

# **Bioinformatics**

## **CS300**

### **Introduction**

**Spring 2021**  
**Oliver BONHAM-CARTER**

# Introduction to Bioinformatics

CMPSC\* 300

Spring 2021

Class: M, W, F 10:40am - 11:40am

Lab: Monday 2:50-4:40pm

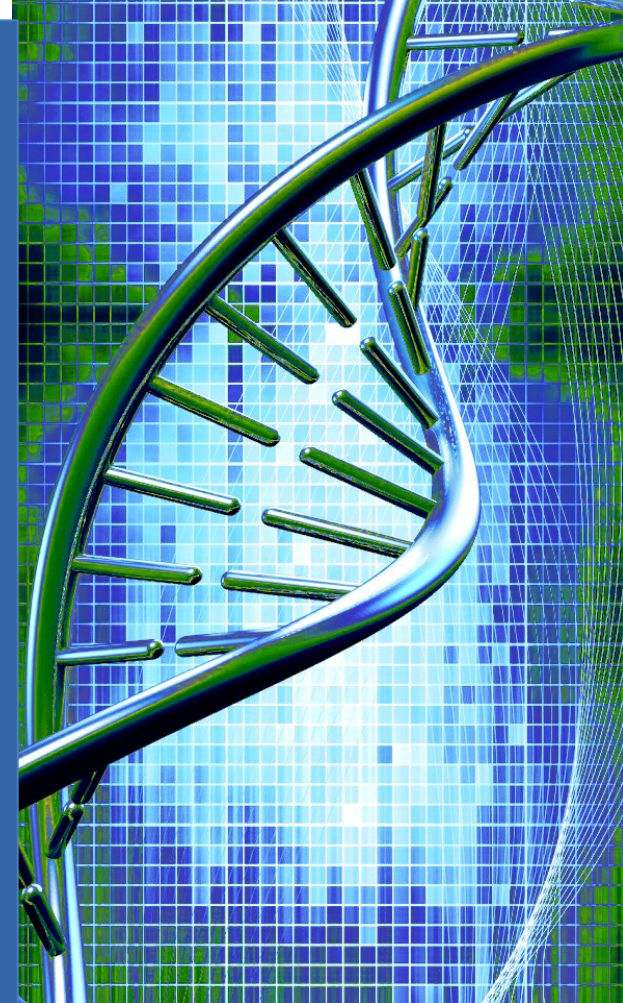
(Alden Hall 109)

Bioinformatics is an emerging, rapidly expanding interdisciplinary discipline that studies how to effectively integrate concepts from computational sciences and biosciences. There is a high demand for scientists who are versed in both biology and computer science in the biomedical industry and research.

BIO 300/CMPSC 300 students will become familiar with the state-of-the-art bioinformatics software and the algorithms behind them. Through hands-on projects, students will explore current biological problems and will develop bioinformatics solutions to these issues.

This course counts as an Area A Biology Course and an Applications Course for Computer Science.

