



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERIA QUÍMICA Y BIOTECNOLOGÍA

# **IDENTIFICACIÓN Y CLONACIÓN DE GLICOSIL HIDROLASAS PARA USO EN PRODUCCIÓN DE BIOETANOL**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL EN BIOTECNOLOGÍA

**ALEJANDRA NATALIA GUERRERO ADAROS**

PROFESOR GUÍA:  
ORIANA SALAZAR AGUIRRE

MIEMBROS DE LA COMISIÓN:  
MARÍA ELENA LIENQUEO CONTRERAS  
RENÉ CARMONA CERDA

SANTIAGO DE CHILE  
ABRIL DE 2009

## Resumen

La hidrólisis de celulosa a azúcares fermentables es una de las etapas clave que deben ser optimizadas para disminuir el costo de producción de bioetanol producido a partir de lignocelulosa, materia prima abundantemente disponible en Chile en desechos agrícolas y forestales. La hidrólisis de celulosa es catalizada por enzimas celulasas, pertenecientes al grupo de las glicosil hidrolasas.

Este Trabajo de Memoria de Título tiene como objetivo la identificación y clonación de glicosil hidrolasas, con énfasis en enzimas potencialmente útiles en procesos de producción de bioetanol. La estrategia general de trabajo consistió en la identificación y selección de cepas capaces de degradar celulosa cristalina, a las cuales se dirigió un método de obtención de fragmentos de DNA codificantes de glicosil hidrolasas, enfocado a la actividad exoglucanasa.

Un análisis *in silico* de la secuencia de los fragmentos de DNA obtenidos indicó que se obtuvieron fragmentos de genes con actividades putativas  $\beta$ -1,3(4)-glucanasa y  $\alpha$ -glucosidasa provenientes de cepas de *Lentinula edodes* y *Peniophora gigantea*, respectivamente. Estos resultados poseen altos niveles de confianza (con e-values del orden de  $10^{-9}$  y  $10^{-30}$ , respectivamente). El fragmento correspondiente a la  $\beta$ -1,3(4)-glucanasa contiene el dominio catalítico conservado característico de este tipo de enzimas. Se estima que los fragmentos encontrados representan un 38% de una  $\beta$ -1,3(4)-glucanasa y un 12% de una  $\alpha$ -glucosidasa.

El método también produjo la obtención de fragmentos de DNA que codificarían enzimas con otras actividades no relacionadas con glicosil hidrolasas, específicamente una transcriptasa reversa y un transportador de urea.

Las enzimas  $\beta$ -1,3(4)-glucanasas y  $\alpha$ -glucosidasas son potencialmente aplicables en la etapa de pretratamiento de lignocelulosa en la producción de bioetanol de segunda generación y en la etapa de degradación de almidón en la producción de bioetanol de primera generación, respectivamente, entre otras aplicaciones industriales de menor interés para este trabajo.

En conclusión, se logró clonar fragmentos de dos glicosil hidrolasas potencialmente aplicables en un proceso de producción de bioetanol. La estrategia utilizada deberá ser perfeccionada para hacerla más específica, logrando el diseño de partidores degenerados más específicos o bien considerando una estrategia global alternativa, tal como utilizar una genoteca de cDNA como templado del método en lugar de DNA genómico.

## Agradecimientos

A mis padres José y Kyra por el amor de toda una vida, su eterno apoyo y fe en mí, y a mis hermanos Catalina y Fernando por su amor incondicional y paciencia.

A Oriana Salazar, mi profesora guía, por su sabiduría, ideas, entusiasmo y, sobre todo, por encontrarse siempre cerca para guiarme en este trabajo.

A mis amigos Rodrigo, Sofía, Cristián, Marcela, Alejandro, Sergio y D.H., por todos aquellos inolvidables momentos compartidos en este largo camino, ustedes hicieron el paso por la Universidad la mejor experiencia de todas.

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 .....	2
Agradecimientos .....	3
Índice de contenidos.....	4
Índice de figuras .....	8
Índice de tablas.....	9
Abreviaturas .....	10
1. Introducción .....	11
1.1 Marco teórico .....	11
1.1.1 Antecedentes generales .....	11
1.1.2 Uso del etanol como combustible .....	12
1.1.2.1 Principales ventajas del etanol.....	12
1.1.3 Materia prima para producir bioetanol.....	13
1.1.3.1 Caracterización del material lignocelulósico.....	14
1.1.4 Producción de bioetanol a partir de lignocelulosa [9]. .....	15
1.1.4.1 Procesos de pretratamiento [4] .....	16
1.1.4.2 Hidrólisis enzimática de celulosa [9].....	17
1.1.5 Sistemas celulasas .....	17
1.1.5.1 Celulasas de hongos .....	18
1.1.6 Clasificación de glicosil hidrolasas .....	19
1.1.7 Dominios conservados de enzimas [15].....	19
1.2 Descripción y justificación del proyecto .....	20
1.3 Objetivos .....	22
1.3.1 Objetivo general .....	22
1.3.2 Objetivos específicos .....	22
2. Materiales y métodos.....	23

2.1	Materiales .....	23
2.1.1	Reactivos .....	23
2.1.2	Equipos.....	23
2.1.3	Herramientas computacionales.....	24
2.1.4	Enzimas de restricción .....	24
2.1.5	Vector de clonación.....	24
2.2	Métodos .....	25
2.2.1	Construcción de micoteca .....	25
2.2.2	Ensayos de actividad .....	25
2.2.2.1	En medio sólido .....	25
2.2.2.2	En medio líquido .....	26
2.2.3	Diseño de partidores degenerados.....	26
2.2.4	Extracción de DNA genómico desde hongos .....	27
2.2.5	Amplificación de fragmentos por PCR .....	27
2.2.6	Secuenciamiento y análisis .....	28
2.2.7	<i>Genome walking</i> .....	28
2.2.7.1	Diseño de partidores .....	29
2.2.7.2	Construcción del adaptador .....	29
2.2.7.3	Digestión del DNA genómico.....	29
2.2.7.4	Construcción de genotecas oligo-adaptador .....	30
2.2.7.5	Primera ronda de PCR.....	30
2.2.7.6	Segunda ronda de PCR.....	31
2.2.8	Secuenciamiento de resultados de <i>genome walking</i> .....	32
3.	Resultados y discusión.....	33
3.1	Construcción de micoteca.....	33
3.2	Identificación y selección de cepas con actividad exoglucanasa. ....	34
3.3	Análisis de regiones de homología para el diseño de partidores degenerados.....	36

3.4	Extracción de DNA genómico desde hongos .....	37
3.5	Amplificación de fragmentos por PCR.....	38
3.6	Secuenciamiento y análisis.....	42
3.7	Obtención de secuencias de regiones aledañas para secuencias seleccionadas .....	48
3.7.1	<i>Genome walking</i> .....	48
3.7.2	Secuenciamiento y análisis .....	50
3.7.3	Reconstrucción de secuencias.....	52
3.7.3.1	Análisis de la presencia de intrones.....	53
3.8	Estimación del porcentaje de secuenciación de las enzimas de interés .....	56
3.9	Aplicaciones de las glicosil hidrolasas encontradas .....	57
3.9.1	$\beta$ -1,3(4)-D-glucanasas.....	57
3.9.2	$\alpha$ -glucosidasas.....	58
3.10	Discusión final.....	59
4.	Conclusiones.....	64
5.	Referencias.....	66
6.	Anexos.....	71
6.1	Micoteca de trabajo .....	71
6.2	Alfabeto degenerado .....	72
6.3	Métodos adicionales.....	72
6.3.1	Verificación de extracción de DNA .....	72
6.3.2	Purificación de DNA desde geles de agarosa.....	72
6.3.3	Ligación de fragmentos en vector de clonación .....	73
6.3.4	Transformación de bacterias <i>E. coli</i> DH5 $\alpha$ .....	73
6.3.5	Extracción de DNA plasmidial desde <i>E. coli</i> DH5 $\alpha$ .....	73
6.3.6	Verificación de inserción del inserto .....	73
6.3.7	Métodos alternativos de extracción de DNA .....	74
6.3.7.1	Método con acetato de potasio [27].....	74

6.3.7.2	Método con reactivo TRI® [28].....	74
6.4	Medios de cultivo.....	75
6.4.1	Medios líquidos.....	75
6.4.2	Medios sólidos.....	75
6.5	Tablas de datos.....	76
6.6	Alineamiento para familia 6.....	77
6.7	Secuencias .....	77

## Índice de figuras

Figura 1.1. Fuentes de azúcares para producción de etanol.....	16
Figura 1.2. Esquema del mecanismo de hidrólisis enzimática de celulosa por celulasas.....	18
Figura 2.1. Vector pGem®-T Easy y sus puntos de referencia.....	25
Figura 2.2. Técnica de <i>genome walking</i> utilizada. ....	31
Figura 3.1. <i>Alternaria alternata</i> y <i>Polysticus sulphureus</i> .....	33
Figura 3.2. <i>Trametes cinnabarinus</i> y <i>Pignosporus cinnabarinus</i> . ....	34
Figura 3.3. PCR de verificación extracción de DNA genómico. ....	38
Figura 3.4. Alineamiento de exoglucanasas de la familia 7 con partidores <i>exoF_fam7_fwd2</i> y <i>exoF_fam7_rev2</i> .....	39
Figura 3.5. Amplificación de DNA genómico de hongos con partidores dirigidos a la familia 6. ....	41
Figura 3.6. Amplificación de DNA genómico de hongos con partidores dirigidos a la familia 7. ....	42
Figura 3.7. Clusters dentro de la superfamilia 16 [31].....	44
Figura 3.8. Resultado de análisis de dominios conservados realizado usando la herramienta CDD [15] para la secuencia 147Ex3A en el marco de lectura +1. ....	44
Figura 3.9. Resultado de análisis de dominios conservados realizado usando la herramienta CDD [15] para la secuencia 147Ex3B en el marco de lectura +1. ....	45
Figura 3.10. Alineamiento de las secuencias 147Ex3B, 147Ex 4A y 147Ex 4B.....	45
Figura 3.11. Resultado del análisis con BLASTX para la secuencia 147Ex5A.....	46
Figura 3.12. Resultado de análisis de dominios conservados realizado usando la herramienta CDD [15] para la secuencia 163Ex6A traducida en el marco -2. Para detalles de la simbología de esta figura referirse a la Figura 3.8.....	47
Figura 3.13. Alineamiento de las secuencias 6A y 6B. ....	47
Figura 3.14. Esquema de los partidores diseñados para la técnica de <i>genome walking</i> . ....	49
Figura 3.15. Segunda ronda de PCR de <i>genome walking</i> hacia el extremo 5'. ....	50
Figura 3.16. Repetición segunda ronda de PCR de <i>genome walking</i> hacia el extremo 5'.....	50
Figura 3.17. Alineamiento en ClustalX de secuencias 127PII31A, 127PII31B y 127EV31A.....	52
Figura 3.18. Esquema de la estructura intrón-exón de un gen genérico. ....	54
Figura 3.19. Secuencias intrónicas de consenso descritas para <i>C. neoformans</i> . ....	55
Figura 6.1. Digestión de DNA plasmidial.....	74
Figura 6.2. Alineamiento secuencias familia 6 con partidores. ....	77



## Índice de tablas

Tabla 1.1. Contenido de celulosa, hemicelulosa y lignina en residuos agrícolas y desechos. ....	15
Tabla 2.1. Listado de reactivos.....	23
Tabla 2.2. Equipos utilizados en laboratorio.....	23
Tabla 2.3. Herramientas bioinformáticas utilizadas. ....	24
Tabla 2.4. Enzimas de restricción utilizadas. ....	24
Tabla 2.5. Partidores degenerados utilizados.....	27
Tabla 2.6. Partidores diseñados para realizar <i>genome walking</i> de los fragmentos seleccionados.....	29
Tabla 3.1. Datos de familias de glicosil hidrolasas que contienen exoglucanasas [13]. ....	36
Tabla 3.2. Bandas seleccionadas para purificación. ....	40
Tabla 3.3. Resumen de resultados parciales.....	47
Tabla 3.4. Porcentaje secuenciado estimado para las secuencias en estudio.....	57
Tabla 6.1. Micoteca de trabajo.....	71
Tabla 6.2. Alfabeto degenerado. ....	72
Tabla 6.3. Datos obtenidos en los ensayos de actividad.....	76

## Abreviaturas

Abreviaturas	Significado
μF	Microfaraday
CDD	Conserved Domain Database
cDNA	DNA complementario
CMC	Carboximetil celulosa
DNA	Ácido desoxirribonucleico
DNS	Ácido dinitrosalicílico
dNTPs	Desoxiribonucleótidos trifosfato
EDTA	Etilendiaminotetraacetato
GHs	Glicosil hidrolasas
Kb	Kilobase
LB	Luria Bertani
LTR	Repeticiones terminales largas
MTBE	metil terbutil éter
mRNA	RNA mensajero
NCBI	National Center for Biotechnology Information
pb	Pares de bases
PCR	Reacción en cadena de polimerasa
RNA	Ácido Ribonucleico
RPM	Revoluciones por minuto
RT	Transcriptasa reversa
SDS	Dodecil sulfato de sodio
ss	Sitio de <i>splice</i>
SSSF	Sodium solute symporter family
T <sub>m</sub>	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-galactopiranósido

# 1. Introducción

## 1.1 Marco teórico

### 1.1.1 Antecedentes generales

Las enzimas catalizan reacciones químicas con gran especificidad y acrecentamiento de la velocidad de reacción. Estas reacciones son la base del metabolismo de todos los organismos vivos, y proveen tremendas oportunidades para la industria, de llevar a cabo conversiones biocatalíticas eficientes y económicas. En el sentido más amplio, la tecnología de enzimas puede ser vista como la aplicación de enzimas libres o biocatalizadores celulares en la producción de bienes y servicios. Ésta es una ciencia interdisciplinaria, reconocida como un importante componente del desarrollo industrial sustentable. Al mismo tiempo, el rápido desarrollo de la ingeniería genética, técnicas de *screening* de alto rendimiento y otras tecnologías establecidas o emergentes, mejoran la economía de los procesos existentes y abren nuevos campos de aplicación [1].

Las enzimas ofrecen ventajas sustanciales por sobre los catalizadores químicos, en cuanto a que son derivadas de recursos renovables, son biodegradables, trabajan bajo condiciones relativamente moderadas de temperatura y pH, y tienden a ofrecer una exquisita selectividad en la estereoquímica tanto de reactantes como de productos. A medida que las preocupaciones medioambientales aumentan por presión de la población y la industrialización del planeta, la necesidad de encontrar métodos más benignos de producción de bienes y provisión de servicios da lugar a mayor interés en las enzimas [2].

En la industria del alcohol, el uso de enzimas para la producción de azúcares fermentables desde almidón está bien establecido. Durante la década pasada, el interés en el alcohol combustible se ha acrecentado producto de un aumento en las consideraciones medioambientales, precios más altos del petróleo crudo y, en mayor medida, la prohibición en ciertas regiones del aditivo de gasolina metil terbutil éter (MTBE), el cual puede ser intercambiado directamente por etanol. Por tanto, se están llevando a cabo intensos esfuerzos por desarrollar enzimas mejoradas, que permitan la utilización de sustratos más baratos y parcialmente utilizados, como lignocelulosa, para hacer del bioetanol un producto competitivo con los combustibles derivados del petróleo (combustibles fósiles). El costo de las enzimas requeridas para transformar lignocelulosa en una materia prima adecuada para

fermentación es un asunto primordial, y los esfuerzos actuales se focalizan tanto en desarrollar enzimas con mejor actividad y estabilidad como en su producción eficiente [3].

#### 1.1.2 Uso del etanol como combustible

El etanol ha sido conocido por largo tiempo, siendo tal vez el producto más antiguo obtenido a través de biotecnología tradicional. Sus aplicaciones actuales incluyen etanol potable, químico y combustible. Los automóviles impulsados por etanol ya habían sido ideados en el siglo XIX, sin embargo, a inicios del siglo XX los combustibles fósiles comenzaron a aparecer y rápidamente dominaron el mercado. Los bajos precios de éstos persistieron hasta el advenimiento de la crisis del petróleo en la década de 1970. Esta crisis subrayó la importancia de fuentes de energía alternativas, entre ellas etanol, dando origen a iniciativas tales como el Programa Nacional de Alcohol en Brazil y el programa de gasohol en Estados Unidos. A medida que los precios del petróleo disminuyeron, también disminuyó el interés por etanol combustible [4]. No obstante, durante las últimas décadas el excesivo consumo de combustibles fósiles, particularmente en las grandes áreas urbanas, ha contribuido enormemente a generar altos niveles de contaminación. Como un paso para resolver este problema, la adición de etanol a la gasolina ha sido ampliamente aplicada en los últimos años, lo cual reduce las emisiones de monóxido de carbono e hidrocarburos no combustionados que forman smog [5].

##### *1.1.2.1 Principales ventajas del etanol.*

Las principales ventajas del uso del bioetanol por sobre combustibles fósiles se comentan a continuación.

**Rendimiento.** El etanol tiene un mayor octanaje que la gasolina, permitiendo que el motor funcione a una mayor tasa de compresión y de este modo aumente su rendimiento neto [5]. Adicionalmente, la presión de vapor del alcohol 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 alcohol. Sin embargo, un galón de etanol puro, debido a su contenido de oxígeno, posee 33% menos de energía que la gasolina [6].

**Niveles de contaminación.** La combustión incompleta de combustibles fósiles emite gases que contribuyen a la formación de smog. Además, la extracción, procesamiento y combustión de éstos resulta en contaminación del aire, agua y suelo y son, por lo tanto, peligrosos para el ambiente y la salud pública. Al utilizar etanol derivado de biomasa, se reducen en forma neta

los niveles de dióxido de carbono (el principal gas invernadero) entre 60-90%, debido a que el CO<sub>2</sub> liberado por combustión se reutiliza en la fotosíntesis. La combustión de etanol tiene la desventaja de aumentar la emisión de aldehídos, la cual es entre dos y cuatro veces más baja al utilizar gasolina [4].

**Agotamiento del petróleo.** El agotamiento de los combustibles fósiles genera serias preocupaciones acerca de las fuentes de energía, y el Instituto de Petróleo Americano ha estimado que el suministro de petróleo se agotará en algún punto del siglo XXI (aún cuando nuevos pozos siguen siendo encontrados) [4].

**Costos.** Los costos directos de producción de bioetanol son mayores que los costos de producción de gasolina. Sin embargo, los costos indirectos de combustibles derivados del petróleo han sido subestimados, ya que entre ellos se cuentan los costos asociados a la gasolina como contaminante del aire (que pueden ascender a US\$45 por barril) y la defensa de los pozos de petróleo en Medio Oriente [4].

### 1.1.3 Materia prima para producir bioetanol

Las tecnologías mejor desarrolladas para la producción de etanol son basadas en cultivos, utilizando sustratos tales como caña de azúcar y almidón de maíz. El etanol obtenido a partir de este tipo de materias primas recibe el nombre de bioetanol de primera generación. Dado que los costos de las materias primas pueden ascender a un 40% del costo total, surgió como alternativa el llamado bioetanol de segunda generación, el cual corresponde a etanol producido a partir de residuos lignocelulósicos. La lignocelulosa es un polímero natural, abundante y potencialmente económico, que se encuentra en los desechos agrícolas (paja de trigo, tallos de maíz, residuos de soya, bagazo de caña de azúcar, etc.), desechos industriales (industria de papel y pulpa), residuos forestales, desechos sólidos municipales, etc. Se ha estimado que la lignocelulosa da cuenta de aproximadamente un 50% de la biomasa del planeta (10-50 billones de toneladas) [4].

En Chile, aún no existe consenso respecto a los volúmenes y ubicación de la biomasa residual producida. Las estimaciones sobre la disponibilidad de residuos generados por la industria del aserrado, asociada a las plantaciones de pino solamente, son de 1,4 millones de metros cúbicos. A esto se debe sumar el volumen correspondiente a los desechos producidos en el manejo de los bosques nativos productivos los cuales se estiman, conservadoramente, en alrededor de 7 millones de toneladas al año. Toda esta biomasa podría abastecer una

industria de bioetanol y biodiesel de segunda generación, el cual es uno de los objetivos señalados claramente por las autoridades nacionales a través de la Comisión Nacional de Energía y que recientemente enfatiza la prioridad que Chile dará a este tipo de biocombustibles [7].

#### *1.1.3.1 Caracterización del material lignocelulósico*

En contraste con los cultivos que contienen azúcares, la utilización de lignocelulosa como sustrato para la producción de etanol presenta la barrera de su compleja estructura, la cual es resistente a la degradación. La lignocelulosa está compuesta por tres fracciones principales: celulosa, hemicelulosa y lignina, las cuales corresponden aproximadamente al 45, 30 y 25% del peso seco de la lignocelulosa, respectivamente [4]. La celulosa y hemicelulosa son polisacáridos que pueden ser hidrolizados a azúcares y eventualmente fermentados a etanol. La lignina no puede ser utilizada para la producción de etanol [8]. A continuación se caracterizan las fracciones de la lignocelulosa.

**Celulosa.** Es el polímero más abundante en la Tierra, compuesta por miles de moléculas de anhidroglucosa unidas por enlaces  $\beta$ -1,4 glicosídicos. La unidad básica repetitiva es la celobiosa disacárido. La conformación secundaria y terciaria de la celulosa, así como su estrecha asociación con lignina, hemicelulosa, almidón y elementos proteicos y minerales, convierten a la celulosa en una molécula resistente a la hidrólisis. El producto de ésta es glucosa, un monosacárido de seis carbonos (hexosa) [4] [8].

**Hemicelulosa.** Es un heteropolímero altamente ramificado, que contiene residuos de azúcares tales como hexosas (D-galactosa, L-galactosa, D-manosa, L-fucosa), pentosas (D-xilosa, L-arabinosa) y ácidos urónicos (ácido D-glucurónico), aunque su composición depende de la fuente de origen. La hemicelulosa es más fácilmente hidrolizada que la celulosa [4].

**Lignina.** Es el polímero aromático más abundante en la naturaleza. De carácter fenólico, 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. Toda la biomasa lignocelulósica posee lignina por lo que siempre será un desecho de los procesos de producción de etanol [4] [8].

En la Tabla 1.1 se presenta el contenido de celulosa, hemicelulosa y lignina en residuos agrícolas y desechos típicos [9].

**Tabla 1.1. Contenido de celulosa, hemicelulosa y lignina en residuos agrícolas y desechos.**

Material lignocelulósico	Celulosa [%]	Hemicelulosa [%]	Lignina [%]
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/A	24-29
Desechos de puercos	6	28	N/A
Estiércol sólido de ganado	1,6-4,7	1,4-3,3	2,7-5,7

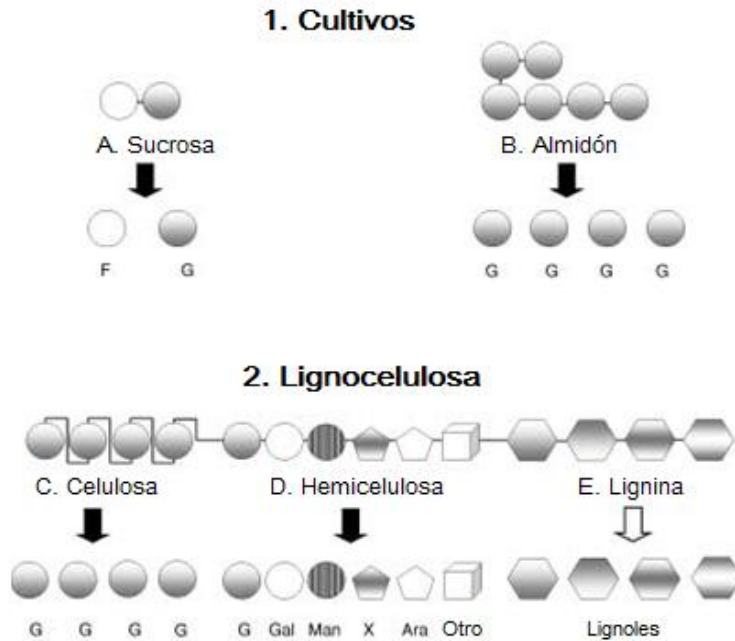
#### 1.1.4 Producción de bioetanol a partir de lignocelulosa [9].

La conversión de lignocelulosa a etanol involucra dos procesos:

1. La hidrólisis de la celulosa en el material lignocelulósico para convertirla en azúcares fermentables.
2. La fermentación de los azúcares a etanol.

En la Figura 1.1 se presenta la obtención de azúcares fermentables a partir de distintas fuentes [4]. La hidrólisis es usualmente catalizada por enzimas celulasas, y la fermentación es llevada a cabo por levaduras o bacterias. Los factores que afectan la hidrólisis de celulosa identificados hasta ahora incluyen la porosidad (área superficial accesible) de los materiales de desecho, la cristalinidad de las fibras de celulosa y el contenido de lignina y hemicelulosa. La presencia de lignina y hemicelulosa dificulta el acceso de las celulasas a la celulosa, reduciendo de este modo la eficiencia de la hidrólisis. La remoción de la lignina y la hemicelulosa, la reducción de la cristalinidad y el aumento de la porosidad en procesos de pretratamiento de la materia prima puede mejorar significativamente la hidrólisis.

El programa Domeyko, en el cual se incluye esta memoria, contempla la realización de un proyecto de optimización del proceso de tratamiento de lignocelulosas para la producción de bioetanol, enfocándose en (i) los procesos de pretratamiento de lignocelulosa y (ii) la hidrólisis de celulosa. Estos procesos se describen con mayor detalle a continuación.



**Figura 1.1. Fuentes de azúcares para producción de etanol.**

1 Cultivos; 2 Lignocelulosa. Las flechas representan hidrólisis (sólo los monómeros generados por hidrólisis representada por flechas negras pueden ser fermentados). *G* glucosa, *Gal* galactosa, *F* fructosa, *Man* manosa, *X* xilosa, *Ara* arabinosa, *Otro* L-ramnosa, L-fucosa, ácidos urónicos.

#### 1.1.4.1 Procesos de pretratamiento [4]

Existen distintos tipos de pretratamientos del material lignocelulósico, los cuales se mencionan a continuación.

**Pretratamiento físico.** Pulverización mecánica, pirolisis.

**Pretratamiento físicoquímico.** Explosión por vapor (autohidrólisis), explosión de fibras por amonio (AFEX), explosión por CO<sub>2</sub>, pretratamiento químico, ozonólisis, hidrólisis ácida, hidrólisis alcalina, delignificación oxidativa, procesos de solventes orgánicos.

