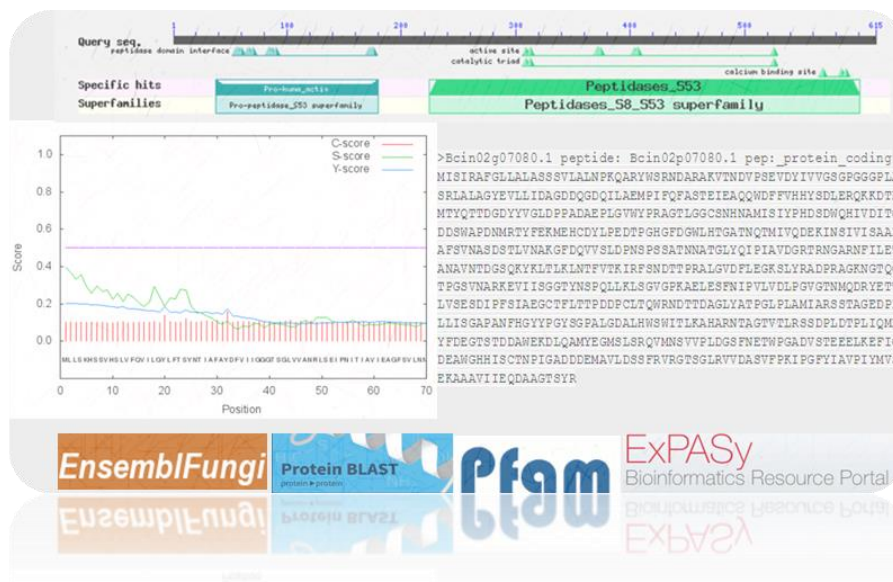


Estudio bioinformático de los secretomas de *Botrytis cinerea*

Bioinformatic analysis of the secretomes from *Botrytis cinerea*



Trabajo de Fin de Grado

FELIPE RUYMÁN ALONSO RAVELO

Tutorizado por Nélida Emilia Brito Alayón y Mario González Carracedo

Grado en Biología. Julio 2017

**Estudio bioinformático de los secretomas de
*Botrytis cinerea***

**Bioinformatic analysis of the secretomes
from *Botrytis cinerea***

Trabajo de Fin de Grado

FELIPE RUYMÁN ALONSO RAVELO

Tutorizado por Nélida Emilia Brito Alayón y Mario González Carracedo

Grado en Biología. Julio 2017

Resumen	1
Abstract	1
Introducción	2
Filogenia	2
Interacción hongo-hospedador.....	2
Secretoma y glicosecretoma	3
Bioinformática	4
Objetivos	5
Material y Métodos	6
Ensemble Fungi	6
Blast (NCBI).....	6
Google Scholar.....	6
PubMed (NCBI).....	6
Urgi.....	7
UniProt	7
CAZy.....	8
Clustal Omega y EMBOSS Needle.....	8
Pfam	8
ProtParam	8
SignalP 4.1	9
NetNGlyc1.0 y NetOGlyc 4.0	9
Resultados y discusión	10
Actualización de las secuencias de las proteínas estudiadas	10
Identificación de las proteínas y análisis de su posible redundancia	11
GMC oxidorreductasa: Bcin02g07080	11
No identificada: Bcin12g03520 y proteínas redundantes.....	12
Lacasas: Bcin14g02510 y proteínas redundantes	13
Serina proteasas: Bcin15g04670 y proteínas redundantes	14
Proteína Ecm33 con anclaje GPI a la pared: Bcin04g01310 y proteínas redundantes	15
No identificada: Bcin12g04640 y proteínas redundantes.....	16
Glucosa oxidasa: Bcin01g11310 y proteínas redundantes	16
Beta-glucosidasas: Bcin10g05590 y proteínas redundantes	17
Xiloglucanasas: Bcin03g03630 y proteínas redundantes.....	18
Metalopeptidasas ácidas: Bcin16g02770 y proteínas redundantes	19
Alfa-1,2-manosidasas: Bcin01g05680 y proteínas redundantes.....	20
Exopoligalacturonasas: Bcin05g02520 y proteínas redundantes	20
Acetilesterasas: Bcin01g08910 y proteínas redundantes.....	21
No identificada: Bcin06g06670 y proteínas redundantes.....	22
Alfa-fucosidasa: Bcin08g05870 y proteínas redundantes.....	22
Conclusiones	23
Bibliografía	24
Anexo	26
Tabla S1. Caracterización <i>in silico</i> de las lacasas codificadas por el genoma de <i>B. cinerea</i>	26
Tabla S2. Caracterización <i>in silico</i> de las posibles serina proteasas codificadas por el genoma de <i>B. cinerea</i>	27
Tabla S3. Caracterización <i>in silico</i> de las beta-glucosidasas codificadas por el genoma de <i>B. cinerea</i>	28
Tabla S4. Caracterización <i>in silico</i> de las alfa-1,2-manosidasas codificadas por el genoma de <i>B. cinerea</i>	29

Tabla S5. Caracterización <i>in silico</i> de las exopoligalacturonas codificadas por el genoma de <i>B. cinerea</i>	30
Tabla S6. Caracterización <i>in silico</i> de las acetilesterasas codificadas por el genoma de <i>B. cinerea</i> .	30

Resumen

Botrytis cinerea es un ascomiceto capaz de infectar más de 1400 especies vegetales, muchas de interés agroeconómico. Durante la interacción con el hospedador, secreta una gran cantidad de proteínas diferentes, dependiendo de las condiciones de crecimiento y del hospedador, que en general, reciben el nombre de secretoma. En este trabajo se ha realizado la caracterización *in silico* de proteínas halladas experimentalmente en dos secretomas. Por un lado, en el “secretoma temprano”, que incluye proteínas secretadas durante las primeras horas tras la germinación de las conidias, y por otro lado, en el “glicosecretoma tardío” o glicoproteínas secretadas por el micelio después de 4 días de crecimiento en medio rico. Las secuencias de las 15 proteínas seleccionadas fueron analizadas utilizando diferentes herramientas bioinformáticas, que permitieran, en primer lugar, su identificación, es decir, su clasificación en una familia proteica concreta, y, en segundo lugar, conocer algunos aspectos que ayudasen a completar la información sobre cada una de ellas. Además, se analizó el grado de redundancia de estas proteínas en *Botrytis*. Se consiguió reconocer una proteína no redundante caracterizada como proteína Ecm33 con anclaje GPI a la pared, que no existe en plantas, que podría servir como diana para el diseño de fármacos destinados al control del hongo.

Abstract

The Ascomycete *Botrytis cinerea* is able to infect more than 1400 plant species, many of agroeconomic interest. During its interaction with the host, it secretes a large amount of proteins, different depending on the growth conditions and the host, which in general, it is named as secretome. In this work, the proteins found experimentally in two specific secretomes were characterized *in silico*. On the one hand, the "early secretome", which includes secreted proteins during the first hours after conidial germination. On the other hand, the "late glycoscretome" or glycoproteins secreted by the mycelium after 4 days of growth in rich medium. The sequences of the 15 selected proteins were analyzed using different bioinformatics tools, which allowed, first, their identification, that is, their classification into a specific protein family and, second, to know features that help to complete their knowledge. In addition, the degree of redundancy of these proteins in *Botrytis* was analyzed. It was able to identify a non-redundant protein as a GPI-anchored Ecm33 protein, which does not exist in plants, and could be used as a target for the design of new drugs to control the fungus.

INTRODUCCIÓN

EnsemblFungi

Protein BLAST
protein • protein

Pfam

ExpASY
Bioinformatics Resource Portal

Introducción

Filogenia

Incluido en la clase Ascomycetes, *Botrytis* es uno de los géneros que conforman la familia Sclerotiniaceae, en la que se incluyen mayormente patógenos vegetales necrotróficos, y en el que destaca la especie *Botrytis cinerea* (1). Este hongo ha sido considerado el segundo hongo patógeno más importante, en función de criterios económicos y científicos (2). Cuenta con un vasto rango de hospedadores en el que se incluyen más de 1400 especies vegetales, muchas de ellas de gran interés agroeconómico, por lo que tiene un profundo impacto económico (3). La reciente secuenciación del genoma de las cepas B05.10 (4) y T4 (<https://urgi.versailles.inra.fr/>), ha arrojado valiosa información acerca de la biología del hongo.

Interacción hongo-hospedador

Dos son las fases que se pueden distinguir, con carácter general, en el ciclo infeccioso del hongo: una primera caracterizada por la germinación de las conidias y penetración de los tejidos vegetales para producir una necrosis a escala local, y una fase final en la que la lesión se expande en las zonas vecinas provocando la maceración del tejido (3).

A lo largo de su ciclo de vida, *Botrytis* secreta al medio gran diversidad de proteínas (5). La familia proteica más abundante es la de las proteasas, responsables de la degradación de las proteínas de la pared celular vegetal y de la hidrólisis de diversas proteínas con el fin de obtener nutrientes (6). Además, el hongo secreta distintas glicosil hidrolasas destinadas a la degradación del resto de los componentes de la pared (celulosas, hemicelulosas y pectina, entre otros) y de otras envueltas celulares (6). De entre ellas, se ha comprobado que la endoxilanasasa Xyn11A contribuye a la virulencia del hongo (7), así como las endopoligalacturonasas BcPG1 y BcPG2 (8). También algunas oxidoreductasas, entre ellas la Cu-Zn-superóxido dismutasa BcSOD1, han sido caracterizadas como factores de virulencia, interviniendo en la respuesta frente a la respuesta oxidativa que se produce durante la interacción planta-patógeno (9). Por otro lado, también el hongo es capaz de secretar proteínas con actividad necrotizante, como las proteínas tipo Nep-1 (10). Sin embargo, hasta el momento, además de las ya mencionadas, pocas son las proteínas secretadas por el hongo que hayan sido analizadas en detalle, siendo muy pocas para las que se ha puesto de manifiesto su papel en la virulencia del hongo. Además, la redundancia de proteínas descrita en *Botrytis* (8,10,11,12), dificulta aún más su estudio. Por todo ello, es necesario profundizar en el

conocimiento del papel de las proteínas secretadas por el hongo que permitan identificar factores de patogenicidad y, con ello, el diseño de nuevos métodos de control de *B. cinerea*.

Secretoma y glicosecretoma

En este trabajo se pretende llevar a cabo la caracterización *in silico* de las proteínas halladas en el secretoma temprano (13) y en el glicosecretoma de *B. cinerea* (14). Las proteínas correspondientes al secretoma temprano fueron obtenidas haciendo germinar al hongo en condiciones que simularan la interacción hongo-hospedador *in planta* (13). Para ello, se crecieron conidias de la cepa silvestre B05.10 en un medio sintético en el que se sumergió una bolsa de diálisis que contenía diversos extractos vegetales (kiwi, fresa o tomate), y al cabo de 16 horas se recogieron las proteínas presentes en el medio de cultivo, se concentraron y fueron analizadas por electroforesis bidimensional y por espectrometría de masas en tándem (13). A partir de este análisis se identificaron un total de 105 proteínas que inicialmente fueron clasificadas en familias atendiendo a sus hipotéticas funciones (13).

