WELCOME
to
Bioinformatics for biologists

BSC4434

Instructor: Dr. Liberles

Assistant Professor, Biological Sciences

Learning assistants

Christian Balbin

Research assistant

cbalbin@fiu.edu

Office hour: Wednesday 3:30-4:30PM

Danny Morales

Undergraduate researcher

dmora127@fiu.edu

Office hour: Thursday 5-6PM

Jessica Gonzalez

Undergraduate researcher

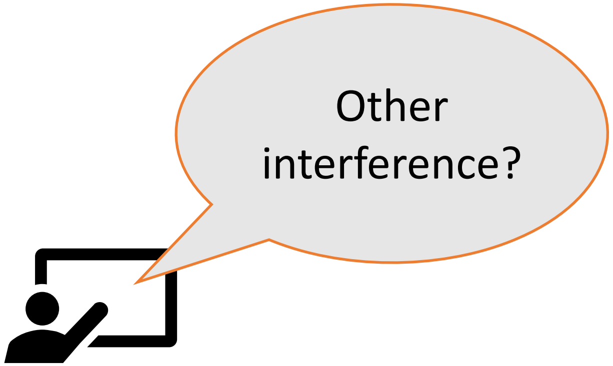
jgonz750@fiu.edu

Office hour: Friday 5-6PM

Liberles
office hour: M/W 6:15-7PM
Tuesday 12:30-3PM

Welcome to the Zoom Classroom!

- We will use Zoom for class.
- Most lectures will be pre-recorded so you can study them before class or they will have enough text so you can read even if there is other interference going on that is outside our control.



- Internet connectivity issues
- Shared workspace/living space
- A barking dog
- Someone mowing the lawn outside your window

- If you (or I) get disconnected at any time for any reason, please use the same meeting link to return to class.

- If that does not work, please send me an email.

Must we have the camera on?

No. While it is very appreciated, it is not necessary.

If you like to have your camera on in the breakout room and not in the main room, that works too.

Must we have the microphone on?

No. In the main room keep your microphone on mute. Chiming in, asking and answering questions is much appreciated. Use the “blue hand” in Zoom to raise your hand or type your questions in the chat.

In the breakout room, on the other hand, it is highly recommended to keep the microphone on, if your background environment allows it. It makes it easier to do group work.

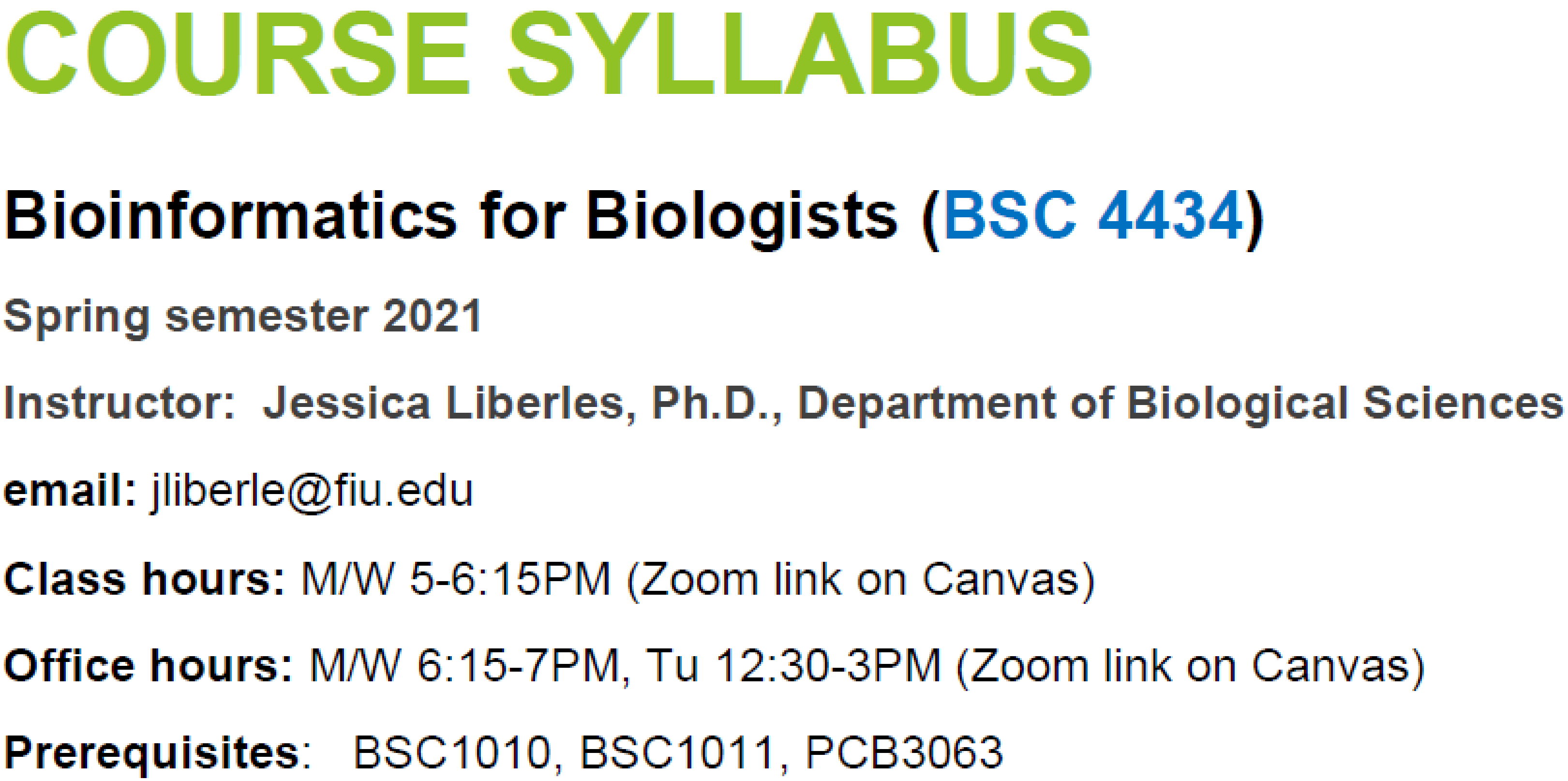
If you are in a noisy environment, you don’t need to have the microphone on, but you can use the chat function.

In the Zoom classroom, in the remote learning environment, with a pandemic threat lurking over us, **we need to be flexible and understanding.**

Unexpected events are to be expected?

If anything happens, **don’t hesitate to reach out** to me or anyone else on the teaching team.

Let’s work together to make this
a fun bioinformatics Zoom class!



