

Problemstellungen der Bioinformatik

Proseminar im Grundstudium, Sommersemester 2003

Themen

1	Genetischer Fingerabdruck	2
2	Sequence Alignment	2
3	FASTA und BLAST	2
4	Sequenzierung	2
5	Proteomics	2
6	Fragmentassemblierung: die „shot gun“-Methode	3
7	Die Gentechnologie als Industriezweig	3
8	Phylogenetische Bäume	3
9	Microarrays I: Technischer Teil	3
8	Microarrays II: Algorithmischer Teil	3
9	Physische und Genetische Karten	4
10	Proteinstrukturvorhersage	4
11	Proteindocking	4

1 Genetischer Fingerabdruck

- Inhalt
- was ist eine genetischer Fingerabdruck, wie wird er erstellt;
 - einige Grundlagen zur Eindeutigkeit ([KS94], Kap. 2 und 4, ev. Kap. 5)
 - historische Fälle (Kaspar Hauser, etc.)
- Literatur
- selbständige Recherche (Internet), guter Startpunkt: www.benecke.com
 - Biologie-Lehrbuch (Abitur?)

2 Sequence Alignment

- Inhalt
- Bewertungsfunktionen
 - lokale/globale Alignments
 - Algorithmen
 - dynamisches Programmieren
- Literatur
- [Gus97] [SM97]: Kap. 3; ev. [Wat95]: Kap. 9;

3 FASTA und BLAST

- Inhalt
- Sequenzdatenbanken, Historisches
 - die approximativen Methoden FASTA und BLAST
 - PAM-Matrizen und BLOSSUM-Matrizen
- Literatur
- [SM97] Kap 3.5 ;[Gus97]

4 Sequenzierung

- Inhalt
- In diesem Vortrag geht es hauptsächlich um die Beschreibung technischer Verfahren, die der Bestimmung von Nukleotidsequenzen und Aminosäuresequenzen dienen.
- Verfahren zur Sequenzierung von DNA und RNA
 - Verfahren zur Proteinsequenzierung
- Literatur
- [Alp98], Kapitel 1,2,3. Ausgewählte Themen aus Kapitel 4,5,7.

5 Proteomics

- Inhalt

6 Fragmentassemblierung: die „shot gun“-Methode

Inhalt	• Biologischer Hintergrund, Zielstellung bei der Fragmentassemblierung • Modelle und Algorithmen • Heuristiken
Literatur	[SM97]: Kap. 4; [Cas]; [BB03], Kap. 8

7 Die Gentechnologie als Industriezweig

Wer sind die “global players” in Bereich Biotechnologie/Gentechnologie? In welchen Produkten gelangen wissenschaftliche Erkenntnisse der letzten Jahre zum Einsatz? Wo werden diese Produkte eingesetzt?

Welche Firmen sind insbesondere in Deutschland mit innovativen Produkten auf dem Markt?

Nehmen Sie Bezug zu ökologischen und ethischen Aspekten.

Startpunkte/ Schlagworte:

- Celera Genomics/ Craig Venter
- <http://finance.yahoo.com>
- <http://biz.yahoo.com/ic>

8 Phylogenetische Bäume

Inhalt	Ermitteln von Stammbäumen, wesentliche Konzepte, Ansätze und Algorithmen
Literatur	[SM97]: Kap. 6; [Gus97]; [Wat95]

9 Microarrays I: Technischer Teil

Inhalt	• Funktionsweise und Herstellung • Typen von Microarrays • Vorverarbeitung und Normalisierung
Literatur	[Bow99]; [KPB ⁺ 98]; [Hac99]; Das Affymetrix Benutzerhandbuch
Weitere Links	www.affymetrix.com

8 Microarrays II: Algorithmischer Teil

Inhalt	• statistischer Hintergrund • Merkmalsextraktion mit TTest und FoldChange, • einfache Klassifikation nach Golub.
Literatur	statistischer Hintergrund: ein einführendes Statistikbuch mit Kapitel über Tests etc. Anwendung: [STM ⁺ 00]; [GST ⁺ 99]

9 Physische und Genetische Karten

folgt

10 Proteinstrukturvorhersage

- Inhalt
- Problemstellung
 - aktuelle Lösungsansätze (homology modeling, fold recognition, ab initio)