El glicosecretoma de *B. cinerea* se estudió a partir de cultivos del hongo crecido en medio rico durante 4 días (14). Los micelios fueron retirados por filtración y las glicoproteínas secretadas fueron purificadas por cromatografía de afinidad utilizando concanavalina A como matriz a partir de los medios de cultivo (14). Las glicoproteínas purificadas fueron analizadas por electroforesis bidimensional y espectrometría de masas en tándem (14). En este caso, se identificaron 158 proteínas clasificadas inicialmente en una gran variedad de familias proteicas, todas ellas con un alto contenido en residuos de serina y treonina, que probablemente sufran un proceso de *O*-glicosilación (15).

La anotación de la posible identidad y función de las proteínas halladas en estas dos aproximaciones experimentales se realizó, en los dos casos, a partir de la identificación en distintas bases de datos, de proteínas con un alto grado de similitud con las de *Botrytis* pero de función conocida, utilizando el software Blast (16). Sin embargo, gran número de proteínas de *Botrytis* no pudieron llegar a ser incluidas en ninguna familia proteica, por no hallar ninguna proteína homóloga o dominio que pudiera caracterizarla, o bien, la identificación de muchas de ellas se basaba en ser homólogas a otras proteínas con función hipotética. Por todo ello, se hace necesario un estudio *in silico* detallado de todas estas proteínas con el fin de llegar a su posible identificación.

Bioinformática

La bioinformática surge de la conjugación entre numerosas disciplinas, tales como la biología, las matemáticas, la informática y la estadística, utilizando la tecnología de la información para el estudio de los sistemas biológicos (17, 18).

La propia naturaleza del análisis que se desee realizar determina qué herramienta bioinformática en particular ha de usarse. Por ejemplo, el sistema Entrez desarrollado por el National Center for Biotechnology Information o NCBI permite buscar y obtener información de un amplio abanico de dominios al acceder a varias bases de datos (19). Los servidores Blast (16) o Clustal Omega (20) establecen alineamientos entre secuencias génicas o proteicas identificando zonas de similitud útiles para estudiar su origen e historia evolutiva. Todas estas herramientas se valen de modelos matemáticos e inferencias estadísticas tales como la programación dinámica, análisis de regresión, análisis cluster, redes neuronales artificiales, entre otras (18), que finalmente, facilitan la caracterización *in silico* de proteínas con un alto grado de fiabilidad. Esta información permitirá planificar y desarrollar la metodología adecuada para su caracterización experimental (18).

OBJETIVOS

EnsemblFungi

Protein BLAST
protein • protein

Pfam

ExPASy
Bioinformatics Resource Portal

Objetivos

Por todo lo expuesto anteriormente, se ha propuesto como objetivo general del presente trabajo:

Identificar y caracterizar *in silico* proteínas del secretoma temprano y del glicosecretoma de *Botrytis cinerea*.

Para ello, se abordarán los siguientes objetivos específicos:

1. Identificación de un grupo de proteínas secretadas por *Botrytis cinerea* en distintas condiciones de crecimiento mediante análisis Blast y reconocimiento en sus secuencias de los correspondientes dominios conservados de cada familia proteica utilizando el servidor Pfam.
2. Análisis del número de genes en el genoma del hongo que codifiquen para proteínas similares a las detectadas en cada secretoma.
3. Caracterización *in silico* de las proteínas objeto de estudio y sus redundantes

MATERIAL Y MÉTODOS

EnsemblFungi

Protein BLAST
protein → protein

Pfam

ExpASY
Bioinformatics Resource Portal

Material y Métodos

Ensemble Fungi: <http://fungi.ensembl.org/index.html>

Esta herramienta bioinformática contiene todos aquellos genomas fúngicos cuyas secuencias y anotaciones han sido estudiadas y enviadas a los distintos servidores o bases de datos que conforman la ‘Colaboración Internacional de Bases de Datos de Secuencias Nucleotídicas’ (servidores tales como GenBank o ENA, entre otros) (21). En esta base de datos se incluye la última secuenciación y anotación del genoma de *Botrytis cinerea* (4).

Blast (NCBI): <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Conocido como BLAST por sus siglas en inglés (Basic Local Alignment Search Tool), esta herramienta bioinformática permite identificar una proteína al comparar su secuencia (*query sequence* o secuencia consulta) con todas las secuencias de proteínas anotadas en distintas bases de datos y encontrar regiones con alto grado de similitud (16). El servidor devuelve una lista de proteínas ordenadas en función de varios parámetros, entre los que destacan el porcentaje de identidad y el denominado “E-value”. El primero da idea del número de coincidencias entre aminoácidos de las secuencias de la proteína problema y la identificada por el servidor (mismos aminoácidos en las mismas posiciones relativas) y el segundo da idea de la significación estadística de la comparación, ya que expresa la probabilidad de que las coincidencias entre las dos secuencias se haya producido al azar (de forma que cuanto más pequeño es el E-value, más fiable es la coincidencia encontrada por el servidor).

Google Scholar: <https://scholar.google.es/>

Propiedad de Google, Google Scholar o Google Académico en español, es un motor de búsqueda lanzado en noviembre de 2004 que permite el acceso a una gran variedad de literatura escolar, tal como resúmenes de conferencias, tesis, informes técnicos, revistas y libros académicos, etcétera. La lista de los artículos o demás información de interés buscada en este servidor se ordena según un ranking que tiene en cuenta diversos factores, con especial peso en el número de citas, de tal manera que los primeros resultados suelen ser artículos que cuentan con gran número de referencias.

PubMed (NCBI): <https://www.ncbi.nlm.nih.gov/pubmed/>

Desarrollado por la Biblioteca Nacional de Medicina de los Estados Unidos, se hizo accesible al público de manera gratuita en junio de 1997 como parte del sistema Entrez. Se

trata de una enorme base de datos que contiene más de 27 millones de citas o referencias de literatura biomédica, revistas de campos de las ciencias de la salud, y libros virtuales, con posibilidad acceder al texto completo a través de la propia central PubMed o de otros sitios web.

Urgi: <https://urgi.versailles.inra.fr/>

La Unité de Recherche Génomique Info desarrollada por el Instituto Nacional de Investigación Agronómica (INRA) está dedicada a la búsqueda e integración de datos relativos a la estructura y dinámica genómica de plantas y patógenos de cultivos. Cuenta con un servidor informático enmarcado dentro del nódulo francés de la red Europea de plataformas bioinformáticas conocida como Elixir. Su sistema de información centralizado almacena gran variedad de datos genómicos que incluyen anotaciones del genoma, marcadores genéticos, polimorfismos, fenotipos, etcétera, existiendo asimismo una gran atención en los transposones, importantes en la función y evolución de los genomas, y que juegan un papel central en la estructura y tamaño de estos.

UniProt: <http://www.uniprot.org/>

Este centro de información universal de proteínas (“Universal Protein Resource” en inglés, de ahí sus siglas) surge de la colaboración entre el Instituto Europeo de Bioinformática, el Instituto Suizo de Bioinformática SIB y el Centro de Información de Proteínas, y consta de tres bases de datos: UniProtKB (UniProt Knowledgebase), UniParc (UniProt Archive) y UniRef (UniProt Reference Clusters).

UniProtKB (<http://www.uniprot.org/uniprot/>) es una base de datos de secuencias proteicas compuesta de dos secciones: “UniProtKB/Swiss-Prot”, una base de datos de secuencias no redundantes anotadas manualmente, y “UniProtKB/TrEMBL”, que contiene secuencias proteicas con anotaciones generadas automáticamente.

UniParc (<http://www.uniprot.org/uniparc/>) almacena cada secuencia proteica y le asigna un identificador único y estable, posibilitando así la identificación de la misma proteína en distintas bases de datos.

UniRef (<http://www.uniprot.org/uniref/>), por su parte, contiene tres bases de datos que surgen de la agrupación de secuencias proteicas de las dos bases anteriores en clusters, teniendo las secuencias de cada cluster entre sí, como mínimo, entre un 50 y un 90% de identidad. Lo que se consigue con esta agrupación de secuencias es reducir el tamaño de la base de datos y, por ende, reducir el tiempo de búsqueda.

CAZy: <http://www.cazy.org/>

Esta base de datos, disponible en la red desde 1998, se dedica al estudio bioquímico y estructural de enzimas con actividad sobre carbohidratos (22). Nuevas secuencias son añadidas regularmente a la base de datos, poco después de que hayan sido recogidas en las publicaciones de GenBank. La lista de familias es constantemente actualizada y solo se añaden nuevas familias a la lista tras haber comprobado que existen evidencias recogidas en la literatura de la actividad de al menos un miembro de la misma. Numerosas bases de datos “accesorias” han surgido a partir de la original, y en la actualidad hay incluso una enciclopedia virtual (CAZypedia) elaborada por miembros de la comunidad científica centrada en el estudio de este tipo de enzimas.

Clustal Omega y EMBOSS Needle: <http://www.ebi.ac.uk/Tools/msa/clustalo/> y http://www.ebi.ac.uk/Tools/psa/emboss_needle/

Bajo el término “Clustal” se agrupa a toda una serie de programas o herramientas de bioinformática destinadas al alineamiento múltiple de secuencias. Clustal Omega permite alinear tres o más secuencias entre sí y “EMBOSS Needle” es uno de los más utilizados para alinear dos secuencias, ambos originados en el Instituto Europeo de Bioinformática (EMBL-EBI). Se identifican, de esta manera, regiones similares que pueden ser indicativas de relaciones funcionales, estructurales y/o evolutivas entre las secuencias de proteínas y ácidos nucleicos. El alineamiento múltiple (Clustal Omega) proporciona una matriz de identidad entre las secuencias analizadas e incluso un árbol filogenético basado en las distancias evolutivas estimadas entre ellas.

Pfam: <http://pfam.xfam.org/>

Esta base de datos de familias proteicas incluye las anotaciones de las mismas y alineamientos de múltiples secuencias generadas usando diversos algoritmos. La última versión, Pfam 31.0, está disponible desde marzo del 2017, y contiene un total de 16712 familias y 604 clanes proteicos. El servidor identifica en la secuencia dominios conservados que permiten enmarcar la proteína en una familia proteica determinada.

ProtParam: <http://web.expasy.org/protparam/>

Esta herramienta permite estimar toda una serie de parámetros físico-químicos de una determinada proteína cuya secuencia sea suministrada por el usuario o bien usando el número de acceso o ID de una proteína que se encuentre recogida en la base de datos Swiss-Prot o TrEMBL.

La información que devuelve el servidor recoge valores de múltiples parámetros proteicos tales como el peso molecular (calculado a partir de la suma de las masas isotópicas medias de los aminoácidos que constituyen la proteína), el punto isoelectrico (utilizando los valores de pK de los aminoácidos que definen la secuencia primaria de la proteína), composición porcentual de cada aminoácido, cantidad de residuos con carga negativa/positiva, composición atómica, su índice alifático (definido como el volumen relativo ocupado por cadenas laterales alifáticas de los aminoácidos que componen la proteína y que da idea de su termoestabilidad) y el valor GRAVY (Grand Average of Hydrophaty, promedio de hidrofobicidad) que determina la solubilidad de la proteína si el valor es negativo, pero que indica cierto carácter hidrófobo si su valor es positivo.

