



Conceptos Básicos de Bioinformática

Otoño 2022

Oscar Salgado S.
oscarsalgado@unach.cl

Laboratorio de Bioinformática
Dirección de Investigación
UnACh

Genómica, Metagenómica y Técnicas de secuenciación masiva

Dr. Andrés Marcoleta

Grupo de Microbiología Integrativa, Departamento de Biología
Facultad de Ciencias, Universidad de Chile

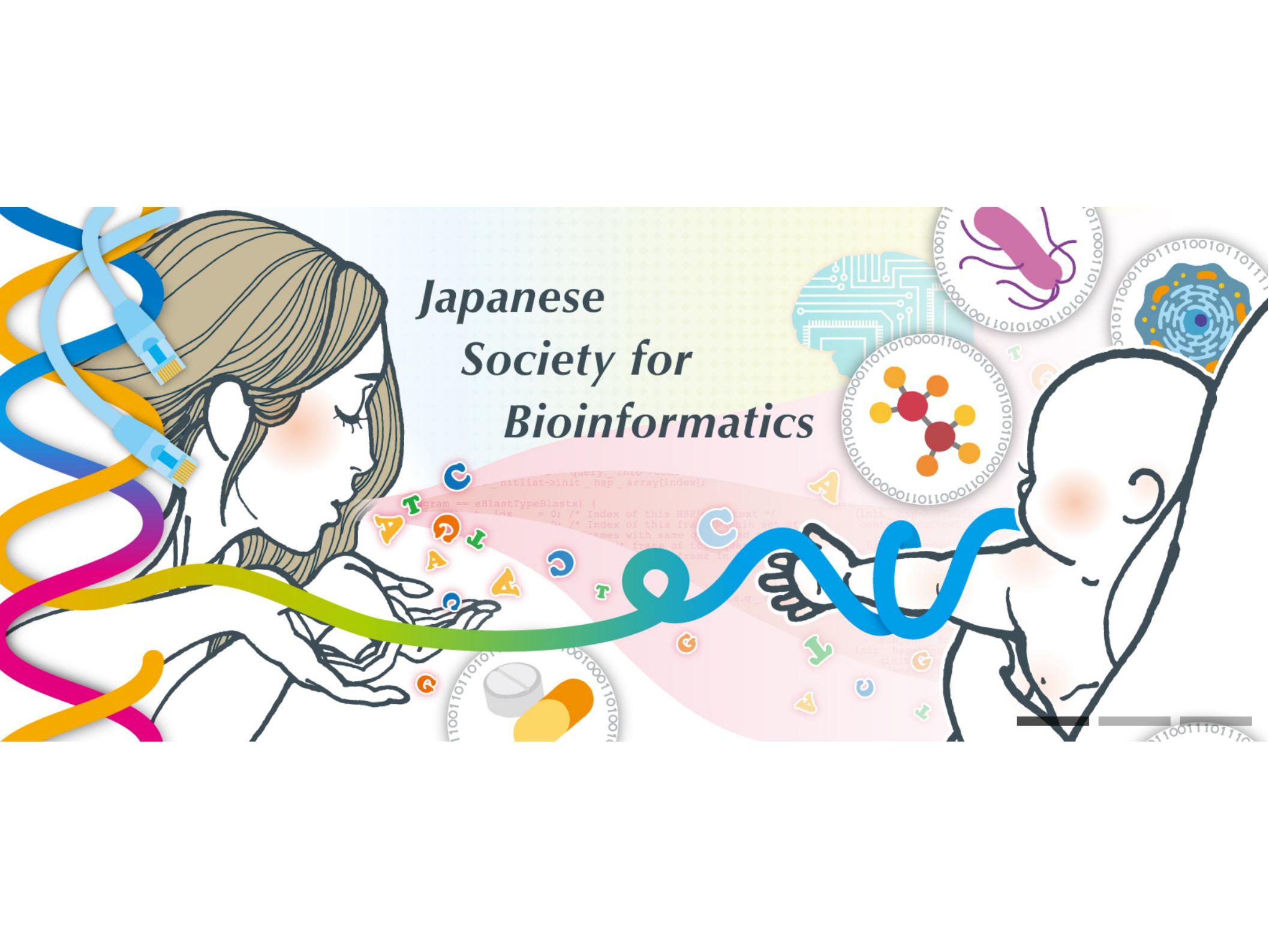
<https://github.com/GMI-Lab/gmi-workshop>

Bioinformatics



<https://es.coursera.org/learn/bioinformatics-pku#syllabus>

Bioinformatics: Introduction and Methods



Japanese Society for Bioinformatics

```
query_info->...  
_initlist->init_hsp_array[index];  
  
if (gram == eBlastTypeBlastx) {  
    iix = 0; /* Index of this HSPM */  
    iox = 0; /* Index of this frame in the thin set of frames with same query */  
    if (frame_in == 0) {  
        init_hsp_array[index]....  
    } else {  
        ...  
    }  
}
```



Bioinformática

- Bioinformatics is an interdisciplinary science that allows one to understand and evaluate knowledge in the field of biology, developing a hypothesis using biological information.

Isea, 2015

Bioinformática

- La bioinformática es una ciencia **interdisciplinaria** que me permite comprender y evaluar el conocimiento en el campo de la biología, desarrollando [probando] una hipótesis con el uso de información biológica.

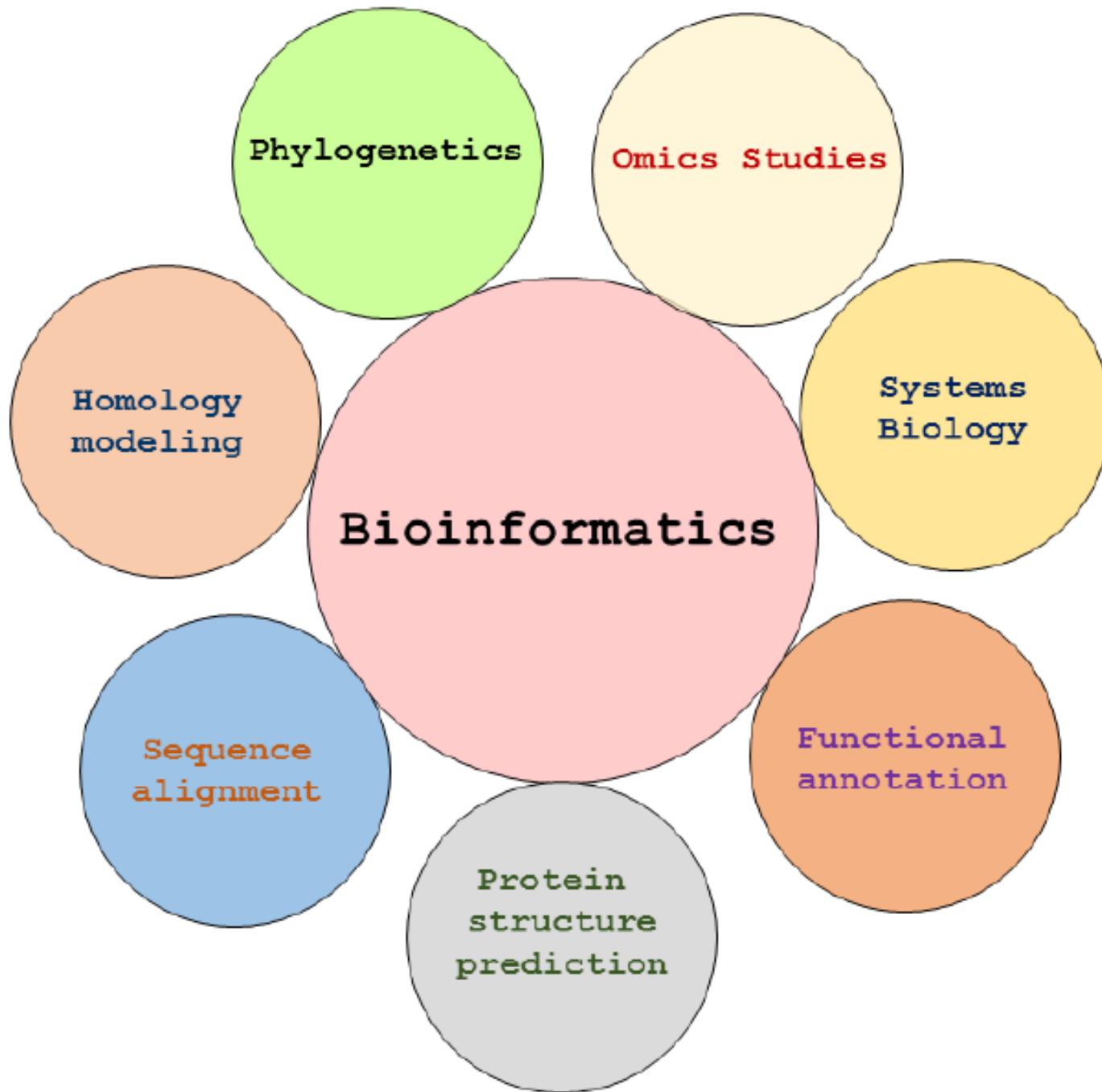
Isea, 2015

Bioinformática

- Biología
- Química
- Física
- Ciencias de la computación
- Ingeniería
- Matemática
- Estadística

Bioinformática

- La bioinformática es una disciplina científica que permite generar conocimiento en ciencias biológicas mediante el uso de herramientas computacionales.



La bioinformática trabaja principalmente con letras

- En el caso del ADN, cada letra representa una base nitrogenada
- Varias bases son una secuencia (puede o no tener significado biológico)

ATGATCGAGATATCCGG

Symbol	Meaning	Explanation
a	a	adenine
c	c	cytosine
g	g	guanine
t	t	thymine in DNA; uracil in RNA
m	a or c	amino
r	a or g	purine
w	a or t	
s	c or g	
y	c or t	pyrimidine
k	g or t	keto
v	a or c or g	not t
h	a or c or t	not g
d	a or g or t	not c
b	c or g or t	not a
n	a or c or g or t	any

[References]

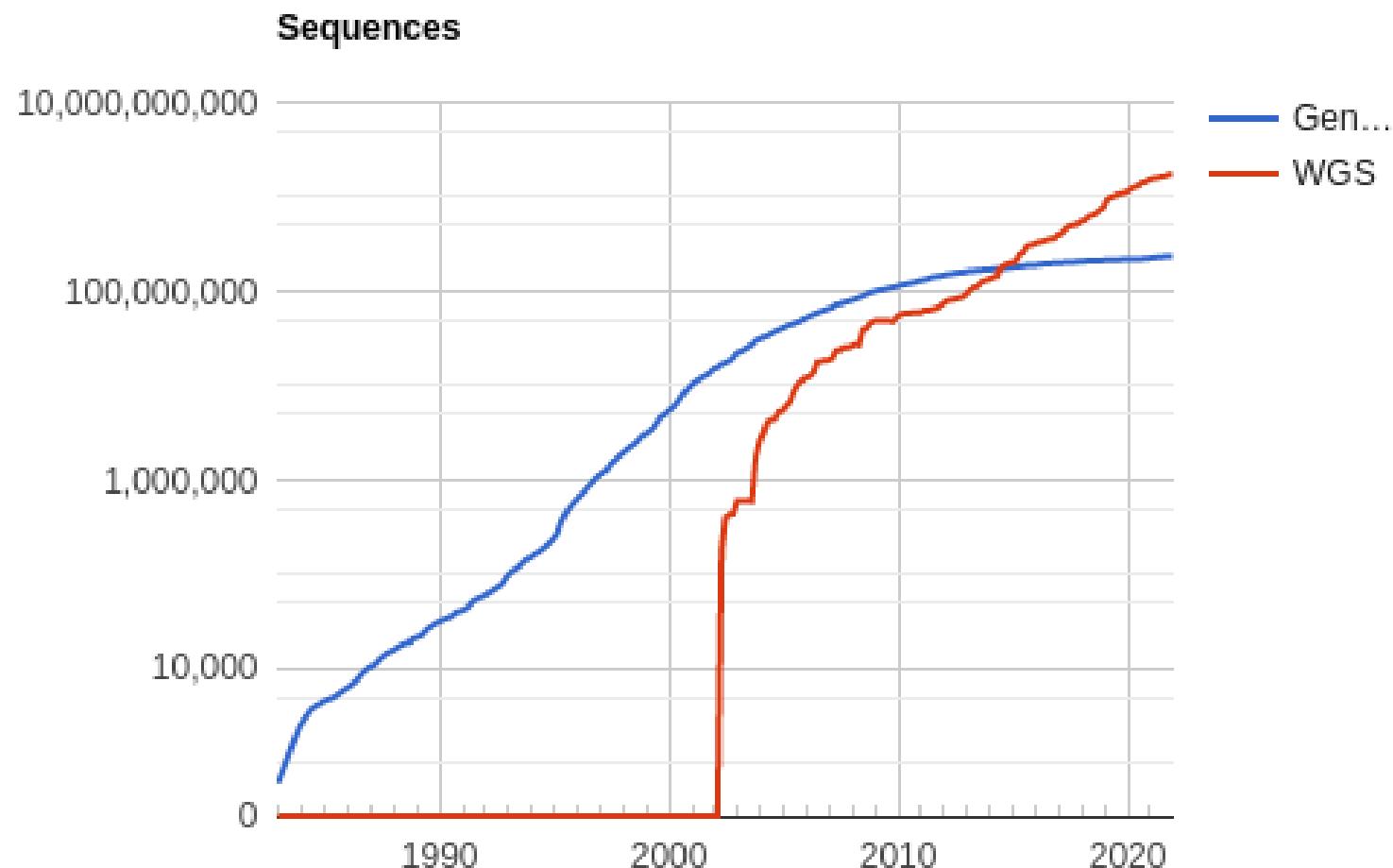
- Cornish-Bowden, A. Nucl Acid Res 13, 3021-3030 (1985)

