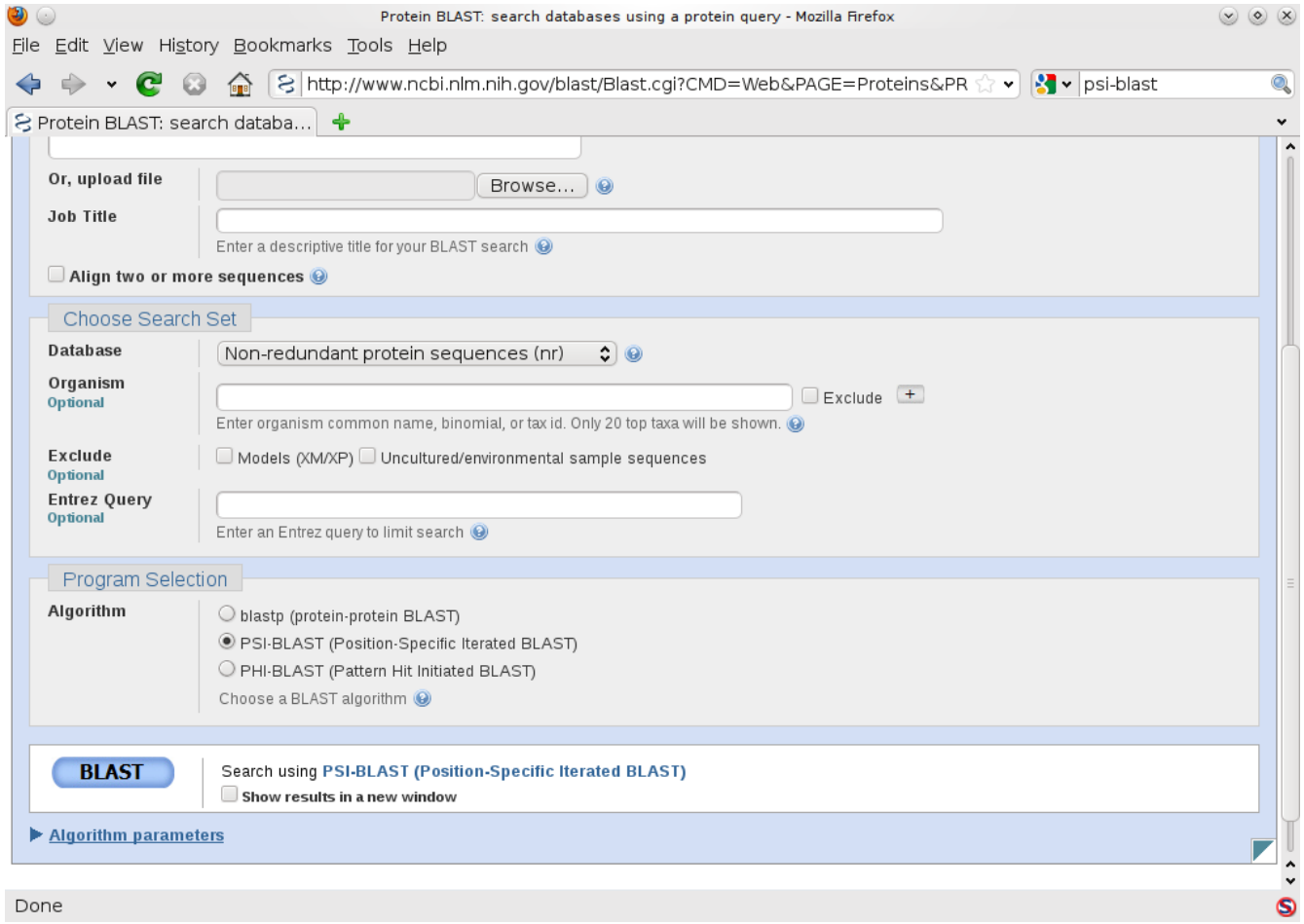
Hidden Markov Models

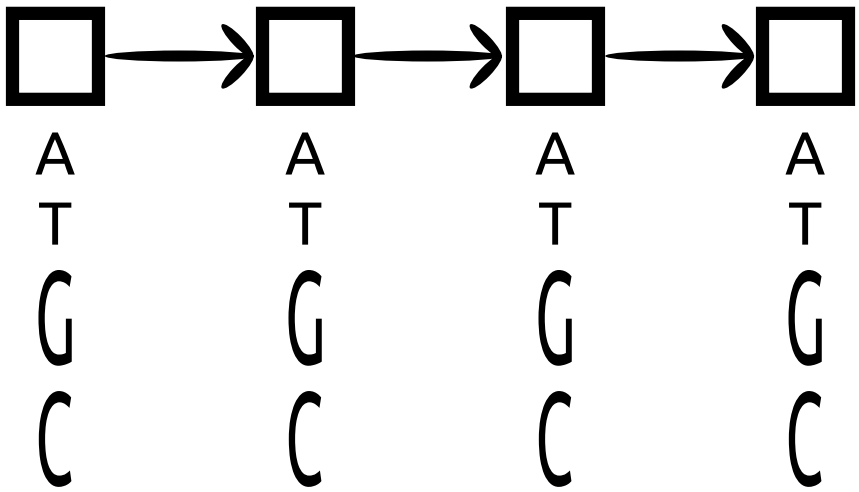
Mark Voorhies

4/2/2012

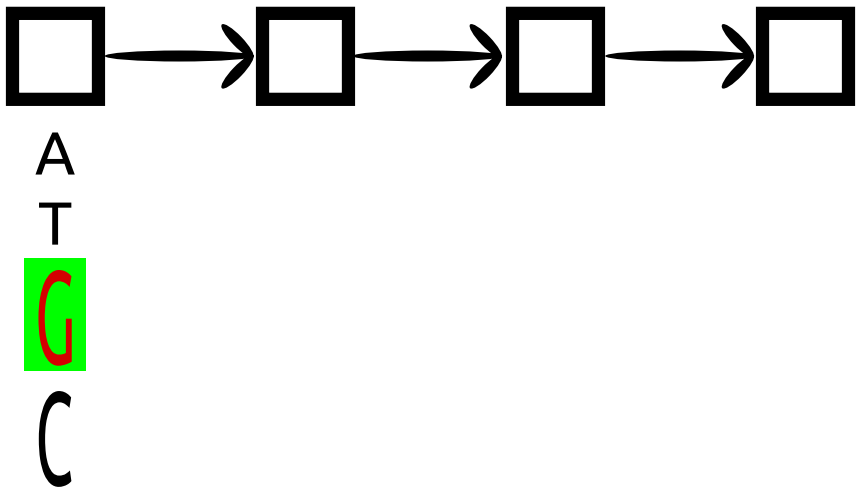
Searching with PSI-BLAST



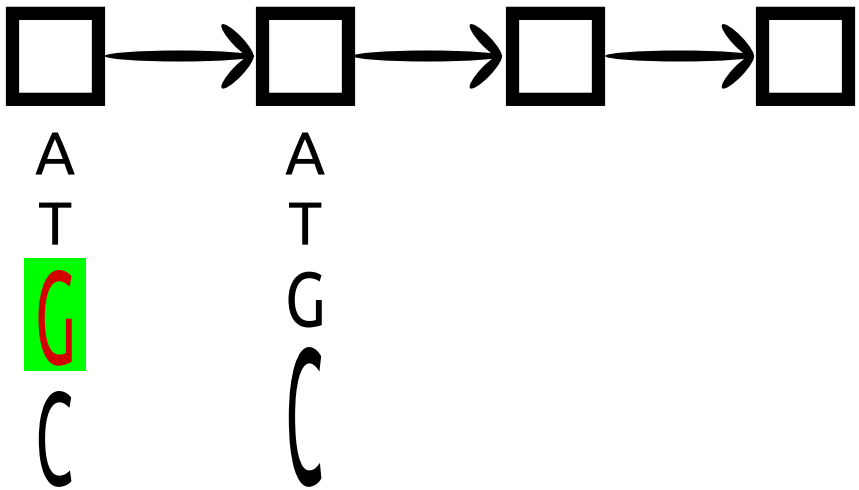
0^{th} order Markov Model



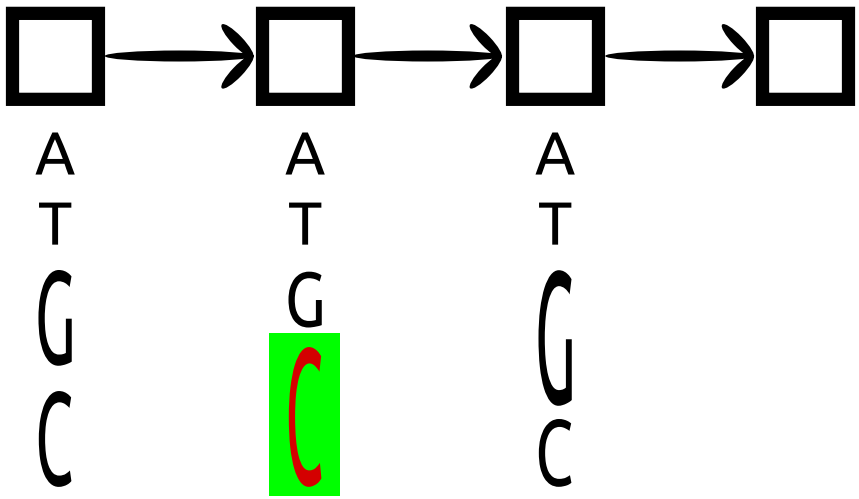
1^{st} order Markov Model



1^{st} order Markov Model



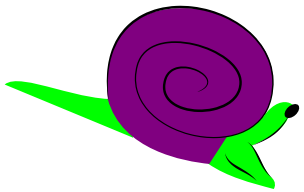
1^{st} order Markov Model

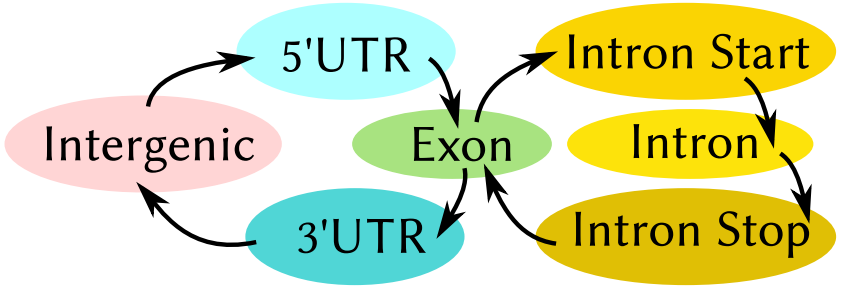


What are Markov Models good for?

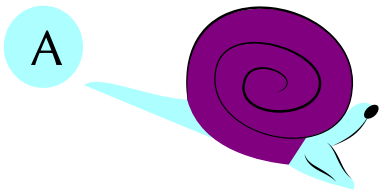
- Background sequence composition
- Spam