SignalP 4.1: <http://www.cbs.dtu.dk/services/SignalP/>

Los péptidos señal son regiones cortas en el extremo amino de la mayor parte de las proteínas que están destinadas a entrar en la vía de secreción y que serán escindidos durante el proceso de maduración de la proteína. SignalP 4.1 Server forma parte de los denominados ‘CBS Prediction Servers’ y permite analizar la posible presencia de una secuencia señal en la secuencia problema, indicando además la posición probable en la que tiene lugar el corte de este péptido señal, así como los aminoácidos que ocupan dichas posiciones.

NetNGlyc1.0 y NetOGlyc 4.0: <http://www.cbs.dtu.dk/services/NetNGlyc> y <http://www.cbs.dtu.dk/services/NetOGlyc/>

Estos dos servidores, al igual que el recién citado *SignalP 4.1 Server*, forman parte de los denominados ‘CBS Prediction Servers’. En concreto, *NetNGlyc 1.0* y *NetOGlyc4.0* predicen sitios potenciales de *N* y *O*-glicosilación. La identificación de las dianas de *N*-glicosilación se basa exclusivamente, por defecto, en secuencias consenso Asn-Xaa-Ser/Thr, aunque existe la posibilidad de basar la predicción en todos los residuos de asparagina. Las predicciones del servidor *NetOGlyc*, por su parte, se basan en la identificación de residuos de serina y treonina en glicoproteínas tipo mucina (23).

RESULTADOS Y DISCUSIÓN

EnsemblFungi

Protein BLAST
protein • protein

Pfam

ExPASy
Bioinformatics Resource Portal

Resultados y discusión

1. Actualización de las secuencias de las proteínas estudiadas

Con posterioridad a la publicación de los dos trabajos en los que experimentalmente fueron halladas las proteínas objeto de estudio (13, 14), se ha completado la secuenciación del genoma de *Botrytis cinerea* (4; <https://urgi.versailles.inra.fr/>). A fin de determinar si como consecuencia de la nueva anotación del genoma las secuencias de las proteínas se han visto modificadas, se procedió a realizar un Blast local usando como secuencia problema o “query” cada una de ellas en el servidor EnsembleFungi (que recoge la última secuenciación del genoma). Las nuevas secuencias se alinearon con las iniciales utilizando la herramienta bioinformática EMBOSS Stretcher, Pairwise Sequence Alignment para identificar estos posibles cambios. Únicamente en la secuencia de la proteína anotada como Bcin12g04640 se detectaron dos sustituciones puntuales de dos aminoácidos: Y287D y T289P.

Las nuevas anotaciones de las proteínas identificadas en el glicosecretoma y en el secretoma temprano se recogen en la Tabla 1.

Tabla 1 Anotación de las proteínas del glicosecretoma (GLY) y del secretoma temprano (TEM) de *B. cinerea* según la nueva anotación del genoma del hongo.

Anotación ^a (glicosecretoma)	Otras anotaciones	secretoma
Bcin02g07080	B0510_10243 ^d	GLY
Bcin12g03520	B0510_2873 ^d	GLY
Bcin14g02510	B0510_5238 ^d	GLY
Bcin15g04670	B0510_2759 ^d	GLY
Bcin04g01310	BC1G_10630.1 ^b	GLY
Bcin12g04640	BC1G_07482.1 ^b	GLY
Bcin01g11310	B0510_10137 ^d /B0510_10138 ^d	GLY
Bcin06g06670	B0510_8903 ^d	GLY
Bcin08g05870	B0510_8264 ^d	GLY
Bcin10g05590	BC1G_10221.1 ^b ; BofuT4_P086700.1 ^c	GLY
Bcin03g03630	BC1G_00594.1 ^b ; BofuT4_P021270.1 ^c	TEM
Bcin16g02770	BC1G_09180.1 ^b ; BofuT4_P112580.1 ^c	TEM
Bcin01g05680	BC1G_00455.1 ^b ; BofuT4_P045420.1 ^c	TEM
Bcin05g02520	BC1G_01617.1 ^b ; BofuT4_P043000.1 ^c	TEM
Bcin01g08910	BC1G_00003.1 ^b	TEM

^a Anotación utilizada en la secuenciación más reciente del genoma de la cepa *B. cinerea* B05.10 (http://fungi.ensembl.org/Botrytis_cinerea/).

^b Anotación utilizada en la primera secuenciación del genoma de *B. cinerea* B05.10 (<http://www.broad.mit.edu>; 2011)

^c Anotación utilizada en la única secuenciación de *B. cinerea* T4 (urgi.versailles.inra.fr; 2012)

^d Anotación utilizada en la segunda secuenciación de *B. cinerea* B05.10 (<http://www.broad.mit.edu>; 2012)

2. Identificación de las proteínas y análisis de su posible redundancia

Con el fin de identificar las proteínas objeto de estudio, se realizó un análisis Blast (NCBI) frente a la base de datos que engloba todas las proteínas identificadas en el taxón Fungi (hongos). Los criterios que se siguieron para analizar el listado de proteínas proporcionado por el servidor fueron que el E-value fuese muy bajo ($<10^{-10}$), el porcentaje de identidad fuera alto y que la identidad de la proteína encontrada por el Blast hubiese sido contrastada experimentalmente, descartando aquellas anotadas como “posibles o hipotéticas” por haber sido inferidas a partir de la homología de su secuencia con otras.

En *B. cinerea* ha sido descrito un alto grado de redundancia para numerosas enzimas extracelulares como aspártico proteasas (11), poligalacturonasas (8), pectín metil esterasas (12), etc. Por ello, en el presente trabajo, se decidió analizar el grado de redundancia de las proteínas identificadas por el análisis Blast en los secretomas de *Botrytis* (13, 14). Para ello, se procedió a realizar un Blast local en la base de datos que recoge la última anotación del genoma de *Botrytis* (<http://fungi.ensembl.org/index.html>), utilizando como secuencia “query” la secuencia de la proteína elegida a partir del análisis Blast detallado en el apartado anterior, caracterizada experimentalmente y con mayor porcentaje de identidad y menor E-value con la secuencia de la proteína de *Botrytis* a identificar inicialmente (Tabla 1). En todos los casos, las nuevas proteínas identificadas en este análisis Blast local fueron confirmadas por medio de un tercer análisis Blast general (NCBI) frente a una base de datos de proteínas no redundantes. También, se estudió el grado de homología entre cada clase de proteínas redundantes realizando un alineamiento mediante la herramienta Clustal Omega.

Por último, una vez elaborada la lista de proteínas de cada tipo codificadas por el genoma de *Botrytis*, se procedió a analizar *in silico* las secuencias de las mismas haciendo uso para ello de los servidores bioinformáticos que se han detallado en la sección de Material y Métodos.

Los resultados obtenidos se detallan a continuación.

GMC oxidorreductasa: *Bcin02g07080*

A pesar de que ninguna de las proteínas homólogas encontradas en otros hongos cuenta con estudios experimentales que corroboren la función predicha para las mismas, todas ellas aparecen anotadas como posibles Glucosa-Metanol-Colina oxidorreductasas. Con la que más homología mostró fue con la colina deshidrogenasa hipotética del Ascomycete *Emmonsia crescens* (KKZ65042.1), aunque esta proteína ha sido anotada también por

homología de secuencia con otras proteínas y no ha sido contrastada experimentalmente a nivel proteico. Al analizar la secuencia de Bcin02g07080 haciendo uso del servidor Pfam, se confirma que contiene un dominio GMC oxidorreductasa N-terminal que interacciona con FAD (PF00732) y otro C-terminal (PF05199), pero no se pudo especificar de qué tipo de oxidorreductasa se trata. Por esta razón, no se analizó la posible redundancia de esta proteína.

Se trata de una proteína de 678 aminoácidos (de los cuales 5 son residuos de cisteína) que contiene una secuencial señal de 19 aminoácidos de extensión, un posible lugar de *N*-glicosilación (posición 221) y otros siete residuos posiblemente *O*-glicosilados. El punto isoeléctrico de 4,54, el índice alifático de 77,88, junto con el valor “GRAVY” (-0,301), clasifican a la proteína como ácida, estable y soluble.

Esta proteína fue identificada experimentalmente en el glicosecretoma de *Botrytis* (14), en la posición vigesimocuarta entre las proteínas más abundantes, por lo que se presupone que debe desempeñar algún papel relevante en la biología del hongo. El análisis *in silico* no ha podido especificar su actividad, por lo que se requiere de un estudio más detallado para lograr su identificación.

No identificada: Bcin12g03520 y proteínas redundantes

Todas las proteínas homólogas a Bcin12g03520 fueron halladas en hongos de la división Ascomycota. Ninguna de ellas cuenta con evidencias experimentales de su actividad recogidas en la literatura. Es más, ninguna de las proteínas con secuencias homólogas encontradas tiene una función caracterizada. Tampoco el análisis de la secuencia de Bcin12g03520 con el servidor *Pfam* identificó ningún dominio conocido que la pudiera clasificar en alguna familia proteica.

Bcin12g03520 es una proteína de 355 residuos, de los que 2 corresponden a cisteínas. Se predijo en su extremo amino péptido señal, de 22 aminoácidos de extensión, 2 posibles sitios de *N*-glicosilación (posiciones 146 y 338), además de 12 dianas potenciales de *O*-glicosilación. Se trata de una proteína ácida (pI 5,38), termoestable (índice alifático de 96,90) y con cierto carácter hidrofóbico (valor “GRAVY” de 0,305). Esta última característica condujo a la búsqueda de una posible región transmembrana (por medio del servidor TMHMM 2.0, <http://www.cbs.dtu.dk/services/TMHMM/>) o de un posible anclaje a membrana por medio de un motivo glicosilfosfatidilinositol (GPI) (utilizando el servidor PredGPI, <http://gpcr.biocomp.unibo.it/predgpi/pred.htm>), pero en ninguno de los casos, se obtuvieron resultados positivos.

En el genoma de *Botrytis* se identificó otra proteína (Bcin06g06090) con una identidad mayor del 40% con Bcin12g03520, con características físico químicas muy parecidas (351 residuos aminoácidos, pI 5,04, y cierto carácter hidrófobo). Tampoco en este caso se pudo llegar a clasificar en ninguna familia proteica esta secuencia.

Bcin12g03520 es lo suficientemente abundante como para ser detectada en el glicosecretoma del hongo (14). Además, proteínas homólogas a Bcin12g03520 sólo fueron identificadas en Ascomycetes, sugiriendo que deben desempeñar funciones claves para el desarrollo de estos organismos. Por otro lado, su ausencia en plantas, las señala como proteínas de interés para valorar si en hongos fitopatógenos podrían identificarse como dianas para el diseño de nuevos fármacos destinados al control de estos organismos, sin afectar la biología de las plantas.

