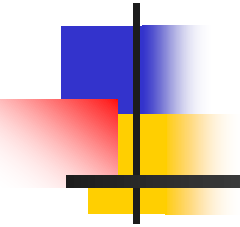


# Bioinformatics





# Course Outline

---

- Introduction to Molecular Biology
- Sequence alignment algorithms
- Genome and Proteome analysis
- Phylogenetic tree analysis
- Bioinformatics Algorithms
- Familiarity with current areas of active bioinformatics research

# What is Bioinformatics?

- The use of computers to collect, analyze, and interpret biological information at the molecular level.
- *"The mathematical, statistical and computing methods for biological problems using DNA and amino acid sequences and related information."*
- A set of software tools for molecular sequence analysis





# What is Bioinformatics?

---

- Bio - Molecular Biology
- Informatics - Computer Science
- Bioinformatics
  - The study of the application of molecular biology, computer science, artificial intelligence, statistics and mathematics
  - To model, organise, understand and discover interesting information associated with the large scale molecular biology databases



# What is Bioinformatics?

---

- “Application of Computer Sciences and allied technologies to answer the questions of biologists, about the mysteries of life”
- Mainly concerned with the various kinds of data emerging from the cells of living organisms
- “Application of computers in solving problems of Molecular Biology”



# Bioinformatics - origins

---

- Driven by experimental molecular biology
  - lab folks generate the data, then need a way to organize and analyze it
- Grabs methods from many different fields
  - biostatistics, machine learning, data mining, linguistics etc.
- Use computer (algorithms) to gain novel biological knowledge.
- Use biological knowledge to construct algorithms.



# Bioinformatics Applications

---

- Comparing sequences
- Detecting patterns in sequences
- Constructing phylogenetic trees
- Determining protein function and metabolic pathways
- Computer Aided Drug Design



# Bioinformatics Algorithms

---

- One biological problem can be solved by various algorithms
- One type of algorithm may be applicable to several different biological problems
- Hard to know if existing algorithms are “best possible” solution
- Research is most active in areas where existing algorithms/software do not give adequate solutions (not correct, or not time efficient)





# Classes of Algorithms

---

- Exhaustive search (i.e. brute force)
- Greedy Algorithms
- Combinatorial Pattern Matching
- Branch-and-bound
- Dynamic Programming
- Graph Algorithms
- Machine Learning
- Clustering



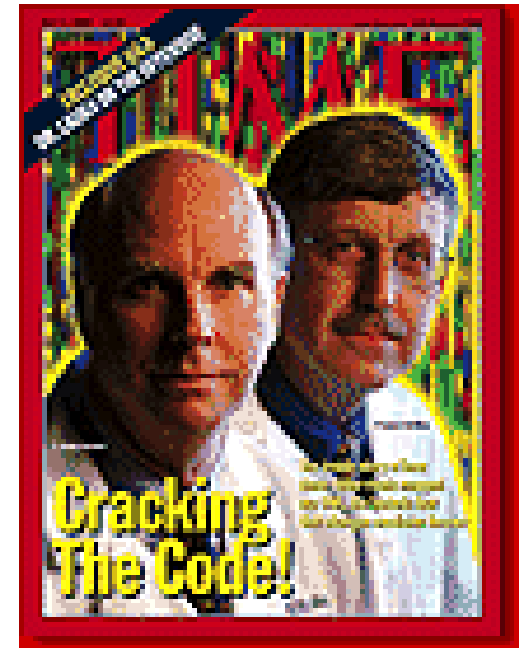
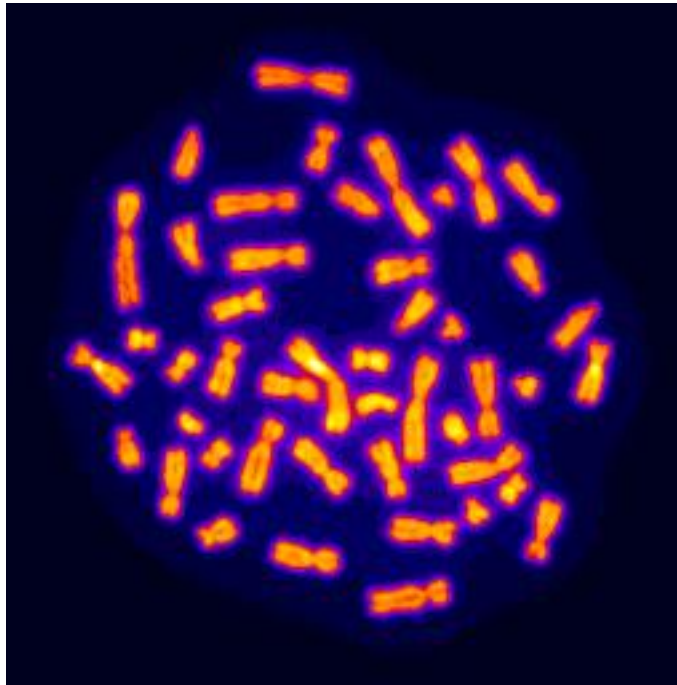
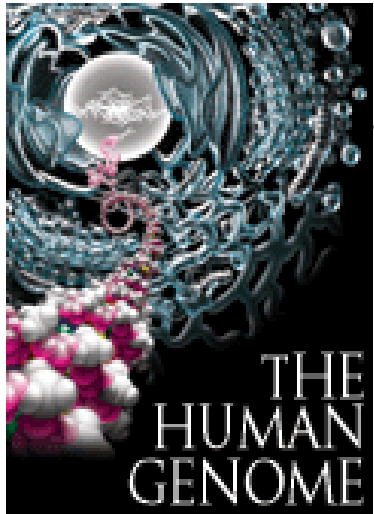
# **A Genome Revolution in Biology and Medicine**