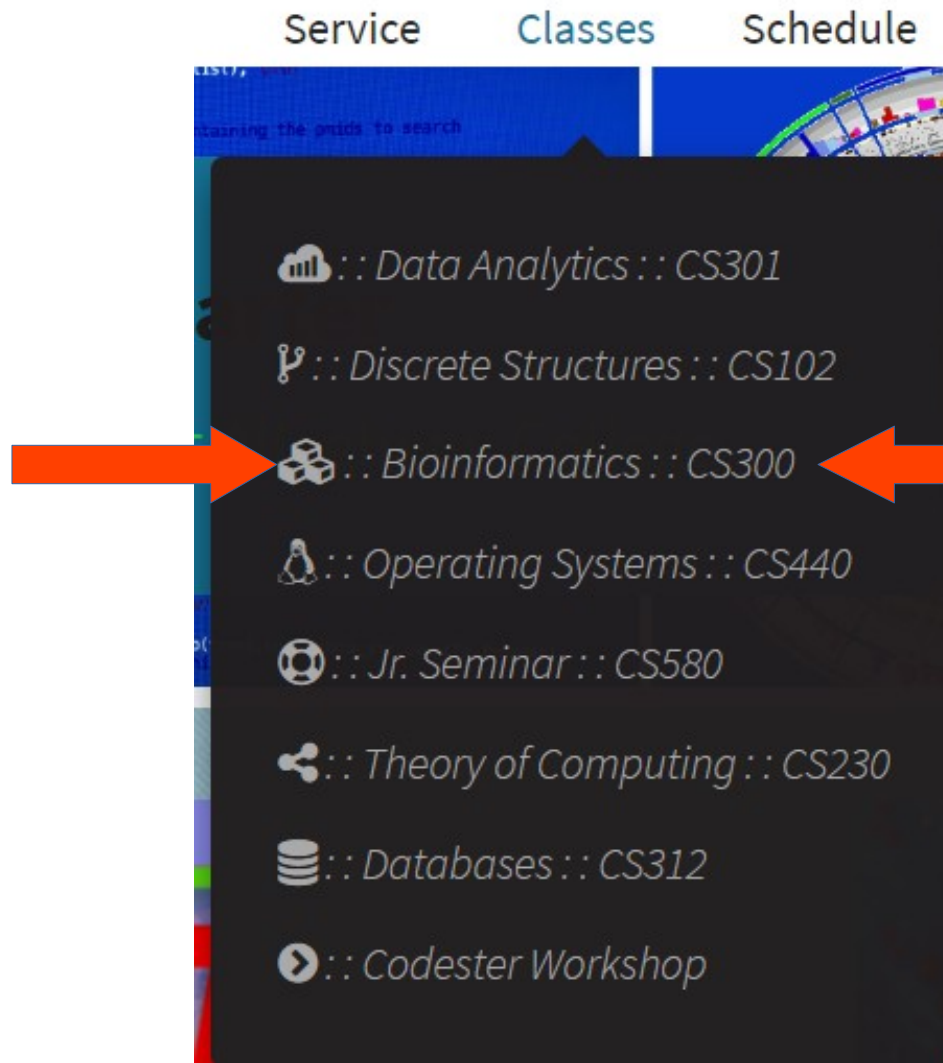
Questions? Contact Dr. BONHAM-CARTER at  
[obonhamcarter@allegheny.edu](mailto:obonhamcarter@allegheny.edu)



	120		130											
L	R	I	C	R	V	L	R	S	L	K	L	L	A	Q
L	R	I	C	R	V	L	R	S	L	K	L	L	A	Q
L	R	I	C	R	V	L	R	S	L	K	L	L	A	Q
L	R	V	C	R	V	L	R	S	L	K	L	L	F	A
T	Y	P	L	R	V	L	R	L	V	H	V	C	M	A
N	Y	T	L	R	A	L	R	L	V	H	V	C	M	A
A	D	G	I	O	S	L	R	I	L	K	L	I	S	Y
A	D	G	I	O	S	L	R	I	L	K	L	I	S	Y
A	D	G	M	O	S	L	R	I	L	K	L	I	G	Y
L	R	A	L	R	A	I	R	V	L	R	L	S	F	
M	R	A	L	R	A	I	R	V	L	R	L	S		



# Class Website



Find:

- Syllabus
- Meeting place and times
  - Class
  - Lab
- Calander
- First slides
- Etc.

<https://www.cs.allegheny.edu/sites/obonhamcarter/>



ALLEGHENY  
COLLEGE

# How To Be Successful

- Participate to classes and lab!!!
- Read the book and handouts!
- Keep up with your lab homework!







