

Máster Universitario en Bioinformática

Generación y mantenimiento de datos ómicos

Curso académico 2024-25



Universidad
Internacional
de Valencia

Dr. Jordi Tronchoni
jordi.tronchoni@professor.universidadviu.com

15/04/2024


De:
 Planeta Formación y Universidades



Tema 1

Introducción a la Bioinformática

15/04/2024

- 
1. ¿Qué es la bioinformática?
 2. Un poco de historia



1. ¿Qué es la bioinformática?

¿Qué es la bioinformática?

The screenshot shows the MeSH (Medical Subject Headings) interface for the term 'Computational Biology'. The page is from the NCBI website. The main heading is 'Computational Biology', with a description: 'A field of biology concerned with the development of techniques for the collection and manipulation of biological data, and the use of such data to make biological discoveries or predictions. This field encompasses all computational methods and theories for solving biological problems including manipulation of models and datasets. Year introduced: 1997'. Below this, there are 'PubMed search builder options' and 'Subheadings' with checkboxes for various categories like classification, economics, education, ethics, history, instrumentation, legislation and jurisprudence, methods, organization and administration, standards, statistics and numerical data, and trends. There are also options to 'Restrict to MeSH Major Topic' and 'Do not include MeSH terms found below this term in the MeSH hierarchy'. The 'Tree Number(s)' are H01.158.273.180, L01.313.124, and the 'MeSH Unique ID' is D019295. The 'Entry Terms' list includes 'Biology, Computational', 'Computational Molecular Biology', 'Biologies, Computational Molecular', 'Biology, Computational Molecular', 'Computational Molecular Biologies', 'Molecular Biologies, Computational', 'Molecular Biology, Computational', 'Bio-Informatics', 'Bio Informatics', 'Bio-Informatic', 'Bioinformatics', and 'Bioinformatic'. The 'Previous Indexing' section lists 'Molecular Biology (1992-1996)'. The 'See Also' section lists 'Medical Informatics'. On the right side, there is a 'PubMed Search Builder' section with a search box and buttons for 'Add to search builder', 'AND', and 'Search PubMed'. Below this is a 'Related Information' section with links to 'PubMed', 'PubMed - Major Topic', 'Clinical Queries', 'NLM MeSH Browser', and 'dbGaP Links'. The 'Recent Activity' section shows a list of recent searches, including 'Computational Biology', 'bioinformatics (1)', 'genome (41)', 'Efficient Enrichment of Bacterial mRNA from Host-Bacteria Total RNA Samples', and 'Lactiplantibacillus pentosus'. At the bottom right, there is a date '15/04/2024'.

Vamos a empezar por lo básico, respondiendo a la pregunta de ¿Qué es la bioinformática?

MeSH funciona a modo de diccionario cuyas entradas definen las búsquedas indexadas de los artículos de PubMed

MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for **PubMed**

¿Qué es la bioinformática?

Computational Biology

A field of biology concerned with the development of techniques for the collection and manipulation of biological data, and the use of such data to make biological discoveries or predictions. This field encompasses all computational methods and theories for solving biological problems including manipulation of models and datasets.

Year introduced: 1997

15/04/2024

Un campo de la biología que se ocupa del desarrollo de técnicas para la recopilación y manipulación de datos biológicos y el uso de dichos datos para realizar descubrimientos o predicciones biológicas. Este campo abarca todos los métodos y teorías computacionales para resolver problemas biológicos, incluida la manipulación de modelos y conjuntos de datos.

¿Qué es la bioinformática?

“La unión de biología e informática: la bioinformática involucra la tecnología que utiliza computadoras para el almacenamiento, manipulación y distribución de información relacionada con macromoléculas biológicas como ADN, ARN y proteínas. El uso de esta tecnología es absolutamente indispensable para la recopilación de información y la creación de conocimientos.”

Luscombe; What is bioinformatics? A proposed definition and overview of the field - ([Luscombe et al. 2001](#))

15/04/2024

La definición actualmente aceptada, pero quizás ya obsoleta por su especificidad, se definió hace 20 años por Luscombe, y se define como la unión entre biología y computación, siendo una ciencia que utiliza los últimos avances tecnológicos para el almacenaje, análisis y distribución de la información relacionada con las macromoléculas como el ADN, ARN o proteínas, siendo la tecnología y su desarrollo indispensables para el manejo eficiente de los datos. Esta primera frase se citó por primera vez en el artículo mencionado.

¿Qué es la bioinformática?

De forma general, se trata de la aplicación de tecnologías computacionales y la estadística a la gestión y análisis de datos biológicos.

Es un área por tanto muy interdisciplinar.

Principales objetivos generales:

1. La organización, curación y acceso del alto volumen de datos biológicos disponibles hoy día.
2. El desarrollo de recursos computacionales (algoritmos, programas, bases de datos, análisis estadísticos, etc.) que permitan el análisis individual y a gran escala de la información biológica disponible.
3. El uso de estos recursos para analizar los datos e interpretar los resultados de una manera útil.

15/04/2024

Personalmente, prefiero una definición más amplia:

De forma general, se trata de la aplicación de tecnologías computacionales y la estadística a la gestión y análisis de datos biológicos. De carácter interdisciplinar con tres objetivos principales:

1. La organización, curación y acceso al alto volumen de datos biológicos disponibles hoy en día.
2. El desarrollo de recursos computacionales (algoritmos, programas, bases de datos, análisis estadísticos, etc.) que permitan el análisis individual y a gran escala de la información biológica disponible.
3. El uso de estos recursos para analizar los datos e interpretar los resultados de una manera útil.

¿Qué es la bioinformática?

La obtención de los datos hoy en día se realiza de varias formas, principalmente mediante *Next Generation Sequencing* **NGS**.

Se utilizan unas máquinas denominadas **secuenciadores** que son capaces de *leer* el material biológico.

Estos secuenciadores proporcionan una enorme cantidad de datos,

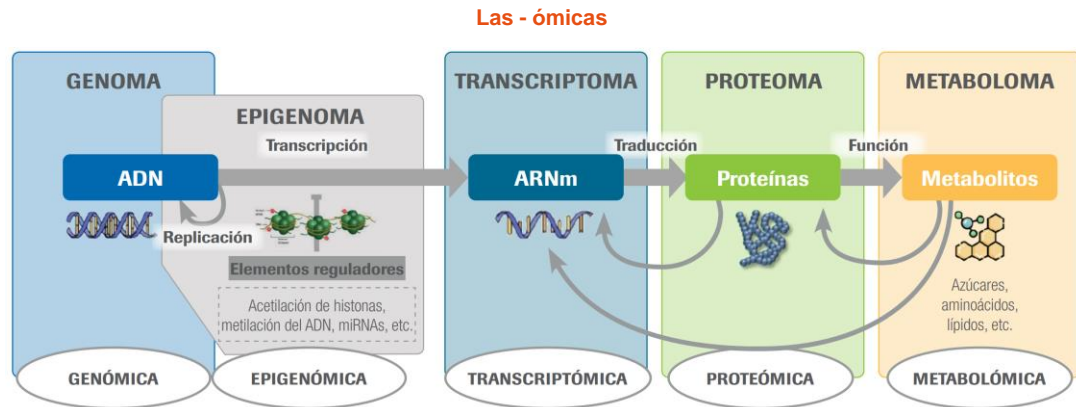
el big data de la biología:
datos enormes que se analizan computacionalmente para identificar patrones, asociaciones y tendencias

15/04/2024

Merece la pena aclarar que actualmente cuando se piensa en bioinformática, solemos estar hablando de bioinformática genómica. Esto es porque esta rama fue la que primero explotó con el boom de los datos masivos y está estrechamente ligada al concepto de secuenciación de próxima generación, secuenciación de alta capacidad, secuenciación de alto rendimiento, secuenciación de datos masivos o *Next Generation Sequencing* o *NGS* por sus siglas en inglés.

¿Qué es la bioinformática?

Sobre el NGS en la bioinformática, podemos dividirlo en función del material biológico que queremos analizar:



Fundación Instituto Roche (2019). Informe "Anticipando Ciencias Ómicas".
Recuperado de https://www.institutoroche.es/static/archivos/Informes_anticipando_CIENCIAS_OMICAS.pdf

15/04/2024

Cuando se habla de NGS, se utiliza indistintamente para citar la técnica o para hablar del objeto de análisis, es decir, las ciencias ómicas.

¿Qué es la bioinformática?