COURSE SYLLABUS

Bioinformatics for Biologists (BSC 4434)

Spring semester 2021

Instructor: Jessica Liberles, Ph.D., Department of Biological Sciences

email: jliberle@fiu.edu

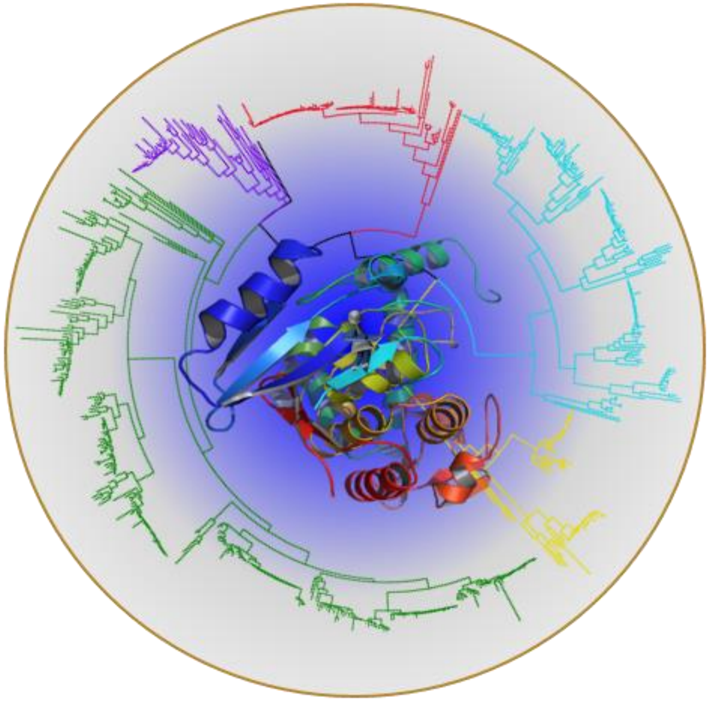
Class hours: M/W 5-6:15PM (Zoom link on Canvas)

Office hours: M/W 6:15-7PM, Tu 12:30-3PM (Zoom link on Canvas)

Prerequisites: BSC1010, BSC1011, PCB3063

3 LAs: Christian, Jessica, and Danny

Bioinformatics for Biologists



Office hours:
after class 6:15-7 PM
Tu 12:30-3

A Zoom room will be open for office hours and you can drop in as needed. Breakout rooms will be used for grade discussions and such.

To contact Dr. Liberles, send an email.
jliberle@fiu.edu

COURSE DESCRIPTION

Introduction to bioinformatic resources/methods for biologists, including development and implementation of a research project. Accessing, searching, retrieving, and analyzing data, including sequence alignment, phylogenetic analysis, and protein structure prediction.

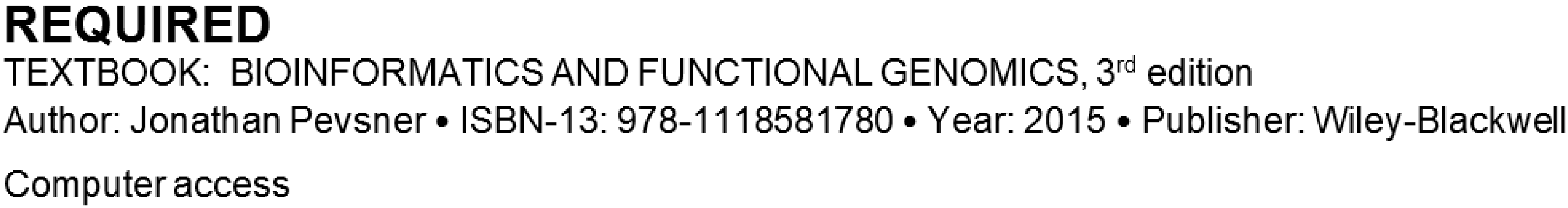
COURSE OBJECTIVES

In Bioinformatics for Biologists you will learn the theory behind fundamental bioinformatics methods, while identifying how to strategically apply these applications. Thus, this course has one theoretical part and one applied part each week. Project based learning will be frequently used.

At the end of the course, you will be able to:

- Use bioinformatics tools to study biology
- Recognize how to correctly apply bioinformatics tools to different situations
- Describe common bioinformatics algorithms
- Determine which combination of data and bioinformatics algorithm is appropriate to address a certain biological question
- Identify the characteristics and limitations of bioinformatics tools to critically analyze the results obtained
- Interpret the results of bioinformatics analyses in a biological context
- Be familiar with peer-review and the importance of reproducible research

Students will be assigned a group project. For the project, groups of 4 students will form a research team and together investigate an assigned specific question using bioinformatics methodology. The project is written up as a research paper and will be peer-reviewed. More details will be provided during the semester.



REQUIRED

TEXTBOOK: BIOINFORMATICS AND FUNCTIONAL GENOMICS, 3rd edition

Author: Jonathan Pevsner • ISBN-13: 978-1118581780 • Year: 2015 • Publisher: Wiley-Blackwell

Computer access

If computer access is a problem, please let me know asap.

COURSE OUTLINE

Part I – Learn to do Bioinformatics (7 weeks)

The first part of the class will be at high pace and based on lectures, activities, and the textbook. The high pace is needed to build a foundation necessary for doing any type of bioinformatics research for part II.

This course follows a “flipped” instructional model, in the sense that lectures and other material for Part I will be presented in online Modules. Students are expected to study these modules before coming to class as specified in the schedule below. Before class each Wednesday, a quiz must be completed in Canvas.

During class, we will use the materials from the modules to actively work on bioinformatics activities in Zoom breakout rooms. The activities are designed to have a mix of tasks to be done individual and in a group of 4. The class also includes interactive lectures based on the modules and the activities. These lectures will integrate the results of the activities in a greater bioinformatics and biological context.

The teaching team (instructor and LAs) will be assisting with activities, providing demos, and discussion in the smaller groups during activity time.

The activities will be completed during class time sometimes by each student and sometimes as a group. These will be checked for a completion grade.

Grading components (Part I)

- Attendance
- Weekly (or biweekly) quizzes at end of Wednesday class
- Applied test 1 – online test, multiple choice
- Theory test 1 – written theory test (multiple choice, short answer, essay questions)
- Activity completion

Part II – Do Bioinformatics to Learn (7 weeks + Finals week)

The second part of the course is a research simulation. Based on a learning profile survey during Part I, you will be placed into groups. Each group will be assigned a research project and each person in the group will have a specific objective to complete. The group project can only be completed if the group collaborates to integrate the different objectives and together writes a research paper that presents, analyzes, and discusses the project and its results.

Peer-review and the reproducibility of research results are two crucial components for the advancement of science through publication. You will perform a peer-review of another group's paper while your paper is being reviewed as well.

Grading components (Part II)

- Project draft 1 – a rough draft
- Project draft 2 – a complete draft
- Project peer review – review another groups paper and have your paper reviewed by another group
- Final project research paper
- Applied test 2
- Final exam

GRADING

PART I

Activity completion	100p
Weekly quizzes (Quiz 1-5: 15p each, Quiz 6: 25p)	100p
Applied test 1	100p
Theory test	140p
Attendance (5p/class)	55p
Total Part I	**495p**

PART II

Project draft 1	30p
Project draft 2	40p
Project peer-review	30p
Project paper final	100p
Applied test 2	100p
Final exam	140p
Attendance (5p/class)	65p
Total Part II	**505p**
FINAL TOTAL (Part I + Part II)	**1000p**

GRADE SCALE

NOTE: The tentative point scale shows the optimal scenario. It is not absolute but serves as a guide. The point scale may need to be adjusted based on difficulty levels of quizzes and tests.

