

# EXPERIMENTACIÓN *IN SILICO* DE COMUNIDADES MICROBIANAS EN PLANTAS DE AGAVE.

*@elracional*



# PROYECTO:

Este proyecto plantea realizar un análisis integral de los factores que estructuran la comunidad microbiana en la rizosfera de dos especies de plantas desérticas, *Agavelechuguilla* Torr, (Agavaceae), comúnmente conocido como lechuguilla, y *Echinocactus platyacanthus* Link et Otto (Cactaceae), comúnmente conocido como biznaga.



# PROBLEMA

Existen pocos estudios que han analizado los factores que afectan la diversidad y composición de la comunidad microbiana de plantas silvestres en zonas áridas (Makhalanyane et al., 2015). De hecho, la incógnita de si las raíces de estas plantas seleccionan comunidades microbianas particulares continúa sin resolverse.

Se quiere demostrar una correlación entre la diversidad microbiana y algunas propiedades del suelo, principalmente pH y el contenido de carbono (Andrew et al., 2012).

# HIPOTESIS

La estructura de las comunidades microbianas rizosféricas de las plantas de ambientes áridos está influenciada por las características del suelo y las variaciones climáticas acorde a la localización geográfica, y ante qué factores ambientales específicos responden, será posible establecer de manera experimental las variaciones en las poblaciones microbianas en respuesta a la modificación de estos factores.

## **Propuesta doctoral:**

Se obtendrá el mismo resultado de la *experimentación insilico* de comunidades microbianas, con respecto a la experimentación en ambiente controlado y en campo.

# OBJETIVO

## Objetivo general

Determinar los factores ambientales (bióticos y abióticos) que influyen sobre la estructura de las comunidades microbianas rizosféricas en plantas de zonas áridas, tomando como modelo a dos especies de amplia distribución, como son Agave lechuguilla y Biznaga.

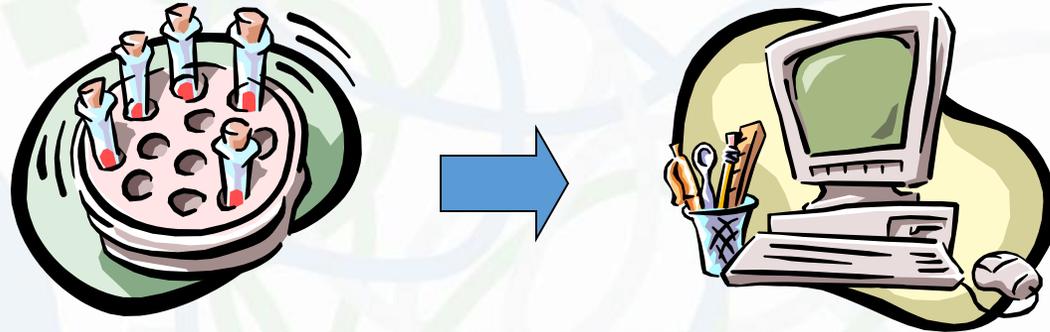
## Objetivos propuesta doctoral:

1. El desarrollo e implementación de herramientas que permitan el acceso, uso y manejo de varios tipos de información para la experimentación de comunidades microbianas in silico .
2. El desarrollo de nuevos algoritmos (fórmulas matemáticas y estadísticas), con los cuales se pueda relacionar partes de un conjunto enorme de datos, como por ejemplo métodos para localizar los factores ambientales, predecir su estructura o función de las comunidades microbianas.

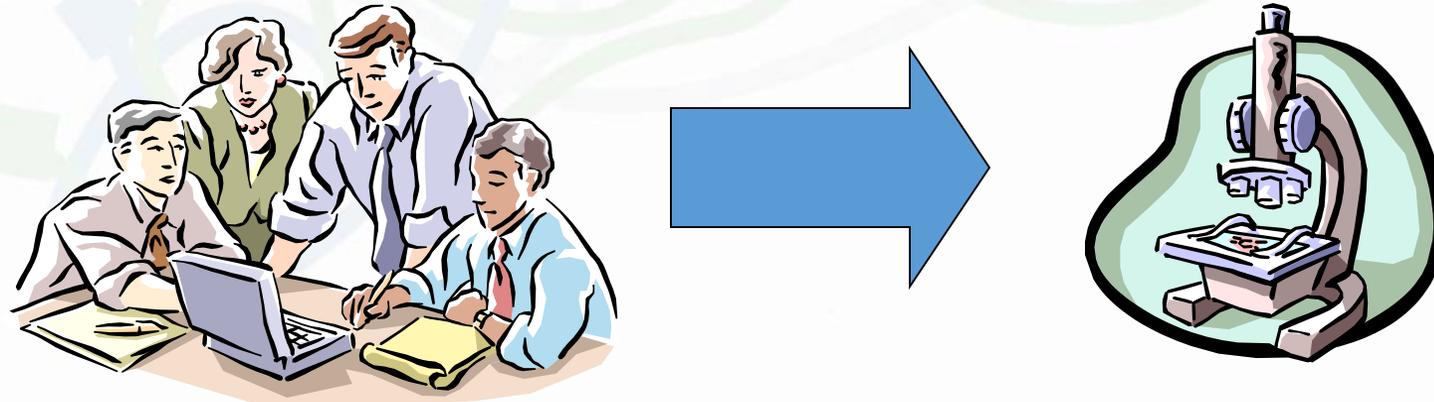
# METAS

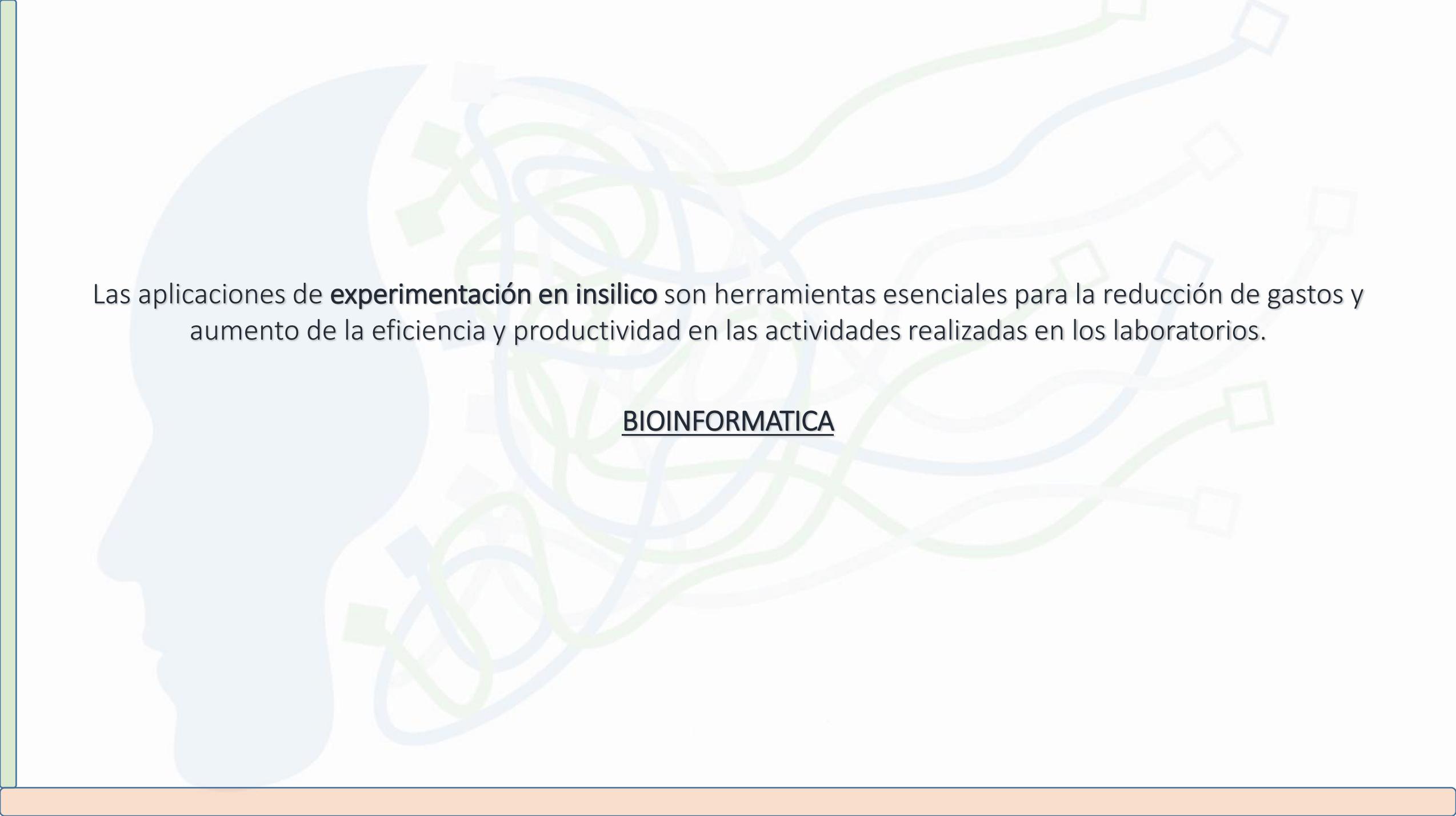
- a) Generar información científica básica que permita comprender cómo se estructuran las comunidades rizosféricas de zonas áridas.
- b) Determinar qué factores influyen dicha estructura para realizar inferencias acerca de las respuestas de estas comunidades ante el estrés ambiental.
- c) A futuro, esta información será importante para desarrollar estrategias de manejo y conservación que incluyan al componente microbiano como una parte importante de los mismos.
- d) Contribuir con información que permita diseñar estrategias para el combate a la erosión del suelo y a la recuperación de los suelos degradados en los desiertos mexicanos.