Lacasas: Bcin14g02510 y proteínas redundantes

Bcin14g02510 ha sido clasificada como un lacasa ya que presentó una identidad del 93,1% y un E-value de 0,0 con la lacasa de *Botrytis aclada* anotada en el UniProt como H8ZRU2 con una estructura tridimensional conocida (24). Las lacasas (EC 1.10.3.2) son miembros de una amplia familia de oxidasas que unen cobre, que catalizan la oxidación de un vasto abanico de sustratos orgánicos e inorgánicos, acompañada de la reducción de oxígeno molecular a agua. El análisis de la secuencia de Bcin14g02510 con el servidor Pfam identificó 3 dominios típicos de las oxidasas con múltiples centros de cobre, desde el extremo amino- al carboxi-terminal, los denominados tipo 3 o acoplado binuclear (PF07732), tipo1 o azul (PF00394) y tipo 2 o normal (PF07731).

Utilizando la secuencia de la lacasa de *Botrytis aclada* (H8ZRU2) como secuencia query, se realizó un Blast local en la base de datos del genoma de *Botrytis cinerea* y se pudieron identificar, además de Bcin14g02510, otros 8 genes que codifican para posibles lacasas (Tabla S1). En todas ellas se identificaron los 3 dominios típicos de las oxidasas de múltiples centros de cobre (tipos 1,2 y 3). Los parámetros resultantes de la caracterización *in silico* de todas estas proteínas se recoge en la Tabla S1 del “Anexo”. Los valores del porcentaje de identidad de las secuencias de estas proteínas con la de *B. aclada* oscilaron entre el 35,5% con Bcin02g07640 y 93,1% con Bcin14g02510. Todas ellas son proteínas ácidas, termoestables y solubles, con alto grado de glicosilación, especialmente Bcin01g00800, en la que se prevén 4 y 102 potenciales residuos *N*- y *O*-glicosilados, respectivamente. Destacar que dos de ellas, Bcin09g02050 y Bcin09g04830, carecen secuencia señal (Tabla S1).

Las lacasas fúngicas participan en el proceso de pigmentación de las esporas, en el proceso de lignificación de las paredes celulares y como factores de virulencia (25). Pueden proteger al hongo del efecto tóxico de las fitoalexinas y taninos que se producen durante la respuesta de defensa del hospedador (25). En el caso de *Botrytis cinerea*, ya se conocía desde hace tiempo la presencia de manera constitutiva de cantidades apreciables en el medio extracelular de proteínas con esta actividad (26), aunque hasta ahora no se había puesto de manifiesto la gran redundancia de estas enzimas en el hongo. Probablemente, esta gran cantidad de isoformas sea necesaria para detoxificar la gran variedad de compuestos de defensa con la que el hongo se encuentra al poder infectar a más de 1400 especies vegetales distintas y para participar en la degradación de la pared celular vegetal (junto con beta-glucosidasas, pectín-metilesterasas, poligalacturonasas y otras) de estructura y composición tan variadas (1).

Serina proteasas: Bcin15g04670 y proteínas redundantes

La secuencia de Bcin15g04670 mostró el mayor porcentaje de identidad (60%) con un E-value de 0,0 con la hipotética tripeptidil-peptidasa I (CZR69025) del Ascomycete *Phialocephala subalpina*. Ni esta peptidasa, ni ninguna con las que el porcentaje de identidad fue alto con Bcin15g04670 ha sido contrastada experimentalmente, y el análisis de su secuencia con el servidor *Pfam* reveló un dominio pro-kumamolisinina de activación (PF09286,) y un dominio peptidasa S8 típico de la familia subtilasa (PF00082). Las subtilasas (EC 3.4.21.62) se incluyen en las serina proteasas (EC 3.4.21), caracterizadas todas ellas por poseer un residuo de serina esencial en su centro activo, pero se diferencia del resto de las familias dentro de este grupo por presentar la triada Asp/His/Ser en su centro catalítico (27). El dominio pro-kumamolisinina suele actuar como un dominio de activación, ya que su corte provoca la activación de la enzima. Debido a que la clasificación de esta peptidasa dentro de una familia en concreto no fue clara, se prefirió clasificarla como una serina proteasa.

En este caso, al no haber encontrado una proteína contrastada experimentalmente en hongos homóloga a Bcin15g04670, el Blast local para analizar la posible redundancia de esta proteína en la base de datos que contiene la última anotación del genoma de *Botrytis*, se realizó utilizando la propia secuencia de Bcin15g04670 como secuencia query. Los resultados mostraron que además de Bcin15g04670, existen otros 9 genes que codifiquen para posibles serina proteasas (Tabla S2). Además de Bcin15g04670, sólo en otras dos (Bcin05g05830 y Bcin06g00330) se identificaron los dominios esperados (dominio de activación pro-kumamolisinina y dominio peptidasa S8). En las 8 restantes, sólo se identificó el dominio de

activación, aunque el análisis Blast de todas ellas, siguió caracterizándolas como peptidasas S8. El análisis de sus secuencias con los distintos servidores bioinformáticos se recoge en la Tabla S2 del “Anexo”. Al alinear las 9 secuencias con el Clustal Omega, se comprobó que el porcentaje de identidad con Bcin15g04670 osciló entre el 23,9% (Bcin03g02040) y el 42,1% (Bcin06g00330). De todas ellas, dos (Bcin06g00620 y Bcin09g01190) parecen ser secretadas por un mecanismo diferente al guiado por un péptido señal, y todas son catalogadas como proteínas termoestables (Tabla S2), sin embargo, dos de ellas (Bcin06g00620 y Bcin15g03150) muestran cierto carácter hidrófobo (Tabla S2).

En numerosos trabajos se ha relacionado el papel de las proteasas en la interacción planta-patógeno (28), aunque su función precisa no se ha elucidado. *Botrytis* podría utilizar estas proteasas para degradar las proteínas de la pared celular vegetal o proteínas con actividad antifúngica producidas por el propio hospedador.

Proteína Ecm33 con anclaje GPI a la pared: Bcin04g01310 y proteínas redundantes

La proteína Bcin04g01310 presentó un porcentaje de identidad del 43,4% con un E-value de $3,3 \times 10^{-92}$ con la proteína de pared Ecm33 (Q4WNS8) de *Aspergillus fumigatus*. El dominio característico de esta familia según el Pfam (PF12454) es relativamente pequeño (de aproximadamente 40 residuos) y en la proteína de *Botrytis* se localizó en el extremo amino de la secuencia, desde el residuo 1 hasta el 41.

Ecm33 es una proteína de anclaje GPI que se identificó por primera vez en *Saccharomyces cerevisiae* y, aunque aún se desconoce la función bioquímica de esta proteína, se ha sugerido una hipotética función en el ensamblaje de la pared celular y en el correcto ensamblaje de la capa de manoproteínas que rodea a esta envuelta celular tanto de los conidios como del micelio (29,30) La mutación del gen que codifica para la proteína Ecm33 de *A. fumigatus* provocó aberraciones morfogénicas tales como defectos en la separación de los conidios, un incremento en el diámetro de los conidios asociado a un incremento de quitina en la pared celular, y una mayor resistencia de los conidios a la destrucción por fagocitos, entre otras (30).

El análisis Blast local para identificar posibles proteínas redundantes con Bcin04g01310 fue infructuoso, por lo que esta proteína es la única proteína de la familia Ecm33 con anclaje GPI a la pared celular que codifica el genoma de *Botrytis*. Se trata de una proteína de 385 aminoácidos, con péptido señal, 3 posibles sitios de *N*-glicosilación y 22 residuos potencialmente *O*-glicosilados. Como muchas de las proteínas secretadas por *Botrytis*, es una proteína con pI cercano a 4 y termoestable (índice alifático de 89,25), aunque

presenta cierto carácter hidrófobo (valor “GRAVY” de 0,148). La búsqueda de un posible motivo de anclaje GPI con el servidor PredGPI identificó un posible motivo en los 26 residuos del extremo carboxilo de la secuencia. Este motivo podría explicar el carácter ligeramente hidrófobo de la proteína.

El análisis de proteínas homólogas a Bcin04g01310 en otros hongos ha puesto de manifiesto el papel de estas proteínas en la biogénesis de la pared celular (29,30). Además, el análisis *in silico* de los genomas fúngicos secuenciados identificó a las proteínas de la familia Ecm33, entre las 4 familias que están presentes en todos los hongos y que de alguna manera, todas ellas, participan en la síntesis, integridad y remodelación de la pared celular (31). En el presente trabajo se ha constatado que Bcin04g01310 es la única proteína de la familia Ecm33 codificada por el genoma de *Botrytis*, por lo que, teniendo en cuenta su probable papel en la biogénesis de la pared celular, se presenta como un posible factor de virulencia del hongo y plantea un estudio detallado del papel que realmente esta proteína desempeña en la biología del hongo y en su patogenicidad.

No identificada: Bcin12g04640 y proteínas redundantes

Los resultados de la búsqueda de secuencias homólogas a la proteína anotada como Bcin12g04640 y el análisis de su secuencia con el servidor Pfam sólo permiten concluir que se trata de una proteína con dominio de unión a FAD (PF01565). El análisis *in silico* de la secuencia (de 483 aminoácidos) permitió identificar péptido señal (de 15 aminoácidos con sitio de corte entre dos residuos de alanina) y 4 posibles residuos glicosilados (2 *N*-glicosilados, y otros 2 *O*-glicosilados). Su pI fue de 4,89, es termoestable y tiene cierto carácter hidrofóbico (valor “GRAVY” de 0,013). Dada la vaga clasificación a la que se ha llegado no se procedió a estudiar su redundancia.

La identificación de esta proteína no se pudo completar con las herramientas de bioinformática utilizadas en el presente trabajo. Sólo se ha podido clasificar como una proteína con dominio de unión a FAD. Sin embargo, dado que se identificó como la vigesimonovena más abundante en el glicosecretoma de *Botrytis* (14), debería desempeñar algún papel relevante en el desarrollo del hongo, y la señala para un futuro estudio en más profundidad.

Glucosa oxidasa: Bcin01g11310 y proteínas redundantes

La proteína Bcin01g11310 ha sido clasificada como una glucosa oxidasa por presentar una identidad del 50% con un E-value de 0,0 con la hipotética glucosa oxidasa de

(A0A100IJG9) de *Aspergillus niger*. Las glucosa oxidasas (EC 1.1.3.4) catalizan la oxidación del primer grupo hidroxilo de la glucosa, utilizando oxígeno molecular como aceptor de electrones. Están incluidas en la familia de las GMC oxidorreductasas, y el análisis de la secuencia de la proteína Bcin01g11310 con el servidor Pfam reveló que, en efecto, contiene un dominio GMC oxidorreductasa N-terminal que interacciona con FAD (PF00732) y otro C-terminal (PF05199). Dada la vaga clasificación a la que se ha llegado, no se procedió a estudiar su redundancia.

Bcin01g11310, de 596 aminoácidos, es una proteína con un alto grado de glicosilación (9 posibles residuos *O*-glicosilados y 7 sitios potenciales de *N*-glicosilación). Para esta proteína el servidor SignalP no predijo péptido señal clásico, pero se prevé como una proteína secretada según el servidor *SecretomeP 2.0*. Se trata de una proteína básica (pI 7,09), estable (índice alifático de 90,18), pero con un valor GRAVY positivo (0,036), indicando cierto carácter hidrófobo.

