Introduction

Dr. Fayyaz ul Amir Afsar Minhas

PIEAS Biomedical Informatics Research Lab

Department of Computer and Information Sciences

Pakistan Institute of Engineering & Applied Sciences

PO Nilore, Islamabad, Pakistan

http://faculty.pieas.edu.pk/fayyaz/

Welcome to the Course!

- Dr. Fayyaz-ul-Amir Afsar Minhas
  - Senior Scientist
    - Pakistan Institute of Engineering & Applied Sciences (PIEAS), Islamabad, Pakistan.
  - PhD Computer Science
    - Research: Machine learning in Bioinformatics
    - Colorado State University, Fort Collins, Colorado, USA
    - Supported by the Fulbright scholarship program
  - http://faculty.pieas.edu.pk/fayyaz/
- Course Webpage:
  - https://piazza.com/pieas.edu.pk/spring2016/cis529/resources

Welcome to the course!

- Objectives
  - What is Bioinformatics?
  - Why Bioinformatics?
  - How is computing solving problems from Biology?
  - How to analyze biological data?
  - Understand the working of existing Bioinformatics algorithms
  - Prepare the participants for advanced research level courses in Bioinformatics

Today's agenda

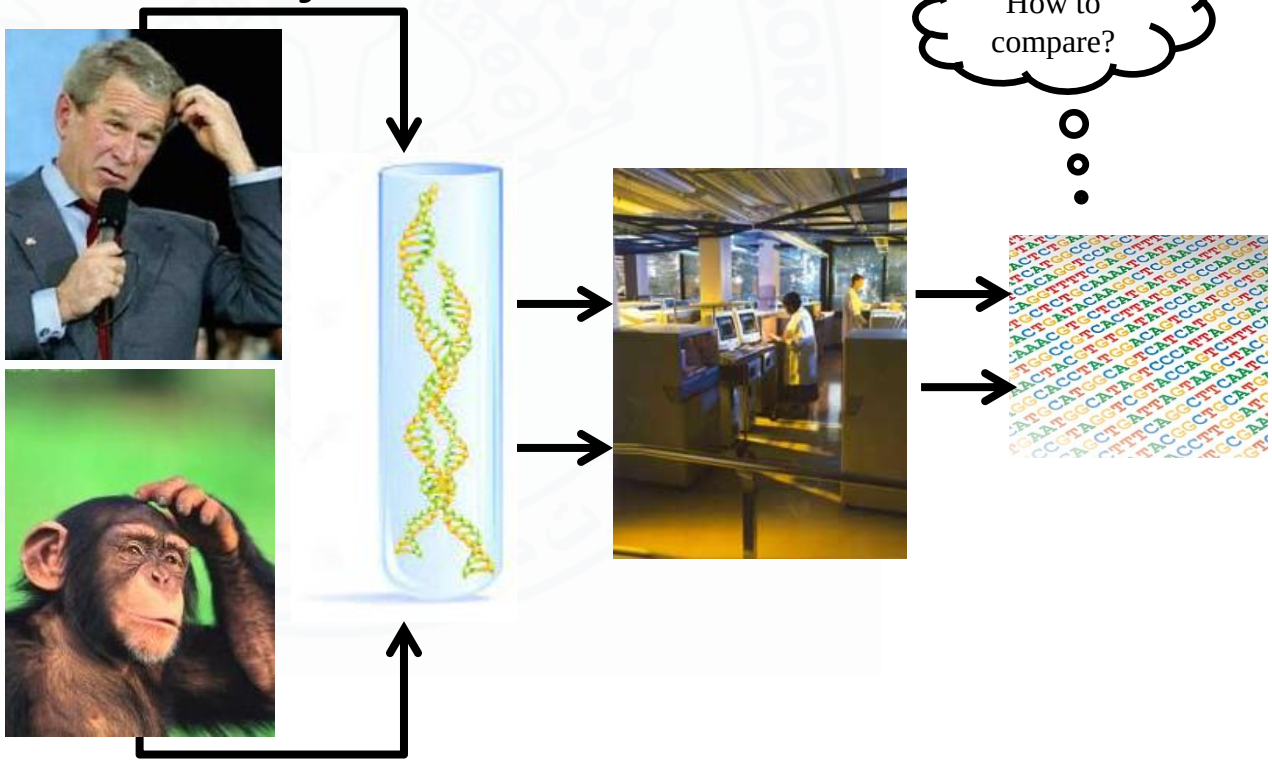
- Why Bioinformatics?
- What is Bioinformatics?
- More motivation
- What is this course really about?
- Who should take this course?
- What can you expect from this course?

What is Bioinformatics?

- Solving problems in Biology using computers

Why Bioinformatics?

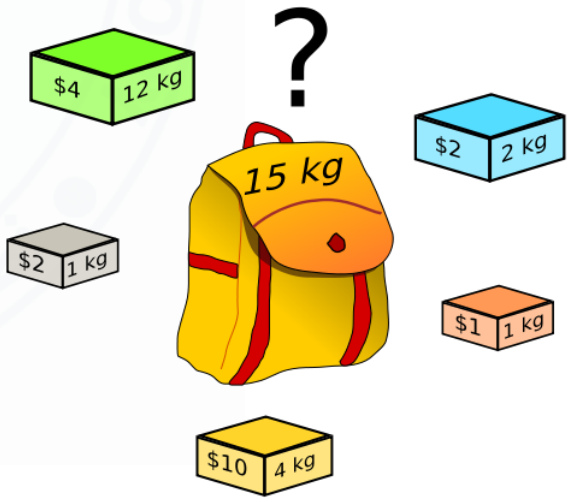
- How do we know that humans and chimpanzees share more than 95% of their DNA?
  - Human Genome Project



Why Bioinformatics?

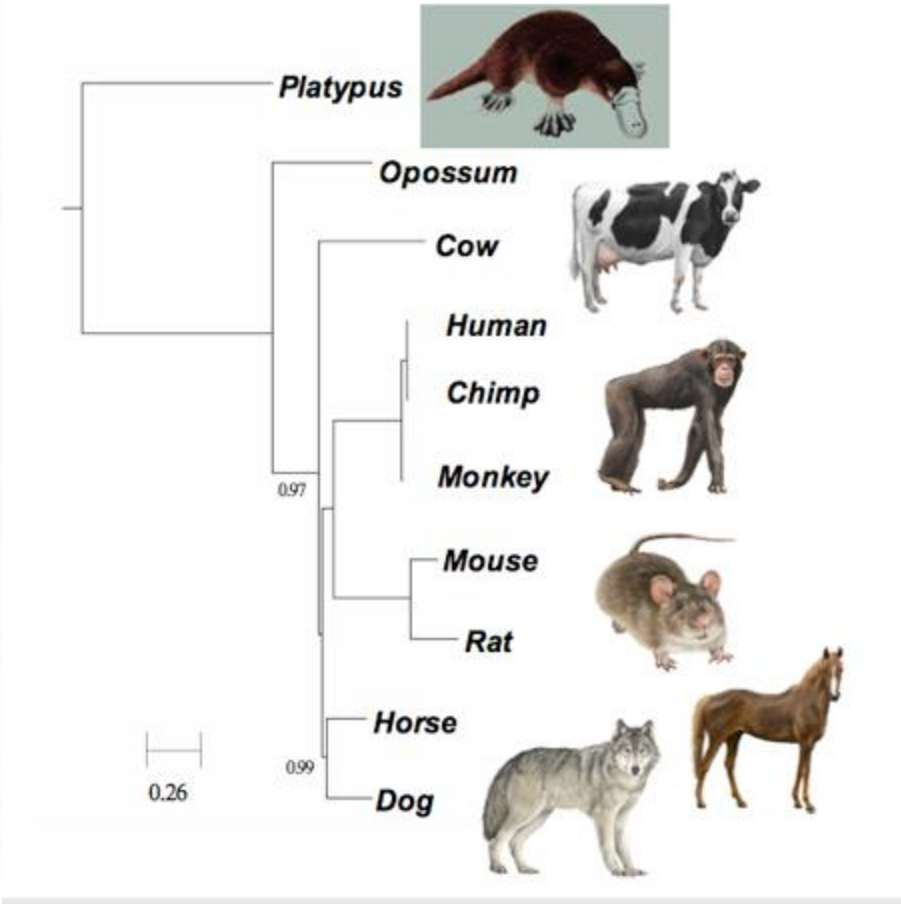
```
AAB24882        TYHMCQFHCRYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881        --------------------YECNQCGKAFAQHSSLKCHYRTHIGEKPYECNQCGKAFSK 40
                                     ****: .***:  * *:** * :****.:* ********..

AAB24882        PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFAQ- 116
AAB24881        HSHLQCHKRTHTGEKPYECNQCGKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
                 **** *:***********:***:**.: .****************    :  *.: :
```

