UNIVERSIDAD DE SALAMANCA

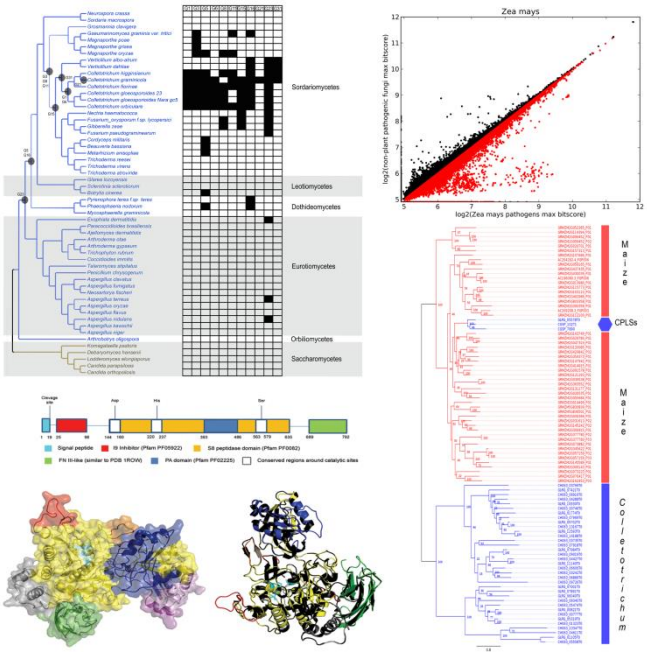
FACULTAD DE BIOLOGÍA

DEPARTAMENTO DE MICROBIOLOGÍA Y GENÉTICA

ÁREA: GENÉTICA

TESIS DOCTORAL

Transferencia horizontal de genes y mimetismo molecular en hongos fitopatógenos del género *Colletotrichum*



Vinicio Danilo Armijos Jaramillo

Salamanca, 2014

UNIVERSIDAD DE SALAMANCA

FACULTAD DE BIOLOGÍA

DEPARTAMENTO DE MICROBIOLOGÍA Y GENÉTICA

ÁREA: GENÉTICA

Transferencia horizontal de genes y mimetismo molecular en hongos fitopatógenos del género *Colletotrichum*

Tesis Doctoral

Programa de Doctorado: Agrobiotecnología

Órgano responsable del Programa de Doctorado:

Departamento de Fisiología Vegetal

Vinicio Danilo Armijos Jaramillo

Salamanca, 2014

UNIVERSIDAD DE SALAMANCA

FACULTAD DE BIOLOGÍA

DEPARTAMENTO DE MICROBIOLOGÍA Y GENÉTICA

ÁREA: GENÉTICA

Horizontal gene transfer and molecular mimicry in phytopathogen fungi of the genus *Colletotrichum*

PhD Thesis

Programa de Doctorado: Agrobiotecnología

Órgano responsable del Programa de Doctorado:

Departamento de Fisiología Vegetal

Vinicio Danilo Armijos Jaramillo

Salamanca, 2014

D. Luis Román Fernández Lago, Director del Departamento de Microbiología y Genética de la Facultad de Biología de la Universidad de Salamanca y **D. Carlos Nicolás Rodríguez**, Directora del Programa de Doctorado en Agrobiotecnología de la Universidad de Salamanca.

CERTIFICAMOS:

Que la presente Memoria titulada "**Transferencia horizontal de genes y mimetismo molecular en hongos fitopatógenos del género *Colletotrichum***", ha sido realizada en el Departamento de Microbiología y Genética de la Facultad de Biología y el Centro Hispano-Luso de Investigaciones Agrarias de la Universidad de Salamanca por **D. Vinicio Danilo Armijos Jaramillo**, bajo la dirección del Prof. D. Michael Ronald Thon y la Dra. Dña. Serenella Ana Sukno y cumple las condiciones exigidas para optar al grado de Doctor por la Universidad de Salamanca.

Para que así conste, firmamos el presente certificado en Salamanca a 31 de octubre de 2014.

Fdo: D. Luis Román Fernández Lago Fdo: D. Carlos Nicolás Rodríguez

D. Michael Ronald Thon, Profesor Titular del Departamento de Microbiología y Genética de la Facultad de Biología de la Universidad de Salamanca, y **Dña. Serenella Ana Sukno**, Personal Investigador Doctor del Centro Hispano-Luso de Investigaciones Agrarias y del Departamento de Microbiología y Genética de la Facultad de Biología de la Universidad de Salamanca

CERTIFICAMOS:

Que la presente Memoria titulada "**Transferencia horizontal de genes y mimetismo molecular en hongos fitopatógenos del género *Colletotrichum***", ha sido realizada en el Departamento de Microbiología y Genética de la Facultad de Biología y el Centro Hispano-Luso de Investigaciones Agrarias de la Universidad de Salamanca, bajo nuestra dirección, por **D. Vinicio Danilo Armijos Jaramillo**, y cumple las condiciones exigidas para optar el grado de Doctor por la Universidad de Salamanca.

Para que así conste, firmamos el presente certificado en Salamanca a 31 de octubre de 2014.

Fdo: Dr. D. Michael R. Thon

Fdo: Dra. Dña. Serenella A. Sukno

Fdo: D. Vinicio D. Armijos

Este trabajo se ha llevado a cabo en el Laboratorio 1 del Centro Hispano-Luso de Investigaciones Agrarias (CIALE), Departamento de Microbiología y Genética, de la Universidad de Salamanca bajo la dirección del Profesor Dr. Michael R. Thon y de la Dra. Serenella A. Sukno. Durante el desarrollo de la Tesis he disfrutado de una beca de de la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador

Publicaciones científicas relacionadas con los resultados obtenidos en esta Tesis Doctoral y otros trabajos publicados durante el periodo de formación predoctoral:

- Sanz-Martín, J.M., **Armijo-Jaramillo, V**., Bello, V., Vargas, W.A., Monod, M., Díaz-Mínguez, J.M., Thon, M.R. and Sukno, S.A. 2014. An effector conserved across the kingdom Fungi has a role in virulence in *C. graminicola* and targets a plant resistance protein (manuscript in preparation for New Phytologist)

- Vargas, W.A., Sanz-Martín, J.M., Rech, G.E., **Armijos-Jaramillo**, **V.,** Rivera, L.P., Echeverria, M.M., Díaz-Mínguez, J.M., Thon, M.R. and Sukno, S.A. 2014. A novel fungal effector with host nuclear localization and DNA-binding properties is required for penetration of the maize epidermis (manuscript submitted to The Plant Cell, MS ID#: PLANTCELL/2014/133629)

- Rech, G.E., Sanz-Martin, J.M., Baroncelli, R., **Armijos-Jaramillo, V.**, Vargas, W.A., Sukno, S.A. and Thon, M.R. 2014. Comparative genomic analysis of *Colletotrcihum graminicola* reveals modularity in genome organization (manuscript in preparation)

- **Armijos-Jaramillo, V**., Sukno, S.A, Thon, M.R. A genome wide survey of molecular mimicry in the maize pathogen *Colletotrichum graminicola* (manuscript in preparation)

- **Armijos-Jaramillo, V**., Sukno, S.A, Thon, M.R. Identification of horizontally transferred genes in the genus *Colletotrichum* reveals a steady tempo of bacterial to fungal gene transfer (manuscript submitted to BMC Genomics, MS ID: 4981206081413646)

- **Armijos Jaramillo, V.D.**, Vargas W.A., Sukno S.A, Thon M.R. (2013). Horizontal transfer of a subtilisin gene from plants into an ancestor of the plant pathogenic fungal genus *Colletotrichum*. PLos ONE; 8:e59078

- **Armijos Jaramillo, V.D.**, Vargas W.A., Sukno S.A, Thon M.R. (2013). New insights into the evolution and structure of *Colletotrichum* plant-like subtilisins (CPLSs). Commun Integr Biol; 6: e25727

CUANTO MÁS COMPRENSIBLE PARECE EL UNIVERSO, MÁS ABSURDO PARECE TAMBIÉN.

Steven Weinberg.

AGRADECIMIENTOS

Quisiera empezar esta sección agradeciendo a mis directores de tesis Mike Thon y Serenella Sukno por darme la oportunidad de realizar este trabajo en su grupo. Asimismo, debo agradecer la dedicación que han tenido para corregir mi trabajo y discutir las ideas plasmadas en este documento.

Agradezco a la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador por concederme la beca que me ha permitido desarrollar esta tesis.

Quisiera reconocer el aporte de mis compañeros de grupo Gabriel, Walter y José, con quienes he colaborado durante este tiempo y de quienes he aprendido un sinfín de cosas, ¡gracias chicos!

Por otro lado, me gustaría agradecer a los Agrobiotecnólogos pseudosalmantinos, mis compañeros de master: Marta, Teresa, Wilson, Irene, Estefanía, Rubén, Daniel, Nieves, Javier y Juan. Junto a ellos he pasado cuatro años estupendos y hemos cultivado una bonita amistad que estoy seguro perdurará en el tiempo y soportará la distancia.

De igual forma quisiera agradecer a mis compañeros del laboratorio 1, con quienes he compartido el día a día y que considero son un grupo excepcional en lo personal y profesional. Debo reconocer especialmente a Ernesto por habernos acogido afectuosamente desde el primer día a Daniela y a mí. Me faltan palabras de agradecimiento y no tengo más que expresarte mi gratitud infinita.

A toda la gente del CIALE, compañeros, profesores y personal de administración y servicios muchas gracias. En este centro me he sentido como en casa y auguro éxitos a cada uno de vosotros. Quisiera ofrecer un agradecimiento especial a Alex, Vicky, Ángel y Jorge con quienes formamos un grupo muy ameno tanto dentro como fuera del CIALE. Además, no me puedo olvidar de la gente del CIALE FC por todos esos buenos momentos, campeonatos, victorias y derrotas, especialmente a Joni y su entusiasmo en cada chilena.

Finalmente quiero terminar este apartado agradeciendo a mi familia por el apoyo incondicional. Muchas gracias a mis hermanos y mis sobrinos y ante todo a mis padres quienes han conseguido formar una familia cálida y unida fruto de su respeto y cariño. Aunque merezca una sección exclusiva, quisiera dedicar estas últimas líneas a Daniela. Durante esta tesis tú has sido mi fortaleza, mi soporte, mi respaldo y mi motivación, me alegra haber emprendido esta aventura junto a ti y simplemente quería decirte ¡un millón de gracias!

Vinicio Danilo Armijos Jaramillo

ÍNDICE TEMÁTICO

RESUMEN

El género *Colletotrichum* engloba especies fitopatógenas de gran importancia a nivel mundial. Prácticamente cada cultivo en el mundo es susceptible al ataque de una o más especies de este género. Las especies del género *Colletotrichum* utilizan estrategias de infección que van desde la hemibiotrofía intracelular (el hongo ingresa en la células huéspedes y las mantiene vivas y se alimenta de ellas para después matarlas y alimentarse de este material) hasta la necrotrofía subcuticular o intramural (después de la penetración de la cutícula de la planta el hongo causa la muerte del tejido epidérmico vegetal manteniéndose por debajo de la cutícula). Sea cual sea la estrategia utilizada el patógeno requiere de una serie de moléculas especializadas para evitar ser reconocido por parte de la planta, manipular a la célula huésped o inhibir la respuesta defensiva. Estas moléculas que son secretadas al espacio extracelular o al interior de la célula vegetal se denominan efectores. Los efectores pueden ser moldeados durante la evolución de las especies por las presiones de selección provenientes de la interacción planta-patógeno, pero se ha observado también que este tipo de moléculas pueden ser adquiridas de especies cercanas o lejanas por el fenómeno conocido como transferencia horizontal de genes.

La transferencia horizontal de genes (THG) es el intercambio de material genético entre especies relacionadas o no relacionadas sin que se produzca apareamiento sexual entre los individuos. Es bien conocido que en procariotas la THG es un fenómeno que ha ocurrido y ocurre con relativa frecuencia, mientras que en eucariotas ha sido catalogado como un evento poco frecuente e incluso aislado. Sin embargo, en los últimos años los casos descritos de THG en eucariotas indican que la trasferencia horizontal es más común de lo que se creía. La THG puede traer beneficios inmediatos al organismo receptor, siempre y cuando el material genético desempeñe una función que incremente la aptitud del individuo receptor. En este posible escenario, el gen transferido se mantendría en el tiempo y los eventos de transferencia ocurridos en el pasado podrían ser detectados en las secuencias de especies actuales. Para detectar estos casos de THG se dispone de métodos de reconstrucción filogenética y de la comparación entre el potencial gen exógeno y el genoma que lo acoge.

Un fenómeno singular que puede ser el resultado de la THG es el mimetismo molecular. Este término se refiere a la cualidad que poseen ciertas moléculas para imitar a otras y con esto efectuar una función biológica precisa. En el caso específico de los patógenos, a las moléculas miméticas se las describe como los factores codificados por los patógenos que se asemejan a los factores del huésped produciendo una ventaja para el patógeno. El mimetismo molecular se puede originar por la THG, por la evolución convergente o por la similitud de familias

génicas en especies no relacionadas. Este tema ha sido ampliamente estudiado en animales pero en patosistemas que involucran plantas es un campo poco explorado.

El objetivo de este trabajo consistió en estudiar los eventos de trasferencia horizontal y mimetismo molecular en especies del género *Colletotrichum*.

El primer capítulo trata sobre la transferencia horizontal de genes en especies fitopatogénicas del género *Colletotrichum*. Para esto se desarrolló un estricto protocolo de reconstrucción filogenética proteína a proteína en tres especies de *Colletotrichum* de las que se dispone de genomas secuenciados. Los resultados fueron escogidos en primera instancia de manera automática (de acuerdo a los resultados de BLAST de cada proteína) y confirmados por varios filtros de curación manual (evaluando patrones de árboles filogenéticos). Como resultado se identificaron 12 eventos de transferencia horizontal (proteínas presentes en una o en varias de las especies estudiadas) con un claro patrón de THG. Once de los 12 eventos de THG identificados se originaron en bacterias mientras que el origen del evento restante fue identificado en plantas. Este último grupo fue investigado en profundidad en el segundo capítulo.

Los 11 eventos de THG de origen bacteriano fueron predichos como proteínas involucradas en el metabolismo de aminoácidos, metabolismo secundario, metabolismos de lípidos y carbohidratos, así como en la interacción con la célula huésped y la virulencia de estas especies. La transferencia de cada uno de estos grupos hacia *Colletotrichum* fue datada por la aproximación del reloj molecular. Este análisis reveló que la transferencia ha sido constante durante la evolución de la sub-división de los Pezizomycota. De igual forma, la comparación con el árbol de especies de varios pezizomycotas reveló que los genes transferidos horizontalmente se han mantenido en unas pocas especies actuales y se han perdido en la mayoría de las especies restantes, lo cual nos indica que no todos los genes transferidos horizontalmente son susceptibles de mantenerse en todas las especies. En conjunto los resultados sugieren que los eventos de THG encontrados en el primer capítulo, a pesar de la antigüedad de las transferencias, han sido retenidos en las especies de *Colletotrichum* para conferirles plasticidad metabólica. Y es posible que dichos genes hayan sido útiles para colonizar nuevos nichos o hayan incrementado la eficacia biológica de estas especies.

El segundo capítulo trata sobre un caso de transferencia horizontal cuyo origen fue identificado en plantas. Este gen que puede ser identificado en los genomas de varias especies

de *Colletotrichum* presenta una remarcable semejanza a un tipo de subtilisinas particulares de plantas. Las subtilisinas son un tipo particular de serin-proteasas ampliamente distribuidas en todos los reinos de la vida. La hipótesis manejada en este capítulo fue que un gen codificante de una subtilisina fue transferido horizontalmente desde las plantas hacia un ancestro de *Colletotrichum* spp. Para poner a prueba esta hipótesis se realizaron análisis filogenéticos así como la caracterización a nivel de estructura primaria, secundaria y terciaria de esta proteína. Debido a la semejanza estructural que presenta esta subtilisina de *Colletotrichum* con las subtilisinas de plantas se propuso que este puede ser un caso de mimetismo molecular. Por esta misma razón a esta subtilisina se la denomino CPLS (por sus siglas en inglés de *Colletotrichum plant-like subtilisin*). Los análisis filogenéticos revelaron que el evento de THG ocurrió aproximadamente hace 150 millones de años, fecha posterior a la divergencia de *Colletotrichum* como un género independiente y anterior a la divergencia entre monocotiledóneas y dicotiledóneas. El análisis de expresión mostró que el gen *CPLS* es modulado durante la infección en maíz por *Colletotrichum graminicola*, sugiriendo que este gen tiene un rol en la patogenicidad del hongo. Además, el pico de expresión de *CPLS* coincide con la represión de varias subtilisinas de maíz. Basados en el conjunto de evidencias mostradas en el capítulo dos, las proteínas CPLSs además de ser un fenómeno evolutivo singular parecen tener una función importante para las especies de *Colletotrichum*.

El tercer capítulo trata sobre la búsqueda de proteínas de *C. graminicola* que potencialmente imiten a proteínas de su huésped *Zea mays*. Para lograr la identificación de dichas proteínas fue necesario desarrollar métodos para el patosistema *C. graminícola*-maíz, ya que la búsqueda de candidatos de mimetismo molecular ha sido realizada para patógenos de humanos pero no para patógenos de plantas. El primer método desarrollado para este propósito estuvo basado en los métodos *in silico* existentes para la detección de mimetismo molecular en patógenos de humanos. Estos métodos se basan en la similitud de secuencia entre la proteína del patógeno y el huésped. Adicionalmente se usaron otros criterios como la colocalización entre los candidatos a proteína imitadora e imitada y/o la coexpresión de los mismos para aumentar la probabilidad de encontrar casos de mimetismo molecular. El segundo método estuvo basado en la búsqueda de dominios casi exclusivos de plantas que son compartidos por unos pocos hongos. Este método permite identificar dominios comunes a plantas y hongos fitopatógenos que no son compartidos por otros hongos no fitopatógenos.

De la aplicación del primer método en los proteomas de *C. graminícola* y *Zea mays* se obtuvieron 30 candidatos que fueron anotados como: Peptidasas, lipasas, proteína de tipo LysM, activadores de moléculas pequeñas de unión a GTP (moléculas bien caracterizadas como imitadoras en bacterias patógenas de humanos), proteasas entre otras. Varias de estas categorías funcionales suelen asociarse a la patogenicidad y la defensa en hongos y plantas respectivamente. Los resultados obtenidos fueron sumamente interesantes dado que aunque no existen casos descritos de mimetismo molecular en hongos fitopatógenos, varios de los candidatos encontrados coinciden con casos de mimetismo molecular descritos en otros sistemas. Así por ejemplo, se encontraron factores de intercambio de guanina (GEF por sus siglas en inglés) y una GTPasa de pequeño tamaño (de tipo ARF) entre las posibles proteínas imitadas en maíz. Sus pares en *C. graminícola* comparten similitud de secuencias y potencialmente el mismo espacio subcelular. Tanto las GEFs como las GTPasas de pequeño tamaño han sido descritas como blanco de mimetismo molecular en bacterias patógenas de animales, lo cual nos lleva a pensar que pueden ser imitadas también por los patógenos de plantas como *C. graminícola*. Además se logró identificar una proteína fúngica LysM (secuestradora de quitina) que aunque no es reconocida en la literatura como un caso de mimetismo molecular, nuestros resultados lo sugieren. La proteína LysM en *C. graminícola* fue identificada por su semejanza a una proteína LysM de su huésped. Además, este par de proteínas coinciden aparentemente en el mismo espacio subcelular cuando se produce la infección.

Con el segundo método utilizado se obtuvieron 3 candidatos. Dentro de estos cabe destacar la presencia de una peroxidasa tipo III típica de plantas. Este candidato fue encontrado en *C. graminícola* y otras especies del género. Asimismo, esta peroxidasa corresponde a uno de los candidatos encontrados por el primer método descrito anteriormente.

Como resultado de este trabajo se puede concluir que las especies fitopatógenas del género *Colletotrichum* registran eventos de transferencia horizontal en sus genomas y que la especie *C. graminicola* podría estar usando varias de sus proteínas para imitar moléculas de su huésped *Zea mays*. Estos eventos ponen de manifiesto los complejos mecanismos evolutivos que han llevado a las especies del género *Colletotrichum* a convertirse al día de hoy en agentes fitopatógenos tan eficaces.

INTRODUCCIÓN

En la actualidad la agricultura tiene que hacer frente a una demanda cada vez mayor de alimento, pienso para animales y combustibles. La necesidad de estos productos por parte de una población que se estima crecerá hasta los 9.1 mil millones de personas para el año 2050 (Godfray et al. 2010) obliga a que las prácticas agronómicas deban ser cada vez más eficientes en términos de producción total y rendimiento por hectárea. Uno de los principales factores limitantes en la producción agrícola son las enfermedades (Strange and Scott 2005; Dodds 2010). Aunque en la práctica resulta complicado estimar la pérdida producida por enfermedades y plagas en los cultivos, a finales del siglo pasado se estimó que dicha pérdida a nivel global asciende a $ 76.1 mil millones (Oerke 1999).

Los hongos se encuentran entre los principales agentes causales de enfermedades en los cultivos (Agrios 2005). Se calcula que solo por las pérdida producidas por hongos en los 5 cultivos más importantes (arroz, trigo, maíz, patata y soja) se podría alimentar alrededor de 596 millones de personas al año (Fisher et al. 2012). Existen más de 10000 especies de hongos con capacidad de atacar plantas (Agrios 2005). Sin embargo, existen unos cuantos géneros y especies particularmente relevantes por su incidencia en importantes cultivos, tal es el caso de *Colletotrichum* spp (Dean et al. 2012).

El género *Colletotrichum*

El género *Colletotrichum* incluye a numerosas especies de hongos fitopatógenos de importancia que afectan tanto a plantas herbáceas como a árboles, aunque virtualmente cada especie cultivada es susceptible al ataque de una o más especies de este género (Dean et al. 2012). Su distribución es predominantemente tropical y subtropical aunque se encuentra también en climas más templados. En cultivos tan importantes como banana, café, aguacate, caña de azúcar, manzana, pepino, nabo, sorgo o maíz (**Figura 1**) las especies de este género pueden producir pérdidas sustanciales (Bailey et al. 1992; Dickman 2000; Cannon et al. 2012). Tras la cosecha, la infección de frutas con *Colletotrichum* puede ocasionar pérdidas del 100% de la producción (Theodoro et al. 2004; Talhinhas et al. 2005; Damasceno e Silva et al. 2007; Dean et al. 2012) . Las especies del género *Colletotrichum* también son capaces de producir otras enfermedades además de la Antracnosis, como la podredumbre roja en caña de azúcar, la mancha marrón de la judía o la podredumbre apical en fresa y banana (Lenné 2002).

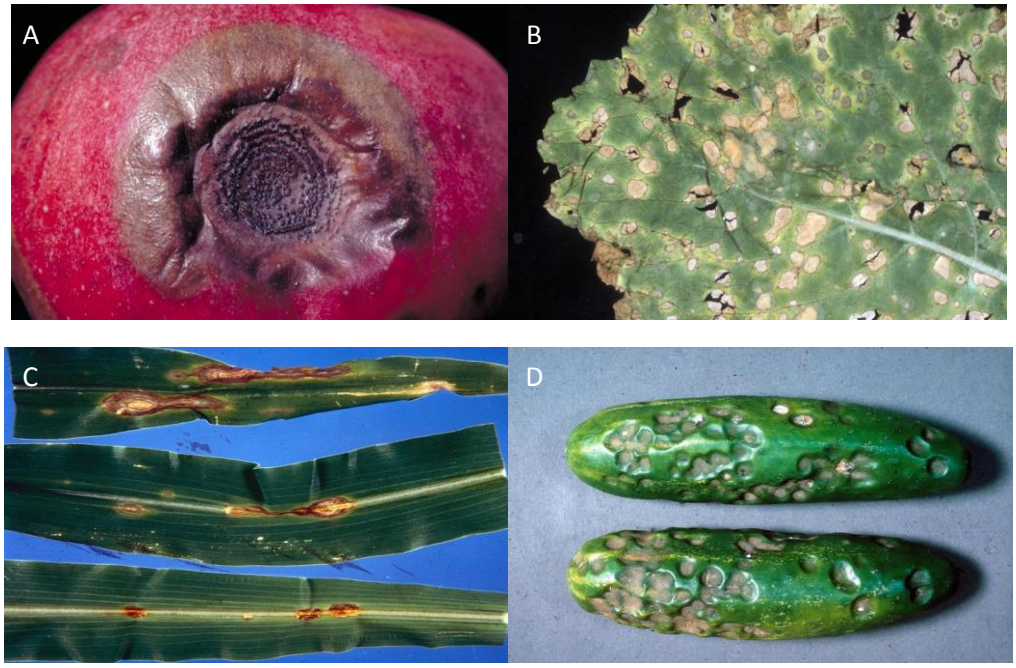


Figure 1. A) Síntomas de antracnosis causados por la infección de *Colletotrichum gloeosporioides* en manzana (Cheryl Kaiser, University of Kentucky, Bugwood.org). B) síntomas causados por la infección de *Colletotrichum higginsianum* en *Brassica rappa ssp. rapa* L. (J.C. Wells, North Carolina State University, Bugwood.org). C) Síntomas causados por la infección de *Colletotrichum graminicola* en hojas de maíz (J.C. Wells, North Carolina State University, Bugwood.org). D) Síntomas de antracnosis causados por la infección de *Colletotrichum orbiculare* en pepino (Charles Averre, North Carolina State University, Bugwood.org). Todas las fotos son usadas bajo licencia Creative Commons Attribution-NonCommercial 3.0 y se editó el código marcado por www.ipmimages.org.

Para desarrollar la infección, las especies de *Colletotrichum* utilizan diferentes estrategias que van desde la hemibiotrofía intracelular (ver debajo) hasta la necrotrofía subcuticular o intramural (el hongo causa la muerte del tejido epidérmico vegetal manteniéndose por debajo de la cutícula) (Bailey et al. 1992). En el caso de la hemibiotrofía intracelular la infección comienza con la germinación de las esporas (conidias) en la superficie de las hojas. A partir de los conidios se producen los tubos germinativos que a su vez desarrollan un apresorio melanizado, que constituye el órgano de penetración del hongo. El apresorio secreta enzimas líticas para la penetración y genera una estructura conocida como estaca o hifa de penetración. La hifa de penetración atraviesa la cutícula y las paredes celulares y forma una hifa primaria (estructura alargada e irregular también llamada hifa de infección) que invagina la membrana plasmática de la célula vegetal. A partir de la hifa primaria se desarrolla una o varias ramificaciones que se dispersan por las células adyacentes (**Figura 2**) (O'Connell et al. 1985; O'Connell et al. 2012). Esta primera etapa se denomina biotrófica ya que la planta no presenta

síntomas de enfermedad y las células se mantienen vivas. En una segunda etapa las hifas primarias se ramifican en una gran cantidad de pequeñas hifas secundarias que invaden la membrana plasmática de la célula vegetal así como el espacio intercelular. En esta etapa el hongo mata a las células vegetales y se alimenta de ellas. Dado que este tipo de organismos pasan por estas dos etapas de nutrición (biotrofía y necrotrofía) se los denomina hemibiotrofos (O'Connell et al. 1986; Münch et al. 2008). La duración de cada etapa y el modo en que colonizan el tejido vegetal las hifas primarias y secundarias son diferentes en las especies de *Colletotrichum* como así se ha demostrado en *C. graminícola* y *C. higginsianum* (Mims and Vaillancourt 2002; O'Connell et al. 2012; Yi and Valent 2013). Sin embargo, en especies patogénicas tales como *C. gloeosporioides* o *C. acutatum* incluso se han descrito relaciones epífitas y endófitas con la planta (Perfect et al. 1999; Freeman et al. 2001).

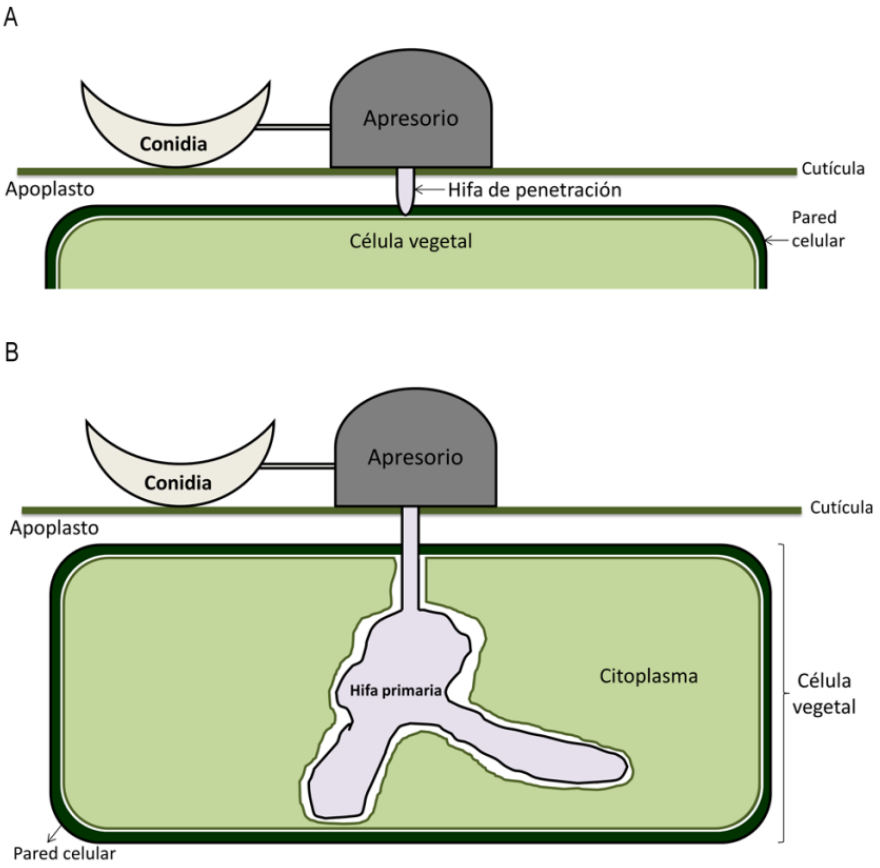


Figura 2. Esquema de la germinación de esporas y proceso de penetración en planta de *Colletotrichum spp*. A) Ingreso de la hifa de penetración a través de la cutícula y la pared celular de la planta. B) Invaginación del citoplasma de la célula vegetal producido por la hifa del hongo.

El estado teleomórfico de *Colletotrichum* se denomina *Glomerella* y se encuentra clasificado dentro de la división ascomycota, sub-división pezizomycotina, clase sordariomycetes, sub-clase hypocreomycetidae, orden glomerellales, familia glomerellaceae (Zhang et al. 2006; Réblová et al. 2011; Cannon et al. 2012; Zhang et al. 2013). En la **Figura 3** se muestra parte del árbol de los pezizomycotina y en él se puede observar la relación del género *Colletotrichum* con otras especies.

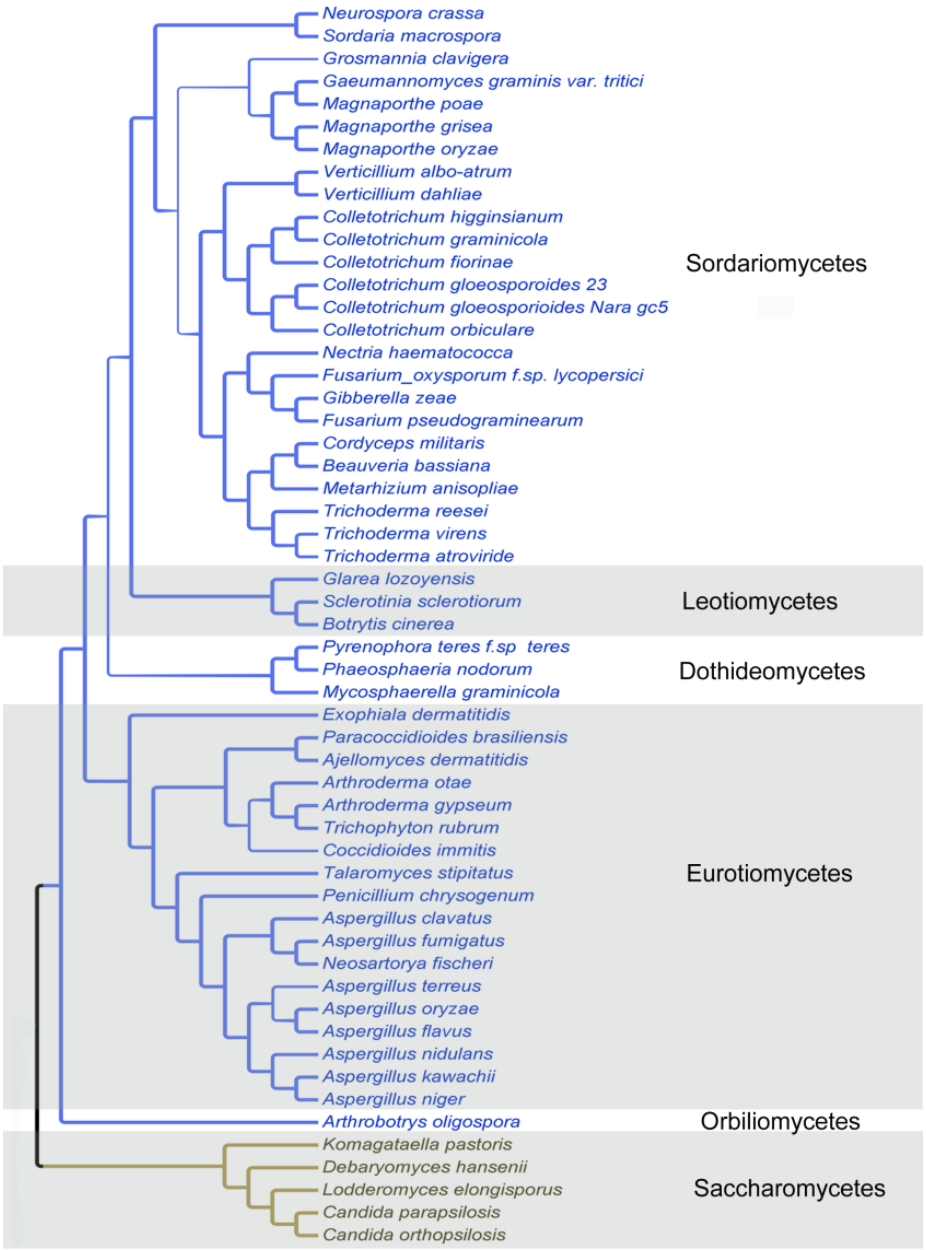


Figura 3. Árbol filogenético de varias especies de Pezizomycotina (en azul) en el que se muestra la ubicación del género *Colletotrichum* dentro de los Sordariomycetes y su relación con otros clases. Especies de la sub-división Saccharomycotina (amarillo) fueron tomadas como grupo externo. Figura modificada del capítulo 1.

El género *Colletotrichum* presenta más de 100 especies diferenciadas (Cannon et al. 2012). En la **Figura 4** se pueden observar el árbol filogenético de algunas de las especies más importantes de este género repartidas en clados. Los hongos de este género son importantes como modelos experimentales en estudios que abarcan diversos aspectos de la fitopatología incluyendo procesos infectivos, resistencia del hospedador y biología molecular de las interacciones planta-patógeno (Perfect et al. 1999; Latunde-Dada 2001; Thon et al. 2002; Sukno et al. 2008). Como resultado del interés por este género en la actualidad existen varias especies secuenciadas y en algunos casos se cuenta con el genoma de más de una cepa (e.g *C. fiorinae* y *C. gloeosporioides*) (**Tabla 1**).

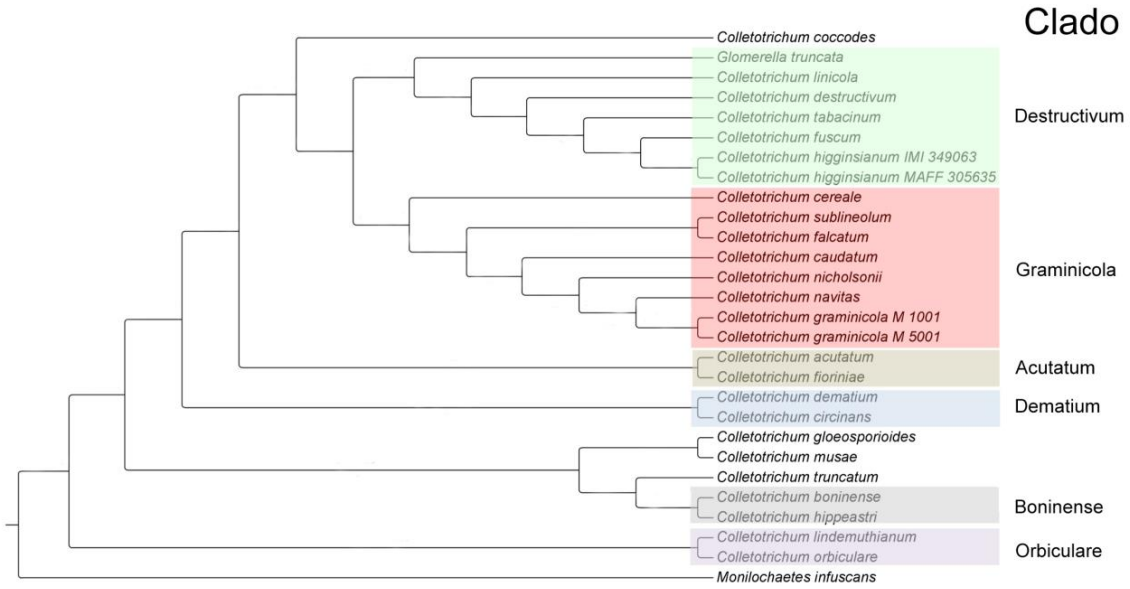


Figura 4. Árbol filogenético de varias especies del género *Colletotrichum* y su división en clados. En el árbol se presenta a la especie *Monilochaetes infuscans* como grupo externo. Modificado de O'Connell et al. (2012).

Tabla 1. Especies secuenciadas de *Colletotrichum spp.* cuyo genoma es público

Especie	Cepa	Referencia	Fuente de descarga
C. graminicola	M1001	(O'Connell et al. 2012)	http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html
C. higginsianum	IMI 349063	(O'Connell et al. 2012)	http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html
C. orbiculare	MAFF 240422	(Gan et al. 2013)	http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AMCV01
C. gloeosporioides	Nara gc5	(Gan et al. 2013)	http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=ANPB01
C. gloeosporioides	Cg-14	(Alkan et al. 2013)	http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AMYD01
C. gloeosporioides	23	Sin publicación	http://genome.jgi.doe.gov/Gloci1/Gloci1.home.html [a]
C. fiorinae	PJ7	(Baroncelli, et al. 2014a)	http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=JARH01
C. fiorinae	MH 18	Sin publicación	http://genome.jgi.doe.gov/Gloac1/Gloac1.home.html [a]
C. sublineola	TX430BB	(Baroncelli, et al. 2014a)	http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=JMSE01

[a] Disponible solo bajo las condiciones del Joint Genome Institute (http://genome.jgi.doe.gov/Gloac1/Gloac1.download.html)

El interés por este género está bien justificado, *C. graminícola* un patógeno exclusivo de maíz se ha convertido en una plaga importante en los principales países productores (Warren 1975; Bergstrom y Nicholson 1999; Palaversic et al. 2009; Costa et al. 2010), causando pérdidas anuales de hasta mil millones de dólares sólo en Estados Unidos (Frey et al. 2011). En España, la antracnosis no ha sido un problema tradicional del cultivo del maíz, pero ha sido detectada recientemente en la Comunidad Autónoma Cataluña (Pioneer Hi-Bred, *comunicación personal*).

El maíz es un cultivo de vital importancia en la alimentación de gran parte del mundo y también es utilizado en la producción de energía y como alimento de ganado. Se estima una superficie de 177 millones de hectáreas cultivadas alrededor del mundo (FAOSTAT 2014) y en España se registraron 368 mil hectáreas cultivadas en 2011 (Ministerio de Agricultura, Alimentación y Medio Ambiente de España 2014). La producción de maíz ocupa en torno al 30% de la superficie agrícola aprovechable de Castilla y León, lo que la convierte en la mayor productora de maíz de España (Subdirección General de Análisis, Prospectiva y Coordinación, 2012). El estudio de este organismo como el de muchas otras especies fitopatógenas del género *Colletotrichum* se vuelve fundamental para diseñar nuevas

estrategias de control fitosanitario y de esta forma intentar garantizar la seguridad alimentaria de las poblaciones.

Efectores en hongos fitopatógenos

Al igual que otros patógenos de plantas, los hongos secretan una variedad de proteínas que modulan la inmunidad innata de las plantas y permiten al patógeno infectar y colonizar a su huésped (Kamoun 2007). A estas moléculas se las denomina efectores, las cuales son proteínas secretadas por los patógenos que poseen la capacidad de manipular a las células de su huésped en forma, función o metabolismo (Giraldo y Valent 2013). Estas moléculas son fundamentales para el éxito de la infección del patógeno. Por tal razón, en los últimos años se ha venido trabajando de manera intensa en el tema al punto de convertirse en uno de los campos más estudiados en la fitopatología molecular. Las plantas poseen sistemas de reconocimiento de patógenos denominados por sus siglas en inglés PRRs (*pattern recognition receptors*), los mismos que reconocen moléculas conservadas en los patógenos que se conocen como PAMPs (*pathogen-associated molecular patterns*) y MAMPs (*microorganism-associated molecular patterns*) (Zipfel y Felix 2005; Dodds y Rathjen 2010; Beck et al. 2012; Liu et al. 2013). En el caso de los hongos los PAMPs mejor conocidos son las moléculas de quitina que conforman la pared celular del patógeno (Kaku et al. 2006; Miya et al. 2007). En cuanto los PRRs reconocen a los PAMPs se desencadena una respuesta conocida como PTI (*PAMP-triggered immunity*) (Zipfel 2008). Entre las respuestas producidas en la planta tras la acción de la PTI tenemos la producción de enzimas hidrolíticas (como proteasas, quitinasas o glucanasas), inhibidores de enzimas líticas del patógeno y especies reactivas de oxígeno (Giraldo y Valent 2013). Es en este punto en el que los efectores actúan de manera efectiva deteniendo los procesos de PTI o paleando los efectos desencadenados por esta. La capacidad de los efectores de suprimir o evadir los efectos de la PTI da paso a una susceptibilidad desencadenada por los efectores o ETS (*effector-triggered susceptibility*) en la planta. Por último, las plantas cuentan con un sistema de detección de efectores (por ejemplo proteínas de tipo NB-LRR) que desencadena la denominada ETI (*effector-triggered immunity*) que da como resultado un respuesta similar a la PTI que normalmente termina con la muerte de la célula (reacción de hipersensibilidad, HR) para frenar el avance del patógeno dentro de los tejidos (Jones y Dangl 2006).

Los efectores pueden ser secretados tanto al apoplasto como al citoplasma de las células del huésped. Entre los efectores apoplásticos mejor conocidos hasta el momento tenemos a los inhibidores de proteasas de la planta, los efectores de unión a quitina y a los inhibidores de peroxidasas (Thomma et al. 2011; Doehlemann y Hemetsberger 2013). Los inhibidores de proteasas y peroxidasas están destinados a neutralizar las defensas de la planta desencadenadas por la PTI mientras que los efectores de unión a quitina intentan evitar el reconocimiento del patógeno atrapando fragmentos de quitina desprendidos por la degradación de la pared celular del hongo o recubriendo la pared celular de este (Giraldo y Valent 2013).

La función de los efectores citoplasmáticos es muy diversa y estos pueden localizarse en distintas partes de la célula vegetal incluyendo el núcleo (Oliva et al. 2010). Dentro de los efectores citoplasmáticos tenemos a AvrPiz-t de *M. oryzae*. Este efector se une a la proteína APIP6, una E3 ubiquitin ligasa de arroz y así AvrPiz-t suprime la PTI desencadenada por la detección de fragmentos de quitina (Park et al. 2012). Este es un ejemplo de cómo pueden interferir los efectores en el sistema de ubiquitinación de la planta para inhibir los procesos de defensa. *Ustilago maydis* posee un efector denominado Cum1 el cual interviene en la ruta del shikimato de la planta. De esta manera Cum1 desvía la ruta metabólica hacia la producción de compuestos aromáticos en lugar de ácido salicílico (fitohormona que aumenta la resistencia contra el hongo) (Djamei et al. 2011). Este es un ejemplo de cómo los efectores pueden alterar el funcionamiento de la planta mediante el desvío de rutas metabólicas.

Aunque no son abundantes en la literatura, también se han detectado efectores que llegan al núcleo de la célula vegetal. En el caso del efector Uf-RTP1p de *Uromyces fabae*, se ha observado que dicha proteína es secretada desde el haustorio del hongo hasta el citoplasma del huésped pero también se la ha podido detectar en el núcleo de la célula vegetal (Kemen et al. 2005). Homólogos de este efector se han encontrado en la especie *Uromyces striatus* y también en 28 especies del orden Pucciniales aunque no se ha detectado señal de localización nuclear en estos genes. Los primeros análisis de la función bioquímica de Uf-RTP1p indican que es capaz de inhibir enzimas proteolíticas y posiblemente actúe en cooperación con otros efectores durante la fase biotrófica del hongo (Pretsch et al. 2013).

Existen evidencias contrastadas de que especies del género *Colletotrichum* spp. producen efectores para anular las respuestas del huésped. De hecho, una de las primeras moléculas efectoras que se identificaron en un hongo filamentoso se describía en *Colletotrichum*

gloeosporioides (Stephenson et al. 2000). Recientemente, nuestro grupo ha identificado varios genes en *C. graminicola* que codifican efectores putativos y se estudia el papel que juegan estos en el desarrollo de la enfermedad en la planta (Sanz-Martín et al. 2012; Vargas et al. 2012b). CgEP1 es un efector nuclear encontrado en *Colletotrichum graminícola*. Lo característico de este efector es la presencia de dos regiones bien definidas, una región N terminal ácida, con presencia de señal de secreción y que se encuentra conservada en otras especies del género *Colletotrichum* y una región C terminal básica, única de la especie. Esta proteína es sintetizada durante las primeras etapas de desarrollo de la enfermedad y es necesaria para que la hifa del hongo pueda crecer a través de la epidermis del huesped. Estudios genéticos, moleculares y bioquímicos confirmaron que este nuevo efector tiene como objetivo el núcleo del huésped, y define una nueva clase de proteínas de unión a ADN con un efecto directo en la expresión génica del huésped durante la patogénesis (Vargas et al, pendiente de publicación).

Dentro del género *Colletotrichum* existen varios casos de efectores caracterizados. Así por ejemplo CgDN3 de *Colletotrichum gloesoporioides*, que es una proteína secretada cuyo mutante nulo es capaz de desencadenar respuesta de hipersensibilidad en su huésped. Esto indica que este efector es capaz de inhibir la respuesta activada por efectores (ETI, por sus siglas en inglés) (Stephenson et al. 2000). Otro caso es el de GLRG_06543, una metaloproteasa, fungalisina, secretada por *C. graminícola* cuyo mutante nulo resulta menos virulento que la cepa silvestre (Sanz-Martín et al., pendiente de publicación). Actualmente varias publicaciones describen efectores putativos en análisis genómicos (Bhadauria et al. 2011; Kleemann et al. 2012; O'Connell et al. 2012) que resaltan la importancia de este tipo de proteínas en las interacciones planta-patógeno. Con toda esta información se podrá develar en un futuro algunos de los secretos que hacen de este género de hongos fitopatógenos uno de los más exitosos de nuestra época.

Transferencia horizontal de genes

La transferencia horizontal de genes (THG) es el intercambio de material genético entre especies emparentadas o no emparentadas sin que se produzca apareamiento sexual entre los individuos (Rosewich y Kistler 2000). Este fenómeno también llamado transferencia lateral de genes se ha convertido en los últimos años en un apasionante campo de estudio con

consecuencias importantes en nuestra forma de concebir la genética y la evolución de las especies.

En procariotas la THG es un fenómeno bien conocido, al punto de considerarse hoy en día una fuerza evolutiva importante para estos organismos (Koonin y Wolf 2008; Wolf et al. 2012; Borziak et al. 2013; Kitahara y Miyazaki 2013). En eucariotas este fenómeno se ha observado con menor regularidad pero en los últimos años los estudios sobre este tema se han incrementado y se ha observado THG en todos los reinos (Richards et al. 2006; Schaack et al. 2010; Nikolaidis et al. 2014).

Si bien los mecanismos por los que se puede dar la THG en procariotas son bien conocidos (conjugación, transducción, transformación) (Gyles y Boerlin 2014), en eucariotas dicho o dichos mecanismos no están del todo claros. Varias de las características de las células eucariotas servirían de barrera para el paso estable de material exógeno hacia los genomas huéspedes. Algunas de estas barreras son: la envoltura del material genético en un núcleo, sistemas de ARN de interferencia, especificidad en el reconocimiento de los promotores, incompatibilidad en los sistemas de corte y empalme del ARN, separación de las líneas celulares reproductivas, entre otros (Irelan and Selker 1996; Richards et al. 2003; Keeling and Palmer 2008; Fitzpatrick 2011). A pesar de la dificultad aparente para que se produzcan estos eventos de transferencia existen abundantes ejemplos de que este fenómeno ha ocurrido en especies eucariotas. Los plásmidos son uno de los vectores propuestos para explicar la THG (Rosewich y Kistler 2000). El caso del paso del ADN-T desde *Agrobacterium tumefaciens* hacia las plantas que infecta es mediada por un plásmido (Chilton et al. 1977), lo cual apoya esta hipótesis. Los transposones son otros de los vectores propuestos para la THG. Estos elementos tienen el potencial de transportar material genético entre especies no relacionadas además de haber sido observados en varios eventos de THG (Kuraku et al. 2012; Broaders et al. 2013; Yuan et al. 2013). La fagositosis también ha sido propuesta como mecanismo de THG. La hipótesis denominada “eres lo que comes” propone que la relación depredador-presa aumenta las probabilidades de que ocurra un evento de THG (Doolittle 1998). De manera similar se ha visto que la asociación de hongos saprófitos con material vegetal en descomposición podría aumentar las probabilidades de adquirir genes de manera lateral en este tipo de hongos (Friesen et al. 2006; Richards et al. 2009; Slot y Rokas 2010).

La evidencia de la THG solo es visible a través de la información dejada en los genes o los genomas. Así, los métodos de detección de THG existentes en la actualidad están basados en

la reconstrucción filogenética o en aproximaciones paramétricas (comparación de parámetros del gen transferido con su genoma huésped) (Azad y Lawrence 2012). Cada una de estas técnicas tiene sus virtudes y debilidades y el combinar ambas aproximaciones es siempre recomendable de ser posible.

Existen diversos métodos paramétricos, la mayoría de ellos desarrollados para detectar casos de THG entre especies. Dentro de estos tenemos: la detección de sesgo en el uso de codones y de nucleótidos al comparar el gen candidato de transferencia horizontal con su genoma huésped, clasificadores Bayesianos, índices de transferencia horizontal, curvas de Z, análisis de correspondencia, agrupamiento por el criterio de información de Akaike, entre otros (Azad y Lawrence 2012). La desventaja en el uso de métodos paramétricos es que solo son útiles en casos de THG recientes. Esto se debe a que los métodos paramétricos están basados en la diferencia entre el gen transferido y el genoma que lo acoge. Para que el gen exógeno "sobreviva" en el genoma huésped necesita adaptarse a sus características (codones preferentes, composición de GC, número de intrones, entre otros) y esto hace que con el tiempo las diferencias entre el gen exógeno y su genoma huésped desaparezcan y sean prácticamente irreconocibles (Lawrence y Ochman 1997).

