

UNIVERSIDAD POLITÉCNICA DE VALENCIA

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA Y DEL MEDIO NATURAL



Análisis integrado del metaboloma y del proteoma de plantas de tomate infectadas con *Pseudomonas syringae*

TRABAJO FINAL DE GRADO EN BIOTECNOLOGÍA

Autor. Jesús Gutiérrez Botella

Tutora. Susana Barceló Cerdá

Curso académico 2017/2018

Valencia, julio de 2018



Análisis integrado del proteoma y del metaboloma de plantas de tomate infectadas con *Pseudomonas syringae*

Autor. Jesús Gutiérrez Botella

Tutora académica. Susana Barceló Cerdá

Resumen.

Pseudomonas syringae es una bacteria causante de muchas enfermedades en el reino vegetal. Este patógeno tiene un amplio espectro de huéspedes y sus síntomas se manifiestan en las hojas, con unas manchas acuosas que pueden necrosar, frecuentemente rodeadas por un halo amarillo. La infección, en caso de que sea compatible, sucede cuando el patógeno penetra en la planta a través de estomas y heridas y es capaz de multiplicarse.

El tomate de la variedad Rio Grande tiene el gen *Pto*, que ofrece resistencia al gen de avirulencia *avrPto* y *avrPtoB* presentes en *P. syringae*. Gracias a un reconocimiento específico entre los genes *Pto* de la planta y *avrPto* del patógeno, la infección es incompatible y la planta es capaz de defenderse antes de desarrollar los síntomas de la infección. Sin embargo, cuando se produce una delección en los genes *avrPto* de la bacteria, este reconocimiento ya no sucede y la planta sufre los síntomas y desarrolla la enfermedad. Este sistema de reconocimiento gen a gen ofrece un buen modelo para comparar las similitudes y diferencias que comparten ambos tipos de infecciones, compatible e incompatible, a nivel de transcriptoma, metaboloma o proteoma.

En este trabajo, mediante un análisis integrado del metaboloma y del proteoma de las plantas del tomate, se caracterizará la respuesta defensiva de esta planta frente a la infección por *P. syringae*. Todo ello, utilizando tres sistemas: (i) control, sin ningún tipo de interacción planta-patógeno; (ii) infección compatible, con una cepa de *P. syringae* con una delección de sus genes *avrPto*, y (iii) infección incompatible, con la cepa salvaje de *P. syringae*.

Para la identificación de los componentes implicados en la respuesta defensiva se parte de datos de concentración de los metabolitos y de las proteínas presentes en cada sistema. Un análisis estadístico multivariante para cada tipo de datos y un análisis integrado permitirá conocer los metabolitos y las proteínas con mayor relevancia en la respuesta frente a la infección.

Por tanto, el objetivo de este estudio es proporcionar un método estadístico que permita conocer: (1) si existen diferencias significativas en el proteoma y el metaboloma de cada tipo de planta, (2) si las proteínas que se expresan diferencialmente en cada tipo de planta están relacionadas directamente con la respuesta que ofrece la planta a la infección, (3) qué metabolitos o rutas metabólicas emplea la planta para defenderse ante la infección por *P. syringae*, y (4) de qué forma es capaz la planta de defenderse del patógeno integrando en un solo análisis los datos metabolómicos y proteómicos.

Palabras clave. Metabolómica, proteómica, tomate, *Pseudomonas*, integración.

València, julio de 2018



Integrated analysis of the proteome and the metabolome of tomato plants infected by *Pseudomonas syringae*.

Author. Jesús Gutiérrez Botella

Academic director. Susana Barceló Cerdá

Abstract.

Pseudomonas syringae is a bacterium source of several diseases in the Kingdom *Plantae*. This pathogen has a broad hosts spectrum and its symptoms are visible in the leaves, showing aqueous stains that can eventually necrose, frequently surrounded by a yellow halo. Infection, in case this is compatible, occurs when the pathogen penetrates inside the plant through the stomas and wounds and it's able to multiply.

Rio Grande tomato variety possess the *Pto* gene, which offers resistance against the virulence genes *avrPto* and *avrPtoB*, found in *P. syringae*. Due to a specific recognition between *Pto* genes in the plant and *avrPto* genes in the pathogen, the infection is incompatible and the plant is capable of defending itself prior to develop infection symptoms. However, when a deletion in bacterium's *avrPto* genes occurs, this recognition no longer happens and the plant suffers from these symptoms, therefore developing the disease. This gen-to-gen recognition system supposes a useful model to compare the similarities and differences that both compatible and incompatible infections share regarding the transcriptome, metabolome or proteome.

In this work, through an integrated analysis of the metabolome and proteome of tomato plants, the tomato defensive response against *P. syringae* will be characterized. All of this will be identified by using three systems: (i) mock, absence of plant-pathogen interaction; (ii) compatible infection, with a *P. syringae* strain that owns a deletion in their *avrPto* genes, and (iii) incompatible infection, with the wild *P. syringae* strain.

So as to identify the implicated components in the response, we will start off the data of the concentration of all the proteins and metabolites that are present in each type of plant. From a multivariate statistical analysis of each kind of data and an integrated analysis, we will get to know what metabolites and proteins have the major relevance in the response against the infection.

Thus, the aim of this study is to provide a statistical method to identify: (1) whether significant differences among the metabolome of each type of plant exist, (2) whether the proteins that are differentially expressed in each type of plant are directly related with the response offered by the plant against the infection, (3) what specific metabolites or metabolic pathways are used by the plant in order to defend itself against the infection produced by *P. syringae*, and (4) in which way the plant defends itself through an integrated analysis of the metabolomic and proteomic data.

Keywords. Metabolomics, Proteomics, *Pseudomonas*, tomato, integration.

València, July 2018



Agradecimientos

En primer lugar, quiero agradecer a Susana toda la ayuda, la implicación y el esfuerzo que ha depositado en mi trabajo. Muchas gracias porque lo que he aprendido contigo ha sido mucho más que lo que se puede leer en estas páginas.

También me gustaría agradecer al IBMCP su colaboración por cederme los datos con los que se ha llevado a cabo el trabajo. Y a Mapi, por toda la ayuda que me ha prestado con la visualización y la interpretación de los resultados.

Además, me gustaría agradecer a mis compañeras y compañeros de Biotecnología, especialmente a Ana, a Mónica y a Ileana, por su amistad y su apoyo durante estos cuatro años. No desistir en algún momento de este camino ha sido mucho más fácil gracias a vosotras.

Por último, a mi familia y a mis amigos, por estar conmigo en cada paso y en cada decisión. Sin vuestro apoyo, todo esto habría sido imposible.

Índice

1. Introducción.....	1
1.1. La genómica funcional. Proteómica y metabolómica.....	1
1.1.1. La proteómica.....	1
1.1.2. La metabolómica.....	2
1.2. Interacciones planta-patógeno.....	2
1.2.1. El reconocimiento del patógeno: el modelo zig-zag.....	3
1.2.2. Las respuestas defensivas de la planta frente a los patógenos.....	3
1.2.3. Algunos sistemas planta-patógeno. El sistema tomate- <i>Pseudomonas syringae</i> pv tomato.....	4
1.3. Análisis estadístico multivariante para datos ómicos.....	5
1.4. Integración de ómicas.....	6
1.5. Estudios ómicos dirigidos. Análisis de la varianza (ANOVA).....	6
2. Objetivos.....	7
3. Materiales y métodos.....	8
3.1. Obtención de datos de metabolitos y proteínas.....	8
3.1.1. Material biológico.....	8
3.1.2. Inoculación de la bacteria.....	8
3.1.3. Extracción de proteínas.....	8
3.1.4. Extracción de metabolitos.....	9
3.2. Estudio estadístico mediante Análisis de Componentes Principales (PCA).....	9
3.3. Estudio estadístico mediante Regresión por Mínimos Cuadrados Parciales (PLS).....	11
3.4. Integración proteómica-metabolómica mediante un análisis <i>Orthogonal two-way Projections to Latent Structures</i> (O2PLS).....	12
3.5. Identificación y representación gráfica de metabolitos y proteínas relacionados con la respuesta defensiva de la planta a la infección.....	14
3.6. Estudio metabolómico dirigido a compuestos potencialmente importantes en la respuesta de la planta mediante un análisis de la varianza (ANOVA).....	14

4. Resultados y discusión del análisis metabolómico	16
4.1. Análisis de los metabolitos polares con carga negativa.....	16
4.2. Análisis de los metabolitos polares con carga positiva	18
4.3. Análisis de los metabolitos apolares	21
4.4. Estudio dirigido a metabolitos apolares importantes mediante un ANOVA.....	21
4.4.1. Lípidos	22
4.4.2. Carotenoides	22
4.4.3. Clorofilas.....	24
4.4.4. Quinonas	25
5. Resultados y discusión del análisis proteómico	27
6. Resultados y discusión de la integración de proteómica y metabolómica.....	30
6.1. Integración de proteínas y metabolitos apolares	30
6.2. Integración de proteínas y metabolitos polares	33
7. Conclusiones	36
8. Referencias	37

Índice de figuras y tablas

Introducción

Figura 1.	Modelo zig-zag en la infección de una planta por un patógeno.....	3
Figura 2.	Esquema de la hipótesis “gen a gen”	4
Tabla 1.	Ejemplos de genes R y sus respectivos genes Avr en diversos sistemas planta-patógeno.....	5

Materiales y métodos

Tabla 2.	Resumen de las muestras tomadas para cada análisis	9
-----------------	----------------------------------------------------------	---

Resultados y discusión del análisis metabolómico

Figura 3.	Gráfico de puntuaciones del PCA obtenido con el análisis de los metabolitos polares con carga negativa	16
Figura 4.	Gráfico de cargas del PCA obtenido con el análisis de los metabolitos polares con carga negativa.....	17
Tabla 3.	Lista de metabolitos polares con carga negativa.....	17
Figura 5.	Estructura del ácido clorogénico (CGA)	18
Figura 6.	Gráfico de puntuaciones del PCA obtenido con el análisis de los metabolitos polares con carga positiva	19
Figura 7.	Gráfico de cargas del PCA obtenido con el análisis de los metabolitos polares con carga positiva	19
Tabla 4.	Lista de metabolitos polares con carga positiva	20
Figura 8.	Respuesta de la planta a la infección compatible por <i>P. syringae</i> según los resultados de la metabolómica	21
Tabla 5.	Resultados del ANOVA para el análisis dirigido a lípidos.....	22
Tabla 6.	Resultados del ANOVA para el análisis dirigido a carotenoides.....	23
Tabla 7.	Resultados del análisis <i>post-hoc</i> para el b-caroteno	23
Tabla 8.	Resultados del ANOVA para el análisis dirigido a clorofilas	24
Tabla 9.	Resultados del análisis <i>post-hoc</i> para la clorofila a y el pheophorbide b.....	25
Tabla 10.	Resultados del ANOVA para el análisis dirigido a quinonas.....	26

Tabla 11.	Resultados del análisis <i>post-hoc</i> para la menaquinone 8.....	26
------------------	--------------------------------------------------------------------	----

Resultados y discusión del análisis proteómico

Figura 9.	Gráfico de puntuaciones del PCA obtenido con el análisis del proteoma de las plantas de tomate	27
Figura 10.	Gráfico de cargas del PCA obtenido con el análisis del proteoma de las plantas de tomate.....	28
Tabla 12.	Lista de proteínas	29

Resultados y discusión de la integración de proteómica y metabolómica

Figura 11.	Gráfico de puntuaciones del PLS de integración obtenido con el análisis conjunto del proteoma (X) y del metaboloma (Y) de las plantas de tomate	30
Figura 12.	Gráfico de puntuaciones del O2PLS de integración obtenido con el análisis conjunto del proteoma (X) y de la fracción apolar del metaboloma (Y) de las plantas de tomate	31
Tabla 13.	Variación desglosada explicada por cada componente del O2PLS de integración de metabolitos apolares y proteínas	32
Tabla 14.	Lista de metabolitos apolares y proteínas	32
Figura 13.	Gráfico de cargas del O2PLS de integración obtenido con el análisis conjunto del proteoma (X) y de la fracción apolar del metaboloma (Y) de las plantas de tomate	33
Tabla 15.	Variación desglosada explicada por cada componente del O2PLS de integración de metabolitos polares y proteínas	33
Figura 14.	Gráfico de puntuaciones del O2PLS de integración obtenido con el análisis conjunto del proteoma (X) y de la fracción polar del metaboloma (Y) de las plantas de tomate	34
Figura 15.	Gráfico de cargas del O2PLS de integración obtenido con el análisis conjunto del proteoma (X) y de la fracción polar del metaboloma (Y) de las plantas de tomate	35
Tabla 16.	Lista de metabolitos polares y proteínas	35

Abreviaturas

ANOVA	Análisis de la varianza
AOC	<i>Allene oxide cyclase</i>
APCI	Ionización química a presión atmosférica
Avr	Gen o factor de avirulencia
<i>avrPto</i>	Gen de avirulencia de <i>Pseudomonas syringae</i> , reconocido por el gen R (<i>Pto</i>) de la planta de tomate.
Δ<i>avrPto</i>	Delección en el gen de avirulencia <i>avrPto</i> de <i>Pseudomonas syringae</i> .
CFU	Unidades formadoras de colonias
CGA	Ácido clorogénico
CT	<i>Coumaroyltyramine</i>
DIGE	<i>Difference gel electrophoresis</i>
ESI	Ionización por electrospray
ETI	<i>Effector-triggered immunity</i>
ETS	<i>Effector-triggered susceptibility</i>
HPLC	Cromatografía líquida de alta presión
HR	Respuesta hipersensible.
HSD	<i>Honestly-significant difference</i> (test <i>post-hoc</i> realizado tras un ANOVA)
JA	Ácido jasmónico
LysoPE	<i>Lysophosphatidylethanolamine</i>
MALDI-TOF	Desorción/ionización láser asistida por matriz acoplada a un detector de iones de tiempo de vuelo.
MS	Espectrometría de masas
MZ	Relación masa:carga
O2PLS	<i>Orthogonal two-way Projections to Latent Structures</i>
OEE	<i>Oxygen evolving enhancer protein</i>
OPLS	<i>Orthogonal Projections to Latent Structures</i>
OSC	<i>Orthogonal Signal Correction</i>
PAMP	<i>Pathogen Associated Molecular Patterns</i>
PCA	<i>Principal Component Analysis</i> (análisis de componentes principales)
PCx / CPx	Componente principal x ($x = 1, 2, 3, \dots$)
PGK	<i>Phosphoglycerate kinase</i>
PLS	<i>Partial Least Squares regression</i> (regresión por mínimos cuadrados parciales)

PR	Proteínas relacionadas con la patogénesis
PRR	<i>Pattern Recognition Receptors</i>
Pst DC3000	<i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000
PTI	<i>PAMP-triggered immunity</i>
R	Gen de resistencia de la planta
R²_X (p. correlación)	Variación de la matriz X debida a la correlación entre las matrices X e Y en un modelo O2PLS.
R²_X	Variación total de la matriz X en un modelo O2PLS.
R²_Y (p. correlación)	Variación de la matriz Y debida a la correlación entre las matrices X e Y en un modelo O2PLS.
R²_Y	Variación total de la matriz Y en un modelo O2PLS.
RG-Pto	Tomate de la variedad Rio Grande, con el gen de resistencia <i>Pto</i> .
ROS	Especies reactivas del oxígeno
SA	Ácido salicílico
SAR	Resistencia sistémica adquirida
THT	<i>Tyramine N-hydroxycinnamoyl transferase</i>
t_R	Tiempo de retención
UV	Varianza unitaria (normalización)

1.

INTRODUCCIÓN

Análisis integrado del metaboloma y del proteoma de plantas de tomate infectadas con *Pseudomonas syringae*

1. Introducción

1.1. La genómica funcional. Proteómica y metabolómica

Desde la publicación en 1995 de la primera secuencia genética completa de un organismo, correspondiente a *H. influenzae* (FLEISCHMANN R, *et al.*, 1995), se inició la revolución genómica que permitió obtener secuencias completas de muchos organismos. Desde este hito y gracias al desarrollo de numerosas tecnologías de secuenciación masiva, se ha conseguido abaratar los costes de secuenciación de tal manera que ya están secuenciadas y almacenadas en bases de datos las secuencias genómicas de un gran número de especies (COOK CE, *et al.*, 2015).

Sin embargo y pese al potencial de la genómica, no permite conocer la modulación de las respuestas de los organismos ante determinadas situaciones. Esta limitación ha supuesto que surjan otras ómicas que se engloban dentro de la genómica funcional, como la transcriptómica, la proteómica o la metabolómica.

1.1.1. La proteómica

El término “proteómica” se acuñó en 1995, y se definió en ese momento como la caracterización a gran escala de las proteínas de una línea celular, tejido u organismo (WASINGER VC, *et al.*, 1995). La proteómica se refiere al estudio global de las proteínas a nivel de composición, estructura, función o incluso interacciones entre complejos proteicos.

El proteoma de un organismo es muy variable a lo largo del tiempo y responde a estímulos del microambiente y del macroambiente.

Tradicionalmente, para el análisis de las proteínas se emplea la electroforesis en dos dimensiones: una de isoelectroenfoque, en la que las proteínas se separan según su punto isoelectroenfoque; y otra de separación por peso molecular. Esta técnica hace que sea sencillo caracterizar cada *spot* como una proteína (MIKKELSEN S, *et al.*, 2004).

En los últimos años se han desarrollado las técnicas de espectrometría de masas, siendo la más utilizada para las proteínas el MALDI-TOF (desorción/ionización láser asistida por matriz acoplada a un detector de iones de tiempo de vuelo) que, combinada con la electroforesis 2D, permite caracterizar cada *spot* del gel (WU KJ, *et al.*, 1994).

Los dos tipos de análisis permiten la informatización y ya existe software bioinformático especializado que permite detectar los *spots* en los geles 2D, y almacenar los datos de espectrometría de masas (relación masa/carga de las proteínas) (SPARKMAN DO, 2000).

1.1.2. La metabolómica

Como ciencia, la metabolómica nació en 1998, cuando Olivier *et al.* llamaron “metaboloma” al conjunto de metabolitos presentes en un organismo en un momento y unas condiciones determinadas (OLIVIER S, *et al.*, 1998), lo que permite conocer cómo la composición bioquímica de un organismo condiciona su fenotipo.

La metabolómica es la *ómica* más reciente, y proporciona una gran información funcional de los organismos (SUMNER L, *et al.*, 2003). Mientras que la transcriptómica o la proteómica aportan también información funcional, la realidad es que los metabolitos representan el último nivel en un sistema biológico, ya que son los metabolitos en sí mismos los que representan entidades funcionales en la célula.

Los avances recientes en las técnicas de espectrometría de masas han permitido que sea posible identificar multitud de metabolitos con pequeñas muestras de tejido (SMITCH C, *et al.*, 2006). Además, el desarrollo de técnicas bioinformáticas y software especializado hace que sea posible estandarizar los procesos de análisis del metaboloma (KATAJAMAA M, *et al.*, 2005).