# EXPERIMENTACIÓN *IN SILICO*



El nuevo paradigma es que todos los genes se conocerán y el punto de partida para una investigación biológica será **TEÓRICO**.





Las aplicaciones de **experimentación en insilico** son herramientas esenciales para la reducción de gastos y aumento de la eficiencia y productividad en las actividades realizadas en los laboratorios.

## BIOINFORMATICA

# ¿QUÉ ES LA BIOINFORMÁTICA?

MATEMÁTICAS

Desarrollo de  
algoritmos

Testeo de hipótesis

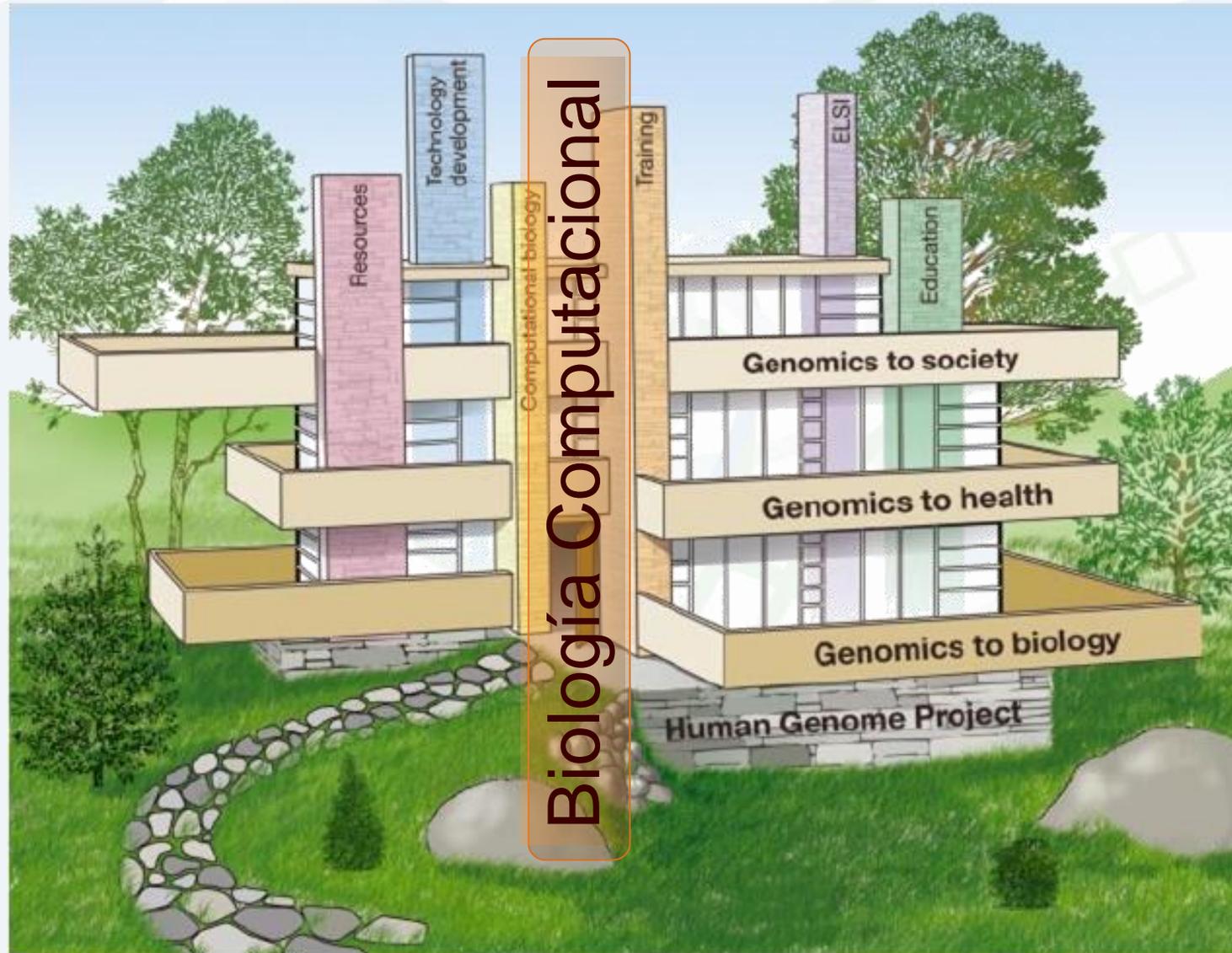
Plataforma  
Web

INFORMÁTICA

Análisis (semi)automatizado  
a gran escala

BIOLOGÍA

# BIOINFORMÁTICA



The future of genomics rests on the foundation of the Human Genome Project.

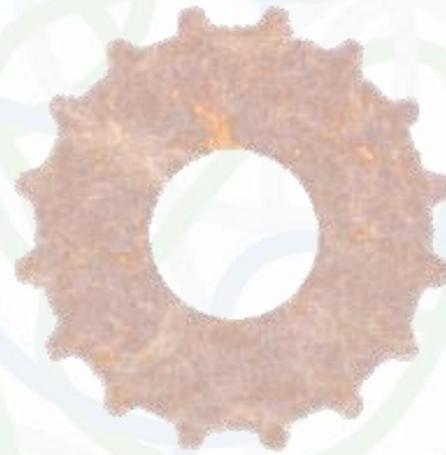
# LA BIOINFORMÁTICA FUNCIONAMIENTO



```
ATGTGCAATGCTT
CGTTACGGCTCAA
TATGCCGCAGTAA
GCTGCAGTATCCG
CCGCAGTAACTGG
GCCGCAG.....
```

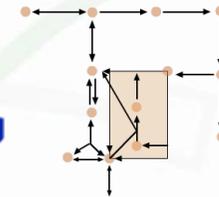
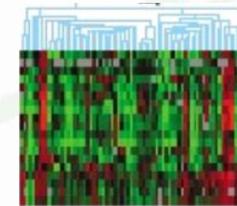
## Datos

Organizar y procesar datos



## Herramientas bioinformáticas

Dar sentido biológico a la información



## Conocimiento

Predicciones, diagnósticos o reconocimientos de comunidades microbianas en estudio.