- The knapsack problem
  - Uses dynamic programming



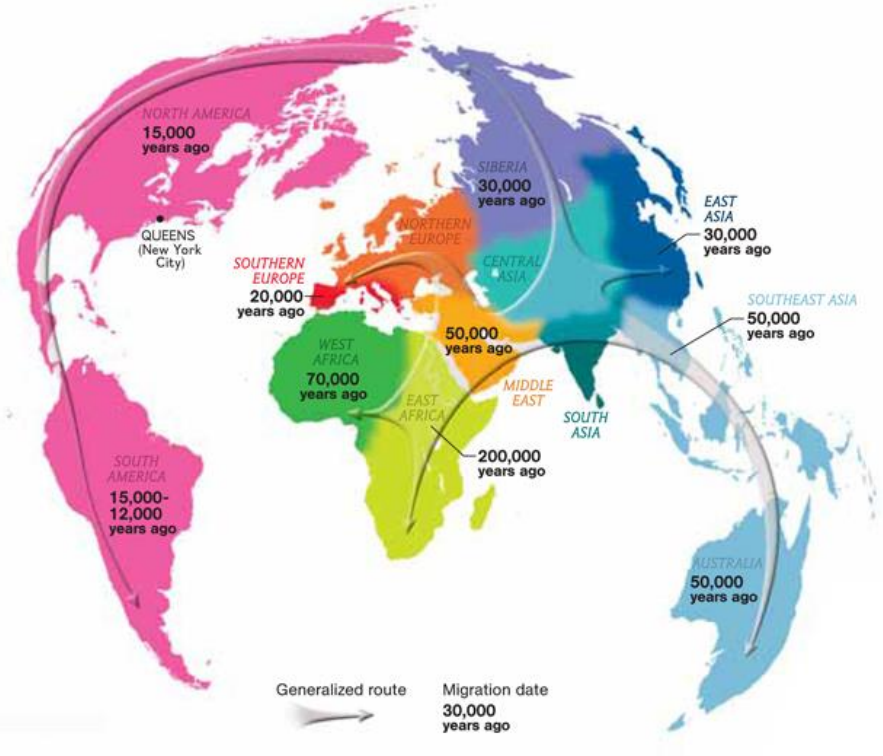
Why Bioinformatics?

- Tree of life



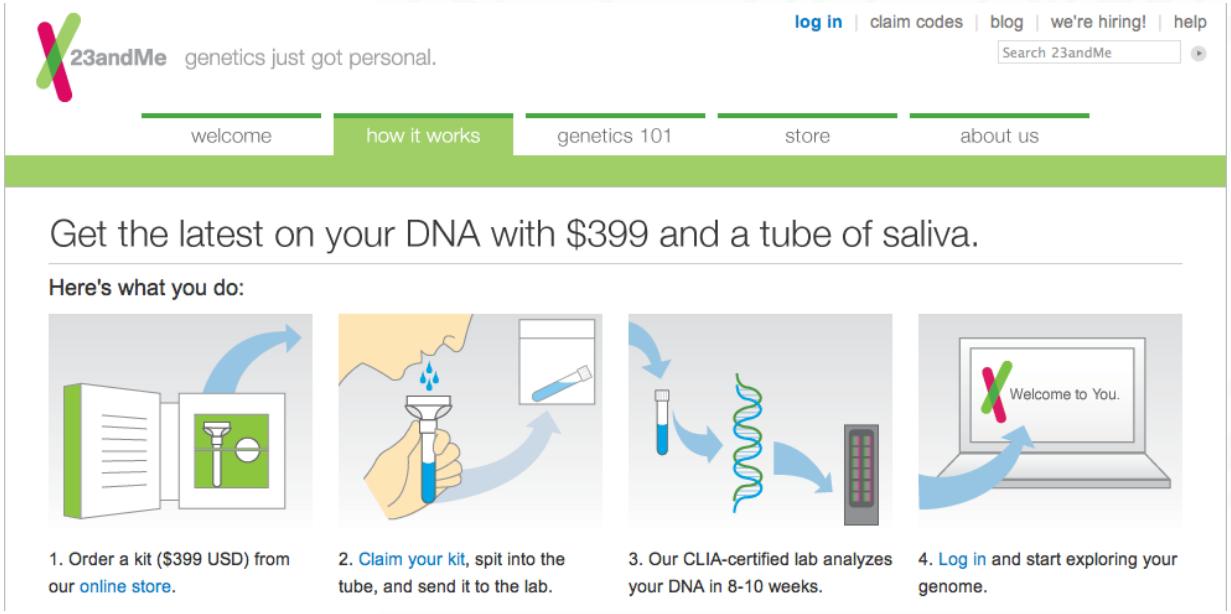
Why Bioinformatics?

- How are humans across the Earth related to each other?
  - Human Genographic project



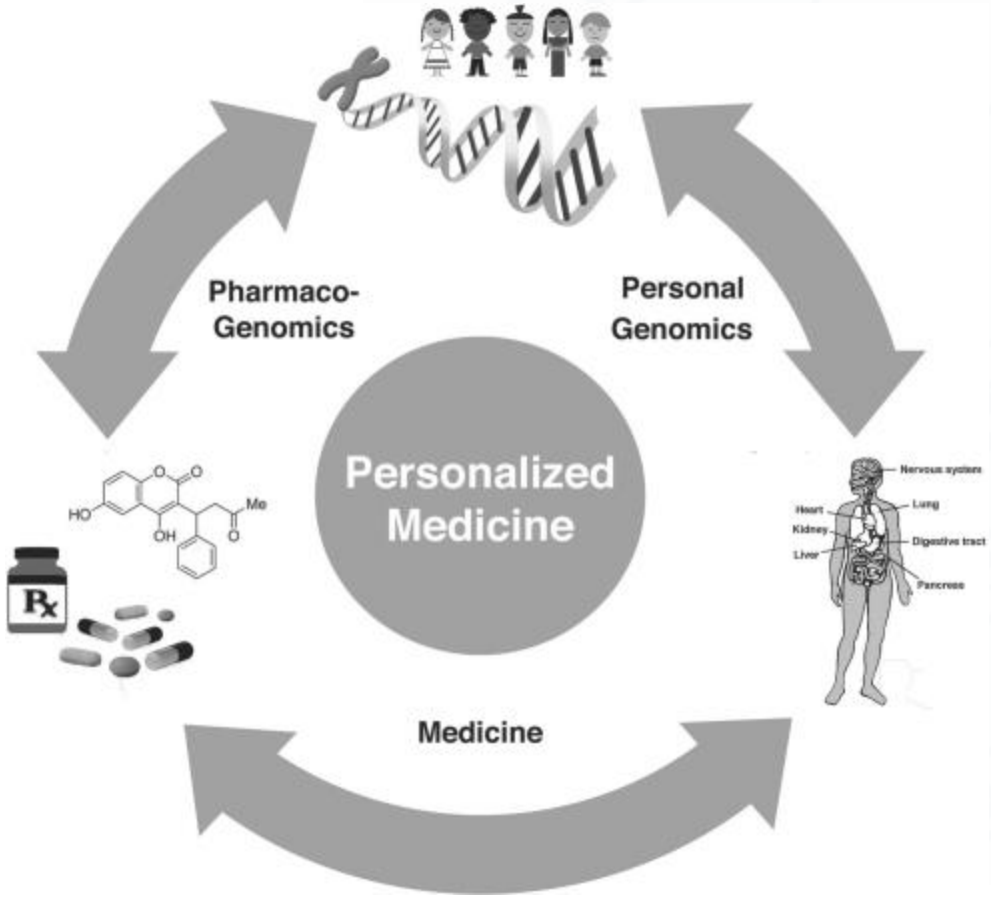
Why Bioinformatics?