Los estudios metabolómicos globales, o *untargeted analysis*, no dirigidos a ningún metabolito concreto son aún más recientes. Aunque requieren el manejo de grandes cantidades de datos al medir tantos metabolitos como sea posible y su análisis es más complicado, proporcionan mucha más información y pueden ayudar a entender la red metabólica completa de los organismos. Algunas aplicaciones de este tipo de análisis son los efectos de algunas infecciones sobre los organismos o la caracterización de genes y proteínas (VINAYAVEKHIN N, *et al.*, 2010).

1.2. Interacciones planta-patógeno

Aunque las plantas se encuentren continuamente en contacto con potenciales microorganismos patógenos, se considera excepcional que sean susceptibles a desarrollar enfermedades.

Las plantas disponen de una “primera línea defensiva”, inespecífica, que puede frenar las invasiones de parásitos, que incluyen barreras poliméricas, cutícula en las hojas, ... Cada patógeno tiene una forma particular de burlar esta línea de defensa (con enzimas específicas, penetrando en la planta a través de estomas, ...) o, incluso, la invasión puede producirse a través de tejidos enfermos o dañados.

Aunque carecen de sistema inmunológico propiamente dicho, las plantas han desarrollado mecanismos de vigilancia ante la posible invasión de patógenos.

En aquellas situaciones en las que el patógeno invada los tejidos de la planta, ésta es capaz de generar unas respuestas si se reconoce específicamente el patógeno, es decir, en caso de un patógeno incompatible. Estos eventos se conocen como respuesta hipersensible (HR) (AZCÓN-BIETO J, TALÓN M, 2009).

En el caso en que este reconocimiento no tenga lugar se producirá una infección de tipo compatible, en la que la planta sufrirá la enfermedad.

1.2.1. El reconocimiento del patógeno: el modelo zig-zag

El modelo zig-zag fue propuesto por JONES J. y DANGL J., 2006, y describe los pasos sucesivos que se dan en la interacción de la planta y el patógeno invasor (**Figura 1**).

El patógeno invasor tiene asociadas una serie de moléculas “patrón”, muy conservadas, llamadas PAMPs (*Pathogen Associated Molecular Patterns*). Estos patrones pueden ser reconocidos por los PRRs de la planta (*Pattern Recognition Receptors*), una serie de receptores de membrana que se hallan en la superficie de la mayoría de las células. Tras este reconocimiento se dispara la PTI (*PAMP-triggered immunity*), asociada al aumento de la producción de especies reactivas del oxígeno, deposición de calosa a los diferentes lugares de infección, fabricación de compuestos antimicrobianos, ... (NÜRNBERGER T, *et al.*, 2004)

Los patógenos pueden producir factores de virulencia denominados comúnmente efectores para potenciar su expansión (DANGL J, MCDOWELL J, 2006), pudiendo actuar extra- o intracelularmente. Estos efectores pueden suprimir la PTI, pero pueden ser reconocidos como factores de avirulencia (Avr) por los genes R de la planta, provocando la ETI (*effector-triggered immunity*). La interacción entre los Avr y los genes R es específica y se explica en el modelo de interacción *gen a gen* (PARKER J, *et al.*, 1997), donde se especifica que una infección es incompatible si hay interacción Avr-R, y es compatible en caso contrario (**Figura 2**). La ETI se caracteriza por la activación de la HR (SPOEL S, DONG X, 2012). Sin embargo, cuando el reconocimiento por los genes R no se produce y no hay interacción Avr-R, se dice que la infección es compatible y por tanto la planta es susceptible. Este fenómeno se conoce como ETS (*effector-triggered susceptibility*).

1.2.2. Las respuestas defensivas de la planta frente a los patógenos

Una vez se ha producido el reconocimiento la planta activa sus respuestas defensivas mediante diversas cascadas. Los mecanismos de defensa dependen de este reconocimiento, y en ocasiones actúan juntos para frenar la expansión del patógeno.

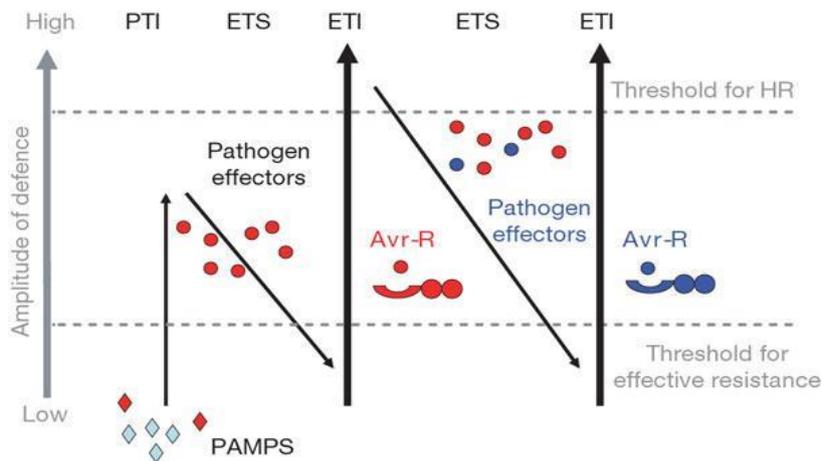


Figura 1. Modelo zig-zag en la infección de una planta por un patógeno. La planta activa las respuestas defensivas propias de la PTI cuando detecta los PAMPs asociados al patógeno. Los efectores pueden interferir con la PTI, y también pueden ser reconocidos por los genes de resistencia, desencadenando una respuesta del tipo ETI (interacción incompatible). En caso de no existir ningún reconocimiento, la planta no es resistente a la infección y se desarrollan los síntomas: se trata de una respuesta ETS (interacción compatible). Tomado de JONES J, DANGL J, *et al.*, 2006.

Estos mecanismos son (COLLINGE J, *et al.*, 1994):

a) *Respuesta hipersensible (HR)*. Ocurre en las interacciones incompatibles. Se caracteriza por necrosis local en la zona de infección, resultado de una muerte celular programada. Este tipo de respuesta va acompañada de producción de fitoalexinas y PR (proteínas relacionadas con la patogénesis). Esta necrosis se produce por la liberación de especies reactivas del oxígeno, entre las que destaca el H₂O₂ (LEVINE A, *et al.*, 1994). También suele ir ligada a una resistencia sistémica adquirida (SAR) inducida por la interacción planta-patógeno.

b) *Resistencia sistémica adquirida (SAR)*. Después de la producción de lesiones necróticas, en la planta puede tener lugar una resistencia sistémica a las siguientes infecciones del patógeno. Esta respuesta tiene que ver con la expresión de genes PR (HEIL M, 2002), y la molécula señalizadora que se cree que puede estar implicada es el ácido salicílico (LAWTON K, *et al.*, 1995).

c) *Producción de barreras físicas*. Es el mecanismo de defensa más común. Es bastante frecuente la lignificación de los tejidos dañados para impedir la entrada de patógenos, causando impermeabilidad y resistencia a dichos tejidos (MOURA J, *et al.*, 2010).

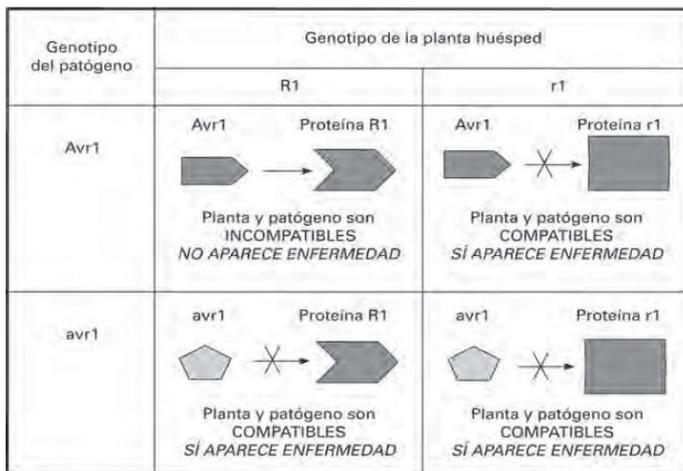


Figura 2. Esquema de la hipótesis “gen a gen”. Para que se produzca incompatibilidad debe existir complementariedad entre el gen Avr y el gen R. Ante cualquier mutación en uno de los genes (*avr*), no existe un reconocimiento específico y la infección es compatible: aparecen los síntomas. Tomado de AZCÓN-BIETO J, TALÓN M, 2009.

Las moléculas señalizadoras, que responden a la infección y extienden esta respuesta defensiva al resto de la planta mediante diversas rutas, son fundamentalmente tres: el etileno y el ácido jasmónico (implicados en la respuesta a patógenos de tipo biótrofo), y el ácido salicílico (de tipo necrótrofo).

1.2.3. *Algunos sistemas planta-patógeno. El sistema tomate-Pseudomonas syringae p.v. tomato*

Tal y como se ha descrito anteriormente, el reconocimiento específico planta-patógeno se produce según la hipótesis “gen a gen”, en la que el gen R de la planta reconoce los Avr del patógeno.

Se han caracterizado multitud de genes R y sus Avr asociados de plantas y patógenos, respectivamente (**Tabla 1**).

El presente trabajo se centrará en la infección que *Pseudomonas syringae* causa en el tomate. Esta bacteria causa unas manchas acuosas en las hojas que pueden necrosar, frecuentemente rodeadas por un halo amarillo (WILLIS D, KINSCHERF T, 2009).

Tabla 1. Ejemplos de genes R y sus respectivos genes Avr en diversos sistemas planta-patógeno.
Adaptado de HAMMOND-KOSACK K, *et al.*, 1997.

Planta	Gen R	Patógeno	Gen Avr
Arabidopsis	<i>RPS2</i>	<i>Pseudomonas syringae</i> p.v. <i>tomato</i>	avrRpt2
Arabidopsis	<i>RPM1</i>	<i>Pseudomonas syringae</i> p.v. <i>maculicola</i>	avrRpm1/avrB
Tabaco	<i>N</i>	Virus del mosaico del tabaco	-
Tomate	<i>Cf-9, Cf-2, Cf-4</i>	<i>Cladosporium fulvum</i>	avr9, avr2, avr4
Tomate	<i>Pto</i>	<i>Pseudomonas syringae</i> p.v. <i>tomato</i>	avrPto

Pseudomonas syringae es un patógeno que causa enfermedades en una gran cantidad de organismos. Aunque numerosas especies han desarrollado resistencias, en un ciclo infeccioso normal *P. syringae* tiene dos formas de infectar: una fase inicial, en la que la bacteria llega a la superficie de una planta sana, y una fase endofítica en la que las bacterias penetran a través de estomas y de algunas heridas (XIN X, *et al.*, 2013). En esta última fase, bajo condiciones que permitan una infección compatible, sucede una multiplicación anormal de la bacteria en el apoplasto y la primera aparición de síntomas.

El cultivar Rio Grande tiene el gen *Pto*, que ofrece resistencia al gen de avirulencia *avrPto* y *avrPtoB*. Gracias a esto, la interacción entre este tipo de tomate y *P. syringae* es incompatible, y la planta no desarrolla la enfermedad (JIA Y, 1997).

Sin embargo, cuando se produce una delección en los genes *avrPto* (Δ *avrPto*/ Δ *avrPtoB*) del patógeno, este reconocimiento específico no se produce y la interacción es compatible: la planta desarrolla la enfermedad. Esto ofrece un sistema modelo para comparar similitudes y diferencias que comparten ambos tipos de infecciones a nivel de proteoma o metaboloma.

1.3. Análisis estadístico multivariante para datos ómicos

Tanto los estudios metabolómicos como los proteómicos no dirigidos generan una gran cantidad de datos. Por ello, el desarrollo de las tecnologías ómicas va ligada al desarrollo de software bioinformático especializado que permita el análisis masivo de este tipo de datos.

Algunos de estos programas son capaces de detectar los spots de los geles 2D o procesar datos de imagen obtenidos en arrays para experimentos proteómicos, o almacenar los datos de concentración de cada metabolito obtenido por espectrometría de masas.

Estos datos masivos requieren métodos estadísticos multivariantes para su análisis ya que, generalmente, en los estudios ómicos no dirigidos hay una gran cantidad de variables (cada uno de los metabolitos, proteínas, ... detectados) y un número reducido de observaciones (cada una de las condiciones experimentales en las que se realiza el estudio).

Una de las herramientas más utilizadas para realizar estudios de datos ómicos es el análisis de componentes principales (PCA, por sus siglas en inglés) o la regresión parcial por mínimos cuadrados (PLS), que permiten conocer qué compuestos, metabolitos, proteínas, ... son los que más respuesta sufren al cambio de condición experimental.

Ambos métodos multivariantes reducen el número de variables original para ofrecer un modelo de regresión similar al modelo original, utilizando un conjunto de nuevas variables llamadas componentes, que se ordenan por la varianza (PCA) o la covarianza entre variables explicativas y variables dependientes (PLS) que explican de los datos originales.

Este abordaje resulta ideal para la observación de ciertos fenómenos biológicos, como la respuesta a una infección propuesta en el presente trabajo, gracias a los gráficos de puntuaciones y de cargas propios de dichos análisis.

Algunos programas estadísticos para realizar el análisis de estos datos son R o SPSS. Sin embargo, existen softwares especializados únicamente en análisis multivariante, como SIMCA.

1.4. Integración de ómicas

Para comprender completamente los fenómenos biológicos que suceden en una determinada situación, como es en este caso la respuesta de una planta a una infección, no es suficiente con estudiar solo un tipo de información cada vez. Este hecho, sumado al enorme crecimiento de la información extraída mediante análisis de todo tipo de ómicas, hace que se requieran métodos que permitan integrar la información de varias de estas ómicas.

Aunque la integración de ómicas es un tipo de estudio relativamente reciente, se han desarrollado métodos multivariantes que permiten este análisis. Uno de estos métodos es el análisis O2PLS (*Orthogonal 2-way Projections to Latent Structures*), una ampliación de la regresión por PLS descrita en TRYGG J., 2002.

La observación de las correlaciones existentes entre varios tipos de ómicas (proteómica-metabolómica, transcriptómica-proteómica, ...) puede proporcionar información que no sería posible extraer en los análisis individuales, por lo que este nuevo método de análisis multivariante puede proporcionar un "extra" de información biológica.

1.5. Estudios ómicos dirigidos. Análisis de la varianza (ANOVA)

En ocasiones, llevar a cabo un estudio no dirigido no es lo más interesante y se opta por el estudio de pequeños grupos de compuestos (por ejemplo, el estudio de la fracción de algunos azúcares, aminoácidos, ...). Estos estudios generan un conjunto de datos mucho más reducido y generalmente no es necesario emplear técnicas de análisis multivariante.

Cuando el número de compuestos considerado es reducido se opta por métodos estadísticos mucho más sencillos. Un ejemplo es el análisis de la varianza (ANOVA). Este método permite realizar un contraste de hipótesis para establecer si las medias de dos o más poblaciones son iguales en respuesta a un factor, que en un análisis de este tipo suele ser la condición experimental del estudio (ROMERO R, ZÚNICA L, 2013).

2.

OBJETIVOS

2. Objetivos

El objetivo principal del presente trabajo es proponer un método que permita estudiar las proteínas, los metabolitos y las relaciones entre ellos asociados a las infecciones compatible e incompatible, para dilucidar los mecanismos de defensa de la planta frente al ataque del patógeno. Todo ello a partir de datos de tres sistemas planta-patógeno: controles, con infección compatible y con infección incompatible.

Los objetivos específicos del estudio serán:

- comprobar si existen diferencias significativas en el proteoma y en el metaboloma en cada tipo de condición.
- conocer si las proteínas y los metabolitos que se expresan diferencialmente en cada tipo de planta están relacionadas directamente con la respuesta de la planta a la infección.
- saber de qué forma la planta es capaz de defenderse del patógeno integrando en un solo análisis los datos metabolómicos y proteómicos.

3.

MATERIALES Y MÉTODOS

3. Materiales y métodos

3.1. Obtención de datos metabolómicos y proteómicos

3.1.1. Material biológico

En este estudio, los investigadores e investigadoras del IBMCP han utilizado la bacteria *Pseudomonas syringae* pv. *tomato* DC3000 (*Pst* DC3000). Además, también utilizan una cepa similar que contiene deleciones en los genes *avrPto* y *avrPtoB* (*Pst* DC3000 Δ *avrPto*/ Δ *avrPtoB*). Estas bacterias han sido crecidas y diluidas a la concentración del inóculo, correspondiente a 10^7 CFU/ml.

Las plantas de tomate empleadas fueron de la variedad Rio Grande. Esta variedad contiene el gen de resistencia *Pto* (RG-*Pto*).

Las semillas han sido crecidas en un invernadero con un fotoperiodo luz/oscuridad de 16/8 h a 26/30°C, una humedad relativa del 65% en macetas de 12 centímetros de diámetro que contienen una mezcla 1:1 de turba y vermiculita.

3.1.2. Inoculación de la bacteria

Las inoculaciones con las bacterias compatibles e incompatibles han sido producidas por inmersión de hojas de plantas de tomate de la variedad Rio Grande de unos 28 días de edad, en una suspensión bacteriana de la concentración mencionada anteriormente (10^7 CFU/ml). Los controles, es decir, las inoculaciones tipo *mock*, se han producido por inmersión de las hojas en una solución 10 mM de $MgCl_2$ sin inóculo.

El procedimiento de inoculación se repite con cuatro hojas diferentes, conservándolas con nitrógeno líquido.

3.1.3. Extracción de proteínas

Las muestras congeladas de hojas inoculadas se machacan en un mortero y se añade tampón de extracción para homogeneizar el material. Tras una centrifugación, el sobrenadante se mezcla con ácido tricloroacético en frío y se obtiene el sedimento proteico, que tras un lavado se disuelve en un tampón de lisis.

Tras este procedimiento, llevado a cabo en el IBMCP, se llevó a cabo una electroforesis bidimensional (2D-DIGE). La identificación de las muestras proteicas en el gel se realizó en el Servicio de Proteómica de la Universidad de Valencia.

La cuantificación de cada *spot* en el gel se lleva a cabo en el IBMCP, utilizando un *western blot* y detectando las proteínas con antisueros contra las proteínas de tomate. Así, se puede conocer la identidad y la concentración de cada proteína.

En total, se obtienen datos de proteínas para 24 muestras u observaciones: *mock*, con infección compatible, y con infección incompatible; a las 18h y a las 36h; con cuatro réplicas por muestra (**Tabla 2**).

3.1.4. Extracción de metabolitos

Las muestras de los metabolitos de las hojas de tomate inoculadas se llevan a una cromatografía líquida acoplada a un espectrómetro de masas (HPLC-MS).

En el caso de los metabolitos apolares, se extraen utilizando como fase móvil metanol y triclorometano (MeOH:CHCl₃) y como fase estacionaria una columna C18 en el HPLC. En la MS acoplada a HPLC se utiliza un ionizador APCI (ionización química a presión atmosférica).

