



Centro de Investigación Científica de Yucatán, A.C.

Posgrado en Ciencias Biológicas

Construcción de un sistema de expresión superficial de
proteínas en *Trichoderma harzianum*

Tesis que presenta

Miguel Angel Canseco Pérez

En opción al título de

DOCTOR EN CIENCIAS

(Ciencias Biológicas: Opción Biotecnología)

Mérida, Yucatán, México

2018

CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A. C.
POSGRADO EN CIENCIAS BIOLÓGICAS



Por medio de la presente hago constar que el trabajo de tesis de **Miguel Angel Canseco Pérez**, titulado “**Construcción de un sistema de expresión superficial de proteínas en Trichoderma harzianum**”, fue realizado en el Laboratorio de Biotecnología Microbiana de la Unidad de Biotecnología, del Centro de Investigación Científica de Yucatán, A.C., en la línea de Biotecnología de Microorganismos, bajo la dirección de la Dra. Blondy Beatriz Canto Canché y del Dr. Lorenzo Felipe Sánchez Teyer, dentro de la opción de Biotecnología, perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro.

Esta tesis tiene orientación al desarrollo socioeconómico de la región porque está orientada hacia la economía basada en conocimiento.

Atentamente:

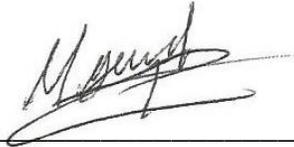


Dr. Manuel Martínez Estévez
Director de Docencia

Mérida, Yucatán, México, a 26 de junio de 2018.

DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en la sección de Materiales y Métodos Experimentales, los Resultados y Discusión de este documento proviene de las actividades de experimentación realizadas durante el período que se me asignó para desarrollar mi trabajo de tesis, en las Unidades y Laboratorios del Centro de Investigación Científica de Yucatán, A.C., y que a razón de lo anterior y en contraprestación de los servicios educativos o de apoyo que me fueron brindados, dicha información, en términos de la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, le pertenece patrimonialmente a dicho Centro de Investigación. Por otra parte, en virtud de lo ya manifestado, reconozco que de igual manera los productos intelectuales o desarrollos tecnológicos que deriven o pudieran derivar de lo correspondiente a dicha información, le pertenecen patrimonialmente al Centro de Investigación Científica de Yucatán, A.C., y en el mismo tenor, reconozco que si derivaren de este trabajo productos intelectuales o desarrollos tecnológicos, en lo especial, estos se registrarán en todo caso por lo dispuesto por la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.



Miguel Angel Canseco Pérez

Este trabajo se llevó a cabo en la Unidad de Biotecnología del Centro de Investigación Científica de Yucatán, y forma parte del proyecto titulado Construcción de un sistema de expresión superficial de proteínas en *Trichoderma harzianum* en el que participé bajo la dirección de los Dres. Blondy Beatriz Canto Canché y Lorenzo Felipe Sánchez Teyer.

AGRADECIMIENTOS

A CONACYT, por la beca otorgada número 242995.

Se reconoce el apoyo de los proyectos CONACYT 220957 y 269833 y FOMIX 247355.

Al CICY, en especial a las unidades de Biotecnología y Bioquímica y Biología Molecular de Plantas, por permitirme usar sus instalaciones.

A la M. C. Ángela Francisca Kú González, por su apoyo en la captura de imágenes en el microscopio confocal.

Al M. C. Bartolomé Chi Manzanero, por su apoyo técnico durante el desarrollo del proyecto.

Al M. C Miguel Alonso Tzec Síma, por su colaboración en el establecimiento de los protocolos de transformación fúngica, toma de fotografías y su apoyo invaluable durante el desarrollo del proyecto. Así como un agradecimiento especial por haberme enseñado que la humildad, paciencia y sabiduría van de la mano con un gran corazón.

A los doctores: Max Apolinar Hernández, Ignacio Islas Flores, Yeny Lizzet Couoh Uicab, Marcela Gamboa Angulo, Lorenzo Felipe Sánchez Teyer y Blondy Canto Canché, por sus aportaciones en la elaboración del manuscrito.

Al comité tutorial conformado por los doctores: Ignacio Islas Flores, Felipe Sánchez Teyer, Yenny Couoh Uicab y Blondy Canto Canché, por sus consejos y aportaciones.

Al comité de revisión de tesis conformado por los doctores: Max Apolinar Hernández, Ignacio Islas Flores, Yeny Lizzet Couh Uicab, Marcela Gamboa Angulo, Aileen O'Connor, Sánchez, Lorenzo Felipe Sánchez Teyer, Blondy Canto Canché, por enriquecer el documento con sus observaciones.

A la Dra. Blondy Canto Canché, por todos los consejos, paciencia y apoyo dado durante el desarrollo del presente proyecto.

A todos los amigos que me brindaron consejo y palabras de aliento durante el desarrollo del presente proyecto.

DEDICATORIAS

A mi familia: mi Madre, Sra. Beatriz Pérez Luna, mi Padre Sr. Miguel Canseco Argüello, mi hermano José Antonio Canseco Pérez y mis sobrinas Juliette y Miranda Canseco Mendoza, por ser mi soporte, motivo y apoyo durante este trayecto.

A mis tíos Yolanda Pérez Luna y Esaú de Jesús Pérez Luna, por ser mi modelo por seguir en mi formación académica.

ÍNDICE

INTRODUCCIÓN.....	1
CAPÍTULO I	3
1.1 ANTECEDENTES.....	3
1.2 JUSTIFICACIÓN.....	20
1.3 HIPÓTESIS.....	20
1.4 OBJETIVOS	21
1.5 ESTRATEGIA EXPERIMENTAL	22
CAPÍTULO II	23
IDENTIFICACIÓN <i>IN SILICO</i> DE GENES QUE CODIFICAN PARA PROTEÍNAS CON ANCLAJE A GPI EN EL GENOMA DE <i>T. REESEI</i> Y <i>T. LONGIBRACHIATUM</i>	23
2.1 INTRODUCCIÓN	23
2.2 MATERIALES Y MÉTODOS.....	26
2.2.1 Generación de base de datos con proteínas con anclaje GPI previamente reportada en hongos y búsqueda de homólogas en el genoma de <i>T. reesei</i>	26
2.2.2 Búsqueda de proteínas con anclaje GPI previamente anotadas en el genoma de <i>T. reesei</i>	26
2.2.3 Análisis bioinformático para la predicción de proteínas con anclaje GPI	26
2.2.4 Análisis <i>in silico</i> para la identificación de funciones putativas y puntos isoeléctricos.....	27
2.2.5 Identificación molecular de cepas de <i>Trichoderma</i> sp	27
2.2.6 Identificación de posible proteínas con anclaje GPI en <i>T. longibrachiatum</i> homólogas a las identificadas en <i>T. reesei</i>	28
2.2.7 Diseño de los cebadores para el aislamiento de proteínas con anclaje GPI en <i>T. longibrachiatum</i>	28
2.2.8 Aislamiento de ARN total y síntesis de ADNc de <i>T. longibrachiatum</i>	29

2.2.9 Amplificación de la región codificante de las proteínas con anclaje GPI	29
2.3 RESULTADOS	30
2.3.1 Identificación <i>in silico</i> de proteínas con anclaje GPI.....	30
2.3.2 Identificación <i>in silico</i> de las posibles funciones de las proteínas predichas con anclaje GPI con base en sus dominios y proteínas homólogas caracterizadas	32
2.3.2 Identificación molecular de cepas de <i>Trichoderma</i> sp	40
2.3.3 Identificación de posibles proteínas con anclaje GPI en <i>T. longibrachiatum</i> , homólogas a las identificadas en <i>T. reesei</i>	41
2.3.4 Diseño de cebadores para el aislamiento de proteínas con anclaje GPI en <i>T. longibrachiatum</i>	42
2.2.7 Aislamiento de ARN total de <i>T. longibrachiatum</i>	43
2.3.6 Amplificación de la región codificante de las proteínas con anclaje GPI	45
2.4 DISCUSIÓN	48
CAPÍTULO III	53
DISEÑO Y CONSTRUCCIÓN DE UN VECTOR PARA LA EXPRESIÓN SUPERFICIAL DE GFP FUSIONADA A PROTEÍNAS CON ANCLAJE GPI.....	53
3.1 INTRODUCCIÓN	53
3.2 MATERIALES Y MÉTODOS.....	58
3.2.1 Diseño de la construcción del vector pAN7.1-GFP-GPI.....	58
3.2.2 Amplificación de secuencias para la introducción de los sitios de restricción	59
3.2.3 Amplificación de las secuencias mediante los cebadores con los adaptadores para el sistema Gibson®.....	59
3.2.4 Ligación múltiple de fragmentos	60
3.2.5 Transformación y confirmación de las transformantes.....	60
3.2.7 Aislamiento de los plásmidos eGFP, pAN7.1 y pAN-GFP-GPI.....	60
3.2.6 Generación de protoplastos para transformación	60
3.2.7 Transformación por electroporación.....	61
3.2.8 Transformación por el Método de PEG y cloruro de calcio	61
3.2.9 Transformación por el protocolo de perlas de cristal	61
3.2.10 Confirmación de las transformantes.....	62

3.2.11 Visualización de las muestras en el microscopio confocal	62
3.2.12 Segunda estrategia para la construcción del Vector usando como plataforma el casete <i>HPH</i> del vector pSilent-1	62
3.2.13 Confirmación de transformantes y perfil de restricción del vector pHyg-GFP-GPI.....	63
3.3.14 Construcción del vector de expresión utilizando enzimas de restricción	64
3.3 RESULTADOS	65
3.3.1 Análisis <i>in silico</i> del constructo del vector de expresión pAN7.1-GFP-GPI	65
3.3.2 Aislamiento y amplificación de las secuencias del constructo	66
3.3.3 Amplificación con los adaptadores diseñados para la metodología de ligación mediante el sistema Gibson®.....	67
3.3.4 Ligación múltiple con el sistema Gibson®.....	70
3.3.5 Transformación y Confirmación de los transformantes	71
3.3.6 Confirmación del constructo pAN7.1-GFP-GPI por patrones de digestión y PCR	73
3.3.7 Establecimiento del protocolo de transformación utilizando el plásmido GFP... 79	
4.3.3 Visualización de la muestras por microscopía.....	82
3.3.7 Ligación múltiple mediante el sistema Gibson® utilizando el vector pSilent-1 como receptor del casete de expresión Ss- GFP-GPI	85
3.3.8 Confirmación de los sitios de inserción en el plásmido pHyg-GFP-GPI.....	87
3.3.9 Construcción del vector de expresión utilizando enzimas de restricción en pSilent-1.....	89
3.4 DISCUSIÓN	92
CAPITULO IV	95
FUNGAL SCREENING ON OLIVE OIL FOR EXTRACELLULAR TRIACYLGLYCEROL LIPASES: SELECTION OF A <i>TRICHODERMA HARZIANUM</i> STRAIN AND GENOME WIDE SEARCH FOR THE GENES	95
4.1 INTRODUCTION.....	97
4.2 MATERIALS AND METHODS	99
4.2.1 Searching of Genes for True Lipases in <i>T. harzianum</i> Genome and <i>In Silico</i> Analysis of the Proteins	99

4.2.2 <i>In Silico</i> Analysis <i>T. harzianum</i> Putative True Lipases.....	100
4.2.3 Phylogenetic Analysis of True Lipases from <i>T. harzianum</i>	100
4.2.4 <i>In Silico</i> Modeling.....	101
4.2.5 RT-PCR Amplification of Lipases.....	102
4.3. RESULTS AND DISCUSSION	103
4.3.1 Searching Genes of True Lipases in <i>T. harzianum</i>	103
4.3.2 <i>In Silico</i> Characterization of <i>T. harzianum</i> True Lipases.....	103
4.3.3 Phylogenetic Analysis	114
4.3.4 Structural Modeling	119
4.3.6 Refining of Selection of Candidates.....	123
4.3.7 RT-PCR Analysis for Validation of Predicting Results.....	123
4.4 CONCLUSIONS	127
CAPÍTULO V	129
5.1 DISCUSIÓN GENERAL	129
CAPÍTULO VI.....	138
CONCLUSIONES Y PERSPECTIVAS	138
6.1 CONCLUSIONES	138
6.2 PERSPECTIVAS	139
BIBLIOGRAFÍA.....	140
ANEXO.....	160

LISTADO DE FIGURAS

Figura 1.1 Representación esquemática de los componentes mayoritarios de la pared celular (Arana et al., 2009).....	4
Figura 1.2 Estructura química de la quitina (Bueter et al., 2013).....	7
Figura 1.3 Características en la secuencias de las proteínas Pir: a) Representación esquemática de los dominios conservados en las proteínas Pir, b) Alineamiento de proteínas tipo Pir de origen fúngico. Se muestra la parte más conservada, iniciando después de la primera cisteína en el dominio característico de cisteínas (Kantún-Moreno et al., 2013; De Groot et al., 2005).....	10
Figura 1.4 Esquematzación de los sitios característicos de las proteínas con anclaje GPI (Klis et al., 2011).....	11
Figura 1.5 Estructura química de la molécula de glicofosfatidilinositol. En gris glicoproteína, ω-sitio de anclaje a GPI, Celeste: manosa, verde: inositol, morado: glucosamina (Modificado de: Vink et al., 2004).....	12
Figura 1.6 Esquematzación de la expresión en superficie celular mediante dos tipos de anclaje: N-terminal y C-terminal. MP, membrana plasmática; PC, pared celular (Basado en Tanaka et al., 2012; Ueda y Tanaka, 2000).	16
Figura 1.7 Visualización por microscopía de fluorescencia de <i>Aspergillus oryzae</i> y cuantificación de actividad de β-glucosidasa desplegada en la pared. a) Cepa recombinante produciendo proteína verde fluorescente (1, 2 y 3). Cepa recombinante expresando fusión de proteína GFP-AOMP1 (4, 5 y 6). b) Actividad enzimática de β-glucosidasa. Círculos negros: cepa expresando la proteína de fusión BGLI-MP1. Cuadrados: Cepa expresando proteína de fusión BGLI-CWP. Rombo: <i>A. oryzae</i> (Adachi et al., 2008).....	18
Figura 2.1 Clasificación de las proteínas de <i>T. reesei</i> con base en los resultados de análisis con los predictores GPI-som, predGPI, biGPI fungal predictor, SignalP, Wolfpsort y Protparam tool. En azul: secuencias negativas (93.4%); En café: (+<30)= 4.9%; En Morado: (+-) = 1% y en Verde: (+>30)= 0.7%.	31

Figura 2.2 Visualización en gel de agarosa del producto de amplificación de los ITS de las 6 cepas de presuntos *Trichoderma* sp. con las que se inició el presente trabajo. a) Cepa B13-1, b) B13-3, c) B13-4, d) AR2-(1), e) Ar3-(1), f) AR7-(1). M) Marcador molecular de 1 Kb plus ladder (ThermoFisher Scientific). 40

Figura 2.3 a) Identificación mediante el programa SignalP de la secuencia correspondiente a la señal de secreción; b) Esquematización de los sitios en los cuales se diseñaron los cebadores para el aislamiento de las secuencias; Flechas rojas: ubicación en la secuencia de los cebadores para el aislamiento de la secuencia señal de secreción. En azul: cebadores para el aislamiento del fragmento codificante para la proteína con anclaje GPI, ω : sitio de anclaje del glicosilfosfatidilinositol (GPI). 43

Figura 3.1 Representación esquemática de los componentes típicos de la unidad de expresión (Tomado de Berka y Barnett, 1989). 54

Figura 3.2 Esquema del método por PCR después de la ligación para la eficiente inserción de múltiples fragmentos de ADN (tomado de An *et al.*, 2010). 55

Figura 3.3 Esquematización del ensamble *in vitro* de múltiples fragmentos mediante la reacción de Gibson® (<https://www.neb.com/products/e5510-gibson-assembly-cloning-kit>). 56

Figura 3.4 Representación circular del vector pAN7.1 con sus principales componentes: promotor *GPD*, terminador *TRPC*, agente de selección higromicina B fosfotransferasa (Hph) y ampicilina (Amp) (Punt *et al.*, 1987). 58

Figura 3.5 Análisis *in silico* del vector pAN7.1-GFP-GPI: a) análisis de marcos de lectura abierto, en donde se observa el marco de lectura esperado en la región comprendida entre los 5.5 Kb y 7Kb que corresponde al constructo de la proteína quimérica Ss-GFP-GPI; b) Esquema de la ubicación de los insertos dentro del vector pAN7.1 modificado. .. 65

Figura 3.6 Análisis *in silico* de los sitios de restricción para el del vector pAN7.1-GFP-GPI, subrayado en rojo los sitios que flanquean el marco de lectura abierto (ORF) que codifica para la proteína quimérica Ss-GFP-GPI, en 5' el sitio *KpnI* y en 3' el sitio *ApaI*. 66

Figura 3.7 Visualización en gel de agarosa al 1% de los amplicones que conforman la

construcción de expresión. a): amplicón correspondiente a la secuencia señal de secreción (Ss) y la secuencia que codifica para la proteína con anclaje GPI reportada por Adachi y colaboradores (2008); b) control negativo; c): amplicón correspondiente al terminador *TRPC* aislado del plásmido pAN7.1 (769 pb); d): amplicón correspondiente al promotor del gen *GPD* de *Aspergillus nidulans* aislado del vector pAN7.1 (2.1 Kb); e): amplicón correspondiente a la región codificante de la proteína verde fluorescente reportada por Maor y colaboradores (1998) del vector peGFP (723 pb); M: marcador de peso molecular 1 Kb plus DNA Ladder (ThermoFisher Scientific)..... 66

Figura 3.8 Amplificación de las secuencias con los cebadores que introducen las secuencias adaptadoras para la ligación con el sistema Gibson®. A) Amplicón del Terminador *TRPC* con los adaptadores Gibson, b) Secuencia señal de secreción con los adaptadores Gibson c) Secuencia codificante para proteína verde fluorescente con los adaptadores Gibson 67

Figura 3.9 Amplificación de la secuencia AOMP1 con los cebadores que introducen los adaptadores para el sistema Gibson®. a) Amplificación usando como templado producto de purificación y el par de cebadores, b) Producto de amplificación usando PCR touchdown, c) amplificación usando como templado el fragmento clonado en TOPO-TA.68

Figura 3.10 Amplificación de la secuencia del promotor *GPD* a) amplificación con el primer par de cebadores que introducen los sitios de restricción en la secuencia. De b) hasta f) corresponden a la amplificación con los cebadores que introducen los adaptadores para la ligación con Gibson®. Se utilizó 10 µM de cada cebador para cada reacción en una PCR de 35 ciclos. Usando las siguientes temperatura de alineamiento (tm) b) 46°C, c) tm 47°C, d) tm 48°C, e) tm 49°C, f) se usó 2.5 µM de cada cebador, tm 47°C y se dieron 40 ciclos de PCR. M: Marcador de peso molecular 1 Kb plus Ladder (ThermoFisher Scientific). La figura g) corresponde a la esquematización de la introducción de los sitios de restricción y las secuencias adaptadoras para el sistema Gibson®, ambos introducidos mediante cebadores de PCR..... 69

Figura 3.11 Representación esquemática del vector pAN7.1 y visualización de la digestión con XbaI a) Esquema con los principales sitios de restricción del vector pAN7-1. b) M: Marcador de peso molecular 1 Kb plus Ladder (ThermoFisher Scientific) ; Carril 1)

pAN7.1 sin digerir; carril 2) pAN7.1 digerido con XbaI. 70

Figura 3.12 Visualización en gel de agarosa de la ligación múltiple con el sistema Gibson®: a) ligación múltiple Gibson® de pAN7.1, GFP y GPI y b) Vector pAn7.1 sin digerir en sus tres conformaciones, M marcador de peso molecular GeneRuler™ 1 Kb plus DNA Ladder (ThermoFisher Scientific). 71

Figura 3.13 Amplificación por PCR en colonia del fragmento codificante para la proteína GFP para verificar transformación con constructo pAN7.1-GFP-GPI en las clonas transformadas; C-: control negativo, C+: amplicón (700 pb) obtenido usando como templado peGFP como control positivo para corroborar la PCR. 72

Figura 3.14 Amplificación por PCR en colonia del fragmento de la región codificante que abarca la secuencia señal de Secreción-GFP-GPI: de a-k): clonas con el posible vector pAN7.1-GFP-GPI. La flecha señala el producto esperado. 73

Figura 3.15 Digestión del plásmido pAN7.1-GFP-GPI con tres diferentes enzimas de restricción: a) Digestión con *Apal*; b) Digestión con *NotI*; c) plásmido sin digerir; d) digestión con *KpnI*; e) plásmido sin digerir y M: Marcador de peso molecular Gene ruler 1 Kb plus DNA Ladder (ThermoFisher Scientific). 74

Figura 3.16 Doble digestión del constructo pAN7.1-GFP-GPI; a) Digestión con las enzimas *NotI* y *KpnI*, b) Digestión con *NotI* y *SpeI*, c) Digestión con *SpeI* y *Apal*, d) plásmido sin digerir y M: marcador de peso molecular 1 Kb plus Ladder (ThermoFisher Scientific). 75

Figura 3.17 Diseño y amplificación de cebadores en sitios internos de la región codificante quimérica Ss-GFP-GPI para la confirmación del orden de los fragmentos de ADN y su secuenciación; a) Esquemmatización de los sitios en donde se diseñaron los, en rojo la ubicación de los cebadores para la amplificación del fragmento Pgfp, en azul para el fragmento gfpghi y en naranja gpiter; b) Amplicóns obtenidos y visualizados en gel de agarosa al 1%. 76

Figura 3.18 Digestión con *EcoRI* de los vectores pAN7.1 y pAN7.1-GFP-GPI. Digestión con *EcoRI*: en a) pAN7.1-GFP-GPI Clona 2 (Ve2), en b) pAN7.1-GFP-GPI clona3 (Ve3),

en e) pAN7.1; plásmidos sin digerir en c) Ve2, d) Ve3 y f) pAN7.1 y M: marcador de peso molecular Gene ruler 1 Kb plus Ladder (ThermoFisher Scientific) 77

Figura 3.19 Amplificación de la secuencia ORF (Ss-GFP-GPI) usando como templado el vector pAN7.1-GFP-GPI: a) Se señala la ubicación de los cebadores utilizados para la amplificación. b) Visualización en gel de agarosa al 1% de los amplicones generados a diferentes temperaturas: 1) 50, 2) 52 3) 56, 4) 58, 5) 60 °C y M: marcador de peso molecular 1 Kb plus ladder (ThermoFisher Scientific). 78

Figura 3.20 Morfología de las cepas de *T. harzianum* transformada con diferentes plásmidos al 4 día post inoculación: a) pEGFP, b) pAN7.1 c) pAN7.1-GFP-GPI..... 81

Figura 3.21 Visualización en gel de agarosa de la amplificación de la secuencia que codifica para el gen de *Hph* a partir de ADN obtenido de tres transformantes de *T. harzianum*: En a) y en b) carril 1 corresponde al marcador de peso molecular Gene ruler 1 Kb plus (ThermmpFisherScientigic); a) peGFP; carril 2 y 3 transformantes, carril 4 control negativo, b) en carril 2 y 3 pAN7.1-GFP-GPI, en 4 y 5 pAN7.1. 82

Figura 3.22 Visualización mediante microscopía de fluorescencia (40X) de cepas transformadas: 1) *T. harzianum* sin transformar a) campo claro, b) campo oscuro; 2) *T. harzianum* transformada con pAN7.1, c) campo claro, d) campo oscuro, 3) *T. harzianum* transformada con peGFP, e) campo claro, f) campo oscuro, 4) *T. harzianum* transformada con pAN7.1-GFP-GPI, g) campo claro, h) campo oscuro..... 83

Figura 3.23 Visualización por microscopía confocal de los diferentes transformantes de *T. harzianum* observado a 60X: arriba: campo normal, medio: campo de fluorescencia, abajo: contraste de fases, a) *T. harzianum* transformado con pAN7.1, b) *T. harzianum* transformado con pAN7.1-GFP-GPI, c) *T. harzianum* transformado con peGFP. 84

Figura 3.24 Visualización en gel de agarosa de la digestión por *Xba*I de pSILENT1. a) Digestión, b) pSILENT1 86

Figura 3.25 Confirmación por PCR en colonia del fragmento que codifica para GFP (752 pb) utilizando como templado las clonas transformadas con el plásmido pHyg-GFP-GPI; Carril 1) Marcador de peso molecular, a-i) clonas 1, 20, 25, 30, 40, 45, 50, 55 y 63

respectivamente, j) control negativo, k) control positivo utilizando como templado el plásmido peGFP. 87

Figura 3.26 Perfil de digestión del plásmido pHyg-Gfp-GPI; Carril 1 marcador de peso molecular, P: plásmido sin digerir, En A) carril 2 digestión con *Apal*, carril 3 digestión con *SpeI*, carril 4 digestión con *XbaI*; En B) Carril 2; doble digestión con *XbaI-SpeI*, carril 3 digestión con *SpeI*; En C) carril 4; doble digestión con las enzimas *KpnI-Apal*. 88

Figura 3.27 Análisis en gel de agarosa al 1% del constructo pCR®2.1TOPO-ORF1 (ORF1 corresponde al casete P_{gpd} -Ss-GFP-GPI- Ter_{trpC} aislado del primer constructo en pAN7.1 y pCR®2.1TOPO-*HPH* (*HPH* es un fragmento de pSilent que contiene el gen que codifica la resistencia a higromicina). En a) carril 1: marcador de peso molecular, carril 2: digestión del plásmido pCR®2.1TOPO-*HPH* con las enzimas de restricción *HindIII* y *Apal*; se observan las bandas esperadas de 1.2 Kb correspondiente al fragmento del gen *HPH* subclonado en el plásmido y 4 Kb correspondiente al plásmido sin inserto, En b) carril 1) y 3), perfil de restricción con *HindIII* de pCR®2.1TOPO-ORF1 clona 2 y 10, 2) y 4) plásmidos pCR®2.1TOPO-ORF1 clona 2 y 10 sin digerir respectivamente, en 5) perfil de restricción de pCR®2.1TOPO con *HindIII* y 6) plásmido sin digerir pCR®2.1TOPO. En c) carril 2 perfil de digestión con las enzimas de restricción *HindIII* y *Apal* de pCR®2.1TOPO-ORF1 de la clona 2 y 10 respectivamente. 90

Figura 3.28 Esquematación del constructo pSmORF1 y su análisis de digestión; a) Esquema general del vector con sus respectivos sitios de restricción; b) Carril 1, Marcador de peso molecular, P: plásmido sin digerir, *KpnI-Apal*: Doble digestión. 91

Figure 4.3.1 Phylogenetic tree of fungal lipases. The tree was constructed with 35 characterized fungal lipases (accessions correspond to GenBank unless another source is specified) and 15 putative triacylglycerol lipases from *T. harzianum*. (*) After the accession numbers are lipases which have been characterized; unlabeled proteins correspond to hypothetical, uncharacterized, predicted lipases. The tree was generated by MAFFT software using the neighbor-joining method (Saitou *et al.*, 1987) with 500 bootstrap resamplings. Clusters IA, IB and II as described by Yadav *et al.* (2012). Highlighted in purple letters, the *Rhizomocur miehei* lipase-like group and in olive green letters, the *Candida rugosa* lipase-like group, according Schmidt-Dannert (1999). Clade I and Clade II, are

consistent with Gupta *et al.* (2015). *T. harzianum* triacylglycerol lipases from this study, highlighted in bold red letters. 117

Figure 4.3.2 RT-PCR analysis of selected putative extracellular TAG lipases from *T. harzianum* in medium without (**B**) or with 1% (v/v) olive oil (**C**) as carbon source. The number of ID at genome portal of each candidate lipase corresponds to: **(1)** 77338; **(2)** 78181; **(3)** 514252; **(4)** 526309; **(5)** 514427; and **(6)** 87496. Lane **(7)** Elongation factor 1, as positive control of PCR. Panel **(A)** corresponds to PCR on genomic DNA, to test the primers 124

Figure 4.3.3 Structural-based multi-alignment of *T. harzianum* 526309, 551811, and 87496 with 2VEO lipase from *C. antarctica* (CALA). Blue line highlights the pentapeptide; orange line, the lid domain; green triangles, the catalytic triad; and yellow triangles, the oxyanion. Alpha helices and beta sheets are indicated at the top. Identical (bold white letters on red background) and similar amino acids (standard red letters) are shown. ... 126

LISTADO DE CUADROS

Cuadro 1.1 Componentes de la pared celular en diferentes hongos (Free, 2013).....	5
Cuadro 1.2 Expresión en la superficie celular en distintos hongos (Tanaka <i>et al.</i> , 2012; Adachi <i>et al.</i> , 2008; Kim <i>et al.</i> , 2002)	17
Cuadro 2.1 Principales programas bioinformáticos para la identificación <i>in silico</i> de proteínas con anclaje a GPI.....	24
Cuadro 2.2 Proteínas con anclaje GPI fúngicas previamente predichas en distintos hongos.....	30
Cuadro 2.3 Lista de las posibles proteínas con anclaje GPI identificadas en el proteoma deducido de <i>Trichoderma reesei</i>	33
Cuadro 2.4 Identificación molecular de 6 cepas de hongos utilizadas en el presente trabajo.	41
Cuadro 2.5 Identificación de proteínas de <i>T. longibrachiatum</i> homólogas de proteínas con anclaje GPI de <i>T. reesei</i>	42
Cuadro 3.1 Metodologías utilizadas en el presente trabajo para la transformación de <i>Trichoderma</i> spp.....	79
Table 4.3.1 List of domains and family description of putative true lipases of <i>T. harzianum</i> *.	106
Table 4.3.2 <i>In silico</i> analysis of putative true lipases of <i>T. harzianum</i>	113
Table 4.3.3 <i>In silico</i> localization of characteristic lipase domains in the candidate lipase proteins from <i>T. harzianum</i>	119

ABREVIATURAS

a.a	Aminoácidos
ADN	Ácido desoxirribonucleico
ADNc	ADN complementario
ADNp	ADN plasmídico
ADNg	ADN genómico
AMP	Ampicilina
ARN	Ácido ribonucleico
CALA	<i>Candida antarctica</i> lipasa A
cm²	Centímetro cuadrado
g/L	Gramos/Litro
GFP	Proteína verde Fluorescente
GPI	Glicosil-fosfatidilinositol
GPI-Cwps	Proteínas con anclaje GPI asociadas a la pared celular
GPI-PMPs	Proteínas con anclaje GPI asociadas a la membrana plasmática
h	Horas
Hph	Higromicina B
Kb	Kilobases
kV	Kilovoltios
kDa	kilodaltons
LB	Luria Bertani caldo
M	Molar
mM	milimolar
min	Minutos
mL	Mililitros
MOPS	3-(n-morfolino)ácido propanesulfónico
nt	Nucleótidos
ORF	Open Reading fragment (Marco de lectura abierto)
pb	Pares de bases
PCR	Reacción en cadena de la polimerasa
PDA	Medio de papa, dextrosa y agar
PDB	Caldo de papa y dextrosa
PEG	Polietilenglicol
pH	Potencial de hidrógeno
Pir	Proteínas con repeticiones internas
rpm	Revoluciones por minuto
p/v	Peso/volumen
RT-PCR	Reacción en cadena de la polimerasa con transcriptasa inversa
spp	Especie
Ser/Thr	Serina/Treonina
TAG	Triacilglicerol

T_m	Temperatura de fusión
U/μL	Unidades/microlitro
v/v	Volumen/volumen
°C	Grados Celsius
μM	micromolar
μL	Microlitros
μg	Microgramos
μg/mL	microgramos/mililitros

RESUMEN

El género *Trichoderma* ha sido ampliamente usado para la expresión de proteínas heterólogas debido a su alta capacidad secretora, condiciones establecidas de transformación y de fermentación, además de la existencia de varios genomas secuenciados. Sin embargo poco se ha investigado sobre las proteínas que se encuentran en su pared y su posible uso para la generación de sistemas de expresión superficial de proteínas de interés industrial.

Por su ubicación dentro de la pared celular fúngica las proteínas con anclaje glicosilfosfatidilinositol han sido propuestas para ser el sistema de anclaje e inmovilización en la expresión superficial, principalmente en levaduras. Se han desarrollado diferentes programas bioinformáticos para la identificación *in silico* de estas proteínas, para su posterior caracterización. En el presente trabajo se identificaron un total de 43 proteínas GPI en el proteoma deducido de *Trichoderma reesei*, se clasificaron dependiendo de su posible ubicación y se seleccionaron 5 de éstas, las cuales también están presentes en el proteoma deducido de *Trichoderma longibrachiatum*. Se obtuvieron las regiones codificantes completas de estas proteínas candidatas.

Paralelamente se diseñó y generó un constructo para la expresión superficial de proteínas; este constructo fue generado mediante el sistema de ensamble Gibson®. En este objetivo se tuvieron muchas complicaciones para lograr el constructo ensamblado correctamente. Al final se procedió a combinar el ensamble Gibson® con el uso de enzimas de restricción.

El tercer elemento fue la elección de una triacilglicerol lipasa para la expresión superficial. Se realizó un análisis del proteoma deducido de *Trichoderma harzianum* para identificar las triacilglicerol lipasas y mediante análisis de expresión por RT-PCR se identificó una triacilglicerol lipasa (ID 526309) que se induce *de novo* cuando el medio de cultivo contiene aceite de oliva. Por finalización del programa de doctorado no se logró ensamblar todas las partes, pero sí se lograron obtener todas las que se necesitan utilizar para la prueba del concepto, lo cual se continuará por el grupo de investigación del CICY.

El análisis funcional de inmovilización superficial se evaluó usando a la proteína GFP como reportera en un constructo control generado también en este trabajo usando la secuencia de una proteína GPI ya evaluadas en la literatura y el resultado fue lo esperado, ya que la fluorescencia se localizó específicamente en los septos de las hifas de *T. harzianum*, mientras que cuando el hongo se transforma con el plásmido GFP la fluorescencia ocurre en todo el micelio.

El constructo de expresión superficial está diseñado de manera modular y se podrá modificar a voluntad cambiando las partes que se quiera por las regiones codificantes de las nuevas proteínas GPIs y por la lipasa u otras enzimas de interés.

ABSTRACT

The genus *Trichoderma* has been largely used for heterologous expression of proteins by its high secretory capacity, and well established conditions for transformation and fermentation, as well as several genomes in this genus have been sequenced. However, little research has been carried out about the proteins occurring on its wall and their potential use to develop superficial protein expression systems with industrial interest.

Because of their location in the fungal cell wall, proteins with glycosylphosphatidylinositol anchoring have been proposed as the anchors for the immobilization of proteins in the surface expression systems, mainly in yeasts so far. Different bioinformatic programs have been developed to identify these proteins for their further characterization. In the present work a total of 43 GPI proteins were identified in the deduced proteome of *Trichoderma reesei*; they were classified in regards of their putative location in the cell and 5 of them were selected, which have homologues in the deduced proteome of *Trichoderma longibrachiatum*. The full coding sequences of these candidate GPI-anchored proteins were obtained by RT-PCR.

At parallel, a vector for protein surface expression was designed and constructed by using the Gibson® assembly system. However many complications were faced to achieve the correct assembly of the construct. At the end, the vector construction was carried out by combination Gibson® system and the use of restriction enzymes.

The third part was the selection of triacylglycerol lipase to test the superficial expression. The quest was performed in the deduced proteome of *Trichoderma harzianum* and by RT-PCR-based analysis it was identified the lipase with ID 526309 which is *de novo* induced when olive oil is included in the culture medium. The Ph D time finished and it was not possible to assembly the different new genetic parts, but all fragments required to assembly the protein surface expression were obtained to further challenge the proof of concept, which will be performed by the CICY research team

The functional analysis of superficial immobilization of proteins was tested using GFP reporter protein in a vector constructed here with a GPI protein tested in the literature and

the result was as expected the fluorescence was related with septums in the hyphae of *T. harzianum*, while fluorescence was diffuse in all hyphae when the GFP plasmid was used to transform the fungus.

The surface expression vector has a modular design and the codifying sequences for the new GPI-anchored proteins and the lipase and other enzymes of interest can be introduced by replacing actual parts of the plasmid.

INTRODUCCIÓN

Una de las principales características de las células fúngicas es que están rodeadas por la pared celular una estructura compleja formada por cinco componentes mayoritarios: 1,3- β -glucano, 1,6- β -glucano, 1,3- α -glucano, quitina y glicoproteínas. Dentro de las glicoproteínas las proteínas con anclaje a pared mediante glicosilfosfatidilinositol (GPI) son las más abundantes y son constituyentes claves de la pared celular. Funcionalmente cumplen diversos roles, tales como: fermentación, reproducción, patogenicidad, floculación o adhesión y en la biosíntesis de esta. A pesar de su importancia son pocas las especies fúngicas en las que se han identificado las proteínas con anclaje GPI, por lo que su estudio es relevante debido a su importancia en la estructuración y funcionalidad de la pared, no solo de especies patogénicas sino también de especies de interés industrial como *Trichoderma*.

Debido a su ubicación en la pared celular las proteínas con anclaje GPI han sido propuestas para su uso en la expresión superficial, siendo el sistema de anclaje en la pared celular para la inmovilización de enzimas de interés industrial. Esta metodología es de gran interés para la industria de la biocatálisis ya que simplifica los pasos de purificación, permite el reciclaje de los biocatalizadores y presenta un alto número de ventajas sobre los sistemas sintéticos de inmovilización (Yuzbasheva *et al.*, 2011). La mayoría de reportes sobre la expresión superficial usando proteínas con anclaje GPI se han enfocado a levaduras, con apenas un reporte de hongos filamentosos, inmovilizando principalmente hidrolasas, dentro de las cuales las lipasas han tomado importancia debido a su amplio uso en las industrias de las bioconversiones y su versatilidad catalítica (Jiang *et al.*, 2008).

La adecuada selección de la proteína de anclaje es de suma importancia para el desarrollo exitoso de un sistema de expresión superficial, las proteínas de anclaje deben ser exportadas eficientemente a la superficie celular y evitar interferir con la actividad de la proteína inmovilizar (Tanaka *et al.*, 2012), por lo que una buena estrategia de identificación y un adecuado sistema de expresión es necesario.

CAPÍTULO I

1.1 ANTECEDENTES

La variabilidad de especies fúngicas en la naturaleza es amplia, se estima que existen 13.5 millones (Abdel-Azeem, 2010). Los hongos crecen de diversas formas, ya sea como levaduras unicelulares y/o hifas ramificadas, que producen una notable variedad de esporas (Adams, 2004). La mayoría de los hongos cumplen papeles fundamentales en el ecosistema y en los ciclos de la vida, establecen simbiosis con plantas o animales, y contribuyen significativamente a la descomposición de la materia orgánica, e.g., la degradación de la madera, de animales muertos o excretas (Grün, 2003). Asimismo, cumplen funciones importantes en la vida diaria de las personas, siendo ampliamente usados en la industria, agricultura, medicina, industria alimentaria, textiles, biorremediación, ciclos biogeoquímicos, bio-fertilización, bio-pesticidas, entre otras (Hasan y Gupta 2012).

Pared celular fúngica

Las células fúngicas están rodeadas por la pared celular, una estructura dinámica localizada en la interfase del microorganismo y su ambiente (Firon *et al.*, 2004), que interviene en un número importante de procesos y es vital para el crecimiento, supervivencia y morfogénesis de los hongos (Free, 2013). La pared celular es una coraza protectora dinámica, bajo continuos cambios, dependiendo de las condiciones de cultivo y estrés ambiental; es una estructura rígida, compuesta por multicapas (Kantún-Moreno *et al.*, 2013; Gastebois *et al.*, 2010; Latgé, 2010; Yamazaki *et al.*, 2008); le da fuerza mecánica a la célula para soportar cambios impuestos por el ambiente (Bowman y Free, 2006). Además de ser única para cada especie de hongo (Grün, 2003), otorga protección contra otros microorganismos, mantiene la forma de la célula, interviene en el apareamiento y sirve como centro de señalización para activar vías de transducción de señales dentro de la célula fúngica. Algunas proteínas presentes en ella, tales como las adhesinas y mucinas, permiten la adhesión de las células fúngicas a superficies hidrofóbicas y permiten al hongo la colonización de nuevos ambientes (Free, 2013; Bowman y Free, 2006; Grün, 2003; Schoffemeer *et al.*, 1999). La pared celular también interviene en la formación de biofilmes, y en el caso

de los hongos patogénicos, es crítica para la patogenicidad y la virulencia (Free, 2013).

La pared fúngica de la mayoría de los hongos consiste en una estructura compleja formada por cinco componentes mayoritarios: β -1,3 glucano, β -1,6 glucano, α -1,3 glucano, quitina y glicoproteínas (Fig. 1.1).

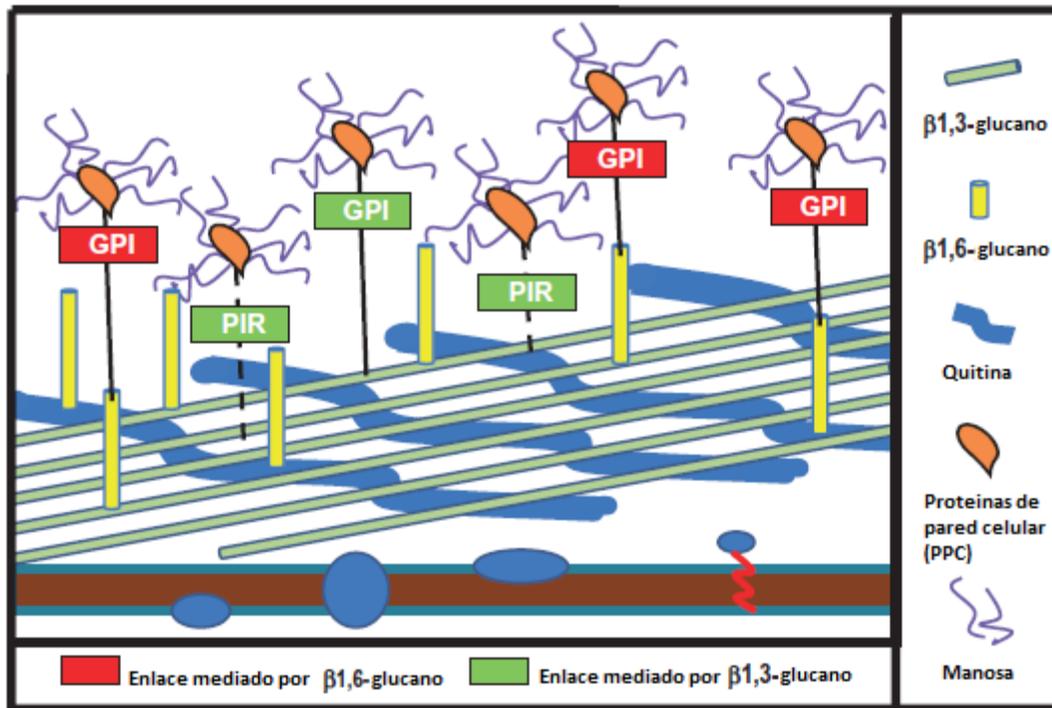


Figura 1.1 Representación esquemática de los componentes mayoritarios de la pared celular (Arana et al., 2009).

La comparación de la pared celular entre distintos hongos evidenció que existe una gran variabilidad en la composición y en la organización. Algunos autores consideran a los mananos y las melaninas como parte también de los componentes principales de la pared (Free, 2013; Adams, 2004). La pared celular mejor estudiada ha sido la de *Saccharomyces cerevisiae* la cual está compuesta de β -1,3 glucano, que es una fracción soluble en álcali (con 20% del total de la proteína de la pared), y otra fracción asociada a quitina, insoluble en álcali (aproximadamente 35%). Además de están presentes β -1,6 glucano (5%), quitina (2%) y manoproteínas (35-40%). Los estudios de microscopía electrónica muestran que los componentes de la pared celular están organizados en capas, en la cual el β -1,3 glucano forma microfibrillas densamente

entretrejadas en la capa más interna, seguidas por β -1,6 glucano y manoproteínas. En las levaduras, las manoproteínas están unidas al β -1,6 glucano vía anclaje con glicosilfosfatidilinositol. Una vez que los glucanos, quitina y las glicoproteínas son liberadas dentro del espacio de la pared celular, varias enzimas se entrecruzan a la pared mediante unión covalentemente con varios de los componentes de la pared celular, generando juntos una matriz tridimensional de quitina/glucano/glicoproteínas; la región más cercana a la membrana plasmática es rica en carbohidratos, mientras que la parte más externa es rica en proteínas (Free, 2013).

La mayor parte de la pared está conformada por polisacáridos. La quitina y glucanos son sintetizados por quitina sintetasas y glucano sintetasas y son extruidos dentro del espacio entre la membrana y la pared celular. Las glicoproteínas son sintetizadas por los ribosomas asociados al retículo endoplasmático y pasan a través de las rutas de secreción canónica (Free, 2013).

Cuadro 1.1 Componentes de la pared celular en diferentes hongos (Free, 2013).

Compo nente de la pared celular	<i>Saccharom yces cerevisiae</i>	<i>Candida albicans</i>	<i>Asper gillus fumig atus</i>	<i>Neurosp ora crassa</i>	<i>Schizosa ccharom yces pombe</i>	<i>Cryptoc occus neoform ans</i>
Quitina	Si (1-2%)	Si (2-6%)	Si (2-6%)	Si (7-15)	Si (4%)	Si
B-1,3 glucano	Si (50-55%)	Si (30-39%)	Si (20-35%)	Si (87%)	Si (46-54%)	Si
Mezcla β - 1,3/1,4 glucano	No	No	Si	Probable mente	No	No
B-1,6 glucano	Si (10-15%)	Si (43-53%)	4%	No	No	Si
A-1,3- glucano	No	No	Si (35-46%)	Solo en la pared de conidios	No	No
Cadena externa de manano	Si (10-20%)	Si (38-40%)	No	No	No	No
Galacto- manano	No	No	Si (20-25%)	Si (12%)	Si (9-14%)	Si
Melanin a	No	Si (Durante infección)	Si (Conidi os)	Si (Ascospo ras y peritecios)	No	Si (Durante infección)

Glucano

Los glucanos son los principales polisacáridos de la pared celular fúngica, compuestos de monómeros de D-glucosa. Estos protegen a la célula de los mecanismos de estrés, determina la forma celular y le otorgan la rigidez a la pared celular (Vaart y Verrips, 1998). De acuerdo al tipo de unión, éstos se pueden dividir en dos grupos: α y β glucanos. Los glucanos más importantes son aquellos vinculados a un enlace β , siendo los de la pared celular de *S. cerevisiae* los mejor estudiados. Se han descrito tres diferentes fracciones; la fracción mayoritaria es la de glucanos insolubles en álcali, los cuales tienen una alta proporción de enlaces β -1,3 y cerca de 3% de enlaces β -1,6. El componente menor es la fracción soluble en álcali con una estructura similar, pero con una mayor cantidad (8-12%) de enlaces β -1,6; mientras que el tercer polímero es altamente ramificado, con 80% de unidades de glucosa unidas por enlace β -1,6 en levaduras, mientras que para hongos filamentosos como *A. fumigatus* estos contienen α -1,3/ 1,4 glucano y galactomananos (Yoshimi *et al.*, 2016; Rodríguez-Iglesias y Schmoll, 2015). La mayoría de estos polímeros son insolubles en agua y están hechos de unidades de glucosa unidas a través de enlaces 1,3, los cuales están presentes en las paredes celulares de un gran número de especies fúngicas (Ruiz-Herrera, 1991). Al igual que lo observado en *S. cerevisiae*, el mayor constituyente de todas las paredes fúngicas caracterizadas es el β -1,3 glucano, que constituye entre el 30% y 80% de la masa de la pared celular; es encontrado en forma de polímero ramificado con ramificaciones unidas a un polímero central por enlaces β -1,6 (Free, 2013) y α -1,3 glucano en el caso de algunos hongos filamentosos y levaduras dimórficas (Fu *et al.*, 2014).

El β -1,3-glucano es sintetizado por β -1,3 glucano sintetasas, las cuales son enzimas asociadas a la membrana plasmática mediante múltiples dominios transmembranales. Las enzimas utilizan como sustrato UDP-glucosa citoplasmático, para la adición de nuevos residuos de glucosa, incrementando la cadena lineal del polímero de glucano. Durante su síntesis éste es extruido dentro del espacio de la pared celular a través de un canal formado por el dominio transmembranal (Free, 2013; Latgé, 2010; Bowman y Free, 2006); cada cadena está compuesta de aproximadamente mil quinientos residuos de glucano conectados vía β -1,3 (Bowman y Free, 2006). El número de genes que codifican para las β -1,3 glucano sintetasas es muy variable entre las diversas especies fúngicas; existen especies que contienen un único gen, tales como *Ustilago*

maydis, *Sporobolomyces roseus* y *Malazessia globosa*, mientras que otras especies como *Schizosaccharomyces pombe* contienen hasta 4 (Ruiz-Herrera y Ortiz-Castellanos, 2010). Los genes que codifican para los componentes de la maquinaria de β -1,3-glucano sintasa fueron identificados por primera vez en *S. cerevisiae*, la cual contiene dos subunidades catalíticas y una proteína regulatoria (Bowman y Free, 2006).

Quitina

La quitina (Fig. 1.2) es un biopolímero lineal con carga neutra, formado por monómeros de N-acetilglucosamina unidas por enlaces β -1.4 (Bueter *et al.*, 2013). La quitina es una parte esencial del esqueleto de carbohidratos de la pared fúngica y su síntesis es realizada por las enzimas quitina sintasas asociadas a la membrana plasmática. Para la síntesis de la molécula lineal estas enzimas usan como sustrato UDP-N-acetilglucosamina, localizado en la parte de la membrana adyacente al citoplasma (Free, 2013). En las levaduras es un componente minoritario de la pared celular, representando solo 1-2% del peso seco de la pared (Cuadro 1.1), mientras que en los hongos filamentosos, tales como *Neurospora* y *Aspergillus* se reportan contenidos de quitinas de 10-20% (Bowman y Free 2006; Klis *et al.*, 2002).

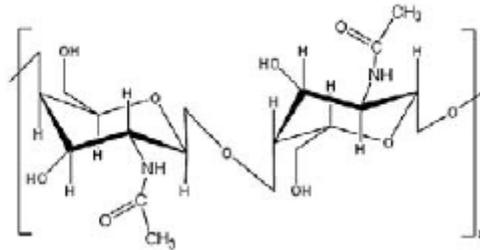


Figura 1.2 Estructura química de la quitina (Bueter *et al.*, 2013)

La síntesis de la quitina está regulada por complejos mecanismos que permiten depositar al polímero en sitios específicos a lo largo del ciclo de vida celular, manteniendo la fuerza constante en la pared y permitiendo rápidas modificaciones hechas bajo condiciones de estrés. El polímero de quitina es el principal componente del septo primario y en levaduras está involucrado en la separación de células madres e hijas (Vaart y Verrips, 1998). En la mayoría de los hongos se ha encontrado que una o más quitina sintasas son las responsables de la mayor parte de la síntesis del

biopolímero. En *S. cerevisiae* se han reportado tres enzimas CHS1, CH2 y CHS3, siendo esta última la responsable de la mayor parte de la síntesis de la quitina (Rolli *et al.*, 2009). En *Aspergillus nidulans* se han caracterizado seis genes de quitina sintasas, demostrando que juegan funciones clave en el crecimiento de las hifas y desarrollo de los conidióforos (Yamazaki *et al.*, 2008). Estudios recientes también han propuesto un papel importante de la quitina en la activación y la atenuación de la respuesta inmune hacia los hongos (Lenardon *et al.*, 2010).

Manoproteínas

Las manoproteínas aparecen en la parte superior de la pared. Determinan la porosidad y regulan el transporte de las proteínas del espacio periplasmático y la entrada de macromoléculas del ambiente (Damveld *et al.*, 2005; Vaart y Verrips, 1998). Muchas de las proteínas funcionan integralmente, en conjunto, en una matriz entrecruzada, importante para la biogénesis de la pared; otras proteínas de pared cumplen funciones diversas tales como sensores de la pared, adhesinas, adquisición de hierro, destoxificación de radicales libres de oxígeno, en la interacción de la célula con su ambiente, en procesos de apareamiento, floculación, contribuyen con la integridad de la pared y limitan el acceso de agentes estresantes, por ejemplo glucanasas de otros organismos (Free, 2013; Yin *et al.*, 2008). La mayoría de estas proteínas comparten la propiedad de tener una secuencia peptídica N-terminal y son transportadas por la vía de secreción hacia la pared celular. Durante el recorrido son altamente glicosiladas en los sitios N- y O-glicosilación. Las cadenas en "O" son oligomanosil cortos, los cuales pueden ser adicionados en cualquier residuo de serina o treonina. Ambas cadenas, tanto N- como O-, son iniciadas en el retículo endoplasmático (Vink *et al.*, 2004).

