



UNIVERSIDAD DE  
**COSTA RICA**

# Introducción a la bioinformática

---

Bradd Mendoza Guido  
Genómica de procariontes B0634-SP8221

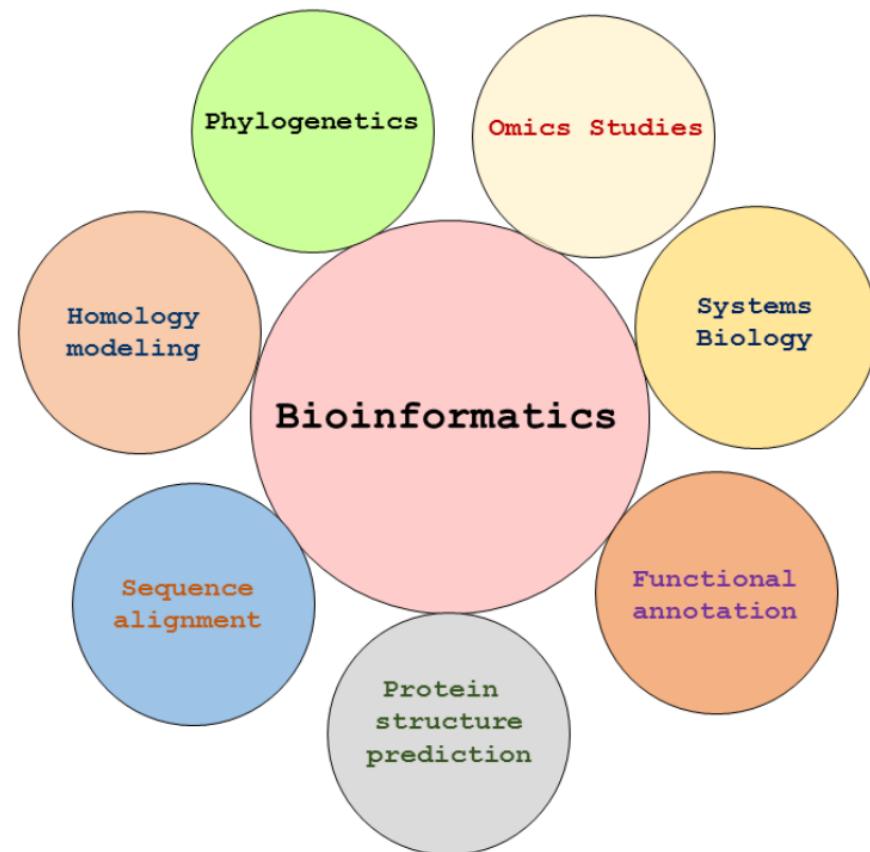
# Introducción a la bioinformática

## Agenda

- Breve historia de la bioinformática
- Tecnologías de secuenciación
- Bases de datos biológicas
- Estructura de un computador y sistemas operativos
- Tipos de archivos

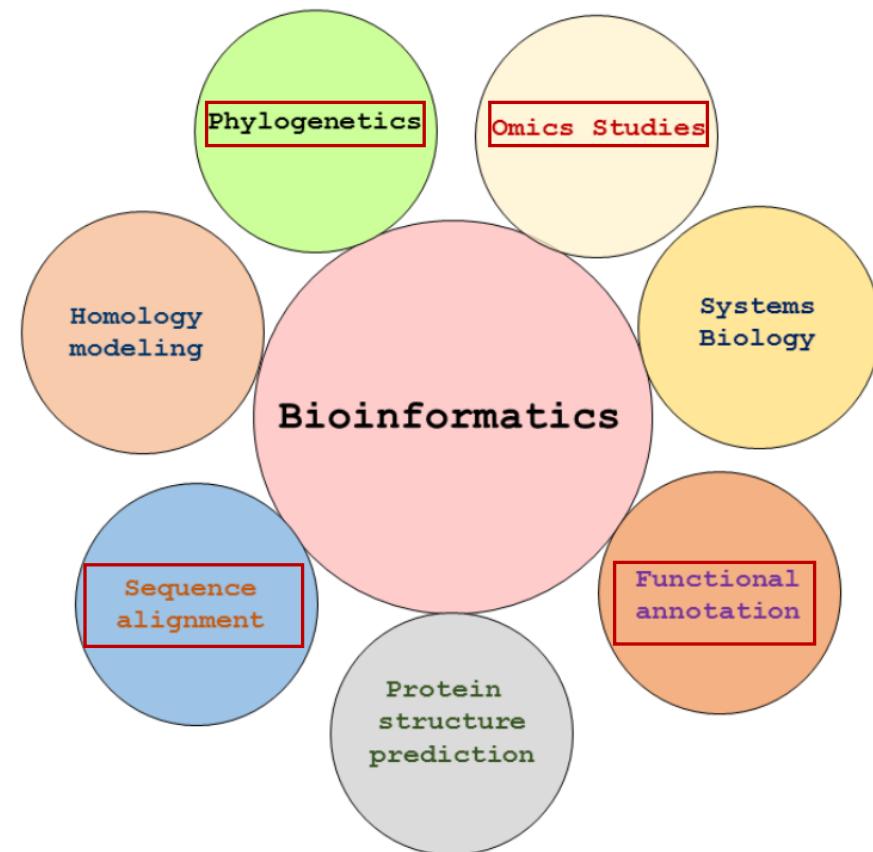


# Bioinformática

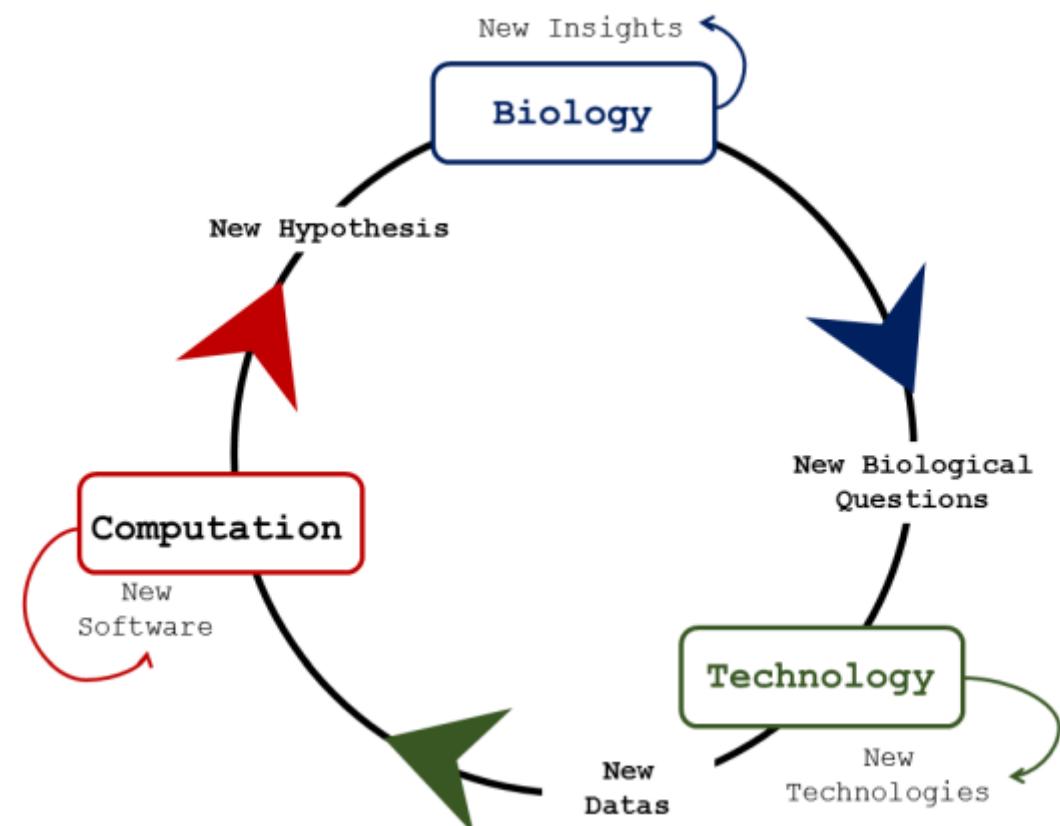


DOI <http://dx.doi.org/10.4238/gmr16019645>

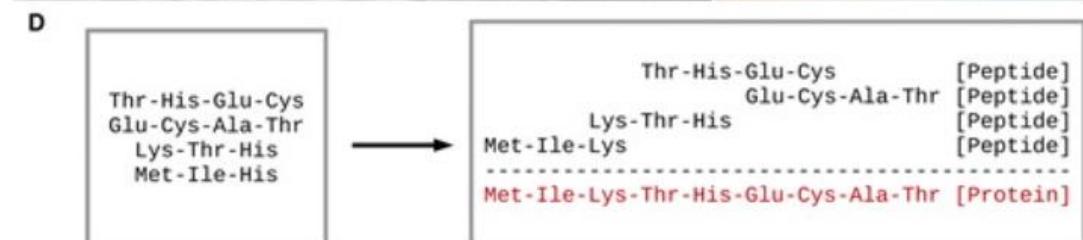
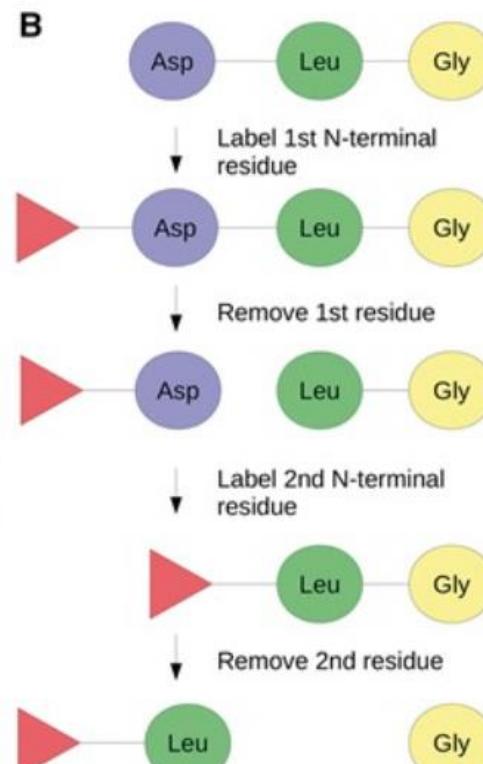
# Algunas aplicaciones del curso