**Pretratamiento biológico.** Se utilizan microorganismos tales como hongos de pudrición café, blanca y blanda, para degradar la lignina y la hemicelulosa en los materiales de desecho. Los hongos de pudrición café atacan principalmente celulosa, mientras que los de pudrición blanca y blanda atacan tanto la celulosa como la lignina. Los de pudrición blanca son los *Basidiomycetos* más efectivos para el pretratamiento biológico. Entre las enzimas que contribuyen a la degradación de la lignina se encuentran lignina peroxidasas, peroxidasas Mn-dependientes, polifenol oxidasas, lacasas, enzimas productoras de H<sub>2</sub>O<sub>2</sub> y enzimas reductoras de quinonas. Para evitar la pérdida de celulosa, se han desarrollado mutantes sin celulasas.



Las ventajas del pretratamiento biológico incluyen un bajo requerimiento de energía y condiciones de reacción suaves. No obstante, la tasa de hidrólisis en la mayoría de los pretratamientos biológicos es muy baja.

#### *1.1.4.2 Hidrólisis enzimática de celulosa [9].*

La hidrólisis enzimática de la celulosa es llevada a cabo por enzimas celulasas altamente específicas. Los productos de hidrólisis son usualmente azúcares reductores, incluyendo glucosa. El costo de la hidrólisis enzimática es por lo general bajo comparado con la hidrólisis ácida o alcalina, debido a que la hidrólisis enzimática es llevada a cabo en condiciones de reacción moderadas (pH 4,8 y temperatura 45-50°C), y no presenta problemas de corrosión. Tanto bacterias como hongos pueden producir celulasas para la hidrólisis de material lignocelulósico.

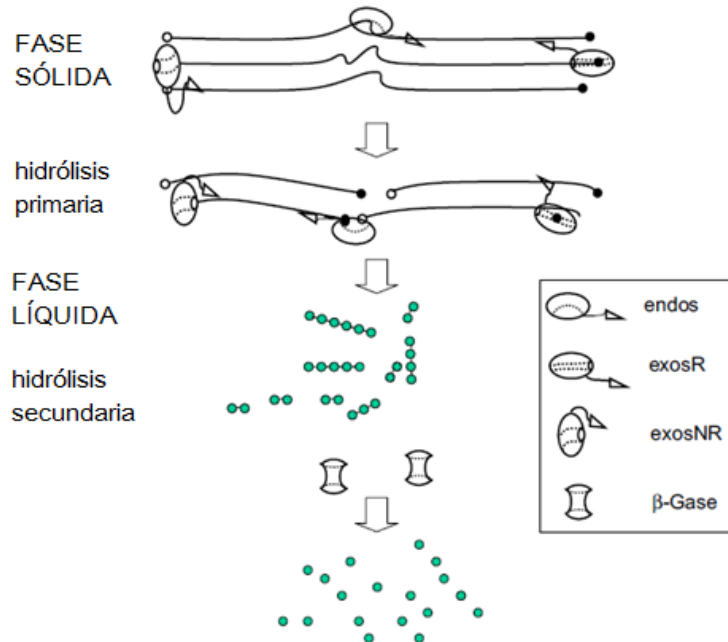
#### 1.1.5 Sistemas celulasas

Los sistemas enzimáticos celulasas están compuestos por múltiples proteínas que interactúan desempeñando funciones complementarias relacionadas con la hidrólisis de celulosa y, en algunos casos, formando complejos multiproteicos [10].

El mecanismo ampliamente aceptado involucra 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 los enlaces  $\beta$ -1,4-glicosídicos intramoleculares de la molécula de celulosa accesibles al azar, produciendo nuevos extremos en la cadena; las exoglucanasas rompen cadenas de celulosa en los extremos, liberando celobiosa soluble o glucosa; y las  $\beta$ -glucosidasas hidrolizan celobiosa a glucosa, evitando la inhibición por celobiosa [11].

La celulosa, como principal componente de las paredes celulares vegetales, determina que la mayoría de las enzimas que la degradan sean extracelulares [11]. Este hecho es clave para los ensayos de actividad que se han desarrollado para su detección, tanto en medio sólido como en medio líquido, como se verá más adelante.

En la Figura 1.2 se presenta un esquema del mecanismo de hidrólisis de celulosa por acción de los sistemas celulasas previamente mencionados [11].



**Figura 1.2. Esquema del mecanismo de hidrólisis enzimática de celulosa por celulasas.** *Endos* endoglucanasas; *exosR* y *exosNR* exoglucanasas que catalizan desde el extremo reductor y no reductor de la molécula de celulosa, respectivamente; *β-Gase* β-glucosidasas.

### 1.1.5.1 Celulasas de hongos

La capacidad de digerir celulosa está ampliamente distribuida entre los muchos géneros de los dominios *Bacteria* y en los grupos fúngicos dentro del dominio *Eukarya*. Los hongos son conocidos agentes de la descomposición de materia orgánica en general, y en particular de sustratos celulósicos, cuya utilización está presente en especies de todo el reino, desde los primitivos *Chytridomycetes* a los evolucionados *Basidiomycetes*. Sin embargo, la taxonomía de hongos tradicionalmente se ha basado en su morfología más que en sus capacidades de utilización de sustratos, razón por la cual a la fecha no es claro cuán ampliamente se extiende la capacidad celulolítica entre los hongos, aunque se pueden mencionar algunos géneros con miembros que han sido objeto de considerable estudio con respecto a su habilidad celulolítica y/o de degradación de madera: *Bulgaria*, *Chaetomium* y *Helotium* (*Ascomycetes*); *Coriolus*, *Phanerochaete*, *Poria*, *Schizophyllum* y *Serpula* (*Basidiomycetes*); y *Aspergillus*, *Cladosporium*, *Fusarium*, *Geotrichum*, *Myrothecium*, *Paecilomyces*, *Penicillium* y *Trichoderma* (*Deuteromycetes*) [10]. De todos los géneros mencionados, *Trichoderma* ha sido el más extensamente estudiado para la producción de celulasas [9].

### 1.1.6 Clasificación de glicosil hidrolasas

Las glicosil hidrolasas (GHs) son enzimas clave del metabolismo de carbohidratos, y se encuentran presentes en los tres reinos principales (arqueas, bacterias y eucariontes). La nomenclatura de enzimas IUB (International Union of Biochemistry, 1984) se basa en los tipos de reacciones que catalizan y en su especificidad de sustrato. Para las GHs, EC 3.2.1.x, los primeros tres dígitos indican enzimas que hidrolizan enlaces O-glicosídicos, mientras que el último número indica el sustrato y a veces refleja el mecanismo molecular. Esta clasificación es muy útil, especialmente para evitar ambigüedades y la proliferación de nombres triviales, y provee una clasificación única. Sin embargo, al menos en el caso de las GHs, tal clasificación no necesariamente refleja las características estructurales de las enzimas [12].

Por lo tanto, se ha propuesto una clasificación en familias basada en la similitud de las secuencias primarias de las GHs. Esta clasificación indicó que (i) existen familias que contienen enzimas con distintos números EC y (ii) enzimas con especificidad de sustrato similares que pertenecen a familias no relacionadas. La clasificación permite relacionar enzimas con similitud en su secuencia, la cual se relaciona con una similitud en su plegamiento y funcionalidad [12]. Actualmente, dicha clasificación comprende la definición de 114 familias de GHs [13].

Además, se han definido clanes, los cuales son grupos de familias que se piensa tienen un ancestro común y son reconocidas por similitudes significativas en su estructura terciaria, junto con conservación de los residuos catalíticos y el mecanismo catalítico [14]. A la fecha se han definido 14 clanes [13].

### 1.1.7 Dominios conservados de enzimas [15]

Los dominios pueden ser vistos como unidades claras de una proteína, funcionales y/o estructurales. De hecho, estas dos clasificaciones coinciden muy a menudo, y lo que corresponde a una unidad de plegamiento independiente de una cadena polipeptídica también conlleva una función específica. Los dominios son frecuentemente identificados como unidades recurrentes (de secuencia o estructura), las cuales pueden existir en varios contextos.

En la evolución molecular, estos dominios podrían haber sido utilizados como “bloques de construcción”, y podrían haber sido recombinados en distintos arreglos para modular la función de proteínas. Se definen los dominios conservados como unidades recurrentes en la

evolución molecular, cuya extensión puede ser determinada por análisis de secuencia y estructural.

Los dominios conservados contienen patrones de secuencia conservados o motivos, los cuales permiten su detección en secuencias polipeptídicas.

Una jerarquía de familia de dominios es un set de dominios relacionados, que comparten un ancestro común, un set de residuos conservados comunes y una función general común, pero que difieren entre sí en su filogenia y función específicas, y en regiones adicionales de residuos conservados. Las jerarquías están presentes en la base de datos de dominios conservados del NCBI para entregar una percepción de cómo los patrones de conservación de residuos y divergencia en una familia se relaciona con las propiedades funcionales.

Un cluster de superfamilia es un set de modelos de dominios conservados que generan una anotación traslapada en la misma secuencia proteica. Se asume que estos modelos representan dominios relacionados evolutivamente y pueden ser redundantes entre sí.

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

En base a los antecedentes expuestos anteriormente, se puede concluir que existe una necesidad urgente de disminuir el uso de combustibles fósiles, los cuales tienen un efecto perjudicial tanto en la población como en el medioambiente. Una alternativa a los combustibles derivados del petróleo se encuentra en el etanol, el cual puede ser obtenido por fermentación de azúcares provenientes de cultivos o desechos lignocelulósicos (bioetanol de primera y segunda generación, respectivamente). En Chile, la abundancia de desechos lignocelulósicos convierte al bioetanol de segunda generación en la alternativa más viable económicamente.

La lignocelulosa está constituida por tres macromoléculas fundamentales: celulosa, hemicelulosa y lignina. La proporción de cada una de éstas está definida por la naturaleza de la lignocelulosa.

La producción de bioetanol de segunda generación consiste en tres etapas principales: el pretratamiento de la lignocelulosa para exponer la celulosa presente, la hidrólisis de celulosa para obtener azúcares fermentables y la fermentación de estos azúcares a etanol. Para aumentar la rentabilidad del proceso de producción, las etapas clave que requieren optimización son el pretratamiento de lignocelulosa y la hidrólisis de celulosa.

La hidrólisis de celulosa es catalizada por enzimas, las cuales son comúnmente conocidas como celulasas. Las celulasas son de tres tipos: endoglucanasas, exoglucanasas y  $\beta$ -glucosidasas, cuya acción sinérgica degrada la molécula de celulosa a pentosas y hexosas que luego pueden ser fermentadas por microorganismos. Las celulasas son enzimas del tipo glicosil hidrolasas, las cuales se han clasificado en familias basándose en similitudes de secuencia aminoacídica. La similitud de secuencia aminoacídica se relaciona estrechamente con una similitud estructural y funcional.

El propósito de esta memoria es identificar y clonar glicosil hidrolasas de hongos, con énfasis en exoglucanasas, para su uso en hidrólisis de celulosa en procesos de producción de bioetanol de segunda generación. Para esto, el primer paso será seleccionar cepas productoras de exoglucanasas mediante un método de *screening* a gran escala, a partir de una colección de hongos, en su mayoría *Basidiomycetos*, facilitada por la Facultad de Ciencias Forestales de la Universidad de Chile. Una vez seleccionadas las cepas, se procederá a extraer el DNA genómico de las mismas, el cual será el templado en una reacción de PCR, con partidores dirigidos a amplificar regiones conservadas de exoglucanasas de hongos. El diseño de estos partidores se realizará utilizando la técnica CODEHOP. Los fragmentos amplificados serán luego secuenciados y analizados *in silico* y a partir de ellos se diseñará partidores específicos para poder completar la secuencia del gen codificante mediante *genome walking*. Los productos serán secuenciados y analizados *in silico* con el fin de reconstruir las secuencias y determinar el producto codificado por los genes encontrados.

## **1.3 Objetivos**

### 1.3.1 Objetivo general

Identificar y clonar genes de glicosil hidrolasas, con énfasis en enzimas potencialmente útiles para el proceso de producción de bioetanol.

### 1.3.2 Objetivos específicos

- Identificar cepas de hongos productoras de enzimas glicosil hidrolasas capaces de degradar sustratos de celulosa insoluble.
- Obtener fragmentos de DNA codificante de las enzimas encontradas y clonarlos en un sistema bacteriano.
- Inferir el producto proteico de las secuencias encontradas mediante alineamientos con secuencias en bases de datos, utilizando herramientas computacionales.

## 2. Materiales y métodos

### 2.1 Materiales

#### 2.1.1 Reactivos

En la Tabla 2.1 se presentan los reactivos utilizados en el trabajo de laboratorio, con el respectivo proveedor.

**Tabla 2.1. Listado de reactivos.**

Laboratorio	Reactivos
Difco	Agar, malta, LB
Ambion	TRI®
CalBiochem	Ampicilina
Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile	N <sub>2</sub> (l)
Fermentas	Proteinasa K, EcoRV, PuvII, X-Gal, IPTG
Fluka	Avicel
IDT	Oligonucleótidos
Invitrogen	dNTPs, <i>E. coli</i> DH5α, kit elongasa, estándar de peso molecular para geles de agarosa (1 Kb).
Lafken	Agarosa
Merck	Papa dextrosa, sales inorgánicas, ácido cítrico, EDTA, SDS, ácido acético glacial, etanol, cloroformo, glucosa
New England Biolabs	EcoRI
Promega	Kit Taq DNA polimerasa, kit pGem®-T Easy
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 equipos utilizados con su respectivo fabricante y modelo.

**Tabla 2.2. Equipos utilizados en laboratorio.**

Equipo	Fabricante y modelo
Agitador orbital	LabLine MaxQ 4000
Centrífuga	BOECO M-24
Thermomixer	Eppendorf Thermomixer comfort
Lector de absorbancia (para placas de 96 pocillos)	Anthos 2010
Termociclador	MJ Research PTC 100

Equipo	Fabricante y modelo
	Eppendorf Mastercycler Gradient
Electroporador	Gibco BRL Cell-Porator
Fuente de poder para electroforesis	BRL Model 500
Transiluminador para geles de agarosa	Vilber Loumat
Cámara fotográfica	Nikon Coolpix 4500

### 2.1.3 Herramientas computacionales

En la Tabla 2.3 se presentan las herramientas bioinformáticas utilizadas para los análisis *in silico* realizados en este trabajo.

**Tabla 2.3. Herramientas bioinformáticas utilizadas.**

Herramienta	Fuente
BLASTX	NCBI [16].
BlockMaker	Fred Hutchinson Cancer Research Center [17].
CDD	NCBI [15]
ClustalX	The Conway Institute of Biomolecular and Biomedical Research, University College Dublin [18].
CODEHOP	Department of Pathobiology, School of Public Health and Community Medicine, University of Washington [19].
OligoCalc	Northwestern University [20]
Translate tool	ExpPASy [21]

### 2.1.4 Enzimas de restricción

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

**Tabla 2.4. Enzimas de restricción utilizadas.**

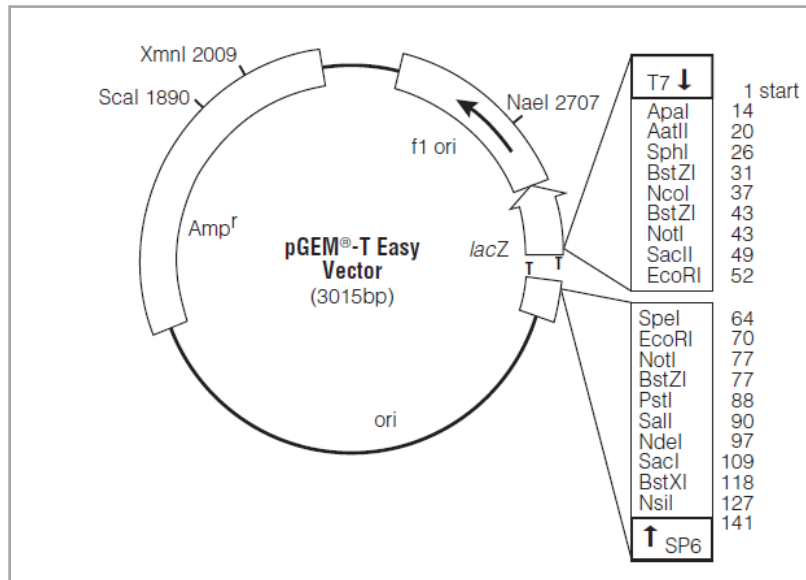
El símbolo ^ indica el sitio de corte.

Enzima	Sitio de restricción
EcoRI	GAT <sup>^</sup> ATC CTA <sup>^</sup> TAG
EcoRV	G <sup>^</sup> AATTC CTTAA <sup>^</sup> G
PuvII	CAG <sup>^</sup> CTG GTC <sup>^</sup> GAC

### 2.1.5 Vector de clonación

En la Figura 2.1 se muestra un esquema del vector de clonación pGem®-T Easy y sus principales puntos de referencia [22].





**Figura 2.1. Vector pGem®-T Easy y sus puntos de referencia.**

Las regiones codificadas de utilidad para este trabajo corresponden a la resistencia a ampicilina  $Amp^r$  y el gen  $lacZ$ , interrumpido por el sitio de múltiple clonamiento. En el recuadro a la derecha se listan los sitios de restricción disponibles en el sitio de múltiple clonamiento.

## 2.2 Métodos

### 2.2.1 Construcción de micoteca

Se recibieron 30 cepas de hongos de diversos orígenes, desde la Facultad de Ciencias Forestales de la Universidad de Chile. En la Tabla 6.1 de los Anexos se presentan las cepas estudiadas. Cada una de las muestras fue cultivada en duplicado en medio agar malta (ver Anexo 6.4) para crear una micoteca de trabajo y una de reserva, en caso de contaminación de las muestras. El cultivo de hongos se realiza en condiciones estériles, se debe cortar un cuadrado de aproximadamente 1 cm de lado del cultivo original con ayuda de un bisturí y ubicarlo en el centro de una placa con medio agar malta virgen. El crecimiento depende mucho de la cepa, pero las muestras utilizadas demoraron aproximadamente entre 5 y 15 días en cubrir la placa. El crecimiento se llevó a cabo a 28°C.

### 2.2.2 Ensayos de actividad

#### 2.2.2.1 En medio sólido

Se cultivaron las cepas en medio agar malta Avicel (ver Anexo 6.4), siguiendo la misma metodología explicada anteriormente. Las placas se crecieron por aproximadamente 5 días a 28°C, tiempo durante el cual los hongos excretarán exoglucanasas al medio extracelular, las cuales degradarán el Avicel. Transcurrido este tiempo, se realiza una tinción de la placa

utilizando rojo Congo al 0,01% durante 30 min. A continuación, las placas se lavan con NaCl 1M durante 15 min con agitación orbital suave. Este procedimiento produce halos sin tinción, los cuales corresponden a zonas donde ha ocurrido degradación de Avicel.

#### *2.2.2.2 En medio líquido*

Las cepas se cultivan en medio líquido para poder utilizar el sobrenadante en ensayos de actividad. Para obtener un cultivo en medio líquido se corta un cuadrado de aproximadamente 1 cm de lado del cultivo en placa y se inocula en un volumen de 50 ml de medio mínimo-Avicel (ver Anexo6.4). Los cultivos se mantienen a 28°C con agitación orbital a 200 RPM por tiempos definidos por los ensayos de actividad, detallado en el capítulo de Resultados.

A continuación se realizan los ensayos de actividad, los cuales fueron conducidos en una serie de días para cada cepa, de modo de detectar el peak de actividad para cada una. El ensayo de actividad se realiza tomando 1 ml del sobrenadante del medio de cultivo, el cual se centrifuga a 14.000 RPM durante 10 min a 4°C. A continuación se lleva a cabo la reacción enzimática, para lo cual se toman 50 µl del sobrenadante y se transfiere a una placa que contiene 100 µl de Avicel al 1% (en buffer citrato 50 mM pH 4,8), lo cual se incuba por 30 min a 37°C, agitando a 1.100 RPM en Thermomixer. Para detener la reacción, se transfieren 90 µl de la mezcla de reacción a una placa que contiene 90 µl de DNS, lo cual se lleva a 100°C por espacio de 10 minutos. Si ha ocurrido degradación de Avicel por actividad enzimática, el medio adquiere un color rojizo, el cual es directamente proporcional a la concentración de azúcares reductores en la mezcla, los cuales provienen de la degradación de la celulosa. Es por esto que se cuantifica la coloración del medio, midiendo la absorbancia de 100 µl a 550 nm.

Paralelamente se realiza un experimento “blanco”, el cual corresponde a la degradación de la celulosa presente en el medio de cultivo, es decir, se pone el sobrenadante centrifugado de cada cultivo en presencia de buffer citrato, el cual **no** contiene Avicel.

Todos los ensayos de actividad fueron realizados en duplicado, ambos valores fueron promediados y al valor obtenido se sustrae el valor promedio del ensayo blanco, también realizado en duplicado.

#### 2.2.3 Diseño de partidores degenerados

Se tomaron las secuencias aminoacídicas de exoglucanasas de hongos de las familias 6 y 7, obtenidas de las bases de datos de EMBL-EBI (<http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz>).

Éstas fueron alineadas utilizando la herramienta ClustalX2 para identificar regiones de homología. A continuación, se diseñaron partidores degenerados siguiendo la estrategia CODEHOP [19], la que involucra la construcción de bloques de homología utilizando la herramienta BlockMaker, los cuales son utilizados como *input* por el software CODEHOP para el diseño de partidores degenerados. Utilizando esta estrategia, se diseñaron partidores *forward* y *reverse* para la familia 6 de las glicosil hidrolasas.

Por otro lado, una revisión bibliográfica permitió encontrar un trabajo en el cual se clonó una celobiohidrolasa del hongo *Penicillium chrysogenum* utilizando partidores degenerados dirigidos a la familia 7 [23]. Estos partidores también fueron utilizados.

Los partidores sintetizados son presentados en la Tabla 2.5.

**Tabla 2.5. Partidores degenerados utilizados.**

Partidor	Familia	Sentido	Secuencia 5' a 3' <sup>a</sup>
exoF_fam6_fwd	6	<i>Forward</i>	cag ccc atc tac gac gar wan mrn ta
exoF_fam6_rev	6	<i>Reverse</i>	tca gca gct gca cga agt agk cnt gra acc a
exoF_fam7_fwd2	7	<i>Forward</i>	ntc att sac gcc awc tgg
exoF_fam7_rev2	7	<i>Reverse</i>	sct stc gag cca caa cat

<sup>a</sup> El alfabeto degenerado utilizado se presenta en el Anexo 6.2.

#### 2.2.4 Extracción de DNA genómico desde hongos

Los hongos fueron cultivados en medio sólido agar papa dextrosa (ver Anexo 6.4). Se toma una muestra de tejido desde estos cultivos en placa, aproximadamente el tejido que se obtiene en la mitad de la superficie de una placa Petri. El tejido se introduce en un tubo Eppendorf de 1,5 ml estéril y se agregan 100 µl de mezcla de proteinasa K (0,1 µg/µl de proteinasa K en buffer 0,25 M Tris HCl, 0,01 M CaCl<sub>2</sub>). El tubo es llevado por 30 s a N<sub>2</sub> (l), luego a 60°C por 2 min y luego a vórtex por 2 min, estos pasos se repiten una vez. A continuación, se lleva por 30 s a N<sub>2</sub> (l) y luego a 95°C por 15 s, 4°C por 30 s y nuevamente vórtex por 2 min. Finalmente, se centrifuga por 10 min a 9.000 RPM y se transfiere el sobrenadante a un tubo Eppendorf de 0,6 ml estéril. La extracción de DNA se verificó haciendo un PCR del DNA extraído con partidores universales de hongos, el protocolo utilizado se detalla en el Anexo 6.3.

#### 2.2.5 Amplificación de fragmentos por PCR

Para la amplificación de fragmentos se utilizó el DNA genómico extraído desde cada cepa de hongo como templado, en un mix de reacción con la siguiente composición: 1 µM de cada primer (*forward* y *reverse*), 0,2 mM dNTPs, 1,5 mM MgCl<sub>2</sub>, buffer Taq 1X, 1 U de Taq

polimerasa y 1 µl de DNA genómico en un volumen de reacción de 100 µl. El programa de PCR utilizado 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 fragmentos generados por este método fueron visualizados en un gel de agarosa al 1,5%, a partir del cual fueron purificados, clonados en *E. coli* DH5α y el DNA plasmidial extraído. El detalle de estos protocolos estándar se entrega en el Anexo 6.3.

#### 2.2.6 Secuenciamiento y análisis

Los fragmentos clonados en pGEM®-T Easy fueron enviados a secuenciar a Macrogen S.A. (Corea). El resultado entregado por Macrogen es analizado *in silico* utilizando la herramienta BLASTX [16]. Además, se utilizó la herramienta de búsqueda de dominios conservados, CDD [15], para analizar los alineamientos encontrados. Los resultados de estos análisis (discutidos en detalle más adelante) permitieron seleccionar secuencias para continuar trabajando.

#### 2.2.7 Genome walking

La estrategia utilizada para realizar *genome walking* de las secuencias seleccionadas fue la técnica desarrollada en el Centro de Ingeniería Bioquímica y Biotecnología (CIByB, Facultad de Ciencias Físicas y Matemática, Universidad de Chile), la cual permite encontrar las secuencias río arriba y río abajo de un fragmento conocido del genoma utilizando el DNA genómico como templado [24]. El método se puede resumir en 6 etapas principales:

1. Diseño de 2 partidores específicos al fragmento de DNA conocido.
2. Construcción de un adaptador a partir de oligonucleótidos genéricos del método.
3. Digestión del DNA genómico para producir genotecas.
4. Construcción de genotecas oligo-adaptador.
5. Primera ronda de PCR utilizando un único partidor específico, para elongación lineal.
6. Segunda ronda de PCR utilizando un partidor genérico y un partidor específico, para amplificación exponencial de la secuencia elongada en el primer PCR.

Las etapas anteriores describen la forma de obtener la secuencia en una dirección. Estas etapas se repiten para obtener la secuencia en la dirección contraria. En la Figura 2.2 se muestra un esquema del método. A continuación se describe en más detalle el trabajo realizado.

### 2.2.7.1 Diseño de partidores

A partir de las secuencias seleccionadas se diseñan partidores dirigidos a las familias 16 y 31 de la clasificación de glicosil-hidrolasas [12]. Los partidores son diseñados considerando que tengan un Tm de 57°C, siguiendo la estrategia de Acevedo, para lo cual se utilizó la herramienta OligoCalc [20], la cual además permite verificar que los oligonucleótidos no formen estructuras de tipo *hairpin* ni dímeros. Para cada una de las secuencias seleccionadas se deben diseñar dos partidores para llevar a cabo el *genome walking* en una dirección, es decir, se diseñan 4 partidores dirigidos a cada una de las familias: dos para completar la secuencia hacia el extremo 5' y dos para completar hacia el extremo 3'. Los partidores diseñados se presentan en la Tabla 2.6.