Las manoproteínas pueden dividirse en tres grupos: manoproteínas extraíbles con dodecil sulfato de sodio (SDS), extraíbles mediante agentes reductores y extraíbles mediante glucanasas (Hamada *et al.*, 1998). En estas últimas se pueden encontrar dos tipos distintos: las proteínas de pared acopladas a glicosilfosfatidilinositol (GPI-CWPs) y las proteínas de pared celular Pir (Pir-CWPs). Las primeras pueden ser específicamente liberadas mediante HF-piridina (Klis *et al.*, 2010), y son el grupo más abundante. Estas proteínas se encuentran unidas al β -1,3 glucano a través de una molécula de β -1,6 glucano en levaduras y galactomananos en el caso de *Neurospora crassa* (Maddi *et al.*, 2012). Las proteínas Pir, están presuntamente unidas directamente al β -1,3 glucano, un enlace que es sensible a álcalis (Vink *et al.*, 2004).

El tercer tipo de proteínas detectadas en la pared son exportadas por rutas alternas, ya que no cuentan con la secuencia canónica de secreción. Dentro de estas proteínas se encuentran: enolasas, aconitasas, piruvato cinasas, fosfoglicerato mutasas, metionina sintasa, entre otras (Martinez, 2004).

Proteínas Pir

Las proteínas con repeticiones internas (Pir, por sus siglas en inglés) son altamente O-glicosiladas y están unidas al 1,3- β -glucano en la pared celular por enlaces alcalinolábiles (Free, 2013; Ecker *et al.*, 2006; Martinez, 2004). No contienen señal de adición de anclaje GPI en la región C-terminal, pero se unen covalentemente a la red de glucano de la pared celular. Todas las proteínas PIR presentan una estructura similar que consiste en un péptido señal en la región N-terminal, un sitio de corte que reconoce la proteasa Kex2 y una región de una o más repeticiones con una secuencia conservada Q[IV][STGNH]DGQ[LIV]Q (Fig. 1.3b). Además, la región del dominio C-terminal en los miembros de la familia de proteínas PIR es altamente conservada y contiene un motivo de 4 cisteínas (Free 2013; Ruiz-Herrera y Ortiz-Castellanos 2010; Vink *et al.*, 2004). En algunos casos también son encontradas repeticiones PIR en proteínas de pared GPI como: CWP1, CWP2, TIR1 y TIR2 de *S. cerevisiae* (Vink *et al.*, 2004).

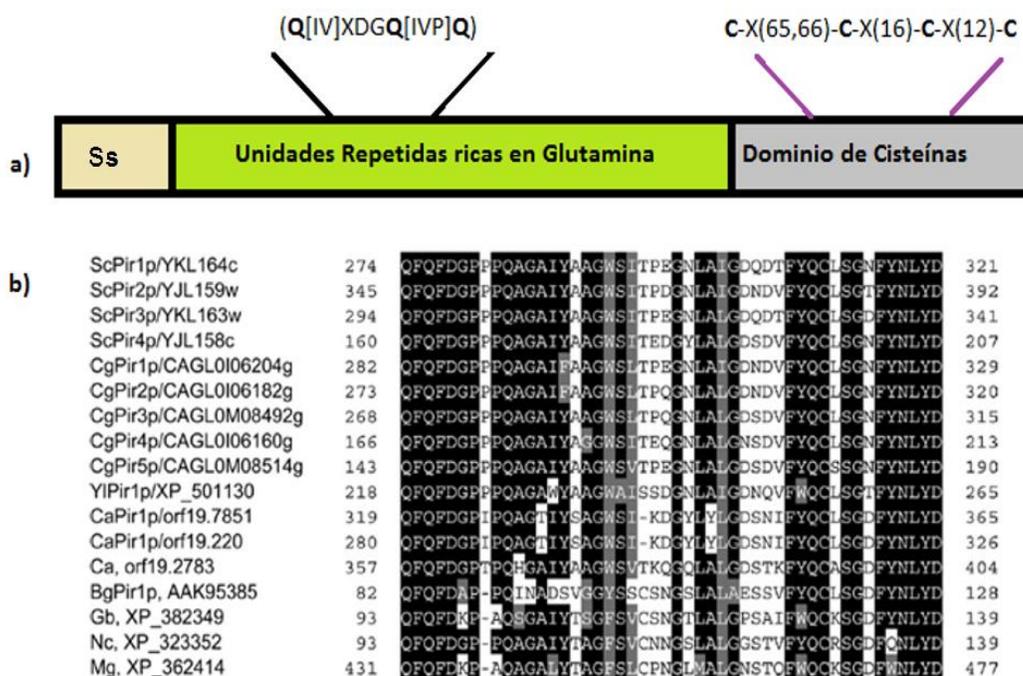


Figura 1.3 Características en la secuencias de las proteínas Pir: a) Representación esquemática de los dominios conservados en las proteínas Pir, b) Alineamiento de proteínas tipo Pir de origen fúngico. Se muestra la parte más conservada, iniciando después de la primera cisteína en el dominio característico de cisteínas (Kantún-Moreno *et al.*, 2013; De Groot *et al.*, 2005).

El primer organismo en el cual fueron descritas estas proteínas fue *S. cerevisiae*, que junto con *Candida albicans* presentan la mayor cantidad en su pared. En trabajos posteriores se demostró la presencia de proteínas Pir en ascomicetos. Las proteínas Pir son importantes en la estructura y estabilidad de la pared celular de los ascomicetos y levaduras; sin embargo no son esenciales para la viabilidad celular (Ecker *et al.*, 2006). Se distribuyen uniformemente en la capa interna de polisacáridos de la pared celular y en caso de daños a la pared, los genes que codifican para estas proteínas son altamente expresados (De Groot *et al.*, 2005).

Proteínas con anclaje GPI

Las proteínas de pared celular ancladas mediante glicosilfosfatidilinositol (GPI) son las proteínas más abundantes del grupo de manoproteínas covalentemente unidas a la pared y son constituyentes claves de ésta (Kantún-Moreno *et al.*, 2013; Frieman y Cormack 2004). Cumplen diversas funciones dependiendo de su ubicación dentro de

la pared (Bruneau *et al.*, 2001). Aquellas que se encuentran vinculadas vía β -1,6-glucano en levaduras, a través de un remanente de anclaje GPI que abarca la molécula de etanolamina-fosfato y varios residuos de manosa, cumplen funciones en la fermentación, reproducción, floculación o adhesión. Al segundo grupo pertenecen aquellas que presenta actividad enzimática y participan en la biosíntesis de la pared (GAS1 en *S. cerevisiae*), las cuales están asociadas a la membrana mediante el anclaje GPI (De Groot *et al.*, 2005).

La secuencia primaria de las proteínas con anclaje GPI tiene tres características en común: una secuencia señal para secreción en el extremo N-terminal, secuencias ricas en Serina/Treonina que son el sitio de glicosilación, y una señal de unión a GPI en el C-terminal. La señal de unión a GPI está compuesta por un sitio de anclaje a GPI denominado sitio omega (ω), la cual es una región espaciadora compuesta de 8-12 aminoácidos polares y un tramo de entre 10-20 aminoácidos hidrofóbicos (Kantún-Moreno *et al.*, 2013; Kapteyn *et al.*, 2002; Eisenhaber *et al.*, 1999)

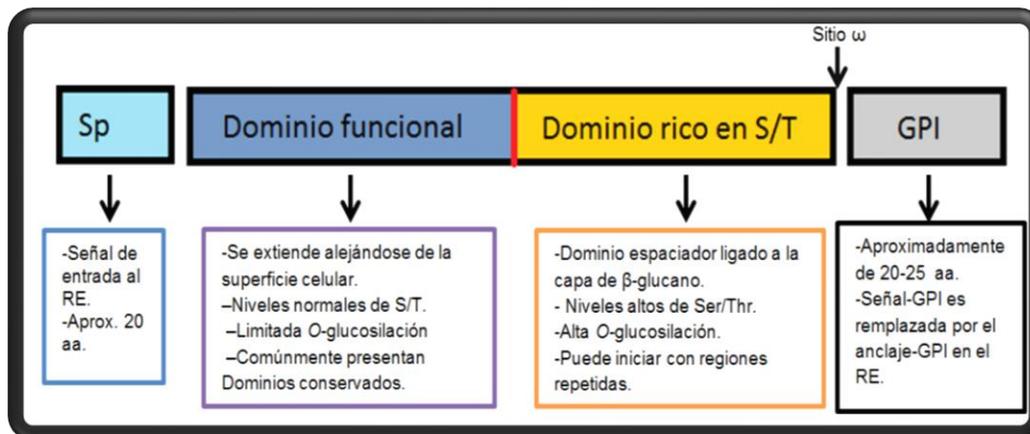


Figura 1.4 Esquematización de los sitios característicos de las proteínas con anclaje GPI (Klis *et al.*, 2011).

El núcleo estructural del anclaje GPI está altamente conservado y consiste en un solo fosfolípido que abarca la membrana y un grupo fosfodiéster enlazado a inositol, el cual se une a una glucosamina, que posteriormente se une a una cadena lineal de tres azúcares de manosa, unida a su vez a una molécula de fosfatidiletanolamina en la manosa terminal (Fig. 1.5).

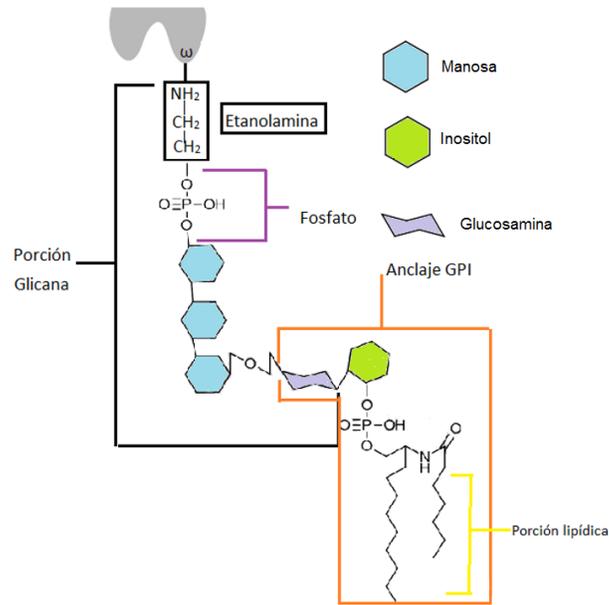


Figura 1.5 Estructura química de la molécula de glicofosfatidilinositol. En gris glicoproteína, ω -sitio de anclaje a GPI, Celeste: manosa, verde: inositol, morado: glucosamina (Modificado de: Vink *et al.*, 2004)

La adición del anclaje GPI a la nueva proteína sintetizada ocurre en el retículo endoplasmático, involucrando al menos 10 reacciones (Fujita y Kinoshita, 2012; Caro *et al.*, 1997). El sitio de unión también llamado sitio omega (ω), está conformado preferentemente por aminoácidos pequeños tales como glicina y asparagina, localizado entre los 17 y 25 aminoácidos río arriba de la región C-terminal, seguido de dos aminoácidos pequeños localizados en la posición $\omega+1$ y $\omega+2$ (Pittet y Conzelmann, 2007; Frieman y Cormack, 2004; Vink, *et al.*, 2004; Caro *et al.*, 1997). Generalmente el anclaje GPI en eucariotas superiores resulta en una unión específica de la proteína a la membrana plasmática; por el contrario la mayoría de proteínas con anclaje GPI de células fúngicas continúan su procesamiento posteriormente (Frieman y Cormack, 2004). Entonces, el anclaje GPI es cortado para llevar a la proteína de la membrana a la pared, después de la cual los remanentes de las proteínas son entrecruzados con el β -1,6-glucano, el cual se conecta al β -1,3-glucano. En el caso de la proteína GPI α -aglutinina de *S. cerevisiae* se ha propuesto que una fosfolipasa C específica de fosfatidilinositol, es la encargada de cortar el anclaje GPI y liberar la proteína de la pared celular (Lu *et al.*, 1994). Caro y colaboradores (1997) propusieron que un motivo dibásico en el 4º residuo río arriba del sitio omega sirve para retener a la proteína en la membrana plasmática. En contraste, aminoácidos hidrofóbicos

ubicados en el sitio 2, 4 y 5 río arriba del sitio de unión GPI parecen tener correlación con la incorporación de la proteína a la pared celular (da Silva *et al.*, 2005). La presencia de valina, isoleucina o leucina en la posición ω -5 o ω -4 y tirosina o asparagina en la posición ω -2 tienden a dirigir a las proteínas con anclaje GPI a la pared celular (Vink *et al.*, 2004) Por otra parte algunos trabajos sugieren que el modelo del motivo dibásico para la distribución de las proteínas con anclaje GPI puede ser simplista. Frieman y Cormack (2003) determinaron el contenido de Ser/Thr en 59 proteínas putativas GPI de *S. cerevisiae*, previamente identificadas por Eisenhaber y colaboradores (2004). Concluyeron que los aminoácidos que se extienden en la región rica en serina y treonina promueven la localización de las proteínas a la pared celular, aún en proteínas cuya señal próxima contiene el motivo dibásico que tiende a asociar a las proteínas a la membrana plasmática (GPI-PMPs). Cuando se fraccionaron los dominios en dos grupos, las proteínas con anclaje GPI-PMPs tenían un contenido menor al 30% de Ser/Thr, mientras que las GPI-Cwps tenían un contenido arriba del 30%. Previamente Caro y colaboradores (1997) reportaron 13 genes de *S. cerevisiae* que codifican para proteínas con anclaje GPI de pared, que además de presentar todas las características estructurales típicas de las proteínas, todas presentaban regiones ricas en serina y treonina. Estas son altamente O-glicosiladas con cadenas cortas de manosa, siendo otro posible motivo que determinará la ubicación de la proteína. Este dato muestra que la célula puede integrar múltiples secuencias para la determinación de la distribución de las proteínas con anclaje GPI entre la pared celular y la membrana.

Identificación de las proteínas con anclaje GPI

Diferentes trabajos se han enfocado en la identificación y caracterización de las proteínas con anclaje GPI, ya sea por su importancia en la biogénesis de la pared o para su aplicación en sistemas de expresión de proteínas heterólogas. La identificación y aislamiento de estas proteínas ha sido mediante técnicas bioquímicas o por el análisis *in silico* de los genomas; Bruneau y colaboradores (2001) a través de la liberación de proteínas con anclaje GPI mediante fosfolipasa C y secuenciándolas mediante espectrometría de masas, identificaron en *Aspergillus fumigatus* nueve proteínas involucradas en la biosíntesis de la pared celular; cinco de ellas presentaron homología con proteínas de levadura previamente reportadas (CSA1, CRH1, CRH2, ECM33 y GAS1).

Varios programas de modelamiento computarizado han sido desarrollados para identificar los sitios potenciales de adición del anclaje GPI en las secuencias primarias de las proteínas candidatas. Ayudados con la secuenciación de los genomas completos, estos programas permiten hacer escrutinios *in silico* para la identificación de las proteínas con anclaje GPI. Caro y colaboradores (1997) determinaron a partir del genoma de *S. cerevisiae* las posibles proteínas con anclaje GPI y su destino hacia la pared o membrana celular, utilizando como criterio la presencia del péptido señal de secreción en N-terminal y la señal de unión del anclaje GPI, clasificando 20 en la membrana plasmática y 38 en la pared celular. Posteriormente Hamada y colaboradores (1998) realizaron nuevamente la búsqueda de posibles proteínas con anclaje GPI en *S. cerevisiae*, basándose en tres criterios: la presencia de la secuencia señal de anclaje GPI, la secuencia señal de secreción y una región que fuese rica en serina o treonina, encontrando al menos 53 proteínas candidatas. El desarrollo de algoritmos y programas más específicos para proteínas con anclaje GPI de origen fúngico como el predictor BIG-PI (Eisenhaber *et al.*, 1999), han permitido análisis *in silico* a gran escala (Bowman y Free, 2006). El BIG-PI predictor fue utilizado para analizar el proteoma de hongos e identificó respectivamente 59, 28, 169, 74 y 87 genes que codifican proteínas-GPI en los genomas de *S. cerevisiae*, *S. pombe*, *C. albicans*, *A. nidulans* y *N. crassa* (Eisenhaber *et al.*, 2004). Por su parte De Groot y colaboradores (2003) generaron un algoritmo propio a partir del reportado por Coyne y colaboradores (1993), dando mayor especificidad mediante ajustes en el momento de la selección del sitio de unión o sitio ω . El algoritmo reconoció proteínas con anclaje GPI conocidas y una vez validado se empleó para hacer escrutinios en diferentes hongos, tanto levaduriformes (*S. cerevisiae*, *C. albicans*, *S. pombe*) como filamentosos (*N. crassa*). Las proteínas identificadas fueron posteriormente analizadas para determinar la secuencia N-terminal, identificando 66 posibles proteínas con anclaje GPI en *S. cerevisiae*, un número mayor que las reportadas en trabajos previos (Hamada *et al.*, 1998; Caro *et al.*, 1997), indicando que este algoritmo permite una identificación más completa de proteínas GPI de *S. cerevisiae*. Recientemente, para la identificación de las proteínas con anclaje GPI, Kantún-Moreno y colaboradores (2013), utilizaron varios programas bioinformáticos con distintos algoritmos, además de una estrategia basada en dos pasos: primero, la búsqueda en distintas bases de datos de proteínas de origen fúngico y con anclaje GPI y posteriormente la búsqueda de sus homólogos en el genoma del hongo patógeno *Pseudocercospora*

(anteriormente conocido como *Mycosphaerella*) *fijiensis*. Posteriormente las proteínas candidatas fueron analizadas con diferentes predictores bioinformáticos para la identificación del sitio ω (BIG-PI, GPI-SOM predictor, predGPI) y la secuencia señal de secreción (SignalIP 3.0 predictor y SOSUlsignal predictor), siendo consideradas positivas aquellas proteínas que tuvieran resultado positivo con al menos tres predictores, de un total de 466 candidatos, 49 fueron positivos.

Expresión en la superficie celular.

La capacidad de las proteínas con anclaje GPI de asociarse a la pared celular ha conllevado a su uso en la retención en la capa más externa de la célula fúngica de proteínas heterólogas. Esta técnica es conocida como “Expresión en la superficie celular” (*Cell surface display*) y permite la expresión de péptidos o proteínas específicas en la superficie celular a través de la unión con una proteína de anclaje, fusionada genéticamente (Tanaka *et al.*, 2012). La inmovilización de proteínas en la superficie celular provee un camino para la generación de reactores enzimáticos novedosos, teniendo numerosas ventajas en comparación con los métodos de inmovilización en soportes sintéticos; el catalizador enzimático pueda ser simplemente removido del medio de la reacción por filtración o centrifugación y si es necesario puede ser regenerado fácilmente (Yuzbasheva *et al.*, 2011; Vaart y Verrips, 1998). La técnica se ha empleado en el desarrollo de vacunas, anticuerpos, bibliotecas de escrutinio, bioconversión, biosoporte, inmovilización, biocatálisis de célula completa y biosensores (Tanaka *et al.*, 2012; Adachi *et al.*, 2008; Vaart y Verrips, 1998). Inicialmente la técnica fue desarrollada en bacteriófagos, en los cuales las proteínas de interés son expresadas e inmovilizadas mediante fusión con una proteína de la cubierta del fago, pero presentaba varios problemas debido a la limitación del tamaño en las proteínas expresadas (Kim *et al.*, 2002). En los hongos es posible su uso mediante la fusión del gen que codifica la proteína de interés con una señal de secreción o una secuencia que codifica generalmente parte de la región C-terminal de una proteína de pared celular que contiene la secuencia señal de anclaje GPI, permitiendo que la proteína se una covalentemente al β -1,6-glucano de la pared celular en el caso de las levaduras y a los galactomananos en el caso de algunos hongos filamentosos como *N. crassa* y *A. fumigatus*. Como una alternativa, la inmovilización de las proteínas objetivo en la superficie de la célula mediante el anclaje C-terminal también puede llevarse a cabo usando el dominio N- de diferentes

proteínas de la pared celular. Entonces, las proteínas pueden ser unidas a proteínas de la pared celular por enlaces covalente al β -1,3-glucano, por enlaces a través de puentes disulfuro a glicoproteínas covalentemente unidas a β -1,6-glucanos, o por adhesión no covalente (Yuzbasheva *et al.*, 2011).

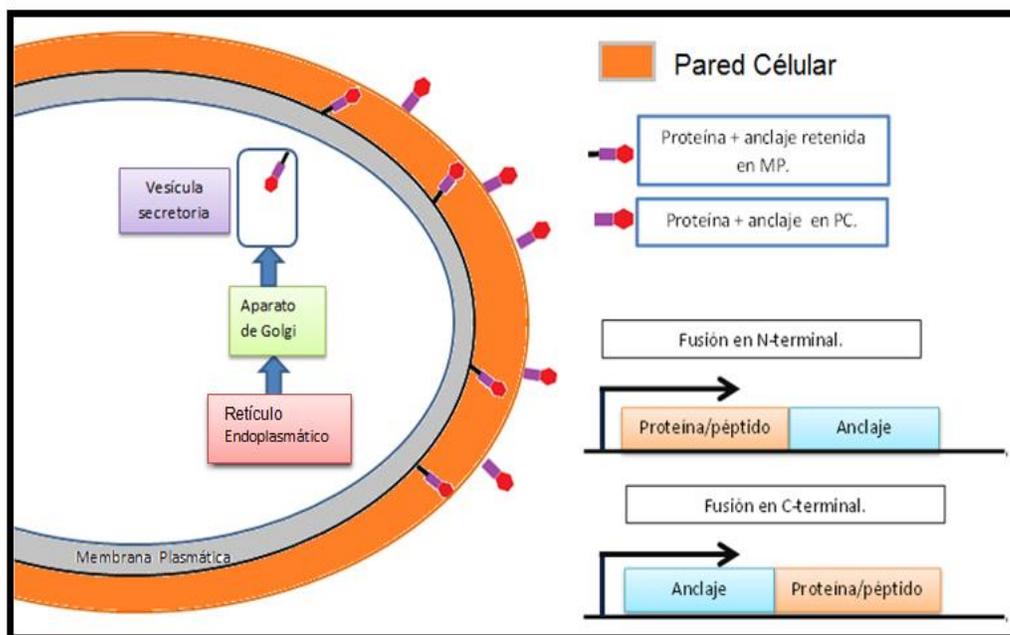


Figura 1.6 Esquematación de la expresión en superficie celular mediante dos tipos de anclaje: N-terminal y C-terminal. MP, membrana plasmática; PC, pared celular (Basado en Tanaka *et al.*, 2012; Ueda y Tanaka, 2000).

El sitio de la fusión entre la proteína de interés y la proteína de anclaje determina si la proteína de interés conserva o no su función cuando son expresadas. Van der Vaart y colaboradores (1998) compararon la capacidad de incorporación de la α -galactosidasa en la pared celular de *S. cerevisiae* mediante la fusión de la mitad correspondiente al C-terminal de la α -aglutinina y otras proteínas de pared celular con dominio GPI como: FLO1, CWP1, CWP2, TIP1 y SED1. Aunque el 80% de estas proteínas fueron fusionadas a pared se observaron grandes diferencias en la retención de la actividad enzimática de la α -galactosidasa, por lo que la selección de la proteína de anclaje puede impactar en la cantidad total de proteína desplegada en la pared y en la actividad recuperada (Tanaka *et al.*, 2012).

Cuadro 1.2 Expresión en la superficie celular en distintos hongos (Tanaka *et al.*, 2012; Adachi *et al.*, 2008; Kim *et al.*, 2002)

Organismo	Anclaje (tamaño)	Proteína objetivo	Aplicación
<i>S. cerevisiae</i>	CWP1, CWP2, AGA1, TIP1, FLO1, SED1, YCR89 Y TIR1P	α -Glucosidasa	Optimización del anclaje
	Fusión del Dominio rico en Ser/Thr (350 a.a)-CWP2 (71aa)	Hemaglutinina	Optimización del sistema de expresión
	FLO1P (42, 102, 146, 318, 428, y 1326 a.a)	<i>Rhizopus oryzae</i> glucoamilasa	Optimización del anclaje
	α -aglutinina	Co-expresión de xilanasas y β xilosidasa	Producción de etanol con xilano
	α -aglutinina	<i>R. oryzae</i>	Biocatalisis de célula completa en solventes orgánicos
	FL y FS	<i>R. oryzae</i> lipasa	Bio catálisis de célula completa
	AGA2	<i>Y. lipolytica</i> lipasa Lip7 y Lip8	Bio catálisis de célula
<i>Pichia pastoris</i>	AGA1 de <i>S. cerevisiae</i>	Enzima amarilla	Bioremediación
	SED1 de <i>S. cerevisiae</i> (338 a.a)	CALB	Biocatalizador de célula completa
	FLO1 (1099 a.a)	Lipasa de <i>R. oryzae</i>	Biocatalizador de célula completa
	AGA2 de <i>S. cerevisiae</i>	Interferón beta de ratón	Desarrollo del sistema de anclaje
<i>Yarrowia lipolytica</i>	Y1CWP110 (110 a.a)	Liasa alginato	Producción de oligosacáridos
	YICWP1-94 (94 a.a)	Lipasa (Lip 2 334 a.a)	Biocatalizador de célula completa
<i>Hansenula polymorpha</i>	YALI0c09031p N-domino (662 a.a)	Lipasa (lip2 334 a.a)	Biocatalizador de célula completa.
	<i>Hp</i> SED1, <i>Hp</i> GAS1, <i>Hp</i> TIP1 y <i>Hp</i> CWP1	Glucosa oxidasa (<i>Aspergillus niger</i>)	Desarrollo del sistema de anclaje
<i>Aspergillus oryzae</i>	MP1 CWP AO090003000648 AO090005001090 AO090120000279	B-glucosidasa	Desarrollo del sistema de anclaje

La mayoría de los trabajos de expresión superficial en células fúngicas se han llevado a cabo en levaduras (Cuadro 1.2), principalmente en *S. cerevisiae*. Sin embargo, los hongos filamentosos representan buenas alternativas para la aplicación de esta técnica debido a su alta producción de enzimas tanto homólogas como heterólogas, y su alta capacidad secretora, alcanzando concentraciones de proteínas de g/L en el medio (Adachi *et al.*, 2008; Mach y Zeilinger, 2003). Adachi y colaboradores (2008) desarrollaron el primer sistema de expresión superficial en hongos filamentosos, realizando primero un análisis de las proteínas de pared celular en el genoma en *Aspergillus oryzae* y enlistando las proteínas con homología con proteínas GPI previamente descritas en *S. cerevisiae*. Seleccionaron cinco candidatas, con anclaje GPI de distintas longitudes y las fusionaron inicialmente con GFP para observar su localización en la célula fúngica (Fig. 8). Aquellas que dieron fluorescencia en la pared se fusionaron posteriormente con β -glucosidasa (BGLI), observando un incremento de la actividad de la enzima con la fusión con GPI-AOMP1, la cual consistía de una región C-terminal de 322 aminoácidos de la proteína con anclaje GPI. El incremento de la actividad permitió asumir que la proteína de fusión (BGLI-AOMP1) se integró correctamente en la pared celular, debido a las características de la proteína de anclaje de la AOMP1 y a la habilidad secretora de *A. oryzae*.

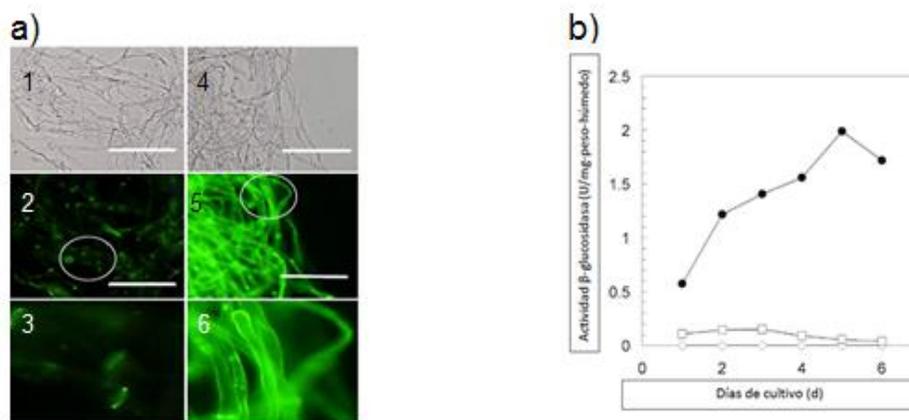


Figura 1.7 Visualización por microscopía de fluorescencia de *Aspergillus oryzae* y cuantificación de actividad de β -glucosidasa desplegada en la pared. a) Cepa recombinante produciendo proteína verde fluorescente (1, 2 y 3). Cepa recombinante expresando fusión de proteína GFP-AOMP1 (4, 5 y 6). b) Actividad enzimática de β -glucosidasa. Círculos negros: cepa expresando la proteína de fusión BGLI-MP1. Cuadrados: Cepa expresando proteína de fusión BGLI-CWP. Rombos: *A. oryzae* (Adachi *et al.*, 2008).

***Trichoderma* como posible sistema de expresión en superficie celular.**

Los hongos filamentosos del género *Trichoderma* tienen una larga historia en la producción industrial de diferentes enzimas hidrolíticas, metabolitos secundarios, biocontrol, biorremediación de metales pesados y producción de proteínas heterólogas (Mukherjee *et al.*, 2013; Schuster y Schmoll, 2010; Keränen y Penttilä, 1995). Algunas cepas mutantes han alcanzado rendimientos en la producción de celulasas en el rango de 40 g/L en el medio de cultivo (Mukherjee *et al.*, 2013).

La excelente producción y capacidad de secreción de este género, junto con las condiciones establecidas de fermentación, lo han llevado a ser uno de los modelos más utilizados para la expresión de proteínas recombinantes (Schuster y Schmoll, 2010; Nevalainen *et al.*, 2005; Keränen y Penttilä, 1995), incentivando entender la biología de la especie. Actualmente se ha llevado a cabo la secuenciación del genoma completo de varias especies de *Trichoderma*, siendo *Trichoderma reesei* el primer miembro secuenciado, seguido por *Trichoderma atroviride*, *Trichoderma virens*, *Trichoderma harzianum*, *Trichoderma asperellum*, *Trichoderma citrinoviride* y *Trichoderma longibrachiatum* (Mukherjee *et al.*, 2013). A pesar de todo esto, existe solo un reporte del uso de *Trichoderma* en la expresión superficial de proteínas, utilizando una proteína de anclaje-GPI perteneciente a *A. fumigatus* (Su *et al.*, 2013). El presente trabajo propone el escrutinio *in silico* de proteínas con anclaje GPI en el género *Trichoderma*, para su comparación en la expresión superficial celular de proteínas o péptidos.

1.2 JUSTIFICACIÓN

La identificación de las proteínas con anclaje GPI ha sido relevante debido a su importancia en la biogénesis y funcionalidad de la pared fúngica. Los métodos computacionales proveen de un punto de inicio exitoso en la identificación *in silico* de potenciales proteínas con anclaje GPI en los genomas, facilitando su identificación y aislamiento. Su posible aplicación en el desarrollo de técnicas tales como la expresión superficial de proteínas o péptidos de interés en la célula, abre una oportunidad para el uso de éstas en la biocatálisis. Sin embargo, la mayoría de los reportes que utilizan a las proteínas con anclaje GPI para la expresión superficial se han llevado a cabo en levaduras, existiendo pocos reportes en hongos filamentosos, a pesar de su alto poder secretor y capacidad de expresión de proteínas heterólogas.

Recientemente, la biología molecular de especies de *Trichoderma* ha sido estudiada extensivamente, incluyendo la determinación de los genomas completos. Estos hongos se han usado de manera industrial para la producción de proteínas heterólogas desde hace más de 20 años, estando, junto con el género *Aspergillus*, entre los hongos filamentosos más comunes para la producción de proteínas heterólogas. El presente trabajo se enfoca en la identificación y comparación de las proteínas con anclaje GPI para el desarrollo de un nuevo sistema de expresión superficial de proteínas o péptidos para futuro interés industrial.

Asimismo, como modelo para prueba de concepto se identificarán triacilglicerol (TAG) lipasas, con capacidad de hidrolizar triacilglicerol con ácidos grasos de cadena larga. Esto con el fin de eventualmente inmovilizar en la pared de *Trichoderma* alguna(s) de esas lipasas.

1.3 HIPÓTESIS

El uso secuencial de las herramientas predictivas disponibles permitirá identificar secuencias de proteínas GPI que permitan la futura expresión superficial de una triacilglicerol lipasa en *Trichoderma harzianum*.

1.4 OBJETIVOS

OBJETIVO GENERAL

Construir un sistema para la expresión superficial en *Trichoderma* sp. de enzimas fúngicas.

OBJETIVOS PARTICULARES

Identificar en la base de datos del genoma de *Trichoderma reesei* proteínas de pared celular.

Identificar mediante programas bioinformáticos las proteínas con anclaje GPI que están asociadas a la pared celular.

Seleccionar proteínas con anclaje GPI, presuntamente asociadas pared celular (GPIp) y diseñar cebadores específicos para su aislamiento por PCR.

Diseñar y construir un vector de expresión para la fusión de proteínas con anclaje GPI y etiqueta GFP (proteína verde fluorescente).

Transformar *Trichoderma harzianum* con los constructos GPIp–GFP.

Analizar mediante microscopía de fluorescencia la expresión superficial en *Trichoderma harzianum* de la proteína GFP fusionada a cada GPIp.

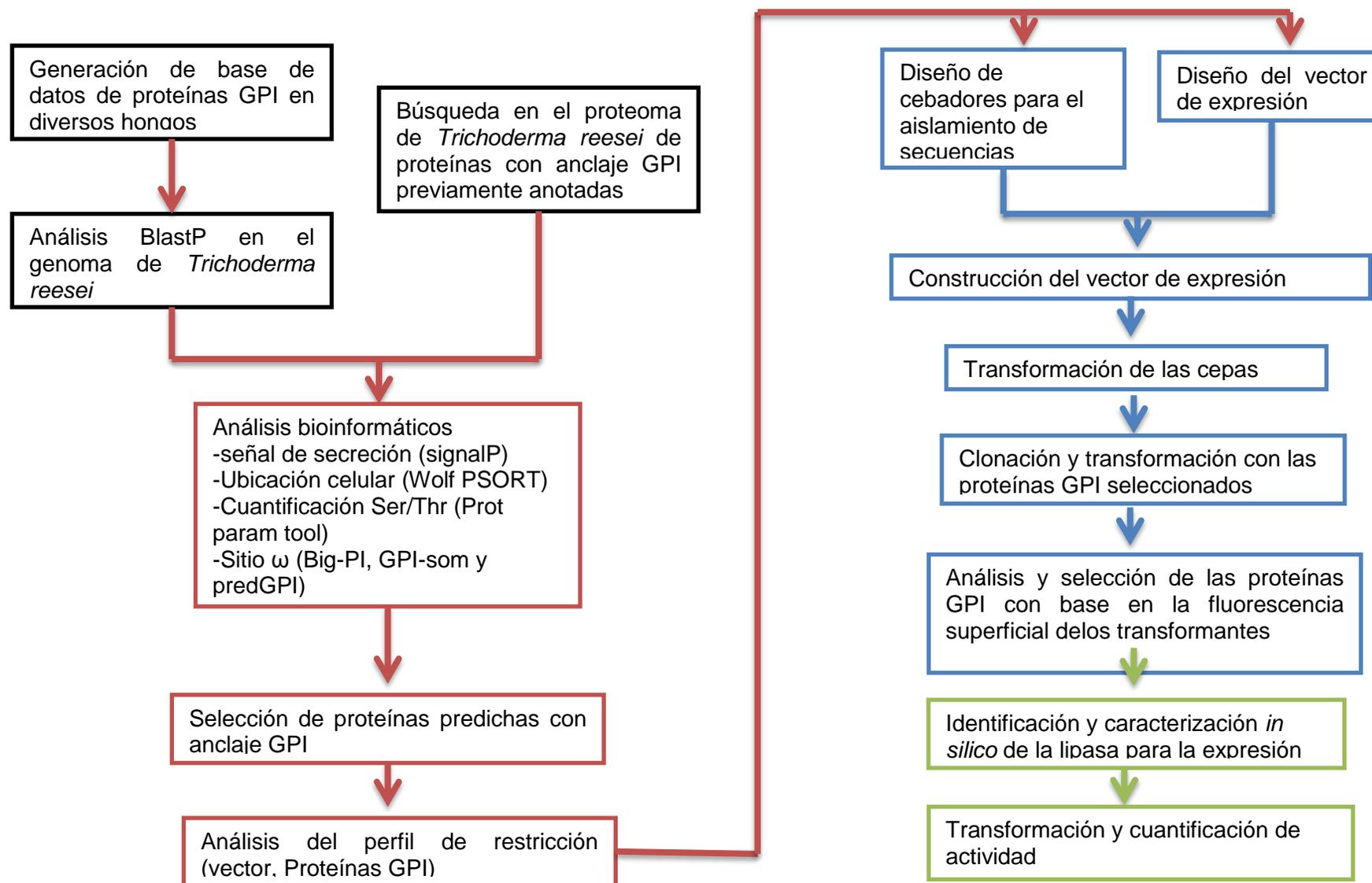
Seleccionar de *Trichoderma harzianum* una triacilglicerol (TAG) lipasa con capacidad de hidrolizar TAG con ácidos grasos de cadena larga.

Fusionar la lipasa al vector de expresión conteniendo la proteína con anclaje GPI.

Transformar *Trichoderma harzianum* con los constructos Lipasa-GPIp.

Cuantificar la actividad de lipasa inmovilizadas en la pared celular.

1.5 ESTRATEGIA EXPERIMENTAL



CAPÍTULO II

IDENTIFICACIÓN *IN SILICO* DE GENES QUE CODIFICAN PARA PROTEÍNAS CON ANCLAJE A GPI EN EL GENOMA DE *T. REESEI* Y *T. LONGIBRACHIATUM*

2.1 INTRODUCCIÓN

Las proteínas con anclaje GPI presentan amplio rango de tamaños y variedad de funciones dentro de la células, entre las que se pueden enlistar receptores de señales externas, receptores de nutrientes, proteasas extracelulares, moléculas de adhesión, etc (Cao *et al.*, 2009; Fankhauser y Mäser, 2005). La expresión de proteínas tipo GPI en sistemas heterólogos han revelado que la señal de anclaje GPI es generalmente reconocida en todos los reinos eucariontes (Fankhauser y Mäser, 2005).

A pesar de que existe una gran variabilidad entre las secuencias de las proteínas con anclaje GPI, éstas presentan características estructurales similares, por ejemplo un péptido señal de secreción en la región N-terminal, con el cual las proteínas son direccionadas al retículo endoplasmático, en donde mediante aminoácidos hidrofóbicos la región C-terminal de la proteína interactúa con el complejo de transaminasa. Este complejo es responsable de la remoción del dominio C-terminal y de la unión de un anclaje GPI a un sitio específico (sitio ω) de la proteína naciente. Sin embargo, la región C-terminal de las proteínas con anclaje GPI no son tan conservadas como para ser alineadas en secuencias consenso (Kantún-Moreno *et al.*, 2013; Fankhauser & Mäser, 2005). El uso de fosfolipasas C o D para la solubilización de las proteínas con anclaje GPI y su posterior identificación, tiene al menos dos problemas según Eisenhaber y colaboradores (2004). El primero es que muchos factores afectan la sensibilidad del corte proteolítico por acción de las fosfolipasas, como la microheterogeneidad del anclaje, y el segundo hace referencia a las fracciones proteicas que no cumplen con la vía canónica de secreción.

La generación de predictores bioinformáticos para la determinación del sitio omega (ω) y el reconocimiento de las proteínas con anclaje GPI ha permitido realizar escrutinios a escala en los proteomas deducidos de los proyectos de secuenciación genómica y

transcriptómica. Varios métodos predictivos están disponibles en línea los cuales utilizan la estructura primaria de las proteínas (Cao *et al.*, 2009). Los programas difieren en los criterios para seleccionar una proteína con un posible anclaje a GPI, lo que lleva a diferencias en sus algoritmos (Cuadro 2.1).

Cuadro 2.1 Principales programas bioinformáticos para la identificación *in silico* de proteínas con anclaje a GPI

Programa	Característica	Referencia
GPI-SOM	Utiliza ambas regiones hidrofóbicas localizadas en las proteínas de anclaje GPI, usa la región C-terminal de las secuencias proteicas y las analiza mediante el uso de mapas de autoorganización y la región N-terminal para la predicción del péptido señal de secreción.	http://gpi.unibe.ch/
BIG-PI	Utiliza una función de predicción, basada en la valoración de la preferencia en la ubicación de cierto tipo de aminoácidos dentro de la secuencia cerca del sitio ω , así como las propiedades físicas que codifican múltiples residuos dentro de la secuencia motive.	http://mendel.imp.ac.at/gpi/fungi_server.html
PRED-GPI	Utiliza un conjunto de algoritmos denominados “máquinas de vectores de soporte” que permite discriminar las señales de anclaje en la proteína, acoplado a un modelo oculto de Markov para la predicción más probable del sitio ω .	http://gpcr.biocomp.unibo.it/pr-edgpi/
FRANGANCHOR	Está basado en el uso en tándem de predictores de redes neuronales y modelos ocultos de Markov, la primera selecciona las secuencias potenciales para el anclaje de la GPI y la segunda clasifica las secuencias seleccionadas de acuerdo a 4 niveles diferentes de precisión de la predicción	http://navet.ics.hawaii.edu/~franganor/NNHMM/NNHMM.html

Estos programas, utilizados en conjunto con otras herramientas bioinformáticas como SignalP o Wolf Psort, pueden aumentar la confiabilidad de la predicción (Kantún *et al.*, 2012).

En el caso de los hongos la ubicación final de las proteínas con anclaje GPI no se limita a la membrana plasmática, algunas continúan su procesamiento y son liberadas hacia la pared celular, a la cual se enlazan mediante el anclaje β -1,6-glucano (Damveld *et al.*, 2005). Sin embargo, aún es desconocida la maquinaria que lo lleva a cabo y cuáles son las señales a reconocer para que este procesamiento se realice. Diferentes estudios acerca de las características presentes en estas proteínas y en la secuencia de los aminoácidos cercanos al sitio ω han propuesto mecanismos que influyen en la ubicación final en la célula de las proteínas con anclaje GPI: a) la presencia de un motivo dibásico río arriba del sitio ω en las proteínas con anclaje GPI podría determinar su ubicación en la membrana plasmática (Frieman y Cormack, 2003), b) la presencia de una región espaciadora rica en serina/treonina entre la secuencia señal de secreción en el N-terminal y la región C-terminal con el sitio ω , aunado a la pérdida del motivo dibásico, favorece la ubicación en la pared celular (Kliss *et al.*, 2009; Caro *et al.*, 1997), c) diferencias significativas en el punto isoeléctrico entre las proteínas con anclaje GPI localizadas en pared (4.865 ± 0.2194) y las de membrana plasmática (6.667 ± 0.9454) podrían servir como parámetros para su clasificación (Pittet y Conzelmann, 2007).

Al inicio del trabajo, se eligió realizar la búsqueda en el genoma de *T. reesei* debido a que se desconocía con que especies de *Trichoderma* contábamos y en la literatura se ha reportado secreción de lipasas en esta especie y es una de las plataformas de expresión de proteínas heterólogas más usadas a nivel industrial (Qin *et al.*, 2012; Rajesh *et al.*, 2010). Con el fin de incrementar la probabilidad de éxito en la identificación correcta de proteínas con anclaje GPI y favorecer la selección de proteínas que se localicen en la pared celular, en el presente trabajo se emplearon varias herramientas bioinformáticas.

2.2 MATERIALES Y MÉTODOS

2.2.1 Generación de base de datos con proteínas con anclaje GPI previamente reportada en hongos y búsqueda de homólogas en el genoma de *T. reesei*

Se realizó una revisión bibliográfica en artículos científicos sobre diferentes hongos que reportan proteínas con anclaje GPI, ya fueran de manera predictiva o confirmadas experimentalmente. Así mismo, se accedió a la base de datos reportada por Eisenhanber y colaboradores (2004), la cual contiene secuencias de proteínas con anclaje GPI de *S. cerevisiae*, *S. pombe*, *C. albicans*, *N. crassa* y *A. nidulans*. Todas las secuencias fueron convertidas a formato FASTA y por medio de análisis con BLASTp (<https://genome.jgi.doe.gov/pages/blast-query.jsf?db=Trire2>) se buscaron sus homólogas en la base de datos de *T. reesei*, tomando como umbral máximo de $E < 1e-3$ (Kantún, 2012).

2.2.2 Búsqueda de proteínas con anclaje GPI previamente anotadas en el genoma de *T. reesei*

Debido a que no se habían identificado las cepas con las que se contaban en el laboratorio, se decidió trabajar con *T. reesei* al ser un hongo modelo y ampliamente utilizado para la expresión de proteínas recombinantes, además de que no se han realizado trabajos previos enfocados a la identificación de las proteínas con anclaje GPI de su pared. Con la versión 2.0 del genoma de *T. reesei* (<http://genome.jgi-psf.org/Trire2/Trire2.home.html>) se realizó una búsqueda de proteínas utilizando las palabras claves: “cell wall protein”, “GPI protein” y “GPI cell wall protein”. Los datos obtenidos se convirtieron a formato FASTA y se analizaron posteriormente.

2.2.3 Análisis bioinformático para la predicción de proteínas con anclaje GPI

Se utilizaron 6 programas bioinformáticos para predecir de manera robusta las proteínas con anclaje GPI; estos predictores se basan en las características estructurales de las proteínas con anclaje GPI:

- Tres predictores para determinar el sitio de anclaje del GPI: Big-PI fungal predictor (http://mendel.imp.ac.at/gpi/fungi_server.html), the GPI-SOM predictor (<http://gpi.unibe.ch/>) y predGPI predictor (<http://gpcr.biocomp.unibo.it/predgpi/>), cada uno con algoritmos distintos para la identificación de las proteínas con anclaje GPI.

- Se utilizó el programa SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) para determinar el péptido señal de secreción dentro de la región N-terminal de la secuencia proteica y el posible sitio de corte;
- Para la determinación de la ubicación de la proteína dentro de célula fúngica se utilizó el programa WolfPSort (<http://wolffpsort.seq.cbrc.jp/>). La cuantificación del contenido de serina y treonina se llevó a cabo mediante la herramienta ProtParam tool (<http://web.expasy.org/protparam/>).

2.2.4 Análisis *in silico* para la identificación de funciones putativas y puntos isoeléctricos

Con el fin de caracterizar las posibles proteínas con anclaje GPI identificadas, se analizaron con la herramienta ProtParam tool (<http://web.expasy.org/protparam/>) para determinar punto isoeléctrico. La identificación de dominios y/o motivos característicos de las proteínas se realizó utilizando los programas Pfam (<https://pfam.xfam.org/>), UniProt (<http://www.uniprot.org/>) y CDD: *conserved domain database* (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), la búsqueda de homólogos caracterizados se realizó utilizando las herramientas de BLASTp del NCBI ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_PR OGRAMS=blastp&PAGETYPE=BlastSearch&DATABASE=refseq_protein&DESCRPTIO NS=100](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGETYPE=BlastSearch&DATABASE=refseq_protein&DESCRIPTIONS=100)) y la descripción KOG se tomó directamente de las anotaciones de las proteínas en la base de datos del genoma de *T. reesei* V2.0 .

2.2.5 Identificación molecular de cepas de *Trichoderma sp*

Se trabajó con 6 cepas que presentaban morfología similar a la reportada para el género de *Trichoderma*, tres aisladas de residuos agroindustriales de plátano (Castillo-Avila 2010) y tres aisladas en medio mínimo adicionado con aceite al 1%. Las cepas se mantuvieron en medio de cultivo PDA a 4 °C.

Para la extracción de ADN genómico las cepas se crecieron en PDA e incubaron por tres días a temperatura ambiente. Posteriormente se tomó un bloque de micelio y se disgregó en PDB previamente esterilizado. Con el micelio macerado se inocularon matraces con 100 mL de medio PDB y se incubaron por tres días a temperatura ambiente a 130 rpm. Se tomaron muestras de cada cepa y se lavaron por triplicado con agua estéril para eliminar los posibles restos del medio de cultivo. Se utilizó el protocolo de extracción reportado por Conde-Ferrández y colaboradores (2008); la pureza de las muestras se cuantificó mediante

Nanodrop (Thermo Scientific, U.S.A., Wilmington) en las longitudes de onda 260/280 y 230/260.

La identificación de los hongos se llevó a cabo mediante la amplificación por la técnica de PCR de la región comprendida entre el ITS1 e ITS4; el producto de PCR se visualizó en gel de agarosa al 1% y se purificó siguiendo las indicaciones del Kit QIAquick Gel extraction (QIAGEN). La secuenciación de los productos amplificados se realizó por el método de sanger en LANBAMA, IPICyT, San Luis Potosí.

2.2.6 Identificación de posible proteínas con anclaje GPI en *T. longibrachiatum* homólogas a las identificadas en *T. reesei*

Debido a que no se encontró una cepa de *T. reesei* dentro de la colección, se decidió hacer la búsqueda en *Trichoderma longibrachiatum* de las proteínas homólogas a las previamente identificadas en el proteoma deducido de *T. reesei* como posibles proteínas con anclaje GPI, ya que este es el más cercano filogenéticamente a *T. reesei* (Joint Genome Institute, 2014; Druzhinina et al. 2005) y no existen estudios sobre proteínas con anclaje GPI en esta especie.

Se seleccionaron cinco proteínas de las analizadas en el genoma de *T. reesei* que dieron positivo con todos los programas utilizados para identificar proteínas con anclaje GPI, tres con un contenido superior del 30% de S/T y dos con un contenido menor al 30%. Las secuencias proteicas se usaron como “query” para un “BLASTp” en la base de datos del genoma de *T. longibrachiatum* (<http://genome.jgi.doe.gov/Trilo1/Trilo1.home.html>). Las nuevas secuencias obtenidas fueron analizadas con los programas bioinformáticos previamente mencionados

2.2.7 Diseño de los cebadores para el aislamiento de proteínas con anclaje GPI en *T. longibrachiatum*

Se seleccionaron 7 proteínas que presentaron homología a las identificadas en *T. reesei*, las cuáles cumplieron con todos los parámetros establecidos. En la base de datos del genoma de *T. longibrachiatum* (<http://genome.jgi.doe.gov/Trilo1/Trilo1.home.html>) se identificó la región codificante de cada proteína. Se diseñó un juego de cebadores para cada una, abarcando la región codificante de la proteína con anclaje GPI madura, es decir, sin la región que codifica el péptido de secreción. Los iniciadores se diseñaron usando el programa oligoanalyzer (<https://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>) y OLIGO Explorer 1.5. La

lista de los cebadores diseñados para el aislamiento de la secuencia se encuentra descrita en el Anexo 1.

2.2.8 Aislamiento de ARN total y síntesis de ADNc de *T. longibrachiatum*

La cepa de *T. longibrachiatum* se cultivó en PDA durante tres días, se tomó un fragmento de micelio y se disgregó por macración en mortero. Se inoculó 500 µL del disgregado en matraces de 250 mL con 50 mL de medio PDB y se incubaron a temperatura ambiente con agitación continua a 110 rpm. La extracción del ARN total se realizó con el paquete Illustra RNAspin Mini RNA Isolation Kit (GE Healthcare); se extrajeron tres tiempos de cultivo, 24 h, 72 h y micelio de 96 h al cual se le aplicó un daño mecánico a las 72 h y se recuperó 24 h después del daño mecánico, ya que las proteínas con anclaje GPI pudieran ser expresadas en el proceso de reparación de la pared. La síntesis del ADNc se llevó a cabo mediante el paquete Maxima First Strand cDNA Synthesis kit for RT-qPCR (Thermo Scientific).

2.2.9 Amplificación de la región codificante de las proteínas con anclaje GPI

Se realizaron PCRs de gradiente para determinar la temperatura óptima en la amplificación de las regiones codificantes de las proteínas con anclaje GPI, usando inicialmente ADN genómico. Obtenida la temperatura óptima se llevó a cabo la amplificación utilizando el ADNc; los amplicones se visualizaron en un gel de agarosa de 1% y la purificación de las bandas para su clonación, se llevaron a cabo mediante el paquete QIAquick Gel extraction Kit (Qiagen).

2.3 RESULTADOS

2.3.1 Identificación *in silico* de proteínas con anclaje GPI

La búsqueda en el genoma de *T. reesei* mediante palabras claves como “*cell wall protein*”, “*GPI protein*” y “*cell wall GPI protein*”, dio como resultado 665 proteínas.

Con base en una revisión bibliográfica se obtuvo una lista de 717 proteínas con anclaje GPI de 7 especies fúngicas (Cuadro 2.2). Estas secuencias fueron usadas para la búsqueda de proteínas homólogas en la base de datos no redundante de *T. reesei* v2.0 (<https://genome.jgi.doe.gov/Trire2/Trire2.home.html> media); el análisis BLASTp, identificó 317 secuencias homólogas.

