



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERIA QUÍMICA Y BIOTECNOLOGÍA

# **IDENTIFICACIÓN Y CLONACIÓN DE NUEVAS CELULASAS PARA LA PRODUCCIÓN DE BIOETANOL DE SEGUNDA GENERACIÓN**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL EN BIOTECNOLOGÍA

**ALEJANDRO ANDRÉS SALINAS VACCARO**

PROFESOR GUÍA:  
ORIANA SALAZAR AGUIRRE

MIEMBROS DE LA COMISIÓN:  
MARÍA ELENA LIENQUEO CONTRERAS  
JUAN ASENJO DE LEUZE

SANTIAGO DE CHILE  
OCTUBRE 2009

## Resumen

Debido a la alta demanda de combustibles y la creciente alza en el precio del petróleo es primordial buscar nuevas alternativas energéticas para satisfacer las necesidades de la población. Dentro de este escenario, considerando la abundancia de residuos lignocelulósicos forestales y agroindustriales en nuestro país, la posibilidad de producir bioetanol a partir de éstos parece ser una alternativa bastante atractiva en los aspectos estratégicos y ambientales.

La hidrólisis de la celulosa es una de las etapas que tiene mayores posibilidades de ser optimizada con el fin de reducir los costos de producción del bioetanol de segunda generación. Ésta se basa en la utilización de una mezcla de celulasas que causan la depolimerización de la celulosa con el fin de generar azúcares simples (glucosa). Considerando lo anterior, el presente Trabajo de Memoria de Título, tuvo como objetivo la identificación de secuencias pertenecientes a nuevas endoglucanasas que puedan ser utilizadas en el proceso de hidrólisis de celulosa.

La estrategia utilizada se dividió en tres etapas principales: (i) la realización de un *screening* de actividad endoglucanasa sobre un conjunto de hongos de pudrición blanca, para identificar cuales poseían enzimas con esta actividad; (ii) la amplificación por PCR desde DNA genómico y cDNA de fragmentos pertenecientes a secuencias que codifican para estas enzimas, utilizando partidores degenerados diseñados a partir de regiones de homología de endoglucanasas de hongos; y (iii) la amplificación de regiones aledañas a las secuencias obtenidas utilizando la estrategia RACE (*Rapid Amplification of cDNA Ends*).

Mediante la utilización de esta metodología, fue posible obtener fragmentos de secuencias de genes de glicosil hidrolasas, los cuales fueron clonados y analizados *in silico*, obteniéndose los siguientes resultados: a partir del cDNA se obtuvieron fragmentos de cuatro endoglucanasas y una exoglucanasa. Una de las secuencias de endoglucanasa pertenece a la familia 5 y procede del hongo *Polystictus versicolor*; la segunda pertenece a la familia 7 y procede de *Fusarium oxysporum*; una tercera pertenece a la familia 61 y procede del hongo *Ganoderma applanatum*, mientras que la cuarta pertenece a la familia 61 y procede del hongo *Fusarium oxysporum*. Las secuencias se compararon con secuencias de endoglucanasas presentes en las bases de datos, encontrándose que poseen un porcentaje de identidad a nivel de secuencia aminoacídica de 93, 100, 58 y 60 % con respecto al mejor alineamiento observado, respectivamente. La secuencia de exoglucanasa proviene del hongo *Polystictus versicolor* y se clasifica dentro de la familia 7, con un 80% de identidad aminoacídica con respecto al mejor alineamiento observado. De todas las endoglucanasas identificadas en este trabajo, sólo la proveniente del hongo *Fusarium oxysporum* se encontraba previamente descrita. A partir de DNA genómico, no fue posible obtener fragmentos con utilidad para el proceso de hidrólisis de celulosa.

En conclusión, la estrategia utilizada fue exitosa, siendo posible obtener fragmentos de secuencia de tres endoglucanasas de hongos no descritas, que pueden ser potencialmente útiles en el proceso de hidrólisis de celulosa para la producción de bioetanol de segunda generación.

# Agradecimientos

A mis padres, Sergio y Gabriela, por su amor y apoyo incondicional, y a mis hermanas, Gabriela y Macarena, por su compañía, paciencia y comprensión.

A mi profesora guía Oriana Salazar, por su entusiasmo inquebrantable y estar siempre ahí cuando la necesité.

A mis amigos Alejandra, Marcela, Sergio, Maricelle, Roberto, Adriana, Rubén, Álvaro, Gabriel, Iván y muchos otros, por entregarme momentos inolvidables, que hicieron esta etapa de mi vida una de las mejores.

Finalmente, agradezco el apoyo del Programa Domeyko (Optimización del Proceso de Tratamiento de Lignocelulosas para obtener Bioetanol) y el Proyecto de Investigación Conjunta en Energías Renovables no Convencionales Chile/Finlandia (Procesos óptimos para el tratamiento de materiales lignocelulósicos para la producción de bioetanol), los cuales hicieron posible la realización de esta memoria.

# Índice de contenidos

Resumen.....	ii
Agradecimientos.....	iii
Índice de contenidos.....	iv
Índice de figuras.....	viii
Índice de tablas.....	x
Abreviaturas.....	xi
1. Introducción.....	1
1.1 Marco teórico.....	1
1.1.1 Antecedentes generales.....	1
1.1.2 Bioetanol de segunda generación como combustible.....	2
1.1.2.1 Ventajas del etanol por sobre el combustible fósil.....	3
1.1.2.2 Ventajas del etanol de segunda generación por sobre del de primera generación.....	4
1.1.2.3 Materia prima para la producción de bioetanol de segunda generación.....	5
1.1.3 Producción de bioetanol de segunda generación.....	6
1.1.4 Mecanismo de hidrólisis enzimática de celulosa.....	10
1.1.5 Microorganismos productores de celulasas.....	12
1.1.6 Regulación de la expresión de celulasas en hongos.....	12
1.1.7 Clasificación de glicosil hidrolasas.....	13
1.1.8 Dominios conservados.....	14
1.2 Descripción y justificación del proyecto.....	14
1.3 Objetivos.....	16
1.3.1 Objetivo General.....	16
1.3.2 Objetivos Específicos.....	16
2. Materiales y métodos.....	17

2.1	Materiales .....	17
2.1.1	Reactivos .....	17
2.1.2	Equipos.....	17
2.1.3	Herramientas computacionales.....	18
2.1.4	Enzimas de restricción .....	18
2.1.5	Vector de clonación.....	19
2.2	Métodos .....	19
2.2.1	Construcción de la micoteca.....	19
2.2.2	Ensayos de actividad endoglucanasa.....	21
2.2.2.1	Utilizando cultivos en medio sólido .....	21
2.2.2.2	Utilizando cultivos en medio líquido .....	21
2.2.2.3	Utilizando cultivos en medios inductores.....	22
2.2.3	Diseño de partidores .....	22
2.2.4	Extracción de DNA genómico desde hongos .....	23
2.2.5	Amplificación de fragmentos por PCR a partir de DNA genómico .....	24
2.2.6	Extracción de RNA desde hongos.....	24
2.2.7	Generación de cDNA a partir del RNA.....	25
2.2.8	Amplificación de fragmentos por PCR a partir de cDNA .....	25
2.2.9	RACE (Rapid Amplification of cDNA Ends).....	26
2.2.9.1	Amplificación del extremo 5' .....	26
2.2.9.2	Amplificación del extremo 3' .....	27
2.2.10	Secuenciamiento y análisis .....	29
3.	Resultados y Discusión .....	30
3.1	Construcción de la micoteca .....	30
3.2	Identificación de hongos productores de endoglucanasas .....	30
3.3	Diseño de partidores degenerados. ....	32
3.4	Amplificación de fragmentos mediante PCR desde DNA genómico .....	35
3.4.1	Análisis de fragmentos clonados provenientes desde DNA genómico.....	40

3.5	Aislamiento de genes de endoglucanasas a partir de RNA/cDNA .....	42
3.5.1	Selección de medio inductor y segunda selección. ....	43
3.5.2	Amplificación de fragmentos mediante PCR desde cDNA .....	44
3.5.3	Secuenciamiento y análisis de fragmentos provenientes desde cDNA. ....	47
3.6	Obtención de secuencias aledañas a los fragmentos seleccionados. ....	55
3.6.1	Amplificación del extremo 5' de los genes de endoglucanasas.....	55
3.6.2	Amplificación del extremo 3' de los genes de endoglucanasas.....	57
3.7	Secuenciamiento y análisis de secuencias obtenidas por RACE .....	59
3.8	Discusión final.....	62
4.	Conclusiones .....	67
5.	Bibliografía.....	68
6.	Anexos.....	73
6.1	Metodología adicional .....	73
6.1.1	Medios de cultivo .....	73
6.1.1.1	Medios sólidos.....	73
6.1.1.2	Medios líquidos.....	73
6.1.2	Determinación de azúcares reductores mediante el método DNS.....	74
6.1.3	Curva de calibración DNS .....	74
6.1.4	Ensayos de actividad en geles de agarosa –CMC .....	74
6.1.5	Alfabeto degenerado .....	75
6.1.6	Verificación de extracción de DNA .....	75
6.1.7	Purificación de fragmentos de DNA desde gel de agarosa.....	75
6.1.8	Ligación de fragmentos de DNA al vector pGEM®-T Easy.....	76
6.1.9	Transformación de células <i>E.coli</i> DH5α.....	76
6.1.10	Extracción del DNA plasmidial desde las células <i>E.coli</i> DH5α transformadas. ....	76
6.1.11	Verificación de presencia de inserto en el vector pGEM®-T Easy.....	76
6.1.12	Extracción de proteínas utilizando buffer de extracción .....	77

6.1.13	Tratamiento contra RNAsas utilizando DEPC.....	77
6.1.14	Tratamiento del RNA con DNAsas .....	77
6.1.15	Secuencias de partidores gen-específico adicionales para RACE .....	77
6.1.16	Secuencia de partidores utilizados para secuenciar .....	77
6.2	Resultados adicionales.....	79
6.2.1	Datos de ensayos de actividad en medio líquido.....	79
6.2.2	Alineamientos de secuencias aminoacídicas adicionales.....	80
6.2.3	Secuencias.....	82
6.2.3.1	Provenientes desde DNA genómico .....	82
6.2.3.2	Provenientes desde cDNA .....	82
6.2.3.3	Provenientes desde RACE .....	83

## Índice de figuras

Figura 1.1 Diagrama del proceso de producción de bioetanol de segunda generación.....	7
Figura 1.2 Distintas configuraciones para el proceso de conversión de biomasa a etanol.....	10
Figura 1.3 Representación esquemática de hidrólisis enzimática de celulosa por celulasas .....	11
Figura 2.1 Mapa del vector pGEM®-T Easy y sus puntos de referencia.....	19
Figura 2.2 Diagrama de la estrategia utilizada para RACE extremo 5' .....	27
Figura 2.3 Diagrama de la estrategia utilizada para RACE extremo 3' .....	28
Figura 3.1 Gráfico de máximos de actividad endoglucanasa sobre CMC para los distintos hongos ....	31
Figura 3.2 Ensayo de actividad endoglucanasa sobre medio sólido .....	32
Figura 3.3 Alineamiento de secuencias de endoglucanasas pertenecientes a la familia 5 de GHs .....	34
Figura 3.4 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 5 de GHs.....	37
Figura 3.5 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 7 de GHs.....	37
Figura 3.6 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 61 de GHs.....	38
Figura 3.7 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 5 de GHs con DMSO.....	38
Figura 3.8 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 12 de GHs con DMSO .....	39
Figura 3.9 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 61 de GHs con DMSO .....	39
Figura 3.10 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f5h162B-1.....	41
Figura 3.11 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f7h162A-2.....	41
Figura 3.12 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f61h110B-2 .....	42
Figura 3.13 Selección de medio inductor de endoglucanasas para los distintos hongos .....	43
Figura 3.14 Amplificación de fragmentos de cDNA utilizando partidores degenerados para la familia 5 de GHs.....	45



Figura 3.15 Amplificación de fragmentos de cDNA utilizando partidores degenerados para la familia 7 de GHs.....	46
Figura 3.16 Amplificación de fragmentos de cDNA utilizando partidores degenerados para la familia 12 de GHs .....	46
Figura 3.17 Amplificación de fragmentos de cDNA utilizando partidores degenerados para la familia 61 de GHs .....	47
Figura 3.18 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f5h106-1.....	48
Figura 3.19 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f7h106-1.....	48
Figura 3.20 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f7h163-1.....	49
Figura 3.21 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f12h117-1 .....	49
Figura 3.22 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f61h117-1 .....	50
Figura 3.23 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f61h163-1 .....	50
Figura 3.24 Alineamiento del fragmento f7h106-1 con secuencia de exoglucanasa de la familia 7....	51
Figura 3.25 Análisis filogenético de los fragmentos de endoglucanasas encontrados.....	53
Figura 3.26 Posición del dominio CBD en secuencias de endoglucanasas de la familia 5 de GHs .....	54
Figura 3.27 RACE extremo 5' utilizando partidore específico 106EN5Rev.....	56
Figura 3.28 RACE extremo 5' utilizando partidore específico 117EN61Rev .....	57
Figura 3.29 RACE extremo 3' utilizando partidore específico 106EN5Fwd.....	58
Figura 3.30 RACE extremo 3' utilizando partidore específico 117En61Fwd.....	59
Figura 3.31 Resultado de búsqueda de dominios conservados para la secuencia f61h117-5' .....	60
Figura 3.32 Resultado de búsqueda de dominios conservados para la secuencia117EN61.....	61
Figura 3.33 Diagrama estrategia 5' modificada .....	64
Figura 3.34 Diagrama estrategia RACE 3' modificada.....	64
Figura 6.1 Curva de calibración DNS .....	74
Figura 6.2 Alineamiento de endoglucanasas pertenecientes a la familia 7 de GHs.....	80
Figura 6.3 Alineamiento de endoglucanasas pertenecientes a la familia 12 de GHs.....	81
Figura 6.4 Alineamiento de endoglucanasas pertenecientes a la familia 61 de GHs.....	81

# Índice de tablas

Tabla 1.1 Contenido de celulosa, hemicelulosa y lignina en residuos lignocelulósicos comunes. ....	6
Tabla 1.2 Condición de salida del material lignocelulósico en distintos procesos de pretratamiento .	8
Tabla 2.1 Reactivos utilizados en el trabajo de laboratorio. ....	17
Tabla 2.2 Equipos utilizados en el trabajo de laboratorio. ....	18
Tabla 2.3 Herramientas computacionales utilizadas. ....	18
Tabla 2.4 Enzimas de restricción utilizadas durante el trabajo. ....	18
Tabla 2.5 Micoteca de trabajo. ....	20
Tabla 2.6 Partidores degenerados utilizados. ....	23
Tabla 2.7 Partidores utilizados para realizar RACE en las secuencias seleccionadas. ....	23
Tabla 3.1 Largo esperado de amplicones para las distintas familias de GHs. ....	35
Tabla 3.2 Resumen de fragmentos amplificados desde DNA genómico de los hongos. ....	36
Tabla 3.3 Resumen de resultados obtenidos para fragmentos encontrados en el DNA genómico. ....	42
Tabla 3.4 Medios de inducción seleccionados para los distintos hongos. ....	44
Tabla 3.5 Resumen de fragmenos obtenidos desde cDNA. ....	45
Tabla 3.6 Resumen de resultados obtenidos para fragmentos encontrados en el cDNA. ....	50
Tabla 3.7 Porcentaje de la secuencia aminoacídica de las endoglucanasas encontradas. ....	54
Tabla 3.8 Porcentaje de identidad y similitud entre fragmentos y enzimas comerciales. ....	55
Tabla 3.9 Resumen bandas obtenidas mediante estrategia RACE. ....	58
Tabla 6.1 Alfabeto degenerado. ....	75
Tabla 6.2 Partidores para gen 18S rRNA de hongos. ....	75
Tabla 6.3 Secuencias de partidores adicionales para RACE. ....	77
Tabla 6.4 Partidores utilizados para la secuenciación. ....	78
Tabla 6.5 Datos de ensayos de actividad en medio líquido. ....	79

## Abreviaturas

Abreviatura	Significado
$\mu$ F	Microfaraday
aa	Aminoácidos
ATP	Adenosín trifosfato
BSA	Albumina de suero bovino
CDD	Conserved Domain Database
cDNA	DNA complementario
CMC	Carboximetil celulosa
dCTP	2-deoxicitidina-5-trifosfato
DEPC	Dietil pirocarbonato
DMSO	Dimetil sulfóxido
DNA	Ácido desoxirribonucleico
DNAsa	Desoxirribonucleasa
DNS	Ácido dinitrosalicílico
dNTPs	Desoxirribonucleótidos trifosfato
EDTA	Etilendiaminotetraacetato
FCFM	Facultad de Ciencias Físicas y Matemáticas
GHs	Glicosil hidrolasas
IPTG	Isopropil-tio- $\beta$ -D-galactopiranosido
Kb	Kilobase
LB	Luria Bertani
mRNA	RNA mensajero
NCBI	National Center for Biotechnology Information
pb	Pares de bases
PCR	Reacción en cadena de polimerasa
PMSF	Fluoruro de fenilmetil sulfonilo
RACE	Rapid Amplification of cDNA Ends
RNA	Ácido Ribonucleico
RNAsa	Ribonucleasa
RPM	Revoluciones por minuto
RT	Transcriptasa reversa
SDS	Dodecil sulfato de sodio
TdT	Desoxinucleotidil-transferasa terminal
Tm	Temperatura de <i>melting</i>
U	Unidad de actividad enzimática
U <sub>abs</sub>	Unidades de absorbancia
X-Gal	5-bromo-4-cloro-3-indolil-beta-D-galactopiranosido

# 1. Introducción

## 1.1 Marco teórico

### 1.1.1 Antecedentes generales

En el último siglo, debido al aumento de la población mundial y que cada vez más países se han industrializado, los requerimientos energéticos se han ido incrementado en forma sostenida. Los combustibles fósiles, como el petróleo, son la mayor fuente energética para satisfacer esta demanda. Sin embargo, las reservas de este combustible deberían alcanzar para proveer de éste sólo durante 70 años más, generando preocupación, tanto por la repercusión en el costo que esto significa, como por la inseguridad generada en cuanto la disponibilidad de este recurso [1]. En el caso de Chile, aproximadamente un 65% del suministro energético proviene de combustibles fósiles no renovables, siendo gran parte de éstos importados [2].

Por otro lado, la utilización de combustibles derivados del petróleo genera altas emisiones de gases contaminantes, los cuales provocan fenómenos muy nocivos para la vida en general, como el conocido efecto invernadero.

Debido a las razones expuestas, nace la necesidad de obtener energía desde otras fuentes, que sean renovables y menos contaminantes. En particular para Chile, considerando la dependencia energética que presenta, es un tema de suma importancia.

Una posible solución a estos problemas es la utilización de biocombustibles, presentando ventajas en cuanto el impacto ambiental y a la naturaleza renovable de las materias primas que se utilizan para su producción [3]. Dentro de éstos, el bioetanol es una interesante alternativa, considerando que es una fuente de energía limpia y renovable, además de ser relativamente fácil de producir y almacenar.

Usualmente, este biocombustible es producido a partir de la fermentación de productos agrícolas de uso alimenticio (maíz, trigo, remolacha, caña de azúcar, entre otros), siendo denominado como bioetanol de primera generación. Pese a las ventajas que presenta, el costo de la materia prima para producirlo ha aumentado en forma ostensible, dificultando su competitividad en cuanto a precio con el combustible fósil. En efecto, se ha estimado que en el mejor de los escenarios, produciendo el biocombustible a partir de maíz, el precio del bioetanol en Chile sería un 85% más alto que el del petróleo [4]. Además, la producción de bioetanol de primera generación, genera una competencia

por suelos entre las plantaciones con fines energéticos y las con fines alimenticios, siendo ésta particularmente crítica en países como Chile, en que la cantidad de suelos que se dispone es muy limitada [5]. Debido a las razones antes presentadas, se han estudiado nuevas fuentes de materia prima para la producción de este biocombustible, como lo son los residuos lignocelulósicos. A los biocombustibles producidos a partir de estas materias primas, se les han llamado biocombustibles de segunda generación [3]. Considerando la abundancia de residuos lignocelulósicos forestales y agroindustriales en nuestro país, se plantea aprovechar éstos para la producción de bioetanol.

Se sabe que la clave para aumentar el rendimiento del bioetanol a partir de residuos lignocelulósicos y a su vez producirlo a precios competitivos, se encuentra en la optimización de las siguientes etapas: pre-tratamiento de residuos, hidrólisis de celulosa y fermentación de azúcares simples, representando el 23, 53,4 y 23,6 % del costo total del proceso respectivamente [6].

En particular, la hidrólisis de la celulosa consiste en la utilización de una mezcla de enzimas celulasas con el fin de causar la despolimerización de los carbohidratos complejos presentes luego del pre-tratamiento, tales como la celulosa y hemicelulosa, llevándolos a azúcares simples fácilmente fermentables [7]. Considerando que las celulasas son enzimas relativamente costosas [8], es fundamental la disminución de los costos relacionados con éstas para poder tener un proceso económicamente viable. Todo lo anterior lleva a la necesidad de producir enzimas en forma abundante y con alta actividad específica.

Considerando los antecedentes presentados, la presente investigación se centró en la identificación y clonación de nuevas enzimas potencialmente útiles para la producción de bioetanol de segunda generación.

### **1.1.2 Bioetanol de segunda generación como combustible**

El interés por combustibles alternativos para el transporte, ha ido en aumento por dos razones principales: la inseguridad causada por el inminente agotamiento de las reservas de petróleo y la necesidad de disminuir las emisiones de dióxido de carbono causadas al utilizar combustibles fósiles, con el fin de evitar los efectos adversos del cambio climático global [9]. Como un paso para resolver estos problemas, la adición de etanol a la gasolina ha sido ampliamente aplicada en los últimos años [10].

El 9 de Mayo de 2008 se publicó en el diario oficial un decreto supremo en el que se define lo que se entenderá por bioetanol en Chile, estableciéndose que éste sólo se podrá mezclar con gasolina

automotriz para uso en motores de ignición por chispa, en un 2% o en un 5% del volumen total de la mezcla [11].

Pese a que en la actualidad no se encuentra vigente ninguna ley que obligue el uso de biocombustibles, se tiene como antecedente un proyecto de ley sobre el fomento de las energías renovables y combustibles líquidos (Boletín N° 4.873-08). Éste establece que la gasolina utilizada en el sector transporte que se expenda al público, deberá contener un porcentaje mínimo obligatorio de biocombustible, decretándose que para el 2013 éste sea de un 5% [12]. Considerando lo anterior, se ha estimado que para el 2013 la demanda de gasolina en Chile será de 3.475.000 m<sup>3</sup>, por lo que para satisfacer el 5% de sustitución sería necesaria la producción de 174.000 m<sup>3</sup> de etanol [13].

#### **1.1.2.1 Ventajas del etanol por sobre el combustible fósil**

Las principales ventajas del bioetanol en comparación a los combustibles fósiles se presentan a continuación.

##### **Rendimiento:**

El etanol presenta un mayor octanaje, habilidad para resistir presión, permitiendo que el motor funcione a una mayor tasa de compresión, obteniendo un rendimiento mayor [10]. Además, la presión de vapor del etanol es mayor y su calor de vaporización es más alto que el de la gasolina, lo que es responsable del aumento de potencia al usar etanol. Sin embargo, debido a su contenido de oxígeno un galón de bioetanol contiene menos energía que uno de gasolina [10].

##### **Impacto ambiental:**

La extracción, procesamiento y combustión de combustibles fósiles generan contaminación del aire, agua y suelo, siendo peligrosos para el ambiente y la salud pública. Al utilizar bioetanol, los niveles netos de CO<sub>2</sub> se reducen en un rango entre 60 y 90%, en comparación a la utilización de gasolina. No obstante, la utilización de bioetanol tiene la desventaja de liberar entre 2 y 4 veces más aldehídos que la gasolina [10].

##### **Disponibilidad:**

El bioetanol es una fuente renovable de energía, por el contrario, se espera que los suministros de combustibles fósiles, al ser una fuente no renovable, se agoten en algún punto del siglo XXI [10].

**Costos:**

Los costos directos asociados a la producción de bioetanol son mayores que los asociados a la producción de gasolina [13]. Sin embargo, no se han considerado los costos indirectos como son el daño ambiental y la protección de los pozos petroleros en Medio Oriente [10].

**1.1.2.2 Ventajas del etanol de segunda generación por sobre el de primera generación**

Por otro lado, la utilización de bioetanol de segunda generación exhibe ventajas comparativas en cuanto a la utilización de bioetanol de primera generación, como las descritas a continuación.

**Competencia por los suelos:**

El bioetanol de primera generación proviene de cultivos utilizados con fines alimenticios, como por ejemplo, el caso del trigo en Estados Unidos y la caña de azúcar en Brasil. Considerando la creciente demanda de estos productos agrícolas para atender la demanda mundial de producción de este biocombustible, existe el temor de que alimentos esenciales aumenten sus precios enormemente [3]. Por otro lado, en países como Chile, la disponibilidad de suelos para cultivos es limitada, lo que genera una competencia entre su utilización para fines alimenticios y energéticos [5]. En el caso del bioetanol de segunda generación la intensificación agrícola no es necesaria, por lo que no existe disputa por los suelos.

**Impacto ambiental:**

Considerando que la materia prima para el bioetanol de primera generación se obtiene a partir de productos agrícolas, son necesarias grandes zonas de cultivo para dar cabida a la creciente demanda por combustible. Esto último ha generado la necesidad de sacrificar bosques tropicales, ricos en biodiversidad, para dar lugar a cultivos con fines energéticos [3]. En consecuencia, la estabilidad ecológica se ve gravemente afectada, contribuyendo en gran medida a la contaminación del medio ambiente [3]. En el caso de biocombustibles de segunda generación, al utilizarse en gran medida materiales que son considerados desechos como materia prima, no ocurre este daño ambiental.

**Costos:**

Los costos de la materia prima para producir el bioetanol son menores en el caso del bioetanol de segunda generación (principalmente desechos). Sin embargo, existen mayores costos de producción para este tipo de combustible, los cuales están asociados principalmente a las etapas de pre-tratamiento e hidrólisis de la celulosa [3].

### **1.1.2.3 Materia prima para la producción de bioetanol de segunda generación**

La biomasa lignocelulósica está compuesta principalmente por tres polímeros: celulosa, hemicelulosa y lignina, los cuales se describen a continuación.

#### **Celulosa:**

La celulosa es un homopolisacárido que está compuesto por monómeros de glucosa (D-glucopiranosas) unidos a través de enlaces  $\beta$ -1,4-glucosídicos. Este polisacárido se encuentra presente en la naturaleza tanto en plantas primitivas como en altamente evolucionadas. Análisis conformacionales indican que la celobiosa (4-O- $\beta$ -D-glucopiranosil- $\beta$ -D-glucopiranosas) es su base estructural y no la glucosa [14].

#### **Hemicelulosa:**

La hemicelulosa es un heteropolisacárido cuya composición química varía entre tejidos y especies. Este polisacárido está formado por una gran variedad de bloques de construcción, incluyendo pentosas, hexosas y ácidos urónicos. A pesar de su complejidad, suele formar una estructura en forma de varilla que facilita su interacción con la celulosa, generando una estrecha asociación que le da gran estabilidad al agregado [14].

#### **Lignina:**

La lignina es una macromolécula aromática de naturaleza fenólica. Es el producto de deshidratación de tres alcoholes monoméricos (lignoles) derivados del ácido *p*-cinámico, que se encuentra incrustada en las paredes celulares y mantiene las células unidas [10]. La propiedad física más importante de esta macromolécula orgánica es la rigidez, no sólo dándole resistencia mecánica a los tejidos de la planta, sino también evita el colapso de los elementos para transportar agua, nutrientes y metabolitos [14].

Exceptuando la lignina, estos polisacáridos de cadena larga pueden ser hidrolizados para producir una mezcla de pentosas (C<sub>5</sub>) y hexosas (C<sub>6</sub>) [3].

Las fuentes de biomasa lignocelulósica para la producción de bioetanol, se pueden dividir en tres grandes categorías: residuos forestales, los cuales incluyen maderas y pajas provenientes de la industria de la pulpa y papel y actividades de tala; cultivos agrícolas, que incluyen desechos de la agroindustria y pastos perennes; desechos provenientes de la agroindustria; y desechos secundarios, tales como desechos sólidos municipales y desechos del procesamiento de alimentos. Todas las fuentes de materia prima antes mencionadas son ricas en lignocelulosa y tienen potencial para ser utilizadas para la producción de bioetanol de segunda generación [3].



En la Tabla 1.1 se presenta el contenido de celulosa, hemicelulosa y lignina de los residuos lignocelulósicos más comunes [15].

**Tabla 1.1 Contenido de celulosa, hemicelulosa y lignina en residuos lignocelulósicos comunes.**

<b>Material lignocelulósico</b>	<b>Celulosa [%]</b>	<b>Hemicelulosa[%]</b>	<b>Lignina[%]</b>
Tallos de madera dura	40-55	24-40	18-25
Tallos de madera blanda	45-50	25-35	25-35
Cáscaras de nueces	25-30	25-30	30-40
Mazorcas de maíz	45	35	15
Pastos	25-40	35-50	10-30
Papel	85-99	0	0-15
Paja de trigo	30	50	15
Hojas	15-20	20	20
Papel de diario	40-55	25-40	18-30
Papel de desecho de pulpas químicas	60-70	10-20	5-10
Sólidos primarios de aguas de desecho	8-15	N/D*	24-29
Desechos de puercos	6	28	N/D*
Estiércol sólido de ganado	1,6-4,7	1,4-3,3	2,7-5,7

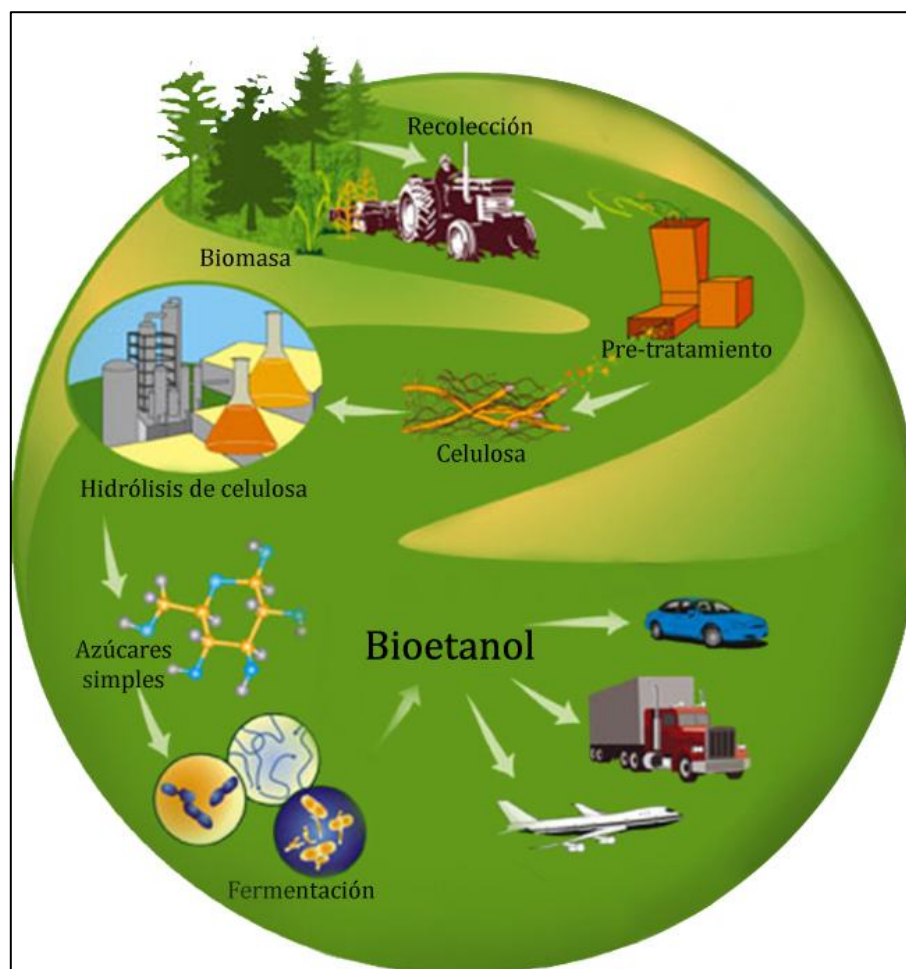
\* Información no disponible

### **1.1.3 Producción de bioetanol de segunda generación**

Este proceso consta de tres etapas principales: pre-tratamiento de los residuos lignocelulósicos, hidrólisis de la celulosa presente en los residuos y fermentación de azúcares a etanol. A continuación se presenta una breve descripción de estas etapas. En la Figura 1.1 se muestra un diagrama del proceso con sus principales etapas.

#### **Pre-tratamiento:**

Tiene como objetivos: remover la lignina, reducir la cristalinidad de la celulosa e incrementar la porosidad del material. Este proceso, debe cumplir con los siguientes requerimientos: (i) mejorar la formación de azúcares o la capacidad de formar posteriormente éstos mediante hidrólisis enzimática; (ii) evitar la degradación o pérdida de carbohidratos; (iii) evitar la formación de subproductos que puedan inhibir los procesos posteriores de hidrólisis y fermentación; y (iv) ser rentable. Se han utilizado procesos físicos, físico-químicos, químicos y biológicos para el pre-tratamiento de los residuos lignocelulósicos [15]. El resultado de esta etapa difiere según el proceso utilizado, en la Tabla 1.1 se presenta un resumen en donde se detalla el estado de salida de los principales componentes del material lignocelulósico luego de utilizar distintos procesos de pre-tratamiento [16].



