

# Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2017

Prof. Dr. Jens Stoye · M.Sc. Tizian Schulz

<https://gi.cebitec.uni-bielefeld.de/teaching/2017summer/sa>

## Präsenzübungsblatt 1, Woche 17/2017

### 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 (Metriken)

1. Drei Eigenschaften müssen erfüllt sein, damit die binäre Operation  $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  auf der Menge  $\mathcal{X}$  eine Metrik ist:

$$d(x, y) = 0 \leftrightarrow x = y \quad (1)$$

$$d(x, y) = d(y, x), \forall x, y \in \mathcal{X} \quad (2)$$

$$d(x, y) \leq d(x, z) + d(z, y), \forall x, y, z \in \mathcal{X} \quad (3)$$

Beweise, dass aus diesen Definitionen die Nichtnegativität folgt:

$$d(x, y) \geq 0, \forall x, y \in \mathcal{X}$$

2. Erläutere, wieso die folgende Metrik auf den Elementen der Potenzmenge  $\mathcal{P}(Z)$  einer Menge  $Z$  auch *Mengentransformationsdistanz* genannt wird:

$$d_\infty(X, Y) = \max\{|X \setminus Y|, |Y \setminus X|\}$$