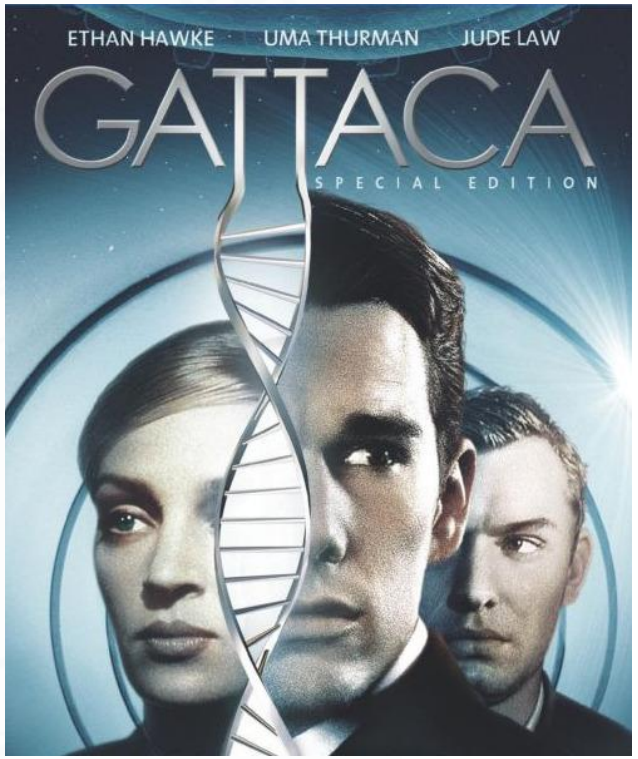
- How can we screen for disease?



Why Bioinformatics?

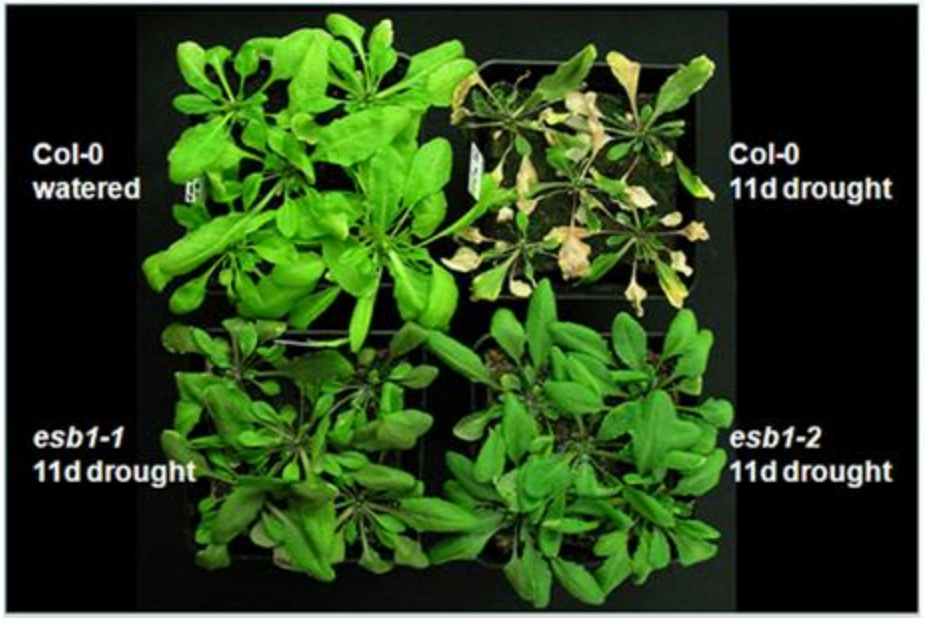
- Personalized medicine





Why Bioinformatics?

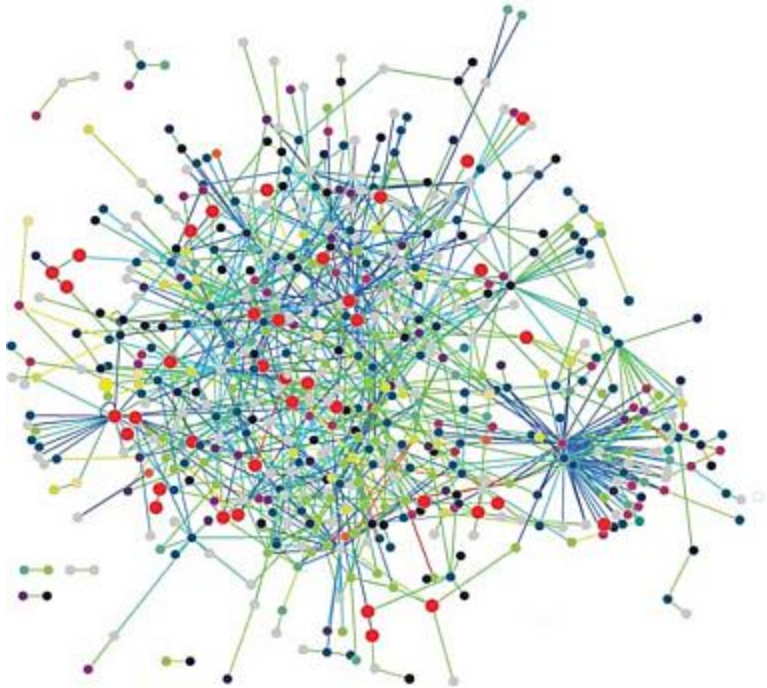
- How can we make better drought resistant crops?

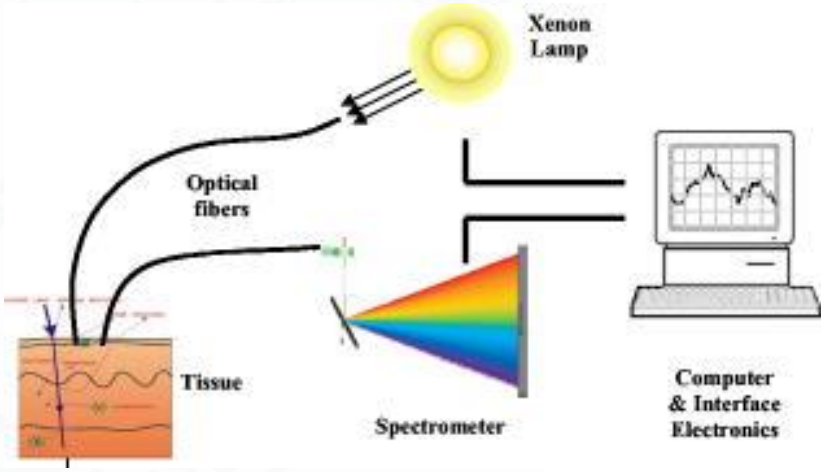


Suberin goes genomics: use of a short living plant to investigate a long lasting polymer

Why Bioinformatics?

