

# **BEST- Binding-site Estimation Suite of Tools**

---

**<http://www.cs.uga.edu/~che/BEST/>**

**<http://www.fas.harvard.edu/~junliu/BEST>**

**Che D., Jensen, S., Cai, L. and Liu JS. BEST: Binding-site Estimation Suite of Tools.**

**To be submitted to the Journal of Bioinformatics**

**Version 1.0 December 2004**

**Dongsheng Che**

**Department of Computer Science**

**The University of Georgia**

**Athens, GA 30602**

**USA**

# Contents

## 1. Overview

## 2. Obtaining and installation of BEST

### 2.1. Obtaining BEST

### 2.2. Unpacking BEST

### 2.3. Installing BEST

## 3. Tutorial

### 3.1. Inputs

### 3.2. Outputs

### 3.3. A simple example

## 4. References

# 1. Overview

Binding-site Estimation Suite of Tools (BEST) is a computational software package which includes four motif-finding programs: AlignACE (Roth et al., 1998), BioProspector (Liu et al., 2001), Consensus (Hertz and Stormo, 1999), and MEME (Bailey and Elkan, 1994), and the optimization program BioOptimizer configured for each of these programs (Jensen and Liu 2004).

BEST was compiled on Linux, and thus it can only be run on Linux machines.

Contact: [dsche@uga.edu](mailto:dsche@uga.edu)

## 2. Obtaining and Installation of BEST

### 2.1. Obtaining BEST

BEST can be downloaded at:

<http://www.cs.uga.edu/~che/BEST/> and <http://www.fas.harvard.edu/~junliu/BEST/>

The distribution contains:

- this documentation (as a PDF file and as a text file)
- INSTALL – a simple installation script
- BEST – an executable program of BEST
- bin – including all motif-finding programs
- data – including a crp dataset for test
- images – supported images and manuals for BEST

### 2.2. Unpacking BEST

After finishing downloading, type

```
gunzip best1.0.tar.gz
```

followed by:

```
tar -xvf best1.0.tar
```

### 2.3. Installing BEST

After finishing unpacking BEST, go to BEST1.0 directory, and type './INSTALL' command, then answer 'no' during the process of installation.

## 3. Tutorial

### 3.1. Inputs

1. Required parameters only include input sequence and motif width field.
2. Input sequence should be FASTA format, BEST can automatically convert FASTA format to other formats satisfying individual programs. For example, 'Consensus' program only accepts consensus format.
3. Both single and multiple motif values are acceptable. For multiple motif width, the following input formats are acceptable: a). 9-12, b) 9, 10, 11, 12 and c). 9 10 11 12.
4. Detailed parameter information and usage of individual programs can be found in the 'help' menu of the main window of BEST.

### 3.2. Outputs

1. All output files are stored in the same directory where input sequence file are stored. Output files can be traced for individual programs. For example, if the input sequence is CRP.fas, and the motif width range is 20-24, then the corresponding output format for individual programs are:

