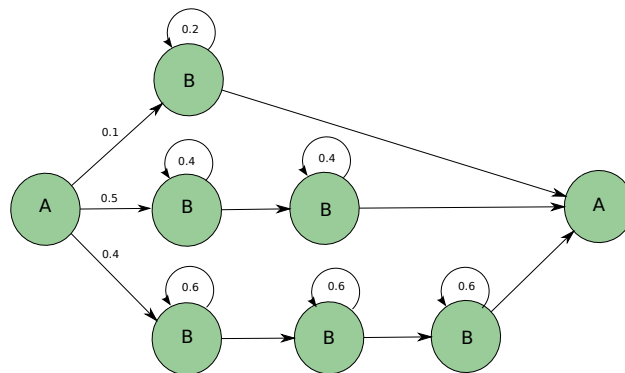


Algorithmische Bioinformatik Übungsblatt 5

Ausgabe: 19. Mai 2016 · Besprechung: 02. Juni 2016

Aufgabe 5.1 Gegeben ist folgender Ausschnitt aus einem HMM. Die Emissionen sind deterministisch (A oder B). Offensichtlich werden Sequenzen der Form AB^+A erzeugt. Berechne für alle sinnvollen Werte von ℓ die Wahrscheinlichkeit für das Ereignis, dass die Länge des erzeugten B-Laufs genau ℓ beträgt.



Aufgabe 5.2 Gegeben ist die Zählmatrix zur PWM des Bindemotivs von SOX-Transkriptionsfaktoren (Abbildung 1A), erstellt aus 31 experimentell verifizierten Bindestellen in Knorpelgewebe. Als Hintergrundverteilung werde die Gleichverteilung angenommen.

1. Berechne die Profilmatrix (ein Pseudocount an jeder Stelle) und daraus die PWM in Zehntelbits, ganzzahlig gerundet. Welcher Score ist maximal erreichbar?
2. Berechne die exakten Höhen der Positionen im Sequenzlogo aus der Profilmatrix *ohne* Pseudocounts.

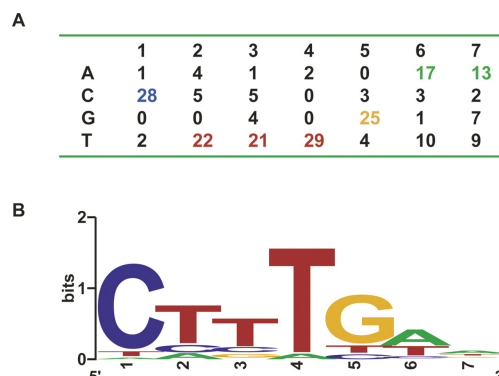


Abbildung 1: A: Zählmatrix; B: Sequenzlogo. [SR Davies et al.: Computational identification and functional validation of regulatory motifs in cartilage-expressed genes. Genome Research 17: 1438–1447 (2007)]