Practical Bioinformatics

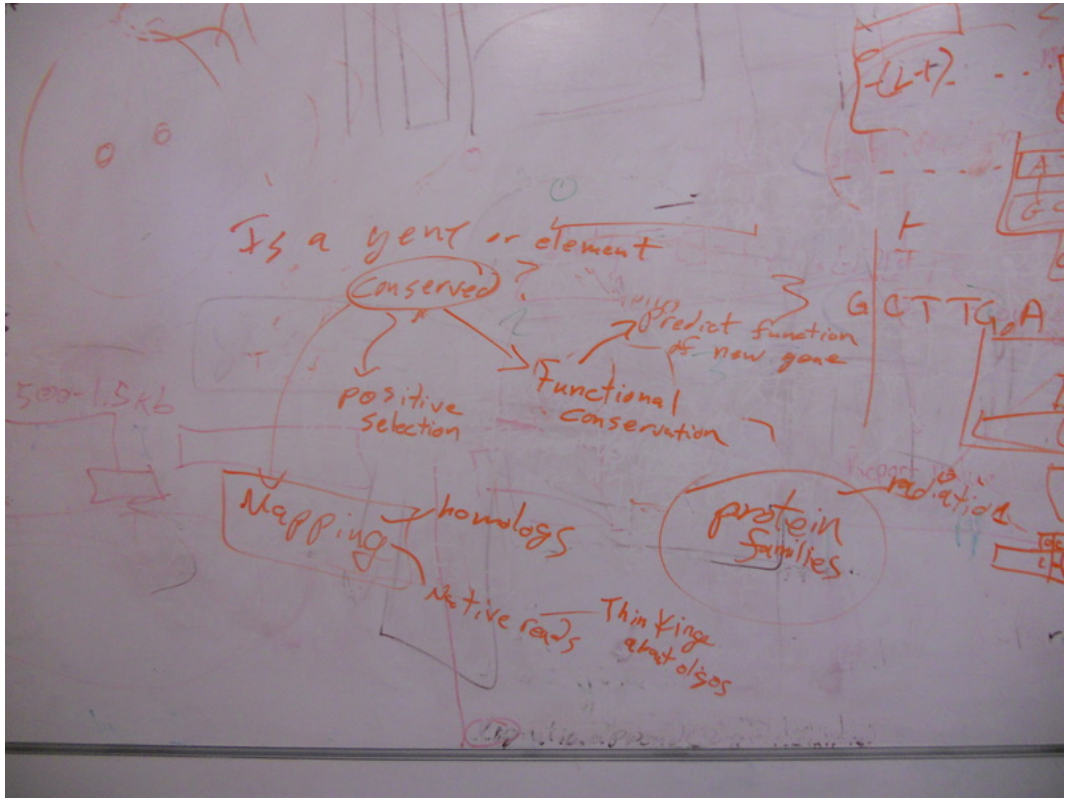
Mark Voorhies

5/26/2015

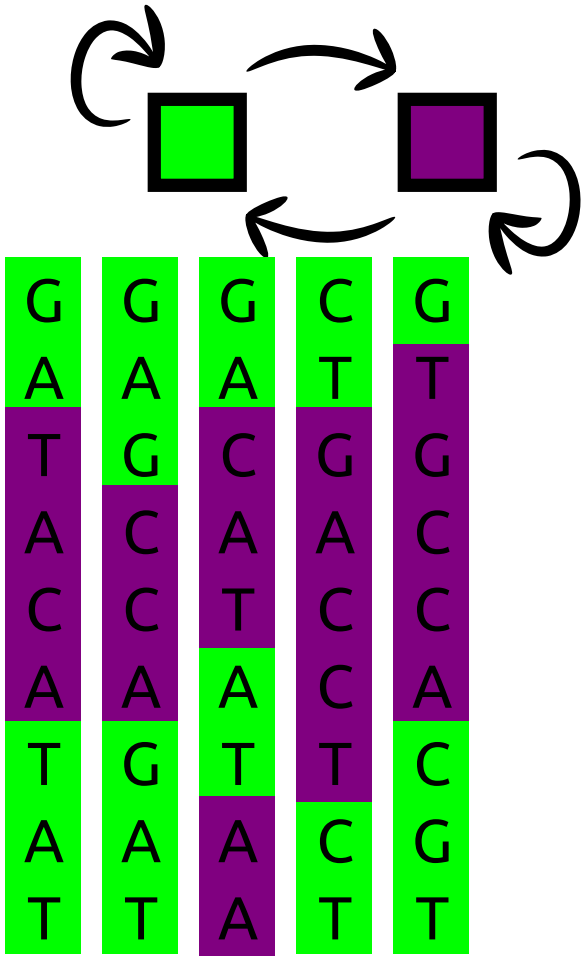
Habits are things you get for free, without requiring any special work.

