



**INSTITUTO POTOSINO DE INVESTIGACIÓN  
CIENTÍFICA Y TECNOLÓGICA, A.C.**

**POSGRADO EN CIENCIAS EN BIOLOGÍA MOLECULAR**

**El gen *phr-1*, que codifica para una fotoliasa de  
*Trichoderma atroviride*, se regula  
transcripcionalmente por elementos de  
respuesta a luz no canónicos**

Tesis que presenta

**Mayte Guadalupe Cervantes Badillo**

Para obtener el grado de

**Doctora en Ciencias en Biología Molecular**

**Director de la Tesis:**

**Dr. J. Sergio Casas Flores**

San Luis Potosí, S.L.P., Junio de 2013



## Constancia de aprobación de la tesis

La tesis “**El gen *phr-1*, que codifica para una fotoliasa de *Trichoderma atroviride*, se regula transcripcionalmente por elementos de respuesta a luz no canónicos**” presentada para obtener el Grado de Doctora en Ciencias en Biología Molecular, fue elaborada por **Mayte Guadalupe Cervantes Badillo** y aprobada el **05 de Junio de 2013** por los suscritos, designados por el Colegio de Profesores de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C.

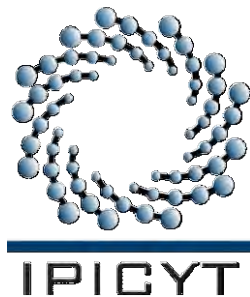
Dr. J. Sergio Casas Flores  
(Director de la tesis)

Dra. Lina Raquel Riego Ruiz  
(Miembro del Comité Tutorial)

Dr. Gerardo Rafael Argüello Astorga  
(Miembro del Comité Tutorial)

Dra. Irene Beatriz Castaño Navarro  
(Miembro del Comité Tutorial)

Dr. Alfredo Heriberto Herrera Estrella  
(Miembro del Comité Tutorial)



## **Créditos Institucionales**

Esta tesis fue elaborada en el Laboratorio de Genómica Funcional y Comparativa de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la dirección del Dr. J. Sergio Casas Flores.

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología (No. de registro 165178) y del Instituto Potosino de Investigación Científica y Tecnológica, A. C (oficio DG-065/2012).

Este trabajo fue apoyado por el proyecto SEP-CONACYT 83798 a Alfredo Herrera-Estrella y Sergio Casas-Flores.



# Instituto Potosino de Investigación Científica y Tecnológica, A.C.

## Acta de Examen de Grado

El Secretario Académico del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta Primero del Libro 055 de Actas de Exámenes de Grado del Programa de Doctorado en Ciencias en Biología Molecular está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 14 días del mes de junio del año 2013, se reunió a las 12:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

<b>Dr. Alfredo Heriberto Herrera Estrella</b>	<b>Presidente</b>	<b>CINVESTAV Irapuato</b>
<b>Dr. Gerardo Rafael Argüello Astorga</b>	<b>Secretario</b>	<b>IPICYT</b>
<b>Dr. Braulio Gutiérrez Medina</b>	<b>Sinodal</b>	<b>IPICYT</b>
<b>Dr. J. Sergio Casas Flores</b>	<b>Sinodal</b>	<b>IPICYT</b>
<b>Dra. Lina Raquel Riego Ruiz</b>	<b>Sinodal</b>	<b>IPICYT</b>

a fin de efectuar el examen, que para obtener el Grado de:

**DOCTORA EN CIENCIAS EN BIOLOGÍA MOLECULAR**

sustentó la C.

**Mayte Guadalupe Cervantes Badillo**

sobre la Tesis intitulada:

*El gen **phr-1**, que codifica para una fotoliasa de **Trichoderma atroviride**, se regula transcripcionalmente por elementos de respuesta a luz no canónicos*

que se desarrolló bajo la dirección de

**Dr. J. Sergio Casas Flores**


El Jurado, después de deliberar, determinó

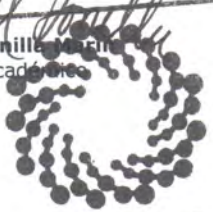
**APROBARLA**

Dándose por terminado el acto a las 14:38 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe el Secretario Académico del Instituto.

A petición de la interesada y para los fines que a la misma convengan, se extiende el presente documento en la ciudad de San Luis Potosí, S.L.P., México, a los 14 días del mes de junio de 2013.

  
**Mtra. Ivonne Lizette Cuevas Vélez**  
Jefa del Departamento del Posgrado

  
**Dr. Marcial Bonilla**  
Secretario Académico

  
INSTITUTO POTOSINO  
DE INVESTIGACIÓN  
CIENTÍFICA Y TECNOLÓGICA, A.C.

**IPICYT**  
SECRETARIA ACADEMICA

## Dedicatorias

Dedico este trabajo, que representa muchos años de esfuerzo, perseverancia y aprendizaje, con todo mi amor a mi familia. Gracias a mis padres, Gloria Badillo y Luis Cervantes, y mis hermanas Melissa y Hada, por toda su alegría, cariño, apoyo y consejos que me dan en cada momento de mi vida, sin importar la distancia.

Para aquellas personas que con su ejemplo han sido una inspiración en varios aspectos de mi vida y la han enriquecido en gran manera: Noemi, Lilí, Yadira, Sergio, Sergio, Martín, Raúl (TM).

*“Un amigo es una persona con la que se puede pensar en voz alta”*

Ralph W. Emerson

*Actúa con humildad, nunca se sabe lo suficiente y nunca es demasiado tarde para aprender algo nuevo y mejorar*

## Agradecimientos

Deseo agradecer al CONACyT por la beca académica otorgada para realizar mi trabajo de doctorado y al IPICYT por las facilidades que se me proporcionaron en cuanto a instalaciones y equipo para la realización del presente trabajo.

Agradezco a los profesores de la División de Biología Molecular por la oportunidad de aprender de su experiencia, tanto en la teoría como en la práctica.

Especialmente a mi director de tesis, el Dr. Sergio Casas, quien ha sido un gran amigo desde mi ingreso al IPICYT. Gracias por toda la paciencia, la asesoría, los tips, las discusiones y la planeación de los experimentos. Pero sobre todo, por la confianza y el apoyo tanto profesional como personal.

Jefesín! como bien dices, este trabajo es producto de una *corresponsabilidad*.

Gracias a todos los miembros de mi comité evaluador, la Dra. Irene Castaño, la Dra. Lina Riego, el Dr. Gerardo Argüello, el Dr. Alfredo Herrera y el Dr. Braulio Gutiérrez, quienes me han proporcionado una excelente retroalimentación científica a través de sus comentarios y observaciones que enriquecieron mi tesis y contribuyeron a mi crecimiento científico.

Me gustaría agradecer al Dr. Rubén López Revilla y doblemente al Dr. Gerardo Argüello y al Dr. Braulio Gutiérrez por compartir conmigo tantas conversaciones interesantes, por su apoyo y sus consejos acompañados siempre con una sonrisa.

Muchas gracias a los técnicos Ing. María Isabel Isordia, Biol. Mireya Sánchez, Biol. Salvador Ambriz y QFB Rosalba Castillo, así como a la asistente de la División de Biología Molecular, Sandra Torres, por su buena disposición para auxiliarme cuando así lo requerí durante mi estancia en el IPICYT.

Durante el desarrollo de este trabajo tuve la fortuna de pertenecer a un grupo que trabaja en equipo, donde el éxito de uno es el éxito de todos. Gracias a mis compañeros Edith Uresti, Miguel Silva, Miguel Salas, Gemita Rose Cristóbal, Tania Muñoz, Jesús (Chuy) Gallardo, Macario Osorio, Magnolia Estrada, Gerardo Bernal, Lizeth Díaz, Aída Rodríguez, Zayra Delgado y Marcelita López, por hacer del trabajo una actividad lúdica y porque he aprendido mucho de cada uno de ustedes.

Gracias a todos mis amigos por su optimismo y por contagiar ese espíritu de superación y perseverancia, aún en situaciones difíciles: Candy Ramírez, Noemí Hernández, Ana Iris Peña, Yadira Díaz, Claudia Peresson, Josef GJ, Nehemías Leija (UASLP), NoHemí Carreras (LANGEBIO) y Sergio Pérez (UNAM).

Gracias por el apoyo de los buenos amigos de la División de Ciencias Ambientales, Dr. Vladimir Escobar y Dr. Luis Felipe Cházaro, y sus estudiantes Areli Bejarano y Mariana Mancera, y por la alegría que irradian en todo momento.

# Contenido

Constancia de aprobación de la tesis.....	ii
Créditos institucionales.....	iii
Acta de examen.....	iv
Dedicatorias.....	v
Agradecimientos.....	vi
Lista de tablas.....	viii
Lista de figuras.....	ix
Anexos.....	x
Abreviaturas.....	xi
Resumen.....	xiii
Abstract.....	xiv
Introducción.....	1
Efecto de la luz en los seres vivos.....	1
Fotorreceptores.....	3
Fitocromos.....	4
Rodopsinas.....	5
Receptores de luz azul.....	6
Fotorreceptores tipo PAS/LOV: WC-1 y VIVID.....	7
Fototropinas.....	9
Criptocromo/Fotoliasa.....	11
BLUF.....	13
Regulación transcripcional mediada por luz.....	14
Respuesta a la luz azul en <i>Trichoderma</i> spp.....	20
Resultados Parte I	
The <i>Trichoderma atroviride</i> photolyase-encoding gene is transcriptionally regulated by non-canonical light response elements.....	28
Resultados Parte II	
Ensayos de cambio de movilidad electroforética.....	64
Identificación del sitio de inicio de la transcripción del gen <i>bld-2</i> , diseño y construcción de fusiones traduccionales con distintas versiones del promotor de <i>bld-2</i> con el gen reportero <i>lacZ</i> .....	68
Discusión.....	74
Materiales y Métodos.....	84
Referencias.....	88
Anexos.....	99

## Lista de tablas

Tabla 1. Número de copias del gen <i>hph</i> en las transformantes, calculado con el método $2^{-\Delta\Delta Ct}$ .....	62
Tabla 2. Oligonucleótidos usados en este trabajo.....	63
Tabla 3. Sondas diseñadas para los EMSA en la PLRR y UCR del promotor de <i>phr-1</i> .....	84
Tabla 4. Secuencia del adaptador y los oligonucleótidos usados en el 5'RLM-RACE.....	85
Tabla 5. Oligonucleótidos usados para fusiones traduccionales de <i>blt-2</i> con el gen <i>lacZ</i> .....	86



## Lista de figuras

Figura 1	Estructura de los fitocromos.....	4
Figura 2	Estructura de la rodopsina.....	6
Figura 3	Fotoquímica de la formación del aducto flavin-cisteinyl en los dominios LOV.....	8
Figura 4	Estructura de los fotorreceptores tipo PAS/LOV.....	9
Figura 5	Estructura de las fototropinas.....	10
Figura 6	Estructura de la familia Criptocromo/Fotoliasa.....	12
Figura 7	Estructura de los fotorreceptores BLUF.....	14
Figura 8	Estructura de WC-1 y WC-2.....	18
Figura 9	Vía de señalización de la luz azul en <i>Neurospora crassa</i> .....	20
Figura 10	Fotoconidiación en <i>Trichoderma atroviride</i> .....	21
Figura 11	Modelo de la vía de señalización de la luz azul en <i>Trichoderma</i> .....	27
Figura 12	La luz induce la unión de proteínas a las regiones PLRR y UCR <i>in vitro</i> .....	66
Figura 13	Características del producto putativo del gen <i>bld-2</i> .....	68
Figura 14	Obtención del amplicón para la identificación del sitio de inicio de la transcripción del gen <i>bld-2</i> de <i>T. atroviride</i> .....	69
Figura 15	Análisis del promotor del gen <i>bld-2</i> de <i>T. atroviride</i> .....	71
Figura 16	Escrutinio de las construcciones P <i>bld-2</i> :: <i>lacZ</i> con enzimas de restricción.....	73

## Anexos

- I The plant growth-promoting fungus *Aspergillus ustus* promotes growth and induces resistance against different lifestyle pathogens in *Arabidopsis thaliana* 99
- II Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma* 110

## Abreviaturas

APE	Albino Proximal Element (Elemento Albino Proximal)
bld	<u>b</u> lue <u>l</u> ight <u>d</u> ownregulated (Regulados negativamente por luz azul)
blu	<u>b</u> lue <u>l</u> ight <u>u</u> pregulated (Regulados positivamente por luz azul)
BLR	Blue Light Regulator (Regulador de luz azul)
BLRC	BLR Complex (Complejo Regulador de Luz Azul)
CPD	Cyclobutane Pyrimidine Dimer (Dímeros de Ciclobutano de Pirimidina)
DA	Dominio de Activación
ELRE	Early Light Response Element (Elemento temprano de respuesta a luz)
ELRG	Early Light Response Genes (Genes tempranos de respuesta a luz)
FAD	Flavin Adenin Dinucleótido
FRQ	Frequency
<i>g</i>	Fuerza de gravedad
kb	Kilobase
kDa	Kilo Dalton
LRE	Light Response Element (Elemento de respuesta a luz)
LLRE	Late Light Response Element (Elemento tardío de respuesta a luz)
LLRG	Late Light Response Genes (Genes tardíos de respuesta a luz)
LOV	Light, Oxygen, Voltage (Luz, Oxígeno, Voltaje)
μg	Microgramo
μl	Microlitro
μm	Micrómetro
NLS	Nuclear Localization Signal (Señal de Localización Nuclear)
ORF	Open Reading Frame (Marco de Lectura Abierto)
PAS	<u>P</u> ER ( <i>Drosophila</i> Period) - <u>A</u> RNT (vertebrate aryl hydrocarbon receptor nuclear translocator) - <u>S</u> IM ( <i>Drosophila</i> single-minded)
pb	Pares de bases
PCR	Polymerase Chain Reaction (Reacción en Cadena de la Polimerasa)

PLRR	Potential Light Responsive Region (Región Potencial Responsiva a la Luz)
RLM-RACE	RNA Ligase Mediated Rapid Amplification of cDNA Ends (Amplificación Rápida de los Extremos de cDNA Mediado por Ligasa de RNA)
TSS	Transcription Start Site (Sitio de Inicio de la Transcripción)
VVD	Vivid
WC	White Collar
WCC	White Collar Complex (Complejo White Collar)
ZnF	Dedos de zinc

## Resumen

**El gen *phr-1*, que codifica para una fotoliasa de *Trichoderma atroviride*, se regula transcripcionalmente por elementos de respuesta a luz no canónicos**

**Resumen** Las proteínas BLR-1 y BLR-2 de *Trichoderma atroviride* son los homólogos de WC-1 y WC-2 de *Neurospora crassa*, dos factores de transcripción involucrados en la regulación de genes que responden a luz azul. BLR-1 y BLR-2 son esenciales para la fotoinducción del gen *phr-1*, que codifica para una fotoliasa, cuyo promotor exhibe secuencias similares elementos de respuesta a luz bien caracterizados en *Neurospora*, incluyendo el Elemento Albino Proximal (APE) y el Elemento de Respuesta a Luz (LRE). No obstante que *phr-1* ha sido utilizado extensamente como marcador de inducción por luz azul en *Trichoderma*, la función de esos presuntos elementos reguladores no se ha examinado experimentalmente. El núcleo del LRE descrito en *N. crassa* está conformado por dos cajas GATA separadas por un número pequeño y variable de nucleótidos, a las cuales se une transitoriamente el complejo WC-1/WC-2 (WCC) tras la aplicación de un estímulo luminoso. Por medio de deleciones seriadas en el extremo 5' del promotor de *phr-1*, así como por mutaciones puntuales del LRE putativo, delimitamos una región de ~50 pb que media la respuesta transcripcional a la luz azul. La región crítica para la fotorrespuesta identificada en este estudio contiene tres motivos CGATC, dos de ellos en polaridad opuesta a los sitios de unión canónicos del WCC. Los experimentos de inmunoprecipitación de cromatina mostraron que la proteína BLR-2 se encuentra unida a lo largo del promotor de *phr-1* en la oscuridad, en tanto que la aplicación de un pulso de luz azul resultó en un decremento en la unión de BLR-2. Nuestros resultados sugieren que BLR-2 y probablemente BLR-1, están localizados sobre el promotor de *phr-1* en la oscuridad, en una conformación propicia para llevar a cabo su función transcripcional en respuesta a la luz.

**PALABRAS CLAVE:** *Trichoderma*; *phr-1*; luz azul; complejo White-Collar; Elemento de Respuesta a la Luz; proteínas BLR

## Abstract

**The *Trichoderma atroviride* photolyase-encoding gene is transcriptionally regulated by non-canonical light response elements**

**Abstract** The BLR-1 and BLR-2 proteins of *Trichoderma atroviride* are the *Neurospora crassa* homologs of White Collar-1 and -2, two transcription factors involved in the regulation of genes by blue-light. BLR-1 and BLR-2 are essential for photoinduction of *phr-1*, a photolyase-encoding gene whose promoter exhibits sequences similar to well-characterized light regulatory elements of *Neurospora*, including the Albino Proximal Element (APE) and the Light Response Element (LRE). However, despite the fact that this gene has been extensively used as a blue light induction marker in *Trichoderma*, the function of these putative regulatory elements has not been proven. The described LRE core in *N. crassa* comprises two close but variably spaced GATA-boxes to which a WC-1/-2 complex binds transiently upon application of a light stimulus. Using 5' serial deletions of the *phr-1* promoter, as well as point mutations of putative LREs, we were able to delimit a ~50 bp long region mediating the transcriptional response to blue-light. The identified light-responsive region contained three CGATC motifs, two of them displaying opposite polarity to canonical WCC binding sites. Chromatin immunoprecipitation experiments showed that the BLR-2 protein binds along the *phr-1* promoter in darkness, whereas the application of a blue light pulse resulted in decreased BLR-2 binding to the promoter. Our results suggest that BLR-2 and probably BLR-1 are located on the *phr-1* promoter in darkness ready to perform their function as transcriptional complex in response to light.

**KEY WORDS:** *Trichoderma*; *phr-1*; blue light; White-Collar-Complex, Light Response Element; BLR proteins

# Introducción

## ***Efecto de la luz en los seres vivos***

Detectar el entorno y garantizar una respuesta celular adecuada son retos cruciales a los que se enfrentan todos los organismos. Estos procesos ocurren de manera secuencial: la célula recibe la señal, la cual se transduce, y se genera una respuesta ante el estímulo (Bahn *et al.*, 2007). La luz del sol es un factor importante del medio ambiente, ya que a lo largo del día varía tanto en calidad como en cantidad, puede ser un estímulo positivo como fuente de energía, o por el contrario, puede tener efectos perjudiciales, particularmente en longitudes de onda corta como la luz ultravioleta (UV). La luz es una fuente esencial de energía para los organismos fotosintéticos como las plantas y algunos microorganismos, pero también sirve como una señal importante del medio ambiente para especies no fotosintéticas; por lo tanto, la percepción de la luz se ha conservado evolutivamente a través de los reinos, desde las arqueas y los hongos hasta los seres humanos (Purschwitz *et al.*, 2006; Bahn *et al.*, 2007; Rodríguez-Romero *et al.*, 2010). En las eubacterias (no fotosintéticas) y cianobacterias (fotosintéticas), por ejemplo, la luz regula diversos procesos como el ciclo circadiano, la actividad transcripcional, el metabolismo, la fototaxis, la fotosíntesis y la adaptación cromática complementaria, un tipo de fotomorfogénesis en cianobacterias (Asayama, 2006; Montgomery, 2007).

En animales, la luz regula el encarrilamiento del ciclo circadiano (del latín *circa*, que significa alrededor de y *dies*, que significa día), un mecanismo de sincronización que funciona para coordinar procesos metabólicos (metabolismo de glucosa, lípidos y drogas), fisiológicos (la frecuencia cardíaca, regulación de hormonas de crecimiento, la remodelación de huesos y el sistema inmune) y de comportamiento (ciclos de sueño-vigilia, regulación del estrés) (Falciatore y Bowler, 2005; Gallego y Virshup, 2007). El sistema circadiano confiere una ventaja

selectiva, ya que permite que los organismos se anticipen a los cambios periódicos que ocurren en el medio ambiente. El comportamiento rítmico controlado por el reloj interno persiste aún bajo condiciones de luz constante (Falciatore y Bowler, 2005). La desregulación de los ritmos circadianos trae como consecuencia la disminución de la aptitud y la supervivencia de los organismos en la naturaleza (Gallego y Virshup, 2007).

En las plantas, las condiciones de luz/oscuridad regulan una gran diversidad de procesos como la fotosíntesis, la floración, el desarrollo del hipocotilo, el ciclo circadiano y la regulación transcripcional de un gran número de genes. Las plántulas que crecen bajo condiciones de luz presentan fotomorfogénesis, desarrollo de hipocotilos cortos, apertura y expansión de los cotiledones, cloroplastos fotosintéticamente activos, y expresión diferencial de genes. Por otra parte, en la oscuridad ocurre la etiolación de la plántula, alargamiento de los hipocotilos, los cotiledones están cerrados y sin dilatar, y los etioplastos son no fotosintéticos (Terzaghi y Cashmore, 1995; Fankhauser y Staiger, 2002; Chen *et al.*, 2004).

En los hongos varios procesos son regulados por la luz, entre estos la regulación del ritmo circadiano, el fototropismo, la conidiación, el metabolismo secundario, la síntesis de pigmentos y el desarrollo sexual (Kritskii *et al.*, 2005; Purschwitz *et al.*, 2006; Herrera-Estrella y Horwitz, 2007; Chen *et al.*, 2010; Rodríguez-Romero *et al.*, 2010). La luz también permite que los hongos orienten su crecimiento celular, ya que permite que el organismo perciba si está en el suelo, dentro de un huésped, o si se expone al aire y/o a otros factores de estrés (Herrera-Estrella y Horwitz, 2007). Los hongos son capaces de detectar la luz en un intervalo amplio del espectro luminoso que cubre longitudes de onda de más de diez órdenes de magnitud, desde el ultravioleta hasta la luz roja lejana. Esto se ve reflejado en una variedad de complejos proteicos asociados a cromóforos (flavinas, pterinas, etc)



denominados fotorreceptores, para monitorear la cantidad y la calidad de la luz en su entorno (Purschwitz *et al.*, 2006).

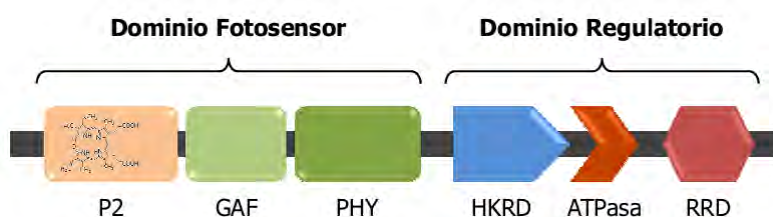
### ***Fotorreceptores***

Las bacterias, los hongos, las plantas y los animales cuentan con una variedad de fotorreceptores que les permiten percibir y responder adecuadamente a la variabilidad espacial, temporal y espectral de la luz en un entorno natural. Los fotorreceptores son proteínas acopladas a cromóforos capaces de captar la luz, generar una señal y propagarla dentro de la célula para estimular una respuesta biológica adecuada (Falciatore y Bowler, 2005; Herrera-Estrella y Horwitz, 2007; Corrochano, 2011). La percepción de la luz causa un cambio conformacional en el cromóforo, seguido de la activación de una cadena de transducción de señales que por último genera una respuesta biológica (Falciatore y Bowler, 2005). La secuencia de cambios conformacionales que experimenta el cromóforo en respuesta a la luz es reversible y se conoce como *fotociclo* (Chen *et al.*, 2010). Los principales cromóforos son flavinas, tetrapirroles y retinal, que absorben longitudes de onda específicas e inducen una reacción en la proteína (Purschwitz *et al.*, 2006; Herrera-Estrella y Horwitz, 2007). Muchos fotorreceptores se han descrito y clasificado en diferentes familias en función de la estructura química de sus cromóforos (Falciatore y Bowler, 2005; Purschwitz *et al.*, 2006; Corrochano, 2007; Herrera-Estrella y Horwitz, 2007).

La vía de transducción de la señal luminosa es equivalente en animales y hongos, por lo que éstos últimos han sido excelentes organismos modelo para estudiar los procesos regulados por luz. Con la creciente disponibilidad de los genomas secuenciados, así como una gran cantidad de herramientas genéticas, moleculares y bioinformáticas, se han identificado nuevos fotorreceptores en los hongos (Galagan *et al.*, 2003; Borkovich *et al.*, 2004; Kubicek *et al.*, 2011), los cuales se describen a continuación.

## Fitocromos

Los fitocromos pertenecen a una amplia familia de fotorreceptores distribuidos en bacterias fotosintéticas y no fotosintéticas, plantas y hongos (Falciatore y Bowler, 2005). Los fitocromos de hongos utilizan a la biliverdina como cromóforo, la cual está unida covalentemente a la proteína a través de un enlace tioéter (Rockwell *et al.*, 2006). El fitocromo tiene dos dominios: un dominio N-terminal fotosensorial (entrada de señal), y un dominio C-terminal (regulatorio) involucrado en dimerización y activación de la transducción de la señal (Falciatore y Bowler, 2005; Rockwell *et al.*, 2006; Rodríguez-Romero *et al.*, 2010) (Figura 1).



**Figura 1. Estructura de los fitocromos.** Los fitocromos presentan una región fotosensora en el extremo N-terminal y una región regulatoria en el extremo C-terminal. El extremo N-terminal, involucrado en la unión del cromóforo de bilina, está compuesto por los dominios PAS (P2), GAF y PHY. En los fitocromos de plantas el cromóforo está unido en el dominio GAF; sin embargo, en hongos y bacterias éste se encuentra unido en el dominio P2. En el C-terminal están presentes un dominio de cinasa de histidinas (HKRD), un dominio de ATPasa y un dominio regulador de la respuesta (RRD) (Modificado de Rodríguez-Romero *et al.*, 2010).

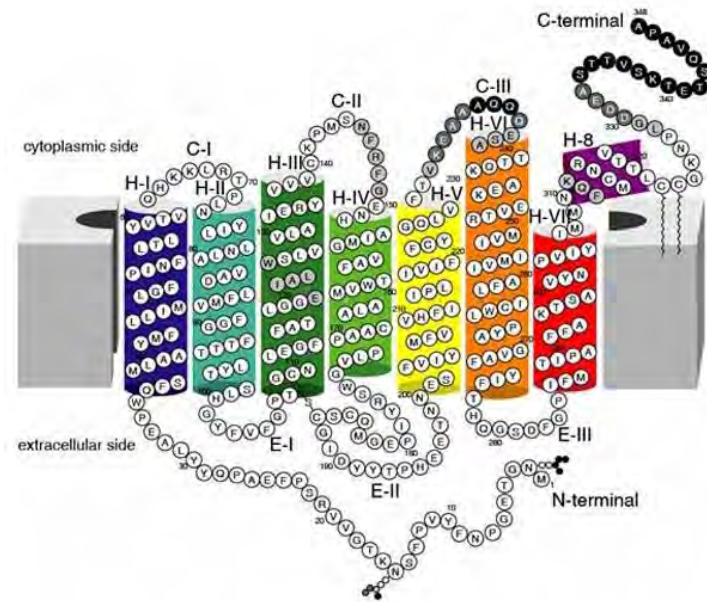
Los fitocromos se encuentran en dos formas relativamente estables, espectralmente distintas e interconvertibles: una forma Pr, cuya absorción máxima es en el rojo ( $\lambda_{\text{max}}$  665 nm) y una forma Pfr, cuyo máximo de absorción es en el rojo lejano ( $\lambda_{\text{max}}$  730 nm). La forma Pr es inactiva para la mayoría de las respuestas, mientras que la forma Pfr es la forma activa (Rockwell *et al.*, 2006). Un ejemplo opuesto ocurre en la cianobacteria *Synechocystis*, cuyo fitocromo es activo en su forma Pr (Yeh *et al.*, 1997).

Los fitocromos participan en procesos de desarrollo, crecimiento y expresión diferencial de genes en las plantas (Mockler *et al.*, 1999; Mazzaella *et al.*, 2001; Tepperman *et al.*, 2001; Nagy y Schafer, 2002; Tepperman *et al.*, 2004), la conidiación, expresión génica y metabolismo secundario en hongos (Blumenstein *et al.*, 2005; Rosales-Saavedra *et al.*, 2006; Atoui *et al.*, 2010), la fotomorfogénesis en cianobacterias, el crecimiento y la fototaxis en procariotes (Falciatore y Bowler 2005; Montgomery 2007), entre otros.

### ***Rodopsinas***

Las rodopsinas pertenecen a un grupo numeroso y antiguo de fotorreceptores distribuidos en bacterias, hongos y animales. Estos fotorreceptores están formados por la apoproteína opsina, la cual consta de siete hélices transmembranales, unida al cromóforo retinal (Bieszke *et al.*, 1999; Purschwitz *et al.*, 2006; Salom *et al.*, 2006; Herrera-Estrella y Horwitz, 2007) (Figura 2). Existen dos tipos de rodopsinas, con estructura similar pero con muy poca similitud en su secuencia. Las rodopsinas tipo I unen *all-trans* retinal y funcionan como transportadores de iones o receptores sensoriales en microbios. La absorción de luz verde ( $\lambda_{\max}$  534 nm) lleva a la isomerización *all-trans* → *11-cis* del retinal (Purschwitz *et al.*, 2006). Las rodopsinas tipo II son los fotorreceptores para la visión en animales y unen *11-cis* retinal (Corrochano, 2007). La absorción de un fotón luz ( $\lambda_{\max}$  498 nm) lleva a la isomerización *11-cis* → *all-trans* del retinal (Palczewski, 2006).

La absorción de la luz por las rodopsinas funciona ya sea para transmitir una señal, como es el caso de las rodopsinas visuales de los ojos animales, o con fines de conservación de energía, como las rodopsinas presentes en arqueas (Falciatore y Bowler, 2005; Purschwitz *et al.*, 2006; Montgomery, 2007).



**Figura 2. Estructura de la rodopsina.** El retinal es un cromóforo que se une a las opsinas, las cuales constan de siete dominios transmembranales (indicados por rectángulos de colores), con el extremo N-terminal hacia el lado extracelular y el extremo C-terminal hacia el lado citoplásmico. La membrana celular se esquematiza en gris (Tomado de Salom *et al.*, 2006).

En los hongos se han descrito opsinas en los ascomicetos y los basidiomicetos (Rodríguez-Romero *et al.*, 2010). En *Neurospora crassa*, *Cryptococcus neoformans* y *Gibberella*, las mutantes no presentan defectos obvios en los procesos regulados por luz (Bieszke *et al.*, 1999; Prado *et al.*, 2004; Idnurm y Heitman, 2005); sin embargo, los trabajos en *Aspergillus nidulans*, *Leptosphaeria maculans* y *Allomyces reticulatus* sugieren que las opsinas podrían participar en la biosíntesis de carotenoides, como bombas de protones o en la fototaxis de las zoosporas, respectivamente (Saranak y Foster, 1997; Idnurm y Howlett, 2001; Waschuk *et al.*, 2005).

### **Receptores de luz azul**

La respuesta de un ser vivo a la luz azul fue reportada por primera vez en 1881 por Darwin, cuando describió el fototropismo inducido por luz azul en las plantas.

Desde entonces, la capacidad de percibir y responder a la luz azul ( $\lambda$  400-500 nm) se ha descrito en los tres dominios: eucariotas, bacterias y arqueas (Falciatore y Bowler, 2005; Purschwitz *et al.*, 2006; Rodríguez-Romero *et al.*, 2010).

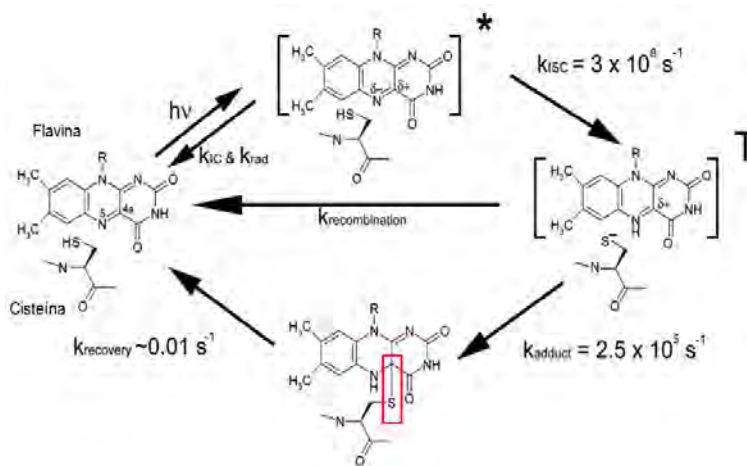
Los fotorreceptores de luz azul son flavoproteínas, ya que utilizan flavinas como cromóforos. Los fotorreceptores de luz azul que se han descrito a la fecha son: los fotosensores tipo PAS/LOV: WC-1 y VIVID y las fototropinas, los criptocromos/fotoliasas y los BLUF (Chen *et al.*, 2004; Falciatore y Bowler, 2005; Herrera-Estrella y Horwitz, 2007).

### ***Fotorreceptores tipo PAS/LOV: WC-1 y VIVID***

Los dominios LOV constituyen una subclase de la familia PAS (Per, Arnt, Sim). Los dominios PAS son módulos de señalización importantes que detectan la luz, potencial redox, oxígeno, fuerza motriz de protones y ligandos pequeños, como una manera de monitorear los cambios del nivel de energía total de una célula (Taylor y Zhulin, 1999). Una vez que el dominio PAS percibe la señal del ambiente, la transduce a un dominio receptor dentro de la misma proteína, o a otra proteína vía interacción proteína-proteína. Los dominios PAS forman heterodímeros con mayor frecuencia y algunas veces homodímeros (Zhong *et al.*, 2003).

Los dominios LOV utilizan una flavina como cromóforo y funcionan como un módulo para percibir luz azul en bacterias, plantas y hongos (Zhong *et al.*, 2002; Crosson *et al.*, 2003; Ogura *et al.*, 2008). La irradiación con luz azul causa la formación de un enlace covalente, llamado aducto, entre la molécula de flavina (C4a) y una cisteína conservada en el dominio LOV (Figura 3). Posteriormente, el aducto decae térmicamente al estado basal, debido a una desprotonación de la posición N5 de la flavina (Corchnoy *et al.*, 2003; Zoltowski *et al.*, 2009). Este cambio estructural dependiente de luz/oscuridad regula la actividad del

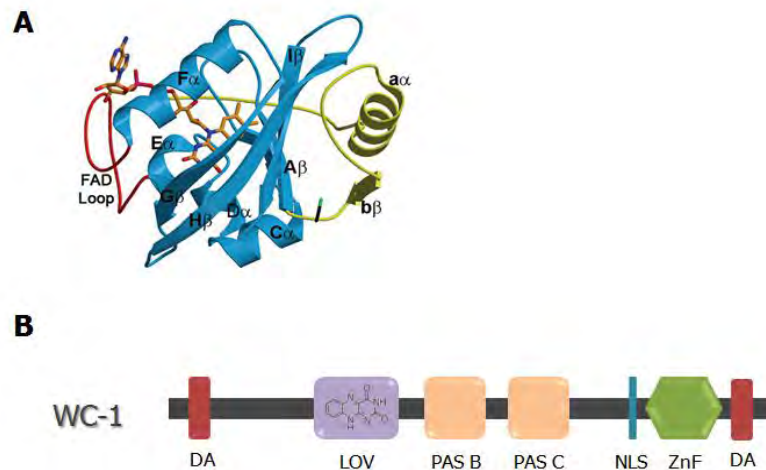
fotorreceptor, el cual inicia la transducción de la señal luminosa en la célula (Crosson *et al.*, 2003; Ogura *et al.*, 2008).



**Figura 3. Fotoquímica de la formación del aducto flavin-cisteinil en los dominios LOV.** La formación del aducto (resaltado en un cuadro rojo) es precedida por un estado triplete de la flavina (T), el cual presenta una estructura electrónica que promueve la protonación del átomo N5 de la flavina y el subsecuente ataque nucleofílico del azufre de la cisteína al átomo C(4a) de la flavina. La energía de un fotón proveniente de la luz azul se denota como  $h\nu$ , la velocidad del decaimiento radiativo del estado singulete fotoexcitado es  $k_{rad}$ , la conversión interna al estado basal es  $k_{IC}$ , y el cruce del intersistema singulete-triplete es  $k_{ISC}$ . El asterisco denota el estado excitado de singulete (Tomado de Crosson *et al.*, 2003).

Los dominios LOV comparten un plegamiento terciario común que crea un hueco (“pocket”) para la unión de un cromóforo de flavina, ya sea FMN (flavin mononucleótido) o FAD (flavín adenín dinucleótido) (Figura 4A). En los hongos, las respuestas mediadas por luz azul dependen principalmente de fotorreceptores tipo LOV. En el ascomiceto *N. crassa* se han descrito los fotorreceptores White Collar (WC)-1 y VIVID (VVD) (Figura 4B). WC-1 es una proteína de 125 kDa que se activa por la luz azul y regula la expresión de genes responsivos a luz y del reloj circadiano (Ballario y Macino, 1997; Crosthwaite *et al.*, 1997; He *et al.*, 2002). VIVID es un fotorreceptor secundario de 21 kDa que consta de un dominio LOV con alta similitud al de WC-1, más una región de “tapa” en el N-terminal (Ncap) y

una inserción tipo bucle (loop) que acomoda el cofactor FAD (Heintzen *et al.*, 2001; Zoltowski *et al.*, 2007). La formación del aducto induce un cambio conformacional en el Ncap de VIVID (Zoltowski *et al.*, 2007). Actualmente, se han identificado proteínas homólogas a WC-1 y VVD en otros hongos. El papel de ambos fotorreceptores en la transducción de la señal luminosa se discutirá más adelante.



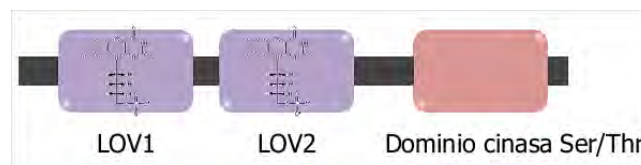
**Figura 4. Estructura de los fotorreceptores tipo PAS/LOV.** A) Estructura tridimensional del dominio LOV de VIVID. El plegamiento típico de los dominios PAS se muestra en azul, la extensión Ncap en amarillo y el loop de unión al cromóforo en rojo (tomado de Zoltowski *et al.*, 2009). B) Estructura general de WC-1. WC-1 tiene tres dominios PAS (morado y salmón), uno de los cuales es un dominio LOV (morado) que une FAD, un dedo de cinc tipo GATA (ZnF), dos dominios de activación transcripcional (DA) y una señal de localización nuclear (NLS) (Modificado de Linden, 2002).

### **Fototropinas**

Las fototropinas son otro tipo de fotorreceptores con dominios LOV con una arquitectura distinta a los receptores de luz azul antes descritos. Aunque el mecanismo de percepción de la señal luminosa a través del dominio LOV es semejante a los de WC-1 y VIVID, el mecanismo de transducción de la señal es distinto, y se describe a continuación. Las fototropinas perciben luz UV-A, UV-B,

azul y verde ( $\lambda$  320-550 nm). Estos fotorreceptores se componen de dos dominios distintos (Figura 5): un dominio N-terminal fotosensor y un dominio C-terminal de proteína cinasa de Ser/Thr (Fankhauser y Staiger, 2002; Chen *et al.*, 2004; Falciatore y Bowler, 2005).

El dominio fotosensor consta de dos dominios LOV asociados a una molécula de FMN. La activación de las fototropinas por la luz azul involucra un fotociclo en el dominio LOV: la luz azul lleva a cabo la formación del aducto entre el FMN y una cisteína conservada en el dominio LOV, dando la activación subsecuente del dominio de cinasa (Falciatore y Bowler, 2005).



**Figura 5. Estructura de las fototropinas.** Las fototropinas comprenden dos dominios LOV (LOV1 y LOV2) y un dominio de cinasa de Ser/Thr. Las fototropinas se autofosforilan en varios residuos de serina y treonina; sin embargo, no se han identificado otros blancos de fosforilación (Modificado de Falciatore y Bowler, 2005).

La actividad de cinasa es necesaria para la transducción de la señal mediada por fototropinas, la cual involucra la autofosforilación en residuos de serina (Salomon *et al.*, 2003). A la fecha, solo se han identificado otros dos blancos de fosforilación de las fototropinas, ATP-BINDING CASSETTE B19 (ABCB19) y PHYTOCHROME KINASE SUBSTRATE 4 (PKS4), y se ha propuesto que se inactivan por fosforilación (Goyal *et al.*, 2013). El proceso de transducción de la señal luminosa mediada por fototropinas también involucra actividad asociada a la membrana, tal como la captación de  $\text{Ca}^{2+}$  y la depolarización de la membrana inducida por luz azul (Falciatore y Bowler, 2005).



