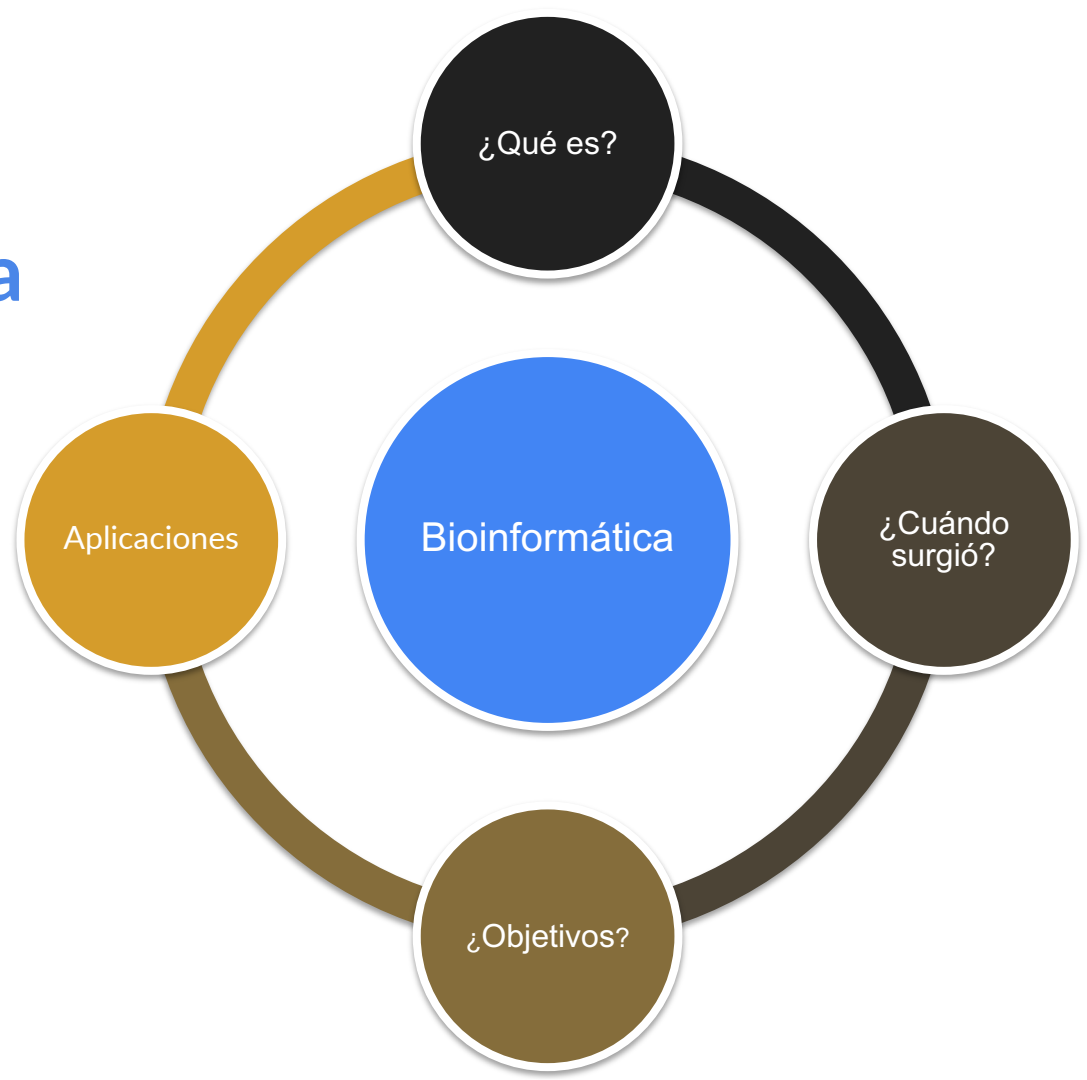


Dia 1. Introducción a la Bioinformática

Introducción a la Bioinformática



¿Cuándo surgió la Bioinformática?

1965

- Margaret Dayhoff desarrolló la primera base de datos de secuencias de proteínas llamada Atlas of Protein Sequence and Structure.