Cuadro 2.2 Proteínas con anclaje GPI fúngicas previamente predichas en distintos hongos.

Referencia	Microorganismo	GPI reportadas
Kliss <i>et al.</i> , 2011	<i>Candida albicans</i>	46
Adachi <i>et al.</i> , 2008	<i>Aspergillus oryzae</i>	5
Kantún-moreno <i>et al.</i> , 2012	<i>Mycosphaerella fijensis</i>	49
	<i>Aspergillus nidulans</i>	
	<i>C. albicans</i>	
Eisenhaber <i>et al.</i> , 2004	<i>Neurospora crassa</i>	616
	<i>Saccharomyces cerevisiae</i>	
	<i>Schizosaccharomyces pombe</i>	
Total		717

Sumando las 665 identificadas con palabras clave más las 317 homólogas a las encontradas en la literatura, se obtuvo una lista de 982 posibles proteínas GPI para *T. reesei*. Estas se analizaron mediante los 6 programas bioinformáticos previamente

seleccionados, identificando 43 candidatas que cumplieron los criterios de selección, es decir, dando positivo en los tres programas predictores del sitio ω , presencia de la secuencia señal de secreción en el N-terminal y localización extracelular.

Las secuencias se clasificaron de la siguiente manera:

- Secuencias proteicas que dieron positivo en los tres programas para predicción del sitio de anclaje GPI, con secuencia señal de secreción, localización extracelular y contenido de treonina/serina > 30% (+>30)
- Secuencias proteicas que dieron positivo en los tres programas para predicción del sitio de anclaje GPI, con secuencia señal de secreción, localización extracelular y contenido de treonina/serina < 30% (+<30)
- Secuencias proteicas que dieron negativo en alguno de los programas bioinformáticos, ya sea para la identificación de sitio ω o la ausencia de la secuencia señal de secreción. (+-)
- Secuencias proteicas que dieron negativo con los programas bioinformáticos utilizados (-)

La figura 2.1 presenta los resultados obtenidos.

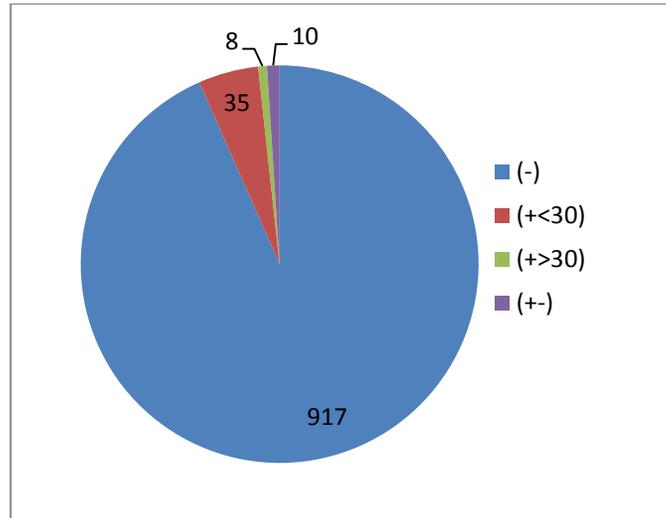


Figura 2.1 Clasificación de las proteínas de *T. reesei* con base en los resultados de análisis con los predictores GPI-som, predGPI, biGPI fungal predictor, SignalP, Wolfpsort y Protparam tool. En azul: secuencias negativas (93.4%); En café: (+<30)= 4.9%; En Morado: (+-) = 1% y en Verde: (+>30)= 0.7%.

2.3.2 Identificación *in silico* de las posibles funciones de las proteínas predichas con anclaje GPI con base en sus dominios y proteínas homólogas caracterizadas

Con el fin de deducir la ubicación en la célula y la posible función de las proteínas seleccionadas se analizaron búsquedas de dominios característicos, y homologías con proteínas conocidas y predicción de sus puntos isoeléctricos. El análisis *in silico* permitió identificar en el proteoma deducido de *T. reesei* proteínas homólogas a proteínas caracterizadas con el anclaje GPI como GAS1, GEL1, GEL2, CFMC, CFMA y CRF1 (cuadro 2.3). Quince de las proteínas inicialmente seleccionadas no presentaron ningún dominio característico dentro de su secuencia; estas tampoco tuvieron algún homólogo caracterizado y algunas tampoco tienen clase asignada en el KOG, dificultando su caracterización *in silico*. De las ocho proteínas predichas con anclaje GPI con un porcentaje mayor del 30% de serina/treonina identificadas, cuatro presentan un dominio CFEM (*Common in Fungal Extracellular Membrane*) el cual es específico para hongos; este presenta 8 residuos de cisteína en una secuencia de aproximadamente 60 amino ácidos, siendo identificado principalmente en proteínas con anclaje GPI localizadas en la pared celular (Vaknin *et al.*, 2014; Kulkarni *et al.*, 2003). En las proteínas predichas con anclaje GPI y con un contenido menor del 30% de serina/treonina se encontraron como dominios más característicos a los de la familia Glyco_hydro 16, éste ha sido identificado en proteínas con anclaje GPI como CRH1 y CRH2 (quitina transglicosilasas), las cuales están relacionadas en la formación de los enlaces en la quitina (Zhang *et al.*, 2013; Cabib *et al.*, 2006); Glyco_hydro 76, presente en Endo- α -1,6-mananasas como DFG5 y DCW1, las cuales son requeridos en la incorporación eficientes de glicoproteínas en la pared de *N. crassa* (Maddi, Fu, y Free 2012). También se encontraron dominios ASP (cl0129) y SAP_Like (cd05474), los cuales están presentes en YPS1, proteasa aspártica identificada en la membrana plasmática de *S. cerevisiae* (Sievi *et al.*, 2001).

Cuadro 2.3 Lista de las posibles proteínas con anclaje GPI identificadas en el proteoma deducido de *Trichoderma reesei*.

Proteína Id	Homólogo con el que se identificó			Homólogo caracterizado o más cercano			Dominio	Clase KOG	pI/ Motivo dibásico
	Organismo (Proteína)	Valor E	Referencia	Organismo (Proteína)	Valor E	Referencia			
Posibles proteínas con anclaje GPI con porcentaje mayor al 30% de treonina/serina									
104401	<i>Mycosphaerella fijiensis</i> (86758)	7.58e ⁻⁰¹⁵	Kantún-Moreno <i>et al.</i> , 2013	<i>Aspergillus fumigatus</i> CFMC (Q4WNE1)	4.2e ⁻²⁶	Vaknin <i>et al.</i> , 2014	CFEM (pfam05730) DUF4115 super family (cl26170)	Transporte y metabolismo de carbohidratos	4.21/ No
102837	<i>M. fijiensis</i> (86075)	2.60e ⁻⁰¹¹	Kantún-Moreno <i>et al.</i> , 2013	<i>A. fumigatus</i> CFMA (Q4WLB9)	7.3e ⁻³²	Vaknin <i>et al.</i> , 2014	CFEM (pfam05730)	Transporte y metabolismo de carbohidratos	4.49/ No
123475	<i>Neurospora crassa</i> (XP_964641.1)	4.14e ⁻⁰¹⁰	Eisenhaber <i>et al.</i> , 2004	<i>Escovopsis weberi</i> hipotética (KOS21701.1)	4e ⁻²⁴	Anotada	CFEM (pfam05730) Atrophin-1 super family (Cl26464)	Mecanismos de defensa	3.54/ No
123039	Identificación por palabras claves (CFEM)			<i>Purpureocillium lilacinum</i> (XP_018175044.1)	2e ⁻³¹	Anotada	CFEM (pfam05730)	Mecanismo de traducción de señales	4.65/ No
122870	<i>Aspergillus nidulans</i> FGSC (AN363.1)	6.66e ^{-0.12}	Eisenhaber <i>et al.</i> , 2004	<i>Metarhizium anisopliae</i> enzima putativa para la biosíntesis de sideróforos (KFG85760.1)	1e ⁻⁴²	Anotada	No identificado	Transporte y metabolismo de carbohidratos	4.99/ No
21412	<i>A. nidulans</i> (AN363.1)	6.04e ^{-0.006}	Eisenhaber <i>et al.</i> , 2004	<i>Penicillium brasilianum</i> hipotética proteína PEBR (OOQ85659.1)	2e ⁻³⁵	Anotada	No identificado	Transporte y metabolismo de iones inorgánicos	4.74/ No
119568	<i>N. crassa</i> proteína 2 de anclaje a pared celular (NCU00957.1)	8.51e ⁻⁰¹³	Eisenhaber <i>et al.</i> , 2004	<i>Fusarium oxysporum</i> proteína hipotética FOTG (EXM27521.1)	2e ⁻⁵⁰	Anotada	GPI-anchored (CL0538)	Transporte y metabolismo de iones inorgánicos	5.57/ No

106223	<i>N. crassa</i> (NCU03773)	1.30e ⁻⁰¹¹	Eisenhaber <i>et al.</i> , 2004	<i>Colletotrichum orbiculare</i> proteína hipotética_09108 (ENH82681.1)	7e ⁻¹⁹	Anotada	Probable pectinesterase (PLN02217) Superantigen-like protein (PRK13335)	Transporte y metabolismo de iones inorgánicos	4.62/ No
Posibles proteínas con anclaje GPI con porcentaje menor al 30% de Treonina/Serina									
81296	<i>M. fijiensis</i> (64178)	5.56e ⁻⁰¹⁶	Kantún-Moreno <i>et al.</i> , 2013	<i>F. oxysporum</i> proteína hipotética FOXB (EGU80058.1)	4e ⁻⁷⁸	Anotada	No identificado	No identificado	8.66/ No
123659	<i>M. fijiensis</i> (86151)	3.05e ⁻⁰⁴⁶	Kantún-Moreno <i>et al.</i> , 2013	<i>Tolypocladium ophioglossoides</i> manoproteína de la pared celular (KND88621.1)	3e ⁻⁸⁴	Anotada	PIR (IPR000420) Atropin-1 super family (cl26464)	Mecanismos de defensa	4.49/ No
120189	<i>M. fijiensis</i> (82178)	9.79e ⁻⁰¹⁴	Kantún-Moreno <i>et al.</i> , 2013	<i>T. ophioglossoides</i> proteína hipotética TOPH (KND93090.1)	2e ⁻³⁹	Anotada	No identificado	Mecanismo de traducción de señales	4.23/ No
106024	<i>M. fijiensis</i> (210373)	4.38e ⁻⁰⁰⁶	Kantún-Moreno <i>et al.</i> , 2013	<i>Pochonia chlamyosporia</i> proteína de adhesión Mad2 (XP_018145462.1)	1e ⁻²⁹	Anotada	DNA_pol3_gamma3 (cl26386)	Estructuras extracelulares	5.29/ No
22564	<i>M. fijiensis</i> (86794)	8.11e ⁻⁰²¹	Kantún-Moreno <i>et al.</i> , 2013	<i>Ophiocordyceps australis</i> proteína hipotética (PHH75080.1)	3e ⁻⁵⁸	Anotada	No identificado	No identificado	4.85/ No
79448		8.00e ⁻⁰¹⁸	Kantún-Moreno <i>et al.</i> , 2013	<i>Nectria haematococca</i> proteína hipotética (XP_003054617.1)	8e ⁻⁶⁷	Anotada	No identificado	No identificado	5.09/ No
119975	<i>Saccharomyces cerevisiae</i> SED1	2.09e ⁻⁰⁰⁷	Hagenet <i>et al.</i> , 2004	<i>Purpureocillium lilacinum</i> proteína hipotética (XP_018175390.1)	5e ⁻³⁵	Anotada	Pecyinesterase super family (cl27601)	Mecanismos de defensa	6.50 / No

108642	<i>Trichoderma reesei</i> (119975)	9.77e ⁻⁰¹¹	Presente estudio	<i>P. chlamydosporia</i> proteína hipotética (XP_018141962.1)	7e ⁻¹⁰	Anotada	No identificado	Mecanismos de defensa	5.63/No
122076	<i>A. nidulans</i> (AN1573.1)	4.61e ⁻⁰⁴⁴	Eisenhaber et al., 2004	<i>Neurospora tetrasperma</i> proteína hipotética (XP_009856979.1)	0.0	Anotada	SAP_like (cd05474) Asp(cl0129)	Modificación postraducciona l, Chaperonas	4.25/No
121306		1.87e ⁻⁰²²	Eisenhaber et al., 2004	<i>Purpureocillium lilacinum</i> peptidasa A1 (XP_018176612.1)	0.0	Anotada	Pepsin_retropepsin_like super family (cl11403) DNA_pol3_gamma3 super family (cl26386)	Modificación postraducciona l, Chaperonas	4.73/No
119876		2.37e ⁻⁰⁴³		<i>Purpureocillium lilacinum</i> peptidasa aspártica (OAQ85013.1)	2e ⁻¹⁴⁰	Anotada	SAP_like (cd05474) COG3577 super family (cl26166) Asp(cl0129)	Modificación postraducciona l, Chaperonas	4.37/No
123538	<i>A. nidulans</i> Gel1 (AN7657.1)	9.91e ⁻¹²⁸	Eisenhaber et al., 2004; Fontaine, 2003	<i>P. lilacinum</i> 1,2-beta-glucanosiltransferasa gel1 (XP_018175330.1)	0.0	Anotada	Glyco_hydro_1 super family (cl23gh) Glyco_hydro 72 (cl0058)	No identificado	4.99/No
77942		6.41e ⁻⁰⁶²		<i>A. fumigatus</i> gel2 (POC954)	1.6e ⁻¹⁵⁴	Mouyna et al., 2005	Glyco_hydro_1 super family (cl2372) Glyco_hydro 72 (cl0058)	No identificado	4.83/No
74807	<i>Candida albicans</i> DFG5 (Q5ACZ2)	1.26e ⁻⁰¹⁰	Eisenhaber et al., 2004 Spreghini et al., 2003	<i>P. lilacinum</i> glucósido hidrolasa (XP_018183717.1)	3e ⁻¹⁶⁸	Anotada	Glyco_hydro 76 (cl0059) LanC_like super family (cl04955)	No identificado	5.51/No
122495		4.81e ⁻⁰⁶⁷		<i>N. haematococca</i> glucósido hidrolasa familia76 (XP_003048447.1)	0.0	Anotada	Glyco_hydro76 (cl0059)	No identificado	5.10/No
70542	<i>N. crassa</i> (NCU06504.1)	2.63e ⁻⁰⁷⁸	Eisenhaber et al.	<i>N. haematococca</i> proteína	1e ⁻¹²²	Anotada	Glyco_hydro 16 (cl0004)	Metabolismo de	4.72/No

58239	<i>A. nidulans</i> CRF1(AN0933.1)	1.06e ⁻⁰⁶⁴	Eisenhaber et al., 2004	hipotética(XP_003049016.1) <i>P. chlamydosporia</i> glucósido hidrolasa (XP_018147080.1)	0.0	Anotada	GH16_fungal_C RH1_transglicosilasa (cd02183) Glyco_hyro 16 (cl0004)	Metabolismo de carbohidratos	4.88/ No
76266	<i>C. albicans</i> CRH11 (orf19.2706)	3.84e ⁻⁰²⁷	Kliss et al., 2011	<i>A. fumigatus</i> crf2 (Q4WI46)	1.5E ⁻¹⁴²	Schütte et al., 2009	ChtBD1 (cl16916) GH16_fungal_C RH1_transglicosilasa (cd02183) Glyco_hydro 16 (CL0004)	Metabolismo de carbohidratos	4.62 /No
105444	Palabras calve (CFEM)			<i>Penicillium expansum</i> proteína hipotética (XP_016600131.1)	7e ⁻⁵⁷	Anotada	No identificado	No identificado	4.91/ No
121251	<i>A. nidulans</i> (AN3141.1)	2.29e ⁻⁰⁴⁰	Eisenhaber et al., 2004	<i>Fusarium verticillioides</i> proteína hipotética (XP_018745891.1)	2e ⁻¹³⁹	Anotada	CFEM (pfam0570) Atrophin-1 (cl26464)	Mecanismos de defensa	4.63/ No
22386	<i>A. nidulans</i> (AN4702.1)	4.37e ⁻⁰¹⁰	Eisenhaber et al., 2004	<i>N. haematococca</i> proteína hipotética (XP_003054568.1)	3e ⁻⁹⁰	Anotada	No identificado	No identificado	4.47/ Si
122874	<i>A. nidulans</i> (AN5290.1)	1.56E ⁻⁰⁰⁹	Eisenhaber et al., 2004	<i>Verticillium dahliae</i> proteína hipotética (XP_009650629.1)	2e ⁻¹⁹	Anotada	No identificado	No identificado	3.96/ No
124256	<i>A. nidulans</i> (AN7782.1)	8.22e ⁻¹⁷³	Eisenhaber et al., 2004	<i>P. lilacinum</i> precursor de lisofosfolipasa (XP_018181292.1)	0.0	Anotada	Patatin_and_cP LA2 (cl11396) PLA2_B (cl0323)	Metabolismo y transporte de lípidos	4.78/ No
53947	<i>A. nidulans</i> (AN9450)	3.82e ⁻⁰¹²	Eisenhaber et al., 2004	<i>F. oxysporum</i> proteína hipotética (XP_018232355.1)	3e ⁻³⁹	Anotada	No identificado	Estructura y dinámica de la cromatina	5.46/ No
120821	<i>N. crassa</i> (ncu09133)	1.27e ⁻⁰¹⁶	Eisenhaber et al., 2004	<i>P. chlamydosporia</i> síntesis de Beta-	4e ⁻⁴⁵	Anotada	GPI-anchored (pfam10342)	Mecanismo de defensa	5.46/ No

			<i>al.</i> , 2004	glucanos de la pared celular (XP_018148390.1)					
81541	<i>N. crassa</i> (ncu09647.1)	3.56e ⁻⁰⁹⁶	Eisenhaber <i>et al.</i> , 2004	<i>P. lilacinum</i> glucósido hidrolasa (XP_018174743.1)	7e ⁻¹⁵²	Anotada	Glyco_hydro_cc (pfam11790)	No identificada	4.84/No
44214	<i>N. crassa</i> (NCU09663)	5.64e ⁻⁰⁴⁸	Eisenhaber <i>et al.</i> , 2004	<i>P. lilacinum</i> carbohidrato esterasa familia 5 (XP_018180086.1)	3e ⁻⁸⁰	Anotada	Cutinase (cl0028)	Función desconocida	4.51/No
21468	<i>N. crassa</i> (NCU09929)	1.43e ⁻⁰²⁸	Eisenhaber <i>et al.</i> , 2004	<i>N. haematococca</i> proteína hipotética (XP_003046002)	3e ⁻⁸⁴	Anotada	No identificado	Biogénesis/desarrollo/pared celular/membrana	4.94/No
111803	<i>N. crassa</i> (NCU06962.1)	2.70e ⁻⁰⁰⁷	Eisenhaber <i>et al.</i> , 2004	<i>F. oxysporum</i> proteína hipotética (XP_018243217)	5e ⁻²⁴	Anotada	No identificado	No identificada	4.47/No
105763	<i>N. crassa</i> (NCU07962.1)	2.18e ⁻⁰⁰⁶	Eisenhaber <i>et al.</i> , 2004	<i>P. lilacinum</i> proteína con dominio CFEM (XP_018182706.1)	2e ⁻¹¹	Anotada	CFEM (pfam05730)	Transcripción	4.89/No
122455	<i>N. crassa</i> (NCU014039)	3.93e ⁻⁰⁵⁴	Eisenhaber <i>et al.</i> , 2004	<i>Metarhizium robertsii</i> proteína hipotética (XP_011411371.1)	5e ⁻¹²³	Anotada	No identificado	Mecanismo de transducción de señales	4.83/No
63416	<i>N. crassa</i> (NCU04496)	5.72e ⁻⁰⁴¹	Eisenhaber <i>et al.</i> , 2004	<i>Sordaria macrospora</i> k-hell proteína hipotética (XP_003350213.1)	2e ⁻⁵⁰	Anotada	Cupredoxin (cd00920)	No identificado	7.74/No
64181	<i>N. crassa</i> (NCU04496)	6.13e ⁻⁰²³	Eisenhaber <i>et al.</i> , 2004	<i>P. chlamydosporia</i> proteína extracelular rica en serina (XP_018142414.1)	4e ⁻⁶²	Anotada	Cupredoxin (cd00920)	No identificado	7.65/No
55443	<i>N. crassa</i> (NCU04496)	5.27e ⁻⁰¹⁹	Eisenhaber <i>et al.</i> , 2004	<i>F. oxysporum</i> proteína hipotética (XP0182427.1)	1e ⁻⁷⁵	Anotada	Cupredoxin (cd00920)	No identificada	4.39/No

81296	<i>N. crassa</i> (NCU02099)	8.63e ⁻¹⁰⁷	Eisenhaber <i>et al.</i> , 2004	<i>F. oxysporum</i> proteína hipotética (XP_018241410)	5e ⁻⁷⁷	Anotada	No identificado	No identificado	8.66/ No
Posibles proteínas con anclaje GPI que dieron negativo en alguno de los programas									
72907	<i>M. fijiensis</i> (89224)	2.39e ⁻⁰⁰⁶	Kantún-Moreno <i>et al.</i> , 2013	<i>P. chlamydosporia</i> glioxal oxidasa (CP_018146738)	9e ⁻⁸⁷	Anotada	WSC domain (pfam01822)	Mecanismos de defensa	5.90/ No
121417	<i>M. fijiensis</i> (119314)	1.03e ⁻⁰⁰⁸	Kantún-Moreno <i>et al.</i> , 2013	<i>Scedosporium apiospermum</i> proteína hipotética (XP_016645747)	9e ⁻⁷⁸	Anotada	Fasciclin (pfam02469)	Biogénesis/des arrollo de pared y membrana celular	8.48/ No
82633	<i>M. fijiensis</i> (209620)	3.40e ⁻¹⁶⁰	Kantún-Moreno <i>et al.</i> , 2013	<i>Arthroderma benhamiae</i> GAS1 (D4ATC3)	2.2e ⁻¹⁸⁰	Tran <i>et al.</i> , 2016	Glyco_hydro_1 (cl23725) X8 (pfam07983)	No identificada	4.42/ No
22914		1.56e ⁻⁰⁵⁵	Kantún-Moreno <i>et al.</i> , 2013	<i>P. chlamydosporia</i> 1,3-beta- glucanosiltransferasa (XP_018143431.1)	0.0	Anotada	Glyco_hydro_1 (cl23725) X8 (pfam07983)	No identificada	4.67/ No
49409	<i>M. fijiensis</i> (204420)	6.77e ⁻⁰⁰⁶	Kantún-Moreno <i>et al.</i> , 2013	<i>Metarhizium robertsii</i> glucósido hidrolasa (XP_007825372)	0.0	Anotada	Glyco_hydro_76 (pfam03663)	No identificada	5.05/ No
55802	<i>M. fijiensis</i> (149838)	1.39e ⁻⁰⁸³	Kantún-Moreno <i>et al.</i> , 2013	<i>N. haematococca</i> proteína hipotética(XP_0030 46699)	0.0	Anotada	Glyco_hydro_ 76 (pfam03663)	No identificada	4.81/ No
67844	<i>M. fijiensis</i> (149838)	6.44e ⁻¹⁰⁶	Kantún-Moreno <i>et al.</i> , 2013	<i>Colletotrichum graminicola</i> glucósido hidrolasa (XP_008090011)	4e ⁻¹⁴⁹	Anotada	Glyco_hydro 76 (pfam03663)	Función desconocida	5.13/ No
69123	<i>M. fijiensis</i> (14983)	3.46e ⁻¹⁰⁶	Kantún-Moreno <i>et al.</i> , 2013	<i>M. robertsii</i> glucósido hidrolasa (XP_007819415)	0.0	Anotada	Glyco_hydro 76 (pfam03663)	No identificada	4.83/ No

109835	<i>M. fijiensis</i> (124305)	6.01e ⁻⁰⁶⁵	Kantún-Moreno <i>et al.</i> , 2013	<i>P. chlamydosporia</i> proteína taumatina (XP_018144149)	0.0	Anotada	Thaumatina (ofan00314)	No identificada	5.25/ No
38441	<i>M. fijiensis</i> (55294)	4.09e ⁻⁰⁰⁷	Kantún-Moreno <i>et al.</i> , 2013	<i>Fusarium fujikuroi</i> relacionado a Uth1p, Nca3p, YIL123w y Sun4p (XP_023425884)	3e ⁻¹⁵⁶	Anotada	SUN (cl04292)	Mecanismo de transducción de señales	4.66/ Si
53495	<i>A. nidulans</i> (AN1330)	7.82e ⁻⁰⁴⁹	Kantún-Moreno <i>et al.</i> , 2013	<i>Metarhizium</i> <i>acridum</i> proteína hipotética (XP_007810071.1)	9e ⁻¹⁴⁷	Anotada	PI_PLCC_GDPD _SF (cl14615)	Metabolismo y transporte de carbohidratos	4.52/ No
66789	<i>A. nidulans</i> (AN4686)	5.47e ⁻⁰¹¹	Eisenhaber <i>et al.</i> , 2004	<i>A. fumigatus</i> CSN- Endo-quitosana (Q4WB37)	4.2e ⁻⁶¹	Eom y Lee, 2003	Glyco_hydro_75 (pfam07335)	No identificada	4.45/ No
21982	<i>A. nidulans</i> (AN4702)	1.51e ⁻⁰⁷⁰	Eisenhaber <i>et al.</i> , 2004	<i>P. lilacinum</i> proteína con anclaje GPI (OAQ78299)	5e ⁻¹¹⁵	Anotada	No identificado	No identificado	5.48/ No
59628	<i>N. crassa</i> (NCU09647)	6.01e ⁻⁰⁸⁸	Eisenhaber <i>et al.</i> , 2004	<i>C. orbiculare</i> proteína con dominio f5.8 (ENH76420.1)	2e ⁻¹²⁴	Anotada	Glyco_hydro_cc (pfam11970)	No identificado	4.92/ No
109235	<i>N. crassa</i> (NCU05609)	1.10e ⁻⁰⁰⁸	Eisenhaber <i>et al.</i> , 2004	<i>P. lilacinum</i> modulo con unión a carbohidrato fam. 18(XP_018183574)	1e ⁻⁶³	Anotada	No identificado	Metabolismo y transporte de carbohidratos	5.88/ No

2.3.2 Identificación molecular de cepas de *Trichoderma* sp

Se obtuvo ADN mediante la extracción de micelio proveniente de medios sumergidos. La figura 2.2 muestra la amplificación de los ITS de las cepas fúngicas con las que se inició este trabajo. Empleando los cebadores ITS1 e ITS4 (Korabecna, 2007; White *et al.*, 1990) se obtuvieron amplicones de aproximadamente 800 pb. Se secuenciaron ambas hebras; los resultados fueron editados para eliminar “ruido” y corregir posibles errores. Para su concatenación la hebra obtenida con el ITS4 fue convertida en la reversa complementaria, obteniéndose una secuencia en la misma dirección y sentido que la secuencia obtenida con el cebador ITS1. Se realizó un análisis BLASTn en la base de datos del NCBI y se confirmó el resultado obtenido con un análisis en el programa TrickOKEY (<http://www.isth.info/tools/molkey/index.php>) de la Subcomisión internacional de taxonomía en *Trichoderma* e *Hipocrea*. Se identificaron 3 especies distintas en las 6 cepas; todas las secuencias mostraron una cobertura $\geq 99\%$ y un porcentaje de identidad $\geq 98\%$ en el BLASTn (Cuadro. 2.4).

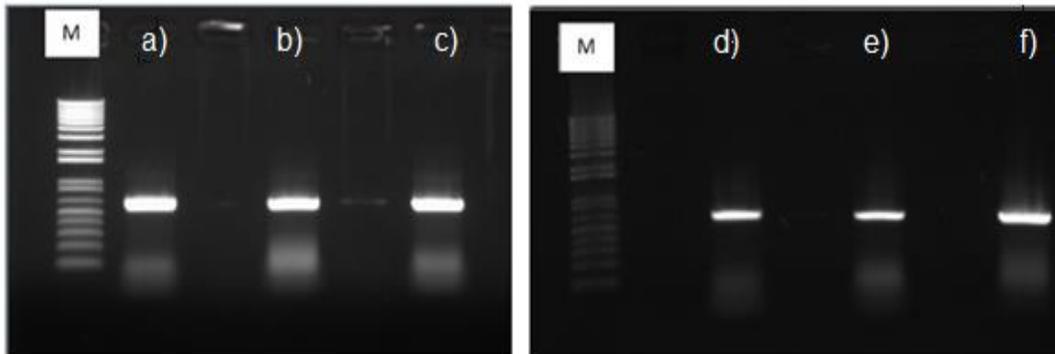


Figura 2.2 Visualización en gel de agarosa del producto de amplificación de los ITS de las 6 cepas de presuntos *Trichoderma* sp. con las que se inició el presente trabajo. a) Cepa B13-1, b) B13-3, c) B13-4, d) AR2-(1), e) Ar3-(1), f) AR7-(1). M) Marcador molecular de 1 Kb plus ladder (ThermoFisher Scientific).

Cuadro 2.4 Identificación molecular de 6 cepas de hongos utilizadas en el presente trabajo.

Cepa	ITS1-ITS4			
	Especie	Query cover	Valor E	Ident
B13-4	<i>T. asperellum</i>	100%	0.0	99%
B13-3	<i>T. longibrachiatum</i>	100%	0.0	99%
B13-1	<i>T. harzianum</i>	99%	0.0	98%
AR7-1	<i>Hypocrea lixii</i>	99%	0.0	100%
AR3-1	<i>T. longibrachiatum</i>	100%	0.0	99%
AR2-1	<i>T. longibrachiatum</i>	100%	0.0	100%

2.3.3 Identificación de posibles proteínas con anclaje GPI en *T. longibrachiatum*, homólogas a las identificadas en *T. reesei*.

Como se mencionó previamente, al inicio del trabajo se eligió realizar la búsqueda en el genoma de *T. reesei* debido a que se desconocía aun con que especies de *Trichoderma* contábamos. Una vez identificadas las especies de *Trichoderma* se seleccionó a *T. longibrachiatum*, cercana filogenéticamente a *T. reesei* y de la cual también se tiene reporte de producción de lipasas (Gochev *et al.*, 2012).

Se seleccionaron del proteoma de *T. reesei* tres proteínas con contenidos de serina/treonina > 30% y que dieron resultados positivos con los programas bioinformáticos de identificación de proteínas de la clase GPIs. También se incluyeron dos proteínas con contenido de serina/treonina entre 20% y 30%. Las 5 proteínas seleccionadas presentaron al menos una secuencia homóloga en *T. longibrachiatum* y sus resultados con los predictores fueron similares a los previamente obtenidos con las proteínas de *T. reesei* con las que tienen identidad. En el cuadro 2.5 se observan las características de las proteínas seleccionadas para el estudio; con el fin de obtener una mayor probabilidad de identificar una proteína con anclaje GPI que funcione para expresión superficial se seleccionaron para continuar el trabajo las siete proteínas de *T. longibrachiatum* que tuvieron homología a las de *T. reesei*.

Cuadro 2.5 Identificación de proteínas de *T. longibrachiatum* homólogas de proteínas con anclaje GPI de *T. reesei*.

Proteína en <i>T. reesei</i>	Proteína en <i>T. longibrachiatum</i>	Valor E	Dominio	Descripción KOG	Tamaño a.a
104401	1401645	4.15 E-59	CFEM	Beta-1,6-N-acetylglucosaminyltransferasa	195
	1436499	1.91E-7			239
119568	2326	1.11E-28	GPI-anchored	Beta-1,6-N-acetylglucosaminyltransferase	245
21412	1397533	9.1E-71	WSC	Ca ²⁺ -modulated nonselective cation channel polycystin	215
	1401002	1.71E-12		Beta-1,6-N-acetylglucosaminultransferase	223
53947	1326581	1.2E-73		Ca ²⁺ -modulated nonselective cation channel polycystin	200
21468	1325672	2E-103		Von Willebrand factor and related coagulation proteins	273

2.3.4 Diseño de cebadores para el aislamiento de proteínas con anclaje GPI en *T. longibrachiatum*

Para el diseño de los cebadores de cada proteína putativa con anclaje GPI, se analizaron las secuencias con el programa SignalP, el cual determina la secuencia señal de secreción (Fig. 2.3. a). Se tomó el codón inmediato al sitio más probable de corte como el inicio del cebador “sentido” y el codón de paro como el sitio de inicio del cebador “antisentido” (Fig. 2.3 b).

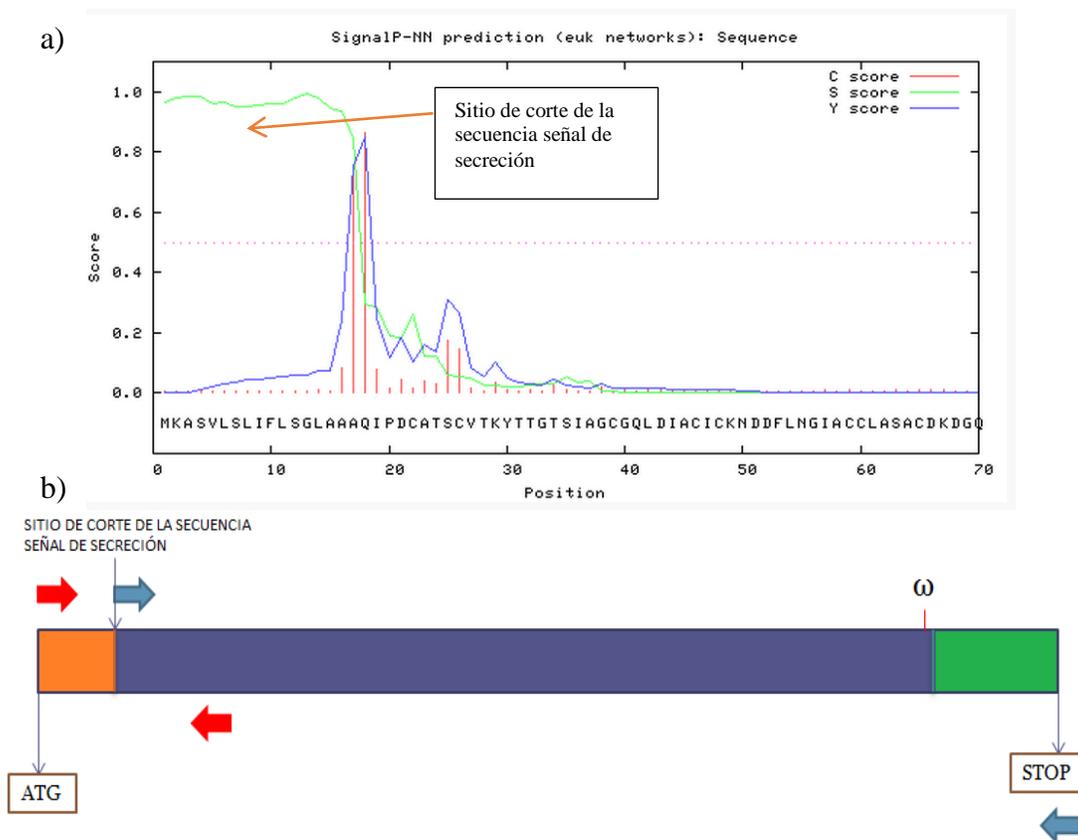


Figura 2.3 a) Identificación mediante el programa SignalP de la secuencia correspondiente a la señal de secreción; b) Esquemmatización de los sitios en los cuales se diseñaron los cebadores para el aislamiento de las secuencias; Flechas rojas: ubicación en la secuencia de los cebadores para el aislamiento de la secuencia señal de secreción. En azul: cebadores para el aislamiento del fragmento codificante para la proteína con anclaje GPI, ω : sitio de anclaje del glicosilfosfatidilinositol (GPI).

2.2.7 Aislamiento de ARN total de *T. longibrachiatum*

Se obtuvo ARN de las tres condiciones propuestas, estas son, micelio de cultivos de 24, 72 y de 72 h al cual se le ocasiono un daño mecánico mediante maceración y se dejó recuperar por 24 h. La integridad del ARN se observa en la figura 2.4, donde se distinguen claramente las bandas 28s y 18s del ARN ribosomal; también se observó posible contaminación por ADN genómico, por lo que las muestras se usaron como templado para amplificación con ITS. Se observó un amplicón de aproximadamente 800 pb

correspondiente a la amplificación con los iniciadores para ITS, lo que confirmó contaminación genómica.

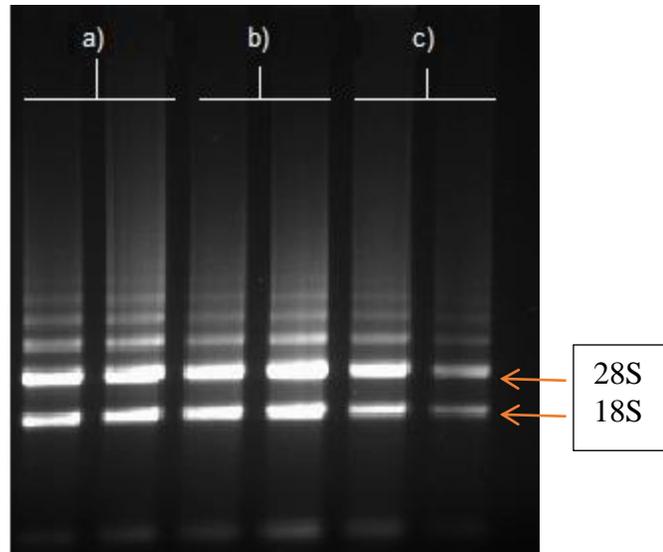


Figura 2.4 Visualización en gel de agarosa al 1% del ARN total aislado de *T. longibrachiatum*: a) ARN total aislado de micelio de 24 h; b) ARN total aislado de micelio de 72 h; c) ARN total aislado de micelio de 96 h (sometido a daño mecánico a las 72 h y posteriormente con 24 h de recuperación).

Se procedió hacer una segunda digestión con ADNasa y se visualizó la integridad mediante gel de agarosa (Fig. 2.5 a), observándose una disminución en la intensidad de las bandas correspondiente a los ARN ribosomales 28S y 18S y un barrido similar al observado en el primer gel. Se usaron nuevamente los iniciadores ITS 1 y 4 para determinar la presencia de ADN genómico; no se observó amplicón en las muestras, pero sí en el control positivo sobre ADN genómico de *T. longibrachiatum* (Fig. 2.5 b). Este resultado indica que el ARN ya no contenía ADNg.

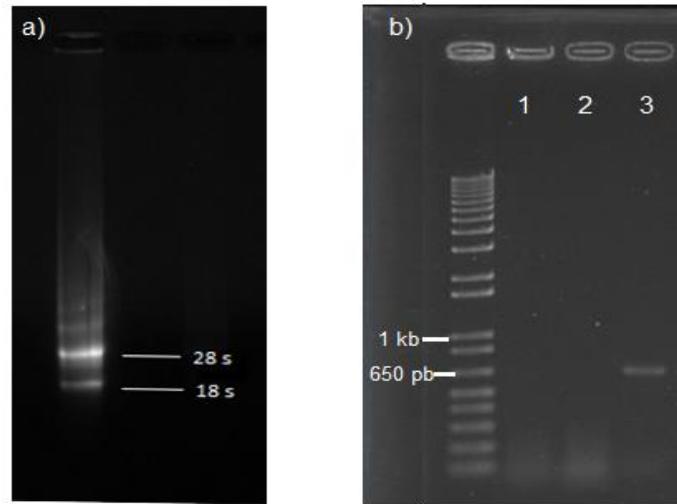


Figura 2.5 Electroforesis para asegurar la integridad y pureza del ARN (toda figura debe llevar un título y posteriormente su descripción. a) Visualización en gel de agarosa de la muestra de ARNt después de segunda digestión con ADNasa; b) Amplificación por PCR usando los iniciadores ITS1 e ITS4 para determinar la contaminación de ADN genómico en las muestras de ARN; 1 y 2) Amplificación de ITS usando como templado muestras de ARN tratadas con ADNasa; 3) amplificación de ITS usando ADN genómico como control positivo.

2.3.6 Amplificación de la región codificante de las proteínas con anclaje GPI

Para descartar por completo contaminación con ADN genómico en las muestras de ADNc se seleccionó los iniciadores para la amplificación del gen 1436399 (Cuadro 2.5), se realizó la amplificación de la secuencia usando ADN genómico y ADNc como templado, observando una diferencia de aproximadamente 200 pb en el tamaño de los amplicones obtenidos debido a la presencia de intrones (Fig. 2.6a), confirmando la ausencia de ADN genómico en las muestra. La ausencia de ADNg se confirmó en todas las muestras siguiendo la misma estrategia.

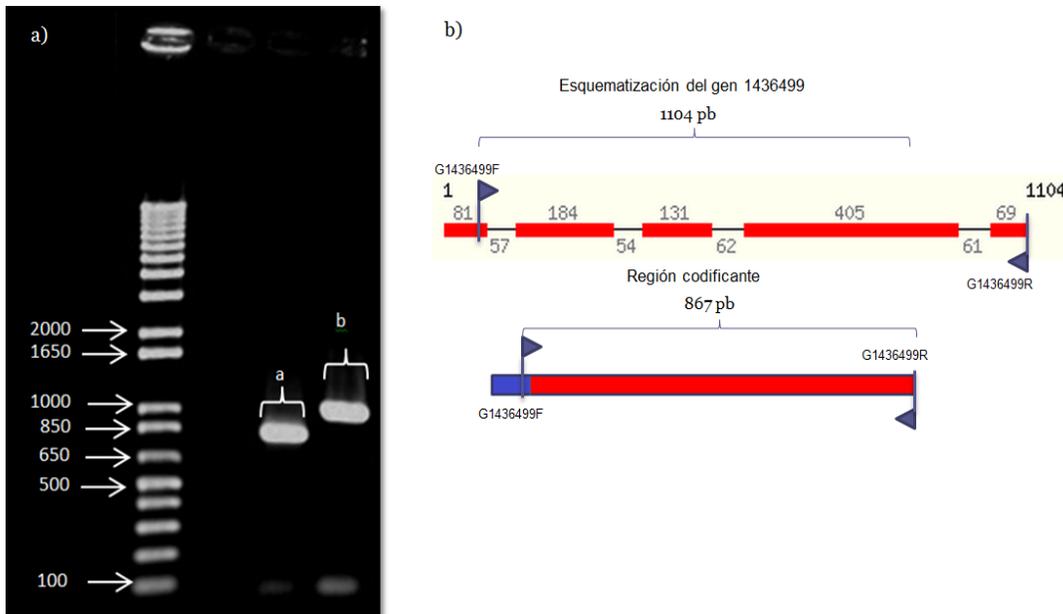


Figura 2.6 Validación de la síntesis de ADNc. a) Amplificación usando los iniciadores para el gen 1436499 que codifica para una proteína putativa con posible anclaje GPI, a: Amplicón usando ADNc con un tamaño de banda esperado de 870 pb aproximadamente; b: amplicón usando ADNg con un tamaño de 1100 pb. b) Representación esquematizada del gen 1436499; parte superior: En rojo: exones; Negro: intrones. Parte inferior: representación de la región codificante, los triángulos representan los sitios donde se diseñaron los cebadores; en azul la secuencia señal de secreción.

Se utilizó como templado para las reacciones de PCR la muestra de ADNc obtenida del micelio con 24 h en cultivo, con la cual se lograron amplificar las 7 regiones codificantes de las proteínas seleccionadas (Fig. 2.7). Todos los amplicones presentaron una banda única del tamaño esperado, exceptuando los amplicones correspondientes a los genes 14011002 y 1326581 que presentaron bandas tenues por encima y de bajo de la banda esperada. Sin embargo, en ambas se observaron los productos de amplificación del tamaño esperado, con mayor intensidad que las bandas inespecíficas.

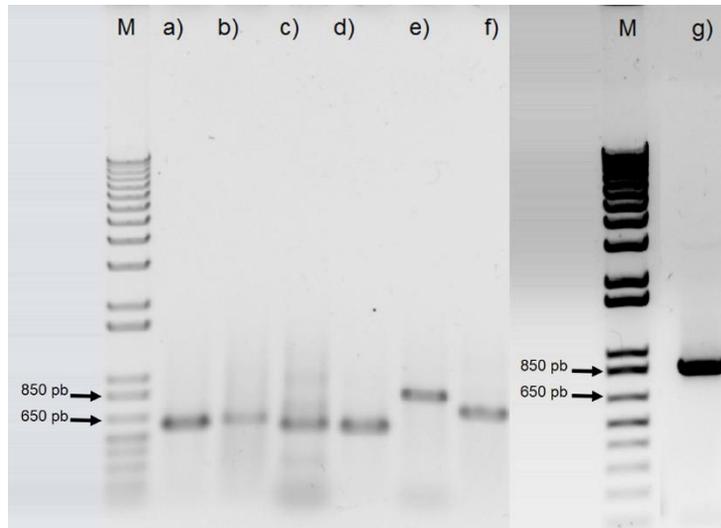


Figura 2.7 Amplificación de las regiones codificantes de las proteínas putativas con anclaje GPI seleccionadas del genoma de *T. longibrachiatum*; a) Amplicón del gen 1397533 (597pb), b) Amplicón del gen 14011002 (630pb), c) Amplicón del gen 1326581 (564 pb), d) Amplicón del gen 1401645 (549 pb), e) Amplicón del gen 1436499 (850 pb) f) Amplicón del gen 2326 (699 pb), g) Amplicón del gen 1325672 (744 pb) y m) marcador de peso molecular.

2.4 DISCUSIÓN

Debido a la importancia de las proteínas con anclaje GPI en el desarrollo y funcionalidad de la célula, la búsqueda de estas proteínas en los diferentes proteomas se ha hecho cada vez más frecuente. Las identificaciones por métodos bioquímicos presentan varios cuellos de botella para la determinación a gran escala, por lo que se han desarrollado análisis bioinformáticos para apoyar la búsqueda.

Para el escrutinio de proteínas con anclaje GPI en el genoma de *T. reesei* se desarrollaron dos estrategias de búsqueda; la primera consistió en obtener las secuencias de proteínas reportadas en diferentes hongos como posibles proteínas con anclaje GPI y su posterior análisis BLASTp para encontrar homólogas en *T. reesei*: La segunda estrategia consistió en el uso de la herramienta “*search*” disponible en la base de datos del genoma de *T. reesei*. Se obtuvieron un total de 717 secuencias, sin embargo, en los análisis bioinformáticas, la segunda estrategia de búsqueda obtuvo un bajo número de secuencias positivas. Por el contrario, la búsqueda de proteínas homólogas en el proteoma de *T. reesei* con aquellas previamente reportadas como GPIs arrojó un mayor número de secuencias potenciales de proteínas con anclaje GPI.

Del total de 982 secuencias analizadas, 43 cumplieron los criterios para identificar proteínas con anclaje GPI, incluyendo la identificación del sitio ω . El resultado obtenido en el presente trabajo es similar al reportado por Kantún-Moreno (2013) la cual analizó el proteoma de *M. fijiensis* empleando los mismos métodos bioinformáticas y con los que obtuvo un total de 49 presuntas proteínas con anclaje GPI. Por su parte (Zhang *et al.*, 2013) realizaron un escrutinio del proteoma predicho de *Pichia pastoris*, utilizando los mismos parámetros para considerar cuales pueden ser proteínas con anclaje GPI y una nueva herramienta bioinformática para la predicción de proteínas con anclaje GPI (ProFASTA). Inicialmente identificaron 61 proteínas que cumplieran con la presencia del sitio ω , de las cuales 11 fueron descartada por la ausencia de la secuencia señal de secreción en la región N-terminal. Otros resultados similares han sido reportado en *S. cerevisiae* con 59 proteínas GPI, *N. crassa* con 55 (Eisenhanber *et al.*, 2004) y *Hansenula polymorpha* con 39 (Cheon *et al.*, 2014). Con base en lo anterior se puede considerar que la metodología aplicada para el escrutinio identifica la mayoría de las proteínas con anclaje GPI de *T. reesei*.

Debido a que en las células fúngicas las proteínas con anclaje GPI se presentan tanto unidas a la membrana plasmática como a la pared celular, se decidió considerar los siguientes criterios para la selección de aquellas que pudieran tener como destino final la pared celular:

La ausencia de un motivo dibásico río arriba del sitio ω promueve la ubicación final de las proteínas con anclaje GPI en la parte más externa de la pared celular, mientras que el motivo dibásico KK o KR inmediatas del sitio ω actúan como señal de anclaje de las proteínas en la membrana plasmática (Cheon *et al.*, 2014; Caro *et al.*, 1997) . En el análisis *in silico* de las proteínas predichas con anclaje GPI en *T. reesei* únicamente la 22386 presentó el motivo dibásico; resultados similares fueron reportados en *H. polymorpha* en donde se analizaron 39 presuntas proteínas con anclaje GPI y ninguna presentó dominio dibásico (Cheon *et al.*, 2014). Se observó que en las proteínas identificadas aquí en *T. reesei* con homología a Gel1, Gel2, Crf2 de *A. fumigatus* se presenta un aminoácido básico inmediatamente después del sitio ω . Estas proteínas de *A. fumigatus* no presentan el dominio dibásico pero se ha demostrado que se localizan en membrana plasmática (Cheon *et al.*, 2014; Zhao *et al.*, 2013; Schütte *et al.*, 2009); esto sugiere que las proteínas homólogas en *T. reesei* pueden anclarse en membrana plasmática.

De acuerdo a Pittet y Conzelmann (2007) el punto isoeléctrico (pI) promedio de proteínas GPI localizadas en la pared celular es 4.87 ± 0.22 y en la membrana plasmática de 6.67 ± 0.95 . Se consideró el punto isoeléctrico como un criterio adicional para seleccionar proteínas que puedan tener como destino final la pared celular. Los pI de las proteínas identificadas como posibles GPIs en *T. reesei* estuvieron en el rango de 3.59 a 8.66, similar a lo observado en 50 proteínas predichas con anclaje GPI en *P. pastoris* cuyos puntos isoeléctricos estuvieron entre 3.26 y 8.79 (Zhang *et al.*, 2013).

Algunos trabajos han establecido una relación entre la abundancia en serina/treonina río arriba del sitio de anclaje a GPI y la ubicación final de la proteína en la parte más externa de la pared (Caro *et al.*, 1997; Eisenhaber *et al.*, 2004). Frieman y Cormack (2004) observaron que al menos 28 de 41 proteínas con anclaje GPI de *S. cerevisiae* con ubicación final en la pared celular tenían un contenido cercano o mayor al 30% de Serina/Treonina; otros estudios realizados en *C. albicans* demuestran que la región rica

en Serina /Treonina en las proteínas con anclaje GPI puede ser la responsable de la unión entre los glucanos y la red de glicoproteínas mediante disulfuros y puentes NaOH (Boisramé *et al.*, 2011). Se analizó el contenido de serina/treonina en la secuencia proteica AO090011000119 reportada por Adachi y colaboradores (Adachi *et al.*, 2008) la cual es una proteína con anclaje GPI localizada en la pared celular y que ha sido utilizada como modelo en la expresión superficial de proteínas. Esta proteína dio resultado positivo con todos los análisis bioinformáticos de predicción de GPI, pero presentó un contenido menor del 30% de serina/treonina (27%). Considerando esto y que solo 28 de las 41 proteínas reportadas por Frieman y Cormack (2004) dieron contenidos arriba del 30% de S/T se seleccionaron cinco proteínas de *T. longibrachiatum* predichas con anclaje GPI que cumplen con tener 30% de serina/treonina pero también se incluyeron dos proteínas con contenidos de S/T entre 20 y 30%.

En la base de datos de *T. reesei* se identificaron varias proteínas homólogas a otras previamente caracterizadas y que son importantes para la síntesis y re-modelamiento de la pared celular, *e. g.*, la proteína con ID82633 presenta 48% de identidad con Gas1 de *S. cerevisiae*, una β -1,3-glucanosiltransferasa que actúa en la formación de las cadenas del β -1,3-glucano (Koch y Pillus, 2009). También se identificaron homólogos a Gel1p (con una identidad del 54% para la proteína 123538) y Gel2p (con una identidad del 49% para 77942), ambas participan activamente en la biosíntesis de la pared celular y la expresión de sus genes es constitutiva durante un amplio rango de condiciones de cultivo (Mouyna *et al.*, 2005). Las proteínas 49409, 69123, 122495, 55802, 67844 y 74807 de *T. reesei* presentaron homología con DFG5 de *C. albicans*, la cual es necesaria para la activación de genes específicos para el crecimiento hifal en medios alcalinos (Spreghini *et al.*, 2003). La proteína de *T. reesei* con ID 104401 presentó homología con la proteína CfmC (Afu6g06690) de *A. fumigatus*, que al ser silenciada junto con dos proteínas similares provocó un incremento en la sensibilidad de la pared hacia agentes desestabilizadores como blanco de calcoflúor y rojo congo (Vaknin *et al.*, 2014). Curiosamente, 104401 de *T. reesei* jaló las proteínas 1401645 y 1436499 de *T. longibrachiatum* pero las homólogas de 104401 en éstas no presentan homología con ninguna secuencia previamente caracterizada.