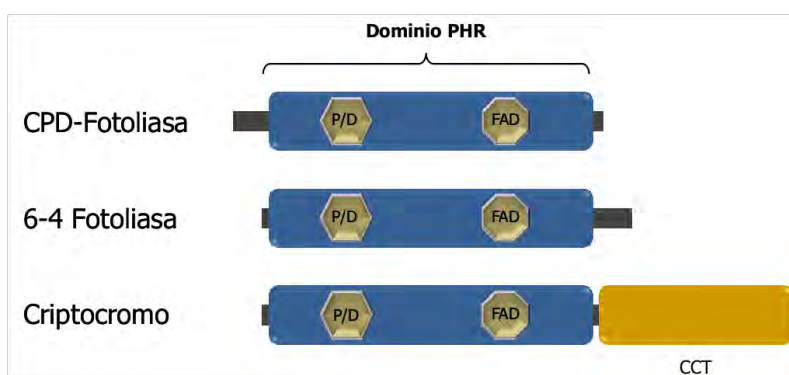
Cuando las semillas germinan en el suelo, la orientación de crecimiento se determina por la gravedad, para asegurar que las plántulas emerjan del suelo. Una vez que la región apical de la plántula (coleoptilo) llega a la superficie, ésta utiliza la luz azul para orientar su crecimiento (Goyal *et al.*, 2013). Las fototropinas se encuentran en la región apical del coleoptilo y participan en los procesos de fototropismo, relocalización de los cloroplastos y apertura de estomas (Fankhauser y Staiger, 2002; Chen *et al.*, 2004; Goyal *et al.*, 2013).

### ***Criptocromo/Fotoliasa***

Todos los miembros de esta familia poseen en su dominio N-terminal una región homóloga de fotoliasas (PHR, por sus siglas en inglés), la cual une un cromóforo catalítico, FAD, y otro cosechador de la luz, que puede ser una pterina (MTHF, meteniltetrahidrofolato) o una deazaflavina (8-HDF, 7,8-didemetil-8-hidroxy-5-deazariboflavina) (Sancar, 2003; Falciatore y Bowler, 2005; Möglich *et al.*, 2010) (Figura 5).

Las fotoliasas catalizan la reparación de DNA dañado por exposición a luz UV-B utilizando luz azul/UV-A ( $\lambda$  350-450 nm) como fuente de energía o como un co-substrato. Hay tres tipos de fotoliasas que reparan específicamente dos lesiones distintas de DNA: los dímeros de ciclobutano de pirimidina (CPD, por sus siglas en inglés) son reparados por las fotoliasas CPD, las cuales están presentes en organismos de todos los reinos, incluyendo algunos virus (Kanai *et al.*, 1997; van Oers *et al.*, 2004). Las 6,4- pirimidina-pirimidona, o fotoproductos (6-4), son reparados por las (6-4) fotoliasas y solo están presentes en eucariontes (Sancar, 2003; Falciatore y Bowler, 2005). La tercera clase se describió recientemente en eubacterias, plantas y animales, con especificidad para CPD en DNA de cadena sencilla (Lucas-Lledó y Lynch, 2009). Sin embargo, cada tipo de fotoliasa repara un fotoproducto en particular y no el otro (Sancar, 2003).

Los criptocromos perciben radiaciones de luz azul y UV-A ( $\lambda$  320-520 nm) y se han encontrado en bacterias, hongos, plantas y animales. Una característica estructural importante de los criptocromos es que la mayoría de ellos, especialmente los de origen vegetal, tienen una extensión C-terminal de 50-250 aa sin homología con las fotoliasas (Sancar, 2003) (Figura 6). Los criptocromos regulan el crecimiento y el ritmo circadiano en plantas, mientras que en animales actúan como fotorreceptores circadianos o represores transcripcionales que controlan el reloj circadiano (Cashmore, 2003; Sancar, 2004).



**Figura 6. Estructura de la familia criptocromo/fotoliasa.** Los miembros de esta familia tienen un dominio PHR que une dos cromóforos, un FAD y una pterina o deazaflavina (P/T). Además del dominio PHR, los criptocromos presentan una extensión en el C-terminal (CCT) poco conservada (Modificado de Schmolle *et al.*, 2010).

Actualmente, la diferencia funcional entre las fotoliasas y los criptocromos se ha cuestionado debido a que recientemente se han encontrado fotoliasas con actividad regulatoria y criptocromos con actividad de fotoliasa. En *Trichoderma atroviride*, la CPD fotoliasa PHR-1 puede regular su propia expresión de manera dependiente de la luz, así como la de otros genes responsivos a luz azul (Berrocal-Tito *et al.*, 2007). En el hongo patógeno *Cercospora zea-maydis*, PHL1 además de su actividad de fotoliasa, afecta la producción de conidias y el metabolismo secundario (Bluhm y Dunkle, 2008). En *Aspergillus nidulans*, CryA presenta una función dual regulando la expresión génica en el hongo, así como una actividad de foto-reparación de DNA en el sistema heterólogo de *E. coli*

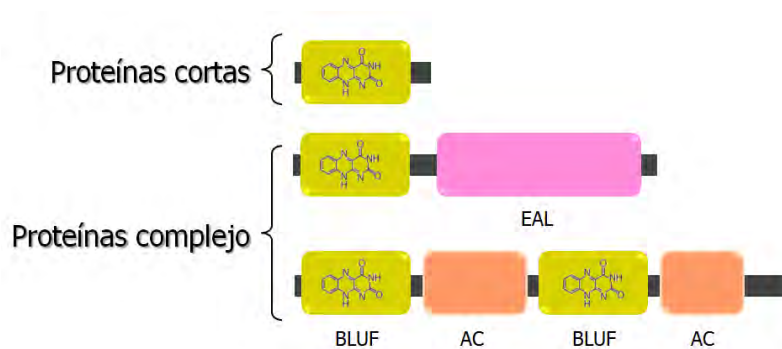
(Bayram *et al.*, 2008). Por otro lado, las fotoliasas PtCPF1 de *Phaeodactylum tricornutum*, así como PHR2 del baculovirus *Chrysodeixis chalcites*, actúan como represores transcripcionales del reloj circadiano en células de mamífero (Coesel *et al.*, 2009; Biernat *et al.*, 2012).

## **BLUF**

Los fotorreceptores que presentan el dominio BLUF están involucrados en la percepción de luz azul (y posiblemente el estado redox) utilizando FAD como cromóforo (de ahí su nombre, por "Blue-Light Using FAD"). Las predicciones basadas en estructura secundaria revelan que el dominio BLUF es un nuevo plegamiento de unión a FAD presente en varias proteínas, principalmente de origen bacteriano. Filogenéticamente, este dominio se encuentra representado en proteobacterias, cianobacterias, protistas y algunos hongos (Gomelsky y Klug, 2002; van der Horst y Hellingwerf, 2004; Herrera-Estrella y Horwitz, 2007).

Las proteínas con dominios BLUF pueden dividirse en dos categorías de acuerdo con la deducción de las estructuras de estos dominios (Figura 7): (i) "complejos", proteínas pertenecientes a multidominios, y (ii) "proteínas cortas" compuestas de un dominio BLUF, más 30-70 residuos adicionales (Gomelsky y Klug, 2002).

Las proteínas con dominios BLUF posiblemente regulan respuestas río abajo mediante interacciones proteína-proteína, ya que carecen de dominios de unión a DNA (Montgomery, 2007). Algunos de estos fotorreceptores presentan dominios de adenilato ciclasa o EAL y GGDEF (abreviaciones que corresponden a los aminoácidos de los motivos conservados en estos dominios). Los dominios GGDEF y EAL participan en la síntesis e hidrólisis de bis-(3',5')-diguanylate cíclico, c-di-GMP, o mononucleótidos cíclicos (Gomelsky y Klug, 2002).



**Figura 7. Estructura de los fotorreceptores BLUF.** Los dominios BLUF se presentan como dominios aislados o unidos covalentemente a dominios efectores, en su mayoría involucrados en el metabolismo de nucleótidos cíclicos. EAL, dominio de fosfodiesterasa; AC, dominio de adenilato ciclasa (Modificado de Möglich *et al.*, 2010).

Se ha descrito la participación de proteínas de esta familia en las respuestas fotofóbicas de *Euglena gracilis* (Iseki *et al.*, 2002), la fototaxis positiva de *Synechocystis* (Okajima *et al.*, 2005) y en la regulación transcripcional de *Rhodobacter sphaeroides* (Masuda y Bauer, 2002), entre otros.

### ***Regulación transcripcional mediada por luz***

Debido a que los organismos que habitan en la superficie de la Tierra están expuestos diariamente a los ciclos de la luz y la temperatura, han evolucionado mecanismos para sincronizar su biología interna con su entorno.

En las plantas, específicamente en *Arabidopsis*, los factores de transcripción tipo Myb CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) y LATE ELONGATED HYPOCOTYL (LHY) actúan en la madrugada para reprimir la expresión del regulador TIMING OF CAB EXPRESSION 1 (TOC1), mediante la unión a un motivo en el promotor de TOC1 llamado elemento nocturno (EE). Los niveles de TOC1 aumentan hacia el final del día y se cree que aumentan, directa o indirectamente, la expresión de CCA1 y LHY (Harmer y Kay, 2005).

Se han reportado cambios en el perfil transcripcional en respuesta a la luz en un grupo numeroso de genes. La abundancia de la mayoría de esos transcritos

incrementa, pero también se han reportado mRNAs cuya abundancia disminuye, como los genes *PHYA*, *PCR*, *GBF3* o *NPR* de *Lemna gibba* (Terzaghi y Cashmore, 1995; Filichkin *et al.*, 2011; Rosales-Saavedra *et al.*, 2006). Para identificar las secuencias que regulan la transcripción en respuesta a la luz, se han realizado un gran número de estudios que combinan microarreglos, análisis *in silico*, deleciones y mutagénesis de promotores de dichos genes fusionados a genes reporteros. Los principales elementos en *cis* encontrados repetidamente en genes regulados por luz en plantas se describen a continuación brevemente (Terzaghi y Cashmore, 1995; Harmer y Kay, 2005):

- Sitios *GT-1*. El núcleo de la secuencia es GGTTAA. Un tetrámero sintético de estas cajas confiere la capacidad de responder a luz al promotor CaMV 35S -90:GUS. Los sitios GT-1 se encuentran normalmente en tándem y su espaciamiento es crítico.
- Cajas *I*. La secuencia que las define es GATAA. Son elementos regulatorios funcionalmente importantes tanto en monocotiledóneas como dicotiledóneas. Muchos genes RBCS (rubisco carboxylase-oxygenase small subunit) presentan una caja I cerca de una caja G, a 100-300 pb río arriba de la caja TATA; mientras que otros genes tienen dos o tres elementos GATA arreglados en tándem y cercanos a la caja TATA.
- Cajas *G*. La secuencia de este elemento es CACGTG. Se encuentran en los promotores de genes que responden a estímulos distintos. Las mutaciones en esta secuencia llevan a una alteración en la expresión del gen que regula.
- *Evening Element* (EE). Su secuencia es de nueve nucleótidos, AAAATATCT. Es importante para la actividad rítmica de varios promotores de fase-nocturna, la cual se suprime cuando se muta este elemento.
- *CBS*. Sitio de unión del factor de transcripción CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) (CBS, por sus siglas en inglés), muy relacionado con el EE. Su secuencia es AAAAATCT. La caja CBS es suficiente para

conferir la expresión rítmica en la fase de madrugada, mediada por la acción positiva de los factores de transcripción CCA1 y LATE ELONGATED HYPOCOTYL (LHY).

- Otras secuencias. Motivos TGACG, cajas *H* (CCTACC), secuencias ricas en AT, elemento CCAAT, caja Gap (ATGAA(A/G)A), elementos ricos en GC, se han identificado en el promotor de genes regulados por luz. Confieren regulación por luz a promotores mínimos, aunque algunos no son esenciales.

La combinación de un número de elementos generales define un promotor inducible por luz, ya que elementos similares pueden presentar efectos diferentes dependiendo del contexto en que se encuentren. En las plantas, las respuestas a la luz están reguladas por varios fotorreceptores que se encuentran interconectados en circuitos positivos y negativos (Terzaghi y Cashmore, 1995; Filichkin *et al.*, 2011).

El conocimiento sobre los mecanismos moleculares de los ritmos circadianos han proporcionado indicios de que las modificaciones post-traduccionales trabajan de la mano con la regulación transcripcional para ajustar finamente nuestros días y noches. El núcleo del reloj es un circuito de retroalimentación negativa compuesto de factores de transcripción que conducen a la expresión de sus propios reguladores negativos (Gallego y Virshup, 2007).

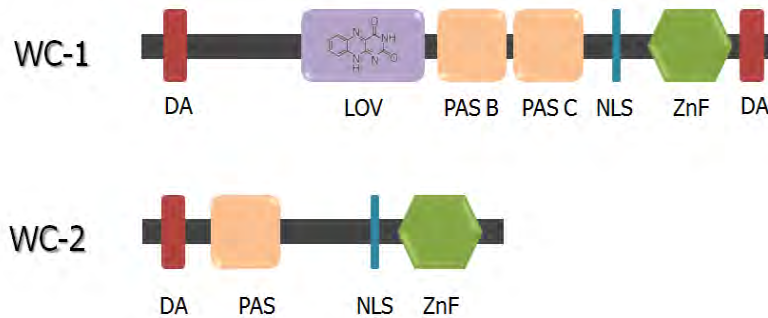
En los animales, los factores de transcripción que regulan positivamente el reloj circadiano son las proteínas CLOCK (Circadian Locomotor Output Cycles Kaput; CLK en *Drosophila melanogaster*) y BMAL1 (Brain and Muscle ARNT-Like 1). Estas proteínas se dimerizan para formar el complejo CLOCK/BMAL1, el cual induce la expresión de un gran número de genes, entre los que se encuentran los reguladores negativos o represores del complejo CLOCK/BMAL1, las proteínas period (PER1 y PER2) y criptocromos (CRY1 y CRY2) (Gallego y Virshup, 2007).

La mayoría de los datos experimentales indican que la secuencia hexamérica CACGTG, llamada caja E, se encuentran comúnmente en el promotor de los genes reconocidos por el complejo CLOCK/BMAL1. Las cajas-E, en general, realizan una función específica y compleja en conjunto con los módulos de información adicional (ej. secuencias tejido específicas) a lo largo del promotor de genes circadianos como *timeless*, *vriille*, *D-element binding protein (dbp)*, *vasopressin (AVP)*, *serotonin N-acetyltransferase (AA-NAT)*, *prokineticin 2 (PK2)* y los *period* (Muñoz y Baler 2003).

En mamíferos, *D. melanogaster* y *Neurospora crassa*, los elementos positivos que regulan la expresión génica en respuesta a la luz son complejos heterodiméricos de proteínas, compuestos de factores de transcripción que presentan dominios PAS. En este sentido, los hongos filamentosos son excelentes modelos para estudiar la percepción y transducción de la señal luminosa, la cual es equivalente a la presente en eucariotas “superiores” como *D. melanogaster* y mamíferos (Gallego y Virshup, 2007).

En el hongo filamentosos *N. crassa*, todas las respuestas a luz conocidas son mediadas por luz azul y casi todas requieren de las proteínas WC-1 y WC-2 (Ballario y Macino 1997; Crosthwaite *et al.* 1997; Linden *et al.* 1997). Aunque la mayoría de las respuestas a la luz azul se pierden en un fondo mutante  $\Delta wc-1$ , en un análisis de microarreglos se identificó una pequeña proporción de genes regulados por luz azul que no son dependientes de *wc-1* (Chen *et al.*, 2009). La proteína WC-1 es el receptor de luz azul y forma un heterodímero con la proteína WC-2 (Figura 8), cuya estructura es similar a WC-1 pero no presenta el dominio LOV (Linden y Macino, 1997; Cheng *et al.*, 2002). Tras un estímulo luminoso, las proteínas WC-1 y WC-2 se heterodimerizan vía sus dominios PAS para formar el Complejo White Collar (WCC) *in vivo* (Ballario y Macino 1997; Crosthwaite *et al.*,

1997; Cheng *et al.*, 2002, 2003), el cual media las respuestas a luz azul, activando la transcripción de genes inducibles por luz (Linden *et al.*, 1997; Liu *et al.*, 2003).



**Figura 8. Estructura de WC-1 y WC-2.** Las proteínas WC son factores de transcripción tipo GATA (ZnF), que presentan un dominio de activación (DA), una señal de localización nuclear (NLS) y dominios de interacción proteína-proteína PAS (color salmón). WC-1 presenta un dominio LOV (color morado) que le da el carácter de fotorreceptor (Modificado de Linden, 2002).

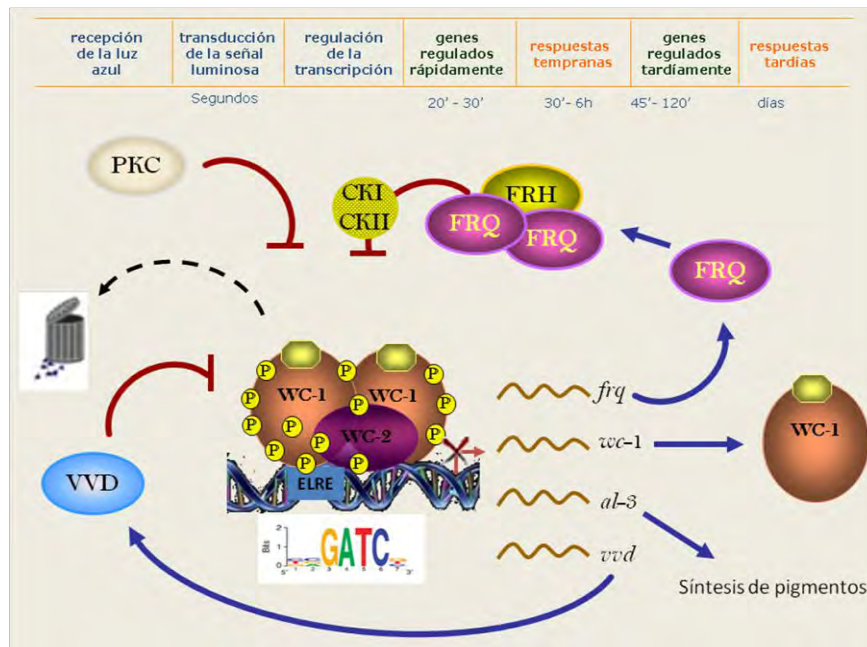
El WCC controla directamente la expresión de un gran número de genes, que a juzgar por la cinética de inducción de la transcripción debida a luz azul, los genes responsivos a luz pueden agruparse en genes tempranos (ELRG, por sus siglas en inglés) y genes tardíos (LLRG, por sus siglas en inglés) (Linden *et al.*, 1997). La inducción de la transcripción de los ELRG se observa desde los 5 min después del pulso de luz, alcanzando su máxima expresión después de 15-30 min. Algunos ejemplos de este grupo de genes incluyen a *wc-1*, frequency (*frq*), *vvd*, los genes *bli*, (inducidos por luz azul), *al* (genes involucrados en síntesis de pigmentos), algunos *ccg*'s (genes controlados por el reloj circadiano), y *con-6* y *con-10* (genes de conidiación -6 y -10). Los LLRG presentan una expresión máxima entre los 45 y 120 min después del pulso de luz. Entre éstos se encuentran *ccg-1*, -2 y -9 (Linden *et al.*, 1997) (Figura 9).

En los estudios iniciales con los promotores de *frq*, *vvd* y *al-3* se comprobó que el WCC se une a los elementos de respuesta a luz (LRE, Light Response Element)



tipo GATA, cuya secuencia consenso es 5'-GATNC---CGATN-3', donde N puede ser cualquier nucleótido, pero es el mismo en ambas cajas, y el espacio entre las cajas GATA es variable (Froehlich *et al.*, 2002; He y Liu 2005). Posteriormente, gracias a estudios de microarreglos se describió una red jerárquica de inducción de genes en respuesta a la luz y se describieron a los LRE tempranos (ELRE; consenso GATC) y a los tardíos (LLRE; consenso TGAYRTCA) (Chen *et al.*, 2009). Entre los ELRG se encuentra el factor transcripcional SUB-1, el cual está involucrado en la regulación de algunos LLRG, presumiblemente a través de la unión a los LLRE (Chen *et al.*, 2009). Recientemente, el consenso del ELRE se extendió a GATCGA, derivado de un análisis de CHIP-seq de todo el genoma de *N. crassa* con un anticuerpo anti-WC-2 (Smith *et al.*, 2010).

La proteína FRQ (producto del gen *frq*) se encuentra en complejo con la RNA helicasa FRH, formando el complejo FRQ/FRH (FFC) (Cheng *et al.*, 2005). El FFC promueve la fosforilación de su activador transcripcional, el WCC, y consecuentemente inhibe su actividad de manera dependiente de la luz, generando un ritmo diario sobre la actividad del WCC que es esencial para mantener el ritmo circadiano en *N. crassa* (Schafmeier *et al.*, 2005). Por otra parte, la proteína cinasa C (PKC) también regula la actividad del WCC, fosforilando principalmente la región C-terminal de WC-1 que contiene el dominio de unión a DNA (Franchi *et al.*, 2005). Finalmente, VVD modula la fotoadaptación del WCC al interactuar con el WCC para regular la activación periódica de la respuesta a la luz azul, permitiendo que el hongo pueda reaccionar a los cambios en las intensidades de luz (Chen *et al.*, 2010a; Malzahn *et al.*, 2010). La fotoadaptación es un mecanismo que utilizan los organismos para ajustar la sensibilidad de sus sistemas de fotopercepción a una exposición prolongada a la luz (Figura 9).



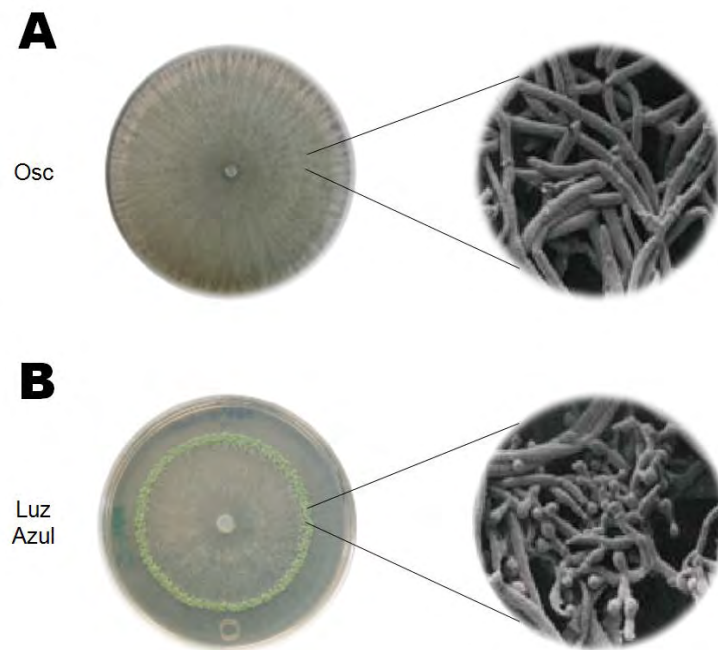
**Figura 9. Vía de señalización de la luz azul en *Neurospora crassa*.**

La luz azul estimula la expresión de genes tempranos y tardíos, la cual es dependiente del WCC y su unión a los ELRE (GATC). En respuesta a la luz, el WCC activa la transcripción de los genes tempranos, como *frq* y *vvd*. FRQ forma un complejo con FRH e inhibe la actividad del WCC a través de interacciones físicas, o reclutando a las cinasas CK I y II que lo fosforilan. Por otra parte, VVD puede heterodimerizar con el WCC, desestabilizándolo. PKC también fosforila el WCC de manera dependiente de la luz. El estado de hiperfosforilación del WCC resulta en la degradación de parte del complejo vía el proteasoma, y una pérdida en la afinidad por el DNA (Tomado y modificado de Linden *et al.*, 1997 y Liu y Bell-Pedersen, 2006).

### **Respuesta a la luz azul en *Trichoderma spp***

*Trichoderma atroviride* es un hongo común del suelo que se emplea en control biológico debido a su capacidad de parasitar a hongos fitopatógenos (Papavizas, 1985). Su modo de propagación en el campo está basado en el uso de conidias, los productos de la reproducción asexual. En muchas especies del género *Trichoderma*, un breve pulso de luz dispara la conidiación (Galun y Gressel, 1966), y se ha mostrado que este proceso está estrechamente regulado por la luz, la limitación de nutrientes y el daño mecánico (Horwitz *et al.*, 1985; Casas-Flores

*et al.*, 2004). *T. atroviride* es empleado como modelo fotomorfogénico ya que en total oscuridad crece indefinidamente como micelio, siempre y cuando tenga todos los nutrientes necesarios en el medio de cultivo. Sin embargo, tras un pulso de luz azul/UVA se induce el desarrollo de conidias en la periferia de una colonia en crecimiento al tiempo de la fotoinducción (Gressel y Galun, 1967) (Figura 10).



**Figura 10. Fotoconidiación en *Trichoderma atroviride*.** Las fotografías muestran el efecto de un pulso de luz azul (5 min) en la morfología de *T. atroviride*. A) Crecimiento de una colonia en oscuridad. B) Morfología de una colonia después de 36 h de recibir el pulso de luz azul, la cual exhibe el anillo de conidias verdes característico del hongo. A la derecha se muestran las micrografías electrónicas de la región indicada (Tomado y modificado de Schmoll *et al.*, 2010 y Carreras *et al.*, 2012).

El espectro de acción en el que ocurre la producción de conidias incluye un pico agudo en la región UV de 350-380 nm y un pico más amplio en la región azul con un máximo a 440-450 nm (Gressel y Hartman, 1968; Kumagai y Oda, 1969), lo cual sugirió la participación de una flavina en la transducción de la señal luminosa. Esta hipótesis fue apoyada posteriormente, al mostrar experimentalmente que la

roseoflavina (un análogo de la riboflavina) interfiere con la fotoconidiación en *Trichoderma* (Horwitz *et al.*, 1984a).

Otros eventos que se han reportado en *Trichoderma* después de la exposición a la luz azul incluyen la síntesis de RNA (Galun y Gressel, 1966; Gressel y Galun, 1967), cambios en la composición de lípidos (Betina y Koman, 1980), cambios en el potencial de membrana (Horwitz *et al.*, 1984b), un incremento en el nivel intracelular de ATP y cAMP (Támová *et al.*, 1995; Grešik *et al.*, 1988), así como un aumento en la fosforilación de las proteínas (Grešik *et al.*, 1989).

Una de las estrategias para identificar el sistema de percepción de luz azul en *Trichoderma* se basó en la relación evolutiva de las fotoliasas y los criptocromos, dado el espectro de acción de la fotoconidiación. Así, se identificó y caracterizó el primer gen responsivo a luz azul en *T. atroviride*, el cual se nombró *phr-1* ya que codifica para una fotoliasa tipo CPD (Berrocal-Tito *et al.*, 1999). La cinética de expresión de *phr-1* mostró que en la oscuridad el mRNA no se detecta, pero luego de un pulso de luz azul se observó una acumulación rápida y transitoria del transcrito, alcanzando un pico máximo entre los 15 y 30 min, y un descenso después de 60 min (Berrocal-Tito *et al.*, 1999). Aunque PHR-1 no fue el fotorreceptor de luz azul de *T. atroviride*, sí presentó actividad de fotoliasa, así como un papel en la regulación su propio transcrito (Berrocal-Tito *et al.*, 2007).

Posteriormente, en *T. atroviride* se identificaron los genes homólogos a *wc-1* y *wc-2*, los cuales se denominaron blue light regulator (*blr-1* y *blr-2*). BLR-1 y BLR-2 presentan dominios similares a los de sus ortólogos en *N. crassa*, a excepción de los dominios de activación presentes en las proteínas WC, lo cual sugiere la participación de una o más proteínas en la activación de la transcripción (Casas-Flores *et al.*, 2004). Tanto *blr-1* como *blr-2* son esenciales para la fotoconidiación y la regulación de genes responsivos a luz azul en *T. atroviride* y *T. reesei*, por lo que se ha propuesto que BLR-1 actúa como fotorreceptor y trabaja en complejo

con BLR-2 (Casas-Flores *et al.*, 2004; Rosales-Saavedra *et al.*, 2006; Castellanos *et al.*, 2010).

Entre los análisis que se han realizado para elucidar la vía de transducción de la señal luminosa, se emplearon una serie de inhibidores que interfieren con la producción de cAMP (Sulová y Farkaš, 1991; Berrocal-Tito *et al.*, 2000). La inhibición de la fosfodiesterasa (PDE) estimuló la conidiación inducida por luz, sin afectar el crecimiento (Sulová y Farkaš, 1991), mientras que la inhibición de la adenilato ciclasa bloqueó la conidiación (Berrocal-Tito *et al.*, 2000). Por otra parte, la adición de un análogo de cAMP provocó que esporulara una colonia que crecía en la oscuridad (Berrocal-Tito *et al.*, 2000), más no se afectó el perfil transcripcional de *phr-1*, lo que sugiere la existencia de dos vías independientes que se activan por luz azul. Se ha observado una activación de la PKA después de un pulso de luz azul, aún en las mutantes  $\Delta blr-1$  y  $\Delta blr-2$ , lo cual confirma un sistema alternativo para la percepción de luz azul ligado a cAMP (Casas-Flores *et al.*, 2006). La activación de la PKA también tiene un impacto directo en la expresión de los genes regulados por luz azul dependientes de las proteínas BLR, ya que una baja actividad de la PKA correlaciona con un bloqueo en la transcripción de dichos genes en *T. atroviride* (Casas-Flores *et al.*, 2006).

La vía de la PKA está relacionada con algunas proteínas G, entre las cuales la GNA3 parece tener un papel en la percepción de la luz, ya que se ha observado un incremento en la transcripción del gen *gna3* debido a la luz en *T. reesei* (Schmoll *et al.*, 2009). Por otra parte, las transformantes que expresan una versión activa de GNA3 presentan una fuerte expresión, en respuesta a la luz, del gen *cbh1*, el cual codifica para una celobiohidrolasa (Schmoll *et al.*, 2009). La GNA1 es otra proteína G analizada en *T. reesei*, la cual también parece estar involucrada en la vía de señalización por luz azul, ya que en la cepa mutante  $\Delta gna1$ , no se observó una inducción por luz del gen *cbh1* (Seibel *et al.*, 2009). El análisis del gen ortólogo de *gna1* en *T. atroviride* (*tga1*) mostró que tiene un papel

como regulador negativo de la conidiación, Las transformantes donde se silenció *tga1* conidieron constitutivamente, mientras que las sobreexpresantes fueron incapaces de producir conidias en respuesta a la luz (Rocha-Ramírez *et al.*, 2002).

Un segundo fotorreceptor de luz azul, ENVOY (ortólogo a VIVID de *N. crassa*), se identificó y caracterizó en *T. reesei* (Schmoll *et al.*, 2005), y también se encuentra codificado en los genomas de *T. atroviride* y *T. virens*. En oscuridad, el gen *env1* se transcribe a un nivel muy bajo, pero luego de recibir el estímulo luminoso la abundancia del transcrito incrementa hasta más de 500 veces y su expresión depende de las proteínas BLR tanto en *T. atroviride* como en *T. reesei* (Rosales-Saavedra *et al.*, 2006; Schuster *et al.*, 2007). Posteriormente, se mostró la participación de ENVOY en la regulación de genes responsivos a luz azul y en la transducción de la señal luminosa, a través de proteínas G y la vía de PKA (Castellanos *et al.*, 2010; Tish *et al.*, 2011). Aunque las proteínas ENV1 y VVD muestran una regulación similar y alta similitud en su secuencia a nivel de aminoácidos entre ellas, el gen *env1* no fue capaz de complementar una cepa  $\Delta vvd$ , lo cual sugiere diferencias en la regulación por luz en ambos hongos (Schmoll *et al.*, 2005).

Entre otros procesos mediados por el complejo BLR y ENVOY, se ha demostrado que hay un incremento en la transcripción de los genes de celulasas bajo luz constante, en comparación con condiciones en oscuridad, lo cual está regulado por ENVOY (Schmoll *et al.*, 2005). Estos experimentos se llevaron a cabo con celulosa como fuente de carbono; sin embargo, se han observado diferencias en las respuestas a luz/oscuridad con distintas fuentes de carbono o nitrógeno (Friedl *et al.*, 2008a, b). Por tanto, la presencia o ausencia de luz parece ser relevante para el uso y/o toma de una fuente de carbono específica. En *T. atroviride*, las mutantes  $\Delta blr-1$  y  $\Delta blr-2$  son incapaces de producir conidias en respuesta a una privación súbita de la fuente de carbono (Casas-Flores *et al.*, 2004).

Entre otros nutrientes, la asimilación de los compuestos de azufre, como cisteína, metionina y S-adenosilmetionina, es dependiente del estado de luz, como se ha visto para *T. reesei*. Dichos compuestos son cruciales para el crecimiento en celulosa en condiciones de luz y no puede compensarse por la adición de metionina (Gremel *et al.*, 2008).

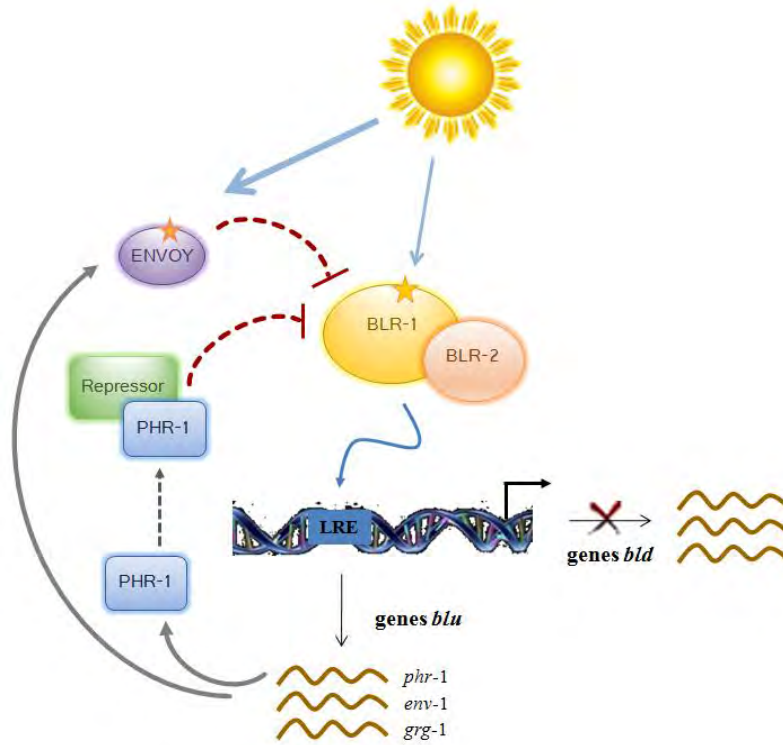
A pesar de las semejanzas entre *Trichoderma* spp. y *N. crassa*, existen diferencias importantes en la respuesta a la luz azul en estos hongos. Se ha demostrado que BLR-2 es el factor limitante en la formación del complejo BLR (BLRC), ya que la sobreexpresión de esta proteína resultó en un incremento en la sensibilidad a la luz azul en *T. atroviride* (Esquivel-Naranjo y Herrera-Estrella, 2007), mientras que en *N. crassa* es WC-1 (Cheng *et al.*, 2001). Por otra parte, las proteínas BLR participan en la activación y represión transcripcional de genes regulados por luz azul, un papel dual que no había sido descrito en otros hongos (Rosales-Saavedra *et al.*, 2006) (Figura 11).

A fin de tener una mayor comprensión sobre la regulación génica en respuesta a la luz en *Trichoderma*, se realizó un primer análisis usando microarreglos de cDNA con 1438 genes representados. Los resultados mostraron que aproximadamente el 3% de los genes fueron responsivos a luz, de los cuales 30 genes se activaron (llamados *blu*, *blue light upregulated*) y 10 se reprimieron (llamados *bld*, *blue light downregulated*) (Rosales-Saavedra *et al.*, 2006). Recientemente, gracias a la disponibilidad de la secuencia de varios genomas, se ha hecho un análisis cuantitativo del transcriptoma con el que se identificaron 331 genes regulados por luz blanca, de los cuales 204 son responsivos específicamente a luz azul (Carreras-Villaseñor *et al.*, 2012). Así mismo, en *T. reesei* se han descrito genes que se expresan o que se reprimen en respuesta a la luz (Schuster *et al.*, 2007).

El análisis *in silico* de la región promotora de los genes regulados por luz azul en *Trichoderma* spp., ha permitido la identificación de secuencias similares a los ELRE descritos para *N. crassa* (Rosales-Saavedra *et al.*, 2006; Schuster *et al.*, 2007), lo que sugiere que el mecanismo de acción del BLRC es similar al reportado para el WCC (Figura 11). Específicamente, en el gen *phr-1* de *T. atroviride* se han descrito dos posibles secuencias reguladoras de la respuesta a la luz, el elemento albino proximal, APE (Albino Proximal Element) (Berrocal-Tito *et al.*, 1999) y un elemento similar al ELRE (Rosales-Saavedra *et al.*, 2006).

Por otra parte, se identificaron dos motivos putativos de unión a proteínas en el promotor de *env1* de *T. reesei*: Envoy Upstream Motifs 1 (EUM1: CTGTGC) y 2 (EUM2: ACCTTGAC) (Schmoll *et al.*, 2005), los cuales se han identificado en otros genes regulados por luz (Schuster *et al.*, 2007). Sin embargo, a la fecha no se ha probado experimentalmente que estos elementos regulatorios putativos sean funcionales *in vivo* en ninguna especie de *Trichoderma*.





**Figura 11. Modelo de la vía de señalización de la luz azul en *Trichoderma*.** Las proteínas BLR-1 (amarillo) y BLR-2 (rosa) forman el complejo fotorreceptor BLRC, el cual percibe la luz del medio y se une a los LRE en el promotor de los genes responsivos a luz azul, reprimiendo (*bld*) o activando (*blu*) su transcripción. ENVOY (morado) participa en la fotoadaptación, reprimiendo la expresión de genes *blu* posiblemente a través de la interacción con el BLRC. Por otra parte, PHR-1 regula su propia expresión y la de otros genes *blu*, posiblemente uniéndose a un represor, que a su vez regula negativamente al BLRC (Tomado y modificado de Berrocal-Tito *et al.*, 2007 y Esquivel-Naranjo y Herrera-Estrella, 2007).

## Resultados parte I

### **The *Trichoderma atroviride* photolyase-encoding gene is transcriptionally regulated by non-canonical light response elements**

Mayte Guadalupe Cervantes-Badillo, Tania Muñoz-Centeno, Edith Elena Uresti-Rivera, Gerardo Rafael Argüello-Astorga, Sergio Casas-Flores\*

IPICYT, División de Biología Molecular. Camino a la presa San José No. 2055, Colonia Lomas 4a sección. C.P. 78216. San Luis Potosí, Mexico.

\*Correspondence

S. Casas-Flores

IPICYT, División de Biología Molecular. Camino a la presa San José No. 2055, Colonia Lomas 4a sección. C.P. 78216. San Luis Potosí, Mexico.

Phone: +52 (444) 8342046; Fax: +52 (444) 8342010;

E-mail: [scasas@ipicyt.edu.mx](mailto:scasas@ipicyt.edu.mx)

Running title: non-canonical LREs drive *phr-1* photoinduction

#### **Abstract**

The BLR-1 and BLR-2 proteins of *Trichoderma atroviride* are the *Neurospora crassa* homologs of White Collar-1 and -2, two transcription factors involved in the regulation of genes by blue-light. BLR-1 and BLR-2 are essential for photoinduction of *phr-1*, a photolyase-encoding gene whose promoter exhibits sequences similar to well-characterized light regulatory elements of *Neurospora*, including the Albino Proximal Element (APE) and the Light Response Element (LRE). However, despite the fact that this gene has been extensively used as a blue light induction marker in *Trichoderma*, the function of these putative regulatory elements has not been proven. The described LRE core in *N. crassa* comprises two close but variably spaced GATA-boxes to which a WC-1/-2 complex binds

transiently upon application of a light stimulus. Using 5' serial deletions of the *phr-1* promoter, as well as point mutations of putative LREs, we were able to delimit a ~50 bp long region mediating the transcriptional response to blue-light. The identified light-responsive region contained three CGATC motifs, two of them displaying opposite polarity to canonical WCC binding sites. Chromatin immunoprecipitation experiments showed that the BLR-2 protein binds along the *phr-1* promoter in darkness, whereas the application of a blue light pulse resulted in decreased BLR-2 binding to the promoter. Our results suggest that BLR-2 and probably BLR-1 are located on the *phr-1* promoter in darkness ready to perform their function as transcriptional complex in response to light.

**Keywords:** *Trichoderma*; *phr-1*; blue light; White-Collar-Complex; Light Response Element; BLR proteins

**Abbreviations:** WC, White Collar; WCC, WC Complex; PAS, PER-ARNT-SIM; LOV, Light, Oxygen and Voltage; BLR, Blue Light Regulator; BLRC, BLR complex; LRE, Light Response Element; ELRE, Early LRE; LLRE, Late LRE; APE, Albino Proximal Element; PLRR, Potential Light Response Region; UCR, upstream conserved region; TSS, Transcription Start Site; ChIP, Chromatin Immunoprecipitation.