–Cory Doctorow *Advice to Writers, 4/5/2012*

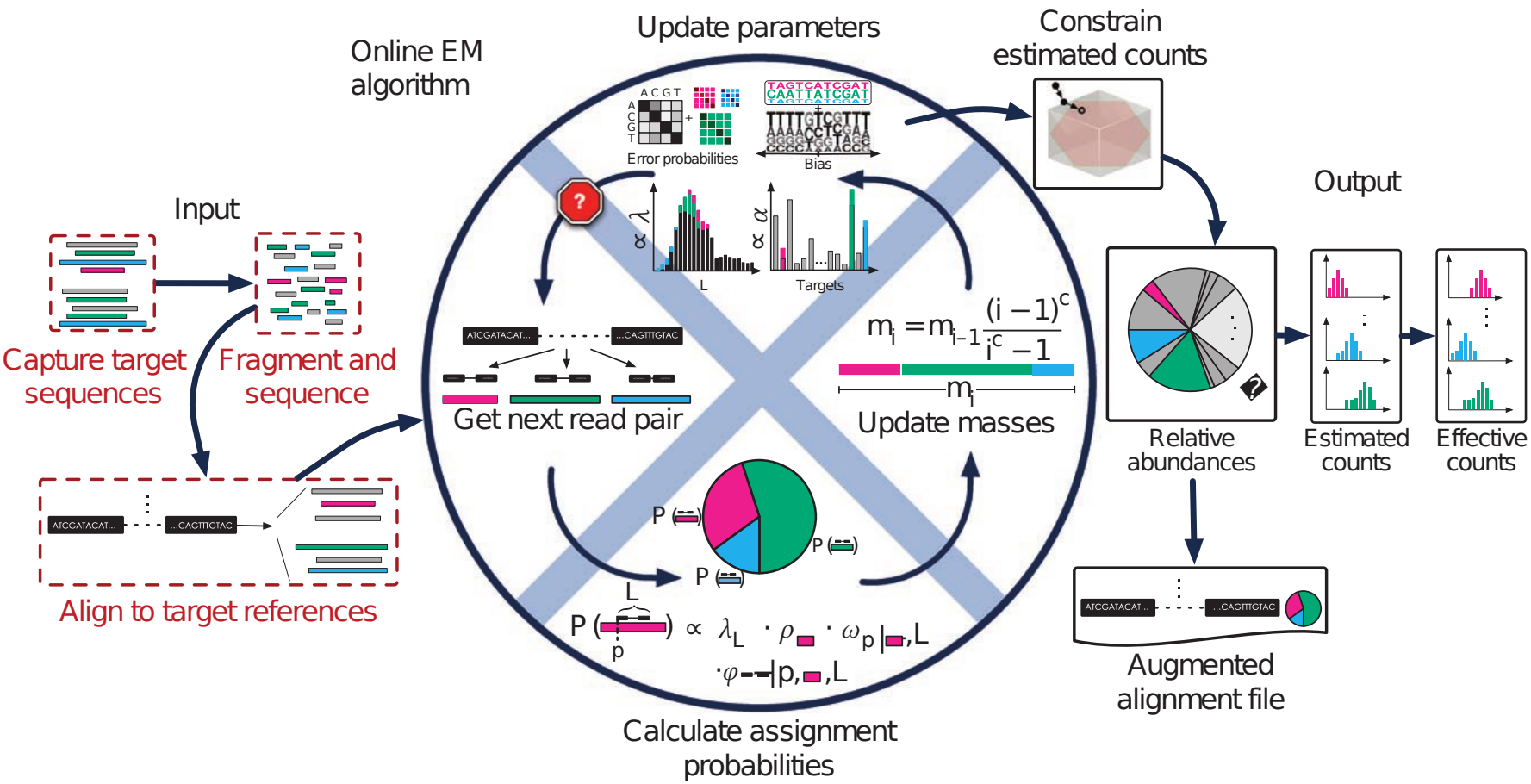
Why compare sequences?



EM: Training an HMM

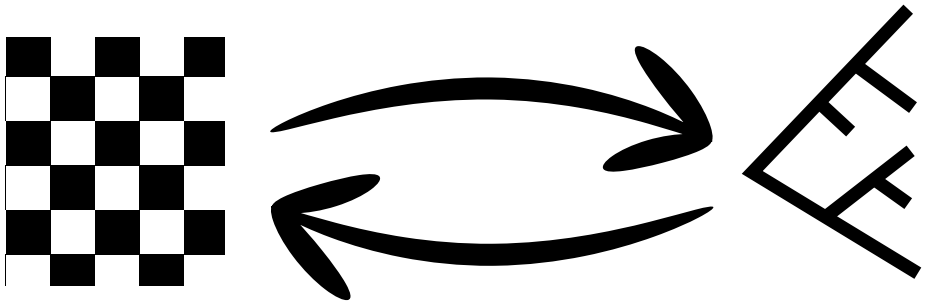


EM: Estimating transcript abundances



Roberts and Pachter, Nature Methods 10:71

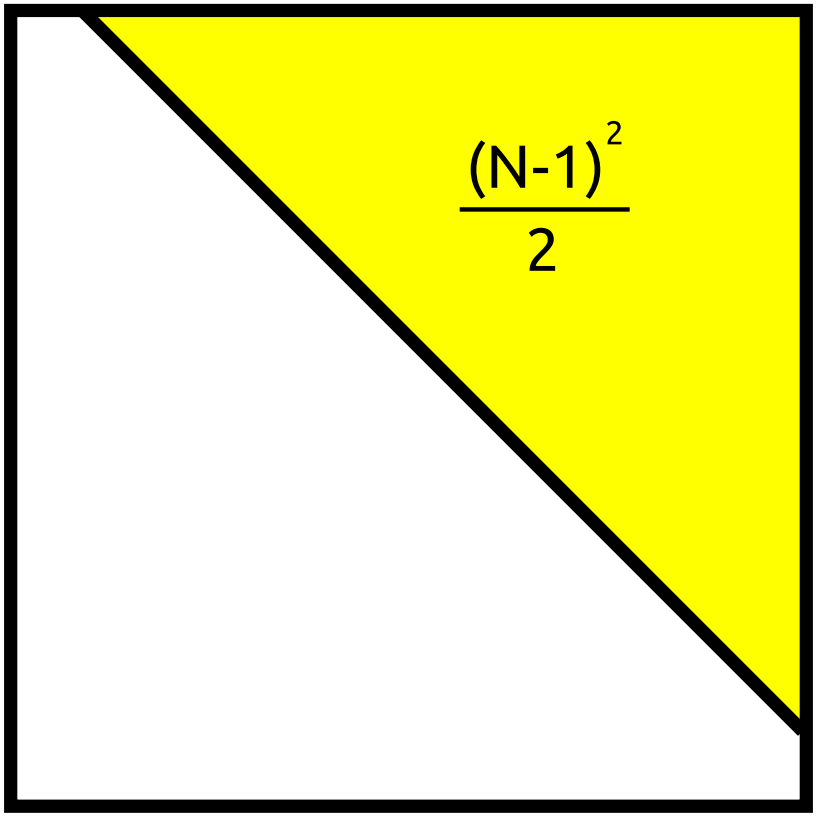
Evolution implies a self-consistent model