El análisis de BLASTp en diferentes bases de datos no logró identificar función para el resto de las proteínas candidatas. La mayoría son proteínas hipotéticas al igual que sus

homólogas en otras especies fúngicas; sin embargo, se observaron dos dominios característicos de proteínas extracelulares de hongos asociadas a la pared celular, el primero de ellos el dominio CFEM que corresponde a un dominio conservado en hongos. Este dominio ha sido encontrados principalmente en proteínas de pared con anclaje GPI que están implicadas en la virulencia y en la interacción hospedante-patógeno (Vaknin *et al.*, 2014). El segundo dominio presente es el WSC, un dominio putativo de unión a carbohidratos (European Bioinformatics Institute, 2018) presente en las 1,3- β -exoglucanasas de *Trichoderma* (Cohen-Kupiec *et al.* (1999) y en la proteína WSC1, sensor importante del estrés de la pared celular de *S. cerevisiae* (Levin, 2005). Por lo que la presencia de estos dominios apoya que la selección de proteínas de anclaje GPI puede efectivamente corresponder a proteínas ubicadas en la pared celular.

CAPÍTULO III

DISEÑO Y CONSTRUCCIÓN DE UN VECTOR PARA LA EXPRESIÓN SUPERFICIAL DE GFP FUSIONADA A PROTEÍNAS CON ANCLAJE GPI

3.1 INTRODUCCIÓN

La versatilidad metabólica de los hongos filamentosos hace de ellos excelentes “fábricas celulares” para la biotecnología. Varios metabolitos fúngicos importantes son ahora producidos a nivel comercial incluyendo productos farmacéuticos, antibióticos, inmunosupresores, productos a granel tales como ácidos orgánicos y enzimas explotadas a nivel industrial (Meyer *et al.*, 2011). En los últimos años se han publicado las secuencias genómicas de los hongos más importantes a nivel industrial, agrícola y médico, facilitando múltiples actividades de investigación (Meyer *et al.*, 2011; Honda y Selker 2009).

En comparación con los vectores de transformación disponibles para usar en *E. coli* o en levaduras, existen relativamente pocos vectores desarrollados para el uso en hongos filamentosos. Por ejemplo están disponibles varios vectores para la expresión en levaduras, con sitios de clonación múltiple flanqueados por diversos promotores y terminadores, pero los vectores desarrollados para hongos filamentosos han sido menos y la mayoría no son del dominio público. Esencialmente los elementos requeridos para el sistema de expresión de genes son similares a los que han sido reportados para levaduras y bacterias, en los que hay dos componentes básicos: el primero denominado “unidad de expresión” la cual comprende la transcripción, traducción y en algunos casos la señal de secreción del organismo hospedero, unido a la secuencia de ADN que codifica el producto de interés; el segundo componente hace referencia al sistema de transformación para la introducción de la unidad del gene y su mantenimiento dentro del organismo hospedero (Berka y Barnett, 1989).

Los componentes básicos de la unidad de expresión incluyen (Figura 3.1):

- a) Un promotor fuerte para la eficiente transcripción: este representa uno de los principales factores limitantes para la obtención de altos niveles de expresión en los sistemas tanto homólogos como heterólogos (He *et al.*, 2013). Existen dos

tipos, los promotores constitutivos y los inducibles que permiten el control de la expresión de los genes introducidos (Moore, 2009).

- b) Una región inicial eficiente de traducción
- c) La secuencia de ADN codificante
- d) Un terminador de la traducción

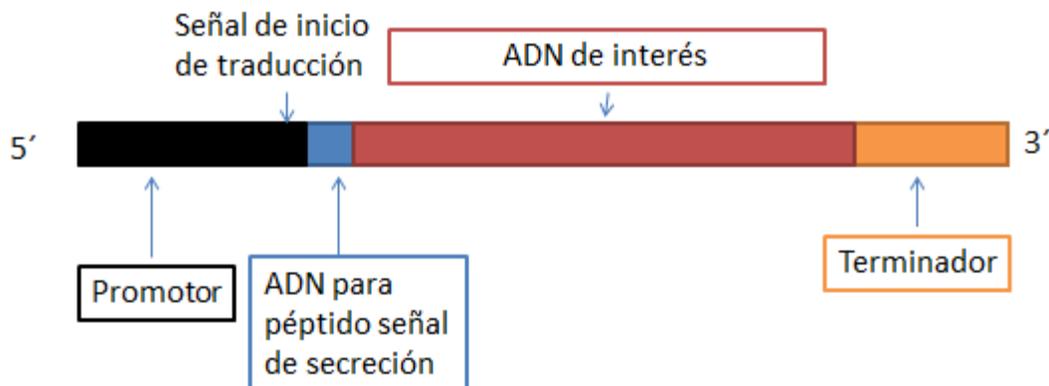


Figura 3.1 Representación esquemática de los componentes típicos de la unidad de expresión (Tomado de Berka y Barnett, 1989).

Sin embargo, la generación del constructo completo de expresión enfrenta limitantes, ya que las reacciones de ligación múltiple son comúnmente ineficientes aun después de la optimización de la reacción (An *et al.*, 2010). En este sentido se han publicado varias metodologías sobre diferentes estrategias para la optimización de la ligación simultánea de múltiples fragmentos o genes. An y colaboradores (2010) establecieron un método de ligación seguido por PCR que abarca desde el inicio del cebador fragmento hasta el final del último lo cual mejora la eficiencia de la clonación simultánea de múltiples insertos. La técnica se basa en insertar sitios específicos de restricción en cada fragmento que será ensamblado, y esos sitios de restricción deben ser compatibles con el fragmento que debe ligarse contiguamente. Los fragmentos de los extremos deben contener sitios de restricción compatibles con sitios de restricción en el vector destino. Una vez digeridos los insertos se mezclan entre sí para ligarse y la mezcla de ligación es usada como templado para llevar a cabo una reacción de PCR usando un par de cebadores que corresponden al inicio del primer inserto y al final del último. El fragmento del tamaño seleccionado es

purificado de un gel, digerido con las enzimas de restricción para hacer sus extremos “pegajosos” y clonado en los sitios respectivos del vector (Fig. 3.2).

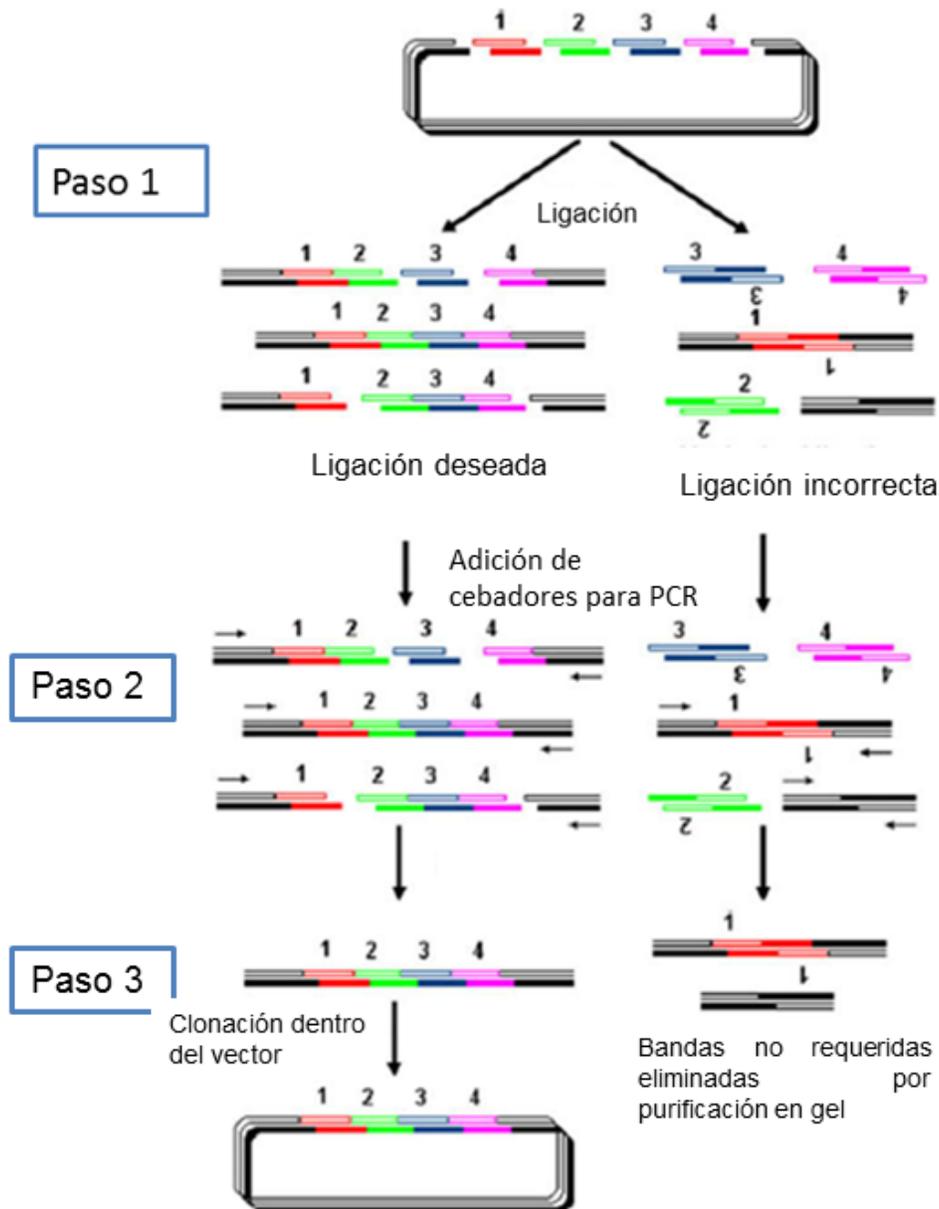


Figura 3.2 Esquema del método por PCR después de la ligación para la eficiente inserción de múltiples fragmentos de ADN (tomado de An *et al.*, 2010).

Gibson *et al.* (2011) desarrollaron una tecnología hoy licenciada a la empresa “New England Biolabs” por “Synthetic Genomics”, y comercializada como “Gibson Assembly

Cloning Kit". Este paquete permite llevar a cabo ligaciones múltiples de fragmentos de ADN en una sola reacción isotérmica, eliminando la necesidad de diseñar sitios de corte para enzimas de restricción dentro de los fragmentos de ADN para el ensamble. La técnica de Gibson® consiste en diseñar cebadores específicos que permitan tener extremos que contengan secuencias de 20 a 40 pb solapadas con el fragmento próximo en el ensamble (Fig. 3.3), posteriormente, una exonucleasa actúa en la región 3' generando extremos salientes, facilitando el alineamiento con el fragmento siguiente, con el cual traslapa. Posteriormente una ADN polimerasa rellena los sitios vacíos dentro de cada fragmento alineado y una ligasa sella los fragmentos, lográndose el ensamblado *in vitro* de la secuencia deseada.

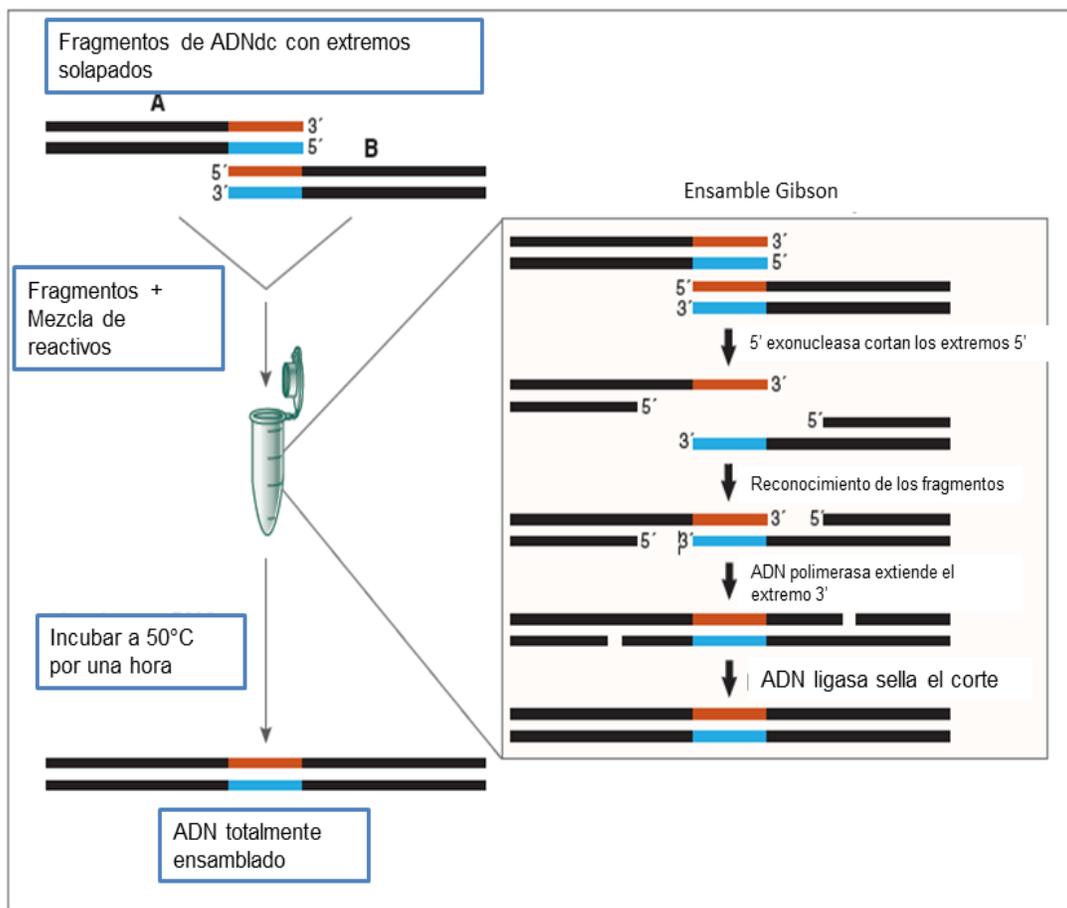


Figura 3.3 Esquematación del ensamble *in vitro* de múltiples fragmentos mediante la reacción de Gibson® (<https://www.neb.com/products/e5510-gibson-assembly-cloning-kit>).

En el presente trabajo se requirió, como esqueleto un vector de expresión. Se procedió a la construcción de éste usando los vectores pAN7.1 y gGFP como donadores de secuencias. La estrategia experimental que se siguió involucró el ensamble de múltiples fragmentos para lograr el casete de expresión superficial y ligarlo simultáneamente en el vector pAN7.1 (ensamble de 6 fragmentos). En total en este trabajo se planeó clonar y expresar 8 proteínas: la GFP, la AMPO1 que es una GPI reportada de *A. oryzae*, 5 proteínas GPIs de *T. longibrachiatum*. Identificadas en el presente trabajo (capítulo II), y una triacilglicerol lipasa descrita en el capítulo 4 (ID 52630). La estrategia es lograr un prototipo que contenga la GFP y la GPI (AOMP1) para posteriormente usarlo de plataforma para lograr los siguientes constructos, reemplazando únicamente la región codificante (GFP por lipasa y la AMPO1 por las proteínas con anclaje GPIs identificadas en el presente trabajo y descritas en el capítulo 2).

3.2 MATERIALES Y MÉTODOS

3.2.1 Diseño de la construcción del vector pAN7.1-GFP-GPI

Para la construcción del vector de expresión se usó como plataforma el vector pAN7.1 (Fig. 3.4).

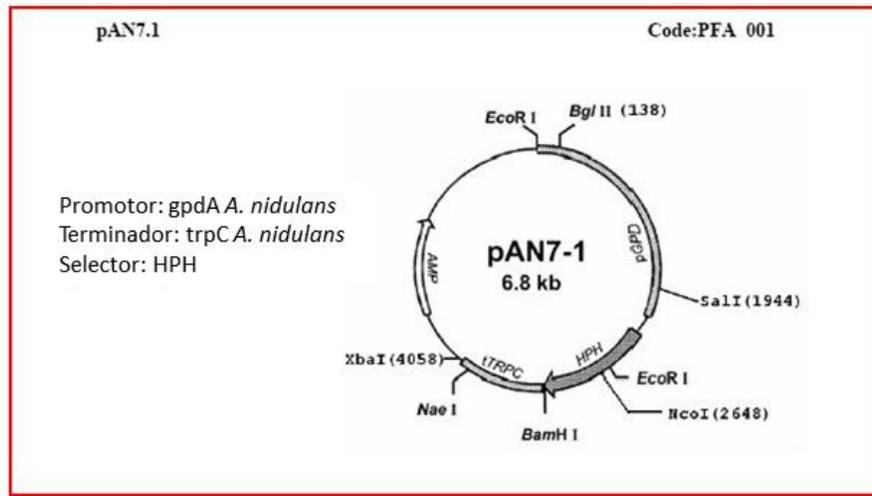


Figura 3.4 Representación circular del vector pAN7.1 con sus principales componentes: promotor *GPD*, terminador *TRPC*, agente de selección higromicina B fosfotransferasa (Hph) y ampicilina (Amp) (Punt *et al.*, 1987).

Con el programa Ugene (<http://ugene.unipro.ru/>) se realizó el perfil de digestión del vector esqueleto y de las secuencias a clonar con todas las enzimas de restricción posibles y se seleccionó el sitio de corte único en *XbaI* para la introducción de un “inserto”, el cual está constituido por los siguientes fragmentos (casete de expresión):

Promotor: Se amplificó el promotor *GPD* del plásmido pAN7.1 (Punt *et al.*, 1987)

Secuencia señal de secreción: fue tomada del trabajo reportado por Adachi y colaboradores (2008), la cual dirige las proteínas hacia el retículo endoplasmático y posteriormente hacia la parte externa de la pared celular (N.28).

Proteína verde Fluorescente: se amplificó del plásmido gGFP (Pöggeler *et al.*, 2003).

Proteína de anclaje GPI: el constructo se realizó inicialmente con la proteína identificada por Adachi y colaboradores (2008) con la clave AO090011000119 (AOMP1). La secuencia del gen fue descargada de la base de datos de *A. oryzae* (http://www.aspgd.org/cache/A_oryzae_RIB40_genomeSnapshot.html); se editó eliminando los 20 primeros tripletes que corresponden a la secuencia señal de secreción propia de esta proteína. La síntesis de la secuencia nucleotídica correspondiente se solicitó a Integrated DNA Technologies (USA)

Terminador: Se utilizó el terminador *TRPC* de *A. nidulans* del vector pAN7.1 (Punt *et al.*, 1987)

3.2.2 Amplificación de secuencias para la introducción de los sitios de restricción

Las cepas de *E. coli* DH5 α transformadas con los vectores pAN7.1 y peGFP se inocularon en medio LB líquido con 100 μ g/mL ampicilina a 37 °C y 180 rpm; se incubaron toda la noche y posteriormente se realizó la extracción del plásmido mediante el paquete QIAprep® Spin Miniprep Kit. Los plásmidos extraídos fueron usados como plantillas para el aislamiento del Promotor, Terminador (pAN7.1) y la secuencia codificadora para eGFP (GFP). La secuencia señal de secreción y la proteína con anclaje GPI AOMP1 reportada por Adachi y colaboradores (2008), fueron sintetizadas por la compañía “*Integrated DNA Technologies*”. Se generaron cebadores para cada secuencia introduciendo en los extremos los sitios de reconocimiento para las enzimas de restricción.

3.2.3 Amplificación de las secuencias mediante los cebadores con los adaptadores para el sistema Gibson®

Para la amplificación de los productos de interés con los cebadores que contienen los adaptadores para el sistema de Gibson®, los amplicones con los sitios de restricción se visualizaron en un gel de agarosa al 0.8% y se cortaron las bandas para su purificación mediante el paquete comercial QIAprep® Gel extraction kit; el producto de purificación se utilizó como plantilla para la amplificación.

En el caso de los amplicones correspondientes al promotor *GPD* y al ADNc que codifica la proteína AOMP1 (con anclaje GPI), se clonaron en el sistema PCR2.1-TOPO-TA para obtener mayor cantidad de ADN plantilla para su amplificación usando los cebadores diseñados para su ensamble con el sistema Gibson®.

3.2.4 Ligación múltiple de fragmentos

Para realizar la ligación múltiple en un solo paso se utilizó el paquete de Gibson® (New England, Biolabs). Los cebadores se diseñaron con el programa NEBuilder (<http://nebuilder.neb.com/>). Los amplicones se cuantificaron por Nanodrop (Thermo Scientific, U.S.A., Wilmington) y se ajustaron a una relación molar 1:1 entre sí, siguiendo las indicaciones del fabricante.

3.2.5 Transformación y confirmación de las transformantes

El constructo ensamblado se tomó de la reacción de Gibson sin purificar y se utilizó para la transformación de células quimio-competentes DH5 α , siguiendo las recomendaciones del proveedor. Las células se cultivaron en medio LB agar suplementado con 100 μ g/mL de ampicilina; las colonias recuperadas se resembraron en medio fresco y se realizó PCR en colonia utilizando los cebadores para la secuencia GFP y los cebadores derecho de la secuencia Ss-gib y reverso de la secuencia GPI-gib. Ss es, por señal de secreción y GPI es la región correspondiente a la proteína GPI a evaluar

Las clonas que dieron positivo para ambas amplificaciones se crio-conservaron en medio LB con glicerol al 50%.

3.2.7 Aislamiento de los plásmidos eGFP, pAN7.1 y pAN-GFP-GPI

Las cepas de *E. coli* que contienen (de manera independiente) los vectores eGFP, pAn7.1 y pAN-GFP-GPI fueron inoculadas en medio LB con 100 μ g/mL ampicilina e incubadas toda la noche a 37 °C a 130 rpm; los plásmidos se extrajeron mediante el protocolo QIAprep Spin Miniprep kit (QIAGEN)

3.2.6 Generación de protoplastos para transformación

Se trabajó con la cepa de *T. harzianum* y *T. longibrachiatum*; éstas se cultivaron en medio PDA y se incubaron tres días a temperatura ambiente. Para cada caso se cortó un fragmento de micelio y se maceró en 50 mL de medio PDB; el micelio se incubó durante tres días, se filtró y se maceró de nuevo. El micelio se inoculó en 50 mL de medio PDB fresco y se incubó otro día a temperatura ambiente a 120 rpm. La biomasa se centrifugó a 4000 rpm por 12 min, se eliminó el sobrenadante y se lavó dos veces con un medio osmótico que contenía 10mM de fosfato de sodio y 1.2 M de sulfato de magnesio a pH

5.8. Se le adicionó una solución lítica compuesta de 150 mg de “*Driselase*” y “*Lysing enzyme*” (Sigma-Aldrich), previamente esterilizadas por filtración y disueltas en 15 mL de medio osmótico. Se incubó toda la noche en oscuridad a 70 rpm y 30 °C. Al final la muestra se filtró con ayuda de dos capas de gasa y una de Miracloth® previamente esterilizadas; el filtrado se dividió en tubos de 1.5 mL y se centrifugaron a 6500 rpm a 4 °C por 12 min. Posteriormente se lavaron dos veces con medio MSC que contiene 10mM de MOPS, 1 M de sorbitol y 50 mM de cloruro de calcio pH 6.5, incubando a 6500 rpm, 4 °C por 12 min. Se realizó una dilución y se contó con cámara de Neubauer, se ajustó la concentración a 1×10^7 protoplastos en 100 μ L.

3.2.7 Transformación por electroporación

Los protoplastos obtenidos se ajustaron a una concentración de 1×10^6 y se lavaron tres veces con una solución a 1.2 M de Sorbitol, la electroporación se realizó de acuerdo al protocolo reportado por Goldman y colaboradores (1990). Los protoplastos electroporados se recuperaron con 200 μ L de PDB y se regeneraron durante dos horas a 29 °C con agitación de 90 rpm. Posteriormente se plaquearon en cajas Petri con PDA 12.5 μ g/mL de Higromicina. Las cajas se incubaron a 30 °C en oscuridad y se observaron cada día durante una semana, para ver el crecimiento de las colonias.

3.2.8 Transformación por el Método de PEG y cloruro de calcio

Una vez obtenidos los protoplastos se continuó con el método de transformación descrito por Kantún-Moreno (2012); se utilizaron diferentes concentraciones de plásmidos para los diferentes transformantes, 6 μ g para GFP, 9 μ g pAn7.1 y 10 μ g pAn7.1-GFP-GPI. Se transformaron las cepas de *T. harzianum* y *T. longibrachiatum*.

3.2.9 Transformación por el protocolo de perlas de cristal

Se buscaron protocolos alternativos de transformación que evitaran la formación de protoplastos, se modificó el propuesto por Pareek y colaboradores (2015) que se basa en el uso de acetato de litio, perlas de cristal y esporas recién germinadas. Se trabajó con la cepa de *T. harzianum* debido que tiene mayor capacidad de generar esporas en comparación con *T. longibrachiatum*. La cepa se cultivó en PDA por 5 días a temperatura ambiente; las esporas se recuperaron con agua destilada de cada caja y se vortexearon

para favorecer su desagregación, posteriormente se centrifugaron a 2000 × g por 10 min a 4 °C y se lavaron dos veces con agua destilada estéril.

Las esporas se cuantificaron en cámara de Neubauer y se inocularon a una concentración de 1×10^6 en PDB y se incubaron por 16 h a 100 rpm. A 400 µL de esporas germinadas se le adicionaron 300 mg de perlas de vidrio previamente estériles y 1 µg de ADN plasmídico linearizado. Posterior se le adicionó 40 µL de acetato de litio 0.1 M y 400 µL de PEG 60% (p/V). Se agitó vigorosamente en vórtex durante 30 s., se dejó sedimentar las perlas, se colectó el líquido en suspensión y se vació en PDA, luego de tres días se le adicionó una capa con el selector.

3.2.10 Confirmación de las transformantes

Desafortunadamente no se logró establecer un protocolo reproducible para la cepa de *T. longibrachiatum* por lo cual se decidió continuar trabajando con la cepa de *T. harzianum*. Las cepas transformadas se sometieron al proceso de selección con 100 µg/mL de higromicina B el medio sólido. Las clonas positivas se cultivaron posteriormente en medio líquido con agente de selección. Se extrajo el ADN utilizando el protocolo reportado por Conde-Ferrández y colaboradores (2008); la pureza de las muestras se cuantificó mediante Nanodrop (Thermo Scientific, U.S.A., Wilmington) en las longitudes de onda 260/280 y 230/260. El ADN obtenido se utilizó como templado para amplificar con los cebadores Hyg (Anexo 2.3) el gen de resistencia para higromicina.

3.2.11 Visualización de las muestras en el microscopio confocal

Las cepas transformadas se cultivaron sobre PDA y se incubaron a temperatura ambiente por dos a tres días dependiendo del desarrollo del hongo. Se cortó un fragmento de aproximadamente 1 cm² en la parte más joven del micelio y se usó el método del cultivo del bloque de agar invertido reportado por Freitag y colaboradores (Freitag *et al.*, 2004).

3.2.12 Segunda estrategia para la construcción del Vector usando como plataforma el casete *HPH* del vector pSilent-1

Debido a que los resultados obtenidos con el constructo pAN7.1-GFP-GPI, usando el vector pAN7.1 como destino no eran del tamaño esperado, se decidió cambiar el vector plataforma de construcción: la secuencia del vector pSilent-1 se analizó con el programa Ugene y se identificaron dos sitios de corte *Xba*I que digieren el vector en dos fragmentos, un fragmento que corresponde a la unidad de expresión del gen de higromicina y los sitios

respectivos de replicación para la bacteria y el hongo , y el segundo fragmento contiene la unidad de silenciamiento, de menor tamaño que el primero.

Se extrajo el plásmido mediante el protocolo de QIAprep Spin Miniprep kit, se cuantificó por Nanodrop (Thermo Scientific, U.S.A., Wilmington) y se digirió 1 µg con la enzima *Xba*I durante toda la noche a 37 °C; la digestión se detuvo a 65 °C por 15 min y se corrió en un gel de agarosa al 0.8%. La banda correspondiente a la unidad de expresión del gen *Hph* se cortó y se purificó usando el kit comercial QIAprep® Gel extraction kit. El producto de purificación se usó para la ligación con el sistema Gibson® con los fragmentos correspondientes al promotor *GPD*, la secuencia señal de secreción (Ss), el fragmento que codifica a la GFP y el fragmento que codifica para la proteína con anclaje GPI (AOMP1) y el terminador *TRPC*, como se describió en el punto 3.4.

3.2.13 Confirmación de transformantes y perfil de restricción del vector pHyg-GFP-GPI

La transformación y confirmación de las clonas se realizó como se describió anteriormente en el apartado 3.5, exceptuando que para la confirmación se amplificó únicamente el gen que codifica para la proteína GFP. Se seleccionó una de las clonas positivas y se cultivó en medio LB 100 µg/mL con ampicilina. El plásmido se extrajo mediante el paquete QIAprep Spin Miniprep, y 1 µg de plásmido se digirió siguiendo las indicaciones del proveedor con las enzimas *Xba*I, *Spe*I, *Apa*I (por separado) y se realizó una doble digestión con las enzimas *Xba*I-*Spe*I y *Kpn*I-*Apa*I, siguiendo las indicaciones del programa

programa	Double	Digest	Calculator
----------	--------	--------	------------

 (<https://www.thermofisher.com/mx/es/home/brands/thermo-scientific/molecular-biology/thermo-scientific-restriction-modifying-enzymes/restriction-enzymes-thermo-scientific/double-digest-calculator-thermo-scientific.html>). Los fragmentos digeridos se visualizan en un gel de agarosa al 1%.

3.3.14 Construcción del vector de expresión utilizando enzimas de restricción

Los plásmidos pHyg-GFP-GPI, pAN7-1-GFP-GPI y pCR@2.1TOPO-*Hph*, fueron digeridos simultáneamente con las enzimas *KpnI* y *Apal* siguiendo las indicaciones del programa “*Double Digest Calculator*”; las digestiones se detuvieron a 65°C por 15 min y la digestión se visualizó en un gel de agarosa al 0.8%. Los fragmentos liberados correspondientes a las regiones codificantes de la proteína Ss-GFP-GPI de los plásmidos pHyg-GFP-GPI (ORF2) y pAN7-1-GFP-GPI (ORF1) y el plásmido digerido pCR@2.1TOPO (sin inserto), se purificaron mediante el paquete comercial QIAprep® Gel extraction kit. Los fragmentos purificados se cuantificaron mediante Nanodrop (Thermo Scientific, U.S.A., Wilmington), y se ligaron utilizando T4 ADN ligasa (5 U/μL) de ThermoFisher, siguiendo las indicaciones del proveedor. De la reacción de ligación se tomaron 4 μL y se transformaron células químico competentes TOPO10. Se seleccionaron dos clonas que fueron positivas y sus plásmidos se extrajeron mediante el protocolo QIAprep Spin Miniprep kit. Los plásmidos pCR@2.1TOPO-ORF1 y pSilent1 se digirieron con las enzimas *HindIII*, *Apal* y una doble digestión con ambas enzimas.

Los fragmentos generados con *HindIII*-*Apal* del ORF1 y del plásmido pSilent1, se cuantificaron y se ligaron usando T4 ADN ligasa (5 U/μL) de ThermoFisher; como control negativo se religó el plásmido pSilent1 sin inserto. Se tomaron 4 μL de la reacción y se transformaron células DH5α químico competentes. Se confirmaron las clonas mediante la amplificación del fragmento que codifica para la GFP con los cebadores descritos en el Anexo 2. 1; las clonas positivas se sometieron a doble digestión con las enzimas *KpnI* y *Apal* con el fin de corroborar la presencia de los respectivos sitios.

3.3 RESULTADOS

3.3.1 Análisis *in silico* del constructo del vector de expresión pAN7.1-GFP-GPI

La secuencia final ensamblada *in silico* del primer constructo se analizó con el programa Ugene para checar el perfil de restricción y la existencia de un solo marco de lectura abierto, considerando el inicio del constructo de expresión (señal de secreción). La figura 3.5a muestra el constructo diseñado: inicia con la señal de secreción, seguido por la proteína fluorescente y al final la proteína con anclaje GPI.

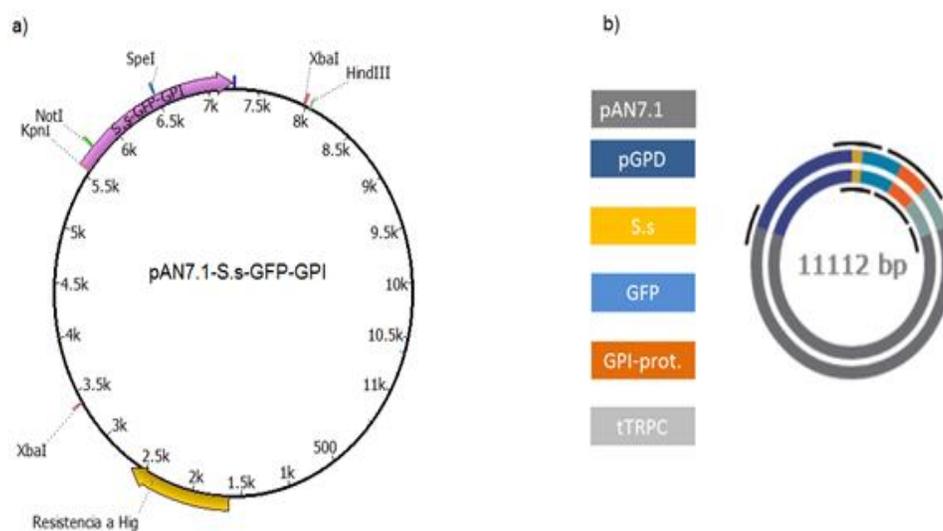


Figura 3.5 Análisis *in silico* del vector pAN7.1-GFP-GPI: a) análisis de marcos de lectura abierto, en donde se observa el marco de lectura esperado en la región comprendida entre los 5.5 Kb y 7Kb que corresponde al constructo de la proteína quimérica Ss-GFP-GPI; b) Esquema de la ubicación de los insertos dentro del vector pAN7.1 modificado.

Con ayuda del programa NEBcutter V2.0 (<http://tools.neb.com/NEBcutter2>) se realizó un perfil de digestión *in silico* del vector. Se observó un marco de lectura abierto que codifica para una proteína de 550 aminoácidos perteneciente a la fusión de la Secuencia señal de secreción-GFP-GPI. Río arriba en el 5´ de la región N-terminal se encuentra el sitio de corte *KpnI* y río abajo de la región C-terminal en el extremo 3´ el sitio de corte *Apa1* insertados mediante adaptadores en los cebadores para la reacción en cadena de la polimerasa (Fig. 3.6)

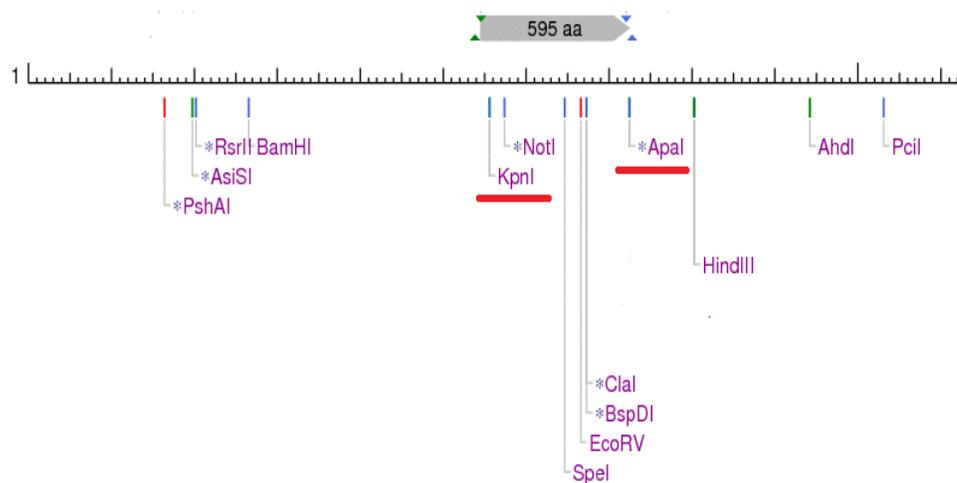


Figura 3.6 Análisis *in silico* de los sitios de restricción para el del vector pAN7.1-GFP-GPI, subrayado en rojo los sitios que flanquean el marco de lectura abierto (ORF) que codifica para la proteína quimérica Ss-GFP-GPI, en 5' el sitio *KpnI* y en 3' el sitio *Apal*.

3.3.2 Aislamiento y amplificación de las secuencias del constructo

Con los cebadores diseñados para cada secuencia se realizaron PCR de gradientes; una vez obtenidas las temperaturas óptimas se amplificaron las secuencias teniendo bandas únicas del tamaño esperado para cada caso (Fig. 3.7).

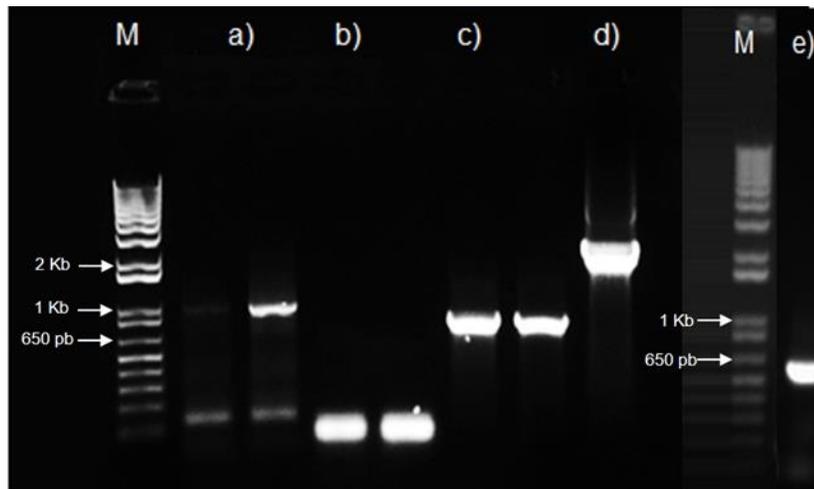


Figura 3.7 Visualización en gel de agarosa al 1% de los amplicones que conforman la construcción de expresión. a): amplicón correspondiente a la secuencia señal de secreción

(Ss) y la secuencia que codifica para la proteína con anclaje GPI reportada por Adachi y colaboradores (2008); b) control negativo; c): amplicón correspondiente al terminador *TRPC* aislado del plásmido pAN7.1 (769 pb); d): amplicón correspondiente al promotor del gen *GPD* de *Aspergillus nidulans* aislado del vector pAN7.1 (2.1 Kb); e): amplicón correspondiente a la región codificante de la proteína verde fluorescente reportada por Maor y colaboradores (1998) del vector peGFP (723 pb); M: marcador de peso molecular 1 Kb plus DNA Ladder (ThermoFisher Scientific).

3.3.3 Amplificación con los adaptadores diseñados para la metodología de ligación mediante el sistema Gibson®

En el caso de los cebadores diseñados para el sistema Gibson®, se obtuvo una amplificación correcta y bandas únicas para los amplicones Ter-Gib (Terminador con los adaptadores para el ensamble Gibson®) y Ss-Gib (secuencia señal de secreción amplificada con los adaptadores para el ensamble Gibson®). En el caso del amplicón eGFP-Gib las bandas eran menos intensas y se observaba un ligero barrido (Fig. 3.8c). Sin embargo, en todos los casos al purificar las bandas de ADN se alcanzó la concentración de ADN requerido para su uso en la ligación múltiple mediante el paquete Gibson®.

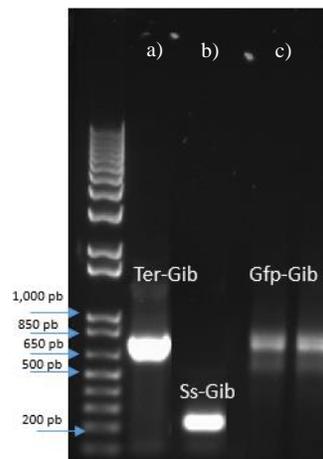


Figura 3.8 Amplificación de las secuencias con los cebadores que introducen las secuencias adaptadoras para la ligación con el sistema Gibson®. A) Amplicón del Terminador *TRPC* con los adaptadores Gibson, b) Secuencia señal de secreción con los adaptadores Gibson c) Secuencia codificante para proteína verde fluorescente con los adaptadores Gibson

Por otra parte la amplificación de la secuencia que codifica la AOMP1, empleando los cebadores que tienen los adaptadores para la ligación con el paquete de Gibson® fue poco eficiente cuando se usó como templado el producto comercial recibido, por lo que fue necesario realizar dos rondas de PCR con los cebadores con los adaptadores Gibson, es decir, el producto de PCR de la primera reacción se reamplificó nuevamente (Fig. 3.9a). Se probó trabajar con PCR *touchdown* con el fin de disminuir la inespecificidad de los cebadores. Sin embargo, se observó un barrido y una banda del tamaño esperado poco nítida (Fig. 3.9b), por lo que se procedió a ligar el producto de la primera amplificación en el vector TOPO-TA y clonarlo en *E. coli* One Shot® TOP10. El plásmido se purificó y se usó como templado para la amplificación con los cebadores GPI-gib, mejorando la amplificación y alcanzando la concentración necesaria para la ligación múltiple (Fig. 3.9).

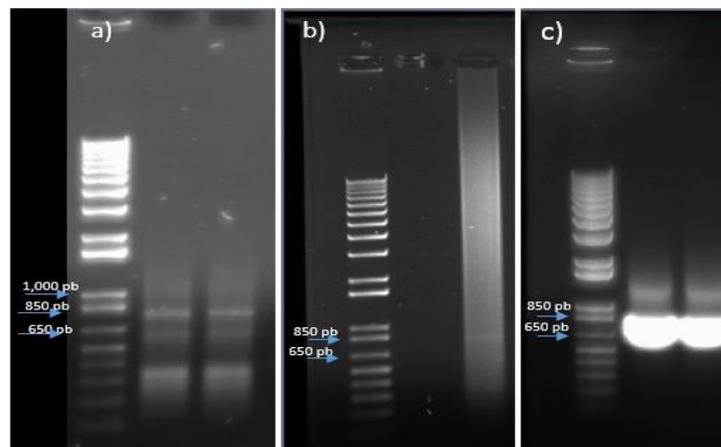


Figura 3.9 Amplificación de la secuencia AOMP1 con los cebadores que introducen los adaptadores para el sistema Gibson®. a) Amplificación usando como templado producto de purificación y el par de cebadores, b) Producto de amplificación usando PCR *touchdown*, c) amplificación usando como templado el fragmento clonado en TOPO-TA.

La amplificación del promotor con los cebadores que introducen los adaptadores para el sistema Gibson® fue también inespecífica cuando se reamplificó sobre el producto de la primera amplificación. Se observaron dos productos de PCR; la banda esperada de más de 2 Kb y una banda de ~200 pb que amplificó a pesar de probar con distintas temperaturas (Fig. 3.10 b, c, d). La mejor amplificación se consiguió a 47°C, disminuyendo la concentración de cebadores (Fig. 3.10).

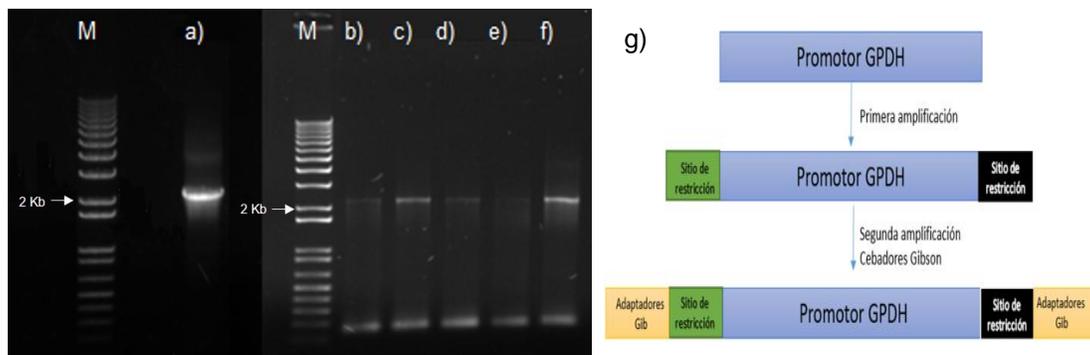


Figura 3.10 Amplificación de la secuencia del promotor *GPD* a) amplificación con el primer par de cebadores que introducen los sitios de restricción en la secuencia. De b) hasta f) corresponden a la amplificación con los cebadores que introducen los adaptadores para la ligación con Gibson®. Se utilizó 10 μ M de cada cebador para cada reacción en una PCR de 35 ciclos. Usando las siguientes temperatura de alineamiento (tm) b) 46°C, c) tm 47°C, d) tm 48°C, e) tm 49°C, f) se usó 2.5 μ M de cada cebador, tm 47°C y se dieron 40 ciclos de PCR. M: Marcador de peso molecular 1 Kb plus Ladder (ThermoFisher Scientific). La figura g) corresponde a la esquemización de la introducción de los sitios de restricción y las secuencias adaptadoras para el sistema Gibson®, ambos introducidos mediante cebadores de PCR.

3.3.4 Ligación múltiple con el sistema Gibson®

En el caso de ligación múltiple (4 a 6 fragmentos) con el sistema Gibson® se requieren 0.2-1 pmoles de ADN de cada amplicón, por lo cual se calcularon las concentraciones necesarias mediante la fórmula:

$$pmoles = \frac{peso\ en\ ng * 1000}{pares\ de\ bases * 650\ Da}$$

Se linearizó el vector pAN7.1, el cual se usó como esqueleto para conferir la resistencia a higromicina y ampicilina; se utilizó *Xba*I que tiene un corte único al final del terminador (*TRPC*); la digestión se llevó a cabo durante 12 h y al final la digestión se analizó en un gel de agarosa al 1% (Fig. 3.11b2).

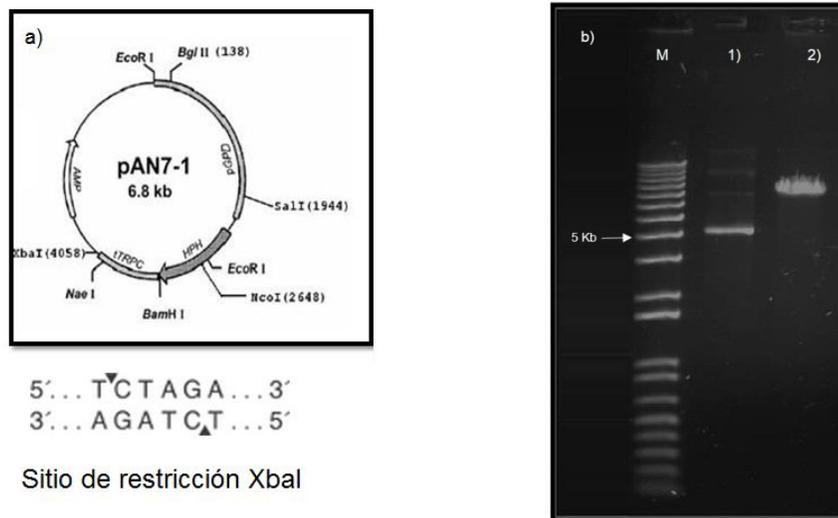


Figura 3.11 Representación esquemática del vector pAN7.1 y visualización de la digestión con *Xba*I a) Esquema con los principales sitios de restricción del vector pAN7-1. b) M: Marcador de peso molecular 1 Kb plus Ladder (ThermoFisher Scientific) ; Carril 1) pAN7.1 sin digerir; carril 2) pAN7.1 digerido con *Xba*I.

En la figura 3.11 b) se observa en el segundo carril las tres conformaciones que presenta un plásmido, la superior corresponde a una forma circular relajada, la segunda corresponde a la forma linearizada y la inferior a la superhelical o superenrollada. En el carril tres se observa la digestión total del vector presentando únicamente la forma linearizada.

Una vez obtenido el plásmido linearizado se procedió a realizar la ligación múltiple siguiendo las especificaciones del paquete Gibson®; posteriormente 5 μ L del producto de ligación se corrieron en un gel de agarosa (Fig. 3.12). Se observaron tres bandas: una banda tenue del tamaño esperado (~11 Kb) que corresponde al constructo final; una banda fuerte correspondiente al vector pAN7.1 linearizado y una banda en la parte inferior correspondiente a los insertos que codifican para GFP o GPI, ya que presentan tamaños similares.

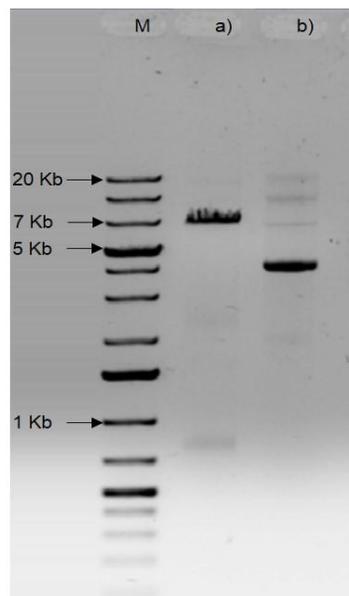


Figura 3.12 Visualización en gel de agarosa de la ligación múltiple con el sistema Gibson®: a) ligación múltiple Gibson® de pAN7.1, GFP y GPI y b) Vector pAn7.1 sin digerir en sus tres conformaciones, M marcador de peso molecular GeneRuler™ 1 Kb plus DNA Ladder (ThermoFisher Scientific).

3.3.5 Transformación y Confirmación de los transformantes

Debido a que en el gel de agarosa se observó una baja cantidad del constructo generado (pAN7.1-GFP-GPI) se decidió utilizar 5 μ L de la reacción (sin purificar) y llevar a cabo la transformación. Se realizaron dos controles: un control negativo sin transformación y un control positivo del proceso de transformación con el vector pUC19. Se obtuvieron un total de 72 transformantes en dos placas y se seleccionaron al azar 44 para sembrarlas y analizar mediante PCR en colonia con los cebadores diseñados para el gen que codifica para la proteína verde fluorescente. Los resultados se presentan en la figura 3.13. No se

pudo calcular la eficiencia de transformación en las células DH5 α debido a que no se cuantificó la cantidad de constructo utilizando en la transformación.

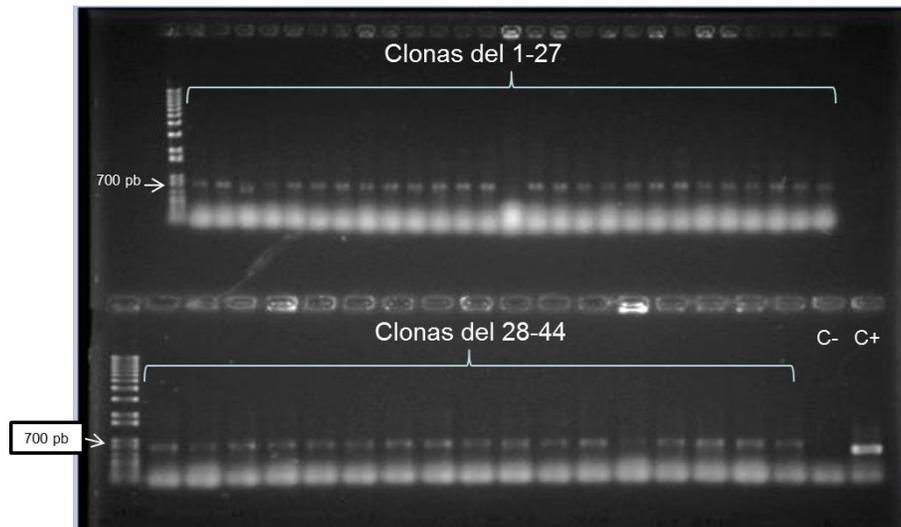


Figura 3.13 Amplificación por PCR en colonia del fragmento codificante para la proteína GFP para verificar transformación con constructo pAN7.1-GFP-GPI en las clonas transformadas; C-: control negativo, C+: amplicón (700 pb) obtenido usando como templado peGFP como control positivo para corroborar la PCR.

En la mayoría de las 44 clonas seleccionadas se observó el amplicón del tamaño esperado (700 pb aprox.). Se seleccionaron las primeras 11 clonas y se realizó una segunda PCR en colonia utilizando una mezcla de dos parejas de cebadores (para Ss-Gib y GPI-gib) para confirmar la presencia de los fragmentos correspondientes a la señal de secreción (Ss) y al anclaje GPI (GPI), los cuales flanquean la región codificante, en este caso, la GFP. El tamaño esperado para la clona positiva es de 1705 pb. Los resultados se presentan en la figura 3.14.