Por su parte, la aproximación filogenética utilizada en la detección de eventos de THG busca descubrir patrones en los árboles de genes (o proteínas) que contradigan radicalmente el árbol de especies aceptado. La detección de THG por este método sufre de las mismas limitaciones que existen en la reconstrucción de cualquier árbol filogenético. Además de las imprecisiones propias de los métodos de filogenia molecular (errores de alineamiento, errores de muestreo, uso incorrecto de los modelos evolutivos existentes, entre otros), diferentes escenarios de transferencia vertical pueden ser confundidos con THG. Por ejemplo, si se compara un árbol de genes con un árbol de especies incorrecto se puede confundir con THG, de igual forma si el estado ancestral de un gen se ha mantenido en un linaje particular pero ha desaparecido en otras especies relacionadas. Otros fenómenos que pueden ser confundidos con THG en una reconstrucción filogenética pueden ser producidos por una tasa de cambio (mutación) desigual en genes de diferentes especies, hibridación interespecífica o comparación de parálogos de diferentes especies y no de ortólogos (**Figura 5)** (Rosewich y Kistler 2000).

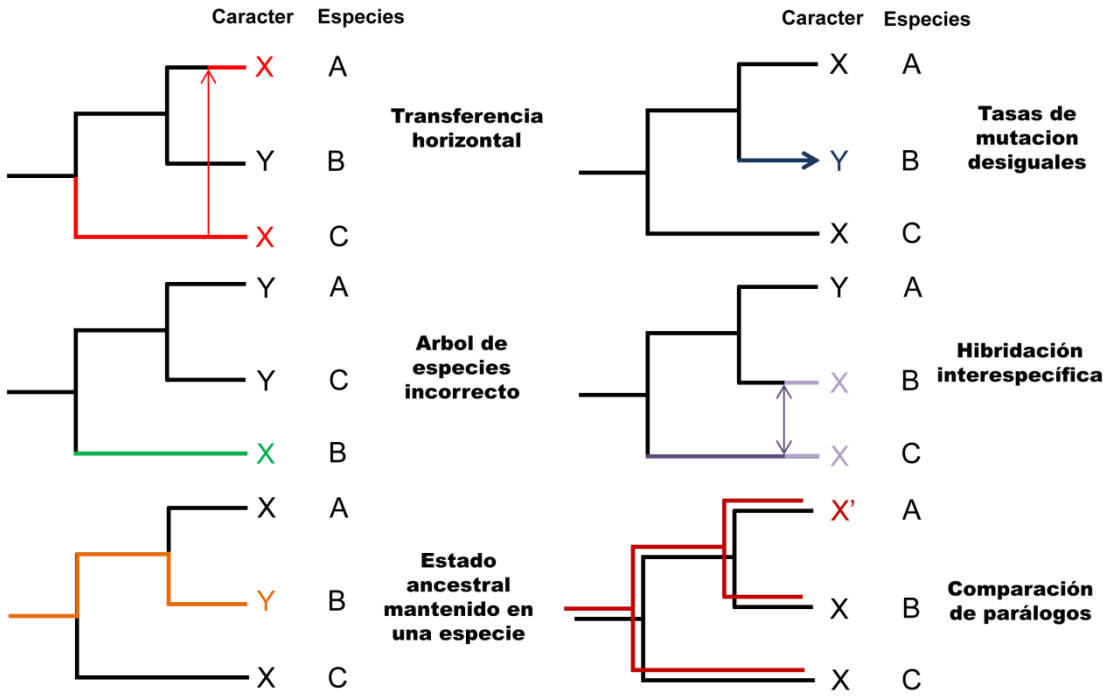


Figura 5. Explicaciones alternativas a la transferencia horizontal de genes en árboles filogenéticos con distribuciones diferentes al árbol de especies. Modificado de Rosewich y Kistler (2000).

A pesar de las limitaciones de los métodos de detección y la dificultad para encontrar el mecanismo exacto de la THG en eucariotas, la bibliografía en este campo recoge cada vez más ejemplos de la presencia y la importancia de estos eventos. La detección de genes transferidos horizontalmente no solo resulta interesante como fenómeno evolutivo sino que ha servido para encontrar genes relacionados con la patogenicidad de ciertos organismos. Por ejemplo, se encontró un grupo de quitinasas en especies del género *Phytophthora* (un género de oomicetos fitopatógenos), cuyo origen parece provenir de bacterias. Estas quitinasas que se encontraban en duplicaciones de entre tres a diez copias en las especies de *Phytophthora* al parecer solo están presentes en oomicetos y tres géneros de actinobacterias. La expresión de estos genes es coincidente con la infección y la germinación de dichos organismos por lo que estos genes parecen ser de importancia en el proceso infectivo del patógeno (Belbahri et al. 2008).

Otro caso remarcable por su importancia y por lo reciente de la transferencia es el paso del factor de virulencia ToxA desde *Stagonospora nodorum* a *Pyrenophora tritici-repentis*. *Pyrenophora tritici-repentis* en la actualidad es conocido como un patógeno de trigo, pero no fue descrito como tal hasta 1941. El gen *ToxA* posee un bajo grado de polimorfismo en las

poblaciones de *P. tritici-repentis* mientras que la variabilidad en *S. nodorum* es alta. Además, la similitud de los genes *ToxA* de estas especies es muy alta (alrededor del 99.7%). Dada esta evidencia, los autores sugieren que *S. nodorum* transfirió horizontalmente el gen *ToxA* a *P. tritici-repentis* poco antes de 1941 y esto ayudó a que esta especie logre infectar a los cultivos de trigo (Friesen et al. 2006).

En los hongos patógenos de Cacao *Moniliophthora perniciosa* y *Moniliophthora roreri* existen tres genes potencialmente implicados en patogenicidad y que tienen su origen en bacterias y oomicetos. Estos genes anotados como una metalohidrolasa, manitol fosfatasa deshidrogenasa y una proteína inductora de necrosis, provendrían de actinobacterias, firmicutes y oomicetos respectivamente. Dado que se sospecha que el ancestro de *M. perniciosa* y *M. roreri* era saprófito se ha especulado que la transferencia de estos genes ha ayudado en la conversión hacia la fitopatogenicidad (Tiburcio et al. 2010).

Genes implicados en procesos metabólicos diversos han sido descritos como transferidos horizontalmente. En levaduras, el grupo de genes de la ruta de la galactosa (genes *GAL*) presenta evidencias de haber sido adquirido en distintas periodos de tiempo por THG. Así, en especies del género *Schizosaccharomyces* los genes *GAL* tendrían su origen en *Candida* y los de *Cryptococcus* serían originarios de una mezcla de genes de *Candida* y *Saccharomyces*. De estos eventos se deduce que en levaduras la THG les ha conferido plasticidad metabólica (Slot y Rokas 2010).

El gen de la beta glucuronidasa (*gus*) permite a los hongos utilizar glucorónidos provenientes de la orina de los vertebrados como fuente de carbón. Según el estudio de Wenzl et al. (2005) el gen de la enzima *gus* fue adquirido en los ascomicetes desde bacterias gram^{+}, lo cual permitió a estos hongos aprovechar una nueva fuente de carbono cuando los vertebrados colonizaron la tierra.

Ciertas racemasas encontradas en especies de los géneros *Fusarium* y *Aspergillus* presentan una alta similitud con genes que codifican estas mismas enzimas en bacterias. Las racemasas catalizan la interconversión de isómeros ópticos de aminoácidos. La función precisa en hongos no ha sido determinada (Marcet-Houben y Gabaldón 2010).

Según el estudio de Richards y colaboradores (2006) la osmotrofía compartida por hongos y oomycetos podría deberse a la THG entre estos linajes. Algunos genes encontrados en este estudio tendrían su origen en bacterias y habrían sido trasferidos a hongos para después pasar

a los oomycetos. Genes como *Araj* (un gen implicado en el transporte de azucares), *CodB* (una permeasa de purinas) o el *PcaH* (una dioxigensa) caerían dentro de este escenario. Estos tres genes tendrían relación con la osmotrofía y podrían explicar en parte la similitud entre estos dos linajes (oomicetos y hongos) tan alejados filogenéticamente pero que comparten similitudes tanto morfológicas como fisiológicas.

Todos estos ejemplos nos dejan ver que la transferencia horizontal de genes podría ser un fenómeno importante en la evolución de los eucariotas y particularmente de los hongos. Estos eventos conferirían nuevas cualidades a los organismos receptores tales como la patogenicidad y la plasticidad de las rutas metabólicas. Al mismo tiempo les permitiría colonizar nuevos nichos o adquirir nuevas herramientas para sobrevivir en sus hábitats naturales. Por tanto la investigación de la transferencia horizontal de genes es una pieza clave para la comprensión de la complejidad de los organismos superiores.

Mimetismo molecular

El término mimetismo molecular (MM) fue empleado originalmente para referirse a la similitud existente entre los antígenos de ciertos patógenos y algunas proteínas humanas (Raymond 1964). Los anticuerpos producidos para combatir estos antígenos en algunos casos reaccionan frente a las proteínas propias del cuerpo humano por el parecido con el antígeno. Esto puede producir enfermedades autoinmunes por la reacción cruzada producida por el antígeno y las proteínas similares en el huésped (Oldstone 1998). Pronto se observó que este fenómeno no estaba restringido a los parásitos animales y que podría ser un mecanismo generalizado que los patógenos usan para infectar a sus huéspedes. En la actualidad se han encontrado todo tipo de organismos que parasitan plantas y animales que usan el mimetismo molecular en sus procesos infecciosos (Alto et al. 2006; Greenbaum et al. 2008; Elde y Malik 2009; Wang et al. 2010).

Las definiciones de mimetismo molecular han ido cambiando con los años para incluir los casos que se han ido descubriendo. Uno de los primeros en conceptualizar este fenómeno fue Oldstone (1989) quien definía el MM como la similitud de estructuras en moléculas que son producidas por genes diferentes. Hall (1994) por su parte propuso diferenciar el MM como adaptativo y de consecuencia. El MM adaptativo es aquel que se produce por las presiones selectivas a las que están sometidas las moléculas del parásito dentro del huésped. Este autor

propone que este tipo de MM tiene por objetivo imitar moléculas de su huésped por razones biológicas tales como la evasión inmunológica o el trastornar los sistemas bioquímicos o fisiológicos del huésped. Por otro lado el MM de consecuencia no implica pero tampoco excluye la adaptación y se refiere principalmente a la semejanza que pueden tener las proteínas del huésped con las del patógeno por la similitud filogenética que presentan (por ejemplo en familias de genes conservadas en especies no relacionadas) o por la similitud en pequeñas partes de dos moléculas no relacionadas. Posteriormente Oldstone (1998) define el MM como estructuras similares que son compartidas por proteínas provenientes de genes distintos. Y Elde y Malik (2009) definen el MM como los factores codificados por los patógenos que se asemejan a los factores del huésped para neutralizar o impedir las funciones del huésped produciendo una ventaja para el patógeno. Estos autores también introducen los conceptos de mimetismo perfecto e imperfecto. En el MM perfecto, una molécula del patógeno imita a una molécula del huésped mientras que en el MM imperfecto una molécula del patógeno puede imitar a más de una molécula del huésped. Como podemos ver todas estas definiciones involucran por lo menos dos especies y la mayoría de ellas toman en cuenta a proteínas o ácidos nucleicos como blancos del mimetismo. Como veremos a continuación esto no siempre es cierto.

Entre los eventos evolutivos que explican la presencia de MM tenemos a la transferencia horizontal de genes y a la evolución convergente (Drayman et al. 2013). En la transferencia horizontal se espera que el organismo patógeno haya adquirido un gen de su huésped para así utilizarlo en la consecución del MM. Este tipo de eventos se caracterizan por la alta similitud de secuencia que presenta la molécula imitadora con la molécula imitada (por ejemplo Olsen y Skriver (2003)). Caso contrario, si la proteína imitadora e imitada no presentan semejanzas a nivel de secuencia puede que las presenten a nivel estructural y/o funcional. De ser este el caso el MM puede ser provocado por evolución convergente, es decir por las presiones selectivas similares a las que están sometidos tanto el patógeno como el huésped o a su vez por presiones selectivas específicas sobre el patógeno para imitar al huésped (Sikora et al. 2005).

Varios tipos de moléculas han sido descritas como blanco de MM, tales como ácidos nucleicos, proteínas y fitohormonas. Algunos de los casos más interesantes de imitación de ácidos nucleicos están relacionados con los virus. Por ejemplo en los virus de la influenza humana A y B (un ARN virus) se ha detectado que los genomas virales presentan frecuencias

bajas de dinucleótidos GpC. Las frecuencias de GpC son características propias de cada genoma. Se sabe que el virus de la influenza humana B ha infectado a los humanos por más tiempo que la versión A. Por esta razón la versión B del virus presenta frecuencias de GpC más bajas (características del genoma humano) que la versión A. Esta última se piensa que migró desde las aves y que su frecuencia de GpC aún se está adaptando a las del genoma humano. Con esta estrategia el virus logra reducir el riesgo de desencadenar la respuesta del sistema inmune y evadir los sistemas de degradación viral (Greenbaum et al. 2008).

Además de los patógenos de animales las bacterias fitopatógenas también utilizan MM para atacar a sus huéspedes. *Pseudomona siringae* posee un efector llamado AvrPtoB el cual es transferido a la planta mediante el sistema de secreción tipo III. Esta proteína imita a las E3 ubiquitin ligasas de la célula vegetal y suprime la muerte celular programada asociada a la inmunidad de la planta. AvrPtoB fue capaz de inhibir la muerte celular programada en organismos a los que no infecta tales como levaduras (Abramovitch et al. 2003; Abramovitch et al. 2006; Janjusevic et al. 2006). En la **Figura 6** se puede observar el parecido entre las estructuras de AvrPtoB, *At*PUB14 una E3 ubiquitin ligasa de *A. thaliana* y Rbx1 (parte del complejo de E3 ubiquitin ligasa) de *H. sapiens*. La ubiquitinación es un sistema exclusivo de eucariotas por lo que asombra el parecido estructural con una proteína de procariota.

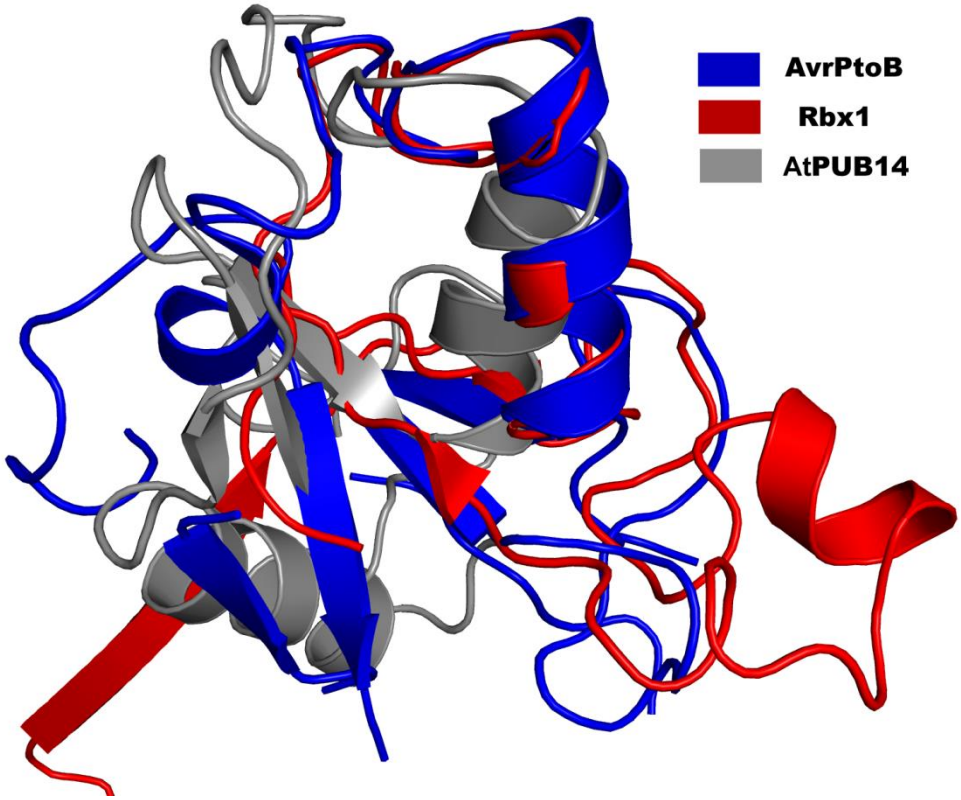


Figura 6. Alineamiento de estructura terciaria entre las proteínas que forman parte del complejo de las E3 ubiquitin ligasas *At*PUB14 (PDB ID 1T1H) y Rbx1 (PDB ID 1LDJ) en *A. thaliana* y *H. sapiens* respectivamente con la proteína AvrPtoB (PDB ID 2FD4) de *P. syringae*. Las bacterias carecen de sistema de ubiquitinación por lo que el parecido con las proteínas eucariotas es remarcable.

En plantas, la familia de proteínas CLE está involucrada en la diferenciación celular de meristemos apicales de raíz y tallo (Clark et al. 1995). De forma inesperada se encontraron motivos CLE (motivos característicos de las proteínas CLE) en el genoma del nematodo patógeno de plantas *Heterodera glycines* (proteínas con estos motivos se las denominó HgCLE) (Olsen y Skriver 2003). En base a estas observaciones se propuso que *H. glycines* utiliza las HgCLE para mantener indiferenciadas las células de las que se alimenta en la planta (células multinucleadas) gracias a su facultad de imitar las proteínas CLE del huésped. Experimentos posteriores revelaron que en efecto la continua secreción de proteínas HgCLE es necesaria para mantener los sitios de alimentación del nematodo (Wang et al. 2005) y que incluso las proteínas inmaduras HgCLE sufren de modificaciones post-traduccionales con la maquinaria celular del huésped. Además, se descubrió que este mecanismo era compartido por otras especies de nematodos que atacan a plantas (Guo et al. 2011). A partir de este caso se han encontrado varios candidatos de MM en nematodos parásitos de plantas. Por ejemplo

la proteína Hs4F01 del nematodo *H. glycines* que presenta un motivo de anexinas de planta. Aunque no se tiene clara la función de esta proteína podría ser usado para desregular el calcio de la membrana celular de la planta (las anexinas regulan el calcio de la membrana de las células) y de esta forma facilitar la penetración del nematodo (Patel et al. 2010). Además de Hs4F01 muchas otras proteínas en *H. glycines* han sido propuestas como candidatas de MM hacia plantas (Bellafiore et al. 2008). Si todos estos candidatos se confirman experimentalmente, podríamos afirmar que el MM es usado con frecuencia por este tipo de patógenos

Además de los ácidos nucleicos y las proteínas, los factores de crecimiento vegetales o fitohormonas también son imitados por los patógenos para lograr infectar a sus huéspedes. A este fenómeno se lo conoce como mimetismo de fitohormonas (Kazan y Lyons 2014). El caso más representativo de este tipo de mimetismo lo encontramos en *Pseudomona syringae*, la cual tiene la habilidad de imitar a la jasmonil isoleucina, una molécula capaz de activar la ruta del ácido jasmónico. La molécula imitadora se denomina coronatina y al parecer tendría el mismo efecto en la planta del que tiene la jasmonil isoleucina (Katsir et al. 2008). Entre los efectos que la jasmonil isoleucina tiene en la planta se cuenta la apertura de estomas y la represión de la respuesta mediada por ácido salicílico. Al parecer *P. syringae* aprovecha los estomas abiertos para ingresar en la planta (Hogenhout et al. 2009) y la supresión de la respuesta de ácido salicílico promueve la infección bacteriana (Brooks et al. 2005). Varios tipos de organismos fitopatógenos son capaces de producir reguladores de crecimiento vegetales. Aunque las moléculas sean exactamente las mismas las rutas por las que se producen en plantas y en patógenos son distintas. En este caso más que de moléculas imitadoras estamos hablando de copias exactas a las usadas por los huéspedes para regular su funcionamiento. En muchos casos se ha propuesto que mediante esta estrategia los fitopatógenos logran manipular el metabolismo vegetal para completar el proceso infectivo. Algunos de los fitopatógenos que logran producir fitohormonas están recogidos en la **Tabla 2**.

Tabla 2. Organismos fitopatógenos que son capaces de producir fitohormonas.

Especie	Tipo de fitohormona	Referencia
Pseudomonas syringae	Auxina	(Yamada 1993)
Agrobacterium tumefaciens	Auxina	(Yamada 1993)
Agrobacterium rhizogenes	Auxina	(Yamada 1993)
Ustilago maydis	Auxina	(Chung et al. 2003)
Colletotrichum acutatum	Auxina	(Chung et al. 2003)
Magnaporthe oryzae	Auxina	(Tanaka et al. 2011)
Pyrenopeziza brassicae	Citoquininas	(Murphy et al. 1997)
Cladosporium fulvum	Citoquininas	(Cooper y Ashby 1998)
Erysiphe graminis	Citoquininas	(Cooper y Ashby 1998)
Lepthospaheria maculans	Citoquininas	(Cooper y Ashby 1998)
Venturia inaequalis	Citoquininas	(Cooper y Ashby 1998)
Ustilago maydis	Citoquininas	(Mills y Van Staden 1978)
Gibberella fujikuroi	Giberelinas	Takahashi et al. (1955) citado por (MacMillan 2001)
***Phaeospheria* sp.**	Giberelinas	Sassa et al. (1989) citado por (MacMillan 2001)
***Sphaceloma* sp.**	Giberelinas	(Rademacher 1992)
Fusarium oxysporum	Ácido jasmónico	(Miersch et al. 1999)

Para finalizar cabe recalcar que no todos los eventos de MM se tienen que producir entre moléculas de diferentes especies como se ha venido señalando hasta ahora. Para poder explicar las complejas interacciones que se dan entre los efectores (proteína secretada por el patógeno que manipula las funciones de la célula huésped) de los patógenos y las proteínas de resistencia (proteínas R) de las plantas se han tenido que elaborar diferentes hipótesis para que se ajusten a las observaciones. Entre las más conocidas tenemos a las hipótesis ligando-receptor (Keen 1990) o la hipótesis del guardia (Jones y Dangl 2006). Sin embargo, a propósito del tema de este apartado nos compete hablar de la hipótesis del señuelo (van der Hoorn y Kamoun 2008). Esta hipótesis propone la existencia de eventos de MM en el que la molécula imitadora y la imitada corresponden al mismo organismo. Para entender este modelo es necesario identificar cuatro componentes clave en el sistema. Por un lado tenemos a los efectores (1), que son proteínas del patógeno que se unen a sus correspondientes proteínas blanco (2) en la planta para de esta forma alterar el funcionamiento de la célula

huésped. Por otro lado están las proteínas R (3) que reconocen a un efector unido a su proteína blanco. Las proteínas R desencadenan una cascada de señalización que termina en una respuesta de defensa por parte de la plantas. El último elemento en este sistema de acuerdo con esta hipótesis es el señuelo (4). Esta molécula imitaría a la proteína blanco e interaccionaría con el efector, llamando de esta forma a las proteínas R para que desencadenen la respuesta en la planta o en el caso de no existir proteína R específica competiría con las proteínas blanco por unirse al efector, con la diferencia de que el señuelo unido al efector no tiene consecuencias sobre el funcionamiento de la célula huésped. El señuelo se habría originado por la duplicación del gen que codifica la proteína blanco o por la evolución convergente de cualquier otro gen.

Entre los ejemplos que apoyan esta teoría tenemos a la proteína de tomate Pto (señuelo) que compite con FLS2 (proteína blanco, funciona como receptor de kinasas en tomate) por la unión al efector de *P. syringae* AvrPto. En este caso la proteína R se denomina Prf (Zhou y Chai 2008). Otro caso es el del efector denominado AvrBs3 de *Xanthomonas campestris* que se une y activa al promotor de *Upa20* (blanco), un regulador del tamaño celular en pimiento. AvrBs3 también tiene la capacidad de unirse al promotor del gen *Bs3* (señuelo) que a su vez codifica la proteína Bs3 (proteína R) que termina desencadenando la reacción de hipersensibilidad en la célula vegetal (Schornack et al. 2008).

Todos estos ejemplos nos permiten apreciar la importancia del MM en la fitopatología y en general en todos los sistemas patógeno-huésped. Es importante por tanto llegar a comprender este fenómeno ya que puede ser pieza clave en el desarrollo de nuevas estrategias en el control de plagas.

HIPÓTESIS Y OBJETIVOS

Hipótesis I

Los genomas de especies del género *Colletotrichum* contienen genes de transferencia horizontal. Estos genes debieron ser retenidos por estas especies por presiones selectivas y por tanto serían relevantes para los genomas huéspedes.

Objetivo I.A

Identificar eventos de transferencia horizontal de genes en los genomas de las especies de *Colletotrichum y* su posible impacto en los genomas huéspedes.

Objetivo I.B

Estimar la antigüedad de los eventos de transferencia así como el origen de los mismos.

Hipótesis II

El gen codificante de la proteína denominada CPLS (*Colletotrichum* plant-like subtilisin) fue transferido horizontalmente desde plantas hacia un ancestro del género *Colletotrichum* y representa un evento evolutivo singular de este género.

Objetivo II.A

Determinar si CPLS ha sido transferido horizontalmente, usando análisis filogenéticos, reconstrucción de la estructura proteica y otras fuentes de evidencia.

Hipótesis III

La especie *Colletotrichum graminicola* posee proteínas que imitan en forma y/o función a proteínas de su huésped *Zea mays*.

Objetivo III.A

Identificar proteínas de *Colletotrichum graminicola* que potencialmente imiten a proteínas de *Zea mays*.

Esquema temático

La presente tesis está dividida en tres capítulos, cada uno de los cuales está redactado en formato de artículo científico en idioma inglés. Las hipótesis y los objetivos fueron elaboradas a partir de los antecedentes recogidos en la introducción de este trabajo.

El primer capítulo trata sobre la búsqueda de eventos de transferencia horizontal de genes entre especies patogénicas del género *Colletotrichum* (***Objetivo I.A***). Para esto fue necesario desarrollar un método bioinformático que permita identificar dichos eventos en genomas completos. Para estudiar el impacto que los eventos de transferencia horizontal de genes han tenido sobre sus genomas huéspedes fue necesario estimar la antigüedad de cada evento así como la posible función de cada gen transferido (***Objetivos I.B***).

El segundo capítulo recoge el caso particular de la familia de genes denominada CPLSs (por sus siglas en inglés: *Colletotrichum* plant-like subtilisins). Para determinar el origen y la distribución de esta familia de genes fueron necesarios diversos análisis filogenéticos. Como modelo de las CPLSs se tomó a la proteína GLRG_05578 de *C. graminícola* para definir su estructura primaria, secundaria y terciaria (***Objetivo II.A***). Asimismo, se analizaron los perfiles de expresión del gen GLRG_05578 y el de algunos de sus homólogos en maíz. Varios resultados independientes resaltan la importancia de las CPLSs y apuntan a que son un claro ejemplo de transferencia horizontal de genes y un posible caso de mimetismo molecular.

El tercer capítulo está dedicado a la búsqueda de proteínas miméticas usando el patosistema *Colletotrichum graminicola*-maíz (***Objetivo III.A***). Para lograr los objetivos de este capítulo se desarrollaron dos metodologías alternativas y complementarias para detectar dichos eventos. La primera consiste en una modificación y ampliación de métodos existentes y la segunda un método original basado en la búsqueda de dominios conservados. Los candidatos de mimetismo molecular obtenidos fueron evaluados utilizando evidencia biológica (análisis de expresión en planta y en el hongo) e inferencias bioinformáticas para determinar la coherencia de los mismos con los casos descriptos en la bibliografía. Finalmente se presentan las conclusiones generales obtenidas de los tres capítulos esbozados aquí.

Nota: Los archivos adicionales (additional files) se encuentran incluidos en la verisón digital de esta tesis

CHAPTER I

Identification of horizontally transferred genes in the genus *Colletotrichum* reveals a steady tempo of bacterial to fungal gene transfer

Introduction

Horizontal gene transfer (HGT, also called lateral gene transfer) is the stable transmission of genetic material between organisms without the use of vertical inheritance mechanisms, mitosis or meiosis (Rosewich and Kistler 2000). HGT is common in Bacteria and Archaea and is considered an important force in their evolution (Koonin and Wolf 2008; Wolf et al. 2012; Borziak et al. 2013; Kitahara and Miyazaki 2013). In eukaryotes HGT is considered to be rare but an increasing number of studies are reporting HGT events in eukaryotes, and it is now beginning to be considered as an important mechanism of eukaryotic evolution (Schaack et al. 2010). In fungi, HGT events have been correlated with the gain of pathogenicity traits (Ma et al. 2010; Mehrabi et al. 2011; Milani et al. 2012; Sun et al. 2013) and with a gain of osmotrophic capacity (Richards et al. 2006; Richards et al. 2011). Genome-wide screens for HGT in fungi have also identified genes related to the metabolism of sugars, nitrogen, amino acids, nucleobases, and macromolecules as well as the acquisition of transporters and secreted proteins (Richards et al. 2011).

The genetic mechanisms that are responsible for HGT are not well known. The nuclear envelope, the storage of DNA in chromatin, RNA interference systems, separate reproductive cell lines, gene promotor specificity, incompatibility of intron splicing systems, alternatives gene codes and others represent barriers against HGT especially in distantly related species (e.g. interkingdom HGT) (Irelan and Selker 1996; Richards et al. 2003; Keeling and Palmer 2008; Fitzpatrick 2011). The mechanisms that make HGT possible across distantly related species are not well known. The transference of genetic material from the mitochondrion to the nuclear genome is one possible explanation of ancient HGT from prokaryotes to eukaryotes. Fungi have a large number of plasmids both inside and outside of fungal mitochondria (Griffiths 1995), and have been implicated as the vector for this type of transfer (Rosewich and Kistler 2000). Transposons are also candidate vectors of HGT. These elements have the potential to transfer genetic material among distantly related species but only in a few cases is there strong evidence to support this type of transfer (Rosewich and Kistler 2000). Another mechanism is phagocytosis, the "you are what you eat" hypothesis which proposes that the predator-prey relationship could increase the chances of an HGT event in microorganisms (Doolittle 1998). Beyond the vectors needed to achieve HGT, the ecological association of fungi with living and dead organisms may increase the chance of transmitting genetic material laterally (Friesen et al. 2006; Richards et al. 2009; Slot and Rokas 2010).

A wide range of methods have been proposed to detect HGT including phylogenetic analysis, and the detection of bias in nucleotide composition and codon usage, using naïve Bayes classifiers, correspondence analysis, or Akaike information criterion clustering (Azad and Lawrence 2012). Each method has its own strengths and weaknesses but in the case of ancient HGT events, phylogenetic approaches have more power to detect HGT and in general it is considered the most robust analysis method (Brown 2003).

The genus *Colletotrichum* contains more than 100 species, many of which are important phytopathogens (Cannon et al. 2012). In the present chapter, I surveyed the genome sequences of three members of the genus *Colletotrichum* to identify additional evidence of HGT to determine the impact of HGT on the evolution of pathogenicity in filamentous fungi. I discuss the potential role of the candidates in pathogenicity and niche adaptation. I found that genes typically transferred by HGT are those that belong to families that are subject to constant gene duplication and loss. I also determined the age of the HGT events, by means of a time calibrated phylogeny, and discuss the timing of HGT events within the context of major geological events. This is the first time that the impact of HGT has been evaluated on a genome-wide scale in *Colletotrichum*, and gives us insight into the evolution of this important genus.

Materials and Methods

Detection of candidates

To identify HGT candidates I created a BLAST database containing all of the complete proteomes deposited in UniProt (www.uniprot.org). Python scripts were used to extract the sequence, description and the two highest taxonomic levels of each sequence. The two highest taxonomic levels of UniProt correspond to kingdom and phylum in Archaea and Bacteria and to superkingdom and kingdom in Eukaryota. The two highest levels were selected to identify the taxonomic assignment at the level of kingdom of the sequences in the database.

With this database a BLASTP (Altschul et al. 1997) search (maximum e-value 10^{-5}) was performed with all putative proteins predicted in the genomes of *C. graminicola* M 1001, *C. higginsianum* IMI 349063 (O'Connell et al. 2012) and *C. gloeosporioides* (Alkan et al. 2013). HGT candidates were identified as those having 80% BLAST hits from a taxonomic classification other than Fungi. To choose this threshold, I developed a set of true positive

HGT genes by combining the HGT candidates from Richards et al. (2009), Schmitt and Lumbsch (2009) and Richards et al. (2011). Using the true positive set of proteins, I performed BLAST searches and then constructed phylogenetic trees with the BLAST hits and the query sequence. I evaluated the trees for topologies consistent with HGT of the query sequence. Not all the cases showed HGT patterns after the tree evaluation. 80% was the threshold that enabled us to identify the maximum number of the true positives HGT candidates.

The proteins with 80% or more BLAST hits from a taxonomic classification other than Fungi were taken for the next step. Sixty best BLAST hits were aligned using MAFFT (Katoh et al. 2002) and the alignments were edited with Gblocks (Talavera and Castresana 2007) to remove ambiguous regions from alignments. Next, the alignments were used to construct phylogenetic trees using PhyML (Guindon et al. 2010) using the default parameters. Each tree was evaluated to identify those with topologies consistent with HGT. The proteins identified in this step were used in a second BLASTP (maximum e-value 10^{-5}) search versus the nr database of NCBI (www.ncbi.nlm.nih.gov). The 20 best BLAST hits from each taxonomic domain (Archaea, Bacteria and Eukaryota) were extracted to perform a new cycle of alignment and tree reconstruction with the same procedures mentioned previously. These new trees were evaluated to detect topologies consistent with HGT.

Homologs to the HGT candidates were identified in the *C. graminicola*, *C. higginsianum* and *C. gloeosporioides* genomes using BLAST. When no evidence of homology in one or more species was found, I additionally searched the EST database of NCBI and the RNA-seq sequences of O'Connell et al. (2012) to find evidence of homologous sequences. To verify the presence of homologous sequences in *C. graminicola*, BLAST searches were performed in the genome of 5 sequenced strains of this species by Rech et al. (2014). Additionally, four new available *Colletotrichum* genomes were used to verify the absence/presence of each candidate in other members of the genus. This verification was performed through BLAST searches. The genomes used for this purpose were: *Colletotrichum gloeosporioides* 23 (teleomorph, *Glomerella cingulata*) (www.jgi.doe.gov), *Colletotrichum fiorinae* MH 18 (teleomorph, *Glomerella acutatum*) (www.jgi.doe.gov), *Colletotrichum gloeosporioides* Nara gc5 (Gan et al. 2013) and *Colletotrichum orbiculare* MAFF 240422 (Gan et al. 2013).

To detect possible false positives, several additional experiments were performed. First, a new BLASTP search was performed taking each protein of the HGT group as a query. Then, the

best BLAST results of each member of the group were compared and fused to get the homologous set of proteins for the group. The BLAST hits with high similarity and coverage were used for the next experiments. The cut-off of coverage and similarity was chosen individually for each group based on manual inspection of the BLAST results and a review of the level of conservation expected for each gene family.

To detect evidence of long branch attraction a different sets of BLAST results with different number of sequences were aligned with MAFFT and reconstructed with PhyML to detect important changes of tree topology when sequences are added or excluded from the analysis. TREE-PUZZLE (Schmidt et al. 2002) unresolved quartets calculation was used as indicator of the suitability of the data for phylogenetic analysis. Proteins with more than 10% of unresolved quartets were removed from the sequence set and the remaining proteins were aligned again. The final set of sequences was edited manually. Then, the alignment was used to reconstruct a tree. This tree was compared with the original tree obtained from the non-edited alignment to determine big changes in topology. MODELGENERATOR (Keane et al. 2006) was used to predict the best evolutionary model and the appropriate substitution matrix for each dataset. In cases when *Colletotrichum* HGT candidates had homologs in other fungal species, statistical topology tests were used to test alternative tree conformations. To generate the alternative topology (without HGT) MrBayes (Ronquist and Huelsenbeck 2003) was used to constrain all fungal proteins in one branch and the bacterial proteins in other branch. The last topology included in the tests was the star phylogeny. TREE-PUZZLE was used to perform the Expected Likelihood Weight (ELW) test and Shimodaira and Hasegawa (SH) tests. Finally, when a group of sequences was validated for all described procedures, the final set of homologous proteins was used to reconstruct a phylogenetic tree in the programs PhyML and MrBayes. In PhyML 100 non-parametric bootstrap were performed using the model selected by MODELGENERATOR. The parameters alpha, percentage of the invariant sites and amino acids frequency were calculated by PhyML. A maximum posterior tree was constructed with MrBayes, performing 2,000,000 generations of samples, using the substitution matrix and model predicted by MODELGENERATOR but allowing the program to calculate the proportion of invariable sites and the alpha parameter for gamma distribution. Two Multiple Chain Markov Chain Monte Carlo (MC^3) searches were conducted with four chains each (three heated and one cold). The convergence between them was checked using a sample frequency of 1000 generations. A burn-in of 25% of generations was excluded to reconstruct the Bayesian consensus tree. The topology differences between maximum

likelihood and Bayesian trees were quantified with the ratio of identical nodes performed in T-coffee (Notredame et al. 2000). The percentage of bootstrap and the posterior probability index were joined in the maximum likelihood tree using the program TreeGraph2 (Stöver and Müller 2010).

To determine if HGT candidates were the result of contamination a BLASTP (vs the nr database of NCBI) search of the upstream and downstream protein of each candidate in the genome of *C. graminicola*, *C. gloeosporioides* and *C. higginsianum* was performed. This search was done if the neighbor gene was located in the same contig of the candidate, otherwise the analysis was not performed. Additionally, to detect putative mitochondrial to nucleus gene transfers, a BLASTP search against all mitochondrion proteins of RefSeq database (Pruitt et al. 2012) was performed. Also all the candidates were submitted to CENSOR (Kohany et al. 2006) to detect the presence of repetitive or transposable elements. The entire pipeline is summarized in the **Figure 7**.

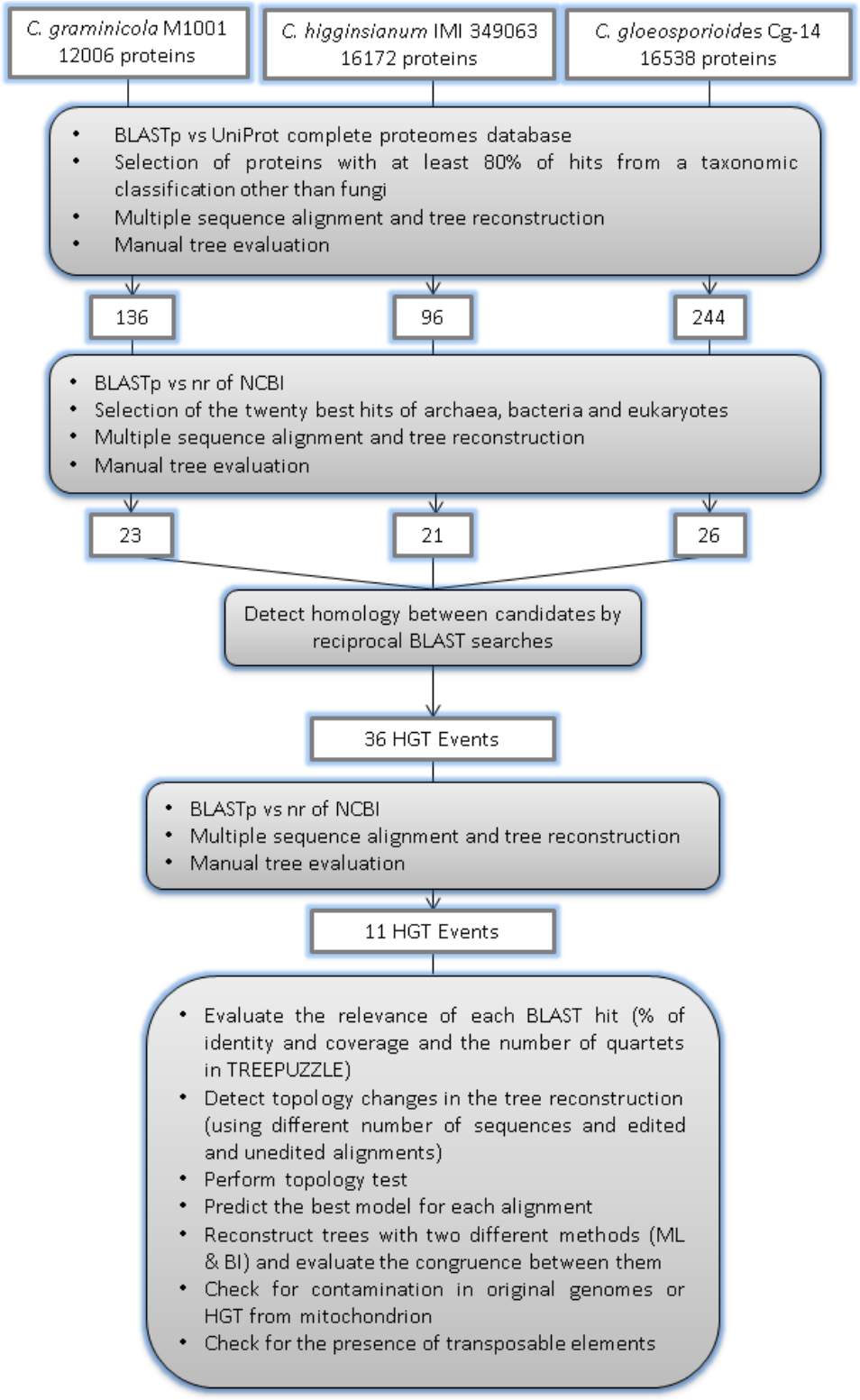


Figure 7. Diagram of the pipeline used in this work to detect HGT candidates in three species of the genus *Colletotrichum*.

Functional annotation of HGT candidates

The program Blast2GO (Conesa et al. 2005) was used to annotate the HGT candidates with Gene Ontology terms. The annotations were verified with InterProScan (Zdobnov and Apweiler 2001) and the putative biochemical functions of candidates were predicted with BRENDA (Schomburg et al. 2013), KEGG (Kanehisa et al. 2012) and MetaCyc (Caspi et al. 2010). Additionally, MEROPS (Rawlings et al. 2011) and CAZy (Lombard et al. 2013) databases were explored to further annotate the functions of proteases and carbohydrate active enzymes. To predict the cellular localization of candidates WOLF PSORT (Horton et al. 2007) and SignalP 4.1 server (Petersen et al. 2011) were used.

BLASTP searches of the HGT candidates were performed against PHI-base (Winnenburg et al. 2008), Virulence Factors Database (VFDV) (Chen et al. 2012) and Database of Fungal Virulence Factors (DFVF) (Lu et al. 2012) to detect proteins implied in pathogenicity.

The expression data were extracted from O'Connell et al. (2012) for *C. graminicola* and *C. higginsianum*.

GC content and intron content of the candidates

To determine the presence/absence of introns in HGT candidates a manual inspection of genes was made. To compare the difference of introns content among the candidates and *C. graminicola* a Mann-Whitney-Wilcoxon test was performed using R v2.13.1 (R Development Core Team 2011). The intron content of each *C. graminicola* gene was calculated from the information available on the Broad institute website (www.broadinstitute.org).

The program CodonW (Peden 1999) was used to calculate the GC content of the whole *Colletotrichum* genome and HGT candidates. These indices were used to calculate the differences of GC content among candidates and whole genomes.

Species tree reconstruction

To estimate the phylogenetic relationship of the species involved in HGT events, all fungal species found in BLAST searches of the HGT candidates were used (54 species in total). The complete proteme of these species was obtained from the UniProt (www.uniprot.org),

GenBank (www.ncbi.nlm.nih.gov/genbank/), Joint genome institute (www.genome.jgi.doe.gov/programs/fungi/index.jsf) and Broad Institute (www.broadinstitute.org/) databases. To reconstruct the species tree, the amino acids inferred from six nuclear genes were chosen. The proteins selected (FG533, FG570, FG832, MS277, MS413 and MS456) from FunyBase (Marthey et al. 2008) demonstrate to be good phylogenetic markers for fungi species trees reconstruction (Aguileta et al. 2008) and for that reason these were selected as a query to perform the BLAST searches in the proteomes of the 54 fungal species. Homology was verified making a multiple sequence alignment and tree reconstruction with MAFFT and PhyML respectively to make clear the orthology relationship between taxa. The topological congruence among the protein trees was checked before accepting the protein in the analysis. The six proteins were concatenated for each taxa. The concatenated proteins were aligned with three different programs, MAFFT, MUSCLE (Edgar 2004) and CLUSTALW (Larkin et al. 2007) to evaluate the differences in the phylogenetic reconstruction when different software is used. When the alignment was chosen TrimAl (Capella-Gutiérrez et al. 2009), GBLOCKS (Talavera and Castresana 2007) and Guidance (Penn et al. 2010) were used to edit it. A tree was reconstructed with each edited and unedited alignment with PhyML with 100 bootstrap repetitions. The best tree was selected by the alignment that produced the tree with the highest bootstrap values. PartitionFinder (Lanfear et al. 2012) was used to detect accurate models for the final alignment, using each protein as a partition. The models predicted by PartitionFinder were used in RaxML to calculate the maximum likelihood tree starting with 100 random trees. Finally, a non-parametric bootstrap analysis with 100 replications was performed and the results were summarized in the maximum likelihood tree.

Molecular clock analysis

To estimate the putative age of the transferred genes a fossil calibrated molecular clock analysis was performed. To avoid problems of convergence in the calculations of the calibrated tree, I selected 15 species from the 54 taxa used in the species tree. The same matrix of 6 concatenated proteins of the species tree was used. The analysis was performed with the BEAST v1.7.5 software package (Drummond et al. 2012). For each partition (each protein) the LG+I+G model was used. To allow uncorrelated rates of evolution across the tree I use a lognormal relaxed clock model, implementing a Yule process as a tree prior. I use a maximum likelihood tree estimated in PhyML under the LG+I+G model as starting tree. To

calibrate the tree I use the estimation of the *Paleopyrenomycites devonicus* fossil age of 400 million years ago (mya) as the lower bound for the Pezizomycotina crown (Lucking et al. 2009). A lognormal distribution with a mean of 460 (estimated from the results of Lucking et al. (2009)), standard deviation of 1 and offset of 400 was used as a prior for the time to the most recent common ancestor (TMRCA) of the Pezizomycotina. Two independent BEAST runs of 15 million generations each were performed. Data was sampled every 1500 generations. The convergence of two runs was visualized with TRACER v1.5 (Rambaut and Drummond 2007) and the Log files and tree files were combined with LogCombiner v1.7.5 (Drummond et al. 2012) dismissing a percentage of the sample in agree with TRACER plots of each run (23.3% for each one). With the remaining trees a maximum clade probability tree was calculated using TreeAnnotator v1.7.5 (Drummond et al. 2012). The resultant tree was visualized with the FigTree software (www.tree.bio.ed.ac.uk/software/figtree/).

Results

Identification of HGT candidates

The most robust method to detect HGT is phylogenetic analysis (Brown 2003). Therefore, to detect putative HGT events in three *Colletotrichum* species (*C. graminicola* M1001 , *C. higginsianum* IMI 349063 (O'Connell et al. 2012) and *C. gloeosporioides* Cg-14 (Alkan et al. 2013)), I developed a pipeline that consists of a series of BLAST searches and automated filtering designed to reduce the number of unlikely HGT candidates, followed by manual evaluation of phylogenetic trees (**Figure 7**). The first BLAST search was performed using a database of proteins from organisms with complete proteome available in UniProt (www.uniprot.org). Next, I selected proteins that having at least 80% of the top 120 hits (e-value e-5) with a taxonomic classification other than fungi as candidates for further analysis. The threshold of 80% was selected by evaluating previously described HGT candidates reported by Richards et al. (2009), Schmitt and Lumbsch (2009) and Richards et al. (2011) (see methods for details). Next, I subjected the HGT candidates to three phylogenetic analyses using different sets of homologous sequences in each phylogeny. The first phylogeny was constructed with homologous sequences from the UniProt complete proteome database (476 proteins from the 3 species were selected). For the second phylogeny I performed a BLAST search of the GenBank nr database and included the 20 best hits from each kingdom

(Archaea, Bacteria and Eukaryota) to avoid a possible under or overrepresentation for the abundance of sequences from any one kingdom in the BLAST results. This procedure was performed to observe events of inter-kingdom HGT and to observe the place into the tree of sequences with the same taxonomic label of the query but that were not the best hits in the BLAST search and therefore excluded from the first analysis that uses the best 120 hits. The third phylogeny was constructed with the best 100 BLAST hits from the nr database. The phylogenetic trees were evaluated manually, selecting only those that have well-supported topologies that are clearly incongruent with known species relationships among the taxa. Asymmetric or ladder-shape trees were excluded because such tree topologies are often a signal of long branch attraction or lack of phylogenic information (Moreira and Philippe 2000). Additionally, candidates with few homologues or with low sequence similarity to all of their BLAST hits were also excluded. In cases where only hits from two kingdoms were obtained (i.e. Bacteria and Eukaryota), high sequence similarity (a minimum of 30% pairwise similarity) and coverage (over 80% coverage) were required to consider them as candidates. The BLAST searches and tree evaluations were performed serially rather than in parallel to minimize the number of manual phylogenetic tree evaluations required (**Figure 7**).

During our analysis, I identified several proteins from *C. gloeosporioides* that appeared to be bacterial-fungal HGT events, had no homology to proteins in the other *Colletotrichum* spp nor to any other fungus. I examined the genomic contigs encoding these proteins and determined they are generally short contigs encoding only one gene. Using MEGABLAST searches of the nr database, I determined that these contigs are the result of bacterial contamination of the *C. gloeosporioides* genome assembly (data not shown). These HGT candidates were removed from further analysis. I performed another set of BLAST searches to identify homologs to the HGT candidates in the three *Colletotrichum* genomes and then arranged these additional proteins and the HGT candidates into groups of homologous sequences. One group of homologous sequences is comprised of two subtilisin-like serine proteases putatively transferred from a plant ancestor. For that importance this group will be described in the chapter II. The remaining groups appear to be of bacterial origin. The result was a list of 11 groups of homologous sequences representing 11 HGT events (**Table 3**). Homologs to HGT4 and HGT5 were described as HGT candidates by Sun et al. (2013) and Marcet-Houben and Gabaldón (2010) respectively.

Table 3. Summary of tests applied to the final set of HGT candidates

HGT Candidate	Number of homologous sequences used[a]	BLAST identity threshold	BLAST coverage cut-off	Topology changes with different number of sequences[b]	Topology changes after editing alignment	Best model	Topology tests (supporting HGT)[c]	Similarities between ML and BI tree reconstruction (%)
HGT1	44	40	90	No	no	LG+I+G+F	ELW, SH	83
HGT2	42	35	85	Yes	no	LG+G+F	n.a.	92
HGT3	54	41	96	Yes	no	LG+G	ELW	100
HGT4	43	40	90	No	no	LG+G+F	n.a.	93
HGT5	53	54	97	No	no	LG+G+F	n.a.	96
HGT6	46	44	90	Yes	no	WAG+G+F	n.a.	95
HGT7	37	35	85	No	no	LG+I+G+F	ELW, SH	84
HGT8	47	28	92	No	no	LG+I+G+F	ELW, SH	93
HGT9	50	37	90	No	no	LG+G+F	ELW, SH	81
HGT10	39	30	75	No	no	LG+G+F	ELW, SH	97
HGT11	44	30	80	No	no	LG+G+F	ELW, SH	89

[a] All sequences used have less than 10% of ambiguous quartets in TREE-PUZZLE

[b] Taking into account only the relative position of horizontal transfer group into the donor group. Topology changes of branches that do not affect the HGT candidates are not reported

[c] ELW= Expected Likelihood Weight, SH= Shimodaira and Hasegawa, n.a=not applicable. The groups with homologous sequence only in bacteria are not suitable for topology tests

The 11 groups of candidates were subjected to three tests to ensure that there is sufficient phylogenetic signal to support the HGT hypothesis. For each candidate, I manually inspected the results of the BLAST search to the nr database, selecting appropriate percent identity and coverage thresholds to eliminate distantly related and poorly aligning homologs. Each HGT candidate and its homologs were aligned and analyzed with TREE-PUZZLE (Hendy and Penny 1989) to calculate the percentage of unresolved quartets for each sequence. This measure quantifies the phylogenetic signal of each sequence in the phylogeny. Proteins with more than 10% unresolved quartets were discarded from each alignment. Next, I manually edited each multiple sequence alignment, removing regions of low sequence similarity where alignment errors are more likely. A tree was reconstructed for the edited and unedited alignments and the trees were compared to ensure that there was no change in topology due to alignment editing. In no case did I observe a change in topology as a result of alignment editing.