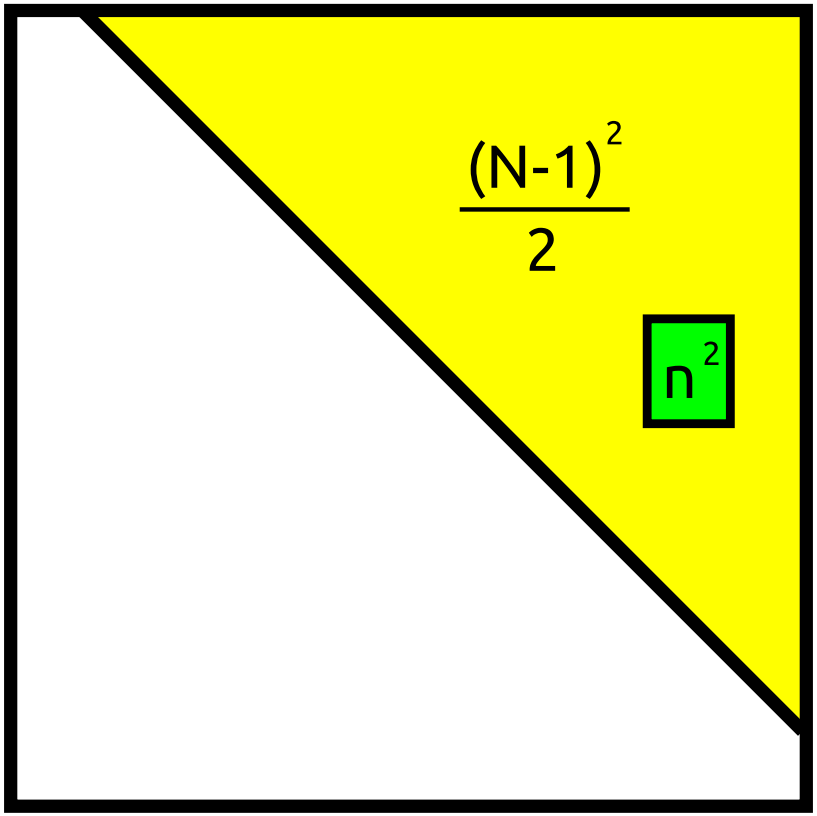
Distances
(Pairwise relationships)

Topology
(Evolutionary history)

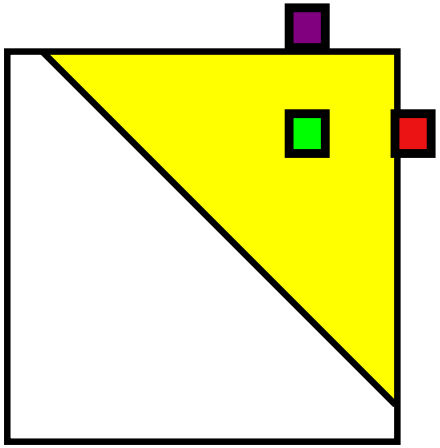
Measure all pairwise distances by dynamic programming