Los metabolitos polares se extraen con una fase móvil de agua y acetonitrilo, con la misma fase estacionaria. En la MS, se utiliza un ESI (ionización por electroespray).

En total, se obtienen tres sets de datos, uno de los polares con carga negativa, con carga positiva, y apolares. Se obtienen datos de metabolitos para 60 muestras u observaciones: *mock*, con infección compatible y con infección incompatible, a las 10h, 18h, 24h, 36h y 48h, con 4 réplicas por muestra (**Tabla 2**).

Tabla 2. Resumen de las muestras tomadas para cada análisis, de metabolómica, proteómica e integración.

	Metabolómica	Proteómica (e integración)
Sistemas	<i>Mock</i> , compatible, incompatible	<i>Mock</i> , compatible, incompatible
Tiempos (h)	10, 18, 24, 36, 48	18, 36
Réplicas	4	4
Total muestras	60	24

Una vez han sido medidos todos los metabolitos de los tres análisis (apolares, polares cargados negativamente, y polares cargados positivamente) y todas las proteínas, los resultados de las medidas se recogen en una hoja de Excel identificando cada *spot* en el DIGE y cada metabolito con un número de identificación, y se procede a su análisis estadístico.

3.2. Estudio estadístico mediante Análisis de Componentes Principales (PCA)

El análisis de componentes principales (PCA) es un método estadístico multivariante utilizado para reducir la dimensión del espacio de las variables de una matriz.

El PCA permite representar dicho espacio con unas pocas variables ortogonales entre sí que capturan la mayoría de la variabilidad. Dichas variables que se denominan componentes principales (CP), son combinaciones lineales de las variables originales y se calculan de tal forma que la primera es la que más variabilidad explica en el sistema, la segunda, la siguiente mayor cantidad de variabilidad, y así sucesivamente. Geométricamente, un PCA se puede describir como una proyección de los datos de las variables originales sobre un hiperplano diferente, el de las componentes principales.

El punto de partida de un PCA es la matriz de datos $X \in \mathbb{R}^{N \times K}$, con N filas (observaciones, en este caso sistemas planta-patógeno) y K columnas (variables, metabolitos y proteínas).

El modelo PCA transforma la matriz de observaciones $X \in \mathbb{R}^{N \times K}$ según la siguiente expresión:

$$X = TP^T = \sum_{k=1}^K t_k p_k^T \quad \text{(Fórm. 1)}$$

A la matriz $T \in \mathbb{R}^{N \times K}$ se le llama matriz de puntuaciones o *scores*, ya que sus filas son los vectores de coordenadas de cada una de las observaciones en el hiperplano definido por el modelo PCA. Las columnas de esta matriz (t_1, t_2, \dots, t_K) son los vectores de *scores* de cada una de las CP.

A la matriz $P \in \mathbb{R}^{K \times K}$ se le llama matriz de pesos o *loadings*, y la componen los vectores de pesos p_1, p_2, \dots, p_K que definen las direcciones sobre las que se proyectan las observaciones. Estas direcciones reflejan la contribución de las variables originales a la formación de las CP, a las que dotan de un posible significado.

Estos vectores p_k son los vectores propios de la matriz de varianzas-covarianzas, correspondientes a los valores propios λ_i , que a su vez son las varianzas de las CP. Así, la primera CP se define como la combinación lineal $t_1 = Xp_1$, que tiene la máxima varianza sujeta a $|p_1|=1$. La segunda, es la combinación lineal $t_2 = Xp_2$ que tiene la siguiente mayor varianza sujeta a $|p_2|=1$, y con la condición de ser ortogonal a la primera CP (t_1). Así, se definen hasta K las posibles CP, todas ellas ortogonales entre sí.

Antes de un PCA, generalmente se lleva a cabo un escalado de la matriz X a varianza unitaria (UV), dividiendo el valor de cada variable entre su desviación típica y además restándoles la media. Esto hace que todas las variables tengan, a priori, el mismo peso en la formación del modelo PCA.

Generalmente, no es necesario calcular todos los posibles K valores propios, ya que la mayor parte de variabilidad de los datos se recoge en las primeras CP. Así, en general, la expresión de la descomposición de X en un modelo PCA reteniendo A componentes quedaría:

$$X = \sum_{a=1}^A t_a p_a^T + E = TP^T + E \quad \text{(Fórm. 2)}$$

Entonces, la aproximación de X (\hat{X}), se daría por $\hat{X} = \sum_{a=1}^A t_a p_a^T$, que es la predicción de X a partir del modelo con A componentes. E representa la matriz de residuos, considerado ruido.

El número de CP a incluir en el modelo PCA debe ser capaz de detectar y ofrecer una representación adecuada de las variaciones existentes en los datos. El algoritmo más utilizado para calcular este número es el método de validación cruzada, que maximiza la capacidad de predicción de las CP (EASTMENT HT, KRZANOWSKI WJ, 1982; KRZANOWSKI WJ, 1987; WOLD S, 1976), y consiste en descartar observaciones de X basándose en la correspondencia entre el PCA y la descomposición en valores singulares de X . Este método predice las observaciones que sucesivamente se van descartando de X , eliminando su fila y columna con el modelo PCA obtenido a partir de la matriz X sin la observación i , $X^{(-i)}$, con la finalidad de utilizar la máxima información recogida en los datos.

Realizar un análisis estadístico de este tipo, que englobe todos los datos metabolómicos y proteómicos, requiere de herramientas capaces de ejecutar análisis multivariantes. En este trabajo se utilizará el software Simca 13.0.3 (Umeå, Suecia).

Para realizar el PCA se establece el valor de concentración de cada metabolito o proteína como variable. Una vez ejecutado el análisis con estos parámetros, se realiza una normalización para escalar las variables a varianza unitaria (UV).

Con el modelo creado y las variables escaladas, se procede a la visualización de los datos. Para ello se recurre a los gráficos de puntuaciones o *score plots*, y a los gráficos de cargas o *loading plots*, en los que se verá reflejado el grado de separación de las condiciones experimentales que hace cada componente y la contribución de cada variable a la formación de las componentes, respectivamente.

3.3. Estudio estadístico mediante Regresión por Mínimos Cuadrados Parciales (PLS)

Utilizando el mismo software también se pueden realizar estudios de regresión, como una regresión por mínimos cuadrados parciales (PLS).

El PLS es un método de proyección en estructuras latentes mediante mínimos cuadrados parciales, presentado por WOLD H, 1973 y posteriormente perfeccionado por WOLD S, *et al.*, 1984. Dicho modelo parte de dos matrices X e Y , $X \in \mathbb{R}^{N \times K}$ o matriz predictora, e $Y \in \mathbb{R}^{N \times M}$ o matriz respuesta (KOURTI T, 2002; KOURTI T, MACGREGOR JF, 1996; KOURTI T, MACGREGOR JF, 2006). Sus ecuaciones, para un modelo con A variables latentes, son:

$$X = \sum_{a=1}^A t_a p_a^T + E = TP^T + E \quad (\text{Fórm. 3})$$

$$Y = \sum_{a=1}^A u_a c_a^T + F = UC^T + F \quad (\text{Fórm. 4})$$

, donde $T \in \mathbb{R}^{N \times A}$ y $U \in \mathbb{R}^{N \times A}$ son matrices de coordenadas (*scores*) de las observaciones originales en las variables latentes, y $P \in \mathbb{R}^{K \times A}$ y $C \in \mathbb{R}^{M \times A}$ son matrices de pesos (*loadings*), que indican cómo las variables latentes están relacionadas con las originales X e Y . $E \in \mathbb{R}^{N \times K}$ y $F \in \mathbb{R}^{N \times M}$ son matrices de residuos para X e Y respectivamente, para un modelo con A variables latentes o componentes determinadas por validación cruzada.

El modelo PLS reduce la dimensión de variables mediante la proyección de los espacios X e Y sobre el espacio T de las variables latentes, que contiene la información más relevante en las A dimensiones consideradas. Trabajando en este espacio de variables latentes de reducidas dimensiones (t_1, t_2, \dots, t_A) se simplifica la identificación e interpretación de las correlaciones.

De la matriz X y de la matriz Y se extraen las variables latentes que expliquen la mayor variabilidad posible de los datos de X y que resultan lo más predictivas posibles de las variables de Y . Por tanto, podemos decir que el PLS es un método que busca las direcciones que maximizan las covarianzas entre X e Y .

La covarianza muestral entre dos variables cualquiera x e y es $\text{cov}(x, y) = r(x, y)s_x s_y$, siendo $r(x, y)$ el coeficiente de correlación y s_x, s_y las respectivas desviaciones típicas. Por tanto, las direcciones que maximicen la covarianza maximizan el producto $r(x, y)s_x s_y$ y, por tanto, dichas direcciones serán aquellas con mayores valores de r (presentando buenas propiedades

predictivas de la y a partir de la x), y que al mismo tiempo maximicen s_x y s_y . Esto quiere decir que dichas direcciones estarán entre las de mayor variabilidad de la matriz X y la matriz Y , lo que supone que explicarán gran parte de la variabilidad de ambos espacios.

La primera variable latente PLS o componente, $t_1 = Xw_1$ es una combinación lineal de las variables x_k que maximiza la covarianza entre t_1 y el espacio de Y . El primer vector de pesos o *loadings*, w_1 , es el primer vector propio de la matriz de varianzas-covarianzas $X^T Y Y^T X$. Una vez que la componente ha sido calculada, las columnas de X son proyectadas sobre t_1 para dar un vector de regresión, $p_1 = \frac{X t_1}{t_1^T t_1}$, y la matriz X es reducida restando los valores predichos $\hat{X} = t_1 - p_1^T$ por el modelo formado por t_1 , w_1 y p_1 para dar los residuos $X_2 = X - t_1 p_1^T$.

Del mismo modo, se hace una regresión de las columnas de Y sobre t_1 , y se obtiene q_1 , el primer vector de pesos en el espacio latente de Y . A continuación, se reduce Y : $Y_2 = Y - t_1 q_1^T$.

A partir de los residuos, se calcula la segunda componente $t_2 = Xw_2$, donde w_2 es el primer vector propio de la matriz de varianzas-covarianzas $X_2^T Y Y^T X_2$, y así sucesivamente.

La obtención de las componentes de forma secuencial es el método utilizado por Simca para los análisis, y se denomina algoritmo NIPALS (GELADY P, KOWALSKI B, 1986).

Tal y como ocurría en el PCA, los vectores de *scores* t_a y de pesos w_a son ortogonales. Sin embargo, mientras que el PCA solo modela espacios simples, PLS maximiza la covarianza entre X e Y , lo que permite explicar sus correlaciones.

Se ha utilizado esta aproximación en aquellos casos en los que no ha sido posible obtener una separación de las condiciones experimentales mediante un PCA. Para este análisis se establece como variable X o independiente el valor de concentración de cada metabolito. La variable Y o dependiente es, en caso de este estudio de metabolómica y proteómica, un valor cualitativo que indica el tipo de muestra o condición experimental. La normalización aplicada también es, como en el PCA, del tipo UV.

En el análisis metabolómico de los metabolitos apolares, realizado con un PLS, se ha establecido como variable Y el tipo de tratamiento que recibe la planta como valor cuantitativo, de 0 a 2: 0 para el *mock*, 1 para la infección incompatible y 2 para la infección compatible, tal y como se realizó en LÓPEZ GRESA M, *et al.*, 2017.

La visualización de los datos es similar a la que se describe para un PCA. Los gráficos de puntuaciones o *score plots* reflejan la separación de las muestras que realizan las componentes, y los gráficos de cargas o *loading plots*, la contribución de la variable a la formación de las componentes.

3.4. Integración proteómica-metabolómica mediante un análisis *Orthogonal 2-way Projections to Latent Structures* (O2PLS)

Debido a que un análisis PLS suele utilizarse para relacionar diferentes conjuntos de datos, también se ha utilizado esta regresión para realizar un análisis integrado de los datos de

metabolómica y de proteómica. En este análisis integrado, se ha utilizado una revisión del método PLS llamada O2PLS (*Orthogonal 2-way Projections to Latent Structures*).

En una regresión PLS se mide la relación entre una matriz descriptora X y una matriz respuesta Y . El método OSC (*orthogonal signal correction*), propuesto por WOLD S, *et al.*, 1998, propone una revisión del PLS en la que se elimina la información sistemática contenida en X , que explica las correlaciones entre las variables de dicho espacio y no correlacionada con las variables de Y .

Siguiendo esta idea, se creó una modificación del algoritmo NIPALS llamada *Orthogonal Projections to Latent Structures* (OPLS), que elimina la variabilidad en X que es ortogonal a Y .

Más tarde, en TRYGG J, 2002 se enunció el análisis *Orthogonal 2-way Projections to Latent Structures*, que además de eliminar la variación en X no relevante para la predicción de Y , también elimina la variación en Y que es ortogonal a X , y resulta predictivo en ambos sentidos: $X \rightarrow Y$ e $Y \rightarrow X$. Así, este método permite que las componentes T ($\sim Y$) también sean componentes principales en X (TW^T).

La idea básica del algoritmo es usar una regresión lineal múltiple para estimar los constituyentes del modelo, para después dividir la parte sistemática en X y en Y en dos partes: una relacionada con X y con Y (covarianza), y otra que no lo está (ortogonal). Es decir, para cada matriz, la siguiente está calculada de una forma que la hace ortogonal o linealmente independiente a la primera.

Por tanto, un O2PLS puede escribirse como un modelo de análisis de factores donde unos factores T son comunes a X e Y :

$$\text{Modelo de X: } X = TW^T + T_{Y\text{-orth}}P_{Y\text{-orth}}^T + E \quad (\text{Fórm. 5})$$

$$\text{Modelo de Y: } Y = UC^T + U_{X\text{-orth}}P_{X\text{-orth}}^T + F \quad (\text{Fórm. 6})$$

$$\text{Relación interna: } U = T + H \quad (\text{Fórm. 7})$$

$$\text{Predicción de Y: } \hat{Y} = TC^T \quad (\text{Fórm. 8})$$

La relación interna para X e Y es la descrita en la **Fórm. 7**, donde H es la matriz residual. Como en el PLS, la ecuación predictiva de Y es la **Fórm. 8**, donde T y $U \in \mathbb{R}^{N \times A}$, son las matrices de *scores* de X e Y respectivamente con A componentes. E y F son los residuos.

$W \in \mathbb{R}^{K \times A}$ y $C \in \mathbb{R}^{M \times A}$ son las matrices de pesos de las componentes conjuntas o correlacionadas. $P_{Y\text{-orth}}^T \in \mathbb{R}^{K \times D}$ es la matriz de cargas con D componentes ortogonales a Y , y $T_{Y\text{-orth}} \in \mathbb{R}^{N \times D}$ se corresponde con la matriz de *scores* ortogonal a Y . De manera análoga se definen $P_{X\text{-orth}}^T$ y $U_{X\text{-orth}}$.

En general, el modelo O2PLS encuentra la variación que altera la proyección de X en W , o la de Y en C . Es decir, encuentra y elimina aquella variación que dificulta la interpretación del modelo en ambas matrices.

Las ventajas de utilizar un análisis O2PLS respecto a un análisis PLS convencional es que los modelos mejoran mucho su interpretabilidad, ya que la variación relevante e irrelevante se separa en X y en Y . De hecho, $T_{Y\text{-orth}} \in \mathbb{R}^{N \times D}$ y $U_{X\text{-orth}} \in \mathbb{R}^{N \times D}$ son generalmente irrelevantes, en el sentido

de que no contribuyen a la predicción de Y y X respectivamente. Aun así, es una buena idea analizar dicha información con un PCA o algún análisis individual para una mejor interpretación del estudio, aunque no se estudien en este caso las correlaciones entre las dos matrices.

Para el análisis integrado O2PLS se ha establecido como variable independiente (X) los datos de la proteómica, y como variable dependiente (Y) los datos de metabolómica, ya que son los metabolitos los que dependen generalmente de la concentración de proteínas, al representar el último nivel de respuesta celular (por debajo de la expresión génica y proteica).

Para realizar el análisis de regresión debemos adaptar los datos de proteómica y metabolómica haciendo que tengan las mismas dimensiones. Al tener los datos de proteómica un número de muestras menor que los de metabolómica (solo se miden a las 18h y a las 36h post-inoculación), únicamente se tienen en cuenta los datos de metabolómica relativos a 18h y 36h, descartando las mediciones a 10h, 24h y 48h. El número total de observaciones para el análisis integrado es, por tanto, de 24 muestras.

La observación, como en los casos anteriores, también es similar, aunque en el gráfico de cargas se puede observar la superposición de los dos tipos de variables, proteínas y metabolitos. En general, un valor similar en el “loading” de dos variables diferentes, X e Y (proteína y metabolito), indica que ambas responden de la misma manera a las condiciones experimentales, por lo que existe la posibilidad de que estén relacionadas.

Esta suposición será el principio que se siga en el análisis de resultados de las integraciones, donde se intentará extraer información sobre aquellas proteínas y metabolitos que estén relacionados entre sí y con una determinada condición experimental.

3.5. Identificación y representación gráfica de metabolitos y proteínas relacionados con la respuesta defensiva de la planta del tomate

La identificación de las proteínas relacionadas con la respuesta defensiva ha dependido directamente de la secuenciación realizada por el IBMCP. Sin embargo, algunos de los metabolitos se han identificado gracias a la base de datos METLIN, del *Scipps Research Institute* y a la relación masa-carga del compuesto y de sus iones precursores.

Una vez identificados proteínas y metabolitos, se ha recurrido al software Cytoscape (SHANNON P, *et al.*, 2003) para representar gráficamente los resultados obtenidos y hacer así más visual e interpretable la información extraída con los análisis estadísticos.

3.6. Estudio estadístico dirigido a metabolitos potencialmente importantes en la planta mediante un análisis de la varianza (ANOVA)

Tras realizar el análisis con los datos de metabolitos apolares, no se obtuvo ningún resultado ni con un modelo PCA ni con una regresión PLS.

Para confirmar esta conclusión, en el IBMCP se realizó un análisis metabolómico dirigido a ciertos metabolitos importantes en la planta. Dichas mediciones se realizaron en plantas *mock* e infectadas con cepas compatibles e incompatibles de *P. syringae*, solo midiendo a las 18h y a las 36h desde la infección. Se consideran datos de 4 grupos de metabolitos: lípidos, carotenoides, clorofilas y quinonas.

Con estos datos amablemente proporcionados por el IBMCP se realiza un análisis de la varianza (ANOVA) de un factor con el software *IBM SPSS Statistics 22*.

Como parámetros se establece un *p-valor* umbral para considerar dicho metabolito como diferencialmente presente de 0,05. El factor tenido en cuenta en el ANOVA es el tipo de infección (*mock*, incompatible y compatible), por lo que el estudio entre grupos tendrá 2 grados de libertad ($n - 1$).