**Figura 1.1 Diagrama del proceso de producción de bioetanol de segunda generación.**

Cabe señalar que la cristalinidad de la celulosa no se ve disminuida como resultado del pre-tratamiento de la biomasa, por lo que es necesario que el sistema enzimático de hidrólisis de la celulosa sea efectivo sobre celulosa cristalina [16].

Por otro lado, los microorganismos deben ser metabólicamente activos en presencia de productos inhibidores generados en la etapa de pre-tratamiento. Estos inhibidores provienen principalmente de la liberación hidrolítica de compuestos presentes en la biomasa no pre-tratada (ácidos orgánicos, extractivos y fenoles), reacción de carbohidratos y otros compuestos solubilizados para formar productos de degradación (furfural y hidroximetil fufural), y la corrosión resultante de la liberación de iones inorgánicos [16]. La cantidad de productos inhibidores depende en gran manera de las condiciones y configuración del proceso utilizado [16].

**Tabla 1.2 Condición de salida del material lignocelulósico en distintos procesos de pretratamiento**

<b>Proceso</b>	<b>Celulosa</b>	<b>Hemicelulosa</b>	<b>Lignina</b>
Pre-tratamiento con ácidos diluidos	Despolimerización baja	80-100% de solubilización, principalmente a monómeros.	Solubilización baja o nula, redistribución extensiva
Explosión de vapor a alta concentración de sólidos	Despolimerización baja	80-100% de solubilización a una mezcla de monómeros, oligómeros, y productos de degradación	Solubilización baja o nula, redistribución extensiva
Procesos hidrotermales	Despolimerización baja	80-100% de solubilización, oligómeros >50%	N/D*
Solventes orgánicos con agua	N/D*	Solubilización substancial	Solubilización parcial, entre 20 a 50%
Explosión de fibra de amoníaco	Descristalización baja	Solubilización entre 0 y 60%, 90% oligómeros	
Pre-tratamiento con hidróxido de sodio	Inflamación substancial	Solubilización substancial (sobre 50%)	Solubilización substancial (sobre 50%)
Pre-tratamiento con cal	N/D*	Solubilización significativa (30%)	Solubilización parcial (40%)

\* Información no disponible

### **Hidrólisis de la celulosa:**

La hidrólisis de la celulosa tiene como propósito la despolimerización de ésta, para así obtener azúcares simples que puedan ser fácilmente fermentables. Los productos de este proceso suelen ser azúcares reductores, incluyendo glucosa. Debido a que el costo del proceso es menor que al utilizar hidrólisis ácida o alcalina, se suele utilizar hidrólisis enzimática, la cual se puede realizar bajo condiciones suaves (pH 4,8 y temperatura entre 45 y 50°C) y no genera problemas de corrosión [15]. No obstante, se presentan algunas desventajas: el tiempo necesario para realizar la hidrólisis es considerablemente mayor que al utilizar hidrólisis ácida y el costo de las enzimas es mucho más alto que por ejemplo el ácido sulfúrico [17].

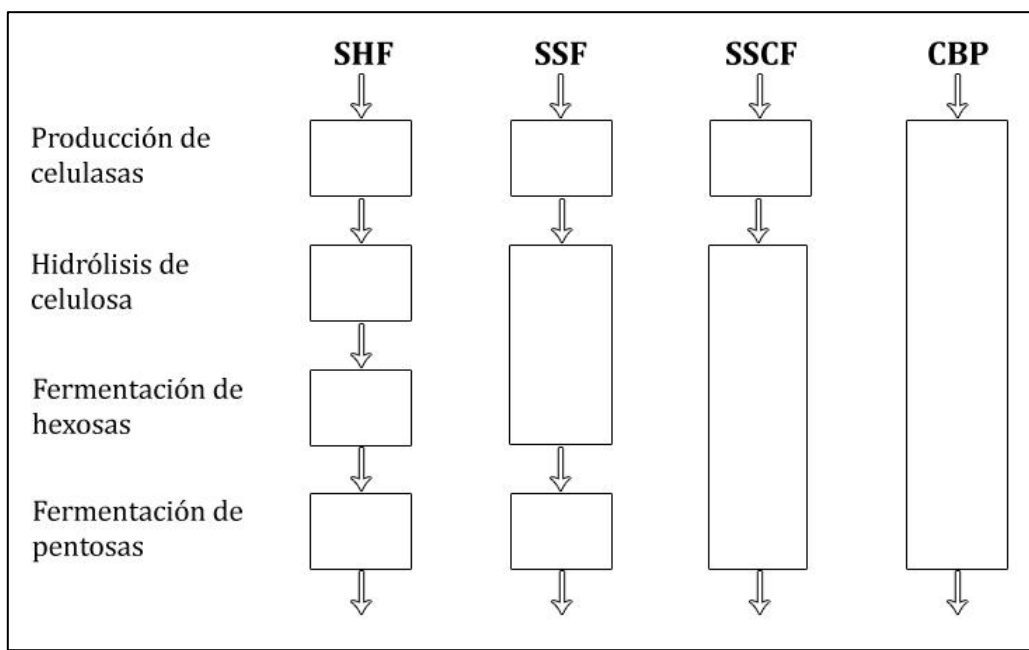
Cabe señalar, que los principales factores que afectan el proceso de hidrólisis enzimática son: la concentración y calidad del sustrato, el pre-tratamiento utilizado, la actividad de las celulasas y las condiciones de la hidrólisis, tales como temperatura, pH y agitación [17].

La reacción de hidrólisis enzimática de celulosa se realiza mediante la utilización de enzimas celulasas, las cuales pueden provenir desde hongos y bacterias [15]. Se ha reportado una gran cantidad de hongos productores de celulasas, siendo el género *Trichoderma* el más estudiado [15].

### **Fermentación:**

En esta etapa, los azúcares simples son fermentados a etanol por medio del uso de microorganismos. El microorganismo más comúnmente utilizado para este proceso a nivel industrial, es la levadura *Saccharomyces cerevisiae*, la cual ha demostrado ser muy resistente y adecuada para la fermentación de material lignocelulósico hidrolizado [18].

Durante el proceso de conversión de la biomasa a etanol, existen cuatro etapas mediadas biológicamente: (i) producción de celulasas, (ii) hidrólisis de celulosa y otros polisacáridos insolubles, (iii) fermentación de los productos solubles provenientes de la hidrólisis de la celulosa, y (iv) fermentación de los productos solubles provenientes de la hidrólisis de la hemicelulosa. Las distintas configuraciones del proceso han sido caracterizadas según su grado de consolidación. Cuando la hidrólisis y la fermentación se realizan en forma separada (SHF), el proceso comprende de cuatro etapas y se utilizan hasta cuatro biocatalizadores distintos. La sacarificación y fermentación simultánea (SSF) consolida la hidrólisis y la fermentación de los productos de la hidrólisis en una etapa, mientras que la producción de celulasas y la fermentación de los productos de la hidrólisis de hemicelulosa ocurren en dos etapas adicionales. La sacarificación y co-fermentación simultánea (SSCF) comprende de dos etapas: la producción de celulasas y una segunda etapa en donde ocurre la hidrólisis de la celulosa y la fermentación tanto de los productos provenientes de las hidrólisis de celulosa y hemicelulosa. En el caso del bioproceso consolidado (CPB), la producción de celulasas, hidrólisis y la fermentación de los productos de la hidrólisis de celulosa y hemicelulosa, se encuentran acoplados en un solo proceso [16]. En la Figura 1.2 se presenta un diagrama resumen con las distintas estrategias para el proceso. Cabe señalar que el utilizar un mayor número de etapas, tiene como ventaja poder mantener las condiciones óptimas en cada uno de los procesos, en cambio, disminuir éstas permite reducir el número de reactores necesarios [18].



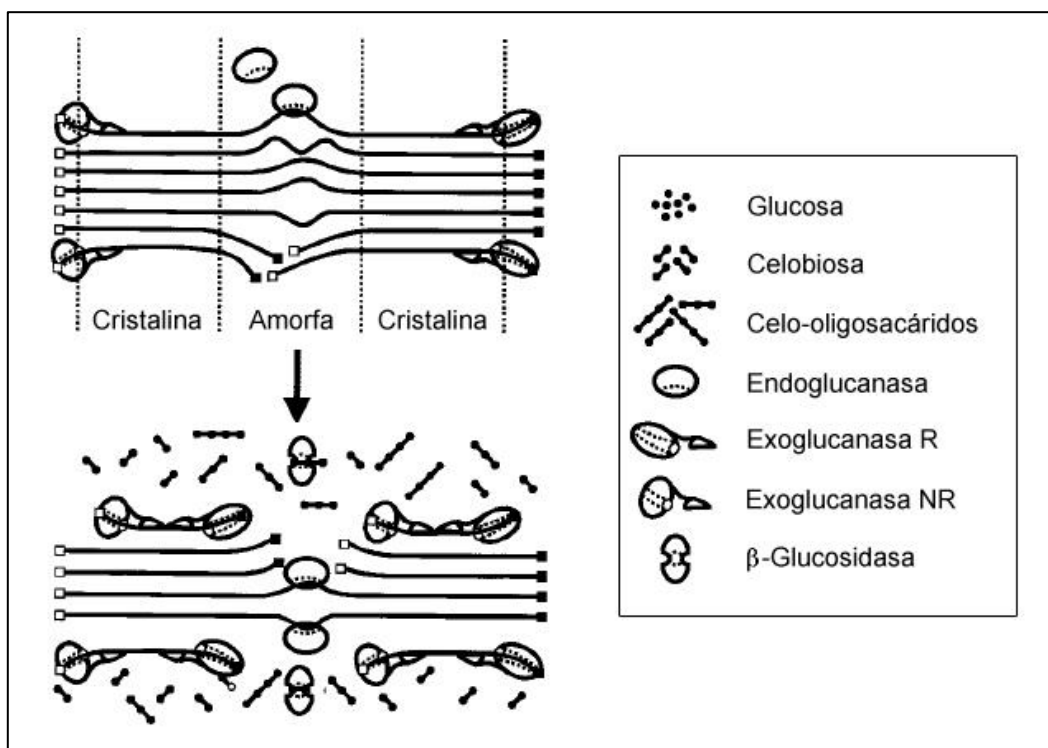
**Figura 1.2 Distintas configuraciones para el proceso de conversión de biomasa a etanol.** A la izquierda de la figura se presentan las etapas mediadas biológicamente. Las cajas representan biorreactores, éstos no se encuentran a escala. SHF: hidrólisis y fermentación separadas; SSF: sacarificación y fermentación simultánea; SSCF: sacarificación y co-fermentación simultánea; CBP: bioproceso consolidado.

#### 1.1.4 Mecanismo de hidrólisis enzimática de celulosa

La celulosa es un polisacárido lineal, formado por la unión de monómeros de D-glucosa a través de enlaces  $\beta$ -1,4-glucosídicos, con un grado de polimerización que va de 100 hasta 20.000 [19].

El mecanismo más aceptado para la hidrólisis de celulosa, comprende la acción sinérgica de endoglucanasas (EC 3.2.1.4), exoglucanasas o celobiohidrolasas (EC 3.2.1.91) y  $\beta$ -glucosidasas (EC 3.2.1.21). Las endoglucanasas hidrolizan enlaces  $\beta$ -1,4-glicosídicos en forma aleatoria desde las regiones amorfas de la cadena del polisacárido, generando nuevos extremos de cadena; las exoglucanasas actúan sobre los extremos reducidos y no reducidos de la cadena, liberando celobiosa o glucosa y finalmente las  $\beta$ -glucosidasas hidrolizan la celobiosa a glucosa con el fin de eliminar la inhibición por ésta [7]. En la Figura 1.3 se presenta un esquema de la hidrólisis de celulosa por el mecanismo enzimático antes mencionado.

Tanto endoglucanasas como exoglucanasas hidrolizan enlaces  $\beta$ -1,4-glicosídicos y su especificidad parece estar determinada por la topología de su sitio activo. Mientras las endoglucanasas poseen sitios activos abiertos en forma de hendidura, las exoglucanasas poseen el sitio activo localizado en túneles formados por largos bucles en la estructura de la proteína [20].



**Figura 1.3 Representación esquemática de hidrólisis enzimática de celulosa por celulasas.** Los extremos reductores y no reductores son representados por cuadrados negros y blancos respectivamente. Las exoglucanasa R y NR, hidrolizan desde los extremos reductores y no reductores respectivamente. Las regiones amorfas y cristalinas se encuentran indicadas.

Por lo general, las celulasas son enzimas bimodulares, con un dominio catalítico extenso y un pequeño dominio de unión a celulosa (CBD), estando éstos ligados por una secuencia proteica altamente glicosilada rica en serina y prolina [21]. El rol específico del enlace no está del todo claro, pero se le han atribuido funciones como: (i) participar en la destrucción de la celulosa interfiriendo con el empaquetamiento regular de las fibrillas; (ii) y/o aumentar la concentración local del dominio catalítico al que están unidos [22]. En tanto el CBD, permite el aumento de la concentración enzimática en un sitio particular de la superficie de la celulosa, dirigiendo a la enzima a localizaciones específicas en el sustrato insoluble, además de hacer el sustrato más accesible mediante la neutralización de asociaciones no covalentes en su estructura [23].

Los CBDs se presentan en todo tipo de celulasas, siendo éstas producidas tanto por bacterias aeróbicas y anaeróbicas, como por hongos. La determinación de la secuencia aminoacídica de más de 100 CBDs, ha permitido su clasificación en 10 familias distintas [24]. Gran parte de los CBDs pertenecen a las familias I, II y III, mientras que las restantes familias poseen pocos miembros y en algunos casos sólo uno. Todos los CBDs de celulasas producidas por hongos pertenecen a la familia I, la cual contiene CBDs pequeños de entre 33 y 44 residuos aminoacídicos [20].

La variabilidad en la complejidad estructural de la celulosa pura y la dificultad para trabajar con sustratos insolubles, han hecho necesario el trabajo con celulosas altamente solubles, como el caso de la carboximetilcelulosa (CMC), polímero de alto grado de polimerización, el cual es comúnmente utilizado en el estudio de la producción de endoglucanasas [25].

### **1.1.5 Microorganismos productores de celulasas**

Variadas especies de bacterias como *Clostridium*, *Cellulomonas*, *Thermomonospora*, *Bacillus*, *Bacteriodes*, *Ruminococcus*, *Erwinia*, *Acetovibrio*, *Microbispora* y *Streptomyces* y de hongos como *Trichoderma*, *Penicillium*, *Fusarium*, *Phanerochaete*, *Humicola* y *Schizophillum* son capaces de producir celulasas y hemicelulasas [17]. En comparación con los hongos, las bacterias producen cantidades pequeñas de enzimas con actividad celulolítica [18].

Las celulasas provenientes del hongo *Trichoderma reesei* han sido las más ampliamente estudiadas y mejor caracterizadas. Otras cepas que poseen actividad celulolítica no producen niveles adecuados de celulasas para su uso práctico, ya que no son capaces de generar una hidrólisis extensiva de grandes concentraciones de celulosa [6].

Las principales ventajas del género *Trichoderma* son: la producción de un sistema de celulasas con todos los elementos requeridos para la hidrólisis de celulosa cristalina, estabilidad bajo condiciones de hidrólisis enzimática (agitación, temperatura 50°C y pH 4,8, por 48 h o más) y resistencia a la inhibición por sustancias químicas. Sus principales desventajas son: la baja producción de  $\beta$ -glucosidasas y que sus celulasas tienen baja actividad específica, son inhibidas por producto y son lentamente inactivadas en reactores a 50°C. No obstante, es el mejor productor de celulasas disponible en la actualidad [6].

Por otro lado, el género *Aspergillus* es un eficiente productor de  $\beta$ -glucosidasas. En varios estudios las celulasas de *Trichoderma* han sido suplementadas con  $\beta$ -glucosidasas adicionales provenientes de *Aspergillus* obteniéndose buenos resultados [18].

### **1.1.6 Regulación de la expresión de celulasas en hongos**

En general, las celulasas de bacterias son producidas en forma constitutiva, no así las celulasas de hongos, que sólo se producen en presencia de celulosa como sustrato. Se cree que la celulosa por sí misma es incapaz de desencadenar la inducción en forma directa debido a su insolubilidad. No obstante, la presencia de un nivel basal de celulasas constitutivas podría degradar la celulosa a celooligosacáridos solubles, capaces de entrar a la célula y ser convertidos en un inductor que

desencadene la inducción. La presencia de celulosa como inductor, aumenta la expresión de los genes de celulasas en al menos 1.100 veces [26].

Variados celo-oligosacáridos y sus derivados han sido ensayados con el propósito de identificar inductores de celulasas. La celobiosa, el celo-oligosacárido más pequeño con enlace  $\beta$ -1,4, ha sido reportada como inductor de celulosa por numerosos investigadores. La soforosa, un disacárido de glucosa con un enlace  $\beta$ -1,2, es capaz de inducir 2.500 veces más celulasas que la celobiosa, en el altamente estudiado hongo *Trichoderma reesei* [27]. Por otro lado, la gentiobiosa, un disacárido de glucosa con un enlace  $\beta$ -1,6, induce 50 veces más la producción de celulasas que la celobiosa en el hongo *Penicillium purpurogenum*, sin embargo, la inducción fue observada sólo en presencia de un represor de  $\beta$ -glucosidasas [28]. Se cree que la inducción por moléculas que no corresponden a celo-oligosacáridos se produce por transglucosilación.

Una vez que el inductor entra a la célula, se desencadena una transcripción a gran escala de los genes de celulasas, la cual es mediada por proteínas activadoras y elementos activadores. Hasta el momento se desconoce como las proteínas activadoras y los elementos activadores son afectados por el inductor. La secreción de grandes cantidades de celulasas permite la degradación de mucha más celulosa a oligosacáridos y glucosa, los cuales son utilizados para el crecimiento del hongo [26].

Luego de la degradación de la celulosa, una gran cantidad de glucosa es liberada, la cual causa represión catabólica. La proteína de represión catabólica es directamente responsable de la regulación negativa a nivel transcripcional, previniendo al hongo de sintetizar un exceso de celulasas en condiciones donde existe una fuente de carbono de fácil asimilación [26]. Este fenómeno se presenta hasta que la glucosa alcanza niveles inferiores a  $\sim 0,1$  mg/ml [6].

### **1.1.7 Clasificación de glicosil hidrolasas**

Las glicosil hidrolasas (GHs) son enzimas claves en el metabolismo de los carbohidratos, encontrándose en los tres reinos principales (bacterias, arqueas y eucariontes). La nomenclatura para enzimas IUB (International Union of Biochemistry, 1984) se basa en el tipo de reacción que la enzima cataliza y en su especificidad de sustrato. En el caso de las GHs (3.2.1.x), los tres primeros dígitos indican la hidrólisis enzimática de enlaces O-glicosídicos, mientras el último dígito representa al sustrato y en algunos casos refleja el mecanismo molecular. Esta clasificación única es muy útil, especialmente para evitar ambigüedades y la proliferación de nombres triviales. Sin embargo, al menos para el caso de las GHs, no es una clasificación que refleje necesariamente las



características estructurales de la enzima. Otro problema con la clasificación EC es que no es apropiada para enzimas que muestran actividad sobre más de un sustrato [29].

Considerando lo anterior y dado que existe una relación directa entre la secuencia y su plegamiento y funcionalidad, se ha propuesto una clasificación en familias basada en la similitud de las secuencias primarias de las GHs. A partir de ésta se encontró (i) enzimas con EC distintos que pertenecían a la misma familia y (ii) enzimas con especificidad de sustrato similares que pertenecían a familias distintas [29]. Actualmente, esta clasificación comprende 115 familias de GHs [30].

### **1.1.8 Dominios conservados**

Los dominios de proteínas son unidades distintivas de la evolución molecular, usualmente asociados con aspectos funcionales y/o estructurales. Estas dos características, suelen encontrarse ligadas, y lo que corresponde a una unidad de plegamiento independiente de una cadena polipeptídica también conlleva una función específica.

Los dominios conservados se definen como unidades recurrentes en la evolución molecular, cuya extensión puede ser determinada por análisis de secuencia y estructural. Éstos contienen patrones de secuencias conservados o motivos, los que permiten su detección desde secuencias polipeptídicas.

Los *cluster* de superfamilia agrupan a conjuntos de dominios conservados procedente de distintas familias que poseen un ancestro evolutivo común, basándose en información estructural, funcional y evolutiva.

La identificación de dominios funcionales en secuencias proteicas, puede dar las primeras claves sobre su función molecular y celular.

## **1.2 Descripción y justificación del proyecto**

Como se mencionó anteriormente, Chile necesita de manera urgente nuevas alternativas para expandir su matriz energética. Dentro de éstas, la utilización de biocombustibles como el bioetanol de segunda generación, el cual es producido a partir de residuos lignocelulósicos, aparece como una buena alternativa. Esto último, debido a la gran cantidad de residuos forestales y agrícolas producidos en nuestro país.

El proceso de producción de bioetanol, está dividido en tres etapas principales: el pre-tratamiento del material lignocelulósico, que tiene como objetivo dejar expuesta la celulosa presente en éste; la hidrólisis de la celulosa, que tiene como finalidad llevar ésta a azúcares simples que puedan ser fácilmente fermentables; y finalmente, la fermentación de estos azúcares a etanol. Dentro de estas etapas, es sabido que el pre-tratamiento de los residuos y la hidrólisis de la celulosa, son las etapas que presentan un mayor potencial para ser optimizadas, con el fin de obtener un proceso económicamente rentable.

La hidrólisis de la celulosa, es catalizada por enzimas endoglucanasas, exoglucanasas y  $\beta$ -glucosidasas, las cuales actúan en forma sinérgica degradando la celulosa a monómeros de glucosa, los cuales luego pueden ser fermentados por microorganismos. Cabe señalar que el alto costo de estas enzimas, las cuales en la actualidad provienen del hongo *Trichoderma reesei* [6], es el mayor cuello de botella para hacer el proceso de producción de bioetanol de segunda generación económicamente viable. Por esto último, es fundamental encontrar nuevas enzimas, que permitan disminuir los costos en cuanto a la hidrólisis se refiere.

En este proyecto se buscó identificar y clonar, a partir de una micoteca compuesta principalmente por hongos de pudrición blanca, nuevas endoglucanasas que puedan ser utilizadas para la hidrólisis de celulosa en el proceso de producción de bioetanol de segunda generación. La utilización de este tipo de microorganismos se debió a que se encuentran presentes en la degradación de material lignocelulósico en la naturaleza y por lo general han sido estudiados más por su capacidad para degradar lignina que por su producción de celulasas [31].

En una primera etapa, se realizó un *screening* para identificar qué hongos poseían actividad endoglucanasa. Luego, a partir del DNA genómico y cDNA de los hongos seleccionados, se realizaron reacciones de PCR, utilizando partidores degenerados dirigidos a regiones de homología presentes en distintas familias de endoglucanasas. El diseño de los partidores se realizó mediante la herramienta CODEHOP. Los fragmentos generados fueron enviados a secuenciar, para luego analizar las secuencias *in silico*. Se procedió a completar las secuencias de los fragmentos pertenecientes a enzimas con actividad endoglucanasa, utilizando la estrategia RACE. Finalmente, a partir de la secuencia nucleotídica encontrada, se infirió el producto proteico utilizando herramientas computacionales.

## **1.3 Objetivos**

### **1.3.1 Objetivo General**

Identificar y clonar secuencias de endoglucanasas nuevas, potencialmente útiles para la hidrólisis de material lignocelulósico.

### **1.3.2 Objetivos Específicos**

- Identificar hongos productores de enzimas con actividad endoglucanasa.
- Aislar los genes codificantes para enzimas con actividad endoglucanasa, presentes en los hongos seleccionados.
- Inferir el producto proteico generado a partir de las secuencias obtenidas, mediante el uso de herramientas computacionales.

## 2. Materiales y métodos

### 2.1 Materiales

#### 2.1.1 Reactivos

En la Tabla 2.1 se presentan los distintos reactivos utilizados en el trabajo de laboratorio y sus respectivos proveedores.

**Tabla 2.1 Reactivos utilizados en el trabajo de laboratorio.**

<b>Proveedor</b>	<b>Reactivo</b>
Ambion	TRI®
CalBiochem	Ampicilina
Difco	Agar, Malta, LB
FCFM, Universidad de Chile	Nitrógeno líquido
Fermentas	Proteinasa K, X-Gal, IPTG
IDT	Oligonucleóticos
Invitrogen	3' RACE System for Rapid Amplification of cDNA Ends, 5' RACE System for Rapid Amplification of cDNA Ends, Desoxirribonucleasa I, ThermoScript™ RNase H- Reverse Transcriptase
Lafken	Agarosa
Merck	Agar papa dextrosa, sales inorgánicas, ácido cítrico, EDTA, SDS, ácido acético glacial, etanol, cloroformo, glucosa.
New England Biolabs	EcoRI, TdT
Promega	Kit Taq DNA polimerasa, kit pGem®-T Easy, kit Reverse transcription system
QIAGEN	Kit QIAEX II, kit QIAprep Spin Miniprep
Sigma	DNS, tris, bromuro de etidio, di-metilformamida

#### 2.1.2 Equipos

En la Tabla 2.2 se presentan los distintos equipos utilizados en el trabajo de laboratorio con sus respectivos fabricantes y modelos.

**Tabla 2.2 Equipos utilizados en el trabajo de laboratorio.**

<b>Equipo</b>	<b>Fabricante y modelo</b>
Agitador orbital	LabLine MaxQ 4000
Cámara fotográfica	Nikon Coolpix 4500
Centrífuga	BOECO M-24 Eppendorf 5804 R
Electroporador	Gibco BRL Cell-Porator
Fuente de poder para electroforesis	BRL Model 500
Lector de absorbancia (placas 96 pocillos)	Anthos 2010 Asys UVM 340
Termociclador	MJ Research PTC 100 Eppendorf Mastercycler Gradient
Transiluminador para geles de agarosa	Vilber Loumat

### 2.1.3 Herramientas computacionales

En la Tabla 2.3 se presentan las herramientas computacionales utilizadas en este trabajo.

**Tabla 2.3 Herramientas computacionales utilizadas.**

<b>Herramienta</b>	<b>Fuente</b>
BLASTX	NCBI [32]
BlockMaker	Fred Hutchinson Cancer Research Center [33]
CDD	NCBI [34]
ClustalX2	NCBI [35]
CODEHOP	Department of Pathobiology, School of Public Health and Community Medicine, University of Washington [36]
Primer3	Whitehead Institute for Biomedical Research [37]
TreeView	Taxonomy and Systematics at Glasgow [38]

### 2.1.4 Enzimas de restricción

En la Tabla 2.4 se observan las enzimas de restricción utilizadas en este trabajo.

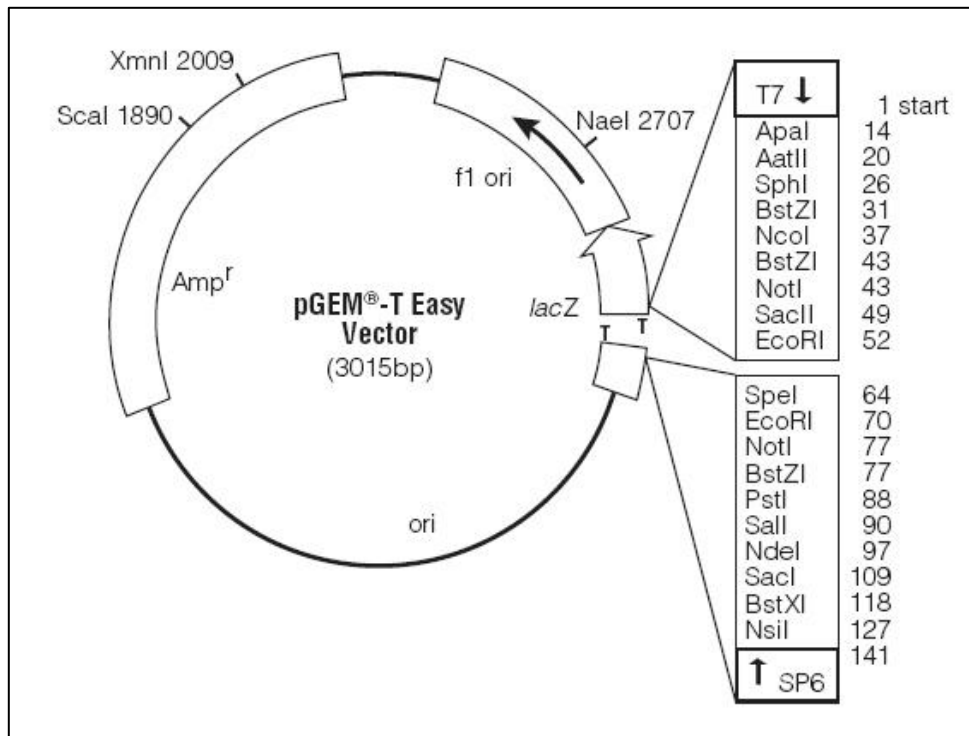
**Tabla 2.4 Enzimas de restricción utilizadas durante el trabajo.**

<b>Enzima</b>	<b>Sitio de restricción*</b>
EcoRI	G <sup>^</sup> AATTC CTTAA <sup>^</sup> G

\* El símbolo <sup>^</sup> representa el sitio de corte de la enzima.

### 2.1.5 Vector de clonación

En la Figura 2.1 se presenta el vector de clonamiento pGEM®-T Easy y sus principales puntos de referencia [39]. Este vector se encuentra linealizado y posee timidinas terminales en los extremos 3' de ambas hebras, permitiendo una eficiente ligación de productos de PCR generados mediante la utilización de la enzima Taq DNA Polimerasa. Por otro lado, posee un gen de resistencia a ampicilina *Amp<sup>r</sup>* y el gen *lacZ*, el cual codifica para la β-galactosidasa, estando este último interrumpido por el sitio de múltiple clonamiento. La utilización de estos genes permite un eficiente método de screening.



**Figura 2.1 Mapa del vector pGEM®-T Easy y sus puntos de referencia.** *Amp<sup>r</sup>*: gen de resistencia a ampicilina. *lacZ*: gen que codifica para la β-galactosidasa, interrumpido por el sitio de múltiple clonamiento. En el recuadro de la derecha se listan los sitios de restricción disponibles en la región de múltiple clonamiento.

## 2.2 Métodos

### 2.2.1 Construcción de la micoteca

Se trabajó a partir de 30 hongos provistos por el Laboratorio de Biodeterioro y Preservación de la Madera de la Facultad de Ciencias Forestales de la Universidad de Chile. Los hongos utilizados se encuentran en la Tabla 2.5. A partir de cada una de éstos, se inocularon dos placas en medio agar malta (ver Anexo 6.1.1.1). Para ello, manteniendo condiciones de asepsia, con ayuda de un bisturí se

transfirió un cuadrado de 1 cm aproximado de lado de muestra (hongo más agar) hacia las nuevas placas. Las placas se mantuvieron a 28°C hasta que se cubrieron completamente con el hongo, luego se retiraron a 4°C para evitar que el hongo siguiera creciendo. Las placas que presentaron contaminación se repitieron. Una de las placas se utilizó para el trabajo posterior, mientras la segunda quedó como respaldo.

**Tabla 2.5 Micoteca de trabajo**

La información de la columna phylum fue obtenida del NCBI. La información de la columna tipo fue proporcionada junto a la micoteca. Las siglas HPB, HPP y N/D, corresponden a hongo de pudrición blanca, parda y información no disponible respectivamente.

Clave*	Especie	Phylum	Tipo
96	<i>Amblyosporium sp.</i>	Microsporidia	N/D
97	<i>Coriolus multicolor</i>	Basidiomycota	HPB
98	<i>Lenzites seiparia</i>	Basidiomycota	N/D
100	<i>Auricularia polytricha</i>	Basidiomycota	N/D
101	<i>Pluteus cervinus</i>	Basidiomycota	N/D
104	<i>Poria monticola</i>	Basidiomycota	HPP
106	<i>Polystictus versicolor</i>	Basidiomycota	HPB
108	<i>Polystictus sulphureus</i>	Basidiomycota	N/D
110	<i>Poria placenta</i>	Basidiomycota	HPP
113	<i>Phanerochaete chrysosporium</i>	Basidiomycota	HPB
114	<i>Agrocybe aegerita</i>	Basidiomycota	HPB
117	<i>Ganoderma applanatum</i>	Basidiomycota	HPB
121	<i>Schizophyllum commune</i>	Basidiomycota	HPB
127	<i>Peniophora gigantea</i>	Basidiomycota	HPB
140	<i>Trametes cinnabarinus</i>	Basidiomycota	N/D
147	<i>Lentinula edodes</i>	Basidiomycota	HPB
151	<i>Trametes versicolor</i>	Basidiomycota	HPB
152	<i>Ungulina fomentaria</i>	Basidiomycota	HPB
162	<i>Coriolus versicolor</i>	Basidiomycota	HPB
163	<i>Fusarium oxysporum</i>	Ascomycota	N/D
164	<i>Gymnopilus spectabilis</i>	Basidiomycota	HPB
167	<i>Pycnoporus cinnabarinus</i>	Basidiomycota	HPB
168	<i>Stereum hirsutum</i>	Basidiomycota	HPB
177	<i>Gymnopilus spectabilis</i>	Ascomycota	N/D
178	<i>Lenzites betulinus</i>	Basidiomycota	HPB
181	<i>Alternaria alternata</i>	Basidiomycota	HPB
183	<i>Grifola gargal</i>	Basidiomycota	HPB
186	<i>Stereum hirsutum</i>	Basidiomycota	HPB
190	<i>Pleurotus ostreatus</i>	Basidiomycota	HPB
202	<i>Pleurotus ostreatus</i>	Basidiomycota	HPB

\*: Con el fin de identificar los aislados en forma más fácil se asignó una clave numérica para identificarlos.