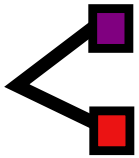


Measure all pairwise distances by dynamic programming

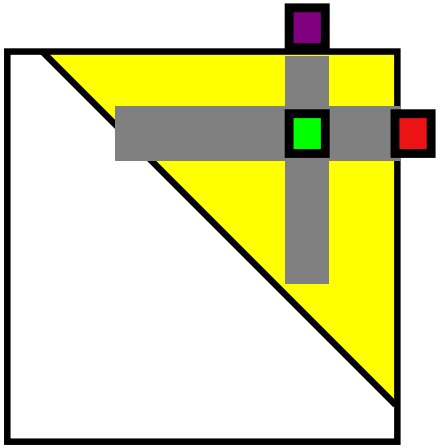


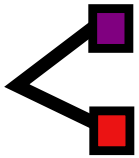
Generate a guide tree by UPGMA



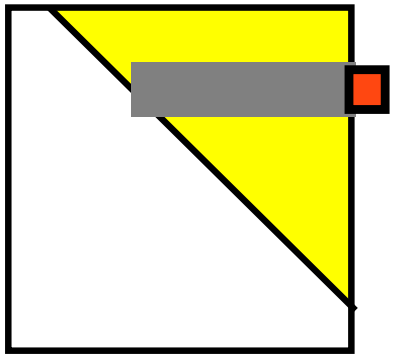


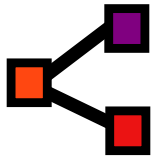
Generate a guide tree by UPGMA



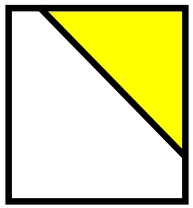


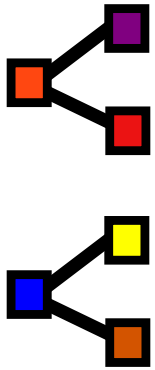
Generate a guide tree by UPGMA



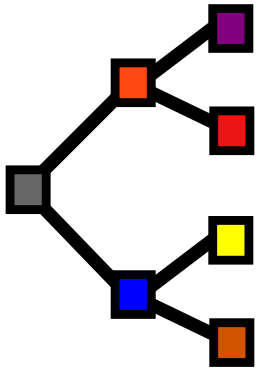


Generate a guide tree by UPGMA

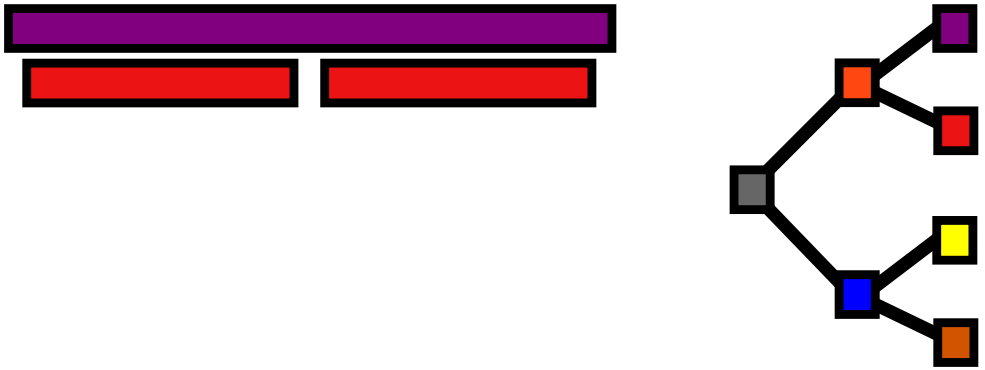




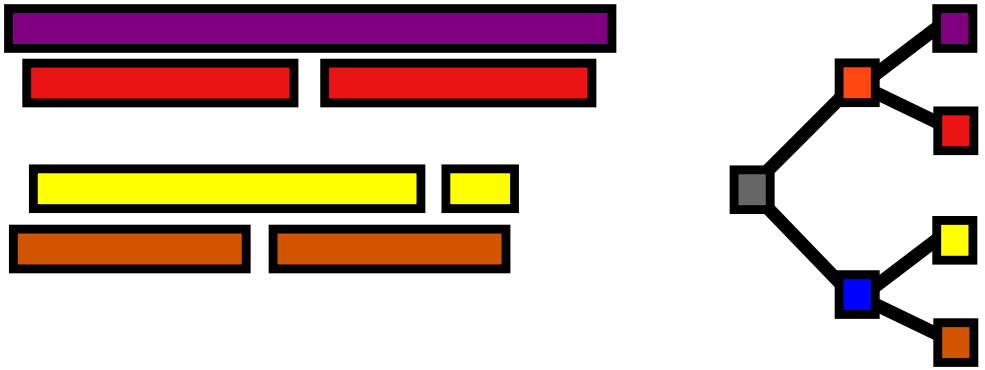
Generate a guide tree by UPGMA



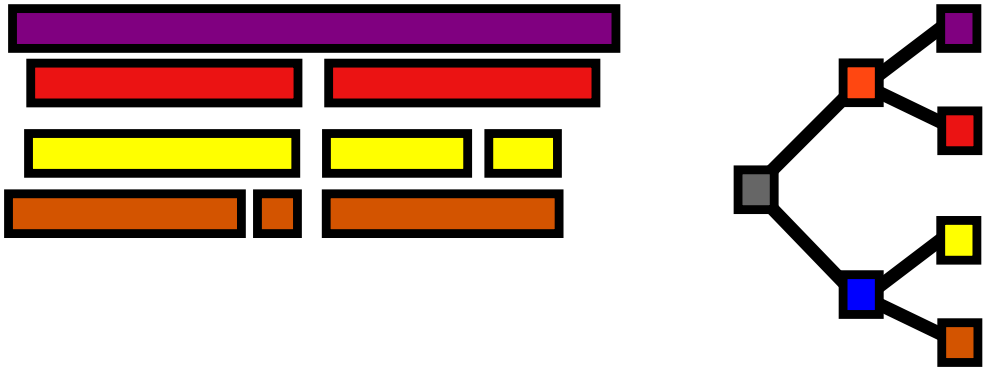
Progressive alignment following the guide tree