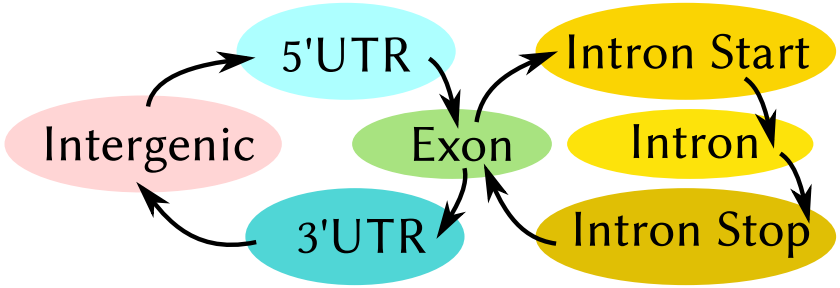
Hidden Markov Models



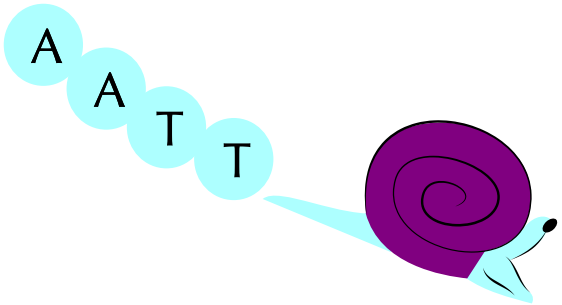


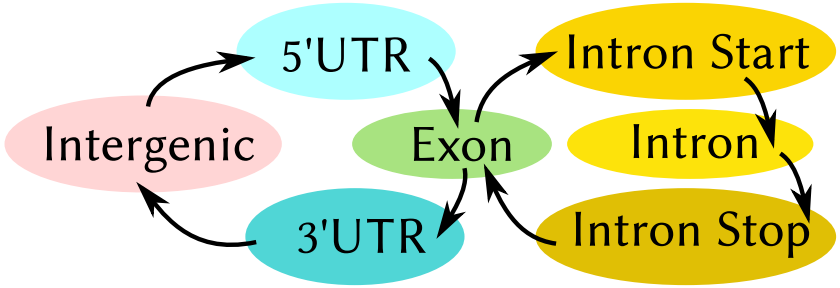
Hidden Markov Models



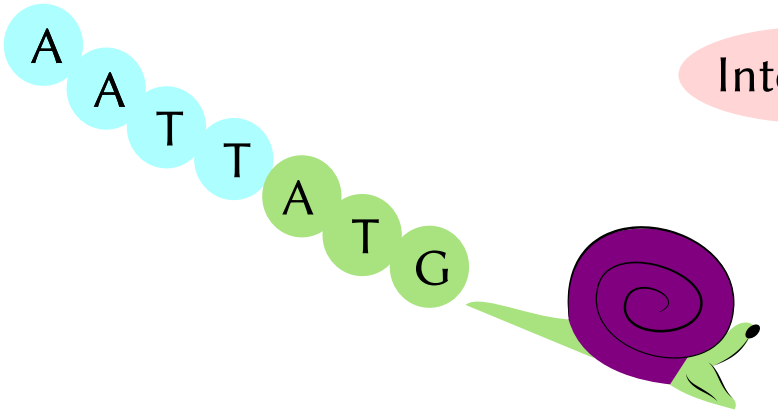


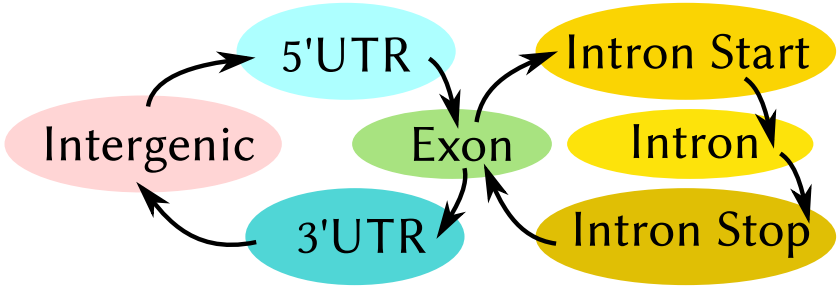
Hidden Markov Models



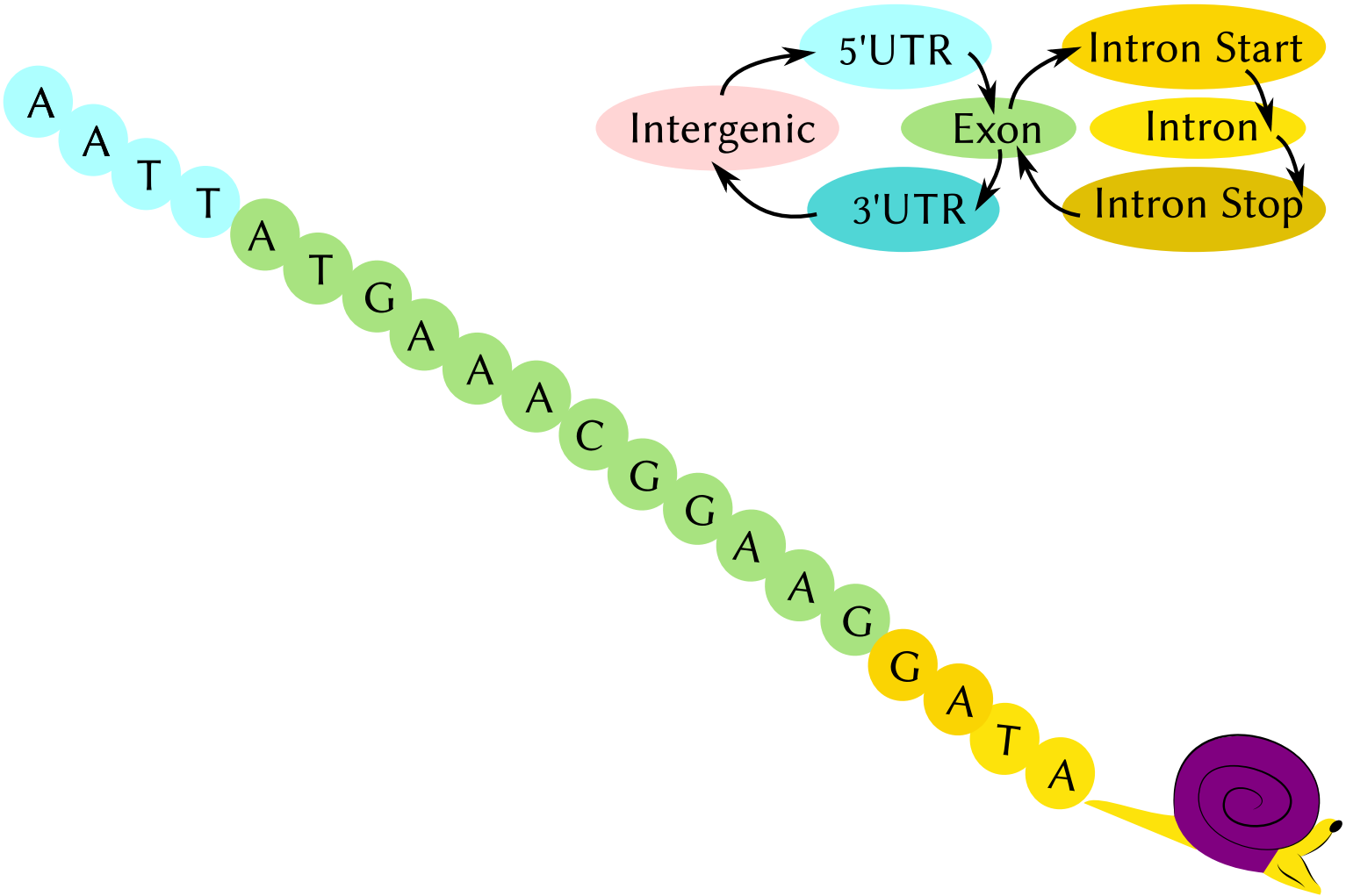


Hidden Markov Models

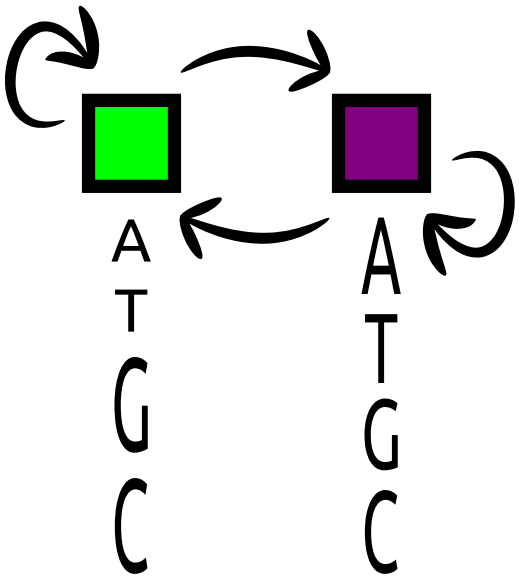




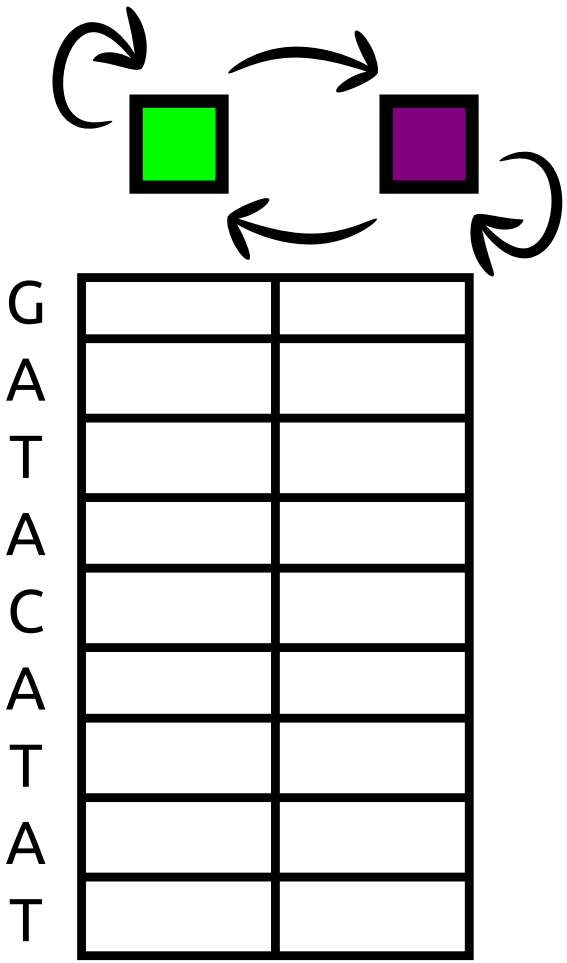
Hidden Markov Models



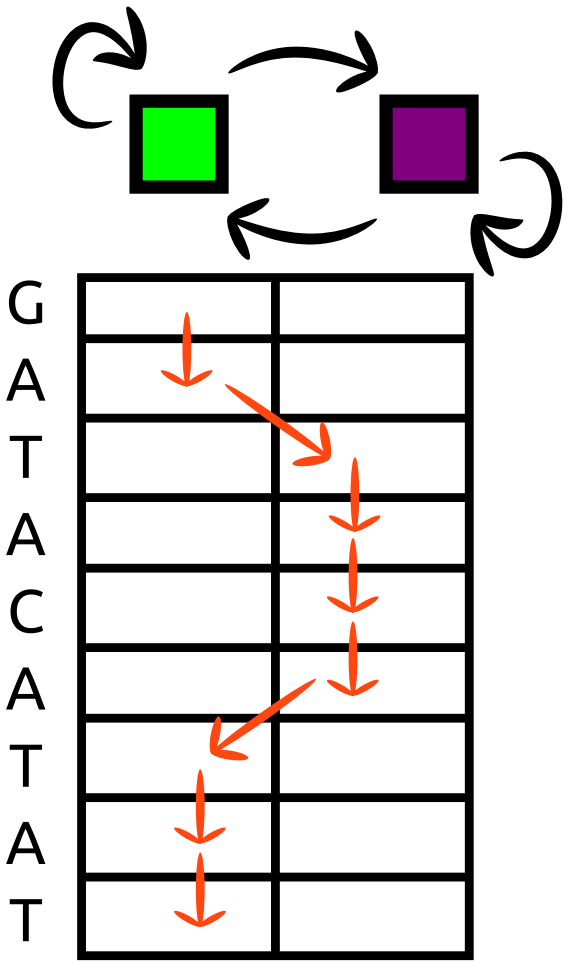
Hidden Markov Model



