

Software Libre en Bioinformática: Abriendo las Puertas a la Ciencia Abierta

A.Carolina Ledezma-Carrizalez

April 21, 2025



Software Libre en Bioinformática: Abriendo las Puertas a la Ciencia Abierta

Objetivo: Ofrecer una visión más amplia del papel del software libre en la bioinformática, destacando su importancia para la investigación colaborativa, la transparencia y el avance de la ciencia.

El software libre es un pilar esencial de la bioinformática moderna y un motor para la ciencia reproducible y colaborativa.



Ante de comenzar es necesario pensar en estos conceptos:

- Ciencia Abierta
- Software Libre
- Bioinformática

Conceptos - Ciencia Abierta - Venezuela

En las Recomendación de la UNESCO (sept.2023) "La ciencia abierta es un conjunto de principios y prácticas que tienen como objetivo hacer que la investigación científica desde todos sea accesible a todos para los beneficios de los científicos y de la sociedad en su conjunto. La ciencia abierta consiste en asegurarse no sólo de que el conocimiento científico sea accesible, sino también de que la producción de ese conocimiento en sí sea inclusiva, equitativa y sostenible."

Conceptos - Ciencia Abierta - Venezuela

"La Ciencia Abierta se define como un constructo inclusivo que combina diversos movimientos y prácticas con el fin de que los conocimientos científicos multilingües estén abiertamente disponibles y sean accesibles para todos, así como reutilizables por todos, se incrementen las colaboraciones científicas y el intercambio de información en beneficio de la ciencia y la sociedad, y se abran los procesos de creación, evaluación y comunicación de los conocimientos científicos a los agentes sociales más allá de la comunidad científica tradicional."

Logros, Iniciativas y Referentes - Ciencia Abierta - Venezuela

- Identificación y sistematización de los principales y referentes de CA en el país.
 - Libro “Ciencia Abierta en Venezuela”
 - Alianza Científico Campesina
 - Fundación INFOCENTRO
 - Sistema Nacional de Orquesta Juveniles e Infantiles de Venezuela.
- Las comunidades de software libre.
- Comunalización de la Ciencia
- recientes experiencias de CA:**
- Universidad Nacional de las Ciencias "Humberto Fernández Moran" teniendo base la CA (postulaciones)
 - CEBISA Centro de Formación en Biotecnología de Semillas

1



Conceptos - Software Libre

El Software Libre y de Código Abierto (SLCA), FOSS (Free and Open Source Software) o FLOSS (Free/Libre Open Source Software)- código fuente está accesible.

¿Qué es el Software Libre? y ¿Por Qué Importa?

Las 4 Libertades esenciales

- **La Libertad 0:** La libertad de ejecutar el programa como se deseé, con cualquier propósito.
- **La Libertad 1:** La libertad de estudiar cómo funciona el programa y modificarlo para que haga lo que deseas.
- **La Libertad 2:** La libertad de redistribuir copias para poder ayudar a tu vecino.
- **La Libertad 3:** La libertad de distribuir copias de tus versiones modificadas a terceros.

FOSS Free and Open Source Software ¿Por qué FOSS en Bioinformática?

Énfasis en la libertad, no solo el costo cero.

- Accesibilidad: Elimina barreras económicas para investigadores, estudiantes e instituciones (especialmente en países en desarrollo).
- Transparencia y Reproducibilidad: ¡Crucial para la ciencia! El código abierto permite verificar algoritmos, entender métodos y replicar resultados.
- Colaboración: Fomenta el desarrollo comunitario, la revisión por pares del código y la construcción sobre trabajos previos.
- Personalización: Permite adaptar herramientas a necesidades específicas de investigación.
- Innovación: Acelera el desarrollo de nuevos métodos y algoritmos.

Canaima

<https://canaima.softwarelibre.gob.ve/>



El software libre es un habilitador clave de la Ciencia Abierta

Valores de Ciencia Abierta

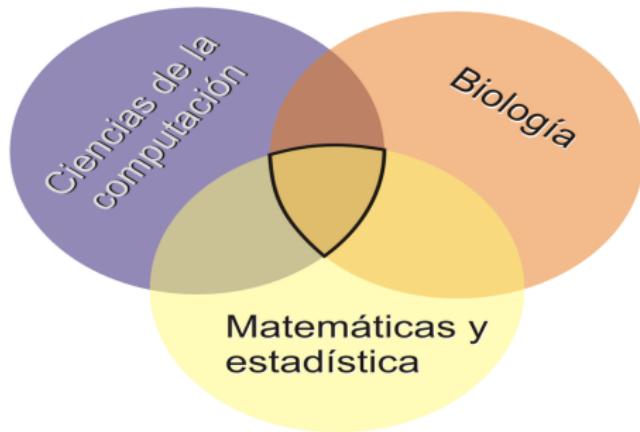
- Calidad e integridad
- Beneficio colectivo
- Equidad y equidad
- Diversidad e inclusión

Principios rectores

- Transparencia, escrutinio, crítica y reproducibilidad
- Igualdad de oportunidades
- Responsabilidad, respeto y rendición de cuentas
- Colaboración, participación e inclusión
- Flexibilidad
- Sostenibilidad

Conceptos - Bioinformática

Intersección entre biología, ciencias de la computación, matemáticas y estadística para analizar e interpretar datos biológicos.



Explosion de Datos

La tecnología abarató la secuenciación y generó una avalancha de datos ómicos. Pero fue el software libre el que nos dio las herramientas gratis y abiertas para poder analizar toda esa información.

Sin ese software, solo unos pocos hubieran podido estudiar los datos, y el avance científico habría sido mucho más lento y cerrado. La explosión de datos y el software libre van totalmente de la mano. El Papel del Software: Es la lente a través de la cual vemos y entendemos los datos biológicos complejos.