I evaluated the stability of the tree topology when different sets of proteins are used. For each candidate I reconstructed several trees by randomly removing several proteins and comparing the resulting tree topology with that of the original. HGT2, HGT6 and HGT3 had different topologies in this analysis but were stable in the rest of the tests to evaluate the phylogenies.

The data sets were used to perform topology tests to verify the clustering of sequences of different kingdoms and support the HGT hypothesis. Some of the candidates did not have high similarity with fungal sequences or sequences of any other kingdom more than bacteria. In that cases the topology tests could not be performed because the putative HGT sequences could not be constrained with the fungal branch in the tree to evaluate the vertical heritage scenario as alternative hypothesis. Most of the candidates have homology to bacterial and fungal proteins. The purpose of this test was evaluate if the candidates (from *Colletotrichum*) have a better explanation in the vertical gene transference scenario (candidates clustering with fungi) or with HGT (candidates clustering with bacteria). A star phylogeny was used as another possible scenario. The ELW (Expected Likelihood Weight) and SH (Shimodaira and Hasegawa) tests were used to test the hypothesis. For all the candidates with sequence homology in bacteria and fungi the HGT trees were chosen as a better explanation (with the highly log-likelihood score) for the data than the vertical heritage tree or the star phylogeny. Finally, for each alignment, a phylogenetic tree was constructed with PhyML (Guindon et al. 2010) and with MrBayes (Ronquist and Huelsenbeck 2003) and the similarity of the two trees was evaluated with T-Coffee (Notredame et al. 2000). These analyses are summarized in **Table 3** and the trees are shown in **Figures 8-18**. In each case, the HGT candidates form a monophyletic group within the bacterial lineage, indicating a single horizontal transfer event.

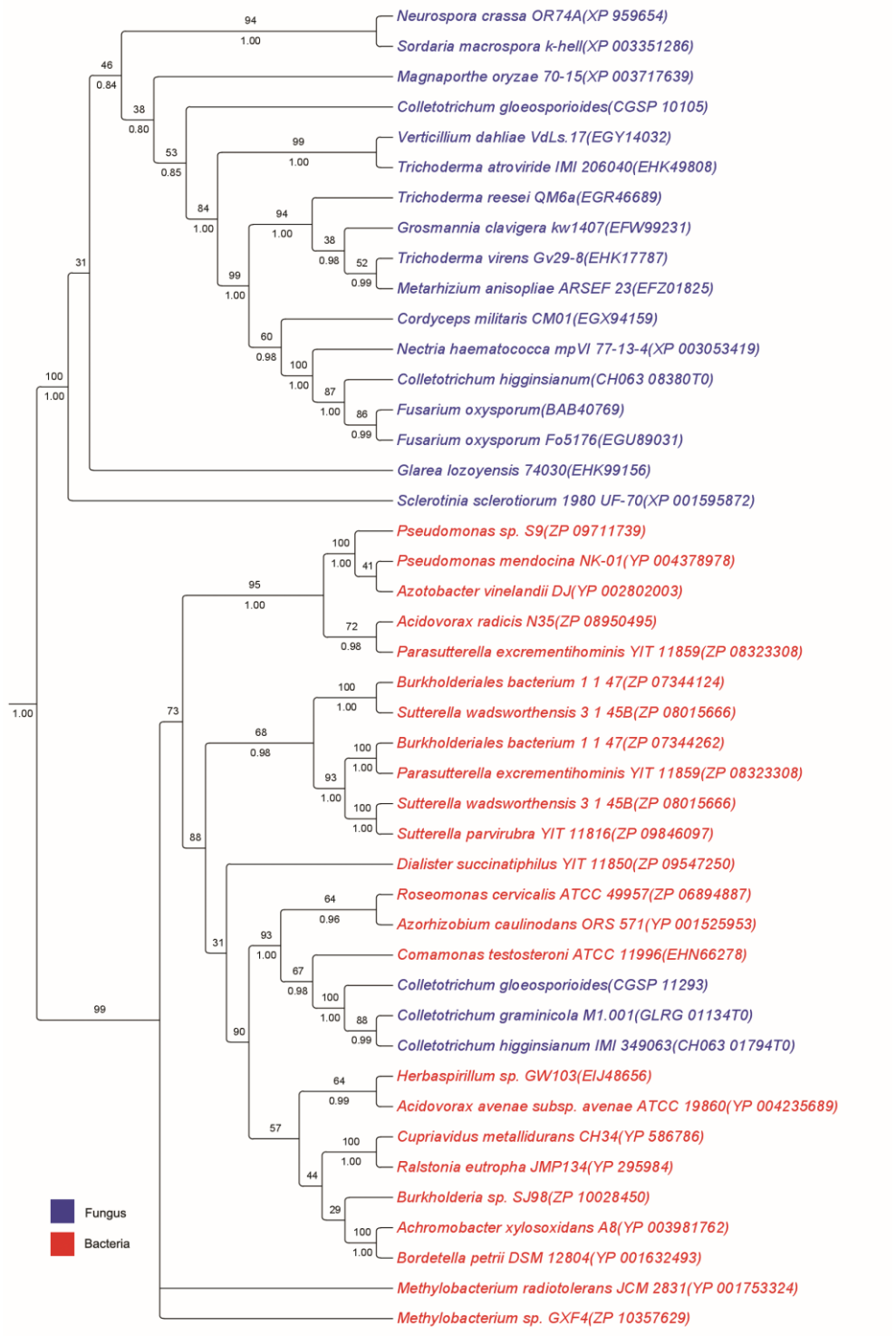


Figure 8. Maximum likelihood tree of HGT1 (GLRG_01134T0, CH063_01794T0 and CGLO_11293). The *Colletotrichum* sequences are not located within the fungal lineage, as is expected for vertically inherited genes. Instead, they are clustered within the bacterial lineage. Bootstrap percentages are shown above the branches and posterior probability is shown below the branches when the ML and BI tree topologies coincide. Accession numbers are shown in parenthesis next to each species name.

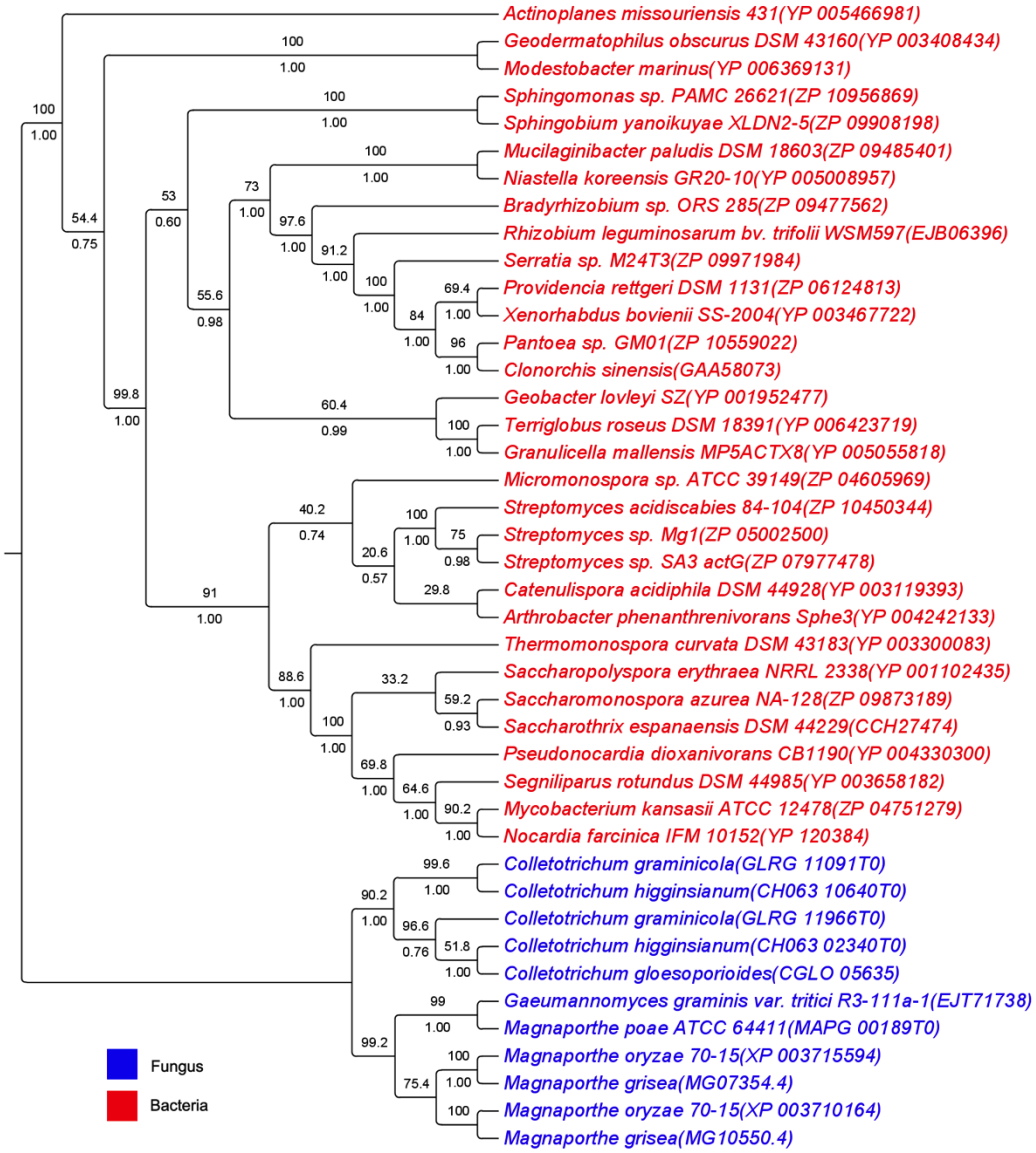


Figure 9. Maximum likelihood tree of HGT2 (GLRG_11091T0, GLRG_11966T0, CH063_02340T0, CH063_10640T0, CGLO_05635). Bootstrap percentage is shown above the branches and posterior probability is shown below the branches. Accession numbers are shown in parenthesis next to the species names.

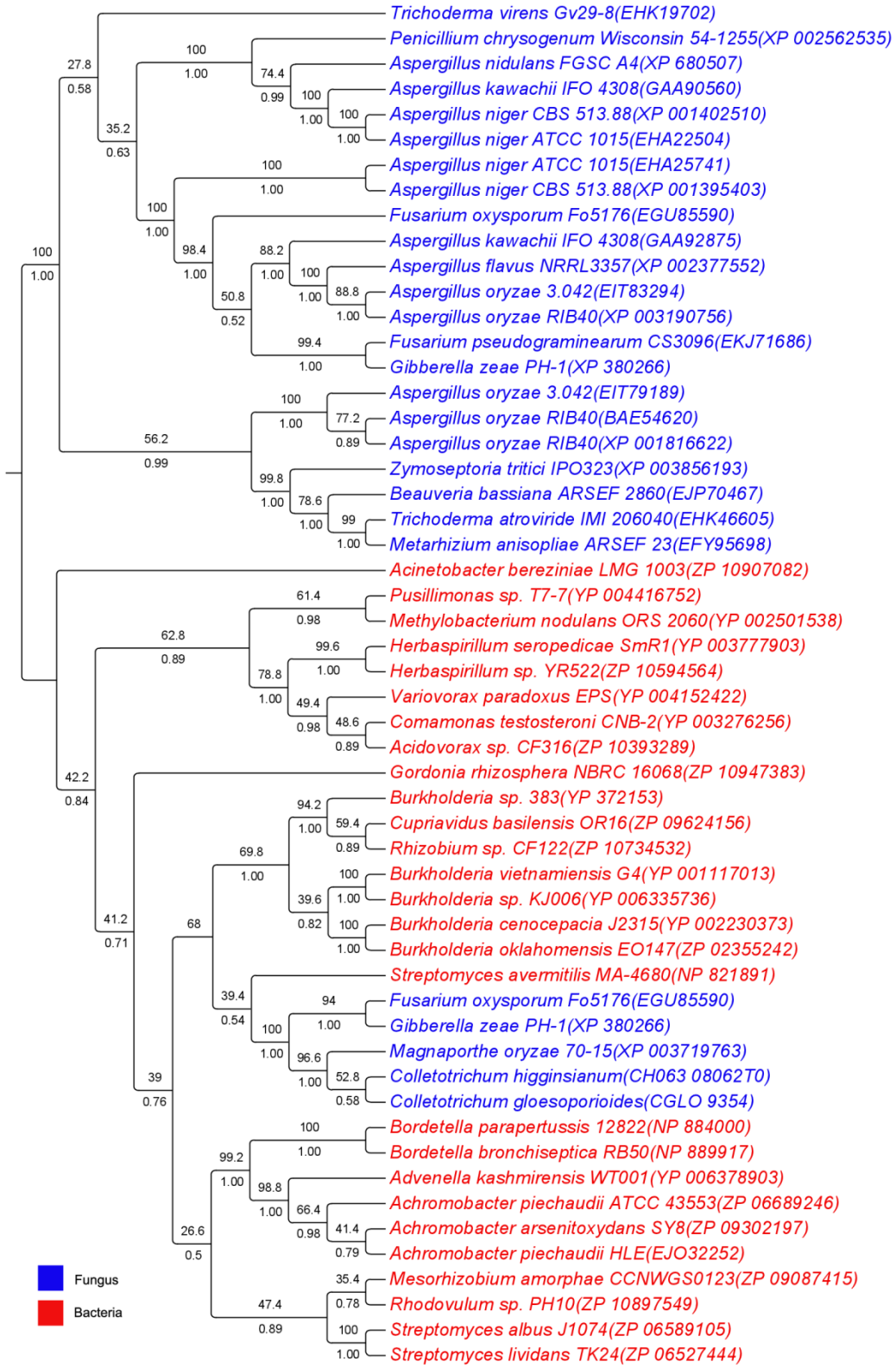


Figure 10. Maximum likelihood tree of HGT3 (CGSP_9354, CH063_08062T0). Bootstrap percentage is shown above the branches and posterior probability is shown below the branches. Accession numbers are shown in parenthesis next to the species names.

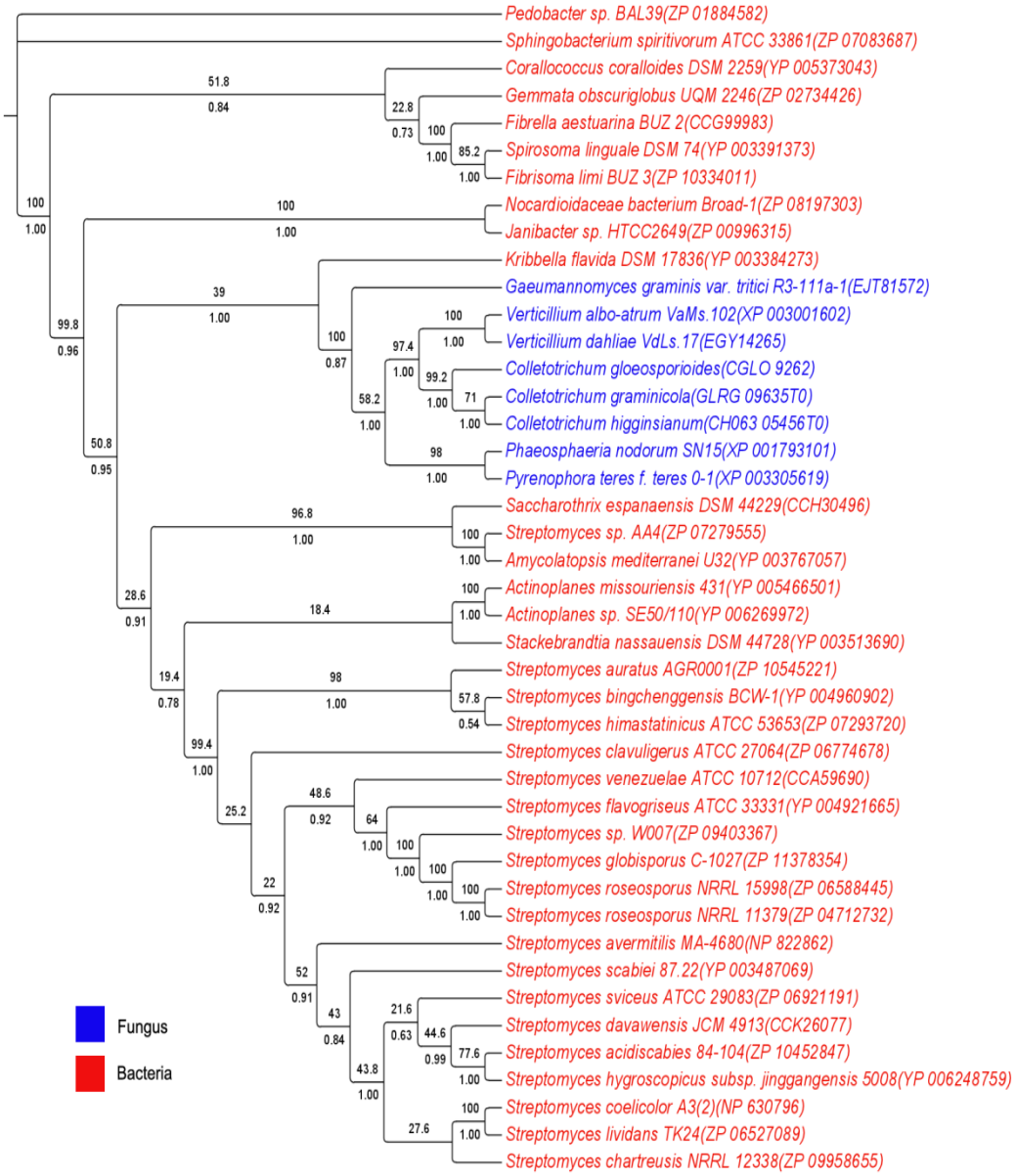


Figure 11. Maximum likelihood tree of HGT4 (CGSP_9262, CH063_05456T0, GLRG_09635T0). Bootstrap percentage is shown above the branches and posterior probability is shown below the branches. Accession numbers are shown in parenthesis next to the species names.

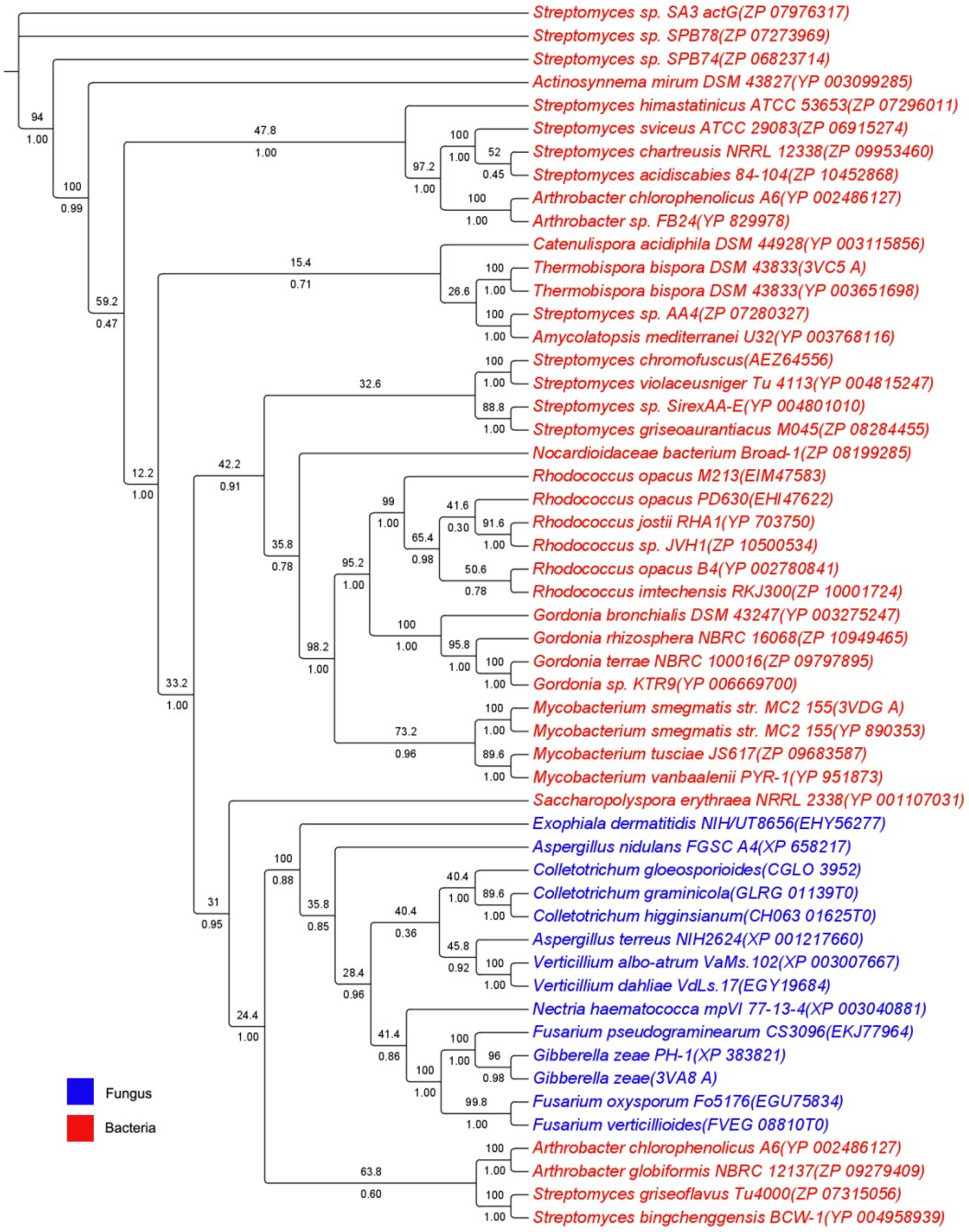


Figure 12. Maximum likelihood tree of HGT5 (CGSP_3952, CH063_01625T0, GLRG_01139T0). Bootstrap percentage is shown above the branches and posterior probability is shown below the branches. Accession numbers are shown in parenthesis next to the species names.

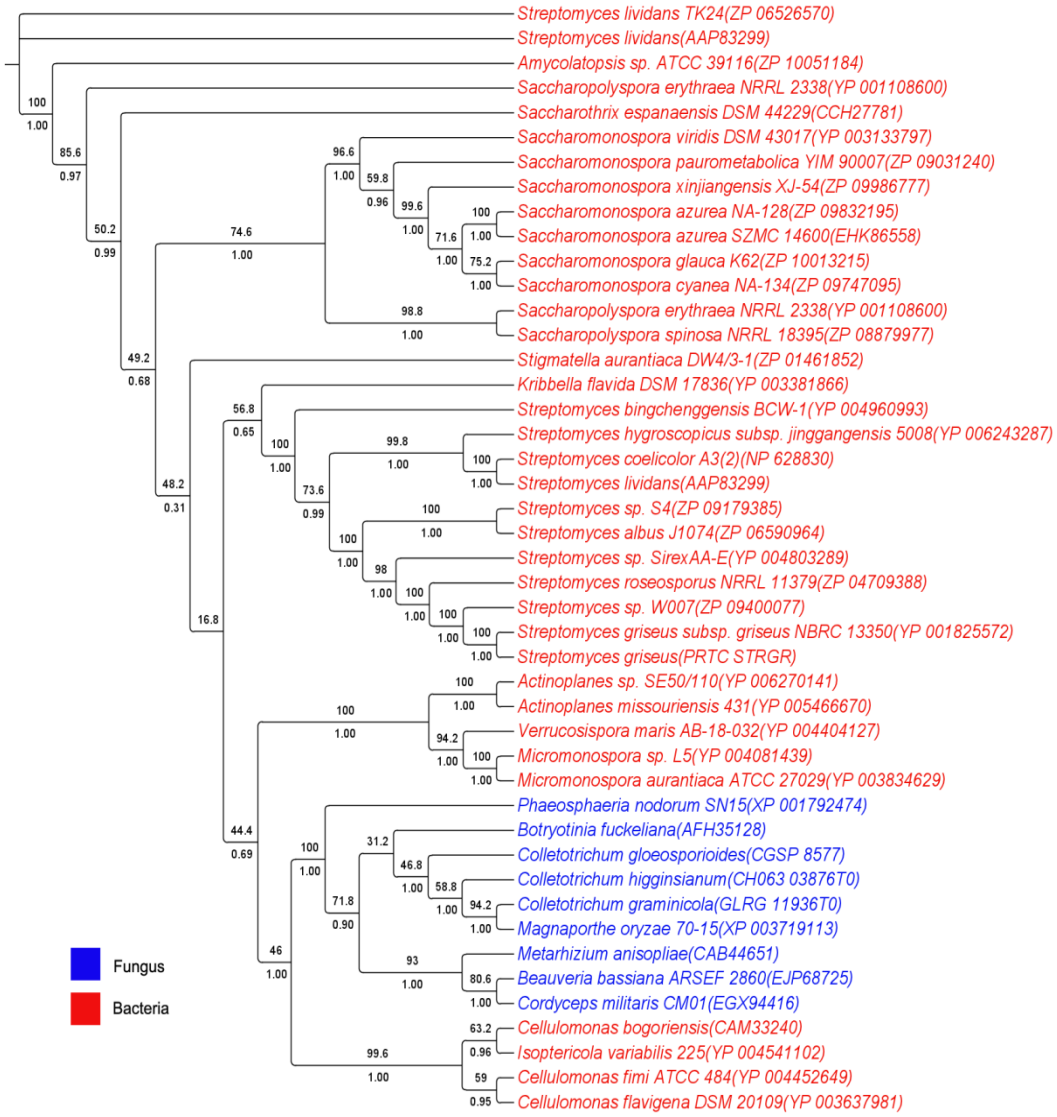


Figure 13. Maximum likelihood tree of HGT6 (CH063_03876T0, CGSP_8577, GLRG_11936T0). Bootstrap percentage is shown above the branches and posterior probability is shown below the branches. Accession numbers are shown in parenthesis next to the species names.

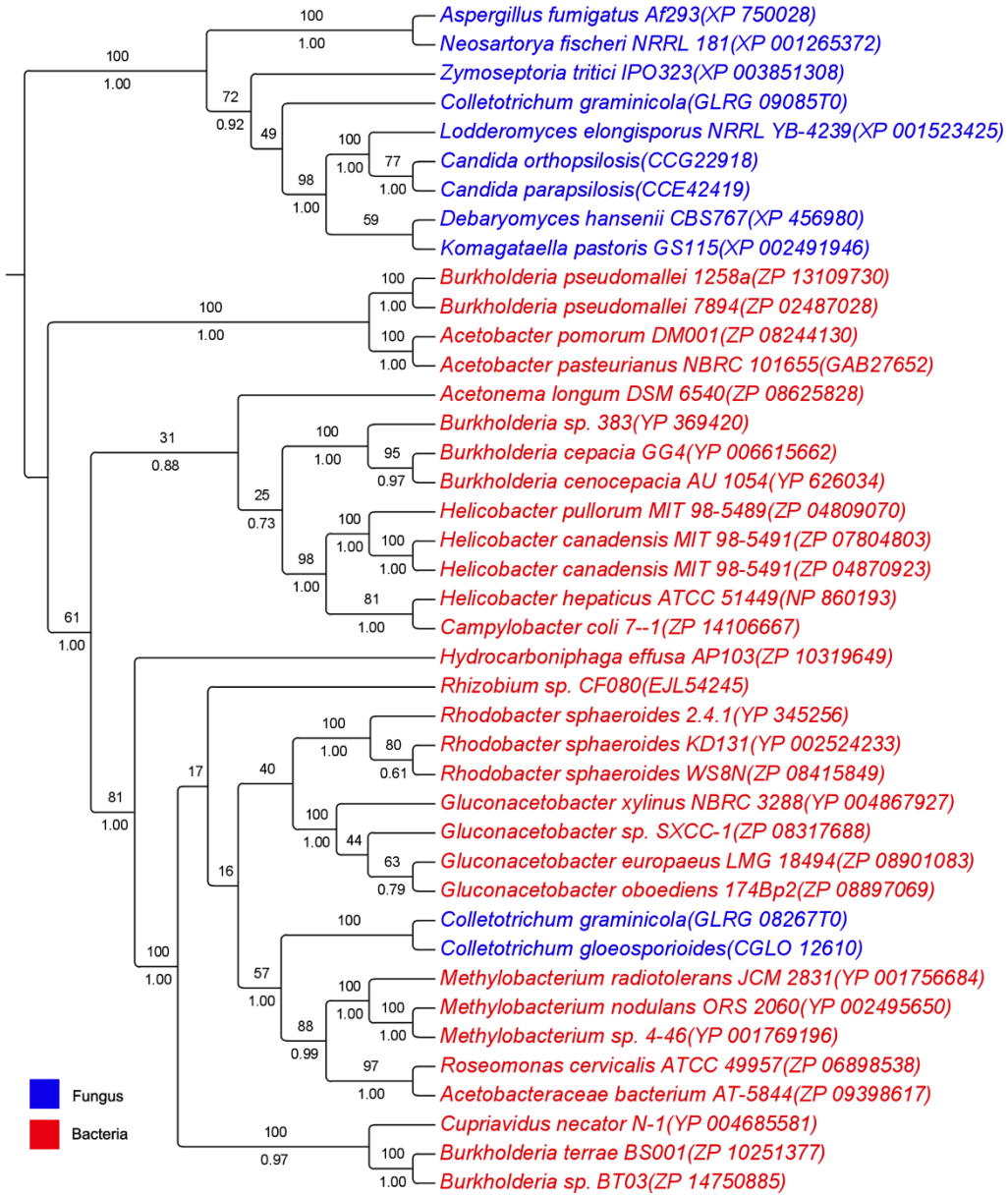


Figure 14. Maximum likelihood tree of HGT7 (CGSP_12610, GLRG_08267T0). Bootstrap percentage is shown above the branches and posterior probability is shown below the branches. Accession numbers are shown in parenthesis next to the species names.

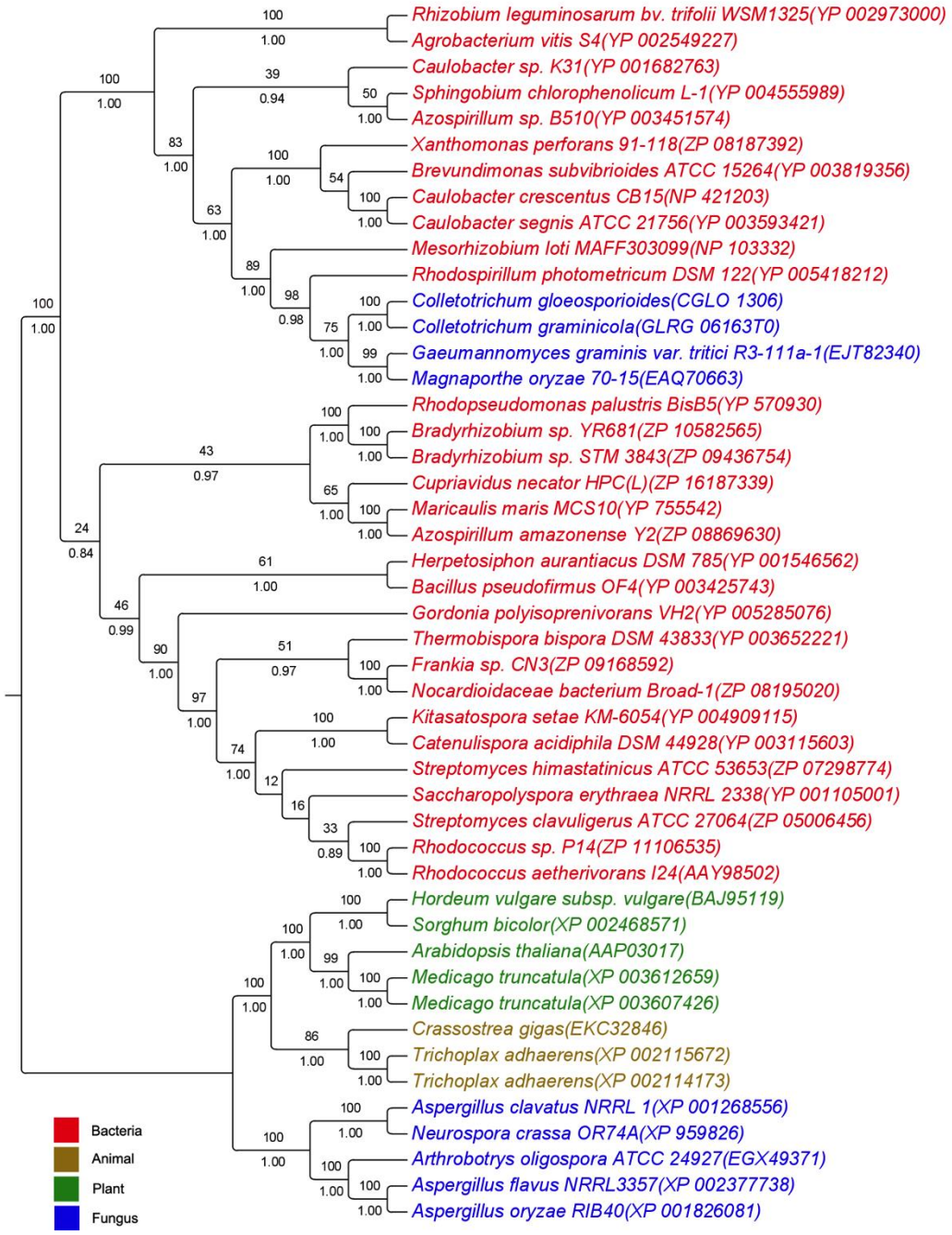


Figure 15. Maximum likelihood tree of HGT8 (CGSP_1306, GLRG_06163T0). Bootstrap percentage is shown above the branches and posterior probability is shown below the branches. Accession numbers are shown in parenthesis next to the species names.

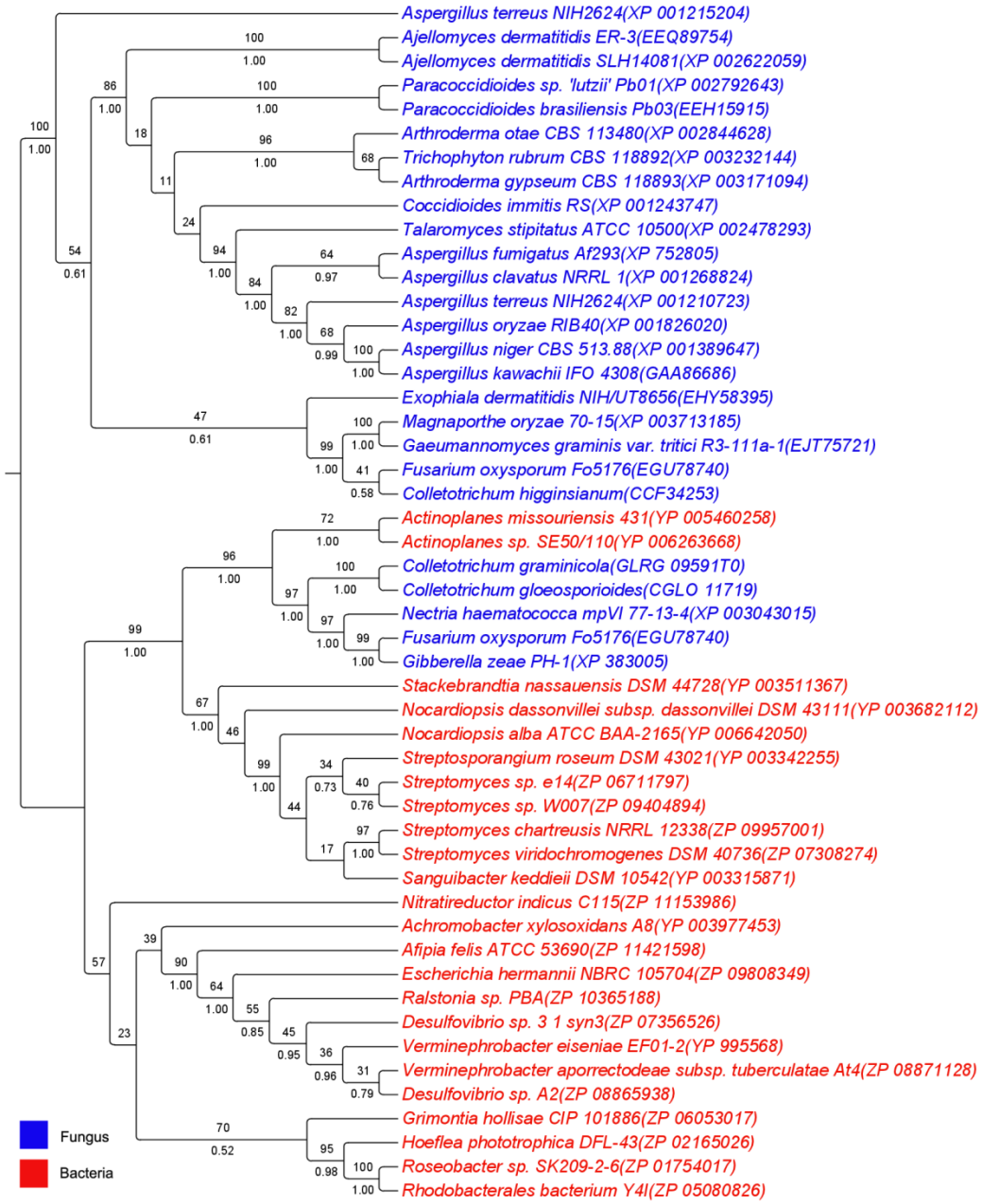


Figure 16. Maximum likelihood tree of HGT9 (CGSP_11719, GLRG_09591T0). Bootstrap percentage is shown above the branches and posterior probability is shown below the branches. Accession numbers are shown in parenthesis next to the species names.

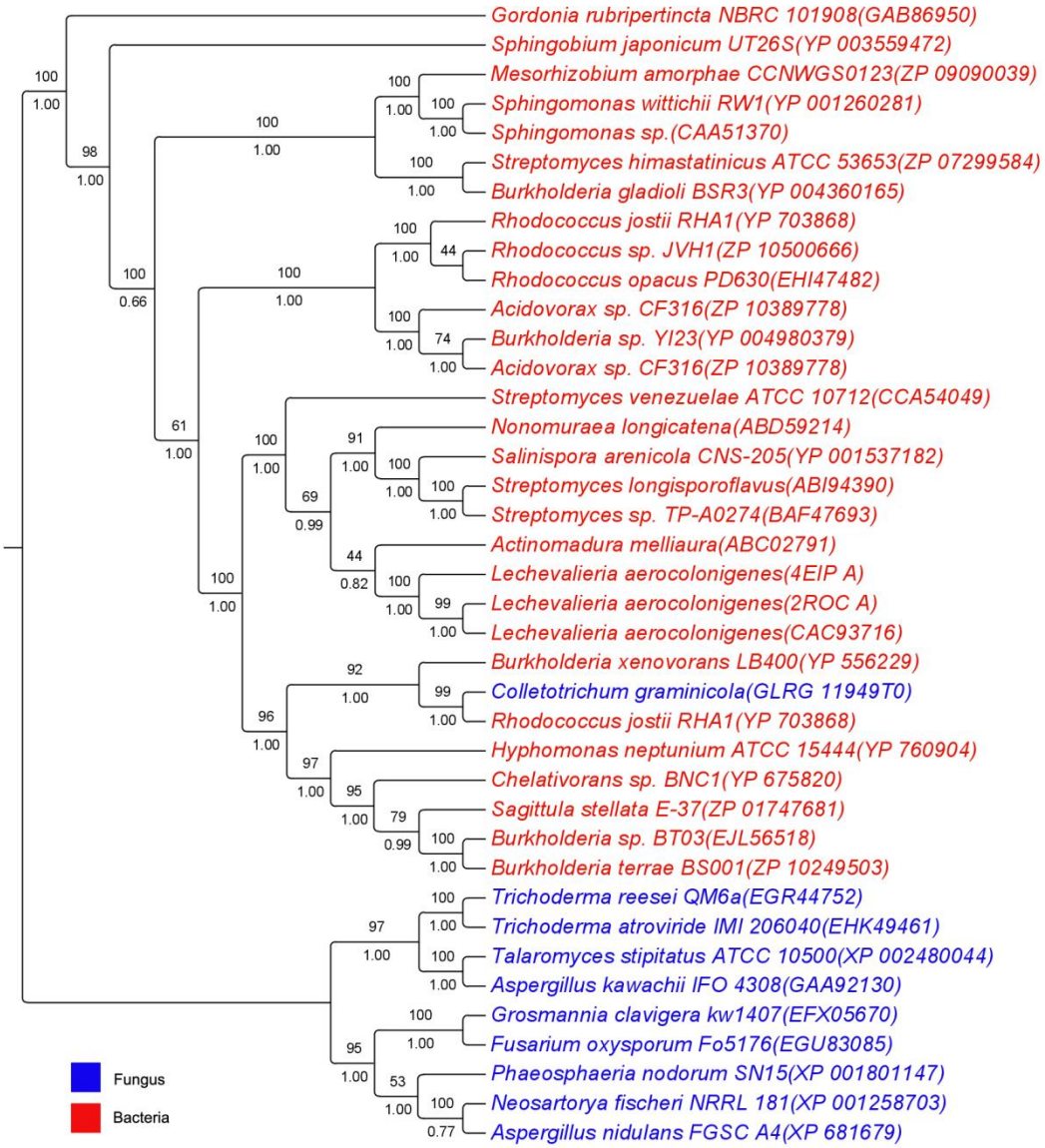


Figure 17. Maximum likelihood tree of HGT10 (GLRG_11949T0). Bootstrap percentage is shown above the branches and posterior probability is show below the branches. Accession numbers are shown in parenthesis next to the species names.

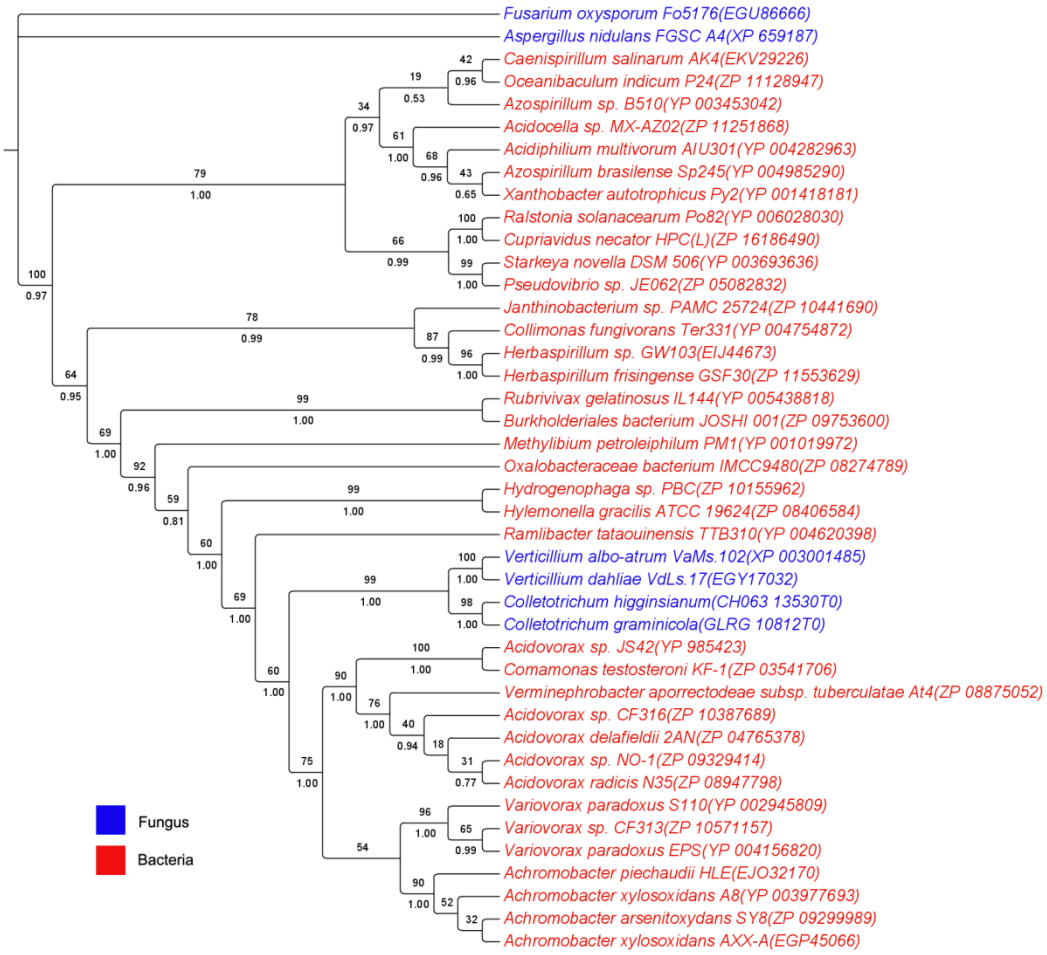


Figure 18. Maximum likelihood tree of HGT11 (GLRG_10812T0, CH063_13530T0). Bootstrap percentage is shown above the branches and posterior probability is shown below the branches. Accession numbers are shown in parenthesis next to the species names.

To determine whether any of the HGT candidates are mitochondrial genes that were incorrectly assembled into the nuclear genome sequence, or are mitochondrion to nuclear HGT events, I performed a BLASTP search vs all mitochondrial proteins available in the RefSeq database (Pruitt et al. 2012). Proteins of HGT9 show high similarity (e-value 6.17e-22) to protein XP_001875307 a 3-hydroxyacyl-CoA-dehydrogenase (EC 1.1.1.35) from the mitochondrion of *Laccaria bicolor*. However, the most similar BLAST hits are from bacteria (e.g e-value of 9.33e-116; YP_005460258 from *Actinoplanes missouriensis*). Therefore, I conclude that this is neither a case of mitochondrion to nucleus transfer nor an error in the genome assembly.

To detect signs of contamination of the genome sequences and identify events of HGT of clusters of genes, a BLASTp search of the nr database of NCBI was performed with the

upstream and downstream proteins of each HGT candidate in the genome of *C. graminicola*, *C. gloeosporioides* and *C. higginsianum*. Most of the neighboring genes showed the expected distribution of BLAST hits with *Colletotrichum* and fungal proteins (**Table 4**). Only proteins CH063_08061T0 and CGSP_9355, the flanking genes of members of HGT3 (CH063_08062T0 and CGSP_9355 respectively) present atypical BLAST hits. Proteins CH063_08061T0 and CGSP_9355 are orthologous and maybe came from the same HGT event of members of HGT3 but the phylogenetic evidence was insufficient to conclude this. Additionally, all the candidate proteins were analyzed in the Web server CENSOR (Kohany et al. 2006) to detect repetitive or transposable elements but no evidence of this kind of element was found. The entire pipeline is summarized in the **Figure 7**.

None of the 11 *C. graminicola* HGT candidates of bacterial origin have introns consistent with their prokaryotic origin. Only 26.4% of the genes encoded in the *C. graminicola* genome lack introns and the probability of selecting 11 intronless genes at random from the *C. graminicola* genome is $4.06e^{-7}$ further supporting the hypothesis that the 11 candidates are of bacterial origin. To corroborate this calculate I performed a Mann-Whitney-Wilcoxon test to determine whether the two samples of genes (the whole genome and the HGT candidates) are the same, with respect to the number of introns in the genes. These two set of genes were significantly different ($p \approx 0$), from which I conclude that the 11 HGT candidates have a different intron distribution than the genome.

I considered that the HGT candidates might lack introns because they are members of gene families that typically lack introns. To test this alternative hypothesis, I first identified 75 homologs to the HGT candidates by searching the proteomes for proteins with the same functional annotation as the candidates. I counted the number of introns in the 75 homologs and compared this sample to the intron content of the rest of the genes in the genome using a Mann-Whitney-Wilcoxon test. The 75 gene sample was not significantly different from the rest of the genes in the genome ($p = 0.33$). Thus, I conclude that the HGT candidates are not from intron-poor gene families. This evidence supports the hypothesis that the lack of introns in the HGT candidates is the consequence of their bacterial origin.

The GC content of horizontally transferred genes can be different than genes within the recipient genome (Azad and Lawrence 2012). I used the Mann-Whitney-Wilcoxon test to determine whether there is a difference in GC content between the 11 HGT candidates in *C.*

graminicola and the rest of the genes encoded in the genome. I found that the GC content of the HGT candidates is not different from the rest of the genome (p= 0.337).

Table 4. BLAST hits of neighboring genes of HGT candidates, associated chromosome location, number of introns and GC content of the candidates.

HGT group	Accession numbers			Taxonomic status of top BLAST hits of neighboring genes		Associated chromosome	Number of introns	GC content (%)
	C. graminicola	*C. higginsianum*	*C. gloeosporioides*	Upstream gene	Downstream gene			
HGT1	GLRG_01134			Fungi	Fungi	8	0	56.9
		CH063_01794		Fungi	Fungi	-	0	61.7
			CGSP_11293	Fungi	Fungi	-	0	54.1
HGT2	GLRG_11091			Fungi	*Colletotrichum*	10	0	52.3
	GLRG_11966			-	*C. graminicola*	-	0	50.3
		CH063_02340		-	Fungi	-	0	57.2
		CH063_10640		-	*Colletotrichum*		0	58.9
			CGSP_05635	Fungi	Fungi	-	0	54
HGT3		CH063_08062		Fungi-bacteria	-	-	0	54.5
			CGSP_09354	Fungi	Fungi-bacteria	-	0	52.7
HGT4	GLRG_09635			Fungi	-	2	0	60.4
		CH063_05456		-	*Colletotrichum*	-	0	67.4
			CGSP_09262	Fungi	Fungi	-	0	59.2
HGT5	GLRG_01139			Fungi	*Colletotrichum*	8	0	54.4
		CH063_01625		Fungi	Fungi	-	0	61.4
			CGSP_03952	Fungi	Fungi	-	0	55.8
HGT6	GLRG_11936			-	-	-	0	61.6
		CH063_03876		-	Fungi	-	0	68
			CGSP_08577	Fungi	Fungi	-	0	65.3

Table 4 continuation

HGT group	Accession numbers			Taxonomic status of top BLAST hits of neighboring genes		Associated chromosome	Number of introns	GC content (%)
	C. graminicola	*C. higginsianum*	*C. gloeosporioides*	Upstream gene	Downstream gene			
HGT7	GLRG_08267			Fungi	Fungi	10	0	52.4
			CGSP_12610	Fungi	Fungi	-	0	53
HGT8	GLRG_06163			Fungi	Fungi	5	0	57
			CGSP_01306	Fungi	Fungi	-	0	53.4
HGT9	GLRG_09591			Fungi	Fungi	2	0	65
			CGSP_11719	Fungi	Fungi	-	0	59.4
HGT10	GLRG_11949			-	-	-	0	53.4
HGT11	GLRG_10812			Fungi	*Colletotrichum*	6	0	71.3
		CH063_13530		-	*Colletotrichum*	-	0	70.5

Functional annotation of HGT candidates

The putative function and the biochemical pathways of the HGT candidates were deduced with BRENDA (Schomburg et al. 2013), KEGG (Kanehisa et al. 2012), MetaCyc (Caspi et al. 2010), MEROPS (Rawlings et al. 2011) and CAZy (Lombard et al. 2013) and are summarized in **Table 5**. Most of the candidates are involved in processes such as carbohydrate metabolism (HGT5, HGT8, HGT9), amino acid metabolism (HGT1, HGT7, HGT11), secondary metabolism (HGT2) or are secreted degrading enzymes (HGT4 and HGT6). All of the candidates are enzymes and except for HGT2 (glutathionylspermidine synthase) all belong to gene families that are also present in vertically transferred genes. For example HGT1 is annotated as an argininosuccinate lyase (EC 4.3.2.1), the enzyme that catalyzes the formation of fumarate and arginine from L-argininosuccinate in the urea cycle. Three other genes in the *C. graminicola* genome share the same annotation.

