

Präsenzübungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WS 2011/2012

Dr. Alexander Sczyrba · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011winter/SequenzAnalyse>

Präsenzübungsblatt 1, Woche 45/2011

Aufgabe 1 (Sequenzwahrscheinlichkeiten)

1. Wie wahrscheinlich ist es, die DNA-Sequenz $s = TATAAA$ per Zufall zu erhalten? Nimm dabei an, dass jede Base mit der selben 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 weitere DNA-Sequenz s_2 der Länge 7. Wie wahrscheinlich ist es, dass genau 3 'As' darin vor kommen? Gebe auch hierzu eine Formel an.
4. Wie wahrscheinlich ist es, dass in der gleichen Sequenz s_2 mindestens ein 'G' vorkommt?

Aufgabe 2 (Substrings und Subsequenzen)

Gegeben sei das Wort $s = \text{KATZENKLO}$.

1. Gib alle Substrings der Länge 5 und eine Subsequenz der Länge 5 von s an.
2. Gib alle Präfixe und Suffixe von s an.
3. Was bedeutet es, wenn eine Menge S von Sequenzen *prefix-closed* ist? Gib ein Beispiel einer Menge an, die *suffix-closed* ist.
4. Was ist ein Superstring?