Aquellos metabolitos cuyo *p-valor* haya mostrado una significación por debajo del umbral establecido ($p\text{-valor} < 0,05$) serán estudiados más a fondo mediante un análisis post-hoc utilizando el mismo software para comprobar las medias de qué condiciones difieren significativamente.

Este análisis post-hoc se realiza con un test HSD (*Honestly-significant difference*) de Tukey. Esta prueba ayuda a “afinar” más la respuesta genérica que ofrece un ANOVA indicando entre qué medias concretas existen diferencias, ya que el ANOVA solo indica si el test F es significativo (si existe, al menos, una diferencia entre las medias, pero no nos dice entre cuáles). El test de Tukey tiene como finalidad indicar entre qué medias existen diferencias y en qué sentido, es decir, cuál de ellas es mayor o menor respecto a las demás (KEPPEL G, WICKENS T, 2004).

4.

RESULTADOS Y DISCUSIÓN DEL ANÁLISIS METABOLÓMICO

4. Resultados y discusión del análisis metabolómico

Tras extraer todos de metabolitos en las diferentes condiciones experimentales, se lleva a cabo el análisis estadístico multivariante para comprobar qué metabolitos se encuentran diferencialmente expresados en cada condición. Esto permitirá caracterizar los metabolitos relacionados con la infección de la planta de tomate por *P. syringae*.

4.1. Análisis de los metabolitos polares con carga negativa

Se registran datos de 95 metabolitos polares con carga negativa. Tras realizar el PCA con estos datos, se consideran las primeras dos componentes del modelo, que explican el 40,9% y el 16,7% de la varianza de los datos, respectivamente. Conjuntamente explican el 57,6% de la varianza total.

En el gráfico de puntuaciones o *score plot* obtenido (**Figura 3**) se observa que la primera componente (PC1) separa las horas transcurridas desde la inoculación (10h, 18h, 24h, 36h y 48h), independientemente del tipo de infección. Por su parte, la segunda componente (PC2) es la que caracteriza la infección por *Pseudomonas*, ya que muestra la alteración metabólica debida al tipo de infección: *mock*, sistémica o compatible (Enferm.) e incompatible (Incomp.).

Los metabolitos que más contribuyen a la formación de ambas componentes aparecen en el gráfico de cargas (**Figura 4**). Como la componente que separa el tipo de infección es la PC2, se seleccionan aquellos con un mayor absoluto en su *loading* para pasar después a su posterior identificación (**Tabla 3**). Para este análisis se seleccionan los 10 metabolitos con un mayor *loading* y los 10 con un *loading* menor.

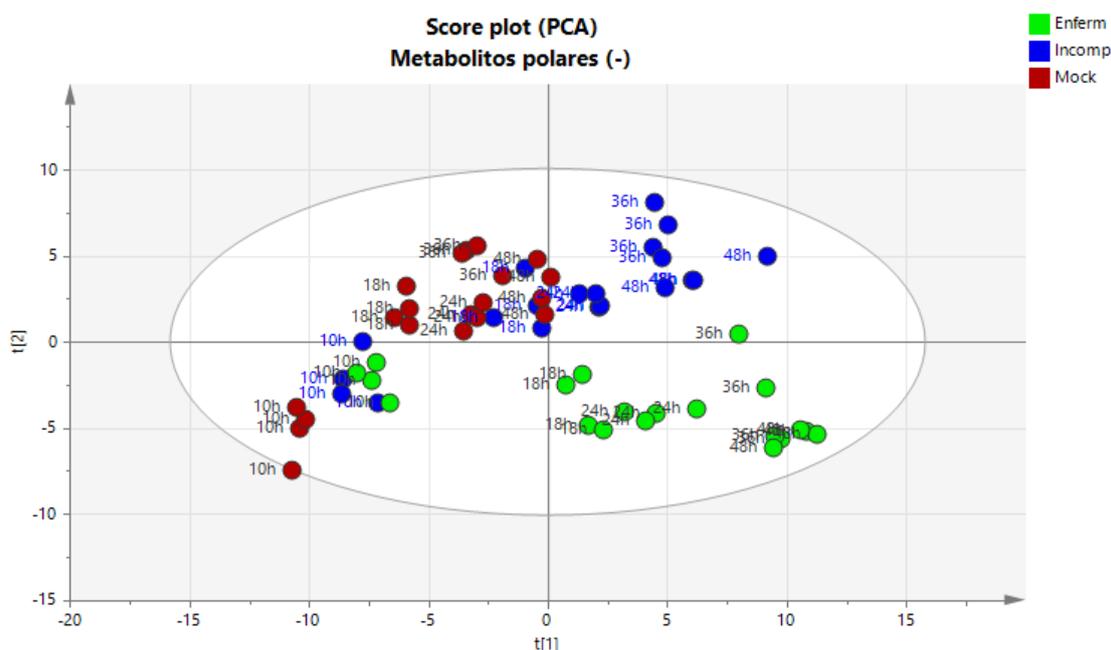


Figura 3. Gráfico de puntuaciones del PCA obtenido con el análisis de los metabolitos polares con carga negativa. En rojo, las plantas inoculadas *mock*; en verde, las inoculadas con una cepa virulenta, con delección; y, en azul, las inoculadas con una cepa silvestre y que producen una infección incompatible.

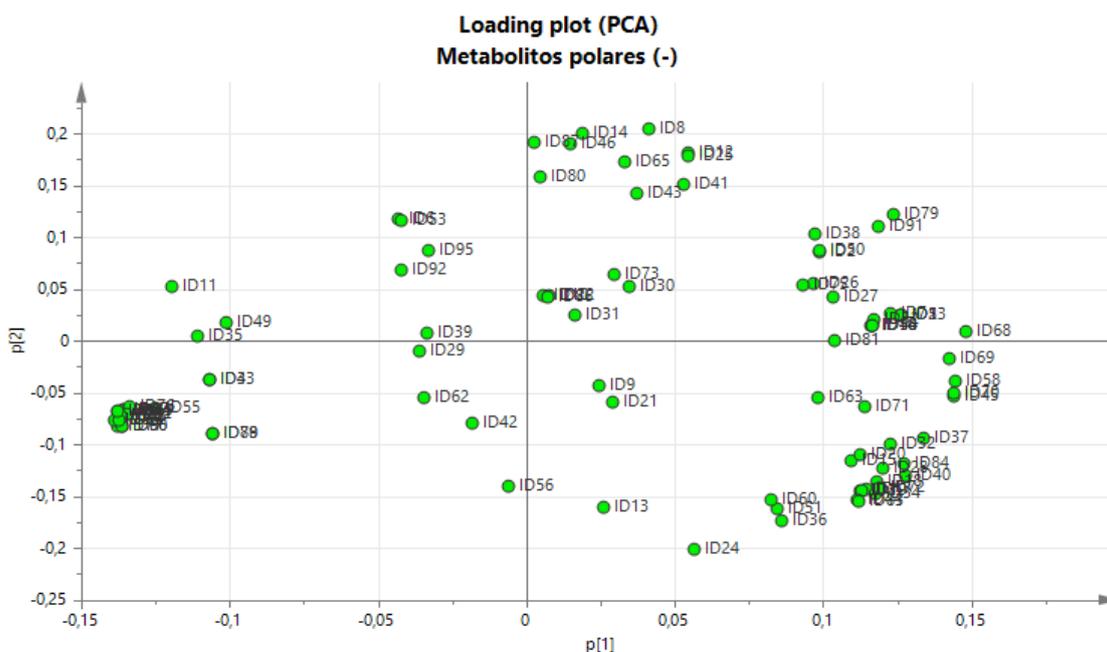


Figura 4. Gráfico de cargas del PCA obtenido con el análisis de los metabolitos polares con carga negativa. Los puntos indican los números de identificación (ID) de los metabolitos. Aquellos con más valor absoluto en la segunda componente son los tenidos en cuenta.

Los metabolitos seleccionados a partir del gráfico de cargas de la PC2 son los más importantes en las infecciones incompatibles (valores positivos) y en las infecciones compatibles (valores negativos).

El PCA llevado a cabo con los metabolitos polares con carga negativa demuestra que la concentración de estos metabolitos varía significativamente en respuesta a la infección de la planta de tomate por *Pseudomonas*. Además, los metabolitos inducidos en ambos tipos de respuesta, compatible e incompatible, son diferentes, sugiriendo un mecanismo metabólico diferenciado en las dos alteraciones.

Tabla 3. Lista de metabolitos polares con carga negativa con más peso en la formación de la PC2 y, por tanto, con un posible papel importante en la respuesta a la infección.

ID	Metabolito	Condición	MZ	t _R	ID	Metabolito	Condición	MZ	t _R
8	no identificado	Incompatible	174,954	1,00547	19	Sucrose or isomer	Compatible	388,116	1,25387
14	no identificado	Incompatible	304,912	0,97305	3	Sucrose or isomer	Compatible	387,113	1,25387
87	no identificado	Incompatible	342,877	1,07310	34	Tetrapeptide	Compatible	537,165	1,25387
46	no identificado	Incompatible	434,869	0,97305	60	Tetrapeptide	Compatible	488,160	1,23151
12	no identificado	Incompatible	288,935	0,96732	44	no identificado	Compatible	729,228	1,24490
25	no identificado	Incompatible	418,892	0,96732	83	Tetrasaccharide	Compatible	683,223	1,24490
65	no identificado	Incompatible	548,850	0,97981	13	Chlorogenic acid	Compatible	353,086	6,97588
80	no identificado	Incompatible	426,838	1,08704	51	Tetrapeptide	Compatible	487,176	1,21960
41	no identificado	Incompatible	158,976	0,99045	36	no identificado	Compatible	439,084	1,51763
43	no identificado	Incompatible	235,925	0,01500	24	Sucrose or isomer	Compatible	377,084	1,51025

En la tabla, de izquierda a derecha: ID (número de identificación del metabolito), nombre del metabolito, condición (la condición experimental en la que el metabolito tiene importancia), MZ (relación masa:carga del metabolito identificada en el espectrómetro de masas) y t_R (tiempo de retención en la cromatografía líquida).

En la infección compatible destaca especialmente la presencia de sacarosa o algún isómero, correspondiente a las ID19, 3 y 24. La detección de este disacárido puede ser indicativo de que se existe una regulación de la actividad “invertasa”, muy común en las interacciones planta-patógeno.

La sacarosa es un metabolito muy importante para el crecimiento, el desarrollo y en el metabolismo de la planta. Además, está implicada en la defensa de la planta activando la respuesta inmune contra varios patógenos y activando cascadas relacionadas con la regulación de la emisión de fitohormonas, refuerzo de la pared celular y algunas rutas de señalización. La presencia de azúcares también induce la expresión de genes PR (TAUZIN, AS, GIARDINA T, 2014).

Otro metabolito que se ha caracterizado en este tipo de infección es el ID13, que se corresponde con el ácido clorogénico (CGA).

Este compuesto (**Figura 5**) es un metabolito muy importante en la respuesta defensiva de la planta, y ya fue detectado en el trabajo realizado por LÓPEZ GRESA M, *et al.*, 2010 en plantas de tomate infectadas con *P. syringae*.

Dicho compuesto parece estar relacionado con la integridad de la membrana celular. En concreto, produce fenómenos de unión y permeabilización, aunque con un mecanismo que todavía se desconoce. La principal hipótesis es que interviene en diferentes procesos intracelulares que desembocan en la pérdida de la

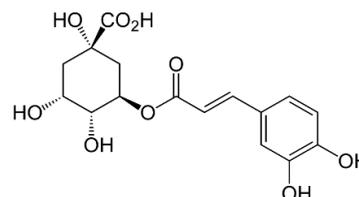


Figura 5. Estructura del CGA.

integridad de la membrana, de la diferencia de potencial e incluso puede tener un papel en la síntesis de nucleótidos. Todos estos factores llevan a la muerte de la bacteria (LOU Z, *et al.*, 2011).

4.2. Análisis de los metabolitos polares con carga positiva

El PCA en este caso se realiza con datos de concentración de 510 compuestos polares con carga positiva. Se consideran las dos primeras componentes del modelo, que explican el 30,9% y el 12,1% de la varianza encontrada en los datos, respectivamente; un total de 43% de varianza explicada por ambas.

En este caso, la PC1 informa sobre el tiempo transcurrido post-infección (10h a 48h), por lo que se descarta del análisis. Son la PC2 y la PC3 las que caracterizan realmente la infección. Como se puede comprobar en el gráfico de puntuaciones (**Figura 6**), la PC2 diferencia entre controles y plantas enfermas, y la PC3 diferencia la infección incompatible en el score negativo.

Por tanto, para caracterizar los metabolitos que intervienen en la enfermedad compatible y la incompatible se toman del gráfico de cargas (**Figura 7**) los metabolitos que más contribuyen a la diferenciación de los dos tipos de enfermedades. Es decir, los que tienen un mayor *loading* positivo en la PC2, y los que tienen un mayor *loading* negativo en la PC3 (**Tabla 4**).

En este caso, debido a la gran cantidad de metabolitos generados, se ha decidido tomar los datos del 5% de los metabolitos con un mayor valor en su *loading* de la PC2, y los del 5% con un valor más pequeño en el *loading* de la PC3, para observar las contribuciones principales.

Entre los metabolitos que caracterizan la infección incompatible, encontramos el *p-coumaroyltyramine* (CT), correspondiente a las ID 42 y 263. Este compuesto es una amida derivada del ácido hidroxicinámico, cuya síntesis está ligada a la inducción del etileno por la respuesta a una infección. Todos los derivados de este ácido comparten actividades antioxidantes y bactericidas, y juegan un papel muy relevante en la respuesta a la infección (ZACARÉS L, *et al.*, 2007).

Otros metabolitos identificados para esta condición, correspondientes al ID 383 y al ID101, es la *feruloyloctopamine* y la *caffeoyltyramine*, también derivados del ácido hidroxicinámico y con un papel similar. Este grupo de metabolitos parece estar más expresado en las plantas enfermas por una sobreexpresión del gen *THT* (*tyramine N-hydroxycinnamoyl transferase*), que responde al etileno y que se encarga de la síntesis del ácido hidroxicinámico (YOGENDRA K, *et al.*, 2015).

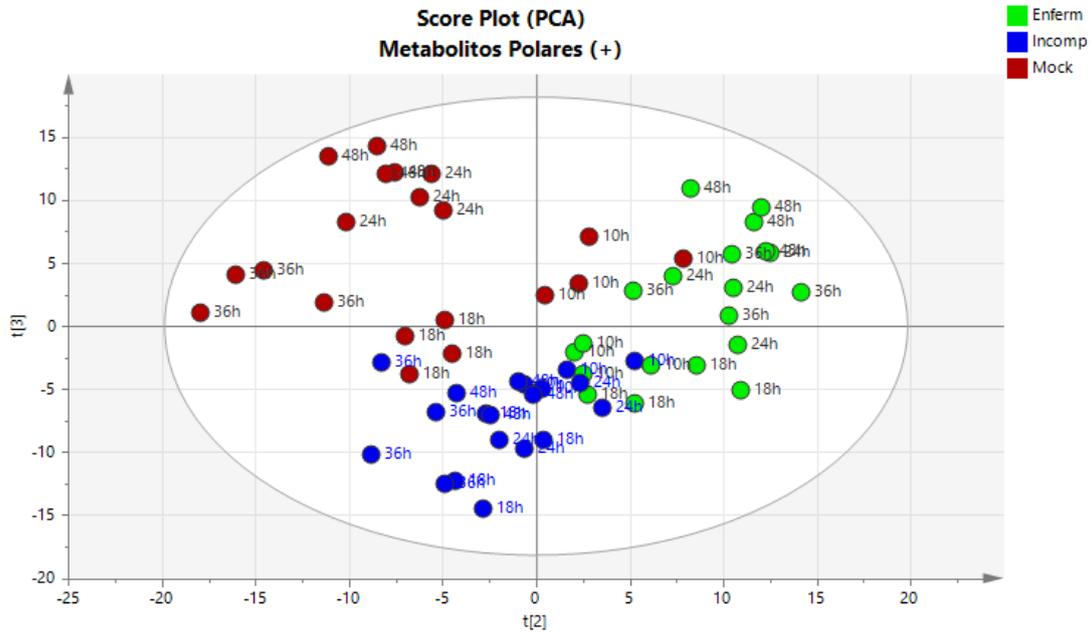


Figura 6. Gráfico de puntuaciones del PCA obtenido con el análisis de los metabolitos polares con carga positiva. En rojo, las plantas inoculadas *mock*; en verde, las inoculadas con una cepa virulenta, con delección; y, en azul, las inoculadas con una cepa silvestre y que producen una infección incompatible.

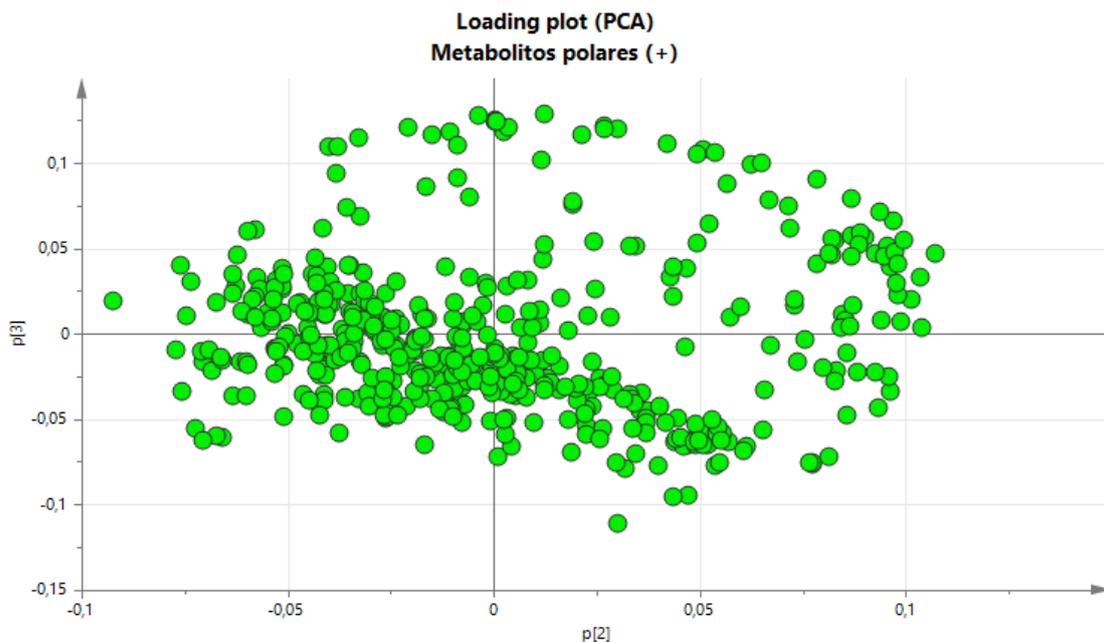


Figura 7. Gráfico de cargas del PCA obtenido con el análisis de los metabolitos polares con carga positiva. Los puntos indican los números de identificación (ID) de los metabolitos. Aquellos con un mayor valor en su *loading* de la PC2 y con un menor valor en la PC3 son los tenidos en cuenta.