**Tabla 2.6. Partidores diseñados para realizar *genome walking* de los fragmentos seleccionados.**

Nombre	Secuencia (5' a 3')	Familia	Dirección	Ronda PCR
GH16B	cat cca gaa tga agg cca gac cta	16	5'	Primera
GH31B	gt gat ttc cgt aga ggt tcg ttc	31	5'	Primera
GH16A2	tga tac cat tat cta atc cct ccc cat t	16	3'	Primera
GH31A2	cta cct cga aca caa ccc gca t	31	3'	Primera
GH16A	aat ggg gag gga tta gat aat ggt atc a	16	5'	Segunda
GH31A	atg cgg gtt gtg ttc gag gta g	31	5'	Segunda
GH16B2	tag gtc tgg cct tca ttc tgg atg	16	3'	Segunda
GH31B2	gaa cga acc tct acg gaa atc acc	31	3'	Segunda

### 2.2.7.2 Construcción del adaptador

Se utilizan dos oligonucleótidos no fosforilados de hebra simple: AdaptF (5'-CTA GGC CAC GCG TCG ACT AGT ACT AGC TT-3') y AdaptR (5'-AGC TAG TAC TAG TCG ACG CGT GGC CTA G-3'), para generar un cassette AdaptT de doble hebra con una timidina protuberante en el extremo 3'. Para esto, AdaptF y AdaptR se hibridan calentando una mezcla 10 µM de los oligonucleótidos en un baño a 100°C utilizando un mechero y enfriando lentamente hasta temperatura ambiente.

### 2.2.7.3 Digestión del DNA genómico

Para las muestras en estudio se realizó una digestión del DNA genómico utilizando las enzimas de restricción EcoRI, EcoRV y PuvII. Para esto, se utilizaron 50 U de EcoRI, EcoRV o PuvII, buffer 1X *ad hoc* a la enzima y DNA genómico para completar un volumen de reacción igual a 100 µl. La digestión se llevó a cabo a 37°C durante 4 horas. En la Tabla 2.4 se presentan los sitios de restricción de cada una de las enzimas utilizadas.

Los fragmentos generados son purificados utilizando el kit QIAEX II, siguiendo las instrucciones del fabricante.

Para completar los extremos 3' recesivos generados por EcoRV y para adenilar los extremos 3' de todos los fragmentos, como lo exige el método, se tratan 20 µl del DNA digerido y purificado con 5 U de Taq DNA polimerasa, 0,2 mM de cada uno de los dNTPs, buffer Taq 1X y 1,5 mM de MgCl<sub>2</sub> en un volumen total de 50 µl por 45 min a 70°C.

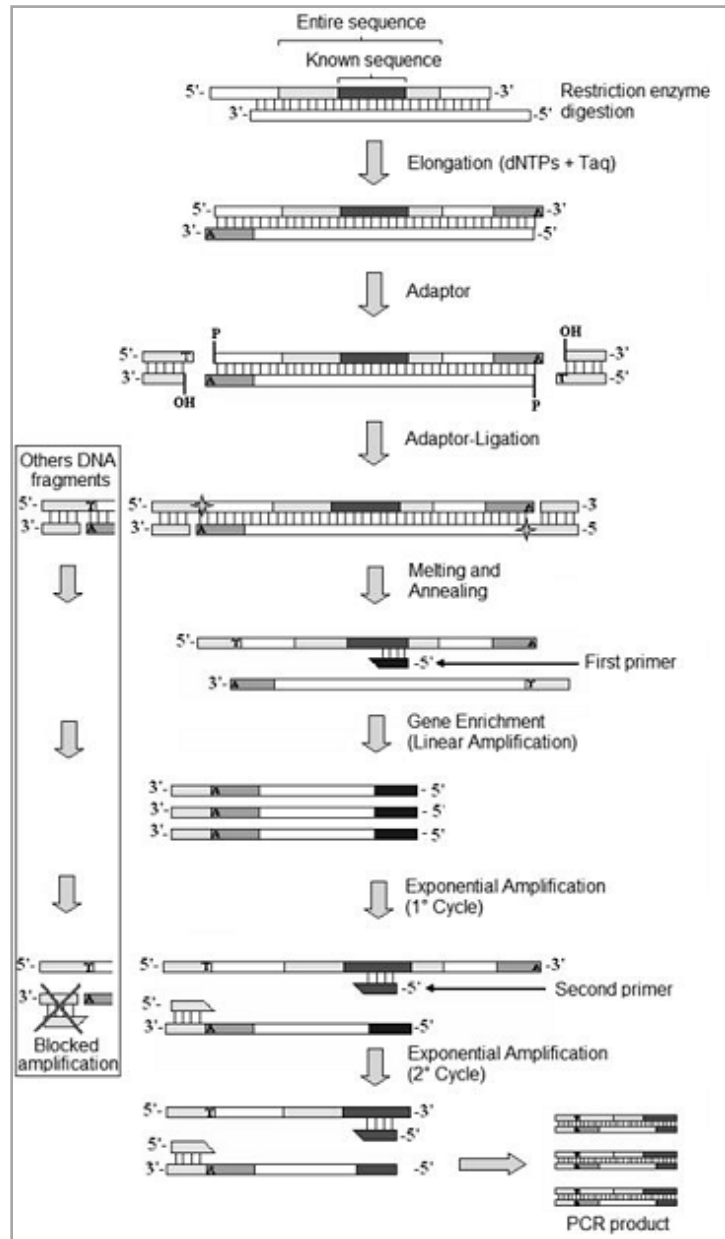
#### 2.2.7.4 Construcción de genotecas oligo-adaptador

Los fragmentos de restricción tienen una adenina protuberante en el extremo 3', la cual facilita la unión de éstos al adaptador AdaptT, el cual a su vez posee una timidina protuberante en el extremo 3'. La ligación de los oligonucleótidos a AdaptT se realiza en las siguientes condiciones: se incuban 7 µl de la solución de oligonucleótidos tratados con Taq polimerasa con 15 pmol de AdaptT, 1 U de DNA ligasa y buffer DNA ligasa 1X en un volumen total de 10 µl. La reacción de ligación se incubó a 16°C toda la noche.

#### 2.2.7.5 Primera ronda de PCR

La reacción de amplificación lineal fue realizada en un volumen total de 50 µl, con mix buffer elongasa 1X, 1,9 mM de MgCl<sub>2</sub>, 0,2 mM de dNTPs, 5 µl del constructo oligonucleótido-adaptador, 1 µl de mix Elongasa y 0,5 µM de partidor específico (GH16B y GH31B para *genome walking* río arriba, GH16A2 y GH31A2 para *genome walking* río abajo, referirse a la Tabla 2.6).

El programa utilizado en el PCR fue el siguiente: 1 ciclo de 94°C por 1 min, 20 ciclos de 94°C por 32 s, 55°C por 1 min y 72°C por 5 min y un ciclo final de 70°C por 7 min. El producto de PCR fue diluido 10 veces y se utilizaron 3 µl como templado para la segunda ronda de PCR.



**Figura 2.2. Técnica de *genome walking* utilizada.**

Se muestra la amplificación del extremo 5' de un gen a partir de una secuencia conocida con los partidores río arriba. Los pasos enmarcados muestran los resultados hipotéticos en caso de amplificación de secuencias inespecíficas.

#### 2.2.7.6 Segunda ronda de PCR.

La segunda ronda de PCR corresponde a una amplificación exponencial. Ésta fue llevada a cabo en un volumen total de 50  $\mu$ l, con 1X buffer elongasa 1X, 1,9 mM de  $MgCl_2$ , 0,2 mM de dNTPs, 3  $\mu$ l del producto del primer PCR diluido 10 veces, 1  $\mu$ l de mix Elongasa, 0,2  $\mu$ M del partidor específico al cassette AdaptF2 (5'-CAC GCG TCG ACT AGT ACT AGC TT-3') y 0,5  $\mu$ M del partidor específico (GH16A y GH31A para *genome walking* río arriba, GH16B2 y GH31B2

para *genome walking* río abajo, referirse a la Tabla 2.6). El programa utilizado en el PCR fue el siguiente: 1 ciclo de 94°C por 1 min, 20 ciclos de 94°C por 32 s, 58°C por 1 min y 72°C por 5 min y un ciclo final de 70°C por 7 min.

#### 2.2.8 Secuenciamiento de resultados de *genome walking*

Los fragmentos obtenidos mediante la técnica de *genome walking* fueron visualizados en geles de agarosa al 1,5%. A partir de esto, fueron seleccionados para su purificación, ligación en vector de clonamiento, transformados en *E. coli* DH5 $\alpha$  y el DNA plasmidial extraído para su secuenciamiento en Macrogen S.A. (Corea). Los protocolos utilizados se detallan en el Anexo 6.3.

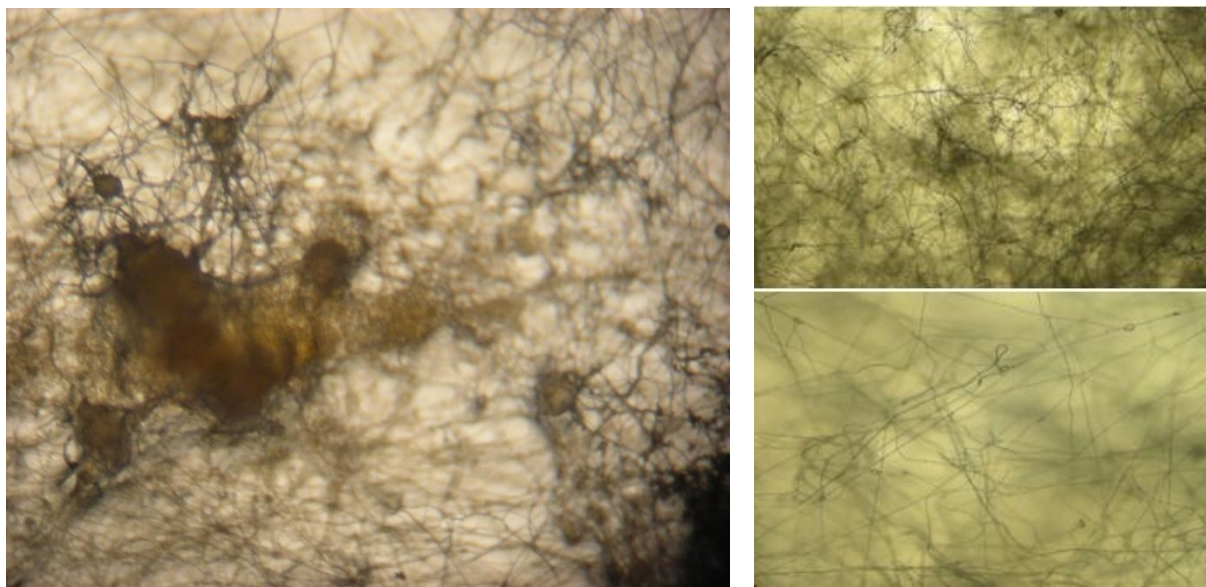


### 3. Resultados y discusión

#### 3.1 Construcción de micoteca

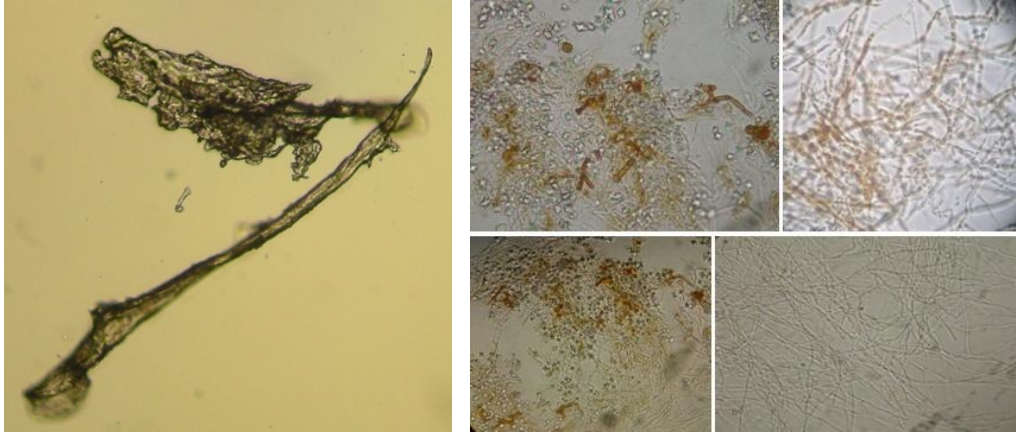
El cultivo de los hongos, a pesar de las precauciones de esterilidad tomadas, fue muy susceptible a contaminaciones. Esto obligó a descartar algunas cepas que no fue posible cultivar libre de contaminaciones. La elaboración de esta micoteca también permitió conocer el ritmo de crecimiento de los hongos, el cual varía ampliamente entre cada cepa. La mayoría demoró aproximadamente 5 días en cubrir la placa; sin embargo, algunas cepas demoraban 14 días o más. Esta observación permitió inferir que los resultados de los ensayos de actividad no serían fácilmente comparables entre cepas, debido a que a un mismo tiempo de crecimiento, cada una presenta una cantidad de biomasa diferente.

En las Figuras 3.1 y 3.2 se presentan algunas fotografías microscópicas de los cultivos obtenidos.



**Figura 3.1. *Alternaria alternata* y *Polysticus sulphureus*.**

Las fotos corresponden a imágenes de microscopía óptica de las placas de cultivo.



**Figura 3.2. Trametes cinnabarinus y Pignosporus cinnabarinus.**  
Las fotos corresponden a imágenes de microscopía óptica de las placas de cultivo.

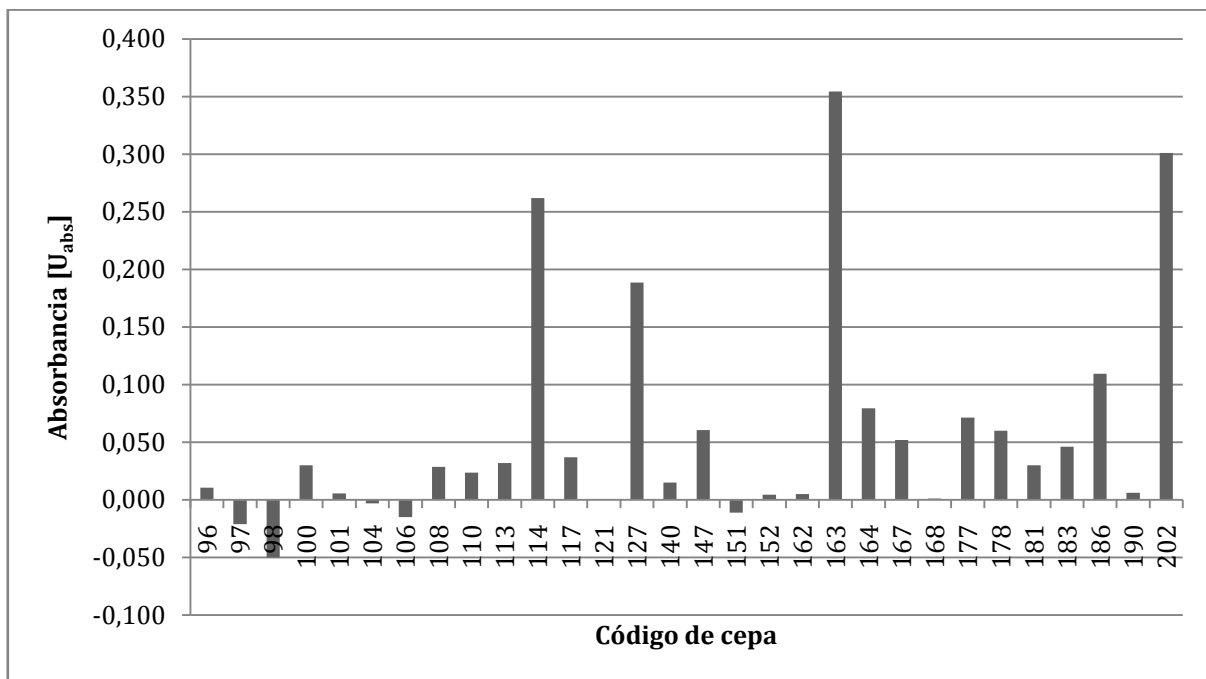
### 3.2 Identificación y selección de cepas con actividad exoglucanasa.

Los ensayos de actividad en medio sólido no fueron concluyentes, debido a que los halos que debían observarse eran poco claros. Inicialmente, este resultado se atribuyó a que el crecimiento en la placa había alcanzado un diámetro demasiado amplio, lo cual no permitía distinguir las zonas con y sin tinción. Sin embargo, al repetir los ensayos en placas con menor diámetro de crecimiento, se obtuvo nuevamente una tinción que no permitió discriminar las cepas que poseían actividad de las que no.

Por esta razón, se procedió a realizar ensayos de actividad en medio líquido. Como se mencionó anteriormente, las cepas tenían velocidades de crecimiento muy dispares, por lo que se optó por realizar ensayos de actividad en una serie de días para cada cepa, con el fin de construir una curva de actividad en función del tiempo y de este modo detectar el peak de actividad de cada cepa. Cabe mencionar que esta estrategia aún no es necesariamente óptima para comparar las actividades de las cepas entre sí, debido a que no permite comparar la actividad específica de las enzimas cuya actividad se detecta. Un hongo que produce mayor cantidad de una enzima con poca actividad arrojará un peak mayor que otro que produzca una enzima altamente activa pero que se produce en poca cantidad. Sin embargo, se optó por esta estrategia a pesar de estas desventajas, ya que se buscaba un método *high-throughput*, es decir, que permitiera el análisis de muchas cepas en un tiempo reducido.

A continuación se presenta el Gráfico 3.1 con los resultados obtenidos. Éstos corresponden a un valor de absorbancia a 520 nm para cada cepa, el cual es directamente proporcional a la

actividad de degradación de Avicel llevada a cabo por la cepa. Los ensayos se realizaron en los días 3, 4, 5 y 7 de crecimiento de los cultivos y, como se explicó anteriormente, los ensayos se realizaron en duplicado y se realizó un experimento “blanco” para normalizar los resultados (para detalles del ensayo y experimento blanco ver página 26). Los datos brutos se presentan en el Anexo 6.5.



**Gráfico 3.1. Máximos de actividad sobre Avicel.**

Los ensayos de actividad fueron realizados en los días 3, 4, 5 y 7, el valor graficado para cada cepa corresponde al máximo obtenido en esos cuatro ensayos. La identidad de cada hongo se presenta en la Tabla 6.1 (ver Anexo 6.1). Tras este análisis las cepas seleccionadas fueron la 114, 127, 147, 163 y 202.

Se observa que algunas barras muestran valores negativos. Esto ocurrió debido a la sustracción del valor obtenido del experimento blanco, el cual en algunos casos fue mayor que el valor obtenido en el ensayo de actividad de la cepa. Esto se atribuye a algunos artefactos inherentes al ensayo, principalmente que la insolubilidad del Avicel producía su decantación, con lo que podría haber interferido en la lectura de la absorbancia. Se tomaron todas las precauciones posibles para evitar esta fuente de error pero siempre existió una mínima presencia de material insoluble en la lectura que condujo a la obtención de estos errores. Sin embargo, se observa que para algunas cepas se logró detectar actividad muy por sobre el error del ensayo.

A partir del resultado anterior, se seleccionaron cinco cepas para continuar trabajando con ellas. Éstas fueron las cepas 114, 127, 147, 163 y 202. Las cepas 114, 127, 163 y 202 muestran

peaks significativamente mayores que el resto de las cepas, sin embargo, después aparecen una serie de muestras que tienen actividades similares. Se seleccionó la cepa 147 porque además de la evidencia experimental, se contaba con evidencia bibliográfica de que es una buena productora de enzimas celulolíticas [25; 26].

### 3.3 Análisis de regiones de homología para el diseño de partidores degenerados

Como se mencionó anteriormente, las glicosil hidrolasas se han clasificado en familias de acuerdo a similitudes en sus secuencias aminoacídicas. La actividad enzimática exoglucanasa corresponde al número E.C. 3.2.1.91; miembros de esta familia están distribuidos en cinco familias de las definidas por Henrissat: éstas son las familias 5, 6, 7, 9 y 48 [12; 13]. Los antecedentes preliminares de algunas de estas familias indican que están conformadas principalmente por endoglucanasas, o bien que su rango taxonómico sólo comprende bacterias, lo cual permitió descartarlas inicialmente de consideración para el diseño de los partidores. Esta información se presenta en la Tabla 3.1.

**Tabla 3.1. Datos de familias de glicosil hidrolasas que contienen exoglucanasas [13].**

Familia	Actividad	Rango taxonómico
5	Principalmente endoglucanasas	Eukarya, bacteria
6	Endoglucanasas y celobiohidrolasas	Eukarya, bacteria
7	Endoglucanasas y celobiohidrolasas	Eukarya
9	Principalmente endoglucanasas	Eukarya, bacteria
48	Endoglucanasas y celobiohidrolasas	Bacteria

La información anterior permitió seleccionar las familias 6 y 7 para que fueran el punto de partida en el diseño de partidores para amplificar regiones de exoglucanasas. Utilizando la base de datos de EMBL-EBI, se obtuvieron las secuencias aminoacídicas de todas las enzimas con actividad exoglucanasa (EC 3.2.1.91) que fueran de hongos y pertenecieran a las familias 6 y 7, con lo cual se obtuvo dos grupos de secuencias, uno para cada familia. Una observación de los alineamientos múltiples de estas secuencias con la herramienta ClustalX permitió identificar la existencia de bloques de homología dentro de cada una de las familias (ver Figura 3.4).

Para el diseño de partidores, se utilizó la estrategia CODEHOP [19], como fue detallado anteriormente. Esta herramienta entregó como resultado una serie de opciones de partidores *forward* y *reverse*. Entre los resultados obtenidos, se seleccionaron parejas de partidores con

Tm similares para que fueran compatibles en una reacción de PCR. Los partidores diseñados fueron presentados en la Tabla 2.5.

En la Figura 3.4 se muestra el alineamiento de las secuencias aminoacídicas de las exoglucanasas seleccionadas con los partidores diseñados a partir de ellas. Se presenta sólo el alineamiento para la familia 7, a modo de ejemplo. El alineamiento para la familia 6 se presenta en el Anexo 6.6.

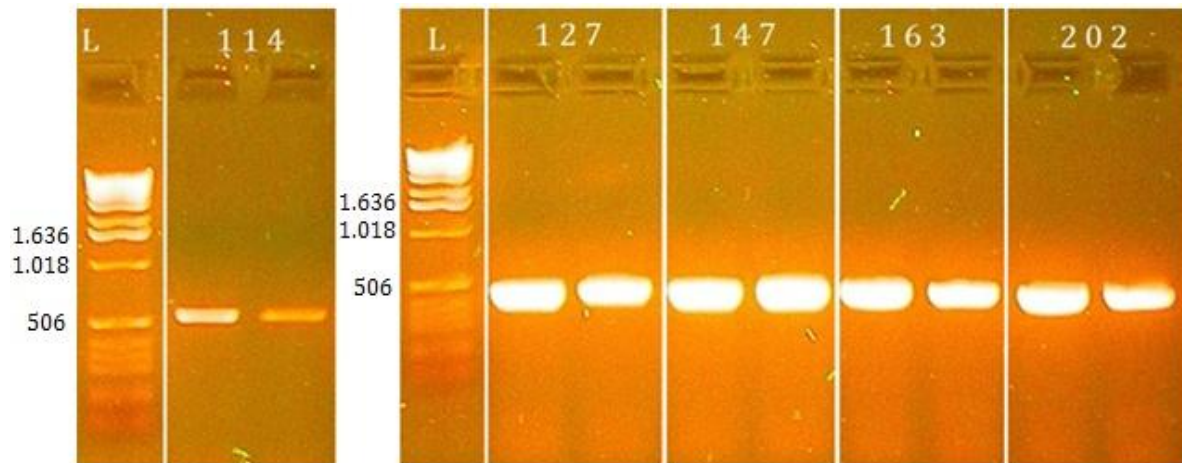
### **3.4 Extracción de DNA genómico desde hongos**

Para realizar la extracción de DNA genómico fue necesario investigar para seleccionar un método de extracción, ya que no se contaba con esta información previamente. Se probaron tres métodos:

1. Método que utiliza proteinasa K y nitrógeno líquido, facilitado por el Laboratorio de Fitopatología del Servicio Agrícola y Ganadero, Chillán. Demora aproximadamente 1 h.
2. Método que utiliza acetato de potasio, alcohol isopropílico y etanol [27]. Demora aproximadamente 3 h.
3. Método que utiliza el reactivo comercial TRI®, etanol y citrato de sodio [28]. Demora aproximadamente 3 h.

Todos los métodos son efectivos para la extracción de DNA genómico para usar como templado para PCR, lo cual fue verificado realizando una reacción de PCR con partidores universales para el gen 18S (ver secuencia en el Anexo 6.3), altamente conservado en hongos [29]. Habiendo probado los tres métodos, se seleccionó el primero, debido principalmente a la rapidez y facilidad con que puede realizarse, lo que permite la extracción de muchas muestras simultáneamente. El resultado del PCR de verificación realizado se muestra en la Figura 3.3.

Los protocolos de los otros dos métodos probados y del PCR de verificación se presentan en el Anexo 6.3.