Table 5. *Colletotrichum* HGT candidates, putative annotation and EC code

HGT	**Locus ID**			**Annotation**	**EC code**
Candidate	*C. graminicola*	*C. higginsianum*	*C. gloeosporioides*		
HGT1	GLRG_01134	CH063_01794	CGSP_11293	Argininosuccinate lyase	EC 4.3.2.1
HGT2	GLRG_11091	CH063_02340	CGSP_05635	Glutathionylspermidine synthase	EC 6.3.1.8
	GLRG_11966	CH063_10640			
HGT3		CH063_08062	CGSP_09354	Hydroxlacyl-CoA dehydrogenase	EC 1.1.1.35
HGT4	GLRG_09635	CH063_05456	CGSP_09262	Oligoxyloglucan reducing-end-specific cellobiohydrolase	EC 3.2.1.150
HGT5	GLRG_01139	CH063_01625	CGSP_03952	Glucarate dehydratase	EC 4.2.1.40
HGT6	GLRG_11936	CH063_03876	CGSP_08577	Serine endopeptidase S1	
HGT7	GLRG_08267		CGSP_12610	L-asparaginase	EC 3.5.1.1
HGT8	GLRG_06163		CGSP_01306	Acetyl-CoA synthetase	EC 6.2.1.1
HGT9	GLRG_09591		CGSP_11719	2-deoxy-D-gluconate3-dehydrogenase	EC 1.1.1.125
HGT10	GLRG_11949			Monooxygenase, FAD-binding	
HGT11	GLRG_10812	CH063_13530		Succinyl-diaminopimelate desuccinylase	EC:3.5.1.18

Four of the HGT candidates are involved in carbohydrate metabolism. Genes in HGT5 encode glucarate dehydratase, the enzyme that transforms D-glucarate to 5-dehydro-4-deoxy-D-glucarate + H2O in the D-glucarate degradation reaction. Genes in HGT9 encode 2-deoxy-D-gluconate 3-dehydrogenase, which is involved in pentose and glucuronate interconversion. Additionally, HGT8 genes encode acetyl-CoA synthetase, which is involved in glucose biosynthesis and in the biosynthesis of the fatty acids and in the Krebs cycle (Jogl et al. 2004). Finally, HGT4 encodes oligoxyloglucan reducing-end-specific cellobiohydrolase enzymes (glycoside hydrolases belonging to CAZy family GH74), which putatively breakdown carbohydrates in the plant cell wall (Zhao et al. 2014).

Candidate HGT2 (glutathionylspermidine synthase) is the only HGT candidate that does not have vertically transmitted homologs in the *Colletotrichum* genomes and this may represent the acquisition of a completely new gene family by HGT. This enzyme catalyzes the synthesis of glutathionylspermidine and ADP+ orthophosphate from glutathione and spermidine. This reaction is well understood in trypanosomatid parasites and *E. coli* and some evidence supports the role of glutathionylspermidine synthase in detoxification of redox reactions (Smith et al. 1992; Chiang et al. 2010).

At least 4 HGT candidates have clear associations with plant interactions and virulence. Members of HGT4 encode secreted glycoside hydrolases belonging to CAZy family GH74 (Lombard et al. 2013) which degrade cellulose in the plant cell wall and are important for virulence in *Magnaporthe oryzae* and other fungi (Van Vu et al. 2012; Zhao et al. 2013; Zhao et al. 2014). Members of HGT9, encoding a short chain dehydrogenase, have significant similarity to virulence factors in the Bacterial Virulence Factors Database and the PHI-Base database of proteins with roles in pathogen/host interactions. A homolog of this protein in *Cochliobolous heterostrophus*, OXI1, is required for biosynthesis of the secondary metabolite T-toxin (Inderbitzin et al. 2010) and null mutants of the gene show reduced virulence. Members of HGT8 have strong similarity to SidI from *Aspergillus fumigatus* which plays a role in siderophore biosynthesis and subsequently, virulence (Yasmin et al. 2012). Finally HGT1, a family of arginosuccinate lyase genes, share homology with ARG1 of *Fusarium oxysporum* f. sp. *melonis* (Namiki et al. 2001). Null mutants in *F. oxysporum* have reduced virulence, linking virulence with arginine biosynthesis.

Many of the HGT candidates belong to functional categories that are described as enriched in 'volatile' genes by Wapinsky et al. (2007). Volatile genes are those that evolve by duplication and loss in contrast to uniform (genes with the same copy number in all species) and persistent (genes with at least one copy per species) genes. Only HGT1 coincides with a category enriched in persistent genes (arginine metabolism, urea cycle) (**Table 6**).

Table 6. Relation among the HGT candidates and the enrichment classes described by Wapinski et al. (2007)

HGT Candidate	Function	Wapinsky et al. enrichment classes	Category
HGT1	Argininosuccinate lyase	Arginine metabolism, urea cycle	Persistent
HGT2	Glutathionylspermidine synthase	Detoxification, stress	Volatile
HGT3	Hydroxlacyl-CoA dehydrogenase	Oxidoreductase	Volatile
HGT4	Oligoxyloglucan reducing-end-specific cellobiohydrolase	Oxidoreductase, cell wall, extracellular region	Volatile
HGT5	Glucarate dehydratase		
HGT6	Serine endopeptidase S1	Extracellular region	Volatile
HGT7	L-asparaginase	AA metabolism	Volatile
HGT8	Acetyl-CoA synthetase		
HGT9	2-deoxy-D-gluconate 3-dehydrogenase	Oxidoreductase	Volatile
HGT10	Monooxygenase, FAD-binding	Oxidoreductase	Volatile
HGT11	Succinyl-diaminopimelate desuccinylase	AA metabolism	Volatile

HGT candidates are expressed during plant infection

I reasoned that if the HGT candidates have roles in virulence then they should be expressed during infection of the host. Therefore, I used the transcriptome profiling experiments reported by O'Connell et al. (2012) to identify differences in the level of expression of the HGT gene candidates from *C. graminicola* and *C. higginsianum* during three stages of infection of their respective hosts, maize and *Arabidopsis*. I also examined the expression of a selection of vertically transmitted homologs to the HGT candidates during the infection process. The transcriptional profiling experiment was conducted at three important time points during the infection process: early (*in vitro* or *in planta* appressoria, a specialized fungal structure used to penetrate into the plant; VA, PA), middle (biotrophic phase; BP) and late (necrotrophic phase; NP). Candidates GLRG_11091 (HGT2) and GLRG_11966 (HGT2) of *C. graminicola* and candidates CH063_02340 (HGT2), CH063_10640 (HGT2) and CH063_01625 (HGT5) of *C. higginsianum* are strongly upregulated at the PA and BP time points which are very early in the infection process, suggesting a role in plant penetration or establishment of infection (**Additional file 1**). The candidates GLRG_11936 (HGT6), GLRG_06163 (HGT8), CH063_01794 (HGT1), CH063_13530 (HGT11) and CH063_05456 (HGT4) are upregulated at the latest stage of infection suggesting roles in nutrient uptake.

A steady tempo of HGT events in the Pezizomycotina

I hypothesized that HGT enables fungi to adopt new ecological niches by giving them access to new nutritional substrates or enabling pathogens or endophytes to jump to new hosts. If this is true then HGT might be more common after major extinction events that may open new niches for colonization by new species. To deduce the age of the HGT events, I reconstructed a species tree with all of the fungal species observed in the HGT detection pipeline. All of the HGT events occurred after the appearance of the Pezizomycotina (**Figure 19**) and all but three (HGT4, HGT5 and HGT6) occurred after the appearance of the Sordariomycetes. I calculated the approximate divergence times of the lineages using a Bayesian time-measured phylogeny using the fossil *Paleopyrenomycites devonicus* to calibrate the tree (**Figure 20**). I estimated the minimum ages of transference from bacteria to ancestral members of the Pezizomycotina based on the lower bound of the highest posterior density (HPD) interval. The HGT events occurred over a broad range of geological periods (from the Siluric to the Tertiary). In contrast to our expectations, there is no evidence of a burst of HGT events coinciding with major geological events. In contrast, HGT appears to be a constant, albeit rare phenomenon in the Pezizomycotina, occurring with a steady tempo during their evolution.

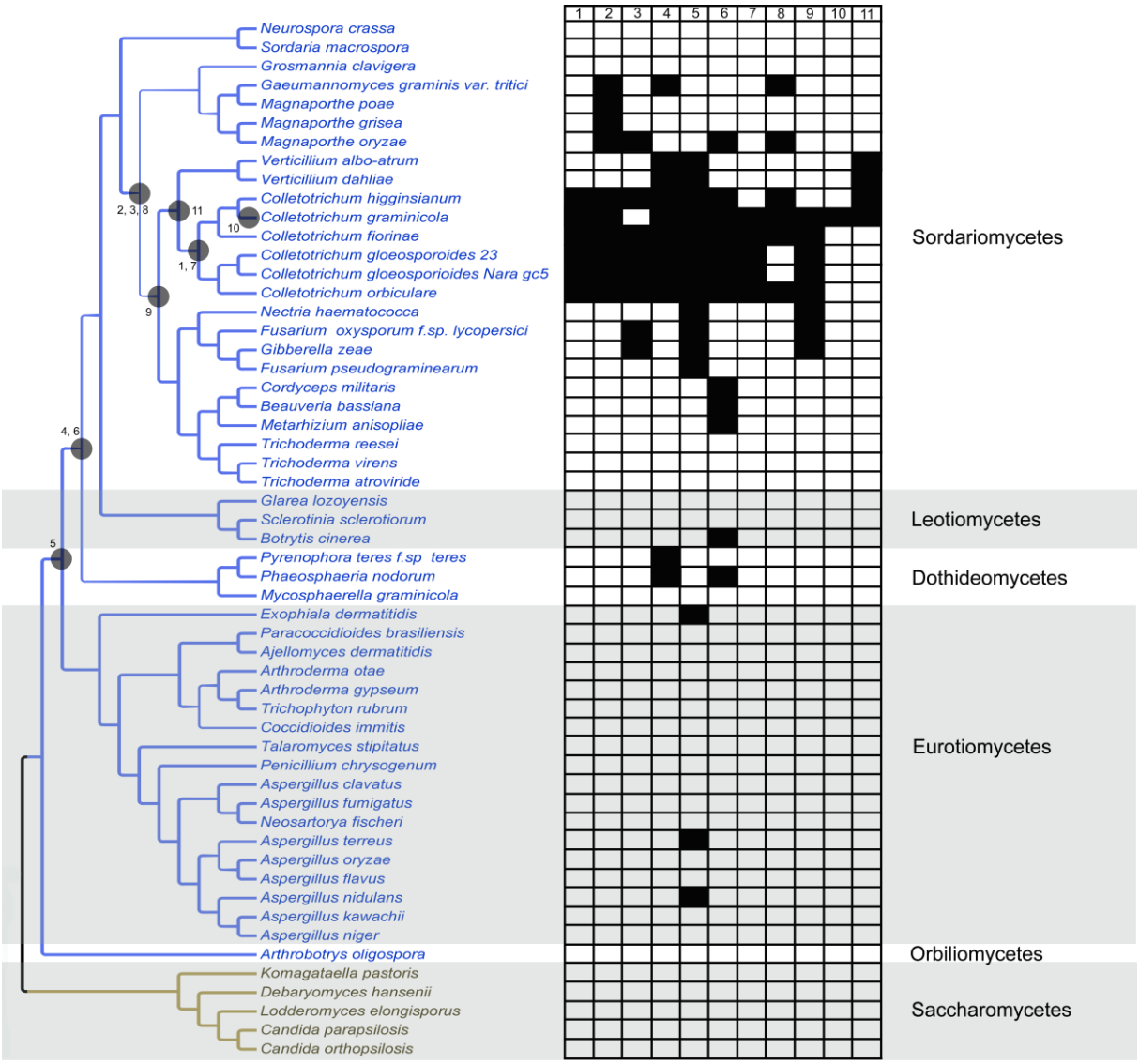


Figure 19. Ascomycetes species tree and HGT events. All species that have draft genome sequences and have homologous sequences to the HGT candidates of this study are presented. Black squares indicate the species that contain an HGT candidate. Line width is proportional to bootstrap support. The gray circles indicate the common ancestor of each HGT event and represents the most recent ancestral node where the HGT could have happened. Blue branches represent members of the Pezizomycotina species and yellow branches members of the Saccharomycotina.

While many of the HGT events are ancient, none of them are broadly distributed in extant species of the Pezizomycotina (**Figure 19**), therefore, many lineages of fungi must have lost the genes. If HGT is associated with niche adaptation then I may find that fungi with certain lifestyles have maintained horizontally transferred genes that are required for the lifestyle.

Node number	Group numbers	Time measure (MY) Median (HPD intervals)
1	HGT1, HGT7	47.68 (26.13-74.28)
2	HGT11	116.33 (75.29-160.17)
3	HGT9	168.8 (124.97-220.84)
4	HGT2, HGT3, HGT8	215.82 (166.45-275.39)
5	HGT6, HGT4	397.96 (351.65-479.67)
6	HGT5	424.44 (401.07-505.69)

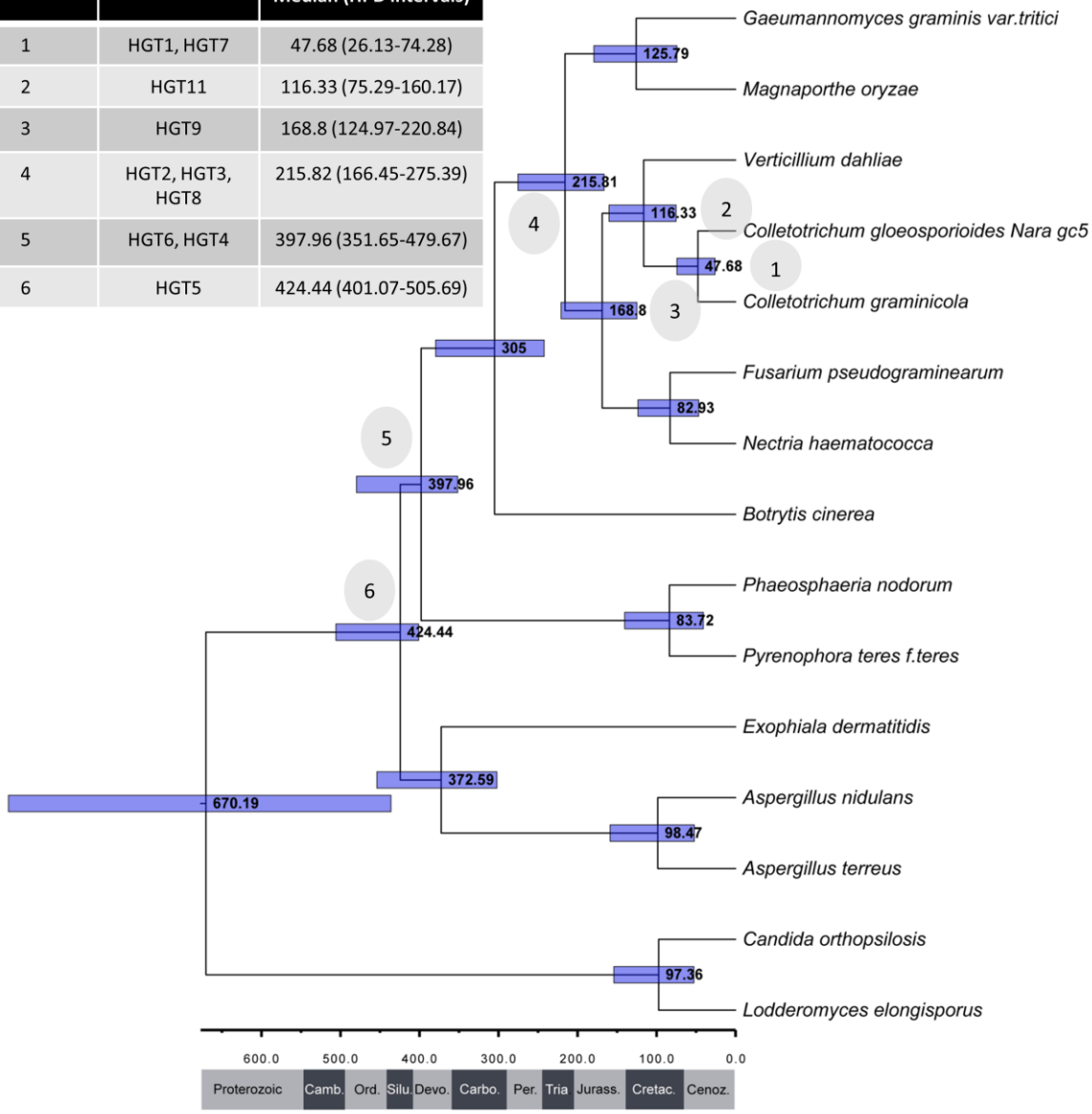


Figure 20. Time measured phylogeny (millions of years) of some of the important species in this study. Blue bars are the highest posterior density (HPD) intervals and the values on internal nodes are the median of the time estimation. Grey circles highlight common ancestor of each HGT event. The table shows the HGT events at each node, the median estimated time measured in millions of year and the HPD intervals. Under the figure a time scale shows a schematic representation of geological periods and Eras.

From **Figure 19** I can also estimate the number of gene losses that occurred following the HGT events, based on the absence of orthologs to the HGT candidates in the whole genome sequences that are available in GenBank. I identified the species with presence/absence of each of eleven groups of candidates (**Table 7**). 88.72% of Pezizomycotina species (with

complete genome available in GenBank) had lost all members of HGT5, 91.73% of species had lost all the members of HGT4 and 93.98% of species had lost all members of HGT6. In contrast, HGT1, HGT2, HGT4, HGT5 and HGT6 are always present in species of the genus *Colletotrichum*. The lifestyle information for each species used in the tree reconstruction was compiled in the **Table 8**. This information was used to determine if there is a correlation between the lifestyle of the Pezizomycotina species and the presence/absence of the HGT candidates. I did not observe a direct relationship between the lifestyle and the presence of HGT candidates except in HGT4 and HGT6, which are exclusively found in pathogenic species including the entomopathogens *Cordyceps militaris* and *Beauveria bassiana* (in the case of HGT6).

Table 7. Number of gene losses of the HGT in different Pezizomycotina lineages, based on whole genome sequences available in GenBank.

HGT Group	Lineage	Number of species available in GenBank	Losses	Percentage of losses (%)
HGT1	*Colletotrichum*	6	0	0
HGT7	*Colletotrichum*	6	0	0
HGT10	*Colletotrichum*	6	0	0
HGT11	*Colletotrichum/Verticillium*	9	1	11.11
HGT2	Sordariomycetes	53	42	79.25
HGT3	Sordariomycetes	53	45	84.91
HGT8	Sordariomycetes	53	47	88.68
HGT9	Sordariomycetes	53	45	84.91
HGT6	Pezizomycotina	133	125	93.98
HGT4	Pezizomycotina	133	122	91.73
HGT5	Pezizomycotina	118	133	88.72

Table 8. List of the species involved in the HGT events, their lifestyles and the presence/absence of the HGT candidates

Species name	Lifestyle	HGT1	HGT2	HGT3	HGT4	HGT5	HGT6	HGT7	HGT8	HGT9	HGT10	HGT11
Botryotinia fuckeliana	Necrotroph (Ohm et al. 2012)						■					
Verticillium albo-atrum	Necrotroph (Ohm et al. 2012)				■	■						■
Verticillium dahliae	Necrotroph (Ohm et al. 2012)				■	■						■
Gibberella zeae	Necrotroph (Ohm et al. 2012)			■		■				■		
Fusarium pseudograminearum	Necrotroph (Akinsanmi et al. 2007)					■						
Phaeosphaeria nodorum	Necrotroph (Ohm et al. 2012)				■		■					
Pyrenophora teres	Necrotroph (Ohm et al. 2012)				■							
Nectria haematococca	Necrotroph/saprotroph (Ohm et al. 2012)					■				■		
Gaeumannomyces graminis	Necrotroph/saprotroph (Freeman and Ward 2004)		■		■				■			
C. orbiculare	Hemibiotroph (Gan et al. 2013)	■	■	■	■	■	■	■	■	■		■
C. gloeosporioides	Hemibiotroph (Gan et al. 2013)	■	■	■	■	■	■	■		■		■
C. graminícola	Hemibiotroph (Vargas et al. 2012a)	■	■		■	■	■	■	■	■	■	■
C. higginsianum	Hemibiotroph (O'Connell et al. 2012)	■	■	■	■	■	■		■			■
Magnaporthe grisea	Hemibiotroph (Ohm et al. 2012)		■									
Metarhizium anisopliae	Hemibiotrophs (Lowe and Howlett 2012)						■					
Magnaporthe oryzae	Hemibiotrophs (Lowe and Howlett 2012)		■	■			■		■			
Fusarium oxysporum	Hemibiotroph/ endophyte/saprotroph (Ohm et al. 2012)			■		■				■		

Magnaporthe poae	Plant pathogen (Landschoot and Jackson 1989)											
C. fiorinae	Plant pathogen, entomopathogen (Shivas and Tan 2009)											
Cordyceps militaris	Entomopathogenic fungus (Zhang and Liang 2013)											
Beauveria bassiana	Entomopathogenic fungus (Clarkson and Charnley 1996)											
Aspergillus terreus	Soil saprophyte-occasional animal pathogen (Rougeron et al. 2013)											
Exophiala dermatitidis	Saprophyte, human pathogen (Döğen et al. 2013; Najafzadeh et al. 2013)											

Discussion

In this study, I detected 12 genes with evidence of HGT from bacteria and one from plants to the genus *Colletotrichum*. The genes were classified into 11 families of homologous sequences (**Table 5**). None of the genes of bacterial origin have introns which is statistically unlikely and consistent with their prokaryotic origin, consistent with other authors have also reported that horizontally transferred genes from bacteria lack typically introns (Garcia-Vallvé et al. 2000; Schmitt and Lumbsch 2009).

Many of the HGT candidates are involved in metabolic processes such as synthesis of amino acids (HGT1, HGT7 and HGT11), lipid metabolism (HGT3), sugar metabolism (HGT8, HGT9 and HGT5) and secondary metabolism (HGT2). HGT1 belongs to a family of argininosuccinate lyase genes responsible for the formation of arginine and fumarate in the urea cycle. In *C. graminicola*, HGT1 member GLRG_01134 as well as the three other arginosuccinate lyase genes in this species are expressed in all stages of the infection process (**Additional file 1**). In contrast, the *C. higginsianum* HGT1 member CH063_01794 is upregulated only late in the infection process. A previous study by Takahara et al. (2012) found that arginine biosynthesis in *C. higginsianum* is essential for the early stages of plant infection but the upregulation of CH063_01794 in the last stage of infection (necrotrophy) suggests a different role in this pathogen. In another study, a mutant of *ARG1*, an arginosuccinate lyase encoding gene in *Fusarium oxysporum* f. sp. *melonis* showed reduced virulence (Namiki et al. 2001). These studies point to an important role of arginine synthesis during the infection process and suggest that the acquisition of an arginosuccinate lyase by HGT may have improved fitness of an ancestral fungal species by increasing virulence.

Other enzymes with roles amino acids metabolism are the members of HGT7, a family of L-asparaginases in the aspartic acid synthesis pathway and HGT11 a family of succinyl-diaminopimelate desuccinylase enzymes in the lysine biosynthesis pathway. HGT7 was found only in the genus *Colletotrichum* and HGT11 is found in *Colletotrichum* as well as the closely related genus *Verticillium*. In bacteria, the horizontal transfer of amino acid metabolic genes have been proposed to provide metabolic plasticity to the recipient species of both non pathogenic (Marri et al. 2007; Boggs et al. 2012) and pathogenic (Rohmer et al. 2011)

species, enabling them to exploit new nutritional sources. This hypothesis has also been proposed for fungal HGT genes (Richards et al. 2011; Soanes and Richards 2014).

Many of the HGT candidates have functions related to sugar and lipid metabolism (HGT8, HGT9, HGT5 and HGT3), consistent previous reports (Richards et al. 2011). Richards et al. (2011) propose that the HGT of genes in these functional categories can increase the osmotrophic capacity of fungi. All fungi are osmotrophs, yet horizontally transferred genes are frequently lost in many lineages indicating that there was a subsequent loss of selective pressure to maintain the gene after its initial acquisition. I suggest that HGT is important in niche adaptation and that during the evolution of the fungi, changes in a species' niche have lead to changes in the selective pressure on the genes required for nutrient uptake.

The putative functions of genes in HGT4 and HGT6 in the degradation of cell wall sugars (HGT4) and proteins (HGT6) and the presence of a predicted secretion signal peptide opens the possibility that they may be secreted lytic enzymes. The time of overexpression of GLRG_11936 (HGT6) in *C. graminicola* (**Additional file 1**) coincides with the necrotrophic phase suggesting that this protein has a bigger role in nutrient acquisition rather than in host penetration. This is also the case with CH063_05456 (HGT4) in *C. higginsianum* which is also overexpressed in the necrotrophic phase. A similar scenario was described by O'Connell et al. (2012) in the transcriptomic analysis of *C. graminicola* and *C. higginsianum*. The authors observed a vast array of lytic enzymes induced at the transition to necrotrophy, the stage at which the pathogen uses dead and dying host cells as a nutrient source to support rapid colonization and sporulation. In contrast, GLRG_09635 (HGT4) of *C. graminicola* does not have changes in expression levels during the infection process, suggesting that its role during infection may be different from that of its *C. higginsianum* counterpart.

The HGT events HGT4 and HGT6 are among the most ancient, having occurred at least 397 million years ago. The transferred genes have been retained in only a few species, all of which are pathogenic, and in the have of HGT6, includes the entomopathogens *Cordyceps militaris* and *Beauveria bassiana*. This suggests that these genes have an important function in host-fungal interactions and that it has maintained this function for millions of years.

A large number of the HGT candidates have functions related to amino acid and carbohydrate metabolism. In the Ascomycetes these two categories were associated with "volatility", a term introduced by Wapinsky et al. (2007) to describe genes that evolve by duplication and loss in

contrast to uniform (genes with the same copy number in all species) and persistent (genes with at least one copy per species) genes. Volatile genes have the ability to evolve by losing or gaining copies without drastic selective consequences, providing new functions or pruning old ones. The functional categories of the HGT candidates reported in the meta analysis of Richards et al. (2011) are also rich in volatile functional categories such as carbohydrate metabolism suggesting that HGT is much more likely if the gene is a member of a volatile family.

The ages deduced for the HGT events in the time calibrated phylogram (**Figure 20**) reflect the antiquity of the HGT events reported in this study. Except for HGT10 the transference of all the candidates was situated between 47.68 million years ago (mya) and 424.44 mya (taking into account the mean of the estimations). That reflects very ancient HGT events and despite the high rate of losses in the Pezizomycotina these proteins probably were useful in the adaptation of the organisms to their niches. From the distribution of transfer events in time I can deduce that the HGT phenomenon happened during all the evolution of Pezizomycotina. Also, given the high percentage of gene losses among the laterally transferred genes I suggest that the number of HGT events detected in the present study represents only a small fraction of the transferences that occurred in the past.

CHAPTER II

Horizontal transfer of a subtilisin gene from plants into an ancestor of the plant pathogenic fungal genus *Colletotrichum*

Introduction

The genus *Colletotrichum* within the Ascomycetes includes a large number of phytopathogenic species that affect a wide range of crops worldwide (Latunde-Dada 2001; O'Connell et al. 2012). Species of this genus are the agents of anthracnose diseases that cause devastating yield losses in agriculture (Perfect and Green 2001). To achieve infection, *Colletotrichum* species employ a diversity of molecules such as effectors, kinases, hydrolytic enzymes and others (Takano 2004; Dunaevsky et al. 2007; Kleemann et al. 2008). Within the molecular arsenal of these organisms, the catalytic enzymes provide a wide range of tools to achieve successful host infection. One of the largest groups of catalytic enzymes is composed of serine proteases, a group of proteins that can be found in all kingdoms of life (Siezen et al. 1991; Withers-Martinez et al. 2012). These enzymes include endopeptidases and exopeptidases organized into 12 clans and 35 families according to the MEROPS peptidase database (Rawlings et al. 2011). The MEROPS S8 family of subtilisins is especially important for the large number of proteins that it contains as well as its broad taxonomic distribution. The S8 family constitutes a heterogeneous group of proteins with a characteristic catalytic triad peptide (Asp, His and Ser), with no other structural resemblance attributable to all members of the family. The MEROPS database subdivides S8 subtilisins in two subfamilies, the real subtilisins as S8A and the S8B kexin subfamily that includes proprotein convertases (Siezen and Leunissen 1997; Tripathi and Sowdhamini 2006).

This extended family of enzymes presents a wide range of functions and their members are involved in a broad spectrum of metabolic processes in plants and fungi, with many members having roles in plant-microbe interactions (**Table 9**). An interesting group of proteins that belong to the S8A subtilisins are pathogenesis-related 7 (PR-7) proteins with roles in plant-pathogen interactions. PR proteins are defined as molecules that are induced in plants under pathological or related situations (van Loon et al. 1994). These proteins form a group with various chemical characteristics and biological functions. For that reason a standardization of the nomenclature was proposed, dividing the proteins by sequence similarity and enzymatic or biological activity (van Loon et al. 1994). One of those groups was named PR-7, represented by tomato P69 proteins (Tornero et al. 1996; Tornero et al. 1997; Jorda et al. 1999) a group of proteins of family S8.

Table 9. Functions of subtilisin family members in plants and fungi

Name	GenBank accession number	Organism	Function	Reference
SbtM1 and SbtM3	BAF95755, BAF95754	*Lotus japonicus*	A role in the development of arbuscular mycorrhiza	(Takeda et al. 2009)
Phytaspase	ACT34764	*Nicotiana tabacum*	Involved in programmed cell death	(Chichkova et al. 2010)
Gm-1	Glyma18g48580.1 *	*Glycine max*	Contains the signal peptide GmSubPep that induces defense-related genes	(Pearce et al. 2010)
P69B and P69C	CAA76725, CAA06412	*Solanum lycopersicum*	Pathogen-related proteins	(Jorda et al. 1999; Meichtry et al. 1999)
Ara12	AAN13181	*Arabidospsis thaliana*	Undefined. Probably involved in the development of plant	(Hamilton et al. 2003)
Pr1	ACZ28128	*Beauveria bassiana*	Virulence factor involved in the pathogenicity against insects	(Donatti et al. 2008)
Mp1	AAD26255	*Magnaporthe poae*	Implicated in the infection process	(Sreedhar et al. 1999)
Ver112	Q68GV9	*Lecanicillium psalliotae*	Nematicidal activity	(Yang et al. 2005)
At1	AAB62277	*Epichloe typhina*	Potential role in symbiosis	(Reddy et al. 1996)

* Genome project accession number

One of the most interesting characteristics of family S8 is their domain variability. The peptidase S8 domain (the most characteristic domain of this family, Pfam:PF00082) is usually found combined with various other domains. The domains typically found in family S8 are: PA (Pfam:PA02225), inhibitor I9 (Pfam:PF05922), alpha-1,3-glucanase (Pfam:PF03659), chitinase class II group (Pfam: PF00704), pectin lyase (SCOP:51133), cyclin domains (InterPro: IPR006670), DUF1034 (Domain of Unknown Function 1034, Pfam:PF06280), DUF1043 (Pfam:PF06280), CytochromeP450 (Pfam:PF00067), P domain (Pfam:PF01483) glyco_hydro_71 (Pfam:PF03659), glyco_hydro_18 (Pfam:PF00704), cyclin (InterPro:IPR006670), Pro-kuma_activ (SMART:SM00944), Sir2 (Pfam:PF02146) and sac_ganp (Pfam:PF03399). This domain combination could be identified in S8 subtilisins of animals, plants, bacteria or fungi (Muszewska et al. 2011).

The subtilisin S8 family represents an important group of proteases. Over 200 members have been identified in bacteria, archaeas, eukaryotes and viruses. In Arabidopsis 56 S8 family members have been identified, and 63 have been reported in maize (Tripathi and Sowdhamini 2006). In fungi, members of the S8 family are also abundant. At least four of the six subfamilies of subtilisins in the classification of (Siezen and Leunissen 1997) and (Saeki et al. 2000) were found in different Ascomycota and Basidiomycota (Bryant et al. 2009; Muszewska et al. 2011). *Colletotrichum higginsianum* contains thirty six subtilisins S8 and *C.*

graminicola twelve. This family is apparently highly expanded in *C. higginsianum* compared with *C. graminicola* and other fungi (O'Connell et al. 2012).

Phylogenetic trees of the S8 family are, in general, congruent with the species tree (Siezen and Leunissen 1997; Bryant et al. 2009; Muszewska et al. 2011) showing that subtilisins are predominantly transmitted vertically to descendants. To my knowledge, an event of horizontal gene transfer (HGT) has not been reported in eukaryotes for this family of proteins.

Cases of HGT were previously taken as isolated incidents and were not considered important, but now have gained enormous interest due to their consequences on species evolution. Numerous cases have been reported in recent years, especially in prokaryotes (Ochman et al. 2000; Dutta and Pan 2002; Heuer and Smalla 2007; Kim et al. 2012; Techtmann et al. 2012). Reports of HGT events in eukaryotes are less abundant, congruent with the idea that HGT is rare in eukaryotic organisms. Barriers such differential intron processing, incompatible gene promoters, unpaired meiotic DNA, eukaryotic membranes, and alternative genetic codes may present obstacles for the horizontal transmission of genes (Keeling and Palmer 2008; Fitzpatrick 2011). However, an increasing number of publications provide evidence of HGT in eukaryotes (Friesen et al. 2006; Choi and Kim 2007; Slot and Rokas 2010; Tiburcio et al. 2010). In contrast, one of the most difficult things to explain is the mechanism by which HGT events occur, especially among unrelated species that come from different kingdoms. There is no direct evidence of a mechanism that enables HGT but some hypotheses have been proposed. For example in fungal HGT, vectors such as mycoviruses, plasmids and transposable elements have been proposed to explain this phenomenon (Rosewich and Kistler 2000). Also, the physical interaction between symbiotic or host-parasite organisms have been suggested as a way for the transfer to occur (Gogarten 2003). Experimental evidence is still needed to confirm or reject these hypotheses.

There are very few reports of HGT from plants to fungi. The *Av1* gene of *Verticillium dahliae* and their homologs in other plant pathogens was presented as potential candidate of HGT from plants to fungi (de Jonge et al. 2012). Strong evidence for four possible events of HGT from plants to fungi were provided by (Richards et al. 2009). These proteins were predicted as Zinc binding alcohol dehydrogenase, DUF239 domain protein, Phosphate-responsive 1 family protein and a hypothetical protein with similarity to zinc finger (C2H2-type) protein.

In this study I provide evidence for the presence of a plant-like S8A subtilisin in the genomes of several species of the genus *Colletotrichum*. These proteins show evidence of lateral gene transfer from plants to a *Colletotrichum* ancestor. This is the first time that evidence is provided for the horizontal transfer of a plant subtilisin to pathogenic fungi. The expression analysis shows that at least two subtilisins of maize are down-regulated when the CPLS is induced. In view of the wide variety of processes that plant subtilisins are involved, it is possible that *Colletotrichum* acquired and use plant-like subtilisins to manipulate the host metabolism.

Materials and Methods

Identification of HGT events

A BLASTp (Altschul et al. 1997) search was done (e-value threshold: 10^{-5}) using the predicted protein sequences from the *Colletotrichum graminicola* M1001, *Colletotrichum higginsianum* IMI 349063 (Broad Institute of Harvard and MIT) and *Colletotrichum gloeosporioides* Cg-14 proteomes (Alkan et al. 2013). Annotated proteins from organisms with complete proteomes deposited in UniProt (www.uniprot.org) were used as the initial BLAST database. The proteins with 80% or more of the BLAST hits from members of the Viridiplantae were selected as candidates. After the first round of candidates was identified, other databases were used to verify the absence of putative homologues not detected in the UniProt database. The NR and EST databases from NCBI (www.ncbi.nlm.nih.gov) and all fungal proteomes of the Broad Institute (www.broadinstitute.org) and the Joint Genome Institute (www.jgi.doe.gov) were used in this second round of searches. The results were analyzed automatically with Python scripts, taking in account the taxonomy of BLAST hits and comparing these with the taxonomy of *Colletotrichum* genera. The percentage of hits with a Viridiplantae taxonomy label was reported.

Phylogenetic analysis

The protein sequences used for the phylogenetic reconstruction came from NR database of NCBI (www.ncbi.nlm.nih.gov), Join Genome Institute (www.jgi.doe.gov) and MaizeSequence (www.maizesequence.org/index.html). Using the candidates as a query,

BLAST hits (e-value threshold: 10^{-10}) with at least 30% of identity and 70% of coverage were used for further analyses. From these, only family members of subtilisins S8A according to MEROPS (Rawlings et al. 2011) were chosen. The sequences were submitted to the PANTHER classification system (Mi et al. 2009) to recover only the members of PANTHER's family PTHR10795 sub family 17. The selected sequences were aligned with MAFFT v6.814b (Katoh et al. 2002) and then manually edited to remove highly divergent alignment columns. Two alternative alignments were also prepared using Gblocks (Talavera and Castresana 2007) and trimAl (Capella-Gutiérrez et al. 2009). The percentage of unresolved quartets was used as measure of the contribution of each sequence to resolve the topology of the phylogenetic tree. Using the program TREE-PUZZLE (Schmidt et al. 2002) the alignment was analyzed to ensure that all sequences had less than 10% unresolved quartets. Any sequences with more than 10% of unresolved quartets were removed of the analysis. MODELGENERATOR (Keane et al. 2006) was used to predict an accurate model of sequence evolution and matrix of substitution from the dataset. A maximum posterior tree was constructed with MrBayes (Ronquist and Huelsenbeck 2003), performing 2,000,000 generations of samples, using the substitution matrix and model predicted by MODELGENERATOR but allowing the program to calculate the proportion of invariable sites and the alpha parameter for gamma distribution. Two Multiple Chain Markov Chain Monte Carlo (MCMCMC) searches were conducted with four chains each (three heated and one cold). The convergence between them was checked using a sample frequency of 1000 generations. A burn-in of 25% of generations was excluded to reconstruct the Bayesian consensus tree.

PhyML (Guindon and Gascuel 2003) was used to reconstruct the maximum likelihood tree and perform 100 non-parametric bootstrap replicates and SH-like branch test support. RAxML (Stamatakis 2006) was used to conduct a rapid Bootstrap analysis with 1000 replicates. Substitution matrix and model selection of MODELGENERATOR were used.

To verify the accuracy of the trees reconstructed, I used statistical topologies test to corroborate the position of the fungal subtilisins inside the plant branches. MrBayes was used to constrain specific groups and generate trees to evaluate different topologies. Expected Likelihood Weight (ELW) test was conducted in TREE-PUZZLE. The AU (Approximately

Unbiased) and SH (Shimodaira and Hasegawa) tests were conducted in CONSEL software (Shimodaira and Hasegawa 2001).

To reconstruct the tree of maize and *Colletotrichum* subtilisins, all the sequences identified as subtilisins S8A in MEROPS (http://merops.sanger.ac.uk/cgi-bin/blast/submitblast/merops/advanced) in the proteomes of *Zea mays*, *Colletotrichum graminicola* and *Colletotrichum higginsianum* were used. The same procedure explained earlier was applied with two differences. Only PhyML bootstrap analysis was used to support the topology and the percentage of quartets was not calculated.

Amplification and sequencing of *Colletotrichum gloeosporioides CGLO_10271* gene

To verify the presence of a premature stop codon in the putative CPLS *CGLO_10271* in the *Colletotrichum gloeosporioides* genome, PCR amplification and sequencing were used. The PCR was performed with PCR extender system Taq polymerase (5 Prime) with forward: AAGCTGCGACGGGGTCAACG and reverse: GCGGCGTCGTCAAGTCTGCT primers for 30 cycles. PCR products were visualized after electrophoresis on agarose gels stained with ethidium bromide. The material for sequencing was isolated from agarose gels, purified, and then sequenced by the Genomics and Proteomics Sequencing Service of the University of Salamanca, using the same primers mentioned above.

Domain determination

The PA domain, I9 inhibitor and peptidase S8 domain were identified by Pfam (Punta et al. 2011). Fn-III like domain was predicted by visual structural homology with SSP-19 (sperm-specific protein, PDB entry 1ROW). Signal peptide and cleavage site were predicted by WoLF PSORT (Horton et al. 2007) and SignalP (Petersen et al. 2011). PANTHER (Mi et al. 2009) was used to classify the proteins in more specific categories. Profile hidden Markov models were constructed using HMMER (Finn et al. 2011).

3D structure determination

The prediction of 3d structures was made in Phyre 2 (Kelley and Sternberg 2009). The manipulation, structural alignment and comparison between 3D models were done with PhyMOL (Schrödinger 2010). Dali pairwise comparison (Hasegawa and Holm 2009) was also used to evaluate the general statistics of the structural alignment. The similarity between the structural regions was evaluated manually and with PhyMOL ColorByRMSD script.

Gene expression assays[1]

Total RNA samples were prepared from maize leaves infected with *C. graminicola* strain M.1001 following the methodology previously described by (Vargas et al. 2012a). Briefly, ten droplets (7.5 ml) containing $3x10^5$ spores/ml were inoculated on the adaxial side (away from the midvein) of the third leaf of maize plants (highly susceptible inbreed line Mo940) in the V3 developmental stage. Plant leaves were harvested 24, 48 and 72 hours post-infection (hpi) and total RNA was prepared using TRIZOL® reagent (Gibco-BRL) according to the protocol provided by the manufacturer.

To assay the gene expression pattern of a set subtilisin S8A genes from maize and *C. graminicola*, semiquantitative RT-PCR experiments were conducted by reverse transcription of RNA followed of PCR reactions using specific primers for each gene. Due to the high sequence identity among the various subtilisin homologs in maize, the specific primers were designed using the predicted 5´UTR region of each sequence. cDNA synthesis was performed using 5 mg of total RNA, Moloney Murine Leukaemia Virus-Reverse Transcriptase (MMLV-RT®, Promega) and oligo-dT primers. Previous to the reverse transcription, RNA samples were treated with Turbo DNA-Free DNAse (Ambion, Austin Texas) to remove trace amounts of genomic DNA.

The amplification of the constitutively expressed beta-tubulin and GAPc genes from *C. graminicola* and maize, respectively, were used as loading and RT controls. PCR reactions

[1] The expression assays were performed by Walter Vargas

were performed in the linear range of product amplification that is between 25 and 35 cycles depending on the abundance of the different target in the samples. To confirm the absence of genomic DNA contaminations RT-PCR assays were performed in reactions where the reverse transcriptase was omitted. PCR products were visualized after electrophoresis on 2% agarose gels and staining with ethidium bromide. Primers used for the PCR reactions are listed in **Table 10**.

Table 10. Primers used for gene expression assays

Gene	Primer	Sequence	Product Size (bp)
GRMZM2G073223_P01	223 Fw	ATTCCGGTCAGTGCGCAGGC	
	223 Rv	TGCCGTCGTGAACAGCCGTC	421
GRMZM2G099452_P02	452 Fw	GCCAGCACCAGCGGAACTGT	
	452 Rv	CAGTGGGCACCGAGGGAGGA	245
GRMZM2G013986_P01	986 Fw	CCAGCTCACCGCCAGTGCTC	
	986 Rv	TAAGCAGCCGCCTTGGCGTT	210
GRMZM2G091578_P01	578 Fw	GCAGTCACGCCTTCCCGTCC	
	578 Rv	TGGCGCCGCATTGTGAGTGA	204
GRMZM2G354373_P01	373 Fw	TTTTCCCGATCCGGCACCGC	
	373 Rv	GGAGGACAGCAGGCCCAGGA	428
GRMZM2G120085_P01	085 Fw	CCGTCTGTGCACCGGACACC	
	085 Rv	CGCCTCCAGATTGCCCGTGG	273
GRMZM2G414915_P01	915 Fw	GCCGCTGTGCCTAGCTCTCG	
	915 Rv	CTCGGCCTCGTCCTCGTCCA	265
AC196090.3_FGP006	090 Fw	CTCCTCCTGCTGCTGTCCGC	
	090 Rv	TCAGTGAGGCTGGCGGCGAA	230
GRMZM2G121293_P01	293 Fw	GCACGAACATGCGTAACATCGGC	
	293 Rv	GAGCATCTTGGCGGCGGAGG	203
GRMZM2G430039_P01	ZmP69Fw	CGGCGACCGCCTAGCATCTG	
	ZmP69Rv	CGGCGACCGCCTAGCATCTG	212
ZmGAPc	GAPc-F	GCTAGCTGCACCACAAACTGC	
	GAPc-R	TAGCCCCACTCGTTGTCGTAC	500
GLRG_05578	CgSLFw	GCCGATCCCTCATCGCTGCC	
	CgSLRv	GCCGAGCAGGGCCGAGTTTT	230
GLRG_01057 (CgTub)	CtubF	CAGTCCCTTGGGCGGCACAG	
	CtubR	TCCCGGGGCAATTGAACGCC	350

Results

The aim of this study was to determine whether genes from plants have been horizontally transferred members of the genus *Colletotrichum*. Our first step to identify potential horizontally transferred genes was to perform a battery of BLAST searches against a database composed of all proteomes available in the UniProt database (www.uniprot.org) using the *C. graminicola*, *C. higginsianum* and *C. gloesoporioides* proteins as query sequences. These BLAST searches resulted in the identification of one protein from *C. graminicola* (locus tag, GLRG_05578; GenBank accession number EFQ30434) and two from *C. gloeosporioides* (locus tags CGLO_07890 GenBank accession number KC544259 and CGLO_10271 GenBank accession number KC544258, genome project number SUB133583) with high percentages of BLAST hits in the kingdom Viridiplantae. I performed additional BLAST searches using these three protein sequences versus other databases (see materials and methods) but no evidence of homology to fungal proteins was found. The proteins GLRG_05578 and CGLO_07890 were identified as members of the subtilisin S8A family and they were designated as *Colletotrichum* plant-like subtilisins (CPLSs). The protein CGLO_10271 appears to be a truncated CPLS. By analyzing the DNA sequence of this gene I determined that a premature stop codon truncates the protein's translation. By aligning the three gene sequences, I identified a thymine at position 1470 downstream of the start codon that caused a frame shift in the open reading frame, which resulted in a premature stop codon. The presence of this premature stop codon was confirmed after PCR amplification and sequencing of the genomic region. I performed TBLASTN searches of the three *Colletotrichum* genomes using the CPLSs as query sequences to identify the presence of possible CPLS pseudogenes, but no evidence was found.

Interestingly, these BLAST searches failed to identify CPLSs in the genome of *C. higginsianum* leading us to speculate that this species lacks a copy of this gene. To confirm that the *C. higginsianum* genome lacks a CPLSs, I performed TBLASTN searches of the *C. higginsianum* RNA-Seq sequence reads (O'Connell et al. 2012) and identified sequences homologous to the *C. graminicola* and *C. gloeosporioides* CPLS. These results indicate that like the other *Colletotrichum* spp. that I examined, *C. higginsianum* also contains a CPLS. In addition, a BLASTP search of the predicted protein sequences of *C. acutatum* (genome sequence kindly provided by R. Baroncelli) revealed the presence of an ortholog in this

species as well. Also, a TBLASTN search of the assembled genome of *C. sublineolum* (Rech and Thon, unpublished data) showed evidence of a CPLS in this species.

I considered the possibility that the CPLSs may in fact belong to contaminating DNA samples in the genome sequencing projects. To determine whether the CPLSs could have been contamination, I examined the position of GLRG_05578 in the genome assembly of *C. graminicola* and the genes in its vicinity. Gene GLRG_05578 is located on supercontig 1.19 in contig 122 of the *C. graminicola* genome project (BioProject: PRJNA37879). Contig 122 is 192 Kb in length and has 64 predicted genes. A BLAST search of the flanking genes (GLRG_05577 and GLRG_05579) revealed that the most similar sequences in GenBank are from other fungi (**Figure 21 and Table 11**). From this result I conclude that contig 122 is, in fact, from the genome of *C. graminicola*. If GLRG_05578 is from contamination, then the contaminating sequence would have to have been aligned and assembled into the fungal genomic sequences during genome assembly. Since transposable elements (TEs) frequently cause misassemblies, I determined whether there are TEs flanking GLRG_05578. The closest annotated TE is located 7 Kb downstream of the gene (data not shown) and is unlikely to have caused misassembly GLRG_05578. Furthermore, the CPLSs are found in the genome sequence of four additional species of *Colletotrichum*, all of which were sequenced by different research groups at different institutions. It is unlikely that the same contaminating sequence would be encountered in all of the genome sequencing projects.

Figure 21. Diagram of contig 122 from the *Collotetrichum graminicola* M 1001 sequencing project showing the genes flanking CPLS GLRG_05578.

Table 11. Best BLASTp hits of the three genes shown in the Figure 21

acc. number	e-value	Species name	Kingdom
		GLRG_05577	
EKV09092	0	*Penicillium digitatum* Pd1	Fungi
XP_002378970	0	*Aspergillus flavus* NRRL3357	Fungi
XP_003043056	0	*Nectria haematococca* mpVI 77-13-4	Fungi
		GLRG_05578	
XP_003608462	0	*Medicago truncatula*	Viridiplantae
XP_002865008	0	*Arabidopsis lyrata* subsp. lyrata	Viridiplantae
XP_003550022	0	*Glycine max*	Viridiplantae
		GLRG_05579	
EIN04521	3,37E-52	*Punctularia strigosozonata* HHB-11173 SS5	Fungi
CCA72850	2,08E-51	*Piriformospora indica* DSM 11827	Fungi
XP_001838268	6,83E-51	*Coprinopsis cinerea* okayama7#130	Fungi

The BLAST searches of the CPLSs to the GenBank nr database revealed that the CPLSs are most similar to plant proteins with the most similar plant BLAST hit having 51.2% identity while the most similar bacterial, archeal and fungal hits were 33.3%, 19.8%, and 24.1% identical respectively. The global multiple sequence alignment between the CPLSs and plant subtilisins reveals that the CPLSs have between 40% and 50% identity to their plant counterparts. In general, subtilisins belonging to the same family have conserved residues at the catalytic site, Asp, His, Ser in all organisms. The plant-like subtilisins identified in *Colletotrichum* spp. also show conserved residues at the catalytic sites when compared to their plant counterparts. In contrast, these same regions were less conserved in other subtilisins from bacterial or fungal origin (**Figure 22**).

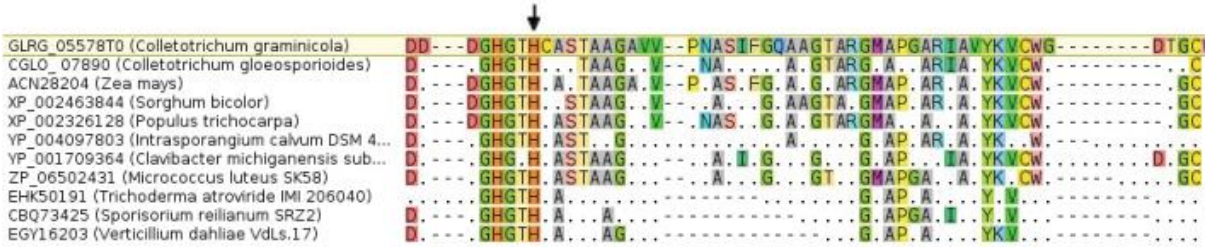


Figure 22. Representative portion of a multiple sequence alignment of CPLSs and subtilisins from plants, bacteria and fungi. The three best BLAST hits to GLRG_05578 from each taxonomic group were used to create the alignment. Amino acid disagreements to GLRG_05578 are represented by dots. Gaps are represented with a dash symbol. The arrow over the alignment indicates the position of the conserved histidine residue of the catalytic site of subtilisins.

Phylogenetic analysis

I constructed phylogenetic trees to test the hypothesis that the CPLSs are derived from plants by HGT. The S8A subtilisins are abundant in all of the kingdoms of life. For that reason I selected a subset of the most similar sequences to our candidates from Bacteria, Archaea, Metazoa, Fungi and Viridiplantae to reconstruct the phylogenetic tree (**Figure 23**).