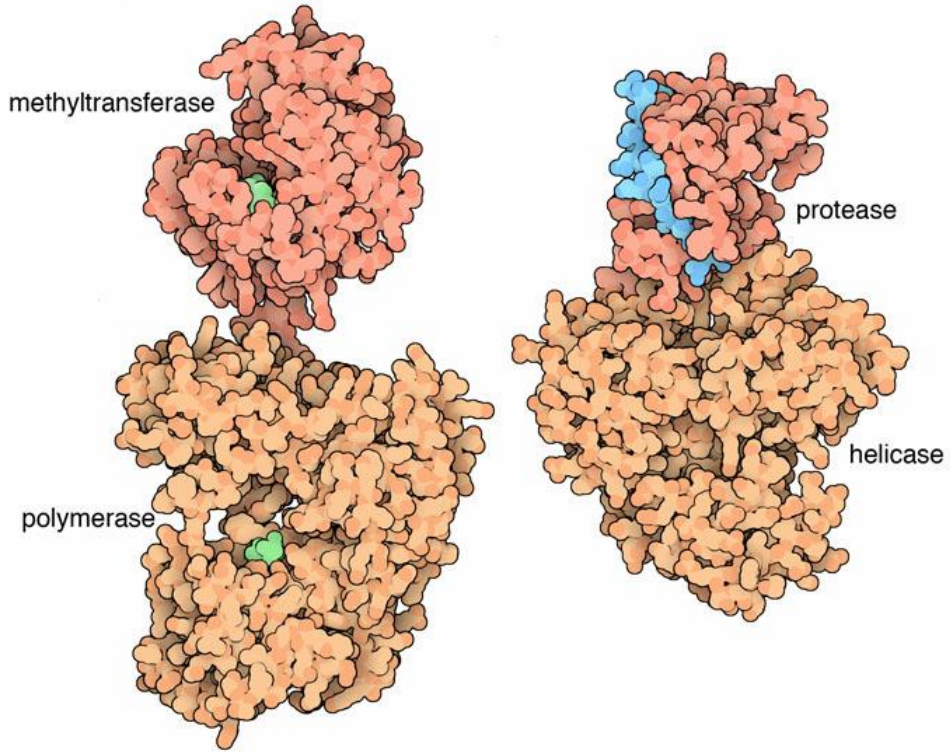
- How can we fight against diseases like Cancer?



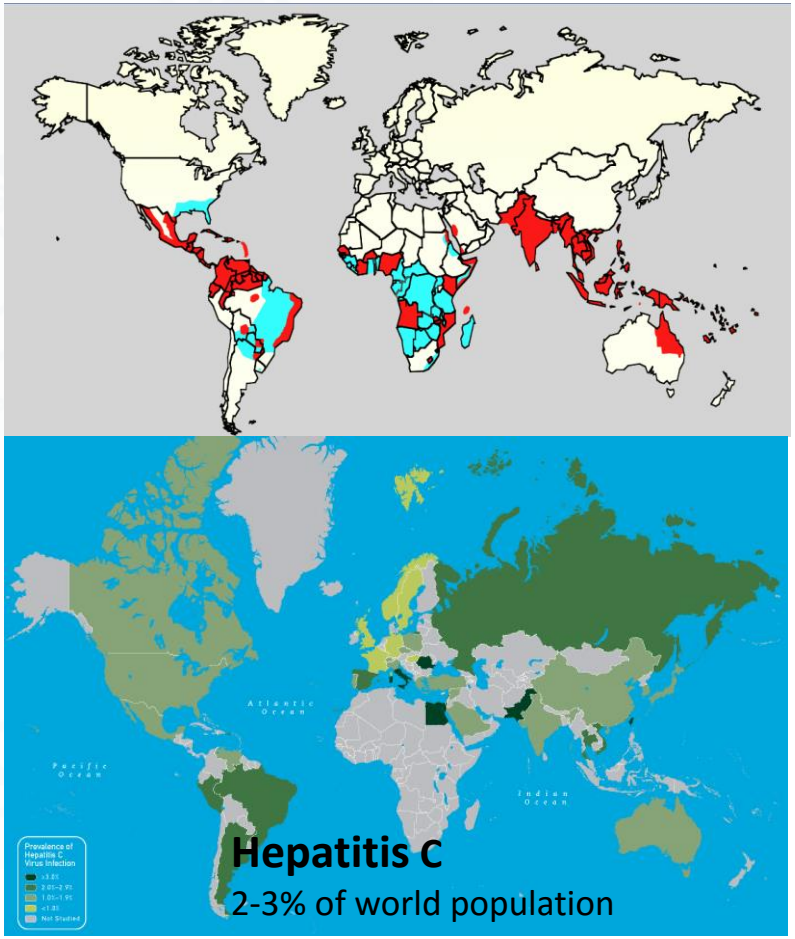


Why Bioinformatics?

- How can we combat viruses?



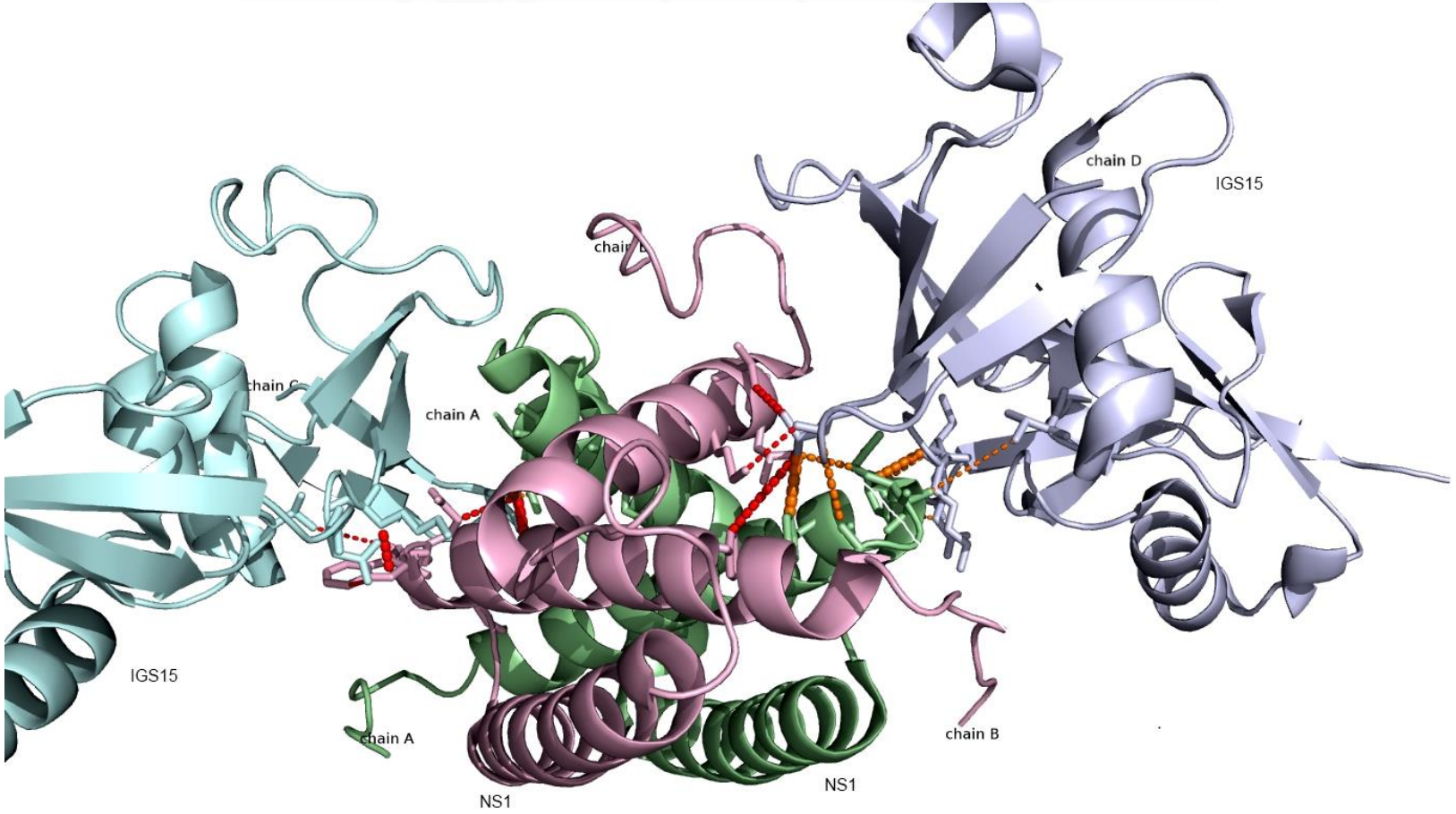
Dengue Fever
Infects: 50 to 528 M/yr
Kills: 25 K/yr



Why Bioinformatics?