# An Ethical Approach

Responsible Computer Science Challenge

*With Great Code Comes  
Great Responsibility*

ResponsibleCS.org

#ResponsibleCS

a partnership of



OMIDYAR NETWORK



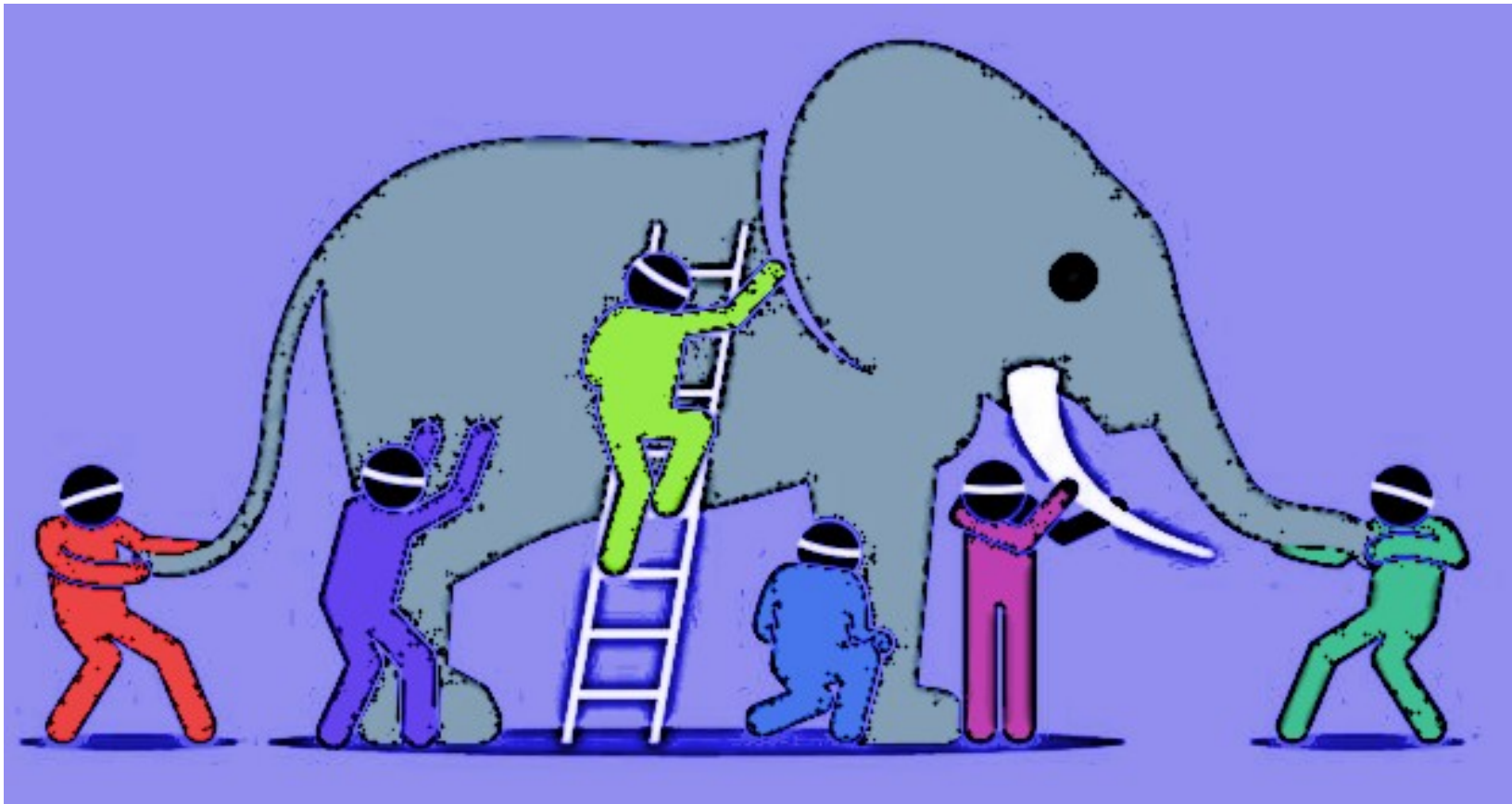
SCHMIDT FUTURES

Craig Newmark Philanthropies

- We will spend time to learn about responsibly in Bioinformatics and in its research
- Ethical considerations for research, building tools, implementing algorithms, working with life, etc ...

# What is this *Bioinformatics*?

- This field has many meanings... depending on the type of research





# Bioinformatics is Many Things...

The science of collecting and analyzing complex biological data such as genetic codes.

A theoretical framework to detect genes which contribute to the onset of unhealthy development.

The field of exploration into data to describe the onset of illness, disease and medical disorder

Is an interdisciplinary field that develops methods and software tools for understanding biological data.

Bioinformatics is both an umbrella term for the body of biological studies that use computer programming as part of their methodology

The study of data from living systems to determine patterns of life and health.

