

# Algorithmische Bioinformatik

Halbkurs

Wintersemester 2005 / 2006

Ulf Leser

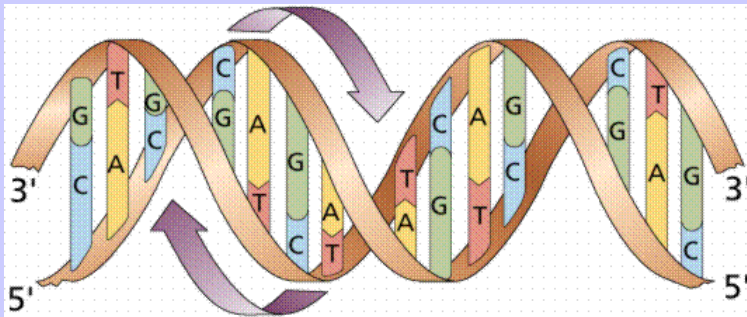
Wissensmanagement in der  
Bioinformatik



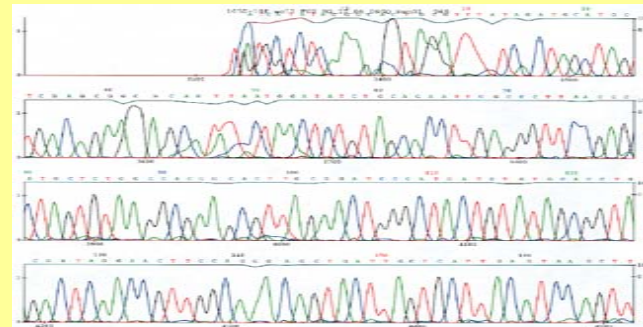
# Bioinformatik



25.4.2003  
50. Jubiläum der  
Entdeckung der  
Doppelhelix durch  
Watson/Crick



14.4.2003  
Humanes Genom zu 99%  
sequenziert mit 99.99%  
Genauigkeit



# Fragen ...

---

- Was heißt das?
- Warum sind 3.000.000.000 Zeichen für uns so wichtig?
- Wenn das Zeichen sind – gibt es Wörter? Eine Sprache?
- Wie wurden die Daten erhoben?
- Wie kann man daraus Wissen gewinnen?
  
- Welche Rolle spielt dabei die Informatik?