Abbreviation	1 letter abbreviation	Amino acid name
Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic acid
Cys	C	Cysteine
Gln	Q	Glutamine
Glu	E	Glutamic acid
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Pyl	O	Pyrrolysine
Ser	S	Serine
Sec	U	Selenocysteine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
Val	V	Valine
Asx	B	Aspartic acid or Asparagine
Glx	Z	Glutamic acid or Glutamine
Xaa	X	Any amino acid
Xle	J	Leucine or Isoleucine
TERM		termination codon

[References]

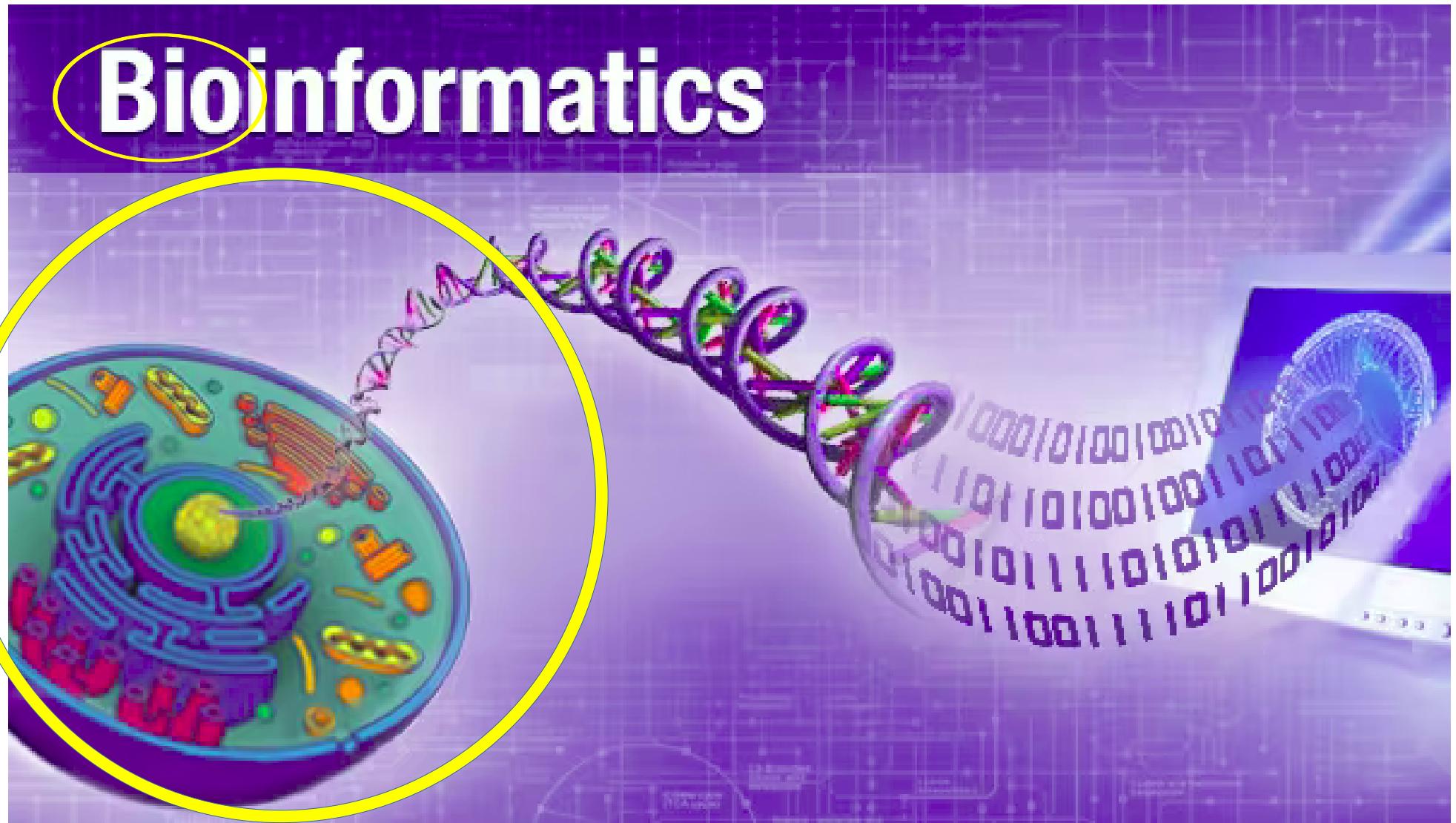
- IUPAC-IUB Joint Commission on Biochemical Nomenclature. Nomenclature and Symbolism for Amino Acids and Peptides. Eur. J. Biochem. 138: 9-37 (1984).

Número de secuencias en dos bases de datos: GenBank y WGS



<https://www.ncbi.nlm.nih.gov/genbank/statistics/>

BIO

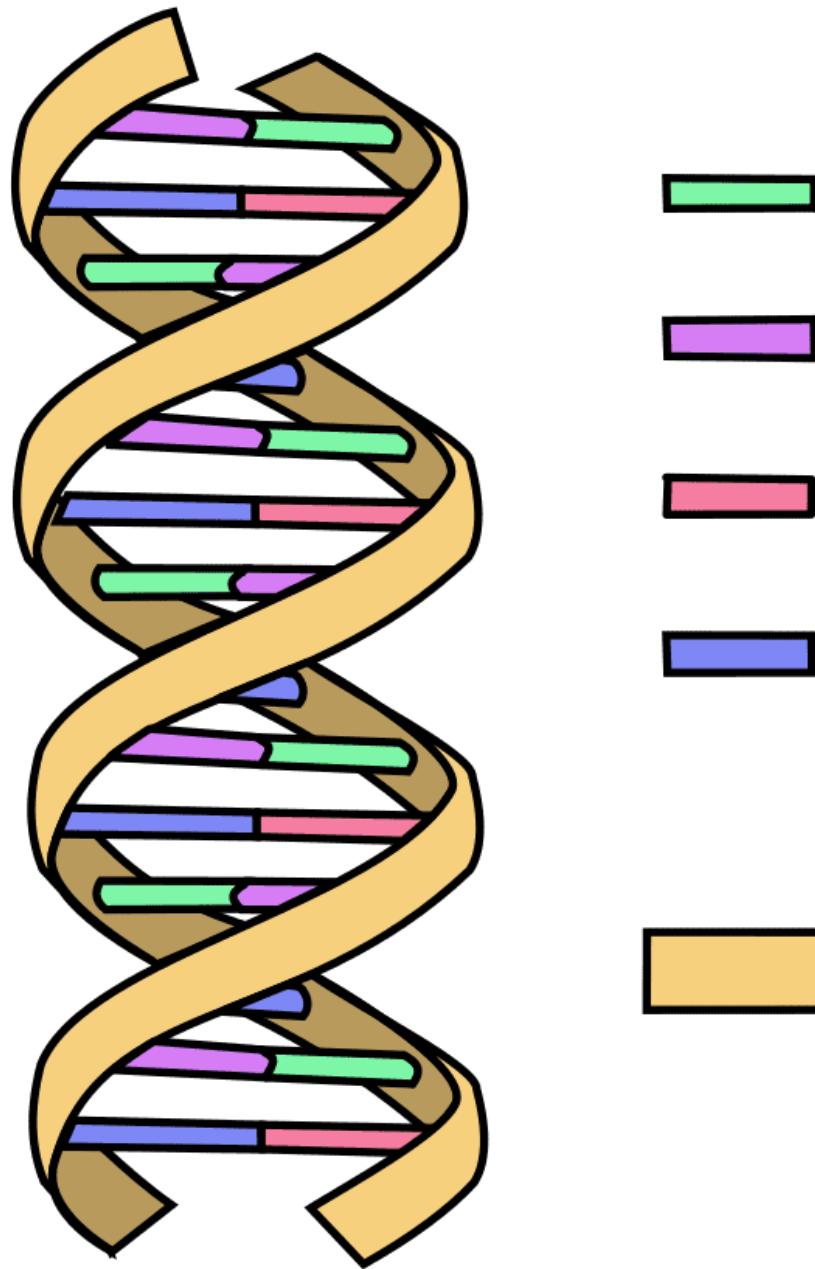


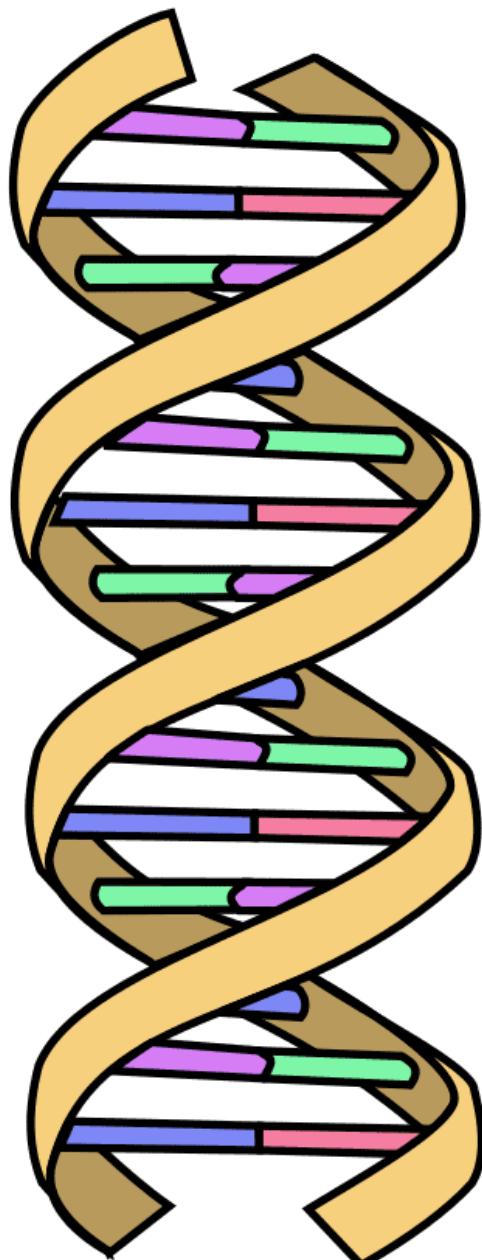
Bioinformatics



*Te advertí que no
le preguntáramos
de dónde vienen
los bebés...*

¿Qué ven?