# DATOS:

## Codificación de la información

- Sistemas vivos: información digital cuaternaria (A,T,G,C) en secuencias unidimensionales.

A = adenina

C = citosina

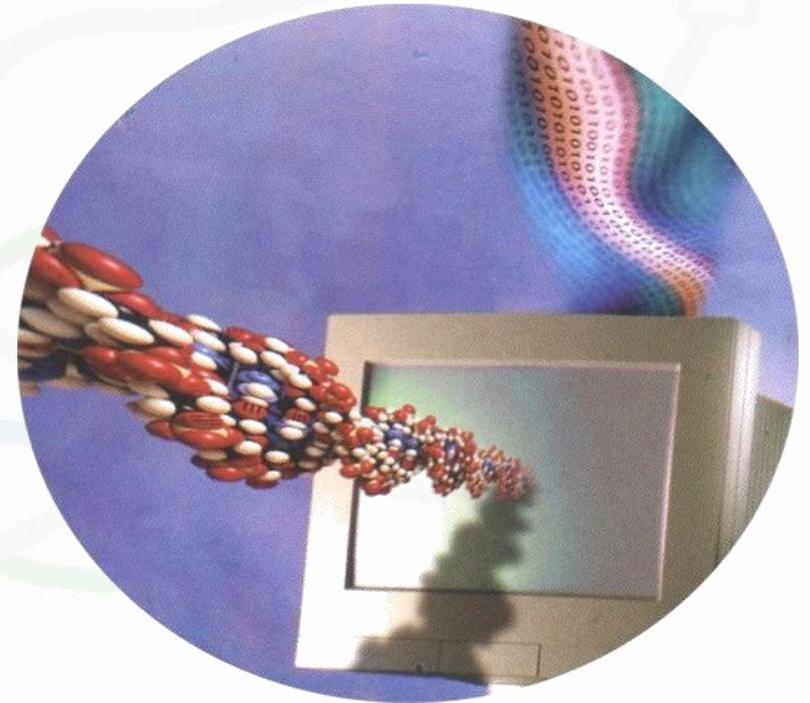
G = guanina

T = timina

- Sistema de Computo: información digital binaria en conjuntos de 0 y 1.

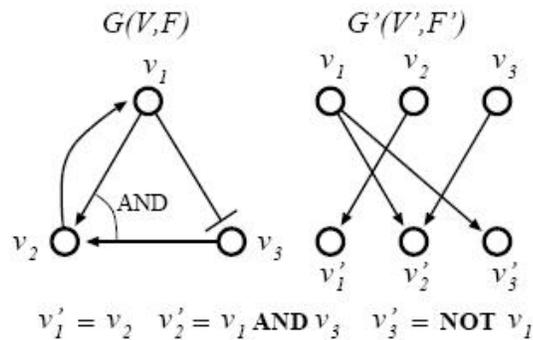
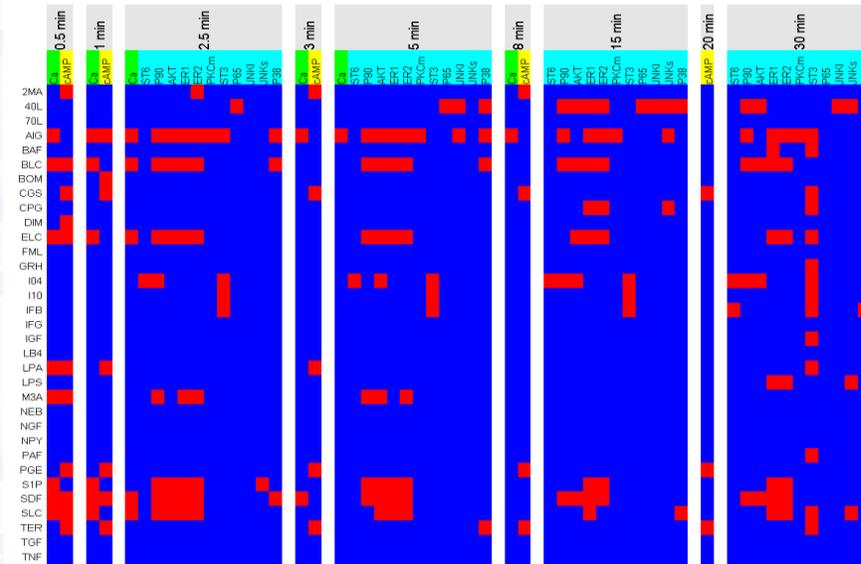
0 = False

1 = True



# INFORMACIÓN

Se pueden aproximar los estados mediante variable booleanas (binarias): encendido o apagado



INPUT			OUTPUT		
$v_1$	$v_2$	$v_3$	$v'_1$	$v'_2$	$v'_3$
0	0	0	0	0	1
0	0	1	0	0	1
0	1	0	1	0	1
0	1	1	1	0	1
1	0	0	0	0	0
1	0	1	0	1	0
1	1	0	1	0	0
1	1	1	1	1	0

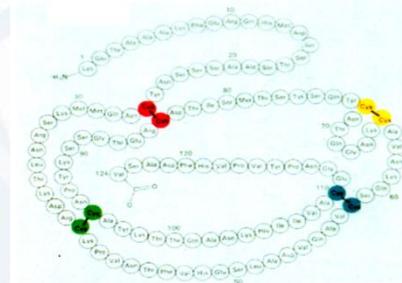
El resultado será un grafo en cuyos nodos se ponen funciones lógicas. Ha sido un modelo muy productivo.