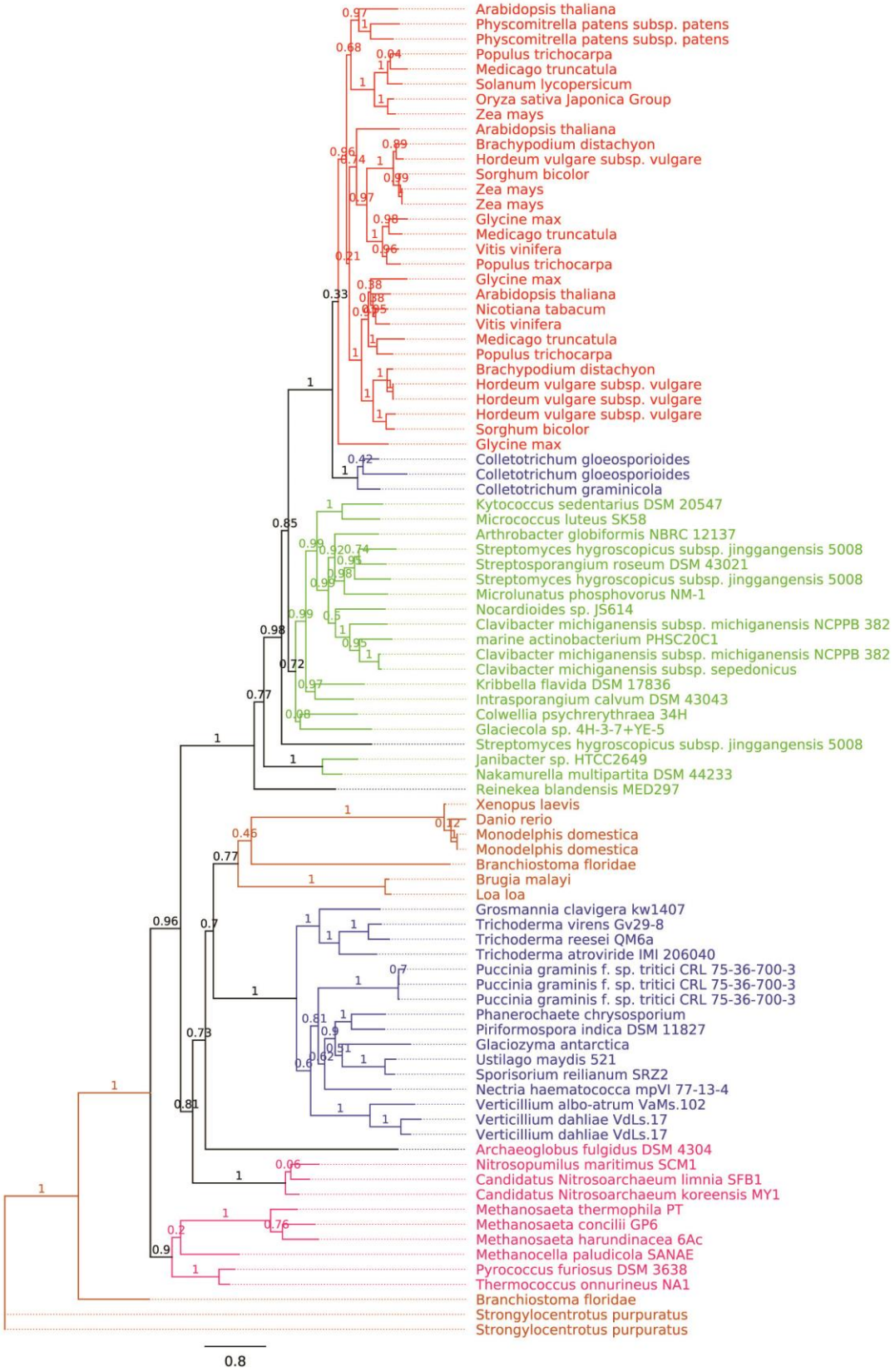


Figure 23. Phylogenetic tree of the protein sequences most similar to GLRG_05578 (of *C. graminicola)* from various kingdoms. Fungi are colored in blue, Viridiplantae in red, Bacteria in green, Archaea in pink and Metazoa in light brown. Internal nodes are labeled with SH-like

support values.

This tree shows that the CPLSs share a common lineage with subtilisins from plants while the remaining *Colletotrichum* subtilisins share a common lineage with other fungal subtilisins. To confirm this, all of the S8A subtilisins from *C. graminicola, C. higginsianum* and *Zea mays* were identified using the MEROPS server (http://merops.sanger.ac.uk/cgi-bin/batch_blast). The S8A protein sequences were aligned with MAFFT (Katoh et al. 2002) and the alignment was manually edited (removing sites with high percentages of gaps) using Geneious 5.5.7 (Drummond et al. 2011). A maximum likelihood tree was reconstructed with PhyML (Guindon and Gascuel 2003), and the tree was tested by performing a non-parametric bootstrap analysis with 100 replications. The maximun likelihood tree shows the separation of the maize subtilisins and *Colletotrichum* subtilisins into two clades with the CPLSs within the clade of maize subtilisins (**Figure 24**). To further confirm these results, I tested the tree with several topology tests by constructing a new tree that forced the monophyly of the fungal subtilisins together with the CPLSs. MrBayes (Ronquist and Huelsenbeck 2003) was used to reconstruct and constrain the tree. TREE-PUZZLE (Schmidt et al. 2002) was used to perform the ELW (Expected Likelihood Weights) topology test and CONSEL (Shimodaira and Hasegawa 2001) was used to perform the AU (Approximately Unbiased) and SH (Shimodaira and Hasegawa) topology tests. In all of the topology tests the unconstrained tree was not rejected and the monophyletic fungal tree was rejected at the 95% confidence level. These results support the hypothesis that the CPLSs share a common ancestry with the plant subtilisins that is distinct from the other subtilisins in the *Colletotrichum* genomes.

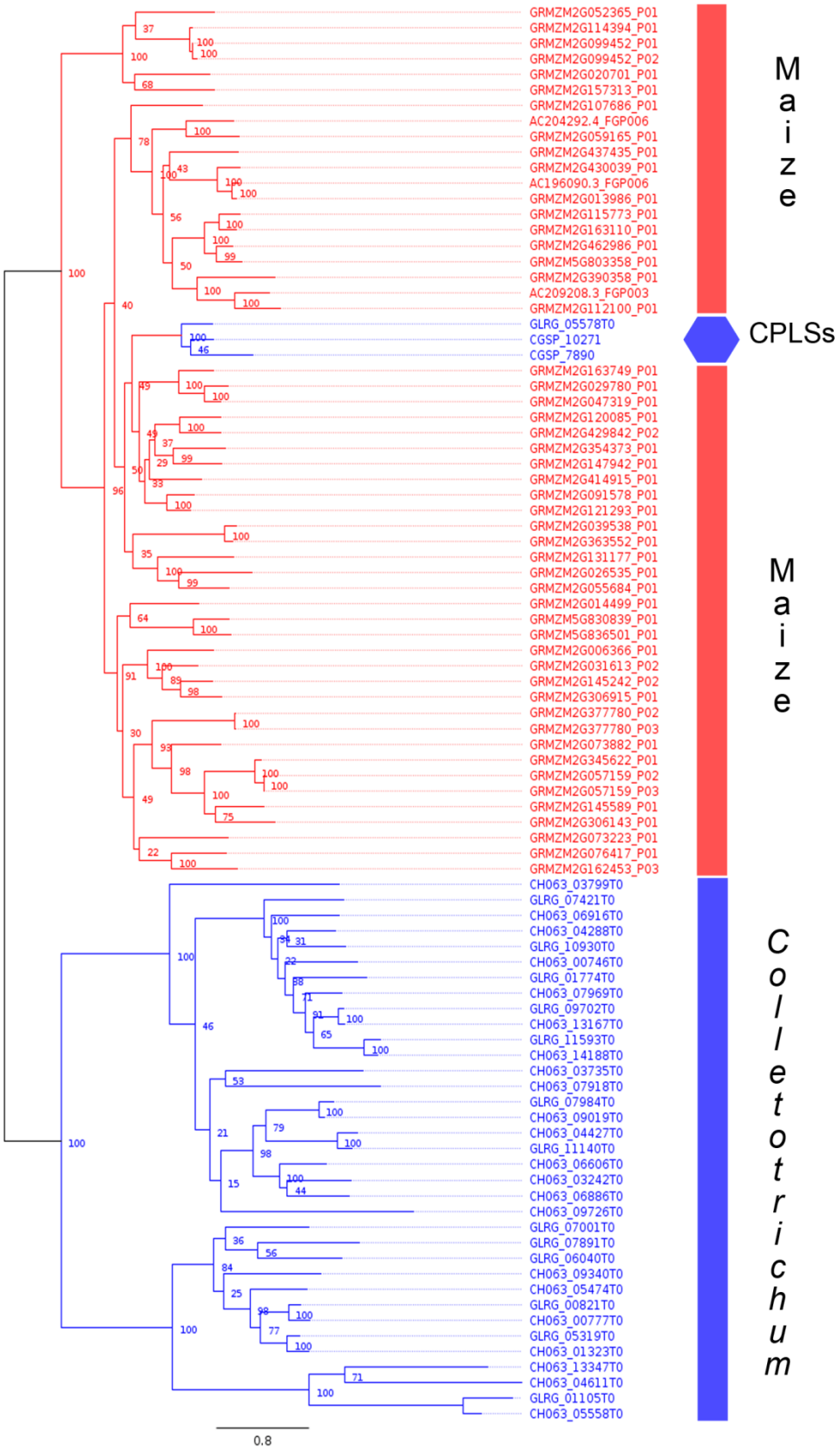


Figure 24. Phylogenetic tree of subtilisins of *Zea mays* (GRMZM2G or AC) colored in red and *Colletotrichum graminicola* (sequence IDs beginning with GLRG) and *C. higginsianum* (CH) colored in blue. Internal nodes are labeled with percentage of bootstrap support.

Subtilisins from bacteria were the best BLAST hits to the CPLSs after those from plants. To determine the relationship between CPLSs and bacterial subtilisins, a second tree was reconstructed with all S8A subtilisins of *C. graminicola, C. higginsianum* and *Zea mays* plus some representatives of the most similar bacterial subtilisins. The resultant tree shows three well defined clades: maize sequences including CPLSs, *Colletotrichum* sequences and bacterial sequences (**Figure 25**). This tree also shows a close clustering between bacterial subtilisins with maize sequences. *Colletotrichum* subtilisins (excluding plant-like subtilisins) form a well-defined clade but this is more distant to the other two.

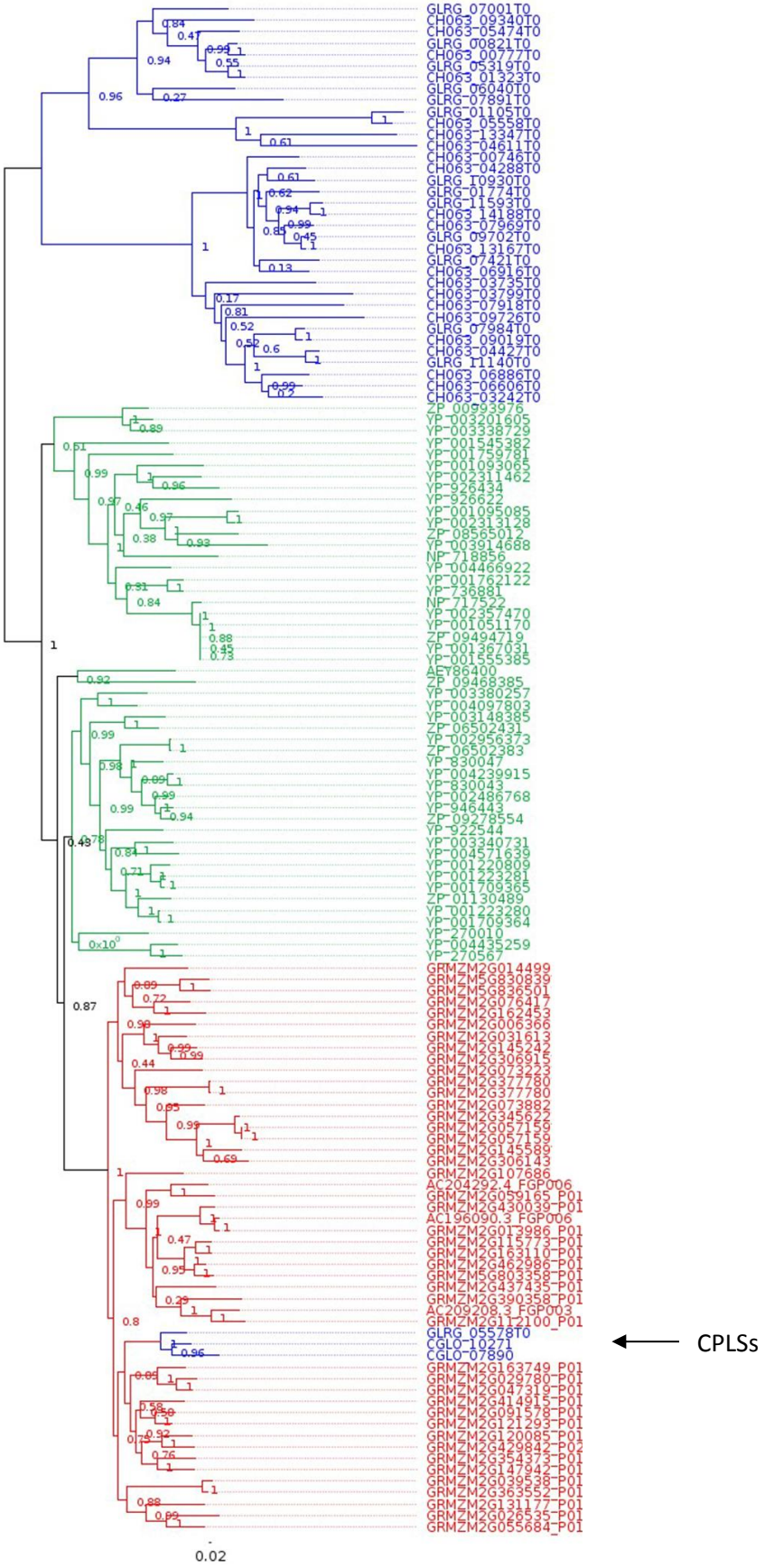

GLRG_07001T0
CH063_09340T0
CH063_05474T0
GLRG_00821T0
CH063_00777T0
GLRG_05319T0
CH063_01323T0
GLRG_06040T0
GLRG_07891T0
GLRG_01105T0
CH063_05558T0
CH063_13347T0
CH063_04611T0
CH063_00746T0
CH063_04288T0
GLRG_I0930T0
GLRG_01774T0
GLRG_11593T0
CH063_14188T0
CH063_07969T0
GLRG_09702T0
CH063_13167T0
GLRG_07421T0
CH063_06916T0
CH063_03735T0
CH063_03799T0
CH063_07918T0
CH063_09726T0
GLRG_07984T0
CH063_09019T0
CH063_04427T0
GLRG_11140T0
CH063_06886T0
CH063_06606T0
CH063_03242T0
ZP_09993976
YP_003201605
YP_003338729
YP_001545382
YP_001759781
YP_001093065
YP_002311462
YP_926434
YP_926622
YP_001095085
YP_002313128
ZP_08565012
YP_003914688
NP_718856
YP_004466922
YP_001762122
YP_736881
NP_717522
YP_002257470
YP_001051170
ZP_09494719
YP_001367031
YP_001555385
AEY86400
ZP_09468385
YP_003380257
YP_004097803
YP_003148385
ZP_06502431
YP_002956373
ZP_06502383
YP_830047
YP_004239915
YP_830043
YP_002486768
YP_946443
ZP_09278554
YP_922544
YP_003340731
YP_004571639
YP_001220809
YP_001523281
YP_001709365
ZP_01130489
YP_001223280
YP_001709364
YP_270010
YP_004435259
YP_270567
GRMZM2G014499
GRMZM5G830839
GRMZM5G836501
GRMZM2G076417
GRMZM2G162453
GRMZM2G006366
GRMZM2G031613
GRMZM2G145242
GRMZM2G306915
GRMZM2G073223
GRMZM2G377780
GRMZM2G377780
GRMZM2G073882
GRMZM2G345622
GRMZM2G057159
GRMZM2G057159
GRMZM2G145589
GRMZM2G306143
GRMZM2G107686
AC204292.4_FGP006
GRMZM2G059165_P01
GRMZM2G430039_P01
AC196090.3_FGP006
GRMZM2G013986_P01
GRMZM2G115773_P01
GRMZM2G163110_P01
GRMZM2G462986_P01
GRMZM5G803358_P01
GRMZM2G437435_P01
GRMZM2G390358_P01
AC209208.3_FGP003
GRMZM2G172100_P01
GLRG_05578T0
CGLO_10271
CGLO_07890
GRMZM2G163749_P01
GRMZM2G029780_P01
GRMZM2G047319_P01
GRMZM2G414915_P01
GRMZM2G091578_P01
GRMZM2G121293_P01
GRMZM2G120085_P01
GRMZM2G429847_P01
GRMZM2G354373_P01
GRMZM2G147942_P01
GRMZM2G039538_P01
GRMZM2G363552_P01
GRMZM2G131177_P01
GRMZM2G026535_P01
GRMZM2G055684_P01
0.02
CPLSs

Figure 25. Tree of all subtilisins found in maize and two species of *Colletotrichum*, including representatives from bacterial phyla. The accession numbers of genome project are shown for maize and *Colletotrichum* sequences and GenBank accession number for bacteria. *Zea mays* (GRMZM2G or AC) colored in red, two species of *Colletotrichum* (*C. graminicola* GLRG and *C. higginsianum* CH) are colored in blue and subtilisins from bacteria in green. Internal nodes are labeled with SH-like support values.

Plant-like subtilisins are absent in all other species of fungi, including *Verticillium* spp. *Verticillium* is estimated to have diverged from *Colletotrichum* approximately 150 million years ago in agree to (O'Connell et al. 2012) or from 160 to 75 million years ago in concordance with the estimation of Chapter I. I constructed a plant subtilisin phylogeny that included the CPLSs, to better understand when, during the evolution of plants, the CPLSs where likely to have been transferred to *Colletotrichum* and to determine if this date occurred after the divergence of *Verticillium* and *Colletotrichum*. The plant S8A subtilisins show evidence for several duplication events but with a considerable level of conservation. I selected the most similar plant subtilisin sequences to CPLSs available in GenBank (Benson et al. 2011) to construct a tree that shows the position of the CPLSs in the plant subtilisins group. The sequences were aligned with MAFFT and edited manually (deleting highly divergent domains). I also prepared alignments by editing the MAFFT alignment with trimAl (Capella-Gutiérrez et al. 2009) and Gblocks (Talavera and Castresana 2007) and constructed four phylogentic trees using PhyML (**Figure 26**). The tree constructed using the manually edited alignment was the only tree that had the same topology as the tree constructed form the unedited alignment. In addition, the bootstrap support values in the tree constructed from the manually edited alignment were higher than the values from the unedited alignment.

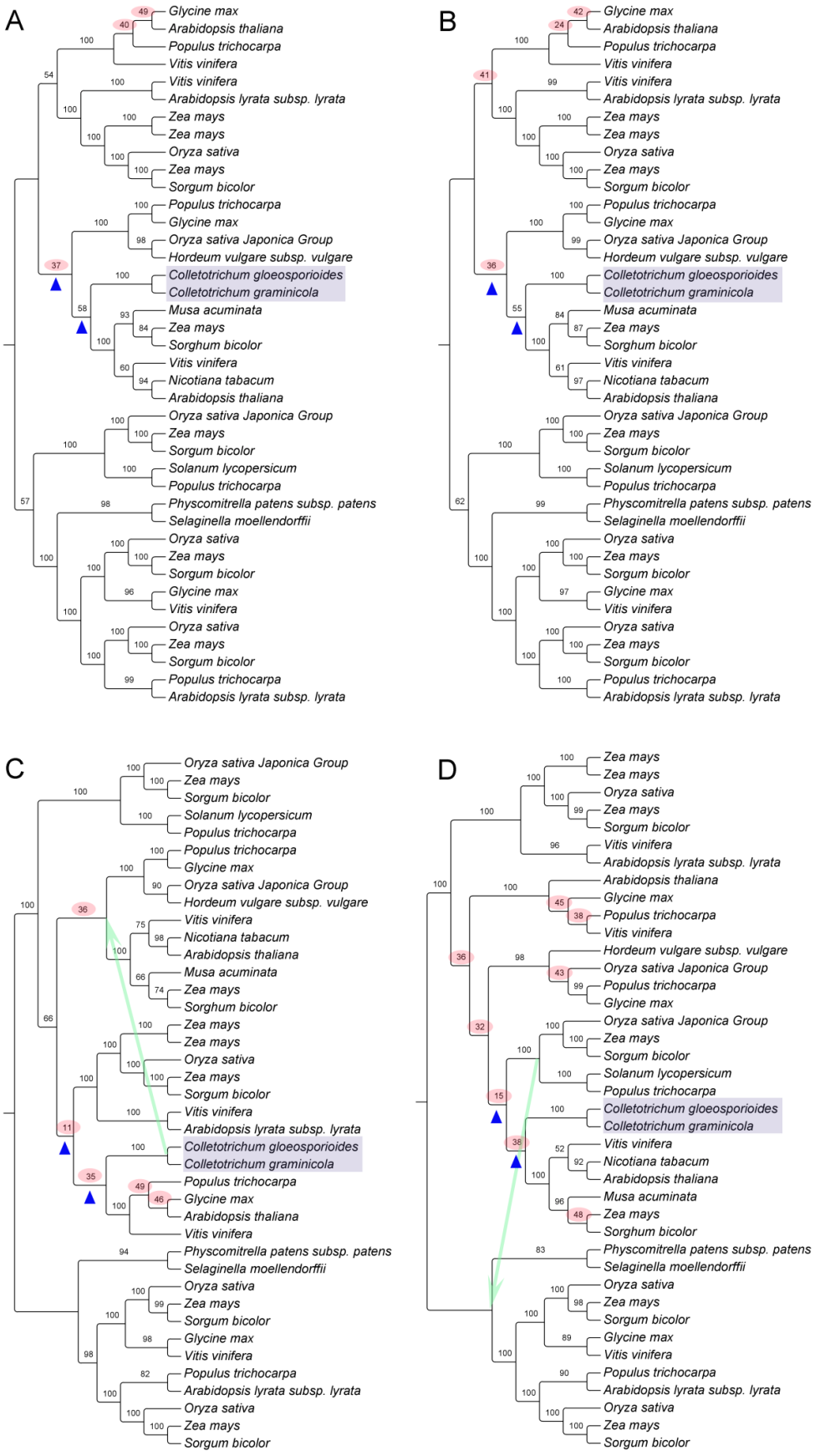


Figure 26. Trees reconstructed in PhyML (100 bootstrap) with the same dataset using four

different alignment editing methods. **A**) Manually edited. **B**) Unedited. **C**) Automatically edited with the trimAl program. **D**) Automatically edited with the Gblocks program. The red circles highlight the bootstrap support values lower than 50%. Blue triangles indicate the support values for the key nodes representing the insertion of CPLSs in the plant lineage. The green arrows in C and D trees show the differences in the position of some clusters taking as a reference the topology of A and B trees. The original alignment was performed with MAFFT.

Therefore, I selected the manually edited alignment for further analysis. Next, I constructed phylogenetic trees using several different methods. A Bayesian tree, supported by posterior probability index, was constructed using MrBayes (Ronquist and Huelsenbeck 2003). A maximum likelihood tree supported with a fast bootstrap approximation was constructed using RAxML (Stamatakis 2006). A maximum likelihood tree was constructed using PhyML (Guindon and Gascuel 2003) with full non-parametric bootstrap and SH-branch tests (Anisimova and Gascuel 2006) to verify the position of branches inside the tree (**Figure 27**). *Colletotrichum* plant-like subtilisins were placed in the same position of the tree in all methods tested. In these trees, the CPLSs are in a position that is ancestral to a lineage that gives rise to monocot and dicot lineages, suggesting that the CPLS were transferred to *Colletotrichum* some time before the divergence of monocots from the angiosperms approx. 134 million years ago (Myr) (Bell et al. 2005) to 200 Myr (Wolfe et al. 1989) with the most recent estimates of 155 Myr to 145 Myr (Chaw et al. 2004; Leebens-Mack et al. 2005; Smith et al. 2010). According to these dates, the HGT event would have occurred approximately 150 to 155 Myr, just before the monocot divergence and just after the *Verticillium-Colletotrichum* divergence.

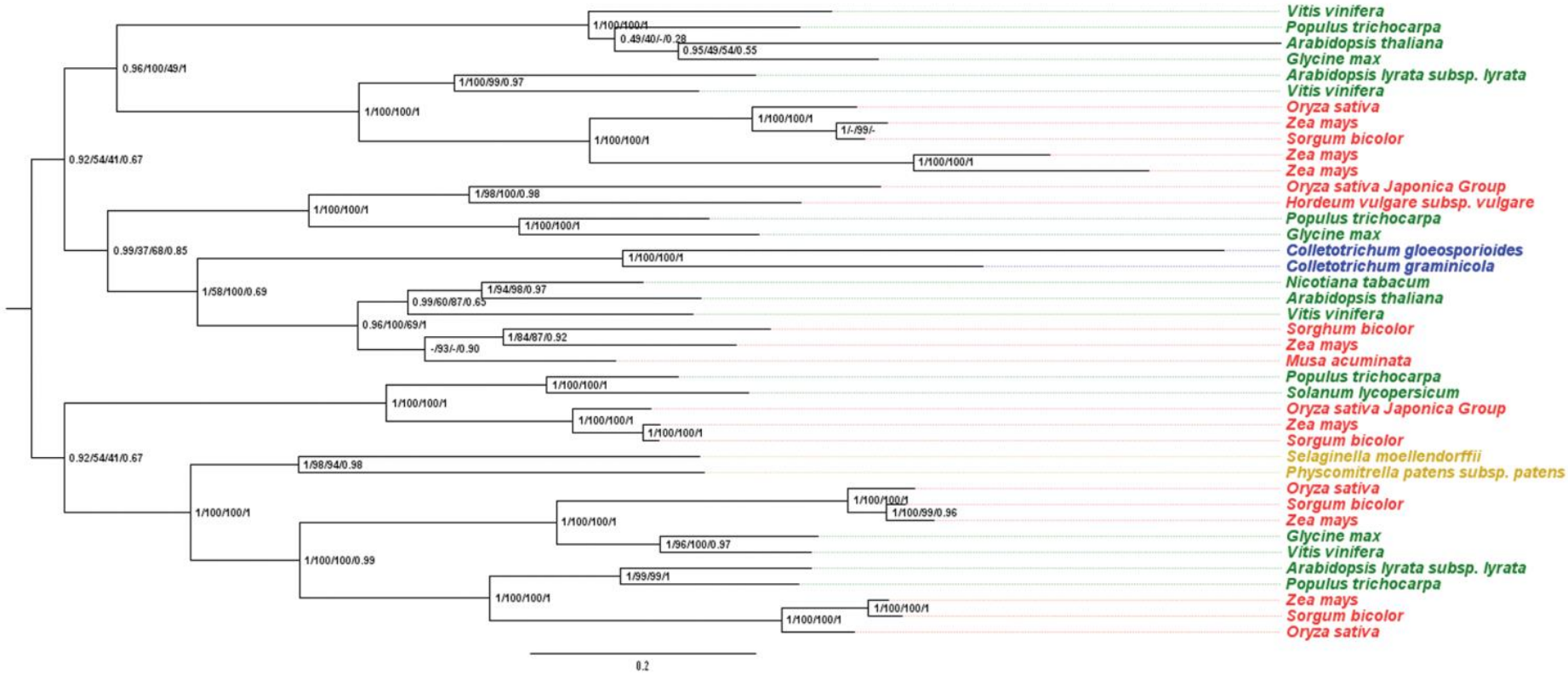


Figure 27. Location of the CPLSs in the phylogenetic tree of plant subtilisins. The tree was rooted on one of the multiple duplication events in this family. The colors represent the taxonomic groups: red for monocots, green for dicots, yellow for embryophytes and blue for *Colletotrichum*. The numbers at each node represent the posterior probability/percentage of bootstrap in PhyML/percentage of bootstrap in RAxML/SH-branch test.

In concordance with data the CPLSs should be present in all of the extant species of *Colletotrichum* and absent in species outside the genus. I previously showed that the CPLSs are present in the genome sequences of three species of *Colletotrichum* but absent in all other fungi. Recently, Gan et al. (Gan et al. 2013) reported the presence of subtilisins, likely resulting from an HGT event, in the genomes of *C. orbiculare* (Cob 12233, Cob 06327) and *C. gloeosporioides* (CGGC5 2662), adding further support to my hypothesis. In addition, genome sequences for several more species of *Colletotrichum* have recently become available, enabling me to further validate the hypothesis. A BLAST search of the GenBank NR database (Benson et al. 2011) revealed the presence of CPLS homologs in the genome of *C. gloeosporioides* Nara gc5 (GenBank accession numbers: ELA37268 and ELA38265). I also performed BLAST searches against the proteomes of two additional species of *Colletotrichum* (*C. gloeosporioides* 23 and *C. fiorinae* MH 18, www.jgi.doe.gov) and found that both species have CPLS homologs. A phylogenetic analysis (data not shown) shows that all of the putative CPLS homologs share common ancestry with the CPLSs identified previously. This idea is consistent with our hypothesis of a plant protein that was transferred to an ancestor of the genus *Colletotrichum*.

Domain content of CPLSs

All of the plant S8A subtilisins that I analyzed contain three domains, the inhibitor I9 domain (PF05922), the PA domain (PF02225) and the peptidase S8 domain (PF00082). The same three domains were observed in CPLSs (**Figure 29a**). Other domains, such as DUFF1034 (PF06280) and Pex16 (PF08610) are present in some plant subtilisins, but are absent in CPLSs. The peptidase S8 domain is always present in fungal subtilisins and is accompanied by either domain PA or inhibitor I9 but rarely with PA and I9 at the same time. Thus, the domain arrangement in CPLSs is more similar to subtilisins from plants and bacteria than to their fungal counterparts. Additionally, in CPLSs a signal peptide and a cleavage site were predicted by WoLF PSORT (Horton et al. 2007) and SignalP (Petersen et al. 2011). This finding suggests that, like many subtilisins, the proteins are secreted.

Using the classification system of PANTHER (Mi et al. 2009), I determined that all proteins related to CPLSs are included in the same sub-family (PTHR10795:SF17), and no other fungal protein was included into this sub-family. Only CPLSs and subtilisins from plants and

bacteria are assigned to PTHR10795:SF17. With all the sequences identified as members of PTHR10795:SF17 a profile hidden Markov model (HMM) was constructed with HMMER (Finn et al. 2011). The profile HMM was used to search the NR database with the tool hmmsearch of the HMMER web server (www.hmmer.janelia.org/search/hmmsearch). This search only resulted in hits from plants, followed by the bacterial phyla Actinobacteria, Gammaproteobacteria and Chloroflexi. These results demonstrate the resemblance of CPLSs to plant proteins.

CPLS structure modeling

I hypothesized that the CPLSs might still share common structural features with their plant counterparts. Common structure may be used to imply common function (Nembaware et al. 2004). Recently, the plant subtilisin SBT3 (PDB 3I6S) from tomato was crystallized (Ottmann et al. 2009) and this was used to predict the structure of different S8A subtilisins of Arabidopsis and the pathogenesis related protein P69B of tomato (Rose et al. 2010). Protein SBT3 shares 46,7% identical sites with the GLRG_05578 protein of *C. graminicola*. I reconstructed the tertiary structure of GLRG_05578 (**Figure 28 and 29b**) using the Phyre 2 server (Kelley and Sternberg 2009) and then used the resulting structure to perform a search using the Dali server (Holm and Rosenstrom 2010). The Dali server returned a match to the tomato SBT3 structure with a root-mean-square deviation (RMSD) of 0.7 Å and a Z-score of 65.9 indicating that the structures are highly similar. I also aligned the two structures using PhyMol (Schrödinger 2010) resulting in a series of RMSD values ranging from a maximum RMSD of 11.68, to a minimum of 0.02 (**Figure 28**).

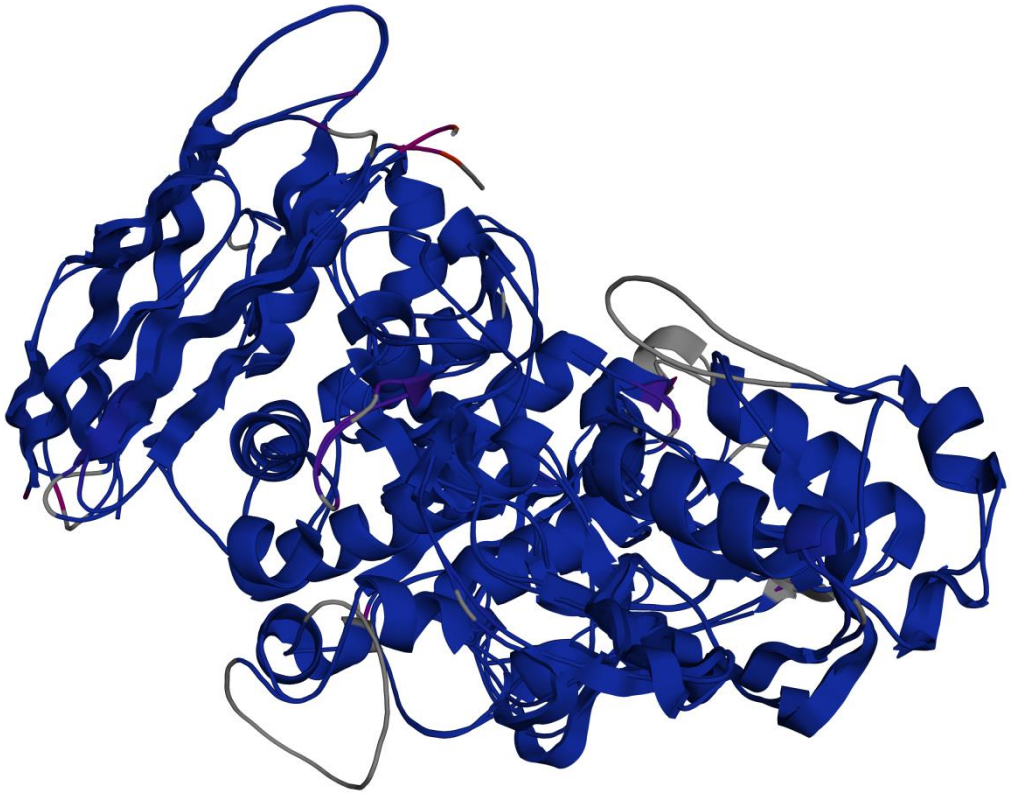


Figure 28. Structural alignment between SBT3 of maize and GLRG_05578 of *C. graminicola*. The color spectrum represents the values of pairwise RMSD, with blue specifying minimum values and red indicating maximum values. Gray sections are not aligned.

The most relevant regions and sites of SBT3 protein described in (Cedzich et al. 2009; Ottmann et al. 2009; Rose et al. 2010) were also observed in GLRG_05578 (**Table 12, Figure 29c**). One of those is the beta hairpin, described in (Ottmann et al. 2009) as an essential structure for the homo-dimerization of SBT3. The GLRG_05578 structure is similar, but with the beta sheet folding not well defined. Three defined regions Ca-1 (Gly-225-Gly-243), Ca-2 (Lys-498) and Ca-3 (Cys-170-Cys-181) were reported as potential responsible sites for the stabilization of SBT3 to high temperatures and alkalinity (only the relevant residues for each site were named in parenthesis). These sites could act in replacement of Ca^{2+}, which is the element commonly present in subtilisins to perform the stabilization (Ottmann et al. 2009).

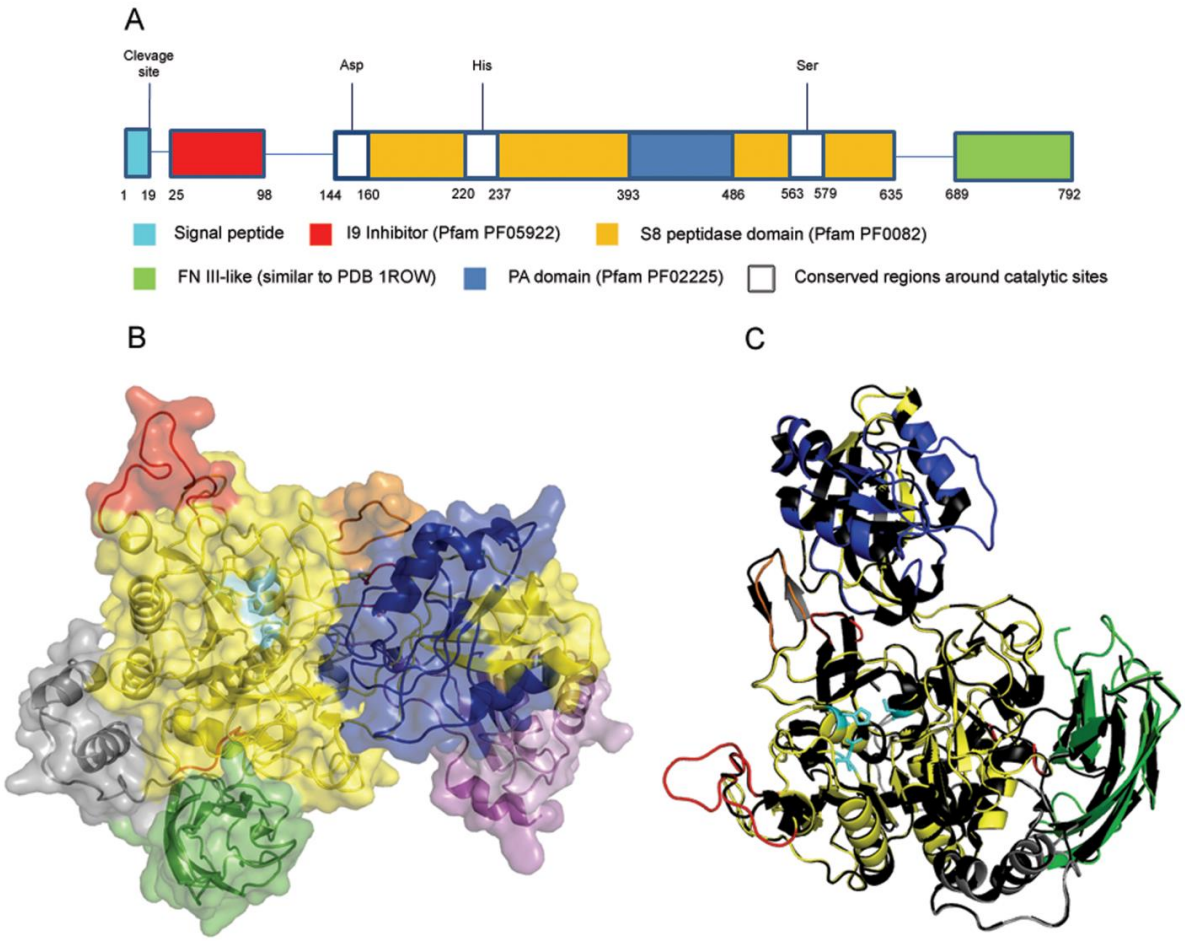


Figure 29. **A**) Schematic view of protein domains found in *C. graminicola* subtilisin GLRG_05578. **B**) 3D surface view of the protein GLRG_05578. The peptidase s8 domain is colored in yellow, PA domain in blue and Fn III-like domain in green. The ß-hairpin like domain is colored in orange, the residues of the catalytic site are colored in cyan and the putative sites of Ca^{+} replacement are in red. The signal peptide and I9 inhibitor are in pink and violet respectively. The gray residues have not been assigned any domain **C**) Alignment between mature forms of subtilisin SBT3 of tomato and GLRG_05578 of *C. graminicola*. The tomato subtilisin is in black and the *C. graminicola* subtilisin is colored as in (B).

Table 12. Comparison of domains and sites between SBT3 (PDB 3I6S) and the predicted tertiary structure of GLRG_05578

Region	**SBT3**		**GLRG_05578**	
	start	end	start	end
Domain peptidase S8	113	599	141	603
PA domain	363	457	393	486
Beta hairpin (like)	519	528	547	558
Domain Fn III-like	600	761	689	792
Hypothetical site of calcium stability (Ca-1)*	225	243	237	254
Hypothetical site of calcium stability (Ca-3)*	170	181	175	186

Relevant sites	**Residue**	**Residue**
Catalytic site	Asp 144	Asp 149
Catalytic site	His 215	His 227
Catalytic site	Ser 538	Ser 566
Hypothetical site of calcium stability (Ca-2)*	Lys 498	Lys 527

*Only relevant residues were named

The lysine 498 in SBT3 is conserved in GLRG_05578 (Lys 527), sharing the same relative position in the peptidase S8 domain (**Figure 30**). Regions Ca-1 and Ca-3 of SBT3 present alignment similarities with GLRG_05578 (RMSD 0.311 for Ca-1 region and RMSD 0.258 for Ca-3 region). But minor shape differences could be observed (**Figure 31a and 31b**). The catalytic triads (Asp 144/ His 215/ Ser 538 in SBT3 and Asp 149/ His 227/ Ser 566 in GLRG_05578) are placed in the same position in both structures. Finally, one of the residues apparently responsible for the union of PA domains in the dimerization of SBT3 (Arg 418) is not present in *C. graminicola* subtilisin. However, GLRG_05578 has some of the elements for the interaction with another monomer, like the hairpin and the PA domain (structures observed in the SBT3 dimer conformation).

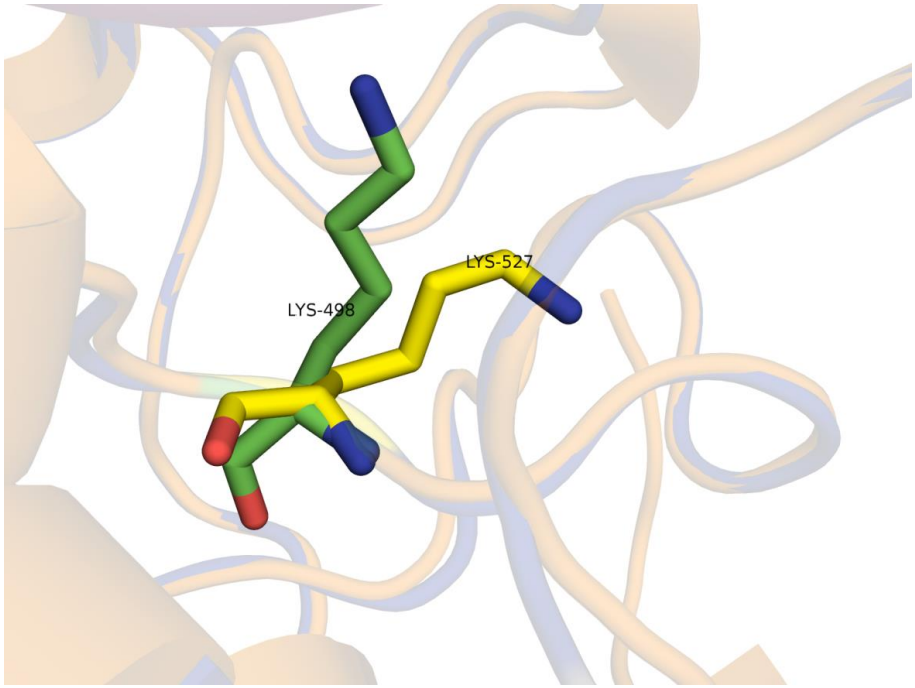


Figure 30.Structural alignment between residue Lys-498 of SBT3 (a putative stabilization site of the proteins to high temperature and alkalinity) and residue Lys-527 of CPLS GLRG_05578. Carbons atoms are represented in green and yellow respectively, oxygen atoms in red and nitrogen atoms in blue. Transparent cartoon representation of nearby structures is colored in blue for GLRG_05578 and in orange for SBT3.

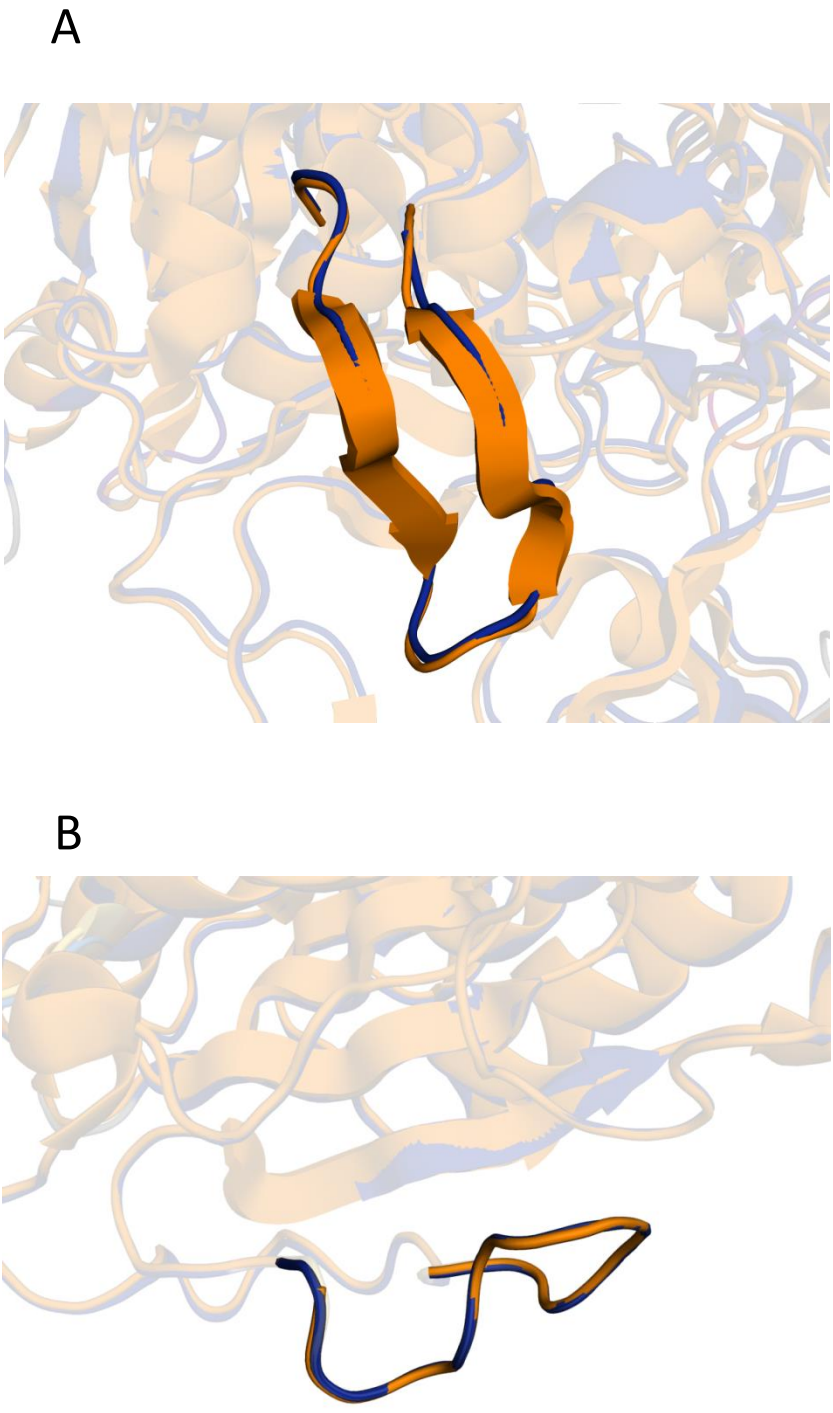


Figure 31. A) Structural alignment of SBT3 and GLRG_5578 in the Ca-1 region. **B)** Structural alignment of SBT3 and GLRG_5578 in the Ca-2 region. SBT3 protein is colored in orange and GLRG_5578 protein is colored in blue.

Another 3D structure of a *C. graminicola* subtilisin was reconstructed. Protein GLRG_07421 is the most similar subtilisin to GLRG_05578 in the *C. graminicola* proteome with 24% amino acid identity. The differences between these proteins are evident in the 3D alignment (**Figure 32**). Only the surrounding regions of the catalytic sites show resemblances. These results show the uniqueness of GLRG_05578 in the *C. graminicola* subtilisin arsenal.

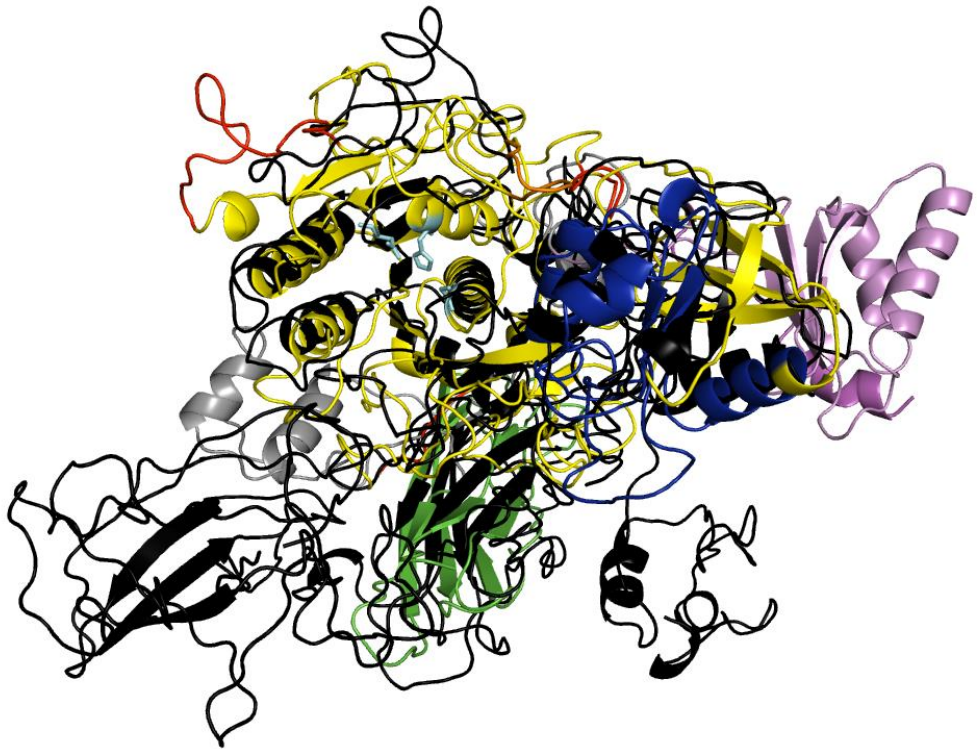


Figure 32. 3D alignment between the CPLS GLRG_05578 and GLRG_07421 of *C. graminicola.* The alignment emphasizes the structural differences between these two *C. graminicola* subtilisins. For GLRG_05578 peptidase s8 domain is colored in yellow, PA domain in blue and Fn III-like in green. The ß-hairpin like is colored in orange, the residues of the catalytic site are colored in cyan (and highlighted in stick view) and the putative sites of Ca^{+} replacement are in red. The signal peptide and I9 inhibitor are in pink and violet respectively. The gray parts have not assigned any characteristic. The structure of GLRG_07421 is entirely in black.

Recently, the crystal structure of the subtilisin Cucumisin of *Cucumis melo* was published (Murayama et al. 2012). With this new structure available, I recalculated the 3-D model of GLRG_05578 (the CPLS of *Colletotrichum graminicola*) using the Phyre 2 server (Kelley and Sternberg 2009). The new structural model has only 57 residues predicted *ab initio* (highly unreliable sites) instead of 73 as reported in the old model based on SBT3. The new GLRG_05578 structural model was aligned to the structural models of SBT3 and Cucumisin to determine the similarities and the differences with the previous model. The structural

alignment of the previous model of GLRG_05578 and SBT3 has a maximum RMSD (root-mean-square-deviation) of 11.68 and a minimum of 0.02. The new model has a highest RMSD value of 12.65 and the lowest value of 0.03. These results show that the new model is less similar to SBT3 than the one showed previously. However, the new model is more similar to the SBT3 subtilisin than to the Cucumisin (**Figure 33**) (maximum RMSD of 17.59 and minimum RMSD of 3.09 with the old model and maximum RMSD of 15.12 and minimum RMSD of 1.84 with the new model). The new model reveals differences in the tertiary protein structures of the PA domain, β-hairpin arrangement and one of the putative Ca^{+2} replacement sites between CPLS and tomato subtilisin. These sites are important in dimerization and stabilization of tomato SBT3 (Ottmann et al. 2009) and the differences with CPLS could imply differences in the quaternary structure (conformation of dimers) and/or the specificity of interaction with other proteins.