## Introduction

Light is an important environmental cue, to which organisms respond in many different ways. Filamentous fungi are able to adapt their physiology to light signals to induce morphogenetic pathways [1]. Blue light regulates photoconidiation, phototropism, entrainment and resetting of circadian rhythms, carotenoid synthesis, sexual, and asexual development, among other processes [1-3]. For decades, the filamentous fungus *Neurospora crassa* has been the classical model system to study blue-light signal transduction. In this fungus, the White Collar (WC) -1 and WC-2 proteins mediate all known responses to blue light [4-6]. Both proteins contain PER-ARNT-SIM (PAS) domains for protein-protein interactions, and GATA type Zn-finger DNA-binding domains. WC-1 functions as a blue-light photoreceptor by means of its specialized PAS domain called LOV (light, oxygen, and voltage), which harbors a flavin as chromophore [4,7]. WC-1 and WC-2 form the White Collar Complex (WCC) via their PAS domains [8]. Upon exposure to light, the WCC binds transiently to promoters of light-inducible genes to activate their transcription [9,10]. In early studies, it was established that WCC binds to the Light Response Element (LRE) core sequence 5'-GATNC---CGATN-3', where N can be any nucleotide but the same nucleotide is used in both repeats, and the space between both GATA boxes is variable [9,10]. Afterwards, a genome-wide microarray analysis revealed a hierarchical network of transcriptional light responses in *Neurospora*, and bioinformatic analyses identified early light response elements (ELREs) as well as late light response element (LLRE) in the promoter of genes exhibiting different temporal light-responsiveness. The defined LLRE core was RTGAYRTCA, whereas the ELRE core was GATCB [11], that was subsequently expanded to GATCGA, a consensus derived from 29 light induced genes identified in a whole genome ChIP-seq analysis using a WC-2 antibody [12].

*Trichoderma atroviride* is used as a photomorphogenic model due to its ability to conidiate upon exposure to light [13]. In total darkness, *T. atroviride* grows indefinitely as a mycelium, whereas a short pulse of blue light induces conidiation forming a discrete ring in the colony perimeter where the light pulse was applied [13]. The corresponding *wc*

orthologous genes in *T. atroviride*, *blr-1* and *blr-2*, are essential for photoinduction of blue-light-responsive genes and photoconidiation in this fungus [14,15]. Essentially, BLR proteins show similar structures to the WC proteins from *N. crassa*, with the exception that the BLR proteins lack an evident transcription activation domain [14]. Based on this, it has been proposed that BLR-1 acts as a photoreceptor together with BLR-2 [14, 15]. Overexpression of *blr-1* and -2 caused a diminished and an increased sensitivity to light, respectively, which led to conclude that BLR-2 is the limiting factor in blue light response in *T. atroviride* [16]. In contrast to *Neurospora*, the putative BLR complex (BLRC) in *T. atroviride* is responsible for induction and repression of several genes. Analysis by cDNA microarrays (representing 1438 genes) revealed 30 up-regulated genes and 10 down-regulated genes by blue light [15]. Subsequently, quantitative genome-wide transcriptome analysis allowed the identification of 331 white light-regulated genes and 204 specifically responsive to blue light, as well as *blr*-independent regulated genes, suggesting involvement of another blue light perception system [17]. BLR-1 and -2 also impact on the utilization of different carbon sources, biosynthesis of secondary metabolites, sulphur metabolism, and cellulase production [18].

A second photoreceptor, the ENV1 protein, orthologous to VIVID in *N. crassa*, was also described in *Trichoderma reesei*. The ENV1 and VIVID proteins contain a PAS/LOV domain and negatively regulate light responses in *Trichoderma* and *Neurospora*, respectively [19-22]. Although ENV1 and VIVID showed similar regulation and high sequence similarity at the amino acid level, the *env1* gene was unable to complement a  $\Delta vvd$  strain [21], which suggests differences in light regulation in both fungi.

The photolyase encoding gene, *phr-1*, was the first blue-light induced gene identified in *T. atroviride* [23] and its expression is dependent on both BLR proteins [14]. The *phr-1* transcript is absent in the dark, and it is strongly and rapidly up-regulated upon exposure to blue light, reaching its maximum level between 15 and 30 min, decreasing 60 minutes after the stimulus [23,24]. Putative light regulatory elements have been described for the *phr-1* promoter, including the Albino Proximal Element (APE) [23], and an LRE sequence [15]

similar to those defined for early light responsive genes in *N. crassa* [10,25]. Moreover, two putative protein-binding motifs, envoy upstream motifs 1 (EUM1: CTGTGC) and 2 (EUM2: ACCTTGAC) were identified on the *env1* and other promoters of blue light regulated genes [21,26]. APE and LRE consensus also have been found on the *phr-1* gene promoter in *T. reesei* [15,26].

Although *phr-1* has been widely used as a molecular marker for light responses in *T. atroviride* [14,15,16,24,27,28], to date no experimental analysis exists to demonstrate the role of the predicted *cis*-acting elements in response to blue-light or to prove the interaction of BLR proteins with the *phr-1* promoter. In this work we identified the *cis*-acting elements responsive to blue light in the *phr-1* promoter, and we analyzed the interaction of BLR-2 protein on this promoter. For this, we examined a set of 5' *phr-1* promoter serial deletions, and *E. coli-lacZ* fusions harboring mutations in some putative regulatory elements, under different light conditions. Our results showed that a short promoter region including several GATN boxes suffices to drive the blue-light induction of *phr-1*, and that two repeats of inverted GATN motifs play an important role in such regulation. Besides, we describe the presence of a putative repressor in darkness and demonstrate the binding of BLR-2 along the *phr-1* promoter by chromatin immunoprecipitation.

## Results

### Delimitation of potential light-responsive elements on the *phr-1* promoter

To determine the intergenic region between *phr-1* (ID: 302457) and its neighbor upstream gene, an informatic search on the *T. atroviride* genome (<http://genome.jgi.doe.gov/Triat2/Triat2.home.html>) was performed. The *phr-1* gene is organized on the genome in a divergent transcription configuration, with the *xlf* gene (Fig. 1A) that encodes for a 485 amino acids protein (ID: 295995) orthologous to the human XRCC4-like factor. This factor belongs to the XLF family and interacts with the XRCC4-DNA ligase IV complex to promote DNA non-homologous end-joining during DNA double strand break repair [29]. The intergenic region between *phr-1* and *xlf* is 1326 bp in length (Fig. 1A).

Sequence alignments of *phr-1* promoter from five different *Trichoderma* species (*T. atroviride*, *T. virens*, *T. reesei*, *T. citrinoviride* and *T. longibrachiatum*) allowed us to identify two conserved regions along the promoters. The first one was located from -107 to -217 position (relative to the *T. atroviride phr-1* TSS), and contained sequences with significant similarity to previously described LREs [9]; consequently, it was termed “potential light responsive region” (PLRR). The second conserved block of sequences was located from -767 to -938 relative to the *phr-1* TSS, and was named “upstream conserved region” (UCR) (Fig. 1A and Fig. S1).

To verify that the *phr-1* gene is regulated in a similar way by blue light in-different *Trichoderma* species, cDNA from *T. virens*, *T. reesei*, and *T. atroviride* mycelia exposed to a pulse of blue light was used for RT-PCR analysis of *phr-1* gene. Indeed, *T. reesei* and *T. virens phr-1* gene was blue light induced as in *T. atroviride*, suggesting that the *cis*-regulatory elements on this gene promoter have remained functional throughout evolution in this fungal genus (Figure 1B).

To analyze more closely the function of the putative *cis*-regulatory elements, a set of 5' serial deletions of the *T. atroviride phr-1* promoter were PCR-amplified and translationally

fused to the *E. coli lacZ* reporter gene by cloning the amplicons in the pCB-*lacZ* vector (see Material and Methods for details). *T. atroviride* wild type strain was transformed with these constructs and stable transformants were selected for further analysis. The integrity of endogenous *phr-1* gene (Fig. S2A) and integration of the construction into the genome were verified (Fig. S2B). Blue-light induced gene expression on generated strains was analyzed by quantitative RT-PCR (qRT-PCR) (Fig. 2A and B). Control strain transformed with empty pCB-*lacZ* vector showed no *lacZ* transcript (Fig. S3). qRT-PCR analyses from -1326 to -196 promoter versions allowed us to determine that the blue-light induction profile was similar to that of the *phr-1* endogenous gene (Fig. 2A and B). Slightly higher expression levels of *lacZ* gene were detected in dark conditions at almost all promoter versions, independently from their transgene copy number (Table S1), when compared to the expression level of *phr-1* endogenous gene, which suggests the presence of a repressor (Fig. 2B). Furthermore, these results indicate that the UCR is dispensable for light transcriptional regulation of *phr-1* under tested conditions (Fig. 2A and B).

In subsequent downstream deletions, the reporter gene expression was constitutive, losing the typical light-regulated expression pattern detected for the *phr-1* endogenous gene (Fig. 2B). Therefore, sequences critical for the transcriptional response to blue light reside within the -196 to -157 promoter segment. An additional observation is that the putative APE sequence located in the -87 truncated promoter does not seem to participate in blue-light regulation of *phr-1* gene (Fig. 2B), as previously suggested [23].

### **The GATA-2 motif is not necessary for transcriptional blue light regulation.**

According to the analysis of *phr-1* promoter serial deletions crucial information required for *phr-1* induction by light resides within the -196 to -156 promoter region. This segment includes a GATA motif (“GATA-2” element) similar to those described for *N. crassa* LREs, which is 42 bp away from a conserved CGATC box (i.e., “GATA-1”) located at positions -141 to -137 (i.e., within the non-light responsive -156 truncated promoter). Consequently we decided to examine the impact of mutations in the GATA-2 element on the regulatory qualities of the -196 *phr-1* promoter. As can be appreciated in Figure 3,



mutations in the GATA-2 box had no effect on photoinduction of *lacZ* gene expression, indicating that sequences essential for transcriptional light responsiveness of the -196 *phr-1* promoter are located downstream the mutated element.

Since the 28 bp-long promoter segment (-184 to 157) where *cis*-acting elements relevant for blue light-responsiveness were mapped does not exhibit any DNA motifs related to described fungal LREs, we performed a careful analysis of this regulatory region and its flanking sequences. As a result two 11-bp imperfect tandem repeats were identified at -178 to -157 nucleotides, which in the antisense strand displays a CGAT(t/g)TCAGTG sequence. The 7-bp core of these repeats, CGAT(t/g)TC, is very similar to the conserved GATA-1 element (i.e., CGATC) found 15 bp downstream (Figure S1 and 4). Therefore, a plausible interpretation of these data is that LREs on the *phr-1* promoter of *Trichoderma* are actually composed of two or more CGATN boxes arranged in identical or alternative polarities. To search for indirect evidence supporting this interpretation, we carried out a thorough comparative analysis of the *phr-1* gene upstream non-coding sequences including five *Trichoderma* species (Figure S1) and distant relatives in the fungal class Sordariomycetes (Figure 4). As a result we delimited *phr-1* promoter sequences of *Fusarium oxysporum*, *Glomerella cingulata* and *N. crassa* that are actually homologs (i.e., evolutionarily equivalent) of the *T. atroviride phr-1* promoter region relevant for blue light transcriptional responses (i.e., the -190 to -106 region). Indeed, the PLRRs of these Sordariomycetes display an analogous arrangement, with several imperfect CGATN repeats in similar or opposite polarities (Figure 4). In the case of *N. crassa*, the putative LRE also included imperfect tandem repeats CGATN, whereas in the *Glomerella cingulata* LRE, two CGATN motifs are part of a 10-bp palindromic element (Fig. 4). Experimental data have demonstrated that *F. oxysporum* and *N. crassa phr-1* genes are also blue light regulated and their transcription is dependent of WC proteins [11,30]. Our bioinformatic analysis indicates that PLRRs were conserved through evolution in Sordariomycetes and that their *phr-1* genes could be regulated in the same way as *phr-1* from *Trichoderma* spp.

### **BLR-2 transiently binds to PLRR and UCR on the *phr-1* promoter**

In *N. crassa*, the WCC binds to the ELREs located in the promoter of blue-light regulated genes, in both darkness and after a light pulse [9-12]. Sequence analysis on several promoters of *T. atroviride* light-induced genes showed that some of them have proximal and distal LREs, whereas others have only one. It was also observed that the length of the sequence separating the GATA motifs of an LRE is variable among these functional elements (Fig. 5A). Since the light induction of *phr-1* in *T. atroviride* is dependent on *blr-1* and -2 products [14], we reasoned that BLR proteins could be interacting with the conserved regions of the *phr-1* promoter. To investigate if the BLR-2 protein is able to bind to the *phr-1* promoter and to map the possible binding sites of such protein, ChIP assays on crosslinked chromatin of *T. atroviride* wild-type strain under different light conditions were performed by using an anti-BLR-2 antibody. BLR-2 binding was mapped through the *phr-1* intergenic region using primers as indicated in Fig. 5B. In darkness, a discrete positioning of BLR-2 was observed on the UCR, the PLRR and the neighboring regions (Fig. 5C). Thirty minutes after a blue-light pulse, a decrease in BLR-2 enrichment was observed. After 2 h of light induction, positioning of BLR-2 on *phr-1* promoter was decreased significantly but still detectable (Fig. 5C). In contrast, there was no BLR-2 enrichment in the intermediate region 5 (around -600 bp) at any time.

## Discussion

Species from the *Hypocrea/Trichoderma* genera have been employed as photomorphogenic models due to their ability to produce a ring of green conidia in response to a pulse of blue light given to a growing colony. After a light pulse, *T. atroviride* also shows a transient biphasic oscillation in intracellular cAMP levels, changes in membrane potential, as well as in ATP levels, and in the induction and repression of a set of genes [15, 31]. Since Berrocal-Tito and co-workers reported *phr-1* as a blue-light induced gene in *T. atroviride* [23], it has been used as a molecular marker to study blue-light responses in this fungus. Sequence analysis of promoters of blue-light regulated genes of *T. atroviride* and *T. reesei* that were identified in studies with expression arrays [15], revealed the presence of elements similar to the LRE core reported in *N. crassa* [10]. We showed that those putative regulatory sequences are highly conserved on the *phr-1* promoter of five *Trichoderma* species (Fig. S1). We also determined that sequences similar to described LREs are present in several light-induced genes of *T. atroviride* (Fig. 5A). Comparative genome analysis of *T. atroviride*, *T. virens*, and *T. reesei* indicated that *T. atroviride* resembles the more ancient state of *Trichoderma* and that the other species evolved later. Besides, *T. reesei* has apparently been evolving faster than *T. virens* and *T. atroviride* since the time of divergence [32]. The molecular mechanism that controls the blue-light transcriptional regulation appears to be conserved in these species, as demonstrated by the sequence (Fig. S1) and transcription analysis (Fig. 1B) of the *phr-1* promoter genes and their transcription profiles.

The performed bioinformatic search for the upstream sequence of the *phr-1* gene of *T. atroviride* revealed that it is organized on the genome in a divergent transcription configuration with the *xf* gene that encodes for an orthologous to the human XRCC4-like factor. This suggests at first glance, a co-regulation of both genes by the fact that *xf* shares *cis*-acting elements with the *phr-1* gene and because of the presumed function of XRCC4-like factor on repairing DNA double strand break. However, the *xf* gene is not induced by white or blue light (Herrera-Estrella A, personal communication).

It is well established that the *phr-1* transcript is absent in darkness, and that after a blue-light stimulus it undergoes up-regulation [23]. However, in qRT-PCR analyses of the reporter gene expression we observed a low but significant transcription of *lacZ* in darkness with almost all the constructs tested (excluding the -600 bp construct), a fact that was in clear contrast with the transcription pattern of endogenous *phr-1*. To explain those odd results two alternative hypotheses can be proposed: 1) the native *phr-1* promoter is actually active in the dark at very low level, though the *phr-1* transcript is rapidly degraded in that condition; 2) presence of a repressor that down-regulates *phr-1* transcription in the dark, but only in its native chromosomal context. If the latter suggestion is correct, then the low level *lacZ* expression in darkness could be attributed to an effect of chromatin organization on the integration site and/or to a chromosomal position effect. Although transformants were not examined in Southern blots, qPCR showed that different constructs could be integrated randomly in the *Trichoderma* genome. In addition, no direct relationship was revealed between *lacZ* expression in the darkness and the copy number of *lacZ* determined by qPCR. These statements led us to suggest that integration of almost all constructs was out of the *phr-1* chromatin context, leading to loss of their repression in the dark.

The promoter of the *T. atroviride phr-1* gene exhibits several sequences similar to described light regulatory elements such as the LRE and APE of *N. crassa* [15, 23]. In this fungus was demonstrated that an LRE conformed by two GATA-boxes is required for induction of blue-light responsive genes and binding of WCC to these sequences in response to light signals was also observed [9,10]. However, in such studies the two GATA boxes comprising the LRE were not studied separately and no point mutations in those elements were generated and examined to prove their function *in vivo*. Derived from expression arrays, a tandemly repeated GATC motif related to the core sequence of the LRE [11] was identified in the promoter of the *frequency (frq)* gene, which is bound by WCC and has been shown to be necessary and sufficient to drive light induction of that gene *in vivo* [9]. This GATC consensus was extended to GATCGA by ChIP-seq analysis [12]. In the present work we demonstrated that a promoter region of *T. atroviride* harboring sequences similar to *Neurospora* LREs but displaying a different arrangement, contains the

necessary information for *phr-1* blue-light induction. The GATA-1 box of the presumed *phr-1* light responsive region and those-GATA-1-like boxes found on the promoters of blue light regulated genes in *T. atroviride* (Fig. 5A) match the two described consensus, taking into account that both of them display variability in the first and last bases [11,12]. On the other hand, our mutagenesis analysis of the GATA-2 motif clearly showed that this element was dispensable for transcriptional light-induction of the *lacZ* gene reporter in the context of the -196 truncated version of the *phr-1* promoter (Fig. 3).

Comparative analysis of the upstream sequences of photolyase-encoding genes from four different genera of Sordariomycetes (*Fusarium*, *Glomerella*, *Neurospora* and *Trichoderma*) revealed the conservation of a ~90 bp promoter region harboring four to six imperfect repeats with a CGATN core sequence. The arrangement of these repeats is variable among the examined fungal species, though the CGATN boxes with the same polarity than the typical LREs from *Neurospora* promoters predominate. In this sense, the light responsive *phr-1* promoter region of *T. atroviride* exhibits an atypical organization because three out of the five CGATN boxes have a different polarity to the canonical WCC binding sites (Fig.4). In spite of this atypical arrangement of the CGATN repeats, the experimental evidence gathered in this study strongly suggests that the inverted elements are functionally equivalent to the direct CGATN repeats, because of the integrity of the -184 to -157 promoter segment, containing two inverted CGATN boxes, was essential to preserve the blue light responsiveness of the truncated or mutated versions of the *phr-1* promoter (see Fig. 2 and 3). Furthermore, there are evidences that the PHRB protein of the basidiomycetous mushroom *Lentinula edodes*, a homologous of WC-2/BLR-2, is able to bind to GATAWWC boxes in both the sense and antisense strand of a photoregulated gene promoter [33]. In this context, it is noteworthy to point out that all known fungal homologs to WC-2 protein bind GATA-related sequences, hence indicating that those regulatory proteins have conserved their DNA-binding specificity since times predating the Ascomycetes-Basidiomycetes evolutionary divergence.

With regard to the issue of the BLRC interaction with the *phr-1* promoter we determined that the BLR-2 protein has the ability to bind to the UCR and PLRR, according to ChIP results. The *in vivo* kinetic enrichment on both regions in response to light suggests that BLR-2 (probably bound to BLR-1) is present in darkness associated with a transcriptional repressor that, upon light stimulus, releases an active BLRC, leading to activation of the *phr-1* transcription immediately after the light pulse (Fig. 6). Taken together these results prompt the hypothesis that BLRC regulates the transcription of blue light-responsive genes by binding to elements analogous to the GATA-1 box (CGATCTC) and the closely associated CGATt/gTC repeats in the antisense strand, that are present in the photoresponsive segment of the *phr-1* promoter of *T. atroviride*. Thus, the LREs in the latter regulatory region resembles the LREs described in *N. crassa* [10-12] because those composite elements are constituted by two or more imperfect CGATN repeats, but differ from the *Neurospora* regulatory elements because some repeats are in the sense DNA strand, whereas others are in the antisense strand. Those differences among fungal species could explain the differences in light-regulated protein complexes and molecular transduction pathways that have been described in other organisms like *N. crassa* and *A. nidulans* [1,3].

Regarding the UCR, we were unable to associate a function of such sequence to the *phr-1* promoter under the tested conditions, even when BLR-2 binds to this region, probably to a GATA-like box on such region; however, we do not discard a role of this sequence in other processes regulated by the BLRC.

This is the first investigation in which blue light responsive elements in a *Trichoderma* gene promoter are delimited and characterized, and it is also the first work demonstrating that a promoter section harboring imperfect CGATN repeats in both the sense and antisense strands is able to drive photoinduction of the *phr-1* and probably of other genes. However, much work needs still to be done to elucidate the regulatory mechanisms that underlie the photoinducible gene expression in *Trichoderma*.

## Materials and methods

### Strains, media and growth conditions

*Trichoderma atroviride* wild-type strain IMI 206040 (ATCC 32173), *Trichoderma virens* wild-type strain Tv 29.8 [34], and *Trichoderma reesei* wild-type strain QM9414 (ATCC 26921) were grown on Potato Dextrose Agar (PDA) (Difco, Franklin Lakes, NJ USA) at 25 °C. *Escherichia coli* strain TOP10F' (Invitrogen, Carlsbad, CA USA) was used for plasmid DNA transformation. Plasmid used was pCB-lacZ (see below).

### Construction of pCB-lacZ and different *phr-1* promoter versions

pCB-lacZ was generated by inserting a 3.0 kb *Sma* I-*Hind* III fragment from pDE1 plasmid [35], containing the *E. coli lacZ* gene fused to the *Aspergillus nidulans trpC* terminator, into their corresponding restriction sites of pCB1004 [36], which harbors the hygromycin B resistance cassette. pCB-lacZ was the parental vector for all *phr-1* promoter versions and chimerical constructs (Fig. S4).

All the promoter versions of *phr-1* gene were PCR amplified using genomic DNA of *T. atroviride* as a template (Fig. 2A). Forward and Reverse-Universal primers used to amplify the different *phr-1* promoter versions included a *Spe* I and *Sma* I restriction sites, respectively, to facilitate their cloning into pCB-lacZ plasmid (Table S2). PCR conditions were 5 min at 94 °C, 25 cycles of 94 °C (30 sec), 58 °C (30 sec), 72 °C (30 sec for < 500 bp products and 60 sec for > 500 bp products), and a final extension of 72 °C for 5 min.

The pCB-lacZ constructs were used to transform *T. atroviride* protoplasts as described [34]. Transformed protoplasts were selected on PDA containing 100 µg/mL of hygromycin. Genomic DNA isolation from *T. atroviride* wild type and transformants was done as described previously [34]. Candidate strains were analyzed by PCR to determine the presence of intact endogenous *phr-1* gene, using -156 *Spe* I F primer on *phr-1* promoter and *phr1trR* primer on *phr-1* ORF. Insertions of translational fusions on *T. atroviride* genome were verified by amplifying the *phr-1* promoter and *lacZ* ORF, using the primer pair -156 *Spe* I F and *lacZ*-R (see Table S2). The copy number of constructs in the transformants genome was calculated by quantitative PCR, using the  $2^{-\Delta\Delta C_t}$  method [37,38] and DNA

from the strain  $\Delta blr-1$  was used as a calibrator, since Southern analysis has shown that it harbors one copy of the *hph* gene [14].

### **Photoinduction experiments**

General growth conditions and manipulations have been described previously [23]. Briefly, *Trichoderma* spp. preinocula were grown on PDA plates in the dark at 25 °C for 48 h. Mycelial plugs were cut and removed from the leading edge of the actively growing colony and placed on the center of PDA plates with a cellophane sheet overlaying the solid media. For DNA extraction, *T. atroviride* was grown for 48 h at 25 °C and the mycelium was scraped off and immediately frozen in liquid nitrogen. For total RNA extraction, cultures were allowed to grow for 36 h in darkness, exposed to a 1200  $\mu\text{mol m}^{-2}$  pulse of blue light (LEE filter no. 183, fluence rate 5  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ), placed back in the dark at 25 °C, and collected at 30 and 120 min after the exposure to blue light. Mycelia grown in darkness were included as control. At the indicated times, mycelia were collected under red safelight (LEE filter no. 106, fluence rate 0.1  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ) and immediately frozen in liquid nitrogen for further total RNA extraction.

### **RNA isolation and RT-PCR analysis**

Total RNA was isolated using Trizol® Reagent (Invitrogen) according to the manufacturer's protocol. RNA quality was assessed using spectrophotometric methods, and formaldehyde-agarose gel electrophoresis, taking into account the 28S/18S rRNA ratio. Five micrograms of total RNA were DNaseI (RNase-free) (Ambion, Austin, TX USA) treated. cDNA synthesis was performed using SuperScript II Reverse Transcriptase (Invitrogen), following manufacturer's recommendations. cDNA was used as a template for quantitative real-time PCR reactions with gene-specific primers and the Fast SYBR Green Master Mix (Applied Biosystems, Foster City, CA USA) according to manufacturer's recommendations. Data were measured with ABI 7500 detection system (Applied Biosystems) and analyzed with the 7500 Software V2.0. We amplified cDNA for *lacZ*, *phr-1*, and *act-1* with specific primers (Table S2). Expression of individual genes was compared and normalized using the  $2^{-\Delta\Delta C_t}$  method [37], against the level of *act-1* mRNA, which was



found to be constant under different light/darkness conditions evaluated in *T. atroviride* [23]. RT-PCR was performed using GoTaq DNA polymerase (Promega, Fitchburg, WI USA) with the following conditions: 5 min at 94 °C, 25 cycles of 94 °C (30 sec), 57 °C (30 sec), and 72 °C (30 sec), and a final extension of 72 °C for 5 min.

### **Chromatin immunoprecipitation (ChIP) Assays**

*T. atroviride* wild-type was grown as described above for blue light induction experiments. Photoinduced mycelia from different times were fixed in cross-linking buffer (50 mM HEPES, pH 7.4, 137 mM KCl, 1 mM EDTA) containing 1% formaldehyde for 15 min at room temperature under red safelight. ChIPs were performed using a BLR-2 antibody (anti-BLR-2, synthesized by Biosynthesis, TX, USA), which recognizes the sequence CEVEEAQRQWAQSRDGRSDI of BLR-2 protein. After DNA extraction, the pellets were resuspended in 50 µL of TE and subjected to PCR with a set of primers along the *phr-1* promoter (see Table S2 and Fig. 5B). The primer pairs used were: ChIP R1 F and ChIP R1 R; ChIP R2 F and ChIP R2 R; ChIP R3 F and ChIP R3 R; ChIP R4 F and ChIP R4 R; ChIP R6 F and ChIP R6 R. PCR program consisted of 5 min at 94 °C, 28-30 cycles of 94 °C (30 sec), 58 °C (30 sec), 72 °C (30 sec), and a final extension of 72 °C for 5 min. To ensure that the amplified PCR products were in the linear range, the PCR conditions were calibrated with different amounts of input DNA (cross-linked chromatin before the immunoprecipitation). Band intensities were quantized by optical density analysis using Quantity-One Software (Bio-Rad Laboratories, Hercules, CA USA). As negative controls, mock precipitations were performed in the absence of antibody. Three independent ChIPs were performed.

### **5. Acknowledgements**

This work was supported by grant SEP-CONACYT 83798 to Alfredo Herrera-Estrella and Sergio Casas-Flores. MGCB, EEUR, and TMC are indebted to CONACYT for Ph. D. and Master of Science scholarships.

## References

1. Rodríguez-Romero J, Hedtke M, Kastner C, Muller S & Fischer R (2010) Fungi, hidden in soil or up in the air: light makes a difference, *Annu Rev Microbiol* **64**, 585-610.
2. Herrera-Estrella A & Horwitz, BA (2007) Looking through the eyes of fungi: molecular genetics of photoreception, *Mol Microbiol* **64**, 5-15.
3. Chen CH, Dunlap JC & Loros, JJ (2010) *Neurospora* illuminates fungal photoreception, *Fungal Genet Biol* **47**, 922-929.
4. Ballario P, Vittorioso P, Magrelli A, Talora C, Cabibbo A & Macino G (1996) White collar-1, a central regulator of blue light responses in *Neurospora*, is a zinc finger protein, *EMBO J* **15**, 1650-1657.
5. Crosthwaite SK, Dunlap JC & Loros, JJ (1997) *Neurospora* wc-1 and wc-2: transcription, photoresponses, and the origins of circadian rhythmicity, *Science* **276**, 763-769.
6. Linden H & Macino G (1997) White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in *Neurospora crassa*, *EMBO J* **16**, 98-109.
7. He Q, Cheng P, Yang Y, Wang L, Gardner KH & Liu Y (2002) White collar-1, a DNA binding transcription factor and a light sensor, *Science* **297**, 840-843.
8. Cheng P, Yang Y, Gardner KH & Liu Y (2002) PAS domain-mediated WC-1/WC-2 interaction is essential for maintaining the steady-state level of WC-1 and the function of both proteins in circadian clock and light responses of *Neurospora*, *Mol Cell Biol* **22**, 517-524.
9. Froehlich AC, Liu Y, Loros JJ & Dunlap JC (2002) White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter, *Science* **297**, 815-819.
10. He Q & Liu Y (2005) Molecular mechanism of light responses in *Neurospora*: from light-induced transcription to photoadaptation, *Genes Dev* **19**, 2888-2899.
11. Chen CH, Ringelberg CS, Gross RH, Dunlap JC & Loros JJ (2009) Genome-wide analysis of light-inducible responses reveals hierarchical light signalling in *Neurospora*, *EMBO J* **28**, 1029-1042.
12. Smith KM, Sancar G, Dekhang R, Sullivan CM, Li S, Tag AG, Sancar C, Bredeweg EL, Priest HD, McCormick RF, Thomas TL, Carrington JC, Stajich JE, Bell-Pedersen D,

Brunner M & Freitag M (2010) Transcription factors in light and circadian clock signaling networks revealed by genome wide mapping of direct targets for *Neurospora* white collar complex, *Eukaryot Cell* **9**, 1549-1556.

13. Gressel J & Galun E (1967) Morphogenesis in *Trichoderma*: photoinduction and RNA, *Dev Biol* **15**, 575-598.

14. Casas-Flores S, Ríos-Momberg M, Bibbins M, Ponce-Noyola P & Herrera-Estrella A (2004) BLR-1 and BLR-2, key regulatory elements of photoconidiation and mycelial growth in *Trichoderma atroviride*, *Microbiol* **150**, 3561-3569.

15. Rosales-Saavedra T, Esquivel-Naranjo EU, Casas-Flores S, Martínez-Hernández P, Ibarra-Laclette E, Cortés-Penagos C & Herrera-Estrella A (2006) Novel light-regulated genes in *Trichoderma atroviride*: a dissection by cDNA microarrays, *Microbiol* **152**, 3305-3317.

16. Esquivel-Naranjo EU & Herrera-Estrella A (2007) Enhanced responsiveness and sensitivity to blue light by *blr-2* overexpression in *Trichoderma atroviride*, *Microbiol* **153**, 3909-3922.

17. Carreras-Villaseñor N, Sánchez-Arreguin JA & Herrera-Estrella A (2012) *Trichoderma*: sensing the environment for survival and dispersal, *Microbiol* **158**, 3-16.

18. Schmoll M, Esquivel-Naranjo EU, Herrera-Estrella A (2010) *Trichoderma* in the light of day-physiology and development, *Fungal Genet Biol* **47**, 909-916

19. Heintzen C, Loros LL & Dunlap JC (2001) The PAS protein VIVID defines a clock-associated feedback loop that represses light input, modulates gating and regulates clock resetting, *Cell* **104**, 453-464.

20. Schwerdtfeger C & Linden H (2001) Blue light adaptation and desensitization of light signal transduction in *Neurospora crassa*, *Mol Microbiol* **39**, 1080-1087.

21. Schmoll M, Franchi L & Kubicek CP (2005) Envoy, a PAS/LOV domain protein of *Hypocrea jecorina* (Anamorph *Trichoderma reesei*), modulates cellulase gene transcription in response to light, *Eukaryot Cell* **4**, 1998-2007.

22. Castellanos F, Schmoll M, Martínez P, Tisch D, Kubicek CP, Herrera-Estrella A & Esquivel-Naranjo, EU (2010) Crucial factors of the light perception machinery and their impact on growth and cellulase gene transcription in *Trichoderma reesei*, *Fungal Genet Biol* **47**, 468-476.

23. Berrocal-Tito G, Sametz-Baron L, Eichenberg K, Horwitz BA & Herrera-Estrella A (1999) Rapid blue light regulation of a *Trichoderma harzianum* photolyase gene, *J Biol Chem* **274**, 14288-14294.
24. Berrocal-Tito GM, Rosales-Saavedra, T, Herrera-Estrella A & Horwitz BA (2000) Characterization of blue-light and developmental regulation of the photolyase gene *phr-1* in *Trichoderma harzianum*, *Photochem Photobiol* **71**, 662-668.
25. Carattoli A, Cogoni C, Morelli G, Macino G (1994) Molecular characterization of upstream regulatory sequences controlling the photoinduced expression of the albino-3 gene of *Neurospora crassa*, *Mol Microbiol* **13**, 787-795.
26. Schuster A, Kubicek CP, Friedl MA, Druzhinina IS & Schmoll M (2007) Impact of light on *Hypocrea jecorina* and the multiple cellular roles of ENVOY in this process, *BMC Genomics* **8**, 449.
27. Casas-Flores S, Rios-Momberg M, Rosales-Saavedra T, Martinez-Hernandez P, Olmedo-Monfil V & Herrera-Estrella A (2006) Cross talk between a fungal blue-light perception system and the cyclic AMP signaling pathway, *Eukaryot Cell* **5**, 499-506.
28. Berrocal-Tito GM, Esquivel-Naranjo EU, Horwitz BA & Herrera-Estrella A (2007) *Trichoderma atroviride* PHR1, a fungal photolyase responsible for DNA repair, autoregulates its own photoinduction, *Eukaryot Cell* **6**, 1682-1692.
29. Callebaut I, Malivert L, Fischer A, Mornon JP, Revy P & de Villartay JP (2006) Cernunnos interacts with the XRCC4 x DNA-ligase IV complex and is homologous to the yeast nonhomologous end-joining factor Nej1, *J Biol Chem* **281**, 13857-13860.
30. Ruiz-Roldán MC, Garre V, Guarro J, Mariné M & Roncero MI (2008) Role of the white collar 1 photoreceptor in carotenogenesis, UV resistance, hydrophobicity, and virulence of *Fusarium oxysporum*, *Eukaryot Cell* **7**, 1227-1230
31. Gresik M, Kolarova N & Farkas V (1988) Membrane potential, ATP, and cyclic AMP changes induced by light in *Trichoderma viride*, *Exp Mycol* **12**, 295–301.
32. Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, Mukherjee M, Kredics L, Alcaraz LD, Aerts A, Antal Z, Atanasova L, Cervantes-Badillo MG, Challacombe J, Chertkov O, McCluskey K, Coulpier F, Deshpande N, von Döhren H, Ebbola DJ, Esquivel-Naranjo EU, Fekete E, Flipphi M, Glaser F, Gómez-Rodríguez EY, Gruber S, Han C, Henrissat B, Hermosa R, Hernández-Oñate M, Karaffá L, Kosti I, Le Crom S, Lindquist E, Lucas S, Lübeck M, Lübeck PS, Margeot A, Metz B, Misra M, Nevalainen H, Omann M, Packer N, Perrone G, Uresti-Rivera EE, Salamov A, Schmoll M, Seiboth B, Shapiro H,

Sukno S, Tamayo-Ramos JA, Tisch D, Wiest A, Wilkinson HH, Zhang M, Coutinho PM, Kenerley CM, Monte E, Baker SE & Grigoriev IV (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*, *Genome Biol* **12**, doi: 10.1186/gb-2011-12-4-r40

33. Sano H, Kaneko S, Sakamoto Y, Sato T & Shishido K (2009) The basidiomycetous mushroom *Lentinula edodes* white collar-2 homolog PHRB, a partner of putative blue-light photoreceptor PHRA, binds to a specific site in the promoter region of the *L. edodes* tyrosinase gene, *Fungal Genet Biol* **46**, 333-341

34. Baek JM & Kenerley CM (1998) The *arg-2* gene of *Trichoderma virens*: cloning and development of a homologous transformation system, *Fungal Genet Biol* **23**, 34-44.

35. Ebbole DJ (1990) Vectors for construction of translational fusions to beta-galactosidase, *Fungal Genet Newsl* **37**.

36. Carroll AM, Sweigard JA & Valent B (1994) Improved vectors for selecting resistance to hygromycin, *Fungal Genet Newsl* **41**, 22.

37. Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  method, *Methods* **25**, 402-408.

38. Bubner B & Baldwin IT (2004) Use of real-time PCR for determining copy number and zygosity in transgenic plants, *Plant Cell Rep* **23**, 263-271.

## Supporting information

**Supplementary Figure S1.** Sequence alignment of *T. atroviride*, *T. virens*, *T. reesei*, *T. citrinoviride*, and *T. longibrachiatum phr-1* intergenic regions.

**Supplementary Figure S2.** Molecular diagnosis of transformant strains.

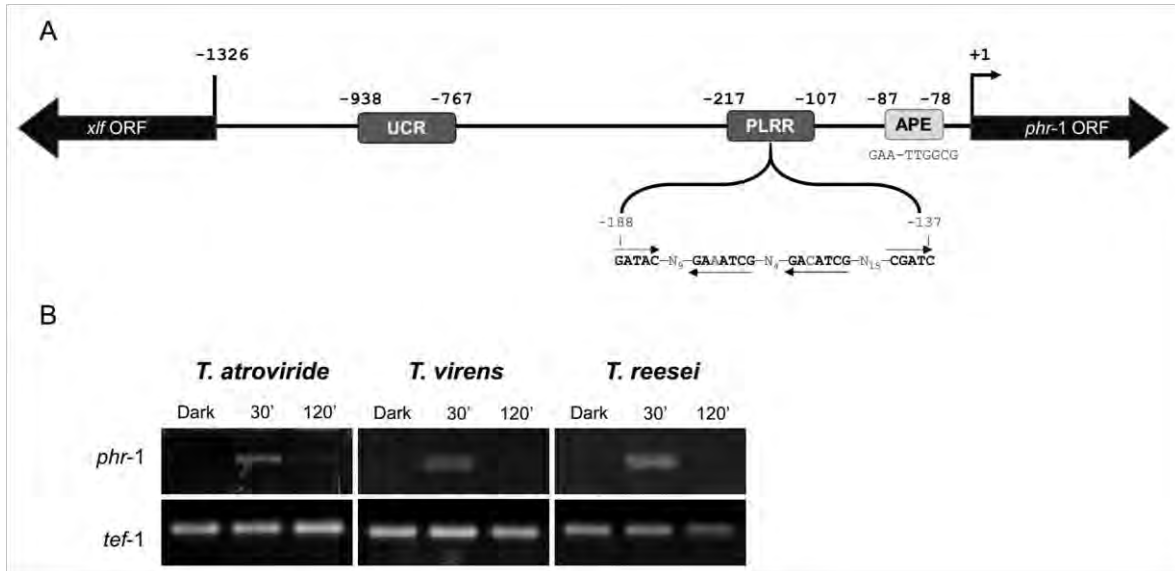
**Supplementary Figure S3.** RT-PCR analysis of control transformant strain carrying the pCB-lacZ empty vector.

