

# 1 Bioinformatik 1 - Sequenzanalyse und Genomik

## 1.1 Sequenz-Alignment-Problem

1.1.1 globale, lokale, semiglobale Alignments

1.1.2 affine Gap-Kosten

1.1.3 Substitutionsmatrizen, PAM und BLOSUM

1.1.4 BLAST, FASTA

## 1.2 Multiple Alignments

1.2.1 Alignment von Profiles

1.2.2 clustalw4

## 1.3 Phylogenierekonstruktion

1.3.1 distanzbasierte Methoden

Neighbor-Joining<sup>11 3</sup>

UPGMA, WPGMA<sup>13</sup>

Split decomposition, Splittree<sup>1</sup>

1.3.2 characterbasierte Methoden

Maximum-Likelihood<sup>8</sup>

Maximum Parsimony (Fitch-Algorithmus)

1.3.3 Tests

Bootstrapping

Quartett-Mapping

## 2 Bioinformatik 2 -

### 2.1 Proteinstrukturen

2.1.1 Gittermodell, self-avoiding walk

2.1.2 Threading

### 2.2 RNA Sekundärstruktur<sup>10</sup>

2.2.1 Darstellungsmethoden

2.2.2 Algorithmus von Nussinov<sup>13</sup>

2.2.3 Algorithmus von Zuker<sup>6</sup>

2.2.4 Partition Function<sup>5</sup>

2.2.5 Stochastic Context Free Grammars<sup>7</sup>, CYK Algorithmus

<http://de.wikipedia.org/wiki/Cocke-Younger-Kasami-Algorithmus>

2.2.6 Sankoff-Algorithmus<sup>12</sup>

## 3 Bioinformatik 3 - Fortgeschrittene Methoden der Bioinformatik

### 3.1 RNA-folding and search for joint structures

Eine Punktmutation hat eine Chance von  $\sim 30\%$  die komplette RNA-Struktur zu zerstören. Nach 100 mya (million years), dies entspricht der Entfernung des LCA (least common ancestors) von Mensch und Maus, wurde im Schnitt jede Sequenzposition einmal mutiert. Wenn trotz geringer Sequenzähnlichkeit zwei oder mehr RNAs die gleiche Struktur aufweisen, dann ist dies ein Nachweis für die Selektion der Struktur.

Pseudoknoten in RNA-Strukturen sind durch normale dynamische Programmierungsansätze nicht auffindbar.

Um die Konsensusstruktur mehrerer Sequenzen zu ermitteln, können drei unterschiedliche Wege beschritten werden.

1. erst ein Sequenzalignment und dann dieses Falten,
2. erst alle RNAs falten und dann Strukturalignment,
3. oder versuchen sowohl Sequenz- als auch Strukturalignment parallel zu berechnen

## 3.2 Vergleich von RNA Sekundärstrukturen

### 3.2.1 Baumdarstellung

### 3.2.2 Tree Alignments, Zhang-Shasha Algorithmus<sup>2</sup>

## 3.3 Hannenhalli-Pevzner Theory, Sorting by Reversals

## 3.4 Zyklische Alignments<sup>9</sup>

## 3.5 BBQ

## 3.6 Hidden Markov Models für Cp G Islands

## 3.7 Image processing of biological data (S. Prohaska)

## 3.8 Präbiotische Evolution

### 3.8.1 Quasispecies und Hypercycle

#### Literatur

1 D. Huson. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* Vol. 14(1): 68-73. [LINK]

2 K. Zhang and D. Shasha. (1989) Simple Fast Algorithms for the Editing Distance between Trees and Related Problems. *SIAM Journal on Computing* Vol. 18(6): 1245-1262.

3 J. Studier and K. Keppler. (1988) A note on the neighbor-joining algorithm of Saitou and Nei. *Mol Biol Evol* Vol. 5(6): 729-731. [LINK]

4 J. Thompson, D. Higgins and T. Gibson. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.. *Nucleic Acids Res.* Vol. 22: 4673-4680.

5 J. McCaskill. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* Vol. 29(6-7): 1105-1119.

6 M. Zuker and P. Stiegler. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucl. Acids Res.* Vol. 9(1): 133-148. [LINK]

7 R. Dowell and S. Eddy. (2004) Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinformatics* Vol. 5(1): 71. [LINK]

8 J. Felsenstein. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* Vol. 17(6): 368-376. [LINK]

9 J. Gregor and M. Thomason. (1993) Dynamic Programming Alignment of Sequences Representing Cyclic Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* Vol. 15(2): 129-135.

10 I. Hofacker, W. Fontana, P. Stadler, L. Bonhoeffer, M. Tacker and P. Schuster. (1994) Fast Folding and Comparison of RNA Secondary Structures. *Monatsh. Chem.* Vol. 125: 167-188.

11 N. Saitou and M. Nei. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* Vol. 4(4): 406-425. [LINK]

12 D. Sankoff. (1985) Simultaneous Solution of the RNA Folding, Alignment and Protosequence Problems. *SIAM Journal on Applied Mathematics* Vol. 45(5): 810-825. [LINK]

13 R. Nussinov, G. Pieczenik, J. Griggs and D. Kleitman. (1978) Algorithms for Loop Matchings. *SIAM Journal on Applied Mathematics* Vol. 35(1): 68-82. [LINK]

14 R. Sokal and C. Michener. (1958) A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* Vol. 38: 1409-1438.