

Germán A. González  
Lionel Uran Landaburu  
Nicolás Palopoli  
(editores)

wida.exe

DESAFÍOS Y AVENTURAS  
DE LA BIOINFORMÁTICA



**GERMÁN A. GONZÁLEZ,  
LIONEL URAN LANDABURU  
Y NICOLÁS PALOPOLI  
(EDITORES)**

**VIDA.EXE**

*Desafíos y aventuras de la bioinformática*

MARTÍN BANCHERO, JUAN PABLO BUSTAMANTE,  
ESTEBAN LANZAROTTI, R. GONZALO PARRA, MARÍA  
VICTORIA REVUELTA, NICOLÁS STOCCHI, ELIN TEPPA Y  
DIEGO JAVIER ZEA



¿De que modo la computación y la informática ampliaron y modificaron los conocimientos sobre las ciencias de la vida? ¿La bioinformática representa realmente una ruptura de paradigmas o es más de lo mismo con otro formato? ¿La matemática puede ayudarnos a entender cómo se propaga una epidemia y qué medidas se pueden tomar para detenerla? ¿El Proyecto Genoma Humano fue el fin de algo o apenas el principio? ¿El ser humano puede crear vida artificial? ¿Es posible diseñarla a través de una computadora? ¿Seremos eternos e inmortales con cuerpos artificiales y mentes digitales?

En *Vida.exe*, un grupo de jóvenes científicas y científicos cuentan cómo la revolución tecnológica impactó en la biología, explican el pasado de ese extraño centauro que es la bioinformática y lo usan de excusa para dar cuenta de cómo se trabaja actualmente desde la ciencia para intentar descifrar los enigmas de la vida.

“Cuando contamos con suficiente información, no solo somos capaces de entender bastante sobre el funcionamiento de las moléculas, las células, los organismos y sus comunidades, sino también de predecirlo y replicarlo. Esto abre la puerta a prácticas que hace veinte años parecían de ciencia ficción, pero hoy son muy reales,

como la biología sintética, la edición de genomas y la medicina personalizada. En un futuro muy próximo, la forma en que entenderemos la biología molecular, la agricultura y gran parte de la medicina estará inexorablemente atravesada por la bioinformática.”

Entender cómo funcionan los seres vivos no solo es un fin en sí mismo, sino también una herramienta para mejorar nuestra vida y la de todas las especies que habitan el planeta.

## Sobre los autores

GERMÁN A. GONZÁLEZ es licenciado en Bioinformática por la Universidad Nacional de Entre Ríos (UNER) y maestrando en Estadística Aplicada en la Universidad Nacional de Córdoba. Se desempeña como profesional adjunto del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Se dedica al análisis de datos biológicos en diferentes escalas, desde lo molecular a lo ecológico.

LIONEL URAN LANDABURU es licenciado en Biotecnología y actualmente realiza su doctorado en el Laboratorio de Genómica y Bioinformática de Trypanosomátidos de la Universidad Nacional de San Martín, en el desarrollo de software para asistir al descubrimiento y la optimización de drogas para el tratamiento de distintas enfermedades infecciosas.

NICOLÁS PALOPOLI es licenciado en Biotecnología y doctor en Ciencias Básicas y Aplicadas por la Universidad Nacional de Quilmes (UNQ). Es docente universitario e investigador

adjunto del CONICET. Se dedica al estudio de estructuras, interacciones y evolución de proteínas.

MARTÍN BANCHERO es licenciado en Ciencias Biológicas por la Universidad de Buenos Aires (UBA) y máster en Bioinformática y Biología de Sistemas por la Uva-vu Amsterdam. Actualmente realiza su doctorado en el campo de la inmunología sistémica.

JUAN PABLO BUSTAMANTE es licenciado en Bioinformática por la UNER y doctor de la UBA en el área de Química Biológica. Trabajó dos años en la empresa uBiome. Es profesor de la UNER y de la Universidad Austral e investigador asistente del CONICET. Lidera un equipo de investigación y desarrollo en el análisis de microbiomas humanos.

ESTEBAN LANZAROTTI es licenciado en Ciencias de la Computación y doctor en Química Biológica por la UBA. Se dedica al estudio de las interacciones entre proteínas. Es jefe de Trabajos Prácticos en la UBA.

R. GONZALO PARRA es licenciado en Bioinformática por la UNER y doctor en Química Biológica por la UBA. Actualmente, es investigador posdoctoral en el Laboratorio Europeo de Biología Molecular (Heidelberg, Alemania). Su experiencia incluye varias ramas de la bioinformática, como la regulación transcripcional, el plegado de proteínas y la transcriptómica de célula única.

