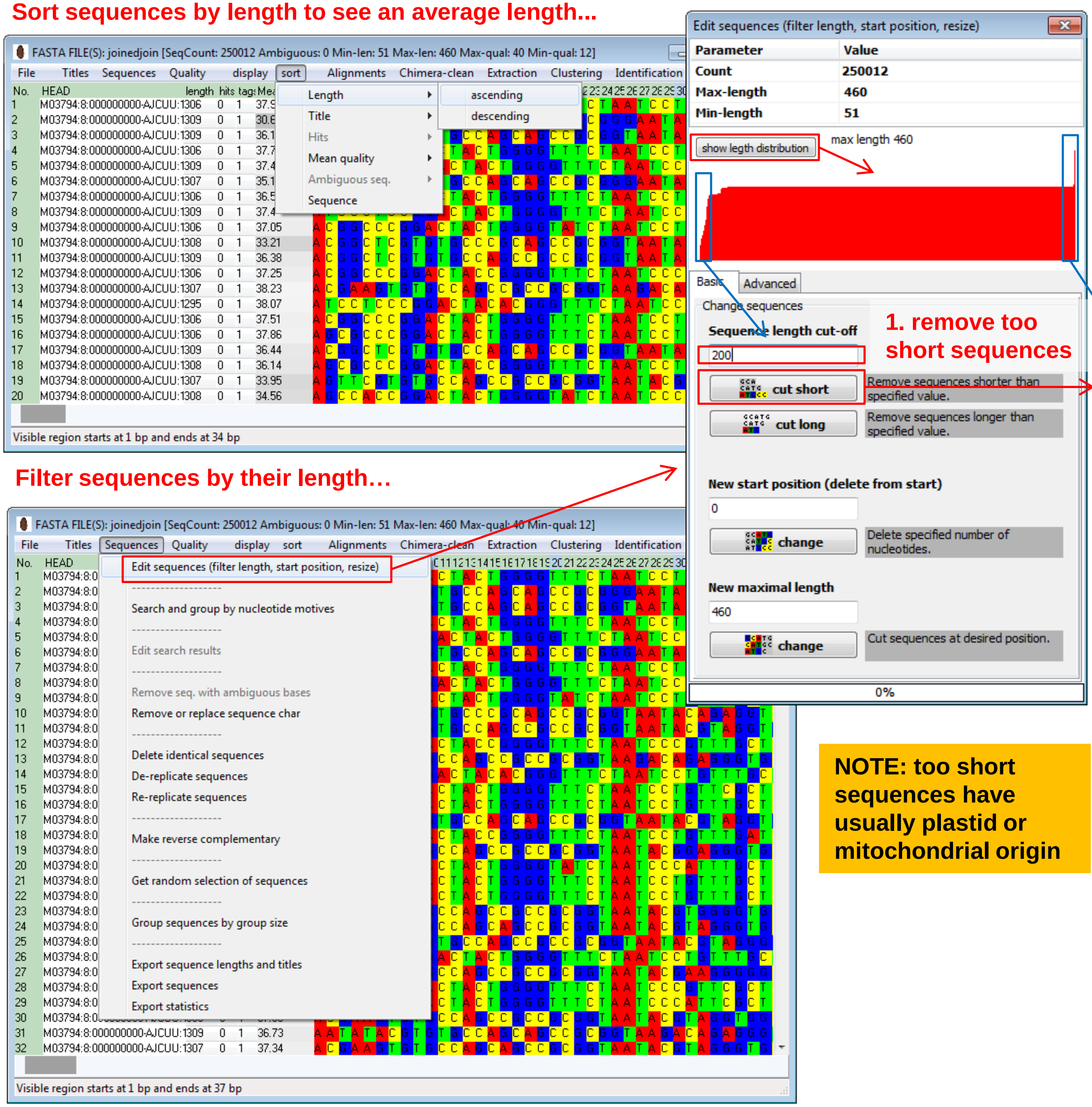


GOOD TO SAVE NOW!

16S_example _joined_qm30

example of file name

Sort sequences by length to see an average length…


FASTA FILE(S): joinedjoin [SeqCount: 250012 Ambiguous: 0 Min-len: 51 Max-len: 460 Max-qual: 40 Min-qual: 12]

No.	HEAD	length	hits	tags	Mea
1	M03794:8:000000000-AJCUU:1306		0	1	37.9
2	M03794:8:000000000-AJCUU:1309		0	1	30.6
3	M03794:8:000000000-AJCUU:1309		0	1	36.1
4	M03794:8:000000000-AJCUU:1306		0	1	37.7
5	M03794:8:000000000-AJCUU:1309		0	1	37.4
6	M03794:8:000000000-AJCUU:1307		0	1	35.1
7	M03794:8:000000000-AJCUU:1306		0	1	36.5
8	M03794:8:000000000-AJCUU:1309		0	1	37.4
9	M03794:8:000000000-AJCUU:1306		0	1	37.05
10	M03794:8:000000000-AJCUU:1308		0	1	33.21
11	M03794:8:000000000-AJCUU:1309		0	1	36.38
12	M03794:8:000000000-AJCUU:1306		0	1	37.25
13	M03794:8:000000000-AJCUU:1307		0	1	38.23
14	M03794:8:000000000-AJCUU:1295		0	1	38.07
15	M03794:8:000000000-AJCUU:1306		0	1	37.51
16	M03794:8:000000000-AJCUU:1306		0	1	37.86
17	M03794:8:000000000-AJCUU:1309		0	1	36.44
18	M03794:8:000000000-AJCUU:1308		0	1	36.14
19	M03794:8:000000000-AJCUU:1307		0	1	33.95
20	M03794:8:000000000-AJCUU:1308		0	1	34.56

Visible region starts at 1 bp and ends at 34 bp

Filter sequences by their length…

FASTA FILE(S): joinedjoin [SeqCount: 250012 Ambiguous: 0 Min-len: 51 Max-len: 460 Max-qual: 40 Min-qual: 12]

No.	HEAD			
31	M03794:8:000000000-AJCUU:1309	0	1	36.73
32	M03794:8:000000000-AJCUU:1307	0	1	37.34

Visible region starts at 1 bp and ends at 37 bp

Edit sequences (filter length, start position, resize)

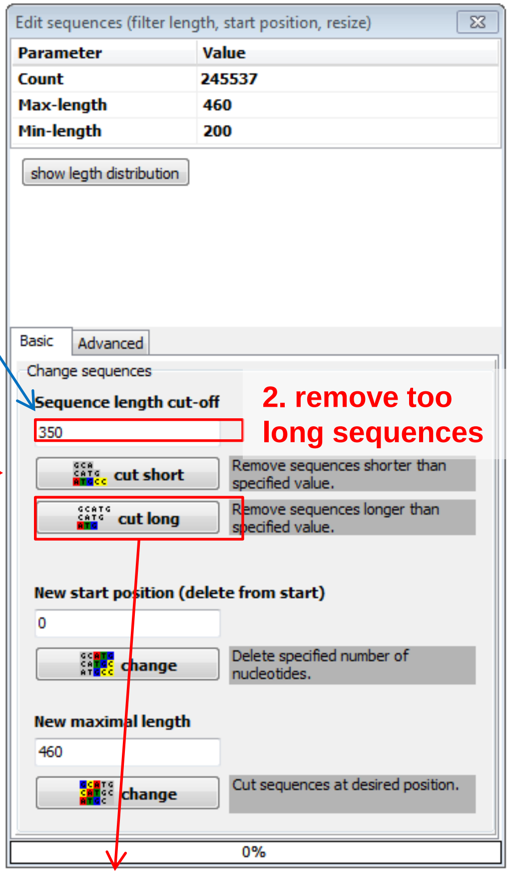
Parameter	Value
Count	250012
Max-length	460
Min-length	51

Sequence length cut-off: 200

New start position (delete from start): 0

New maximal length: 460

1. remove too short sequences


Edit sequences (filter length, start position, resize)

Parameter	Value
Count	245537
Max-length	460
Min-length	200

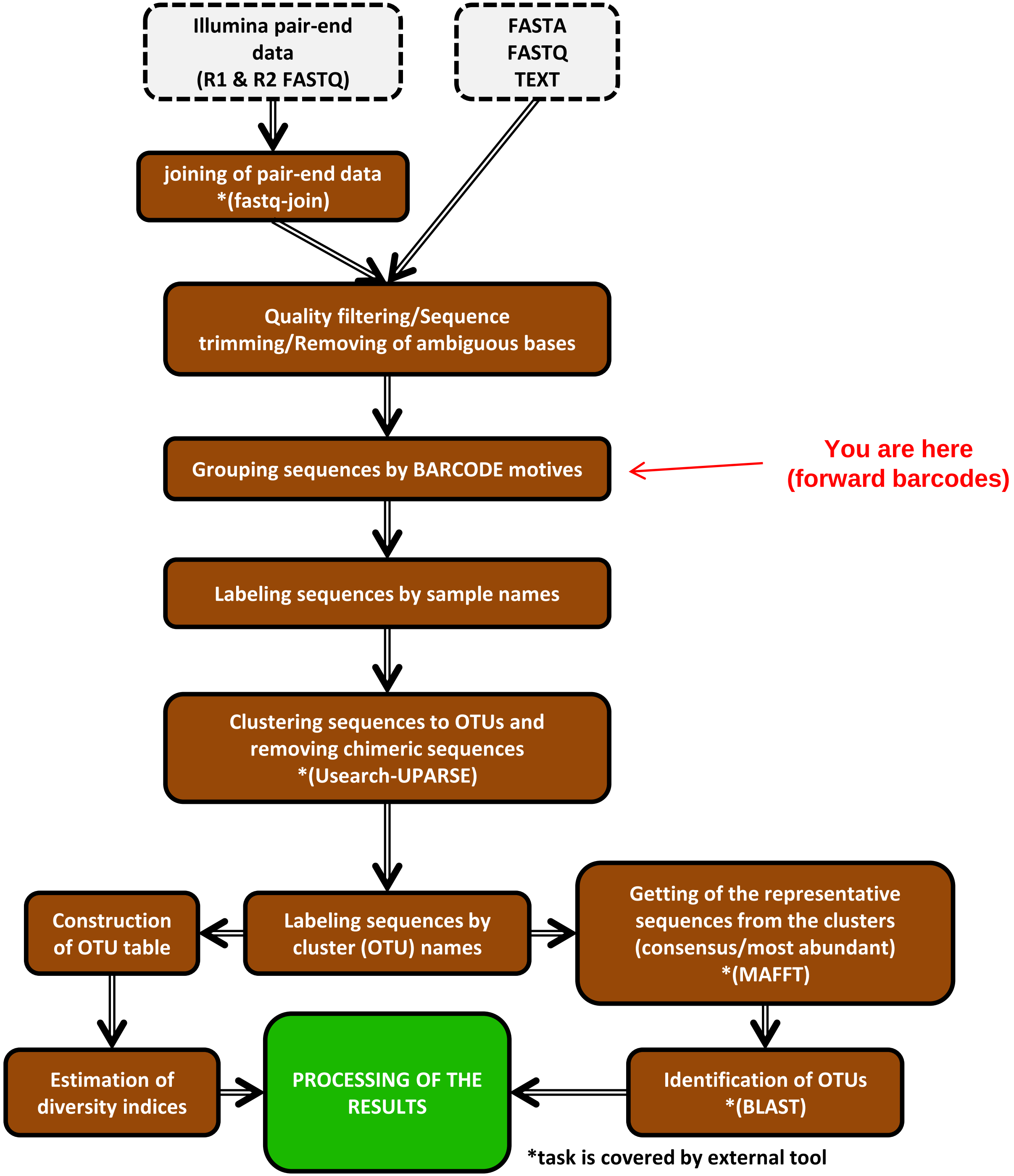
Sequence length cut-off: 350

New start position (delete from start): 0

New maximal length: 460

2. remove too long sequences

NOTE: too short sequences have usually plastid or mitochondrial origin



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

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

515F	GTGCCAGCMGCCGCGGTAA

TAG SPACER PRIMER

Tagged Forward primers

515F_T002	ACGAAGTGTGCCAGCMGCCGCGGTAA
515F_T007	AGCCAGTGTGCCAGCMGCCGCGGTAA
515F_T008	AGTTCGTGTGCCAGCMGCCGCGGTAA
515F_T101	ACGGCTCGTGTGCCAGCMGCCGCGGTAA
515F_T103	AATATACGTGTGCCAGCMGCCGCGGTAA

Sequence motive and tag name to search (TAB delimited)

ACGAAGTGTGC	515F_T002
AGCCAGTGTGC	515F_T007
AGTTCGTGTGC	515F_T008
ACGGCTCGTGTGC	515F_T101
AATATACGTGTGC	515F_T103

Reverse primer

806R	GGACTACHVGGGTWTCTAAT

Tagged Reverse primers

806R_T007	AGCCACCGGACTACHVGGGTWTCTAAT
806R_T011	AACAGCCGGACTACHVGGGTWTCTAAT
806R_T020	ACTGGCCGGACTACHVGGGTWTCTAAT
806R_T029	AGCGCCCGGACTACHVGGGTWTCTAAT
806R_T052	ATCCTCCCGGACTACHVGGGTWTCTAAT

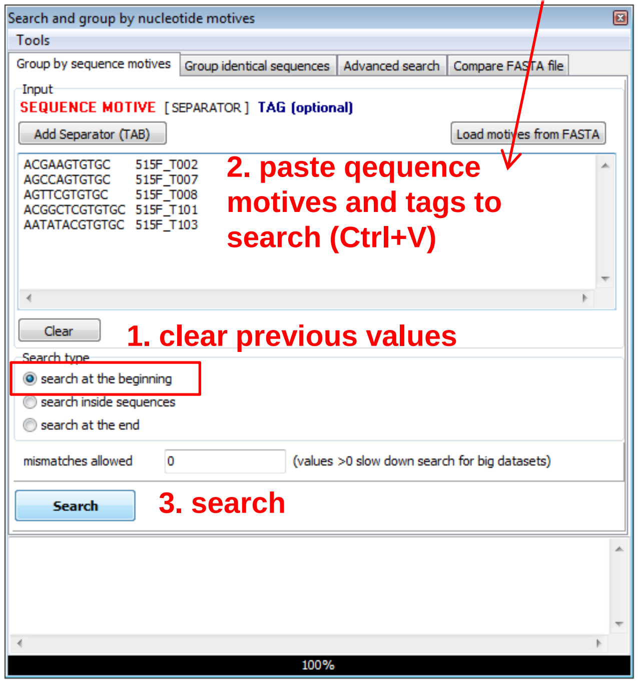
AGCCACCGGAC	806R_T007
AACAGCCGGAC	806R_T011
ACTGGCCGGAC	806R_T020
AGCGCCCGGAC	806R_T029
ATCCTCCCGGAC	806R_T052

FASTA FILE(S): joinedjoin [SeqCount: 245459 Ambiguous: 0 Min-len: 200 Max-len: 350 Max-qual: 40 Min-qual: 12]

30 M03794:8:000000000-AJCUU:1:1114:7992:24422 203 0 1 38.45

Visible region starts at 1 bp and ends at 19 bp

Search for the forward tag motives...



Search and group by nucleotide motives

ACGAAGTGTGC	515F_T002
AGCCAGTGTGC	515F_T007
AGTTCGTGTGC	515F_T008
ACGGCTCGTGTGC	515F_T101
AATATACGTGTGC	515F_T103

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

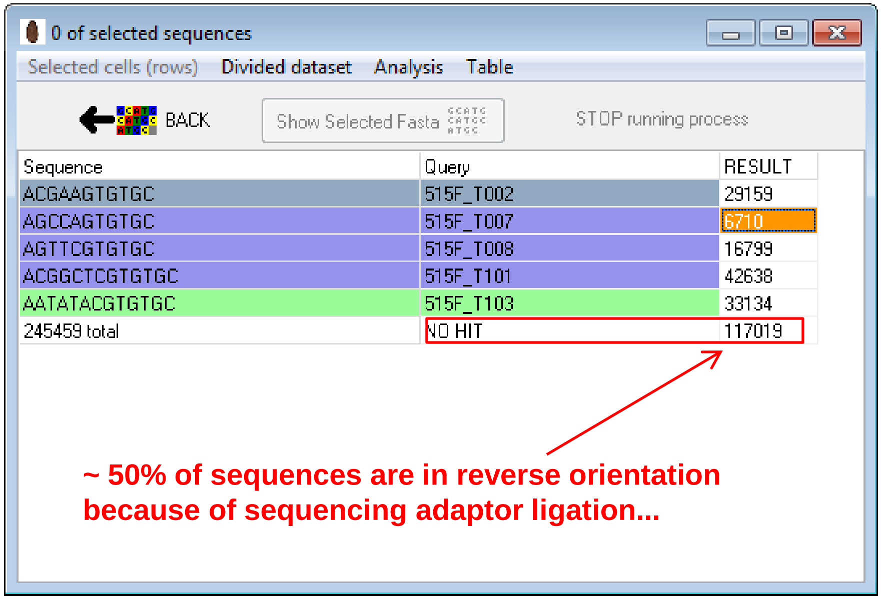
mismatches allowed: 0 (values >0 slow down search for big datasets)

1. clear previous values

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

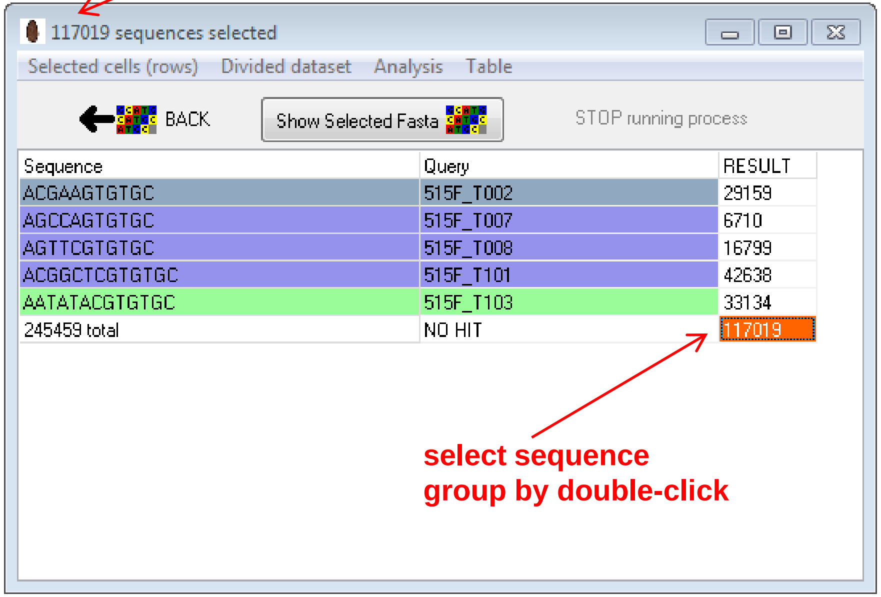
3. search

total nuber of selected sequences



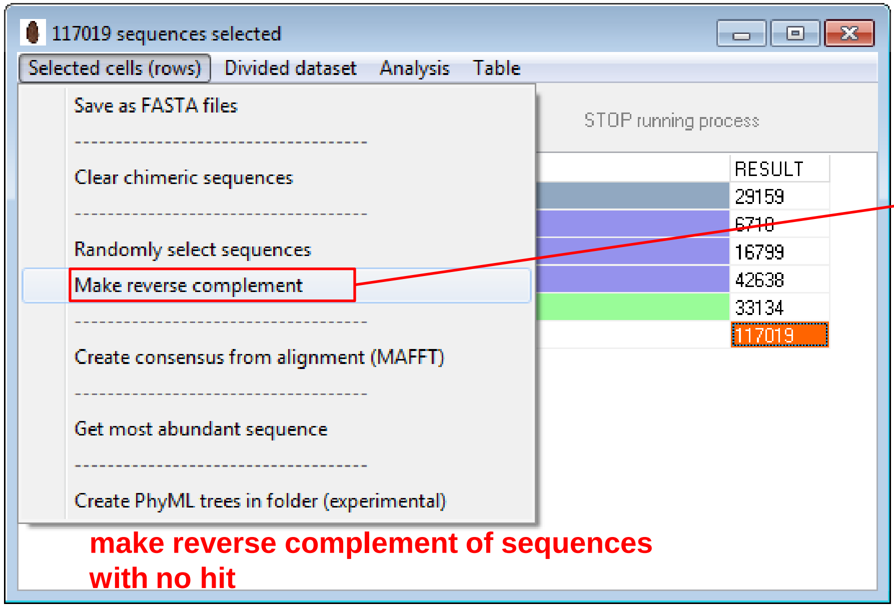
Sequence	Query	RESULT
ACGAAGTGTGC	515F_T002	29159
AGCCAGTGTGC	515F_T007	6710
AGTTCGTGTGC	515F_T008	16799
ACGGCTCGTGTGC	515F_T101	42638
AATATACGTGTGC	515F_T103	33134
245459 total	NO HIT	117019

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



Sequence	Query	RESULT
ACGAAGTGTGC	515F_T002	29159
AGCCAGTGTGC	515F_T007	6710
AGTTCGTGTGC	515F_T008	16799
ACGGCTCGTGTGC	515F_T101	42638
AATATACGTGTGC	515F_T103	33134
245459 total	NO HIT	117019

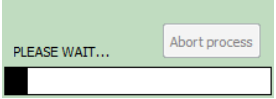
select sequence group by double-click



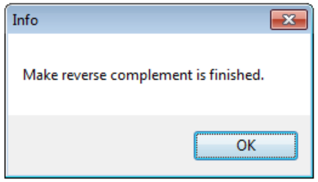
RESULT
29159
6710
16799
42638
33134
117019

make reverse complement of sequences with no hit

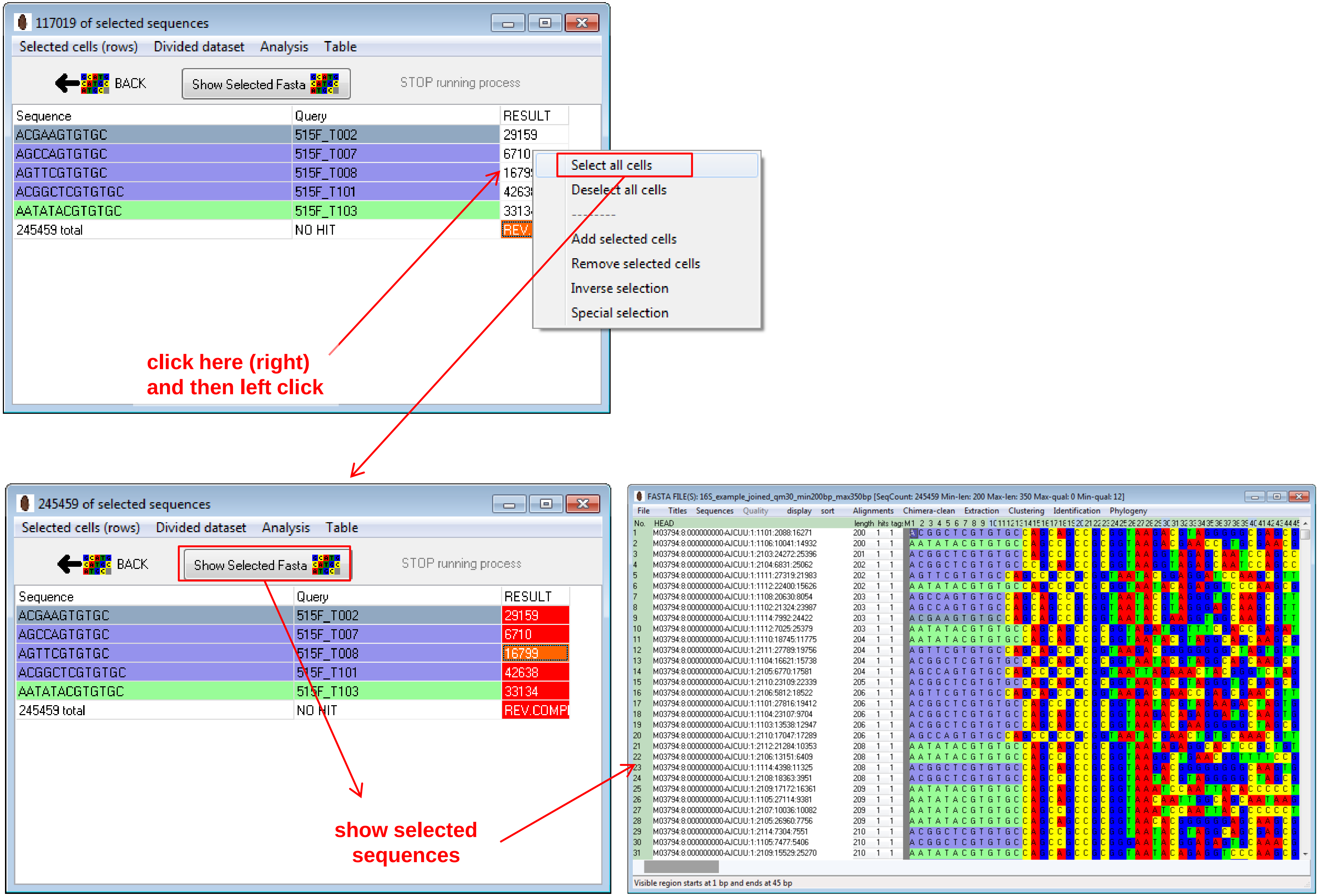
it may take a while...



...and then...



Make reverse complement is finished.



