Logistics

- Problem set 3 due Fri, Sep 27th
- 7Eleven-1 due Fri, Sep 27th

Solution sets posted Saturday

No late assignments will be accepted once solution sets are posted

- In-class Exam, October 1st
  - Covers Sequence alignment,
    Models of sequence substitution
  - Letures through Sept. 19
  - Closed book
  - Two pages of notes

Pairwise sequence alignment
(global, semiglobal and local)

Multiple sequence alignment

global   Local
(PSSMs, HMMs)

Markov Chains and Sequence evolution models

Database searching

BLAST

Substitution matrices

Sequence statistics

Evolutionary tree reconstruction

Two widely used families of Amino Acid Substitution Matrices Parameterized for evolutionary divergence (*N*)

- PAM matrices, Dayhoff *et al*, 1978
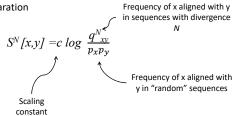
- BLOSUM (Block Sum) matrices, Hennikoff &Hennikoff, 1991

Amino Acid Substitution Matrices
Parameterized for evolutionary divergence (*N*)

Overall strategy for both PAM and BLOSUM

1. Trusted amino acid alignments
2. Obtain amino acid pair counts $\left(A_{xy}^N\right)$ with corrections for
   - Evolutionary divergence
   - Sample biases
3. Estimate substitution frequencies, q_{xy}^N, from pair counts, A_{xy}^N
4. Log odds substitution matrix: $S^N[x,y] = c\ log\ \dfrac{q_{xy}^N}{p_x p_y}$

Log odds substitution matrices

Two sequences have *N* PAMs divergence , if, on average, *N* amino acid replacements per 100 residues occurred since their separation

$$S^N[x,y] = c \; log \; \frac{q^N_{xy}}{p_x p_y}$$

Frequency of x aligned with y in sequences with divergence *N*

Frequency of x aligned with y in "random" sequences

Scaling constant

Two widely used families of Amino Acid Substitution Matrices Parameterized for evolutionary divergence (*N*)

➢ PAM matrices, Dayhoff *et al*, 1978

• BLOSUM (Block Sum) matrices, Hennikoff &Hennikoff, 1991

PAM Matrices

Atlas of Protein Sequence & Structure
1965 - 1978

Examined 1572 changes in 71 groups of closely related proteins

Margaret Dayhoff
PhD in Chemistry, 47
Watson Computing Lab
Fellow 47 - 48

Evolutionary divergence (amino acids)

• PAM: Percent Accepted Mutation
  • *Accepted Mutations* are mutations that are retained and passed on to future generations

• We say the divergence between two sequences is *N* PAMs, if, on average, *N* amino acid replacements per 100 residues (including multiple substitutions) occurred since their separation.