## **2.2.2 Ensayos de actividad endoglucanasa**

### **2.2.2.1 *Utilizando cultivos en medio sólido***

Desde cada placa de medio agar malta destinada para seguir trabajando, se transfirió un cuadrado de 1 cm aproximado de lado de muestra a dos placas con medio agar malta CMC (ver Anexo 6.1.1.1). Estas placas se llevaron a 28°C y se dejaron crecer hasta que el diámetro de crecimiento del hongo fue aproximadamente la mitad del diámetro de la placa. Para determinar la actividad de los hongos, se tiñeron las placas utilizando rojo Congo al 0,1% durante 30 min, para posteriormente ser lavadas con NaCl 1M durante 15 min con agitación orbital suave. Este procedimiento revela halos sin tinción, los cuales corresponden a las zonas donde el CMC ha sido degradado.

### **2.2.2.2 *Utilizando cultivos en medio líquido***

Los hongos se cultivaron en medio líquido para poder utilizar el sobrenadante en ensayos de actividad. Para ello, se cortó un cuadrado de aproximadamente un 1 cm de lado, desde las placas con medio agar malta y se inocularon en 50 ml de medio mínimo-CMC (ver Anexo 6.1.1.2). Los cultivos se mantuvieron a 28°C con agitación orbital de 200 RPM, para luego realizar ensayos de actividad a partir de una muestra de 1 ml obtenida desde éstos. Debido a que la velocidad de crecimiento de los distintos hongos no es homogénea, se obtuvieron muestras luego de 3, 4, 5 y 7 días de cultivo, para así poder determinar el máximo de actividad de cada una de éstos.

Para realizar los ensayos de actividad, la muestra obtenida se centrifugó a 14.000 RPM a 4°C durante 10 min y se recuperó el sobrenadante. Este último fue utilizado para hidrolizar el sustrato soluble CMC. Luego, con el fin de cuantificar los azúcares reductores liberados se utilizó DNS. Finalmente, el producto proveniente de la reducción del DNS fue medido mediante espectrofotometría óptica. La cantidad de azúcares reductores presentes en la mezcla es directamente proporcional a la absorbancia de la muestra, por lo que a mayor absorbancia la degradación del CMC por acción enzimática es mayor. El protocolo utilizado es detallado en el Anexo 6.1.2.

Con el fin de cuantificar la actividad endoglucanasa de cada hongo, se utilizó una curva de calibración confeccionada con distintas concentraciones de glucosa (ver Anexo 6.1.3). Una unidad de actividad enzimática (U), se definió como la cantidad de  $\mu$ moles de azúcares reductores producidos en un minuto a 37°C y pH 4,8.



Debido a la presencia de CMC en el medio de cultivo, fue primordial determinar qué azúcares reductores se produjeron a partir de éste. Para ello, se realizó en forma paralela un blanco para cada muestra, el cual consistió en realizar la reacción en buffer citrato sin CMC. De esta forma, para calcular el valor real de la muestra fue necesario restar la absorbancia del blanco a la absorbancia de la muestra. Todas las muestras fueron realizadas en duplicado y ambos valores fueron promediados.

Por otro lado, a partir del sobrenadante centrifugado, se realizó un ensayo de actividad cualitativo en geles de agarosa con CMC (ver Anexo 6.1.4). Si existen endoglucanasas en el sobrenadante éstas degradaran el CMC presente en el gel, presentándose halos sin tinción.

Para todos los ensayos se utilizó como control negativo una muestra proveniente desde un medio de cultivo sin inocular.

### **2.2.2.3 Utilizando cultivos en medios inductores**

Utilizando la metodología detallada en el punto 2.2.1 los hongos fueron cultivados a 28°C en medio mínimo sólido utilizando CMC, avicel y Lenga como fuente de carbono (ver Anexo 6.1.1.1). Una vez que el hongo cubrió toda la placa, se obtuvieron sus proteínas utilizando un buffer de extracción (detalle del procedimiento en el Anexo 6.1.12). A partir del extracto de proteínas obtenido, se realizaron ensayos de actividad empleando el método DNS (ver Anexo 6.1.2). El cálculo de la actividad endoglucanasa se realizó en forma análoga al punto anterior.

### **2.2.3 Diseño de partidores**

A partir de las bases de datos de EMBL-EBI [40], se obtuvieron las secuencias aminoacídicas de las endoglucanasas presentes en hongos de las familias 5, 7, 12 y 61. La selección de estas familias, se debió a que fueron en las cuales se encontraron una mayor cantidad de secuencias descritas.

Para el diseño de los partidores degenerados se siguió la estrategia CODEHOP[36]. En un primer paso, las secuencias aminoacídicas fueron alineadas utilizando la herramienta ClustalX2 [35]. A partir de los alineamientos, mediante la utilización de la herramienta BlockMaker [33] se generaron bloques de homología, los que finalmente son utilizados como *input* por el programa para el diseño de partidores degenerados CODEHOP. Utilizando esta estrategia se diseñaron partidores *forward* y *reverse*, para las 4 familias antes mencionadas, los que se presentan en la Tabla 2.6.

**Tabla 2.6 Partidores degenerados utilizados.**

<b>Nombre</b>	<b>Dirección</b>	<b>Familia</b>	<b>Secuencia 5´a 3´*</b>
celF_endo_fwd5	<i>Forward</i>	5	tgatcttcgacatcatgaacgarymhcayra
celF_endo_rev5	<i>Reverse</i>	5	ccccaccaggagccdkydkcca
celF_endo_fwd7	<i>Forward</i>	7	cgacgaggagacctgcsnvanaaytg
celF_endo_rev7	<i>Reverse</i>	7	cacgtagcactggcgctrcartancc
celF_endo_fwd12	<i>Forward</i>	12	cgcccctagatcgtgtayaabawbcwrt
celF_endo_rev12	<i>Reverse</i>	12	cgtcaggtagttgaacaactccttnabrtcnsc
celF_endo_fwd61	<i>Forward</i>	61	tggacaagaccaccaacaagtkbktnaarat
celF_endo_rev61	<i>Reverse</i>	61	acttctagatggtgatgaagatgccngsrctngt

\* El alfabeto degenerado se presenta en el Anexo 6.1.5.

Por otra parte, para la estrategia RACE, fue necesario diseñar partidores específicos. Éstos se diseñaron utilizando la herramienta Primer3 [37], la cual permite verificar la formación de estructuras secundarias indeseadas como horquillas y dímeros. Para cada secuencia seleccionada se diseñó un partidor en sentido *forward* y otro en sentido *reverse*, con un T<sub>m</sub> de 65°C por recomendación del fabricante del kit. En la Tabla 2.7 se presentan los partidores específicos utilizados.

**Tabla 2.7 Partidores utilizados para realizar RACE en las secuencias seleccionadas.**

<b>Nombre</b>	<b>Dirección</b>	<b>Familia</b>	<b>Hongo</b>	<b>Secuencia 5´a 3´</b>
106EN5Fwd	<i>Forward</i>	5	106	ggacccgaacaacaacattgcta
106EN5Rev	<i>Reverse</i>	5	106	gcgaggatgagctggctggt
117EN61Fwd	<i>Forward</i>	61	117	gtgcagcaggaaatcatgaacg
117EN61Rev	<i>Reverse</i>	61	117	cgtgcacgacgggtagaactc

Además, se diseñaron partidores para las secuencias provenientes del hongo 163 los cuales se presentan en el Anexo 6.1.15.

#### **2.2.4 Extracción de DNA genómico desde hongos**

A partir de los hongos crecidos en medio agar malta, se inocularon placas con medio sólido agar papa dextrosa (ver Anexo 6.1.1.1) utilizando el mismo procedimiento detallado en el punto 2.2.1. Éstas se cultivaron a 28°C hasta que su superficie estuviese completamente cubierta por el hongo. Con ayuda de un bisturí se tomó una muestra de micelio equivalente a la mitad de la superficie de la placa y se introdujo en un tubo Eppendorf de 1,5 ml estéril con 100 µl de proteinasa K (µg/µl en buffer 0,25 M Tris HCl, 0,01 M CaCl<sub>2</sub>). El tubo se congeló durante 30 s utilizando N<sub>2</sub> (l), luego se llevó a 60°C por 2 min y posteriormente a vórtex por 2 min, estos 3 pasos se repitieron una vez. A continuación, la muestra se volvió a congelar durante 30 s con N<sub>2</sub> (l), se llevó a 95°C por 15 s, 4°C por 30 s y vórtex por 2 min. Finalmente, se centrifugó por 10 min a 9.000 RPM y se transfirió el

sobrenadante a un tubo Eppendorf de 0,6 ml estéril. La extracción de DNA fue verificada realizando una reacción de PCR con partidores universales para el gen 18S (ver detalle en el Anexo 6.1.6), altamente conservado en hongos [41].

### **2.2.5 Amplificación de fragmentos por PCR a partir de DNA genómico**

Utilizando el DNA genómico obtenido como templado y los partidores degenerados diseñados para cada familia, se realizaron una serie de reacciones de PCR según se explica a continuación. La mezcla de reacción utilizada fue la siguiente: 1  $\mu\text{M}$  de cada partidor (forward y reverse), 0,2 mM de dNTPs, 1,4 mM de  $\text{MgCl}_2$ , buffer Taq 1X, 1 U Taq polimerasa y 0,5  $\mu\text{l}$  de DNA genómico en un volumen total de 50  $\mu\text{l}$ . El programa utilizado para el PCR fue el siguiente: un ciclo de 95°C por 5 min, 5 ciclos de 95°C por 1 min, 50°C por 1 min y 72°C por 1 min, 35 ciclos de 95°C por 1 min, 72°C por 1 min con gradiente de 4°C y 72°C por 1 min y finalmente un ciclo de 72°C por 10 min.

Además, con el fin de aumentar la eficiencia y especificidad, se realizaron reacciones de PCR utilizando DMSO como agente adyuvante [42]. La mezcla de reacción utilizada fue la siguiente: 1  $\mu\text{M}$  de cada partidor, 0,2 mM de dNTPs, 1,4 mM de  $\text{MgCl}_2$ , buffer Taq 1X, 1 U Taq polimerasa, 1% DMSO y 0,5  $\mu\text{l}$  de DNA genómico en un volumen total de 50  $\mu\text{l}$ . El programa utilizado para el PCR fue el siguiente: un ciclo de 95°C por 5 min, 5 ciclos de 95°C por 1 min, 50°C por 1 min y 72°C por 1 min, 35 ciclos de 95°C por 1 min, 62°C por 1 min con gradiente de 8°C y 72°C por 1 min y finalmente un ciclo de 72°C por 10 min.

Los resultados de los distintos PCR se visualizaron en un gel de agarosa al 1,5%. Los fragmentos seleccionados fueron purificados desde el gel, para posteriormente ser ligados al vector pGEM®-T Easy. Luego, con el fin de obtener un mayor número de copias de los constructos vector-inserto, éstos fueron clonados en células *E.coli* DH5 $\alpha$ . Finalmente, se realizó una extracción del DNA plasmidial desde las células, el cual fue digerido mediante enzimas de restricción para verificar que el inserto era el deseado. Los detalles de estos protocolos se presentan en el Anexo 6.1.

### **2.2.6 Extracción de RNA desde hongos**

La extracción de RNA se realizó mediante la utilización del reactivo TRI® según se indica a continuación. Con ayuda de un bisturí se obtuvo el micelio presente en la superficie de dos placas Petri, se puso en un mortero y se le agregó  $\text{N}_2(\text{l})$ . La muestra congelada se homogenizó y luego se disolvió en 1 ml de TRI REAGENT. Se incubó durante 5 min a temperatura ambiente y se centrifugó a 12000 g a 4°C por 10 min. A continuación, se transfirió el sobrenadante a un tubo nuevo, se le agregaron 200  $\mu\text{l}$  de cloroformo y se agitó vigorosamente por 15 s. Se incubó durante 10 min a

temperatura ambiente, para luego centrifugar a 12000 g a 4°C por 15 min. La fase acuosa superior (fase sin color) se transfirió a un tubo nuevo, al cual se le agregó 500 µl de isopropanol, se agitó con vórtex por 10 s y se incubó a temperatura ambiente por 10 min. Luego, se centrifugó a 12000 g a 4°C por 8 min y se descartó el sobrenadante. El sedimento se lavó con 1 ml de etanol 75%, se centrifugó a 7500 g por 5 min, se removió el etanol y se dejó secar a temperatura ambiente. Finalmente, el sedimento se resuspendió en 50 µl de agua libre de RNAsas.

Con el fin de inhibir la acción de RNAsas, el material utilizado para la extracción de RNA fue previamente tratado con DEPC (ver Anexo 6.1.13). La extracción de RNA se verificó en un gel de agarosa al 1,5%.

Debido a la posibilidad de que la muestra se encontrara contaminada con DNA, se realizó una reacción de PCR utilizando partidores universales para el gen 18S (ver Anexo 6.1.6). El resultado de la reacción se verificó en un gel de agarosa al 1,5%. La presencia de una banda de aproximadamente 500 pb implicaría que la muestra se encuentra contaminada, haciendo necesario realizar un tratamiento con DNAsas como se indica en el Anexo 6.1.14.

### **2.2.7 Generación de cDNA a partir del RNA**

Para la generación de cDNA a partir de RNA extraído, se utilizó el kit reverse transcription system. En primer lugar, se calentó la muestra de RNA a 70°C por 10 min, se centrifugó brevemente y se mantuvo en hielo. A continuación, se preparó la mezcla de reacción como se indica a continuación: MgCl<sub>2</sub> 5mM, RT buffer 10X, dNTPs 1mM, rRNAsin® (inhibidor de RNAsas) 2U/µl, partidor Oligo dT 25 µg/ml, AMV RT 22 U/µl, RNA 46.7%. La mezcla se llevó a 42°C por una hora, 90°C por 5 min y finalmente se incubó a 4°C por 5 min.

### **2.2.8 Amplificación de fragmentos por PCR a partir de cDNA**

Utilizando el cDNA generado como templado y los partidores degenerados diseñados para cada familia, se realizaron una serie de reacciones de PCR según se explica a continuación. La mezcla de reacción utilizada fue la siguiente: 1 µM de cada partidor (forward y reverse), 0,2 mM de dNTPs, 1,4 mM de MgCl<sub>2</sub>, buffer Taq 1X, 1 U Taq polimerasa y 0,5 µl de cDNA en un volumen total de 25 µl. El programa utilizado para el PCR fue el siguiente: un ciclo de 95°C por 5 min, 5 ciclos de 95°C por 1 min, 50°C por 1 min y 72°C por 1 min, 35 ciclos de 95°C por 1 min, 65°C por 1 min con gradiente de 5°C y 72°C por 1 min y finalmente un ciclo de 72°C por 10 min.

Los resultados de los distintos PCR se visualizaron en un gel de agarosa al 1,5%. Los fragmentos seleccionados fueron clonados en el vector pGEM®-T Easy en forma análoga a como se indica en la sección 2.2.5.

### **2.2.9 RACE (Rapid Amplification of cDNA Ends)**

Con el fin de conocer las secuencias río arriba y río abajo de los fragmentos encontrados desde el cDNA, se utilizaron con algunas modificaciones los kits 3' RACE System for Rapid Amplification of cDNA Ends y 5' RACE System for Rapid Amplification of cDNA Ends respectivamente.

A continuación se detalla el procedimiento realizado para obtener las secuencias faltantes.

#### **2.2.9.1 Amplificación del extremo 5'**

Se sintetizó cDNA a partir del RNA utilizando los partidores gen-específico dirección *Reverse* (ver Tabla 2.7) y la enzima transcriptasa reversa ThermoScript™. En un tubo estéril se agregaron 1,5 µl de partidador específico 10 mM, 2 µl de dNTPs 10 mM y 8,5 µl de RNA. El tubo se incubó a 65°C por 5 min y luego se colocó en hielo. Posteriormente, se centrifugó brevemente y se agregó la siguiente mezcla de reacción: 5 µl de buffer de síntesis de cDNA 5X, 1 µl DTT 0,1 M, 2,25 µl de H<sub>2</sub>O, y 0,75 de ThermoScript™. Finalmente, la mezcla se llevó a 60°C por una hora y luego a 85°C por 5 min.

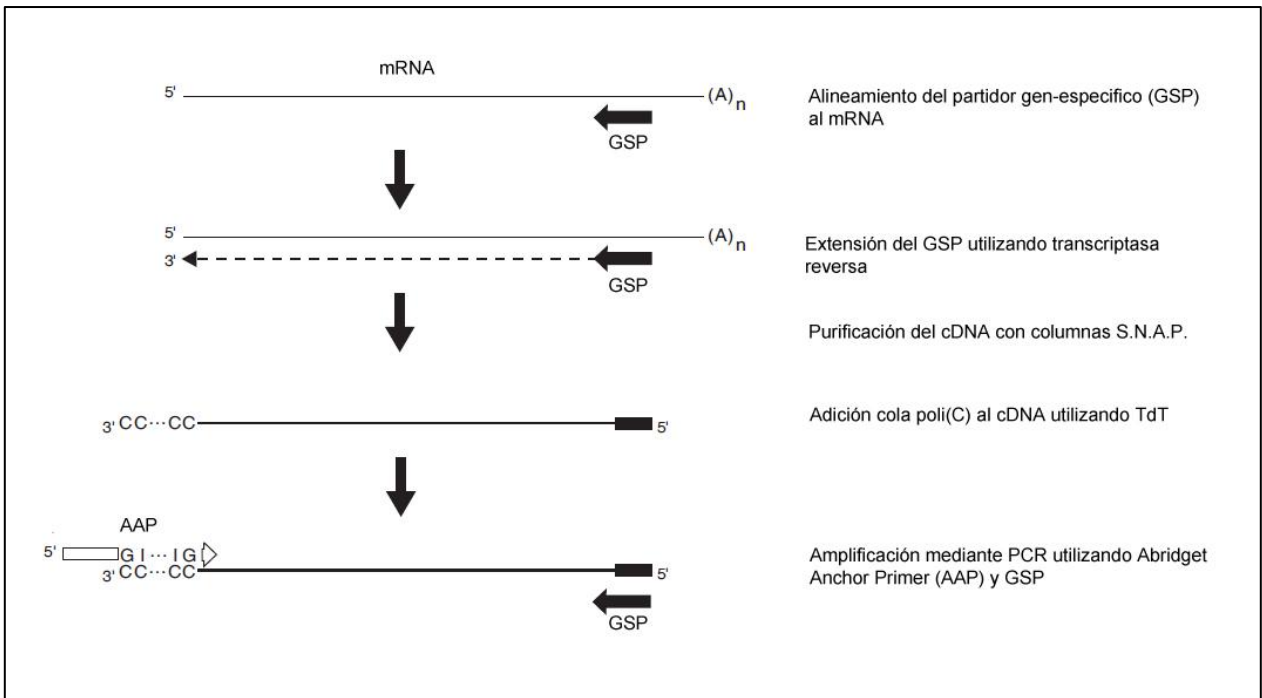
Con el fin de eliminar el exceso de partidores, dNTPs y proteínas que pudiesen entorpecer los procesos posteriores, se realizó una purificación del cDNA utilizando columnas S.N.A.P. según las instrucciones del fabricante.

Luego de la purificación, mediante la utilización de la enzima desoxinucleotidil-transferasa terminal (TdT), se agregó una cola poli(C) en el extremo 3' del cDNA. La mezcla de reacción fue la siguiente: 5 µl de buffer TdT 10X, 5 µl de CoCl<sub>2</sub> 2,5 mM, 0,5 µl dCTP 10 mM, 0,5 µl de TdT y 10 µl de cDNA en un volumen final de 50 µl. Se colocaron todos los componentes salvo la enzima en un tubo Eppendorf de 0,6 ml, el cual se llevó a 94°C por 3 min y luego se puso en hielo durante 1 min. El tubo se centrifugó brevemente para que bajase el líquido condensado en las paredes y se agregó la enzima TdT. Se llevó a 37°C durante 30 min y luego a 70°C durante 10 min.

Finalmente, se realizó una reacción de PCR utilizando como templado el cDNA con la cola poli(C). La mezcla de reacción utilizada fue la siguiente: 10 µl buffer Taq 5X, 3 µl MgCl<sub>2</sub> 25 mM, 1 µl dNTPs 10 mM, 2 µl partidador gen-específico dirección *reverse* 10 mM, 2 µl de partidador *Abridged Anchor Primer* (5'-GGCCACGCGTTCGACTAGTACGGGIIGGGIIGGGIIG-3') 10mM, 0,5µl de Taq DNA polimerasa 5 U/µl y 2,5 µl de cDNA con cola poli(C) en un volumen total de 50 µl. El programa utilizado para la reacción

fue: un ciclo de 94°C por 2 min, 5 ciclos de 94°C por 1 min, 50°C por 1 min y 72°C por 1 min, 31 ciclos de 94°C por 1 min, 68°C por 1 min con gradiente de 5°C y 72°C por 2 min y finalmente un ciclo de 72°C por 10 min. Con el fin de aumentar la eficiencia de la reacción, en algunas reacciones de PCR se adicionó BSA a una concentración final de 0,56 mg/ml [43].

En la Figura 2.2 se presenta un diagrama resumen de las principales etapas antes descritas para la amplificación del extremo 5'.

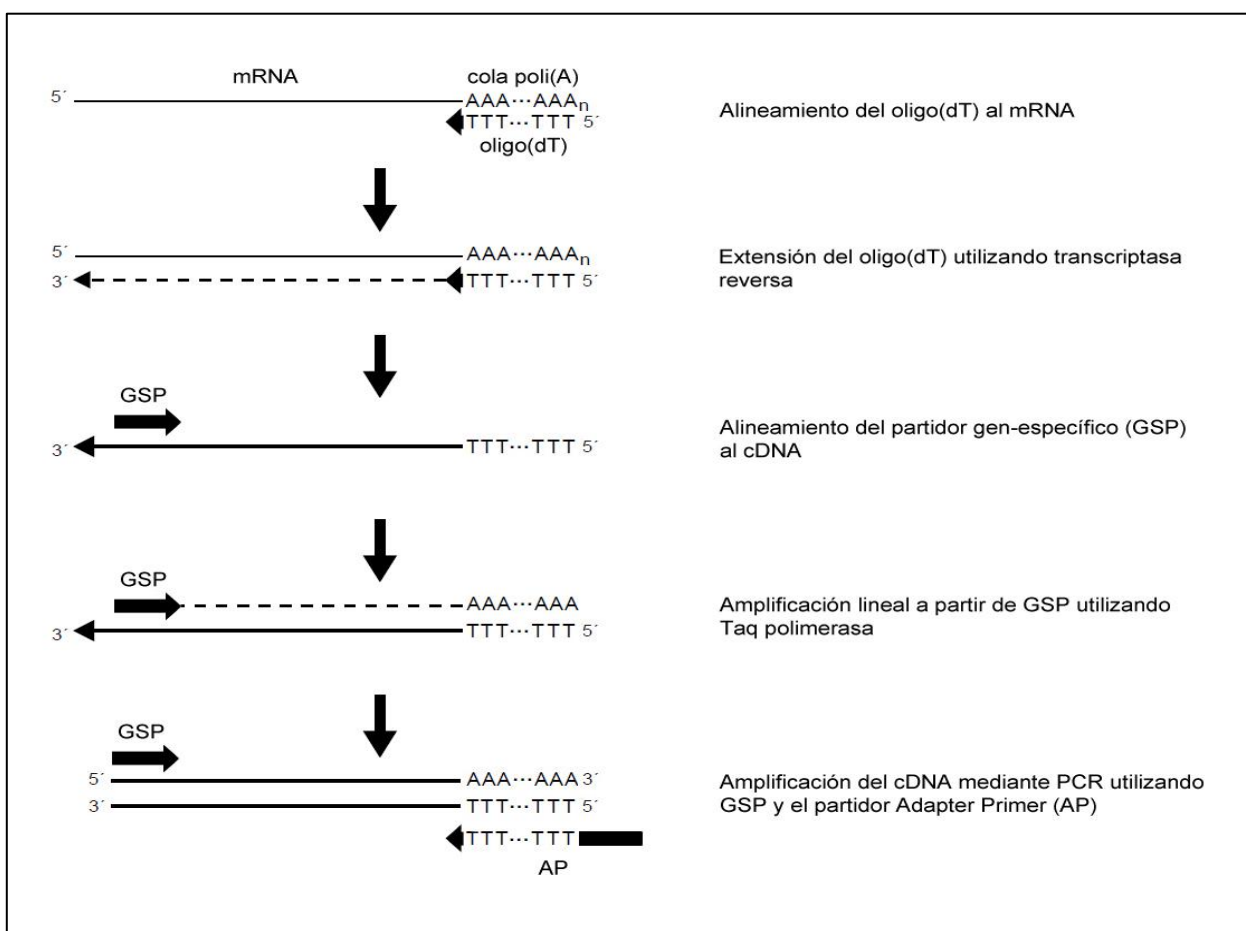


**Figura 2.2 Diagrama de la estrategia utilizada para RACE extremo 5'.** En el diagrama se presentan las principales etapas realizadas para la amplificación del extremo 5'. GSP corresponde a un partidor gen-especifico diseñado a partir de una secuencia nucleotídica conocida.

### 2.2.9.2 Amplificación del extremo 3'

A partir del mRNA de los hongos seleccionados, se sintetizó cDNA en forma análoga al punto 2.2.7. El cDNA generado fue utilizado para una amplificación lineal a partir del partidor específico dirección *forward* (ver Tabla 2.7). La mezcla de reacción se detalla a continuación: 5 µl de buffer Taq polimerasa 5X, 1,5 µl MgCl<sub>2</sub> 25 mM, 0,5 µl dNTPs 10 mM, 0,5 µl partidor específico dirección *forward* 10 mM, 0,5 µl Taq polimerasa 5 U/µl y 0,5 µl de cDNA en un volumen total de 25µl. El programa utilizado para la reacción fue el siguiente: un ciclo de 94°C por 1 min, 20 ciclos de 94°C por 32 s, 58,5°C por 32 s y 68°C por 5 min y un ciclo final de 70°C por 7 min.

Utilizando la amplificación lineal como templado, se realizó una segunda reacción de PCR correspondiente a una amplificación exponencial. La mezcla de reacción fue la siguiente: 10 µl de buffer Taq polimerasa 5X, 3 µl de MgCl<sub>2</sub> 25mM, 1 µl de dNTPs 10 mM, 1µl de partidor *Adapter Primer* (5'-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTTT-3') 10 mM, 1 µl partidor específico dirección *forward* 10 mM, 0,5 µl Ta polimerasa 5 U/ul y 1 µl del producto de la amplificación lineal diluida 10 veces en un volumen total de 50 µl. El programa utilizado fue el siguiente: un ciclo de 95°C por 5 min, 5 ciclos de 95°C por 1 min, 42°C por 1 min y 72°C por 1 min, 35 ciclos de 95°C por 1 min, 65°C por 1 min con un gradiente de 10°C y 95°C por 1 min, y un ciclo final de 72°C por 10 min. Al igual que en el caso del extremo 5', con el fin de aumentar la eficiencia de la reacción, en algunas reacciones de PCR se adicionó BSA a una concentración final de 0,56 mg/ml [43].



**Figura 2.3 Diagrama de la estrategia utilizada para RACE extremo 3'.** En el diagrama se presentan las principales etapas realizadas para la amplificación del extremo 3'. GSP corresponde a un partidor gen-específico diseñado a partir de una secuencia nucleotídica conocida.

En la Figura 2.3 se presenta un diagrama resumen de las principales etapas antes descritas para la amplificación del extremo 3'.

Los resultados de la reacción de PCR para ambos extremos se visualizaron en un gel de agarosa al 2%. Los fragmentos seleccionados se purifican desde el gel, para posteriormente ser ligados al vector pGEM®-T Easy. Luego, con el fin de obtener un mayor número de copias de los constructos vector-inserto, éstos fueron clonados en células *E.coli* DH5α. Finalmente, se realizó una extracción del DNA plasmidial desde las células y luego una digestión de éste con el fin de verificar que los fragmentos clonados son los deseados. Los detalles de estos protocolos se presentan en el Anexo 6.1.

#### **2.2.10 Secuenciamiento y análisis**

Las muestras con los fragmentos clonados en pGEM®-T Easy provenientes del DNA genómico, cDNA y RACE, fueron enviadas a secuenciar a Macrogen S.A. (Corea), para lo cual se utilizaron los partidores M13F-pUC y M13R-pUC (secuencias en el Anexo 6.1.16).

Con el fin de encontrar regiones de similitud con otras secuencias, los resultados fueron analizados con la herramienta BLASTX [32]. Además, se utilizó la herramienta CDD [34] para verificar la presencia de dominios conservados en las secuencias.

La herramienta BLASTX, utiliza como entrada una secuencia nucleotídica la cual compara contra una base de datos de proteínas, entregando como resultado un parámetro estadístico (e-value) el cual indica la probabilidad de que el alineamiento haya sido obtenido producto del azar. Mientras menor es el valor del e-value más significativo es el alineamiento. Además, la herramienta BLASTX permite conocer el marco de lectura con el cual se obtuvo el alineamiento, el cual puede ser útil para realizar análisis posteriores a partir de la secuencia aminoacídica. La herramienta CDD, utiliza como entrada una secuencia aminoacídica la cual comparara con una base de datos de dominios conservados de proteínas, entregando como resultado el grado de similitud que posee la secuencia con los dominios de la base de datos. El parámetro estadístico utilizado al igual que la herramienta BLASTX es el e-value.

Se consideraron como significativos todos los alineamientos que poseyesen un e-value de un orden menor a  $10^{-3}$ .



## **3. Resultados y Discusión**

### **3.1 Construcción de la micoteca**

La construcción de la micoteca permitió conocer los distintos ritmos de crecimiento de los hongos analizados. Se pudo observar que la gran mayoría de éstos lograban cubrir la placa en aproximadamente 7 días, sin embargo, existieron casos en que pasados 14 días la placa aún no se encontraba completamente cubierta. Cabe señalar, que debido a la necesidad de analizar una gran cantidad de hongos distintos al mismo tiempo, la temperatura utilizada para el crecimiento de éstos fue la misma (28°C). Considerando que las temperaturas óptimas de crecimiento de los distintos hongos son distintas, el utilizar una única temperatura no permitió que todos crecieran a su máxima velocidad.

Por otro lado, pese a extremar las precauciones para mantener las condiciones de esterilidad, existieron hongos que no fue posible cultivar libre de contaminaciones, por lo que fueron descartados. Se cree que estos últimos provenían de placas que ya se encontraban contaminadas y por ello no fue posible su cultivo en forma pura.

### **3.2 Identificación de hongos productores de endoglucanasas**

Con el fin de discriminar entre los hongos que poseían actividad y los que no, se realizaron ensayos de actividad endoglucanasa en medio sólido. Sin embargo, los resultados obtenidos a partir de éstos no fueron concluyentes. En algunos casos, la presencia del hongo en la placa no permitía visualizar en forma clara el resultado del ensayo (un halo sin tinción si es que existía actividad endoglucanasa). Por otro lado, el ensayo dependía en demasía de la velocidad de crecimiento del hongo, que en algunos casos era bastante baja.

Debido a lo anterior, se procedió a realizar ensayos de actividad a partir del sobrenadante de cultivos de hongos crecidos en medio líquido. Con el objetivo de inducir la producción de endoglucanasas, los hongos fueron cultivados en ausencia de glucosa y utilizando CMC como fuente de carbono.

Considerando que existía una gran variabilidad en la velocidad de crecimiento de los distintos hongos, se realizaron ensayos de actividad en una serie de días, lo cual permitió determinar el valor máximo de actividad para cada uno de éstos.