DOI <http://dx.doi.org/10.4238/gmr16019645>



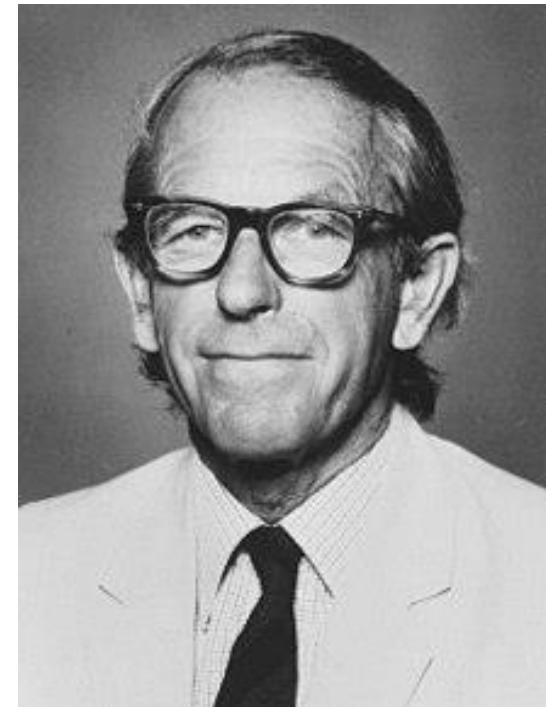
# Análisis de péptidos



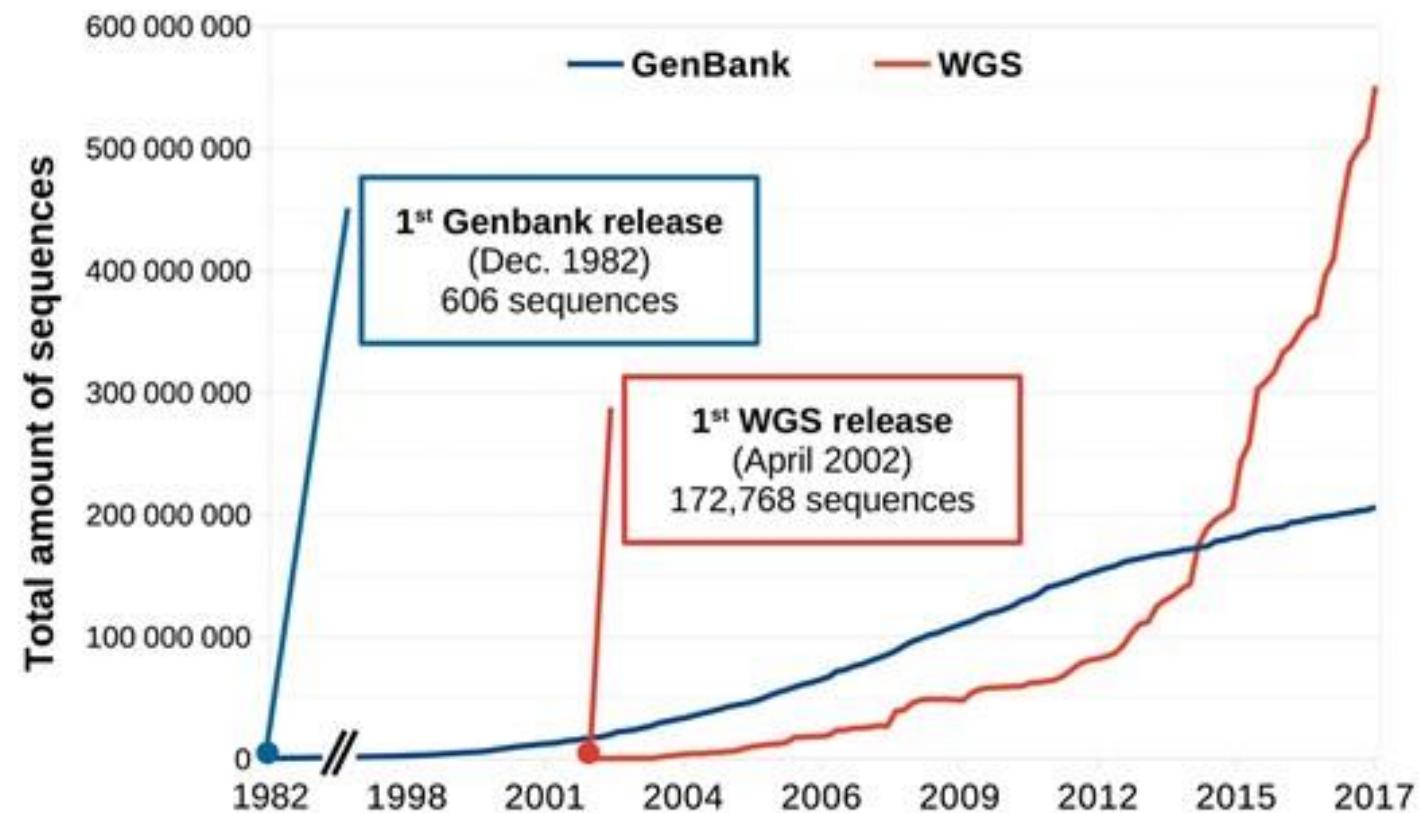
# De las proteínas al ADN

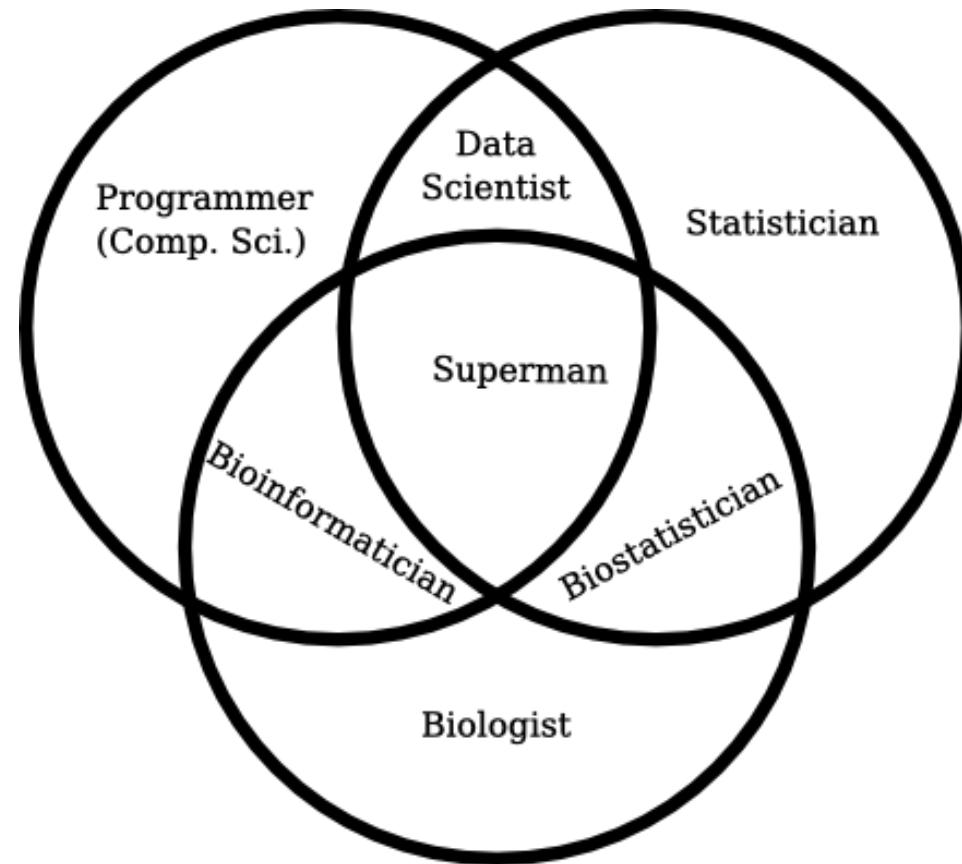


Hank Morgan//SPL v1.0

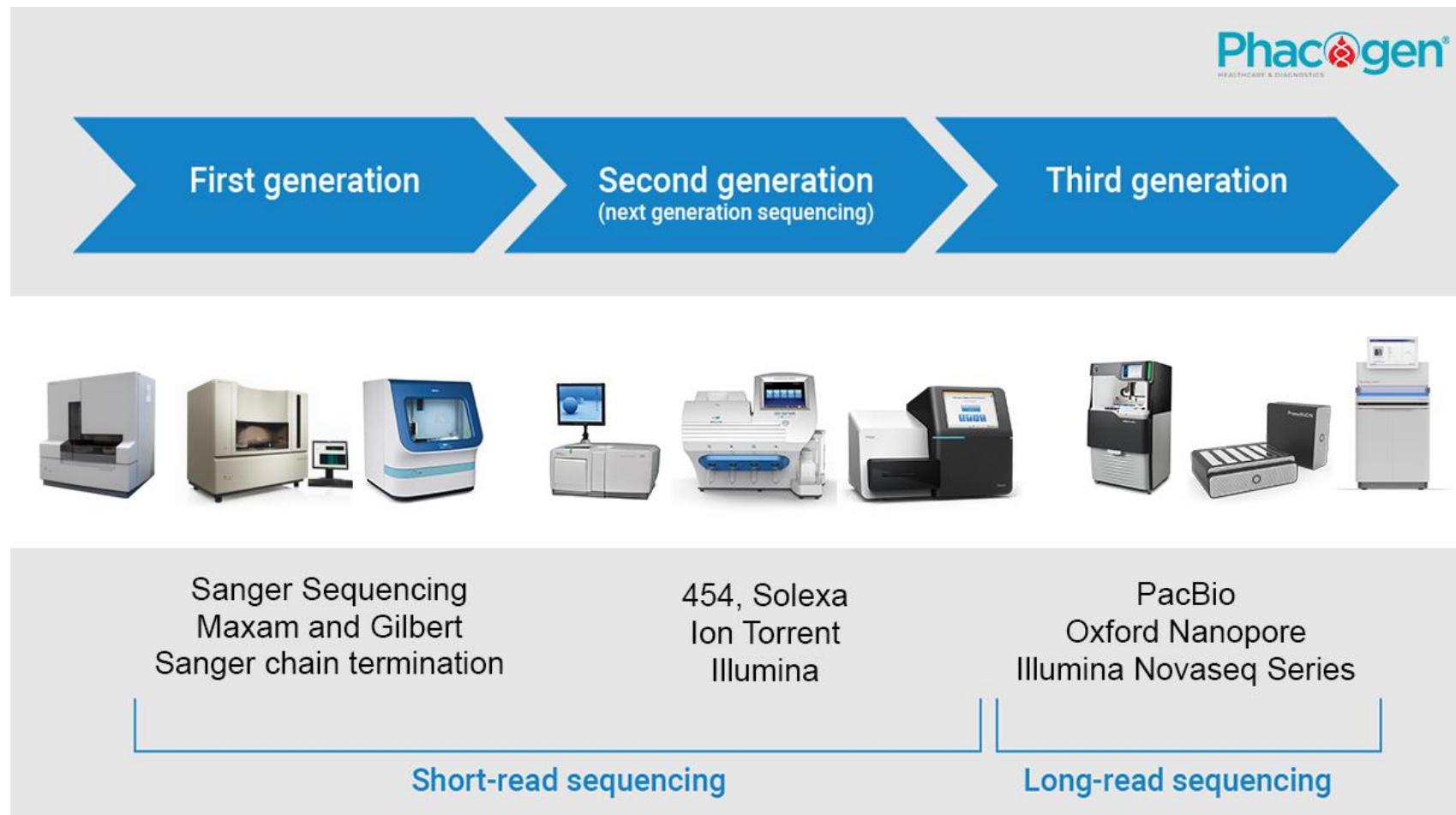


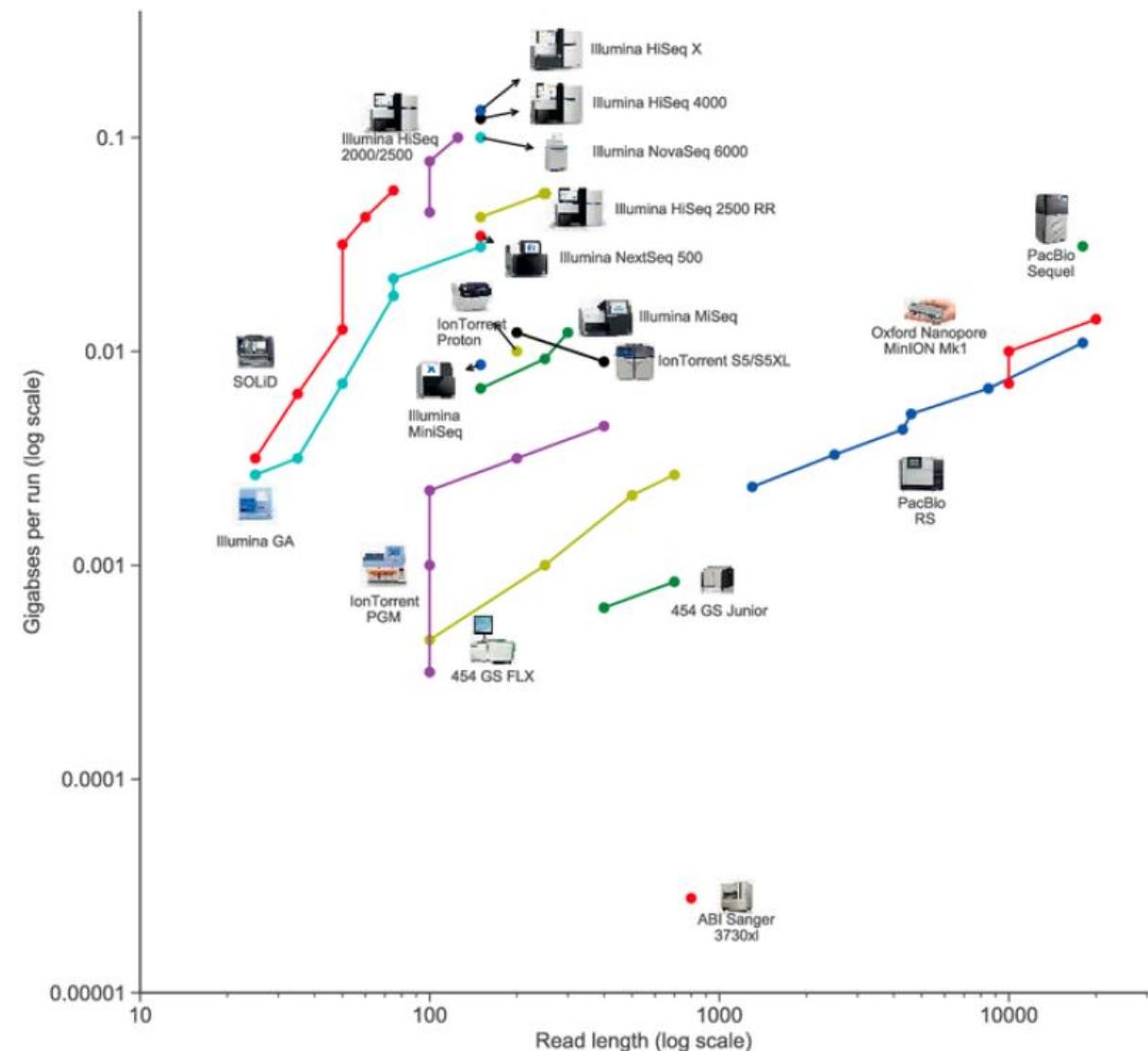
# Crecimiento exponencial





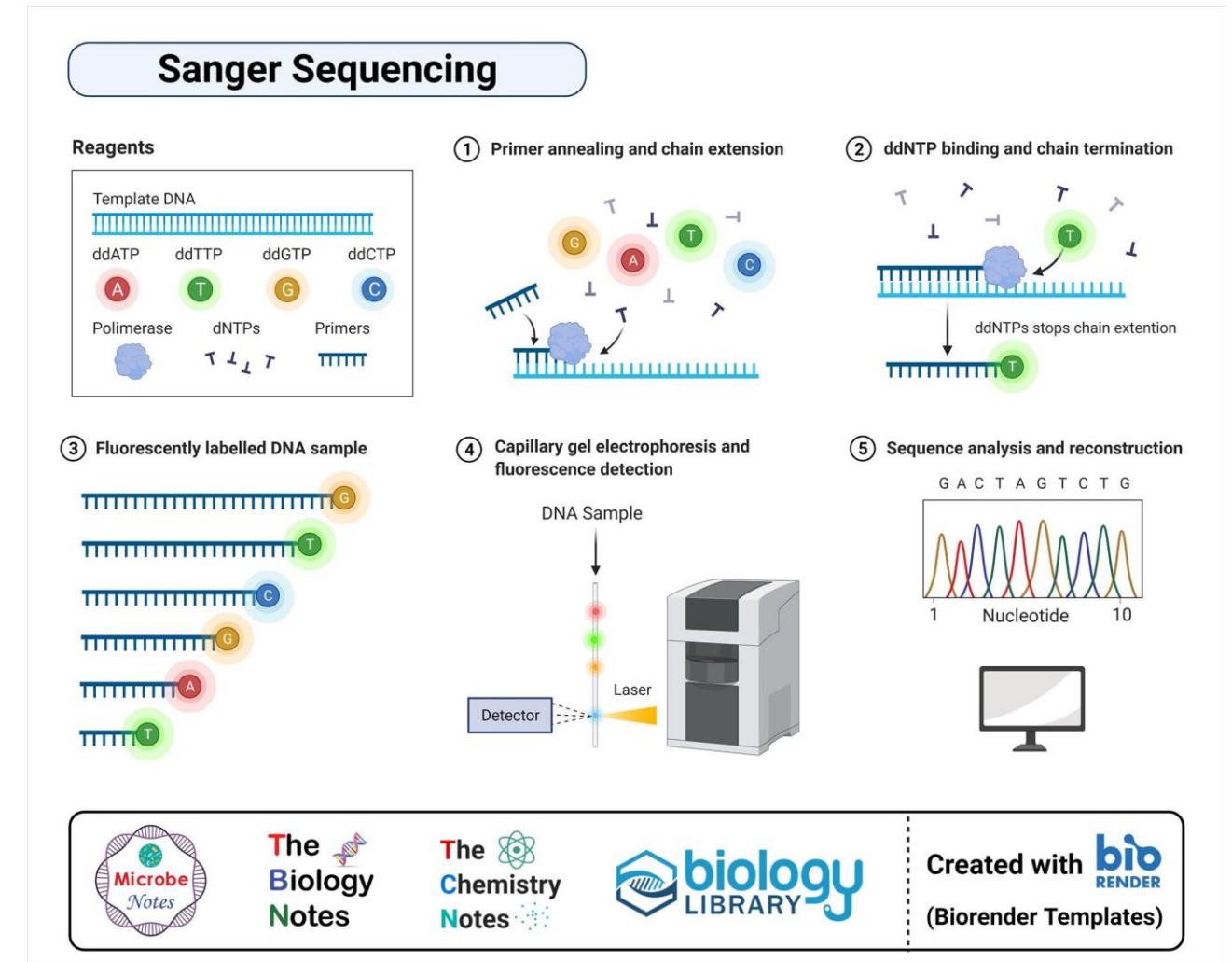
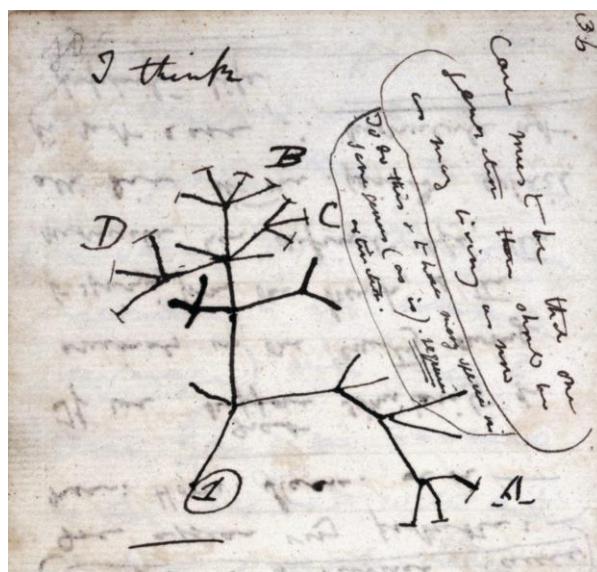
# Tecnologías de secuenciación





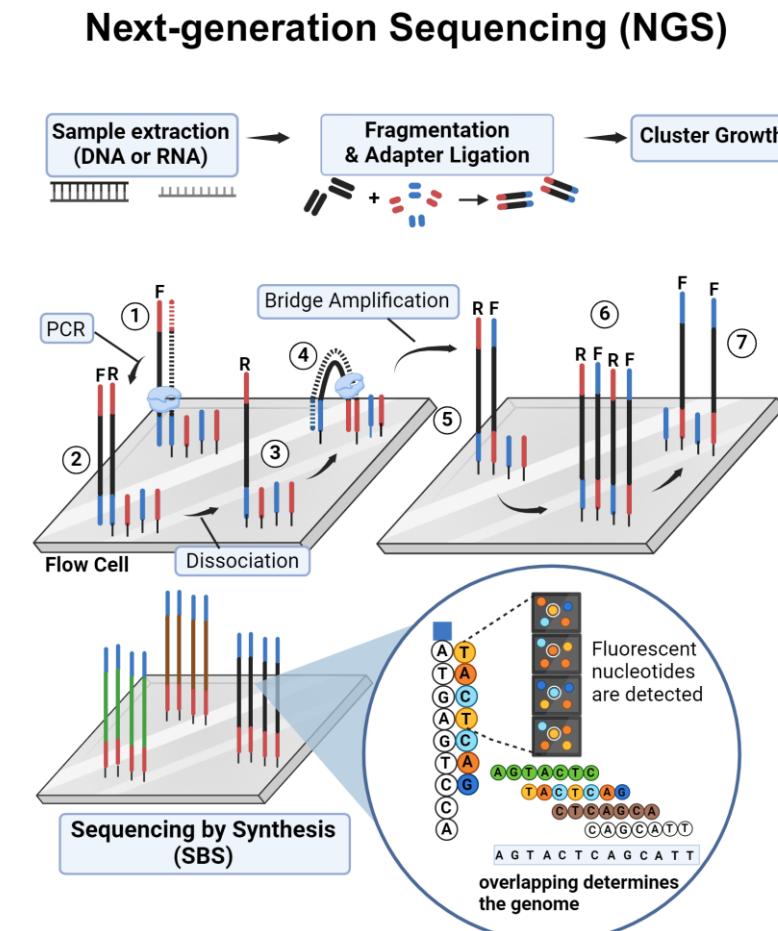
# Secuenciación Sanger

- Fragmentos menores a 1000pb
- Técnica que aún mantiene mayor fidelidad
- Se utiliza mucho en barcoding y análisis de variantes

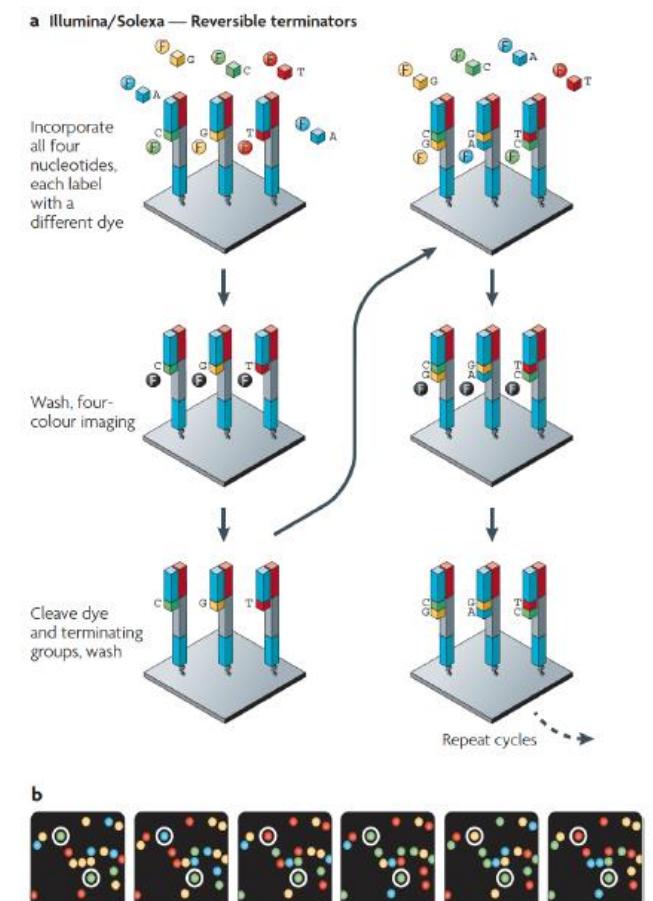