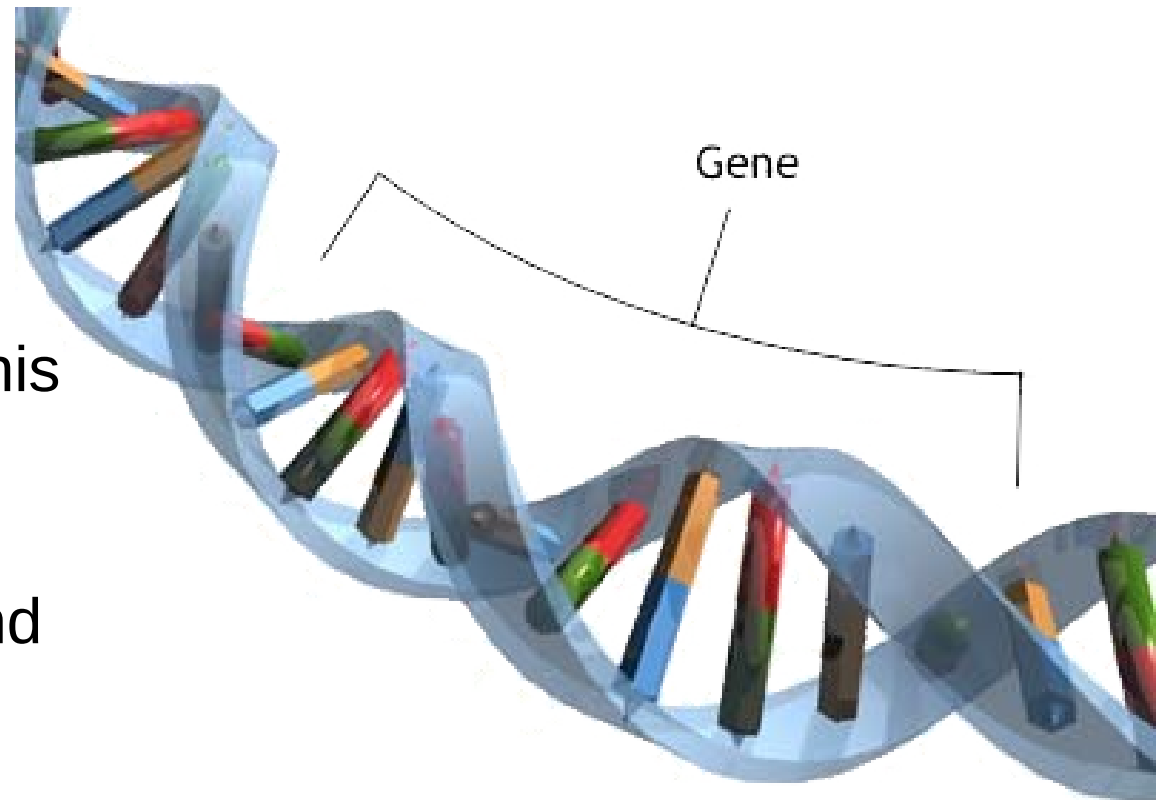
The development of tools to aid in the comparison of genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology.

A framework used to determine the relatedness between people, dogs, cats, mice, rats, rabbits, or any living thing!

# Bioinformatics:

## The Study of (*all things*) DNA

- DNA is the genetic material that houses genetic information.
- Genes are written in this language
- Understanding DNA allows us to understand how genes work.



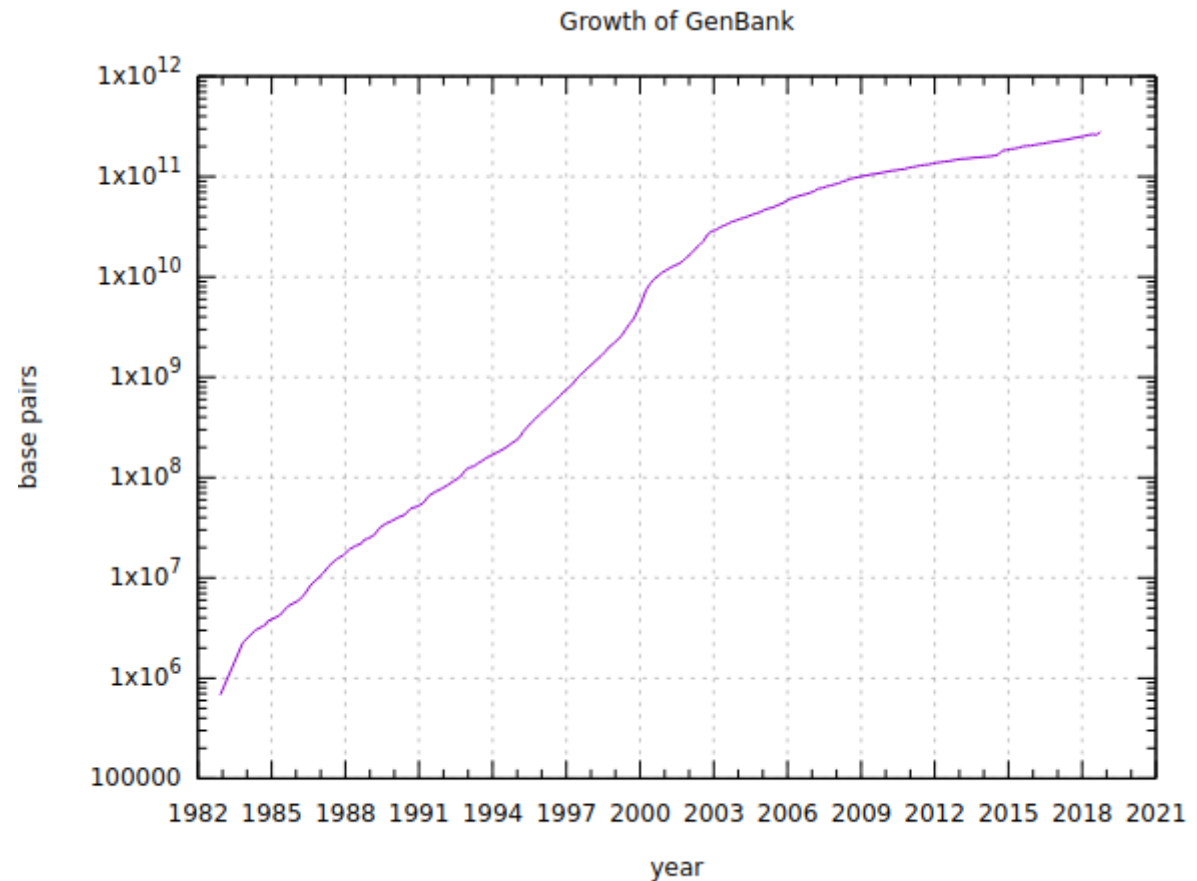
```
1 attaaagggtt tataccttcc caggtaacaa accaaccaac ttctgatctc ttgtagatct
61 gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcatgc ttagtgcact
121 cacgcagtat aattaataac taattactgt cgttgacagg acacgagtaa ctctgtctatc
181 ttctgcaggc tgcttacggt ttctgtccgtg ttgcagccga tcatcagcac atctagggttt
241 cgtccgggtg tgaccgaaag gtaagatgga gagccttgtc cctgggtttca acgagaaaaac
```