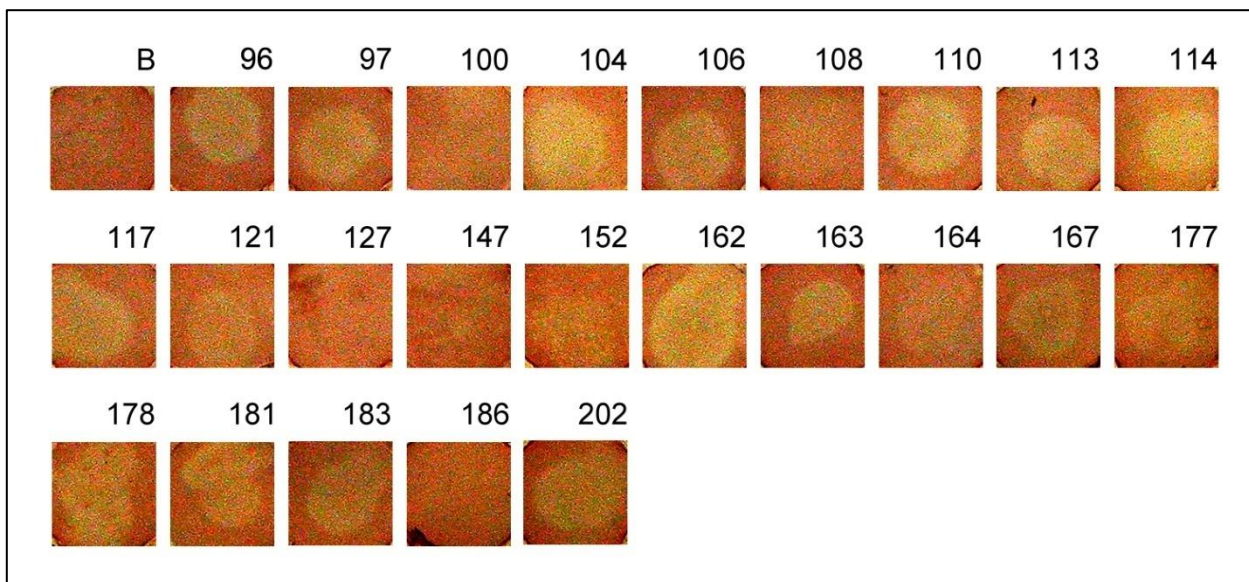
En la Figura 3.1 se presentan los valores máximos de actividad para cada hongo. El ensayo consistió en la medición de los azúcares reductores producidos luego de una reacción de hidrólisis sobre CMC. La elección de este sustrato se debió a que al poseer un alto grado de polimerización, se suele utilizar para la identificación de endoglucanasas, además, su solubilidad permite trabajar con éste en forma fácil [25]. Para cuantificar los azúcares reductores se utilizó el método DNS. La absorbancia obtenida es directamente proporcional a la cantidad de azúcares reductores generados por cada hongo (ver detalle del método en el punto 2.2.2.2). Los ensayos se realizaron luego de 3, 4, 5 y 7 días de crecimiento de los cultivos. Los datos brutos se presentan en el Anexo 6.2.1.



**Figura 3.1 Gráfico de máximos de actividad endoglucanasa sobre CMC para los distintos hongos.** Máximo valor de actividad endoglucanasa obtenido para cada hongo. B corresponde a una muestra proveniente de un medio de cultivo que no fue inoculado, mientras los distintos hongos se presentan con sus respectivas claves.

Además, utilizando el sobrenadante obtenido desde los hongos crecidos en medio líquido, se realizaron ensayos de actividad en geles de agarosa utilizando CMC al 1% como sustrato. Luego de una reacción de hidrólisis, la placa se tiñó con rojo Congo, para luego ser lavada con NaCl (el detalle del procedimiento se presenta en el Anexo 6.1.4). Los halos sin tinción corresponden a las zonas en que el CMC fue degradado. Se debe mencionar que este ensayo es de índole cualitativo, debido a que no es posible cuantificar la cantidad de CMC que es degradado. Se realizaron ensayos luego de 5 y 7 días de crecimiento de los cultivos, presentándose resultados similares. Los resultados correspondientes al séptimo día, se presentan en la Figura 3.2. Se puede observar que las muestras 100, 108, 127, 147, 152 y 186 no presentan generación de halos, por lo que no existió degradación de CMC.

Los resultados obtenidos a partir de ambos ensayos son concordantes, presentándose halos en todas las muestras que poseen sobre 30 U/l de actividad endoglucanasa. A partir de los resultados obtenidos, considerando que poseían máximos de actividad considerablemente superiores al resto, se seleccionaron los siguientes hongos para seguir trabajando: 106, 110, 113, 117 y 162.



**Figura 3.2 Ensayo de actividad endoglucanasa sobre medio sólido.** Ensayo de actividad sobre CMC al 1%, utilizando el sobrenadante proveniente del séptimo día de cultivo. Los halos sin tinción corresponden a las zonas en que el CMC fue degradado. B corresponde a una muestra proveniente de un medio de cultivo que no fue inoculado, mientras los distintos hongos se presentan con sus respectivas claves.

Cabe señalar, que los ensayos utilizados no permiten determinar cuáles son los hongos que poseen las enzimas con mayor actividad endoglucanasa. Esto debido a que para realizarlos, se utiliza el sobrenadante de los cultivos, encontrándose presente en éstos todas las enzimas extracelulares producidas por el hongo. Por lo que por ejemplo, un hongo que produce una alta cantidad de enzima de baja actividad específica, puede presentar una actividad mayor que uno que produce poca enzima de alta actividad específica. Para poder determinar el hongo que produce las enzimas de mayor actividad, sería necesario purificar la enzima para poder calcular su actividad específica, sin embargo, para esta etapa lo importante era asegurar la presencia de enzimas con actividad endoglucanasa, para lo cual la metodología utilizada dio resultados satisfactorios.

### 3.3 Diseño de partidores degenerados.

El diseño de partidores degenerados se basó en la utilización de regiones de homología presentes en secuencias aminoacídicas de enzimas ya descritas. Se realizó una búsqueda de secuencias aminoacídicas, utilizando como criterio de búsqueda el EC 3.2.1.4, correspondiente a la actividad

enzimática endoglucanasa. Se encontraron secuencias pertenecientes a 13 familias, esto según la clasificación de glicosil hidrolasas basada en la similitud de secuencias aminoacídicas descrita por Henrissat [29]. Las familias fueron las siguientes: 5, 6, 7, 8, 9, 10, 12, 44, 45, 48, 51, 61 y 74. Sin embargo, en varias de éstas sólo se encontraron secuencias aminoacídicas correspondientes a enzimas de bacterias, por lo que las familias 8, 44, 48, 51 y 74 fueron descartadas.

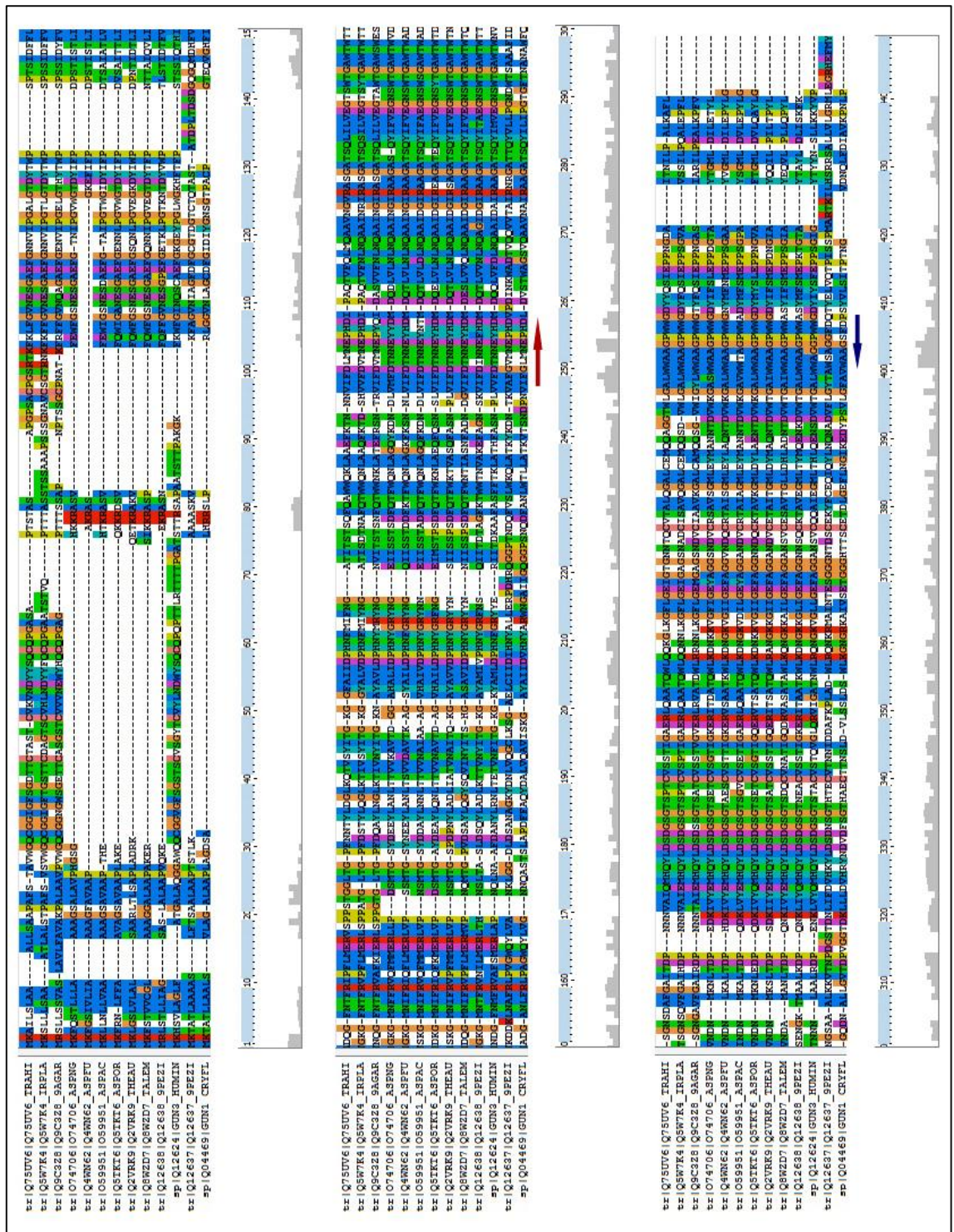
A partir de la base de datos de EMBL-EBI, se obtuvieron todas las secuencias aminoacídicas que cumplieran el siguiente criterio de búsqueda: poseer actividad endoglucanasa (EC 3.2.1.4), ser proveniente de hongos y pertenecer a una de las siguientes familias de: 5, 6, 7, 9, 10, 12, 45 y 61. Tras la realización de la búsqueda, las familias que poseían un mayor número de secuencias descritas fueron seleccionadas para seguir trabajando, esto con el fin de asegurar regiones de homología más representativas. Las familias 5, 7, 12 y 61 fueron las seleccionadas.

Mediante la utilización de la herramienta ClustalX2 [35], se realizó un alineamiento de las secuencias aminoacídicas correspondientes a cada familia de GHs. A partir del alineamiento, utilizando la estrategia CODEHOP [36] descrita en el punto 2.2.3, se diseñaron partidores degenerados para cada familia. Se obtuvieron como resultado una serie de partidores para las direcciones *forward* y *reverse*, seleccionándose parejas de partidores que tenían Tm similares y generaban un amplicón de entre 300 y 600 pb, de modo de facilitar la reacción de PCR. Los partidores diseñados se muestran en la Tabla 2.6.

En la Figura 3.3 se presenta el alineamiento correspondiente a las secuencias aminoacídicas seleccionadas pertenecientes a la familia 5 de GHs. Además, en la misma figura, se presentan las regiones de homología para las cuales fueron diseñados los partidores degenerados. Los alineamientos generados para las familias de GHs 7, 12 y 61 y las regiones de homología a partir de las cuales fueron diseñados los partidores, se presentan en el Anexo 6.2.2.

Como se puede observar en los alineamientos, las zonas utilizadas para diseñar los partidores degenerados presentan un alto grado de homología, sin embargo, en el resto de la secuencia la homología es menor. Considerando lo anterior, se espera que la utilización de esta estrategia permita obtener nuevas proteínas que posean actividad endoglucanasa, sin necesariamente tener un alto grado de homología con las enzimas ya descritas.

A partir de los alineamientos generados, fue posible calcular un largo esperado de amplicón para las distintas familias, éstos se presentan en la Tabla 3.1 .



**Figura 3.3 Alineamiento de secuencias de endoglucanasas pertenecientes a la familia 5 de GHs.** Alineamiento realizado utilizando la herramienta ClustalX2. Debajo de cada alineamiento se presenta una regla que indica la posición de los aminoácidos. Las columnas coloreadas indican conservación del residuo aminoacídico en esa posición. Las flechas roja y azul corresponden a la región de homología a partir de las cuales se diseñaron los partidores degenerados *celF\_endo\_fwd5* y *celF\_endo\_rev5* respectivamente.

**Tabla 3.1 Largo esperado de amplicones para las distintas familias de GHs.**

<b>Familia</b>	<b>Largo amplicón [pb]</b>
5	500
7	360
12	530
61	300

### **3.4 Amplificación de fragmentos mediante PCR desde DNA genómico**

La estrategia utilizada para realizar la reacción de PCR consistió en un programa con ciclos iniciales a una temperatura de alineamiento baja, para luego pasar a ciclos a una temperatura de alineamiento más alta.

Considerando que se estaban utilizando partidores degenerados, los cuales no eran perfectamente complementarios al DNA templado, los ciclos iniciales a baja temperatura permitían enriquecer la muestra con templado proveniente de hibridaciones inespecíficas. Luego, al aumentar la temperatura de alineamiento, la especificidad de la hibridación de los partidores aumentaba, promoviendo la amplificación de los fragmentos de interés.

Cabe señalar, que debido a la utilización de partidores degenerados, no fue fácil determinar la temperatura de alineamiento para los segundos ciclos. Esto, ya que la  $T_m$  de los partidores degenerados no era un valor fijo, sino que se encontraba dentro de un rango amplio. Debido a lo anterior, se utilizaron gradientes de temperatura, los cuales permitían analizar una gran cantidad de condiciones al mismo tiempo. La utilización de gradientes, también permitió analizar muestras pertenecientes a distintos hongos en forma simultánea, conllevando a una disminución en el tiempo utilizado en la realización de las reacciones.

Los principales resultados obtenidos utilizando la estrategia antes descrita, se presentan a continuación. En las Figuras 3.4, 3.5 y 3.6 se pueden observar los productos de PCR obtenidos utilizando los partidores degenerados diseñados para las familias de GHs 5, 7 y 61 respectivamente.

Por otro lado, como se mencionó anteriormente, con el fin de aumentar la eficiencia y especificidad de las reacciones de PCR, se utilizó DMSO como agente adyuvante. Éste facilita la separación de las hebras de DNA al impedir el apareamiento de las bases [44]. Su utilización permitió la aparición de nuevas bandas en algunos casos, sin embargo, en otros éstas desaparecieron. En las Figuras 3.7, 3.8 y 3.9, se presentan los resultados para las familias de GHs 5, 12 y 61 respectivamente.

Las bandas seleccionadas para su purificación se encuentran señaladas con flechas blancas, las cuales están acompañadas de un número para facilitar su identificación. El criterio utilizado para realizar la selección, fue la intensidad de éstas, priorizando por las que fuesen más intensas. En la Tabla 3.2 se presenta un resumen de las bandas seleccionadas para ser purificadas.

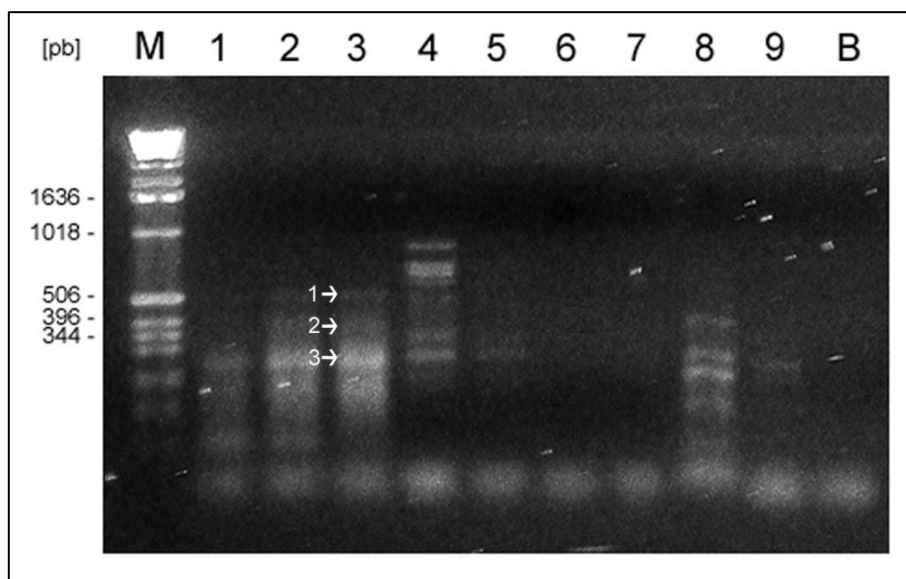
**Tabla 3.2 Resumen de fragmentos amplificados desde DNA genómico de los hongos.**

<b>Fragmento</b>	<b>Partidores*</b>	<b>Hongo</b>	<b>Largo aproximado [pb]</b>	<b>Nombre</b>
1	Familia 5	106	500	f5h106A
2	Familia 5	106	400	f5h106B
3	Familia 5	106	280	f5h106C
4	Familia 7	162	400	f7h162A
5	Familia 61	162	380	f61h162A
6	Familia 61	162	280	f61h162B
7	Familia 61	162	220	f61h162C
8	Familia 5	117	900	f5h117A
9	Familia 5	117	750	f5h117B
10	Familia 5	117	500	f5h117C
11	Familia 5	117	350	f5h117D
12	Familia 5	117	300	f5h117E
13	Familia 5	162	530	f5h162A
14	Familia 5	162	450	f5h162B
15	Familia 5	162	300	f5h162C
16	Familia 12	110	400	f12h110A
17	Familia 61	110	530	f61h110A
18	Familia 61	110	450	f61h110B

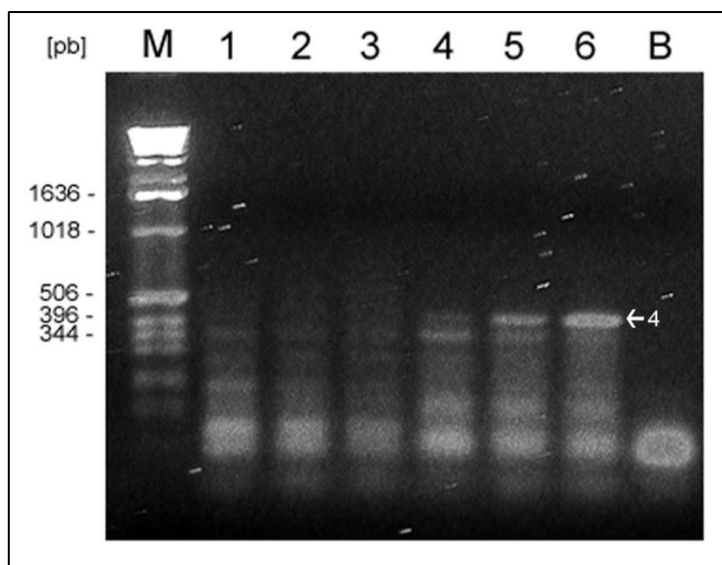
\* Familias de glicosil hidrolasas según Henrissat [29].

En la Tabla 3.2 se puede observar que los fragmentos poseen largos variados, sin coincidir necesariamente con los largos esperados para las distintas familias (ver Tabla 3.1). Esto se puede deber a que como la amplificación se realizó desde DNA genómico proveniente de organismos eucariontes, es probable que las secuencias presenten intrones, los cuales aumentarían el largo esperado para ellas. Por otro lado, existe la posibilidad de hibridación inespecífica por parte de los partidores, uniéndose a regiones que no corresponden a los sitios de homología para los cuales fueron diseñados, esto implicaría la amplificación de fragmentos que no son de interés.

Desde cada banda purificada, se obtuvo dos clones con el fragmento inserto en el vector pGEM®-T Easy. A partir de éstos, se extrajo el DNA plasmidial, al cual se le realizó una restricción para verificar que el inserto era el correcto. Las muestras verificadas fueron enviadas a secuenciar a Macrogen S.A. (Corea) para su posterior análisis. El detalle del procedimiento se encuentra en el Anexo 6.1.

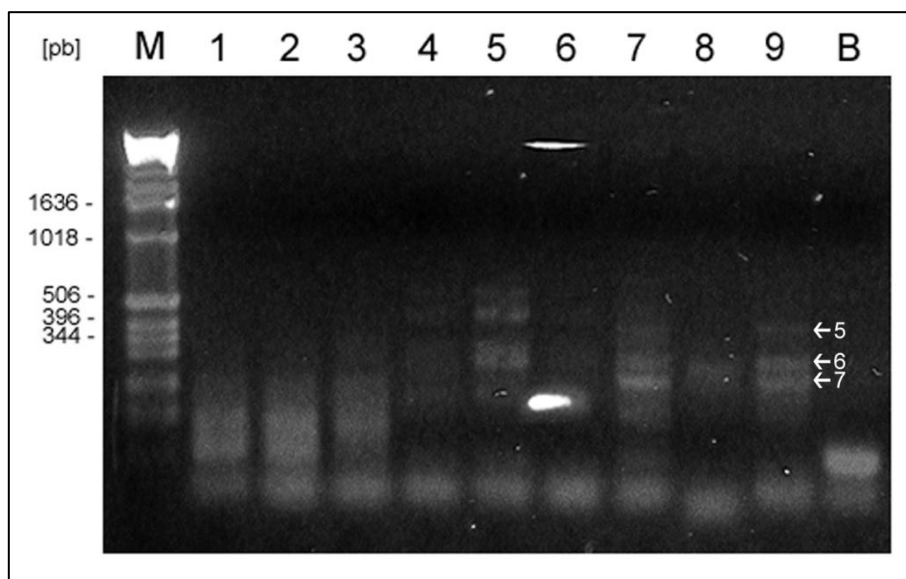


**Figura 3.4 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 5 de GHs.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 106; columnas del 4 al 6: hongo 117; columnas del 7 al 9: hongo 162; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1, 4 y 7: 70,3°C; columnas 2, 5 y 8: 72,5°C; columnas 3, 6 y 9: 74,6°C. Los productos seleccionados (indicados con flechas blancas) fueron purificados para luego ser secuenciados.

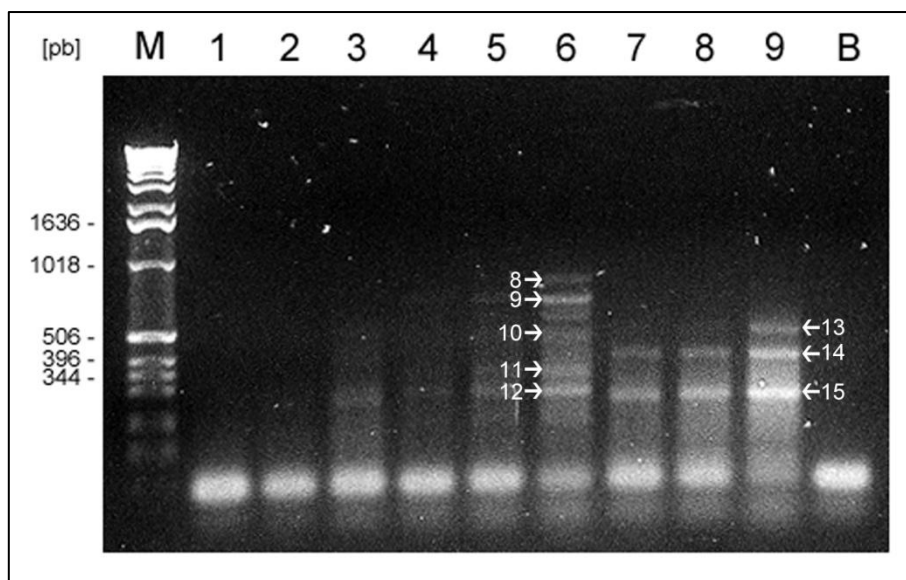


**Figura 3.5 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 7 de GHs.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 110; columnas del 4 al 6: hongo 162; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1 y 4: 70,3°C; columnas 2 y 5: 72,5°C; columnas 3 y 6: 74,6°C. El producto seleccionado (indicado con una flecha blanca), fue purificado para luego ser secuenciado.

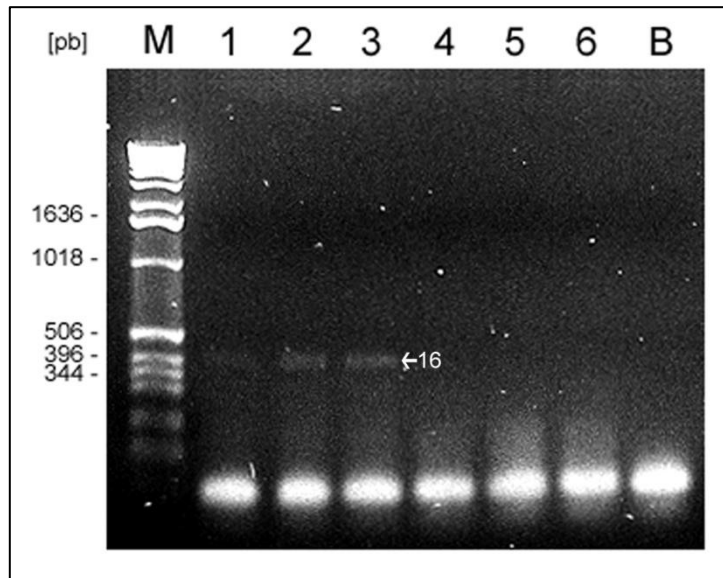




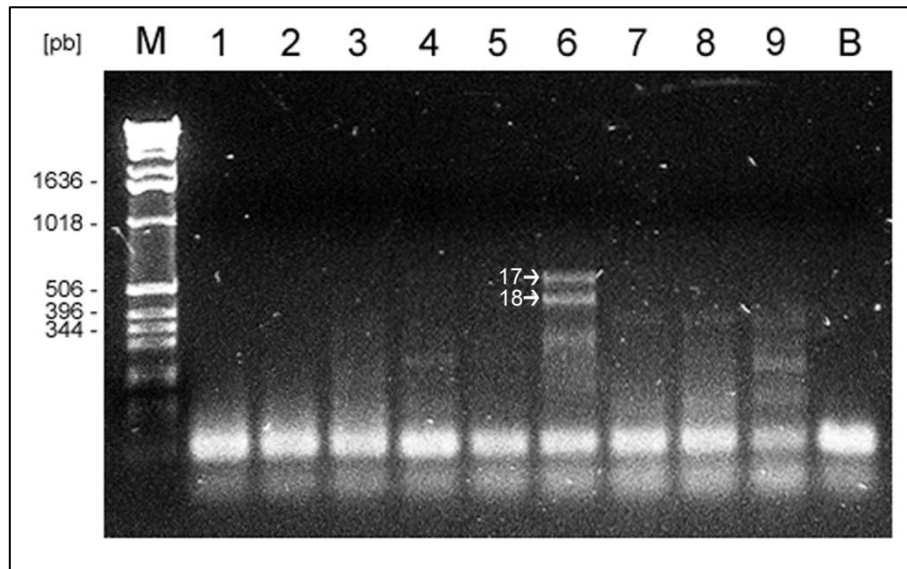
**Figura 3.6 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 61 de GHs.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 106; columnas del 4 al 6: hongo 110; columnas del 7 al 9: hongo 162; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1, 4 y 7: 70,3°C; columnas 2, 5 y 8: 72,5°C; columnas 3, 6 y 9: 74,6°C. Los productos seleccionados (indicados con flechas blancas) fueron purificados para luego ser secuenciados.



**Figura 3.7 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 5 de GHs con DMSO.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 106; columnas del 4 al 6: hongo 117; columnas del 7 al 9: hongo 162; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1, 4 y 7: 62,7°C; columnas 2, 5 y 8: 64,9°C; columnas 3, 6 y 9: 66,9°C. Los productos seleccionados (indicados con flechas blancas) fueron purificados para luego ser secuenciados.



**Figura 3.8 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 12 de GHs con DMSO.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 110; columnas del 4 al 6: hongo 162; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1 y 4: 58,4°C; columnas 2 y 5: 60,5°C; columnas 3 y 6: 62,7°C. El producto seleccionado (indicado con una flecha blanca), fue purificado para luego ser secuenciado.



**Figura 3.9 Amplificación de fragmentos de DNA genómico utilizando partidores degenerados para la familia 61 de GHs con DMSO.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 106; columnas del 4 al 6: hongo 110; columnas del 7 al 9: hongo 162; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1, 4 y 7: 62,7°C; columnas 2, 5 y 8: 64,9°C; columnas 3, 6 y 9: 66,9°C. Los productos seleccionados (indicados con flechas blancas) fueron purificados para luego ser secuenciados.

### 3.4.1 Análisis de fragmentos clonados provenientes desde DNA genómico.

Los resultados de los secuenciamientos fueron analizados utilizando las herramientas BLASTX [32] y CDD[34], como se indicó en el punto 2.2.10 de Metodología. La herramienta BLASTX, permite compara secuencias nucleotídicas con bases de datos de proteínas, entregando como resultado un parámetro estadístico (e-value) que indica qué tan probable es que la analogía presentada sea producto del azar; mientras más cercano se encuentre este valor a 0, el alineamiento tiene un mayor grado de confianza. La herramienta CDD, permite comparar secuencias aminoacídicas con una base de datos de dominios conservados, utilizando al igual que la herramienta BLASTX, el parámetro e-value para evaluar el grado de confianza del alineamiento.

Previo a realizar los análisis, las secuencias debieron ser “limpiadas”, lo cual consistió en eliminar de éstas las secuencias correspondientes al vector de clonación que flanqueaban el inserto.

A continuación se presentan los resultados obtenidos. El número que acompaña al nombre, corresponde al clon del cual se obtuvo el inserto.

#### **f5h117A-1:**

El análisis con BLASTX indicó que esta secuencia tiene alta similitud (e-value del orden de  $10^{-42}$ ) con proteínas predichas para el hongo *Postia placenta*, sin embargo, se desconoce la función de éstas. Al utilizar la herramienta CDD, no se encontraron dominios conservados.

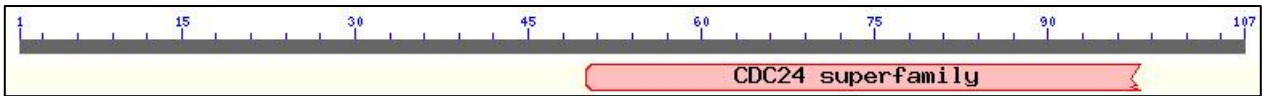
#### **f5h117C-1:**

El análisis con BLASTX indicó que esta secuencia posee alta similitud (e-value del orden de  $10^{-28}$ ) con la proteína hipotética CC1G-03926 proveniente del hongo *Coprinopsis cinerea*, la cual posee dominios conservados de proteína hamartin (supresor de tumores) y proteínas de segregación cromosómica SMC. Al realizar una búsqueda con la herramienta CDD para la muestra f5h117C-1, no se encontraron dominios conservados.

#### **f5h162B-1:**

El análisis con BLASTX indicó que esta secuencia posee alta similitud (e-value del orden de  $10^{-19}$ ) con una proteína predicha para el hongo *Postia placenta*, sin embargo, se desconoce la función de ésta. Al realizar una búsqueda de dominios conservados utilizando el marco de lectura -3, se encontró que la secuencia tiene similitud con la superfamilia CDC24 (e-value del orden de  $10^{-6}$ ). Particularmente, posee un dominio de homología a calponin, el cual se encuentra presente en proteínas de unión a actina, las cuales tienen rol importante en la contracción muscular [45]. En Figura 3.10 la se presenta la región de similitud con la superfamilia, no obstante, el dominio

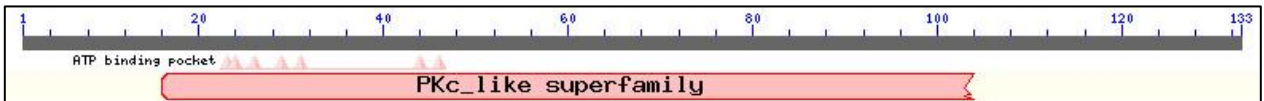
catalítico no se encuentra completo, esto se deduce a partir de los bordes irregulares presentes en el extremo derecho de la barra que representa a la superfamilia.



**Figura 3.10 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f5h162B-1.** Se utilizó el marco de lectura -3. La barra gris representa a la secuencia f5h162B-1, mientras que la superfamilia CDC24 es representada con una barra roja. La regla azul indica la posición de los residuos aminoacídicos.

#### **f7h162A-2:**

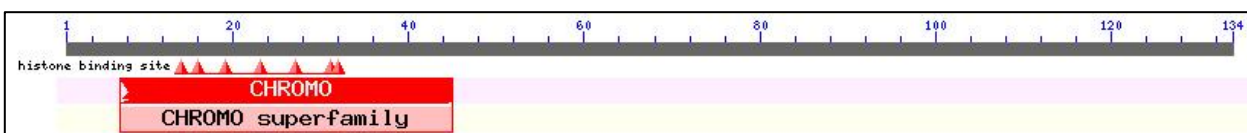
El análisis utilizando la herramienta BLASTX indicó que esta secuencia posee alta similitud (e-value del orden de  $10^{-20}$ ) con una proteína predicha para el hongo *Laccaria bicolor*, sin embargo, se desconoce el rol de ésta. Al realizar una búsqueda de dominios conservados utilizando el marco de lectura +1 mediante la herramienta CDD, se encontró que la secuencia posee alta similitud (e-value del orden de  $10^{-15}$ ) con la superfamilia PKc-like. Esta familia está compuesta principalmente por los dominios catalíticos de proteínas kinasas, las cuales catalizan la transferencia del grupo  $\gamma$ -fosforil desde el ATP a grupos hidroxilos de sustratos específicos. En la Figura 3.11 se presenta la región de similitud con la superfamilia, pudiéndose apreciar que el dominio catalítico no se encuentra completo, esto se deduce del borde irregular presente en el extremo derecho de la barra que representa a la superfamilia. Además, se puede observar la posición de 7 pocillos de unión a ATP asociados a la superfamilia, éstos se encuentran indicados con triángulos rojos.



