

Capítulo 22

Las ómicas
en el control biológico

Chapter 22

Omics in
biological control

Gloria Patricia Barrera,¹ Pablo Daniel Ghiringhelli,²
Stephen Mosher,¹ Alejandro Caro-Quintero,¹
Sebastien Massart,³ Mariano Nicolás Belaich²

¹ Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA)

² Laboratorio de Ingeniería Genética y Biología Celular y Molecular - Área Virosis de Insectos (LIGBCM-AVI), Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes

³ Laboratory of Integrated and Urban Phytopathology, Gembloux Agro-Bio Tech, University of Liège

Contenido

Introducción954

Contexto histórico954

Tecnologías biológicas de alto rendimiento en genómica y transcriptómica955

Tecnologías biológicas de alto rendimiento en proteómica957

Tecnologías biológicas de alto rendimiento en metabolómica959

Bioinformática960

 La bioinformática y las disciplinas ómicas961

 La bioinformática y la transformación de datos en información963

Impacto de las tecnologías biológicas de alto rendimiento en el control biológico968

 Genómica y control biológico969

 Transcriptómica y control biológico970

 Proteómica y control biológico974

 Metabolómica y control biológico976

 Perspectivas y desafíos conceptuales978

Agradecimientos981

Referencias982

Resumen

En los últimos años los avances tecnológicos han posibilitado explorar la naturaleza como nunca y, en consecuencia, posicionar la biología como una ciencia de la información. Entre estas nuevas aproximaciones metodológicas de las ciencias de la vida se destacan las *disciplinas ómicas*, cuyo objeto de estudio son conjuntos biomoleculares particulares (ADN, ARN, proteínas o metabolitos secundarios) analizados mediante procedimientos de alto rendimiento. En este sentido, la genómica (con sus derivados: pangenómica y metagenómica), la transcriptómica, la proteómica y la metabolómica —entre otras disciplinas ómicas— conforman líneas de estudio imprescindibles para describir la naturaleza y potenciar la producción de bienes y servicios biotecnológicos. En este capítulo se hace una recopilación de los procedimientos empleados para hacer estudios ómicos y se resalta su utilidad en la investigación y el desarrollo de agentes de control biológico para invertebrados plaga y fitopatógenos. Se reseña, en consecuencia, la importancia del conocimiento genómico, transcriptómico, proteómico y metabolómico para la comprensión de las características moleculares de hongos, bacterias y virus utilizados en el control biológico. Además, se aporta información sobre el funcionamiento de las relaciones multipartitas para la implementación de la agricultura como un sistema productivo sostenible, amigable con el ambiente y resiliente.

Palabras clave

Genómica, metabolómica, proteómica, transcriptómica

Abstract

Recently, the technological advances have made it possible to explore nature as never before, and to position biology as an information science. Among these new methodological approaches are the disciplines known as "omics", which include all areas of life sciences that take particular biomolecular sets as study objects (eg, DNA, RNA, proteins or secondary metabolites), and that use high-throughput procedures for their approach. Thus, genomics (including derivatives known as pangenomics and metagenomics), transcriptomics, proteomics and metabolomics, among other omics, form important study lines to describe nature and from there, to enhance the generation of products and services in biotechnology. In this context, this chapter compiles the procedures used to carry out omic studies, with special emphasis on the research and development of biological control agents for invertebrate pest and phytopathogens. Thus, the importance of genomic, transcriptomic, proteomic and metabolomic knowledge is underlined in order to understand the molecular reasons that make bacteria, fungi and viruses useful in biological control, as well as to gather information about how multipartite relations work, all central aspects to turn agriculture into a productive, sustainable, environmentally friendly and resilient system.

Keywords

Genomics, metabolomics, proteomics, transcriptomics

Introducción

Las ciencias biológicas se han beneficiado de la revolución tecnológica ocurrida en las últimas décadas, gracias a la cual es más sencillo y productivo hoy en día generar evidencias útiles para la comprensión de la naturaleza. Si algo caracteriza y unifica al siglo XXI es que este representa el inicio de una era en la que la información parece ser la base de todo avance y desarrollo para la humanidad. Nadie duda de que tanto el acceso como el usufructo de la información en sus distintos niveles y dimensiones —desde los progresos en las comunicaciones interpersonales e interinstitucionales, hasta la mejora indiscutible de las ciencias informáticas— han adquirido buenos niveles de democratización y han permeado numerosas disciplinas y actividades humanas. En tal sentido, el beneficio de ello para las ciencias y tecnologías de la vida permitirá mejorar los sistemas agrícolas y transformarlos en más sustentables, resilientes y productivos, lo cual será de gran provecho para la seguridad alimentaria mundial.

Contexto histórico

Desde el siglo XIX las ciencias biológicas han incrementado su marco conceptual de manera significativa. El inicio de este sorprendente camino se debe a las ideas y demostraciones realizadas por Charles Darwin y Gregor Mendel (Charlesworth & Charlesworth, 2009; Liu & Li, 2015). Para el siglo XX, se alcanzó el descubrimiento de que la información contenida en la materia viva está almacenada en el ADN y se halla escrita con una sintaxis particular, con ciertos rasgos de universalidad entre especies. Esta información se expresa como ARN y proteínas, moléculas que concentran la mayor diversidad de actividades útiles en la construcción de células, organismos y, en mayor escala, poblaciones, comunidades y ecosistemas.

A partir de la década de los setenta, la ingeniería genética surgió como un nuevo e imprescindible brazo metodológico de la biotecnología, lo cual generó una

auténtica revolución socioproductiva (Miller, 2017). De cierta manera, es posible entender la ingeniería genética como la disciplina que asume el control de la información biológica, que desarrolla estrategias y procedimientos para adquirirla, entenderla e intervenirla con el fin de producir bienes y servicios útiles para la humanidad.

En este devenir histórico, como subproducto de lo anterior, cabe mencionar el nacimiento de las disciplinas ómicas, terminología derivada del sufijo *-oma*, que hace referencia a conjuntos de algo en estudio (Hasin et al., 2017; Moshelion & Altman, 2015). Así, hoy en día se habla con naturalidad de *genómica* (información genética de un individuo) (Thao & Tran, 2016), *pangenómica* (información genética de una especie) (Udaondo, Duque, & Ramos, 2017) y *metagenómica* (información genética de distintas especies conviviendo en un mismo sitio) (Crovadore et al., 2017). También se menciona la *transcriptómica* (moléculas de ARN expresadas por un tipo celular de un individuo) (Nawaz et al., 2017) y la *metatranscriptómica* (moléculas de ARN expresadas por la materia viva que habita en un sitio dado) (Newman et al., 2016). Finalmente, es usual describir evidencias derivadas de la *proteómica* (proteínas expresadas por materia viva) (Lande et al., 2017), la *interactómica* (redes de vínculos entre ADN, ARN y proteínas) (Rodríguez-Llorente et al., 2009) y la *metabolómica* (colección de metabolitos que ayuda a comprender funciones celulares) (Najafabadi, Naghavi, Farahmand, & Abbasi, 2017).

En todos los casos, se hace referencia a conjuntos de información cuya comprensión, manejo y control se realiza mediante aproximaciones de ingeniería genética. Dicha información es fundamental para el desarrollo de tecnologías esenciales para la salud humana o para actividades cruciales de nuestra subsistencia, como la agricultura y la ganadería. Hasin et al. (2017) muestran cómo el diseño de los estudios ómicos requiere de una planificación y ejecución cuidadosas, que suele demandar un gran número de comparaciones, análisis estadísticos e inversión de tiempo y dinero para la aplicación de tecnologías particulares.

Para lograr estos avances significativos ha sido necesario, por tanto, el desarrollo de nuevos procedimientos, conocidos como tecnologías biológicas de alto rendimiento (*high-throughput*), las cuales se caracterizan por incluir metodologías que permiten realizar de miles a millones de experimentos en paralelo y en tiempos cortos, para poder dominar conjuntos informativos complejos. La aplicación de estos métodos se evidencia en numerosos casos, por

ejemplo, en las respuestas de la soya a condiciones de estrés tanto de origen biótico (infecciones por *Fusarium oxysporum*, *Heterodera glycines* y el virus del mosaico) como abiótico (sequía, anegamiento, frío, salinidad y deficiencia de fósforo) (Tyczewska, Gracz, Kuczyński, & Twardowski, 2016). Todo este avance tecnológico vino acompañado del progreso de la bioinformática, disciplina que colabora en la comprensión y análisis de la información biológica.

Tecnologías biológicas de alto rendimiento en genómica y transcriptómica

Los primeros “proyectos genoma” estuvieron basados en la metodología de secuenciación propuesta por Frederick Sanger en 1975 (hoy reconocida como la primera generación de métodos para estudiar secuencias de ADN) (Sanger, Nicklen, & Coulson, 1977b). Después, se conformó un consorcio internacional, a finales de la década de los ochenta, para conocer el genoma humano (Moraes & Góes, 2016), objetivo que recién se comenzó a alcanzar a comienzos de este siglo (Lander et al., 2001). Paralelo a este enorme desafío, el mismo propósito se replicó para algunos organismos modelo como *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998), *Drosophila melanogaster* (Adams et al., 2000) o *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000), con los cuales se obtuvieron resultados exitosos antes que en nuestra especie.

Cabe destacar que el primer genoma secuenciado fue el de un bacteriófago (Sanger et al., 1977a), mientras que el primer ser vivo del que se conoció su secuencia genómica fue la bacteria *Haemophilus influenzae* RD (Fleischmann et al., 1995); ambos estudios fueron llevados a cabo mediante el método desarrollado por Frederick Sanger. Sin embargo, los costos y dificultades derivados de estas empresas a finales del siglo XX plantearon la necesidad de innovar tecnológicamente. Ante tales desafíos, en la primera década de este siglo surgieron nuevos procedimientos conocidos como “tecnologías de secuenciación de siguiente generación” (*next generation sequencing* o NGS) (Ambardar, Gupta, Trakroo, Lal, & Vakhlu, 2016), las cuales incluyen las siguientes tecnologías de *segunda generación*: 454

(comercializada por Hoffmann-La Roche, actualmente descontinuada), Illumina (comercializada por Illumina, Inc.) Solid y Ion Torrent (ambas comercializadas por Thermo Fisher Scientific).

Las NGS también incluyen las tecnologías de *tercera generación* o SMRT (*single molecule real time sequencing*), de las cuales, PacBio (comercializada por Pacific Biosciences of California, Inc.) y Nanopore (comercializada por Oxford Nanopore Technologies Limited) son sus representantes disponibles en la actualidad. Tanto las de segunda como de tercera generación son consideradas tecnologías biológicas de alto rendimiento, gracias a cuya aplicación es posible secuenciar genomas o transcriptomas parciales o completos en plazos cortos, así como disponer de información metagenómica y metatranscriptómica, todo a costos más accesibles.

Básicamente, el flujo de trabajo para emplear estas tecnologías consiste en aislar el ácido nucleico de la muestra problema: ADN para genómica o ARN para transcriptómica, provenientes, por ejemplo, de un individuo, de un conjunto de individuos de una misma especie o de una muestra ambiental (suelo o agua) que contiene distintas especies. Después, dicho ácido aislado se convierte en un ADN copia (si originalmente era ARN), se fragmenta y se une a segmentos de ADN de doble cadena de secuencia conocida (adaptadores) para generar una biblioteca (*DNA library*). Una vez obtenida la biblioteca, estas “quimeras” son capturadas en soportes sólidos (generalmente perlas esféricas o celdas de flujo) y amplificadas *in vitro* con ADN polimerasas (solo para

el caso de 454, Illumina, Solid y Ion Torrent, mediante cebadores que hibridan en los adaptadores), para luego adquirir la información de secuencia en paralelo para la totalidad de los fragmentos mediante diferentes métodos según la NGS en cuestión. Estos métodos suelen ser pirosecuenciación y detección de luz blanca (454), terminador de síntesis reversible y fluorimetría

(Illumina), ligación de sondas y fluorimetría (Solid), detección con semiconductores de protones derivados de síntesis (Ion Torrent), empleo de la guía de onda de “modo cero” (ZMW, *zero-mode waveguide*) y fluorimetría (PacBio); o detección de cambios de voltaje en una membrana con poros por donde pasan biomoléculas (Nanopore) (Ambardar et al., 2016) (figura 22.1).

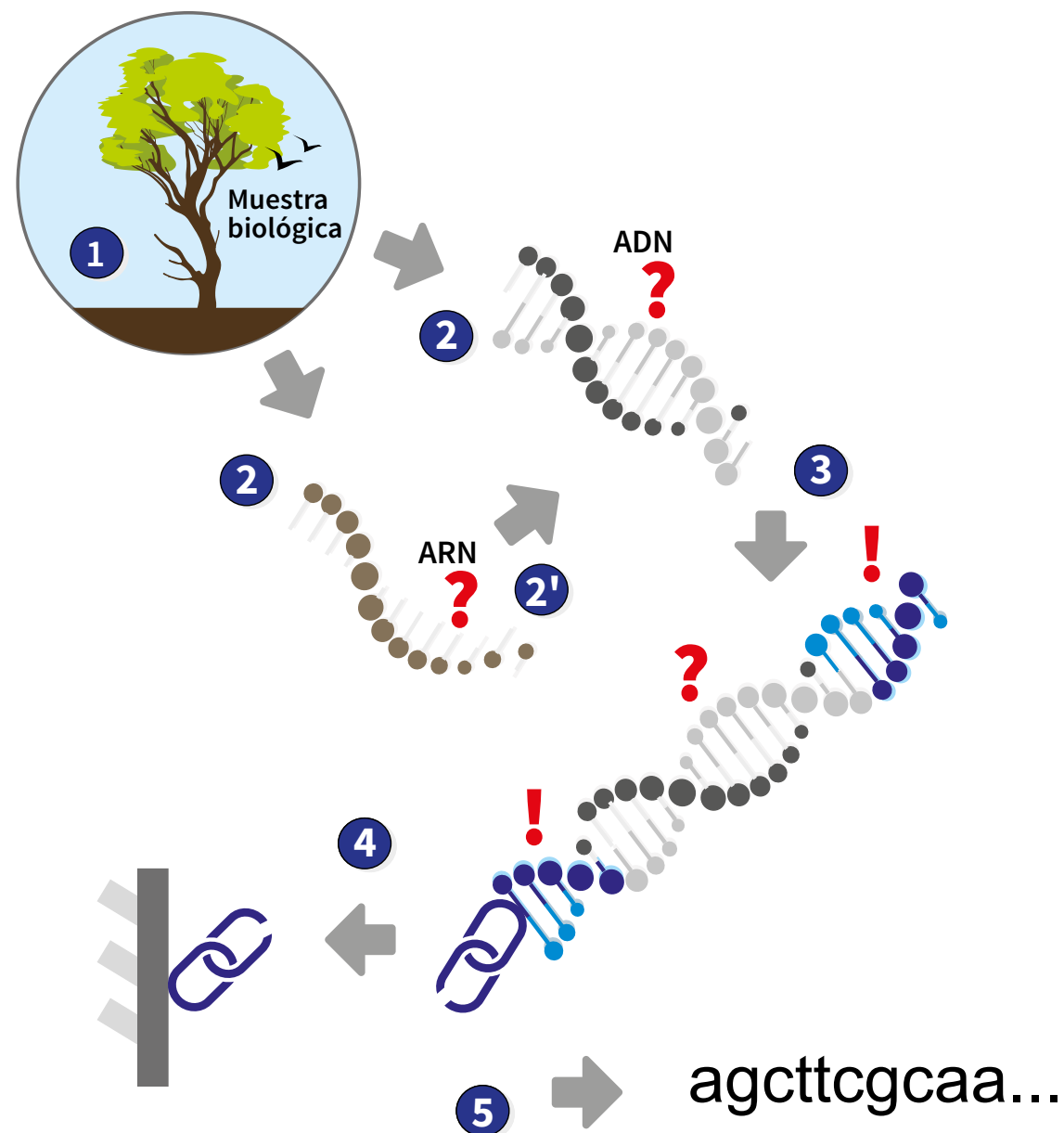


Figura 22.1. Flujo de trabajo para secuenciación de ADN y ARN. 1. Toma de muestra (de individuos, poblaciones y muestras complejas). 2. Aislamiento de ADN o ARN. 2'. El ARN se convierte a ADN copia. 3. Fragmentación y unión de adaptadores de ADN de doble cadena de secuencia conocida para generar una biblioteca (en color azul). 4. Captura en soporte sólido (en algunas tecnologías existe un paso posterior de amplificación *in vitro* con ADN polimerasas). 5. Proceso de secuenciación de cada uno de los fragmentos.