**Supplementary Figure S4.** Schematic representation of pCB-lacZ.

**Table S1.** Copy number of *hph* gene in transformant strains, calculated with the  $2^{-\Delta\Delta Ct}$  method.

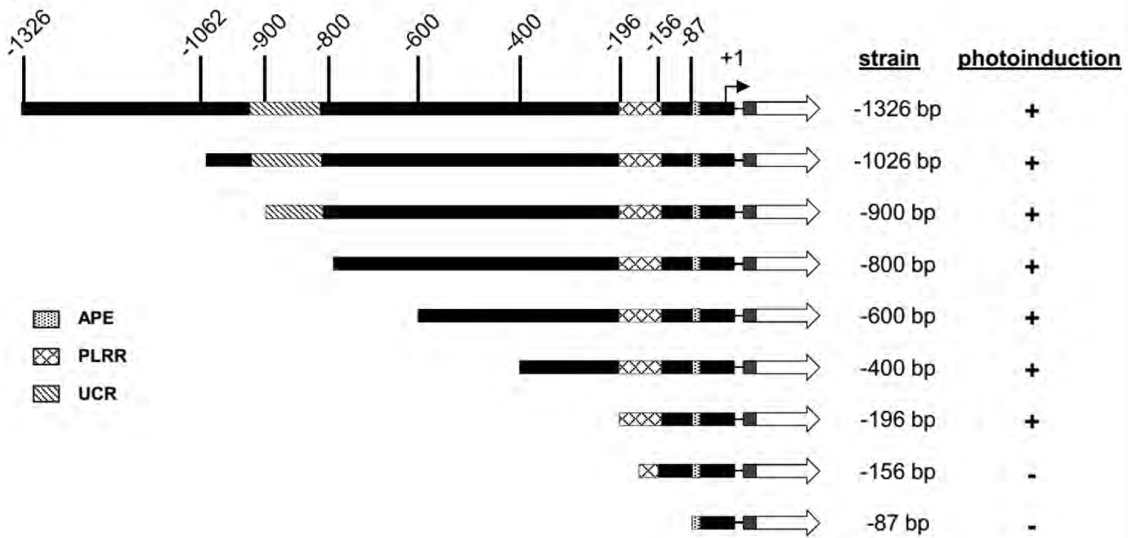
**Table S2.** Oligonucleotides used in this work

## Figures

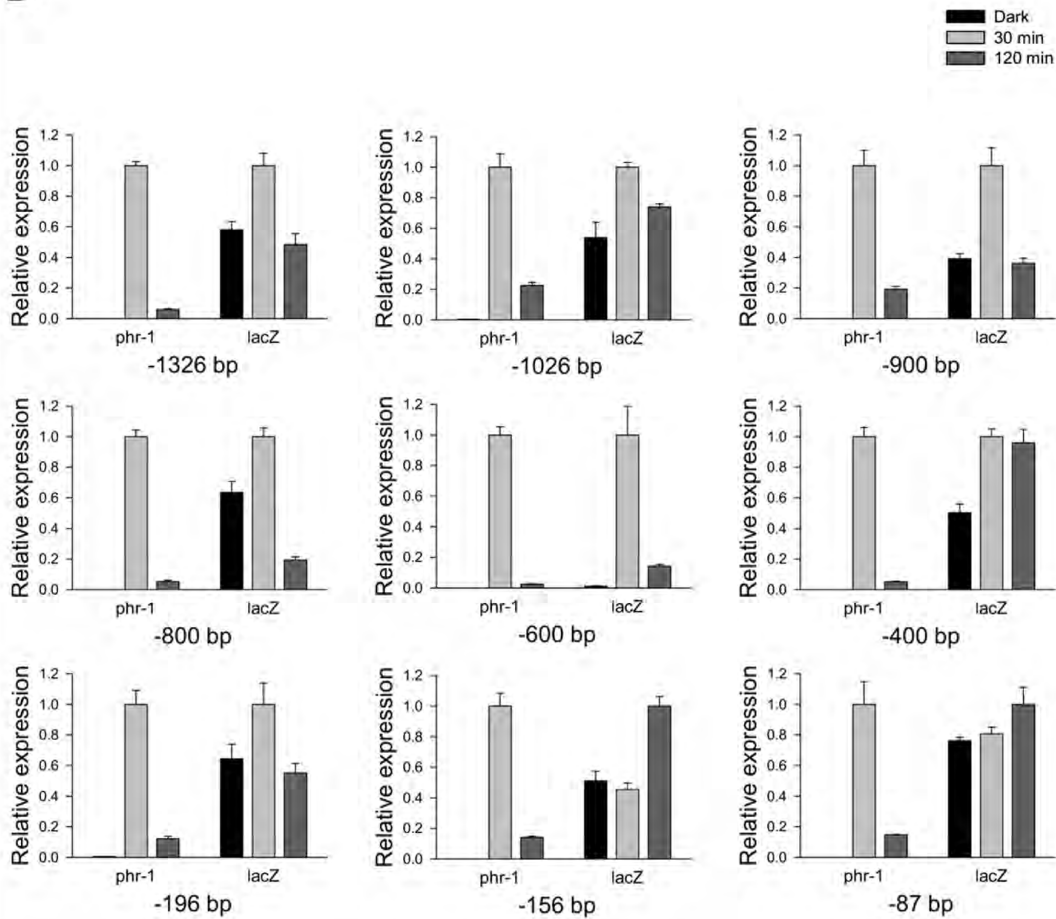


**Figure 1.** Promoter elements of the photoinducible gene *phr-1* are conserved among *Trichoderma* spp. (A) Schematic representation of *phr-1* gene promoter architecture, which resulted from *T. atroviride*, *T. virens*, *T. reesei*, *T. citrinoviride*, and *T. longibrachiatum* *phr-1* promoter sequence alignment (Fig. S1). The upstream conserved region (UCR) and the Potential Light Response Region (PLRR) are depicted by dark grey rectangles. The PLRR contains two inverted CGATN-boxes between the GATA-1 and -2 boxes of the core LRE. The albino proximal element (APE) is shown as a light grey rectangle. The *phr-1* open reading frame (ORF) and the putative upstream *xlf* ORF are indicated in black solid arrows. Location and sequence of potential *cis* acting elements on the promoter are indicated below rectangles for each. Bold numbers on top represent the location of the different elements along the promoter from the transcriptional start site (+1) (thin arrow). (B) RT-PCR analysis of *T. atroviride*, *T. virens*, and *T. reesei* *phr-1* gene expression under dark (control), or 30 and 120 min after a blue light pulse. Elongation factor 1 (*tef-1*) gene was used as housekeeping control gene.

**A**

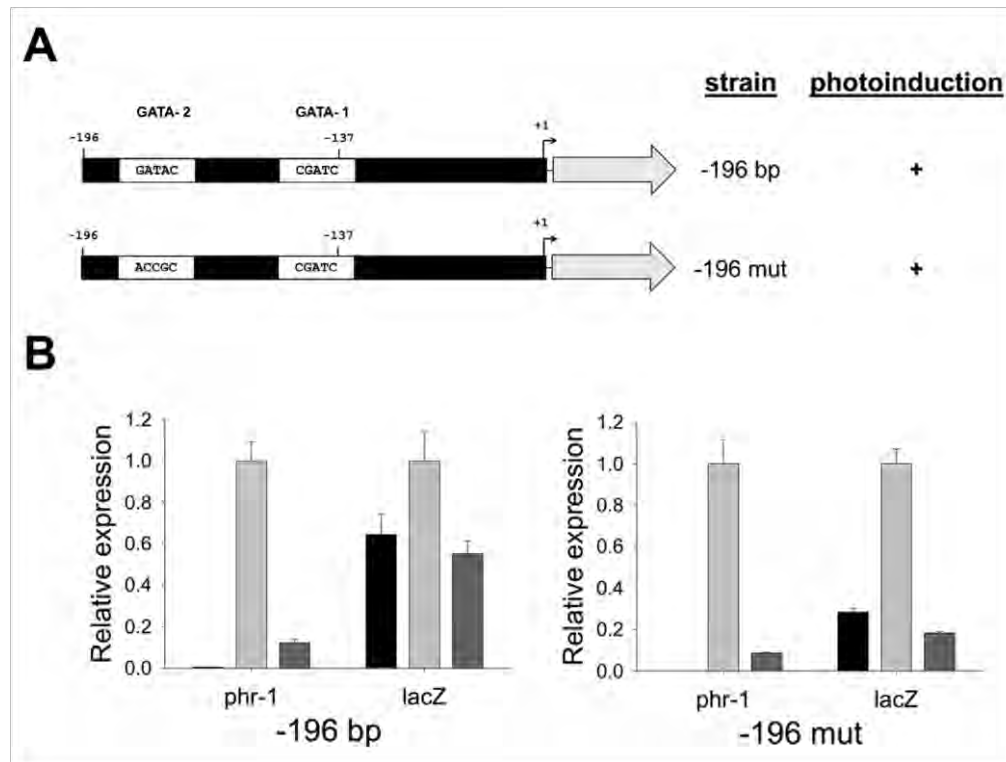


**B**

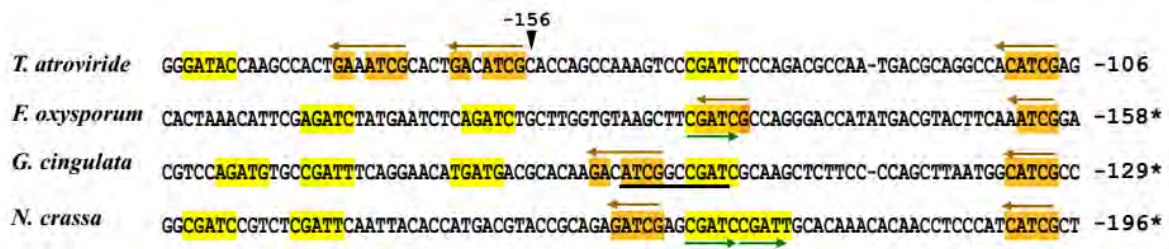




**Figure 2.** A conserved region between -156 and -196 from the transcription start site is responsible for photoinduction of *lacZ* gene. (A) Schematic representation of 5' serial deletions of the *phr-1* promoter translationally fused to the *lacZ* reporter gene (white arrow). Putative APE, PLRR, and UCR *cis*-acting elements are represented by dotted, diamond, and diagonal lines filled rectangles, respectively. Numbers on top represent the distance from the transcription start site (+1) (thin arrow). The ability to photoinduce *lacZ* transcription is denoted by +, whereas - denotes lack of photoinduction of the reporter gene. (B) qRT-PCR analysis of *phr-1* and *lacZ* transcripts in darkness (control) (black bars), or 30 (light grey bars) and 120 min (dark grey bars) after a blue light pulse. The *phr-1* and *lacZ* expressions were normalized to the expression level of *act-1*. The resulting values were normalized to the 30 min expression level after light treatment. Data represent the mean  $\pm$  standard deviation (n = 2 for each strain).



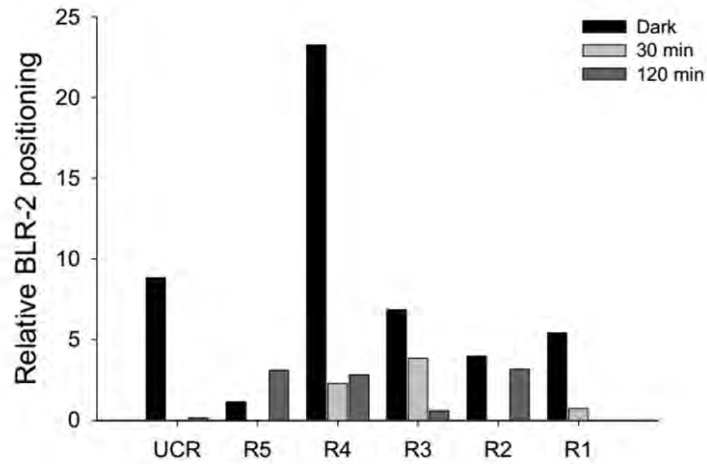
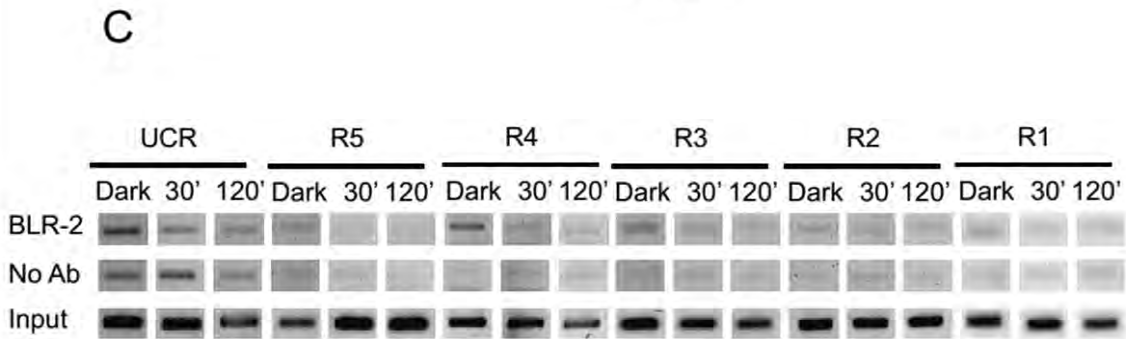
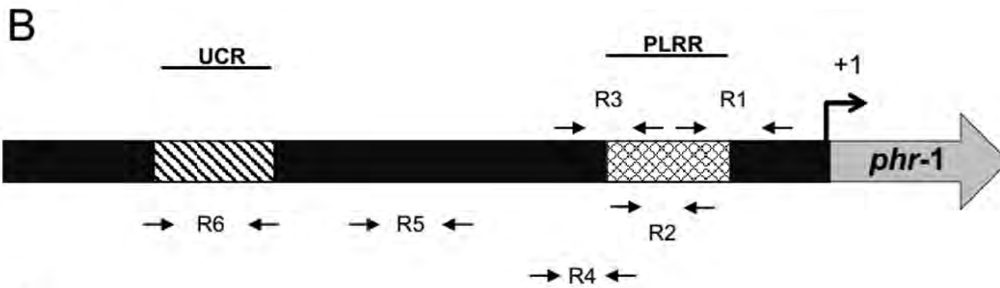
**Figure 3.** Effect of mutations of GATA-2 box on photoinduction of *lacZ* transcription. (A) Schematic representation of wild type (-196 bp) and GATA-2 box mutated (-196 mut) versions of PLRR. A grey arrow represents the *lacZ* reporter gene, and the PLRR is represented by a black rectangle, whereas white rectangles inside the PLRR depict the GATA-1 and GATA-2 boxes. Numbers on top represent the distance from the transcription start site (+1) (thin arrow). The ability to photoinduce *lacZ* transcription is denoted by +, whereas - denotes lack of photoinduction of the reporter gene. (B) qRT-PCR analysis of *phr-1* and *lacZ* transcripts after 30 (light grey bars) or 120 min (dark grey bars) of the application of a blue light pulse or maintained in the dark (black bars) as control. The *phr-1* and *lacZ* expressions were normalized to the expression level of *act-1* and the resulting values were normalized to the 30 min expression level after light treatment. Data represent the mean  $\pm$  standard deviation (n = 2 for each strain).



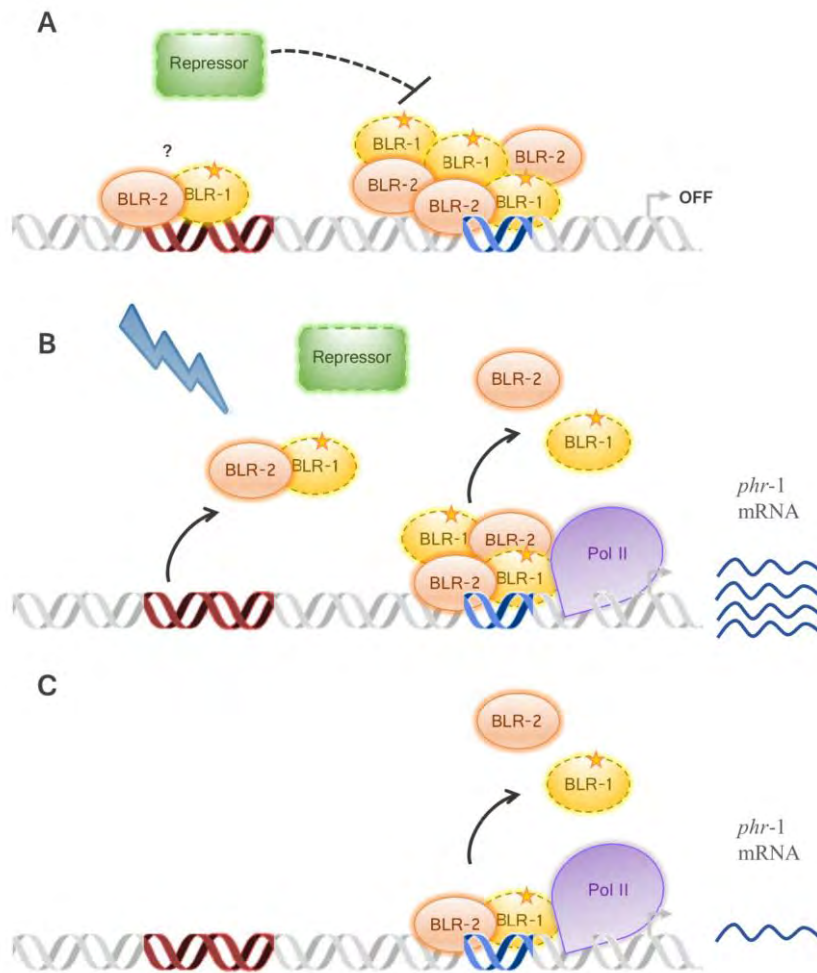
**Figure 4.** Sequence comparisons of the blue light responsive region of *T. atroviride* *phr-1* promoter with homologous non-coding sequences of three species of Sordariomycetes. Sequence alignment of the region -106 to -190 of the *phr-1* promoter of *T. atroviride* with the homologous regulatory regions of three distant relatives. The number on the right side of the *T. atroviride* sequence indicates the nucleotide position relative to the transcription start site, whereas numbers with an asterisk indicates the nucleotide positions relative to the start codon of *Fusarium oxysporum*, *Neurospora crassa* and *Glomerella cingulata* (GenBank accession number AF500083.1 and X58713.1, and ID: 1829160\* respectively) *phr-1* genes. Number 156 on top and the black arrow below it marks the 5'border of the truncated version of the *phr-1* promoter that did not activates the *lacZ* expression after a blue light pulse. Putative light responsive motifs (CGATC or imperfect GATN repeats) displaying the polarity of canonical WCC binding sites are yellow highlighted, whereas GATN motifs showing opposite polarity (indicated by brown arrows) are dark orange colored. Green arrows depict tandem repeats of a CGATN motif in *N. crassa*. A 10-bp perfect palindrome in the *Glomerella* sequence is underlined. \* <http://genome.jgi-psf.org/Gloci1/Gloci1.home.html>

**A**

<i>phr-1</i>	-191	AGGGATACC-N <sub>39</sub> -CC <b>CGATC</b> TC	-134
<i>al-3</i>	-860	AG <b>CGATC</b> TT-N <sub>19</sub> -AA <b>CGATC</b> TC	-829
<i>grg-1</i>	-464	CG <b>CGATC</b> TC-N <sub>10</sub> -CAAGATGCA	-435
<i>env-1 p</i>	-477	CTCGATGGT-N <sub>19</sub> -TGT <b>GATC</b> TG	-440
<i>env-1 d</i>	-674	CC <b>CGATC</b> GT-N <sub>23</sub> -CC <b>CGATC</b> GC	-633
<i>frq p</i>	-335	CT <b>CGATCGA</b> - - - <b>TCGATC</b> GC	-319
<i>frq d</i>	-1180	GG <b>CGATC</b> AA	-1171

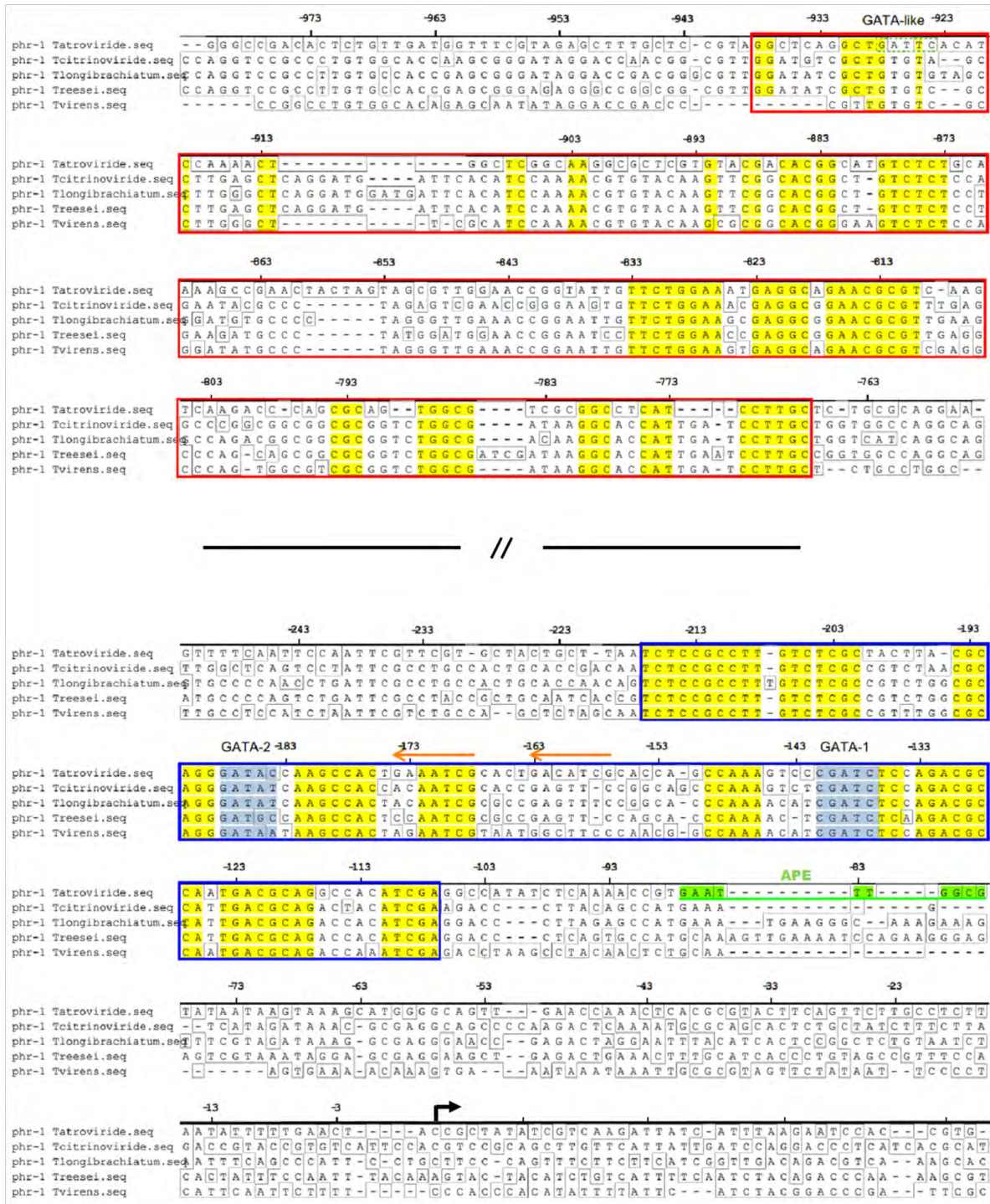


**Figure 5.** Binding of BLR-2 through the *phr-1* promoter in *T. atroviride*. (A) Promoter sequences alignment of representative light responsive genes in *T. atroviride*; putative LRE are highlighted in dash boxes, and putative GATA boxes consensus is indicated in bold letters. Nucleotide positions are numbered relative to the TSS (+1). (B) Scheme of *phr-1* promoter. Six pairs of primers were designed to scan the binding of BLR-2 to the *phr-1* promoter by ChIP assays using an antibody against BLR-2. Location of each primer is indicated along the promoter scheme (see supplementary table 2 for details). PLRR region is filled with diamonds whereas UCR region is filled with diagonal lines. (C) ChIP assay was carried out from wild-type strain mycelia grown in darkness (Dark), or exposed to a blue light pulse and collected 30 (light grey bars) and 120 min (dark grey bars) after the light stimulus. Upper panel, representative PCR amplification of purified DNA samples of BLR-2 immunoprecipitation (BLR-2), no antibody (No Ab) and input (IP). Lower panel, histograms represent the difference between the values for BLR-2 and No Ab PCR products immunoprecipitated, divided by the PCR value with input DNA. Three independent ChIP experiments were performed.



**Figure 6.** Model of *phr-1* photoactivation by the BLR complex. A) Our results showed that BLR-2 and probably BLR-1 proteins (BLRC) are positioned along the promoter of *phr-1*, including the PLRR (DNA strand blue colored) and the UCR (DNA strand red colored) in the dark, remaining repressed by a putative repressor (green rectangle). B) Light treatment of *Trichoderma* leads to derepression of the BLRC, enabling activation and rapid transcription of *phr-1* by RNA polymerase II (Pol II). C) As time elapses (in the dark conditions), the BLR proteins leave the promoter leading to the decrease of the *phr-1* transcript, system desensitization, and repositioning of the BLR proteins on the promoter to be ready to activate the system upon light application. Solid lines indicate supporting evidence from experimental data, and dotted lines (including proteins) indicate hypothetical steps.

# Supplementary Figure S1



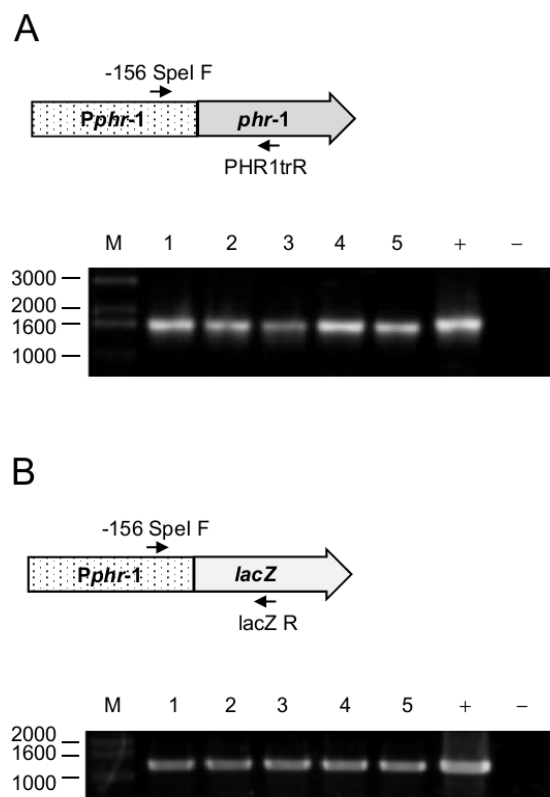
**Supplementary Figure S1.** Sequence alignment of *T. atroviride*, *T. virens*, *T. reesei*, *T. citrinoviride*, and *T. longibrachiatum phr-1* intergenic regions. The highly conserved regions of *phr-1* promoters are highlighted in yellow. The Upstream Conserved Region (UCR) is enclosed with a red rectangle, which contains a GATA-like box (green dashed rectangle). The Potential Light Response Region (PLRR) is located from -107 to -217 position (relative to the *T. atroviride phr-1* TSS) and is enclosed with a blue rectangle. The LRE core similar to that described in *N. crassa* consists of the GATA1 and GATA2 boxes (highlighted in light blue). Two 11-bp imperfect tandem repeats to GATA-1 box are located at -178 to -157, which in the antisense strand displays a CGAT(t/g)TCAGTG sequence (orange arrows). The Albino Proximal Element (APE), which is present only in *T. atroviride*, is highlighted in green. Top numbers indicate the distances relative to transcription start site (black arrow) of the *T. atroviride phr-1* promoter.

**Cervantes-Badillo *et al.*, FEBS Journal (DOI: 10.1111/febs.12362)**

**Supplementary Figure S1**



## Supplementary Figure S2

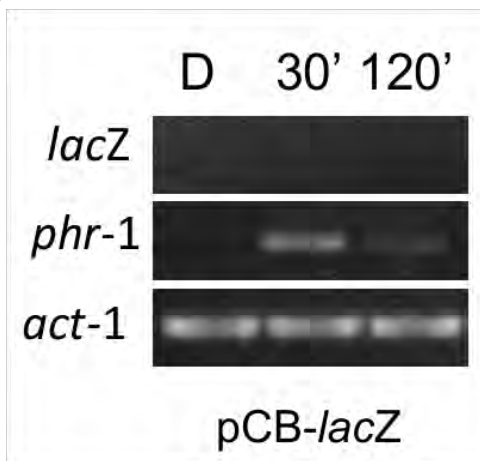


**Supplementary Figure S2.** Molecular diagnosis of transformant strains. (A) Schematic representation of the *phr-1* gene. The dotted rectangle represents the *phr-1* promoter and the gray arrow depicts the *phr-1* ORF. Small black arrows indicate the primers designed to detect the *phr-1* gene. The PCR expected product size is of 1630 bp. Upper numbers (1-5) indicate representative transformants. DNA from the wild-type strain was used as positive control (+). (B) Schematic representation of *Pphr-1::lacZ* constructs. The dotted box represents *phr-1* promoter, whereas the white arrow indicates the *lacZ* ORF. Small black arrows indicate the primers used to detect the construct. The PCR expected product size is of 1247 bp. Upper numbers (1-5) indicate representative transformants. *Pphr1-1326::lacZ* plasmid was used as positive control (+), and the PCR mix without DNA was used as negative control (-); 1 kb DNA ladder molecular size markers (M).

Cervantes-Badillo *et al.*, FEBS Journal (DOI: 10.1111/febs.12362)

Supplementary Figure S2

### Supplementary Figure S3

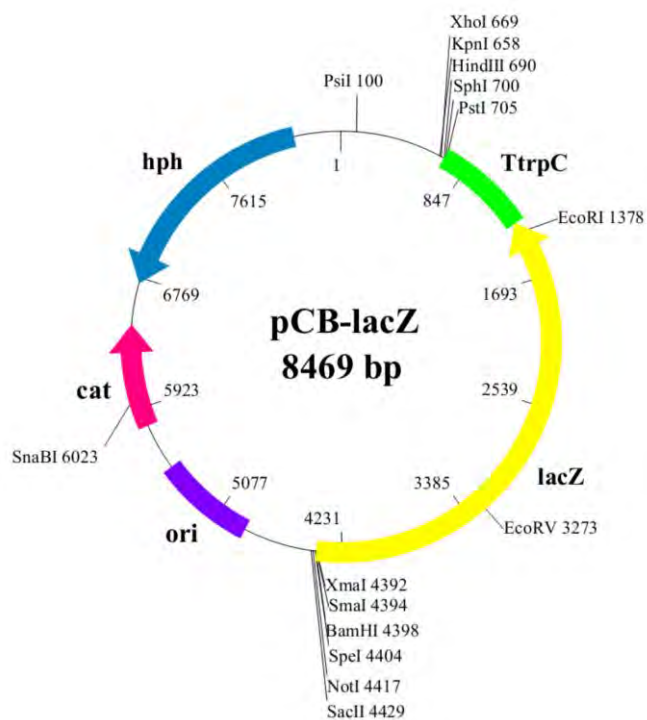


**Supplementary Figure S3.** RT-PCR analysis of control transformant strain carrying the pCB-lacZ empty vector. cDNA generated from mycelia of pCB-lacZ transformant grown in the dark (D), or exposed to a pulse of blue light, and collected at the indicated times were subjected to RT-PCR. The *lacZ* reporter transcript is shown in the first line (*lacZ*), whereas the *phr-1* endogenous transcript was used as light-induced gene control (*phr-1*). Actin was used as housekeeping control gene (*act-1*).

Cervantes-Badillo *et al.*, FEBS Journal (DOI: 10.1111/febs.12362)

Supplementary Figure S3

#### Supplementary Figure S4.



**Supplementary Figure S4.** Schematic representation of pCB-lacZ. pCB-lacZ was generated by inserting a 3.0 kb *Sma* I-*Hind* III fragment from pDE1 plasmid [35], containing the *E. coli lacZ* reporter gene (yellow arrow) fused to the *Aspergillus nidulans* trpC terminator (green rectangle), into the corresponding restriction sites of pCB1004 [36], which harbors the hygromycin B resistance cassette (blue arrow). The plasmid also carries the *cat* gene (pink arrow), which encodes for the chloramphenicol acetyltransferase enzyme, to select resistant *Escherichia coli* transformants. Translational fusion can be inserted into the multiple cloning sites, which have *Xma* I, *Sma* I, *Bam* HI, *Spe* I, *Not* I and *Sac* II restriction enzyme sites available. pCB-lacZ vector was developed for construction of translational fusions of filamentous fungal promoters to the bacterial reporter gene  $\beta$ -galactosidase. Unique restriction sites and locations are indicated.

Cervantes-Badillo *et al.*, FEBS Journal (DOI: 10.1111/febs.12362)

Supplementary Figure S4

**Table S1.** Copy number of *hph* gene in transformant strains, calculated with the  $2^{-\Delta\Delta Ct}$  method.

Strain	$\Delta Ct (Ct_{hph} - Ct_{tef-1})$	$\Delta\Delta Ct^{a,b}$	Copy number ( $2^{-\Delta\Delta Ct}$ )	S.D. <sup>c</sup>
-1326	-2.302	-1.520	3	0.205
-1026	-2.186	-1.404	3	0.659
-900	2.720	-1.726	3	0.022
-800	2.269	-2.177	5	0.031
-600	2.595	-1.851	4	0.011
-400	-3.498	-2.971	8	0.359
-196	-4.282	-3.500	11	2.067
-156	-4.663	-3.881	15	0.358
-87	2.655	-1.791	3	0.010
-196mut	2.658	-1.788	3	0.042
pCBlacZ	2.682	-1.764	3	0.004

<sup>a</sup>  $\Delta\Delta Ct = \Delta Ct_{\text{sample}} - \Delta Ct_{\text{calibrator}}$

<sup>b</sup>  $\Delta blr-1$  was used as the calibrator strain

<sup>c</sup> S.D. Standard deviation

Cervantes-Badillo *et al.*, FEBS Journal (DOI: 10.1111/febs.12362)

Supplementary Table S1

**Table S2. Oligonucleotides used in this work**

Primer name	Sequence (5'-3') <sup>a,b</sup>	Gene amplified	Position <sup>c</sup>
-1326 SpeI F	<u>GACTAGT</u> ATGGCGTCACAAAAGGGG	<i>phr-1</i>	-1326
-900 SpeI F	<u>GACTAGT</u> GATTCACATCCAAAAGGGCTC GG	<i>phr-1</i>	-900
-800 SpeI F	GG <u>ACTAGT</u> TGCGCAGGAATAGTGG	<i>phr-1</i>	-800
-600 SpeI F	CC <u>ACTAGT</u> GCTTTCTAGGTCATCCAA	<i>phr-1</i>	-600
-400 SpeI F	GG <u>ACTAGT</u> CGTGGTTGCTCTTGCAT	<i>phr-1</i>	-400
-196 SpeI F	<u>ACTAGT</u> TACGCAGGGATAACCAAGCCA	<i>phr-1</i>	-196
-196mut SpeI F	<u>ACTAGT</u> TACGCAGG <b>ACCC</b> CCAAGCCA	<i>phr-1</i>	-196
-156 SpeI F	<u>ACTAGT</u> CACCAGCCAAAGTCCCGATC	<i>phr-1</i>	-156
UPphr1 SmaIR	<u>CCCGGGG</u> ACGCTCCTCGCGAGCATTT	<i>phr-1</i>	+63
lacZ-F	AAAACCCGAAACTGTGGAGC	<i>lacZ</i>	+869
lacZ-R	TGCTCATCCATGACCTGACC	<i>lacZ</i>	+1091
phr1trF	TCATGTTTGGATCAGCGAAGTCG	<i>phr-1</i>	+1339
phr1trR	TACTCGGGCTTGAATGGCTTGT	<i>phr-1</i>	+1474
phr-1 TA-F	TTATGATCGGCCGAGAAAAAT	<i>phr-1</i>	+993
phr-1 TA-R	CAGCGTCGACAATGGGAAATC	<i>phr-1</i>	+1557
ChIP R1 F	ACCAGCCAAAGTCCCGATC	<i>phr-1</i>	-317
ChIP R1 R	GAGGCAAGAAGTGAAGTAC	<i>phr-1</i>	-17
ChIP R2 F	CTCGCTACTTACGCAGGGATAC	<i>phr-1</i>	-205
ChIP R2 R	ATATGGCCTCGATGTGGC	<i>phr-1</i>	-98
ChIP R3 F	TCATCCAGCTGATGGCTTATGC	<i>phr-1</i>	-317
ChIP R3 R	ATTCAGTGGCTTGGTATCCC	<i>phr-1</i>	-169
ChIP R4 F	CGACTGCATTATATGTAGGGTCATC	<i>phr-1</i>	-337
ChIP R4 R	TGGCTTGGTATCCCTGCGTAAGTAG	<i>phr-1</i>	-177
ChIP R5 F	ATCCAACACGTCGCTTGCCTCTTTC	<i>phr-1</i>	-550
ChIP R5R	GCGGTACCTTGTACCTTTACATAGT	<i>phr-1</i>	-426
ChIP R6 F	CTCAGGCTGATTCACATCCAAAAGTGGC	<i>phr-1</i>	-936
ChIP R6 R	CGTTCTGCCTCATTTCCAGAACAATACCGG	<i>phr-1</i>	-813
ACTINtr-F	TCACCGAGGCCCCCATCAACC	<i>act-1</i>	
ACTINtr-R	CGACCGGAAGCGTACAGGGACAGA	<i>act-1</i>	
tef-1 F	AGGCCGAGCGTGAGCGTGGTAT	<i>tef-1</i>	
tef-1 R	ATGGGGACGAAGGCAACGGTCTT	<i>tef-1</i>	

<sup>a</sup> Restriction sites are indicated as underlined letters

<sup>b</sup> Mutations are indicated in bold letters

<sup>c</sup> Positions were numbered relative to the transcriptional start site (+1) of the *phr-1* gene or relative to ATG for *lacZ* gene

## Resultados parte II

### ***Ensayos de cambio de movilidad electroforética***

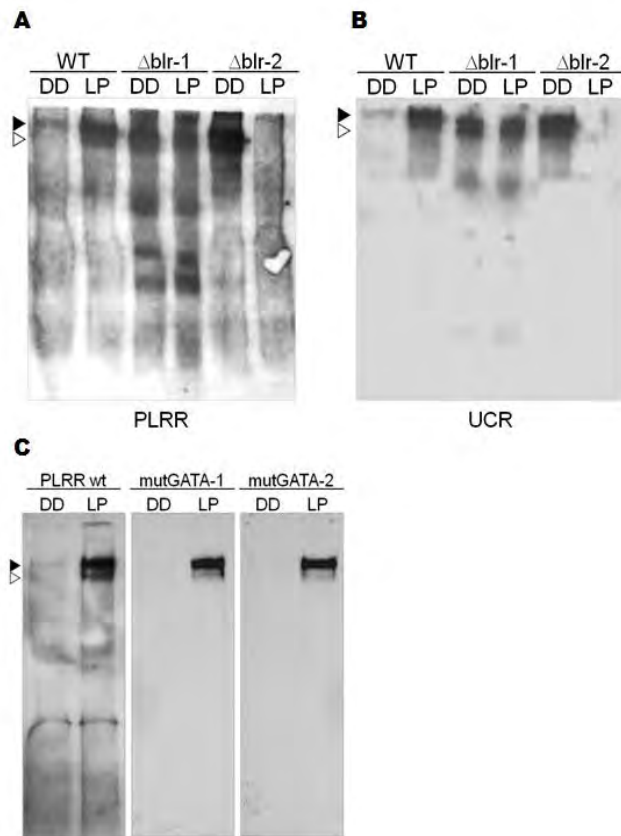
En *N. crassa*, el complejo WCC se une a los ELRE localizados en el promotor de los genes regulados por luz, en condiciones de oscuridad y luz (Froehlich *et al.*, 2002; He y Liu, 2005; Smith *et al.*, 2010). El análisis de las secuencias de los promotores de varios genes regulados por luz en *T. atroviride* mostró que algunos genes presentan dos regiones similares a un LREs, mientras que otros tienen solo una. Debido a que en *T. atroviride* la inducción por luz de *phr-1* es dependiente de BLR-1 y BLR-2 (Casas-Flores *et al.*, 2004), inferimos que las proteínas BLR podrían interactuar con las regiones conservadas del promotor de *phr-1*. Para probar esta hipótesis, realizamos una serie de ensayos de cambio de movilidad electroforética (EMSA) con extractos de proteína nuclear de las cepas silvestre y mutantes  $\Delta blr-1$  y  $\Delta blr-2$  de *T. atroviride*, junto con sondas de las regiones PLRR y UCR del promotor de *phr-1*. Adicionalmente, se diseñaron sondas con mutaciones en cada una de las cajas GATA de la región responsiva a luz del promotor de *phr-1* (ver Tabla 3; Materiales y Métodos).

Los EMSA se realizaron como se ha descrito previamente (Froehlich *et al.*, 2002; He y Liu, 2005) con extractos de proteínas nucleares que se mantuvieron en oscuridad o a las que se expuso a luz azul durante la reacción de unión. Utilizando la sonda PLRRwt, la incubación con la fracción de proteínas nucleares de la cepa silvestre en oscuridad resultó en una señal débil (Fig. 12A, flecha negra). Mientras que el tratamiento con luz azul indujo el enriquecimiento del complejo PLRR/proteína y la aparición de un segundo complejo de menor peso molecular (Fig. 12A, flecha blanca). Estos datos nos sugieren *i*) una mayor afinidad de las proteínas por el DNA en respuesta a la luz, *ii*) el desensamble de una o más proteínas del complejo DNA/proteína en respuesta a la luz. Se realizaron EMSAs similares con la sonda UCR y se observó un patrón de cambio de movilidad

electroforética similar al obtenido con la sonda PLRRwt (Fig. 12B). Estos datos sugieren que estas dos regiones podrían ser reconocidas por el BLRC.