Beta-glucosidasas: Bcin10g05590 y proteínas redundantes

La proteína Bcin10g05590 mostró un porcentaje de identidad del 58,7% y un E-value de 0,0 con la beta-glucosidasa A (P87076) de *Aspergillus kawachii* (P87076). Las beta-glucosidasas (EC 3.2.1.21), incluidas en la familia 3 de las Glicosil Hidrolasas, catalizan la hidrólisis de residuos terminales no reductores de beta-D-glucósidos, liberando D-glucosa. Se caracterizan en sus secuencias dos dominios globulares, uno C-terminal (PF01915) implicado en la catálisis y, en ocasiones, en la unión al beta-glucano, y otro N-terminal (PF00933). Además, es muy frecuente que se distinga otro dominio con estructura tipo fibronectina III (“Fn3-like”, PF14310) de función desconocida. Estos 3 dominios se han identificado, utilizando el servidor *Pfam*, en la secuencia de Bcin10g05590.

El análisis Blast local con la secuencia de la beta-glucosidasa A de *Aspergillus kawachii* (P87076), permitió identificar en el genoma de *Botrytis* otros 11 genes que posiblemente codifiquen proteínas con actividad beta-glucosidasa (Tabla S3). Los porcentajes de identidad de las doce secuencias con la de la beta-glucosidasa A de *A. kawachii* oscilaron entre un 50,2% (Bcin09g06990) y un 58,7% (Bcin10g05590), siendo la identidad estimada entre estas dos proteínas de *Botrytis* del 36,9%.

La caracterización *in silico* de las posibles 12 beta-glucosidasas se recoge en la Tabla S3 en el “Anexo” de la presente memoria. Destacar que en todas ellas se identificaron los dos dominios globulares típicos de esta familia de proteínas, además del tercer dominio “Fn3-like”, ya citados anteriormente. También todas ellas se prevén como hidrofílicas (valor

GRAVY inferior a -0,049), excepto Bcin15g02440, que mostró cierto carácter hidrófobo. Curiosamente, esta es la única secuencia en la que se ha identificado un dominio de unión a celulosa implicado en la unión de la enzima al sustrato (32) (PF00734) y además es en la que se predicen mayor número de *O*-glicosilaciones (un total de 54), de los que más de la mitad se encuentran agrupados en la región bisagra entre el dominio semjante a fibronectina tipo 3 y el dominio de unión a celulosa. En dos proteínas (Bcin05g07270 Bcin14g00480) se localizó un cuarto dominio, el dominio PA14 (PF07691), que crea un inserto en ciertas beta-glucosidasas y que parece desempeñar un papel en la unión al sustrato, más que una función catalítica (32).

Es importante también destacar que de las 12 proteínas, sólo en 4 de ellas se predice un péptido señal responsable de la incorporación de las proteínas a la vía de secreción. El resto parece utilizar vías de señalización alternativas (Tabla S3). Precisamente, la beta-glucosidasa A de *Aspergillus kawachii* es una proteína anclada a la pared celular (33), pero que surge de la expresión del gen *bglA*, que dependiendo del procesamiento post-traducciona que sufra, da lugar a otras dos beta-glucosidasas extracelulares, denominadas en este caso, EX-1 y EX-2, respectivamente (33). Las condiciones de crecimiento del hongo se han propuesto como las moduladoras del destino final de la beta-glucosidasa en la célula. Curiosamente, para la enzima Bcin10g05590 de *Botrytis*, en la que no se predijo péptido señal, aparecen anotadas en la base de datos de la última anotación del genoma del hongo, dos isoformas, diferenciadas en la localización del ATG en cada una de ellas y en la inclusión o no de un intrón en la secuencia, muy cerca del extremo 5' del gen. Estas modificaciones darán lugar precisamente a diferencias en el extremo amino de las dos isoformas, que es justamente la región de la proteína que determinará su localización en la célula. La gran redundancia de estas enzimas en el hongo y las posibles modificaciones postraduccionales de las mismas que podrían aún más aumentar la variabilidad de estas enzimas en el hongo, ponen de manifiesto la complejidad que supone el estudio del papel de las mismas en la biología del hongo.

Xiloglucanasas: Bcin03g03630 y proteínas redundantes

La proteína Bcin03g03630 mostró un porcentaje de identidad del 55,6% con un E-value de $1,1 \times 10^{-90}$ con la endo-beta-1,4-glucanasa A específica de xiloglucano (O94218) de *Aspergillus aculeatus*. Estas enzimas (EC 3.2.1.151), incluidas en la familia 12 de las glicosil hidrolasas, denominadas también xiloglucanasas, catalizan la hidrólisis de enlaces endo-1,4-beta-D-glucosídicos en el xiloglucano (34). El xiloglucano es la principal hemicelulosa presente en la pared primaria de dicotiledóneas y monocotiledóneas no gramíneas. Actúa creando puentes cruzados entre microfibrillas de celulosas adyacentes para formar una red

celulosa-xiloglucano que constituye la principal estructura de soporte de la pared primaria (35). Analizando la secuencia de Bcin03g03630 con el servidor Pfam, se distinguió un único dominio conservado que permitió clasificarla precisamente en la familia 12 de las glicosil hidrolasas (PF01670).

Se determinó la existencia de sólo una proteína redundante con Bcin03g03630 codificada por el genoma de *Botrytis* anotada como Bcin13g02320, estimándose un porcentaje de identidad entre ellas del 42,9%. Para ambas se predijo péptido señal de 18 residuos, así como el dominio conservado que las incluye como glicosil hidrolasas de la familia 12. Sin embargo, en Bcin13g02320 también se identificó un posible dominio de unión a celulosa tipo 1, unido a la región catalítica, por una región bisagra altamente *O*-glicosilada (un total de 48 residuos). Por su parte, en Bcin03g03630 no se prevé ningún residuo con este tipo de modificación post-traduccional. Estas diferencias probablemente hacen que esta xiloglucanasa tenga un mayor carácter hidrófobo que Bcin03g03630.

Metalopeptidasas ácidas: Bcin16g02770 y proteínas redundantes

A pesar de que ninguna de las secuencia homólogas a la secuencia de la proteína Bcin16g02770 halladas en otros hongos cuenta con apoyo experimental a su posible actividad, la inmensa mayoría aparecen anotadas como posibles homólogas a metalopeptidasas ácidas (EC 3.4.24.39). El estudio de la secuencia de la proteína Bcin16g02770 con el servidor Pfam reveló un dominio conservado que la incluye en la familia peptidasa M35 (PF02102), incluida en la base de datos de peptidasas MEROPS (<https://www.ebi.ac.uk/merops/>) en el grupo de metalopeptidasas. Esta familia de proteínas utiliza zinc como cofactor.

La secuencia Bcin16g02770 se utilizó para realizar un Blast local en la base de datos de la última anotación del genoma de *Botrytis* y este análisis identificó una única proteína redundante (Bcin12g06300), con un porcentaje de identidad entre ellas del 32,4%. En ambas secuencias se identificó péptido señal (de 17 residuos en la primera secuencia y de 19, en la segunda), un número total similar de residuos (336 y 335, respectivamente) y también en ambas el grado de glicosilación fue muy similar (de 16 a 18 residuos, respectivamente, *O*-glicosilados). Sin embargo, Bcin16g02770 parece ser una proteína soluble (valor GRAVY de -0,158), mientras que Bcin12g06300 tiene cierto carácter hidrófobo (valor GRAVY de 0,045).

Alfa-1,2-manosidasas: Bcin01g05680 y proteínas redundantes

La manosil-oligosacárido-alfa-1,2-manosidasa 1B (Q12563) de *Aspergillus phoenicis* fue la proteína identificada en el análisis Blast que ha sido caracterizada experimentalmente con mayor porcentaje de identidad y E-value más bajo con Bcin01g05680. Estas enzimas (EC 3.2.1.113), incluidas en las glicosil hidrolasas 47, hidrolizan residuos terminales del alfa-D-manosa unidos por enlaces (1->2) de alfa-D-manósidos, importantes en la formación de nuevos *N*-glicanos y en la degradación de glicoproteínas (36). El análisis de la secuencia de Bcin01g05680 con el servidor Pfam, identificó un dominio conservado que la clasifica precisamente en la familia 47 de las glicosil hidrolasas (PF01532).

Siete fueron las proteínas encontradas redundantes con Bcin01g05680, codificadas por el genoma de *Botrytis*. Los parámetros físico-químicos de todas ellas obtenidos tras la caracterización *in silico* se recogen en la Tabla S4 del “Anexo”. Los porcentajes de identidad de las secuencias de estas proteínas con la secuencia de la proteína Q12563 oscilaron entre el 28,4% (Bcin16g04160) y el 57,5% (Bcin01g05680), siendo la identidad estimada entre estas dos últimas proteínas de *Botrytis* del 30,3%.

Exopoligalacturonasas: Bcin05g02520 y proteínas redundantes

La proteína Bcin05g02520 fue identificada como una exopoligalacturonasa ya que mostró una mayor identidad (58,3%) con un E-value de $1,1 \times 10^{-90}$ con la exopoligalacturonasa X (Q00293) de *Aspergillus tubingensis*. Estas enzimas (EC 3.2.1.67), incluidas en la familia 28 de las glicosil hidrolasas, catalizan la hidrólisis del último enlace glicosídico presente en el alfa-1,4-D-galacturónido, disminuyendo la longitud de este en un monómero y liberando D-galacturonato. Estos galacturonatos son monosacáridos con carga negativa que forman el esqueleto de los componentes pécticos de la pared celular, presente en plantas y algunos procariotas (37). En la secuencia de Bcin05g02520 se identificó un único dominio conservado (PF00295) que permitió incluir a la proteína en la familia 28 de las glicosil hidrolasas.

Por otro lado, se determinó la presencia de tres proteínas redundantes con Bcin05g02520 codificadas por el genoma de *Botrytis*. Los parámetros correspondientes a la caracterización *in silico* de todas ellas se plasman en la Tabla S3 del “Anexo”. La identidad de las secuencias de todas estas proteínas con la exopoligalacturonasa de *A. tubingensis* osciló entre el 24,3 % (Bcin05g08240) y el 58,3% (Bcin05g02520), siendo la identidad entre estas dos últimas del 24,0%. Para todas ellas se prevé secuencia señal que las dirija a la vía de secreción, posibles residuos *N*-glicosilados, y en todas, salvo en Bcin15g05030, también se

identifican posibles residuos *O*-glicosilados. Precisamente, esta proteína es la más pequeña y la que presenta un pI más bajo (Tabla S5). Todas tienen un marcado carácter hidrofílico, que las sitúa en el medio extracelular (Tabla S5).