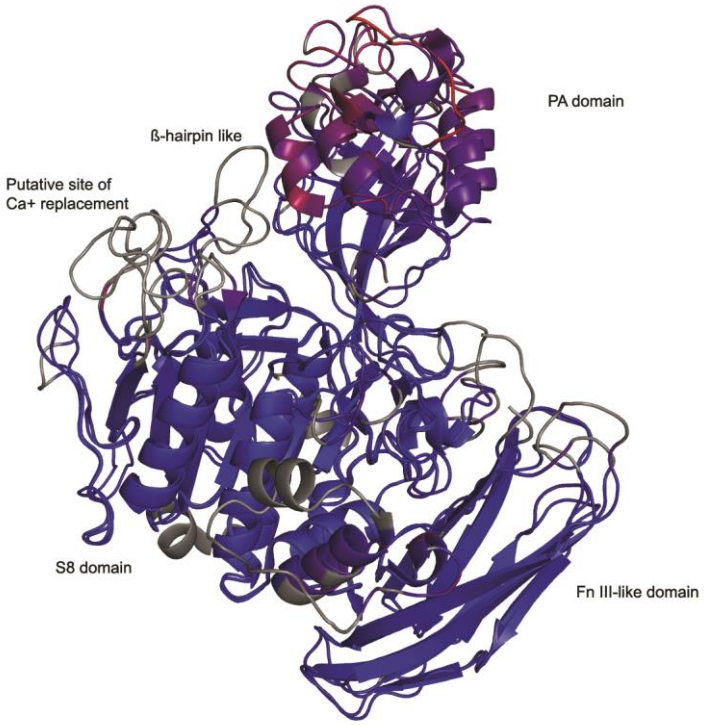


Figure 33. Structural alignment of the new model predicted for the tertiary conformation of CPLS GLRG_05578 of *Colletotrichum graminicola* and one monomer of the crystallized subtilisin Cucumisin (3VTA) of *Cucumis melo*. The color spectrum represents the values of pairwise RMSD, with blue specifying minimum values and red indicating maximum values. Gray sections are not aligned. Labels indicate the relevant domains and sites of the proteins conserved in subtilisin SBT3 of tomato in Ottmann et al. (2009)

CPLS Gene Expression

Subtilisin is a large family in plants, and in the case of maize I identified 53 members of the subtilisin S8 family. A recent study described the expression pattern of a set of subtilisin S8A genes, also called PR-7, in maize leaves infected with *Ustilago maydis* (Doehlemann et al. 2008). To further investigate the participation of subtilisin-encoding genes during anthracnose development, I followed the expression of ten putative PR-7 genes in the maize genome as well as GLRG_05578 during infection by *C. graminicola*. The selection of the ten genes was based on the identity with CPLSs, the identity with P69s (a well-known group of PR-7 genes) in tomato, (van Loon et al. 1994; Jorda et al. 1999) and the expression pattern after *U. maydis* infection. The expression assays revealed that the CPLS GLRG_05578 of *C. graminicola* is induced at late stages of biotrophic infection (48 hours post-infection) and continues to be up-regulated 72 hours post-infection (hpi) (**Figure 34a**), suggesting the importance of the protein product during the transition from biotrophic to necrotrophic stages of the fungal infection.

In the case of the maize genes tested a heterogenic behavior was detected. For instance no amplification product was detected for five of the sequences (GRMZM2G430039, GRMZM2G121293, GRMZM2G354373, GRMZM2G120085, AC196090.3_FGP006) suggesting that the protein product of these genes is not needed during infection by *C. graminicola* (**Figure 34b**). In contrast, GRMZM2G013986 displays constitutive expression with no changes between mock-inoculated leaves and leaves from plants infected with *C. graminicola*.

Gene GRMZM2G073223 displays a similar expression pattern as other maize PR genes such as PR-1, PR-4 and PR-5 during anthracnose development (Vargas et al. 2012a) (**Figure 34b**). The expression results also revealed interesting expression profiles for two maize genes GRMZM2G091578 and GRMZM2G414915, which display increased levels of expression at early stages of infection that decreased with the progress of the disease (**Figure 34b**). Among the putative subtilisin-encoding genes from maize GRMZM2G091578 and GRMZM2G414915 are among the most similar to CPLS GLRG_05578 of *C. graminicola*. The fact that these maize genes are down-regulated at the time the fungal homolog is induced might suggest a compensation of the enzymatic activity where the fungus is hijacking the

plant subtilisins, interfering with the normal proteolytic activities in the host cells and the biochemical processes associated with specific forms of subtilisins.

While the RT-PCR experiments were designed to study the expression of the CPLS in *C. graminicola*, they also provide additional evidence that GLRG_05578 is not a product of foreign DNA contamination in the *C. graminicola* genome project. I detected the expression of GLRG_05578 in infected maize leaves (**Figure 34a**). A Primer-BLAST search (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) failed to detect any potential primer binding sites in the maize genome. In addition, I amplified the gene fragment from genomic DNA obtained from axenic culture of *C. graminicola* (**Figure 34a**).

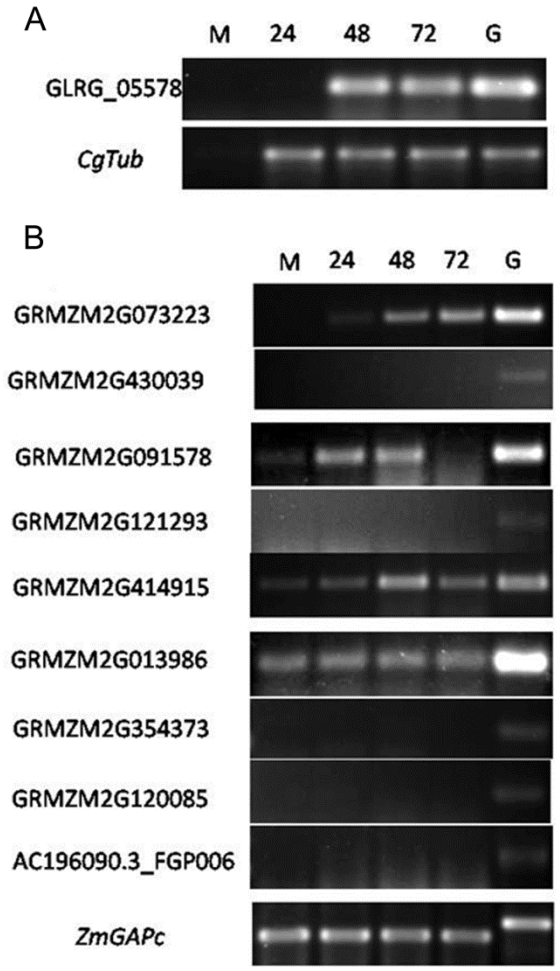


Figure 34. Gene expression during anthracnose development. Due to the low representation of fungal mRNA in the samples, semi-quantitative RT-PCR assays were conducted to test the expression of CPLS GLRG_05578 of *C. graminicola* and the selected maize putative subtilisins. The amount of total RNA used in each PCR reaction was adjusted to the amount needed to provide equal amplification levels of CgTub in all samples. PCR products were visualized after electrophoresis on 2% agarose gel and ethidium bromide staining. **A**) RT-

PCR products for GLRG_05578 and CgTub. **B**), RT-PCR products of nine genes encoding putative subtilisins in maize. *ZmGAPc* was amplified as an internal loading control. The number of cycles in PCR reactions was optimized to be in the linear amplification range of each gene. These assays were repeated two times with similar results. In both panels, the numbers over the lanes indicate the time-point at which RNA samples were taken. M indicates RNA samples from mock-inoculated leaves and G indicates genomic DNA.

Interactions of plant subtilisins could reveal the function of CPLSs

Although the protein targets of the CPLSs are unknown, analysis of the most similar plant proteins could reveal potential clues. A BLASTP search in the UniProt/SwissProt database shows that the most similar proteins to CPLSs are three subtilisins of *Arabidopsis thaliana* (O65351, O64495, Q9LLL8) and the Cucumisin of *Cucumis melo* (Q39547). Information about the interaction of the *Arabidopsis* subtilisins is available in the STRING database (Franceschini et al. 2013), but it is not available for Cucumisin. This database shows that O64495 is co-expressed with several proteins involved in the development of stomata, including AT1G63700, AT5G62230, AT1G80080 (Tricker et al. 2012), AT5G53210, AT3G06120 (Pillitteri et al. 2011) and others. The protein O65351 (ARA12) is predicted to interact with AT1G22300 and AT4G09000, two general regulatory factors (Swatek et al. 2011). Gene Q9LLL8 is co-expressed with AT4G35350 and AT1G20850 (Zhao et al. 2000) which have peptidase activity, cysteine-type peptidase activity, respectively. The interaction and co-expression of these proteins could help to explain the role of CPLSs in the host and help to plan future studies. Plant subtilisins with high percentage of sequence similarity to CPLSs are involved in varied, and sometimes unrelated biological process. The CPLS homologs in *Arabidopsis* are a good example of this phenomenon. Protein O65351 (ARA12) is essential for mucilage release from seed coats whereas O64495 is involved in stomatal development and distribution, and Q9LLL8 is involved in negative regulation of catalytic activity. The functions of the CPLSs are not clear yet, but they could be involved in the miss-regulation or mimicry of any of those processes. Alternatively, novel functions might also be expected for CPLS.

Discussion

Phylogenetic analysis, domain content and tertiary structure prediction allowed us to identify the presence of a plant-like member of the subtilisin S8A family in the genome of

Colletotrichum graminicola M1001 and *Colletotrichum gloeosporioides* Cg 14. I have also evidence of CPLSs in *C. higginsianum, C. acutatum, C. orbiculare, C. gloesoporioides* Nara gc5, *C. gloeosporioides* 23, *C. fiorinae* MH 18 and *C. sublineolum*. Several independent lines of evidence support the hypothesis that CPLSs are part of *Colletotrichum* genomes and not a product of contamination from plants or other foreign DNA. Most importantly, the presence of CPLSs in the genomes of nine different species of *Colletotrichum*, all of which were sequenced by different research groups in different laboratories using different samples and methodologies support the hypothesis that the CPLSs are not a result of contamination and are, in fact, components of the *Colletotrichum* spp. genomes.

The most similar sequences to the CPLSs are found in plants followed by bacteria. The phylogenetic reconstruction shows that CPLSs are within a clade of plant subtilisins. This suggests that the CPLSs originated in plants, and were not transferred vertically from fungi. On the other hand, bacterial S8A subtilisins (mainly from Actinobacteria, Cloroflexi and Gammaproteobacteria) were observed in BLAST searches as the most similar proteins after plants. Phylogenetic reconstructions place a monophyletic bacterial branch near to the plant-CPLS lineage (**Figures 24 and 26**). This reveals a complex evolutionary history behind these proteins. In fact, CPLSs, plant subtilisins and some bacterial subtilisins (of the three phyla named earlier) are recognized as members of PTHR10795:SF17 subfamily in the PANTHER database (Mi et al. 2009). Most of the bacterial proteins identified as S8A subtilisins in MEROPS belong to the PTHR10795 family in PANTHER but only a few belong to the PTHR10795:SF17 subfamily. The classification system in PANTHER uses experimental data and evolutionary relationships to create families. Functional divergence evidence from the ancestors is used to classify proteins in subfamilies (Mi et al. 2009). In MEROPS, the proteins are classified on the basis of sequence comparison (BLAST, FastA, HMMER) to a reference sequence (Rawlings et al. 2011). Subfamily S8A is equivalent to PTHR10795, but PTHR10795 sub-family 17 has no equivalent in the MEROPS classification system.

Members of subfamily PTHR10795:SF17 only could be identified in three phyla of bacteria. A systematic loss of subfamily PTHR10795:SF17 in bacteria and a cross-kingdom HGT event could explain these observations. On the other hand, complex events of HGT from plants to, at least three different, bacterial ancestors could be an alternative but less plausible explanation. In any case, the relationship between bacterial and plants subtilisins cannot be

determined with the evidence provided in this investigation. However, other examples of HGT that involve three different kingdoms have been reported in the past (Richards et al. 2006). Therefore, I do not discard the hypothesis that subtilisins were transferred horizontally from bacteria to plants, and subsequently to fungi. The lateral transfer of an ancestral subtilisin from bacteria to plants would have happened early in the evolution of plants. Subtilisins from sub-family PTHR10795:SF17 were found in all Viridiplanteae members except in Chlorophyta. This observation reflects the ancient origin of this subfamily in plants.

Likewise, the lateral transfer of a plant subtilisin to a *Colletotrichum* ancestor should be ancient, at least before the divergence between monocots and dicots. Based on the draft genome sequences of members of *Colletotrichum* available to us at this time, CPLSs homologs were found in all nine species that I examined. If our hypothesis about the HGT to an ancestor of the *Colletotrichum* is correct, then I expect that all members of the genus to contain a CPLS. If the HGT event occurred earlier, then I expect to find CPLS homologs in other fungal genera. To explain the presence of CPLSs only in *Colletotrichum* without HGT requires us to accept that it is a very ancient gene family that was conserved only in *Colletotrichum* and was lost in all other fungal lineages. I believe that this is unlikely and that the body of evidence supports the HGT hypothesis. The period of time proposed for the HGT transfer is congruent with molecular clock estimation for the monocot divergence from the angiosperms and *Colletotrichum* divergence from other related genera. The estimated divergence date for the monocots ranges from 200 Myr (Wolfe et al. 1989) to 134 Myr (Bell et al. 2005). The most recent calculations propose intermediate values between 155 Myr to 145 Myr (Chaw et al. 2004; Leebens-Mack et al. 2005; Smith et al. 2010). For *Colletotrichum* genera the divergence from other members of the class Sordaryomycetes is calculated to be approximately 150 Myr ago (O'Connell et al. 2012) and 160 to 75 Myr ago from the estimations of the Chapter I. The hypothesis of a lateral transference from an angiosperm that predates the monocot divergence to a *Colletotrichum* ancestor explains the lack of homologous sequence of CPLSs in other fungal species and the abundance in plants. Multiple duplication events prior the monocot divergence are evident in the plant subtilisins tree (**Figure 27**), but CPLSs are not placed inside any branch of a specific group. The CPLSs are placed at a node ancestral to the monocot divergence. These kinds of ancient HGT events were reported in the past (Brown 2003; Richards et al. 2006; Rolland et al. 2009; Marcet-Houben and Gabaldón 2010; de Jonge et al. 2012). Despite the age of the HGT event, the

level of conservation of protein sequences is remarkable. This fact also denotes the possibility of protein function conservation in the fungi, probably favored by selection. In fact, expansions of S8A serine proteases have been observed in *C. higginsianum* (O'Connell et al. 2012), suggesting an important selection pressure for new copies of this gene in some *Colletotrichum* spp.

The broad spectra of metabolic processes that subtilisins are involved in make it difficult to predict a specific function of the CPLSs. However, the remarkable level of structural conservation of CPLSs with plant subtilisins and the differences with the rest of *Colletotrichum* subtilisins suggests the possibility of molecular mimicry. For parasitic organisms a mimetic molecule is defined as a factor that resembles the host molecules for the pathogen's advantage (Elde and Malik 2009). In bacteria, some cases of plant protein mimicry have been reported. AvrPtoB is a protein from *Pseudomonas syringae* than mimics E3 ubiquitin ligase of its plant host (Abramovitch et al. 2006). The molecule suppresses programmed cell death in compatible interactions, enabling the pathogen to avoid the hypersensitive reaction. AvrPtoB can also suppress program cell death in yeast, demonstrating that the molecule has different functions in different eukaryotic models (Abramovitch et al. 2003). But, at the moment, no cases of fungal proteins that mimic plant proteins have been reported.

Our analysis reveals that GLRG_05578 is up-regulated during the infection of maize at 48 hpi and 72 hpi. At the same time, the putative PR-7 genes of maize GRMZM2G091578 and GRMZM2G414915 were induced in the first hours post-infection and then repressed when GLRG_05578 was induced. Putative proteins of GRMZM2G091578 and GRMZM2G414915 genes are S8A subtilisins within the subfamily PTHR10795:SF17 and show the highest similarity to GLRG_05578. In equivalent expression experiments only a few predicted PR-7 proteins were induced or repressed in maize after *Ustilago maydis* infection (Doehlemann et al. 2008). In these experiments GRMZM2G091578 expression was detected 4-8 days post-infection (dpi) and GRMZM2G414915 was repressed 4 dpi. This observation shows a differential behavior of maize subtilisins in the presence of two different fungal pathogens. The low and delayed expression of PR-7 in maize in the presence of *U. maydis* is consistent with the behavior of the maize-*U. maydis* pathosystem, because *U. maydis* is an obligate biotroph with longer periods of symptomless colonization compared with *C. graminicola* (a

hemibiotrophic fungus). On the other hand, GLRG_05578 was expressed towards the transition from biotrophic to necrotrophic stages of fungal infection and coincide with the down-regulation of two putative maize PR-7s with high sequence similarity to CPLSs. Whether the maize proteins are down-regulated by the effect of GLRG_05578 or by the effect of any other stimulus is not yet known. But the synchrony of induction-repression patterns and the level of similarity suggest an important role of GLRG_05578 in the infection process, perhaps by the repression of pathogen-related proteins in the host. In consequence, the acquisition of a PR-like protein would be important for the fungal cells to interfere with plant immune systems.

The structural similarity of CPLSs with plant subtilisins and the pattern of expression in plant infections suggest an important function in plant-fungal interactions. The direct interaction of proteins has been shown in plant and fungal chorismate mutases of maize and *Ustilago maydis* (Djamei et al. 2011). The interaction between CPLSs and plant S8A subtilisins is also possible. For example the crystallized subtilisin SBT3 of tomato reveals the conformation of homodimers (Ottmann et al. 2009). Some of the regions involved in the dimerization of SBT3 are present in GLRG_05578 (see **Table 12**). Whether the CPLSs form dimers is not yet known, but if they do, it is possible that they form heterodimers with their plant counterparts. On the other hand, pre-processed subtilisins are known to inhibit the activity of mature subtilisins. This was studied in a heterologous system, in which the authors determined that the immature form of ARA12 (a subtilisin of *Arabidopsis*) can inhibit the activity of cucumisin (the CPLS model was also compared with this protein), a subtilase of *Cucumis melo* (Nakagawa et al. 2010). The CPLSs encode a signal peptide and the pre-domain inhibitor I9. These domains of the immature proteins are normally removed in the endoplasmic reticulum by auto-catalysis, according to the behavior observed in several studies (Ikemura and Inouye 1988; Ohta and Inouye 1990; Bryan et al. 1995; Coffeen and Wolpert 2004). To use the inhibitory property of immature subtilisins observed in ARA12, would require that *Colletotrichum* skips the preprocessing of CPLS that is normally observed in other serine proteases.

The case of the *C. gloeosporioides* gene CGLO_10271 is particularly interesting. An insertion of one nucleotide truncates this gene by causing a shift of the reading frame. The DNA sequence is very similar to CGLO_07890 and GLRG_05578 (60.5% and 63%, respectively).

Thus, apart from the shift of the reading frame no other nonsense mutations could be identified before or after the nucleotide insertion. Also if the reading frame is corrected, the resulting translated protein shows a high percentage of identity with CGLO_07890 and GLRG_05578 (57.1% and 65.4% respectively). These data suggest that the frameshift mutation is recent. Probably the presence of a second copy of a CPLS is not essential in the genome of *C. gloeosporioides* or the nonsense mutation was a casual event.

Only a few examples of HGT from plants to fungi have been described to date demonstrating that HGT events are very rare (Richards et al. 2009; de Jonge et al. 2012). In the case of the CPLSs, the functions of known subtilisins coupled with the expression pattern during plant infection suggest that they have important roles in plant disease. It is interesting to speculate that the role in plant disease provided a selective advantage to the *Colletotrichum* ancestor providing it with improved fitness, possibly with improved ability to invade its host.

CHAPTER III

A genome wide survey of molecular mimicry in the maize pathogen *Colletotrichum graminicola*

Introduction

Parasitic organisms have the ability to manipulate the host for their benefit. To achieve this some microbial pathogens have evolved the capacity to mimic characteristics of the host. For example, the conidia of the fungus *Monilinia vaccinii-corymbosi* mimics the germinating pollen of the blueberry enabling it to gain access to the plant ovary (Ngugi and Scherm 2006). Extrapolating from morphological mimicry, the ability of one molecule to imitate the function and/or the shape of another molecule is called molecular mimicry, a term that was used for the first time by Damian [5] to refer to the sharing of antigenic determinants between animal pathogens and hosts. Since then, several additional examples of molecular mimicry have been reported. For example, the human poxviruses and herpesviruses acquired the gene *Bcl-2* by horizontal gene transfer from their host. Therefore, the viral protein v-Bcl-2 mimics the function of the native protein and interferes with the normal apoptotic process of the host cell benefiting the replication of the viral molecules (Gangappa et al. 2002; Elde and Malik 2009). Bacteria, including *Salmonella enterica* and *Escherichia coli,* have an effector called IpgB2 that mimics the host Rho-GTPase enzyme and manipulates the cytoskeletal dynamics to increase the pathogen's fitness. IpgB2 and Rho-GTPase do not share structural similarities and it is considered a case of functional mimicry (Alto et al. 2006; Alto and Dixon 2008; Klink et al. 2010). CLE (CLAVATA3 (CLV3)/Endosperm Surrounding Region (ESR)) proteins are involved in cell-fate determination in shoot and root apical meristems of *Arabidopsis* (Clark et al. 1995). Surprisingly, a gene encoding a CLE-like proteins was found in the genome of the plant pathogenic nematode *Heterodera glycines* (HgCLE proteins)(Olsen and Skriver 2003). Subsequently, it has been shown that nematode CLEs can cause root and shoot meristem termination and can complement *Arabidopsis* CLE mutants. Also, it has been shown that the continuous secretion of HgCLE proteins may be required to maintain the feeding cell in a dedifferentiated state (Wang et al. 2005; Wang et al. 2010; Wang et al. 2011). Annexins are a heterogeneous group of proteins involved in calcium regulation of membrane surfaces of the plant cell. The *H. glycines* protein Hs4F01 has an annexin motif, suggesting that it may mimic plant annexins. Functional analyses revealed that homozygous lines of *A. thaliana* transformed with *Hs4F01* are hyper-susceptible to *Heterodera* spp. nematodes. The *A. thaliana* mutant *AnnAt1* deficient in annexin 1 was complemented with *Hs4F01* and the susceptibility to salt stress of the mutant was reverted. This evidence suggests that *Hs4F01*

modulates the stress responses in the plant. Additionally, yeast two hybrid assays confirmed the interaction of *Hs4F01* with a member of 2OG-Fe(II) oxygenase family (the same that promotes the susceptibility to oomycetes pathogens) of the plant, suggesting that this oxidoreductase is the target of Hs4F01 (Patel et al. 2010). Another interesting example of molecular mimicry in plants is the case of E3 ubiquitin ligases. Ubiquitination is a mechanism exclusive to eukaryotes which alters the activity or the localization of the proteins. This mechanisms uses the 26S-proteasome to degrade proteins (Kerscher et al. 2006). In some plant pathogenic bacteria E3 ubiquitin ligases mimic the ubiquitination functions in their host cells. Thus, AvrPtoB of *Pseudomonas syringae* (Abramovitch et al. 2006; Janjusevic et al. 2006; Rosebrock et al. 2007) and XopD of *Xanthomonas campestris* pv. *vesicatoria* (Singer et al. 2013) function like the plant E3 ubiquitin ligases and change the cell host response to the pathogen's attack. According to Rosebrock et al. (2007), AvrPtoB ubiquitinates the plant proteins Fen (kinase protein) that normally leading the effector-triggered immunity (ETI). Without this kinase ETI is not activated and *P. syringae* continues with the infection. AvrPtoB is a multifunctional effector protein and other proposed targets of AvrPtoB are Pto/FLS2/BAK1/CERK1/RIN4 (Deslandes and Rivas 2012). The crystal structure of AvrPtoB revealed strong similarities with E3 ubiquitin ligases and confirm the mimicry role of this protein inside of the host cell (Janjusevic et al. 2006). These examples show the efficacy of molecular mimicry strategies employed by pathogens to avoid the host's defenses and to perform a successful colonization.

The evolution of molecular mimicry is usually explained by horizontal gene transfer or convergent evolution. In the case of horizontal gene transfer, sequence similarity is expected between the mimicking and imitated proteins. After the lateral transference of DNA from the host to the pathogen strong selective pressure may maintain high sequence similarity in cases where high similarity is required for mimicry or may decrease it in order to adapt the gene sequence to the pathogen's genome. In contrast, when molecular mimicry is the result of convergent evolution, high sequence similarity is not expected and to the contrary, structural similarities (tertiary conformation) are awaited. The events of convergent evolution may evolve via random mutations and natural selection (Ludin et al. 2011; Doxey and McConkey 2013; Drayman et al. 2013). Nevertheless, Hall (1994) suggests the concept of "consequential mimicry" to include the cases of phylogenetically preserved molecules. In this case the molecular mimicry is the consequence of the conservation of sequence in a family of proteins that is used by the pathogen to imitate their host target. This scenario does not imply

adaptation behind the molecular mimicry event but there is not restriction for the adaptation either.

Most of the examples of molecular mimicry at this time are the result of functional experiments in the laboratory (Alto et al. 2006; Christensen and Kolomiets 2011; Wang et al. 2011). However, computational methods have recently been developed to detect potential cases of molecular mimicry through the analysis of genome sequences. The first method was developed to detect molecular mimicry in eukaryotic endoparasites of humans (Ludin et al. 2011) and the second one was developed to detect cases in human pathogenic bacteria (Doxey and McConkey 2013). Both methods are based on the detection of sequence similarity among host and parasite proteins. Another *in silico* methodology was developed to detect structural mimicry in several kinds of human parasites (Drayman et al. 2013). This method uses the structures deposited in the protein data bank (PDB) to compare microbe surface proteins with host ligands. This method has the advantage to explore the molecular mimicry performed by convergent evolution but this is limited by the number of structures deposited in the PDB.

Colletotrichum graminicola (Ces.) G. W. Wils. is an ascomycete of the class Sordariomycetes that causes the anthracnose disease of maize. This species is able to infect all parts of the maize plants and at all stages of development (Bergstrom and Nicholson 1999). Anthracnose is one of the most economically important diseases of maize worldwide (Bergstrom and Nicholson 1999; Venard and Vaillancourt 2007; Sukno et al. 2008). In the United States alone *C. graminicola* produces more than 1 billion dollars in yield losses (Frey et al. 2011) and it is a potential risk for the agricultural ecosystems (Kamenidou et al. 2013). Because of its economic importance, *C. graminicola* was the first species of the genus *Colletotrichum* to be sequenced. The result was a high-quality reference genome of 57.4Mb distributed in 13 chromosomes and 12006 predicted protein coding genes (O'Connell et al. 2012).

Very little is known about molecular mimicry in plant pathogenic fungi. In this manuscript, I utilize sequence similarity and functional domain searches to identify molecular mimicry candidates in *C. graminicola*. I identified 30 putative candidates in *C. graminicola* and identified their putative imitated proteins in *Zea mays*. Additionally, the domain search technique revealed the presence of potential candidates of molecular mimicry in other species of the genus *Colletotrichum* including *C. gloeosporioides* and *C. orbiculare*. This is the first time that an *in silico* procedure is developed and applied to detect molecular mimicry candidates in a plant pathogenic fungus.

Materials and Methods

Identification of molecular mimicry candidates by sequence similarity

The proteomes of *C. graminicola* and several non-plant pathogenic fungi were downloaded from NCBI (www.ncbi.nlm.nih.gov) and the Broad Institute (www.broadinstitute.org), **Table 13**. The *Z. mays* proteome was downloaded from the Gramene website (http://ensembl.gramene.org/Zea_mays/Info/Index).

For the first comparison (search based on the pathogen (SBP)) a BLASTP (Altschul et al. 1997) (e-value 1E-1) search was performed with all proteins predicted in the genome of *C. graminicola* M1.001 versus the database made of the non-plant pathogenic fungi proteomes (**Table 13**) plus the maize proteome. For the second comparison (search based on the host (SBH)) a BLASTP (e-value 1E-1) search was performed with all the proteins predicted in the genome of maize versus the database comprised of non-plant pathogenic fungi and *C. graminicola* M1.001.

Table 13. Non-plant pathogenic fungal species used in this study.

Non-plant pathogenic fungi	Source
***Acremonium alcalophilum* JCM 7366**	NCBI: PRJNA33785
***Aspergillus nidulans* FGSC A4**	NCBI: PRJEA40559
***Aspergillus terreus* NIH2624**	NCBI: PRJNA17637
Beauveria bassiana	JGI
***Candida albicans* WO1**	NCBI: PRJNA16373
***Chaetonium globosum* CBS 148.51**	NCBI: PRJNA12795
***Cordiceps militaris* CM01**	NCBI: PRJNA41129
***Exophiala dermatitidis* NIH/UT8656**	NCBI: PRJNA64935
***Myceliophthora thermophila* ATCC 42464**	NCBI: PRJNA32775
***Neurospora crassa* OR74A**	NCBI: PRJNA132
Neurospora discreta	JGI
***Neurospora tetrasperma* FGSC 2508**	NCBI: PRJNA65273
***Saccharomyces cerevisiae* RM11-1a**	Broad Institute
	NCBI: PRJNA13674
***Thielavia terrestres* NRRL 8126**	NCBI: PRJNA79337

From the SBP all the query proteins with higher hit bitscores in maize proteins than non-plant pathogenic fungi but with at least 10 points of difference between the two groups were selected. From the SBH the query proteins with higher hit bit scores in *C. graminicola* than non-plant pathogenic fungi but with at least 10 points of difference between the two groups were selected. All these steps were performed by Python scripts.

Prediction of subcellular localization

SignalP 4.1 (Petersen et al. 2011), WolfPsort v0.2 (Horton et al. 2007) and TMHMM 2.0 (Krogh et al. 2001) were used to annotate the putative subcellular localization of the candidates of SBP and SBH. When more than two transmembrane sites were predicted in the protein it was annotated as a transmembrane protein without taking into account the other annotations. When a signal peptide was predicted (by SignalP or WolfPsort) with two or less transmembrane sites the proteins were predicted as extracellular. Finally, if the protein was predicted to contain two or less transmembrane sites and without signal peptide, this was annotated with the WolfPsort category.

The mimicry candidates and their imitated maize proteins were filtered based on their predicted subcellular localizations. Two combinations of subcellular localization were considered. Extracellular proteins of *C. graminicola* may colocalize with any maize protein and the transmembrane proteins of *C. graminicola* may be able to interact with the extracellular proteins of maize. The rest of combinations would be highly unlikely to occur. The colocalized candidates were annotated with InterPro (Hunter et al. 2012), MEROPS batch BLAST analysis (Rawlings et al. 2011), CAT (CAZymes analysis toolkit) (Park et al. 2010; Lombard et al. 2013), BLAST2GO (Conesa et al. 2005) and BLAST analysis to PHI-base (Winnenburg et al. 2008) and DFVF database (Lu et al. 2012). Additionally, the presence/absence of the candidates in other 5 sequenced strains (i318, i13649, i47511, i63127, i113173) of *C. graminicola* (Rech et al. 2014) was evaluated by BLAST searches.

Gene enrichment analysis, domain interaction prediction and phylogenetic analysis

The pairwise alignments (*C. graminicola*-maize pairs) were performed with MAFFT (Katoh et al. 2002). The search for tandem repeats was performed with RADAR (Heger and Holm 2000). The proteins were considered repetitive when RADAR detected more than 4 repeats with a total score greater than 100. Protein domain interactions were identified candidate by candidate with the Domain Interaction Map (DIMA) (Luo et al. 2011), 3did (Stein et al. 2009) and iPfam (Finn et al. 2005).

To reconstruct the phylogenies of the mimicry candidates, BLASTP searches versus the NR database of NCBI (www.ncbi.nlm.nih.gov) was performed with each *C. graminicola* and

maize protein. When more than one kingdom was observed in the BLAST hits of a candidate (e.g fungi, animal, bacteria and plants) a phylogenetic tree was reconstructed using PhyML (Guindon et al. 2010) to evaluate the possibility of horizontal gene transfer.

The enrichment tests were performed for the final set of mimicry colocalized candidates using the topGO module (Alexa et al. 2006) of Bioconductor (Gentleman et al. 2004) and GOEAST Customized Microarray platform (Zheng and Wang 2008). The GO terms for the entire proteome of *C. graminicola* (metabolic function, biological process and cellular component) were obtained from AgBase GORetriever (McCarthy et al. 2006). The false discovery rate FDR value (≤ 0.15) was used to evaluate the enrichment of each GO term.

Gene expression analysis

The RNA-Seq experiments of O'Connell et al. (2012) were retrieved from the GEO (Gene Expression Omnibus) database (http://www.ncbi.nlm.nih.gov/geo/) of NCBI and used for the expression analysis of the candidates. This experiment contains data of maize leaves infected with *C. graminicola* with samples collected 24, 36 and 60 hours post infection. For the *C. graminicola* mimicker candidates the data were extracted directly from O'Connell et al. (2012). The differential expression of the maize genes was calculated from the raw data available in GEO database under the accession number GSE34632. Briefly, the raw data in SRA format were transformed to fastq format with SRAtoolkit (http://www.ncbi.nlm.nih.gov/Traces/sra/?view=software). Then, the fastq data were evaluated with FastQC software (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) and the reads were preprocessed with FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Next, the reads were indexed with Bowtie 2 (Langmead and Salzberg 2012) and then aligned to the *Z. mays* genome with TopHat 2 (Kim et al. 2013) (anchor length 10 and max multihits 5). SAMStats (Lassmann et al. 2011) was used to evaluate the quality of the read mapping. To extract the information of the alignment to the reference genome SAMTools (Li et al. 2009) was used. The differential expression of the data was calculated with the package DESseq (Anders and Huber 2010) (based on the negative binomial distribution) of R (R Development Core Team 2011).

An overview of the pipeline is shown in **Figure 35**.

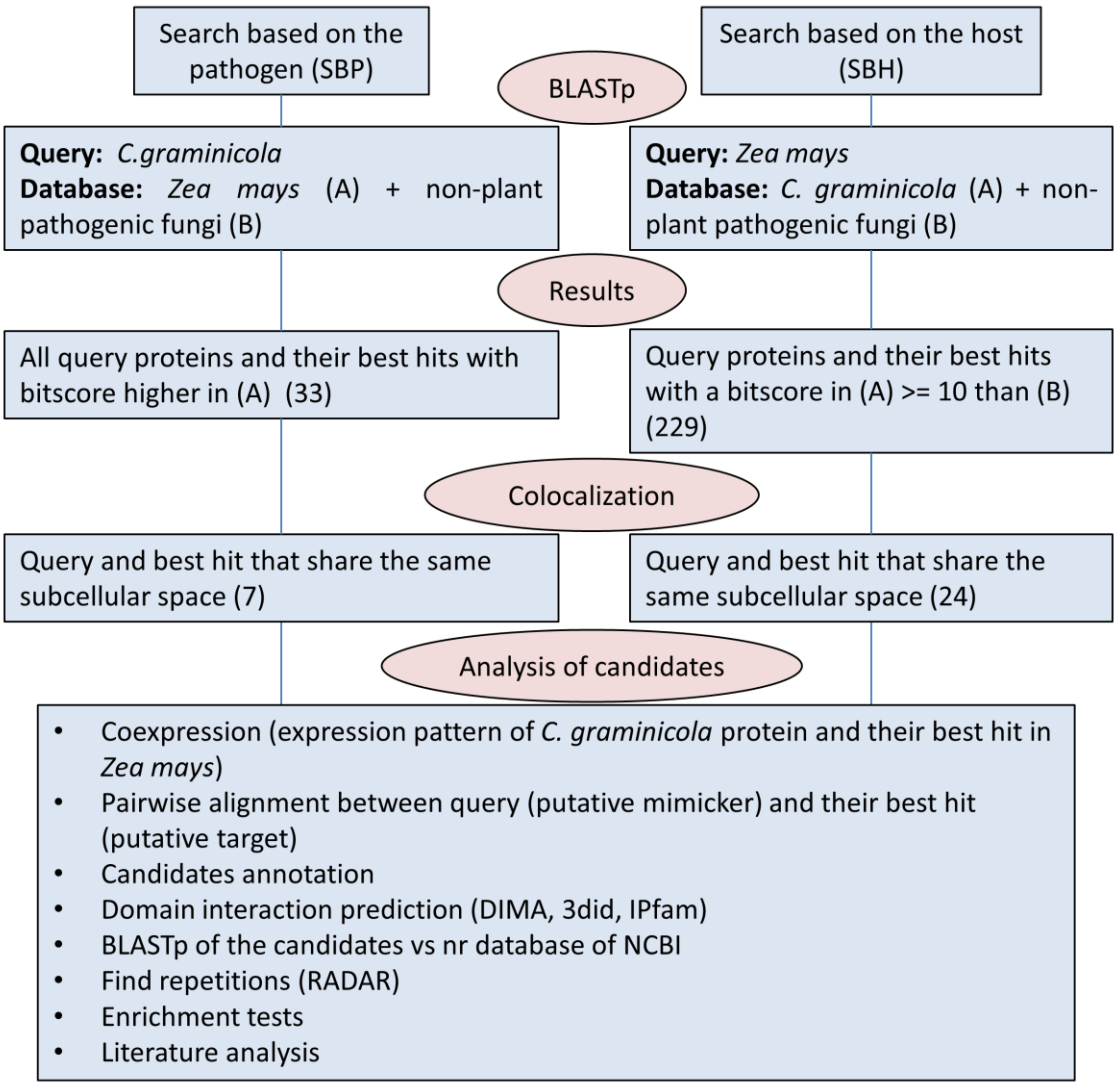


Figure 35. General pipeline of the sequence similarity methods used to detect molecular mimicry candidates in *C. graminicola*.

Identification of molecular mimicry candidates by domain content

All InterPro terms annotated in the complete proteomes from the UniProt database (www.uniprot.org) were extracted. A Python script (available upon request) was used to extract taxonomic information of the occurrence of each InterPro term and prepare a table with all InterPro terms, and the number of occurrences in the plants, fungi and the genus *Colletotrichum*. In addition, the relative percentage of each group (plants, fungi and *Colletotrichum*) to the total number of occurrences of each term (**Additional file 2**) was calculated.

The InterPro terms in which the sum of the percentage of plants and fungi were equal or larger than 98% and also present in *Colletotrichum* spp. were chosen as candidates. The proteins of *Colletotrichum* spp. with the selected InterPro terms were used as a query for BLAST searches vs the NR database of NCBI to detect evidence of horizontal gene transfer. The subcellular localization of these proteins were predicted with SignalP 4.1 (Petersen et al. 2011), WolfPsort v0.2 (Horton et al. 2007) and TMHMM 2.0 (Krogh et al. 2001).

Results

Identification of mimicry candidates by sequence similarity

Molecular mimicry candidates were identified based on the sequence similarity methods of Ludin et al. (2011) and Doxey and McConkey (2013). In these studies the authors compare the proteins of pathogens with those of the hosts and, as a control, to proteins of non-host species that are phylogenetically more closely related to the pathogen than the host. It is expected that mimicry candidates will have more sequence similarity to the proteins of its host than to the control species. The result is a pair of candidate proteins that represents the mimicker protein (encoded in pathogen's genome) and the imitated protein (encoded in the host's genome). To perform this analysis I retrieved the complete proteomes of 14 non-plant pathogenic fungal species (controls) (**Table 13**), *C. graminicola* and the proteome of *Z. mays*.

A BLAST search was performed using the proteins of *C. graminicola* (search based on the pathogen; SBP) as a query versus a database comprised of proteins from non-plant pathogenic fungi and the complete proteome of *Z. mays*. I selected as mimicry candidates the *C. graminicola* proteins with best hits to *Z. mays* proteins and having bit scores at least 10 points greater to any proteins of non-plant pathogenic fungi. 33 *C. graminicola* proteins were obtained for this method (**Figure 36a**).

A second BLAST search was performed using the proteins of *Z. mays* (search based on the host; SBH) as a query versus a database comprised of the proteomes of *C. graminicola* and non-plant pathogenic fungi. I selected proteins with best hits to *C. graminicola* proteins that had bit scores at least 10 points greater to proteins of non-plant pathogenic fungi. I obtained 544 maize proteins in this search **(Figure 36b**). Since several *Z. mays* proteins had as their best hits, the same *C. graminicola* protein, the 544 *Z. mays* proteins correspond to 229 *C.*

graminicola proteins. From the SBP and SBH searches I obtained 262 *C. graminicola* proteins with high similarity to maize proteins.

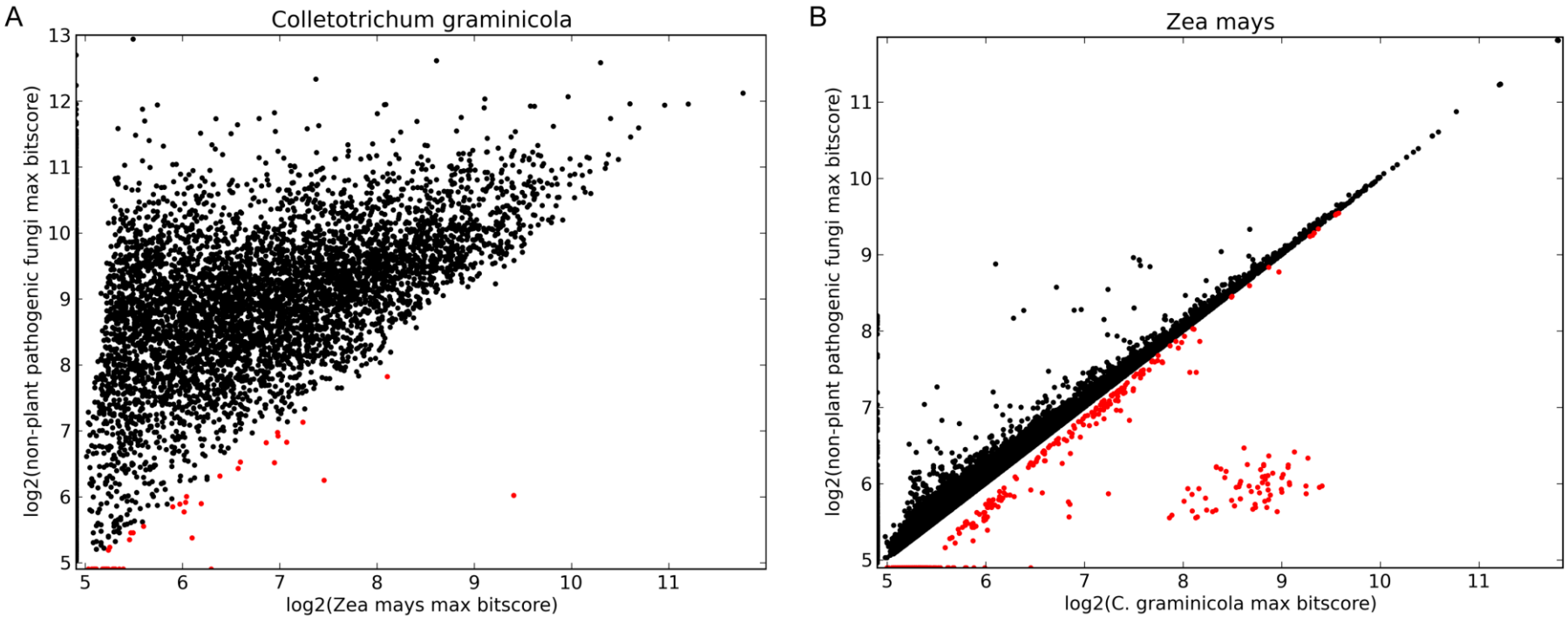


Figure 36. Dot plots of the best BLAST hits for the whole proteome of *C. graminicola* (pathogen) and *Z. mays* (host) vs non-plant pathogenic fungi (control). A) Red dots represent *C. graminicola* proteins with higher bit scores to *Z. mays* proteins than to non-plant pathogenic fungi. B) Red dots represent *Z. mays* proteins with higher bit scores to *C. graminicola* than non-plant pathogenic fungi.

Proteins that mimic proteins of the host frequently are co-localized to the same subcellular space as the imitated protein (Gangappa et al. 2002; Alto et al. 2006; Drayman et al. 2013). For example, the proteins SopE and SopE2 of the human enteric pathogen *Salmonella* are secreted through the type III secretion system to the host cytoplasm. Inside the host, these proteins mimic guanine nucleotide exchange factors (GEFs) that activate Rac1 and Cdc4 (GTPases) leading to the rearrangement of the actin cytoskeleton and subsequent bacterial internalization (Galán and Zhou 2000). These events require that the mimicking proteins (SopE and SopE2) and the imitated proteins (human GEFs) share the same subcellular space, the host cytoplasm. For that reason I predicted the subcellular localization for each pair of candidates (*C. graminicola-Z. mays* proteins) obtained in the BLAST searches described above. I deduced that the subcellular localization of the pair is congruent only if the *C. graminicola* protein is secreted or if the protein of *Z. mays* is secreted and its pair in *C. graminicola* is a transmembrane protein. A schematic view of the interacting proteins between *C. graminicola* and *Z. mays* is shown in **Figure 37**. I obtained 7 pairs of candidates from the SBP search (**Table 14**) and 24 pairs of candidates from the SBH search after the application of the colocalization criterion. In the SBH search, some maize proteins have the same best hit in the *C. graminicola* proteome, thus, 24 pairs of candidates represent one *C. graminicola* protein with a family of maize proteins (**Table 15**). In total I obtained 30 mimicry candidate proteins in *C. graminicola*. The results from the SBP and the SBH searches only share one protein, the subtilisin of *C. graminicola* GLRG_05578T0. The rest of the candidates are unique to each search strategy.

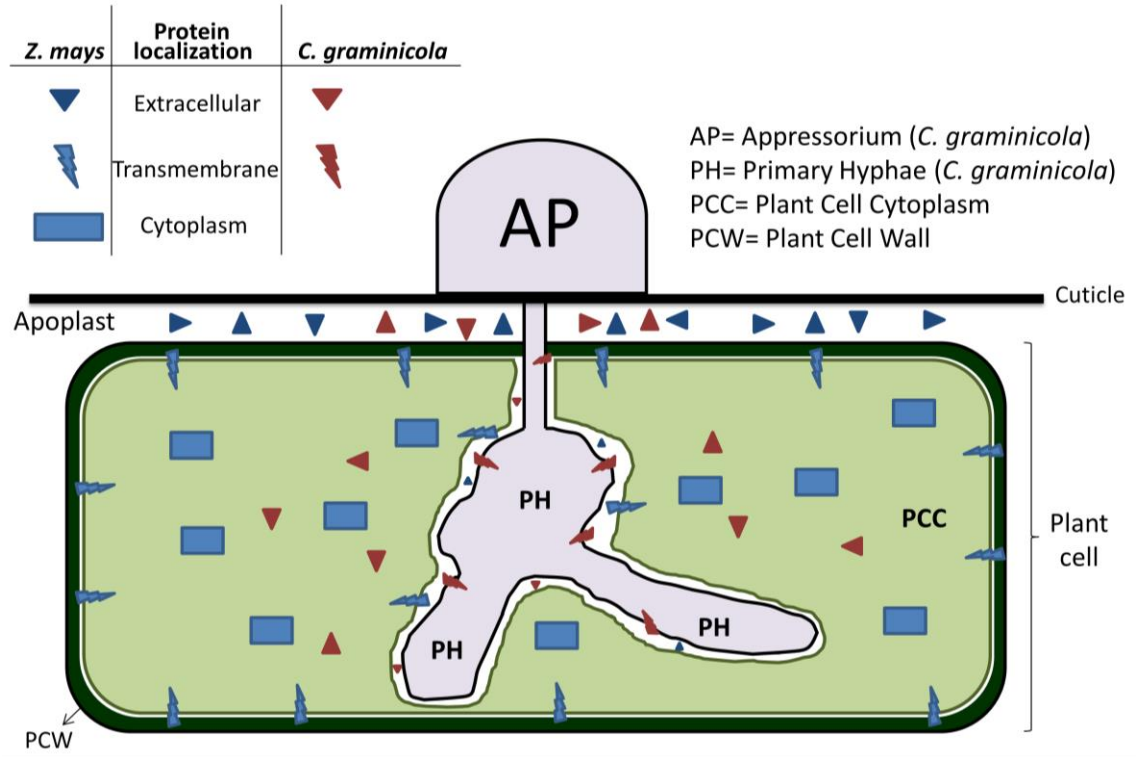


Figure 37. Schematic view of the subcellular localization of mimicry candidates and imitated *Z. mays* proteins. Secreted *C. graminicola* proteins may be in the apoplast or in the plant cell cytoplasm, consequently these proteins could imitate plant proteins inside or outside the plant cell. The transmembrane proteins of *C. graminicola* could interact with the extracellular proteins of *Z. mays* but not with the transmembrane proteins of the plant cell because the fungal cell wall (not represented in the figure) separates the space between them.

Table 14. *C. graminicola* mimicry candidates from the search based on the pathogen (SBP) and that are co-localized to the same subcellular space as the putative imitated maize protein.

C. graminicola (locus number)	Subcellular localization	*Z. mays* (locus number)	Subcellular localization	Pairwise similarity (%)	InterPro domains
GLRG_07383T0	extracellular	GRMZM2G066120_P01	chloroplast	19.6	Protein kinase domain[b]
GLRG_03516T0	extracellular	GRMZM2G106795_P03	cytoplasm-nucleus	19.9	Small GTPase superfamily, ARF type[b]
GLRG_02022T0	extracellular	GRMZM2G110289_P01	chloroplast	34.3	N.A[c]
GLRG_05578T0	extracellular	GRMZM2G121293_P01	extracellular	63.1	Peptidase S8/S53 domain
GLRG_02201T0	extracellular	GRMZM2G143725_P02	chloroplast	9.8	Sec23/Sec24, trunk domain[b]
GLRG_06086T0	extracellular	GRMZM2G150984_P02	plasma membrane	9.1	Kinesin, motor domain[b]
GLRG_02070T0	extracellular	GRMZM2G399183_P01	chloroplast	24.1	Glycine cleavage H-protein[b]

[a] Domain shared by the *C. graminicola* and the *Z. mays* proteins

[b] Domain detected in *Z. mays* protein but not in *C. graminicola* protein

[c] Not available for either of the two proteins

Table 15. *C. graminicola* mimicry candidates from the search based in the host that are colocalized to the same subcellular space as the putative imitated maize protein.