Sequence	Query	RESULT
ACGAAGTGTGC	515F_T002	29159
AGCCAGTGTGC	515F_T007	6710
AGTTCGTGTGC	515F_T008	1679
ACGGCTCGTGTGC	515F_T101	4263
AATATACGTGTGC	515F_T103	3313
245459 total	NO HIT	REV.

click here (right) and then left click

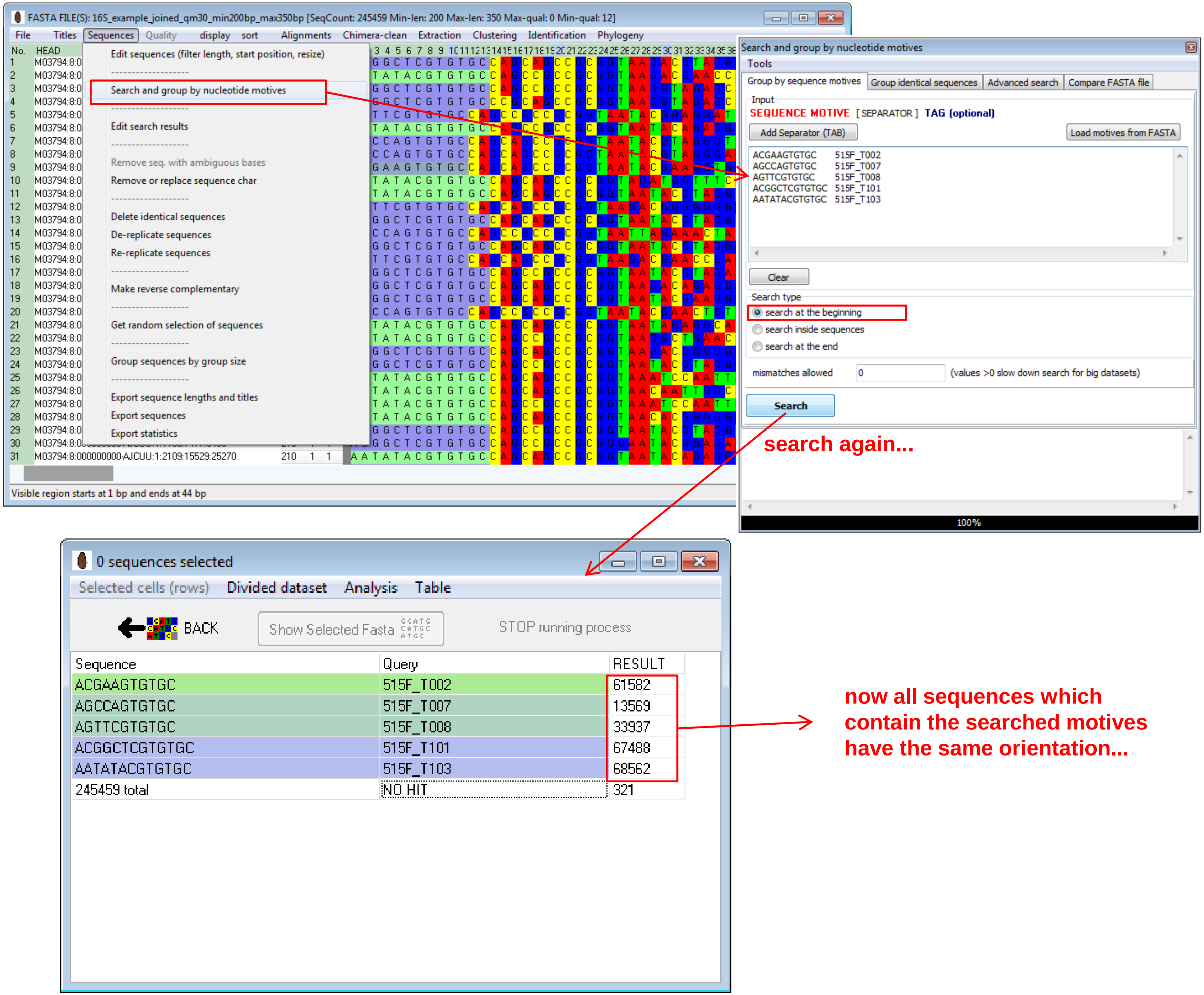
Sequence	Query	RESULT
ACGAAGTGTGC	515F_T002	29159
AGCCAGTGTGC	515F_T007	6710
AGTTCGTGTGC	515F_T008	16799
ACGGCTCGTGTGC	515F_T101	42638
AATATACGTGTGC	515F_T103	33134
245459 total	NO HIT	REV.COMP

show selected sequences

FASTA FILE(S): 16S_example_joined_qm30_min200bp_max350bp [SeqCount: 245459 Min-len: 200 Max-len: 350 Max-qual: 0 Min-qual: 12]

No.	HEAD	length	hits	tags
1	M03794:8:000000000-AJCUU:1:1101:2088:16271	200	1	1
2	M03794:8:000000000-AJCUU:1:1106:10041:14932	200	1	1
3	M03794:8:000000000-AJCUU:1:2103:24272:25396	201	1	1
4	M03794:8:000000000-AJCUU:1:2104:6831:25062	202	1	1
5	M03794:8:000000000-AJCUU:1:1111:27319:21983	202	1	1
6	M03794:8:000000000-AJCUU:1:1112:22400:15626	202	1	1
7	M03794:8:000000000-AJCUU:1:1108:20630:8054	203	1	1
8	M03794:8:000000000-AJCUU:1:1102:21324:23987	203	1	1
9	M03794:8:000000000-AJCUU:1:1114:7992:24422	203	1	1
10	M03794:8:000000000-AJCUU:1:1112:7025:25379	203	1	1
11	M03794:8:000000000-AJCUU:1:1110:18745:11775	204	1	1
12	M03794:8:000000000-AJCUU:1:2111:27789:19756	204	1	1
13	M03794:8:000000000-AJCUU:1:1104:16621:15738	204	1	1
14	M03794:8:000000000-AJCUU:1:2105:6770:17581	204	1	1
15	M03794:8:000000000-AJCUU:1:2110:23109:22339	205	1	1
16	M03794:8:000000000-AJCUU:1:2106:5812:18522	206	1	1
17	M03794:8:000000000-AJCUU:1:1101:27816:19412	206	1	1
18	M03794:8:000000000-AJCUU:1:1104:23107:9704	206	1	1
19	M03794:8:000000000-AJCUU:1:1103:13538:12947	206	1	1
20	M03794:8:000000000-AJCUU:1:2110:17047:17289	206	1	1
21	M03794:8:000000000-AJCUU:1:2112:21284:10353	208	1	1
22	M03794:8:000000000-AJCUU:1:2106:13151:6409	208	1	1
23	M03794:8:000000000-AJCUU:1:1114:4398:11325	208	1	1
24	M03794:8:000000000-AJCUU:1:2108:18363:3951	208	1	1
25	M03794:8:000000000-AJCUU:1:2109:17172:16361	209	1	1
26	M03794:8:000000000-AJCUU:1:1105:27114:9381	209	1	1
27	M03794:8:000000000-AJCUU:1:2107:10036:10082	209	1	1
28	M03794:8:000000000-AJCUU:1:2105:26960:7756	209	1	1
29	M03794:8:000000000-AJCUU:1:2114:7304:7551	210	1	1
30	M03794:8:000000000-AJCUU:1:1105:7477:5406	210	1	1
31	M03794:8:000000000-AJCUU:1:2109:15529:25270	210	1	1

Visible region starts at 1 bp and ends at 45 bp



FASTA FILE(S): 16S_example_joined_qm30_min200bp_max350bp [SeqCount: 245459 Min-len: 200 Max-len: 350 Max-qual: 0 Min-qual: 12]

31 M03794:8:000000000-AJCUU:1:2109:15529:25270 210 1 1

Visible region starts at 1 bp and ends at 44 bp

Search and group by nucleotide motives

ACGAAGTGTGC	515F_T002
AGCCAGTGTGC	515F_T007
AGTTCGTGTGC	515F_T008
ACGGCTCGTGTGC	515F_T101
AATATACGTGTGC	515F_T103

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

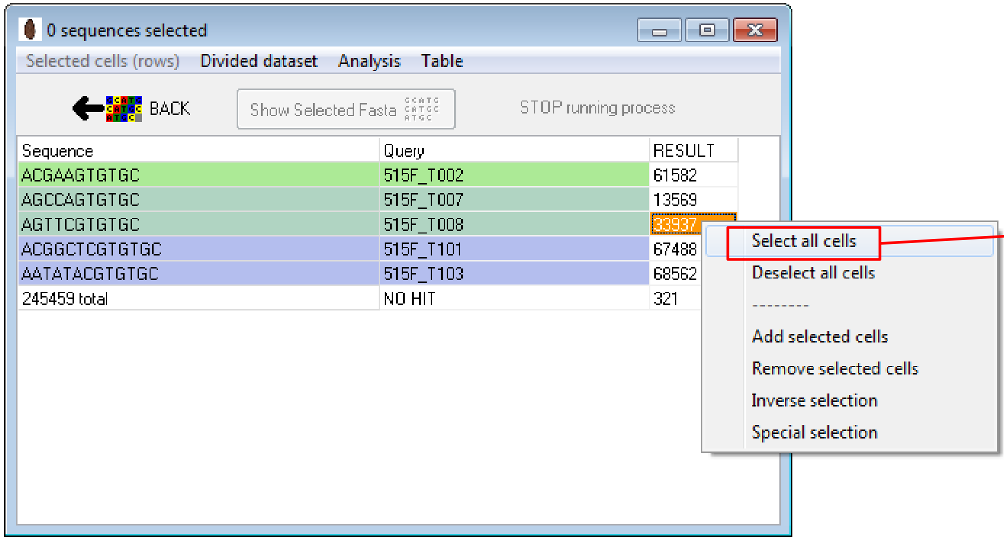
mismatches allowed: 0 (values >0 slow down search for big datasets)

search again...

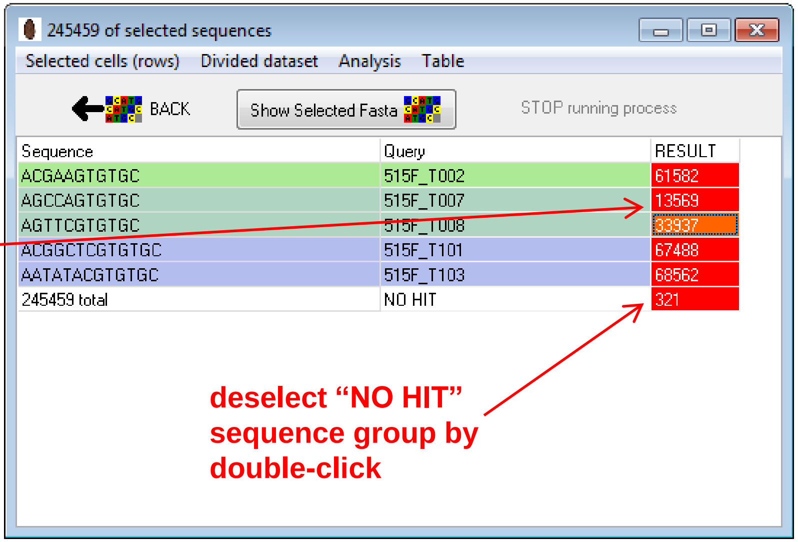
0 sequences selected

Sequence	Query	RESULT
ACGAAGTGTGC	515F_T002	61582
AGCCAGTGTGC	515F_T007	13569
AGTTCGTGTGC	515F_T008	33937
ACGGCTCGTGTGC	515F_T101	67488
AATATACGTGTGC	515F_T103	68562
245459 total	NO HIT	321

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

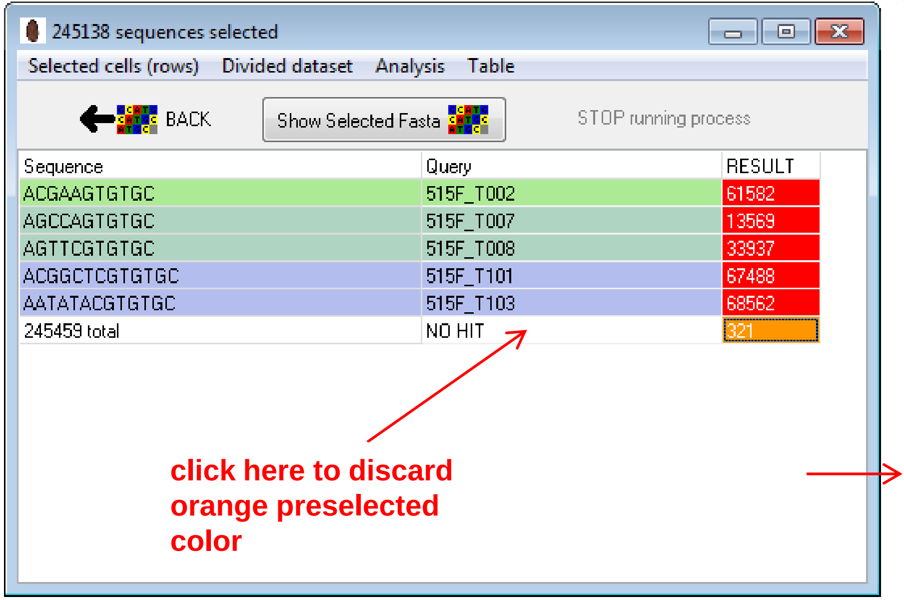


Sequence	Query	RESULT
ACGAAGTGTGC	515F_T002	61582
AGCCAGTGTGC	515F_T007	13569
AGTTCGTGTGC	515F_T008	33937
ACGGCTCGTGTGC	515F_T101	67488
AATATACGTGTGC	515F_T103	68562
245459 total	NO HIT	321



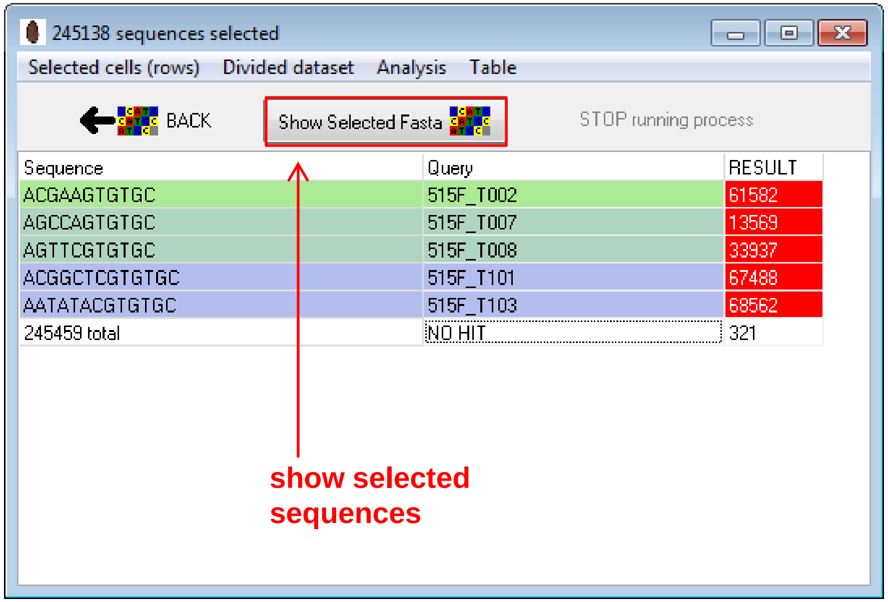
Sequence	Query	RESULT
ACGAAGTGTGC	515F_T002	61582
AGCCAGTGTGC	515F_T007	13569
AGTTCGTGTGC	515F_T008	33937
ACGGCTCGTGTGC	515F_T101	67488
AATATACGTGTGC	515F_T103	68562
245459 total	NO HIT	321

deselect "NO HIT" sequence group by double-click



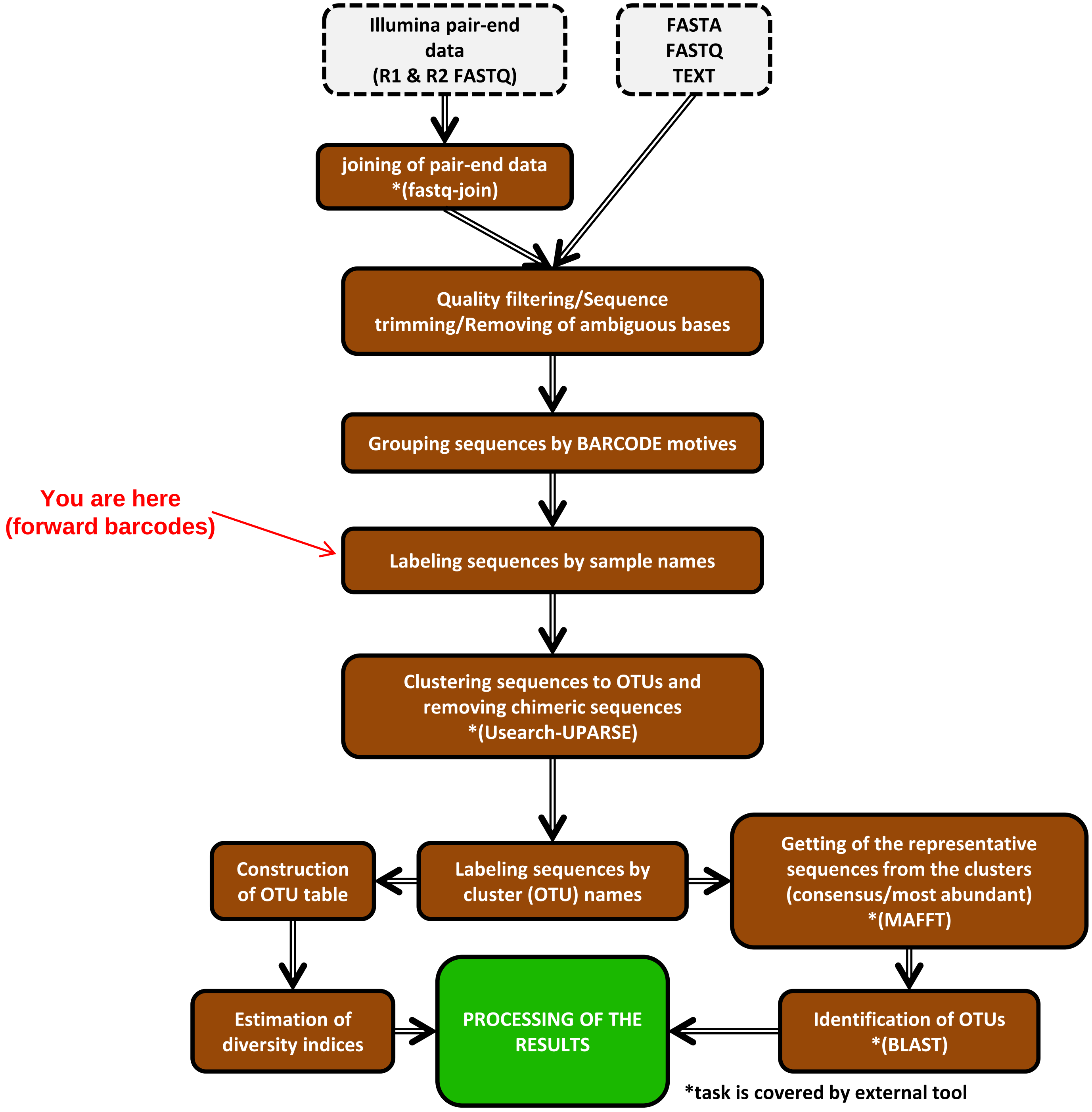
Sequence	Query	RESULT
ACGAAGTGTGC	515F_T002	61582
AGCCAGTGTGC	515F_T007	13569
AGTTCGTGTGC	515F_T008	33937
ACGGCTCGTGTGC	515F_T101	67488
AATATACGTGTGC	515F_T103	68562
245459 total	NO HIT	321

click here to discard orange preselected color



Sequence	Query	RESULT
ACGAAGTGTGC	515F_T002	61582
AGCCAGTGTGC	515F_T007	13569
AGTTCGTGTGC	515F_T008	33937
ACGGCTCGTGTGC	515F_T101	67488
AATATACGTGTGC	515F_T103	68562
245459 total	NO HIT	321

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

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)

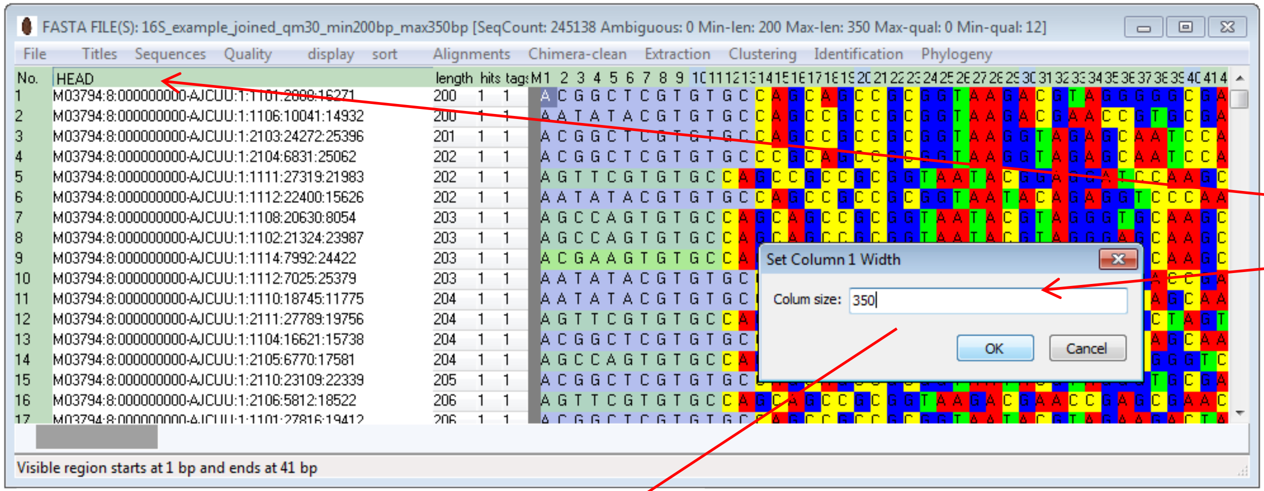
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool

Resize title column...



FASTA FILE(S): 16S_example_joined_qm30_min200bp_max350bp [SeqCount: 245138 Ambiguous: 0 Min-len: 200 Max-len: 350 Max-qual: 0 Min-qual: 12]

No.	HEAD	length	hits	tags
1	M03794:8:000000000-AJCUU:1:1101:2088:16271	200	1	1
2	M03794:8:000000000-AJCUU:1:1106:10041:14932	200	1	1
3	M03794:8:000000000-AJCUU:1:2103:24272:25396	201	1	1
4	M03794:8:000000000-AJCUU:1:2104:6831:25062	202	1	1
5	M03794:8:000000000-AJCUU:1:1111:27319:21983	202	1	1
6	M03794:8:000000000-AJCUU:1:1112:22400:15626	202	1	1
7	M03794:8:000000000-AJCUU:1:1108:20630:8054	203	1	1
8	M03794:8:000000000-AJCUU:1:1102:21324:23987	203	1	1
9	M03794:8:000000000-AJCUU:1:1114:7992:24422	203	1	1
10	M03794:8:000000000-AJCUU:1:1112:7025:25379	203	1	1
11	M03794:8:000000000-AJCUU:1:1110:18745:11775	204	1	1
12	M03794:8:000000000-AJCUU:1:2111:27789:19756	204	1	1
13	M03794:8:000000000-AJCUU:1:1104:16621:15738	204	1	1
14	M03794:8:000000000-AJCUU:1:2105:6770:17581	204	1	1
15	M03794:8:000000000-AJCUU:1:2110:23109:22339	205	1	1
16	M03794:8:000000000-AJCUU:1:2106:5812:18522	206	1	1
17	M03794:8:000000000-AJCUU:1:1101:27816:19412	206	1	1

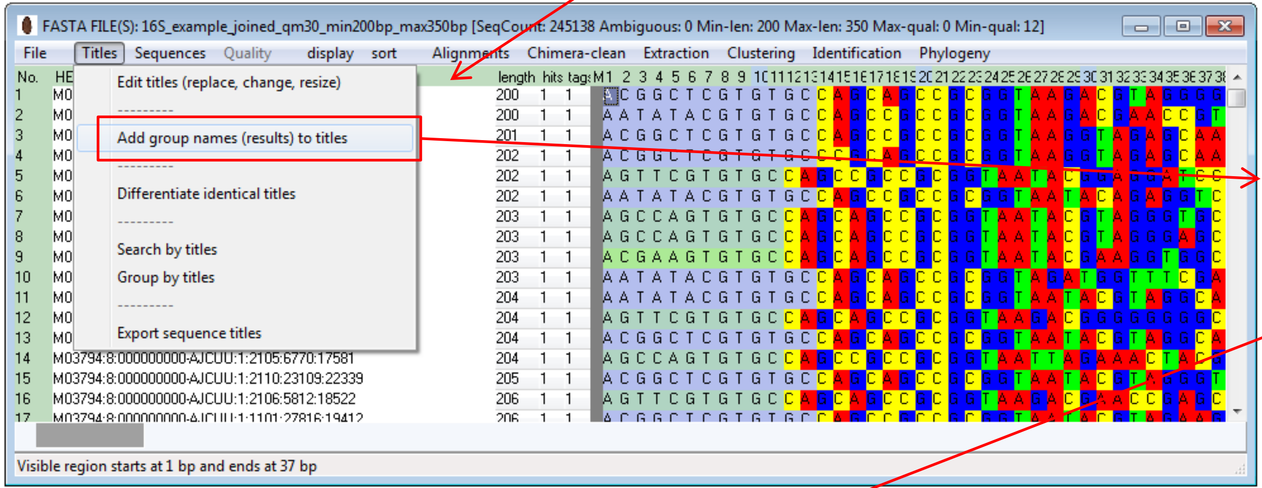
Visible region starts at 1 bp and ends at 41 bp

click here to set the width

Set Column 1 Width

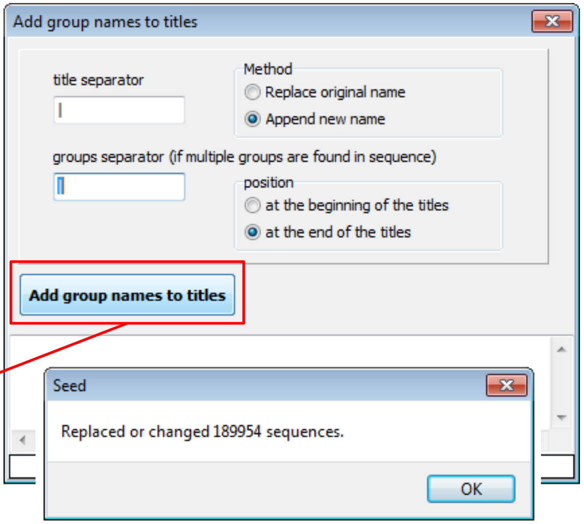
Colum size: 350

Add group (tag) name to title...



FASTA FILE(S): 16S_example_joined_qm30_min200bp_max350bp [SeqCount: 245138 Ambiguous: 0 Min-len: 200 Max-len: 350 Max-qual: 0 Min-qual: 12]

Visible region starts at 1 bp and ends at 37 bp



Add group names to titles

title separator: |

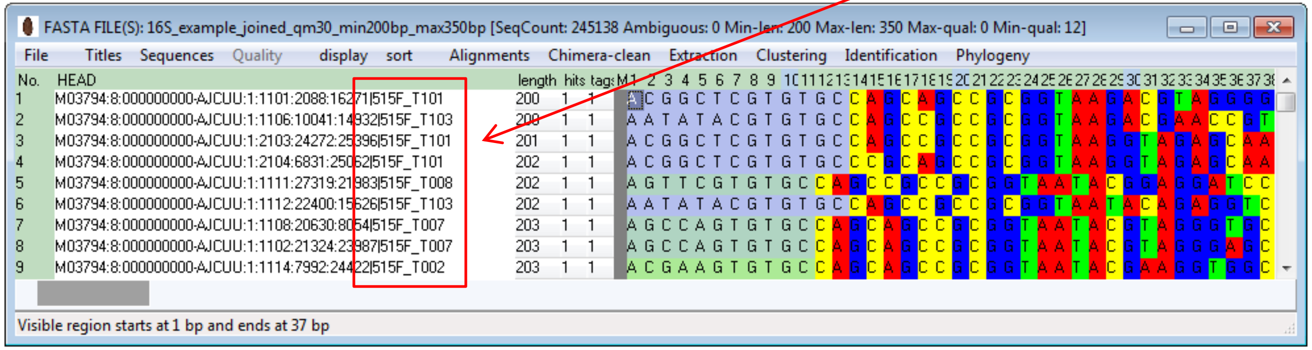
Method
○ Replace original name
● Append new name

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

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

Seed

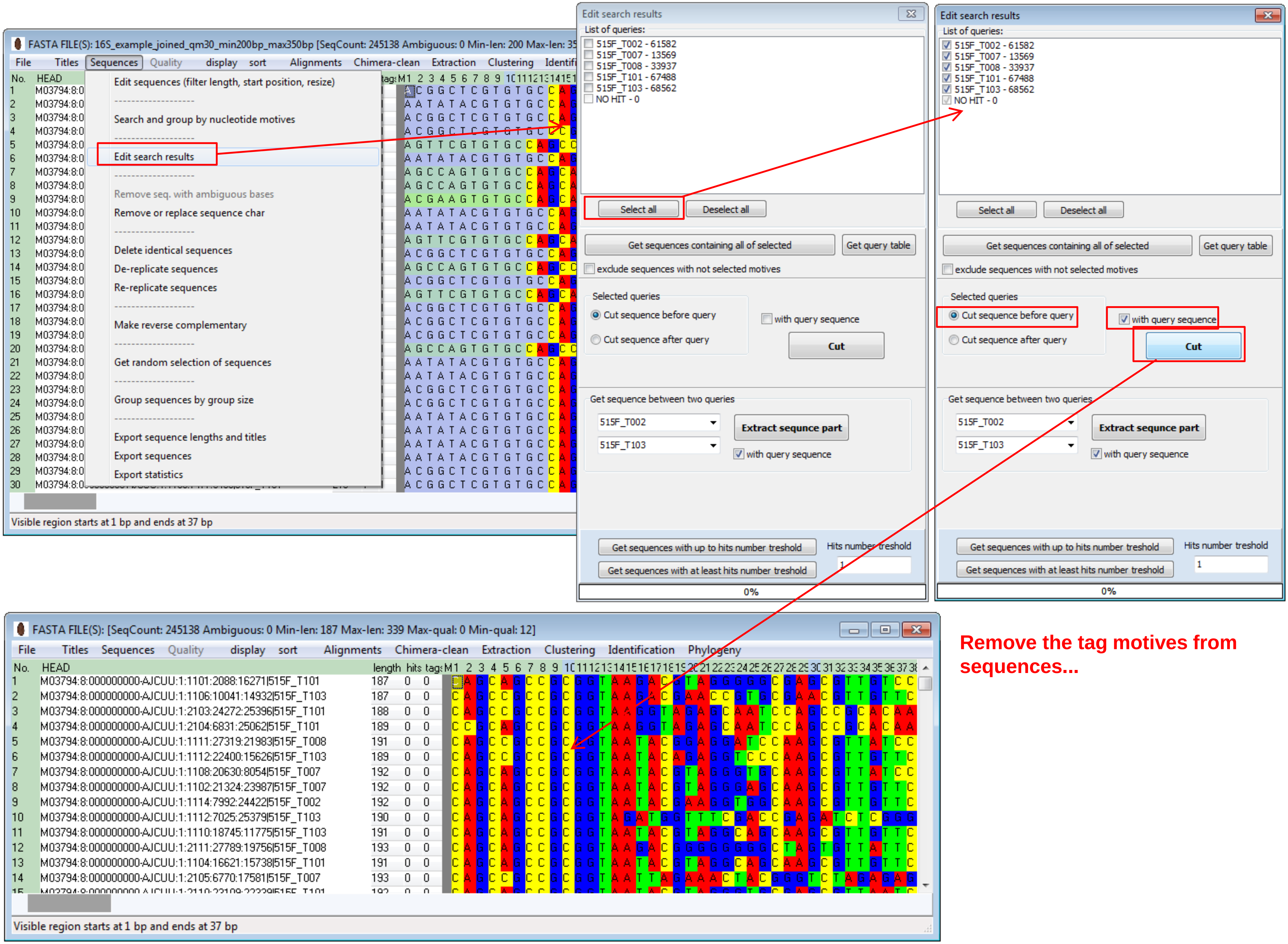
Replaced or changed 189954 sequences.