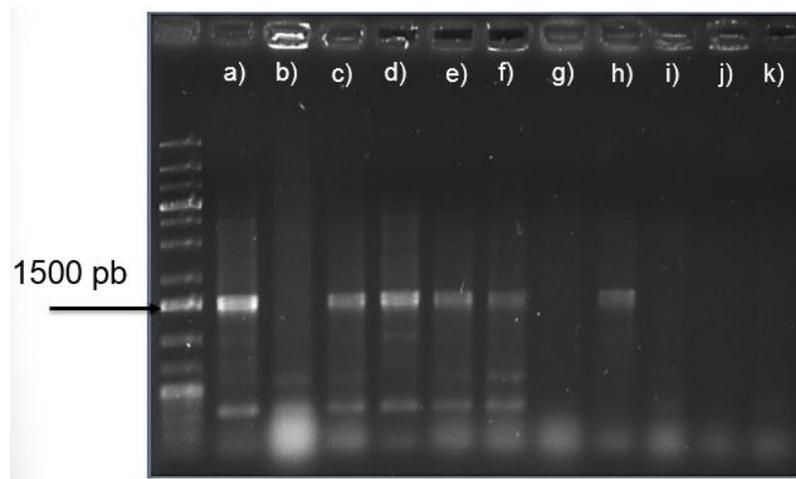


Figura 3.14 Amplificación por PCR en colonia del fragmento de la región codificante que abarca la secuencia señal de Secreción-GFP-GPI: de a-k): clonas con el posible vector pAN7.1-GFP-GPI. La flecha señala el producto esperado.

Como se observa en la figura 3.14, el amplicón esperado se obtuvo en las clonas a), c), d), e), f) y h). Aunque se aprecian al menos tres bandas de diferente tamaño en algunas clonas, las cuales podrían ser productos inespecíficos o dímeros generados por los cebadores, para la PCR se utilizó la temperatura de alineamiento del cebador que tiene la T_m más baja, ya que es la temperatura en la que ambos cebadores coinciden.

3.3.6 Confirmación del constructo pAN7.1-GFP-GPI por patrones de digestión y PCR

Para confirmar los resultados, el constructo fue analizado con enzimas de restricción. Se utilizaron tres enzimas de corte único en el constructo (*Apal*, *NotI* y *KpnI*) estos sitios se introdujeron por medio de la PCR con los adaptadores de los cebadores antes del ensamble con el sistema Gibson®. En los casos de las enzimas *Apal* y *KpnI* se observó una linearización total (Fig. 3.15), mientras que con la enzima *NotI* la digestión fue parcial observándose un incremento en la banda intermedia que corresponde al plásmido linearizado y una disminución en la banda inferior, que corresponde al plásmido superenrollado. La linearización del vector indica que el constructo no tienen inserciones o deleciones en los sitios de corte y por lo tanto los sitios son funcionales para ser utilizados en la excisión de la secuencia GPI usada como control (AMPO1) y reemplazarla más

adelante con las secuencias de GPIs identificadas en el análisis bioinformático en el capítulo II.

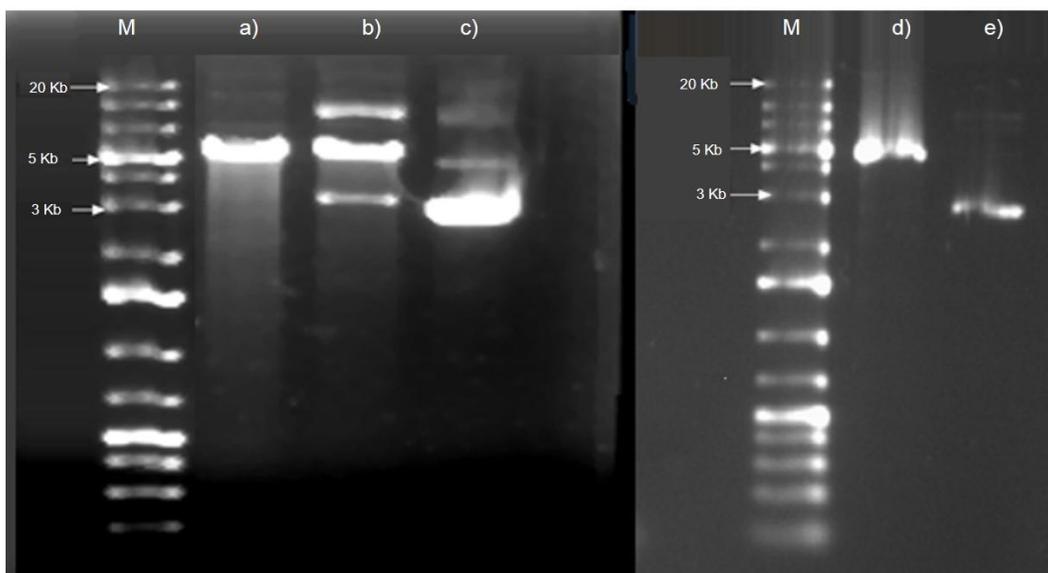


Figura 3.15 Digestión del plásmido pAN7.1-GFP-GPI con tres diferentes enzimas de restricción: a) Digestión con *Apal*; b) Digestión con *NotI*; c) plásmido sin digerir; d) digestión con *KpnI*; e) plásmido sin digerir y M: Marcador de peso molecular Gene ruler 1 Kb plus DNA Ladder (ThermoFisher Scientific).

Sin embargo se observó que el tamaño del plásmido linearizado no corresponde con lo estimado de ~11 Kb; en cambio se observa una banda única cercana a 5 Kb.

Debido a lo observado anteriormente, se realizó una doble digestión. Para ello se seleccionaron enzimas de restricción que flanquearan los módulos que componen el inserto Ss-GFP-GPI. Una de las enzimas fue *NotI*, esta enzima se encuentra entre la secuencia de Ss y GFP, por lo cual se utilizó para realizar la primera digestión. Posteriormente se tomó una alícuota de esta digestión y se colocó en un segundo amortiguador de reacción para la digestión con la enzima *KpnI*

No se observó liberación del fragmento que codifica la señal de secreción (Fig. 3.16a), a pesar que en la digestión sólo con *NotI* o con *KpnI* se observó la presencia de los sitios respectivos de corte; en el caso de la región codificante para la GFP si se observó el producto de digestión esperado de aproximadamente 750 pb (Fig. 3.16b), aunque la banda era tenue y se observó bandas superiores que correspondían al plásmido en sus tres conformaciones y una quinta banda cercana a 5 Kb que podría corresponder al

plásmido doblemente digerido. La secuencia que codifica para la proteína con anclaje GPI también se observó liberada del plásmido (Fig. 3.16c) en la región de ~800 pb y se observó una banda correspondiente al plásmido doblemente digerido, similar a una de las bandas observadas en la Fig. 3.16b.

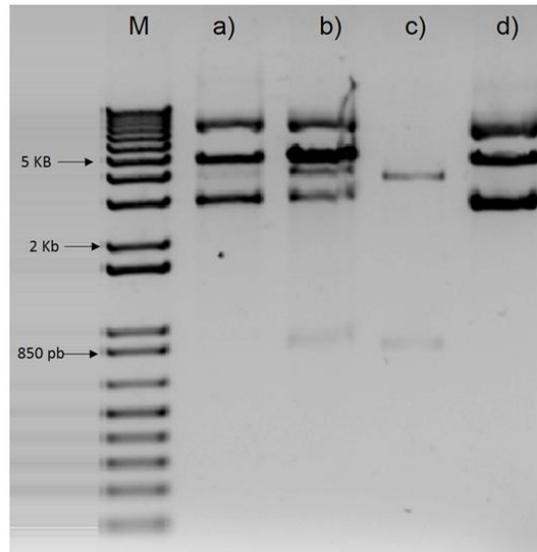


Figura 3.16 Doble digestión del constructo pAN7.1-GFP-GPI; a) Digestión con las enzimas *NotI* y *KpnI*, b) Digestión con *NotI* y *SpeI*, c) Digestión con *SpeI* y *ApaI*, d) plásmido sin digerir y M: marcador de peso molecular 1 Kb plus Ladder (ThermoFisher Scientific).

Se diseñaron cebadores específicos para amplificar fragmentos internos de la región codificante del inserto, con el fin de corroborar que los módulos no se hayan ligado de forma errónea e impidan la traducción correcta de la proteína quimérica diseñada (Ss-GFP-GPI). Los cebadores amplifican regiones entre dos fragmento continuos, e.g. los cebadores P_{gfp} F y R amplifican una secuencia que inicia en la región final del promotor hasta la región inicial del fragmento que codifica para la proteína GFP; el traslape de los cebadores entre un fragmento y el contiguo permite, confirmar el orden de los insertos (Fig. 3.17a).

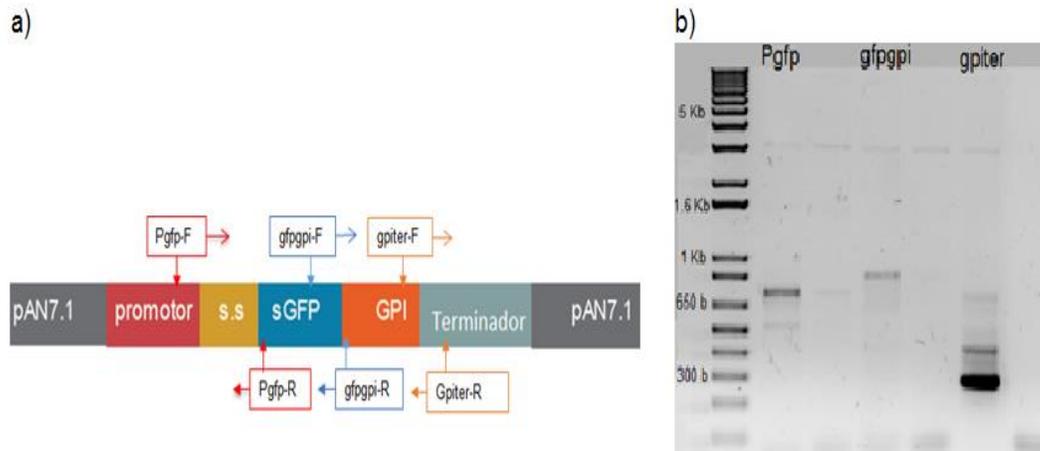


Figura 3.17 1 Diseño y amplificación de cebadores en sitios internos de la región codificante quimérica Ss-GFP-GPI para la confirmación del orden de los fragmentos de ADN y su secuenciación; a) Esquemización de los sitios en donde se diseñaron los, en rojo la ubicación de los cebadores para la amplificación del fragmento Pgfp, en azul para el fragmento gfgppi y en naranja gpiter; b) Amplicones obtenidos y visualizados en gel de agarosa al 1%.

En el caso de los cebadores para la GFP en 5' y un fragmento de la secuencia de GPI en 3' amplifican una única banda del tamaño esperado de 840 pb. Pero para las dos parejas de cebadores que están flanqueadas por una región ya sea del promotor o el terminador se obtienen múltiples bandas. En el caso de los cebadores diseñados entre el promotor y la región que codifica la GFP se observó el amplicón de interés (750 Kb) y una segunda banda cercana a 500 pb; mientras que con los cebadores para la región que va de la secuencia GPI al terminador se observan múltiples bandas entre 300 y 650 pb; el tamaño esperado del amplicón es 550 pb.

La amplificación de las bandas confirma que al menos las secuencias que están flanqueadas por el promotor y el terminador cumplen con el tamaño esperado y con la dirección correcta, sugiriendo que el menor tamaño del constructo (de 5 Kb cuando se esperaba de 11 Kb) puede deberse a algún daño en las regiones de los promotores y/o terminadores.

Se realizó la digestión con la enzima *EcoRI* la cual debería dar un patrón de digestión de tres fragmentos si el constructo pAN7.1-GFP-GPI estuviera completo y dos bandas para

el pAN7.1 vacío. Se observó que el patrón de digestión no correspondía a lo esperado con 2 en vez de 3 bandas (Fig. 3.18) en dos clonas analizadas, clona 2 (Ve2) y clona 3 (Ve3), los constructos pAN7.1-GFP-GPI, mientras que el vector pAN7.1 vacío sí presenta las bandas esperadas.

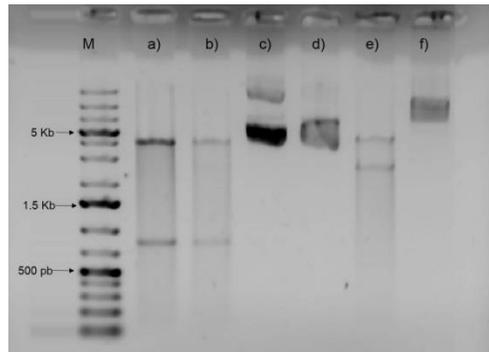


Figura 3.18 Digestión con *EcoRI* de los vectores pAN7.1 y pAN7.1-GFP-GPI. Digestión con *EcoRI*: en a) pAN7.1-GFP-GPI Clona 2 (Ve2), en b) pAN7.1-GFP-GPI clona3 (Ve3), en e) pAN7.1; plásmidos sin digerir en c) Ve2, d) Ve3 y f) pAN7.1 y M: marcador de peso molecular Gene ruler 1 Kb plus Ladder (ThermoFisher Scientific)

Debido a estos resultados se decidió construir de nuevo el vector, evitando ensamblar mediante Gibson® el vector pAN7.1 y el casete construido (pGPD-Ss-GFP-GPI-tTRPC), ya que este vector posee en el casete de la higromicina el mismo promotor (para la gliceraldehído 3 fosfato deshidrogenasa) y terminador (para triptófano) de los usados aquí en la construcción del casete. Por ello se consideró ensamblar el casete pGPD-Ss-GFP-GPI-tTRPC mediante tecnología Gibson® y una vez hecha la ligación usarla como templado para una reacción de PCR con una TAQ polimerasa de alta fidelidad, digerir con *XbaI* el amplicón y también el vector pAN7.1 con *XbaI* y ligar ambos.

Para ello se amplificaron nuevamente las secuencias del promotor y terminador a partir del vector pAN7.1 vacío (original), y por otro lado se utilizaron los cebadores que flanquean la región codificante (Secuencia señal de secreción forward-GPI reverse) para amplificar esta secuencia (Ss-GFP-GPI) del constructo pAN7.1-GFP-GPI. La PCR usando diferentes temperaturas produjo múltiples bandas pero también se observó una banda del tamaño esperado, aproximadamente de 1700 pb (Fig. 3.19). El fragmento de interés se purificó y la concentración se ajustó para realizar la ligación por el sistema Gibson® con el promotor y terminador, pero ahora sin el vector esqueleto, como se mencionó en el párrafo anterior. Después del ensamble se realizó una reacción de PCR usando la enzima TAQ polimerasa de alta fidelidad pfx, para amplificar el fragmento de aproximadamente 4800 pb. , digerirlo con la enzima *Xba*I y hacer la ligación en el vector pAN7.1 digerido con la misma enzima.

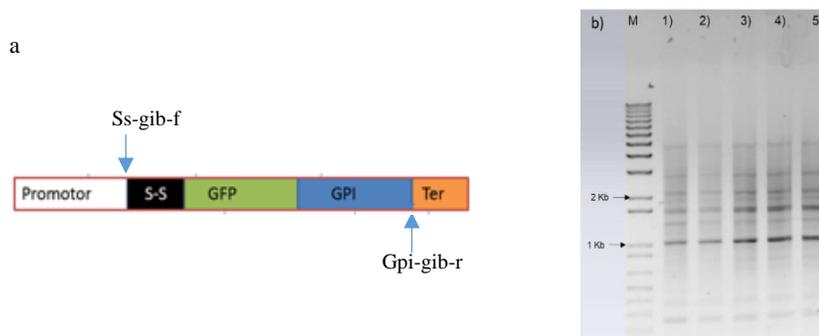


Figura 3.19 Amplificación de la secuencia ORF (Ss-GFP-GPI) usando como template el vector pAN7.1-GFP-GPI: a) Se señala la ubicación de los cebadores utilizados para la amplificación. b) Visualización en gel de agarosa al 1% de los amplicones generados a diferentes temperaturas: 1) 50, 2) 52 3) 56, 4) 58, 5) 60 °C y M: marcador de peso molecular 1 Kb plus ladder (ThermoFisher Scientific).

Lamentablemente, no se obtuvo el producto esperado después de la ligación mediante Gibson® y la amplificación del casete completo, aun utilizando gradiente de temperaturas. Así mismo se intentó hacer ligación por PCR, utilizando la mezcla de los cebadores que generan los fragmentos individuales para ensamble Gibson® (es decir, los cebadores que traslapan una región del fragmento anterior y una región del fragmento siguiente), pero los resultados fueron negativos todas las veces. Se consideró clonar el casete en el vector TOPO-TA para digerirlo con *Xba*I pero no fue posible porque no se logró amplificar el casete (~4.8 Kb) mediante esta estrategia.

3.3.7 Establecimiento del protocolo de transformación utilizando el plásmido GFP.

A pesar de los resultados negativos en la generación del plásmido que contiene el casete pGPD-Ss-GFP-GPI-tTRPC de expresión superficial, se procedió a establecer el protocolo de transformación, ya que es una metodología que se requería en el presente trabajo y que se podía desarrollar de manera paralela. Los primeros ensayos se realizaron utilizando una cepa de *T. longibrachiatum* por ser muy cercano a *T. reesei* que es la especie con la que se inició la búsqueda de proteínas GPIs; se probaron diferentes protocolos de transformación (Tabla 3.1) utilizando tres plásmidos distintos: pAN7.1, peGFP (donados por sus creadores) y pAN7.1-GFP-GPI (el primer vector construido en este trabajo). Con esta cepa todos los esfuerzos resultaron infructuosos y no se lograron obtener transformantes. En varias ocasiones las cepas presuntamente transformadas lograban sobrevivir sobre el agente de selección cuando se cultivaban en sólido, pero no crecían en medio líquido con las mismas concentraciones del fungicida selector. Se decidió probar con una cepa de *T. harzianum*, ya que es una de las especies más utilizadas de *Trichoderma* y también porque se observó que genera más esporas que la cepa de *T. longibrachiatum*.

Cuadro 3.1 Metodologías utilizadas en el presente trabajo para la transformación de *Trichoderma* spp

Protocolo	Condiciones	Concentración de adn y protoplastos	Resultado
Electorporación	Voltaje: 2.8 kv Amperaje: 800 ohms Resistencia: 25 F	2 µg ADN 400 µL (1 * 10 ⁶ protoplastos/mL)	Los protoplastos no fueron viables después del choque eléctrico
	Voltaje: 2.0 kv y 2.2 Kv Amperaje: 800 ohms Resistencia: 25F	1 y 2 µg ADN 400 µL 1 * 10 ⁶ protoplastos/mL	Crecimiento de <i>T. longibrachiatum</i> en 12.5 µL de Higromicina en medio sólido pero no en medio líquido con el

			selector
	Voltaje: 2.0 kv Amperaje: 800 ohms Resistencia: 25F	2 µg ADN 400 µL 1 * 10 ⁶ protoplastos/mL de <i>T. harzianum</i>	No se obtuvieron transformantes
Electroporación	Voltaje 2.0 KV Tiempo constante 5 ms	2 µg ADN 400 µL 1 * 10 ⁶ protoplastos/mL de <i>T. harzianum</i> ó <i>T. longibrachiatum</i>	Los protoplastos no fueron viables después del choque eléctrico
Peg y cloruro de calcio	PEG 60% (p/v)	2 µg ADN 100 µL 1 * 10 ⁶ protoplastos/mL de <i>T. harzianum</i> ó <i>T. longibrachiatum</i>	No se obtuvieron transformantes
		3, 4 y 5 µg ADN 100 µL 5 * 10 ⁶ protoplastos/mL de <i>T. harzianum</i>	Se obtuvo un transformante con el vector pAN7.1
		6 µg para GFP, 9 µg pan7.1 y 10 µg pAn7.1-GFP-GPI	Se obtuvieron transformantes con los tres vectores
Acetato de litio y perlas de cristal	Acetato de litio 0.1 M Esporas germinadas	1 µg ADN 400 µL 1 * 10 ⁶ esporas/mL de <i>T. harzianum</i>	No se obtuvieron transformantes

Mediante el protocolo de transformación de PEG/cloruro de calcio, empleando de 5 a 10 µg de plásmido se obtuvieron transformantes de la cepa de *T. harzianum*; cada transformante se aisló al tercer día y se colocó en medio líquido y sólido, ambos con higromicina B (100 µg/mL), creciendo el hongo en ambas condiciones.

Se observó diferencia en la velocidad de crecimiento de las cepas; mientras que aquellas transformadas con pEGFP y pAN7.1 crecían a velocidad similar a la silvestre, la cepa transformada con pAN7.1-GFP-GPI crecía más lento tanto en los cultivos sólido como líquido (Fig. 3.20). Esto podría estar relacionado con el posible daño en los promotores que flanquean dos de los casetes de expresión del constructo: el casete de expresión superficial y el casete de resistencia a higromicina. Es decir, la expresión del inserto Ss-GFP-GPI no sería óptima, pero tampoco la expresión del gen de resistencia a la higromicina.

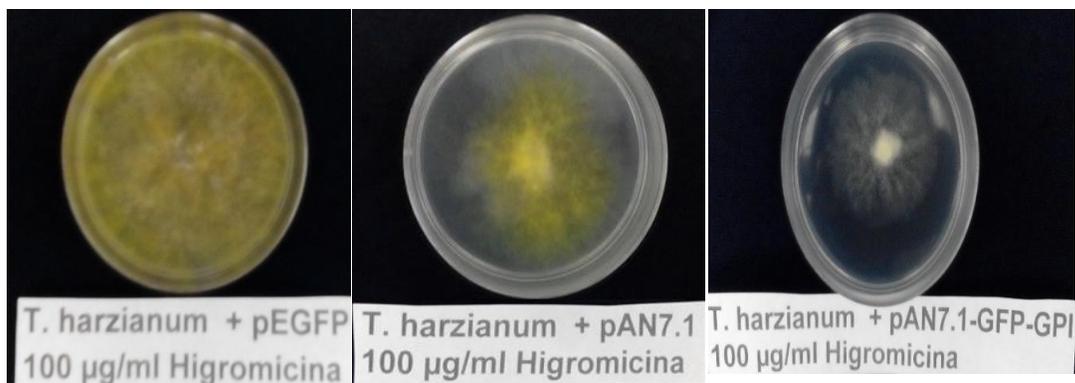


Figura 3.20 Morfología de las cepas de *T. harzianum* transformada con diferentes plásmidos al 4 día post inoculación: a) pEGFP, b) pAN7.1 c) pAN7.1-GFP-GPI.

Para corroborar la transformación se extrajo el ADN y se amplificó la región que codifica para el gen *Hph* (Fig. 3.21).

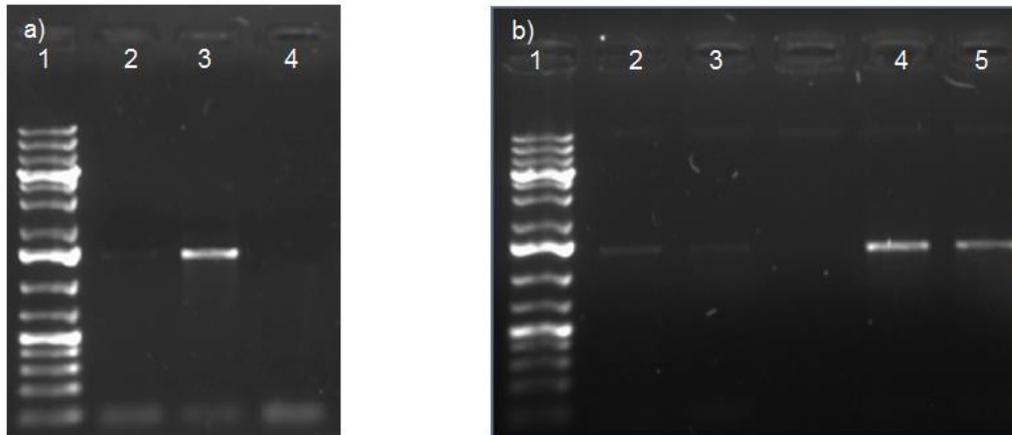


Figura 3.21 Visualización en gel de agarosa de la amplificación de la secuencia que codifica para el gen de *Hph* a partir de ADN obtenido de tres transformantes de *T. harzianum*: En a) y en b) carril 1 corresponde al marcador de peso molecular Gene ruler 1 Kb plus (ThermpFisherScientific); a) peGFP; carril 2 y 3 transformantes, carril 4 control negativo, b) en carril 2 y 3 pAN7.1-GFP-GPI, en 4 y 5 pAN7.1.

Se observó que el amplicón correspondiente al fragmento del gen *Hph* en las cepas transformadas con el plásmido pAN7.1-GFP-GPI (Fig. 3.21 b1 y 2) presentaban una señal relativamente baja comparándola con los transformantes con peGFP (Fig 3.21 a3) y con los de pAN7.1 (Fig. 3.21 b4 y 5).

4.3.3 Visualización de la muestras por microscopía

Una vez confirmada la generación de algunas clonas transformadas de *T. harzianum* se procedió a observar la fluorescencia de las transformantes con el plásmido pAN7.1-GFP-GPI para evaluar si se observa fluorescencia, lo que indicaría que la proteína GFP fusionada a la proteína con anclaje GPI si se está expresando y determinar si esta fluorescencia se localiza en la superficie. Las cepas se incubaron en líquido y se tomó micelio de tres días de edad para colocar en el portaobjetos estéril, se incubaron una noche a temperatura ambiente.

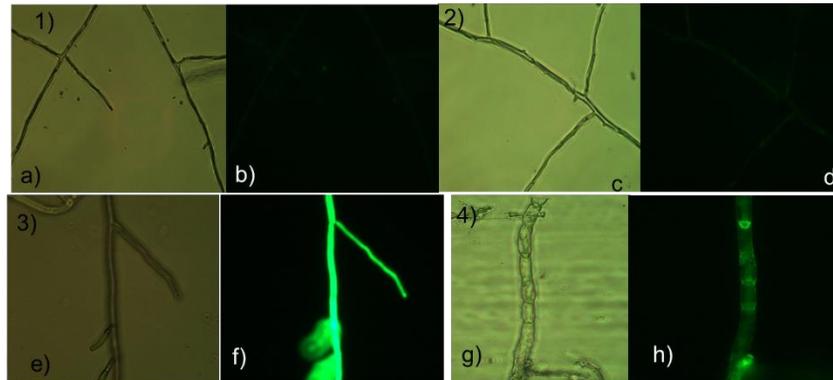


Figura 3.22 Visualización mediante microscopía de fluorescencia (40X) de cepas transformadas: 1) *T. harzianum* sin transformar a) campo claro, b) campo oscuro; 2) *T. harzianum* transformada con pAN7.1, c) campo claro, d) campo oscuro, 3) *T. harzianum* transformada con peGFP, e) campo claro, f) campo oscuro, 4) *T. harzianum* transformada con pAN7.1-GFP-GPI, g) campo claro, h) campo oscuro.

La fluorescencia se observó fuerte y en todo el micelio en la transformante con el plásmido peGFP, mientras que en la transformante con el plásmido pAN7.1-GFP-GPI, se observó fluorescencia tenue y localizada en los septos, lo que sería congruente con la localización superficial esperada en la pared celular (Fig. 3.22d). Sin embargo, las cepas que no expresan GFP tienen una débil fluorescencia basal que se observa difusa en la hifa, siendo más fuerte en micelio con mayor edad (imagen no mostrada). El resultado observado es similar a algunos reportados en la literatura que indican que *Trichoderma* spp. presenta auto fluorescencia (Lu et al., 2004). Para tener resultados más concluyentes las mismas clonas se observaron en microscopio confocal, el cual permite eliminar ruido o señal que no provenga de la proteína GFP.

Se realizaron modificaciones al protocolo previamente mencionado para la visualización en el microscopio confocal; en vez de usar hifas crecidas en medio sólido se creció en medio líquido y se montaron con glicerol al menos 2 horas antes de la visualización.

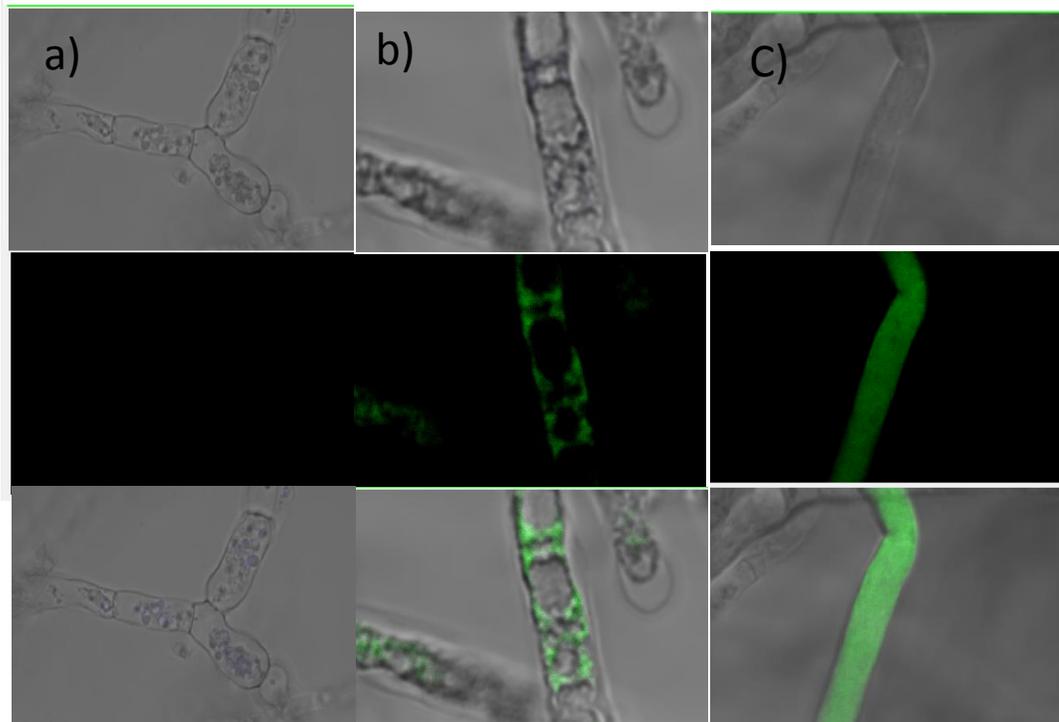


Figura 3.23 Visualización por microscopía confocal de los diferentes transformantes de *T. harzianum* observado a 60X: arriba: campo normal, medio: campo de fluorescencia, abajo: contraste de fases, a) *T. harzianum* transformado con pAN7.1, b) *T. harzianum* transformado con pAN7.1-GFP-GPI, c) *T. harzianum* transformado con peGFP.

Los resultados fueron similares a los observados con el microscopio de epifluorescencia; la fluorescencia de la cepa transformada con el vector peGFP fue intensa y cubría toda la hifa (debido a que la localización de la GFP es citosólica), mientras que la cepa transformada con el vector pAN7.1-GFP-GPI presentó fluorescencia localizada en los extremos de la hifa y/o en los septos. Por su parte la cepa transformada con pAN7.1 muestra una ligera fluorescencia que se pierde en el momento de enfocar las hifas (Fig. 3.23a). La presencia dirigida hacia la parte externa de la célula fúngica en la cepa transformada con pAN7.1-GFP-GPI indica que a pesar de que el tamaño del vector no es el esperado, se está llevando a cabo la expresión de la proteína quimérica (Ss-GFP-GPI) y aparentemente la secreción y ubicación en la pared no se perdieron en su totalidad por la pérdida de fragmentos en el vector. Sin embargo la dificultad de crecimiento de la clona sí puede estar asociada con el daño ocurrido en el vector: si la secuencia del promotor no está completa, la expresión de GFP-GPI pudiera ser menor de lo que se diera si el

constructo se hubiera logrado como originalmente se planeó. Asimismo, la expresión del gen de selección también puede ser ineficiente y eso explicaría el fenotipo de la cepa que crece muy lentamente sobre la higromicina. Aunque el resultado obtenido es alentador y apoya que el diseño del casete Ss-GFP-GPI fue correcto y la localización de la proteína es la esperada, el fenotipo de lento crecimiento de la clona y la posible subexpresión de la proteína recombinante son incompatibles con una aplicación biotecnológica, por lo que se prosiguió a una tercera generación del vector, cambiando el vector receptor (“backbone”) por otro que difiera en el promotor que usa en sus casetes de expresión. Se optó por usar el vector pSilent en lugar del pAN7.1.

3.3.7 Ligación múltiple mediante el sistema Gibson® utilizando el vector pSilent-1 como receptor del casete de expresión Ss- GFP-GPI

Se decidió repetir la ligación con el sistema Gibson® para la generación del casete pero ligando éste en un vector que no contuviera la secuencia promotora *PGD* y que presentara el mismo sitio de corte único en *Xba*I en el terminador. El vector pSilent-1 cumple con estos criterios: en el casete de resistencia a higromicina (*Hph*) usa al promotor *TRPC* de *A. fumigatus*. La digestión con *Xba*I del vector pSilent-1 libera dos fragmentos (fig. 3.24), uno de ~4.8 Kb en el cual se ubica el sitio de origen de replicación en bacteria, el origen de replicación en hongos y la unidad de expresión para el gen *HPH*; en el segundo fragmento (de ~3 Kb) se ubica el casete de silenciamiento. Se purificó el fragmento de mayor tamaño, se religó con el casete Ss- GFP-GPI y el constructo se utilizó para transformar TOPO10 *E. coli*, con el fin de corroborar que esté tuviera las secuencias necesarias para su replicación en la bacteria así como todos los fragmentos de ADN que se esperan en el constructo.

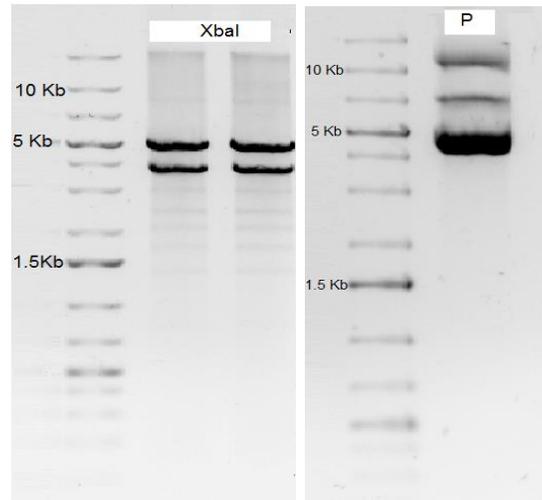


Figura 3.24 Visualización en gel de agarosa de la digestión por *XbaI* de pSILENT1. a) Digestión, b) pSILENT1

Después del ensamble mediante reacción de Gibson® del casete de expresión superficial, se tomaron 5 μ L de la reacción y se transformó la cepa de *E. coli* DH5 α ; se obtuvieron 63 colonias, de las cuales se analizaron por PCR 9 seleccionadas al azar. Para la PCR en colonia se utilizaron los cebadores que amplifican el gen GFP (Figura 3.25); el plásmido generado fue nombrado pHyg-GFP-GPI.

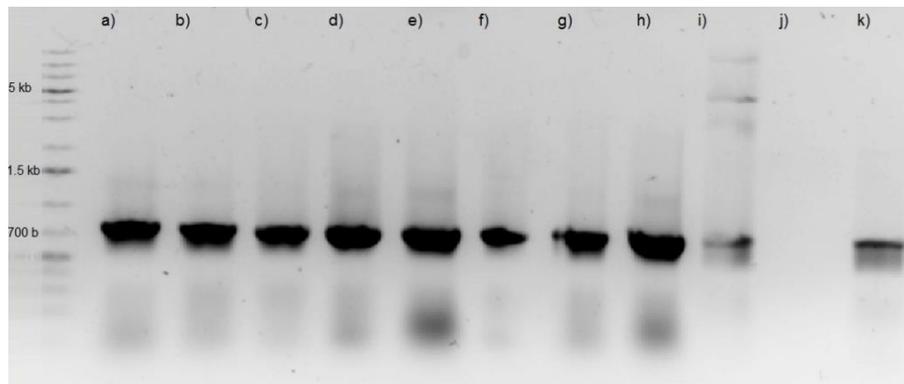


Figura 3.25 Confirmación por PCR en colonia del fragmento que codifica para GFP (752 pb) utilizando como templado las clonas transformadas con el plásmido pHyg-GFP-GPI; Carril 1) Marcador de peso molecular, a-i) clonas 1, 20, 25, 30, 40, 45, 50, 55 y 63 respectivamente, j) control negativo, k) control positivo utilizando como templado el plásmido peGFP.

3.3.8 Confirmación de los sitios de inserción en el plásmido pHyg-GFP-GPI

Con el fin de corroborar que el ensamble se realizó de manera correcta y que el plásmido sea del tamaño esperado, se seleccionó una de las clonas positivas que amplificaron el fragmento del gen GFP; el plásmido se extrajo y se digirió con tres enzimas de restricción, dos de ellas (c/u) que lo linealizarían con sitios de cortes únicos introducidos con los cebadores; *Apal* que flanquea en el 3' el fragmento que codifica para la GPI, *SpeI* que flanquea en el 3' el fragmento que codifica para GFP. También se usó *XbaI* que digiere en dos fragmentos el constructo, de tamaños esperados 4.5 y 4.8 Kb.

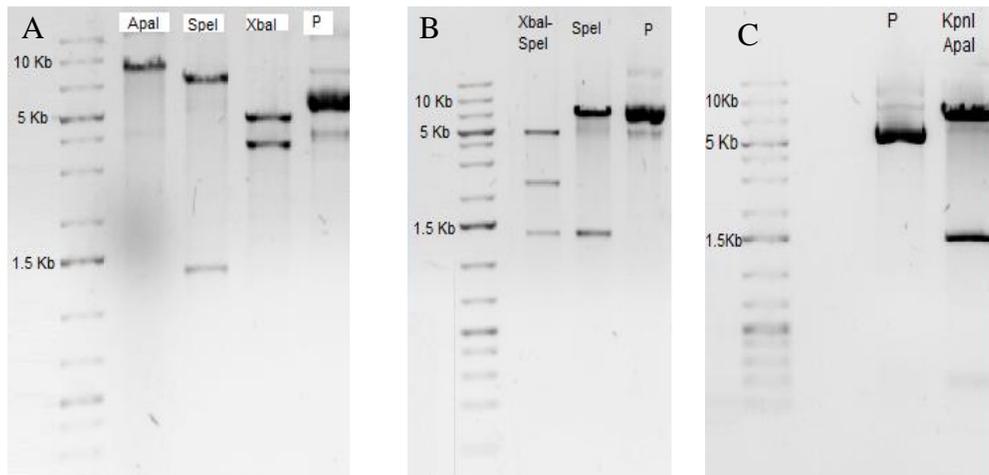


Figura 3.26 Perfil de digestión del plásmido pHyg-Gfp-GPI; Carril 1 marcador de peso molecular, P: plásmido sin digerir, En A) carril 2 digestión con *Apal*, carril 3 digestión con *Spel*, carril 4 digestión con *Xbal*; En B) Carril 2; doble digestión con *Xbal-Spel*, carril 3 digestión con *Spel*; En C) carril 4; doble digestión con las enzimas *KpnI-Apal*.

En la figura 3.26A se observan los perfiles de digestión del plásmido pHyg-GFP-GPI con las tres enzimas seleccionadas: *Apal* presentó un sitio de corte único, linearizando el constructo y produciendo una banda bien definida cerca de 10 Kb, siendo el tamaño estimado *in silico* del constructo de 9.3 Kb. Por su parte *Xbal* genera un perfil de doble banda, ya que la enzima digiere el plásmido en las dos unidades de expresión: un fragmento de 4.8 Kb que contiene el casete de expresión del gen *Hph* y de ampicilina (*Amp*), y otro fragmento de 4.5 Kb que contiene el casete de expresión de la proteína con anclaje GPI, unida a la GFP. Sin embargo, *Spel* digirió en dos fragmentos el vector, un fragmento cercano a los 1.5 Kb y un fragmento de mayor tamaño cercano a 7 Kb; este perfil de doble banda indica la aparición de un nuevo sitio *Spel* dentro de la secuencia, que no se había diseñado ni predicho *in silico*. Esto sugiere algún rearrreglo en el constructo, se analizó otra clona positiva y produjo los mismos resultados.

Se realizó una doble digestión utilizando las enzimas *Xbal* y *Spel* con el fin de determinar en cuál de las dos unidades de expresión se localizaba el nuevo sitio de restricción *Spel* (Fig.3.26B). El perfil de digestión reveló que la banda intermedia correspondiente al casete de expresión de Ss-GFP-GPI ya que ésta disminuye su tamaño cuando se realiza la doble digestión, mientras que la banda de 1.5 Kb que corresponde a la banda generada

por *SpeI* sobre el otro fragmento del plásmido se mantiene del mismo tamaño. Con el fin de determinar si los sitios de restricción que flanqueaban la región codificante Ss-GFP-GPI se habían insertado correctamente, el plásmido se sometió a una doble digestión con las enzimas *KpnI* y *ApaI*, las cuales generaron el perfil de bandeado esperado: un fragmento de aproximadamente 1.5 Kb el cual corresponde a la región codificante y uno mayor de 7 Kb que corresponde al resto del plásmido (Fig. 3.26 C).

3.3.9 Construcción del vector de expresión utilizando enzimas de restricción en pSilent-1

Debido a que los dos vectores que se generaron mediante el sistema Gibson® presentaban alguna característica no deseada, pero ambos tienen a *KpnI* y *ApaI* flanqueando el ORF que codifica la proteína quimérica Ss-GFP-GPI, se decidió subclonar éstos desde los vectores pAN7.1-GFP-GPI y pHyg-GFP-GPI al vector pCR®2.1 TOPO y posteriormente subclonar el ORF1 en el vector pSILENT1. Los tres constructos resultantes (pAN7.1-GFP-GPI y pHyg-GFP-GPI) y el pCR®2.1 TOPO *Hph* fueron sometidos a una digestión doble con las enzimas *KpnI*-*ApaI*; el vector pCR®2.1 TOPO vacío y digerido (Figura 3.27a) y los fragmentos digeridos correspondientes a los ORF de Ss-GFP-GPI de pAN7.1-GFP-GPI y pHyg-GFP-GPI se purificaron y se utilizaron para la ligación mediante una T4 ligasa.

Únicamente se obtuvieron clonas de la ligación del pCR®2.1TOPO y el fragmento de pAN7.1-GFP-GPI (ORF1). La clonación en TOPO-TA adiciona un sitio *HindIII* que queda río arriba en el extremo 5' del ORF1. Se seleccionaron dos clonas y se procedió a la digestión con *HindIII*. En ambas clonas se observó la linearización del vector (Figura 3.27b).

Una vez corroborada la presencia del sitio *Hind*III, se procedió a realizar una doble digestión usando las enzimas *Hind*III y *Apa*I con el fin de liberar el inserto del vector pCR®2.1TOPO-ORF1. Se observó una banda de ~1 Kb que es congruente con el tamaño del fragmento liberado del plásmido pAN7-1-GFP-GPI cuando es digerido con las enzimas *Kpn*I-*Apa*I. El ORF1 pCR®2.1TOPO está flanqueado por los sitios *Hind*III en 5' y *Apa*I en 3'-lo que permitió extraerlo y ligarlo en esos sitios en el plásmido pSilent-1, lo que elimina de éste la región que presenta el intron de cutinasa y permitiendo la expresión del ORF1. Se recuperó únicamente una clona positiva la cual se nombró pSmodORF1. La ligación del fragmento de pSilent (que contiene el casete de higromicina) con el casete Ss-GFP-GPI construido aquí genera un sitio *Kpn*I; la clona se sometió a una doble digestión con las enzimas *Kpn*I-*Apa*I confirmando la presencia del sitio *Kpn*I y la liberación del fragmento esperado de 900 pb (Figura 3. 28b).

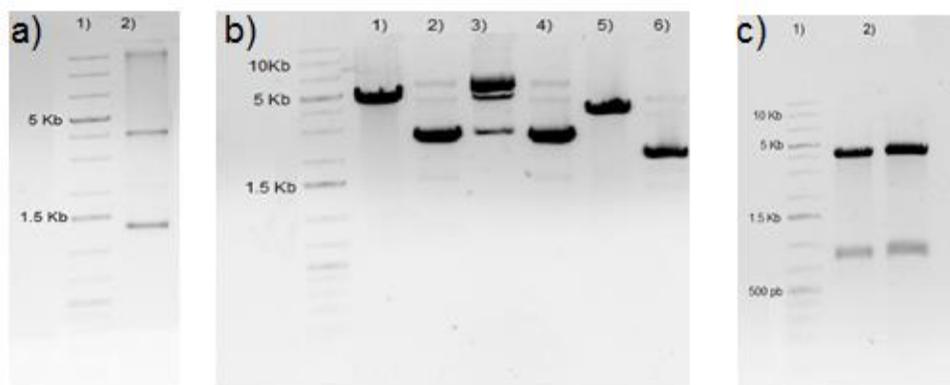


Figura 3.27 Análisis en gel de agarosa al 1% del constructo pCR®2.1TOPO-ORF1 (ORF1 corresponde al casete P_{gpd} -Ss-GFP-GPI- Ter_{trpC} aislado del primer constructo en pAN7.1 y pCR®2.1TOPO-*HPH* (*HPH* es un fragmento de pSilent que contiene el gen que codifica la resistencia a higromicina). En a) carril 1: marcador de peso molecular, carril 2: digestión del plásmido pCR®2.1TOPO-*HPH* con las enzimas de restricción *Hind*III y *Apa*I; se observan las bandas esperadas de 1.2 Kb correspondiente al fragmento del gen *HPH* subclonado en el plásmido y 4 Kb correspondiente al plásmido sin inserto, En b) carril 1) y 3), perfil de restricción con *Hind*III de pCR®2.1TOPO-ORF1 clona 2 y 10, 2) y 4) plásmidos pCR®2.1TOPO-ORF1 clona 2 y 10 sin digerir respectivamente, en 5) perfil de restricción de pCR®2.1TOPO con *Hind*III y 6) plásmido sin digerir pCR®2.1TOPO. En c) carril 2) perfil de digestión con las enzimas de restricción *Hind*III y *Apa*I de pCR®2.1TOPO-ORF1 de la clona 2 y 10 respectivamente.

Estos resultados sugieren que el constructo de pSmodORF1 fue exitoso. Es importante resecuenciar el inserto para corroborar que no hubo ningún desfase de los ORFs, y no se introdujo ni eliminó accidentalmente ningún sitio de restricción, como sucedió en el caso anterior para *SpeI*.

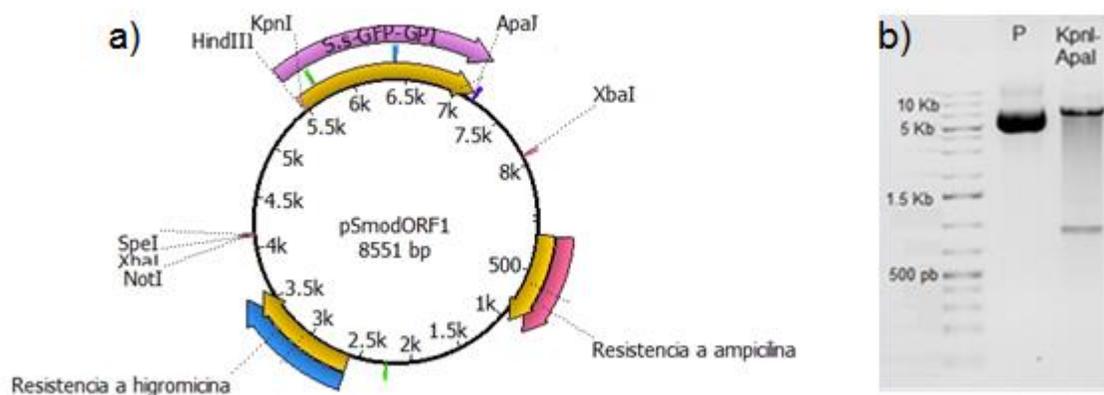


Figura 3.28 Esquemización del constructo pSmodORF1 y su análisis de digestión; a) Esquema general del vector con sus respectivos sitios de restricción; b) Carril 1, Marcador de peso molecular, P: plásmido sin digerir, *KpnI-ApaI*: Doble digestión.

La generación de los constructos se tuvo que detener aquí por terminarse el programa de Doctorado. El resultado final fue alentador pero tiene que corroborarse transformando *E. coli* y posteriormente *T. harzianum*. En caso de confirmarse la obtención correcta del constructo se podrá validar la expresión superficial de GFP y de la enzima lipasa mediada por las proteínas GPI identificadas en este trabajo y descritas en el capítulo 2 como prueba de concepto del desarrollo tecnológico planteado en este trabajo de tesis.

3.4 DISCUSIÓN

La necesidad del estudio de proteínas bajo condiciones específicas *in vivo*, el rediseño de sistemas biológicos existentes o la construcción de nuevos sistemas de expresión, ha llevado a desarrollar un sinnúmero de estrategias novedosas que permitan generar los fragmentos o la información genética necesaria de manera sintética o mediante sistemas de ensamblaje. (Juhas y Ajioka, 2017). A pesar que las tecnologías en la síntesis de ADN mejoran continuamente, todavía presentan deficiencias o complicaciones en la síntesis de fragmentos largos *ex vivo*, siendo costosos y pocos accesibles (Casini *et al.*, 2015).

Por su parte los sistemas de ensamblaje de ADN se han diversificado y mejorado en sus capacidades del número de secuencias a ensamblar. Actualmente los sistemas se pueden clasificar en tres categorías: a) basados en el uso de enzimas de restricción tipo IIs las cuales difieren del tipo II por su capacidad de cortar el ADN a unas pocas bases a distancia de su sitio de reconocimiento. Sistema como MoClo®, GoldenGate® y GoldenBraid® utilizan esta tecnología. b) basados en serina integrasas o en sitios específicos de recombinación, los cuales omiten el uso de endonucleasas mediante sitios específicos que son reconocidos por enzimas de bacteriófagos (ejemplo *att* del fago lambda) que llevan a cabo el rearrreglo entre las secuencias de ADN, c) métodos basados en “End-homology”, los cuales permiten el ensamblaje de múltiples secuencias en una sola reacción utilizando la homología en los finales entre las partes cercanas, como el sistema Gibson®. En el presente trabajo se lograron ensamblar entre 4 y hasta 6 fragmentos, siendo los productos finales de 5 Kb y 11 Kb.

El sistema de Gibson® ha sido un método exitoso en la generación y ensamblaje de múltiples fragmentos de ADN de tamaños largos (Pollak *et al.*, 2018; Anderson *et al.*, 2010) y ha sido ampliamente adoptado por su versatilidad, relativa eficiencia y la generación de fragmentos “libres de cicatrices” (Pollak *et al.*, 2018). La dependencia del sistema para el ensamblaje de regiones homólogas traslapadas es uno de los principales puntos a considerar debido a la posibilidad de introducir errores en los oligonucleótidos diseñados (Pollak *et al.*, 2018; Anderson *et al.*, 2010), así como también la composición de las secuencias *e.g.* contenido alto de G y C (Li *et al.*, 2018; Casini *et al.*, 2015) y elementos repetitivos que dificultan o disminuyen la eficiencia de la ligación y pueden llevar a cabo deleciones y/o rearrreglos en el sistema. No hay artículos que describan lo

que sucedió en el presente trabajo. Sin embargo, los resultados de las amplificaciones y los perfiles de digestión en ambos constructos generados con el sistema Gibson® apuntan a un re arreglo provocado por usar secuencias promotoras y terminadoras idénticas entre el casete de expresión superficial y el vector destino pAN7-1 o la aparición de un nuevo sitio SpeI cuando se usó el fragmento de pSilent-1 que contiene el casete de expresión del gen *HPH* y *AMP*.

A pesar de que el tamaño del constructo pAN7.1-GFP-GPI no correspondió a lo esperado, el casete de expresión que otorga resistencia a la higromicina se mantuvo funcional, al menos parcialmente; el análisis de las mutantes por microscopía confocal, reveló fluorescencia asociada a regiones de la pared celular, confirmando que el casete de expresión de la proteína quimérica Ss-GFP-GPI es funcional. El tamaño del constructo fue significativamente más pequeño de lo esperado (~5 Kb en vez de 11 Kb) posiblemente porque en el proceso de ensamble mediante el sistema Gibson® ocurrió recombinación por homología entre las secuencias promotoras del casete construido aquí y el casete de higromicina del vector receptor. Cada casete posiblemente mantuvo una región mínima funcional que permitió respectivamente la expresión basal de la proteína quimérica Ss-GFP-GPI y del gen de resistencia a la higromicina. Es ampliamente conocido que los promotores pueden funcionar con regiones mínimas, *e.g.* eliminaciones en el promotor LIP1 de *Fusarium graminearum* identificaron una región mínima de 563 pb río arriba del codón de inicio; esta región permitió la expresión del gen reportero GFP; se identificaron también elementos inducibles entre la región 563-372 pb (Feng *et al.*, 2011).