Las áreas más utilizadas actualmente:

- **Genómica**

Bioinformática Genómica

workflows/pipelines específicos para:

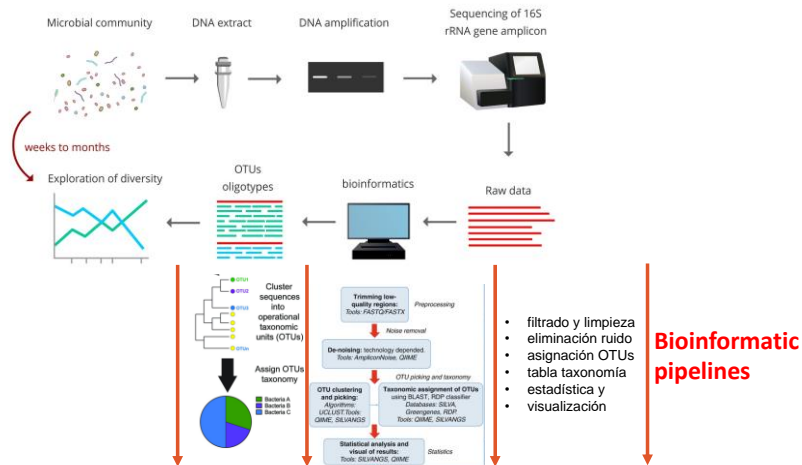
- **DNAseq**
- **RNAseq**
- **Metagenómica**
- **Metataxonómica**
- **etc...**

15/04/2024

En la actualidad, solemos pensar en bioinformática genómica cuando se cita la bioinformática. La bioinformática genómica es una rama de aplicabilidad de la bioinformática. Una rama muy amplia, en la que podemos encontrar expertos en distintos flujos de trabajo. Lo cual no implica ser experto en el análisis de distintos organismos para ese flujo de trabajo, es decir, veremos que distintos problemas conllevan distintas soluciones. Habitualmente esto significará, usar distintas herramientas, ya sea por el organismo, la técnica de secuenciación (muy habitual), el objetivo de la secuenciación, etc....

¿Qué es la bioinformática?

Flujo de trabajo o **workflow**: desde el diseño del experimento a la interpretación de resultados



15/04/2024

¿A qué nos referimos cuando hablamos de **workflow**? Generalmente nos referiremos al flujo de trabajo que existe desde la recogida de la muestra, hasta la obtención de conclusiones.

Vamos a verlo en un ejemplo, en este caso el objetivo es conocer la diversidad bacteriana (microbiana) de una muestra de suelo, podría ser de cualquier otro nicho ecológico pero como veremos esto tendría que tenerse en cuenta ya que variaría el flujo de trabajo.

Este es por tanto un ejemplo de bioinformática genómica, concretamente de metataxonomía.

Dentro de este flujo de trabajo o workflow, hablaremos de protocolos de extracción de ADN o protocolos de preparación de librerías de ADN a la parte de “wetlab” o de laboratorio de biología molecular. Pudiendo utilizarse workflow para concretar entre la parte en la que se obtienen las secuencias en crudo hasta la parte en la que se obtienen resultados analizados, figuras, gráficas etc...

En cambio “pipeline” suele utilizarse específicamente para aquella parte en la que interconectamos programas, algoritmos o datos cuya salida es la entrada del siguiente programa.

En muchas ocasiones, ambos términos se utilizarán indistintamente.

¿Qué es la bioinformática?

Las áreas más utilizadas actualmente:

- **Genómica**

Bioinformática Genómica

workflows/pipelines específicos para:

- **DNAseq**
- **RNAseq**
- **Metagenómica**
- **Metataxonómica**
- **etc...**

- **Proteómica**

Bioinformática Estructural

- **Farmacológica**

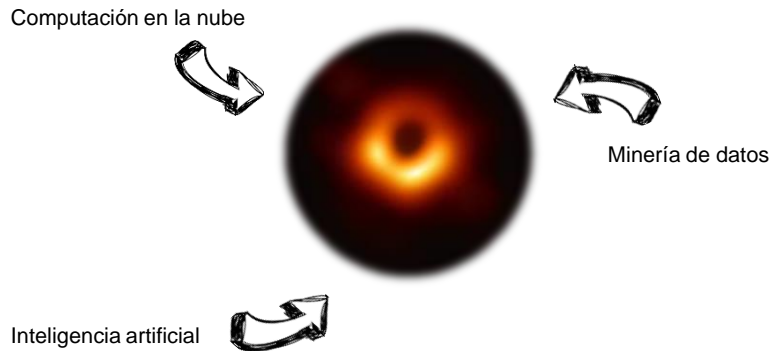
Bioinformática Farmacológica

15/04/2024

Además de la bioinformática genómica, las áreas más utilizadas actualmente son la proteómica diferencial y estructural, o la aplicabilidad de la estructural, la farmacológica.

Pero la bioinformática no maneja únicamente estas áreas que hemos comentado.

¿Qué es la bioinformática?



15/04/2024

Debido al componente computacional de la bioinformática, esta área crece y se desarrolla por los avances en computación. Avances como los que estamos viendo en los últimos años en computación en la nube. Un ejemplo de ello son nuestros WorkSpaces, la minería de datos o la inteligencia artificial. Todos ellos elementos en constante avance actualmente, lo que hacen de la bioinformática un área de trabajo interesante hoy en día.

¿Qué es la bioinformática?

CIENCIA

La IA resuelve un enigma básico de la biología: cómo se forman proteínas

- Lo ha conseguido el programa AlphaFold, creado por la compañía DeepMind a partir de algoritmos de aprendizaje profundo

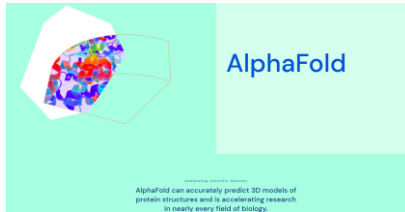


Las células inmunitarias liberan proteínas llamadas citoquinas que alertan al resto del sistema inmunitario de la presencia de un virus (Servicio Ilustrado (Automático) / EP)

15/04/2024

Entre los muchos frentes abiertos y excitantes de la bioinformática, uno de ellos es sin duda la aplicabilidad de la inteligencia artificial a la proteómica estructural. AlphaFold ha supuesto un cambio de paradigma en la bioinformática estructural, y ese cambio de paradigma está sucediendo ahora.

AlphaFold



Copilot

Your AI pair programmer

GitHub Copilot uses the OpenAI Codex to suggest code and entire functions in real-time, right from your editor.



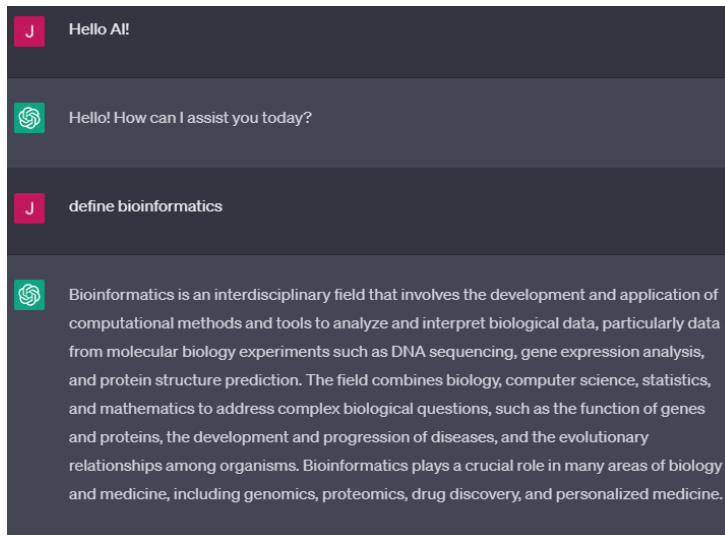
GitHub
Copilot

```
1 import matplotlib.pyplot as plt
2
3 def draw_scatterplot(x_values, y_values):
4     plt.scatter(x_values, y_values, s=20)
5     plt.title("Scatter Plot")
6     plt.xlabel("x values")
7     plt.ylabel("y values")
8     plt.show()
```

<https://github.com/features/copilot>

15/04/2024

ChatGPT (Mar 23 Version)



15/04/2024

Otros ejemplos de inteligencia artificial, nótese que lo primero que destaca es la interdisciplinaridad:

se define como un campo interdisciplinario que implica el desarrollo y la aplicación de métodos y herramientas computacionales para analizar e interpretar datos biológicos, particularmente datos de experimentos de biología molecular como la secuenciación de ADN, el análisis de expresión génica y la predicción de la estructura de proteínas (hasta aquí la definición).

¿Qué es la bioinformática?