# Secuenciación Illumina

- Genera fragmentos pequeños (200-500pb)
- Puede generar millones de fragmentos por corrida
- Método con poca tasa de error
- El gold estándar de secuenciación a nivel mundial



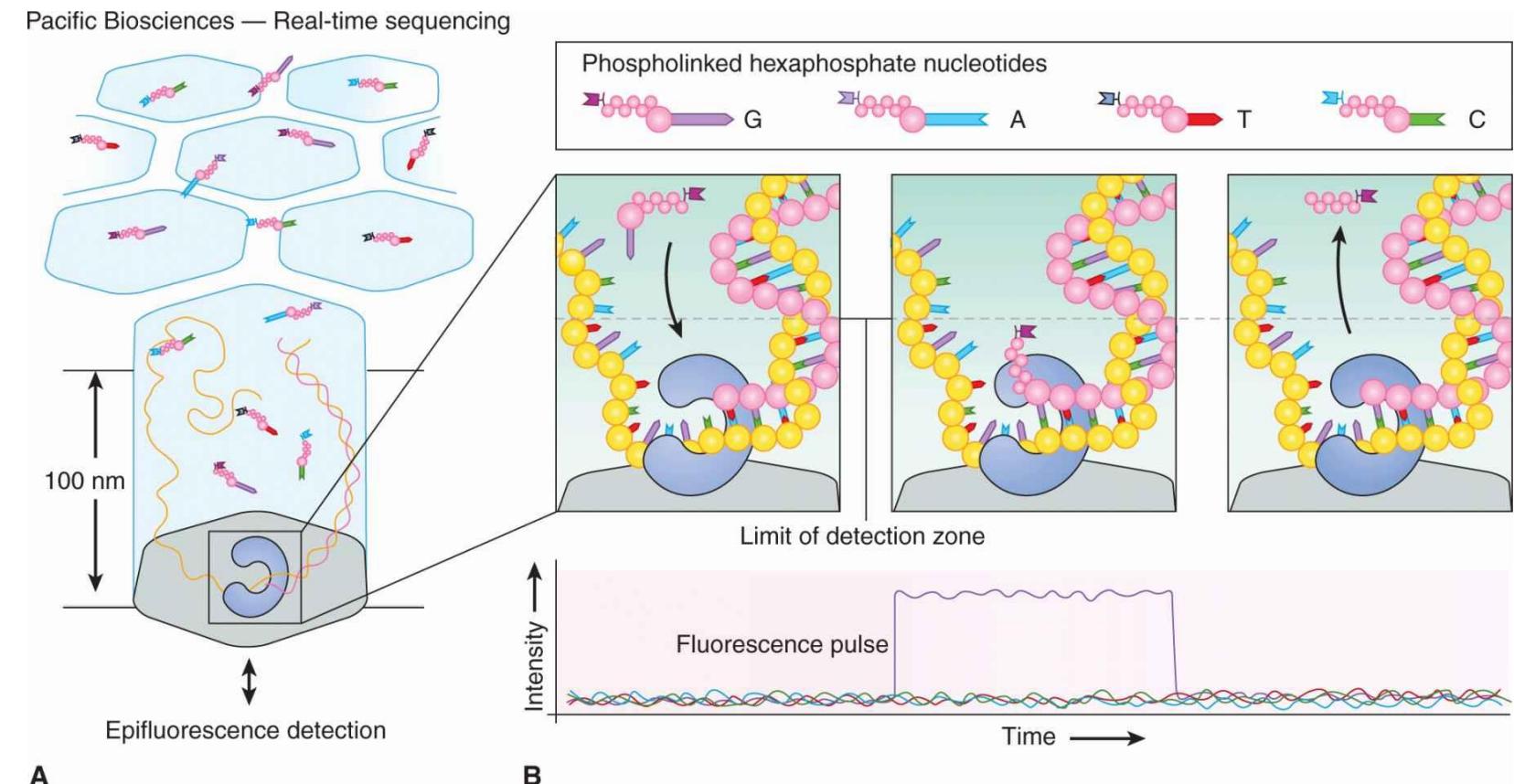
<https://www.biorender.com/template/next-generation-sequencing-workflow-ee448>



[https://doi.org/10.1007/978-3-030-76469-2\\_9](https://doi.org/10.1007/978-3-030-76469-2_9)

# Secuenciación PacBio

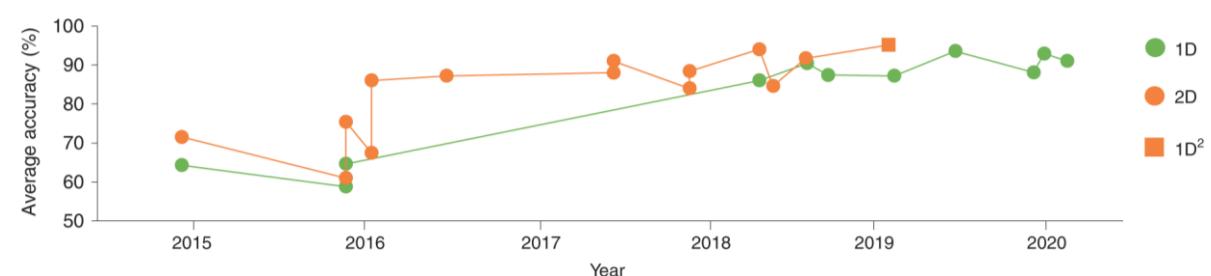
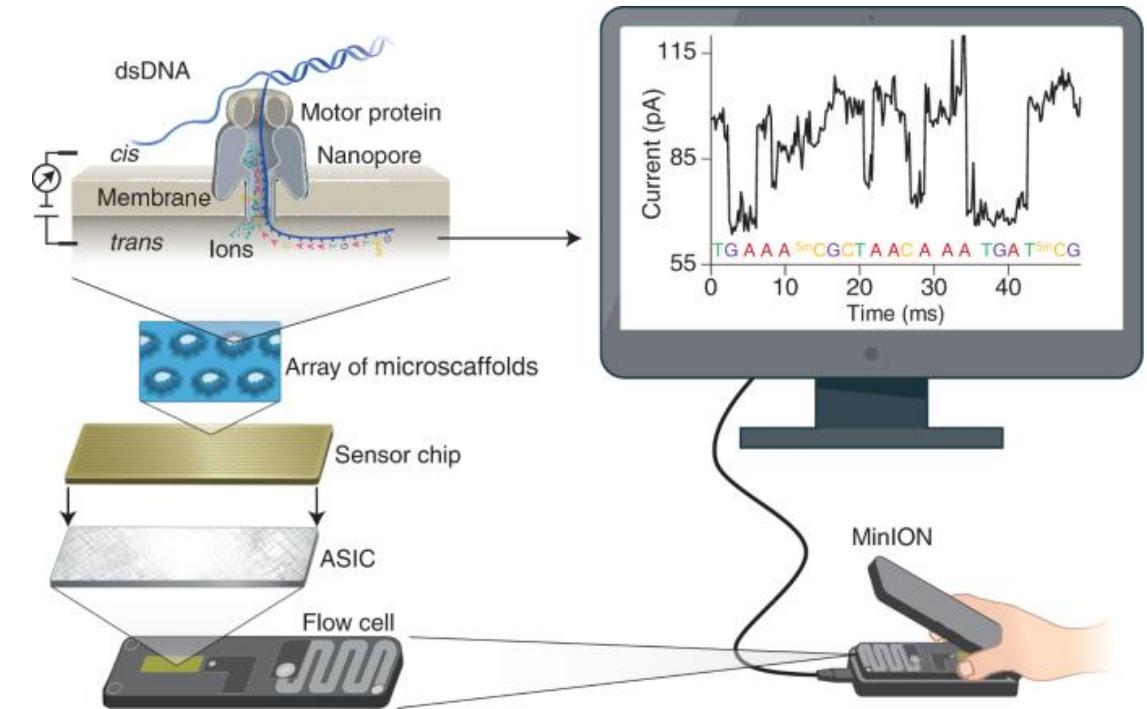
- Produce fragmentos largos (15-20kpb)
- Tiene una tasa de error considerable pero bastante buena
- Elevado precio, un solo genoma bacteriano puede rondar los 800 dólares