Fuente: Elaboración propia

Las NGS presentan diferentes costos y utilidades según sus capacidades diferenciales, y es común combinarlas para lograr un mayor poder de captura de información o seleccionarlas de acuerdo con el objetivo deseado (Hinrichs et al., 2015; Liu et al., 2012; Wei, Bemmels & Dick, 2014). Por ejemplo, para contemplar una gran cantidad de lecturas o profundidad de la secuenciación es recomendable utilizar Illumina o Solid, mientras que PacBio se destaca en la longitud individual de cada lectura y, por lo tanto, es útil para el ensamblado de las secuencias. Por su parte, Ion Torrent se destaca por mejorar los costos por megabases, pero si se incluye la variable del costo y la portabilidad de los equipamientos, el más económico y pequeño resulta ser el derivado de la tecnología Nanopore. En consecuencia, es necesario proponer el mejor algoritmo de trabajo según el fin particular deseado: conocer la secuencia de un genoma, conocer la diversidad de un pangenoma, disponer de la lista de especies presentes en una muestra o conocer las actividades expresadas en determinadas condiciones ambientales, entre otros.

Otra tecnología biológica de alto rendimiento que se emplea como complemento para la adquisición de secuencias genómicas y transcriptómicas por medio de las NGS es la hibridación en microarreglos (Douglas & Ehrling, 2005; Marzancola, Sedighi, & Li, 2016; Williams, Monge-Monge, Otvos, Reardon, & Ragenovich, 2011). Esta aproximación metodológica se difundió antes de la comercialización de las NGS, por lo cual fue el disparador del desarrollo y aplicación de la inquietud científica en la exploración de la variabilidad pangenómica de las especies y de los patrones de expresión génica derivados.

Estos estudios fueron necesarios para enriquecer y expandir la información genómica producida a

partir de los “proyectos genoma” realizados mediante tecnología Sanger, la cual resulta ahora inadecuada por costos y tiempos. Básicamente, la hibridación en microarreglos incluye el aislamiento de una muestra de ácidos nucleicos de interés (del mismo modo que lo indicado para el ADN en estudios genómicos o el ARN en análisis transcriptómicos) y su posterior fraccionamiento y marcación con grupos químicos fluorescentes, para luego enfrentarlos a pequeños chips que contienen miles de oligonucleótidos ordenados y de secuencia conocida (microarreglo) (figura 22.2). Así, en función de los segmentos retenidos (por complementariedad de las bases nitrogenadas), se podrá revelar qué secuencias estaban presentes en la muestra (lo cual indica las variantes genómicas del individuo en estudio o cuáles genes están siendo transcritos).

A pesar de que esta tecnología surgió previamente a los nuevos métodos de secuenciación, aún hoy se sigue utilizando con los mismos fines. Existen numerosas empresas que ofrecen el servicio para el diseño de chips a medida, y su empleo, por ser todavía más económico que el de las NGS, continúa presentándose como una excelente alternativa para enriquecer el conocimiento derivado de las disciplinas ómicas. En particular, la hibridación en microarreglos permite realizar estudios amplios de variabilidad pangenómica —generalmente expresada en SNP (*single nucleotide polymorphisms*) y CNV (*copy number variations*)— que incluyen numerosos individuos, con el posterior análisis de asociación a fenotipos particulares. Esto permite la determinación de biomarcadores útiles, por ejemplo, en el mejoramiento vegetal y animal o en la identificación de secuencias con valor patológico (Chang, Lipka, Domier, & Hartman, 2016; Evans et al., 2008).

Tecnologías biológicas de alto rendimiento en proteómica

Si bien la información de la materia viva está almacenada en moléculas de ADN y es posible revelar los indicios de su expresión a partir de las moléculas de ARN, las actividades centrales están representadas en la fracción proteica. Por ello, a veces resulta imprescindible conocer cuáles actividades biológicas ocurren en un momento dado, para lo cual, es de gran

importancia disponer de tecnologías de aproximación proteómica. Cabe destacar que, infortunadamente, no existen procedimientos tecnológicos para la aplicación en polipéptidos tan precisos y con tanta capacidad de recuperación de la información como los ofrecidos por las tecnologías NGS antes mencionadas. Sin embargo, se encuentran opciones que ayudan a su tratamiento,

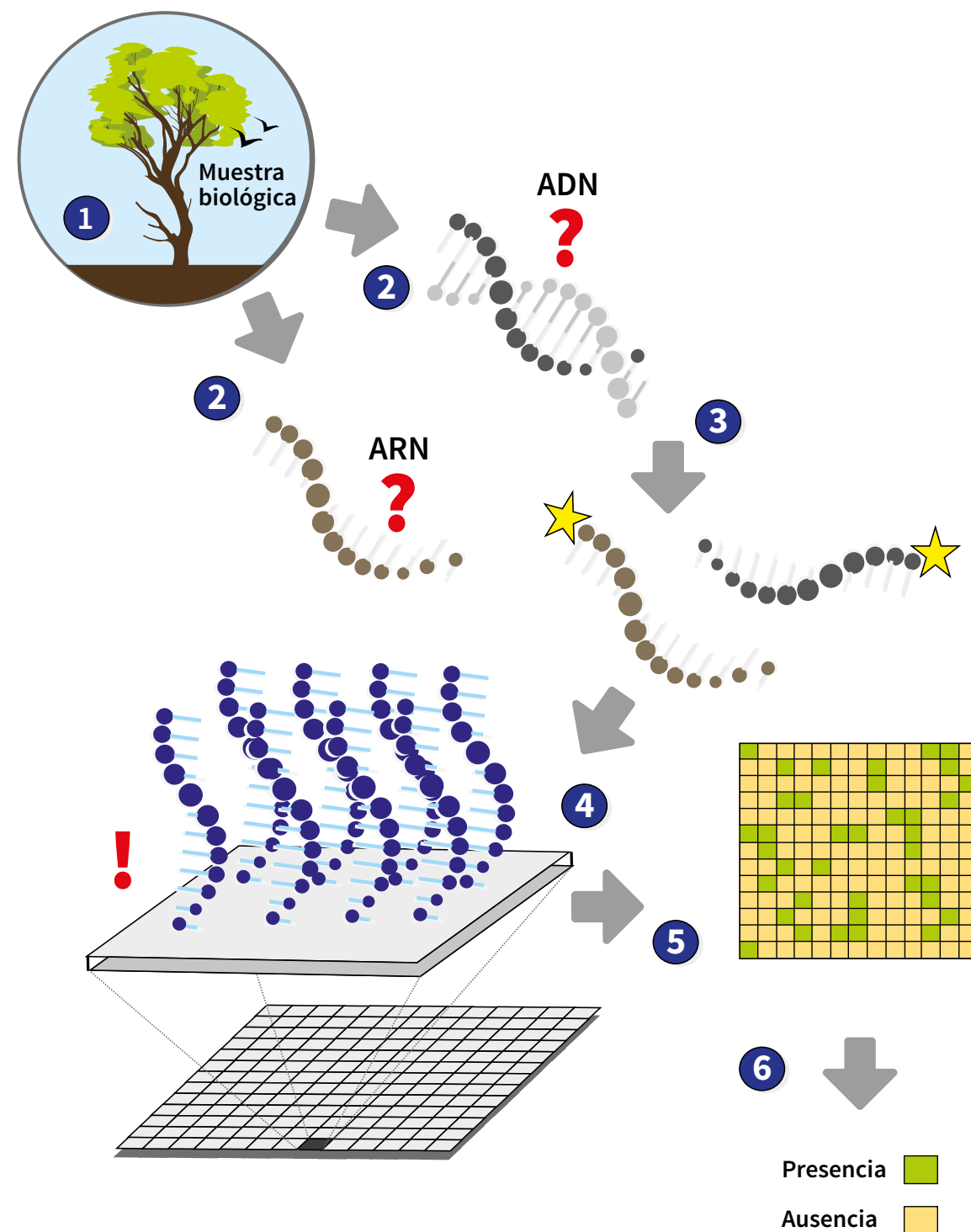


Figura 22.2. Flujo de trabajo para hibridación en microarreglos. 1. Toma de muestra (de individuos). 2. Aislamiento de ADN o ARN. 3. Fragmentación, desnaturalización y marcado con fluorocromos (es posible también amplificar con polimerasas). 4. Mezcla de los ácidos nucleicos de la muestra con el microarreglo en condiciones adecuadas. 5. Revelado. 6. Determinación de secuencias presentes y ausentes en la muestra. También, según como se realice el experimento, es posible extraer información cuantitativa.

Fuente: Elaboración propia

como la espectrometría de masas (Vidova & Spacil, 2017). Esta tecnología permite analizar con gran precisión la distribución de moléculas en función de su masa y carga, de modo tal que, sin secuenciar, se puede aportar evidencia significativa para inferir la presencia de determinadas proteínas y reconocer sus modificaciones postraduccionales. Para ello, la muestra de partida es ionizada mediante diferentes aproximaciones: por *electrospray* (ESI o *electrospray ionization*) o por desorción/ionización con láser asistida por una matriz (Maldi, o *matrix assisted laser desorption/ionization*), entre otras. Después, la muestra se somete a un campo electromagnético y se analizan las masas, también por diferentes aproximaciones: tiempo de vuelo (TOF o *time of flight*) y trampa de iones con cuadrupolo, entre otras. Existen numerosos ejemplos de su aplicación, que incluyen su uso exitoso en la identificación de bacterias patógenas de humanos (Bizzini & Greub, 2010) y de plantas

(Kazemi-Pour, Condemine, & Hugouvieux-Cotte-Pattat, 2004), así como de hongos fitopatógenos (Brun et al., 2013).

Debido al gran número de polipéptidos diferentes expresados por una materia viva, suele recomendarse un fraccionamiento de esta, antes de su análisis por espectroscopía de masas. Es frecuente emplear, entonces, electroforesis bidimensional (Najafabadi, Naghavi, Farahmand, & Abbasi, 2017) o técnicas de cromatografía líquida (Qu et al., 2016) para disponer de submuestras menos complejas que faciliten una más completa determinación (figura 22.3). Es importante señalar que, para un mejor aprovechamiento de las aproximaciones proteómicas, es conveniente primero conocer las secuencias genómicas de las materias vivas en estudio y disponer, así, de bases de datos que contengan las masas de las proteínas teóricas que podrían expresar esos organismos, de forma que se les pueda asignar una identidad.

Tecnologías biológicas de alto rendimiento en metabolómica

Todas las funciones que un organismo puede llevar a cabo están escritas en las moléculas de ADN que porta como genoma. Sin embargo, los genes no se expresan todo el tiempo ni se expresan en una proporción similar, por lo cual, si bien conocer esa información es de ayuda, esta no revela cuáles funciones están presentes. Aunque gracias al conocimiento actual que tenemos sobre la diversidad de motivos proteicos es posible hipotetizar —con base en los polipéptidos detectados experimentalmente en una muestra— cuáles serían las funciones activas, es necesario recurrir a otras fuentes de información para tener un conocimiento completo sobre genómica funcional. Es por esta razón que la metabolómica resulta de capital importancia para una mejor comprensión de la fisiología de una célula, de un tejido o de un organismo. Comparativamente, la información pangenómica o metagenómica suele ser poco variable a lo largo del tiempo (aunque para algunos ecosistemas o ciertas especies de vida corta sí pueden suceder variaciones contundentes, producto de la ocurrencia de cambios ambientales, muchos de ellos antropogénicos), mientras que las dimensiones transcriptómicas y proteómicas son consideradas *per se*

como de alto dinamismo. Por su parte, la metabolómica podría ser definida como la ventana biológica más dinámica de todas.

Como se mencionó antes, básicamente la metabolómica pretende describir el conjunto de metabolitos (moléculas orgánicas pequeñas, como hormonas, antibióticos, compuestos de señalización o metabolitos secundarios) presentes en una muestra biológica dada, lo cual puede traducirse en un perfil metabólico particular o huella que revela cuestiones fisiopatológicas puntuales del momento en que se tomó la muestra (Gogna, Singh, Sheeba, & Dorai, 2015).

Para su aplicación, primero se deben separar los metabolitos presentes mediante electroforesis capilar o cromatografías en fase gaseosa o líquida, como la cromatografía líquida de alto rendimiento (HPLC, *high performance liquid chromatography*). Luego se realiza su determinación y potencial cuantificación empleando técnicas que incluyen la espectroscopía de masas y de resonancia magnética nuclear (RMN). Todos estos procedimientos hacen posible, por ejemplo, el estudio

de los metabolitos que producen los microbios o la identificación de los cambios fisiológicos que suceden en insectos sometidos a estreses térmicos (Covington, McLean, & Bachmann, 2016; Chou, Pathmasiri, Deese-Spruill, Sumner, & Buchwalter, 2017). El estudio experimental presenta similitudes con el estudio proteómico (figura 22.3).

Los metabolitos se pueden clasificar, por ejemplo, en primarios y secundarios en virtud de las funciones en las cuales participan (funciones fisiológicas esenciales o auxiliares, respectivamente), y en endógenos y exógenos según el lugar en el que se hayan producido (en el mismo organismo o en otros del entorno, respec-

tivamente). De este modo, su presencia y cantidad relativa puede transformarse en una huella metabólica útil para revelar cuáles funciones biológicas (tanto fisiológicas como patológicas) se encuentran activas.

Bioinformática

La bioinformática es una disciplina científica emergente que se nutre de datos, conocimientos y estrategias provenientes de múltiples disciplinas, entre ellas, la biología, la bioquímica, la biofísica, la biotecnología,

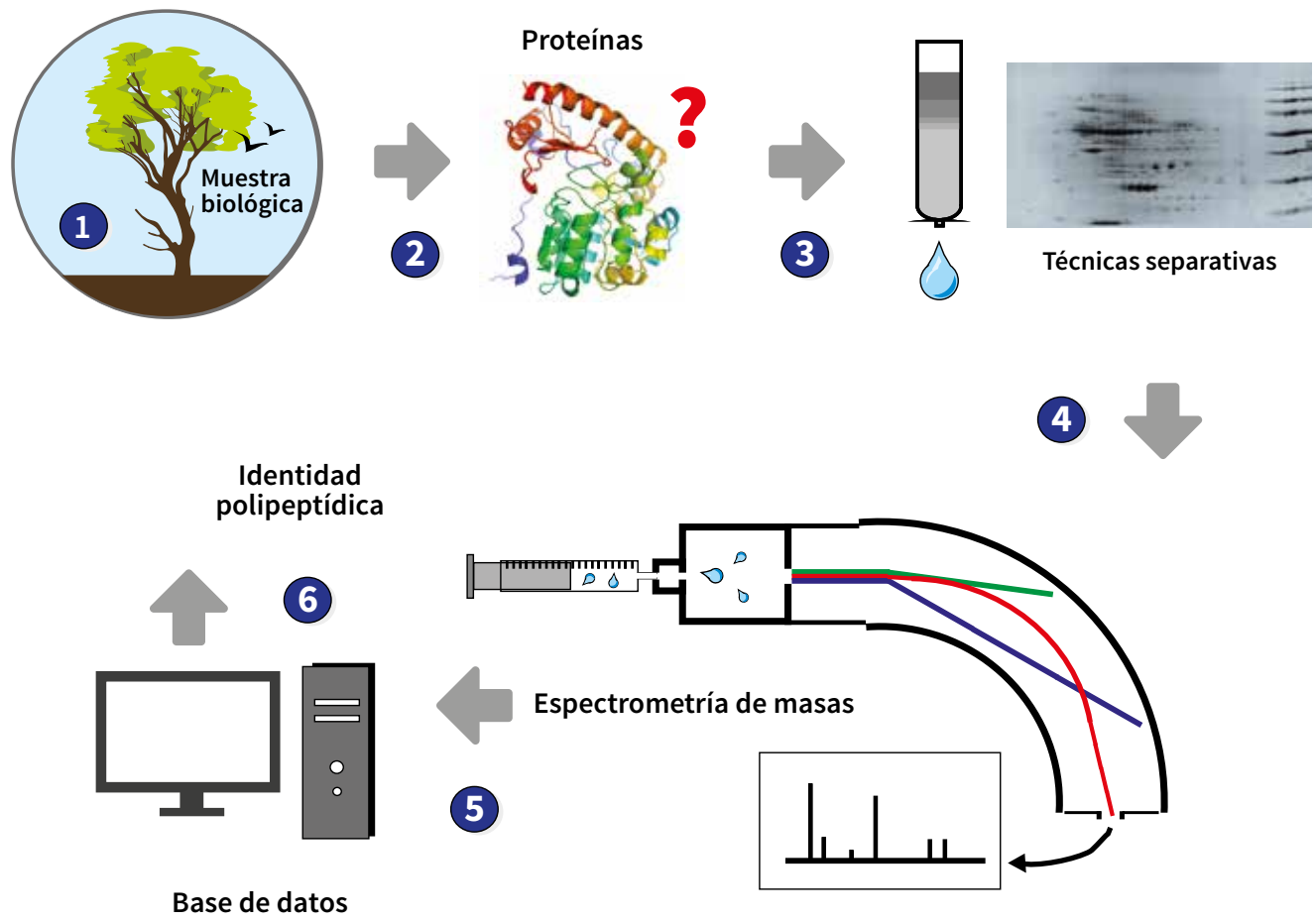


Figura 22.3. Flujo de trabajo para estudios proteómicos. 1. Toma de muestra (generalmente de individuos o tejidos de individuos, también de secretomas). 2. Aislamiento de proteínas. 3. Separaciones mediante cromatografía o electroforesis bidimensional (esta etapa puede llevarse a cabo de manera comparativa entre situaciones). 4. Fragmentación de polipéptidos y análisis mediante espectrometría de masas. 5. Análisis de resultados mediante aplicación de distintos tipos de *software* y comparación con bases de datos. 6. Determinación de la identidad de los polipéptidos presentes.