# Wofür kann man sie benutzen?

- Kommerzielle Gentests im WWW
- 2 Firmen
- 32 Tests
- Preise: 100–1400€

**Tabelle**  
**Gentests, die im Internet in Deutschland bestellbar sind (Stand: Juli 2002)**

Indikation*	Anbieter**	Untersuchungsgegenstand	Preis (inkl. MwSt.)
Alkoholverträglichkeit	2	keine Angaben (k. A.)	207,79 €
Alzheimer	2	k. A.	134,06 €
Alzheimer <sup>2)</sup>	1	E4-Allel des Apolipoprotein-E-Gens auf Chromosom 10	650,00 €
Angelman-Syndrom <sup>2)</sup>	1	Deletion auf dem Chromosom 15	850,00 €
Anti-Aging-Risikoprofil	2	k. A.	653,61 €
Arteriosklerose/Herzinfarkt/Schlaganfall	2	k. A.	512,81 €
Azotämie	1	31 Mutationen einschließlich einer 5T-Variante auf dem CFTR-Gen auf dem Chromosom 7	850,00 €
Bluthochdruck	2	k. A.	127,40 € 439,24 €
Diabetes Typ 2	2	k. A.	127,40 € 194,39 €
Dickdarmkrebs <sup>3)</sup>	1	MLH1- und MSH2-Mutationen	1600,00 €
Entgiftungsfähigkeit	2	k. A.	811,10 €
Faktor V Leiden-Mutation	1	Gerinnungsfaktor-V auf dem langen Arm von Chromosom 1	400,00 €
Familiäre Hypercholesterinämie	1	Mutationen im Low-Density-Lipoprotein-Rezeptor-Gen und im Exon 26 Apolipoprotein-B-Gen	850,00 €
Familiäre Hyperlipoproteinämie Typ III	1	E2-Allel des Apolipoprotein-E-Gens auf Chromosom 19	500,00 €
Familiärer Brustkrebs <sup>3)</sup>	1	BCRA1- und BCRA2-Mutationen	1400,00 €
Fettgen/Adipositas	2	k. A.	241,35 € 576,44 €
Fettstoffwechsel/Cholesterin	2	k. A.	395,48 €
Fragiles X-Syndrom <sup>4)</sup>	1	FMRI-(fragile X mental retardation-)Gen des X-Chromosoms (Region Xq27.3)	950,00 €
Hämochromatose	2	k. A.	207,84 €
Hämochromatose	1	Austausch der DNS-Basen Guanin zu Adenin an der Position 845 und von Cytosin zu Guanin an der Position 187 des HFE-Gens auf dem Chromosom 6	500,00 €
Hyperhomocysteinämie	1	k. A.	550,00 €
Mukoviszidose (Cystische Fibrose)	1	Mutation eines Gens auf Chromosom 7	850,00 €
Muskeldystrophie	1	Deletionen (Verlust von DNA-Teilsequenzen) im Dystrophin-Gen auf dem X-Chromosom	850,00 €
Osteoporose	2	k. A.	103,89 € 191,01 €
Osteoporose	1	Mutation (Basenaustausch von Guanin zu Thymin) im Intron 1 des Kollagen Typ I Alpha 1-Gens	650,00 €
Ovarialkarzinom <sup>3)</sup>	1	BCRA1- und BCRA2-Mutationen	850,00 €
Persönliches Ernährungsprofil	2	k. A.	841,32 €
Prader-Willi-Syndrom	1	Deletion oder Translokation auf dem langen Arm des Chromosoms 15 (15q11)	850,00 €
Prothrombin-Mutation	1	Austausch der DNS-Basen Guanin zu Adenin an der Position 20210 des Prothrombingens auf dem Chromosom 11	550,00 €
Risiko Alkohol- und Drogenabhängigkeit	2	k. A.	274,86 €
Thrombose	2	k. A.	134,06 € 281,52 €

# Halbkurs Algorithmisch Bioinformatik

---

- Vorlesung 4SWS
- Übung 2SWS
- Sprechstunde: Nach Vereinbarung  
Ulf Leser  
Raum: IV.103  
Tel: (030) 2093 – 3902  
eMail: [leser@informatik.hu-berlin.de](mailto:leser@informatik.hu-berlin.de)

