

Problemstellungen der Bioinformatik

Proseminar im Grundstudium, Sommersemester 2003

Themen

1	Sequence Alignment	2
2	Multiple Sequence Alignment	2
3	FASTA und BLAST	2
4	Sequenzierung	2
5	Genetische und physikalische Karten	2
6	Fragmentassemblierung: die shot gun Methode	3
7	Human Genome Project: Sequenzierung des menschlichen Genoms	3
8	Phylogenetische Bäume	3
9	Microarrays: Datengewinnung, Vorverarbeitung und Normalisierung	3
10	Analyse von Microarrays: Clustern	4
11	Analyse von Microarrays: Merkmalsauswahl und Klassifikation	4
12	Analyse von Microarrays: neuronale Netze zur Klassifikation	4
13	Röntgenkristallographie, NMR und Proteindatenbanken	4
14	Vorhersage von Proteinstrukturen	4
15	Protein-Ligand-Docking	5
16	DNA-Computing	5
17	Genetischer Fingerabdruck	5
18	Ethische Aspekte der Bioinformatik	5

1 Sequence Alignment

Inhalt Vorstellen der Algorithmen zur Berechnung von lokalen und globalen optimalen Alignments zwischen zwei Sequenzen (vor allem dynamisches Programmieren sowie verschiedene Bewertungsfunktionen für die Ähnlichkeit).

Literatur [Gus97] [Set97]: Kap. 3; ev. [Wat95]: Kap. 9;

2 Multiple Sequence Alignment

Inhalt Vorstellen der Algorithmen zur Berechnung von lokalen und globalen optimalen Alignments zwischen mehreren Sequenzen (dynamisches Programmieren und heuristische Verfahren)

Literatur [Gus97] [Set97]: Kap. 3.4; [Wat95]: Kap. 10; ev. [Lee02]

3 FASTA und BLAST

Inhalt Vorstellen von Sequenzdatenbanken, sowie PAM und BLOSSUM und genaue Erläuterung des Sequenzvergleiches, der dort angewendet wird

Literatur [Set97] Kap 3.5 ;[Gus97]

4 Sequenzierung

Inhalt Erklären der üblichen Verfahren zur Sequenzierung von DNA und RNA (chemische Methode, Kettenabbruchmethode) und der Aminosäuresequenzierung. Beschreiben der Funktionsweise der Gelelektrophorese. Weitere eventuelle Themen: 2D-gelelektrophorese und PAGE.

Literatur [Alp98]

5 Genetische und physikalische Karten

Inhalt

- Genetische Karten: Bestimmung der (relativen) Lage von Genen auf den Chromosomen (nur kurz)
- Physikalische Karten: Bestimmen der Lage von größeren DNA-Teilen und/oder Markern

Verfahren bei der Annahme fehlerfreier Daten und unter Berücksichtigung von Fehlern (vor allem: double digest, single digest, restriction site mapping, hybridisation mapping)

Literatur [Cas]; [Wat95]: Kap. 6; [Set97]: Kap. 1.5 und 5; [Gus97]: Teil aus Kap. 16, [Böc03], Kap. 7

Weitere Links <http://www.cs.technion.ac.il/Labs/cbl/teaching/bab/>

6 Fragmentassemblierung: die shot gun Methode

Inhalt Zusammensetzen von Teilsequenzen eines größeren DNA-Stücks unter Berücksichtigung von Fehlern sowie unter Annahme der Korrektheit.
(heuristischer und exakter Algorithmus)

Literatur [Set97]: Kap. 4; [Cas]; [Böc03], Kap. 8

7 Human Genome Project: Sequenzierung des menschlichen Genoms

Als Startpunkt: [Con01]

8 Phylogenetische Bäume

Inhalt Ermitteln von Stammbäumen, die dabei auftretenden Problem und approximative Lösungen dafür.

Literatur [Set97]: Kap. 6; [Gus97]; [Wat95]

9 Microarrays: Datengewinnung, Vorverarbeitung und Normalisierung

Inhalt Vorstellung verschiedener Arten von Microarrays, ihrer Herstellung und der Datengewinnung.

weitere Themen zur Vertiefung:

- Erläuterung eines genetischen Algorithmus' zur Bestimmung von Sonden
- Normalisierung der Daten, sodass man mehrere Microarrays miteinander vergleichen kann

Literatur [Bow99]; [Kel98]; [Hac99]; Das Affymetrix Benutzerhandbuch

Weitere Links www.affymetrix.com

10 Analyse von Microarrays: Clustern

Inhalt Distanzmaße und Linkage-Verfahren, sowie hierarchisches Clustern am Beispiel des Eisen-Programms. Weitere Themen: SOMs oder k-means.