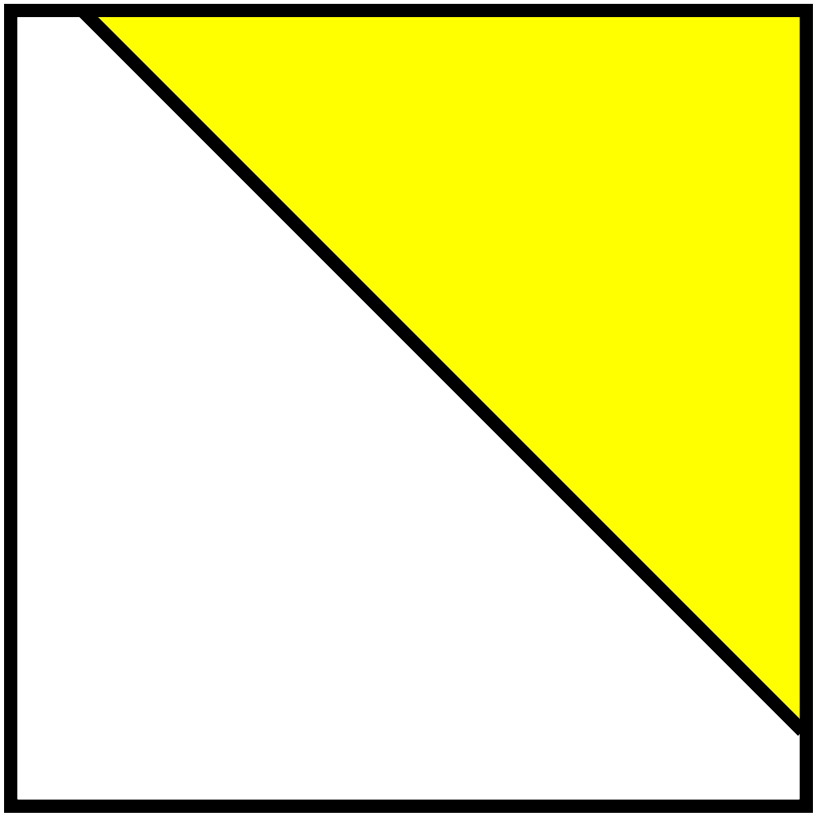
Progressive alignment following the guide tree



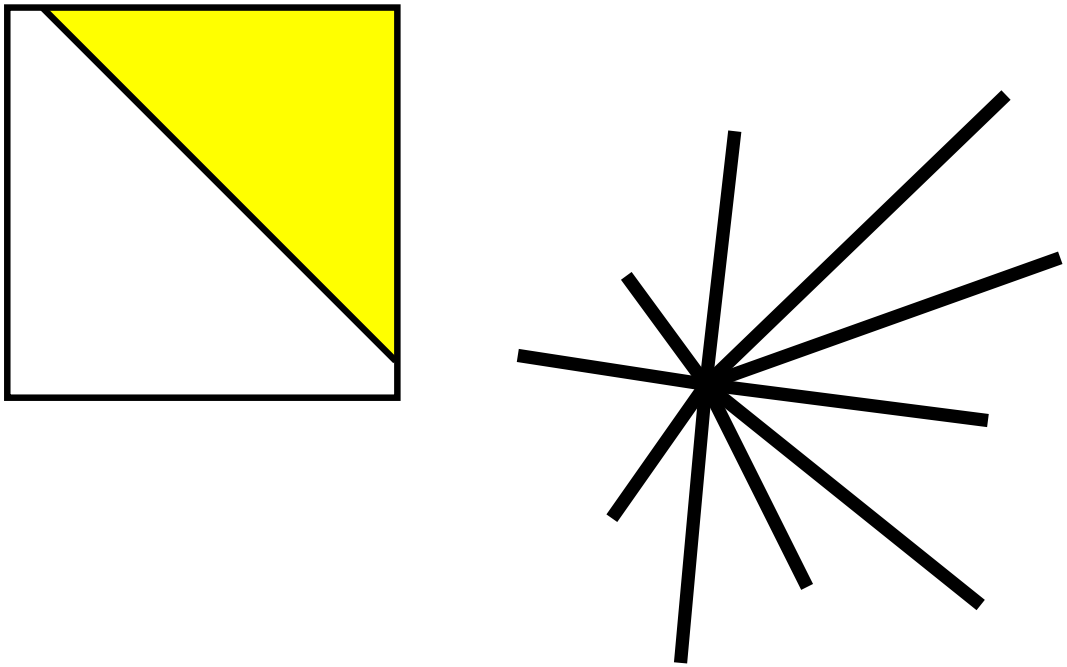
Progressive alignment following the guide tree



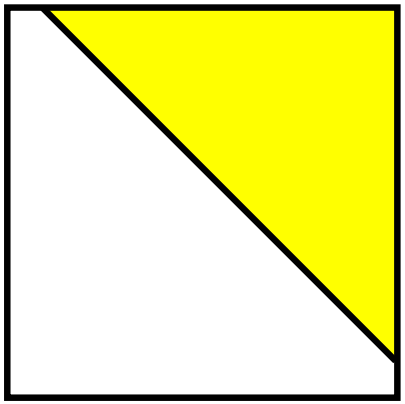
Measure distances directly from the alignment