# How Much DNA to Study?

- Gen Bank: Public repository of DNA sequences and related data.
- Seemingly exponential growth in amount of sequence data available for study



As of release 228, retrieved 2018-11-30

# An Important DNA Study That I Might Know??

- The Human Genome Project
- Large scale project to determine the genetic sequence of human DNA.
- 20<sup>th</sup> anniversary of landmark HGP publications



# Another Important DNA Study ??

- Washington Post Article: Scientists extract from Siberian mammoth molars the **oldest DNA** ever recovered

Mammoth  
Molar  
Munchers!



Bringing them back from **extinction**?

# Use DNA to Compare...

- DNA sequences
- Genes
- Proteins
- Organisms

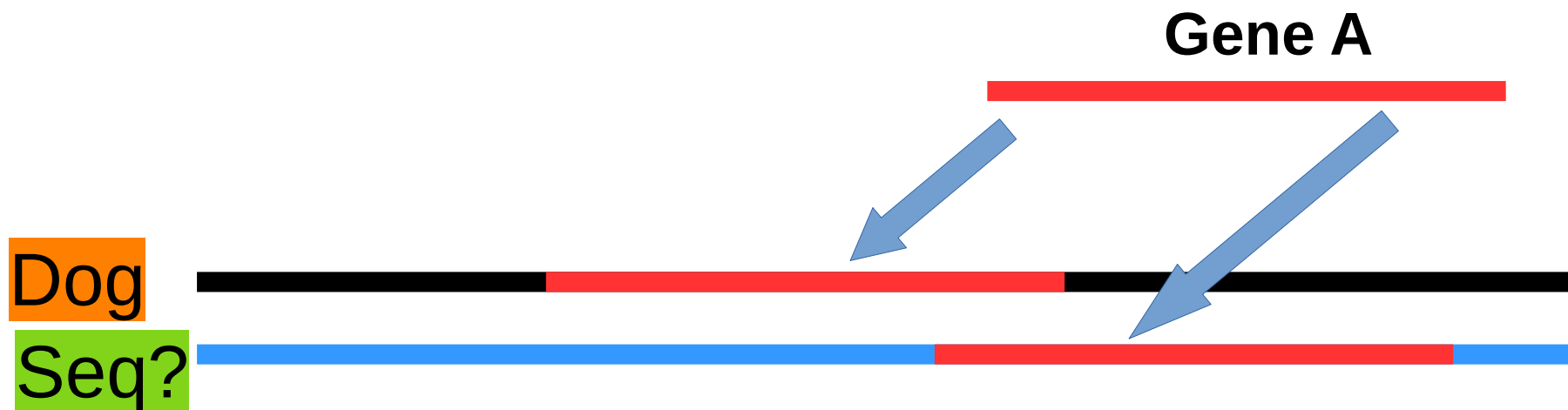


- **Why do we *compare* these things?**
- **What do we learn when *things* have similar DNA?**
  - **Or do not have similar genetic material?**