Datos Ómicos- Las Ómicas

Es un grupo de disciplinas en biología como ...
Estas disciplinas tienen como objetivo la identificación, caracterización y cuantificación de conjuntos de biomoléculas y procesos moleculares que originan la estructura, función y dinámicas de células, tejidos y organismos

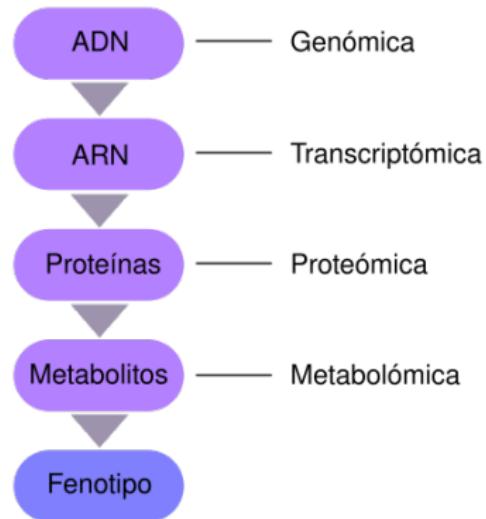


Figure: Multiómica.

El Ecosistema FOSS en Bioinformática - Un Vistazo General

Diversidad de Herramientas: Cubren todo el espectro del análisis bioinformático. Categorías Principales :

- Análisis de Secuencias (BLAST, Bowtie, STAR)
- Ensamblaje de Genomas (SPAdes, Canu)
- Análisis Filogenético (PhyML, RAxML)
- Análisis Estructural (PyMOL - versión libre disponible, VMD)
- Estadística y Visualización (¡Aquí entra R!)
- Gestión de Flujos de Trabajo (Nextflow, Snakemake)

R - El Corazón Estadístico y Gráfico de la Bioinformática



¿Qué es R? Lenguaje y entorno para computación estadística y gráficos. Es una herramienta fundamental en el mundo de la ciencia de datos, la estadística, la investigación académica y la industria para el análisis, la manipulación, la visualización y el reporte de datos. Es potente, flexible y, al ser de código abierto, accesible para todos. Es software libre y de código abierto.

Capacidades Gráficas: Potentes librerías como ggplot2 para visualizaciones de alta calidad.

Comunidad Enorme: Gran soporte, tutoriales, foros.

<https://github.com/z3tt>



Figure: Cédric Scherer

Reproducibilidad:
Facilita
la creación de
informes y análisis
reproducibles (R
Markdown, Quarto).

dplyr::case_when()

IF ELSE...
(but you love it?)

df %>% ADD COLUMN "danger"
mutate(danger = case_when(type == "kraken" ~ "extreme!",
TRUE ~ "high"))
OTHERWISE, danger is high.

IF type is kraken THEN danger is extreme!

danger is extreme!

OTHERWISE, danger is high.

type	age	danger
kraken	baby	extreme!
dragon	adult	high
cyclops	teen	high
kraken	adult	extreme!
dragon	teen	high

©Allison Horst

Figure: Allison Horst <https://github.com/allisonhorst>

¿Por qué es tan dominante en Bioinformática?

Bioconductor- R

¡La clave! Es un proyecto de código abierto (open source) y desarrollo abierto que se dedica a crear y compartir herramientas de software para el análisis de datos biológicos de alto rendimiento, especialmente en áreas como la genómica y la biología molecular. Está basado principalmente en R. Ejemplos de paquetes Bioconductor: DESeq2/edgeR (expresión diferencial), limma, GenomicRanges.



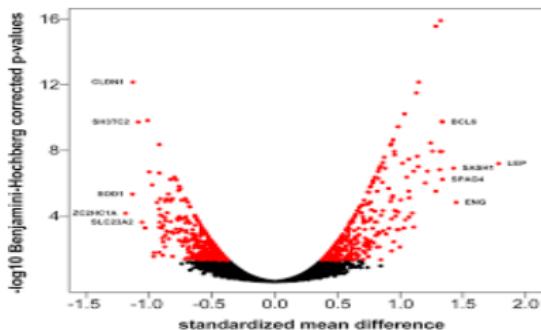
Figure: Logo Bioconductor.

Más Allá de R - Otras Herramientas FOSS Imprescindibles

- Python: Otro gigante en bioinformática. Librerías de Python
 - Biopython (manipulación de secuencias y formatos)
 - NumPy/SciPy (cálculo científico)
 - Pandas (manipulación de datos)
 - Matplotlib Seaborn (visualización)
 - Scikit-learn (Machine Learning)
- Herramientas de Línea de Comandos (CLI): Fundamentales para procesar grandes volúmenes de datos y automatizar tareas.
 - SAMtools / BCFtools: Manipulación de alineamientos y variantes genéticas. Indispensables.
 - BEDTools: Herramientas para trabajar con intervalos genómicos.

- BLAST (NCBI National Center for Biotechnology Information): Aunque no es FOSS en el sentido estricto de licencia para modificar todo, la herramienta es libremente utilizable y distribuible, y es un estándar de facto.
- GATK Genome Analysis Toolkit(parcialmente): Históricamente complejo en licencias, pero partes clave y versiones anteriores son accesibles y ampliamente usadas.
- Gestores de Flujos de Trabajo (Workflows): Nextflow / Snakemake: Esenciales para crear pipelines de análisis complejos, reproducibles y escalables. Utilizan herramientas FOSS subyacentes.

por favor recuerda que vale la pena ser libre, aunque es
comodo pulsar next



Muchas gracias por su atencion!