The screenshot shows the NIH National Human Genome Research Institute website. The main heading is "Glosario parlante de términos genómicos y genéticos". Below this, it states: "El glosario presenta casi 250 términos explicados de manera fácil de entender por científicos y profesionales destacados del Instituto Nacional de Investigación del Genoma Humano." There is a search bar with the placeholder text "¿Qué le interesa?". On the right side, there is a large graphic of a DNA double helix and a chromosome. Below the main heading, there is a section titled "Popular" with two sub-sections: "Gene Therapy" and "Genetic Code".

NIH National Human Genome Research Institute

ABOUT GENOMICS RESEARCH FUNDING RESEARCH AT NIH/NIH ABOUT HEALTH CAREERS & TRAINING NEWS & EVENTS ABOUT NIH/NIH

Begin your search here

Glosario parlante de términos genómicos y genéticos

El glosario presenta casi 250 términos explicados de manera fácil de entender por científicos y profesionales destacados del Instituto Nacional de Investigación del Genoma Humano.

¿Qué le interesa?

Navegar alfabéticamente | English

- A
- B
- C
- D
- E
- F
- G
- H

Ácido nucleico
Adenina
ADN (Ácido Desoxirribonucleico)
ADN mitocondrial
ADN no codificante
ADN recombinante (rADN)

Popular

Gene Therapy
Genetic Code

15/04/2024

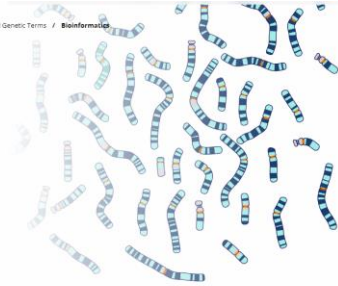
Otro buen lugar donde intentar responder a nuestra pregunta es el Instituto Nacional de Investigación del Genoma Humano, donde podemos encontrar un glosario de términos genómicos y genéticos.

¿Qué es la bioinformática?

[Home](#) / [About Genomics](#) / [Educational Resources](#) / [Talking Glossary of Genomic and Genetic Terms](#) / **Bioinformatics**

BIOINFORMATICS

updated: April 9, 2024



[En Español](#)

Definition



Bioinformatics, as related to genetics and genomics, is a scientific subdiscipline that involves using computer technology to collect, store, analyze and disseminate biological data and information, such as DNA and amino acid sequences or annotations about those sequences. Scientists and clinicians use databases that organize and index such biological information to increase our understanding of health and disease and, in certain cases, as part of medical care.

15/04/2024

Además lo podemos encontrar en castellano y en inglés.

IA, ¿Qué es la bioinformática?

Usted
¿Que es la bioinformática?

Copilot
✓ Generando respuestas...

La **bioinformática** es una disciplina científica que **permite investigar, desarrollar y aplicar herramientas informáticas** para mejorar el manejo de datos biológicos. A través de estas herramientas, se pueden reunir, almacenar, organizar, analizar e interpretar datos relacionados con la biología, como secuencias de **ADN, ARN** o anotaciones sobre esas secuencias^{[1][2][3][4]}. En el contexto de la genética y la genómica, la bioinformática se utiliza para recopilar, almacenar, analizar y diseminar información biológica, lo que incluye secuencias de ADN y aminoácidos. En resumen, la bioinformática es un subconjunto del campo más amplio de la biología computacional, aplicando técnicas analíticas cuantitativas al modelado de sistemas biológicos^{[3][4]}. Si tienes más preguntas o necesitas más detalles, no dudes en preguntar. 🗨️

Más información [1](#) [2](#) [3](#) [4](#) [es.wikipedia.org](#) [genome.gov](#) [es.wikipedia.org](#) +2 más

Usted
What is bioinformatics?

Copilot
✓ Generando respuestas...

Bioinformatics is an interdisciplinary field of science that **develops methods and software tools** for understanding biological data, especially when dealing with large and complex datasets. It combines principles from biology, chemistry, physics, computer science, computer programming, information engineering, mathematics, and statistics to **analyze and interpret biological data**^[1]. The subsequent process of analyzing and interpreting data is referred to as **computational biology**. Here are some key aspects of bioinformatics:

- 1. Sequence Analysis:** Bioinformatics involves analyzing and aligning DNA, RNA, and protein sequences. This helps identify genes, single nucleotide polymorphisms (SNPs), and other variations within genomes^[1].
- 2. Proteomics:** Bioinformatics plays a role in understanding the organizational principles within nucleic acid and protein sequences. It aids in annotating genomes and studying observed mutations^[1].
- 3. Text Mining and Ontologies:** Bioinformatics includes text mining of biological literature and the development of biological and gene ontologies. These tools help organize and query biological data^[1].
- 4. Gene and Protein Expression:** Bioinformatics tools assist in comparing, analyzing, and interpreting genetic and genomic data. They also contribute to understanding evolutionary aspects of molecular biology^[1].
- 5. Structural Biology:** In structural biology, bioinformatics aids in simulating and modeling DNA, RNA, proteins, and biomolecular interactions^[1].

The term "bioinformatics" was first coined in 1970 by Paulien Hogeweg and Ben Hesper to refer to the study of information processes in biotic systems. It parallels biochemistry and has become essential for advancing our understanding of biological processes^{[1][2]}. If you have further questions or need more details, feel free to ask!

Más información [1](#) [2](#) [en.wikipedia.org](#) [genome.gov](#) [britannica.com](#)

🗨️ 📄 📌 📄 📌 2 de 90 respuestas

2024

Respuestas de Copilot en Abril del 24:

Observad que la respuesta es dependiente del idioma, más acertada en su versión inglesa al incorporar en la definición la interdisciplinaridad, la estadística. A pesar de ello vemos como indica que el subsecuente proceso de analizar e interpretar datos se conoce como biología computacional.

IA, ¿Qué es la bioinformática?

Usted

What is computational biology?

Copilot

✓ Generando respuestas...

Computational biology is a fascinating interdisciplinary field that combines **biology**, **computer science**, and **big data**. Its primary goal is to **understand biological systems and relationships** by leveraging data analysis, mathematical modeling, and computational simulations ¹ ². Let's delve into the details:

- History:**
 - The roots of computational biology trace back to the early 1970s when researchers began using computers to analyze and compare large biological datasets.
 - The **Human Genome Project**, which started in 1990, is perhaps the most well-known example of computational biology. By 2003, it had mapped around 85% of the human genome, achieving its initial goals. Work continued, and by 2021, the project reached a "complete genome" level, with only 0.3% remaining bases covered by potential issues ³.
- Applications:**
 - **Anatomy:** Computational anatomy focuses on modeling and simulating biological structures at the visible or gross anatomical scale. It extracts anatomical coordinate systems from dense 3D measurements, such as those obtained through magnetic resonance imaging ⁴.
 - **Data and Modeling:** Computational biology builds models for various types of experimental data (e.g., concentrations, sequences, images) and biological systems (e.g., molecules, cells, tissues, organs). It draws from mathematical and computational fields like complexity theory, algorithmics, machine learning, and robotics ⁵.
- Achievements:**
 - Computational biology has contributed to:
 - Creating accurate models of the **human brain**.
 - Mapping the **3D structure of genomes**.
 - Modeling complex **biological systems** ⁶.

In summary, computational biology plays a pivotal role in unraveling the mysteries of life by integrating computational methods with biological insights. If you have more questions or need further details, feel free to ask! 🤖

Usted

What are the differences between bioinformatics and computational biology?

15/04/2024

Busquemos biología computacional...

IA, ¿Qué es la bioinformática?

Copilot

✓ Generando respuestas...

Certainly! Let's explore the distinctions between **bioinformatics** and **computational biology**:

1. Bioinformatics:

- **Focus:** Bioinformatics primarily deals with **managing biological data** using computational tools and techniques.
- **Data Types:** It involves analyzing biological sequences (such as DNA, RNA, and proteins), annotating genomes, predicting gene functions, and understanding molecular structures.
- **Applications:**
 - **Sequence Analysis:** Bioinformatics aligns and compares sequences to identify genes, regulatory elements, and variations.
 - **Genome Annotation:** It annotates genomes by identifying genes, promoters, and other functional elements.
 - **Protein Structure Prediction:** Bioinformatics predicts protein structures and interactions.
- **Example Task:** If you want to find similarities between two DNA sequences, bioinformatics tools can help.
- **Key Tools:** BLAST, FASTA, and genome browsers.
- **Overall Goal:** Extract meaningful information from biological data.