Para conocer un poco más sobre el papel de BLR-1 y BLR-2 en la formación de los complejos DNA/proteína observados en estos ensayos, se realizaron EMSAs con proteínas nucleares de las mutantes  $\Delta blr-1$  y  $\Delta blr-2$ . La incubación con las proteínas nucleares de la cepa  $\Delta blr-1$  en oscuridad generaron dos señales principales similares a las observadas con la cepa silvestre, pero de menor peso molecular, y con un patrón electroforético similar en luz y oscuridad. Adicionalmente, observamos un conjunto de bandas de menor peso molecular, respecto de las dos señales principales con velocidad electroforética baja (Fig. 12A), lo cual sugiere que BLR-2 podría estar presente en el complejo DNA/proteína (por los datos de los ChIPs), pero que se requiere de BLR-1 activo para formar los complejos en respuesta a la luz. Para el caso de la sonda UCR, también se observaron dos señales principales de baja movilidad electroforética semejantes a las obtenidas con la sonda PLRRwt, y solo una señal adicional de menor peso (Fig. 12B).

Los EMSA realizados con las proteínas nucleares de  $\Delta blr-2$  y las sondas PLRRwt y UCR, mostraron dos complejos DNA/proteína en oscuridad semejantes a los que ya se habían observado con los otros ensayos, pero con un peso intermedio (Fig. 12A y B). Sin embargo, en la condición de luz ya no se observa ningún retardo en el corrimiento de las sondas, lo cual sugiere *i*) que posiblemente BLR-1 esté unido en oscuridad y una vez activo pierda la capacidad de unirse al DNA y *ii*) que se requiera de BLR-2 para mantener la integridad del complejo que se une al DNA.



**Figura 12. La luz induce la unión de proteínas a las regiones PLRR y UCR *in vitro*.** A-B) Unión de proteínas nucleares de distintos fondos genéticos con la sonda LREwt (A) o UCR (B). C) EMSA con sondas del LRE mutado en la caja GATA-1 o en la caja GATA-2, con proteínas nucleares de la cepa silvestre, bajo distintas condiciones de luz. La flecha negra indica el complejo principal observado en condición de oscuridad (DD), y la flecha blanca indica el complejo principal observado en luz (LP). La sonda libre corrió fuera del gel. Se muestra uno de dos experimentos independientes.

Debido a que no observamos un papel evidente de la UCR en la regulación transcripcional de *phr-1*, nos enfocamos en probar si la mutación de las cajas GATA-1 y GATA-2 (ver Tabla 3; Materiales y Métodos) afecta la unión de las proteínas nucleares de la cepa silvestre. El complejo que se observaba en la oscuridad con la sonda PLRRwt no se observa al introducir mutaciones en las cajas GATA (Fig. 12C), ya que fue muy bajo (sonda mutGATA-2) o nulo (sonda

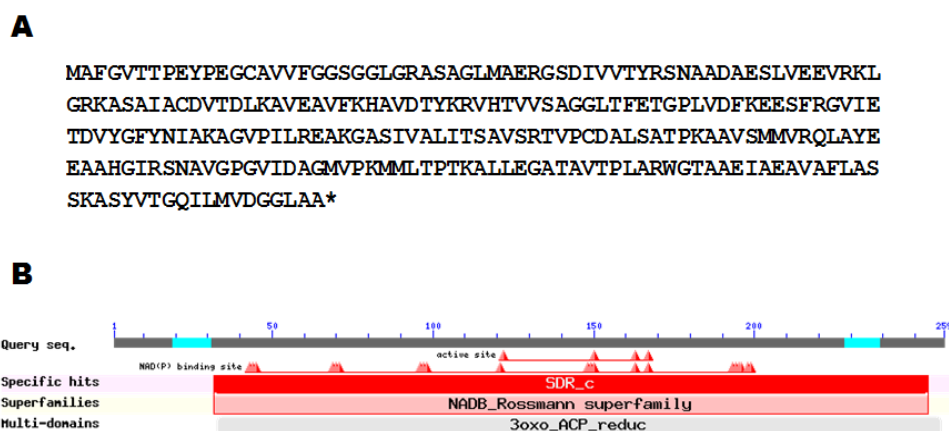


mutGATA-1). Aún cuando se observaron complejos DNA/proteína en la condición de luz, la señal fue más débil respecto de la sonda LREwt, lo que podría explicarse por la presencia de las cajas GATA repetidas invertidas entre la GATA-1 y la GATA-2.

## ***Identificación del sitio de inicio de la transcripción del gen bld-2, diseño y construcción de fusiones traduccionales de distintas versiones del promotor de bld-2 con el gen reportero lacZ***

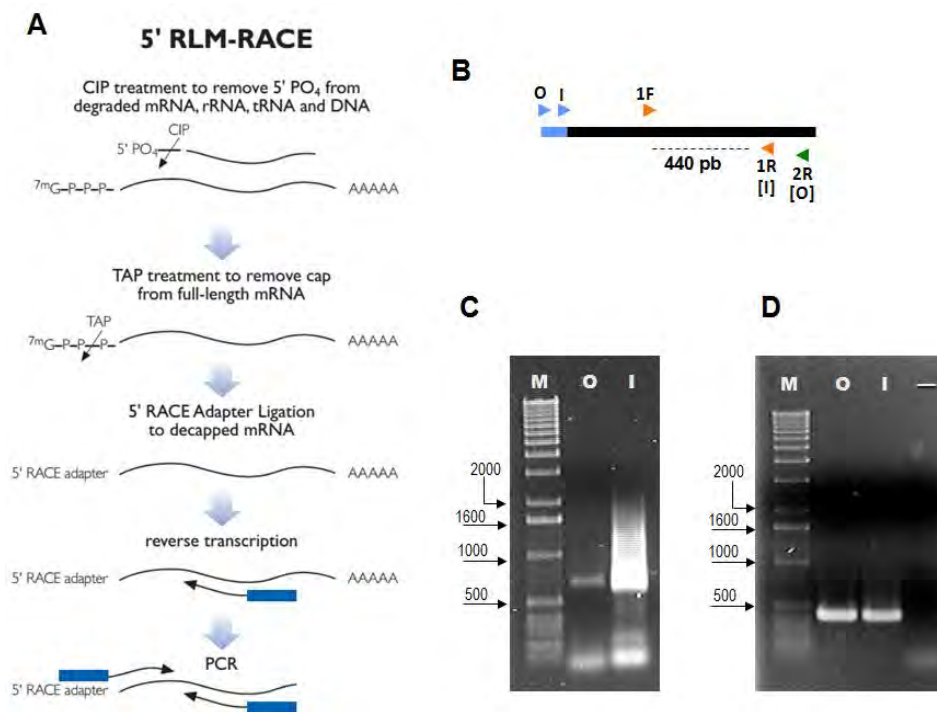
Se identificó el sitio de inicio de la transcripción (TSS, por sus siglas en inglés) del gen *bld-2* (Protein ID: 301399; <http://genome.jgi-psf.org/Triat2/Triat2.home.html>) mediante la técnica 5'-RLM-RACE, a fin de delimitar el promotor de dicho gen, y estudiar su regulación en respuesta a la luz azul.

El gen *bld-2* se expresa en la oscuridad y comienza a bajar su expresión a partir de los 30 min después de que se incide un pulso de luz azul a una colonia de *T. atroviride* (Rosales-Saavedra *et al.*, 2006). El marco de lectura abierto (ORF, por sus siglas en inglés) codifica para una proteína de 259 aminoácidos (Fig. 13A) con posible actividad de deshidrogenasa de la familia NADB-Rossmann (Fig. 13B), las cuales participan en numerosas vías metabólicas.



**Figura 13. Características del producto putativo del gen *bld-2*.** A) Secuencia de aminoácidos predicha para la proteína BLD-2 de *T. atroviride*. BLD-2 consta de 259 aminoácidos. B) Dominios teóricos de BLD-2. Un análisis tipo BLAST con la secuencia de aminoácidos muestra que BLD-2 pertenece a la familia de deshidrogenasas NADB-Rossmann (bloque rosa).

El TSS se identificó con el kit comercial FirstChoice 5' RLM-RACE (Ambion) siguiendo las instrucciones del proveedor (Fig. 14A). De la reacción *outer* se obtuvo un amplicón de 700 pb aproximadamente, la cual se enriqueció en la reacción *inner* (Fig. 14B y C). Para corroborar que el amplicón obtenido fuese específico del gen *bld-2*, se realizó un PCR con los oligonucleótidos específicos que amplifican 440 pb del ORF de este gen (Fig. 14B). Como se observa en la figura 14D, los amplicones obtenidos de las reacciones *outer* e *inner* son específicos de *bld-2*.

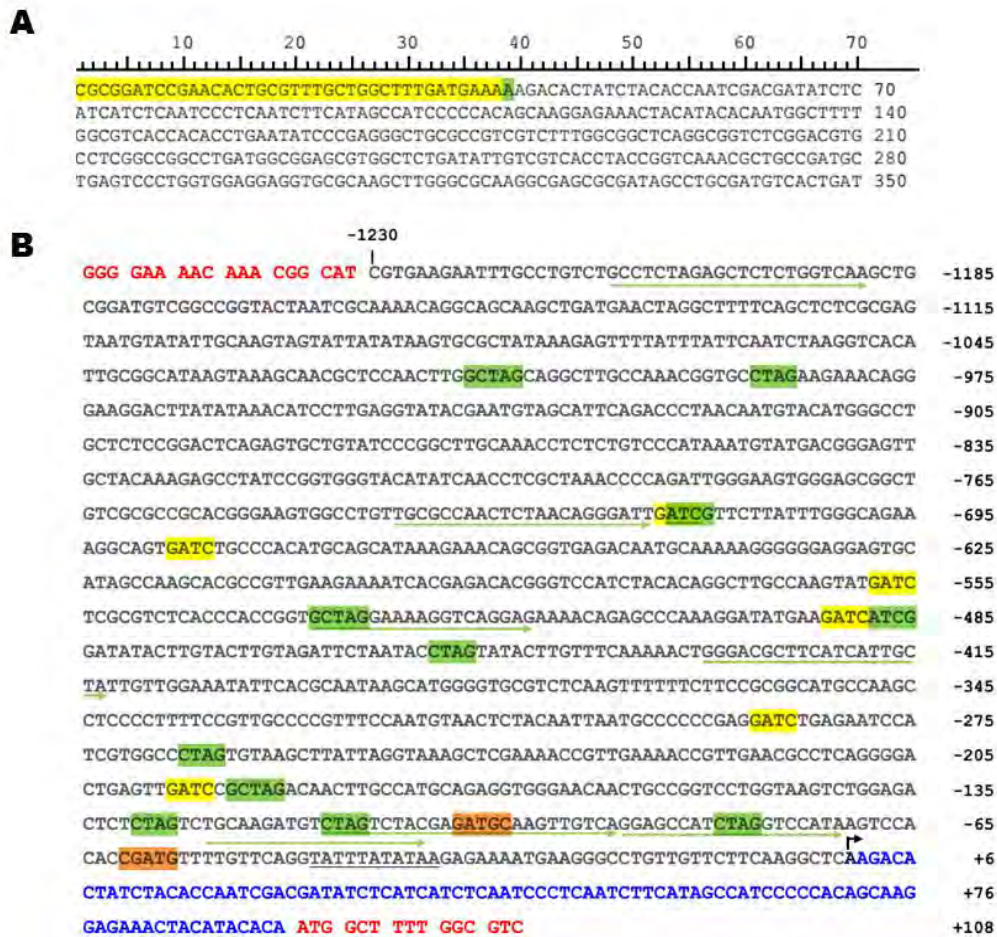


**Figura 14. Obtención del amplicón para la identificación del sitio de inicio de la transcripción del gen *bld-2* de *T. atroviride*.** A) Esquema general del protocolo de 5' RLM-RACE (tomado de la ficha técnica del FirstChoice® RLM-RACE kit, Ambion 2008). B) Esquema de los primers utilizados para cada PCR y tamaño del amplicón específico en el OFR de *bld-2*. O, 5' RACE Outer; I, 5' RACE Inner; 1F, *bld-2* TA Fwd; 1R, *bld-2* TA Rev; 2R, *bld-2* tr Rev. C) Amplicones obtenidos de las reacciones de PCR *outer* (O) e *inner* (I). D) PCR para corroborar la especificidad de los amplicones obtenidos de las reacciones *outer* (O) e *inner* (I), y control negativo (-) con los primers específicos para *bld-2* 1F/1R. M, marcador de 1 kb.

Se clonó el producto de la reacción *inner* en pGEM y se secuenciaron cuatro clonas por el método de Sanger. Una vez obtenida la secuencia, se identificó el adaptador 5' RACE y la siguiente base se tomó como el TSS del gen *bld-2* (Fig. 15A). Se obtuvo la secuencia del gen en la base de datos del genoma de *T. atroviride* y se delimitó el promotor, ubicando el TSS (Fig. 15B). El gen *bld-2* presenta una caja putativa TATA de -45 a -35 del TSS, un 5'-UTR de 93 pb, con el primer codón ATG a +94 del TSS. Además identificamos varias cajas GATC, CGAT o CGATC a lo largo de la región intergénica tanto en la cadena sentido como en la antisentido (Fig. 15B). Observamos la presencia de varias cajas tipo GATA en dos regiones: de -57 a -197 pb y de -451 a -558 pb, lo que sugiere que estas regiones podrían ser importantes para la regulación de *bld-2* por luz azul.

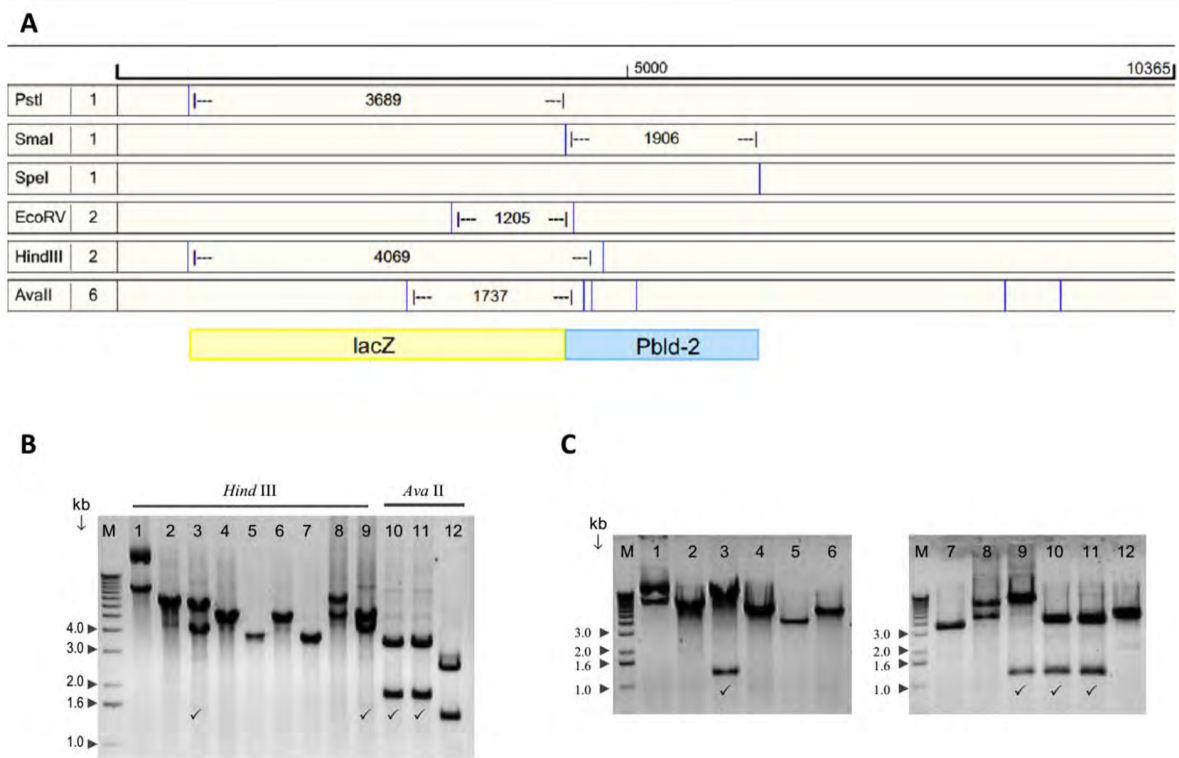
Se realizó un análisis de la región río arriba del TSS de *bld-2* para delimitar la región intergénica. Para ello, se tomaron 4 kb y se analizaron en BLAST, identificando un ORF a -1230 pb del TSS de *bld-2* (Fig. 15B), que codifica para una proteína de actividad desconocida (No. de Acceso en el GeneBank EHK40562.1).

Se diseñaron oligonucleótidos para amplificar fragmentos a varias distancias del TSS de *bld-2* para realizar fusiones traduccionales con el gen reportero *lacZ* (Fig. 15B). Los oligonucleótidos *forward* incluyen el sitio de restricción de *Spe* I y el universal reverso lleva el sitio *Sma* I, a fin de clonar en los mismos sitios del vector pCBlacZ (Tabla 5). Los productos de PCR se clonaron en pGEM-T Easy y se secuenciaron por el método de Sanger.



**Figura 15. Análisis del promotor del gen *bld-2* de *T. atroviride*.** A) Identificación del sitio de inicio de la transcripción de la secuencia obtenida con la técnica de 5'RLM-RACE. El adaptador 5'RACE está resaltado en amarillo y la primera base posterior (verde) se tomó como el sitio de inicio de la transcripción de *bld-2*. B) Delimitación del promotor de *bld-2* en la secuencia genómica. El nucleótido identificado en (A) se indica con una flecha negra, delimitando la región del promotor (letras grises), el 5'-UTR (letras azules) y los primeros 5 codones del marco de lectura abierto (letras rojas). La putativa caja TATA está subrayada. Las cajas GATC en la cadena sentido se resaltan en color amarillo, y en la antisentido en color verde. El consenso igual al LRE reportado por He y Liu (2005) para *N. crassa* se resalta en naranja. A la derecha se muestran las coordenadas respecto al sitio de inicio de la transcripción. Los oligonucleótidos diseñados para las construcciones en pCBlacZ se muestran en flechas verdes.

Los plásmidos pGEM y pCBlacZ se digirieron con *Sma* I y *Spe* I para realizar las fusiones del promotor de *bld-2* con el gen *lacZ*. Las ligaciones se transformaron en la cepa *E. coli* TOP10F' y se realizó un escrutinio de las construcciones correctas utilizando enzimas de restricción (Fig. 16). Inicialmente, se intentó generar otra construcción con la versión de -1794 y usando el sitio *Nhe* I (G<sup>+</sup>CTAG<sub>↓</sub>C) que genera sitios compatibles con *Spe* I (A<sup>+</sup>CTAG<sub>↓</sub>T); sin embargo, no fue posible obtener esa construcción con esta estrategia. En la figura 16 se muestran los análisis de restricción de algunas de las construcciones *Pbld-2::lacZ* generadas.



**Figura 16. Escrutinio de las construcciones *Pbid-2::lacZ* con enzimas de restricción. A)** Mapa de restricción de la versión -1794 del promotor *bid-2* de *T. atroviride* fusionado a *lacZ*. A la izquierda se muestran las enzimas de restricción que se utilizaron, el número de sitios en el plásmido, así como el tamaño del fragmento que se libera por la fusión del promotor con el gen reportero. *Pst* I delimita el extremo 3' de *lacZ*, *Sma* I está en la intersección de *lacZ* con el promotor y *Spe* I delimita el extremo 5' del promotor de *bid-2*. El esquema de la fusión se muestra en la parte inferior del mapa. **B)** Patrón de restricción con las enzimas *Hind* III y *Ava* II. Se digirieron clonas de las versiones -1797 (líneas 1-3), -1797dig *Nhe* I (líneas 4-6) y -738 (líneas 7-9) con *Hind* III y de la versión -113 (líneas 10-12) con *Ava* II. **C)** Patrón de restricción con la enzima *Eco* RV. Se digirieron clonas de las versiones -1797 (líneas 1-3), -1797dig *Nhe* I (líneas 4-6), -738 (líneas 7-9) y -113 (líneas 10-12). Las clonas con el patrón de restricción esperado se indican con una (✓). M, marcador de peso molecular de 1 kb (Invitrogen).

## Discusión

Las especies del género *Trichoderma* se han empleado como modelos fotomorfogénicos porque producen un anillo de conidias verdes en respuesta a la luz, el cual se forma en el perímetro de una colonia en crecimiento al momento en que el estímulo fue aplicado. Después de un pulso de luz, *T. atroviride* también muestra una oscilación bifásica transitoria en los niveles intracelulares de cAMP y ATP, cambios en el potencial de membrana, así como en la inducción y represión de un conjunto de genes, entre otras respuestas (Grešik *et al.*, 1988; Rosales-Saavedra *et al.*, 2006). Desde que Berrocal-Tito y colegas reportaron a *phr-1* como un gen inducible por luz azul en *T. atroviride* (Berrocal-Tito *et al.*, 1999), éste se ha utilizado como un marcador molecular para estudiar las respuestas a la luz azul en éste hongo. Sin embargo, la regulación de la transcripción de este gen es compleja, ya que se expresa por luz y por desarrollo (Berrocal-Tito *et al.*, 1999; Berrocal-Tito *et al.*, 2000). Además, se mostró que la proteína PHR-1 puede regular la expresión de su gen de manera dependiente de la luz, posiblemente a través de la modulación negativa del complejo BLR (Berrocal-Tito *et al.* 2007).

El análisis de los promotores de varios genes regulados por luz azul, identificados por estudios con microarreglos, revelaron la presencia de elementos de respuesta a luz, similares a los reportados en *N. crassa* (He y Liu, 2005) en *T. atroviride* y *T. reesei* (Rosales-Saavedra *et al.*, 2006). Nosotros mostramos que esas secuencias son altamente conservadas en el promotor de *phr-1* de cinco especies de *Trichoderma* (Fig. S1, publicación), e incluso se encuentran en el promotor de varios genes ortólogos de la fotoliasa en otros Sordariomicetos (Fig. 4, publicación). También, determinamos que este núcleo del LRE está presente en varios genes inducidos por la luz azul en *T. atroviride* (Fig 5A, publicación). Un análisis de genómica comparativa de *T. atroviride*, *T. virens* y *T. reesei* indica que *T. atroviride* se parece al estado más ancestral de *Trichoderma* y que tanto *T. virens* como *T. reesei* evolucionaron después. Además, aparentemente *T. reesei*



ha evolucionado más rápido que *T. virens* y *T. atroviride* desde que inició la divergencia evolutiva en éste grupo (Kubicek *et al.*, 2011). El mecanismo molecular que controla la regulación transcripcional debida a la luz azul parece estar conservado en estas especies, como se mostró en el análisis de la secuencia del promotor de *phr-1* (Fig. S1, publicación) y con el perfil transcripcional del gen (Fig. 1B, publicación).

El gen *phr-1* está organizado en el genoma en una configuración transcripcional divergente con el gen *xlf*, lo cual a primera vista sugiere una co-regulación de ambos genes, ya que comparten elementos en *cis*, tal como el LRE y la UCR, y por su función: PHR-1 es una fotoliasa que repara el DNA dañado por luz UV-B, mientras que el factor XRCC4-like (el producto del gen *xlf*) tiene una función putativa en la reparación de cortes en cadenas dobles de DNA. Sin embargo, el gen *xlf* no se induce por luz blanca o azul (Herrera-Estrella A, comunicación personal).

El papel transcripcional de los elementos APE (Berrocal-Tito *et al.*, 1999) y LRE (Rosales-Saavedra *et al.*, 2006) descritos previamente en el promotor del gen *phr-1* de *T. atroviride*, no han sido analizados experimentalmente. En este estudio, nosotros mostramos que el promotor de *phr-1* tiene dos regiones conservadas, la que contiene el LRE (PLRR) y una nueva que nombramos UCR. El perfil transcripcional de *phr-1* en respuesta a la luz muestra que no se expresa en la oscuridad, y luego de un estímulo de luz azul se incrementa su transcripción (Berrocal-Tito *et al.*, 1999). Sin embargo, observamos una ligera expresión en oscuridad del gen reportero *lacZ* en casi todas las construcciones probadas (salvo por la versión de -600 pb), un hecho que contrasta con el patrón de expresión del gen *phr-1* endógeno. Para explicar estos datos, planteamos dos hipótesis: 1) el promotor nativo de *phr-1* tiene una actividad muy baja en la oscuridad; sin embargo, el transcrito se degrada rápidamente en esta condición. O, 2) existe un represor que reprime la transcripción de *phr-1* en la oscuridad, pero solo en su

contexto cromosomal nativo. Si la segunda hipótesis es correcta, entonces los bajos niveles de expresión de *lacZ* en oscuridad pueden atribuirse a un efecto en la organización de la cromatina en el sitio de integración y/o a un efecto de posición en el cromosoma. Aunque las transformantes no se examinaron por Southern blot, el análisis con qPCR mostró que distintas construcciones podían estar integradas aleatoriamente en el genoma de *T. atroviride*. Por otra parte, no se observó una relación directa entre la expresión de *lacZ* en la oscuridad con el número de copias del transgén de resistencia a higromicina (*hph*) que se determinó por qPCR. Estas evidencias nos permitieron sugerir que la integración de casi todas las construcciones ocurrió fuera del contexto de *phr-1* en el cromosoma, lo que llevó a la pérdida de la represión del gen reportero en la oscuridad en casi todas las cepas examinadas.

En este sentido, nuestros datos refuerzan las conclusiones derivadas de un análisis de la transcripción de *phr-1* en una cepa mutante nula  $\Delta phr-1$  y una sobreexpresante de éste gen, donde se sugiere por primera vez la presencia de un represor en la regulación transcripcional de *phr-1* (Berrocal-Tito *et al.*, 2007). La expresión de un gen reportero regulado por 2 kb río arriba del sitio de inicio de la transcripción de *phr-1* fue baja en un fondo mutante, mientras que en una cepa sobreexpresante fue menor la cantidad de energía requerida para llevar a cabo la expresión génica (Berrocal-Tito *et al.*, 2007). Cabe resaltar que no solo la regulación transcripcional de *phr-1* se vio afectada en ambos fondos genéticos, sino que otros genes responsivos a luz también mostraron un patrón similar, lo que llevó a pensar que PHR-1 puede modular la transcripción de otros genes regulados por luz azul a través de la interacción con el BLRC, reclutando un represor al complejo unido sobre promotor en la oscuridad (Berrocal-Tito *et al.*, 2007).

Previamente se demostró que en *N. crassa* un LRE consiste de dos cajas GATA, las cuales son requeridas para la transcripción de genes responsivos a luz, así

como la unión del WCC a dichas secuencias en respuesta a la luz azul (Froehlich *et al.*, 2002; He y Liu 2005). Sin embargo, en estos trabajos no se estudió por separado cada caja GATA que conforma un LRE, ni se analizaron mutaciones puntuales en esas secuencias para probar su función *in vivo*. En este trabajo, mostramos que la región responsiva a luz del promotor de *phr-1* de *T. atroviride* exhibe secuencias similares al LRE descrito en *N. crassa*: la caja GATA1 altamente conservada y dos repeticiones de ésta en la polaridad opuesta. Nuestros resultados sugieren que estas cajas GATA trabajan en conjunto en la regulación de *phr-1*, ya que la eliminación de las cajas GATA repetidas invertidas resultó en una pérdida de la respuesta a luz del gen reportero *lacZ*. Nuestros análisis de mutagénesis en la caja GATA-2 mostraron claramente la relevancia de la región responsiva a luz en dirigir la fotoinducción del gen reportero *lacZ*. Basado en ello, y considerando el análisis de secuencia de los genes responsivos a luz azul en *Trichoderma* (Fig. 5A, publicación), proponemos que el consenso del LRE para este hongo es CGATC. Los cambios en los nucleótidos de la primera y/o quinta posición de este consenso en *T. atroviride* pueden llevar a la pérdida de función. Nuestros resultados, junto con los últimos trabajos descritos sobre los LRE de *N. crassa* (Chen *et al.*, 2009; Smith *et al.*, 2010), sugieren fuertemente que el consenso se encuentra altamente conservado y puede dirigir la transcripción de genes fotoinducibles tanto en *N. crassa* como en *T. atroviride*. Sin embargo, es necesario realizar más estudios de mutagénesis de las cajas GATA individualmente, que son parte del núcleo del LRE, de otros promotores de genes regulados por luz en *T. atroviride* y otros Sordariomycetes para probar tal hipótesis.

Un análisis comparativo de la secuencia promotora de los genes de fotoliasas de cuatro géneros distintos de Sordariomicetos (*Fusarium*, *Glomerella*, *Neurospora* y *Trichoderma*) mostró la conservación de una región de ~90 pb, la cual presenta de cuatro a seis repeticiones imperfectas de la secuencia CGATN. El arreglo de esas repeticiones es variable entre las especies de los hongos examinados, aunque

predominan las cajas CGATN con la misma polaridad que los LRE típicos de los promotores de *Neurospora*. En este sentido, la región responsiva a luz del promotor de *phr-1* de *T. atroviride* presenta una organización atípica, ya que tres de las cinco cajas CGATN tienen una polaridad distinta a los sitios de unión canónicos del WCC (Fig. 4, publicación). Este arreglo atípico de las repeticiones CGATN junto con la evidencia experimental obtenida en este estudio, sugieren fuertemente que los elementos invertidos son funcionalmente equivalentes a las repeticiones directas CGATN, ya que la integridad del segmento del promotor -157 a -184 (el cual contiene dos cajas CGATN invertidas) es esencial para preservar la respuesta a la luz azul, respecto de las versiones truncas o mutadas del promotor de *phr-1* (Fig. 2 y 3, publicación). Por otra parte, existe evidencia de que la proteína PHRB del hongo basidiomiceto *Lentinula edodes*, la cual es homóloga a WC-2/BLR-2, es capaz de unirse a cajas GATAWWC en sentido y antisentido del promotor de un gen fotoregurable (Sano *et al.*, 2009). En este contexto, es importante señalar que todos los homólogos fúngicos conocidos de WC-2 se unen a secuencias GATA, lo que indica que estas proteínas reguladoras han conservado su especificidad de unión a DNA desde el tiempo previo a la divergencia evolutiva de los Ascomicetos y Basidiomicetos.

El análisis de otros promotores de genes responsivos a la luz azul en *T. atroviride* sugiere que las cajas CGATC son importantes para dirigir la expresión en respuesta a este estímulo. Un ejemplo de ello se ilustra en el gen *frq*, el cual presenta dos secuencias LRE: la proximal está conformada por tres cajas CGATC sobrelapadas, mientras que la distal cuenta con una sola caja CGATC (Fig. 5A, publicación).

El promotor de *phr-1* no presenta los elementos putativos EUM1 y EUM2 identificados en *T. reesei*. Sin embargo, ambos motivos han sido identificados en el promotor de algunos otros genes regulados por luz en *T. atroviride*, *T. virens* y *T. reesei* (Rosales-Saavedra *et al.*, 2006). Ninguno de estos elementos está

conservado entre los genes ortólogos, lo cual sugiere que podrían no tener un papel en la fotoinducción de la transcripción en *Trichoderma* spp., tal como se observó para el APE, el cual solo se encontró en el promotor de *phr-1* de *T. atroviride* y no presentó un papel en la regulación de la transcripción en respuesta a la luz en este hongo.

Previamente, el grupo de Schmoll identificó secuencias EUM1 en el promotor de los genes *env1* y *gna3* de *T. reesei* (Schuster *et al.*, 2007), y por ensayos de cambio de movilidad electroforética (EMSA, de sus siglas en inglés) mostraron la unión de un complejo a esta secuencia bajo distintas condiciones de luz (Schuster y Schmoll, 2009). Sin embargo, no se analizaron mutaciones del EUM1 que permitieran verificar la especificidad del complejo por estas secuencias. Por otra parte, detectamos un consenso GATC en todas las sondas empleadas, por lo que los complejos detectados en luz y oscuridad podrían estar interactuando con el GATC en lugar de con el EUM1. Aunque se realizaron ensayos de competencia, el diseño experimental no fue el más apropiado (ya que se emplearon extractos de proteína total y no una fracción nuclear), por lo que se observó la pérdida de varias señales, incluida la del mencionado complejo que se une en respuesta a la luz, y esto hace difícil interpretar la importancia del EUM1 en la regulación transcripcional mediada por luz, por lo que se requiere de ensayos adicionales para probar la funcionalidad de estas secuencias.

Adicionalmente, examinamos la posible interacción de las proteínas BLR-1 y BLR-2 con la secuencia CGATC y el promotor del gen *phr-1*. Tanto en *N. crassa* como en *T. atroviride*, las proteínas WC-2 y BLR-2, respectivamente, no son responsables de la percepción de la luz azul, pero son un componente esencial en la transducción de la señal luminosa (Denault *et al.*, 2001; Casas-Flores *et al.*, 2004). En *T. atroviride*, BLR-2 es el factor limitante en la formación del BLRC (Esquivel-Naranjo y Herrera-Estrella, 2007), mientras que en *N. crassa* se ha mostrado que el dominio de unión a DNA de WC-2, más que el de WC-1, es

esencial para la transcripción de varios genes regulados por luz, como *al-3* y *vvd* (Collet et al., 2002; Cheng et al., 2003). En este sentido, iniciamos el análisis del posicionamiento del BLRC sobre el promotor de *phr-1* por medio de ensayos de inmunoprecipitación de la cromatina (ChIP) utilizando un anticuerpo que reconoce a BLR-2, y encontramos que esta proteína se une principalmente en las dos regiones conservadas del promotor de *phr-1* (LRE y UCR) *in vivo*. La cinética de unión sobre ambas regiones en respuesta a la luz sugiere que BLR-2, y probablemente BLR-1, están unidos al promotor en la oscuridad, asociados con un represor transcripcional putativo y tras un estímulo de luz, se libera el complejo BLRC activo, lo que da lugar a la expresión de *phr-1* inmediatamente después que se dio el pulso de luz (Fig. 6, publicación).

La posible interacción de BLR-1 y BLR-2 en ambas regiones conservadas se analizó *in vitro* mediante una serie de EMSAs, utilizando extractos nucleares de *T. atroviride* de las cepas silvestre y las mutantes  $\Delta blr-1$  y  $\Delta blr-2$  obtenidos bajo condiciones de luz y oscuridad. Tanto la sonda PLRRwt como la UCR mostraron un retardo electroforético cuando se incubaron con los extractos de oscuridad de la cepa silvestre, y se observó la formación de un segundo complejo de menor peso molecular en la condición de luz, lo que sugiere una pérdida de subunidades y soporta nuestros resultados obtenidos mediante ChIPs. De modo interesante, la cinética de los complejos formados en *T. atroviride* contrasta con la reportada para *N. crassa*, donde se ha descrito un complejo pequeño en la oscuridad, formado por una proteína WC-1 y una proteína WC-2, y la formación de un complejo grande debido al reclutamiento de una segunda proteína WC-1 en respuesta a la luz (complejo en luz) (Froehlich et al., 2002; Cheng et al., 2003; He y Liu, 2005). Nosotros observamos un complejo grande en la oscuridad y la aparición de uno más pequeño en luz, lo cual sugiere diferencias en los componentes del complejo fotorreceptor y por tanto, mecanismos distintos. En este sentido, el análisis con las mutantes  $\Delta blr-1$  y  $\Delta blr-2$  permitió una mayor comprensión de la dinámica que ocurre sobre el promotor de *phr-1* en *T. atroviride*. Tomando estos resultados junto

con los del ChIP, nuestros datos sugieren la liberación de BLR-1 en respuesta a la luz y la presencia de una pequeña proporción de BLR-2 con respecto al estado de oscuridad. Sin embargo, para conocer la naturaleza de los diferentes complejos bajo distintas condiciones de luz es necesario complementar nuestros estudios con otros experimentos.

La cinética del enriquecimiento de BLR-2 *in vivo* sobre las regiones conservadas del promotor de *phr-1* en respuesta a la luz, así como los ensayos con las cepas  $\Delta blr-1$  y  $\Delta blr-2$ , nos permitieron proponer una hipótesis alternativa sobre el papel de las proteínas BLR en la fotoinducción, donde BLR-1 y BLR-2 están unidas al promotor en oscuridad, funcionando como represores transcripcionales en vez de activadores (como se pensó inicialmente por su similitud con las proteínas WC), lo cual explicaría la ausencia de un dominio de activación en ambas proteínas BLR (Casas-Flores *et al.*, 2004), y la detección del transcrito de *phr-1* inmediatamente después de un pulso de luz (Berrocal-Tito *et al.*, 1999). Alternativamente, podría ser posible que solo una de las proteínas BLR funcione como represora, regulando la función de la otra, ya que se ha visto un patrón de proteínas diferencial entre las cepas  $\Delta blr-1$  y  $\Delta blr-2$  en condiciones de luz y oscuridad, lo cual apoya nuestra hipótesis como resultado de los análisis tipo EMSA (Sánchez-Arreguín *et al.*, 2012; éste trabajo).

Tomando en conjunto nuestros resultados y lo reportado por otros grupos, al parecer las proteínas BLR trabajan en conjunto como un complejo, BLRC, regulando directamente a los genes *blu* y *bld* de manera dependiente de la luz, uniéndose a las regiones conservadas del promotor, donde las cajas CGATC son los elementos importantes que dirigen la transcripción de estos genes.

En este sentido, la región responsiva a luz que identificamos sobre el promotor de *phr-1* en las cinco especies de *Trichoderma* presenta elementos similares al ELRE reportado en *N. crassa* (He y Liu, 2005; Chen *et al.*, 2009; Smith *et al.*, 2010), pero

las sutiles diferencias entre las distintas especies de hongos podrían explicar las diferencias en las vías de transducción de la señal que se ha descrito en los modelos *N. crassa* y *A. nidulans* (Rodríguez-Romero *et al.*, 2010; Chen *et al.*, 2010b). En cuanto a la UCR, no encontramos una función de esta secuencia asociada con la fotoinducción de *phr-1*, al menos bajo las condiciones que probamos; sin embargo, no descartamos que esta secuencia pueda tener un papel en otros procesos y otras condiciones. De modo interesante, sí observamos unión de BLR-2 a esta región en la oscuridad, donde identificamos una caja GATA semejante a las de la región responsiva a luz, la cual podría ser el sitio de unión de BLR-2.

Finalmente, se inició el trabajo para analizar la regulación transcripcional de un gen reprimido por la luz azul, *bld-2*, el cual codifica para una deshidrogenasa con diferentes especificidades (Rosales-Saavedra *et al.*, 2006). El perfil de expresión es contrario al de *phr-1*, ya que se expresa en condiciones de oscuridad y el nivel de su transcrito disminuye progresivamente después del pulso de luz azul. La expresión de este gen también es mediada por las proteínas BLR, ya que en las mutantes nulas de cualquiera de estos factores de transcripción presentan niveles del transcrito tal como si estuviera en oscuridad, aun cuando se haya estimulado la colonia con luz azul (Rosales-Saavedra *et al.*, 2006).

Para analizar el promotor del gen *bld-2*, identificamos su sitio de inicio de la transcripción (TSS) mediante la técnica de 5'-RACE. Aún cuando el consenso propuesto para un TSS en eucariontes es muy degenerado (5'-Y-Y-A<sup>+1</sup>-N-T/A-Y-Y-Y-3') (Lodish *et al.*, 2000), la posición +1 del TSS de *bld-2* determinado experimentalmente es una adenina, y algunos nucleótidos aledaños ajustan con el consenso. El análisis de la región 5' muestra una posible caja TATA localizada a -36 pb. También identificamos varias cajas GATA, seis en la cadena sentido y once en la cadena antisentido, y un consenso como el reportado por He y Liu (2005): (-103) GATGC-N<sub>37</sub>-CGATG (-57).



El gen *bld-2* se encuentra organizado en el genoma en una configuración divergente, de la misma manera que *phr-1*, con una región intergénica de 1230 pb. El análisis del gen río arriba de *bld-2* por medio de BLAST no arrojó datos que nos permitieran asignarle una función y solo podemos calificarlo como una proteína putativa. Para identificar las secuencias que regulan la expresión de *bld-2* en respuesta a la luz, se clonaron ocho versiones del promotor al gen reportero *lacZ*, que corresponden a deleciones seriadas del extremo 5'.

A la fecha, este es el primer trabajo donde se analizan e identifican elementos de respuesta a luz en genes responsivos a luz en *Trichoderma*. Sin embargo, aún queda mucho trabajo por realizar para elucidar los mecanismos que regulan la transcripción de genes responsivos a luz azul en *Trichoderma*.