1970

- Laboratorio Nacional de Brookhaven estableció el Banco de Datos de Proteínas para archivar estructuras tridimensionales de proteínas

1970

- El primer algoritmo de alineación de secuencias fue desarrollado por Needleman y Wunsch en 1970

1980

- Se estableció GenBank y se desarrollaron algoritmos rápidos de búsqueda en bases de datos como FASTA por William Pearson y BLAST por Stephen Altschul y sus colegas

1980

- El inicio del proyecto del genoma humano

1990

- El desarrollo y el uso cada vez más extendido de Internet en la década de 1990 hicieron posible el acceso instantáneo, el intercambio y la difusión de datos biológicos.

¿Qué es la Bioinformática?

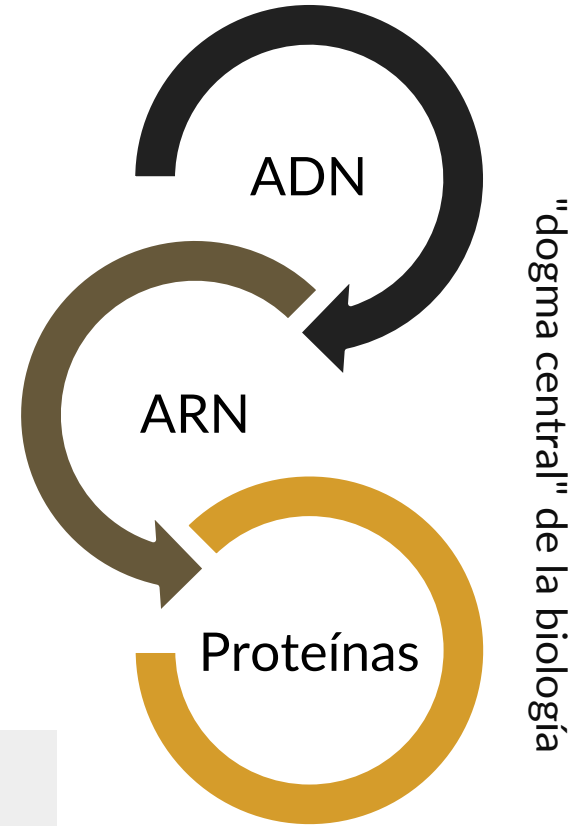


La bioinformática es la tecnología que utiliza **computadoras** para el almacenamiento, recuperación, manipulación y distribución de información relacionada con las **macromoléculas biológicas** como el **ADN, ARN y proteínas**.

¿Cuáles son los objetivos de la Bioinformática?


- El desarrollo de herramientas computacionales y bases de datos
- La aplicación de estas herramientas y bases de datos en la generación de conocimiento biológico

El objetivo final de la bioinformática es comprender mejor una célula viva y cómo funciona a nivel molecular.




¿Objetivos de la Bioinformática?

El desarrollo de herramientas y bases de datos computacionales



la aplicación de estas herramientas y bases de datos en la generación de conocimiento biológico