DNA

-  = Adenine
-  = Thymine
-  = Cytosine
-  = Guanine
-  = Phosphate backbone

The structure of DNA by direct imaging

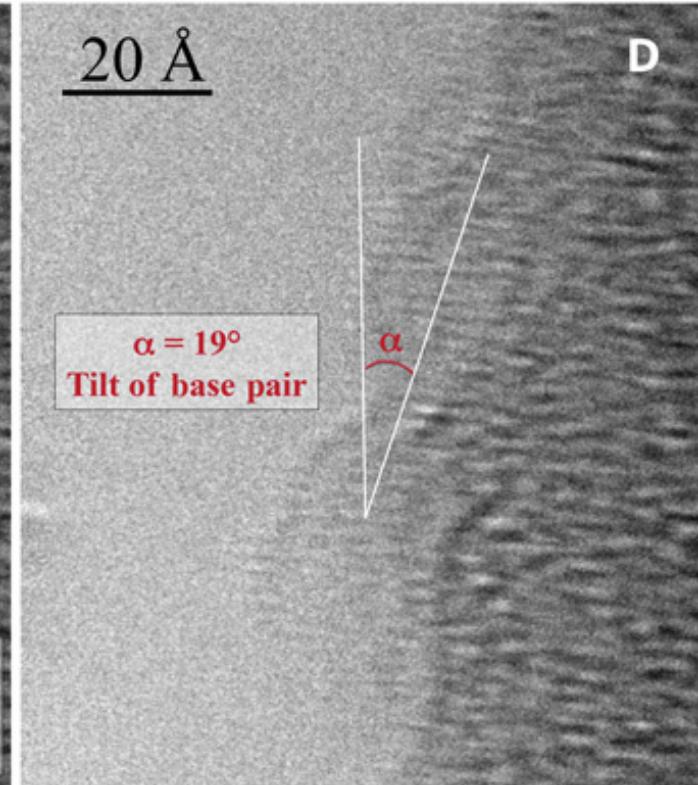
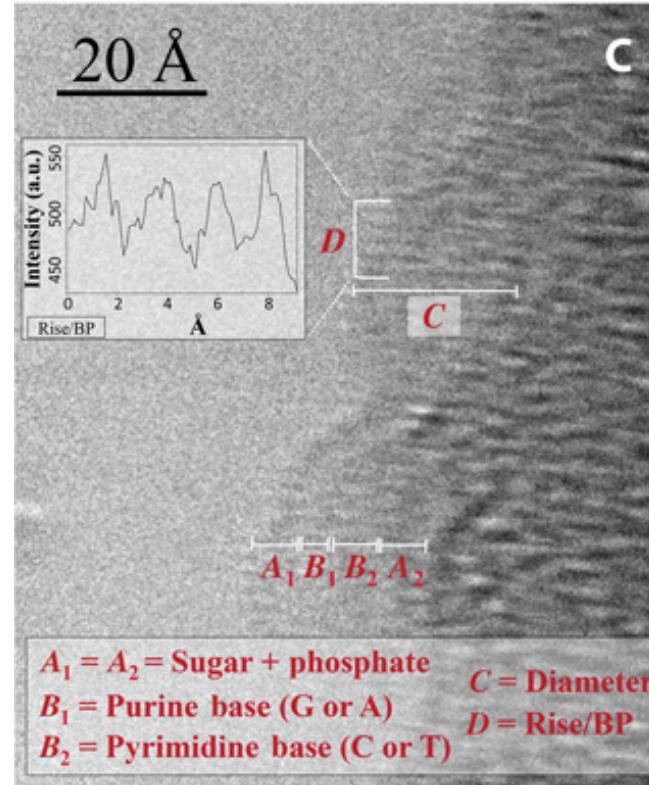
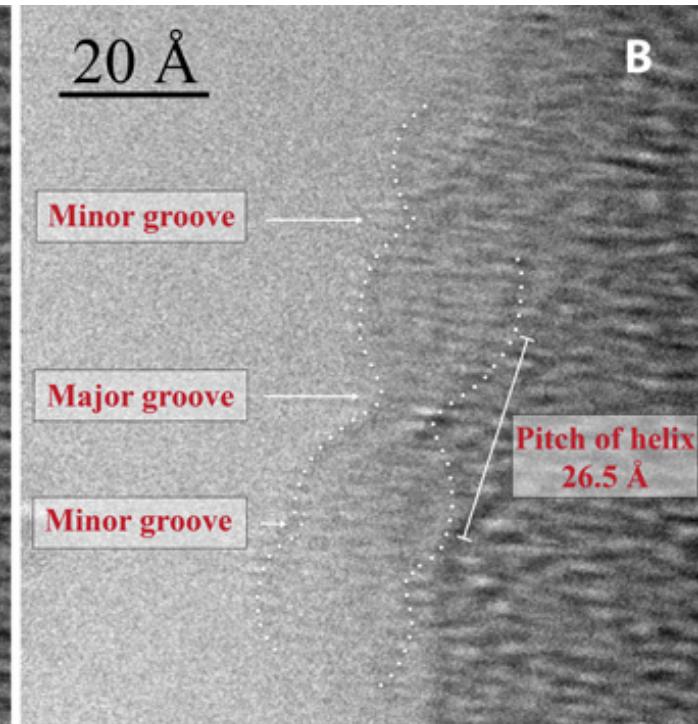
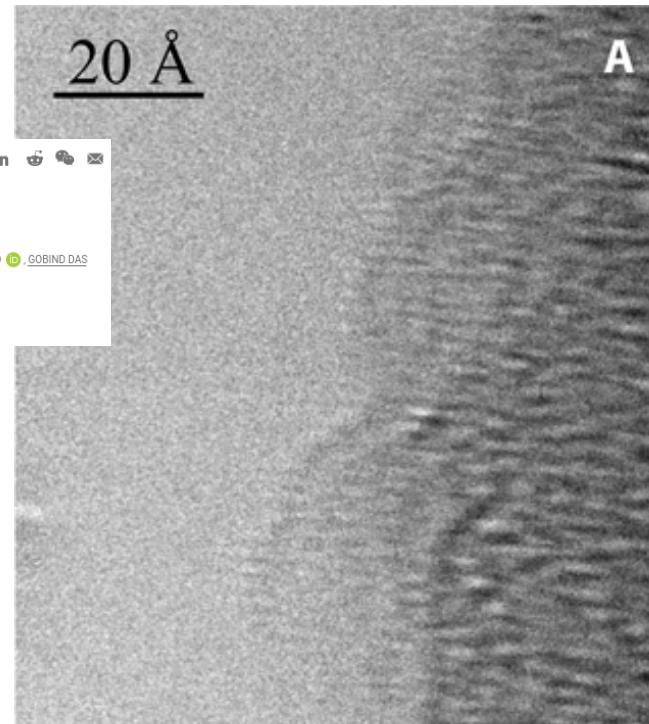
MONICA MARINI, ANDREA FALQUI, MANOLA MORETTI, TANIA LIMONGI, MARCO ALLIONE, ALESSANDRO GENOVESE, SERGEI LOPATIN, LUCA TIRINATO, GOBIND DAS

ENZO DI FABRIZIO

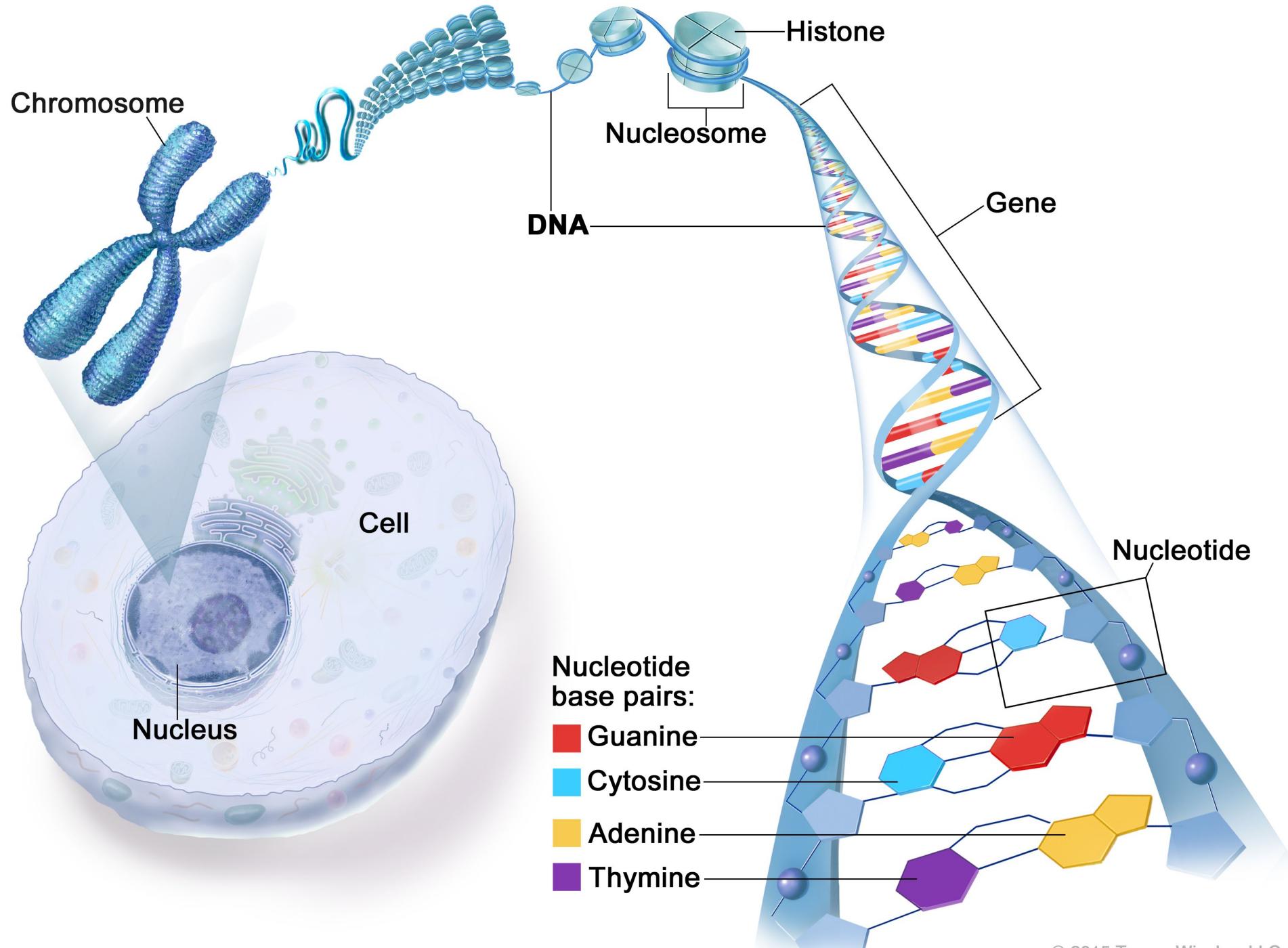
+5 authors

Authors Info & Affiliations

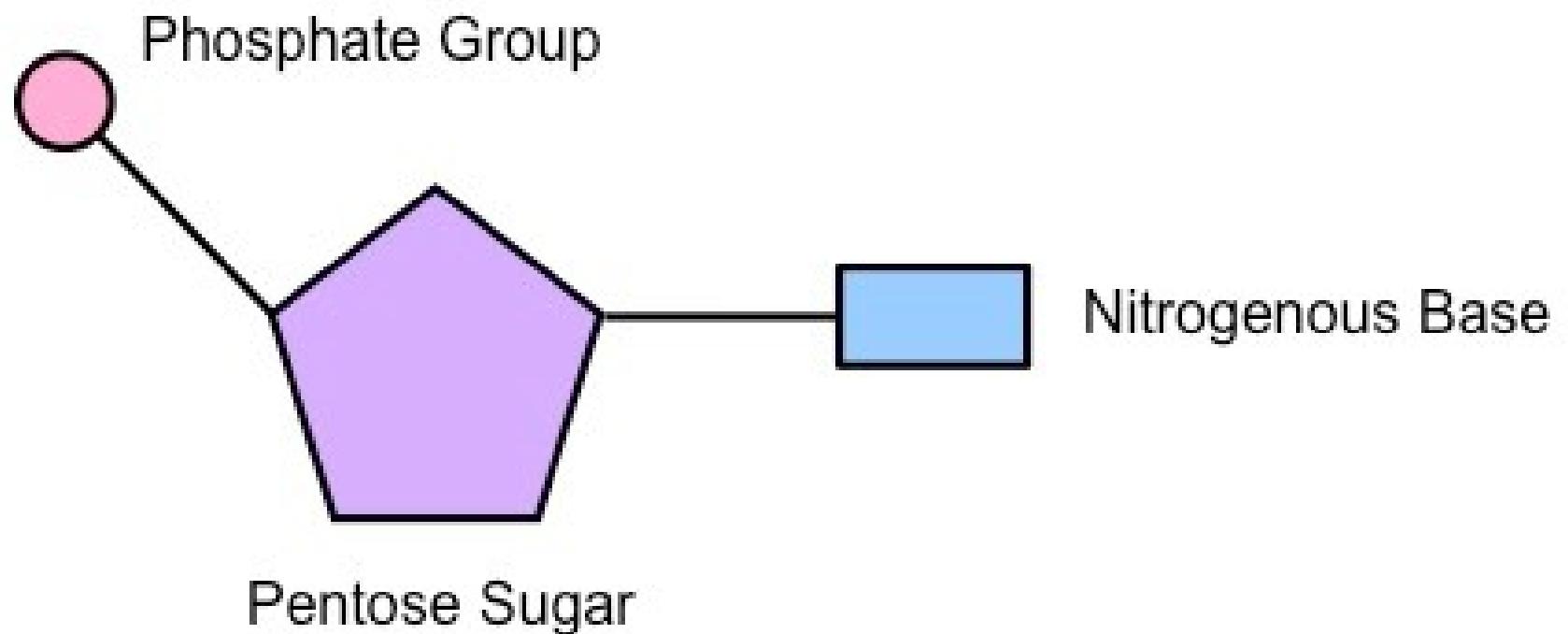
SCIENCE ADVANCES • 28 Aug 2015 • Vol 1, Issue 7 • DOI: 10.1126/sciadv.1500734



