

Agosto 2015

Tesis de doctorado, PEDECIBA Biología

**OBTENCIÓN DE NUEVAS  
ENZIMAS PARA LA PRODUCCIÓN  
DE BIOCOMBUSTIBLES**

**Inés Loaces**

**Departamento de Bioquímica y Genómica Microbianas**

**Instituto de Investigaciones biológicas “Clemente Estable”**

**Tutores:**

**Dr. Francisco Noya**

**Dra. Elena Fabiano**

**Q.F. Alicia Arias**

**Tribunal:**

**Dra. Claudia Etchebere**

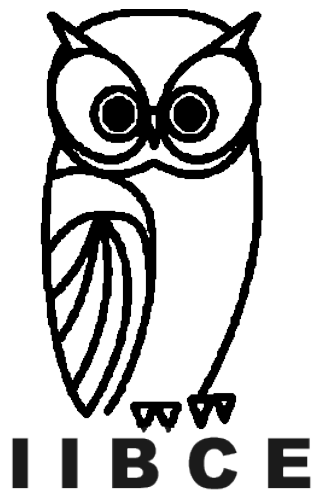
**Dra. Susana Castro**

**Dra. Beatriz Álvarez**

"Alcohol makes a beautiful, clean and efficient fuel... Alcohol can be manufactured from corn stalks, and in fact from almost any vegetable matter capable of fermentation... We need never fear the exhaustion of our present fuel supplies so long as we can produce an annual crop of alcohol to any extent desired."

**Alexander Graham Bell**

National Geographic, Vol. 31, Feb. 1917, p. 131



Instituto de Biotecnología  
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO



**ANII**  
Agencia Nacional de  
Investigación e Innovación



## Agradecimientos

En primer lugar quiero agradecer a Francisco la oportunidad que me brindó de trabajar a su lado, ha sido una experiencia enriquecedora la cual siempre recordaré con cariño y admiración.

Este trabajo no hubiera sido posible sin la colaboración del Dr. Alfredo, el cual me recibió amablemente en su laboratorio dos veces, compartiendo conmigo todo lo que se hoy sobre las fermentaciones y dándome la posibilidad de conocer un hermoso país del cual me enamoré.

Quisiera agradecer al Ing. Nikolai Guchin de ANCAP por proveernos bagacillo de caña de azúcar y la glicerina cruda y por las charlas que nos mostraron el “otro lado” de este proceso.

A las agencias financiadoras de este trabajo: la Agencia nacional de investigación e Innovación (ANII), el Instituto Nacional de Investigaciones Agropecuarias (INIA), el Programa de Ciencias Básicas (PEDECIBA), Inter American Network of Academies of Sciences (IANAS) y la Agencia Mexicana de Cooperación Internacional para el Desarrollo, de la Secretaría de Relaciones Exteriores (AMEXCID-SRE).

Agradezco también al Dr. Lonnie Ingram y al Dr. Ismael Nieves de la Universidad de Florida en Gainesville (EE.UU.) por la cepa LY180 y el plásmido pLOI297 así como recibirme en su laboratorio.

A la gente de BIOGEM, LEM y MICROMOL, especialmente a Vane, Rufo y Rosco que siempre tuvieron alguna solución a la mano para tantos imprevistos. Y a Alicia por la mano enmerme que nos dio!

A mi viejo, que desde el día en que se me ocurrió esta loca idea de hacer Ciencias me acompañó creyendo siempre en mí. A mi familia que escucharon mil veces sin entender demasiado que es lo que hago y supieron comprender mi dedicación al trabajo, sobre todo a mi hermana, Joa y a mis sobrinos Santi, Nahu, Joaco, Joel, Adrián y la pequeña Marley.

Arlen, Kike, Nadia, Luz, Lili... México no sería lo mismo sin ustedes!!!

A mis amigos de la vida, sin los cuales nada sería igual.

Ferchu, Ari... gracias por hacer más amena las jornadas de labo por estar siempre para escuchar, compartir, poner buena onda y fe a todo lo que nos ha pasado. Gracias por ser mis amigas!!! Y principalmente por hacer posible que exista vida “después de las 14 horas” ☺

## Índice

RESUMEN .....	10
Resumen.....	11
INTRODUCCIÓN GENERAL.....	13
Biocombustibles.....	14
Biomasas para la producción de bioetanol.....	15
Biomasa para la producción de biodiesel.....	17
Uso de biocombustibles en el transporte.....	19
Casos particulares .....	20
Desafíos en la producción de biocombustibles .....	24
Biorrefinerías .....	26
Estructura y composición de la biomasa lignocelulósica.....	27
Hemicelulosa.....	27
Lignina .....	29
Celulosa.....	29
Celulasas, hemicelulasas y lacasas.....	34
Hemicelulasas .....	35
Enzimas modificadoras de lignina .....	36
Celulasas .....	36
Producción de etanol a partir de celulosa.....	41
Pretratamiento .....	41
Sacarificación.....	46
Fermentación.....	47
Tecnologías de conversión de biomasa .....	49
Sacarificación y fermentación por separado (SHF) .....	49
Sacarificación y fermentación en simultáneo (SSF) .....	49
Sacarificación y co-fermentación (SScF).....	51
Bioproceso consolidado (CBP).....	52
Bioprospección .....	54
Metagenómica .....	55
Aportes de la metagenómica a la biotecnología.....	58
Identificación de celulasas mediante metagenómica funcional .....	59
Estructura de la tesis .....	61
Plan de trabajo seguido .....	62

Hipótesis .....	63
Objetivo general.....	63
Objetivos específicos .....	63
Capítulo 1.....	64
CONSTRUCCIÓN DE DOS METAGENOTECAS, SCREENING Y SELECCIÓN DE CLONES CON ACTIVIDAD CELULOLÍTICA .....	64
Resumen.....	65
Objetivos .....	65
Metodología .....	66
Cepas y plásmidos.....	66
Obtención de muestras ambientales y procesamiento de ADN metagenómico .....	67
Obtención y procesamiento de ADN metagenómico .....	67
Construcción de bibliotecas metagenómicas .....	68
Evaluación en medios indicadores .....	68
Caracterización fenotípica de los clones seleccionados.....	68
Pretratamiento de bagazo de caña.....	69
Análisis <i>in silico</i> de los fósmidos.....	70
Clonado y expresión de hidrolasas.....	70
Resultados y discusión.....	74
Obtención de ADN y construcción de bibliotecas metagenómicas.....	74
Caracterización fenotípica.....	78
Análisis <i>in silico</i> de las secuencias e identificación de genes .....	81
Clonado, expresión y caracterización de hidrolasas .....	90
Conclusiones .....	93
Capítulo 2.....	94
CARACTERIZACIÓN FUNCIONAL Y ANÁLISIS DEL GENOTIPO DEL CLON M9F1.....	94
Resumen.....	95
Objetivos .....	95
Principales resultados y conclusiones .....	96
Contribución personal al trabajo .....	96
INTRODUCCIÓN .....	97
MATERIALES Y MÉTODOS .....	98
Obtención de muestras .....	98
Obtención de ADN metagenómico .....	98

Construcción de las metagenotecas y selección de actividades enzimáticas de interés .....	99
Pretratamiento de bagazo de caña .....	99
Secuenciación y análisis bioinformático.....	99
Producción y detección de etanol.....	99
Determinación de glucosa y celobiosa.....	99
RESULTADOS.....	100
Construcción de metagenotecas.....	100
Selección de las actividades de interés .....	100
Capacidad de crecimiento en papel de filtro como única fuente de carbono .....	100
Producción de etanol utilizando papel de filtro .....	102
Utilización del bagacillo de caña de azúcar .....	102
Secuenciación y análisis bioinformático del clon M9F1 .....	104
CONCLUSIONES Y RECOMENDACIONES.....	106
AGRADECIMIENTOS .....	108
BIBLIOGRAFÍA .....	108
Ensayos complementarios: Microscopia electrónica de barrido.....	121
Capítulo 3.....	123
LA INCORPORACIÓN DE UN FRAGMENTO DE ADN METAGENÓMICO EN LA CEPA DE <i>Escherichia coli</i> MS04 MEJORA LA PRODUCCIÓN DE ETANOL A PARTIR DE BIOMASA ...	123
Resumen .....	124
Objetivos .....	124
Principales resultados y conclusiones .....	125
Contribución personal al trabajo .....	125
Capítulo 4.....	138
CARACTERIZACIÓN DE UNA CELULASA NOVEDOSA CON POTENCIAL PARA SER UTILIZADA EN LA PRODUCCIÓN INDUSTRIAL DE BIOETANOL .....	138
Resumen .....	139
Objetivos .....	139
Principales resultados y conclusiones .....	139
Contribución personal al trabajo .....	140
Ethanol production from <i>E. coli</i> MS04 expressing a multifunctional halotolerant glucanase isolated from cow rumen.....	141
Abstract.....	141
Introduction.....	142

Materials and methods .....	143
Cloning, expression of <i>endoG</i> and protein purification .....	143
Sequence and phylogenetic analyses .....	143
Enzymatic assays .....	143
pH and temperature effects on rEndoG activity and enzyme thermostability .....	143
Effect of metals and potential inhibitors on EndoG activity .....	144
Determination of substrate specificity .....	144
Nucleotide sequence accession number .....	144
Conversion of cellulose into ethanol.....	145
Results.....	145
Origin and phylogeny of EndoG.....	145
Expression and purification of EndoG.....	145
Enzymatic properties of rEndoG.....	146
EndoG tolerance to industrial inhibitors .....	146
Substrate specificity .....	147
Consolidated bioprocessing .....	147
Discussion .....	147
Acknowledgements.....	148
Figures .....	149
Tables .....	153
Supplementary Figures .....	155
References.....	156
Capítulo 5.....	159
PRODUCCIÓN DE ETANOL A PARTIR DE GLICEROL CRUDO DERIVADO DE LA INDUSTRIA DEL BIODIESEL .....	159
Resumen.....	160
Objetivos .....	160
Principales resultados y conclusiones.....	160
Contribución personal al trabajo.....	161
Ethanol production using <i>E. coli</i> from biodiesel derived glycerol.....	162
Abstract.....	162
Introduction.....	162
Materials and methods .....	163
Metagenomic library construction .....	163



Screening of metagenomic library .....	164
G1 sequence and in silico analysis.....	164
Gene identification by mutagenesis .....	165
Ethanol production assay .....	165
Analyses.....	166
Results and Discussion .....	166
Metagenomic library screening.....	166
G1 sequence analysis .....	167
Ethanol production using Epi300 strains .....	167
Ethanol production using LY180 strain .....	169
Conclusions.....	170
Figures .....	171
Tables .....	175
References.....	178
DISCUSIÓN GENERAL, CONCLUSIONES Y PERSPECTIVAS.....	180
Discusión general y conclusiones .....	181
Perspectivas.....	184
REFERENCIAS BIBLIOGRÁFICAS.....	186
ANEXOS .....	203
Anexo 1.....	204
COMPOSICIÓN DE SOLUCIONES Y MEDIOS DE CULTIVO.....	204
Soluciones.....	204
Medios de cultivo.....	206
Anexo 2.....	208
ABREVIATURAS .....	208
Anexo 3. ....	201
TABLAS DE ORFs ASIGNADOS A LOS NUEVE FÓSMIDOS SECUENCIADOS.....	201
Anexo 4. ....	211
ALINEAMIENTO DE LA SECUENCIA DE AMINOÁCIDOS IDENTIFICADOS COMO RESPONSABLES DE LA ACTIVIDAD CELULOLÍTICA O XILANOLÍTICA EN LOS CLONES SELECCIONADOS .....	211

# RESUMEN

---

## Resumen

Con el fin último de contribuir a diversificar la matriz energética del país, en esta Tesis de Doctorado se planteó el objetivo de identificar y caracterizar elementos genéticos, a partir de una aproximación de metagenómica funcional, que confieran a cepas de *Escherichia coli* la habilidad de degradar desechos agro-industriales para la producción de bioetanol combustible.

Actualmente la producción de etanol carburante se debe principalmente a la utilización de cultivos como caña de azúcar, sorgo dulce y remolacha azucarera. La utilización de desechos agronómicos (rastros o bagazos) y de cultivos energéticos, se vislumbra como la alternativa natural al uso de cultivos convencionales para la producción de biocombustibles. Pero el principal desafío en la producción de etanol de segunda generación es que estas biomásas son altamente recalcitrantes, requiriendo un coctel complejo de enzimas hidrolíticas de celulosa y hemicelulosa, las cuales actualmente son costosas y poco eficientes. Con esto en mente, nos planteamos evaluar dos ambientes ricos en microorganismos degradadores de biomasa vegetal con el fin de obtener nuevas enzimas, eficientes y estables, capaces de ser incorporadas en la cadena de producción de bioetanol.

Se contruyeron dos bibliotecas metagenómicas a partir de contenido ruminal y lodos de un digestor anaerobio, conteniendo 1.2 Gpb y 4.2 Gpb, respectivamente. A partir de éstas se obtuvieron 58 clones con actividad celulasa y/o hemicelulasa. El estudio exhaustivo de 9 clones reveló la capacidad de hidrolizar diversos sustratos, como xilano, avicel, carboximetil-celulosa, papel de filtro y bagazo de caña de azúcar. El estudio *in silico* de la secuencia de los fósmidos de estos clones, mostró que su origen filogenético podría estar relacionado con bacterias del filo Bacteroidetes, pero los organismos parentales no habrían sido cultivados aún.

A nivel funcional, se identificaron más de 10 enzimas implicadas en la degradación de biomasa vegetal, tales como endoglucanasas,  $\beta$ -xilosidasas, arabinofuranasas y lacasas con diversos dominios catalíticos de glicosil-hidrolasas (GH) y dominios de unión a carbohidratos (CBM). El análisis bioinformático de la secuencia de aminoácidos de estas enzimas mostró que la mayoría posee homología con proteínas no caracterizadas previamente, confirmando la utilidad de la herramienta utilizada con el fin de obtener enzimas novedosas.

Teniendo en mente que el proceso más eficiente y económicamente rentable de producción de etanol a partir de biomasa vegetal es aquel en el cual un mismo microorganismo es capaz de expresar las enzimas implicadas en la degradación de celulosa a la vez que es capaz de producir etanol a partir de los azúcares liberados, los fósmidos seleccionados, o las celulasas codificadas en ellos, fueron

incorporados a cepas de *E. coli* productoras de etanol. Esto permitió la utilización de celulosa como fuente de carbono para la producción de etanol. A su vez, se verificó que la carga de enzimas celulolíticas comerciales requeridas en el proceso, es menor cuando se utilizan las cepas complementadas con los fósmidos estudiados.

Actualmente, la producción de biodiesel se ve acompañada por la generación de 10% (p/v) de glicerina cruda, la cual es considerada un desecho industrial. La obtención de productos químicos con valor agregado a partir de este desecho, mejoraría la ecuación económica de la producción de biocombustibles a nivel nacional, siendo una de las alternativas la producción de bioetanol. Con el fin de identificar fósmidos que mejoren la utilización de este sub- producto, las metagenotecas fueron evaluadas en medio mínimo conteniendo glicerina cruda como fuente de carbono y se identificó un clon cuyo crecimiento se vio favorecido. El análisis a nivel genómico de este clon (denominado G1) mostró que posee 38 ORFs, pero no se encontraron enzimas implicados en la captación o metabolización de glicerol. Por mutagénesis *in vitro* se identificó un ORF que codifica para una proteína de membrana que posiblemente esté involucrada en el fenotipo observado. La incorporación de el fósrido G1 en la cepa etanologénica de *E. coli* LY180 permitió la generación de 75 g/L de etanol a partir de glicerina en 12 horas.

Los resultados obtenidos muestran que mediante metagenómica funcional es posible acceder a la diversidad genética microbiana, aumentando el espectro de catalizadores con potencial para ser utilizados a nivel industrial. A complementación de cepas etanologénicas con los fósmidos seleccionados permitió la producción de etanol a partir de desechos agro-industriales.

# **INTRODUCCIÓN**

## **GENERAL**

---

## Biocombustibles

Los biocombustibles son aquellos combustibles que derivan de materiales biológicos, tales como plantas y animales, obtenidos directamente o indirectamente (a partir de residuos agrícolas, comerciales, domésticos, y/o industriales). También se los puede definir como los combustibles creados por la conversión de biomasa (organismos recientemente vivos) mediante procesos térmicos, químicos o biológicos, resultando en combustibles líquidos, sólidos o gaseosos.

Los biocombustibles actuales, biodiesel, bioetanol y biogás, se caracterizan tanto por su capacidad para mezclarse con los combustibles derivados del petróleo, como por la capacidad de ser utilizados en los motores de combustión interna existentes y ser distribuidos a través de la infraestructura actual (1). La producción comercial de biocombustibles líquidos es de casi 90 billones de litros por año a nivel mundial. Por su parte, el biogás, derivado del tratamiento anaeróbico de estiércol y otros desechos, se obtiene en volúmenes relativamente pequeños para ser utilizado en el transporte.

El bioetanol combustible se obtiene mediante fermentación de azúcares simples obtenidos de vegetales ricos en azúcar o en almidón (1) usualmente mediante levaduras (utilizando cepas de *Saccharomyces*) en condiciones moderadas. La energía volumétrica del etanol es aproximadamente 66% que la de la gasolina, sugiriendo una reducción de un tercio en el total de kilómetros por volumen consumido. Pero la comparación de las propiedades físico- químicas del etanol sugieren que la combustión del etanol es 15% más eficiente respecto a las naftas (2).

El biodiesel es una mezcla de ésteres de ácidos grasos de cadena larga, los cuales son sintetizados mediante esterificación o transesterificación de ácidos grasos libres o triglicéridos, respectivamente. Este proceso requiere de la presencia de un alcohol, como metanol o etanol y de un catalizador, siendo los más comunes hidróxido de sodio o de potasio (3). Las materias primas más comunes para la producción de biodiesel incluyen las grasas de origen animal y aceites vegetales. La composición química de biodiesel depende de la materia prima de la que se produce, ya que los aceites vegetales y las grasas animales de diferente origen tienen composiciones de ácidos grasos diferentes (4). La composición de ésteres grasos del biodiesel es esencialmente idéntica a la del aceite o la grasa de partida. Convencionalmente, en el proceso de producción, los lípidos son mezclados con metanol en presencia de hidróxido de sodio, el cual actúa como catalizador del proceso (3). La mezcla es agitada y mantenida a la temperatura de ebullición del metanol (65°C). Cuando se detiene la agitación, la mezcla de reacción se separa en dos fases. La capa superior son los éster de metilo de ácidos grasos, o sea el biodiesel, y la capa inferior es el denominado “glicerol crudo” que contiene el metanol sin reaccionar, las sales y otras sustancias sólidas presentes en las materias primas. El glicerol crudo generalmente

contiene 75% de glicerol, sin embargo, el contenido de glicerol en los residuos de fabricación de biodiesel depende del proceso de fabricación y las materias primas empleadas (3). Cada 10 volúmenes de biodiesel producidos se genera un volumen de glicerol como sub-producto (5). El biodiésel puede mezclarse con gasoil procedente del refinado de petróleo en diferentes cantidades. Se utilizan notaciones abreviadas según el porcentaje por volumen de biodiesel en la mezcla: B100 en caso de utilizar sólo biodiesel, o B5, B15, B30 y B50, donde la numeración indica el porcentaje por volumen de biodiesel en la mezcla.

## **Biomasa para la producción de bioetanol**

Materias primas biológicas ricas en azúcar, o materiales que puedan ser convertidos en azúcares, pueden ser fermentados para producir bioetanol. Las materias primas se pueden clasificar en tres tipos: biomasa rica en sacarosa, como remolacha azucarera, sorgo dulce y caña de azúcar; materiales que contienen almidón (por ejemplo, trigo, maíz y cebada) y la biomasa lignocelulósica, como madera, paja y pastos. Un problema importante con la producción de bioetanol es la disponibilidad de materias primas para su producción. Ésta puede variar considerablemente de una estación a otra y dependen de la ubicación geográfica. El precio de las materias primas también es volátil, lo que puede afectar los costos de la producción de bioetanol (6). Debido a que las materias primas típicamente representan un tercio de los costos de producción, maximizar el rendimiento de bioetanol es imperativo para que el proceso sea económicamente viable.

### **Biomasa azucarada**

Las principales materias primas para la producción de bioetanol son la caña de azúcar y la remolacha azucarera. Dos tercios de la producción mundial de azúcar se obtienen de caña de azúcar y un tercio a partir de remolacha. Estos dos cultivos se producen en regiones geográficamente distintas. La caña de azúcar se cultiva en países tropicales y subtropicales, mientras que la remolacha azucarera solamente se cultiva en países de clima templado (6). La caña de azúcar es un cultivo muy eficiente para la producción de biomasa, representando el cultivo con mayor biomasa por área sembrada de cualquier cultivo comercial importante, incluyendo el maíz. Esto es resultado de la capacidad de la caña de azúcar para incorporar compuestos C3 y C4 en su ruta fotosintética, mientras que la mayoría de plantas sólo incorporan compuestos C3 (2).

En Asia (India, Tailandia, Filipinas) la caña de azúcar se produce en pequeños campos, propiedad de pequeños agricultores. Por ejemplo la India tiene alrededor de 7 millones de pequeños productores con campos de caña de azúcar de alrededor de 0,25 hectáreas en promedio (6). En los países

Europeos, la remolacha azucarera es la biomasa más utilizada. Ésta se cultiva en la mayoría de los países de la UE, y el rendimiento de bioetanol por hectárea es sustancialmente más alto que para el trigo (6). Las ventajas de la remolacha azucarera son un ciclo de producción corto, mayor rendimiento y alta tolerancia de una amplia gama de variaciones climáticas, bajos requerimientos de agua y fertilizantes.

El sorgo dulce (*Sorghum bicolor* L.) es uno de los cultivos agrícolas más resistentes a la sequía, ya que tiene la capacidad de permanecer latente durante los períodos secos. De los muchos cultivos que están siendo investigados para la obtención de la energía, el sorgo dulce es uno de los candidatos más prometedores, sobre todo para la producción de bioetanol en los países en desarrollo (6).

El proceso de obtención de alcohol etílico a partir de caña de azúcar, remolacha o sorgo dulce requiere como primer paso la extracción del jugo rico en azúcares. A continuación los azúcares son fermentados por cepas de levadura, en condiciones leves. Los pasos subsiguientes son la separación de la biomasa del caldo resultante y la concentración del etanol obtenido mediante destilación (1).

### Biomasa amilácea

Otro tipo de materias primas que pueden ser usados para la producción de bioetanol, son los materiales a base de almidón. El almidón está compuesto por dos polisacáridos, la amilosa y la amilopectina. Ambos están formados por unidades de D-glucosa, unidas entre ellas por enlaces  $\alpha$ -1-4 lo que da lugar a una cadena lineal. En el caso de la amilopectina, aparecen ramificaciones mediante enlaces  $\alpha$ -1-6. Este tipo de materia prima es la más utilizada para la producción de bioetanol en América del Norte y Europa, siendo el maíz y el trigo los principales cultivos empleados. En Estados Unidos la materia prima destinada a la producción de etanol es el maíz, utilizándose para este fin un 15% de su cosecha (2).

Para la producción de bioetanol a partir de almidón es necesario romper las cadenas de este hidrato de carbono para la obtención del jarabe de glucosa, que se puede convertir en bioetanol gracias a la fermentación realizada por levaduras. La hidrólisis de almidón por amilasas a temperaturas relativamente altas es un proceso conocido industrialmente como licuefacción. Los factores que afectan la hidrólisis enzimática del almidón incluyen la carga de sustratos, la actividad enzimática y las condiciones de reacción (temperatura, pH, así como otros parámetros) (6). La industria del bioetanol a base de almidón ha sido comercialmente viable desde hace unos 30 años; tiempo en el cual se han obtenido mejoras respecto a la eficiencia de la enzima, lo que reduce los costos y el tiempo de proceso, aumentando los rendimientos de bioetanol (1,6).



## Biomasa lignocelulósica

La biomasa lignocelulósica, como los residuos agrícolas (paja de trigo, bagazo de caña de azúcar, rastrojo de maíz), los residuos forestales (aserrín, restos de raleo), así como de cultivos energéticos (Salix, Switchgrass), es un material atractivo para la producción de combustibles, ya que es el recurso biológico más abundante en la Tierra. Estas materias primas son suficientemente abundantes y generan emisiones netas de gases de efecto invernadero muy bajas, lo que reduce el impacto ambiental (1,6). Por ejemplo, la paja de arroz es uno de los materiales de desecho lignocelulósicos más abundantes en el mundo. Se producen anualmente alrededor de 731 millones de toneladas, que se distribuyen en África (20,9 millones de toneladas), Asia (667,6 millones de toneladas), Europa (3,9 millones de toneladas), América (37,2 millones de toneladas) y Oceanía (1,7 millones de toneladas) (6). Los cultivos perennes lignocelulósicos (por ejemplo, los leñosos de ciclo corto y gramíneas no comestibles) son prometedores como materia prima debido a los altos rendimientos, bajos costos, buena adecuación a tierras de baja calidad y bajo impacto ambiental (1,6).

La biomasa lignocelulósica consta de tres unidades estructurales principales: celulosa, hemicelulosa y lignina. La celulosa es un polímero cristalino de glucosa, la hemicelulosa es un polímero amorfo compuesto principalmente por xilosa y arabinosa, mientras que la lignina es un polímero de compuestos aromáticos. La conversión de biomasa lignocelulósica en alcohol es más difícil en comparación con la conversión de materias primas ricas en almidón. Requiere tres procesos básicos: el pretratamiento de la biomasa, la hidrólisis ácida o enzimática y la fermentación de los azúcares. Debido a las operaciones unitarias requeridas para la conversión de biomasa en combustibles líquidos, el costo de producción de bioetanol a partir de lignocelulosa es relativamente alto cuando se usan las tecnologías actuales. Los principales desafíos son el bajo rendimiento y el alto costo del proceso de hidrólisis enzimática (6,7). Debido a que la materia prima puede representar más del 40% del costo total del proceso, se requiere una transformación económicamente viable de la biomasa en bioetanol, lo cual depende de la conversión rápida y eficiente de todos los azúcares presentes en la celulosa y la hemicelulosa.

## **Biomasa para la producción de biodiesel**

Las materias primas para la producción de biodiesel varían con la ubicación geográfica, según el clima y la disponibilidad. En general, los aceites o grasas de las materias primas más abundantes en una región en particular son las más comunes para la producción local de biodiesel. Un rápido aumento de la capacidad de producción de biodiesel y los mandatos gubernamentales para el uso de combustibles alternativos en todo el mundo en los últimos años han hecho necesaria la implementación de materias primas alternativas para biodiesel. Se pueden diferenciar cuatro grupos de biomasa aptas para la

producción de biodiesel: semillas oleaginosas, grasas animales, algas y diversos materiales de poco valor tales como aceites de cocina usados.

### Oleaginosas

Cultivos de oleaginosas tradicionales son la soja, la colza, la canola y la palma. Por su parte, el maní, las semillas de mostaza, de girasol y de algodón son materias primas consideradas potenciales (4). Los aceites de colza y girasol se utilizan principalmente en Europa, el aceite de palma predomina en los países tropicales, y el aceite de soja se utiliza mayoritariamente en los Estados Unidos (4).

La *Jatropha curcas* (*Jatropha*) no comestible ha atraído recientemente interés como materia prima para la producción de biodiesel en la India y otras regiones climáticamente similares del mundo (4). El árbol de *Jatropha* es una planta perenne oleaginosa (de hasta 5 metros), perteneciente a la familia Euphorbiaceae, sus semillas contienen hasta un 30% de aceite. Éste se puede encontrar en las regiones tropicales y subtropicales de América Central, África, el subcontinente indio y otros países de Asia (4).

### Grasas animales

Las grasas animales pueden incluir materiales de una variedad de animales domesticados, como vacas, gallinas, cerdos y otros animales, así como peces e insectos. Las grasas animales se caracterizan normalmente por un mayor porcentaje de ácidos grasos saturados en comparación con los aceites obtenidos a partir del reino vegetal. Esta biomasa es generalmente considerada como un producto de desecho, por lo que normalmente es menos costosa que los aceites vegetales, lo que la hace atractiva como materia prima para la producción de biodiesel (4).

### Aceites usados

Los aceites usados pueden incluir una variedad de materiales de poco valor tales como aceites de cocina usado o de fritura, grasas para jabones de origen vegetal, aceites ácidos, aceite de resina y otros materiales de desecho. Los aceites usados se caracterizan normalmente por un número relativamente alto de ácidos grasos libres, alto contenido de agua y la presencia de diversos materiales sólidos que deben eliminarse por filtración antes de la conversión a biodiesel. En el caso de los aceites usados de cocina o de fritura la hidrogenación para aumentar la vida útil en la cocción puede resultar en alteraciones de los constituyentes, lo que influye en las propiedades físicas del combustible resultante

(4). Los aceites usados se obtienen principalmente de los restaurantes y pueden costar 60% menos que los aceites vegetales crudos, dependiendo de la fuente y la disponibilidad.

## Algas

El aceite obtenido de microalgas representa otra opción atractiva para la producción de biodiesel, ya que el cultivo de las microalgas requiere menos tierra cultivable que las materias primas de origen vegetal. La productividad de estos microorganismos fotosintéticos para convertir dióxido de carbono en lípidos, es muy superior a la de los cultivos oleaginosos agrícolas. Sin embargo, la producción basada en algas plantea desafíos en lo que respecta a los rendimientos y los procesos de separación, los cuales aún no son económicamente viables (8).

## Uso de biocombustibles en el transporte

El inventor alemán Nicolaus August Otto diseñó el primer motor de automóvil en 1876. El motor de combustión interna de cuatro tiempos que desarrolló utilizaba alcohol como combustible, el cual era abundante y no estaba gravado con impuestos en Europa (9). A su vez, Rudolph Diesel construyó su primer motor en 1900 el cual funcionaba con aceite de maní. El Ford Modelo T, disponible por primera vez en 1908, también fue diseñado para funcionar con etanol, el combustible más utilizado en esa época en los Estados Unidos. Pero el descubrimiento de abundantes pozos de petróleo en Texas en 1901 propició el desarrollo de las refinerías y la implantación de combustibles de origen fósil (9). Mientras, comenzó a gravarse al etanol con fuertes tasas impositivas, lo cual hizo que su utilización se redujera notablemente.