La generación de un vector pcrTOPO2.1 que contiene únicamente los sitios de restricción *HindIII*, *KpnI* en 5' y *Apal* en 3' flanqueando la región codificante Ss-GFP-GPI permitió su subclonación en el vector pSilent1; éste fue previamente modificado eliminando la región para el silenciamiento y utilizando el resto del vector de expresión sin la necesidad de subclonar fragmentos que codifiquen terminadores o promotores para Ss-GFP-GPI.

Lamentablemente por finalizarse el tiempo del Doctorado no se pudo evaluar la expresión del ORF en el constructo final, pero la validación mediante PCR de la presencia y orden de los fragmentos, así como el tamaño de los fragmentos obtenidos con diferentes enzimas de restricción apoyan que finalmente se logró el vector diseñado. Al transformar *T. harzianum*, de lograrse una clona de crecimiento más rápido que el logrado en los

intento anteriores, y que la fluorescencia estuviera fuerte y localizada en la superficie, ya se tendría el vector listo para recibir la enzima para la prueba de concepto de expresión superficial. Para la prueba se decidió usar una triacilglicerol lipasa de *T. harzianum*, que sea capaz de actuar sobre triacilgliceroles con ácidos grasos de cadena larga. Para ello en el siguiente capítulo se presenta la estrategia que se siguió para identificar lipasas y seleccionar para su futura expresión superficial de una lipasa de *T. harzianum*.

CAPITULO IV**FUNGAL SCREENING ON OLIVE OIL FOR EXTRACELLULAR TRIACYLGLYCEROL LIPASES: SELECTION OF A *TRICHODERMA HARZIANUM* STRAIN AND GENOME WIDE SEARCH FOR THE GENES**

El capítulo IV corresponde al artículo publicado en GENES <http://www.mdpi.com/2073-4425/9/2/62/s1>. A partir de una cepa de *Trichoderma harzianum* identificada como productora de lipasas extracelulares, se decidió hacer una búsqueda en el genoma de los genes responsables de esta actividad y su caracterización *in silico* mediante análisis filogenéticos, búsqueda de homólogos, modelamiento tridimensional e identificación de dominios y sitios catalíticos. La estrategia consistió en una búsqueda en el genoma por palabras claves, seguido de la caracterización *in silico*; aquellas lipasas que fueran secretadas y tuvieran homología con triacilglicerol lipasas o dominios de éstas, fueron seleccionadas para analizar por RT-PCR sus perfiles de expresión en el hongo cultivado en medio mínimo suplementado con aceite de oliva.

Se identificó una lipasa que se induce por el aceite de oliva y la región codificante de ésta será usada para la expresión superficial, sustituyendo la secuencia que codifica la proteína GFP en el constructo generado en el capítulo III.

La lista de autores y su afiliación se describe a continuación:

Miguel Angel Canseco-Pérez 1, Genny Margarita Castillo-Avila 1,2, Bartolomé Chi-Manzanero 1, Ignacio Islas-Flores 3, Max M. Apolinar-Hernández 1, Gerardo Rivera-Muñoz 4, Marcela Gamboa-Angulo 1, Felipe Sánchez-Teyer 1, Yeny Couoh-Uicab 5 and Blondy Canto-Canché 1,*

1 Unidad de Biotecnología, Centro de Investigación Científica de Yucatán, A.C., Calle 43 No. 130 X 32 y 34, Col. Chuburná de Hidalgo, C.P. 97205 Mérida, México

2 Centro de Estudios Tecnológicos del Mar No. 17, km 1.5 Carr. Antigua Chelem, C.P. 97320 Yucalpetén, Mexico

3 Unidad de Bioquímica y Biología Molecular de Plantas, Centro de Investigación Científica de Yucatán, A.C., Calle 43 No. 130 X 32 y 34, Col. Chuburná de Hidalgo, C.P. 97205 Mérida, México

4 Departamento de Ingeniería Química y Bioquímica, Tecnológico Nacional de México, Campus Instituto Tecnológico de Mérida, Km. 5 Carr. Mérida-Progreso S/N, C.P. 97118 Mérida, México

5 División de Ingeniería Bioquímica, Instituto Tecnológico Superior de Purísima del Rincón. Blvd del Valle #2301, Guardarrayas, C.P. 36413 Purísima del Rincón, Guanajuato, México

* Correspondence: cantocanche@cicy.mx; Tel.: +52-999-942-8330

Received: 7 November 2017 / Accepted: 20 December 2017 / Published: 25 January 2018

Abstract: A lipolytic screening with fungal strains isolated from lignocellulosic waste collected in banana plantation dumps was carried out. A *Trichoderma harzianum* strain (B13-1) showed good extracellular lipolytic activity (205 UmL⁻¹). Subsequently, functional screening of the lipolytic activity on Rhodamine B enriched with olive oil as the only carbon source was performed. The successful growth of the strain allows us to suggest that a true lipase is responsible for the lipolytic activity in the B13-1 strain. In order to identify the gene(s) encoding the protein responsible for the lipolytic activity, *in silico* identification and characterization of triacylglycerol lipases from *T. harzianum* is reported for the first time. A survey in the genome of this fungus retrieved 50 lipases; however, bioinformatic analyses and putative functional descriptions in different databases allowed us to choose seven lipases as candidates. Suitability of the bioinformatic screening to select the candidates was confirmed by reverse transcription polymerase chain reaction (RT-PCR). The gene codifying 526309 was expressed when the fungus grew in a medium with olive oil as carbon source. This protein shares homology with commercial lipases, making it a candidate for further applications. The success in identifying a lipase gene inducible with olive oil and the suitability of the functional screening and bioinformatic survey carried out herein, support the premise that the strategy can be used in other microorganisms with sequenced genomes to search for true lipases, or other enzymes belonging to large protein families.

Keywords: enzymes; lipases; triacylglycerol lipases; true lipases; olive oil induced lipases; protein bioinformatic analyses.

4.1 INTRODUCTION

Lipases are serine hydrolases defined as triacylglycerol acyl hydrolases (EC 3.1.1.3), which hydrolyze glycerol esters of long chain fatty acids (water insoluble). Many enzymes have activity on water soluble medium-chain triacylglycerols (<12 carbon atoms), long chain monoacyl glycerols or long-chain nitrophenyl acyl esters but are not able to act against triacylglycerols (TAGs) with long-chain fatty acids, such as vegetable oils. Those lipases are not considered true lipases but they are esterases (EC 3.1.1.1), which hydrolyze esters of short chain fatty acids (water soluble). For lipase catalysis there is a phenomenon called interfacial activation, which does not occur in esterases (Gopinath *et al.*, 2013). Substrate emulsion generates a lipid interface that activates the lipase: a loop

which covers the active site (lid) changes conformation and the enzyme shifts to the open conformation, where the active site is accessible to the substrates (Verger, 1997). In addition to hydrolyses, lipases are involved in conversion reactions in non-aqueous media: esterification, interesterification, transesterification, alcoholysis, acidolysis and aminolysis (Gopinath *et al.*, 2013; Nwuche y Ogbonna 2011).

Lipases have been known since 1930 and have captured the interest of scientists and industrial businessmen to the present day. These enzymes are used in the detergent industry, in the organic synthesis of pharmaceuticals, pesticides and insecticides, the production of emollient for personal care in cosmetics and more recently in transesterification reactions for biodiesel production and in the emerging oil and fat industry, for example production of cocoa butter equivalent (chocolate) and human milk fat substitute (Daiha *et al.*, 2015; Gopinath *et al.*, 2013; Andualema y Gessesse 2012). Recently, Daiha *et al.* (2015) reviewed contemporary research publications and patent applications dealing with lipases and concluded that these enzymes are of much relevance in modern industry; however, further research is still needed to reach their full potential.

Lipases are ubiquitous in living beings. In particular, microbial lipases are of great importance and have diverse commercial applications (Yadav *et al.*, 2012). Bacteria, yeasts, filamentous fungi and a few protozoa produce extracellular lipases for the digestion of lipid materials (Gopinath *et al.*, 2013). Screening for extracellular lipases is currently performed on isolated microorganisms, by microbial genome-wide mining, or by metagenomic mining. Lipolytic microorganisms have been isolated or identified in collections from soils, plant matter, plant compost, oleaginous fruits, domestic vegetable oil wastes, pig manure, organic dumps, garbage collectors and industrial effluents, among many other sources. Microbes are highly successful in adapting and surviving in a wide range of environments, by exploiting their trophic niche; thus, the ability to secrete enzymes is of great survival value.

There are many screening methods for lipolytic activity (Lanka y Latha, 2015). The most common is solid media added with different lipid substrates and lipase activity can be detected as clear or turbid zones around the colonies or by the production of crystals on the agar surface. Combination with Rhodamine B has proved to be a fast method; Rhodamine B binds fatty acids, mono- and diglycerides and develops fluorescence under ultra violet (UV) light (Lanka y Latha 2015; Gopinath *et al.*, 2013). Tributyrin-agar plate is a widely used method, but this substrate cannot distinguish between true lipases and

esterases. Olive oil-agar plates are a good test for screening true lipase-positive colonies (Ortiz-Lechuga *et al.*, 2016; Gopinath *et al.*, 2013). Lipase production is often induced by the presence of vegetable oils and the best results have been obtained with olive oil (Ortiz-Lechuga *et al.*, 2016; Coradi *et al.*, 2013; Falony *et al.*, 2006; Lima *et al.*, 2003).

In the present work, fungi isolated from banana plantation waste were screened for production of extracellular lipases on olive oil-Rhodamine B plates. *Trichoderma harzianum* and *Trichoderma longibrachiatum* were identified as the most prominent producers, the former being the best strain in our collection. We hypothesize that the lipase responsible for the extracellular activity on olive oil is a true lipase. Ülker *et al.* (2011) performed the first characterization of the extracellular lipolytic activity of *T. harzianum*. They found that the activity is optimum at 40 °C and that it is thermostable, showing 55% of activity when it was incubated at 50 °C and 20% with incubation at 70 °C. Coradi *et al.* (2013) precipitated the protein and found that the lipolytic enzyme worked on long-chain TAGs as triolein and olive oil, i.e., it is a true lipase.

To date there is no description regarding genes for lipases in *T. harzianum*. The goal of this work is to identify the gene or genes that encode the protein responsible for the extracellular lipolytic activity of this fungus on olive oil.

4.2 MATERIALS AND METHODS

4.2.1 Searching of Genes for True Lipases in *T. harzianum* Genome and *In Silico* Analysis of the Proteins

The deduced proteome of *T. harzianum* in the JGI genome portal (Nordberg *et al.*, 2014) was searched for lipases, followed by refining the quest for true lipases with the keywords triglyceride lipase, triacylglycerol lipase, and lipase class 3.

To support the quest of true lipases, BLASTp in the *T. harzianum* deduced filter models of proteins was performed by using as queries the amino acid sequences of reported lipases with activity on long-chain triacylglycerides (vegetable oils): *Fusarium graminearum* FGL1 AAQ23181 (Voigt *et al.*, 2005), *Fusarium heterosporum* AAB34680 (Nagao *et al.*, 1994); and the long-chain triacylglyceride lipases enlisted in the US 9476008 B2 patent (Haraldsson y Kristinsson, 2016): *Rhizomucor miehei* (P19515), *Thermomyces lanuginosus* (CAB58509), *Candida antarctica* LipA (2VEO) and LipB (P41365) and *Rhizopus oryzae* (AER14043).

For each protein hit retrieved, the model name, the scaffold location of their genes in the *T. harzianum* portal and their automatic annotations at portal were recorded.

4.2.2 *In Silico* Analysis *T. harzianum* Putative True Lipases

Lipases found in *T. harzianum* with the keywords triglyceride lipase, and those annotated as secreted lipases were selected for further bioinformatic analyses. Amino acid sequence in Fasta format was downloaded for each of the retrieved hits. Lipase candidates were examined at different platforms to search domains/motif information: the Common Domains database in the non-redundant fungal database at NCBI (Marchler-Bauer *et al.*, 2017), the Pfam databases using the sequence search tool (Finn *et al.*, 2016), the HHpred (Homology detection and structure prediction by HMM-HMM) (Alva *et al.*, 2016) and the InterPro database (Finn *et al.*, 2017).

For identification of extracellular lipases, amino acid sequences were analyzed with SignalP 4.1 server (Petersen *et al.*, 2011) to predict N-terminal secretion signal. Prediction of transmembrane helices was performed with TMHMM 2.0 server (Krogh *et al.*, 2001). The putative location in the cell was predicted with WoLF PSORT (Horton *et al.*, 2007). Molecular size was *in silico* calculated with Protein Molecular Weight Calculator—Science Gateway and ProtParam tool (Gasteiger *et al.*, 2005).

BLASTp 2.6.0 was performed for rapid assignment (Marchler-Bauer *et al.*, 2017; Marchler-Bauer *et al.*, 2015). Gene Ontology categories, i.e., Biological Process, Molecular Function and Cellular Component were retrieved from the different bioinformatics tools used to search homology, conserved domains, motifs and signatures (Pfam, InterPro, Superfamily, Protein Database, etc.). Each selected candidate was also examined in the genomic portal (clicking on the ID number) to search automatic annotations.

The complete amino acid sequence of each candidate was submitted by BLASTp to the LED (Lipase Engineering Database) (Fischer y Pleiss, 2003) and ESTHER Database (ESTerases and alpha/beta-Hydrolase Enzymes and Relatives) (Hotelier *et al.*, 2004) to inference class, parent family and probable function through the use of annotation transfer.

4.2.3 Phylogenetic Analysis of True Lipases from *T. harzianum*

A list of representative fungal lipases was taken from Yadav *et al.*, (2012) and the amino acid sequence from each was downloaded from the GenBank. The putative true lipases

retrieved in this study from the deduced proteome of *T. harzianum* and their characterized homologues or best hits in PSI-BLAST (Altschul *et al.*, 1997): *Tolypocladium ophioglossoides* (KND87948), *Metarhizium anisopliae* (KFG82987), *Nectria haematococca* (CAC19602), *Purpureocillium lilacinum* (XP_018176823), *N. haematococca* (XP_003053404), *Claviceps purpurea* (CCE34466), *Fusarium graminearum* (XP_01138453), *Hirsutella minnesotensis* (KJZ78693), *Fusarium oxysporum* (EWZ47945), *Saccharomyces cerevisiae* (AAA50367), *T. ophioglossoides* (KND93486), *Talaromyces cellulolyticus* (GAM42581), *Colletotrichum fioriniae* (XP_007591165), *T. ophioglossoides* (KND89897), *S. cerevisiae* (AJS39941), *P. lilacinum* (OAQ83707) and *Drechmeria coniospora* (ODA76916). The known true lipases (from literature) mentioned above were included, as well as Lip1 (P20261) and Lip5 (P32949) from *Candida rugosa* (Schmidt-Dannert, 1999), the proteins Lip1 (EAA67628, ACE80261 and XP_001800960) from the pathogenic fungi *Gibberella zeae* (Feng, 2005), *Blumeria graminis* (Feng *et al.*, 2009) and *Stagonospora nodorum* (Feng *et al.*, 2011), *Malassezia restricta* MrLip1 (ALG03641) (Park *et al.*, 2015), *Malassezia globosa* (XP_001732206) (Xu *et al.*, 2007), *Acremonium alcalophilum* LipA (POCT91) (Pereira *et al.*, 2013), the *S. cerevisiae*'s Tgl3p (AJS99689) and Tgl5p (DAA10860) which are the major TAG lipases in this yeast (Klein *et al.*, 2016), the *Ophiostoma piceae* sterol esterase (4UPD) (Cedillo *et al.*, 2012) and the esterase *Bjerkandera adusta* (APW29213), the latter with no activity on long-chain glycerides (Sánchez-Carbente *et al.*, 2017) which was included as one reference for non-true lipases. The lipases CAA00250 from *R. miehei* and Lip1 and Lip5 from *C. rugosa* are fungal phylogenetic references (Schmidt-Dannert, 1999).

A multi-alignment was performed in MAFFT (Katoh y Toh, 2008) with default parameters. A phylogenetic tree was constructed using the neighbor-joining method (Saitou y Nei, 1987) with 500 bootstrap.

4.2.4 *In Silico* Modeling

Three-dimensional modeling was performed with SWISS-MODEL (Biasini *et al.*, 2014), HHpred (Söding *et al.*, 2005) and I-TASSER (Yang *et al.*, 2015), using the complete amino sequence of each candidate. The retrieved hit with the highest score was selected in each case as the best model. The catalytic site and the lid domain were identified at IPBA web

server (Gelly *et al.*, 2011) by 3D superposition with the best template. Amino acid composition in the conserved motifs in the template were identified and labeled with PyMOL Molecular Graphics System program (DeLano, 2002) and then identified in the candidate by structure-based and sequence-based comparisons.

The PSI-BLAST (Position-Specific Iterated) at the NCBI (Altschul *et al.*, 1997) was conducted to search homologues in the protein structure database.

For TAG lipase which was de novo expressed on olive oil, a structure-based multi-alignment was conducted by ClustalW (Thompson *et al.*, 1994). The alignment was represented using ESPript (Robert y Gouet, 2014) .

4.2.5 RT-PCR Amplification of Lipases

To validate the selection of putative true lipases and to identify the gene/protein responsible for the extracellular lipolytic activity of *T. harzianum* strain B13-1 on olive oil, reverse transcription polymerase chain reaction (RT-PCR) was performed on RNA obtained from the fungus grown in olive oil-free and olive oil-containing medium (1% v/v).

The nucleotide coding regions of the selected lipases were downloaded from the *T. harzianum* database (Nordberg *et al.*, 2014) and sequence primers were designed with the Primer Quest Tool software (Nordberg *et al.*, 2014). The list and expected sizes of the fragments is shown in Table S 3.1.

Mycelium was harvested at 1, 3, 5 and 7 days. RNA was extracted with TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), pooled and cDNA was synthesized with Maxima First Strand cDNA Synthesis Kit (Thermo Scientific, Vilnius, Lithuania) following supplier's recommendations. Each reaction mixture (total volume of 15 µL) contained 1× PCR buffer, 2.0 mM MgCl₂, 1.6 µM dNTPs, 400 µM of each primer and 1.0 U of GoTaq DNA polymerase (Promega, Madison, WI, USA); 1 µL of cDNA was used as template per reaction. The PCR parameters were 3 min of 95 °C; followed by 30 cycles of 94 °C for 30 s, 53 °C for 30 s and 72 °C for 30 s; and a final elongation at 72 °C for 5 min. The PCR products were purified from gel using QIAquick Gel Extraction Kit (Qiagen, Germantown, MD, USA); for sequencing the samples were sent to LANBAMA. Sequences were edited as above and then used for BLAST in the *T. harzianum* genome and NCBI non-redundant protein database.

4.3. RESULTS AND DISCUSSION

4.3.1 Searching Genes of True Lipases in *T. harzianum*

A survey of the genome was conducted to explore the possibility of identifying the respective gene among the *T. harzianum* TAG lipases. To date, there have been no reports on the lipase genes in this microorganism. The quest in the *T. harzianum* deduced proteome at genomic portal (Nordberg *et al.*, 2014) using “lipase” as keyword resulted in 50 hits (Table S3.2), showing that the lipase family is large in this fungus. The family comprises alpha-beta hydrolases, phospholipases, carbohydrate esterases, isoamyl acetate-hydrolyzing esterase, among others. Four lipases (551811, 87496, 526309 and 514427) are annotated as secreted in the *T. harzianum* genome portal and are therefore plausible candidates to explain the extracellular lipolytic activity measured here in strain B13-1. To expand the list of candidates, the search in the genome for true lipases was performed by using the keywords “triglyceride lipase” and it retrieved 9 hits (IDs 510832, 79895, 92423, 135964, 492160, 77338, 78181, 502433 and 514252). The last four belong to lipase class 3, which is a lipase family of true lipases (NCBI Resource Coordinators 2016; Voigt *et al.*, 2005). Lipases with activity on long-chain TAG were used as queries to search for more true lipases in *T. harzianum*: *F. graminearum* FGL1 AAQ23181, *F. heterosporum* AAB34680, *R. miehei* P19515 and *R. oryzae* AER14043 retrieved all of the *T. harzianum* proteins 78181 and 77338; *T. lanuginosus* CAB58509 retrieved these two lipases and the protein 502433. These three lipases are in the list found previously. Therefore, the list comprises 13 candidates in the search for the putative true lipase responsible for the extracellular lipolytic activity on olive oil in *T. harzianum*.

4.3.2 *In Silico* Characterization of *T. harzianum* True Lipases

BLASTp 2.6.0 was performed for rapid assignment of lipase candidates, but for most of them the closest homologue was uncharacterized predicted lipase, or hypothetical protein, arising from genomic sequencing projects. Therefore, this analysis was not of much help in supporting or eliminating candidates. Genomic data is so abundant in the GenBank that usually 50-100 hits corresponded to uncharacterized predicted lipase, although some of the lipase candidates under analysis were retrieved by BLASTp from *T. harzianum* portal

using a characterized TAG lipase as query. However in the huge sequencing data, those queries ranked so far in comparison with many genomic sequences.

PSI-BLAST increases the opportunities to predict function, especially when *Trichoderma* species are excluded, because the searching result rules out the large list of predicted, hypothetical proteins that come from genomic sequencing data from this genus. Table 4.3.1 shows a summary of sequence-based function transfer based on PSI-BLAST.

Although BLASTp hits were not suitable to predict function by homology, the conserved domain searching tool hosted in BLAST server enabled us to find motifs and conserved domains in the protein candidates (Table 4.3.1). The analysis in the conserved domain database (CDD) confirms the lipase 3 domain in 77338, 78181, 514252, and 502433, consistent with the automatic annotation for these proteins in the genomic portal of *T. harzianum*. The Lipase 3 domain contains the flip/lid domain (TYITNTIIDLS in 77338, SNLRNFITDVV in 78181, TSTNDKVNDNL in 514252, and TLFEDVLADLT in 502433), the catalytic triad (Ser - His - Asp), and the nucleophilic elbow (GX SXG consensus motif) which is GHSLG in these four lipases. In addition to these domains, 502433 has one AF-4 domain. This domain is present in AF4 and FMR2 nuclear proteins. In *Drosophila* AF4 protein homologue acts in cytoskeleton regulation, segmentation and morphogenesis.

Since the survey performed in this work is for identifying the protein responsible for extracellular activity in B13-1 strain on olive oil, molecular size and cellular localization were analyzed. One of the putative lipase class 3 gave a negative result for signal peptide in SingalP program (502433), and its size is larger (115 kDa) than expected for extracellular lipases, which ranks from molecular weights less than 20 to 65 kDa (Mehta *et al.*, 2016; Sharma *et al.*, 2011). The other three class 3 lipases were predicted with signal peptide and as being extracellular, making them good candidates (Table 4.3.2).

Table 4.3.1 List of domains and family description of putative true lipases of *T. harzianum**.

Protein id	Superfamily	Domains	Domain description	Active site domain	Substrate binding pocket	Blast hits (search homologues in the first 50 hits)	Psi-blast (homologues characterized)
551811	Abhydrolase family cl21494	LIP pfam03583 Mal_quin-oxido TIGR01320	Secretory lipase: related with lipases from <i>C. albicans</i> . Malate:quinone-oxidoreductase: Membran-associated enzyme as part of the TCA cycle	Active site Ser - His - Asp/Glu. Nucleophilic attack on a carbonyl carbon atom.	Substrate binding pocket related to pfam03583	Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp.) as hypothetical protein and Predicted lipases	Lipase 1 <i>Candida albicans</i> (KGR02689) Query cover 94% Identity 33%
87496	Abhydrolase family cl21494	LIP pfam03583	Secretory lipase: Related with lipases from <i>C. albicans</i>	Active site Ser - His - Asp/Glu. Nucleophilic attack on a carbonyl carbon atom.	Substrate binding pocket related to pfam03583	Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp.) as hypothetical protein and Predicted lipases	Lipase 1 <i>C. albicans</i> (KGR02689) Query cover 93% Identity 32%

526309	Abhydrolase family cl21494	LIP pfam03583	Secretory lipase: Related with lipases from <i>C. albicans</i>	Active site Ser - His - Asp/Glu. Nucleophilic attack on a carbonyl carbon atom.	Substrate binding pocket related to pfam03583	GH16 protein [<i>Trichoderma guizhouense</i>]. Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp., <i>Pochonia</i> sp., <i>Cordyceps</i> sp., etc) as hypothetical protein. Secretory lipase. Probable lipase precursor	Secretory lipase 5 <i>C. albicans</i> (ADP21191.1) Query cover 92% Identity 39%
514427	Abhydrolase family cl21494	LIP pfam03583 DAP2	Secretory lipase; Related with lipases from <i>C. albicans</i> Dipeptidyl aminopeptidase/acyla minoacyl peptidase:Amino acid transport and metabolism Acetyl xylan esterase (AXE1).	Active site Ser - His - Asp/Glu. Nucleophilic attack on a carbonyl carbon atom.	Substrate binding pocket related to pfam03583	Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp.) as hypothetical protein, and prolyl aminopeptidase (secreted protein).	No characterized protein among the first 500 hits
510832	Alpha/Beta hydrolase	PLN02872	Triacylglycerol lipase	-----	-----	Genome Sequence and	Yeh2p <i>S. cerevisiae</i>

	fold cl26327	Abhydro_lipase (pfam04083)	Partial alpha/beta hydrolase lipase region:			Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp., <i>Colletotrichum</i> sp., etc) as hypothetical protein, triglyceride lipase-cholesterol esterase, alpha/beta- hydrolase, Sterol esterase 2.	YJM1419 (AJV63174). Query cover 58% Identity 35%
		Mhpc (COG0596)	Pimeloyl-ACP methyl ester carboxylesterase				
92423	Alpha/Beta hydrolase fold cl26327	PLN02872 Abhydro_lipase (pfam04083) MhpC (Cpg0596)	Triacylglycerol lipase Partial alpha/beta-hydrolase lipase region Pimeloyl-ACP methyl ester carboxylesterase	-----	-----	Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp., <i>Colletotrichum</i> sp., etc) as steryl ester lipase TPL1, lysosomal acid lipase/cholesteryl ester hydrolase, triglyceride lipase-cholesterol esterase, hypothetical protein,	Tgl1p <i>S.</i> <i>cerevisiae</i> YJM1083 (AJS39941) Query cover 80% Identity 43%

79895	WD40 CI25539	WD40 (cd00200) WD40 (COG2319) WD40 (smart00320) Wd40 (pfam00400) PHA03247 Atrophin-1 (pfam03154) PLN00171 Amelogenin (smart00818) PABP-1234 (TIGR01628)	WD40 domain, found in many eukaryotic proteins with a wide variety of functions. Ancestral coatmer element 1 (ACE1) of COPII with role in vesicular traffic. Atrophin-1 family domain Protein SPA1-related Cell adhesion proteins Polyadenylate binding protein:	-----	-----	alpha/beta hydrolase fold-1 Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp., <i>Colletotrichum</i> sp., etc) as hypothetical protein, Vesicle coat complex COPII, Sec31, transport protein sec31, transporter.	Web1p S. <i>cevevisiae</i> (AAA50367) Query cover 99% Identity 31%
135964	Abhydrolase family CI21494.	Esta (cog1075) Pgap1 (pfam07819)	Triacylglycerol esterase/lipase Pgap1-like protein	Active site ser - his - asp/glu. Nucleophilic attack on a carbonyl carbon atom.	-----	Genome sequence and annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>hirsutella</i> sp., <i>stachybotrys</i> sp., <i>fusarium</i> sp., <i>ophiocordyceps</i> sp., <i>colletotrichum</i> sp., etc) as hypothetical protein, triacylglycerol lipase, lipase 2, related to tgl2- triacylglycerol	No characterized protein among the first 500 hits

						lipase, gpi inositol-deacylase pgap1-like protein, pgap1-like protein lipase.	
492160	DRF_GBD Superfamily CI05720 DRF_FH3 Superfamily CI05717 FH2 Superfamily cl19758	Drf_GBD (pfam06371) Drf_FH3 (pfam06367) FH2 (pfam02181) FH2 (smart00498) PHA03307	Diaphanous GTPase-binding domain; Rho proteins, leading to activation of the Drf protein. Formin Homology 2 Domain Involved in rearrangements of the actin cytoskeleton. Transcriptional regulator ICP4-like	-----	-----	Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Tolypocladium</i> sp., <i>Pochonia</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp., <i>Colletotrichum</i> sp., etc) as cytokinesis protein sepA, Rho-GTPase effector BNI1 and related formin, involved in mating, karyogamy and meiosis, hypothetical protein, SepA/Bni1	PSI-BLAST analysis did not run
77338	Abhydrolase family cl21494.	Lipase_3 (cd00519) Lipase_3 (pfam00176 4) PLN02310 Lip2 (COG3675)	Lipases (Class 3). Lipase that can hydrolyze long-chain acyl-triglycerides into di- and monoglycerides, glycerol, and free fatty acids.	Active site Ser - His - Asp/Glu. Nucleophilic elbow on conserved domain Lipase_3: GHSLG	Active site flap/lid on conserved domain Lipase_3 (11 amino acids).	Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Tolypocladium</i> sp., <i>Pochonia</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp., <i>Colletotrichum</i> sp., etc) as lipase, hypothetical protein, triacylglycerol lipase, Mono- and diacylglycerol	Triacylglycerol lipase FGL2 <i>F. graminearum</i> (ABW74155) Query cover 89% Identity 62%

514252	Lipase (class 3). Alpha/beta hydrolase cd00519	Lipase_3 (pfam01764) Lipase_3 (cd00519) CVT17 (COG5153) PRK11071	Lipase that can hydrolyze long-chain acyl-triglycerides into di- and monoglycerides, glycerol, and free fatty acids. Putative lipase essential for disintegration of autophagic bodies inside the vacuole Esterase YqiA	Active site Ser - His - Asp/Glu. Nucleophilic elbow on conserved domain Lipase_3: GHSLG	Active site flap/lid on conserved domain Lipase_3 (11 amino acids).	lipase, Putative feruloyl esterase A, Alpha/Beta hydrolase protein. Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Tolypocladium</i> sp., <i>Pochonia</i> sp., <i>Metharhizium</i> sp., <i>Fusarium</i> sp., <i>Colletotrichum</i> sp., etc) as autophagy related lipase Atg15, triacylglycerol lipase, hypothetical protein, alpha/beta-hydrolase, related to starvation induced protein PSI-7, autophagy lipase.	Hypothetical protein FGSG_02519 <i>F. graminearum</i> PH-1 (XP_011318453) Query cover 92% Identity 70%
502433	Predicted triacylglycerol lipase activity, Lipase_3.	Lipase 3 (cd00519) Lipase_3 (pfam01764) PLN02847 AF-4 (pfam05110)	Lipase that can hydrolyze long-chain acyl-triglycerides into di- and monoglycerides, glycerol, and free fatty acids. AF-4 Proto-oncoprotein; Nuclear proteins linked to human disease	Active site Ser - His - Asp/Glu. Nucleophilic elbow on conserved domain Lipase_3: GHSLG	Active site flap/lid on conserved domain Lipase_3 (11 amino acids).	Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Tolypocladium</i> sp., <i>Pochonia</i> sp., <i>Metharhizium</i> sp., <i>Ophiocordyceps</i> sp., <i>Fusarium</i> sp., <i>Colletotrichum</i> sp., etc) as esterase/lipase, hypothetical protein, Sn1-	No characterized protein among the first 500 hits

78181	Predicted triacylglycerol lipase activity, Lipase_3. Lipid transport and metabolism.	Lipase_3 (cd00519) Lipase_3 (pfam01764) PLN00413 Lip2	Lipase that can hydrolyze long-chain acyl-triglycerides into di- and monoglycerides, glycerol, and free fatty acids.	Active site Ser - His - Asp/Glu. Nucleophilic elbow on conserved domain Lipase_3: GHSLG	Active site flap/lid on conserved domain Lipase_3 (11 amino acids).	specific diacylglycerol lipase beta, Lipase, Lipase, class 3, Genome Sequence and Annotation of different fungi (e.g. <i>Trichoderma</i> sp., <i>Purpureocillium</i> sp., <i>Fusarium</i> sp., <i>Tolypocladium</i> sp., etc) as extracellular lipase-like protein, hypothetical protein, Lipase, probable triacylglycerol lipase precursor, Chain A, Crystal Structure Of Lipase	Lipase F. <i>heterosporum</i> (AAB34680)) Query cover 98% Identity 52%
-------	---	--	--	---	--	--	---

*Annotation is summary of information retrieved from Gene ontology, Conserved Domain tool at NCBI, Pfam, and Superfamily.

Table 4.3.2 *In silico* analysis of putative true lipases of *T. harzianum*.

Protein id	Aa residues	Molecular weight (kDa*)	SignalP (secreted)	TMHMM Domains	Putative cellular localization	Continue for next analysis
551811	379	41.185	No	No	Cytoplasmic	Yes
87496	430	46.62	Yes	No	Extracellular	Yes
526309	452	47.99	Yes	No	Extracellular	Yes
514427	454	49.25	Yes	No	Extracellular	Yes
510832	722	82.80	No	No	Cytoplasmic	No
92423	552	62.361	No	1	Ambiguous location (plasmatic, endoplasmic reticulum, Golgi Apparatus)	Yes
79895	1254	134.85	No	No	Mitochondria	No
135964	339	37.51	No	No	Endoplasmic reticulum	no
492160	1782	198.07	No	No	Nuclear	No
77338	404	44.5	Yes	1	Extracellular	Yes
514252	613	65.4	Yes	No	Extracellular	Yes
502433	1059	115.6	No	No	Cytosolic/ nuclear	Yes
78181	340	36.2	Yes	No	Extracellular	Yes

*Kilodalton (kDa).

The analysis in the CDD identified the alpha/beta hydrolase fold on the putative TAG lipases 510832, 92423, and 135964. Top hits at BLASTp retrieved sterol esterases for 510832, and GPI inositol-deacylase PGAP1-like protein for 135964, the last one probably involved in vesicular traffic from endoplasmatic reticulum.

92423 shares homology with lysosomal acid and other acid lipases. Consistent with intracellular roles, any of these three lipases are predicted extracellular (Table 4.3.2). Regarding the other two putative TAG lipases, 79895 and 492160, neither lipase nor

hydrolase domains were identified, but WD40 and Sec domains in the former, and Diaphanous FH2/FH3 domains in the later (Table 4.3.1), suggesting roles with other functions than lipases. Deduced sizes are 134 and 198 kDa respectively; it is not prognosticated secretion for these proteins and in agreement, signal peptide was not identified.

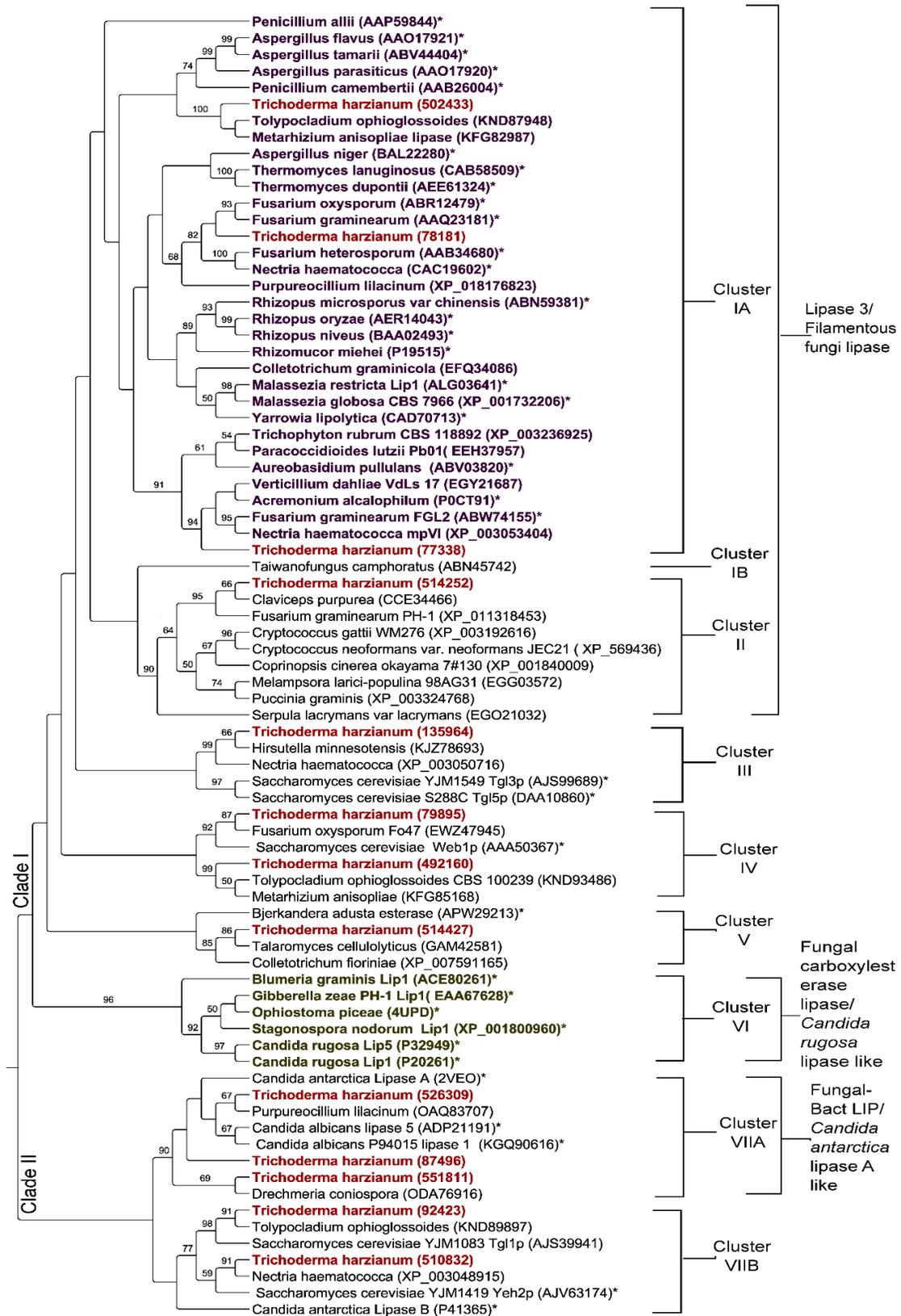
The four automatically annotated secreted lipases (514427, 526309, 551811, and 87496), show Abhydrolase fcl21494 domain, containing the catalytic triad Ser – His – Asp, and contain the Pfam domain 03583 which is related with a lipase from *C. albicans*. Curiously, one of those proteins, 551811, gave negative prediction of signal peptide with SignalP, and WoLF PSORT predicts cytoplasmic localization. A Rossmann-fold NAD(P)H/NAD(P)(+) binding (NADB) domain was identified in 551811, a domain found in numerous redox enzymes. 514427 shares homology with the acetyl xylan esterase (AXE1) domain, related with acetyl xylan esterase. Signal P analysis supports prediction of secretion for protein: 514427, 526309, and 87496, and in congruence WoLF PSORT deduced that they are extracellular proteins (Table 4.3.2).

Rapid analyses and deduction of intracellular location and large size can exclude some of these candidates, but we decided to perform further bioinformatics analyses on all these proteins in order to achieve the most effective selection of final candidates. In accordance with the search criteria, the order for extracellular TAG lipase in *T. harzianum* is 77338, 78181, 514427, 526309, 87496, 514252, 510832, 92423, 135964, 502433, 551811, 79895 and 492160.

4.3.3 Phylogenetic Analysis

In order to find a relationship with other fungal lipases, a phylogenetic tree was constructed. We used 66 lipase sequences, 35 of which have been characterized, plus the 13 sequences of the candidates. The characterized fungal (Ascomycete and Basidiomycete) lipases/esterases were included as guides for functions. As far as we know, this is the first genome survey to retrieve true lipases in *T. harzianum*, and the first phylogeny analysis including its lipases. Seven clusters were defined (Figure 4.3.1). Clusters 1A, 1B and II are largely consistent with the topology of the phylogenetic tree published by Yadav *et al.* (2012) where filamentous Ascomycete (cluster I) and Basidiomycete (cluster II) are apart (Figure 4.3.1), but differs in the branching structure

found by Feng *et al.* (Feng *et al.*, 2011) and Barriuso and Martínez (Barriuso y Martínez, 2017) for Lip1 type lipases and Basidiomycete lipases. The reports presented above placed Lip1-type and Basidiomycete lipases in sister clusters, and here Lip1 type lipases placed in the cluster of *C. rugosa* type lipases (cluster VI), apart from Basidiomycetes. We obtained the same branching structure as those authors when a few amino acid sequences were aligned (data not shown); many phylogenetic trees of fungal lipases have been constructed focusing on a narrow group of lipases (Park *et al.*, 2015; Feng *et al.*, 2011, Feng *et al.*, 2005), which forces unrelated proteins to group together validates it and supports the phylogeny of the candidate proteins.



0.0 0.5

Figure 4.3.1 Phylogenetic tree of fungal lipases. The tree was constructed with 35 characterized fungal lipases (accessions correspond to GenBank unless another source is specified) and 15 putative triacylglycerol lipases from *T. harzianum*. (*) After the accession numbers are lipases which have been characterized; unlabeled proteins correspond to hypothetical, uncharacterized, predicted lipases. The tree was generated by MAFFT software using the neighbor-joining method (Saitou *et al.*, 1987) with 500 bootstrap re-samplings. Clusters IA, IB and II as described by Yadav *et al.* (2012). Highlighted in purple letters, the *Rhizomucor miehei* lipase-like group and in olive green letters, the *Candida rugosa* lipase-like group, according Schmidt-Dannert (1999). Clade I and Clade II, are consistent with Gupta *et al.* (2015). *T. harzianum* triacylglycerol lipases from this study, highlighted in bold red letters.

Our result is congruent with early phylogeny proposed by Schmidt-Dannert (Schmidt-Dannert, 1999) by placing the “*Rhizomucor miehei* lipase family” clearly apart (highlighted in the phylogenetic tree in purple letters) and the “*Candida rugosa* lipase family” (highlighted in the phylogenetic tree in olive green letters, and corresponds to cluster VI; see Figure 4.3.1).

Cluster III comprises Tgl lipases and homologues; cluster IV is a branch comprising uncharacterized proteins, probably related with intracellular protein traffic; cluster V contains *B. adusta* and homologues. Cluster VII has two sub clusters where VIIA comprises *C. antarctica* lipase A and homologues.

The topology of the phylogenetic tree reconstructed here is largely congruent with the phylogenetic tree published recently by Gupta *et al.* (2015). These authors carried out a comprehensive analysis on lipases available in the LED data base and then conducted the phylogenetic analysis of 61 representative lipases with different combinations of oxyanion and pentapeptide sequences in fungal and yeast lipases; their tree defined two clades. From our phylogenetic tree, the clusters I, III, V and VI belong to clade I and the cluster VII belongs to clade II. Clusters II and IV were not represented in their phylogenetic tree. Predicted proteins in cluster IV share low homology with known lipases and no member in this cluster has been characterized to date but reasons for exclusion of Basidiomycetes (cluster II) are unclear. Other differences between both phylogenetic trees is that they included 14 lipase sequences from *Y. lipolytica* (most of them placed in the clade I) and 10 lipase sequences from *C. albicans* (most of them placed in the clade II, resulting in a larger clade II in their phylogenetic tree, in comparison with the cluster II in the phylogenetic tree presented here). The most important difference is that class 3 lipases comprise a single cluster in our phylogenetic tree but these lipases split between clade I and clade II in

theirs. This difference is because all *T. harzianum* class 3 lipases and their homologues, are lipases with the pentapeptide GHSLG, which is the most frequent in this class of lipases. However, class 3 lipases can contain the pentapeptides GTSAG, GHSFG and GHSYG (Zan *et al.*, 2016). Gupta *et al.* (2015) included class 3 lipases with different pentapeptides. So, the differences between both phylogenetic trees are apparent but they are actually quite consistent. The consistency in the topology of our phylogenetic tree with, Yadav *et al.* (2012), Schmidt-Dannert (1999) and Gupta *et al.* (2015) validates it and supports the phylogeny of the candidate proteins

The four lipases class 3 place in different branches in the large cluster I of “Filamentous fungi lipase” (Figure 4.3.1). Class 3 family comprises true lipases (Marchler-Bauer *et al.*, 2017, 2015) and they are characterized as having an active site flap/lid and work on long-chain acyl-triglycerides. In consequence, these four class 3 lipases became strong candidates. The protein 78181 placed in the same cluster as *F. graminearum* FGL1 AAQ23181, *F. heterosporum* AAB34680 and *T. laniginosus* CAB58509.

The protein 77338 clustered with P0CT91, LipA of *Acremonium alcalophilum*, a protein which is a lipase/acetylxylylan esterase activity that works on long-chain pNP esters and xylans (Pereira *et al.*, 2013). The protein 502433 grouped with uncharacterized lipases, in a sister clade that grouped lipases involved in aflatoxin B-producing in *Aspergillus flavus*, *Aspergillus tamari* and *Aspergillus parasiticum*. 502433 is probably related with the production of toxic metabolites against *T. harzianum* antagonists.

T. harzianum 514252 grouped in the cluster II with lipases from *C. purpurea* and *F. graminearum*, in a sister clade to lipases from Basidiomycetes (*Cryptococcus gatti*, *Cryptococcus neoformans*, *Coprinus cinereus*, *Melampsora larici-populina* and *Puccinia graminis*); however functional information regarding these lipases has not been elucidated.

Three of the four proteins automatically annotated as “secreted lipases” in the *T. harzianum* portal (526309, 551811 and 87496), clustered with the *C. antarctica* Lipase A (CALA), a TAG lipase (cluster VII). The fourth predicted secreted lipase (514427) grouped with *B. adusta* (cluster V) fungal esterase APW29213, which has no activity on long-chain glycerides (Sánchez-Carbente *et al.*, 2017).

The protein 135964 grouped with Tgl3p and Tgl5p (cluster III) and 92423 with Tgl1 (cluster VIIB), the major TAG lipases of the yeast *S. cerevisiae* (Klein *et al.*, 2016). 510832 also placed close to Tgl1.

The protein 492160 and 79895 grouped with *F. oxysporum* EWZ47945, *S. cerevisiae* Web1p AAA50367, *T. ophioglossoides* KND93486 and *M. anisopliae* KFG85168 in cluster IV. All of which are uncharacterized proteins originating from genomic sequencing projects and none predicted as lipase

4.3.4 Structural Modeling

Currently there are a number of modeling programs available and practically all have pros and cons. In this work Swiss model, HHpred and I-TASSER were simultaneously used for analyzing the lipase candidates.

For most of the putative lipases under analysis, the protein database (PDB) accession for the best model coincided for two of the three types of modeler software and the third software retrieved a PDB accession which corresponded to another version of the same, or similar protein. For example, the proteins 526309, 87496, 551811 and 514427 hit PDB-3GUU in Swiss model, corresponding to *C. antarctica* Lipase A and in HHpred server retrieved top hit PDB-2VEO, which is the crystal structure of *C. antarctica* lipase A in its closed state. BLAST searching in specialized α/β -hydrolases/lipase-esterase databases was consistent to support homology-based and structure-based transfer annotation for most of the *T. harzianum* lipases revised here. The summary is presented in Table 4.3.3.

Table 4.3.3 *In silico* localization of characteristic lipase domains in the candidate lipase proteins from *T. harzianum*.

Protein ^a	Template	Reference art	Pentapeptide	Lid	Catalytic triad	Oxyanion	
551811	3guu	Ericsson <i>et al.</i> (2008)	GYSGG	Asn195-	Ser162,	Asp92	
			Gly160-Gly164	Ser287	Asp309,	Gly163	
						His341	
87496			GYSGG	Leu223-	Ser190,	Asp96	
			Gly188-Gly192	Glu313	Asp334,	Gly191	
				His368			
526309			GYSGG	Asn243-	Ser210,	Asp116	
			Gy208-Gly212	Asp332	Asp352,	Gly211	
					His384		
514427			GHSQG	Ala258-	Ser228,	Ile146	
			Gly226-Gly230	Phe349	Asp373,	Gln229	
					His405		
510832	1k8q	Roussel <i>et al.</i>	CHSQG	Phe496-	Ser433,	Leu349	
			Cys431-Gly435	Glu550	Asp634,	Gln434	

92423	(2002) and Selvan <i>et al.</i> (2010)	GFSQG Gly224-Gly228	Ile286- Ile321	His667 Ser226, Asp396, His422	Leu139 Gln227
79895 [#]	No structural homologue of lipase was identified				
135964	4x6u Dror <i>et al.</i> (2015)	AHSMG Ala-152- Gly156	Phe189- Pro200	Ser154, Asp275, His297	Leu70 Met155
492160 [#]	No structural homologue of lipase was identified				
77338	3o0d Bordes <i>et al.</i> (2010)	GHSLG Gly214-Gly218	Thr137- Tyr154	Ser216, Asp282, His343	Thr137 Leu217
514252		GHSLG Gly316-Gly320	Thr230- Trp262	Ser318, Asp379, His457	Thr230 Leu319
502433*	Complete protein sequence does not model with lipase				
78181	3ngm Lou <i>et al.</i> (2010)	GHSLG Gly174-Gly178	Asn115- Phe126	Ser176, Asp230, His289	Ser114 Leu177

*Lipase domains restricted to a short fragment of the protein. #No modeling with lipases.
&Three-dimensional models are available as Figure S3.1.

General characteristics in lipases comprise α/β hydrolase fold, a catalytic triad (Ser-Asp/Glu-His) and the pentapeptide motif GX SXG (Casas-Godoy *et al.*, 2012). The catalytic triad found here was Ser-Asp-His. Most of the predicted proteins under revision here have the pentapeptide GX SXG but two other versions were identified, AHSMG and CHSQQ. The pentapeptide AX SXG is in the lipase family 1.4 found in many bacteria and it has been found in two yeasts, *Trichospora asahii* lipase TALipA, with activity on p-NPC-18 (Kumari y Gupta, 2015) and *S. cerevisiae* TGL2, which prefers short chain-TAG but retains 20% of relative activity on long chain C18:1 TAG (Ham *et al.*, 2010). CHSQQ pentapeptide has not been identified in any characterized lipase so far.

Three-dimensional superposition with IPBA software of each *T. harzianum* lipase model with its respective best template model permitted the identification of putative lid domain, catalytic triad, the pentapeptide sequence and the oxyanion hole in most of the candidate proteins. Specific position of each domain and motif in each *T. harzianum* TAG lipase is described in Table 4.3.3.

Identity of lipase was not supported for two protein candidates: models for protein 79895 in Swiss model and HHpred were SEC3-type transport proteins, involved in the structure of the COPII coat transport-vesicles on membranes. I-TASSER identified an RNA-dependent

RNA polymerase of Cypovirus. The search for homologues in ESTHER and LED data bases retrieved no hits.

The other unsupported candidate was 492160. Swiss model resulted in hits with low coverage (288 to 743 amino acids) over this large protein (which is 1782 amino acids). The first hit was PDB-1Y64 (30.38% identity), which corresponds to Bni1p formin from *S. cerevisiae*, complexed with ATP-actin. This is a calcium-dependent and ATP-dependent protein binding, involved in actin filament bundle assembly and cytoskeleton remodeling. HHpred showed homology with different proteins at N- and C-ends. At N end, the first hit was 3EG5, a Rho-Diaphanus-binding protein involved in cytoskeleton assembly and cell division. At C-end the first hit was PDB-2J1D, actin binding for actin assembly. No hit was retrieved in any of the modeler servers belonging to lipase or alpha-beta hydrolase families. I-TASSER was also unable to model 492160 with a non-lipase protein. BLASTp in ESTHER and LED data bases retrieved no homologues for this candidate.

4.3.5 Functional Prediction of Candidates

Different strategies have been used to predict function. These strategies comprise sequence *homology* (e.g., BLAST), phylogenetic relationship, structure homology (i.e., detection of classic domains and motifs) and three-dimensional homology. Conducting a search with any one of them results in satisfactory functional prediction for proteins with large conservation with characterized proteins; however, more difficulties arise when proteins show low identity with known proteins, as in the case of lipases. To overcome this problem, function prediction was based on the sum of those function-transfer strategies.

The proteins 77338, 78181, are related with extracellular TAG lipases. 77338 modeled with PDB-3O0d (Table 4.3.3), an olive oil- induced lipase from the non-pathogenic yeast *Y. lipolitica* (Najjar *et al.*, 2011), with barely 35% identity and it placed closer to LipA (Figure 4.3.1), a lipase/acetylxylylan esterase from *A. alcalophilum* which works on long-chain pNP esters and xylans (Pereira *et al.*, 2013).

78181 is close to lipases involved in pathogenicity such as CAC19602 from *N. haematococca* (Eddine *et al.*, 2001), *F. heterosporum* AAB34680 (Nagao *et al.*, 1994) and AAQ23181 (FGL1) from *F. graminearium* (Voigt *et al.*, 2005) which are able to work on vegetable oils, including olive oil. The latter being identified as the best 3D model for 78181 (Table 4.3.3), and FGL1 expression is induced by olive oil (Voigt *et al.*, 2005). 77338 and 78181 therefore are plausible candidates for what is being sought.