2. Computational Biology:

2. Computational Biology:


- **Scope:** Computational biology encompasses a broader range of topics beyond data management.
- **Approach:** It combines biology, mathematics, and computer science to model and simulate biological systems.
- **Applications:**
 - **Modeling:** Computational biology builds mathematical models for biological processes.
 - **Systems Biology:** It studies interactions within biological networks (e.g., metabolic pathways, gene regulatory networks).
 - **Biological Simulations:** Computational models simulate biological phenomena.
- **Example Task:** Predicting how a specific gene mutation affects an entire cellular pathway.
- **Key Tools:** Mathematical modeling, simulations, and statistical analysis.
- **Overall Goal:** Understand biological systems through quantitative approaches.


In summary, bioinformatics focuses on data handling and sequence analysis, while computational biology extends to modeling and understanding complex biological processes. Both fields are essential for advancing our knowledge of life sciences! 🧬


15/04/2024

Actualmente, a pesar de ser muy cómodo y proporcionar respuestas aparentemente sólidas, no recomiendo su uso como respuesta final.

ChatGPT (Mar 23 Version)

 write a python script for the game rock, scissors, paper

 Sure! Here is an example Python script for a simple game of rock, scissors, paper. In this script, the user plays against the computer, and the first player to win three rounds wins the game.

```
python  Copy code

import random

choices = ['rock', 'scissors', 'paper']
user_score = 0
computer_score = 0


print("Let's play rock, scissors, paper! First player to win three rounds wi

while user_score < 3 and computer_score < 3:
    # get user's choice
    user_choice = input("Enter 'rock', 'scissors', or 'paper': ")
    while user_choice not in choices:
        user_choice = input("Invalid input. Please enter 'rock', 'scissors',

    # get computer's choice
    computer_choice = random.choice(choices)

    # determine the winner
    if user_choice == computer_choice:
        print("Tie!")
    elif (user_choice == 'rock' and computer_choice == 'scissors') or \
         (user_choice == 'scissors' and computer_choice == 'paper') or \
         (user_choice == 'paper' and computer_choice == 'rock'):
        user_score += 1
        print("You win this round!")
    else:
        computer_score += 1
        print("Computer wins this round!")

    # display the current score
    print(f"Your score: {user_score}\nComputer's score: {computer_score}\n")

# determine the winner of the game
if user_score > computer_score:
    print("Congratulations!")
else:
     Regenerate response
```

También para código



Algunas fechas históricas

Siglo XX

Pioneros

ADN (1953), Watson y Crick

Rosalind Franklin



Secuenciación insulina (1953), Fred Sanger (1901–1994)



Publicación del primer *Atlas of Protein Sequences* (1965), Margaret Dayhoff (1925–1983)



Avances posteriores

Nacimiento del proyecto ARPANET (actual Internet), EE.UU. (1969)



Interpretación código gen. (~1975), Har Gobind Khurana (1922–2011)



Método secuenciación (1977), Fred Sanger (1901–1994)



Siglo XXI

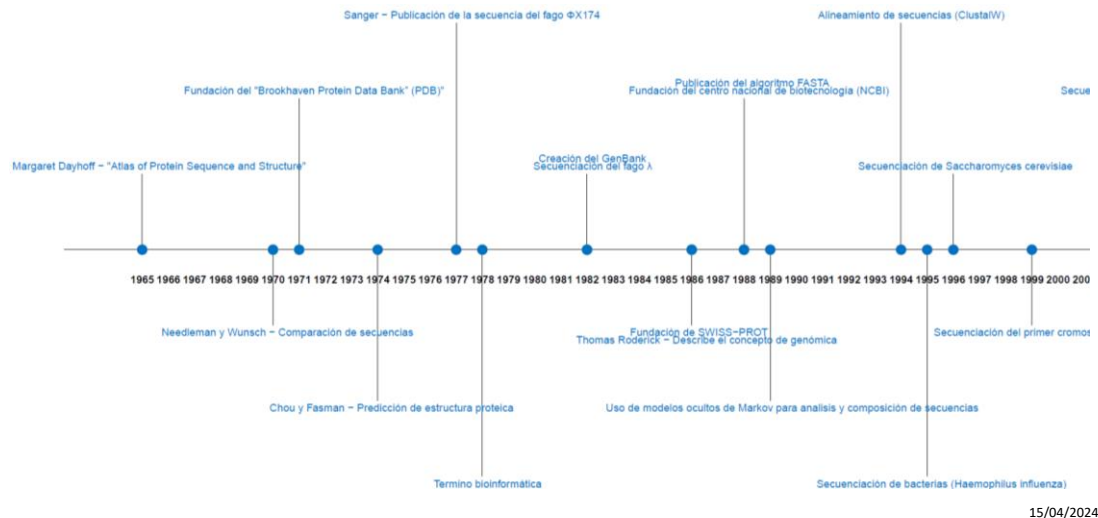
Secuenciación completa del genoma humano y publicación
National Human Genome Research Institute (NHGRI) (2003)



15/04/2024

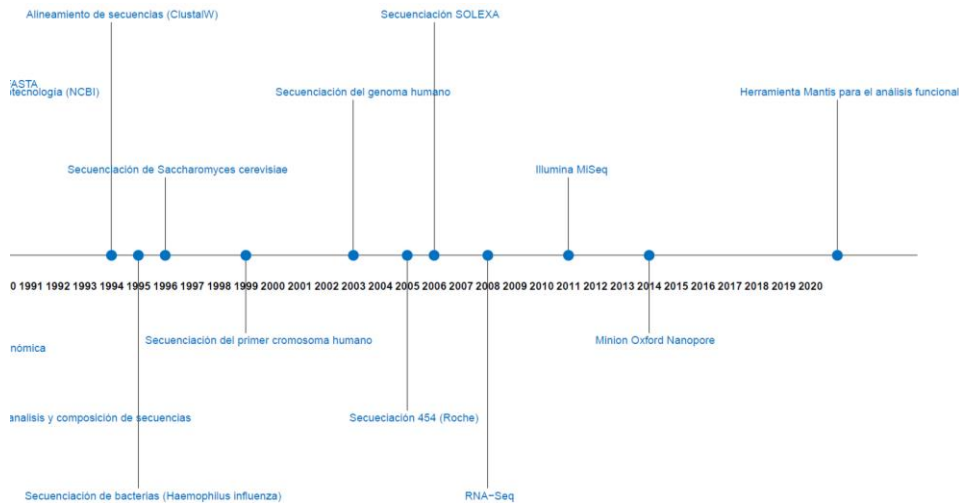
Rosalind Franklin (1958) fue una química y cristalógrafa británica. Sus trabajos con imágenes por difracción de rayos X tuvieron gran importancia en el progreso de campos diversos. Fueron clave para revelar la estructura de los carbones, el grafito, el ARN y varios virus, aunque la mayor trascendencia la tuvo su aporte para entender la estructura del ADN, gracias a la imagen llamada Fotografía 51, que impactó en los avances científicos de la genética. Sus investigaciones sobre la estructura del carbón y de los virus fueron reconocidos en vida. Su muerte prematura impidió que pudiera disfrutar del reconocimiento por su aporte al descubrimiento de la estructura del ADN.

Algunas fechas históricas



El primer gran proyecto de bioinformática fue emprendido por Margaret Dayhoff en 1965, quien desarrolló una primera base de datos de secuencias de proteínas llamada "Atlas of Protein Sequence and Structure" ([Dayhoff 1965](#)). Posteriormente, el primer algoritmo para comparar secuencias de proteínas o ADN fue publicado por Needleman y Wunsch en 1970 ([Needleman and Wunsch 1970](#)) y el primer algoritmo de predicción de la estructura de la proteína fue desarrollado por Chou y Fasman en 1974 ([Chou and Fasman 1974](#)). En 1971 se fundó el "Brookhaven Protein Data Bank" (PDB). El PDB es una base de datos para el almacenamiento de datos cristalográficos de proteínas. El desarrollo de la bioinformática procedió muy lentamente al principio hasta que se publicó la secuencia completa del gen del virus del bacteriófago ϕ X174 en 1977 ([Sanger et al. 1977](#)), a través del método descrito con el mismo nombre, secuenciación Sanger. El primer término que define la bioinformática no se acuñó hasta 1978 ([Hogeweg 1978](#)), con la publicación de un modelo de análisis interactivo. En 1986 se fundó la base de datos SWISS-PROT y Thomas Roderick acuñó el término genómica, que describe la disciplina científica de secuenciación y descripción de genomas completos ([Kuska 1998](#)). Dos años después, se creó el Centro Nacional de Biotecnología (NCBI); siendo a día de hoy una de las bases de datos biológicas más importantes. El mismo año también vio el inicio de la Iniciativa Genoma Humano y la publicación del algoritmo FASTA ([Pearson and Lipman 1988](#)).