The Viterbi algorithm: Alignment

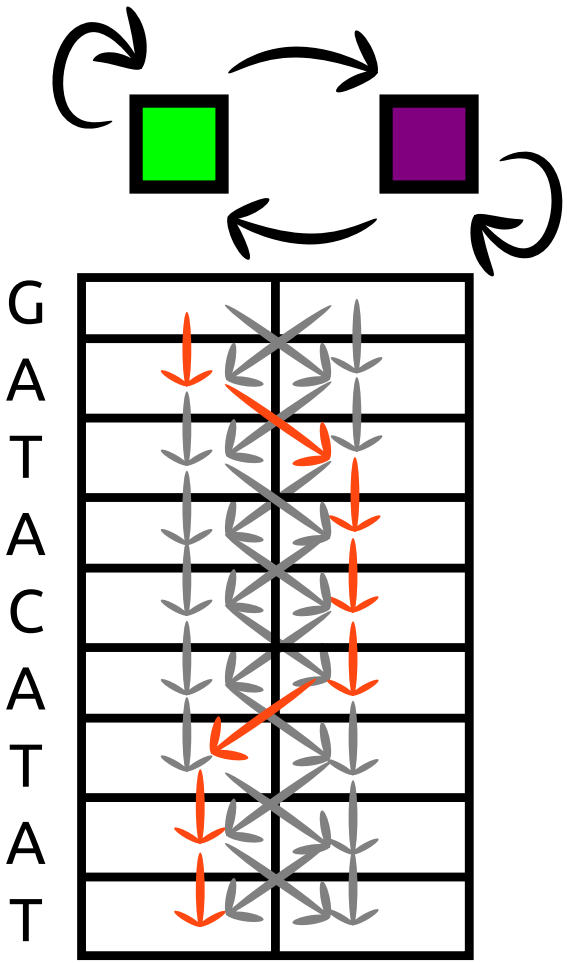


The Viterbi algorithm: Alignment



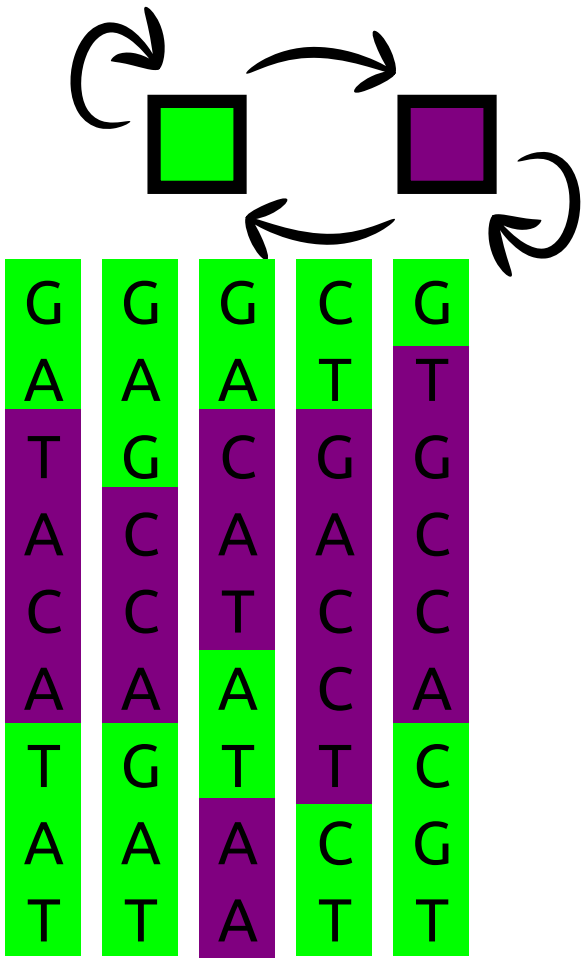
- Dynamic programming, like Smith-Waterman
- Sums *best* log probabilities of emissions and transitions (*i.e.*, multiplying independent probabilities)
- Result is most likely annotation of the target with hidden states

The Forward algorithm: Net probability



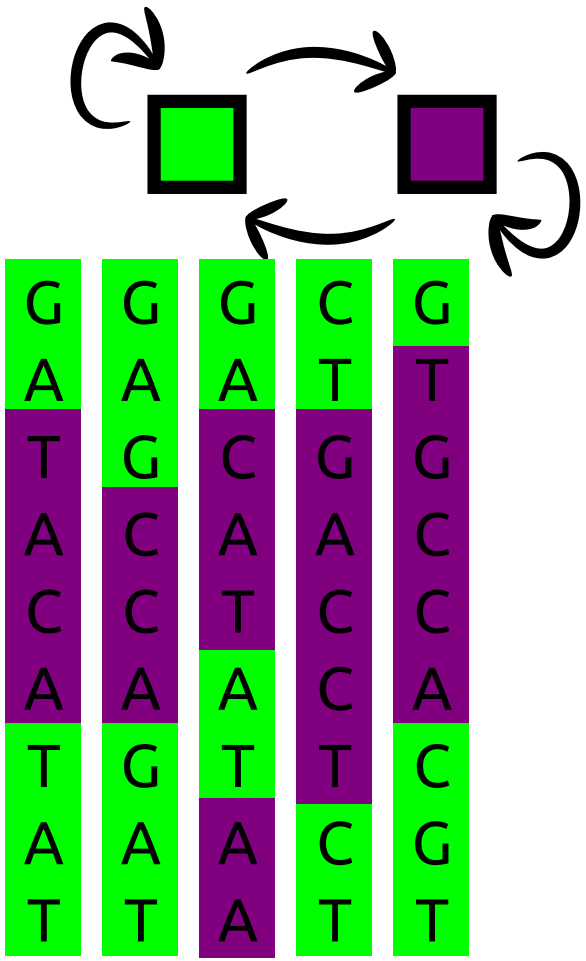
- Probability-weighted sum over all possible paths
