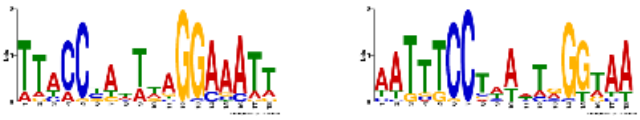
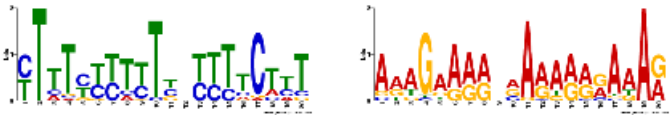
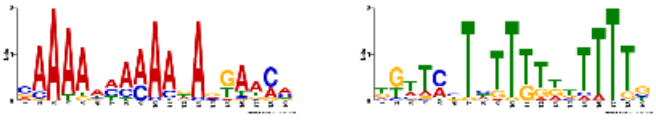
MEME
Multiple Em for Motif Elicitation

For further information on how to interpret these results or to get a copy of the MEME software please access http://meme.nbcr.net.

DISCOVERED MOTIFS | BLOCK DIAGRAMS OF MOTIFS | PROGRAM INFORMATION | EXPLANATION

DISCOVERED MOTIFS

Motif Overview

Motif 1	• 5.3e-017 • 23 sites	
Motif 2	• 1.5e-011 • 23 sites	
Motif 3	• 1.7e+006 • 23 sites	

Further Analysis

Submit all motifs to MAST FIMO GOMO BLOCKS Mouse-over buttons for more information.

MOTIF 1

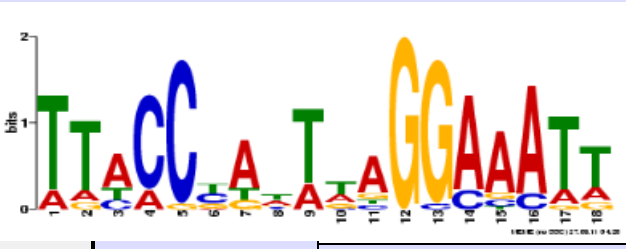
Next Top

Summary

E-value	5.3e-017
Width	18
Sites	23

show more

Sequence Logo



Standard | Reverse Complement

Download LOGO Orientation: standard SSC: off Format: web (png) Width: 18 cm Height: 7.5 cm Download

Regular expression

TT[AT]CC[CT]A[TA][TA][TA]AGGAAAT[TA]

Further Analysis

Submit this motif to MAST FIMO TOMTOM GOMO BLOCKS Mouse-over buttons for more information.

Data Formats

View the motif in PSPM Format PSSM Format BLOCKS Format FASTA Format Raw Format or Hide

Sites

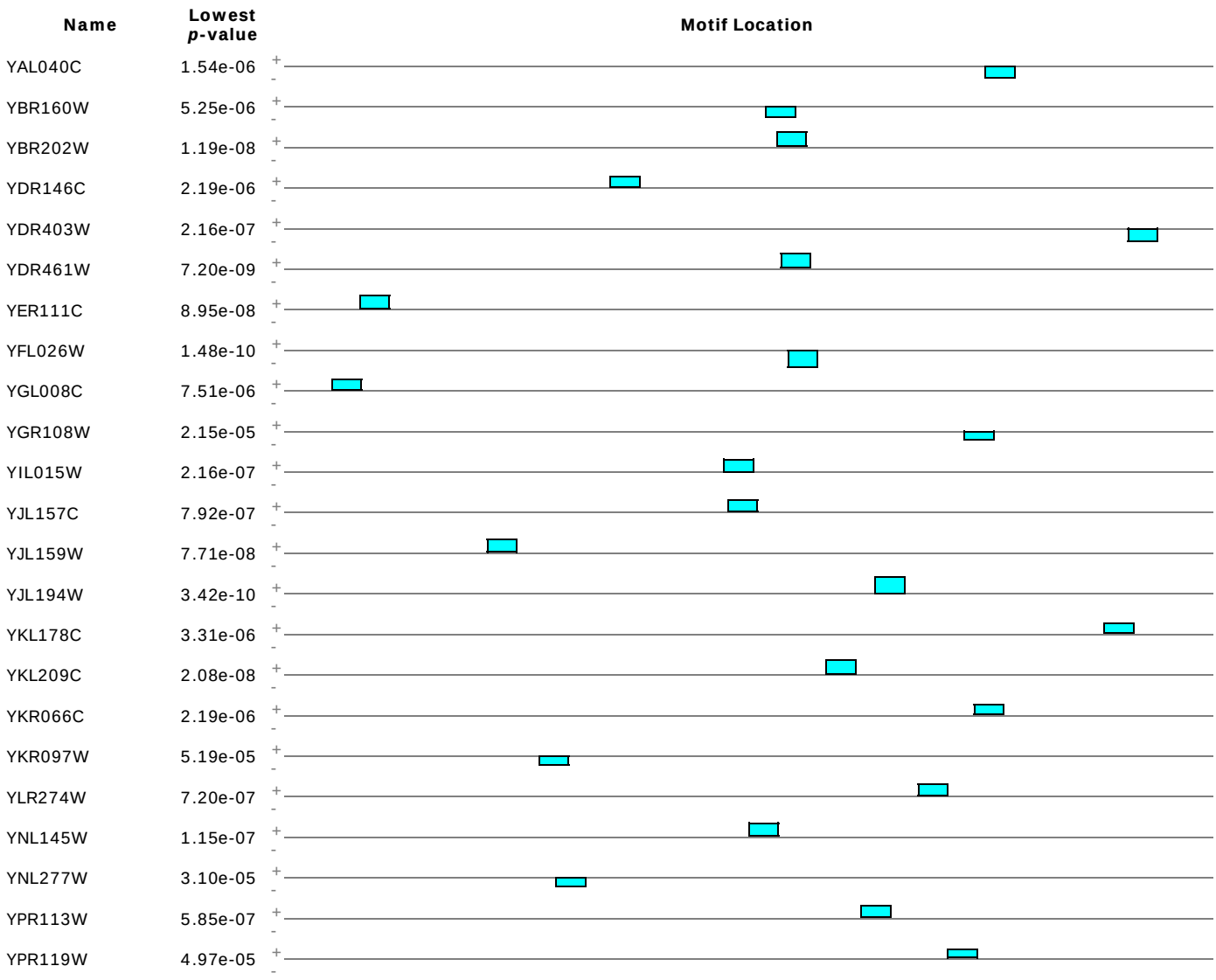
Click on any row to highlight sequence in all motifs.