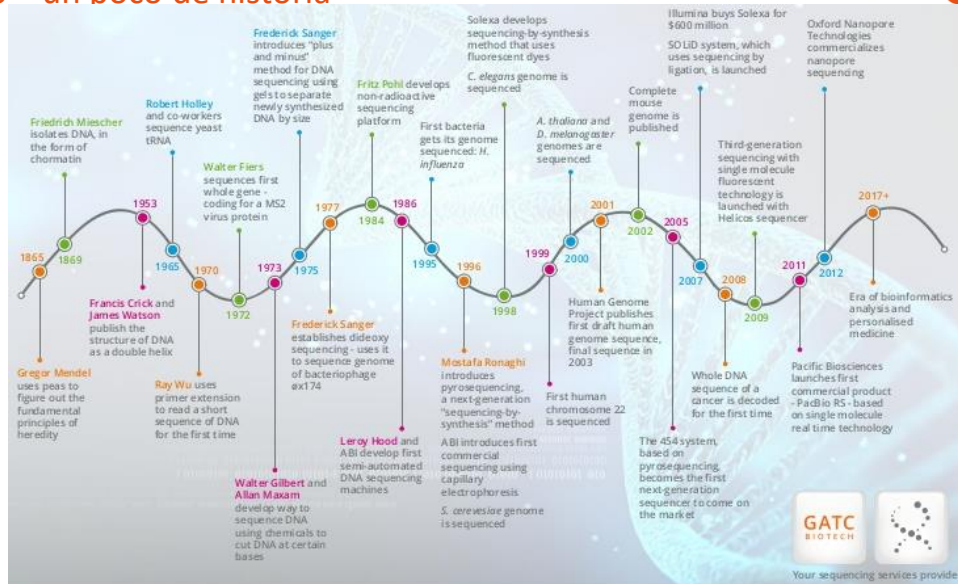
Algunas fechas históricas



15/04/2024

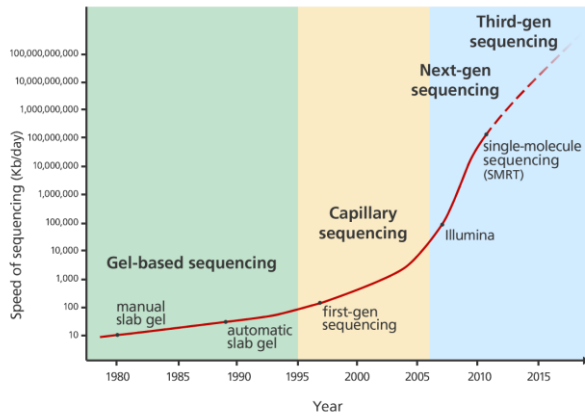
A partir de la secuenciación del genoma humano se originaron muchos otros proyectos relacionados con otros organismos. En el año 2005 se presentó la secuenciación 454 (Roche), la primera técnica de secuenciación de próxima generación (NGS), seguida en 2006 por la secuenciación de Solexa. En el año 2008 se introdujo la técnica RNA-Seq dando lugar a una serie de nuevas disciplinas como la farmacogenética y proteogenómica. Posteriormente se han desarrollado nuevos métodos de secuenciación, como MiniON de Oxford Nanopore. Como hemos visto con los ejemplos anteriores, la bioinformática es una disciplina creciente en términos de evolución metodológica.

NGS – un poco de historia

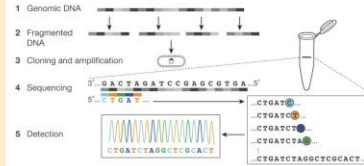


15/04/2024

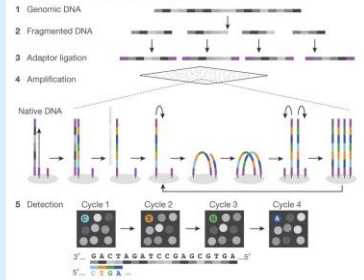
Estamos en la tercera generación de secuenciadores



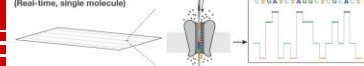
First generation sequencing (Sanger)



Second generation sequencing (massively parallel)

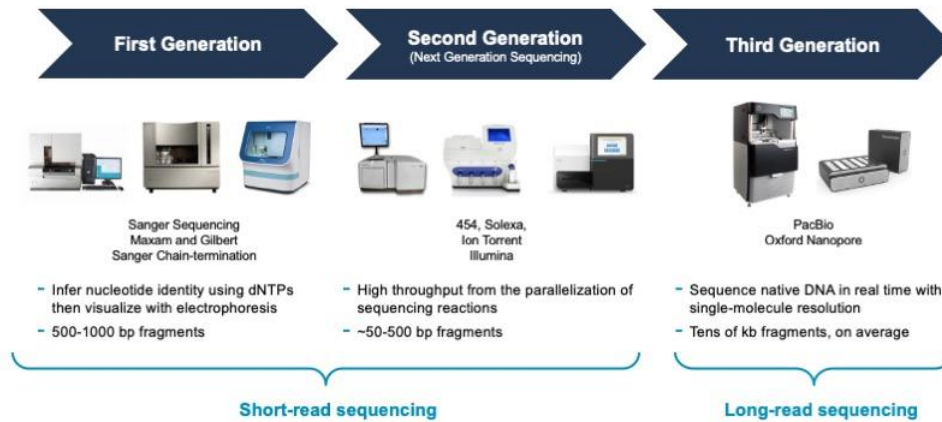


Third generation sequencing (Real-time, single molecule)



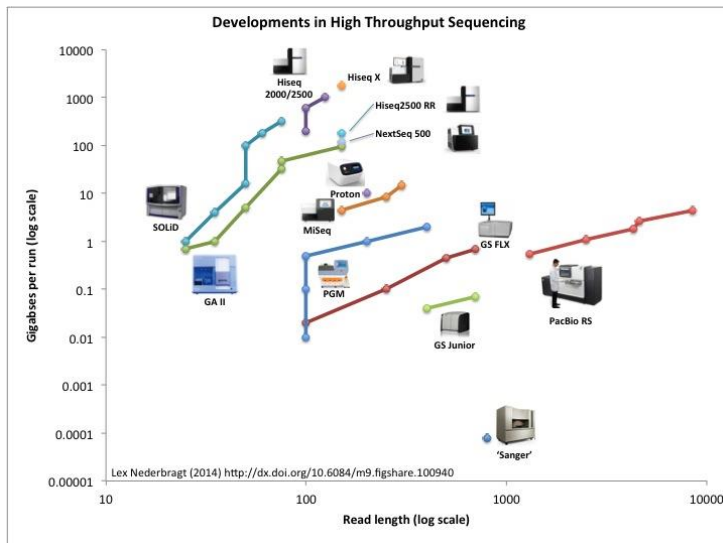
15/04/2024

estamos en la tercera generación de secuenciadores



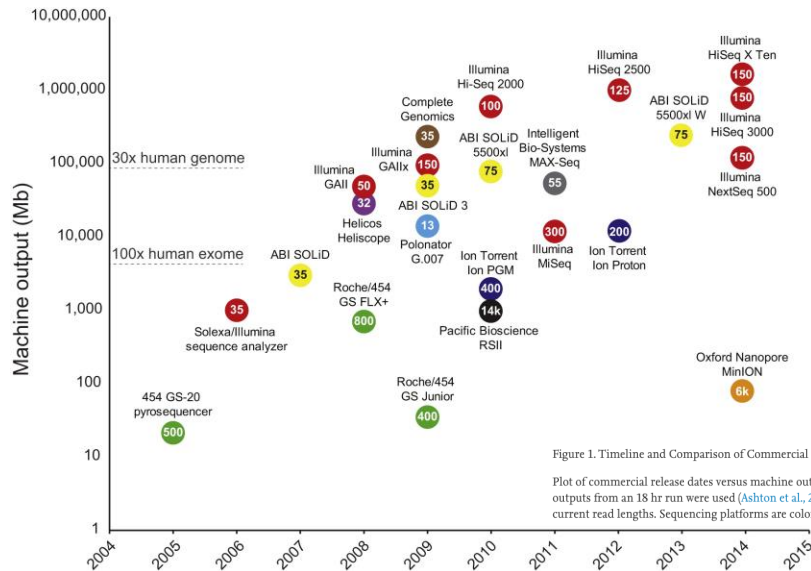
15/04/2024

cantidad / longitud de lectura
(también importa la calidad o fiabilidad)



15/04/2024

Cantidad/tiempo



Reuter et al., 2015

Before:



Hemos pasado de necesitar encargar la secuenciación a compañías externas que se dedicaban específicamente a ello,

After:



HiSeq2000
Illumina



MinION
Oxford Nanopore

15/04/2024

A poder realizarlo “in the house”.

