

Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2004/05

Utz J. Pape · Ben Rich · Dr. Stefan Röpcke · Prof. Dr. Martin Vingron

Blatt 8 · Ausgabe am 6.12.2004

Abgabe am 13.12.2004 vor Beginn der Vorlesung

Aufgabe 28 (Genvorhersage). Machen Sie sich etwas mit den Genvorhersageprogrammen Genscan und GeneMark vertraut.

<http://opal.biology.gatech.edu/GeneMark>

<http://genes.mit.edu/GENSCAN.html> Nutzen Sie dazu folgende humane Sequenz, in der das Insulingen kodiert ist:

<http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=23986710> und die *E. coli* Sequenz U00096.2.gbk von 2825000 bis 2829000 (siehe Programmieraufgabe). Interpretieren Sie die Outputs.

Aufgabe 29 (Programmieraufgabe zur Genvorhersage). Für die Genvorhersage in prokaryontischen Genomen wird vorrangig das Ungleichgewicht der Basenverteilungen in proteinkodierenden Sequenzen ausgenutzt. Zum einen sind die einzelnen Aminosäuren nicht gleichverteilt und zum zweiten sind in einer Spezies bestimmte Kodons bevorzugt anzutreffen (Codon Usage). Ihre Aufgabe besteht darin, das in der Vorlesung vorgestellte Verfahren zur Berechnung des Kodierungspotentials zu implementieren. Als Datenbasis verwenden Sie bitte *Escherichia coli* K-12 Stamm MG1655. Sie finden die komplette annotierte Sequenz im Genbank-Format unter <http://www.genome.wisc.edu/pub/sequence/U00096.2.gbk> oder über die Veranstaltungsseite.

1. Vergleichen Sie die von Ihnen gefundenen Werte mit publizierten Daten, zu finden z. B. unter <http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=Escherichia+coli+%5Bgbct%5D>
2. Wählen Sie exemplarisch eine Region (so, dass man etwas interessantes sieht) und stellen Sie das Kodierungspotential in den sechs Leserastern graphisch dar.
3. Symbolisieren Stopkodons und offene Leseraster.

Aufgabe 30 (*Evolution). Die meisten Aminosäuren lassen sich bekanntlich durch mehrere Triplets kodieren. Untersuchen Sie die Hypothese, dass die Wahl der Triplets mit dem GC-Gehalt der umgebenden Sequenz korrespondiert. Dahinter steckt die Annahme, dass der lokale GC-Gehalt funktionelle Relevanz hat und die frei mutierbaren Basen in der kodierenden Sequenz sich dem GC-Gehalt anpassen.

Aufgabe 31 (Testen, Markovketten). Ab Dienstag morgen finden Sie auf der Vorlesungsseite die Aufgabe mit den entsprechenden Daten.