# Comparing Regions?

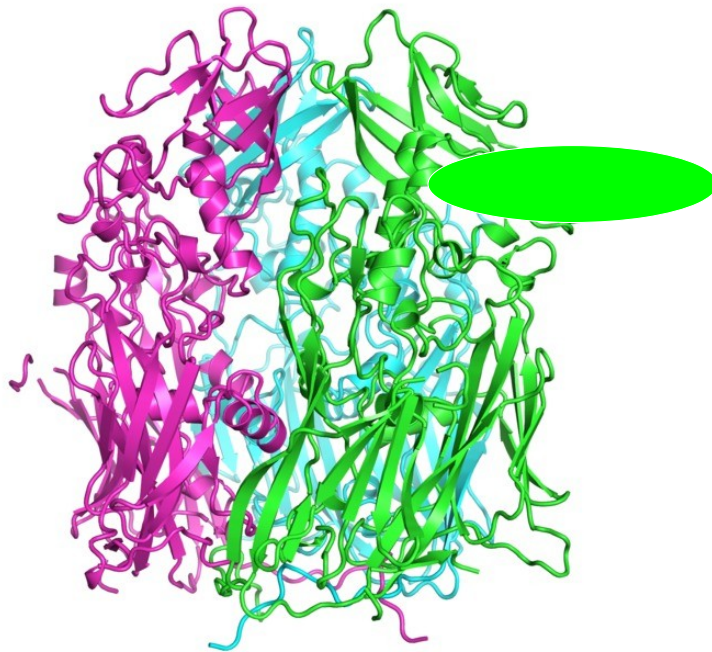
- We scan millions of unknown DNA sequences to find familiar embedded genes.
- Does a gene have the same function each time we find it in a sequence? Same origins? Do the same disorders affect the copies of the gene?



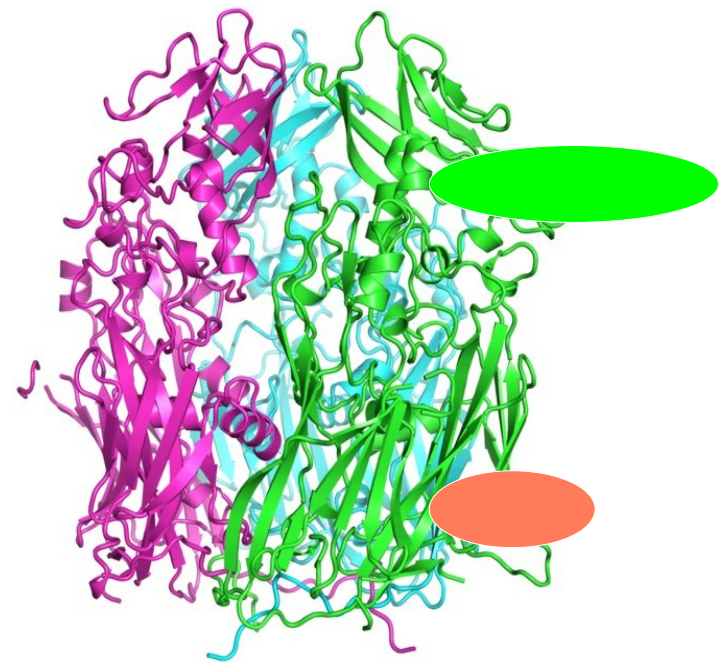
**Gene A is found in this unknown sequence**

# Comparing Protein

- Two proteins (wildtype, non-wildtype) are compared to find causes of disorder.