Grade	Points Per Credit Hour	Tentative point scale
A	4.00	>925
A-	3.67	>895-925
B+	3.33	>865-895
B	3.00	>825-865
B-	2.67	>795-825
C+	2.33	>765-795
C	2.00	>695-765
D	1.00	>595-695
F	0.00	<595

TENTATIVE SCHEDULE

⬜ University closed
🟨 Quiz
🟦 Test
🟩 Project deadline

	Week	Dates	Focus	Online module
Part I	1	Jan 11	Introduction	Module 1
		Jan 13	Databases	
	2	⬜ Jan 18		Module 2
		🟨 Jan 20	BLAST Quiz1-2	
	3	Jan 25	Multiple Sequence Alignments Phylogenetic trees Tree analysis Quiz 3	Module 3
		🟨 Jan 27		
	4	Feb 1	Tree analysis Protein domains and structure Protein modeling Quiz 4	Module 4
		🟨 Feb 3		
	5	Feb 8	Protein modeling Protein predictions Quiz 5	Module 5
		🟨 Feb 10		
	6	Feb 15	Pathways & interactions	Module 6
		🟨 Feb 17	Quiz 6	
	7	🟦 Feb 22	**Theory test**	
		🟦 Feb 24	**Applied test 1**	
Part II	8	Mar 1	**Project launch**	
		Mar 3	Project	
	9	Mar 8	Project	
		Mar 10	Project	
	10	Mar 15	Project	**Draft 1 due noon on Mar 17.** Draft 1 feedback returned Mar 22.
		🟩 Mar 17	Project Draft 1	
	11	Mar 22	Project	**Draft 2 due noon on Mar 31** Draft 2: For feedback from another group including reproducibility. Also feedback from instructor and LAs. Peer review: is the writing sound, does the protocol work, anything missing or unclear? Rubric provided.
		Mar 24	Project	
	12	Mar 29	Project	
		🟩 Mar 31	Project Draft 2/ Peer review	
	13	Apr 5	Project	
		Apr 7	Project	
	14	🟦 Apr 12	**Applied test 2**	
		🟩 Apr 14	Project deadline (in class)	
	Finals week	🟦 Apr 19 5:00-7:00PM	**Final exam (2h)**	

MODULE material on class webpage

Flipped class:
Some material (videos, textbook reading, and so on) to complete before class on Monday and Wednesday.

This material is important for getting the most out of the activities in class.

A 10-15 min quiz will be given at the end of Wednesday’s class every week. A study guide is in the module.

Attendance is mandatory

- Problem solving together enhances your learning and prepares you for the research project that follows in Part II

Activities:

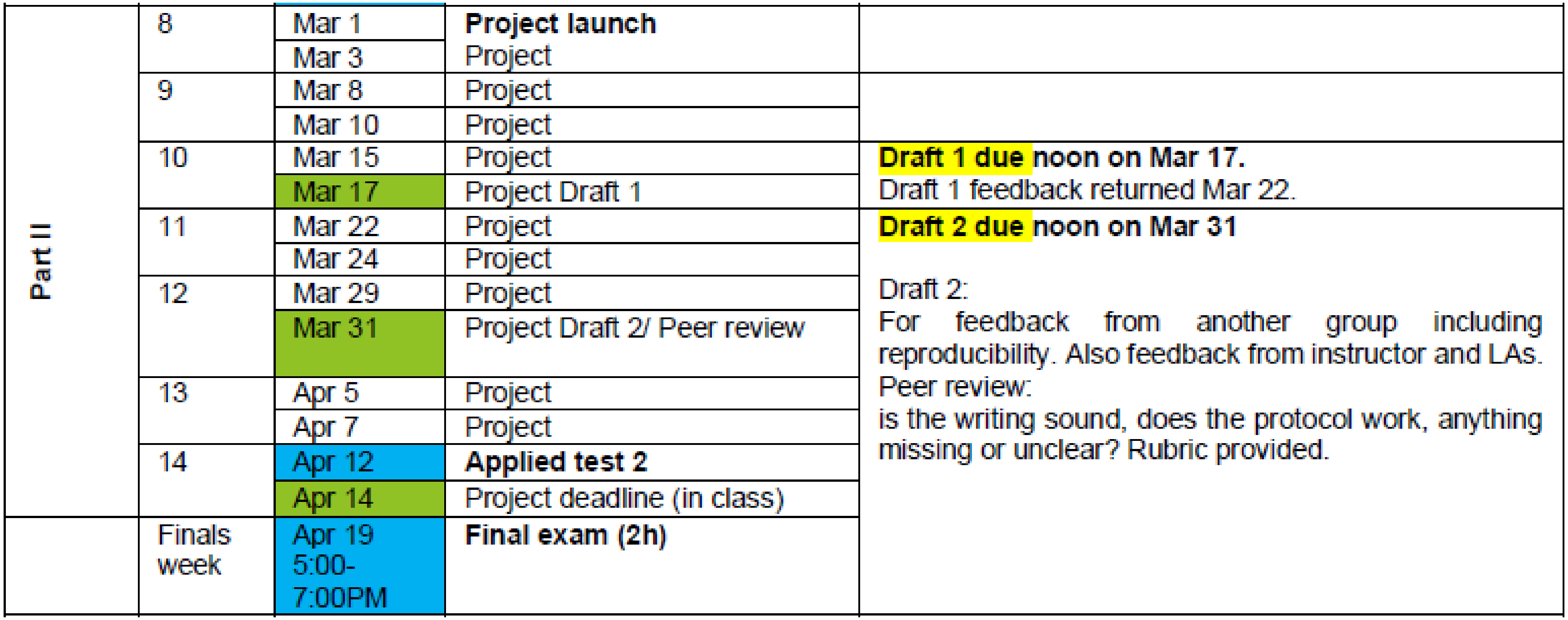
- Bioinformatics is best learned “hands on” so you will work on bioinformatics activities during most of class time.

- Some activities will be completed independently and some as a group. Some will have a mix.

- At the end of class, save as pdf and upload to canvas.

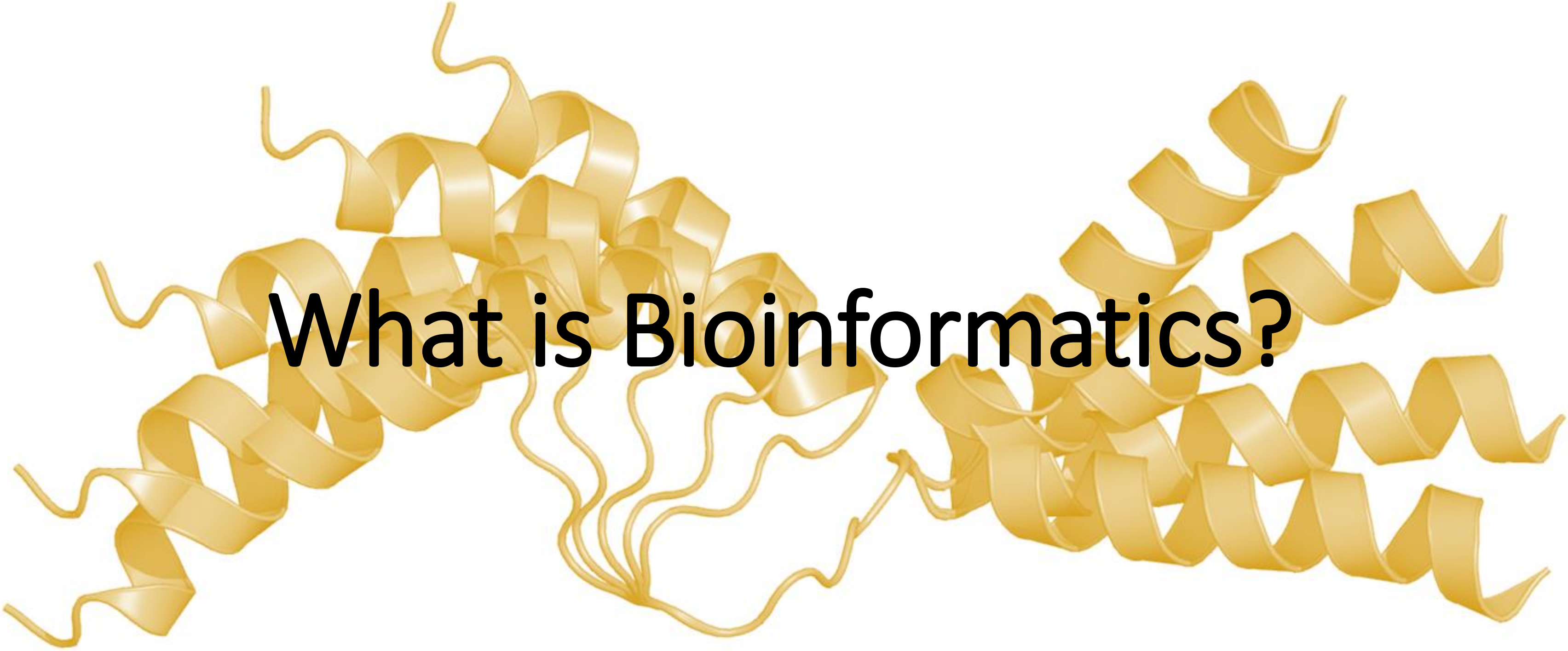
The Research Project

- Cancer genomics?
- COVID-19 related?



