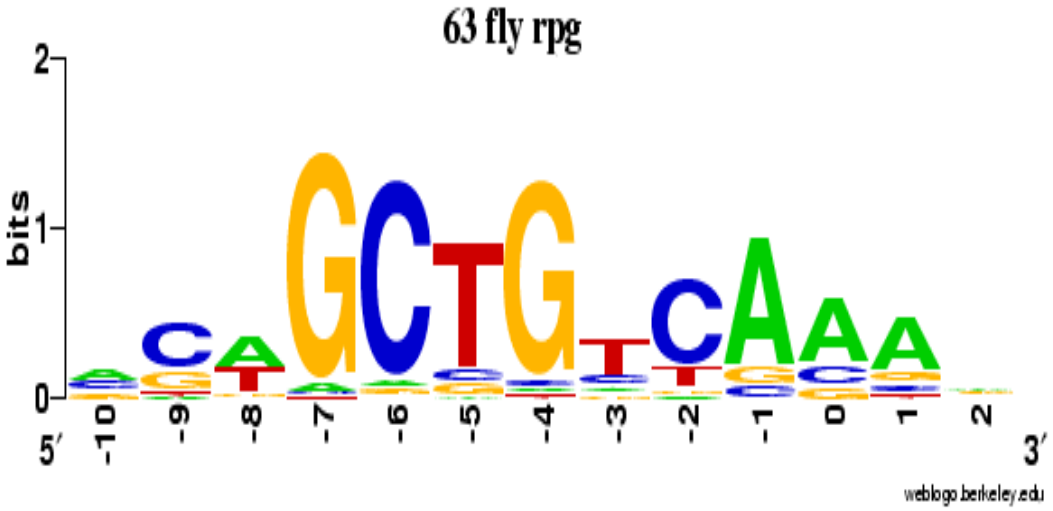
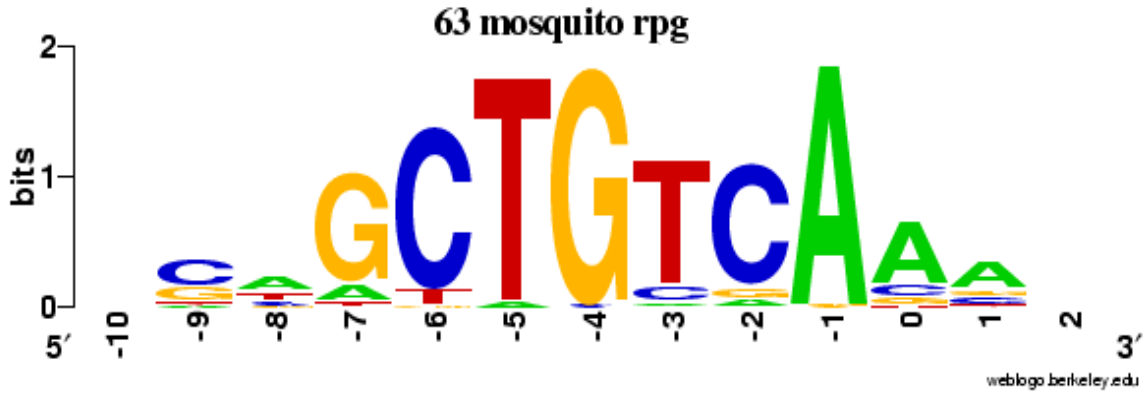
Appendix 3: Motif Instances in Insect Ribosomal Protein Genes

In this file, we list the motifs found by the Gibbs sampler on the insect ribosomal dataset. Here we show the two motif logos first.





In the following, the consensus for fly and mosquito are on the left and right, respectively, of the first row. Then the weight matrices for the two species are listed. Finally, the motif instances are given for each gene we used (the 800 bp upstream of the ribosomal protein gene sequences for the two insect species are retrieved from Ensembl at www.ensembl.org/). The number of matches in the two instances is provided at the end of the alignment.

ACAGCTGTCAAAA	ACAGCTGTCAAAA
0.377 0.300 0.269 0.054	0.300 0.238 0.285 0.177
0.054 0.577 0.269 0.100	0.069 0.531 0.300 0.100
0.485 0.054 0.100 0.362	0.469 0.115 0.100 0.315
0.054 0.023 0.900 0.023	0.131 0.023 0.792 0.054
0.054 0.869 0.054 0.023	0.008 0.869 0.023 0.100
0.038 0.100 0.085 0.777	0.038 0.008 0.008 0.946
0.038 0.054 0.869 0.038	0.008 0.023 0.962 0.008
0.131 0.177 0.100 0.592	0.038 0.115 0.023 0.823
0.054 0.700 0.069 0.177	0.069 0.823 0.085 0.023
0.777 0.085 0.115 0.023	0.962 0.008 0.023 0.008
0.638 0.192 0.146 0.023	0.700 0.146 0.100 0.054
0.638 0.131 0.146 0.085	0.577 0.162 0.177 0.085
0.377 0.131 0.300 0.192	0.285 0.254 0.208 0.254

