

Übungsblatt 6

Sequenzmotive

Deadline: 15.6.2022 um 20:00 MEZ

Bioinformatik für Biochemiestudierende

Dr. Florian Klimm

Sommersemester 2022

Aufgabe 6.1 Laplace Succession (10%)

Die letzten sieben Vorlesungen haben stattgefunden. Nutzen Sie die Laplace Succession um die Wahrscheinlichkeit, dass die nächste Vorlesung auch stattfindet, zu bestimmen.

Laplace Succession

$$Pr = \frac{(s+1)}{(n+2)} \quad \Rightarrow n \hat{=} \text{Beobachtungen}$$

$$\Rightarrow s \hat{=} \text{Erfolge}$$

$$n = 7 \quad ; \quad s = 7$$

$$Pr = \frac{(7+1)}{(7+2)} = \frac{8}{9} = 0,8 = 80\% \quad \Rightarrow \text{mit } 80\% \text{ Wahrscheinlichkeit findet die nächste Vorlesung auch statt}$$

Aufgabe 6.2 Sequenzmotive (20%)

Sie haben die folgenden Motive gefunden

$$\begin{array}{ccccc} G & A & T & A & C \\ G & A & T & C & A \\ A & T & T & G & A \\ G & C & T & C & G \end{array} \quad (1)$$

Berechnen Sie das Profil dieser Motive. Berechnen Sie auch das Profil der Pseudo-counts und vergleichen Sie die Ergebnisse.

Motiv =>

A:	1	2	0	1	2
C:	0	1	0	2	1
G:	3	0	0	1	1
T:	0	1	4	0	0

Profil =>

A:	$\frac{1}{4}$	$\frac{2}{4}$	0	$\frac{1}{4}$	$\frac{2}{4}$
C:	0	$\frac{1}{4}$	0	$\frac{2}{4}$	$\frac{1}{4}$
G:	$\frac{3}{4}$	0	0	$\frac{1}{4}$	$\frac{1}{4}$
T:	0	$\frac{1}{4}$	$\frac{4}{4}$	0	0

A: 1+1 2+1 0+1 1+1 2+1
 C: 0+1 1+1 0+1 2+1 1+1
 G: 3+1 0+1 0+1 1+1 1+1
 T: 0+1 1+1 4+1 0+1 0+1

=> Pseudocount-Matrix:

A: 2 3 1 2 3
 C: 1 2 1 3 2
 G: 4 1 1 2 2
 T: 1 2 5 1 1

=> Pseudocount-Profil:

A: $\frac{2}{8}$ $\frac{3}{8}$ $\frac{1}{8}$ $\frac{2}{8}$ $\frac{3}{8}$
 C: $\frac{1}{8}$ $\frac{2}{8}$ $\frac{1}{8}$ $\frac{3}{8}$ $\frac{2}{8}$
 G: $\frac{4}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{2}{8}$ $\frac{2}{8}$
 T: $\frac{1}{8}$ $\frac{2}{8}$ $\frac{5}{8}$ $\frac{1}{8}$ $\frac{1}{8}$

Aufgabe 6.3 Profile (20%)

Sie haben das folgende Profil gefunden

A	1/2	1/4	0	1
C	1/2	1/4	0	0
T	0	1/4	1/2	0
G	0	1/4	1/2	0

Bestimmen Sie die Wahrscheinlichkeit der folgenden Strings

1. ACTA => $\frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot 1 = \frac{1}{16} = 0,0625$

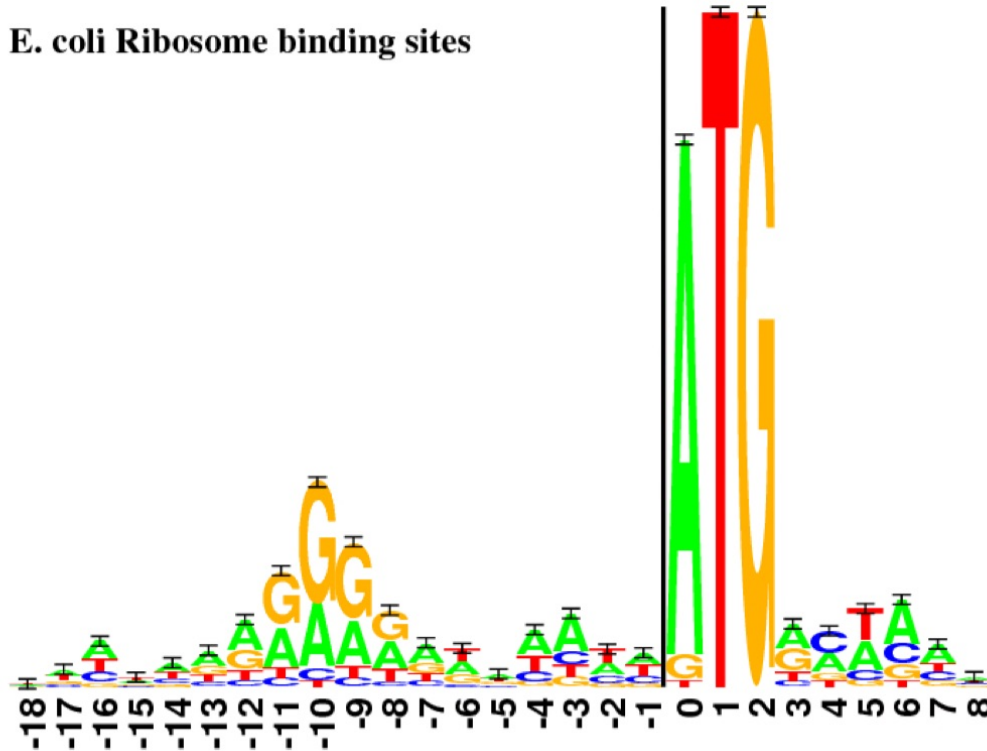
2. ACTG => $\frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot 0 = 0$

3. CCTAA => $\frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot 1 \cdot \frac{1}{4} = \frac{1}{64} = 0,0156$

Aufgabe 6.4 Sequence Logo (20 %)

Interpretieren Sie das folgende Sequence Logo.

E. coli Ribosome binding sites



- > Startcodon ATG an Position 0, 1, 2
- > Wellenförmige Peaks repräsentieren DNA-Windungen
 - => z.B. bei -11, -10, -9 & 0, 1, 2
 - => Windungen der DNA alle 10,4 Basenpaare
- > Größe d. Buchstaben repräsentiert Häufigkeit, in der sie an bestimmter Stelle vorkommen
 - => z.B. an Stelle -11, -10, -9 sind G & A am häufigsten