FASTA FILE(S): 16S_example_joined_qm30_min200bp_max350bp [SeqCount: 245138 Ambiguous: 0 Min-len: 200 Max-len: 350 Max-qual: 0 Min-qual: 12]

No.	HEAD	length	hits	tags
1	M03794:8:000000000-AJCUU:1:1101:2088:16271|515F_T101	200	1	1
2	M03794:8:000000000-AJCUU:1:1106:10041:14932|515F_T103	200	1	1
3	M03794:8:000000000-AJCUU:1:2103:24272:25396|515F_T101	201	1	1
4	M03794:8:000000000-AJCUU:1:2104:6831:25062|515F_T101	202	1	1
5	M03794:8:000000000-AJCUU:1:1111:27319:21983|515F_T008	202	1	1
6	M03794:8:000000000-AJCUU:1:1112:22400:15626|515F_T103	202	1	1
7	M03794:8:000000000-AJCUU:1:1108:20630:8054|515F_T007	203	1	1
8	M03794:8:000000000-AJCUU:1:1102:21324:23987|515F_T007	203	1	1
9	M03794:8:000000000-AJCUU:1:1114:7992:24422|515F_T002	203	1	1

Visible region starts at 1 bp and ends at 37 bp



No.	HEAD	length	hits	tags
1	M03794:8:000000000-AJCUU:1:1101:2088:16271|515F_T101	187	0	0
2	M03794:8:000000000-AJCUU:1:1106:10041:14932|515F_T103	187	0	0
3	M03794:8:000000000-AJCUU:1:2103:24272:25396|515F_T101	188	0	0
4	M03794:8:000000000-AJCUU:1:2104:6831:25062|515F_T101	189	0	0
5	M03794:8:000000000-AJCUU:1:1111:27319:21983|515F_T008	191	0	0
6	M03794:8:000000000-AJCUU:1:1112:22400:15626|515F_T103	189	0	0
7	M03794:8:000000000-AJCUU:1:1108:20630:8054|515F_T007	192	0	0
8	M03794:8:000000000-AJCUU:1:1102:21324:23987|515F_T007	192	0	0
9	M03794:8:000000000-AJCUU:1:1114:7992:24422|515F_T002	192	0	0
10	M03794:8:000000000-AJCUU:1:1112:7025:25379|515F_T103	190	0	0
11	M03794:8:000000000-AJCUU:1:1110:18745:11775|515F_T103	191	0	0
12	M03794:8:000000000-AJCUU:1:2111:27789:19756|515F_T008	193	0	0
13	M03794:8:000000000-AJCUU:1:1104:16621:15738|515F_T101	191	0	0
14	M03794:8:000000000-AJCUU:1:2105:6770:17581|515F_T007	193	0	0

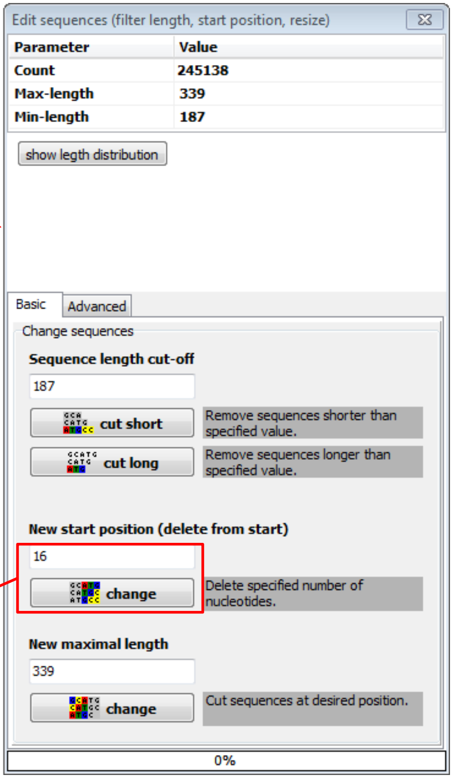
Remove the tag motives from sequences...

Remove the rest of the forward primer sequence...

FASTA FILE(S): [SeqCount: 245138 Ambiguous: 0 Min-len: 187 Max-len: 339 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tag
31	M03794:8:000000000-AJCUU:1:2109:15529:25270|515F_T103	197	0	0

Visible region starts at 1 bp and ends at 29 bp



Edit sequences (filter length, start position, resize)

Parameter	Value
Count	245138
Max-length	339
Min-length	187

Change sequences

Sequence length cut-off: 187

Remove sequences shorter than specified value.

Remove sequences longer than specified value.

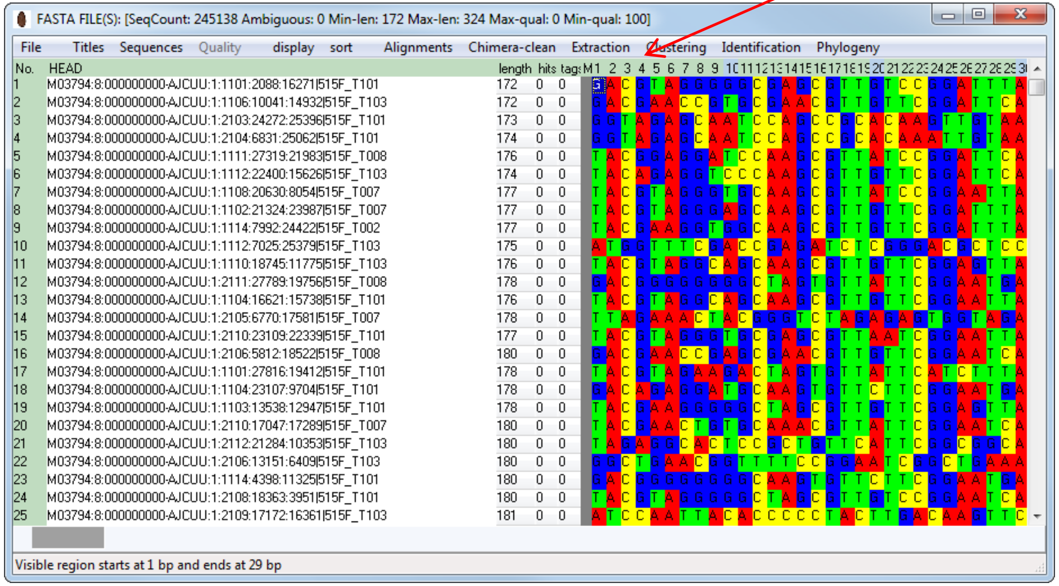
New start position (delete from start): 16

Delete specified number of nucleotides.

New maximal length: 339

Cut sequences at desired position.

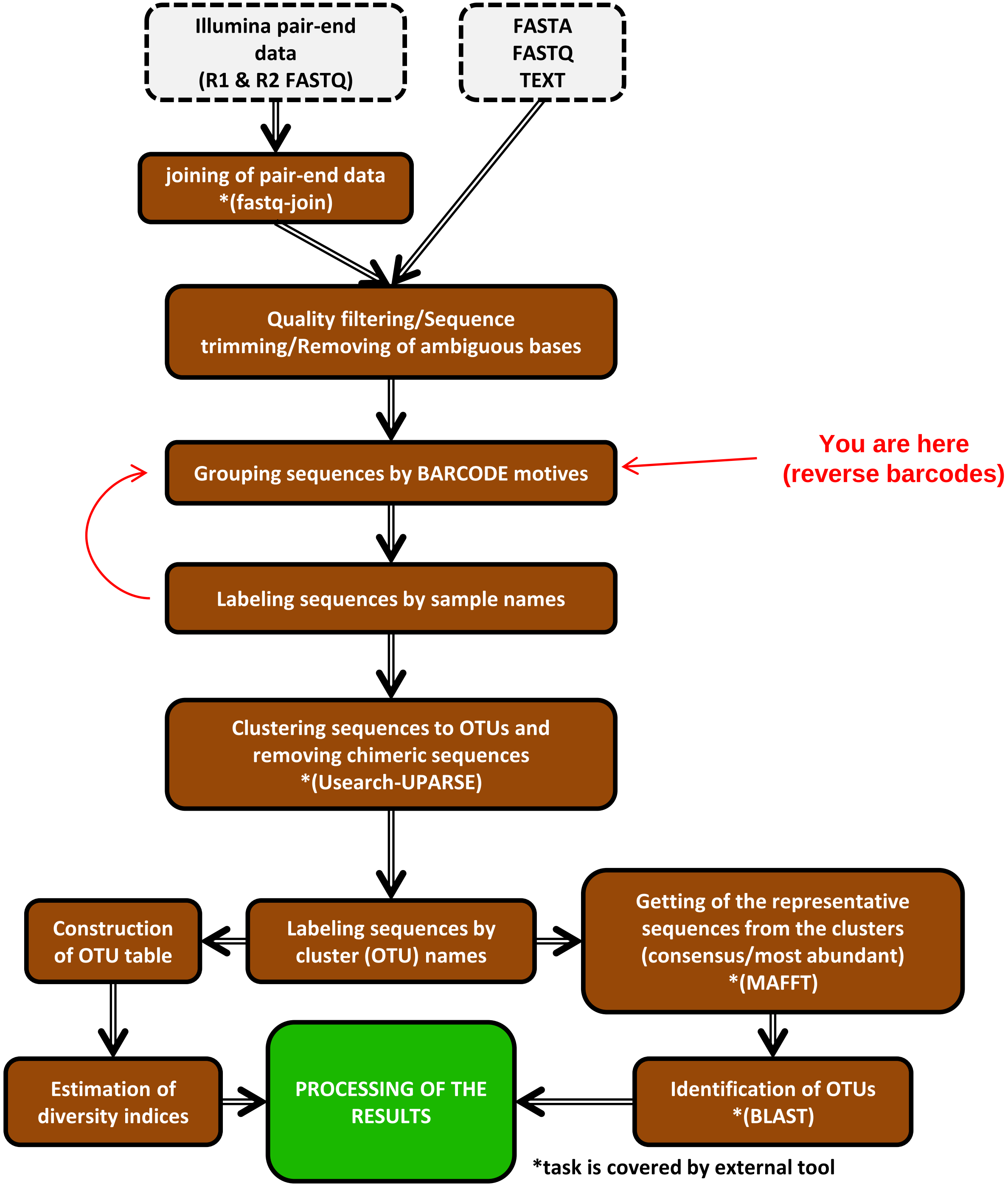
0%



FASTA FILE(S): [SeqCount: 245138 Ambiguous: 0 Min-len: 172 Max-len: 324 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tag
1	M03794:8:000000000-AJCUU:1:1101:2088:16271|515F_T101	172	0	0
2	M03794:8:000000000-AJCUU:1:1106:10041:14932|515F_T103	172	0	0
3	M03794:8:000000000-AJCUU:1:2103:24272:25396|515F_T101	173	0	0
4	M03794:8:000000000-AJCUU:1:2104:6831:25062|515F_T101	174	0	0
5	M03794:8:000000000-AJCUU:1:1111:27319:21983|515F_T008	176	0	0
6	M03794:8:000000000-AJCUU:1:1112:22400:15626|515F_T103	174	0	0
7	M03794:8:000000000-AJCUU:1:1108:20630:8054|515F_T007	177	0	0
8	M03794:8:000000000-AJCUU:1:1102:21324:23987|515F_T007	177	0	0
9	M03794:8:000000000-AJCUU:1:1114:7992:24422|515F_T002	177	0	0
10	M03794:8:000000000-AJCUU:1:1112:7025:25379|515F_T103	175	0	0
11	M03794:8:000000000-AJCUU:1:1110:18745:11775|515F_T103	176	0	0
12	M03794:8:000000000-AJCUU:1:2111:27789:19756|515F_T008	178	0	0
13	M03794:8:000000000-AJCUU:1:1104:16621:15738|515F_T101	176	0	0
14	M03794:8:000000000-AJCUU:1:2105:6770:17581|515F_T007	178	0	0
15	M03794:8:000000000-AJCUU:1:2110:23109:22339|515F_T101	177	0	0
16	M03794:8:000000000-AJCUU:1:2106:5812:18522|515F_T008	180	0	0
17	M03794:8:000000000-AJCUU:1:1101:27816:19412|515F_T101	178	0	0
18	M03794:8:000000000-AJCUU:1:1104:23107:9704|515F_T101	178	0	0
19	M03794:8:000000000-AJCUU:1:1103:13538:12947|515F_T101	178	0	0
20	M03794:8:000000000-AJCUU:1:2110:17047:17289|515F_T007	180	0	0
21	M03794:8:000000000-AJCUU:1:2112:21284:10353|515F_T103	180	0	0
22	M03794:8:000000000-AJCUU:1:2106:13151:6409|515F_T103	180	0	0
23	M03794:8:000000000-AJCUU:1:1114:4398:11325|515F_T101	180	0	0
24	M03794:8:000000000-AJCUU:1:2108:18363:3951|515F_T101	180	0	0
25	M03794:8:000000000-AJCUU:1:2109:17172:16361|515F_T103	181	0	0

Visible region starts at 1 bp and ends at 29 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

You are here (reverse barcodes)

Labeling sequences by sample names

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

Construction of OTU table

Labeling sequences by cluster (OTU) names

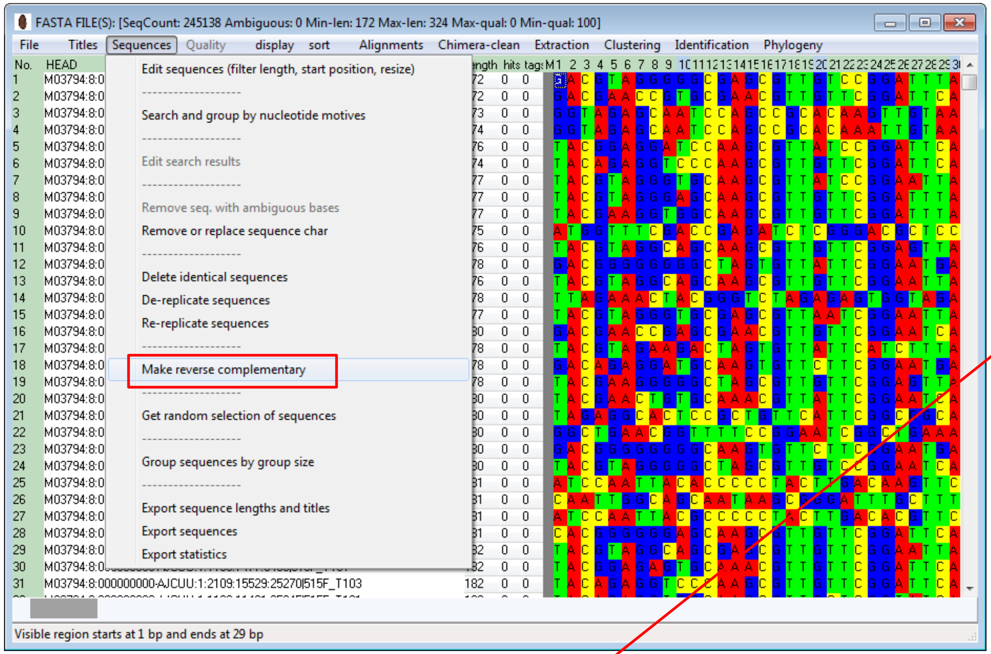
Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

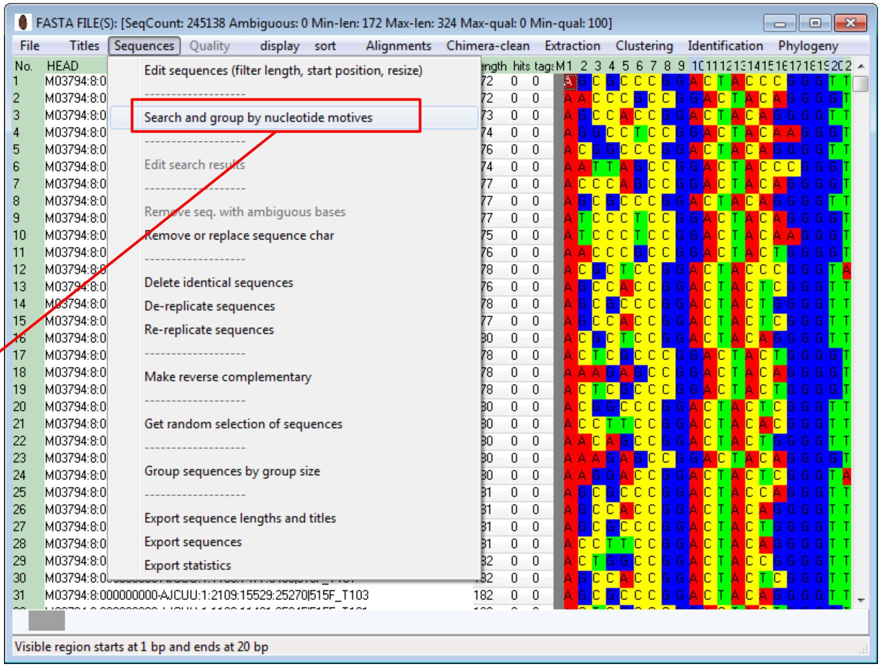
Estimation of diversity indices

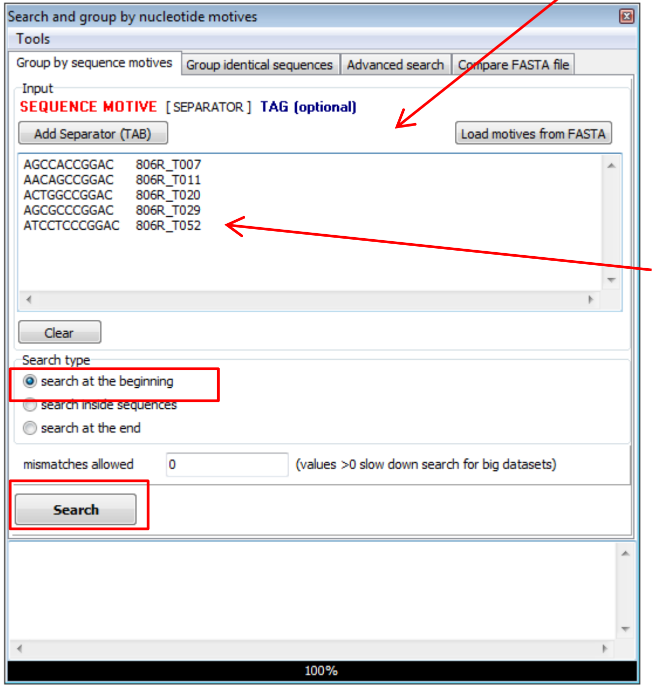
PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool







```
AGCCACCGGAC    806R_T007
AACAGCCGGAC    806R_T011
ACTGGCCGGAC    806R_T020
AGCGCCCGGAC    806R_T029
ATCCTCCCGGAC   806R_T052
```

● 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...

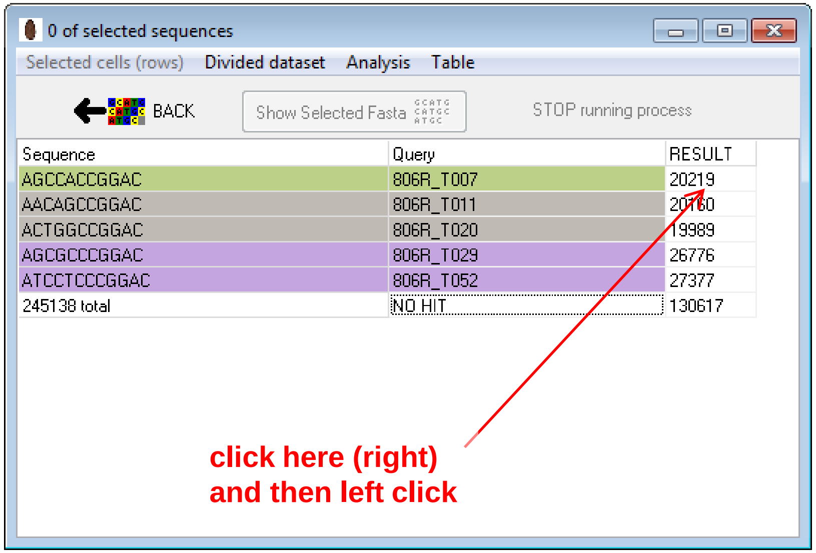
AGCCACCGGAC	806R_T007
AACAGCCGGAC	806R_T011
ACTGGCCGGAC	806R_T020
AGCGCCCGGAC	806R_T029
ATCCTCCCGGAC	806R_T052

reverse primer

806R	GGACTACHVGGGTWTCTAAT

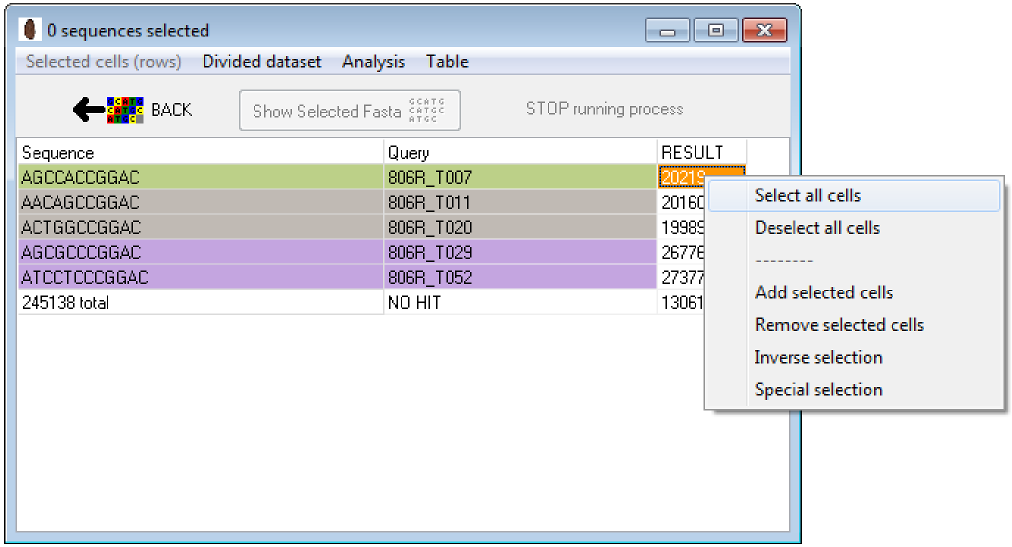
tagged reverse primers

806R_T007	**AGCCACC**GGACTACHVGGGTWTCTAAT
806R_T011	**AACAGCC**GGACTACHVGGGTWTCTAAT
806R_T020	**ACTGGCC**GGACTACHVGGGTWTCTAAT
806R_T029	**AGCGCCC**GGACTACHVGGGTWTCTAAT
806R_T052	**ATCCTCCC**GGACTACHVGGGTWTCTAAT

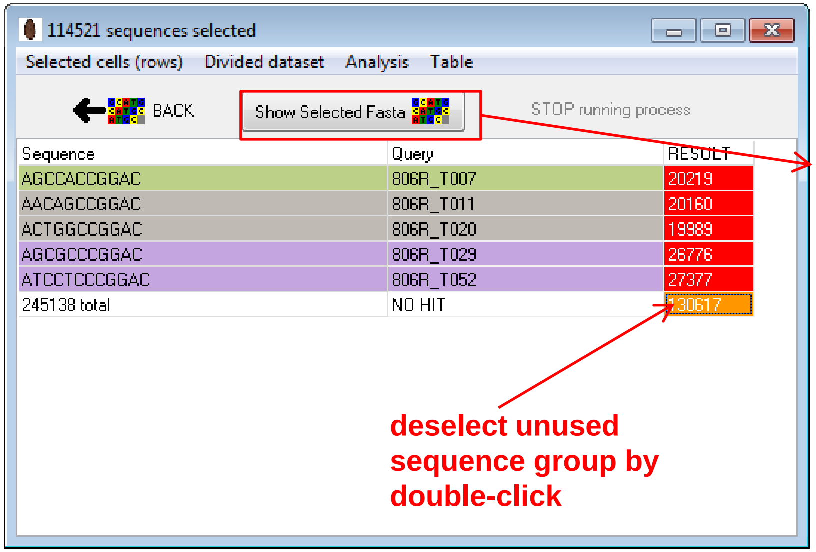


Sequence	Query	RESULT
AGCCACCGGAC	806R_T007	20219
AACAGCCGGAC	806R_T011	20160
ACTGGCCGGAC	806R_T020	19989
AGCGCCCGGAC	806R_T029	26776
ATCCTCCCGGAC	806R_T052	27377
245138 total	NO HIT	130617

click here (right) and then left click

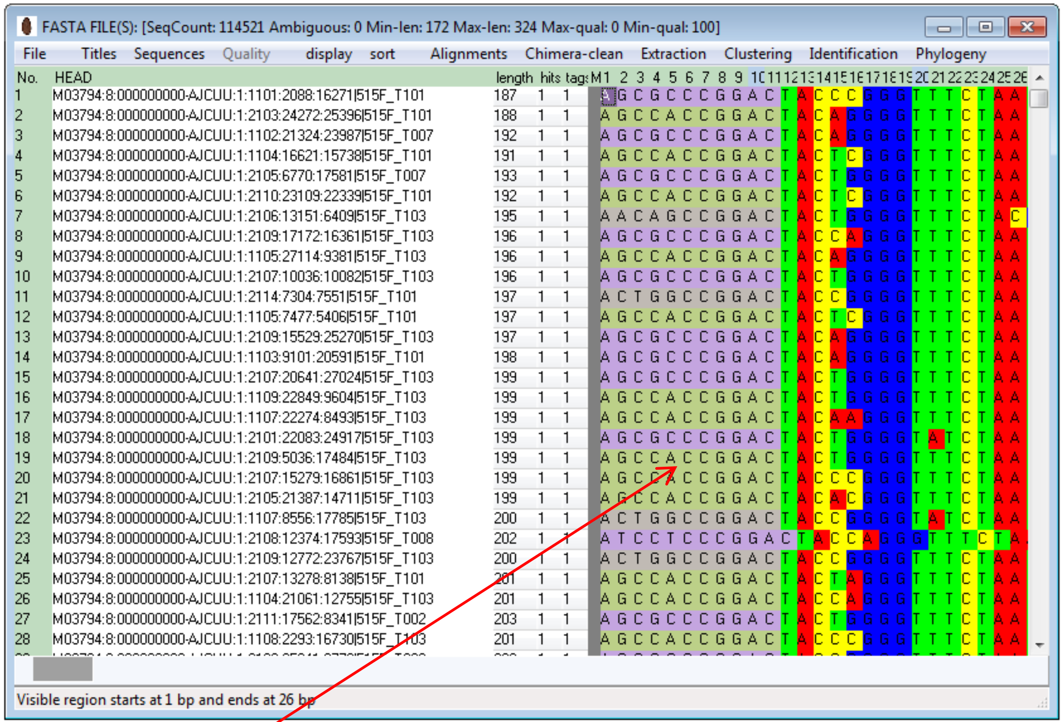


Sequence	Query	RESULT
AGCCACCGGAC	806R_T007	20219
AACAGCCGGAC	806R_T011	20160
ACTGGCCGGAC	806R_T020	19989
AGCGCCCGGAC	806R_T029	26776
ATCCTCCCGGAC	806R_T052	27377
245138 total	NO HIT	13061