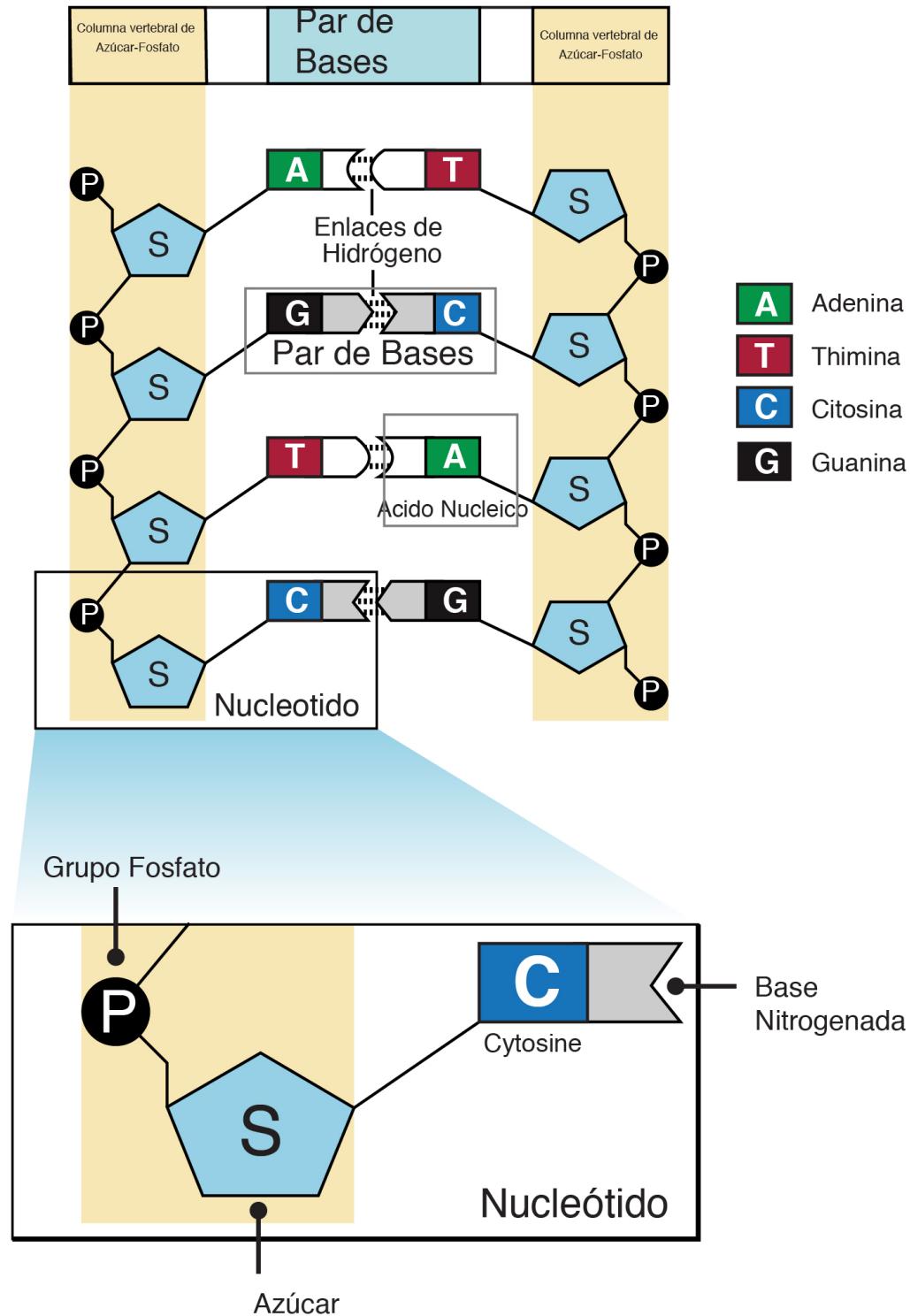
DNA Structure



NUCLEÓTIDO

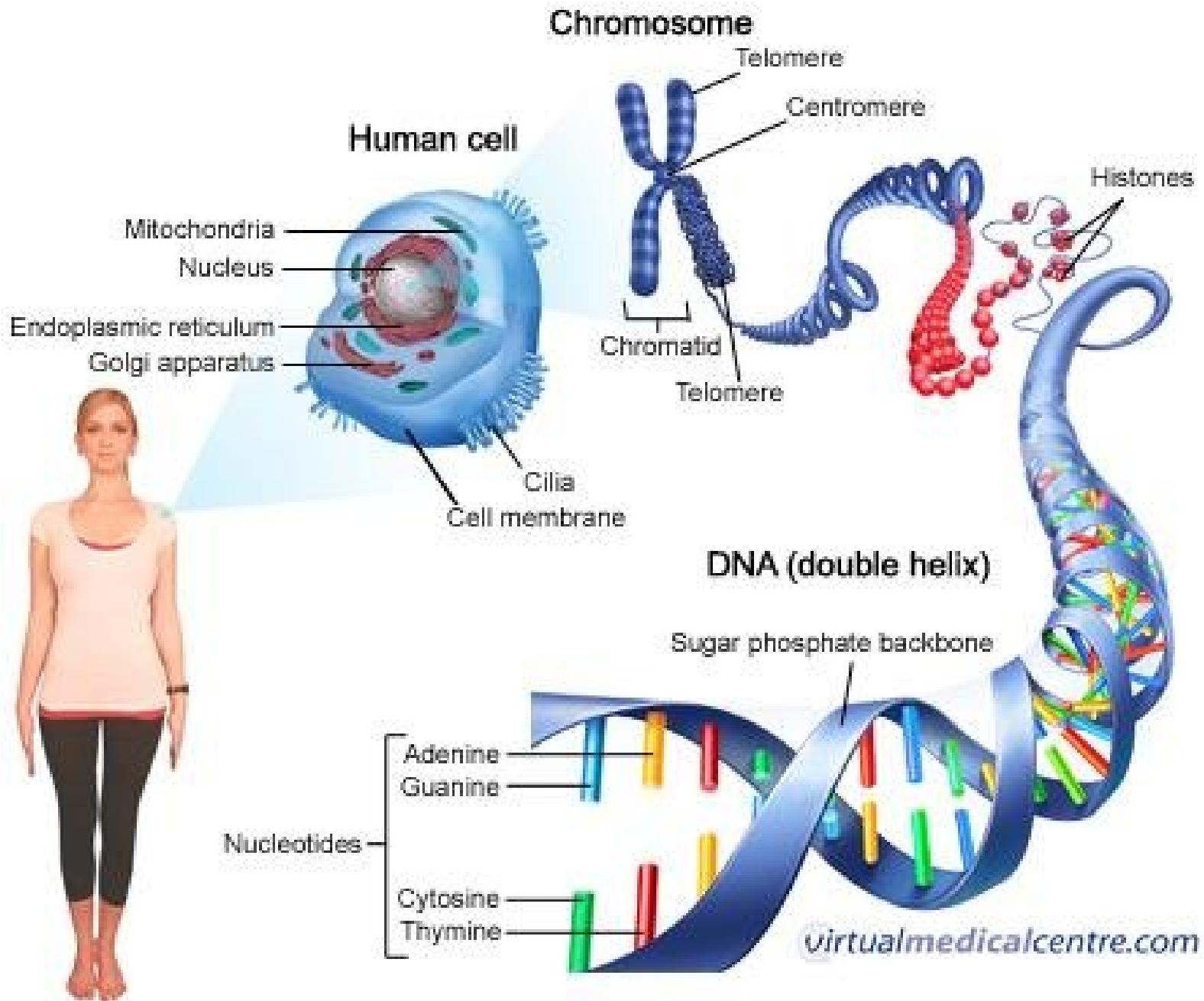


Acido Desoxirribonucleico (ADN)