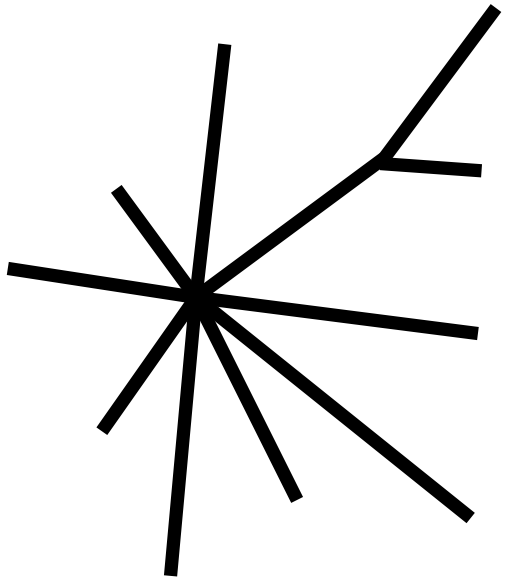


Generate neighbor-joining tree from new distances

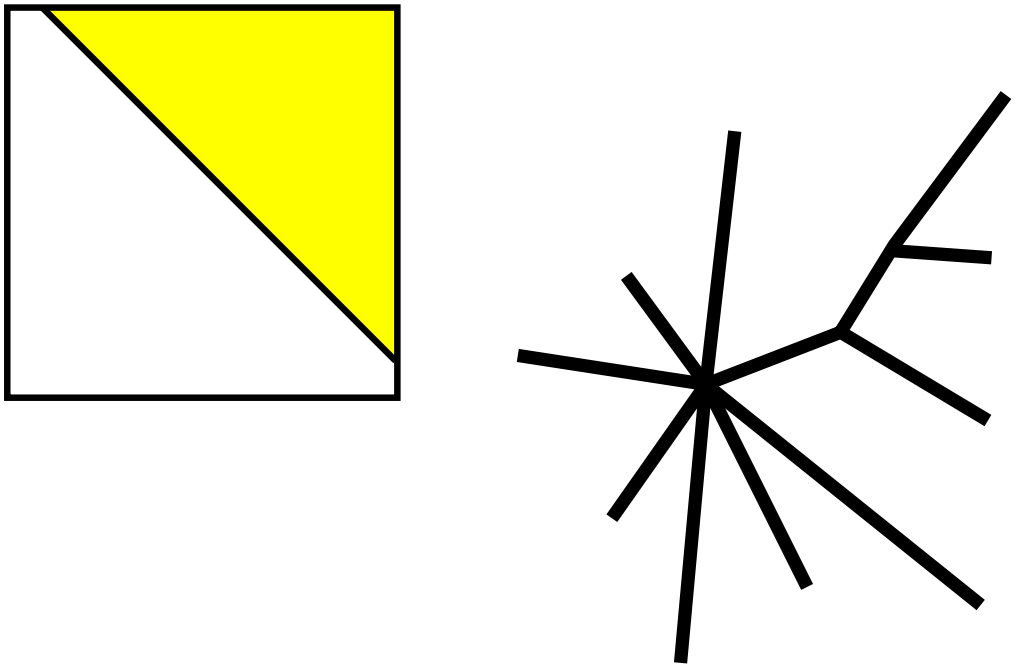


Generate neighbor-joining tree from new distances

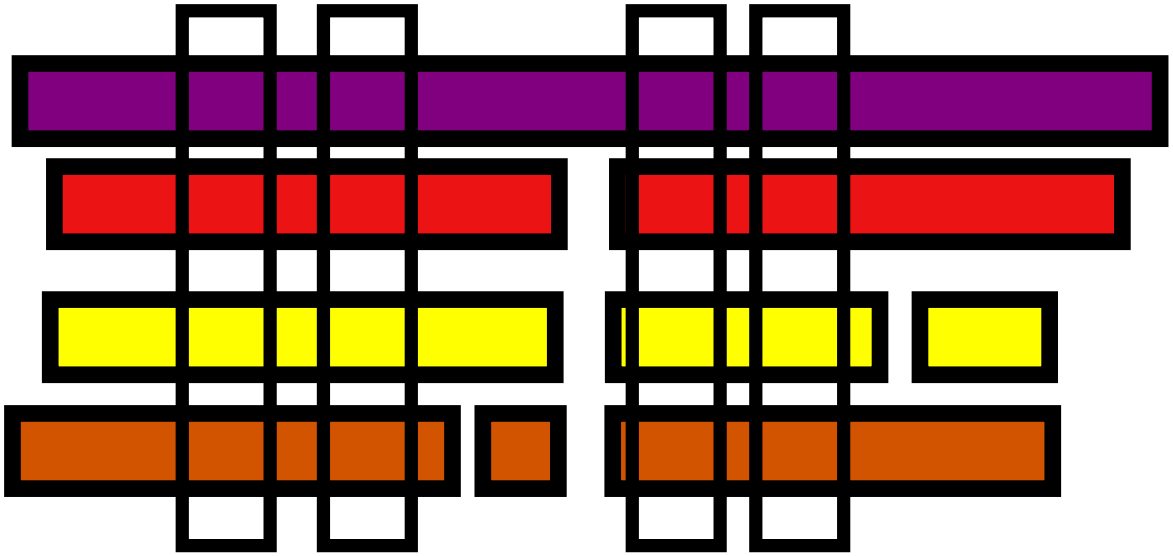