Sequence	Query	RESULT
AGCCACCGGAC	806R_T007	20219
AACAGCCGGAC	806R_T011	20160
ACTGGCCGGAC	806R_T020	19989
AGCGCCCGGAC	806R_T029	26776
ATCCTCCCGGAC	806R_T052	27377
245138 total	NO HIT	130617

deselect unused sequence group by double-click

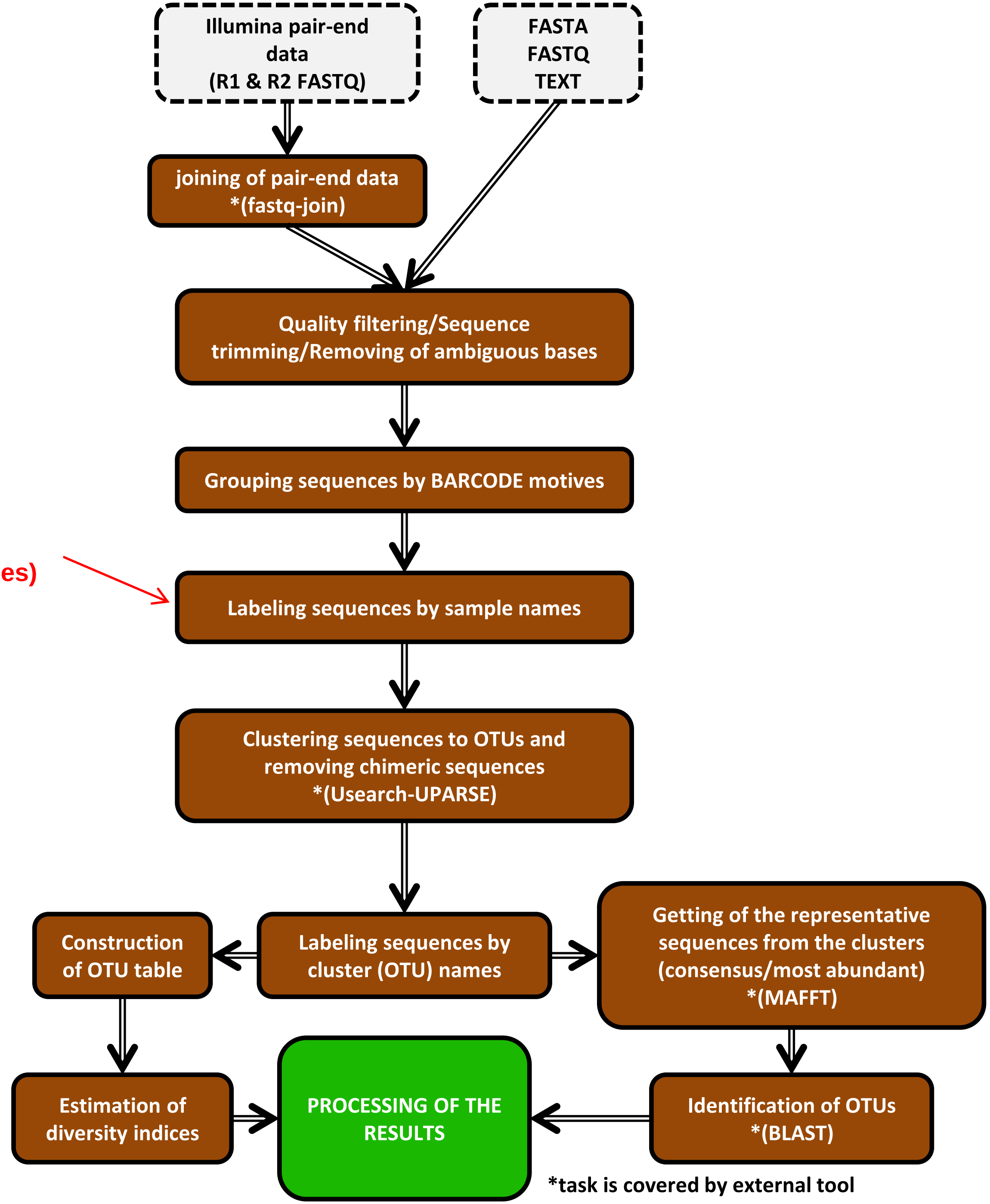


FASTA FILE(S): [SeqCount: 114521 Ambiguous: 0 Min-len: 172 Max-len: 324 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tag:	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
1	M03794:8:000000000-AJCUU:1:1101:2088:16271|515F_T101	187	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	C	C	G	G	G	T	T	T	C	T	A	A
2	M03794:8:000000000-AJCUU:1:2103:24272:25396|515F_T101	188	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	A	G	G	G	G	T	T	T	C	T	A	A
3	M03794:8:000000000-AJCUU:1:1102:21324:23987|515F_T007	192	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	A	G	G	G	G	T	T	T	C	T	A	A
4	M03794:8:000000000-AJCUU:1:1104:16621:15738|515F_T101	191	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	T	C	G	G	G	T	T	T	C	T	A	A
5	M03794:8:000000000-AJCUU:1:2105:6770:17581|515F_T007	193	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A
6	M03794:8:000000000-AJCUU:1:2110:23109:22339|515F_T101	192	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	T	C	G	G	G	T	T	T	C	T	A	A
7	M03794:8:000000000-AJCUU:1:2106:13151:6409|515F_T103	195	1	1		A	A	C	A	G	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	C
8	M03794:8:000000000-AJCUU:1:2109:17172:16361|515F_T103	196	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	C	A	G	G	G	T	T	T	C	T	A	A
9	M03794:8:000000000-AJCUU:1:1105:27114:9381|515F_T103	196	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	A	G	G	G	G	T	T	T	C	T	A	A
10	M03794:8:000000000-AJCUU:1:2107:10036:10082|515F_T103	196	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A
11	M03794:8:000000000-AJCUU:1:2114:7304:7551|515F_T101	197	1	1		A	C	T	G	G	C	C	G	G	A	C	T	A	C	C	G	G	G	G	T	T	T	C	T	A	A
12	M03794:8:000000000-AJCUU:1:1105:7477:5406|515F_T101	197	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	T	C	G	G	G	T	T	T	C	T	A	A
13	M03794:8:000000000-AJCUU:1:2109:15529:25270|515F_T103	197	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	A	G	G	G	G	T	T	T	C	T	A	A
14	M03794:8:000000000-AJCUU:1:1103:9101:20591|515F_T101	198	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	A	G	G	G	G	T	T	T	C	T	A	A
15	M03794:8:000000000-AJCUU:1:2107:20641:27024|515F_T103	199	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A
16	M03794:8:000000000-AJCUU:1:1109:22849:9604|515F_T103	199	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A
17	M03794:8:000000000-AJCUU:1:1107:22274:8493|515F_T103	199	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	A	A	G	G	G	T	T	T	C	T	A	A
18	M03794:8:000000000-AJCUU:1:2101:22083:24917|515F_T103	199	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	A	T	C	T	A	A
19	M03794:8:000000000-AJCUU:1:2109:5036:17484|515F_T103	199	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A
20	M03794:8:000000000-AJCUU:1:2107:15279:16861|515F_T103	199	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	C	C	G	G	G	T	T	T	C	T	A	A
21	M03794:8:000000000-AJCUU:1:2105:21387:14711|515F_T103	199	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	A	C	G	G	G	T	T	T	C	T	A	A
22	M03794:8:000000000-AJCUU:1:1107:8556:17785|515F_T103	200	1	1		A	C	T	G	G	C	C	G	G	A	C	T	A	C	C	G	G	G	G	T	A	T	C	T	A	A
23	M03794:8:000000000-AJCUU:1:2108:12374:17593|515F_T008	202	1	1		A	T	C	C	T	C	C	C	G	G	A	C	T	A	C	C	A	G	G	G	T	T	T	C	T	A
24	M03794:8:000000000-AJCUU:1:2109:12772:23767|515F_T103	200	1	1		A	C	T	G	G	C	C	G	G	A	C	T	A	C	C	G	G	G	G	T	T	T	C	T	A	A
25	M03794:8:000000000-AJCUU:1:2107:13278:8138|515F_T101	201	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	T	A	G	G	G	T	T	T	C	T	A	A
26	M03794:8:000000000-AJCUU:1:1104:21061:12755|515F_T103	201	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	C	A	G	G	G	T	T	T	C	T	A	A
27	M03794:8:000000000-AJCUU:1:2111:17562:8341|515F_T002	203	1	1		A	G	C	G	C	C	C	G	G	A	C	T	A	C	T	G	G	G	G	T	T	T	C	T	A	A
28	M03794:8:000000000-AJCUU:1:1108:2293:16730|515F_T103	201	1	1		A	G	C	C	A	C	C	G	G	A	C	T	A	C	C	C	G	G	G	T	T	T	C	T	A	A

Visible region starts at 1 bp and ends at 26 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

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

Construction of OTU table

Labeling sequences by cluster (OTU) names

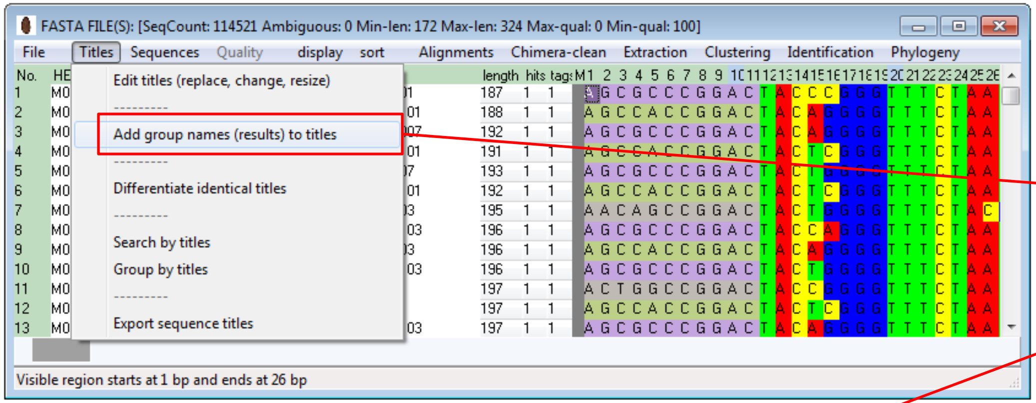
Getting of the representative sequences from the clusters (consensus/most abundant) *(MAFFT)

Estimation of diversity indices

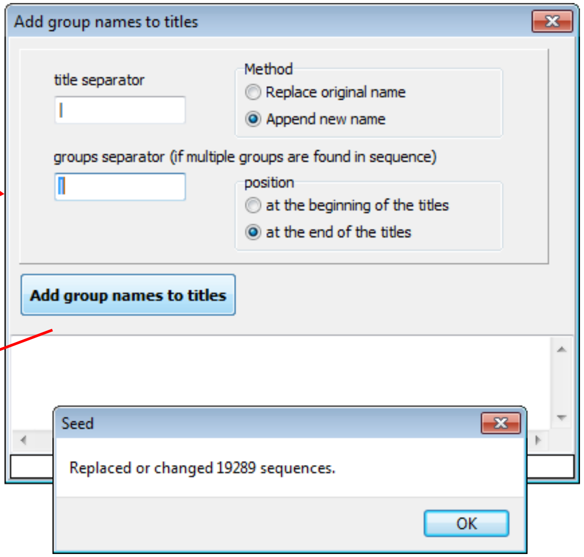
PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool


FASTA FILE(S): [SeqCount: 114521 Ambiguous: 0 Min-len: 172 Max-len: 324 Max-qual: 0 Min-qual: 100]

Visible region starts at 1 bp and ends at 26 bp


Add group names to titles

title separator: |

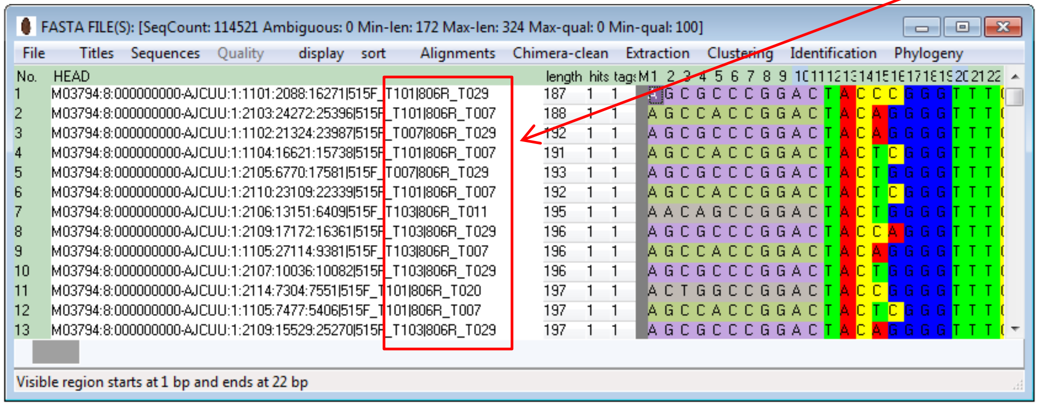
Method
○ Replace original name
● Append new name

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

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

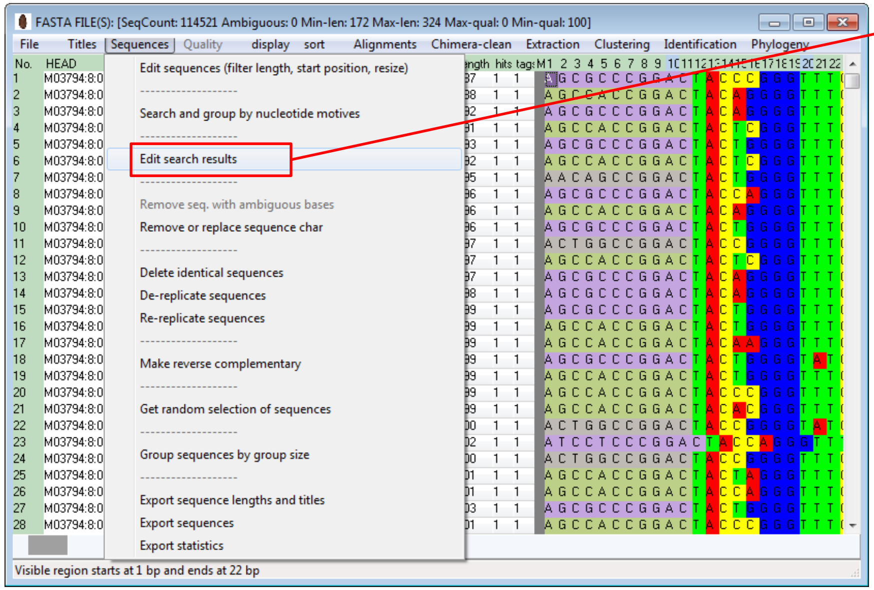
Seed

Replaced or changed 19289 sequences.

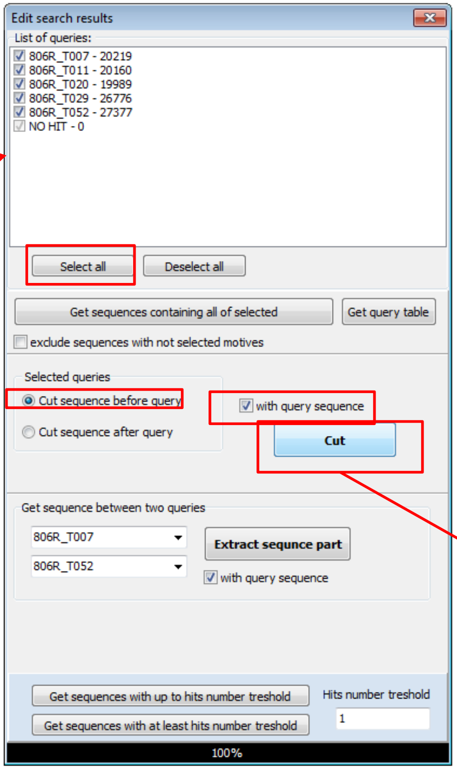

FASTA FILE(S): [SeqCount: 114521 Ambiguous: 0 Min-len: 172 Max-len: 324 Max-qual: 0 Min-qual: 100]

```
No. HEAD
1   M03794:8:000000000-AJCUU:1:1101:2088:16271|515F_T101|806R_T029
2   M03794:8:000000000-AJCUU:1:2103:24272:25396|515F_T101|806R_T007
3   M03794:8:000000000-AJCUU:1:1102:21324:23987|515F_T007|806R_T029
4   M03794:8:000000000-AJCUU:1:1104:16621:15738|515F_T101|806R_T007
5   M03794:8:000000000-AJCUU:1:2105:6770:17581|515F_T007|806R_T029
6   M03794:8:000000000-AJCUU:1:2110:23109:22339|515F_T101|806R_T007
7   M03794:8:000000000-AJCUU:1:2106:13151:6409|515F_T103|806R_T011
8   M03794:8:000000000-AJCUU:1:2109:17172:16361|515F_T103|806R_T029
9   M03794:8:000000000-AJCUU:1:1105:27114:9381|515F_T103|806R_T007
10  M03794:8:000000000-AJCUU:1:2107:10036:10082|515F_T103|806R_T029
11  M03794:8:000000000-AJCUU:1:2114:7304:7551|515F_T101|806R_T020
12  M03794:8:000000000-AJCUU:1:1105:7477:5406|515F_T101|806R_T007
13  M03794:8:000000000-AJCUU:1:2109:15529:25270|515F_T103|806R_T029
```

Visible region starts at 1 bp and ends at 22 bp


FASTA FILE(S): [SeqCount: 114521 Ambiguous: 0 Min-len: 172 Max-len: 324 Max-qual: 0 Min-qual: 100]

Visible region starts at 1 bp and ends at 22 bp


Edit search results

List of queries:
☑ 806R_T007 - 20219
☑ 806R_T011 - 20160
☑ 806R_T020 - 19989
☑ 806R_T029 - 26776
☑ 806R_T052 - 27377
☑ NO HIT - 0

☐ exclude sequences with not selected motives

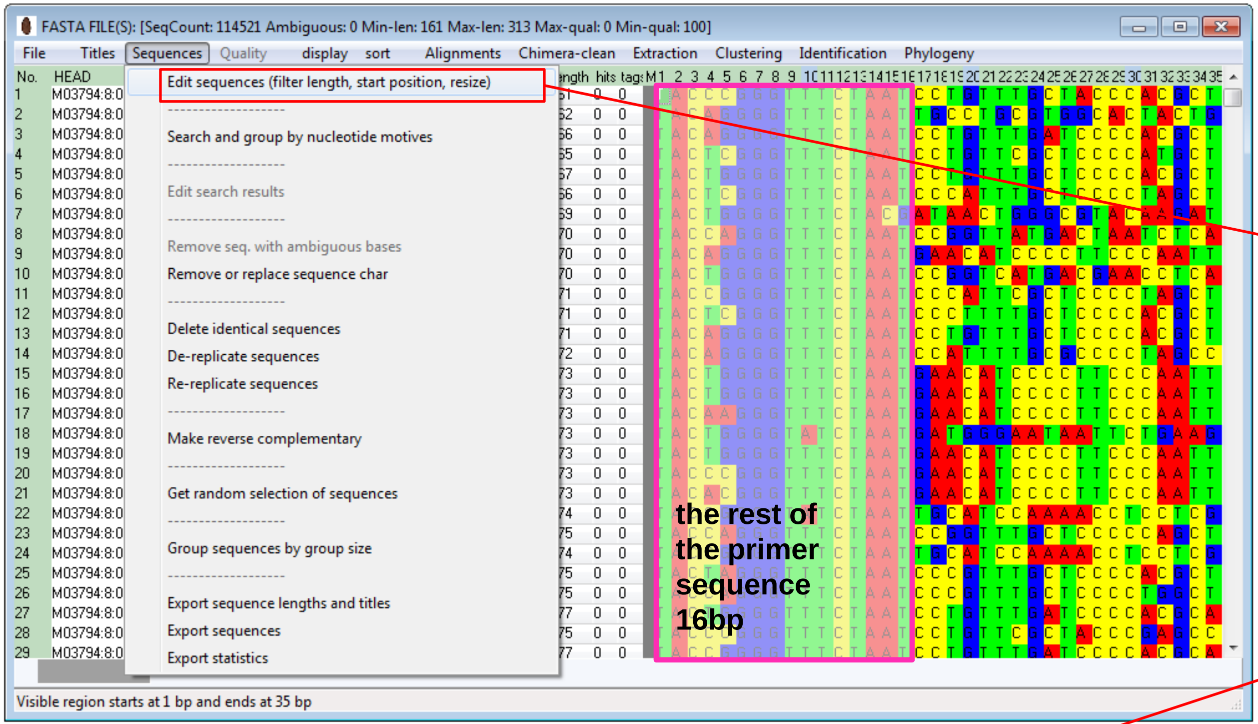
Selected queries
● Cut sequence before query
○ Cut sequence after query
☑ with query sequence

Get sequence between two queries
806R_T007
806R_T052
☑ with query sequence

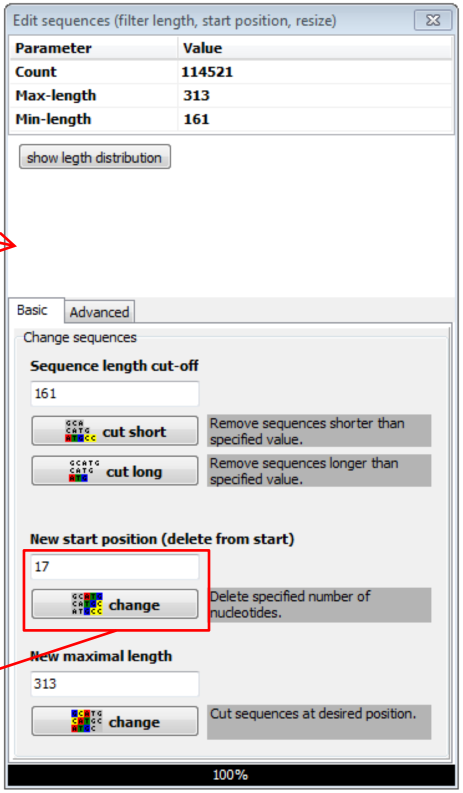
Hits number treshold: 1

Remove the tag motives from sequences...

Remove the rest of the reverse primer sequence...


FASTA FILE(S): [SeqCount: 114521 Ambiguous: 0 Min-len: 161 Max-len: 313 Max-qual: 0 Min-qual: 100]

Visible region starts at 1 bp and ends at 35 bp


Edit sequences (filter length, start position, resize)

Parameter	Value
Count	114521
Max-length	313
Min-length	161

Sequence length cut-off

161

Remove sequences shorter than specified value.

Remove sequences longer than specified value.

New start position (delete from start)

17

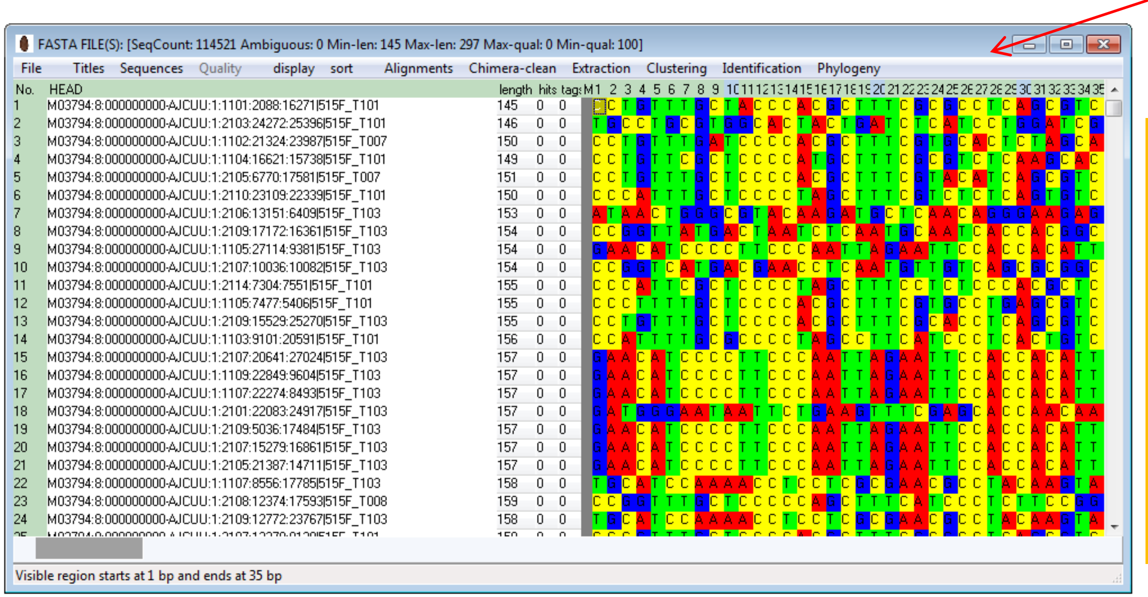
Delete specified number of nucleotides.

New maximal length

313

Cut sequences at desired position.

100%


FASTA FILE(S): [SeqCount: 114521 Ambiguous: 0 Min-len: 145 Max-len: 297 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tags
1	M03794:8:000000000-AJCUU:1:1101:2088:16271|515F_T101	145	0	0
2	M03794:8:000000000-AJCUU:1:2103:24272:25396|515F_T101	146	0	0
3	M03794:8:000000000-AJCUU:1:1102:21324:23987|515F_T007	150	0	0
4	M03794:8:000000000-AJCUU:1:1104:16621:15738|515F_T101	149	0	0
5	M03794:8:000000000-AJCUU:1:2105:6770:17581|515F_T007	151	0	0
6	M03794:8:000000000-AJCUU:1:2110:23109:22339|515F_T101	150	0	0
7	M03794:8:000000000-AJCUU:1:2106:13151:6409|515F_T103	153	0	0
8	M03794:8:000000000-AJCUU:1:2109:17172:16361|515F_T103	154	0	0
9	M03794:8:000000000-AJCUU:1:1105:27114:9381|515F_T103	154	0	0
10	M03794:8:000000000-AJCUU:1:2107:10036:10082|515F_T103	154	0	0
11	M03794:8:000000000-AJCUU:1:2114:7304:7551|515F_T101	155	0	0
12	M03794:8:000000000-AJCUU:1:1105:7477:5406|515F_T101	155	0	0
13	M03794:8:000000000-AJCUU:1:2109:15529:25270|515F_T103	155	0	0
14	M03794:8:000000000-AJCUU:1:1103:9101:20591|515F_T101	156	0	0
15	M03794:8:000000000-AJCUU:1:2107:20641:27024|515F_T103	157	0	0
16	M03794:8:000000000-AJCUU:1:1109:22849:9604|515F_T103	157	0	0
17	M03794:8:000000000-AJCUU:1:1107:22274:8493|515F_T103	157	0	0
18	M03794:8:000000000-AJCUU:1:2101:22083:24917|515F_T103	157	0	0
19	M03794:8:000000000-AJCUU:1:2109:5036:17484|515F_T103	157	0	0
20	M03794:8:000000000-AJCUU:1:2107:15279:16861|515F_T103	157	0	0
21	M03794:8:000000000-AJCUU:1:2105:21387:14711|515F_T103	157	0	0
22	M03794:8:000000000-AJCUU:1:1107:8556:17785|515F_T103	158	0	0
23	M03794:8:000000000-AJCUU:1:2108:12374:17593|515F_T008	159	0	0
24	M03794:8:000000000-AJCUU:1:2109:12772:23767|515F_T103	158	0	0

Visible region starts at 1 bp and ends at 35 bp

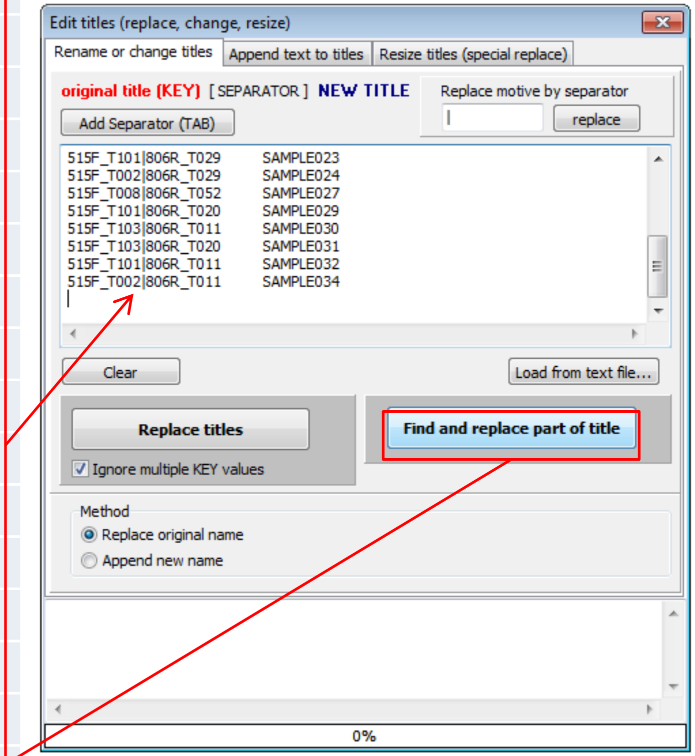
NOTE: Now, you can also remove too short or long sequences (see page 8) to remove potential plastid, mitochondrial or other contaminants


GOOD TO SAVE NOW!

