

Log Odds Matrix

ASE Übung

03.05.2019

BLOCKS Datenbank

- Multiple Alignments stark konservierter Regionen in Proteinfamilien
- Keine Lücken
- Klassifizierung von Proteinen
- Blocks Substitution Matrix (BLOSUM)

Log Odds Matrix

$$M_{ij} = \left[2 \times \log_2 \frac{q_{ij}}{e_{ij}} \right]$$

Erwartete Häufigkeit von Substitutionen

$$e_{ij} = \begin{cases} 2p_i p_j, & i \neq j \\ p_i^2, & i = j \end{cases}$$

Relative Häufigkeiten von Aminosäuren

$$p_A = \frac{3}{5}$$

$$p_B = \frac{2}{5}$$

A

A

A

B

B

Erwartete Häufigkeit von Substitutionen (cont.)

$$e_{AA} = \frac{3}{5} \times \frac{3}{5}$$

A

A

$$e_{BB} = \frac{2}{5} \times \frac{2}{5}$$

A

B

$$e_{AB} = 2 \times \frac{3}{5} \times \frac{2}{5}$$

B

Beobachtete Häufigkeit von Substitutionen

$$\begin{array}{rcl} q_{AA} & = & \frac{3}{10} & \text{A} \\ & & & \text{A} \\ q_{BB} & = & \frac{1}{10} & \text{A} \\ & & & \text{B} \\ q_{AB} & = & \frac{6}{10} & \text{B} \end{array}$$

$$\text{Anzahl von Paaren: } k \frac{n(n-1)}{2}$$

Halb-Matrix

E_{ij}	A	I	L	W
A	?	-	-	-
I	?	?	-	-
L	?	?	?	-
W	?	?	?	?

Logarithmen

$$\log_b x = \frac{\log_a x}{\log_a b}$$

Block 1: AAIA
AALA
AAIL
LALA
AAIW

Block 2: IWIA
IWLA
LWAA
LWIA
AWLA

Berechnung der Matrix

1. Relative Häufigkeit der Aminosäuren p_i berechnen
2. Erwartete relative Häufigkeit der Aminosäuren-Substitutionen berechnen: $e_{ij} = p_i^2$ bzw. $e_{ij} = 2p_i p_j$
3. Aminosäuren-Substitutionen zählen
4. Aminosäuren-Substitutionen normalisieren: q_{ij}
5. Log-Odds Matrix-Elemente berechnen: $\left[(2 \times \log_2 \frac{q_{ij}}{e_{ij}}) \right]$

Frage 1

Wie kommt ein $e_{ij} = 0$ bzw. $q_{ij} = 0$ Wert zustande?

Welches Problem tritt dann bei der Berechnung der Matrix auf?

Wie kann man dieses Problem vermeiden?

Frage 2

Nennen Sie Anforderungen an einen für die Erstellung einer Substitutionsmatrix nach vorgestelltem Verfahren verwendeten Datensatz.

Wie viele Proteine sollten verwendet werden?

Wie homolog sollten sie sein?

Sollten lokale oder globale Alignments verwendet werden?

Wie sollten die Alignments aussehen?

Frage 3

Wie würden Substitutions-Matrizen aussehen, welche für Alignments unterschiedlich verwandter Proteinpaare verwendet werden können?

Nennen Sie einen Ansatz diese verschiedenen Matrizen zu berechnen.