Tabla 4. Lista de metabolitos polares con carga positiva con más peso en la formación de la PC2 y la PC3 y, por tanto, con un posible papel importante en la respuesta a la infección.

ID	Metabolito	Condición	MZ	t _R	ID	Metabolito	Condición	MZ	t _R
185	Sucrose or isomer	Compatible	382,082	1,27138	5	Caffeoylputrescine	Incompatible	251,139	2,90965
203	<i>no identificado</i>	Compatible	180,102	7,43217	46	Caffeoylputrescine	Incompatible	252,140	2,88669
218	Tetrapeptide	Compatible	490,176	1,24343	49	4-hydroxybenzoylcholine	Incompatible	266,162	4,08459
156	D-fructose 1-pho.	Compatible	293,063	1,27138	2	Tripeptide	Incompatible	348,186	8,01012
454	Sucrose or isomer	Compatible	383,076	1,27138	92	<i>no identificado</i>	Incompatible	696,367	8,10416
65	Tripeptide	Compatible	400,146	1,80907	321	<i>no identificado</i>	Incompatible	266,664	4,25280
489	Sucrose or isomer	Compatible	360,150	1,35038	423	<i>no identificado</i>	Incompatible	164,067	0,95458
202	<i>no identificado</i>	Compatible	411,098	1,23487	233	Caffeoylputrescine	Incompatible	252,141	3,43106
496	<i>no identificado</i>	Compatible	355,108	1,30307	6	<i>no identificado</i>	Incompatible	696,366	8,01012
210	<i>no identificado</i>	Compatible	354,609	1,29922	483	<i>no identificado</i>	Incompatible	500,720	7,54359
23	Sucrose or isomer	Compatible	381,079	1,27138	450	4-hydroxycoumarin	Incompatible	163,039	3,29884
77	<i>no identificado</i>	Compatible	354,108	1,26934	117	<i>no identificado</i>	Incompatible	253,154	2,00132
42	N-coumaroyl tyramine	Compatible	284,128	12,16357	395	<i>no identificado</i>	Incompatible	192,062	0,95600
101	Caffeoyltyramine	Compatible	282,112	9,73919	314	Coumarin	Incompatible	189,073	5,10678
109	Tetrapeptide	Compatible	533,155	1,29922	461	LysoPE(16:1(9Z,12Z))	Incompatible	475,263	7,03089
172	<i>no identificado</i>	Compatible	362,598	1,29922	373	L-tryptophan	Incompatible	206,100	5,17903
390	2-methylcitrate	Compatible	224,076	1,27786	351	<i>no identificado</i>	Incompatible	254,157	1,92907
280	Tetrapeptide	Compatible	448,166	1,27138	29	Coumarin	Incompatible	188,070	5,10678
71	L-Tyrosine	Compatible	182,081	1,79008	21	L-tryptophan	Incompatible	205,097	5,10678
45	<i>no identificado</i>	Compatible	362,096	1,29922	34	LysoPE(16:1(9Z,12Z))	Incompatible	474,259	7,05925
383	Feruloyloctopamine	Compatible	312,123	10,02057	131	LysoPE(16:1(9Z,12Z))	Incompatible	475,262	7,05925
95	Sucrose or isomer	Compatible	381,080	1,16957	41	<i>no identificado</i>	Incompatible	253,154	1,92907
364	Tetrapeptide	Compatible	476,161	1,28396	253	LysoPE(18:3(9Z,12Z))	Incompatible	518,323	20,63712
397	Valine	Compatible	156,042	1,27138	287	1-lynenoyl-sn-glycero-3-phosphate	Incompatible	520,339	22,00680
263	N-coumaroyl tyramine	Compatible	285,131	12,16357	230	LysoPE(18:2(9Z,12Z))	Incompatible	478,292	21,94857

Dicha red de respuesta a la infección compatible aparece representada en la **Figura 8**, integrada además con la red de respuesta del CGA (HE L, *et al.*, 2013). Los cuatro compuestos identificados son metabolitos fenólicos y además son derivados del aminoácido Tyr, que también aparece como inducido en la respuesta compatible.

En cuanto a los metabolitos que caracterizan la infección incompatible encontramos a las cumarinas (ID29, 314 y 450), un grupo de compuestos clasificados como fenilpropanoides y que tienen propiedades antibióticas (OJALA T, *et al.*, 2000). Además, su síntesis deriva del ácido cinámico, lo que podría indicar que ambas infecciones, compatible e incompatible, y a la vista de los resultados anteriores, podrían activar la biosíntesis de compuestos derivados de este ácido.

De hecho, la presencia de *caffeoylputrescine*, en las ID5, 46 y 233, lleva a la misma conclusión al tener como precursor este mismo ácido.

Además, en la lista se encuentran algunos compuestos fosfolipídicos, *LysoPE* (*lysophosphatidylethanolamine*), en las ID34, 131, 253 y 230. Los compuestos de este tipo juegan un papel importante en la señalización celular y la comunicación planta-patógeno, y en la síntesis y ampliación de la membrana celular como respuesta a una infección (SIEBERS M, *et al.*, 2016).

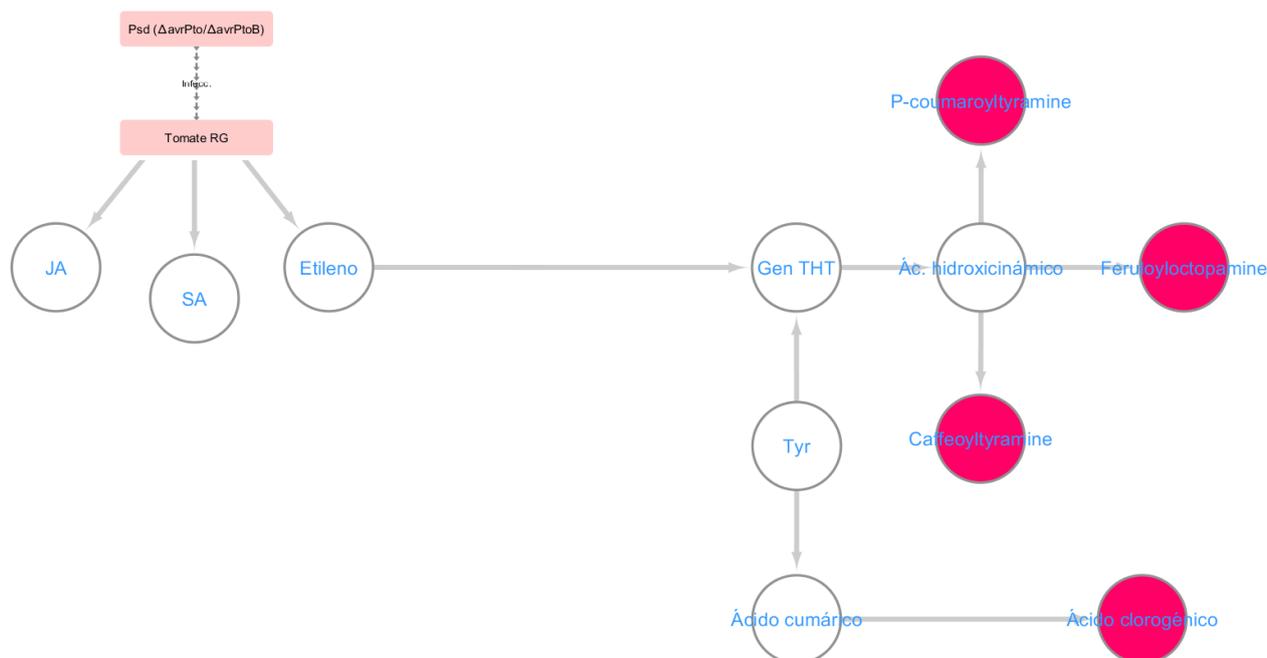


Figura 8. Respuesta de la planta a la infección compatible por *P. syringae* según los resultados de la metabolómica. Los metabolitos identificados aparecen en color rosa.

4.3. Análisis de los metabolitos apolares

Como en los análisis anteriores, con los datos en este caso de 416 compuestos apolares, se realiza un PCA. Sin embargo, el análisis no logró ningún tipo de separación de las condiciones experimentales en las componentes.

Se recurre a realizar después un análisis de regresión por mínimos cuadrados (PLS) como otro sistema para intentar caracterizar la infección. Para realizar el PLS se toma como variable independiente o X la concentración de los metabolitos, mientras que la variable dependiente, variable respuesta o Y son las tres condiciones experimentales cuantificadas, de 0 a 2 (0 para la inoculación *mock*, 1 para la infección incompatible y 2 para la compatible). Tras la regresión PLS, sin embargo, las condiciones experimentales tampoco aparecen diferenciadas.

Como ningún análisis de los planteados lograba separar el tipo de infección en función de la concentración de los metabolitos, se concluye que ni el análisis PCA ni el PLS han permitido explicar las condiciones experimentales en función de la concentración de metabolitos apolares.

4.4. Estudio dirigido a metabolitos apolares potencialmente importantes mediante un análisis de la varianza (ANOVA)

Tras comprobar que los metabolitos apolares no parecen jugar un papel importante en la respuesta de la planta a la infección, desde el IBMCP se midieron las concentraciones de algunos metabolitos potencialmente importantes en la respuesta a la infección en los tres sistemas.

Con los datos de concentración, se agrupan dichos metabolitos según su estructura y se lleva a cabo un análisis de la varianza (ANOVA).

4.4.1. Lípidos

Los resultados del análisis de lípidos aparecen resumidos en la **Tabla 5**.

Las concentraciones de los lípidos considerados no muestran diferencias significativas para ninguna de las condiciones experimentales, al ser en todos los casos el *p*-valor mayor al valor crítico (0,05).

Tabla 5. Resultados del análisis de la varianza (ANOVA) para el análisis dirigido a lípidos.

Lípido		Suma de cuadrados	gl*	Media cuadrática	F	Sig.**
13(E-Z)-docosenamide	Entre grupos	9,216E-05	2	4,608E-05	,647	,584
	Dentro de grupos	,00021	3	7,119E-05		
	Total	,00031	5			
DG(19:1(9Z)-22:0)	Entre grupos	1,492E-06	2	7,461E-07	,320	,748
	Dentro de grupos	6,987E-06	3	2,329E-06		
	Total	8,479E-06	5			
DG(12:0-22:0)	Entre grupos	2,313E-08	2	1,157E-08	,313	,753
	Dentro de grupos	1,110E-07	3	3,698E-08		
	Total	1,341E-07	5			
linoleic acid	Entre grupos	4,278E-05	2	2,139E-05	2,816	,205
	Dentro de grupos	2,279E-05	3	7,597E-06		
	Total	6,557E-05	5			
linolenic acid	Entre grupos	,03884	2	,01942	2,888	,200
	Dentro de grupos	,02017	3	,00672		
	Total	,05902	5			
oxo-heneicosanoic acid	Entre grupos	,01949	2	,00975	,068	,935
	Dentro de grupos	,42758	3	,14253		
	Total	,44707	5			
oxo-heptadecanoic acid	Entre grupos	2,916E-05	2	1,458E-05	,446	,677
	Dentro de grupos	9,806E-05	3	3,269E-05		
	Total	,00013	5			
oxo-nonadenoic acid	Entre grupos	,15413	2	,07706	2,975	,194
	Dentro de grupos	,07770	3	,02590		
	Total	,23183	5			
PS(P-16:0-12:0)	Entre grupos	,01174	2	,00587	4,931	,113
	Dentro de grupos	,00357	3	,00119		
	Total	,01531	5			

*gl: grados de libertad.

**Sig.: nivel de significación o *p*-valor. Solo se consideran en análisis posteriores aquellos metabolitos con un nivel de significación menor a 0,05.

4.4.2. Carotenoides

Los resultados del análisis de los carotenoides aparecen resumidos en la **Tabla 6**.

En este análisis, la concentración del b-caroteno muestra diferencias significativas entre las poblaciones de plantas *mock*, con enfermedad compatible e incompatible. Para determinar en qué población este metabolito difiere se recurre a un análisis *post-hoc*: la prueba de Tukey, cuyos resultados aparecen resumidos en la **Tabla 7**.

Tabla 6. Resultados del ANOVA para el análisis dirigido a carotenoides.

Carotenoide		Suma de cuadrados	gl	Media cuadrática	F	Sig.
a-carotene	Entre grupos	4,106E-09	2	2,053E-09	2,228	,255
	Dentro de grupos	2,764E-09	3	9,215E-10		
	Total	6,870E-09	5			
lutein	Entre grupos	,00039	2	,00020	3,530	,163
	Dentro de grupos	,00017	3	5,580E-05		
	Total	,00056	5			
b-carotene	Entre grupos	1,552E-06	2	7,762E-07	45,381	,006
	Dentro de grupos	5,131E-08	3	1,710E-08		
	Total	1,604E-06	5			
violaxanthin	Entre grupos	2,345E-07	2	1,173E-07	1,039	,454
	Dentro de grupos	3,385E-07	3	1,128E-07		
	Total	5,731E-07	5			
neoxanthin	Entre grupos	2,406E-08	2	1,203E-08	,343	,734
	Dentro de grupos	1,052E-07	3	3,507E-08		
	Total	1,293E-07	5			

Tabla 7. Resultados del análisis *post-hoc* para el b-caroteno utilizando la prueba de Tukey.

Variable dependiente	Diferencia de medias (I-J)	Error estándar	Sig.	95% de interv. de confianza				
				Límite inferior	Límite superior			
b-carotene	Mock	Sist.	-,00112544*	,00013078	,007	-,0016719	-,0005789	
		HR	-,00009972	,00013078	,748	-,0006462	,0004468	
		Sist.	Mock	,00112544*	,00013078	,007	,0005789	,0016719
			HR	,00102572*	,00013078	,009	,0004792	,0015722
		HR	Mock	,00009972	,00013078	,748	-,0004468	,0006462
			Sist.	-,00102572*	,00013078	,009	-,0015722	-,0004792

Los resultados del análisis *post-hoc* muestran que la concentración del b-caroteno está alterada en las muestras de plantas con la enfermedad compatible, existiendo diferencias entre compatible, con mock y HR. No hay diferencias significativas entre las plantas control y las que tienen la infección incompatible.

El b-caroteno es un pigmento de coloración rojiza abundante en plantas y frutos, perteneciente al grupo de los carotenos (terpenos). Este compuesto tiene una importante función antioxidante, que reacciona con ROS e impide que ejerzan efectos nocivos (MCELROY J, KOPSELL D, 2009), por lo que su concentración depende generalmente de factores externos como las invasiones de patógenos.

Además, el b-caroteno es un precursor del ácido abscísico, ABA, (CAZZONELLI C, POGSON B, 2010), una importante fitohormona que regula la respuesta y la tolerancia de la planta ante diversos estreses bióticos y abióticos (CAO F, *et al.*, 2011). Sin embargo, la síntesis de ABA no parece que sea la finalidad del b-caroteno, ya que en la ruta biosintética de este ácido aparecen intermediarios como la violaxanthin o neoxanthin, compuestos que no han mostrado ninguna diferencia significativa en el ANOVA.

4.4.3. Clorofilas

Los resultados obtenidos para las clorofilas aparecen resumidos en la **Tabla 8**.

Como se ha hecho con el análisis de carotenos, se recurre a un análisis *post-hoc* con la prueba de Tukey para identificar las diferencias significativas entre medias para cada metabolito. Se analizan los metabolitos *Pheophorbide b* (p -valor = 0,015) y *Chlorophyll a* (p -valor = 0,05) por estar por debajo o igualar el nivel de significación establecido en $p < 0,05$ (**Tabla 9**).

Tabla 8. Resultados del ANOVA para el análisis dirigido a clorofilas.

Clorofila		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Chlorophyll a	Entre grupos	,82614	2	,41307	9,423	,050
	Dentro de grupos	,13150	3	,04383		
	Total	,95764	5			
Chlorophyll b	Entre grupos	,01628	2	,00814	1,682	,324
	Dentro de grupos	,01452	3	,00484		
	Total	,03079	5			
Chlorophyllide a	Entre grupos	2,992E-05	2	1,496E-05	1,681	,324
	Dentro de grupos	2,670E-05	3	8,900E-06		
	Total	5,662E-05	5			
Chlorophyllide b	Entre grupos	2,269E-05	2	1,134E-05	,893	,496
	Dentro de grupos	3,810E-05	3	1,270E-05		
	Total	6,079E-05	5			
Divynilprotochlorophyllide a	Entre grupos	3,734E-09	2	1,867E-09	,841	,513
	Dentro de grupos	6,658E-09	3	2,219E-09		
	Total	1,039E-08	5			
Divynilprotochlorophyllide b	Entre grupos	2,089E-05	2	1,044E-05	1,140	,428
	Dentro de grupos	2,747E-05	3	9,158E-06		
	Total	4,836E-05	5			
Pheophorbide a	Entre grupos	,05753	2	,02877	,891	,497
	Dentro de grupos	,09690	3	,03230		
	Total	,15443	5			
Pheophorbide b	Entre grupos	4,633E-06	2	2,316E-06	22,933	,015
	Dentro de grupos	3,030E-07	3	1,010E-07		
	Total	4,936E-06	5			
Pheophytin a	Entre grupos	,63099	2	,31550	,633	,590
	Dentro de grupos	1,49492	3	,49831		
	Total	2,12591	5			
Pheophytin b	Entre grupos	,00386	2	,00193	2,136	,265
	Dentro de grupos	,00271	3	,00090		
	Total	,00656	5			
Primary fluorescent chlorophyll catabolite	Entre grupos	6,616E-08	2	3,308E-08	,665	,577
	Dentro de grupos	1,492E-07	3	4,975E-08		
	Total	2,154E-07	5			
Pyropheophorbide a	Entre grupos	,03218	2	,01609	4,638	,121
	Dentro de grupos	,01041	3	,00347		
	Total	,04259	5			

La prueba *post-hoc* muestra que la clorofila se encuentra en una concentración mayor en las plantas que sufren una infección compatible respecto a mock y HR, sin diferencias significativas; y que el pheophorbide b se encuentra en una menor concentración en las plantas infectadas con una infección incompatible (HR) respecto a mock y compatible, que no tienen diferencias entre sí.

La clorofila A es el pigmento mayoritario en las hojas de la planta del tomate y es la causante de su color. Además de intervenir en el desarrollo de la fotosíntesis al estar acoplada a los fotosistemas, es un potente antioxidante (LANFER-MARQUEZ U, *et al.*, 2005). Esto puede explicar que se encuentre en mayor proporción en las plantas que han sufrido una infección compatible.

El pheophorbide es un compuesto antioxidante obtenido por la degradación de la clorofila, cuya acumulación puede producir muerte celular independiente de la luz (HIRASHIMA M, *et al.*, 2009). El hecho de que en plantas con infección incompatible no se acumule este compuesto explica que la planta resista a la infección.