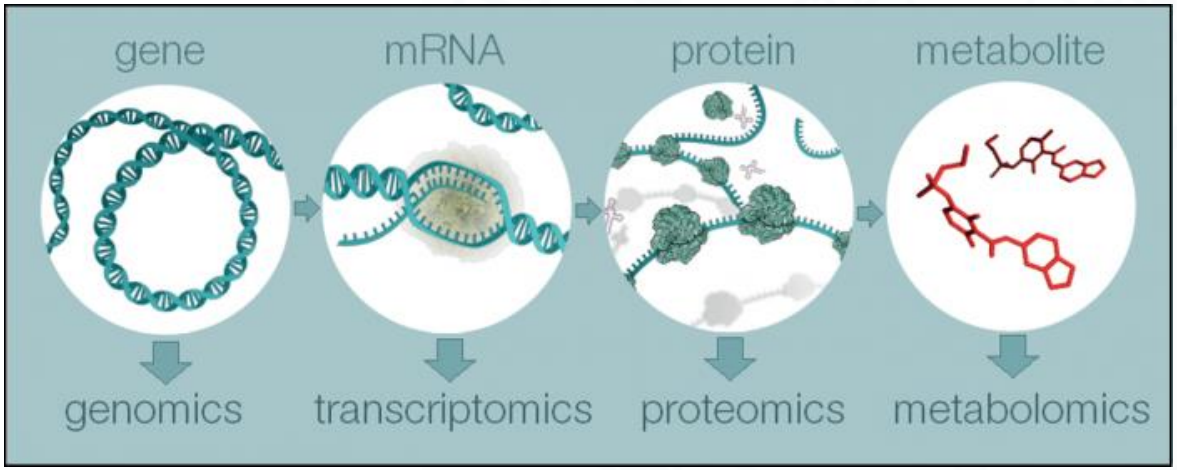
Why Bioinformatics?

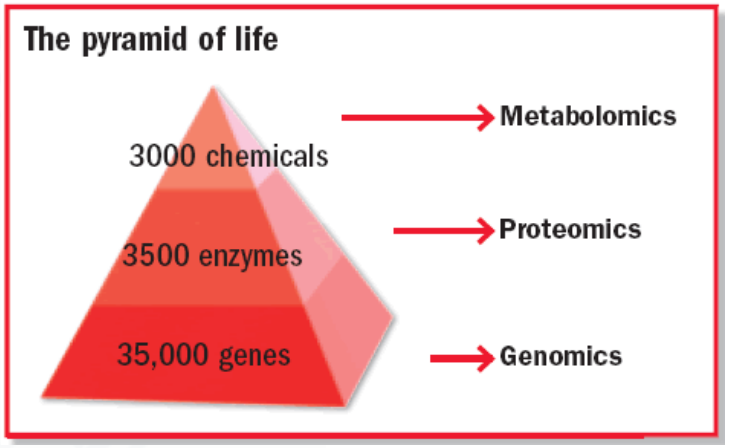
- Understanding viruses with computers



Why Bioinformatics?

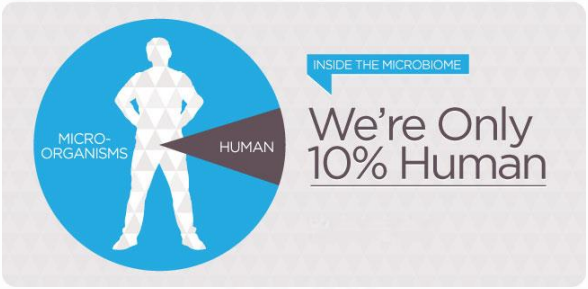
- How can we find out what are the effects of a certain disease?





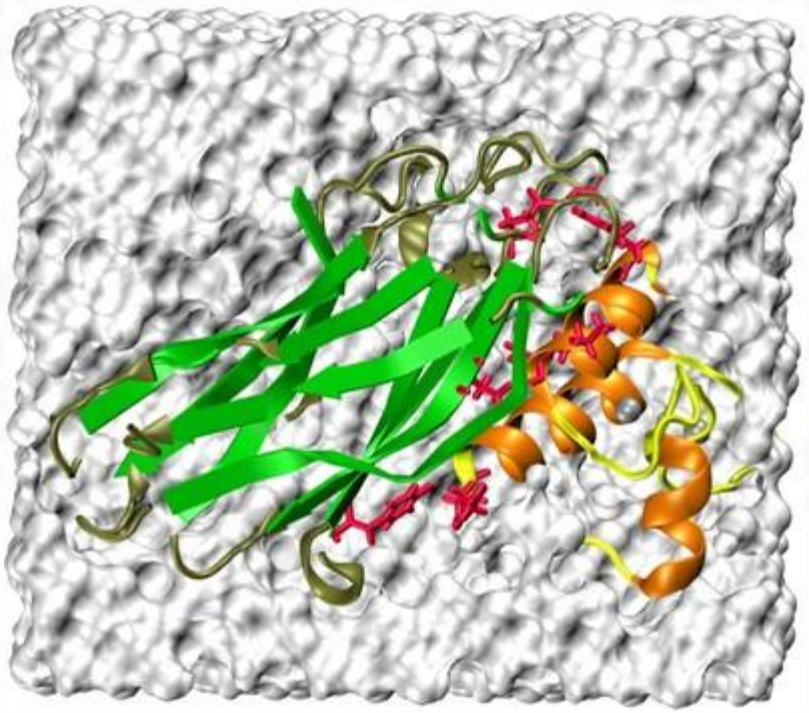
Why study Bioinformatics?

- How is life linked? Is there symbiosis?
  - Human Microbiome project
  - Metagenomics