Name	Strand	Start	*p*-value	Sites
YFL026W	-	276	1.48e-10	CATCATGTAC TTACCCAATTAGGAAATT TACATGGTTG

YJL194W	+	322	3.42e-10	GATCAGGAAA TTACCCACTTAGGAAATT GAGTTGGCTG
YDR461W	+	272	7.20e-09	GAGTGTGTAA TTACCCAAAAAGGAAATT TACATGTTAA
YBR202W	+	270	1.19e-08	CTTAACTAAT TTACCCAGAAAGGAAATT TCCTTATAAG
YKL209C	+	296	2.08e-08	TGCCATGTAA TTACCTAATAGGGAAATT TACACGCTGC
YJL159W	+	116	7.71e-08	GATCTTACTA TTTCCTATTTCGGAAATT ATTAAAGACA
YER111C	+	48	8.95e-08	CGCCTGCTTA TTTCCCGTTTAGGAAATA AATGGTTAAC
YNL145W	+	255	1.15e-07	AAGCATGTAT TTACCTATTCGGGAAATT TACATGACAT
YIL015W	+	242	2.16e-07	TGGCATGTAA TTACCGTAAAAGGAAATT ACATGGCGAG
YDR403W	-	457	2.16e-07	CGACAAAGCT TTACCGAATTAGGATATT AACAAATTGT
YPR113W	+	315	5.85e-07	CCGCACGTAA ATCCCTATTACGGAAATT ATTTCTCAAT
YLR274W	+	345	7.20e-07	GTGTGGCGTG TTTCCCTTTTAGGAAAAG TGATTACTTG
YJL157C	+	244	7.92e-07	TTCACCGAAA ATACCTAAAAAGGAAATA GGCAATAAAA
YAL040C	-	381	1.54e-06	CAGAGGAGTT AAACCGAATGAGGAAATG CAATGGATAC
YKR066C	+	375	2.19e-06	AAGCCGGCTG TTACCAATTTAGGAAAGG ACCCTGGAGC
YDR146C	+	181	2.19e-06	TTTCGTACTT TAACCTGTTTAGGAAAAA GGTAAACAAT
YKL178C	+	444	3.31e-06	TTATTGCTTT TTAACTTTAGAGGCAATT AAATTTGTGT
YBR160W	-	264	5.25e-06	TGGCGTTTGT TGCCCTAATAAGGCCATA TATGTATATA
YGL008C	+	33	7.51e-06	TTTTTTTTTT TTTACCTCTGTGGAAATT GTTACTCTCA
YGR108W	-	370	2.15e-05	TCTAAACAGC TGCACCAGTTAGCAACTT TAATAGTTTT
YNL277W	-	152	3.10e-05	GTGTGGCGCC TTCCGCAGTGAGGCGCAT CACGTGAAAA
YPR119W	+	361	4.97e-05	TTAGATTCTT TTAACGACATTGGAGAAA TTTTAATCAT
YKR097W	-	143	5.19e-05	TAAGCTTTAC AATCCTGTTTGGGCCCTT TATCCGCCTA

Block Diagrams

The height of the motif "block" is proportional to -log(p-value), truncated at the height for a motif with a p-value of 1e-10.
Click on any row to highlight sequence in all motifs. Mouse over the center of the motif blocks to see more information.



Time 13.8 secs.

MOTIF 2

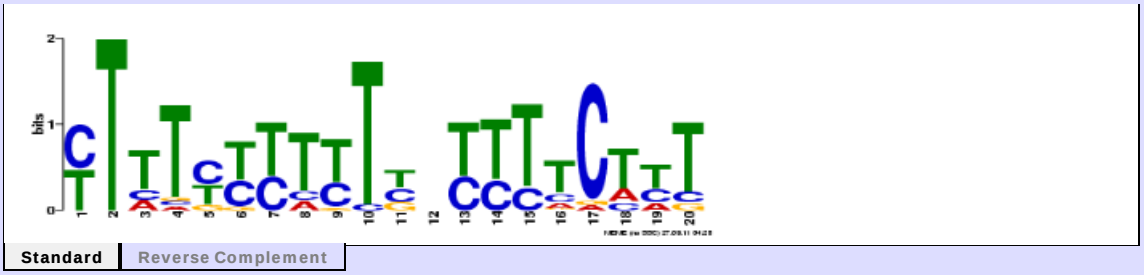
Previous Next Top

Summary

E-value	1.5e-011
Width	20
Sites	23

show more

Sequence Logo



Standard | Reverse Complement

Download LOGO Orientation: standard SSC: off Format: web (png) Width: 20 cm Height: 7.5 cm Download

Regular expression

[CT]TTT[CT][TC][TC]T[TC]T[TCG][TCG][TC][TC][TC]TC[TA][TC]T

Further Analysis

Submit this motif to MAST FIMO TOMTOM GOMO BLOCKS Mouse-over buttons for more information.