16S_example_joined_qm30_min200bp_max350bp_fwdtag_revtag

Replace tag names by sample name…

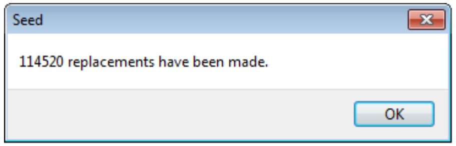
name	FW Primer	REV Primer		
SAMPLE001	515F_T103	806R_T007	515F_T103|806R_T007	SAMPLE001
SAMPLE002	515F_T002	806R_T052	515F_T002|806R_T052	SAMPLE002
SAMPLE005	515F_T007	806R_T029	515F_T007|806R_T029	SAMPLE005
SAMPLE006	515F_T101	806R_T052	515F_T101|806R_T052	SAMPLE006
SAMPLE010	515F_T101	806R_T007	515F_T101|806R_T007	SAMPLE010
SAMPLE012	515F_T103	806R_T029	515F_T103|806R_T029	SAMPLE012
SAMPLE019	515F_T002	806R_T020	515F_T002|806R_T020	SAMPLE019
SAMPLE020	515F_T103	806R_T052	515F_T103|806R_T052	SAMPLE020
SAMPLE021	515F_T002	806R_T007	515F_T002|806R_T007	SAMPLE021
SAMPLE023	515F_T101	806R_T029	515F_T101|806R_T029	SAMPLE023
SAMPLE024	515F_T002	806R_T029	515F_T002|806R_T029	SAMPLE024
SAMPLE027	515F_T008	806R_T052	515F_T008|806R_T052	SAMPLE027
SAMPLE029	515F_T101	806R_T020	515F_T101|806R_T020	SAMPLE029
SAMPLE030	515F_T103	806R_T011	515F_T103|806R_T011	SAMPLE030
SAMPLE031	515F_T103	806R_T020	515F_T103|806R_T020	SAMPLE031
SAMPLE032	515F_T101	806R_T011	515F_T101|806R_T011	SAMPLE032
SAMPLE034	515F_T002	806R_T011	515F_T002|806R_T011	SAMPLE034



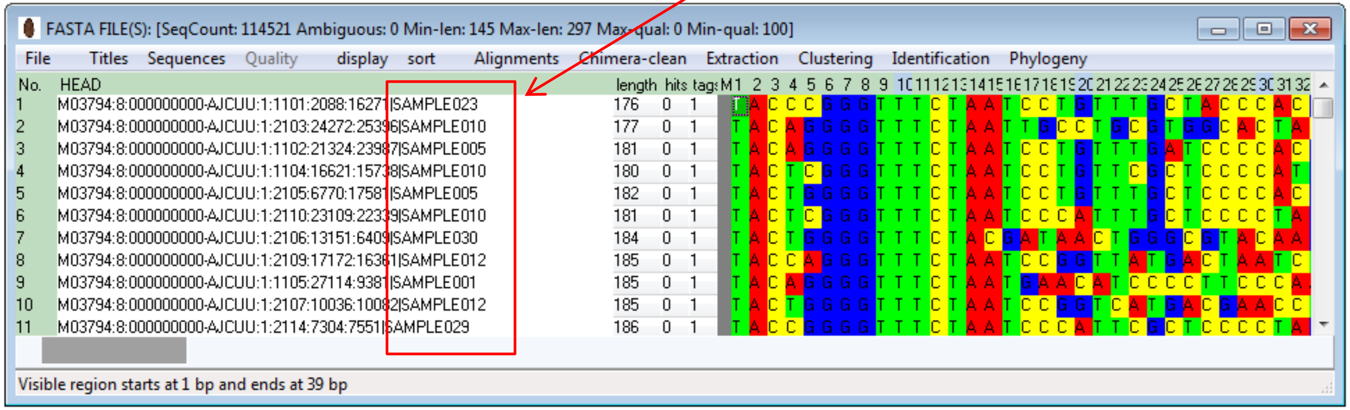
```
515F_T101|806R_T029	SAMPLE023
515F_T002|806R_T029	SAMPLE024
515F_T008|806R_T052	SAMPLE027
515F_T101|806R_T020	SAMPLE029
515F_T103|806R_T011	SAMPLE030
515F_T103|806R_T020	SAMPLE031
515F_T101|806R_T011	SAMPLE032
515F_T002|806R_T011	SAMPLE034
```

☑ Ignore multiple KEY values

● Replace original name
○ Append new name

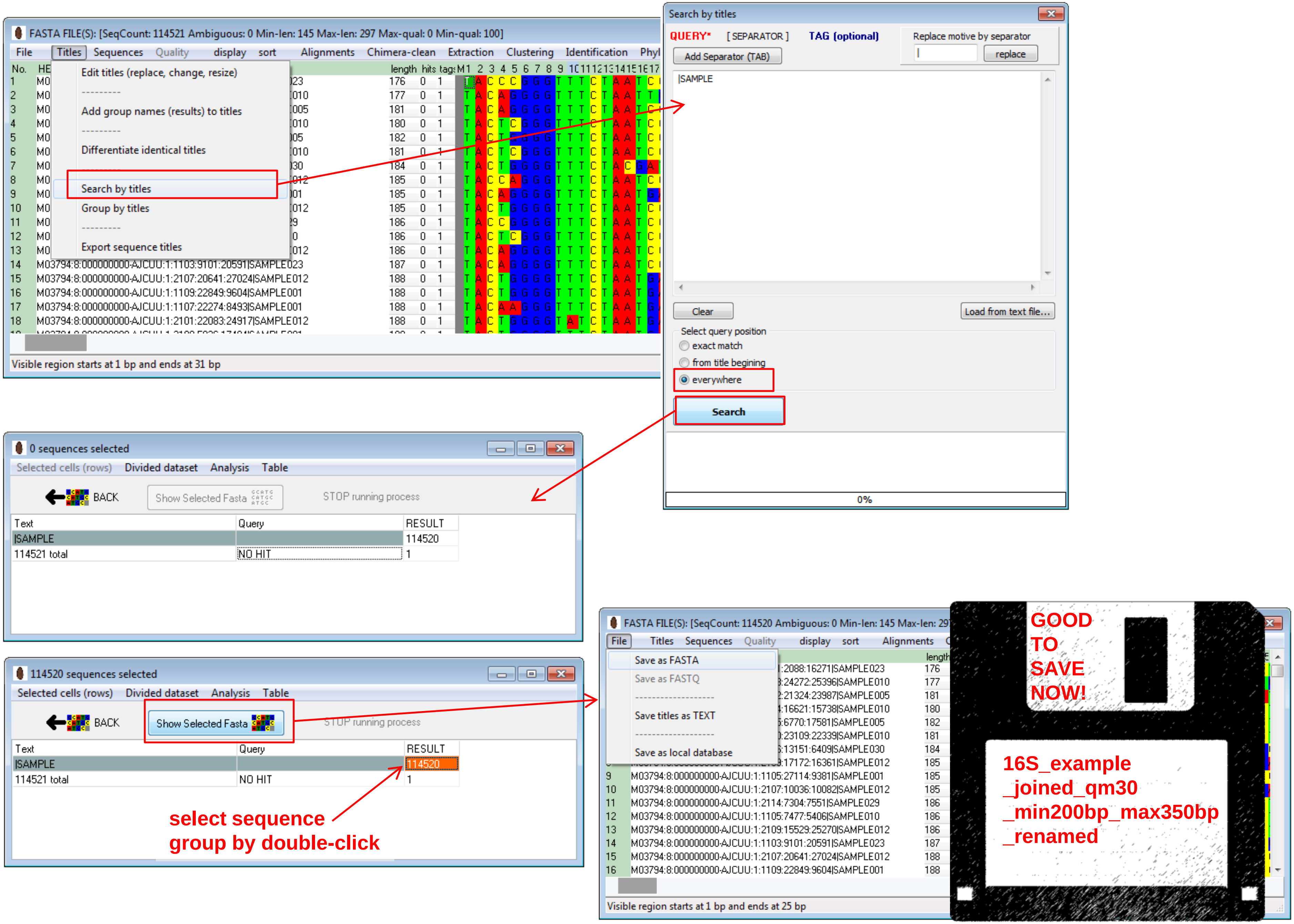


114520 replacements have been made.



No.	HEAD	length	hits	tags
1	M03794:8:000000000-AJCUU:1:1101:2088:16271|SAMPLE023	176	0	1
2	M03794:8:000000000-AJCUU:1:2103:24272:25396|SAMPLE010	177	0	1
3	M03794:8:000000000-AJCUU:1:1102:21324:23987|SAMPLE005	181	0	1
4	M03794:8:000000000-AJCUU:1:1104:16621:15738|SAMPLE010	180	0	1
5	M03794:8:000000000-AJCUU:1:2105:6770:17581|SAMPLE005	182	0	1
6	M03794:8:000000000-AJCUU:1:2110:23109:22339|SAMPLE010	181	0	1
7	M03794:8:000000000-AJCUU:1:2106:13151:6409|SAMPLE030	184	0	1
8	M03794:8:000000000-AJCUU:1:2109:17172:16361|SAMPLE012	185	0	1
9	M03794:8:000000000-AJCUU:1:1105:27114:9381|SAMPLE001	185	0	1
10	M03794:8:000000000-AJCUU:1:2107:10036:10082|SAMPLE012	185	0	1
11	M03794:8:000000000-AJCUU:1:2114:7304:7551|SAMPLE029	186	0	1

Visible region starts at 1 bp and ends at 39 bp



FASTA FILE(S): [SeqCount: 114521 Ambiguous: 0 Min-len: 145 Max-len: 297 Max-qual: 0 Min-qual: 100]

No.	HE	length	hits	tag
1	M0 …23	176	0	1
2	M0 …010	177	0	1
3	M0 …005	181	0	1
4	M0 …010	180	0	1
5	M0 …05	182	0	1
6	M0 …010	181	0	1
7	M0 …30	184	0	1
8	M0 …012	185	0	1
9	M0 …01	185	0	1
10	M0 …012	185	0	1
11	M0 …9	186	0	1
12	M0 …0	186	0	1
13	M0 …012	186	0	1
14	M03794:8:000000000-AJCUU:1:1103:9101:20591|SAMPLE023	187	0	1
15	M03794:8:000000000-AJCUU:1:2107:20641:27024|SAMPLE012	188	0	1
16	M03794:8:000000000-AJCUU:1:1109:22849:9604|SAMPLE001	188	0	1
17	M03794:8:000000000-AJCUU:1:1107:22274:8493|SAMPLE001	188	0	1
18	M03794:8:000000000-AJCUU:1:2101:22083:24917|SAMPLE012	188	0	1

Visible region starts at 1 bp and ends at 31 bp

QUERY* [SEPARATOR] **TAG (optional)**

Replace motive by separator: |

|SAMPLE

Select query position
- ○ exact match
- ○ from title begining
- ● everywhere

0%

0 sequences selected

Text	Query	RESULT
|SAMPLE		114520
114521 total	NO HIT	1

114520 sequences selected

Text	Query	RESULT
|SAMPLE		114520
114521 total	NO HIT	1

select sequence group by double-click

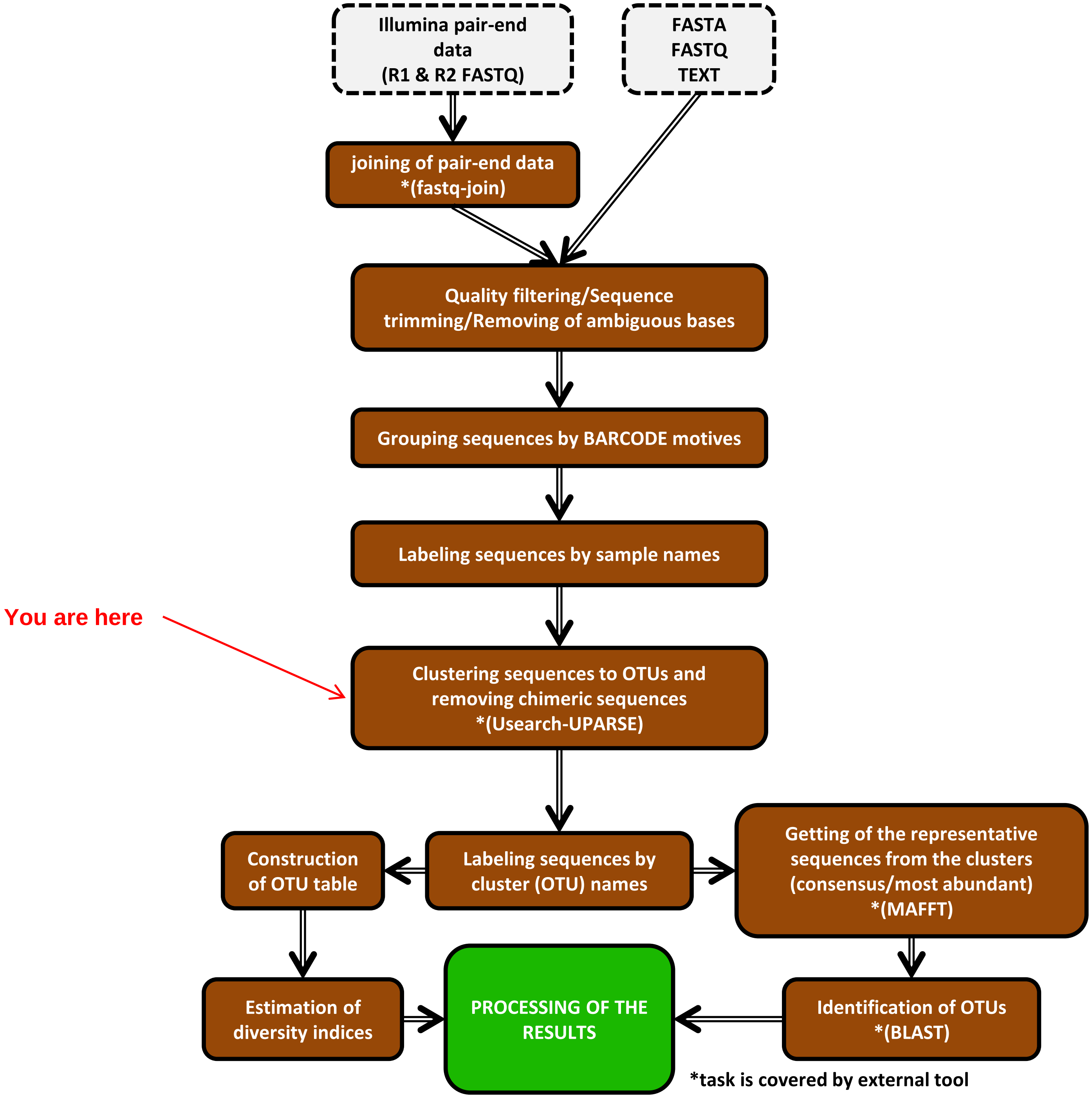
FASTA FILE(S): [SeqCount: 114520 Ambiguous: 0 Min-len: 145 Max-len: 29

No.	HE	length
	…:2088:16271|SAMPLE023	176
	…:24272:25396|SAMPLE010	177
	…:21324:23987|SAMPLE005	181
	…:16621:15738|SAMPLE010	180
	…:6770:17581|SAMPLE005	182
	…:23109:22339|SAMPLE010	181
	…:13151:6409|SAMPLE030	184
	…:17172:16361|SAMPLE012	185
9	M03794:8:000000000-AJCUU:1:1105:27114:9381|SAMPLE001	185
10	M03794:8:000000000-AJCUU:1:2107:10036:10082|SAMPLE012	185
11	M03794:8:000000000-AJCUU:1:2114:7304:7551|SAMPLE029	186
12	M03794:8:000000000-AJCUU:1:1105:7477:5406|SAMPLE010	186
13	M03794:8:000000000-AJCUU:1:2109:15529:25270|SAMPLE012	186
14	M03794:8:000000000-AJCUU:1:1103:9101:20591|SAMPLE023	187
15	M03794:8:000000000-AJCUU:1:2107:20641:27024|SAMPLE012	188
16	M03794:8:000000000-AJCUU:1:1109:22849:9604|SAMPLE001	188

Visible region starts at 1 bp and ends at 25 bp

GOOD TO SAVE NOW!

16S_example_joined_qm30_min200bp_max350bp_renamed



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

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)

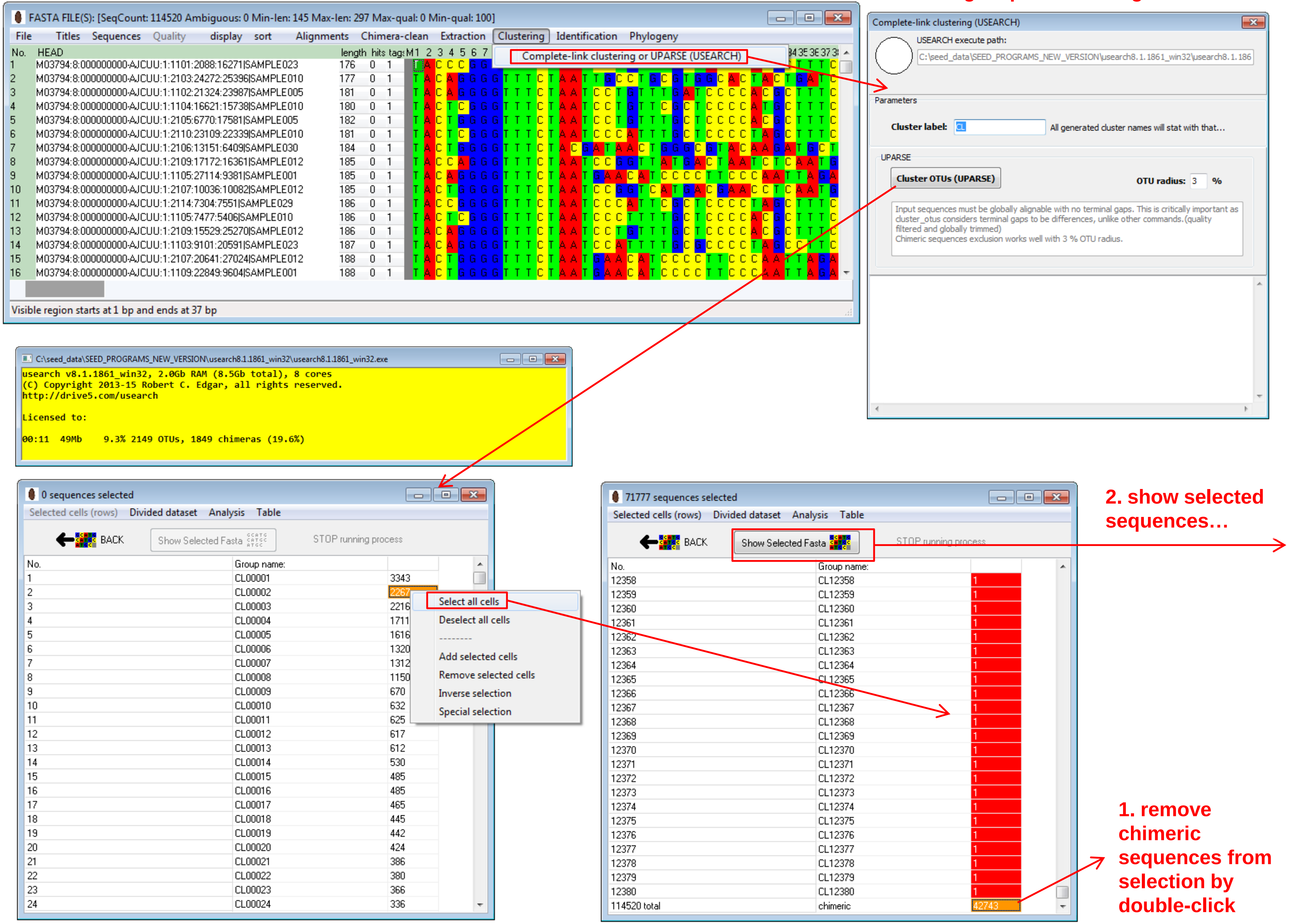
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool

Clustering sequences using USEARCH…



FASTA FILE(S): [SeqCount: 114520 Ambiguous: 0 Min-len: 145 Max-len: 297 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tag:	M1 2 3 4 5 6 7
1	M03794:8:000000000-AJCUU:1:1101:2088:16271|SAMPLE023	176	0	1	TACCCGG
2	M03794:8:000000000-AJCUU:1:2103:24272:25396|SAMPLE010	177	0	1	TACAGGGGTTTCTAATTGCCTGCGTGGCACTACTGATC
3	M03794:8:000000000-AJCUU:1:1102:21324:23987|SAMPLE005	181	0	1	TACAGGGGTTTCTAATCCTGTTTGATCCCCACGCTTTC
4	M03794:8:000000000-AJCUU:1:1104:16621:15738|SAMPLE010	180	0	1	TACTCGGGTTTCTAATCCTGTTCGCTCCCCATGCTTTC
5	M03794:8:000000000-AJCUU:1:2105:6770:17581|SAMPLE005	182	0	1	TACTGGGGTTTCTAATCCTGTTTGCTCCCCACGCTTTC
6	M03794:8:000000000-AJCUU:1:2110:23109:22339|SAMPLE010	181	0	1	TACTCGGGTTTCTAATCCCATTTGCTCCCCTAGCTTTC
7	M03794:8:000000000-AJCUU:1:2106:13151:6409|SAMPLE030	184	0	1	TACTGGGGTTTCTACGATAACTGGGCGTACAAGATGCT
8	M03794:8:000000000-AJCUU:1:2109:17172:16361|SAMPLE012	185	0	1	TACCAGGGTTTCTAATCCGGTTATGACTAATCTCAATG
9	M03794:8:000000000-AJCUU:1:1105:27114:9381|SAMPLE001	185	0	1	TACAGGGGTTTCTAATGAACATCCCCTTCCCAATTAGA
10	M03794:8:000000000-AJCUU:1:2107:10036:10082|SAMPLE012	185	0	1	TACTGGGGTTTCTAATCCGGTCATGACGAACCTCAATG
11	M03794:8:000000000-AJCUU:1:2114:7304:7551|SAMPLE029	186	0	1	TACCGGGGTTTCTAATCCCATTCGCTCCCCTAGCTTTC
12	M03794:8:000000000-AJCUU:1:1105:7477:5406|SAMPLE010	186	0	1	TACTCGGGTTTCTAATCCCTTTTGCTCCCCACGCTTTC
13	M03794:8:000000000-AJCUU:1:2109:15529:25270|SAMPLE012	186	0	1	TACAGGGGTTTCTAATCCTGTTTGCTCCCCACGCTTTC
14	M03794:8:000000000-AJCUU:1:1103:9101:20591|SAMPLE023	187	0	1	TACAGGGGTTTCTAATCCATTTTGCGCCCCTAGCCTTC
15	M03794:8:000000000-AJCUU:1:2107:20641:27024|SAMPLE012	188	0	1	TACTGGGGTTTCTAATGAACATCCCCTTCCCAATTAGA
16	M03794:8:000000000-AJCUU:1:1109:22849:9604|SAMPLE001	188	0	1	TACTGGGGTTTCTAATGAACATCCCCTTCCCAATTAGA

Visible region starts at 1 bp and ends at 37 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

Cluster OTUs (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.

C:\seed_data\SEED_PROGRAMS_NEW_VERSION\usearch8.1.1861_win32\usearch8.1.1861_win32.exe

```
usearch v8.1.1861_win32, 2.0Gb RAM (8.5Gb total), 8 cores
(C) Copyright 2013-15 Robert C. Edgar, all rights reserved.
http://drive5.com/usearch

Licensed to:

00:11  49Mb    9.3% 2149 OTUs, 1849 chimeras (19.6%)
```

0 sequences selected

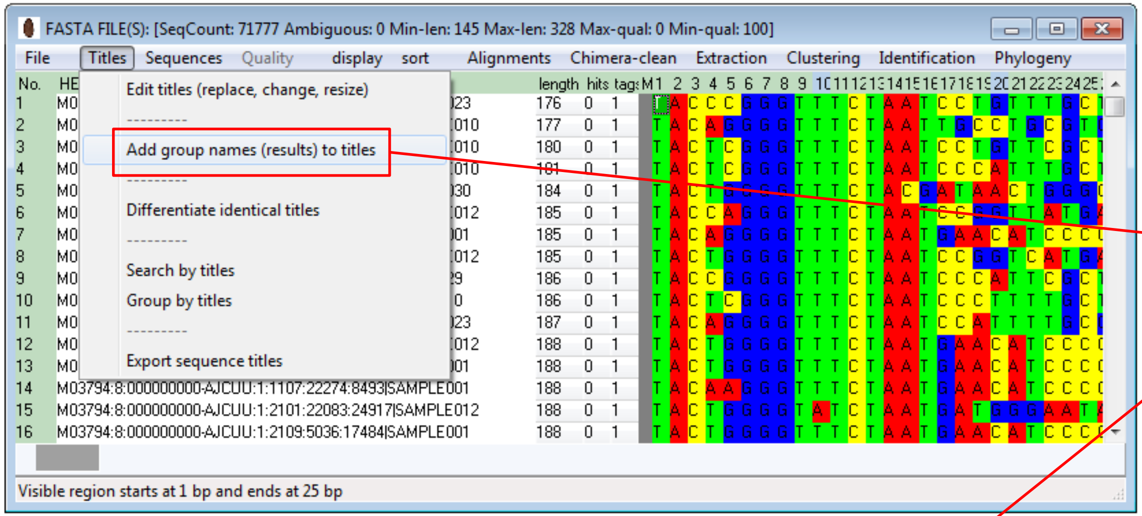
No.	Group name:	
1	CL00001	3343
2	CL00002	2267
3	CL00003	2216
4	CL00004	1711
5	CL00005	1616
6	CL00006	1320
7	CL00007	1312
8	CL00008	1150
9	CL00009	670
10	CL00010	632
11	CL00011	625
12	CL00012	617
13	CL00013	612
14	CL00014	530
15	CL00015	485
16	CL00016	485
17	CL00017	465
18	CL00018	445
19	CL00019	442
20	CL00020	424
21	CL00021	386
22	CL00022	380
23	CL00023	366
24	CL00024	336

71777 sequences selected

No.	Group name:	
12358	CL12358	1
12359	CL12359	1
12360	CL12360	1
12361	CL12361	1
12362	CL12362	1
12363	CL12363	1
12364	CL12364	1
12365	CL12365	1
12366	CL12366	1
12367	CL12367	1
12368	CL12368	1
12369	CL12369	1
12370	CL12370	1
12371	CL12371	1
12372	CL12372	1
12373	CL12373	1
12374	CL12374	1
12375	CL12375	1
12376	CL12376	1
12377	CL12377	1
12378	CL12378	1
12379	CL12379	1
12380	CL12380	1
114520 total	chimeric	42743

2. show selected sequences…

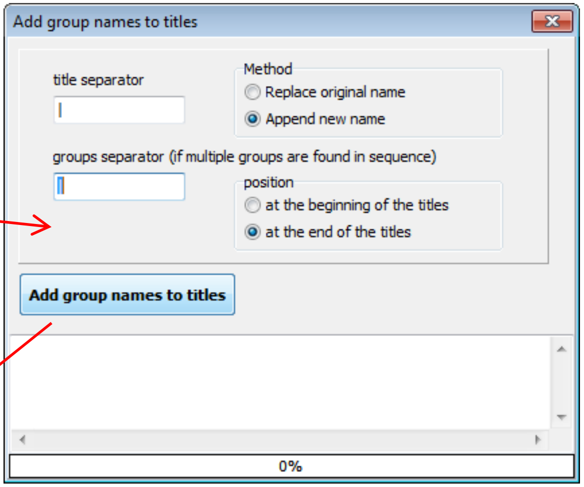
1. remove chimeric sequences from selection by double-click



FASTA FILE(S): [SeqCount: 71777 Ambiguous: 0 Min-len: 145 Max-len: 328 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tags
14	M03794:8:000000000-AJCUU:1:1107:22274:8493|SAMPLE001	188	0	1
15	M03794:8:000000000-AJCUU:1:2101:22083:24917|SAMPLE012	188	0	1
16	M03794:8:000000000-AJCUU:1:2109:5036:17484|SAMPLE001	188	0	1