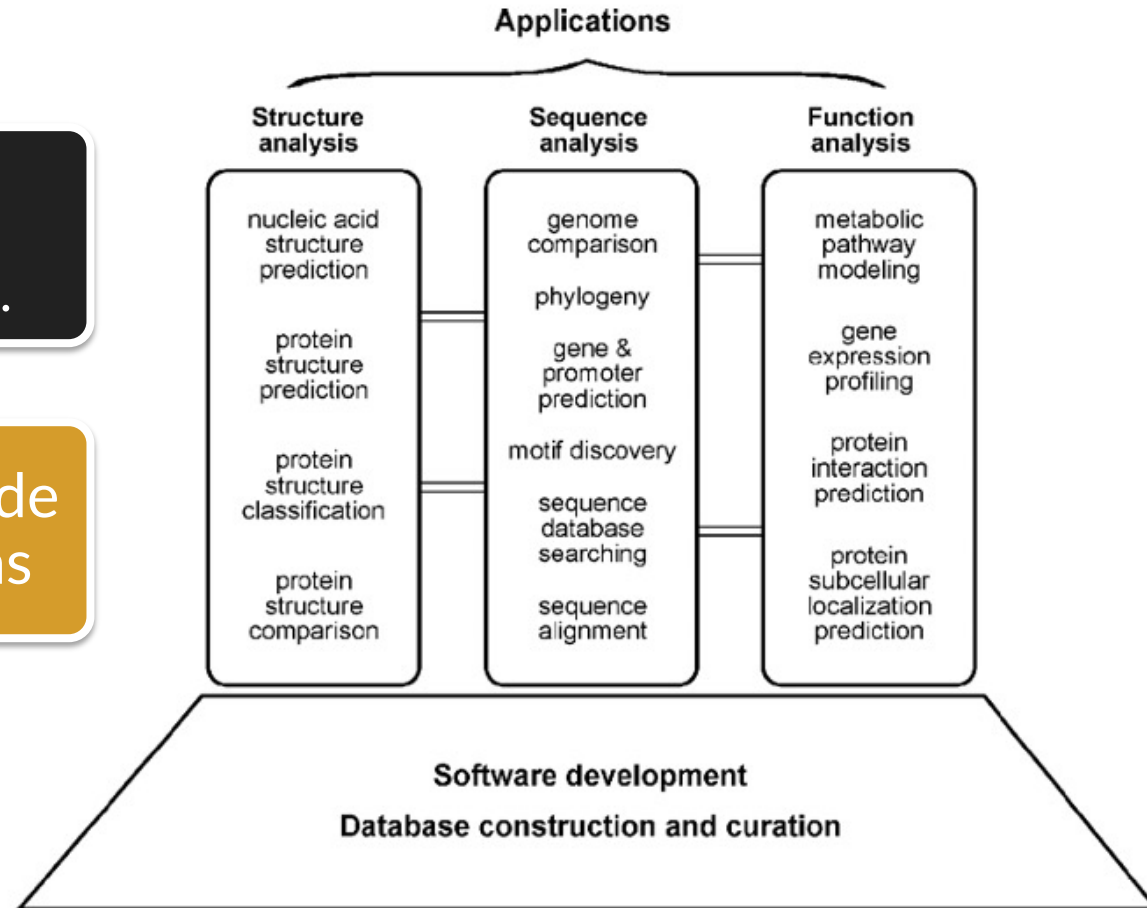
comprender mejor los sistemas vivos

¿Objetivos de la Bioinformática?

Escritura de software para análisis de secuencias, estructurales y funcionales.



Construcción y curación de bases de datos biológicas



Bases de datos primarias del ADN

Colaboración Internacional de Bases de Datos de Secuencia de Nucleótidos (INSDC)

- GenBank (Estados Unidos)

(<http://www.ncbi.nlm.nih.gov/Genbank/index.html>)

- EMBL (Europa)

