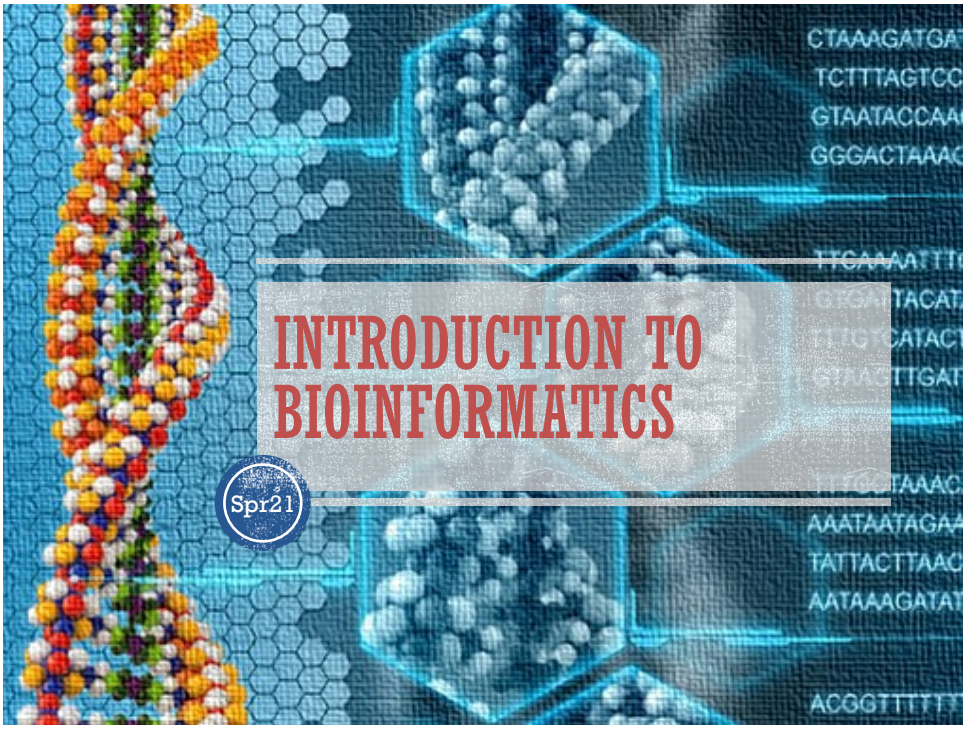


1

1. Introduction

- Icebreaking
- Course Info
- Introduction :
  - The human genome project
  - DNA Sequencing
  - Bioinformatics: what & why
  - Applications

Topics

2

Who is your instructor?

Dr. Amal Khalifa

FOLLOW

Assistant Professor of Computer Science, Purdue University Fort Wayne
Verified email at pfw.edu

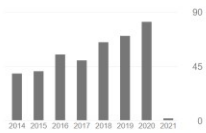
Data Analytics Computational Biology Steganography High Performance Computing

GET MY OWN PROFILE

TITLE	CITED BY	YEAR
Video steganography: a comprehensive review MM Sadek, AS Khalifa, MGM Mostafa Multimedia tools and applications 74 (17), 7063-7094	119	2015
High-capacity DNA-based steganography A Khalifa, A Alito 2012 8th International Conference on Informatics and Systems (INFOS), BIO-76 …	53	2012
High capacity image steganography using wavelet-based fusion MF Tolba, MAS Ghonemy, IAH Taha, AS Khalifa Proceedings. ISCC 2004. Ninth International Symposium on Computers And …	52	2004
Using integer wavelet transforms in colored image steganography MF Tolba, MA Ghonemy, IA Taha, AS Khalifa	51	2004

Cited by VIEW ALL

	All	Since 2016
Citations	493	324
h-index	10	9
i10-index	11	7



PURDUE FORT WAYNE

Princess Nora Bint Abdul Rahman University

UNIVERSITY OF CONNECTICUT 1881

AIN SHAMS UNIVERSITY

3

4

5 Course Info

5

Contact info.