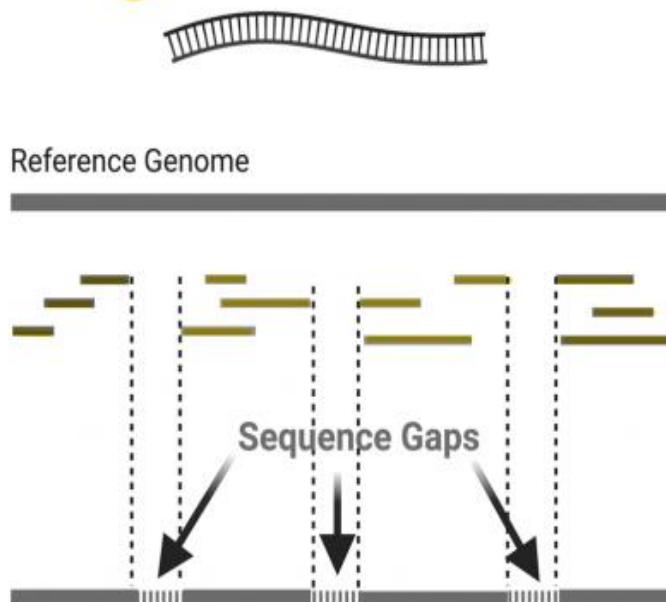
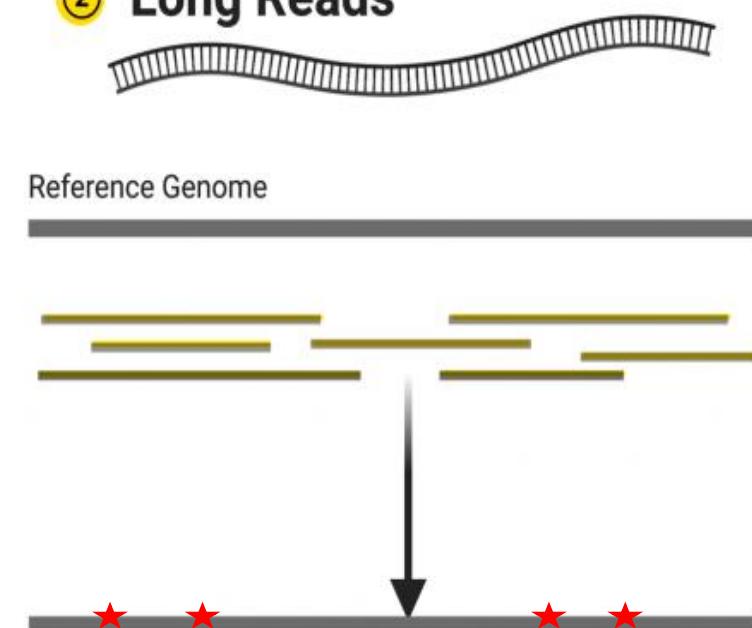
<https://oncohemakey.com/methods-of-molecular-analysis/>

# Secuenciación Nanopore

- Tasas de error muy altas aún
- Relativamente económico
- El secuenciador puede ser fácilmente movilizable



<https://doi.org/10.1038/s41587-021-01108-x>

**① Short Reads****② Long Reads**

<https://thegep.org/sequencing/>

# Estructura de archivos

## FASTA - genes and genomes

```
>NG_008679.1:5001-38170 Homo sapiens paired box 6 (PAX6)
ACCTCTTCTTATCATTGACATTAAACTCTGGGCAGGTCTCGCTAGAACCGGGCTGTCAGATCT
GCCACTTCCCTGCCAGCGCGGTGAGAAGTGTGGAACCGCGCTGCCAGGCTCACCTGCTCCCCGC
CCTCCGCTCCAGGTAAACGCCCGGCTCCGGCCCGCTCGGGCCCGCAGGGCTCTCCGCTG
CCAGCGACTGCTGTCCCCAAATCAAAGCCGCCCAAGTGGCCCGGGCTTGATTTGCTTTAAAAG
GAGGCATACAAAGATGGAAGCGAGTTACTGAGGGAGGGATAGGAAGGGGGTGGAGGAGGGACTTGTCTT
TGCGAGTGTCTCTGCAAAGTAGCAAATGTTCACTCTAAAGTGGACTTCCAGTCCGGCCCT
GAGCTGGGAGTAGGGGGCGGGAGTCTGCTGCTGTCTGCTAAAGCCACTCGCAGCGAAAAATGCA
GGAGGTGGGGACGCACTTGCATCCAGACCTCTGCAATCGCAGTTCACGACATCCACGCTGGAAAG
TCCGTACCCCGCCTGGAGCGCTAAAGACACCTGCCGGTCGGCGAGGTGCAGCAGAAGTTCCC
GCGGTTGCAAAGTGCAGATGGCTGGACCGCAACAAAGTCTAGAGATGGGTTCTCAGAAAGACGC
```

## FASTQ - sequencing reads

Identifier → @HWI-EAS209\_0006\_FC706VJ:5:58:5894:21141#ATCACG/1  
 Sequence → TTAATTGGTAAATAATCTCCTAATAGCTTAGATNTTACCTTNNNNNNNNNTAGTTCTTGAGA  
 + sign & identifier → +HWI-EAS209\_0006\_FC706VJ:5:58:5894:21141#ATCACG/1  
 Quality scores → efcfffffcfeffffcfffffd`feed)`\_)\_Ba\_`\_\_[YBBBBBBBBBRRTT\])[]dddd`

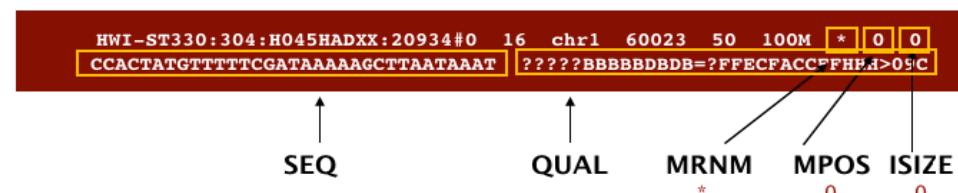
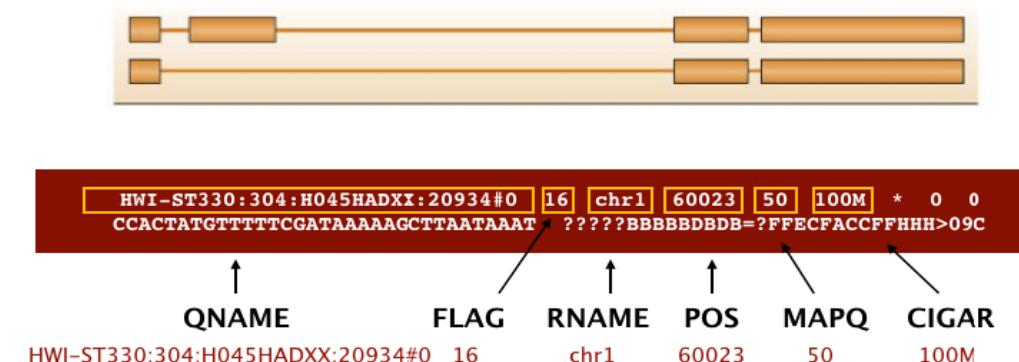
Base T

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

Tomado de: Gabriel Wallau

# Archivos de mapeo SAM/BAM

SAM/BAM



GFF

```

0 ##gff-version 3.2.1
1 ##sequence-region ctg123 1 1497228
2 ctg123 . gene      1000 9000 . + . ID=gene00001;Name=EDEN
3 ctg123 . TF_binding_site 1000 1012 . + . ID=tfbs00001;Parent=gene00001
4 ctg123 . mRNA      1050 9000 . + . ID=mRNA00001;Parent=gene00001;Name=EDEN.1
5 ctg123 . mRNA      1050 9000 . + . ID=mRNA00002;Parent=gene00001;Name=EDEN.2
6 ctg123 . mRNA      1300 9000 . + . ID=mRNA00003;Parent=gene00001;Name=EDEN.3
7 ctg123 . exon      1300 1500 . + . ID=exon00001;Parent=mRNA00003;Windows
    
```

# Variant Calling Format (VCF)

CP000819.1	1521	.	C	T	207	.	DP=9;VDB=0.993024;MQSB=-0.662043;MQ0F=0;AC=1;AN=1;DP4=0,0,4,5;MQ=60
CP000819.1	1612	.	A	G	225	.	DP=13;VDB=0.52194;MQSB=-0.676189;MQ0F=0;AC=1;AN=1;DP4=0,0,6,5;MQ=60
CP000819.1	9092	.	A	G	225	.	DP=14;VDB=0.717543;MQSB=-0.670168;MQ0F=0;AC=1;AN=1;DP4=0,0,7,3;MQ=60
CP000819.1	9972	.	T	G	214	.	DP=10;VDB=0.022095;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,2,8;MQ=60
CP000819.1	10563	.	G	A	225	.	DP=11;VDB=0.958658;MQSB=0.952347;MQ0F=0;AC=1;AN=1;DP4=0,0,5,5;MQ=60
CP000819.1	22257	.	C	T	127	.	DP=5;VDB=0.0765947;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,2,3;MQ=60
CP000819.1	38971	.	A	G	225	.	DP=14;VDB=0.872139;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,4,8;MQ=60
CP000819.1	42306	.	A	G	225	.	DP=15;VDB=0.969686;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,5,9;MQ=60
CP000819.1	45277	.	A	G	225	.	DP=15;VDB=0.470998;MQSB=0.680642;MQ0F=0;AC=1;AN=1;DP4=0,0,7,5;MQ=60
CP000819.1	56613	.	C	G	183	.	DP=12;VDB=0.879703;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,8,3;MQ=60
CP000819.1	62118	.	A	G	225	.	DP=19;VDB=0.414981;MQSB=0.691153;MQ0F=0;AC=1;AN=1;DP4=0,0,8,10;MQ=59
CP000819.1	64042	.	G	A	225	.	DP=18;VDB=0.451328;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,7,9;MQ=60

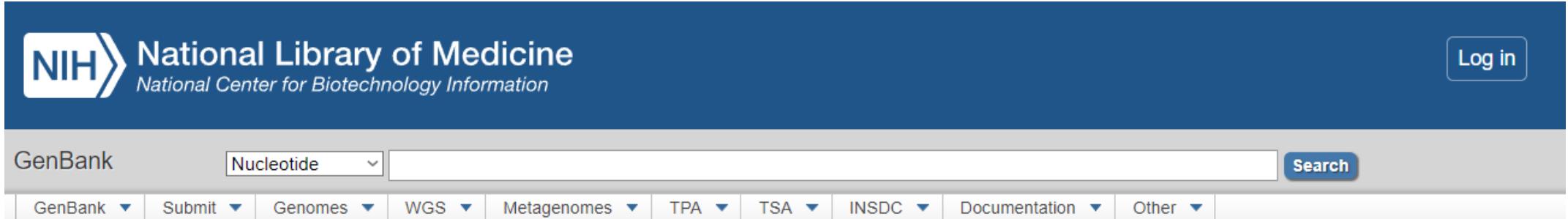
column	info
CHROM	contig location where the variation occurs
POS	position within the contig where the variation occurs
ID	a . until we add annotation information
REF	reference genotype (forward strand)
ALT	sample genotype (forward strand)
QUAL	Phred-scaled probability that the observed variant exists at this site (higher is better)
FILTER	a . if no quality filters have been applied, PASS if a filter is passed, or the name of the filters this variant failed

GTTACTGTGCGTTGAATACTCCACaATGTC  
 GTTACTGTGCGTTGAATAATCCACGATGTC  
 GTTACTGTGCGTTGAATATACTCCACGATGTC  
 GTTACTGTGCGTTGAATgCTCCACGATGTC  
 GTTACTGTGCGTTGAATATACTCCACAATGTC  
 GTTACTGTGCGTTGAATATACTCCACGATGTC  
 GTTACTGTGCGTGTAATATCTCCACaATGTC  
 GTTACTGTGCGTTGAATATACTCCACaATGTC  
 GTTAaTGTCGTTGAATATACTCCACGATGTC  
 GTTACTGTGCGTTGAcTACTCCACGATGTC  
 GTTACTGTGCGTTGAATATACTCCACaATGTC



sequencing errors                                    SNP

# Bases de datos biológicas



NIH > National Library of Medicine  
National Center for Biotechnology Information