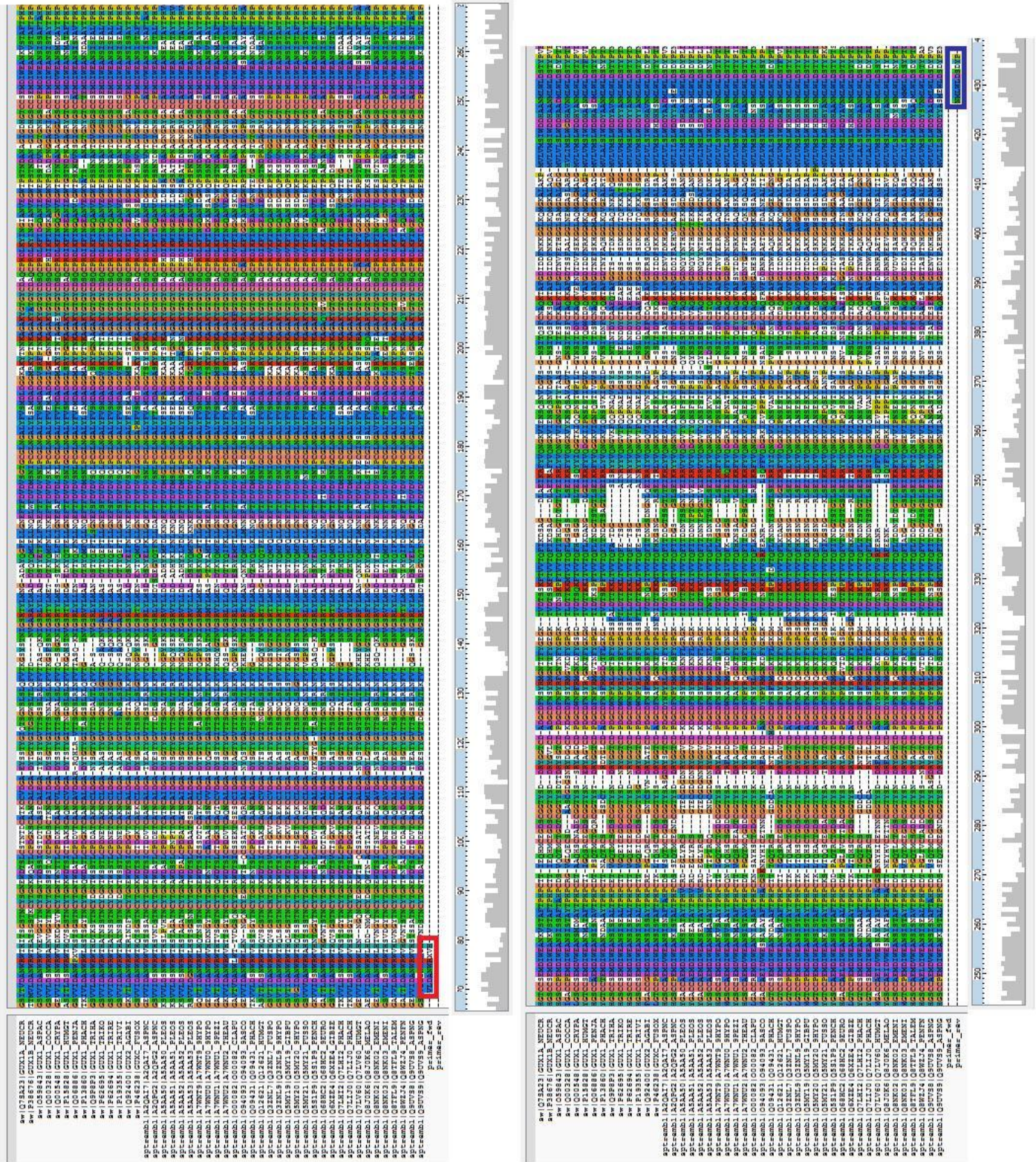
**Figura 3.3. PCR de verificación extracción de DNA genómico.**

Se realizó una electroforesis en gel de agarosa al 1,5% con bromuro de etidio. La columna marcada "L" corresponde a un marcador de peso molecular de 1 kb, el amplicón obtenido es de aproximadamente 500 pb. Las extracciones fueron realizadas en duplicado por lo que hay dos carriles por cada muestra.

### 3.5 Amplificación de fragmentos por PCR

El programa de PCR utilizado para la amplificación de fragmentos con los partidores degenerados requirió de una estrategia en la que se realizó un PCR con dos ciclos de denaturación, alineamiento y extensión, utilizando temperaturas distintas de alineamiento en cada ciclo. El segundo ciclo proveía una temperatura más alta a la etapa de hibridación, a lo que se suma que se utilizó un gradiente de temperaturas. Esta estrategia persigue dos objetivos principales:

1. Se debe tener presente que los partidores fueron diseñados a partir de bloques de homología por lo cual éstos no son perfectamente complementarios al DNA templado. Es por esto que utilizar una temperatura más baja en los primeros ciclos permite la hibridación inespecífica de los partidores, lo cual permite aumentar la población de DNA templado en un comienzo. Al aumentar la temperatura en el segundo ciclo, se restringe la hibridación por lo cual se promueve la amplificación de los fragmentos que sean más similares a los bloques de homología que sirvieron para diseñar los partidores.
2. La utilización de un gradiente de temperatura para la hibridación en el segundo ciclo permitió explorar una serie de condiciones en un solo experimento. Esto fue especialmente relevante debido a que se estaban usando distintos DNA templado en un mismo experimento (se realizó un PCR para las cinco muestras seleccionadas en forma simultánea).



**Figura 3.4. Alineamiento de exoglucanasas de la familia 7 con partidores *exoF\_fam7\_fwd2* y *exoF\_fam7\_rev2*.**

Se muestran las secuencias de exoglucanasas de la familia 7 alineadas con los partidores en el programa ClustalX. El alineamiento se presenta en dos secciones para mayor comodidad. El rectángulo rojo enmarca el partidor *exoF\_fam7\_fwd2* y el rectángulo azul enmarca a *exoF\_fam7\_rev2*. Las columnas coloreadas indican conservación del residuo aminoacídico en esa posición. Debajo de cada alineamiento se muestra una regla que indica la posición de los residuos.

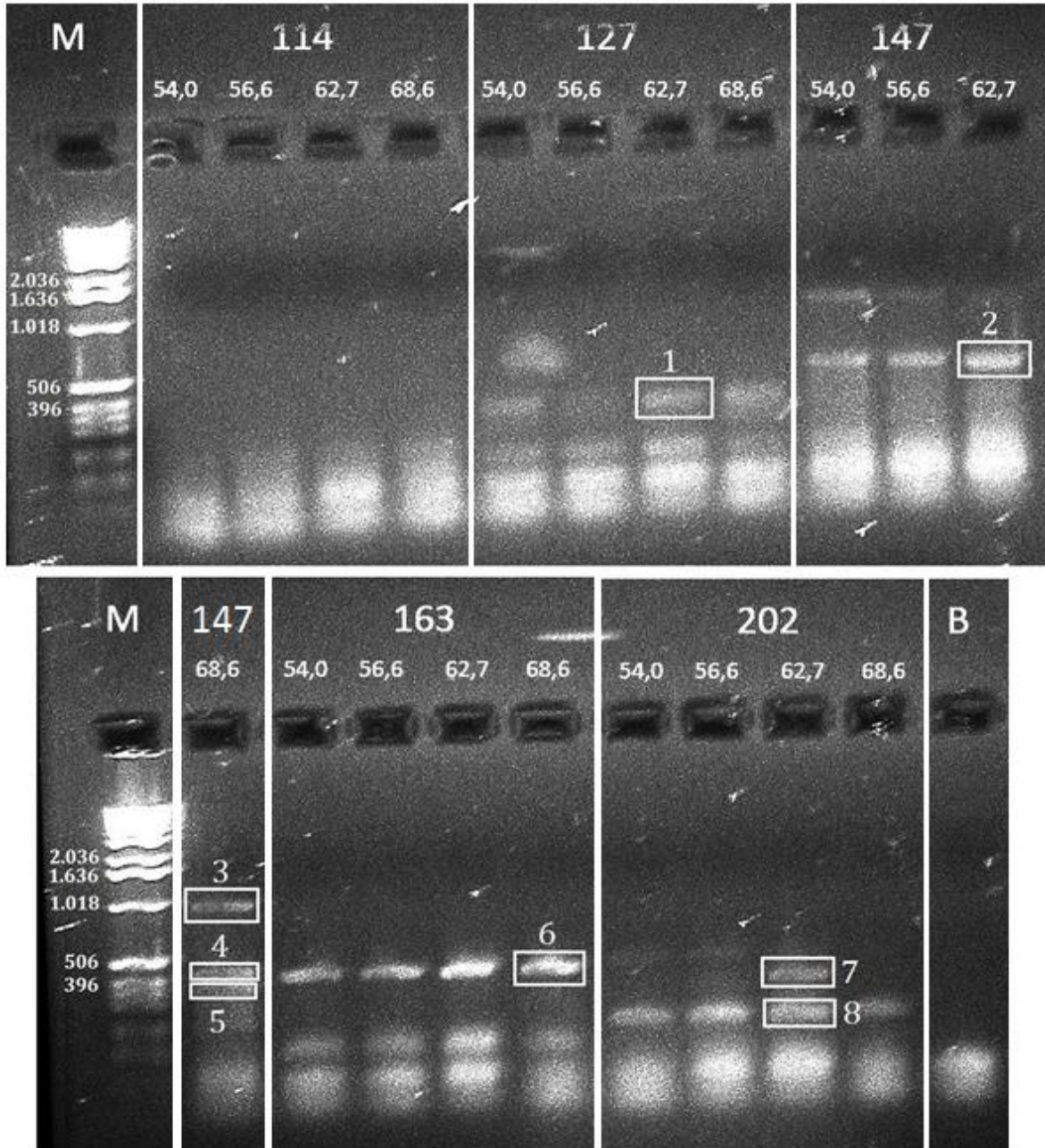
Se realizó una electroforesis con los productos de PCR obtenidos utilizando la estrategia anterior, en el cual se observó una población de fragmentos amplificados. En las Figuras 3.5 y 3.6 se muestra una fotografía con el resultado. Se debe tener presente que estas bandas pueden corresponder efectivamente a fragmentos de enzimas relacionadas con aquéllas que se utilizaron como base para diseñar los partidores degenerados, sin embargo, también puede ocurrir que existan hibridaciones inespecíficas de los partidores con el DNA genómico templado, y que se produzcan bandas con cierto grado de homología o bien que no presenten homología con las proteínas de las familias 6 y 7.

En la Tabla 3.2 se presenta un resumen de las bandas seleccionadas para purificar, junto con una breve caracterización de éstas.

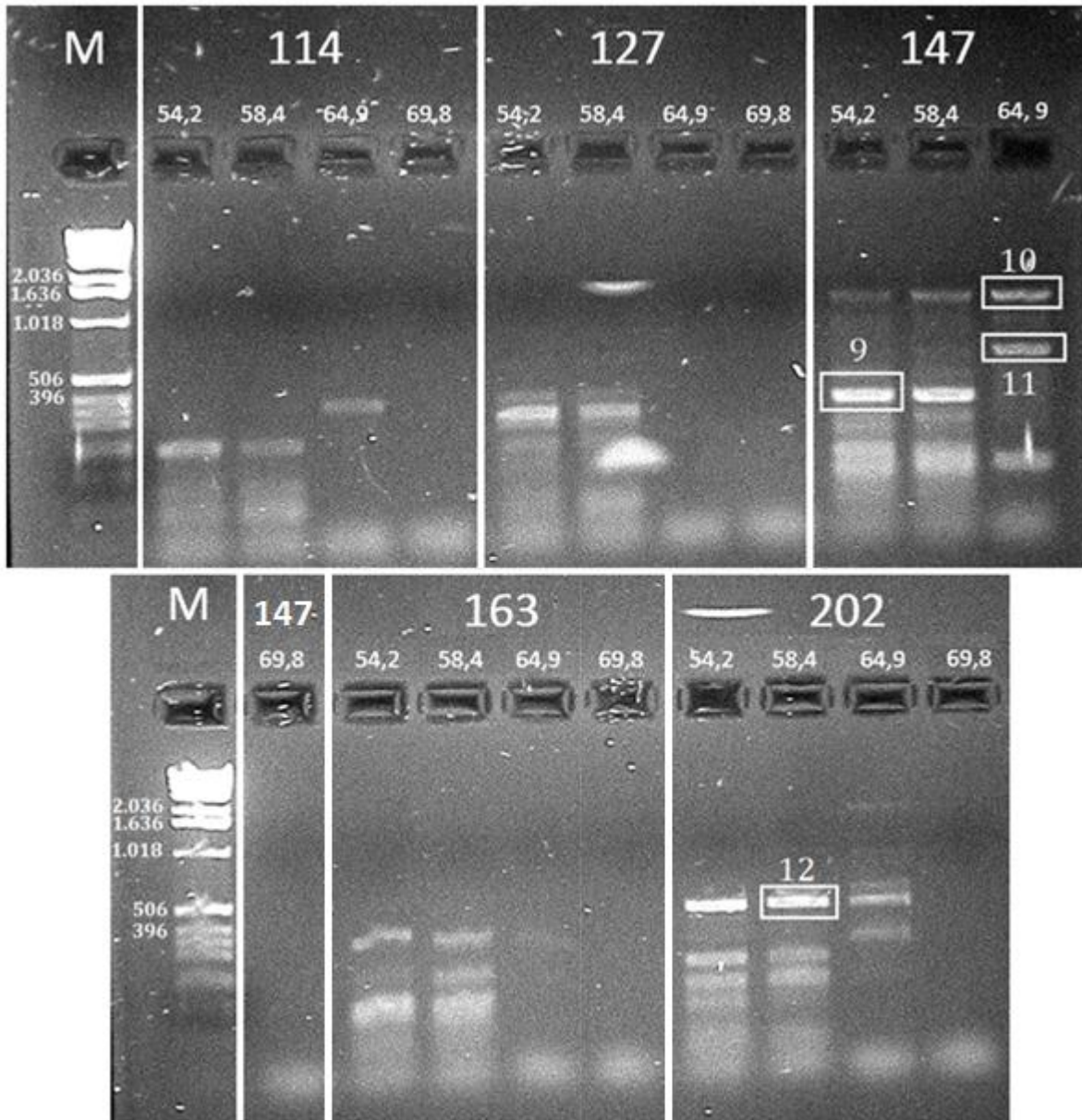
**Tabla 3.2. Bandas seleccionadas para purificación.**

Muestra	Partidores utilizados	Largo aproximado [pb]	Cepa
1	Fam. 6	400	127
2	Fam. 6	700	147
3	Fam. 6	1.000	147
4	Fam. 6	400	147
5	Fam. 6	350	147
6	Fam. 6	450	163
7	Fam. 6	400	202
8	Fam. 6	200	202
9	Fam. 7	400	147
10	Fam. 7	1.600	147
11	Fam. 7	800	147
12	Fam. 7	500	202





**Figura 3.5. Amplificación de DNA genómico de hongos con partidores dirigidos a la familia 6.** Se realizó electroforesis en gel de agarosa al 1,5% con bromuro de etidio. Las columnas marcadas "M" corresponden a marcador de peso molecular de 1 kb. Se realizaron cuatro experimentos de PCR para cada muestra, que corresponden a las cuatro columnas bajo cada muestra y además en cada columna se indica la temperatura de hibridación utilizada en el segundo ciclo. Las bandas encuadradas corresponden a las bandas seleccionadas para purificación, las cuales se marcaron con un número para identificarlas fácilmente. La columna marcada "B" corresponde al experimento control negativo, la mancha de bajo peso molecular que se observa corresponde a los partidores.



**Figura 3.6. Amplificación de DNA genómico de hongos con partidores dirigidos a la familia 7.**  
 Para detalles de la simbología de esta figura referirse a la figura anterior.

### 3.6 Secuenciamiento y análisis

Cada una de las bandas seleccionadas dio origen a cuatro clones (identificados con letras A a la D a continuación del número de la banda) que poseen el fragmento inserto en el vector de clonación pGem®-T Easy. El DNA plasmidial de estos clones fue extraído y sometido a digestión para verificar la presencia del fragmento, lo cual permitió identificar clones en los cuales falló la inserción del fragmento y éstos no fueron secuenciados. Para más detalles de este protocolo referirse al Anexo 6.3.

Las secuencias obtenidas fueron analizadas utilizando las herramientas BLASTX [16] y CDD [15]. La herramienta BLASTX permite comparar secuencias nucleotídicas con bases de datos de secuencias aminoacídicas. Estos resultados generan un valor de e-value, el cual es un parámetro que indica el número de hits que cabe esperar obtener por azar cuando se busca en una base de datos de un tamaño dado. Esencialmente, el e-value describe el ruido de fondo de la búsqueda, por lo tanto, mientras más bajo el valor del e-value (mientras más cercano a cero), el match es más significativo [30]. Además, el resultado de BLASTX señala un marco de lectura donde ocurre el alineamiento, el cual es un punto de partida para realizar la búsqueda de dominios conservados con la herramienta CDD, que se realiza a partir de secuencias aminoacídicas. La herramienta CDD corresponde a una base de datos de dominios conservados de proteínas disponible para consulta en línea (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>). La consulta se realiza a partir de una secuencia aminoacídica y se obtiene como resultado el grado de similitud que posee la secuencia con los dominios conservados almacenados en la base de datos, si es que la hay. Para mayor detalle sobre dominios conservados proteicos referirse a la Introducción de este trabajo (página 19).

Para esto, el primer paso fue “limpiar” las secuencias, esto es, buscar las secuencias del vector que flanquean el inserto y eliminarlas de modo de no considerarlas en el análisis. A continuación se presentan los resultados obtenidos de estos análisis para las muestras que fueron efectivamente secuenciadas. El texto de las secuencias se presenta en el Anexo 6.6.

- Secuencias 127Ex1C y 127Ex1D

Ambas secuencias del fragmento 1 presentan alta similitud con  $\beta$ -galactosidasa de vectores de clonación. Esto unido al hecho de que no se encontraron los flancos del vector pGemT®-Easy en las secuencias lleva a concluir que los fragmentos no se insertaron correctamente.

- Secuencia 147Ex3A

El análisis con BLASTX indica que la secuencia tiene similitud con glicosil hidrolasas de la familia 16, en los marcos de lectura +1 y +2. De los alineamientos obtenidos se puede decir con alta confianza (e-value del orden de  $10^{-10}$ ) que la enzima pertenece a la familia 16 de glicosil hidrolasas. Los alineamientos que siguen, con confianza del orden de  $10^{-5}$ , indican que la secuencia presenta similitud con glucanasas tipo MLG (*mixed-linked glucans*, glucanos de enlace mixto), específicamente con endo- $\beta$ -1,3(4)-glucanasa.

La familia 16 incluye varios tipos de glicosil hidrolasas, entre ellas las de tipo MLG (ver Figura 3.7). Las enzimas tipo MLG (o MLGAs) son aquéllas que pueden degradar glucanos de enlaces mixtos, también llamadas  $\beta$ -1,3-1,4-glucanasas,  $\beta$ -glucanasas, o liquenasas. Algunas MLGAs pueden degradar también otros glucanos como  $\beta$ -1,3-glucanos y  $\beta$ -1,4-glucanos [31].

Haciendo el análisis con la herramienta CDD, se observa que el marco +2 no muestra dominios conservados. El marco +1, por el contrario, tiene similitud con la superfamilia de glicosil hidrolasas 16, ver Figura 3.8. En ésta se observa la región de similitud con la superfamilia, sin embargo, no se encuentra presente la región completa (indicado por los bordes irregulares de la barra que representa la secuencia de la superfamilia). El resultado indica que la secuencia 147Ex3A presenta tres de los tres residuos catalíticos asociados a la superfamilia, indicado por los triángulos (marcados como *catalytic site*).

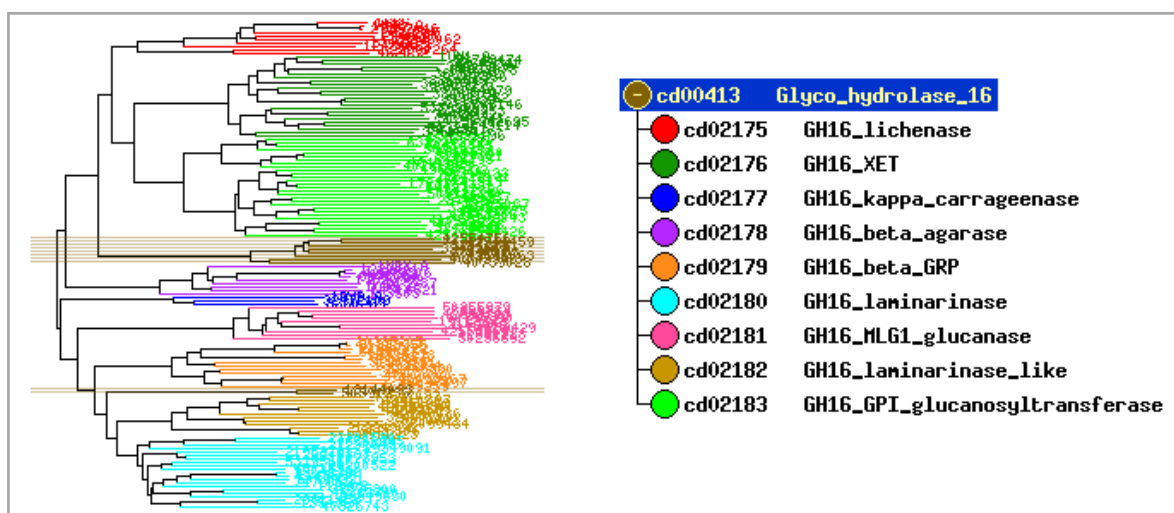


Figura 3.7. Clusters dentro de la superfamilia 16 [31].



Figura 3.8. Resultado de análisis de dominios conservados realizado usando la herramienta CDD [15] para la secuencia 147Ex3A en el marco de lectura +1.

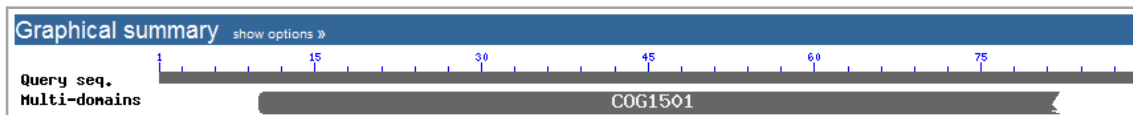
La banda gris representa la secuencia 147Ex3A (*query*). La cinta roja indica la posición donde se encontró similitud con el dominio conservado de GHs de la familia 16. Los triángulos rojos indican la posición de los residuos catalíticos en la secuencia *query*. La posición se refiere a residuos aminoácidos.

- Secuencia 147Ex3B

El análisis con BLASTX indica alta similitud con  $\alpha$ -glucosidasas, pertenecientes a la familia 31 de GHs, con e-values del orden de  $10^{-25}$ , en el marco +1.

La familia 31 comprende enzimas con varias actividades conocidas, entre ellas  $\alpha$ -glucosidasas,  $\alpha$ -galactosidasas, glucoamilasas, sucroisomaltasas,  $\alpha$ -xilosidasas y  $\alpha$ -glucanoliasas. Existe evidencia de que un residuo de ácido aspártico se encuentra involucrado la actividad catalítica de sucrasa, isomaltasa y  $\alpha$ -glucosidasa [13].

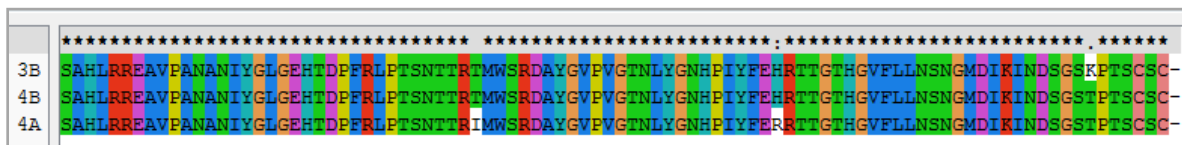
En cuanto a dominios conservados, presenta similitud con el modelo COG 1501, con e-value del orden de  $10^{-4}$ , que consiste en  $\alpha$ -glucosidasas de la familia 31 (ver Figura 3.9). No se encuentran presentes los residuos del dominio catalítico.



**Figura 3.9. Resultado de análisis de dominios conservados realizado usando la herramienta CDD [15] para la secuencia 147Ex3B en el marco de lectura +1.**

- Secuencias 147Ex4A y 147Ex4B

La secuencia 147Ex4A traducida en el marco de lectura +1 y la secuencia 147Ex4B traducida en el marco -1 son idénticas a la secuencia 147Ex3B, lo cual se puede observar a simple vista en un alineamiento de las secuencias traducidas utilizando la herramienta ClustalX (ver Figura 3.10).



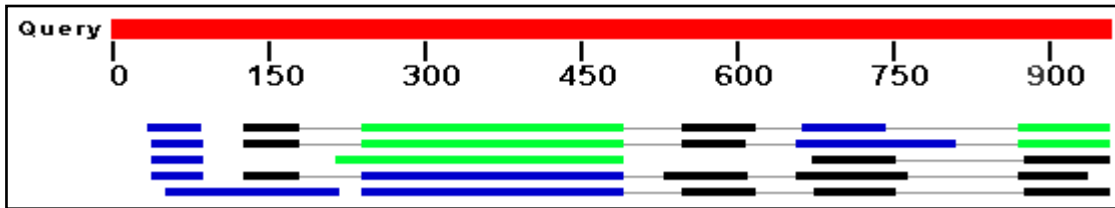
**Figura 3.10. Alineamiento de las secuencias 147Ex3B, 147Ex4A y 147Ex4B.**

Resultado obtenido usando el programa ClustalX [18]. Los asteriscos en la fila superior indican identidad en las tres secuencias en esa posición.

- Secuencia 147Ex5A y 147Ex5C

Estas secuencias son idénticas entre sí, confirmado por alineamiento con ClustalX, en forma similar al caso anterior.

El análisis con BLASTX presenta similitud con transportador de urea, con e-value del orden de  $10^{-10}$ , en los marcos de lectura +1, +2 y +3, sin embargo, no se encuentra la presencia de dominios conservados al utilizar la herramienta CDD.



**Figura 3.11. Resultado del análisis con BLASTX para la secuencia 147Ex5A.**

Se presentan los resultados más significativos. La banda roja representa la secuencia 147Ex5A, la posición es relativa a nucleótidos. Las bandas inferiores representan los resultados más significativos obtenidos (e-value menor a  $10^{-10}$ ). Los colores representan la significancia del alineamiento, siendo verde el más similar, luego azul y finalmente negro. Las líneas negras que unen las bandas coloreadas indican regiones que no alinean con la secuencia consultada.

Particularmente, la secuencia alinea con proteínas de la familia SSSF (*sodium solute symporter family*), la cual contiene cientos de miembros de origen tanto procarionte como eucarionte. Las proteínas de esta familia participan en el mecanismo de transporte de solutos a través de la membrana citoplasmática, el cual es usado por la célula para dirigir la acumulación de solutos en contra del gradiente de concentración. Las proteínas de la familia SSSF poseen entre 11 y 15 dominios transmembrana en conformación  $\alpha$ -hélice [32; 33].

