

## Gedächtnisprotokoll Klausur Bioinformatik vom 27.03.2017 (bei Dr. Hartmann)

- 1) Bewerten Sie die folgenden Aussagen entweder als Wahr (W) oder Falsch (F)
  - a. In der GenBank finden sich sowohl ausführlich nach der jeweiligen Funktion des Genes beschriebene Einträge als auch Einträge mit nur sehr knapper funktionaler Beschreibung.
  - b. Das Verfahren des Homology Modellings beruht auf dem Vorhandensein der query-Sequenz homologer Sequenzen, von denen die 3D-Struktur bekannt ist.
  - c. Eine phylogenetische Analyse via Maximum Parsimony liefert einen perfekten phylogenetischen Baum.
  - d. U.v.m.
- 2) a) Schreiben Sie eine mögliche Proteinsequenz auf, die durch den folgenden Prosite regex-Syntax erkannt werden würde.

C-X(1,2)-C-[LISF]-X(1,4)-H-X(2)-[FLISE]-H-X(1,2)-E-Q

- b) Nennen Sie ein Problem, das bei der Identifikation von Domänen aus Proteinsequenzen mittels regular expressions auftreten kann.
- 3) Definieren Sie folgende Begriffe
  - a. Transkriptom
  - b. Hub
  - c. SNP
- 4) Nennen Sie ein Beispiel dafür, wo in der Biologie Graphen eine Rolle spielen. Was wird in Ihrem Beispiel durch die Knoten, was durch die Kanten dargestellt? Sagen und begründen Sie, ob der Graph (un)gerichtet bzw. (un)gewichtet ist.
- 5) Nennen Sie für folgende bioinformatische Verfahren jeweils den input und den output.
  - a. Re-Sequenzierung (mapping)
  - b. Phylogenetische Analyse mittels der Neighbor Joining Methode
  - c. OTU-Bestimmung
- 6) Welche Annahmen liegen der 3D-Strukturbestimmung via Threading zugrunde? Erläutern sie kurz das Prinzip des Threadings.
- 7) Es ist ein Auszug aus einem BLAST-Ergebnis gegeben (Mit E-value, raw score, bit score und gaps). Der Auszug zeigt den besten Treffer in der entsprechenden Datenbank.
  - a. Erläutern Sie in Worten, was der E-value aussagt.
  - b. Interpretieren Sie obiges Ergebnis
  - c. 1 Jahr später wird der gleiche BLAST noch einmal durchgeführt. Die gleiche Sequenz wie zuvor ist wieder der beste Treffer. Allerdings hat sich die Gesamtmenge der Sequenzen in der Datenbank erheblich erhöht. Wie verändern sich jeweils der raw score und der E-value?

8) Im Folgenden ist ein R-Skript (ohne Ausgaben und Grafiken) gegeben. Erläutern Sie in Stichworten, was durch die Befehle erreicht wird.

1) `> x<-read.table("genomics.txt")`

2) `> mean(x[,2])`

3) `> par(mfrow=c(1,2))`

4) `> plot(x[,2],x[,3])`

5) `> boxplot(x[,2])`

6) `> par(mfrow=c(1,1))`