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 or GLAM2 on groups of related DNA or protein sequences.
▶ Search sequence databases using motifs.
▶ Compare a motif to all motifs in a database of motifs.
▶ Associate motifs with gene ontology terms via their putative target genes.

Click on the following link to download a commercial-use license for the MEME Suite. Simply complete and mail the completed/signed license and you are good to go!

TECHNOLOGY DESCRIPTION

The MEME Suite is an integrated collection of tools for discovering and characterizing sequence motifs in collections of DNA 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 16 years and is now probably the most widely used motif discovery tool, with a total of 1676 citations in Google Scholar and 800 users per month submitting over 200 queries per day to the MEME web service. 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, 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.

APPLICATIONS

MEME 4.x is the latest version of the MEME Suite and contains many improvements and new features. The output of the MEME program has been redesigned to make it easier to interpret the results. For example, motifs discovered by MEME are now displayed as "sequence LOGOS," making it much easier to recognize and understand the motifs. With a single click, MEME motifs can be sent for analysis by other programs in the MEME Suite, such as Tomtom and MAST. The output of the MAST motif-scanning algorithm has also been greatly enhanced in readability.

There are two new motif-scanning tools—FIMO and AMA—that add new ways to scan sequence databases to predict where motifs occur. FIMO finds
individual motif occurrences. AMA estimates the DNA binding affinity of DNA-binding motifs (such as transcription factor motifs). In addition, the MCAST algorithm extends motif scanning to the prediction of clusters of DNA binding sites, rounding out the motif scanning features of the MEME Suite.

The gapped-motif discovery and scanning programs GLAM2 and GLAM2SCAN have been added to the MEME Suite to complement MEME, MAST, and FIMO, which are designed for non-gapped motifs. The new MEME Suite also contains the GOMO algorithm, which allows the biological function of DNA motifs to be predicted. MEME motifs can be sent for analysis by GOMO with a single click on the MEME output form.

SpaMo
SpaMo does inference of transcription factor complexes by looking for significant spacings between binding sites. The inputs are a set of many short sequences, a primary motif, and one or more databases of secondary motifs. It searches for the strongest primary motif binding site and then searches in the area around it looking for the strongest secondary motif binding site. The relative spacings of the primary and secondary motif in all the sequences are tallied and the probability of the close spacings happening by chance is calculated.

After all the calculations are done, SpaMo outputs the non-redundant secondary motifs in order of significance, provided they had a bin that passed the significance threshold. Similar secondary motifs are grouped together and listed in order of significance on the secondary motif they were redundant to. If the bin size is one, then an alignment for each of the similar secondary motifs is created.

MEME-ChIP
MEME-ChIP performs several motif analysis steps on a set of DNA sequences that you provide. It is especially appropriate for analyzing the bound genomic regions identified in a transcription factor (TF) ChIP-seq experiment.

MEME-ChIP can:
- Discover novel DNA-binding motifs.
- Analyze them for similarity to known binding motifs.
- Visualize the arrangement of the predicted motif sites in your input sequences.
- Detect very subtly enriched known motifs in your sequences.
- Provide an estimate of the amount of binding of each novel motif to each of your sequences.

MCAST
Search a sequence database for clusters of known motifs. MCAST employs a motif-based hidden Markov model, using a star topology and a novel scoring algorithm. The motifs may appear in any order.

RELATED MATERIALS

▶ For further information on the purpose and workings of MEME ChIP: The MEME Suite Motif-based sequence analysis tools.