



# Introduction to Bio-Informatics

## Lecture 7

Lecture by: Ahmad R. Naghsh-Nilchi, PhD

Department of Artificial Intelligence

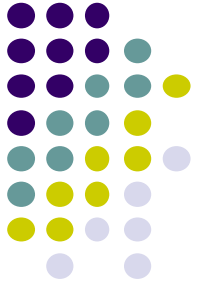
Faculty of Computer Engineering

University of Isfahan



## In the last session, we discussed:

- Not much difference between prokaryotic X01714 and eukaryotic U90223 (mRNA sequence)!
- But significant differences between X01714 or U90223 with eukaryotic **AH005568**, a genomic one.
- Gene-Centric Database is beneficial.
  - Main advantage: it returns results that are more synthetic than a long list of GenBank entries.
  - For that search NCBI “**Gene database**” rather than GenBank.



# Non-coding Regions:

- Non-coding regions of nucleotide sequences, often referred to as non-coding DNA, encompass segments of the genome that do not encode protein sequences but play crucial roles in gene regulation and chromosomal structure.
- These regions include introns, enhancers, silencers, and other regulatory elements, which contribute to the intricate modulation of gene expression and the maintenance of genome integrity.



# Guanine (G) and Cytosine (C)

- "GC" typically refers to the GC content of a DNA sequence, which is the proportion of guanine (G) and cytosine (C) bases in the DNA.
- High GC content can influence the stability and structure of DNA, as well as affect gene expression



# Viral Genomics

- **Viral genomics is the study of the genetic material of viruses.**
- **They are distinct from the genomic structure of cellular organisms.**
- **The complexity of viral evolution is observed by examining the molecular and structural distinctions between viral and cellular genomes.**

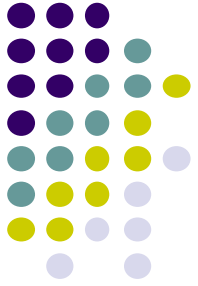


# **Viral Genomics Characteristics**

- **One of the most noticeable features of viral genomics is its diversity in genetic material.**
- **Viruses may possess either RNA or DNA as their genetic material, and it can exist in various forms, including single-stranded or double-stranded configurations.**
- **In contrast, cellular genomes, consisting of double-stranded DNA, are relatively uniform across different organisms.**



# Characteristics ....



- **The variability in viral genomes allows for rapid adaptation and evolutionary changes.**
- **High mutation rates and mechanisms such as recombination and reassortment cause this phenomenon.**
  - **Recombination is responsible for viral resistance to antiviral therapies.**
  - **Reassortment occurs predominantly in segmented viruses, wherein co-infection of a host cell leads to the exchange of genomic segments, generating new viral strains with unique phenotypic characteristics.**
  - **Both pose challenges for public health surveillance and vaccine development.**



# Viral Genomes Size:



- **Viral genomes are much smaller than cellular genomes, from a few thousand to several hundred thousand base pairs:**
  - For comparison, the human genome contains approximately 3 billion base pairs, much larger and more complex.
- **This compactness in viral genomes leads to the need for encoding multifunctional proteins:**
  - A single viral gene may serve multiple roles, contrasting with the more modular and specialized gene functions observed in cellular systems.





# Viral Genomics Sequences



- **The structure of viral genomic sequences is also notably different.**
- **Viruses often have a linear or circular setting of their genetic material, while cellular genomes are organized into chromosomes with a nucleus.**
- **Viral genomes often are in non-coding regions utilized for regulation or packaging.**
- **This feature differs from the eukaryotic genes often marked by intron-exon structures that require merging.**



# Genomes Replication



- **Viral genomes need host cellular machinery for propagation**
- **So, viral genomes have a critical dependency on cellular organisms.**
- **However, cellular genomes are autonomous.**
- **While cellular organisms replicate their DNA through well-understood processes such as mitosis and meiosis, viruses use rapid and sometimes error-prone methods of replication, often leading to significant genetic diversity.**



# **A Viral Genomeic Example: SRAS-CoV-2**



- **The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is the virus responsible for COVID-19.**
- **SARS-CoV-2 was first identified in December 2019, and rapid sequencing of its genome has allowed scientists to understand its structure, mutations, and evolutionary relationships.**
- **Its accession number is NC045512.**
- **Searching on GenBank, the first page is in next slide.**



Nucleotide

Nucleotide

NC045512

Search

Advanced

Help

GenBank

Send to:

Change region shown

Customize view

# Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC\_045512.2

[FASTA](#) [Graphics](#)

Go to: ☒

LOCUS NC\_045512 29903 bp ss-RNA linear VRL 18-JUL-2020  
DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome.  
ACCESSION NC\_045512  
VERSION NC\_045512.2  
DBLINK BioProject: [PRJNA485481](#)  
KEYWORDS RefSeq.

Analyze this sequence

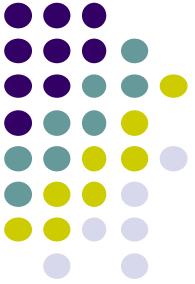
Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

NCBI Virus



# Thank you

for your attention