Tabla 9. Resultados del análisis *post-hoc* para la clorofila a y el pheophorbide b utilizando la prueba de Tukey.

Variable dependiente	Diferencia de medias (I-J)	Error estándar	Sig.	95% de interv. de confianza		
				Límite inferior	Límite superior	
Chlorophyll a	Mock Sist.	-,79108368	,20936598	,065	-1,6659650	,0837977
	HR	-,00792935	,20936598	,999	-,8828107	,8669520
	Sist. Mock	,79108368	,20936598	,065	-,0837977	1,6659650
	HR	,78315433	,20936598	,066	-,0917270	1,6580357
	HR Mock	,00792935	,20936598	,999	-,8669520	,8828107
	Sist.	-,78315433	,20936598	,066	-1,6580357	,0917270
Pheophorbide b	Mock Sist.	-,00016863	,00031782	,863	-,0014967	,0011595
	HR	,00177401*	,00031782	,023	,0004459	,0031021
	Sist. Mock	,00016863	,00031782	,863	-,0011595	,0014967
	HR	,00194264*	,00031782	,018	,0006146	,0032707
	HR Mock	-,00177401*	,00031782	,023	-,0031021	-,0004459
	Sist.	-,00194264*	,00031782	,018	-,0032707	-,0006146

4.4.4. Quinonas

Los resultados del análisis de las quinonas aparecen resumidos en la **Tabla 10**.

Las quinonas, como los lípidos, tampoco muestran diferencias significativas en sus concentraciones entre las diferentes poblaciones, al no superar en ningún caso el umbral crítico del p -valor $< 0,05$. Sin embargo, el compuesto menaquinone 8 muestra un nivel de significación cercano al nivel crítico, por lo que las diferencias entre las medias son prácticamente significativas. Por ello, se realiza la prueba de Tukey (**Tabla 11**).

La menaquinone 8 se encuentra a una concentración más baja en las plantas con una respuesta incompatible a la infección. La menaquinone 8 es un tipo de vitamina K, cuyas propiedades como nutriente son muy interesantes: tienen un gran poder antioxidante, se especula con su potencial antineoplásico, ... (GRÖBER U, *et al.*, 2014) Sin embargo, el papel que puede ofrecer dicho compuesto a este tipo de estrés no se encuentra definido en la bibliografía.

Tabla 10. Resultados del ANOVA para el análisis dirigido a quinonas.

Quinonas		Suma de cuadrados	gl	Media cuadrática	F	Sig.
a-tocopherol quinone	Entre grupos	2,383E-08	2	1,191E-08	,516	,642
	Dentro de grupos	6,933E-08	3	2,311E-08		
	Total	9,315E-08	5			
Menaquinone 8	Entre grupos	7,258E-08	2	3,629E-08	8,642	,057
	Dentro de grupos	1,260E-08	3	4,199E-09		
	Total	8,518E-08	5			
phyloquinone	Entre grupos	,00023	2	,00011	1,793	,307
	Dentro de grupos	,00019	3	6,412E-05		
	Total	,00042	5			
Plastochromanol 8	Entre grupos	1,483E-09	2	7,417E-10	,958	,477
	Dentro de grupos	2,323E-09	3	7,745E-10		
	Total	3,807E-09	5			
Plastoquinol 9	Entre grupos	9,589E-10	2	4,795E-10	,873	,503
	Dentro de grupos	1,648E-09	3	5,495E-10		
	Total	2,607E-09	5			
Ubiquinone 10	Entre grupos	4,609E-07	2	2,305E-07	,486	,657
	Dentro de grupos	1,424E-06	3	4,747E-07		
	Total	1,885E-06	5			

Tabla 11. Resultados del análisis *post-hoc* para la menaquinone 8 utilizando la prueba de Tukey.

Variable dependiente	Diferencia de medias (I-J)	Error estándar	Sig.	95% de interv. de confianza		
				Límite inferior	Límite superior	
Menaquinone 8	Mock Sist.	,00004304	,00006480	,798	-,0002277	,0003138
	HR	,00025184	,00006480	,060	-,0000189	,0005226
	Sist. Mock	-,00004304	,00006480	,798	-,0003138	,0002277
	HR	,00020880	,00006480	,095	-,0000620	,0004796
	HR Mock	-,00025184	,00006480	,060	-,0005226	,0000189
	Sist.	-,00020880	,00006480	,095	-,0004796	,0000620

5.

RESULTADOS Y DISCUSIÓN DEL ANÁLISIS PROTEÓMICO

5. Resultados y discusión del análisis proteómico

Tras el análisis metabolómico, se procede al análisis del proteoma de las plantas de tomate. Como anteriormente, se lleva a cabo un análisis multivariante (en concreto un PCA) para comprobar qué proteínas son importantes en la defensa de la planta ante la infección por *P. syringae*.

El proteoma completo que se considera para este análisis contiene 1107 proteínas, a partir de las que se realiza un PCA. En este caso, se reduce el número de muestras respecto al análisis metabolómico: se consideran 18h y 36h post-inoculación en las tres condiciones.

Se consideran las dos primeras componentes del modelo, que en este caso explican el 27,9% y el 20,2% de la varianza de los datos: en total, ambas explican el 48,1% de la varianza total.

En el gráfico de puntuaciones (**Figura 9**) se observa que la PC1 explica, además del tiempo post-inoculación (18 y 36 horas), la diferencia existente en el proteoma de plantas que no han tenido contacto con la bacteria (*mock*) y las que sí lo han tenido (incompatibles y compatibles).

La PC2, que será la que se considere en la posterior identificación al ser la más importante, explica la respuesta defensiva de la planta. Esta componente separa la respuesta defensiva hipersensible y la enfermedad compatible que sufren aquellas plantas que están infectadas por la cepa de *P. syringae* con la delección en su gen de virulencia.

A partir del gráfico de cargas (**Figura 10**) se seleccionan aquellas proteínas con un mayor valor absoluto en el *loading* de la PC2. Se seleccionan alrededor de un 5% de proteínas con un mayor valor positivo (50 proteínas), y la misma cantidad de proteínas con mayor valor negativo.

En la **Tabla 12** aparecen recogidas 50 proteínas representativas para cada tipo de infección.

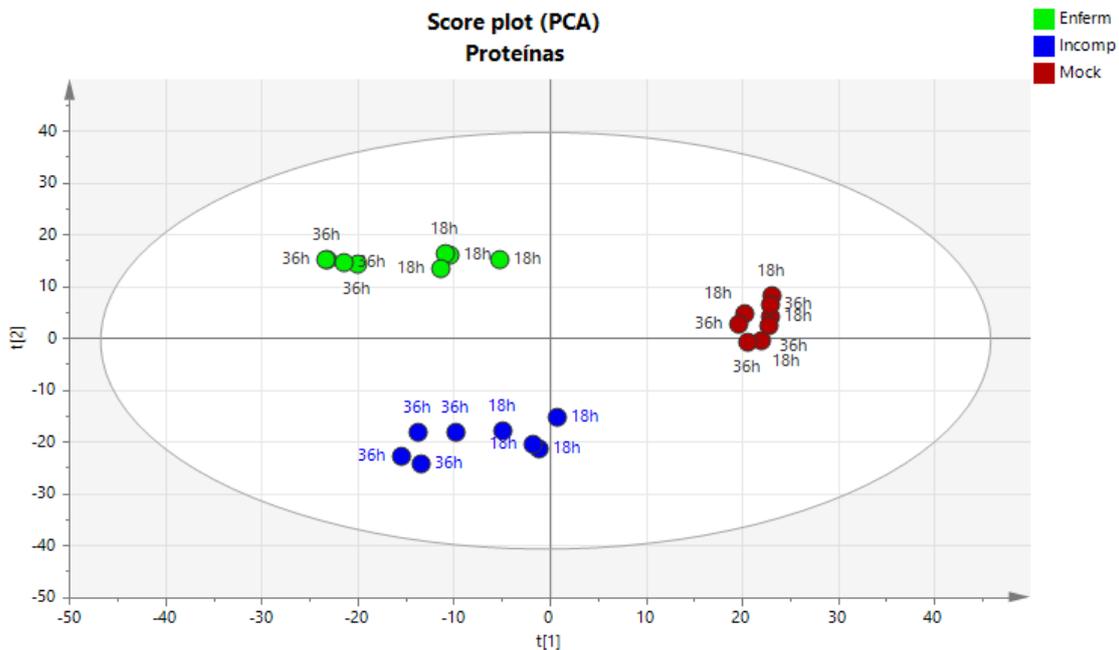


Figura 9. Gráfico de puntuaciones del PCA obtenido con el análisis del proteoma de las plantas de tomate. En rojo, las plantas inoculadas *mock*; en verde, las inoculadas con una cepa virulenta, con delección; y, en azul, las inoculadas con una cepa silvestre y que producen una infección incompatible.

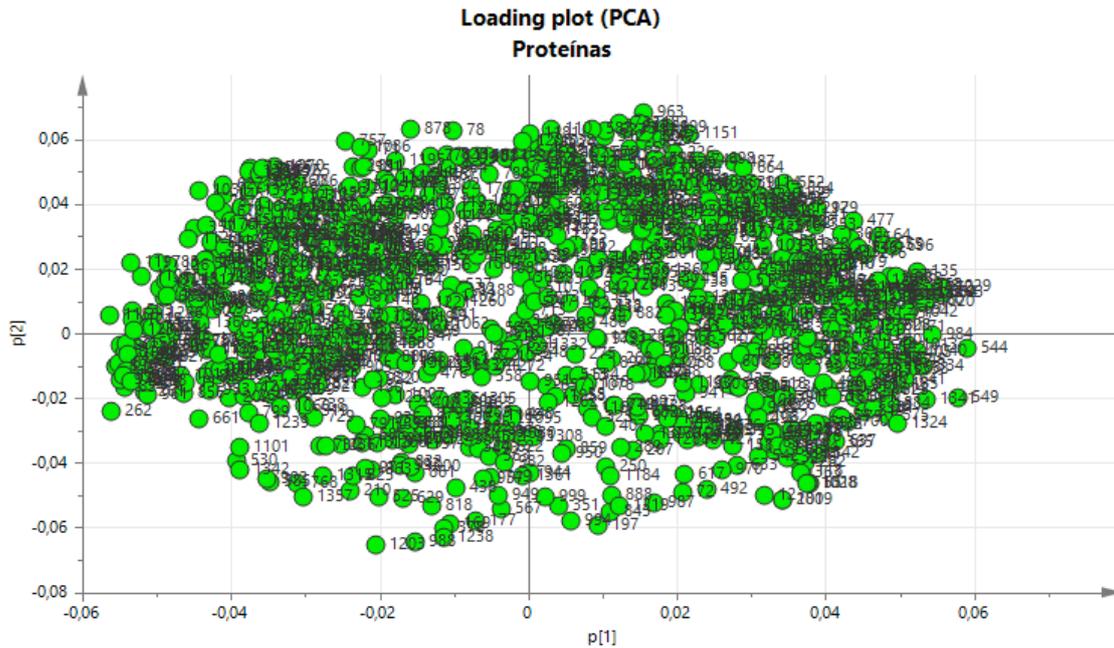


Figura 10. Gráfico de cargas del PCA obtenido con el análisis del proteoma de las plantas de tomate. Los puntos indican los números de identificación (ID) de los metabolitos. Aquellos con más valor absoluto en la segunda componente son los que se tienen en cuenta en el análisis.

Entre las proteínas más importantes en la infección compatible aparece la Rubisco (ribulosa 1,5 carboxilasa oxigenasa), con un papel fundamental en la fotosíntesis. Sin embargo, generalmente la presencia de este complejo enzimático se debe a posibles contaminaciones debido a que está presente en grandes cantidades en la planta.

En la infección incompatible, la planta resistente, se han caracterizado dos proteínas. La aleno óxido ciclasa (AOC) es una proteína que interviene en la síntesis de ácido jasmónico (JA) en la planta (RIEMANN M, *et al.*, 2013), molécula señal de una infección. Aunque *P. syringae* es un patógeno biótrofo, el JA comúnmente es conocido como marcador de patógenos necrótrofos (GLAZEBROOK J, 2005).

La shikimato kinasa, por su parte, es una enzima que interviene en la biosíntesis del ácido shikímico, precursor de todos los compuestos fenólicos y aminoácidos aromáticos de la planta. El shikímico también es el precursor del ácido salicílico (SA), molécula que juega un papel fundamental en la repuesta a patógenos biótrofos (HAVAT S, *et al.*, 2013).

Por su parte, la *Oxygen evolving enhancer protein 2* (OEE2) parece estar implicada en la producción de especies reactivas de oxígeno (ROS) como H_2O_2 o superóxidos en los cloroplastos en determinados estreses, entre ellos el estrés biótico. Esta producción y acumulación de ROS es propia de la respuesta de las plantas ante infecciones incompatibles (YANG E, *et al.*, 2003).

En resumen, las proteínas identificadas como sobreexpresadas en la infección incompatible tienen como función la respuesta a estrés biótico, ya sea sintetizando moléculas señalizadoras (como el SA o el JA) o produciendo ROS, que causan la muerte del microorganismo en las zonas de infección.

Tabla 12. Lista de proteínas con más peso en la formación de la PC2 y, por tanto, con un posible papel en la defensa contra la infección.

Loading positivo (compatible)			Loading negativo (incompatible)		
ID	Loading	Proteína	ID	Loading	Proteína
963	0,0683891	(No secuenciada)	1203	-0,0651181	(No secuenciada)
732	0,065057		988	-0,0644212	
1182	0,0647757		1238	-0,0631863	
613	0,0638091		378	-0,0603072	
969	0,0635717		197	-0,0593501	
199	0,0635413		169	-0,0586888	
878	0,0629721		994	-0,0580093	
13	0,0629461		177	-0,057999	
587	0,062916		843	-0,0551005	
110	0,062912		567	-0,0540019	
527	0,0628605		1219	-0,0533137	
78	0,0626737		818	-0,0533041	
87	0,0626599		351	-0,0531823	
1125	0,0623958		987	-0,0525318	
1181	0,061817		1019	-0,0514238	
1151	0,0616077		280	-0,0514216	
20	0,0608162		629	-0,0509923	
312	0,0607102		525	-0,0505971	
240	0,0607087		999	-0,0504714	
252	0,0598075		1357	-0,0503494	
538	0,0596086		949	-0,0502255	
652	0,059367		12	-0,0499893	
757	0,0593015		888	-0,0498986	
1283	0,0592741		72	-0,0489028	
1024	0,0587733		210	-0,0487309	
1086	0,0575578		492	-0,0483548	
650	0,057114		438	-0,0478679	
131	0,0570412		768	-0,0464735	
591	0,0568768		1018	-0,046358	
71	0,0568627		1028	-0,0462311	
673	0,0562238		55	-0,0461569	
593	0,0558576	Subunidad de RuBisCo.	585	-0,0458768	
226	0,0558277		902	-0,0451551	
579	0,0557534		957	-0,0447134	
1025	0,0556409		749	-0,0445873	
822	0,0554676		1361	-0,0442073	
881	0,0549947		1315	-0,043992	
776	0,054853		1184	-0,0439564	
230	0,0548497		825	-0,0439141	
1143	0,054829		617	-0,0435065	
602	0,054785		363	-0,0430447	
924	0,0547397		944	-0,0429883	-oxygen-evolving enhancer protein 2, chloroplastic -probable shikimate kinase like 1, chloroplastic-like
17	0,0546712		861	-0,0429812	
301	0,0545325		118	-0,0426563	
499	0,0542491		970	-0,0421599	
620	0,054096		863	-0,0421358	
99	0,0540399		1342	-0,0421037	
255	0,0539463		971	-0,0419083	
692	0,0538302		1000	-0,0413021	Allene oxide cyclase

6.

RESULTADOS Y DISCUSIÓN DE LA INTEGRACIÓN DE PROTEÓMICA Y METABOLÓMICA

6. Resultados y discusión de la integración de proteómica y metabolómica

Tras analizar los metabolitos y las proteínas por separado, se procede a realizar un análisis global de las proteínas y metabolitos utilizando una regresión PLS modificada, llamada O2PLS. Esto permitirá conocer si los cambios de concentraciones de metabolitos de la planta se correlacionan con los cambios en las concentraciones de proteínas en la respuesta celular.

Para ello, en primer lugar, se realiza un PLS de comprobación conjunto, con los datos de todas las proteínas y todos los metabolitos (variables X e Y, respectivamente). Los resultados se presentan en el gráfico de puntuaciones (**Figura 11**), donde se puede observar que efectivamente existe una correlación entre ambas variables que se relaciona con las condiciones experimentales.

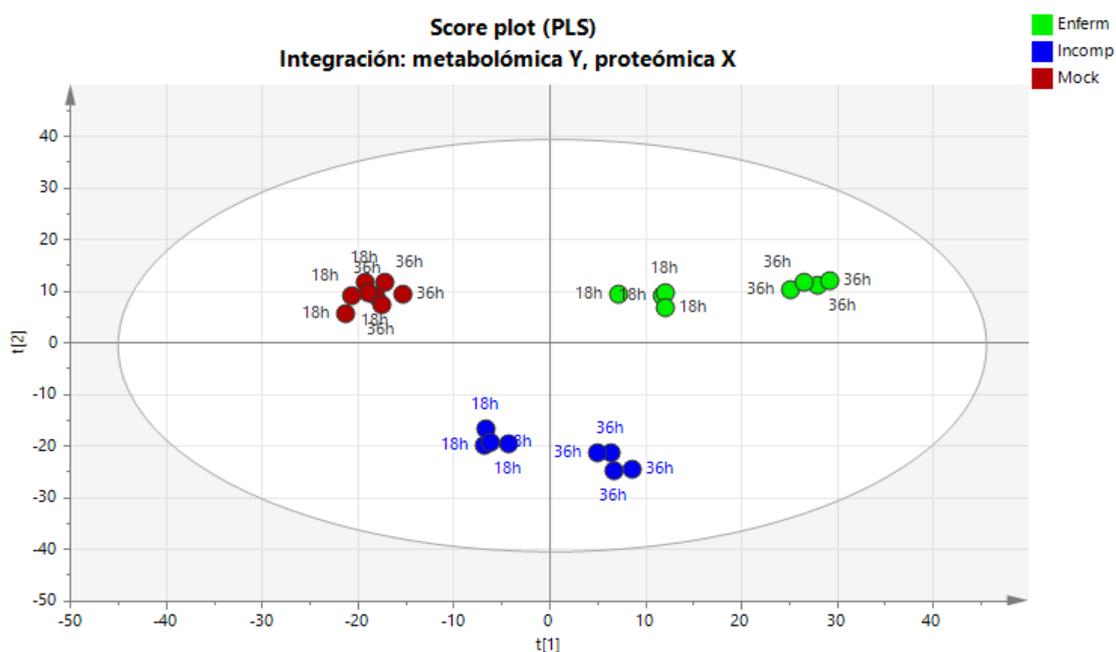


Figura 11. Gráfico de puntuaciones del PLS de integración obtenido con el análisis conjunto del proteoma (X) y del metaboloma (Y) de las plantas de tomate. En rojo, las plantas inoculadas *mock*; en verde, las inoculadas con una cepa virulenta, con delección; y, en azul, las inoculadas con una cepa silvestre y que producen una infección incompatible.