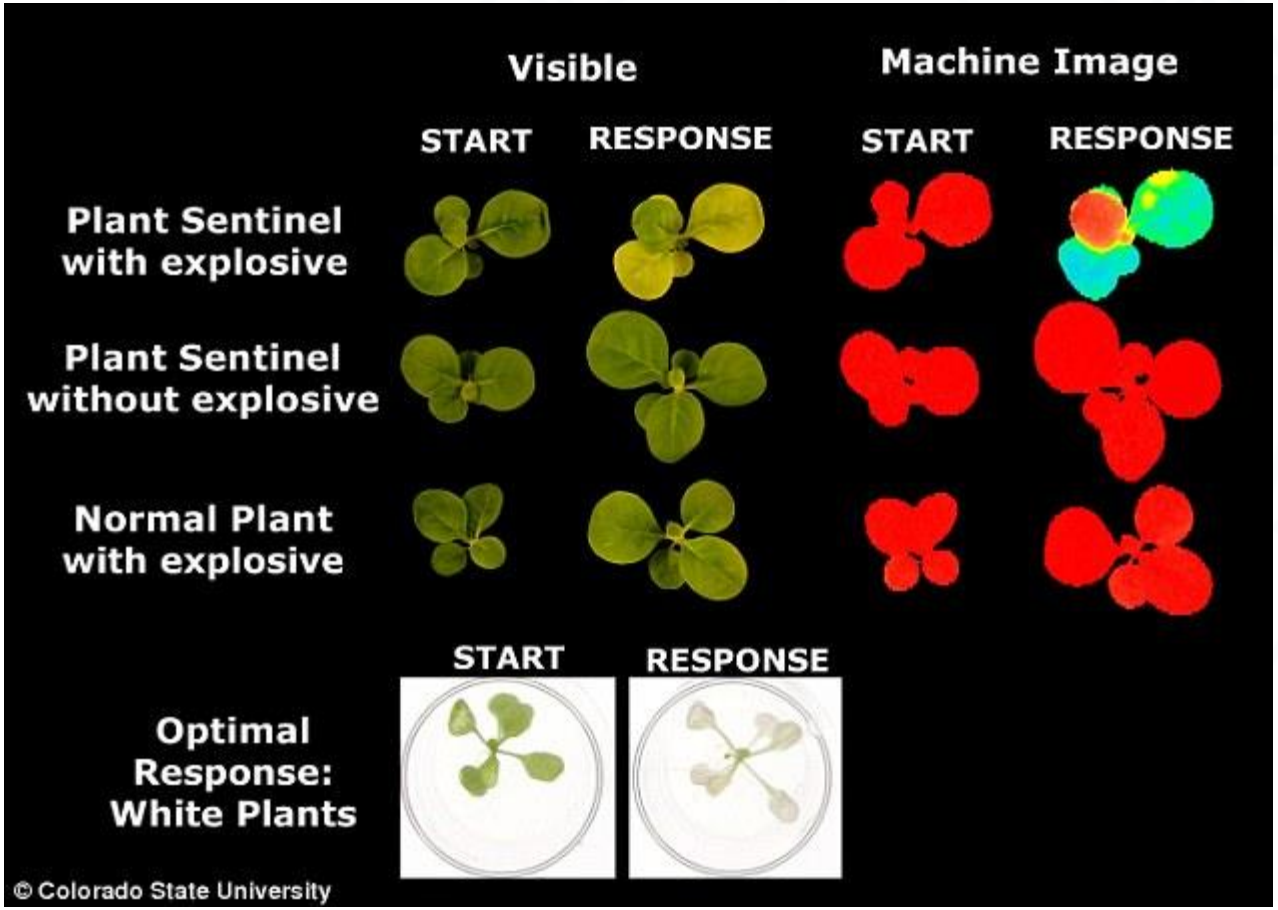
Why Bioinformatics?

- How can we make more efficient/greener fuels?



Why Bioinformatics?

- How can we combat terrorism?



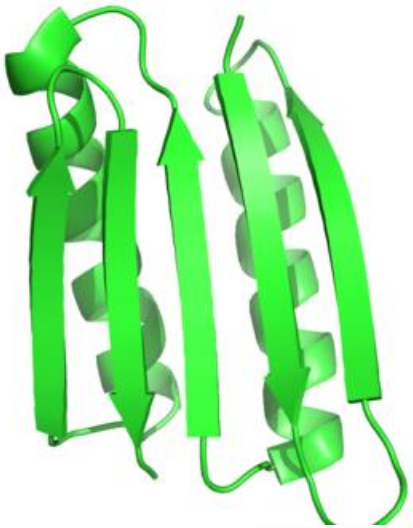
Dr. Jane Medford

Why Bioinformatics?

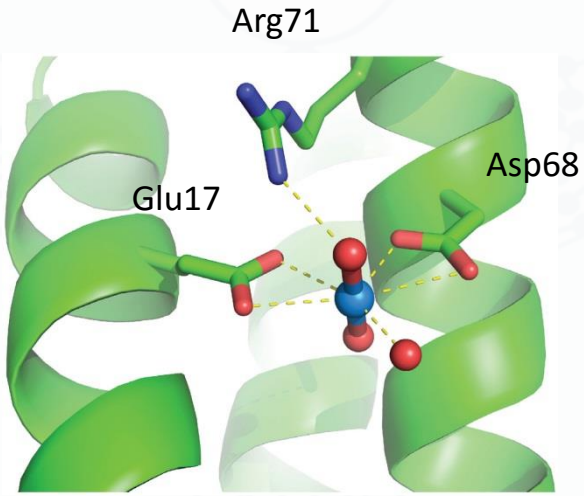
- How can we combat terrorism?

Why Bioinformatics?

- How can we design a molecular motor or the world's 'smallest' flag or biomaterials?

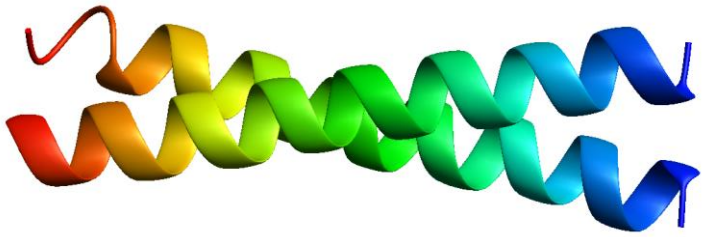


TOP7: Fold does not exist in nature



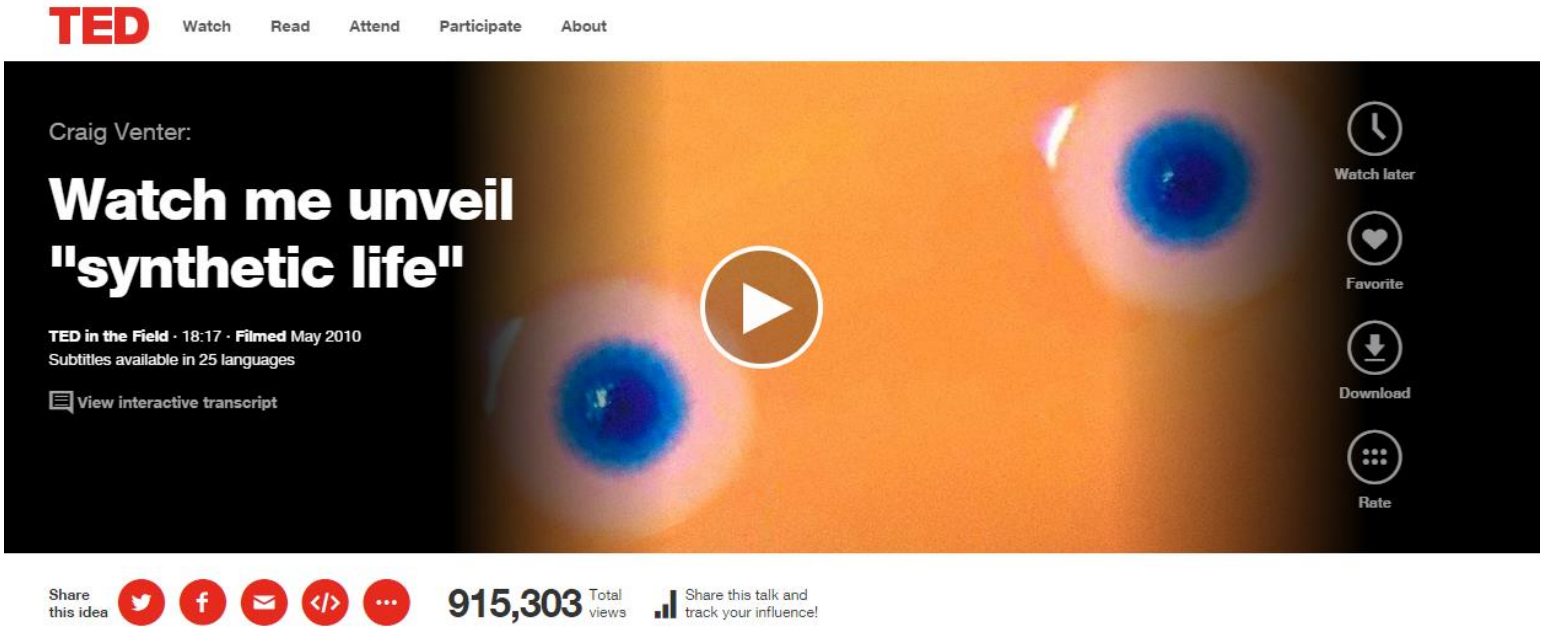
Zhou, Lu, Mike Bosscher, Changsheng Zhang, Salih Özçubukçu, Liang Zhang, Wen Zhang, Charles J. Li, et al. 2014. "A Protein Engineered to Bind Uranyl Selectively and with Femtomolar Affinity." *Nature Chemistry* 6 (3): 236–41.

