

Prof. Dr. Knut Reinert
Dr. Roland Krause
Matthias Winkelmann
Patrick Pett

Institut für Informatik
AG Algorithmische Bioinformatik

Algorithmische Bioinformatik WS 11/12

Übungsblatt 4

Abgabe am 14.11.2011, 12 Uhr

Aufgabe 1: Expectation Maximization

In der Vorlesung und im Script werden erwähnt, dass die relative Entropie immer nicht-negativ ist. Beweisen Sie dies.

Aufgabe 2: Motif Finding

Gegeben sind die folgenden Sequenzabschnitte:

	1	2	3	4	5	6
x_1	A	T	A	T	G	G
x_2	C	T	G	G	A	G
x_3	T	G	T	G	G	C

Und eine initiale Wahrscheinlichkeitsverteilung:

	0	1	2	3
A	0.25	0.1	0.3	0.2
C	0.25	0.1	0.2	0.2
G	0.25	0.3	0.3	0.3
T	0.25	0.5	0.2	0.3

Führen Sie zwei Schritte des in der Vorlesung besprochenen MEME-Modells zur Motivsuche durch. Sie können dabei davon ausgehen, dass das Motiv in den Sequenzen jeweils nur einmal vorkommt. Beschreiben Sie kurz Ihre Beobachtungen.

Aufgabe 3: Gene Finding

In der Vorlesung wurden einige Algorithmen zur Genvorhersage besprochen. Sie sollen nun zeigen, dass es auch einfacher geht.

- a) Berechnen Sie hierzu die Verteilung der Länge von *Open Reading Frames* in *E. coli*¹. und stellen Sie diese dar.
- b) Bestimmen Sie anhand der Ergebnisse ein geeignetes Kriterium für die Klassifizierung von Nukleotiden und wenden Sie dieses auf das komplette Genom an.
- c) Vergleichen Sie Ihre Ergebnisse mit den Gendaten im Genbank-File und bewerten Sie Ihre Vorhersage.

¹ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia_coli_K_12_substr_MG1655_uid57779/NC_000913.gbk