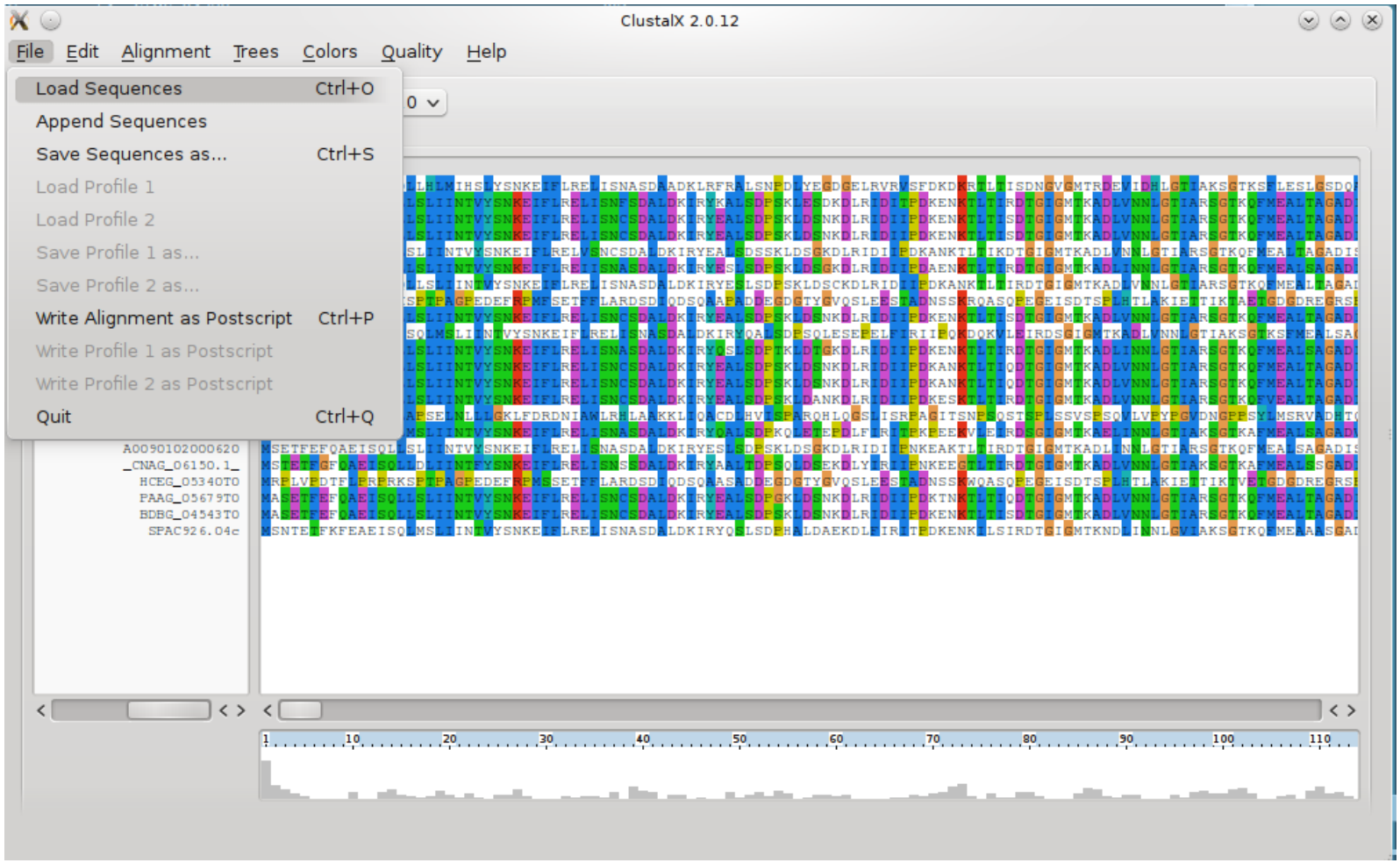
Generate neighbor-joining tree from new distances



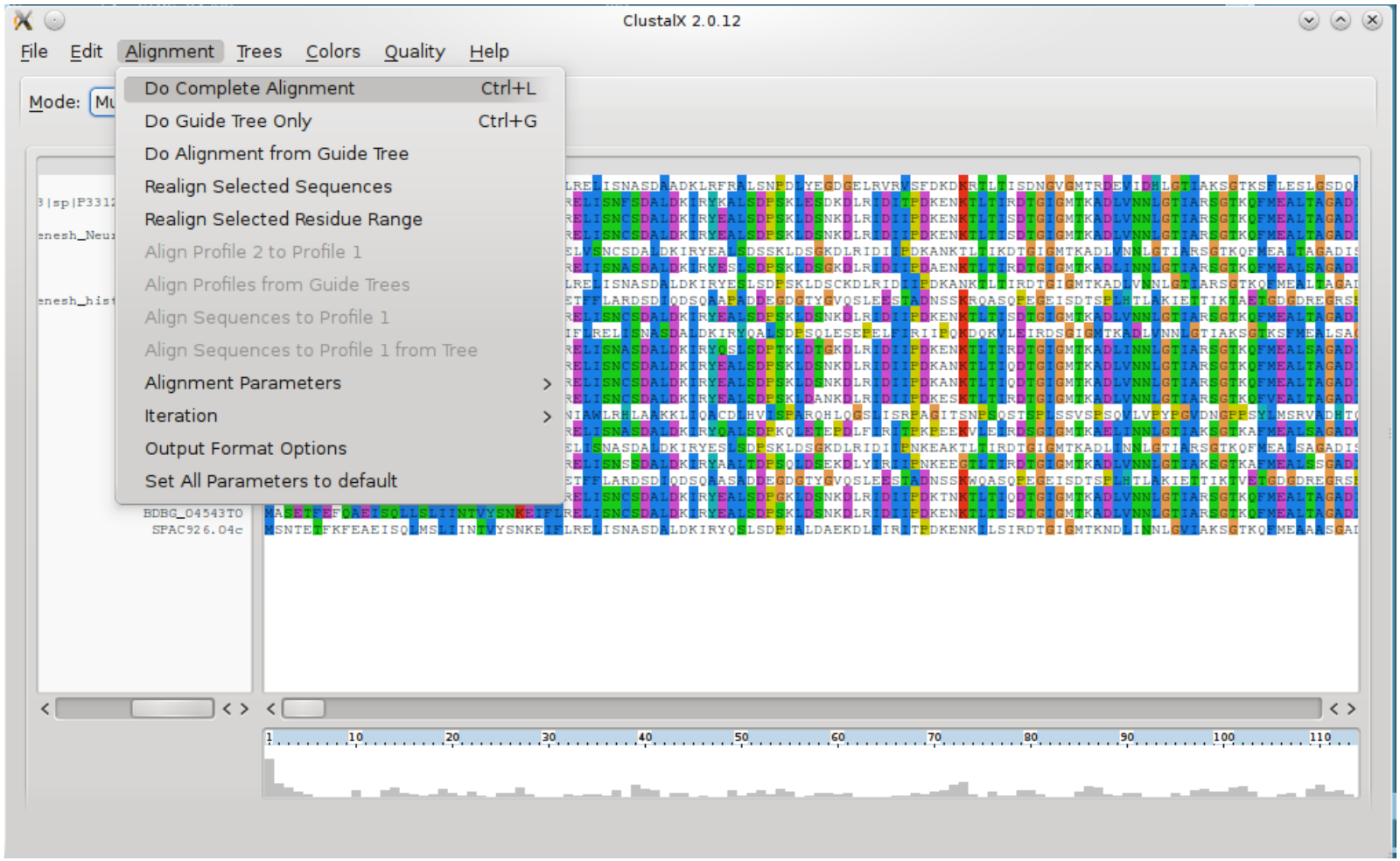
Generate bootstrap values from subsets of the alignment



