

Exercise 1 From multiple alignment to motif search.

- Download file upstreams.fasta from seminar folder.
- Run T-Coffee (<http://tcoffee.crg.cat/apps/tcoffee/do:mcoffee>) to get MSA
- Look up for most conservative region. Copy this region to the most conservative region to separate alignment file:

>1 GGTCAATTCACCTGCCTT

>2 GGCCAATTTACGGCCTT

>3 GGTCAGTTCACGGCATT

>4 TCCTAATTTACAGCAGC

>5 GGTCAGTTCACGGCATT

>6 GGCCAATTTACTGCGTT

>7 AACCAGCTTGAGACAGC

Consensus sequence and PWM from an alignment

- Go to RSAT web tool: <http://embnet.ccg.unam.mx/rsat/>, Procaryotes RSAT, select Matrix tools from left menu, then convert matrix.
- Select input data type “sequences”, paste alignment in fasta format to the field. Then check the boxes “consensus”, “counts”, “frequencies”, “weights”.
- Press “GO”

Output

Output:

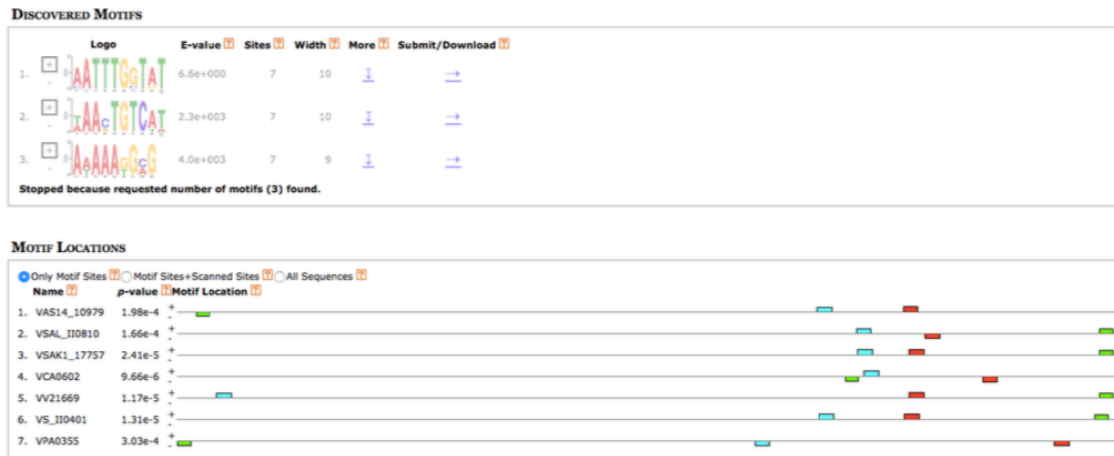
```

; Matrix type: counts
; Pos 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
;-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
a 1 1 0 0 7 4 0 0 0 6 1 1 1 0 4 0 0
c 0 1 4 6 0 0 1 0 3 0 6 0 0 7 2 0 2
g 5 5 0 0 0 3 0 0 0 1 0 4 6 0 1 2 0
t 1 0 3 1 0 0 6 7 4 0 0 2 0 0 0 5 5
//
;
; Matrix type: frequencies
;P 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
;-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
a 0.2 0.2 0.0 0.0 0.9 0.5 0.0 0.0 0.0 0.8 0.2 0.2 0.2 0.0 0.5 0.0 0.0
c 0.0 0.2 0.5 0.8 0.0 0.0 0.2 0.0 0.4 0.0 0.8 0.0 0.0 0.9 0.3 0.0 0.3
g 0.7 0.7 0.0 0.0 0.0 0.4 0.0 0.0 0.0 0.2 0.0 0.5 0.8 0.0 0.2 0.3 0.0
t 0.2 0.0 0.4 0.2 0.0 0.0 0.8 0.9 0.5 0.0 0.0 0.3 0.0 0.0 0.0 0.7 0.7
//
;
; Matrix type: weights
;P 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
;-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
a -0.5 -0.5 -2.1 -2.1 1.3 0.8 -2.1 -2.1 -2.1 1.1 -0.5 -0.5 -2.1 0.8 -2.1 -2.1
c -2.1 -0.5 0.8 1.1 -2.1 -2.1 -0.5 -2.1 0.5 -2.1 1.1 -2.1 -2.1 1.3 0.1 -2.1 0.1
g 1.0 1.0 -2.1 -2.1 -2.1 0.5 -2.1 -2.1 -2.1 -0.5 -2.1 0.8 1.1 -2.1 -0.5 0.1 -2.1
t -0.5 -2.1 0.5 -0.5 -2.1 -2.1 1.1 1.3 0.8 -2.1 -2.1 0.1 -2.1 -2.1 -2.1 1.0 1.0
//
; consensus ggycArTtYAcKGCmky
; consensus.rc RMKGCMGTAAATGRCC
convert-matrix_2017-11-14.161338_HgWbEH_input_m1
...matrix_2017-11-14.161338_HgWbEH_input_m1 Rev. cpl.

```

Exercise 2 MEME (<http://meme-suite.org/>)

- Upload the upstreams.fasta file, press “
- Advanced options”, set “How wide can motifs be?” from 5 to 10.
- Submit and wait for results.
- Output

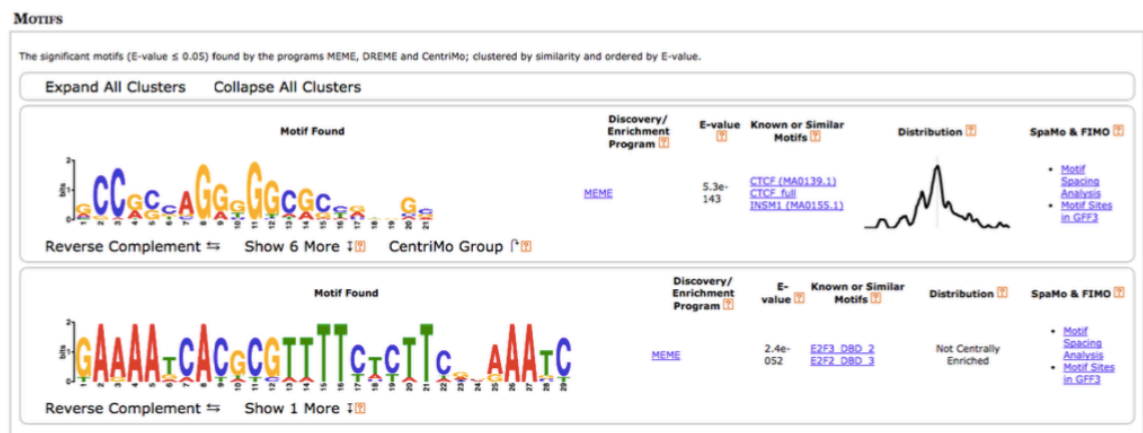


Find what this motif could be?

- TOMTOM <http://meme-suite.org/tools/tomtom>

Exercise 3. Motif search in ChIP-Seq data

- Download peaks.fasta (This is fasta file with chicken ChIP-Seq data for CTCF architecture protein.)
- MEME-ChIP for motif search (<http://meme-suite.org/> -> MEME-ChIP)



According to the suggestion, it is the CTCF motif.

The second one is E2f3 binding factor colocalized with CTCF. It is not necessarily associated with peaks center.