Literatur Clustern allgemein: Ein Buch übers Clustern (z.B. [Gor99], bei Stefan Posch)
Anwendungen: [Eis98]; [Tam99]

11 Analyse von Microarrays: Merkmalsauswahl und Klassifikation

Inhalt Merkmalsextraktion mit TTest und FoldChange, einfache Klassifikation nach Golub.
Vor allem: statistischer Hintergrund

Literatur statistischer Hintergrund: ein einführendes Statistikbuch mit Kapitel über Tests etc.
Anwendung: [Slo00]; [Gol99]

12 Analyse von Microarrays: neuronale Netze zur Klassifikation

Inhalt Was sind neuronale Netze und wie funktionieren sie, einfache Klassifikation am Beispiel der Golub-Daten, wie im Khan-Paper beschrieben

Literatur [Gol99], [Kha01], [Ber02]
ein Buch über neuronale Netze, z.B. [Cal03]

13 Röntgenkristallographie, NMR und Proteindatenbanken

Literatur [Gla95]: S.389-411; [Leh94]: S.194, 367ff; [Ber77]; [Sus98]

14 Vorhersage von Proteinstrukturen

Inhalt Allgemein das Problem der Vorhersage einer Struktur aus einer bekannten Sequenz und aktuelle Lösungsansätze (homology modeling, fold recognition, ab initio)

Literatur *Biochemische Hintergründe:*
[Bro69]
Allgemein zur Vorhersageproblematik:
[Kön97]; [Len96]
[Fas90]: Kap. 17; [Gla95]: Kap. 9.III
Genetische Algorithmen:
[Col92] [Jon97]

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

15 Protein-Ligand-Docking

Literatur FlexX <http://cartan.gmd.de/flexx/>; [Jon97]

16 DNA-Computing

[Bra02, Rub00]

17 Genetischer Fingerabdruck

was ist eine genetischer Fingerabdruck, wie wird er erstellt; einige Grundlagen zur Eindeutigkeit ([Kra94], Kap. 2 und 4, ev. Kap. 5)

möglicherweise Anwendungen in der Forensik, oder historische Fälle (Kaspar Hauser)

Weitere Infos: www.benecke.com

18 Ethische Aspekte der Bioinformatik

Diskussion (gern auch interaktiv) über obiges Thema

Möglicher Startpunkt:

http://www.ornl.gov/sci/techresources/Human_Genome/elsi/elsi.shtml

www.ornl.gov/hgmis/publicat/genechoice/yourgenes.pdf

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

Literatur

- [Alp98] L. Alphey: *DNA-Sequenzierung*, Spektrum Akademischer Verlag, Heidelberg, Berlin, 1998.
- [Ber77] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. J. Meyer, M. Brice, J. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi: *The Protein Data Bank: a computer-based archival file for macromolecular structures*, *Journal of Molecular Biology*, Bd. 112, 1977, S. 535–542.
- [Ber02] D. Berrar, C. Downes, W. Dubitzky: *Multiclass Cancer Classification Using Gene Expression Profiling and Probabilistic Neural Networks*, University of Ulster, Northern Ireland, 2002.
- [Böc03] H.-J. Böckenhauer, D. Bongartz: *Algorithmische Grundlagen der Bioinformatik*, Teubner, 2003.
- [Bow99] D. D. Bowtell: *Options available – from start to finish – for obtaining expression data by microarray*, *Nature Genetics*, Bd. 21, 1999, S. 25–32.
- [Bra02] R. S. Braich, N. Chelyapov, C. Johnson, P. W. Rothmund, L. Adleman: *Solution of a 20-Variable 3-SAT Problem on a DNA Computer*, *Science*, Bd. 296, 2002, S. 499–502.
- [Bro69] W. J. Browne, A. C. T. North, D. C. Phillips: *A Possible Three-dimensional Structure of Bovine α -Lactalbumin based on that of Hen’s Egg-White Lysozyme*, *Journal of Molecular Biology*, Bd. 42, 1969, S. 65–86, Historisches Paper mit handgemachter Homologievorhersage, Drahtmodell und Stereophotos.
- [Cal03] R. Callan: *Neuronale Netze im Klartext*, Pearson Studium, 2003.
- [Cas] D. Casey: *Primer on Molecular Genetics*, http://www.ornl.gov/sci/techresources/Human_Genome/p
- [Col92] T. Coleman, D. Shalloway, W. Zhijun: *Isotropic Effective Energy Simulated Annealing Searches for Low Energy Molecular Cluster States*, 92-113, 1992.
- [Con01] I. H. G. S. Consortium: *Initial sequencing and analysis of the human genome*, *Nature*, Bd. 409, Nr. 6822, February 2001, S. 860–921, (Dan Brown is a member of IHGSC).
- [Eis98] M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein: *Cluster analysis and display of genome-wide expression patterns*, *Proc. Natl. Acad. Sci. USA*, Bd. 95, 1998, S. 14863–14868.
- [Fas90] G. D. Fasman (Hrsg.): *Prediction of protein structure and the principles of protein conformation*, Plenum Press, 1990, Einführung in warum-Proteine-wie-falten, bersicht ber mehrere Proteinstrukturvorhersagemethoden.
- [Gla95] J. A. Glasel (Hrsg.): *Introduction to biophysical methods for protein and nucleic acid research*, Academic Press, 1995, Physikalische Beschreibung von Rntgenstrukturanalyse und Beschreibung von Faltungsvorhersage.

