

Additional Results and Tables

Table A1 - Percentage of correctly identified sites for different methods, varying number of starting points and data set length L

The rows display the results for each method: 1) MEME software with the options `-nmotifs 1 -dna -mod oops -brief -noshorten -b 0 -adj none` or 2) the variable change point algorithm with normal prior, $\lambda = 0$ and motif width and motif specification (uni- or bi-modal) listed in Table 1. The first column shows the number of starting points used in the algorithm. Starting points were selected according to the MEME procedure as described in the text. The next columns correspond to the different values of L . Entries labeled with an asterisk correspond to a trial in which the correct motif was not found, but one or two sites were correctly predicted by chance, with a spurious motif. Note for *rap1* and *reb1*, we do not start at $L = 100$. Both data sets have a sequence that contains two sites separated by more than 100 bases. For these two cases, we started at the shortest 100 base increment such that the two sites were contained: $L = 200$ for *rap1* and $L = 300$ for *reb1*.

rap1

Method	# of SP	L								
		200	300	400	500	600	700	800	900	1000
MEME	1	87	87	87	87	73	73	0	0	0
$\lambda = 0$	1	93	87	87	80	73	80	0	0	0
$\lambda = 0$	100	93	87	87	87	73	87	80	80	80

Method	# of SP	1100	1200	1300	1400	1500	1600	1700	1800
MEME	1	0	0	0	0	0	0	0	0
$\lambda = 0$	1	0	0	0	0	0	0	0	0
$\lambda = 0$	100	80	0	80	80	0	0	0	0

abf1

Method	# of SP	L									
		100	200	300	400	500	600	700	800	900	1000
MEME	1	83	83	83	61	6*	50	6*	0	11*	0
$\lambda = 0$	1	83	83	61	67	0	56	0	0	0	0
$\lambda = 0$	100	83	83	83	67	0	56	0	0	0	0

reb1

Method	# of SP	L						
		300	400	500	600	700	800	900
MEME	1	86	79	0	0	0	0	0
$\lambda = 0$	1	64	71	0	0	0	0	0
$\lambda = 0$	100	86	71	86	0	0	0	0

Results of Gibbs Motif Sampler on Real Data

Gibbs Motif Sampler (GMS) is also popular software for finding motifs and was the basis for BioProspector. Table A2 lists the results of our test sets using GMS. For the uni-modal motifs, the real motif was never discovered for *rap1*. For *reb1*, several different combinations were adequate for its discovery up to length 700. For the bi-modal examples, the closest option in GMS that matches our model specification is the fragmentation model, with fragmentation from center. For *crp*, the values are $W = 16$ and $J = 10$, while for *abf1* the values are $W = 12$ and $J = 7$. We also tried different values for W and J . In both cases, fragmentation from center was not successful (Table A2). The original fragmentation model was able to find the motif for *crp* with $W = 20$ and 24, but only up to $L = 500$. The results were similar when we fixed $W = 20$ or 24 and used different values for J . For *abf1*, neither the fragmentation model nor the “fragment from center” version found the motif. Because GMS with the fragmentation model was not successful for *abf1*, we also tried the original model without fragmentation using several different values for W . The correct motif (10 predicted sites out of 18) was discovered only for $W = 12$ at $L = 600$.

Table A2 - Percentage of correctly identified sites using different options in Gibbs Motif Sampler

The basic model requires the motif width W . The fragmentation model, with or without the option “fragment from center”, requires both W and J . Entries labeled with an asterisk correspond to a trial in which the correct motif was not found, but one or two sites were correctly predicted by chance, with a spurious motif.

crp

Options	L			
	400	500	600	700
Fragmentation Model				
$W = 16, J = 10$	65	0	0	0
$W = 18, J = 10$	65	0	0	0
$W = 20, J = 10$	70	65	0	0
$W = 22, J = 10$	65	0	0	0
$W = 24, J = 10$	70	70	0	0
Fragment from Center				
$W = 16, J = 10$	0	0	0	0
$W = 18, J = 10$	0	0	0	0
$W = 20, J = 10$	0	0	0	0
$W = 22, J = 10$	0	0	0	0
$W = 24, J = 10$	0	0	0	0

abf1

Options	L			
	600	700	800	900
Fragmentation Model				
$W = 12, J = 7$	0	0	0	0
$W = 14, J = 7$	0	0	0	0
$W = 16, J = 7$	0	0	0	0
$W = 18, J = 7$	0	0	0	0
$W = 20, J = 7$	0	0	0	0
Fragment from Center				
$W = 12, J = 7$	0	0	0	0
$W = 14, J = 7$	0	0	0	0
$W = 16, J = 7$	0	0	0	0
$W = 18, J = 7$	0	0	0	0
$W = 20, J = 7$	0	0	0	0

rap1

Options	L				
	1400	1500	1600	1700	1800
Basic Model					
$W = 5$	0	0	7*	13*	7*
$W = 6$	0	0	7*	0	0
$W = 7$	0	0	7*	7*	7*
$W = 8$	0	0	7*	7*	7*
Fragmentation Model					
$W = 13, J = 5$	7*	0	7*	0	0
$W = 13, J = 6$	0	0	0	0	0
$W = 13, J = 7$	0	0	7*	7*	7*
$W = 13, J = 8$	0	0	0	0	7*

reb1

Options	L		
	500	600	700
Basic Model			
$W = 5$	86	0	0
$W = 6$	64	79	79
$W = 7$	79	79	79
$W = 8$	79	0	0
Fragmentation Model			
$W = 13, J = 5$	0	0	0
$W = 13, J = 6$	64	79	71
$W = 13, J = 7$	71	0	0
$W = 13, J = 8$	79	86	86