

# Algorithms in Bioinformatics II, SS2003

## Assignment sheet # 4

Daniel Huson

May 19, 2003

### 1 Viterbi for Genscan's model (3 points)

First summarize how Genscan's model differs from a simple HMM.

Then, based on the pseudo-code version of the Viterbi algorithm for HMMs on page 202 of the script, discuss which modifications are necessary to adapt the algorithm to work for the Genscan model.

### 2 Viterbi for Twinscan's model (1 point)

Show what additional modification(s) is/are necessary to extend the adaption to the Twinscan model.

To solve the following assignment, please download the file

[www-ab.informatik.uni-tuebingen.de/teaching/ss03/abi2/java/assign04.zip](http://www-ab.informatik.uni-tuebingen.de/teaching/ss03/abi2/java/assign04.zip).

### 3 A WMM program for detecting sites (3 points)

The class `WMMBase.java` provides a basic implementation of a WMM. Write a new class `WMM.java` that extends `WMMBase.java` and provides a method `reportSites(String seq, double threshold)` that, given a sequence and a threshold, reports all sites in the sequence for which the sum of the logs of the probabilities centered at that position exceeds the given threshold.

Add a `main` method that reads a WMM, a threshold and a sequence specified on the command line and then reports all sites returned by the analysis method.

### 4 Application of the WMM program (1 point)

Please apply the WMM program to the sequence given in file `human.fa`, using the following WMMs and

```
start.wmm      -17
thresholds: donar.wmm  -10
              acceptor.wmm -20
```

This should generate a large number of each type of site.

### 5 Apply Genscan (2 points)

Apply the Genscan program (<http://genes.mit.edu/GENSCAN.html>) to the sequence in `human.fa`. Under the assumption that the Genscan prediction is true, how many true start, donar and acceptor sites exist in the given sequence `human.fa`? What does the WMM program suffer from, lack of sensitivity or lack of specificity?

Assignments due: **Monday, May 26, 10am**