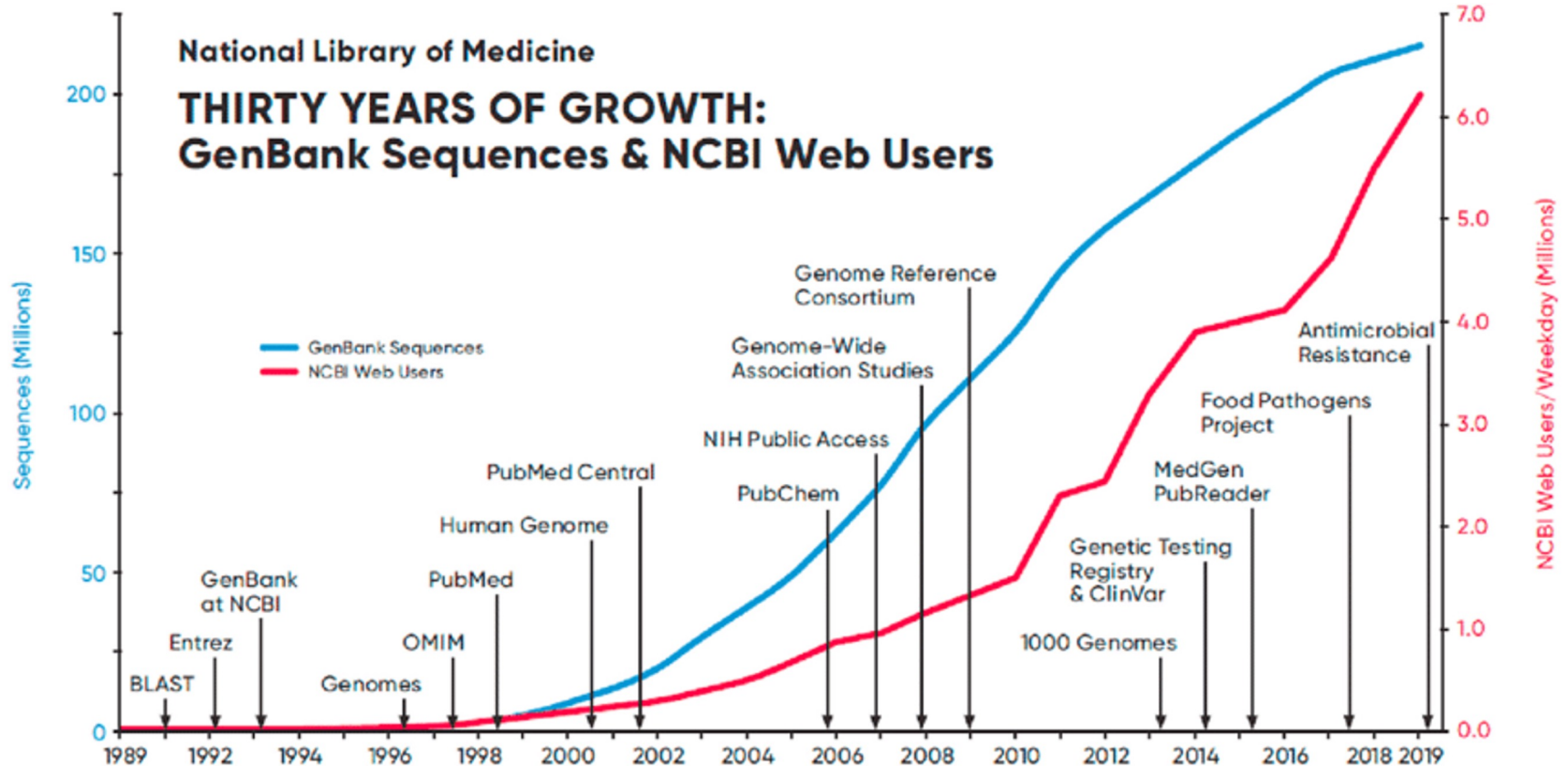
(<http://www.ebi.ac.uk/embl/>)

- DDBJ (Japón)

(<http://www.ddbj.nig.ac.jp/>)

Las secuencias de ADN son idénticas

Bases de datos primarias del ADN



Bases de datos específicas

Pseudomonas Genome DB

Current Strain (Clear/Search All) Gene/Product Name Return
Pseudomonas aeruginosa PAO1 (Stover et al.) Exact matches
Complete Genomes Only Current DB Version: 22.1 (2023-10-06)

Home News Acknowledgements About Search Annotations Search Sequences Annotations by Category Browsers

The Pseudomonas Genome Database collaborates with an international panel of expert *Pseudomonas* researchers to provide high quality updates to the PAO1 genome annotation and make cutting edge genome analysis data available

Search Gene Annotations by Name

Gene/Product name
Exact Name Name Contains
Choose a specific strain (optional)
Start typing and select from the list
OR choose a species (optional)
Start typing and select from the list
Complete genomes only
Include draft genomes
Search Annotations

Database Statistics

Complete Genomes	1324
Draft Genomes	12906
Manually-curated annotation updates	8755

Latest News

October 10, 2023
The original genome annotation for *P. fluorescens* SBW25 has been re-added after being erroneously deleted (NCBI had tagged it as an anomalous assembly).

October 6, 2023
Pseudomonas Genome Database version 22.1 released.
4883 curated updates to annotations.
New *Pseudomonas* genomes added and older versions of assemblies have been removed.
Database now has 7982 *Pseudomonas* aeruginosa genomes and 14229 *Pseudomonas* genomes overall.
Updated antimicrobial resistance gene predictions based on the *Resistance Gene Identifier* (RGI) and CARD version 3.2.5.
Updated functional domain predictions based on InterPro version 5.63-95.6.
Updated 3D structures: NCBI Protein Database structures (downloaded Aug 24, 2023) linked to Pseudomonas proteins using sequence similarity.
New link to search for AlphaFold 2-predicted 3D structures: Previously, *P. aeruginosa* PAO1 links were available but now it is possible to link other genomes using the UniProtKB accession.