Data Formats

View the motif in PSPM Format PSSM Format BLOCKS Format FASTA Format Raw Format or Hide

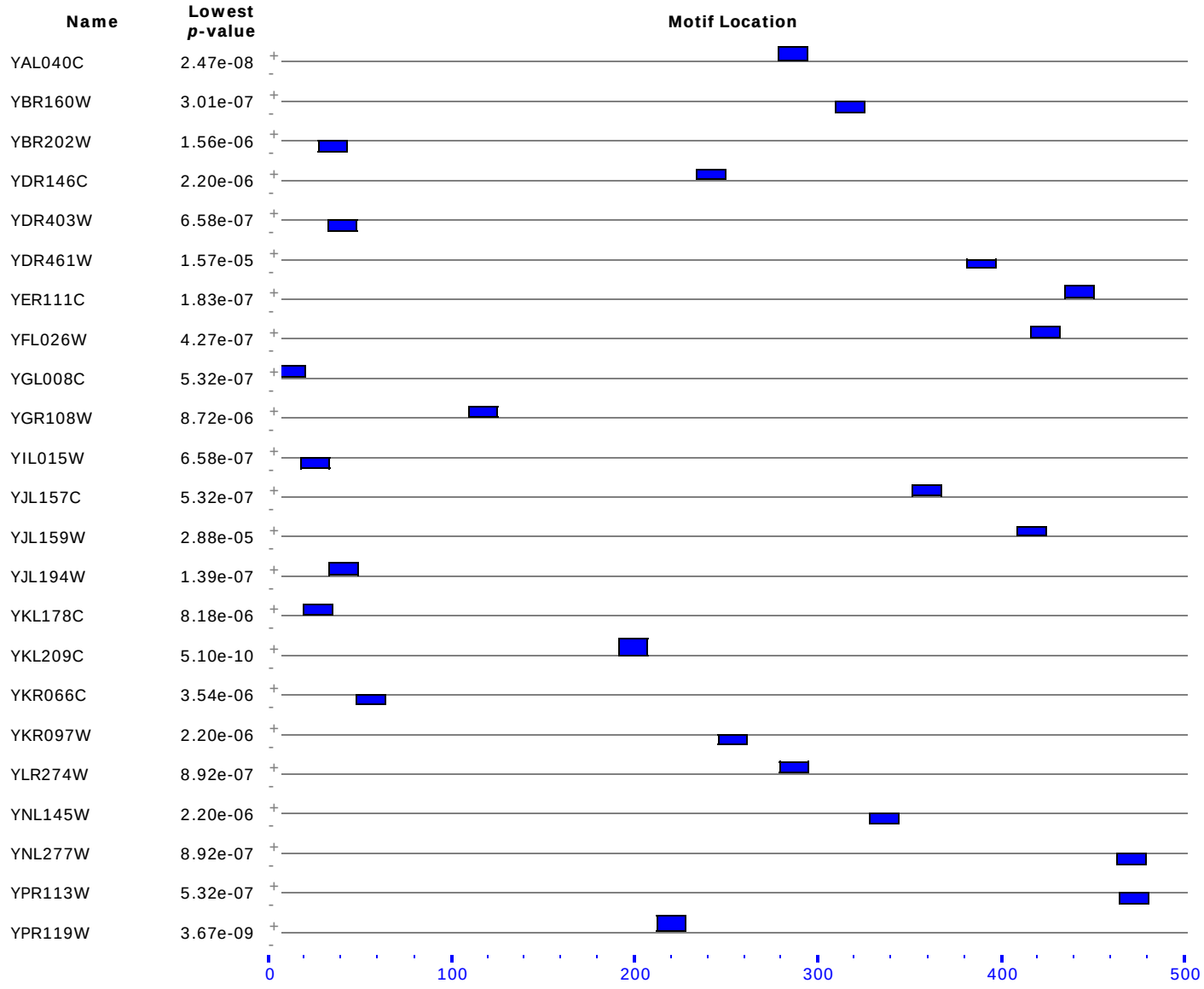
Sites

Click on any row to highlight sequence in all motifs.