Las exopoligalacturonasas intervienen en la degradación de la pectina, componente muy abundante en la pared de la mayoría de los hospedadores de *Botrytis*. Este hecho explicaría la redundancia de estas proteínas en el hongo. Las enzimas degradadoras de la pectina han sido ampliamente estudiadas en *B. cinerea* y, aunque el esfuerzo se ha centrado en el estudio de las endopoligalacturonasas, también se dispone de datos sobre las exopoligalacturonasas. Así, se ha descrito la expresión diferencial de dos exopoligalacturonasas (en su momento denominadas como exo-PG I y exo-PG II) durante la infección de hojas de pepino por el hongo (38) que podrían jugar un papel importante en la patogénesis del hongo en las etapas tempranas de la infección –lo que concuerda con el hecho de que la proteína objeto de estudio haya sido encontrada en el secretoma temprano-, así como en la maceración de los tejidos de la planta hospedadora (39).

Acetilesterasas: Bcin01g08910 y proteínas redundantes

La acetilesterasa (T2FKR1) de *Myceliophthora thermophila* mostró el mayor porcentaje de identidad y el E-value más bajo (34,2% y 4×10^{-56} , respectivamente) con Bcin01g08910, por lo que probablemente esta proteína pueda ser clasificada como acetilesterasa. La proteína T2FKR1 en particular, actúa como una exo-desacetilasa, catalizando la hidrólisis de ésteres acéticos para liberar alcohol y acetato (40).

Tres fueron también las proteínas redundantes con Bcin05g02520 codificadas por el genoma de *Botrytis*. Los parámetros correspondientes a la caracterización *in silico* de todas ellas se resumen en la Tabla S7 del “Anexo”. Las identidades entre las secuencias de estas proteínas y la secuencia de la acetilesterasa (T2FKR1) de *Myceliophthora thermophila* oscilan entre el 22,9% (Bcin07g03280) y 43% (Bcin08g00950), siendo la identidad entre estas dos del 20,0%. En las cuatro se predijo péptido señal, a excepción de la correspondiente a la proteína Bcin07g03280, que, según el servidor SecretomeP 2.0 Server, poseía algún mecanismo alternativo de entrada en la vía de secreción. El número de dianas de *N*-glicosilación encontradas en todas las secuencias fue bajo, entre uno y dos. El número de sitios potenciales de *O*-glicosilación mostró una variación notable entre secuencias, oscilando entre ninguno y treinta y tres. Todas las proteínas resultaron tener puntos isoeléctricos ácidos (entre 3,85 y 5,35) y resultaron ser hidrofílicas, a excepción de la proteína Bcin06g03920, que resultó ser hidrofóbica.

No identificada: Bcin06g06670 y proteínas redundantes

No se encontró ninguna proteína homóloga en otros organismos con una función caracterizada y recogida en la literatura con la secuencia de la proteína anotada como Bcin06g06670. De las 209 entradas que se recogen en el UniProt, 205 corresponden a organismos eucariotas –todos ellos hongos; entre ellos 203 pertenecientes al subreino Dikarya-, mientras que son tres las entradas referentes a organismos bacterianos, y una referente a arqueas.

Se predijo la existencia de péptido señal de 20 aminoácidos de extensión, con sitio de corte entre dos residuos de alanina. Sólo se halló un residuo que pudiera sufrir *N*-glicosilación y también un posible sitio de *O*-glicosilación. La proteína de 170 aminoácidos (2 de ellos, residuos de cisteínas), mostró un punto isoeléctrico de 4,33, índice alifático de 60,76 y un valor “GRAVY” de -0,327. No se encontraron dominios conservados en la secuencia. Una sola proteína redundante parece ser codificada por el genoma de *Botrytis*: Bcin03g08030. La secuencia de esta proteína poseía péptido señal de 21 aminoácidos de longitud. La longitud de la proteína madura resultó ser de 208 aminoácidos (4 cisteínas), con punto isoeléctrico 4,32, índice alifático 67,12 y valor “GRAVY” -0,224. No se encontró, sin embargo, ningún dominio conservado en su secuencia. La identidad de la proteína con Bcin06g06670 resultó ser del 24,5%.

Alfa-fucosidasa: Bcin08g05870 y proteínas redundantes

La proteína Bcin08g05870 ha sido identificada como una alfa-fucosidasa ya que en el análisis Blast mostró mayor identidad (50,8%) con la alfa-fucosidasa A (Q5AU81) de *Emericella nidulans*. Estas enzimas (3.2.1.51), incluidas en las glicosil hidrolasas 65, son activas sobre oligómeros de xiloglucano, y están involucradas en la hidrólisis de alfa-L-fucósidos, liberando L-fucosa y alcohol.

Sólo se identificó una proteína redundante con Bcin08g05870 codificada por el genoma de *Botrytis*, la anotada como Bcin07g06710. Estas proteínas son muy similares, de entre 771 y 747 aminoácidos, con 4-5 residuos potencialmente *N*-glicosilados, termoestables y solubles. Se diferencian, sin embargo, en que en la primera se identificó una posible secuencia señal y en la segunda, no; y en que Bcin08g05870 se prevén mayor número de residuos *O*-glicosilados (un total de 11, frente a los 4 de Bcin07g06710).

CONCLUSIONES

EnsemblFungi

Protein BLAST
protein ► protein

Pfam

ExpASY
Bioinformatics Resource Portal

Conclusiones:

1.- Las herramientas bioinformáticas utilizadas en este trabajo han sido útiles para identificar 10 de las 15 proteínas secretadas por *Botrytis cinerea* en diversas condiciones ambientales. Tres de ellas sólo han podido ser incluidas en superfamilias proteicas, y otras dos han sido catalogadas como de “función desconocida”.

2.- Las serina proteasas y las beta-glucosidasas aparecen como proteínas muy redundantes por lo que, para analizar su función en la virulencia de *Botrytis*, se harán necesarias herramientas de biología molecular que permitan estudiar múltiples genes de forma simultánea.

3.- *Botrytis cinerea* codifica para una única proteína Ecm33 con anclaje GPI a la pared, que en otros hongos interviene en la integridad de la pared celular. La ausencia de esta familia proteica en plantas, la convierte en una clara candidata como posible diana para el diseño de nuevos métodos de control de *Botrytis*.

Conclusions:

1. The bioinformatic tools used in this work have been useful to identify 10 of the 15 proteins secreted by *Botrytis cinerea* in various environmental conditions. Two of them have been included in protein superfamilies and other two have been classified as "unknown".

2. Serine proteases and beta-glucosidases appear as very redundant proteins, so that, in order to analyze their role in the *Botrytis* virulence, molecular biology tools will be necessary to allow the study of multiple genes simultaneously.

3. *Botrytis cinerea* encodes a single GPI-anchored protein Ecm33, which in other fungi is involved in cell wall integrity. The absence of this protein family in plants suggests that it may be a putative candidate target for the design of new methods for *Botrytis* control.

BIBLIOGRAFÍA

EnsemblFungi

Protein BLAST
protein ► protein

Pfam

ExPASy
Bioinformatics Resource Portal

Bibliografía

1. **Hahn, M., Viaud, M. y van Kan, J.A.L.** 2014. The genome of *Botrytis cinerea*, a ubiquitous broad host range necrotroph. *En: Dean, R., Lichens-Park, A. y Kole, C. (eds.), Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens.* Springer, Heidelberg. , p.19-44
2. **Dean, R., van Kan, J.A.L, Pretorius, ZA, Hammond-Kosack, KE, Pietro D, A., et al.** 2012. The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology* **13**: 414-430.
3. **Elad, Y., Vivier, M. y Fillinger, S.** 2016. *Botrytis*, the Good, the Bad and the Ugly. *En: Fillinger, S. and Elad, Y. (eds.), Botrytis – the fungus, the pathogen and its management in agricultural systems.* Springer, Heidelberg. p. 1-15.
4. **Van Kan, Jan A. L, Stassen, J.H.M., Mosbach, A., Van Der Lee, Theo A. J., Faino, L., et al.** 2017. A gapless genome sequence of the fungus *Botrytis cinerea*. *Molecular Plant Pathology* **18**: 75-89.
5. **De Miccolis Angelini, R.M., Pollastro, S. y Faretra, F.** 2016. Genetics of *Botrytis cinerea*. *En: Fillinger, S. and Elad, Y. (eds.), Botrytis – the fungus, the pathogen and its management in agricultural systems.* Springer, Heidelberg. p. 35-53
6. **González, C., Brito, N. y Sharon, A.** 2016. Infection Process and Fungal Virulence Factors. *En: Fillinger, S. and Elad, Y. (eds.), Botrytis – the fungus, the pathogen and its management in agricultural systems.* Springer, Heidelberg. p. 229-246
7. **Brito, N., Espino, JJ y González, C.** 2006. The endo-beta-1,4-xylanase xyn11A is required for virulence in *Botrytis cinerea*. *Molecular plant-microbe interactions* , **19**: 25-32.
8. **Kars, I., Krooshof, GH, Wagemakers, L., Joosten, R., Benen, JAE, et al.** 2005. Necrotizing activity of five *Botrytis cinerea* endopolygalacturonases produced in *Pichia pastoris*. *The Plant Journal* **43**: 213-225.
9. **Rolke, Y.,Liu, S.,Quidde,T.,Williamson,B.,Schouten,A.,et al.** 2004.Functional analysis of H2O2-generating systems in *Botrytis cinerea*: the major Cu-Zn-superoxide dismutase (BCSOD1) contributes to virulence on French bean, whereas a glucose oxidase (BCGOD1) is dispensable.*Molecular Plant Pathology* **5**:17-27.
10. **Schouten, A., van Baarlen, P. y van Kan, J.A.L.** 2008. Phytotoxic Nep1-like Proteins from the Necrotrophic Fungus *Botrytis cinerea* Associate with Membranes and the Nucleus of Plant Cells. *New Phytologist* **177**: 493-505.
11. **Have, At., Espino, JJ, Dekkers, E., Van Sluyter, SC, Brito, N., et al.** 2010. The *Botrytis cinerea* aspartic proteinase family. *Fungal Genetics and Biology* **47**: 53-65.
12. **Kars, I., McCalman, M., Wagemakers, L. y van Kan, J.A.L.** 2005. Functional analysis of *Botrytis cinerea* pectin methylesterase genes by PCR-based targeted mutagenesis: Bcpme1 and Bcpme2 are dispensable for virulence of strain B05.10. *Molecular Plant Pathology* **6**: 641-652.
13. **Espino, JJ, Gutiérrez-Sánchez, G., Brito, N., Shah, P., Orlando, R., et al.** 2010. The *Botrytis cinerea* early secretome. *Proteomics* **10**: 3020-3034.
14. **González, M., Brito, N. y González, C.**2014.Identification of glycoproteins secreted by wild-type *Botrytis cinerea* and by protein *O*-mannosyltransferase mutants. *BMC Microbiology* **14**,doi: 10.1186/s12866-014-0254-y
15. **González, M., Brito, N. y González, C.** 2012. High abundance of Serine/Threonine-rich regions predicted to be hyper-*O* -glycosylated in the secretory proteins coded by 8 fungal genomes. *BMC Microbiology* **12**, doi: 10.1186/1471-2180-12-213
16. **Altschul, SF, Madden,TL, Schäffer, A.A., Zhang, J., Zhang, Z., et al.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389-3402
17. **Hogeweg, P.** 2011. The Roots of Bioinformatics in Theoretical Biology. *PLoS Computational Biology* **7**, e1002021
18. **Aamer Mehmood, M.** 2014. Use of Bioinformatics Tools in Different Spheres of Life Sciences. *Journal of Data Mining in Genomics & Proteomics* **5**(2), <http://dx.doi.org/10.4172/2153-0602.1000158>
19. **Geer, RC y Sayers, EW.** 2003. Entrez: making use of its power. *Briefings in bioinformatics* **4**: 179-184.
20. **Thompson, JD, Higgins, DG y Gibson, TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.