RPS3A	ACAGGTCTCAAAT	
	CGCGCTGTCAAAA	7
RPL29	CGAGCTGTGGAAT	
	AGAACTGTCAATA	7
RPL18A	CCAGCTATCGAAC	
	CGAGCTGTCAAAG	9
RPL31	CCAGGAGGCACGC	
	CCAGCTGTCAAGC	9
RPS16	GCAGCTGTCAGTG	
	GCTGCTGTCAAAA	9
RPL23	AGAGATGTGAAAA	
	AGAGCTCTGAAAC	10
RPS24	GCAGCTGGCAGGA	
	GCAACTGACAGCG	9
RPL11	AGTGCTGTCCAAG	
	ACTGCTGTTGCAA	8

RPLP1	CCAGCTGTTAACT	
	GCTGCAGTAAAAA	7
RPS15	AATGCCGCCAAAA	
	ACTGCTGCAAGCT	7
RPL12	CTAGCTGAAGAAA	
	CTAGCTGTCAAAG	9
RPS18	CGTGCTGTTAATA	
	AGACCTGTCAAAG	7
RPL19	CCGGCTGTCAAAT	
	TCAGCTGTCAAGG	9
RPL37	CCTGCGGACAAAA	
	TCGGCTGTCAAAA	9
RPS23	AGTGCTGACAAAT	
	CAAGCTGTCAAAG	8
RPL23A	CCTCCTGTCAAAA	
	CATGCTGGCAAAG	9
RPL26	TGAGCTATCAAAA	
	GGCGCTGTCACGT	7
RPS12	GCTGCTGTCGTCG	
	GCTGCTGCGAAGA	7
RPS4	ACAGCTGACAAAA	
	GAAGTTGTCAAAC	8
RPL10A	AGAGCAGCCAGAA	
	ACAACTGTCAAAT	7
RPL28	AGAGCTGGAAAAA	
	AGCGCTGCCAGTA	8
RPLP0	ACAGCTGCTACGA	
	TCTACTGCAACAA	8
RPL14	CGAGCTGTCAAAA	
	CGAGCTGTCATTG	10
RPL18	CGTGCTGTTAAGG	

	CCAGCTGTCAAAT	8
RPS17	TCTGCTGCCACCG	
	CGTGTTGCCAACA	8
RPL8	CCAGACGCCAAAC	
	ACAGCTGTCAAAC	9
RPS20	ACTGCTGGCAATT	
	ACAACTGTCAATG	9
RPS8	ACTGCTGTCGAAA	
	GGTGCTGTCAAGT	8
RPL34	GCGGCGGTCAAGG	
	CGGGCTGTCAAAA	8
RPL13A	AGTGCTGACGGAA	
	GGAGCTGTCACTC	6
RPL35A	AAATCTGTCAGAA	
	AGAGCTGCCAGAC	9
RPS3	GCAGCTCTCCAAC	
	GCAGTTGTCAAAA	9
RPL27	ACGGCTGCCACTG	
	ACAGCTGTCAAAC	8
RPS27	CCAGCTGTTCAGC	
	TTTGCTGTCAACA	6
RPL3	CGTACTGCGAAAA	
	CGTACTGTCAAAC	10
RPS25	GCAGCTGTAAGAT	
	GCAGCTGTCAAGC	9
RPS7	GCAATTGTTACCA	
	GCAACTGTCAACT	8
RPL32	AAAGCTGACAGAA	
	ATAGCTGTCACAT	9
RPL24	ACAACTGTCAACT	
	ACAACAGTGAAAA	9

RPLP2	ATAGCCGGCAAGG	
	AAAGCTGTCAAGA	9
RPL13	ACCGCTGTCCAAG	
	ACTGCTGTCACCC	8
RPS2	ATTGCTGTTAAAT	
	TCTGTTGTAAACA	7
RPL21	ACAGCTGTTTAAG	
	TGAGCTGTCAAAC	8
RPL30	TGAGCCGTCACAA	
	GGTTCTGTCAAAA	8
RPL27A	GGTGCTGTTAAGC	
	GCCGCTGTCAAAC	9
RPL9	GGTGCTGTTAACG	
	TTCGCTGTCACAT	6
RPS27A	GCCGCCGTCGAAG	
	GTCGTTGTCAAAA	7
RPS13	AGAGCTGTCAACG	
	ACTGGTGTGAAAT	7
RPS21	CCAGCTGGCAAAG	
	ACAGCTGTCAAAT	10
RPL7	CGGGCTGTCAAAT	
	CGGGCTGTCAAAT	13
RPS26	ACTGCTTACACTA	
	TCTGCTGTCACGG	8
RPL37A	GCGGCGGTCACAC	
	TGATCTGTCAAAC	7
RPS9	ACAGACGTGAAAG	
	ACCGCTGTGATCG	7
RPS19	GCCGCGGCCAAAT	
	GCGGCTGTCAGAG	8
RPS14	GTAGCTGTCAAAA	

	CCGGTTGTCAAAT	8
RPS6	CCTGGTCTCACAA	
	GCAGCTGTCAAGC	6
RPS15A	GTTGCTGTCAGCA	
	GCGGCTGTCACGG	8
RPL7A	GCTGCTTCCAAAG	
	GCTGCTGTCATGT	8
RPS10	ATTGCTGACACAT	
	TTATCTGTCAAAT	7
RPL22	GCTGCGGCCACAC	
	ACTGCTGCCAACT	8
RPL36	CCAGCTGTCCAAG	
	CGTGCTGTCAAAC	9
RPL35	GCTGCTGTTACAG	
	TCTGCTGACAACT	7
RPS5	CCGGCTGCCAGGG	
	CCAGCTGTCAGAC	9