MARÍA VICTORIA REVUELTA es licenciada y doctora en Ciencias Biológicas por la Universidad Nacional de Mar del Plata (UNMDP). Es investigadora posdoctoral en Weill Cornell Medicine (Nueva York) y actualmente trabaja en el análisis de datos de metabolómica, transcriptómica y epigenómica en tumores hematológicos.

NICOLÁS STOCCHI es licenciado en Ciencias Biológicas y doctorando en Biología por la UNMDP. Actualmente, es becario doctoral del CONICET. Se dedica al desarrollo de software para la anotación de superfamilias de proteínas y proteomas completos.

ELIN TEPPA es licenciada en Biotecnología y doctora en Ciencias Básicas y Aplicadas por la UNQ, y se convirtió en bioinformática durante el doctorado en el Instituto Leloir. Es investigadora del CONICET. Se dedica a estudiar evolución molecular y cómo los virus se hacen resistentes a las drogas.

DIEGO JAVIER ZEA es licenciado en Biotecnología y doctor en Ciencias Básicas y Aplicadas por la UNQ. Actualmente, es investigador posdoctoral en la Universidad de la Sorbona, Francia. Se dedica al estudio de la evolución y la estructura de proteínas.

# Índice

Cubierta

Portada

Sobre este libro

Sobre los autores

Agradecimientos

Prólogo

I. Sopa de letras, por Germán A. González

II. Guía para escalar una montaña de datos, por Germán A. González

III. El árbol de la vida, por Germán A. González, Nicolás Stocchi y María Victoria Revuelta

IV. La vida secreta de las moléculas, por Juan Pablo Bustamante, R. Gonzalo Parra y Nicolás Palopoli

V. El mundo es un pañuelo, por Diego Javier Zea

VI. Sobre gallinas y engranajes, por R. Gonzalo Parra, Juan Pablo Bustamante y Nicolás Palopoli

VII. ¿Big bang? No, Bio bang, por Esteban Lanzarotti, Lionel Uran Landaburu y Martín Banchero

VIII. Diseñado para vos, por Germán A. González

IX. El genoma después del genoma, por Germán A. González y R. Gonzalo Parra

X. El poder de las masas, por Elin Teppa

XI. El rey desnudo y sin corona (virus), por Germán A.

González

Epílogo

Créditos

## *Agradecimientos*

QUEREMOS agradecer especialmente a RSG-Argentina, capítulo local de la asociación internacional de estudiantes en bioinformática (*International Society for Computational Biology Student Council*, ISCB-SC), de la cual formamos parte y desde donde surgió el proyecto de este libro.

Y a quienes participaron del primer boceto de este trabajo: Estefanía Mancini, Santiago Carmona, Alexander Monzón e Inti Pagnuco. Sin ellos, no existiría esta versión.

Aviso al lector: ante cualquier problema, mantenga la calma y presione Ctrl+Alt+Supr. Si eso no funciona, pruebe cerrar el libro y volver a abrirlo.

## *Prólogo*

QUIENES escribimos este libro (todos modelo ochenta y pico, joya, nunca taxi) pertenecemos a la generación *millennial*, esa que, en general, no nació con una computadora bajo el brazo, sino que aprendió a usarla en la niñez o en la adolescencia. Conocemos cómo era la vida cuando la música se escuchaba en un *walkman* y para hacer planes con un amigo tenías que llamarlo al fijo, pero también nos adaptamos con facilidad al mundo digital e hiperconectado de las últimas décadas. Es muy probable que encontremos diferencias en la forma en que vivimos la revolución tecnológica si consultamos a una persona de la generación de nuestros padres (que llegaron más tarde al mundo digital) o de la que nos sigue (los nativos digitales). Sin embargo, todos estaremos de acuerdo en que la tecnología ha transformado de manera radical la forma en que vivimos y pensamos. Las ciencias en general, y las ciencias de la vida en particular, no han sido la excepción.