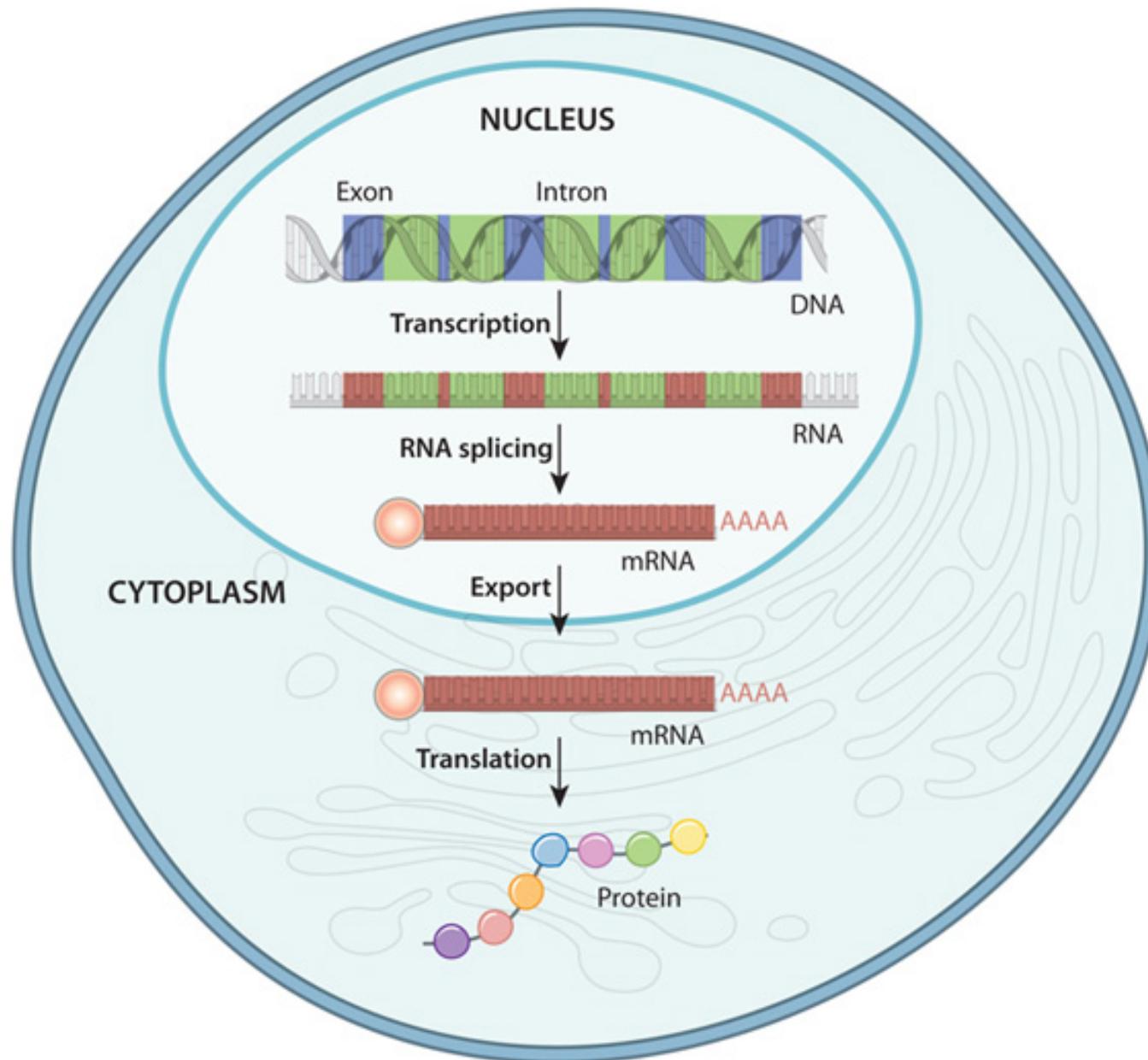


<https://www.genome.gov/es/genetics-glossary/Nucleotido>

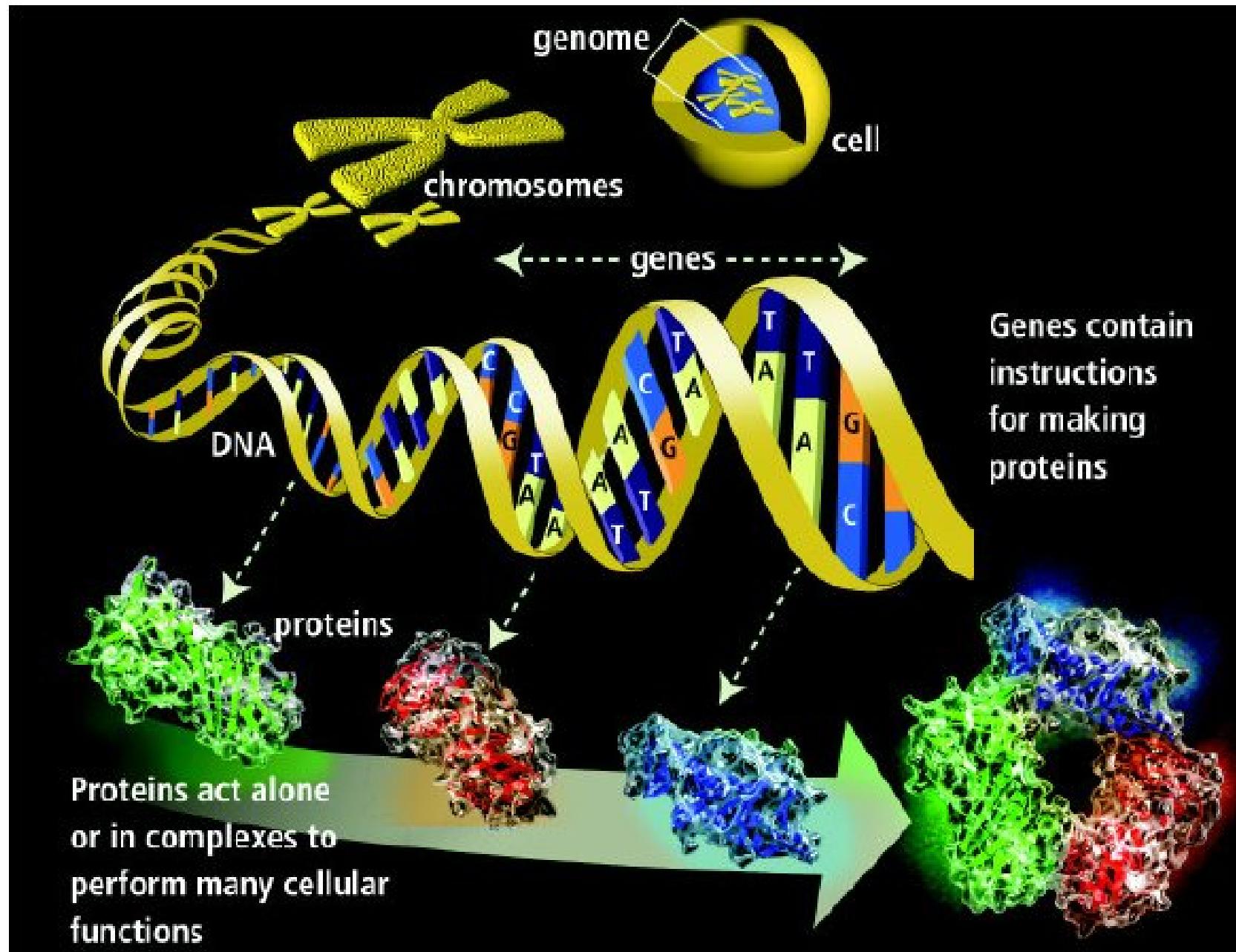
Ácido desoxirribonucleico



Ácido desoxirribonucleico



Ácido desoxirribonucleico





Aequorea victoria

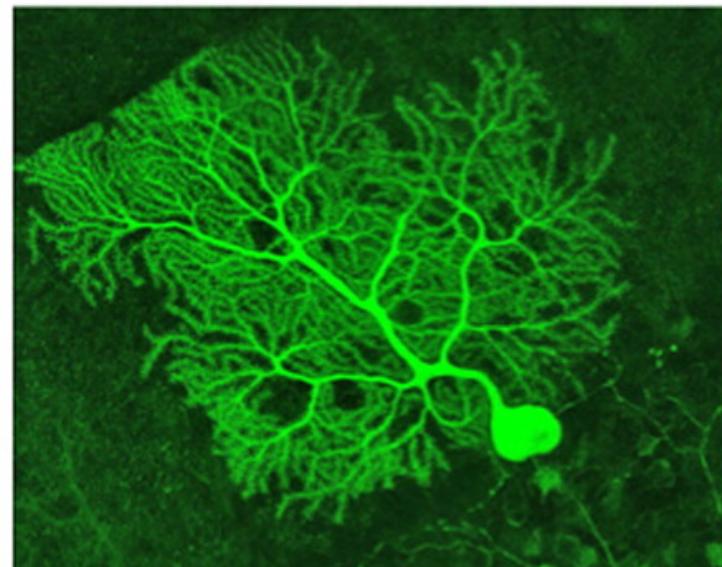
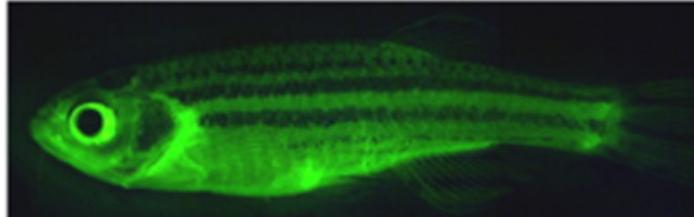
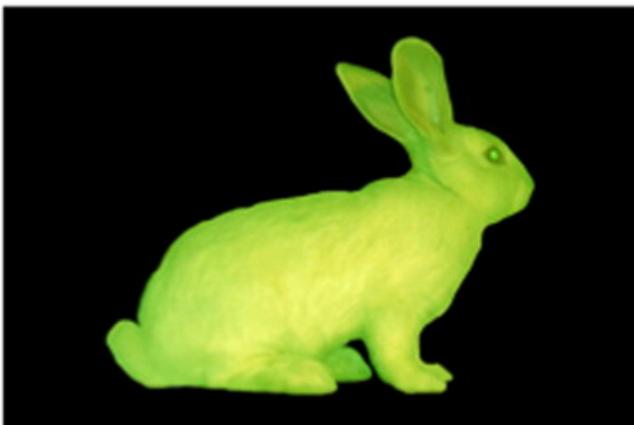
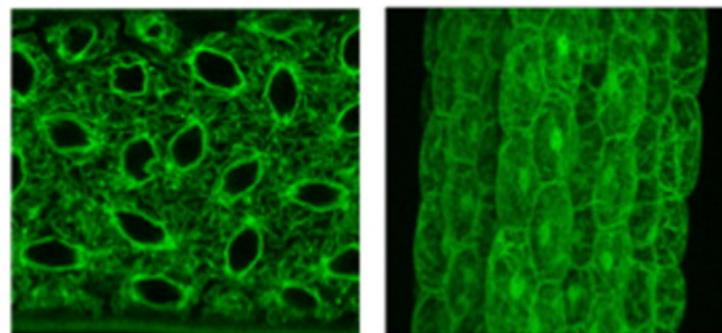
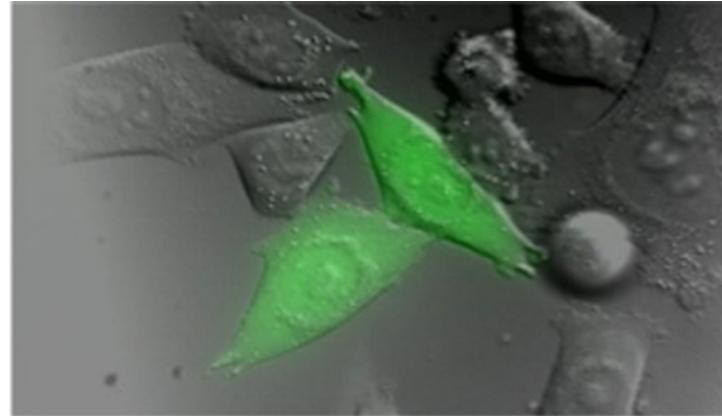
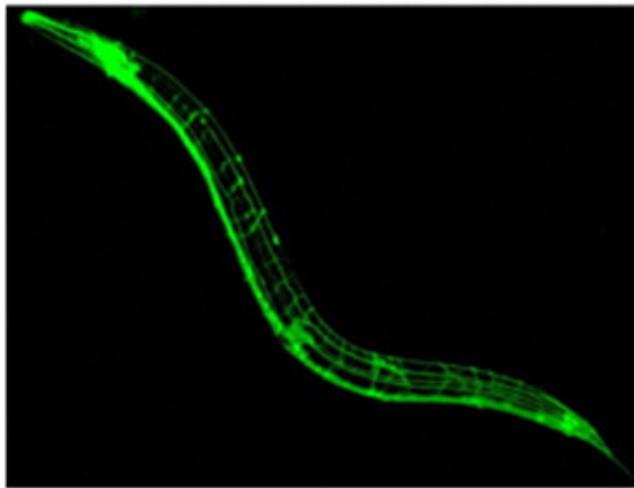
Proteína verde fluorescente Green fluorescent protein (*gfp*)

- Genes = genotipo
- Ambiente = entorno donde vive
- Genotipo + Ambiente = Fenotipo