- Simple modification of Viterbi (although *summing* probabilities means we have to be more careful about rounding error)
- Result is the probability that the observed sequence is explained by the model
- In practice, this probability is compared to that of a null model (*e.g.*, random genomic sequence)

Training an HMM



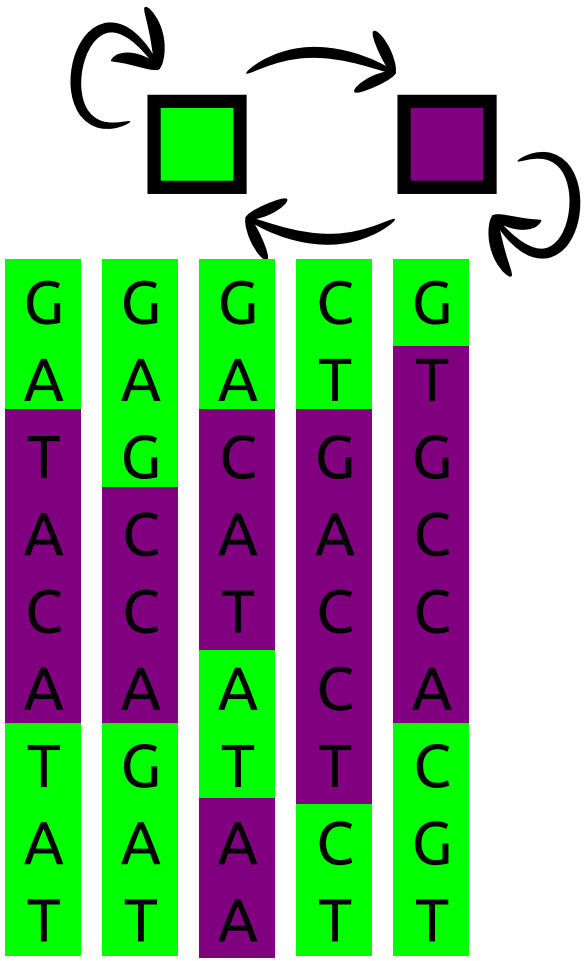
- If we have a set of sequences with known hidden states (*e.g.*, from experiment), then we can calculate the emission and transition probabilities directly