## Materiales y Métodos

### *Ensayos de cambio de movilidad electroforética*

Se realizaron una serie de ensayos de cambio de movilidad electroforética (EMSA) con extractos de proteína nuclear de las cepas silvestre y mutantes  $\Delta blr-1$  y  $\Delta blr-2$  de *T. atroviride*, junto con sondas de las regiones PLRR y UCR del promotor de *phr-1* (Tabla 3). Adicionalmente, se diseñaron sondas con mutaciones en cada una de las cajas GATA de la región responsiva a luz del promotor de *phr-1*

Tabla 3. Sondas diseñadas para los EMSA en la PLRR y UCR del promotor de *phr-1*

Nombre	Secuencia (5'- 3') <sup>a, b</sup>	Descripción
PLRRwt	GCTACTTACGCAGG <b>GATAC</b> CAAGCCACTGAAAT CGCACTGACATCGCACCAGCCAAAGTCC <b>CGATC</b> TCCAGACGCCAATGACGCAGGCCACATCGAGGC CATAT	PLRR con las dos cajas GATA intactas
mutGATA-1	GCTACTTACGCAGG <b>GATAC</b> CAAGCCACTGAAAT CGCACTGACATCGCACCAGCCAAAGTCC <b>GTCCT</b> TCCAGACGCCAATGACGCAGGCCACATCGAGGC CATAT	PLRR con la caja GATA-1 mutada
mutGATA-2	GCTACTTACGCAGG <b>ACCGC</b> CAAGCCACTGAAAT CGCACTGACATCGCACCAGCCAAAGTCC <b>CGATC</b> TCCAGACGCCAATGACGCAGGCCACATCGAGGC CATAT	PLRR con la caja GATA-2 mutada
UCR	CTCAGGCT <b>GATTC</b> ACATCCAAAACCTGGCTCGGCA AGGCGCTCGTGACGACACGGCATGTCTCTGCA AAAGCCGAACACTAGTAGCGTTGGAACCGGTAT TGTTCTGGAAATGAGGCAGAACG	Región conservada río arriba (UCR)

<sup>a</sup> Las cajas GATA silvestres están resaltadas en negro y en gris

<sup>b</sup> Las mutaciones en las cajas GATA de la PLRR están subrayadas

Los EMSA se realizaron como se ha descrito previamente (Froehlich *et al.*, 2002; He y Liu, 2005) con extractos de proteínas nucleares que se mantuvieron en oscuridad o a las que se expuso a luz azul durante la reacción de unión. Para ello, se utilizaron 2  $\mu$ g de proteína nuclear, 2 nM de sonda marcada con biotina en el extremo -3' (Thermo Scientific), 50 ng/ $\mu$ l Poly dI•dC, buffer de unión buffer [20 mM

HEPES pH 7.9, 1 mM EDTA, 2 mM MgCl<sub>2</sub>, 10%(v/v) glicerol, 20 μM ZnCl<sub>2</sub>]. Las reacciones de unión se visualizaron con el kit LightShift Chemiluminescent EMSA (Thermo Scientific), siguiendo el protocolo del proveedor.

### **Identificación del sitio de inicio de la transcripción del gen *bld-2***

Se identificó el sitio de inicio de la transcripción (TSS, por sus siglas en inglés) del gen *bld-2* (Protein ID: 301399; <http://genome.jgi-psf.org/Triat2/Triat2.home.html>) mediante la técnica 5'-RLM-RACE con el kit comercial FirstChoice 5' RLM-RACE (Ambion) siguiendo las instrucciones del proveedor para una reacción estándar. La reacción se llevó a cabo con 10μg de RNA total proveniente de una muestra de *T. atroviride* incubado en oscuridad. La secuencia del adaptador 5', así como los oligonucleótidos empleados se muestran en la Tabla 4.

Tabla 4. Secuencia del adaptador y los oligonucleótidos usados en el 5'RLM-RACE

<b>Componente</b>	<b>Secuencia</b>
Adaptador 5'RACE	5'-GCUGAUGGCGAUGAAUGAACACUGCGUUUGCUGGCUUUGAUGAAA-3'
5' RACE Outer Primer	5'-GCTGATGGCGATGAATGAACACTG-3'
5' RACE Inner Primer	5'-CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG-3'
<i>bld-2</i> TA Fwd	5'-CGTCGTCTTTGGCGGCTCAGG-3'
<i>bld-2</i> TA Rev	5'-AACGGGTACACGGCACAGTCC-3'
<i>bld-2</i> tr Rev	5'-AGTTGGCGCACCATCATACTGACA-3'

Para la reacción *outer*, se utilizó 1 μl del cDNA y se usó el par de oligonucleótidos 5' RACE Outer (O)/*bld-2* tr Rev con el siguiente programa de PCR: 94°C 3 min; 35 ciclos de 94°C (30 seg), 60°C (30 seg), 72°C (30 seg); extensión final 72°C 7 min. Para la reacción *inner*, se tomó como DNA molde 2 μl de la reacción *outer* y se usó el par de oligonucleótidos 5' RACE Inner/*bld-2* TA Rev con el mismo programa de PCR que la reacción *outer*.

Se tomaron 10  $\mu$ l de cada reacción y se corrieron en un gel de agarosa al 1%. De la reacción *outer* se obtuvo un amplicón de 700 pb aproximadamente, la cual se enriqueció en la reacción *inner*.

Se clonó 1  $\mu$ l de la reacción *inner* en el vector pGEM-T Easy (Promega) y se secuenciaron cuatro clonas por el método de Sanger. Se identificó el adaptador 5' RACE en la secuencia obtenida y la siguiente base se tomó como el TSS del gen *bld-2*

### ***Diseño y construcción de fusiones traduccionales de distintas versiones del promotor de bld-2 con el gen reportero lacZ***

Se diseñaron oligonucleótidos para amplificar fragmentos a varias distancias del TSS de *bld-2* (Tabla 5), para realizar fusiones traduccionales con el gen reportero *lacZ*. Los oligonucleótidos *forward* incluyen el sitio de restricción de *Spe* I y el universal reverso lleva el sitio *Sma* I, a fin de clonar en los mismos sitios del vector pCBlacZ.

Tabla 5. Oligonucleótidos usados para fusiones traduccionales de *bld-2* con el gen *lacZ*

<b>Nombre</b>	<b>Secuencia (5'→3')<sup>a</sup></b>
Pbld2-1794 <i>Spe</i> I F	5'- <b>ACTAGT</b> TTAGCTGAACCGGCG-3'
Pbld2-1210 <i>Spe</i> I F	5'- <b>ACTAGT</b> GCCTCTAGAGCTCTCTGGTCAA-3'
Pbld2-738 <i>Spe</i> I F	5'- <b>ACTAGT</b> TGCGCCAACTCTAACAGGGATT-3'
Pbld2-535 <i>Spe</i> I F	5'- <b>ACTAGT</b> GCTAGGAAAAGGTCAGGAG-3'
Pbld2-432 <i>Spe</i> I F	5'- <b>ACTAGT</b> GGGACGCTTCATCATTGCTA-3'
Pbld2-125 <i>Spe</i> I F	5'- <b>ACTAGT</b> CTGCAAGATGTCTAGTCTAC-3'
Pbld2-113 <i>Spe</i> I F	5'- <b>ACTAGT</b> CTACGAGATGCAAGTTGTCA-3'
Pbld2-89 <i>Spe</i> I F	5'- <b>ACTAGT</b> GGAGCCATCTAGGTCCATA-3'
UPbld-2 <i>Sma</i> I R	5'- <b>CCCGGG</b> GACGCCAAAAGCCATTGTGT -3'

<sup>a</sup> Los sitios de restricción están indicados en negritas

Todas las versiones del promotor de *bld-2* se amplificaron por PCR, utilizando DNA genómico de *T. atroviride* como molde. Las condiciones de PCR fueron 5 min a 94°C, 25 ciclos of 94°C (30 seg), 58°C (30 seg), 72°C (30 seg para amplicones < 500 pb, y 60 seg para amplicones > 500 pb), con una extensión final de 72°C por 5 min. Los productos de PCR se clonaron en pGEM-T Easy y se secuenciaron.

Los plásmidos pGEM y pCBlacZ se digirieron con *Sma* I y *Spe* I para realizar las fusiones del promotor de *bld-2* con el gen *lacZ*. Las ligaciones se transformaron en la cepa *E. coli* TOP10F' y se realizó un escrutinio de las construcciones correctas utilizando enzimas de restricción

## Referencias

- Asayama M. (2006) "Regulatory system for light-responsive gene expression in photosynthesizing bacteria: *cis*-elements and *trans*-acting factors in transcription and post-transcription" *Biosci Biotechnol Biochem* **70**:565-573
- Atoui A, Kastner C, Larey CM, Thokala R, Etxebeste O, Espeso EA, Fischer R, Calvo AM. (2010) "Cross-talk between light and glucose regulation controls toxin production and morphogenesis in *Aspergillus nidulans*" *Fungal Genet Biol* **47**:962-972
- Bahn YS, Xue C, Idnurm A, Rutherford JC, Heitman J, Cardenas ME. (2007) "Sensing the environment: lessons from fungi" *Nat Rev Microbiol* **5**(1):57-69
- Ballario P, Macino G. (1997) "White collar proteins: PASSing the light signal in *Neurospora crassa*" *Trends Microbiol* **5**:458-462
- Bayram Ö, Biesemann C, Krappmann S, Galland P, Braus GH. (2008) "More than a repair enzyme: *Aspergillus nidulans* Photolyase-like CryA is a regulator of sexual development" *Mol Biol Cell* **19**:3254-3262
- Berrocal-Tito G, Sametz-Baron L, Eichenberg K, Horwitz BA, Herrera-Estrella A. (1999) "Rapid blue light regulation of a *Trichoderma harzianum* photolyase gene" *J Biol Chem* **274**:14288-14294
- Berrocal-Tito GM, Rosales-Saavedra T, Herrera-Estrella A, Horwitz BA. (2000) "Characterization of blue-light and developmental regulation of the photolyase gene *phr-1* in *Trichoderma harzianum*" *Photochem Photobiol* **71**:662-668
- Berrocal-Tito GM, Esquivel-Naranjo EU, Horwitz BA, Herrera-Estrella A. (2007) "*Trichoderma atroviride* PHR1, a fungal photolyase responsible for DNA repair, autoregulates its own photoinduction" *Eukaryot Cell* **6**:1682-1692
- Berrocal-Tito GM, Rosales-Saavedra T, Herrera-Estrella A, Horwitz BA. (2000) "Characterization of blue-light and developmental regulation of the photolyase gene *phr1* in *Trichoderma harzianum*" *Photochem Photobiol* **71**:662-668
- Betina V, Koman V. (1980) "Changes in the lipid composition during the photo-induced conidiation of *Trichoderma viride*" *Folia Microbiol* **25**:295-300
- Biernat MA, Eker AP, van Oers MM, Vlak JM, van der Horst GT, Chaves I. (2012) "A baculovirus photolyase with DNA repair activity and circadian clock regulatory function" *J Biol Rhythms* **27**:3-11

- Bieszke JA, Braun EL, Bean LE, Kang S, Natvig DO, Borkovich KA. (1999) "The nop-1 gene of *Neurospora crassa* encodes a seven transmembrane helix retinal-binding protein homologous to archaeal rhodopsins" *Proc Natl Acad Sci USA* **96**:8034-8039
- Bluhm BH, Dunkle LD. (2008) "PHL1 of *Cercospora zea-maydis* encodes a member of the photolyase/cryptochrome family involved in UV protection and fungal development" *Fungal Genet Biol* **45**:1362-1372
- Blumenstein A, Vienken K, Tasler R, Purschwitz J, Veith D, Frankenberg-Dinkel N, Fischer R. (2005) "The *Aspergillus nidulans* phytochrome FphA represses sexual development in red light" *Curr Biol* **15**:1833–1838
- Borkovich KA, Alex LA, Yarden O, Freitag M, Turner GE, Read ND, Seiler S, Bell-Pedersen D, Paietta J, Plesofsky N, Plamann M, Goodrich-Tanrikulu M, Schulte U, Mannhaupt G, Nargang FE, Radford A, Selitrennikoff C, Galagan JE, Dunlap JC, Loros JJ, Catchside D, Inoue H, Aramayo R, Polymenis M, Selker EU, Sachs MS, Marzluf GA, Paulsen I, Davis R, Ebbole DJ, Zelter A, Kalkman ER, O'Rourke R, Bowring F, Yeadon J, Ishii C, Suzuki K, Sakai W, Pratt R. (2004) "Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism" *Microbiol Mol Biol Rev* **68**:1-108
- Casas-Flores S, Rios-Momberg M, Bibbins M, Ponce-Noyola P, Herrera-Estrella A. (2004) "BLR-1 and BLR-2, key regulatory elements of photoconidiation and mycelial growth in *Trichoderma atroviride*" *Microbiology* **150**:3561-3569
- Casas-Flores S, Rios-Momberg M, Rosales-Saavedra T, Martinez-Hernandez P, Olmedo-Monfil V, Herrera-Estrella A. (2006) "Cross talk between a fungal blue light perception system and the cyclic AMP signaling pathway" *Eukaryot Cell* **5**:499–506
- Castellanos F, Schmoll M, Martínez P, Tisch D, Kubicek CP, Herrera-Estrella A, Esquivel-Naranjo EU. (2010) "Crucial factors of the light perception machinery and their impact on growth and cellulase gene transcription in *Trichoderma reesei*" *Fungal Genet Biol* **47**:468-476
- Carreras-Villaseñor N, Sánchez-Arreguín JA, Herrera-Estrella AH. (2012) "*Trichoderma*: sensing the environment for survival and dispersal" *Microbiology* **158**:3-16
- Chen CH, Ringelberg CS, Gross RH, Dunlap JC, Loros JJ. (2009) "Genome-wide analysis of light-inducible responses reveals hierarchical light signalling in *Neurospora*" *EMBO J* **28**:1029-1042
- Chen CH, DeMay BS, Gladfelter AS, Dunlap JC, Loros JJ (2010a) "Physical interaction between VIVID and white collar complex regulates photoadaptation in *Neurospora*" *Proc Natl Acad Sci U S A* **107**:16715-16720

- Chen CH, Dunlap JC, Loros JJ. (2010b) "*Neurospora* illuminates fungal photoreception" *Fungal Genet Biol* **47**:922-929
- Chen M, Chory J, Fankhauser C. (2004) "Light signal transduction in higher plants" *Annu Rev Genet* **38**:87-117
- Cheng P, Yang Y, Liu Y. (2001) "Interlocked feedback loops contribute to the robustness of the *Neurospora* circadian clock" *Proc Natl Acad Sci USA* **98**:7408–7413
- Cheng P, Yang Y, Gardner KH, Liu Y. (2002) "PAS domain-mediated WC-1/WC-2 interaction is essential for maintaining the steady-state level of WC-1 and the function of both proteins in circadian clock and light responses of *Neurospora*" *Mol Cell Biol* **22**:517-524
- Cheng P, Yang Y, Wang L, He Q, Liu Y. (2003) "WHITE COLLAR-1, a multifunctional *Neurospora* protein involved in the circadian feedback loops, light sensing, and transcription repression of *wc-2*" *J Biol Chem* **278**: 3801-3808
- Cheng P, He Q, He Q, Wang L, Liu Y. (2005) "Regulation of the *Neurospora* circadian clock by an RNA helicase" *Genes Dev* **19**:234-241
- Coesel S, Mangogna M, Ishikawa T, Heijde M, Rogato A, Finazzi G, Todo T, Bowler C, Falciatore A. (2009) "Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity" *EMBO Rep* **10**:655-661
- Collett, MA, Garceau N, Dunlap JC, Loros JJ. (2002) "Light and clock expression of the *Neurospora* clock gene *frequency* is differentially driven by but dependent on WHITE COLLAR-2" *Genetics* **160**:149-158
- Corchnoy SB, Swartz TE, Lewis JW, Szundi I, Briggs WR, Bogomolni RA. (2003) "Intramolecular proton transfers and structural changes during the photocycle of the LOV2 domain of Phototropin 1" *J Biol Chem* **278**:724–731
- Corrochano LM. (2007) "Fungal photoreceptors: sensory molecules for fungal development and behaviour" *Photochem Photobiol Sci* **6**:725-736
- Corrochano LM. (2011) "Fungal photobiology: a synopsis" *IMA Fungus* **2**:25-28.
- Crosson S, Rajagopal S, Moffat K. (2003) "The LOV domain family: photoresponsive signaling modules coupled to diverse output domains" *Biochemistry* **42**:2-10
- Crosthwaite SK, Dunlap JC, Loros JJ. (1997) "*Neurospora wc-1* and *wc-2*: transcription, photoresponses, and the origins of circadian rhythmicity." *Science* **276**:763-769



- Denault DL, Loros JJ, Dunlap JC. (2001) "WC-2 mediates WC-1-FRQ interaction within the PAS protein-linked circadian feedback loop of *Neurospora*" *EMBO J* **20**:109-117
- Falciatore A, Bowler C. (2005) "The evolution and function of blue and red light photoreceptors" *Curr Top Dev Biol* **68**:317-350
- Fankhauser, C., Staiger D. (2002) "Photoreceptors in *Arabidopsis thaliana*: light perception, signal transduction and entrainment of the endogenous clock" *Planta* **216**:1-16
- Filichkin SA, Breton G, Priest HD, Dharmawardhana P, Jaiswal P, Fox SE, Michael TP, Chory J, Kay SA, Mockler TC. (2011) "Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules" *PLoS One* **6**: e16907
- Franchi L, Fulci V, Macino G. (2005) "Protein kinase C modulates light responses in *Neurospora* by regulating the blue light photoreceptor WC-1" *Mol Microbiol* **56**:334-345
- Friedl MA, Kubicek CP, Druzhinina IS. (2008a) "Carbon source dependence and photostimulation of conidiation in *Hypocrea atroviridis*" *Appl Environ Microbiol* **74**:245-250
- Friedl MA, Schmoll M, Kubicek CP, Druzhinina IS. (2008b) "Photostimulation of *Hypocrea atroviridis* growth occurs due to a cross-talk of carbon metabolism, blue light receptors and response to oxidative stress" *Microbiology* **154**:1229-1241
- Froehlich AC, Liu Y, Loros JJ, Dunlap JC. (2002) "White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter" *Science* **297**:815-819
- Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, Staben C, Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N, Barrett R, Gnerre S, Kamal M, Kamvysselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S, Rasmussen C, Metzenberg RL, Perkins DD, Kroken S, Cogoni C, Macino G, Catcheside D, Li W, Pratt RJ, Osmani SA, DeSouza CP, Glass L, Orbach MJ, Berglund JA, Voelker R, Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig DO, Alex LA, Mannhaupt G, Ebbole DJ, Freitag M, Paulsen I, Sachs MS, Lander ES, Nusbaum C, Birren B. (2003) "The genome sequence of the filamentous fungus *Neurospora crassa*" *Nature* **422**:859-868

- Gallego M, Virshup DM. (2007) "Post-translational modifications regulate the ticking of the circadian clock" *Nat Rev Mol Cell Biol* **8**:139-148
- Galun E, Gressel J. (1966) "Morphogenesis in *Trichoderma*: Suppression of photoinduction by 5-fluorouracil" *Science* **151**:696-698
- Gomelsky M, Klug G. (2002) "BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms" *Trends Biochem Sci* **27**:497-500
- Goyal A, Szarzynska B, Fankhauser C. (2013) "Phototropism: at the crossroads of light-signaling pathways" *Trends Plant Sci* doi: 10.1016/j.tplants.2013.03.002
- Gremel G, Dorrer M, Schmoll M, (2008) "Sulphur metabolism and cellulase gene expression are connected processes in the filamentous fungus *Hypocrea jecorina* (anamorph *Trichoderma reesei*)" *BMC Microbiol.* **8**:174
- Grešik M, Kolarova N, Farkaš V. (1988) "Membrane potential, ATP, and cyclic AMP changes induced by light in *Trichoderma viride*" *Exp Mycol* **12**:295-301
- Grešik M, Kolarova N, Farkaš V. (1989) "Light-stimulated phosphorylation of proteins in cell-free extracts from *Trichoderma viride*" *FEBS Lett.* **248**:185- 187
- Gressel J, Galun E (1967) "Morphogenesis in *Trichoderma*: photoinduction and RNA" *Dev Biol* **15**:575-598
- Gressel J, Hartmann KM. (1968) "Morphogenesis in *Trichoderma*: Action spectrum of photoinduced sporulation" *Planta (Berl.)* **79**:271-274
- Harmer SL, Kay SA. (2005) "Positive and negative factors confer phase-specific circadian regulation of transcription in *Arabidopsis*" *Plant Cell* **17**:1926-40
- He Q, Cheng P, Yang Y, Wang L, Gardner KH, Liu Y. (2002) "White collar-1, a DNA binding transcription factor and a light sensor" *Science* **297**:840–843
- He Q, Liu Y. (2005) "Molecular mechanism of light responses in *Neurospora*: from light-induced transcription to photoadaptation" *Genes Dev* **19**:2888-2899
- Heintzen C, Loros JJ, Dunlap JC. (2001) "The PAS protein VIVID defines a clock-associated feedback loop that represses light input, modulates gating, and regulates clock resetting" *Cell* **104**:453–464.
- Herrera-Estrella A, Horwitz BA. (2007) "Looking through the eyes of fungi: molecular genetics of photoreception" *Mol Microbiol* **64**:5-15

- Horwitz BA, Malkin S, Gressel J. (1984a) "Roseoflavin inhibition of photoconidiation in a *Trichoderma* riboflavin auxotroph: indirect evidence for flavin requirement for photoreactions" *Photochem Photobiol* **40**:763–769
- Horwitz BA, Weisenseel MH, Dorn A, Gressel J. (1984b) "Electric currents around growing *Trichoderma* hyphae, before and after photoinduction of conidiation" *Plant Physiol* **74**:912-916
- Horwitz BA, Gressel J, Malkin S, Epel BL. (1985). "Modified cryptochrome *in vivo* absorption in *dim* photosporulation mutants of *Trichoderma*" *Proc Natl Acad Sci USA* **82**:2736-2740
- Idnurm A, Howlett BJ. (2001) "Characterization of an opsin gene from the ascomycete *Leptosphaeria maculans*" *Genome* **44**:167–171
- Idnurm A, Heitman J. (2005) "Light controls growth and development via a conserved pathway in the fungal kingdom" *PLoS Biol* **3**:e95
- Iseki M, Matsunaga S, Murakami A, Ohno K, Shiga K, Yoshida K, Sugai M, Takahashi T, Hori T, Watanabe MA. (2002) "Blue-light-activated adenylyl cyclase mediates photoavoidance in *Euglena gracilis*" *Nature* **415**:1047-1051
- Kanai S, Kikuno R, Toh H, Ryo H, Todo T. (1997) "Molecular evolution of the photolyase blue-light photoreceptor family" *J Mol Evol* **45**:535-548
- Kritskii MS, Belozerskaia TA, Sokolovskii V, Filippovich S. (2005) "Photoreceptor apparatus of a fungus *Neurospora crassa*" *Mol Biol* **39**:602-617
- Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, Mukherjee M, Kredics L, Alcaraz LD, Aerts A, Antal Z, Atanasova L, Cervantes-Badillo MG, Challacombe J, Chertkov O, McCluskey K, Couplier F, Deshpande N, von Döhren H, Ebbole DJ, Esquivel-Naranjo EU, Fekete E, Flippi M, Glaser F, Gómez-Rodríguez EY, Gruber S, Han C, Henrissat B, Hermosa R, Hernández-Oñate M, Karaffa L, Kosti I, Le Crom S, Lindquist E, Lucas S, Lübeck M, Lübeck PS, Margeot A, Metz B, Misra M, Nevalainen H, Omann M, Packer N, Perrone G, Uresti-Rivera EE, Salamov A, Schmoll M, Seiboth B, Shapiro H, Sukno S, Tamayo-Ramos JA, Tisch D, Wiest A, Wilkinson HH, Zhang M, Coutinho PM, Kenerley CM, Monte E, Baker SE, Grigoriev IV. (2011) "Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*" *Genome Biol* **12**(4):R40
- Kumagai T, Oda Y. (1969) "An action spectrum for photoinduced sporulation in the fungus *Trichoderma viride*" *Plant Cell Physiol* **10**:387-392
- Linden H, Ballario P, Macino G. (1997) "Blue light regulation in *Neurospora crassa*" *Fungal Genet Biol* **22**:141-150

- Linden H, Macino G. (1997) "White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in *Neurospora crassa*" *EMBO J* **16**:98-109
- Linden H. (2002) "Circadian rhythms. A white collar protein senses blue light" *Science* **297**:777-778
- Liu Y, Bell-Pedersen D. (2006) "Circadian rhythms in *Neurospora crassa* and other filamentous fungi" *Eukaryot Cell* **5**:1184–1193
- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. (2000) "Regulatory Sequences in Eukaryotic Protein-Coding Genes", en *Molecular Cell Biology* 4th edition. New York: W. H. Freeman. Section 10.4
- Lucas-Lledó JI, Lynch M. (2009) "Evolution of mutation rates: Phylogenomic analysis of the Photolyase/Cryptochrome family" *Mol Biol Evol* **26**:1143-1153
- Malzahn E, Ciprianidis S, Káldi K, Schafmeier T, Brunner M. (2010) "Photoadaptation in *Neurospora* by competitive interaction of activating and inhibitory LOV domains" *Cell* **142**:762-772
- Masuda S, Bauer CE. (2002) "AppA is a blue light photoreceptor that antirepresses photosynthesis gene expression in *Rhodobacter sphaeroides*" *Cell* **110**:613-623
- Mazzella MA, Cerdan PD, Staneloni RJ, Casal JJ. (2001) "Hierarchical coupling of phytochromes and cryptochromes reconciles stability and light modulation of *Arabidopsis* development" *Development* **128**:2291–2299
- Mockler TC, Guo H, Yang H, Duong H, Lin C. (1999) "Antagonistic actions of *Arabidopsis* cryptochromes and phytochrome B in the regulation of floral induction" *Development* **126**:2073–2082
- Möglich A, Yang X, Ayers RA, Moffat K. (2010) "Structure and function of plant photoreceptors" *Annu Rev Plant Biol* **61**:21-47
- Montgomery BL. (2007) "Sensing the light: photoreceptive systems and signal transduction in cyanobacteria" *Mol Microbiol* **64**:16-27
- Muñoz E, Baler R. (2003) "The circadian E-box: when perfect is not good enough" *Chronobiol Int* **30**:371-388
- Nagy F, Schafer E. (2002) "Phytochromes control photomorphogenesis by differentially regulated, interacting signaling pathways in higher plants" *Annu Rev Plant Biol* **53**:329–355

- Ogura Y, Tokutomi S, Wada M, Kiyosue T. (2008) "PAS/LOV proteins" *Plant Signal Behav* **3**: 966–968
- Okajima K, Yoshihara S, Fukushima Y, Geng X, Katayama M, Higashi S, Watanabe M, Sato S, Tabata S, Shibata Y, Itoh S, Ikeuchi M. (2005) "Biochemical and functional characterization of BLUF-type flavin-binding proteins of two species of cyanobacteria" *J Biochem* **137**:741–750
- Palczewski K. (2006) "G protein-coupled receptor rhodopsin" *Annu Rev Biochem* **75**: 743–767
- Papavizas GC. (1985). "*Trichoderma* and *Gliocladium*: biology, ecology, and potential for biocontrol" *Annu Rev Phytopathol* **23**:23–54.
- Prado MM, Prado-Cabrero A, Fernandez-Martin R, Avalos J. (2004) "A gene of the opsin family in the carotenoid gene cluster of *Fusarium fujikuroi*" *Curr Genet* **46**:47–58
- Purschwitz J, Müller S, Kastner C, Fischer R. (2006) "Seeing the rainbow: light sensing in fungi" *Curr Opin Microbiol* **9**:566-571
- Rocha-Ramirez V, Omero C, Chet I, Horwitz BA, Herrera-Estrella A. (2002) "*Trichoderma atroviride* G-protein alpha-subunit gene *tga1* is involved in mycoparasitic coiling and conidiation" *Eukaryot. Cell* **1**:594–605
- Rockwell NC, Su YS, Lagarias JC. (2006) "Phytochrome structure and signaling mechanisms" *Annu Rev Plant Biol* **57**:837-858
- Rodriguez-Romero J, Hedtke M, Kastner C, Müller S, Fischer R. (2010) "Fungi, hidden in soil or up in the air: light makes a difference" *Annu Rev Microbiol* **64**:585-610
- Rosales-Saavedra T, Esquivel-Naranjo EU, Casas-Flores S, Martínez-Hernández P, Ibarra-Laclette E, Cortes-Penagos C, Herrera-Estrella A. (2006) "Novel light-regulated genes in *Trichoderma atroviride*: a dissection by cDNA microarrays" *Microbiology* **152**:3305-3317
- Salom D, Lodowski DT, Stenkamp RE, Le Trong I, Golczak M, Jastrzebska B, Harris T, Ballesteros JA, Palczewski K. (2006) "Crystal structure of a photoactivated deprotonated intermediate of rhodopsin" *Proc Natl Acad Sci USA* **103**:16123-16128
- Sancar A. (2003) "Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors" *Chem Rev* **103**:2203-2237
- Sancar A. (2004) "Regulation of the mammalian circadian clock by cryptochrome" *J Biol Chem* **279**:34079–34082

- Sánchez-Arreguín A, Pérez-Martínez AS, Herrera-Estrella A. (2012) "Proteomic analysis of *Trichoderma atroviride* reveals independent roles for transcription factors BLR-1 and BLR-2 in light and darkness" *Eukaryot Cell* **11**:30-41
- Sano H, Kaneko S, Sakamoto Y, Sato T, Shishido K (2009) "The basidiomycetous mushroom *Lentinula edodes* white collar-2 homolog PHRB, a partner of putative blue-light photoreceptor PHRA, binds to a specific site in the promoter region of the *L. edodes* tyrosinase gene" *Fungal Genet Biol* **46**:333-341
- Saranak, J, Foster KW. (1997) "Rhodopsin guides fungal phototaxis" *Nature* **387**:465–466
- Schafmeier T, Haase A, Káldi K, Scholz J, Fuchs M, Brunner M. (2005). "Transcriptional feedback of *Neurospora* circadian clock gene by phosphorylation-dependent inactivation of its transcription factor" *Cell* **122**:235–246
- Schmoll M, Franchi L, Kubicek CP. (2005) "Envoy, a PAS/LOV domain protein of *Hypocrea jecorina* (Anamorph *Trichoderma reesei*), modulates cellulase gene transcription in response to light" *Eukaryot Cell* **4**:1998-2007
- Schmoll M, Schuster A, do Nascimento Silva R, Kubicek CP. (2009) "The G-alpha protein GNA3 of *Hypocrea jecorina* (anamorph *Trichoderma reesei*) regulates cellulase gene expression in the presence of light" *Eukaryot. Cell* **8**:410–420
- Schuster A, Kubicek CP, Friedl MA, Druzhinina IS, Schmoll M. (2007) "Impact of light on *Hypocrea jecorina* and the multiple cellular roles of ENVOY in this process" *BMC Genomics* **8**:449
- Schwerdtfeger C, Linden H (2001) "Blue light adaptation and desensitization of light signal transduction in *Neurospora crassa*" *Mol Microbiol* **39**:1080-1087
- Seibel C, Gremel G, Silva RD, Schuster A, Kubicek CP, Schmoll M. (2009) "Light dependent roles of the G-protein subunit GNA1 of *Hypocrea jecorina* (anamorph *Trichoderma reesei*)" *BMC Biol* **7**:58
- Smith KM, Sancar G, Dekhang R, Sullivan CM, Li S, Tag AG, Sancar C, Bredeweg EL, Priest HD, McCormick RF, Thomas TL, Carrington JC, Stajich JE, Bell-Pedersen D, Brunner M & Freitag M (2010) "Transcription factors in light and circadian clock signaling networks revealed by genome wide mapping of direct targets for *Neurospora* white collar complex" *Eukaryot Cell* **9**, 1549-1556
- Salomon M, Knieb E, von Zeppelin T, Rudiger W. (2003) "Mapping of low- and high-fluence autophosphorylation sites in phototropin 1" *Biochemistry* **42**:4217–4225
- Sulová Z y Farkaš V. (1991) "Photoinduced conidiation in *Trichoderma viride*: a study with inhibitors" *Folia Microbiol* **36**:267-270

- Támová G, Farkaš V, Betina V. (1995) "Relationship between light intensity and the changes in the intracellular level of ATP during photoinduction in *Trichoderma viride*" *Folia Microbiol* **40**:516-518
- Taylor BL, Zhulin IB. (1999) "PAS domains: internal sensors of oxygen, redox potential, and light" *Microbiol Mol Biol Rev* **63**:479-506
- Tepperman JM, Hudson ME, Khanna R, Zhu T, Chang SH, Wang X, Quail PH. (2004) "Expression profiling of phyB mutant demonstrates substantial contribution of other phytochromes to red light regulated gene expression during seedling de-etiolation" *Plant J.* **38**:725-739
- Tepperman JM, Zhu T, Chang HS, Wang X, Quail PH. (2001) "Multiple transcription-factor genes are early targets of phytochrome A signaling" *Proc Natl Acad Sci USA* **98**:9437-9442
- Terzaghi WB, Cashmore AR. (1995) "Photomorphogenesis. Seeing the light in plant development" *Curr Biol* **5**:466-468
- Tisch D, Kubicek CP, Schmoll M. (2011) "New insights into the mechanism of light modulated signaling by heterotrimeric G-proteins: ENVOY acts on *gna1* and *gna3* and adjusts cAMP levels in *Trichoderma reesei* (*Hypocrea jecorina*)" *Fungal Genet Biol* **48**:631-640
- van der Horst MA, Hellingwerf KJ. (2004) "Photoreceptor proteins, 'star actors of modern times': a review of the functional dynamics in the structure of representative members of six different photoreceptor families" *Acc Chem Res* **37**:13-20
- van Oers M, Herniou EA, Usmany M, Messelink GJ, Vlak JM. (2004) "Identification and characterization of a DNA photolyase-containing baculovirus from *Chrysodeixis chalcites*" *Virology* **330**:460-470
- Waschuk SA, Bezerra AG, Jr, Shi L, Brown LS. (2005) "*Leptosphaeria* rhodopsin: bacteriorhodopsin-like proton pump from a eukaryote" *Proc Natl Acad Sci USA* **102**:6879-6883
- Yeh K-C, Wu S-H, Murphy JT, Lagarias JC. (1997) "A cyanobacterial phytochrome two-component light sensory system" *Science* **277**:1505-1508
- Zhong X., Hao B, Chan MK. (2003) "Structure of the PAS fold and signal transduction mechanism" En: PAS proteins, regulators and sensors of development and physiology, S. Crews (editor), Kluwer Academic Publishers, 1-16
- Zoltowski, BD, Schwerdtfeger C, Widom J, Loros JJ, Bilwes AM, Dunlap JC, Crane BR. (2007) "Conformational switching in the fungal light sensor Vivid" *Science* **316**:1054-1057

Zoltowski BD, Vaccaro B, Crane BR. (2009) "Mechanism-based tuning of a LOV domain photoreceptor" *Nat Chem Biol* **5**:827–834





## The Plant Growth-Promoting Fungus *Aspergillus ustus* Promotes Growth and Induces Resistance Against Different Lifestyle Pathogens in *Arabidopsis thaliana*

Salas-Marina, Miguel Angel<sup>1†</sup>, Miguel Angel Silva-Flores<sup>1†</sup>, Mayte Guadalupe Cervantes-Badillo<sup>1†</sup>, Maria Teresa Rosales-Saavedra<sup>1</sup>, Maria Auxiliadora Islas-Osuna<sup>2</sup>, and Sergio Casas-Flores<sup>1\*</sup>

<sup>1</sup>*División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica, Camino a la Presa San José No. 2055, Colonia Lomas 4a Sección, C. P. 78216, San Luis Potosí, S.L.P., Mexico*

<sup>2</sup>*Laboratorio de Genética y Biología Molecular de Plantas, Centro de Investigación en Alimentación y Desarrollo. P. O. Box 1735, Hermosillo, Sonora, 83000, Mexico*

Received: January 10, 2011 / Revised: April 18, 2011 / Accepted: April 19, 2011

To deal with pathogens, plants have evolved sophisticated mechanisms including constitutive and induced defense mechanisms. Phytohormones play important roles in plant growth and development, as well as in the systemic response induced by beneficial and pathogen microorganisms. In this work, we identified an *Aspergillus ustus* isolate that promotes growth and induces developmental changes in *Solanum tuberosum* and *Arabidopsis thaliana*. *A. ustus* inoculation on *A. thaliana* and *S. tuberosum* roots induced an increase in shoot and root growth, and lateral root and root hair numbers. Assays performed on *Arabidopsis* lines to measure reporter gene expression of auxin-induced/repressed or cell cycle controlled genes (*DR5* and *CycB1*, respectively) showed enhanced GUS activity, when compared with mock-inoculated seedlings. To determine the contribution of phytohormone signaling pathways in the effect elicited by *A. ustus*, we evaluated the response of a collection of hormone mutants of *Arabidopsis* defective in auxin, ethylene, cytokinin, or abscisic acid signaling to the inoculation with this fungus. All mutant lines inoculated with *A. ustus* showed increased biomass production, suggesting that these genes are not required to respond to this fungus. Moreover, we demonstrated that *A. ustus* synthesizes auxins and gibberellins in liquid cultures. In addition, *A. ustus* induced systemic resistance against the necrotrophic fungus *Botrytis cinerea* and the hemibiotrophic bacterium *Pseudomonas syringae* DC3000, probably through the induction of the expression of salicylic acid, jasmonic acid/ethylene, and camalexin defense-related genes in *Arabidopsis*.

**Keywords:** *Aspergillus*, plant growth-promoting fungus, *Arabidopsis*, salicylic acid, jasmonic acid, systemic resistance

In their natural setting, plants have to deal with a whole range of environmental changes that determine plant growth and development. Hormones and many endogenous signals regulate plant growth and development, which in combination with the genetic information determine the plant's shape (development). Auxins and cytokinins (CK) regulate cell division and expansion, lateral-root development, and apical dominance [15, 20, 21, 33]. Gibberellins (GAs) and brassinosteroids (BS) promote germination, stem elongation, and flowering, and regulate photomorphogenesis [44]. Abscisic acid (ABA) is involved in several stress signaling pathways and promotes seed dormancy [2].

Microorganisms interact with plants either beneficially or as pathogens and may influence plant growth and development. Perception of microorganisms by plants is highly coordinated through cellular processes that determine the final outcome of the relationship, ranging from parasitism to mutualism [3, 36]. A number of plant-associated microorganisms have been described to synthesize phytohormones, which are necessary to mediate communication between the plant and microorganisms. Free-living microorganisms are also able to produce phytohormones [6]. It is well known that many bacteria are able to stimulate plant growth through direct or indirect mechanisms; these are called plant growth-promoting rhizobacteria (PGPR) [6]. PGPR increase plant growth by improving mineral nutrition, disease suppression, or phytohormones production, such as auxins, CK, GAs, and volatile organic compounds [11].

Like PGPR, some rhizosphere fungi are able to promote plant growth and development. Fungi described in the

\*Corresponding author

Phone: +52 444 8342000; Fax: +52 444 8342010;  
E-mail: scasas@ipicyt.edu.mx

<sup>†</sup>These authors contributed equally to this work.

RESEARCH

Open Access

# Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*

Christian P Kubicek<sup>1\*</sup>, Alfredo Herrera-Estrella<sup>2</sup>, Verena Seidl-Seiboth<sup>1</sup>, Diego A Martinez<sup>3</sup>, Irina S Druzhinina<sup>1</sup>, Michael Thon<sup>4</sup>, Susanne Zeilinger<sup>1</sup>, Sergio Casas-Flores<sup>5</sup>, Benjamin A Horwitz<sup>6</sup>, Prasun K Mukherjee<sup>7</sup>, Mala Mukherjee<sup>6</sup>, László Kredics<sup>8</sup>, Luis D Alcaraz<sup>2</sup>, Andrea Aerts<sup>9</sup>, Zsuzsanna Antal<sup>8</sup>, Lea Atanasova<sup>1</sup>, **Mayte G Cervantes-Badillo**<sup>5</sup>, Jean Challacombe<sup>9</sup>, Olga Chertkov<sup>9</sup>, Kevin McCluskey<sup>10</sup>, Fanny Couplier<sup>11</sup>, Nandan Deshpande<sup>12</sup>, Hans von Döhren<sup>13</sup>, Daniel J Ebbole<sup>14</sup>, Edgardo U Esquivel-Naranjo<sup>2</sup>, Erzsébet Fekete<sup>15</sup>, Michel Flippi<sup>16</sup>, Fabian Glaser<sup>6</sup>, Elida Y Gómez-Rodríguez<sup>5</sup>, Sabine Gruber<sup>1</sup>, Cliff Han<sup>9</sup>, Bernard Henrissat<sup>17</sup>, Rosa Hermosa<sup>4</sup>, Miguel Hernández-Oñate<sup>2</sup>, Levente Karaffa<sup>15</sup>, Idit Kosti<sup>6</sup>, Stéphane Le Crom<sup>11</sup>, Erika Lindquist<sup>9</sup>, Susan Lucas<sup>9</sup>, Mette Lübeck<sup>18</sup>, Peter S Lübeck<sup>18</sup>, Antoine Margeot<sup>19</sup>, Benjamin Metz<sup>1</sup>, Monica Misra<sup>9</sup>, Helena Nevalainen<sup>12</sup>, Markus Omann<sup>1</sup>, Nicolle Packer<sup>12</sup>, Giancarlo Perrone<sup>20</sup>, Edith E Uresti-Rivera<sup>5</sup>, Asaf Salamov<sup>9</sup>, Monika Schmoll<sup>1</sup>, Bernhard Seiboth<sup>1</sup>, Harris Shapiro<sup>9</sup>, Serenella Sukno<sup>4</sup>, Juan Antonio Tamayo-Ramos<sup>21</sup>, Doris Tisch<sup>1</sup>, Aric Wiest<sup>10</sup>, Heather H Wilkinson<sup>14</sup>, Michael Zhang<sup>9</sup>, Pedro M Coutinho<sup>17</sup>, Charles M Kenerley<sup>14</sup>, Enrique Monte<sup>4</sup>, Scott E Baker<sup>9,22</sup> and Igor V Grigoriev<sup>9</sup>

## Abstract

**Background:** Mycoparasitism, a lifestyle where one fungus is parasitic on another fungus, has special relevance when the prey is a plant pathogen, providing a strategy for biological control of pests for plant protection. Probably, the most studied biocontrol agents are species of the genus *Hypocrea/Trichoderma*.