Visible region starts at 1 bp and ends at 25 bp



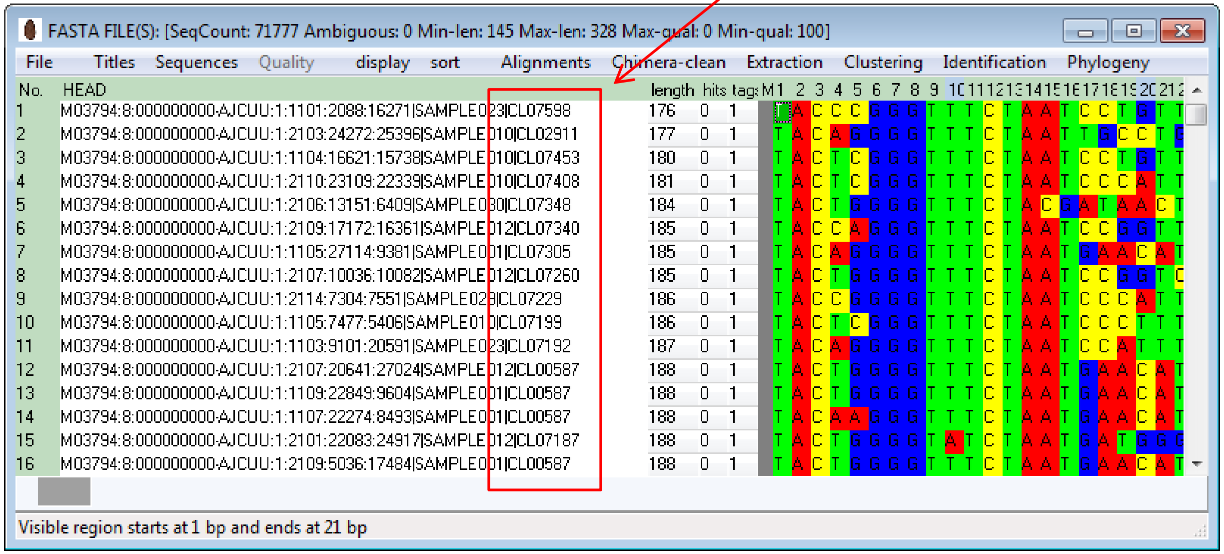
title separator: |

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…

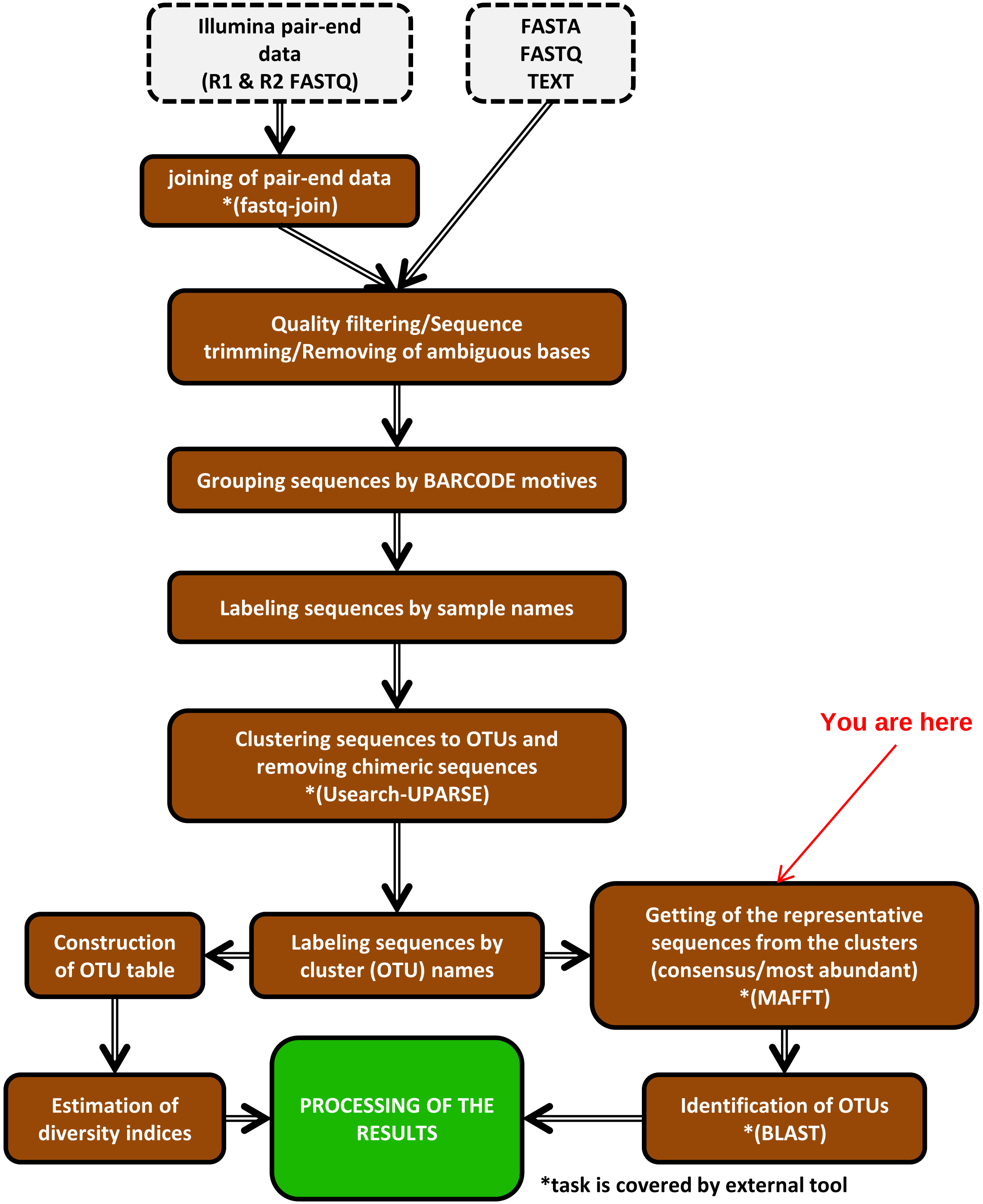


FASTA FILE(S): [SeqCount: 71777 Ambiguous: 0 Min-len: 145 Max-len: 328 Max-qual: 0 Min-qual: 100]

No.	HEAD	length	hits	tags
1	M03794:8:000000000-AJCUU:1:1101:2088:16271|SAMPLE023|CL07598	176	0	1
2	M03794:8:000000000-AJCUU:1:2103:24272:25396|SAMPLE010|CL02911	177	0	1
3	M03794:8:000000000-AJCUU:1:1104:16621:15738|SAMPLE010|CL07453	180	0	1
4	M03794:8:000000000-AJCUU:1:2110:23109:22339|SAMPLE010|CL07408	181	0	1
5	M03794:8:000000000-AJCUU:1:2106:13151:6409|SAMPLE030|CL07348	184	0	1
6	M03794:8:000000000-AJCUU:1:2109:17172:16361|SAMPLE012|CL07340	185	0	1
7	M03794:8:000000000-AJCUU:1:1105:27114:9381|SAMPLE001|CL07305	185	0	1
8	M03794:8:000000000-AJCUU:1:2107:10036:10082|SAMPLE012|CL07260	185	0	1
9	M03794:8:000000000-AJCUU:1:2114:7304:7551|SAMPLE029|CL07229	186	0	1
10	M03794:8:000000000-AJCUU:1:1105:7477:5406|SAMPLE010|CL07199	186	0	1
11	M03794:8:000000000-AJCUU:1:1103:9101:20591|SAMPLE023|CL07192	187	0	1
12	M03794:8:000000000-AJCUU:1:2107:20641:27024|SAMPLE012|CL00587	188	0	1
13	M03794:8:000000000-AJCUU:1:1109:22849:9604|SAMPLE001|CL00587	188	0	1
14	M03794:8:000000000-AJCUU:1:1107:22274:8493|SAMPLE001|CL00587	188	0	1
15	M03794:8:000000000-AJCUU:1:2101:22083:24917|SAMPLE012|CL07187	188	0	1
16	M03794:8:000000000-AJCUU:1:2109:5036:17484|SAMPLE001|CL00587	188	0	1

Visible region starts at 1 bp and ends at 21 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

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)

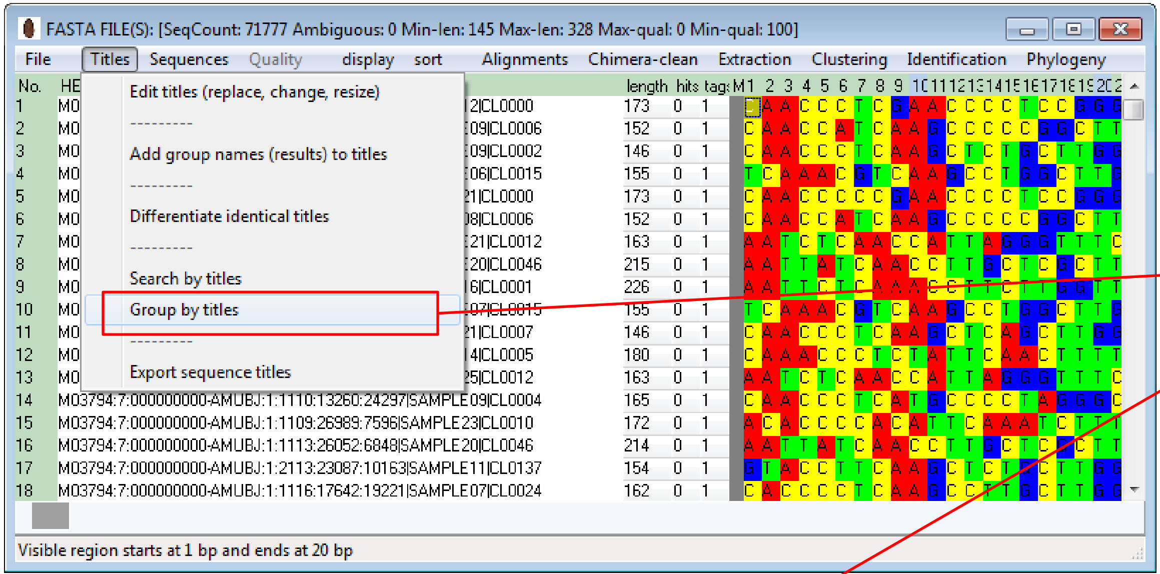
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

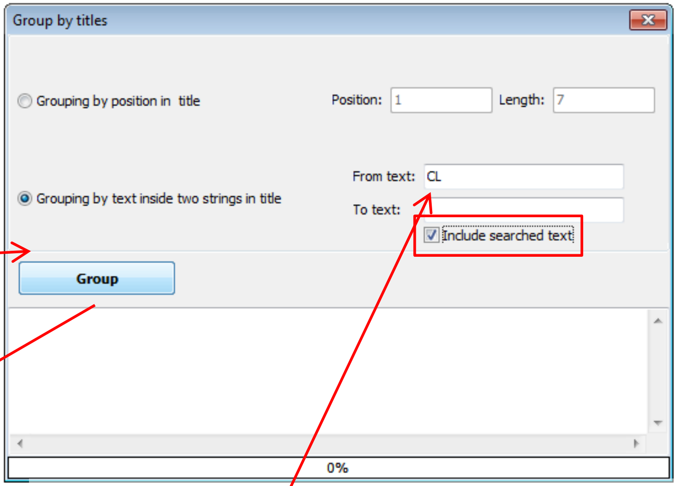
*task is covered by external tool

Get clusters (OTUs) representative sequences…


FASTA FILE(S): [SeqCount: 71777 Ambiguous: 0 Min-len: 145 Max-len: 328 Max-qual: 0 Min-qual: 100]

No.	HE	length	hits	tag:
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|CL0046	215	0	1
9	M0 16|CL0001	226	0	1
10	M0 07|CL0015	155	0	1
11	M0 21|CL0007	146	0	1
12	M0 14|CL0005	180	0	1
13	M0 25|CL0012	163	0	1
14	M03794:7:000000000-AMUBJ:1:1110:13260:24297|SAMPLE09|CL0004	165	0	1
15	M03794:7:000000000-AMUBJ:1:1109:26989:7596|SAMPLE23|CL0010	172	0	1
16	M03794:7:000000000-AMUBJ:1:1113:26052:6848|SAMPLE20|CL0046	214	0	1
17	M03794:7:000000000-AMUBJ:1:2113:23087:10163|SAMPLE11|CL0137	154	0	1
18	M03794:7:000000000-AMUBJ:1:1116:17642:19221|SAMPLE07|CL0024	162	0	1

Visible region starts at 1 bp and ends at 20 bp

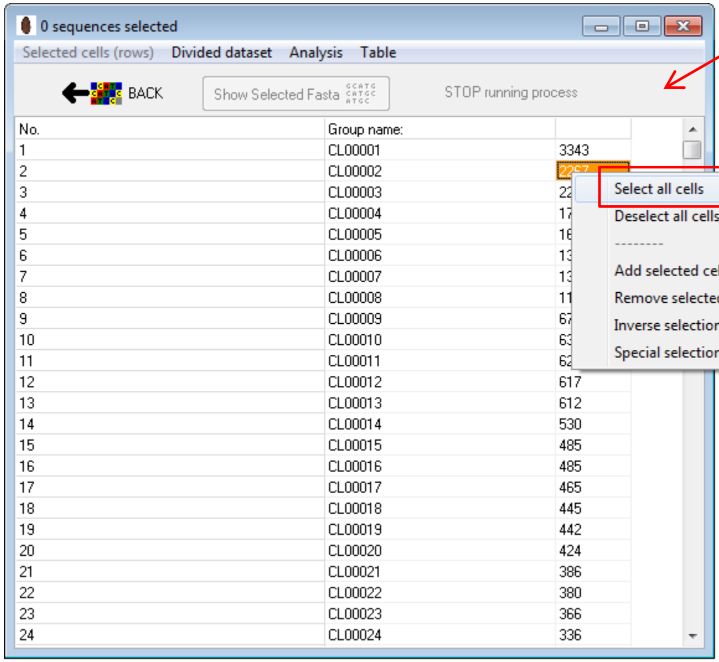

Group by titles

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

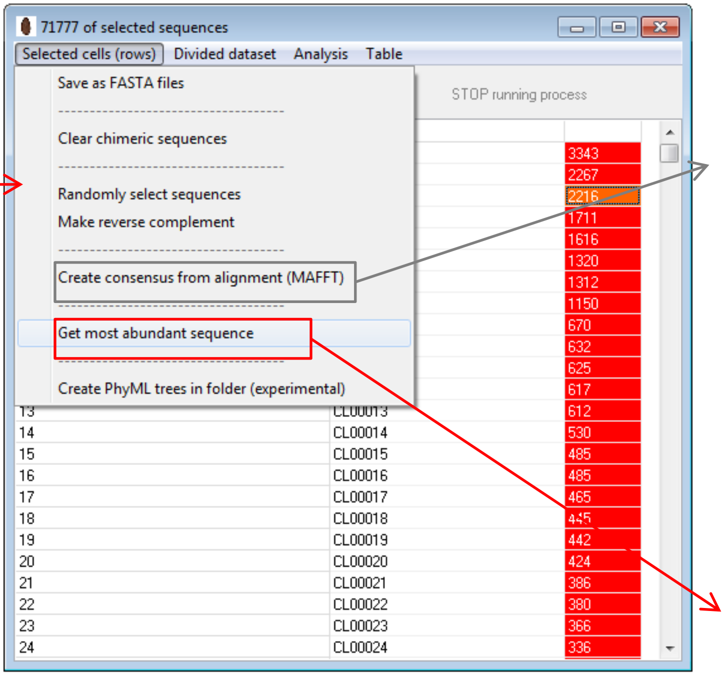
● Grouping by text inside two strings in title — From text: CL — To text:

☑ Include searched text

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


0 sequences selected

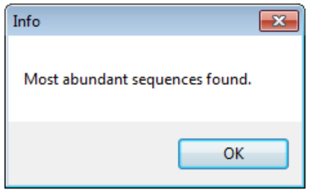
No.	Group name:	
1	CL00001	3343
2	CL00002	2267
3	CL00003	22
4	CL00004	17
5	CL00005	16
6	CL00006	13
7	CL00007	13
8	CL00008	11
9	CL00009	67
10	CL00010	63
11	CL00011	62
12	CL00012	617
13	CL00013	612
14	CL00014	530
15	CL00015	485
16	CL00016	485
17	CL00017	465
18	CL00018	445
19	CL00019	442
20	CL00020	424
21	CL00021	386
22	CL00022	380
23	CL00023	366
24	CL00024	336


71777 of selected sequences

		3343
		2267
		2216
		1711
		1616
		1320
		1312
		1150
		670
		632
		625
		617
13	CL00013	612
14	CL00014	530
15	CL00015	485
16	CL00016	485
17	CL00017	465
18	CL00018	445
19	CL00019	442
20	CL00020	424
21	CL00021	386
22	CL00022	380
23	CL00023	366
24	CL00024	336

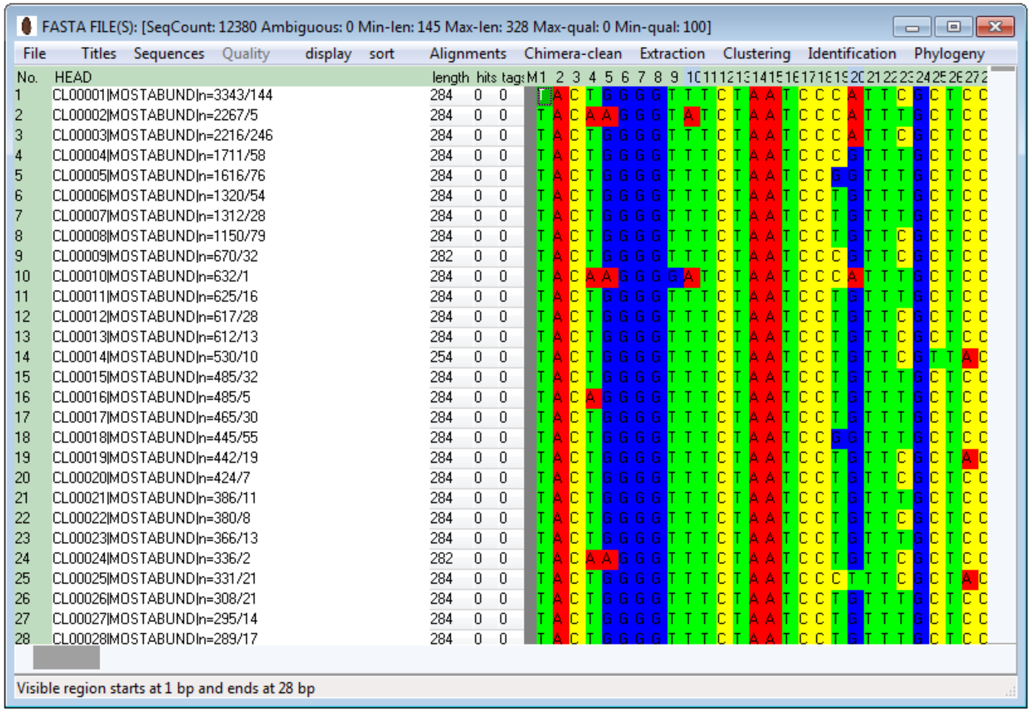
(alternative) compute a consensus from aligned sequences using „mafft„ aligner (it may take a long while)

get the most abundant sequence from each cluster (fast)


Info

Most abundant sequences found.

OUTs representative sequences - the most abundant sequences…



FASTA FILE(S): [SeqCount: 12380 Ambiguous: 0 Min-len: 145 Max-len: 328 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	2
1	CL00001|MOSTABUND|n=3343/144	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	C	A	T	T	C	G	C	T	C	C
2	CL00002|MOSTABUND|n=2267/5	284	0	0		T	A	C	A	A	G	G	G	T	A	T	C	T	A	A	T	C	C	C	A	T	T	T	G	C	T	C	C
3	CL00003|MOSTABUND|n=2216/246	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	C	A	T	T	C	G	C	T	C	C
4	CL00004|MOSTABUND|n=1711/58	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	C	G	T	T	T	G	C	T	C	C
5	CL00005|MOSTABUND|n=1616/76	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	G	G	T	T	T	G	C	T	C	C
6	CL00006|MOSTABUND|n=1320/54	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
7	CL00007|MOSTABUND|n=1312/28	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
8	CL00008|MOSTABUND|n=1150/79	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	C	G	C	T	C	C
9	CL00009|MOSTABUND|n=670/32	282	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	C	G	T	T	C	G	C	T	C	C
10	CL00010|MOSTABUND|n=632/1	284	0	0		T	A	C	A	A	G	G	G	G	A	T	C	T	A	A	T	C	C	C	A	T	T	T	G	C	T	C	C
11	CL00011|MOSTABUND|n=625/16	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
12	CL00012|MOSTABUND|n=617/28	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	C	G	C	T	C	C
13	CL00013|MOSTABUND|n=612/13	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	C	G	C	T	C	C
14	CL00014|MOSTABUND|n=530/10	254	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	C	G	T	T	A	C
15	CL00015|MOSTABUND|n=485/32	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
16	CL00016|MOSTABUND|n=485/5	284	0	0		T	A	C	A	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
17	CL00017|MOSTABUND|n=465/30	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
18	CL00018|MOSTABUND|n=445/55	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	G	G	T	T	T	G	C	T	C	C
19	CL00019|MOSTABUND|n=442/19	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	C	G	C	T	A	C
20	CL00020|MOSTABUND|n=424/7	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	C	G	C	T	C	C
21	CL00021|MOSTABUND|n=386/11	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
22	CL00022|MOSTABUND|n=380/8	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	C	G	C	T	C	C
23	CL00023|MOSTABUND|n=366/13	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
24	CL00024|MOSTABUND|n=336/2	282	0	0		T	A	C	A	A	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	C	G	C	T	C	C
25	CL00025|MOSTABUND|n=331/21	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	C	T	T	T	C	G	C	T	A	C
26	CL00026|MOSTABUND|n=308/21	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
27	CL00027|MOSTABUND|n=295/14	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C
28	CL00028|MOSTABUND|n=289/17	284	0	0		T	A	C	T	G	G	G	G	T	T	T	C	T	A	A	T	C	C	T	G	T	T	T	G	C	T	C	C

Visible region starts at 1 bp and ends at 28 bp

CL00001|MOSTABUND|n=3343/144

cluster name

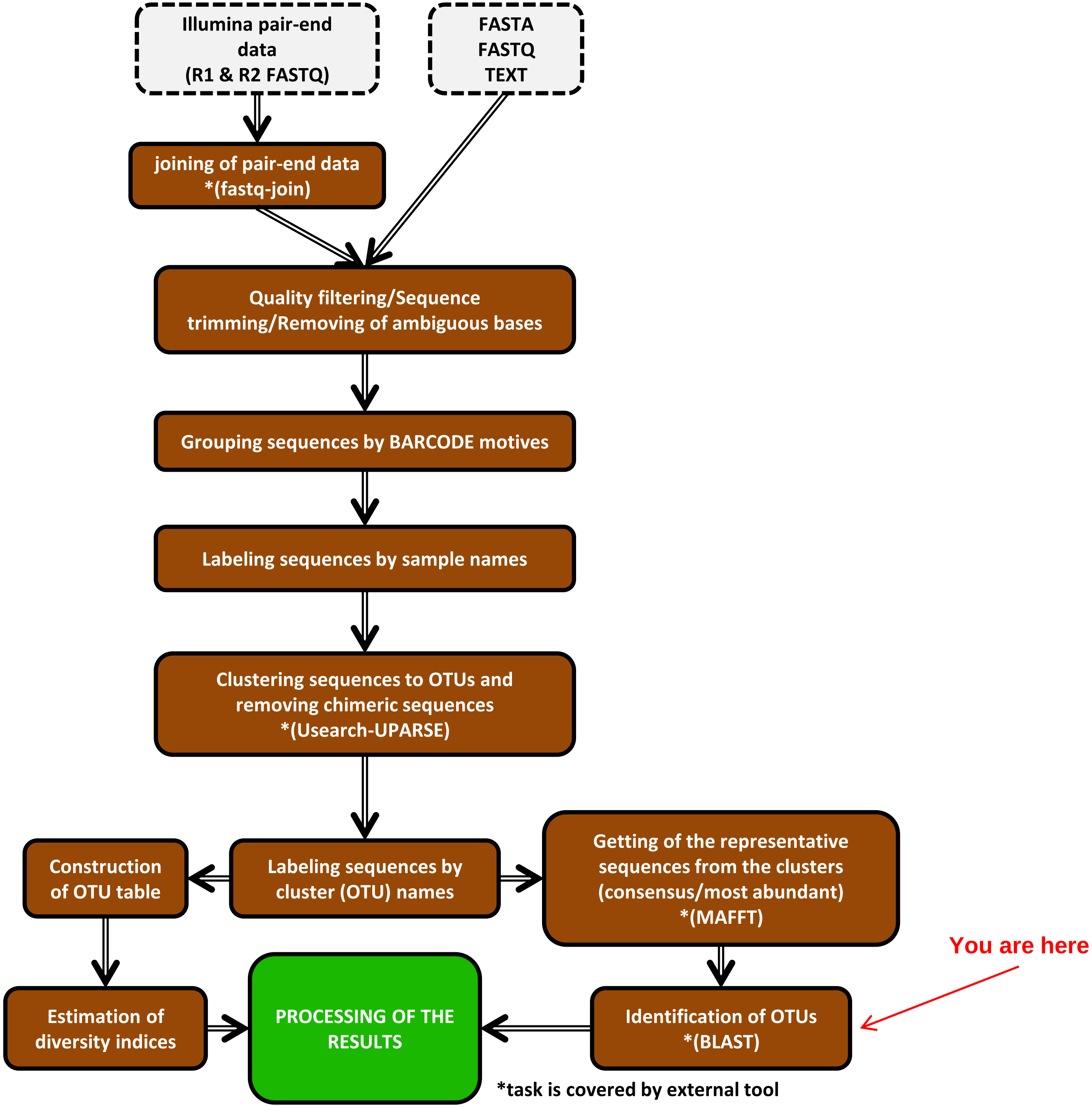
type of selection

number of sequences in group

number of most abundant identical sequences in group

GOOD TO SAVE NOW!

16S_example_joined_qm30_min200bp_max350bp_renamed_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

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

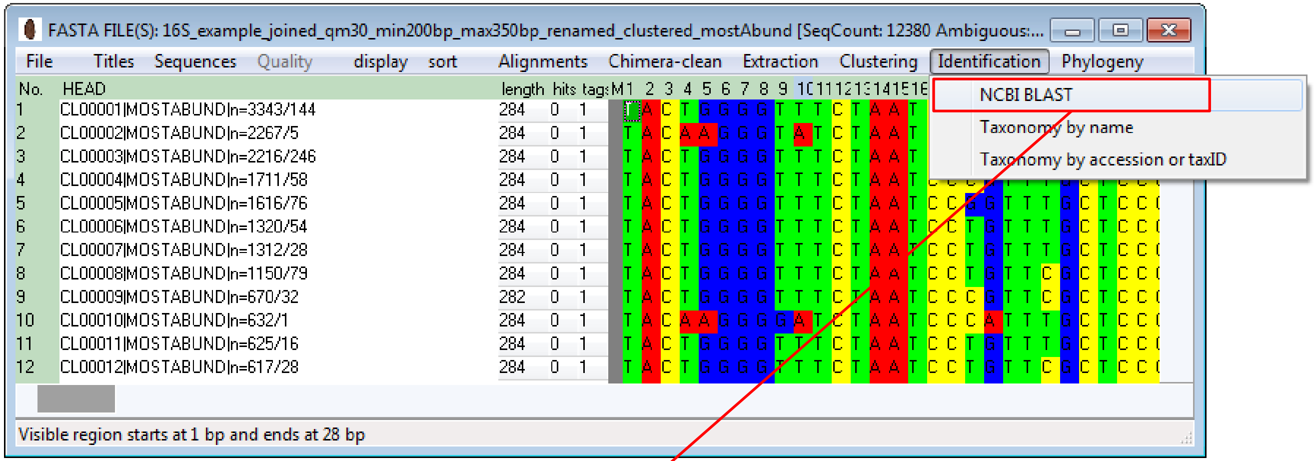
PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

You are here

*task is covered by external tool

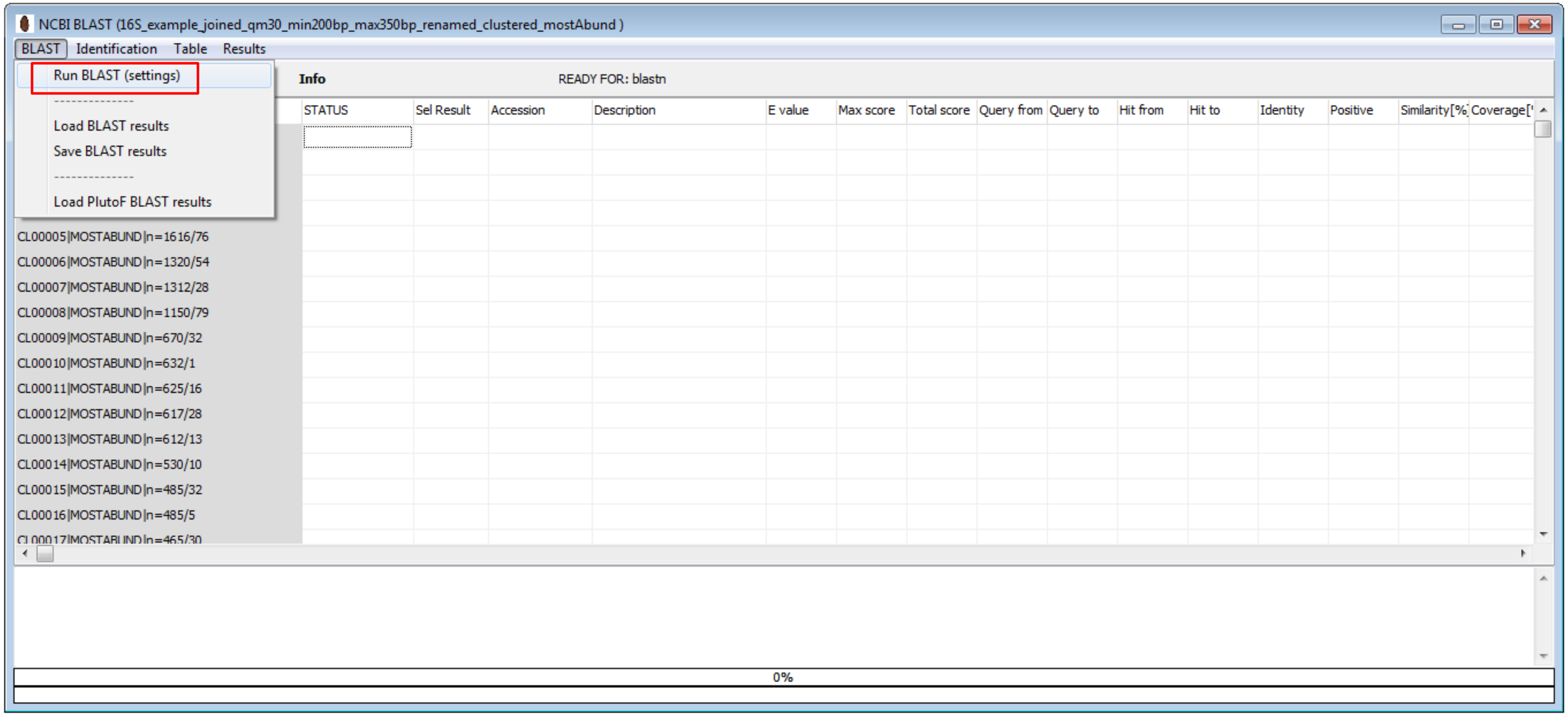
Identification of representative sequences…



FASTA FILE(S): 16S_example_joined_qm30_min200bp_max350bp_renamed_clustered_mostAbund [SeqCount: 12380 Ambiguous:...

