Hidden Markov Models (HMM) are flexible statistical models that have gained tremendous popularity in Bioinformatics over the past decade. One of the primary advantages in the use of HMMs for modeling sequence information is that the sequential structure can often be represented in the topology of the model (i.e. Profile-HMMs). There are a variety of HMM based tools that are currently used by bioinformatisits for sequence analysis tasks including protein family classification (PFAM), gene finding (GLIMMER), and Splice Junction prediction (GeneSplicer), and new HMM tools are continuing to be developed.

HMMs are typically constructed with a supervised training approach using known observations (sequence samples) of the system to be modeled. Training involves adjusting the transition and emission probabilities of the HMM itself so that the likelihood of the model generating the observations is maximized (i.e find the λ=(A,B,π) that maximizes $P(O|\lambda)$ where λ represent the HMM and O represents the set of known observations). The Baum-Welch algorithm is one of the most widely used methods for HMM training. In general, satisfactory results are obtained using BW training as evidenced by the tools mentioned above. Never the less it is often the case that the training phase is iterated numerous times from different initial starting points to ensure that the resulting model has not convergence to a sub-optimal solution.

As the amount of sequence data available for analysis continues to grown exponentially the trend for developing useful HMM based tools is likely to continue. Thus the problem of how to avoiding training HMMs that converge to sub-optimal parameters is an increasingly important problem for HMM developers. Alternative training methods for HMMs that overcome this problem need to be identified as the variety of and number of problems addressed with these models increases. The HMM training problem can be viewed as a search for an optimal set of HMM parameters in which the search space is, computationally speaking, intractable for even a small HMM. Evolutionary Algorithms (EAs) are a class of population based heuristic search methods often considered to be an excellent choice for large search space problems, especially those in which sub-optimal solutions may dominate the search space (e.g. high-dimensional multi-model fitness landscapes).

We have developed a framework for systematic investigation of the use of Evolutionary Algorithms (EAs) as an alternative method for training HMMs. Our initial results indicate that EAs can be successfully used for training HMMs when a systematic approach to identification of the EA parameters is utilized. Our approach also provides information about the resulting HMM parameters that may increase researchers ability to extract biologically useful information from the trained HMM itself. Previous attempts to use EAs for training HMMs for particular problem instances and have been somewhat successful but are not easily transferred to new problems. Our approach has the long term goal of being able to generalize EA training in such a way as to make it readily available as a general training alternative to the BW approach.