C. graminicola (locus number)	Subcellular localization[§]	*Z. mays* (locus number)	Subcellular localization[§]	Pairwise similarity (%)	InterPro domains
GLRG_01658	extr	GRMZM2G374248_P01	extr	35.8	RlpA-like double-psi beta-barrel domain
		GRMZM2G117942_P01	extr	23.6	RlpA-like double-psi beta-barrel domain
		GRMZM2G117971_P01	extr	23.7	RlpA-like double-psi beta-barrel domain
		GRMZM2G169967_P01	extr	34.2	RlpA-like double-psi beta-barrel domain
		GRMZM2G303937_P01	extr	34.2	RlpA-like double-psi beta-barrel domain
		GRMZM2G450546_P01	extr	34.2	RlpA-like double-psi beta-barrel domain
GLRG_01795	extr	GRMZM2G090652_P01	extr	34.9	Lipase, GDSL
GLRG_01826	extr	GRMZM2G081841_P01	extr	39.2	Glucose-methanol-choline oxidoreductase N y C terminal
GLRG_02947	extr	GRMZM2G168588_P01	extr	29.6	Peptidoglycan-binding lysin domain
GLRG_04104	extr	GRMZM2G120779_P01	extr	40	Pectinesterase, catalytic & Pectin lyase fold/virulence factor
		GRMZM2G012328_P01	extr	28.6	Pectinesterase, catalytic & Pectin lyase fold/virulence factor
		GRMZM2G043943_P01	extr	37.3	Pectinesterase, catalytic & Pectin lyase fold/virulence factor
		GRMZM2G177940_P01	extr	39.7	Pectinesterase, catalytic & Pectin lyase fold/virulence factor
		GRMZM2G177940_P03	extr	34.3	Pectinesterase, catalytic & Pectin lyase fold/virulence factor
GLRG_05347	extr	GRMZM2G147780_P02	extr	34.1	FAD linked oxidase, N-terminal & FAD-binding, type 2[a] PRONE domain[b]
		GRMZM2G147780_P01	extr	30.2	PRONE domain[b]
GLRG_05368	extr	GRMZM2G312226_P01	extr	22.3	FAD dependent
GLRG_05395	extr	GRMZM2G181259_P02	cyto	42.6	Glycoside hydrolase, superfamily
GLRG_05589	extr	GRMZM2G063435_P01	extr	38.7	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G004984_P01	extr	17.593	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G023840_P01	extr	22.289	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G024234_P01	extr	21.429	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G040638_P01	extr	21.646	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G042347_P01	extr	20.552	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G043855_P01	extr	24.057	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G044049_P01	extr	20.219	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G061230_P01	extr	12.987	Plant/fungal/bacteria peroxidase & Haem peroxidase

		GRMZM2G068699_P01	extr	19.88	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G076562_P01	extr	21.893	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G095404_P01	extr	21.006	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G095404_P02	extr	20.71	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G108123_P01	extr	22.154	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G108207_P01	extr	22.961	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G108219_P01	extr	21.557	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G116902_P01	extr	21.429	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G122816_P01	extr	25.705	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G133475_P01	extr	26.087	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G136158_P02	extr	19.412	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G144648_P01	extr	19.585	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G177792_P01	extr	19.764	Plant/fungal/bacteria peroxidase & Haem peroxidase
		GRMZM2G427815_P01	extr	23.457	Plant/fungal/bacteria peroxidase & Haem peroxidase
GLRG_06236	extr	GRMZM2G054224_P01	extr	47.8	Prolyl 4-hydroxylase, alpha subunit & Oxoglutarate/iron-dependent dioxygenase
		GRMZM2G168506_P01	extr	24.16	Prolyl 4-hydroxylase, alpha subunit & Oxoglutarate/iron-dependent dioxygenase
		GRMZM2G459063_P01	extr	24.27	Prolyl 4-hydroxylase, alpha subunit & Oxoglutarate/iron-dependent dioxygenase
		GRMZM2G459063_P02	extr	31.39	Prolyl 4-hydroxylase, alpha subunit & Oxoglutarate/iron-dependent dioxygenase
		GRMZM2G459063_P03	extr	29.41	Prolyl 4-hydroxylase, alpha subunit & Oxoglutarate/iron-dependent dioxygenase
		GRMZM2G520535_P01	extr	25.27	Prolyl 4-hydroxylase, alpha subunit & Oxoglutarate/iron-dependent dioxygenase
GLRG_06664	extr	GRMZM2G014793_P02	plas	32.4	GDP-fucose protein O-fucosyltransferase[b]
GLRG_07065	extr	GRMZM2G006658_P01	extr	22.5	WD40/YVTN repeat-like-containing domain [a]
		GRMZM2G006019_P01	extr	18.5	Mini-chromosome maintenance complex-binding protein[b]
		GRMZM2G343058_P01	extr	22.5	Mini-chromosome maintenance complex-

					binding protein[b]
GLRG_07266	extr	GRMZM2G090411_P01	cyto	37.4	CMP/dCMP deaminase, zinc-binding & Cytidine deaminase-like
GLRG_07361	extr	AC233903.1_FGP002	chlo	12.5	Zinc finger, RING/FYVE/PHD-type & Zinc finger, RING-type
GLRG_08031	extr	GRMZM2G170839_P01	extr	29.1	Glycoside hydrolase, family 10
		GRMZM2G170839_P02	extr	29	Glycoside hydrolase, family 10
GLRG_08648	extr	GRMZM2G086628_P01	extr	26.2	Cupredoxin
GLRG_09115	extr	GRMZM2G105682_P01	extr	32.5	Cupredoxin
GLRG_10053	extr	GRMZM2G031572_P01	extr	40.6	Peptidase S28
	extr	GRMZM2G082502_P01	extr	26.2	Peptidase S28
GLRG_10499	extr	GRMZM5G851862_P01	chlo	42.2	Cytochrome P450
		GRMZM2G140448_P01	chlo	41.9	Cytochrome P450
		GRMZM2G159353_P01	chlo	39.1	Cytochrome P450
		GRMZM2G349187_P02	chlo	21.6	Cytochrome P450
GLRG_10770	extr	GRMZM2G167637_P02	cyto	28.9	Pectinesterase, catalytic & Pectin lyase fold/virulence factor
GLRG_10984	extr	GRMZM2G124307_P01	cyto	39.3	Lipocalin/cytosolic fatty-acid binding domain & Calycin-like
GLRG_11050	extr	GRMZM2G019373_P01	extr	16.5	NA
GLRG_11959	extr	GRMZM5G861077_P01	extr	41.8	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A
GLRG_05578	extr	GRMZM2G121293_P01[c]	extr	63.1	Peptidase S8/S53 domain

[§] extr = extracellular, cyto = cytoplasm, chlo = chloroplast

[a] This InterPro domain is only present in the *C. graminicola* protein but not in the maize protein

[b] This InterPro domain is only present in the *Z. mays* protein but not in the *C. graminicola* protein

[c] GLRG_05578T0 was the best hit of 77 different subtilisins of maize. Only the best hit of maize is presented in this table

The mimicry candidates were annotated with the information available in MEROPS (Rawlings et al. 2011), CAZy (Lombard et al. 2013) and through BLAST searches in PHI-base (Winnenburg et al. 2008) and database of fungal virulence factors (DFVF) (Lu et al. 2012). From the 30 candidates of *C. graminicola* 8 were identified as CAZy enzymes or CAZy associated modules, 2 proteins as peptidases and 5 candidates were highly similar to genes related with reduced virulence or loss of pathogenicity in fungal pathogens. I also searched the genome sequences of 5 other strains of *C. graminicola* i318, i13649, i47511, i63127, i113173 for the presence of the candidates (Rech et al. 2014). Almost all the candidates were found in the 5 isolates of *C. graminicola* with the exception of GLRG_07383T0, GLRG_02201T0 and GLRG_09115T0 that were found only in 4 of the isolates and GLRG_10984T0 that was found only in *C. graminicola* M1.001 and i318 (**Table 16 and 17)**.

Table 16. Functional annotation of the mimicry candidates

C. graminicola (locus number)	CAZy[a]	MEROPS[b]	GOs annotations[c]	Presence in other strains of *C. graminicola*[d]
GLRG_07383				4
GLRG_03516				5
GLRG_02022				5
GLRG_05578		S08A	P:biological_process; P:protein metabolic process; P:catabolic process; F:protein binding; F:peptidase activity	5
GLRG_02201				4
GLRG_06086				5
GLRG_02070				5
GLRG_01658				5
GLRG_01795			P:lipid metabolic process; F:hydrolase activity	5
GLRG_01826	AA3		F:nucleotide binding; F:catalytic activity; P:metabolic process	5
GLRG_02947			P:catabolic process	5
GLRG_04104	CE8		F:hydrolase activity; P:cellular component organization; C:cell wall	5
GLRG_05347	AA7		F:nucleotide binding; F:catalytic activity; P:metabolic process	5
GLRG_05368	AA3		P:metabolic process; F:catalytic activity	5
GLRG_05395	GH3		P:carbohydrate metabolic process; F:hydrolase activity	5
GLRG_05589	AA2		F:binding; F:antioxidant activity; F:catalytic activity; P:metabolic process; P:response to stress	5
GLRG_06236			F:binding; F:catalytic activity; P:metabolic process	5
GLRG_06664				5
GLRG_07065			F:protein binding	5
GLRG_07266			F:hydrolase activity; F:binding; P:metabolic process	5
GLRG_07361			F:binding; F:protein binding	5
GLRG_08031	GH10		P:carbohydrate metabolic process; F:hydrolase activity	5
GLRG_08648				5
GLRG_09115				4
GLRG_10053		S28	F:peptidase activity; P:protein metabolic process; P:catabolic process	5
GLRG_10499			F:binding; F:catalytic activity; F:electron carrier activity; P:metabolic process	5
GLRG_10770	CE8		F:hydrolase activity; P:cellular component organization; C:cell wall	5
GLRG_10984				1
GLRG_11050			P:translation; F:translation factor activity, nucleic acid binding	5
GLRG_11959				5

[a] CAZY enzymes, family (letters), subfamily (numbers): CT= Glycosyltransferase , CBM= Carbohydrate-Binding Module, AA= Associated Modules, CE= Carbohydrate Esterase, GH= Glycoside Hydrolases

[b] MEROPS, code family (first letter + number), subfamily (last letter): S= Serine peptidases

[c] Gene ontology domains: C= cellular component, P= biological process, F= molecular function

[d] Presence of the candidate in one or more of the sequenced strains of *C. graminicola* (i318, i13649, i47511, i63127, i113173). The original search was performed in *C. graminicola* M1.001.

Table 17. Mimicry candidates with BLAST similarity to proteins in the Pathogen-Host Interaction database (PHI-base) and the Database of Fungal Virulence Factors (DFVF).

C. graminicola (query sequence)	e-value [a]	Accession number (UniProt)	PHI-base number	DFVF number	Species name	Annotation	Phenotype of the mutant	Reference
GLRG_01826	2.12 e^{-63}	Q4P8E8	922	Q4P8E8_USTMA	*Ustilago maydis*	Uncharacterized protein	Unaffected pathogenicity	(Kämper et al. 2006)
GLRG_02947	7.87 e^{-28}	B3VBK9_CLAFU	N.A[b]	B3VBK9_CLAFU	*Cladosporium fulvum*	LysM domain	Reduced virulence	(Bolton et al. 2008)
GLRG_04104	4.15 e^{-72}	Q9C2Y1	1028	Q9C2Y1_BOTFU	*Botrytis cinerea*	Pectinesterase	Reduced virulence	(Valette-Collet et al. 2003)
GLRG_05347	2.31 e^{-46}	A0ST43	1046	A0ST43_CERNC	*Cercospora nicotianae*	Oxidoreductase	Reduced virulence	(H.-Q. Chen et al. 2007)
GLRG_05395	2.08 e^{-35}	Q00903	24	Q99324_SEPLY	*Gaeumannomyces graminis*	Avenacinase	Loss of pathogenicity	(Bowyer et al. 1995)
GLRG_08031	9.23 e^{-23}	Q01176	568	N.A[b]	*Magnaporthe oryzae*	Endo-1,4-beta-xylanase 2	Unaffected pathogenicity	(Wu et al. 1997)
GLRG_10770	1.04 e^{-81}	Q9C2Y1	278	Q9C2Y1_BOTFU	*Botryotinia fuckeliana*	Pectinesterase	Reduced virulence	(Valette-Collet et al. 2003)

[a] e-value threshold 10^{-10}

[b] Not available

Several of the molecular mimicry candidates have similarity to mimicry proteins reported in other biological systems, including one LysM protein, a GEF protein, and a small ARF-type GTP-binding protein. In plant pathogens, LysM effector proteins represent a clear case of imitation of plant defensive mechanism to avoid recognition. The fungal LysM proteins compete with plant LysM proteins for the chitin molecules that are released from the fungal cell wall by plant chitinases. The chitin molecules are recognized as pathogen-associated molecular patterns (PAMP) and trigger the defense response of the plant. Thus, the sequestering of chitin by fungal LysM proteins prevents the activation of plant defenses (Bolton et al. 2008; de Jonge and Thomma 2009; de Jonge et al. 2010; Mentlak et al. 2012). I detected one fungal LysM protein (GLRG_02947) and its putative imitated protein in maize (GRMZM2G168588_P01) (both with LysM motifs). The molecular mimicry phenomenon was also observed in GTP-binding molecules of various animal pathogens (Alto et al. 2006). I detected a small GTP-binding molecule ARF type of maize as candidates to imitated protein (GRMZM2G106795_P03) but in their counterpart (GLRG_03516) the small GTPase domain was not detected, nevertheless, these two protein share sequence similarity and potentially the same subcellular localization. Likewise, I detected two maize proteins with PRONE domains (GEF's typical domains of plants) GRMZM2G147780_P02 and GRMZM2G147780_P01. In their counterpart of *C. graminicola* (GLRG_05347) a PRONE domain was not identified but

like in the case of GLRG_03516 this sequence shares sequence similarity and the same potential subcellular location with the maize proteins.

The phylogenetic distribution of the *C. graminicola* candidates and the *Z. mays* ones were investigated to determine whether they are the product of horizontal gene transfer. I performed BLASTP searches vs the NR database of NCBI (www.ncbi.nlm.nih.gov). For all the candidates with hits in more than one kingdom (e.g. fungi, bacteria and animals) I reconstructed a phylogenetic tree. Proteins GLRG_11050T0 and GLRG_05578T0 have incongruent phylogenetic distributions compared to the species tree as expected with horizontal gene transfer. The evidence of the plant origin of GLRG_05578T0 was described previously in the Chapter II. The protein GLRG_11050T0 has a few homologues in the NR database and for that reason the tree derived from this protein is inconclusive (data not shown). Several other candidates are unique to *C. graminicola* (GLRG_07383T0, GLRG_01040T0) or to the genus *Colletotrichum* (GLRG_03516T0, GLRG_02201T0) making analysis of HGT impossible. The other candidates (in *C. graminicola* and in maize) have a phylogenetic distribution congruent with the species tree, thus it is unlikely that these genes evolved by horizontal gene transfer.

I used the DIMA (Domain interactions map) tool (Luo et al. 2011) to determine whether the mimicry candidates interact with their imitated *Z. mays* proteins or other proteins. The prediction of interacting domains can also give us clues to the putative target proteins of the mimicking candidate. More than 70% of the domains in the candidates are predicted to interact with other domains and at least 14 candidates have the ability to interact with the same domain (to form dimers) (**Table 18**). It is also possible that such interacting proteins could form host-pathogen heterodimers as was reported by Chan et al. (2009). Based on the Doxey and McConkey 2013) study several of mimicry candidates of human parasites show intragenic tandem repeats. The presence of repetitive proteins was investigated with the tool RADAR (Heger and Holm 2000). Only GLRG_11050T0 presented evidence of a repetitive structure (**Table 18)**.

Table 18. InterPro domain interaction and repetitive pattern observed in the candidates.

C. graminicola (locus number)	InterPro term with predicted interactions as reported by DIMA	Number of interacting domains reported in DIMA	Interaction with the same domain	Repetitive pattern
GLRG_07383T0	-	-	-	No
GLRG_03516T0	-	-	-	No
GLRG_02022T0	-	-	-	No
GLRG_05578T0	IPR000209	15	Yes	No
GLRG_02201T0	-	-	-	No
GLRG_06086T0	-	-	-	No
GLRG_02070T0	-	-	-	No
GLRG_01658T0	IPR009009	-	-	No
GLRG_01795T0	IPR001087	6	Yes	No
GLRG_01826T0	IPR000172	2	Yes	No
GLRG_02947T0	IPR018392	5	No	No
GLRG_04104T0	IPR000070	2	Yes	No
GLRG_05347T0	IPR006094	24	Yes	No
GLRG_05368T0	IPR006076	50	Yes	No
GLRG_05395T0	IPR001764	9	Yes	No
GLRG_05589T0	IPR002016	3	Yes	No
GLRG_06236T0	IPR005123	8	Yes	No
GLRG_06664T0	-	-	-	No
GLRG_07065T0	-	-	-	No
GLRG_07266T0	IPR002125	13	No	No
GLRG_07361T0	IPR001841	-	-	No
GLRG_08031T0	IPR001000	4	Yes	No
GLRG_08648T0	IPR008972	-	-	No
GLRG_09115T0	IPR008972	-	-	No
GLRG_10053T0	IPR008758	1	Yes	No
GLRG_10499T0	IPR001128	34	Yes	No
GLRG_10770T0	IPR000070	2	Yes	No
GLRG_10984T0	IPR000566	6	Yes	No
GLRG_11050T0	-	-	-	Yes
GLRG_11959T0	IPR021102	-	-	No

Additionally, an enrichment test (in GO terms) for the *C. graminicola* candidates was performed. The category cell wall organization was detected as enriched (candidates GLRG_04104T0 and GLRG_10770T0; Pectin lyases) (**Table 19**).

The overall pipeline of the sequence similarity method was summarized in the **Figure 35.**

Table 19. GO terms enrichment of the molecular mimicry candidates.

GO.ID	Term	Count	False discovery rate (FDR) value
GO:0042545	cell wall modification	2	0.01
GO:0045229	external encapsulating structure organization	2	0.02
GO:0071555	cell wall organization	2	0.02
GO:0006816	calcium ion transport	1	0.03
GO:0070588	calcium ion transmembrane transport	1	0.03
GO:0043086	negative regulation of catalytic activity	1	0.04
GO:0044092	negative regulation of molecular function	1	0.04
GO:0070838	divalent metal ion transport	1	0.05
GO:0072511	divalent inorganic cation transport	1	0.06
GO:0016998	cell wall macromolecule catabolic process	1	0.08
GO:0006979	response to oxidative stress	1	0.11
GO:0044036	cell wall macromolecule metabolic process	1	0.12
GO:0071554	cell wall organization or biogenesis	3	0.15

Identification of mimicry candidates by domain content

I also reasoned that fungal proteins that contain domains that are typically found in plants, may be mimicking plant genes. For that reason a method based on domain searching was performed to compare with the sequence similarity methods. Therefore, I identified protein domains that are frequent in plant proteins but rare in fungal proteins. At first step the InterPro annotation of all proteins of the species with a complete genome deposited in UniProt (www.uniprot.org) was extracted. With this information the percentage of appearance of each domain in plants, fungi and species of genus *Colletotrichum* was calculated. Then, I sort the results from the most common domains in plants to the less common ones. The chosen molecular mimicry candidates were the proteins with very common domains in plants, uncommon in fungi and that are present in species of *Colletotrichum* (see methods). The proteins that fit with these criteria are shown in the **Table 20** and the entire list of InterPro terms is available in the **Additional file 2**.

Table 20. The most frequent InterPro terms of plants that are also present in *Colletotrichum spp.* proteins.

InterPro ID	Total count	Description	% of occurrence in plants	% of occurrence in fungi	% of occurrence in *Colletotrichum*	Proteins in *Colletotrichum* species with the InterPro term[a]	Secretion signal in *Colletotrichum protein*
IPR000823	4369	Plant peroxidase	99.61	0.25	0.09	CH063_00654	Yes
						GLRG_05589	Yes
						CGGC5_13024	Yes
						Cob_11760	Yes
IPR013210	8581	Leucine rich repeat	99.42	0.03	0.01	Cob_09963	Yes
IPR001929	1339	Germin	93.95	4.63	0.37	CGGC5_8308	Yes
						CH063_15189	Yes
						CGGC5_8928	Yes
						CH063_01707	Yes
						CH063_04639	Yes

[a] CH = *Colletotrichum higginsianum*, GLRG = *Colletotrichum graminicola*, CGGC5= *Colletotrichum gloeosporioides* gc5, Cob = *Colletotrichum orbiculare*

With this method the only candidate protein of *C. graminicola* obtained was GLRG_05589T0 (a putative peroxidase). This protein was also obtained with the SBH protein similarity method and represents the only coincidence between these two methods and reflects the importance of this protein as a mimicry candidate. The rest of candidates belong to different species of *Colletotrichum* spp. For all the proteins detected with this method BLASTP

searches vs the NR database of NCBI were performed. The majority of the blast hits are from fungi, as expected in vertically inherited proteins. The only exception is Cob_09963 of *C. orbiculare* in which the majority of the blast hits were from plants. The tree reconstruction of Cob_09963 with their BLAST hits suggests that this gene may have been inherited as the result of a horizontal gene transfer from plants into an ancestor of *Colletotrichum*, *Bipolaris* and *Pyrenosphora* (**Figure 38**). This observation explains the lack of the domain IPR013210 in other fungi. The third interesting result obtained by this method was the *Colletotrichum* proteins with the InterPro term IPR001929 (Germin). This domain was more extended in the fungal kingdom (more than 60 fungal proteins have this domain in the InterPro database) than the other results showed in the **Table 20**. The term Germin is strongly associated with plants and these proteins were described as developmentally regulated proteins and could also be involved in defense of cereals (Lane 1994; Patnaik and Khurana 2001). There is no evidence that fungal proteins with Germin domain had been transferred by horizontal gene transfer. Therefore, it is possible that the similarities between Germin of plants and fungi came from convergent evolution. Overall the results obtained for the domain search give us interesting molecular mimicry candidates. This method in theory has the ability to detect candidates that evolve for convergent evolution (because the InterPro searches are not only based in sequence similarity) and not only by horizontal gene transfer or consequential mimicry.

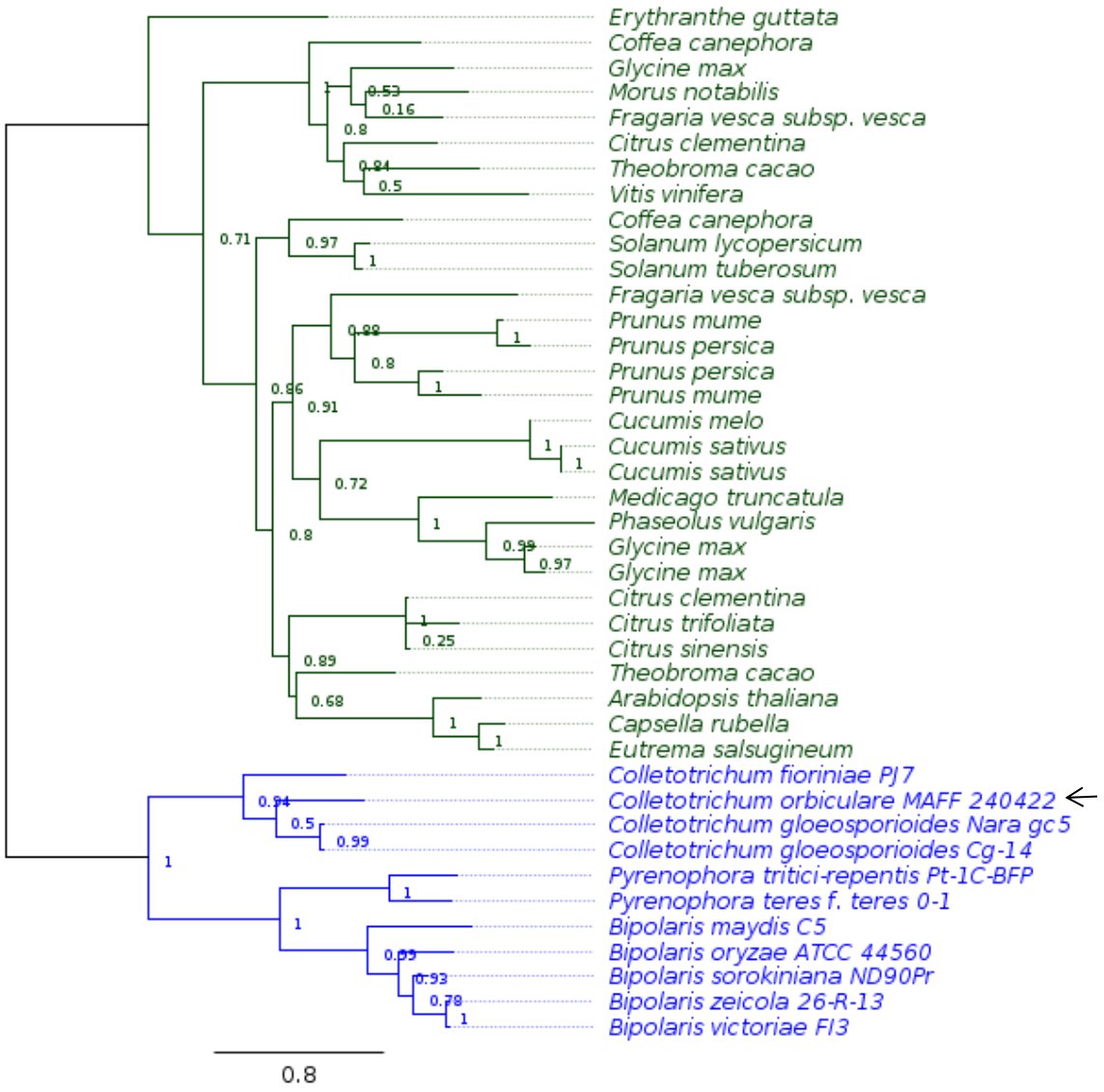


Figure 38. Phylogenetic tree of the *C. orbiculare* protein Cob_09963 (arrow). This family has a narrow distribution in fungi (blue) and are most similar to plants (green). Internal nodes are labeled with SH-branch support.

Coexpression of mimicry candidates and maize genes

The mimicking and the imitated genes may share a similar expression pattern throughout the infection process. This behavior could increase the opportunities of the pathogen's protein to imitate their counterpart because it is localized in the correct place (colocalized) and at the correct time (coexpressed). Evidently, this is not the only option to perform the molecular mimicry and a different pattern of expression of the putative mimicry candidate compared with the imitated gene candidate does not exclude the possibility of molecular mimicry. But

the coincidence of expression patterns could reflect a fine-tuning between two phylogenetically unrelated organisms. The expression of each candidate on infected maize leafs by *C. graminicola* after 24, 36 and 60 hours post infection (hpi) was explored. The *C. graminicola* expression data were obtained from the RNA-seq analysis of O'Connell et al. (2012) and the maize expression data were obtained from the raw RNA-seq data deposited in GEO (http://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE34632 (see methods). The data for all pairs of candidates are showed in the **Additional file 3** and the entire expression profile of maize genome is showed in the **Additional file 4**. Four candidates in *C. graminicola* (GLRG_03516T0, GLRG_05347T0, GLRG_05589T0 and GLRG_10499T0; **Table 15**) have a similar expression profile to their pairs in maize (**Figure 39)**. These candidates share sequence similarity, similar subcellular localization and expression patterns during infection, making them the strongest candidates for molecular mimicry.

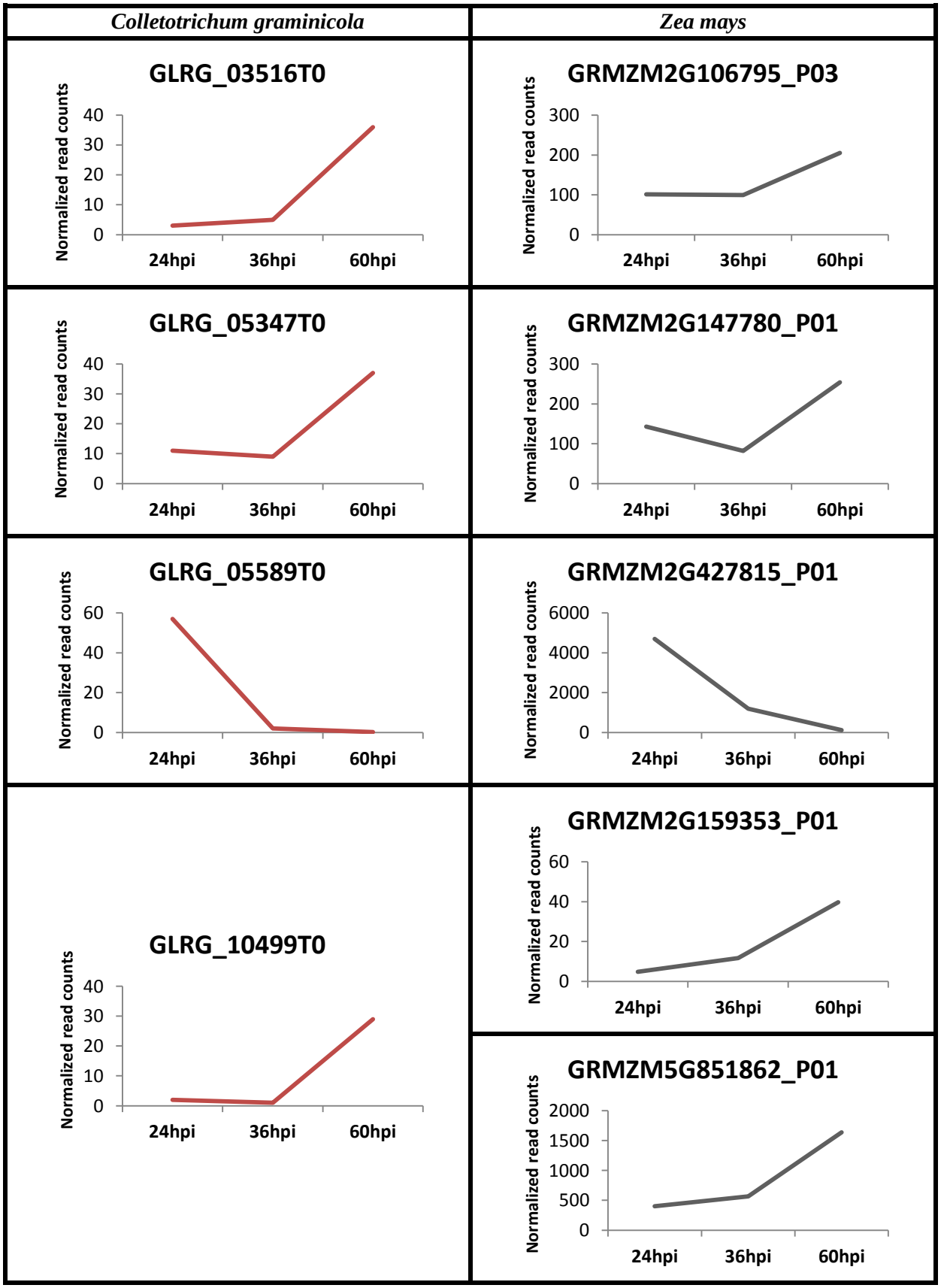


Figure 39. Expression profile of pairs of genes (*C. graminicola* – *Z. mays*) with similar expression patterns during infection. Expression was measured 24, 36 and 60 hours post infection (hpi) of *C. graminicola* in maize leafs using the transcription profiling experiments of O'Connell et al. (2012).

Discussion

To identify potential molecular mimicry proteins encoded in the *C. graminicola* genome I applied two approaches that are based on sequence similarity between the mimicking protein and the imitated protein that were originally developed for detecting mimicry in human pathogens (Ludin et al. 2011)(Doxey and McConkey 2013). I also include an additional criterion that requires the mimicry candidate and its imitated protein in maize to be colocalized to the same subcellular space. Using these methods, I identified 30 candidate proteins in the *C. graminicola* genome that putative imitate proteins encoded in the maize genome.

Several of the candidates are carbohydrate-active enzymes (CAZy) and peptidases (see **Table 16**). These families are usually associated with pathogenicity in fungi, particularly those with roles in plant cell wall modification (van den Brink and de Vries 2011; O'Connell et al. 2012; Ohm et al. 2012). Secreted peptidases were reported in several functions like, nutrient transport (Thon et al. 2002), degradation of host plant tissues (Carlile et al. 2000) and digestion of proteins involved in the plant response against pathogens (Olivieri et al. 2002). In addition, five of the candidates show high similarity (e-values < 10^{-10}) to known virulence factors in other fungi (**Table 16 and 17**) (Winnenburg et al. 2008) suggesting that they may have a similar role in *C. graminicola.* One of the most intriguing candidates is GLRG_02947 a LysM protein with similarity to Ecp6 of *Cladosporium fulvum*. A gene knockout of *Ecp6* reduced the virulence of the fungus (Bolton et al. 2008). Another interesting candidate is GLRG_05395 which has similarity to a saponin detoxifying enzyme of *Gaeumannomyces graminis*. The disruption of this gene caused the total loss of pathogenicity in *G. graminis* (Bowyer et al. 1995). Based on these results, I suggest that molecular mimicry is important in the adaptation of *C. graminicola* to it host plant, maize. I also analyzed the relative abundance of InterPro terms in difference taxonomic groups and identified three that are abundant in plants but very rare in *Colletotrichum* spp. One of the most intriguing results is IPR000823, which represents a family of class III peroxidases (EC 1.11.1.7) present in all plants (Passardi et al. 2004). The *C. graminicola* member of this family, GLRG_05589, shares subcellular localization and patterns of expression with GMRZM2G427815_P01 of maize (**Figure 39**). Class III peroxidases are involved in several plant processes including abiotic and biotic stress, wounding, leaf expansion, nutrient deficiency, senescence, abscission and many others

(Cosio and Dunand 2009). They are known to be involved in ROS (Reactive Oxygen Species) production during the early stages of the response of maize and other plants response to pathogen attack (Lamb and Dixon 1997; Almagro et al. 2009) . In maize, the class III peroxidase *pox12* interacts with the *Ustilago maydis* peroxidase inhibitor Pep1 preventing infection. The silencing of *pox12* led to a substantial increase in penetration efficiency of Δpep1 *U. maydis* as well as a reduction of visible plant defense responses (Hemetsberger et al. 2012). Because of the large number of copies and functions of these enzymes in plants, it is believed that each copy is specialized for a different process (Cosio and Dunand 2009). In *Arabidopsis thaliana* some specific functions of peroxidases were studied by transcriptional profiling and gene silencing. The most similar *A. thaliana* protein to the maize proteins GMRZM2G427815_P01 is AT5g05340 which has roles in ozone stress (Ludwikow et al. 2004), insect defense (Little et al. 2007), infection by *Pseudomonas syringae* (Mohr and Cahill 2007) and stamen abscission (Cai and Lashbrook 2008). As such, peroxidases may be excellent targets for molecular mimicry because of their importance and specificity in plants. Important processes for the plant defenses such as ROS production or signal transduction pathways could be disrupted by fungal effector proteins that imitate plant produced peroxidases.

The LysM motif (Lysine motif) is a modular cassette (usually 42-48 amino acids in length) found in proteins of every living organism except Archaea (Bateman and Bycroft 2000; Zhang et al. 2007). In plant pathogenic and symbiotic fungi proteins with LysM motifs have a specific role in sequestration of chitin molecules detached from the fungal cell wall by the action of plant chitinases (Bolton et al. 2008; de Jonge and Thomma 2009; de Jonge et al. 2010; Mentlak et al. 2012) or to protect the fungal cell wall from attack by hydrolytic enzymes (Marshall et al. 2011). Chitin is one of the major components of the fungal cell wall and function as pathogen-associated molecular patterns (PAMPs). PAMPs activate PAMP-triggered immunity (PTI) that is associated with the secretion of ROS, hydrolytic enzymes and inhibitors of pathogen hydrolytic enzymes (Giraldo and Valent 2013). The plant receptors of PAMPS are called PRRs (pattern recognition receptors). In rice, receptors with LysM motifs called CEBiP have been reported to bind chitin fragments (Shimizu et al. 2010). PRRs with LysM motifs (including CEBiP) compete with the LysM fungal effector of *Magnaporthe oryzae* called Slp1 to bind to chitin oligomers in the plant apoplast. In that sense, it was proposed that fungal LysM effectors are able to inhibit the PTI of their hosts by the competition between the plant LysM receptors and the fungal LysM effectors for the chitin

oligomers (Mentlak et al. 2012). This case fits very well with the definition of molecular mimicry by Elde and Malik (2009). They define molecular mimicry as pathogen-encoded factors (LysM fungal effectors) that resemble host factors (LysM plant receptors) in order to co-opt or disrupt host functions for the pathogen's benefit (to inhibit PTI). LysM proteins with affinity to chitin oligomers have been found in several pathogenic fungi including ECP6 of *Cladosporium fulvum* (Bolton et al. 2008), Mg3LysM of *Mycosphaerella graminicola* (Marshall et al. 2011), Slp1 of *Magnaporthe oryzae* (Mentlak et al. 2012) and ChELP1 of *Colletotrichum higginsianum* (Kombrink 2014). The interference with the activation of chitin-induced host immunity was demonstrated for all these LysM proteins. The presence of LysM effector proteins in all these species suggests a conserved trait among fungal plant pathogens. This pipeline identified a LysM protein (GLRG_02947) for its resemblance to a maize LysM protein GRMZM2G168588_P01 (**Table 15**). The protein GLRG_02947 is highly similar to other fungal LysM like ECP6, Mg3LysM, Mg1LysM, ChELP1, ChELP2 and Slp1 (data not shown) suggesting that the same role of binding to chitin oligomers could be used by *C. graminicola*. Despite that the LysM fungal-plant interaction has not previously been described as molecular mimicry I consider that this is in fact a case of molecular mimicry.

Rop proteins are small GTP-binding molecules that work as signaling switches for controlling growth, development and plant responses to various environmental stimuli (Yang 2002) including pathogen-induced H_2O_2 production (Li et al. 2001) and are key regulators of the immune system of both plant and animals (Kawano et al. 2014) These proteins turn on or turn off from a GDP-bound inactive to a GTP-bound active conformation. The key molecules that catalyze the GDP dissociation for subsequent GTP binding are called guanine nucleotide exchange factors (GEFs). An exclusive family of GEFs was characterized in plants by the presence of the domain PRONE (plant-specific Rop nucleotide exchanger) (Berken et al. 2005). A PRONE domain was detected in maize protein GRMZM2G147780_P02. Its best hit in *C. graminicola* was GLRG_05347 with 34.1% of similarity, but a PRONE domain was not detected. This pair of proteins (GRMZM2G147780_P02 - GLRG_05347) shares the same subcellular space (**Table 15**) and very similar expression patterns during infection (**Figure 39**) (both are differentially expressed 60 hours post infection, see **Additional file 3**). Because of their ubiquitous role in eukaryotic cells, the small GTP-binding molecules are typical targets for bacterial pathogens (Boquet 2000). In *Salmonella*, the proteins SopE and SopE2 work as GEFs activating human Rac1 and Cdc4 which in turn rearrange the host cell cytoskeleton leading to bacterial internalization into the intestinal epithelial cells (Galán and

Zhou 2000). Because GEFs have been shown to be molecular mimics on other systems, the finding of a plant-like GEFs protein in *Colletotrichum* suggests that they too may be mimicking GEFs of their hosts.

Several proteins related with small GTP-binding molecules were detected as molecular mimicry in bacteria like *E. coli* or *Shigella* (Alto et al. 2006; Alto and Dixon 2008). A small GTP-binding protein (of type ARF) was identified among the mimicry candidates (maize protein GRMZM2G106795_P03). In plants, these proteins are involved in the trafficking between the plasma membrane and the cytoplasm of an auxin efflux carrier (Geldner et al. 2001). The best BLAST hit of GRMZM2G106795_P03 in *C. graminicola* (GLRG_03516) shares a low similarity with the maize protein but the expression pattern is very similar and the subcellular localization is congruent with a possible interaction (**Figure 39** and **Table 15**). The possibility that the fungus takes control of these kinds of molecules could be of enormous relevance to disrupt plant cell functions for the pathogen's benefit.

The lipase GDSL (GLRG_01795) may also have an important role in hijacking the plant's immune system (Lee et al. 2009). Lipases function in the synthesis of signaling lipids that have important roles in plant defenses. The lipase GLIP1 of *Arabidopsis* is required for resistance to *Alternaria brassicicola* and overexpression of GLIP1 enhances resistance (Oh et al. 2005). When GLIP1 is applied to *Arabidopsis* plants, disease resistance is induced throughout the plant, suggesting that it has a role in systemic defense signaling (Kwon et al. 2009). The *Arabidopsis* GLIP2 protein, another member of the GDSL family of lipases also has a role in plant defense. A knockout of GLIP2 resulted in increased susceptibility to *Erwinia carotovora*. A GLIP2 knockout had increased proliferation of lateral roots as well as increased expression of auxin responsive genes, suggesting that auxin signaling is deregulated in the knockout mutant (Navarro et al. 2006; Z. Chen et al. 2007; Lee et al. 2009). Auxin signaling usually reduces plant defenses and many pathogens are known to synthesize auxin (Spaepen et al. 2007). Thus GLRG_01795 may mimic plant GDSL to enhance auxin production and promote susceptibility.

Horizontal gene transfer and the convergent evolution are two mechanisms by which molecular mimicry may evolve (Stebbins and Galán 2001; Ludin et al. 2011; Doxey and McConkey 2013; Drayman et al. 2013). In the case of convergent evolution, the two proteins may have low similarity at the sequence level, although they should have structural similarity. In the case of horizontal gene transfer, sequence similarity is expected (Drayman et al. 2013).

These two options are incongruent with the results of Ludin et al. (2011) and Doxey and McConkey (2013) because these authors found molecular mimicry candidates by sequence similarity but without evidence of horizontal gene transfer in most of the cases. This observation contradicts the expectation that molecular mimicry should evolve by HGT. To detect candidates of molecular mimicry by sequence resemblance between the mimicker and the imitated protein another evolutionary explanation is needed. The presence of conserved gene families in unrelated organisms and the presence of selective pressures (e.g the need of the pathogen to infect the host) may explain these results. A similar evolutionary scenario was described by Hall (1994) in the use of the term "consequential mimicry". This term in contrast to adaptative mimicry (produced by horizontal gene transfer or convergent evolution) is the result of sharing phylogenetically conserved genes in distantly related organisms. Consequential mimicry does not imply adaptation but equally does not exclude it. I propose that the method based on sequence similarity is able to detect cases of molecular mimicry that fit with the concept of consequential mimicry (with or without adaptation) or cases that are the result of horizontal gene transfer. The molecular mimicry cases produced by convergent evolution cannot be detected with this method. On the other hand, I developed a method based on the abundance of certain domains that are scarce in unrelated lineages which can detect potential cases of molecular mimicry regardless of their evolutionary history.

In this study, I utilized subcellular localization as an additional criterion to identify mimicry candidates. Several factors such as an insufficient number of unrelated control species or the loss of gene families in unrelated control species could mislead mimicry candidate identification, resulting in false positives. I used the colocalization criterion to filter the list of result of the BLAST searches. Two colocalization scenarios were proposed. First, the secreted proteins of *C. graminicola* may imitate the extracellular or cytoplasmic proteins of the host (Kamoun 2007; Koeck et al. 2011; Rafiqi et al. 2013). Second, the extracellular proteins of maize could interact with the extracellular domain of transmembrane proteins of *C. graminicola*. All the candidates obtained in this study fit with the first scenario. I considered that the transmembrane proteins of maize and *C. graminicola* are not able to interact due the presence of fungal cell wall in the primary infecting hypha that separates them (Rafiqi et al. 2012; Yi and Valent 2013).

The analysis of domain interaction revealed the ability of several mimicry candidates to form dimers. The capacity to form dimers may be exploited by mimicry proteins by forming a

heterodimer with the imitated host protein. This is the case of the Toll-like receptor proteins (TLR) of humans that interact with the myeloid differentiation primary response gene 88 (MyD88) to trigger the innate immune response (Chan et al. 2009). The heterodimerization of TLR through TIR domains is necessary to regulate the synthesis of pro-inflammatory cytokines. However, a TIR domain structurally similar to TLR was discovered in *Paracoccus denitrificans*, and it has the ability to interact with MyD88. Apparently, with this mechanism *Paracoccus denitrificans* decrease the inflammatory response avoiding the dimerization of TLR or interacting with MyD88 (Chan et al. 2009). At least 14 of the *C. graminicola* mimicry candidates were predicted with the capability to form dimers (**Table 18**). These proteins have the potential to form heterodimers like TIR domains proteins of *P. denitrificans* to disrupt their host counterparts.

I used the transcriptional profiling experiments of O'Connell et al. (2012) to study the expression of maize and *C. graminicola* genes under infection conditions. In candidates like GLRG_05589 at least one of their pairs in maize (GRMZM2G427815_P01) shares the expression pattern (**Figure 39** and **Additional file 3**). This data is congruent with an initial overexpression of the gene in *C. graminicola* and their maize pairs with a posterior fall in the biotrophic and necrotrophic phases. This behavior is congruent with molecular mimicry since these proteins localized to the same subcellular space and are expressed at the same time. Other examples of synchronized expression are in GLRG_03516T0, GLRG_10499T0 and GLRG_05347T0 with some of their putative imitated maize proteins (**Figure 39**). It is important to emphasize that the coexpression is not the only possible scenario for the molecular mimicry. If the mimicker protein inhibits the expression of their target (as in negative feedback systems) the expression pattern expected in plant's genes is the opposite of the pathogen's genes.

To test the horizontal gene transfer (HGT) hypothesis, a BLASTP search of the candidates versus the NR database was performed. From the results of the Chapter I, I know that the candidates of this study were not considered horizontal transferred genes (taking into account interkingdom HGT only) except for GLRG_05578T0 that is a member of the family that I call *Colletotrichum* plant-like subtilisin (CPLS) discussed extensively in the Chapter II. Nevertheless, one of the candidates shows an unusual phylogenetic distribution. GLRG_11050T0 is very uncommon in fungi and I found homologous in metazoan and euglenozoa but with a very low similarity. Another characteristic of GLRG_11050T0 is their

repetitive nature, the only one with this feature in the candidates. The repetitive proteins were related with molecular mimicry before (Doxey and McConkey 2013). The lack of homology in other species was also observed in other repetitive *C. graminicola* protein described in Vargas et al. (manuscript in review). The rest of the candidates of this study (from *Colletotrichum* and maize) shows a non-HGT phylogenetic distribution.

The study of molecular mimicry opens a new perspective in the knowledge of the plant-microbe interactions. The capacity of fungi to mimic plant proteins involved in defense suggests the evolution of very specific tactics to invade their host. This gives us the opportunity to identify new components of the plant's immune system which can lead to the development of new disease control strategies. Likewise, the improvement of *in silico* molecular mimicry detection methods is an important task for the future and also opens an exciting field of research for molecular plant pathology.

CONCLUSIONES

Como último punto de este trabajo, a continuación se enumeran las conclusiones obtenidas en esta tesis y que responden a los objetivos e hipótesis formuladas al inicio de la misma.

1. El árbol filogenético calibrado en el tiempo realizado en este trabajo muestra que los genes de transferencia horizontal identificados en la actualidad se han mantenido durante millones de años en *Colletotrichum* spp. y en varios hongos de la sub-división pezizomycotina. Asimismo, la evidencia filogenética muestra que el flujo de genes procedentes de bacterias hacia especies del género *Colletotrichum* ha sido constante en la historia evolutiva de este género.

2. Las categorías funcionales inferidas a los candidatos de transferencia horizontal coinciden con las categorías funcionales asignadas a familias de genes volátiles (genes que evolucionan por pérdidas y duplicaciones). Esto sugiere que no todas los genes son susceptibles de ser mantenidos de manera estable tras la transferencia horizontal, sino que además deben poseer funciones específicas, compatibles con la evolución del genoma huésped.

3. Dada la evidencia filogenética y estructural se puede afirmar que la familia de genes denominada como *CPLSs* (por sus siglas en inglés *Colletotrichum plant-like subtilisins)* es el resultado de la transferencia de un gen de plantas hacia un ancestro del género *Colletotrichum* y que esta transferencia fue anterior a la división entre monocotiledóneas y dicotiledóneas. De acuerdo a los análisis realizados en los proteomas de diversos organismos, CPLS no presenta ortólogos fuera del género *Colletotrichum* pero sí xenólogos en plantas y bacterias. Los niveles de expresión de *CPLS* detectados durante la infección de *C. graminicola* a maíz resaltan la importancia de esta familia de genes en la patogenicidad de estos hongos.

4. De acuerdo a los experimentos de similitud de secuencia, colocalización y coexpresión realizados, es altamente probable que existan proteínas de *C. graminicola* que imiten a proteínas de su huésped *Zea mays*. Los métodos diseñados para este propósito fueron capaces de discriminar entre miles de secuencias los pares de proteínas que concuerdan con el concepto de molécula imitadora e imitada. Varios de los candidatos encontrados concuerdan con casos de mimetismo molecular descritos en otros sistemas biológicos.

5. Se encontraron varios genes de *C. graminicola* que comparten similitud de secuencia, ubicación subcelular (proteínas) y patrones de expresión bajo condiciones de infección con genes de *Zea mays*. Esto refleja un alto nivel de sincronía entre pares de genes que pertenecen a organismos no relacionados filogenéticamente pero que comparten un estrecho vínculo biológico (interacción planta-patógeno).

REFERENCIAS

Abramovitch RB, Janjusevic R, Stebbins CE, Martin GB. 2006. Type III effector AvrPtoB requires intrinsic E3 ubiquitin ligase activity to suppress plant cell death and immunity. Proc Natl Acad Sci USA. 103:2851–2856.

Abramovitch RB, Kim Y-J, Chen S, Dickman MB, Martin GB. 2003. Pseudomonas type III effector AvrPtoB induces plant disease susceptibility by inhibition of host programmed cell death. EMBO J. 22:60–69.

Agrios G. 2005. Plant Pathology. 5ta ed. New York: Academic Press

Aguileta G, Marthey S, Chiapello H, Lebrun M-H, Rodolphe F, Fournier E, Gendrault-Jacquemard A, Giraud T. 2008. Assessing the Performance of Single-Copy Genes for Recovering Robust Phylogenies. Syst Biol. 57:613–627.

Akinsanmi OA, Chakraborty S, Backhouse D, Simpfendorfer S. 2007. Passage through alternative hosts changes the fitness of *Fusarium graminearum* and *Fusarium pseudograminearum*. Environ. Microbiol. 9:512–520.

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 22:1600–1607.

Alkan N, Meng X, Friedlander G, Reuveni E, Sukno S, Sherman A, Thon M, Fluhr R, Prusky D. 2013. Global Aspects of pacC Regulation of Pathogenicity Genes in *Colletotrichum gloeosporioides* as Revealed by Transcriptome Analysis. Mol. Plant Microbe Interact. 26:1345–1358.

Almagro L, Gómez Ros LV, Belchi-Navarro S, Bru R, Ros Barceló A, Pedreño MA. 2009. Class III peroxidases in plant defence reactions. J. Exp. Bot. 60:377–390.

Alto NM, Dixon JE. 2008. Analysis of Rho-GTPase mimicry by a family of bacterial type III effector proteins. Meth. Enzymol. 439:131–143.

Alto NM, Shao F, Lazar CS, Brost RL, Chua G, Mattoo S, McMahon SA, Ghosh P, Hughes TR, Boone C, et al. 2006. Identification of a bacterial type III effector family with G protein mimicry functions. Cell. 124:133–145.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol. 11:R106.

Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 55:539–552.

Azad RK, Lawrence JG. 2012. Detecting laterally transferred genes. Methods Mol. Biol. 855:281–308.

Bailey JA, O'Connell RJ, Pring RJ, Nash C. 1992. *Colletotrichum*: biology, pathology and control. In: Bailey JA, Jeger MJ, editors. Infection strategies of *Colletotrichum* species. Wallingford, U.K: CAB International. p. 88–120.

Baroncelli R, Sanz-Martín JM, Rech GE, Sukno SA, Thon MR. 2014a. Draft genome sequence of *Colletotrichum sublineola*, a destructive pathogen of cultivated sorghum. Genome Announc. 2:e00540–14.

Baroncelli R, Sreenivasaprasad S, Sukno SA, Thon MR, Holub E. 2014b. Draft genome sequence of *Colletotrichum acutatum* sensu lato (*Colletotrichum fioriniae*). Genome Announc. 2:e00112–e00114.

Bateman A, Bycroft M. 2000. The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). J. Mol. Biol. 299:1113–1119.

Beck M, Heard W, Mbengue M, Robatzek S. 2012. The INs and OUTs of pattern recognition receptors at the cell surface. Curr. Opin. Plant Biol. 15:367–374.

Belbahri L, Calmin G, Mauch F, Andersson JO. 2008. Evolution of the cutinase gene family: evidence for lateral gene transfer of a candidate *Phytophthora* virulence factor. Gene. 408:1–8.

Bellafiore S, Shen Z, Rosso M-N, Abad P, Shih P, Briggs SP. 2008. Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. PLoS Pathog. 4:e1000192.

Bell CD, Soltis DE, Soltis PS. 2005. The age of the angiosperms: a molecular timescale without a clock. Evolution. 59:1245–1258.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2011. GenBank. Nucleic Acids Res. 39:D32–D37.

Bergstrom GC, Nicholson RL. 1999. The Biology of Corn Anthracnose: Knowledge to Exploit for Improved Management. Plant Disease. 83:596–608.

Berken A, Thomas C, Wittinghofer A. 2005. A new family of RhoGEFs activates the Rop molecular switch in plants. Nature. 436:1176–1180.

