Sum-of-Pairs

Motivation

- Multiple sequence alignment

- Solves problem with score function S_c.

Problem with S_c

- In PAM: $S_c(a,b) = \dfrac{Pr[alignmtents\ are\ related]}{Pr[unrelated\ random\ sequences]} = \prod \dfrac{p_{a_i,b_i}}{q_{a_i} q_{bi}}$

- Use instead of multiplication the sum over logs

- What if $Sc(a,\ldots,n)$?

- Not enough data available to compute probabilities

Sum-of-Pairs

- align every sequence pairwise

- sum over the scores

- Example:

$$S_C(x, y, z) = S_C(x, y) + S_C(x, z) + S_C(y, z)$$

Problem

- Scoring can be incorrect: $S_C(x, y, z) = log(\frac{p_{xyz}}{q_x q_y q_z})$

- but:

$$S_C(x, y, z) = S_C(x, y) + S_C(x, z) + S_C(y, z)$$

$$log(\frac{p_{xyz}}{q_x q_y q_z}) \neq log(\frac{p_{xy}}{q_x q_y}) + log(\frac{p_{xz}}{q_x q_z}) + log(\frac{p_{yz}}{q_y q_z})$$

List of references

- Lecture "Multiples Sequence Alignment - Basics" Bioinformatics I, Prof. Backofen

  URL: http://www.bioinf.uni-freiburg.de//Lehre/Courses/2014_SS/V_Bioinformatik_1/multiple-alignment.pdf, visited: 13/11/2014