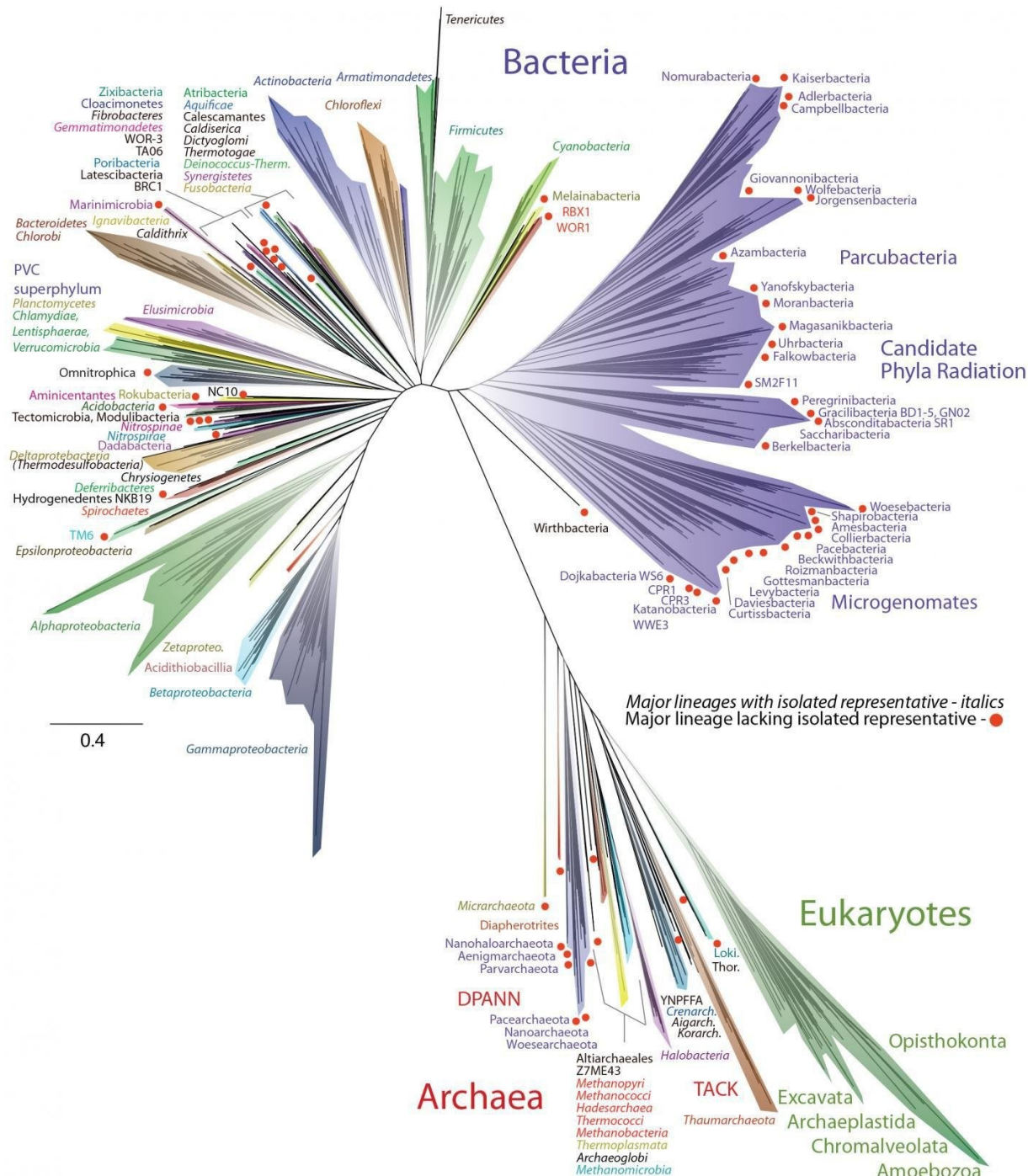
Healthy



Unhealthy



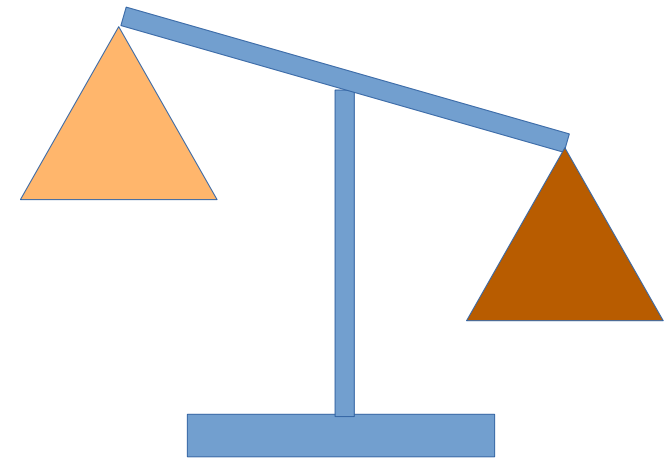
ALLEGHENY  
COLLEGE



Comparing  
organismal  
DNA  
to study the  
Tree of Life  
to  
Determining  
Relatedness

# So, Why Computer Science?

- What is different between sequences?
  - Compare seq A and seq B
  - Now try seq C and seq D