4.5 billion metric tons of uranium, diluted down to a minuscule 3.3 parts per billion



Why Bioinformatics?

- How can we design new life?



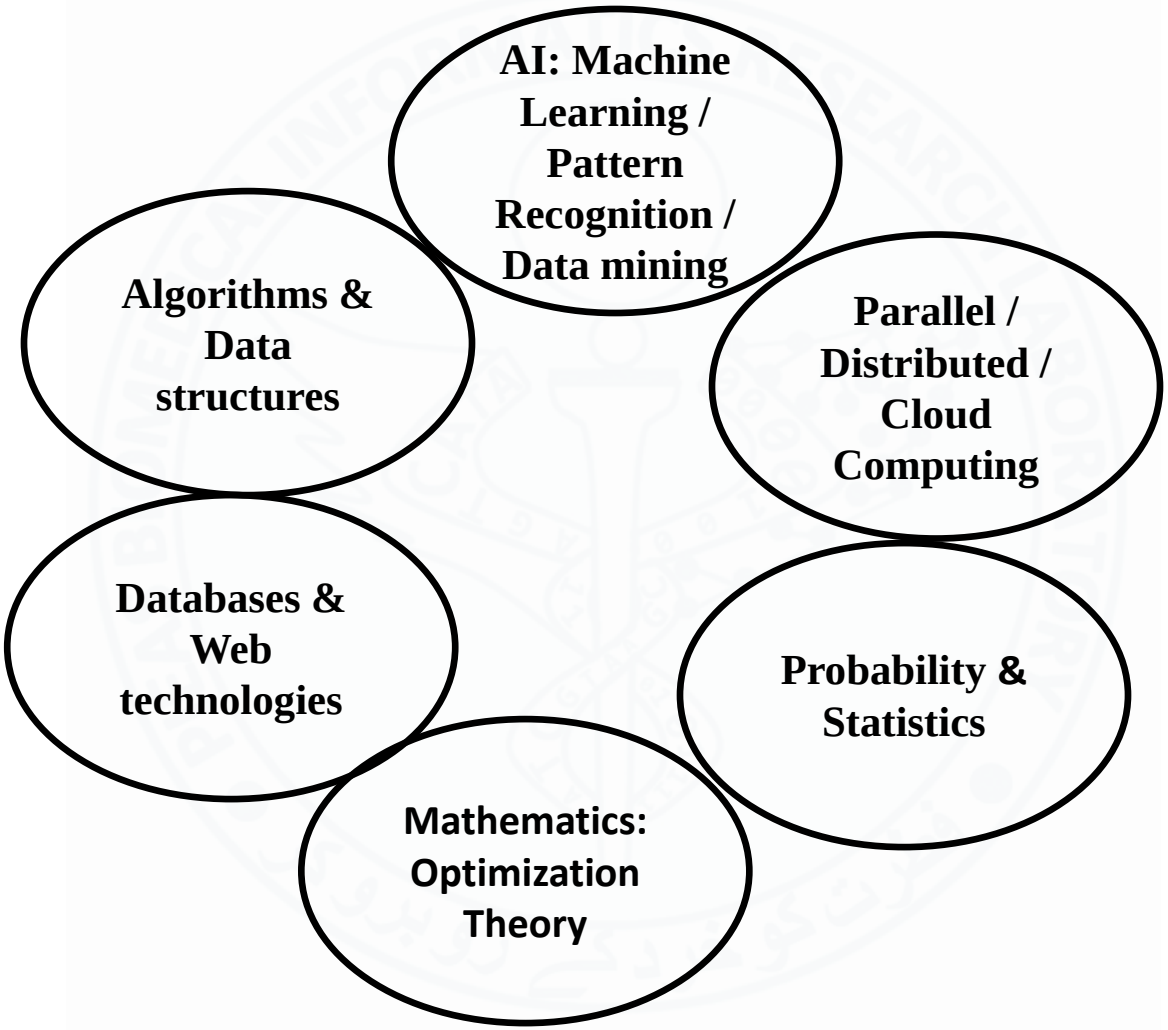
https://www.ted.com/talks/craig_venter_unveils_synthetic_life

Why computer scientists should study it?

- Biology easily has 500 years of exciting problems to work on
- Most of Bioinformatics is younger than me
- Its cross-disciplinary
- Raises interesting computational problems
- Its full of very interesting machine learning/data mining problems
- Global impact
- Funding & Career opportunities

Knuth

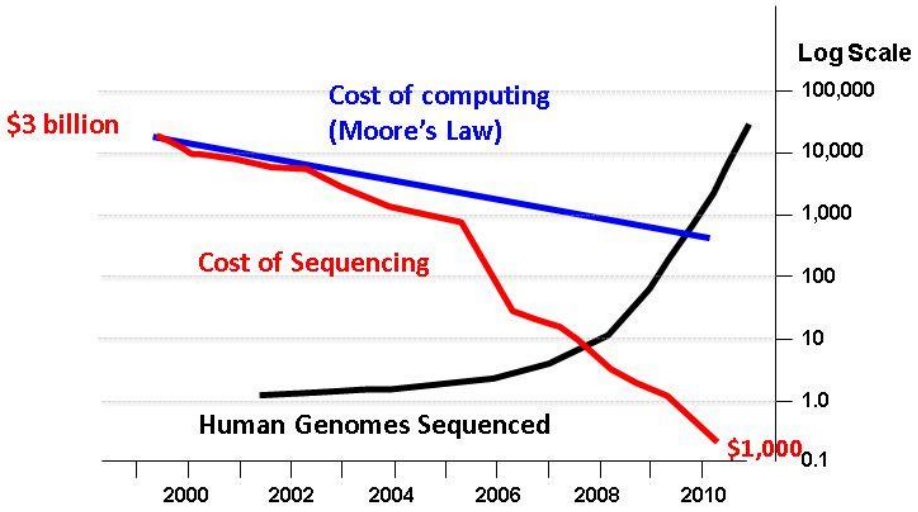
Why computer scientists should study it?