Al observarse que existe dicha correlación entre proteínas y metabolitos, se procede a estudiar por separado la integración de proteínas con los metabolitos polares y con los apolares.

6.1. Integración de proteínas y metabolitos apolares

El análisis de la regresión del tipo O2PLS en el gráfico de puntuaciones (**Figura 12**) muestra que los metabolitos apolares y las proteínas se correlacionan en las condiciones experimentales (PC1) y en el tiempo post-inoculación (PC2).

Esta separación ha tenido lugar analizando los metabolitos apolares y las proteínas en conjunto, pero el análisis individual por PCA y PLS de los metabolitos apolares no permitió caracterizar ninguna condición experimental.

Tras la integración, esta separación sí ha sucedido. Esto puede explicarse observando el gráfico de cargas (**Figura 13**), donde se observa que las relaciones más importantes entre

proteínas y metabolitos apolares se dan en la infección incompatible, y no tanto en la compatible. Es la ausencia de las proteínas en el análisis de los metabolitos apolares aisladamente, por tanto, la que ha causado que la separación de las condiciones experimentales no se produjese, lo que refleja la importancia del enfoque de realizar un análisis integrado de ómicas.

La varianza explicada por el modelo O2PLS aparece resumida en la **Tabla 13**.

Como se puede observar en la variación explicada por cada componente, el modelo realizado con 2 componentes permite explicar el 51,3% de la variación en la proteómica (R^2_X) y el 56,6% de la variación en la fracción de metabolitos apolares (R^2_Y).

De esta variación total, la correlación entre proteínas y metabolitos explica alrededor de un 77% de la variación total de X ($R^2_{X(p. \text{ correlación})}/R^2_X$), y un 83% de la variación total de Y ($R^2_{Y(p. \text{ correlación})}/R^2_Y$).

Por tanto, con este modelo O2PLS aplicado a los datos de proteínas y compuestos apolares, se puede asegurar que la mayor parte de variación presente en los datos puede explicarse por la correlación entre la proteómica y la metabolómica.

La componente que se considere para establecer las correlaciones será la PC1, ya que separa las diferentes condiciones experimentales y permite diferenciar entre plantas enfermas (valor de *loading* positivo) y plantas cuya respuesta ha hecho que la infección sea incompatible (valor de *loading* negativo).

Aproximadamente el 5% de metabolitos y proteínas más representativos se presenta en la **Tabla 14**.

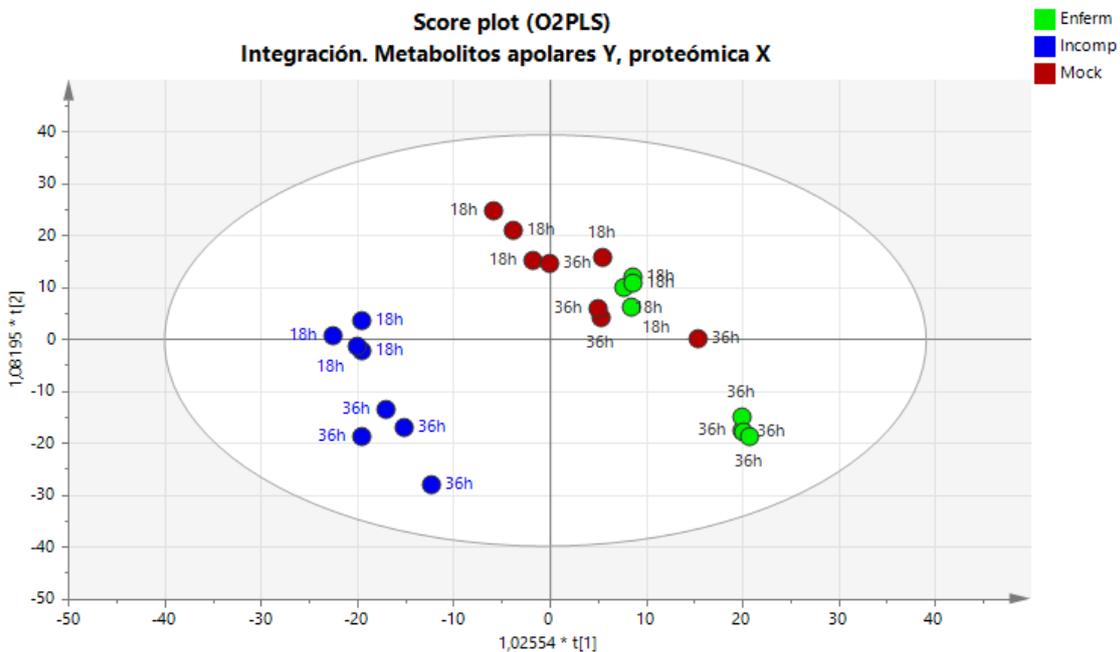


Figura 12. Gráfico de puntuaciones del O2PLS de integración obtenido con el análisis conjunto del proteoma (X) y de la fracción apolar del metaboloma (Y) de las plantas de tomate. En rojo, las plantas inoculadas *mock*; en verde, las inoculadas con una cepa virulenta, con delección; y, en azul, las inoculadas con una cepa silvestre y que producen una infección incompatible.

Tabla 13. Variación desglosada explicada por cada una de las dos primeras componentes del modelo O2PLS de integración de metabolitos apolares y proteínas.

Componente	R ² _X	R ² _X (p. correlación)	R ² _Y	R ² _Y (p. correlación)
Comp. 1	0,271	0,201	0,408	0,348
Comp. 2	0,242	0,194	0,158	0,121
Total (modelo)	0,513	0,395	0,566	0,469

En la tabla, de izquierda a derecha: la componente de modelo, R²_X la variación de X explicada por cada componente, R²_X (p. correlación) la variación de X debida a la correlación entre X e Y, y R²_Y la variación de Y explicada por cada componente, R²_Y (p. correlación) la variación de Y debida a la correlación entre X e Y.

Tabla 14. Lista de metabolitos apolares y proteínas con más peso en la formación de la PC1 en el modelo O2PLS y, por tanto, posiblemente correlacionados entre sí.

Infección incompatible					Infección compatible				
Proteínas		Metabolitos apolares			Proteínas		Metabolitos apolares		
ID	Proteína	ID +/-	MZ	t _R	ID	Proteína	ID +/-	MZ	t _R
1238	(No sec.)	212+	585,449	5,64961	230	(No sec.)	40-	515,244	2,97377
988		99+	613,480	5,01489	822		192+	300,121	0,74178
1203		189+	614,483	5,01489	692		21-	593,270	2,97377
994		190+	335,256	5,01489	650		194+	517,258	2,97943
197		217+	405,262	0,93383	301		320+	330,132	0,74178
169		313+	405,261	4,35278	75		94-	534,260	2,97377
351		221+	423,272	4,35278	1195		176+	298,106	0,74178
378		177+	586,451	4,35278	673		59-	533,255	2,97377
818		86+	585,448	4,35278	776		62+	284,127	0,77408
177		306+	353,267	2,33988	252		115-	312,124	0,77551
843		247-	1095,770	5,28202	579		44-	282,113	0,78992
629		298+	567,438	4,35278	638		233+	285,130	0,77408
567		142+	793,561	2,30627	1204		118+	282,111	0,74178
		76+	792,558	2,30627	50		112-	298,108	0,77551
		228-	1067,738	4,71160	1181		4+	593,273	2,96831
		243+	307,225	4,35278	881		18+	594,276	2,97943
		155+	423,272	0,93383	981		297+	536,271	2,94917
		60-	1065,723	4,35806	664		135+	535,268	2,94917
		81-	1066,725	4,34487	1198		161+	314,137	0,77408
		302+	337,272	3,36511	924		77+	595,279	2,97943
		185-	1067,729	4,34487	201		238+	558,530	5,40346
		100-	1093,753	5,00943	499		121+	557,526	5,40346
		126-	1094,756	5,00943	594	Eukaryotic transl. factor			
		22+	353,267	1,03193	1086				
		286+	587,464	2,06633					
		80+	354,271	1,03193					
		259-	812,560	3,41430					
		135-	811,556	3,41430					
		209-	1044,742	5,66689					
		142-	1043,739	5,66689					
		281-	1072,773	6,05515					
		141-	249,185	4,34487					
		171-	277,217	3,58887					
		202-	1071,769	6,01971					
		90+	335,256	3,19439					
		280+	336,259	3,19439					
		58-	277,216	5,00943					
		226-	278,220	5,00943					

En la tabla, de izquierda a derecha: datos de la infección incompatible y compatible. Para cada una, se señalan las proteínas, con su número ID en el PAGE y con su identificación si se ha secuenciado; y los metabolitos, con su ID, su carga (+ positiva y - negativa), su relación masa-carga y su tiempo de retención en la columna de cromatografía.

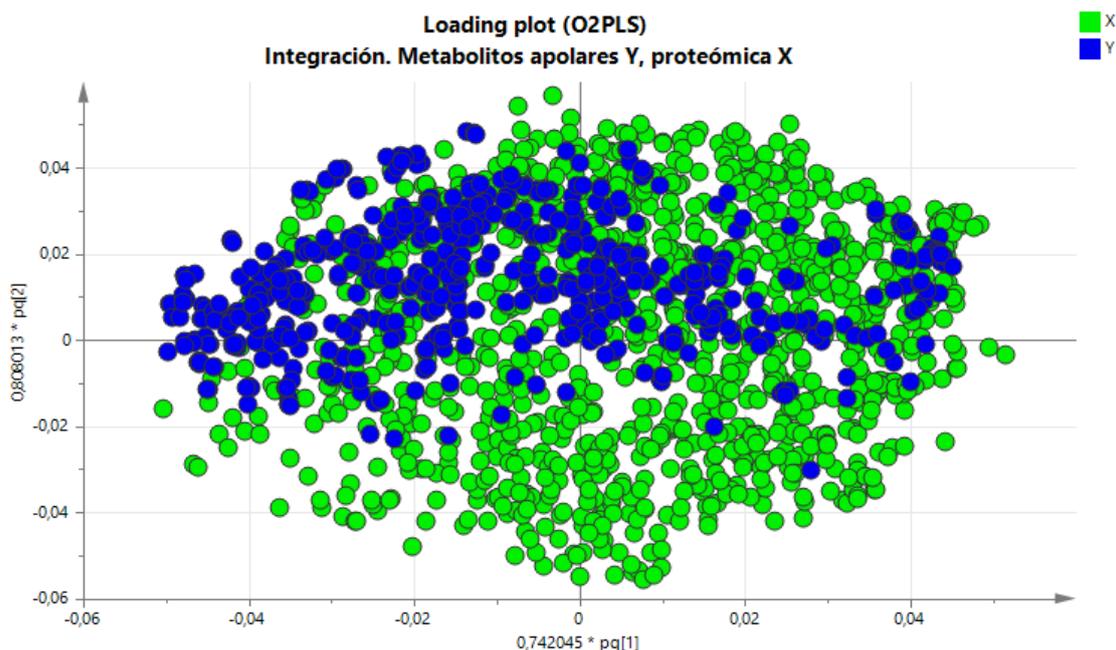


Figura 13. Gráfico de cargas del O2PLS de integración obtenido con el análisis conjunto del proteoma (X) y de la fracción apolar del metaboloma (Y) de las plantas de tomate. En azul, se representan los metabolitos apolares; y, en verde, las proteínas.

Pese a que los resultados han sido muy prometedores, no ha sido posible dar con la identificación de los metabolitos apolares que el análisis integrado ha sacado a la luz.

Sin embargo, y dado que el enfoque de este trabajo es proponer un método de análisis integrado de ómicas, proponemos a los investigadores e investigadoras del IBMCP la identificación de dichos metabolitos para averiguar, sobre todo, cómo actúan en la respuesta que ofrece la planta a una infección incompatible.

6.2. Integración de proteínas y metabolitos polares

El último análisis de integración O2PLS involucra a las proteínas (X) y a los metabolitos polares, con cargas positivas y negativas (Y).

El gráfico de puntuaciones (**Figura 14**) muestra que también se ha producido una separación de las condiciones experimentales: la PC1, además de las horas post-inoculación, parece separar las plantas *mock* de las que han sido inoculadas con la bacteria. Por su parte, la PC2 separa las plantas con infección compatible (*loadings* positivos) y con incompatible (*loadings* negativos).

La variación explicada por el modelo O2PLS aparece resumida en la **Tabla 15**.

Tabla 15. Variación desglosada explicada por cada una de las dos primeras componentes del modelo O2PLS de integración de metabolitos polares y proteínas.

Componente	R^2_X	$R^2_{x(p. \text{ correlación})}$	R^2_Y	$R^2_{y(p. \text{ correlación})}$
Comp. 1	0,338	0,273	0,285	0,239
Comp. 2	0,223	0,181	0,184	0,152
Total (modelo)	0,561	0,454	0,469	0,391

En este caso, las dos componentes explican el 56,1% y el 46,9% de la variación de la proteómica (R^2_X) y la metabolómica polar (R^2_Y), respectivamente. Además, la correlación entre proteínas y compuestos polares del metaboloma explica un 81% de la variación total de X ($R^2_{X(p. \text{ correlación})}/R^2_X$), y un 83% de la variación total de Y ($R^2_{Y(p. \text{ correlación})}/R^2_Y$).

Es la PC2 la que se considerará para estudiar la infección. Tras la obtención del gráfico de cargas (**Figura 15**), se seleccionan aquellos metabolitos y proteínas más influyentes (**Tabla 16**) para intentar proceder a la identificación.

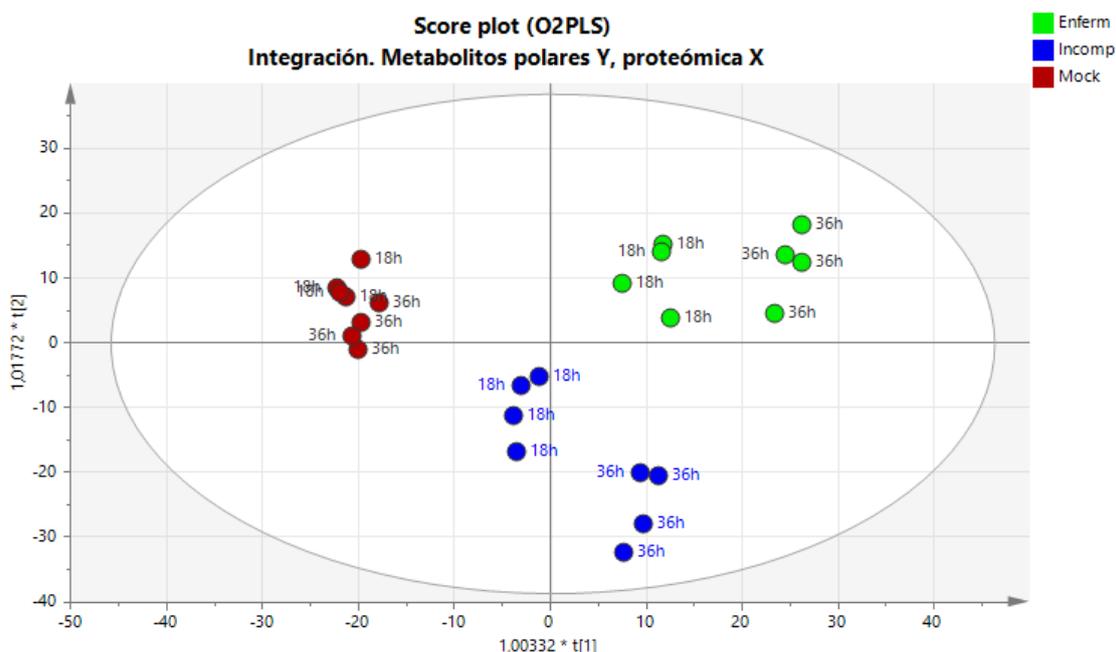


Figura 14. Gráfico de puntuaciones del O2PLS de integración obtenido con el análisis conjunto del proteoma (X) y de la fracción polar del metaboloma (Y) de las plantas de tomate. En rojo, las plantas inoculadas *mock*; en verde, las inoculadas con una cepa virulenta, con delección; y, en azul, las inoculadas con una cepa silvestre y que producen una infección incompatible.

Como se observa en la **Tabla 16**, muchos metabolitos identificados para la infección compatible se habían detectado previamente en los análisis individuales por PCA.

Los metabolitos identificados para este tipo de infección son, en su mayoría, azúcares (sacarosa o isómeros), tal y como podemos observar con las identificaciones en las **Tablas 3 y 4**. De hecho, comparten esta identificación los compuestos con carga positiva con ID 489, 95, 23 y 454; y el ID 24 de los que tienen carga negativa.

Esto apoya la hipótesis de que en la infección compatible el metabolismo de azúcares juega un papel fundamental, como se había expuesto en la discusión de resultados de metabolitos polares.

Además, en la lista de proteínas aparece la fosfoglicerato kinasa, una enzima clave en la glicólisis, la principal vía de metabolismo de azúcares, que cataliza la transferencia reversible de un grupo fosfato del ADP al ATP, pasando del 1,3-bisfosfoglicerato al 3-fosfoglicerato. Ambos compuestos son importantes intermediarios de esta ruta metabólica e intervienen en procesos relacionados con la fotosíntesis como el ciclo de Calvin, por lo que también pueden tener un papel relacionado con el metabolismo energético de la planta (DIXON WL, KEES T, 1980).

