Projektmanagement im Softwarebereich: SeqAn

Marjan Faizi
Mathematik und Informatik Fachbereich
Freien Universität Berlin

May 23, 2013

1 Projektplan

Meine Gruppe sollte in SeqAn eine BlastX Variante implementieren. BlastX übersetzt eine gegebene Nukleotidsequenz mit sechs möglichen frames in ihre Aminosäuresequenzen, um anschließend diese mit einer Proteindatenbank zu vergleichen. Wir haben versucht ein schnelleres Programm zu entwickeln, indem wir zum Einen nicht das normale Aminosäurealphabet genommen haben, sondern zuerst ein reduzierte Alphabet erstellt haben. Was die Suche außerdem noch schneller machen sollte, war das Erstellen zweier Indizes sowohl ber den Seeds als auch ber die Datenbank, dies geschieht parallel. Danach sollten wir die Matches noch verifizieren und nur die statistisch signifikanten in eine Datei schreiben. Anschließend wurde das Programm getestet.

2 Verlauf

Um die reduzierten Alphabete zu erstellen, war es sehr hilfreich, dass in SeqAn bereits Scorematrizen gespeichert sind. Dadurch ist unser Programm flexibler, da man zwischen acht verschiedenen Matrizen auswählen kann.
Die Methode GET_ ALPHABET() ist verbesserungswürdig, da sie das Alphabet leider nur minimal auf sieben Zeichen clustern kann, bei niedrigeren Werten gelangt es in eine Endlosschleife, jedoch blieb nicht genug Zeit um dies zu ändern.
Die Konstanten K und λ sind nicht optimal gewählt, weil sie für die Blosum62 berechnet wurden. Wir mussten aber für die Scoreberechnung auf die Blosum30 zurückgreifen, da die andere nicht in SeqAn implementiert war bzw. nicht funtioniert hat.
Bei den Tests am Ende konnte ich nicht alle acht Scorematrizen testen, daher habe mich für drei entschieden und auch nicht alle möglichen Alphabetlängen, da meine Rechnerleistung das nicht hergab.