

Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2016/17

Martin Vingron · Annalisa Marsico · Alena van Bömmel · Edgar Steiger · Thimo Wellner

Blatt 9 · Ausgabe am 12.12.2016

Abgabe am 9.1.2017 vor Beginn der Vorlesung

Name:

Matrikelnummer:

Übungsgruppe:

Aufgabe 1 (20 Punkte; Theorie). Der Nussinov-Algorithmus ist einer der wichtigsten Algorithmen zur Berechnung von RNA-Sekundärstrukturen. Folgende Matrix ist beim Anwenden des Nussinov-Algorithmus auf eine gegebene RNA entstanden.

	A	C	U	C	U	A	A	A	U	U	U	A	G	A	U	G	U
A	0	0	1	1	1	2	2	2	3	4	4	5	6				
C	0	0	0	0	0	1	2	2	3	3	3	4	5				
U	0	0	0	0	0	1	2	2	3	3	3	4	5				
C	0	0	0	0	0	1	1	1	2	3	3	4	5				
U	0	0	0	0	0	1	1	1	2	3	3	4	4	4	5	5	5
A	0	0	0	0	0	0	0	0	1	2	3	3	3	3	4	4	5
A	0	0	0	0	0	0	0	0	1	2	2	3	3	3	4	4	4
A	0	0	0	0	0	0	0	0	1	1	1	2	2	3	3	3	3
U	0	0	0	0	0	0	0	0	0	0	0	1	1	2	2	2	2
U	0	0	0	0	0	0	0	0	0	0	0	1	1	2	2	2	2
U	0	0	0	0	0	0	0	0	0	0	0	1	1	1	2	2	2
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

- Vervollständigen Sie die Matrix mit dem Nussinov-Algorithmus. Geben Sie dabei die Rekursionsgleichung und die Scoring-Funktion an.
- Geben Sie den Traceback-Pfad durch die Matrix an und skizzieren Sie die erhaltene Sekundärstruktur.

Aufgabe 2 (35 Punkte; Praxis/Programmieren). In folgender Aufgabe sollen Sie den Nussinov-Algorithmus auf ein komplexeres Beispiel anwenden und Ihre Ergebnisse mit denen vom Zuker-Algorithmus vergleichen.

1. Schreiben Sie ein Programm, das eine FASTA-Datei mit einer RNA-Sequenz einliest, den Nussinov-Algorithmus auf diese Sequenz anwendet und mit dem Traceback die Menge der Basenpaare mit optimalem Score findet. Sie können auch das unten angegebene Python-Skript¹ verwenden.
2. Modifizieren Sie Ihr Skript so, dass folgende Nebenbedingungen erfüllt werden:
 - (a) Hairpin-Loops müssen mindestens drei ungepaarte Basen enthalten
 - (b) Die Basenpaare sollen unterschiedlich mit einer Scoring-Funktion $S(a, b)$ bewertet werden:

$$S(a, b) = S(b, a) = \begin{cases} 3 & \text{wenn } a = G \ \& \ b = C \\ 2 & \text{wenn } a = A \ \& \ b = U \end{cases}$$

3. Demonstrieren Sie die praktische Anwendung auf der Testsequenz der ribosomalen rRNA 16S². Behandeln Sie eventuelle Ts als Us. Welchen Score erhalten Sie mit dem Nussinov-Algorithmus?
4. Stellen Sie die erhaltene Struktur in regulärer Darstellung dar. Sie können z.B. das VARNA-Applet³ oder `forna webapp`⁴ verwenden. Wählen Sie eine Perspektive, in der man die Struktur gut erkennen kann.
5. Vergleichen Sie Ihr Ergebnis mit der Ausgabe des Zuker-Algorithmus. Verwenden Sie dazu die `RNAfold`-Implementierung vom Vienna-RNA-Package⁵.
6. Vergleichen und diskutieren Sie die Ergebnisse.



¹Nussinov: <http://forrestbao.blogspot.de/2007/11/python-implementation-of-nussinov.html>

²Material 1: <https://molgen.iedit.mpg.de/3715667/rna16s.txt>

³<http://varna.lri.fr/>

⁴<http://nibiru.tbi.univie.ac.at/forna/>

⁵<http://www.tbi.univie.ac.at/RNA/index.html>

Aufgabe 3 (20 Punkte; Theorie). RNA-Strukturvorhersage kann durch die Nutzung von schon bekannten RNA-Strukturen unterstützt werden. Dazu betrachtet man ein multiples Sequenzalignment von RNAs und versucht konservierte Interaktionen zwischen Positionen zu identifizieren. Gegeben sei nun das folgende multiple Sequenzalignment:

```

    UAAACUCUAAAUUUAGAUGUCC
    UAAACUCACGAUAGGGAUGUCC
    UAAACUCUACAUGUGGAUGUCC
    UAAACUCUCCAUGAGAUGUCC
  
```

1. Müssen Sie alle Paare von Positionen betrachten, um konservierte Interaktionen zu finden? Begründen Sie Ihre Antwort.
2. Berechnen Sie die paarweise Mutual Information der Positionen i und j im MSA wie folgt:

$$MI_{i,j} = \sum_{(x,y) \in \{A,C,G,T\}^2} f(x_i, y_j) \log_2 \frac{f(x_i, y_j)}{f(x_i) f(y_j)} \quad (1)$$

3. Stellen Sie die Mutual Information graphisch dar. Welche Positionen sind besonders interessant für die Strukturvorhersage?



⁶Christmas icons created by Hand-drawn goods