

BMI/CS 576 – Day 11

- Today
 - Substitution matrices, statistical significance of alignments, BLAST
- Thursday
 - Multiple alignment

Tentative HW schedule

- Now available in “Course Overview” module on Canvas

UW-Madison Multi-factor authentication

- Use secondary device (e.g., phone) to prove identity when logging in with NetID
- Required for students after Oct 31
- Enroll as soon as possible to avoid problems
- Won't be able to log in to any wisc services (including Canvas) without it!
- <https://it.wisc.edu/learn/guides/getting-started-multi-factor-authentication-students/>

HW2

- posted last week
- due Oct 23
- There was a slight typo in recurrences for H_x and H_y
 - Thanks to the student who pointed this out
 - HW2 template has been fixed
 - Correct recurrences are:

$$H_x(i, j) = \delta(x[i] \neq x[i - 1]) + \max \begin{cases} M(i - 1, j) + h + t \\ H_x(i - 1, j) + t \end{cases}$$

$$H_y(i, j) = \delta(y[j] \neq y[j - 1]) + \max \begin{cases} M(i, j - 1) + h + t \\ H_y(i, j - 1) + t \end{cases}$$

Quiz

With the definition of a substitution score between characters a and b given by

$$s(a, b) = \log \left(\frac{p_{ab}}{q_a q_b} \right)$$

then if $a \neq b$ (i.e., a mismatch), the value of $s(a, b)$ will be

Positive, negative, or zero! (any of the above in the quiz)

Depends on how p_{ab} compares to $q_a \times q_b$

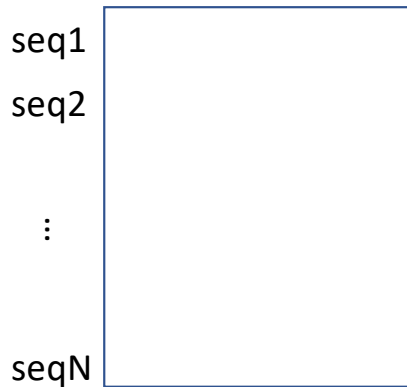
BLOSUM62

[illegible]

Constructing the BLOSUM matrices

Block

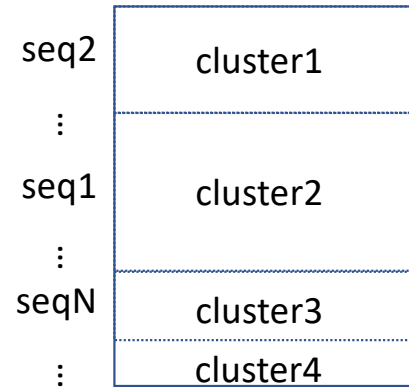
(multiple alignment of structurally conserved region of homologous proteins)



PARTITION

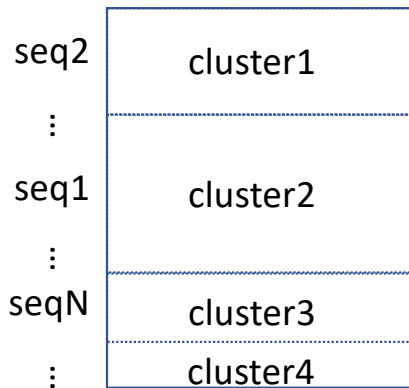
For a BLOSUMX matrix: two sequences must be in the same cluster if they are at least X% identical

-> implies that a pair of sequences that are from different clusters are < X% identical



Constructing the BLOSUM matrices

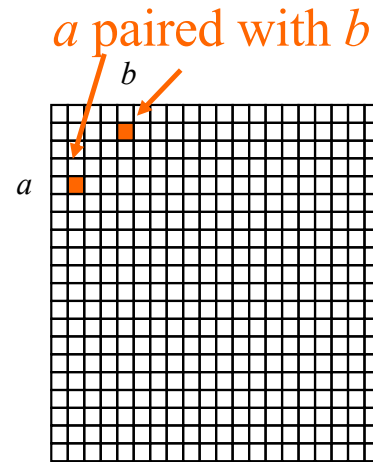
Partitioned sequences



count aligned pairs of
characters only from
pairs of sequences in
different clusters



Character pair counts (A_{ab})



BLOSUMX: The lower the value of X, the more divergence between the sequences being compared