# Termine und Prüfung

---

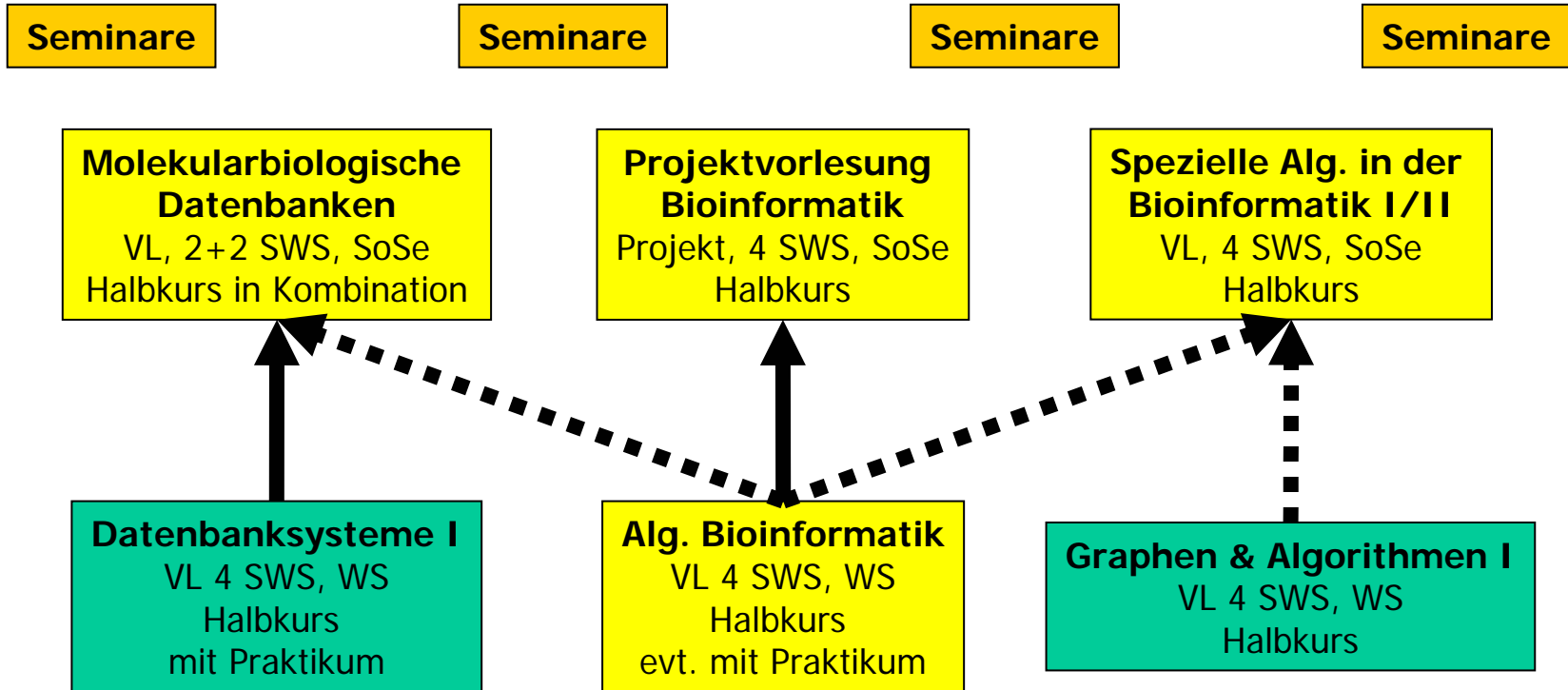
- Vorlesung
  - Donnerstag, 9 – 11 Uhr, RUD26, Raum 1'303
  - Freitag, 13 – 15 Uhr, RUD26, Raum 1'303
- Übung (Hakenberg, Bauckmann)
  - Donnerstag, 11 – 13 Uhr, RUD26, Raum 1'303
- Weihnachten: 17.12.2005 – 1.1.2006
- Voraussetzung für Prüfung
  - Je nach Studienordnung: Vordiplom
  - Bestehen der Übung
  - **Verständnis der Algorithmen**

# Gäste

---

- Die nächsten 5 Doppelstunden
  - Prof. Dr. Thomas Börner
  - Professor für Genetik
  - Zellen, Chromosomen, Gene, Transcription und Translation, differentielles Splicen, ...
  
- Im Januar/Februar
  - Dr. Piepenbrock, Epigenomics AG
  - Dr. Klein, Metanomics GmbH

# Bioinformatik@Informatik



Voraussetzung



Günstige, aber nicht notwendige  
Voraussetzung



Bioinformatik



Voraussetzung

# Literatur

---

- Primär
  - Dan Gusfield: „Algorithms on Strings, Trees, and Sequences“, Cambridge University Press, 1997 (ca. 60 Euro)
- Weitere
  - Böckenhauer, Bongartz: „Algorithmische Grundlagen der Bioinformatik“, Teubner, 2003 (ca. 40 Euro)
  - David Mount: „Bioinformatics. Sequence and Genome Analysis“, Cold Spring Harbour Press, 2001 (ca. 70 Euro)
  - Merkl & Waack: „Bioinformatik interaktiv“, Wiley-VCH, 2002 (Ca. 80 Euro)
  - Setubal & Meidanis: „Introduction to Computation Molecular Biology“, PWS Publishing, 1997 (ca. 90 Euro)
  - Gibson & Muse: „A primer of genome sciences“, Sinauer Associates, 2001 (ca. 50 Euro)