Name	Strand	Start	*p*-value	Sites
YKL209C	+	192	5.10e-10	AAATATGTTC CTTTCCTTTTTCTTTTCTCT ATTCTGCCAC
YPR119W	+	213	3.67e-09	GTTATTTATT CTTTCTTTTTTTCTTTCTTT TTGATTGAGC
YAL040C	+	279	2.47e-08	TTCTATTTCT CTTTTCTTTTCTCCCTCTTT TTTCTCTGTA
YJL194W	+	34	1.39e-07	ATACTTCTAC CTCTTTTTCTGGCTTCCTTT CTTTCCCTTA
YER111C	+	435	1.83e-07	ATTAACTCTT CTTTCTTTCTCATTTCCCTT CTGTCCCTCT
YBR160W	-	310	3.01e-07	TTTTCCAAAT TTTTCTCTTTTTTTCTCATT TAGGTTCGTT
YFL026W	+	417	4.27e-07	AATACGATAC CTTTTCTTTTCACCTGCTCT GGCTATAATT
YPR113W	-	465	5.32e-07	TACTATCACA CTTTCCCTCTTATTTACATT ACTTGAAGGC
YJL157C	+	352	5.32e-07	GAAAATAAAA CTATCCTTTTTGTTCACTCT GTCTTGAGAG
YGL008C	+	5	5.32e-07	GCTTT TTTTGTTTCTGTTTTCCTTT TTTTTTTTTT
YIL015W	-	18	6.58e-07	TTTATGACGC CTTTCCTTTTTTCCTTCAAC GGCCTTCTTA
YDR403W	-	33	6.58e-07	ACCCCTTAAA TTTTCTCATTGCTCTTCATT GAGGTTGCAT
YNL277W	-	464	8.92e-07	TTATACTACG TTTTTTTCTTTCTTTTCTAT TATTCCAAAT
YLR274W	+	280	8.92e-07	GCCTTTAATT TTCTTCTCCTCCTCTTCACT CCCAGGTTAA
YBR202W	-	28	1.56e-06	TTAGAATTTA CTTACTCTTTTACCCTCTCT GTCTCTCGTA
YNL145W	-	329	2.20e-06	ATGCAACAGC TTTTTCCTCTCGCTTTGTTC TTTTTCCTCT
YKR097W	-	246	2.20e-06	CTTATATACC TTTTTTTTGTGGTCTTCTAT ACTCTGCCCA
YDR146C	+	234	2.20e-06	TAGCGACTAA TTATCTTTCTTTCTCCCTAT AGATCTGCTC
YKR066C	-	49	3.54e-06	TCGATCCAAG CTCGTCTTTTGGCTTTCTTG GCTATTCTTA
YKL178C	+	20	8.18e-06	ACCAGAAGCG TTCTGGCTCCCCTTTTCTCT AAACGTTAAA
YGR108W	+	110	8.72e-06	TTGTTTGTGA TTTCCTCATTCGTCTTCCTC TGCAGAAACG
YDR461W	-	382	1.57e-05	TATGAAGACT CTATTTCATTTTTTTACTTG ATTTTCAGCA
YJL159W	+	409	2.88e-05	TCAATTTCGA TTATGTCCTTCCTTTTACCT CAAAGCTCAA

Block Diagrams

The height of the motif "block" is proportional to -log(p-value), truncated at the height for a motif with a p-value of 1e-10.
Click on any row to highlight sequence in all motifs. Mouse over the center of the motif blocks to see more information.



Time 27.1 secs.

MOTIF 3

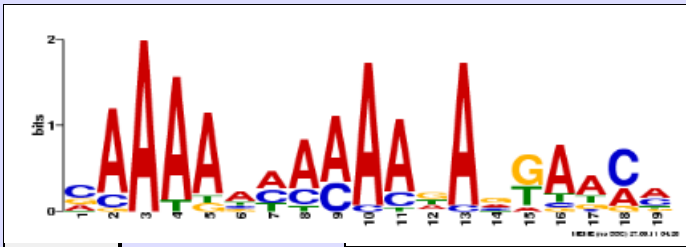
Previous Top

Summary

E-value	1.7e+006
Width	19
Sites	23

show more

Sequence Logo



Standard | Reverse Complement

Download LOGO Orientation: standard SSC: off Format: web (png) Width: 19 cm Height: 7.5 cm Download

Regular expression

[CGA]AAAAA[ACT][AC][AC]AA[GAT]A[GAC][GT]A[AT][CA][AC]

Further Analysis

Submit this motif to MAST FIMO TOMTOM GOMO BLOCKS Mouse-over buttons for more information.

Data Formats

View the motif in PSPM Format PSSM Format BLOCKS Format FASTA Format Raw Format or Hide

