

# Molekularbiologische Datenbanken

Sommersemester 2004

Ulf Leser

Wissensmanagement in der  
Bioinformatik



# 14. April 2003



Quelle: Focus,  
14.4.2003

- Was heißt das?
- Ist das nicht schon 2-mal verkündet worden?
- Was sind das für Daten?
- Wie wurden sie erhoben?
- Wo kann ich sie finden?
- Wie werden sie verwaltet?
- Was gibt es noch für molekularbiologische Datensammlungen?

# Molekularbiologische Datenbanken

---

- Vorlesung 2SWS
- Übung 2SWS
  
- Sprechstunde
  - Nach Vereinbarung / Offene Türen
  - IV.103
  - (030) 2093 – 3901
  - [leser@informatik.hu-berlin.de](mailto:leser@informatik.hu-berlin.de)

# Termine

---

- Vorlesung: Mi, 9.00 – 11.00 Uhr
- Übung: Mi, 11.00 – 13.00 Uhr
  
- Ausfallende Termine
  - 5.5.2004 (Tag der Informatik)
  
- Folien im Web jeweils vorab verfügbar

# Einbettung

---

- Voraussetzung: DBS-I
  - Relationalenmodell, ER, funkt. Abhängigkeiten
  - Joins, Anfrageübersetzung, Anfrageoptimierung
  - SQL
- Vorlesung Bioinformatik ist **keine** Voraussetzung
  - Grundlagen der Molekularbiologie kommen gleich
  - Stringalgorithmen brauchen wir hier nicht
- Diese Vorlesung ist „bunter“
  - Molekularbiologie – Experiment – Algorithmen – Datenbank
  - Algorithmen werden eher skizziert (da oft sehr heuristisch)

# Ziel der Vorlesung

---

- **Verständnis molekularbiologischer Fakten**
  - Gene, Proteine, Transkription, Expression, ...
- **Verständnis wichtiger experimenteller Techniken**
  - Hybridisierung, Clonierung, Kartierung, Sequenzierung, ESTs, Genexpression, Proteomics, ...
  - Konzeptionelles Verständnis, keine Biochemie
- **Kenntnis wichtiger Algorithmen zur Datenauswertung**
  - Von experimentellen Rohdaten zu „Wissen“
  - Content und Clone-Mapping, Base Calling und Sequenzassembly, EST-Clustering, Microarrayanalyse, ...
- **Kenntnis wichtiger Datenmodelle und Datenbanken**
  - Mapping-, Sequenz-, Expressions- und Proteindatenbanken
  - Semantik und Qualität der Daten, Modelle, Zugriffsmethoden, Verwendung

# Inhaltsübersicht –1-

---

- Heute
  - Vorstellung der Vorlesung & Übung
  - Grundbegriffe der Molekularbiologie
- Einführung in MDB
  - Human Genome Project
  - Beispieldatenbanken
  - Klassen von MDB
  - Typische Zugriffsmethoden
  - Anforderungen

# Inhaltsübersicht –2–

---

- Datenmodelle
  - Flatfile, relational, objektorientiert, XML & ACeDB
  - Ausgesuchte Modellierungsaspekte
    - Identifikation, Versionierung, widersprüchliche Daten
- Kartierungsdatenbanken
  - Genomkarten und Kartierung
  - Arten von Karten: physikalisch, genetisch, ...
  - Experimentelle Grundlagen: Clonierung, PCR, Hybridisierung
  - Algorithmen: Content und Clonemapping, kombinatorische Algorithmen, Umgang mit Fehlern
  - Datenmodell: OMG Standard „Genome Maps“
  - Konkrete Datenbanken

# Inhaltsübersicht –3–

---

- Sequenzdatenbanken
  - Sequenzierungsstrategien und –technik
  - Datenaufbereitung: Base-Calling, Assembly, Finishing
  - EST und cDNA: Zweck und Verwendung, EST Clustering
  - Datenmodelle: EMBL versus BioSQL
  - Konkrete Datenbanken
- Genexpression
  - Experimenteller Ansatz, Datenqualität, Vergleichbarkeit
  - Differentielle Expression und Co-Regulation
  - MAGE, MIAME, MGED – Standards und Datenmodelle
  - Expressionsdatenbanken

# Inhaltsübersicht –4–

---

- Proteindatenbanken
  - Proteinidentifikation: 2D-Gelelektrophorese, Massenspektroskopie, Spektren und Identifikation
  - Proteinstruktur: Faltungsarten und Typen, Klassifikationen, Chains, Proteins, Compounds
  - Proteininteraktion: Yeast2Hybrid, Modellierung von Zellsystemen, Systembiologie
  - Proteinsequenz: Domänen, Motive, Sites
  - SWISS-PROT, PDB, DIP, BIND, MINT, InterPro, Prosite, ...

# Übung –1–

---

- 2SWS
- Maximale Teilnehmerzahl: 16 Studenten
- Termin: Direkt nach der Vorlesung
  - Raum: RUD 26, Raum 0'313
- Erster Termin: Heute
- Anmeldung: GOYA, per eMail, per Erscheinen
- Inhalt: Entwicklung einer Microarraydatenbank
  - Modellierung
  - Datenupload
  - Patternmatching mit PRINTS
  - Annotation mit der GeneOntology

# Übung -2-

---

- 6 Aufgaben
- Gruppen a ? Leute
- Jeweils 1-3 Wochen Zeit
- Präsentation der Ergebnisse (und Arbeitsschritte)
- Entwicklung auf PostGres und Perl/Java/Python/...
- Übungsschein
  - Gruppe hat alle Aufgaben gelöst
  - Jedes Mitglied hat mindestens einmal die Lösung vorgestellt
- Wer kommt in die Übung?

# Anrechenbarkeit

---

- Zusammen mit der Vorlesung „Data Warehouses“ ist es ein Halbkurs
- Dazu braucht man einen Übungsschein
  - Entweder Molekularbiologische Datenbanken
  - Oder Data Warehouses
- Weitere Kombinationen auf Anfrage

# Literatur

---

- Primär
  - A. Baxevanis, F. Ouellette: „Bioinformatics – a Practical Guide ...“, John Wiley & Sons, 2001 (ca. 70 Euro)
  - C. Sensen (ed): „Essentials of Genomics and Bioinformatics“, Wiley-VCH, 2002 (ca. 80 Euro)
  - Jede Januarausgabe von *Nucleic Acid Research*,  
Online unter <http://nar.oupjournals.org/>
- Weitere
  - M. Campbell, L. Heyer: „Discovering Genomics, Proteomics, & Bioinformatics“, Benjamin Cummings, 2003 (ca. 70 Euro)
  - M. Kanehisa: „Post-Genome Informatics“, Oxford University Press, 2000
  - S. Letovsky: „Bioinformatics: Databases and Systems“, Kluwer Academic, 1999

---

# Fragen ?

# Fragen

---

- Diplominformatiker ?
- Semester ?
- DBS-I ?
- Übungsteilnahme ?
- Prüfung ?