# Webseite

The screenshot shows a Mozilla Firefox browser window with the title 'Algorithmische Bioinformatik - Mozilla Firefox'. The address bar contains the URL 'http://www.informatik.hu-berlin.de/wbi/teaching/ws0506/hk\_bioinformatik/index.html'. The browser's menu bar includes 'File', 'Edit', 'View', 'Go', 'Bookmarks', 'Tools', and 'Help'. The bookmarks bar shows folders for 'Nachsehen', 'DBLP', 'MyStuff', 'Google', 'Lehre', 'News', 'Suchen', 'Buecher kaufen', 'Reisen', 'Paper suchen', and 'MyStuff'. The page content is on a blue background and includes the following text:

[Humboldt Universität zu Berlin](#) [Institut für Informatik](#)

## Algorithmische Bioinformatik

**Halbkurs im Wintersemester 2005/2006**  
**Professor Ulf Leser**

Der Halbkurs "Algorithmische Bioinformatik" behandelt Algorithmen zur Lösung grundlegender Fragestellungen moderner Molekularbiologie. Nach einer ausführlichen Einführung in die Grundlagen der Molekularbiologie (Gene und Genome, Expression, Proteine, Regulation und Transkription) werden die folgenden algorithmischen Probleme behandelt: Exaktes Stringmatching, Stringmatching mit mehreren Pattern, approximatives Matching, Indexstrukturen für Sequenzdatenbanken, Editabstand und Alignment, Multiples Alignment, Phylogenetische Bäume. Die Algorithmen werden jeweils anhand der zugrundeliegenden biologischen Fragestellung erklärt, wie z.B. Patternsuche in DNA- und Proteinsequenzen, Assembly von Teilsequenzen, Homologiesuche in Sequenzdatenbanken, und Berechnung evolutionärer Stammbäume.

**Voraussetzungen**  
Voraussetzung für den Besuch sind grundlegende Kenntnisse in Algorithmen. Kenntnisse in der Molekularbiologie werden nicht vorausgesetzt, sondern vermittelt.

**Prüfungen**  
Prüfungen sind mündlich. Die Vorlesung ist als Halbkurs der praktischen Informatik anrechenbar.

**Ort / Zeit:**

- Donnerstag, 9.00 - 11.00, RUD26, Raum 1'303
- Freitag, 13.00 - 15.00, RUD26, Raum 1'303

**Literatur zur Vorlesung**  
Dan Gusfield: "Algorithms on Strings, Trees, and Sequences", Cambridge University Press.  
Die Vorlesung folgt in grossen Teilen diesem Buch. Zusätzliche Literatur wird in den jeweiligen Stunden angegeben.

**Themen und Termine im Einzelnen**  
(Folien sind hier jeweils vor der Vorlesung als PDF verfügbar. Änderungen möglich).

• Administrative, Einleitung und Überblick

Done

# Evaluation 2004/2005

---

- Vorlesung
  - Durch die Geschwindigkeit des Vortrages ist es schwer, zu folgen
  - Script wäre praktisch
  - Wiederholung der letzten VL zum Anfang ist hilfreich.
  - Es ist verdammt viel Stoff, aber es wäre auch verdammt schade, etwas wegzulassen. Das sollte dann bei der Prüfung berücksichtigt werden.
- Übung
  - Anforderungen von 50 % pro Übungsblatt, 2x Fehlschuss -> Mist. traurig, falls das der einzige Weg sein soll, die Leute zur Bearbeitung zu motivieren
  - Zu wenig Beispiele in der Übung
  - Vorrechnen der Aufgaben ist "langweilig", wenn man sie selber gelöst hat
  - 12/14 ist ein bisschen zu hart zum bestehen. Aber nur ein bisschen
- ...

---

Fragen ?

# Fragen meinerseits

---

- Diplominformatiker?
- Semester?
- Prüfung?
  
- Spezielle Erwartungen?