Fuente: Elaboración propia

la matemática, la estadística y la computación. La bioinformática es una ciencia de la información, ya que participa activamente en la organización y la transformación de datos, a partir de lo cual se renueva el conocimiento que se tiene sobre el objeto o el sistema en estudio.

La mayor parte de las ciencias que proveen datos a la bioinformática tienen una base esencialmente experimental. La interdisciplinariedad ha llegado a todas las áreas del quehacer “biológico”, lo cual ha generado importantes cambios en paradigmas básicos. Como antes se describió, en las últimas décadas se ha incrementado de forma exponencial la capacidad de recoger datos en distintos niveles de organización: desde lo molecular hasta lo sistémico. El conocimiento de secuencias genómicas, por ejemplo, es técnicamente accesible de forma rápida y relativamente económica. Sin embargo, la información contenida en estas debe ser decodificada para alcanzar el conocimiento pleno y postular su aplicación en distintas áreas de interés.

Por otra parte, la bioinformática no solo tiene a su cargo el análisis de datos complejos mediante el uso de herramientas computacionales, sino que también se involucra en el diseño y mantenimiento de bases de datos que reúnen, organizan y almacenan de forma estructurada la información biológica. Al mismo tiempo, la bioinformática define los estándares de los

procedimientos que permiten acceder y recuperar de forma eficiente la información biológica almacenada. En la actualidad, prácticamente no existen procesos biológicos cuyo estudio no implique etapas previas o posteriores de análisis bioinformático.

La bioinformática y las disciplinas ómicas

Es incuestionable que la secuenciación mediante síntesis e incorporación de terminadores de cadena (dideoxiNTPs), también conocida como secuenciación de Sanger (Sanger et al., 1977a), dominó el proceso de secuenciación durante más de 20 años. Sin embargo, la necesidad del acceso al conocimiento de nuevas secuencias genómicas superó ampliamente los avances tecnológicos desarrollados durante esos años. Así, en paralelo con el auge creciente de lo que se llama “la era posgenómica”, comenzaron a emerger las nuevas tecnologías de secuenciación antes mencionadas. En pocos años se fueron ofreciendo en el mercado nuevos métodos y equipos de secuenciación basados en tecnologías alternativas (tabla 22.1). Todas ellas permiten generar grandes volúmenes de datos (de millones a miles de millones de lecturas) en poco tiempo

Tabla 22.1. Tecnologías de secuenciación de segunda y tercera generación

Método	Medición	Nombre comercial	Referencia
Pirosecuenciación	Luz blanca (fotón)	FLX 454*	Margulies et al. (2006)
Síntesis (terminadores reversibles)	Fluorescencia	Illumina	Bentley (2006)
Ligación (octámeros fluorescentes)	Fluorescencia	Solid*	McKernan et al. (2009)
Síntesis (fluoróforos en 3')	Fluorescencia	Helicos*	Harris et al. (2008)
Síntesis (fluoróforo en los fosfatos)	Fluorescencia	PacBio	Eid et al. (2009)
Síntesis	Voltaje	Ion Torrent	Rothberg et al. (2011)
Flujo a través de un nanoporo	Voltaje	MinION	Stoddart, Heron, Mikhailova, Maglia, & Bayley et al. (2009)

* Descontinuada

Fuente: Elaboración propia

(desde unas pocas horas hasta un máximo de alrededor de diez días) y de forma económica (aproximadamente entre US\$300 y US\$1.500).

Los grandes volúmenes de datos producidos por NGS significan un reto enorme para la bioinformática: ¿cómo ordenar adecuadamente la mayor parte de las lecturas en información genómica de buena calidad? A este proceso se le llama *ensamblaje*, del cual existen dos alternativas procedimentales. Por un lado, se pueden ensamblar las lecturas respecto de un genoma de referencia para después mapearlas sobre la plantilla seleccionada con el fin de determinar su orden y solapamiento; por otro, se pueden ensamblar *de novo*. Para ello, se utilizan solo las lecturas obtenidas en la secuenciación masiva para generar los *contigs* (segmentos de secuencia derivados de lecturas contiguas) y los *scaffolds* más probables (unión de *contigs* vinculados por proximidad con base en variantes de la tecnología de secuenciación), sin compararlos *a priori* con ningún genoma de referencia.

En cualquiera de las dos alternativas, el paso final (denominado *finalización* o *terminación del genoma*)

es el más laborioso, e implica la resecuenciación con métodos alternativos (por ejemplo, inicialmente con Illumina o Ion Torrent y después con PacBio) o el uso de PCR (*polymerase chain reaction*, ‘reacción en cadena de la polimerasa’) y secuenciación estándar de Sanger si las regiones desconocidas o de baja calidad son pocas y pequeñas en longitud.

La figura 22.4 resume el flujo de trabajo de un proceso de secuenciación genómica. El cambio de paradigma alcanzado con el advenimiento de la era posgenómica facilita la obtención inicial de los datos ómicos en múltiples niveles, pero el camino para su transformación en información útil y confiable no es corto (Aguilar-Bultet & Falquet, 2015; Glenn, 2011).

En la era posgenómica, las principales áreas de investigación con metodologías *high throughput* son las siguientes: la genómica, la transcriptómica, la proteómica y la metabolómica, aunque el universo de las ómicas ha crecido tanto que se ha propuesto un lenguaje específico para identificar subáreas de trabajo (Greenbaum, Luscombe, Jansen, Qian, & Gerstein, 2001).

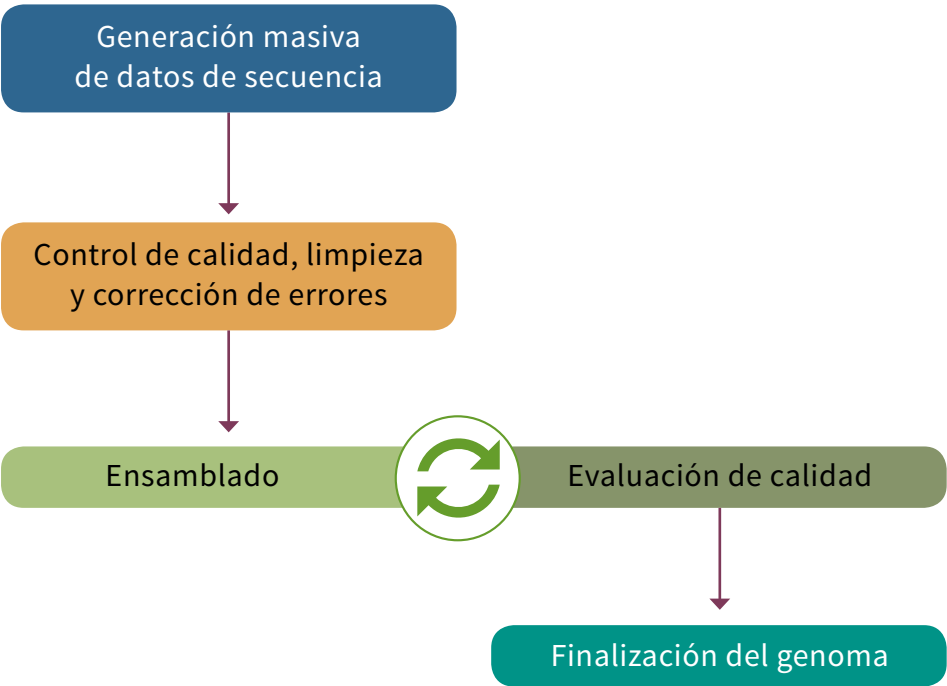


Figura 22.4. Esquema del procesamiento de datos obtenidos mediante secuenciación masiva. Las etapas de ensamblado y evaluación de calidad son iterativas hasta que se alcanza la calidad adecuada.

Fuente: Elaboración propia

La bioinformática y la transformación de datos en información

Con el objetivo de incrementar la eficacia de los potenciales agentes de control biológico, las tecnologías masivas actuales permiten recolectar grandes volúmenes de datos vinculados con los siguientes propósitos:

- 1. Identificar las características moleculares del agente de biocontrol.
- 2. Indagar acerca de las bases moleculares de propiedades diferenciales de cepas y variantes genéticas con diferente capacidad de control biológico.
- 3. Identificar y caracterizar las bases moleculares (genes, transcriptos, proteínas y rutas metabólicas) en cuanto a los modos de acción del agente de biocontrol.
- 4. Estudiar en forma integrada el sistema biológico de intervención, conformado por el agente de biocontrol, el hospedero y el resto de organismos presentes.

Sin embargo, aun cuando la obtención de datos masivos se ha facilitado (en tiempo y en dinero), el camino posterior para su organización e interpretación es una ardua tarea que requiere de la interacción y colaboración de diferentes grupos de investigación y organizaciones vinculadas con el mundo académico. Para facilitar esto, se ha logrado universalizar el acceso a la información obtenida a partir de trabajos previos, mediante la organización de bases de datos. Estas bases de datos cuentan con estructuras y contenidos alternativos que mejoran la calidad de las inferencias de “significado biológico” a partir de los datos recolectados en experimentos masivos. En síntesis, permiten una mejor transformación de los datos crudos en información biológica precisa. En la tabla 22.2, se listan algunas de las bases de datos que se han ido creando en el transcurso de las últimas tres décadas.

Con base en lo discutido, es posible mencionar que para todos estos temas existen múltiples ejemplos de trabajos de investigación en los cuales el aporte de las disciplinas ómicas y su interpretación bioinformática ha sido central (Conesa et al., 2016; Lorito, Woo, Harman, & Monte, 2010; Massart, Perazzolli, Höfte, Pertot, & Jijakli, 2015; Oshlack, Robinson, & Young, 2010; Van Lenteren, Bolckmans, Köhl, Ravensberg, & Urbaneja, 2017).

Tabla 22.2. Bases de datos de genómica, transcriptómica, proteómica y metabolómica

Denominación y breve descripción		URL
Genómica	DDBJ (DNA Data Base of Japan). Repositorio de secuencias de ADN a partir de cualquier tipo de sistema biológico y también de secuencias artificiales (sintéticas o patentes)	http://www.ddbj.nig.ac.jp/
	EMBL (European Molecular Biology Laboratories). Repositorio de secuencias de ADN a partir de cualquier tipo de sistema biológico y también de secuencias artificiales (sintéticas o patentes)	http://www.embl.org/
	GenBank . Repositorio de secuencias de ADN a partir de cualquier tipo de sistema biológico y también de secuencias artificiales (sintéticas o patentes)	http://www.ncbi.nlm.nih.gov/GenBank/

(Continúa)

(Continuación tabla 22.2)

Denominación y breve descripción		URL
Genómica	Genomas	Ensembl. Repositorio de información biológica de genomas grandes (humano, ratón, mosca de la fruta, pez cebra, etc.) http://www.ensembl.org
		Ensembl Genomes. Repositorio de información biológica de otros sistemas biológicos (invertebrados, plantas, hongos, protistas, etc.) http://ensemblgenomes.org/
		Patric (PathoSystems Resource Integration Center). Repositorio de genomas de bacterias patógenas http://patricbrc.vbi.vt.edu/portal/portal/patric/Home
		NMPDR (National Microbial Pathogen Data Resource). Repositorio de genomas de bacterias patógenas, principalmente del tipo B http://www.nmpdr.org
		SGD (Saccharomyces Genome Database). Repositorio del genoma de <i>Saccharomyces cerevisiae</i> , genes y productos codificados, fenotipos mutantes, etc. http://www.yeastgenome.org
		VBRC (Viral Bioinformatics Resource Center). Almacena genomas virales. Contiene datos sobre siete familias de patógenos emergentes o reemergentes: Arenaviridae, Bunyaviridae, Filoviridae, Flaviviridae, Paramyxoviridae, Poxviridae y Togaviridae; y cinco familias adicionales: Adenoviridae, Asfarviridae, Baculoviridae, Herpesviridae e Iridoviridae. http://www.biovirus.org/
		VectorBase. Repositorio del genoma de invertebrados que son vectores de patógenos humanos http://www.vectorbase.org/
Transcriptómica	ARN	Rfam. Repositorio de información acerca de familias de ncRNA (non-coding RNA) y otros elementos de RNA estructurado http://rfam.xfam.org
		miRbase. Repositorio de secuencias de micro-RNA y sus correspondientes anotaciones http://www.mirbase.org/
		lncRNAdb. Contiene anotaciones sobre lncRNA (<i>eukaryotic long non-coding RNAs</i>) http://www.lncrnadb.org/
		RDP (Ribosomal Database Project). Permite clasificar secuencias de rRNA 16S rápidamente y con certeza, de acuerdo con las últimas reglas taxonómicas propuestas en el <i>Manual de Bergey (Taxonomic outline of the Prokaryotes)</i> http://rdp.cme.msu.edu/
		SRPDB (Signal Recognition Particle Database). Provee secuencias filogenéticamente ordenadas, alineadas y con anotaciones correspondientes a la estructura y función de las SRP http://rth.dk/resources/rnp/SRPDB/
		tmRDB (Transfer-Messenger RNA Database). Provee secuencias filogenéticamente ordenadas, alineadas y con anotaciones correspondientes a la estructura y función de las tmRNP http://www.ag.auburn.edu/mirror/tmRDB/

(Continúa)

(Continuación tabla 22.2)

Denominación y breve descripción		URL
Proteómica	Expresión de proteínas	Expression Atlas. Repositorio de resultados de expresión génica a partir de más de 3.000 experimentos asociados con 40 organismos diferentes, incluyendo metazoarios y plantas http://www.ebi.ac.uk/gxa/home/
		GEO (Gene Expression Omnibus). Repositorio público internacional que archiva y distribuye libremente datos de genómica funcional derivados de tecnologías de <i>high-throughput</i> https://www.ncbi.nlm.nih.gov/geo/
		UniProt (Universal Protein Resource). Repositorio central de datos proteicos, secuencias y anotaciones funcionales. Es la combinación de las bases de datos Swiss-Prot, TrEMBL y PIR-PSD http://www.uniprot.org/
		PIR (Protein Information Resource). Repositorio para asistir a los investigadores en la identificación e interpretación de información relacionada con secuencias proteicas http://pir.georgetown.edu/
		Swiss-Prot. Repositorio que provee secuencias proteicas con anotación exhaustiva (descripción de función, estructura de sus dominios, modificaciones postraduccionales, variantes proteicas, etc.), con un nivel mínimo de redundancia y con integración con otras bases de datos http://www.expasy.org/sprot
		Pedant (Protein Extraction, Description and Analysis Tools). Provee un análisis automatizado exhaustivo de secuencias genómicas con una gran variedad de herramientas bioinformáticas http://pedant.gsf.de
		Prosite (Protein Sites). Es una base de datos de proteínas que describe familias proteicas, dominios y sitios funcionales, así como los patrones y perfiles aminoacídicos presentes en ellas http://prosite.expasy.org
		Pfam (Protein families). Base de datos que incluye dominios y familias proteicas. Pfam-A almacena un alineamiento múltiple y un modelo oculto de Markov. Pfam-B se generó automáticamente y contiene un gran número de pequeñas familias. Pfam-B es útil cuando no se detecta ninguna familia en Pfam-A http://pfam.xfam.org
		InterPro. Base de datos de familias proteicas, dominios y sitios funcionales. El contenido de InterPro está basado en firmas moleculares y en las proteínas que tienen similitud significativa con ellas. Las firmas moleculares consisten en modelos simples (expresiones regulares) o más complejos (modelos ocultos de Markov) http://www.ebi.ac.uk/interpro/
		Prints (Protein Fingerprints). Colección de <i>fingerprints</i> con anotación detallada para las familias proteicas y con herramientas de diagnóstico para usar con secuencias nuevas http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/
		ProDom (Protein Domains). Base de datos de familias proteicas construida automáticamente mediante el agrupamiento de segmentos homólogos (<i>clustering</i>) http://prodom.prabi.fr/prodom/current/html/home.php

(Continúa)

(Continuación tabla 22.2)