Part II	8	Mar 1	**Project launch**	
		Mar 3	Project	
	9	Mar 8	Project	
		Mar 10	Project	
	10	Mar 15	Project	**Draft 1 due noon on Mar 17.** Draft 1 feedback returned Mar 22.
		Mar 17	Project Draft 1	
	11	Mar 22	Project	**Draft 2 due noon on Mar 31** Draft 2: For feedback from another group including reproducibility. Also feedback from instructor and LAs. Peer review: is the writing sound, does the protocol work, anything missing or unclear? Rubric provided.
		Mar 24	Project	
	12	Mar 29	Project	
		Mar 31	Project Draft 2/ Peer review	
	13	Apr 5	Project	
		Apr 7	Project	
	14	Apr 12	**Applied test 2**	
		Apr 14	Project deadline (in class)	
	Finals week	Apr 19 5:00-7:00PM	**Final exam (2h)**	

What is Bioinformatics?



[Bioinformatics: What? Why? Who? - YouTube]

Important milestones for the field of bioinformatics

time

1953
Watson & Crick propose the double-helix model of DNA.
Why is this so important?

1958
Kendrew & Perutz solve the first high-resolution protein structure

1970
Needleman-Wunsch Algorithm for aligning protein sequences.

1977
Fred Sanger develops Sanger sequencing.

1984
Epstein-Barr virus genome sequenced.

1988
The term ‘bioinformatics’ is coined.

1990
The Human Genome project is launched.

1990
BLAST – the Basic Local Alignment Search Tool

1998
C. elegans genome sequenced.

2000
Drosophila melanogaster and Arabidopsis thaliana genomes sequenced.

2001
A draft of Human Genome project is published.

time

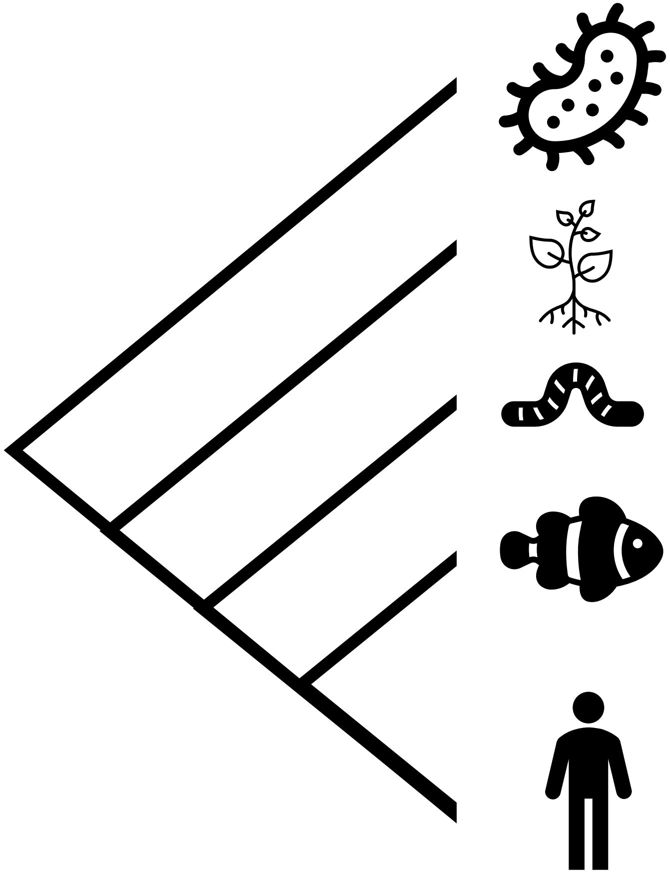
Francis Collins, NIH

Craig Venter, Celera genomics

(Venter Collins)

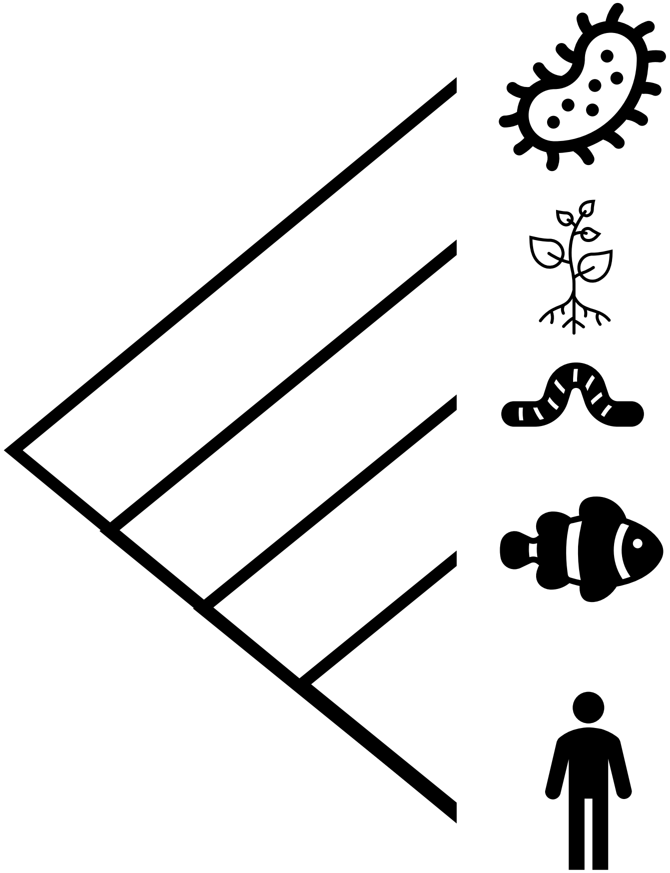
The sequence is known but the code has not yet been cracked. Bioinformatics is working on this in various ways…

How many genes in the human genome?



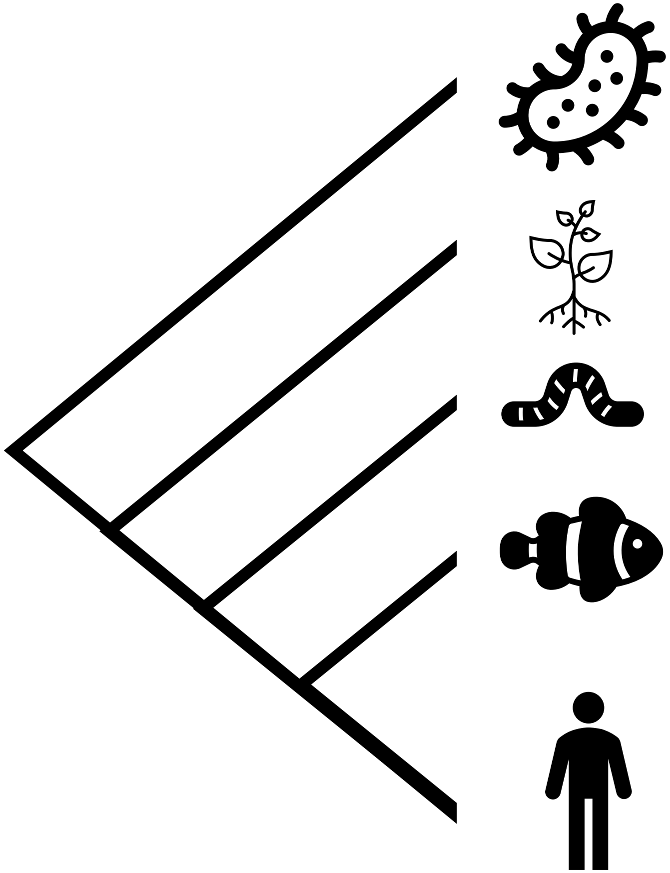
Species name	Number of genes (approx.)
E. coli	5,000-6,000
Arabidopsis thaliana	27,000
C. elegans	22,000
Zebrafish	
Human	