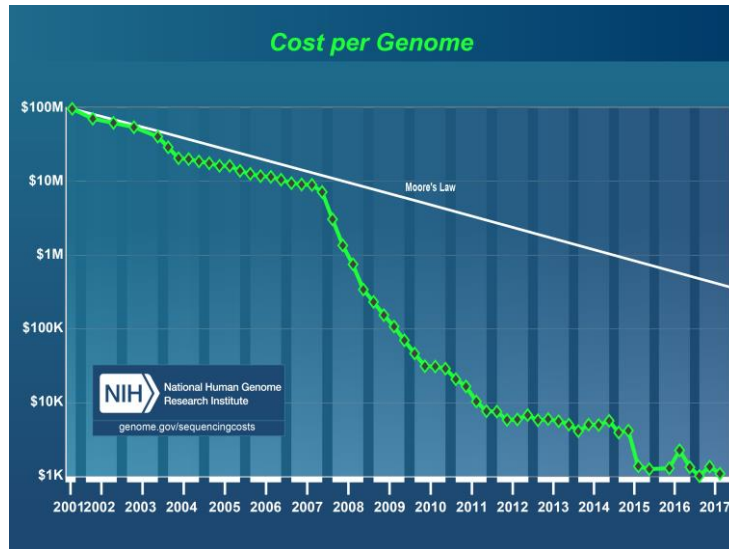
En la imagen podemos ver dos ejemplos de secuenciadores, el de la izquierda es muy similar al de la diapositiva anterior pero de un centro privado, el de la derecha es uno de los más asequibles hoy en día, 1000 € aproximadamente.

secuenciadores



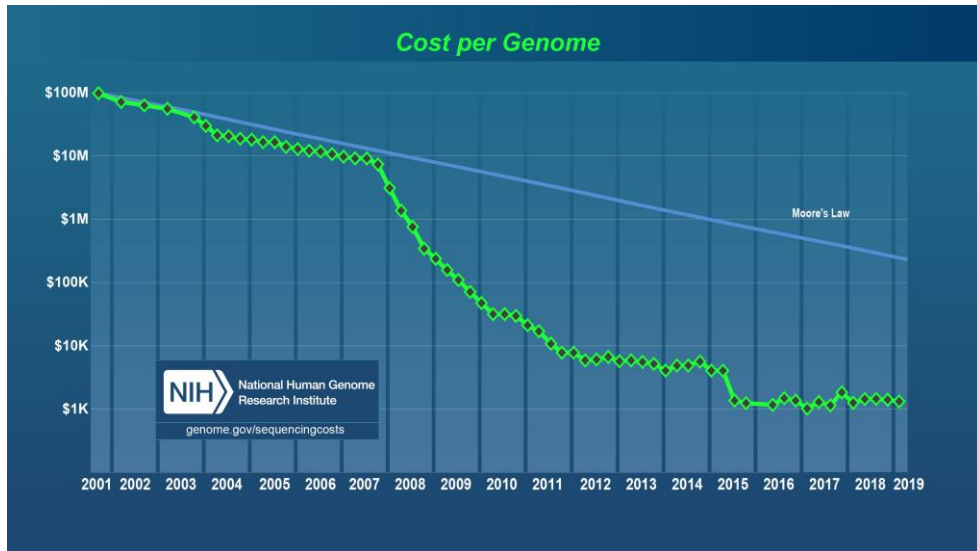
	Illumina (HiSeq 4000)	PacBio (Sequel)	Oxford Nanopore (MinION)
Read length	Up to 150 bp	10-15kb	Up to 900kb
Number of reads	2.5-5 Million	500 K	Up to 1 M
Processing time	<1-3.5 days	Up to 10 hours	~ 6 hours
Error rate	<1%	10-15%	5-15%
Cost per run	~\$3000	~\$850	\$500-\$900
Instrument price	\$900 K	\$350K	\$1K
Advantages	Highly accurate	Sequence long reads	Sequence long reads Portable device

Coste por Genoma

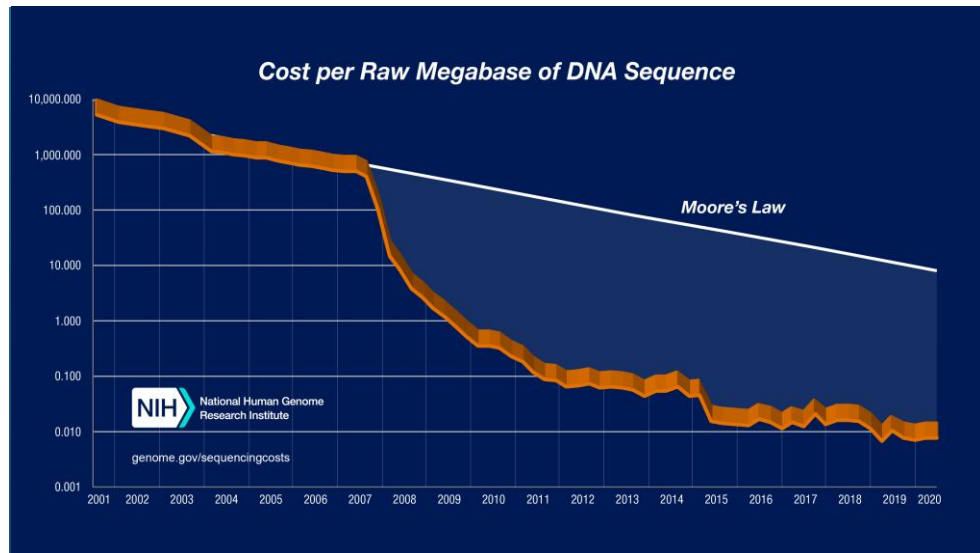


15/04/2024

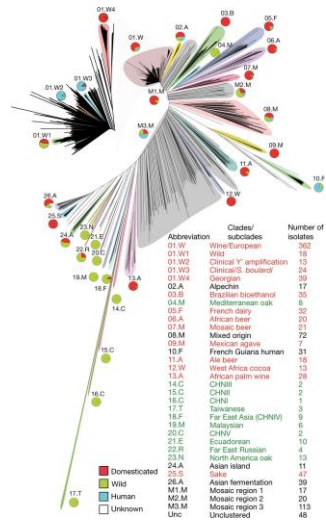
Coste por Genoma



Coste por Genoma

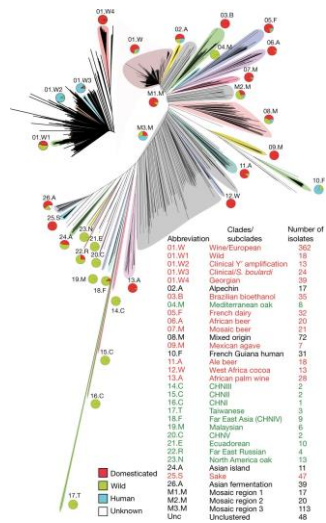


1,011 *Saccharomyces cerevisiae* genomes



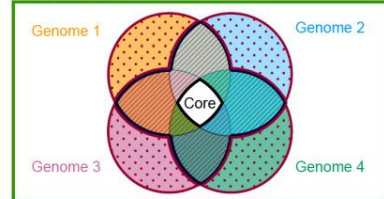
15/04/2024

1,011 *Saccharomyces cerevisiae* genomes



pangenoma

Pangenome



Cloud genome
Shell genome

15/04/2024

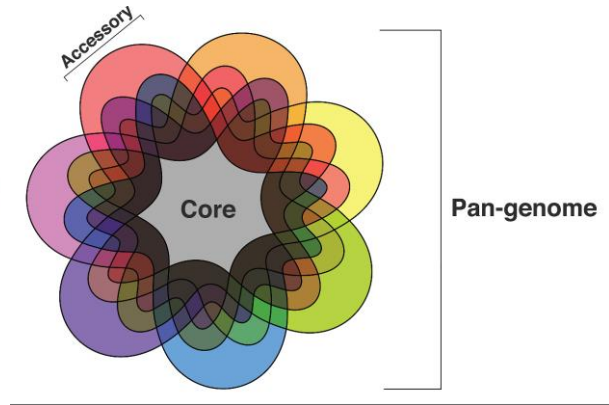
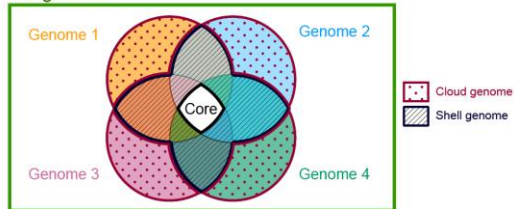


En 2023 se publicó un borrador de referencia del pangenoma humano, basado en 47 genomas de personas de diversas etnias. Se está planeando una referencia mejorada que capture aún más biodiversidad de una muestra aún más amplia.