Durante la Segunda Guerra Mundial, se registró la primera gran demanda de biocombustibles, buscando una alternativa para la importación de combustibles de origen fósil en Europa. En este periodo, Alemania fue uno de los países que sufrió una grave escasez de carburantes, propiciando el uso de mezclas de gasolina junto con alcohol obtenido de papa. Gran Bretaña fue el segundo país que utilizó alcohol de granos mezclado con gasolina. Los periodos de guerra fueron decisivos para el impulso de diversas tecnologías que permitían el uso de biocombustibles, pero el descubrimiento de los mayores yacimientos de petróleo del mundo, en la década del 40 en Arabia Saudí y en la del 60 en Siberia provocaron un gran aumento de la oferta de combustibles fósiles con lo que el interés geopolítico y económico en los biocombustibles se desvaneció.

Una grave crisis energética golpeó distintos países durante el período de 1973 y 1979, a causa de la denominada Crisis del Petróleo. La organización de Países Exportadores de Petróleo (OPEP) hizo

recortes importantes de las exportaciones de crudo, especialmente a los países no OPEP (9). La escasez generada de combustible y el aumento de precios atrajo la atención de la academia y los gobiernos a los problemas de la crisis energética y el desarrollo de biocombustibles a gran escala. Así, el siglo XXI llegó con la atención puesta hacia el uso de los biocombustibles. Además de los crecientes precios del crudo, y las crecientes demandas mundiales de energía, las emisiones de gases de efecto invernadero y el interés en el desarrollo rural que acompaña la generación de biocombustibles han estimulado avances en la tecnología y la implementación de los mismos a gran escala.

## Casos particulares

### Brasil

Ya en 1931 el gobierno de Brasil estableció un decreto que obligaba a mezclar 5% de etanol en la nafta importada (10). A partir de los conflictos geopolíticos de 1973, Brasil, en una actitud aislada a nivel internacional, creó el programa "Pró-álcool", y así el etanol recibió atención como biocombustible de gran potencial. El gobierno promovió estudios económicos para la producción de alcohol a gran escala, realizando transferencia de tecnología y proporcionando subsidios a las plantas destiladoras de alcohol.

Mientras, las industrias automovilísticas instaladas en Brasil en aquella época - Volkswagen, Fiat, Ford y General Motors - adaptaron los motores de los vehículos para funcionar con alcohol como combustible. Así surgieron en el mercado dos versiones de vehículos: con motor funcionando con alcohol 100% y la versión convencional de gasolina. El primer automóvil en funcionar con 100% de alcohol fue el Fiat 147, lanzado al mercado en 1978. Desde entonces y hasta 1986, el automóvil de alcohol ganó popularidad entre los brasileños, al punto que la mayoría de los vehículos fabricados por la industria automovilística utilizaban ese combustible. En 1986 la producción llegó a 619.854 automóviles, representando el 76% de todos los vehículos fabricados en ese año (2,10).

En 2004, el gobierno organizó el "Programa Nacional de Produção e Uso do Biodiesel", para fomentar la producción de este biocombustible y su introducción en el mercado interno. En 2005, el Congreso de Brasil emitió la Ley 11.097/2005 en la cual se estableció que a partir del 2013, el diésel tradicional debe contener un mínimo de 5% de biodiesel. También fue emitida la Ley 11.116/2005 estableciendo exenciones de impuestos a las compañías productoras de biodiesel cuya materia prima provenga de granjas familiares en situación de pobreza.

Actualmente, Brasil es el segundo mayor productor de etanol del mundo (28 billones de litros en 2013) y el mayor exportador mundial. En 2014 su producción de biodiesel alcanzó las 3,1 millones

de toneladas métricas. Por esto, es considerado el líder internacional en materia de biocombustibles y la primera economía en alcanzar un uso sostenible de los mismos.

## Estados Unidos

En 1980 el presidente Jimmy Carter creó un programa de incentivo fiscal para la producción de etanol en Estados Unidos mediante la firma de un proyecto de ley que subsidiaba 54 centavos de dólar por galón de etanol y, al mismo tiempo, marcó la exclusión de la importación de etanol desde latinoamericana al mercado estadounidense. Estas dos leyes protegieron la industria del etanol de Estados Unidos desde sus inicios hasta mediados de la década de 1990, cuando creció a más de tres mil millones de litros por año (10). En 1974, el Departamento de Energía fundó el "National Renewable Energy Laboratory" (NREL), el cual comienza a trabajar en 1977. El NREL es el único laboratorio federal dedicado a la investigación, el desarrollo, la comercialización y la implementación de tecnologías de energías renovables y de eficiencia energética.

En el año 2000, fue publicada la Ley "Biomass Research and Development Act" en la que los departamentos gubernamentales de energía y de agricultura tienen como mandato la coordinación a nivel federal de las actividades de investigación y desarrollo relacionados con los biocombustibles y subproductos asociados a la transformación de la biomasa.

En 2005, entró en vigor la Ley de Política Energética (Energy Policy Act, EPAct) cuya sección 1501 estableció como norma sobre combustibles renovables (Renewable Fuel Standard, RFS) la incorporación obligatoria de 7,5 mil millones de galones (28,4 mil millones de litros) de biocombustibles a la gasolina a partir de 2012 (10). Además, esta ley provee una serie de incentivos fiscales a la producción y uso de biocombustibles, entre los cuales están el Incentivo al Pequeño Productor de Etanol (Small Ethanol Producer Tax Credit), el Incentivo al Pequeño Productor de Agro biodiesel (Small Agri-Biodiesel Producer Tax Credit) y un crédito a la compra de equipo surtidor de combustibles alternativos.

El nuevo RFS establece metas altas para la industria de biocombustibles de Estados Unidos. Fija la producción de 136 mil millones de litros de biocombustibles -principalmente de etanol y biodiesel para el año 2022, con 80 mil millones de litros proveniente de los llamados "biocombustibles avanzados", que pueden ser producidos usando una variedad de nuevas materias primas y tecnologías. De esto, se espera que unos 60 mil millones de litros sean de biocombustibles celulósicos.

En 2013 Estados Unidos produjo 13 mil millones de litros de etanol y 4,2 millones de toneladas métricas de biodiesel, lo cual lo posiciona como referente en la producción y desarrollo de biocombustibles.

## Uruguay

Uruguay integra el grupo de países destacados en materia de inversión en energías renovables (11). En 2012, ocupó el primer lugar a nivel mundial con más alto porcentaje del PIB invertido, un 3% (12). En 2013, ocupó el cuarto lugar como el país que atrajo la mayor inversión de América Latina, alrededor de 1,1 mil millones de dólares (11). Mientras que en 2014, fue el país de América Latina con la mayor tasa de crecimiento de las inversiones en energías limpias (11). Pero, de los 1.600 millones de dólares destinados para incorporación de fuentes renovables de energía durante 2013 cerca del 90% se destinó únicamente a la energía eólica.

Uruguay, siendo un pequeño país fuertemente dependiente de combustibles líquidos derivados del petróleo y con amplio potencial agrícola, ha legislado la producción y utilización de biocombustibles. La Ley de Combustibles Alternativos, N° 17.567, de octubre del 2002 (“Combustibles alternativos, renovables y sustitutivos de los derivados del petróleo elaborados con materia prima nacional de origen animal o vegetal”), declaró su producción de interés nacional y la exoneró de los impuestos que gravan a los derivados del petróleo. Posteriormente, la Ley de Agrocombustibles, N° 18.195, de noviembre del 2007, fijó 5% como mínimo obligatorio para la mezcla, a partir del 2012 para biodiesel y en 2015 para bioetanol, con gas oil y gasolinas, respectivamente. Estos mínimos obligatorios establecidos en tan corto plazo estimularon la producción nacional, siendo 2010 el primer año de incorporación de biocombustibles a la matriz energética (13). Su participación, principalmente en el sector transporte, fue de 1% en 2010 y 2% en 2011-2012 y de 3% en 2013. El consumo pasó de 7,0 ktep (miles de toneladas equivalentes de petróleo) a 37,8 ktep en los 4 años, considerando en conjunto el bioetanol y el biodiesel (14).

## Biodiesel

La producción de biodiesel en nuestro país comenzó con diversos emprendimientos de carácter privado que implicaron inversiones de poca magnitud (15). Como la ley limita la cantidad máxima que cada empresa puede producir, para muchos actores privados representa un impedimento de escala para el desarrollo de la actividad (16). A su vez, su viabilidad económica está fuertemente ligada a la evolución del precio de las materias primas, condicionando la permanencia de dichos emprendimientos.

A partir de 2010 ALUR comienza a producir biodiesel en el clúster formado por ANCAP, ALUR, COUSA, CUTCSA y CONAPROLE; asegurando el acceso a la materia prima la cual es producida en un radio de 90 Km de Montevideo incluyendo los departamentos de San José, Florida, Flores, Maldonado, Canelones y Montevideo, fomentando el desarrollo de los pequeños productores (menos de 200 has). Mientras tanto, COUSA aporta servicios de molienda de granos y producción de

aceites (16) y CONAPROLE utiliza las harinas de la molienda como alimento para el ganado de producción lechera. De esta manera, se viabiliza económicamente la producción conjunta de aceites, biocombustible y harinas (15). A partir de 2014, ALUR está recolectando y utilizando para la producción de biodiesel aceites usados generados en diversas cadenas de restaurantes de la capital mientras promueve un plan de recolección domiciliaria en Montevideo, Canelones y Maldonado (17).

Al año 2013, la capacidad instalada de biodiesel fue de 90.000 toneladas/año producido a partir de soja, girasol, colza y sebo (14), teniendo ALUR la capacidad de producir 50.000 toneladas anuales. Como co-producto de la producción de biodiesel, se genera un 10 % (p/p) de glicerina cruda por año, constituido principalmente por glicerol (>80%), agua y ácidos grasos. La mayor parte de la glicerina cruda generada por ALUR (más de 2.000 toneladas/año) es quemada en los hornos de Cementos Artigas como combustible. Además, es utilizada como materia prima en la producción de biogás. Actualmente, ALUR está desarrollando tratativas para la exportación de glicerina cruda a Alemania (17).

## Bioetanol

En 2005 ANCAP, a través de ALUR S.A. adquirió el ingenio azucarero ubicado en Bella Unión, departamento de Artigas, convirtiéndose en el principal productor de etanol carburante. El proyecto de producción de bioetanol está inspirado en el modelo brasilero (ANCAP ha recibido cooperación técnica de Petrobrás) ya que combina una inversión industrial y una agrícola, así como la producción de azúcar, etanol y la generación eléctrica.

Todo el proyecto fue concebido a una escala acorde con la extensión del área agrícola -10 mil hectáreas de caña de azúcar y 4 mil hectáreas de sorgo dulce-, dado que existen restricciones para el incremento de las plantaciones (15,18) . Esto permite la producción de 45 mil toneladas métricas de azúcar y 20.000 m<sup>3</sup> de etanol, lo que representa un 5% de la gasolina consumida (19). En este sentido, para que sea posible producir etanol para mezclar en un porcentaje mayor al 5% establecido por ley, sería necesario producir alcohol en otros lugares del país con distinta materia prima (18) .

ALUR S.A. tiene un segundo proyecto en Paysandú, que incorporará 70.000 m<sup>3</sup> de etanol al año y demandará para su funcionamiento el equivalente a 50.000 hectáreas de granos, en particular sorgo, pero también otros como maíz, trigo y cebada. Además, a partir de vinaza la planta elaborará 50.000 toneladas anuales de alimento animal. El proyecto propone también una planta de cogeneración de energía eléctrica con biomasa como combustible, con una potencia de 12 MW, 7 de los cuales se colocarán en la red pública de UTE (16).

## Desafíos en la producción de biocombustibles

Para ser un sustituto viable de un combustible fósil, un combustible alternativo no sólo debe tener beneficios ambientales superiores sobre los combustibles fósiles, ser económicamente competitivo y ser producible en cantidades suficientes para tener un impacto significativo en la demanda de energía, sino que también debe proporcionar una ganancia neta de energía sobre las fuentes utilizadas para producirlo (20). A su vez, la utilización de biocombustibles debe ser compatible con la infraestructura existente de distribución y con la flota de vehículos, lo que se denominada "drop-in" (21).

Varios trabajos reportan que los biocombustibles de primera generación, además de competir directa o indirectamente con la producción de alimentos, poseen niveles de emisión de gases de efecto invernadero iguales o hasta más altos que los producidos por los combustibles fósiles (22–24). El análisis realizado por Zah et al. (25) resulta muy interesante para visualizar la comparación de diversos biocombustibles con los de origen fósil (Figura 1). En dicho trabajo los autores comparan la gasolina, el diésel y el gas natural con 26 biocombustibles producidos a partir de diversos orígenes. Los autores evalúan el impacto ambiental total de cada combustible mediante la agregación del agotamiento de los recursos naturales y los daños a la salud humana y los ecosistemas en un solo indicador, utilizando dos métodos diferentes<sup>1</sup>.

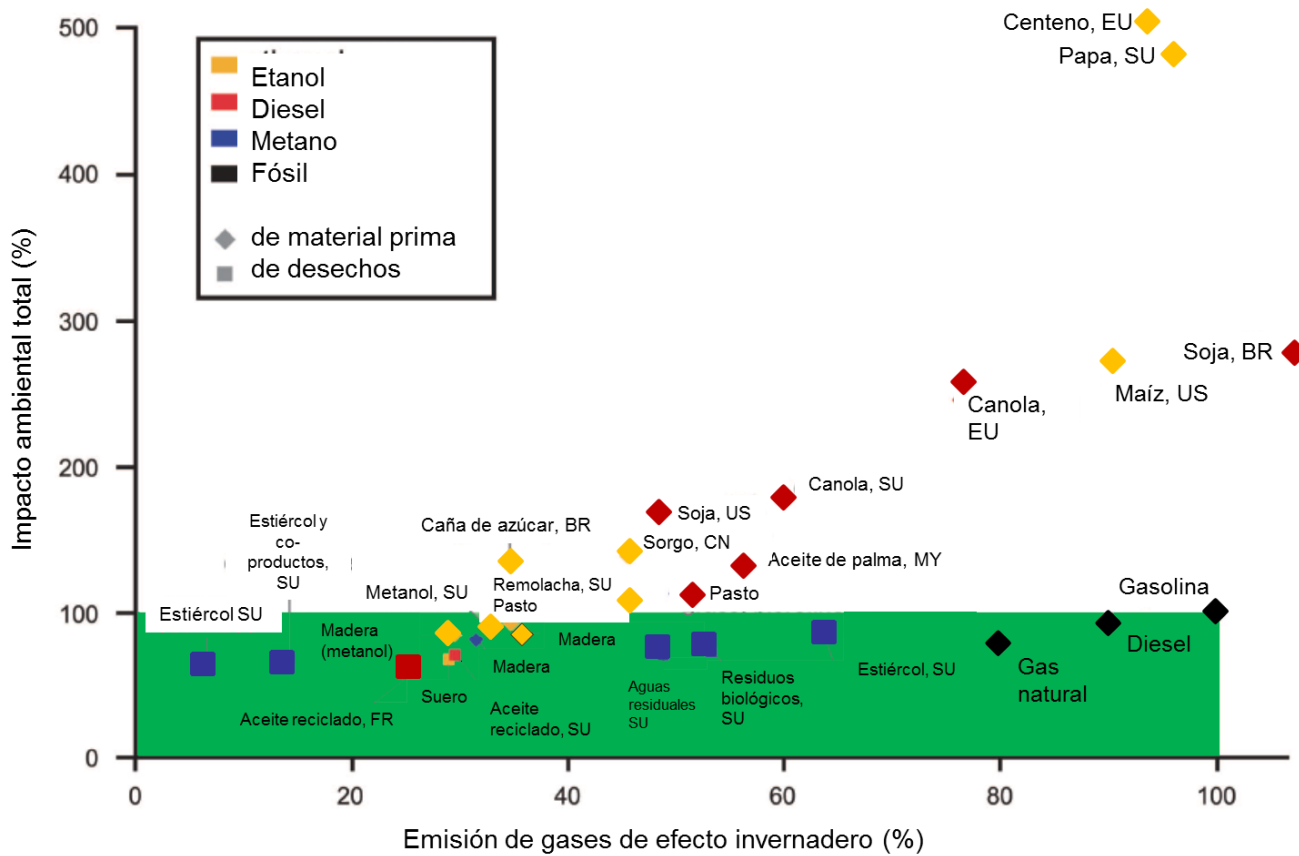
El segundo criterio tenido en cuenta es el nivel de emisiones de gases de efecto invernadero para cada combustible con respecto a la gasolina.

Como se puede apreciar, la mayoría de los biocombustibles (21 de 26) presentan emisiones de gases de efecto invernadero al menos 30% menores a la gasolina. Pero casi la mitad, incluidos los biocombustibles mayormente producidos en la actualidad, tendrían un balance negativo con respecto a los combustibles fósiles (25). A su vez, éstos derivan de cultivos que compiten directamente con la producción de alimentos (22,26). Los biocombustibles que muestran mejores balances son aquellos producidos a partir de desechos, tales como los residuos orgánicos o aceite de cocina reciclado, así como el etanol obtenido a partir de pastos o madera.

---

<sup>1</sup> Los autores utilizan cómo métodos los indicadores suizos y los europeos. Los parámetros suizos de impacto ambiental, miden cuánto excede el daño ambiental los límites legales (ver [www.esu-services.ch/download/Frischknecht-2006-EcologicalScarcity-Paper.pdf](http://www.esu-services.ch/download/Frischknecht-2006-EcologicalScarcity-Paper.pdf)); mientras que los eco-indicadores europeos, cuantifican el daño a la salud humana y a los ecosistemas (ver [www.pre.nl/eco-indicator99/default.htm](http://www.pre.nl/eco-indicator99/default.htm)).





**Figura 1.** Emisiones de gases de efecto invernadero versus impacto ambiental total de 28 combustibles, relativos a la gasolina. El origen de los biocombustibles se indica: Brasil (BR), China (CN), Unión Europea (EU), Malasia (MY) y Suiza (SU). Los combustibles en la zona sombreada son considerados ventajosos para ambos parámetros. Modificado de (22).

Las materias primas no alimenticias ofrecen ventajas tanto en criterios energéticos, ambientales y económicos. Los biocombustibles de segunda generación, por ejemplo el etanol o el metanol producido a partir de lignocelulosa, son energéticamente más eficientes y más flexibles en cuanto al origen de la materia prima. La posibilidad de utilizar biomasa celulósica heterogénea sugiere menores costos y un mejor desempeño ambiental (27). Posibles materias primas son los desechos agropecuarios, como los rastrojos de maíz, trigo o arroz y el bagazo de la caña de azúcar (28). A su vez, tanto el Switchgrass (*Panicum virgatum*), diversos pastos y malezas o plantas leñosas con potencial como materia prima para la generación de hidrocarburos combustibles, se pueden producir en tierras agrícolas marginales sin o con baja aplicación de fertilizantes, pesticidas, y consumo de energía. Además de la producción de etanol celulósico, la combustión de los residuos de estas biomásas, como la fracción de lignina, podría generar energía para alimentar las plantas de producción de los biocombustibles (20).

## **Biorrefinerías**

El remplazo de los químicos derivados del petróleo por los provenientes de biomasa representa un factor clave en el crecimiento de la industria química. La biomasa vegetal está compuesta por una variedad de compuestos como celulosa, hemicelulosa, lignina, aceites, almidón y proteínas. Cada constituyente puede ser transformado en alimentos, combustibles, productos agro-industriales, fármacos o diversos productos con valor agregado. Según el NREL, una biorrefinería integra los procesos de conversión y los equipamientos necesarios para producir combustibles, químicos y energía a partir de biomasa. El objetivo entonces, es la transformación de la biomasa en diversos productos valiosos mediante la aplicación de procesos eficientes y amigables con el medio ambiente (29). Esto requiere del desarrollo de tecnologías específicas para la valorización de cada fracción, las que pueden implicar transformaciones químicas, enzimáticas y/o microbiológicas (30). Para la implementación de biorrefinerías tecnológicamente sostenibles, toda la cadena de producción de la biomasa (la siembra, el cultivo y la cosecha), el pretratamiento y la conversión a productos deben ser considerados. El contenido relativamente bajo de energía, la estacionalidad y la disponibilidad geográfica discreta de las materias primas se han propuesto como barreras a los grandes volúmenes demandados para la generación de energía y de combustibles (30). De todos modos, la aplicación de las biorrefinerías se ha planteado como un medio para aumentar la viabilidad económica de la industria de los biocombustibles (31), sobre todo por la co-generación de productos de alto valor. Una alternativa, que valoriza aún más la biorrefinería, es la utilización de sub-productos o residuos generados durante la producción del biocombustible. Por ejemplo, los residuos ricos en lignina pueden ser quemados en la caldera o convertidos en productos con alto valor (6,32). La producción de biodiesel puede verse beneficiada por la utilización de la fracción rica en glicerol como punto de partida para la producción de 1,3-propanodiol, etanol (5,33), hidrógeno (3) y aminoácidos (34), entre otros productos con valor agregado.

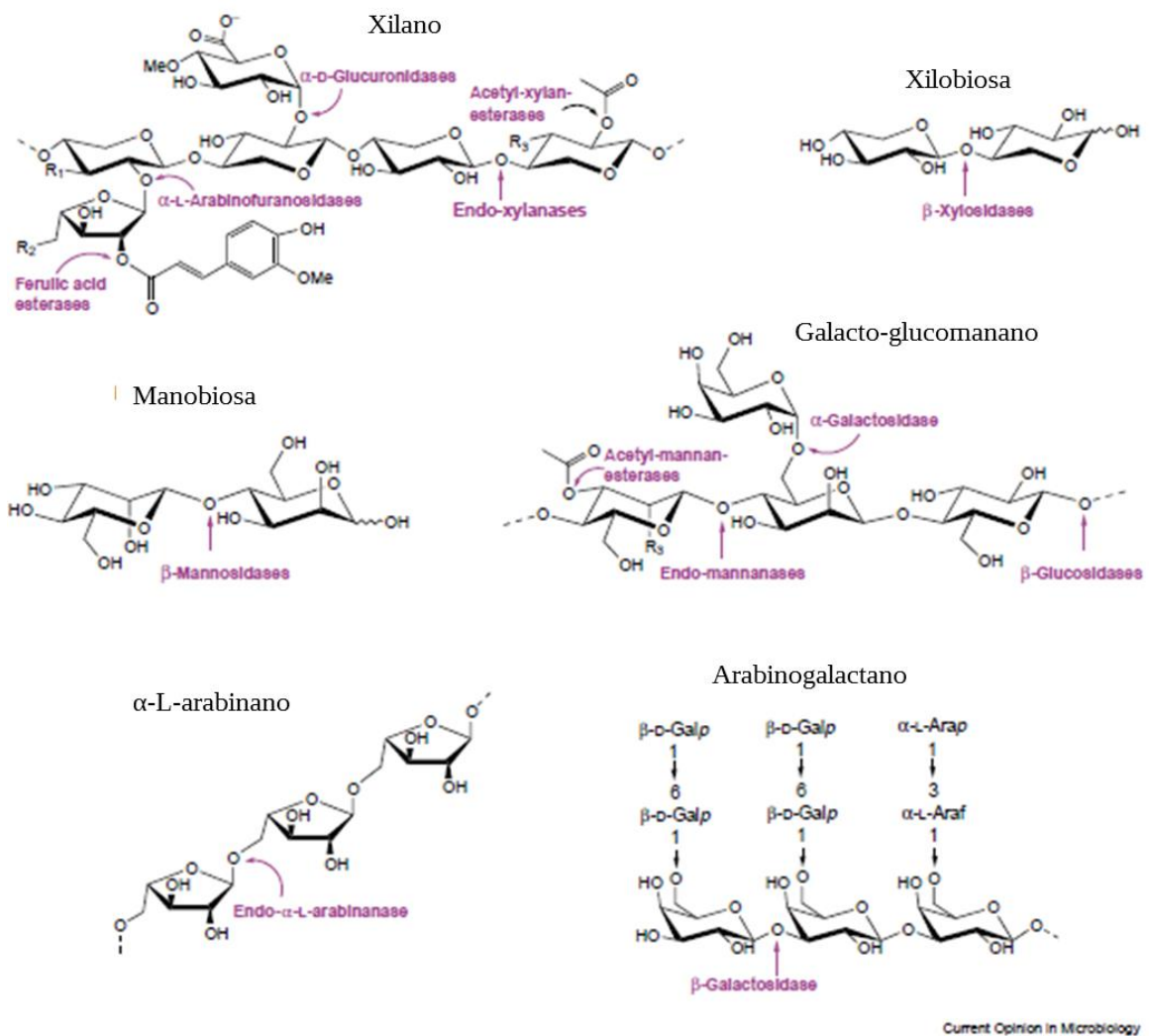
## Estructura y composición de la biomasa lignocelulósica

La lignocelulosa es un material orgánico renovable y el mayor componente de todas las paredes celulares vegetales. Está formado por tres componentes principales: la celulosa (40-50%), la hemicelulosa (20-40%) y la lignina (20-30%) formando una estructura compleja. Los componentes menores de las paredes vegetales son las pectinas, las cuales forman un gel de polisacáridos altamente ramificado de ácido galacturónico y sus metil-ésteres (35,36), siendo más abundantes en frutas y en tejidos de plantas jóvenes.

### Hemicelulosa

La hemicelulosa constituye entre un 20 y un 40% de la biomasa vegetal. Es un heteropolímero compuesto por pentosas ( $\beta$ -D-xilosa,  $\alpha$ -L-arabinosa), hexosas ( $\beta$ -D-manosa,  $\beta$ -D-glucosa,  $\alpha$ -D-galactosa) y ácidos urónicos (ácidos  $\alpha$ -D-glucurónico,  $\alpha$ -D-4-O-metilgalacturónico) (37). Otros azúcares como  $\alpha$ -L-ramnosa y  $\alpha$ -L-fucosa pueden estar presentes en pequeñas cantidades y los grupos hidroxilo de los azúcares pueden estar parcialmente substituidos por grupos acetilo (38). Los componentes de la hemicelulosa más relevantes son los xilanos y los glucomananos, siendo los primeros los más abundantes (Figura 2). El xilano es un heteropolisacárido complejo con un esqueleto de  $\beta$ -1,4-xilosa rica en cadenas laterales de arabinosa, mientras que los glucomananos consisten en una columna vertebral lineal de unidades  $\beta$ -D-glucopiranosidos y  $\beta$ -D-manopiranosidos, unidos por enlaces glucosídicos  $\beta$ - (1,4), parcialmente acetilados en C2 o C3 y substituidos por unidades de  $\alpha$ -D-galactopiranosido unidas a la glucosa y manosa mediante enlaces  $\alpha$ - (1,6).

Los xilanos son los componentes principales de la hemicelulosa de las paredes celulares secundarias de la biomasa de las maderas duras y las plantas herbáceas, alcanzando hasta un 50% en algunos tejidos de cereales y pastos. La hemicelulosa del tipo mananos, como glucomananos y galactoglucomananos son los componentes mayoritarios de la hemicelulosa de las paredes secundarias de las maderas blandas, estando presentes en maderas duras en cantidades menores (38).



**Figura 2.** Estructura de los componentes principales de la hemicelulosa: xilano, xilobiosa, manobiosa, glucomananos y arabinanos. Se indican las hemicelulasas responsables de la degradación de cada tipo de componente y los enlaces sobre los que actúan. Modificado de (39).

## Lignina

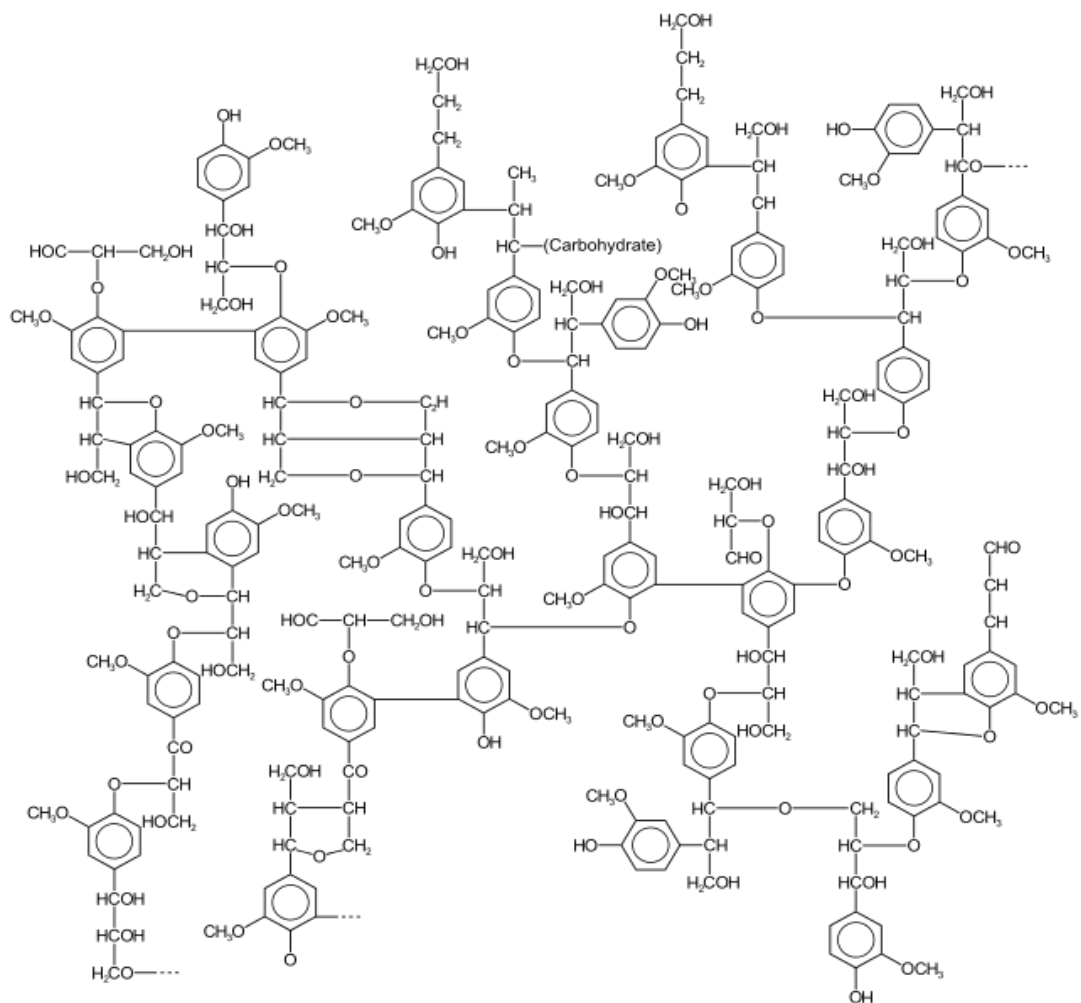
La lignina es el tercer componente principal de las plantas vasculares, constituyendo entre un 15 y un 25% de la pared celular. En general, las ligninas son co-polímeros que derivan principalmente de tres unidades de fenilpropano monoméricas (monolignoles) básicas: alcohol p-cumarílico, alcohol coniferílico y alcohol sinapílico (40) (Figura 3). Los tres principales monolignoles (Figura 4) se forman en el citoplasma, teniendo a la fenilalanina como intermediario clave. Los monolignoles se generan mediante reacciones de desaminación, hidroxilación, reducción y metilación catalizadas por diversas enzimas. Estos monolignoles interactúan en la pared celular, a través de reacciones de oxidación catalizadas por peroxidasas (intermedios radicalarios) para formar finalmente los polímeros de lignina (41).