Denominación y breve descripción		URL	
Proteómica	Estructuras proteicas	PDBe & PDBj (Protein Data Bank Europe o Protein Data Bank Japan). Repositorio de datos de estructuras 3D de moléculas biológicas grandes (especialmente proteínas)	http://www.pdbe.org
		CATH (CATH Protein Structure Classification). Contiene una clasificación jerárquica de dominios proteicos basada en sus estructuras	http://www.cathdb.info/
		SCOP2 (Structural Classification of Proteins, sucesor de la base de datos scop). Su principal objetivo son las proteínas estructuralmente caracterizadas y depositadas en el PDB. Las proteínas están organizadas con base en sus relaciones estructurales y evolutivas	http://scop2.mrc-lmb.cam.ac.uk/
		RCSB PDB. Base de datos orientada a la biología estructural. Provee herramientas de búsqueda y análisis de las estructuras	http://www.rcsb.org/pdb/home/home.do
	Modelos proteicos	Swiss-Model. Servidor web de bioinformática estructural dedicado al modelado por homología de estructuras proteicas 3D	http://swissmodel.expasy.org
		ModBase. Base de datos de modelos proteicos prolijamente anotados. Contiene más de 3,8 millones de secuencias proteicas únicas	http://salilab.org/modbase
		PMP (Protein Model Portal). Desarrollado para facilitar el acceso a un uso efectivo de los datos de modelos 3D en la investigación biomédica	http://www.proteinmodelportal.org/
	Identificación proteica	Pride (Proteomics Identifications). Repositorio centralizado de datos proteómicos. Incluye la identificación de proteínas y péptidos, modificaciones postraduccionales y evidencias espectrofotométricas de soporte	http://www.ebi.ac.uk/pride/archive/
		GelMap. Herramienta universal para la visualización de <i>spots</i> en la imagen de un gel 2D. Almacena mapas de referencia y permite identificar proteínas por sus características de migración en geles 2D	http://www.genetik.uni-hannover.de/pflanzenproteomik.html
	Interacciones proteína-proteína y otras	BIND (Biomolecular Interaction Network Database). Repositorio de datos de interacción proteica acompañados de información funcional e información bibliográfica	http://bond.unleashedinformatics.com/
BioGRID (Biological General Repository for Interaction Datasets). Colección manualmente corregida de interacciones entre macromoléculas		http://thebiogrid.org/	

(Continúa)

(Continuación tabla 22.2)

Denominación y breve descripción		URL	
Proteómica	Interacciones proteína-proteína y otras	CCSB. Repositorio de interacciones binarias proteína-proteína de varios sistemas: <i>H. sapiens</i>, <i>A. thaliana</i>, <i>S. cerevisiae</i>, <i>C. elegans</i>, virhostoma (polyomavirus/hospedador, papillomavirus/hospedador, adenovirus/hospedador y Epstein-Barr/hospedador) y fragmentoma. El fragmentoma se basa en proteínas modulares (con dominios que se pliegan independientemente) que pueden establecer un mismo tipo de interacción aun cuando estén acopladas en dominios diferentes.	http://interactome.dfci.harvard.edu/
	DIP (Database of Interacting Proteins). Base de datos biológicos que organiza y cataloga interacciones proteína-proteína experimentalmente determinadas	http://dip.doe-mbi.ucla.edu/	
	IntAct. Base de datos de interacciones moleculares, que incluyen proteína-proteína, proteína-pequeñas moléculas y proteína-ácidos nucleicos	http://www.ebi.ac.uk/intact/site/index.jsf	
	String (Search Tool for Recurring Instances of Neighbouring Genes). Es una base de datos de interacciones proteína-proteína, determinadas experimentalmente o predichas	http://string.embl.de/	
	MINT (Molecular INTeraction). Está enfocada exclusivamente en interacciones proteína-proteína verificadas de forma experimental	http://mint.bio.uniroma2.it/mint/Welcome.do	
	iRefWeb. Web de interfaz que permite acceder a datos de interacción proteica consolidados a partir de diez bases de datos públicas: BIND, BioGRID, Corum, DIP, IntAct, HPRD, MINT, MPact, MPPI y OPHID	http://wodaklab.org/iRefWeb/	
	BioCyc Database Collection. Conjunto de más de 5.000 bases de datos con información específica de genomas y rutas metabólicas. Comprende MetaCyc, HumanCyc, EcoCyc, AraCyc, YeastCyc, LeishCyc y TrypanoCyc	http://biocyc.org/	
Metabolómica	Enzimas	Brenda. Fuente electrónica de información molecular y bioquímica sobre las enzimas que han sido clasificadas por la IUBMB (International Union of Biochemistry and Molecular Biology)	http://www.br-end-enzymes.org
	Manet Database (Molecular Ancestry Network). Base de datos bioinformáticos que mapea las relaciones evolutivas de la arquitectura proteica directamente sobre las redes biológicas	http://www.manet.uiuc.edu/	

(Continúa)

(Continuación tabla 22.2)

Denominación y breve descripción		URL	
Metabolómica	Redes metabólicas	Kegg Pathway Database (Kyoto Encyclopedia of Genes and Genomes). Colección de bases de datos relacionadas con genomas, rutas metabólicas, enfermedades, drogas y sustancias químicas	http://www.kegg.jp
		Metabolights . Base de datos de experimentos metabolómicos y la información derivada	http://www.ebi.ac.uk/metabolights/
		MetaNetX . Servidor que posibilita la construcción automatizada de modelos de redes metabólicas a gran escala y la anotación genómica correspondiente	http://metanetx.org/
		SMPDB (Small Molecule Pathway Database). Base de datos interactiva que contiene más de 600 rutas metabólicas de moléculas pequeñas detectadas en <i>Homo sapiens</i>	http://smpdb.ca/
	Metabolitos pequeños	HMDB (Human Metabolome Database). Contiene información detallada acerca de metabolitos pequeños hallados en el cuerpo humano	http://www.hmdb.ca/
		YMDB (Yeast Metabolome Database). Contiene información detallada y manualmente verificada de metabolitos pequeños hallados en o producidos por <i>Saccharomyces cerevisiae</i>	http://www.ymdb.ca/
		ECMDB (Escherichia Coli Metabolome Database). Contiene información detallada y manualmente verificada de metabolitos pequeños hallados en o producidos por <i>Escherichia coli</i>	http://www.ecmdb.ca/

Fuente: Elaboración propia

Impacto de las tecnologías biológicas de alto rendimiento en el control biológico

Las disciplinas ómicas han tenido un impacto trascendente en todas las ramas científico-tecnológicas. Los avances en el modelado de los sistemas biológicos, en los estudios de genómica funcional, en las herramientas bioinformáticas que apoyan la producción de información a partir de datos experimentales y en los procedimientos de ingeniería genética de la era posgenómica han producido una extraordinaria revolución en la agricultura y le permiten al ser humano una cantidad importante de aplicaciones concretas (Thao & Tran, 2016). Por ejemplo, gracias a las NGS se puede conocer la relación ancestral de las variedades de plantas que cultivamos (Brozynska et al., 2017),

describir el microbioma de la rizosfera que favorece la fijación de nutrientes (Mitter, De Freitas, & Germida, 2017) o ayudar en la completa identificación de virus que comprometen la sanidad vegetal (Rozado-Aguirre, Adams, Fox, Dickinson, & Boonham, 2017). También, es posible conocer el genoma de los invertebrados plaga (con toda la variabilidad pangenómica de las especies circulantes) (Xie et al., 2017) y de sus enemigos naturales (otros invertebrados, bacterias, hongos y virus biocontroladores), como es el caso de un nucleopoliedrovirus colombiano que infecta larvas de *Spodoptera frugiperda*, cuyo análisis genómico reveló la ocurrencia en la naturaleza de fenómenos de

transferencia horizontal de genes (Barrera, Belaich, Patarroyo, Villamizar, & Ghiringhelli, 2015).

Además, se puede comprender cómo atacan y cómo se defienden los organismos presentes en nuestros agroecosistemas, lo cual revela qué moléculas y funciones se expresan ante determinadas circunstancias fisiopatológicas mediante diferentes aproximaciones: transcriptómicas, por ejemplo, que ayudaron en el descubrimiento de los genes asociados a la resistencia insecticida en *Bactrocera dorsalis* (Hsu et al., 2012); proteómicas, que sirven, por ejemplo, para controlar el crecimiento y el desarrollo de las plantas, así como para entender sus respuestas a estreses bióticos y abióticos (Tan, Lim, & Lau, 2017); y metabolómicas, que, entre otras aplicaciones, ayudaron a entender los efectos de algunos pesticidas sobre los vegetales que consumimos (Zhao, Huang, Adeleye, & Keller, 2017).

En referencia particular al control biológico, las herramientas ómicas posibilitan caracterizar en términos moleculares los ingredientes activos que conforman formulaciones bioplaguicidas, es decir, ingredientes activos de los entomopatógenos y antagonistas de fitopatógenos (por ejemplo, virus, bacterias, hongos o nematodos). De este modo, se facilita la elección de los mejores candidatos, su posterior registro y su seguimiento en campo. Las herramientas ómicas también son de vital importancia para el reconocimiento de las especies o biotipos de los organismos plaga en los cultivos (fitopatógenos y fitófagos) y para la descripción de las huellas metabólicas derivadas de las interacciones entre pares de una misma especie o con las especies que utilizan como alimento.

Genómica y control biológico

La genómica es el estudio integral del funcionamiento, el contenido, la evolución y el origen de los elementos genéticos de un organismo (cromosomas y plásmidos). El estudio de los genomas de microorganismos ha permitido identificar los mecanismos genéticos responsables de sus características fisiológicas relevantes: las capacidades enzimáticas de degradación, factores de virulencia, mecanismos de reconocimiento de hospederos, producción de fitohormonas y control de

patógenos, entre otros. Esta identificación del potencial fisiológico se basa en los principios de la ruta central de la biología molecular, que establece que el flujo de la información en los organismos vivos parte de los genes en los cromosomas y se traduce en forma de función (proteínas) y estructura (células) (Crick, 1970). Por lo tanto, si se conoce la composición genética, es posible determinar el potencial del organismo.

Los genomas también pueden utilizarse para identificar los cambios genéticos, como la adquisición y pérdida de genes, o las ganancias o deleciones microestructurales, factores que pueden alterar de manera drástica o sutil las funciones celulares. Dichos cambios modulan las capacidades metabólicas de los organismos y determinan su funcionalidad en los ambientes, por ejemplo, la patogenicidad. Además, debido a que el genoma contiene todas las mutaciones acumuladas a través de la historia evolutiva del organismo, puede usarse para dilucidar de un mejor modo su procedencia.

El uso de aproximaciones de anotación bioinformática multinivel permite reconstruir las capacidades metabólicas primarias de los organismos a partir de la información genómica. Estas reconstrucciones son posibles gracias a la caracterización experimental de múltiples genes microbianos, la gran acumulación de información genómica en las bases de datos y el desarrollo de herramientas que permiten predecir funciones por medio de búsqueda de homología y algoritmos *ab initio* (como InterProScan) o crear mapas metabólicos (como KEGG) (tabla 22.2). La base de datos de KEGG, por ejemplo, es un recurso que permite el entendimiento de alto nivel de las funciones de los sistemas biológicos. Por otro lado, BioCyc es una colección de cerca de 7.667 rutas metabólicas basadas en genomas generadas con el programa de reconstrucción Pathway Tools de la compañía SRI International (Caspi et al., 2009). Estas aproximaciones aún no son perfectas, pero permiten generar hipótesis que, en conjunto con la información experimental, posibilitan la identificación y potenciación de genes asociados a una función de interés.

La reconstrucción funcional y estructural a partir de la información genómica ha permitido, entre otras cosas, guiar la manipulación genética para favorecer metabolitos de interés biotecnológico, entender las capacidades del metabolismo central y estructural que

distinguen un aislamiento nuevo de aquellos más cercanos, fortalecer la descripción de nuevas especies (Konstantinidis & Tiedje, 2005) y mejorar las técnicas de aislamiento y cultivo. A pesar de la enorme ventaja que ofrece la caracterización genómica en el control biológico, es importante señalar que la reconstrucción fisiológica representa una hipótesis de la funcionalidad y que este es un borrador con varios sesgos. Uno de los mayores inconvenientes en la asignación de funciones es la falta de caracterizaciones de genes correspondientes al metabolismo secundario que puedan ser relacionados con funciones biocontroladoras.

Por otro lado, la secuenciación genómica puede servir como una hoja de ruta para la caracterización profunda de organismos élites con potencial de bioprospección en el control biológico. Esto se ha demostrado en microorganismos para el control de insectos plaga (Barrera et al., 2015; Cuartas et al., 2015) y en el control de fitopatógenos (revisado por Massart et al., 2015) como *Pseudomonas protegens* Pf-5 (Paulsen et al., 2005). Esta bacteria, por ejemplo, es reconocida en el campo del control biológico como uno de los controladores de referencia por la gran variedad de enfermedades que suprime. *Pseudomonas* Pf-5 se aisló por primera vez de la rizosfera del algodón, y se ha estudiado ampliamente debido a su capacidad supresora de enfermedades causadas en las plántulas de algodón por *Rhizoctonia solani* (Howell & Stipanovic, 1979) y por *Pythium ultimum* (Howell & Stipanovic, 1980). Estos patógenos son causantes de la podredumbre de las semillas y la muerte de las plántulas del algodón y de muchas otras plantas. Además de los patógenos ya mencionados, Pf-5 suprime otros hongos patógenos transmitidos por el suelo o por residuos. Por ejemplo, en el residuo de la paja de trigo, Pf-5 suprime la formación de ascocarpos del patógeno de manchas de trigo (*Pyrenophora tritici-repentis*), lo cual disminuye la carga de inóculo disponible para la reinfección.

La amplia evidencia de *Pseudomonas protegens* Pf-5 como agente biocontrolador ha despertado un enorme interés en la caracterización de los mecanismos genéticos involucrados en la actividad controladora. Estudios iniciales permitieron la identificación de una región genómica requerida para la producción de antibióticos como la pioluteorina (Kraus & Loper, 1995). Este es un antibiótico poliquétido aromático producido por varias especies de *Pseudomonas* que suprimen las en-

fermedades de las plantas causadas por hongos fitopatógenos. Esta caracterización mostró el gran potencial derivado del estudio genético y señaló la necesidad de analizar la secuencia del genoma completo, lo que se logró casi una década después (Paulsen et al., 2005).

El genoma de Pf-5 tiene un tamaño de 7,07 Megabases (Mb), lo cual lo ubica como uno de los más grandes descritos para la especie. Esta dimensión refleja la adaptación al estilo de vida como comensal de otro organismo y muestra una buena capacidad de respuesta a los cambios ambientales. El genoma de Pf-5 tiene una extensa colección de genes reguladores, algunos de los cuales se han caracterizado por sus papeles en la regulación de la producción de metabolitos secundarios o en el control biológico. De manera consecuente con su estilo de vida comensal, Pf-5 parece carecer de una serie de factores de virulencia y de patogenicidad que se encuentran en los patógenos de las plantas. Además de las seis rutas de metabolitos secundarios conocidos producidos por Pf-5, tres nuevos clústeres podrían contribuir al control biológico.

Así, la secuencia genómica ha proporcionado numerosas pistas sobre los mecanismos utilizados por la bacteria para sobrevivir en la espermosfera y en la rizosfera. Estas características incluyen una amplia capacidad catabólica y de transporte para la utilización de exudados de semillas y raíces, una gran colección de sistemas de eflujo para la defensa contra el estrés ambiental y la competencia microbiana, y la presencia de 45 receptores de membrana externa que deberían permitir la absorción de hierro a partir de una amplia gama de sideróforos producidos por microorganismos del suelo (Paulsen et al., 2005). Estudios como este muestran el enorme potencial que posee la secuenciación genómica y su posterior interpretación bioinformática, para dilucidar los mecanismos eventualmente responsables de las actividades de control biológico.

Transcriptómica y control biológico

En la naturaleza, las plantas hospederas y sus comunidades microbianas circundantes tienden a adaptarse a medida que se producen cambios ambientales y

ecológicos. Cuando los microorganismos y las células individuales de los organismos superiores perciben, mediante la perturbación de receptores, las diferentes señales químicas o físicas del ambiente, desencadenan una respuesta funcional en la que intervienen principalmente los reguladores de transcripción y los actuadores apropiados. Los mecanismos asociados con la percepción, transmisión y procesamiento de la información ambiental se basan en complejas interacciones moleculares. Si bien nuestra comprensión de estas respuestas es limitada, es lógico suponer que la mala regulación en cualquiera de las etapas del procesamiento de la información ambiental podría conducir a resultados adversos para el organismo en cuestión.

Debido a esto, el perfil transcripcional de las respuestas orgánicas al medioambiente puede ser muy informativo, pues proporciona una plataforma para desentrañar las bases moleculares de la adaptación de los seres vivos.

Ningún organismo se desarrolla en un estado aislado, y las asociaciones entre las plantas y microorganismos son muy diversas: estas suceden en el suelo, en las superficies foliares y también dentro de la planta. Dichas interacciones pueden ser beneficiosas, neutrales o perjudiciales con respecto al crecimiento vegetal, la salud y el desarrollo; además, en los sistemas agrícolas, también afectan el rendimiento. Los avances recientes en las tecnologías de secuenciación han acelerado la capacidad de rastrear los cambios globales en las transcripciones y de identificar las intrincadas interacciones que pueden correlacionarse con los resultados fisiológicos de las relaciones planta-microbios.