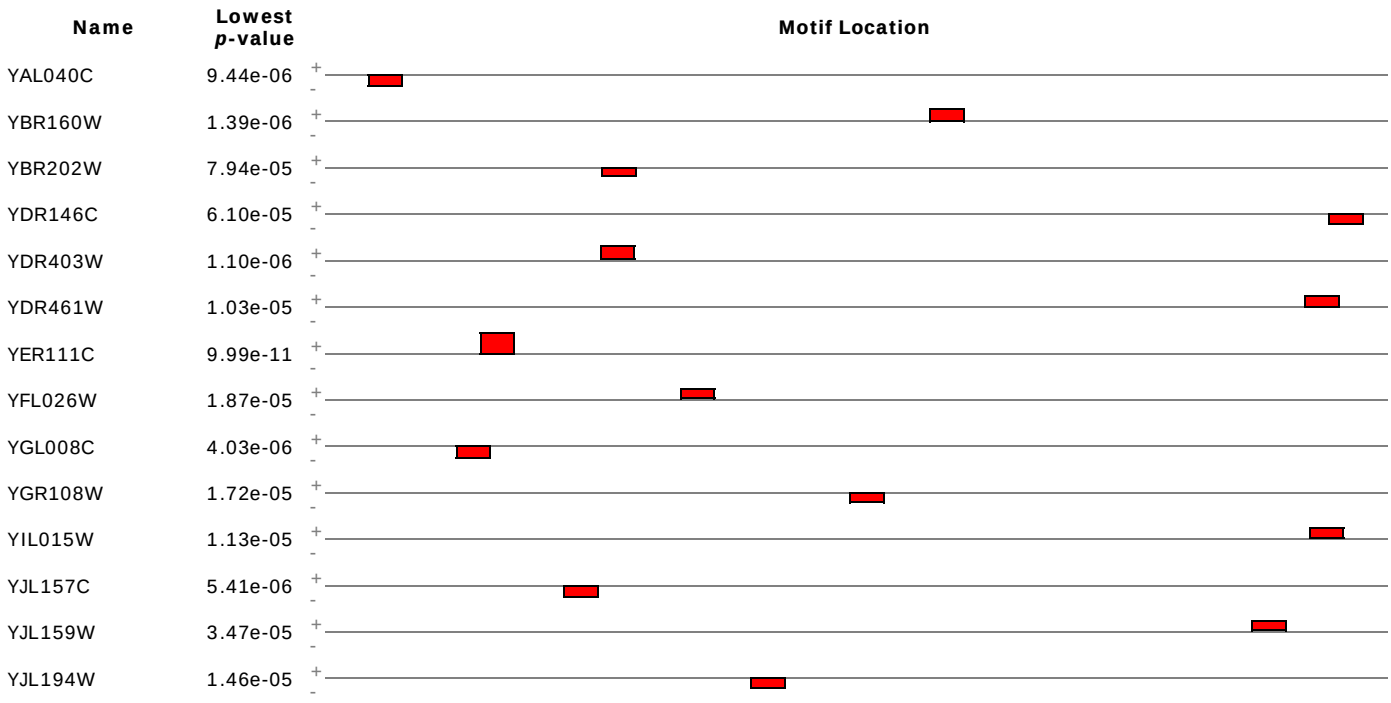
Sites

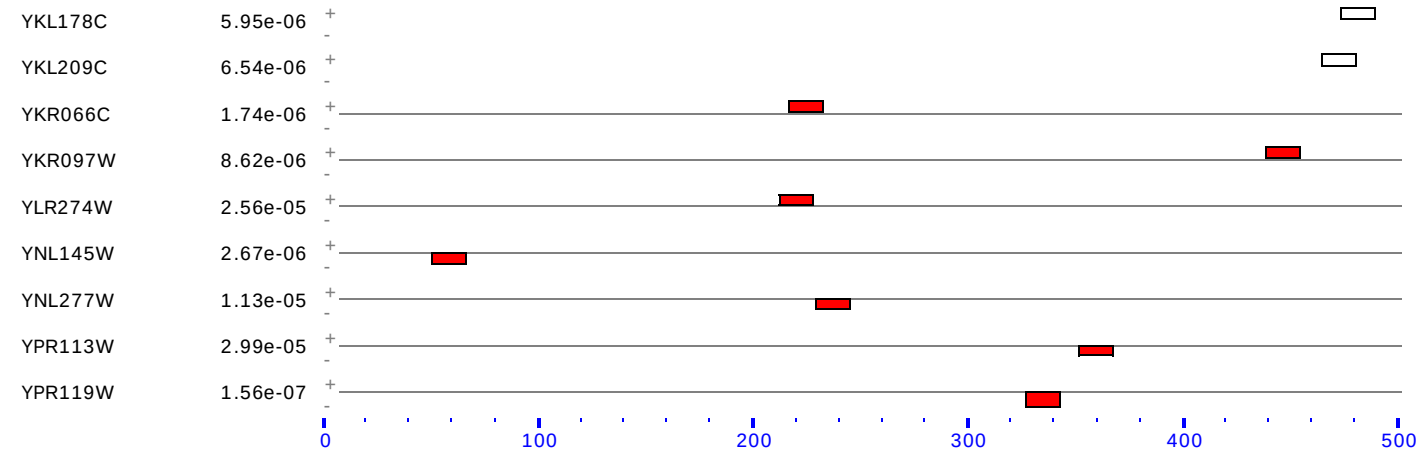
Click on any row to highlight sequence in all motifs.

Name	Strand	Start	*p*-value	Sites
YER111C	+	80	9.99e-11	GTTAACCTAA CAAAAAAAAAAGAGGAACA AGGACTAAGA
YPR119W	-	327	1.56e-07	TCTAAAAGGA CAAAAATACAAAAGGAAAA CAGATGCTCA
YDR403W	+	136	1.10e-06	AATTGATCGA AAAAAGCAAACCAGGAACC CCTCACAAAT
YBR160W	+	289	1.39e-06	GCAACAAACG CCAAAAATAAAAACGAACC TAAATGAGAA
YKR066C	+	217	1.74e-06	AATGGAAATT CCAAAGCCAAAGAATATCC TGCTTTGTCA
YNL145W	-	51	2.67e-06	TGATGAAACT GAAAAATAAAAAAATAAAC AGTCTGGTCT
YGL008C	-	69	4.03e-06	TTGGAATAAA CAAAACAAACAAACGAACT AAAGAGTGTG
YJL157C	-	119	5.41e-06	TGCCCGCCTA ACAAAACCAAAGAGGTAAA TGAAAGTAAA
YKL178C	+	474	5.95e-06	ATTTGTGTAG GAAAGGCAAAATACTATCA AAATTTTCA
YKL209C	+	465	6.54e-06	AAGTGCCGCT GAAAATTCCACTAGGAAAC AAAGAACAAG
YKR097W	+	439	8.62e-06	TTGTTTAACT CAAAAATAAAAAAAAAAAA CCAAACTCAC
YAL040C	-	28	9.44e-06	GTAGGAGGAA GAAAAAAAAAAGATGAGGG CAAAAACCCA
YDR461W	+	464	1.03e-05	GTTTCTCGGA TAAAACCAAAATAAGTACA AAGCCATCGA
YNL277W	-	230	1.13e-05	GGGTGTGTGC CAAATCCAAACGATTAACA TGACTTAGTT
YIL015W	+	466	1.13e-05	AAAATCATAC CAAAATAAAAAGAGTGTCT AGAAGGGTCA
YJL194W	-	206	1.46e-05	CAACTAGTTT CAAAACACAACAAATCGCA AGCCAGACAT
YGR108W	-	252	1.72e-05	GGCCGAATGT AAAAAACACAATACAACCA AAATTGCCTA
YFL026W	+	173	1.87e-05	TCCTGCTTTG CAATGAAACAATAGTATCC GCTAAGAATT
YLR274W	+	213	2.56e-05	AGTAAACTGT CAATATCAAAAACGTAACA ATTGATTCAA
YPR113W	-	352	2.99e-05	TCCTCGAGAC GGAAAAAACAATAGGAGGG GCTTTCCCTA
YJL159W	+	439	3.47e-05	CAAAGCTCAA AAAAATATCAATAAGAAAC TCATATTCCT
YDR146C	-	475	6.10e-05	TCTTTTTC GAAAAAACAATGACGTTAA TATTTTATTT
YBR202W	-	137	7.94e-05	TGTTATTGAG AAAATATACAAGAAGCGCT TATTATATTC

Block Diagrams

The height of the motif "block" is proportional to -log(p-value), truncated at the height for a motif with a p-value of 1e-10.
Click on any row to highlight sequence in all motifs. Mouse over the center of the motif blocks to see more information.





