Sequence Comparison

1. A method to compare individual letters → BLOSUM

2. Comparing the sequence of letters.

2.a. Dotplot method

\[ W_i : \begin{array}{cccc} A & T & G & C \\ G & A & C & G \end{array} \]

\[ W_j : \begin{array}{cccc} A & T & G & C \\ G & A & C & G \end{array} \]

\[ S(W_i, W_j) \]

\[ S = S_1 + S_2 + S_3 + S_4 + S_5 \]
Fill up the whole table with numbers.

Threshold

7

S1

S1

\[
\begin{array}{c}
1.15 \\
.21 \\
6.7 \\
\end{array}
\]
Quantitative sequence comparison:
(a) Alignment between the sequences
(b) A score representing that alignment

How to do that?

"This class is boring" 19 letters with

"This class is really boring" 26 letters assuming no possible insertions or deletions will lead to these 2 sentences to be different.

We need to account for possible "gaps"

This class is ( ) boring

This class is (really) boring
$S_1$: $N$ letters

$S_2$: $M$ letters

Number of match: $1 \leq k \leq \min(N, M)$

For a given $k$:

$$\sum_{k=1}^{\min(N,M)} \binom{N}{k} \binom{M}{k}$$

too large to be tested exhaustively.

"Simple" CS algorithm:

1. Dynamic programming:

A1: scores for comparing individual letters are additive.

A2: Score of the optimal alignment that ends at a given pair of positions in the sequences is the score of the best alignment previous to these positions plus the score of aligning their point