**Results:** Here we report an analysis of the genome sequences of the two biocontrol species *Trichoderma atroviride* (teleomorph *Hypocrea atroviridis*) and *Trichoderma virens* (formerly *Gliocladium virens*, teleomorph *Hypocrea virens*), and a comparison with *Trichoderma reesei* (teleomorph *Hypocrea jecorina*). These three *Trichoderma* species display a remarkable conservation of gene order (78 to 96%), and a lack of active mobile elements probably due to repeat-induced point mutation. Several gene families are expanded in the two mycoparasitic species relative to *T. reesei* or other ascomycetes, and are overrepresented in non-syntenic genome regions. A phylogenetic analysis shows that *T. reesei* and *T. virens* are derived relative to *T. atroviride*. The mycoparasitism-specific genes thus arose in a common *Trichoderma* ancestor but were subsequently lost in *T. reesei*.

**Conclusions:** The data offer a better understanding of mycoparasitism, and thus enforce the development of improved biocontrol strains for efficient and environmentally friendly protection of plants.

\* Correspondence: ckubicek@mail.zserv.tuwien.ac.at

<sup>1</sup>Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering Vienna University of Technology, Getreidemarkt 9, 1060 Vienna, Austria

Full list of author information is available at the end of the article

## Background

Mycoparasitism is the phenomenon whereby one fungus is parasitic on another fungus, a lifestyle that can be dated to at least 400 million years ago by fossil evidence [1]. This has special relevance when the prey is a plant pathogen, providing a strategy for biological control of pests for plant protection ('biocontrol'). The movement toward environmentally friendly agricultural practices over the past two decades has thus accelerated research in the use of biocontrol fungi [2]. Probably the most studied biocontrol agents are species of the genus *Hypocrea*/*Trichoderma*, *Trichoderma atroviride* (*Ta*) and *Trichoderma virens* (*Tv*) - teleomorphs *Hypocrea atroviridis* and *Hypocrea virens*, respectively - being among the best mycoparasitic biocontrol agents used in agriculture [3]. The beneficial effects of *Trichoderma* spp. on plants comprise traits such as the ability to antagonize soil-borne pathogens by a combination of enzymatic lysis, secretion of antibiotics, and competition for space and substrates [4,5]. In addition, it is now known that some *Trichoderma* biocontrol strains also interact intimately with plant roots, colonizing the outer epidermis layers, and acting as opportunistic, avirulent plant symbionts [6].

Science-based improvement of biocontrol agents for agricultural applications requires an understanding of the biological principles of their actions. So far, some of the molecular aspects - such as the regulation and role of cell wall hydrolytic enzymes and antagonistic secondary metabolites - have been studied in *Trichoderma* [3-5]. More comprehensive analyses (for example, by the use of subtractive hybridization techniques, proteomics or EST approaches) have also been performed with different *Trichoderma* species, but the interpretation of the data obtained is complicated by the lack of genome sequence information for the species used (reviewed in [7]).

Recently, the genome of another *Trichoderma*, *Trichoderma reesei* (*Tr*, teleomorph *H. jecorina*), which has a saprotrophic lifestyle and is an industrial producer of plant biomass hydrolyzing enzymes, has been sequenced and analyzed [8]. Here we report the genome sequencing and comparative analysis of two widely used biocontrol species of *Trichoderma*, that is, *Ta* and *Tv*. These two were chosen because they are distantly related to *Tr* [9] and represent well defined phylogenetic species [10,11], in contrast to *Trichoderma harzianum sensu lato*, which is also commonly used in biocontrol but constitutes a complex of several cryptic species [12].

## Results

### Properties of the *T. atroviride* and *T. virens* genomes

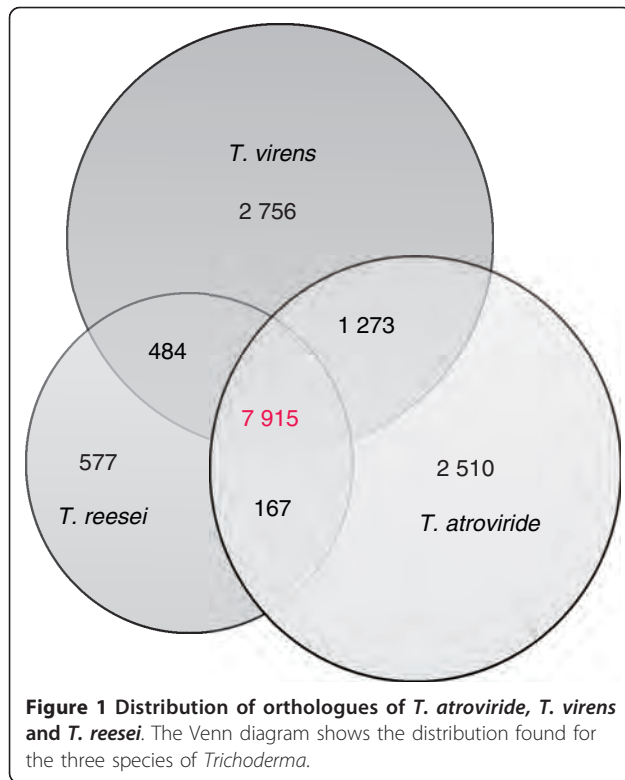
The genomes of *Ta* IMI 206040 and *Tv* Gv29-8 were sequenced using a whole genome shotgun approach to approximately eight-fold coverage and further improved using finishing reactions and gap closing. Their genome sizes were 36.1 (*Ta*) and 38.8 Mbp (*Tv*), and thus larger than the 34 Mbp determined for the genome of *Tr* [8]. Gene modeling, using a combination of homology and *ab initio* methods, yielded 11,865 gene models for *Ta* and 12,428 gene models for *Tv*, respectively (Table 1), both greater than the estimate for *Tr* (9,143). As shown in Figure 1, the vast majority of the genes (7,915) occur in all three *Trichoderma* species. Yet *Tv* and *Ta* contain about 2,756 and 2,510 genes, respectively, that have no true orthologue in any of the other species, whereas *Tr* has only 577 unique genes. *Tv* and *Ta* share 1,273 orthologues that are not present in *Tr*, which could thus be part of the factors that make *Ta* and *Tv* mycoparasites (for analysis, see below).

With respect to other ascomycetes, *Tr*, *Ta* and *Tv* share 6,306/7,091, 6,515/7,549, and 6,564/7,733 orthologues with *N. crassa* and *Gibberella zeae*, respectively.

**Table 1 Genome assembly and annotation statistics**

	<i>T. atroviride</i>	<i>T. virens</i>	<i>T. reesei</i>
Genome size, Mbp	36.1	38.8	34.1
Coverage	8.26x	8.05x	9.00x
Assembly gaps, Mbp	0.1 (0.16%)	0.2(0.4%)	0.05 (0.1%)
Number of scaffolds	50	135	89
Number of predicted genes	11865	12518	9143
Gene length, bp	1747.06	1710.05	1793,25
Protein length, amino acids	471.54	478.69	492,27
Exons per gene	2,93	2,98	3,06
Exon length, bp	528.17	506.13	507,81
Intron length, bp	104.20	104.95	119,64
Supported by homology, NR	10,219 (92%)	10,915 (94%)	8409 (92%)
Supported by homology, Swissprot	8,367(75%)	8,773 (75%)	6763 (74%)
Has PFAM domain	5,883 (53%)	6,267 (54%)	5096 (56%)

NR, non-redundant database; PFAM, protein families.



Thus, approximately a third of the genes in the three *Trichoderma* species are not shared in even the relatively close relative *G. zeae* and are thus unique to *Trichoderma*.

### Genome synteny

A comparison of the genomic organization of genes in *Ta*, *Tv* and *Tr* showed that most genes are in synteny: only 367 (4%) genes of *Tr*, but 2,515 (22%) of genes of *Tv* and 2,690 (21%) genes of *Ta* are located in non-

**Table 2** Occurrence of orthologues, paralogues and singletons in the genomes of the three *Trichoderma* spp

Genome	Synteny	Total genes	Orthologs <sup>a</sup>	Non-orthologs	P-value <sup>b</sup>
<i>T. atroviride</i>	Syntenic	9,350	7,326	2,024	2.2e-16
	Non-syntenic	2,515	1,265	1,250	
<i>T. virens</i>	Syntenic	9,828	7,326	2,502	2.2e-16
	Non-syntenic	2,690	1,532	1,158	
<i>T. reesei</i>	Syntenic	8,776	7,326	1,450	2.2e-16
	Non-syntenic	367	153	214	

<sup>a</sup>Orthologs that are in all three genomes. <sup>b</sup>Null hypothesis that the proportion of non-orthologs that are syntenic is less than the proportion of non-orthologs that are non-syntenic. P-value: null hypothesis that the proportion of paralogs that are syntenic is less than the proportion of paralogs that are non-syntenic.

syntenic regions (identified as a break in synteny by a series of three or more genes (Table 2); a global visual survey can be obtained at the genome websites of the three *Trichoderma* species (see Materials and methods) by clicking 'Synteny' and 'Dot Plot'). As observed for other fungal genomes [13-15], extensive rearrangements have occurred since the separation of these three fungi but with the prevalence of small inversions [16]. The numbers of the synteny blocks increased with their decreased size, compatible with the random breakage model [14] as in aspergilli [15,17]. Sequence identity between syntenic orthologs was 70% (*Tr* versus *Ta*), 78% (*Tr* versus *Tv*), and 74% (*Tv* versus *Ta*), values that are similar to those calculated for aspergilli (for example, *Aspergillus fumigatus* versus *Aspergillus niger* (69%) and versus *Aspergillus nidulans* (68%) and comparable to those between fish and man [17,18].

### Transposons

A scan of the genome sequences with the *de novo* repeat finding program 'Piler' [19] - which can detect repetitive elements that are at least 400 bp in length, have more than 92% identity and are present in at least three copies - was unsuccessful at detecting repetitive elements. The lack of repetitive elements detected in this analysis is unusual in filamentous fungi and suggests that, like the *Tr* genome [8], but unlike most other filamentous fungi, the *Ta* and *Tv* genomes lack a significant repetitive DNA component.

Because of the paucity of transposable elements (TEs) in the *Trichoderma* genomes, we wondered whether simple sequence repeats and minisatellite sequences may also be rare. To this end, we surveyed the genomes of the *Trichoderma* species using the program Tandem Repeat Finder [20]. We also included the genomes of three additional members of the Sordariomycetes and one of the Eurotiomycetes as reference (Table S1 in Additional file 1). Satellite DNA content varied from as little as 2,371 loci (0.53% of the genome) in *A. nidulans* to 9,893 (1.46% of the genome) in *Neurospora crassa*. Satellite DNA content of the *Trichoderma* genomes ranged from 5,249 (0.94%) in *Ta* to 7,743 (1.54%) in *Tr*. Since these values are within the range that we found in the reference species, we conclude that there is no unusual variation in the satellite DNA content of the *Trichoderma* genomes.

We also scanned the genomes with RepeatMasker and RepeatProteinMask [21] to identify sequences with similarity to known TEs from other organisms. Thereby, sequences with significant similarity to known TEs from other eukaryotes were identified (Table 3). In most cases, the TE families that we detected were fragmented and highly divergent from one another, suggesting that they did not arise from recent transposition events.

**Table 3 The major classes of transposable elements found in the *Trichoderma* genomes**

Class	<i>T. atroviridae</i>		<i>T. reesei</i>		<i>T. virens</i>	
	Copy number	Total length (bp)	Copy number	Total length (bp)	Copy number	Total length (bp)
DNA	372	39,899	446	50,448	370	52,358
LTR	533	64,534	559	76,482	541	67,484
Helitrons	40	9,235	45	9,962	34	8,547
LINE	561	65,202	530	54,928	349	59,414
Total <sup>a</sup>		178,870 (0.49%)		191,820 (0.57%)		187,803 (0.48%)

<sup>a</sup>Total in base pairs and percentage of genome of transposable elements found in the genomes. LINE, long interspersed nuclear element; LTR, long terminal repeat.

Based on these results, we conclude that no extant, functional TEs exist in the *Trichoderma* genomes. The presence of ancient, degenerate TE copies suggests that *Trichoderma* species are occasionally subject to infection, or invasion by TEs, but that the TEs are rapidly rendered unable to replicate and rapidly accumulate mutations.

#### Evidence for the operation of repeat-induced point mutation in *Trichoderma*

The paucity of transposons in *Trichoderma* could be due to repeat-induced point mutation (RIP), a gene silencing mechanism. In *N. crassa* and many other filamentous fungi, RIP preferentially acts on CA dinucleotides, changing them to TA [22]. Thus, in sequences that have been subject to RIP, one should expect to find a decrease in the proportion of CA dinucleotides and its complement dinucleotide TG as well as a corresponding

increase in the proportion of TA dinucleotides. The RIP indices TA/AT and (CA + TG)/(AC + GT) developed by Margolin *et al.* [22] can be used to detect sequences that have been subject to RIP. Sequences that have been subjected to RIP are expected to have a high TA/AT ratio and low (CA + TG)/(AC + GT) ratio, with values >0.89 and <1.03, respectively, being indicative of RIP [22,23].

To identify evidence for RIP in the TE sequences, we computed RIP indices for four of the most prevalent TE families in each of the three species (Table 4). Since many of the sequences are very short, we computed the sum of the dinucleotide values within each TE family within each species, and used the sums to compute the RIP ratios. In only one of the 12 families did we find that both RIP indices were within the ranges that are typically used as criteria for RIP. Most of the TE sequences that we identified in the *Trichoderma* genomes are highly degenerate and have likely continued to accumulate mutations after the RIP process has acted on them. We suspect that these mutations have masked the underlying bias in dinucleotide frequencies, making the RIP indices ineffective at identifying the presence of RIP. To overcome this, we also prepared manually curated multiple sequence alignments of the TE families, selecting only sequences that had the highest sequence similarity, and thus should represent the most recent transposon insertion events in the genomes. We were able to prepare curated alignments for all four of the test TE families of *Tr* and *Tv* only for the long terminal repeat element Gypsy and the long interspersed nuclear element R1 in *Ta* (Table S2 in Additional file 1). Among DNA sequences that make up these ten alignments, we detected RIP indices within the parameters that are indicative of RIP in seven alignments. In addition, all seven alignments have high transition/transversion ratios, as is expected in sequences that are subject to RIP.

Finally, screening of the genome sequences of *Tr*, *Ta* and *Tv* identified orthologues of all genes required for RIP in *N. crassa* (Table 5).

**Table 4 Repeat-induced point mutation ratios for four of the most abundant transposable element families in the three *Trichoderma* species**

Sequence	TA/AT ratio	CT+AT/AC+GT ratio	RIP <sup>a</sup>
<i>T. atroviride</i>	0.70	1.35	
LTR Copia	0.42	1.50	
LTR Gypsy	0.97	1.21	
LINE R1	1.86	1.67	
LINE Tad1	0.82	1.32	
<i>T. reesei</i>	0.71	1.28	
LTR Copia	1.04	1.31	
LTR Gypsy	1.01	1.28	
LINE R1	0.99	2.40	
LINE Tad1	0.33	1.30	
<i>T. virens</i>	0.71	1.33	
LTR Copia	0.77	1.48	
LTR Gypsy	0.95	1.16	
LINE R1	0.75	2.14	
LINE Tad1	1.33	0.99	*

<sup>a</sup>The asterisk indicates the family Tad1 from *T. virens* in which the RIP ratios fall within values that are typically associated with RIP. LINE, long interspersed nuclear element; LTR, long terminal repeat; RIP, repeat-induced point mutation; TE, transposable element.

**Table 5 Presence of genes in *Trichoderma* known to be required in *N. crassa* for repeat-induced point mutation**

<i>N. crassa</i> protein <sup>a</sup>	Accession number <sup>a</sup>	Function <sup>a</sup>	<i>Trichoderma</i> orthologue (ID number)		
			<i>T. atroviride</i>	<i>T. virens</i>	<i>T. reesei</i>
<b>RIP</b>					
RID	XP_959047.1	Putative DMT, essential for RIP and for MIP			
Dim-5	XP_957479.2	Histone 3-K9 HMT essential for RIP; RdRP	152017	55211	515216
<b>Quelling</b>					
QDE-1	XP_959047.1	RdRP, essential for quelling	361	64774	67742
QDE-2	XP_960365.2	Argonaute-like protein, essential for quelling	79413	20883	49832
QDE-3	XP_964030.2	RecQ helicase, essential for quelling	91316	30057	102458
DCL1	XP_961898.1	Dicer-like protein, involved in quelling	20162	20212	69494
DCL2	XP_963538.2	Dicer-like protein, involved in quelling	318	47151	79823
QIP	CAP68960.1	Putative exonuclease protein, involved in quelling	14588	41043	57424
<b>MSUD</b>					
SAD-1	XP_964248.2	RdRP essential for MSUD	465	28428	103470
SAD-2	XP_961084.1	Essential for MSUD	No	No	No

<sup>a</sup>*N. crassa* gene information and abbreviations taken from [36]. DMT, cytosine DNA methyltransferase; HMT, histone methyltransferase; MIP, methylation induced premeiotically; MSUD, meiotic silencing of unpaired DNA; RdRP, RNA-dependent RNA-polymerase.

#### Paralogous gene expansion in *T. atroviride* and *T. virens*

We used Marcov cluster algorithm (MCL) analysis [24] and included ten additional ascomycete genomes present in the Joint Genome Institute (JGI) genome database (including Eurotiomycetes, Sordariomycetes and Dothidiomycetes) to identify paralogous gene families that have become expanded either in all three *Trichoderma* species or only in the two mycoparasitic *Trichoderma* species. Forty-six such families were identified for all three species, of which 26 were expanded only in *Ta* and *Tv*. The largest paralogous expansions in all three *Trichoderma* species have occurred with genes encoding Zn(2)Cys(6) transcription factors, solute transporters of the major facilitator superfamily, short chain alcohol dehydrogenases, S8 peptidases and proteins bearing ankyrin domains (Table 6). The most expanded protein sets, however, were those that were considerably smaller in *Tr* ( $P < 0.05$ ). These included ankyrin proteins with CCHC zinc finger domains, proteins with WD40, heteroincompatibility (HET) and NACHT domains, NAD-dependent epimerases, and sugar transporters.

#### Genes with possible relevance for mycoparasitism are expanded in *Trichoderma*

Mycoparasitism depends on a combination of events that include lysis of the prey's cell walls [3,4,7]. The necessity to degrade the carbohydrate armor of the prey's hyphae is reflected in an abundance of chitinolytic enzymes (composing most of the CAZy (Carbohydrate-Active enZYmes database) glycoside hydrolase (GH) family GH18 fungal proteins along with more rare endo- $\beta$ -N-acetylglucosaminidases) and  $\beta$ -1,3-glucanases (families GH17, GH55, GH64, and GH81) in

*Trichoderma* relative to other fungi. Family GH18, containing enzymes involved in chitin degradation, is also strongly expanded in *Trichoderma*, but particularly in *Tv* and *Ta*, which contain the highest number of chitinolytic enzymes of all described fungi (Table 7). Chitin is a substantial component of fungal cell walls and chitinases are therefore an integral part of the mycoparasitic attack [3,25]. It is conspicuous that not only was the number of chitinolytic enzymes elevated but that many of these chitinases contain carbohydrate binding domains (CBMs). Mycoparasitic *Trichoderma* species are particularly rich in subgroup B chitinases that contain CBM1 modules, historically described as cellulose binding modules, but binding to chitin has also been demonstrated [26]. *Tv* and *Ta* each have a total of five CBM1-containing GH18 enzymes. Subgroup C chitinases possess CBM18 (chitin-binding) and CBM50 modules (also known as LysM modules; described as peptidoglycan- and chitin-binding modules). Interestingly, CBM50 modules in *Trichoderma* are found not only in chitinases but also frequently as multiple copies in proteins containing a signal peptide, but with no identifiable hydrolase domain. In most cases these genes can be found adjacent to chitinases in the genome.

Together with the expanded presence of chitinases, the number of GH75 chitosanases is also significantly expanded in all three analyzed *Trichoderma* species. As with plant pathogenic fungi [27,28], we have also observed an expansion of plant cell wall degrading enzyme gene families. A full account of all the carbohydrate active enzymes is presented in Tables S3 to S8 in Additional file 1. Additional details about the *Trichoderma* CAZome (the genome-wide inventory of CAZy) are given in Chapter 1 of Additional file 2.

**Table 6 Major paralogous gene expansions in *Trichoderma***

PFAM domain	<i>T. reesei</i>	<i>T. virens</i>	<i>T. atroviride</i>	Other fungi <sup>a</sup>
Unknown protein with ankyrin (PF00023), CCHC zinc finger (PF00098; C-X2-C-X4-H-X4-C) and purine nucleoside phosphorylase domain (01048)	19	<b>38</b>	<b>45</b>	4
<b>Zn(II)Cys6 transcription factor (00172) cluster 1-5</b>	20	<b>43</b>	<b>42</b>	5,1
Peptidase S8 subtilisin cluster 1-4	10	<b>33</b>	<b>36</b>	9,6
<b>Unknown protein with WD40, NACHT and HET domain</b>	13	<b>38</b>	<b>35</b>	3,4
<b>Short chain alcohol dehydrogenase (PF00106) cluster 1 and 2</b>	20	32	34	4,7
<b>Unknown protein family 1-4</b>	12	<b>25</b>	<b>28</b>	5
<b>NAD-dependent epimerase (PFAM 01370)</b>	10	<b>21</b>	<b>23</b>	5,8
<b>Isoflavon reductase, plus PAPA-1 (INO80 complex subunit B), epimerase and Nmr1 domain</b>	9	<b>18</b>	<b>19</b>	6
<b>Ankyrin domain protein</b>	10	17	19	8
<b>Sugar transporters</b>	11	24	18	10,8
<b>GH18 chitinases</b>	6	<b>11</b>	<b>16</b>	2
<b>Protein kinase (00069) plus TPR domain</b>	2	<b>24</b>	<b>15</b>	4,7
Unknown major facilitator subfamily (PF07690) domain	9	15	15	5,5
F-box domain protein	7	10	11	1,7
Ankyrin domain protein with protein kinase domain	6	8	11	2,7
Amidase	4	11	11	2,8
Epoxide hydrolase (PF06441) plus AB hydrolase_1 (PF00561)	5	14	11	3,2
FAD_binding_4, plus HET and berberine bridge enzymes (08031) domain	5	13	11	6,1
FMN oxidoreductases	2	8	10	2,5
Unknown protein with DUF84 (NTPase) and NmrA domain	5	19	10	3,7
Protein with GST_N and GST_C domains	6	12	10	4,6
<b>Class II hydrophobins</b>	6	8	9	1,1
<b>Proteins with LysM binding domains</b>	6	7	9	1,2
<b>Unknown protein family with NmrA domain</b>	2	<b>11</b>	<b>8</b>	0,2
<b>Pro_CA</b>	5	9	8	1,3
<b>WD40 domain protein</b>	5	11	8	2,2
<b>C2H2 transcription factors</b>	1	<b>5</b>	<b>7</b>	1,4
GFO_IDH_MocA (01408 and 02894) oxidoreductase	3	9	7	1,5
<b>Protein kinase (00069)</b>	4	6	6	0,7
<b>Nonribosomal peptide synthase</b>	3	4	5	1
<b>SSCP ceratoplatanin-family</b>	3	4	5	1
<b>GH75 chitosanase</b>	3	5	5	1,1
<b>SNF2, DEAD box helicase</b>	3	5	5	1,3
Nitrilase	3	6	5	2,2
<b>GH65 trehalose or maltose phosphorylase (PFAM 03632)</b>	4	4	4	0,8
AAA-family ATPase (PF00004)	4	3	4	1
Pyridoxal phosphate dependent decarboxylase (00282)	2	3	4	1,2
Unknown protein	3	4	4	1,3

<sup>a</sup>Results are from MCL analysis of the three *Trichoderma* species (*Tr*, *Ta*, *Tv*) and mean values from ten other ascomycetes whose genomes are present in the JGI database [63]. Eurotiomycetes: *Aspergillus carbonarius*, *Aspergillus niger*. Sordariomycetes: *Thielavia terrestris*, *Chaetomium globosum*, *Cryphonectria parasitica*, *Neurospora discreta*, *Neurospora tetrasperma*. Dothidiomycetes: *Mycosphaerella graminicola*, *Mycosphaerella fijiensis*, *Cochliobolus heterostrophus*. The number of genes present in the "other fungi" is averaged. Data were selected from a total of 28,919 clusters, average cluster number 5.8 (standard deviation 15.73). PFAM categories printed in bold specify those that are significantly ( $P < 0.05$ ) expanded in all three *Trichoderma* species; numbers in bold and italics specify genes that are significantly more abundant in *Ta* and *Tv* versus *Tr* ( $P < 0.05$ ). GH, glycosyl hydrolase family; GST, glutathione-S transferase; SSCP, small secreted cysteine-rich protein.

Another class of genes of possible relevance to myco-parasitism are those involved in the formation of secondary metabolites (Chapter 2 of Additional file 2). With respect to these, the three *Trichoderma* species contained a varying assortment of non-ribosomal

peptide synthetases (NRPS) and polyketide synthases (PKS) (Table 8; see also Tables S9 and S10 in Additional file 1). While *Tr* (10 NRPS, 11 PKS and 2 NRPS/PKS fusion genes [8]) ranked at the lower end when compared to other ascomycetes, *Tv* exhibited the highest

**Table 7 Glycosyl hydrolase families involved in chitin/chitosan and  $\beta$ -1,3 glucan hydrolysis that are expanded in mycoparasitic *Trichoderma* species**

	Taxonomy	Glycosyl hydrolase family					Total $\beta$ -glucan <sup>b</sup>	
		Chitin/chitosan <sup>a</sup>		$\beta$ -glucan <sup>a</sup>				
		GH18	GH75	GH17	GH55	GH64		GH81
<i>Trichoderma atroviride</i>	S	<b>29</b>	<b>5</b>	5	<b>8</b>	<b>3</b>	2	18
<i>Trichoderma virens</i>	S	<b>36</b>	<b>5</b>	4	<b>10</b>	<b>3</b>	1	18
<i>Trichoderma reesei</i>	S	<b>20</b>	<b>3</b>	4	<b>6</b>	<b>3</b>	2	15
<b>Pezizomycota</b>								
<i>Nectria haematococca</i>	S	28	2	6	5	2	1	14
<i>Fusarium graminearum</i>	S	19	1	6	3	2	1	12
<i>Neurospora crassa</i>	S	12	1	4	6	2	1	13
<i>Podospira anserina</i>	S	20	1	4	7	1	1	13
<i>Magnaporthe grisea</i>	S	14	1	7	3	1	2	13
<i>Aspergillus nidulans</i>	E	19	2	5	6	0	1	12
<i>Aspergillus niger</i>	E	14	2	5	3	0	1	9
<i>Penicillium chrysogenum</i>	E	9	1	5	3	2	1	11
<i>Tuber melanosporum</i>	P	5	1	4	2	0	3	9
<b>Other ascomycetes</b>								
<i>Saccharomyces cerevisiae</i>	SM	2	0	4	0	0	2	6
<i>Schizosaccharomyces pombe</i>	SS	1	0	1	0	0	1	2
<b>Basidiomycota</b>								
<i>Phanerochaete chrysosporium</i>	A	11	0	2	2	0	0	4
<i>Laccaria bicolor</i>	A	10	0	4	2	0	0	6
<i>Postia placenta</i>	A	20	0	4	6	0	0	10

<sup>a</sup>Main substrates for the respective enzyme families. <sup>b</sup>Number of all enzymes that can act on  $\beta$ -glucan as a substrate. Taxonomy abbreviations: S, Sordariomycetes; E, Eurotiomycetes; P, Pezizomycetes; S, Saccharomycetes; SS, Schizosaccharomycetes; A, Agaricomycetes. The bold numbers indicate glycosyl hydrolase (GH) families that have a statistically significant expansion in *Trichoderma* ( $P < 0.05$ ) or *Ta* and *Tv* (GH18). This support was obtained only when *N. haematococca* and *T. melanosporum* were not included in the analysis of GH18 and GH81, respectively.

number (50) of PKS, NRPS and PKS-NRPS fusion genes, mainly due to the abundance of NRPS genes (28, twice as much as in other fungi). A phylogenetic analysis showed that this was due to recent duplications of genes encoding cyclodipeptide synthases, cyclosporin/enniatin

**Table 8 The number of polyketide synthases and non-ribosomal peptide synthetases of *Trichoderma* compared to other fungi**

Fungal species	PKS	NRPS	PKS-NRPS NRPS-PKS	Total
<i>Trichoderma virens</i>	<b>18</b>	<b>28</b>	<b>4</b>	<b>50</b>
<i>Aspergillus oryzae</i>	26	14	4	44
<i>Aspergillus nidulans</i>	26	13	1	40
<i>Cochliobolus heterostrophus</i>	23	11	2	36
<i>Trichoderma atroviride</i>	<b>18</b>	<b>16</b>	<b>1</b>	<b>35</b>
<i>Magnaporthe oryzae</i>	20	6	8	34
<i>Fusarium graminearum</i>	14	19	1	34
<i>Gibberella moniliformis</i>	12	16	3	31
<i>Botryotinia fuckeliana</i>	17	10	2	29
<i>Aspergillus fumigatus</i>	13	13	1	27
<i>Nectria haematococca</i>	12	12	1	25
<i>Trichoderma reesei</i>	<b>11</b>	<b>10</b>	<b>2</b>	<b>23</b>
<i>Neurospora crassa</i>	7	3	0	10

synthase-like proteins, and NRPS-hybrid proteins (Figure S1 in Additional file 3). Most of the secondary metabolite gene clusters present in *Tr* were also found in *Tv* and *Ta*, but about half of the genes remaining in the latter two are unique for the respective species, and are localized on non-syntenic islands of the genome (see below). Within the NRPS, all three *Trichoderma* species contained two peptaibol synthases, one for short (10 to 14 amino acids) and one for long (18 to 25 amino acids) peptaibols. The genes encoding long peptaibol synthase lack introns and produce an mRNA that is 60 to 80 kb long that encodes proteins of approximately 25,000 amino acids, the largest fungal proteins known.

Besides PKS and NRPS, *Ta* and *Tv* have further augmented their antibiotic arsenal with genes for cytolytic peptides such as aegerolysins, pore-forming cytolsins typically present in bacteria, fungi and plants, yeast-like killer toxins and cyanovirins (Chapter 2 of Additional file 2). In addition, we found two high molecular weight toxins in *Ta* and *Tv* that bear high similarity (E-value 0 for 97% coverage) to the Tc ('toxin complex') toxins of *Photorhabdus luminescens*, a bacterium that is mutualistic with entomophagous nematodes [29] (Table S11 in Additional file 1). Apart from *Trichoderma*, they are



also present in *G. zeae* and *Podospora anserina*. Yet there may be several more secondary metabolite genes to be detected: *Trichoderma* species contain expanded arrays of cytochrome P450 CYP4/CYP19/CYP26 subfamilies (Table S12 in Additional file 1), and of soluble epoxide hydrolases that could act on the epoxides produced by the latter (Figure S2 in Additional file 3).

The *Hypocrea/Trichoderma* genomes also contain an abundant arsenal of putatively secreted proteins of 300 amino acids or less that contain at least four cysteine residues (small secreted cysteine-rich proteins (SSCPs); Chapter 3 of Additional file 2). They contained both unique and shared sets of SSCP, with a higher complexity in *Tv* and *Ta* than in *Tr* (Table S13 in Additional file 1).

#### Genes present in *T. atroviride* and *T. virens* but not in *T. reesei*

As mentioned above, 1,273 orthologous genes were shared between *Ta* and *Tv* but absent from *Tr*. When the encoded proteins were classified according to their PFAM domains, fungal specific Zn(2)Cys(6) transcription factors (PF00172, PF04082) and solute transporters (PF07690, PF00083), all of unknown function, were most abundant (Table S14 in Additional file 1). However, the presence of several PFAM groups of oxidoreductases and monooxygenases, and of enzymes for AMP activation of acids, phosphopatheine attachment and synthesis of isoquinoline alkaloids was also intriguing. This suggests that *Ta* and *Tv* may contain an as yet undiscovered reservoir of secondary metabolites that may contribute to their success as mycoparasites.

We also annotated the 577 genes that are unique in *T. reesei*: the vast majority of them (465; 80.6%) encoded proteins of unknown function or proteins with no homologues in other fungi. The remaining identified 112 genes exhibited no significant abundance in particular groups, except for four Zn(2)Cys(6) transcription factors, four ankyrins, four HET-domain proteins and three WD40-domain containing proteins.

#### Evolution of the non-syntenic regions

A search for overrepresentation of PFAM domains and Gene Ontology terms in the non-syntenic regions described above revealed that all retroposon hot spot repeat domains [30] are found in the non-syntenic regions. In most eukaryotes, these regions are located in subtelomeric areas that exhibit a high recombination frequency [31]. In addition, the genes for the protein families in *Tv* and *Ta* that were significantly more abundant compared to *Tr* were enriched in the non-syntenic areas (Table 9). In addition, the number of paralogous genes was significantly increased in the non-syntenic regions. We considered three possible explanations for

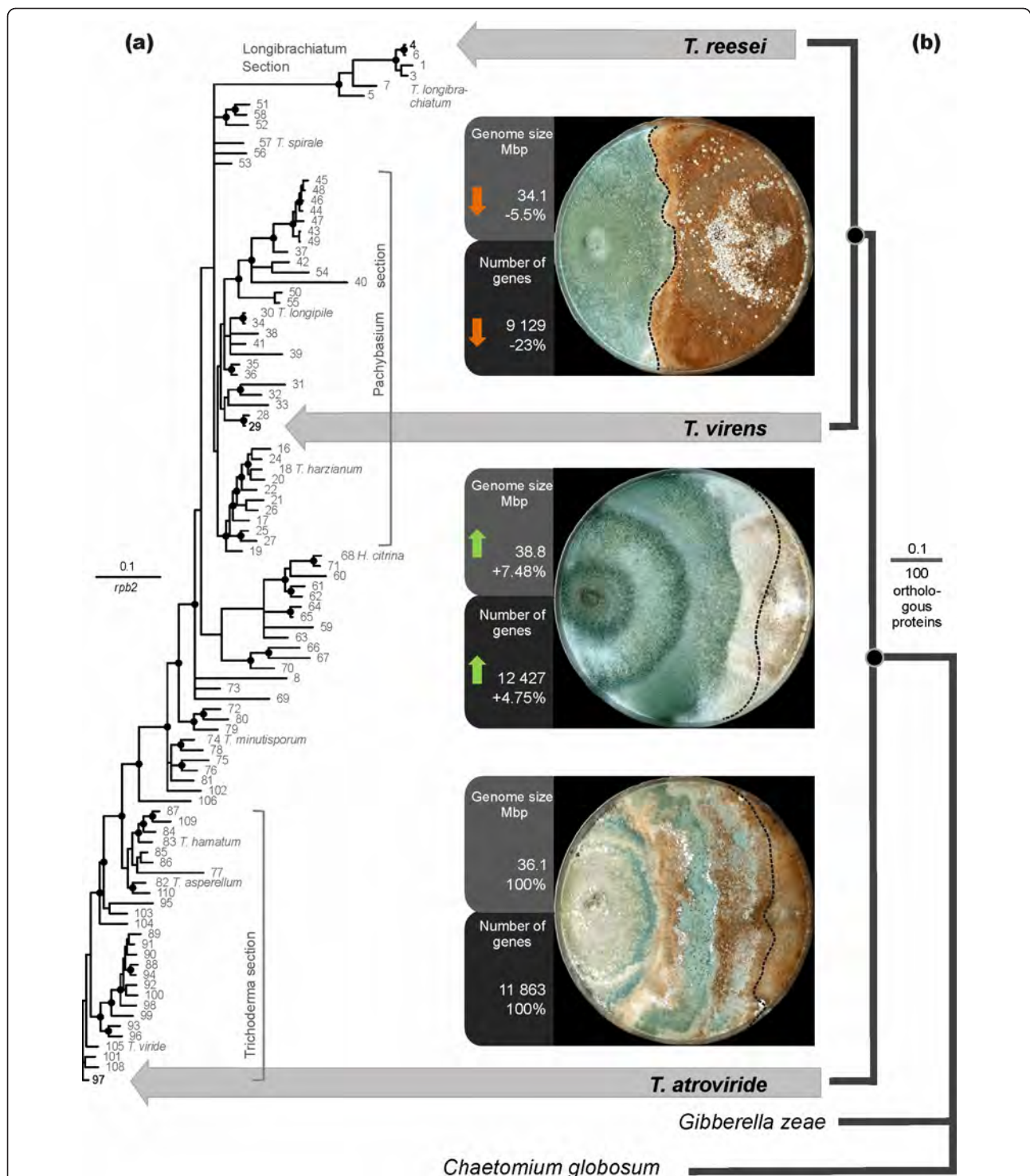
**Table 9 Number of PFAM domains that are enriched among paralogous genes in non-syntenic areas**

	<i>T. reesei</i>	<i>T. virens</i>	<i>T. atroviride</i>
Zn2Cys6 transcription factors	9	<b>95</b>	69
WD40 domains	1	11	14
Sugar transporters	<b>0</b>	18	13
Proteases	2	28	23
Cytochrome P450	7	<b>40</b>	15
NmrA-domains	2	19	21
Major facilitator superfamily	7	52	60
HET domains	3	26	27
Glycoside hydrolases	3	33	26
FAD-binding proteins	2	28	24
Ankyrins	4	44	37
Alcohol dehydrogenases	<b>4</b>	51	71
$\alpha/\beta$ -fold hydrolases	2	26	15
ABC transporters	4	14	3
Number of genes	50	485	418
Total gene number in NS areas	92	686	1012

Boxed numbers are those that are significantly ( $p < 0.05$ ) different from the two other species when related to the genome size. PFAM, protein family; NS, non-syntenic; HET, heteroincompatibility.

this: the non-syntenic genes were present in the last common ancestor of all three *Trichoderma* species but were then selectively and independently lost; the non-syntenic areas arose from the core genome by duplication and divergence during evolution of the genus *Trichoderma*; and the non-syntenic genes were acquired by horizontal transfer. To distinguish among these hypotheses for their origin, we compared the sequence characteristics of the genes in the non-syntenic regions to those present in the syntenic regions in *Trichoderma* and genes in other filamentous fungi. We found that the majority (>78%) of the syntenic as well as non-syntenic encoded proteins have their best BLAST hit to other ascomycete fungi, indicating that the non-syntenic regions are also of fungal origin. Also, a high number of proteins encoded in the non-syntenic regions of *Ta* and *Tv* have paralogs in the syntenic region. Finally, codon usage tables and codon adaptation index analysis [32] indicate that the non-syntenic genes exhibit a similar codon usage (Figure S3 in Additional file 3). Taken together, the most parsimonious explanation for the presence of the paralogous genes in *Ta* and *Tv* is that the non-syntenic genes arose by gene duplication within a *Trichoderma* ancestor, followed by gene loss in the three lineages, which was much stronger in *Tr*.

*Tr*, *Ta* and *Tv* each occupy very diverse phylogenetic positions in the genus *Trichoderma*, as shown by a Bayesian *rpb2* tree of 110 *Trichoderma* taxa (Figure 2). In order to determine which of the three species more likely resembles the ancestral state of *Trichoderma*, we performed a Bayesian phylogenetic analysis [33] using a



**Figure 2** Mycoparasitism is an ancient life style of *Trichoderma*. **(a)** Position of *Ta*, *Tv* and *Tr* within the genus *Hypocrea/Trichoderma*. The positions of *Tr*, *Tv* and *Ta* are 4, 29 and 97, respectively - shown in bold, and a few hallmark species are given by their names. For the identities of the other species, see the gene accession numbers (Materials and methods). **(b)** Bayesian phylogram based on the analysis of amino acid sequences of 100 orthologous syntenic proteins (MCMC, 1 million generations, 10,449 characters) in *Tr*, *Tv*, *Ta*, *Gibberella zeae* and *Chaetomium globosum*. Circles above nodes indicate 100% posterior probabilities and significant bootstrap coefficients. The numbers in the boxes between (a) and (b) indicate the genome sizes and gene counts and percentage net gain regarding *Ta*. Photoplates show the mycoparasitic reaction after the contact between *Trichoderma* species and *Rhizoctonia solani*. *Trichoderma* species are always on the left side; dashed lines indicate the position of *Trichoderma* overgrowth of *R. solani*.

concatenated set of 100 proteins that were encoded by orthologous genes in syntenic areas in the three *Trichoderma* species and also *G. zae* and *Chaetomium globosum*. The result (Figure 2) shows that *Ta* occurs in a well-supported basal position to *Tv* and *Tr*. These data indicate that *Ta* resembles the more ancient state of *Trichoderma* and that both *Tv* and *Tr* evolved later. The lineage to *Tr* thus appears to have lost a significant number of genes present in *Ta* and maintained in *Tv*. The long genetic distance of *Tr* further suggests that it was apparently evolving faster than *Ta* and *Tv* since the time of divergence.

