

Übungen zur Vorlesung  
“Algorithmen der Bioinformatik II”  
Wintersemester 2005/2006

---

**Blatt 5**

1. Zu einer DNA-Eingabesequenz  $S$  sei eine Menge von Exonkandidaten gegeben. Jedem Exonkandidaten  $e = (a, b, r, f, s)$  sei seine Anfangsposition  $1 \leq a \leq |S|$  und seine Endposition  $1 \leq b \leq |S|$  mit  $b > a$  zugeordnet.  $r \in \{\rightarrow, \leftarrow\}$  gibt an, ob er auf dem Vorwärts- bzw. Rückwärtsstrang ist. Der Leserahmen  $f \in \{0, 1, 2\}$  des Exonkandidatens sei die Anzahl der Basen vor der in Leserichtung ersten Kodongrenze in  $e$ . Die Sorte  $s \in \{\text{single, initial, internal, terminal}\}$  gibt an, ob  $e$  das einzige (single) Exon des Gens ist, oder ob es in Leserichtung, das erste (initial), eines der inneren (internal) oder das letzte (terminal) Exon ist. Gib an, wie man die Menge von Typen  $T$ , und die Funktionen  $\text{vor}(e)$  und  $\text{nach}(e)$  in Abschnitt 1.4.3 definieren kann, so dass die konsistenten Ketten gerade die sind, die biologisch möglichen Exon-Abfolgen entsprechen.

**6 Punkte**

2. Wie muss man den Gene-Assembly-Algorithmus aus der vorhergehenden Aufgabe abändern, so dass nur vollständige Gene gefunden werden?

**4 Punkte**

3. Schreibe ein Programm, das versucht, den Leserahmen von kurzen Exons  $\sigma$  zu erraten. Benutze hierzu eine 3-periodische Markow-Kette zweiter Ordnung wie in Abschnitt 1.5.2. Zur Vereinfachung der Aufgabe, lasse die Berechnung der Wahrscheinlichkeit für die ersten  $k = 2$  Nukleotide der Sequenz  $\sigma$  weg.

Zum Schätzen der Parameter verwende die menschlichen Exonsequenzen mit bekanntem Leserahmen unter <http://www.gobics.de/mario/genomanalyse>. Versuche, mit deinem Programm den Leserahmen der 10 Testsequenzen auf dieser Seite zu erraten.

**10 Punkte**

Abgabe bis Dienstag, den 29. November, bei Mario Stanke. Quelltext von Programmen an [mstanke@gwdg.de](mailto:mstanke@gwdg.de). Lösungen werden am Dienstag, den 29. November, besprochen.