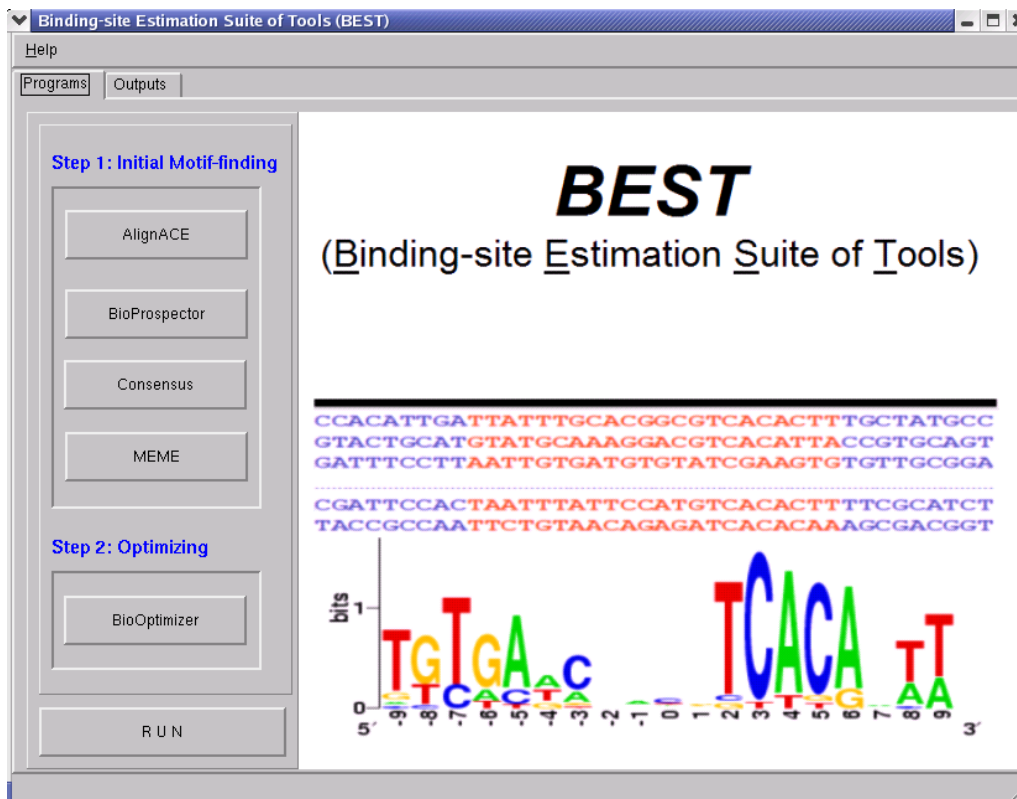
- a). AlignACE: CRP.aa\_20, CRP.aa\_21, ... CRP.aa\_24
- b). BioProspector: CRP.biop\_20, CRP.biop\_21, ... CRP.biop\_24
- c). Consensus: CRP.con\_20, CRP.con\_21, ... CRP.con\_24
- d). Meme: CRP.meme\_20, CRP.meme\_21, ... CRP.meme\_24
- e). BioOptimizer:  
For AlignACE: CRP.aa\_20.opt.all, CRP.aa\_20.opt.sum, CRP.aa\_20.opt.best  
CRP.aa\_21.opt.all, CRP.aa\_21.opt.sum, CRP.aa\_21.opt.best  
...  
CRP.aa\_24.opt.all, CRP.aa\_24.opt.sum, CRP.aa\_24.opt.best  
similarly for BioProspector, Consensus and Meme.

2. When running both motif-finding programs and the BioOptimizer program together in BEST, BEST outputs a summary table after evaluating all predicted sites. The summary file can also be in the same directory

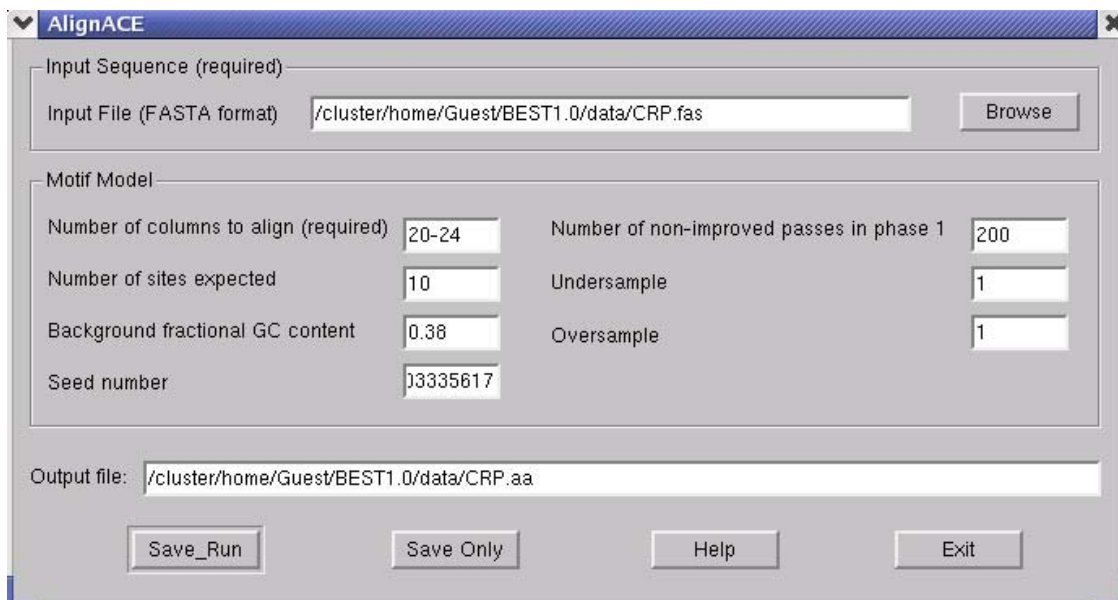
under the input sequence file is stored. For example, if the input sequence file is 'CRP.fas', the summary file will be 'CRP\_BEST.Summary'.