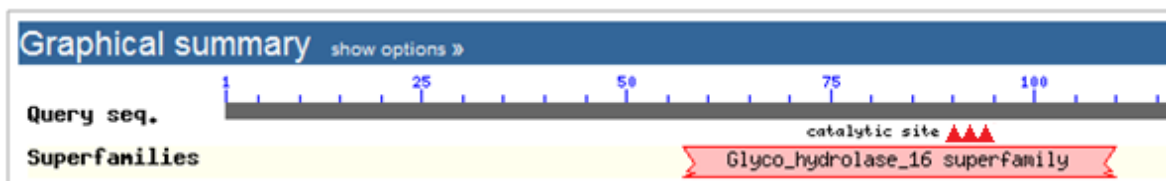
- Secuencia 147Ex5D

Presenta alta similitud con  $\beta$ -galactosidasa de vectores de clonación. Esto unido al hecho de que no se encontraron los flancos del vector pGemT-Easy lleva a concluir que el fragmento no se insertó correctamente.

- Secuencia 163Ex6A

El análisis con BLASTX indica similitud con GHs de la familia 16 en los marcos de lectura -1 y -2, con e-values del orden de  $10^{-10}$ . También presenta similitud con enzimas que poseen el dominio del tipo MLG ( $\beta$ -1,3-1,4-glucanasas), con e-values del orden de  $10^{-5}$ .

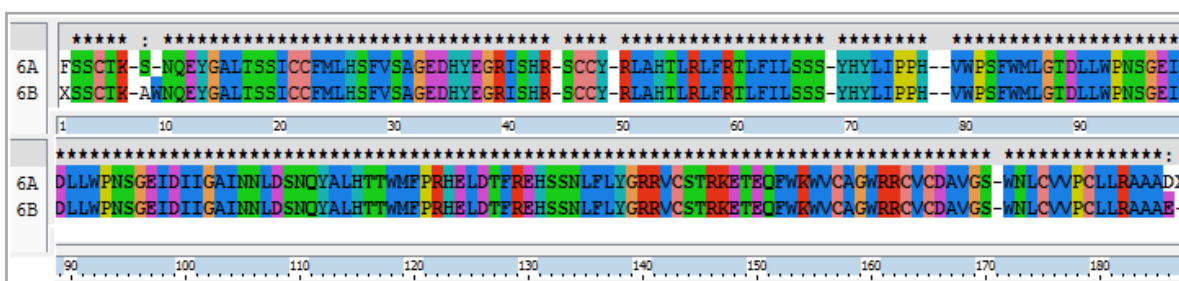
En el análisis CDD, se encontró que en el marco de lectura -2 no se presenta ningún dominio conservado. En el marco de lectura -1 se encuentra un alineamiento con la superfamilia de glicosil hidrolasas 16, en la Figura 3.12 se presenta este resultado. El análisis con CDD indicó que la secuencia 163Ex6A presenta los tres residuos del sitio catalítico que caracteriza a esta superfamilia, éstos se encuentran indicados en la Figura 3.12.



**Figura 3.12. Resultado de análisis de dominios conservados realizado usando la herramienta CDD [15] para la secuencia 163Ex6A traducida en el marco -2.**  
Para detalles de la simbología de esta figura referirse a la Figura 3.8.

- Secuencia 163Ex6B

Esta secuencia, traducida en el marco de lectura +3 es idéntica a la secuencia 163Ex6A traducida en el marco -1, el alineamiento que lo confirma se presenta en la Figura 3.13.



**Figura 3.13. Alineamiento de las secuencias 163Ex6A y 163Ex6B.**

Resultado obtenido usando el programa ClustalX [18]. Los asteriscos en la fila superior indican identidad en las tres secuencias en esa posición.

- Secuencia 202Ex7B

La herramienta BLASTX no arroja ningún resultado significativo (todos los alineamientos presentan e-values mayores a 0,5).

- Secuencias 202Ex7C y 202Ex8B

En estas secuencias no se encuentran los flancos del vector de clonación. Por otro lado, BLASTX arroja alineamientos con el gen *lacZ* de vectores de clonación. En conjunto, estos resultados indican que podrían no haberse insertado los fragmentos correctamente.

En resumen, los resultados anteriores indican que se tiene lo siguiente:

**Tabla 3.3. Resumen de resultados parciales.**

Resultado	Secuencias
Secuencias que no poseen inserto	127Ex1C, 127Ex1D, 202Ex7C, 202Ex8B
Secuencias similares a $\beta$ -1,3(4)-endoglucanasa (Fam. 16)	147Ex3A, 163Ex6A, 163Ex6B
Secuencias similares a $\alpha$ -glucosidasa (Fam. 31)	147Ex3A, 147Ex4A, 147Ex4B
Secuencias sin alineamientos significativos	202Ex7B

Como se puede observar, las secuencias obtenidas no corresponden con las secuencias a las cuales se había dirigido el diseño de los partidores (exoglucanasas de las familias 6 y 7). Esto se explica porque la estrategia utilizada no es altamente específica. Los partidores diseñados son degenerados, esto significa que lo que se tiene es un pool de partidores con muchas secuencias diferentes, a los cuales además se les permiten condiciones relajadas de hibridación en los primeros ciclos del PCR, todo lo cual promueve amplificaciones inespecíficas que podrían favorecer obtener el resultado esperado pero que también deja espacio a resultados menos específicos como los obtenidos.

Como se introdujo en este trabajo, la clasificación de glicosil hidrolasas de Henrissat también introduce el concepto de clanes, los cuales se han definido en base a enzimas que pertenecen a familias diferentes pero que exhiben relación en cuanto a su secuencia aminoacídica. En particular, las familias 7 y 16 son los únicos miembros del clan B y si bien la familia 6 no se ha descrito como perteneciente a ningún clan, alineamientos preliminares realizados para diseñar los partidores CODEHOP revelaron que también existen bloques conservados entre la familia 6 y 7. Todo esto explica la aparición de familias diferentes a las esperadas.

### **3.7 Obtención de secuencias de regiones aledañas para secuencias seleccionadas**

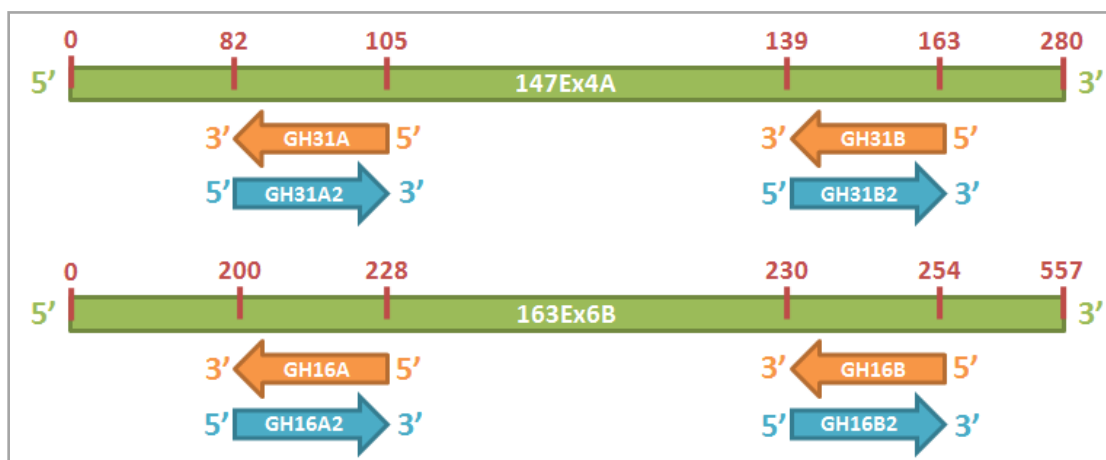
De acuerdo a los resultados obtenidos en el secuenciamiento de los fragmentos, se procedió a diseñar partidores para hacer *genome walking* de los fragmentos con similitud a las familias 16 y 31. Para esto, se utilizaron como templado las secuencias de los fragmentos 163Ex6B y 147Ex4A, para diseñar partidores dirigidos a las familias 16 y 31, respectivamente. Como se mencionara anteriormente, se diseñaron cuatro partidores para cada familia, como lo requiere el método utilizado. La posición de estos partidores en las respectivas secuencias se presenta en la Figura 3.14.

#### *3.7.1 Genome walking*

El método se aplicó a las cinco cepas seleccionadas en la primera etapa excepto la cepa 116, para la cual no se amplificó ninguna banda en los PCR con partidores degenerados, por lo que se trabajó con las cepas 127, 147, 163 y 202. La digestión del DNA genómico de las cepas analizadas se realizó con las enzimas EcoRI, EcoRV y PuvII. Todas estas muestras (12 en total) pasaron a la etapa de PCR, donde se probaron los sets de partidores de las familias 16 y 31 para todas ellas. La nomenclatura adoptada fue n° de muestra, enzima y partidores usados (p.



ej. la muestra 147EV31 corresponde al DNA genómico de la cepa 147 digerido con EcoRV y PCR con los partidores dirigidos a la familia 31).



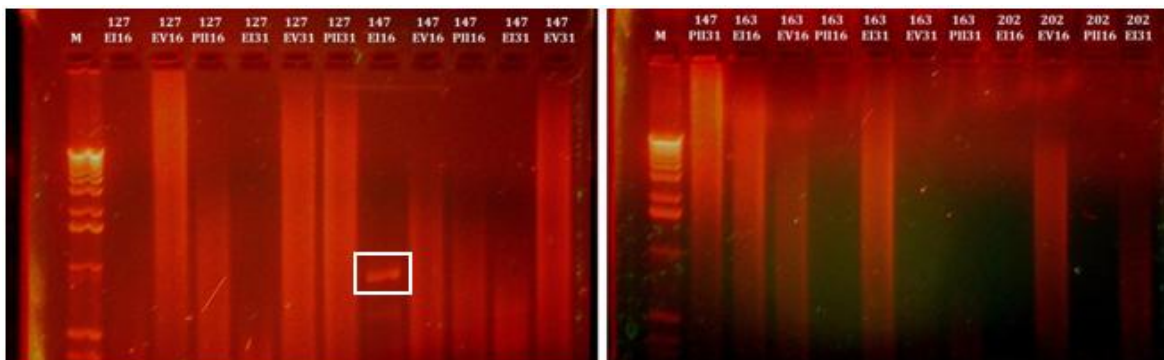
**Figura 3.14. Esquema de los partidores diseñados para la técnica de genome walking.**

Las secuencias utilizadas como base para el diseño se muestran en verde (147Ex4A para la familia 31 y 163Ex6B para la familia 16). En rojo se indica la posición donde alinean los partidores relativa a la secuencia templado.

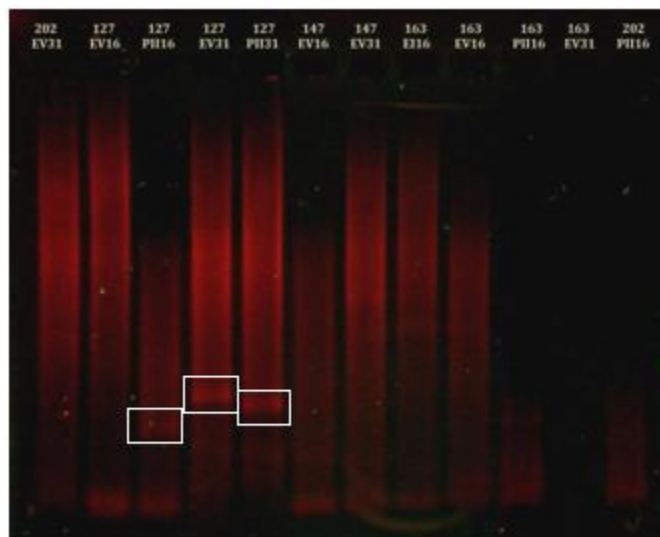
Primero, se realizaron las rondas de PCR para hacer el *genome walking* hacia el extremo 5' de los fragmentos, cuyo resultado se muestra en la Figura 3.15. Como se puede observar, sólo para la muestra 147EI16 se obtuvo una banda clara que pudo ser purificada adecuadamente. Para las demás muestras se observa un chorreo, lo que indica que ocurrió mucha amplificación inespecífica. Por esta razón, se realizó nuevamente la segunda ronda de PCR, aumentando la temperatura de la etapa de hibridación (de 55°C a 58°C) para así aumentar la especificidad. En la Figura 3.16 se muestra el resultado de este nuevo intento, el cual dio como resultado bandas claras para las muestras 127PII16, 127EV31 y 127PII31.

El *genome walking* hacia el extremo 3' se realizó utilizando inmediatamente una temperatura de hibridación de 58°C en la segunda ronda de PCR, debido a la experiencia del PCR hacia el extremo 5'. Sin embargo, en la electroforesis de la segunda ronda de PCR se observó un patrón de migración característico de amplificación inespecífica para todas las muestras. Como ya se había usado una temperatura alta en la segunda ronda, se repitió el primer PCR usando 58°C para la hibridación y se subió el de la segunda ronda de 58°C a 59°C. Esta estrategia tampoco dio los resultados esperados, por lo que finalmente se probó utilizar un gradiente de temperaturas en el segundo PCR, utilizando como templado el producto de la primera ronda a 59°C. La temperatura de hibridación usada fue de 60°C con un gradiente de 4°C. Este último intento tampoco dio los resultados esperados, presentando un chorreo en todos los casos en

la electroforesis. Por consideraciones de tiempo no se pudo seguir intentando mejorar las condiciones de los PCR y por lo tanto no se pudo realizar el *genome walking* hacia el extremo 3'.



**Figura 3.15. Segunda ronda de PCR de *genome walking* hacia el extremo 5'.**  
El fragmento obtenido se muestran enmarcados en un rectángulo blanco.



**Figura 3.16. Repetición segunda ronda de PCR de *genome walking* hacia el extremo 5'.**  
Los fragmentos obtenidos se muestran enmarcados en un rectángulo blanco.

### 3.7.2 Secuenciamiento y análisis

Las cuatro bandas purificadas (147EI16, 127PII16, 127EV31 y 127PII31) fueron procesadas según los protocolos explicados en el capítulo de Métodos. Nuevamente, cada muestra dio origen a cuatro clones que deben poseer el vector con el inserto (identificadas con letras A a la D a continuación del nombre de la muestra). Una digestión del DNA plasmidial de estos clones permitió identificar las mejores candidatas para secuenciar.

Al igual que con las primeras secuencias obtenidas, se realizó un análisis de éstas utilizando las herramientas BLASTX y CDD. A continuación se presentan los resultados obtenidos tras estos análisis.

- Secuencia 147EI16A y 147EI16C

Las secuencias son idénticas entre sí. La herramienta BLASTX arroja como resultado similitud con genes de transcriptasa reversa con un altísimo grado de confianza, e-values del orden de  $10^{-120}$ , en el marco de lectura +1.

La traducción de la secuencia fue luego sometida a análisis con la herramienta CDD, encontrándose el dominio conservado RT\_LTR con e-value igual a  $2 \times 10^{-47}$ .

El dominio RT\_LTR corresponde a transcriptasas reversas (RT) de retrotransposones y retrovirus que tienen repeticiones terminales largas (LTRs). Las RT catalizan la replicación del DNA desde un templado de RNA, y son responsables de la replicación de retroelementos. Los genes RT están presentes en una variedad de elementos móviles, incluyendo retrotransposones, retrovirus, intrones del grupo II, entre otros. [34].

- Secuencia 147EI16B

La herramienta BLASTX indica similitud con GHs de la familia 16, con e-values del orden de  $10^{-5}$ , en marcos de lectura +1 y +2.

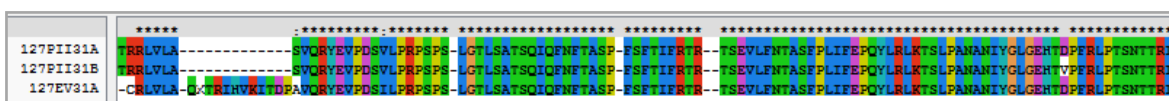
De acuerdo a la herramienta CDD, no se encuentran dominios conservados en ninguno de los marcos de lectura.

- Secuencia 127EV31A

La herramienta BLASTX indica alta similitud con  $\alpha$ -glucosidasa de hongos de la familia 31, con alto nivel de confianza (e-value del orden de  $10^{-30}$ ). Estos alineamientos corresponden al marco de lectura +1. Sin embargo, la herramienta CDD no encuentra dominios conservados en esta secuencia.

Esta secuencia es casi idéntica a las secuencias 127PII31A y 127PII31B, las cuales son idénticas entre sí y presentan un fragmento menos (12 aminoácidos) que la secuencia 127EV31A (ver alineamiento de las secuencias traducidas en Figura 3.17). Éstas presentan los

mismos alineamientos con e-value levemente mayor y tampoco se detectan dominios conservados en ellas.



**Figura 3.17. Alineamiento en ClustalX de secuencias 127PII31A, 127PII31B y 127EV31A.**  
Las secuencias son 127PII31A, 127PII31B y 127EV31A, de arriba hacia abajo.

- Secuencia 127EV31C

También presenta similitud con  $\alpha$ -glucosidasas de la familia 31, con un nivel de confianza bastante menor que las secuencias anteriores (e-value del orden de  $10^{-3}$ ). El alineamiento ocurre en el marco de lectura +3. La herramienta CDD no detecta dominios conservados en esta secuencia.

### 3.7.3 Reconstrucción de secuencias

La única secuencia que presentó similitud con GHs de la familia 16 fue la secuencia 147E116B. Se intentó encontrar el punto donde ésta se superpone con la secuencia 147Ex3A, de modo de reconstruir un fragmento más largo que corresponda a la unión de ambas secuencias. Las secuencias brutas son:

```
>147Ex3A
TTCAGCAGCTGCACGAAGTAGTCTTGGAACCAGGAATACGGTGCCTGACCTCATCCATTTGTTGTTTTATGCTCCACTCA
TTTGTCTTCTGCAGGTGAAGATCACTACGAAGGACGCATATCCATCGGTAGTCTTGCTGTTATTGACGTTTGGCACATACA
CTACGGTTGTTCCGTACGTTATTTATCCTATCTAGTTCATGATACCATTATCTAATCCCTCCCCATTAATAGGTCTGGCCT
TCATTCGGATGCTCGGCACCGACCTCCTCTGGCCAAATTCCGGCGAAATCGACATCATCGGAGCCATCAACAACCTCGAT
TCAAACCAATACGCCCTGCACACCCTGGATGTTTTCTAGGCAGAACTCGATACATTCAGGGAGCATTCTTCAAACCTTA
TTTCTCTACGGGCGCAGGTGTGTAGTACAAGAAAAGAAACCGAGTAGTTTTGGAAAGTGGGTTTGGCAGGCTGGCCGAGG
TGTGTTTGCAGCGCAGTTGGATCTTAGTGGAATCTATGTGTGTTCCAAGCCTACTTCGTGCAGCTGCTGAA
```

```
>147E116B
TCACGCGTCGACTAGTACTAGCTTACAGCTGCACACCTTGAGCGAGGGAATGGCAGCCACGGTTGAGGAAAGGCATCTTTC
GGAGGTTTATATGCGGCTAGATTTTCGTAGACACCTTGACGGAAGAAAGACGACTAAGAAGGGGGCAAATCCAGCAATTAAT
CGTTCGAGTCACTTTTTGATTCCCTTTGTGAGAGTTCCAAGATTCCGGTTTCTTTCCAGGTCATTAGTGGACAACCTGAGT
TTCCAGGTCCTCAATTCGCACATCTGATTGAGAATGATATCCTAGATAAATTCGCGAGCCAAAAGGTAAATTCGCCGG
TTGAGTTTGATTTTACTTTTGGAGGACATGGTAGCATGAGCTTCTAGAATATACTGTAGCAGTCGCAGAGGCCAGTCTGGT
GAGAGTGAGGATCGCTACAACGGTCATGGTTAGAAGCAGATTTCCGCAATTATTGGAAGCCGACAGCCGAGAATGATCACA
TGACGATAAGCAGCTCGGTGCTAGCCAGACTCTCATCAGTCTCAATCCCGGTTGAATTTACTCTCAACGACTCCTTCTC
AAGGCCCGTCTCTGCTGCCGCTCCATCATCTGACTTACATCGACCCAAACCAACCATGCGATAATCCGGGTCGATAATGC
GACAAATATAACGGCAGCTGCGACAGTGGAACCAGGAATACGGTGCCTGACCTCATCCATTTGTTGTTTTATGCTCCACT
CATTTGTTCTGTAGGTGAAGATCACTACGAAGGACGCATATCTCATCGGTAGTCTTGCTGTTATTGACGTTTGGCACATA
CACTACGGTTGTTCCGTACGTTATTTATCCTATCTAGTTCATGATACCATTATCTAATCCCTCCCCATCCCTCCCCATTA
```

En rojo se muestran las regiones comunes de ambos fragmentos, se observa que la secuencia 147E116B presenta las zonas en común hacia el extremo 3', por lo tanto uniendo ambas (147E116B hacia el extremo 5' y 147Ex3A hacia el extremo 5'), se obtiene una secuencia conjunta de 1.218 pb, a la que se ha llamado 147GH16 y que se muestra a continuación. La

unión de las secuencias se realizó de tal modo de conservar los marcos de lectura de las secuencias originales (marco +1 en el caso de ambas).

>147GH16

```
TCACGCGTCGACTAGTACTAGCTTACAGCTGCACACCTTGAGCGAGGGAATGGCAGCCACGGTTGAGGAAAGGCATCTTTC
GGAGTTTATATGCGGCTAGATTTTCGTAGACACCTTGACGGAAGAAAGACGACTAAGAAGGGGGCAAATCCAGCAATTAAT
CGTTTCGAGTCACCTTTTGGATTCCCTTGTGAGAGTTCCAAGATTCGCGTTTTCTTCCAGGTCATTAGTGACACAACCTGAGT
TTCCCAGGTCCTCAATTCGCACATCTGATTGAGAATGATATCCTAGATAAATTCGCGAGCCAAAAGGTAAATTCGCCGG
TTGAGTTGATTTTACTTTTGGAGGACATGGTAGCATGAGCTTCTAGAATATACTGTAGCAGTCGCAGAGGCCAGTCTGGT
GAGAGTGAGGATCGCTACAACGGTCAATGGTTAGAAGCAGATTTCCGCAATTATTGGAAGCCGACAGCCGAGAATGATCACA
TGACGATAAGCAGCTCGGTGCTAGCCCAGACTCTCATCAGTCTCAATCCCCGGTTGAATTTACTCTCAACGACTCCTTCTC
AAGGCCCGTCTGCTGCCGCTCCATCATCTGACTTACATCGACCCAAACACCAACCATGCGATAATCCGGGTCGATAATGC
GACAAATATAACGGCAGCTGCGACAGTGGAAACCAGGAATACGGTTCGCTGACCTCATCCATTTGTTGTTTATGCTCCACT
CATTGTTTCTGTAGGTGAAGATCACTACGAAGGACGCATATCTCATCGGTAGTCTTGCTGTTATTGACGTTTGGCACATA
CACTACGGTTGTTCCGTACGTTATTTATCCTATCTAGTTTCATGATACCATTATCTAATCCCTCCCCAATCCCTCCCCATTA
TAGGTCGCGCCTTCATTTCTGGATGCTCGGCACCGACCTCCTCTGGCCAAATTCGGCGAAATCGACATCATCGGAGCCATC
AACAACCTCGATTCAAACCAATACGCCCTGCACACCACCTGGATGTTTTCCTAGGCACGAACTCGATACATTAGGGAGCAT
TCTTCAAACCTTATTTCTCTACGGGCGCAGGGTGTGTAGTACAAGAAAAGAAACCGAGTAGTTTTGGAAAGTGGGTTTGGCGA
GGCTGGCGGAGGTGTGTTTGGCAGCGAGTTGGATCTTAGTGGAATCTATGTGTGGTTCCAAGCCTACTTCGTGCAGCTGCT
GAA
```

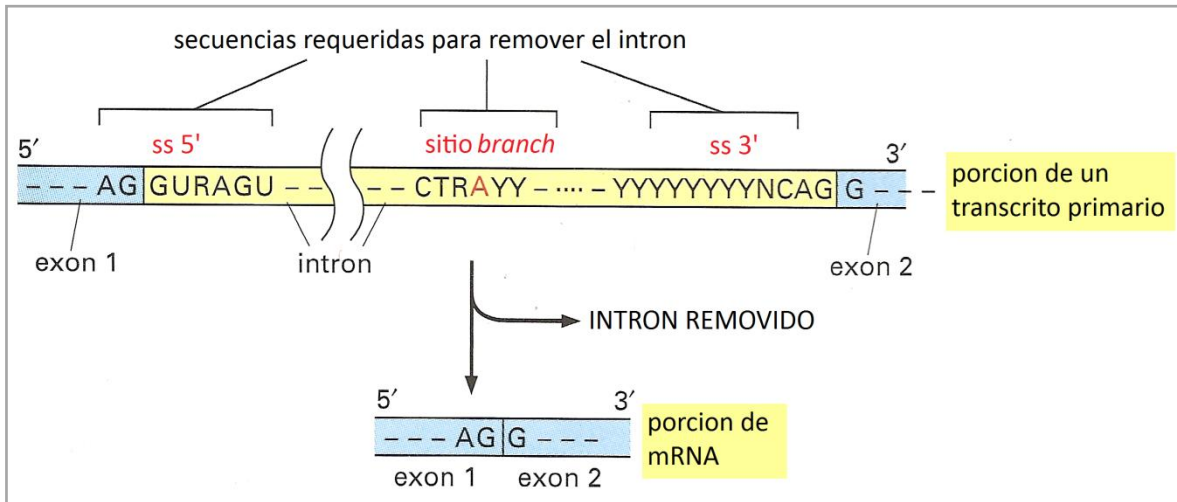
En azul se muestra la región 147E116B y en verde, la región 147Ex3A. Esta secuencia reconstituida se sometió a análisis BLASTX, obteniéndose similitud con  $\beta$ -1,3(4)-glucanasas con e-value del orden de  $10^{-5}$ . En el análisis de dominios conservados sólo se encuentra el que corresponde a la región 147Ex3A, con un e-value igual a  $10^{-10}$  (ver Figura 3.8).

Los análisis de las secuencias derivadas del *genome walking* dirigido a la familia 31 demostraron que existen básicamente dos secuencias que presentan similitud con enzimas de esta familia: la secuencia 127EV31A, que además representa a las secuencias 127PII31A y 127PII31B, y la secuencia 127EV31C. Sin embargo, todas estas secuencias fueron obtenidas a partir del DNA genómico de la muestra 127, mientras que los fragmentos a partir de los cuales se diseñaron los partidores pertenecían a la muestra 147 (fragmento 147Ex4A, ver Tabla 3.2). Por esta razón, no es posible unir las secuencias y obtener un fragmento más extenso.