Los enfoques de secuenciación NGS, tales como la secuenciación de ARN dual y la metatranscriptómica, permiten un análisis imparcial para identificar los genes expresados diferencialmente e incluso permiten identificar cambios en los eventos de regulación transcripcional. Massart et al. (2015) realizaron una revisión sobre estudios relevantes en el control biológico de fitopatógenos y la importancia de las aproximaciones ómicas en modelos *in vitro* e *in planta*. En los siguientes apartados de esta sección se destacan los estudios que han utilizado el análisis de perfiles de transcripción global para descubrir fenómenos naturales que conducen al control biológico tanto de patógenos microbianos como de insectos plaga.

Transcriptómica en el control de fitopatógenos

Las investigaciones relacionadas con las interacciones planta-patógeno son un campo en rápida expansión debido a las pérdidas de rendimiento que dichas interacciones pueden generar en la producción agrícola. Aunque la resistencia a las enfermedades de las plantas suele ser robusta en la protección que el genoma proporciona contra los invasores microbianos, las adaptaciones coevolutivas en los patógenos de las plantas pueden propiciar la aparición de enfermedades (Anderson et al., 2010).

Por ejemplo, el cultivo de *Brassica napus* (canola) se ve fuertemente afectado por el patógeno fúngico *Sclerotinia sclerotiorum*, agente causante de la pudrición del tallo de canola. En la actualidad, no existe un germoplasma resistente a esta enfermedad y su ataque depende de la existencia de genes para resistencia cuantitativa o incompleta (Margulies et al., 2006; Wu et al., 2013), lo que convierte en una prioridad el descubrimiento y caracterización de agentes de control biológico. Recientemente, Duke et al. (2017) describieron los efectos de *Pseudomonas chlororaphis* PA23 sobre *B. napus* en presencia de *S. sclerotiorum*, cuyas tasas de infección se redujeron en un 91 %. Este trabajo documentó la reprogramación transcripcional en canola mediante el análisis de ARN, después del tratamiento con el patógeno, el controlador biológico o la interacción de ambos. Así, demostraron que el tratamiento con *S. sclerotiorum* indujo la regulación positiva de aproximadamente 8.000 genes en la canola, mientras que el pretratamiento con el biocontrolador *P. chlororaphis* PA23 redujo 16 veces la expresión de los genes regulados (aproximadamente 500). En ausencia del patógeno, se observó que *P. chlororaphis* PA23 indujo redes de defensa en plantas y también reguló la resistencia sistémica adquirida y la producción de especies reactivas de oxígeno. Este trabajo proporcionó un marco para establecer las bases moleculares de la estimulación de resistencia en la canola con *P. chlororaphis* y abrió una ventana para usar la resistencia sistémica adquirida en el manejo de la enfermedad.

Otro ejemplo de análisis funcional de agentes de control biológico es el logrado en el control de la enfermedad poscosecha conocida como *moho azul*, la

cual es causada por el hongo filamentoso *Penicillium expansum* en frutos (manzanas y peras, por ejemplo). Durante el progreso de esta enfermedad, se produce la micotoxina patulina, que se asocia con efectos genotóxicos, teratogénicos e inmunotóxicos en los mamíferos, por lo cual es un riesgo para la salud de los que consuman las frutas infectadas (Moake et al., 2005; Puel et al., 2010).

El método actual para controlar las infecciones por *P. expansum* es el tratamiento con fungicidas. Pero debido a que la ingesta de estas sustancias también está relacionada con riesgos, el tratamiento con agentes de control biológico, como la levadura detoxificante de patulina *Sporobolomyces* sp., es una estrategia prometidora (Mahunu et al., 2016). Para obtener información sobre los mecanismos asociados con la degradación de la patulina, Ianiri et al. (2016) monitorearon los cambios transcripcionales de *Sporobolomyces* sp. (cepa IAM 13481) tras la incubación con patulina, mediante secuenciación de ARN. Los resultados mostraron que este tratamiento condujo a la producción de especies reactivas de oxígeno y a respuestas al estrés oxidativo. Además, documentaron la regulación transcripcional de genes asociados con procesos de oxidación-reducción y transporte durante la detoxificación y expulsión de la patulina de la célula. Estos autores también propusieron que la patulina se conjuga con el glutatión por proteínas del tipo glutatión S-transferasas, las cuales son reguladas positivamente antes de la carga por transportadores ABC específicos de patulina-glutatión para su degradación no específica en vacuolas. No obstante, a pesar de lo detallado y complejo de estos mecanismos, aún no se ha determinado la evidencia experimental de tales actividades. Sin embargo, la secuenciación de ARN es una poderosa herramienta para plantear rápidamente hipótesis conducentes a desarrollar de un mejor modo la experimentación en el laboratorio.

Dentro de otros casos, cabe mencionar los hallazgos con especies del género *Trichoderma* descritas como simbiontes avirulentos de plantas que funcionan como antagonistas de muchos hongos fitopatógenos, lo cual mejora, en consecuencia, la sanidad del vegetal. Este hongo —que tiene una amplia distribución mundial (Vinale et al., 2008)— incluye muchas especies que han sido bien documentadas tanto por sus propiedades de control biológico como por su actividad promotora del

crecimiento vegetal (Hermosa et al., 2012; Matarese et al., 2012; Ryder et al., 2012). Aunque se ha demostrado que varias cepas de este hongo son funcionales para uno u otro objetivo, la cepa GD12 de *Trichoderma hamatum* mostró tener ambas propiedades (Ryder et al., 2012).

En tal sentido, Shaw et al. (2016) realizaron un análisis de la progresión de las respuestas transcripcionales de este microorganismo mediante la secuenciación del ARN. La evaluación se realizó durante la promoción del crecimiento de la lechuga y el control biológico del patógeno *S. sclerotiorum*, durante 15 días, con 6 puntos de muestreo en el tiempo.

Este experimento se realizó en bandejas tipo semillero de seis pozos (macetas) con suelo, y los tratamientos fueron suelo con *T. hamatum*, suelo con *S. sclerotiorum* y suelo con *T. hamatum* y *S. sclerotiorum*. En todos los casos, cada pozo contenía una sola plántula de lechuga. La secuenciación de ARN se realizó los días 1, 2, 4, 5, 10 y 15. Los resultados mostraron que *T. hamatum* GD12 tiene un perfil transcripcional distinto cuando se cultiva en presencia de *S. sclerotiorum*, con expresión de genes que incluyen transportadores, factores de oxidación-reducción y una destacada representación de sideróforos. Además, durante el antagonismo de *S. sclerotiorum*, se encontró que un grupo de proteínas secretadas ricas en cisteína funcionaron como moléculas efectoras potenciales, las cuales podrían afectar el control biológico y la promoción del crecimiento de las plantas, también reguladas positivamente. Este estudio proporcionó un informe detallado de la cinética de la reprogramación transcripcional durante el control biológico y la promoción del crecimiento de las plantas, información que ha abierto una puerta a la caracterización funcional de rasgos microbianos de importancia agrícola.

Transcriptómica en el control de insectos

Además de los agentes de control biológico que se estudian y desarrollan para combatir los patógenos vegetales, también se buscan activamente microorganismos para el control de insectos dañinos para los cultivos. Las especies del hongo cosmopolita del suelo *Metarhizium* han sido reconocidas por su potencial

para controlar artrópodos. De hecho, Novozymes ha comercializado un producto deno-minado Met52, el cual es una formulación de la cepa F52 de *Metarhizium anisopliae* que tiene actividad insecticida, sin riesgos toxicológicos o restricciones de cosecha. En algunos países, el Met52 está incluso registrado para su uso en la agricultura ecológica. Algunos insectos blanco de Met52 incluyen varias garrapatas y escarabajos, gorgojos de las raíces, moscas, mosquitos, trips, langostas y saltamontes como hospederos.

En un informe de Gao et al. (2011), los enfoques de secuenciación de ARN se utilizaron para comparar las primeras respuestas transcripcionales de *M. anisopliae* y *Metarhizium acridum* durante la infección de langostas y cucarachas, respectivamente. El objetivo del estudio fue comprender mejor los eventos tempranos en la germinación de esporas de hongos, la adhesión a la cutícula de los insectos, la penetración y la formación de apresorios. Si bien existen algunas superposiciones en patrones de expresión, tales como transcripciones asociadas con traducción y modificaciones del tipo postraduccional, los dos hongos difirieron entre sí.

Los autores propusieron que las dos especies de *Metarhizium* spp. estudiadas tienen mecanismos finamente perfeccionados para regular la diferenciación celular en presencia de diferentes insectos y para reconocer diversos estímulos relacionados con las cutículas del hospedero. Estos eventos de reconocimiento utilizarían vías de señalización que implican receptores acoplados a proteína G, vías de señalización dependientes de calcio y MAP quinasas que dirigen la expresión de proteasas secretadas y grupos de genes asociados con la síntesis de metabolitos secundarios.

Se ha encontrado que *M. anisopliae* tiene una fuerte actividad biocontroladora de insectos plaga, sin embargo, existe evidencia de que langostas migratorias gregarias tienen mayor sobrevivencia en respuesta al tratamiento con *M. anisopliae* que aquellas que viven en solitario (Wilson et al., 2002). Para ampliar la comprensión de este fenómeno, Wang et al. (2013) llevaron a cabo experimentos de secuenciación de ARN a fin de resaltar la expresión diferencial que sustenta la longevidad de las langostas gregarias después de la infección con *M. anisopliae*. Al analizar los datos

de secuenciación, encontraron que las langostas solitarias susceptibles dedicaban casi el doble de la cantidad de genes para combatir el patógeno fúngico en comparación con sus pares gregarios. Por otra parte, las langostas gregarias exhibieron niveles elevados en la expresión de genes relacionados con la defensa, tales como receptores de reconocimiento, inhibidores de proteasas y factores antioxidantes. Por ejemplo, la proteína de reconocimiento GGBP3 se une a β - (1,3) glucanos, un componente importante de la pared celular de hongos, y activa las respuestas sistémicas antifúngicas del hospedero (Matskevich et al., 2011).

La caracterización posterior por Wang et al. (2013) reveló que GGBP3 aumenta la susceptibilidad a la proteólisis durante la infección por *M. anisopliae*, presumiblemente liberando PAMP (patrones moleculares asociados a patógenos). Estas son moléculas inductoras de respuestas de defensa en las plantas, que en este caso actúan contra las langostas. Por otro lado, se ha observado que el silenciamiento mediado por ARN de interferencia contra ARNm producidos por el gen GGBP3 reduce efectivamente la vida media de las langostas gregarias durante la infección con *M. anisopliae*, lo cual indica que la respuesta inmune dependiente de GGBP3 es crítica para la inducción de la resistencia en estos animales.

Conclusiones

Antes del advenimiento de las tecnologías de secuenciación del ADN, los genetistas y los biólogos moleculares se basaron en el cribado de grandes colecciones de organismos mutantes para identificar genes implicados en funciones biológicas específicas. Esto se logró a menudo mediante la inducción química de mutaciones aleatorias y la caracterización de los mutantes, seguidas por amplios procedimientos de clonación basada en mapas para aislar genes específicos implicados en el proceso de interés. En conjunto, estas técnicas tomaban años de duro trabajo.

Hoy en día, afortunadamente, mediante la ejecución de experimentos controlados de manera adecuada, seguidos por la extracción y secuenciación de ácidos nucleicos, se puede revelar rápidamente cuáles son los genes (la información) implicados en cada uno de los

fenómenos biológicos. Los beneficios de la transcripción son claros, dado que una vez identificados los genes críticos para un proceso biológico en particular, se pueden desarrollar tamizajes de alto rendimiento basados en la expresión de o en actividades asociadas con genes particulares.

A pesar de las grandes ventajas, los perfiles transcripcionales presentan desafíos, que incluyen la necesidad de confirmar experimentalmente las funciones de los genes identificados para ser considerados biológicamente relevantes en un proceso determinado. Además, este trabajo requiere de científicos con capacitación especial, cuya atención a la bioinformática, la estadística y la manipulación de grandes conjuntos de datos es fundamental. Sin embargo, la disponibilidad de bases de datos y de programas computacionales para la manipulación de la secuencia y su análisis estadístico puede facilitar el proceso, lo cual hace accesible la herramienta para muchos científicos de todo el mundo.

Proteómica y control biológico

En los seres vivos, las proteínas son las moléculas protagonistas, encargadas de llevar a cabo las principales funciones celulares. Esta particularidad se debe, sin duda, a la diversidad de aminoácidos que conforman su alfabeto. De hecho, la diferente composición secuencial y la longitud de los polipéptidos naturales determinan la conformación de estructuras complejas y diversas, cuyos residuos constitutivos exponen sus distintos grupos químicos, modifican los entornos y otorgan oportunidades de interacción con grados diferenciales de afinidad con otras moléculas y átomos.

Esta característica composicional de las proteínas hace que puedan intervenir en numerosas y variadas actividades (incluso variables de acuerdo con las condiciones de contorno y la presencia de otras proteínas). También hace posible el establecimiento de vínculos estrechos con diferentes biomoléculas o iones metálicos para conformar vías metabólicas, rutas de señalización o estructuras complejas, entre otras posibilidades que han evolucionado en la naturaleza. Es de destacar que, en muchos organismos, un solo tipo de proteína individual (o su cantidad relativa)

es el responsable de un rasgo fenotípico útil para una aplicación biotecnológica. En cambio, en otras situaciones, la proteína solo es un miembro de una red diversa que, en su trabajo coordinado y conjunto, define el aspecto morfológico o fisiológico de interés científico.

Por tales razones, el estudio proteómico es esencial para la postulación y posterior formulación de agentes de biocontrol en agricultura, dado que puede ayudar en la comprensión de los mecanismos de interés.

Proteómica en el control de fitopatógenos

Existen reportes sobre el uso de herramientas proteómicas encaminadas a elucidar los mecanismos de acción de los agentes de control biológico de fitopatógenos. Estas técnicas han sido aplicadas para identificar proteínas asociadas a la función biocontroladora y así comprender y determinar cuáles son los procesos involucrados. Un ejemplo reciente en esta línea de estudios es el trabajo llevado a cabo con el micoparásito *Aspergillus aculeatus* Asp-4, un agente de biocontrol que coloniza y degrada esclerocios de *S. sclerotiorum*, reduce la germinación de estas estructuras y, por ende, la enfermedad causada por este patógeno que afecta alrededor de 400 especies de plantas en todo el planeta (Hu et al., 2017).

Además de emplear metodologías transcriptómicas, los autores aplicaron un estudio proteómico consistente en la extracción de proteínas totales a partir *A. aculeatus* Asp-4 cultivados en presencia o ausencia de esclerocios triturados de *S. sclerotiorum*, para luego proceder a su separación mediante electroforesis bidimensional y a la posterior identificación de los polipéptidos diferenciales empleando espectroscopía de masas del tipo MALDI-TOF. Estos estudios comparativos de los perfiles proteicos permitieron determinar que 93 proteínas habían mostrado un incremento en su expresión cuando el micoparásito había crecido sobre los esclerocios triturados (quitinasas, proteasas, lipasas, porinas, proteínas del metabolismo energético y oxidoreductasas, entre otras), mientras que 73 habían disminuido su presencia (proteínas de respuesta a estreses bióticos y abióticos). De este modo, la información proteómica mostró indicios claros de cuáles son los mecanismos que se activan cuando el agente controlador afecta al fitopatógeno.

Se encuentran también otros casos relevantes en el empleo de tecnologías proteómicas. Se destaca, por ejemplo, la aplicación de estas metodologías para comprender mejor el modo de acción de los agentes biocontroladores de aplicación en poscosecha sobre la superficie de las frutas. En tal sentido, las levaduras antagonistas son microorganismos utilizados con estos fines, dado que se ha visto su efecto benéfico mediado por diversos factores como la competencia por los nutrientes, la exclusión de nicho y la inducción en el fruto de respuestas de defensa ante los patógenos (Spadaro & Droby, 2016).

Zhang et al. (2017), por ejemplo, estudiaron los efectos de *Yarrowia lipolytica* (levadura antagonista) sobre manzanas para evitar el desarrollo de la enfermedad del moho azul causada por *Penicillium expansum* (Calvo et al., 2007), un importante productor de patulina, que, como se mencionó anteriormente es una micotoxina con actividad citotóxica, genotóxica e inmunosupresora (Puel et al., 2010). Para revelar cuáles mecanismos podrían estar implicados en los beneficios que otorgan levaduras y bacterias, los autores compararon los perfiles proteicos derivados de heridas en manzanas en presencia o ausencia de *Y. lipolytica*, mediante aislamiento de proteínas totales, fraccionamiento por electroforesis bidimensional e identificación posterior con espectroscopía de masas en tándem y el *software* de búsqueda Mascot. Así, mostraron que 12 proteínas se encontraban más expresadas, mientras que otras 23 disminuían su cantidad cuando la levadura estaba presente. De las primeras, destacaron la presencia polipéptidos de defensa y de respuesta al estrés, algunos de ellos previamente descritos en otros sistemas o por métodos diferentes. *Y. lipolytica* estimuló la actividad de polifenoloxidasas, peroxidasa, quitinasa y L-fenilalanina amonio liasa, todas involucradas en el aumento de respuestas de defensa en la manzana. De este modo, entender cuáles son las proteínas que defienden al fruto de *P. expansum* ayudará en la selección de variedades vegetales más resistentes, así como en la selección de nuevos agentes de biocontrol que estimulen vías similares.