Liao WW, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, et al. (May 2023). ["A draft human pangenome reference"](#). *Nature*. **617** (7960): 312–324. [Bibcode:2023Natur.617..312L](#). [doi:10.1038/s41586-023-05896-x](#). [PMC 10172123](#). [PMID 37165242](#).

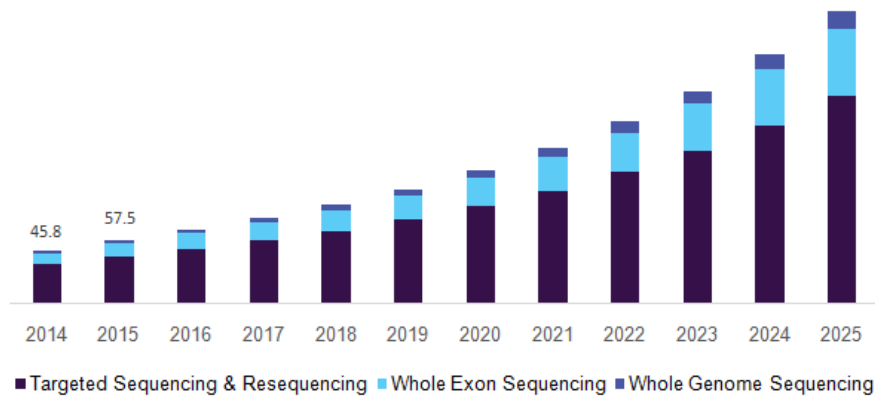
pangenoma

Pangenome



15/04/2024

pronósticos de mercado para la secuenciación



precio al alcance de todos



23andMe Ancestry Service - DNA Test Kit with Personalized Genetic Reports Including Ancestry Composition with 2500+ Geographic Regions, Family Tree, DNA Relative Finder and Trait Reports

Visit the 23andMe Store
4.7 ★★★★★ 33,500 ratings | 990 answered questions
Amazon's Choice in Genetic Tests by 23andMe

Currently unavailable.
We don't know when or if this item will be back in stock.

- **WHAT YOU GET:** At-home DNA test kit with access to 80+ personalized reports. Our innovative ancestry composition estimates your ancestry to the 0.1% across 2500+ geographic regions. Learn when your most recent ancestors from each population lived. Opt in to find DNA relatives and automatically build your family tree. Find out what makes you unique with personalized traits reports.
- **ANCESTRY FEATURES:** Dig deeper into your ancestry with the most comprehensive ancestry breakdown. Go back in time with the Ancestry Timeline to gain a clearer picture of where your ancestors lived and when they lived there. Discover your family origins with the Haplogroup feature. Opt-in to DNA Relative Finder to discover and connect with people who share your DNA. The automatic Family Tree feature makes it easy to see your DNA relationships.
- **TRAIT REPORTS:** Do you have freckles, a taste aversion to cilantro, perfect

Currently unavailable.
We don't know when or if this item will be back in stock.

See Similar Items

Deliver to Spain

Add to List

Have one to sell?
Sell on Amazon

everlywell

Food Sensitivity Test

Everlywell Food Sensitivity Test -...

★★★★★ 6,115

\$199.00 prime

Save 20% with coupon

Sponsored

La bioinformática cada vez más cercana



Los cuatro sencillos pasos para
implantar el Censo Genético Canino en
tu municipio

ADN Canino

Què és el cens caní ?

La Mancomunitat Intermunicipal de l'Horta Sud, en col·laboració amb ADN Canino, ha creat un registre censal caní comarcal per a tindre identificats a tots els gossos de les localitats que la conformen. La identificació dels gossos ha de realitzar-se tant per microxip homologat com per patró d'ADN, que identifica de manera única a cada animal i s'obté a través d'una mostra de saliva totalment indolora.

Per què un cens genètic del teu gos ?

Es una eina fiable per a protegir els animals, ja que el cens permetrà localitzar-los si es perden, són maltractats o abandonats. A més, permet mantindre el teu municipi net i més saludable, ja que amb aquest mètode és possible identificar al propietari/ària de l'animal que ha incomplert la seua obligació de recollida d'ecrecments.

Quins animals han d'estar inscrits ?

En el registre censal caní municipal han d'estar inscrits tots els gossos del terme municipal de la localitat segons la seua ordenança.

Quan he de censar al meu gos ?

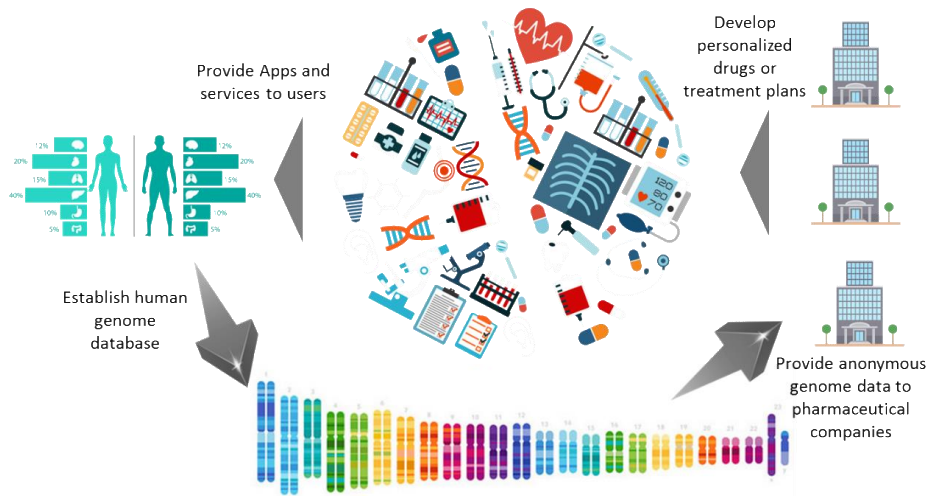
El període d'inscripció inicial és obligatori i subvencionat i comprén de l'1 fins al 31 de desembre de 2022. Una vegada tancat el termini, hi serà de pagament.

Quin cost té la inscripció ?

La Mancomunitat Intermunicipal de l'Horta Sud subvencionarà el 100% durant el període bonificat.

15/04/2024

BIG data is your friend (?)



15/04/2024

Ética: Big Data, la tecnología ni es buena ni es mala

EMAILED ON SEPTEMBER 20, 2018 BY CONOR GRANT

156-year old insurer John Hancock now requires customers to use health wearables

John Hancock, a life insurance company that was founded during the Civil War, **announced** yesterday that it will require all policyholders to record fitness and health data using wearable devices.

A new lease on life

Hancock has offered "interactive policies" since 2015, offering discounts for healthy behaviors such as exercising or buying healthy foods — saving itself money by paying out fewer claims.

To run the new program, Hancock partnered with **Vitality Group**, a platform that already operates widely in South Africa and the UK. Vitality **claims** policyholders using wearables live 13 to 21 years longer than "normally" insured couch potatoes.

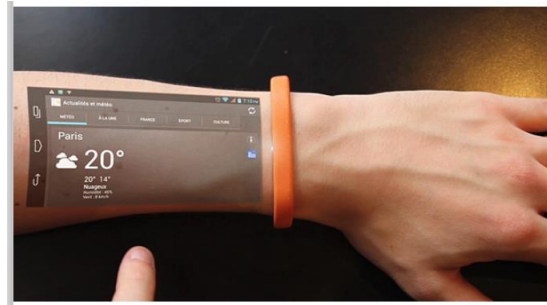
As other insurers roll out wearable discounts, critics warn that wearable incentives could disadvantage some people — like anyone with an existing condition or anyone without an Apple Watch.

But Apple is excited about health — and healthy profit margins

John Hancock offers financing for health wearables — incentivizing companies like Apple and Fitbit to ramp up production to capitalize on insurance's data-driven future.

Last week, Apple **rolled out** a new Watch with an EKG designed to offer new health-monitoring options — and just yesterday morning, Fitbit **launched** a program called Fitbit Care that provides medical coaching that integrates with doctors and insurers.

So now more than ever, the wearables industry is really one to Watch: Forecasters **expect** the market for wearables to increase to \$25B by next year, thanks in part to increased demand driven by insurers.