### 3.7.3.1 Análisis de la presencia de intrones

Los genes de eucariontes usualmente presentan intrones que interrumpen la secuencia codificante de éstos. Dentro de los hongos, existe una gran diversidad en este aspecto: *S. cerevisiae* presenta en promedio menos de un intrón por gen, muchos *Ascomycetos* tienen en promedio 1-2 intrones por gen y algunos *Basidiomycetos* presentan hasta aproximadamente 7 por gen [35]. Por lo tanto, un análisis de la presencia de intrones es imprescindible en este caso, debido a que la cepa 147 es un *Basidiomyceto* y, además, los resultados de BLASTX presentaron alineamientos en dos marcos de lecturas, lo que sugiere que la secuencia codificante podría estar interrumpida por intrones.

*In vivo*, la remoción de las secuencias de intrones (o *splicing*) desde un transcrito primario de RNA, involucra el reconocimiento de tres regiones en el mismo: el sitio de *splice* (ss) 5', el ss 3' y el sitio *branch* dentro del intrón que forma la base de la región removida (ver Figura 3.18 [36]); cada uno de estos sitios posee una secuencia de consenso que sirve de señal para el proceso de *splicing* [36].



**Figura 3.18. Esquema de la estructura intrón-exón de un gen genérico.**

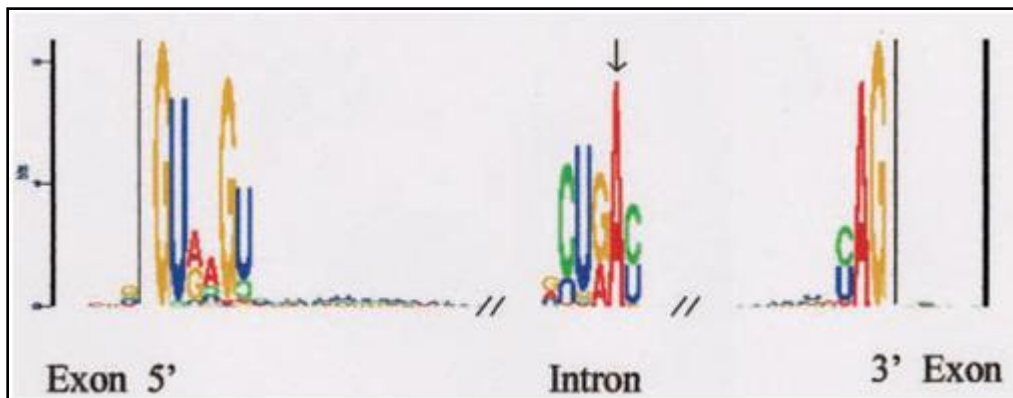
Para analizar la presencia de intrones en las secuencias, se buscaron herramientas que permitieran hacer la búsqueda en forma automática, sin embargo, se encontró que las herramientas disponibles para organismos eucariontes se centran en localizar intrones en los organismos más comúnmente estudiados (*S. cerevisiae*, *A. thaliana*, *H. sapiens*, entre otros). Se ha sugerido que los genes de hongos presentan diferencias con los genes de eucariontes superiores, basado principalmente en que los hongos exhiben intrones más cortos y exones más largos que aquéllos [37]. Es por esto que el análisis se realizó centrándose en un estudio realizado específicamente en hongos [37], en el cual se describen las secuencias de consenso para estos organismos. El estudio se centra en pocos casos y es por esto que este análisis no pretende ser concluyente sino más bien sugerir posibles intrones en las secuencias en estudio.

Las secuencias de consenso presentadas en el estudio son las siguientes (las bases degeneradas se encuentran de acuerdo al alfabeto en el Anexo 6.2).

- ss 5': G|GURAGU (la barra vertical indica el inicio del intrón)
- Sitio *branch*: RCURAY (la A subrayada corresponde al punto *branch*)

- ss 3': UNYAG| (la barra vertical indica el fin del intrón)

Se procedió a buscar estos consensos en las secuencias, pero éstos no pudieron ser localizados, por tanto, se procedió a relajar los criterios de búsqueda, buscando las secuencias de consenso que se describen para uno de los hongos en particular, *C. neoformans* (ver Figura 3.19), el cual es el único *Basidiomyceto*, al igual que las cepas 127 (*Peniophora gigantea*), 147 (*Lentinula edodes*) y 202 (*Pleurotus ostreatus*), a partir de las cuales se obtuvieron las secuencias con resultados significativos.



**Figura 3.19. Secuencias intrónicas de consenso descritas para *C. neoformans*.**

Logos de estructura de secuencia para el extremo 3' del exón adyacente al intrón, el ss 5', el sitio *branch*, el ss 3' y el extremo 5' del exón adyacente al intrón. El eje x muestra las posiciones nucleotídicas y el eje y refleja la frecuencia de ocurrencia de una base. Sólo se muestran las bases cuya frecuencia es significativamente mayor que una variación al azar.

Considerando sólo este organismo, las secuencias de consenso cambian a las siguientes:

- ss 5': |GTRNGT
- Sitio *branch*: CTRAY
- ss 3': YAG|

Realizando esta búsqueda en todas las secuencias que presentaron alineamientos significativos, sólo se encontraron los consensos en la secuencia 147GH16. En ella se encontraron tres posibles intrones, los cuales se muestran a continuación.

```
TCACGCGTCGACTAGTACTAGCTTACAGCTGCACACCTTGAGCGAGGGAATGGCAGCCACGGTTGAGGAAAGGCATCTTTC
GGAGGTTTATATGCGGCTAGATTTTCGTAGACACCTTGACGGAAGAAAGACGACTAAGAAGGGGGCAAATCCAGCAATTAAT
CGTTCGAGTCACCTTTTGGATTCCCTTGTGAGAGTTCCAAGATTCGCGTTTCTTTCCAGGTCATTAGTGGACAACCTGAGT
TTCCAGGTCCTCAATTCGCACATCTGATTGAGAAATGATATCCTAGATAATTTTCGCGAGCCAAAAGGTAAATTTTCGCCGG
TTGAGTTTGATTTTACTTTTGGAGGACATGGTAGCATGAGCTTCTAGAATATACTGTAGCAGTCGCAGAGGCCAGTCTGGT
GAGAGTGAGGATCGCTACAACGGTTCATGGTTAGAAGCAGATTTTCGGCAATTATTGGAGCCGACAGCCGAGAATGATCACA
TGACGATAAGCAGCTCGGTGCTAGCCCAGACTCTCATCAGTCTCAATCCCCGGTTGAATTTACTCTCAACGACTCCTTCTC
AAGGCCCGTCTCTGCTGCCGCTCCATCATCTGACTTACATCGACCCAAACACCAACCATGCGATAATCCGGGTCGATAATGC
GACAAATATAACGGCAGCTGCGACAGTGAACAGGAATACG GTGCGT GACCTCATCCATTTGTTGTTTATGCTCCACT
```

CATTTGTTTCT **CTAGGT**CAAGATCACTACGAAGGACGCATATCTCATCGGTAGTCTTGCTGTTATTGACGTTTGGCACATA  
 CACTACGGTTGTTCC **CTACGT**ATTTATCCTATCTAGTTTCATGATACCATTAT **CTAAT**CCTCCCAATCCCTCCCATTA  
**TAG**CTCTGGCCTTCATCTCTGGATGCTCGGCACCGACCTCCTCTGGCCAAATTCGGCGAAATCGACATCATCGGAGCCATC  
 AACAACTCGATTCAAACCAATACGCCCTGCACACCACCTGGATGTTTCCTAGGCACGAACTCGATACATTAGGGGAGCAT  
 TCTTCAAACCTTATTTCTCTACGGGCGCAGGGTGTGTAGTACAAGAAAAGAAACCGAGTAGTTTTGGAGTGGGTTTGC  
 GGCTGGCGGAGGTGTGTTGCGACGCAGTTGGATCTTAGTGGAATCTATGTGTGGTTCCAAGCCTACTTCGTGCAGCTGCT  
 GAA

En rojo se muestran las tres posibles regiones ss 5', en azul el único sitio *branch* encontrado y en verde el único ss 3' encontrado. Esto da origen a tres posibles *splicings*.

- Intrón de 204 pb
- Intrón de 154 pb
- Intrón de 69 pb

La secuencia de éstos se presenta en el Anexo 6.6.

Las tres secuencias generadas fueron sometidas a análisis BLASTX y CDD, encontrándose que la última secuencia mejora los puntajes de los resultados en ambas herramientas: BLASTX presenta alineamientos con  $\beta$ -1,3(4)-glucanasas con e-values del orden de  $10^{-9}$  (versus  $10^{-5}$  de la secuencia original) y CDD presenta similitud con la superfamilia GH16 con e-value igual a  $2 \times 10^{-14}$  (versus  $10^{-5}$  de la secuencia original). Otra evidencia que apoyaría la selección del tercer intrón presentado es el largo de éste, debido a que se tiene que el tamaño de los intrones de hongos son más bien pequeños, con un peak dominante entre 50 y 70 pb.

Sin embargo, la nueva secuencia no resuelve el problema del marco de lectura ya que igualmente se presentan alineamientos en los marcos +1 y +2; sin embargo, se debe tener presente que este análisis de intrones es completamente preliminar.

### 3.8 Estimación del porcentaje de secuenciación de las enzimas de interés

Cabe suponer que las secuencias con similitud a  $\beta$ -1,3(4)-glucanasas y  $\alpha$ -glucosidasas corresponden a una fracción del gen que las codifica, basado en los resultados obtenidos con BLASTX. Éstos permiten realizar una estimación del porcentaje del gen que ya se tiene secuenciado, considerando los resultados más significativos en cada caso. Un alineamiento en BLASTX indica la posición de la secuencia en la base de datos donde comienza y termina el alineamiento ( $BD_{fin}$  y  $BD_{fin}$ , respectivamente) y el largo de ésta ( $BD_{largo}$ ). Para esto, se consideraron los resultados con e-value menor que  $10^{-10}$  (con un máximo de 10 resultados considerados para cada secuencia). Este porcentaje fue calculado según la siguiente ecuación:



$$\% = \frac{BD_{fin} - BD_{ini}}{BD_{largo}} \times 100$$

Los resultados obtenidos para las secuencias analizadas se presentan en la Tabla 3.4.

**Tabla 3.4. Porcentaje secuenciado estimado para las secuencias en estudio.**

Secuencia	Porcentaje estimado
127EV31A	12%
127PII31B	12%
147GH16	17%

### 3.9 Aplicaciones de las glicosil hidrolasas encontradas

#### 3.9.1 $\beta$ -1,3(4)-D-glucanasas

Los  $\beta$ -glucanos son un componente principal de la pared celular de cebada, centeno, arroz y trigo, y consiste en aproximadamente 1.200 unidades de glucosa unidas por enlaces glicosídicos  $\beta$ -1,3 y  $\beta$ -1,4. Los enlaces glicosídicos  $\beta$ -1,3 dan cuenta de alrededor del 25-30% de la molécula de  $\beta$ -glucano [38].

En fuentes microbianas, hasta ahora se han descrito cuatro tipos principales de enzimas con actividad *endo*, capaces de depolimerizar 1,3-1,4- $\beta$ -D-glucanos. Éstas corresponden a (i) liquenasas (EC 3.2.1.73, 1,3-1,4- $\beta$ -D-glucano-4-glucanohidrolasa), encontradas principalmente en especies de *Bacillus*, (ii) 1,3(4)- $\beta$ -D-glucanasas (EC 3.2.1.6), activas en 1,3-1,4- $\beta$ -D-glucanos y 1,3- $\beta$ -D-glucanos, (iii) 1,3- $\beta$ -D-glucanasas (EC 3.2.1.39, 1,3- $\beta$ -D-glucano-3-glucanohidrolasa), las cuales presentan baja actividad en 1,3-1,4- $\beta$ -D-glucanos y (iv) 1,4- $\beta$ -D-glucanasas (EC 3.2.1.4, 1,4- $\beta$ -D-glucano-4-glucanohidrolasa o celulasa) [39].

**Aplicación en producción de bioetanol.** Se ha estudiado el uso de las enzimas producidas por el hongo *Talaromyces emersonii* para la conversión de material lignocelulósico en azúcares fermentables. En la etapa de hidrólisis de hemicelulosa se detectó la actividad de  $\beta$ -1,3(4)-glucanasas, entre otras enzimas. Su producción se vio inducida por crecimiento en sustratos ricos en hemicelulosa [40].

**Aplicación en la industria ganadera.** El glucano de cebada es un factor anti-nutricional especialmente para pollos y lechones, y el glucano en una dieta basada en cereales tiene

efectos perjudiciales en las heces. La adición de  $\beta$ -1,3(4)-D-glucanasa a estos alimentos mejora la digestibilidad de dietas basadas en cebada y reduce la enteritis [38]

**Aplicación en la industria alimenticia.** Durante el proceso de malteado, la  $\beta$ -1,3(4)-glucanasa se inactiva por temperatura, dejando un alto contenido de  $\beta$ -glucanos en el mosto de cerveza. Esto puede producir severos problemas, como baja tasa de filtración y rendimiento de extracción, al igual que una apariencia indeseable de precipitados gelatinosos en la cerveza. Por esta razón, a menudo se agrega  $\beta$ -1,3(4)-glucanasa para suplementar las enzimas de la malta, con el fin de reducir la viscosidad y obtener malta de alta calidad durante la maceración [38; 41].

Además, las enzimas que degradan glucanos de enlace mixto tienen potencial significancia en la industria médica ya que dichos glucanos son un componente importante de la fibra soluble en la dieta humana [31].

Como se puede ver, existen muchas aplicaciones para  $\beta$ -1,3(4)-D-glucanasas con la termoestabilidad y propiedades de pH apropiadas. Es por esto que se ha hecho necesaria la ingeniería de proteínas para diseñar enzimas con características adecuadas para estas aplicaciones [39]. En el caso de la secuencia encontrada en esta memoria, esto es particularmente atractivo debido a que el fragmento contiene el dominio catalítico conservado de la enzima, por lo que se pueden intentar estrategias de expresión de la enzima la cual, al contar con el dominio catalítico, podría encontrarse activa. En tal caso se pueden realizar intentos de evolución dirigida con el fin de aumentar la actividad específica y la estabilidad de la proteína recombinante. Una enzima de estas características podría ser aplicada en la fase de pretratamiento, específicamente en la degradación de hemicelulosa, previo a la hidrólisis de celulosa, en el proceso de producción de bioetanol de segunda generación.

### 3.9.2 $\alpha$ -glucosidasas

Las  $\alpha$ -glucosidasas (3.2.1.20) catalizan la liberación de glucosa desde los extremos no-reductores de  $\alpha$ -glucósidos, oligosacáridos con enlace  $\alpha$  y  $\alpha$ -glucanos, hidrolizando específicamente enlaces glicosídicos  $\alpha$ -1,4 y  $\alpha$ -1,6 [42; 43]. Estas enzimas presentan especificidad de sustrato muy diversa: algunas prefieren di-, oligo- y/o poliglucanos con enlace  $\alpha$ , mientras otras hidrolizan preferentemente sustratos heterogéneos tales como

sacarosa y sucrosa [44]. Algunos ejemplos de sustratos incluyen maltosa, maltotriosa, isomaltosa, panosa, kojibiosa, almidón soluble y granos de almidón [42].

Las  $\alpha$ -glucosidasas son un importante subgrupo de las amilasas, tanto en el metabolismo primario como en biosíntesis y procesamiento de glicoconjugados. Estas enzimas se encuentran principalmente en las familias 13 y 31 y, en menor grado, en las familias 4 y 63. Las  $\alpha$ -glucosidasas de la familia 31 han sido escasamente estudiadas, a pesar de su importancia y el número de distintas actividades representadas en un amplio rango de organismos, incluyendo animales, plantas y microorganismos. Esta familia contiene  $\alpha$ -glucosidasas tan importantes como la  $\alpha$ -glucosidasa lisosomal humana (cuya deficiencia resulta en la acumulación de glicógeno en el corazón y la musculatura esquelética), la  $\alpha$ -glucosidasa II del retículo endoplásmico que juega un rol clave en el procesamiento y plegamiento de glicoproteínas y la enzima digestiva sucrasa-isomaltasa, la cual es el blanco de las drogas antidiabetes acarbosa y miglitol [45].

La actividad de las  $\alpha$ -glucosidasas ha sido aplicada industrialmente para producir isomaltooligosacáridos y también para conjugar azúcares y convertirlos en materiales biológicamente útiles, con el fin de mejorar sus propiedades químicas y funciones fisiológicas [44]. A pesar de la variedad de aplicaciones industriales de estas enzimas, el mayor volumen se utiliza en la industria de almidón, para producción de jarabes con alto contenido de fructosa y etanol. Para que el almidón contenido en el maíz se encuentre disponible para hidrólisis, éste se somete a cocimiento bajo presión. Para lograr la licuefacción del almidón, se añaden amilasas termoestables, usualmente optimizadas mediante ingeniería de proteínas [2]. Al ser enzimas activas en almidón, las  $\alpha$ -glucosidasas poseen potenciales aplicaciones en la producción de bioetanol de primera generación.

### **3.10 Discusión final**

Se logró clonar fragmentos de dos glicosil hidrolasas, con actividad putativa  $\beta$ -1,3(4)-glucanasa y  $\alpha$ -glucosidasa, respectivamente. Estas enzimas no poseen la actividad a la cual se había dirigido el diseño de los partidores en una primera etapa (exoglucanasa). La principal explicación de esto se encuentra en la naturaleza degenerada de los partidores utilizados, diseñados a partir de regiones de homología de exoglucanasas.

La clasificación de glicosil hidrolasas de Henrissat ha definido también clanes, los cuales agrupan familias que presentan secuencias relacionadas [13]. Por otro lado, se tiene que

enzimas con diferente especificidad de sustrato se encuentran a veces en una misma familia, indicando una divergencia evolutiva para adquirir nuevas especificidades, como se observa por ejemplo en las familias 1, 13 y 16. Por otro lado, las enzimas que hidrolizan el mismo sustrato a veces se encuentran en distintas familias [46]. Todo lo anterior implica que puede suceder que al diseñar partidores basados en la homología de una familia, éstos alineen con secuencias en el genoma que corresponden a enzimas de otras familias, que tengan actividades distintas pero posean secuencias similares. Incluso se ha reportado que celulasas de la familia 7 tienen una disposición de los residuos catalíticos y un plegamiento similar a  $\beta$ -1,3-glucanasas y  $\beta$ -1,3(4)-glucanasas de la familia 16 [46].

Por otro lado, la estrategia de utilizar partidores degenerados da lugar a que puedan amplificarse secuencias completamente no relacionadas a las secuencias *target*, debido a que los partidores consisten en un pool de oligonucleótidos, entre los cuales pueden existir algunos con una secuencia que sea complementaria a otras regiones del genoma, producto del azar (como en el caso del fragmento con alta similitud a transcriptasa reversa encontrado).

Considerando lo anterior, es imprescindible hacer un análisis del trabajo realizado y destacar las principales ventajas y desventajas de la estrategia adoptada. Entre las ventajas se encuentra que es una estrategia *high-throughput*, es decir, permite el análisis de muchas muestras en forma rápida, sobre todo con el *screening* de actividad inicial que permitió descartar cepas no productoras de exoglucanasas. El trabajo posterior de biología molecular es relativamente sencillo, la extracción de DNA genómico con el método seleccionado es un proceso rápido y fácil, al igual que la técnica de PCR. Al ser técnicas de fácil utilización, también permitió el trabajo con todas las cepas seleccionadas en forma simultánea. Por otro lado, el utilizar regiones de homología de proteínas ya descritas permite trabajar con las cepas sin tener ningún conocimiento previo de sus genes en particular. En cuanto a las desventajas, la principal consistió en que es una estrategia poco selectiva, como ya se discutió anteriormente, los partidores degenerados no son altamente específicos para la actividad buscada y por lo tanto se obtienen amplificaciones de muchas regiones que no son las buscadas, ya sea que tienen homología en su secuencia o que son completamente no relacionadas (en este trabajo se dieron ambos casos). Éstas podrían tener mejor amplificación que las regiones de interés, yendo en desmedro de las últimas, las cuales pueden no ser detectadas. Además, al basarse en la homología de enzimas conocidas, la técnica no podrá detectar enzimas que posean secuencias completamente nuevas. Otra dificultad la presenta el

hecho de trabajar con DNA genómico. Si bien éste es de fácil extracción y manipulación, al tratarse de organismos eucariontes se debe considerar la presencia de intrones, los cuales interrumpen la secuencia codificante de los genes. Lamentablemente, no existe mucha información ni herramientas disponibles para identificar las secuencias intrónicas en hongos por lo que en este caso sólo se pudo realizar un análisis preliminar con la información disponible. Una última desventaja, que resultó ser fundamental, consistió en la dificultad de poner a punto las reacciones de PCR. Esta técnica es muy eficiente pero depende de un fino ajuste de la temperatura que se utiliza en la etapa de alineamiento. Al estar trabajando con varias cepas simultáneamente y con partidores degenerados en la primera etapa de PCR, la temperatura no fue necesariamente la óptima para todas. Se intentó barrer el mayor espectro de posibilidades al utilizar gradientes de temperatura pero de todos modos cabe la posibilidad de no haber utilizado la mejor temperatura de hibridación en cada caso.

Para poder lograr los objetivos planteados inicialmente, se pueden sugerir distintos caminos a seguir. Una opción sería mejorar el diseño de los partidores degenerados, los cuales fueron la principal causa de poca selectividad de la estrategia adoptada. Los partidores en este trabajo fueron diseñados a partir de regiones de homología de muchas exoglucanasas de hongos, algunos tal vez muy distantes evolutivamente a los hongos de la micoteca. Una alternativa es disminuir el número de hongos que se usan como modelo, limitándolo por ejemplo únicamente a Basidiomycetos, que corresponde a la mayoría de los hongos de la micoteca. Además, podrían probarse otras herramientas de apoyo para el diseño distintas a CODEHOP, la cual no siempre entregaba partidores en las regiones más conservadas, identificadas a simple vista, ya que el programa posee parámetros de diseño que pueden no ser los óptimos en este caso.

Otra opción es adoptar una estrategia global alternativa que podría llevar a obtener mejores resultados. Para esto, se podría clonar exoglucanasas a partir de genotecas de cDNA. Brevemente, esta estrategia debería desarrollarse de acuerdo a la siguiente metodología general:

1. Crecimiento de las cepas en un medio inductor de la expresión de celobiohidrolasas (alternativas: lactosa, soforosa, celobiosa [47]).
2. Extracción de transcritos desde las cepas (extracción de mRNA).
3. Producción de cDNA a partir del mRNA (transcripción reversa).
4. Clonamiento en un sistema de expresión (cDNA clonado en un vector de expresión).

5. Selección de clones productores de exoglucanasa por medio de ensayos de actividad en un sustrato apropiado (Avicel).
6. Secuenciamiento de los fragmentos seleccionados.

Este enfoque resuelve varios de los problemas que se presentaron en este trabajo, tales como la baja selectividad, ya que se ensaya la actividad de cada clon por lo que sólo se amplifican aquellos fragmentos que efectivamente poseen la actividad deseada; la presencia de intrones, ya que el cDNA se construye a partir de los transcritos de la célula, los cuales ya han sido procesados y no poseen intrones en su secuencia; se pueden encontrar enzimas que no presenten alto grado de homología con las enzimas ya conocidas; la puesta a punto de las temperaturas de hibridación tampoco es un problema ya que se pueden utilizar partidores del vector en el cual se inserten los cDNA; si los hongos son cultivados en un medio inductor, los transcritos de las enzimas que degradan el sustrato se encontrarán enriquecidos en comparación con otros transcritos no relacionados.

Sin embargo, esta estrategia presenta sus propias dificultades, la principal es la correcta expresión de los cDNA en el sistema de expresión. 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 que los cDNA provienen de organismos eucariontes, por lo tanto el procesamiento post-traducciona l de la estructura primaria de las proteínas podría no llevarse a cabo correctamente. Sin embargo, se sabe que en general las celobiohidrolasas son enzimas con poca glicosilación [48], lo cual podría facilitar su correcta expresión en un sistema bacteriano. La expresión de enzimas de origen fúngico en *E. coli* es escasa; sin embargo, existe información de expresión de enzimas de *T. reesei* en esta bacteria [49; 50]. De lo contrario, debería considerarse el uso de un sistema eucarionte, como por ejemplo *Saccharomyces cerevisiae*. *S. cerevisiae* es reconocidamente un mal secretor de enzimas al medio extracelular, debido a obstáculos tales como la hiperglicosilación y el transporte a través de la pared celular. A pesar de lo anterior, existen casos de genes de origen fúngico donde se ha logrado la secreción de proteína al medio extracelular en forma exitosa, utilizando señales de exportación nativas que median el efectivo procesamiento post-traducciona l y la secreción de proteínas funcionales [10]. En particular, se ha logrado la expresión y secreción eficiente de celobiohidrolasas de *Trichoderma reesei* en *S. cerevisiae* [51]. También se cuenta con el antecedente de la expresión y secreción eficiente de la celobiohidrolasa II de *T. reesei* en la levadura *Schizosaccharomyces*

*pombe* [52]. Cabe mencionar que las enzimas, aún cuando se expresaran correctamente, también deben ser secretadas al medio extracelular para llevar a cabo el ensayo de actividad que se utilizó en este trabajo. De lo contrario se presenta el desafío de la extracción de las enzimas desde el medio intracelular.

Como puede verse, el uso de una estrategia alternativa queda estrictamente sujeto a resolver los problemas de la expresión en forma activa de las enzimas en estudio. Además, este enfoque requiere de la selección de clones mediante ensayos de actividad que, *a priori*, podría ser el mismo ensayo de azúcares reductores utilizado en este estudio. No obstante, este ensayo presentó algunas dificultades que valdría la pena intentar solucionar para mejorar su aplicación en el futuro. El principal problema consistió en la insolubilidad del Avicel en el medio de reacción. Al ser insoluble, no se puede tener certeza de que éste se encuentra en suspensión en el medio de reacción y de este modo disponible para las enzimas. Por otro lado, al transferir una alícuota del medio de reacción para la lectura de absorbancia en placa, muchas veces era inevitable tomar algo de Avicel el cual se deposita en el fondo del pocillo y genera artefactos en la lectura, conduciendo a posibles errores. Estos problemas podrían ser solucionados si se aumenta el volumen de reacción; de este modo, podría asegurarse una mejor agitación del medio de reacción al ser más visible y, a la vez, facilitar la transferencia a una placa de lectura. Esto se logra ya que se permite una buena decantación del Avicel al cabo del tiempo de reacción, con un volumen sobrenadante adecuado para pipetearlo libre de Avicel. Una segunda alternativa es realizar ensayos de actividad directamente desde los medios de cultivo. Los clones en estudio pueden ser cultivados en un medio que contenga Avicel, en el cual se mida la presencia de azúcares reductores al tiempo cero (experimento “blanco” para corregir los valores) y al cabo de  $n$  días. Esta opción debe ser puesta a prueba para determinar su viabilidad, ya que puede ocurrir que el cambio en la concentración de azúcares reductores no sea perceptible o bien que el medio imponga alguna restricción al ensayo.