21. Kersey ,PJ, Allen, JE, Armean, I., Boddu, S., Bolt, BJ, *et al.* 2016. Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Research* **44**: D574-D580.
22. Cantarel, BL, Coutinho, PM, Rancurel, C., Bernard, T., Lombard, V., *et al.* 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37**: D233-D238.
23. Hansen, J., Lund, O., Tolstrup, N., Gooley, A., Williams, K., *et al.* 1998. NetOglyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate Journal* **15**: 115-130.
24. Osipov, E., Polyakov, K., Kittl, R., Shleev, S., Dorovatovsky, P., *et al.* 2014. Effect of the L499M mutation of the ascomycetous *Botrytis aclada* laccase on redox potential and catalytic properties. *Acta Crystallographica Section D* **70**: 2913-2923.
25. Mayer, A. M. y Staples, R. C. 2002. Laccase: new functions for an old enzyme. *Phytochemistry* **60**: 551-565.
26. Viterbo, A., Yagen, B. y Mayer, AM. 1992. Cucurbitacins, 'attack' enzymes and laccase in *Botrytis cinerea*. *Phytochemistry* **32**: 61-65.
27. Wlodawer, A., Li, M., Oyama, H., Gustchina, A., Oda, K., *et al.* 2001. Carboxyl proteinase from *Pseudomonas* defines a novel family of subtilisin-like enzymes. *Nature Structural Biology* **8**: 442-446.
28. Karimi Jashni, M., Mehrabi, R., Collemare, J., Mesarich, CH y de Wit, P.J.G.M. 2015. The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant-pathogen interactions. *Frontiers in Plant Science* **6**, <https://doi.org/10.3389/fpls.2015.00584>
29. Lussier, M., White, AM, Sheraton, J., di-Paolo, T., Treadwell, J., *et al.* 1997. Large Scale Identification of Genes Involved in Cell Surface Biosynthesis and Architecture in *Saccharomyces cerevisiae*. *Genetics* **147**: 435-450.
30. Chabane, S., Sarfati, J., Ibrahim-Granet, O., Du, C., Schmidt, C., *et al.* 2006. GPI-Anchored Ecm33p Influences Conidial Cell Wall Biosynthesis in *Aspergillus fumigatus*. *Applied and Environmental Microbiology* **72**: 3259-3267.
31. Latgé, J. P. y Beauvais, A. 2014. Functional duality of the cell wall. *Current Opinion in Microbiology* **20**: 111-117.
32. Roberts, S.A., Weichsel, A., Grass, G., Thakali, K., Hazzard, J.T., *et al.* 2002. Crystal Structure and Electron Transfer Kinetics of CueO, A Multicopper Oxidase Required for Copper Homeostasis in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 2766-2771.
33. Iwashita, K., Nagahara, T., Kimura, H., Takano, M., Shimoi, H., *et al.* 1999. The *bglA* gene of *Aspergillus kawachii* encodes both extracellular and cell wall-bound beta-glucosidases. *Applied Environmental Microbiology* **65**: 5546-5553
34. Pauly, M., Andersen, LN, Kauppinen, S., Kofod, LV, York, WS, *et al.* 1999. A xyloglucan-specific endo-beta-1,4-glucanase from *Aspergillus aculeatus*: expression cloning in yeast, purification and characterization of the recombinant enzyme. *Glycobiology* **9**: 93-100.
35. Eckardt, N.A. 2008. Role of Xyloglucan in Primary Cell Walls. *The Plant Cell* **20**: 1421-1422.
36. Inoue, T., Yoshida, T. y Ichishima, E. 1995. Molecular cloning and nucleotide sequence of the 1,2-alpha-D-mannosidase gene, *msdS*, from *Aspergillus saitoi* and expression of the gene in yeast cells. *Biochimica et Biophysica Acta* **1253**: 141-145.
37. Molhoj, M., Verma, R. y Reiter, W. 2004. The Biosynthesis of D-Galacturonate in Plants. Functional Cloning and Characterization of a Membrane-Anchored UDP-D-Glucuronate 4-Epimerase from *Arabidopsis*. *Plant Physiology* **135**: 1221-30
38. Rha, E., Park, HJ, Kim, JW, Kim, MO y Chung YR, L.C. 2001. Expression of exo-polygalacturonases in *Botrytis cinerea*. *FEMS Microbiology Letters* **201**: 105-109.
39. Blanco-Ulate, B., Morales-Cruz, A., Amrine, K.C.H., Labavitch, JM, Powell, A.L.T., *et al.* 2014. Genome-wide transcriptional profiling of *Botrytis cinerea* genes targeting plant cell walls during infections of different hosts. *Frontiers in Plant Science* **5**, [doi: 10.3389/fpls.2014.00435](https://doi.org/10.3389/fpls.2014.00435)
40. Koutaniemi, S., van Gool, MP, Juvonen, M., Jokela, J., Hinz, SW, *et al.* 2013. Distinct roles of carbohydrate esterase family CE16 acetyl esterases and polymer-acting acetyl xylan esterases in xylan deacetylation. *Journal of Biotechnology* **168**: 684-692.

ANEXO

EnsemblFungi

Protein BLAST
protein ► proteins

Pfam

ExPASy
Bioinformatics Resource Portal

Tabla S1. Caracterización *in silico* de las lacasas codificadas por el genoma de *B. cinerea*.

Anotación ^a	aa ^b	Péptido señal ^c	Dominios conservados ^d	N-glyc ^e	O-glyc ^f	Cys ^g	pI ^h	AI ^h	GRAVY ^h
Bcin14g02510	562	19-20 (A-T)	Tipo 3: 74-191 Tipo 1: 197-353 Tipo 2: 406-547	3 (93, 120, 283)	18	9	4,44	80,73	-0,141
Bcin01g00800	692	18-19 (A-S)	Tipo 3: 213-330 Tipo 1: 336-490 Tipo 2: 524-681	4 (232, 259, 371, 383)	102	9	5,25	74,10	-0,197
Bcin01g07190	571	22-23 (S-Q)	Tipo 3: 81-200 Tipo 1: 206-369 Tipo 2: 409-560	1 (123)	18	7	5,09	67,46	-0,419
Bcin01g07950	545	20-21 (A-L)	Tipo 3: 72-189 Tipo 1: 195-341 Tipo 2: 400-529	5 (91, 118, 230, 288, 331)	34	7	5,72	81,80	-0,083
Bcin02g07640	644	19-20 (P-A)	Tipo 3: 147-265 Tipo 1: 271-422 Tipo 2: 489-628	2 (194, 308)	72	10	5,63	69,64	-0,285
Bcin07g06780	571	18-19 (A-Q)	Tipo 3: 71-190 Tipo 1: 196-358 Tipo 2: 396-556	4 (47, 51, 250, 364)	15	7	5,09	77,15	-0,277
Bcin09g02050	604	Si ¹	Tipo 3: 74-194 Tipo 1: 200-363 Tipo 2: 412-560	3 (33, 116, 364)	13	9	5,56	75,38	-0,327
Bcin09g04830	602	Si ¹	Tipo 3: 70-188 Tipo 1: 194-345 Tipo 2: 418-560	3 (117, 392, 466)	8	11	4,88	77,71	-0,271
Bcin13g00110	589	18-19 (P-S)	Tipo 3: 103-210 Tipo 1: 216-372 Tipo 2: 441-569	4 (50, 92, 228, 251)	8	7	5,14	78,64	-0,277

^aAnotación correspondiente a la última secuenciación del genoma de la cepa B05.10 de *B. cinerea*.

^bNúmero de residuos de aminoácidos de la proteína madura.

^cPunto de corte del péptido señal según el servidor SignalP4.1. Se indica las posiciones que limitan el punto de corte y los aminoácidos en cada una de ellas.

^dDominios oxidasa con multiples centro de cobre tipo 1 (PF00394), tipo 2 (PF07731) y tipo 3 (PF07732) según el servidor PFAM. Se indican las posiciones de inicio y fin de cada dominio.

^eNúmero de posibles residuos N-glicosilados y sus posiciones en la secuencia según el servidor NetNGlyc1.0.

^fNúmero de posibles residuos O-glicosilados y sus posiciones en la secuencia según el servidor NetOGlyc4.0.

^gNúmero de residuos de cisteínas.

^hParámetros físico-químicos estimados con el servidor ProtParam (pI, punto isoeléctrico; AI, índice alifático; GRAVY, valor medio de hidropatía).

¹Sistema no clásico de secreción, según el servidor SecretomeP 2.0.

Tabla S2. Caracterización *in silico* de las posibles serina proteasas codificadas por el genoma de *B. cinerea*.

Anotación ^a	aa ^b	Péptido señal ^c	Dominios conservados ^d	N-glyc ^e	O-glyc ^f	Cys ^g	pI ^h	AI ^h	GRAVY ^h
Bcin15g04670	596	19-20 (S-S)	Pro-kumamolisinina: 38-180 Peptidasa S8 : 336-551	0	10	6	5 ,07	73 ,37	-0 ,368
Bcin03g02040	622	20-21 (T-A)	Pro-kumamolisinina: 38-181	1 (47)	11	10	4 ,71	72 ,94	-0 ,277
Bcin05g05830	632	17-18 (A-A)	Pro-kumamolisinina: 36-179 Peptidasa S8: 371-593	0	12	9	4 ,69	71 ,79	-0 ,362
Bcin06g00330	613	19-20 (A-S)	Pro-kumamolisinina: 39-181 Peptidasa S8:3 16-574	2 (223, 518)	17	6	5 ,86	73 ,39	-0 ,359
Bcin06g01010	711	19-20 (A-S)	Pro-kumamolisinina: 97-293	0	9	10	5 ,69	65 ,79	-0 ,425
Bcin06g00620	596	Si ¹	Pro-kumamolisinina: 32-169	1 (252)	16	7	4 ,44	81 ,91	-0 ,045
Bcin08g01020	655	18-19 (A-A)	Pro-kumamolisinina:44-187	4 (226, 342, 535, 582)	15	9	5 ,04	71 ,95	-0 ,361
Bcin09g01190	601	Si ¹	Pro-kumamolisinina: 33-175	2 (204, 298)	10	8	5 ,68	78 ,57	-0 ,380
Bcin10g01890	651	18-19 (A-A)	Pro-kumamolisinina: 36-179	3 (222, 505, 615)	13	9	5 ,16	69 ,86	-0 ,424
Bcin15g03150	569	18-19 (A-L)	Pro-kumamolisinina: 36-181	2 (116, 576)	29	9	4 ,81	79 ,40	-0 ,009

^aAnotación correspondiente a la última secuenciación del genoma de la cepa B05.10 de *B.cinerea*.

^bNúmero de residuos de aminoácidos de la proteína madura.