**Figura 3.11 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f7h162A-2.** Se utilizó el marco de lectura +1. La barra gris representa a la secuencia f7h162A-2, mientras la superfamilia PKc\_like es representada con una barra roja. Los triángulos rosados corresponden a sitios de unión a ATP. La regla azul indica la posición de los residuos aminoacídicos.

#### **f61h110B-2:**

El análisis con BLASTX indicó que esta secuencia posee alta similitud (e-value del orden de  $10^{-15}$ ) con la proteína hipotética CC1G\_07263 proveniente del hongo *Coprinopsis cinerea*, sin embargo, se desconoce el rol de ésta. Se realizó una búsqueda de dominios conservados con la herramienta CDD a partir del marco de lectura +1, obteniéndose similitud (e-value del orden de  $10^{-9}$ ) con la superfamilia CHROMO. El dominio CHROMO se encuentra en una variedad de proteínas cromosomales, teniendo aparentemente un rol en la organización funcional del núcleo eucariota. En la Figura 3.12 se presenta la región de similitud con la superfamilia, correspondiendo los triángulos rojos a sitios de unión de histonas.



**Figura 3.12 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f61h110B-2.** Se utilizó el marco de lectura +1. La barra gris representa a la secuencia f61h110B-2, mientras la superfamilia CHROMO es representada con una barra roja. Los triángulos rojos corresponden a sitios de unión a histonas. La regla azul indica la posición de los residuos aminoacídicos.

Cabe señalar, que los fragmentos obtenidos, se generaron a partir de DNA genómico, existiendo la posibilidad de que los dominios conservados se vean interrumpidos por la presencia de intrones.

En la Tabla 3.3 se presenta un resumen del análisis realizado de los fragmentos obtenidos desde el DNA genómico. Para los fragmentos restantes, no se encontraron alineamientos significativos.

**Tabla 3.3 Resumen de resultados obtenidos para fragmentos encontrados en el DNA genómico.**

Fragmento	Mejor alineamiento			Dominio conservado	
	Organismo	e-value	Marco	Superfamilia	e-value
f5h117A-1	<i>Postia placenta</i>	10 <sup>-42</sup>	+1	No se encontró	-
f5h117C-1	<i>Coprinopsis cinerea</i>	10 <sup>-28</sup>	-3	No se encontró	-
f5h162B-1	<i>Postia placenta</i>	10 <sup>-19</sup>	-3	CDC24	10 <sup>-6</sup>
f7h162A-2	<i>Laccaria bicolor</i>	10 <sup>-20</sup>	+1	PKc-like	10 <sup>-15</sup>
F61h110B-2	<i>Coprinopsis cinerea</i>	10 <sup>-15</sup>	+1	CHROMO	10 <sup>-9</sup>

Como se puede observar, no se obtuvieron secuencias pertenecientes a las familias de endoglucanasas hacia las cuales estaban dirigidos los partidores degenerados. Una posible explicación para esto, es la baja especificidad de la estrategia utilizada. Al utilizarse partidores degenerados, sumado a las bajas condiciones de astringencia a las cuales se realizaron las reacciones de PCR, es probable que se produjesen hibridaciones demasiado inespecíficas, promoviendo la amplificación de fragmentos de DNA no deseados. Otra posible explicación, es que considerando que se utilizó DNA genómico como templado y que los partidores se diseñaron a partir de secuencias de homología aminoacídica, es posible que estas últimas se encontrasen interrumpidas por intrones, impidiendo la hibridación de los partidores. Otro factor que puede ser importante, es que los genes se encuentren presentes en un bajo número de copias, dificultando el encuentro con los partidores.

### 3.5 Aislamiento de genes de endoglucanasas a partir de RNA/cDNA

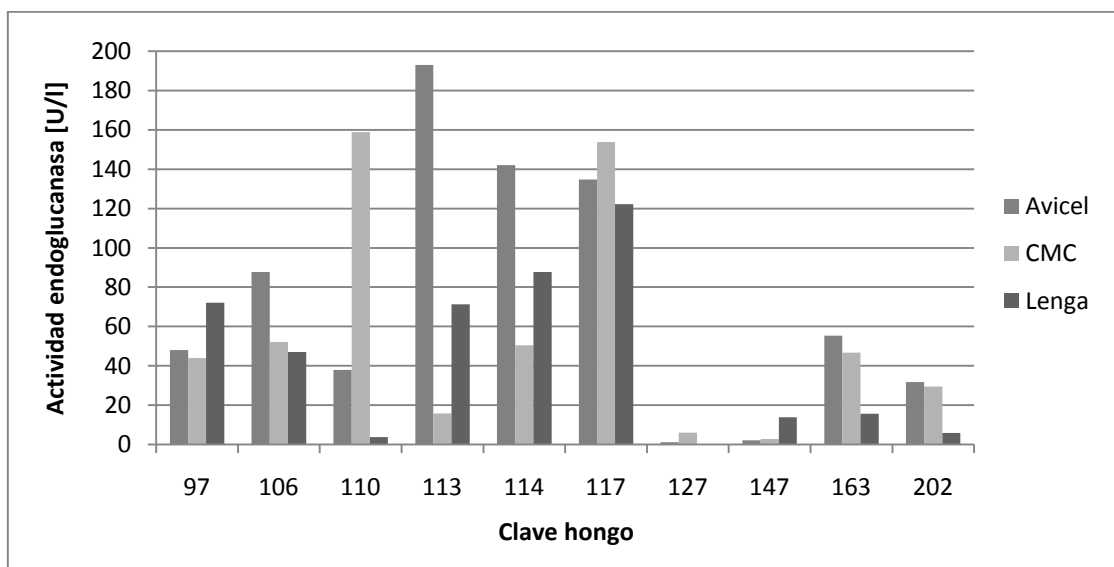
Teniendo en consideración los resultados obtenidos a partir del DNA genómico, se planteó una nueva estrategia para poder obtener las secuencias de las endoglucanasa presentes en los hongos.

Ésta se basa en la utilización de cDNA como templado para las reacciones de PCR. Sus principales ventajas son: la posibilidad de tener un alto número de copias de las secuencias que se desea amplificar (aumentando la cantidad de mRNA codificante utilizando un medio inductor de endoglucanasas) y la ausencia de intrones debido a que se parte desde el mRNA.

### 3.5.1 Selección de medio inductor y segunda selección.

Con el fin de determinar el mejor medio inductor de endoglucanasas para cada hongo, se realizaron ensayos de actividad a partir de cultivos crecidos utilizando distintas fuentes de carbono. Las fuentes de carbono testeadas fueron el sustrato soluble CMC y los sustratos insolubles Avicel y Lengua en astillas. La utilización de estos sustratos se debió a que para que el hongo produzca celulasas en forma considerable, es necesaria la presencia de celulosa en su medio de cultivo [26]. En particular, la utilización de Lengua se debió a que las enzimas que se buscan serán utilizadas en sustratos naturales como éste.

Las proteínas presentes en los cultivos fueron obtenidas mediante la utilización de un buffer de extracción (ver Anexo 6.1.12). La actividad de éstas fue testada mediante un ensayo de actividad endoglucanasa en medio líquido, el cual consistió en la medición de los azúcares reductores producidos luego de una reacción de hidrólisis de CMC.



**Figura 3.13 Selección de medio inductor de endoglucanasas para los distintos hongos.** Hongos crecidos utilizando Avicel, CMC y Lengua como fuente de carbono. Ensayo de actividad en medio líquido sobre CMC. Absorbancia medida a 550 nm, siendo ésta proporcional a la cantidad de azúcares reductores producidos en la reacción de hidrólisis de CMC

En la Figura 3.13 se presentan los resultados del ensayo. Se puede observar que los hongos 127, 147 y 202 obtuvieron bajos niveles de actividad en todos los sustratos, por lo que fueron descartados. En

el caso del hongo 110, se observa un alto nivel de actividad cuando se cultivó en CMC, sin embargo, al utilizar Lenga como sustrato casi no presentó actividad, por lo que fue descartado. Esta última decisión fue tomada considerando que se espera que las enzimas sean capaz de hidrolizar este sustrato en forma eficiente. Para el resto de los hongos el mejor medio de inducción fue Avicel, salvo los hongos 97 y 117 que alcanzaron su mayor actividad al cultivarse en Lenga y CMC respectivamente. A partir de lo anterior, los hongos seleccionados para seguir trabajando fueron: 97, 106, 113, 114, 117 y 163. En la Tabla 3.4 se presenta un resumen con los hongos seleccionados y sus respectivos medios inductores.

**Tabla 3.4 Medios de inducción seleccionados para los distintos hongos.**

<b>Clave hongo</b>	<b>Medio inductor</b>
97	Lenga
106	Avicel
113	Avicel
114	Avicel
117	CMC
163	Avicel

### **3.5.2 Amplificación de fragmentos mediante PCR desde cDNA**

Los hongos seleccionados fueron cultivados en sus respectivos medios inductores, con el fin de enriquecer la producción de mRNA codificante para endoglucanasas. Una vez que el crecimiento de éstos fue el deseado, se procedió a la extracción del RNA. A partir de éste, se sintetizó cDNA utilizando la enzima transcriptasa reversa. Para ver el detalle de estos procedimientos, consultar el punto 2.2 de Metodología. No se pudo obtener RNA a partir del hongo 97, por lo que fue descartado para el análisis.

La estrategia para realizar la reacción de PCR fue análoga a la utilizada para la amplificación desde el DNA genómico, es decir, se utilizó un programa con ciclos iniciales con baja temperatura de alineamiento y luego ciclos con una temperatura de alineamiento más alta (utilizando un gradiente de temperaturas, con el fin de ensayar varias condiciones en forma simultánea).

Los resultados de las reacciones de PCR se presentan a continuación. Las bandas seleccionadas, se indican con una flecha blanca acompañada de un número para facilitar su identificación. En este caso, la amplificación se llevó a cabo a partir de cDNA, por lo que no debiese existir presencia de intrones. Considerando esto último, el criterio de selección se basó tanto en la intensidad de la banda, como en su tamaño, el cual debiese ser acorde a lo estimado en la Tabla 3.1.

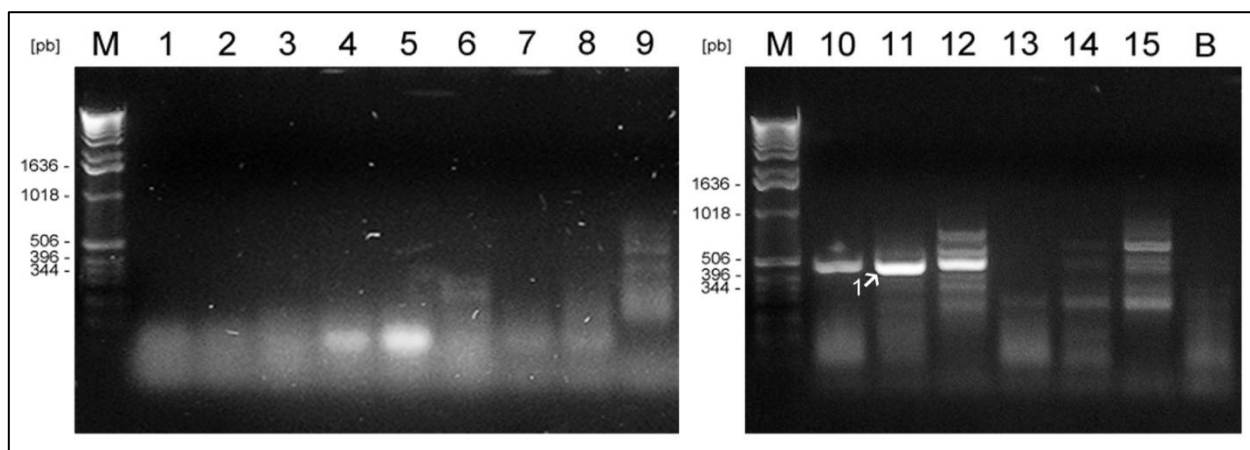
**Tabla 3.5 Resumen de fragmentos obtenidos desde cDNA.**

Fragmentos	Partidores*	Hongo	Largo aproximado [pb]	Nombre
1	Familia 5	106	500	f5h106
2	Familia 7	106	340	f7h106
3	Familia 7	163	340	f7h163
4	Familia 12	117	350	f12h117
5	Familia 61	117	320	f61h117
6	Familia 61	163	320	f61h163

\* Familias de glicosil hidrolasas según Henrissat [29].

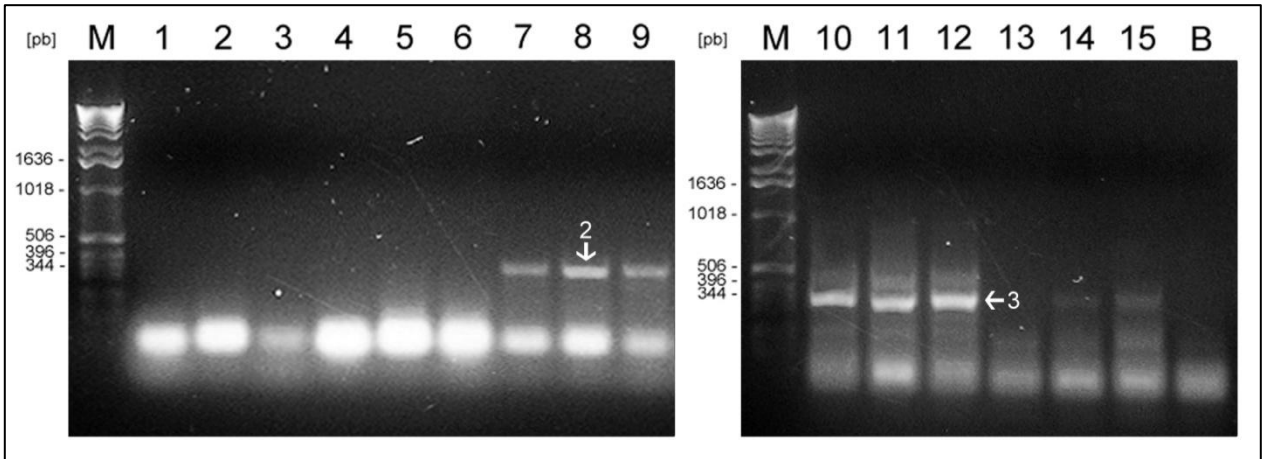
En la Tabla 3.5 se presenta un resumen con las bandas seleccionadas para ser purificadas.

A partir de cada banda purificada, se obtuvieron dos clones con el fragmento inserto en el vector pGEM®-T Easy. El DNA plasmidial de éstos fue extraído y enviado a secuenciar a Macrogen S.A. (Corea) para su posterior análisis. Para mayor detalle del procedimiento consultar Anexo 6.1.

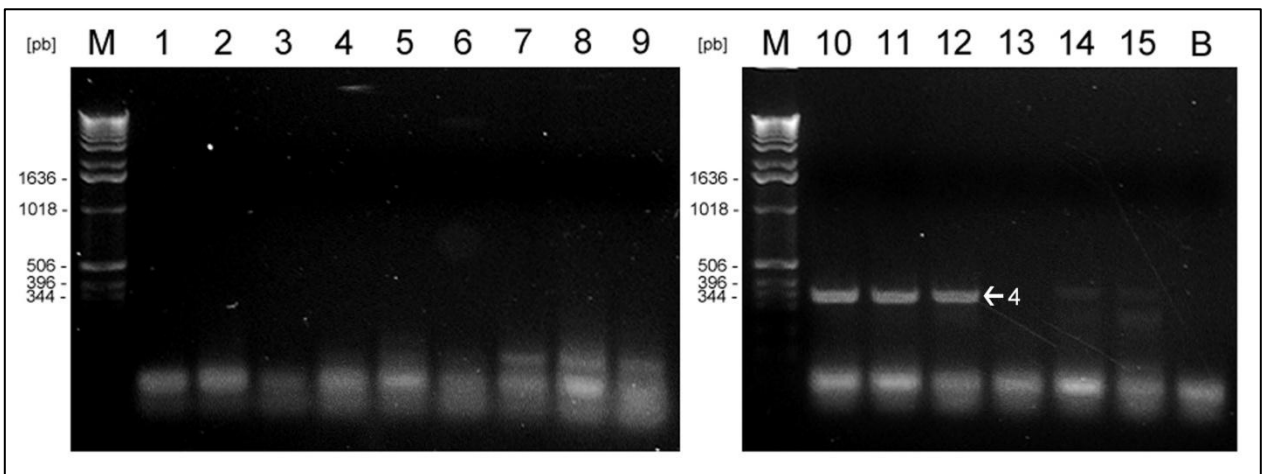


**Figura 3.14 Amplificación de fragmentos de cDNA utilizando partidores degenerados para la familia 5 de GHs.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 113; columnas del 4 al 6: hongo 114; columnas del 7 al 9: hongo 163; columnas del 10 al 12: hongo106; columnas del 13 al 15: hongo 117; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1, 4, 7, 10 y 13: 60,5°C; columnas 2, 5, 8, 11 y14: 64,9°C; columnas 3, 6, 9, 12 y 15: 70,5°C. El producto seleccionado (indicado con flecha blanca) fue purificado para luego ser secuenciado.

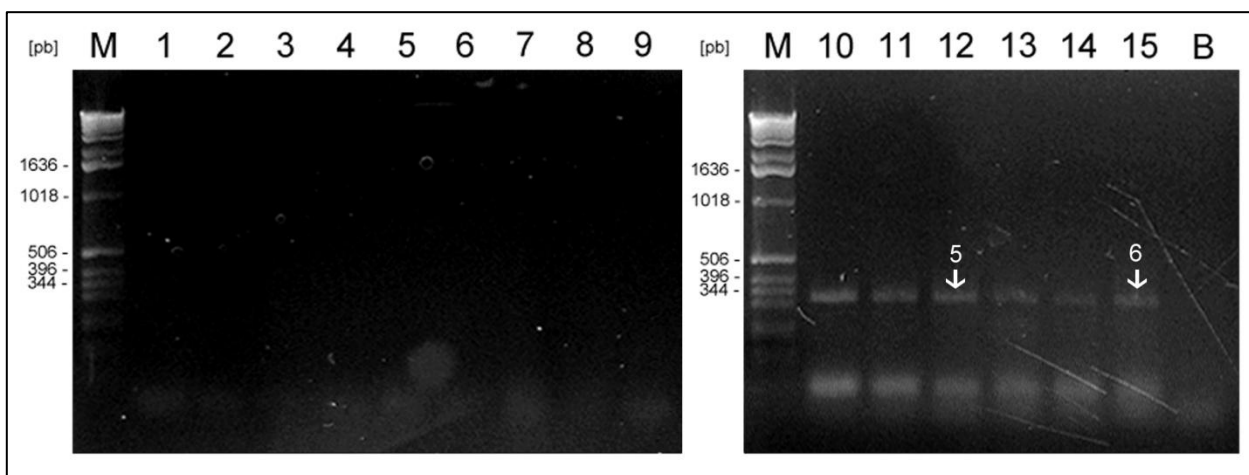




**Figura 3.15 Amplificación de fragmentos de cDNA utilizando partidores degenerados para la familia 7 de GHs.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 113; columnas del 4 al 6: hongo 114; columnas del 7 al 9: hongo 163; columnas del 10 al 12: hongo 106; columnas del 13 al 15: hongo 117; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1, 4, 7, 10 y 13: 60,5°C; columnas 2, 5, 8, 11 y 14: 64,9°C; columnas 3, 6, 9, 12 y 15: 70,5°C. Los productos seleccionados (indicados con flechas blancas) fueron purificados para luego ser secuenciados.



**Figura 3.16 Amplificación de fragmentos de cDNA utilizando partidores degenerados para la familia 12 de GHs.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 106; columnas del 4 al 6: hongo 113; columnas del 7 al 9: hongo 114; columnas del 10 al 12: hongo 117; columnas del 13 al 15: hongo 163; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1, 4, 7, 10 y 13: 60,5°C; columnas 2, 5, 8, 11 y 14: 64,9°C; columnas 3, 6, 9, 12 y 15: 70,5°C. El producto seleccionado (indicado con flecha blanca) fue purificado para luego ser secuenciado.



**Figura 3.17 Amplificación de fragmentos de cDNA utilizando partidores degenerados para la familia 61 de GHs.** Resultado de electroforesis de productos de PCR en un gel de agarosa al 1,5%. Columna M: Marcador de peso molecular 1 kb; columnas del 1 al 3: hongo 106; columnas del 4 al 6: hongo 113; columnas del 7 al 9: hongo 114; columnas del 10 al 12: hongo 117; columnas del 13 al 15: hongo 163; B: control negativo. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1, 4, 7, 10 y 13: 60,5°C; columnas 2, 5, 8, 11 y 14: 64,9°C; columnas 3, 6, 9, 12 y 15: 70,5°C. Los productos seleccionados (indicados con flechas blancas) fueron purificados para luego ser secuenciados.

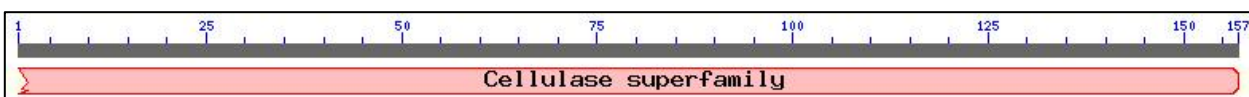
### 3.5.3 Secuenciamiento y análisis de fragmentos provenientes desde cDNA.

Para realizar el análisis, tal como en el caso de las secuencias encontradas a partir del DNA genómico, se utilizaron las herramientas BLASTX y CDD.

A continuación se presentan los resultados obtenidos. El número que acompaña al nombre, corresponde al clon del cual se obtuvo el inserto.

#### f5h106-1:

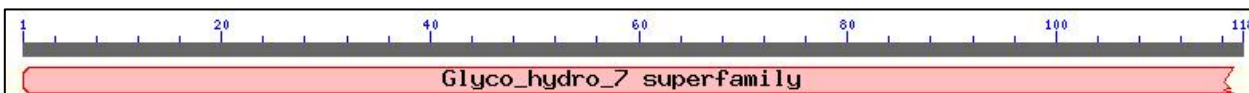
Mediante la utilización de la herramienta BLASTX, se obtuvo que la secuencia posee un alto grado de homología con endoglucanasas, siendo una del hongo *Trametes hirsuta* con la cual obtuvo el mejor alineamiento (e-value del orden de  $10^{-70}$ ), con un 93% de identidad a nivel de aminoácidos. Además, se obtuvo que los mejores alineamientos se generaron con el marco de lectura -1, por lo que se procedió a realizar una búsqueda de dominios conservados con la herramienta CDD utilizando este marco. Se obtuvo que la secuencia posee un dominio de la superfamilia de celulasas, específicamente de GHs de la familia 5 con un alto grado de seguridad (e-value del orden de  $10^{-19}$ ). En la Figura 3.18 se puede observar que la secuencia íntegramente es parte del dominio conservado, el cual no se encuentra completo (borde irregular del extremo izquierdo de la barra que representa al dominio).



**Figura 3.18 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f5h106-1.** Se utilizó el marco de lectura -1. La barra gris representa a la secuencia f5h106-1, mientras la barra roja representa a la superfamilia de celulastas. La regla azul muestra la posición de los residuos aminoacídicos.

#### f7h106-1:

Utilizando la herramienta BLASTX, se obtuvo que la secuencia posee un alto grado de homología con secuencias de celobiohidrolasas, obteniendo el mejor alineamiento con una proteína proveniente del hongo *Polyporus arcularius* (e-value del orden  $10^{-51}$ ), con un 80% de identidad a nivel de aminoácidos. Utilizando el marco de lectura +2, se realizó una búsqueda de dominios conservados con la herramienta CDD, obteniéndose como resultado la presencia de un dominio de la superfamilia Glyco\_hydro\_7 (e-value del orden de  $10^{-58}$ ), correspondiente a GHs de la familia 7. En la Figura 3.19 se puede observar que el dominio de GHs de la familia 7 comprende toda la secuencia, sin embargo, éste se encuentra incompleto, lo que se ve reflejado en el extremo irregular de la barra que lo representa.



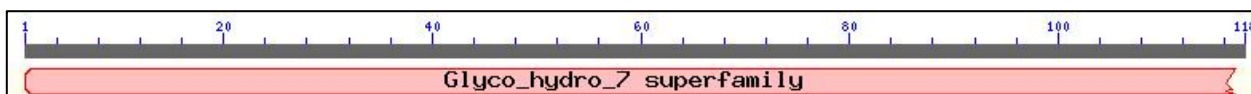
**Figura 3.19 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f7h106-1.** Se utilizó el marco de lectura +2. La barra gris representa a la secuencia f7h106-1, mientras la banda roja representa a la superfamilia de GHs de la familia 7. La regla azul muestra la posición de los residuos aminoacídicos.

#### f7h163-1 y f7h163-2:

Utilizando la herramienta BLASTX, se obtuvo que la secuencia posee alta similitud con endoglucanasas de la familia 7, en particular, su mejor alineamiento se registró con una proteína proveniente del mismo organismo del cual se obtuvo el fragmento, el hongo *Fusarium oxysporum* (e-value del orden de  $10^{-54}$ ), con un 97% de identidad a nivel de aminoácidos. Debido a lo anterior, se revisó el alineamiento entre las secuencias nucleotídicas, notando que diferían sólo en 3 nucleótidos, encontrándose éstos situados en un extremo del alineamiento. Considerando que se utilizaron partidores degenerados, los cuales pudieron haber insertado alguna mutación, las regiones correspondientes a ellos fueron removidas de la secuencia. A partir de la secuencia sin partidores, se volvió a utilizar la herramienta BLASTX, obteniéndose un 100% de identidad con la endoglucanasa antes mencionada (código de acceso P46237), siendo altamente probable que se trate de la misma enzima. Pese al aumento en el porcentaje de identidad, el e-value aumentó a un valor en torno a  $10^{-50}$  (es decir disminuyó la probabilidad de que sea la misma secuencia), esto se

debió a que al eliminar los partidores, el largo de la secuencia disminuyó, siendo éste un factor importante para el cálculo del parámetro.

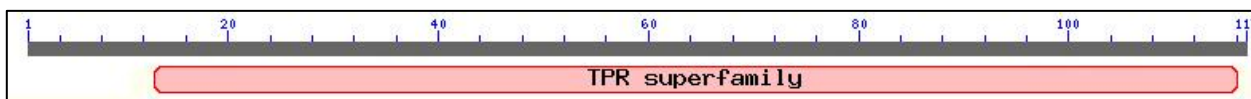
Al realizar una búsqueda de dominios conservados con la herramienta CDD, utilizando el marco de lectura +2, se obtuvo como resultado la presencia de un dominio de la superfamilia Glyco\_hydro\_7 (e-value del orden de  $10^{-41}$ ), correspondiente a GHs de la familia 7. En la Figura 3.20 se puede observar que el dominio de GHs de la familia 7 comprende toda la secuencia, sin embargo, éste se encuentra incompleto, lo cual se deduce del borde irregular presente en el extremo derecho de la barra que lo representa.



**Figura 3.20 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f7h163-1.** Se utilizó el marco de lectura +2. La barra gris representa a la secuencia f7h163-1, mientras la banda roja representa a la superfamilia de GHs de la familia 7. La regla azul muestra la posición de los residuos aminoacídicos.

#### f12h117-1:

El análisis utilizando la herramienta BLASTX, indica que esta secuencia posee alta similitud (e-value del orden de  $10^{-31}$ ) con una proteína predicha para el hongo *Laccaria bicolor*, sin embargo, se desconoce el rol de ésta. Al realizar una búsqueda de dominios conservados mediante la utilización de la herramienta CDD, a partir del marco de lectura -3, se encontró que la secuencia posee similitud (e-value del orden de  $10^{-3}$ ) con la superfamilia TPR. Ésta está compuesta por dominios repetidos de tetratricopéptidos, los cuales son importantes para las interacciones proteína-proteína. En la Figura 3.21 se puede observar, que el dominio TPR se encuentra en forma íntegra, siendo éste gran parte de la secuencia.

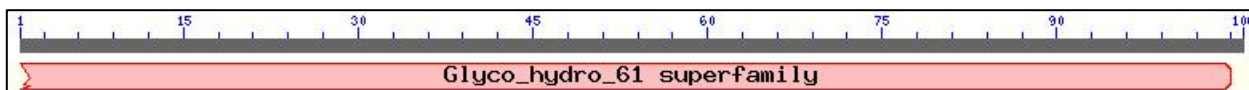


**Figura 3.21 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f12h117-1.** Se utilizó el marco de lectura -3. La barra gris representa a la secuencia f12h117-1, mientras la banda roja representa a la superfamilia TPR. La regla azul muestra la posición de los residuos aminoacídicos.

#### f61h117-1:

Mediante la utilización de la herramienta BLASTX, se obtuvo que la secuencia posee un alto grado de homología con endoglucanasas de la familia 61, siendo una del hongo *Laccaria bicolor* con la cual obtuvo el mejor alineamiento (e-value del orden de  $10^{-28}$ ), con un 58% de identidad a nivel de aminoácidos. El marco de lectura que generó los mejores alineamientos fue el -1, por lo que se procedió a realizar una búsqueda de dominios conservados con la herramienta CDD utilizando este

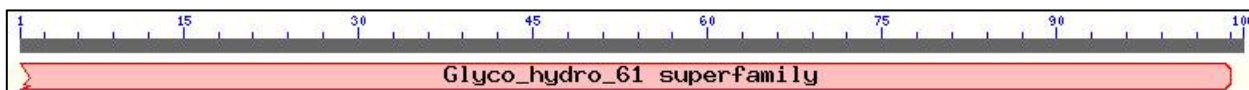
marco. Se obtuvo que la secuencia posee un dominio de la superfamilia Glyco\_hydro\_61, compuesta por GHs de la familia 61 (e-value del orden de  $10^{-22}$ ). En la Figura 3.22 se puede observar que la secuencia íntegramente es parte del dominio conservado, el cual no se encuentra completo (borde irregular del extremo izquierdo de la barra que representa al dominio).



**Figura 3.22 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f61h117-1.** Se utilizó el marco de lectura -1. La barra gris representa a la secuencia f61h117-1, mientras la banda roja representa a la superfamilia de GHs de la familia 61. La regla azul muestra la posición de los residuos aminoacídicos.

### f61h163-2:

El análisis utilizando la herramienta BLASTX, indica que esta secuencia posee alta similitud con endoglucanasas de la familia 61, en particular, con la proteína hipotética CHGG\_10043 proveniente del hongo *Chaetomium globusum*, obteniendo un e-value del orden de  $10^{-32}$  y un porcentaje de identidad del 60%. Al realizar una búsqueda de dominios conservados mediante la utilización de la herramienta CDD, a partir del marco de lectura -1, se encontró que la secuencia posee similitud con la superfamilia Glyco\_hydro\_61, compuesta por GHs de la familia 61 (e-value del orden de  $10^{-29}$ ). En la Figura 3.22 se puede observar que la secuencia íntegramente es parte del dominio conservado, el cual no se encuentra completo (borde irregular del extremo izquierdo de la barra que representa al dominio).



**Figura 3.23 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f61h163-1.** Se utilizó el marco de lectura -1. La barra gris representa a la secuencia f61h163-1, mientras la banda roja representa a la superfamilia de GHs de la familia 61. La regla azul muestra la posición de los residuos aminoacídicos.

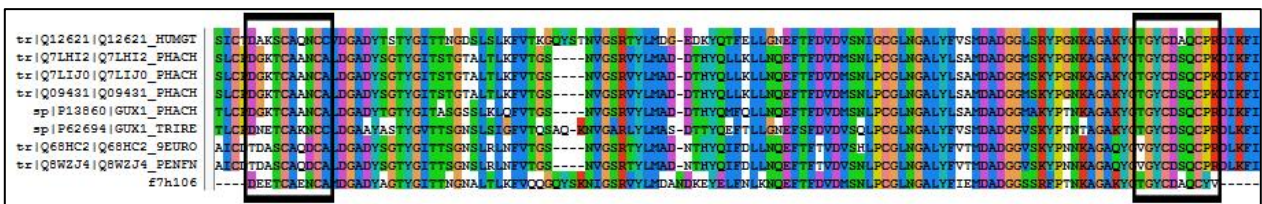
En la Tabla 3.6 se presenta un resumen del análisis realizado para los fragmentos obtenidos.

**Tabla 3.6 Resumen de resultados obtenidos para fragmentos encontrados en el cDNA.**

Fragmento	Mejor Alineamiento			Dominio conservado		
	Organismo	% identidad	e-value	Marco	superfamilia	e-value
f5h106	<i>Trametes hirsuta</i>	93	$10^{-70}$	-1	Cellulase	$10^{-19}$
f7h106	<i>Polyporus arcularius</i>	80	$10^{-51}$	+2	Glyco_hydro_7	$10^{-58}$
f7h163	<i>Fusarium oxysporum</i>	100	$10^{-50}$	+2	Glyco_hydro_7	$10^{-50}$
f12h117	<i>Laccaria bicolor</i>	62	$10^{-31}$	-3	TPR	$10^{-3}$
f61h117	<i>Laccaria bicolor</i>	58	$10^{-28}$	-1	Glyco_hydro_61	$10^{-22}$
f61h163	<i>Chaetomium globusum</i>	60	$10^{-32}$	-1	Glyco_hydro-61	$10^{-29}$

Como se puede observar, al contrario de lo ocurrido cuando se utilizó como templado DNA genómico, fue posible encontrar 4 secuencias con fragmentos de dominios catalíticos pertenecientes a endoglucanasas, hallándose éstas en las familias de GHs para las cuales se habían diseñado los partidores degenerados.

