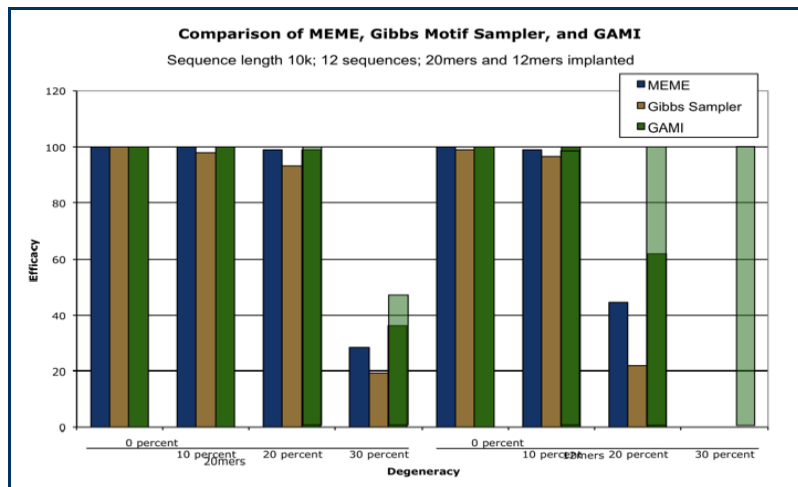


A Genetic Algorithms Approach to Regulatory Motif Discovery

Congdon, CB^{1,2}, Aman, J² Leighton, B³ and Mattingly, C². 1: Department of Computer Science, University of Southern Maine, Portland, ME 04104; 2: Mount Desert Island Biological Laboratory, Salisbury Cove, ME 04672; 3: Department of Computer Science, Colby College, Waterville, ME 04901

Background and Objective: GAMI, Genetic Algorithms for Motif Inference, is designed to find putative conserved regulatory regions in non-coding sequence. Several studies suggest that comparative analysis of evolutionarily diverse organisms will help to predict functionally important non-coding regions. GAMI is an effective tool for searching large datasets (long sequence lengths and many sequences) of divergent species, finding known transcription factor binding sites (TFBS) referenced in other published work and finding the best motifs identified by exhaustive search. In this work, we evaluate GAMI's inference abilities as compared to MEME and Gibbs Sampler.

Methods: Thousands of artificial datasets were constructed by generating randomized sequences implanted with motifs of varying lengths and degeneracy. The three systems were run on these variations.



Results: Fig 1. shows results from the datasets with 10k long sequences and implanted 20mers and 12mers, and varying levels of degeneracy. Green shading (as opposed to solid green) indicates that GAMI found a stronger pattern in the background sequence.

Fig. 1. Comparison of MEME, Gibbs Motif Sampler, and GAMI on artificial data with implanted motifs.

Discussion and Conclusions: All three systems found implanted motif less often as the degeneracy increased; GAMI and MEME found the implanted motif more often than Gibbs Sampler. More work is needed to better compare these systems.

Support: P20 RR-016463 NIH/NCRR.