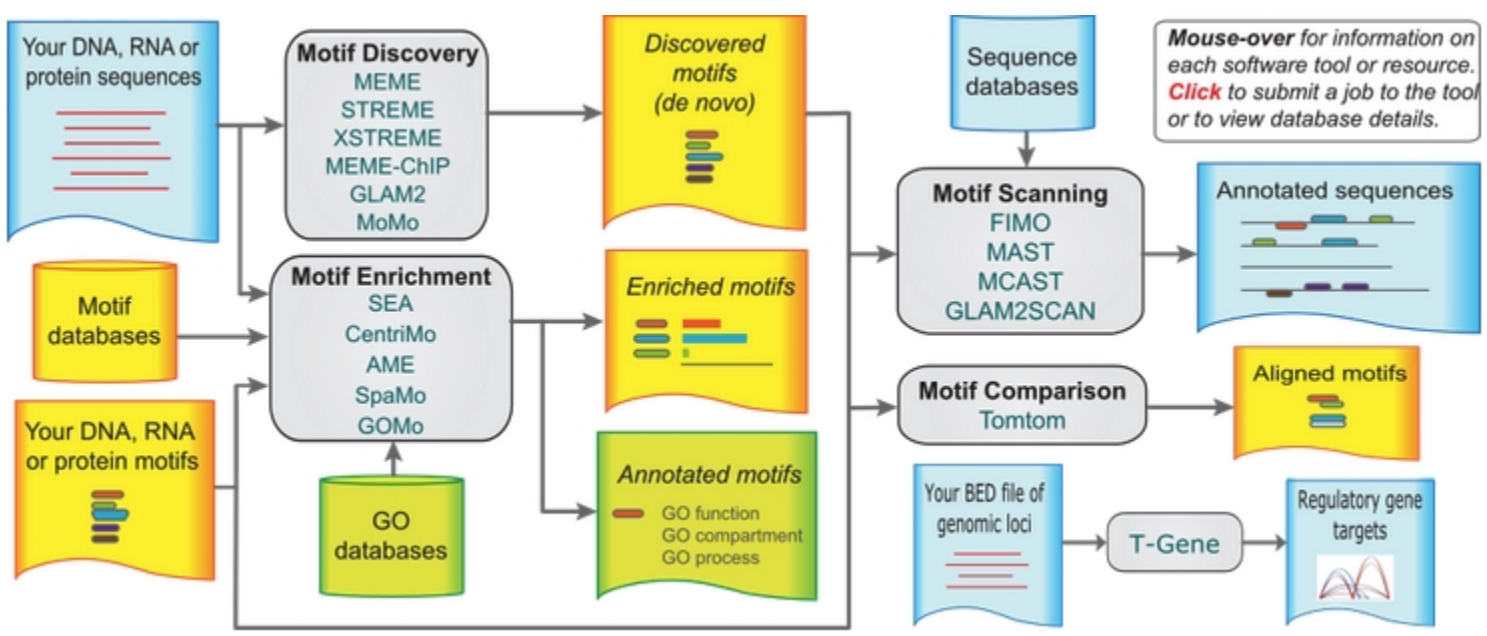
The MEME Suite: Motif-Based Sequence Analysis Tools

Tech ID: 20911 / UC Case 2010-808-0

BACKGROUND

The MEME Suite allows you to:

- Discover motifs using MEME, STREME, XSTREME, MEME-ChIP, GLAM2, or MoMo on groups of related DNA, RNA or protein sequences.
- Predict enrichment of motifs (or motif pairs) in DNA, RNA or protein sequences with SEA, CentriMo, AME and SpaMo (DNA only).
- Search sequence databases using motifs using FIMO, MAST, MCAST and GLAM2Scan.
- Compare a motif to all motifs in a database of motifs using Tomtom.
- Associate motifs with gene ontology terms via their putative target genes using GOMo.
- Predict regulatory links between chromosome regions and genes using T-Gene.
- Many of the above tools also allow for user-defined sequence alphabets.



TECHNOLOGY DESCRIPTION

The MEME Suite is an integrated collection of tools for discovering and characterizing sequence motifs in collections of DNA, RNA or protein sequences.

The flagship program in the suite is MEME, which finds motifs in unaligned collections of DNA and sequence motifs. Initially described in 1994, MEME has been continually maintained and improved in the ensuing years and is now probably the most widely used motif discovery tool, with over 32,000 citations according to Google Scholar and over 65,000 users per year submitting over 1500 queries per day to the MEME Suite website. MEME has been used to discover novel motifs encoding many types of biological signals, including transcription factor binding sites, long/short interspersed nuclear elements (LINES and SINES), protein family markers, and allergenic protein markers, to name a few.

In addition, the MEME Suite contains tools for searching sequence databases using single motifs or groups of motifs discovered by MEME or taken from curated motif databases. This kind of motif search has many uses in biology, including identifying transcription factor binding sites,

CONTACT

University of California, San Diego
Office of Innovation and Commercialization
innovation@ucsd.edu
tel: 858.534.5815.