Por otro lado, se encontró una secuencia para una celobiohidrolasa (f7h106-1), la cual pertenece a la familia de GHs para la cual estaban diseñados los partidores que se utilizaron (familia 7). Realizando un alineamiento de este fragmento con secuencias pertenecientes a celobiohidrolasas de la familia 7 (ver Figura 3.24), se pudo apreciar, que pese a no ser óptima, existe cierta homología entre el fragmento obtenido y las demás secuencias en la región correspondiente a los partidores. Considerando que se utilizaron partidores degenerados, existe la posibilidad de que alguna de las conformaciones de éstos pudiese ser complementaria a la región de homología, lo cual sumado a condiciones relajadas de hibridación de los primeros ciclos del PCR, pudo haber promoviendo la amplificación del fragmento en cuestión.



**Figura 3.24 Alineamiento del fragmento f7h106-1 con secuencia de exoglucanasa de la familia 7.** Alineamiento realizado utilizando la herramienta ClustalX2. Las columnas coloreadas indican conservación del residuo aminoacídico en esa posición. Las regiones correspondientes a los partidores se encuentran encuadradas en negro.

Al utilizar los partidores diseñados para la familia 7 de GHs, se obtuvieron tanto endoglucanasas como exoglucanasas, lo cual se debió a que las regiones de homología a partir de las cuales fueron diseñados los partidores se encontraban conservadas en toda la familia. Considerando lo anterior, si se desea obtener secuencias pertenecientes a enzimas con una actividad en particular, se recomienda verificar que las regiones de homología se conserven sólo en proteínas de la familia con la actividad buscada y no en la familia completa. Para el caso de las familias 5, 12 y 61 de GHs, no se presentó este problema, debido a que para ellas sólo existían secuencias descritas correspondientes a endoglucanasas.

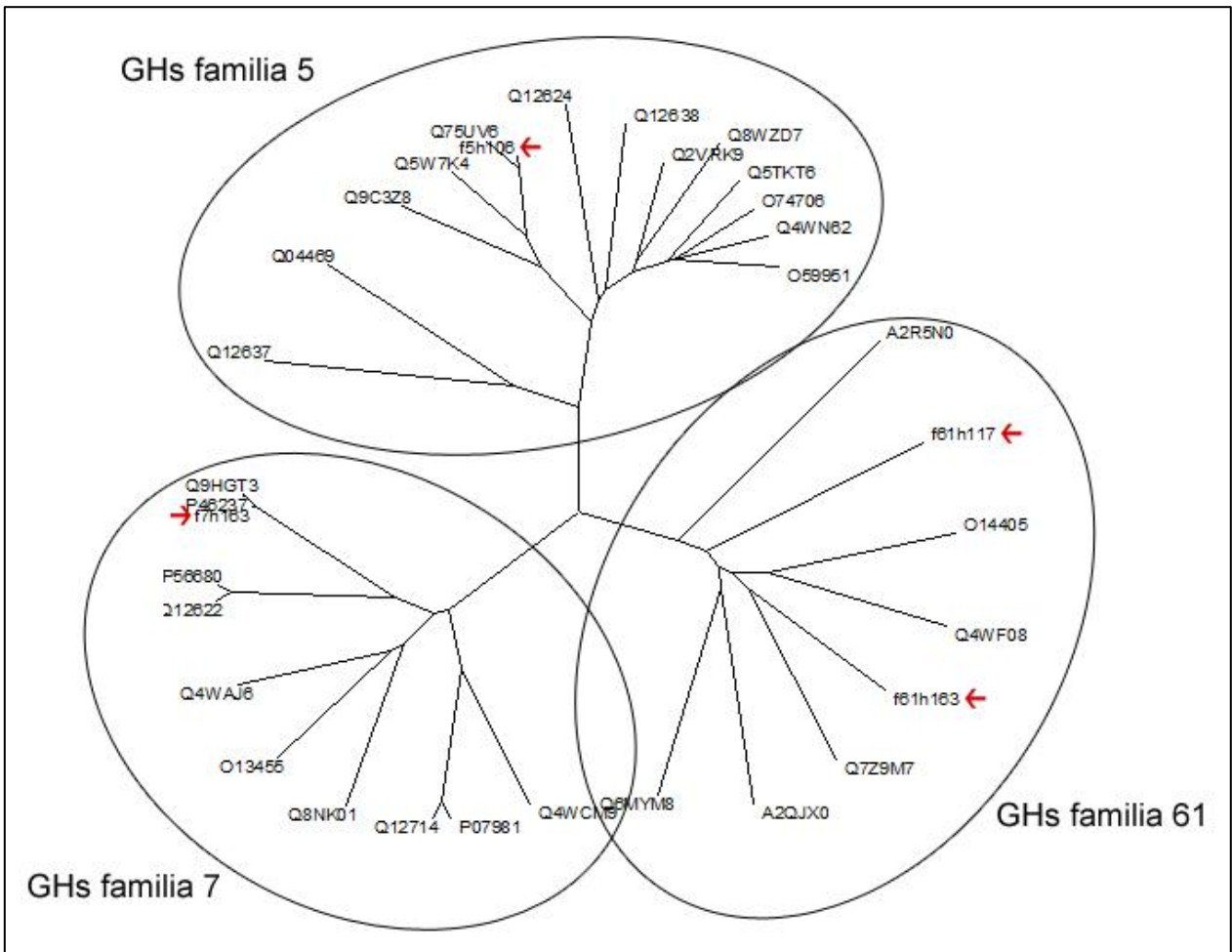
Se realizó un alineamiento utilizando la herramienta BLASTX, con el fin de determinar el grado de homología entre las secuencias aminoacídicas correspondientes a los fragmentos obtenidos utilizando los partidores diseñados para la familia 61 de GHs (f61h117 y f61h163). Los resultados del análisis entregaron un 44% de identidad y un 60% de similitud entre las secuencias, lo que

indica que pese a utilizar los mismos partidores, se pueden obtener secuencias considerablemente distintas con esta estrategia.

Además de las celulasas, se encontró un fragmento que posee un dominio de tetratricopéptido, siendo altamente probable que éste se obtuviese debido a una hibridación inespecífica. Cabe destacar, que éste era el único fragmento en que el tamaño de la banda a partir del cual fue purificado no concordaba con el tamaño esperado. De lo anterior, se desprende que la utilización del tamaño esperado como criterio para seleccionar las bandas es adecuado.

Se realizó un estudio filogenético de los fragmentos de endoglucanasas encontrados. Para ello, mediante la herramienta ClustalX2, se alinearon éstos con todas las secuencias que habían sido utilizadas para el diseño de los partidores degenerados de las familias 5, 7 y 61. El resultado del alineamiento, presentando en la Figura 3.25, fue visualizado con la herramienta TreeView [38]. Se puede observar que se formaron 3 grandes ramas, cada una correspondiente a una familia de endoglucanasas. En cuanto a los fragmentos encontrados, éstos se situaron en las ramas pertenecientes a las familias de las cuales se le habían encontrado dominios conservados, confirmando el análisis anterior. En particular, se observa que el fragmento f7h163 se encuentra solapado con la enzima de código de acceso P46237, siendo ésta la endoglucanasa proveniente de *Fusarium oxysporum* con la cual posee un 100% de identidad. Los fragmentos se encuentran indicados con flechas rojas, para facilitar su identificación.

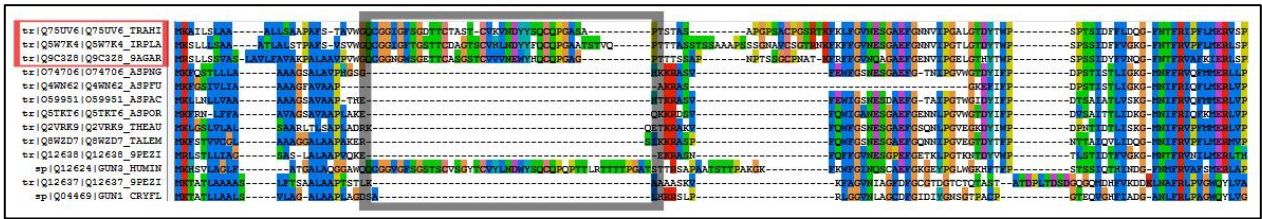
Basándose en el árbol filogenético generado, se realizó un estudio preliminar para determinar la presencia de dominios CBD en las endoglucanasas encontradas, considerando las secuencias de mayor cercanía.



**Figura 3.25 Análisis filogenético de los fragmentos de endoglucanasas encontrados.** Alineamiento generado con el programa ClustalX2 y visualizado mediante el programa TreeView. Las flechas rojas indican los fragmentos de endoglucanasas encontrados. Las distintas familias de GHs se encuentran encerradas en elipses. Las secuencias fueron obtenidas desde la base de datos de la EMBL-EBI.

Se observó que gran parte de las secuencias procedentes de la familia 5 no poseen dominio CBD; sin embargo, las que tienen un mayor grado de similitud con el fragmento f5h106, sí lo poseen. Esto se puede observar en el alineamiento presentado en la Figura 3.26, en donde se indica la posición del dominio CBD en la secuencia, encontrándose cercano al extremo 5'. Además, se tiene que las 4 secuencias que presentan este dominio son las que más cercanas se encuentran del fragmento f5h106 en el árbol filogenético. De lo anterior, se puede inferir que es probable que la enzima de la cual procede el fragmento f5h106 posea un dominio CBD y que al continuar la secuenciación hacia el extremo 5' debiese encontrarse el CBD.





**Figura 3.26 Posición del dominio CBD en secuencias de endoglucanasas de la familia 5 de GHs.** Las secuencias con mayor grado de similitud con el fragmento f5h106 se encuentran encuadradas en rojo, mientras la región en el alineamiento donde se presenta el dominio CBD, se encuentra encuadrada en gris. Además, se presenta una regla celeste, la cual indica la posición de los residuos aminoacídicos.

Según lo discutido anteriormente, con un alto grado de certeza se tiene que el fragmento f7h163 pertenece a la enzima de código de acceso P46237, la cual no posee un dominio CBD.

Las secuencias pertenecientes a la familia 61 poseen el dominio CBD indistintamente, encontrándose situado cercano al extremo 3' de éstas. Debido a que las secuencias de la familia 61 poseen largos muy disímiles (ver Figura 6.4), no fue posible determinar una región en la cual se encontrara éste. Las secuencias vecinas en el árbol al fragmento f61h117 (A2R5N0 y O14405 o Q4WF08) poseen CBD, en cambio, sólo una de las que flanquea a la secuencia f61h163 posee el dominio (O14405 o Q4WF08). Considerando lo anterior, es difícil determinar si las enzimas a las cuales corresponden los fragmentos poseen CBD.

Se hizo una estimación del porcentaje identificado de las enzimas de endoglucanasas encontradas. Para ello, se realizó un alineamiento de secuencias aminoacídicas mediante la herramienta ClustalX2 y se basó en la que presentaba mayor similitud. En la Tabla 3.7 se presentan los resultados obtenidos.

**Tabla 3.7 Porcentaje de la secuencia aminoacídica identificado de las endoglucanasas encontradas.**

Nombre	aa identificados	aa faltantes extremo N	aa faltantes extremo C	Identificado [%]
f5h106	157	158	26	46
f7h163	117	88	255	25
f61h117	105	128	127	29
f61h163	105	128	127	29

Finalmente, con el fin de determinar el porcentaje de identidad y similitud entre las secuencias encontradas y enzimas comerciales utilizadas en el proceso de hidrólisis de celulosa, se realizó un análisis con la herramienta BLASTX. La búsqueda se restringió sólo a secuencias pertenecientes a *Trichoderma reesei*, dado que las celulasas utilizadas hoy en día en este proceso provienen de este hongo [6]. En la Tabla 3.8 se presentan los mejores alineamientos para cada fragmento, cabe señalar que las secuencias obtenidas pertenecen a las mismas familias de GHs que los fragmentos

analizados. Se puede observar que los porcentajes de identidad y similitud presentados entre los fragmentos encontrados y las secuencias de *Trichoderma reesei* son bastante bajos. Cabe destacar que las enzimas de *T. reesei* han sido estudiadas y optimizadas por años, por lo que de encontrar secuencias similares a éstas, difícilmente poseerán mejores características para el proceso de hidrólisis de celulosa. Considerando lo anterior, el haber encontrado secuencias con un bajo porcentaje de similitud permite prever éstas enzimas poseen diferencias estructurales considerables y por ende diferente funcionalidad, existiendo la posibilidad que sean más adecuadas para el proceso de hidrólisis de celulosa, dependiendo de las condiciones utilizadas para realizar este proceso.

**Tabla 3.8 Porcentaje de identidad y similitud entre secuencias encontradas y enzimas comerciales.**

Nombre	Secuencia de <i>Trichoderma reesei</i> *	Identidad [%]	Similitud [%]
f5h106	ABA64553	33	50
f7h163	AAX28897	47	66
f61h117	Q7Z9M7	37	55
f61h163	Q7Z9M7	54	70

\* Código de acceso de la proteína.

### 3.6 Obtención de secuencias aledañas a los fragmentos seleccionados.

A partir de los resultados anteriores, las secuencias seleccionadas para ser completadas fueron: f5h106-1 y f61h117-1. Esto debido a que son secuencias provenientes de hongos de los cuales no se cuentan endoglucanasas descritas. Con el fin de obtener las secuencias aledañas a los fragmentos ya conocidos, se utilizó la estrategia RACE (para el detalle de ésta, ver punto 2.2.9 de Metodología). Para ésta, fue necesario diseñar partidores específicos a partir de las secuencias que se deseaban completar, los cuales se muestran en la Tabla 2.7.

#### 3.6.1 Amplificación del extremo 5' de los genes de endoglucanasas.

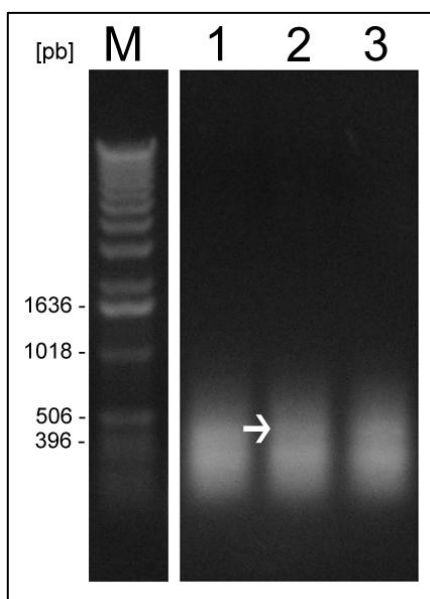
La estrategia se basa en la adición de una cola poli(c) al extremo 5' de todo el cDNA perteneciente al organismo. Luego, utilizando un partidor que posee una secuencia complementaria a ésta y otro diseñado a partir de la secuencia de interés (el cual otorga la especificidad a la reacción), se realiza un PCR para obtener el fragmento deseado.

Según la posición en la secuencia del partidor específico diseñado y la estimación de aminoácidos faltantes presentada en la Tabla 3.7, tras la realización de la reacción de PCR, se esperaban fragmentos de aproximadamente 594 y 600 pb, utilizando los partidores 106EN5Rev y 117EN61Rev respectivamente. A continuación, se presentan los principales resultados obtenidos.

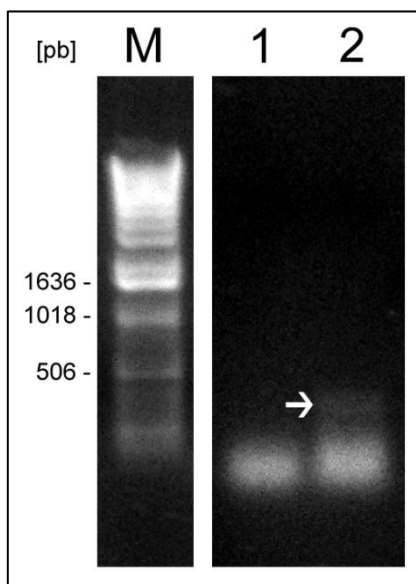
En la Figura 3.27, que corresponde a la utilización del partidador 106EN5Rev en contra del cDNA del hongo 106, se puede observar que se obtuvo una banda tenue de aproximadamente 450 pb. Ésta posee un largo menor al esperado, no obstante, no fue posible obtener una banda de mayor tamaño, por lo que fue purificada para su posterior análisis.

Por otro lado, al utilizar el partidador 117EN61Rev sobre el cDNA del hongo 117, tampoco fue posible obtener una banda del largo esperado, produciéndose una poco definida de aproximadamente 400 pb, la cual se indica con un flecha blanca en la Figura 3.28. Al igual que en el caso anterior, la banda fue purificada para su posterior análisis.

Cabe señalar, que para poder obtener la banda en el primer caso, se utilizó como agente adyuvante BSA, el cual estabiliza las interacciones proteína-DNA [44], sin embargo, al utilizarlo en el segundo caso, no se obtuvo banda alguna.



**Figura 3.27 RACE extremo 5' utilizando partidador específico 106EN5Rev.** Resultado de electroforesis de productos de PCR con BSA utilizando como templado el cDNA del hongo 106, en un gel de agarosa al 2%. Columnas M: Marcador de peso molecular 1 kb; columna 1: temperatura de alineamiento 60°C; columna 2: temperatura de alineamiento 63,3°C; columna 3: temperatura de alineamiento 66°C. El producto seleccionado para purificar se indica con una flecha de color blanco.



**Figura 3.28 RACE extremo 5' utilizando partidor específico 117EN61Rev.** Resultado de electroforesis de productos de PCR utilizando como templado el cDNA del hongo 117, en un gel de agarosa al 2%. Columnas M: Marcador de peso molecular 1 kb; columna 1: temperatura de alineamiento 55°C; columna 2: temperatura de alineamiento 63,2°C. El producto seleccionado para purificar se indica con una flecha de color blanco.

### 3.6.2 Amplificación del extremo 3' de los genes de endoglucanasas.

La estrategia se basa en el aprovechamiento de la cola poli(A) presente en todo el cDNA del organismo. Luego, utilizando un partidor que posee una secuencia complementaria a ésta y otro diseñado a partir de la secuencia de interés (el cual da la especificidad a la reacción), se realiza un PCR para obtener el fragmento deseado.

Utilizando como referencia la estimación de aminoácidos faltantes presentada en la Tabla 3.7 y la posición en la secuencia para la cual fueron diseñados los partidores gen-específico, tras la realización de la reacción de PCR, se esperaban bandas de 320 y 580 pb aproximadamente, utilizando los partidores 106EN5Fwd y 117EN61Fwd respectivamente. Los principales resultados obtenidos, se presentan a continuación.

En la Figura 3.29, correspondiente a la utilización del partidor 106EN5Fwd sobre el cDNA del hongo 106, se puede observar la presencia de una banda de aproximadamente 250 pb, la cual aumenta notoriamente su intensidad al utilizar BSA en la mezcla de reacción de PCR. Esta banda posee un largo inferior al esperado, no obstante, fue purificada para su posterior análisis (la banda purificada se indica con un flecha blanca).

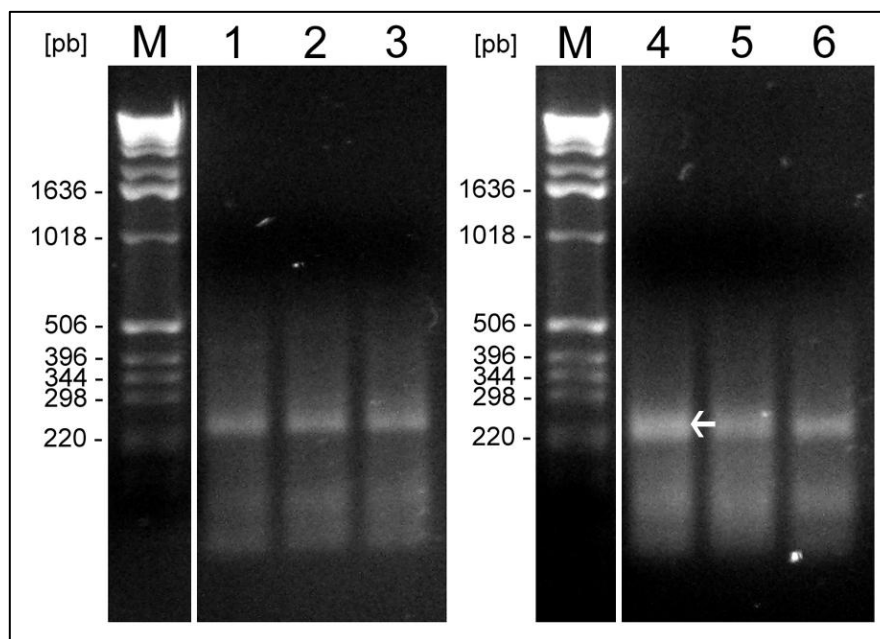
Al utilizar el partidador 117EN61Fwd sobre el cDNA del hongo 117, no se lograron obtener bandas definidas. En la Figura 3.30 se presentan los resultados obtenidos utilizando las mismas condiciones que el caso anterior. Se puede observar, que se genera un chorreo, el cual sólo se ve intensificado al utilizar BSA en la mezcla de reacción.

Por consideraciones de tiempo no fue posible seguir intentando mejorar las condiciones para la reacción de PCR, obteniéndose sólo las bandas antes mencionadas. En la Tabla 3.9 se presenta un resumen de las bandas obtenidas.

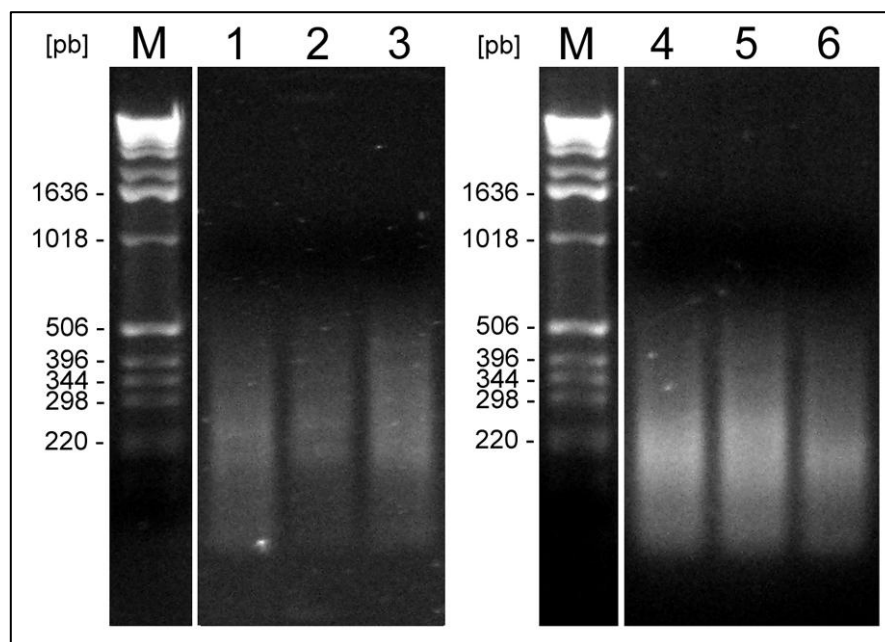
**Tabla 3.9 Resumen bandas obtenidas mediante estrategia RACE**

Nombre	Dirección	Largo [pb]
f5h106-5'	5'	450
f61h117-5'	5'	400
f5h106-3'	3'	250

Las bandas obtenidas fueron purificadas, para luego ser enviadas a secuenciar.



**Figura 3.29 RACE extremo 3' utilizando partidador específico 106EN5Fwd.** Resultado de electroforesis de productos de PCR utilizando como templado el cDNA del hongo 106, en un gel de agarosa al 2%. Columnas M: Marcador de peso molecular 1 kb; columnas del 1 al 3: sin BSA; columnas del 4 al 6: con BSA. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1 y 4: 51°C; columnas 2 y 5: 53.2°C; columnas 3 y 6: 55°C. El producto seleccionado para purificar se indica con una flecha de color blanco



**Figura 3.30 RACE extremo 3' utilizando partidor específico 117En61Fwd.** Resultado de electroforesis de productos de PCR utilizando como templado el cDNA del hongo 117, en un gel de agarosa al 2%. Columnas M: Marcador de peso molecular 1 kb; columnas del 1 al 3: sin BSA; columnas del 4 al 6: con BSA. Las temperaturas de alineamiento para los segundos ciclos fueron las siguientes: columnas 1 y 4: 51°C; columnas 2 y 5: 53.2°C; columnas 3 y 6: 55°C.

### 3.7 Secuenciamiento y análisis de secuencias obtenidas por RACE

Las secuencias obtenidas fueron analizadas utilizando las herramientas BLASTX y CDD. Los resultados de este análisis se presentan a continuación.

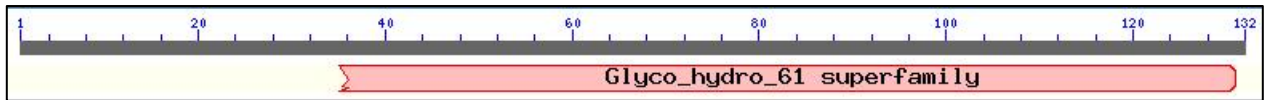
#### f5h106-5':

Al utilizar la herramienta BLASTX, se obtuvo que posee alta similitud con una cadena pesada de una miosina tipo 2 (e-value en torno a  $10^{-23}$ ), proveniente del hongo *Laccaria bicolor*. Esta proteína, participa en la contracción muscular [46]. Al realizar una búsqueda de dominios conservados con la herramienta CDD, no se obtuvo similitud con ningún dominio.

#### f61h117-5':

Mediante la utilización de la herramienta BLASTX, se obtuvo que posee alta similitud con una GH de la familia 61 (e-value en torno a  $10^{-32}$ ), proveniente del hongo *Laccaria bicolor*. Al realizar una búsqueda de dominios conservados con la herramienta CDD, utilizando el marco de lectura -3, se obtuvo que posee alta similitud con un dominio de la superfamilia Glyco\_hydro\_61 (e-value en torno a  $10^{-21}$ ), correspondiente a GHs de la familia 61. En la Figura 3.31 se puede observar que dominio

conservado, representado por una banda roja, se encuentra situado en el extremo 5' de la secuencia, existiendo un fragmento de aproximadamente 35 aa que no posee ningún dominio asociado.



**Figura 3.31 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia f61h117-5'.** Se utilizó el marco de lectura -3. La barra gris representa a la secuencia f61h117-5', mientras la banda roja representa a la superfamilia de GHs de la familia 61. La regla azul muestra la posición de los residuos aminoacídicos.

### f5h106-3':

Al utilizar la herramienta BLASTX, no se obtuvieron alineamientos significativos.

A partir de lo anterior, se tiene que sólo la secuencia correspondiente a la amplificación hacia el extremo 5' de la secuencia f61h117 presentó similitud con dominios conservados de GHs, por lo que se procedió analizar si realmente correspondía a la región aleadaña a esta secuencia. Para esto, se compararon las secuencias involucradas con el fin de detectar si existían zonas en común.

A continuación se presentan las secuencias, las cuales se encuentran orientadas de tal forma que coincidan con el marco de lectura entregado por el buscador de dominios conservados.

```
>f61h117-1
TTGGACAAGACCACCAACAAGTGGGTGAAGATCGACGAGGTCGGGCGCGAGGCCAACGGCGGCGACTGGGTGCAGCAGGAAATCATG
AACGGCGGCACCTACACCGTGACGCTCCCGTCGAACATCGCGCCCGGCGACTACCTCATCCGGCACGAGATCATCGCGCTGCACCTG
GGCATGACCGAGGACGGCGGGAGTTCTACCCGTCGTGCACGCAGGTGCGCATCACCGCAACGGGAGCGGCACGCCAACCCAGACC
GTGTCCTTCCCCGGCGGTACTCAGACACTGACCAAGGCATCTTCATCAACATCTAGAAGT
```

```
>f61h117-5'
TGGCCACGCGTCTGACTAGTACGGGGGGGGGGGGGCGAGAACGTGAGTGCTTGTGTTGCTTGCTTTGGCTGCGCGGGAGGATATGATAG
GTGCTGATGTGGATACGTGGGTGTGCGTGTAGTGGCCGCACAATACGGGCCCGCTGATGACTTATATGGGCGCGTGCAGGGGACGA
CGTGTGACAAGTACACCGCGACGGACGCAAGTGGTTCAAGATCGACGAGGTCGGGCGCGAGGCCAACGGCGGCGACTGGGTGCAGC
AGGAAATCATGAACGGCGGCACCTACACCGTGACGCTCCCGTCGAACATCGCGCCCGGCGACTACCTCATCCGGCACGAGATCATCG
CGCTGCACCTGGGCATGACCGAAGGCGGCGGGAGTTCTACCCGTCGTGCACG
```

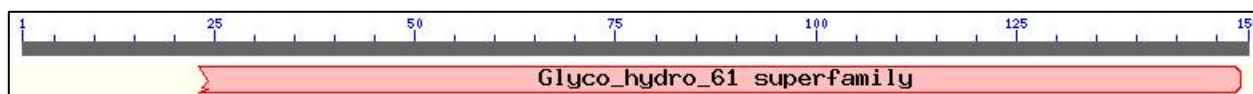
Los nucleótidos marcados de color rojo, corresponden a las regiones en común entre las dos secuencias. Los nucleótidos marcados de color azul, corresponden al partidor *Abridged Anchor Primer*, utilizado al realizar la estrategia RACE 5'. Las regiones que se encuentra subrayadas corresponden a los partidores degenerados utilizados en la amplificación a partir de cDNA.

Se puede observar que el fragmento f61h117-1 posee la región de homología hacia el extremo 5', mientras el fragmento f61h117-5' la presenta hacia el 3', lo cual permite la unión entre estas dos secuencias. No obstante, en el extremo 5' del fragmento f61h117-1 se observan una serie de diferencias en los nucleótidos, éstas se deben a que esa región corresponde al sitio de unión de uno de los partidores degenerados utilizados para la amplificación desde cDNA, por lo que probablemente existen mutaciones asociadas a éste.

A partir de la unión de las secuencias (considerando los nucleótidos pertenecientes al fragmento f61h117-5' en la región asociada al partidor degenerado) y eliminando los nucleótidos correspondientes al partidor *Abridged Anchor Primer*, se obtuvo una nueva secuencia la cual fue denominada 117EN61 y se presenta a continuación.

```
>117EN61
AACGTGAGTGCTTGTGCTTGGCTGCGCGGGAGGATATGA TAGGTGCTGATGTGGATACGTGGGTGTGCGTGTAGTGGCCG
CACAATACGGGCCCCGCTGATGACTTATATGGGCGCGTGCGAGGGGACGACGTGTGACAAGTACACCGCGACGGACGCAAAGTGGTTC
AAGATCGACGAGGTGCGGCGCGAGGCCAACGGCGGCGACTGGGTGCAGCAGGAAATCATGAACGGCGGCACCTACACCGTGACGCTC
CCGTGGAACATCGCGCCCGGCGACTACCTCATCCGGCAGGAGATCATCGCGCTGCACCTGGGCATGACCGAAGGCGCGCGGAGTTC
TACCCGTGCTGCACGCAGGTGCGCATCACCGGCAACGGGAGCGGCACGCCAACCCAGACCGTGTCTTCCCCGGCGCGTACTCAGAC
ACTGACCAAGGCATCTTCATCAACATCTAGAAGT
```

Al utilizar la herramienta BLASTX, se obtiene que posee un 55% de identidad con una GH de la familia 61 (e-value en torno a  $10^{-42}$ ), perteneciente al hongo *Laccaria bicolor*. Mientras que al utilizar la herramienta CDD, se obtuvo que posee alta similitud con un dominio de GHs de la familia 61 (e-value en torno a  $10^{-30}$ ), como se observa en la Figura 3.32. Si se compara este dominio con el obtenido desde la secuencia f61h117-1 (ver Figura 3.22), se tiene que su longitud aumentó en aproximadamente 30 aminoácidos, lo que indica que la amplificación hacia el extremo 5' fue exitosa, sin embargo, el dominio obtenido no se encuentra completo.



**Figura 3.32 Resultado de búsqueda de dominios conservados utilizando la herramienta CDD para la secuencia 117EN61.** Se utilizó el marco de lectura +1. La barra gris representa a la secuencia 117EN61, mientras la banda roja representa a la superfamilia de GHs de la familia 61. La regla azul muestra la posición de los residuos aminoacídicos.

Si se revisa la secuencia 117EN61, cerca del comienzo de ésta se presenta un codón de stop (TAG, indicado con color rojo), el cual justamente coincide con el inicio del dominio conservado. Se cree que éste se pudo haber generado por una mutación introducida por la enzima Taq polimerasa. Considerando lo anterior, se recomienda realizar nuevamente RACE hacia el extremo 5', utilizando un partidor gen-específico diseñado a partir de la secuencia aledaña encontrada, esto con el fin de asegurar que se posee la secuencia completa hacia ese extremo. Considerando sólo los nucleótidos a partir del codón de stop, el conocimiento de la secuencia aumentó de un 29 a un 38%.

Por otro lado, se cree que existe la posibilidad de formación de estructuras secundarias en el RNA hacia el extremo 5', lo que pudo haber impedido la síntesis completa del cDNA. Se recomienda realizar la síntesis de éste a mayor temperatura, para evitar la formación de este tipo de estructuras.



### 3.8 Discusión final

A continuación se presenta un análisis del trabajo realizado, haciendo hincapié en sus principales ventajas y desventajas.

