Kevin Zhang
CMSC 27600
Project 3 Writeup
May 29, 2013

The following are the calculated motifs for different values of W, with all other input parameters the same.

Motif length (W)	Motif sequence
2	CG
3	GAC
4	CGAT
5	**CGATA**
6	GATACG
7	AGATCCG
8	GATACGTC

Proposed method for finding optimal W

One way to find a good value of W is to run the algorithm across a range of W, align the reported motif against each of the input sequences, and choose the W that yields the highest-scoring alignment. In this particular case, a local alignment would be most appropriate since we want to find a subsequence common to all of the input sequences, which are otherwise dissimilar.

Another method, as proposed by Lawrence et al., is to use the incomplete-data log-likelihood measure and information-per-parameter criterion. As in the first proposed method, we would iterate over W and find the value that yields the highest information-per-parameter.