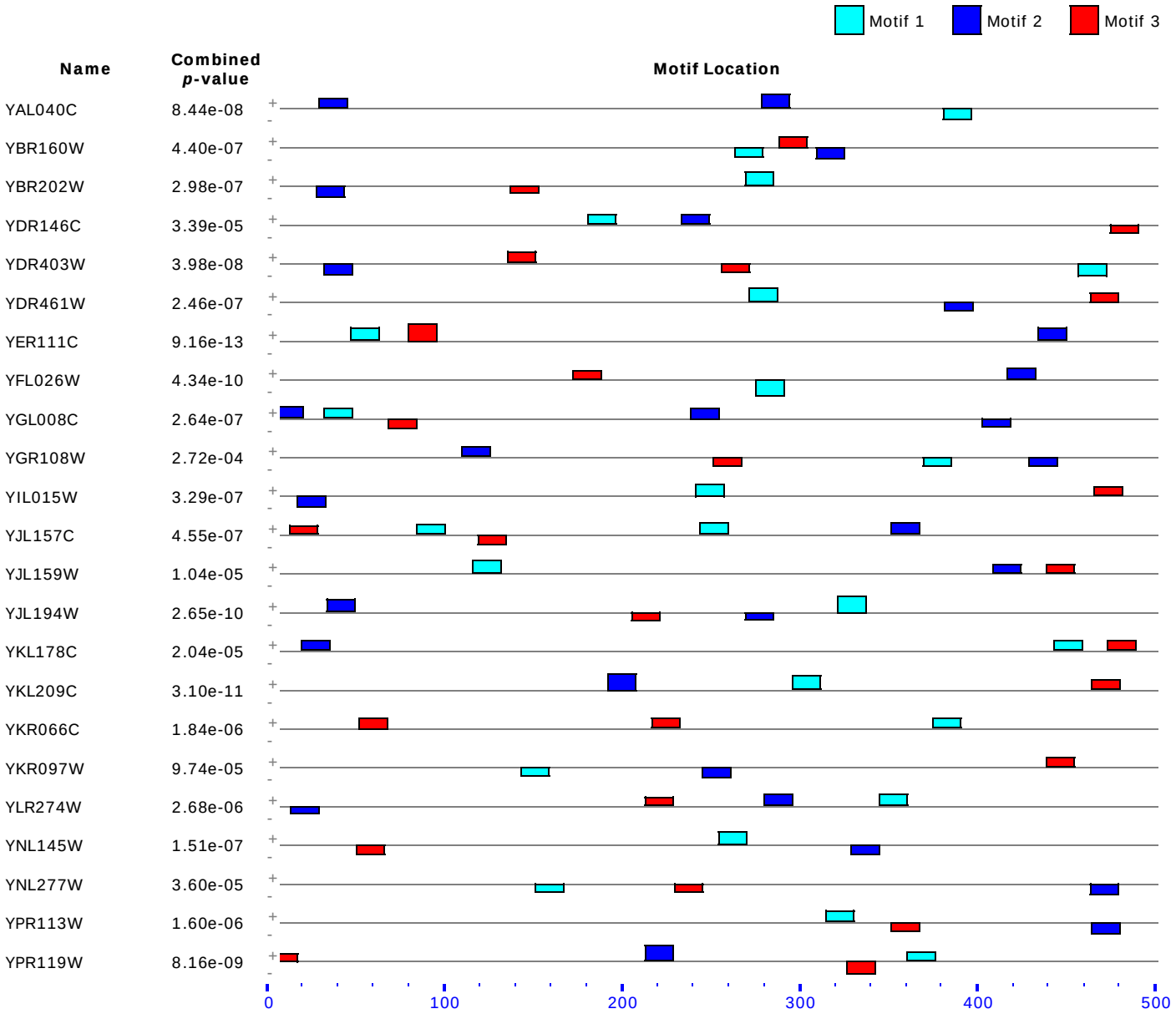
Time 40 secs.

ALL MOTIFS

Top

Combined Block Diagrams

Non-overlapping sites with a *p*-value better than 0.0001.
The height of the motif "block" is proportional to -log(p-value), truncated at the height for a motif with a p-value of 1e-10.
Click on any row to highlight sequence in all motifs. The motif blocks have tool tips with more information.



Top

MEME version
4.6.1 (Release date: Mon Mar 21 13:41:45 EST 2011)

Reference
Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

show training set...

Command line summary

```
meme sequences -sf promoters.txt -dna -mod oops -nmotifs 3 -minw 10 -maxw 20 -time 7200 -maxsize 60000 -revcomp -oc . -
```

Letter frequencies in dataset
A: 0.325 C: 0.175 G: 0.175 T: 0.325

Background letter frequencies (from dataset with add-one prior applied):
A: 0.325 C: 0.175 G: 0.175 T: 0.325

Stopping Reason
Stopped because nmotifs = 3 reached. Program ran on *compute-0-8.local*.

show model parameters...