Information for 514252 is not consistent. Fragments of this protein share homology with lipases class 3 and carboxyl esterases. 514252 has fragments of sequences with homology with domains of lipases involved in the disintegration of autophagic bodies inside the vacuole at starvation (Table S3.2 and the closest homologue (75% identity) is XP_011318453 and autophagy related lipase from *F. graminearum* (Nguyen *et al.*, 2011), reinforcing an intracellular role. However, the protein 514252 is predicted extracellular, which is not congruent with the expected location of this protein in the cell.

502433 is a large protein (115 kDa); it shares homology with class 3 lipases, it is predicted intracellular (Table 4.3.2) and probably function in relation to the binding of heat-shock proteins (Table S3.2). This kind of putative lipases/calmodulin binding heat shock proteins are predicted in few genomic data but so far, they have not been characterized at all.

The proteins 514427, 526309, 87496 and 551811 share homologies with CALA, a lipase with interfacial activation. CALA is a secreted lipase from *C. antarctica* that catalyzes hydrolysis on long-chain TAGs such as triolein and olive oil (Coradi *et al.*, 2013). For 551811 Signal P and WoLF PSORT give results that do not support its secretion (Table 4.3.2), decreasing its eligibility. The other three are predicted extracellular, as expected for the activity in the fungal B13-1 strain. Gene Ontology relates 514427 with iron ion transport and searching by PSI-BLAST found homology of this protein with acetyl xylan esterases.

92423 and 510832 are probably related to each other. Both share homology with acidic lipases (Table 4.3.3), with cholesterol esterases and pimeloyl-ACP methyl ester carboxylesterases (Table 4.3.1). Congruent with these probable functions, both are predicted intracellular (Table 4.3.2).

135964 has a small size but seems to be related with intracellular protein transport from endoplasmic reticulum (Tables S4.2, 4.3.1 and 4.3.2).

Analyses of 79895 and 492160 do not support their association with lipases; only small fragments of their sequences share homology with lipases. The former is related with proteins involved in the traffic of proteins from endoplasmic reticulum to the Golgi apparatus, and the latter with GTP-binding and organization of actin in cytoskeleton (Table 4.3.1). Congruent with their putative roles, both of them are predicted intracellular (Table 4.3.2).

4.3.6 Refining of Selection of Candidates

Based on sequence analysis (protein size and prediction of cellular location), structure analysis (identification of domain and motifs), phylogenetic analysis (clustering with known true lipases or known non-true lipases), 3D-modeling and function prediction, the list of initial 13 candidates was reduced to a final list of candidates. Since the quest is for the gene/protein responsible for *T. harzianum*'s extracellular olive oil-lipolytic activity and as we hypothesized that the responsible lipase(s) is or are true lipases, the criteria for selecting the final list of candidates are: small size (<80 kDa) (Mehta *et al.*, 2016; Sharma *et al.*, 2011), predicted secreted, preferably with lid domain (which means interfacial activation, common in true lipases), preferably with homology with lipases active on long-chain acyl-triglycerides, including vegetable oils, such as olive oil. The candidates that meet these criteria are five: 77338, 78181, 514427, 526309 and 87496. It was decided to include 514252 in the list, since it shows domains of lipase class 3 (this class has interfacial activation) and it is predicted extracellular, although its function could not be proposed. However, sufficient information is not yet available to support its exclusion.

4.3.7 RT-PCR Analysis for Validation of Predicting Results

All primer pairs, except the primer for 87496, specifically amplified the expected PCR product (Figure 4.3.2A). We were unable to screen the expression of 87496 because the primers failed to amplify even on DNA template, despite multiple attempts.

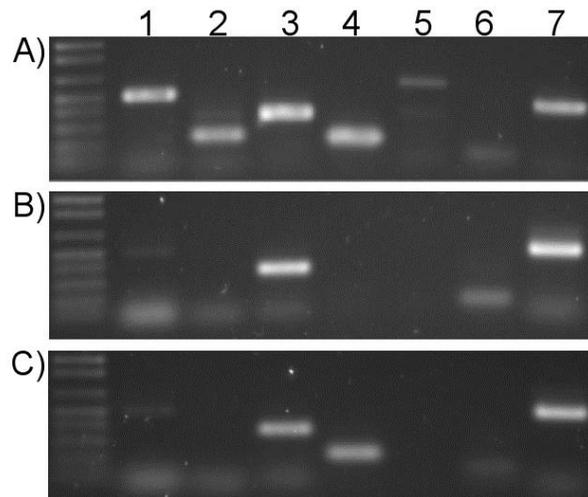


Figure 4.3.2 RT-PCR analysis of selected putative extracellular TAG lipases from *T. harzianum* in medium without (**B**) or with 1% (v/v) olive oil (**C**) as carbon source. The number of ID at genome portal of each candidate lipase corresponds to: (1) 77338; (2) 78181; (3) 514252; (4) 526309; (5) 514427; and (6) 87496. Lane (7) Elongation factor 1, as positive control of PCR. Panel (**A**) corresponds to PCR on genomic DNA, to test the primers

RT-PCR analysis evidenced the olive oil-responding expression of 526309 in *T. harzianum* B13-1 strain (Figure 4.3.2C). Lipases 77338 and 514252 were also expressed, but only 526309 was expressed de novo (Figures 3B and 3C). Thus, the RT-PCR demonstrates that the bioinformatic analyses performed here worked successfully to select candidates to explain the lipolytic activity in the strain B13-1. This result gives support to the possibility that the gene `estExt_fgenesh1_pg.C_1_t10183`, located at scaffold 1:494922-496430 (-) and which codifies the protein 526309, is responsible for the extracellular lipolytic activity of the *T. harzianum* B13-1 strain on olive oil. The closest homologue for 526309 is CALA, with 36% identity. This low conservation between 526309 and CALA does not discard functional relationship. Widmann *et al.* (2010) reported 32 lipases from different sources phylogenetically related with CALA and they share a similar tridimensional structure with this protein but low identity at sequence level (as low as 15%), as found for homologues in *T. harzianum*. Conservation is usually low in lipases (Fischer y Pleiss 2003; Jaeger *et al.*, 1994). Figure 4.3.3 shows the multi-alignment of 526309 and the other two predicted secreted lipases in *T. harzianum* (this study), as well as CALA protein. They share

conservation at catalytic triad and pentapeptide sequences, while lids are composed of amino acids with similar characteristics.

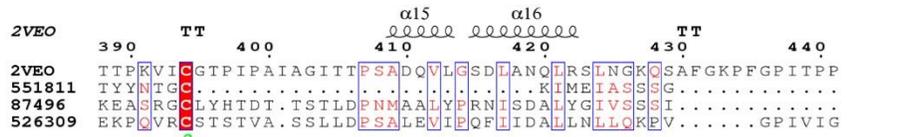
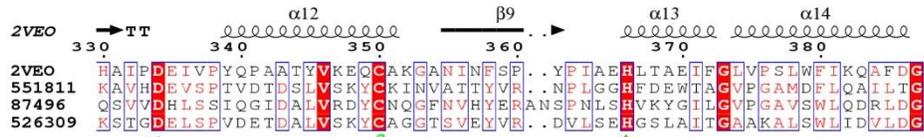
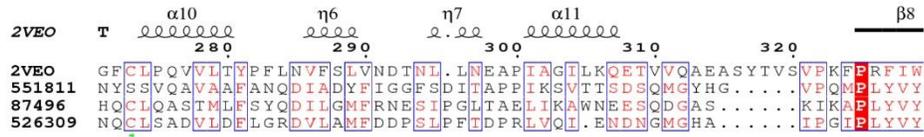
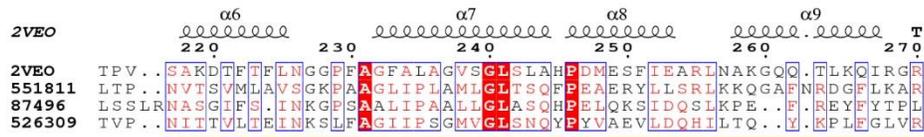
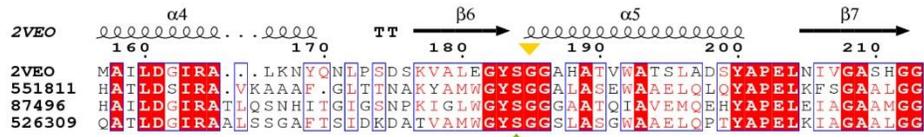
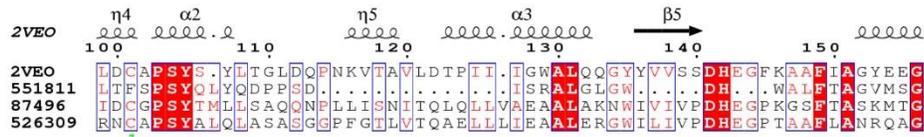
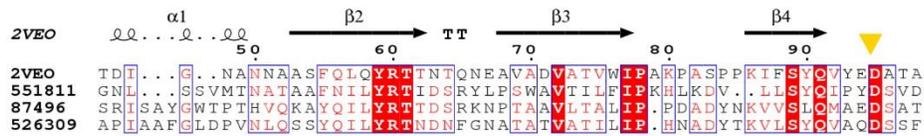
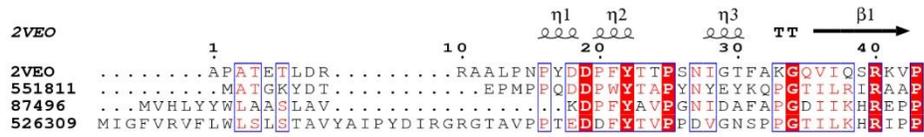


Figure 4.3.3 Structural-based multi-alignment of *T. harzianum* 526309, 551811, and 87496 with 2VEO lipase from *C. antarctica* (CALA). Blue line highlights the pentapeptide; orange line, the lid domain; green triangles, the catalytic triad; and yellow triangles, the oxyanion. Alpha helices and beta sheets are indicated at the top. Identical (bold white letters on red background) and similar amino acids (standard red letters) are shown.

T. harzianum has different ecological niches, i.e., plant endophyte, antagonist of a wide range of plant pathogens and it is also a cosmopolitan soil-borne fungus frequently found on decaying wood. Different TAG lipases can be involved in the different ecological niches of the fungus. In the case of the lipase 526309, which is induced by olive oil, it can play role in lipid-rich environments.

Interestingly, 526309 does not belong to class 3 of the lipases and is not even annotated as TAG lipase in the genome database of *T. harzianum* but Brennes and Baeck (Brenneis y Baeck, 2012) showed that CALA, its homologue, works on long-chain TAGs. Therefore, transfer function based on structure-homology predicts triacylglycerol lipolytic activity for 526309.

It is important to clarify that class 3 lipases are true lipases but not all true lipases belong to class 3 (Messaoudi *et al.*, 2011). The most abundant amino acids in 526309 are Alanine (10.8%), Leucine (11.5%), Glycine (8.4%) and Serine (7.5%), congruent with those found in true lipases by Messaoudi *et al.* (2011). Three-dimensional modeling identified on 526309 the lid domain from Asn243 to Asp332 (Table 4.3.3), which is compatible with the criteria we proposed in our search.

The survey emphasized proteins annotated in the genome database of *T. harzianum* as TAG lipases but was not restricted to them and that was a success which permitted the identification of this lipase. Cloning of the full-length cDNA of lipase 526309 is in progress in our laboratory for heterologous expression and further characterization of this protein, to challenge this proposal.

This is the first report on the identification and *in silico* characterization of TAG lipases from *T. harzianum* and expands the catalogue of potential enzymes available for industrial applications. The combined strategy of functional screening of isolated microorganisms

and bioinformatic analyses, if the genome sequence is available, can be applied to identify lipases in other microorganisms, or to identify other particular enzymes among a large family of protein candidates. To date, there are a large number of sequenced genomes of fungi and bacteria.

4.4 CONCLUSIONS

In this work, we reported one *T. harzianum* strain as the best producer of extracellular lipolytic activity among the fungal strains analyzed here. A search in the deduce proteome at *T. harzianum* genomic portal retrieved 50 lipases. Bioinformatics survey managed to reduce the list to seven candidate proteins and RT-PCR analysis identified de novo expression of one of those candidates when the fungus grew on olive oil-containing medium, supporting the suitability of the procedure followed here. 526309 shares homology with CALA lipase and has potential for further application in industry.

Supplementary Materials:

The following are available online at <http://www.mdpi.com/2073-4425/9/2/62/s1>.

Table S3.1 List of primers used in this study.

Table S3.2 List of all lipase proteins identified in the *T. harzianum* genome homepage

Figure S3.1 Three-dimensional model proposed for (A) 551811; (B) 87496; (C) 526309; (D) 514427; (E) 510832; (F) 92423; (G) 135964; (H) 77338; (I) 514252; (J) 78181. In all protein models: (1) Red, α -helices; yellow, β -strands. (2) Close up showing the lid (yellow), the catalytic triad (in red) and the oxyanion (in blue). (3) Superposition with best template protein model (see Table 4.3.3). (4) Superposition of catalytic triad in both proteins. All models were generated by I-TASSER and visualizations were performed in PyMOL.

Acknowledgments: MACP acknowledges a scholarship granted by CONACyT-Mexico (242995). This work was partially supported by the CONACYT Grants 220957 and 269833, and FOMIX 247355.

Authors thank Miguel Tzec-Sima for formatting the figures.

Author Contributions: Miguel Canseco and Blondy Canto conceived, designed and wrote the paper; M. Canseco performed the experiments and analyzed the data; Marcela

Gamboa contributed with support in fungal isolation and manipulation; Ignacio Islas and Gerardo Rivera contributed by assisting M. Canseco in lipolytic screening and measurement of the lipolytic activity; Bartolome Chi contributed with the gene expression analysis; Max Apolinar contributes with support in phylogenetic analysis and the 3D-modeling; B. Canto supervised all the work. All authors participated in the data analysis and writing of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

CAPÍTULO V

5.1 DISCUSIÓN GENERAL

Cada vez es más deseable el uso de herramientas biotecnológicas amigables. En este sentido, el uso de enzimas ofrece una alternativa para resolver de forma ecosostenible las necesidades industriales y ambientales de la humanidad. Sin embargo, el costo en la producción y purificación de las enzimas es uno de los mayores problemas (Robinson, 2015), por lo que su empleo se limita a la preparación de detergentes, preparación de alimentos, generación de fármacos y otras aplicaciones cuya rentabilidad compensa el costo del catalizador biológico. Sus usos sin embargo son más amplios, por lo que es necesario buscar alternativas que permitan a la larga disminuir los costos de producción.

Una forma barata de producción enzimática es la fermentación en fase sólida, y ésta ha sido explotada sobre todo para producir enzimas para bio-remediación. En este caso se han usado técnicas de bio-aumentación y bio-estimulación para la multiplicación *in situ* de microorganismos deseables, usando como sustrato rastrojos y residuos vegetales. Esta estrategia ha sido empleada en países asiáticos y ha sido ampliamente recomendada para países en vías de desarrollo debido a su bajo costo. Sin embargo, no es aplicable para muchas aplicaciones biotecnológicas modernas que requieren el uso de enzimas puras como el procesamiento de alimentos y bio-transformación de fármacos. Un problema común tanto en los extractos enzimáticos como en las enzimas puras es la rápida inactivación enzimática. Para alargar la vida útil de las enzimas éstas suelen ser inmovilizadas en polímeros como alginato (Zhang *et al.*, 2014; Shen *et al.*, 2011). La última generación de enzimas inmovilizadas utiliza la expresión superficial en organismos vivos. Esta tiene la ventaja de la protección enzimática por la inmovilización, y también la ventaja de que el organismo se puede multiplicar, lo que facilita la producción (Yuzbasheva *et al.*, 2011). La primera expresión superficial se realizó en bacteriófagos, en donde la proteína de interés fue expresada fusionada con la proteína de la cubierta del fago, pero su uso se ha limitado a expresar anticuerpos. Actualmente la mayoría de los trabajos de expresión superficial se han llevado a cabo en levaduras (Moura *et al.*, 2015; Zhang *et al.*, 2013; Yuzbasheva *et al.*, 2011; Jiang *et al.*, 2008; Kim *et al.*, 2002; Matsumoto *et al.*, 2002;).

Para la expresión superficial es necesario que la proteína se procese como proteína dirigida al exterior (con péptido de secreción), pero que posea un motivo proteico que lo obligue a permanecer en la pared celular. Los hongos filamentosos tienen alta capacidad secretora, que alcanzan concentraciones de gramos por litro de proteína en el medio, por lo que son una buena alternativa para expresar superficialmente enzimas (Adachi *et al.*, 2008; Mach y Zeilinger, 2003). Adachi y colaboradores (2008) fueron los primeros en reportar la expresión superficial de proteínas en un hongo filamentoso. Para ello buscaron en el genoma de *Aspergillus oryzae* proteínas con homología con proteínas con anclaje GPI descritas en *S. cerevisiae* y fusionándolas a GFP realizaron análisis funcional de 5 candidatas; en 2 de ellas observaron localización superficial de la fluorescencia. La fusión de las GPIs que localizaron a la GFP en la pared se fusionaron con la β -glucosidasa (BGLI) y observaron actividades variables con los diferentes constructos, lo que evidencia que es necesario evaluar varias proteínas para elegir la más apropiada.

En el presente trabajo se propuso expresar superficialmente enzimas en un hongo filamentoso del género *Trichoderma*. Éste género ha sido ampliamente reportado como un hongo con alta capacidad secretora (Zhang *et al.*, 2014; Mukherjee *et al.*, 2013). Como enzima para prueba de concepto se consideró usar una triacilglicerol lipasa, dado que éstas son importantes en reacciones de transesterificación, por ejemplo para producir biodiesel, o de alguna lacasa para usar el hongo en bio-remediación. Se intentó originalmente conseguir alguno de los vectores reportados (Ronglin He *et al.* 2013; Adachi *et al.* 2008; MAOR *et al.* 1998) pero ninguno de los autores proporcionó su vector. En consecuencia se consideró construir un vector propio, lo que además daría la posibilidad de establecer un desarrollo tecnológico que pudiera someterse a protección intelectual.

La búsqueda de anclajes propios se basó en la búsqueda de proteínas GPI de manera global en el proteoma deducido de *T. reesei* debido a que inicialmente no se tenía la identidad de las especies de *Trichoderma* con las que se contaba en el grupo, y se seleccionó esa especie por ser hipersecretora. Se identificaron todas las potenciales proteínas GPI tanto usando softwares de acceso libre, especializados (GPI: Big-PI fungal predictor (http://mendel.imp.ac.at/gpi/fungi_server.html), GPI-SOM predictor (<http://gpi.unibe.ch/>) y predGPI predictor (<http://gpcr.biocomp.unibo.it/predgpi/>), como mediante búsqueda de proteínas homólogas a proteínas caracterizadas con anclaje GPI. Estos softwares usan algoritmos similares, enfocados a la identificación de péptido señal en el extremo amino y el dominio ω en el extremo carboxilo. Además de esos criterios,

que han sido usados por muchos autores (Zhang *et al.*, 2013; da Silva Castro *et al.*, 2005; Kim *et al.*, 2002; Hamada *et al.*, 1998) incluyendo nuestro grupo (Kantún-Moreno *et al.*, 2013), se reforzó la predicción extracelular mediante el análisis de la ubicación de la proteína en la célula con el programa WolfPSort (<http://wolfpsort.seq.cbrc.jp/>). También se analizó el contenido de serina/treonina, clasificando las proteínas GPIs como >30% S/T y ≤30% S/T, ya que de acuerdo a Eisenhaber y colaboradores (2004) las primeras corresponderían a proteínas localizadas en la pared celular. Asimismo se analizó el punto isoeléctrico (pI) con base en que Pittet y Conzelmann (2007) observaron que las proteínas GPI localizadas en pared celular presentan pI de 4.865 ± 0.2194 mientras que las que son retenidas en membrana plasmática poseen pI de 6.667 ± 0.9454 . Por último, Caro y colaboradores (1997) y Kliss, y colaboradores (2009) reportaron la identificación de un grupo de dos aminoácidos básicos (grupo dibásico) en la región C-terminal, cerca del sitio ω que favorece la ubicación en membrana plasmática y que en las proteínas que llegan a pared celular está ausente. Todos estos criterios robustecen la predicción de que una proteína con anclaje GPI esté ubicada en pared celular. Cabe mencionar que el objetivo del trabajo no es identificar todas las proteínas GPIs de *T. reesei*; la astringencia de los criterios puede dejar fuera muchas proteínas GPIs de pared celular, pero el objetivo es seleccionar proteínas que tengan la mayor probabilidad posible de anclarse en la pared celular para lograr la inmovilización superficial de las proteínas que se fusionen con estas etiquetas (“tags”).

Los resultados de identificación *in silico* del presente trabajo son similares en cuanto a las características de las proteínas reportadas en otros trabajos como el de Cheon y colaboradores (2014), Zhang y colaboradores (2013) y de Adachi y colaboradores (2008) para expresión superficial (en pared celular) de hongos filamentosos (o levadura). Es decir, las mejores proteínas (Ccw14 para *Hansenula polymorpha*, Gcw61p para *Pichia pastoris* y AMPO1 para *A. oryzae*) cumplen con los criterios, incluyendo que no poseen el grupo dibásico en el carboxilo terminal y poseen pI menores de cinco, exceptuando Gcw61p que presenta un pI de 8.54. Esto muestra que puede haber proteínas en la pared celular que no cumplen todos los criterios, pero aquellas proteínas que los cumplen efectivamente se localizan en pared celular, lo cual apoya la estrategia seguida en el presente trabajo. Dos de las proteínas que fallaron en localizarse superficialmente en el trabajo de Adachi y colaboradores (2008) no son identificadas como GPI por los programas GPI-SOM predictor y predGPI predictor. Estos resultados también apoyan que

la estrategia de identificación *in silico* implementada es robusta para seleccionar proteínas que sean localizadas en la pared celular fúngica. Siguiendo esta estrategia se identificaron 4 proteínas (ID 1401645, 1436499, 2326 y 1397533 que poseen más de 30% de serina/treonina), pero también se decidió incluir las proteínas con ID 1401002, 1326581 y 1325672 que cumplen todos los criterios, excepto que poseen menos de 30% de serina/treonina; esto se decidió porque algunas de las proteínas que han sido ubicadas en pared (Zhang *et al.*, 2013) poseen 20% de serina y treonina.

Para construir el vector prototipo de expresión superficial de proteínas, sobre el que se generarán los propios, se utilizó la secuencia de la proteína AMPO1 (AO090011000119) de *A. oryzae* que fue la que mejor resultado dio en cuanto a anclaje en pared celular en el trabajo de Adachi y colaboradores (2008).

El casete diseñado en el presente trabajo incluye el promotor del gen Gliceraldehído-3-fosfato deshidrogenasa (*GPDA*), terminador del gen triptófano (*TRPC*), ambos de *Aspergillus nidulans* y la señal de secreción y el dominio GPI de la proteína AMPO1 de *A. oryzae*. El P_{gpDA} y el T_{trpC} fueron amplificados del vector pAN7.1 y la síntesis de la secuencia codificante de la proteína AMPO1 se solicitó a una empresa. Se diseñaron cebadores para que la unión de todos los fragmentos del casete se realizara mediante la tecnología Gibson®. El manual de éste kit reporta que puede ensamblar hasta 12 fragmentos sumando un tamaño total de 15-20 Kb. En el presente trabajo se ensamblaron cinco (promotor, señal de secreción, GFP, GPI, terminador) o seis fragmentos (los anteriores, más el vector receptor). Los cinco fragmentos tienen un tamaño total de ~4.8 Kb y repetidamente se logró ensamblar. Desafortunadamente los intentos de ligar el casete en pAN7.1 mediante tecnología Gibson® fueron infructuosos y cada vez se obtuvo un constructo de ~5 Kb que es menos de la mitad de lo esperado (~11 Kb). Mediante PCRs parciales, esto es, que amplificaban parte de un inserto y parte del siguiente inserto, se confirmó que el casete se logró ensamblar en todos sus fragmentos y en el orden correcto. También se confirmó mediante digestión con enzimas de restricción. Sin embargo, cuando se repetía el intento de ensamblar con pAN7.1 se obtenía el mismo resultado a pesar de iniciar con el casete de tamaño correcto (4.8 Kb) y con el pAN7.1 linearizado, de 6.8 Kb. Todas las veces se recuperaba un constructo de ~5 Kb.

Después de múltiples intentos y de confirmar que los elementos del inicio estaban bien, y que el resultado era *reproducibile*, se consideró que esto se debe a que el promotor y terminador usados en el casete fueron obtenidos del pAN7.1; el vector tiene estas

regiones regulatorias en su casete de resistencia a higromicina, por lo que posiblemente el sistema Gibson®, el cual está basado en recombinación homóloga, produce de manera reproducible el re-arreglo de las secuencias, recombinando las regiones comunes entre el casete y el pAN7.1. Esto suena lógico una vez que se analiza la situación, pero no era fácil de predecir al inicio. No hay reportes que describan este problema en los trabajos que han usado la tecnología Gibson® (Wagner *et al.* 2013; Gibson *et al.* 2010, 2009), ni tampoco en su manual. Muchos posibles problemas están enlistados y que en un principio se enfrentaron y resolvieron en el presente trabajo, como no recuperar células transformantes si las concentraciones de los fragmentos para ligar no alcanzan las concentraciones recomendadas, o recuperar únicamente colonias con el vector religado, y que fue resuelto evaluando de manera independiente el porcentaje de vector digerido y purificando el vector digerido. Sin embargo, a pesar de que la lista de posibles problemas es larga (Gibson Assembly® Master Mix, New England, Biolabs), en ningún caso alerta que no debe haber secuencias comunes entre el inserto y otros fragmentos a ligar o con el vector receptor. Lo que es más, en preguntas frecuentes incluye “¿es este método aplicable para el ensamble de secuencias repetidas?”, la respuesta que indica es. “Sí. Sin embargo, se debe asegurar que cada inserto tenga secuencias de traslape únicas con su contiguo para que las secuencias se puedan ensamblar apropiadamente. El ensamble se puede hacer en una etapa o en un proceso de dos pasos”. En el caso de los fragmentos aquí ensamblados, cada fragmento tuvo secuencias de traslape únicos. El casete completo tenía fragmentos en sus extremos que traslapan de manera específica con los extremos respectivos del pAN7.1 linearizado, para que se pudiera ligar por Gibson®. Se intentó ensamblar todo en un solo paso (6 fragmentos), y también ensamblando primero el casete y luego ligarlo mediante Gibson® al pAN7.1. Como se ha mencionado, ambos dieron el mismo constructo de tamaño menor al esperado. Las evaluaciones por PCR y digestión de estos constructos apoyan que hubo recombinación entre el casete de expresión superficial construido *in vitro* y el casete de higromicina del pAN7.1.

Cabe mencionar que algunos vectores para hongos usan en sus diferentes casetes el mismo promotor y terminador; por ejemplo el vector GFP (Maor *et al.*, 1998). En este caso los casetes de resistencia a higromicina y el casete para expresión de la proteína GFP tienen ambos el mismo promotor *GPD* y terminador *TRPC*. Este plásmido fue construido mediante la tecnología convencional con enzimas de restricción.

La recombinación homóloga es menos eficiente en los hongos en comparación con las bacterias y las levaduras, lo que explica por qué el vector GFP no tiene problema en la transformación de los hongos filamentosos; de hecho, el knock out en hongos requiere de más de 1 Kb en cada extremo para que se dé la recombinación homóloga, y la generación de mutantes knock out ocurre con baja eficiencia (Choquer *et al.*, 2008; Chaverocche *et al.*, 2000). ¿Por qué el vector GFP puede ser clonado en bacteria y no es reorganizado ni reestructurado durante este proceso? La respuesta puede provenir del hecho de que el vector GFP es circular y superenrollado. La reacción de Gibson® debe supuestamente circularizar el constructo final ya que la mezcla contiene una exonucleasa, una ADN polimerasa y una ADN ligasa, pero si queda algún corte (“nick”), éste usualmente puede ser reparado dentro de la bacteria, pero si el sistema bacteriano encuentra en el ADN foráneo (con “nicks”) regiones repetidas, el resultado pudiera ser diferente al deseado. Se sabe que la integración del ADN foráneo al endógeno se favorece cuando el vector es linearizado con una enzima de restricción, en comparación con la menor eficiencia de introducir el vector superenrollado. La metodología de linearizar el vector es conocida como REMI, por “Restriction Enzyme Mediated Integration” (Kuspa 2006; Riggle y Kumamoto, 1998).

En cuanto a los ensayos de transformación de especies de *Trichoderma*, no se pudo transformar *Trichoderma longibrachiatum*. La cepa presentó alta tolerancia a la higromicina y mostró muchos escapes. No se recuperaron clonas transformadas en ninguno de los intentos lo que hace suponer que los protoplastos de *T. longibrachiatum* no se volvían competentes. Una posibilidad menos plausible es que las regiones regulatorias de los casetes de expresión en el vector GFP no fueran funcionales en *T. longibrachiatum*. Paralelamente se estuvo tratando transformar *T. harzianum* y en este caso la transformación se confirmó mediante PCR en colonia y mediante aislamiento y digestión del plásmido GFP, por lo que siendo una especie del mismo género se decidió continuar con éste ya que *T. harzianum* también es una especie con alta capacidad secretoria (Ferreira-Filho *et al.*, 2017). Las clonas de *T. harzianum* transformadas con el constructo de expresión superficial crecieron muy lentamente sobre higromicina, pero significativamente mejor que la cepa silvestre. La observación al microscopio confocal mostró fluorescencia en los septos de las hifas del hongo, mientras que una transformante con el vector GFP mostró, como se esperaba, fluorescencia en toda la hifa (Figs. 3.22 y 3.23). Esta prueba funcional indica que tanto el casete de expresión superficial como el

casete de higromicina retuvieron al menos una actividad mínima del promotor, lo que permitió observar la fluorescencia y se confirmó que la AMPO1 se relaciona con localización en regiones superficiales de la célula. Sin embargo, para que sea biotecnológicamente útil el vector, ambos casetes deben estar funcionando adecuadamente porque no es posible trabajar con una cepa que crece mal y/o que expresa de manera subóptima la proteína de interés.

Se realizó un último intento, basado en el conocimiento descrito anteriormente y se decidió hacer el constructo ligando el casete a un fragmento del vector pSilent-1. En este vector el casete de higromicina (*HPH*) contiene al promotor *TRPC*, aunque el terminador desafortunadamente es el mismo. Se digirió el vector pSilent-1 con *XbaI* lo que produce 2 fragmentos; se purificó el fragmento 4.8 Kb que contiene el sitio de origen de replicación en bacteria y en hongo y el casete de expresión del gen *HPH*. El constructo fue validado por PCR y digestión y coincide con el tamaño esperado. El único hecho inexplicable en este caso fue la aparición de un sitio *SpeI* que no estaba predicho. Se generó otro constructo ligando el casete de expresión superficial en pSilent-1 en *HindIII* y *ApaI* y todas las pruebas de ese vector coinciden con lo esperado. Falta transformar *T. harzianum* con esos constructos para corroborar que tanto el casete de la higromicina como el de la expresión superficial están funcionando adecuadamente. De confirmarse se tendrá listo el vector de expresión superficial.

Las tecnologías basadas en recombinación homóloga son cada vez más frecuentes ya que la ligación resulta más eficiente al no depender de la reacción de una ligasa; y las subclonaciones son más fáciles porque no dependen de sitios de restricción que frecuentemente resultan incompatibles entre el vector de salida y el vector de entrada. Estos sistemas se basan en las maquinarias de recombinación homóloga de diferentes bacteriófagos como el Lambda y el PAC1, por mencionar algunos (Sharan *et al.*, 2009) y llevan a cabo tanto recombinación *in vitro* (Hartley *et al.*, 2000) como *in vivo* (Thomason *et al.*, 2009). Estas metodologías son tremendamente útiles en la ingeniería genética moderna pero en ocasiones se presentan problemas como los enfrentados en el presente trabajo.

El tercer punto a desarrollar fue la selección de la lipasa. El interés en el presente trabajo fue por una triacilglicerol lipasa que sea capaz de hidrolizar ácidos grasos de cadena larga. Estas lipasas son de interés industrial en la modificación de alimentos, en detergentes, para la producción de fármacos y para la transesterificación para generar

biodiesel (Andulema y Gessesse, 2012). Estas lipasas, también llamadas lipasas verdaderas, pueden hidrolizar aceites vegetales como el de olivo, y esta fue la estrategia que se siguió para tamizar una colección de hongos. La cepa más lipolítica resultó ser una de *Trichoderma harzianum*. Después de revisar la literatura se observó que muchos reportes han encontrado cepas de *T. harzianum* como lipolíticas pero al momento no se había descrito el gen o proteína responsable de esa actividad. En el presente trabajo se realizó la búsqueda en el proteoma deducido de *T. harzianum* de todas sus presuntas triacilglicerol lipasas y se realizó una detallada caracterización *in silico*, identificando los motivos importantes de estas enzimas como son el bolsillo y la tríada catalítica, el oxianión y el dominio "lid". Mediante búsqueda blast en bases de datos especializadas como LED ("Lipase Engineering Database") y ESTHER Database ("ESTerases and alpha/beta-Hydrolase Enzymes and Relatives"), se predijeron sus posibles funciones y el análisis se apoyó con el modelamiento de las proteínas y el análisis filogenético. La lista de proteínas candidatas se fue tamizando y al final se analizó la expresión de 6 lipasas en presencia y ausencia de aceite de olivo en el medio de cultivo, observando que la lipasa con número de identificación 52630 se induce *de novo*. Esta proteína es homóloga a la lipasa CalA de *Candida antarctica*. Es posible que la proteína 52630 posea características fisicoquímicas similares a CalA. Ésta última es tolerante a disolventes y a alta temperatura (Kirk y Christensen, 2002), lo que coincide con las características descritas por Coradi y colaboradores (2013) y Ülker colaboradores (2011) para la lipasa de *T. harzianum* cuando éstos autores midieron su actividad y la purificaron a partir del hongo cultivado en medio con aceite de olivo. Sin embargo, esos autores no aislaron el transcrito ni secuenciaron la proteína que purificaron. Para caracterizar la proteína 52630 se plantea clonarla en un vector comercial para su expresión heteróloga en levadura, pero esto último ya no formó parte del presente trabajo.

El trabajo de generar un vector completo a partir de su partes es complejo, retador y una tarea que requiere mucho tiempo de trabajo. Es común la inversión de varios años de trabajo (Wagner *et al.*, 2013). Máxime en el presente caso en el que gran cantidad de ensayos fueron negativos. Sin embargo se logró avanzar en la construcción del vector propio. Asimismo se han generado clonas que contienen los ORF de las proteínas GPIs. Cada una está flanqueada por las mismas secuencias de restricción que la actual GPI en el vector de expresión superficial, lo que permitirá cortarlas del vector de clonación y

subclonaras en el lugar de la actual GPI (AMPO1) en el constructo de expresión superficial generado en el presente trabajo.

La región codificante de la lipasa también ya se encuentra clonada y confirmada que está en marco de lectura abierta. El ORF está flanqueado por los sitios de restricción que flanquean la GFP en el vector de expresión superficial, por lo que se sustituirá la GFP por la lipasa. A pesar de todas las dificultades y de no haberse logrado evaluar en el constructo de expresión superficial, todas y cada una de las secuencias nuevas de proteínas GPIs y de la lipasa, se logró preparar todos los elementos que se necesitan para generar el desarrollo tecnológico que se planteó. El grupo de trabajo continuará para terminar este desarrollo y de confirmarse los resultados esperados, solicitar la protección intelectual de dicho desarrollo.

CAPÍTULO VI

CONCLUSIONES Y PERSPECTIVAS

6.1 CONCLUSIONES

En el presente trabajo se lograron identificar y caracterizar *in silico* 43 posibles proteínas con anclaje GPI en el genoma de *T. reesei* y 7 del genoma de *T. longibrachiatum*, aunque la mayoría de ellas presentan funciones desconocidas, se identificaron algunas con homología a proteínas encargadas de la síntesis y remodelamiento de la pared celular, así como varias que presentaban el dominio CFEM característico de proteínas extracelulares relacionado con la patogenicidad y virulencia.

A pesar de que no fue posible generar el vector de expresión para la proteína quimérica Ss-GFP-GPI de manera correcta mediante el sistema Gibson®, se generaron dos ORF los cuales presentan los sitios de restricción necesarios para subclonar los fragmentos que codifican las proteínas con anclaje GPI y el fragmento que codifica la lipasa de interés para la generación del vector mediante el uso de enzimas de restricción y su posterior uso para la generación del bio-catalizador de célula completa. El sistema de ensamble Gibson® resulta eficiente en la ligación de múltiples fragmentos (5 para el presente trabajo). Sin embargo, se debe tener en alta consideración la homología existente entre los fragmentos a ligar y el vector plataforma ya que pueden generar procesos de re arreglos.

Se probaron diferentes protocolos para la transformación de las cepas de *T. harzianum* y *T. longibrachiatum*, siendo el protocolo de Polietilenglicol (PEG) y cloruro de calcio el de mejores resultados y el cual fue reproducible usando diferentes plásmidos para la transformación y diferentes concentraciones, siendo de 5 a 10 µg de ADNp suspendido en máximo 10 µL de agua, la concentración con mejores resultados en la obtención de los transformantes.

El escrutinio *in silico* del gen de lipasas de interés en *T. harzianum* permitió identificar de un total de 50 candidatos, siete como posibles triacilglicerol lipasas, de los cuales la 526309 es inducida por presencia de aceite de olivo en el medio, la cual presenta homología con CALA una lipasa de gran interés para la bio catálisis. Se identificó *in silico*

un nuevo pentapéptido para lipasas de hongos filamentosos descrito únicamente en bacterias y en dos levaduras.

6.2 PERSPECTIVAS

El presente trabajo es el primer análisis en identificar proteínas con anclaje GPI en el género de los *Hypocreales* e identificar proteínas de interés ya sea por su función en la síntesis de la pared o para su posible uso en los sistemas de expresión en la superficie de la célula. Aunque desafortunadamente no se logró comparar la posición y capacidad de anclaje de las proteínas seleccionadas, se establecieron los protocolos tanto de subclonación de las secuencias, como para la transformación de las cepas con el fin de realizar el análisis a la brevedad posible.

El análisis de la capacidad de anclaje de las proteínas seleccionadas permitirá identificar una o varias proteínas que puedan ser utilizadas para el anclaje de enzimas de interés en la superficie celular en *Trichoderma harzianum*, pudiéndose extrapolar a otros hongos filamentosos de interés industrial.

Por su parte el análisis y caracterización *in silico* de lipasas en el proteoma deducido de *T. harzianum* permitió identificar al menos 4 lipasas de interés para su posterior caracterización en el laboratorio; las 77338 y 78181, por su homología con lipasas relacionadas a patogenicidad y nutrición previamente descritas en otros hongos, ambas ya se presentan en proceso de clonación en el sistema Gateway, la 135964 por la presencia del pentapéptido AHSMG descrito para bacterias y únicamente para dos; y la 526309 debido a su homología con CALA y a que es inducida por la presencia de aceite de olivo en el medio, lo que significaría su capacidad para trabajar con triacilglicéridos de cadena larga los cuales son importantes en la industria alimenticia y el desarrollo de biocombustibles como el biodiesel.

BIBLIOGRAFÍA

- Abdel-Azeem, A. M (2010). The history, fungal biodiversity, conservation, and future perspectives for mycology in Egypt. *IMA Fungus*, *The Global Mycological Journal*, 1, 123-142.
- Adachi, T., J. Ito, K. Kawata, M. Kaya, H. Ishida, H. Sahara, Y. Hata, C. Ogino, H. Fukuda y A. Kondo (2008). Construction of an *Aspergillus oryzae* cell-surface display system using a putative GPI-anchored protein. *Applied Microbiology and Biotechnology*, 81, 711-719.
- Adams, D. J. (2004). Fungal cell wall chitinases and glucanases. *Microbiology*, 150, 2029-2035.
- Altschul, S. F, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller y D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST, a new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389-3402.
- Alva, V., S. Z. Nam, J. Söding y A. N. Lupas (2016). The MPI Bioinformatics Toolkit as an Integrative Platform for Advanced Protein Sequence and Structure Analysis. *Nucleic Acids Research*, 44, W410-W415.
- An, Y., W. Wu y A. Lv (2010). A PCR-after-ligation method for cloning of multiple DNA inserts. *Analytical Biochemistry*, 402, 203-205.
- Anderson, JC., J. E. Dueber, M. Leguia, G. C. Wu, J. A. Goler, A. P. Arkin y J. D. Keasling (2010). BglBricks, A flexible standard for biological part assembly. *Journal of Biological Engineering*, 4, 1.
- Andualema, B. y A. Gessesse (2012). Microbial Lipases and Their Industrial Applications, Review. *Biotechnology (Faisalabad)*, 11, 100-118.
- Barriuso, J. y M. J. Martínez (2017). Evolutionary History of Versatile-Lipases from Agaricales through Reconstruction of Ancestral Structures. *BMC Genomics*, 18.

- Berka, R. M. y C. C. Barnett (1989). The development of gene expression systems for filamentous fungi. *Biotechnology Advances*, 7, 127-154.
- Biasini, M., S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer T. G. Cassarino, M. Bertoni, L. Bordoli y T. Schwede (2014). SWISS-MODEL, modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42, W252-258.
- Boisramé, A., A Cornu, G. Da Costa y M. L. Richard (2011). Unexpected Role for a Serine/Threonine-Rich Domain in the *Candida albicans* Iff Protein Family. *Eukaryotic Cell*, 10, 1317-1330.
- Bordes, F., S. Barbe, P. Escalier, L. Mourey, I. André, A. Marty y S. Tranier (2010). Exploring the Conformational States and Rearrangements of *Yarrowia lipolytica* Lipase. *Biophysical Journal*, 99, 2225-2234.
- Bowman, S. M. y S. J. Free (2006). The structure and synthesis of the fungal cell wall. *BioEssays*, 28 (8), 799-808.
- Brenneis, R. y B. Baeck (2012). Esterification of Fatty Acids Using *Candida antarctica* Lipase A in Water-Abundant Systems. *Biotechnology Letters*, 34, 1459-1463.
- Bruneau, J. M., T. Magnin, E. Tagat, R. Legrand, M. Bernard, M. Diaquin, C. Fudali y J. P. Latgé (2001). Proteome analysis of *Aspergillus fumigatus* identifies glycosylphosphatidylinositol-anchored proteins associated to the cell wall biosynthesis. *ELECTROPHORESIS*, 22, 2812-2823.
- Bueter, C. L., C. A. Specht y S. M. Levitz (2013). Innate Sensing of Chitin and Chitosan. *PLOS Pathogens*, 9, e1003080.
- Cao, W., J. Maruyama, K. Kitamoto, K. Sumikoshi, T. Terada, S. Nakamura y K. Shimizu (2009). Using a new GPI-anchored-protein identification system to mine the protein databases of *Aspergillus fumigatus*, *Aspergillus nidulans*, and *Aspergillus oryzae*. *The Journal of General and Applied Microbiology*, 55, 381-93.

- European Bioinformatics Institute, 2018. Carbohydrate-binding WSC (IPR002889) < InterPro < EMBL-EBI. [Online] (Actualizado 2018). Disponible en: <https://www.ebi.ac.uk/interpro/entry/IPR002889>. [Acceso 10 mayo 2018].
- Caro, L. H. P., H Tettelin, J. H. Vossen, A. F. J. Ram, H. Van Den Ende y F. M. Klis (1997). In Silicio Identification of Glycosyl-Phosphatidylinositol-Anchored Plasma-Membrane and Cell Wall Proteins of *Saccharomyces cerevisiae*. *Yeast*, 13, 1477-1489.
- Casas-Godoy, L., S. Duquesne, F. Bordes, G. Sandoval y A. Marty (2012). Lipases, An Overview, en *Lipases and Phospholipases*, Sandoval, G. (ed). NJ, Humana Press. Totowa. pp. 3-30.
- Casini, A., M. Storch, G. S. Baldwin y T. Ellis (2015). Bricks and Blueprints, Methods and Standards for DNA Assembly. *Nature Reviews Molecular Cell Biology*, 16, 568-576.
- Castillo-Avila, G. (2010). Aislamiento de hongos lignocelulolíticos a partir de residuos agroindustriales de plátano. Centro de investigacion científica de Yucatán. México. pp.108
- Cedillo, V., F. J. Plou y M. Martínez (2012). Recombinant Sterol Esterase from *Ophiostoma piceae*, An Improved Biocatalyst Expressed in *Pichia pastoris*. *Microbial Cell Factories*, 11, 73.
- Chaverroche, M., J. Ghigo y C. d'Enfert (2000). A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. *Nucleic Acids Research*, 28, e97-e97.
- Cheon, S. A., J. Jung, J. H. Choo, D. Oh y H. A. Kang (2014). Characterization of Putative Glycosylphosphatidylinositol-Anchoring Motifs for Surface Display in the Methylophilic Yeast *Hansenula polymorpha*. *Biotechnology Letters*, 36, 2085-94.
- Choquer, M., G. Robin, P. Le Pêcheur, C. Giraud, C. Levis y M. Viaud (2008). *Ku70* or *Ku80* Deficiencies in the Fungus *Botrytis cinerea* Facilitate Targeting of Genes That

- Are Hard to Knock out in a Wild-Type Context. *FEMS Microbiology Letters*, 289, 225–232.
- Conde-Ferrández, L. R. Grijalva y A.C. James-Kay (2008). A rapid DNA extraction method from mycelium which is suitable for PCR. *Revista Latinoamericana de Microbiología*, 50, 86-88.
- Coradi, G. V., V. L. da Visitação, E. A. de Lima, L. Y. T. Saito, D. A. Palmieri, M. A. Takita, P. de Oliva Neto y V. M. G. de Lima (2013). Comparing Submerged and Solid-State Fermentation of Agro-Industrial Residues for the Production and Characterization of Lipase by *Trichoderma harzianum*. *Annals of Microbiology*, 63, 533-540.
- Daiha, K., R. Angeli, S. Dias y R. Volcan (2015). Are Lipases Still Important Biocatalysts? A Study of Scientific Publications and Patents for Technological Forecasting. Editado por Willem J.H. van Berkel. *PLOS ONE*, 10, e0131624.
- Damveld, R. A., M. Arentshorst, P. A. VanKuyk, F. M. Klis, C. A.M.J.J. van den Hondel y A.F.J. Ram (2005). Characterisation of CwpA, a putative glycosylphosphatidylinositol-anchored cell wall mannoprotein in the filamentous fungus *Aspergillus niger*. *Fungal Genetics and Biology*, 42, 873-885.
- De Groot, P. W.J., A. F. Ram y F. M. Klis (2005). Features and functions of covalently linked proteins in fungal cell walls. *Fungal Genetics and Biology*, 42 (8), 657-675.
- DeLano, W.L. (2002). *The PyMol Molecular Graphics System*, Vol. 30.
- Dror, A., M. Kanteev, I. Kagan, S. Gihaz, A. Shahar y A. Fishman (2015). Structural Insights into Methanol-Stable Variants of Lipase T6 from *Geobacillus stearothermophilus*. *Applied Microbiology and Biotechnology*, 99, 9449-9461.
- Druzhinina, I. S., A. G. Kopchinskiy, M. Komoń, J. Bissett, G. Szakacs y C. P. Kubicek (2005). An oligonucleotide barcode for species identification in *Trichoderma* and *Hypocrea*. *Fungal Genetics and Biology*, 42, 813-828.

- Ecker, M., R. Deutzmann, L. Lehle, V. Mersa y W. Tanner (2006). Pir Proteins of *Saccharomyces Cerevisiae* Are Attached to β -1,3-Glucan by a New Protein-Carbohydrate Linkage. *Journal of Biological Chemistry*, 281, 11523-11529.
- Eddine, A. N., F. Hannemann y W. Schafer (2001). Cloning and expression analysis of NhL1, a gene encoding an extracellular lipase from the fungal pea pathogen *Nectria haematococca* MP VI (*Fusarium solani* f. sp. pisi) that is expressed in planta. *Mol Genet Genomics*, 265, 215-224.
- Eisenhaber, B., G. Schneider, M. Wildpaner y F. Eisenhaber (2004). A Sensitive Predictor for Potential GPI Lipid Modification Sites in Fungal Protein Sequences and its Application to Genome-wide Studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Journal of Molecular Biology*, 337, 243-253.
- Ericsson, D. J., A. Kasrayan, P. Johansson, T. Bergfors, A. G. Sandström, J. E. Bäckvall y S. L. Mowbray (2008). X-Ray Structure of *Candida antarctica* Lipase A Shows a Novel Lid Structure and a Likely Mode of Interfacial Activation. *Journal of Molecular Biology*, 376, 109-119.
- Falony, G., J. C. Armas, J. C. Dustet y J. Martínez-Hernández (2006). Production of Extracellular Lipase from *Aspergillus niger* by Solid-State Fermentation. *Food Technology & Biotechnology*, 44, 235-240.
- Fankhauser, N. y P. Mäser (2005). Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics*, 21, 1846-1852.
- Feng, J. (2005). A Secreted Lipase Encoded by LIP1 Is Necessary for Efficient Use of Saturated Triglyceride Lipids in *Fusarium graminearum*. *Microbiology*, 151, 3911-3921.
- Feng, J., R. Hwang, S.-F. Hwang, D. Gaudet y S. E. Strelkov (2011). Molecular Characterization of a *Stagonospora nodorum* Lipase Gene LIP1, A Triglyceride Lipase from *Stagonospora Nodorum*. *Plant Pathology*, 60, 698-708.

- Feng, J., F.Wang, G. Liu, D. Greenshields, W. Shen, S. Kaminskyj, G. R. Hughes, Y. Peng, G. Selvaraj, J. Zou y Y. Wei (2009). Analysis of a *Blumeria graminis*-Secreted Lipase Reveals the Importance of Host Epicuticular Wax Components for Fungal Adhesion and Development. *Molecular Plant-Microbe Interactions*, MPMI, 22, 1601–1610.
- Ferreira-Filho, J. A., M. A. Crivelente, L. L. B., C. A. dos Santos y A. P. de Souza (2017). Carbohydrate-active enzymes in *Trichoderma harzianum*: a bioinformatic analysis bioprospecting for key enzymes for the biofuels industry. *BMC Genomics*, 18, 779.
- Finn, R. D., T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork, A. J. Bridge, H. Chang, Z. Dostányi, S. El-Gebali, M. Fraser, J. Gough, D. Haft, G. Holliday, H. Huang, X. Huang, I. Letunic, R. Lopez, S. Lu, A. Marchler-Bauer, H. Mi, J. Mistry, D. NAtale, M. Necci, G. Nuka, C. Orengo, Y. Park, S. Pesseat, D. Piovesan, S. Potter, N. Rawlings, N. Redaschi, L. Richardson, C. Rivoire, A. Sagrador-Vegas, C. Sigrist, I. Sillitoe, B. Smithers, S. Squizzato, G. Sutton, N. Thanki, P. Thomas, S. Tosatto, C. Wu, I. Xenarios, L. Yeh, S. Young y A. Mitchell (2017). InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Research*, 45, D190–199.
- Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. Salazar, J. Tate y A. Bateman (2016). The Pfam Protein Families Database, Towards a More Sustainable Future. *Nucleic Acids Research*, 44, D279-285.
- Firon, A., G. Lesage y H. Bussey (2004). Integrative studies put cell wall synthesis on the yeast functional map. Growth and development. *Current Opinion in Microbiology*, 7, 617-623.
- Fischer, M. y J. Pleiss (2003). The Lipase Engineering Database, a navigation and analysis tool for protein families. *Nucleic Acids Research*, 31, 319-321.
- Fontaine, T. (2003). Structures of the glycosylphosphatidylinositol membrane anchors from *Aspergillus fumigatus* membrane proteins. *Glycobiology*, 13, 169-177.