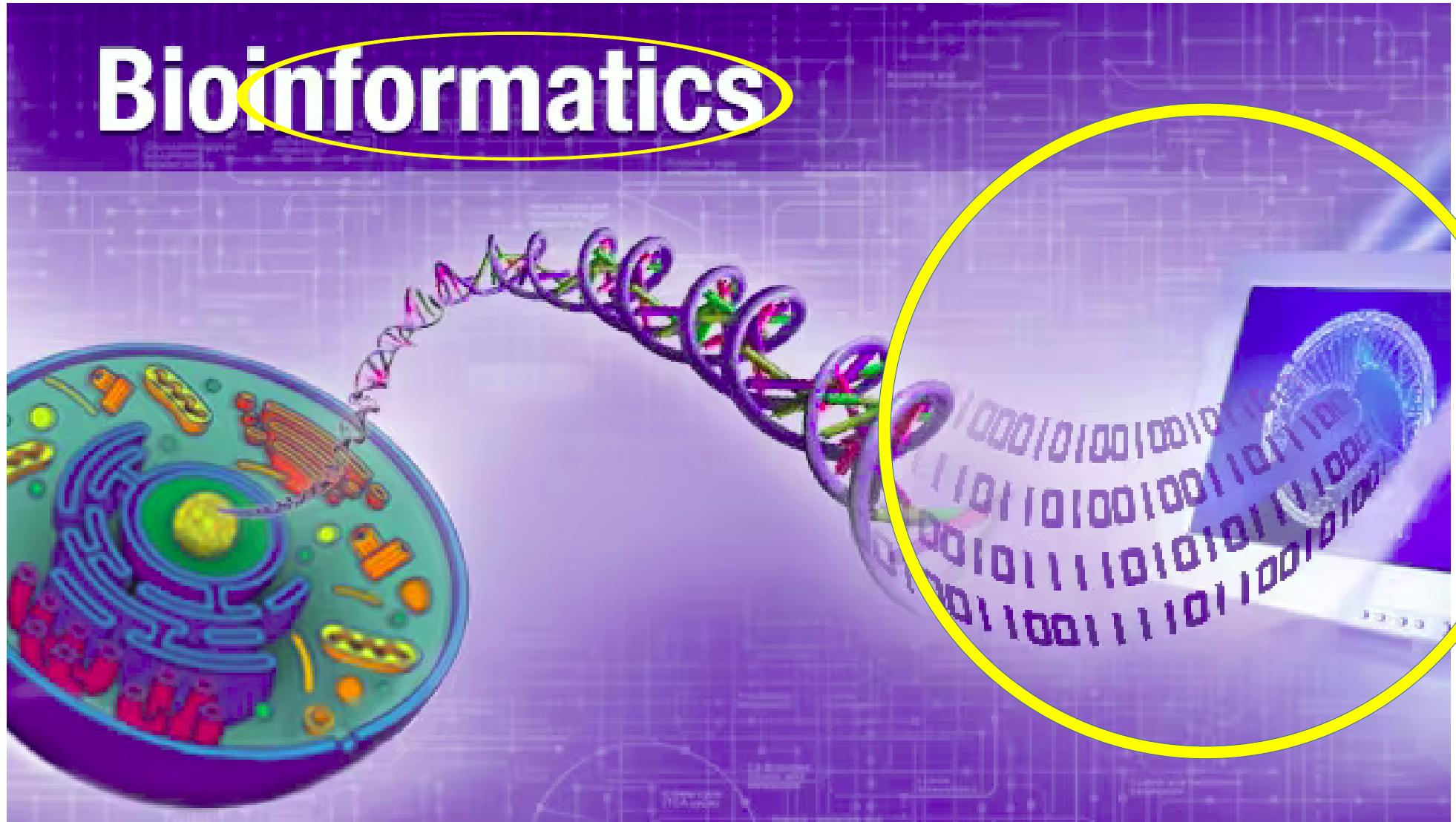
Cambio genético → Cambio fenotípico

<https://www.pnas.org/content/106/25/10073>

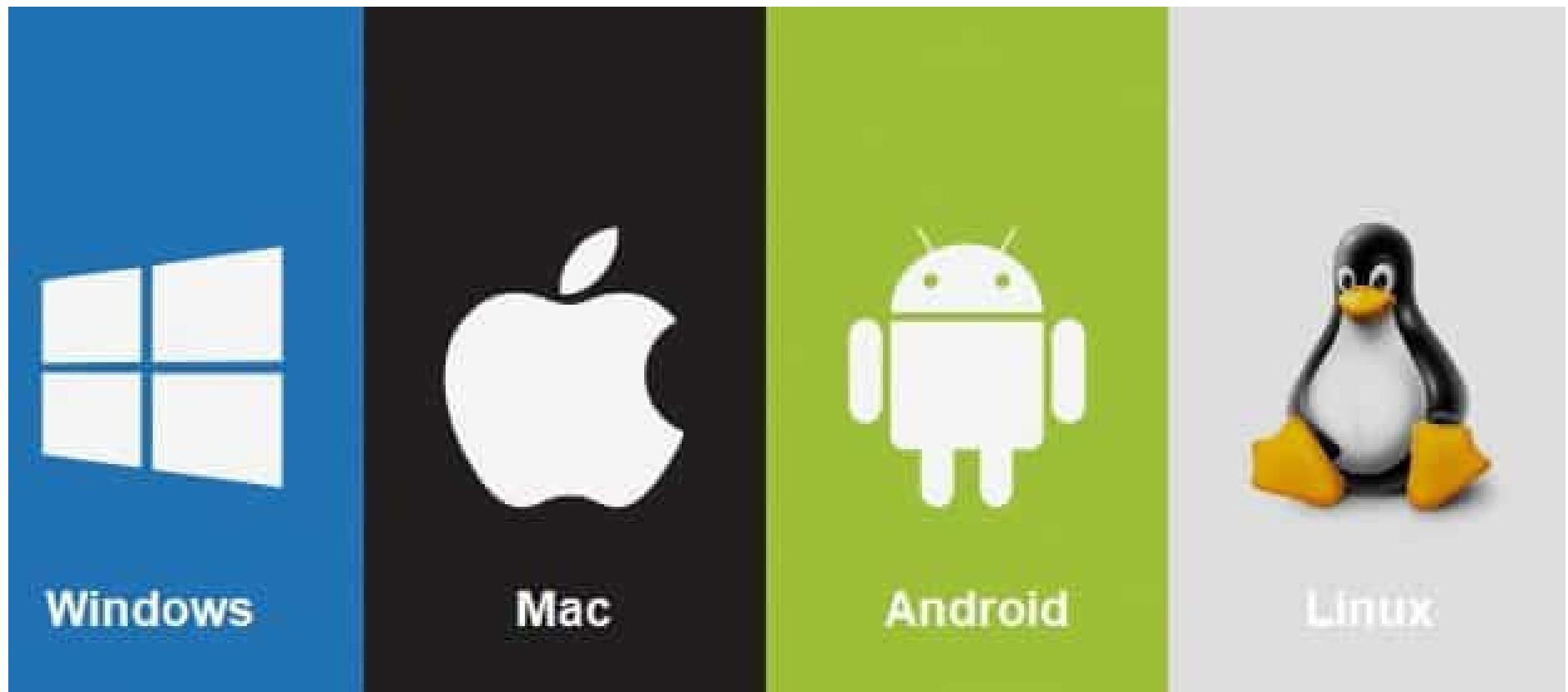


Informática

Bioinformatics



Sistemas operativos



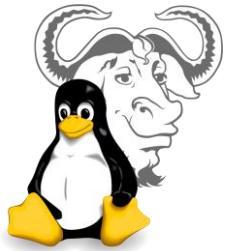


Curso Internacional de Genómica y Metagenómica
14 – 16 de diciembre 2021

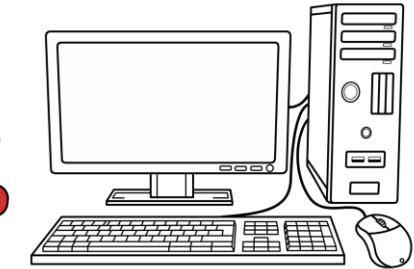
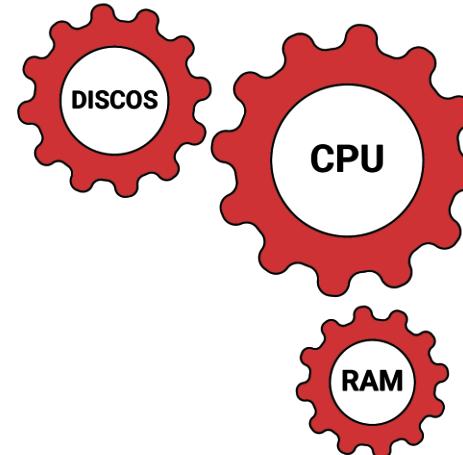
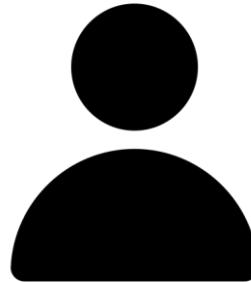
Introducción al trabajo en línea de comandos en UNIX

Dr.(c) Camilo Berríos P.

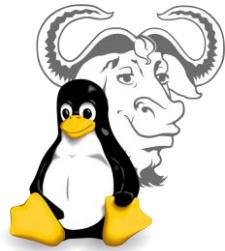




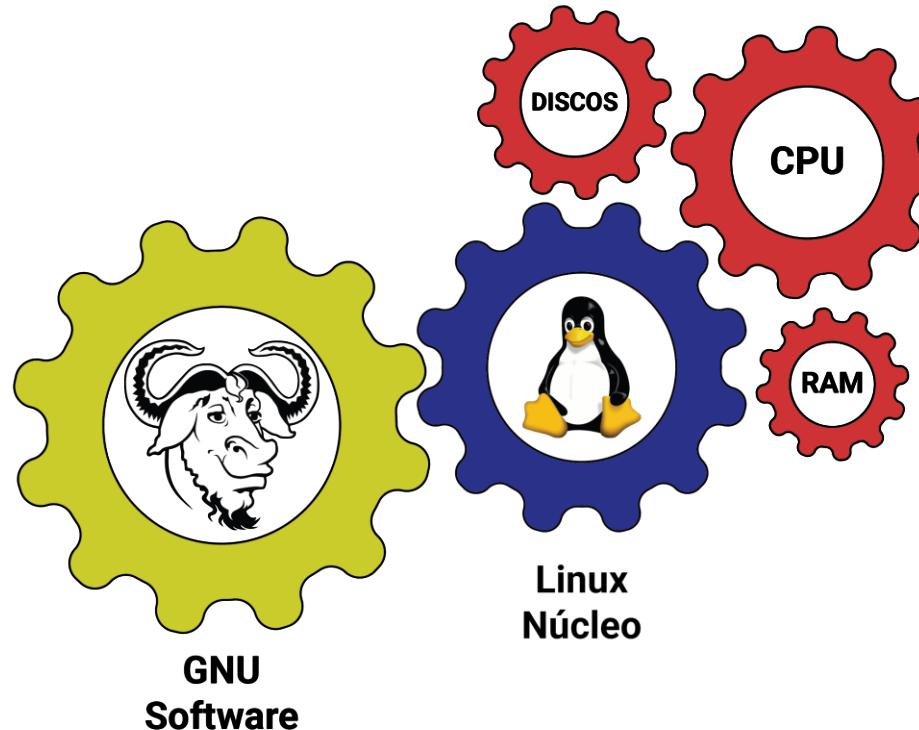
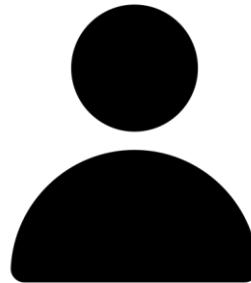
¿Por qué aprender **GNU/Linux**?



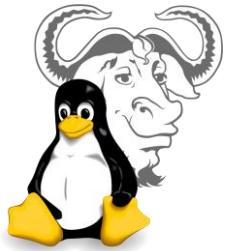
- **Proyecto GNU.** Anunciado por Richard Stallman en 1983. Proyecto colaborativo de **software libre** con el objetivo de crear un sistema operativo.



¿Por qué aprender GNU/Linux?



- **Núcleo (kernel) Linux.** Concebido por Linus Torvalds en 1991.



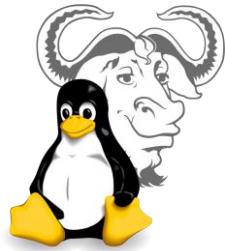
¿Por qué aprender **GNU/Linux**?

Multiusuario y Multitarea. Capaz de permitir servicios a más de un usuario a la vez. También, permite a los usuarios realizar varias labores al mismo tiempo.

Seguridad. Los sistemas basados en GNU/Linux son seguros por naturaleza.

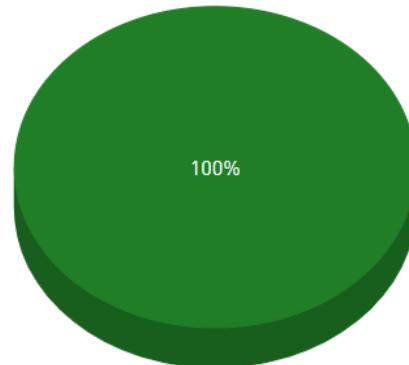
Ideal para programadores. Existen diversas herramientas para el desarrollo de programas con los lenguajes de programación más populares.

Fiabilidad. Código Libre. Privacidad...



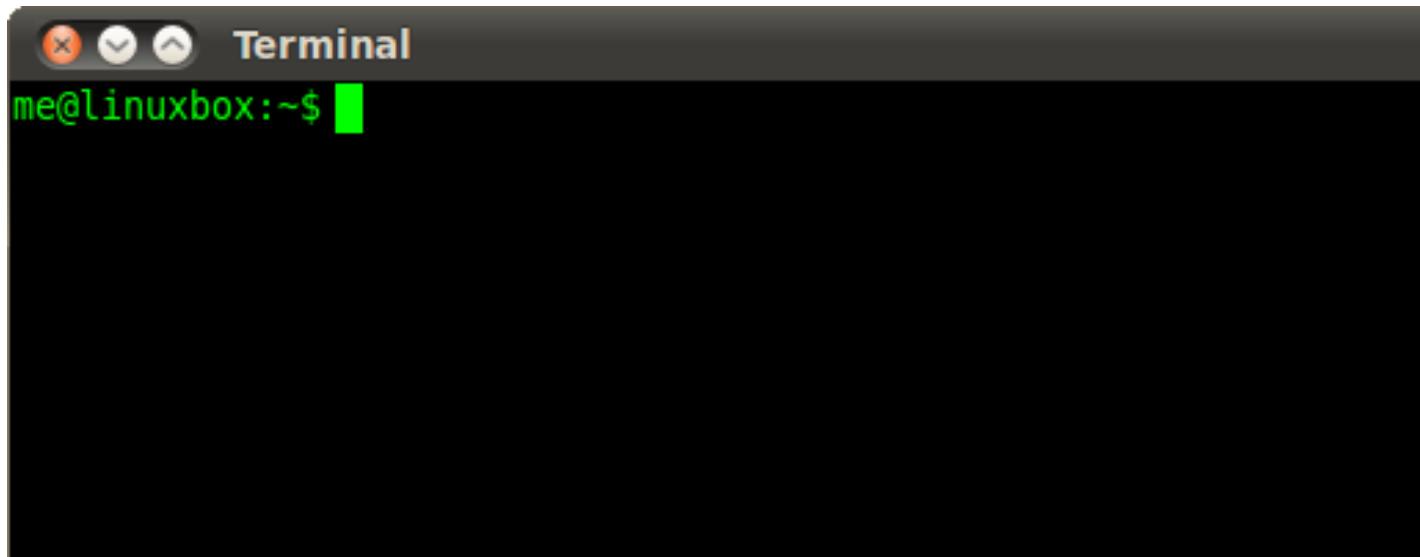
¿Por qué aprender GNU/Linux?

Diversas distribuciones.