---

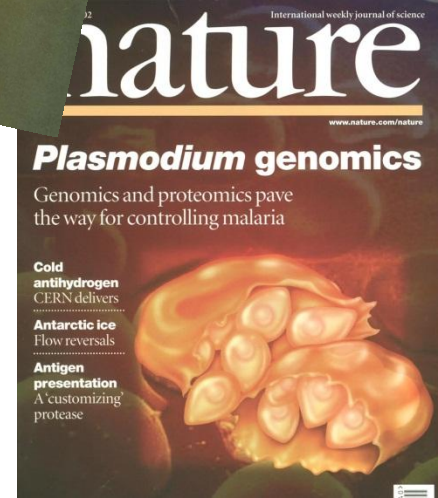
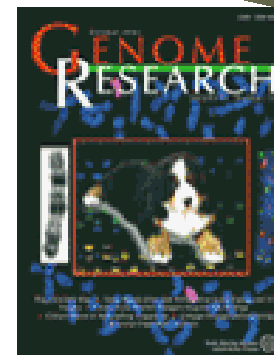
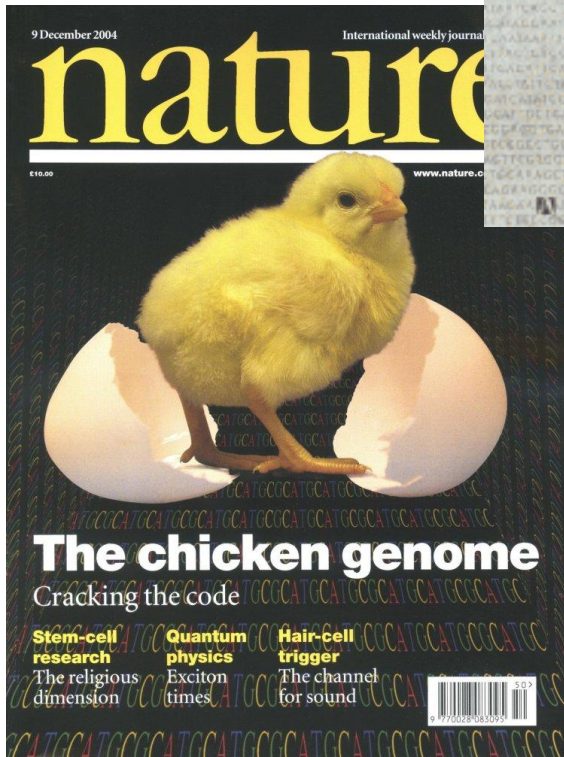
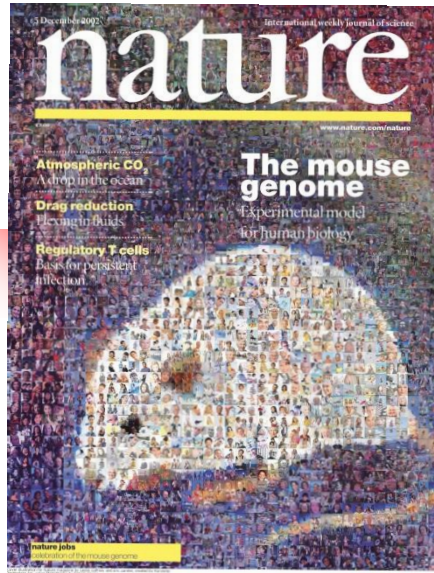
- We are in the midst of a "Golden Era" of biology
- The Human Genome Project has produced a huge storehouse of data that will be used to change every aspect of biological research and medicine
- The revolution is about treating biology as an information science, not about specific biochemical technologies.