Bhadauria V, Banniza S, Vandenberg A, Selvaraj G, Wei Y. 2011. EST mining identifies proteins putatively secreted by the anthracnose pathogen *Colletotrichum truncatum*. BMC Genomics. 12:327.

Boggs JM, South AH, Hughes AL. 2012. Phylogenetic analysis supports horizontal gene transfer of L-amino acid oxidase gene in *Streptococcus oligofermentans*. Infect. Genet. Evol. 12:1005–1009.

Bolton MD, van Esse HP, Vossen JH, de Jonge R, Stergiopoulos I, Stulemeijer IJE, van den Berg GCM, Borrás-Hidalgo O, Dekker HL, de Koster CG, et al. 2008. The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. Mol. Microbiol. 69:119–136.

Boquet P. 2000. Small GTP binding proteins and bacterial virulence. Microbes Infect. 2:837–843.

Borziak K, Fleetwood AD, Zhulin IB. 2013. Chemoreceptor gene loss and acquisition via horizontal gene transfer in *Escherichia coli*. J. Bacteriol. 195:3596–3602.

Bowyer P, Clarke BR, Lunness P, Daniels MJ, Osbourn AE. 1995. Host range of a plant pathogenic fungus determined by a saponin detoxifying enzyme. Science. 267:371–374.

Van den Brink J, de Vries RP. 2011. Fungal enzyme sets for plant polysaccharide degradation. Appl. Microbiol. Biotechnol. 91:1477–1492.

Broaders E, Gahan CGM, Marchesi JR. 2013. Mobile genetic elements of the human gastrointestinal tract: potential for spread of antibiotic resistance genes. Gut Microbes. 4:271–280.

Broad Institute of Harvard and MIT. *Colletotrichum* Sequencing Project. Available from: http://www.broadinstitute.org/

Brooks DM, Bender CL, Kunkel BN. 2005. The *Pseudomonas syringae* phytotoxin coronatine promotes virulence by overcoming salicylic acid-dependent defences in *Arabidopsis thaliana*. Mol. Plant Pathol. 6:629–639.

Brown JR. 2003. Ancient horizontal gene transfer. Nat Rev Genet. 4:121–132.

Bryan P, Wang L, Hoskins J, Ruvinov S, Strausberg S, Alexander P, Almog O, Gilliland G, Gallagher T. 1995. Catalysis of a protein folding reaction: Mechanistic implications of the 2.0 ANG. structure of the subtilisin-prodomain complex. Biochemistry. 34:10310–10318.

Bryant MK, Schardl CL, Hesse U, Scott B. 2009. Evolution of a subtilisin-like protease gene family in the grass endophytic fungus *Epichloë festucae*. BMC Evol Biol. 9:168.

Cai S, Lashbrook CC. 2008. Stamen abscission zone transcriptome profiling reveals new candidates for abscission control: enhanced retention of floral organs in transgenic plants overexpressing *Arabidopsis* ZINC FINGER PROTEIN2. Plant Physiol. 146:1305–1321.

Cannon PF, Damm U, Johnston PR, Weir BS. 2012. *Colletotrichum* - current status and future directions. Stud Mycol. 73:181–213.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.

Carlile AJ, Bindschedler LV, Bailey AM, Bowyer P, Clarkson JM, Cooper RM. 2000. Characterization of SNP1, a cell wall-degrading trypsin, produced during infection by *Stagonospora nodorum*. Mol. Plant Microbe Interact. 13:538–550.

Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, et al. 2010. The MetaCyc database of metabolic

pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucl. Acids Res. 38:D473–D479.

Cedzich A, Huttenlocher F, Kuhn BM, Pfannstiel J, Gabler L, Stintzi A, Schaller A. 2009. The protease-associated domain and c-terminal extension are required for zymogen processing, sorting within the secretory pathway, and activity of tomato subtilase 3 (SLSBT3). J Biol Chem. 284:14068–14078.

Chan SL, Low LY, Hsu S, Li S, Liu T, Santelli E, Le Negrate G, Reed JC, Woods VL Jr, Pascual J. 2009. Molecular mimicry in innate immunity: crystal structure of a bacterial TIR domain. J. Biol. Chem. 284:21386–21392.

Chaw S-M, Chang C-C, Chen H-L, Li W-H. 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. J Mol Evol. 58:424–441.

Chen H-Q, Lee M-H, Chung K-R. 2007. Functional characterization of three genes encoding putative oxidoreductases required for cercosporin toxin biosynthesis in the fungus *Cercospora nicotianae*. Microbiology. 153:2781–2790.

Chen L, Xiong Z, Sun L, Yang J, Jin Q. 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res. 40:D641–D645.

Chen Z, Agnew JL, Cohen JD, He P, Shan L, Sheen J, Kunkel BN. 2007. *Pseudomonas syringae* type III effector AvrRpt2 alters *Arabidopsis thaliana* auxin physiology. Proc Natl Acad Sci USA. 104:20131–20136.

Chiang B-Y, Chen T-C, Pai C-H, Chou C-C, Chen H-H, Ko T-P, Hsu W-H, Chang C-Y, Wu W-F, Wang AH-J, et al. 2010. Protein S-thiolation by Glutathionylspermidine (Gsp): the role of *Escherichia coli* Gsp synthetASE/amidase in redox regulation. J. Biol. Chem. 285:25345–25353.

Chichkova NV, Shaw J, Galiullina RA, Drury GE, Tuzhikov AI, Kim SH, Kalkum M, Hong TB, Gorshkova EN, Torrance L, et al. 2010. Phytaspase, a relocalisable cell death promoting plant protease with caspase specificity. EMBO J. 29:1149–1161.

Chilton MD, Drummond MH, Merio DJ, Sciaky D, Montoya AL, Gordon MP, Nester EW. 1977. Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. Cell. 11:263–271.

Choi I-G, Kim S-H. 2007. Global extent of horizontal gene transfer. Proc Natl Acad Sci USA. 104:4489–4494.

Christensen SA, Kolomiets MV. 2011. The lipid language of plant-fungal interactions. Fungal Genet. Biol. 48:4–14.

Chung KR, Shilts T, Ertürk U, Timmer LW, Ueng PP. 2003. Indole derivatives produced by the fungus *Colletotrichum acutatum* causing lime anthracnose and postbloom fruit drop of citrus. FEMS Microbiol. Lett. 226:23–30.

Clark SM, Running MP, Meyerowitz EM. 1995. CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. Development. 121:2057–2067.

Clarkson JM, Charnley AK. 1996. New insights into the mechanisms of fungal pathogenesis in insects. Trends in Microbiology. 4:197–203.

Coffeen WC, Wolpert TJ. 2004. Purification and characterization of serine proteases that exhibit caspase-like activity and are associated with programmed cell death in *Avena sativa*. Plant Cell. 16:857–873.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21:3674–3676.

Cooper SJ, Ashby AM. 1998. Comparison of cytokinin and cytokinin-O-glucoside cleaving beta-glucosidase production in vitro by *Venturia inaequalis* and other phytopathogenic fungi with differing modes of nutrition in planta. Physiol Mol Plant Pathol. 53:61–72.

Cosio C, Dunand C. 2009. Specific functions of individual class III peroxidase genes. J. Exp. Bot. 60:391–408.

Costa RV da, Silva DD da, Cota LV, Parreira DF, Ferreira A da S, Casela CR. 2010. Incidence of *Colletotrichum graminicola* in stalk from maize genotypes. Summa Phytopathologica. 36:122–128.

Damasceno e Silva KJ, De Souza EA, Ishikawa FH. 2007. Characterization of *Colletotrichum lindemuthianum* isolates from the state of ,Minas Gerais, Brazil. J Phytopat. 155:241–247.

Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, et al. 2012. The Top 10 fungal pathogens in molecular plant pathology. Mol. Plant Pathol. 13:414–430.

Deslandes L, Rivas S. 2012. Catch me if you can: bacterial effectors and plant targets. Trends Plant Sci. 17:644–655.

Dickman MB. 2000. *Colletotrichum*. In: Fungal Pathology. The Netherlands, Dordrecht: Kluwer Academic. p. 127–148.

Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, Osorio S, Tohge T, Fernie AR, Feussner I, et al. 2011. Metabolic priming by a secreted fungal effector. Nature. 478:395–398.

Dodds PN. 2010. Genome Evolution in Plant Pathogens. Science. 330:1486–1487.

Dodds PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. Nat. Rev. Genet. 11:539–548.

Doehlemann G, Hemetsberger C. 2013. Apoplastic immunity and its suppression by filamentous plant pathogens. New Phytol. 198:1001–1016.

Doehlemann G, Wahl R, Horst RJ, Voll LM, Usadel B, Poree F, Stitt M, Pons-Kühnemann J, Sonnewald U, Kahmann R, et al. 2008. Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph *Ustilago maydis*. Plant J. 56:181–195.

Döğen A, Kaplan E, Ilkit M, de Hoog GS. 2013. Massive contamination of *Exophiala dermatitidis* and *E. phaeomuriformis* in railway stations in subtropical Turkey. Mycopathologia. 175:381–386.

Donatti AC, Furlaneto-Maia L, Fungaro MHP, Furlaneto MC. 2008. Production and regulation of cuticle-degrading proteases from *Beauveria bassiana* in the presence of *Rhammatocerus schistocercoides* cuticle. Curr Microbiol. 56:256–260.

Doolittle WF. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. 14:307–311.

Doxey AC, McConkey BJ. 2013. Prediction of molecular mimicry candidates in human pathogenic bacteria. Virulence. 4:453–466.

Drayman N, Glick Y, Ben-nun -shaul O, Zer H, Zlotnick A, Gerber D, Schueler-Furman O, Oppenheim A. 2013. Pathogens use structural mimicry of native host ligands as a mechanism for host receptor engagement. Cell Host Microbe. 14:63–73.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, et al. 2011. Geneious v5.5. Available from: http://www.geneious.com

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

Dunaevsky YE, Matveeva AR, Beliakova GA, Domash VI, Belozersky MA. 2007. Extracellular alkaline proteinase of *Colletotrichum gloeosporioides*. Biochemistry Mosc. 72:345–350.

Dutta C, Pan A. 2002. Horizontal gene transfer and bacterial diversity. J Biosci. 27:27–33.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Elde NC, Malik HS. 2009. The evolutionary conundrum of pathogen mimicry. Nature Rev Microbiol. 7:787–797.

FAOSTAT. 2014. Food and Agriculture Organization of the United Nations. Available from: http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39:W29–W37.

Finn RD, Marshall M, Bateman A. 2005. iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. Bioinformatics. 21:410–412.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature. 484:186–194.

Fitzpatrick DA. 2011. Horizontal gene transfer in fungi. FEMS Microbiol Lett. 329:1–8.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, et al. 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41:D808–D815.

Freeman J, Ward E. 2004. *Gaeumannomyces graminis*, the take-all fungus and its relatives. Mol. Plant Pathol. 5:235–252.

Freeman S, Horowitz S, Sharon A. 2001. Pathogenic and Nonpathogenic Lifestyles in *Colletotrichum acutatum* from Strawberry and Other Plants. Phytopathology. 91:986–992.

Frey TJ, Weldekidan T, Colbert T, Wolters PJCC, Hawk JA. 2011. Fitness evaluation of Rcg1, a locus that confers resistance to *Colletotrichum graminicola* (Ces.) G.W. Wils. using near-isogenic maize hybrids. Crop Science. 51:1551.

Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. Nat Genet. 38:953–956.

Galán JE, Zhou D. 2000. Striking a balance: modulation of the actin cytoskeleton by *Salmonella*. Proc Natl Acad Sci USA. 97:8754–8761.

Gangappa S, van Dyk LF, Jewett TJ, Speck SH, Virgin HW 4th. 2002. Identification of the *in vivo* role of a viral bcl-2. J. Exp. Med. 195:931–940.

Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, Takano Y, Kubo Y, Shirasu K. 2013. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol. 197:1236–1249.

Garcia-Vallvé S, Romeu A, Palau J. 2000. Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. Mol. Biol. Evol. 17:352–361.

Geldner N, Friml J, Stierhof YD, Jürgens G, Palme K. 2001. Auxin transport inhibitors block PIN1 cycling and vesicle trafficking. Nature. 413:425–428.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5:R80.

Giraldo MC, Valent B. 2013. Filamentous plant pathogen effectors in action. Nat. Rev. Microbiol. 11:800–814.

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C. 2010. Food Security: The Challenge of Feeding 9 Billion People. Science. 327:812–818.

Gogarten JP. 2003. Gene transfer: gene swapping craze reaches eukaryotes. Curr Biol. 13:R53–R54.

Greenbaum BD, Levine AJ, Bhanot G, Rabadan R. 2008. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS Pathog. 4:e1000079.

Griffiths AJ. 1995. Natural plasmids of filamentous fungi. Microbiol. Rev. 59:673–685.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.

Guo Y, Ni J, Denver R, Wang X, Clark SE. 2011. Mechanisms of molecular mimicry of plant CLE peptide ligands by the parasitic nematode *Globodera rostochiensis*. Plant Physiol. 157:476–484.

Gyles C, Boerlin P. 2014. Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. Vet. Pathol. 51:328–340.

Hall R. 1994. Molecular mimicry. In: Baker, J.R, Muller, R, Rollinson, D, editors. Advances in Parasitology. Academic Press. p. 314.

Hamilton JMU, Simpson DJ, Hyman SC, Ndimba BK, Slabas AR. 2003. Ara12 subtilisin-like protease from *Arabidopsis thaliana*: purification, substrate specificity and tissue localization. Biochem J. 370:57–67.

Hasegawa H, Holm L. 2009. Advances and pitfalls of protein structural alignment. Curr Opin Struct Biol. 19:341–348.

Heger A, Holm L. 2000. Rapid automatic detection and alignment of repeats in protein sequences. Proteins. 41:224–237.

Hemetsberger C, Herrberger C, Zechmann B, Hillmer M, Doehlemann G. 2012. The *Ustilago maydis* effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity. PLoS Pathog. 8:e1002684.

Hendy MD, Penny D. 1989. A Framework for the quantitative study of evolutionary trees. Syst Biol. 38:297–309.

Heuer H, Smalla K. 2007. Horizontal gene transfer between bacteria. Environ Biosafety Res. 6:3–13.

Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S. 2009. Emerging concepts in effector biology of plant-associated organisms. Mol. Plant Microbe Interact. 22:115–122.

Holm L, Rosenstrom P. 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38:W545–W549.

Van der Hoorn RAL, Kamoun S. 2008. From guard to decoy: A new model for perception of plant pathogen effectors. Plant Cell. 20:2009–2017.

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35:W585–W587.

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, et al. 2012. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 40:D306–D312.

Ikemura H, Inouye M. 1988. In vitro processing of pro-subtilisin produced in *Escherichia coli*. J Biol Chem. 263:12959–12963.

Inderbitzin P, Asvarak T, Turgeon BG. 2010. Six new genes required for production of T-toxin, a polyketide determinant of high virulence of *Cochliobolus heterostrophus* to maize. Mol. Plant Microbe Interact. 23:458–472.

Irelan JT, Selker EU. 1996. Gene silencing in filamentous fungi: RIP, MIP and quelling. J. Genet. 75:313–324.

Janjusevic R, Abramovitch RB, Martin GB, Stebbins CE. 2006. A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase. Science. 311:222–226.

Jogl G, Hsiao Y-S, Tong L. 2004. Structure and function of carnitine acyltransferases. Ann. N. Y. Acad. Sci. 1033:17–29.

Jones JDG, Dangl JL. 2006. The plant immune system. Nature. 444:323–329.

De Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, van der Krol S, Shibuya N, Joosten MHAJ, Thomma BPHJ. 2010. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science. 329:953–955.

De Jonge R, van Esse HP, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, Zhang Z, Usami T, Lievens B, Subbarao KV, et al. 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc Natl Acad Sci USA. 109:5110–5115.

De Jonge R, Thomma BPHJ. 2009. Fungal LysM effectors: extinguishers of host immunity? Trends in Microbiology. 17:151–157.

Jorda L, Coego A, Conejero V, Vera P. 1999. A genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. J Biol Chem. 274:2360–2365.

Kaku H, Nishizawa Y, Ishii-Minami N, Akimoto-Tomiyama C, Dohmae N, Takio K, Minami E, Shibuya N. 2006. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. Proc Natl Acad Sci USA. 103:11086–11091.

Kamenidou S, Jain R, Hari K, Robertson JM, Fletcher J. 2013. The Microbial Rosetta Stone Central Agricultural Database: An Information Resource on High-Consequence Plant Pathogens. Plant Disease. 97:1097–1102.

Kamoun S. 2007. Groovy times: filamentous pathogen effectors revealed. Curr. Opin. Plant Biol. 10:358–365.

Kämper J, Kahmann R, Bölker M, Ma L-J, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Müller O, et al. 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature. 444:97–101.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40:D109–D114.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Katsir L, Schilmiller AL, Staswick PE, He SY, Howe GA. 2008. COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. Proc Natl Acad Sci USA. 105:7100–7105.

Kawano Y, Kaneko-Kawano T, Shimamoto K. 2014. Rho family GTPase-dependent immunity in plants and animals. Front. Plant Sci. 5:522.

Kazan K, Lyons R. 2014. Intervention of phytohormone pathways by pathogen effectors. Plant Cell. 26:2285–2309.

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that *ad hoc* assumptions for choice of matrix are not justified. BMC Evol Biol. 6:29.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 9:605–618.

Keen NT. 1990. Gene-for-gene complementarity in plant-pathogen interactions. Annu. Rev. Genet. 24:447–463.

Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 4:363–371.

Kemen E, Kemen AC, Rafiqi M, Hempel U, Mendgen K, Hahn M, Voegele RT. 2005. Identification of a protein from rust fungi transferred from haustoria into infected plant cells. Mol. Plant Microbe Interact. 18:1130–1139.

Kerscher O, Felberbaum R, Hochstrasser M. 2006. Modification of proteins by ubiquitin and ubiquitin-like proteins. Annu. Rev. Cell Dev. Biol. 22:159–180.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36.

Kim SE, Moon JS, Choi WS, Lee SH, Kim SU. 2012. Monitoring of horizontal gene transfer from agricultural microorganisms to soil bacteria and analysis of microbial community in soils. J Microbiol Biotechnol. 22:563–566.

Kitahara K, Miyazaki K. 2013. Revisiting bacterial phylogeny: Natural and experimental evidence for horizontal gene transfer of 16S rRNA. Mob Genet Elements. 3:e24210.

Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, Ver Loren van Themaat E, van Themaat EVL, van der Does HC, Hacquard S, Stüber K, Will I, et al. 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. PLoS Pathog. 8:e1002643.

Kleemann J, Takahara H, Stüber K, O'Connell R. 2008. Identification of soluble secreted proteins from appressoria of *Colletotrichum higginsianum* by analysis of expressed sequence tags. Microbiology. 154:1204–1217.

Klink BU, Barden S, Heidler TV, Borchers C, Ladwein M, Stradal TEB, Rottner K, Heinz DW. 2010. Structure of *Shigella* IpgB2 in complex with human RhoA: implications for the mechanism of bacterial guanine nucleotide exchange factor mimicry. J. Biol. Chem. 285:17197–17208.

Koeck M, Hardham AR, Dodds PN. 2011. The role of effectors of biotrophic and hemibiotrophic fungi in infection. Cell. Microbiol. 13:1849–1857.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. 7:474.

Kombrink A. 2014. Functional analysis of LysM effectors secreted by fungal plant pathogens. *PhD tesis*. Wageningen University. Available from: http://www.wageningenur.nl/nl/show/Functional-analysis-of-LysM-effectors-secreted-by-fungal-plant-pathogens.htm

Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36:6688–6719.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567–580.

Kuraku S, Qiu H, Meyer A. 2012. Horizontal transfers of Tc1 elements between teleost fishes and their vertebrate parasites, lampreys. Genome Biol Evol. 4:929–936.

Kwon SJ, Jin HC, Lee S, Nam MH, Chung JH, Kwon SI, Ryu C-M, Park OK. 2009. GDSL lipase-like 1 regulates systemic resistance associated with ethylene signaling in *Arabidopsis*. Plant J. 58:235–245.

Lamb C, Dixon RA. 1997. The oxidative burst in plant disease resistance. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48:251–275.

Landschoot PJ, Jackson N. 1989. *Magnaporthe poae* sp. nov., a hyphopodiate fungus with a *Phialophora* anamorph from grass roots in the United States. Mycol Res. 93:59–62.

Lane BG. 1994. Oxalate, germin, and the extracellular matrix of higher plants. FASEB J. 8:294–301.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23:2947–2948.

Lassmann T, Hayashizaki Y, Daub CO. 2011. SAMStat: monitoring biases in next generation sequencing data. Bioinformatics. 27:130–131.

Latunde-Dada AO. 2001. *Colletotrichum*: tales of forcible entry, stealth, transient confinement and breakout. Mol Plant Pathol. 2:187–198.

Lawrence JG, Ochman H. 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. 44:383–397.

Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, depamphilis CW. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Mol Biol Evol. 22:1948–1963.

Lee DS, Kim BK, Kwon SJ, Jin HC, Park OK. 2009. Arabidopsis GDSL lipase 2 plays a role in pathogen defense via negative regulation of auxin signaling. Biochem. Biophys. Res. Commun. 379:1038–1042.

Lenné J. 2002. Some major plant diseases. In: Waller J., Lenné J., Waller S., editors. Plant Pathologist's Pocketbook. 3ra ed. UK: CABI.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078–2079.

Li H, Shen JJ, Zheng ZL, Lin Y, Yang Z. 2001. The Rop GTPase switch controls multiple developmental processes in *Arabidopsis*. Plant Physiol. 126:670–684.

Little D, Gouhier-Darimont C, Bruessow F, Reymond P. 2007. Oviposition by pierid butterflies triggers defense responses in *Arabidopsis*. Plant Physiol. 143:784–800.

Liu W, Liu J, Ning Y, Ding B, Wang X, Wang Z, Wang G-L. 2013. Recent progress in understanding PAMP- and effector-triggered immunity against the rice blast fungus *Magnaporthe oryzae*. Mol Plant. 6:605–620.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2013. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42:D490-D495

Van Loon L, Pierpoint W, Boller T, Conejero V. 1994. Recommendations for naming plant pathogenesis-related proteins. Plant Mol Biol Rep. 12:245–264.

Lowe RGT, Howlett BJ. 2012. Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. PLoS Pathog. 8:e1002515.

Lucking R, Huhndorf S, Pfister DH, Plata ER, Lumbsch HT. 2009. Fungi evolved right on track. Mycologia. 101:810–822.

Ludin P, Nilsson D, Mäser P. 2011. Genome-wide identification of molecular mimicry candidates in parasites. PLoS ONE. 6:e17546.

Ludwikow A, Gallois P, Sadowski J. 2004. Ozone-induced oxidative stress response in *Arabidopsis*: transcription profiling by microarray approach. Cell. Mol. Biol. Lett. 9:829–842.

Luo Q, Pagel P, Vilne B, Frishman D. 2011. DIMA 3.0: Domain Interaction Map. Nucleic Acids Res. 39:D724–D729.

Lu T, Yao B, Zhang C. 2012. DFVF: database of fungal virulence factors. Database. 2012:bas032.

MacMillan J. 2001. Occurrence of gibberellins in vascular plants, fungi, and bacteria. J. Plant Growth Regul. 20:387–442.

Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. 464:367–373.

Marcet-Houben M, Gabaldón T. 2010. Acquisition of prokaryotic genes by fungal genomes. Trends Genet. 26:5–8.

Marri PR, Hao W, Golding GB. 2007. The role of laterally transferred genes in adaptive evolution. BMC Evol. Biol. 7 Suppl 1:S8.

Marshall R, Kombrink A, Motteram J, Loza-Reyes E, Lucas J, Hammond-Kosack KE, Thomma BPHJ, Rudd JJ. 2011. Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. Plant Physiol. 156:756–769.

Marthey S, Aguileta G, Rodolphe F, Gendrault A, Giraud T, Fournier E, Lopez-Villavicencio M, Gautier A, Lebrun M-H, Chiapello H. 2008. Funybase: a Fungal phylogenomic database. BMC Bioinformatics. 9:456.

McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, et al. 2006. AgBase: a functional genomics resource for agriculture. BMC Genomics. 7:229.

Mehrabi R, Bahkali AH, Abd-Elsalam KA, Moslem M, Ben M'barek S, Gohari AM, Jashni MK, Stergiopoulos I, Kema GHJ, de Wit PJGM. 2011. Horizontal gene and chromosome transfer in plant pathogenic fungi affecting host range. FEMS Microbiol. Rev. 35:542–554.

Meichtry J, Amrhein N, Schaller A. 1999. Characterization of the subtilase gene family in tomato (*Lycopersicon esculentum* Mill.). Plant Mol Biol. 39:749–760.

Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, Terauchi R, Nishizawa Y, Shibuya N, Thomma BPHJ, et al. 2012. Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. Plant Cell. 24:322–335.

Miersch O, Bohlmann H, Wasternack C. 1999. Jasmonates and related compounds from *Fusarium oxysporum*. Phytochemistry. 50:517–523.

Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. 2009. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res. 38:D204–D210.

Milani NA, Lawrence DP, Arnold AE, VanEtten HD. 2012. Origin of pisatin demethylase (PDA) in the genus *Fusarium*. Fungal Genet. Biol. 49:933–942.

Mills LJ, Van Staden J. 1978. Extraction of cytokinins from maize, smut tumors of maize and *Ustilago maydis* cultures. Physiological Plant Pathology. 13:73–80.

Mims CW, Vaillancourt LJ. 2002. Ultrastructural characterization of infection and colonization of maize leaves by *Colletotrichum graminicola*, and by a *C. graminicola* pathogenicity mutant. Phytopathology. 92:803–812.

Ministerio de Agricultura, Alimentación y Medio Ambiente de España. 2014. Cereales - Cultivos herbáceos e industriales - Producciones agrícolas - Agricultura - magrama.es. Available from: http://www.magrama.gob.es/es/agricultura/temas/producciones-agricolas/cultivos-herbaceos/cereales/default.aspx#para2

Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, Narusaka Y, Kawakami N, Kaku H, Shibuya N. 2007. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. Proc Natl Acad Sci USA. 104:19613–19618.

Mohr PG, Cahill DM. 2007. Suppression by ABA of salicylic acid and lignin accumulation and the expression of multiple genes, in *Arabidopsis* infected with *Pseudomonas syringae* pv. tomato. Funct. Integr. Genomics. 7:181–191.

Moreira D, Philippe H. 2000. Molecular phylogeny: pitfalls and progress. Int. Microbiol. 3:9–16.

Münch S, Lingner U, Floss DS, Ludwig N, Sauer N, Deising HB. 2008. The hemibiotrophic lifestyle of *Colletotrichum* species. J. Plant Physiol. 165:41–51.

Murayama K, Kato-Murayama M, Hosaka T, Sotokawauchi A, Yokoyama S, Arima K, Shirouzu M. 2012. Crystal structure of cucumisin, a subtilisin-like endoprotease from *Cucumis melo* L. J Mol Biol. 423:386–396.

Murphy AM, Pryce-Jones E, Johnstone K, Ashby AM. 1997. Comparison of cytokinin production in vitro by *Pyrenopeziza brassicae* with other plant pathogens. Physiol Mol Plant Pathol. 50:53–65.

Muszewska A, Taylor JW, Szczesny P, Grynberg M. 2011. Independent subtilases expansions in fungi associated with animals. Mol Biol Evol. 28:3395–3404.

Najafzadeh MJ, Dolatabadi S, Saradeghi Keisari M, Naseri A, Feng P, de Hoog GS. 2013. Detection and identification of opportunistic *Exophiala* species using the rolling circle amplification of ribosomal internal transcribed spacers. J. Microbiol. Methods. 94:338–342.

Nakagawa M, Ueyama M, Tsuruta H, Uno T, Kanamaru K, Mikami B, Yamagata H. 2010. Functional analysis of the cucumisin propeptide as a potent inhibitor of its mature enzyme. J Biol Chem. 285:29797–29807.

Namiki F, Matsunaga M, Okuda M, Inoue I, Nishi K, Fujita Y, Tsuge T. 2001. Mutation of an arginine biosynthesis gene causes reduced pathogenicity in *Fusarium oxysporum* f. sp. *melonis*. Mol. Plant Microbe Interact. 14:580–584.

Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JDG. 2006. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science. 312:436–439.

Nembaware V, Seoighe C, Sayed M, Gehring C. 2004. A plant natriuretic peptide-like gene in the bacterial pathogen *Xanthomonas axonopodis* may induce hyper-hydration in the plant host: a hypothesis of molecular mimicry. BMC Evol Biol. 4:10.

Ngugi HK, Scherm H. 2006. Mimicry in plant-parasitic fungi. FEMS Microbiol. Lett. 257:171–176.

Nikolaidis N, Doran N, Cosgrove DJ. 2014. Plant expansins in bacteria and fungi: evolution by horizontal gene transfer and independent domain fusion. Mol. Biol. Evol. 31:376–386.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302:205–217.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature. 405:299–304.

O'Connell RJ, Bailey JA, Richmond DV. 1985. Cytology and physiology of infection of *Phaseolus vulgaris* by *Colletotrichum lindemuthianum*. Physiol Plant Pathol. 27:75–98.

O'Connell RJ, Bailey JA, Vose IR, Lamb CJ. 1986. Immunogold labelling of fungal antigens in cells of *Phaseolus vulgaris* infected by *Colletotrichum lindemuthianum*. Physiol Mol Plant Pathol. 28:99–105.

O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, Damm U, Buiate EA, Epstein L, Alkan N, et al. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat Genet. 44:1060–1065.

Oerke E-C. 1999. Estimated crop losses due to pathogens, animal pests and weeds. In: WEBER E-CO-WDS, editor. Crop Production and Crop Protection. Amsterdam:

Elsevier. p. 72–741. Available from: http://www.sciencedirect.com/science/article/pii/B9780444820952500099

Oh IS, Park AR, Bae MS, Kwon SJ, Kim YS, Lee JE, Kang NY, Lee S, Cheong H, Park OK. 2005. Secretome analysis reveals an *Arabidopsis* lipase involved in defense against *Alternaria brassicicola*. Plant Cell. 17:2832–2847.

Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F, et al. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog. 8:e1003037.

Ohta Y, Inouye M. 1990. Pro-subtilisin E: purification and characterization of its autoprocessing to active subtilisin E *in vitro*. Mol Microbiol. 4:295–304.

Oldstone MB. 1989. Molecular mimicry as a mechanism for the cause and a probe uncovering etiologic agent(s) of autoimmune disease. Curr. Top. Microbiol. Immunol. 145:127–135.

Oldstone MB. 1998. Molecular mimicry and immune-mediated diseases. FASEB J. 12:1255–1265.

Oliva R, Win J, Raffaele S, Boutemy L, Bozkurt TO, Chaparro-Garcia A, Segretin ME, Stam R, Schornack S, Cano LM, et al. 2010. Recent developments in effector biology of filamentous plant pathogens. Cell. Microbiol. 12:705–715.

Olivieri F, Zanetti ME, Oliva CR, Covarrubias AA, Casalongué CA. 2002. Characterization of an extracellular serine protease of *Fusarium eumartii* and its action on pathogenesis related proteins. Eur J Plant Pathol. 108:63–72.

Olsen AN, Skriver K. 2003. Ligand mimicry? Plant-parasitic nematode polypeptide with similarity to CLAVATA3. Trends Plant Sci. 8:55–57.

Ottmann C, Rose R, Huttenlocher F, Cedzich A, Hauske P, Kaiser M, Huber R, Schaller A. 2009. Structural basis for Ca2+-independence and activation by homodimerization of tomato subtilase 3. Proc Natl Acad Sci USA. 106:17223–17228.

Palaversic B, Jukic M, Buhinicek I, Vragolovic A, Kozic Z. 2009. Breeding maize for resistance to stalk anthracnose. Maydica. 54:229–232.

Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC. 2010. CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology. 20:1574–1584.

Park C-H, Chen S, Shirsekar G, Zhou B, Khang CH, Songkumarn P, Afzal AJ, Ning Y, Wang R, Bellizzi M, et al. 2012. The *Magnaporthe oryzae* effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. Plant Cell. 24:4748–4762.

Passardi F, Longet D, Penel C, Dunand C. 2004. The class III peroxidase multigenic family in rice and its evolution in land plants. Phytochemistry. 65:1879–1893.

Patel N, Hamamouch N, Li C, Hewezi T, Hussey RS, Baum TJ, Mitchum MG, Davis EL. 2010. A nematode effector protein similar to annexins in host plants. J. Exp. Bot. 61:235–248.

Patnaik D, Khurana P. 2001. Germins and germin like proteins: an overview. Indian J. Exp. Biol. 39:191–200.

Pearce G, Yamaguchi Y, Barona G, Ryan CA. 2010. A subtilisin-like protein from soybean contains an embedded, cryptic signal that activates defense-related genes. Proc Natl Acad Sci USA. 107:14921–14925.

Peden JF. 1999. Analysis of codon usage. *PhD tesis*. University of Nottingham. Available from: http://codonw.sourceforge.net/JohnPedenThesisPressOpt_water.pdf

Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. Nucleic Acids Res. 38:W23–W28.

Perfect SE, Green JR. 2001. Infection structures of biotrophic and hemibiotrophic fungal plant pathogens. Mol Plant Pathol. 2:101–108.

Perfect SE, Hughes HB, O'Connell RJ, Green JR. 1999. *Colletotrichum*: A model genus for studies on pathology and fungal-plant interactions. Fungal Genet. Biol. 27:186–198.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8:785–786.

Pillitteri LJ, Peterson KM, Horst RJ, Torii KU. 2011. Molecular profiling of stomatal meristemoids reveals new component of asymmetric cell division and commonalities among stem cell populations in *Arabidopsis*. Plant Cell. 23:3260–3275.

Pretsch K, Kemen A, Kemen E, Geiger M, Mendgen K, Voegele R. 2013. The rust transferred proteins-a new family of effector proteins exhibiting protease inhibitor function. Mol. Plant Pathol. 14:96–107.

Pruitt KD, Tatusova T, Brown GR, Maglott DR. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 40:D130–D135.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. 2011. The Pfam protein families database. Nucl Acids Res. 40:D290–D301.

Rademacher W. 1992. Occurrence of gibberellins in different species of the fungal genera *Sphaceloma* and *Elsinoe*. Phytochemistry. 31:4155–4157.

Rafiqi M, Ellis JG, Ludowici VA, Hardham AR, Dodds PN. 2012. Challenges and progress towards understanding the role of effectors in plant-fungal interactions. Curr. Opin. Plant Biol. 15:477–482.

Rafiqi M, Jelonek L, Akum NF, Zhang F, Kogel K-H. 2013. Effector candidates in the secretome of *Piriformospora indica*, a ubiquitous plant-associated fungus. Front Plant Sci. 4:228. doi: 10.3389/fpls.2013.00228.

Rambaut A, Drummond AJ. 2007. Tracer v1.4. Available from: http://beast.bio.ed.ac.uk/Tracer

Rawlings ND, Barrett AJ, Bateman A. 2011. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucl Acids Res. 40:D343–D350.

Raymond D. 1964. Molecular mimicry: antigen sharing by parasite and host and its consequences. Am Naturalist. 98:129–149.

R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org/

Réblová M, Gams W, Seifert KA. 2011. Monilochaetes and allied genera of the Glomerellales, and a reconsideration of families in the Microascales. Stud. Mycol. 68:163–191.

Rech GE, Sanz-Martín JM, Anisimova M, Sukno SA, Thon MR. 2014. Natural selection on coding and noncoding DNA sequences is associated with virulence genes in a plant pathogenic fungus. Genome Biol Evol. 6:2368–2379.

Reddy PV, Lam CK, Belanger FC. 1996. Mutualistic fungal endophytes express a proteinase that is homologous to proteases suspected to be important in fungal pathogenicity. Plant Physiol. 111:1209–1218.

Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. 2006. Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. Curr Biol. 16:1857–1864.

Richards TA, Hirt RP, Williams BAP, Embley TM. 2003. Horizontal gene transfer and the evolution of parasitic protozoa. Protist. 154:17–32.

Richards TA, Leonard G, Soanes DM, Talbot NJ. 2011. Gene transfer into the fungi. Fungal Biol Rev. 25:98–110.

Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. 2009. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. Plant Cell. 21:1897–1911.

Rohmer L, Hocquet D, Miller SI. 2011. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. Trends Microbiol. 19:341–348.

Rolland T, Neuvéglise C, Sacerdot C, Dujon B. 2009. Insertion of horizontally transferred genes within conserved syntenic regions of yeast genomes. PLoS ONE. 4:e6515.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Rosebrock TR, Zeng L, Brady JJ, Abramovitch RB, Xiao F, Martin GB. 2007. A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity. Nature. 448:370–374.

Rose R, Schaller A, Ottmann C. 2010. Structural features of plant subtilases. Plant Signal Behav. 5:180–183.

Rosewich UL, Kistler HC. 2000. Role of horizontal gene transfer in the evolution of fungi. Annu Rev Phytopathol. 38:325–363.

Rougeron A, Giraud S, Razafimandimby B, Meis JF, Bouchara J-P, Klaassen CHW. 2013. Different colonization patterns of *Aspergillus terreus* in patients with cystic fibrosis. Clin. Microbiol. Infect. 20:327-333

Saeki K, Okuda M, Hatada Y, Kobayashi T, Ito S, Takami H, Horikoshi K. 2000. Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic *Bacillus* spp.: enzymatic properties, sequences, and evolutionary relationships. Biochem Biophys Res Commun. 279:313–319.

Sanz-Martín JM, Vargas WA, Thon MR, Sukno S. 2012. Identificación y caracterización funcional de proteínas efectoras de *C. graminicola*. In: San Feliu de Guixols, Spain: MICELIO Meeting: Spanish Network of Filamentous Fungi Researchers.

Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol. Evol. 25:537–546.

Schmidt HA, Strimmer K, Vingron M, Von Haeseler A. 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18:502–504.

Schmitt I, Lumbsch HT. 2009. Ancient horizontal gene transfer from bacteria enhances biosynthetic capabilities of fungi. PLoS ONE. 4:e4437.

Schomburg I, Chang A, Placzek S, Söhngen C, Rother M, Lang M, Munaretto C, Ulas S, Stelzer M, Grote A, et al. 2013. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res. 41:D764–D772.

Schornack S, Minsavage GV, Stall RE, Jones JB, Lahaye T. 2008. Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. New Phytol. 179:546–556.

Schrödinger L. 2010. The PyMOL molecular graphics system, Version 1.3r1.

Shimizu T, Nakano T, Takamizawa D, Desaki Y, Ishii-Minami N, Nishizawa Y, Minami E, Okada K, Yamane H, Kaku H, et al. 2010. Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. Plant J. 64:204–214.

Shimodaira, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–1247.

Shivas R., Tan Y. 2009. A taxonomic re-assessment of *Colletotrichum acutatum*, introducing *C. fioriniae* comb. et stat. nov. and *C. simmondsii* sp. nov. Fungal Divers. 39:111–122.

Siezen RJ, Leunissen JA. 1997. Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci. 6:501–523.

Siezen RJ, de Vos WM, Leunissen JA, Dijkstra BW. 1991. Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. Protein Eng. 4:719–737.

Sikora S, Strongin A, Godzik A. 2005. Convergent evolution as a mechanism for pathogenic adaptation. Trends Microbiol. 13:522–527.

Singer AU, Schulze S, Skarina T, Xu X, Cui H, Eschen-Lippold L, Egler M, Srikumar T, Raught B, Lee J, et al. 2013. A pathogen type III effector with a novel E3 ubiquitin ligase architecture. PLoS Pathog. 9:e1003121.

Slot JC, Rokas A. 2010. Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. Proc Natl Acad Sci USA. 107:10136–10141.

Smith K, Nadeau K, Bradley M, Walsh C, Fairlamb AH. 1992. Purification of glutathionylspermidine and trypanothione synthetases from *Crithidia fasciculata*. Protein Sci. 1:874–883.

Smith SA, Beaulieu JM, Donoghue MJ. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. Proc Natl Acad Sci USA. 107:5897–5902.

Soanes D, Richards TA. 2014. Horizontal gene transfer in eukaryotic plant pathogens. Annu Rev Phytopathol. 52:583–614.

Spaepen S, Vanderleyden J, Remans R. 2007. Indole-3-acetic acid in microbial and microorganism-plant signaling. FEMS Microbiol. Rev. 31:425–448.

Sreedhar L, Kobayashi DY, Bunting TE, Hillman BI, Belanger FC. 1999. Fungal proteinase expression in the interaction of the plant pathogen *Magnaporthe poae* with its host. Gene. 235:121–129.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.

Stebbins CE, Galán JE. 2001. Structural mimicry in bacterial virulence. Nature. 412:701–705.

Stein A, Panjkovich A, Aloy P. 2009. 3did Update: domain-domain and peptide-mediated interactions of known 3D structure. Nucleic Acids Res. 37:D300–D304.

Stephenson SA, Hatfield J, Rusu AG, Maclean DJ, Manners JM. 2000. CgDN3: an essential pathogenicity gene of *Colletotrichum gloeosporioides* necessary to avert a hypersensitive-like response in the host *Stylosanthes guianensis*. Mol. Plant Microbe Interact. 13:929–941.

Stöver BC, Müller KF. 2010. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics. 11:7.

Strange RN, Scott PR. 2005. Plant disease: a threat to global food security. Annu Rev Phytopathol. 43:83–116.

Sukno SA, García VM, Shaw BD, Thon MR. 2008. Root infection and systemic colonization of maize by *Colletotrichum graminicola*. Appl. Environ. Microbiol. 74:823–832.

Sun B-F, Xiao J-H, He S, Liu L, Murphy RW, Huang D-W. 2013. Multiple interkingdom horizontal gene transfers in *Pyrenophora* and closely related species and their contributions to phytopathogenic lifestyles. PLoS ONE. 8:e60029.

Swatek KN, Graham K, Agrawal GK, Thelen JJ. 2011. The 14-3-3 isoforms chi and epsilon differentially bind client proteins from developing *Arabidopsis* seed. J. Proteome Res. 10:4076–4087.

Takahara H, Huser A, O'Connell R. 2012. Two arginine biosynthesis genes are essential for pathogenicity of *Colletotrichum higginsianum* on *Arabidopsis*. Mycology. 3:54–64.

Takano Y. 2004. Molecular genetic studies on infection mechanism in *Colletotrichum lagenarium*. J Gen Plant Pathol. 70:390–390.

Takeda N, Sato S, Asamizu E, Tabata S, Parniske M. 2009. Apoplastic plant subtilases support arbuscular mycorrhiza development in *Lotus japonicus*. Plant J. 58:766–777.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

Talhinhas P, Sreenivasaprasad S, Neves-Martins J, Oliveira H. 2005. Molecular and phenotypic analyses reveal association of diverse *Colletotrichum acutatum* groups and a low level of *C. gloeosporioides* with olive anthracnose. Appl. Environ. Microbiol. 71:2987–2998.

Tanaka E, Koga H, Mori M, Mori M. 2011. Auxin poduction by the rice blast fungus and its localization in host tissue. J Phytopat. 159:522–530.

Techtmann SM, Lebedinsky AV, Colman AS, Sokolova TG, Woyke T, Goodwin L, Robb FT. 2012. Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. Front Microbiol. 3:132.

Theodoro GF, Peres N a. R, Verona L a. F. 2004. Outbreak of postbloom fruit drop of citrus, caused by *Colletotrichum acutatum*, in Santa Catarina State, southern Brazil. Plant Pathol. 53:253–253.

Thomma BPHJ, Nürnberger T, Joosten MHAJ. 2011. Of PAMPs and effectors: the blurred PTI-ETI dichotomy. Plant Cell. 23:4–15.

Thon MR, Nuckles EM, Takach JE, Vaillancourt LJ. 2002. CPR1: a gene encoding a putative signal peptidase that functions in pathogenicity of*Colletotrichum graminicola* to maize. Mol. Plant Microbe Interact. 15:120–128.

Tiburcio RA, Costa GGL, Carazzolle MF, Mondego JMC, Schuster SC, Carlson JE, Guiltinan MJ, Bailey BA, Mieczkowski P, Meinhardt LW, et al. 2010. Genes acquired by

horizontal transfer are potentially involved in the evolution of phytopathogenicity in *Moniliophthora perniciosa* and *Moniliophthora roreri*, two of the major pathogens of cacao. J Mol Evol. 70:85–97.

Tornero P, Conejero V, Vera P. 1996. Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: Similarity of functional domains to subtilisin-like endoproteases. Proc Natl Acad Sci USA. 93:6332–6337.

Tornero P, Conejero V, Vera P. 1997. Identification of a new pathogen-induced member of the subtilisin-like processing protease family from plants. J Biol Chem. 272:14412–14419.

Tricker PJ, Gibbings JG, Rodríguez López CM, Hadley P, Wilkinson MJ. 2012. Low relative humidity triggers RNA-directed de novo DNA methylation and suppression of genes controlling stomatal development. J. Exp. Bot. 63:3799–3813.

Tripathi LP, Sowdhamini R. 2006. Cross genome comparisons of serine proteases in *Arabidopsis* and rice. BMC Genomics. 7:200.

Valette-Collet O, Cimerman A, Reignault P, Levis C, Boccara M. 2003. Disruption of *Botrytis cinerea* pectin methylesterase gene *Bcpme1* reduces virulence on several host plants. Mol. Plant Microbe Interact. 16:360–367.

Vargas WA, Sanz-Martín JM, Rech GE, Rivera LP, Benito EP, Díaz-Mínguez JM, Thon MR, Sukno SA. 2012a. Plant defense mechanisms are activated during biotrophic and necrotrophic development of *Colletotricum graminicola* in maize. Plant Physiol. 158:1342–1358.

Vargas WA, Sanz-Martín JM, Thon MR, Sukno S. 2012b. Identification and functional characterization of CgEP1, a novel pathogenicity factor from *Colletotrichum graminicola*. In: Marburg, Germany: European Conference on Fungal Genetics. p. 160.

Venard C, Vaillancourt L. 2007. Penetration and colonization of unwounded maize tissues by the maize anthracnose pathogen *Colletotrichum graminicola* and the related nonpathogen C. sublineolum. Mycologia. 99:368–377.

Van Vu B, Itoh K, Nguyen QB, Tosa Y, Nakayashiki H. 2012. Cellulases belonging to glycoside hydrolase families 6 and 7 contribute to the virulence of *Magnaporthe oryzae*. Mol. Plant Microbe Interact. 25:1135–1141.

Wang J, Lee C, Replogle A, Joshi S, Korkin D, Hussey R, Baum TJ, Davis EL, Wang X, Mitchum MG. 2010. Dual roles for the variable domain in protein trafficking and host-specific recognition of *Heterodera glycines* CLE effector proteins. New Phytol. 187:1003–1017.

Wang J, Replogle A, Hussey R, Baum T, Wang X, Davis EL, Mitchum MG. 2011. Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode *Heterodera schachtii*. Mol. Plant Pathol. 12:177–186.

Wang X, Mitchum MG, Gao B, Li C, Diab H, Baum TJ, Hussey RS, Davis EL. 2005. A parasitism gene from a plant-parasitic nematode with function similar to CLAVATA3/ESR (CLE) of *Arabidopsis thaliana*. Mol. Plant Pathol. 6:187–191.

Wapinski I, Pfeffer A, Friedman N, Regev A. 2007. Natural history and evolutionary principles of gene duplication in fungi. Nature. 449:54–61.

Warren HL. 1975. Kernel infection, seedling blight, and wilt of maize caused by *Colletotrichum graminicola*. Phytopathology. 65.

Wenzl P, Wong L, Kwang-won K, Jefferson RA. 2005. A functional screen identifies lateral transfer of beta-glucuronidase (gus) from bacteria to fungi. Mol. Biol. Evol. 22:308–316.

Winnenburg R, Urban M, Beacham A, Baldwin TK, Holland S, Lindeberg M, Hansen H, Rawlings C, Hammond-Kosack KE, Köhler J. 2008. PHI-base update: additions to the pathogen host interaction database. Nucleic Acids Res. 36:D572–D576.

Withers-Martinez C, Suarez C, Fulle S, Kher S, Penzo M, Ebejer J-P, Koussis K, Hackett F, Jirgensons A, Finn P, et al. 2012. Plasmodium subtilisin-like protease 1 (SUB1): insights into the active-site structure, specificity and function of a pan-malaria drug target. Int J Parasitol. 42:597–612.

Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA. 86:6201–6205.

Wolf YI, Makarova KS, Yutin N, Koonin EV. 2012. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. Biol Direct. 7:46.

Wu S-C, Ham K-S, Darvill AG, Albersheim P. 1997. Deletion of two *Endo* -β-1,4-xylanase genes reveals additional isozymes secreted by the rice Blast fungus. Mol. Plant Microbe Interact. 10:700–708.

Yamada T. 1993. The Role of Auxin in Plant-Disease Development. Annu Rev Phytopathol. 31:253–273.

Yang J, Huang X, Tian B, Wang M, Niu Q, Zhang K. 2005. Isolation and characterization of a serine protease from the nematophagous fungus, *Lecanicillium psalliotae*, displaying nematicidal activity. Biotechnol Lett. 27:1123–1128.

Yang Z. 2002. Small GTPases: versatile signaling switches in plants. Plant Cell. 14 Suppl:S375–S388.

Yasmin S, Alcazar-Fuoli L, Gründlinger M, Puempel T, Cairns T, Blatzer M, Lopez JF, Grimalt JO, Bignell E, Haas H. 2012. Mevalonate governs interdependency of ergosterol and siderophore biosyntheses in the fungal pathogen *Aspergillus fumigatus*. Proc Natl Acad Sci USA. 109:E497–E504.

Yi M, Valent B. 2013. Communication between filamentous pathogens and plants at the biotrophic interface. Annu Rev Phytopathol. 51:587–611.

Yuan J-B, Zhang X-J, Liu C-Z, Wei J-K, Li F-H, Xiang J-H. 2013. Horizontally transferred genes in the genome of Pacific white shrimp, *Litopenaeus vannamei*. BMC Evol. Biol. 13:165.

Zdobnov EM, Apweiler R. 2001. InterProScan-an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 17:847–848.

Zhang G, Liang Y. 2013. Improvement of fruiting body production in *Cordyceps militaris* by molecular assessment. Arch. Microbiol. 195:579–585.

Zhang N, Castlebury LA, Miller AN, Huhndorf SM, Schoch CL, Seifert KA, Rossman AY, Rogers JD, Kohlmeyer J, Volkmann-Kohlmeyer B, et al. 2006. An overview of the systematics of the Sordariomycetes based on a four-gene phylogeny. Mycologia. 98:1076–1087.

Zhang N, Rossman A, Seifert K, Bennett J, Cai G, Cai L, Hillman B, Hyde K, Luo J, Manamgoda D, et al. 2013. Impacts of the International Code of Nomenclature for algae, fungi and plants (Melbourne Code) on the scientific names of plant pathogenic fungi. Online.APSnet Feature. American Phytopathological Society. Available from: http://www.apsnet.org/publications/apsnetfeatures/Pages/Melbourne.aspx

Zhang X-C, Wu X, Findley S, Wan J, Libault M, Nguyen HT, Cannon SB, Stacey G. 2007. Molecular evolution of lysin motif-type receptor-like kinases in plants. Plant Physiol. 144:623–636.

Zhao C, Johnson BJ, Kositsup B, Beers EP. 2000. Exploiting secondary growth in *Arabidopsis*. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. Plant Physiol. 123:1185–1196.

Zhao Z, Liu H, Wang C, Xu J-R. 2013. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 14:274.

Zhao Z, Liu H, Wang C, Xu J-R. 2014. Correction: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 15:6.

Zheng Q, Wang X-J. 2008. GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. Nucleic Acids Res. 36:W358–W363.

Zhou J-M, Chai J. 2008. Plant pathogenic bacterial type III effectors subdue host responses. Curr. Opin. Microbiol. 11:179–185.

Zipfel C. 2008. Pattern-recognition receptors in plant innate immunity. Curr. Opin. Immunol. 20:10–16.

Zipfel C, Felix G. 2005. Plants and animals: a different taste for microbes? Curr. Opin. Plant Biol. 8:353–360.