**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 GGTCAATTCACTGCCTT

>2 GGCCAATTTACGGCCTT

>3 GGTCAGTTCACGGCATT

>4 TCCTAATTTACAGCAGC

>5 GGTCAGTTCACGGCATT

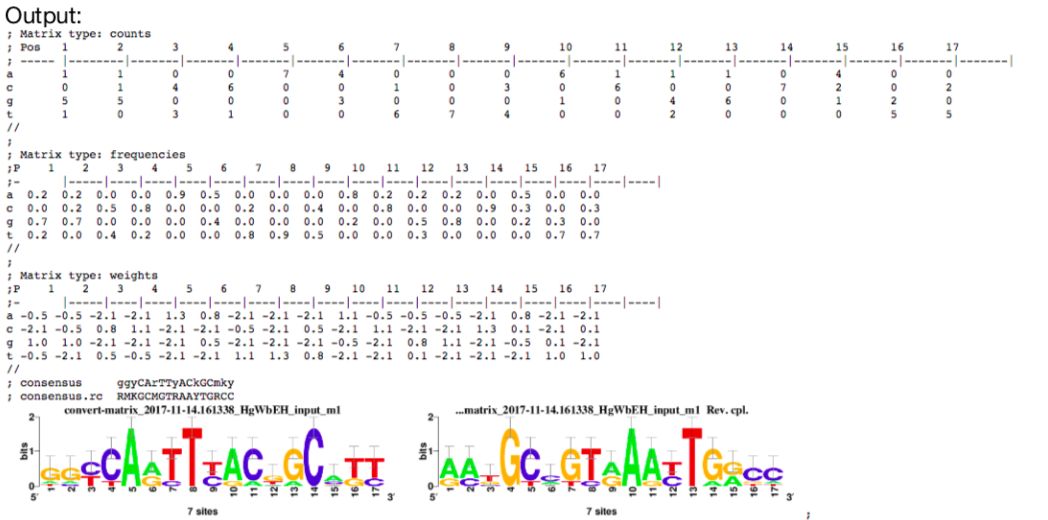
>6 GGCCAATTTACTGCGTT

>7 AACCAGCTTGAGACAGC

Consensus sequence and PWM from an alignment

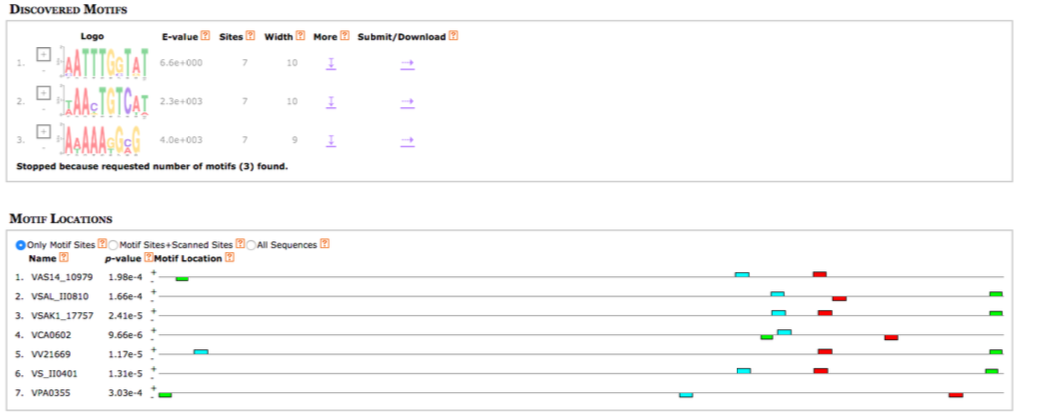
* Go to RSAT web tool: http://embnet.ccg.unam.mx/rsat/, Procaryotes RSAT, select Matrix tools from left menu, then convert matrix.
* Select input data type “sequences”, paste alignment in fasta format to the field. Then check the boxes “consensus”, “counts”, “frequencies”, “weights”.
* Press “GO”

Output



**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

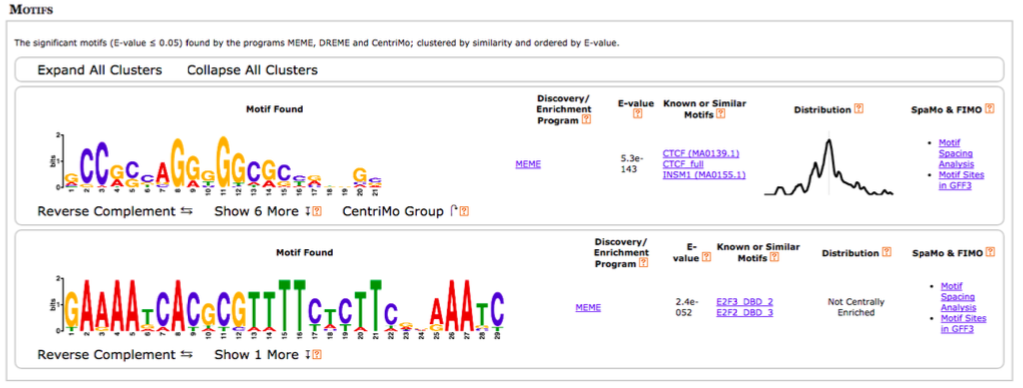


Find what this motif could be?

* 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.