- Free, Stephen J. 2013. Chapter Two - Fungal Cell Wall Organization and Biosynthesis, en *Advances in Genetics*, Friedmann, T., J. C. Dunlap, y S. F. Goodwin (ed), Academic Press. Oxford. pp. 33-82
- Freitag, M., P. C. Hickey, N. B. Raju, E. U. Selker y N. D. Read (2004). GFP as a tool to analyze the organization, dynamics and function of nuclei and microtubules in *Neurospora crassa*. *Fungal Genetics and Biology*, 41, 897-910.
- Frieman, M. B. y B. P. Cormack (2003). The ω -site sequence of glycosylphosphatidylinositol-anchored proteins in *Saccharomyces cerevisiae* can determine distribution between the membrane and the cell wall. *Molecular Microbiology*, 50, 883-896.
- Frieman, M. B. y B. P. Cormack (2004). Multiple sequence signals determine the distribution of glycosylphosphatidylinositol proteins between the plasma membrane and cell wall in *Saccharomyces cerevisiae*. *Microbiology*, 150, 3105-3114.
- Fu, C., E. Sokolow, C. B. Rupert y S. J. Free (2014). The *Neurospora Crassa* CPS-1 Polysaccharide Synthase Functions in Cell Wall Biosynthesis. *Fungal Genetics and Biology*, 69, 23-30.
- Fujita, M. y T. Kinoshita (2012). GPI-anchor remodeling, Potential functions of GPI-anchors in intracellular trafficking and membrane dynamics. *Lipids and Vesicular Transport*, 1821 (8), 1050-1058.
- Gastebois, A., T. Fontaine, J. Latgé y I. Mouyna. 2010. β (1-3) Glucanosyltransferase Gel4p Is Essential for *Aspergillus fumigatus*. *Eukaryotic Cell*, 9, 1294-98.
- Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel y A. Bairoch (2005). Protein Identification and Analysis Tools on the ExPASy Server, en *The Proteomics Protocols Handbook*, Walker, J. M. (ed). NJ, Humana Press, Totowa. pp. 571–607.
- Gelly, J., A. Praveen, N. Srinivasan y A. G de Brevern (2011). iPBA, a tool for protein structure comparison using sequence alignment strategies. *Nucleic Acids Research*, 39 (Web Server issue), W18–23.

- Gibson, D. G., H. O. Smith, C. A. H. Iii, J. C. Venter y C. Merryman. 2010. "Chemical Synthesis of the Mouse Mitochondrial Genome". *Nature Methods* 7: 901-903.
- Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. H. Iii y H. O. Smith. (2009). Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nature Methods* 6, 343-45.
- Goldman, G. H., M. Van Montagu y A. Herrera-Estrella (1990). Transformation of *Trichoderma harzianum* by High-Voltage Electric Pulse. *Current Genetics*, 17, 169-174.
- Gopinath, S. C. B., P. Anbu, T. Lakshmipriya y A. Hilda (2013). Strategies to Characterize Fungal Lipases for Applications in Medicine and Dairy Industry. *BioMed Research International*, 2013, 1–10.
- Grün, C. H. (2003). Structure and Biosynthesis of Fungal Alpha-Glucans. Tesis de Doctorado. Utrecht University. Países Bajos. pp. 144
- Gupta, R., A. Kumari, P. Syal y Y. Singh (2015). Molecular and Functional Diversity of Yeast and Fungal Lipases, Their Role in Biotechnology and Cellular Physiology. *Progress in Lipid Research*, 57, 40-54.
- Ham, H. J., H. J. Rho, S. K Shin y H. J. Yoon (2010). The TGL2 Gene of *Saccharomyces cerevisiae* Encodes an Active Acylglycerol Lipase Located in the Mitochondria. *Journal of Biological Chemistry*, 285, 3005-3013.
- Hamada, K, S Fukuchi, M Arisawa, M Baba y K Kitada (1998). Screening for glycosylphosphatidylinositol (GPI)-dependent cell wall proteins in *Saccharomyces cerevisiae*. *Mol Gen Genet*, 258, 53-59.
- Haraldsson, G. G. y B. Kristinsson (2016). Process for separating polyunsaturated fatty acids from long chain unsaturated or less saturated fatty acids. US9476008 B2, subido el 23 de junio de 2011, y emitido el 25 de octubre de 2016. <http://www.google.ch/patents/US9476008>.
- Hartley, J. L., G. F. Temple y M. A. Brasch (2000). DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Research*, 10, 1788-1795.

- Hasan, S. y G. Gupta (2012). Fungal Biodiversity , Evolution & Distribution-A Review International Journal of Applied Research & Studies, 1, 1-8.
- He, R., C. Zhang, W. Guo, L. Wang, D. Zhang y S. Chen (2013). Construction of a plasmid for heterologous protein expression with a constitutive promoter in *Trichoderma reesei*. Plasmid, 70, 425-429.
- Honda, S. y E. U. Selker (2009). Tools for Fungal Proteomics, Multifunctional Neurospora Vectors for Gene Replacement, Protein Expression and Protein Purification. Genetics, 182, 11-23.
- Horton, P., K. Park, T. Obayashi, N. Fujita, H. Harada, C.J. Adams-Collier y K. Nakai (2007). WoLF PSORT, protein localization predictor. Nucleic Acids Research, 35, W585–587.
- Hotelier, T., L. Renault, X. Cousin, V. Negre, P. Marchot y A. Chatonnet (2004). ESTHER, the database of the α/β -hydrolase fold superfamily of proteins. Nucleic Acids Research, 32, D145-147.
- Jaeger, K. E., S. Ransac, B. W. Dijkstra, C. Colson, M. van Heuvel y O. Misset (1994). Bacterial Lipases. FEMS Microbiology Reviews 15, 29-63.
- Jiang, Z., B.Gao, R. Ren, X. Tao, Y. Ma y D. Wei (2008). Efficient display of active lipase LipB52 with a *Pichia pastoris* cell surface display system and comparison with the LipB52 displayed on *Saccharomyces cerevisiae* cell surface. BMC Biotechnology, 8, 4.
- Joint Genome Institute, 2014. Home - *Trichoderma longibrachiatum* ATCC 18648 v1.0. [Online] (Acceso 18 marzo 2018). Disponible en: <https://genome.jgi.doe.gov/Trilo1/Trilo1.home.html>.
- Juhas, M. y J. W. Ajioka (2017). High molecular weight DNA assembly in vivo for synthetic biology applications. Critical Reviews in Biotechnology, 37, 277-286.
- Kantún-Moreno, N. (2012). Estudio de genes que codifican proteínas de pared celular del hongo *Mycosphaerella fijiensis*. Tesis de Doctorado. Centro de Investigación Científica de Yucatán. México. pp. 117

- Kantún-Moreno, N., R. Vázquez-Euán, M. Tzec-Simá, L. Peraza-Echeverría, R. Grijalva-Arango, C. Rodríguez-García, A. C. James, J. Ramírez-Prado, I. Islas-Flores y B. Canto-Canché (2013). Genome-wide in silico identification of GPI proteins in *Mycosphaerella fijiensis* and transcriptional analysis of two GPI-anchored β -1,3-glucanotransferases. *Mycologia*, 105, 285-296.
- Katoh, K. y H. Toh (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, 9, 286-298.
- Keränen, S. y M. Penttilä (1995). Production of recombinant proteins in the filamentous fungus *Trichoderma reesei*. *Current Opinion in Biotechnology*, 6, 534-537.
- Kim, S. Y., J. H. Sohn, Y. R. Pyun y E. S. Choi (2002). A cell surface display system using novel GPI-anchored proteins in *Hansenula polymorpha*. *Yeast*, 19, 1153-1163.
- Kirk, O. y M. W. Christensen (2002). Lipases from *Candida antarctica*: Unique Biocatalysts from a Unique Origin. *Organic Process Research & Development*, 6, 446-451
- Klein, I., L. Klug, C. Schmidt, M. Zandl, M. Korber, G. Daum y K. Athenstaedt (2016). Regulation of the Yeast Triacylglycerol Lipases Tgl4p and Tgl5p by the Presence/Absence of Nonpolar Lipids. *Molecular Biology of the Cell*, 27, 2014-2024.
- Klis, F. M., C. G. de Koster y S. Brul (2011). A mass spectrometric view of the fungal wall proteome. *Future Microbiology*, 6, 941-951.
- Klis, F. M., P. Mol, K. Hellingwerf y B. Stanley (2002). Dynamics of cell wall structure in *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews*, 26, 239-256.
- Kliss, F. M., B. Stanley y P. W. J. De Groot (2009). Covalently linked wall proteins in ascomycetous fungi. *Yeast*, 27, 489-493.
- Koch, M. R. y L. Pillus (2009). The glucanotransferase Gas1 functions in transcriptional silencing. *Proceedings of the National Academy of Sciences*, 106, 11224.

- Korabecna, M. (2007). The Variability in the Fungal Ribosomal DNA (ITS1, ITS2, and 5.8 S RRNA Gene), Its Biological Meaning and Application in Medical Mycology. Badajoz: Formatex Research Center, 783-787.
- Krogh, A., B. Larsson, G. von Heijne y E. L.L. Sonnhammer (2001). Predicting transmembrane protein topology with a hidden markov model, application to complete genomes. *Journal of Molecular Biology*, 305, 567-580.
- Kumari, A. y R. Gupta (2015). Functional Characterisation of Novel Enantioselective Lipase TALipA from *Trichosporon asahii* MSR54, Sequence Comparison Revealed New Signature Sequence AXSXG Among Yeast Lipases. *Applied Biochemistry and Biotechnology*, 175, 360-371.
- Kuspa, A. (2006). Restriction Enzyme-Mediated Integration (REMI) Mutagenesis. En *Dictyostelium discoideum* Protocols, de Eichinger Ludwig, Rivero-Crespo F., y Rivero F., New Jersey: Humana Press, 346,
- Lanka, S. y J. N. L. Latha (2015). Purification and characterization of a new cold active lipase, EnL A from *Emericella nidulans* NFCCI 3643. *Africal Journal of Biotechnology*, 14, 1897-1909.
- Latgé, J. P. (2010). Tasting the Fungal Cell Wall. *Cellular Microbiology*, 12, 863-872.
- Lenardon, M. D., C. A. Munro y N. A. R. Gow (2010). Chitin synthesis and fungal pathogenesis. *Current Opinion in Microbiology* 13 (4), 416-423.
- Li, L., W. Jiang y Y. Lu (2018). A Modified Gibson Assembly Method for Cloning Large DNA Fragments with High GC Contents, en: *Synthetic Metabolic Pathways. Methods in Molecular Biology*, Jensen, M. K. y J. D. Keasling (ed). Humana Press, New York, NY. pp. 203–209.
- Lima, M. G. V., N. Krieger, M. I. M Sarquis, D. A. Mitchell, L. P. Ramos, y J. D. Fontana. 2003. Effect of Nitrogen and Carbon Sources on Lipase Production by *Penicillium aurantiogriseum*. *Food Technology and Biotechnology*, 412, 105-110.

- Lou, Z., M. Li, Y. Sun, Y. Liu, Z. Liu, W. Wu y Z. Rao (2010). Crystal Structure of a Secreted Lipase from *Gibberella zeae* Reveals a Novel 'Double-Lock' Mechanism. *Protein & Cell*, 1, 760-770.
- Lu, C. F., J. Kurjan y P. N. Lipke (1994). A pathway for cell wall anchorage of *Saccharomyces cerevisiae* alpha-agglutinin. *Molecular and Cellular Biology*, 14, 4825-4833.
- Lu, Z., R. Tombolini, S. Woo, S. Zeilinger, M. Lorito y J. K. Jansson (2004). In Vivo Study of *Trichoderma*-Pathogen-Plant Interactions, Using Constitutive and Inducible Green Fluorescent Protein Reporter Systems. *Applied and Environmental Microbiology*, 70, 3073-3081.
- Mach, R. y S. Zeilinger (2003). Regulation of Gene Expression in Industrial Fungi, *Trichoderma*. *Applied Microbiology and Biotechnology*, 60, 515-522.
- Maddi, A., C. Fu y S. J. Free (2012). The *Neurospora crassa* dfg5 and dcw1 Genes Encode α -1,6-Mannanases That Function in the Incorporation of Glycoproteins into the Cell Wall. *PLOS ONE*, 7, e38872.
- Maor, R., M. Puyesky, B. A. Horwitz y A. Sharon (1998). Use of green fluorescent protein (GFP) for studying development and fungal-plant interaction in *Cochliobolus heterostrophus*. *Mycological Research*, 102, 491-96.
- Marchler-Bauer, A., Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. Derbyshire, R. Geer, N. Gonzales, M. Gwadz, D. Hurwitz, F. Lu, G. Marchler, J. Song, N. Thanki, Z. Wang, R. Yamashita, D. Zhang, C. Zheng, L. Geer y S. Bryant (2017). CDD/SPARCLE, functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45, D200-203.
- Martinez, A. I. (2004). Role of Pir1 in the Construction of the *Candida albicans* Cell Wall. *Microbiology*, 150, 3151-3161.
- Matsumoto, T., H. Fukuda, M. Ueda, A. Tanaka y A. Kondo (2002). Construction of Yeast Strains with High Cell Surface Lipase Activity by Using Novel Display Systems

- Based on the Flo1p Flocculation Functional Domain. *Applied and Environmental Microbiology*, 68, 4517-4522.
- Mehta, A., N. Kumar-Saun y R. Gupta (2016). Purification and Characterization of Lipase from *Thermophilic geobacillus* Sp. *Current Biotechnology*, 5, 81-89.
- Messaoudi, A., H. Belguith, I. Ghram y J. Ben (2011). LIPABASE, A Database for 'true' Lipase Family Enzymes. *International Journal of Bioinformatics Research and Applications*, 7, 390.
- Meyer, V., F. Wanka, J. van Gent, M. Arentshorst, C. A. M. J. J. van den Hondel y Arthur F J Ram (2011). Fungal Gene Expression on Demand, an Inducible, Tunable, and Metabolism-Independent Expression System for *Aspergillus niger*. *Applied and Environmental Microbiology*, 77, 2975-2983.
- Moura, M. V. H., G. P. da Silva, A. C. de Oliveira, F. A. Gonçalves, D. M. G. Freire y R. V. Almeida (2015). Displaying Lipase B from *Candida antarctica* in *Pichia pastoris* Using the Yeast Surface Display Approach: Prospection of a New Anchor and Characterization of the Whole Cell Biocatalyst. *PLOS ONE*, 10, e0141454.
- Mouyna, I., W. Morelle, M. Vai, M. Monod, B. Léchenne, T. Fontaine, A. Beauvais, J. Sarfati, M. Prévost, C. Henry y J. Latgé (2005). Deletion of GEL2 Encoding for a $\beta(1-3)$ Glucanoyltransferase Affects Morphogenesis and Virulence in *Aspergillus fumigatus*. *Molecular Microbiology*, 56, 1675-1688.
- Mukherjee, P. K., B. A. Horwitz, A. Herrera-Estrella, M. Schmoll y C. M. Kenerley (2013). *Trichoderma* Research in the Genome Era. *Annual Review of Phytopathology*, 51, 105-129.
- Nagao, T., Y. Shimada, A. Sugihara y Y. Tominaga (1994). Cloning and Nucleotide Sequence of cDNA Encoding a Lipase from *Fusarium heterosporum*. *The Journal of Biochemistry*, 116, 536-540.
- Najjar, A., S. Robert, C. Guérin, M. Violet-Asther y F. Carrière (2011). Quantitative study of lipase secretion, extracellular lipolysis, and lipid storage in the yeast *Yarrowia*

- lipolytica* grown in the presence of olive oil, analogies with lipolysis in humans. Applied Microbiology and Biotechnology, 89, 1947-1962.
- NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 44 (Database issue), D7–19.
- Nevalainen, K.M. H., V. S. J. Te'o y P. L. Bergquist (2005). Heterologous protein expression in filamentous fungi. Trends in Biotechnology, 23, 468-474.
- Nguyen, L. N, J. Bormann, G. T. T. Le, C. Stärkel, S. Olsson, J. D. Nosanchuk, H. Giese y W. Schäfer (2011). Autophagy-Related Lipase FgATG15 of *Fusarium graminearum* Is Important for Lipid Turnover and Plant Infection. Fungal Genetics and Biology, 48, 217-224.
- Nordberg, H., M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, I. Shabalov, T. Smirnova, I. V Grigoriev y I. Dubchak (2014). The genome portal of the Department of Energy Joint Genome Institute, 2014 updates. Nucleic Acids Research, 42, D26–31.
- Nwuche, C. O y J. C. Ogbonna (2011). Isolation of lipase producing fungi from palm oil Mill effluent (POME) dump sites at Nsukka. Brazilian Archives of Biology and Technology, 54, 113-116.
- Ortiz-Lechuga, E. G., I. Quintero-Zapata y A. Katiushka (2016). Detection of extracellular enzymatic activity in microorganisms isolated from waste vegetable oil contaminated soil using plate methodologies. African Journal of Biotechnology, 15, 408-416.
- Pareek, M., M. Sachdev, M. Tetorya y M. Rajam (2015). Glass-Bead and Agrobacterium-Mediated Genetic Transformation of *Fusarium oxysporum*, en Genetic Transformation Systems in Fungi, van den Berg M. y K. Maruthachalam (ed). Fungal Biology. Springer, Cham. Heidelberg New York Dordrecht London. pp. 169-174
- Park, M., W. H. Jung, S. H. Han, Y. H. Lee y Y. W. Lee (2015). Characterisation and Expression Analysis of MrLip1, a Class 3 Family Lipase of *Malassezia restricta*. Mycoses, 58, 671-678.

- Pereira, E., A. Tsang, T. A McAllister y R. Menassa (2013). The Production and Characterization of a New Active Lipase from *Acremonium alcalophilum* Using a Plant Bioreactor. *Biotechnology for Biofuels*, 6, 111.
- Petersen, T. Nordahl, S. Brunak, G. von Heijne y H. Nielsen (2011). SignalP 4.0, discriminating signal peptides from transmembrane regions. *Nature Methods*, 8, 785.
- Pittet, M. y A. Conzelmann (2007). Biosynthesis and function of GPI proteins in the yeast *Saccharomyces cerevisiae*. *Regulation of Lipid Metabolism in Yeast*, 1771, 405-420.
- Pöggeler, S., S. Masloff, B. Hoff, S. Mayrhofer y U. Kück (2003). Versatile EGFP reporter plasmids for cellular localization of recombinant gene products in filamentous fungi. *Curr Genet*, 43, 54-61.
- Pollak, B., A. Cerda, M. Delmans, S. Álamos, T. Moyano, A. West, R. A. Gutiérrez, N. Patron, F. Federici y J. Haseloff (2018). Loop Assembly, a simple and open system for recursive fabrication of DNA circuits. *bioRxiv*, 1-19.
- Punt, P.J., R. P. Oliver, M. A. Dingemans, P. H. Pouwels y C. A.M.J.J. van den Hondel (1987). Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. *Gene* 56, 117-124.
- Riggle, P. J. y C. A. Kumamoto (1998). Genetic Analysis in Fungi Using Restriction-Enzyme-Mediated Integration. *Current Opinion in Microbiology*, 1, 395-399.
- Robert, X. y P. Gouet (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42, W320–324.
- Robinson, P. K. (2015). *Enzymes: principles and biotechnological applications*. *Essays in Biochemistry*, 59, 1-41.
- Rolli, E., E. Ragni, J. Calderon, S. Porello, U. Fascio y L. Popolo (2009). Immobilization of the Glycosylphosphatidylinositol-anchored Gas1 Protein into the Chitin Ring and Septum Is Required for Proper Morphogenesis in Yeast. Editado por Howard Riezman. *Molecular Biology of the Cell*, 20, 4856-4870.

- Rodriguez-Iglesias y A., M. Schmoll (2015). Protoplast Transformation for Genome Manipulation in Fungi, en: Genetic Transformation Systems in Fungi, Vol 1, van den Berg, M. y K. Maruthachalam (ed). Fungal Biology. Springer Cham. pp. 21-40
- Roussel, A., N. Miled, L. Berti-Dupuis, M. Rivière, S. Spinelli, P. Berna, V. Gruber, R. Verger y C. Cambillau (2002). Crystal Structure of the Open Form of Dog Gastric Lipase in Complex with a Phosphonate Inhibitor. *Journal of Biological Chemistry*, 277, 2266-2274.
- Ruiz-Herrera, J. (1991). Biosynthesis of β -Glucans in Fungi. *Antonie van Leeuwenhoek*, 60, 73-81.
- Ruiz-Herrera, J. y L. Ortiz-Castellanos (2010). Analysis of the phylogenetic relationships and evolution of the cell walls from yeasts and fungi. *FEMS Yeast Research*, 10, 225-243.
- Saitou, N. y M. Nei (1987). The Neighbor-Joining Method, A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, 4, 406-425.
- Sánchez-Carbente, M. R., R. A. Batista-García, A. Sánchez-Reyes, A. Escudero-García, Catalina Morales-Herrera, L. I. Cuervo-Soto, L. French-Pacheco, A. Silva-Fernández, C. Amero, E. Castillo y J. Folch-Mallol (2017). The First Description of a Hormone-Sensitive Lipase from a Basidiomycete, Structural Insights and Biochemical Characterization Revealed *Bjerkandera adusta* Ba EstB as a Novel Esterase. *Microbiology Open*, 6, e00463.
- Schmidt-Dannert, C. (1999). Recombinant Microbial Lipases for Biotechnological Applications. *Bioorganic & Medicinal Chemistry*, 7, 2123-2130.
- Schoffemeer, E. A. M., F. M. Klis, J. H. Sietsma y B. J. C. Cornelissen (1999). The Cell Wall of *Fusarium oxysporum*. *Fungal Genetics and Biology*, 27, 275-282.
- Schuster, A. y M. Schmoll (2010). Biology and biotechnology of *Trichoderma*. *Applied Microbiology and Biotechnology*, 87, 787-799.
- Schütte, M., P. Thullier, T. Pelat, Xenia Wezler, P. Rosenstock, D. Hinz, M. I. Kirsch, M. Hasenberg, R. Frank, T. Schirrmann, M. Gunzer, M. Hust y S. Dübel (2009).

- Identification of a Putative Crf Splice Variant and Generation of Recombinant Antibodies for the Specific Detection of *Aspergillus fumigatus*. PLoS ONE, 4, e6625.
- Selvan, A., C. Seniya, S. N. Chandrasekaran, N. Siddharth, S. Anishetty y G. Pennathur (2010). Molecular Dynamics Simulations of Human and Dog Gastric Lipases, Insights into Domain Movements. FEBS Letters, 584, 4599-4605.
- Sharan, S. K, L. C. Thomason, S. G. Kuznetsov y D. L. Court (2009). Recombineering: A Homologous Recombination-Based Method of Genetic Engineering. Nature protocols, 4, 206-223.
- Sharma, D., B. Sharma y A. K. Shukla (2011). Biotechnological approach of microbial lipase, a review. Biotechnology, 10, 23-40.
- Shen, Q., R. Yang, X. Hua, F. Ye, W. Zhang y W. Zhao (2011). Gelatin-Templated Biomimetic Calcification for β -Galactosidase Immobilization. Process Biochemistry, 46, 1565-1571.
- Silva Castro, N. da, Z. A. Maia, M. Pereira y C. M. de Almeida Soares (2005). Screening for glycosylphosphatidylinositol-anchored proteins in the *Paracoccidioides brasiliensis* transcriptome. Genetics and Molecular Research, 4, 326-345.
- Söding, J., A. Biegert y A. N. Lupas (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research, 33, W244-248.
- Spreghini, E., D. A. Davis, R. Subaran, M. Kim y A. P. Mitchell (2003). Roles of *Candida albicans* Dfg5p and Dcw1p Cell Surface Proteins in Growth and Hypha Formation. Eukaryotic Cell, 2, 746-755.
- Su, J., H. Ouyang, W. Zhao, Z. Dong y C. Jin (2013). Construction of a cell-surface expression system in *Trichoderma reesei*. Wei sheng wu xue bao = Acta microbiologica Sinica, 53, 38-46.

- Tanaka, T., R. Yamada, C. Ogino y A. Kondo (2012). Recent Developments in Yeast Cell Surface Display toward Extended Applications in Biotechnology. *Applied Microbiology and Biotechnology*, 95, 577-591.
- Thomason, L. C., A. B. Oppenheim, y D. L. Court (2009). Modifying Bacteriophage lambda with Recombineering, en *Bacteriophages*, Clokie, M. R.J. y A. M. Kropinski (ed). NJ: Humana Press, Totowa. pp. 239-251.
- Thompson, J D, D G Higgins y T J Gibson (1994). CLUSTAL W, improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673-4680.
- Ueda, M. y A. Tanaka (2000). Genetic immobilization of proteins on the yeast cell surface. *Biotechnology Advances*, 18, 121-140.
- Ülker, S., A. Colak y S. A. Karaoglu (2011). Isolation, production and characterization of an extracellular lipase from *Trichoderma harzianum* isolated from soil. *Turkish Journal of Biology*, 35, 543-550.
- Vaart, J. M. y C. T. Verrips (1998). Cell Wall Proteins of *Saccharomyces cerevisiae*. *Biotechnology and Genetic Engineering Reviews*, 15, 387-412.
- Vaknin, Y., Y. Shadkchan, E. Levdansky, M. Morozov, J. Romano y N. Osherov (2014). The Three *Aspergillus Fumigatus* CFEM-Domain GPI-Anchored Proteins (CfmA-C) Affect Cell-Wall Stability but Do Not Play a Role in Fungal Virulence. *Fungal Genetics and Biology*, 63, 55-64.
- Verger, R. (1997). 'Interfacial Activation' of Lipases, Facts and Artifacts. *Trends in Biotechnology*, 15, 32-38.
- Vink, E., R. J. Rodriguez-Suarez, M. Gérard-Vincent, J. C. Ribas, H. de Nobel, H. van den Ende, A. Durán, F. M. Klis y Bussey J. C. (2004). An *in vitro* assay for (1 → 6)-β-D-glucan synthesis in *Saccharomyces cerevisiae*. *Yeast*, 21, 1121-1131.

- Voigt, C. A., W. Schäfer y S. Salomon (2005). A Secreted Lipase of *Fusarium graminearum* Is a Virulence Factor Required for Infection of Cereals, Lipase as a Virulence Factor. *The Plant Journal*, 42, 364-375.
- Wagner, J. C., S. J. Goldfless, S. M. Ganesan, M. C. S. Lee, D. A. Fidock y J. C. Niles (2013). An integrated strategy for efficient vector construction and multi-gene expression in *Plasmodium falciparum*. *Malaria Journal*, 12, 373.
- White, T. J., Thomas Bruns, SJWT Lee y J. L. Taylor (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols, a guide to methods and applications* 18, 315-322.
- Widmann, Michael, P Benjamin Juhl y Jürgen Pleiss (2010). Structural Classification by the Lipase Engineering Database, A Case Study of *Candida antarctica* Lipase A. *BMC Genomics*, 11, 123.
- Xu, J., C. W. Saunders, P. Hu, R. A. Grant, T. Boekhout, E. E. Kuramae, J. W. Kronstad, Y. M. DeAngelis, N. L. Reeder, K. R. Johnstone, M. Leland, A. M. Fieno, W. M. Begley, Y. Sun, M. P. Lacey, T. Chaudhary, T. Keough, L. Chu, R. Sears, B. Yuan y T. L. Dawson (2007). Dandruff-Associated *Malassezia* Genomes Reveal Convergent and Divergent Virulence Traits Shared with Plant and Human Fungal Pathogens. *Proceedings of the National Academy of Sciences*, 104, 18730-18735.
- Yadav, S, A. K. Dubey, S. Yadav, D. Bisht, N. Singh-Darmwal y D. Yadav (2012). Amino acid sequences based phylogenetic and motif assessment of lipases from different organisms. *Online Journal of Bioinformatics*, 13.
- Yamazaki, H., A. Tanaka, J. Kaneko, A. Ohta y H. Horiuchi (2008). *Aspergillus Nidulans* ChiA Is a Glycosylphosphatidylinositol (GPI)-Anchored Chitinase Specifically Localized at Polarized Growth Sites. *Fungal Genetics and Biology*, 45, 963-972.
- Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson y Y. Zhang (2015). The I-TASSER Suite, protein structure and function prediction. *Nature methods*, 12, 7-8.
- Yin, Q. Y., P. W. J. de Groot, C. G. de Koster y F. M. Klis (2008). Mass spectrometry-based proteomics of fungal wall glycoproteins. *Trends in Microbiology*, 16, 20-26.

- Yoshimi, A., K. Miyazawa y K. Abe (2016). Cell wall structure and biogenesis in *Aspergillus* species. *Bioscience, Biotechnology, and Biochemistry*, 80, 1700-1711.
- Yuzbasheva, E. Y., T. V. Yuzbashev, I. A. Laptev, T. K. Konstantinova y S. P. Sineoky (2011). Efficient Cell Surface Display of Lip2 Lipase Using C-Domains of Glycosylphosphatidylinositol-Anchored Cell Wall Proteins of *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology*, 91, 645-654.
- Zan, X., X. Tang, L. Chu, Lina Zhao, H. Chen, Y. Q. Chen, W. Chen y Y. Song (2016). Lipase Genes in *Mucor circinelloides*, Identification, Sub-Cellular Location, Phylogenetic Analysis and Expression Profiling during Growth and Lipid Accumulation. *Journal of Industrial Microbiology & Biotechnology*, 43, 1467-1480.
- Zhang, S., W. Shang, X. Yang, X. Zhang, Y. Huang, S. Zhang y J. Chen (2014). Immobilization of Lipase with Alginate Hydrogel Beads and the Lipase-Catalyzed Kinetic Resolution of α -Phenyl Ethanol. *Journal of Applied Polymer Science*, 131, 40178.
- Zhang, L., S. Liang, X. Zhou, Z. Jin, F. Jiang, S. Han, S. Zheng y Y. Lin (2013). Screening for Glycosylphosphatidylinositol-Modified Cell Wall Proteins in *Pichia pastoris* and Their Recombinant Expression on the Cell Surface. *Applied and Environmental Microbiology*, 79, 5519-5526.
- Zhao, W., Y. Lu, H. Ouyang, H. Zhou, J. Yan, T. Du y C. Jin (2013). N-Glycosylation of Gel1 or Gel2 Is Vital for Cell Wall β -Glucan Synthesis in *Aspergillus fumigatus*. *Glycobiology*, 23, 955-968.

ANEXO

Anexo 1: Lista de cebadores generados para el aislamiento de las secuencias que codifican las posibles proteínas con anclaje GPI seleccionadas en *Trichoderma longibrachiatum*.

Identificador ^a	Cebador ^b	T _m (°c) ^c	Tamaño amplicón (pb)
1401645	F: 5'- <u>ACTAGT</u> CAGATCCCGGATTGCGCTACC-3' R: <u>GGGCC</u> TTACAAGGCAGCAGCCAGCAC-3'	60.3 60.7	537
1436499	F: 5' <u>ACTAGT</u> TCTTCCAACCTCGCCGCCT-3' R: 5' <u>GGGCC</u> TTAGAGCATGAGGAGGGCAACACC-3'	61.8 60.9	813
2326	F: 5' <u>ACTAGT</u> AAGCCCGAGTTCCTCAACTCC-3' R = <u>GGGCC</u> TTAAGAGAAGAAGGCAATGGCAATGG-3'	59.2 57.7	699
1397533	F: 5' <u>ACTAGT</u> GCTTGCGAGCACTACAGCTTAC-3' R: 5' <u>GGGCC</u> CTACAGCATCATGATGACTCCGAGTG-3'	60.6 60.7	597
1401002	F: 5' <u>ACTAGT</u> AAGACCGACCTCGCCGGCTGCA-3' R: 5' <u>GGGCC</u> TTACAAAAGACCAACGGCCGCAACACCG-3'	67.8 64.9	630
1326581	F: 5' <u>ACTAGT</u> CAGACCTCCAGCTCCGCTGCC-3' R: 5' <u>GGGCC</u> TTACAAAGCCAGCAGGGCAGCAAC-3'	64.9 62.4	564
1325672	F: 5' <u>ACTAGT</u> GTCCTGGGCCAGAGCCTCG-3' R: 5' <u>GGGCC</u> TTACAGAAGAGCAATGACACCGGCG-3'	63.1 61.2	744

^a Las secuencias del ADNc fueron tomados de la base de dato de *T. longibrachiatum* ATCC 18648 v3.0 (<https://genome.jgi.doe.gov/Trilo3/Trilo3.home.html>)

^b Subrayado: sitios de restricción introducidos en la secuencia, para F: Spel y R: Apal.

^c Las T_m de los cebadores fueron calculados con el programa Oligoanalyzer 3.1DT (<http://www.idtdna.com/calc/analyzer>)

Anexo 2.1: Lista de cebadores diseñados para el aislamiento de las secuencias correspondientes al vector de expresión.

SECUENCIA A AISLAR	CEBADORES	TM (°C)	TAMAÑO AMPLICIÓN (PB)
PROMOTOR	F: TCTAGAAGTCCTCGTGTACTGTGTAAGCGCCCACTCCACATCTCCACTCGA CGAATTCCCTTGTATCTCTAC	50	2130
	R: GGTACCGGGAAAAGAAAGAGAAAAGAAAA	49.7	
TERMINADOR	F:GGGCCCCGATCCACTTAACGTTACTGAA	49.8	769
	R:ATCGATTGAGTGGAGATGTGGAG	53.4	
S.S	F: GGTACC ATGCATCTTGCTATCAAGTCTCTC	54.7	179
	R: TGCGGCCGCGGCGCTGTTGGTAGAGCTAG	57.8	
GFP	F: GCGGCCGCAATGGTGAGCAAGGGCGAGGAGC	64.1	723
	R: ACTAGTCTTGTACAGCTCGTCCATGCCGAGA	62.3	
GPI	F: ACTAGTACCCCCCTCGTTGAGCGTGCTGGTTCC	69.5	771
	R: GGGCCCTTAGACAGCGACGGCAATGGCGGCCATGGC	70.4	
PGFP	F: GCCATATTTTCCTGCTCTCCC	55.8	758
	R: GTGGCGGATCTTGAAGTTCACC	58.3	
GFPGPI	F: ATATCATGGCCGACAAGCAG	55.2	840
	R: AGGCAGAACCGGTAGTAGAGG	58	
GPITER	F. GGTCAAGGACTCCCTGAACC	57.4	552
	R: GTCTCAACTCCGGAGCTGAC	57.5	

Anexo 2.2 Lista de los cebadores diseñados para introducir los adaptadores del sistema Gibson®.

Sitio de la sobre posición	Cebador	Secuencia que reconoce	F/R	Tm (°c)
PAN7.1	TTTAGAGGTAATCCTTCTTTCTAGAAGTCCTCGTGTAC	Promotor	F	48.5
SS	GCATGGTACCGGAAAAGAAAGAGAAAAGAAAAG	Promotor	R	50.7
PROMOTOR	TTCTTTTCCCGGTACCATGCATCTTGCTATCAAGTCTCTCTTTGTCTCTCTCC	S.s	F	65
SGFP	TGCTCACCATTGCGGCCGCGGCGCTGTT	S.s	R	69.7
SS	CGCGGCCGCAATGGTGAGCAAGGGCGAG	sGFP	F	57.9
GPI-PROMT	CCGGGATCTGACTAGTCTTGTACAGCTCGTC	sGFP	R	53.6
SGFP	CAAGACTAGTCAGATCCCGGACTGTGCTAC	GPI-prot	F	57.3
TERMINADOR	TTAAGTGGATGGGCCCTTACAGGGCAGC	GPI-prot	R	61.7
GPI-PROMT	GTAAGGGCCCATCCACTTAACGTTACTGAAATC	Term	F	50.8
PAN7.1	TACACAGTACACGAGGACTTCTAGAAAGAAGGATTACCTCTAAAC	Term	R	50.7

Anexo 2.3 Lista de cebadores para la amplificación del gen que codifica para la resistencia a higromicina

Secuencia a aislar	Cebadores	Tm (°c)	Tamaño amplicón (pb)
<i>HPH</i>	F: GACCTGCTGAGGTCCCTC	53	1342
	R: GCCCTCGGACGAGTGCTG	53	

Table 3.1 List of primers used in this study

Protein id in <i>t. Harzianum</i> Database	Model name	Forward and reverse primers	Sequences ^a	Expected size of the product (bp) ^b
77338	e_gw1.3.2527.1	77338F 77338R	CCAGATCACGGAGGAGAATTAC GGGCAGACTTGGAGATGAATA	445
78181	e_gw1.3.156.1	78181F 78181R	CTCCGCAACTTCATCACAGA CAATGGCGAAACCCTGAGTA	241
514252	fgenes1_kg.36_#_14_#_Locus1356v1rpkm114.02	514252F 514252R	CACTGGCTATGAGTGTGTCTATG GGTGTGTCGTCTTGTCTTT	387
526309	estExt_fgenes1_pg.C_1_t10183	526309F 526309R	ATCCACACAACGCAGACTA GCTTGCCGATTAGCCAGAAA	239
514427	fgenes1_kg.83_#_1_#_Locus825v1rpkm186.18	514427F 514427R	GAGTTGCCGTAGCTCCTAGT ATGGTTCGGGAGGTCAAAGT	277
87496	e_gw1.7.1721.1	87496F 87496R	CCTCTCTCGCCGTTAAGGAT TCTTCCGGCTATCTGTCGTC	177
Elongation factor	estExt_Genemark1.C_1_t10440	EF1-983F EF1-1567R	GCY CCY GGH CAY CGT GAY TTY AT ACH GTR CCR ATA CCA CCR ATC TT	400

^a: Degenerate codes: H = A, C, or T, R = A or G, Y = C or T,

^bbp: Base pair

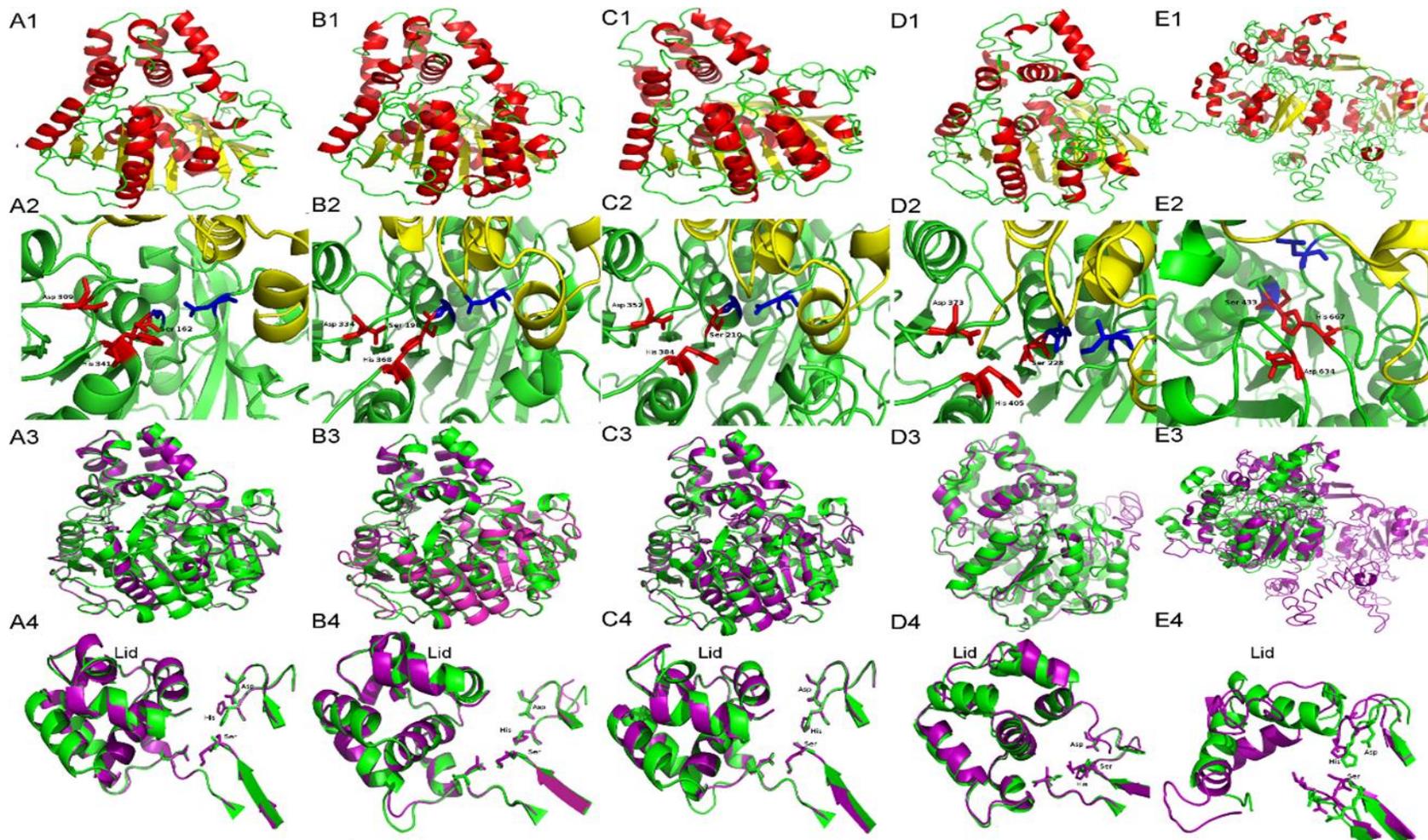
Table S3.2 List of all lipase proteins identified in the *T. harzianum* genome homepage

Protein id	Model name	Chromosome location	Group	Annotations
510832	fgenes1_kg.11_#_179_#_Locus5819v1rpkm1.05	scaffold_11:472451-475242 (-)	Triacylglycerol lipase	Triglyceride lipase-cholesterol esterase
79895	e_gw1.4.2035.1	scaffold_4:1760760-1764647 (+)	Triacylglycerol lipase	Vesicle coat complex COPII, subunit SEC31
21954	estExt_Genemark1.C_420014	scaffold_42:42833-44151 (-)	Acylglycerol lipase	Lysophospholipase
155645	E0stExt_Genewise1.C_250246	scaffold_25:244434-245914 (-)	Acylglycerol lipase	Lysophospholipase
92423	e_gw1.11.791.1	scaffold_11:316020-317968 (+)	Triacylglycerol lipase	Triglyceride lipase-cholesterol esterase
551811	MIX15484_2_17	scaffold_45:15003-16167 (+)	Lipase_secreted	Secretory lipase
87496	e_gw1.7.1721.1	scaffold_7:463722-465074 (-)	Lipase_secreted	Secretory lipase
77338	e_gw1.3.2527.1	scaffold_3:1045025-1046239 (-)	Triacylglycerol lipase	Lipase (class 3)
514252	fgenes1_kg.36_#_14_#_Locus1356v1rpkm11.4.02	scaffold_36:53708-56143 (+)	Triacylglycerol lipase	Lipase (class 3), Putative lipase essential for disintegration of autophagic bodies inside the vacuole
511179	fgenes1_kg.12_#_53_#_Locus7951v1rpkm3.2	scaffold_12:152256-153471 (+)	Acylglycerol lipase	Lysophospholipase
526309	estExt_fgenes1_pg.C_1_t10183	scaffold_1:494922-496430 (-)	Lipase_secreted	Secretory lipase
135964	estExt_Genewise1.C_3_t10273	scaffold_3:330453-331832 (-)	Triacylglycerol lipase	Intrinsic component of endoplasmic reticulum membrane; intracellular protein transport

492160	fgenes1_pm.3_#_685	scaffold_3:1875111-1880526 (-)	Triacylglycerol lipase	Rho GTPase binding, actin cytoskeleton organization
78181	e_gw1.3.156.1	scaffold_3:1634109-1635186 (-)	Triglyceride lipase	Lipase (class 3)
502433	fgenes1_kg.1_#_333_#_Locus3731v1rpkm31.19	scaffold_1:1178847-1182507 (-)	Triglyceride lipase	Lipase (class 3) Predicted lipase/calmodulin-binding heat-shock protein
1117	gm1.1117_g	scaffold_1:3774719-3775867 (+)	Alpha/beta hydrolase	Hormone-sensitive lipase HSL
545835	MIX9508_25_78	scaffold_21:72679-74243 (+)	Lipase_GDXG_HIS_A S, AB_hydrolase	Arylacetamide deacetylase
95431	e_gw1.14.395.1	scaffold_14:945848-948060 (-)	Alpha/beta hydrolase	Hormone-sensitive lipase HSL
183965	CE14116_52606	scaffold_1:1725128-1726866 (+)	Alpha/beta hydrolase	Hormone-sensitive lipase HSL
503030	fgenes1_kg.1_#_930_#_Locus4686v1rpkm20.07	scaffold_1:2789190-2793166 (+)	Predicted alpha/beta hydrolase	DUF676_lipase-like
525479	estExt_fgenes1_pm.C_260025	scaffold_26:94511-96195 (-)	GDSL-like Lipase/Acylhydrolase	Hydrolase activity on ester bonds
554838	MIX18511_21_34	scaffold_7:437392-440685 (-)	GDSL-like Lipase/Acylhydrolase	Carbohydrate Esterase Family 3
43373	gw1.10.153.1	scaffold_10:997206-998495 (+)	GDSL-like Lipase/Acylhydrolase	hydrolase activity on ester bonds
72612	e_gw1.1.2278.1	scaffold_1:1349400-1350411 (-)	GDSL-like Lipase/Acylhydrolase	Carbohydrate Esterase Family 3
92434	e_gw1.11.353.1	scaffold_11:106578-1067065 (-)	GDSL-like Lipase/Acylhydrolase	Carbohydrate Esterase Family 3
509409	fgenes1_kg.8_#_31_#_Locus8614v1rpkm2.33	scaffold_8:113072-113949 (-)	GDSL-like Lipase/Acylhydrolase	Isoamyl acetate-hydrolyzing esterase

510248	fgenes1_kg.10_#_8_#_Locus10596v1rpkm1.04	scaffold_10:53541-54634 (+)	GDSL-like Lipase/Acylhydrolase	Carbohydrate Esterase Family 3
516582	estExt_fgenes1_pm.C_2_t30018	scaffold_2:2892227-2893894 (+)	GDSL-like Lipase/Acylhydrolase	Isoamyl acetate-hydrolyzing esterase
85395	e_gw1.6.337.1	scaffold_6:679883-681325 (-)	Predicted alpha/beta hydrolase	DUF676_lipase-like
100838	e_gw1.29.72.1	scaffold_29:106446-107981 (+)	Lipase_GDXG_HIS_A S, AB_hydrolase	Arylacetamide deacetylase
118553	estExt_Genewise1Plus.C_8_t10171	scaffold_8:191543-195600 (-)	GDSL-like Lipase/Acylhydrolase	Carbohydrate Esterase Family 3
157348	estExt_Genewise1.C_490011	scaffold_49:20234-23599 (+)	GDSL-like Lipase/Acylhydrolase	Carbohydrate Esterase Family 3
514427	fgenes1_kg.83_#_1_#_Locus825v1rpkm186.18	scaffold_83:2335-3893 (+)	Lipase_secreted	Secretory lipase; iron ion transport
77612	e_gw1.3.1367.1	scaffold_3:1583837-1587250 (-)	Phospholipase C	Phosphoinositide-specific phospholipase C
222426	CE52577_10851	scaffold_11:194867-198458 (+)	Phospholipase C	Phosphoinositide-specific phospholipase C
482782	fgenes1_pg.7_#_388	scaffold_7:1132150-1133895 (+)	Phospholipase C	Phosphoinositide-specific phospholipase C
490471	fgenes1_pm.2_#_392	scaffold_2:1112245-1114326 (-)	Phospholipase C	Phosphoinositide-specific phospholipase C
536904	MIX577_51_94	scaffold_1:739432-741727 (-)	Lipase_GDXG_HIS_A S, CarbesteraseB	Carboxylesterase family
91073	e_gw1.10.999.1	scaffold_10:814005-815072 (-)	Phospholipase C	Glycosylphosphatidylinositol-specific phospholipase C
481511	fgenes1_pg.5_#_682	scaffold_5:2279112-2280468 (+)	Phospholipase C	PLipase_C_PInositol
504276	fgenes1_kg.2_#_820_#_Locus1919v1rpkm78.8	scaffold_2:2472619-2474504 (+)	Phospholipase C	PLipase_C_PInositol

509413	fgenesh1_kg.8_#_35_#_Locus4260v1rpkm24.39	scaffold_8:129174-131043 (+)	Phospholipase C	Glycosylphosphatidylinositol-specific phospholipase C
72504	e_gw1.1.1.3340.1	scaffold_1:3946901-3949351 (-)	Acyl_Trfase/lysoPLipase	Lysophospholipase
350643	CE180794_97	scaffold_3:1591594-1592361 (-)	Phospholipase A2	Phospholipid metabolic process
443708	CE273859_156	scaffold_7:624825-625362 (-)	Phospholipase A2	Phospholipid metabolic process
449569	CE279720_76	scaffold_7:1520373-1520871 (-)	Phospholipase A2	Phospholipid metabolic process
549088	MIX12761_545_96	scaffold_3:3077745-3079308 (+)	Phosphatidylinositol phospholipase	Lipid metabolic process, oxidoreductase activity
77540	e_gw1.3.2811.1	scaffold_3:1671663-1673855 (+)	Sphingomyelin phosphodiesterase	Acid sphingomyelinase and PHM5 phosphate metabolism
93268	e_gw1.12.65.1	scaffold_12:840535-842776 (-)	Acyl_Trfase/lysoPLipase	Lysophospholipase
514118	fgenesh1_kg.31_#_48_#_Locus5221v1rpkm15.36	scaffold_31:139768-142584 (+)	Acyl_Trfase/lysoPLipase	Lysophospholipase



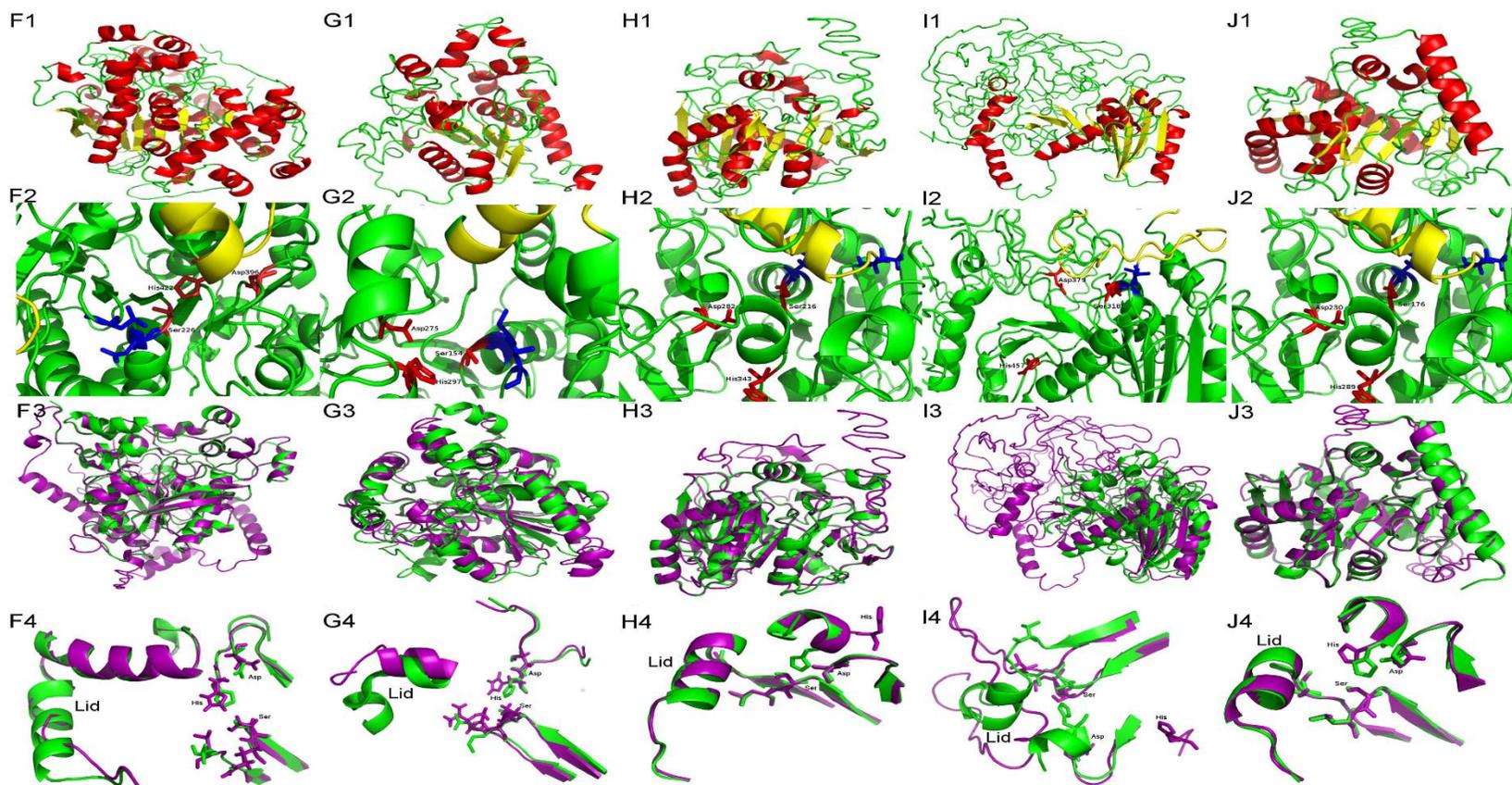


Figure S3.1 Three-dimensional model proposed for (A) 551811; (B) 87496; (C) 526309; (D) 514427; (E) 510832; (F) 92423; (G) 135964; (H) 77338; (I) 514252; (J) 78181. In all protein models: (1) Red, α -helices; yellow, β -strands. (2) Close up showing the lid (yellow), the catalytic triad (in red) and the oxyanion (in blue). (3) Superposition with best template protein model (see Table 4). (4) Superposition of catalytic triad in both proteins. All models were generated by I-TASSER and visualizations were performed in PyMOL.