Literatur

Biochemische Hintergründe:
[BNP69]
Allgemein zur Vorhersageproblematik:
[KD97]; [LTZ96]
[Fas90]: Kap. 17; [Gla95]: Kap. 9.III
Genetische Algorithmen:
[CSZ92] [JWG⁺97]

Weitere Links www.tcm.phy.cam.ac.uk/~mmlk2/report13/report13.html

11 Proteindocking

- Inhalt
- Problemstellung
 - physikalische und chemische Grundlagen für Wechselwirkungen
 - Modelle molekularer Oberflächen
 - Algorithmen zur Lösung des Dockingproblems
- Literatur
- F. Ackermann, *Prädiktion und Bewertung von Protein-Protein-Dockingpositionen*, Verlag Hans Jacobs, 1997
 - http://web.informatik.uni-bonn.de/FGL/_vorlesung25/vorlesung25.html

Bei der angegebenen Literatur handelt es sich um eine “Basisausrüstung” – es können und sollen auch andere Quellen hinzugezogen werden.

Literatur

- [Alp98] Luke Alphey. *DNA-Sequenzierung*. Spektrum Akademischer Verlag, Heidelberg, Berlin, 1998.
- [BB03] Hans-Joachim Böckenhauer and Dirk Bongartz. *Algorithmische Grundlagen der Bioinformatik*. Teubner, 2003.
- [BNP69] W. J. Browne, A. C. T. North, and D. C. Phillips. A possible three-dimensional structure of bovine α -lactalbumin based on that of hen’s egg-white lysozyme. 42:65–86, 1969. Historisches Paper mit handgemachter Homologievorhersage, Drahtmodell und Stereophotos.
- [Bow99] David D.L. Bowtell. Options available – from start to finish – for obtaining expression data by microarray. *Nature Genetics*, 21:25–32, 1999.
- [Cas] D. Casey. Primer on molecular genetics. <http://www.ornl.gov/hgmis/publicat/primer/intro.html>.
- [CSZ92] T. Coleman, D. Shalloway, and W. Zhijun. Isotropic effective energy simulated annealing searches for low energy molecular cluster states. Technical Report 92-113, 1992.
- [Fas90] Gerald D. Fasman, editor. *Prediction of protein structure and the principles of protein conformation*. Plenum Press, 1990. Einführung in warum-Proteine-wie-falten, blickt über mehrere Proteinstrukturvorhersagemethoden.
- [Gla95] Jay A. Glasel, editor. *Introduction to biophysical methods for protein and nucleic acid research*. Academic Press, 1995. Physikalische Beschreibung von Röntgenstrukturanalyse und Beschreibung von Faltungsvorhersage.
- [GST⁺99] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286:531–537, 1999.
- [Gus97] Dan Gusfield. *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*. Cambridge University Press, Cambridge, 1997.
- [Hac99] J.G. Hacia. Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genetics*, 21:42–47, 1999.
- [JWG⁺97] Gareth Jones, Peter Willett, Robert C. Glen, Andrew Leach, and Robin Taylor. Development and validation of a genetic algorithm for flexible docking. *Journal Molecular Biology*, 267:727–748, 1997.
- [KD97] R. König and T. Dandekar. Computational methods for the prediction of protein folds. *Biochemica et Biophysica acta*, 1343(1):1, 1997.
- [KPB⁺98] A. Kel, A. Ptitsyn, V. Babenko, S. Meier-Ewert, and H. Lehrach. A genetic algorithm for designing gene family-specific oligonucleotide sets used for hybridization: the G protein-coupled receptor protein superfamily. *Bioinformatics*, 14(3):259–270, 1998.
- [KS94] M. Krawczak and J. Schmidtke. *DNA-Fingerprinting*. Spektrum Akademischer Verlag, 1994.

- [LTZ96] T. Lengauer, R. Thiele, and R. Zimmer. Modellierung von proteinstrukturen. *Der GMD-Spiegel*, 2/3:14–18, 1996.
- [SM97] Joao Setubal and Joao Meidanis. *Introduction to Computational Molecular Biology*. PWS Publishing, Boston, Mass., 1997.
- [STM⁺00] Donna K. Slonim, Pablo Tamayo, Jill P. Mesirov, Todd R. Golub, and Eric S. Lander. Class prediction and discovery using gene expression data. pages 263–272, 2000.
- [Wat95] Michael S. Waterman. *Introduction to Computational Biology: Maps, Sequences and Genomes*. Chapman & Hall, London, 1995.