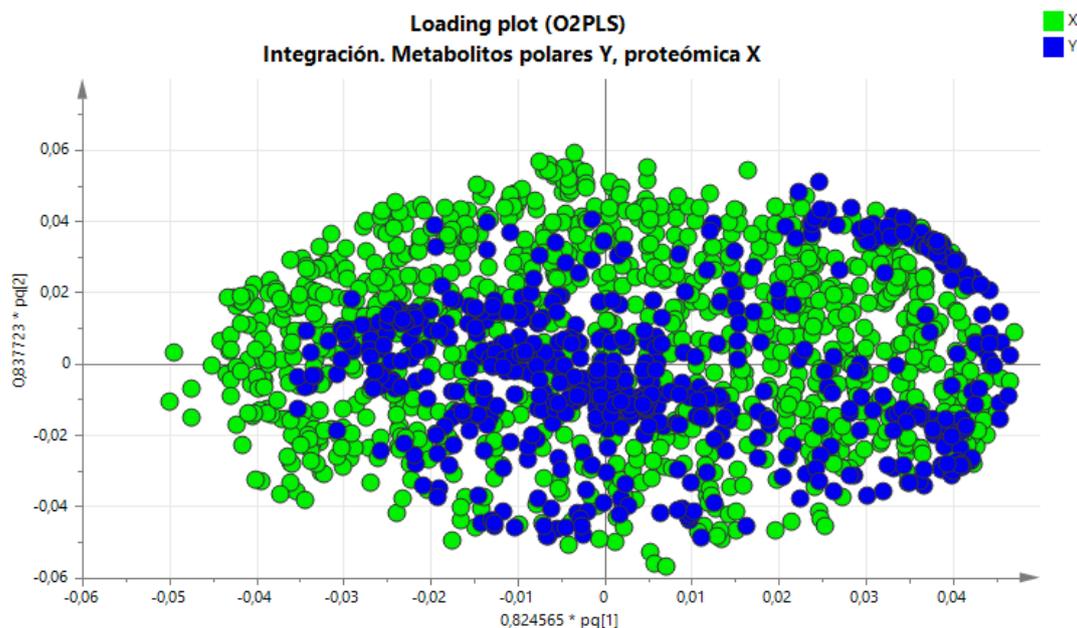


Figura 15. Gráfico de cargas del O2PLS de integración obtenido con el análisis conjunto del proteoma (X) y de la fracción polar del metaboloma (Y) de las plantas de tomate. En azul, se representan los metabolitos apolares; y, en verde, las proteínas.

Tabla 16. Lista de metabolitos polares y proteínas con más peso en la formación de la PC1 en el modelo O2PLS y, por tanto, posiblemente correlacionados entre sí.

Infección incompatible					Infección compatible				
Proteínas		Metabolitos polares			Proteínas		Metabolitos polares		
ID	Proteína	ID +/-	MZ	tr	ID	Proteína	ID +/-	MZ	tr
1203	(No sec.)	102+	253,153	1,89779	1086	(No sec.)	280+	448.165	1,27138
988		32+	255,944	0,97313	878		342+	444.139	1,26570
378		198+	385,902	0,97313	1195		378+	489.193	1,24645
1238		174+	214,917	0,93366	757		499+	525.667	1,26570
169		129+	232,928	0,97313	581		489+	360.149	1,35038
197		8+	167,013	0,98981	71		364+	476.161	1,28396
177		65-	548,851	0,97982	213		325+	273.582	1,23158
863		14-	304,913	0,97306	1370		486+	325.113	1,30506
993		270+	204,000	1,40549	1145		97+	273.081	1,26570
831		25-	418,893	0,96732	760		302+	365.115	1,55987
1357		46+	252,140	2,88669	714		95+	381.080	1,16957
971		84+	296,970	1,02513	1335		128+	452.137	1,29922
825		46-	434,870	0,97306	27		236+	192.055	1,24117
818		43+	239,966	0,98981	408	Phospho-glycer. kinase	23+	381.079	1,27138
888		197+	273,954	1,02513	811		51-	487.176	1,21960
832		12-	288,935	0,96732	666		24-	377.084	1,51025
351		171+	369,924	1,02633	873		13-	353.087	6,97588
994		350+	282,990	1,02513	1122		454+	383.076	1,27138
768		357+	241,964	0,97313	1282		208+	365.105	1,26156
949		182+	218,983	1,44256	669	RuBisCo	47+	281.071	1,23151
987		58+	257,977	0,98981	431		60-	488.161	1,26570
902		377+	240,967	0,98981			36-	439.084	1,51763
567							414+	214.179	9,49256
944	Oxygen-evolving enhancer prot. 2						243+	525.165	1,26156
1268							210+	354.611	1,29922
72							218+	490.167	1,24343
1315	Allene oxide cyc						425+	281.071	1,16663
							436+	377.084	6,36698
							86+	355.112	6,36698

7.

CONCLUSIONES

7. Conclusiones

- Los análisis multivariantes PCA y PLS han demostrado su solvencia en los análisis ómicos, al hacer fácilmente interpretable la información masiva que genera un ensayo de proteómica o de metabolómica como el planteado en este trabajo.
- De acuerdo con los resultados, se puede asegurar que la planta de tomate responde a la infección de ambas cepas, compatible e incompatible, de *P. syringae*. Los cambios que se dan en las plantas tienen lugar tanto a nivel proteómico como metabolómico.
- A nivel de metabolómica y a pesar de que gran parte de los metabolitos no se han podido identificar, los que sí se han identificado, generalmente son compuestos fenólicos con un gran potencial antibiótico. Además, abundan los compuestos fenólicos y los derivados del ácido hidroxicinámico, además de algunos azúcares y fosfolípidos.
- En cuanto a proteómica, se han identificado proteínas que intervienen en la infección incompatible. Generalmente se trata de proteínas de síntesis de compuestos de respuesta a la infección (ácido jasmónico) o de respuesta a especies reactivas de oxígeno, muy importantes en la respuesta hipersensible.
- El PCA realizado en todas las fracciones de metabolómica polar y proteómica ha permitido destacar los metabolitos y proteínas propios de cada tipo de respuesta. Sin embargo, ha sido la integración de ambas ómicas con el análisis O2PLS la que ha sacado a la luz que los metabolitos apolares guardan también relación con la infección, y que la relación de dichos compuestos y las proteínas es muy importante en la enfermedad incompatible. De aquí la importancia de realizar los análisis integrados de diferentes tipos de ómicas, tal y como se ha puesto de manifiesto en el presente trabajo.

8.

REFERENCIAS

8. Referencias

- AZCÓN-BIETO J, TALÓN M. Fundamentos de fisiología vegetal. Madrid: McGraw-Hill; 2009.
- BINO R, HALL R, FIEHN O, KOPKA J, SAITO K, DRAPER J *et al.* Potential of metabolomics as a functional genomics tool. *Trends in Plant Science*. 2004;9(9):418-425.
- BOUHADDANI, E.S., HOUWING DUISTERMAAT, J., SALO, P., PEROLA, M., JONGBLOED, G., AND UH, H.-W. (2016). Evaluation of O2PLS in Omics data integration. *BMC Bioinformatics* 17(S2).
- CAO, F., YOSHIOKA, K., & DESVEAUX, D. (2011). The roles of ABA in plant–pathogen interactions. *Journal Of Plant Research*, 124(4), 489-499. doi: 10.1007/s10265-011-0409-y
- CAZZONELLI, C., & POGSON, B. (2010). Source to sink: regulation of carotenoid biosynthesis in plants. *Trends In Plant Science*, 15(5), 266-274. doi: 10.1016/j.tplants.2010.02.003
- CHRISTOPHE B.Y. CORDELLA. PCA: The Basic Building Block of Chemometrics, *Analytical Chemistry* (2012). Ira S. Krull, IntechOpen.
- COLLINGE J, WHITTINGTON M, SIDLE K, SMITH C, PALMER M, CLARKE A *et al.* Prion protein is necessary for normal synaptic function. *Nature*. 1994;370(6487):295-297.
- COOK C, BERGMAN M, FINN R, COCHRANE G, BIRNEY E, APWEILER R. The European Bioinformatics Institute in 2016: Data growth and integration. *Nucleic Acids Research*. 2015;44(D1):D20-D26.
- DANGL J, MCDOWELL J. Two modes of pathogen recognition by plants. *Proceedings of the National Academy of Sciences*. 2006;103(23):8575-8576.
- DIXON WL AND KEES T. Identification of the regulatory steps in glycolysis in potato tubers. *Phytochemistry*. 1980;19:1297-1301.
- EASTMENT, H. T. and KRZANOWSKI, W. J. (1982), "Crossvalidatory choice of the number of components from a principal component analysis," *Technometrics*, 24, 73-77.
- FLEISCHMANN R, ADAMS M, WHITE O, CLAYTON R, KIRKNESS E, KERLAVAGE A *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*. 1995;269(5223):496-512.
- GELADY, P. AND KOWALSKI, B. Partial least square regression: a tutorial. *Anal. Chem. Acta*. 1986;185,1-17.
- GLAZEBROOK, J. (2005). Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. *Annual Review Of Phytopathology*, 43(1), 205-227. doi: 10.1146/annurev.phyto.43.040204.135923
- GRÖBER, U., REICHRATH, J., HOLICK, M., & KISTERS, K. (2014). Vitamin K: an old vitamin in a new perspective. *Dermato-Endocrinology*, 6(1), e968490. doi: 10.4161/19381972.2014.968490
- HAMMOND-KOSACK K, JONES J. Plant disease resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology*. 1997;48(1):575-607.
- HAYAT, S., AHMAD, A., & ALYEMENI, M. (2013). *Salicylic acid* (pp. 3-5). Dordrecht: Springer.
- HE, L., XU, X., LI, Y., LI, C., ZHU, Y., YAN, H. *et al.* (2013). Transcriptome Analysis of Buds and Leaves Using 454 Pyrosequencing to Discover Genes Associated with the Biosynthesis of Active Ingredients in *Lonicera japonica* Thunb. *Plos ONE*, 8(4), e62924
- HEIL M. Induced Systemic Resistance (ISR) Against Pathogens in the Context of Induced Plant Defences. *Annals of Botany*. 2002;89(5):503-512.
- HIRASHIMA, M., TANAKA, R., & TANAKA, A. (2009). Light-Independent Cell Death Induced by Accumulation of Pheophorbide a in *Arabidopsis thaliana*. *Plant And Cell Physiology*, 50(4), 719-729. doi: 10.1093/pcp/pcp035
- JONES J, DANGL J. The plant immune system. *Nature*. 2006;444(7117):323-329.
- KATAJAMAA M, OREŠIČ M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics*. 2005;6(1):179.
- KEPPEL G, WICKENS T. Chapter 6, Simultaneous Comparisons and the Control of Type I Errors, en *Design and Analysis* (2004), pp. 5-6, Ed. Pearson.
- KOURTI, T. (2002), "Process analysis and abnormal situation detection: from theory to practice," *IEEE Control Systems Magazine*, 22, 10-25.

Referencias

- KOURTI, T. and MACGREGOR, J. F. (1996), "Multivariate SPC Methods for Process and Product Monitoring," *Journal of Quality Technology*, 28, 409-428.
- KOURTI, T. and MACGREGOR, J. F. (2006), "Process Analytical Technology Beyond Real-Time Analyzers: The Role of Multivariate Analysis," *Crit.Rev.Anal.Chem.*, 36, 257-278.
- KRZANOWSKI, W. J. (1987), "Cross-validation in Principal Component Analysis," *Biometrics*, 43, 575-584.
- LANFER-MARQUEZ, U., BARROS, R., & SINNECKER, P. (2005). Antioxidant activity of chlorophylls and their derivatives. *Food Research International*, 38(8-9), 885-891. doi: 10.1016/j.foodres.2005.02.012
- LAWTON K. Systemic Acquired Resistance in Arabidopsis Requires Salicylic Acid but Not Ethylene. *Molecular Plant-Microbe Interactions*. 1995;8(6):863.
- LEVINE A, TENHAKEN R, DIXON R, LAMB C. H₂O₂ from the oxidative burst orchestrates the plant hypersensitive disease resistance response. *Cell*. 1994;79(4):583-593.
- LÓPEZ-GRESA M, LISÓN P, CAMPOS L, RODRIGO I, RAMBLA JL, *et al*. A Non-targeted Metabolomics Approach Unravels the VOCs Associated with the Tomato Immune Response against *Pseudomonas syringae*. *Frontiers in Plant Science*. 2017;8:1188
- LÓPEZ-GRESA, M., TORRES, C., CAMPOS, L., LISÓN, P., RODRIGO, I., BELLÉS, J., & CONEJERO, V. (2011). Identification of defence metabolites in tomato plants infected by the bacterial pathogen *Pseudomonas syringae*. *Environmental And Experimental Botany*, 74, 216-228. doi: 10.1016/j.envexpbot.2011.06.003
- LOU, Z., WANG, H., ZHU, S., MA, C., & WANG, Z. (2011). Antibacterial Activity and Mechanism of Action of Chlorogenic Acid. *Journal Of Food Science*, 76(6), M398-M403. doi: 10.1111/j.1750-3841.2011.02213.x
- MCELROY, J., & KOPSELL, D. (2009). Physiological role of carotenoids and other antioxidants in plants and application to turfgrass stress management. *New Zealand Journal Of Crop And Horticultural Science*, 37(4), 327-333. doi: 10.1080/01140671.2009.9687587
- MELOTTO M, UNDERWOOD W, HE S. Role of Stomata in Plant Innate Immunity and Foliar Bacterial Diseases. *Annual Review of Phytopathology*. 2008;46(1):101-122.
- MIKKELSEN S, CORTÓN E, *Bioanalytical Chemistry*. John Wiley & Sons, Inc. p. 224 (2004).
- MOURA J, BONINE C, DE OLIVEIRA FERNANDES VIANA J, DORNELAS M, MAZZAFERA P. Abiotic and Biotic Stresses and Changes in the Lignin Content and Composition in Plants. *Journal of Integrative Plant Biology*. 2010;52(4):360-376.
- NURNBERGER T, BRUNNER F, KEMMERLING B, PIATER L. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunological Reviews*. 2004;198(1):249-266.
- OJALA, T., REMES, S., HAANSUU, P., VUORELA, H., HILTUNEN, R., HAAHTELA, K., & VUORELA, P. (2000). Antimicrobial activity of some coumarin containing herbal plants growing in Finland. *Journal Of Ethnopharmacology*, 73(1-2), 299-305. doi: 10.1016/s0378-8741(00)00279-8
- OLIVER S. Systematic functional analysis of the yeast genome. *Trends in Biotechnology*. 1998;16(9):373-378.
- PARKER J, COLEMAN M. Molecular intimacy between proteins specifying plant-pathogen recognition. *Trends in Biochemical Sciences*. 1997;22(8):291-296.
- RIEMANN, M., HAGA, K., SHIMIZU, T., OKADA, K., ANDO, S., & MOCHIZUKI, S. *et al*. (2013). Identification of rice Allene Oxide Cyclase mutants and the function of jasmonate for defence against *Magnaporthe oryzae*. *The Plant Journal*, 74(2), 226-238. doi: 10.1111/tpj.12115
- ROMERO R, ZÚNICA L. Capítulo 9: Introducción al análisis de la varianza, en *Métodos estadísticos para ingenieros* (2013), pp. 157-172, Ed. Universitat Politècnica de València
- SHANNON P, MARKIEL A, OZIER O, BALIGA NS, WANG JT, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003; 13(11):2498-504.
- SIEBERS, M., BRANDS, M., WEWER, V., DUAN, Y., HÖLZL, G., & DÖRMANN, P. (2016). Lipids in plant-microbe interactions. *Biochimica Et Biophysica Acta (BBA) - Molecular And Cell Biology Of Lipids*, 1861(9), 1379-1395. doi: 10.1016/j.bbalip.2016.02.021
- SMITH C, WANT E, O'MAILLE G, ABAGYAN R, SIUZDAK G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry*. 2006;78(3):779-787.
- SPARKMAN DAVID O, *Mass spectrometry desk reference*. Pittsburgh: Global View Pub. 2000.

Referencias

- SPOEL S, DONG X. How do plants achieve immunity? Defence without specialized immune cells. *Nature Reviews Immunology*. 2012;12(2):89-100.
- SUMNER L, MENDES P, DIXON R. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*. 2003;62(6):817-836.
- TAUZIN AS, GIARDINA T. Sucrose and invertases, a part of the plant defense response to the biotic stresses. 2014;5:293.
- TRYGG, J., (2002), O2-PLS for Qualitative and Quantitative Analysis in Multivariate Calibration, *Journal of Chemometrics*, 16, 283-293.
- VINAYAVEKHIN N, SAGHATELIAN A. Untargeted Metabolomics. *Current Protocols in Molecular Biology*. 2010;Chapter 30(Unit 30.1):1-24.
- WASINGER VC, CORDWELL SJ, CERPA-POLJAK A, YAN JX, GOOLEY AA, WILKINS MR, DUNCAN MW, HARRIS R, WILLIAMS KL, HUMPHERY-SMITH I, *Electrophoresis* 16 (1995) 1090-1094.
- WEINBERGER F. Pathogen-Induced Defense and Innate Immunity in Macroalgae. *The Biological Bulletin*. 2007;213(3):290-302.
- WILLIS D, KINSCHERF T. Population Dynamics of *Pseudomonas syringae* pv. tomato Strains on Tomato Cultivars Rio Grande and Rio Grande-Ptounder Field Conditions. *Journal of Phytopathology*. 2009;157(4):219-227.
- WOLD, H. (1973), "Non-linear iterative partial least squares (NIPALS) Modelling. Some current developments," in *Multivariate Analysis*, P. R. Krishnaiah (ed.), New York: Academic Press.
- WOLD, S. (1976), "Pattern Recognition by Means of Disjoint Principal Component Models," *Pattern Recognition*, 8, 127-139.
- WOLD, S., ANTTI, H., LINDGREN F. & ÖHMAN J. Orthogonal signal correction of near-infrared spectra. 1998, 44(2), 175-185.
- WOLD, S., SJÖSTRÖM, M., & ERIKSSON, L. PLS-regression: a basic tool of chemometrics. *Chemometrics And Intelligent Laboratory Systems*. 2009, 58(2), 109-130. doi: 10.1016/s0169-7439(01)00155-1
- WOLD, S., WOLD, H., DUNN, W. J., and RUHE, A. (1984), "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses," *SIAM Journal on Scientific and Statistical Computing*, 5, 735-743.
- WU KJ., SHALER TA, BECKER CH, *Analytical Chemistry* 66 (1994) 1637-1645.
- XIN X, HE S. *Pseudomonas syringae* pv. tomato DC3000: A Model Pathogen for Probing Disease Susceptibility and Hormone Signaling in Plants. *Annual Review of Phytopathology*. 2013;51(1):473-498.
- YANG, E., OH, Y., LEE, E., PARK, A., CHO, S., YOO, Y., & PARK, O. (2003). Oxygen-evolving enhancer protein 2 is phosphorylated by glycine-rich protein 3/wall-associated kinase 1 in *Arabidopsis*. *Biochemical And Biophysical Research Communications*, 305(4), 862-868. doi: 10.1016/s0006-291x(03)00851-9
- YOGENDRA, K., KUMAR, A., SARKAR, K., LI, Y., PUSHPA, D., & MOSA, K. *et al.* (2015). Transcription factor StWRKY1 regulates phenylpropanoid metabolites conferring late blight resistance in potato. *Journal Of Experimental Botany*, 66(22), 7377-7389. doi: 10.1093/jxb/erv434
- ZACARÉS, L., LÓPEZ-GRESA, M., FAYOS, J., PRIMO, J., BELLÉS, J., & CONEJERO, V. (2007). Induction of p-Coumaroyldopamine and Feruloyldopamine, Two Novel Metabolites, in Tomato by the Bacterial Pathogen *Pseudomonas syringae*. *Molecular Plant-Microbe Interactions*, 20(11), 1439-1448. doi: 10.1094/mpmi-20-11-14