Proteómica en el control de insectos

En el control de insectos plaga también se han aplicado las técnicas de investigación en proteómica.

Recientemente, un acercamiento desde este campo disciplinar ha sido empleado en el estudio de una cepa altamente proteolítica de *B. bassiana*, entomopatógeno útil para el control de *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae), considerada la plaga más limitante en el cultivo de tomate (Borgi et al., 2016). En este trabajo, la cepa P2 (un mutante espontáneo de la cepa de *B. bassiana* denominada P1) se caracterizó por mostrar una significativa producción de actividades proteasa y quitinasa, las cuales tienen capacidad degradadora de la cutícula de los insectos, propiedad que la convertía en un mejor agente de biocontrol.

En particular, los autores analizaron las proteínas extracelulares derivadas de cultivos de seis días de las cepas P1 y P2, las cuales fueron recuperadas y separadas por electroforesis bidimensional, para luego, a partir de las señales diferenciales, aislar los polipéptidos, fragmentarlos y analizarlos mediante un nano-HPLC en capilares acoplado a un espectrómetro de masas. Luego, mediante la utilización de diferentes tipos de *software* y análisis estadísticos, concluyeron que ambas cepas producían polipéptidos similares, en su gran mayoría, enzimas hidrolíticas, pero con una abundancia diferencial. Entre ellas, se destacó una proteasa similar a la subtilisina, que evidenció cinco isoformas distintas, de las cuales una se expresaba igual en ambas cepas, mientras que otras dos fueron exclusivas de P2. Los resultados de este trabajo mostraron que el perfil de enzimas extracelulares, especialmente de las proteasas, podría considerarse como una huella dactilar que refleja la patogenicidad de los hongos hacia los insectos y, en consecuencia, este patrón podría ser utilizado para la selección de cepas con potencial en control biológico.

Conclusiones

Como se discutió antes, los beneficios otorgados por las técnicas ómicas para una mejor comprensión de los agentes de control biológico de fitopatógenos y de insectos plaga en la agricultura son incuestionables. Cabe aquí resaltar que, si bien el análisis genómico revela todo el potencial que posee un agente destinado para estos fines —dado que a partir de su aplicación es posible inferir los factores de virulencia que contiene (entre otras características deseables)—, es necesario e imprescindible también abordar las

dimensiones transcriptómicas y proteómicas para evidenciar cuándo y en qué medida esos factores y sus vías asociadas se expresan.

Estas variaciones dependen de regulaciones pre- y pos-transcripcionales y traduccionales que son consecuencia de numerosas variables que aún son difíciles de determinar mediante el análisis de secuencias génicas únicamente. Pequeñas variaciones en los promotores, los UTR (*untranslated regions* o 'regiones no traducidas') y otros subcomponentes genéticos, así como la presencia de genes que expresan moléculas de ARN reguladoras, entre otras múltiples causas, pueden hacer que dos genomas con una gran similitud en sus secuencias expresen fenotipos claramente diferenciales. En función de ello, el análisis de los transcriptos es muy útil, dado que facilita el muestreo tamizado para identificar cuáles genes se estarían expresando ante una situación dada. Sin embargo, dicho análisis no revela información de forma directa sobre las cantidades relativas de las actividades expresadas (téngase presente que principalmente son ejercidas por proteínas) ni sobre sus variaciones producto de sutiles modificaciones sufridas por los polipéptidos y que alteran sus funciones. Es en este punto que los estudios proteómicos revelan un valor único y superior a los anteriores, dado que permiten identificar con gran precisión cuáles son los factores que una entidad expresa ante una situación dada, ya sea el patógeno o su hospedero, y, en consecuencia, las causas de los fenotipos observables, como la virulencia o la capacidad de defensa, respectivamente.

Metabolómica y control biológico

Los microorganismos y los virus que habitan un cierto espacio natural se vinculan de múltiples maneras en complejas redes de interacción. En estos sistemas, cada individuo de una especie dada mostrará un fenoma particular, en virtud de las potencialidades que exprese su genoma y a causa de los vínculos que establezca con los otros miembros presentes de su población (el pangenoma que allí se encuentre), con los representantes de las otras especies (visualizable por el metagenoma) y por las condiciones físicas y químicas del ambiente donde se halle (que modularán

en modo significativo cada uno de sus fenotipos). Tal es el grado de complejidad que se requiere de un estudio exhaustivo en distintos niveles informativos, para la comprensión del funcionamiento global de un sistema biológico en estas escalas y, así, poder plantear modelos matemáticos que permitan su comprensión, sostenimiento y proyección.

En tal camino, la metabolómica surge como una capa importante de información biológica que es deseable y necesario explorar, puesto que a partir de ella es posible generar una lectura eficiente de las consecuencias que los estreses ambientales o las interacciones biológicas (incluyendo las relaciones patógeno-hospedero) causan sobre la materia viva. Estos enfoques son centrales para una correcta selección y formulación de agentes de control biológico, ya que otorgan evidencias útiles y permiten identificar moléculas interesantes para la sanidad vegetal.

Metabolómica en el control de fitopatógenos

Las relaciones tripartitas comprendidas por las plantas, sus fitopatógenos y los organismos benéficos han sido muy estudiadas con el fin de promoverlas y usarlas para optimizar el éxito de los sistemas agrícolas. En general, los considerados como "organismos benéficos" son las entidades responsables de reducir en los vegetales los efectos negativos causados por los fitopatógenos y de promover otras respuestas positivas. De todas estas vías funcionales producidas a partir de las interacciones múltiples, se derivan numerosos metabolitos, cuya determinación puede desentrañar información útil para la comprensión global de los ecosistemas, pero, sobre todo, para desarrollar o mejorar bioproductos que les permitan un mayor grado de resiliencia y de productividad a los cultivos. En dicha dirección se han realizado numerosos estudios metabolómicos.

Trichoderma es un género de hongos de amplia distribución mundial, cuyas distintas especies han sido descritas como simbioses avirulentos de plantas, que funcionan como parásitos y antagonistas de muchos hongos fitopatógenos, y que mejoran, en consecuencia, la sanidad del vegetal con el que interactúan (Vinale et al., 2008). Este hongo produce una gran cantidad

de metabolitos secundarios asociados a las funciones que expresa, lo que incluye antibióticos y otros compuestos con diversas actividades útiles para controlar fitopatógenos. En virtud de ello, los estudios metabolómicos son útiles para seleccionar las mejores cepas de *Trichoderma* spp.

Kang et al. (2011), por ejemplo, evaluaron los perfiles de metabolitos secundarios de 33 cepas pertenecientes a 7 especies de este género, para lo cual utilizaron cromatografía líquida acoplada a espectroscopía de masas (LC-ESI-MS-MS) y métodos estadísticos multivariantes. Así, detectaron diversos compuestos como ácido hetelídico, sorbicilinol, trichodermanona C, ácido gicládico, bisorbicilinol y otras sustancias, algunas de las cuales fueron producidas exclusivamente por ciertas cepas. Lo importante de este estudio es que los autores confrontaron un análisis genético con uno químico derivado del estudio metabolómico, para así inferir la clasificación taxonómica de las cepas estudiadas. De hecho, encontraron que la quimiotaxonomía basada en metabolitos secundarios mostró un mayor poder discriminador que el estudio basado en ITS (*Internal Transcribed Spacers*).

Por otro lado, la exploración metabolómica no solo ayuda en la identificación de cepas destacadas, sino que permite el descubrimiento de nuevos compuestos con gran valor en agricultura y en otras aplicaciones biotecnológicas. Recientemente, fue hallado y caracterizado un estereoisómero del ácido arziánico, denominado ácido isoharziánico, producido por una cepa de *Trichoderma* (Vinale et al., 2014). Esta sustancia demostró inhibir el crecimiento de los fitopatógenos *S. sclerotiorum* y *R. solani*, además de mejorar la germinación de semillas de tomate e inducir resistencia en las plantas. De este modo, la metabolómica muestra su gran utilidad en la optimización de estrategias de control contra fitopatógenos.

Metabolómica en el control de insectos

Los insectos herbívoros también establecen interesantes relaciones multipartitas, que incluyen tanto a la planta que consumen como a los entomopatógenos que los afectan. En este contexto, los insectos pueden inducir respuestas de defensa que actúan contra

sus entomopatógenos o, por el contrario, la planta puede defenderse de sus predadores invertebrados promoviendo el desarrollo de microorganismos biocontroladores del insecto, generalmente, a través de la emisión de compuestos orgánicos volátiles (COV) que desencadenan las diferentes respuestas. Para comprender estas interacciones y el diálogo molecular detrás de ellas, es de gran importancia identificar los metabolitos secundarios producidos por cada uno de los protagonistas de estas relaciones, ya que aquellos evidencian las vías empleadas por cada entidad y, a partir de allí, ofrecen una guía para optimizar los procesos deseados y evitar los procesos perjudiciales en los sistemas agrícolas (Beck et al., 2014; Shikano, 2017).

Se ha visto, por ejemplo, que las plantas expresan un perfil de metabolitos diferente según el agente que les provoca un daño. Por ejemplo, el metaboloma de ejemplares de *Solanum lycopersicon* (tomate) se altera de maneras significativamente diferentes cuando la planta es atacada por *Spodoptera littoralis* frente a lo que sucede cuando ocurre un daño mecánico (Zebelo et al., 2012). También existen reportes que muestran que los microbiomas del suelo impactan el metaboloma de la planta y afectan el comportamiento alimentario de los insectos herbívoros. Particularmente, Badri, Zolla, Bakker, Manter, & Vivanco (2013) estudiaron los efectos de diferentes consorcios microbianos presentes en el suelo de plantas de la especie modelo *Arabidopsis thaliana* y su influencia sobre larvas fitófagas de *Trichoplusia ni*. Para su análisis, los compuestos fueron extraídos a partir de diferentes tejidos de las plantas que habían sido expuestas a los distintos tratamientos, para luego ser analizados mediante cromatografía gaseosa acoplada a espectroscopía de masas. Así, las sustancias fenólicas y los aminoácidos, alcoholes y azúcares mostraron variaciones significativas según el microbioma presente, lo cual estuvo correlacionado, en algunos casos, con la disminución en el consumo de material vegetal por parte de las larvas.

En síntesis, la mirada metabolómica puede ser de gran utilidad para el perfeccionamiento de los sistemas agropecuarios, dado que hace posible identificar numerosos disparadores químicos que podrían utilizarse en forma directa para potenciar la eficacia de agentes de control biológico de insectos plaga.

Conclusiones

De la información que se ha presentado se concluye que ninguna técnica ómica puede otorgar por sí sola toda la información que se requiere para comprender y optimizar un sistema biológico, aunque cada una de ellas ofrece informaciones relevantes que pueden ser directamente aprovechadas sin necesidad de su integración con las otras dimensiones. Por ejemplo, el conocimiento genómico permite hacer un seguimiento preciso de una entidad y, a partir de estudios proteómicos, es posible identificar factores proteicos útiles para diversas actividades humanas. Estos datos se pueden producir en sistemas heterólogos y llegar a ser útiles *per se*. Del mismo modo, la metabolómica no solo es valiosa porque dé indicios firmes sobre las vías funcionales que un organismo vivo se encuentre ejecutando en una condición dada (la cual incluye sus relaciones multipartitas), sino porque también puede ofrecer una colección de compuestos naturales con diferentes actividades (útiles en la agricultura, la biomedicina, etc.). Una vez determinada la composición de estas moléculas, se pueden sintetizar químicamente y explotarse como productos en diferentes campos industriales.

Perspectivas y desafíos conceptuales

En la era posgenómica varios aspectos merecen la atención de los investigadores para la selección y formulación de los microorganismos biocontroladores. Es de esperar que, en un futuro cercano, las tecnologías ómicas continúen con un desarrollo exponencial a un ritmo impredecible y proporcionen evidencias que contribuyan al conocimiento holístico de los sistemas biológicos que se desean intervenir. Este enfoque integrador es particularmente interesante para identificar características específicas y, a veces, inesperadas de los posibles agentes de control biológico y de sus relaciones tróficas con los otros organismos del ecosistema en estudio (Massart et al., 2015). Las investigaciones en esta área aplicarán cada vez más la diversidad de tecnologías ómicas disponibles y otras que serán desarrolladas, dado que son funcionales para proporcionar un mejor conocimiento sobre los modos de acción y sobre aspectos ecológicos de los

agentes de control. Sin duda, el conocimiento derivado del uso de estas tecnologías guiará el desarrollo de bioproductos más robustos y de mayor eficiencia que los que existen hoy en el mercado.

Hasta ahora, los enfoques ómicos aplicados a la investigación en control biológico se han llevado a cabo principalmente de manera individual o secuencial, y difícilmente se ha efectuado una integración de estos, en la que yace, tal vez, una sus mayores potencialidades. De hecho, la combinación de la información ómica de manera multiescala podría colaborar en aspectos predictivos sobre la dinámica temporal y espacial de un modelo biológico. Como recientemente proponen varios autores (Barah & Bones, 2014; Yugi et al., 2016), nos aproximamos a un escenario de estudios y proyecciones transómicas, consistente en la integración de la información ómica individual recolectada mediante el empleo de tecnologías de alto rendimiento asociadas a cada dimensión biológica abordable (genómica, epigenómica, transcriptómica, proteómica, interactómica y metabolómica), para comprender, a nivel molecular y con un enfoque integral, el funcionamiento de los sistemas naturales (figura 22.5).

Este análisis integrado brinda la posibilidad de comprender los organismos y los virus, más el conjunto de sus relaciones tróficas, como complejos algoritmos computacionales que pueden ser útiles para revelar aspectos fenomícos trascendentes con enorme impacto en diferentes áreas tecnológicas, incluyendo los agroecosistemas. Un ejemplo en esta dirección se ha logrado con la producción de vinos, dado que una misma variante de vid se desarrolla y produce frutos con una expresión diferencial en función del *terroir*, de las prácticas de cultivo empleadas y de las otras entidades biológicas presentes en el lugar (Fabres et al., 2017). De este modo, la mirada transómica puede ser funcional para el mejor aprovechamiento de un sistema biológico de interés.

Por otra parte, las tecnologías ómicas requieren un mayor desarrollo de conocimientos y habilidades para recolectar información útil y evitar miradas subjetivas en su análisis. Los sesgos pueden introducirse en cualquier etapa del proceso, ya sea en el diseño experimental y preparación de las muestras, ya en los análisis posteriores. Por ejemplo, actualmente no existe un protocolo estándar para el procesamiento

bioinformático de los datos generados por las herramientas ómicas, factor que se convierte en el cuello de botella en la aplicación de estas tecnologías. Las líneas de análisis en bioinformática pueden variar de acuerdo con el *software* elegido y con sus parámetros de aplicación, así como con las bases de datos que se empleen para hacer comparaciones. De esta forma, cuando las estrategias son demasiado astringentes, puede suceder que se deseche información útil; mientras que en líneas de análisis muy relajadas es posible, por el contrario, que se generen falsos positivos.

En estos enfoques, por tanto, es altamente recomendable que se realice un análisis cuidadoso de la información global, para luego proponer metodologías que confirmen individualmente los aspectos seleccionados como útiles para la caracterización del sistema. Por ejemplo, si se detecta en un estudio determinado la presencia de transcritos, proteínas o metabolitos diferenciales como signo de una interacción en evaluación, cada una de estas moléculas debería ser luego analizada mediante aproximaciones individuales para confirmar las evidencias previas, empleando métodos como la PCR, la RT-PCR o el Western Blot.