Training an HMM



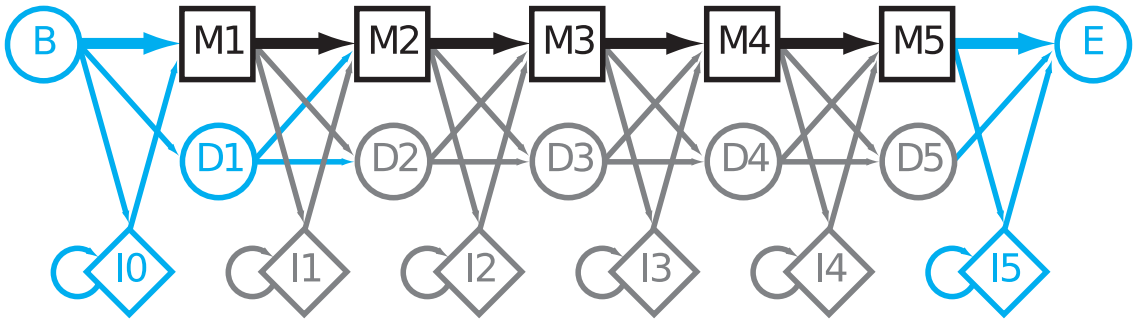
- If we have a set of sequences with known hidden states (*e.g.*, from experiment), then we can calculate the emission and transition probabilities directly
- Otherwise, they can be iteratively fit to a set of unlabeled sequences that are known to be true matches to the model

Training an HMM



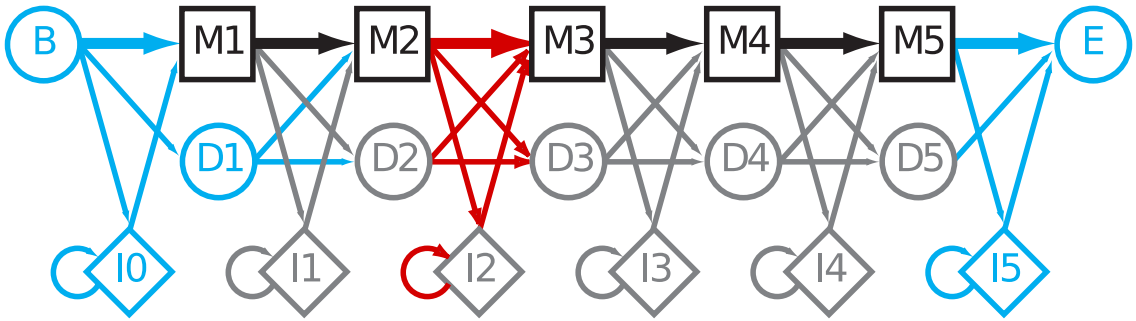
- If we have a set of sequences with known hidden states (*e.g.*, from experiment), then we can calculate the emission and transition probabilities directly
- Otherwise, they can be iteratively fit to a set of unlabeled sequences that are known to be true matches to the model
- The most common fitting procedure is the Baum-Welch algorithm, a special case of expectation maximization (EM)

Profile Alignments: Plan 7



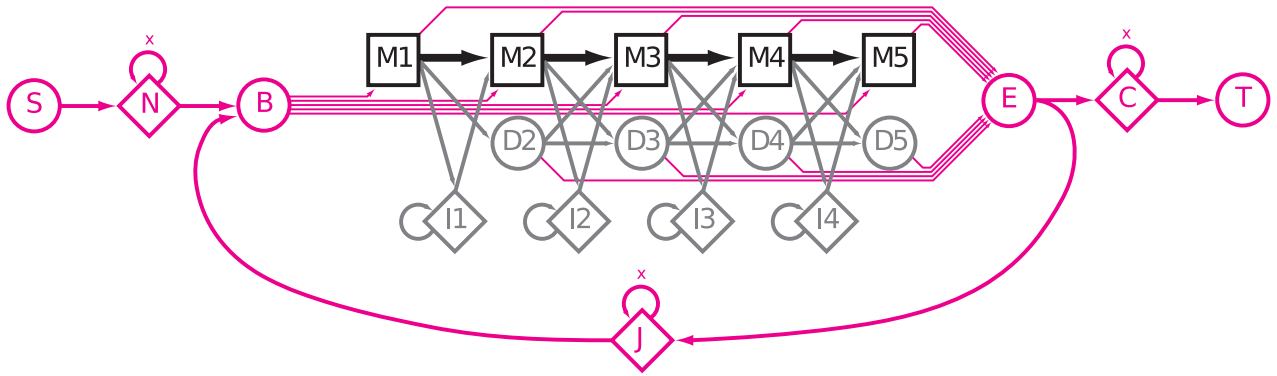
(Image from Sean Eddy, PLoS Comp. Biol. 4:e1000069)

Profile Alignments: Plan 7 (from Outer Space)



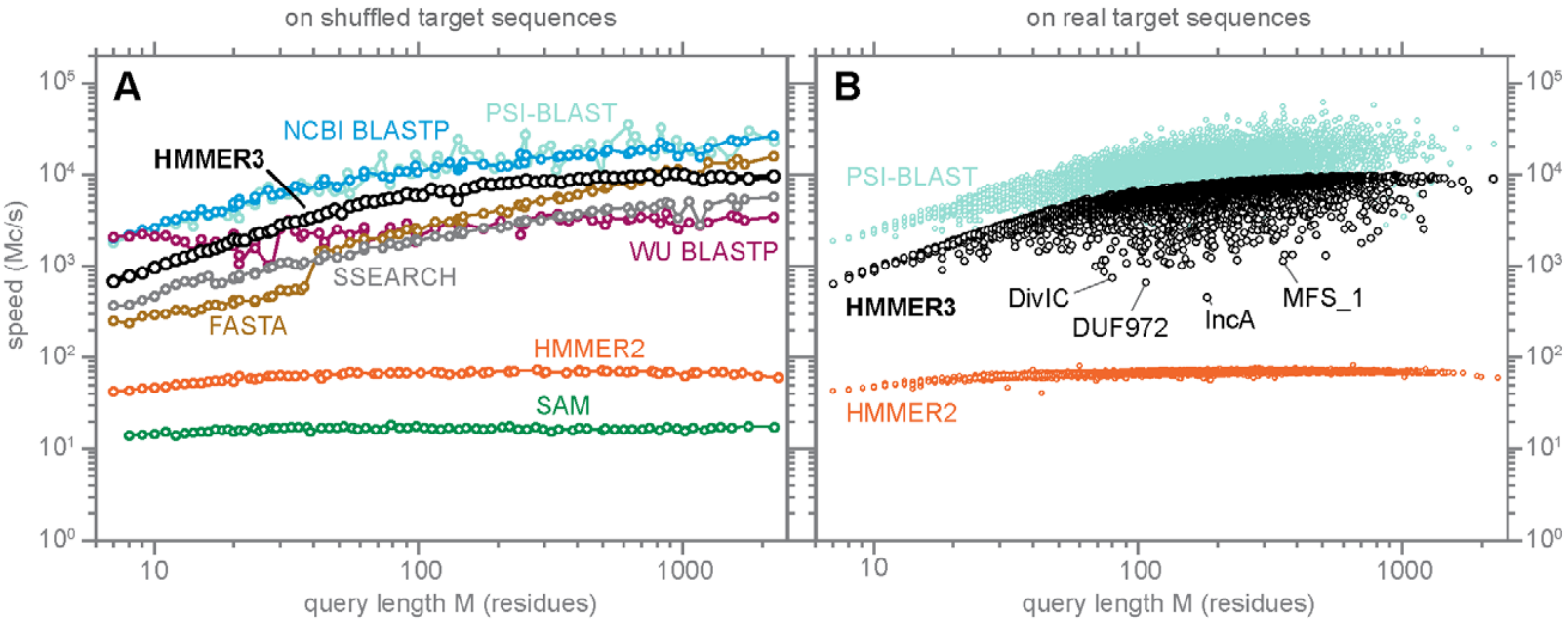
(Image from Sean Eddy, PLoS Comp. Biol. 4:e1000069)