Log in

GenBank Nucleotide Search

GenBank Submit Genomes WGS Metagenomes TPA TSA INSDC Documentation Other

## GenBank Overview

### What is GenBank?

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research](#), 2013 Jan;41(D1):D36-42). GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the [ftp site](#). The [release notes](#) for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for [previous GenBank releases](#) are also available. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An [annotated sample GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

### GenBank Resources

[GenBank Home](#)

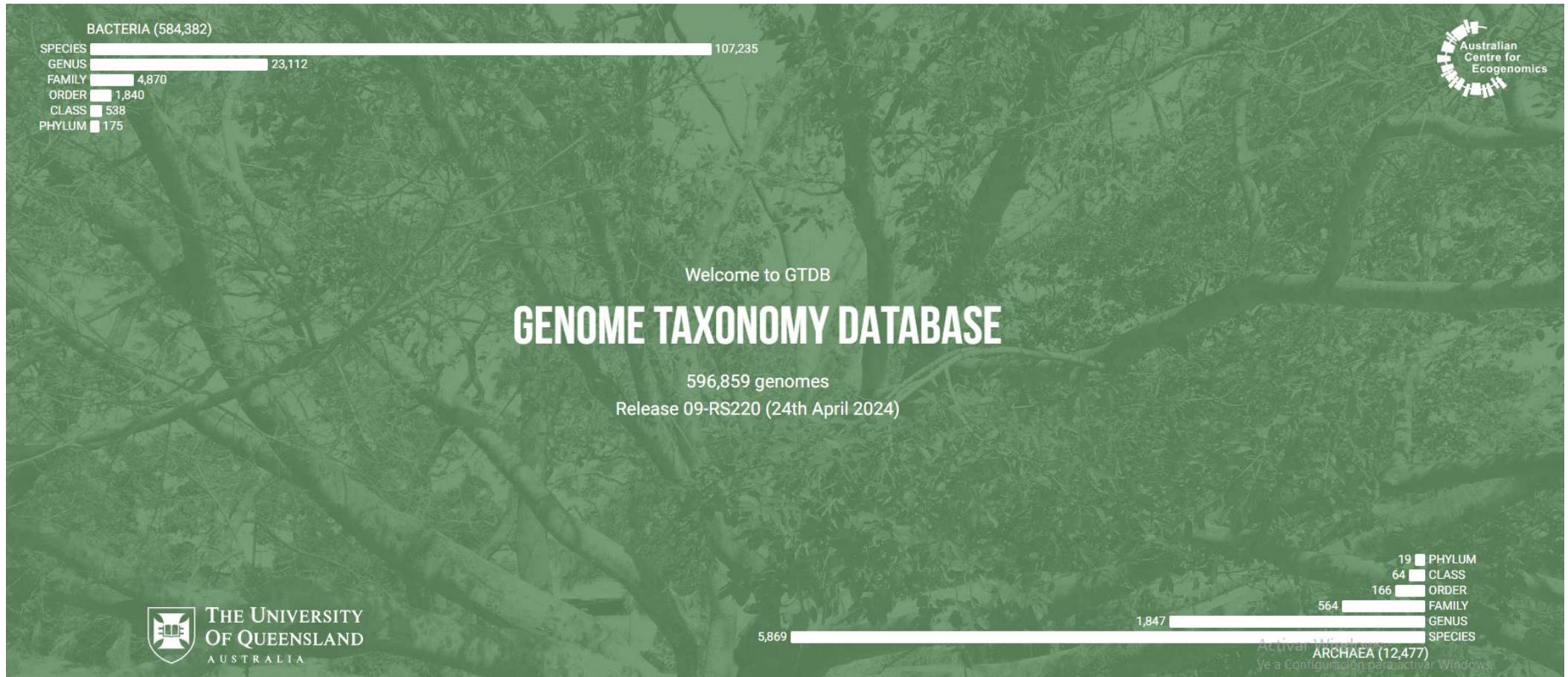
[Submission Types](#)

[Submission Tools](#)

[Search GenBank](#)

[Update GenBank Records](#)

# Bases de datos biológicas



**BACTERIA (584,382)**

Taxonomic Rank	Count
SPECIES	23,112
GENUS	4,870
FAMILY	1,840
ORDER	538
CLASS	175
PHYLUM	107,235

Welcome to GTDB

## GENOME TAXONOMY DATABASE

596,859 genomes  
Release 09-RS220 (24th April 2024)

**ARCHAEA (12,477)**

Taxonomic Rank	Count
PHYLUM	19
CLASS	64
ORDER	166
FAMILY	564
GENUS	1,847
SPECIES	5,869

Australian Centre for Ecogenomics

The University of Queensland Australia



KEGG Search Help  
» Japanese

**KEGG: Kyoto Encyclopedia of Genes and Genomes**

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. See [Release notes](#) (July 1, 2024) for new and updated features.

**Main entry point to the KEGG web service**

<a href="#">KEGG2</a>	<a href="#">KEGG Table of Contents</a>	<a href="#">[Update notes]</a>	<a href="#">[Release history]</a>
-----------------------	----------------------------------------	--------------------------------	-----------------------------------

**Data-oriented entry points**

<b>KEGG PATHWAY</b>	KEGG pathway maps	Pathway
<b>KEGG BRITE</b>	BRITE hierarchies and tables	Brite
<b>KEGG MODULE</b>	KEGG modules	Brite table
<b>KEGG ORTHOLOGY</b>	KO functional orthologs	Module
<b>KEGG GENES</b>	Genes and proteins [Annotation]	Network
<b>KEGG GENOME</b>	Genomes [KEGG Virus   Syntax]	KO (Function)
<b>KEGG COMPOUND</b>	Small molecules	Organism
<b>KEGG GLYCAN</b>	Glycans	Virus
<b>KEGG REACTION</b>	Biochemical reactions [RModule]	Compound
<b>KEGG ENZYME</b>	Enzyme nomenclature	Disease (ICD)
<b>KEGG NETWORK</b>	Disease-related network variations	Drug (ATC)
<b>KEGG DISEASE</b>	Human diseases	Drug (Target)
<b>KEGG DRUG</b>	Drugs [New drug approvals]	Antimicrobials

**KEGG MEDICUS** Health information resource [Drug labels search]

**Organism-specific entry points**

<b>KEGG Organisms</b>	Enter org code(s) <input type="text"/>	<input type="button" value="Go"/>	hsa	hsa eco
-----------------------	----------------------------------------	-----------------------------------	-----	---------

**Analysis tools**

<b>KEGG Mapper</b>	PATHWAY/BRITE/MODULE mapping tools
<b>KEGG Web Apps</b>	Pathway viewer with coloring features, etc.
<b>KEGG Syntax</b>	Genome alignment to analyze conserved synteny
<b>BlastKOALA</b>	BLAST-based KO annotation and KEGG mapping
<b>GhostKOALA</b>	HOSTX-based KO annotation and KEGG mapping
<b>BLAST/Fasta</b>	Sequence similarity search
<b>SIMCOMP</b>	Chemical structure similarity search

JGI **IMG/M** INTEGRATED MICROBIAL GENOMES & MICROBIOMES

[JGI HOME](#) [CONTACT US](#) [LOGIN / SIGN-ON](#)

# IMG/M Integrated Microbial Genomes & Microbiomes

## Bridging the Gap from Sequence to Biology

**Learn About IMG** **Explore** **Find & Analyze** **Save & Manage**

## JOURNAL ARTICLE

# NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes

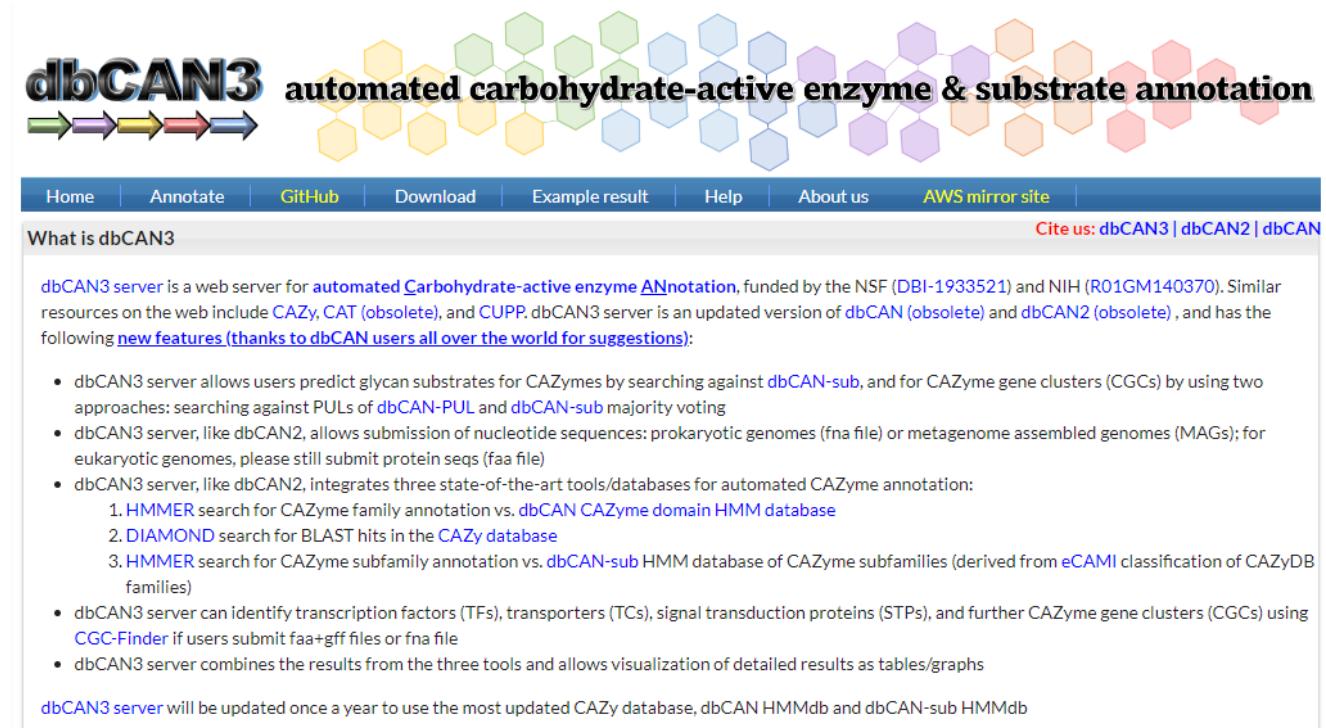
Qichao Tu ✉, Lu Lin ✉, Lei Cheng, Ye Deng, Zhili He

*Bioinformatics*, Volume 35, Issue 6, March 2019, Pages 1040–1048,

<https://doi.org/10.1093/bioinformatics/bty741>

Published: 28 August 2018

Article history ▾



The image shows the dbCAN3 web server homepage. At the top, there is a decorative header featuring a series of colored hexagons (yellow, green, blue, purple) arranged in a staggered pattern. To the left of the hexagons, the dbCAN3 logo is displayed with three arrows pointing right below it. To the right of the hexagons, the text "automated carbohydrate-active enzyme & substrate annotation" is written in a bold, black, sans-serif font. Below the header, a navigation bar contains links for "Home", "Annotate", "GitHub", "Download", "Example result", "Help", "About us", "AWS mirror site", and "Cite us: dbCAN3 | dbCAN2 | dbCAN". The main content area is titled "What is dbCAN3?" and describes the server as a web server for automated CAZyme annotation. It mentions funding from the NSF and NIH, and compares it to previous versions dbCAN and dbCAN2. A bulleted list details the features of dbCAN3, including its ability to predict glycan substrates, submit nucleotide sequences, integrate three state-of-the-art tools, identify transcription factors, transporters, signal transduction proteins, and further CAZyme gene clusters using CGC-Finder. The footer notes that dbCAN3 will be updated once a year.

dbCAN3 is a web server for automated Carbohydrate-active enzyme ANnotation, funded by the NSF (DBI-1933521) and NIH (R01GM140370). Similar resources on the web include CAZy, CAT (obsolete), and CUPP. dbCAN3 server is an updated version of dbCAN (obsolete) and dbCAN2 (obsolete), and has the following new features (thanks to dbCAN users all over the world for suggestions):

- dbCAN3 server allows users predict glycan substrates for CAZymes by searching against dbCAN-sub, and for CAZyme gene clusters (CGCs) by using two approaches: searching against PULs of dbCAN-PUL and dbCAN-sub majority voting
- dbCAN3 server, like dbCAN2, allows submission of nucleotide sequences: prokaryotic genomes (fna file) or metagenome assembled genomes (MAGs); for eukaryotic genomes, please still submit protein seqs (faa file)
- dbCAN3 server, like dbCAN2, integrates three state-of-the-art tools/databases for automated CAZyme annotation:
  1. HMMER search for CAZyme family annotation vs. dbCAN CAZyme domain HMM database
  2. DIAMOND search for BLAST hits in the CAZy database
  3. HMMER search for CAZyme subfamily annotation vs. dbCAN-sub HMM database of CAZyme subfamilies (derived from eCAMI classification of CAZyDB families)
- dbCAN3 server can identify transcription factors (TFs), transporters (TCs), signal transduction proteins (STPs), and further CAZyme gene clusters (CGCs) using CGC-Finder if users submit faa+gff files or fna file
- dbCAN3 server combines the results from the three tools and allows visualization of detailed results as tables/graphs

dbCAN3 server will be updated once a year to use the most updated CAZy database, dbCAN HMMdb and dbCAN-sub HMMdb



