

Vorlesung Algorithmische Bioinformatik WS 2008/09

December 19, 2008

Themen zur Vorbereitung der Zwischenklausur

Diese Stichpunkte sollen Anhaltspunkte zur Vorbereitung auf die Zwischenklausur am 5. Januar 2009 dienen.

Paarweises Alignment

- dotplots
- Needleman-Wunsch und Smith-Waterman-Algorithmen
- Initialisierung der DP-Matrix und Rekursionsformeln
- Scorefunktionen und Score-Matrizen für AA Sequenzen
- Edit-Distanz / Hamming-Distanz
- lineare und linear-affine Gap-Funktionen
- Idee beim Algorithmus mit linearem Speicherbedarf (Hirshberg)
- Idee beim Gotoh Algorithmus
- free end gaps bzw. overlap-Alignment

Datenbanksuchen und Alignmentstatistik

- Strategien zur Datenbanksuche
- Vorgehen bei BLAST
- Poisson- und Extremwertverteilung
- Geometrische und negative Binomialverteilung
- Idee zur analytischen Berechnung von p -Werten
- Vorgehen zur empirischen Berechnung von p -Werten
- E -Werte
- Profilsuche
- Definition false positive, true positive etc.
- Was bedeuten die Begriffe Spezifität, Sensitivität?

Multiples Alignment

- Sum of pairs (SP) Score
- Komplexität beim SP-optimalen Alignment
- Carillo-Lipman-Schranke
- Star alignment
- Prinzipien des progressives Alignment (clustal/muscle)

Molekulare Phylogenetik

- Merkmals- und distanzbasierte Verfahren
- Maximum Parsimony (Fitch-Algorithmus, Long-Branch-Attraction)
- Additive Metrik (Vierpunktbedingung), Ultrametrik (Dreipunktbedingung)
- UPGMA (wann ok/wann schlecht?) / Clustering
- Idee/Komplexität bei Neighbor Joining
- Bootstrapping
- Maximum Likelihood Schätzung
- Markov Modell (Markov-Eigenschaft)
- Dayhoffs PAM Matrizen
- Jukes-Cantor Modell / Jukes-Cantor Korrekturformel
- Markov-Prozesse, Matrixexponential
- Stationäre Verteilung, Reversibilität
- Was ist die Likelihood eines Baumes?

Hidden Markov Modelle

- Was ist log odds score / bzw. log likelihood ratio?
- Definition eines HMM und Anwendungen in der Bioinformatik
- Viterbi-Algorithmus, Vorwärts-Algorithmus
- Idee Baum-Welch Algorithmus
- Was ist ein Profil-HMM?

Genomevolution

- Homologie, Orthologie, Paralogie
- In-Paraloge, Out-Paraloge, Co-Orthologe
- Idee des tree reconciliation

Kartierung und Sequenzierung

- Sequenzierstrategien, Motivation für STS
- Consecutive ones (CO) property und Ansätze zur Lösung
- Lander-Waterman Theorem

RNA

- Sekundärstrukturelemente und ihre Darstellung
- Zielfunktion bei Nussinov- und bei Zuker-Algorithmus

Motivsuche

- Was ist ein Profil?
- Pseudocounts
- PSSMs

Genvorhersage

- Lösungen in Prokaryoten und Eukaryoten
- Dynamic programming-Ansatz nach Krogh