Los porcentajes de distribución de los principales monolignoles dependen del tipo de planta. El monolignol más abundante en las maderas blandas es el alcohol coniferílico, que puede llegar a superar el 95% del total de monolignoles presentes; mientras que en las maderas duras coexisten fundamentalmente los alcoholes coniferílico y sinapílico. En el caso de plantas del tipo herbáceas, puede haber proporciones similares de los tres monolignoles principales (40).

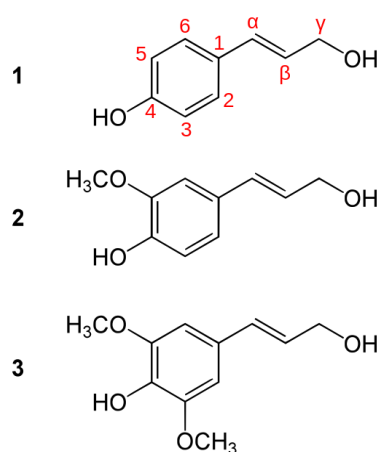
La lignina es necesaria ya que aporta soporte mecánico a tallos y hojas y da rigidez y fuerza a las paredes vegetales. Existe evidencia que muestra que la lignina junto con otros componentes de la pared celular proveen resistencia a enfermedades, insectos, bajas temperaturas y otras formas de estrés biótico o abiótico (42). Se ha reportado que no solo un aumento en el contenido de lignina repercute en una menor degradabilidad de los tejidos vegetales, sino que también están implicados la composición molecular de la lignina, su hidrofobicidad y las interacciones con otros componentes de la matriz (43).

## Celulosa

Más allá de diferencias en la composición y la estructura de las variedades vegetales, la celulosa representa entre un 40 y un 50% del peso seco vegetal. En algunos casos, como en las cápsulas de algodón, está casi en estado puro, pero en la mayoría, las fibras de celulosa están formando parte de una matriz más compleja (7) (Figura 5).



**Figura 3.** Modelo parcial de la estructura de lignina de la madera. Tomado de (44).

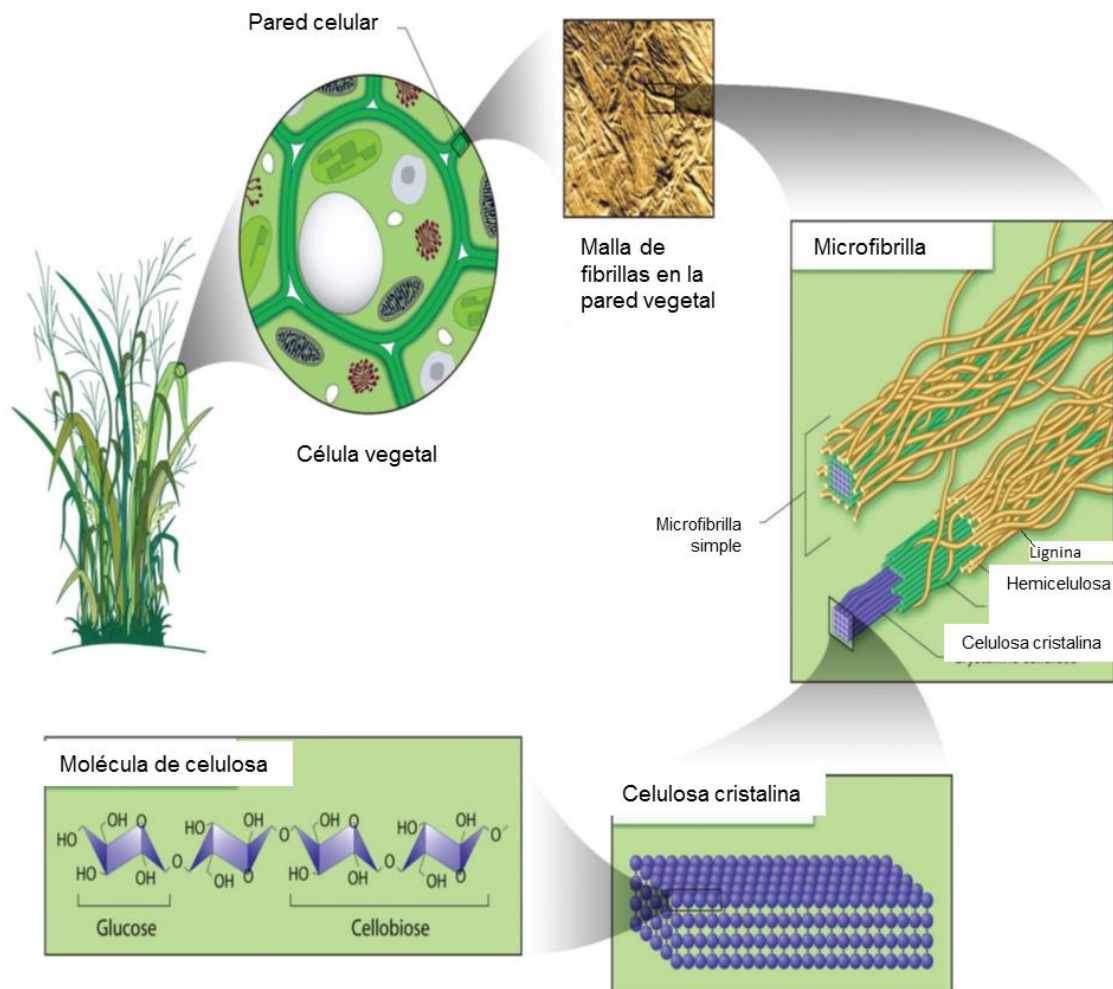


**Figura 4.** Tipos de monolignoles más comunes encontrados en los bloques de la lignina: 1: alcohol p-cumarílico, 2: alcohol coniferílico, 3: alcohol sinapílico.

Una característica importante de la celulosa es su estructura cristalina. Este polisacárido es sintetizado como cadenas lineales de residuos de D-glucosa unidos por enlaces  $\beta$  (1-4), los cuales se auto-ensamblan en el sitio de biosíntesis (45). En las plantas, la síntesis es llevada a cabo por las proteínas sintetasas de celulosa (CESA), las cuales están embebidas en la membrana plasmática (46). Para la formación de las fibrillas elementales o "protofibrillas" 36 moléculas de celulosa individuales se asocian, las cuales que se agrupan en unidades más grandes llamadas "microfibrillas", los cuales a su vez se ensamblan formando las fibras de celulosa. Las cadenas de celulosa se asocian mediante puentes de hidrógeno inter- e intra- catenarios (Figura 6). Hojas adyacentes se superponen entre sí y se unen mediante fuerzas de Van der Waals. A pesar de la debilidad de este tipo de interacciones, su efecto total sobre los residuos en la fibrilla elemental es considerable, asegurando que todos los átomos estén en posiciones discretas con respecto a otro, lo cual define la naturaleza cristalina de la celulosa. Una característica importante de la matriz de celulosa es que las moléculas de las microfibrillas individuales se asocian firmemente, lo cual evita la penetración incluso de pequeñas moléculas tales como agua (7). Pero en la naturaleza las fibras de celulosa no son puramente cristalinas. Existen regiones amorfas y diversos tipos de irregularidades, como giros o pliegues en las microfibrillas e incluso microporos superficiales y capilares (43). Esta heterogeneidad en la estructura hace que las fibras estén parcialmente hidratadas y que algunas macromoléculas puedan acceder a los microporos y los capilares.

La estructura y las interacciones de los componentes de la matriz varían con el tipo de pared celular y la madurez del cultivo (42) y representan una gran limitante para la degradación de la biomasa vegetal. A su vez, la incorporación de hemicelulosa durante la formación de las paredes de las células vegetales es uno de los determinantes en los patrones de agregación de la celulosa, actuando como moderador de la estructura terciaria de la celulosa en la pared celular (47).

El denominado índice de cristalinidad (IC) se utiliza para describir la cantidad relativa de material cristalino en la celulosa. Generalmente, el IC está en el rango de 40 a 95% en materiales naturales, siendo el resto celulosa amorfa (48). El grado de polimerización (DP) es el número de unidades monoméricas en una molécula de polímero, el cual para la celulosa varía de 500 a 15.000 dependiendo del material (Tabla 1). A su vez, dos grupos terminales diferentes se encuentran al final de cada cadena de celulosa. En un extremo un grupo no-reductor, o sea con el anillo cerrado, está presente. Un grupo reductor, con un grupo aldehído se encuentra en el otro extremo (Figura 6). Esto hace que la fibra de celulosa sea una molécula polarizada, donde nuevos residuos de glucosa pueden añadirse al extremo no-reductor permitiendo la elongación de la cadena (49).



**Figura 5.** Esquema de la estructura molecular de las paredes celulares vegetales. Se detallan la estructura cristalina de la celulosa y la interacción con la hemicelulosa y la lignina (microfibrillas). Las moléculas de celulosa están compuestas por unidades de glucosa unidas mediante enlaces  $\beta$ -1,4. Fuente: Office of Biological and Environmental Research of the U.S. Department of Energy Office of Science. [science.energy.gov/ber/](http://science.energy.gov/ber/).

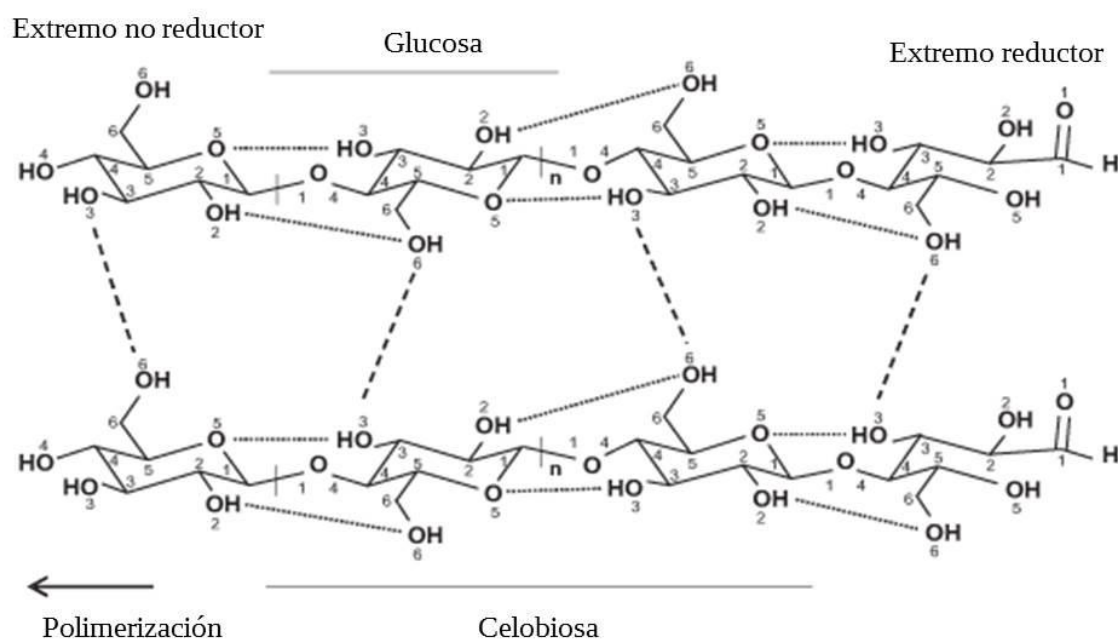


**Tabla 1.** Propiedades físicas de algunos sustratos celulósicos

Sustrato <sup>1</sup>	Índice de cristalinidad	Grado de polimerización
Carboximetil-celulosa (CMC)	NA	100-2000
Cellodextrinas	NA	2—60
Avicel	0.5-0.6	300
CB	0.76-0.95	2000
PASC	0-0.04	100
Algodón	0.81-0.95	1000-3000
Papel de filtro	0-0.45	750
Pulpa de madera	0.5-0.7	500-1500
Avicel PH-101 Fluka	0.56-0.91	200-400*
Celulosa Fluka	0.48-0.82	280*
$\alpha$ -celulosa Sigma	0.64	2140-2420*

<sup>1</sup> CB, celulosa bacteriana; PASC, avicel tratado con ácido sulfúrico

\*. Según datos del fabricante. NA, no aplicable.



**Figura 6.** Estructura de los puentes de hidrógeno inter- e intra-catenarios en la celulosa. Las líneas discontinuas representan los enlaces inter-catenarios, mientras que las líneas punteadas a los enlaces intra-catenarios. El extremo reductor y no- reductor están indicados. Modificado de (46).

## Celulasas, hemicelulasas y lacasas

Más allá de su naturaleza recalcitrante la biomasa lignocelulósica no se acumula debido a la acción de hongos y bacterias celulolíticas y lignolíticas que lentamente degradan los componentes de las paredes celulares vegetales. Tanto las bacterias como los hongos expresan enzimas como lacasas, hemicelulasas y celulasas, las que degradan eficientemente lignina, hemicelulosa y celulosa, respectivamente. Otro grupo de proteínas, con actividad de disrupción de celulosa (expansinas, "swollenins" y "loosenins") tienen la capacidad de relajar las paredes celulares por disrupción de los puentes de hidrógeno que unen las fibrillas de celulosa entre sí y con los otros polisacáridos (50–52).

Según la clasificación de enzimas de la IUBMB (International Union of Biochemistry and Molecular Biology) la cual se basa en la especificidad de sustrato, las enzimas celulolíticas se agrupan con las hemicelulasas y otras polisacaridasas en O-glicosil hidrolasas (EC 3.2.1.x). Esta clasificación ha resultado ineficiente debido a la diversidad estructural y al gran número de sustratos que poseen estas enzimas, por lo que surgió una clasificación alternativa (53), en la cual las glicosil hidrolasas (GH) se agrupan en familias definidas en base a su similitud en la secuencia de aminoácidos (54). Esta clasificación en familias es complementaria a la determinada por la IUBMB, porque refleja las características estructurales de la enzima, y porque el rango de sustratos rara vez es determinado en forma completa. La base de datos CAZy (Carbohydrate-Active enZYmes, <http://www.cazy.org/>) describe los dominios catalíticos y de unión a carbohidratos (CBM) de las enzimas implicadas en la degradación, modificación o creación de enlaces glicosídicos. Es así que se encuentran, además de las GH, las glicosil transferasas (GT), las polisacárido liasas (PL), las carbohidrato esterasas (CE) y diversas enzimas auxiliares (AA), como las enzimas degradadoras de lignina y las mono-oxigenasas líticas de polisacáridos (LPMO, lytic polysaccharide mono-oxygenases) (36).

Actualmente existen 133 familias de GH, agrupadas en 14 clanes (definidos en base a similitud en la estructura tridimensional). Algunas familias contienen enzimas con diferente especificidad de sustrato, por ejemplo la familia GH5 incluye enzimas capaces de degradar celulosa, xilanos, arabinoglucanos, mananos, liquenanos, entre muchos otros sustratos. Esto sugiere una evolución divergente de una estructura básica del sitio activo para acomodar diferentes sustratos. Al mismo tiempo, se encuentran endoglucanasas en varias familias (GH 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74 y 124) sugiriendo evolución convergente de diferentes estructuras capaces de interactuar con el mismo sustrato.

En la mayoría de las celulasas y hemicelulasas libres la unión a los sustratos está mediada por módulos de unión a carbohidratos (CBM, carbohydrate-binding module), los que han sido clasificados en 39 familias CBM (<http://afmb.cnrs-mrs.fr/CAZY/>) de acuerdo a su homología de secuencia de

aminoácidos. La unión a la superficie de celulosa mediante el CBM presumiblemente facilita la hidrólisis de la celulosa, ya que permite la proximidad del dominio catalítico al sustrato. La presencia de los CBM es particularmente importante para la iniciación y la procesividad de las exoglucanasas. A nivel funcional, los CBM han sido divididos en tres grandes categorías: tipo A, que se unen a la superficie de polisacáridos insolubles, CBM tipo B, que se unen a cadenas de glucanos solubles, y los tipo C, los cuales unen pequeños sacáridos (55).

## Hemicelulasas

Debido a la variedad de componentes de la hemicelulosa y su compleja organización, su degradación requiere de la acción de varias enzimas. Las hemicelulasas son un grupo de diversas proteínas modulares con dominios catalíticos y accesorios, estos últimos principalmente involucrados en la interacción con el sustrato o con otras enzimas. Los dominios catalíticos son o bien glicosil hidrolasas (GH), los cuales hidrolizan enlaces glicosídicos, o carbohidrato esterasas (CE), los que hidrolizan las uniones éster de los grupos laterales acetato o de ácido ferúlico (39).

Las xilanasas (EC 3.2.1.8) hidrolizan los enlaces  $\beta$ -1,4 en la cadena del xilano, generando xilooligómeros. Aunque la mayoría de las xilanasas conocidas pertenecen a las familias GH 10 y 11, existen representantes de estas enzimas en las familias GH 5, 8, 9, 12, 16, 26, 30, 43, 44, 51 y 62.

Las  $\beta$ -mananasas (EC 3.2.1.78) hidrolizan la hemicelulosa de tipo manano, liberando oligómeros cortos de  $\beta$ -1,4-mananos, los que luego pueden ser degradados a manosa por  $\beta$ -manosidasas (EC 3.2.1.25). Hay unas 50 secuencias de  $\beta$ -mananasas en las familias GH 5 y GH 26, y unas 15  $\beta$ -manosidasas en las familias GH 1, 2 y 5.

Las  $\alpha$ -L-arabinofuranosidasas (EC 3.2.1.55) y las  $\alpha$ -L-arabinanasas (EC 3.2.1.99) degradan la fracción de hemicelulosa compuesta por arabinofuranósidos, los xilooligómeros y los arabinanos, y se encuentran en las familias 3, 43, 51, 54 y 62. A su vez, las  $\alpha$ -D-glucuronidasas rompen los enlaces 1,2-glicosídicos del ácido 4-O-metil-D-glucorónico presente en las cadenas laterales de los xilanos. Estas enzimas están exclusivamente presentes en las familias de glicosil hidrolasas 4 y 67.

Las  $\beta$ -xilosidasas (EC 3.2.1.37) son exoglicosidasas que hidrolizan xilooligómeros de cadena corta en unidades de xilosa. Están agrupadas en las GH 1, 3, 5, 30, 39, 43, 51, 52, 54, 116 y 120. La similitud espacial entre la D-xilopiranososa y L-arabinofuranosa conduce a que existan enzimas bifuncionales con actividad xilosidasa y arabinosidasa, las que se encuentran principalmente en las familias 3, 43 y 54.

Las esterasas hemicelulolíticas incluyen las xilano esterasas (EC 3.1.1.72) que rompen las sustituciones acetil en los motivos xilosa, y las ferulil esterasas (EC 3.1.1.73), que hidrolizan los enlaces éster entre la arabinosa y el ácido ferúlico. Este tipo de enlace está implicado en el entrecruzamiento entre el xilano y la lignina.

## **Enzimas modificadoras de lignina**

Las ligninasas son un conjunto de enzimas oxidativas, principalmente de origen fúngico, que catalizan la descomposición de la lignina (56). El término correcto para este grupo de proteínas es enzimas modificadoras de lignina (LME), ya que estas enzimas no son hidrolíticas, siendo su mecanismo enzimático de tipo oxidativo. Las LMEs incluyen peroxidasas, como las peroxidasas de lignina (EC 1.11.1.14), y manganeso peroxidasa (EC 1.11.1.13) y muchas de las fenol-oxidasas de tipo lacasa. Las lacasas (benzenediol /oxígeno oxidoreductasas, EC 1.10.3.2) catalizan la reducción del oxígeno molecular a agua con la oxidación de una amplia variedad de compuestos tanto fenólicos como no fenólicos, incluyendo hidrocarburos aromáticos policíclicos, pesticidas, colorantes industriales y alquenos (57).

## **Celulasas**

Existen dos grandes tipos de sistemas de degradación de celulosa microbianos. En uno, el organismo produce un conjunto de enzimas que son secretadas al espacio extracelular, y actúan de forma sinérgica. En el segundo grupo, las enzimas hidrolíticas se organizan en un complejo que se localiza en la superficie celular, los denominados celulosomas.

La degradación eficiente de la celulosa requiere de la acción conjunta de tres tipos de enzimas (Figura 7): endoglucanasas (1,4- $\beta$ -D-glucan-4-glucanohidrolasas, EC 3.2.1.4), exoglucanasas (4- $\beta$ -D-glucan glucanohidrolasas, también llamadas celodextrinasas (EC 3.2.1.74) y 1,4- $\beta$ -D-glucan celobiohidrolasas o celobiohidrolasas (EC 3.2.1.91)) y las  $\beta$ -glucosidasas ( $\beta$ -glucósido glucohidrolasas, EC 3.2.1.21), las cuales actúan sinérgicamente para hidrolizar la celulosa. Se ha propuesto como modelo de acción que las endoglucanasas atacan las regiones amorfas de las fibras de celulosa generando extremos en la cadena de celulosa. Las exoglucanasas degradan de forma procesiva las cadenas de celulosa, tanto desde los extremos reductores como desde los no reductores, liberando celobiosa (celobiohidrolasas) o glucosa (glucanohidrolasas). Estas últimas son capaces también de hidrolizar las regiones cristalinas de la celulosa (7). Las unidades de celobiosa y las celodextrinas de cadena corta son convertidas a glucosa por la acción de las  $\beta$ -glucosidasas (Figura 7).

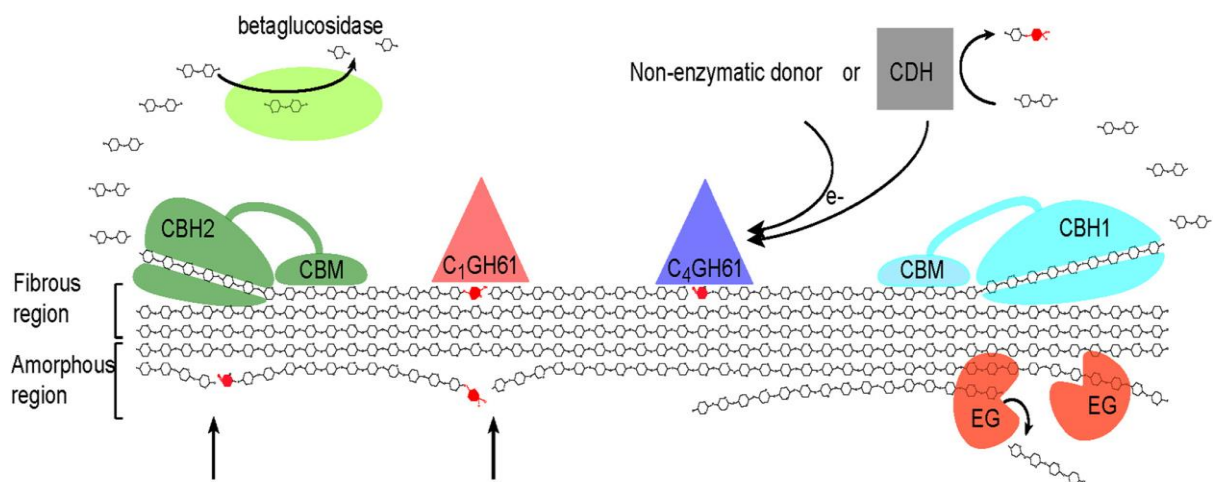
Se han descrito endoglucanasas en las familias GH 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74 y 124. Por su parte, las celodextrinasas se agrupan en las familias GH 1, 3, 5 y 9 y celobiohidrolasas en las GH 5, 6 y 9. Las  $\beta$ -glucosidasas forman parte de las GH 1, 3, 5, 9, 30 y 116.

Se ha demostrado sinergismo entre diversas celulasas y proteínas auxiliares (AA) de la familia AA10 (antes CBM33) y la familia AA9 (antes GH61) (58). Estos dos tipos de enzimas son monooxigenasas dependientes de cobre capaces de romper las cadenas de celulosa generando nuevos extremos (36,59). Las enzimas pertenecientes a la familia AA9 pueden oxidar celulosa en presencia de un agente reductor e incrementar significativamente la eficiencia hidrolítica de celulasas fúngicas (60,61). Estas enzimas se han identificado en el genoma de diversos hongos degradadores de madera (36,60). Por su parte, las enzimas de la familia AA10 logran aumentar la accesibilidad de enzimas hidrolíticas al sustrato, actuando sinérgicamente con quitinasas y celulasas (62). Las proteínas AA10 no están restringidas a un reino específico, pero están predominantemente en las bacterias (36).

## Celulosomas

Los celulosomas son complejos multi-enzimáticos capaces de degradar la celulosa de forma muy eficiente (63) (Figura 8). Estos son producidos principalmente por bacterias anaerobias, principalmente por especies de *Clostridium*, *Ruminococcus*, *Acetivibrio* y *Bacteroides*. Se ha reportado que varios hongos anaerobios son capaces de secretar complejos tipo-celulosoma (64). Una característica central de los celulosomas es la presencia de una sub-unidad no catalítica de andamiaje (o "scaffolding") (65), encargado de varias funciones, como la incorporación de enzimas al complejo por interacciones específicas cohesina-doquerina, el anclaje del complejo a la membrana celular y la orientación del complejo hacia el sustrato celulósico, mediante dominios de unión a celulosa (CBM) (64,66). Los módulos de cohesina son proteínas cuya función es la de unirse a módulos de doquerina presentes en unidades catalíticas u otras unidades de scaffolding. La interacción cohesina-doquerina es fundamental para la formación del celulosoma.

Hasta la fecha, todos los módulos de unión a la celulosa de los "scaffolding" pertenecen a la familia CBM3 (66). Este tipo de CBM se une a la superficie de la celulosa cristalina, reteniendo el celulosoma cerca del sustrato. A su vez, para que los diferentes componentes del celulosoma interactúen con los diferentes sustratos, las enzimas celulosomales poseen varios tipos de CBMs de diferentes familias, los cuales exhiben especificidad para los diferentes sustratos (65). Cuando el celulosoma está unido a la superficie celular, el mismo CBM guía a la célula completa hacia el sustrato (65). El avance en el conocimiento de estos complejos ha evidenciado el hecho de que la presencia del CBM no es una característica intrínseca de los celulosomas, ya que puede no estar presente (65).

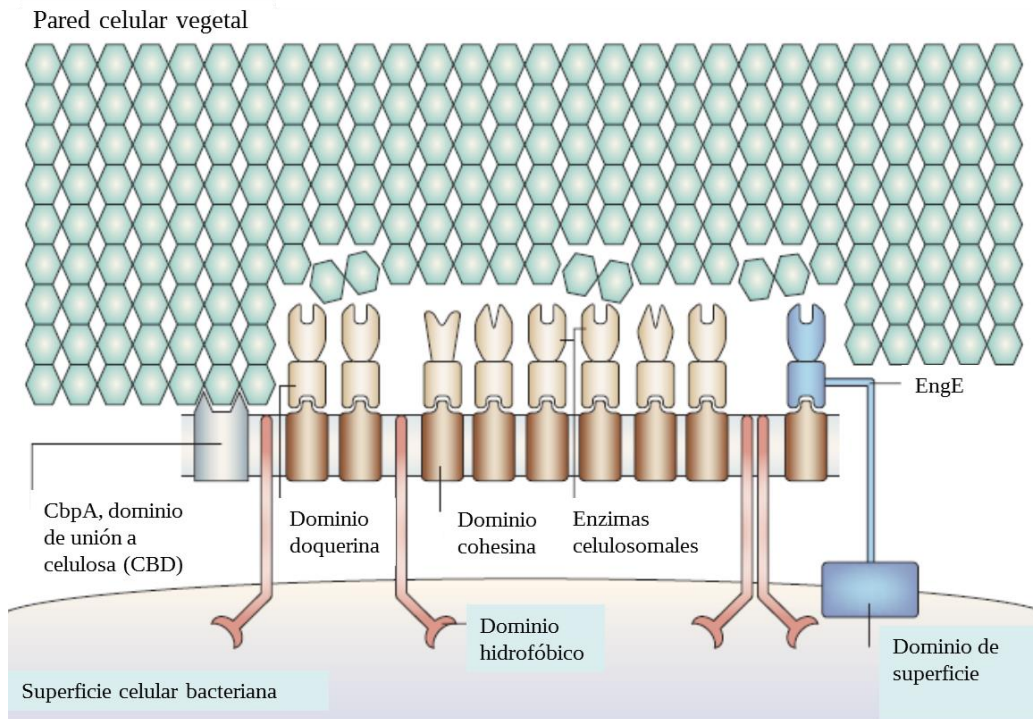


**Figura 7.** Esquema del modo de acción de los tres tipos de celulasas: EG: endoglucanasas, CBH: exocelulasa y  $\beta$ -glucosidasas. Las EG son capaces de hidrolizar los enlaces  $\beta$  (1-4) internos de las fibrillas de celulosa, generando extremos reductores y no reductores. Las CBH liberan unidades de celobiosa a partir de dichos extremos. Por su parte, las  $\beta$ -glucosidasas hidrolizan las unidades de celobiosa o celodextrinas generando glucosa. Las proteínas auxiliares (GH61 y CDH) oxidan residuos de glucosa internos a las fibras de celulosa o celobiosa. CMB: Dominio de unión a celulosa, CDH: Celobiosa deshidrogenasa, GH61: Polisacárido monooxigenasa. Las flechas muestran nuevos puntos posibles de ataque de las CBHs.

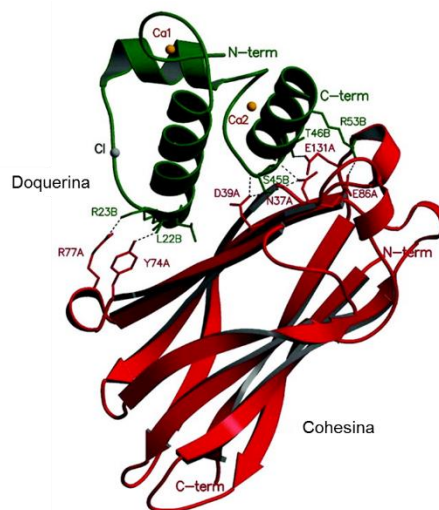
Las interacciones cohesina-doquerina son de las más potentes del tipo proteína-proteína en la naturaleza. La interacción entre estos dos componentes puede verse como una especie de arreglo "plug-and-socket", donde la doquerina encaja en el módulo cohesina (Figura 9). La secuencia y estructura de estos módulos determinan la interacción cohesina- doquerina. Esta interacción es la que establece las posibles conformaciones que puede adoptar el complejo enzimático.