OTHER INFORMATION

CATEGORIZED AS

- Research Tools
  - Nucleic Acids/DNA/RNA

RELATED CASES

2010-808-0

locating cis-regulatory modules, and predicting protein family members and their interrelationships. The suite also contains a tool, Tomtom, for comparing newly discovered DNA motifs to known transcription factor binding site motif databases. Tomtom allows the biologist to make predictions about the identity of transcription factors regulating a set of genes.

The MEME Suite comes packaged with its own Web server, providing an intuitive, integrated, graphical-user interface. The interface allows the output of one tool to be sent to another tool by simply clicking a button on the Web-based output. This design allows the biologist to, for example, use motifs discovered by MEME to easily search a sequence database or a database of known motifs. The software also comes with extensive documentation and tutorials and is supported by an active development team and user community.

Licensees can install the MEME Suite either from source or as a pre-packaged Docker image. System requirements are a modern computer running MacOS, Linux or Windows.

APPLICATIONS

MOTIF DISCOVERY

- *MEME* discovers novel, ungapped motifs (recurring, fixed-length patterns) in your sequences. MEME splits variable-length patterns into two or more separate motifs.
- *STREME* discovers ungapped motifs (recurring, fixed-length patterns) that are enriched in your sequences or relatively enriched in them compared to your control sequences.
- *XSTREME* performs comprehensive motif analysis (including motif discovery) on sequences where the motif sites can be anywhere in the sequences. The input sequences may be of any length, and their lengths may vary.
- *MEME-ChIP* performs comprehensive motif analysis (including motif discovery) on sequences where the motif sites tend to be centrally located, such as ChIP-seq peaks. The input sequences should be centered on a 100 character region expected to contain motifs, and each sequence should ideally be around 500 letters long.
- *GLAM2* discovers novel, gapped motifs (recurring, variable-length patterns) in your DNA or protein sequences.
- *MoMo* discovers sequence motifs associated with different types of protein post-translational modifications (PTMs). The program takes as input a collection of PTMs identified using protein mass spectrometry. For each distinct type of PTM, MoMo uses one of three algorithms to discover motifs representing amino acid preferences flanking the modification site.

Motif Enrichment

- *SEA* identifies known or user-provided motifs that are relatively enriched in your sequences compared with shuffled sequences or your control sequences.
- *CentriMo* identifies known or user-provided motifs that show a significant preference for particular locations in your sequences. CentriMo can also show if the local enrichment is significant relative to control sequences.
- *AME* identifies known or user-provided motifs that are either relatively enriched in your sequences compared with control sequences, that are enriched in the first sequences in your input file, or that are enriched in sequences with small values of scores that you can specify with your input sequences.
- *SpaMo* identifies significantly enriched spacings in a set of sequences between a primary motif and each motif in a set of secondary motifs. Typically, the input sequences are centered on ChIP-seq peaks, and are each 500bp long.
- *GOMo* scans all promoters using nucleotide motifs you provide to determine if any motif is significantly associated with genes linked to one or more Genome Ontology (GO) terms. The significant GO terms can suggest the biological roles of the motifs.

Motif Scanning

- *FIMO* scans a set of sequences for individual matches to each of the motifs you provide.
- *MAST* searches sequences for matches to a set of motifs, and sorts the sequences by the best combined match to all motifs.
- *MCAST* searches sequences for clusters of matches to one or more nucleotide motifs.
- *GLAM2Scan* searches sequences for matches to gapped DNA or protein GLAM2 motifs.

Motif Comparison

- *Tomtom* compares one or more motifs against a database of known motifs (e.g., JASPAR). Tomtom will rank the motifs in the database and produce an alignment for each significant match.

Gene Regulation

▶ *T-Gene* predicts regulatory links between each of the loci (genomic locations, typically TF ChIP-seq peak locations) that you provide and each of the organism's genes.

Other Tools

▶ The MEME Suite includes many additional tools for manipulating biological sequence and motif files.

ADDITIONAL INFORMATION

Licensing Information:

Click on the following link to review a copy of our commericial-use license for the MEME Suite.

Please use the following link to a Word version of the agreement. Please make any suggested changes using track changes and email to OIC-MEMESuite@ucsd.edu. Once received someone from our licensing team will get back to you as soon as possible.

If you have no suggested edits and are ready to complete and sign the license agreement, simply click on the following URL link to be directed to our DocuSign Powerform. Click Here

If you have any questions regarding this agreement, please reach out to our licensing team at OIC-MEMESuite@ucsd.edu.

RELATED MATERIALS

▶ Further information on the MEME Suite of Motif-based sequence analysis tools.

University of California, San Diego
Office of Innovation and Commercialization
9500 Gilman Drive, MC 0910, ,
La Jolla,CA 92093-0910

Tel: 858.534.5815
innovation@ucsd.edu
https://innovation.ucsd.edu
Fax: 858.534.7345