## 4. Conclusiones

Se concluye que el método de *screening* a gran escala fue útil para seleccionar en forma preliminar cepas de hongos productoras de glicosil hidrolasas que degradan el sustrato insoluble Avicel. El ensayo, sin embargo, posee algunos aspectos que podrían mejorarse, principalmente provocados por la naturaleza insoluble de la molécula de Avicel. Un aumento del volumen de reacción sería una alternativa factible, sin embargo, se debe intentar mantenerlo en un volumen pequeño que permita mantener el carácter de alto rendimiento del método aplicado.

Se concluye que la etapa de clonamiento de secuencias codificantes de glicosil hidrolasas con aplicación en la producción de bioetanol se logró en forma parcial. Se logró clonar secuencias parciales de glicosil hidrolasas en bacterias *E. coli* DH5 $\alpha$ , las cuales han sido clasificadas putativamente como enzimas del tipo  $\beta$ -1,3(4)-D-glucanasa (secuencia 147GH16, proveniente de *Lentinula edodes*) y  $\alpha$ -glucosidasa (secuencias 127EV31A, 127PII31A, 127PII31B, provenientes de *Peniophora gigantea*), identificadas como tales usando las herramientas bioinformáticas BLASTX y CDD, con altos niveles de confianza (e-values del orden de  $10^{-9}$  y  $10^{-30}$ , respectivamente). Estas enzimas tienen potenciales aplicaciones en la producción de bioetanol. Las  $\beta$ -1,3(4)-D-glucanasas podrían tener un rol en la degradación de hemicelulosa en la etapa de pretratamiento de lignocelulosas y las  $\alpha$ -glucosidasas podrían ser aplicadas en la degradación de almidón, en la producción de bioetanol de primera generación. Cabe hacer énfasis en que para la  $\beta$ -1,3(4)-glucanasa se logró clonar el sitio catalítico característico de esta familia de enzimas, lo que constituye un promisorio punto de partida para intentar su expresión en un sistema adecuado ya que cabe esperar que su correcta expresión podría resultar en una enzima activa. Además, al contar con el dominio catalítico, podrían iniciarse esfuerzos de mejoramiento de estabilidad y actividad, ya sea por métodos racionales o al azar.

Se concluye que el método utilizado de partidores degenerados diseñados a partir de regiones de homología no es un método con alta especificidad, como quedó demostrado con la obtención de otras secuencias no relacionadas con las secuencias *target*, siendo los dos ejemplos más claros las secuencias obtenidas que codifican transcriptasa reversa (secuencias 147EI16A y 147EI16C) y transportador de urea (secuencias 147Ex5A y 147Ex5C), todas estas secuencias provenientes de *Lentinula edodes*.



Por otro lado, en cuanto a las técnicas particulares aplicadas, la experiencia adquirida en este trabajo indica que en general las etapas de PCR deberán ser optimizadas para cada caso particular, lo cual puede bajar el rendimiento del método en general. También se concluye que el ensayo de actividad en sustrato insoluble puede ser mejorado para evitar artefactos que conducen a errores en los resultados que se obtienen.

Finalmente, se concluye que la estrategia global utilizada es útil como un método *high-throughput* para la identificación y clonación de genes de glicosil hidrolasas en general, sin embargo, no es factible dirigir la metodología para la obtención de secuencias de enzimas con actividades particulares, debido a que no se puede contar con la especificidad de los productos obtenidos ya que las glicosil hidrolasas constituyen una amplia gama de enzimas que presentan un importante grado de homología entre enzimas con actividades diferentes.

## 5. Referencias

- [1] van Beilen, J. B. (2002). Enzyme technology: an overview. *Current Opinion in Biotechnology* , 13, 338–344.
- [2] Cherry, J. R. (2003). Directed evolution of industrial enzymes: an update. *Current Opinion in Biotechnology* , 14, 438-443.
- [3] Kirk, O. (2002). Industrial enzyme applications. *Current Opinion in Biotechnology* , 13, 345–351.
- [4] J. Zaldívar, J. N. (2001). Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Appl Microbiol Biotechnol* , 56, 17-34.
- [5] Wyman, C. E. (1996). Ethanol production from lignocellulosic biomass: overview. En C. E. Wyman, *Handbook on bioethanol: production and utilization* (págs. 1-18). Washington DC: Taylor and Francis.
- [6] Kosaric, N. (1996). Ethanol – potential source of energy and chemical products. En G. R. H. J. Rehm, *Biotechnology* (págs. 122-203). Weinheim: VCH.
- [7] Carmona, R. *Biocombustibles Líquidos en Chile*. Laboratorio de Química de la Madera, Departamento de Ingeniería de la Madera, Facultad, de Ciencias Forestales, Universidad de Chile , Santiago.
- [8] Hamelinck, C. N. (2005). Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. *Biomass and Bioenergy* , 28, 384–410.
- [9] Y. Sun, J. C. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technology* , 83 (1), 1-11.
- [10] L. Lynd, P. W. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews* , 506-577.
- [11] P. Zhang, M. H. (2006). Outlook for cellulase improvement: screening and selection strategies. *Biotechnology Advances* 24 , 452-481.
- [12] Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J.* , 280, 309-316.

- [13] Cantarel, B. L. (2008). *Carbohydrate Active Enzymes database*. Recuperado el 6 de Marzo de 2009, de <http://www.cazy.org>
- [14] Henrissat, B. B. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem J.*, 316, 695-696.
- [15] Marchler-Bauer, A. (2007). CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res*, 35, 237-240.
- [16] Altschul, S. F. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25, 3389-3402.
- [17] Fred Hutchinson Cancer Research Center. (Oct de 2003). *BlockMaker*. Recuperado el Sep de 2008, de [http://blocks.fhcrc.org/blocks/make\\_blocks.html](http://blocks.fhcrc.org/blocks/make_blocks.html)
- [18] Larkin, M. A. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23 (21), 2947-2948.
- [19] Rose, T. M. (2003). CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Research*, 31, 3763-3766.
- [20] Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, 35, W43-W46.
- [21] SIB Switzerland. (8 de Feb de 1999). *Translate Tool*. Recuperado el Ene de 2009, de <http://www.expasy.ch/tools/dna.html>
- [22] Promega. (2007). *pGEM®-T and pGEM®-T Easy Vector Systems, Technical Manual*.
- [23] Hou, Y. (2007). Cloning, Sequencing and Expression Analysis of the First Cellulase Gene Encoding Cellobiohydrolase 1 from a Cold-adaptive *Penicillium chrysogenum* FS010. *Acta Biochimica et Biophysica Sinica*, 39 (2), 101-107.
- [24] Acevedo, J. P. (2008). Cloning of complete genes for novel hydrolytic enzymes from Antarctic sea water bacteria by use of an improved genome walking technique. *Journal of Biotechnology*, 133, 277-286.
- [25] Lee, C. (2001). Cloning and characterization of two cellulase genes from *Lentinula edodes*. *FEMS Microbiology Letters*, 205, 355-360.

- [26] Sakamoto, Y. (2005). Isolation and characterization of a fruiting body-specific exo- $\beta$ -1,3-glucanase-encoding gene, *exg1*, from *Lentinula edodes*. *Current Genetics*, *47*, 244–252.
- [27] Liu, D. (2000). Rapid Mini-Preparation of Fungal DNA for PCR. *Journal of Clinical Microbiology*, *38*, 471.
- [28] Molecular Research Center Inc. (Diciembre de 2007). *RNA & DNA isolation reagents for RT-PCR, PCR, microarrays*. Recuperado el 04 de Marzo de 2009, de <http://www.mrcgene.com/tri.htm>
- [29] Eiinsele, H. (1997). Detection and identification of fungal pathogens in blood by using molecular probes. *Journal of Clinical Microbiology*, *35*, 1353–1360.
- [30] NCBI. (s.f.). *Blast: Frequently Asked Questions*. Recuperado el Marzo de 2009, de [http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\\_TYPE=BlastDocs&DOC\\_TYPE=FAQ#expect](http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect)
- [31] Görlach, J. M. (1998). Cloning and Targeted Disruption of *MLG1*, a Gene Encoding Two of Three Extracellular Mixed-Linked Glucanases of *Cochliobolus carbonum*. *Applied and Environmental Microbiology*, *64* (2), 385–391.
- [32] Reizer, J. (1994). A functional superfamily of sodium/solute symporters. *Biochim. Biophys. Acta*, *1197* (2), 133-166.
- [33] Jung, H. (2002). The sodium/substrate symporter family: structural and functional features. *FEBS Lett.*, *529* (1), 73-77.
- [34] Xiong, Y. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J*, *9* (10), 3353-3362.
- [35] Nielsen, C. (2004). Patterns of Intron Gain and Loss in Fungi. *PLoS Biol*, *2* (12), 2234-2242.
- [36] Alberts, B. (2002). From RNA to protein. En B. Alberts, *Molecular Biology of the Cell* (págs. 318-319). New York: 4th ed.
- [37] Kupfer, D. M. (2004). Introns and Splicing Elements of Five Diverse Fungi. *Eukaryotic Cell*, *3* (5), 1088-1100.

- [38] Teng, D. (2006). Cloning of  $\beta$ -1,3-1,4-glucanase gene from *Bacillus licheniformis* EGW039 (CGMCC 0635) and its expression in *Escherichia coli* BL21 (DE3). *Appl Microbiol Biotechnol*, 72, 705–712.
- [39] McCarthy, T. (2003). Catalytic properties and mode of action of three endo- $\beta$ -glucanases from *Talaromyces emersonii* on soluble  $\beta$ -1,4- and  $\beta$ -1,3;1,4-linked glucans. *International Journal of Biological Macromolecules*, 33, 141–148.
- [40] Fernandes, S. (2008). Enzyme Systems from the Thermophilic fungus *Talaromyces emersonii* for sugar beet conversion. *BioResources*, 3 (3), 898-909.
- [41] Louw, M. E. (1993). Characterization, cloning and sequencing of a thermostable endo-(1,3-1,4)- $\beta$ -glucanase-encoding gene from an alkalophilic *Bacillus brevis*. *Appl Microbiol Biotechnol*, 38, 507-513.
- [42] Beers, E. P. (1990). Partial Characterization and Subcellular Localization of Three  $\alpha$ -Glucosidase Isoforms in Pea (*Pisum sativum* L.) Seedlings. *Plant Physiol.*, 94, 738-744.
- [43] van der Maarel, M. (2002). Properties and applications of starch-converting enzymes of the  $\alpha$ -amylase family. *Journal of Biotechnology*, 94, 137-155.
- [44] Kato, N. (2001). Novel  $\alpha$ -Glucosidase from *Aspergillus nidulans* with Strong Transglycosylation Activity. *Applied and Environmental Microbiology*, 68 (3), 1250–1256.
- [45] Lovering, A. L. (2005). Mechanistic and Structural Analysis of a Family 31  $\alpha$ -Glycosidase and its Glycosyl-Enzyme Intermediates. *The Journal of Biological Chemistry*, 280 (3), 2105-2115.
- [46] Dies, G. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, 3, 853-859.
- [47] Ilmén, M. (1997). Regulation of Cellulase Gene Expression in the Filamentous Fungus *Trichoderma reesei*. *Applied and Environmental Microbiology*, 63 (4), 1298-1306.
- [48] Baldrian, P. (2008). Degradation of cellulose by basidiomycetousfungi. *FEMS Microbiol. Review*, 32, 501-521.
- [49] O. Hirofumi, K. T. (1998). Molecular Characterization and Heterologous Expression of the Gene Encoding a Low-Molecular-Mass Endoglucanasefrom *Trichoderma reesei* QM9414. *Applied and Environmental Biotechnology*, 555-563.

[50] Laymon, R. (1996). Cloning and expression of full-length *Trichoderma reesei* cellobiohydrolase I cDNAs in *Escherichia coli*. *Appl. Biochem. Biotechnol.*, *Spring*, 389-397.

[51] M. E. Penttila, L. A. (1988). Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*. *Gene*, *63*, 103-112.

[52] Hirofumi, O. (1998). Efficient secretion of *Trichoderma reesei* cellobiohydrolase II in *Schizosaccharomyces pombe* and characterization of its products. *Appl. Microbiol. Biotechnol.*, *49*, 301-308.

## 6. Anexos

### 6.1 Micoteca de trabajo

**Tabla 6.1. Micoteca de trabajo.**

La columna *Clave* corresponde al número asignado a cada muestra que se utilizó para identificarlas durante el trabajo de memoria. La información en la columna género fue obtenida del NCBI. En la columna Tipo, *HPB*, *HPC* y *HPBL* corresponde a hongo de pudrición blanca, café y blanda, respectivamente, información que fue proporcionada junto con la micoteca.

Clave	Especie	Género	Tipo
96	<i>Amblyosporium sp.</i>	Microsporidia	
97	<i>Coriolus multicolor</i>	Basidiomycota	HPB
98	<i>Lenzites seiparia</i>	Basidiomycota	
100	<i>Auricularia politricha</i>	Basidiomycota	
101	<i>Pluteus cervinus</i>	Basidiomycota	
104	<i>Poria monticola</i>	Basidiomycota	HPC
106	<i>Polystictus versicolor</i>	Metazoa	HPB
108	<i>Polystictus sulphureus</i>	Metazoa	
110	<i>Poria placenta</i>	Basidiomycota	
113	<i>Phanerochaetes chrysosporium</i>	Basidiomycota	HPB
114	<i>Agrocybe aegerita</i>	Basidiomycota	
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	
147	<i>Lentinula edodes</i>	Basidiomycota	
151	<i>Trametes versicolor</i>	Basidiomycota	HPB
152	<i>Ungulina fomentaria</i>	Basidiomycota	HPB
162	<i>Coriolus versicolor</i>	Basidiomycota	HPB
163	<i>Fusarium oxisporiumsp</i>	Metazoa	
164	<i>Gymnopilus spectabilis</i>	Basidiomycota	HPB
167	<i>Pignosporus cinnabarinus</i>	Basidiomycota	HPB
168	<i>Stereum hirsutum</i>	Basidiomycota	
177	<i>Gymnopilus spectabilis</i>	Ascomycota	
178	<i>Lenzites betulinus</i>	Basidiomycota	HPB
181	<i>Alternaria alternata</i>	Basidiomycota	HPB
183	<i>Grifola gargal</i>	Basidiomycota	
186	<i>Stereum hirsutum</i>	Basidiomycota	
190	<i>Pleurotus ostreatus</i>	Basidiomycota	
202	<i>Pleurotus ostreatus</i>	Basidiomycota	HPB

## 6.2 Alfabeto degenerado

**Tabla 6.2. Alfabeto degenerado.**

Letra	Base(s)
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

## 6.3 Métodos adicionales

### 6.3.1 Verificación de extracción de DNA

La verificación fue realizada mediante un PCR con partidores universales de hongos, los cuales amplifican una región conservada del gen multicopia 18S [29]. 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. Los partidores requeridos para el método son:

- F18S\_fwd: 5'-ATTGGAGGGCAAGTCTGGTG
- F18S\_rev: 5'-CCGATCCCTAGTCGGCATAG

### 6.3.2 Purificación de DNA desde geles de agarosa

Se utilizó el kit QIAEX II, siguiendo las instrucciones del fabricante. En el paso final, el DNA fue eluído en 20 µl de buffer TE, del mismo fabricante.



### 6.3.3 Ligación de fragmentos en vector de clonación

Los fragmentos purificados desde agarosa son ligados en el vector de clonación en la siguiente mezcla de reacción: 50 ng de pGem®-T Easy, buffer de ligación 1X, 1 U de DNA ligasa de T4 y 3,5 µl de inserto, en un volumen final de 10 µl. La mezcla de ligación se deja reaccionar toda la noche a 4°C.

### 6.3.4 Transformación de bacterias *E. coli* DH5α

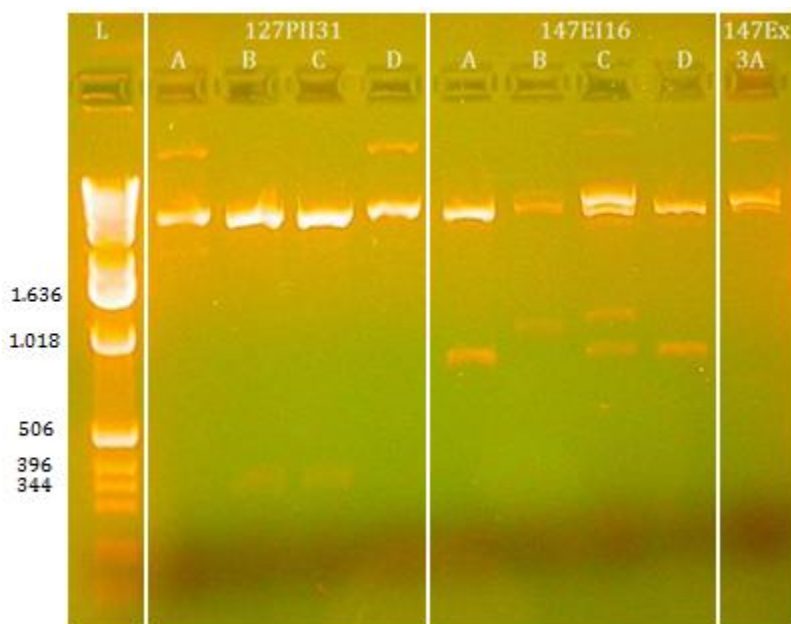
Para obtener un mayor número de copias de los constructos plasmidio-inserto, éstos son clonados en células electrocompetentes *E. coli* DH5α. Para esto, se mezcla 1 µl de la mezcla de ligación descrita anteriormente con 20 µl de células. Para la electroporación se utilizó un equipo Cell-Porator, con 420 V, 330 µF, baja impedancia, una tasa de carga rápida y un tiempo máximo de 2,5 ms. Las células transformadas se recuperan por espacio de 1 h a 37°C en 1 ml de medio de recuperación. Las células se siembran en placas con medio LB selectivo, las cuales se incuban a 37°C toda la noche. Al día siguiente, se han obtenido colonias blancas y azules, correspondiendo las blancas a colonias de clones transformantes. Se seleccionan 4 colonias transformantes por placa y se traspasan a tubos de ensayo con 4 ml de medio líquido LB-amp, lo cual se incuba a 37°C, con agitación orbital a 200 RPM, durante toda la noche. Las composiciones de los medios se presentan en el Anexo 6.4.

### 6.3.5 Extracción de DNA plasmidial desde *E. coli* DH5α

Los cultivos líquidos de *E. coli* DH5α se centrifugan a 13.000 RPM durante 2 min para formar un pellet de células. Este pellet se utiliza para extraer el DNA plasmidial de las células utilizando el kit QIAprep Spin Miniprep, siguiendo las instrucciones del fabricante.

### 6.3.6 Verificación de inserción del inserto

Para verificar la presencia del inserto en el vector de clonamiento antes de mandar a secuenciar, se realiza una digestión del DNA plasmidial extraído. Para esto, se digieren 5 µl del DNA plasmidial con 1 U de EcoRI y buffer NE EcoRI 1X en un volumen total de 10 µl, durante 4 h a 37°C. El producto de la digestión se visualiza en un gel de agarosa. En la Figura 6.1 se presenta la fotografía de una de estas electroforesis a modo de ejemplo.



**Figura 6.1. Digestión de DNA plasmidial**

Los carriles marcados *L* corresponden a marcador de 1 Kb. Los carriles se encuentran marcados con la muestra que contienen. Se observa que para las muestras 127PII31 A y D y la muestra 147Ex3A no hay presencia de inserto. Además, se observan dos bandas de alto peso molecular, lo que indica una digestión incompleta. En el caso de las muestras 127PII31B y C se observa un inserto de aproximadamente 350 pb. Para la muestras 147EI16x se observa una población de fragmentos de alrededor de 1 Kb.

### 6.3.7 Métodos alternativos de extracción de DNA

#### 6.3.7.1 Método con acetato de potasio [27]

Poner una muestra de micelio en 500  $\mu$ l de buffer de lisis (400 mM Tris-HCl pH 8,0, 60 mM EDTA pH 8,0, 150 mM NaCl y 1% SDS) en un tubo Eppendorf de 1,5 ml. Romper el micelio con ayuda de un mondadientes y dejar a temperatura ambiente durante 10 min. Agregar 150  $\mu$ l de acetato de potasio (acetato de potasio 3 M, 11,5% ácido acético glacial, pH 4,8) y aplicar vórtex brevemente. Centrifugar a 10.000 RPM durante 1 min y transferir el sobrenadante a un tubo nuevo. Agregar 1 volumen de alcohol isopropílico, invirtiendo el tubo para mezclar, y centrifugar a 10.000 RPM durante 2 min. Descartar el sobrenadante y lavar el pellet con 300  $\mu$ l de etanol al 70%. Centrifugar a 10.000 RPM durante 1 min, descartar el sobrenadante y secar el pellet al aire. Finalmente, resuspender el pellet en buffer Tris-EDTA 1X.

#### 6.3.7.2 Método con reactivo TRI® [28]

Tomar 50-100 mg de tejido y agregar 1 ml de TRI® (Sigma-Aldrich), agitar vigorosamente y dejar 5 min a temperatura ambiente. Agregar 0,2 ml de cloroformo y agitar vigorosamente por 15 min. Incubar a temperatura ambiente 5 min y centrifugar a 12.000 g durante 15 min a 4°C.

A continuación, retirar la fase acuosa (superior) y agregar 0,3 ml de etanol 100%, invirtiendo para mezclar. Incubar a temperatura ambiente 2-3 min y centrifugar a 2.000 g durante 5 min a 4°C. Descartar el sobrenadante y lavar el pellet con 1 ml de citrato trisodio 0,1 M en 10% de etanol, dejar actuar durante 30 min, agitando periódicamente. Centrifugar a 2.000 g durante 5 min a 4°C y repetir los últimos pasos de lavado con citrato trisodio y la centrifugación. Resuspender en 1,5 ml de etanol 75% e incubar a temperatura ambiente durante 15 min, mezclando periódicamente. Centrifugar a 2.000 g durante 5 min a 4°C, descartar el sobrenadante y secar el pellet al aire, el cual luego debe ser resuspendido en 0,3-0,6 ml de NaOH 8 mM, pasando lentamente por una pipeta. Finalmente, centrifugar a 12.000 g durante 10 min y transferir el sobrenadante a un tubo nuevo.

#### **6.4 Medios de cultivo**

Todos los medios se prepararon en agua destilada MilliQ y fueron autoclavados para asegurar esterilidad.

##### 6.4.1 Medios líquidos

- Medio mínimo-Avicel  
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%, Avicel 2%.
- Medio LB  
15,5 g/L de LB, lleva a 10g/L de NaCl (agregar 9,5 g/L de NaCl).
- Medio 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.

##### 6.4.2 Medios sólidos

- Medio agar malta  
Agar 1,5%, malta 2%.
- Medio agar malta Avicel  
Agar 1,5%, malta 2%, Avicel 1%
- Medio agar papa dextrosa  
39 g/L de agar papa dextrosa

- Medio LB selectivo

Agar 1,5%, ampicilina 100 µg/ml, X-Gal 40 µg/ml (en di-metilformamida), IPTG 0,1 mM.

## 6.5 Tablas de datos

**Tabla 6.3. Datos obtenidos en los ensayos de actividad.**

El valor de actividad corresponde al promedio de los ensayos en duplicado, al cual se sustrae el promedio de los experimentos blanco en duplicado. Los ensayos fueron realizados en una serie de días de los cultivos y se consideró el máximo para cada cepa para elaborar el Gráfico 3.1.

Cepas	Actividad exoglucanasa en Avicel [ $U_{abs}$ ]				Máximo de actividad [ $U_{abs}$ ]
	Día 3	Día 4	Día 5	Día 7	
96	0,011	-0,008	-0,070	-0,359	0,011
97	-0,021	-0,071	-0,141	-0,435	-0,021
98	-0,366	-0,050	-0,068	-0,091	-0,050
100	0,030	-0,009	-0,072	-0,066	0,030
101	-0,497	-0,008	0,005	-0,131	0,005
104	-0,003	-0,084	-0,085	-0,082	-0,003
106	-0,015	-0,369	-0,100	-0,093	-0,015
108	0,029	-0,014	-0,022	-0,079	0,029
110	0,024	-0,125	-0,086	-0,101	0,024
113	0,032	-0,031	-0,147	-0,126	0,032
114	0,262	-0,060	-0,086	-0,049	0,262
117	0,037	-0,127	-0,120	0,024	0,037
121	0,001	-0,019	-0,049	-0,048	0,001
127	0,090	0,189	0,133	0,132	0,189
140	0,015	-0,055	-0,075	-0,020	0,015
147	0,005	0,051	0,048	0,061	0,061
151	-0,011	-0,125	-0,173	-0,098	-0,011
152	0,005	0,000	0,005	-0,078	0,005
162	0,000	-0,026	0,005	-0,049	0,005
163	0,073	-0,038	-0,006	0,355	0,355
164	0,080	-0,157	0,025	-0,077	0,080
167	0,052	0,003	-0,008	-0,027	0,052
168	0,001	-0,364	-0,085	-0,124	0,001
177	0,002	0,072	0,012	-0,113	0,072
178	0,060	0,049	0,006	-0,076	0,060
181	0,021	0,026	-0,029	0,030	0,030
183	0,046	0,013	0,001	-0,120	0,046

Cepas	Actividad exoglucanasa en Avicel [ $U_{abs}$ ]				Máximo de actividad [ $U_{abs}$ ]
	Día 3	Día 4	Día 5	Día 7	
186	0,025	0,043	0,110	0,052	0,110
190	0,006	-0,117	-0,048	-0,015	0,006
202	-0,006	0,301	0,111	-0,374	0,301

## 6.6 Alineamiento para familia 6.

A continuación se presenta el alineamiento de secuencias de exoglucanasas de la familia 6 con los partidores diseñados por CODEHOP.