Rigging Plan 7 for Multi-Hit Alignment



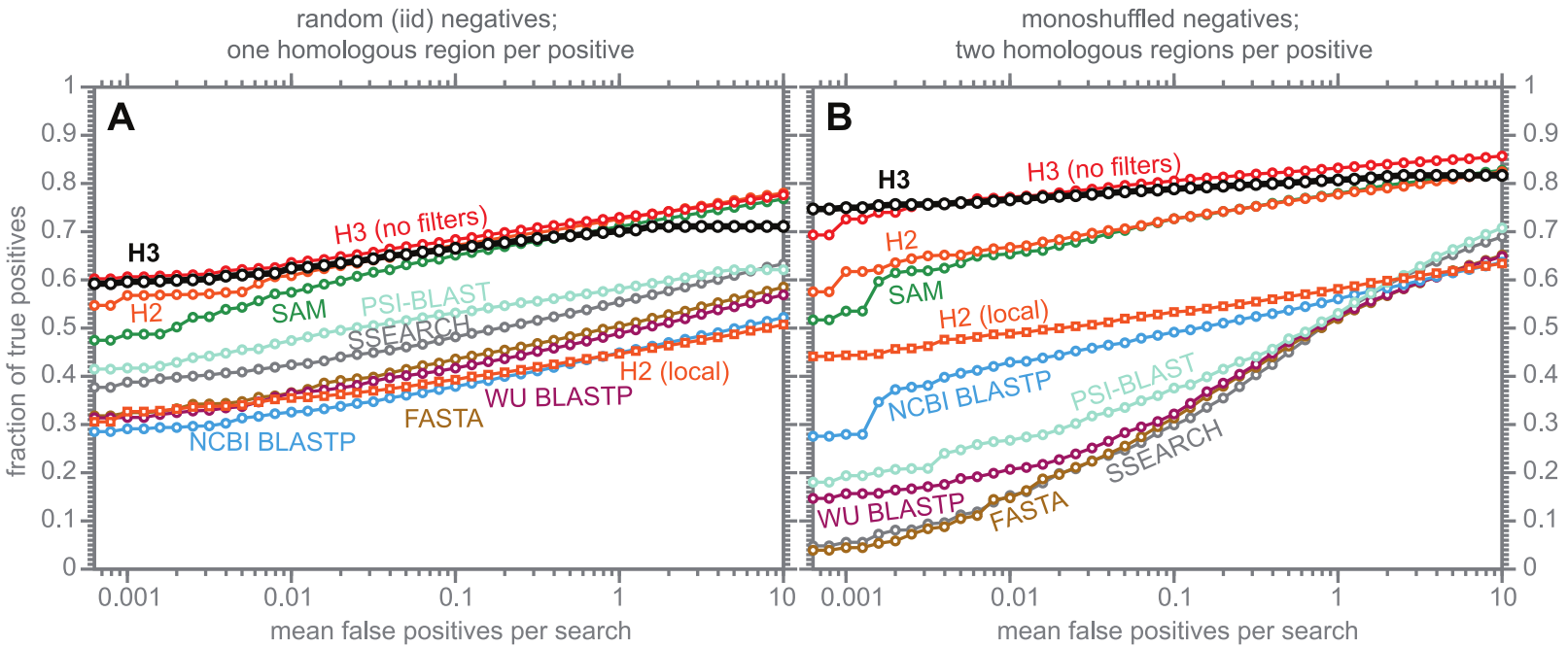
(Image from Sean Eddy, PLoS Comp. Biol. 4:e1000069)

HMMer3 speeds



Eddy, PLoSCompBiol 7:e1002195

HMMer3 sensitivity and specificity

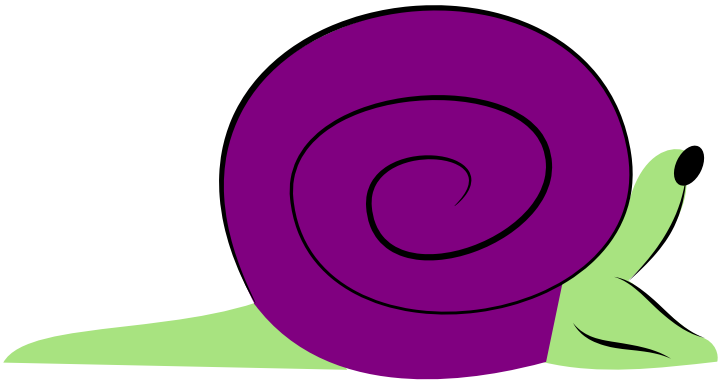


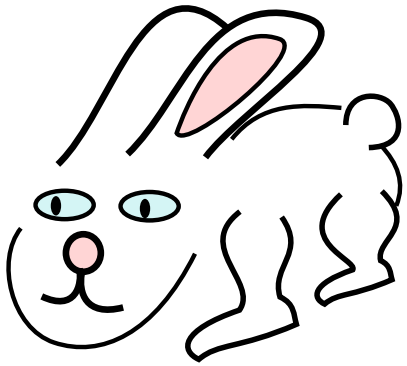
Eddy, PLoSCompBiol 7:e1002195

Homework

- Compare the performance of BLASTP, PSI-BLAST, phmmer, and jackhmmer on a difficult sequence such as AGA1p (CAA96325.1). Use the shuffling tool on the course website to generate negative controls with the same composition. For positive controls, see Euk. Cell 5:628.
- Download Cluster3 and JavaTreeView
- Read PNAS 95:14863

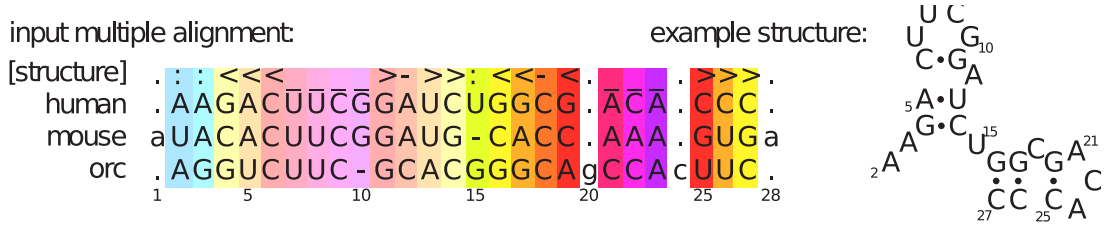
Stochastic Context Free Grammars

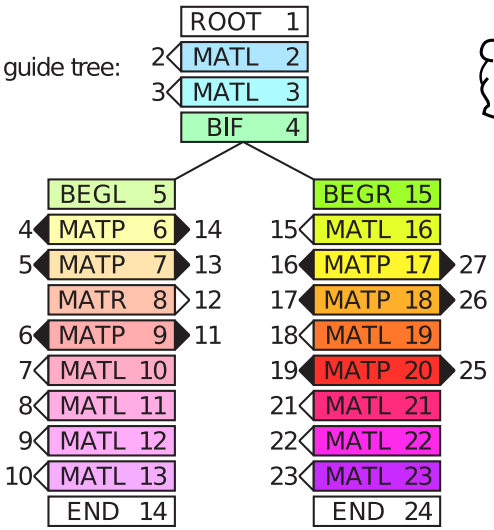




- Can emit from both sides $\rightarrow$ base pairs
- Can duplicate emitter $\rightarrow$ bifurcations