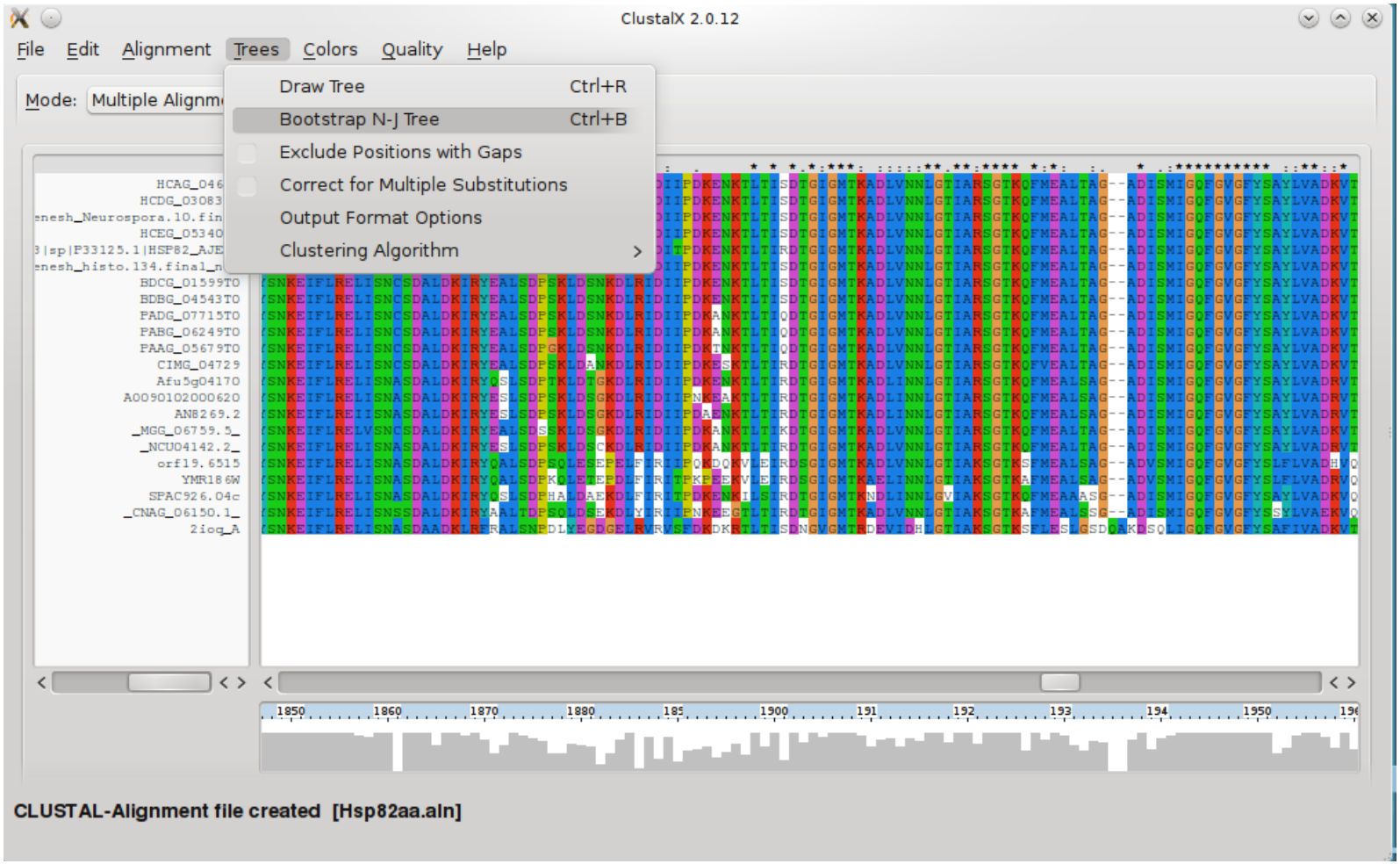
Generating a multiple alignment in CLUSTALX



Generating a multiple alignment in CLUSTALX



Generating a neighbor joining tree in CLUSTALX



Viewing the alignment and tree in JALVIEW

Related tools

- Protein Multiple Alignment
  - MUSCLE
  - Clustal Omega
  - Probcons
  - hmmalign (HMMer3)
- Tree Building
  - MrBayes (Bayesian MCMC)
  - PhyML (maximum likelihood)
  - RaxML (fast maximum likelihood)
  - FastTree2 (very large heuristic trees)

Homework

Finish your dynamic programming implementation.