Noviembre 2021
<https://www.top500.org/statistics/list/>

La línea de comandos



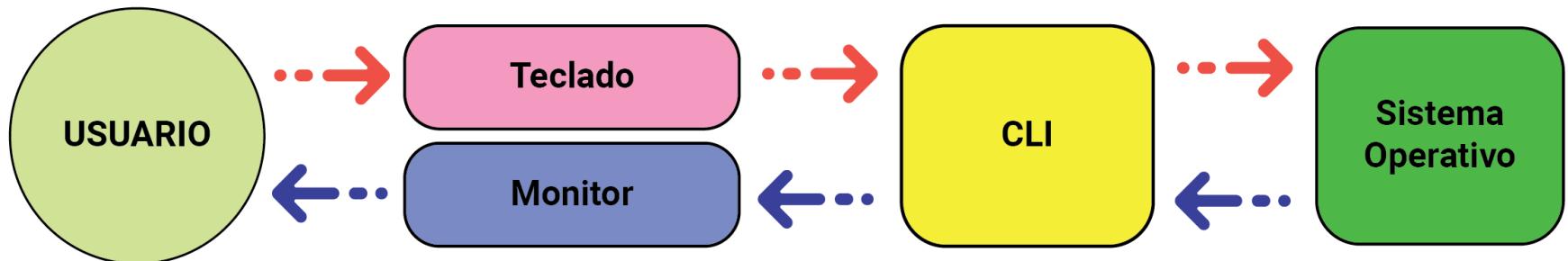
CLI, CLP, prompt, consola, terminal

La línea de comandos

```
me@linuxbox:~$
```

CLI, CLP, prompt, consola, terminal

La línea de comandos



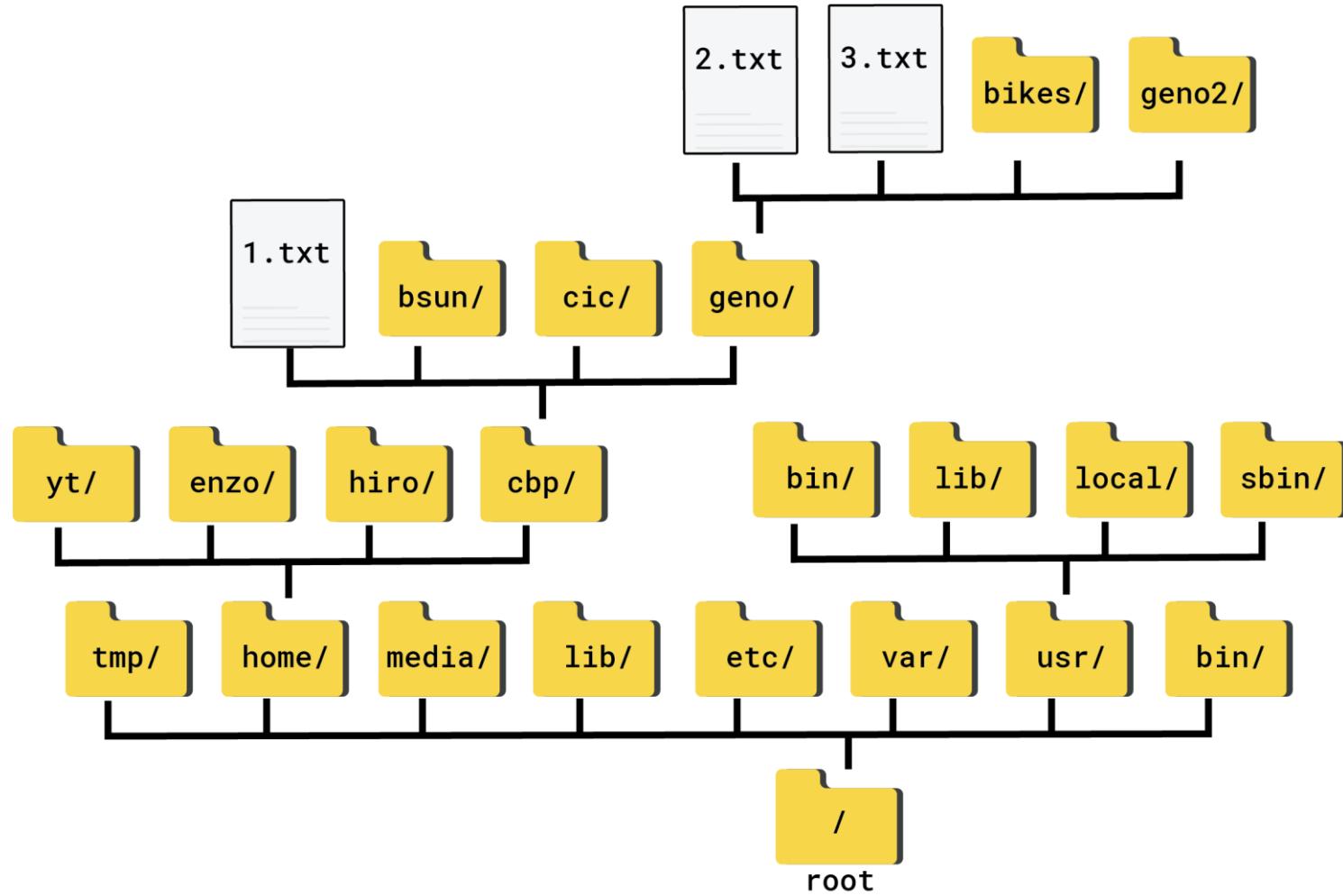
Comandos y opciones

```
camilo@Ada:~$ whoami
```

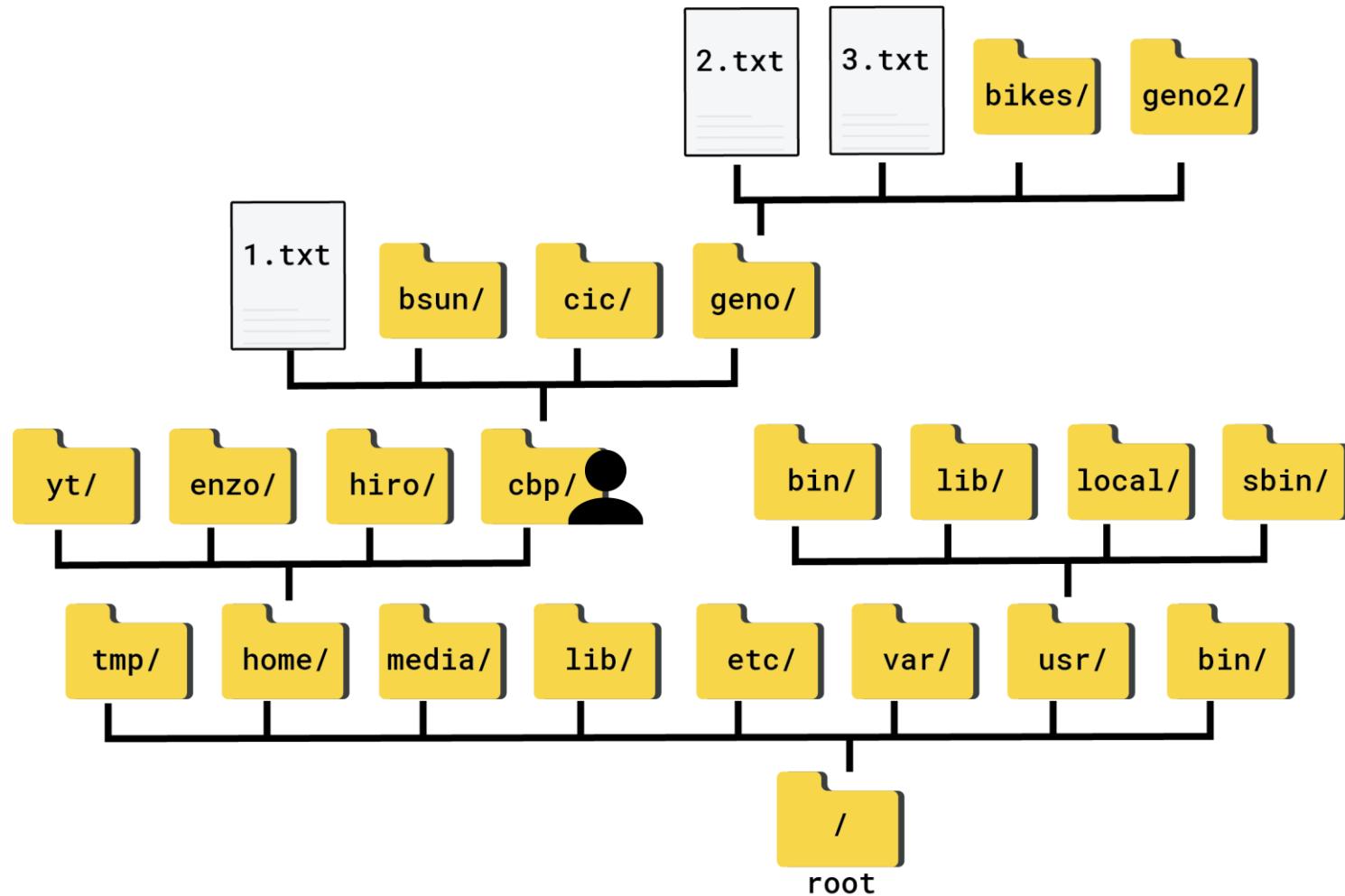
```
camilo@Ada:~$ who
```

```
camilo@Ada:~$ pwd
```

Estructura de Directorios



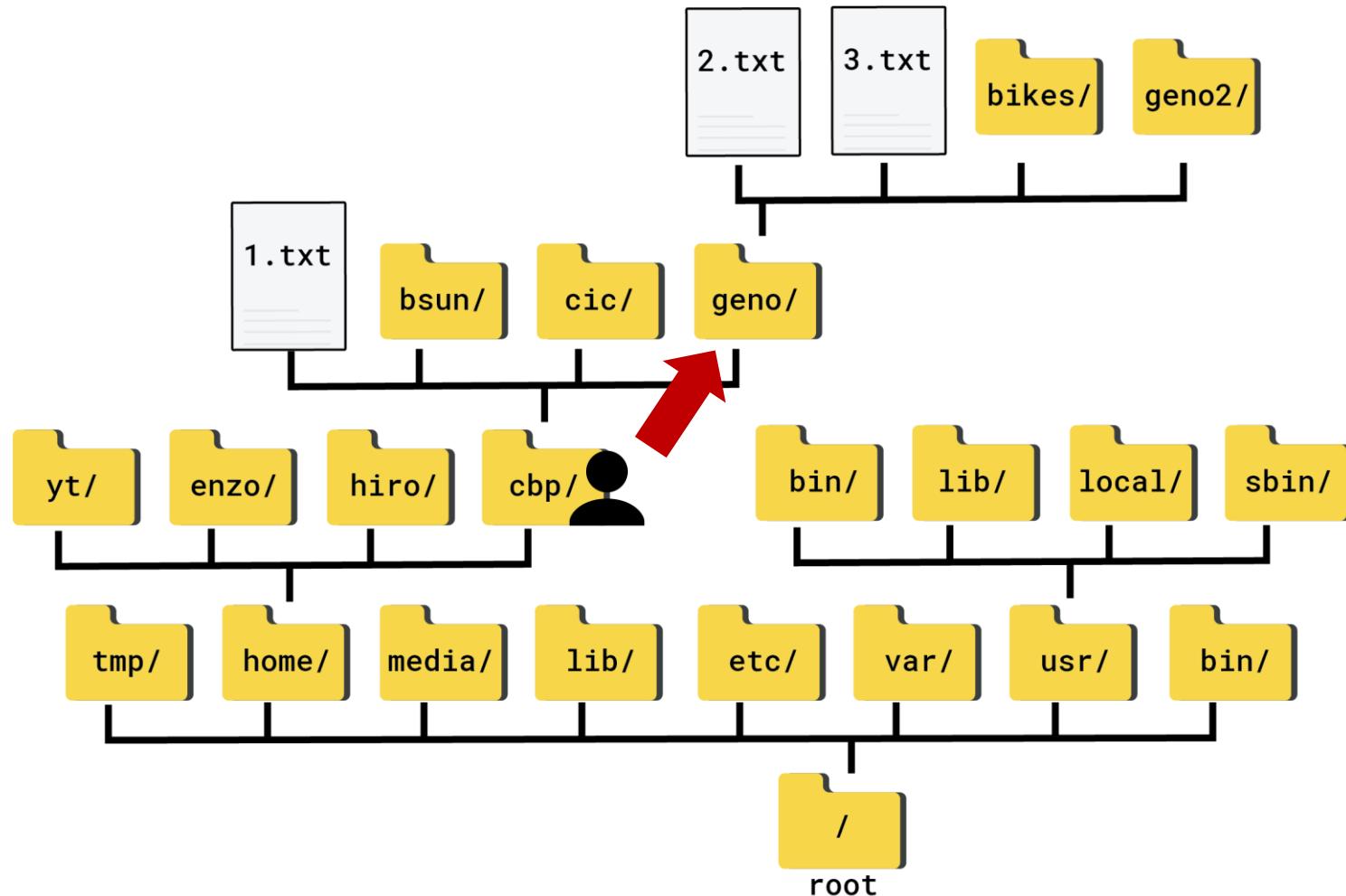
Estructura de Directorios



Comandos y opciones

```
cbp@tarro:~$ pwd  
/home/cbp/
```

Estructura de Directorios

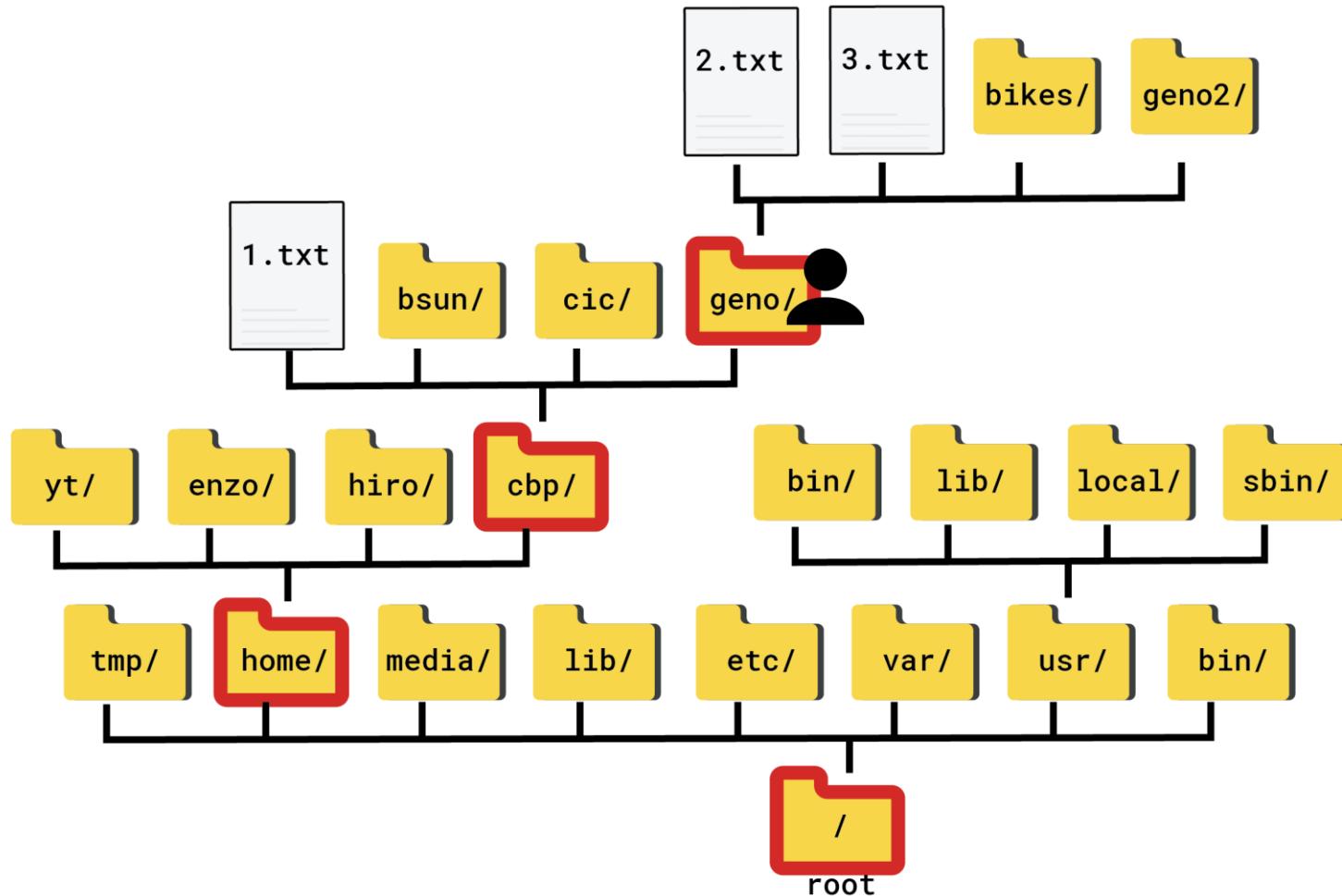


Comandos y opciones

```
cbp@tarro:~$ pwd  
/home/cbp/
```

```
cbp@tarro:~$ cd /home/cbp/geno/
```