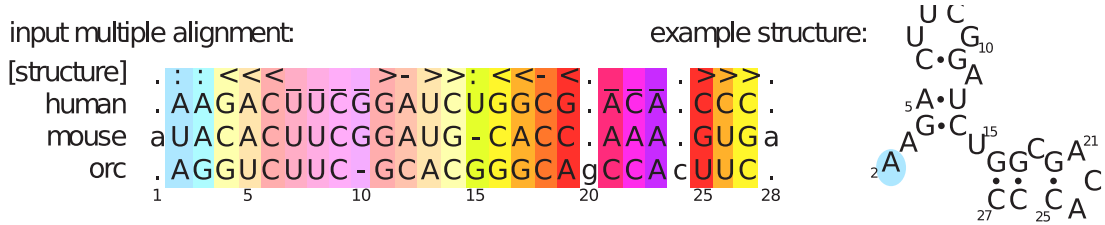
INFERNAL/Rfam

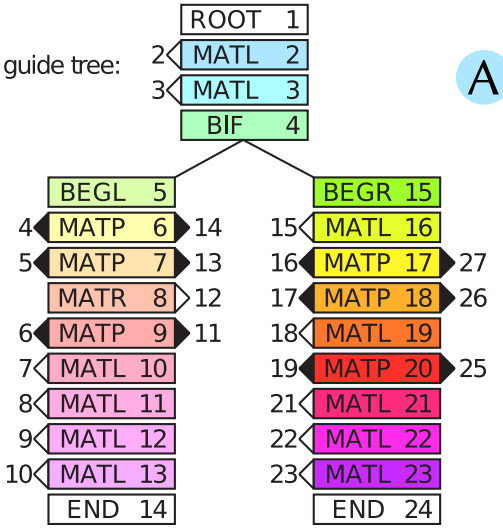


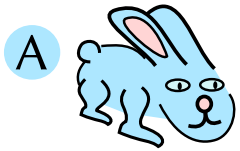


Modified from the INFERNAL User Guide – Nawrocki, Kolbe, and Eddy

INFERNAL/Rfam

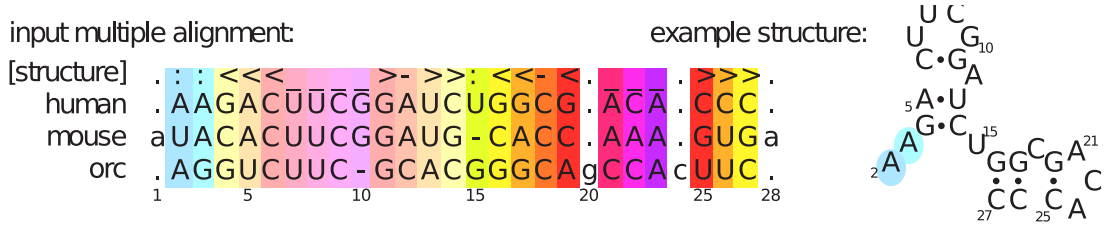


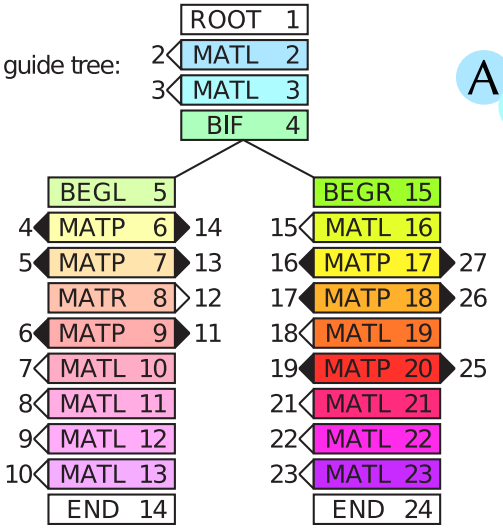




Modified from the INFERNAL User Guide – Nawrocki, Kolbe, and Eddy

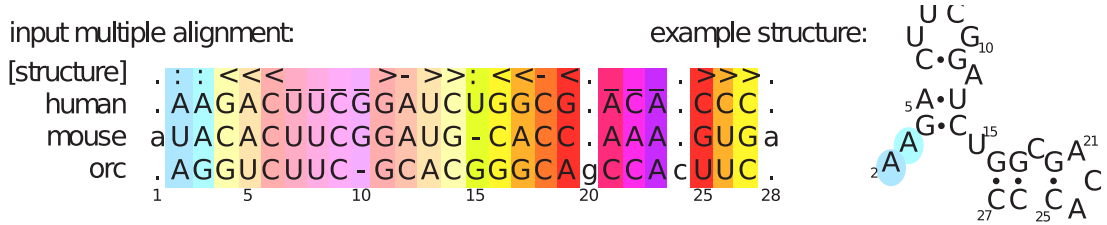
INFERNAL/Rfam

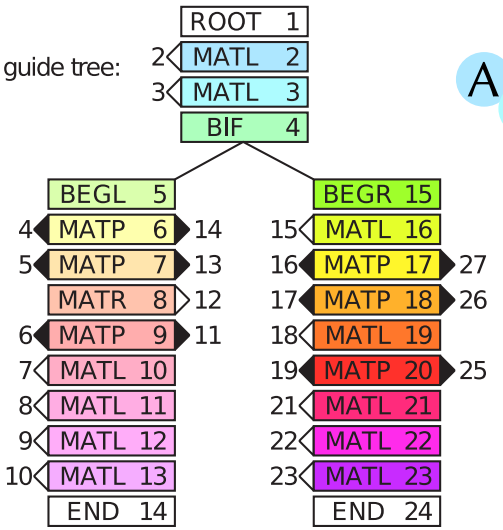


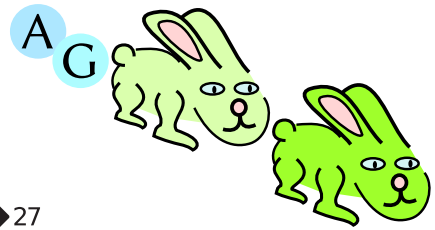


Modified from the INFERNAL User Guide – Nawrocki, Kolbe, and Eddy

INFERNAL/Rfam

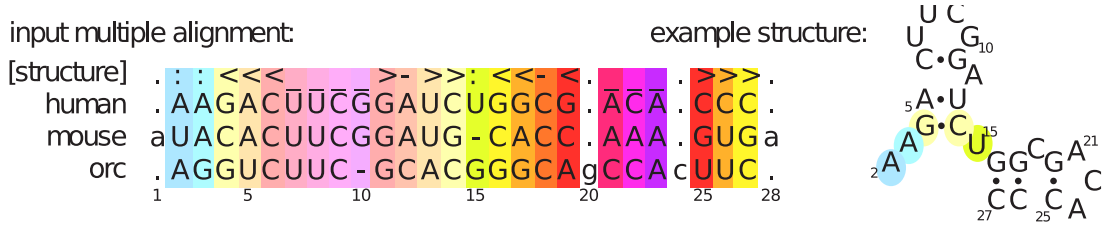