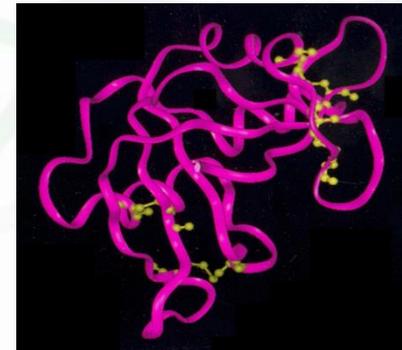
# REPRESENTACION GRAFICA



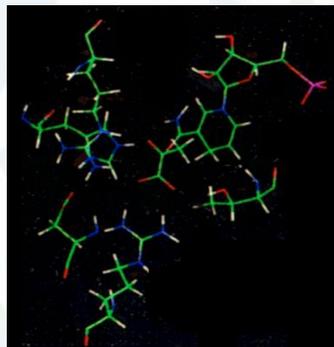
Secuencia DNA



Secuencia Proteína



Estructura 3D



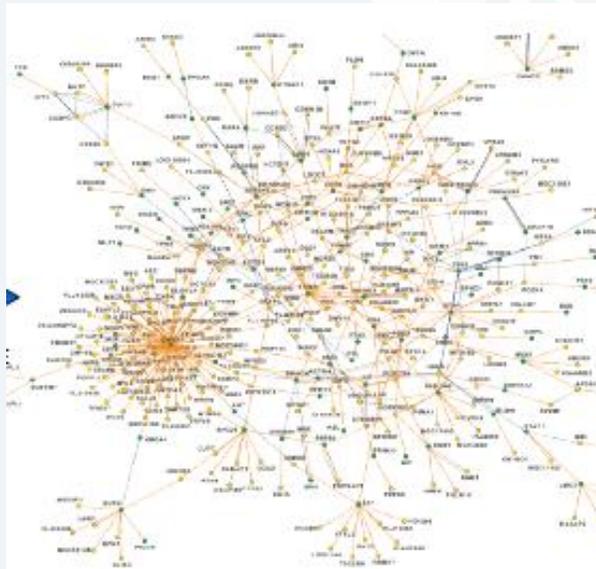
Reconocimiento

# REALIDAD...

>gi|261252063|ref|NZ\_ACZV0100005.1| *Vibrio orientalis* CIP 102891 VIA.Contig80, whole genome shotgun sequence

```
ACGCGTTAAGTAGACCGCCTGGGGAGTACGGTCGCAAGATTAACAACTCAAATGAATTGACGGGGCCCCG
ACAAGCGGTGGAGCATGTGGTTAATTCGATGCAACGCGAAGAACCCTTACCTACTCTTGACATCCAGAGA
AGCCGGAAGAGATTCTGGTGTGCCCTTCGGGAACCTCTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTG
TTGTGAAATGTGGGTTAAGTCCCAGCAACGAGCGCAACCCTTATCCTTGTGGCAGCGAGTAATGTCCGG
GAACCTCCAGGGAGACTGCCGGTGATAAACCGGAGGAAGGTGGGGACGACGTCAGTCAAGTCATCATGGCCCTTA
CGAGTAGGGCTACACACGTGCTACAATGGCGCATACAGAGGGCAGCCAACTTGGCAAAGTGAGCGAATCC
CAAAAAGTGCCTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAAATCGCTAGTAATCG
TGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGG
CTGCAAAAAGAAGTAGGTAGTTAACCCTTCGGGAGAACGCTTACCCTTGTGGTTCATGACTGGGGTGAA
GTCGTAACAAGGTAGCCCTAGGGGAACCTGGGGCTGGATCACCTCCTTATACGATGATTACTCAGCATGA
GTGTCCACACAGATTGATATGCTTTATTAGAGCTTTGAGGGCTATAGCTCAGCTGGGAGAGCGCTTCC
```

Secuencia DNA



Reconocimiento

```
>P03176|KITH_HHV11 Thymidine kinase - Human herpesvirus 1 (st
MASYPCHQHASAFDQAARSRGHNNRRRTALRPRRQQEATEVRPEQKMPDLLRVYIDGPHGM
GKTTTTQLLVALGSRDDIVYVPEPMTYWRVLGASETIANIYTTQHRLDQGEISAGDAAVV
MTSAQITMGMPYAVTDAVLAPHIGGEAGSSHAPPPALTLIFDRHPAIALLCYPAARYLMG
SMTFQAVLAFVALIPPTLPGTNIVLGLALPEDRHIDRLAKRQRPGERLDLAMLAAIRRVYG
LLANTVRYLQCGGSWREDWGLSGTAVPPQGAEPQSNAGPRPHIGDTLFTLFRAPPELLAP
NGDLYNVFAWALDVLAKRLRSMHVFIIDYDQSPAGCRDALLQLTSGMVQTHVTTPGSIPT
ICDLARTFAREMGEAN
```

Secuencia Proteína

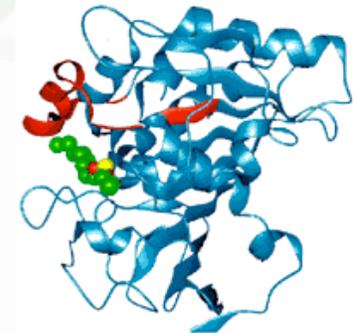
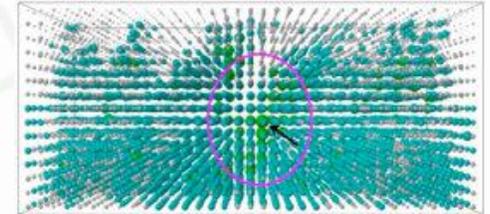
ATOM	95	CE2	TRP	115	28.381	8.071	33.915	1.00	10.00
ATOM	96	CE3	TRP	115	27.500	9.825	32.526	1.00	10.00
ATOM	97	CZ2	TRP	115	27.750	7.155	33.103	1.00	10.00
ATOM	98	CZ3	TRP	115	26.888	8.895	31.705	1.00	10.00
ATOM	99	CH2	TRP	115	27.053	7.584	32.002	1.00	10.00
ATOM	100	N	ASP	116	26.290	11.255	36.778	1.00	10.00
ATOM	101	CA	ASP	116	25.763	10.825	38.096	1.00	10.00
ATOM	102	C	ASP	116	24.689	11.802	38.607	1.00	10.00
ATOM	103	O	ASP	116	24.564	12.103	39.797	1.00	10.00
ATOM	104	CB	ASP	116	26.872	10.617	39.142	1.00	50.00
ATOM	105	CG	ASP	116	26.368	10.397	40.557	1.00	50.00
ATOM	106	OD1	ASP	116	25.812	9.294	40.721	1.00	50.00
ATOM	107	OD2	ASP	116	26.590	11.276	41.416	1.00	50.00
ATOM	108	N	PHE	117	23.915	12.348	37.709	1.00	10.00
ATOM	109	CA	PHE	117	22.766	13.148	38.156	1.00	10.00