Number of genes does not explain organismal complexity



Species name	Number of genes (approx.)
E. coli	5,000-6,000
Arabidopsis thaliana	27,000
C. elegans	22,000
Zebrafish	26,000
Human	21,000

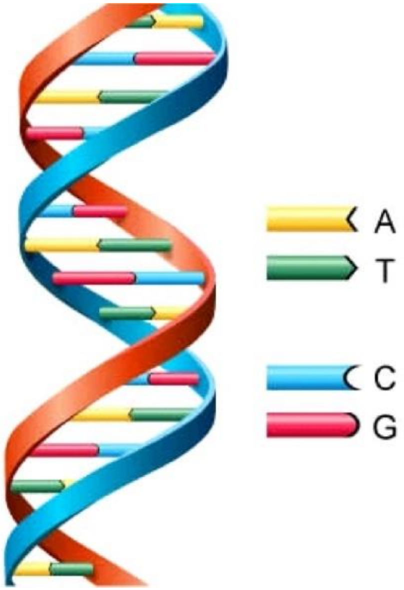
How many genes in the genome?



Species name	Number of genes (approx.)
E. coli	5,000-6,000
Arabidopsis thaliana	27,000
C. elegans	22,000
Zebrafish	26,000
Human	21,000

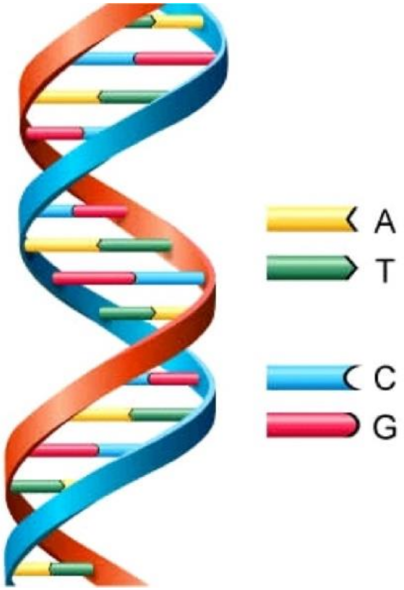
The number of genes is similar.
Diversity?
Complexity?

The Human Genome



The entire string of DNA sequence (A, C, G, T)

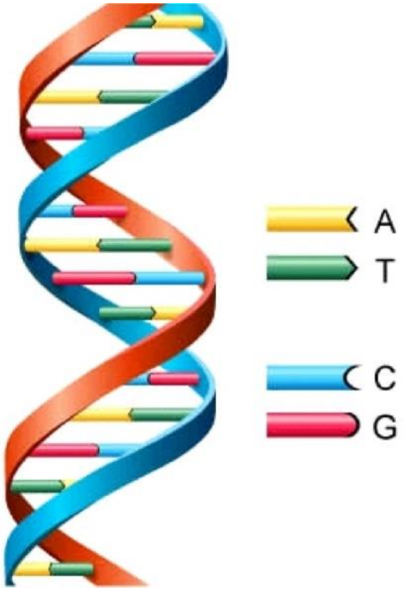
The Human Genome



The entire string of DNA sequence (A, C, G, T)

We have the code, but we still need to decipher what it means!

The Human Genome

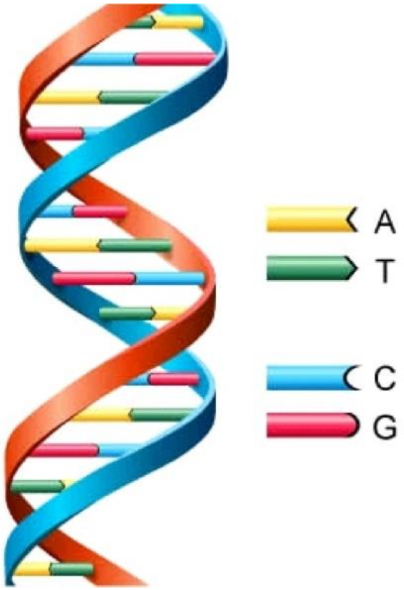


The entire string of DNA sequence

Genes

Not coding for genes.

The Human Genome



The entire string of DNA sequence

A gene can be divided into

Introns: 🟨

and

Exons: 🟧

The Human Genome

The entire string of DNA sequence (3 billion base pairs)

Genes

Introns

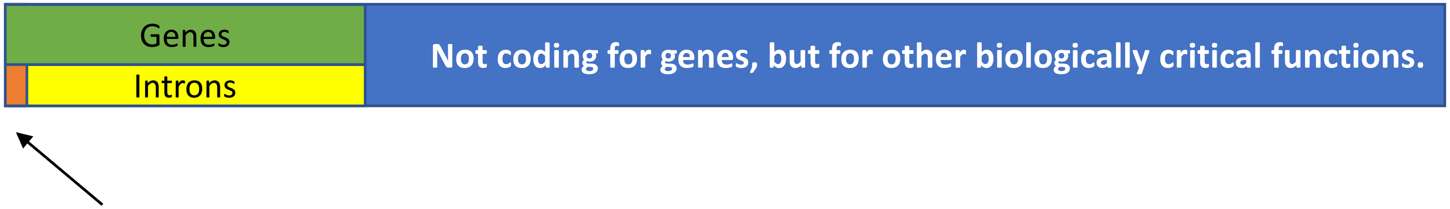
Not coding for genes.

Exons

Only a tiny part (1.5%) of the human genome encode exons.

The Human Genome

The entire string of DNA sequence (3 billion base pairs)



Genes

Introns

Not coding for genes, but for other biologically critical functions.

Exons

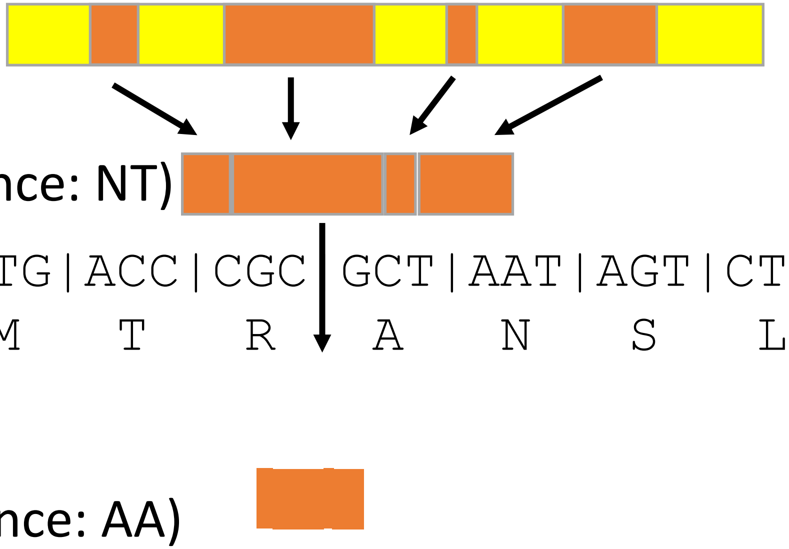
Only a tiny part (1.5%) of the human genome encode exons.