^cPunto de corte del péptido señal según el servidor SignalP4.1. Se indica las posiciones que limitan el punto de corte y los aminoácidos en cada una de ellas.

^dDominio de activación Pro-kumamolisinina (PF09286) y dominio peptidasa S8 (PF00082,) según el servidor PFAM. Se indican las posiciones de inicio y fin de cada dominio.

^eNúmero de posibles residuos N-glicosilados y sus posiciones en la secuencia según el servidor NetNGlyc1.0.

^fNúmero de posibles residuos O-glicosilados y sus posiciones en la secuencia según el servidor NetOGlyc4.0.

^gNúmero de residuos de cisteínas.

^hParámetros físico-químicos estimados con el servidor ProtParam (pI, punto isoeléctrico; AI, índice alifático; GRAVY, valor medio de hidropatía).

¹Sistema no clásico de secreción, según el servidor SecretomeP 2.0.

Tabla S3. Caracterización *in silico* de las beta-glucosidasas codificadas por el genoma de *B. cinerea*.

Anotación ^a	aa ^b	Péptido señal ^c	Dominios conservados ^d	N-glyc ^e	O-glyc ^f	Cys ^g	pI ^h	AI ^h	GRAVY ^h
Bcin10g05590	887	Si ¹	GH3-N:78-367_GH3-C:426-680 Fn3-like :809-877	6 (30, 88, 238, 342, 349, 468)	14	8	4,51	76,79	-0,213
Bcin03g08710	824	21-22 (A-Q)	GH3-N:77-336_GH3-C:395-651 Fn3-like :767-835	4 (64, 311, 318, 563)	14	6	5,16	74,72	-0,315
Bcin05g07270	888	No	GH3-N:47-329_GH3-C:364-756 Fn3-like :795-869 PA14 :447-593	1 (713)	17	7	5,12	85,38	-0,298
Bcin06g04840	777	No ¹	GH3-N:46-373_GH3-C:413-659 Fn3-like :696-765	4 (111, 132, 212, 673)	8	4	4,69	87,12	-0,049
Bcin08g06830	805	22-23 (A-S)	GH3-N:53-316_GH3-C:357-683 Fn3-like :740-809	5 (24, 78, 195, 480, 598)	44	6	4,43	73,12	-0,200
Bcin09g02640	784	20-21 (G-Q)	GH3-N:93-349_GH3-C:413-663 Fn3-like :722-793	7 (58, 325, 332, 385, 571, 598, 610)	15	7	4,65	78,90	-0,155
Bcin09g04850	961	Si ¹	GH3-N:155-422_GH3-C:491-760 Fn3-like :876-952	3 (188, 404, 697)	9	7	5,10	78,20	-0,377
Bcin09g5460	829	Si ¹	GH3-N:120-371_GH3-C:425-686 Fn3-like :745-817	5 (188, 414, 655, 854, 889)	24	7	5,43	84,79	-0,134
Bcin09g06990	759	Si ¹	GH3-N:48-295_GH3-C:341-623 Fn3-like :679-745	7 (4, 28, 86, 172, 271, 390, 405)	12	8	4,87	82,49	-0,154
Bcin12g02830	820	Si ¹	GH3-N:105-312_GH3-C:391-682 Fn3-like :738-808	3 (59, 272, 334)	15	8	4,78	82,18	-0,084
Bcin14g00480	833	No	GH3-N:11-280_GH3-C:315-708 Fn3-like :741-812 PA14 :433-526	5 (59, 272, 334, 599, 626)	4 (45, 261, 486, 496)	9	5,34	87,24	-0,214
Bcin15g02440	807	17-18 (A-A)	GH3-N:58-310_GH3-C:354-599 Fn3-like :650-720 CBM :792-819	2 (223, 337)	54	11	5,62	82,64	0,018

^aAnotación correspondiente a la última secuenciación del genoma de la cepa B05.10 de *B. cinerea*.

^bNúmero de residuos de aminoácidos de la proteína madura.

^cPunto de corte del péptido señal según el servidor SignalP4.1. Se indica las posiciones que limitan el punto de corte y los aminoácidos en cada una de ellas.

^dDominios Glicosil Hidrolasa 3 C-terminal (GH3-C, PF01915); Glicosil Hidrolasa 3 N-terminal (GH3-N, PF00933); y dominio tipo fibrotecnica 3 (Fn3-like, PF14310), Dominio PA14 (PF07691) y dominio de unión a celulosa 1 (CBM) según el servidor PFAM. Se indican las posiciones de inicio y fin de cada dominio.

^eNúmero de posibles residuos N-glicosilados y sus posiciones en la secuencia según el servidor NetNGlyc1.0.

^fNúmero de posibles residuos O-glicosilados y sus posiciones en la secuencia según el servidor NetOGlyc4.0.

^gNúmero de residuos de cisteínas.

^hParámetros físico-químicos estimados con el servidor ProtParam (pI, punto isoeléctrico; AI, índice alifático; GRAVY, valor medio de hidropatía).

¹Sistema no clásico de secreción, según el servidor SecretomeP 2.0.

Tabla S4. Caracterización *in silico* de las alfa-1,2-manosidasas codificadas por el genoma de *B. cinerea*.

Anotación ^a	aa ^b	Péptido señal ^c	Dominios conservados ^d	N-glyc ^e	O-glyc ^f	Cys ^g	pI ^h	AI ^h	GRAVY ^h
Bcin01g05680	510	20-21 (A-A)	GH_47:54-525	1 (327)	14	2	4,52	73,73	-0,289
Bcin01g06120	731	No	GH_47:193-686	2 (28, 241)	12	3	5,54	70,52	-0,583
Bcin05g00140	1043	41-42 (A-M)	GH_47:57-600	2 (101, 696)	32	8	5,58	84,37	-0,291
Bcin06g04560	592	Si ¹	GH_47:104-590	0	11	6	5,45	80,08	-0,319
Bcin06g05970	568	No	GH_47:101-563	1 (164)	10	6	7,63	83,13	-0,243
Bcin07g05530	633	No	GH_47:105-630	1 (287)	14	6	5,60	79,98	-0,406
Bcin15g00880	996	No	GH_47:217-993	2 (419, 714)	43	7	5,28	72,14	-0,514
Bcin16g04160	573	No	GH_47:105-571	1 (171)	11	5	5,59	78,87	-0,308

^aAnotación correspondiente a la última secuenciación del genoma de la cepa B05.10 de *B. cinerea*.

^bNúmero de residuos de aminoácidos de la proteína madura.

^cPunto de corte del péptido señal según el servidor SignalP4.1. Se indica las posiciones que limitan el punto de corte y los aminoácidos en cada una de ellas.

^dDominio Glicosil Hidrolasa 47 (PF01532) según el servidor PFAM. Se indican las posiciones de inicio y fin de cada dominio.

^eNúmero de posibles residuos N-glicosilados y sus posiciones en la secuencia según el servidor NetNGlyc1.0.

^fNúmero de posibles residuos O-glicosilados y sus posiciones en la secuencia según el servidor NetOGlyc4.0.

^gNúmero de residuos de cisteínas.

^hParámetros físico-químicos estimados con el servidor ProtParam (pI, punto isoeléctrico; AI, índice alifático; GRAVY, valor medio de hidropatía).

¹Sistema no clásico de secreción, según el servidor SecretomeP 2.0.

Tabla S5. Caracterización *in silico* de las exopoligalacturonas codificadas por el genoma de *B. cinerea*.

Anotación ^a	aa ^b	Péptido señal ^c	Dominios conservados ^d	N-glyc ^e	O-glyc ^f	Cys ^g	pI ^h	AI ^h	GRAVY ^h
Bcin01g07250	420	25-26 (C-A)	GH_28: 94-441	2 (94, 436)	8 (52, 53, 56, 57, 62, 66, 67, 422)	10	7,08	81,71	-0,065
Bcin05g02520	437	19-20 (A-S)	GH_28: 107-448	4 (75, 146, 314, 440)	5 (42, 53, 56, 60, 65, 67)	10	4,84	76,52	-0,278
Bcin05g08240	438	19-20 (A-G)	GH_28: 95-386	4 (45, 85, 102, 173)	4 (36, 39, 43, 349)	9	6,18	73,52	-0,327
Bcin15g05030	395	19-20 (A-K)	GH_28: 50-347	7 (69, 80, 86, 95, 111, 228, 408)	0	10	4,56	80,20	-0,033

^aAnotación correspondiente a la última secuenciación del genoma de la cepa B05.10 de *B.cinerea*.

^bNúmero de residuos de aminoácidos de la proteína madura.

^cPunto de corte del péptido señal según el servidor SignalP4.1. Se indica las posiciones que limitan el punto de corte y los aminoácidos en cada una de ellas.

^dDominio Glicosil Hidrolasa 28 (PF00295) según el servidor PFAM. Se indican las posiciones de inicio y fin de cada dominio.

^eNúmero de posibles residuos N-glicosilados y sus posiciones en la secuencia según el servidor NetNGlyc1.0.

^fNúmero de posibles residuos O-glicosilados y sus posiciones en la secuencia según el servidor NetOGlyc4.0.

^gNúmero de residuos de cisteínas.

^hParámetros físico-químicos estimados con el servidor ProtParam (pI, punto isoeléctrico; AI, índice alifático; GRAVY, valor medio de hidropatía).

¹Sistema no clásico de secreción, según el servidor SecretomeP 2.0.

Tabla S6. Caracterización *in silico* de las acetilsterasas codificadas por el genoma de *B. cinerea*.

Anotación ^a	aa ^b	Péptido señal ^c	Dominios conservados ^d	N-glyc ^e	O-glyc ^f	Cys ^g	pI ^h	AI ^h	GRAVY ^h
Bcin01g08910	320	21-22 (G-A)	--	1 (62)	3 (67, 69, 293)	4	5,01	71,97	-0,257
Bcin06g03920	383	16-17 (A-S)	--	2 (158, 260)	33	4	3,85	76,48	0,158
Bcin07g03280	402	Si ¹	--	2 (101, 280)	5 (99, 103, 113, 182, 183)	5	5,15	75,22	-0,163
Bcin08g00950	285	18-19 (A-F)	--	1 (206)	0	2	5,35	71,96	-0,062

^aAnotación correspondiente a la última secuenciación del genoma de la cepa B05.10 de *B.cinerea*.

^bNúmero de residuos de aminoácidos de la proteína madura.

^cPunto de corte del péptido señal según el servidor SignalP4.1. Se indica las posiciones que limitan el punto de corte y los aminoácidos en cada una de ellas.

^dNo se identificó ningún dominio conservado según el servidor PFAM.

^eNúmero de posibles residuos N-glicosilados y sus posiciones en la secuencia según el servidor NetNGlyc1.0.

^fNúmero de posibles residuos O-glicosilados y sus posiciones en la secuencia según el servidor NetOGlyc4.0.

^gNúmero de residuos de cisteínas.

^hParámetros físico-químicos estimados con el servidor ProtParam (pI, punto isoeléctrico; AI, índice alifático; GRAVY, valor medio de hidropatía).

¹Sistema no clásico de secreción, según el servidor SecretomeP 2