[More news...](#)

Strains and Isolates in the Database

Searches apply to the currently selected organism database:
E. coli K-12 substr. MG1655

Click **Change Current Database** to search one of NaN other organism databases.



BioCyc
Genome Database Collection

Bacteria **Eukaryota** **MetaCyc**
19,435 databases 39 databases Metabolic Encyclopedia

Change Current Database Current Database: *Escherichia coli* K-12 substr. MG1655 reference genome (EcoCyc)

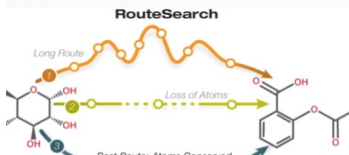
Search in Current Database: Enter a gene, protein, metabolite or pathway...

[SmartTables](#) | [Genome Browser](#) | [Metabolic Map](#) | [Omics Dashboard](#)

outeSearch: Search for Paths Through the etabolic Network

Search for lowest-cost paths through the metabolic network of the selected organism.

[am More](#)



20,050 Pathway/Genome Databases to Search

BioCyc is a collection of 20,050 Pathway/Genome Databases (PGDBs) for model eukaryotes and for thousands of microbes, plus software tools for exploring them. BioCyc is an encyclopedic reference that contains curated data from 146,000 publications.

Our EcoCyc and *Arthrosira platensis* NIES-39 databases are freely available, but access to the remaining BioCyc databases requires a paid subscription.

[BioCyc Intro](#) [Request Free Classroom Use](#)

The following websites are part of BioCyc. Opening these sites in your browser is a convenient way to pre-select that database for searches.

Tools Sites Pathway Tools Help Subscribe to BioCyc Welcome: Francine Amaral

EcoCyc

Change Current Database Current Database: *Escherichia coli* K-12 substr. MG1655 reference genome (EcoCyc)

Search in Current Database: Enter a gene, protein, metabolite or pathway...

Summary of *Escherichia coli* K-12 substr. MG1655, version 28.0
Tier 1 Highly Curated Database

Database Authors: Peter D. Karp¹, Anamika Kohli², Suzanne Paley³, Markus Krummenacker⁴, Ian Paulsen⁵, Amanda Mackie⁶, Lisa Moore⁷, Julio Collado-Vides⁸, Cesar Bonavides-Martinez⁹, Socorro Gama-Castro¹⁰, Alberto Santos-Zavala¹¹, Victor H Tierrafra¹², Paloma Lara Figueroa¹³

¹SRI International, ²Macquarie University, ³Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México

Summary
EcoCyc is a model-organism database for *Escherichia coli* K-12 MG1655. For more information, see URL EcoCyc.org.

The *E. coli* genome sequence used in EcoCyc has been updated from version 2 to version 3. GenBank accession number U00096.3, for release 28.0 of EcoCyc. Despite the involvement of EcoCyc staff in ongoing updates to the U00096 record, some annotation differences may be found between U00096 and EcoCyc, such as due to recent updates to EcoCyc.

Citations: [Kasler21]

Taxonomic Lineage: cellular organisms, Bacteria → bacterium, Pseudomonadota, Gammaproteobacteria, Enterobacterales, Enterobacteriaceae, *Escherichia*, *Escherichia coli*, *Escherichia coli* K-12, *Escherichia coli* K-12 substr. MG1655

Synonyms: *Escherichia coli* K-12 substr. MG1655

Unification Links: GOLD:Gp0072957, NCBI:Taxonomy:511145

Organism or Sample Properties	Relationship to Oxygen	Temperature Range	NCBI Genome Type
	anaerobic	mesophilic	reference

RefIcon	Total Genes	Protein Genes	RNA Genes	Pseudogenes	Size (bp)	NCBI Link
Chromosome	4,703	4,328	229	148	4,641,652	NCBI-GenBank:U00096.3
Genes without a physical map position	3					
Ortholog data available?	Yes					

Base de Datos de Genomas de *Escherichia coli* (EcoCyc).
Base de Datos del Genoma de *Saccharomyces cerevisiae* (SGD).
Base de Datos del Genoma de *Mycobacterium tuberculosis* (TBDB).
Pseudomonas Genome Database (Pseudomonas Genome DB)

¿Aplicaciones de la Bioinformática?