Para comprender la relación entre la biología y la tecnología, debemos remontarnos a un tiempo atrás. Si nos

aventuramos a pensar en los orígenes de la biología (cuando todavía no se llamaba biología), nos vamos a encontrar con que los primeros acercamientos respondieron a la necesidad de conservar los grandes volúmenes de frutas y granos generados en el inicio de la agricultura, hace más de diez mil años. Más adelante, se desarrollaron procesos que utilizamos hasta hoy, como la fermentación para fabricar cerveza o pan. Bastante más tarde (alrededor del 2000 a. C.) comienza a tomar forma en diferentes puntos geográficos (Mesopotamia, China y Egipto) otra de las disciplinas de la ciencia de la vida: la medicina. En ese momento, estaba lejos aún de su forma actual y era una especie de mezcla entre la magia y una cierta ciencia racional. El cambio de paradigma llegaría a fines del Renacimiento europeo y comienzos de la Era Moderna (entre los siglos XVII y XVIII), cuando la biología se transformó en algo parecido a lo que estudiamos actualmente en escuelas y universidades. En este punto los biólogos (entonces denominados médicos, botánicos o naturalistas), sin buscar una aplicación inmediata del conocimiento, dedicaron sus obras a describir “la vida” con rigor científico.

Las ciencias de la vida se cristalizaron en manuales de anatomía, herbarios y bestiarios. Dichas obras fueron posibles gracias al desarrollo de mecanismos y aparatos que permitieron expandir los límites de lo conocido. Y, a su vez, este nuevo conocimiento llevó a producir mejores

dispositivos. La indispensable sinergia entre ciencia y tecnología.

Recién con la llegada del siglo XIX, la biología se transformaría definitivamente en una ciencia moderna. En la misma época, dos grandes científicos trabajaban en paralelo para tratar de entender cómo se había diversificado la vida en nuestro planeta desde los primeros organismos formados por una sola célula hasta las aves, los mamíferos y los demás seres vivos que existen hoy.

Por un lado, Charles Darwin embarcaba a bordo del *Beagle* para su travesía por la costa atlántica de América del Sur. Por el otro, Alfred Wallace se encontraba en un periplo por el archipiélago malayo. Ninguno sabía en qué trabajaba el otro (¡faltaban casi doscientos años para que existieran las redes sociales!), lo que hace más maravilloso aún que hayan llegado a la misma conclusión. Esas ideas dieron lugar al paradigma de la evolución de las especies, que atravesó las ciencias de la vida desde entonces.

A esta altura, cabe preguntarse: ¿a qué viene tanto preámbulo?

El motivo de este breve repaso por la historia de las ciencias de la vida no es otro que el de mostrar las múltiples transformaciones que ha sufrido su razón y forma de estudio a lo largo del tiempo. La visión naturalista de hace dos siglos ha dado paso a otra manera de estudiar la biología más orientada hacia los datos. Este cambio de paradigma vuelve preponderante el rol de las tecnologías

de la información, de las que la biología se nutre para formar esta nueva disciplina que es la bioinformática.

La bioinformática es el eje del libro, pero también, en cierto modo, es una excusa para contar cómo los científicos trabajan para intentar comprender los enigmas de la vida.

En las próximas páginas, analizaremos el recorrido que han realizado las ciencias de la vida hasta convertirse en ciencias de la información, asistidas por algunas disciplinas jóvenes (como la informática) y otras bastante veteranas (como la matemática). Veremos cómo este grupo heterogéneo de ciencias nos permite acercarnos un poquito más hacia la comprensión de algún aspecto puntual del funcionamiento de los seres vivos. Podemos explorar lo que pasa dentro de una célula con la simulación del movimiento de moléculas y sus interacciones desde la pantalla de la computadora. O estudiar cómo células de diferentes tipos se agrupan para formar un tejido y un órgano. Incluso podemos ir a una escala mucho más grande para analizar cómo la introducción de una especie exótica en un área natural puede disminuir las poblaciones de especies autóctonas.

A medida que crece nuestro conocimiento de los sistemas biológicos, se vuelve más factible uno de los argumentos clásicos de la ciencia ficción: ¿el ser humano puede crear vida? O, en este caso, ¿podemos diseñar la vida desde una computadora? Aunque parezca lejano, cada día estamos un poco más cerca de que esto sea factible (¡no lo intenten en sus casas!).

Trabajar con computadoras tiene una ventaja adicional: es relativamente barato cuando se lo compara con los costos de tener un laboratorio en funcionamiento. Durante el desarrollo de un nuevo medicamento, usualmente miles de compuestos candidatos quedan en el camino, porque no sirven para lo que deberían o tienen demasiados efectos adversos. Si cada uno de esos candidatos tuviera que probarse en el laboratorio, estaríamos hablando de una inversión de tiempo y dinero considerable, pero afortunadamente gran parte de las pruebas pueden realizarse en computadoras, para que solo los más prometedores lleguen a los ensayos preclínicos y clínicos.

En resumen, la bioinformática se ocupa de la creación de modelos simplificados de la realidad que nos permiten entender mejor un fenómeno y realizar predicciones sobre el comportamiento del sistema en caso de que algo cambie.