Tanto las enzimas libres (no celulosomales) como las que forman el complejo enzimático, poseen dominios catalíticos similares, frecuentemente pertenecientes a las mismas familias de glicosil hidrolasas. La diferencia entre ellas es que las enzimas libres suelen poseer un módulo de unión a sustrato, que en las enzimas celulosomales es sustituido por un módulo doquerina (65). El espectro de enzimas asociadas a los celulosomas incluye numerosas endo- y exoglucanasas, muchas de las cuales suelen ser procesivas. Otras enzimas degradadoras de polisacáridos, como hemicelulasas (67,68), quitinasas (69) y liasas de pectina (70) también se pueden asociar al celulosoma. Más allá de las diferencias en los sistemas celulolíticos de los hongos filamentosos y las bacterias, las celulasas celulosomales de las familias GH 5, 9 y 48 son especialmente prevalentes, aunque existen miembros que contienen los módulos de doquerina en al menos 20 familias adicionales (65). Por otro lado, no se han descrito celulasas de las familias GH 6 o GH 7 en ninguno de estos sistemas.

El desarrollo y utilización de celulosomas artificiales fue sugerido por Bayer (71) y su primer demostración fue realizada en 2002 (72). En un celulosoma diseñado, cada enzima está equipada con un módulo doquerina que interacciona específicamente con un módulo cohesina presente en el scaffoldin quimérico. Éstos últimos sirven de esqueleto de acoplamiento y contienen un módulo de unión a carbohidratos específico para el sustrato que se quiere degradar y uno o más módulos cohesina a los que se unen las diferentes unidades celulosomales (73). Los celulosomas secretados por *Clostridium cellulolyticum* han servido de base para el desarrollo de celulosomas artificiales o quimeras (74).



**Figura 8.** Modelo esquemático del celulosoma de *Clostridium cellulovorans*. La proteína “scaffoldin” se muestra con el dominio de unión a celulosa (CBD), las cohesinas, cuatro dominios hidrofóbicos y nueve enzimas celulosomales unidas al scaffoldin a través de las doquerinas, EngE traba el celulosoma a la superficie celular (64).



**Figura 9.** Estructura cristalina tridimensional del complejo cohesina-doquerina tipo I de *C. thermocellum*. El complejo está formado entre una molécula de cohesina tipo 2 (en rojo) y una doquerina unida a  $\text{Ca}^{2+}$  (en verde). Los residuos que participan en los contactos de dominio se muestran como modelos de palo. Los dos iones de  $\text{Ca}^{2+}$  asociados al dominio doquerina se representan como esferas de color naranja. Tomado de (75).



## Producción de etanol a partir de celulosa

El esquema convencional de producción de etanol de segunda generación consiste en cuatro procesos individuales: el pretratamiento de la biomasa, la hidrólisis enzimática (denominada sacarificación), la fermentación de los azúcares y la purificación del producto, generalmente por destilación del etanol o filtrado molecular (Figura 9).

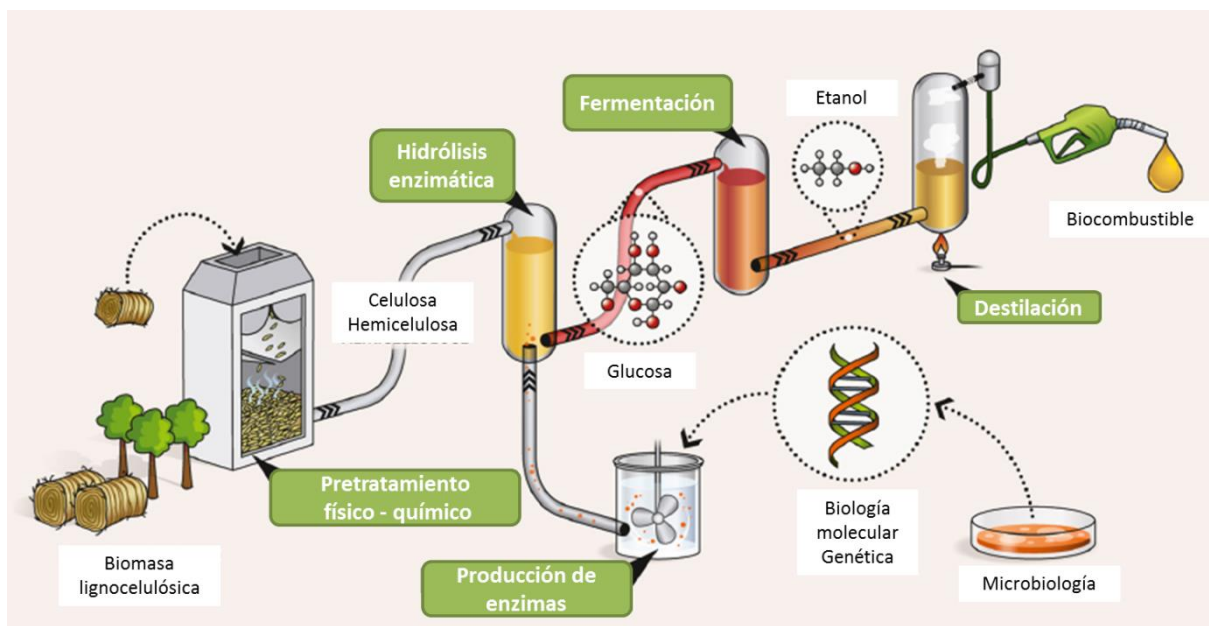
### Pretratamiento

Un pretratamiento adecuado implica la ruptura de la matriz de hemicelulosa y lignina, de los puentes de hidrógeno que mantienen unida las fibrillas de celulosa, el aumento en la porosidad y el área superficial de la celulosa (76) con bajos requerimientos de energía y el reciclado de los catalizadores utilizados (77). Durante el pretratamiento se intenta evitar la degradación completa de la lignina, debido a que es utilizada para la generación de otros productos con valor agregado, y la formación de compuestos inhibidores de la fermentación, como son: vanilina, acetato, siringaldehído, hidroxibenaldehído, hidroximetilfurfural y furfural (78). Los pretratamientos suelen clasificarse en físicos, químicos o biológicos. Los más aplicados actualmente suelen involucrar una combinación entre componentes físicos y químicos, aunque el desarrollo de nuevas metodologías con mejoras en los rendimientos y disminución de los costos está en pleno auge (76,77).

#### Métodos físicos

El principal objetivo de los pretratamientos físicos es la disminución del tamaño de las partículas de lignocelulosa y a su vez, disminuir el grado de interacción entre los componentes de la matriz. Por lo general la biomasa suele molerse, reduciendo el tamaño de las partículas a 0,2 –30 mm dependiendo del mecanismo utilizado. Aunque este proceso suele realizarse previo a los siguientes tratamientos de la biomasa, algunos trabajos lo proponen como único paso previo a la sacarificación enzimática (79).

El pretratamiento de extrusión es un pretratamiento termo-físico en el que los materiales se someten a mezcla, calentamiento y disrupción que conduce a la alteración física y química de su estructura (80).



**Figura 9.** Esquema de producción de etanol a partir de biomasa lignocelulósica. El proceso implica 4 grades etapas: el pretratamiento de la biomasa, la hidrólisis enzimática de la fracción sólida rica en celulosa y hemicelulosa, la fermentación de los azúcares generados en etanol y la purificación del alcohol mediante destilación. La etapa de sacarificación requiere de enzimas hidrolíticas las cuales son producidas de manera exógena por microorganismos. El mejoramiento genético de las cepas utilizadas y de las enzimas utilizadas puede contribuir a mejorar los rendimientos del proceso.

Por su parte, la irradiación con microondas ha sido evaluada como un posible pretratamiento, ya que podría ser una alternativa a los tratamientos térmicos convencionales con el fin de alterar la ultraestructura de la celulosa, principalmente a través de la colisión molecular. Este proceso, comparado con el calentamiento convencional, requiere menos tiempo y energía y suele ser más uniforme (81).

Recientemente se desarrolló un método novedoso para el pretratamiento físico mediante la congelación de la biomasa y se encontró que aumenta significativamente la digestibilidad enzimática de rastrojo de arroz. A pesar de los altos costos involucrados y que sólo unos pocos estudios se han llevado a cabo utilizando este pretratamiento hasta el momento, parece tener algunas características positivas las cuales han atraído una gran atención (81).

### Tratamientos químicos

El pretratamiento ácido, en particular el uso de ácido sulfúrico, es el más común de los tratamientos químicos. En éste, la hemicelulosa es hidrolizada a monosacáridos, permitiendo el acceso a la celulosa. El pretratamiento ácido puede realizarse a bajas concentraciones de ácido con altas temperaturas o a altas concentraciones de ácido con bajas temperaturas (76). Las principales desventajas asociadas son la corrosión de los equipos, la recuperación del ácido y la formación de diversos inhibidores, cuya naturaleza depende de la combinación de concentración y temperatura aplicada a cada material.

El pretratamiento alcalino se lleva a cabo a más baja temperatura y presión que la hidrólisis ácida. Cuando un material lignocelulósico se empapa en una solución alcalina, por ejemplo, hidróxido de sodio, de potasio o de amonio, se rompen los enlaces entre la lignina y los polisacáridos y también se puede solubilizar parte de la lignina. El efecto de esto es un aumento en el área de superficie disponible, mientras que el grado de polimerización se reduce. En general, los métodos de pretratamiento utilizando álcalis son eficaces en ciertos tipos de materias primas, por ejemplo, sobre los residuos agrícolas y cultivos herbáceos. La razón es probablemente la menor cantidad relativa de lignina en estos tipos de materiales (77).

Los líquidos iónicos (IL) han ganado mucha atención en los últimos años ya que pueden disolver los carbohidratos y la lignina simultáneamente. Los IL son sales (punto de fusión  $< 100^{\circ}\text{C}$ ) compuestas por un catión orgánico y un anión pequeño inorgánico que son líquidos a temperatura ambiente (82). El modo de acción de los IL es la interrupción de los enlaces entre la celulosa, la hemicelulosa y la lignina, lo cual disuelve las macromoléculas sin la formación de inhibidores que afecten a los microorganismos fermentadores. Aunque la utilización de estos "solventes verdes"

presenta muchas ventajas, el costo de los mismos y su recuperación están entre sus limitantes. Por otro lado, la búsqueda de enzimas hidrolíticas compatibles con los IL es un área de investigación activa (83).

Los solventes orgánicos, como metanol, etanol, acetona y etilenglicol, con o sin la adición de un agente catalizador pueden ser utilizados en el proceso denominado organosolv. Los catalizadores utilizados incluyen ácidos (clorhídrico y sulfúrico) o bases (hidróxido de sodio, amoníaco). Este pretratamiento es capaz de romper los enlaces internos entre la lignina y la hemicelulosa, siendo especialmente eficiente para biomasa con alto contenido de lignina. Por otra parte, lignina de alta calidad se puede obtener de manera relativamente pura como un subproducto del proceso (84). Obviamente, la eliminación de lignina conduce a una mayor área de superficie en la celulosa, haciéndola más accesible a la degradación enzimática. Los principales inconvenientes de este método son el bajo punto de ebullición de los solventes orgánicos, el alto riesgo de operación a alta presión, la inflamabilidad y la volatilidad de tales disolventes. Además, los solventes deben ser reciclados con el fin de disminuir el costo de operación y prevenir sus efectos inhibidores sobre la hidrólisis enzimática y los microorganismos (76).

### Pretratamiento físico-químico

Uno de los métodos más utilizados es el pretratamiento con vapor de agua, llamado comúnmente "explosión de vapor". En este proceso la descomposición de los componentes estructurales se debe a la aplicación de calor en forma de vapor de agua, lo cual genera la hidrólisis de enlaces debido a la expansión de la humedad. Típicamente, el material lignocelulósico es tratado a temperaturas entre 160-240°C durante 1 a 20 minutos (77). Antes del tratamiento con vapor, se puede o no impregnar el material con algún tipo de catalizador. El vapor generado condensa debido a la alta presión, impregnando el material. Cuando la presión dentro del reactor se libera rápidamente, las fibras se abren, rompiendo los enlaces inter e intra moleculares (38). Se ha demostrado que el principal mecanismo es bastante similar a la hidrólisis con ácido diluido, aunque realizado con mayor carga de sólidos y en vapor en lugar de en un líquido caliente. La mayoría de los residuos agrícolas y de árboles de madera dura liberan ácidos orgánicos (principalmente ácido acético) cuando son tratados térmicamente en medio acuoso, causando la auto-hidrólisis de la hemicelulosa. Ésta suele ser más bien leve, ya que el pH de partida es casi neutro y el pH al final del proceso es de aproximadamente 3,5-4 dependiendo de la cantidad de unidades de ácidos liberadas. Sin embargo, para algunos materiales, tales como madera blanda, la cantidad de ácidos orgánicos intrínsecos no es suficiente para causar la degradación de la estructura principal de la hemicelulosa. La adición de un ácido mineral, por ejemplo,  $H_2SO_4$ ,  $HCl$  o  $H_3PO_4$ , reduce el pH inicial por debajo de 2, lo que resulta en una hidrólisis más eficiente de la hemicelulosa, viéndose

también favorecida la hidrólisis enzimática subsiguiente (77). Sin embargo, tal como es el caso del pretratamiento ácido, la degradación de los hidratos de carbono y parte de la lignina puede resultar en la formación de compuestos tóxicos, que pueden afectar a la hidrólisis enzimática y la fermentación.

En el pretratamiento de explosión de fibras por amonio (AFEX, ammonia fiber explosion), la biomasa se trata con amoníaco líquido bajo el concepto del proceso de explosión de vapor. Cuatro parámetros incluyendo la carga de amoníaco, la carga de agua, la temperatura de reacción y el tiempo de residencia se pueden variar con el fin de optimizar este pretratamiento. El proceso incluye alta presión (1,72 a 2,06 Mpa) y temperaturas moderadas (60-120°C) durante varios minutos (<30 minutos) seguido de una liberación repentina de presión. La rápida expansión del gas amoníaco provoca la ruptura del complejo lignina-carbohidratos y la consiguiente ruptura física de fibras de la biomasa (76).

Al contrario que la explosión de vapor, que produce una suspensión, en el AFEX, debido a la baja temperatura de ebullición del amoníaco, produce solo material sólido. Tampoco se liberan azúcares directamente debido a la baja solubilización de la hemicelulosa, pero abre la estructura de la biomasa y aumenta la superficie de los polímeros y, en consecuencia, la digestibilidad enzimática.

Al igual que con otros métodos alcalinos el pretratamiento AFEX funciona mejor en materiales con bajo contenido de lignina, tales como residuos agrícolas, mientras que el efecto sobre la madera blanda es menos potente. El principal inconveniente de este método es que el amoníaco debe ser reciclado y manejado con precaución para que el proceso sea amigable con el medio ambiente y rentable desde un punto de vista económico (77).

### Métodos biológicos

El pretratamiento de biomasa lignocelulósica por microorganismos, especialmente hongos, se trata básicamente de un pretratamiento amigable con el medio ambiente. En contraste con la mayoría de los métodos de pretratamiento que requieren alto capital y costo operativo, este método aprovecha los hongos blancos, marrones y de podredumbre blanda para deslignificar y mejorar la hidrólisis enzimática de la biomasa lignocelulósica. La eficiencia más alta entre los métodos de pretratamiento biológicos se ha logrado mediante el uso de hongos de la podredumbre blanca degradadores de lignina, destacando entre los más eficientes *Phanerochaete chrysosporium* debido a su alta tasa de crecimiento y la capacidad de biodegradación de la lignina (76). La tasa de conversión del pretratamiento biológico es mucho menor comparado con la mayoría de los otros métodos de pretratamiento, lo cual durante muchos años ha sido considerado como un factor que impide su uso a nivel industrial (77).

## Sacarificación

La sacarificación es el proceso de hidrolizar un carbohidrato complejo (como el almidón o la celulosa) en los monosacáridos que lo componen. Los agentes empleados en la sacarificación son químicos o enzimáticos o bien una combinación de los dos. Entre los primeros está la utilización de ácidos diluidos, particularmente el HCl. Entre los procedimientos enzimáticos están la utilización de amilasas o celulasas de origen microbiano.

Las diversas enzimas implicadas en la hidrólisis de los polisacáridos son producidas a nivel comercial por cepas modificadas genéticamente de *Trichoderma reesei*, *Fusarium venenatum*, *Aspergillus oryzae* y *Aspergillus niger*. A una carga de 20% de sólidos en esta etapa, la dosis usual de enzimas suele ser de 10 FPU/g celulosa (FPU, unidades de papel de filtro), lo que equivale a 20 mg de enzima/ g de celulosa (85). Obviamente, la dosis ideal de enzimas depende de la biomasa, la carga de sólidos y el pretratamiento utilizado. Más allá de las variables intrínsecas a esta etapa, se considera al costo de las enzimas, y de esta etapa en particular, un cuello de botella en el proceso de producción de etanol celulósico. El análisis realizado por Klein y colaboradores (85) muestra que, en general, la gran mayoría de la literatura ha subestimado la contribución del costo de las enzimas en la producción de biocombustibles. Según los autores, el costo de las mismas se puede bajar mediante un cambio en las materias primas, la reducción de los tiempos de fermentación y la reducción de la complejidad del proceso. Si bien la mejora de la actividad enzimática ha sido el centro de diversas investigaciones científicas, estas mejoras contribuyen a la reducción de los tiempos de residencia de sacarificación, pero no hacen frente a los costos de producción de la enzima.

Diversas estrategias se están llevando a cabo para mejorar la eficiencia de las celulasas y mejorar los rendimientos de producción de los microorganismos que secretan estas enzimas. Algunas estrategias que pueden reducir el costo del proceso radican en la inmovilización de las enzimas o de las células productoras (78). También el reciclaje de las enzimas se plantea como una alternativa, aunque esta estrategia está limitada por la adsorción de las enzimas a la lignina. Se ha probado que el agregado de detergentes o surfactantes reduce considerable la adhesión de las enzimas a la lignina (86).

El diseño de enzimas con mejor actividad y estabilidad es otra alternativa que busca hacer más eficiente la sacarificación (87). El diseño racional implica modificaciones específicas de las enzimas, y requiere un conocimiento detallado de la relación entre las secuencias de aminoácidos y la actividad enzimática. Desafortunadamente, esta relación entre la secuencia y la función sigue siendo un enigma para la gran mayoría de las enzimas, por lo que a menudo se prefiere la evolución dirigida. En particular, los enfoques más difíciles de diseño racionales son aquellos que implican la modificación de aminoácidos específicos ya que es muy difícil predecir cómo pequeños cambios moleculares afectan a

la función deseada. Muchas de las propiedades enzimáticas no se limitan a uno o unos pocos aminoácidos ya que por lo general se distribuyen a través de una proteína.

La evolución dirigida es una técnica utilizada para detectar mutantes de las enzimas con características más favorables. Esta estrategia es más efectiva cuando la enzima de partida tiene propiedades interesantes, para que mediante unas pocas mutaciones, se puedan obtener enzimas con mejoras sustanciales (78). La mutagénesis aleatoria, la estrategia más básica evolución dirigida, es una manera de descubrir enzimas mejoradas a través de mutaciones al azar. La mutagénesis dirigida puede ser más eficaz porque son seleccionados unos pocos residuos clave para la mutación. La técnica también ha sido una herramienta fundamental en la caracterización de celulasas, especialmente en la identificación de los residuos catalíticos y de unión (78). Alternativamente, la recombinación genética mediante la combinación aleatoria de ADN puede ser utilizada para combinar enzimas estructuralmente similares, a veces resultando en un mayor grado de alteraciones estructurales. Para cada técnica, es fundamental establecer una conexión directa entre el rendimiento de la enzima durante la selección y el desempeño real en la materia prima.

## **Fermentación**

Tanto el pretratamiento como la hidrólisis están diseñados para liberar azúcares que puedan ser fermentados en el proceso de producción de alcohol. Las levaduras industriales, como *Saccharomyces cerevisiae* han sido utilizadas en la producción de alcohol principalmente en las industrias del vino y la cerveza durante cientos de años, siendo también los organismos más utilizados en la producción de biocombustibles derivados de azúcar o almidón (86). *S. cerevisiae* fermenta hexosas mediante la vía Embden-Meyerhof en condiciones anaeróbicas. Las fermentaciones basadas en levaduras siempre conllevan la formación de CO<sub>2</sub> como co-producto y la necesidad de suplementar con nitrógeno para mejorar los rendimientos. Esta cepa, cuya temperatura óptima son los 30°C, tolera alta presión osmótica y bajo pH, generando etanol con altos rendimientos (90% del teórico) a partir de hexosas (86). Pero la producción de etanol a partir de biomasa lignocelulósica requiere también la utilización de las pentosas derivadas de la hemicelulosa. A su vez, un microorganismo adecuado para este proceso debe tolerar los inhibidores generados durante el pretratamiento y altas concentraciones de etanol. La modificación genética de *S. cerevisiae* ha permitido el desarrollo de cepas capaces de metabolizar tanto arabinosa como xilosa (88).

*Escherichia coli* es una bacteria, Gram-negativa, que tiene la habilidad de fermentar un amplio rango de sustratos, incluyendo pentosas, hexosas, ácidos urónicos y polioles. Pero su utilización está

finamente regulada. Si azúcares como glucosa, sacarosa y fructosa están presentes en el medio de cultivo, se reprime la expresión de las enzimas necesarias para la utilización de xilosa, arabinosa, maltosa y lactosa (30). El desarrollo de cepas recombinantes capaces de utilizar de forma eficiente una mezcla de azúcares ha sido el objetivo de diversos trabajos (38,89). A su vez, las cepas de laboratorio e industriales no patogénicas pueden crecer en medio mínimo simple (89). Aunque muchos de los compuestos tóxicos que se generan durante el pretratamiento de la biomasa pueden inhibir el desempeño de *E. coli* durante la fermentación, algunas cepas han sido modificadas genéticamente para tolerar la presencia de furfural, hidroximetilfurfural y acetato (32,89). Pero *E. coli* no es tan eficiente en la fermentación de hexosas, como lo es *Zymomonas mobilis*, una bacteria etanologénica obligada. Una de las principales estrategias desarrolladas para mejorar el comportamiento de *E. coli* es la incorporación de la vía de producción de etanol de *Z. mobilis* (90). La vía metabólica de producción de etanol en *Z. mobilis* es un proceso de dos pasos en el cual la piruvato decarboxilasa (PDC) convierte directamente el piruvato a acetaldehído y CO<sub>2</sub>, para que luego la alcohol deshidrogenasa (ADH) convierta el acetaldehído en etanol. La expresión de un plásmido con el operón artificial de producción de etanol (pET) conteniendo los genes *pdh* y *adhB* de *Z. mobilis* en *E. coli*, permite que el etanol sea el producto de fermentación mayoritario (90). La incorporación de estos genes en el genoma de *E. coli* junto con la disrupción de genes clave implicados en la síntesis fermentativa de co-productos ha permitido el desarrollo de cepas capaces de utilizar diversos carbohidratos con altos rendimientos de etanol como único producto (89,91).

Existen otros microorganismos con potencial para ser utilizados en la fermentación de azúcares. Por ejemplo, *Pichia stipitis* es una levadura capaz de fermentar pentosas naturalmente que se ha evaluado como productora de etanol (9). *Klebsiella oxytoca*, una bacteria mesófila que tiene enzimas nativas para la captación y fermentación de celooligosacáridos, xilanos, hexosas y pentosas, ha sido evaluada en la producción de etanol (78,92). Por su parte, *Clostridium thermocellum* es una bacteria anaerobia termófila que degrada naturalmente celulosa y también es capaz de fermentar los azúcares resultantes principalmente en etanol, lactato, acetato, dióxido de carbono e hidrógeno (78). Esta cepa ha sido evaluada en co-cultivo con otros clostridios termofílicos capaces de fermentar pentosas, ya que *C. thermocellum* sólo puede fermentar hexosas (93).



## **Tecnologías de conversión de biomasa**

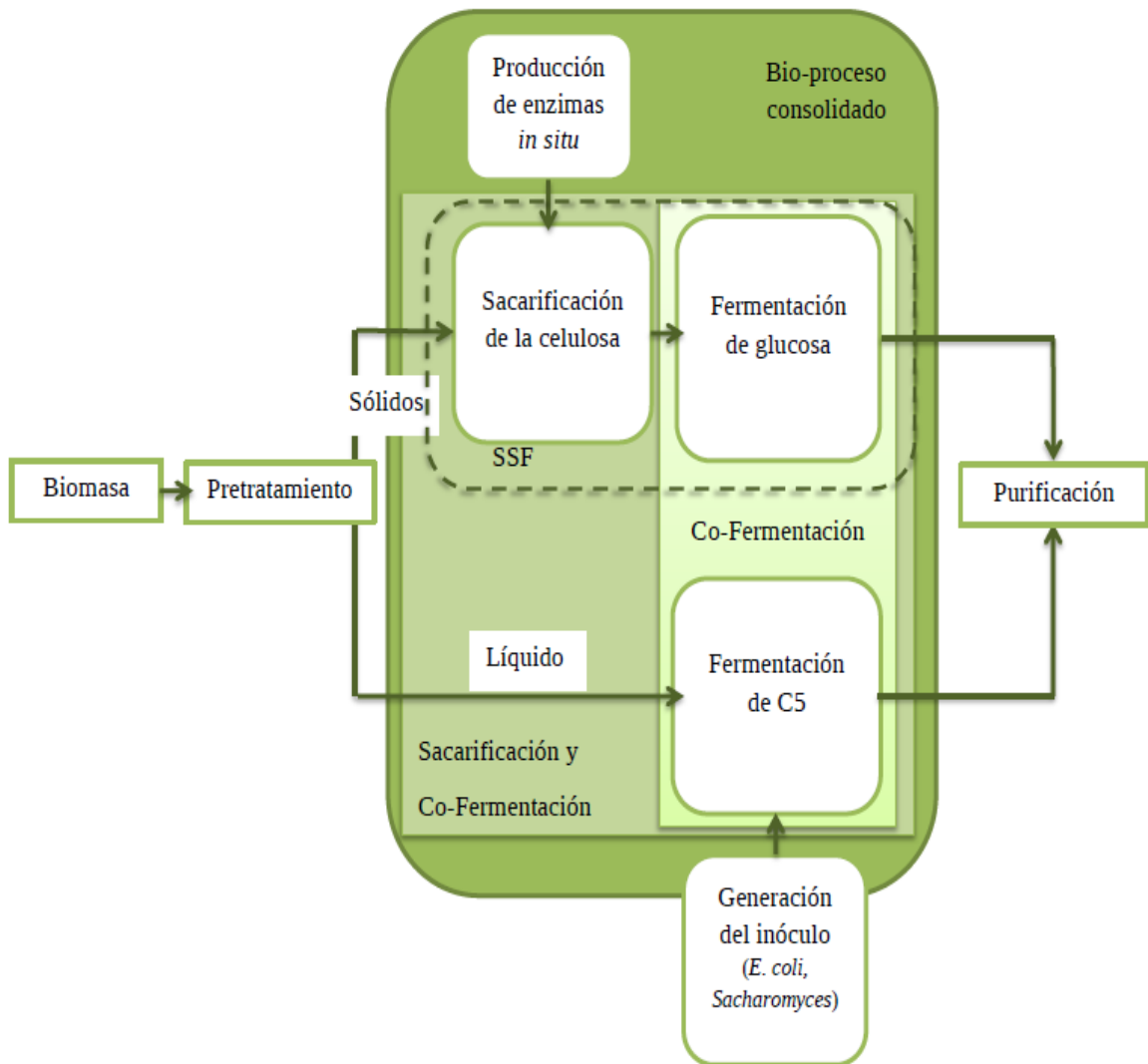
### **Sacarificación y fermentación por separado (SHF)**

Cuando la hidrólisis enzimática se realiza por separado de la etapa de fermentación el proceso es conocido como sacarificación y fermentación por separado (SHF, Separate hydrolysis and fermentation). En la configuración de SHF, el flujo líquido de los reactores de hidrólisis se mezcla con la fracción líquida generada con el pretratamiento y entra al reactor de fermentación de la glucosa. Luego de la misma, la mezcla se destila para purificar el bioetanol, dejando la xilosa no convertida en el reactor. En un segundo paso, xilosa se fermenta en etanol y éste se destila (6). La principal ventaja de SHF es que la hidrólisis y la fermentación se producen en condiciones óptimas y que las células pueden ser recicladas (94); la desventaja es que las enzimas celulolíticas son inhibidas por el producto final, de manera que la velocidad de hidrólisis se reduce progresivamente cuando la glucosa y celobiosa se acumulan (6) (Tabla 2).

### **Sacarificación y fermentación en simultáneo (SSF)**

Un proceso alternativo para la fermentación de los hidrolizados de la biomasa lignocelulósica es la sacarificación y fermentación en simultáneo (SSF), en el cual ambas etapas se llevan a cabo en un mismo reactor (28) (Figura 10). Esto se debe a que se requiere menos equipamiento, lo que representa una inversión menor. A su vez, la remoción de los productos finales de la hidrólisis (glucosa y celobiosa) deriva en un mayor rendimiento de conversión de la celulosa y la hemicelulosa (Tabla 2). Pero hacer compatibles las condiciones óptimas de las dos transformaciones (pH, temperatura, concentración de sustrato, etc) es la clave para el éxito del proceso (30).

La hidrólisis enzimática suele realizarse a 50°C, temperatura óptima de acción de las enzimas implicadas en este proceso, mientras que la fermentación se realiza a temperaturas entre 37 y 42°C, dependiendo de la cepa utilizada.



**Figura 10.** Resumen de las distintas configuraciones para la producción de etanol lignocelulósico.

La desventaja del SSF puede superarse utilizando microorganismos termófilos o enzimas adaptadas a bajas temperaturas. La identificación de cepas termo-tolerantes capaces de producir cantidades significativas de etanol a temperaturas más favorables para la sacarificación ha sido clave para la implementación de este proceso (95).