EXPLANATION OF MEME RESULTS

Top

The MEME results consist of

- The **overview** of all discovered motifs.
- Information on each of the **motifs** MEME discovered, including:
  1. A **summary table** showing the width, number of contributing sites, log likelihood ratio, statistical significance, information content and relative entropy of the motif.
  2. A **sequence LOGO**.
  3. Downloadable **LOGO files** suitable for publication.
  4. A **regular expression** describing the motif.
  5. Some **further analysis** that can be performed on the motif.
  6. A list of **data formats** describing the motif.
  7. The **contributing sites** of the motif sorted by *p*-value and aligned with each other.
  8. The **block diagrams** of the contributing sites of the motif within each sequence in the training set.
- A **combined block diagram** showing an optimized (non-overlapping) tiling of all of the motifs onto each of the sequences in the training set.
- The **version** of MEME and the date it was released.
- The **reference** to cite if you use MEME in your research.
- A description of the **sequences** you submitted (the "training set") showing the name, "weight" and length of each sequence.
- The **command line summary** detailing the parameters with which you ran MEME.
- The reason why MEME **stopped** and the name of the CPU on which it ran.
- This **explanation** of how to interpret MEME results.

Motifs

For each motif that it discovers in the training set, MEME prints the following information:

Summary Table

This summary table gives the main attributes of the motif.

E-value
: The statistical significance of the motif. MEME usually finds the most statistically significant (low *E*-value) motifs first. The *E*-value of a motif is based on its log likelihood ratio, width, sites, the background letter frequencies (given in the command line summary), and the size of the training set. The *E*-value is an estimate of the expected number of motifs with the given log likelihood ratio (or higher), and with the same width and site count, that one would find in a similarly sized set of random sequences. (In random sequences each position is independent with letters chosen according to the background letter frequencies.)

Width
: The width of the motif. Each motif describes a pattern of a fixed with as no gaps are allowed in MEME motifs.

Sites
: The number of sites contributing to the construction of the motif.

Log Likelihood Ratio
: The log likelihood ratio of the motif.The log likelihood ratio is the logarithm of the ratio of the probability of the occurrences of the motif given the motif model (likelihood given the motif) versus their probability given the background model (likelihood given the null model). (Normally the background model is a 0-order Markov model using the background letter frequencies, but higher order Markov models may be specified via the **-bfile** option to MEME.)

Information Content
: The information content of the motif in bits. It is equal to the sum of the **uncorrected** information content, `R()`, in the columns of the LOGO. This is equal relative entropy of the motif relative to a uniform background frequency model.

Relative Entropy
: The relative entropy of the motif, computed in bits and relative to the background letter frequencies given in the command line summary. It is equal to the log-likelihood ratio (llr) divided by the number of contributing sites of the motif times `1/ln(2)`,

re = llr / (sites * ln(2)).

Sequence LOGO

MEME motifs are represented by position-specific probability matrices that specify the probability of each possible letter appearing at each possible position in an occurrence of the motif. These are displayed as "sequence LOGOS", containing stacks of letters at each position in the motif. The total height of the stack is the "information content" of that position in the motif in bits. The height of the individual letters in a stack is the probability of the letter at that position multiplied by the total information content of the stack.

Note: The MEME LOGO differs from those produced by the Weblogo program because a **small-sample correction is NOT applied**. However, MEME LOGOs in PNG and encapsulated postscript (EPS) formats **with small-sample correction (SSC)** are available by clicking on the download button with "SSC" set to "on" under Download LOGO. The MEME LOGOs without small sample correction are similarly available. Error bars are included in the LOGOs with small-sample correction.

Modern web browsers supporting the canvas element and it's text manipulation functions as described in the html 5 standard, can render the sequence LOGOs without needing the images. The browsers which work with this feature are:

- Firefox 3.5 and above
- Safari 4 and above
- Google Chrome 4 and above

Unfortunately Internet Explorer 8 does not support any html 5 features.

The information content of each motif position is computed as described in the paper by Schneider and Stephens, "Sequence Logos: A New Way to Display Consensus Sequences" but **the small-sample correction, e(n), is set to zero for the LOGO displayed in the MEME output.** The corrected information content of position i is given by

```
R(i) for amino acids   = log2(20) - (H(i) + e(n))   (1a)
R(i) for nucleic acids =    2    - (H(i) + e(n))   (1b)
```

where H(i) is the entropy of position i,

```
H(l) = - (Sum f(a,i) * log2[ f(a,i) ]).            (2)
```

Here, f(a,i) is the frequency of base or amino acid a at position i, and e(n) is the small-sample correction for an alignment of n letters. The height of letter a in column i is given by

```
height = f(a,i) * R(i)                             (3)
```

The approximation for the small-sample correction, e(n), is given by:

```
e(n) = (s-1) / (2 * ln(2) * n),                    (4)
```

where s is 4 for nucleotides, 20 for amino acids, and n is the number of sequences in the alignment.

The letters in the logos are colored as follows.
For DNA sequences, the letter categories contain one letter each.

NUCLEIC ACIDS	COLOR
A	RED
C	BLUE
G	ORANGE
T	GREEN

For proteins, the categories are based on the biochemical properties of the various amino acids.

AMINO ACIDS	COLOR	PROPERTIES
A, C, F, I, L, V, W and M	BLUE	Most hydrophobic[Kyte and Doolittle, 1982]
NQST	GREEN	Polar, non-charged, non-aliphatic residues
DE	MAGENTA	Acidic
KR	RED	Positively charged
H	PINK	
G	ORANGE	
P	YELLOW	
Y	TURQUOISE	

J. Kyte and R. Doolittle, 1982. "A Simple Method for Displaying the Hydropathic Character of a Protein", J. Mol Biol. 157, 105-132.