# Genome Projects

The Human Genome sequence is complete  
approximately 3.2 billion base



# More Genomes

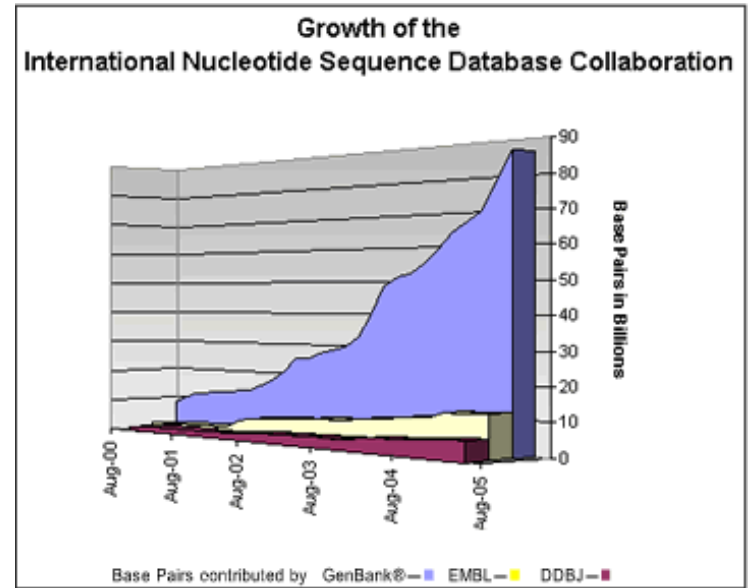




# Bioinformatics Challenges

## The huge dataset

- Lots of new sequences being added
- GenBank has over 100 **Billion** bases and is doubling every year!!
  - problem of exponential growth
  - how can computers keep up?
  - hard drives are cheaper, but processor speeds are not keeping up



# Genomics

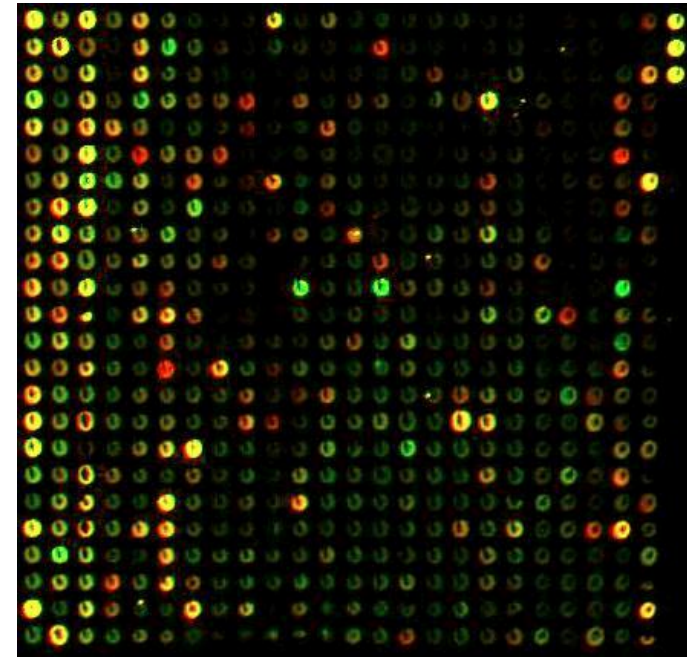


---

- The application of high-throughput automated technologies to molecular biology.
- The experimental study of complete genomes.