Cepas de levadura de los géneros *Kluyveromyces*, *Saccharomyces* y *Fabospora* que producen más de 5% (p/v) de etanol a temperaturas elevadas (>40°C) han sido identificadas (95). *Kluyveromyces marxianus* es una de las especies más prometedoras. Muchas de sus cepas crecen a temperaturas entre 45 y 50°C, siendo capaces de producir etanol a 38-45°C. A su vez, *K. marxianus* tiene la habilidad de utilizar arabinosa, galactosa, manosa y xilosa a temperaturas altas. Gracias a estas ventajas, *K. marxianus* a sido utilizada para producir etanol a partir de diversos materiales lignocelulósicos. Álamo, eucalipto, bagazo de sorgo dulce, paja de trigo, y residuos de *Brassica carinata*, tratados previamente por explosión de vapor, fueron convertidos a etanol a través de un proceso SSF utilizando *K. marxianus* y 15 PFU/ g sustrato de celulasas comerciales, obteniendo entre 50-72% del máximo teórico (96). Otra especie de levadura prometedora es *Pichia kudriavzevii*, una especie tolerante a altas temperaturas, ácidos y etanol. Kitagawa y colaboradores (97) desarrollaron un sistema de transformación para esta levadura, con el cual pudieron expresar una  $\beta$ -glucosidasa (BGL) de *Aspergillus aculeatus* en *P. kudriavzevii*. Esta cepa fue capaz de producir 29 g/L de etanol en 72 horas a 40°C sin el agregado exógeno de  $\beta$ -glucosidasa, a partir de avicel. Oberoi *et al.* (98) utilizaron la cepa HOP-1 de *P. kudriavzevii* para producir etanol a partir de rastrojo de arroz pretratado. La biomasa fue incubada 4 horas a 50°C con 20 PFU/g biomasa de celulasas, 50 U/g biomasa de  $\beta$ -glucosidasa y 15 U/g biomasa de pectinasa. Luego fue adicionada la cepa y se mantuvo durante 24 horas a 40°C. Mediante este procedimiento se produjeron 25 g/L de etanol (85% del máximo teórico).

## **Sacarificación y co-fermentación (SScF)**

La sacarificación y co-fermentación (SScF) es una de las opciones más convenientes para producir etanol a partir de biomasa rica en hemicelulosa. En este esquema, la sacarificación de la fracción sólida, rica en celulosa, es hidrolizada en el mismo reactor en que es fermentada la fracción líquida obtenida luego del pretratamiento de la biomasa. La fracción líquida contiene principalmente las pentosas derivadas de la hemicelulosa. Pero se ha determinado que la utilización de xilosa en hidrolizados de lignocelulosa, es el mayor cuello de botella que afecta los rendimientos de etanol tanto para *E. coli* KO11, *S. cerevisiae* 424A(LNH-ST) y *Z. mobilis* AX101 (99). Por lo tanto, el empleo de cepas modificadas genéticamente y el ajuste de los parámetros del proceso es necesario para viabilizar este esquema de producción de etanol. Por ejemplo, para la cepa recombinante de *S. cerevisiae* (TMB3400) capaz de metabolizar xilosa, se ha visto que el agregado de xilanasas y celulasas junto con

el sustrato mejora la utilización de xilosa (de 40 a 80%) y aumenta el rendimiento de etanol a partir de rastrojo de trigo pretratado (100). Por su parte, Zhang y Lynd (101) probaron dos cepas recombinantes capaces de fermentar la xilosa: *Z. mobilis* 8b y *S. cerevisiae* RWB222. En un proceso de SScF a partir de desechos de papel conteniendo 10 PFU de celulasas /g celulosa y 60 U  $\beta$ -glucosidasa/g celulosa, se obtuvieron 40 g/L de etanol (0.39 g etanol /g carbohidrato) a 30°C o 37°C para *S. cerevisiae* y *Z. mobilis*, respectivamente.

**Tabla 2.** Comparación entre las distintas configuraciones del proceso de producción de etanol

Proceso	Ventaja	Desventaja
Sacarificación y fermentación por separado	Optimización de cada etapa	Inhibición de celulasas por producto final
Sacarificación y fermentación en simultáneo	Menores dosis de enzimas Mayores rendimientos	Riesgos de contaminación Necesidad de enfriar luego de la sacarificación La hidrólisis y la fermentación no pueden optimizarse
Bioproceso consolidado	Mejores rendimientos Menores requisitos energéticos	Falta de organismos adecuados Dificultad en el control del proceso

(28,30,94)

## Bioproceso consolidado (CBP)

La producción de etanol a partir de lignocelulosa mediante sistemas microbianos incluye tres procesos biológicos: la producción de las enzimas hidrolíticas, la hidrólisis de la celulosa y la hemicelulosa a azúcares simples y la fermentación de los mismos (Figura 10). Estos ocurren en un único reactor en el proceso denominado bioproceso consolidado (CBP, consolidated bioprocessing). El CBP tiene el potencial de reducir los costos de producción con una mayor eficiencia que los procesos que implican un paso dedicado a la producción de celulasas (78,102). La aplicación del CBP para la producción de etanol requiere de un microorganismo (o un consorcio microbiano) capaz de utilizar los sustratos y generar el producto de interés. Dentro de las propiedades que se buscan está la utilización del sustrato mediante la producción de un sistema de enzimas hidrolíticas que permita altas velocidades de hidrólisis y la utilización de los productos resultantes de la hidrólisis en condiciones anaerobias en un medio de crecimiento sencillo (95). A su vez, se pretende obtener selectividad del producto formado y altas concentraciones del mismo. Un microorganismo que combine las características celulolíticas deseadas junto con altos títulos de etanol no ha sido descrito aún. Poder optimizar las condiciones para la sacarificación y la fermentación es clave para el éxito de este proceso.

El desarrollo de un microorganismo para la conversión de celulosa mediante CBP puede obtenerse por dos estrategias: la estrategia celulolítica nativa involucra la modificación genética de organismos que naturalmente degraden celulosa para que sean capaces de producir los productos de interés con altos rendimientos y buena tolerancia a los mismos. Por otro lado, la estrategia celulolítica recombinante implica la ingeniería de cepas no-celulolíticas que poseen altos rendimientos de productos y tolerancia a los mismos para que sean capaces de utilizar celulosa como resultado de un sistema heterólogo celulolítico (30).

Los microorganismos blanco para la primera estrategia suelen ser bacterias anaerobias, como *Clostridium* y *Thermoanaerobacterium*, las cuales son capaces de degradar celulosa o hemicelulosa y fermentar diversos azúcares. Un obstáculo importante para el uso de estos cultivos para la producción de etanol es que presentan un rendimiento de etanol reducido debido a la formación de otros productos de fermentación, en particular acetato y lactato (103). A su vez, uno de los factores que ha limitado la modificación de bacterias celulolíticas, es la ausencia de técnicas de transferencia de genes adecuados para estos microorganismos (102), aunque se desarrollaron métodos de transformación para *Clostridium cellulolyticum* y *Clostridium thermocellum* (104,105).

El hongo *Fusarium oxysporum* es otro candidato para CBP ya que secreta altos niveles de celulasas y fermenta pentosas y hexosas. Se ha visto que aunque este hongo es eficiente degradando celulosa y celobiosa, la glucosa actúa como inhibidor competitivo de las enzimas. La producción de etanol es acompañada por la formación de ácido acético, no siendo capaz de tolerar altos niveles de etanol. Aunque posee limitantes, *F. oxysporum* ofrece una fuente potencial de información genética para organismos CBP transgénicos con capacidades fermentativas más fuertes debido a su diverso arsenal de enzimas celulolíticas (78).

Incluso, se ha planteado la producción de etanol a partir de biomasa lignocelulósica (Switchgrass) sin pretratamiento. La cepa celulolítica *Caldicellulosiruptor bescii* fue modificada genéticamente para eliminar la generación de co-productos de fermentación a la vez que fue complementada con la acetaldehído/alcohol deshidrogenasa de *C. thermocellum* (106). Utilizando 2% (m/v) de biomasa obtuvieron 0,59 g/L de etanol. Aunque los rendimientos son bajos, la conversión directa de biomasa en etanol representa un nuevo paradigma para el desarrollo del bioproceso consolidado.

*Z. mobilis* es uno de los microorganismos prometedores para el desarrollo de cepas celulolíticas recombinantes. Se ha descrito la expresión heteróloga y secreción de 2 endo-1,4- $\beta$ -glucanasas de *Acidothermus cellulolyticus* utilizando señales nativas de secreción de *Z. mobilis* (107). Por su parte, la incorporación de una endoglucanasa de *Enterobacter cloacae* permitió la producción de etanol a partir de CMC y bagaso pretratado con NaOH (108).

Debido a la gran cantidad de productos químicos y alcoholes producidos a partir de glucosa mediante cepas de *E. coli* obtenidas mediante ingeniería metabólica, el desarrollo de un sistema celulolítico eficiente en este microorganismo es una estrategia prometedora para el desarrollo de una cepa CBP (109). Utilizando la proteína de anclaje PgsA, se expresaron en la superficie de *E. coli* una endoglucanasa, una celobiohidrolasa y una  $\beta$ -glucosidasa de *C. cellulolyticum*, logrando la producción de etanol a partir de avicel tratado con ácido sulfúrico (PASC) (110). Otro ejemplo es la secreción de la  $\beta$ -glucosidasa (BglC) de *Thermobifida fusca* fusionada al sistema de secreción tipo V (111) en la cepa de *E. coli* etanologénica MS04, la cual es capaz de producir etanol con altos rendimientos a partir de celobiosa.

Un fenómeno que impacta positivamente en la viabilidad y el rendimiento de CBP es la sinergia enzima-microorganismo, es decir, el hecho de que la eficacia de la hidrólisis de la celulosa mejora cuando está presente el complejo microorganismo-enzima-celulosa (CEM) en comparación con el complejo enzima-celulosa (CE). Diversos trabajos han mostrado que las tasas de hidrólisis son sustancialmente mayores cuando es mediada por el crecimiento de cepas celulolíticas en comparación con los preparados enzimáticos, incluso bajo condiciones optimizadas (102). Un buen ejemplo es el trabajo de Lu y colaboradores (112). *C. thermocellum* secreta un celulosoma que consiste de una endoglucanasa (CelA), una celobiohidrolasa (CelK), una xilanas (XynC) y una exoglucanasa (CelS) (78). Cuando el celulosoma fue separado y usado para hidrolizar celulosa con otra bacteria etanologénica en un proceso SSF, su actividad fue significativamente disminuida (112). Esto refleja la sinergia entre las enzimas celulolíticas y la bacteria secretora, cuya evaluación cuantitativa podría proporcionar una prueba más apoyo la conveniencia de CBP.

## Bioprospección

El número total de células procariotas en la tierra se ha estimado en  $4 \times 10^{30}$  (113). Esta diversidad presenta una enorme reserva genética y biológica que puede ser explotada para la recuperación de nuevos genes, vías metabólicas y sus productos (114). La bioprospección es la búsqueda sistemática, la clasificación e investigación de nuevas fuentes de compuestos químicos, genes, proteínas, microorganismos y otros productos con valor económico actual o potencial, que forman parte de la biodiversidad. Mediante técnicas clásicas de aislamiento y caracterización de microorganismos se han identificado, por ejemplo, microorganismos con capacidad de degradar hidrocarburos (115), cepas de microalgas híper productoras de lípidos (116), bacterias termófilas capaces de degradar

eficientemente la celulosa (117–119), hongos celulolíticos en diversos ambientes (120), así como bacterias psicrófilas degradadoras de lípidos o celulosa (121).

Está ampliamente aceptado que la aplicación de métodos microbiológicos clásicos, dependientes del cultivo en condiciones de laboratorio, para la recuperación de microorganismos del medio ambiente, ha tenido un éxito limitado en el acceso a la verdadera magnitud de la biodiversidad microbiana. De ello se desprende que gran parte de la diversidad genética microbiana existente permanece sin ser explotada, una cuestión de gran relevancia para la comprensión de las comunidades microbianas y de gran importancia para la industria biotecnológica (113). El reciente desarrollo de tecnologías diseñadas para acceder a esta gran cantidad de información genética a través de la extracción de ácidos nucleicos del medio ambiente ha proporcionado un medio de evitar las limitaciones los métodos dependientes del cultivo (113,122,123).

## **Metagenómica**

La "metagenómica" es el análisis, independiente de cultivo, de los genomas presentes en una comunidad microbiana, el denominado metagenoma (124). Las aproximaciones de metagenómica pueden basarse en el estudio funcional y/o en el análisis de las secuencias de los genomas microbianos contenidos en una muestra ambiental.

### **Metagenómica basada en el análisis de secuencia**

Uno de los abordajes posibles a realizar sobre un metagenoma es la secuenciación completa del ADN y su posterior análisis mediante herramientas bioinformáticas (113). Los avances en las tecnologías de próxima generación (NGS, next generation sequencing) han permitido un aumento significativo en la secuenciación de genomas individuales e impulsó una revolución en la secuenciación y análisis de metagenomas (125).

Con los avances recientes en la automatización, las tecnologías de secuenciación de alto rendimiento y el desarrollo de algoritmos para el ensamble de las secuencias, los proyectos de secuenciado masivo son técnicamente posibles (113), aunque los estudios de metagenómica utilizando NGS requieren recursos computacionales significativos y un grupo de investigadores especializados en el uso de estas herramientas.

Aunque el objetivo final es reconstruir todos los genomas pertenecientes a un ambiente, esto no es posible debido a la complejidad computacional involucrada. En su lugar hay dos tipos generales

de análisis realizados para la reconstrucción de genomas, con sus limitaciones inherentes: 1) el ensamble de las lecturas en "contigs", la clasificación taxonómica y la asignación de funciones, o 2) la reconstrucción de los componentes funcionales y taxonómicos del metagenoma basándose en los "reads" obtenidos (126).

El ensamblado de metagenomas presenta diversos desafíos. Debido a la composición desigual de la comunidad microbiana, muchos genomas presentan mala o nula cobertura, mientras que otros se ven sobre-representados. Para superar este inconveniente, son necesarios más datos para cubrir los diversos genomas, a costa de los requisitos de memoria y tiempo de cálculo computacional (125,127).

Los enfoques basados en "reads" también pueden permitir la exploración de las funciones genéticas y la clasificación taxonómica de los organismos. Sin embargo, sólo un subconjunto de los datos se puede anotar y clasificar, y el grado de confianza de la asignación puede variar en función de las bases de datos de referencia utilizadas y el nivel de cobertura de las lecturas (127).

Otra alternativa es la utilización de cebadores diseñados para blancos específicos. Esto es posible gracias a la conservación de regiones en las secuencias, lo que permite el aislamiento de genes sin conocimiento previo de la secuencia completa (128). Pero esto es válido principalmente para genes específicos, denominados de anclaje filogenético como el que codifica para subunidad ribosomal 16S para bacterias (129). Por el contrario, los genes que son de gran interés para aplicaciones biotecnológicas son demasiado divergentes respecto al consenso, por lo que son los menos propensos a ser detectado por sondas o cebadores diseñados en base a las secuencias depositadas en las bases de datos (128).

El ensamblado de genomas de microorganismos no cultivables ha sido posible utilizando esta estrategia, por ejemplo, a partir de muestras del mar de los Sargazos (130), una mina ácida (131) y agua superficial del estrecho de Puget (132).

### Metagenómica funcional

La aproximación de metagenómica funcional tiene como objetivo la caracterización de actividades enzimáticas presentes en una comunidad microbiana (133). Se basa en el clonado del ADN aislado en un vector adecuado, la incorporación del ADN foráneo en un hospedero seleccionado y la evaluación de los clones en medios selectivos (114). Este análisis permite la identificación rápida de clones con actividad de interés en diversas áreas.

En el año 2000, Rondon y colaboradores publicaron la primer "metagenoteca" a partir de ADN bacteriano de suelo utilizando cromosomas bacterianos artificiales (BACs, bacterial artificial



chromosome) (134). Los autores destacan entre sus conclusiones que esta estrategia de clonado directo de los ácidos nucleicos (ADN) a partir de una muestra ambiental puede ser una ruta para expandir el conocimiento de la diversidad microbiana, siendo también una vía de acceso a las propiedades filogenéticas, físicas y funcionales de los microorganismos presentes. En el trabajo, reportan la identificación de clones con actividades de interés, cuyas secuencias resultaron ser novedosas respecto a las reportadas por métodos clásicos, por lo que reconocen que esta aproximación es una fuente de genes novedosos plausibles de ser identificados mediante expresión en cepas como *E. coli* (128,134).

A su vez, el clonado de grandes fragmentos de ADN genómico, mediante la utilización de BACs o fósmidos, permite capturar operones o clúster de genes que codifican para vías metabólicas que pueden dirigir la síntesis de moléculas complejas, como, por ejemplo, los antibióticos (128). La secuencia de los elementos génicos que flanquean a un gen en particular puede, potencialmente, proporcionar información sobre el entorno genómico del gen o la filiación filogenética del organismo del que deriva.

La limitante de esta aproximación es que requiere la expresión de los genes, una correcta traducción, plegado y localización de la enzima con actividad de interés en el hospedero seleccionado. A su vez, en muchos casos se requiere de genes adicionales o co-factores, los cuales deben ser aportados por el microorganismo hospedero o estar codificados en el contexto del fragmento de ADN clonado (128). Para mitigar estas limitantes diversas estrategias se han desarrollado. Los sistemas de expresión heteróloga están siendo mejorados, mediante el desarrollo de nuevos vectores (133,135) o el desarrollo de hospederos alternativos a *E. coli* como *Streptomyces lividans* y *Pseudomonas putida* (136) o diversas Proteobacterias (137). De todos modos, se ha reportado que la maquinaria transcripcional de *E. coli* puede ser relativamente promiscua, siendo capaz de reconocer señales de expresión foráneas, incluso de genes de Firmicutes, Crenarchaeota, Euryarchaeota, Crenarchaeota y Proteobacteria (138).

Otra de las limitantes de esta aproximación es la necesidad de utilizar un buen método de selección de los clones en función de las actividades que se estén buscando. Por ejemplo, existen diversos sustratos que permiten la identificación de clones con actividad celulolítica. La carboximetil celulosa (CMC) es un sustrato utilizado comúnmente para la detección de actividad de endocelulasas (139). La degradación del mismo es fácilmente visualizada luego de la tinción de las placas con Rojo Congo, gracias a la formación de halos claros alrededor de las colonias con actividad positiva (140). El avicel, por su parte, es una forma de celulosa microcristalina. Durante su preparación con ácido diluido se elimina la hemicelulosa y la mayoría de la celulosa amorfa (7). Al evaluarse la utilización de este sustrato en medio mínimo, se presume que los clones capaces de crecer en avicel tendrán, además de hidrolasas, los sistemas de transporte necesarios para secretar estas enzimas al medio extracelular y para internalizar los glúcidos liberados. Por su parte, el papel de filtro está compuesto esencialmente por

celulosa y su conversión en azúcares reductores es un ensayo comúnmente utilizado para determinar la actividad celulasa en enzimas purificadas (141).

Aunque la degradación de la hemicelulosa requiera de múltiples enzimas, el xilano es un sustrato ampliamente utilizado para la detección y caracterización de enzimas implicadas en la degradación de dicha biomasa (142). Por ejemplo, uno de los sustratos más utilizados es el xilano derivado de madera de haya, el cual es muy insoluble (aproximadamente un 95%), está compuesto por un 87.7% de xilosa, pequeñas cantidades de glucosa, trazas de arabinosa y galactosa, y un 3% de ácido ascórbico (143). La incorporación de este sustrato en placas permite la detección de clones con actividad xilanólítica luego del revelado de las mismas con Rojo Congo.

## **Aportes de la metagenómica a la biotecnología**

El uso a gran escala de la biotecnología en lugar, o como complemento de los procesos tradicionales de producción industrial, sobre todo en el sector químico, es visto como una oportunidad y también como una necesidad (144). Esto, sumado al desarrollo de nuevas aplicaciones biotecnológicas, impulsa el mercado de enzimas a nivel comercial, el cual se estima rondará los 7 billones de dólares en 2017 (145).

La metagenómica es considerada una fuente potencial de catalizadores capaces de cubrir las actividades requeridas por la industria (146). Diversas enzimas novedosas se han detectado y caracterizado a partir de metagenotecas (123,146,147), muchas de las cuales presentan un novedoso rango de sustratos y/o productos o buena estabilidad en condiciones extremas (147).

De la gran cantidad de clones positivos obtenidos en diversos "screening", unos pocos genes han sido clonados y sus productos sobre-expresados, purificados y caracterizados bioquímicamente. Se atribuye que la causa de esto es el alto número de nuevas enzimas que pueden detectarse en una librería metagenómica y la cantidad de tiempo y esfuerzo que requiere la caracterización completa de cada una de estas enzimas (123). El desarrollo de ensayos de selección de alto rendimiento permite sobrellevar este inconveniente (148).

El primer reporte de un screening funcional exitoso de ADN ambiental se basó en la evaluación de un consorcio enriquecido de un digestor anaerobio alimentado con lignocelulosa (149), a partir del cual cuatro clones con actividad en celulosa o xilosa fueron identificados y caracterizados. Gracias al uso de la metagenómica funcional se han detectado enzimas con actividad celulasa (149–154), quitinasa (155), amilasa (156), DNAsa (134), esterasa/lipasa (134,157–159), proteasa (160), oxigenasa (161) y factores con actividad antimicrobiana, cuya expresión posiblemente involucre la expresión de múltiples

genes (134,162). Detalles de la diversidad de las enzimas detectadas, su origen y la estrategia de detección se listan en (163,164).

## **Identificación de celulasas mediante metagenómica funcional**

Las celulasas poseen múltiples aplicaciones industriales. Se usan en la industria textil para el ablandamiento del algodón, en el mercado de los detergentes para el acabado del color y para evitar la deposición de los detergentes en polvo; en la industria del papel y la pulpa de celulosa para la modificación de las fibras y la eliminación de tintas (165); en la industria de los alimentos y bebidas, principalmente cerveza y vino y en alimentación animal (166).

Tanto los hongos como las bacterias producen una amplia variedad de celulasas y hemicelulasas. Los hongos han sido más utilizados como productores de enzimas, debido a su capacidad de secretar grandes cantidades al medio extracelular, lo cual facilita su purificación (93). La mayor parte de las celulasas comerciales (incluyendo las  $\beta$ -glucosidasas) son producidas por cepas de *Trichoderma* y *Aspergillus* (93,167), representando aproximadamente un 20% del mercado mundial de enzimas (166).

El aislamiento y la caracterización de nuevas glicosil-hidrolasas (GH) bacterianas está en auge. Esto se debe principalmente a que las bacterias ofrecen una tasa de crecimiento mayor que los hongos, y por consiguiente, una mayor tasa de producción de enzimas recombinantes (168). A su vez, las GH bacterianas son más complejas, lo que deriva en que posean mayor funcionalidad y sinergia (93). Otro aspecto muy importante, es que las bacterias habitan ambientes con diversas condiciones, lo cual implica que las celulasas producidas por bacterias que habitan ambientes extremos sean también tolerantes a condiciones ambientales extremas (169) lo que las hace útiles a nivel industrial.

La búsqueda de celulasas novedosas utilizando aproximaciones metagenómicas puede aplicarse, virtualmente, a cualquier ambiente, principalmente aquéllos en donde habitan microorganismos especializados en la degradación de lignocelulosa. Las variaciones a nivel de la secuencia de nucleótidos en los genes que codifican para GH no permite el diseño de cebadores universales que permitan la amplificación masiva de genes de celulasas a partir de ADN ambiental (170). Existen, de todos modos, algunos ejemplos en donde esto se ha logrado para sub-conjuntos de celulasas (171,172).

Estudios de análisis de secuencias han mostrado la diversidad de celulasas presentes en diversos ambientes, principalmente el rumen bovino (153,173–175), el intestino de termitas (176), el microbioma de hormigas (177) y gusanos (178,179).

El ambiente celulolítico mejor estudiado es el rumen, el cual es básicamente un quimióstato degradador de paredes celulares vegetales. El rumiante se encarga de la molienda de la biomasa, la cual es luego digerida por una mezcla compleja y densa de microorganismos anaerobios (180). Aunque la celulosa es la mayor fuente de carbono en el rumen, solo el 10% de las  $10^{10}$ /mL bacterias que lo habitan son celulolíticas (181), lo cual refleja lo eficiente que son estos microorganismos en la degradación de biomasa. Existen degradadores de celulosa tanto adheridos a las partículas de alimento como en el líquido ruminal, aunque los microorganismos de las dos fracciones son diferentes (175,182). Utilizando herramientas de secuenciado masivo, la diversidad de enzimas degradadoras de biomasa en el rumen se ha caracterizado (153,174,175) y se ha visto que la mayoría de estos genes comparte poca homología con los depositados en las bases de datos (183). La aplicación de metagenómica funcional en este ambiente ha revelado la presencia de glucanasas novedosas (152,183–185), pero la diversidad microbiana y enzimática del rumen no está aun totalmente caracterizada.

## Estructura de la tesis

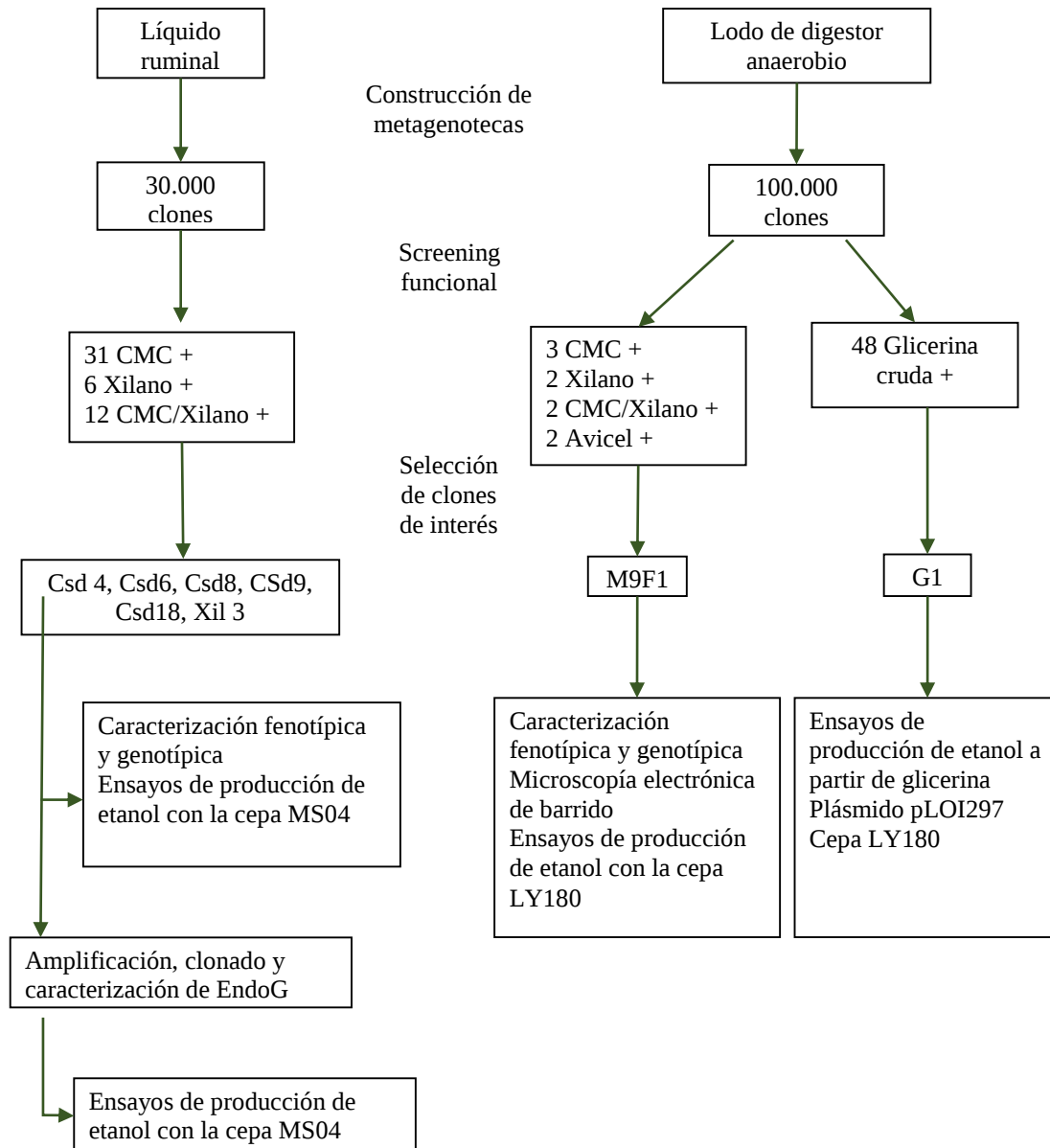
Luego de la introducción general se presentan la hipótesis y los objetivos de trabajo. A continuación la metodología empleada, así como los resultados alcanzados se dividen en 5 capítulos:

- 1) Se presenta la estrategia general de trabajo, la construcción de dos metagenotecas y la selección de clones con actividad de interés. La caracterización fenotípica y el análisis bioinformático del genotipo de cada uno de ellos algunos clones seleccionados.
- 2) Artículo publicado en la revista FPTA-INIA. En éste se muestra que el fósforo denominado M9F1 es necesario y suficiente para que *E. coli* sea capaz de degradar papel de filtro, bagazo de caña y avicel de manera eficiente y utilizar estos sustratos como fuente de carbono.
- 3) Artículo publicado en la revista Applied Microbiology & Biotechnology, en el cual se detalla la caracterización funcional y genotípica del fósforo Csd4. Se presentan también los ensayos que permiten concluir que este fósforo confiere a cepas de *E. coli* etanológicas la producción de etanol a partir de carboximetil celulosa.
- 4) Manuscrito enviado a la revista arbitrada Enzyme & Microbial Technology en el cual se presenta la caracterización bioquímica de una endoglucanasa con potencial para ser utilizada en procesos industriales.
- 5) Manuscrito en preparación. En este apartado se presenta la producción de etanol a partir de glicerina cruda, obtenida como sub-producto en la producción de diésel. Se seleccionó un fósforo que permite la utilización de forma eficiente de glicerol y su fermentación en condiciones microaerofílicas.

Luego se presenta la discusión general del trabajo de tesis y las perspectivas que surgen de la misma.

# Plan de trabajo seguido

Esquema de trabajo seguido durante el desarrollo de este trabajo de Tesis.



## Hipótesis

---

Mediante aproximaciones de metagenómica funcionales es posible acceder a enzimas novedosas que viabilizan el desarrollo de sistemas más eficientes, para la producción de biocombustibles a partir de biomasa vegetal y de sub-productos industriales.

## Objetivo general

---

Identificar y caracterizar elementos genéticos que confieran a cepas de *Escherichia coli* la habilidad de degradar desechos agro-industriales para la producción de bioetanol combustible, utilizando una aproximación de metagenómica funcional.