### 3.1. A simple example

1. Start BEST by typing BEST. This command opens the main window of BEST, which contains four buttons for setting up the GUI of the four motif-finding programs, one button for the GUI of BioOptimizer, and one global 'run' button to run all programs.



2. Set up required fields in GUIs of motif-finding programs. Open GUIs by clicking buttons. For example, the AlignACE program is selected for motif-finding by clicking the 'AlignACE' button. Then fill out all required fields, including the input sequence file in FASTA format and motif width. We use CRP dataset as the example, with the width range from 20 to 24. Save the parameters by clicking 'save' button. We can set up other motif-finding programs similarly.



3. Set up the GUI of BioOptimizer since it usually can improve the prediction accuracy based on the predicted results of motif-finding programs. After saving the GUI for BioOptimizer, all buttons in the main window turn green, indicating all programs have been set up.

4. Click the 'run' button in the main window. When all invoked programs finish running, a summary of all predicted results from different programs is displayed as the followings. At this point, you might open a specific file to look into the detailed information based on this summary table.

	Initial score	W	A	Consensus	Final score	Null score	W	A	Consensus	From File
1	-5154.16	21	11	gTGTGaaaTcggtcacatAat	-5103.05	-5139.3	22	38	ataTGTGaaactagTCACAtat	/cluster/home/Guest/BEST1.0/data/CRP.biop_21.opt
2	-5149.13	20	16	tatgTGatcgattTCACAAA	-5103.09	-5139.25	22	39	ataTGTGaaactagatCACAAat	/cluster/home/Guest/BEST1.0/data/CRP.biop_20.opt
3	-5161.31	22	18	aaTgTGatcgaggTCACActta	-5104.5	-5139.16	22	38	ttaTgTGaaactagTCACAAat	/cluster/home/Guest/BEST1.0/data/CRP.con_22.opt.s
4	-5160.31	24	18	aaTgTGatcgaggTCACActata	-5104.52	-5139.13	22	38	ttaTgTGaaactagTCACAAat	/cluster/home/Guest/BEST1.0/data/CRP.con_24.opt.s
5	-5159.57	21	18	aTgTGatcgaggTCACActta	-5104.54	-5139.16	22	38	ttaTgTGaaactagTCACAAat	/cluster/home/Guest/BEST1.0/data/CRP.con_21.opt.s

## 4. References

Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, Stanford, CA, AAAI Press, Bethesda, MD, pp. 28–36.

Hertz, G.Z. and Stormo, G.D. (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics*, **15**, 563–577.

Jensen, S.T. and Liu, J.S. (2004) BioOptimizer: a Bayesian scoring function approach to motif discovery. *Bioinformatics*, **20**, 1557–1564

Liu, X., Brutlag, D.L. and Liu, J.S. (2001) BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac. Symp. Biocomput.*, **6**, 127–138.

Roth, F.R., Hughes, J. D., Estep, P. E. and Church G. M. (1998). Finding DNA regulatory motifs within unaligned non-coding sequences clustered by whole-Genome mRNA quantitation. *Nature Biotechnology* 16:939-45.