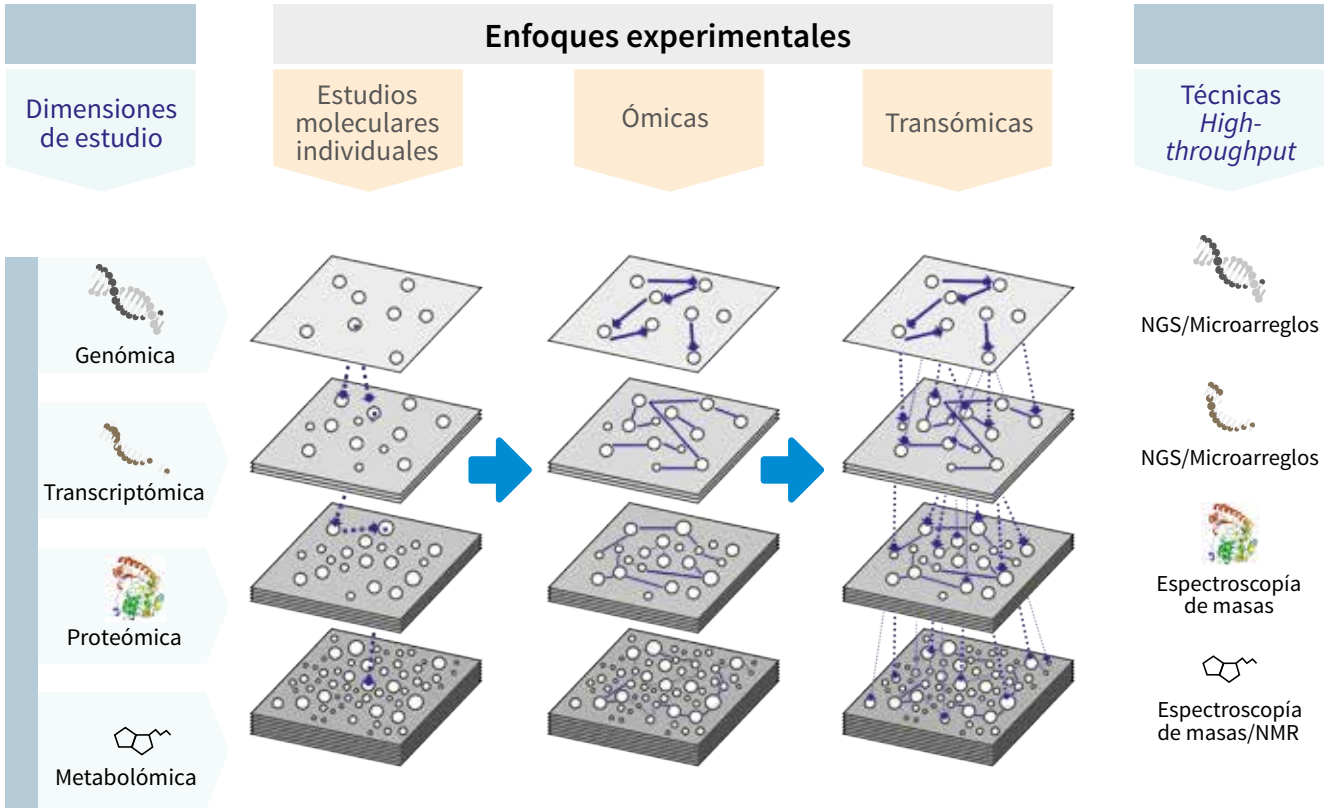


Figura 22.5. Enfoques holísticos en biología. Los fenotipos integrados de un organismo (fenoma) pueden ser abordados mediante su análisis en dimensiones separadas: genómica, transcriptómica, proteómica y metabolómica (aunque existen otras ventanas biológicas informativas, como la epigenómica, la interactómica, etc.). En dichas aproximaciones experimentales se pueden realizar estudios moleculares individuales (como PCR, RT-PCR, secuenciación automática Sanger, Southern/Northern/Western Blot, HPLC, etc.), ómicos (se destacan en la imagen las tecnologías de alto rendimiento más utilizadas hoy en día) y transómicos. Esta última estrategia busca integrar todas las biomoléculas en redes funcionales. Ahora bien, cada una de las dimensiones ómicas está representada por capas, que son múltiples en aquellas ventanas biológicas que son dinámicas por las condiciones de contorno (transcriptómica, proteómica y metabolómica). Las capas contienen círculos blancos que indican las biomoléculas (para el ADN los círculos referencian las secuencias aisladas no concatenadas). Las flechas azules con líneas punteadas representan probables caminos metabólicos en la célula. Las flechas azules con líneas continuas indican los conjuntos de biomoléculas detectadas mediante un enfoque *high-throughput*.

Fuente: Adaptado de Yugi et al. (2016)

También se deberían utilizar herramientas de ingeniería genética, como la interrupción de genes o su sobreexpresión, para apoyar con mayor solidez los conocimientos generados. De esta forma, la complejidad del análisis de la información ómica, más las consideraciones con respecto a las fuentes potenciales de sesgo, hacen necesaria una formación profunda y multidisciplinar del investigador en control biológico o una estrecha colaboración para este por parte de profesionales expertos en áreas emergentes como la bioinformática (Massart et al., 2015).

Otro elemento clave para lograr el éxito de una investigación en esta área con la aplicación de tecnologías ómicas es el diseño de un modelo experimental adecuado que responda satisfactoriamente a las preguntas científicas planteadas. Por ejemplo, muchos estudios transcriptómicos y proteómicos sobre agentes de biocontrol no virales comparan su comportamiento (en condiciones *in vitro*) sujeto a la variación en las fuentes de carbono para su alimentación. En el caso de los virus, también se realizan comparaciones mediante ensayos de exposición de los hospederos (criados en laboratorio) al agente en cuestión (también producido en el laboratorio mediante diferentes estrategias), en entornos distantes a los naturales. En general, estos experimentos son fáciles de llevar a cabo y de controlar en condiciones de laboratorio, pero pueden producir resultados sesgados en comparación con las condiciones de campo, por lo cual pueden resultar poco informativos para mejorar el empleo tecnológico de estos recursos en la agricultura.

Por otra parte, la interpretación biológica de los datos puede ser en apariencia contradictoria. Como ejemplo, Mgbeahuruike et al. (2013) encontraron diferentes evidencias sobre el efecto de las moléculas secretadas por el agente de biocontrol *Phlebiopsis gigantea* sobre el perfil de expresión de genes del patógeno *Heterobasidion annosum* en sentido amplio, que dependen de la etapa de crecimiento del controlador y el momento de interacción. Los genes que codifican proteínas de defensa como las hidrofobinas mostraron una baja expresión, mientras que aquellos relacionados con la percepción de la señal o la tolerancia al estrés fueron altamente expresados. Esto resalta la complejidad subyacente en los mecanismos asociados al control biológico y la importancia de encontrar un diseño experimental adecuado con tiempos de muestreo apropiados.

Como ya se ha mencionado, la secuenciación del genoma de un ACB probablemente sea uno de los primeros pasos en su caracterización. Esta acción puede proporcionar una columna vertebral para apoyar estudios adicionales, cualesquiera que sean las tecnologías y metodologías que se utilicen. Como una aproximación a este objetivo es posible encarar secuenciaciones parciales, lo que incluye segmentos informativos clave para reconocer especies o biotipos, más aquellos asociados al rol biocontrolador deseado, como factores de virulencia y de interacción con los hospederos. Sin embargo, es previsible proyectar que en un futuro cercano se solicite la secuencia genómica completa de un ACB para su registro, dada la disminución de costos y de tiempos asociados al uso de las tecnologías NGS y a la existencia de numerosos procedimientos estandarizados.

Para un mejor desarrollo de las tecnologías ómicas también es importante que se incremente su interacción con las otras metodologías vigentes que, en general, son reconocidas como de “bajo rendimiento” por su rol individual en el estudio de biomoléculas y organismos. Así, es recomendable que el inicio de la caracterización de un ACB con valor potencial para su empleo en campo deba llevarse a cabo mediante la ejecución de metodologías tradicionales, para luego introducir el enfoque ómico (pangenómico, transcriptómico, proteómico, metabolómico o de otra índole). Se recomienda entonces que las primeras hipótesis que se formulen asociadas a la exploración del valor biocontrolador de un ACB estén apoyadas por evidencias microbiológicas, bioquímicas y de imágenes logradas por microscopía.

Con este valioso conocimiento preliminar se puede aplicar un modelo experimental más apropiado para luego planificar el uso de los enfoques ómicos, útiles para aportar un mayor grado de precisión en la caracterización del sistema en estudio, precisión tendiente a su mejor aprovechamiento y potenciación.

En síntesis, es factible visualizar un futuro cercano en el cual las técnicas ómicas serán una parte obligada para el desarrollo de los agentes de control biológico, dado el enorme potencial que la aplicación de estas ha evidenciado en otras áreas científico-tecnológicas y el que están demostrando en la agricultura, como se describió en este capítulo.

Agradecimientos

Los autores agradecen a Alba Marina Cotes por la invitación a participar en este libro. Pablo Daniel Ghiringhelli y Mariano Nicolás Belaich agradecen a la Universidad Nacional de Quilmes, a Conicet y a la Agencia Nacional de Promoción Científica y Tecnológica de Argentina por financiar sus investigaciones en agrobiotecnología. Gloria Barrera, Stephen Mosher y Alejandro Caro agradecen al Ministerio de Agricultura y Desarrollo Rural de Colombia, a Colciencias y a AGROSAVIA.



Referencias

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., ... Galle, R. F. (2000). The genome sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185-2195. doi:10.1126/science.287.5461.2185.
- Aguilar-Bultet, L., & Falquet, L. (2015). Secuenciación y ensamblaje de novo de genomas bacterianos: una alternativa para el estudio de nuevos patógenos. *Revista de Salud Animal*, 37(2), 125-132.
- Ambardar, S., Gupta, R., Trakroo, D., Lal, R., & Vakhlu, J. (2016). High throughput sequencing: An overview of sequencing chemistry. *Indian Journal of Microbiology*, 56(4), 394-404. doi:10.1007/s12088-016-0606-4.
- Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., & Singh, K. B. (2010). Plants versus pathogens: an evolutionary arms race. *Functional Plant Biology*, 37(6), 499-512. doi:10.1071/FP09304.
- Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), 796-815. doi:10.1038/35048692.
- Badri, D. V., Zolla, G., Bakker, M. G., Manter, D. K., & Vivanco, J. M. (2013). Potential impact of soil microbiomes on the leaf metabolome and on herbivore feeding behavior. *New Phytologist*, 198(1), 264-273. doi:10.1111/nph.12124.
- Barah, P., & Bones, A. M. (2014). Multidimensional approaches for studying plant defence against insects: from ecology to omics and synthetic biology. *Journal of Experimental Botany*, 66, 479-493. doi:10.1093/jxb/eru489.
- Barrera, G. P., Belaich, M. N., Patarroyo, M. A., Villamizar, L. F., & Ghiringhelli, P. D. (2015). Evidence of recent interspecies horizontal gene transfer regarding nucleopolyhedrovirus infection of *Spodoptera frugiperda*. *BMC Genomics*, 16, 1008. doi:10.1186/s12864-015-2218-5.
- Beck, J. J., Smith, L., & Baig, N. (2014). An overview of plant volatile metabolomics, sample treatment and reporting considerations with emphasis on mechanical damage and biological control of weeds. *Phytochemical Analysis*, 25(4), 331-341. doi:10.1002/pca.2486.
- Bentley, D. R. (2006). Whole-genome re-sequencing. *Current Opinion in Genetics and Development*, 16(6), 545-552. doi:10.1016/j.gde.2006.10.009.
- Bizzini, A., & Greub, G. (2010). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. *Clinical Microbiology and Infection*, 16(11), 1614-1619. doi:10.1111/j.1469-0691.2010.03311.x.
- Borgi, I., Dupuy, J. W., Blibech, I., Lapaillerie, D., Lomenech, A. M., Rebai, A., ... Gargouri, A. (2016). Hyper-proteolytic mutant of *Beauveria bassiana*, a new biological control agent against the tomato borer. *Agronomy for Sustainable Development*, 36, 60. doi:10.1007/s13593-016-0394-6.
- Brozynska, M., Copetti, D., Furtado, A., Wing, R. A., Crayn, D., Fox, G., ... Henry, R. J. (2017). Sequencing of Australian wild rice genomes reveals ancestral relationships with domesticated rice. *Plant Biotechnology Journal*, 15(6), 765-774. doi:10.1111/pbi.12674.
- Brun, S., Madrid, H., Gerrits Van Den Ende, B., Andersen, B., Marinach-Patrice, C., Mazier, D., & De Hoog, G. S. (2013). Multilocus phylogeny and MALDI-TOF analysis of the plant pathogenic species *Alternaria dauci* and relatives. *Fungal biology*, 117(1), 32-40. doi:10.1016/j.funbio.2012.11.003.
- Calvo, J., Calvente, V., De Orellano, M. E., Benuzzi, D., & Tosetti, M. I. S. (2007). Biological control of postharvest spoilage caused by *Penicillium expansum* and *Botrytis cinerea* in apple by using the bacterium *Rahnella aquatilis*. *International Journal of Food Microbiology*, 113(3), 251-257. doi:10.1016/j.ijfoodmicro.2006.07.003.
- Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., ... Krummenacker, M. (2009). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, 40(Database issue), D742-D753. doi:10.1093/nar/gkr1014.
- Chang, H.-X., Lipka, A. E., Domier, L. L., & Hartman, G. L. (2016). Characterization of disease resistance loci in the USDA Soybean Germplasm Collection using Genome-wide association studies. *Phytopathology*, 106(10), 1139-1151. doi:10.1094/PHYTO-01-16-0042-FI.
- Charlesworth, B., & Charlesworth, D. (2009). Darwin and genetics. *Genetics*, 183(3), 757-766. doi:10.1534/genetics.109.109991.
- Chou, H., Pathmasiri, W., Deese-Spruill, J., Sumner, S., & Buchwalter, D. B. (2017). Metabolomics reveal physiological changes in mayfly larvae (*Neocloeon triangulifer*) at ecological upper thermal limits. *Journal of Insect Physiology*, 101, 107-112. doi:10.1016/j.jinsphys.2017.07.008.
- Conesa, A., Madrigal, P., Tarazona, S., Gómez-Cabrero, D., Cervera, A., McPherson, A., ... Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17, 13-19. doi:10.1186/s13059-016-0881-8.
- Covington, B. C., McLean, J. A., & Bachmann, B. O. (2016). Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites. *Natural Product Reports*, 34(1), 1-19. doi:10.1039/C6NP00048G.
- Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227, 561-563.
- Crovadore, J., Xu, S., Chablais, R., Cochard, B., Lukito, D., Calmin, G., & Lefort, F. (2017). Metagenome-Assembled Genome Sequence of *Rhodopseudomonas palustris* Strain ELI 1980, Commercialized as a Biostimulant. *Genome Announcements*, 5(18), 1-3. doi:10.1017/CBO9781107415324.004.
- Cuartas, P. E., Barrera, G. P., Belaich, M. N., Barreto, E., Ghiringhelli, P. D., & Villamizar, L. F. (2015). The complete sequence of the first *Spodoptera frugiperda* betabaculovirus genome: a natural multiple recombinant virus. *Viruses*, 7(1), 394-421. doi:10.3390/v7010394.
- Douglas, C. J., & Ehrling, J. (2005). *Arabidopsis thaliana* full genome longmer microarrays: A powerful gene discovery tool for agriculture and forestry. *Transgenic Research*, 14(5), 551-561. doi:10.1007/s11248-005-8926-x.
- Duke, K. A., Becker, M. G., Girard, I. J., Millar, J. L., Fernando, W. G. D., Belmonte, M. F., & De Kievit, T. R. (2017). The biocontrol agent *Pseudomonas chlororaphis* PA23 primes *Brassica napus* defenses through distinct gene networks. *BMC Genomics*, 18, 1-16. doi:10.1186/s12864-017-3848-6.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., ... Turner, S. (2009). Single polymerase molecules. *Science*, 323(5910), 133-138. doi:10.1126/science.1162986.
- Evans, A., Forde, N., O'Gorman, G., Zielak, A., Lonergan, P., & Fair, T. (2008). Use of microarray technology to profile gene expression patterns important for reproduction in cattle. *Reproduction in Domestic Animals*, 43(s2), 359-367. doi:10.1111/j.1439-0531.2008.01185.x.
- Fabres, P. J., Collins, C., Cavagnaro, T. R., & Rodríguez López, C. M. (2017). A concise review on Multi-Omics data integration for terroir analysis in *Vitis vinifera*. *Frontiers in Plant Science*, 8, 1065. doi:10.3389/fpls.2017.01065.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., ... Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269(5223), 496-512. doi:10.1126/science.7542800.
- Gao, Q., Jin, K., Ying, S. H., Zhang, Y., Xiao, G., Shang, Y., ... Wang, C. (2011). Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genetics*, 7(1), e1001264. doi:10.1371/journal.pgen.1001264.
- Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 11(5), 759-769. doi:10.1111/j.1755-0998.2011.03024.x.
- Gogna, N., Singh, V. J., Sheeba, V., & Dorai, K. (2015). NMR - based investigation of the *Drosophila melanogaster* metabolome under the influence of daily cycles of light and temperature. *Molecular BioSystems*, 11(12), 3305-3315. doi:10.1039/c5mb00386e.
- Greenbaum, D., Luscombe, N. M., Jansen, R., Qian, J., & Gerstein, M. (2001). Interrelating different types of genomic data, from proteome to secretome: coming in on function. *Genome Research*, 11(9), 1463-1468. doi:10.1101/gr.207401.necessary.
- Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., ... Xie, Z. (2008). Single-molecule DNA sequencing of a viral genome. *Science*, 320(5872), 106-109. doi:10.1126/science.1150427.
- Hasin, Y., Seldin, M., Lusi, A., Guo, W., Go, J., Shi, H., ... Landray, M. (2017). Multi-omics approaches to disease. *Genome Biology*, 18, 83. doi:10.1186/s13059-017-1215-1.
- Hermosa, R., Viterbo, A., Chet, I., & Monte, E. (2012). Plant-beneficial effects of *Trichoderma* and of its genes. *Microbiology*, 158(Pt1), 17-25. doi:10.1099/mic.0.052274-0.
- Hinrichs, J. W. J., Van Blokland, W. T. M., Moons, M. J., Radersma, R. D., Van Loon, J. H. R., De Voij, C. M. A., ... De Weger, R. A. (2015). Comparison of next-generation sequencing and mutation-specific platforms in clinical practice. *American Journal of Clinical Pathology*, 143(4), 573-578. doi:10.1309/AJCP40XETVYAMJPY.
- Howell, C. R., & Stipanovic, R. D. (1979). Control of *Rhizoctonia solani* on cotton seedlings with *Pseudomonas fluorescens* and with an antibiotic produced by the bacterium. *Phytopathology*, 69, 480-482.
- Howell, C. R., & Stipanovic, R. D. (1980). Suppression of *Pythium ultimum*-induced damping-off of cotton seedlings by *Pseudomonas fluorescens* and its antibiotic, pyoluteorin. *Phytopathology*, 70, 712-715.