## Objetivos específicos

---

- ☞ Seleccionar y caracterizar de nuevas enzimas hidrolíticas capaces de degradar diversos materiales lignocelulósicos.
- ☞ Seleccionar y caracterizar de elementos genéticos que confieran a cepas de *E. coli* la habilidad de crecer en glicerina cruda derivada de la producción de biodiesel.
- ☞ Evaluar el potencial de las enzimas seleccionadas en la producción de etanol a partir de celulosa mediante su incorporación en cepas de *E. coli* etanologénicas.

**Capítulo 1.**

**CONSTRUCCIÓN DE DOS  
METAGENOTECAS, SCREENING Y  
SELECCIÓN DE CLONES CON ACTIVIDAD  
CELULOLÍTICA**

---



## Resumen

La parte inicial de este trabajo consistió en la construcción de dos bibliotecas metagenómicas, una a partir de líquido ruminal (conteniendo 1.2 Gpb) y otra de contenido de un reactor anaerobio (4.2 Gpb). Estas fueron evaluadas en medios indicadores para actividad celulasa o xilanasas, obteniendo una relación de un clon positivo cada 20 Mpb y cada 600 Mpb, respectivamente. De 58 clones positivos, se seleccionaron nueve cuyo análisis funcional exhaustivo reveló que confieren a *Escherichia coli* la habilidad de utilizar avicel, xilano, bagazo de caña de azúcar o celobiosa como fuente de carbono. El análisis bioinformático de estos fósmidos reveló que, para siete de ellos, la actividad se debía a la expresión de enzimas con homología a celulasas o xilanasas no caracterizadas provenientes, principalmente, de microorganismos no cultivables. En los otros dos fósmidos no se pudo atribuir, por homología de secuencia, a ninguno de los genes anotados las cualidades fenotípicas detectadas. Estos resultados confirman la hipótesis de que las aproximaciones de metagenómica funcional permiten acceder a enzimas novedosas.

## Objetivos

- ☞ Detectar clones con actividad celulolítica y/o xilanolítica dentro de dos metagenotecas construidas a partir de líquido ruminal y lodos de un reactor anaerobio.
- ☞ Caracterizar fenotípicamente clones seleccionados en función de sus características. Determinar su capacidad de degradar papel de filtro, avicel y bagazo de caña.
- ☞ Caracterizar, mediante herramientas bioinformáticas, el genotipo de estos clones, identificar los ORFs que codifican para las enzimas implicadas en los fenotipos observados y determinar *in silico* la presencia de dominios GH y CBM en las enzimas detectadas.

## Metodología

### Cepas y plásmidos

Para el desarrollo de esta tesis se utilizaron una serie de vectores y cepas de *E. coli* las cuales se listan en la tabla 3. Como se puede apreciar, la combinación del vector pCC1FOS con la cepa Epi300 T-1R, utilizada para la construcción de las metagenotecas, permite el mantenimiento de los fósidos de manera estable en bajo número de copia dentro del hospedero. El agregado de arabinosa al medio de cultivo permite la expresión del factor TrfA, codificado en el genoma de la cepa, el cual es necesario para la replicación del vector mediante el origen de replicación *OriV*. Éste asegura la presencia de aproximadamente 100 copias del elemento extracromosómico, facilitando el screening funcional de las metagenotecas.

**Tabla 3.** Características de los vectores y cepas utilizados

Plásmido	Características
pCC1FOS	<i>OriF</i> , <i>OriV</i> (dependiente del factor TrfA) Sitios cos, permiten el empaquetado en fagos $\lambda$
Vectores pET	Gen <i>lacI</i> , codificante para la proteína represora lac Promotor T7, específico para ARN polimerasa T7 <i>OriF</i> . Amp <sup>R</sup> Secuencia codificante para His-Tag
pLOI279	Operón PET ( <i>ptrC<sub>Zm</sub></i> , <i>adhA<sub>Zm</sub></i> ) Tet <sup>R</sup>
Cepa	Características
Epi300 T-1R	F- <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) (StrR) $\phi$ 80 <i>dlacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74</i> <i>recA1</i> <i>endA1</i> <i>araD139</i> $\Delta$ ( <i>ara</i> , <i>leu</i> )7697 <i>galU</i> <i>galK</i> $\lambda$ - <i>rpsL</i> <i>nupG</i> <i>trfA</i> <i>tonA</i> <i>dhfr</i> Expresión del factor TrfA bajo la inducción de arabinosa Resistencia a la infección por bacteriófago
Bl21(pLysS) DE3	F- <i>ompT</i> <i>gal</i> <i>dcm</i> <i>lon</i> <i>hsdSB</i> (rB- mB-) $\lambda$ (DE3) pLysS(cmR) Expresión de T7 RNA polimerasa inducible por IPTG Carente de proteasas OmpT y Lon
MS04	<i>E. coli</i> MG1655 $\Delta$ <i>pflB</i> , $\Delta$ <i>adhE</i> , $\Delta$ <i>frdA</i> , $\Delta$ <i>xyIFGH</i> , $\Delta$ <i>ldhA</i> , <i>PpflB::pdc<sub>Zm</sub>-adhB<sub>Zm</sub></i> Homofermentativa, productora de etanol. Metaboliza xilosa y glucosa eficientemente. Tolerante a acetato Km <sup>R</sup>
LY180	<i>PfrdBC::</i> ( <i>Zm</i> frg <i>celY<sub>Ec</sub></i> ) <i>ldhA::</i> ( <i>Zm</i> frg <i>casA</i> B <sub>Ko</sub> ) <i>adhE::</i> ( <i>Zm</i> frg <i>estZ<sub>Pp</sub></i> FRT) <i>ackA::FRT</i> <i>rrlE::</i> ( <i>pdc</i> <i>adhA</i> <i>adhB</i> FRT) <i>mgsA::FRT</i> Fermentativa, productora de etanol y ácido fórmico. Tolerante a furfural

Para los ensayos de producción de etanol se utilizaron dos cepas de *E. coli* etanologénicas. Estas cepas fueron modificadas por ingeniería metabólica con el fin de que todo el flujo metabólico, en condiciones fermentativas, derive en la producción de etanol. Por esto no poseen las enzimas necesarias para la producción de lactato, succinato y ácido acético. A su vez, poseen diferentes características que favorecen su utilización en la producción de etanol a partir de azúcares derivados de lignocelulosa, como son: utilización de xilosa, resistencia a furfural o acetato. Difieren en que, la cepa LY180 puede metabolizar, en condiciones fermentativas, glicerol. Para la utilización de este sustrato es necesaria la co-formación de formato, y dicha cepa posee el gen *pflB*, el cual codifica para la piruvato formato liasa I, que cataliza la reacción entre el piruvato y el coenzima A para dar acetilCoA y formato.

### **Obtención de muestras ambientales y procesamiento de ADN metagenómico**

De un novillo Holando de 480 kg alimentado en praderas naturales, sacrificado en el Frigorífico Pando (Canelones, Uruguay) se obtuvo una muestra de contenido ruminal. A su vez, una muestra de lodos anaerobios fue obtenida de un digestor alimentado con contenido ruminal [sólidos totales (ST) 19,3% (p/p) y sólidos volátiles (SV) 17,3% (p/p)], grasa de trinchado de curtiembre (ST 25,1% y SV 17,6%) y lodos de purga de lagunas anaerobias de una aceitera (ST 25,6% y SV 21,7%) (186) el cual opera en el Departamento de Reactores de Facultad de Ingeniería (Universidad de la República, Montevideo, Uruguay). Todas las muestras fueron mantenidas a 4°C hasta su procesamiento.

### **Obtención y procesamiento de ADN metagenómico**

Se aisló ADN metagenómico a partir de fluido ruminal y de lodo de reactores anaerobios según (187) con modificaciones. Para esto, 500µl de muestra se diluyeron en 3mL de amortiguador fosfato salino (PBS, ver anexo 1), se colocaron en 3mL de Percoll (Sigma-Aldrich) y se centrifugaron durante 20 minutos a 10.000 g, conservando la capa opalescente. Se verificó la presencia de células bacterianas en la misma observando al microscopio óptico luego de realizar tinción de Gram (188).

La fracción obtenida se suspendió en 1mL de solución de lisis (ver anexo 1). Se agregaron bolitas de zirconia (aproximadamente 100 µl) y se agitó horizontalmente durante 30 segundos. Luego la muestra se incubó a 70°C durante 15 minutos mezclando cada 5 minutos y se centrifugó a 4°C durante 5 minutos a 16.000 g. El sobrenadante se transfirió a un tubo limpio, se agregaron 15 µl de proteinasa-K (20 mg/mL) y se incubó 1 hora a 37°C. Posteriormente se agregaron 160 µL de CTAB 10% en NaCl 0,7 M y se incubó 10 minutos a 65°C. El ADN se extrajo dos veces con igual volumen de fenol y dos veces con igual volumen de cloroformo y luego se precipitó con 0,6 volúmenes de isopropanol y 1/10 volúmenes de AcNa 3 M pH = 5,2; posteriormente se incubó en hielo durante 30 minutos. Se centrifugó a 10.000 rpm a 4°C durante 15 minutos, se eliminó el sobrenadante y el pellet se lavó con etanol 70%.

Se dejó secar a temperatura ambiente y se suspendió en 200 µL de agua miliQ. Se verificó la calidad del ADN obtenido mediante electroforesis en gel de agarosa 0,8% (p/v) en amortiguador TAE 1x y se conservó a -20°C.

### **Construcción de bibliotecas metagenómicas**

Se seleccionó ADN de entre 30 y 50 Kpb por electroforesis en gel de agarosa 0,8% (p/v) el cual fue clonado mediante la utilización de CopyControl Fosmid Library Production Kit (Epicentre Biotechnologies) para la generación de fósmidos de aproximadamente 40 Kpb según las recomendaciones del fabricante. Se infectaron células de Epi300-T1<sup>R</sup> y se seleccionaron en placas de Luria Bertani (LB)- Cm<sub>12,5</sub>. Las transfectantes se ordenaron en grupos de 10 clones en placas de 96 pocillos en 150 µL de LB- Cm<sub>12,5</sub>, se crecieron a 37°C durante 16 horas, se agregó igual volumen de glicerol al 50% (v/v) y se conservaron a -80°C.

### **Evaluación en medios indicadores**

Las actividades hidrolíticas se detectaron en LB- Cm<sub>12,5</sub> agar con el agregado de 0,2% (p/v) CMC sal disódica, 0,2% CMC fibrosa, 0,2% CMC microgranulada o 0.5% de xilano de madera de haya (beechwood-xylan). Con el fin de aumentar el número de copias del vector, se adicionó 0,01% (p/v) de arabinosa al medio. Luego de 7 días de incubación a 30°C las placas fueron lavadas con agua estéril y teñidas con Rojo Congo 0,05% (p/v) durante 20 minutos (140). Se eliminó el exceso de colorante y se asignaron como positivos los clones que mostraron la formación de un halo amarillo alrededor de la colonia.

A su vez se evaluó la capacidad de los clones de crecer en medio mínimo M9 (ver anexo 1) con 0,5% de avicel PH-101 como única fuente de carbono. Luego de ser recuperados los "pooles" conservados a -80°C en placas de LB- Cm<sub>12,5</sub>, los clones fueron transferidos a placas de medio M9 suplementado con 0,5% de glucosa, se incubaron durante 16 horas a 37°C y se transfirieron al medio final, manteniéndose a 30°C durante 10 días.

### **Caracterización fenotípica de los clones seleccionados**

Los fósmidos presentes en los clones positivos en al menos uno de los medios de selección fueron aislados utilizando el QIAprep<sup>R</sup> Miniprep kit (Quiagen) y transferidos a células de Epi300-T1<sup>R</sup> electrocompetentes frescas (189). Los fenotipos originales fueron verificados y sólo se conservaron los clones que demostraron mantener el fenotipo original. Para descartar los posibles clones duplicados que

hayan surgido de seleccionar dos o más veces el mismo fósido, se realizaron perfiles de restricción de los fósidos con la enzima *EcoRV*. En caso de constatarse idénticos perfiles de restricción en dos o más clones, se eligió un clon representante de la clase para continuar el estudio descartándose los duplicados. Los clones así seleccionados fueron sometidos a ensayos complementarios para indagar aún más en su capacidad celulolítica.

Con el fin de evaluar la capacidad de los clones de utilizar otras fuentes de carbono, cinco mL de medio líquido mínimo M9 fueron suplementados con papel de filtro (Whatmann N<sup>o</sup>2), celobiosa o xilano a una concentración final de 0,5% (p/v) como fuente de carbono. Se inocularon estos medios con los clones seleccionados con una DO<sub>620 nm</sub> de 0,1 y se incubaron a 30°C durante 10 días midiendo el crecimiento celular cada 24 horas. El crecimiento celular, se siguió utilizando el equipo UV Spectrophotometer UV-1800 (Shimadzu).

Como indicador de actividad hidrolítica se utilizó el sustrato fluorogénico 4-MUC (190). Para esto, los clones se crecieron en 150 µL de LB- Cm<sub>12,5</sub>-Ara en placas de 96 pocillos durante 16 horas, se centrifugaron las células, se re-suspendieron en 50 µL de buffer de lisis y se lisaron por ciclos de congelado-descongelado (-80°C/ 37°C). A los lisados celulares se les adicionó el sustrato 4-MUC (0,05% (p/v) final), se incubaron durante 1 hora a 37°C y luego se iluminaron con luz ultravioleta. Se consideraron positivos los clones capaces de emitir luz en dichas condiciones. A los clones seleccionados en este ensayo se les realizaron ensayos complementarios con el mismo sustrato para estudiar la cinética de la actividad glucolítica. 200 µL de LB- Cm<sub>12,5</sub>-Ara suplementado con 4-MUC 0,05% fueron inoculado con las células a una DO<sub>620nm</sub> inicial de 0,05 y suplementado con 0,5% de glucosa, xilosa, celobiosa, CMC, avicel o sin suplementar. El crecimiento celular y la fluorescencia (longitud de onda de excitación 365nm y emisión 445nm) fueron medidos cada 8 horas durante 60 horas en placas de 96 pocillos utilizando el equipo Varioskan Flash Multimode Reader (Thermo Scientific).

Los sustratos CMC, xilano, avicel y 4-MUC fueron obtenidos de Sigma-Aldrich.

### **Pretratamiento de bagazo de caña**

Bagazo de caña molido con un contenido de humedad del 54% fue obtenido de la empresa productora de alcoholes ALUR. Este sustrato fue sometido a dos tipos de pretratamiento: incubación con agitación durante 2 horas con ácido fosfórico 1% o con ácido sulfúrico 1%, seguido de 1 hora de autoclave a 121°C. Luego de este tratamiento, la fracción líquida fue neutralizada con NaOH hasta lograr pH= 6,5 y la fracción sólida fue lavada 3 veces con abundante agua destilada. La biomasa obtenida se mantuvo a 4 °C hasta su incorporación en el medio M9 en una concentración final de 1% (p/v).

## **Análisis *in silico* de los fósmidos**

De los fósmidos con actividad positiva nueve se seleccionaron para su secuenciación completa mediante la construcción de librería Shotgun seguido por secuenciación por Sanger de los clones individuales y posterior ensamblado (Macrogen Inc.). Alternativamente, se utilizó la tecnología Ion Torrent (Plataforma de Secuenciación Masiva, IIBCE). En este caso, se filtraron los reads que correspondían a ADN de *E. coli* contaminante usando el programa Mirabait y el genoma de la cepa K12. Posteriormente se procedió al ensamblado utilizando el programa Mira 4.0 rc5, indicando que el secuenciador era un Ion Torrent. Se determinó que la cobertura óptima es 40X. Por lo tanto, se tomaron muestras al azar de 7500 reads (con un largo promedio de 235bp cada uno). El curado manual de los contigs obtenidos se realizó con el programa Gap5. Una vez ensamblados los contigs, se procedió a su análisis *in silico* en el servidor WebMGA (191). Para predecir marcos de lectura abiertos (Open Reading Frame, ORFs) en los fósmidos se utilizó el algoritmo MetaGene (192). Las posibles funciones de ORFs identificados fueron anotadas basándose en su homología con agrupamientos de grupos de ortólogos (COGs) y con las familias de proteínas Pfam utilizando BLASTp. Las secuencias obtenidas se compararon con las depositadas en la base de datos mediante BLASTX y BLASTP (BLAST, “basic local alignment search tool”) del NCBI (National Center for Biotechnology Information, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Las secuencias de aminoácidos de cada enzima de interés identificada y las más cercanas seleccionadas luego de la comparación mediante BLASTX se usaron para realizar análisis filogenéticos. Para los alineamientos se utilizó el software Mega 6.0 (193) por el método “Muscle”. A su vez, se realizó la anotación de las proteínas codificadas basándose en la presencia de dominios específicos de enzimas activas en carbohidratos (CAZymes, Carbohydrate-active enzymes) (54) en el servidor: <http://csbl.bmb.uga.edu/dbCAN/annotate.php> (194). Con igual fin también se utilizó el paquete CAZymes Analysis Toolkit (CAT) del BESC Data & Tools 2.0. del BioEnergy Science Center (195) disponibles en el sitio web: <http://mothra.ornl.gov/cgi-bin/cat/cat.cgi?tab=ORTHOLOGS>.

Las características físico químicas de las enzimas caracterizadas se inferidas utilizando el servidor ProtParam (<http://web.expasy.org/protparam/>) (196).

## **Clonado y expresión de hidrolasas**

Una vez identificados los genes que codifican para las enzimas de interés, se procedió al diseño de cebadores adecuados para la amplificación y posterior clonado de los mismos en vectores de la familia pET fusionados a la secuencia poli-histidina (His-tag). Se seleccionaron 3 vectores pET teniendo en cuenta la presencia de sitios de restricción internos en los genes de interés y las características particulares de cada uno de ellos (Tabla 4).

Los genes identificados se amplificaron por PCR usando Pfu ADN polimerasa (Fermentas, USA). Los cebadores utilizados y el ADN molde se detallan en la tabla 5. Se agregaron sitios de restricción adecuados para el clonado de los insertos en los vectores seleccionados. Las condiciones de PCR fueron las siguientes: 95°C durante 5 minutos seguido de 30 ciclos de 93°C durante 1 minuto, 65°C por 30 segundos y 68°C durante 90 segundos, con una extensión final a 68°C durante 10 minutos.

**Tabla 4.** Vectores seleccionados para el clonado y expresión heteróloga de genes seleccionados

Vector	Resistencia	His-tag <sup>1</sup>	Péptido señal	Enzimas de restricción seleccionadas para el clonado
pET-14b	Ampicilina	N	No	<i>NdeI/ BamHI (NdeI/XhoI)</i>
pET-22b	Ampicilina	C	Si *	<i>NcoI/HindIII</i>
pET-28a	Kanamicina	N/C	No	<i>NcoI/BamHI</i>

<sup>1</sup> Posición de la cola de poli-histidinas. N, extremo amino-terminal; C, extremo carboxi-terminal

\* Secuencia señal pelB en N-terminal para potencial localización periplasmática

Los productos de PCR se purificaron utilizando el kit Isolate PCR and Gel Kit (Bioline, UK). Una vez purificados se ligaron al vector pBluescript + (pBK+) digerido previamente con la enzima *EcoRV*. Una vez confirmada la presencia del inserto esperado se escindió del vector por restricción utilizando las enzimas correspondientes. Los vectores pET14a, pET22b y pET28a fueron sometidos a restricción con las mismas enzimas de restricción, se purificaron de gel de agarosa luego de ser sometidos a electroforesis usando el kit Isolate PCR and Gel Kit (Bioline, UK). Los insertos se ligaron a los vectores pET usando T4 DNA Ligase (Fermentas) y se transformaron células competentes de *E. coli* DH5 $\alpha$  preparadas según (189).

Las células obtenidas se sembraron en placas de LB-Km<sub>50</sub> o Amp<sub>50</sub>-IPTG 100  $\mu$ M- X-Gal<sub>20</sub>, seleccionando aquellas con inserto por coloración azul en dicho medio. A partir de estas se purificó la construcción y se confirmó la presencia del inserto por secuenciación. Se transfirieron las construcciones a la cepa de *E. coli* BL21 (DE3) pLysS para la expresión de las proteínas recombinantes. Los clones fueron seleccionados en placas de LB-Cm<sub>12,5</sub>-Km<sub>50</sub>/Amp<sub>50</sub> y conservados a -80°C.

Para determinar las condiciones óptimas de expresión de las enzimas se inoculó un cultivo de 5mL de LB con antibiótico correspondiente (Cm<sub>12,5</sub> + Km<sub>50</sub> o Amp<sub>50</sub>) con las transformantes de BL21 obtenidas y se incubó a 37°C durante toda la noche a 150 rpm. Luego, este cultivo fue utilizado para inocular 5 mL de 2xYT (ver anexo 1) con Cm + Km/Amp según corresponda, se incubó a 37°C hasta alcanzar una DO<sub>620nm</sub> de 0,6. La expresión fue inducida con el agregado de IPTG 0,5 mM e incubando a 25, 30 y 37°C durante 16 horas, con agitación. Se tomaron muestras de cultivo y sobrenadante a las

0, 2, 4 y 16 horas de incubación. Las células fueron resuspendidas en buffer de lisis e incubadas a 100°C durante 5 minutos. Se midió la actividad a los lisados celulares y al sobrenadante y se visualizaron las proteínas mediante SDS-PAGE en geles de poliacríamida, usando el sistema Mini Protean (Bio-Rad). La concentración total de proteínas se midió utilizando el método el kit Quick Start Bradford Protein Assay (Bio-Rad).

**Tabla 5.** Lista de cebadores utilizados para la amplificación de ORFs seleccionados y los vectores utilizados en cada caso

Gen	Cebador de evance	Cebador reverso	Vector
EndoG	5'-AGGCACCATGGCTATGAAAA	5'-ATATTAAGCTTTTTATCAAGCC	pET28a
	AGATTCTTTATTTCTTTGC-3'	CTTTGGCACCC-3'	
Xyl4	5'-AAAAGGCCATGGATATGCTA	5'-TTCGGATCCGAAGGTTTTTCTT	pET22b
	CTGGTTATCGCAATCC-3'	GATAGTCGAAGTAAAGG-3'	
Lac4	5'-ATAAGGATCCGATGACACAA	5'-CCCCCAAAGCTTACTGTTTGT	pET22b
	TTGCCCAACTCC-3'	GACAAGTAGATTCC-3'	
Endo18	5'-ATCATATGGCCCGCATCAAG	5'-ATGGATCCGTTGTCATCGGCCA	pET14b
	GAGATGG-3'	CTACG-3'	
Xyl3-43a	5'-ATCATATGCTGTATCGCGCGG	5'-ATCTCGAGCAACCAAACCCAAT	pET14b
	CCGGTG-3'	TCATCC-3'	
Xyl3-43b	5'-ATCATATGGATGGCAAATATG	5'-ATGGATCCGAACGCAATGGACA	pET14b
	CCCTTTATACG-3'	TCACC-3'	
Bacon6	5'-ATCATATGAAACAATATCTCA	5'-ATGGATCCTGTCTCCGATGGTGTA	pET14b
	TCTGCC-3'	GTCG-3'	
Endo9	5'-ATCATATGCCATTGTATCAAA	5'-ATGGATCCTGTGGTGAAGCAAG	pET14b
	AACG-3'	ATCAGC-3'	
Xyl8	5'-ATCATATGGTCGAGAAGGTG	5'-ATGGATCCCGAGGCTTTAGCTC	pET14b
	GGCCCGG-3'	AGTTGG-3'	
LysM-8	5'-ATCATATGCTGAGCTTCTTTGC	5'-ATGGATCCCGATTTTCAGGAGC	pET14b
	GGCCG-3'	CATCC-3'	

La actividad enzimática fue medida incubando 100  $\mu$ L de muestra en buffer fosfato 0,05 M, pH 5.5 con el sustrato pNPC (Sigma) a una concentración de 0,5 mM, en un volumen final de 500  $\mu$ L. Esta mezcla se incubó a 50°C durante 30 minutos y la reacción se detuvo por el agregado de igual volumen de NaCO<sub>3</sub> 1 M. La liberación del p-nitrofenil se cuantificó espectrofotométricamente a 410 nm ( $\epsilon$ = 18.3 mM<sup>-1</sup>cm<sup>-1</sup>).



En los casos en que se logró la sobre-expresión de la enzima, se procedió a la inducción en un volumen final de 50 mL de 2xYT en las mismas condiciones en que se obtuvo una buena actividad (tiempo y temperatura) para proceder con la purificación de la proteína. Luego de realizado el cultivo, las células fueron recolectadas por centrifugación a 5000 g durante 15 minutos y el pellet fue congelado a -20°C.

Las células se suspendieron en 5mL en tampón de unión conteniendo urea y luego sonicadas (5 ciclos de 30 segundos alternando con incubaciones en hielo de igual duración) y centrifugadas a 15.000 g durante 15 minutos, conservando el sobrenadante. Se aplicaron 6 mL de la fracción soluble a 1mL de resina de Ni-NTA His-Bind (Quiagen) previamente equilibrada con el buffer nativo. Se incubó 1 hora con agitación suave a 4°C, se eliminó la fracción no retenida en la columna y se realizaron lavados con buffer sin urea con concentraciones crecientes de imidazol (20 mM, 50 mM y 100 mM) en el igual buffer (ver anexo 1). La proteína de interés se liberó de la resina con imidazol 250 mM (ver anexo 1). Se siguió la elución de las proteínas mediante SDS-PAGE en geles de poliacríamida (ver anexo 1), usando el sistema Mini Protean (Bio-Rad) y midiendo actividad enzimática. La concentración total de proteínas se midió utilizando el método el kit Quick Start Bradford Protein Assay (Bio-Rad).

Una vez confirmada la presencia de la enzima activa en el eluido se realizó el cambio de buffer. Para esto se realizaron dos tandas de diálisis de 1 L cada una en buffer de unión. Una vez finalizado se agregó a las suspensión de la proteína DTT 10 mM, PMSF 0,5 mM y EDTA 5 mM. Se agregó glicerol (100 µl cada 500 µl) y se conservaron fracciones a -80°C.

## Resultados y discusión

### Obtención de ADN y construcción de bibliotecas metagenómicas

El rumen está habitado por una gran diversidad de bacterias, acompañadas por hongos y protozoarios anaerobios (197). Gran parte de esta microbiota está especializada en la degradación de lignocelulosa (174,197), presumiblemente debido a la co-evolución de los microorganismos presentes y las variedades vegetales de las cuales se alimentan los rumiantes (175). Es por esto que se construyó la primer metagenoteca a partir de líquido ruminal.

El reactor del cual se tomó una muestra para la construcción de la segunda biblioteca metagenómica, fue alimentado con contenido ruminal de origen vacuno, residuos de descarte de curtiembre y lodo biológico secundario de una planta aceitera. Debido a que el contenido ruminal tiene una alta concentración de bacterias fermentativas, pero no contiene metanogénicas acetoclásticas en concentración significativa, el reactor fue inoculado con lodo metanogénico. Se utilizó para esto lodo proveniente de un reactor UASB empleado para el tratamiento de efluente de maltería (186).

El primer objetivo planteado es la selección y caracterización de enzimas novedosas implicadas en la degradación de lignocelulosa, por lo cual se seleccionaron dichos ambientes para la construcción de las bibliotecas metagenómicas. El primer paso fue fragmentar las muestras con el fin de obtener la fracción enriquecida en células procariotas. La porción obtenida a partir del gradiente de Percoll estuvo constituida mayoritariamente por bacterias Gram negativas y no se encontraron levaduras u otros microorganismos que pudieran afectar la extracción de ADN.

A partir de los dos ambientes fue posible obtener ADN metagenómico de calidad para la construcción de las dos metagenotecas. Este ADN fue clonado en el vector pCC1FOS<sup>TM</sup> y utilizado para la transformación de la cepa de *E. coli* Epi300-T1<sup>R</sup>. Se obtuvieron 30.000 clones para la biblioteca construida a partir de líquido ruminal y 100.000 para la biblioteca obtenida a partir de lodos de digestor anaerobio. Cada uno de estos clones alberga un fragmento de ADN metagenómico de 42 Kpb en promedio, lo que representa un total de 1,2 Gpb y 4,2 Gpb para la biblioteca de rumen y del digestor, respectivamente.

Las bibliotecas funcionales construidas a partir de contenido ruminal de diversos animales varían mucho en su tamaño, aunque generalmente no se reporta una cobertura tan grande como la obtenida en este trabajo (Tabla 6). El uso de fósmidos como vector de expresión permite la detección de operones o clúster de genes implicados en una actividad microbiana, bajo el control de promotores nativos. El clonado de grandes fragmentos de ADN permite tener más información que los insertos

pequeños, por lo tanto, menos clones dentro de una biblioteca deberían someterse a las pruebas de selección para las actividades de interés. A su vez, el origen de replicación basado en el factor F de *E. coli* presente en cósmidos, BACs y fósidos, permite el mantenimiento estable de grandes fragmentos de ADN incluso en presencia de genes tóxicos. La activación selectiva del aumento de número de copias (mediante el origen de replicación *oriV*) asegura la presencia de aproximadamente 100 copias del ADN exógeno por célula. Este alto número de replicones permite un aumento de la frecuencia de la detección de genes funcionales, incluso aunque sean débilmente expresados (133).

**Tabla 6.** Características de metagenotecas de rumen evaluadas para actividades celulolíticas

Origen o ambiente	Tipo de librería	Inserto (Kpb)	Nº clones	Biblioteca (Mpb)	Clones con actividad Celulasa	Clones con actividad Xilanasas	Genes positivos/Mpb	Referencia
Líquido ruminal	Fósidos	42	30.000	1200	43	18	1/20	Este trabajo
Digestor anaerobio	Fósidos	42	100.000	4200	5	2	1/600	Este trabajo
Contenido rumen bovino	Plásmidos	15	6.000	90	1		1/90	(198)
Contenido rumen bovino	Fagos		14.000	77	7		1/50	(199)
Líquido ruminal	Fagos	3	50.000	150	1		1/150	(200)
Contenido rumen búfalo	Cósmidos	35	15.000	525	61		1/8.6	(182)
Rumen de yak	Cósmidos	ND	4.000	--	1		--	(152)
Líquido ruminal	Fósidos	40	20.000	700	1		1/700	(201)

### Evaluación en medios indicadores

Las metagenotecas fueron evaluadas en medio rico conteniendo carboximetil-celulosa (CMC) o xilano de avena y en medio mínimo con avicel, con el fin de seleccionar clones con actividad celulolítica (CMCasas o Avicelasas) y xilanasas. El CMC es un sustrato utilizado comúnmente para la detección de actividades de endo-celulasas (139). La degradación del mismo es fácilmente visualizada luego de la tinción de las placas con rojo congo, gracias a la formación de halos claros alrededor de las colonias con actividad positiva (140). El avicel, por su parte, es una forma de celulosa microcristalina.