BLAST Align Peptide search ID mapping SPARQL

Release 2024\_04 | Statistics Help

# Find your protein

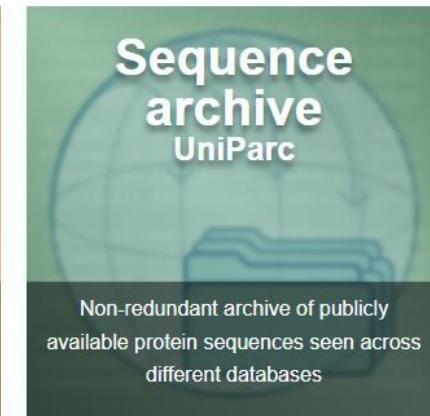
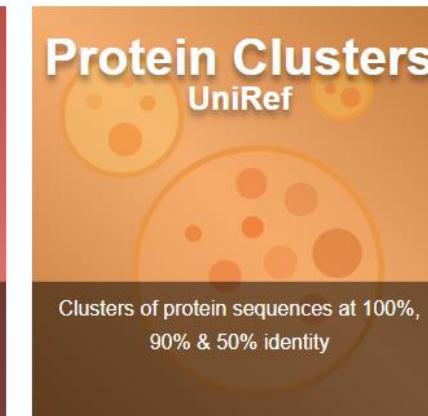
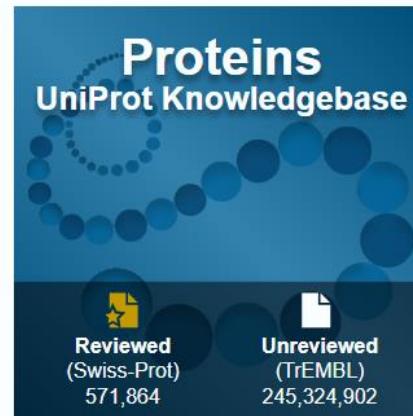
UniProtKB ▾

Examples: Insulin, APP, Human, P05067, organism\_id:9606

Advanced | List

Search

UniProt is the world's leading high-quality, comprehensive and freely accessible resource of protein sequence and functional information. [Cite UniProt™](#)



Feedback

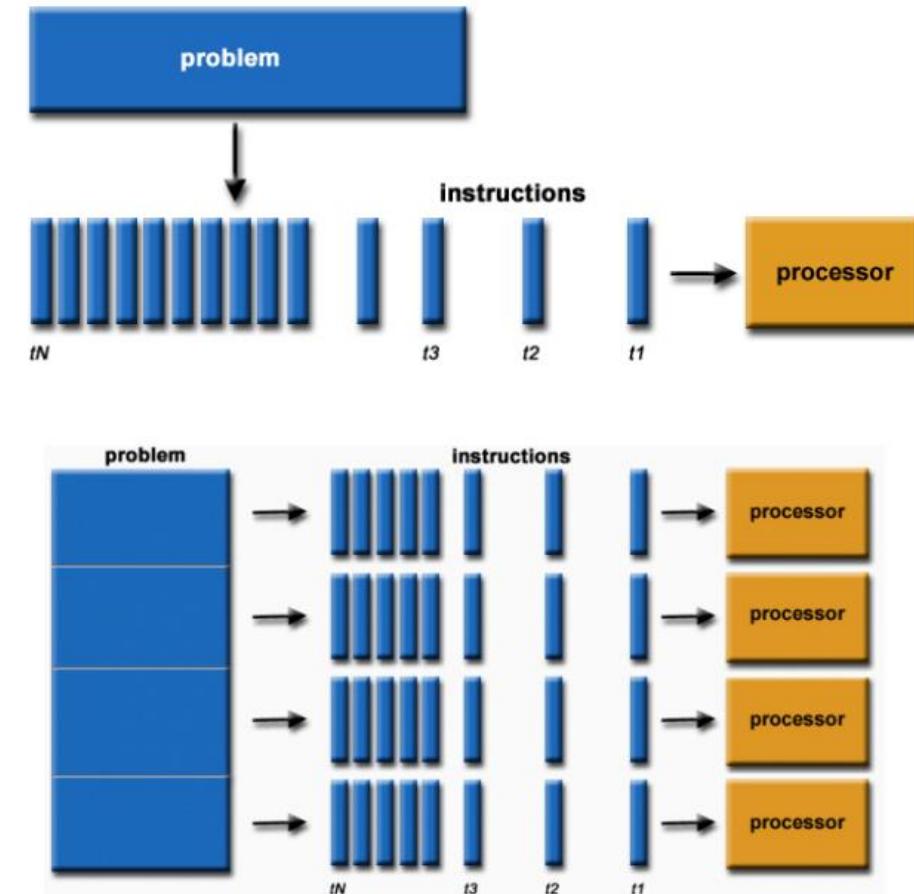
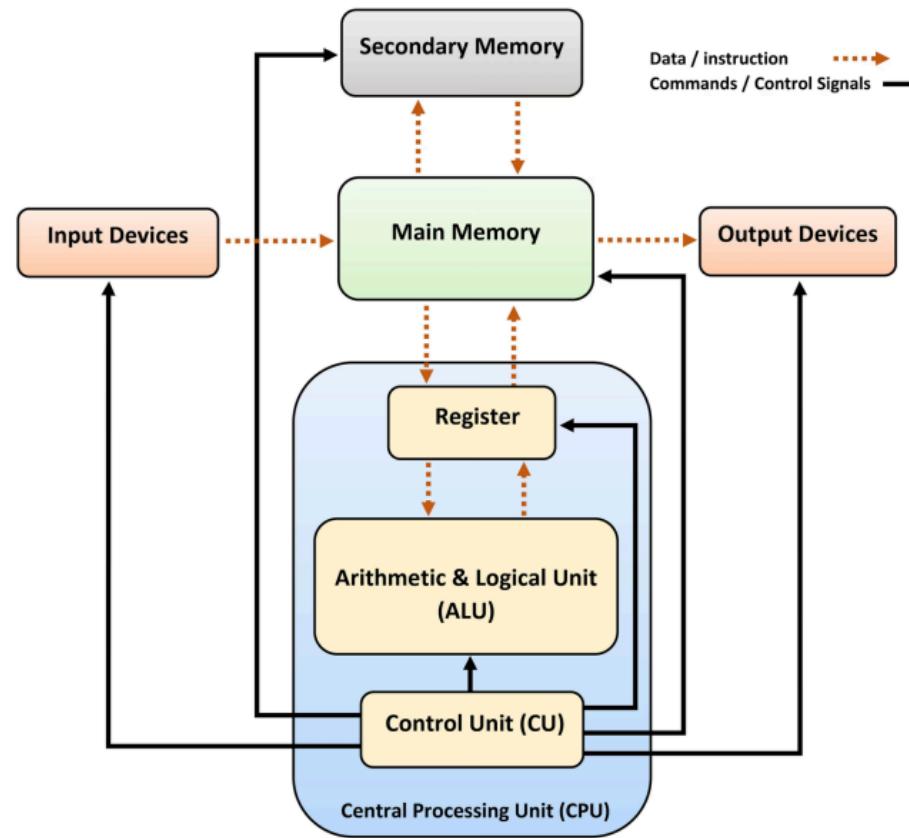


Help



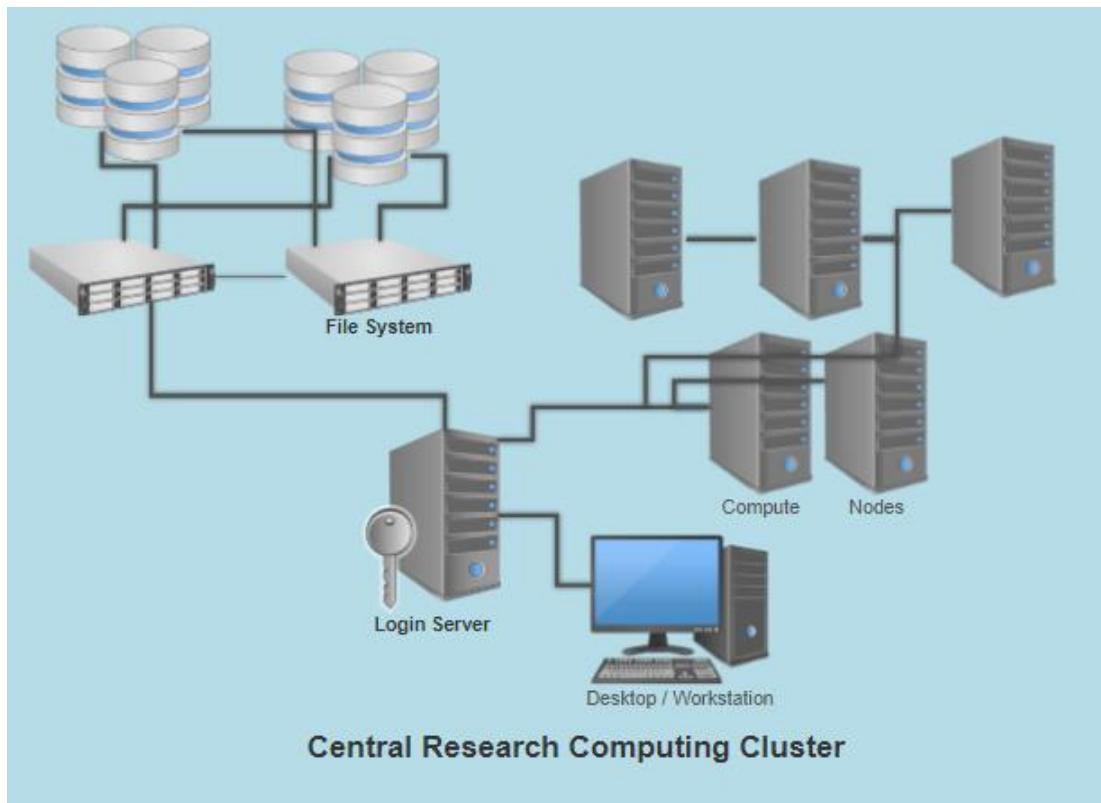
Activar Windows

# Estructura de un computador y SO

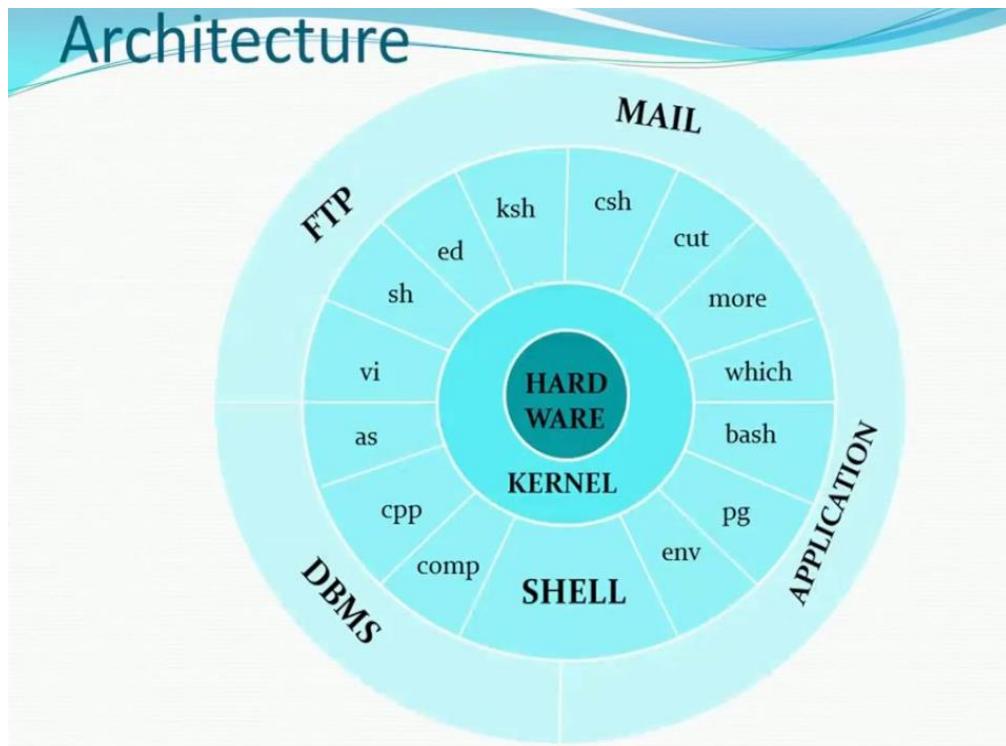


Tomado de: Gabriel Wallau

# Clúster computacionales



# Sistemas Operativos



```
bmendoza@login-0:/work/bm ~ braddmg@Braddmg: ~ + v
Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 5.15.153.1-microsoft-standard-WSL2 x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/pro

 * Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s just raised the bar for easy, resilient and secure K8s cluster deployment.

https://ubuntu.com/engage/secure-kubernetes-at-the-edge

This message is shown once a day. To disable it please create the /home/braddmg/.hushlogin file.
(base) braddmg@Braddmg:~$
```