From Gene to Protein

A gene can be divided into

Introns: 🟨

Exons: 🟧



Gene DNA code

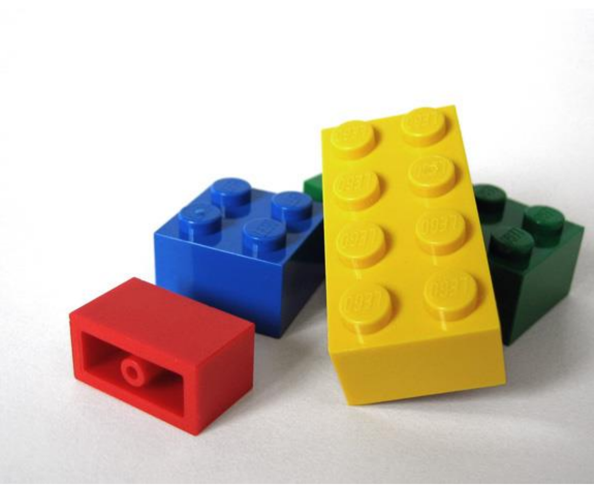
Exon DNA code
(nucleotide sequence: NT)

NT: ATC | GCT | ATG | ACC | CGC | GCT | AAT | AGT | CTT | GCG | ACG | GAA | GAT

AA: I A M T R A N S L A T E D

Protein
(amino acid sequence: AA)

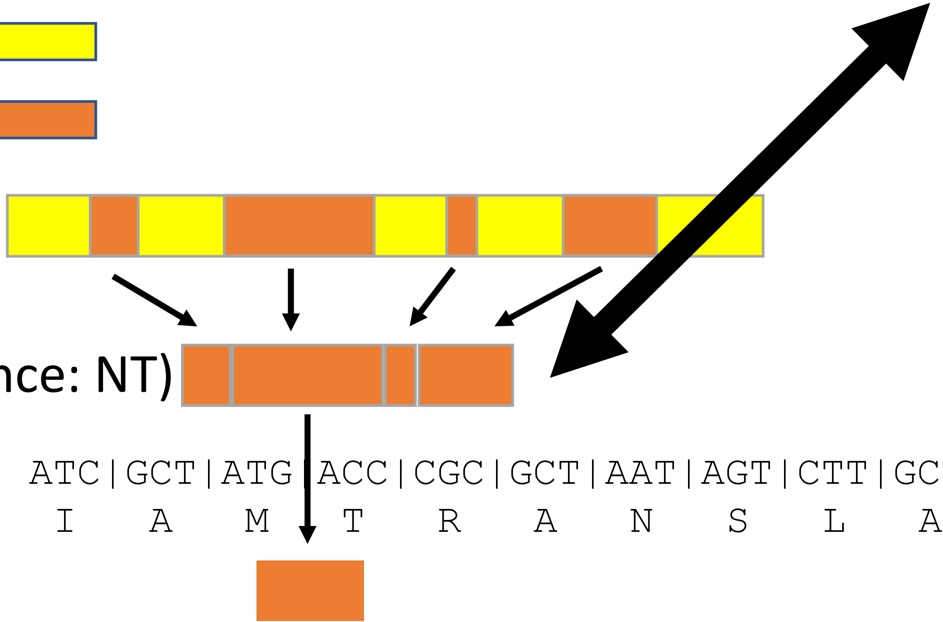
From Gene to Protein



A gene can be divided into

Introns: 🟨

Exons: 🟧



Gene DNA code

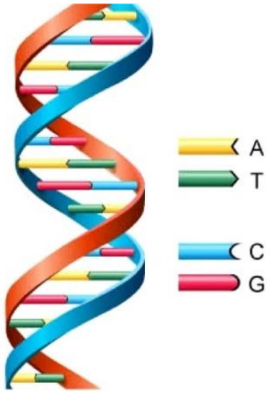
Exon DNA code
(nucleotide sequence: NT)

```
NT: ATC|GCT|ATG|ACC|CGC|GCT|AAT|AGT|CTT|GCG|ACG|GAA
AA:  I   A   M   T   R   A   N   S   L   A   T   E
```

Protein
(amino acid sequence: AA)

The exons can be combined in different ways (like LEGO blocks) to make different proteins.

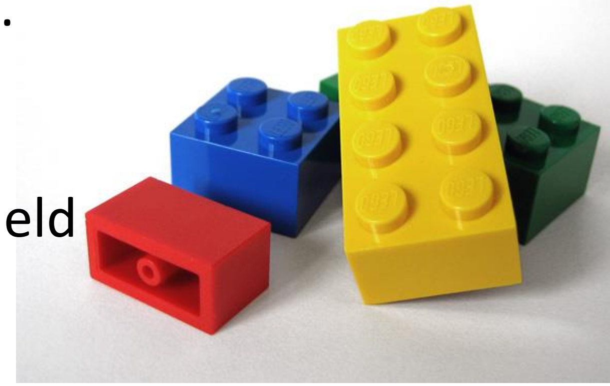
The Human Genome



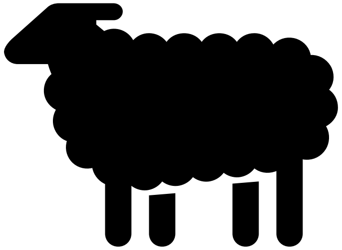
The entire string of DNA sequence

Only a tiny part of the human genome encode exons.

Alternative splicing:
Combining the exons in a gene in different ways to yield different mRNAs that will become different proteins.



Genomic sequence data -



2001 – the first rough draft of human genome sequence

2003 – a "complete" human genome sequence

13 years

$3 billion (almost)

Now: the bulk price for a human genome sequence

1-2 days

SALE $1,000

More sequence than experimental validation.

Other omics fields are also moving fast, but not fast enough to keep up with the sequence data accumulation.