- Hsu, J. C., Chien, T. Y., Hu, C. C., Chen, M. J. M., Wu, W. J., Feng, H. T., ... Chen, C. Y. (2012). Discovery of genes related to insecticide resistance in *Bactrocera dorsalis* by functional genomic analysis of a *De novo* assembled transcriptome. *PLoS One*, 7(8), doi:10.1371/journal.pone.0040950.
- Hu, X., Qin, L., Roberts, D. P., Lakshman, D. K., Gong, Y., Maul, J. E., ... Liao, X. (2017). Characterization of mechanisms underlying degradation of sclerotia of *Sclerotinia sclerotiorum* by *Aspergillus aculeatus* Asp-4 using a combined qRT-PCR and proteomic approach. *BMC Genomics*, 18(1), 674. doi:10.1186/s12864-017-4016-8.
- Ianiri, G., Idnurm, A., & Castoria, R. (2016). Transcriptomic responses of the basidiomycete yeast *Sporobolomyces* sp. to the mycotoxin patulin. *BMC Genomics*, 17, 210. doi:10.1186/s12864-016-2550-4.
- Kang, D., Kim, J., Choi, J. N., Liu, K.-H., & Lee, C. H. (2011). Chemotaxonomy of *Trichoderma* spp. using mass spectrometry-based metabolite profiling. *Journal of Microbiology and Biotechnology*, 21(1), 5-13.
- Kazemi-Pour, N., Condemine, G., & Hugouvieux-Cotte-Pattat, N. (2004). The secretome of the plant pathogenic bacterium *Erwinia chrysanthemi*. *Proteomics*, 4(10), 3177-3186. doi:10.1002/pmic.200300814.
- Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7), 2567-2572. doi:10.1073/pnas.0409727102.
- Kraus, J., & Loper, J. E. (1995). Characterization of a genomic region required for production of the antibiotic pyoluteorin by the biological control agent *Pseudomonas fluorescens* Pf-5. *Applied and Environmental Microbiology*, 61(3), 849-854.
- Lande, N. V., Subba, P., Barua, P., Gayen, D., Keshava Prasad, T. S., Chakraborty, S., & Chakraborty, N. (2017). Dissecting the chloroplast proteome of chickpea (*Cicer arietinum* L.) provides new insights into classical and non-classical functions. *Journal of Proteomics*, 165, 11-20. doi:10.1016/j.jjprot.2017.06.005.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., ... Fitz-Hugh, W. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921 doi:10.1038/35057062.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., ... Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, 2012, 11. doi:10.1155/2012/251364.
- Liu, Y., & Li, X. (2015). Darwin and Mendel today: a comment on “Limits of imagination: the 150th Anniversary of Mendel’s Laws, and why Mendel failed to see the importance of his discovery for Darwin’s theory of evolution”. *Genome*, 59(1), 75-77. doi:10.1139/gen-2015-0155.
- Lorito, M., Woo, S. L., Harman, G. E., & Monte, E. (2010). Translational research on *Trichoderma*: from omics to the field. *Annual Review of Phytopathology*, 48, 395-417. doi:10.1146/annurev-phyto-073009-114314.
- Mahunu, G. K., Zhang, H., Yang, Q., Li, C., & Zheng, X. (2016). Biological control of patulin by antagonistic yeast: A case study and possible model. *Critical Reviews in Microbiology*, 42(4), 643-655. doi:10.3109/1040841X.2015.1009823.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., ... Rothberg, J. M. (2006). Genome sequencing in open microfabricated high density picoliter reactors. *Nature*, 437(7057), 376-380. doi:10.1038/nature03959.
- Marzancola, M. G., Sedighi, A., & Li, P. C. H. (2016). DNA Microarray-Based Diagnostics. En P. C. H. Li, A. Sedighi, & L. Wang, (Eds.), *Microarray Technology: Methods and Applications* (pp. 161-178). Nueva York, EE. UU.: Springer.
- Massart, S., Perazzolli, M., Höfte, M., Pertot, I., & Jijakli, M. H. (2015). Impact of the omic technologies for understanding the modes of action of biological control agents against plant pathogens. *BioControl*, 60(6), 725-746. doi:10.1007/s10526-015-9686-z.
- Matarese, F., Sarrocco, S., Gruber, S., Seidl-Seiboth, V., & Vannacci, G. (2012). Biocontrol of Fusarium head blight: Interactions between *Trichoderma* and mycotoxigenic *Fusarium*. *Microbiology*, 158(Pt1), 98-106. doi:10.1099/mic.0.052639-0.
- Matskevich, A. A., Quintin, J., & Ferrandon, D. (2011). Pathway Activation Function. *European Journal of Immunology*, 40(5), 1244-1254. doi:10.1002/eji.200940164.
- McKernan, K. J., Peckham, H. E., Costa, G., McLaughlin, S., Tsung, E., Fu, Y., ... Blanchard, A. P. (2009). Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation. *Genome Research*, 19(9), 1527-1541. doi:10.1101/gr.091868.109.
- Mgbeahuruike, A. C., Kohler, A., & Asiegbu, F. O. (2013). Expression analysis of the impact of culture filtrates from the biocontrol agent, *Phlebiopsis gigantea* on the conifer pathogen, *Heterobasidion annosum* ss transcriptome. *Microbial Ecology*, 66(3), 669-681. doi:10.1007/s00248-013-0255-7.
- Miller, H. I. (2017). Genetic engineering applied to agriculture has a long row to hoe. *GM Crops & Food*, 0, 1-4. doi:10.1080/21645698.2017.1378840.
- Mitter, E. K., De Freitas, J. R., & Germida, J. J. (2017). Bacterial root microbiome of plants growing in oil sands reclamation covers. *Frontiers in Microbiology*, 8, 849. doi:10.3389/fmicb.2017.00849.
- Moake, M. M., Padilla-Zakour, O. I., & Worobo, R. W. (2005). Comprehensive review of patulin control methods in foods. *Comprehensive Reviews in Food Science and Food Safety*, 4(1), 8-21. doi:10.1111/j.1541-4337.2005.tb00068.x.
- Moraes, F., & Góes, A. (2016). A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. *Biochemistry and Molecular Biology Education*, 44(3), 215-223. doi:10.1002/bmb.20952.
- Moshelion, M., & Altman, A. (2015). Current challenges and future perspectives of plant and agricultural biotechnology. *Trends in Biotechnology*, 33(6), 337-342. doi:10.1016/j.tibtech.2015.03.001.
- Najafabadi, A. S., Naghavi, M. R., Farahmand, H., & Abbasi, A. (2017). Transcriptome and metabolome analysis of *Ferula gummosa* Boiss. to reveal major biosynthetic pathways of galbanum compounds. *Functional & Integrative Genomics*, 17(6), 1-13. doi:10.1007/s10142-017-0567-7.
- Nawaz, M. A., Rehman, H. M., Baloch, F. S., Ijaz, B., Ali, M. A., Khan, I. A., ... Yang, S. H. (2017). Genome and transcriptome-wide analyses of cellulose synthase gene superfamily in soybean. *Journal of Plant Physiology*, 215, 163-175. doi:10.1016/j.jplph.2017.04.009.
- Newman, M. M., Lorenz, N., Hoilett, N., Lee, N. R., Dick, R. P., Liles, M. R., ... Kloepper, J. W. (2016). Changes in rhizosphere bacterial gene expression following glyphosate treatment. *The Science of the Total Environment*, 553, 32-41. doi:10.1016/j.scitotenv.2016.02.078.
- Oshlack, A., Robinson, M. D., & Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome Biology*, 11, 220. doi:10.1186/gb-2010-11-12-220.
- Paulsen, I. T., Press, C. M., Ravel, J., Kobayashi, D. Y., Myers, G. S. A., Mavrodi, D. V., ... Madupu, R. (2005). Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nature Biotechnology*, 23(7), 873. doi:10.1038/nbt1110.
- Puel, O., Galtier, P., & Oswald, I. P. (2010). Biosynthesis and toxicological effects of patulin. *Toxins*, 2(4), 613-631. doi:10.3390/toxins2040613.
- Qu, M., An, B., Shen, S., Zhang, M., Shen, X., Duan, X., ... Qu, J. (2016). Qualitative and quantitative characterization of protein biotherapeutics with liquid chromatography mass spectrometry. *Mass Spectrometry Reviews*, 36(6), 734-754. doi:10.1002/MAS.21500.
- Rodríguez-Llorente, I., Caviedes, M. A., Dary, M., Palomares, A. J., Cánovas, F. M., & Peregrín-Álvarez, J. M. (2009). The Symbiosis Interactome: a computational approach reveals novel components, functional interactions and modules in *Sinorhizobium meliloti*. *BMC Systems Biology*, 3, 63. doi:10.1186/1752-0509-3-63.
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., ... Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475, 348-352. doi:10.1038/nature10242.
- Rozado-Aguirre, Z., Adams, I., Fox, A., Dickinson, M., & Boonham, N. (2017). Complete sequence and genomic annotation of carrot torradovirus 1. *Archives of Virology*, 162(9), 2815-2819. doi:10.1007/s00705-017-3410-5.
- Ryder, L. S., Harris, B. D., Soanes, D. M., Kershaw, M. J., Talbot, N. J., & Thornton, C. R. (2012). Saprotrophic competitiveness and biocontrol fitness of a genetically modified strain of the plant-growth-promoting fungus *Trichoderma hamatum* GD12. *Microbiology*, 158(Pt 1), 84-97. doi:10.1099/mic.0.051854-0.
- Sanger, F., Coulson, A., Friedmann, T., Air, G., Barrell, B., Brown, N., ... Smith, M. (1977a). The nucleotide sequence of bacteriophage φX174DNA. *Nature*, 265, 687-695. doi:10.1038/265687a0.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977b). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences of the United States of America*, 74(12), 5463-5467. doi:10.1073/pnas.74.12.5463.
- Shaw, S., Le Cocq, K., Paszkiewicz, K., Moore, K., Winsbury, R., De Torres Zabala, M., ... Grant, M. R. (2016). Transcriptional reprogramming underpins enhanced plant growth promotion by the biocontrol fungus *Trichoderma hamatum* GD12 during antagonistic interactions with *Sclerotinia sclerotiorum* in soil. *Molecular Plant Pathology*, 17(9), 1425-1441. doi:10.1111/mpp.12429.
- Shikano, I. (2017). Evolutionary ecology of multitrophic interactions between plants, insect herbivores and entomopathogens. *Journal of Chemical Ecology*, 43(6), 1-13. doi:10.1007/s10886-017-0850-z.
- Spadaro, D., & Droby, S. (2016). Development of biocontrol products for postharvest diseases of fruit: The importance of elucidating the mechanisms of action of yeast antagonists. *Trends in Food Science and Technology*, 47, 39-49. doi:10.1016/j.tifs.2015.11.003.
- Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G., & Bayley, H. (2009). Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proceedings of the National Academy of Sciences of the United States of America*, 106(19), 7702-7707. doi:10.1073/pnas.0901054106.
- Tan, B. C., Lim, Y. S., & Lau, S.-E. (2017). Proteomics in commercial crops: An overview. *Journal of Proteomics*, 169, 30185-30189. doi:10.1016/j.jjprot.2017.05.018.

Thao, N. P., & Tran, L.-S. (2016). Enhancement of plant productivity in the post-genomics era. *Current Genomics*, 17(4), 295-296. doi:10.2174/138920291704160607182507.

The *C. elegans* Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, 282(5396), 2012-2018. doi:10.1126/science.282.5396.2012.

Tyczewska, A., Gracz, J., Kuczyński, J., & Twardowski, T. (2016). Deciphering the soybean molecular stress response via high-throughput approaches. *Acta Biochimica Polonica*, 63(4), 631-643. doi:10.18388/abp.2016_1340.

Udaondo, Z., Duque, E., & Ramos, J.-L. (2017). The pangenome of the genus *Clostridium*. *Environmental Microbiology*, 19(7), 2588-2603. doi:10.1111/1462-2920.13732.

Van Lenteren, J. C., Bolckmans, K., Köhl, J., Ravensberg, W. J., & Urbaneja, A. (2017). Biological control using invertebrates and microorganisms: plenty of new opportunities. *BioControl*, 63(1), 1-21. doi:10.1007/s10526-017-9801-4.

Vidova, V., & Spacil, Z. (2017). A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Analytica Chimica Acta*, 964, 7-23. doi:10.1016/j.aca.2017.01.059.

Vinale, F., Manganiello, G., Nigro, M., Mazzei, P., Piccolo, A., Pascale, A., ... Lanzuise, S. (2014). A novel fungal metabolite with beneficial properties for agricultural applications. *Molecules*, 19(7), 9760-9772. doi:10.3390/molecules19079760.

Vinale, F., Sivasithamparam, K., Ghisalberti, E. L., Marra, R., Barbeti, M. J., Li, H., ... Lorito, M. (2008). A novel role for *Trichoderma* secondary metabolites in the interactions with plants. *Physiological and Molecular Plant Pathology*, 72(1-3), 80-86. doi:10.1016/j.pmpp.2008.05.005.

Wang, Y., Yang, P., Cui, F., & Kang, L. (2013). Altered immunity in crowded locust reduced fungal (*Metarhizium anisopliae*) pathogenesis. *PLoS Pathogens*, 9(1), e1003102. doi:10.1371/journal.ppat.1003102.

Wei, N., Bemmels, J. B., & Dick, C. W. (2014). The effects of read length, quality and quantity on microsatellite discovery and primer development: From Illumina to PacBio. *Molecular Ecology Resources*, 14(5), 953-965. doi:10.1111/1755-0998.12245.

Wilson, K., Thomas, M. B., Blanford, S., Doggett, M., Simpson, S. J., & Moore, S. L. (2002). Coping with crowds: density-dependent disease resistance in desert locusts. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8), 5471-5475. doi:10.1073/pnas.082461999.

Williams, H. L., Monge-Monge, K. S., Orvos, I. S., Reardon, R., & Ragenovich, I. (2011). Genotypic variation among Douglas-fir tussock moth nucleopolyhedrovirus (OpNPV) isolates in the western United States. *Journal of Invertebrate Pathology*, 108(1), 13-21. doi:10.1016/j.jip.2011.06.004.

Wu, J., Cai, G., Tu, J., Li, L., Liu, S., Luo, X., ... Zhou, Y. (2013). Identification of QTLs for resistance to *Sclerotinia* stem rot and BnaC.IGMT5.a as a candidate gene of the major resistant QTL SRC6 in *Brassica napus*. *PLoS One*, 8(7), e67740. doi:10.1371/journal.pone.0067740.

Xie, W., Chen, C., Yang, Z., Guo, L., Yang, X., Wang, D., ... Zhang, Y. (2017). Genome sequencing of the sweetpotato whitefly *Bemisia tabaci* MED/Q. *GigaScience*, 6(5), 1-7. doi:10.1093/gigascience/gix018.

Yugi, K., Kubota, H., Hatano, A., & Kuroda, S. (2016). Trans-omics: how to reconstruct biochemical networks across multiple 'omic' layers. *Trends in Biotechnology*, 34(4), 276-290. doi:10.1016/j.tibtech.2015.12.013.

Zebelo, S. A., Matsui, K., Ozawa, R., & Maffei, M. E. (2012). Plasma membrane potential depolarization and cytosolic calcium flux are early events involved in tomato (*Solanum lycopersicon*) plant-to-plant communication. *Plant Science*, 196, 93-100. doi:10.1016/j.plantsci.2012.08.006.

Zhang, H., Chen, L., Sun, Y., Zhao, L., Zheng, X., & Yang, Q. (2017). Investigating proteome and transcriptome defense response of apples induced by *Yarrowia lipolytica*. *Molecular Plant-Microbe Interactions Journal*, 30(4), 301-311. doi:10.1094/MPMI-09-16-0189-R.

Zhao, L., Huang, Y., Adeleye, A. S., & Keller, A. A. (2017). Metabolomics reveals Cu(OH)₂ nanopesticide activated antioxidative pathways and decreased beneficial antioxidants in spinach leaves. *Environmental Science and Technology*, 51(17), 10184-10194. doi:10.1021/acs.est.7b02163.