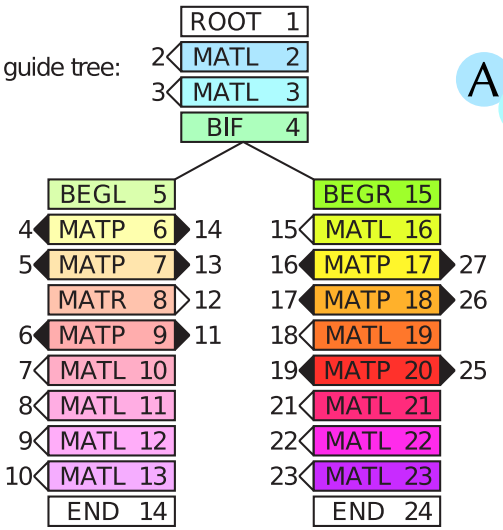


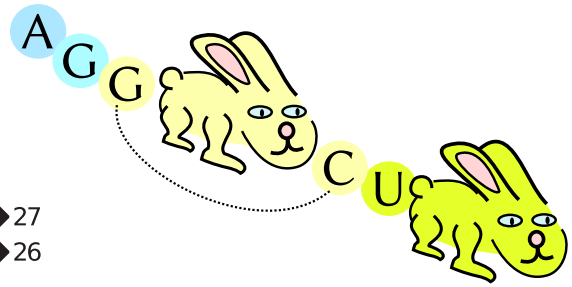


Modified from the INFERNAL User Guide – Nawrocki, Kolbe, and Eddy

INFERNAL/Rfam

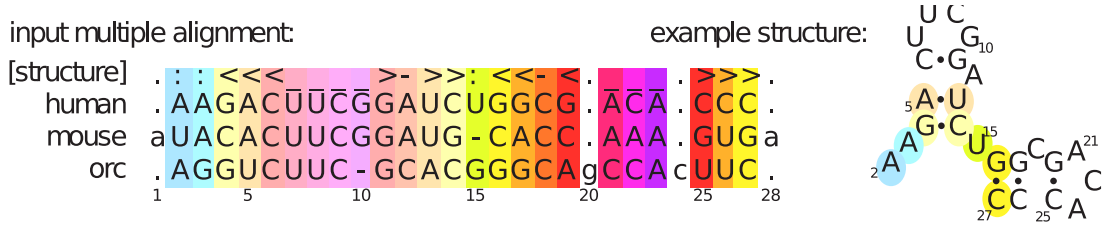


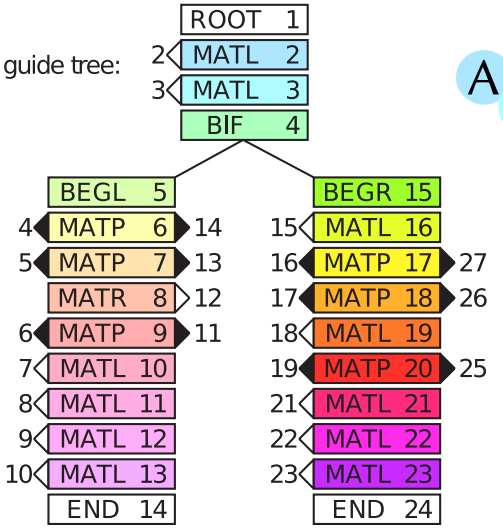


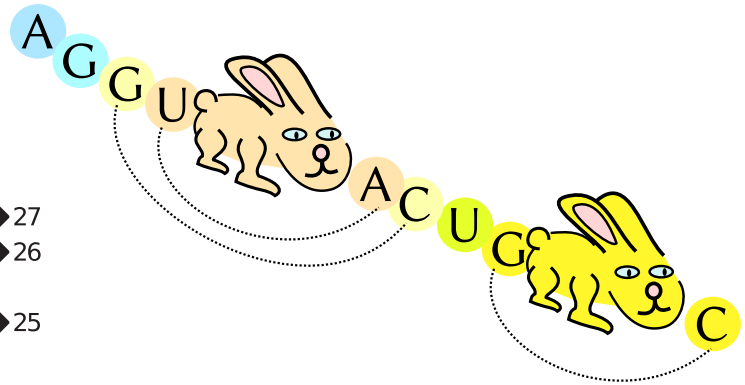


Modified from the INFERNAL User Guide – Nawrocki, Kolbe, and Eddy

INFERNAL/Rfam

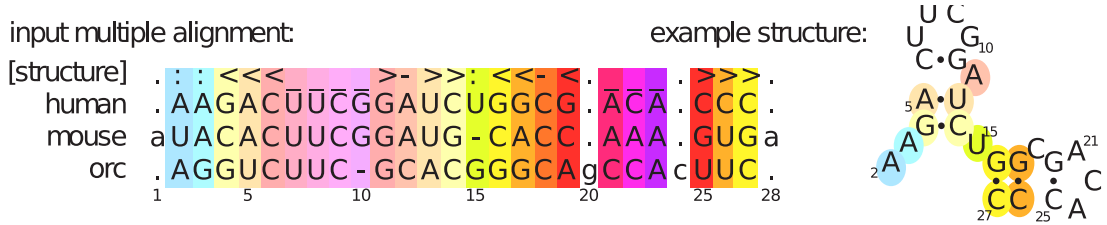


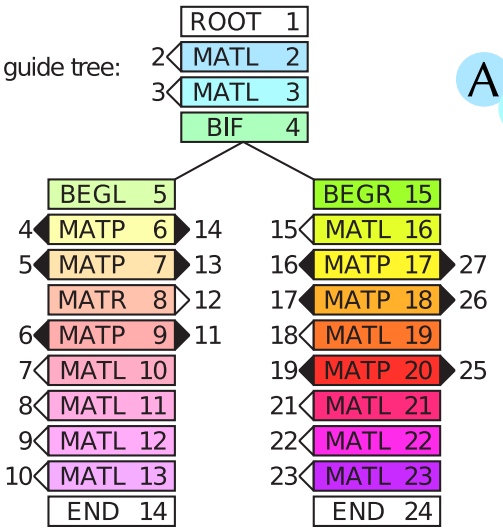


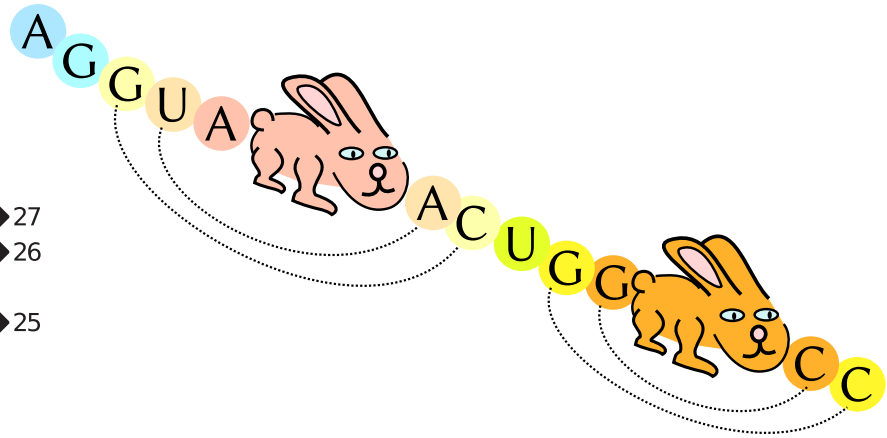


Modified from the INFERNAL User Guide – Nawrocki, Kolbe, and Eddy

INFERNAL/Rfam







Modified from the INFERNAL User Guide – Nawrocki, Kolbe, and Eddy