**Seq A: *This is a goat!***

**Seq B: *This ia s boat!***

**Seq C: *actcgaattt ctcgcattta ctttgtttt gaattcgcgc***

**Seq D: *actcgaactt ctcgcattta ctttagttg gatttagcgc***

**What if these sequences get *really, really* long!**





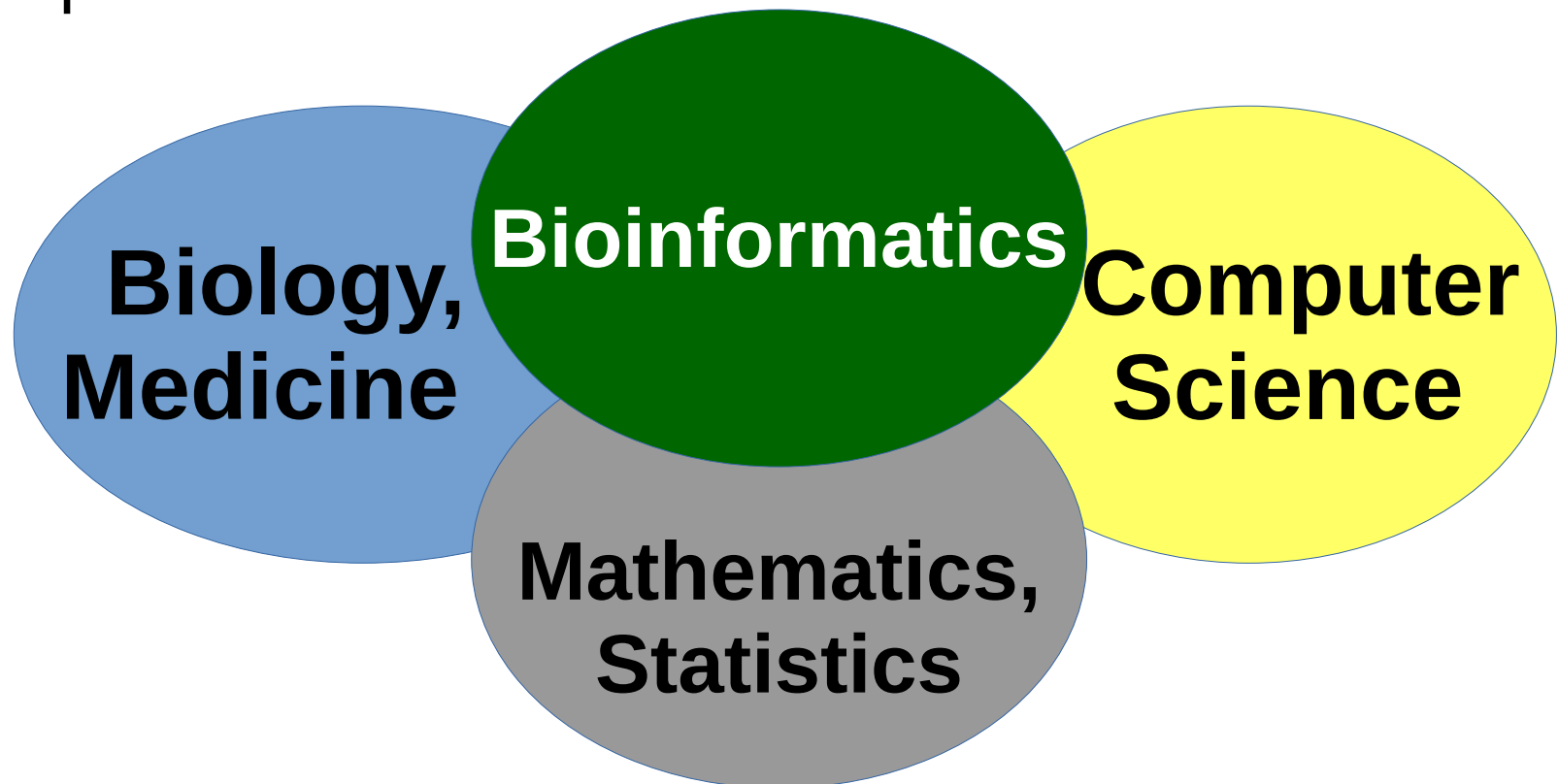
# General Objectives in Bioinformatics

- Sequence analysis
- Make/ use software tools to perform analyses
- Add (publicly) available bio data
- Identification of sequence similarities
- Learn the function of sequences
- Use DNA to answer questions
- And more!



# As a Discipline

- Discipline: a branch of knowledge, typically one studied in higher education.
- Bioinformatics is formed out of three or four disciplines.





# Consider This!

- By group-work? Think about / Discuss the following.
- List some potential applications of bioinformatics (inferring from today's class)
- Why else would you want to compare genetic material? How could this knowledge be meaningful?
- How could knowing the relationships between organisms improve science and medicine?

**THINK**