To test this assumption, we compared the evolutionary rates of the 100 orthologous and syntenic gene families between the three *Trichoderma* species. The median values of the evolutionary rates ( $K_s$  and  $K_a$ ) of *Ta-Tr* and *Tv-Tr* were all significantly higher (1.77 and 1.47, and 1.33 and 1.19, respectively) than those of *Ta-Tv* (1.13 and 0.96; all  $P$  values <0.05 by the two-tailed Wilcoxon rank sum test). This result supports the above suggestion that *Tr* has been evolving faster than *Ta* and *Tv*.

## Discussion

Comparison of the genomes of two mycoparasitic and one saprotrophic *Trichoderma* species revealed remarkable differences: in contrast to the genomes of other multicellular ascomycetes, such as aspergilli [15,17], those of *Trichoderma* appear to have the highest level of synteny of all genomes investigated (96% for *Tr* and still 78/79% for *Tv* and *Ta*, respectively, versus 68 to 75% in aspergilli), and most of the differences between *Ta* and *Tv* versus *Tr* or other ascomycetes occur in the non-syntenic areas. Nevertheless, at a molecular level the three species are as distant from each other as apes from *Pices* (fishes) or *Aves* (birds) [17], suggesting a mechanism maintaining this high genomic synteny. Espagne et al. [13] proposed that a discrepancy of genome evolution between *P. anserina*, *N. crassa* and the aspergilli and saccharomycotina yeasts is based on the difference between heterothallic and homothallic fungi: in heterothallics the presence of interchromosomal translocation could result in chromosome breakage during meiosis and reduced fertility, whereas homothallism allows translocations to be present in both partners and thus have fewer consequences on fertility. Since *Trichoderma* is heterothallic [34], this explanation is also applicable to it. However, another mechanism, meiotic silencing of unpaired DNA [35] - which has also been proposed for *P. anserina* [13], and which eliminates progeny in crosses involving rearranged chromosomes in one of the partners - may not function in *Trichoderma* because one of the essential genes (*SAD2* [36]) is missing.

Our data also suggest that the ancestral state of *Hypocrea/Trichoderma* was mycoparasitic. This supports an earlier speculation [37] that the ancestors of *Trichoderma* were mycoparasites on wood-degrading basidiomycetes and acquired saprotrophic characteristics to follow their prey into their substrate. Indirect evidence for this habitat shift in *Tr* was also presented by Slot and Hibbett [38], who demonstrated that *Tr* - after switching to a specialization on a nitrogen-poor habitat (decaying wood) - has acquired a nitrate reductase gene (which was apparently lost earlier somewhere in the Sordariomycetes lineage) by horizontal gene transfer from basidiomycetes.

Furthermore, the three *Trichoderma* species have the lowest number of transposons reported so far. This is unusual for filamentous fungi, as most species contain approximately 10 to 15% repetitive DNA, primarily composed of TEs. A notable exception is *Fusarium graminearum* [27], which, like the *Trichoderma* species, contains less than 1% repetitive DNA [8]. The paucity of repetitive DNA may be attributed to RIP, which has been suggested to occur in *Tr* [8] and for which we have here provided evidence that it also occurs in *Ta* and *Tv*. It is likely that this process also contributes to prevent the accumulation of repetitive elements.

The gene inventory detected in the three *Trichoderma* species reveals new insights into the physiology of this fungal genus: the strong expansion of genes for solute transport, oxidoreduction, and ankyrins (a family of adaptor proteins that mediate the anchoring of ion channels or transporters in the plasma membrane [39]) could render *Trichoderma* more compatible in its habitat (for example, to successfully compete with the other saprotrophs for limiting substrates). In addition, the expansion of WD40 domains acting as hubs in cellular networks [40] could aid in more versatile metabolism or response to stimuli. These features correlate well with a saprotrophic lifestyle that makes use of plant biomass that has been pre-degraded by earlier colonizers. The expansion of HET proteins (proteins involved in vegetative incompatibility specificity) on the other hand suggests that *Trichoderma* species may frequently encounter related yet genetically distinct individuals. In fact, the presence of several different *Trichoderma* species can be detected in a single soil sample [41]. Unfortunately, vegetative incompatibility has not yet been investigated in any *Trichoderma* species, and based on the current data, should be a topic of future research.

Finally, the abundance of SSCPs in *Trichoderma* may be involved in rhizosphere competence: the genome of the ectomycorrhizal basidiomycete *Laccaria bicolor* also encodes a large set of SSCPs, which accumulate in the hyphae that colonize the host root [42].

Gene expansions in *Tv* and *Ta* that do not occur in *Tr* may comprise genes specific for mycoparasitism.

As a prominent example, proteases have expanded in *Ta* and *Tv*, supporting the hypothesis that the degradation of proteins is a major trait of mycoparasites [43]. Likewise, the increase in chitinolytic enzymes and some  $\beta$ -glucanase-containing GH families is remarkable and illustrates the importance of destruction of the prey's cell wall in this process. With respect to the chitinases, the expansion of those bearing CBM50 modules was particularly remarkable: proteins containing these modules were recently classified into several different groups by de Jonge and Thomma [44]. Proteins that consist solely of CBM50 modules are type-A LysM proteins, and there is evidence for the role of these as virulence factors in plant pathogenic fungi. The high numbers of LysM proteins that are found in *Trichoderma*, however, indicate other/additional roles for these proteins in fungal biology that are not understood yet. Also, the expansion of the GH75 chitosanases was intriguing: chitosan is a partially deacetylated derivative of chitin and, depending on the fungal species and the growth conditions, in mature fungal cell walls chitin is partially deacetylated. It has also been reported that fungi deacetylate chitin as a defense mechanism [45,46]. Chitosan degradation may therefore be a relevant aspect of mycoparasitism and fungal cell wall degradation that has also not been regarded yet. Overall, the carbohydrate-active enzyme machinery present in *Trichoderma* is compatible with saprophytic behavior but, interestingly, the set of enzymes involved in the degradation of 'softer' plant cell wall components, such as pectin, is reduced. A possible plant symbiotic relationship [3] might rely on a mycoparasitic capacity along with a reduced specificity for pectin, minimizing the plant defense reaction.

Although the genes encoding proteins for the synthesis of typical fungal secondary metabolites (PKS, NRPS, PKS-NRPS) are also abundant, they are not significantly more expanded than in some other fungi. An exception is *Tv* and its 28 NRPS genes. However, our genome analysis revealed also a high number of oxidoreductases, cytochrome P450 oxidases, and other enzymes that could be part of as yet unknown pathways for the synthesis of further secondary metabolites. In support of this, several of these genes were found to be clustered in the genome (data not shown), and were more abundant in the two mycoparasitic species *Ta* and *Tv*. Together with the expanded set of oxidoreductases, monooxygenases, and enzymes for AMP activation of acids, phosphothetheine attachment, and synthesis of isoquinoline alkaloids in *Ta* and *Tv*, these genes may define new secondary metabolite biosynthetic routes.

## Conclusions

Our comparative genome analysis of the three *Trichoderma* species now opens new opportunities for the

development of improved and research-driven strategies to select and improve *Trichoderma* species as biocontrol agents. The availability of the genome sequences published in this study, as well as of several pathogenic fungi and their potential host plants (for example, [47]) provides a challenging opportunity to develop a deeper understanding of the underlying processes by which *Trichoderma* interacts with plant pathogens in the presence of living plants within their ecosystem.

## Materials and methods

### Genome sequencing and assembly

The genomes of *T. virens* and *T. atroviride* each were assembled from shotgun reads using the JGI (USA Department of Energy) assembler Jazz (see Table S15 in Additional file 1 for summary of assembly statistics). Each genome was annotated using the JGI Annotation pipeline, which combines several gene prediction, annotation and analysis tools. Genes were predicted using Fgenesh [48], Fgenesh+ [49], and Genewise programs [50]. ESTs from each species (Chapter 4 of Additional file 2) were clustered and either assembled and converted into putative full-length genes directly mapped to genomic sequence or used to extend predicted gene models into full-length genes by adding 5' and/or 3' untranslated regions to the models. From multiple gene models predicted at each locus, a single representative model was chosen based on homology and EST support and used for further analysis. Gene model characteristics and support are summarized in Tables S16 and S17 in Additional file 1.

All predicted gene models were functionally annotated by homology to annotated genes from a NCBI non-redundant set and classified according to Gene Ontology [51], eukaryotic orthologous groups (KOGs) [52], and Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways [53]. See Tables S18 and S19 in Additional file 1 for a summary of the functional annotation. Automatically predicted genes and functions were further refined by user community-wide manual curation efforts using web-based tools at [54,55]. The latest version gene set containing manually curated genes is called GeneCatalog.

Assembly and annotation data for *Tv* and *Ta* are available through JGI Genome Portals homepage at [54,55]. The genome assemblies, predicted gene models, and annotations were deposited at GenBank under project accessions [GenBank: ABDF00000000 and ABDG00000000], respectively. GenBank public release of the data described in this paper should coincide with the manuscript publication date.

### Genome similarity analysis and genomic synteny

Orthologous genes, as originally defined, imply a reflection of the history of species. In recent years, many

studies have examined the concordance between orthologous gene trees and species trees in bacteria. With the purpose of identifying all the orthologous gene pairs for the three *Trichoderma* species, a best bidirectional blast hit approach as described elsewhere [56,57] was performed, using the predicted translated gene models for each of the three species as pairwise comparison sets. The areas of relationship known as syntenic regions or syntenic blocks are anchored with orthologs (calculated as mutual best hits or bi-directional best hits) between the two genomes in question, and are built by controlling for the minimum number of genes, minimum density, and maximum gap (genes not from the same genome area) compared with randomized data as described in [56]. While this technique may cause artificial breaks, it highlights regions that are dynamic and picking up a large number of insertions or duplications.

Orthologous and paralogous gene models were identified by first using BLAST to find all pairwise matches between the resulting proteins from the gene models. The pairwise matches from BLAST were then clustered into groups of paralogs using MCL [58]. In parallel we applied orthoMCL [59] to the same pairwise matches to identify the proteins that were orthologous in all of the three genomes. By subtracting all the proteins that were identified as orthologs from the groups of paralogs and unique genes, we were left with only the protein products of gene models that have expanded since the most recent common ancestor (MRCA) of the three *Trichoderma* genomes. We then calculated the *P*-value under the null hypothesis that the number of non-orthologous genes that are non-syntenic is less than the number of non-orthologous genes that are syntenic.

#### Identification of transposable elements

We scanned the *Trichoderma* genomes with the *de novo* repeat finding program Piler [19]. Next, we searched for sequences with similarity to known repetitive elements from other eukaryotes with the program RepeatMasker [21] using all eukaryotic repetitive elements in the RepBase (version 13.09) database. After masking repetitive sequences that matched the DNA sequence of known repetitive elements, we scanned the masked genome sequences with RepeatProteinMask (a component of the RepeatMasker application). This search located additional degenerate repetitive sequences with similarity to proteins encoded by TEs in the RepBase database.

#### CAZome identification and analysis

All protein models for *Ta* and *Tv* were compared against the set of libraries of modules derived from CAZy [60,61]. The identified proteins were subjected to manual analysis for correction of the protein models, for full modular annotation and for functional inference

against a library of experimentally characterized enzymes. Comparative analysis was made by the enumeration of all modules identified in the three *Trichoderma* species and 14 other published fungal genomes.

#### Phylogenetic and evolutionary analyses

One-hundred genes were randomly selected from *Ta*, *Tv*, *Tr* and *C. globosum* based on their property to fulfill two requirements: they were in synteny in all four genomes, and they were true orthologues (no other gene encoding a protein with amino acid similarity >50% present). After alignment, the concatenated 10,449 amino acids were subjected to Bayesian analysis [33] using 1 million generations. The respective cDNA sequences (31,347 nucleotides) were also concatenated, and Ks/Ka ratios determined using DNAsp5 [62]. The same file was also used to determine the codon adaptation index [32]. In addition, 80 non-syntenic genes were also selected randomly for this purpose.

The species phylogram of *Trichoderma/Hypocrea* was constructed by Bayesian analysis of partial exon nucleotide sequences (824 total characters from which 332 were parsimony-informative) of the *rpb2* gene (encoding RNA polymerase B II) from 110 *ex*-type strains, thereby spanning the biodiversity of the whole genus. The tree was obtained after 5 million MCMC generations sampled for every 100 trees, using burnin = 1200 and applying the general time reversible model of nucleotide substitution. The NCBI ENTREZ accession numbers are: 1 [HQ260620]; 3 [DQ08724]; 4 [HM182969]; 5 [HM182984]; 6 [HM182965]; 7 [AF545565]; 8 [AF545517]; 16 [FJ442769]; 17 [AY391900]; 18 [FJ179608]; 19 [FJ442715]; 20 [FJ442771]; 21 [AY391945]; 22 [EU498358]; 23 [DQ834463]; 24 [FJ442725]; 25 [AF545508]; 26 [AY391919]; 27 [AF545557]; 28 [AF545542]; 29 [FJ442738]; 30 [AF545550]; 31 [AY391909]; 32 [AF545516]; 33 [AF545518]; 34 [AF545512]; 35 [AF545510]; 36 [AF545514]; 37 [AY391921]; 38 [AF545513]; 39 [AY391954]; 40 [AY391944]; 41 [AF545534]; 42 [AY391899]; 43 [AY391907]; 44 [AF545511]; 45 [AY391929]; 46 [AF545540]; 47 [AY391958]; 48 [AY391924]; 49 [AF545515]; 50 [AY391957]; 51 [AF545551]; 52 [AF545522]; 53 [FJ442714]; 54 [AF545509]; 55 [AY391959]; 56 [DQ087239]; 57 [AF545553]; 58 [AF545545]; 59 [DQ835518]; 60 [DQ835521]; 61 [DQ835462]; 62 [DQ835465]; 63 [DQ835522]; 64 [AF545560]; 65 [DQ835517]; 66 [DQ345348]; 67 [AF545520]; 68 [DQ835455]; 69 [AF545562]; 70 [AF545563]; 71 [DQ835453]; 72 [FJ179617]; 73 [DQ859031]; 74 [EU341809]; 75 [FJ179614]; 76 [DQ087238]; 77 [AF545564]; 78 [FJ179601]; 79 [FJ179606]; 80 [FJ179612]; 81 [FJ179616]; 82 [EU264004]; 83 [FJ150783]; 84 [FJ150767]; 85

[FJ150786]; 86 [EU883559]; 87 [FJ150785]; 88 [EU248602]; 89 [EU241505]; 90 [FJ442762]; 91 [FJ442741]; 92 [FJ442783]; 93 [EU341805]; 94 [FJ442723]; 95 [FJ442772]; 96 [EU2415023]; 97 [EU341801]; 98 [EU248600]; 99 [EU341808]; 100 [EU3418033]; 101 [EU2485942]; 102 [AF545519]; 103 [EU248603]; 104 [EU248607]; 105 [EU341806]; 106 [DQ086150]; 107 [DQ834460]; 108 [EU711362]; 109 [EU883557]; 110 [FJ150790].

## Additional material

### Additional file 1: Comparative properties and gene inventory of *T. reesei*, *T. virens* and *T. atroviride*.

This file contains additional information on genomic properties and selected gene families from the three *Trichoderma* species comprising 19 tables. Table S1 summarizes the satellite sequences identified in the *Trichoderma* genomes and four other fungal genomes. Table S2 summarizes manually curated sequence alignments of transposable element families from the *Trichoderma* genomes. Table S3 lists the total number of CAZy families in *Trichoderma* and other fungi. Table S4 lists the glycoside hydrolase (GH) families in *Trichoderma* and other fungi. Table S5 lists the glycosyltransferase (GT) families in *Trichoderma* and other fungi. Table S6 lists the polysaccharide lyase (PL) families in *Trichoderma* and other fungi. Table S7 lists the carbohydrate esterase (CE) families in *Trichoderma* and other fungi. Table S8 lists the carbohydrate-binding module (CBM) families in *Trichoderma* and other fungi. Table S9 lists the NRPS, PKS and NRPS-PKS proteins in *T. atroviride*. Table S10 lists NRPS, PKS and NRPS-PKS proteins in *T. virens*. Table S11 lists the putative insecticidal toxins in *Trichoderma*. Table S12 lists the cytochrome P450 CYP4/CYP19/CYP26 class E proteins in *Trichoderma*. Table S13 lists the small-cysteine rich secreted protein from *Trichoderma* spp. Table S14 lists the most abundant PFAM domains in those genes that are unique to *T. atroviride* and *T. virens* and not present in *T. reesei*. Table S15 surveys the assembly statistics. Table S16 provides gene model support. Table S17 summarizes gene model statistics. Table S18 provides numbers of genes with functional annotation according to KOG, Gene Ontology, and KEGG classifications. Table S19 lists the largest KOG families responsible for metabolism.

**Additional file 2: Additional information on selected gene groups of *Trichoderma*, methods used for genome sequencing, and legends for the figures in Additional file 3.** Chapter 1: Carbohydrate-Active enzymes (CAZymes). Chapter 2: Aegerolysins and other toxins. Chapter 3: Small secreted cysteine rich proteins (SSCPs). Chapter 4: EST sequencing and analysis. Chapter 5: Legends to figures.

**Additional file 3: Figures that illustrate selected aspects of the main text.** Figure S1 provides a phylogeny of *Trichoderma* NPRSs. Figure S2 compares the numbers of epoxide hydrolase genes in *Trichoderma* with that in other fungi. Figure S3 compares the codon usage in genes from syntenic and nonsyntenic regions of the genomes of *Trichoderma reesei*, *T. atroviride* and *T. virens*.

## Abbreviations

CAZy: Carbohydrate-Active enZymes; CBM: carbohydrate binding module; EST: expressed sequence tag; GH: glycosyl hydrolase; HET: heteroincompatibility; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: clusters of eukaryotic orthologous groups; NRPS: non-ribosomal peptide synthase; PKS: polyketide synthase; RIP: repeat-induced point mutation; SSCP: small secreted cysteine-rich protein; *Ta*: *Trichoderma atroviride*; TE: transposable element; *Tr*: *Trichoderma reesei*; *Tv*: *Trichoderma virens*.

## Acknowledgements

Genome sequencing and analysis was conducted by the US Department of Energy Joint Genome Institute and supported by the Office of Science of

the US Department of Energy under contract number DE-AC02-05CH11231. MGC-B, EYG-R, MH-O, and EEU-R are indebted to Conacyt for doctoral fellowships. SLC and FC was supported by the Infraestructuras en Biología Santé et Agronomie (IBISA). EM and RH work was supported by the grants Junta de Castilla y León GR67, MICINN AGL2008-0512/AGR and AGL2009-13431-C02. The work of ISD, VS-S, LA, BS, BM, SZ, MS, and CPK was supported by the Austrian Science Foundation (grants FWF P17895-B06, P20559, T390, P18109-B12, P-19421, V139B20 and P-19340). The work of PMC and BH was supported by project number AANR-07-BIOE-006 from the French national program PNRB. MF was the recipient of a postdoctoral contract Ramón y Cajal from the Spanish Ministry of Science and Innovation (MCINN: RYC-2004-003005). SZ acknowledges support from the Vienna Science and Technology Fund (WWTF LS09-036).

## Author details

<sup>1</sup>Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering Vienna University of Technology, Getreidemarkt 9, 1060 Vienna, Austria. <sup>2</sup>Laboratorio Nacional de Genómica para la Biodiversidad, Cinvestav Campus Guanajuato, Km. 9.6 Libramiento Norte, Carretera Irapuato-León, 36821 Irapuato, Mexico. <sup>3</sup>Broad Institute of MIT and Harvard, 301 Binney St, Cambridge, MA 02142, USA. <sup>4</sup>Centro Hispanoluso de Investigaciones Agrarias (CIALE), Department of Microbiology and Genetics, University of Salamanca, Calle Del Duero, 12, Villamayor 37185, Spain. <sup>5</sup>División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica, Camino a la Presa San José, No. 2055, Colonia Lomas 4a Sección, San Luis Potosí, SLP., 78216, México. <sup>6</sup>Department of Biology, Technion - Israel Institute of Technology, Neve Shaanan Campus, Technion City, Haifa, 32000, Israel. <sup>7</sup>Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Mumbai 400085, India. <sup>8</sup>Department of Microbiology, Faculty of Science and Informatics, University of Szeged, Közép fasor 52, Szeged, H-6726, Hungary. <sup>9</sup>DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. <sup>10</sup>School of Biological Sciences, University of Missouri- Kansas City, 5007 Rockhill Road, Kansas City, MO 64110, USA. <sup>11</sup>Institut de Biologie de l'École normale supérieure (IBENS), Institut National de la Santé et de la Recherche Médicale U1024, Centre National de la Recherche Scientifique UMR8197, 46, rue d'Ulm, Paris 75005, France. <sup>12</sup>Chemistry and Biomolecular Sciences, Macquarie University, Research Park Drive Building F7B, North Ryde, Sydney, NSW 2109, Australia. <sup>13</sup>TU Berlin, Institut für Chemie, FG Biochemie und Molekulare Biologie OE2, Franklinstr. 29, 10587 Berlin, Germany. <sup>14</sup>Department of Plant Pathology and Microbiology Building 0444, Nagle Street, Texas A&M University College Station, TX 77843, USA. <sup>15</sup>Department of Biochemical Engineering, Faculty of Science and Technology, University of Debrecen, Egyetem tér 1, Debrecen, H-4010, Hungary. <sup>16</sup>Instituto de Agroquímica y Tecnología de Alimentos, Consejo Superior de Investigaciones Científicas, Apartado de Correos 73, Burjassot (Valencia) E-46100, Spain. <sup>17</sup>Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS, Université de la Méditerranée, Case 932, 163 Avenue de Luminy, 13288 Marseille 13288, France. <sup>18</sup>Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Lautrupvang 15, DK-2750 Ballerup, Denmark. <sup>19</sup>Biotechnology Department, IFP Energies nouvelles, 1-4 avenue de Bois Préau, Rueil-Malmaison, 92852, France. <sup>20</sup>Institute of Sciences of Food Production (ISPA), National Research Council (CNR), Via Amendola 122/O, 70126 Bari, Italy. <sup>21</sup>Wageningen University, Systems and Synthetic Biology, Fungal Systems Biology Group, Dreijenplein 10, 6703 HB Wageningen, The Netherlands. <sup>22</sup>Chemical and Biological Process Development Group, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352, USA.

## Authors' contributions

CPK, IVG, BH, EM, SEB, CMK, and AHE contributed equally to this work as senior authors. AA, JC, MM, AS, and IVG performed global annotation and analysis, MZ and HS did the assembly, OC and CH finished the assembly, and EL and SL performed the genome and EST sequencing. SEB, AH-E, CMK and CPK designed the study, and coordinated and supervised the analysis; CPK drafted and submitted the paper. All other authors contributed research (annotations and/or analyses). All authors read and approved the final manuscript.

Received: 31 December 2010 Revised: 28 March 2011

Accepted: 18 April 2011 Published: 18 April 2011

## References

1. Taylor TN, Hass H, Kerp H, Krings M, Hanlin RT: **Perithecial ascomycetes from the 400 million year old Rhynie chert: an example of ancestral polymorphism.** *Mycologia* 2005, **97**:269-285.
2. Vincent C, Goettel MS, Lazarovits G: *Biological Control: A Global Perspective: Case Studies from Around the World* Wallingford, UK: CAB International; 2007.
3. Harman GE, Howell CR, Viterbo A, Chet I, Lorito M: **Trichoderma species-opportunistic, avirulent plant symbionts.** *Nat Rev Microbiol* 2004, **2**:43-56.
4. Howell CR: **Mechanisms employed by Trichoderma species in the biological control of plant diseases: the history and evolution of current concepts.** *Plant Disease* 2003, **87**:4-10.
5. Harman GE: **Overview of mechanisms and uses of Trichoderma spp.** *Phytopathology* 2006, **96**:190-194.
6. Shores M, Harman GE, Mastouri F: **Induced systemic resistance and plant responses to fungal biocontrol agents.** *Annu Rev Phytopathol* 2010, **48**:21-43.
7. Lorito M, Woo SL, Harman GE, Monte E: **Translational research on Trichoderma: from 'omics to the field.** *Annu Rev Phytopathol* 2010, **48**:395-417.
8. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, et al: **Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina).** *Nat Biotechnol* 2008, **26**:553-560.
9. Druzhinina IS, Kopchinskiy A, Kubicek CP: **The first one hundred Trichoderma species characterized by molecular data.** *Mycoscience* 2006, **47**:55-64.
10. Chaverri P, Samuels GJ, Stewart EL: **Hypocrea virens sp. nov., the teleomorph of Trichoderma virens.** *Mycologia* 2001, **93**:1113-1124.
11. Dodd SL, Lieckfeldt E, Samuels GJ: **Hypocrea atroviridis sp. nov., the teleomorph of Trichoderma atroviride.** *Mycologia* 2003, **95**:27-40.
12. Druzhinina IS, Kubicek CP, Komoń-Zelazowska M, Mulaw TB, Bissett J: **The Trichoderma harzianum demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages.** *BMC Evol Biol* 2010, **10**:94.
13. Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, Couloux A, Aury JM, Séguens B, Poulain J, Anthouard V, Grossetete S, Khalili H, Coppin E, Déquard-Chablat M, Picard M, Contamine V, Arnaise S, Bourdais A, Berteaux-Lecellier V, Gautheret D, de Vries RP, Battaglia E, Coutinho PM, Danchin EG, Henrissat B, Khoury RE, Sainsard-Chanet A, Boivin A, Pinan-Lucarré B, et al: **The genome sequence of the model ascomycete fungus Podospora anserina.** *Genome Biol* 2008, **9**:R77.
14. Fischer G, Rocha EP, Brunet F, Vergassola M, Dujon B: **Highly variable rates of genome rearrangements between hemiascomycetous yeast lineages.** *PLoS Genet* 2006, **2**:e32.
15. Galagan JE, Calvo S, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Baştürkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scacciocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, et al: **Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae.** *Nature* 2005, **438**:1105-1115.
16. Seoighe C, Federspiel N, Jones T, Hansen N, Bivolarovic V, Surzycki R, Tamse R, Komp C, Huizar L, Davis RW, Scherer S, Tait E, Shaw DJ, Harris D, Murphy L, Oliver K, Taylor K, Rajandream MA, Barrell BG, Wolfe KH: **Prevalence of small inversions in yeast gene order evolution.** *Proc Natl Acad Sci USA* 2000, **97**:14433-14437.
17. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva JC, Badger JH, Albarraq A, Angiuoli S, Bussey H, Bowyer P, Cotty PJ, Dyer PS, Egan A, Galens K, Fraser-Liggett CM, Haas BJ, Inman JM, Kent R, Lemieux S, Malavazi I, Orvis J, Roemer T, Ronning CM, Sundaram JP, Sutton G, Turner G, Venter JC, et al: **Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus.** *PLoS Genet* 2008, **4**:e1000046.
18. Nadeau J, Taylor B: **Lengths of chromosomal segments conserved since divergence of man and mouse.** *Proc Natl Acad Sci USA* 1984, **81**:814-818.
19. Edgar RC, Myers EW: **PILER: identification and classification of genomic repeats.** *Bioinformatics* 2005, **21**(Suppl 1):i152-i158.
20. Benson G: **Tandem repeats finder: a program to analyze DNA sequences.** *Nucleic Acids Res* 1999, **27**:573-580.
21. **RepeatMasker Open-3.0.** [http://www.repeatmasker.org].
22. Margolin BS, Garrett-Engle PW, Stevens JN, Fritz DY, Garrett-Engle C, Metzberg RL, Selker EU: **A methylated Neurospora 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation.** *Genetics* 1998, **149**:1787-1797.
23. Selker EU, Tountas NA, Cross SH, Margolin BS, Murphy JG, Bird AP, Freitag M: **The methylated component of the Neurospora crassa genome.** *Nature* 2003, **422**:893-897.
24. Enright AJ, van Dongen S, Ouzounis CA: **An efficient algorithm for large-scale detection of protein families.** *Nucleic Acids Res* 2002, **30**:1575-1584.
25. Seidl V: **Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions.** *Fungal Biol Rev* 2008, **22**:36-42.
26. Limon MC, Chacón MR, Mejías R, Delgado-Jarana J, Rincón AM, Codón AC, Benítez T: **Increased antifungal and chitinase specific activities of Trichoderma harzianum CECT 2413 by addition of a cellulose binding domain.** *Appl Microbiol Biotechnol* 2004, **64**:675-682.
27. Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang YL, Decaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Maucci E, Mewes HW, Mitterbauer R, Muehlbauer G, et al: **The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization.** *Science* 2007, **317**:1400-1402.
28. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeier C, Li W, Harding M, Kim S, Lebrun MH, Bohnert H, Coughlan S, Butler J, Calvo S, Ma LJ, et al: **The genome sequence of the rice blast fungus Magnaporthe grisea.** *Nature* 2005, **434**:980-986.
29. Münch A, Stingl L, Jung K, Heermann R: **Photorhabdus luminescens genes induced upon insect infection.** *BMC Genomics* 2008, **9**:229.
30. Wellinger RJ, Sen H: **The DNA structures at the ends of eukaryotic chromosomes.** *Eur J Cancer* 1997, **33**:735-749.
31. Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, Nehrbass U, Wellem TE, Scherf A: **Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum.** *Nature* 2000, **407**:1018-1022.
32. Sharp PM, Li WH: **The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications.** *Nucleic Acids Res* 1987, **15**:1281-1295.
33. Yang Z, Rannala B: **Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method.** *Mol Biol Evol* 1997, **14**:717-724.
34. Seidl V, Seibel C, Kubicek CP, Schmoll M: **Sexual development in the industrial workhorse Trichoderma reesei.** *Proc Natl Acad Sci USA* 2009, **106**:13909-13914.
35. Shiu PK, Metzberg RL: **Meiotic silencing by unpaired DNA: properties, regulation and suppression.** *Genetics* 2002, **161**:1483-1495.
36. Borkovich KA, Alex LA, Yarden O, Freitag M, Turner GE, Read ND, Seiler S, Bell-Pedersen D, Pajetta J, Plesofsky N, Plamann M, Goodrich-Tanrikulu M, Schulte U, Mannhaupt G, Nargang FE, Radford A, Selitrennikoff C, Galagan JE, Dunlap JC, Loros JJ, Catchside D, Inoue H, Aramayo R, Polymenis M, Selker EU, Sachs MS, Marzluf GA, Paulsen I, Davis R, Ebbole DJ, et al: **Lessons from the genome sequence of Neurospora crassa: tracing the path from genomic blueprint to multicellular organism.** *Microbiol Mol Biol Rev* 2004, **68**:1-108.
37. Rossmann AY, Samuels GJ, Rogerson CT, Lowen R: **Genera of Bionectriaceae, Hypocreaceae and Nectriaceae (Hypocreales, Ascomycetes).** *Stud Mycol* 1999, **42**:1-83.
38. Slot JC, Hobbitt DS: **Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study.** *PLoS One* 2007, **2**:e1097.
39. Bennett V, Baines AJ: **Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues.** *Physiol Rev* 2001, **81**:1353-1392.
40. Stimimann CU, Petsalaki E, Russell RB, Müller CW: **WD40 proteins propel cellular networks.** *Trends Biochem Sci* 2010, **35**:565-574.
41. Migheli Q, Balmas V, Komoń-Zelazowska M, Scherm B, Fiori S, Kopchinskiy AG, Kubicek CP, Druzhinina IS: **Soils of a Mediterranean hot spot of biodiversity and endemism (Sardinia, Tyrrhenian Islands) are inhabited by pan-European, invasive species of Hypocrea/Trichoderma.** *Environ Microbiol* 2009, **11**:35-46.

42. Martin F, Aerts A, Ahrén D, Brun A, Danchin EG, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V: **The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis.** *Nature* 2008, **452**:88-92.
43. Seidl V, Song L, Lindquist E, Gruber S, Koptchinskiy A, Zeilinger S, Schmolli M, Martínez P, Sun J, Grigoriev I, Herrera-Estrella A, Baker SE, Kubicek CP: **Transcriptomic response of the mycoparasitic fungus *Trichoderma atroviride* to the presence of a fungal prey.** *BMC Genomics* 2009, **10**:567.
44. de Jonge R, Thomma BP: **Fungal LysM effectors: extinguishers of host immunity?** *Trends Microbiol* 2009, **17**:151-154.
45. Baker LG, Specht CA, Donlin MJ, Lodge JK: **Chitosan, the deacetylated form of chitin, is necessary for cell wall integrity in *Cryptococcus neoformans*.** *Eukaryotic Cell* 2007, **6**:855-862.
46. El Gueddari NE, Rauchhaus U, Moerschbacher BM, Deising HB: **Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi.** *New Phytol* 2002, **156**:103-111.
47. diArk: **a resource for eukaryotic genome research.** [http://www.diark.org/diark/search].
48. Salamov AA, Solovyev W: **Ab initio gene finding in *Drosophila* genomic DNA.** *Genome Res* 2000, **10**:516-522.
49. Birney E, Durbin R: **Using GeneWise in the *Drosophila* annotation experiment.** *Genome Res* 2000, **10**:547-548.
50. Zdobnov EM, Apweiler R: **InterProScan - an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**:847-848.
51. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene Ontology: tool for the unification of biology The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
52. Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Rogozin IB, Smirnov S, Sorokin AV, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA: **A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes.** *Genome Biol* 2004, **5**:R7.
53. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome.** *Nucleic Acids Res* 2004, **32**:D277-D280.
54. ***Trichoderma virens* Gv29-8 v2.0.** [http://www.jgi.doe.gov/Tvirens].
55. ***Trichoderma atroviride* v2.0.** [http://www.jgi.doe.gov/Tatroviride].
56. Castillo-Ramirez S, Gonzalez V: **Factors affecting the concordance between orthologous gene trees and species tree in bacteria.** *BMC Evol Biol* 2008, **8**:300.
57. Moreno-Hagelsieb C, Janga SC: **Operons and the effect of genome redundancy in deciphering functional relationships using phylogenetic profiles.** *Proteins* 2008, **70**:344-352.
58. **MCL - a cluster algorithm for graphs.** [http://mclans.org/mcl].
59. Li L, Stoeckert CJ Jr, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome Res* 2003, **13**:2178-2189.
60. **CAZY - Carbohydrate-Active EnZYmes database.** [http://www.cazy.org].
61. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The carbohydrate-active EnZYmes database (CAZy): an expert resource for glycogenomics.** *Nucleic Acids Res* 2009, **37**:D233-D238.
62. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25**:1451-1452.
63. **JGI genome portal.** [http://genome.jgi-psf.org].
64. Latgé JP: **The cell wall: a carbohydrate armour for the fungal cell.** *Mol Microbiol* 2007, **66**:279-290.
65. de Groot P, Brandt BW, Horiuchi H, Ram AF, de Koster CG, Klis FM: **Comprehensive genomic analysis of cell wall genes in *Aspergillus nidulans*.** *Fungal Genet Biol* 2009, **46**(Suppl 1):S72-S81.
66. Lieckfeldt E, Kullnig CM, Samuels GJ, Kubicek CP: **Sexually competent, sucrose- and nitrate-assimilating strains of *Hypocrea jecorina* (*Trichoderma reesei*, *Hypocreales*) from South American soils.** *Mycologia* 2000, **92**:374-384.
67. Vargas WA, Mandawe JC, Kenerley CM: **Plant-derived sucrose is a key element in the symbiotic association between *Trichoderma virens* and maize plants.** *Plant Physiol* 2009, **151**:792-797.
68. Martens-Uzunova ES, Schaap PJ: **Assessment of the pectin degrading enzyme network of *Aspergillus niger* by functional genomics.** *Fungal Genet Biol* 2009, **46**(Suppl 1):S170-S179.
69. Berne S, Lah L, Sepčić K: **Aegerolysins: structure, function, and putative biological role.** *Protein Sci* 2009, **18**:694-706.
70. Goodrich-Blair H, Clarke DJ: **Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: two roads to the same destination.** *Mol Microbiol* 2007, **64**:260-268.
71. Hares MC, Hinchliffe SJ, Strong PC, Eleftherianos I, Dowling AJ, French-Constant RH, Waterfield N: **The *Yersinia pseudotuberculosis* and *Yersinia pestis* toxin complex is active against cultured mammalian cells.** *Microbiology* 2008, **154**:3503-3517.
72. Karasova D, Havlickova H, Sisak F, Rychlik I: **Deletion of *sodCI* and *spvBC* in *Salmonella enterica* serovar *Enteritidis* reduced its virulence to the natural virulence of serovars Agona, Hadar and infantis for mice but not for chickens early after infection.** *Vet Microbiol* 2009, **139**:304-309.
73. McNulty C, Thompson J, Barrett B, Lord L, Andersen C, Roberts IS: **The cell surface expression of group 2 capsular polysaccharides in *Escherichia coli*: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell.** *Mol Microbiol* 2006, **59**:907-922.
74. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
75. **SignalP 3.0 Server.** [http://www.cbs.dtu.dk/services/SignalP/].
76. Brotman Y, Briff E, Viterbo A, Chet I: **Role of swollenin, an expansin-like protein from *Trichoderma*, in plant root colonization.** *Plant Physiol* 2008, **147**:779-789.
77. **Max Planck Institute of Developmental Biology: Bioinformatics toolkit.** [http://toolkit.tuebingen.mpg.de/blastclust].
78. Talbot NJ: **Growing into the air.** *Curr Biol* 1997, **7**:R78-R81.
79. Wösten HA, van Wetter MA, Lugones LG, van der Mei HC, Busscher HJ, Wessels JG: **How a fungus escapes the water to grow into the air.** *Curr Biol* 1999, **9**:85-88.
80. Kubicek CP, Baker S, Gamauf C, Kenerley CM, Druzhinina IS: **Purifying selection and birth-and-death evolution in the class II hydrophobin gene families of the ascomycete *Trichoderma/Hypocrea*.** *BMC Evol Biol* 2008, **8**:4.
81. Viterbo A, Chet I: **TasHyd1, a new hydrophobin gene from the biocontrol agent *Trichoderma asperellum*, is involved in plant root colonization.** *Mol Plant Pathol* 2006, **7**:249-258.
82. Djonovic S, Pozo MJ, Dangott LJ, Howell CR, Kenerley CM: **Sm1, a proteinaceous elicitor secreted by the biocontrol fungus *Trichoderma virens* induces plant defense responses and systemic resistance.** *Mol Plant Microbe Interact* 2006, **19**:838-853.
83. Djonovic S, Vargas WA, Kolomiets MV, Horndeski M, Wiest A, Kenerley CM: **A proteinaceous elicitor Sm1 from the beneficial fungus *Trichoderma virens* is required for induced systemic resistance in maize.** *Plant Physiol* 2007, **145**:875-889.
84. Rep M: **Small proteins of plant-pathogenic fungi secreted during host colonization.** *FEMS Microbiol Lett* 2005, **253**:19-27.
85. Mukherjee PK, Hadar R, Pardovitz-Kedmi E, Trushina N, Horwitz BA: **MRSF1, encoding a novel *Trichoderma* secreted protein, is negatively regulated by MAPK.** *Biochem Biophys Res Commun* 2006, **350**:716-722.
86. Armaleo D, Gross SR: **Structural studies on *Neurospora* RNA polymerases and associated proteins.** *J Biol Chem* 1985, **260**:16174-16180.
87. **Vogels 50x salts.** [http://www.fgsc.net/methods/vogels.html].
88. Jones JDG, Dunsmuir P, Bedbrook J: **High-level expression of introduced chimaeric genes in regenerated transformed plants.** *EMBO J* 1985, **4**:2411-2418.
89. Berrocal-Tito G, Sametz-Baron L, Eichenberg K, Horwitz BA, Herrera-Estrella A: **Rapid blue light regulation of a *Trichoderma harzianum* photolyase gene.** *J Biol Chem* 1999, **274**:14288-14294.
90. Dettler JC, Jett JM, Lucas SM, Dalin E, Arellano AR, Wang M, Nelson JR, Chapman J, Lou Y, Rokhsar D, Hawkins TL, Richardson PM: **Isothermal strand displacement amplification applications for high-throughput genomics.** *Genomics* 2002, **80**:691-698.
91. Huang X, Madan A: **CAP3: A DNA Sequence Assembling Program.** *Genome Res* 1999, **9**:868-877.
92. Papadopoulos JS, Agarwala R: **COBALT: constraint-based alignment tool for multiple protein sequences.** *Bioinformatics* 2007, **23**:1073-1079.

doi:10.1186/gb-2011-12-4-r40

**Cite this article as:** Kubicek et al.: Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biology* 2011 **12**:R40.