# Por qué Linux?

- Open source
- Totalmente gratis
- Uno de los SO más estables



# Línea de comandos vs Interfaz gráfica

## CLI VS GUI

Comparison Chart

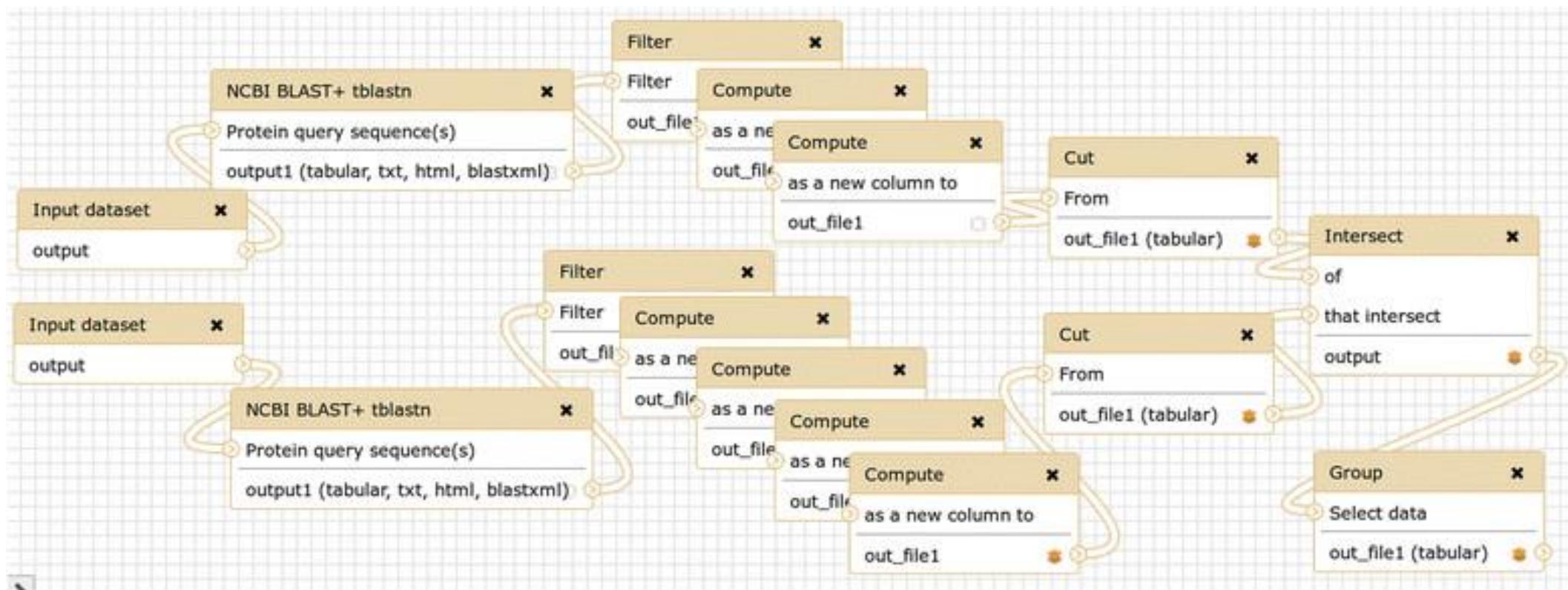
CLI	GUI		
Users interact with the system using various commands in the command window.	Users interact with the system with graphical elements such as icons, menus, images, etc.	CLI is not very user-friendly because the user needs to memorize a lot of commands.	GUI is relatively simpler to use and is more user-friendly.
CLI requires only a keyboard to enter commands.	GUI requires various input devices to interact with the system, such as a keyboard, mouse, etc.	CLI is relatively faster and excels in efficiency for professional users.	GUI is slower and tends to prioritize ease of use over speed.

CLI consumes less RAM and processing power.

GUI consumes more RAM and processing power.

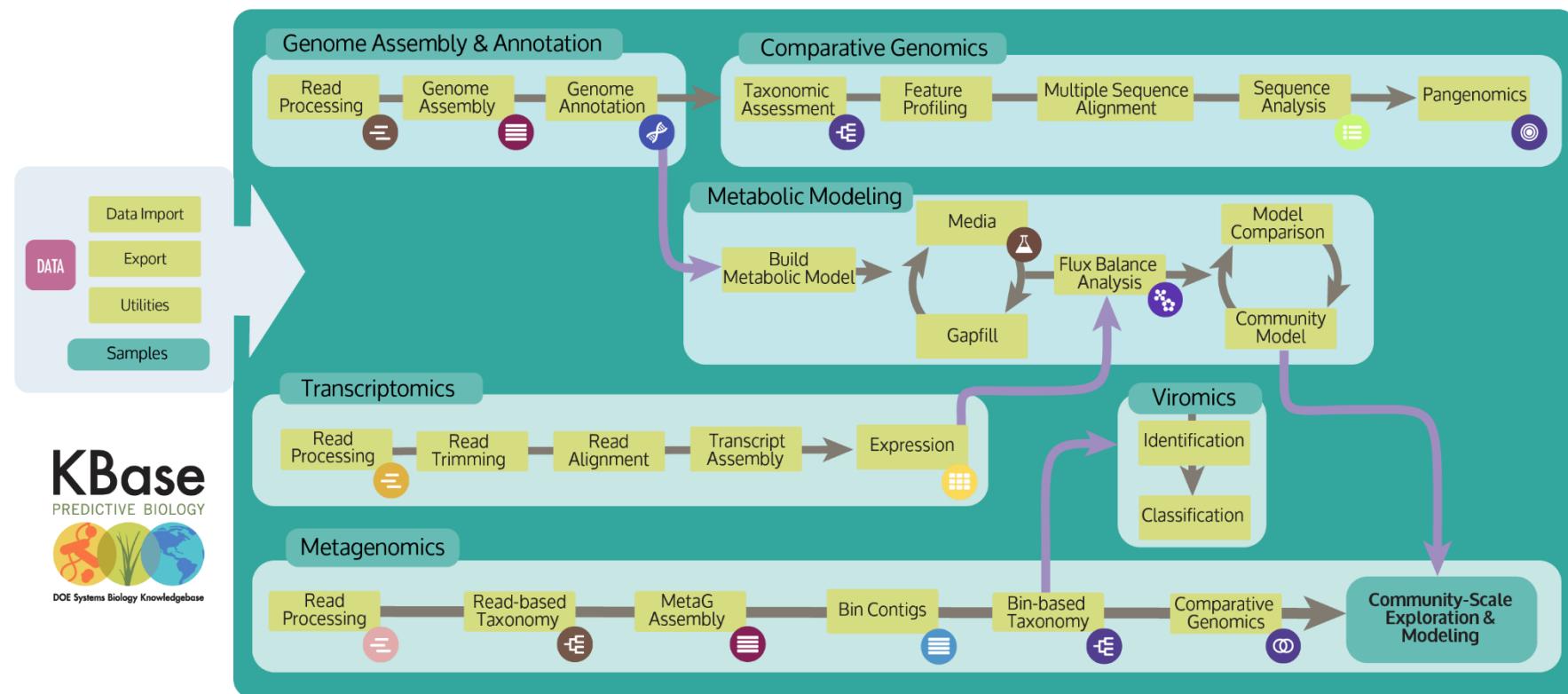
# Análisis bioinformáticos en línea

- **GALAXY**

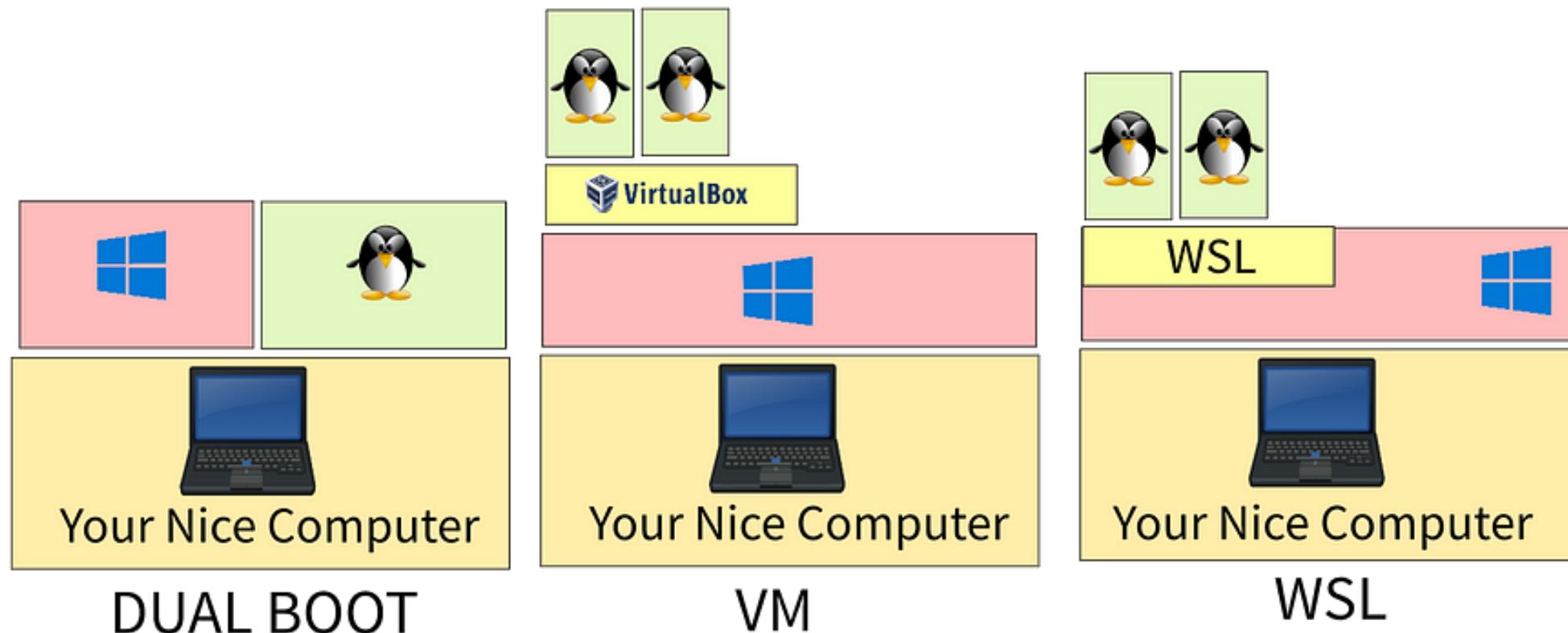


# Análisis bioinformáticos en línea

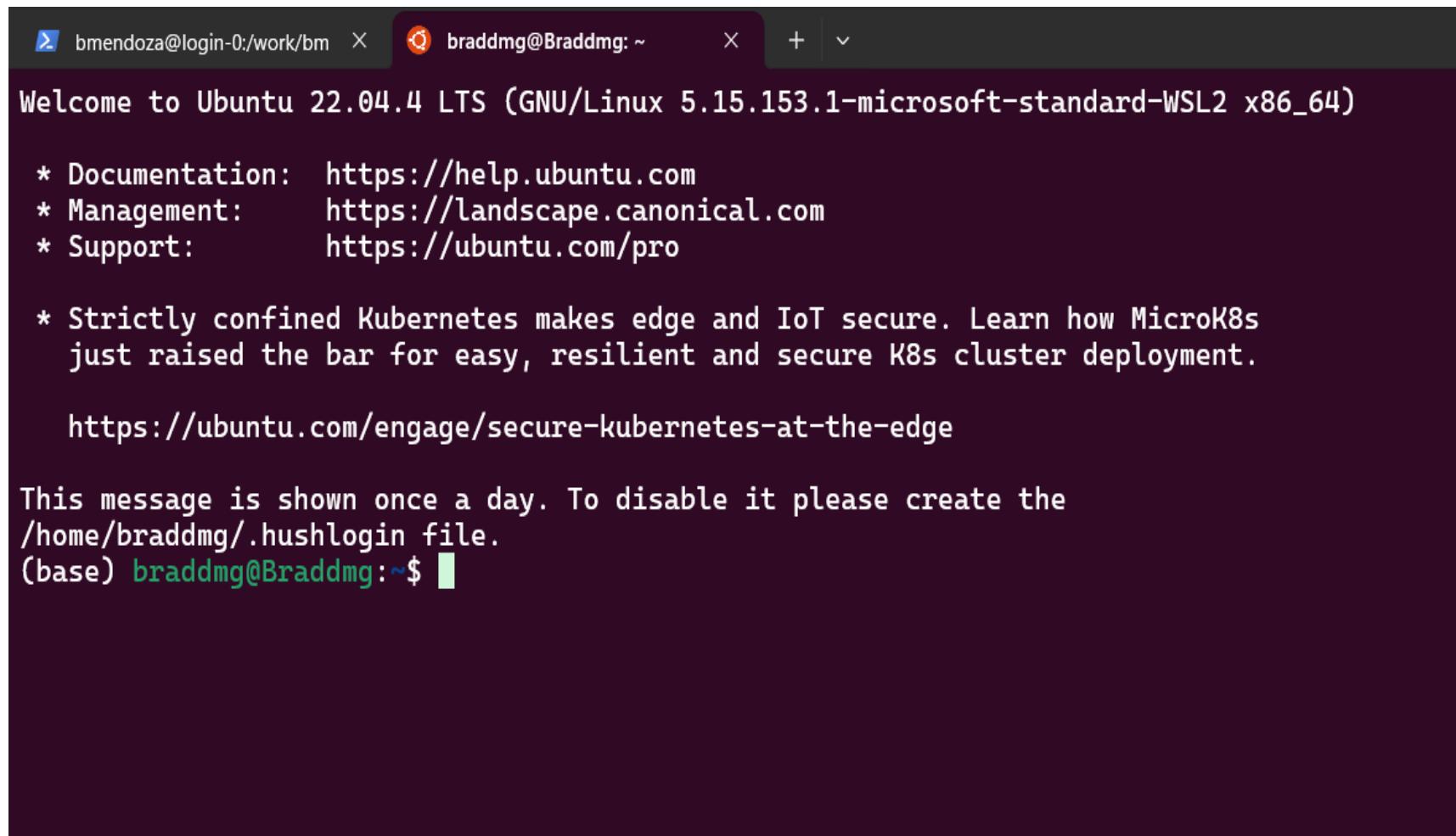
- KBASE



# Cómo instalar Linux en mi computadora?



# Windows Subsystem for Linux



The screenshot shows a Windows Terminal window with two tabs open. The active tab is titled "braddmg@Braddmg: ~" and displays a terminal session for Ubuntu 22.04.4 LTS. The session starts with a welcome message and documentation links:

```
Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 5.15.153.1-microsoft-standard-WSL2 x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/pro

 * Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s
 just raised the bar for easy, resilient and secure K8s cluster deployment.

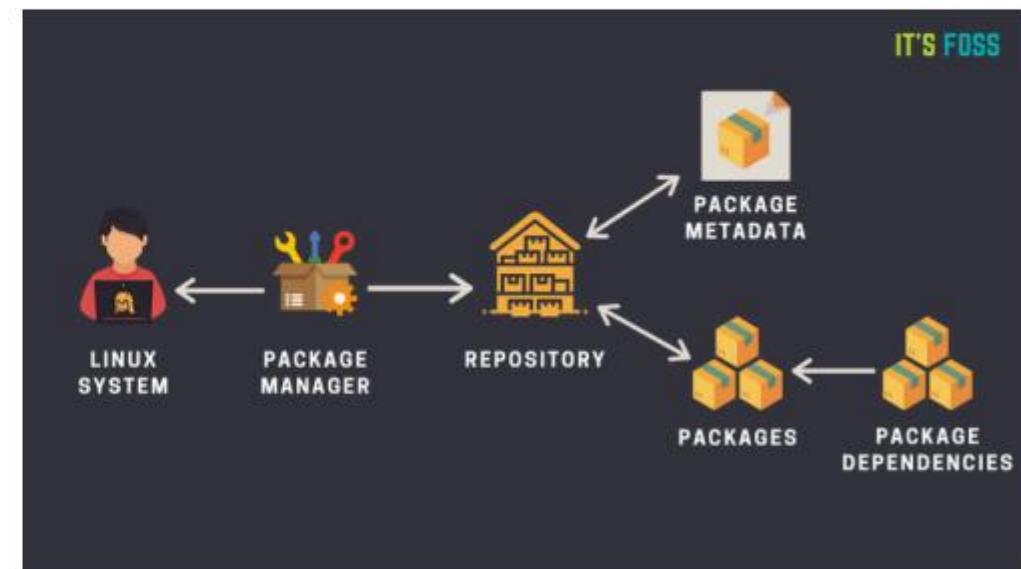
https://ubuntu.com/engage/secure-kubernetes-at-the-edge

This message is shown once a day. To disable it please create the
/home/braddmg/.hushlogin file.
(base) braddmg@Braddmg:~$
```