El método de *screening* de actividad utilizado, permitió identificar la presencia de endoglucanasas en forma fácil y relativamente rápida en una gran cantidad de hongos a la vez, permitiendo reducir el número de éstos para el trabajo posterior. Sin embargo, éste método no permite comparar entre enzimas, por lo que existe la posibilidad de que se descarten hongos que producen proteínas con gran actividad específica, pero en baja cantidad. Por otro lado, al utilizar condiciones comunes de crecimiento, es posible que algunos hongos no crezcan de manera óptima, presentando menor actividad de la que podrían poseer. Pese a todo lo anterior, se cree que el método utilizado fue adecuado para este trabajo, considerando que el interés era encontrar enzimas endoglucanasas, sin ser prioridad la actividad específica de éstas.

En cuanto al trabajo de biología molecular, se utilizaron procedimientos que permitieron el análisis de todos los hongos seleccionados al mismo tiempo. Una de las grandes ventajas de utilizar como estrategia el uso de partidores degenerados, fue el poder realizar una búsqueda de enzimas sin poseer antecedentes previos sobre éstas. No obstante, al ser una técnica poco específica, el encontrar las condiciones óptimas para realización de las reacciones de PCR presentó gran dificultad, siendo necesario probar una gran cantidad de condiciones para cada hongo.

Al utilizar DNA genómico como templado para las reacciones de PCR, no fue posible obtener secuencias de endoglucanasas pertenecientes a las familias para las cuales se habían diseñado los partidores degenerados. Sin embargo, utilizando los mismos partidores pero sobre cDNA, sí se obtuvieron secuencias, con una gran selectividad en cuanto a actividad enzimática y familias de GHs se refiere. De lo anterior se desprende que la causa principal de la amplificación inespecífica al utilizar DNA genómico, no fue un mal diseño de los partidores y probablemente se deba a un bajo número de copias de los genes o la presencia de intrones que pudiesen interrumpir las regiones de homología para las cuales fueron diseñados los partidores. Una forma de dilucidar esto último, sería una vez obtenida la secuencia completa, diseñar partidores específicos que flanqueen ésta y realizar una reacción de PCR utilizando éstos sobre el DNA genómico. Otra opción sería verificar si el sitio de los intrones se conserva en las secuencias de mayor similitud y utilizar esa información para predecir su ubicación en la secuencia obtenida. Se encontró que en el caso de la familia 12 de GHs, existen 4 intrones que conservan su posición en las secuencias [47], encontrándose uno de éstos

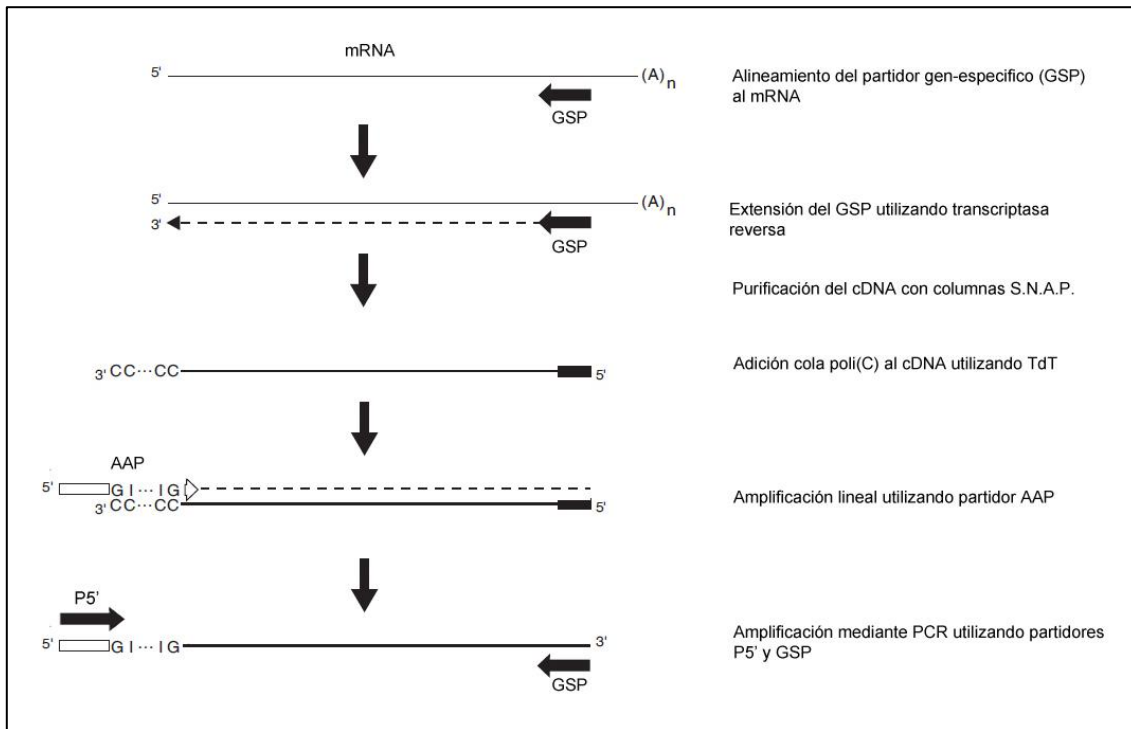
interrumpiendo la secuencia nucleotídica para la cual se encontraba diseñado el partidor degenerado en sentido *reverse*. No se encontró información para el resto de las familias de GHs. Dado los antecedentes encontrados, se cree que la presencia de intrones en los genes de endoglucanasas pudiese haber sido la razón por la cual la identificación no fue efectiva al utilizar DNA genómico como templado para el PCR.

Cabe señalar, que las regiones de las secuencias encontradas que corresponden a los partidores degenerados, no pueden ser consideradas como parte de la secuencia nucleotídica de la enzima, esto debido a que la presencia de los partidores pudo haber introducido mutaciones indeseadas.

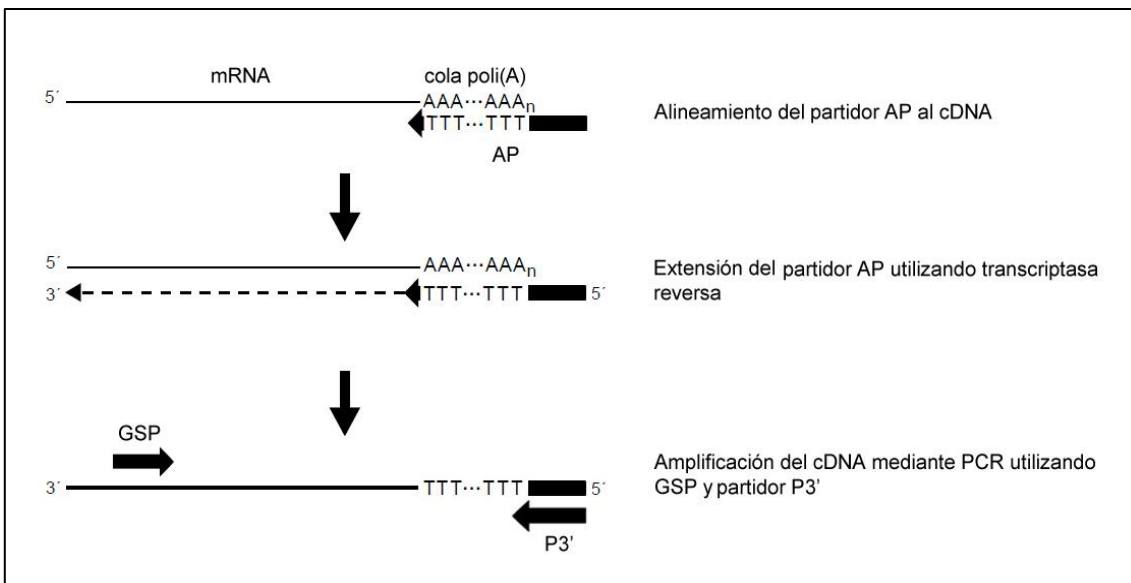
La utilización de la estrategia RACE, salvo el caso del extremo 5' en el fragmento f61h117, no dio los resultados esperados en cuanto la obtención de la secuencias nucleotídica completas de las enzimas que se encontraron. Se cree que esto se debió a la poca especificidad del método, el cual se basa en la utilización de sólo un partidor específico en la reacción de PCR, mientras el otro es común para todas las secuencias del cDNA, teniendo estos dos partidores grandes diferencias en sus  $T_m$ . Es posible seguir modificando las condiciones de los PCR utilizados, no obstante, es poco probable que los resultados varíen en demasía. Se propone realizar pequeñas modificaciones en la estrategia utilizada, con el fin de aumentar la selectividad de la reacción de PCR, éstas se detallan a continuación:

En el caso de la amplificación del extremo 5', se propone realizar una amplificación lineal (utilizando el partidor AAP) previa a la reacción de PCR exponencial, para luego realizar esta última, utilizando como partidor uno que corresponda a la región aledaña a la poli(G) del AAP y un partidor gen-específico de  $T_m$  similar. En la Figura 3.33 se presenta el diagrama para el RACE 5' modificado.

En cuanto a la amplificación del extremo 3', se propone realizar la síntesis del cDNA a partir del partidor AP, para luego realizar una reacción de PCR exponencial utilizado como partidor uno que corresponda a la región aledaña a la poli(T) del AP (P3') y un partidor específico para el gen de  $T_m$  similar. En la Figura 3.34 se presenta el diagrama para el RACE 3' modificado.



**Figura 3.33 Diagrama estrategia 5' modificada.** Se presentan las principales etapas de la estrategia. GPS: partidor gen específico; AAP: partidor Adapter Anchor Primer; P5': partidor AAP sin poli(G).



**Figura 3.34 Diagrama estrategia RACE 3' modificada.** Se presentan las principales etapas de la estrategia. GPS: partidor gen específico; AP: partidor Adapter Primer; P3': partidor AP sin poli(T).

Considerando las altas posibilidades de que el fragmento f7h163 pertenezca a la endoglucanasa de código de acceso P46237, se propone diseñar partidores específicos a partir de ésta, con el fin de obtener su secuencia completa utilizando el cDNA del hongo 163. Pese a que la secuencia de esta

proteína ya se encuentra descrita, ésta puede utilizarse para pruebas de expresión recombinante o como material de inicio para una posible evolución dirigida, la cual permita hacerla más adecuada para el proceso de hidrólisis de material lignocelulósico.

Dado el éxito en la identificación de endoglucanasas utilizando partidores degenerados sobre el cDNA, se plantea la utilización de esta técnica para obtener fragmentos de mayor longitud. Para que lo anterior sea factible es necesario verificar la existencia de regiones de homología cercanas a los extremos de la secuencia que permitan el diseño de partidores degenerados para las distintas familias de GHs. Con el fin de aumentar la especificidad de la reacción, una opción sería realizar una reacción de PCR utilizando el partidador degenerado diseñado en un sentido y un partidador específico diseñado a partir de las secuencias ya encontradas en el sentido contrario.

Una estrategia alternativa para el proceso global de identificación de nuevas endoglucanasas, sería la utilización de genotecas de cDNA. Las principales etapas para esta estrategia son las siguientes: (i) generación de una genoteca de cDNA para cada hongo, (ii) clonamiento de la genoteca en un sistema de expresión recombinante y (iii) selección de los clones productores mediante ensayos de actividad. La principal ventaja de esta estrategia, es la posibilidad de encontrar enzimas con actividad endoglucanasa totalmente nuevas, ya que estas no son encontradas basándose en regiones de homología presentes en las enzimas ya descrita, como en el caso de la estrategia utilizada en este trabajo. Por otro lado, de encontrar alguna enzima con la actividad buscada, se tendría toda la secuencia de ésta y se sabría que se expresa en forma funcional. Sin embargo, posee desventajas importantes como es la dificultad para producir las enzimas en forma activa, esto debido a que las enzimas provienen de un organismo eucarionte y por ende poseen modificaciones post-traduccionales. Además, para poder identificar una enzima, es necesario analizar un gran número de clones, lo que conlleva a gastos en cuanto a tiempo y recursos elevados.

Una vez que las secuencias nucleotídicas pertenecientes a las endoglucanasas encontradas se encuentren completamente identificadas, se espera poder producir éstas en un sistema de expresión recombinante. Esto último, con el fin de obtener grandes cantidades de enzima en forma pura, lo cual permita poder caracterizarlas. Considerando que los genes provienen de organismos eucariontes, la selección de un correcto sistema de expresión es de suma importancia. Una alternativa es utilizar un sistema bacteriano (por ejemplo el vector de expresión pET22 en la cepa *E. coli* BL21), sin embargo, la expresión no necesariamente se producirá correctamente debido a la carencia de modificaciones post-traduccionales en este tipo de huésped. Se ha reportado la expresión de endoglucanasas de *Trichoderma reesei* en *E. coli*, sin embargo, sólo una pequeña

fracción de las proteínas se expresó en forma activa [48]. La utilización de un sistema eucarionte, como la levadura *Saccharomyces cerevisiae*, parece ser una opción más adecuada. Existe información de la expresión de un número importante de celulasas de origen fúngico utilizando este sistema, pudiendo ser secretadas al medio extracelular en forma exitosa pese a la hiperglicosilación de las proteínas [16]. Otro sistema de expresión eucarionte que se ha utilizado exitosamente para la expresión activa de endoglucanasas de hongos, es la levadura *Pichia pastoris* [49] .

Luego, con el fin de mejorar las propiedades de las enzimas y hacerlas más apropiadas para el proceso de hidrólisis de celulosa, se plantea la realización de una evolución dirigida. La elección de esta estrategia, se debe principalmente a que el desconocimiento de la estructura de las enzimas encontradas impide un enfoque racional. Una posible técnica para realizar esta estrategia, es la utilización de PCR con error para generar enzimas mutantes, las cuales luego son testeadas para determinar si sus propiedades se vieron mejoradas. Se ha reportado la utilización de evolución dirigida en celulasas para alterar propiedades tales como actividad, termoestabilidad y la estabilidad a pHs alcalinos [8]. Para la selección de la propiedad a optimizar es necesario tomar en cuenta bajo qué condiciones se utilizará la enzima, por ejemplo si el proceso se lleva a cabo a altas temperaturas se requiere que la celulasa posea una mayor termoestabilidad, por lo que la evolución dirigida será enfocada a mejorar esa propiedad.

## 4. Conclusiones

La metodología utilizada para realizar el *screening* de actividad endoglucanasa, empleando el sustrato soluble CMC, permitió identificar de manera eficiente hongos que poseían enzimas con esta actividad, permitiendo realizar una selección preliminar entre éstos.

La etapa correspondiente al aislamiento de los genes codificantes para endoglucanasa presentes en los hongos seleccionados, se cumplió en forma exitosa, siendo posible clonar fragmentos de secuencia correspondiente a cinco glicosil hidrolasas en células de *E.coli* DH5 $\alpha$ . Sin embargo, no fue posible completar las secuencias de los genes encontrados. Cabe destacar que la estrategia adecuada fue la utilización de cDNA como templado y no DNA genómico, ya que a partir de este último no fue posible encontrar ningún fragmento con utilidad para el proceso de hidrólisis de celulosa.

Además, el uso de partidores degenerados diseñados a partir de regiones de homología, es altamente selectivo en cuanto a familias de glicosil hidrolasas y actividad se refiere. Todas las secuencias encontradas, pertenecen a las familias de glicosil hidrolasas para las cuales estaban diseñados los partidores con las que fueron amplificadas.

Debido al alto grado de identidad (100%) presentando entre la secuencia f7h106 y la proteína de código de acceso P46237 y considerando que provienen del mismo hongo (*Fusarium oxysporum*), existen altas probabilidades de que se trate de la misma proteína.

Los fragmentos obtenidos, fueron clasificados como endoglucanasas (secuencias f5h106, procedente de *Polystictus versicolor*; f61h117, procedente de *Ganoderma applanatum*; y f7h163 y f61h163, procedentes de *Fusarium oxysporum*) y exoglucanasas (secuencia f7h106, procedente de *Polystictus versicolor*), mediante la utilización de las herramientas bioinformáticas BLASTX y CDD, obteniendo resultados con un alto grado de confianza (e-value en torno  $10^{-28}$  y  $10^{-22}$  como máximo, respectivamente). Se obtuvo que éstos no poseen un porcentaje de identidad significativo con enzimas que se utilizan actualmente en el proceso de hidrólisis de celulosa, siendo éste inferior al 54%.

Finalmente, considerando todo lo anterior, se concluye que la estrategia global utilizada permitió identificar y clonar fragmentos de tres secuencias nuevas pertenecientes a endoglucanasas, las cuales son potencialmente útiles para ser utilizadas en el proceso de hidrólisis de celulosa.

## 5. Bibliografía

- [1] Almada, M. (2006). Análisis de la producción de materias primas para la elaboración de bioetanol y biodiesel, y de estos combustibles, presente y esperada hasta el 2020, en países potencialmente proveedores de Chile. [en línea]. Recuperado el 05 de Junio, de 2008, de [http://www.enap.cl/opensite\\_20051129160918.asp](http://www.enap.cl/opensite_20051129160918.asp).
- [2] Comisión Nacional de Energía. (2006). Hidrocarburos [en línea]. Recuperado el 24 de Marzo de 2008, de [http://www.cne.cl/fuentes\\_energeticas/f\\_energeticas.html](http://www.cne.cl/fuentes_energeticas/f_energeticas.html).
- [3] Tan, K.T., Lee, K.T. Mohamed, A.R. (2008). Role of energy policy in renewable energy accomplishment: The case of second-generation bioethanol. *Energy Policy*, 36, 3360-3365.
- [4] Muñoz, R., Ortega, R., Acosta, L., González, R. (2009). Potencial de producción de biocombustibles en base a cultivos agrícolas en Chile. Parte II: Evaluación económica de elaboración de biocombustibles. Universidad Técnica Federico Santa María, en proceso de publicación.
- [5] CAMPOCOOP LTDA, (2008). Chile apuesta a biocombustibles de segunda generación. [en línea]. Recuperado el 21 de Julio de 2008, de <http://www.campocoop.cl/?page=articulo&id=653>.
- [6] Ryu, D., Mandels, M. (1980). Cellulases: biosynthesis and applications. *Enzyme Microbiology Technology*, 2, 91-102.
- [7] Bayer, E., Lamed, R., Himmel, M. (2007). The potential of cellulases and cellulosomes for cellulosic waste management. *Current Opinion in Biotechnology*, 18, 237-245.
- [8] Zhang, Y.H.P., Himmel, M., Mielenz, J. (2006). Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances*, 24, 452-481.
- [9] Solomon, B., Barnes, J., Halvorsen, K. (2007). Grain and cellulosic ethanol: History, economics, and energy policy. *Biomass and Bioenergy*, 31, 416-425.
- [10] Zaldivar, J., Nielsen, J., Olson, L. (2001) Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Applied Microbiology and Biotechnology*, 56, 17-34.

- [11] Diario oficial de la República de Chile. (2008). Decreto Supremo N°11. *Subsecretaría de Economía, Fomento y Reconstrucción*. [en línea]. Recuperado el 23 de Octubre de 2009, de [http://www.cne.cl/archivos\\_bajar/DS11.pdf](http://www.cne.cl/archivos_bajar/DS11.pdf).
- [12] Biblioteca del Congreso Nacional. (2007). Boletín N° 4.873-08. [en línea]. Recuperado el 23 de Octubre de 2009, de [http://www.bcn.cl/actualidad\\_legislativa/temas\\_portada.2007-01-29.7882319251/boletin\\_4873\\_actualidad.pdf](http://www.bcn.cl/actualidad_legislativa/temas_portada.2007-01-29.7882319251/boletin_4873_actualidad.pdf).
- [13] Muñoz, R., Ortega, R., Acosta, L., González, R. (2009). Potencial de producción de biocombustibles en base a cultivos agrícolas en Chile. Parte III: Localización simultánea de centros de acopio y plantas productoras de biocombustibles. Universidad Técnica Federico Santa María, en proceso de publicación.
- [14] Pereira, L. (2003). The chemistry involved in the steam treatment of lignocellulosic materials. *Química Nova*, 26, 863-871.
- [15] Sun, Y., Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technology*, 83, 1-11.
- [16] Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S. (2002). Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiology and Molecular Biology Reviews*, 66, 506-577.
- [17] Taherzadeh, M., Karimi, K. (2007). Enzyme-based hydrolysis for processes for ethanol from lignocellulosic materials, a review. *BioResources*, 2, 707-738.
- [18] Galbe, M., Zacchi, G. (2002). A review of the production of ethanol from softwood. *Applied Microbiology and Biotechnology*, 59, 618-628.
- [19] O'Sullivan, A. C. (1997). Cellulose: the structure slowly unravels. *Cellulose*, 4, 173-207.
- [20] Linder, M., Teeri, T. (1997). The roles and function of cellulose-binding domain. *Journal of Biotechnology*, 57, 15-28.
- [21] Kaur, J., Chadha, B.S., Kumar, B., Saini, H. (2007). Purification and characterization of two endoglucanases from *Melanocarpus sp.* MTCC 3922. *Bioresource Technology*, 98, 74-81.



- [22] Brun, E., Gans, P., Marion, D., Barras, F. (1995). Overproduction, purification and characterization of the cellulose-binding domain of the *Erwinia chrysanthemi* secreted endoglucanase EGZ. *European Journal of Biochemistry*, 231, 142-148.
- [23] Tomme, P., Warren, R.A.J., Gilkes, N.R. (1995). Cellulose hydrolysis by bacteria and fungi. *Advances in Microbial Physiology*, 37, 3-81.
- [24] Tomme, P., Warren, R.A.J., Miller, R.C., Kilburn, D.G., Gilkes, N.R. (1995). Cellulose-binding domains: Classification and properties. *ACS Symposium Series*, 618, 143-163.
- [25] Fierobe, H.P., Bagnara-Tarif, C., Gaudin, C., Guerlesquin, F., Sauve, P., Belaich, A., Belaich, J.P. (1993) Purification and characterization of endoglucanase C from *Clostridium cellulolyticum*. Catalytic comparison with endoglucanase A. *European Journal of Biochemistry*, 217, 557-565.
- [26] Suto, M., Tomita, F. (2001). Induction and Catabolite Repression Mechanisms of Cellulase in Fungi. *Journal of Bioscience and Bioengineering*, 92, 305-311.
- [27] Mandels, M., Parrish, F.W., Reese, E.T. (1962). Sophorose as an inducer of cellulose in *Trichoderma viride*. *Journal of Bacteriology*, 83, 400-408.
- [28] Kurasawa, T., Yachi, M., Suto, M., Kamagata, Y., Takao, S., Tomita, F. (1992). Induction of cellulase by gentiobiose and its sulfur-containing analog in *Penicillium purpurogenum*. *Applied and Environmental Microbiology*, 58, 106-110.
- [29] Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*, 280, 309-316.
- [30] Cantarel, B.L. (2008). Carbohydrate Active Enzymes database. [en línea]. Recuperado el 01 de Septiembre de 2009, de <http://www.cazy.org>.
- [31] Leonowicz, A., Matuszewska, A., Luterek, J., Ziegenhagen, D., Wojtaś -Wasilewska, M., Cho, N., Hofrichter, M., Rogalski, J. (2002). Biodegradation of Lignin by White Rot Fungi. *Fungal Genetic and Biology*, 27, 175-185.
- [32] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, L.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389-3402.

- [33] Fred Hutchinson Cancer Research Center. (2003). *BlockMaker*. [en línea] Recuperado el 12 de Septiembre de 2008, de [http://blocks.fhcrc.org/blocks/make\\_blocks.html](http://blocks.fhcrc.org/blocks/make_blocks.html).
- [34] Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., et al. (2007). CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Research*, 35, 237-240.
- [35] Larkin, M. A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-2948.
- [36] Rose, T., Henikoff, J., Henikoff, F. (2003). CODEHOP (Consensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Research*, 31, 3763-3766.
- [37] Rozen, S., Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Humana Press, Totowa, NJ, 365-386.
- [38] Page, R.D.M. (1996). TreeView: An application to display phylogenetic trees on personal computer. *Computer Applications in Biosciences*, 12, 357-358.
- [39] Promega. (2007). *pGEM®-T and pGEM®-T Easy Vector Systems, Technical Manual*.
- [40] Brooksbank, C., Camon, E., Harris, M., Magrane, M., Martin, M.J., Mulder, N., et al. (2003) The European Bioinformatics Institute's data resources. *Nucleic Acids Research*, 31, 43-50.
- [41] Einsele, H., Hebart, H., Roller, G. (1997). Detection and identification of fungal pathogens in blood by using molecular probes. *Journal of Clinical Microbiology*, 35, 1353-1360.
- [42] Frackman, S., Kobs, G., Simpson, D., Stort, D. (1998). Betain and DMSO: Enhancing Agents for PCR. *Promega Notes*, 65, 27-30.
- [43] Nagai, M., Yoshida, A., Sato, N. (1998). Additive effects of bovine serum albumin, dithiothreitol and glycerol on PCR. *Biochemistry and Molecular Biology International*, 44, 157-163.
- [44] Zimmer, E.A., Roalson, E.H. (2005). Molecular Evolution: Producing the Biochemical Data, Part B. *Methods in Enzymology*, 395, 87-103.

- [45] Korenbaum, E., Rivero, F. (2002). Calponin homology domains at a glance. *Journal of Cell Science*, 115, 3543-3545.
- [46] Lehman, W., Szent-Gyorgyi, A.G. (1997). Regulation of muscular contraction. Distribution of actin control and myosin control in the animal kingdom. *The Journal of General Physiology*, 66, 1-30.
- [47] Goedegebuur, F., Fowler, T., Phillips, J., van der Kley, P., van Solingel, P., Dankmeyer, L., Power, S. (2002). Cloning and relational analysis of 15 novel fungal endoglucanases from family 12 glycosyl hydrolase. *Current Genetics*, 41, 89-98.
- [48] Okada, H., Tada, K., Sekiya, T., Yokoyama, K., Takahashi, A., Tohda, H., Kumagai, H., Morikawa Y. (1998). Molecular Characterization and Heterologous Expression of the Gene Encoding a Low-Molecular-Mass Endoglucanase from *Trichoderma reesei* QM9414. *Applied and Environmental Biotechnology*, 64, 555-563.
- [49] Wonganu, B., Pootanakit, K., Boonyapakron, K., Champreda, V., Tanapongpipat, S., Eurwilaichitr, L. (2008). Cloning, expression, and characterization of a thermotolerant endoglucanase from *Syncephalastrum racemosum* in *Pichia pastoris*. *Protein Expression and Purification*, 58, 78-86.

## 6. Anexos

### 6.1 Metodología adicional

#### 6.1.1 Medios de cultivo

Todos los medios son preparados en agua destilada MilliQ y son esterilizados utilizando autoclave.

##### 6.1.1.1 Medios sólidos

- Medio agar malta  
Agar 1,5%, malta 2%.
- Medio agar malta CMC  
Agar 1,5%, malta 2%, CMC 1%.
- Medio agar papa dextrosa  
Agar papa dextrosa 39 g/l.
- Medio LB selectivo  
Agar 1,5%, ampicilina 100 µg/ml, X-Gal 40 µg/ml, IPTG 0,1 mM.
- Medio mínimo sólido CMC  
Agar 1,6%, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 0,05%, KH<sub>2</sub>PO<sub>4</sub> 0,1%, KCl 0,05%, MgSO<sub>4</sub>×7H<sub>2</sub>O 0,041%, CaCl<sub>2</sub> 0,005%, extracto de levadura 0,05%, CMC 0,5%.
- Medio mínimo sólido avicel  
Agar 1,6%, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 0,05%, KH<sub>2</sub>PO<sub>4</sub> 0,1%, KCl 0,05%, MgSO<sub>4</sub>×7H<sub>2</sub>O 0,041%, CaCl<sub>2</sub> 0,005%, extracto de levadura 0,05%, avicel 0,5%.
- Medio mínimo sólido Lenga  
Agar 1,6%, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 0,05%, KH<sub>2</sub>PO<sub>4</sub> 0,1%, KCl 0,05%, MgSO<sub>4</sub>×7H<sub>2</sub>O 0,041%, CaCl<sub>2</sub> 0,005%, extracto de levadura 0,05%, lenga 0,5%. Características físicas de la lenga: 1-2 mm de espesor, 1 mm de ancho y 5-7 mm de largo, con una densidad aparente de 156 kg/m<sup>3</sup>.

##### 6.1.1.2 Medios líquidos

- Medio mínimo-CMC  
KH<sub>2</sub>PO<sub>4</sub> 1,5%, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 0,5%, MgSO<sub>4</sub> 0,06%, CaCl<sub>2</sub> 0,06%, FeSO<sub>4</sub> × 7H<sub>2</sub>O 0,0005%, MnSO<sub>4</sub>×H<sub>2</sub>O 0,00016%, ZnSO<sub>4</sub>×7H<sub>2</sub>O 0,00014%, CoCl<sub>2</sub>×6H<sub>2</sub>O 0,00037%, CMC 2%.
- Medio LB  
LB 15,5 g/l, NaCl 9,5 g/l.
- Medio LB de recuperación

Glucosa 0,02 M, MgCl<sub>2</sub> 0,05 M en medio LB.

- Medio LB-amp

Ampicilina 0,1 mg/ml en medio LB.

### 6.1.2 Determinación de azúcares reductores mediante el método DNS

En primer lugar, para realizar la reacción enzimática, se transfieren 50 µl de la muestra con enzimas a una placa que contiene 100 µl de CMC al 1% (en buffer citrato 50 mM, pH 4.8) y se incuba a 37°C por 30 min. Para detener la reacción, se transfieren 90 µl a una placa que contiene 90 µl de DNS. Esta mezcla es calentada durante 10 min a 100°C y luego enfriada en hielo durante 5 min. Finalmente, para cuantificar actividad, se transfieren 100 µl a una placa Falcon de 96 pocillos y se mide la absorbancia a 550 nm.

### 6.1.3 Curva de calibración DNS

Se utilizaron distintas diluciones de una solución de glucosa de concentración 1 g/l. A 90 µl de cada solución se le agregó 90 µl de DNS. Esta mezcla fue calentada durante 10 min a 100°C y luego enfriada en hielo durante 5 min. Finalmente, se transfieren 100 µl a una placa Falcon de 96 pocillos y se mide la absorbancia a 550 nm.

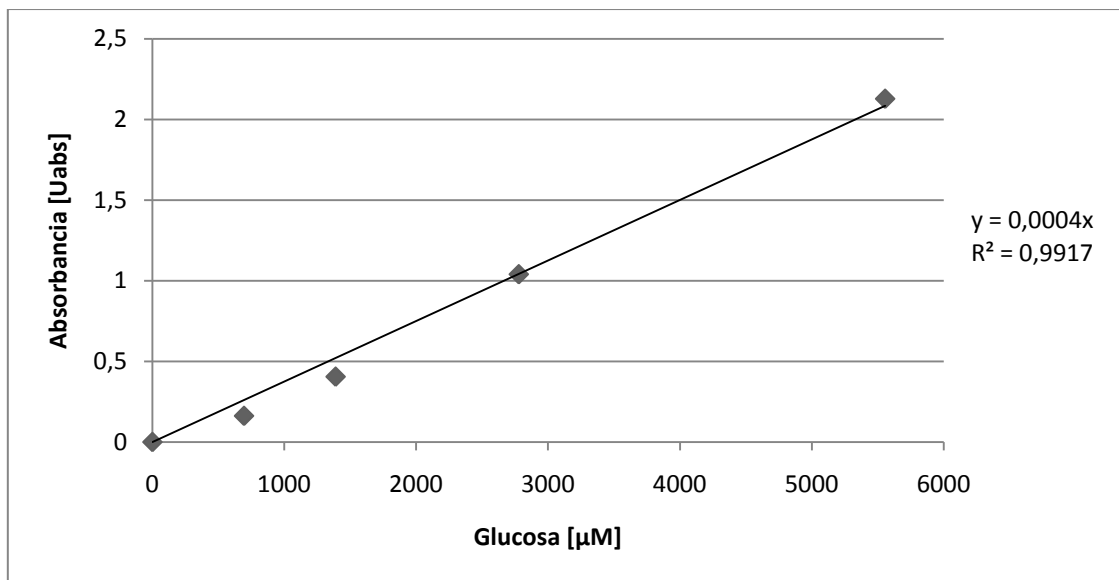


Figura 6.1 Curva de calibración DNS

### 6.1.4 Ensayos de actividad en geles de agarosa -CMC

El gel se prepara utilizando agarosa al 1% y CMC al 1% en buffer citrato (50mM en agua destilada MilliQ y ph 4.8). Se calienta y luego se vacía en una placa Petri para realizar el ensayo. Una vez gelificado, se agregan 3µl del sobrenadante previamente centrifugado proveniente de los hongos

crecidos en medio líquido. Se incubó la placa durante 2hr a 37°C para que se lleve a cabo la reacción. Se tiñe la placa con rojo Congo al 0,1% a temperatura ambiente durante 15 min con agitación orbital suave. Se lava varias veces con NaCl 1M a temperatura ambiente y agitación orbital suave, hasta notar claramente halos sin tinción.