- **Instructor**: Dr. Amal Khalifa
- **Office hours**: (ET-125D) MW 10:00 AM- 12:00 PM, or by appointment
- **Email**: khalifaa@pfw.edu

6

References

7

Assessment methods and grading

- **Programming Assignments**: 50%
- **Quizzes**: 30%
- **Term Project :** 20%

	A	90 – 100
	B	80 – 89
Passing grade →→	C	70 – 79
	D	60 – 69
	F	Below 59

8

Policies

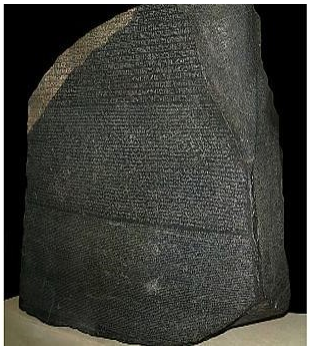
- Class attendance
  - Stay at home in case you feel sick!
  - Assigned seats
  - QR code
- Assignments should be done individually
  - 20% penalty for late assignments
- Take notes, review material, please!
- Active engagement, share knowledge from which both peers and instructor can learn.
- Missed exam regulations
- Ensure that you have access to the course web site on Brightspace.
- You need to bring you laptop for both interactive and coding exercises

9

10 Introduction

Ciphering a language!

- The Rosetta Stone is found by French troops the Napoleon's Egyptian Campaign in the village of Rosetta, in Egypt.
- Dated back to 196 BC and written in three scripts: Ancient Egyptian hieroglyphs, Demotic script, and Ancient Greek.
- allowed for the modern understanding of hieroglyphs.



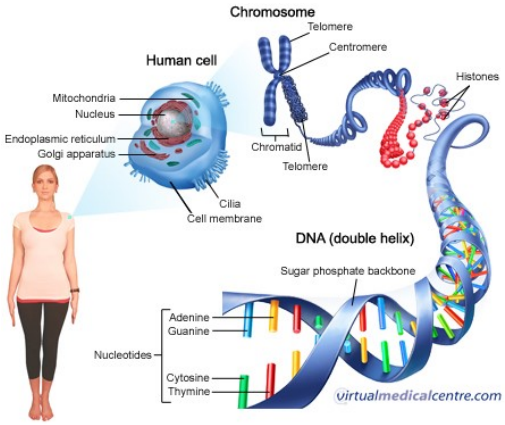
The language of life

- The Human Genome Project (HGP) was one of the great achievements of exploration in history.
- Objective : reveal the secrets of life
- An inward voyage of discovery started in 1990 and led by an international team of researchers
- Completed in 2003 by identifying the genetic blueprint for our species, Homo sapiens.

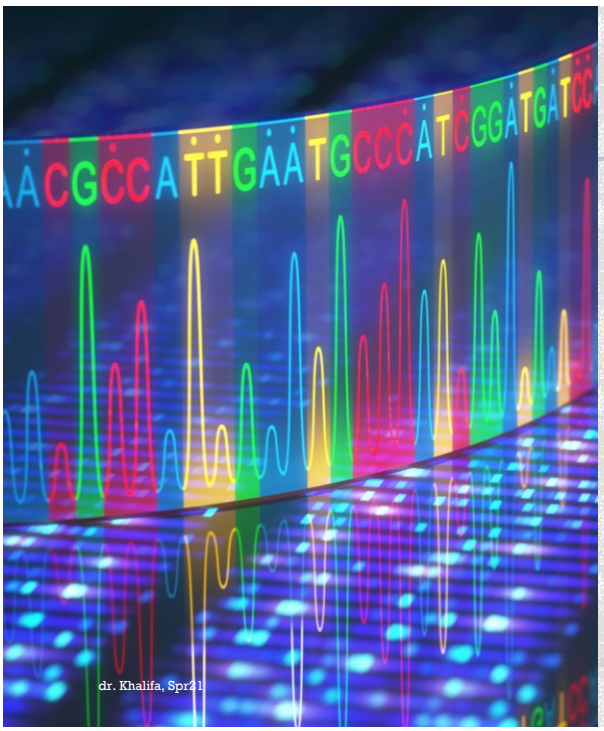
https://www.genome.gov/human-genome-project

Genetic information

- Every cell in a person's body has the same DNA (deoxyribonucleic acid)
  - 46 (23-pairs) human Chromosome
  - Packed in the nucellus
  - Half of a person's DNA is inherited from the mother, and the other half from the father
  - contains all necessary information required to produce/create/direct activities any living organism.
- Human genome
  - 3 billion DNA bases
  - ~ 30,000 genes [ref]