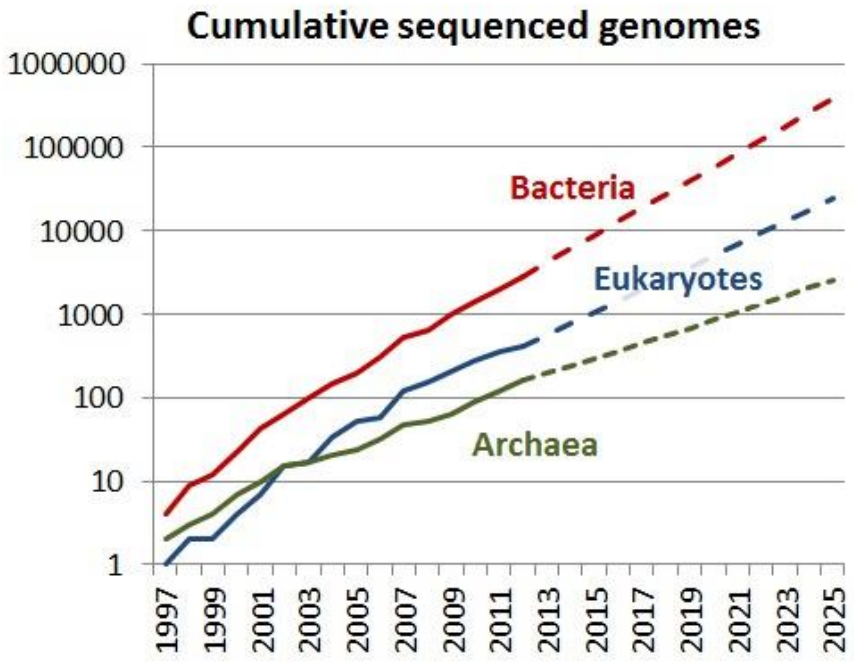


Why Bioinformatics? The Economics

Adapted from The Economist

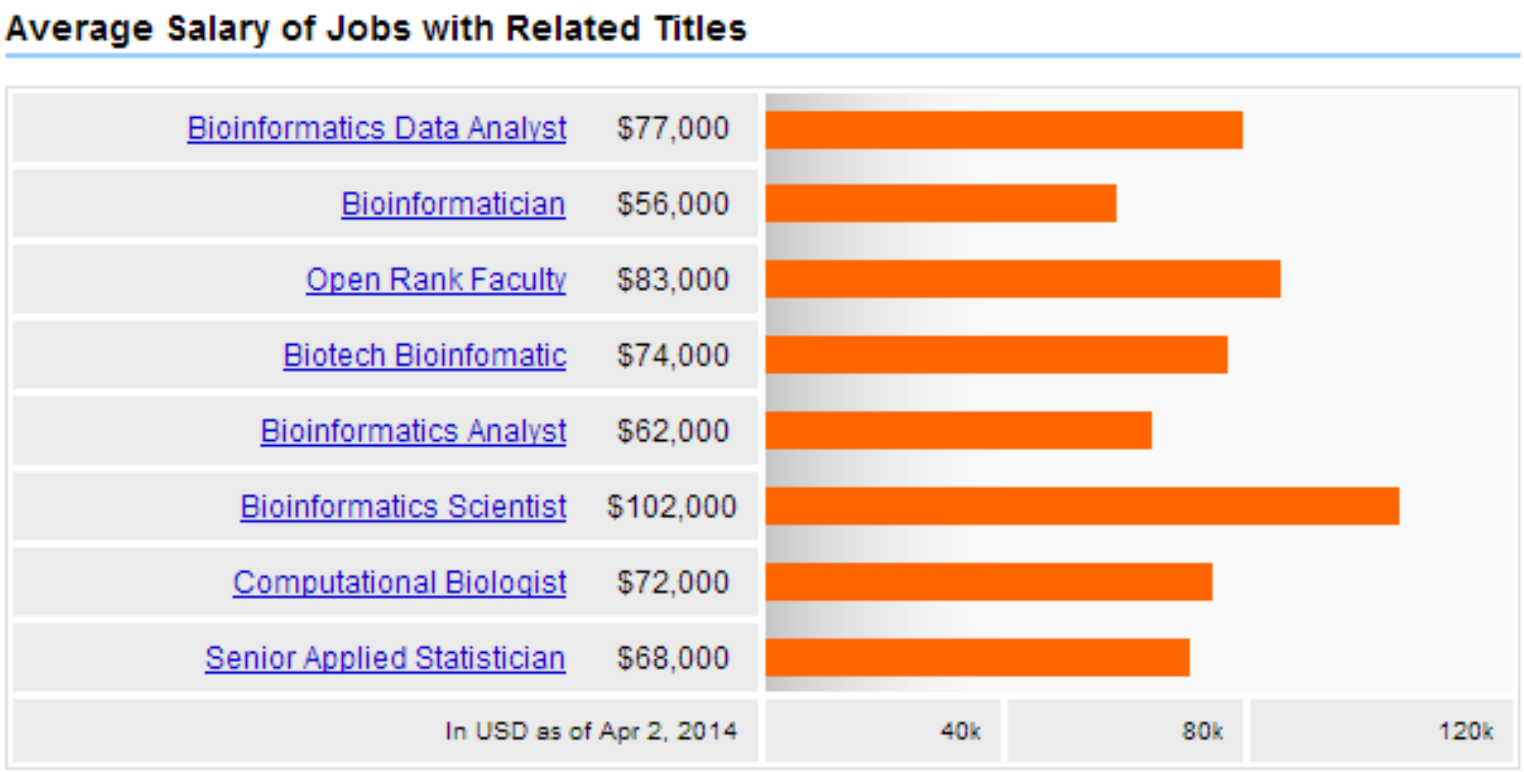
The Sequencing Explosion





http://sulab.org/2013/06/sequenced-genomes-per-year/

Why Bioinformatics? The Economics



Bioinformatics pays? http://www.homolog.us/blogs/blog/2014/04/02/what-is-really-the-salary-of-a-bioinformaticiancomputational-biologist/

Bioinformatics in Pakistan

- It's still nascent
- Plenty of room
- Will give you applied skills to handle any computer science jobs
- And there are jobs!

Who should take this course?

- Computer scientists graduate and senior graduates who:
  - Know high level language programming
  - Know algorithms & data structures
  - Are curious to learn bioinformatics!
  - Are looking to some challenging computing research problems
- Biologists who:
  - have good computational (algorithmic & programming) skills
  - Are looking to expand their breadth

Back to course

- Course contents
  - Week 1: Assignment 1 of Python Programming
    - Introduction, Biology Primer, Python Beginning
  - Week 2-6: Assignments 2,3 - Programming Alignments with Biopython and UGENE, Sessional -1
    - Sequence Alignments
  - Week 7-8: Assignment 4 – Generating MSAs and Trees (UGENE)
    - Multiple Sequence Alignments, Profiles and Phylogenetic trees
  - Week 9-11: Sessional - 2
    - Sequencing and Assembly
  - Week 12-13
    - Proteins and Protein Structures
  - Week 14-16: Assignment 5
    - Structural Alignments
    - Structure Prediction

Course Evaluation

- Quizzes/Reading Assignments: 5%
- Assignments: 25% (tentatively 5 Assignments)
  - Programming/software based
- Sessional-I: 10%
- Sessional-II: 10%

- Final Exam: 50%

- BONUS: class questions and http://rosalind.info/

- Effort Required: *Avg.* 2 hours per 1 class hour

Logistics

- Course Webpage
  - Piazza
  - https://piazza.com/pieas.edu.pk/spring2016/cis529/resources
  - Please register, signup or send me an email so
    - afsar at pieas dot edu dot pk
  - Please use a single email address for all interaction and be sure to check it daily for updates
- Office Hours
  - (Free Tuition Time!) Fridays 0930-1030 in B-216
  - By Email Appointment
- Attendance: PIEAS Policy

Resources

- Books
  - \\172.30.10.2\FacultyShare\Fayyaz ul Amir Afsar Dr\CIS529 Bioinformatics
- Python Help
  - \\172.30.10.2\FacultyShare\Fayyaz ul Amir Afsar Dr\PYTHON
- Online Help
  - Scipy: http://www.scipy-lectures.org/
  - Biopython: http://biopython.org/wiki/Main_Page
- UGENE: http://ugene.net/
- PyMOL

Self-Learning Requirements

- Python
  - Install Anaconda Python Distribution in Windows or Spyder in Ubuntu Linux
  - What you need to understand:
    - Installation and Administration
      - Using package manager (pip, easy_install, conda)
    - Basic Constructs: Variables, Control Flow, Object Oriented concepts, Mutable and Immutable Types, Lists and Dictionaries
    - Using Scipy (Matplotlib for plotting and Numpy)
    - Debugging (pdb)
- Reading pointers will be given
  - Can discuss issues in the office hours

TO DO

- Required Reading
  - Cohen, Jacques. “Bioinformatics: An Introduction for Computer Scientists.” *ACM Comput. Surv.* 36, no. 2 (June 2004): 122–58. doi:10.1145/1031120.1031122.
  - https://en.wikipedia.org/wiki/Bioinformatics

- Optional Reading
  - Jones and Pevzner 2004 – Chapter 1 (required if you do not have a background in programming)

End of Lecture-1

We want to make a machine that will be proud of us.

- Danny Hillis