No.	HEAD	length	hits	tags	M1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
1	CL00001|MOSTABUND|n=3343/144	284	0	1	T A C T G G G G T T T C T A A T
2	CL00002|MOSTABUND|n=2267/5	284	0	1	T A C A A G G G T A T C T A A T
3	CL00003|MOSTABUND|n=2216/246	284	0	1	T A C T G G G G T T T C T A A T
4	CL00004|MOSTABUND|n=1711/58	284	0	1	T A C T G G G G T T T C T A A T
5	CL00005|MOSTABUND|n=1616/76	284	0	1	T A C T G G G G T T T C T A A T C C G G T T T G C T C C
6	CL00006|MOSTABUND|n=1320/54	284	0	1	T A C T G G G G T T T C T A A T C C T G T T T G C T C C
7	CL00007|MOSTABUND|n=1312/28	284	0	1	T A C T G G G G T T T C T A A T C C T G T T T G C T C C
8	CL00008|MOSTABUND|n=1150/79	284	0	1	T A C T G G G G T T T C T A A T C C T G T T C G C T C C
9	CL00009|MOSTABUND|n=670/32	282	0	1	T A C T G G G G T T T C T A A T C C C G T T C G C T C C
10	CL00010|MOSTABUND|n=632/1	284	0	1	T A C A A G G G G A T C T A A T C C C A T T T G C T C C
11	CL00011|MOSTABUND|n=625/16	284	0	1	T A C T G G G G T T T C T A A T C C T G T T T G C T C C
12	CL00012|MOSTABUND|n=617/28	284	0	1	T A C T G G G G T T T C T A A T C C T G T T C G C T C C

Visible region starts at 1 bp and ends at 28 bp

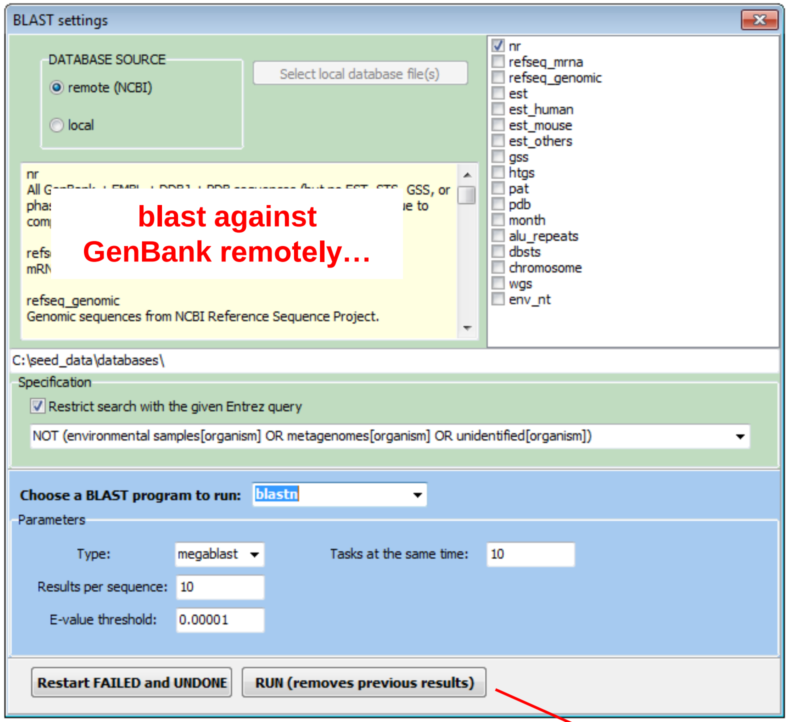


NCBI BLAST (16S_example_joined_qm30_min200bp_max350bp_renamed_clustered_mostAbund)

Info 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[%	Coverage['

CL00005|MOSTABUND|n=1616/76
CL00006|MOSTABUND|n=1320/54
CL00007|MOSTABUND|n=1312/28
CL00008|MOSTABUND|n=1150/79
CL00009|MOSTABUND|n=670/32
CL00010|MOSTABUND|n=632/1
CL00011|MOSTABUND|n=625/16
CL00012|MOSTABUND|n=617/28
CL00013|MOSTABUND|n=612/13
CL00014|MOSTABUND|n=530/10
CL00015|MOSTABUND|n=485/32
CL00016|MOSTABUND|n=485/5



DATABASE SOURCE

● remote (NCBI)

○ local

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

nr

refseq_genomic
Genomic sequences from NCBI Reference Sequence Project.

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

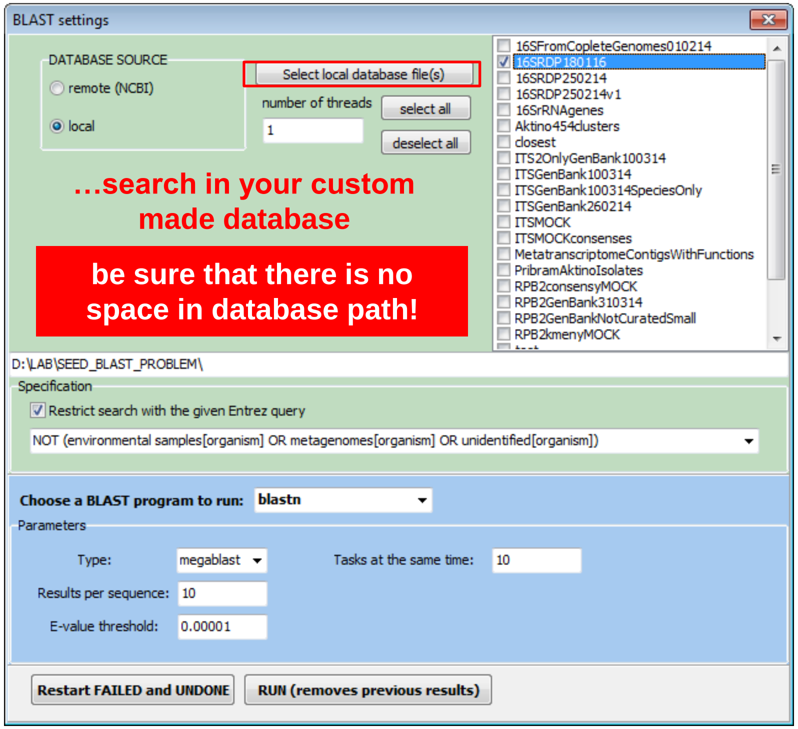
Tasks at the same time: 10

Results per sequence: 10

E-value threshold: 0.00001

blast against GenBank remotely…

or



DATABASE SOURCE

○ remote (NCBI)

● local

number of threads: 1

☐ 16SFromCompleteGenomes010214
☑ 16SRDP180116
☐ 16SRDP250214
☐ 16SRDP250214v1
☐ 16SrRNAgenes
☐ Aktino454clusters
☐ closest
☐ ITS2OnlyGenBank100314
☐ ITSGenBank100314
☐ ITSGenBank100314SpeciesOnly
☐ ITSGenBank260214
☐ ITSMOCK
☐ ITSMOCKconsenses
☐ MetatranscriptomeContigsWithFunctions
☐ PribramAktinoIsolates
☐ RPB2consensyMOCK
☐ RPB2GenBank310314
☐ RPB2GenBankNotCuratedSmall
☐ RPB2kmenyMOCK

D:\LAB\SEED_BLAST_PROBLEM\

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

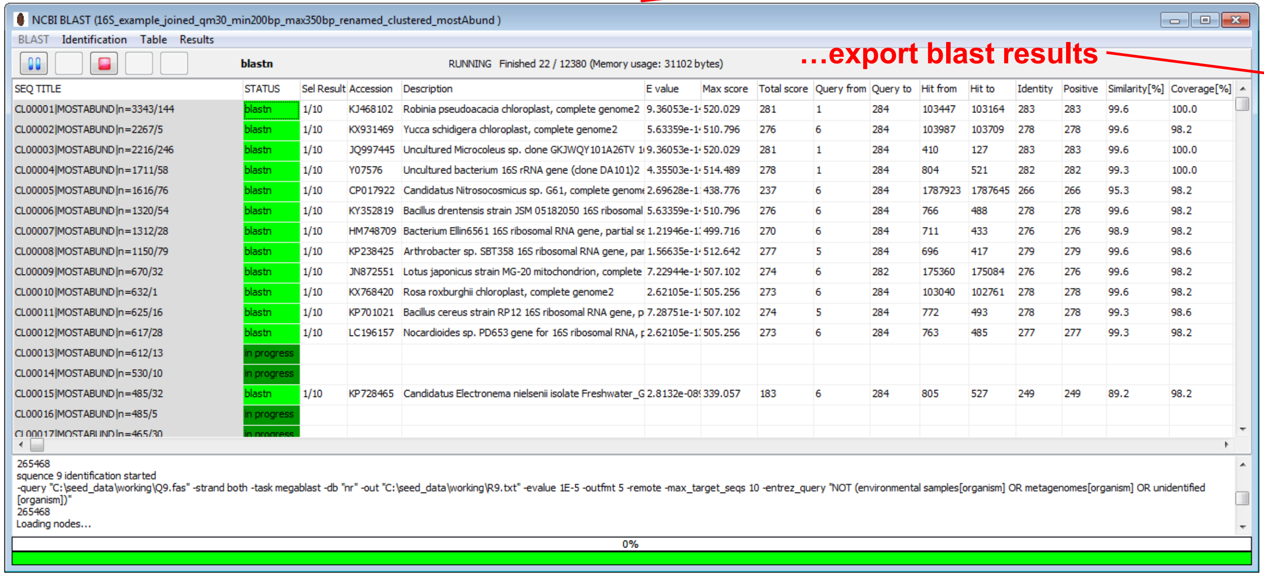
Tasks at the same time: 10

Results per sequence: 10

E-value threshold: 0.00001

…search in your custom made database

be sure that there is no space in database path!



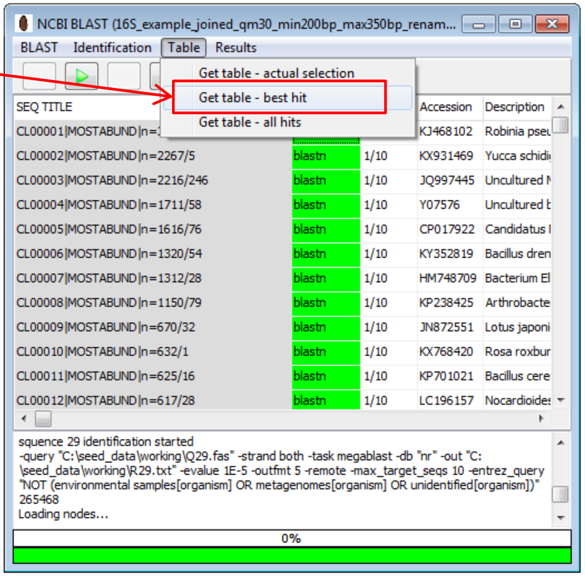
blastn RUNNING Finished 22 / 12380 (Memory usage: 31102 bytes)

SEQ TITLE	STATUS	Sel Result	Accession	Description	E value	Max score	Total score	Query from	Query to	Hit from	Hit to	Identity	Positive	Similarity[%]	Coverage[%]
CL00001|MOSTABUND|n=3343/144	blastn	1/10	KJ468102	Robinia pseudoacacia chloroplast, complete genome2	9.36053e-1	520.029	281	1	284	103447	103164	283	283	99.6	100.0
CL00002|MOSTABUND|n=2267/5	blastn	1/10	KX931469	Yucca schidigera chloroplast, complete genome2	5.63359e-1	510.796	276	6	284	103987	103709	278	278	99.6	98.2
CL00003|MOSTABUND|n=2216/246	blastn	1/10	JQ997445	Uncultured Microcoleus sp. clone GKJWQY101A26TV 1	9.36053e-1	520.029	281	1	284	410	127	283	283	99.6	100.0
CL00004|MOSTABUND|n=1711/58	blastn	1/10	Y07576	Uncultured bacterium 16S rRNA gene (clone DA101)2	4.35503e-1	514.489	278	1	284	804	521	282	282	99.3	100.0
CL00005|MOSTABUND|n=1616/76	blastn	1/10	CP017922	Candidatus Nitrosocosmicus sp. G61, complete genome	2.69628e-1	438.776	237	6	284	1787923	1787645	266	266	95.3	98.2
CL00006|MOSTABUND|n=1320/54	blastn	1/10	KY352819	Bacillus drentensis strain JSM 05182050 16S ribosomal	5.63359e-1	510.796	276	6	284	766	488	278	278	99.6	98.2
CL00007|MOSTABUND|n=1312/28	blastn	1/10	HM748709	Bacterium Ellin6561 16S ribosomal RNA gene, partial se	1.21946e-1	499.716	270	6	284	711	433	276	276	98.9	98.2
CL00008|MOSTABUND|n=1150/79	blastn	1/10	KP238425	Arthrobacter sp. SBT358 16S ribosomal RNA gene, par	1.56635e-1	512.642	277	5	284	696	417	279	279	99.6	98.6
CL00009|MOSTABUND|n=670/32	blastn	1/10	JN872551	Lotus japonicus strain MG-20 mitochondrion, complete	7.22944e-1	507.102	274	6	282	175360	175084	276	276	99.6	98.2
CL00010|MOSTABUND|n=632/1	blastn	1/10	KX768420	Rosa roxburghii chloroplast, complete genome2	2.62105e-1	505.256	273	6	284	103040	102761	278	278	99.6	98.2
CL00011|MOSTABUND|n=625/16	blastn	1/10	KP701021	Bacillus cereus strain RP12 16S ribosomal RNA gene, p	7.28751e-1	507.102	274	5	284	772	493	278	278	99.3	98.6
CL00012|MOSTABUND|n=617/28	blastn	1/10	LC196157	Nocardioides sp. PD653 gene for 16S ribosomal RNA, p	2.62105e-1	505.256	273	6	284	763	485	277	277	99.3	98.2
CL00013|MOSTABUND|n=612/13	in progress														
CL00014|MOSTABUND|n=530/10	in progress														
CL00015|MOSTABUND|n=485/32	blastn	1/10	KP728465	Candidatus Electronema nielsenii isolate Freshwater_G	2.8132e-08	339.057	183	6	284	805	527	249	249	89.2	98.2
CL00016|MOSTABUND|n=485/5	in progress														
CL00017|MOSTABUND|n=465/30	in progress														

```
265468
squence 9 identification started
-query "C:\seed_data\working\Q9.fas" -strand both -task megablast -db "nr" -out "C:\seed_data\working\R9.txt" -evalue 1E-5 -outfmt 5 -remote -max_target_seqs 10 -entrez_query "NOT (environmental samples[organism] OR metagenomes[organism] OR unidentified
[organism])"
265468
Loading nodes...
```

0%

…export blast results

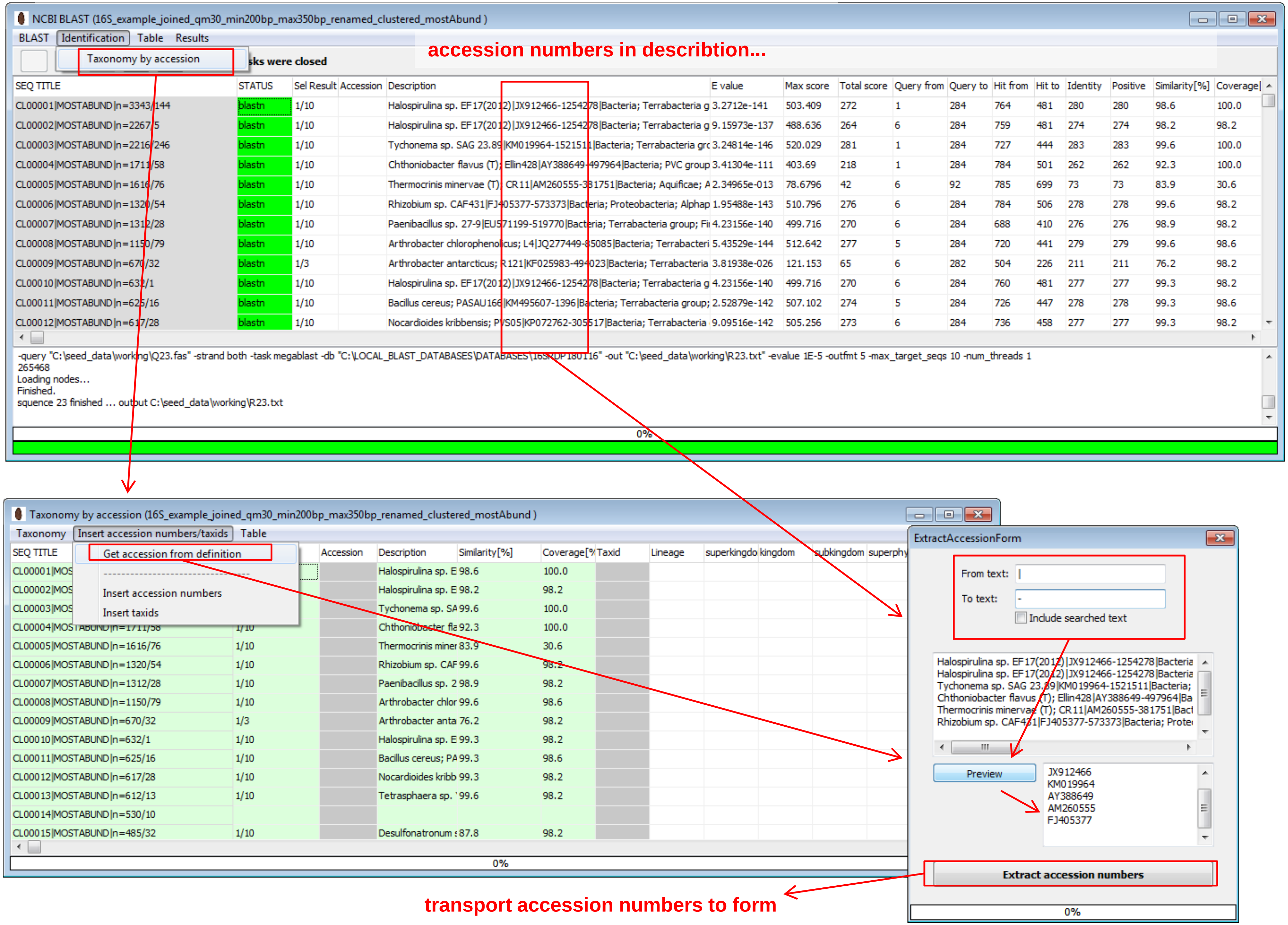


SEQ TITLE			Accession	Description
CL00001|MOSTABUND|n=			KJ468102	Robinia pseu
CL00002|MOSTABUND|n=2267/5	blastn	1/10	KX931469	Yucca schidi
CL00003|MOSTABUND|n=2216/246	blastn	1/10	JQ997445	Uncultured M
CL00004|MOSTABUND|n=1711/58	blastn	1/10	Y07576	Uncultured b
CL00005|MOSTABUND|n=1616/76	blastn	1/10	CP017922	Candidatus I
CL00006|MOSTABUND|n=1320/54	blastn	1/10	KY352819	Bacillus dren
CL00007|MOSTABUND|n=1312/28	blastn	1/10	HM748709	Bacterium El
CL00008|MOSTABUND|n=1150/79	blastn	1/10	KP238425	Arthrobacte
CL00009|MOSTABUND|n=670/32	blastn	1/10	JN872551	Lotus japoni
CL00010|MOSTABUND|n=632/1	blastn	1/10	KX768420	Rosa roxbur
CL00011|MOSTABUND|n=625/16	blastn	1/10	KP701021	Bacillus cere
CL00012|MOSTABUND|n=617/28	blastn	1/10	LC196157	Nocardioides

```
squence 29 identification started
-query "C:\seed_data\working\Q29.fas" -strand both -task megablast -db "nr" -out "C:
\seed_data\working\R29.txt" -evalue 1E-5 -outfmt 5 -remote -max_target_seqs 10 -entrez_query
"NOT (environmental samples[organism] OR metagenomes[organism] OR unidentified[organism])"
265468
Loading nodes...
```

0%

Get taxonomic classification (custom made database example)…



accession numbers in describtion...

transport accession numbers to form

SEQ TITLE	STATUS	Sel Result	Accession	Description	E value	Max score	Total score	Query from	Query to	Hit from	Hit to	Identity	Positive	Similarity[%]	Coverage[
CL00001|MOSTABUND|n=3343/144	blastn	1/10		Halospirulina sp. EF17(2012)|JX912466-1254278|Bacteria; Terrabacteria g	3.2712e-141	503.409	272	1	284	764	481	280	280	98.6	100.0
CL00002|MOSTABUND|n=2267/5	blastn	1/10		Halospirulina sp. EF17(2012)|JX912466-1254278|Bacteria; Terrabacteria g	9.15973e-137	488.636	264	6	284	759	481	274	274	98.2	98.2
CL00003|MOSTABUND|n=2216/246	blastn	1/10		Tychonema sp. SAG 23.89|KM019964-1521511|Bacteria; Terrabacteria grc	3.24814e-146	520.029	281	1	284	727	444	283	283	99.6	100.0
CL00004|MOSTABUND|n=1711/58	blastn	1/10		Chthoniobacter flavus (T); Ellin428|AY388649-497964|Bacteria; PVC group	3.41304e-111	403.69	218	1	284	784	501	262	262	92.3	100.0
CL00005|MOSTABUND|n=1616/76	blastn	1/10		Thermocrinis minervae (T); CR11|AM260555-381751|Bacteria; Aquificae; A	2.34965e-013	78.6796	42	6	92	785	699	73	73	83.9	30.6
CL00006|MOSTABUND|n=1320/54	blastn	1/10		Rhizobium sp. CAF431|FJ405377-573373|Bacteria; Proteobacteria; Alphap	1.95488e-143	510.796	276	6	284	784	506	278	278	99.6	98.2
CL00007|MOSTABUND|n=1312/28	blastn	1/10		Paenibacillus sp. 27-9|EU571199-519770|Bacteria; Terrabacteria group; Fi	4.23156e-140	499.716	270	6	284	688	410	276	276	98.9	98.2
CL00008|MOSTABUND|n=1150/79	blastn	1/10		Arthrobacter chlorophenolicus; L4|JQ277449-85085|Bacteria; Terrabacteri	5.43529e-144	512.642	277	5	284	720	441	279	279	99.6	98.6
CL00009|MOSTABUND|n=670/32	blastn	1/3		Arthrobacter antarcticus; R121|KF025983-494023|Bacteria; Terrabacteria	3.81938e-026	121.153	65	6	282	504	226	211	211	76.2	98.2
CL00010|MOSTABUND|n=632/1	blastn	1/10		Halospirulina sp. EF17(2012)|JX912466-1254278|Bacteria; Terrabacteria g	4.23156e-140	499.716	270	6	284	760	481	277	277	99.3	98.2
CL00011|MOSTABUND|n=625/16	blastn	1/10		Bacillus cereus; PASAU166|KM495607-1396|Bacteria; Terrabacteria group;	2.52879e-142	507.102	274	5	284	726	447	278	278	99.3	98.6
CL00012|MOSTABUND|n=617/28	blastn	1/10		Nocardioides kribbensis; PVS05|KP072762-305517|Bacteria; Terrabacteria	9.09516e-142	505.256	273	6	284	736	458	277	277	99.3	98.2

```
-query "C:\seed_data\working\Q23.fas" -strand both -task megablast -db "C:\LOCAL_BLAST_DATABASES\DATABASES\16SRDP180116" -out "C:\seed_data\working\R23.txt" -evalue 1E-5 -outfmt 5 -max_target_seqs 10 -num_threads 1
265468
Loading nodes...
Finished.
squence 23 finished ... output C:\seed_data\working\R23.txt
```

0%

SEQ TITLE		Accession	Description	Similarity[%]	Coverage[%	Taxid	Lineage	superkingdo	kingdom	subkingdom	superphy
CL00001|MOS			Halospirulina sp. E	98.6	100.0						
CL00002|MOS			Halospirulina sp. E	98.2	98.2						
CL00003|MOS			Tychonema sp. SA	99.6	100.0						
CL00004|MOSTABUND|n=1711/58	1/10		Chthoniobacter fla	92.3	100.0						
CL00005|MOSTABUND|n=1616/76	1/10		Thermocrinis miner	83.9	30.6						
CL00006|MOSTABUND|n=1320/54	1/10		Rhizobium sp. CAF	99.6	98.2						
CL00007|MOSTABUND|n=1312/28	1/10		Paenibacillus sp. 2	98.9	98.2						
CL00008|MOSTABUND|n=1150/79	1/10		Arthrobacter chlor	99.6	98.6						
CL00009|MOSTABUND|n=670/32	1/3		Arthrobacter anta	76.2	98.2						
CL00010|MOSTABUND|n=632/1	1/10		Halospirulina sp. E	99.3	98.2						
CL00011|MOSTABUND|n=625/16	1/10		Bacillus cereus; PA	99.3	98.6						
CL00012|MOSTABUND|n=617/28	1/10		Nocardioides kribb	99.3	98.2						
CL00013|MOSTABUND|n=612/13	1/10		Tetrasphaera sp. '	99.6	98.2						
CL00014|MOSTABUND|n=530/10											
CL00015|MOSTABUND|n=485/32	1/10		Desulfonatronum :	87.8	98.2						

0%

From text: |

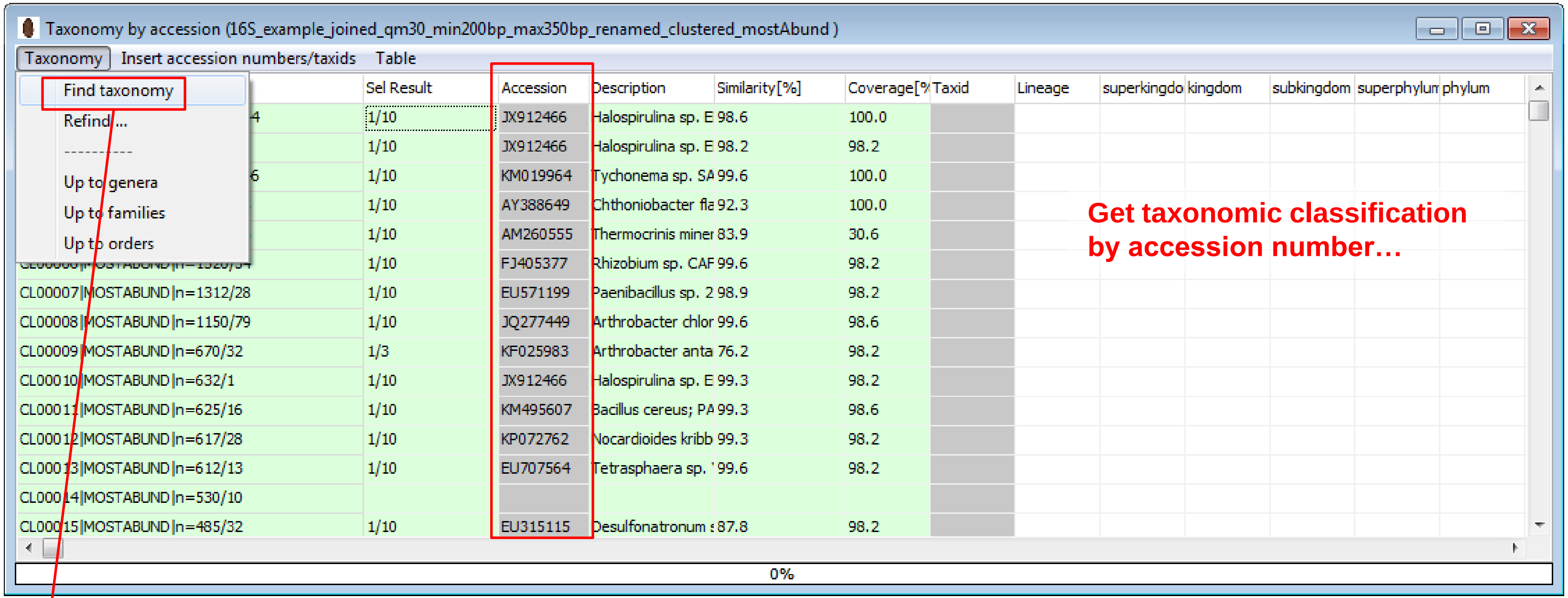
To text: -

☐ Include searched text

```
Halospirulina sp. EF17(2012)|JX912466-1254278|Bacteria
Halospirulina sp. EF17(2012)|JX912466-1254278|Bacteria
Tychonema sp. SAG 23.89|KM019964-1521511|Bacteria;
Chthoniobacter flavus (T); Ellin428|AY388649-497964|Ba
Thermocrinis minervae (T); CR11|AM260555-381751|Bact
Rhizobium sp. CAF431|FJ405377-573373|Bacteria; Prote
```