Durante su preparación con ácido diluido se elimina la hemicelulosa y la mayoría de la celulosa amorfa (7). Al evaluarse la utilización de este sustrato en medio mínimo, se presume que los clones capaces de crecer en avicel tendrán, además de hidrolasas, los sistemas de transporte necesarios para secretar a estas enzimas al medio extracelular y para internalizar a los glúcidos liberados.

El componente principal de la hemicelulosas, el xilano, es un heteropolisacárido complejo con un esqueleto de  $\beta$ -1,4-xilosa rica en cadenas laterales de arabinosa. Las endo-xilanasas (o simplemente xilanasas) son las enzimas que atacan la estructura del esqueleto principal por escisión aleatoria de enlaces internos, mientras que sus productos son hidrolizados adicionalmente por  $\beta$ -xilosidasas. Sin embargo, para acceder a la estructura primaria rica en arabinosa, se requieren enzimas adicionales, por ejemplo,  $\alpha$ -L-arabinofuranasas (202). De todos modos, el xilano es un sustrato ampliamente utilizado para la detección y caracterización de enzimas implicadas en la degradación de la hemicelulosa (142). El xilano derivado de madera de haya es muy insoluble (aproximadamente un 95%), está compuesto por un 87.7% de xilosa, pequeñas cantidades de glucosa, trazas de arabinosa y galactosa, y un 3% de ácido ascórbico (143).

Utilizando estos sustratos se seleccionaron 58 clones positivos que presentaron actividad en placas de LB o medio mínimo (Tabla 7). Las aproximaciones de metagenómica funcional tienen diversas limitantes que pueden determinar la cantidad de clones con actividad detectados. Obviamente, los sustratos seleccionados para realizar los screening son un factor determinante de las actividades que pueden o no ser detectadas. Cómo se ve en los resultados, la mayoría de los clones positivos para "celulasas" tuvieron actividad en CMC, un sustrato soluble, lo cual no asegura su actividad en celulosa cristalina. Incluso se ha visto que algunos microorganismos pueden degradar CMC gracias a la expresión de una mezcla de  $\beta$ -glucanasas (7,139). Otro obstáculo puede deberse a la necesidad de co-factores, o que la actividad enzimática requiera una temperatura y o pH diferentes a los empleados en el screening (133).

Cómo se puede observar la mayoría de los clones con actividad positiva para celulasas o hemicelulasas fue obtenida en la biblioteca de rumen, aunque ésta fue la que contuvo el menor número de clones. A su vez, en ésta también se obtuvieron clones con actividad en otros sustratos. Se destaca la obtención de 10 clones con actividad esterasa/ lipasa, una de ellas perteneciente a una familia recientemente descrita de enzimas lipolíticas, caracterizada por nuestro grupo de trabajo (158) y tres clones con actividad antimicrobiana.

**Tabla 7.** Clones positivos detectados en las evaluaciones de las metagenotecas

Actividad	Medio de selección	Clones positivos para biblioteca de rumen	Clones positivos para biblioteca de lodo
Celulasa	LB-CMC, revelado con Rojo Congo	31	3
	Avicel, crecimiento en medio mínimo		2
Hemicelulasa	LB-Xilano, revelado con Rojo Congo	6	2
Celulasa/ Hemicelulasa	LB-CMC, LB-Xilano, revelado con Rojo Congo	12	2
Total		<b>49</b>	<b>9</b>
			<b>58</b>

Para explicar la diferencia en el número de clones positivos entre las dos bibliotecas se plantean tres posibles razones:

1) los grupos de clones generados para la biblioteca obtenida a partir de lodos del reactor anaerobio contienen aproximadamente 250 clones cada uno, mientras que los de rumen poseen sólo 10. Seguramente el alto número de clones en cada grupo sembrado en la placa de “screening” dificulte la visualización de las actividades de interés en los medios de selección. La metodología utilizada para identificar el clon responsable del fenotipo observado para un grupo consistió en re-aislar los integrantes del grupo y volver a analizarlos en los medios de selección. El alto número de clones por grupo en la biblioteca del reactor anaerobio dificultó, en algunos casos, la identificación del clon responsable del fenotipo observado para el grupo, ya que era necesario repetir la selección para más de 250 clones.

2) La baja frecuencia de detección de clones positivos en la biblioteca de lodos anaerobios, también puede atribuirse a que *E. coli* posee limitaciones para la expresión heteróloga de genes foráneos si su origen es filogenéticamente distante (133,138,203). En general, la conversión metanogénica de residuos de celulosa en un sistema de digestión húmeda es llevada a cabo por bacterias pertenecientes al filo *Firmicutes*, siendo los degradadores de celulosa dominantes en estos ambientes *Clostridium thermocellum*, *Clostridium stercorarium*, *Bacteroides cellulosolvens*, *Halocella cellulosilytica* y *Acetivibrio spp* (204,205).

La transcripción de genes exógenos requiere del reconocimiento de las señales de transcripción de los mismos por la maquinaria transcripcional del hospedero (206). A su vez, el inicio de la traducción

depende de la presencia de sitios de unión ribosomal en el ADN foráneo adecuadas (207). El uso de codones es uno de los factores a los que se ha atribuido la traducción ineficiente de proteínas recombinantes, debido a las diferencias en la frecuencia de uso en el hospedero y el microorganismo del cual proviene el ADN, principalmente si estos son filogenéticamente distantes (133,207). El correcto plegado de las proteínas (208), la agregación en cuerpos de inclusión (207) y la secreción al espacio extracelular (209) son considerados factores que pueden condicionar el éxito de la expresión heteróloga de proteínas.

3) La comunidad microbiana presente en el reactor anaerobio está enriquecida en bacterias metanogénicas acetoclásticas debido al inóculo utilizado. Contrariamente, el rumen es rico en microorganismos especializados en la degradación de biomasa lignocelulósica (114,153), por lo cual era probable obtener más clones con actividad celulolítica/ xilanolítica en este ambiente.

### **Caracterización fenotípica**

Los 58 clones seleccionados fueron sometidos a una serie de ensayos complementarios para profundizar en su caracterización fenotípica y reducir el número de clones con los cuales trabajar. Para esto, se evaluó su capacidad de crecer en medio mínimo con avicel o papel de filtro como fuente de carbono. Sin embargo, estos sustratos no contienen lignina por lo que no constituyen un sustrato lignocelulósico. Además, no representan una materia prima relevante desde el punto de vista del proceso industrial de una biorrefinería. Por estas razones, se realizaron también ensayos empleando un sustrato de mayor valor agroindustrial como es el bagazo de caña de azúcar. Estos experimentos mostraron que los fósmidos Csd4, Csd23 y M9F1 son necesarios y suficientes para viabilizar el crecimiento de *E.coli* en papel de filtro y bagazo de caña como única fuente de carbono (Tabla 8).

Como se comentó antes, el CMC es un sustrato específico para endocelulasas (139). Para dilucidar si los clones con esta actividad también poseían actividad exocelulasa se utilizó el sustrato 4-MUC (210), el cual permite, de manera sensible, detectar clones con esta actividad gracias a la liberación del grupo fluorogénico 4-metilumbeliferona. De los 58 clones estudiados 5 mostraron actividad 4-MUCasa en extractos celulares totales (Tabla 8), mientras que dos, Epi (Csd4) y Epi (Csd18), fueron capaces también de hidrolizar dicho sustrato con células completas.

**Tabla 8.** Características fenotípicas de clones seleccionados

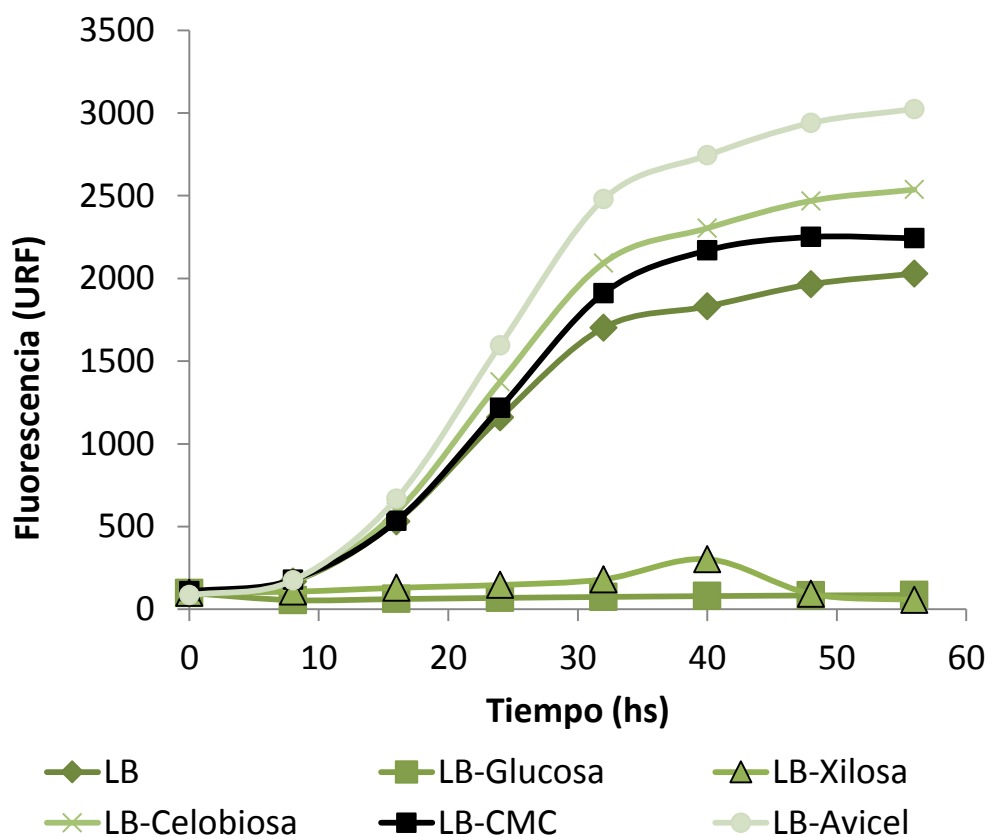
Fósmido	Actividad hidrolítica detectada en			Crecimiento en medio mínimo <sup>c</sup>			
	CMC <sup>a</sup>	Xilano <sup>a</sup>	4-MUC <sup>b</sup>	Avicel	PF	Celobiosa	Bagazo de caña
Csd4	++	+	Extracto /Células completas	+	+	++	+
Csd6	+	-	Extracto	-	-	-	-
Csd8	+	-	Extracto	-	-	-	-
Csd9	++	-	Extracto	+	+	-	-
Csd18	+	-	Extracto /Células completas	-	-	+	+
Csd23	+++	-	-	+	+	-	+
Xil3	-	+	-	-	Nd	Nd	Nd
Xil5	-	+	-	-	Nd	Nd	Nd
M9F1	-	-	-	+	+	++	+

<sup>a</sup> Halo de degradación visualizado en placas de LB suplementado con el sustrato. - sin actividad; + 1 a 4 mm, ++ 4 a 10 mm; +++ más de 10 mm de halo alrededor de la colonia.

<sup>b</sup> Degradación del sustrato 4-β-D-metilumbeliferil-celobiosido en extractos (células lisadas) o en cultivos de LB (células completas) luego de 1 hora a 37°C.

<sup>c</sup> Crecimiento del clon correspondiente en M9- Cm<sub>12,5</sub> suplementadas con 1% de avicel o bagazo de caña o con una tira de papel (PF) (40 x 100 mm) o celobiosa al 1%. - sin actividad; + crecimiento después de 5 días de incubación a 37°C ; ++ crecimiento después de 3 días de incubación a 37°C y 150 rpm; Nd. no determinado.

Con los clones Epi (Csd4) y Epi (Csd18) se hicieron ensayos complementarios para indagar en la regulación de la expresión de las enzimas involucradas en la actividad exocelulasa (4-MUCasa). Para esto, se realizaron cultivos en LB- Cm<sub>12,5</sub> suplementado con 4-MUC y glucosa, xilosa, celobiosa, avicel o CMC. Para el clon epi (Csd18) se observó que la presencia de glucosa o xilosa determina la pérdida completa de la actividad. Mientras que el agregado de celobiosa, CMC o avicel no afecta la actividad (Figura 11). Los resultados de este ensayo para el clon Epi300 (Csd4) se presentan en el capítulo 3. Por el diseño del experimento no es posible distinguir si el efecto negativo de estos sustratos se deba a la inhibición a nivel de la expresión de los genes que codifican para las enzimas de interés o si la glucosa y la xilosa actúan como inhibidores alostéricos de las mismas. Se ha reportado que la glucosa y la celobiosa, los productos finales de la degradación de la celulosa, actúan como inhibidores de diversas celulasas (7,211). Pero no se encontraron reportes de que la xilosa sea capaz de inhibir la actividad de las exoglucanasas. Seguramente, el efecto negativo de estos dos azúcares simples se deba a la regulación en la expresión de los genes que no permite la síntesis de enzimas involucradas en la degradación de sustratos complejos teniendo disponibilidad de azúcares simples.



**Figura 11.** Actividad 4-MUCasa in vitro del fósido Csd18 en células de epi300. 200  $\mu$ L de LB-Cm<sub>12,5</sub>-0.05% 4-MUC se inocularon (DO<sub>i</sub>= 0.05) e incubaron a 37°C durante 60 horas, midiendo la fluorescencia (excitación 365 nm, emisión 445 nm) cada 8 horas. Todos los ensayos se hicieron por triplicado. Glucosa, xilosa, celobiososa, CMC o avicel fue adicionado a una concentración final de 0,5% (p/v).



## Análisis in silico de las secuencias e identificación de genes

En base a la caracterización fenotípica presentada en las secciones anteriores, se seleccionaron nueve fósidos (listados en la tabla 6) los cuales fueron secuenciados completamente utilizando los servicios de MacroGen o la plataforma de secuenciación masiva del IIBCE. Los marcos abiertos de lectura (ORFs) fueron predichos mediante el algoritmo MetaGene en el servidor en línea WebMGA y anotados mediante BLASTp utilizando la base de datos del NCBI. Éstos se listan en el Anexo 3. Para cada una de las proteínas de interés se realizó un análisis de asignación de homólogos por similitud de secuencia utilizando la base de datos de proteínas del NCBI (algoritmo BLASTp); con las secuencias más similares se contruyeron alineamientos mediante Mega 6.0 (ver anexo 4). A su vez, par cada una de las secuencias de aminoácidos de las enzimas de interés se realizó la búsqueda de dominios conservados de glicosil hidrolasas (GH), glicosiltransferasas (GT) o dominios de unión a carbohidratos (CBM) mediante dbCAN, CAT y CD-Search del NCBI.

### M9F1

Los resultados obtenidos para el fósido M9F1 se presentaron en la publicación incluida en el capítulo 1. Complementariamente se realizó una comparación la región identificada que codifica para la sub unidad del ARNr 16S con la base de datos de secuencias de ARN ribosomal 16S (tanto de Bacteria como de Archaea) usando Megablast en el NCBI. Este análisis de blastn mostró que este gen posee homología con genes ribosomales de organismos no cultivables provenientes de diversos reactores anaerobios (Tabla 9), lo cual es coherente con el origen del fósido M9F1. Uno de los homólogos proviene de un reactor especializado en la digestión anaerobia seca de lodos para la metanización de celulosa (204). Mientras que otro fue identificado en un reactor anaerobio termofílico alimentado con basura doméstica orgánica mezclada con residuos de jardín (212).

**Tabla 9.** Genes que codifican para la sub unidad ribosomal 16S realcionados al detectado en M9F1

GeneBank ID	Organismo [Origen]	E-value	Identidad (%)	Referencia
<a href="#">FJ825457.1</a>	Bacteria no cultivable [planta mesófila de biogas]	0	99	No publicado
<a href="#">JF417934.1</a>	Bacteria no cultivable [digestor metanogénico]	0	97	(204)
<a href="#">AM947547.1</a>	Bacteria no cultivable [reactor anaerobio termófilo]	0	97	(212)
<a href="#">GQ134349.1</a>	Bacteria no cultivable [digestor anaerobio]	0	98	No publicado
<a href="#">CP000448.1</a>	<i>Syntrophomonas wolfei</i> subsp. wolfei	0	94	
<a href="#">NR_122058.1</a>	<i>Syntrophomonas wolfei</i> strain Goettingen G311	0	94	
<a href="#">KF550879.1</a>	Bacteria no cultivable [digestor anaerobio de residuos de alimentos]	0	94	No publicado

## Csd23

Dentro de la secuencia del fósido Csd23 se identificó un gen que codifica para una proteína ribosomal. Los análisis de homología del mismo en GreenGene sugieren que Csd23 proviene de una bacteria del filo Bacteroidetes, siendo posiblemente una Sphingobacteria relacionada con *Pontibacter*, *Olivibacter* o *Pedobacter*. La comparación de este gen con la base de datos de nucleótidos (nr/ nt) del NCBI utilizando el algoritmo Megablast mostró que el mismo posee una alta identidad de secuencia con genes anotados a partir de bacterias no cultivadas de origen ruminal o intestinal (Tabla 10). Destaca la coherencia entre el ambiente del cual fue obtenido este fósido y el origen de las secuencias más similares.

El resto de la secuencia del fósido no mostró homología con los genomas depositados en las bases de datos.

**Tabla 10.** Genes que codifican para una proteína ribosomal relacionados al identificado en Csd23

GeneBank ID	Organismo [Origen] <sup>(a)</sup>	E-value	Identidad (%)	Referencia
<a href="#">EU472286.1</a>	Bacteria no cultivable [intestino mamífero]	0	98	(213)
<a href="#">AB746584.1</a>	Bacteria no cultivable [rumen bovino]	0	96	No publicado
<a href="#">KC163072.1</a>	Bacteria no cultivable [rumen de yac]	0	93	No publicado
<a href="#">AB616493.1</a>	Bacteria no cultivable [rumen vacuno]	0	94	No publicado
<a href="#">JX218522.1</a>	Bacteria no cultivable [rumen de buey]	0	93	No publicado

<sup>(a)</sup> Asignación de homólogos realizada mediante el algoritmo megablast utilizando la base de datos de nucleótidos (nr/nt) del NCBI.

A nivel funcional, dentro de los 23 ORFs anotados en el fósido Csd23 (Tabla 8 en Anexo 3) no se pudo identificar uno como responsable del fenotipo conferido por este fósido. Los ORFs Csd23\_1, 5, 12, 14, 22 y 23 corresponden a proteínas hipotéticas, todos pertenecientes al filo Bacteroidetes, tal vez relacionadas con las actividades de interés.

Los ORFs restantes están relacionados con la modificación del ADN (ORFs Csd23\_2, 6, 15 y 16), el metabolismo de pentosas (Csd23\_8), el metabolismo de aminoácidos (Csd23\_2, 7, 18, 19, 20 y 21), la modificación de nucleótidos (Csd23\_4) y el procesamiento de los ARN mensajeros (Csd23\_10).

## Csd4

Como se presentará en profundidad el capítulo 3, en el fósido denominado Csd4 se identificaron genes que codifican para enzimas con homología con diversas glicosil hidrolasas,

posiblemente responsables de los fenotipos observados en este clon (ver Tabla 8). La proteína codificada por el ORF Csd4\_23, denominada EndoG, posee un dominio perteneciente a la familia GH5 (Tabla 11 y Figura 12). En Csd4, además de una endoglucanasa (descrita en profundidad en el capítulo 4) se identificaron una xilanasa (codificada por el ORF Csd4\_16) y una lacasa (ORF Csd4\_22). La xilanasa, denominada Xyl4 posee un dominio de la familia GH 43 (Tabla 9 y Figura 12). Esta familia es de las más abundantes en el rumen bovino tanto en la fracción líquida como la adherida a las fibras, junto con las familias GH 2, GH 3, GH 31 y GH 51 (175). Estas familias presentan actividad sobre las cadenas laterales de la hemicelulosas y las pectinas y sobre los extremos no reductores de glucanos de cadena corta (únicamente GH 2).

La proteína codificada por el ORF Csd4\_22, denominada Lac4, presenta un dominio conservado Cu-oxidasa\_4 (pfam02578). Las lacasas son multi-cobre oxidoreductasas capaces de oxidar un amplio rango de compuestos fenólicos y no fenólicos que están ampliamente distribuidas entre procariontes y eucariotes (214). Estas enzimas forman parte de las enzimas modificadoras de lignina.

## Csd6

En el fósido Csd6 se identificó un operón implicado en la degradación de polisacáridos (Figura 13). En él se pueden reconocer dos genes que codifican para  $\alpha$ -amilasa (ORFs Csd6\_c1.24 y Csd6\_c1.26) con dominios pertenecientes a la súper familia AmyAc (cl07893), una celulasa (ORF Csd6\_c1.25) y una  $\beta$ -1,4-mananasa con un dominio conservado GH 26 (ORF Csd6\_c1.26). Las enzimas de la familia de  $\alpha$ -amilasa actúan sobre el almidón, el glucógeno y oligo- y polisacáridos. Estas proteínas catalizan la ruptura de los enlaces glucosídicos  $\alpha$ -1,4 y  $\alpha$ -1,6 con retención del centro anomérico. Las glicosil hidrolasas de la familia 26 son principalmente endo- $\beta$ -1,4-mananasas, aunque recientemente se describió una con acción exo- $\beta$ -mananasa. La familia también contiene enzimas que muestran actividad  $\beta$ -1,3: 1,4-glucanasa y  $\beta$ -1,3-xilanasa.

La celulasa identificada en la secuencia del fósido Csd6 (codificada por el ORF Csd6\_c1.25), denominada Bacon6, posee un dominio de unión a carbohidratos N-terminal de Bacteroidetes (BACON, "Bacteroidetes-Associated Carbohydrate-binding Often N-terminal"). Este dominio de unión, descrito a partir de análisis de secuencias metagenómicas (215), se ha visto asociado a enzimas que degradan carbohidratos y a proteasas. Los análisis realizados mediante dbCAN y el servidor BESC infieren, a su vez, la presencia de un dominio catalítico GH 5 en la secuencia de este ORF (Figura 12).

**Tabla 11.** ORFs identificados como codificantes para enzimas con actividad celulasa/ xilanasas

Fósmido	ORF	Enzima codificada (nombre)	COG <sup>(a)</sup>	Dominio GH <sup>(b)</sup>			Anotación CAZY <sup>(c)</sup>		
				Familia	Cobertura	E-value	Dominio Pfam	GenBank ID	Organismo
Csd 4	Csd4_23	Endoglucanasa (EndoG)	2730	GH 5	0.8255	2.10E-22	Cellulase	<a href="#">AGK74985.1</a>	uncultured organism
	Csd4_16	Xylosidasa/arabinosidasa (Xyl4)	3507	GH 45	0.9435	6.30E-65	Glyco_hydro_43	<a href="#">ADV43017.1</a>	<i>Bacteroides helcogenes</i> ATCC 35417
	Csd4_22	Lacasa tipo YfiH (Lac4)	1496	NC					
Csd 6	Csd6_c1.25	Celulasa (BACON6)	2730	GH 5	0.3855	2.10E-22	Cellulase	<a href="#">ADY35478.1</a>	<i>Bacteroides salanitronis</i> DSM 18170
Csd 8	Csd8_c5.2	Beta-xylosidasa (Xyl8)	3507	GH 43	0.6895	4.20E-32	Glyco_hydro_43	<a href="#">ADE83588.1</a>	<i>Prevotella ruminicola</i> 23
				CBM6	0.6667	0.00011			
Csd 9	Csd9_c1.5	Endoglucanasa (Endo9)	2730	GH 5	0.8145	6.30E-40	Cellulase	<a href="#">AFN57695.1</a>	uncultured bacterium r_05
Csd 18	Csd18_c1.8	Endoglucanasa (Endo18)	2730	GH 5	0.7127	1.20E-43	Cellulase	<a href="#">AGK74985.1</a>	uncultured organism
Xil 3	Xil3_c1.8	Glicosil hidrolasa (Xil3-43a)	2152	GH 130	0.8716	6.30E-91	DUF377	<a href="#">CBK63193.1</a>	<i>Alistipes shahii</i> WAL 8301
	Xil3_c1.25	Glicosil hidrolasa (Xil3-43b)	2152	GH 130	0.5169	1.10E-49	DUF377	<a href="#">ADB80102.1</a>	uncultured microorganism
	Xil3_c1.15	Celulasa	2730	GH 5	0.7818	8.90E-39	Cellulase	<a href="#">ADY35478.1</a>	<i>Bacteroides salanitronis</i> DSM 18170
	Xil3_c1.5	Glicosil hidrolasa	4124	GH 26	0.9274	5.40E-59	Glyco_hydro_26	<a href="#">CBK67266.1</a>	<i>Bacteroides xylanisolvens</i> XB1A

<sup>a</sup> Asignado por MetaGene

<sup>b</sup> Asignado por dbCAN. NC, No corresponde.

<sup>c</sup> Anotación utilizando el servidor BESC Data & Tools (<http://bobcat.ornl.gov/besc/index.jsp>). Algoritmo usado: anotación de secuencia & Pfam (Bit Score=55, rules support=40, E-value=0.01). La asignación de homólogos se hace frente a la base de datos de CAZy ([www.cazy.org](http://www.cazy.org)), por lo cual se ven diferencias con las secuencias más similares asignadas mediante BLASTp (ver Anexo 4).

## Csd8

Por otra parte, aunque el fósforo Csd8 confiere a *E. coli* la capacidad de degradar CMC y 4-MUC, en su secuencia no se identificó ningún ORF con homología con endo- o exo-celulasas depositadas en las bases de datos. Sorprendentemente, sí se identificó un gen (ORF Csd8\_c5.2) que codifica para una  $\beta$ -xilosidasa, denominada Xyl8, con un dominio catalítico GH 43. Los análisis de la secuencia de esta proteína mostraron que posee también un dominio CBM6 (Tabla 11, Figura 12). Los CBM son dominios no catalíticos que facilitan la unión del dominio catalítico (GH) al sustrato insoluble. Los dominios CBM de la familia 6 por lo general están junto a dominios catalíticos GH 3, GH 11 y GH 43. Los CBM6 representan una quimera de dos sitios de unión con distintos modos de unión: el sitio I dentro de las regiones "loop" y el sitio II, en la cara cóncava del pliegue beta. La combinación entre dominios catalíticos GH 43 y de unión CBM6 se ha visto en *Fibrobacter succinogenes* y en *Penicillium purpurogenum* (216,217). Los CBM, al igual que las doquerinas y las cohesinas, son poco abundantes en el rumen bovino lo cual infiere que la degradación de celulosa mediante celulosomas es poco habitual en este ambiente (175).

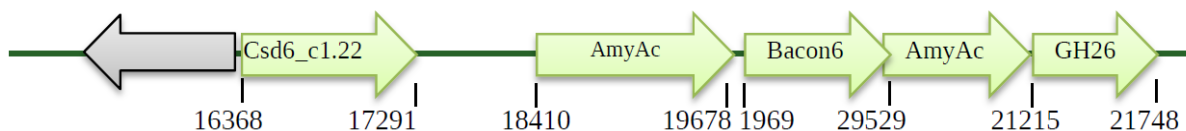
Acompañando al ORF que codifica para Xyl8, se identificó una hidrolasa asociada a la pared celular (ORF Csd8\_c5.4) denominada LysM-8, (Tabla 3 en Anexo 3) que codifica para una proteína hipotética perteneciente a la familia NLP/P60 (pfam00877). Las proteínas NLP/P60 definen una familia de peptidasas de pared celular que están ampliamente representados en varios linajes bacterianos. Actualmente los miembros caracterizados son conocidos por hidrolizar D- $\gamma$ -glutamil-meso-diaminopimelato o uniones N-acetilmuramato-L-alanina (218). Hacia el extremo C-terminal, LysM-8 posee un dominio Lisina de 33 aminoácidos (cd00118) o CBM50 (Figura 12), reportado como involucrado en la unión a peptidoglicanos en bacterias y a quitina en eucariotas. Este dominio se identificó originalmente en enzimas que degradan las paredes celulares bacterianas, pero diversas proteínas implicadas en otras funciones biológicas también contienen este dominio (219). Se ha reportado que los dominios Lisina pueden actuar como una señal para el reconocimiento específico en la interacción planta-bacterias en la patogénesis bacteriana y en la simbiosis con rizobios (220).

## Csd18

En el fósforo Csd18, que confiere a *E. coli* la habilidad de degradar CMC, 4-MUC y crecer en medio mínimo con celobiosa o bagazo de caña de azúcar como fuente de carbono, se identificó un ORF, Csd18\_c1.8, que codifica para una enzima putativa implicada en la degradación de carbohidratos. Esta enzima, denominada Endo18, posee un dominio catalítico GH 5 (Tabla 11 y Figura 12). Los homólogos más cercanos son enzimas identificadas mediante aproximaciones de metagenómica en el rumen bovino (153,201).



**Figura 12.** Posición relativa de los dominios encontrados en la secuencia de aminoácidos de las proteínas seleccionadas para estudiar en profundidad. Los dominios catalíticos de glicosil hidrolasas (GH) se muestran en verde, mientras que los dominios de interacción con carbohidratos se muestran en rojo.



**Figura 13.** Esquema de la región final del contig 1 del fósforo Csd6 en donde se infiere la presencia de un operón implicado en la degradación de polisacáridos. Las flechas representan los ORFs, e indican su posición relativa y su sentido. Según el servidor Genome2D, existe un promotor en la región 18383-18414 (Score= 8.9), no se detectaron promotores en la región interna del operón.

## Xil3

En el fósido Xil 3, seleccionado por conferir a *E. coli* la habilidad de degradar xilano, se detectaron varios ORFs que codifican para enzimas involucradas en la degradación de celulosa y hemicelulosa (Tabla 7 del Anexo 3). A su vez, parece haber un operón implicado en la degradación de diversos polisacáridos (a partir de la posición 28490 pb) formado por una proteína de unión a almidón, una de unión a colágeno, un transportador de azúcares y una glicosidasa.

Las glicosil hidrolasas detectadas en la secuencia de Xil3 fueron una celulasa, con dominio catalítico GH 5 (ORF Xil3\_c1.15) y una glicosil hidrolasa GH 26 (ORF Xil3\_c1.5) con homología a una  $\beta$ -mananasa de *Alistipes sp.*, y dos glicosidasas no caracterizadas (codificadas por los ORFs Xil3\_c1.8 y Xil3\_c1.25). Aunque estos 4 ORFs no están formando un operón, observando el origen de las secuencias más similares a este fósido, parece evidente que este fragmento de ADN proviene de un organismo especializado en la modificación de polisacáridos complejos probablemente perteneciente al filo Bacteroidetes que aún no se ha descrito.