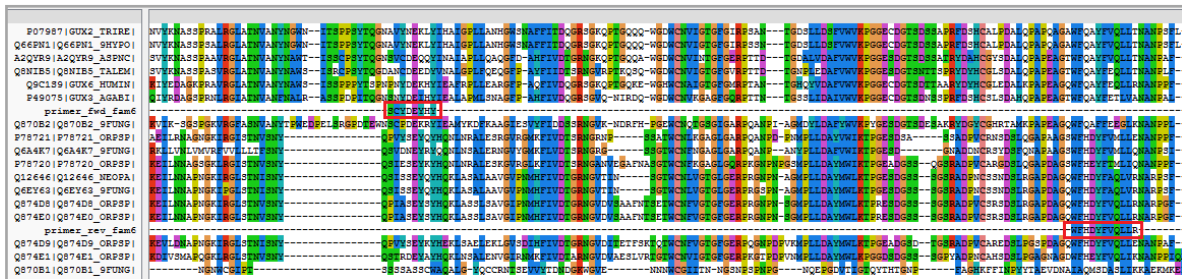


Figura 6.2. Alineamiento secuencias familia 6 con partidores.

A la izquierda de la foto se muestra enmarcado en rojo el partidore *exoF\_fam6\_fwd* y a la izquierda, el partidore *exoF\_fam6\_rev*.

## 6.7 Secuencias

A continuación se presentan las secuencias obtenidas en el trabajo, ya habiendo eliminado las secuencias del vector que las flanquean. No se presentan las secuencias que correspondían al vector de clonamiento sin inserto, ni aquéllas que no presentaron alineamientos significativos.

>147Ex3A

```
TTCAGCAGCTGCACGAAAGTAGTCTTGGAAACCAGGAATACGGTGCCTTGACCTCATCCATTTGTTGTTTTATGCTCCACTCA
TTTGTTCCTGCAGGTGAAGATCACACAGGACGCATATCCCATCGGTAGTCTTGTGTTATTGAGCTTTGGCACATACA
CTACGGTTGTTCCGTCAGTTATTTATCCTATCTAGTTCATGATACCATATCTAATCCCTCCCCATTAATAGGTCTGGCCT
TCATTCGGATGCTCGGCACCGACCTCCTCTGGCCAAATTCGGCGAAATCGACATCATCGGAGCCATCAACAACCTCGAT
TCAAACCAATACGCCCTGCACACCACCTGGATGTTTCTAGGCACGAACTCGATACATTCAGGGAGCATTCTTCAAACCTTA
TTTCTCTACGGGCGCAGGTGTGTAGTACAAGAAAAGAAACCGAGTAGTTTTGGAAGTGGGTTTGGCAGGCTGGCGGAGG
TGTGTTTGGCAGCGAGTTGGATCTTAGTGGAAATCTATGTGTGGTTCCAAGCCTACTTCGTGCAGCTGCTGAA
```

>147Ex3B

```
TCAGCCCATCTACGACGAGAAGCGGTACCTGCCAACGCCAACATATATGGTCTAGGCGAACACACAGATCCCTTCAGGCTA
CCTACCTCGAACACAACCCGCACAATGTGGTCCCGAGATGCATATGGTGTTCGGTTGGAACGAACCTCTACGGAAATCAC
CCCATAATTTTGAACATCGGACTACTGGTACCCATGGCGTTTTCTTGTCTCAACTCGAATGGAATGGACATCAAGATCAAC
GATTCTGGTTCCAAGCCTACTTCGTGCAGCTGCTGAA
```

>147Ex4A

```
TCAGCCCATCTACGACGAGAAGCGGTACCTGCCAACGCCAACATATATGGTCTAGGCGAACACACAGATCCCTTCAGGCTA
CCTACCTCGAACACAACCCGCACAATGTGGTCCCGAGATGCATATGGTGTTCGGTTGGAACGAACCTCTACGGAAATCAC
```

CCCATATATTTTGAACGTCGGACTACTGGTACCCATGGCGTTTTCTTGTCTCAACTCGAATGGAATGGACATCAAGATCAAC  
GATTCTGGTTCCACGCCTACTTCGTGCAGCTGCTGAA

>147Ex4B

TTCAGCAGCTGCACGAAGTAGGCGTGGAACCAGAATCGTTGATCTTGATGTCCATTCCATTTCGAGTTGAGCAAGAAAACGC  
CATGGGTACCAGTAGTCCGATGTTCAAATATATGGGGTGATTTCCGTAGAGGTTTCGTTCCAACCGGAACCATATGCAT  
CTCGGGACCACATTGTGCGGGTTGTGTTCGAGGTAGGTAGCCTGAAGGGATCTGTGTGTTTCGCTTAGACCATATATGTTGG  
CGTTGGCAGGTACCGCTTCTCGTCGTAGATGGGCTGA

>147Ex5A

GGNNNNNTANNNTNTTTANGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTTGGGAGCTCTCCCATATGGTTCGA  
CCTGCAGGCGCGCGGAATTCAGTAGTGATTTTCAGCAGCTGCACGAAGTAGGCGTGAAACCACAGTCAAAGCGTATCTTGT  
AGGGGGATTAGCTTGGTACATTCCCCGTTACCATATCATGAACGGACTGACATTATCTCTACATGTGTCCATCCAGGTTTG  
CCATCCCGTTTCACATTGCCACCACCTCGGATTGTCAGCGGTTCGATTGCGAGGAGACCCGGATATGAGACCGCTCTCCG  
CGGCAGATGTGTCTGCGGGGCTTCCCCCGCTGCAGCTGCAGCCGCGCCCATGGGTCAAGCCGGGGCCGCGGCTCTCTGG  
TGCTCTTGTTCCTTGTGTGACGAGTGCCACTTTCAGCAGAATTAATTGCTGTGACGAGCTTGTAAACATATGATGTGTTCA  
AGGTACGGATTATGCGGTACCGCAGAAGGCTGCTAGGCGCGCTTACTGTGCTTGTGTTTTCCAGAAATATATCAAGCC  
TAATGCCACGGATAGTCAACTTTTGCAGTGGGACATATACATGGTTGGTCTGGTGAATCCTCATTTCGCTTTCCCTCTCTT  
TAACATGAACAGGTTGGCCTTTATGGTCTTGTGATGGCGCTTGCAGGCCAATCTTCTTCTATATTTGGCGTATCGATGGGC  
TGTTTATACGTAAGTCTTTCTCCAACGGTGACTCTTCTTTCCGAAACGCGTTGATTGTAGCCTCCATCACTTTTCCCTTAT  
TCTCGTACCATGACTATGACTCATTAGATTGTCTCTCTTGTCTCTTCTAGACATTCATGGGTGAATTCATAGGCTCCG  
GGTCTGTTCCCATTCGACTTTGTATCACATGGAGTAAAGCTAACAAGTGGGGATGTATA

>147Ex5C

GGNNNNNTAGGCTNCTTAGGTGAACATAGAATACTCAAGCTATGCATCCAACGCGTTGGGAGCTCTCCCATATGGTTCGAC  
CTGCAGGCGCGCGGAATTCAGTAGTGATTTTCAGCAGCTGCACGAAGTAGGCTTGGAAACCACAGTCAAAGCGTATCTTGT  
GGGGGATTAGCTTGGTACATTCCCCGTTACCATATCATGAACGGACTGACATTATCTCTACATGTGTCCATCCAGGTTTG  
CATCCCGTTTCACATTGCCACCACCTCGGATTGTCAGCGGTTCGATTGCGAGGAGACCCGGATATGAGACCGCTCTCCGC  
GGCAGATGTGTCTGCGGGGCTTCCCCCGCTGCAGCTGCAGCCGCGCTCATGGGTCAAGCCGGGGCCGCGGCTCTCTGGT  
GCTCTTGTTCCTTGTGTGACGAGTGCCACTTTCAGCAGAATTAATTGCTGTGACGAGCTTGTAAACATATGATGTGTTCAA  
GGTACGGATTATGCGGTACCGCAGAAGGCTGCTAGGCGCGCTTACTGTGCTTGTGTTTTCCAGAAATATATCAAGCCT  
AATGCCACGGATAGTCAACTTTTGCAGTGGGACATACATGGTTGGTCTGGTGAATTCATTCGCTTTCCCTCTCTTT  
AACATGAACAGGTTGGCCTTTATGGTCTTGTGATGGCGCTTGCAGGCCAATCTTCTTCTATATTTGGCGTATCGATGGGCT  
GGTTATACGTAAGTCTTTCTCCAACGGTGACTCTTCTTTCCGAAACGCGTTGATTGTAGCCTCCATCACTTTTCCCTTATT  
CTCGTACCATGACTATGACTCATTAGATTGTCTCTCTTGTCTCTTCCAGACATTCATGGGTGAATTCATAGGCTCCGG  
GGTCTGTTCCCATTCGACTTTGTATCACATGGAGTAAAGCTAACAAGTGGGGATGTATA

>163Ex6A

TCAGCAGCTGCACGAAGTAGGCATGGAACCACACATAGATTCCACTAAGATCCAACGCGTTCGAAACACACCTCCGCCAG  
CCTGCGCAAACCCACTTCCAAAACGCTCGGTTTTCTTTCTTGTACTACACACCCTGCGCCCGTAGAGAAATAAGTTTGAA  
GAATGCTCCCTGAATGTATCGAGTTCTGTGCTTAGGAAACATCCAGGTGGTGTGAGGGCGTATTGGTTTGAATCGAGGTTG  
TTGATGGCTCCGATGATGTGATTTTCGCCGAATTTGGCCAGAGGAGGTCGGTTCGCGAGCATCCAGAATGAAGGCCAGACC  
TATTAATGGGGAGGATTAGATAATGGTATCATGAACATAGATAGGATAAATAACGTACGGAACAACCGTAGTGTATGTGCC  
AAACGTCATAACAGCAAGACTACCGATGGGATATGCGTCTTTCGATGATCTTCCCTGCAGAAACAAATGAGTGGAGC  
ATAAAACAACAATGGATGAGGTCAACGCACCGTATTCCTGGTTTCAAGACTACTTCGTGCAGCTGCTGAA

>163Ex6B

TCAGCAGCTGCACGAAGTAGGCCTGGAACCAGGAATACGGTGCCTGACCTCATCCATTTGTTGTTTTATGCTCCACTCAT  
TTGTTTCTGCAGGTGAAGATCACTACGAAGGACGCATATCCCATCGGTAGTCTTGTGTTATTGACGTTTGGCACATACAC  
TACGGTTGTTCCGTACGTTATTTATCCTATCTAGTTTCATGATACCATTTATCTAATCCCTCCCCATTAATAGGCTGGCCTT  
CATTCTGGATGCTCGCACCGACCTCCTCTGGCCAAATTCGCGGAAATCGACATCATCGGAGCCATCAACAACCTCGATT  
CAAACCAATACGCCCTGCACACCACCTGGATGTTTCTTAGGCACGAACCTCGATACATTCAGGGAGCATTTCTTCAAACCTTAT  
TTCTCTACGGGCGCAGGTTGTGTAGTACAAGAAAAGAACTGAGCAGTTTTGGAAGTGGGTTTGGCAGGCTGGCGGAGGT  
GTGTTTGGCAGCAGTTGGATCTTAGTGAATCTATGTGTGGTTCCATGCCTACTTCGTGCAGCTGCTGAA

>127EV31A

TGTCGACTAGTACTAGCTTAACAGLAAACAAGGATACATGTCAAGATCACCGATCCTGCCGTCCAACGCTACGAAGTCCCC  
GATTCATCCTTCCGCGACCCAGTCCCTCACTGGGAACTCTCTCGGCAACCTCTCAAATTCATTCAACTTTACTGCCTCT  
CCCTTCTCCTTCAACTCTTTTCGCACGCGCACCTCTGAAGTCTTATCAATACTGCTTCTTTCCCTCTGATTTTTGAGCCA  
CAATATCTACGACTCAAGACCTCGCTACCTGCCAACGCCAATATATGGTCTAGGCGAACACACAGATCCCTTCAGGCTA  
CCTACCTCGAACACAACCCGCATA

>127EV31C  
TCACGCGTCGACTAGTACTAGCTTGGCCACGCGTCGACTAGTACTAGCTTAGGCCACGCGTCGACTAGTACTAGCTAGGTC  
ACGCGTCGACTAGTATAGCTTGGACACGCGTCGACTAGTACTAGCTTGGCCACGCGTCGACTAGTACTAGCTAGGCCACGC  
GTCGACTAGTACTAGCTAGGCCACGCGTCGACTAGTACTAGCTTAGGCCACGCGTCGACTAGTACTAGCTTCCCTGCCAACG  
CCAACATATATGGTCTAGGCCAACACACAGATCCCTTCAGGCTACCTACCTCGAACACAACCCGCATA

>127PII31B  
TCACGCGTCGACTAGTACTAGCTTCCGTCCAACGCTACGAAGTCCCCGATTTCAGTCTTCCGCGACCCAGTCCCTCACTGG  
GAACTCTCTCGGCAACCTCTCAAATTCAAATTCAACTTTACTGCCTCTCCCTTCTCCTTCACAATCTTTTCGCACGCGCACCT  
CTGAAGTCTTCAATACTGCTTCTTCCCTCTGATTTTCGAGCCACAATATCTACGACTCAAGACCTCGCTACCTGCGCA  
ACGCCAACATATATGGTCTAGGCCAACACACAGTTCCTTTCAGGCTACCTACCTCGAACACAACCCGCATA

>147EI16A  
TCATCCAGAATGAAGGCCAGACCTAGAGGAGAATGAAATCATAACAGCAACGGAGGAATCCCCAATACACCGAATACGAGC  
CAATCACAAAACGCGTCGGGCATGGGTGAAAGCGGGGATATTGGAAGAACAACAGAGGAAGTATGGTGCTCAGCAGGGTT  
CACATATCTCAACAACCTAGCAGAGGAGGCTAACCCGCGATAAACCCATTAAAACCTTTGAGGAGATGGTCCCTGAGCAGTA  
CCGAGACTCAAGAAGGTTTTCTCCGAATCTGCCTTGAGCGATTACCTGCCATCAGCCCTGGGATCAGCTATCGATCT  
TGTACCCGGAGCACCAACCATGCGGACAAAGATCTACCAATGTCTTGAACGAGCAGGAGGAACCTCGACCCGCTTCT  
AGAAGAAAACCTACGGAAGGCTACATCGTTCCTCGAAGTCACCCATCTCATCTCCAGTCTTCTTCGTGAAGAAGAAGGA  
TGGGAAACTCCGTTTCGTACAGGACTACCGGAAACTGAATGAGTATACTGTGAAGAACCGATACCCCTCTCCATTAGTTGC  
TGATATCATTAGCCGGCTTCAGGGAGCGGATATTTACCAAGTTTGACGTTTCGCTGGGGCTACAATAATGTGCGGATTA  
GAAGGGGCATGAATGGAAAGGCGCATTGCAACAACCTCGGGGCTTGTTTCGAGCCTAAGGTTATGTTCTTCGGTCTCACGAA  
TTCCCTGCCACATTTCAAGCACTTATGAATGCGATATTTGCTGATCTCATTGCAGCTGGCGAGGTGGCAGTCTACCTAGA  
CGACATCCTCATATTCAGTAACGACCTCAAGGAACCCGACAAGTGGTACGGAAGTATTAACCGGACTAGAAAAGCAGCA  
CCTCTATTTGCGCCAGAAAGCTAGTACTAGTTCGACGCGTGA

>147EI16B  
TCACGCGTCGACTAGTACTAGCTTACAGCTGCACACCTTGAGCGAGGAATGGCAGCCACGGTTGAGGAAAGGCATCTTTC  
GGAGGTTTATATGCGGCTAGATTTTCGTAGACACCTTGACGGAAGAAAGACGACTAAGAAGGGGGCAAATCCAGCAATTAAT  
CGTTCGAGTACCTTTTTGATTCCTTTGTGAGAGTTCCAAGATTCGCGTTTCTTTCCAGGTCATTAGTGGAACAACCTGAGT  
TTCCAGGTCCCTCAATTCGCACATCTGATTGAGAATGATATCCTAGATAAATTTTCGCGAGCCAAAAGGTAAATTTTCGCGG  
TTGAGTTGATTTTACTTTTGGAGGACATGGTAGCATGAGCTTCTAGAATATACTGTAGCAGTCGCAGAGGCCAGTCTGGT  
GAGAGTGAGGATCGCTACAACGGTTCATGGTTAGAAGCAGATTTCCGCAATTAATGGAAGCCGACAGCCGAGAATGATCACA  
TGACGATAAGCAGCTCGGTGCTAGCCAGACTCTCATCAGTCTCAATCCCCGGTTGAATTTACTCTCAACGACTCCTTCTC  
AAGGCCCGTCTGCTGCGCTCCATCATCTGACTTACATCGACCCAAACACCAACCATGCGATAATCCGGGTGCGATAATGC  
GACAAATATAACGGCAGCTGCGACAGTGAACCAGGAATACGGTGCCTTGACCTCATCCATTTGTTGTTTTATGCTCCACT  
CATTTGTTCTGTAGGTGAAGATCACTACGAAGGACGCATATCTCATCGGTAGTCTTGCTGTTATTGACGTTTGGCACATA  
CACTACGGTTGTTCCGTACGTTATTTATCCTATCTAGTTTCATGATACCATTATCTAATCCCTCCCCAATCCCTCCCCATTA

>147EI16C  
TCATCCAGAATGAAGGCCAGACCTAGAGGAGAATGAAATCATAACAGCAACGGAGGAATCCCCAATACACCGAATACGAGC  
CAATCACAAAACGCGTCGGGCATGGGTGAAAGCGGGGATATTGGAAGAACAACAGAGGAAGTATGGTGCTCAGCAGGGTT  
CACATATCTCAACAACCTAGCAGAGGAGGCTAACCCGCGATAAACCCATTAAAACCTTTGAGGAGATGGTCCCTGAGCAGTA  
CCGAGACTCAAGAAGGTTTTCTCCGAATCTGCCTTGAGCGATTACCTGCCATCAGCCCTGGGATCAGCTATCGATCT  
TGTACCCGGAGCACCAACCATGCGGACAAAGATCTACCAATGTCTTGAACGAGCAGGAGGAACCTCGACCCGCTTCT  
AGAAGAAAACCTACGGAAGGCTACATCGTTCCTCGAAGTCACCCATCTCATCTCCAGTCTTCTTCGTGAAGAAGAAGGA  
TGGGAAACTCCGTTTCGTACAGGACTACCGGAAACTGAATGAGTATACTGTGAAGAACCGATACCCCTCTCCATTAGTTGC  
TGATATCATTAGCCGGCTTCAGGGAGCGGATATTTACCAAGTTTGACGTTTCGCTGGGGCTACAATAATGTGCGGATTA  
GAAGGGGCATGAATGGAAAGGCGCATTGCAACAACCTCGGGGCTTGTTTCGAGCCTAAGGTTATGTTCTTCGGTCTCACGAA  
TTCCCTGCCACATTTCAAGCACTTATGAATGCGATATTTGCTGATCTCATTGCAGCTGGCGAGGTGGCAGTCTACCTAGA  
CGACATCCTCATATTCAGTAACGACCTCAAGGAACCCGACAAGTGGTACGGAAGTATTAACCGGACTAGAAAAGCAGCA  
CTAGAAAAGCAGACCTCTATTTGCGCCAGAAAGCTAGTACTAGTTCGACGCGTGA

A continuación se presentan las secuencias intrónicas de los tres posibles *splicings* para la secuencia 147GH16

>Intrón de 204 pb  
TCACGCGTCGACTAGTACTAGCTTACAGCTGCACACCTTGAGCGAGGAATGGCAGCCACGGTTGAGGAAAGGCATCTTTC  
GGAGGTTTATATGCGGCTAGATTTTCGTAGACACCTTGACGGAAGAAAGACGACTAAGAAGGGGGCAAATCCAGCAATTAAT  
CGTTCGAGTACCTTTTTGATTCCTTTGTGAGAGTTCCAAGATTCGCGTTTCTTTCCAGGTCATTAGTGGAACAACCTGAGT

TTCCCAGGTCCCTCAATTTCGCACATCTGATTGAGAAATGATATCCTAGATAAATTCGCGAGCCAAAAGGTAAATTTTCGCCGG  
TTGAGTTTGATTTTACTTTTGGAGGACATGGTAGCATGAGCTTCTAGAATATACTGTAGCAGTCGCAGAGGCCAGTCTGGT  
GAGAGTGAGGATCGCTACAACGGTCAATGGTTAGAAGCAGATTTCCGGCAATTATTGGAAGCCGACAGCCGAGAATGATCACA  
TGACGATAAGCAGCTCGGTGCTAGCCAGACTCTCATCAGTCTCAATCCCCGGTTGAATTTACTCTCAACGACTCCTTCTC  
AAGGCCCGTCTGCTGCCGCTCCATCATCTGACTTACATCGACCCAAACACCAACCATGCGATAATCCGGGTTCGATAATGC  
GACAAATATAACGGCAGCTGCGACAGTGAACCAGGAATACGGTCTGGCCTTCATTCTGGATGCTCGGCACCGACCTCCTC  
TGGCCAAATTCGGCGAAATCGACATCATCGGAGCCATCAACAACCTCGATTCAAACCAATACGCCCTGCACACCACCTGG  
ATGTTTCTAGGCACGAACTCGATACATTCAGGGAGCATTCTTCAAACCTTATTTCTCTACGGGCGCAGGGTGTGTAGTACA  
AGAAAAGAAACCGAGTAGTTTTGGAAGTGGGTTTGGCAGGCTGGCGGAGGTGTGTTTGGCAGCGAGTTGGATCTTAGTGG  
AATCTATGTGTGGTTCCAAGCCTACTTCGTGCAGCTGCTGAA

>Intrón de 154 pb

TCACGCGTCGACTAGTACTAGCTTACAGCTGCACACCTTGAGCGAGGGAATGGCAGCCACGGTTGAGGAAAGGCATCTTTC  
GGAGGTTTATATGCGGCTAGATTTTCGTAGACACCTTGACGGAAGAAAGACGACTAAGAAGGGGGCAAATCCAGCAATTAAT  
CGTTCGAGTCACCTTTTGGATTCCCTTGTGAGAGTTCCAAGATTCGCGTTTCTTTCCAGGTCATTAGTGGACAACCTGAGT  
TTCCCAGGTCCCTCAATTTCGCACATCTGATTGAGAAATGATATCCTAGATAAATTCGCGAGCCAAAAGGTAAATTTTCGCCGG  
TTGAGTTTGATTTTACTTTTGGAGGACATGGTAGCATGAGCTTCTAGAATATACTGTAGCAGTCGCAGAGGCCAGTCTGGT  
GAGAGTGAGGATCGCTACAACGGTCAATGGTTAGAAGCAGATTTCCGGCAATTATTGGAAGCCGACAGCCGAGAATGATCACA  
TGACGATAAGCAGCTCGGTGCTAGCCAGACTCTCATCAGTCTCAATCCCCGGTTGAATTTACTCTCAACGACTCCTTCTC  
AAGGCCCGTCTGCTGCCGCTCCATCATCTGACTTACATCGACCCAAACACCAACCATGCGATAATCCGGGTTCGATAATGC  
GACAAATATAACGGCAGCTGCGACAGTGAACCAGGAATACGGTGCCTTGACCTCATCCATTTGTTGTTTTATGCTCCACT  
CATTTGTTTCTGTCTGGCCTTCATTCTGGATGCTCGGCACCGACCTCCTCTGGCCAAATTCGGCGGAAATTCGACATCATCG  
GAGCCATCAACAACCTCGATTCAAACCAATACGCCCTGCACACCCTGGATGTTTCTAGGCACGAACTCGATACATTCAG  
GGGAGCATTTCAAACCTTATTTCTCTACGGGCGCAGGGTGTGTAGTACAAGAAAAGAAACCGAGTAGTTTTGGAAGTGGG  
TTTGGCAGGCTGGCGAGGTGTGTTTGGCAGCGAGTTGGATCTTAGTGGAATCTATGTGTGGTTCCAAGCCTACTTCGTG  
CAGCTGCTGAA

>Intrón de 69 pb

TCACGCGTCGACTAGTACTAGCTTACAGCTGCACACCTTGAGCGAGGGAATGGCAGCCACGGTTGAGGAAAGGCATCTTTC  
GGAGGTTTATATGCGGCTAGATTTTCGTAGACACCTTGACGGAAGAAAGACGACTAAGAAGGGGGCAAATCCAGCAATTAAT  
CGTTCGAGTCACCTTTTGGATTCCCTTGTGAGAGTTCCAAGATTCGCGTTTCTTTCCAGGTCATTAGTGGACAACCTGAGT  
TTCCCAGGTCCCTCAATTTCGCACATCTGATTGAGAAATGATATCCTAGATAAATTCGCGAGCCAAAAGGTAAATTTTCGCCGG  
TTGAGTTTGATTTTACTTTTGGAGGACATGGTAGCATGAGCTTCTAGAATATACTGTAGCAGTCGCAGAGGCCAGTCTGGT  
GAGAGTGAGGATCGCTACAACGGTCAATGGTTAGAAGCAGATTTCCGGCAATTATTGGAAGCCGACAGCCGAGAATGATCACA  
TGACGATAAGCAGCTCGGTGCTAGCCAGACTCTCATCAGTCTCAATCCCCGGTTGAATTTACTCTCAACGACTCCTTCTC  
AAGGCCCGTCTGCTGCCGCTCCATCATCTGACTTACATCGACCCAAACACCAACCATGCGATAATCCGGGTTCGATAATGC  
GACAAATATAACGGCAGCTGCGACAGTGAACCAGGAATACGGTGCCTTGACCTCATCCATTTGTTGTTTTATGCTCCACT  
CATTTGTTTCTGTAGGTGAAGATCACTACGAAGGACGCATATCTCATCGGTAGTCTTGCTGTTATTGACGTTTTGGCACATA  
CACTACGGTTGTTCCGCTGGCCTTCATTCTGGATGCTCGGCACCGACCTCCTCTGGCCAAATTCGGCGGAAATTCGACATC  
ATCGGAGCCATCAACAACCTCGATTCAAACCAATACGCCCTGCACACCCTGGATGTTTCTAGGCACGAACTCGATACA  
TTCAGGGAGCATTCTTCAAACCTTATTTCTCTACGGGCGCAGGGTGTGTAGTACAAGAAAAGAAACCGAGTAGTTTTGGAAG  
TGGGTTTGGCAGGCTGGCGGAGGTGTGTTTGGCAGCGAGTTGGATCTTAGTGGAATCTATGTGTGGTTCCAAGCCTACTT  
CGTGCAGCTGCTGAA