13

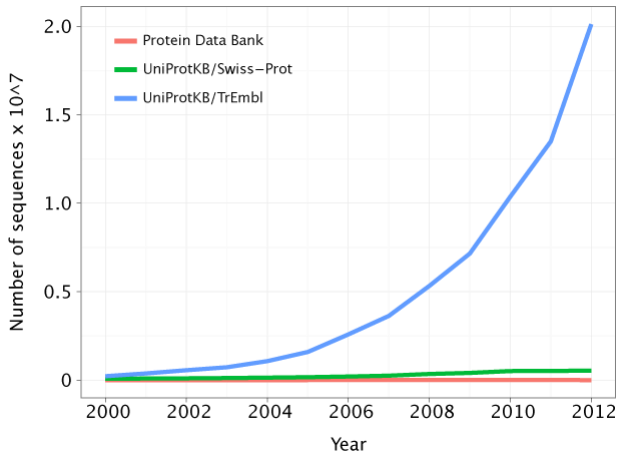


DNA Sequencing

- DNA sequencing is the process of determining the nucleic acid sequence – the order of nucleotides in DNA.
- "next-generation" sequencing (NGS) methods were developed in the 1990s and were implemented in commercial DNA sequencers by the year 2000 → drop in the cost

14

Growth of Sequence Data



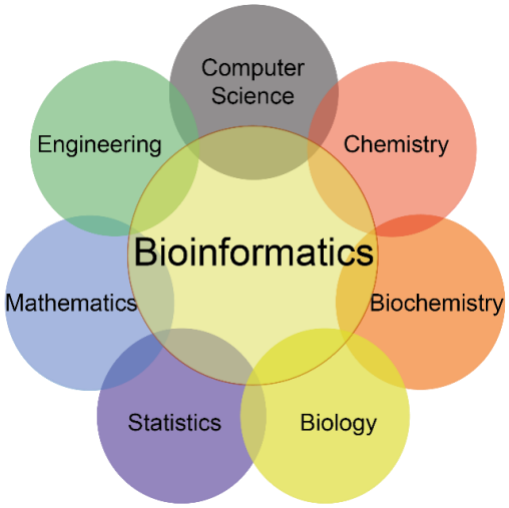
Revealing structure does not really mean understandi ng function!

15

Computation is needed!

- The production of raw sequence data is only the beginning of its detailed bioinformatical analysis
  - Data needs to be organized and stored efficiently
  - Analysis of data requires efficient algorithms
  - Exponential growth requires scalable methods

16

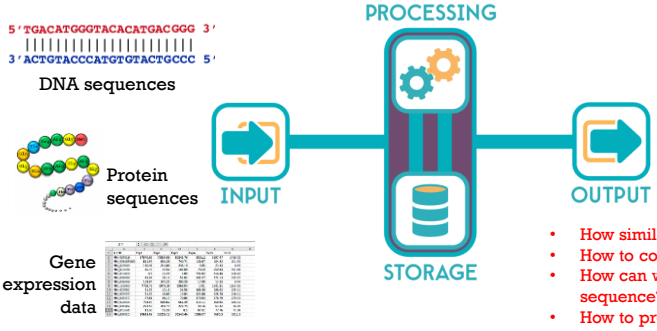


Bioinformatics

- Two parts:
  - Bio: biology
  - Informatics : the science of data processing
- an interdisciplinary field that develops and applies computational technologies and methods to study and analysis of complex biological data

17

From data to Information



- How similar two sequences are?
- How to compare sequences?
- How can we find a gene in a DNA sequence?
- How to predict the structure and functionality of a gene?
- Can I use the known function of a gene to guide my study of a new gene?
- How genes are differentially expressed and why? How is this related to diseases?
- How can we construct protein interaction networks?

18

Applications

- Vaccine design
  - Using bioinformatics tools for the identification of structures from bacteria, virus, parasites, cancer cells, or allergens that could induce an immune response capable of protecting against a specific disease [ref]
- Drug development
  - Bioinformatics has helped experts accelerate drug target identification, drug candidate screening and refinement, characterization of side effects and predict drug resistance. [ref]

19

Applications

- Gene therapy
  - It is possible to single out and analyze the genes that are directly related to illness and disease based on information collected on all of the human genes and proteins. This helped in the detection, diagnosis, treatment, and prevention of diseases. [ref]
- Personalized medicine
  - Molecular profiles generated from a patient's genomic information could help accurately drive the diagnostic, prognostic, and therapeutic plans, tailored to the patient's physiological status. [ref]

20

Questions?