Note: the "text" output format of MEME preserves the historical MEME format where LOGOS are replaced by a simplified probability matrix, a relative entropy plot, and a multi-level consensus sequence.

Download LOGO

Logos can be generated on the fly by the meme webservice and you may specify a number of options to customize them to your needs. The options are:

Orientation
: Only valid for nucleotide motifs. Generate the standard view or the reverse complemented view of the motif.

SSC
: Use small sample correction and show errorbars on the image. Small sample correction is used by the Weblogo program.

Format

The format of the generated image. If the image is to be used on the web then png is recommend. If the image is to be published then eps is recommended.

Width
: The width of the generated image in centimetres.

Height
: The height of the generated image in centimetres.

Regular Expression

This is a regular expression (RE) describing the motif. In each column, all letters with observed frequencies greater than 0.2 are shown; less-frequent letters are not included in the RE. MEME regular expressions are interpreted as follows: single letters match that letter; groups of letters in square brackets match any of the letters in the group. Regular expressions can be used for searching for the motif in sequences (using, for example, PatMatch) but the search accuracy will usually be better with the PSSM (using, for example MAST.)

Further Analysis

Either as a group or individually the motifs have a number of options for further analysis.

MAST
: Finds the best matching positions for a set of motifs in each sequence provided to it, ranked by the combined score of each sequence. For more information about MAST please read the introduction.

FIMO
: Finds all matches for a motif. For more information about FIMO please read the introduction.

TOMTOM
: Compares a single motif to a database of motifs. For more information about TOMTOM please read the introduction.

GOMO
: Identifies possible roles of DNA binding motifs. For more information about GOMO please read the introduction.

BLOCKS
: Submit to Blocks Multiple Alignment Processor where you can do several things like create phylogeny trees and search the blocks against a database of other blocks (protein only). For more information about BLOCKS Multiple Alignment Processor please visit the website.

Data Formats

The extracted data is avaliable in the following formats.

PSPM Format
: The motif itself is a position-specific probability matrix giving, for each position in the pattern, the observed frequency ("probability") of each possible letter. The probability matrix is printed "sideways"--columns correspond to the letters in the alphabet (in the same order as shown in the simplified motif) and rows corresponding to the positions of the motif, position one first. The motif is preceded by a line starting with "letter-probability matrix:" and containing the length of the alphabet, width of the motif, number of occurrences of the motif, and the *E*-value of the motif.
Note: Earlier versions of MEME gave the posterior probabilities--the probability after applying a prior on letter frequencies--rather than the observed frequencies. These versions of MEME also gave the number of *possible* positions for the motif rather than the actual number of occurrences. The output from these earlier versions of MEME can be distinguished by "n=" rather than "nsites=" in the line preceding the matrix.

PSSM Format
: The position-specific scoring matrix corresponding to the motif is printed for use by database search programs such as MAST. This matrix is a log-odds matrix calculated by taking 100 times the log (base 2) of the ratio p/f at each position in the motif where p is the probability of a particular letter at that position in the motif, and f is the background frequency of the letter (given in the command line summary section.) This is the same matrix that is used above in computing the p-values of the occurrences of the motif in the Sites and Block Diagrams sections. The scoring matrix is printed "sideways"--columns correspond to the letters in the alphabet (in the same order as shown in the simplified motif) and rows corresponding to the positions of the motif, position one first. The scoring matrix is preceded by a line starting with "log-odds matrix:" and containing the length of the alphabet, width of the motif, number of characters in the training set, the scoring threshold (obsolete) and the motif *E*-value.
Note: The probability p used to compute the PSSM is *not* exactly the same as the corresponding value in the Position Specific Probability Matrix (PSPM). The values of p used to compute the PSSM take into account the motif prior, whereas the values in the PSPM are just the *observed* frequencies of letters in the motif sites.

BLOCKS Format
: For use with BLOCKS tools.

FASTA Format
: The FASTA format as described here.

Raw Format
: Just the sites of the sequences that contributed to the motif. One site per line.

Sites

MEME displays the occurrences (sites) of the motif in the training set. The sites are shown aligned with each other, and the ten sequence positions preceding and following each site are also shown. Each site is identified by the name of the sequence where it occurs, the strand (if both strands of DNA sequences are being used), and the position in the sequence where the site begins. When the DNA strand is specified, '+' means the sequence in the training set, and '-' means the reverse complement of the training set sequence. (For '-' strands, the 'start' position is actually the position on the **positive** strand where the site ends.) The sites are **listed in order of increasing statistical significance** (p-value). The p-value of a site is computed from the the match score of the site with the position specific scoring matrix for the motif. The p-value gives the probability of a random string (generated from the background letter frequencies) having the same match score or higher. (This is referred to as the **position *p*-value** by the MAST algorithm.)

Block Diagrams

The occurrences of the motif in the training set sequences are shown as coloured blocks on a line. One diagram is printed for each sequence showing all the sites contributating to that motif in that sequence. The sequences are **listed in the same order as in the input** to make it easier to compare multiple block diagrams. Additionally the best p-value for the sequence/motif combination is listed though this may not be in ascending order as with the sites. The p-value of an occurrence is the probability of a single random subsequence the length of the motif, generated according to the 0-order background model, having a score at least as high as the score of the occurrence. When the DNA strand is specified '+', it means the motif appears from left to right on the sequence, and '-' means the motif appears from right to left on the complementary strand. A sequence position scale is shown at the end of each table of block diagrams.

Combined Block Diagrams

The motif occurrences shown in the motif summary **may not be exactly the same as those reported in each motif section** because only motifs with a position *p*-value of 0.0001 that don't overlap other, more significant motif occurrences are shown.

See the documentation for MAST output for the definition of position and combined *p*-values.