



Weichun Huang

OMiMa system is a computational tool for motif identification in biological sequences. It is based on our Optimized Mixed Markov Model, which incorporates both local and long-range dependencies within a motif to improve prediction accuracy. In comparison with other leading methods, OMiMa often shows better prediction performance, particularly when available training samples are limited. For more details of the underlying methods of OMiMa system, please see our following publication:

Weichun Huang, David M Umbach, Uwe Ohler, Leping Li. Optimized mixed Markov models for motif identification. *BMC Bioinformatics* 2006, **7**:279

OMiMa system can be used to search motifs in different biological sequences including, but not limited to, DNA, RNA and protein sequences. OMiMa is also capable of searching for multiple different motifs simultaneously. The outputs of OMiMa include both the scores and locations of motif sites as well as their distributions. OMiMa is a command line tool with two standard usages:

OMiMa input-seq-file configure-file

The configure-file is used to change the default parameter values of OMiMa.

OMiMa input-seq-file

In this case, the standard configure file named `OMiMa.conf` must exist. An example of OMiMa configure file is given in the following.

An example of OMiMa configure file

```
# 1= scan original sequence from left to right, 2=scan original and complementary strands
# of DNA, 4=scan both strands in both directions. Default=2
WAY 2

# The directory of training motif files with extension ".mf". The first line of a motif
# file must be ">motifName" with the remaining lines are the alignment of motif sequences
motifDir /To/Train/motif/directory

# The false motif or background sequence file. The bases for masking repeats should be
# removed. The file should be in the FASTA format
bgseqFile /Dir/background_seq.fa

# The output directory. All output files except the main one are put in this directory
outDir /output/directory

# The main output file
outFile /directory/motif_score.out

# The order of the Markov model. In case of selecting the best model, this is the maximum
```

```

# Markov order. Default=2
MaxMcOrder 1

# The percentage of training data as the pseudo count for flat Dirichlet prior.
mcPrior 0.1

# The cutoff of prediction accuracy for training data itself (containing true motif). It
# is used to select cutoff value for positive sites. Default=1.00
testRate 1.00

# Markov model structure: 1=linear, 2=circle, 0=select the best among all. Default=0
model 1

# Score calculation method. 1=log likelihood ratio, 0=log likelihood. Default=1
LogRatio 1

# Print out model selection information, such as the Chi-square tests and other model
# selection information. Default=0, i.g. does not output
printModel 0

# The model selection criterion: A=AIC, B=BIC. Default=A
criteria

# The interval of histogram for motif scores. Default=0.1
hisInterval

# The minimum cutoff value of  $\log P(s|M_s)$ .
minLogProb

# The minimum cutoff value of  $\log(P(s|M_s)/P(s|M_b))$ , default=2.0
minLogRatio

#-----
#The cutoff thresholds for different motifs, with the following format
motif_name_1 cutoff_1
motif_name_2 cutoff_2

```

Copyright© 2005, *Weichun Huang*