Estructura 3D

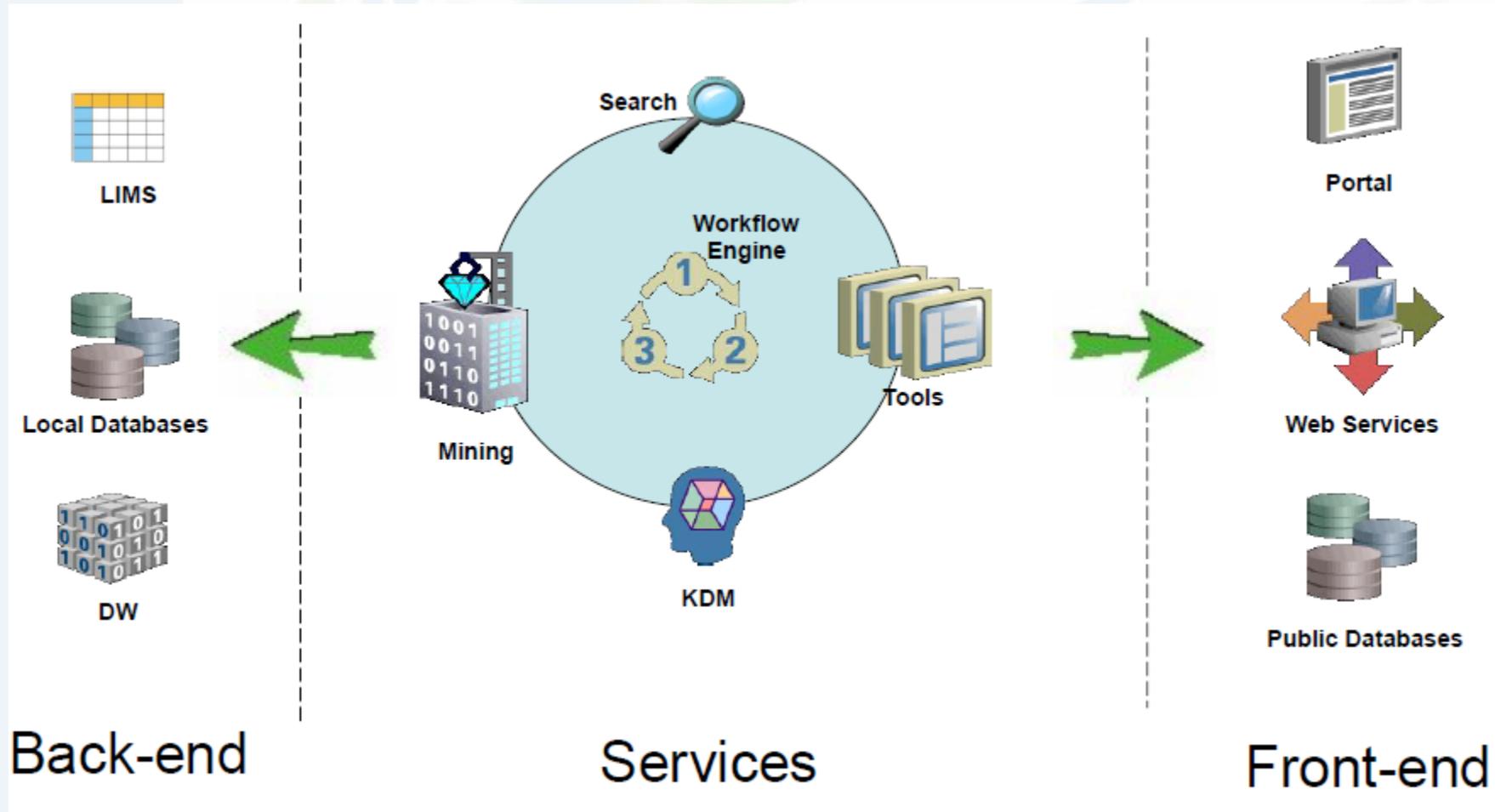
# TECNOLOGIAS DE LA INFORMACION APLICADAS A LA BIOINFORMATICA



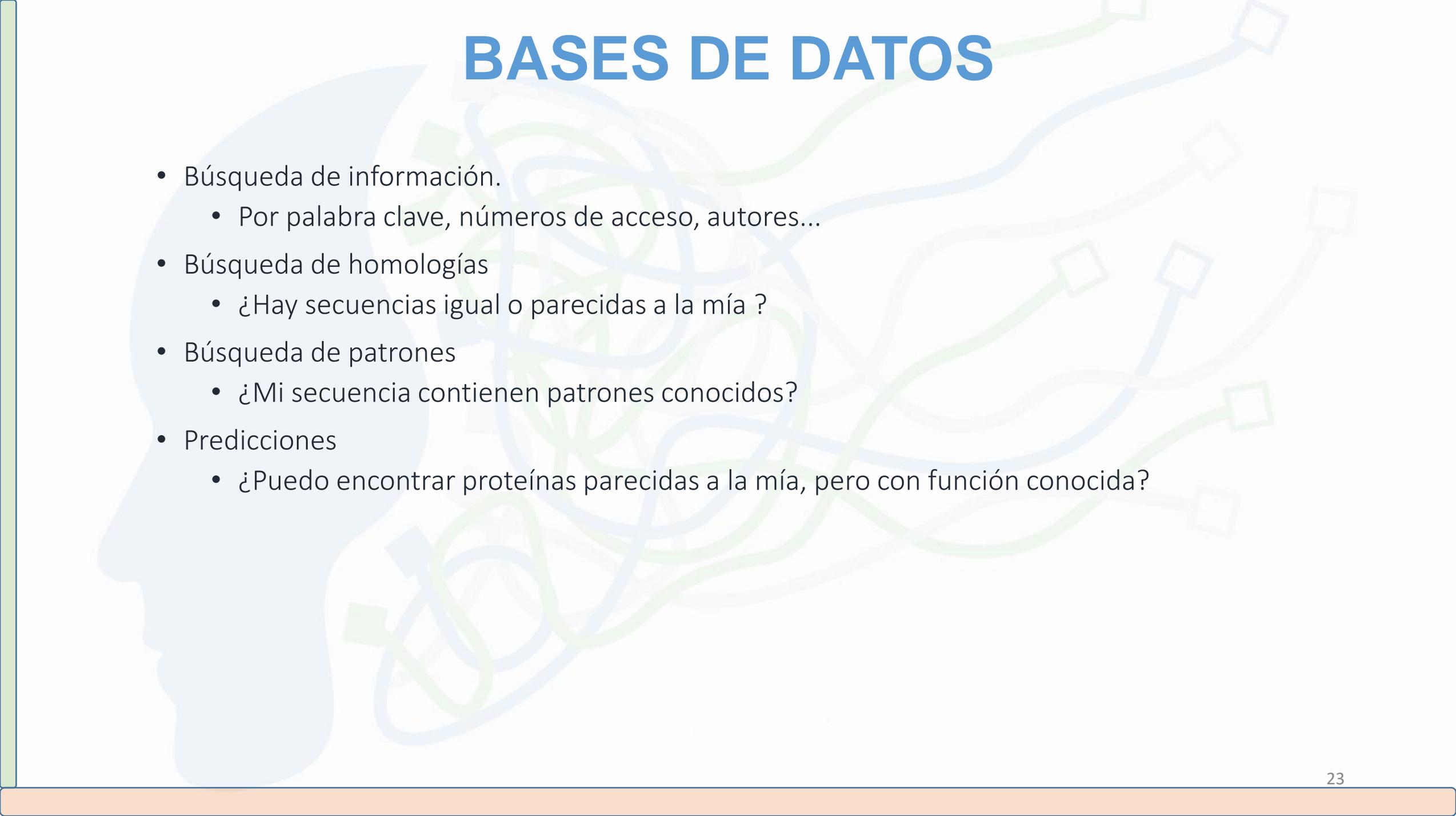
- Bases de Datos
- Computación Gráfica
- Minería de Datos
- Colaboración y Redes
- Modelado y Simulación
- Reconocimiento de patrones