### 6.1.5 Alfabeto degenerado

**Tabla 6.1 Alfabeto degenerado.**

<b>Letra</b>	<b>Base(s)</b>
A	A
C	C
G	G
T	T
R	AG
Y	CT
M	AC
K	GT
W	AT
S	CG
B	CGT
D	AGT
H	ACT
V	ACG
N	ACGT

### 6.1.6 Verificación de extracción de DNA

Para verificar una correcta extracción de DNA, se realizó un PCR utilizando partidores universales para hongos, diseñados para amplificar un fragmento del gen 18S rRNA [41]. El programa de PCR utilizado fue: 34 ciclos de 94°C por 30 s, 62°C por 1 min y 72°C por 2 min, seguidos por una extensión terminal a 72°C por 5 min. Los partidores utilizados se presentan en la Tabla 6.2. El fragmento generado debe poseer un largo entre 482 y 503 pb dependiendo del hongo.

**Tabla 6.2 Partidores para gen 18S rRNA de hongos.**

<b>Nombre</b>	<b>Dirección</b>	<b>Secuencia (5' a 3')</b>
F18S_fwd	Forward	attggagggcaagtctggtg
F18S_rev	Reverse	ccgatccctagtcggcatag

### 6.1.7 Purificación de fragmentos de DNA desde gel de agarosa

Se utilizó el kit QIAEX II, siguiendo las instrucciones del fabricante. En el paso final, el DNA fue eluído en 15 µl de buffer EB, debido a la baja concentración de DNA presente en la muestra.

### **6.1.8 Ligación de fragmentos de DNA al vector pGEM®-T Easy.**

En un tubo Eppendorf de 0,6 ml estéril, se realiza una mezcla de ligación como se indica a continuación: 50 ng de pGem®-T Easy, buffer de ligación 2X, 1 U de DNA ligasa de T4 y 3 µl de DNA proveniente de la purificación desde gel de agarosa, en un volumen final de 10 µl. La mezcla de reacción se deja toda la noche a 4°C.

### **6.1.9 Transformación de células *E.coli* DH5α**

Con el fin de obtener un alto número de copias del vector en el cual se encuentra el inserto, éstos se clonan en células electrocompetentes *E.coli* DH5α mediante electroporación. Para ello, a 20 µl de células se le agrega 1 µl del producto de la ligación. La electroporación se lleva a cabo utilizando un equipo Cell-Porator, con 420 V, 330 µF, baja impedancia, una tasa de carga rápida y un tiempo máximo de 2,5 ms. Tras la transformación, las células se mezclan con 1 ml de medio de recuperación (ver Anexo 6.1.1.2) y se incuban durante 1 hr a 37°C. Luego, la mezcla es centrifugada a 13000 RPM durante 2 min, se descartan 950 µl de sobrenadante y el sedimento es resuspendido en el volumen restante, el cual se utiliza para sembrar placas con medio LB selectivo (ver Anexo 6.1.1.1). Estas últimas, se incuban a 37°C durante toda la noche. Una vez terminada la incubación, se tendrán colonias blancas y azules, correspondiendo las blancas a colonias que poseen el inserto. Se seleccionan 4 colonias blancas por placa y se traspasan a tubos de ensayo con 4 ml de medio LB-amp (ver Anexo 6.1.1.2), los cuales se dejan incubando a 37°C con agitación orbital de 200 RPM durante toda la noche.

### **6.1.10 Extracción del DNA plasmidial desde las células *E.coli* DH5α transformadas.**

Los cultivos en LB-amp son centrifugados a 13000 RPM por 2 min para obtener un sedimento de células *E.coli* DH5α. A partir de éste, utilizando el kit QIAprep Spin Miniprep según las instrucciones del fabricante, se obtiene el DNA plasmidial.

### **6.1.11 Verificación de presencia de inserto en el vector pGEM®-T Easy.**

Para verificar la presencia de inserto en el vector, se realiza una digestión sobre éste utilizando la enzima de restricción EcoRI. La reacción se lleva a cabo según se indica a continuación: 5 µl del DNA plasmidial, 1 U de EcoRI y buffer NE EcoRI 1X en un volumen final de 10 µl, durante 4 h a 37°C. El resultado de la digestión se observa en un gel de agarosa al 1,5%.

### 6.1.12 Extracción de proteínas utilizando buffer de extracción

Se prepara un buffer de extracción compuesto por: buffer acetato de sodio (0,1 M, pH 5,5), NaCl 0,02M y azida de sodio 0,02%. Con ayuda de un bisturí se corta el agar en cuadrados de 2 mm de lado y se mezclan en un tubo falcon con 15 ml de buffer de extracción. La mezcla se incuba a temperatura ambiente durante 2 hr, para luego ser centrifugada a 4400 RPM por 10 min a 4°C. El sobrenadante es guardado con PMSF (inhibidor de proteasas) a una concentración final de 1mM para ser utilizado en pruebas posteriores.

### 6.1.13 Tratamiento contra RNAsas utilizando DEPC

Los morteros a utilizar para la extracción de RNA se dejan inmersos en agua con DEPC al 0,1 % a 37°C durante toda la noche. Al otro día, éstos se forran en papel kraft y se esterilizan por autoclave. Además, es necesario esterilizar agua con DEPC al 0,1%, la cual se utiliza para la dilución del etanol y la elución final del RNA.

### 6.1.14 Tratamiento del RNA con DNAsas

En un tubo Eppendorf de 0,6 ml estéril, se realiza la siguiente mezcla de reacción: 1 µg de RNA, 1 µl de buffer de reacción DNasa I 10X y 1 µl de DNasa I 1U/µl llevando a un volumen total de 10 µl utilizando agua tratada con DEPC. La mezcla se incuba por 15 min a temperatura ambiente. Con el fin de inactivar la DNasa I, se agrega 1 µl de EDTA 20 mM. Finalmente, la mezcla se calienta durante 10 min a 65°C.

### 6.1.15 Secuencias de partidores gen-específico adicionales para RACE

En la Tabla 6.3 se presentan los partidores gen-específicos diseñados para la estrategia RACE.

**Tabla 6.3 Secuencias de partidores adicionales para RACE**

<b>Nombre</b>	<b>Dirección</b>	<b>Familia</b>	<b>Hongo</b>	<b>Secuencia 5´a 3´</b>
163EN7Fwd	Forward	7	163	caggatggtgtaagagcacgag
163EN7Rev	Reverse	7	163	ccacaaggaagcttctccatctc
163EN61Fwd	Forward	61	163	caaccgaacatggacgacaaag
163EN61Rev	Reverse	61	163	ttggcagttcccgacctataa

### 6.1.16 Secuencia de partidores utilizados para secuenciar

En la Tabla 6.4 se presentan los partidores utilizados para la secuenciación a partir del vector pGEM®-T Easy.



**Tabla 6.4 Partidores utilizados para la secuenciación.**

<b>Nombre</b>	<b>Secuencia 5'a 3'</b>
M13F-pUC	GTTTCCCAGTCACGAC
M13R-pUC	CAGGAAACAGCTATGAC

## 6.2 Resultados adicionales

### 6.2.1 Datos de ensayos de actividad en medio líquido

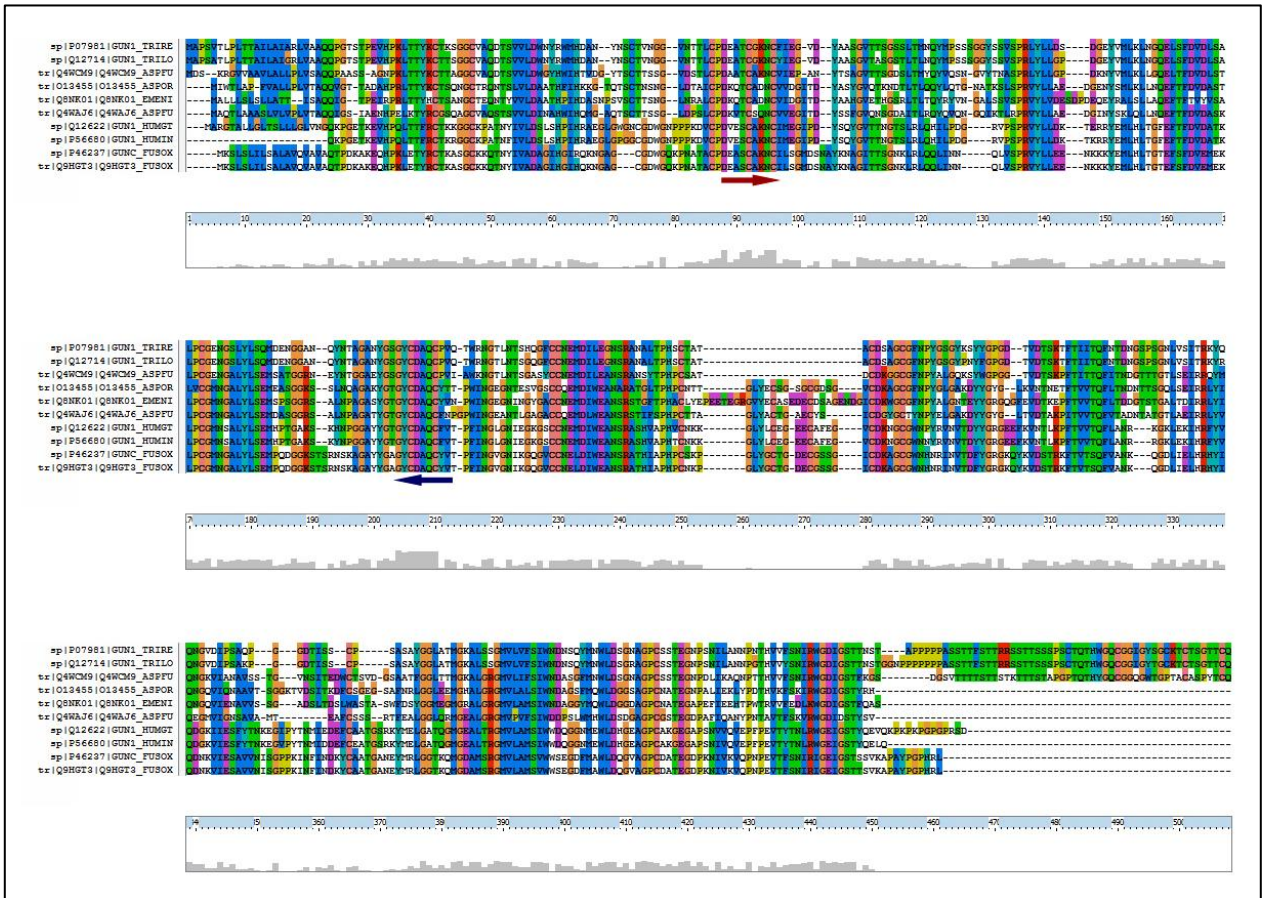
**Tabla 6.5 Datos de ensayos de actividad en medio líquido.**

Los valores de absorbancia corresponden al promedio de los ensayos en duplicado, al cual se le sustrae el promedio del valor del experimento blanco en duplicado. La columna máximo corresponde al día en el cual se obtuvo el máximo de absorbancia. La fila B corresponde a los valores obtenidos desde un medio de cultivo que no fue inoculado.

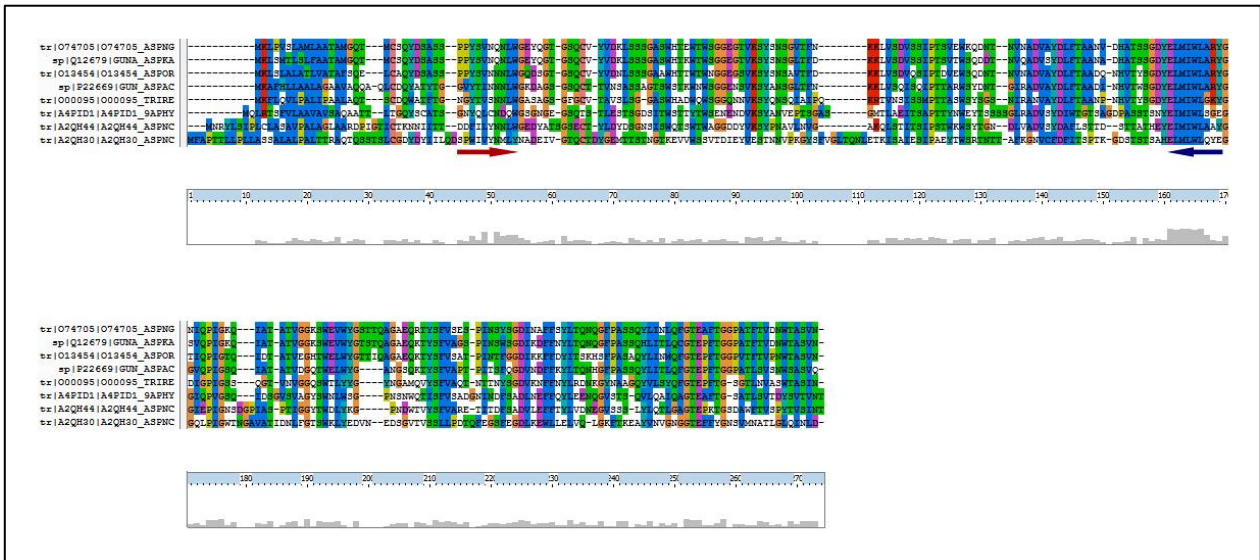
Clave hongo	Absorbancia [U <sub>abs</sub> ]				Máximo	Máx. actividad endoglucanasa [U/l]
	Día 3	Día 4	Día 5	Día 7		
B	0,098	0,073	0,232	0,089	0,232	14,625
96	0,361	0,424	0,312	0,319	0,424	30,667
97	0,561	0,486	0,335	0,331	0,561	42,042
100	0,243	0,267	0,187	0,216	0,267	17,583
104	0,483	0,470	0,342	0,405	0,483	35,583
106	0,602	0,753	0,368	0,478	0,753	58,042
108	0,253	0,351	0,280	0,364	0,364	25,625
110	0,624	0,976	0,404	0,417	0,976	76,667
113	0,589	0,899	0,384	0,456	0,899	70,250
114	0,636	0,453	0,388	0,375	0,636	48,333
117	0,818	0,997	0,370	0,425	0,997	78,417
121	0,224	0,273	0,218	0,277	0,277	18,417
127	0,144	0,155	0,087	0,111	0,155	8,208
147	0,125	0,111	0,089	0,074	0,125	5,750
152	0,122	0,073	0,069	0,287	0,287	19,208
162	0,870	0,966	0,541	0,593	0,966	75,792
163	0,321	0,300	0,387	0,393	0,393	28,042
164	0,378	0,486	0,307	0,337	0,486	35,792
167	0,633	0,523	0,280	0,195	0,633	48,083
177	0,287	0,466	0,247	0,247	0,466	34,167
178	0,376	0,447	0,299	0,358	0,447	32,583
181	0,426	0,315	0,342	0,399	0,426	30,792
183	0,220	0,233	0,261	0,415	0,415	29,875
186	0,145	0,110	0,075	0,078	0,145	7,417
202	0,663	0,546	0,251	0,403	0,663	50,542

## 6.2.2 Alineamientos de secuencias aminoacídicas adicionales

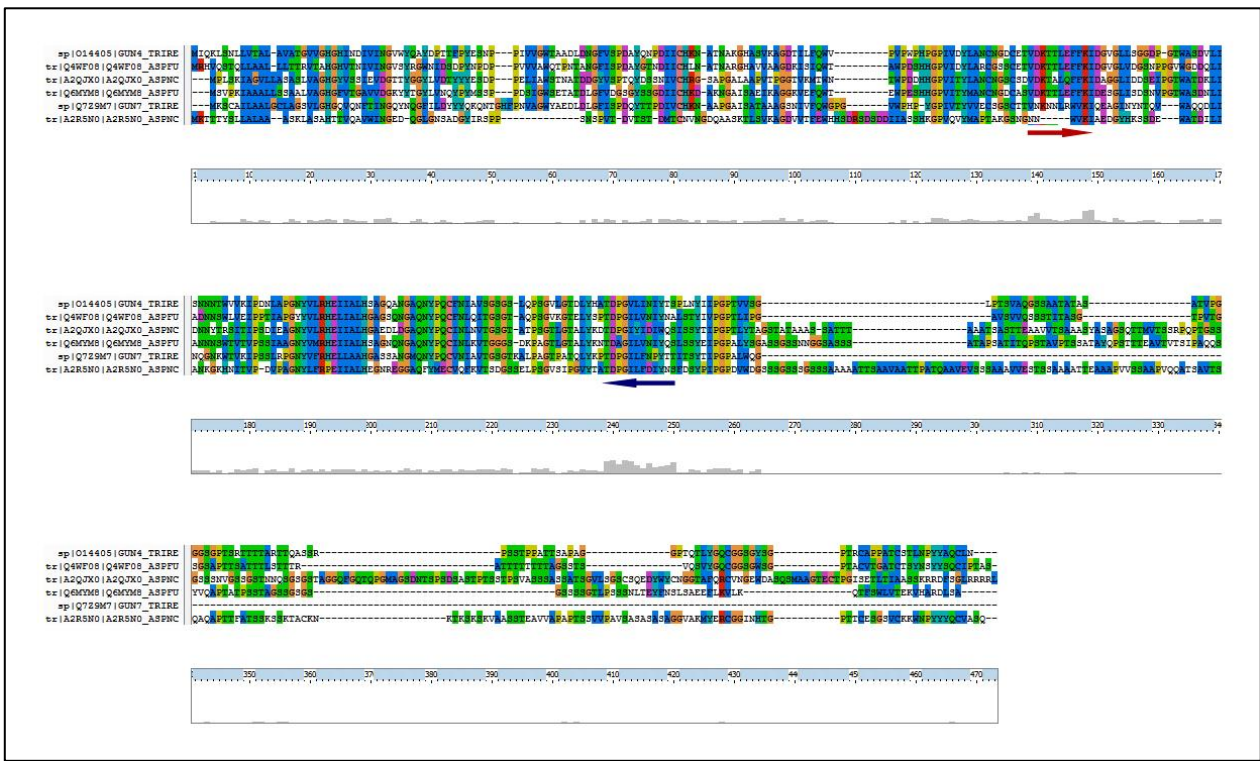
A continuación se presentan los alineamientos de las secuencias aminoacídicas pertenecientes endoglucanasas de las familias de GHs 7, 12 y 61. Además, se presentan la posición en la secuencias de los partidores degenerados diseñados para cada familia.



**Figura 6.2** Alineamiento de endoglucanasas pertenecientes a la familia 7 de GHs. Resultado del alineamiento de las secuencias aminoacídicas de la familia 7 utilizando la herramienta ClustalX2. Debajo de cada alineamiento se presenta una regla que indica la posición de los aminoácidos. La flechas roja y azul representan la posición en la secuencia de los partidores degenerados *celF\_endo\_fwd7* y *celF\_endo\_rev7* respectivamente.



**Figura 6.3 Alineamiento de endoglucanas pertenecientes a la familia 12 de GHs.** Resultado del alineamiento de las secuencias aminoacídicas de la familia 12 utilizando la herramienta ClustalX2. Debajo de cada alineamiento se presenta una regla que indica la posición de los aminoácidos. La flecha roja y azul representan la posición en la secuencia de los partidores degenerados celF\_endo\_fwd12 y celF\_endo\_rev12 respectivamente.



**Figura 6.4 Alineamiento de endoglucanas pertenecientes a la familia 61 de GHs.** Resultado del alineamiento de las secuencias aminoacídicas de la familia 61 utilizando la herramienta ClustalX2. Debajo de cada alineamiento se presenta una regla que indica la posición de los aminoácidos. Las flechas roja y azul representan la posición en la secuencia de los partidores degenerados celF\_endo\_fwd61 y celF\_endo\_rev61 respectivamente.

### 6.2.3 Secuencias

A continuación se presentan las secuencias que exhibieron alineamientos significativos. Las secuencias correspondientes al vector de clonación fueron removidas.

#### 6.2.3.1 Provenientes desde DNA genómico

>f5h117A-1

```
CCCCCACCAGGAGCCGGTAGCCAGACATTCCCCTTCGAGAAACCATCCAAAGTCGTATCGCGCAGCGTCCAACCCGGAGTCCGGA
GCCCTTAACGTTTCCCATCCATTTATCCCATCGCCAATACAGTCAGAGTCCATTCCCTCATCATCGTCCGGCCTCGTCGAAATTCCC
TCCAACAATATCAGCACTCCGGAGTTAATTCAGCCGCCCTCCATCGTACTGTCACAATGTGGACACTCTACCCCTCGCTACGAC
AAGCACGCACCTCGCTTCGATCCGGAACAGCCGCGGACCTACTTCGCTATTTTCGAGGACCTGGCGGACTTATTTGCCGCTCATCCA
AATGCGGTCCCGAATGAAGCAACAAAAAGGCGGTAGTGGTAAAATATGTGCCTATGTATGAGGAGGAGCTATGGAAGGCGTCCCG
GAATTTTCGGACGACTCCAAAAGCTTTGAGGACTTCAAGGGAGCAGTCTTGGCGTTATACCCAGCCGTTAAGGAGGACCAGAGATAT
AGCGTCGGGGATATGGACCGACTAGTGGGGGAACGGCAACGAGTGGGCATCCACAATTTGGCGGACCTGGCAGCATTTTCATCGCGAC
TTCTACTTATCACTCGATATCTGCGCAAGAACGACATCATTTCCGTTCGAGAGCAGGGGAGGGCATCCAGCGAGGGTTTCAGCCG
GAGTTGTGGAACAAAAATATTCGGGCGACTCCAGATCAAGGACGTAGACCACCAACCAGACACTCCATATGGCGTGNACGAGGTTTAC
AAGGCAGCGGAGTTTATTTCTTCATGGGCCCGCGCTCCTGGTGGGGGA
```

>f5h117C-1

```
CCCCCACCAGGAGCCAGCGGCCATTCCAAGACGTCTCACCCGATATCTCGTTATGAACGTCTAGCAAATATAGATCCATTAGACG
CCGCCGAAAGTCTTTGATCTTTTCGTAGTAGTCATCGGTTCTTCTCTGGAGGGCTGATAAGATGAGCTCCTTTGCACGGTTGACGGTA
CCGTCCGGGAGCTTAGGGTCCCGCAGCGCAGGGCGGAGTATGATGTGAACCAAGTAGAAATCAACGAGCAAGGGGAGAGAATCGGA
GACAACCTCTGGAGTACGGTCAGAAATGAGTCCACATGCCGCGGCGATGTGTGGTTCGACGAGGTCTTGTGAATGGTGAACCTCG
TCGTCCAGACGGGTCTGTAGGACGTGGGCTCTGAGGACGTCTGTATTCGTGGATGAAGGCATCAATAGCAGGGCATAGCTGGGCC
GCCGGCTCCTGGTGGGGGA
```

>f5h162B-1

```
CCCCCACCAGGAGCCTGCGGCCATGGCGAAGTGTGGATGGCACGCTTGCAGCTTCTCGTTCGTTCGGATCAAACGACGCCGTGT
CGACATCGAGCGGAGTGATGTTGTTGAGCTGGTTGAAGAGGTAGCAGAGCGGGACACCCAAAGCGAAGCAGTCCCAGAGCTGGGTGA
CGGGGTCGGTTCGAGCGTCAGGACGACGAGGACTCGGGTTCGGGTGAGAGGCTAAAGAAGTCGGAAGTTCGTGCACACGCATCAAGC
GGTTGCGGAGAGCGGAGCATTGCTGTTAGAGGGACGTGACTGGGCCACAGGCTCCTGGTGGGGGA
```

>f7h162A-2

```
CCCCCACCAGGAGCCGGCGGCCACCATCGGAGGGCTTATACGCGCACTACGACGTGACCGAGGAGCTCGGCAAGGGCTCCTTTGCC
ACAGTCATGAAGGCCATGAGCAAGGCCAACGGTCACTGGTACGCCGTGAAGATGATCCACAACAAGAACCTCAAGCGTTCGGTCACA
AAGTACATCAACGGCAATGGCGAGCGTGTACGAGGACGCGTCCAAAGCGCTGCACAAGGAGGTACAGATCCTGCAGCGGCTGCAG
CACCCGAACATCTGTGAGCTGAAGGAAGTGTCTACGAGGAGAATACATCCGTGAGTTGAGTTTTGTCCAGCTGAATGCATGTGCT
CACTATTGGGACAGATATCGTCTCGAGTGGGCAGCCGCTCCTGGTGGGGGA
```

>f61h110B-2

```
ACTTCTAGATGTTGATGAAGATGCCGGGGTTCGGTTGCAGTACCTCGTCAAATGGAAGGTTATGACGACTCGCATACTCATGGGAA
CCGGCGCGCAACGTTGCCAATGCGCCGCCCTCGTAGCCGACTTCCACCGGAAGAACCAGAAATGCCCTCGCCGCTCAATGCTACT
ACCTTTGGTGTACTCAACTTCCAGCCAACGCCTTCGCCGCTGACTGTAGCGCGTGCAGCTCCGGGTGGGAGATGGGCCGGGGGGCC
TCGCATGAGGTTCATGCCCGTTGAGGGGGGGTAATGTCACGAGTTGGTACGGTGTCTAGCCGCTTTTTAATTTTTTCTTTTCGCG
GGCTCGAACCCGCGACCTCTGTACTACATGCCTGATACGGACCCGGGCATCTTCATCAACATCTAGAAGTA
```

#### 6.2.3.2 Provenientes desde cDNA

>f5h106-1

```
CCCCCACCAGGAGCCGGCGGCCACCTGAGCGCGCCGAGCCACGTCCCGCCGGCTGCTGCATCTCGCAGAGCGCGCCCTGCAGCGC
GGTCACGCACTGCGTGTGTTGCCCCGTGCCGATCTCGCCGAGGAAGCCCTTGGAGGCCCTTCTGCTGCAGCCACTGCGTCGCCGCTG
CAGGCGTCCGCGCCGATCGTGTCCGACACGCAGATCGGGGACGTGCCCGAGCCGTCCGAGTCCAGGTAAGTGGTGCATCTGGATAGC
AATGTTGTTGTTCCGGTTCCTTGATCGCGCCGAACGCGTTCGCTGTTGCCCGAGGTCGTCACGTCACGCCCCAGTCCAGCTCGTGCC
CTCGCGAGGATGAGTGGCTGGTTCGCGCCGCTCGCGCACGCCGTTGACGGCCGCTTGCATGAGCTGGAAGACGGTCTGCGCGGG
GATGTCATGTGACTCGTTCATGATGTCGAAGATCAA
```

>f7h106-1  
CGACGAGGAGACCTGCGCGGAGAAGCTGCGCGATGGACGGCGCTGACTACGCCGGCACCTACGGTATCACGACCAACGGCAACGGCGCT  
CACGCTCAAGTTTCGTGCAACAGGGTCAGTACTCGAAGAACATCGGCTCGCGCGTGTACCTCATGGAGCGGAACGACAAGGAGTACGA  
GCTCTTCAACCTGAAGAACCAGGAGTTCACCTTCGACGTTGACATGTCGAACCTCCCTGCGGCCTCAACGGCGCGCTCTACTTCAT  
CGAGATGGACGCCGACGGTGGCTCGTCCCGCTTCCCCACCAACAAGGCTGGCGCCAAGTACGGAACCGGTTACTGCGACGCCAGTG  
CTACGTGA

>f7h163-1  
CGACGAGGAGACCTGCGCGGATAACTGTATCCTCAGTGGTATGGACTCAAACGCTTACAAGAACGCTGGTATCACTACTTCTGGCAA  
CAAGCTTCGTCTTCAGCAGCTTATCAACAACAGCTTGTCTCCTCGAGTTTATCTGCTTGAGGAGAACAAGAAGAGTATGAGAT  
GCTTCACCTCACTGGCACTGAGTTCTCTTTGATGTTGAGATGGAGAAGCTTCCTTGTGGTATGAATGGTGTCTTTGTACCTTTCTGA  
GATGCCCCAGGATGGTGGTAAGAGCACGAGCCGAAACGCAAGGCTGGCGCCTACTATGGTGTGGTACTGCGACGCCAGTGCTA  
CGTGA

>f7h163-1 sin partidores  
TATCCTCAGTGGTATGGACTCAAACGCTTACAAGAAGCTGGTATCACTACTTCTGGCAACAAGCTTCGTCTTCAGCAGCTTATCAA  
CAACCAGCTTGTCTCCTCGAGTTTATCTGCTTGAGGAGAACAAGAAGAGTATGAGATGCTTCACCTCACTGGCACTGAGTTCTC  
TTTTGATGTTGAGATGGAGAAGCTTCCTTGTGGTATGAATGGTGTCTTTGTACCTTTCTGAGATGCCCCAGGATGGTGGTAAGAGCAC  
GAGCCGAAACAGCAAGGCTGGCGCCTACTATGGTGTCT

>f7h163-2  
CGACGAGGAGACCTGCGCGGAAAACCTGTATCCTCAGTGGTATGGACTCAAACGCTTACAAGAACGCTGGTATCACTACTTCTGGCAA  
CAAGCTTCGTCTTCAGCAGCTTATCAACAACAGCTTGTCTCCTCGAGTTTATCTGCTTGAGGAGAACAAGAAGAGTATGAGAT  
GCTTCACCTCACTGGCACTGAGTTCTCTTTGATGTTGAGATGGAGAAGCTTCCTTGTGGTATGAATGGTGTCTTTGTACCTTTCTGA  
GATGCCCCAGGATGGTGGTAAGAGCACGAGCCGAAACGCAAGGCTGGCGCCTACTATGGTGTGGCTATTGCGACGCCAGTGCTA  
CGTGA

>f12h117-1  
CGTCAGGTAGTTGAACAACCTCCTTGATGTGCGCCATCTCCACCAGTGCCTTTGCTTTTCTGAAGTGCCCTTACTCCATGGCCGCTT  
GAGCGCGATGACGGTCTCTGCGTCTACGAGAGCACCAATCAGGTCGCCCGATTTCGAGGTACGCAGCGGACCGGTTGGAGATGACAGT  
GGAGAGCTCGTCTCGCATGAGTTGACTGGCTTCCCAAGGTGGCGCTGCATAGCGTAGTTCGCGGCCAAGGTGTATTTGGCGACGGC  
CTTGTCTGGCTGGCTGCTTTGAACAGTTGGTTTCTCCTCCTTCGTGTTGTTCACTACCTGCGACAGCTTCTTGTACACGATCTA  
GGGGCGA

>f61h117-1  
ACTTCTAGATGTTGATGAAGATGCCTTGGTCACTGTCTGAGTACGCGCCGGGAAGGACACGGTCTGGTTGGGCGTGCCGCTCCCGT  
TGCCGGTATGCGCACCTGCGTGCACGACGGGTAGAACTCCGCGCCGCTCCTCGGTCATGCCAGGTGCAGCGCGATGATCTCGTGCC  
GGATGAGGTAGTCCCGGGCGCGATGTTTCGACGGGAGCGTCACGGTGTAGGTGCCCGGTTTCATGATTTCTGCTGCACCCAGTCCG  
CGCCGTTGGCTCGCGCCGACCTCGTCGATCTTACCACCTTGTGGTGGTCTTTGTCAA

>f61h163-2  
ACTTCTAGATGTTGATGAAGATGCCGGGTCCTGGGAGTGTAGAGCTTAGTGCCCAAGACACCTCAGGCTTGGCAGTTCCCGACC  
CTATAATCTCGATATTGAAGCATTGAGGGTAGTTCTGTGCGCCGCCAATCCTGTTTCGACCCGTGAACAGCGATTGTTTCGTGACGGA  
AGACATATTTGCCGCGAGTTAGCGTCGAGGGTATCTTTGTGCTCCATGTTTCGGTTGTTGTTGACCATCTTTCTAGCCGCCACACTT  
GAGTGTCAAATCGATCCCGCTTCGTCAATCTTACCACCTTGTGGTGGTCTTTGTCAA

### 6.2.3.3 *Provenientes desde RACE*

>f61h117-5'  
CGTGCACGACGGGTAGAAGTCCGCGCCGCTTCCGGTATGCCAGGTGCAGCGCGATGATCTCGTGCCGGATGAGGTAGTCCCGGG  
CGCGATGTTTCGACGGGAGCGTCACGGTGTAGGTGCCCGGTTTCATGATTTCTGCTGCACCCAGTCCCGCCGTTGGCTCGCGCCC  
GACCTCGTCGATCTTGAACCACTTTGGCTCCGTCGCGGTGACTTGTACACGTCGTCCTCCCTCGCAGCGCCCATATAAGTCATCAG  
CGGGCCCGTATTGTGCGGCCACTACACGCACACCCAGTATCCACATCAGCACCTATCATATCTCCCGCGCAGCCAAAGCAAGCAA  
ACAAGCACTCACGTTCTGCCCCCCCCCCCCCTACTAGTCGACGCGTGGCCA

>117EN61  
AACGTGAGTGTCTGTTTGGCTTGGCTTGGCTGCGCGGGAGGATATGATAGGTTGCTGATGTGGATACGTGGGTGTGCGTGTAGTGGCCG  
CACAATACGGGCCCCGTGATGACTTATATGGGCGGTGCGAGGGGACGAGTGTGACAAGTACACCCGACGGACGCAAAGTGGTTC  
AAGATCGACGAGGTGCGGCGCGAGGCCAACGGCGGCGACTGGGTGCAGCAGGAAATCATGAACGGCGGCACCTACACCGTGACGCTC  
CCGTGCAACATCGCGCCCGGCGACTACCTCATCCGGCACGAGATCATCGCGCTGCACCTGGGCATGACCGAAGGCGGCGGGAGTTC  
TACCCGTCGTGCACGAGGTGCGCATCACCGCAACGGGAGCGGCACGCCCAACCAGACCGTGTCTTCCCCGGCGGCTACTCAGAC  
ACTGACCAAGGCATCTTCATCAACATCTAGAAGT