# Compilación de programas con Conda



CONDA®

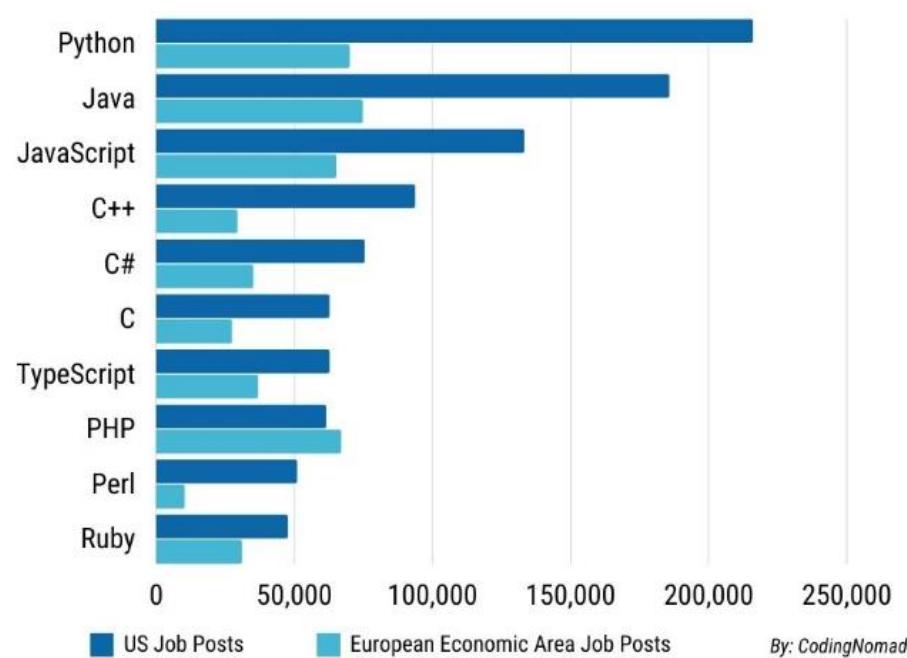


Tomado de: Gabriel Wallau

# Múltiples softwares y múltiples lenguajes

## Most in-demand programming languages of 2022

Based on LinkedIn job postings in the USA & Europe



By: CodingNomads

```
source("https://bioconductor.org/biocLite.R")  
biocLite("DESeq2")  
  
COPY
```

```
install.packages("matrixpls")
```

```
easy_install snakemake
```

```
./configure --prefix=/usr/local  
make  
make install
```

```
apt-get install bwa
```

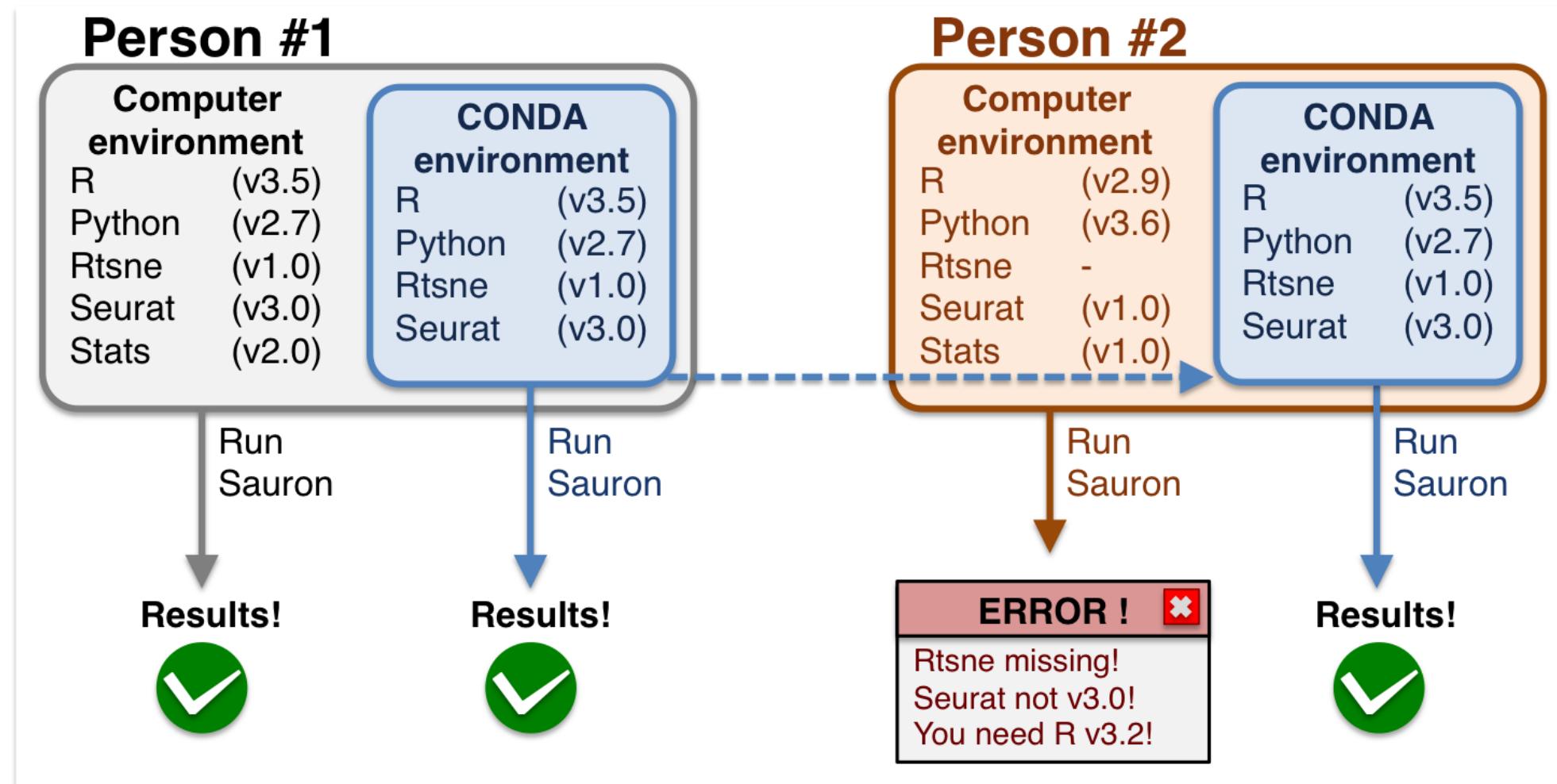
```
cp lib/amd64/jli/*.so lib  
cp lib/amd64/*.so lib  
cp * $PREFIX
```

```
cmake ../../my_project \  
-DCMAKE_MODULE_PATH=~/devel/seqan/util/cmake \  
-DSEQAN_INCLUDE_PATH=~/devel/seqan/include  
make  
make install
```

```
cpan -i bioperl
```

```
yum install python-h5py
```

# Conda



 ANACONDA.ORG

Search Anaconda.org

[About Anaconda](#) [Help](#) [Download Anaconda](#)[Sign In](#)

# bioconda / packages / spades 4.0.0



SPAdes (St. Petersburg genome assembler) is intended for both standard isolates and single-cell MDA bacteria assemblies.

[Conda](#)[Files](#)[Labels](#)[Badges](#)

-  License: GPLv2
-  Home: <http://cab.spbu.ru/software/spades/>
-  Development: <https://github.com/ablab/spades>
-  Documentation: <http://cab.spbu.ru/files/release4.0.0/manual.html>
-  417522 total downloads
-  Last upload: 1 month and 28 days ago



# Linux Cheat Sheet

Basic Commands	
<code>ls</code>	List all files in the present working directory
<code>ls -R</code>	Lists files in sub-directories
<code>ls -a</code>	Lists hidden files
<code>ls -al</code>	Lists files and directories with detailed information
<code>cd or cd ~</code>	Go back to home directory
<code>cd ..</code>	Move one directory level up
<code>cd</code>	Change to a particular directory
<code>cd /</code>	Go to the root directory
<code>touch &lt;filename&gt;</code>	Create a new file
<code>cat filename</code>	Display the file content
<code>mv filename new_file_name</code>	Renames the file to a new filename
<code>rm filename</code>	Deletes a file
<code>sudo</code>	Allow a user to run programs with the security privileges of the admin

File Permissions	
<code>ls -l</code>	Show file type and access permission
<code>r</code>	Read
<code>w</code>	Write
<code>x</code>	Execute
<code>-=</code>	No permission
<code>chown user</code>	Change the ownership of a file/directory
<code>chown user:group filename</code>	Change the user as well as group for a file or directory

File Compression	
<code>tar xvzf</code>	Create or extract .tar or .tgz files
<code>gzip, gunzip, zcat</code>	Create or extract .gz files
<code>uuencode, uudecode</code>	Create or extract .Z files
<code>zip, unzip -v</code>	Create or extract .ZIP files
<code>rpm</code>	Create or extract .rpm files
<code>rar</code>	Create or extract .rar files

Networking	
<code>ssh username@ip-address or hostname</code>	Login into a remote Linux machine
<code>ping hostname="" or =""</code>	Ping and analyze network and host connections
<code>dir</code>	Display files in the current directory of a remote computer
<code>cd "dirname"</code>	Change directory to "dirname" on a remote computer
<code>put file</code>	Upload 'file' from local to a remote computer
<code>get file</code>	Download 'file' from remote to a local computer
<code>quit</code>	Logout

Misc	
<code>pwd -P</code>	Print current working directory

# 30 Bash Commands Cheat Sheet

## Basic File and Directory Operations

ls Command	lists all files and directories in the current directory	<code>ls [options] [file dir]</code>
cd Command	changes the current directory to the specified one	<code>cd [directory]</code>
pwd Command	prints the path of the current working directory	<code>pwd [options]</code>

## File Manipulation

cat Command	print a file's contents in the terminal	<code>cat [options] [file path]</code>
touch Command	create a new, empty file	<code>touch [file]</code>
rm Command	removes files or directories	<code>rm [options] [file dir path]</code>
mkdir Command	creates a new, empty directory	<code>mkdir [options] [dir]</code>
rmdir Command	removes only empty directories	<code>rmdir [options] [dir]</code>
cp Command	copy files and directories using the CLI	<code>cp [options] [source] [destination]</code>
mv Command	move files and directories	<code>mv [options] [source] [destination]</code>

## Changing Permissions

chmod Command	modifies file and directory permissions	<code>chmod [options] [mode] [file]</code>
chown Command	changes the ownership of files and directories	<code>chown [options] [mode] [file]</code>

## Archiving and Compression

tar Command	creates, extracts, and manipulates archive files	<code>tar [options] [archive-file] [file   dir...]</code>
gzip Command	gzip compression utility	<code>gzip [options] [file...]</code>
gunzip Command	utility for decompressing gzip files	<code>gunzip [options] [file...]</code>

## Searching and Sorting

find Command	searches for files or a specific string of characters	<code>find [location] [expression] [options]</code>
grep Command	searches files for lines that match a regular expression	<code>grep [options] [search pattern] [file]</code>
sort Command	sorts the lines in text files in a specific order	<code>sort [options] [file]</code>

## Control Operations

kill Command	terminates processes by sending signals to them	<code>kill [options] pid...</code>
history Command	displays a list of recently executed commands	<code>history [options]</code>
exit Command	ends a terminal session or script	<code>exit [n]</code>
pipe Utility	connects the standard output of one command to the standard input of another	
Redirect Operator	controls the input and output of commands	>, >>, 2>, <, <<

## Viewing File and System Details

head Command	shows the beginning section of a text file or input stream	<code>head [options] [file...]</code>
tail Command	shows the ending portion of a text file or input stream	<code>tail [options] [file...]</code>
echo Command	displays text or variables on the terminal	<code>echo [options] [string]</code>
date Command	prints the system date and time	<code>date [options] [+format]</code>
df Command	displays information about disk space usage	<code>df [options] [filesystem]</code>
du Command	displays the disk space usage of files and directories	<code>du [options] [file]</code>
ps Command	provides information about currently running processes	<code>ps [options]</code>

Let's make bioinformatics!



<https://braddmg.github.io/GenPro/Preparacion-de-ambiente-linux/>