# NIVELES DE PLATAFORMA BIOINFORMÁTICA



# BASES DE DATOS



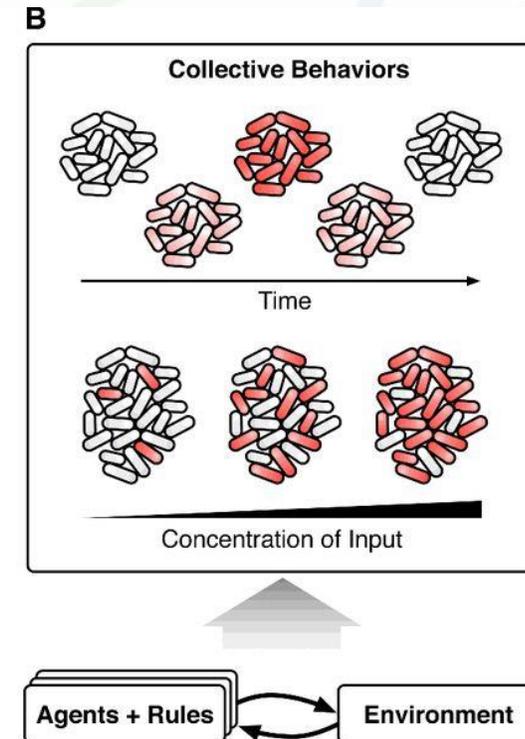
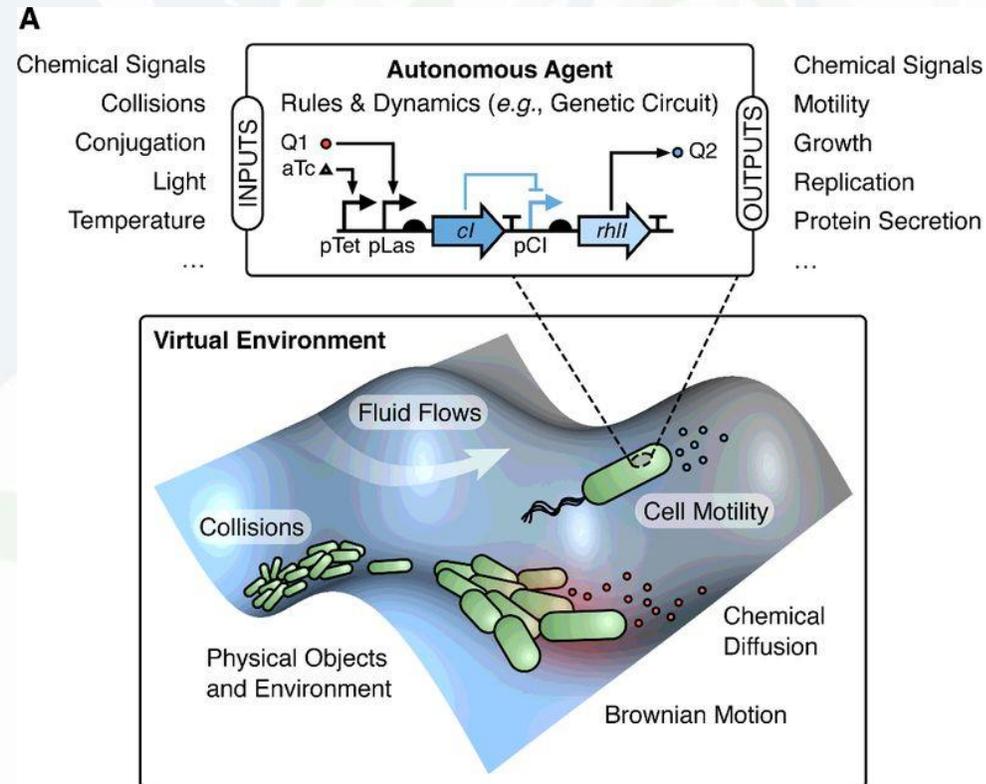
- Búsqueda de información.
  - Por palabra clave, números de acceso, autores...
- Búsqueda de homologías
  - ¿Hay secuencias igual o parecidas a la mía ?
- Búsqueda de patrones
  - ¿Mi secuencia contienen patrones conocidos?
- Predicciones
  - ¿Puedo encontrar proteínas parecidas a la mía, pero con función conocida?

# TIPOS DE BASES DE DATOS

- Existen cientos de BD en número tan elevado que no es práctico enumerarlas:
  - Bases de datos [bibliográficas](#)
  - Bases de datos [taxonómicas](#)
  - Bases de datos de [nucleótidos](#)
  - Bases de datos [genómicas](#)
  - Bases de datos de [proteínas](#)
  - Bases de datos de [microarrays](#)

# MODELAMIENTO Y SIMULACIÓN

- Para modelar eventos que simulen interacciones entre proteínas hasta el análisis de caminos biológicos potenciales se requiere de modelos de manejo de eventos tales como:
- Event-Driven (programación van determinados por los sucesos que ocurran en el sistema, definidos por el usuario o que ellos mismos provoquen)
- Hybrid simulation

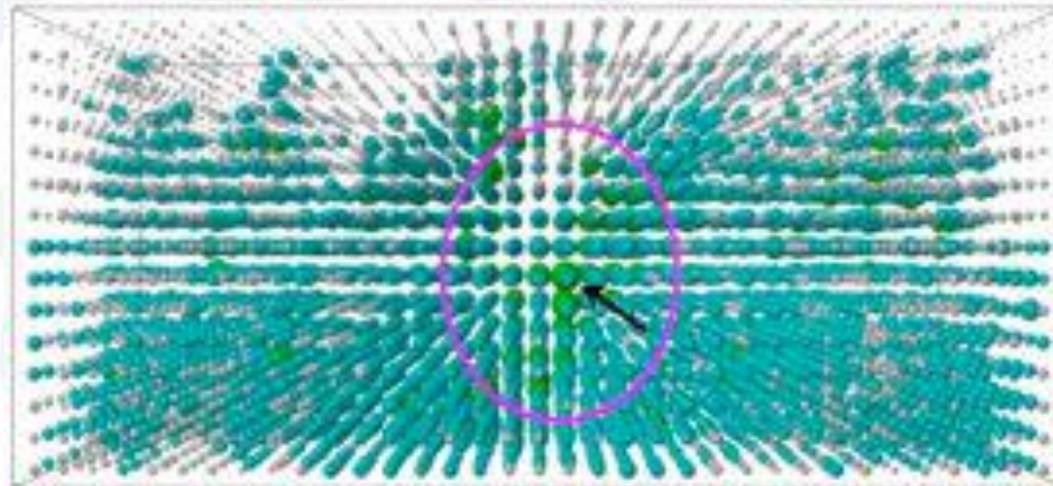


# RECONOCIMIENTO DE PATRONES

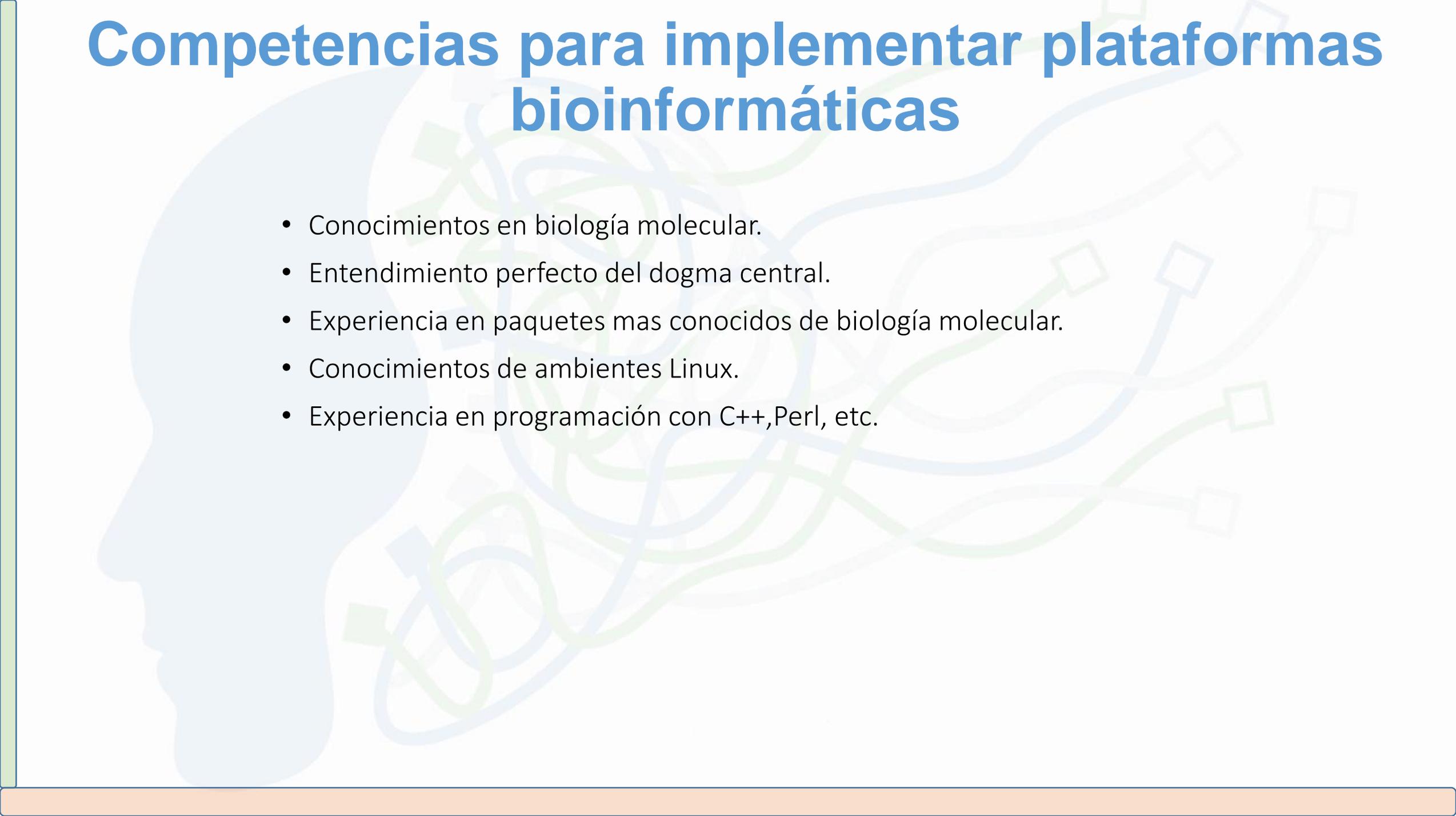
Tienen aplicación en investigaciones de biología molecular y básicamente se requiere de aplicar tecnologías de sistemas Expertos e Inteligencia Artificial en:

