

Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2018

Dr. Daniel Dörr

<https://gi.cebitec.uni-bielefeld.de/teaching/2018summer/sa>

Präsenzübungsblatt 1, Woche 16/2018

Aufgabe 1 (Sequenzwahrscheinlichkeiten)

1. Wie wahrscheinlich ist es, die DNA-Sequenz $s = \text{TATAAA}$ per Zufall zu erhalten? Nimm dabei an, dass jede Base mit derselben Wahrscheinlichkeit auftritt. Gib auch die Formel an, die du zur Berechnung benutzt hast.
2. Wieso wird diese Sequenz nicht nur aufgrund von Zufall in einem Eukaryoten-Genom auftauchen?
3. Gegeben sei nun eine DNA-Sequenz der Länge 7. Wie wahrscheinlich ist es, dass kein A darin vorkommt? Gib auch hierzu eine Formel an.
4. Wie wahrscheinlich ist es, dass in einer Sequenz der Länge 7 mindestens zwei Ts vorkommen?

Aufgabe 2 (GC-Gehalt)

Das Genom von *Streptomyces coelicolor* besitzt einen GC-Gehalt von etwa 70% und eine Genomgröße von ca. 9 Mb.

1. Wie oft würde man hier die Schnittstelle des Restriktionsenzym EcoRI mit der Erkennungssequenz GAATTC erwarten?
2. Wie hoch ist die Wahrscheinlichkeit, dass ein zufälliger Abschnitt der Länge 6, der zwei As an dieser Position, aber keiner anderen enthält, auftritt? (Also: BAABBB, mit B nicht A)
3. Wie viele Möglichkeiten gibt es in einem zufälligen Abschnitt der Länge 6 zwei As zu platzieren?
4. Wie hoch ist die Wahrscheinlichkeit genau zwei As in einem zufälligen Abschnitt der Länge 6 zu erhalten?

Aufgabe 3 (Anzahl von Teilsequenzen und Teilworten mit Fehlern)

Bei der Übertragung einer Sequenz der Länge n über einen rauschenden Kanal hat sich ein Fehler eingeschlichen: ein Zeichen wurde durch ein anderes ersetzt. Wieviele Teilsequenzen und wieviele Teilworte der Länge k hat die ursprüngliche mit der übertragenen Sequenz *mindestens* gemeinsam?

Aufgabe 4 (O-Notation)

Sortiere die O-Klassen der Funktionen $a(n)$ bis $e(n)$ anhand ihrer Untermengenrelation \subset , z.B. $O(g(n)) \subset O(h(n))$.

- $a(n) = n^2 \cdot \log_2(n) + 42$
- $b(n) = 2^n + n^4$
- $c(n) = 2^{2n}$
- $d(n) = 2^{n+3}$
- $e(n) = \sqrt{n^5}$