Estructura de Directorios



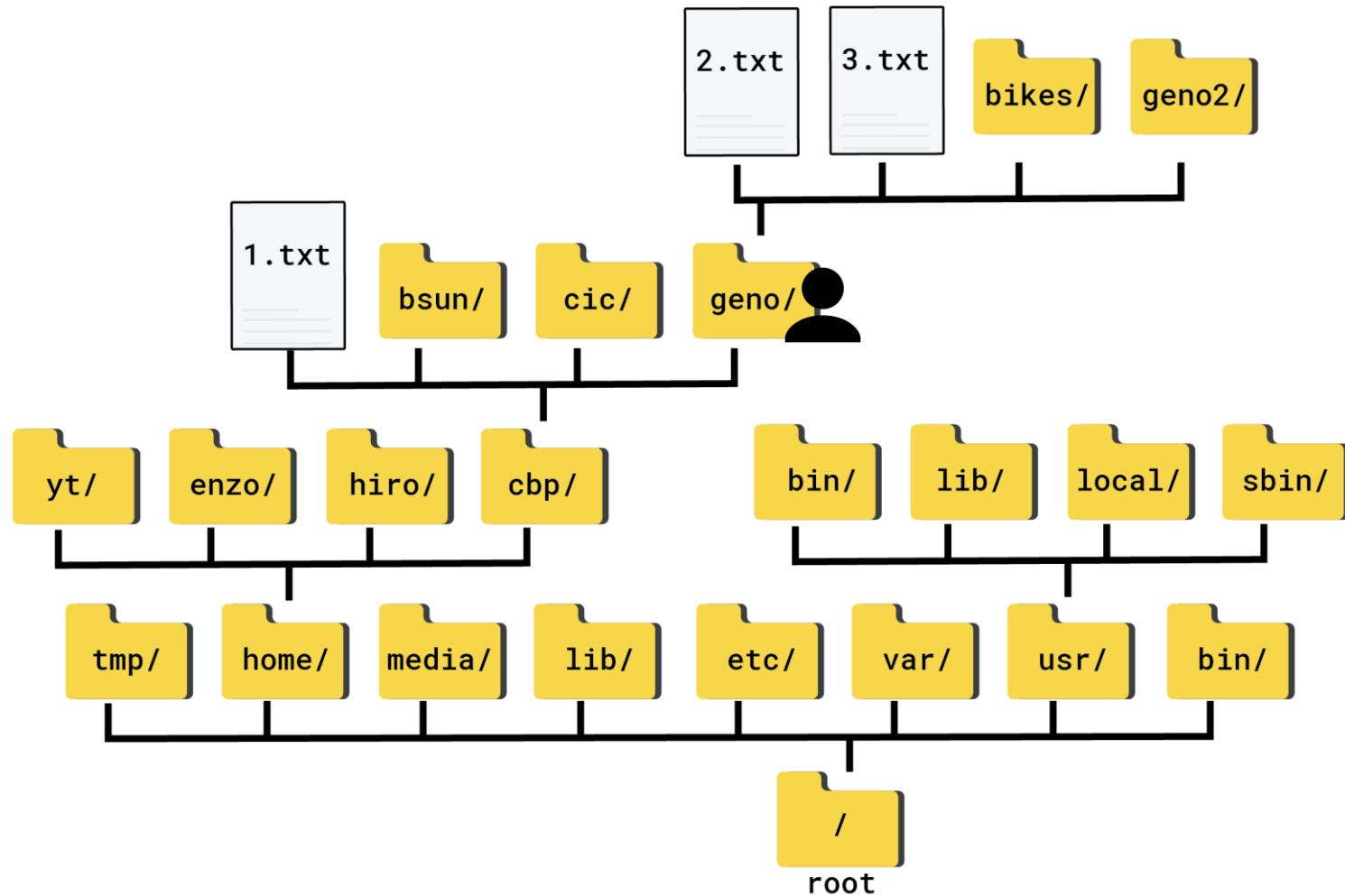
Comandos y opciones

```
cbp@tarro :~$ pwd  
/home/cbp/
```

```
cbp@tarro :~$ cd /home/cbp/geno/
```

```
cbp@tarro :~$ pwd  
/home/cbp/geno/
```

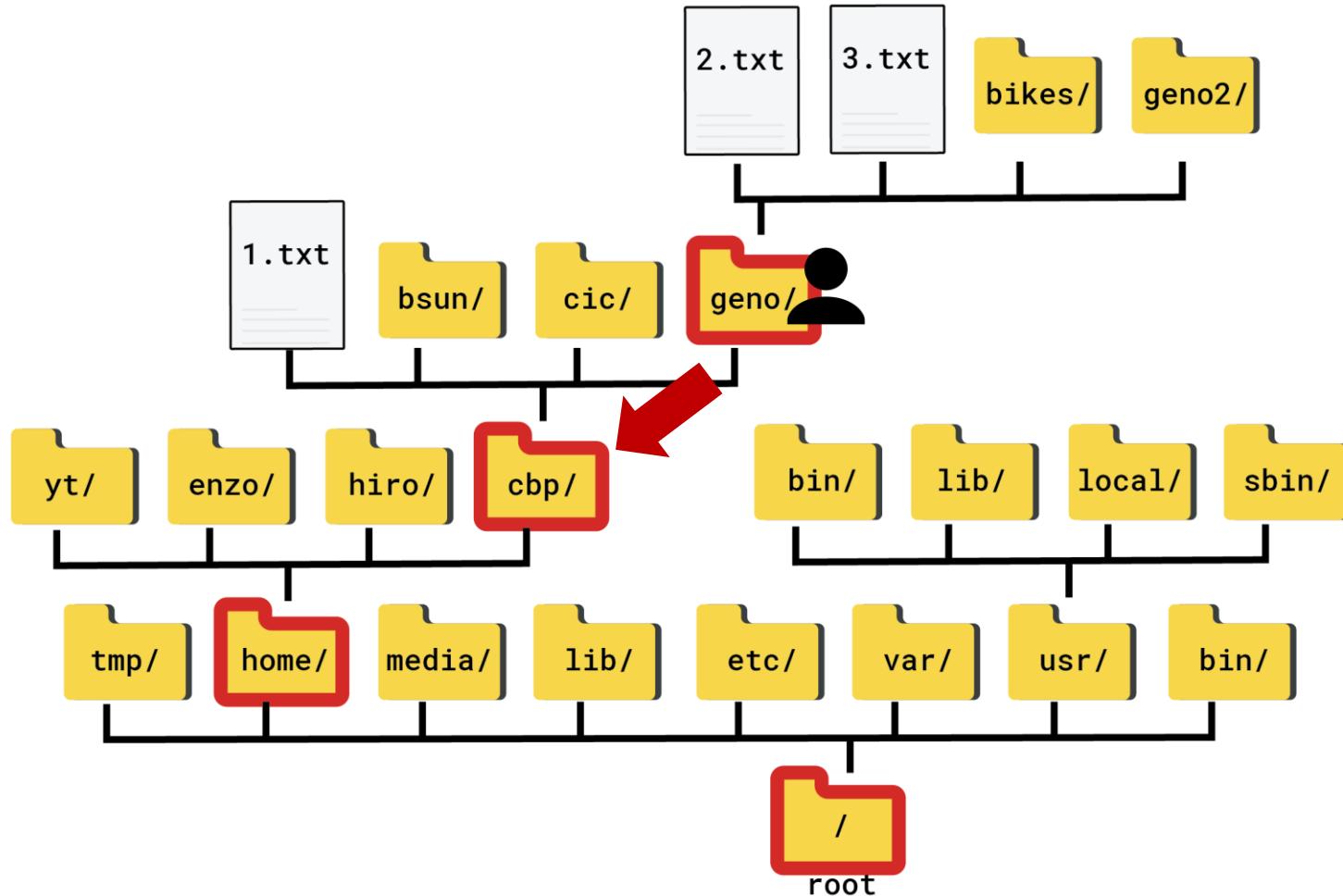
Estructura de Directorios



Listar archivos y directorios

```
cbp@tarro :~$ ls
bikes/
geno2/
2.txt
3.txt
```

Estructura de Directorios



Rutas (*paths*) absolutas y relativas

```
cbp@tarro :~$ pwd  
/home/cbp/geno/
```

Ruta absoluta

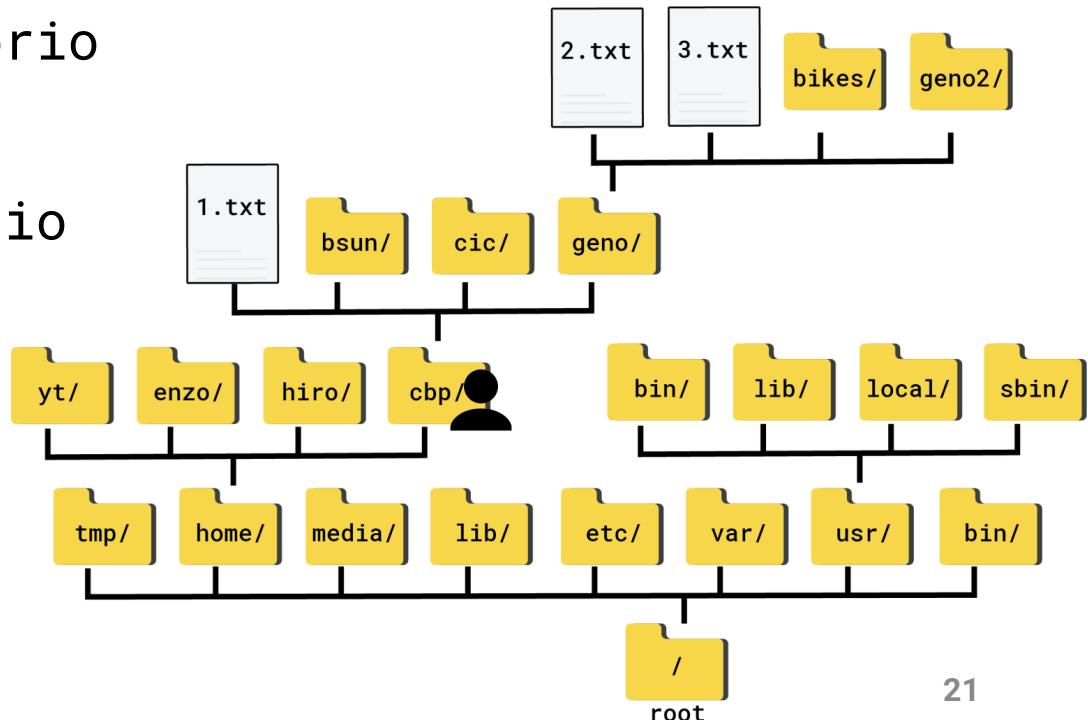
```
cbp@tarro :~$ cd /home/cbp/
```

Ruta relativa

```
cbp@tarro :~$ cd ..
```

Rutas (*paths*) absolutas y relativas

- aquí
- .. atrás
- último directorio
- ~ home del usuario



Manipulación básica de archivos

```
cbp@tarro :~$ touch <archivo>
```

```
cbp@tarro :~$ cp <origen> <destino>
```

```
cbp@tarro :~$ mv <origen> <destino>
```

```
cbp@tarro :~$ mkdir <nombre carpeta>
```

```
cbp@tarro :~$ rm <archivo>
```

```
cbp@tarro :~$ rm -r <carpeta>
```

Descargar datos desde la WEB

```
cbp@tarro :~$ wget <url>
```

```
cbp@tarro :~$ wget http://example.com/sample.php
```

Documentación de una herramienta

```
cbp@tarro :~$ man wget
```

Conexión a un servidor remoto

```
cbp@tarro:~$ ssh usuario@maquina -XC
```

```
cbp@tarro:~$ ssh camilo@172.16.138.98 -XC
```





Curso Internacional de Genómica y Metagenómica
14 – 16 de diciembre 2021

Introducción al trabajo en línea de comandos en UNIX

Dr.(c) Camilo B.P.