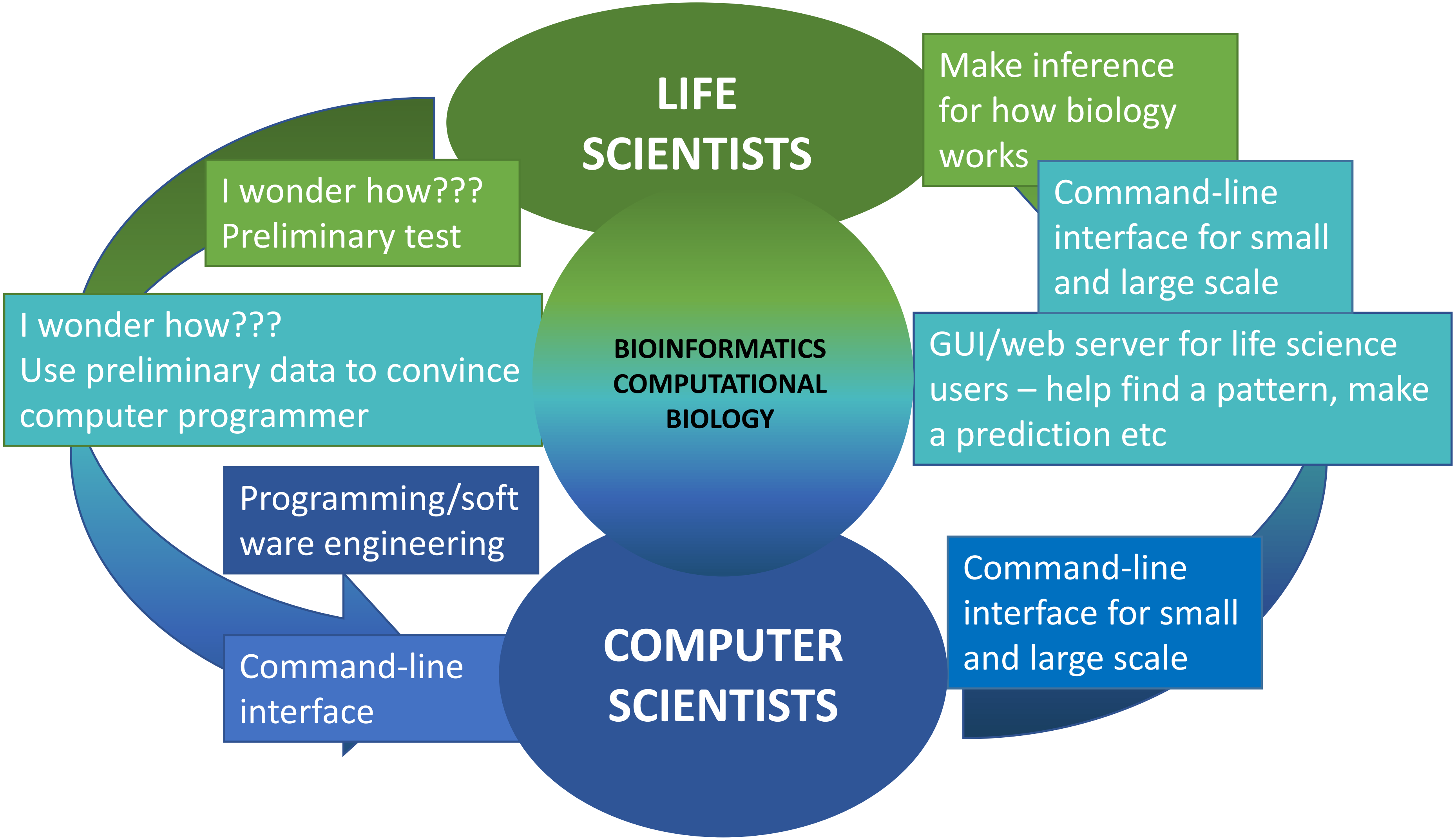
Predicting features from the sequence data makes it more useful.

Computer science is critical for making efficient algorithms, but experts from other disciplines are important for driving the science forward by generating prototypes or observing patterns.

Depending on what to predict and study, different disciplines are important. For instance,

- Biophysics for protein folding and 3D modeling
- Evolutionary biology for comparative genomics
- Cell and molecular biology for function…
- Statistics for genome wide associations

Bioinformatics: two sub-cultures?



LIFE SCIENTISTS

I wonder how???
Preliminary test

I wonder how???
Use preliminary data to convince computer programmer

Programming/software engineering

Command-line interface

COMPUTER SCIENTISTS

Command-line interface for small and large scale

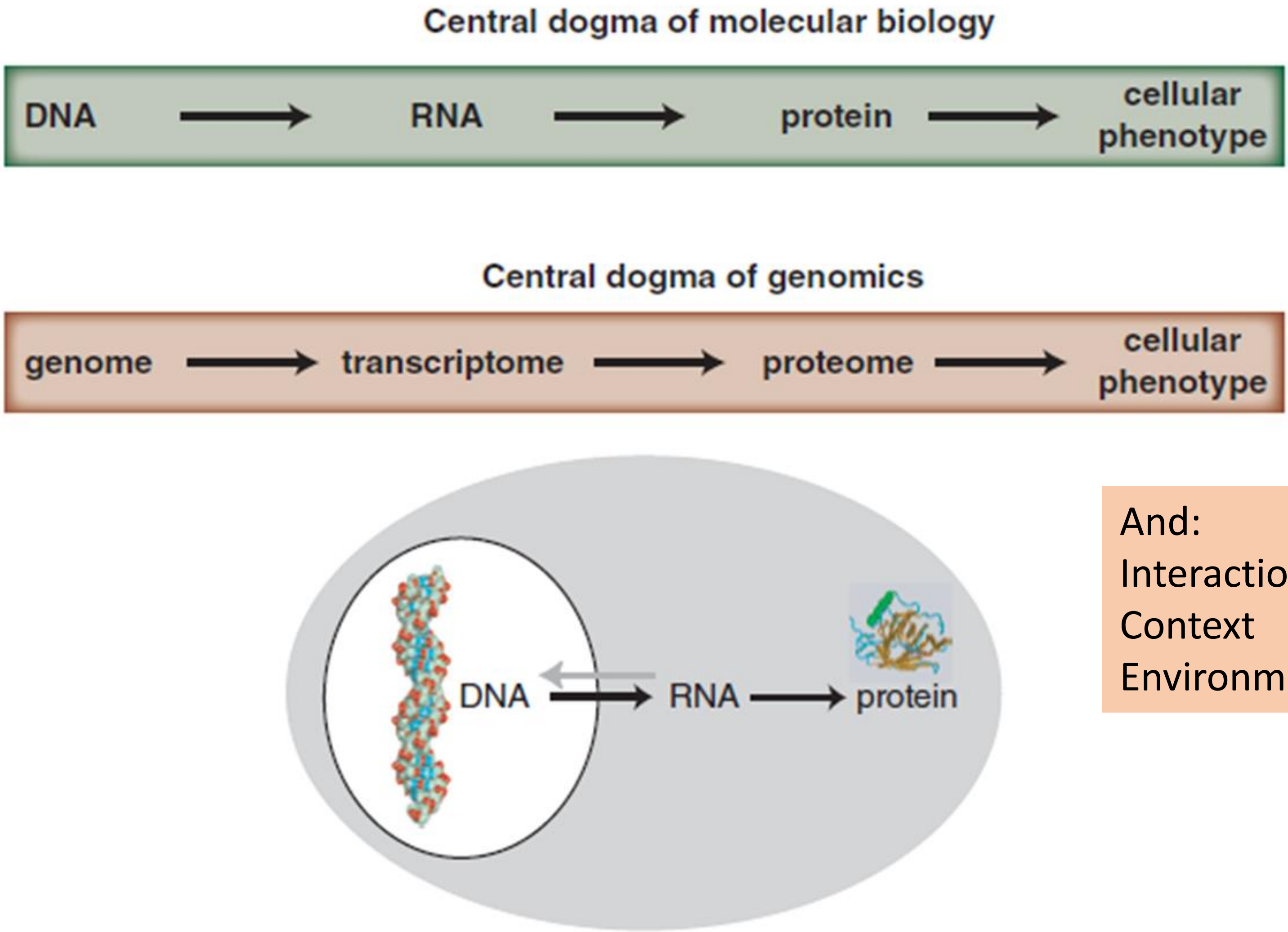
GUI/web server for life science users – help find a pattern, make a prediction etc