- Comparación de secuencias
- Predicción de estructuras y modelación molecular.
- Evolución molecular

Ejemplo Dr. Lulu Qian en el Instituto de Tecnología de California (Caltech) han creado circuitos bioquímicos sintéticos que pueden realizar el procesamiento de la información a nivel molecular, una red neuronal artificial que consiste en ADN en lugar de hardware y software.



# Competencias para implementar plataformas bioinformáticas



- Conocimientos en biología molecular.
- Entendimiento perfecto del dogma central.
- Experiencia en paquetes mas conocidos de biología molecular.
- Conocimientos de ambientes Linux.
- Experiencia en programación con C++,Perl, etc.

# Tools



- Herramientas de análisis
  - Clustal, Blast, Phylip, Rasmol, etc.
- Módulos de lenguajes
  - Bioperl, Biopython, Biojava, etc.
- Frameworks bioinformáticos
  - Web Services, Databases, Portals, Workflows, etc.

# Identificación de Componentes

## *hard*

- Clustering
- GigaSwitch
- Broadband Internet

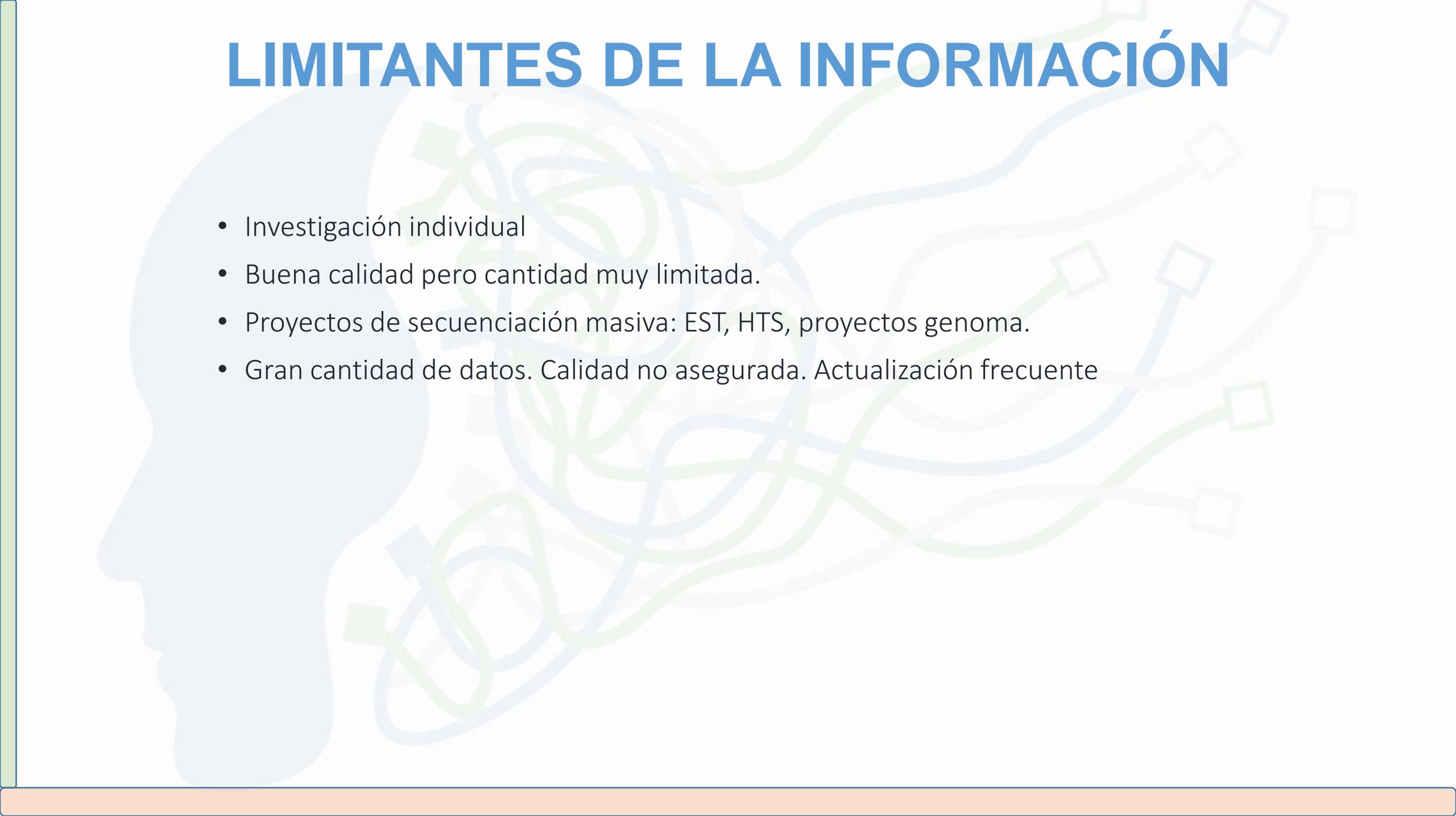


## *Soft*

- Software Architecture
  - Database Engines
    - Analyses Tools
  - Middleware tool