15/04/2024

Implicaciones bioéticas

- Actualmente se genera un gran volumen de datos
- ¿El paciente está totalmente informado?
 - Consentimiento informado → Comité de bioética
 - Confidencialidad
- Hackeo de las bases de datos

Empresas dedicadas al análisis de datos filogenéticos

Podrían requerirse los datos por el país, según las leyes

Selección de personas en función de los resultados



15/04/2024

Otro de los grandes problemas derivados del análisis y gestión de grandes bases de datos de información biológica concierne a la bioética aplicada al análisis de datos.

En la era del Big Data, el gran volumen de datos disponibles desafían las normas sociales y éticas. Se debe prestar atención a los problemas conocidos con el análisis de Big Data, que cubren temas como el consentimiento informado, la privacidad, la confidencialidad, la diversidad, la propiedad de los datos, las brechas digitales y los derechos colectivos (Mittelstadt y Floridi 2016; Mittelstadt 2017; Taylor, Floridi y van der Sloot 2017). El consentimiento informado es un documento básico revisado cuidadosamente por comités de revisión ética (ERC) que el paciente debe firmar, donde se tienen en cuenta todos los puntos que se van a realizar con, por ejemplo, la muestra que extraigamos del mismo, a la vez que se firma una cláusula de confidencialidad mediante la cual sea difícil identificar al paciente en las posteriores fases del análisis.

Uno de los principales problemas del almacenaje de esta información en bases de datos es la posible filtración de esta información. Si el valor de un número de teléfono, dirección o una cuenta de correo es muy alto para determinadas empresas, la información genética puede llegar a tener mucha más relevancia.

Este punto está relacionado con el sistema de negocio de muchas empresas que dan información sobre la filogenia de nuestros genes a partir de una muestra de saliva o hisopo de mucosa. Más allá de la fiabilidad de los resultados, la información queda almacenada en sus bases de datos que podrían en última instancia ser requeridas legalmente por el país donde se ubique dicha empresa, como puede pasar con los datos en redes sociales, por ejemplo, quedando esta información expuesta.

Otro ejemplo de problema bioético es la implicación de nuestra información genética a la hora de contratar seguros privados. Existen marcadores genéticos que predisponen a distintos tipos de afecciones y que pueden encarecer o denegar la contratación de un seguro privado.

Lectura: Big Data, Biomedical Research, and Ethics. Review: New Challenges for IRBs. Agata Ferretti, Marcello Ienca, Samia Hurst, and Effy Vayena

El ataque cibernético a 23andMe expuso los datos personales de judíos asquenazíes

Al menos un millón de puntos de información de cuentas de 23andMe parecen haberse revelado en BreachForums. Aunque se desconoce la magnitud del ataque, la compañía afirma que está trabajando para confirmar la autenticidad de los datos.



La filial de la firma 23andMe anuncia con un anuncio en su página web la filtración de "datos genéticos". BREXID PAUL WEBER/GETTY IMAGES

La compañía de pruebas genéticas 23andMe confirmó el viernes pasado que los datos de un subconjunto de sus usuarios se vieron comprometidos. La empresa declaró que sus sistemas no fueron vulnerados y que los atacantes recopilaban la información advirtiendo las credenciales de inicio de sesión de un determinado grupo de clientes y, posteriormente, extrajeron los datos de más personas de una función conocida como **DNA Relatives** (Parientes de ADN), a través de la que eligen compartir su información para que otros la vean.

Los hackers publicaron una muestra inicial de la información en la plataforma BreachForums, a principios de esta semana, asegurando que contenía un millón de puntos de información exclusivamente sobre judíos asquenazíes. También parece haber cientos de miles de usuarios de ascendencia china afectados por la filtración. El

15/04/2024

Bioinformática y Biología computacional

Ciencia interdisciplinar

- Biología
 - Secuenciación de genomas
- Química y bioquímica
 - Interacción entre proteínas y metabolitos
- Medicina
 - Aplicaciones biotecnológicas, como test diagnósticos

Bioinformática ≠ Biología computacional

- Biología computacional - modelaje y simulaciones de sistemas biológicos
 - Dinámica de poblaciones
 - Teoría de juegos en estudios de comportamiento
 - Predicción.
- Bioinformática - Herramientas y bases de datos para la gestión de datos biológicos
 - Alineamiento de secuencias
 - Estudios filogenéticos
 - Ensamblajes



- Desarrollo de fármacos
- Análisis forense
- Desarrollo en la agricultura

15/04/2024

A lo largo de la historia hemos visto que la bioinformática es una ciencia interdisciplinar, que abarca ramas como la biología, química o la medicina, siguiendo un camino que va desde el análisis de los genomas de interés hasta el desarrollo de test diagnósticos o vacunas, siendo clave por tanto en aplicaciones biotecnológicas.

Como nota aclaratoria, cuando hablamos de bioinformática hay que tener en cuenta que comúnmente nos encontraremos con el término biología computacional, como hemos visto en la primera definición. Ambos conceptos se pueden usar de forma indistinta o no en función del profesional al que le preguntas. Generalmente, el **término biología computacional**, hoy en día, **hace referencia a los procesos matemáticos sobre el modelaje o la simulación de sistemas biológicos con el objetivo de desarrollar futuras herramientas bioinformáticas. Usa modelos como dinámica de poblaciones o teoría de juegos para entender cómo funcionan los entornos y entender los sistemas.**

Por su parte, como hemos visto previamente, la bioinformática son el conjunto de herramientas y bases de datos para analizar y gestionar todo el conocimiento biológico. Dentro de estas herramientas encontramos alineadores de secuencias, ensambladores, sistemas de agregación (clustering) y otros muchos métodos que nos permitirán entender en su totalidad un sistema biológico. Por otra parte, toda la

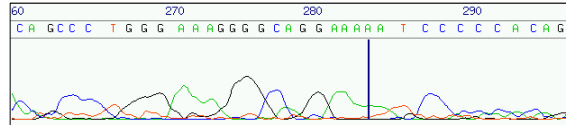
información generada se almacena en bases de datos de acceso público, a través de las cuales podemos realizar distintos análisis comparativos de nuestra información respecto al conocimiento generado previamente.

La bioinformática tiene infinidad de aplicaciones dentro de multitud de campos, por ejemplo en el diseño de fármacos y vacunas, el análisis forense de ADN, estudiar en mayor profundidad los problemas asociados a la rápida evolución de los genes de resistencia a antibióticos (ARGs) en bacterias o el desarrollo de medicina personalizada. Fuera del mundo de la salud también podemos encontrarle aplicabilidad en la agricultura, a través del desarrollo de nuevas variedades de cultivos que tengan una mayor productividad y más resistencia a las enfermedades.

Limitaciones y problemas

Limitaciones y problemas

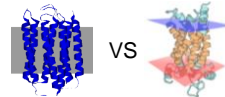
- **Calidad** de la información en las bases de datos
 - Secuencia de ADN con errores → Alineamientos de baja calidad



- Dependencia de un **gran número de información** disponible
 - No siempre tenemos toda la información disponible

FEATURES	Location/Qualifiers
source	1..4503
	/organism="Pseudomonas putida LF54"
	/mol_type="genomic DNA"
	/submitter_seqid="seq9_NODE_9"
	/strain="Pseudomonas"
	/isolation_source="lotus field"
	/db_xref="taxon:1293312"
	/country="Japan: Ibaraki Tsukuba"
	/collection_date="20-Jul-2009"
	/breeds="putida"

- Usar **más de un mismo método/programa** para una misma tarea



15/04/2024

Sin embargo, la bioinformática también presenta a día de hoy una serie de limitaciones inherentes que hay que tener en cuenta desde el primer momento. La calidad de las predicciones bioinformáticas depende de la calidad de los datos almacenados en las bases de datos y la eficiencia de los algoritmos que se utilizan, es decir, las secuencias de los genomas secuenciados a menudo contienen errores, dando lugar a anotaciones incorrectas que generan resultados incorrectos en análisis estructurales o filogenéticos.

Otro de los problemas es la dependencia de un gran número de información relativa a cada secuencia que se almacena en las bases de datos. Comúnmente al añadir información a las bases de datos esta es incompleta y nos olvidamos de datos tan básicos como la fecha de extracción de la muestra o la procedencia.

Además, siempre es positivo utilizar varios programas para una misma tarea, si están disponibles, y realizar varias evaluaciones. A menudo se puede obtener una predicción más precisa si se llega a un consenso comparando los resultados de diferentes algoritmos.

¡Gracias!



Universidad
Internacional
de Valencia

universidadviu.com

De:
🌐 Planeta Formación y Universidades