



















## HMMS and Gene Prediction The accuracy of HMM gene prediction depends on emission probabilities and transition probabilities. Emission probabilities are calculated based on the base composition in that particular state in the training data. Transition probabilities are calculated based on the average lengths of that particular state in the training data. Exon length boxplots (ICEB, Dosophila melangosiser Exon Dalabase) Evon length boxplots (ICEB, Dosophila melangosiser Exon Dalabase) Homework Question: How do transition probabilities affect the length of predicted ORFs? Weissing et al. A Marks-on Introduction to

## Conclusions

- Hidden Markov Models have proven to be useful for finding genes in unlabeled genomic sequence. HMMs are the core of a number of gene prediction algorithms (such as Genscan, Genemark, Twinscan).
- Hidden Markov Models are machine learning algorithms that use transition probabilities and emission probabilities.
- Hidden Markov Models label a series of observations with a state path, and they can create multiple state paths.
- It is mathematically possible to determine which state path is most likely to be correct.

Weisstein et al. A Hands-on Introduction to