# Genomics Technologies

- Automated DNA sequencing
- Automated annotation of sequences
- DNA microarrays
- Protein-protein interactions





# Impact on Bioinformatics

---

- Genomics produces high-throughput, high-quality data, and bioinformatics provides the analysis and interpretation of these massive data sets.
- It is impossible to separate genomics laboratory technologies from the computational tools required for data analysis.





# Aim of research in Bioinformatics

---

Understand the functioning of living things - to “improve the quality of life”.

- drug design
- identification of genetic risk factors
- gene therapy
- genetic modification of food crops and animals, etc.
- application to biotechnology



# Prior knowledge

---

- The course will focus on computing techniques used to analyse biological data
- You do not need to have a biological background to do the course- Specific knowledge required will be discussed in the class.



# Why Bioinformatics?

---

- *“The number of radically new results in pure Computer Science is likely to decrease, while scientists will continue working on biological challenges for the next 500 years”*

**Donald Knuth**

Dan Doenberg, *Computer Literacy Interview with Donald Knuth*,  
Computer Literacy Book, Dec. 1993, Available online at

<http://www.literateprogramming.com/clb93.pdf>



# Evaluation Pattern

---

- Grading: Absolute

(But 'S' and 'A' reserved for truly 'outstanding' and 'excellent' performances!)

- Weightage

- Test 1: 20%
- Test 2: 20%
- Assignments/group project: 20%
- End-Sem Exam: 40%



# Support material

---

## **Course texts and required reading:**

- **An introduction to Bioinformatics Algorithms 4th Ed: Neil James and Pavel A Pevzner, OUPress (2014)**
- **Bioinformatics : Principles and Applications: Zhumur Ghosh, Bibekanand Mallick: OUPress(2015)**
- **Building Bioinformatics Solutions: Concord Bessant, Darren Oakley, Ian Shadforth : OU press(2014)**
- **Computational Molecular Biology-An introduction (1st Ed): Peter Clote and Rolf Backofen, Wiley Series (2000)**



# Support material

---

## **Other texts of interest:**

- Bioinformatics: An Introduction for Computer Scientists, J. Cohen, ACM Computing Surveys, 36(2), 122-158, 2004.
- Bioinformatics- A Beginner's Guide, Jean-Michel Claverie, Cedric Notredame, Publisher: Wiley-Dreamtech India Pvt. Ltd., Year 2003
- Fundamental Concepts of Bioinformatics Krane & Raymer. Publisher: Pearson Education
- Introduction to bioinformatics - Arthur Lesk. Publisher: Oxford University Press. Year 2002
- An Introduction to Bioinformatics- Attwood & Parry-Smith, Publisher: Pearson Education
- Bioinformatics: Sequence and Genome Analysis – David. W. Mount, Publisher: CBS
- An Introduction to Bioinformatics Algorithms – Jones and Pevsner, MIT, Publisher (in India): Ane books



# Useful Websites

---

- <http://www.ncbi.nlm.nih.gov/>  
National Council for Biotechnology Information
- <http://bioinformatics.oxfordjournals.org/>  
Oxford Journal on Bioinformatics (Public access available for selected articles)
- [http://www.cs.brandeis.edu/~jc/bio\\_informatics.html](http://www.cs.brandeis.edu/~jc/bio_informatics.html)  
Introductory paper by Jacques Cohen
- IEEE/ACM Transactions on Computational Biology and Bioinformatics
- SIGKDD and BIODKDD Proceedings

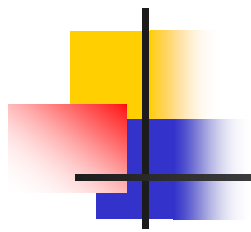


# Reading for the week

---

- [1] Jacques Cohen, *Computer Science and Bioinformatics*, Communications of the ACM, March 2005/Vol. 48 No.3, Pages 73-78
- [3] Jacques Cohen, *Bioinformatics- An Introduction to Computer Scientists*, ACM Computing Surveys, Vol. 96, No.2, June 2004, Pages 122-158





---

Thank You