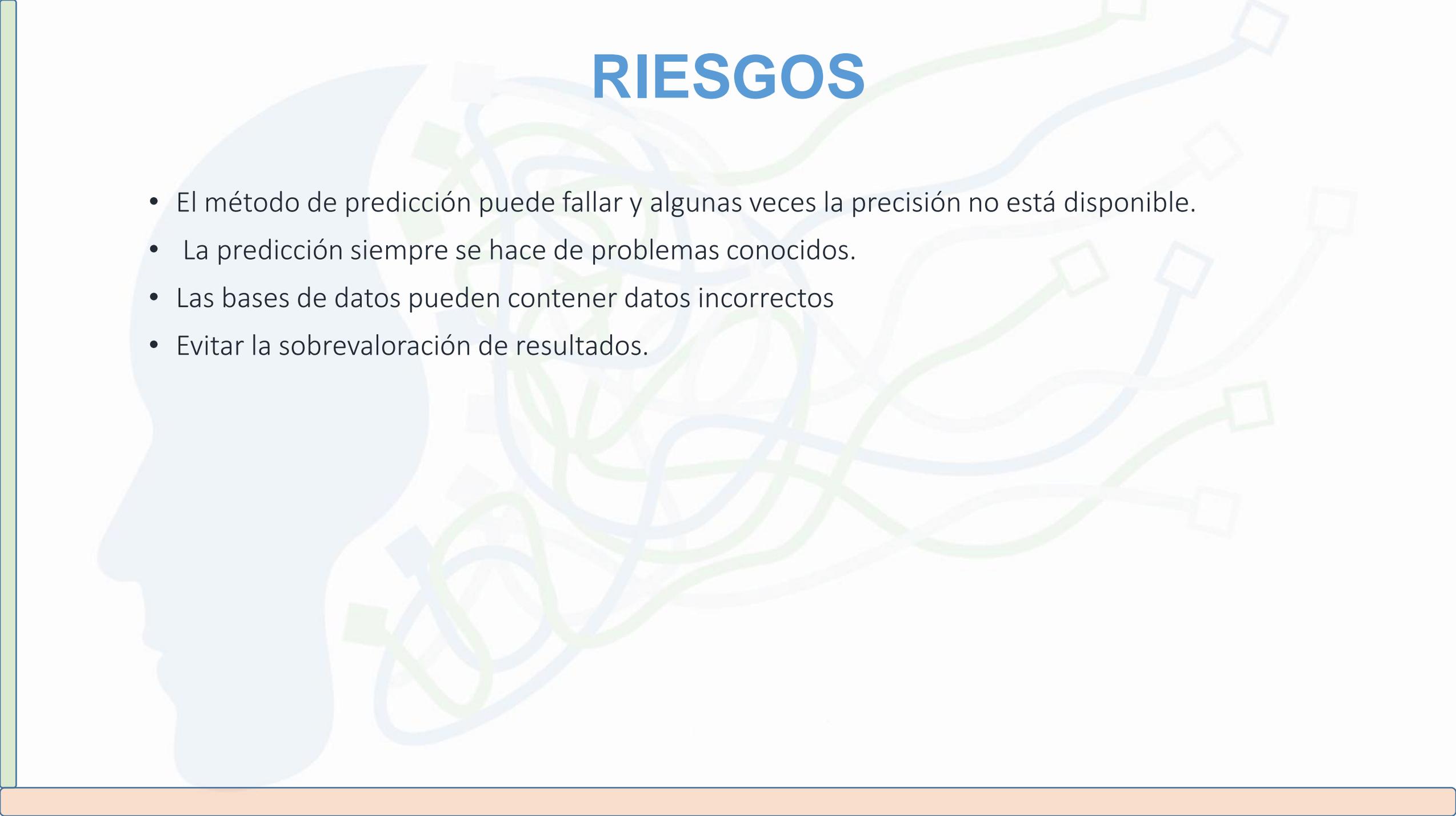


# LIMITANTES DE LA INFORMACIÓN



- Investigación individual
- Buena calidad pero cantidad muy limitada.
- Proyectos de secuenciación masiva: EST, HTS, proyectos genoma.
- Gran cantidad de datos. Calidad no asegurada. Actualización frecuente

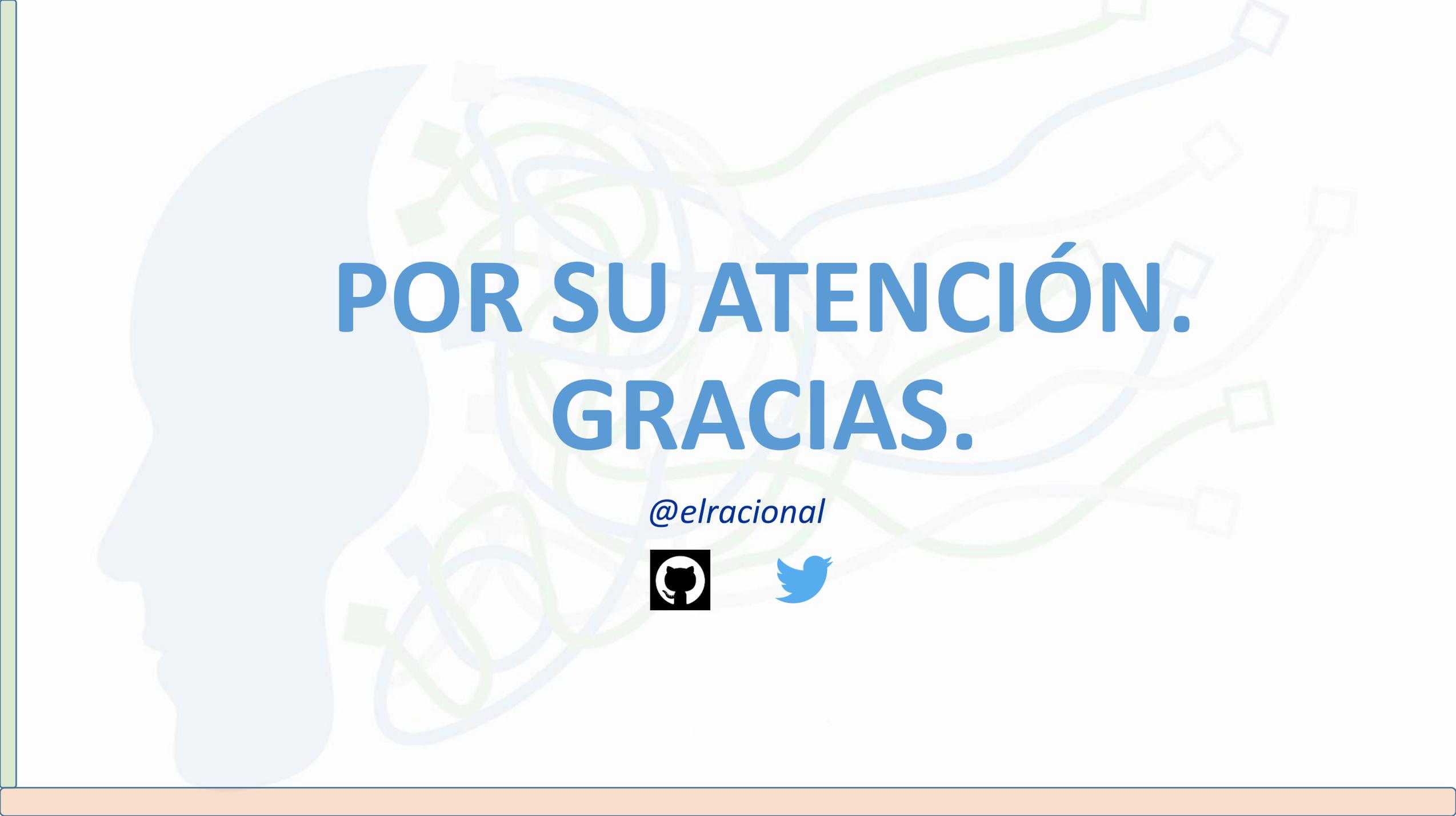
# RIESGOS



- El método de predicción puede fallar y algunas veces la precisión no está disponible.
- La predicción siempre se hace de problemas conocidos.
- Las bases de datos pueden contener datos incorrectos
- Evitar la sobrevaloración de resultados.

# UN RESUMEN A MODO DE CONCLUSIONES

- La biología y la medicina no son las mismas desde que tuvieron relaciones con la computación.
- Se pueden buscar respuestas, formular hipótesis y hasta hacer experimentos sin un laboratorio ni “salir de casa”.
- Los profesionales de las áreas de salud, química, biología y afines, necesitan, *ineludiblemente* de las herramientas bioinformáticas, no como una habilidad opcional sino básica.
- Bioinformática es un campo donde la computación y la ingeniería de software tienen un alto potencial de aplicación.
- El desarrollo de la bioinformática requiere de avances significativos en investigación en temas de ingeniería.
- Existen posibilidades de proyectos en diferentes campos en tecnologías de la información.



**POR SU ATENCIÓN.  
GRACIAS.**

*@elracional*

