

## Algorithmen zur Sequenzanalyse

Wintersemester 2015/2016  
Besprechung am 23.01.2015

### Übungsblatt 6

Prof. Dr. E. Ohlebusch, Doktoranden  
Institut für Theoretische Informatik

---

#### Aufgabe 6.1.

Die Burrows-Wheeler-Transformationen der Strings  $S$  und  $S^{rev}$  sind jeweils in einem Wavelet-Baum gespeichert, siehe Abbildung 1 und 2. Ermitteln Sie mittels Algorithmus 29 im Skript, wie oft der String  $GTA$  in  $S$  vorkommt.

#### Aufgabe 6.2.

Erweitern Sie die Vorkommen (beginnend mit den gefundenen Vorkommen aus Aufgabe 1) jeweils mit einem Rückwärtssuchschritt um den Buchstaben  $c$  und versuchen Sie mit einem Vorwärtssuchschritt, das Muster auf der anderen Seite mit dem Watson-Crick-Komplement von  $c$  zu erweitern. Wie lautet das längste Muster, das Sie auf diese Weise finden können? Die Watson-Crick-Paarungen sind  $A-T$  und  $C-G$ .

#### Aufgabe 6.3.

Beweisen Sie, dass folgendes gilt:

$$\sum_{i=0}^k \binom{m}{i} (|\Sigma| - 1)^i \in O(m^k |\Sigma|^k)$$

#### Aufgabe 6.4.

Beweisen Sie formal, dass die links-rechts Zerlegung von  $P$  bzgl.  $S$  eindeutig ist.

#### Aufgabe 6.5.

Geben Sie einen Algorithmus an, der mit Hilfe der BWT von  $S^{rev}$  das  $M_{lr}$ -Array berechnet. Wie ist die Laufzeit Ihres Algorithmus? Wie kann das  $M_{lr}$ -Array effizient gespeichert werden?

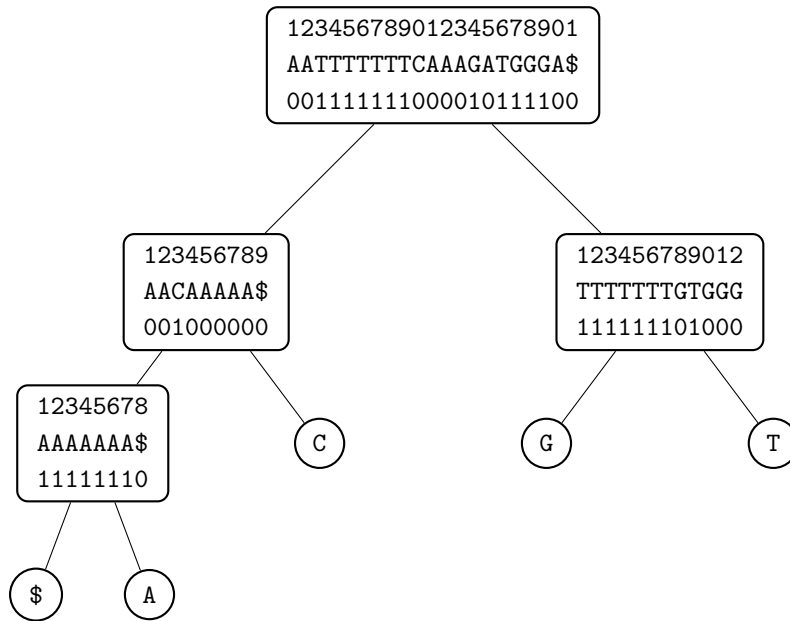


Abbildung 1: Wavelet-Baum von *AATTTTTTCAAAGATGGGA\$*

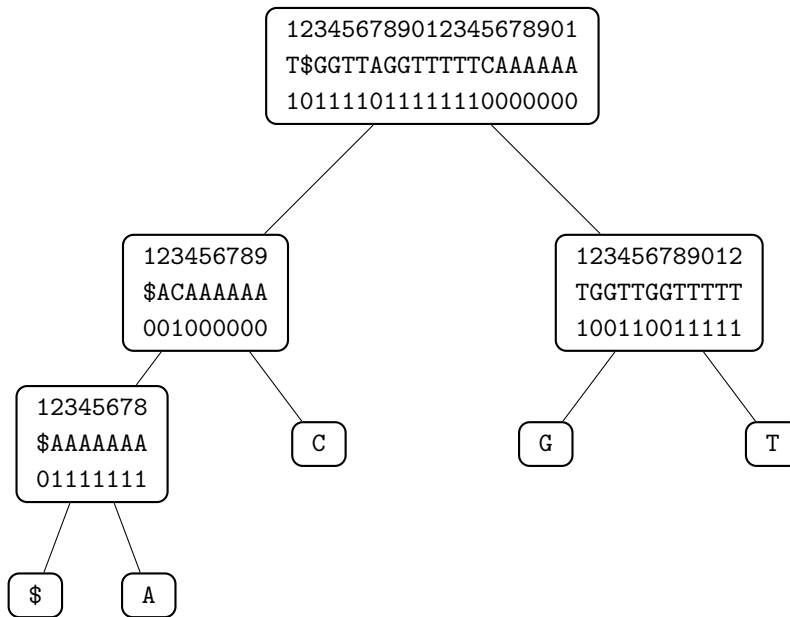


Abbildung 2: Wavelet-Baum von *T\$GGTTAGGTTTTCAA AAAA*