

# **Bioninformatics and Comparative Genome Analyses**

<http://www.pasteur.fr/~tekaia/ACBCGA.html>

## **Genome comparisons : practical sessions**

### **Markov Cluster algorithm or MCL algorithm**

**Stijn van Dongen**

**Mcl web page :**

<http://micans.org/mcl/>

**Animation example :**

<http://micans.org/mcl/ani/mcl-animation.html>

**(use s/d : step forward/backward)**

**Course :**

<http://www.pasteur.fr/~tekaia/BCGA/TALKS/svd/whale2/index.html>

<http://www.pasteur.fr/~tekaia/BCGA/TALKS/svd/lizard/index.html>

<http://www.pasteur.fr/~tekaia/BCGA/TALKS/svd/lizard/26.html>

## 14) classification of non unique genes in a given genome using mcl see mcl.scr script

/home/usr/tekaia/acbcga/ACBCGA/bin (available scripts)  
/home/usr/tekaia/acbcga/ACBCGA/data (available data)  
/home/usr/tekaia/acbcga/ACBCGA/genanal (precomputed results)

Consider the genomes MYLE and MYUL.

Make your own directory: GCOMP

```
mkdir MYLE
mkdir MYLEseqnew
cd MYLEseqnew
```

```
cp /home/user/tekaia/acbcga/ACBCGA/genanal/MYLE/MYLEseqnew/allmyleseqnew .
cp /home/user/tekaia/acbcga/ACBCGA/genanal/MYLE/MYLEseqnew/bestmyleseqnew .
```

-create a file including HS (highly significant) hits in allhitsmyle :

```
grep -w HS allmyleseqnew > allmyleseqnew.HS
```

- create a file including sequence identifications listed according to their corresponding frequencies and sorted in decreasing order.

```
cat allmyleseqnew.HS | nom.pl | sort > seqnames
freqsortednames.pl seqnames
```

output: seqnames.freq

```
mv seqnames freqmyle.myle
```

```
sort -k 2 -n -r freqmyle.myle | more (to see the output)
```

```
sort -k 2 -n -r freqmyle.myle | nom.pl > nomorf
```

-Make a directory MCL.

```
cd MCL
ln -s ../nomorf
ln -s ../allmyleseqnew.HS
```

Procedure to build clusters of non unique proteins:

- 1- *mcltabform.pl nomorf > MYLE.tab*
- 2- *mclall2num.pl MYLE.tab allmyleseqnew.HS > allmylenum*
- 3- *mclall2cmi.pl allmylenum MYLE.tab & (output allmylenum.cmi )*
- 4- *mcl allmylenum.cmi -I 3.0 -progress 100 -o MYLE.clusters & (clustering)*
- 5- *mcltribefamilies.pl MYLE.clusters MYLE.tab > MYLE.clusters-tribe &*
- 6- *mclclustsize.pl MYLE.clusters-tribe > MYLE.mclclusters &*  
(Note change: *\$size=\$NTAB[1]; par \$size=\$NTAB[0];* )

#renumber classes in increasing order for each size

- 7- *renumclass.pl MYLE.mclclusters &*

Browse the output file look at the different families.

```
mkdir MYUL
```

```
mkdir MYULseqnew
```

```
cd MYULseqnew
```

```
cp /home/user/tekaia/acbcga/ACBCGA/genanal/MYUL/MYULseqnew/allmyulseqnew .
cp /home/user/tekaia/acbcga/ACBCGA/genanal/MYUL/MYULseqnew/bestmyulseqnew .
```

Construct classes of non unique proteins in MYUL.  
(follow the steps used for MYLE).

15 ) MEME/MAST (clusters of proteins).  
See meme.scr script

**Fredj Tekaiia (tekaia@pasteur.fr)**