Command-line interface for small and large scale

Make inference for how biology works

BIOINFORMATICS
COMPUTATIONAL BIOLOGY

The Central Dogma



Central dogma of molecular biology

DNA → RNA → protein → cellular phenotype

Central dogma of genomics

genome → transcriptome → proteome → cellular phenotype

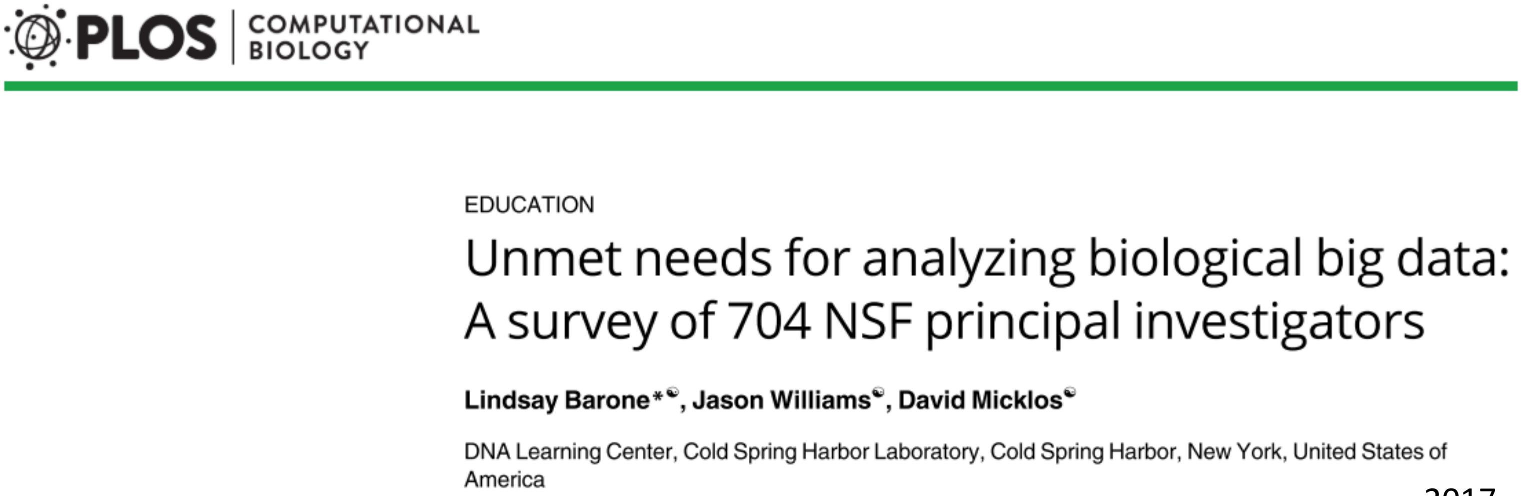
And:
Interactions
Context
Environment

Biology is in transition

High-throughput and large-scale experiments are no longer too expensive: they are becoming/have become the norm.

- Understanding and being able to handle large datasets becomes an expectation.
- Using bioinformatics algorithms appropriately is key.
- Experimental design benefits from being informed by bioinformatics.
- The type of experiments that can be run can focus on entire systems, from multiple angles, not just a gene at the time…

From “What is?” to “what isn’t?” computational biology….



PLOS | COMPUTATIONAL BIOLOGY

EDUCATION

Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators

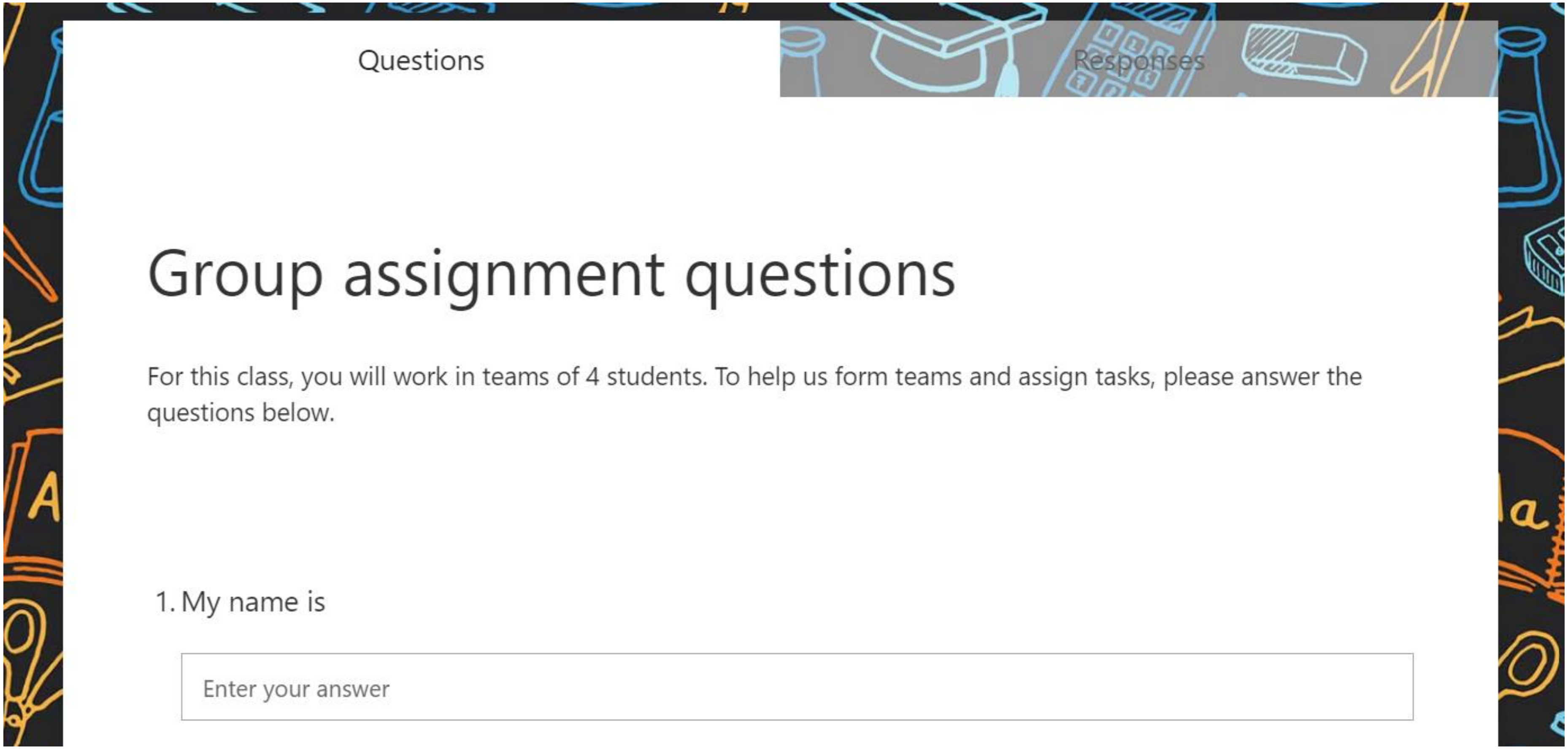
Lindsay Barone*, Jason Williams, David Micklos

DNA Learning Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America

2017

- 87% PIs were currently, or would in the near future, work with big data sets.
- Almost 70% would work with genomic sequence data.
- Unmet need? Trained workforce!

Today’s activity



Group assignment questions

For this class, you will work in teams of 4 students. To help us form teams and assign tasks, please answer the questions below.

1. My name is

https://forms.office.com/Pages/ResponsePage.aspx?id=qOV5rOTgS0OikiyJtcKDZtmYSj5NX4BJg98SuEa9hVdUMktMVFdCUkpFRkEwMjBGRERHME5JU1dONi4u

We will try to make well-functioning groups, but if you ever feel that your group is not working, please let us know so we can adjust the groups.

The activity sheet is here: https://drive.google.com/file/d/1OGnCBe-7L3zI6HYMnt63L-FaawaYeAmg/view?usp=sharing (link on class website)

- Make a copy of the sheet.
- Go through the steps of the activity and write your answers in your copy of the sheet
- Talk to your peers in your random breakout room.
- Help each other.
- Recommended to keep the microphone and video on, if possible.
- We will circle through, but you can also use the Ask for Help button.
- Submit to Canvas Day 1 activity when completed, no later than 1 PM on Wednesday.
- This deadline is important so I can follow up on the activity at the beginning of next class.

So, where is the class website?

- Can you find it in Canvas?

- https://www.siltbergliberleslab.com/bfb-sp2021
- The password is panda