Este es un libro sobre la vida... ¡pero en bits!

Posdata: *Vida.exe* estaba a punto de imprimirse cuando la historia (y, con ella, las imprentas) pareció detenerse por la pandemia del nuevo coronavirus. Como no se podía gambetear el tema, que fue una bisagra (también) para la ciencia, hemos agregado un último capítulo que reúne al menos una pequeña parte de las relaciones entre las disciplinas que abundan en el resto de los capítulos y ese fragmento de no vida bautizado SARS-COV-2: qué se sabe de su origen, del análisis matemático de las pandemias y de

las respuestas humanas a un problema que es, al menos en parte, humano, demasiado humano.

# *I. Sopa de letras*

Germán A. González

Somos el resultado final de más de mil millones de años de ajustes evolutivos, y nuestros genes contienen las costuras y remiendos que revelan la historia.

SPENCER WELLS

DICEN que el tiempo tiene sus propias maneras de develar la verdad. Por suerte, cuando ese tiempo es aprovechado con una mirada científica, algo de talento y mucho trabajo, podemos acercarnos un poquito a descubrir cómo funciona la naturaleza.

Además de buenas ideas, los científicos necesitan herramientas concretas que les permitan resolver las preguntas que se plantean, ya sea el compás de Darwin o las máquinas secuenciadoras de ADN del Proyecto Genoma Humano. La transformación tecnológica que ocurrió durante el último siglo tuvo un impacto inconmensurable en nuestra forma de estudiar a los seres vivos, convirtiendo las computadoras en la herramienta esencial que une todas las disciplinas científicas modernas.

Este pequeño viaje a través de la historia de la biología en sus cruces con la informática será un recuento que, si bien no planea ser exhaustivo, se detendrá en algunos momentos

clave para el desarrollo de la ciencia. Pónganse los cinturones: ¡ahí vamos!

## LOS LADRILLOS DE LA VIDA

La primera parada en este viaje por el tiempo nos deja en Berlín, a fines del siglo XIX. Los contextos favorables suelen fomentar el progreso, y esta no sería la excepción: en la capital del Imperio alemán, uno de los principales centros de la ciencia contemporánea, el bioquímico Albrecht Kossel trabajaba sobre el ADN descubierto años antes mientras se estudiaban los glóbulos blancos. Kossel lo bautizó con el nombre por el que lo conocemos actualmente y detalló por primera vez su composición. Esas largas cadenas que se había logrado aislar en el laboratorio estaban formadas por cuatro compuestos químicos con nombre y abreviatura propios: adenina (A), timina (T), citosina (C) y guanina (G). La combinación de apenas cuatro moléculas era suficiente para conformar la cadena de ADN, cuya función se desconocía. Este trabajo sobre los ácidos nucleicos le valió en 1910 el Premio Nobel de Medicina.

Fueron necesarias varias décadas y el trabajo de muchos científicos para llegar a comprender que el ADN es el “dispositivo de almacenamiento” donde se aloja la información que se hereda de padres a hijos. Sin embargo, había mucho que no se sabía. ¿Qué forma tiene el ADN? ¿Cómo usa el organismo esa información? Pasaron casi cincuenta años antes de que los científicos pudieran contestar estas preguntas.

## *Época de cambios*

Al terminar la Segunda Guerra Mundial, se produjeron algunas transformaciones en la manera de hacer ciencia. Por un lado, aparecieron los transistores y circuitos integrados, que permitieron progresivamente desarrollar computadoras más pequeñas y accesibles para los centros de investigación (hasta entonces, eran del tamaño de una habitación, como se ve en la película *El código enigma*, que retrata las hazañas del matemático Alan Turing).

Por el lado de la biología, comenzó lo que se llamaría “molecularización” de la disciplina, que significó prestar más atención a lo que el ADN, las proteínas y otros pequeños componentes de la célula podían decirnos sobre los organismos vivos.

### DESTAPAR LA DOBLE HÉLICE

La posguerra encontró a varios grupos de investigación en competencia para ser los primeros en descubrir la estructura del ADN; entre ellos, James Watson y Francis Crick en Cambridge, Maurice Wilkins y Rosalind Franklin en Londres y Linus Pauling en California. Se habían hecho varias propuestas sobre cómo podía ser la estructura, pero ninguno de los grupos había conseguido pruebas concretas. Tras muchos ensayos y esfuerzo, Franklin obtuvo una imagen por rayos x donde se podía observar claramente la doble hélice que hoy conocemos. Wilkins, un colega del instituto, le mostró la impactante foto (conocida como foto