- [Gol99] T. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing, M. Caligiuri, C. Bloomfield, E. Lander: *Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring*, *Science*, Bd. 286, 1999, S. 531–537.
- [Gor99] A. Gordon: *Classification*, Chapman and Hall, 1999.
- [Gus97] D. Gusfield: *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*, Cambridge University Press, Cambridge, 1997.
- [Hac99] J. Hacia: *Resequencing and mutational analysis using oligonucleotide microarrays*, *Nature Genetics*, Bd. 21, 1999, S. 42–47.
- [Jon97] G. Jones, P. Willett, R. C. Glen, A. Leach, R. Taylor: *Development and Validation of a Genetic Algorithm for Flexible Docking*, *Journal Molecular Biology*, Bd. 267, 1997, S. 727–748.
- [Kel98] A. Kel, A. Ptitsyn, V. Babenko, S. Meier-Ewert, H. Lehrach: *A genetic algorithm for designing gene family-specific oligonucleotide sets used for hybridization: the G protein-coupled receptor protein superfamily*, *Bioinformatics*, Bd. 14, Nr. 3, 1998, S. 259–270.
- [Kha01] J. Khan, J. Wei, P. S. Meltzer: *Classification and diagnostic prediction of cancer using gene expression profiling and artificial neural networks*, *Nature Medicine*, Bd. 7(6), 2001, S. 673–679.
- [Kön97] R. König, T. Dandekar: *Computational methods for the prediction of protein folds*, *Biochimica et Biophysica acta*, Bd. 1343, Nr. 1, 1997, S. 1.
- [Kra94] M. Krawczak, J. Schmidtke: *DNA-Fingerprinting*, Spektrum Akademischer Verlag, 1994.
- [Lee02] C. Lee, C. Grasso, M. Sharlow: *Multiple sequence alignment using partial order graphs*, *Bioinformatics*, Bd. 18(3), 2002, S. 452–464.
- [Leh94] A. L. Lehninger, D. L. Nelson, M. M. Cox: *Prinzipien der Biochemie*, Springer, 1994, Dickes Biochemie Lehrbuch.
- [Len96] T. Lengauer, R. Thiele, R. Zimmer: *Modellierung von Proteinstrukturen*, *Der GMD-Spiegel*, Bd. 2/3, 1996, S. 14–18.
- [Rub00] A. J. Ruben, L. F. Landweber: *The past, present and future of molecular computing*, *Nature Reviews Molecular Cell Biology*, Bd. 1, 2000, S. 69–72.
- [Set97] J. Setubal, J. Meidanis: *Introduction to Computational Molecular Biology*, PWS Publishing, Boston, Mass., 1997.
- [Slo00] D. K. Slonim, P. Tamayo, J. P. Mesirov, T. R. Golub, E. S. Lander: *Class prediction and discovery using gene expression data*, in *Proceedings of the Annual International Conference on Research in Computational Molecular Biology (RECOMB)*, 2000, S. 263–272.
- [Sus98] J. L. Sussman, D. Lin, J. Jiang: *Protein Data Bank (PDB): Database of Three-Dimensional Structural Information of Biological Macromolecules*, *Acta Crystallographica Section D Biological Crystallography*, Bd. 6, 1998, S. 1078–.

- [Tam99] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander, T. R. Golub: *Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation*, *Proc. Natl. Acad. Sci. USA*, Bd. 96, Nr. 6, 1999, S. 2907–2912.
- [Wat95] M. S. Waterman: *Introduction to Computational Biology: Maps, Sequences and Genomes*, Chapman & Hall, London, 1995.