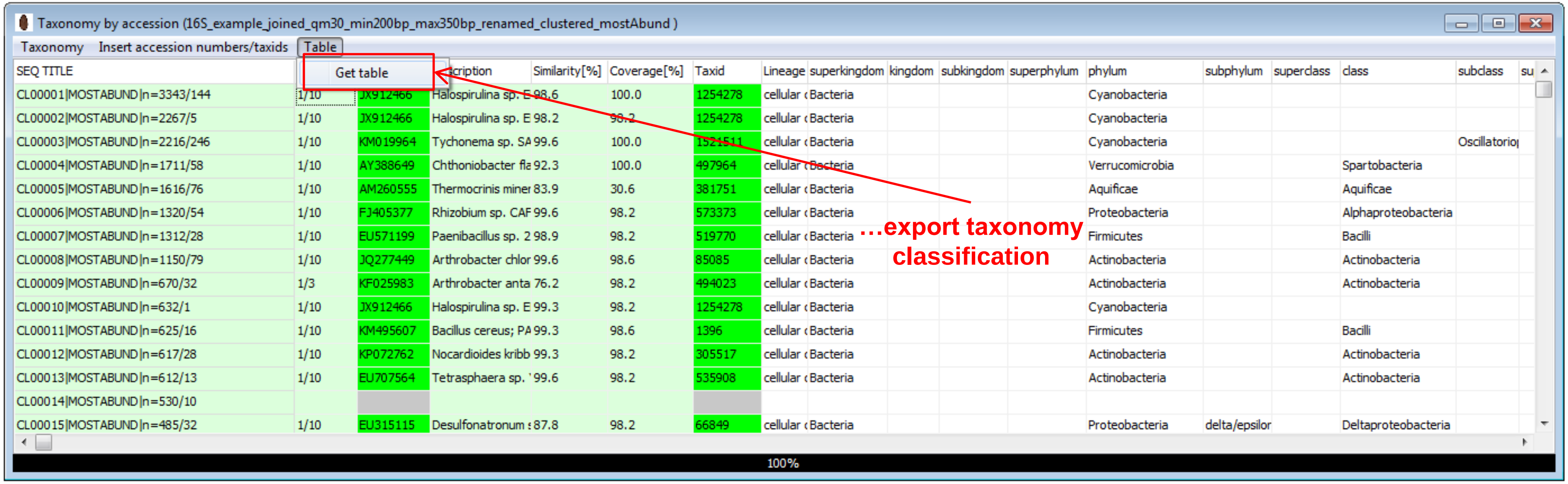
```
JX912466
KM019964
AY388649
AM260555
FJ405377
```

0%



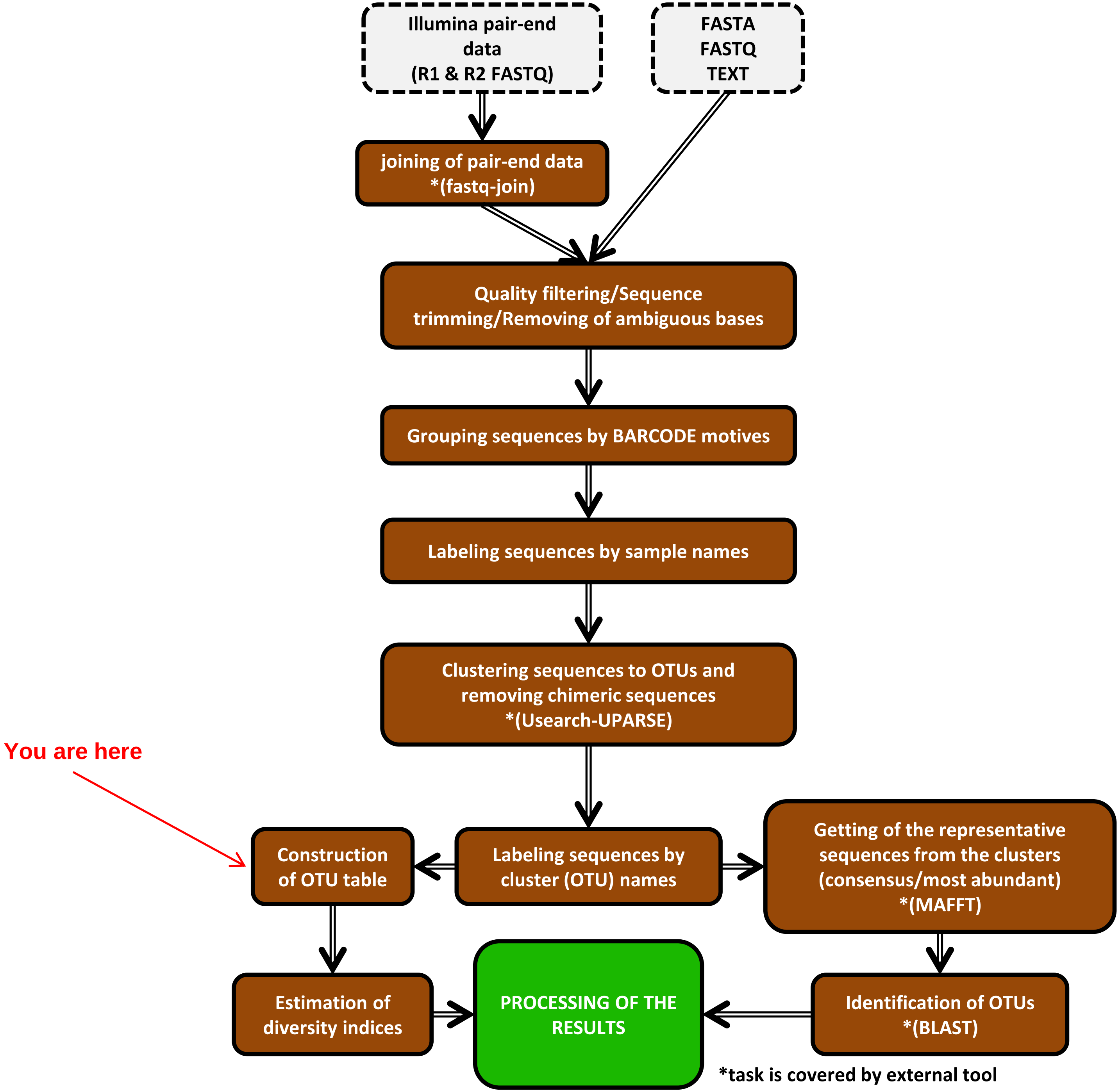
	Sel Result	Accession	Description	Similarity[%]	Coverage[%	Taxid	Lineage	superkingdo	kingdom	subkingdom	superphylun	phylum
4	1/10	JX912466	Halospirulina sp. E	98.6	100.0							
	1/10	JX912466	Halospirulina sp. E	98.2	98.2							
6	1/10	KM019964	Tychonema sp. SA	99.6	100.0							
	1/10	AY388649	Chthoniobacter fla	92.3	100.0							
	1/10	AM260555	Thermocrinis miner	83.9	30.6							
	1/10	FJ405377	Rhizobium sp. CAF	99.6	98.2							
CL00007|MOSTABUND|n=1312/28	1/10	EU571199	Paenibacillus sp. 2	98.9	98.2							
CL00008|MOSTABUND|n=1150/79	1/10	JQ277449	Arthrobacter chlor	99.6	98.6							
CL00009|MOSTABUND|n=670/32	1/3	KF025983	Arthrobacter anta	76.2	98.2							
CL00010|MOSTABUND|n=632/1	1/10	JX912466	Halospirulina sp. E	99.3	98.2							
CL00011|MOSTABUND|n=625/16	1/10	KM495607	Bacillus cereus; PA	99.3	98.6							
CL00012|MOSTABUND|n=617/28	1/10	KP072762	Nocardioides kribb	99.3	98.2							
CL00013|MOSTABUND|n=612/13	1/10	EU707564	Tetrasphaera sp. '	99.6	98.2							
CL00014|MOSTABUND|n=530/10												
CL00015|MOSTABUND|n=485/32	1/10	EU315115	Desulfonatronum :	87.8	98.2							

Get taxonomic classification by accession number…



SEQ TITLE			cription	Similarity[%]	Coverage[%]	Taxid	Lineage	superkingdom	kingdom	subkingdom	superphylum	phylum	subphylum	superclass	class	subclass	su
CL00001|MOSTABUND|n=3343/144	1/10	JX912466	Halospirulina sp. E	98.6	100.0	1254278	cellular (	Bacteria				Cyanobacteria					
CL00002|MOSTABUND|n=2267/5	1/10	JX912466	Halospirulina sp. E	98.2	98.2	1254278	cellular (	Bacteria				Cyanobacteria					
CL00003|MOSTABUND|n=2216/246	1/10	KM019964	Tychonema sp. SA	99.6	100.0	1521511	cellular (	Bacteria				Cyanobacteria				Oscillatorio	
CL00004|MOSTABUND|n=1711/58	1/10	AY388649	Chthoniobacter fla	92.3	100.0	497964	cellular (	Bacteria				Verrucomicrobia			Spartobacteria		
CL00005|MOSTABUND|n=1616/76	1/10	AM260555	Thermocrinis miner	83.9	30.6	381751	cellular (	Bacteria				Aquificae			Aquificae		
CL00006|MOSTABUND|n=1320/54	1/10	FJ405377	Rhizobium sp. CAF	99.6	98.2	573373	cellular (	Bacteria				Proteobacteria			Alphaproteobacteria		
CL00007|MOSTABUND|n=1312/28	1/10	EU571199	Paenibacillus sp. 2	98.9	98.2	519770	cellular (	Bacteria				Firmicutes			Bacilli		
CL00008|MOSTABUND|n=1150/79	1/10	JQ277449	Arthrobacter chlor	99.6	98.6	85085	cellular (	Bacteria				Actinobacteria			Actinobacteria		
CL00009|MOSTABUND|n=670/32	1/3	KF025983	Arthrobacter anta	76.2	98.2	494023	cellular (	Bacteria				Actinobacteria			Actinobacteria		
CL00010|MOSTABUND|n=632/1	1/10	JX912466	Halospirulina sp. E	99.3	98.2	1254278	cellular (	Bacteria				Cyanobacteria					
CL00011|MOSTABUND|n=625/16	1/10	KM495607	Bacillus cereus; PA	99.3	98.6	1396	cellular (	Bacteria				Firmicutes			Bacilli		
CL00012|MOSTABUND|n=617/28	1/10	KP072762	Nocardioides kribb	99.3	98.2	305517	cellular (	Bacteria				Actinobacteria			Actinobacteria		
CL00013|MOSTABUND|n=612/13	1/10	EU707564	Tetrasphaera sp. '	99.6	98.2	535908	cellular (	Bacteria				Actinobacteria			Actinobacteria		
CL00014|MOSTABUND|n=530/10																	
CL00015|MOSTABUND|n=485/32	1/10	EU315115	Desulfonatronum :	87.8	98.2	66849	cellular (	Bacteria				Proteobacteria	delta/epsilor		Deltaproteobacteria		

…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

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)

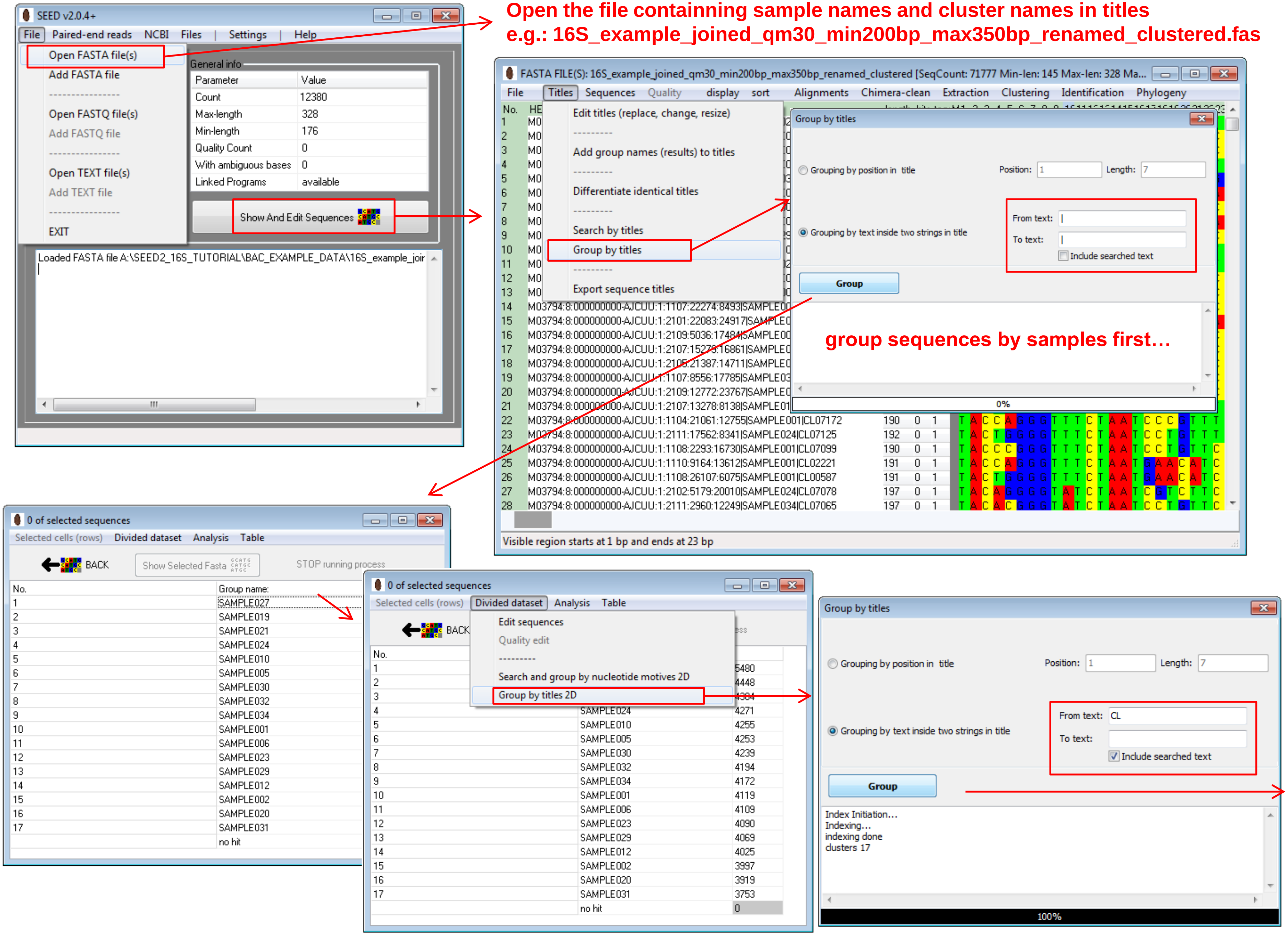
Estimation of diversity indices

PROCESSING OF THE RESULTS

Identification of OTUs *(BLAST)

*task is covered by external tool

OTU table construction…



Open the file containning sample names and cluster names in titles
e.g.: 16S_example_joined_qm30_min200bp_max350bp_renamed_clustered.fas

Parameter	Value
Count	12380
Max-length	328
Min-length	176
Quality Count	0
With ambiguous bases	0
Linked Programs	available

```
Loaded FASTA file A:\SEED2_16S_TUTORIAL\BAC_EXAMPLE_DATA\16S_example_joir
```

FASTA FILE(S): 16S_example_joined_qm30_min200bp_max350bp_renamed_clustered [SeqCount: 71777 Min-len: 145 Max-len: 328 Ma...

```
No.  HE
14   M03794:8:000000000-AJCUU:1:1107:22274:8493|SAMPLE00
15   M03794:8:000000000-AJCUU:1:2101:22083:24917|SAMPLE0
16   M03794:8:000000000-AJCUU:1:2109:5036:17484|SAMPLE00
17   M03794:8:000000000-AJCUU:1:2107:15279:16861|SAMPLE0
18   M03794:8:000000000-AJCUU:1:2105:21387:14711|SAMPLE0
19   M03794:8:000000000-AJCUU:1:1107:8556:17785|SAMPLE03
20   M03794:8:000000000-AJCUU:1:2109:12772:23767|SAMPLE0
21   M03794:8:000000000-AJCUU:1:2107:13278:8138|SAMPLE01
22   M03794:8:000000000-AJCUU:1:1104:21061:12755|SAMPLE001|CL07172   190  0  1  TACCAGGGTTTCTAATCCCGTTT
23   M03794:8:000000000-AJCUU:1:2111:17562:8341|SAMPLE024|CL07125    192  0  1  TACTGGGGTTTCTAATCCTGTTT
24   M03794:8:000000000-AJCUU:1:1108:2293:16730|SAMPLE001|CL07099    190  0  1  TACCCGGGTTTCTAATCCTGTTC
25   M03794:8:000000000-AJCUU:1:1110:9164:13612|SAMPLE001|CL02221    191  0  1  TACCAGGGTTTCTAATGAACATC
26   M03794:8:000000000-AJCUU:1:1108:26107:6075|SAMPLE001|CL00587    191  0  1  TACTGGGGTTTCTAATGAACATC
27   M03794:8:000000000-AJCUU:1:2102:5179:20010|SAMPLE024|CL07078    197  0  1  TACAGGGGTATCTAATCGTCTTC
28   M03794:8:000000000-AJCUU:1:2111:2960:12249|SAMPLE034|CL07065    197  0  1  TACACGGGTATCTAATCCTGTTC
```

Visible region starts at 1 bp and ends at 23 bp

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	SAMPLE027
2	SAMPLE019
3	SAMPLE021
4	SAMPLE024
5	SAMPLE010
6	SAMPLE005
7	SAMPLE030
8	SAMPLE032
9	SAMPLE034
10	SAMPLE001
11	SAMPLE006
12	SAMPLE023
13	SAMPLE029
14	SAMPLE012
15	SAMPLE002
16	SAMPLE020
17	SAMPLE031
	no hit

0 of selected sequences

No.		
1		5480
2		4448
3		4304
4	SAMPLE024	4271
5	SAMPLE010	4255
6	SAMPLE005	4253
7	SAMPLE030	4239
8	SAMPLE032	4194
9	SAMPLE034	4172
10	SAMPLE001	4119
11	SAMPLE006	4109
12	SAMPLE023	4090
13	SAMPLE029	4069
14	SAMPLE012	4025
15	SAMPLE002	3997
16	SAMPLE020	3919
17	SAMPLE031	3753
	no hit	0

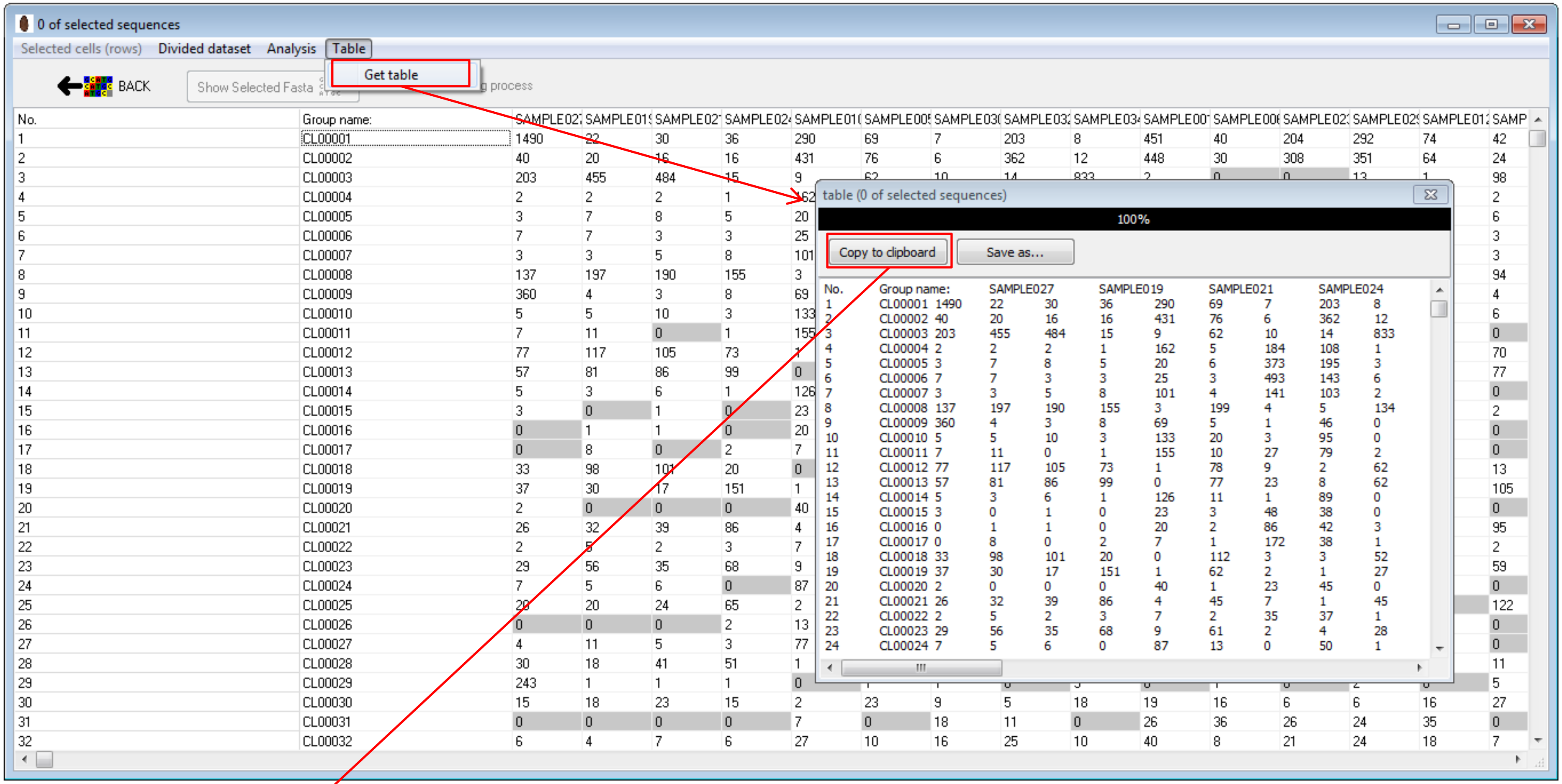
Group by titles

○ 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

```
Index Initiation…
Indexing…
indexing done
clusters 17
```

OTU table is done…



No.	Group name:	SAMPLE02	SAMPLE01	SAMPLE02	SAMPLE02	SAMPLE01	SAMPLE00	SAMPLE03	SAMPLE03	SAMPLE03	SAMPLE00	SAMPLE00	SAMPLE02	SAMPLE02	SAMPLE01	SAMP
1	CL00001	1490	22	30	36	290	69	7	203	8	451	40	204	292	74	42
2	CL00002	40	20	16	16	431	76	6	362	12	448	30	308	351	64	24
3	CL00003	203	455	484	15	9	62	10	14	833	2	0	0	13	1	98
4	CL00004	2	2	2	1	162										2
5	CL00005	3	7	8	5	20										6
6	CL00006	7	7	3	3	25										3
7	CL00007	3	3	5	8	101										3
8	CL00008	137	197	190	155	3										94
9	CL00009	360	4	3	8	69										4
10	CL00010	5	5	10	3	133										6
11	CL00011	7	11	0	1	155										0
12	CL00012	77	117	105	73	1										70
13	CL00013	57	81	86	99	0										77
14	CL00014	5	3	6	1	126										0
15	CL00015	3	0	1	0	23										2
16	CL00016	0	1	1	0	20										0
17	CL00017	0	8	0	2	7										0
18	CL00018	33	98	101	20	0										13
19	CL00019	37	30	17	151	1										105
20	CL00020	2	0	0	0	40										0
21	CL00021	26	32	39	86	4										95
22	CL00022	2	5	2	3	7										2
23	CL00023	29	56	35	68	9										59
24	CL00024	7	5	6	0	87										0
25	CL00025	20	20	24	65	2										122
26	CL00026	0	0	0	2	13										0
27	CL00027	4	11	5	3	77										0
28	CL00028	30	18	41	51	1										11
29	CL00029	243	1	1	1	0										5
30	CL00030	15	18	23	15	2	23	9	5	18	19	16	6	6	16	27
31	CL00031	0	0	0	0	7	0	18	11	0	26	36	26	24	35	0
32	CL00032	6	4	7	6	27	10	16	25	10	40	8	21	24	18	7

```
No.     Group name:     SAMPLE027       SAMPLE019       SAMPLE021       SAMPLE024
1       CL00001 1490    22      30      36      290     69      7       203     8
2       CL00002 40      20      16      16      431     76      6       362     12
3       CL00003 203     455     484     15      9       62      10      14      833
4       CL00004 2       2       2       1       162     5       184     108     1
5       CL00005 3       7       8       5       20      6       373     195     3
6       CL00006 7       7       3       3       25      3       493     143     6
7       CL00007 3       3       5       8       101     4       141     103     2
8       CL00008 137     197     190     155     3       199     4       5       134
9       CL00009 360     4       3       8       69      5       1       46      0
10      CL00010 5       5       10      3       133     20      3       95      0
11      CL00011 7       11      0       1       155     10      27      79      2
12      CL00012 77      117     105     73      1       78      9       2       62
13      CL00013 57      81      86      99      0       77      23      8       62
14      CL00014 5       3       6       1       126     11      1       89      0
15      CL00015 3       0       1       0       23      3       48      38      0
16      CL00016 0       1       1       0       20      2       86      42      3
17      CL00017 0       8       0       2       7       1       172     38      1
18      CL00018 33      98      101     20      0       112     3       3       52
19      CL00019 37      30      17      151     1       62      2       1       27
20      CL00020 2       0       0       0       40      1       23      45      0
21      CL00021 26      32      39      86      4       45      7       1       45
22      CL00022 2       5       2       3       7       2       35      37      1
23      CL00023 29      56      35      68      9       61      2       4       28
24      CL00024 7       5       6       0       87      13      0       50      1
```

…paste the table to excel…

Combine the obtained information…

blast identification

Taxonomy classification

OTU table

OTU	**Organism**	**Phylum**	**SAMPLE027**	**SAMPLE019**	**SAMPLE021**	**SAMPLE024**	**SAMPLE010**	**SAMPLE005**	**SAMPLE030**	**SAMPLE032**	**SAMPLE034**
CL00001	Halospirulina sp. EF17(2012)	Cyanobacteria	1490	22	30	36	290	69	7	203	8
CL00002	Halospirulina sp. EF17(2012)	Cyanobacteria	40	20	16	16	431	76	6	362	12
CL00003	Tychonema sp. SAG 23.89	Cyanobacteria	203	455	484	15	9	62	10	14	833
CL00004	Chthoniobacter flavus (T); Ellin428	Verrucomicrobia	2	2	2	1	162	5	184	108	1
CL00005	Thermocrinis minervae (T); CR11	Aquificae	3	7	8	5	20	6	373	195	3
CL00006	Rhizobium sp. CAF431	Proteobacteria	7	7	3	3	25	3	493	143	6
CL00007	Paenibacillus sp. 27-9	Firmicutes	3	3	5	8	101	4	141	103	2
CL00008	Arthrobacter chlorophenolicus; L4	Actinobacteria	137	197	190	155	3	199	4	5	134
CL00009	Arthrobacter antarcticus; R121	Actinobacteria	360	4	3	8	69	5	1	46	0
CL00010	Halospirulina sp. EF17(2012)	Cyanobacteria	5	5	10	3	133	20	3	95	0
CL00011	Bacillus cereus; PASAU166	Firmicutes	7	11	0	1	155	10	27	79	2
CL00012	Nocardioides kribbensis; PVS05	Actinobacteria	77	117	105	73	1	78	9	2	62
CL00013	Tetrasphaera sp. YC6726	Actinobacteria	57	81	86	99	0	77	23	8	62
CL00015	Desulfonatronum sp. Su2	Proteobacteria	3	0	1	0	23	3	48	38	0
CL00016	Rhodoplanes sp. 303	Proteobacteria	0	1	1	0	20	2	86	42	3
CL00017	Bacillus sp. PS1-5	Firmicutes	0	8	0	2	7	1	172	38	1
CL00018	Hydrogenobaculum sp. Y04AAS1	Aquificae	33	98	101	20	0	112	3	3	52
CL00019	Micrococcus endophyticus; DT20X	Actinobacteria	37	30	17	151	1	62	2	1	27
CL00020	Gaiella occulta (T); F2-233	Actinobacteria	2	0	0	0	40	1	23	45	0
CL00021	Pseudomonas sp. III-116-17	Proteobacteria	26	32	39	86	4	45	7	1	45
CL00022	Gaiella occulta (T); F2-233	Actinobacteria	2	5	2	3	7	2	35	37	1

+ additional metadata

Process the results…