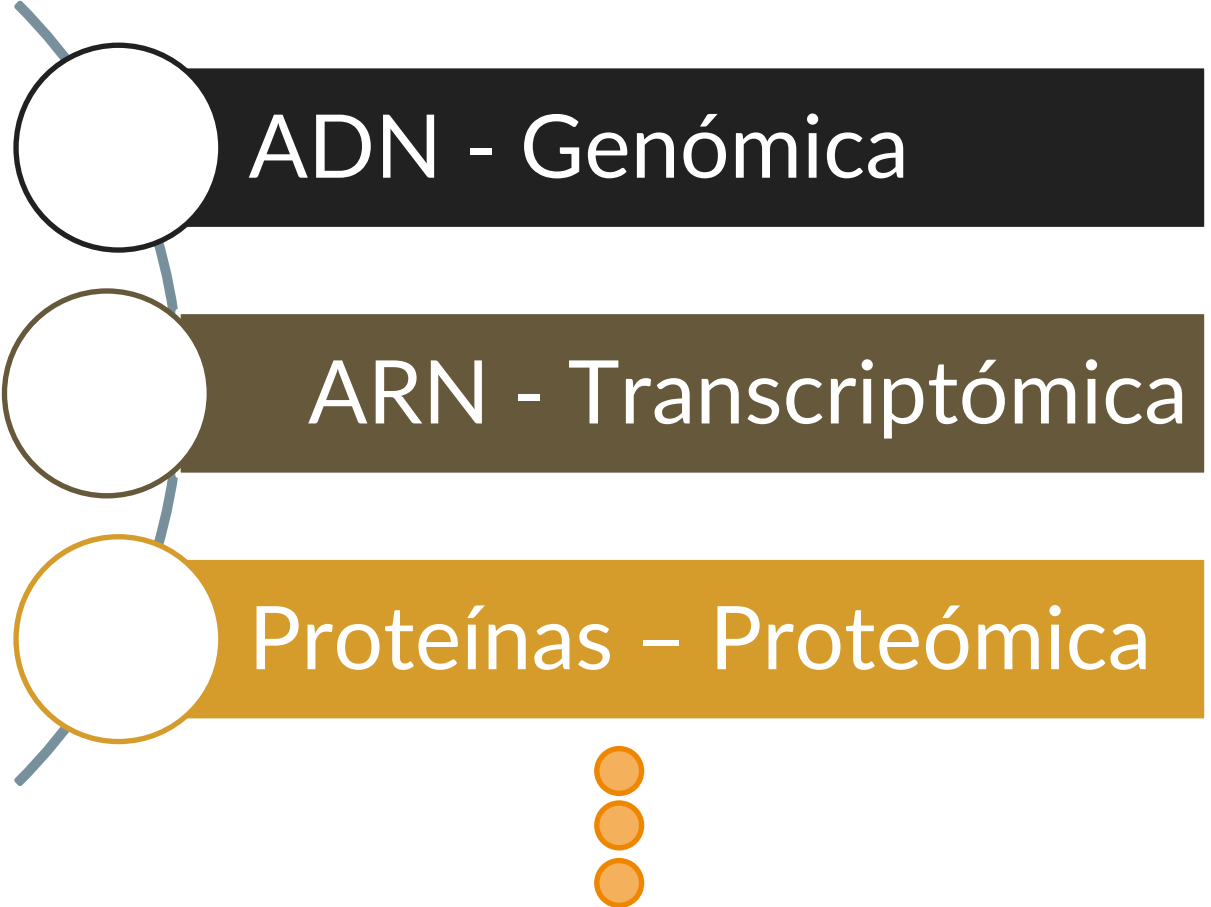
TECNOLOGIA



Secuenciación de nueva generación (NGS en sus siglas en inglés)



BAJO COSTE



¿Aplicaciones de la Bioinformática?

TECNOLOGÍA



Secuenciación de nueva
generación (NGS en sus
siglas en inglés)



BAJO COSTE



ANÁLISES



DATOS



NECESIDAD
DE
ANALISAR
LOS DATOS
GENERADOS

BIOINFORMÁTICA