La anotación de las enzimas codificadas por los ORFs Xil3\_c1.8 y Xil3\_c1.25, denominadas X3-43a y X3-43b respectivamente, reveló que ambas podrían pertenecer a la familia Pfam DUF377 (COG2152). Esta familia, de función desconocida, está presente en muchas proteínas hipotéticas, algunas de las cuales se prevé que sean glicosil hidrolasas (221). Según la comparación con la base de datos de dominios conservados del NCBI, ambos tienen dominios de la super familia GH 43\_62\_32\_68 (cl14647). Mientras que, la anotación de estos ORFs utilizando BESC o dbCAN asigna sus dominios a la familia GH 130 (Tabla 11, Figura 12). Los miembros de las familias de glicosil hidrolasas 32, 43, 62 y 68 poseen una arquitectura similar (un dominio de 5 hojas beta). La mayoría de las enzimas GH 43 muestran tanto actividad  $\alpha$ -L-arabinofuranosidasa y  $\beta$ -D-xilosidasa, usando aril-glicósidos como sustratos. La familia GH 62 contiene  $\alpha$ -L-arabinofurasidasas (EC 3.2.1.55) que cortan tanto  $\alpha$ -1,2 o  $\alpha$ -1,3-L-arabinofuranósidos de las cadenas laterales del xilano. GH 32 comprende sacarosa-6-fosfato hidrolasas, invertasas, inulinasas, levanasas fructosiltransferasas y fructanotransferasas, mientras que la familia GH 68 está compuesta por fructosiltransferasas.

Por su parte, la familia GH 130 contiene fosforilasas que catalizan la fosforólisis de uniones  $\beta$ -manosídicos en el extremo no reductor de los sustratos. Las actividades reportadas para esta familia son 4-O- $\beta$ -D-mannosil-D-gluco fosforilasa de *Bacteroides fragilis* (222),  $\beta$ -1,4-mannooligosaccárido fosforilasa de *Ruminococcus albus* (223), 2- $\beta$ -oligomanan fosforilasa y  $\beta$ -1,2-mannobiosa fosforilasa de *Thermoanaerobacter sp. X-514* (224). Aunque es posible inferir que estos genes sean los responsables del fenotipo observado, la falta de homólogos caracterizados no permite discernir a cuál de las dos familias pertenecen. La caracterización exhaustiva de las proteínas codificadas permitirá esclarecer la naturaleza de las mismas y su rango de sustratos.

## Csd9

El ORF Csd9\_c1.5, del fósido Csd9 codifica para una posible celulasa, denominada Endo9. La asignación de dominios conservados mediante dbCAN mostró que posee un dominio catalítico perteneciente a la familia GH 5 (Tabla 9 y Figura 12). Tanto los estudios de homología utilizando la base de datos del NCBI o de CAZy, mostraron similitud de secuencia con celulasas no caracterizadas de origen ruminal (Tabla 11 y Tabla 4 en Anexo 3). Se han reportado un gran número de celulasas con dominios GH 5 y GH 9 obtenidos mediante aproximaciones de metagenómica funcional a partir de suelo, rumen e intestino de termitas (122,182,199,201,225). La predominancia de estos dominios puede que se deba a que estas celulasas serían más fáciles de expresar en *E. coli* o a que sean las más abundantes en los genomas bacterianos (122). En el microbioma del rumen las enzimas pertenecientes a las familias GH 5 y GH 9 están presentes pero son poco abundantes (175).

## Origen filogenético de los fósidos

A nivel de nucleótidos no fue posible asignar un origen filogenético a ninguno de los fragmentos de ADN clonados y secuenciados. Como se comentó, solo en los fósidos M9F1 y Csd23 se identificaron secuencias que codifican para proteínas ribosomales, pero el resto de la secuencia de ADN de éstos no mostró homología con genomas de las especies bacterianas inferidas para estos fósidos.

Los análisis de homología de las proteínas codificadas en los restantes fósidos secuenciados permiten suponer que estarían vinculados a *Prevotella* (Csd4, Csd9), *Alistipes* (Xil3) y *Bacteroides* (Csd8, Csd18). Estos tres grupos se han reportado como ubicuos en las heces de ganado (226,227), y en el líquido ruminal (227). La dificultad a la hora de cultivar representantes del filo Bacteroidetes, hacen que no se cuente con suficiente información genómica de este grupo (228), por lo cual se podría suponer que estos fragmentos de ADN provienen de organismos no descritos aún.

La comparación de la frecuencia en el uso de tetranucleótidos es una aproximación que permite inferir si un fragmento de ADN está emparentado al genoma de un microorganismo determinado. Utilizando TETRA (229) se determinó el z-score (que representa una medida para la relación en el uso de tetranucleótidos de fragmentos genómicos, Figura 14) para los distintos fósidos analizados y los genomas de *P. ruminicola*, *Bacteroides fragilis* y *Alistipes finegoldii*, los candidatos que surgieron como probables organismos parentales para los fósidos en los análisis de asignación de ORFs (Tabla 12).

Aunque se considera menos robusto que la frecuencia en el uso de tetranucleótidos o de codones (230), el contenido CG de una secuencia de nucleótidos también puede ser indicativo del origen de la misma, el porcentaje de dichos nucleótidos en las secuencias analizadas se muestran en la Tabla 12.



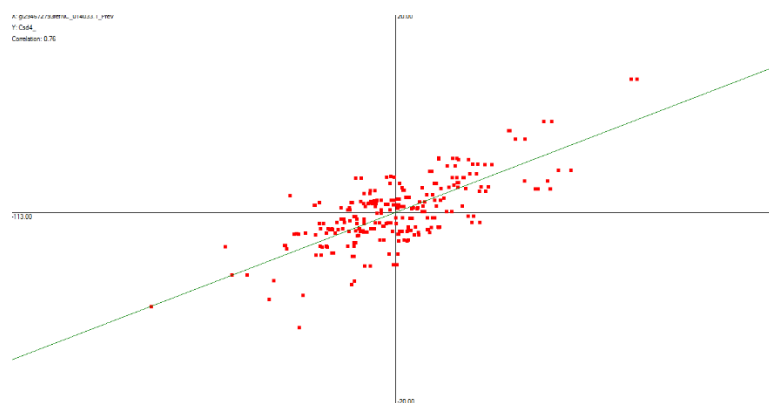
**Tabla 12.** Contenido GC y comparación de la frecuencia del uso de tetranucleótidos de los fósmidos caracterizados

Fósmido	GC % <sup>a</sup>	Coeficiente de correlación de Z-score <sup>b</sup>			Origen de los ORF asignados <sup>c</sup>
		<i>P. ruminicola</i>	<i>Bacteroides</i>	<i>Alistipes</i>	
Csd4	49	0,76	0,57	0,54	<i>Prevotella</i> (21/27) <i>Bacteroides</i> (5/27)
Csd6	58	0,53	0,4	0,74	<i>Prevotella</i> (6/36) <i>Bacteroides</i> (8/36) <i>Alistipes</i> (6/36)
Csd8	63/55	0,4	0,18	0,84	<i>Prevotella</i> (2/37) <i>Bacteroides</i> (25/37) <i>Alistipes</i> (4/37)
Csd9	49	0,75	0,71	0,54	<i>Prevotella</i> (19/32)
Csd18	48	0,55/ 0,08	0,37/0,27	0,26	<i>Prevotella</i> (7/27) <i>Bacteroides</i> (11/27) <i>Alistipes</i> (2/27)
Xil3	61	0,54	0,34	0,77	<i>Prevotella</i> (4/25) <i>Bacteroides</i> (8/25) <i>Alistipes</i> (10/25)
Csd23	49	0,76	0,6	0,61	<i>Bacteroides</i> (3/23)
M9F1	45	0,4	0,54	0,28	<i>Acinetobacter</i> (15/25) <i>Syntrophomonas</i> (2/25)
<i>Prevotella</i>	41				
<i>Bacteroides</i>	43				
<i>Alistipes</i>	55				

<sup>a</sup> Asignado por GC-Profile (231)

<sup>b</sup> Calculado en TETRA. Coeficiente de correlación de Pearson para los Z-score asignados entre la frecuencia de los tetranucleótidos de los fósmidos y las cepas correspondientes

<sup>c</sup> Obtenidos del análisis de homología de los ORFs codificados en cada secuencia, ver Anexo 3



**Figura 14.** Gráfico de dispersion de los Z-scores para la frecuencia de tetranucleótidos de *P. ruminicola* (x) y Csd4 (y). Se muestra la línea de regresión. El coeficiente de correlación de Pearson es de 0.76.

## Clonado, expresión y caracterización de hidrolasas

Las características fisicoquímicas inferidas *in silico* para las enzimas codificadas por los ORFs amplificados se detallan en la tabla 13.

**Tabla 13.** Características y situación actual de las enzimas codificadas en los fósmidos en estudio

Enzima	aa <sup>a</sup>	Peso molecular (g/mol) <sup>b</sup>	Coef. Extinción (M-1cm-1) <sup>b</sup>	Situación actual
EndoG	392	46044	106000	Caracterizada
Xyl4	505	56368.7	102595	Clonada en vector pET22a – en BL21
Lac4	263	29215.9	31775	Clonada en vector pET22a/ no hay expresión
Bacon6	292	31591.5	50670	Clonada en vector pET14b- en DH5α
Xil8	429	47621.7	78185	Clonada en vector pBK+
LysM8	277	30351.2	26025	
Endo9	411	46135.1	104530	Clonada en vector pET14b / Expresión en cuerpos de inclusión
Endo18	315	37199.3	91120	Clonada en vector pET14b
X3-43a	303	33637.9	63050	Clonada en vector pBK+
X3-43b	189	20917.7	31525	

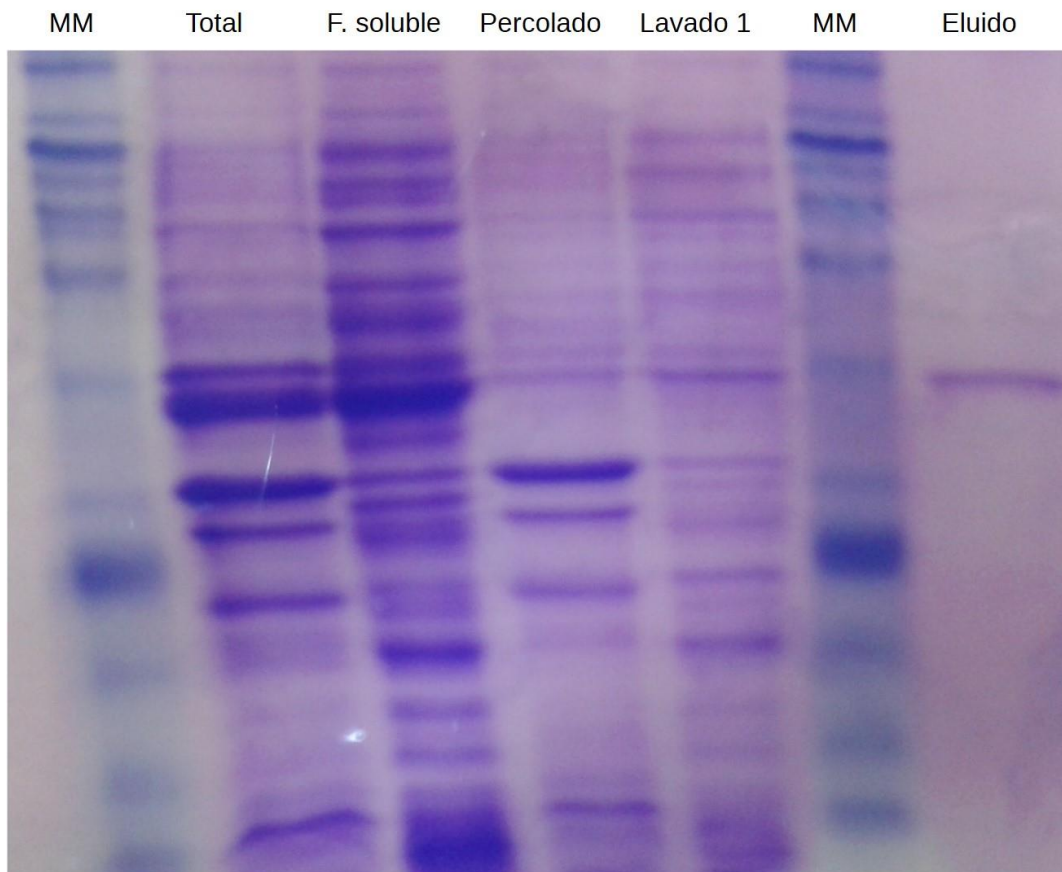
<sup>a</sup> Tamaño en aminoácidos

<sup>b</sup> Parámetros determinados mediante ProtParam (<http://web.expasy.org/protparam/>).

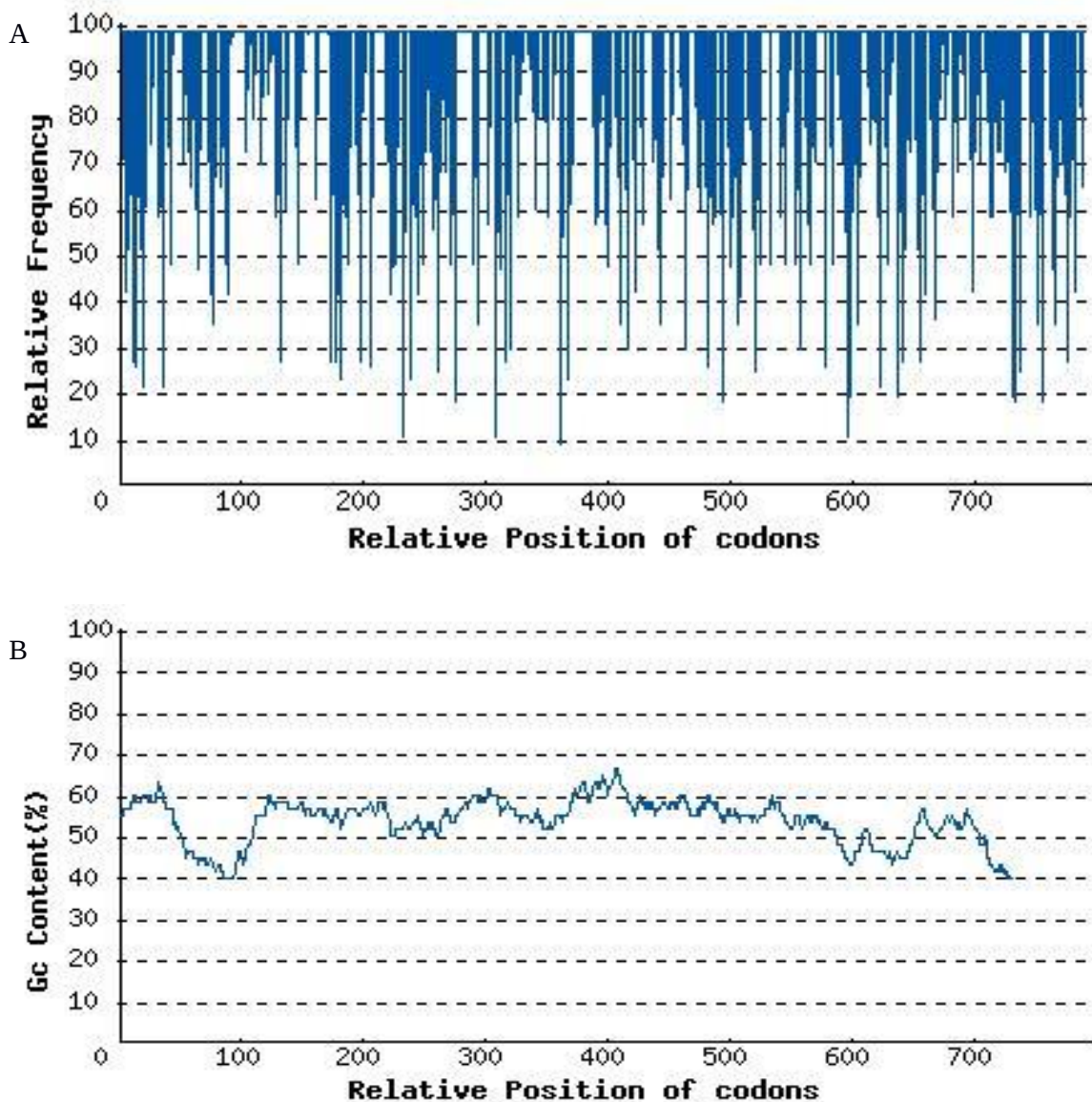
La enzima codificada por el ORF Csd4\_23, que codifica para una endoglucanasa con un dominio catalítico GH 5, EndoG, se amplificó por PCR y se clonó en el vector pET28a. Esta construcción fue transferida a la cepa BL21 (pLysS) DE3 y se determinó que las condiciones óptimas para su expresión son: incubación a 37°C, con agitación hasta alcanzar una DO de 0,3, luego agregado de IPTG 0,5 mM e incubación 4 horas más a igual temperatura. Luego de verificar por SDS-PAGE y ensayos de actividad que la enzima era activa, se procedió a su purificación mediante hibridación con resina de Ni-Agarosa. Para esto se probaron varios protocolos de purificación, determinando que los mejores rendimientos se obtenían mediante un protocolo “híbrido” en el cual se adiciona urea en el buffer desnaturizante, previo al sonicado del cultivo, la fracción soluble obtenida se incubó durante 4 horas con la resina previamente equilibrada, para luego proceder con las etapas de lavado y elución sin urea. En la figura 15 se visualiza, en un gel SDS-PAGE, la sobre-expresión y purificación de EndoG recombinante.

La caracterización bioquímica y funcional de EndoG se detalla en el Capítulo 4.

No fue posible la sobre-expresión de Lac4 en ninguna de las condiciones estudiadas (4, 25, 30 y 37°C; durante 2, 4, 12 y 16 horas). Tomando en cuenta que ni el clon Epi (Csd4) ni BL21 (pLysS) pET14b-Lac4 mostraron actividad lacasa positiva, se decidió hacer un análisis bioinformático del uso de codones para la secuencia de dicho gen. El resultado del mismo mostró que la región codificante para Lac4 posee una frecuencia en el uso de codones que difiere con la empleada por *E. coli* (Figura 16), por lo cual es posible plantear que no sea posible la expresión de este gen en *E. coli*.



**Figura 15.** SDS-PAGE (12% acrilamida, 0,8% bisacrilamida) de las etapas de purificación de EndoG. Las condiciones de incubación de la cepa BL21 (pLysS) DE3-pET28a-EndoG fueron: 37°C, 150 rpm, 4 horas con 0,5 mM de IPTG. MM, Marcador de peso molecular Thermo Scientific, Page Ruler Unstained Protein Ladder, #26614; FT, percolado,



**Figura 16.** A, Distribución de la frecuencia en el uso de codones a lo largo de la secuencia del ORF Csd4\_22 en relación al óptimo para ser expresado en *E. coli*. Idealmente, para una buena expresión en el organismo hospedero el índice de adaptación de codones (CAI, Codon Adaptation Index) debe ser mayor a 0,8, para este gen el valor promedio es 0,62. B, Contenido GC (%) dentro de la secuencia codificante para Lac4, el valor promedio es de 53.84. Obtenido mediante Rare Codon Analysis Tools, de GenScript ([http://www.genscript.com/cgi-bin/tools/rare\\_codon\\_analysis](http://www.genscript.com/cgi-bin/tools/rare_codon_analysis)).

## Conclusiones

La metodología empleada para la construcción de las bibliotecas metagenómicas permitió la obtención de 30.000 clones a partir de líquido ruminal y 100.000 a partir del contenido de un digestor anaerobio con un inserto de 42 Kpb, aproximadamente. El screening funcional de las dos metagenotecas reveló la presencia de 58 clones con actividad celulasa o xilanasa. En base a las características fenotípicas de los mismos se seleccionaron 9 clones, los cuales resultaron positivos para actividad exocelulasa, endocelulasa o xilanasa en diversos sustratos.

Sus fósmidos fueron secuenciados completamente, y los análisis de similitud de secuencia de nucleótidos revelaron que ninguno proviene de organismos cuyo genoma se encuentre actualmente disponible en las bases de datos. Mediante análisis de similitud de las proteínas codificadas y estudios de comparación en la frecuencia de uso de tetranucleótidos, se puede inferir que su origen estaría vinculado a bacterias del género *Bacteroidetes*, pero no se pudo atribuir su origen filogenético.

La anotación de los ORFs detectados permitió la identificación de al menos 10 glicosil hidrolasas novedosas en siete de los fósmidos secuenciados. Estas enzimas mostraron similitud con enzimas no caracterizadas previamente, provenientes de la comunidad microbiana del rumen. Además de conferir diferentes fenotipos a *E. coli* estas enzimas poseen diversos dominios catalíticos (GH) y de unión a carbohidratos (CBM), lo cual demuestra la diversidad de enzimas detectadas.

En los fósmidos M9F1 y Csd23 no se pudo identificar un gen o clúster de genes responsable de los fenotipos observados, por lo que es posible suponer que las enzimas codificadas tendrían muy baja homología con las familias de celulasas conocidas.

La caracterización funcional de las enzimas detectadas proveerá de información suficiente para determinar el potencial de las mismas de participar en procesos industriales de sacarificación de biomasa lignocelulósica y producción de etanol. Un ejemplo de esto se presenta en el capítulo siguiente.

**Capítulo 2.**  
**CARACTERIZACIÓN FUNCIONAL Y ANÁLISIS**  
**DEL GENOTIPO DEL CLON M9F1**

---

Loaces, I., Rodríguez, C., Amarelle, V., Fabiano, E. & Noya, F. (2013) **Búsqueda de nuevas enzimas para la conversión de biomasa en etanol**. Serie FPTA, N° 47 del Instituto Nacional de Investigaciones Agropecuarias (INIA).

## Resumen

Como se describió en la parte inicial de este trabajo el análisis funcional de las metagenotecas construídas reveló la presencia de varios clones con actividad de interés. Un fósmino en particular, denominado M9F1, confiere a *Escherichia coli* la habilidad de utilizar avicel, papel de filtro, bagazo de caña de azúcar o celobiosa como fuente de carbono. Mediante el análisis bioinformático de este fósmino no se pudo atribuir, por homología de secuencia, las cualidades fenotípicas detectadas a ninguno de los genes anotados. La incorporación de este fósmino en la cepa de *E. coli* etanologénica LY180 permitió la producción de etanol a partir de papel de filtro.

## Objetivos

- ☞ Caracterizar, fenotípicamente, al clon Epi300 (M9F1), seleccionado por su capacidad de crecer en medio mínimo con avicel como única fuente de carbono.
- ☞ Estudiar el genotipo del fósmino M9F1, determinando los genes responsables de las actividades detectadas y el origen filogenético del mismo.
- ☞ Evaluar el potencial del fósmino M9F1 en el proceso de producción de etanol a partir de celulosa.

## Principales resultados y conclusiones

Con el fin de obtener enzimas novedosas implicadas en la degradación de biomasa lignocelulósica, se contruyeron dos bibliotecas metagenómicas, a partir de ambientes especializados en su degradación. La evaluación funcional de los clones obtenidos reveló la presencia de 58 con actividades prometedoras frente a glucanos o xilanos. El estudio en profundidad del fósido denominado M9F1 reveló que es capaz de conferir a *E. coli* la capacidad de crecer en medio mínimo con papel de fitro, bagazo de caña de azúcar o celobiosa como única fuente de carbono. El análisis de la secuencia del fósido mostró que probablemente el organismo de origen de este fragmento de ADN esté emparentado con el género *Syntrophomonas*, aunque la cepa parental no estaría secuenciada actualmente. A su vez, mediante la anotación de los marcos de lectura abiertos (Open Reading Frame, ORFs) identificados no fue posible identificar qué genes están implicados en los fenotipos observados. Estos resultados confirman la hipótesis de que las aproximaciones de metagenómica funcional permiten acceder a elementos génicos novedosos, los cuales pueden contribuir al desarrollo de cepas aptas para el bioproceso consolidado.

## Contribución personal al trabajo

En este trabajo realicé la construcción de las librerías metagenómicas y el screening de las mismas con la colaboración de V. Amarelle. Llevé a cabo los ensayos fenotípicos de los clones seleccionados. Participé en los análisis filogenéticos y de asignación de funciones de los ORFs detectados en la secuencia de M9F1. Los ensayos de fermentación los realicé en las instalaciones Microbial Physiology and Biotechnology Lab, Department of Microbiology and Cell Science, University of Florida, Estados Unidos, bajo la tutoría del Dr. Lonnie Ingram.

El trabajo presentado en este capítulo fue financiado parcialmente por un proyecto FPTA del Instituto Nacional de Investigación Agropecuaria (INIA), por un proyecto FSE de la Agencia Nacional de Investigación e Innovación (ANII) y una beca IANAS-Fellowship Short Visit Program. Inter American Network of Academies of Science (IANAS) Fellowship. Sponsored by the U.S. Department of State and the National Research Council of the National Academies.



Inés Loaces\*  
Cecilia Rodríguez\*  
Vanesa Amarelle\*  
Elena Fabiano\*  
Francisco Noya\*

\*Departamento de Bioquímica y  
Genómica Microbianas

## Búsqueda de nuevas enzimas para la conversión de biomasa en etanol

Proyecto FPTA 256

Período de Ejecución: May. 2009-Jul. 2012

### RESUMEN

El uso de combustibles fósiles constituye actualmente una de las mayores fuentes de emisión de gases con efecto invernadero. La búsqueda de fuentes de energía renovables se presenta como uno de los desafíos cruciales con los que se enfrenta actualmente la Humanidad. Una nueva matriz energética basada en fuentes de energía renovables debe contemplar entre sus principales elementos a los biocombustibles líquidos. El Uruguay ya ha iniciado este camino y hoy se encuentra produciendo los dos principales biocombustibles líquidos: bioetanol y biodiesel. La diversificación de las materias primas utilizadas para obtener estos combustibles es indispensable para asegurar la independencia energética. La biomasa vegetal lignocelulósica es una alternativa abundante, ubicua y conveniente ya que se obtiene como un subproducto de distintas cadenas agroindustriales. En la actualidad, el paso limitante en la conversión de lignocelulosa en etanol es la sacarificación enzimática de la biomasa para obtener azúcares fermentables a partir de los cuales se produce el etanol. Este trabajo planteó desarrollar un microorganismo capaz de facilitar la producción de etanol lignocelulósico sin necesidad de enzimas exógenas. Para ello se utilizaron técnicas de metagenómica funcional utilizando como material de partida las comunidades microbianas aisladas de hábitats naturales como ser el rumen bovino o los lodos de reactores anaerobios de tratamiento de efluentes. El material genético de estas comunidades fue seleccionado por su capacidad para favorecer el crecimiento de *Escherichia coli* en material lignocelulósico como única fuente de carbono. Uno de los fragmentos obtenidos, denominado M9F1, fue capaz de permitir el crecimiento de *E.coli* tanto en papel de filtro como en bagazo de caña de azúcar. Además, cuando se lo introdujo en una cepa etanológica, se pudo constatar la fermentación de etanol a partir de papel. La secuencia de M9F1 reveló genes relacionados con el metabolismo de los polisacáridos pero ninguno de ellos presenta homologías significativas con celulasas,  $\beta$ -glucosidasas, xilanasas u otras glicosilhidrolasas conocidas

### INTRODUCCIÓN

La búsqueda de nuevas fuentes de energía renovables se presenta como uno de los desafíos cruciales con los que se enfrenta actualmente la Humanidad. La alarmante acumulación atmosférica de gases de efecto invernadero producida principalmente por la quema de combustibles fósiles amenaza con producir un cambio climático global capaz de desencadenar un nuevo suceso de extinción masiva. En este contexto, las energías renovables son una alternativa viable, confiable y factible tanto para los países desarrollados como para aquellos en vías de desarrollo. La Agencia Internacional de la Energía (IEA) estima que para lograr mantener el calentamiento global en no más de 2 °C por encima de

las temperaturas medias previas a la Revolución Industrial, será necesario aumentar cuatro veces la producción de biocombustibles avanzados para el año 2020.

Uno de los principales biocombustibles es el bioetanol. Su producción consiste esencialmente en la degradación de polisacáridos complejos en azúcares simples y en su posterior fermentación alcohólica. Por lo tanto, las materias primas de este proceso pueden ser sumamente diversas. En la actualidad, las más utilizadas industrialmente son el almidón del maíz y la sacarosa de la caña de azúcar. Sin embargo, es la lignocelulosa el polisacárido de glucosa más abundante, más barato y mejor distribuido del mundo. La

lignocelulosa es el principal componente de las paredes vegetales, en general, y de la madera, en particular. Es el polímero orgánico más abundante y el principal constituyente de la biomasa terrestre. Está compuesta esencialmente por celulosa, hemicelulosa y lignina en una estructura supramolecular altamente resistente a la degradación y oxidación. Debido a esta naturaleza recalcitrante solamente algunos microorganismos con sistemas enzimáticos especializados pueden digerirla. Lignocelulosa apta para ser utilizada en la producción de bioetanol puede obtenerse de las partes no comestibles de cultivos tradicionales como el maíz, el trigo y el arroz, desechos forestales como aserrín y cortezas, desechos domésticos e industriales como papel y cartón, o de cultivos destinados a su producción como eucaliptus, álamos y mijo. Cualquiera sea su origen, esta materia prima no constituye un alimento para el hombre lo que representa una ventaja sobre el almidón del maíz o la sacarosa de la caña de azúcar. El Uruguay cuenta con una buena capacidad de producción de lignocelulosa ya sea que ésta provenga de productos secundarios de la actividad agropecuaria (cáscara de arroz, rastrojo de trigo o maíz, bagazo de caña de azúcar, corteza de árboles) o constituya el producto principal de cultivos especializados. Mediante el tratamiento hidrolítico adecuado, la lignocelulosa puede ser utilizada en la producción de bioetanol. El principal obstáculo para producir bioetanol a partir de lignocelulosa es el alto costo que actualmente tienen las enzimas lignocelulolíticas comerciales. Además, a escala industrial, estas enzimas son producidas y comercializadas por un pequeño número de proveedores, todos ellos extranjeros, lo que crearía una nueva dependencia de la matriz energética del país. En este estudio nos hemos propuesto utilizar la metagenómica funcional como herramienta para aislar nuevas funciones metabólicas que puedan ser aplicadas a mejorar la producción de bioetanol lignocelulósico. Esencialmente, la metagenómica funcional consiste en la búsqueda de actividades de interés en el conjunto de los genomas integrantes de una comunidad microbiana natural como puede ser, por ejemplo, la que habita una muestra de suelo fértil. El primer paso para lograr este objetivo consiste en aislar el ADN total de la comunidad, fragmentarlo y clonarlo en vectores apropiados para obtener una colección de clones de *Escherichia coli* cada uno de los cuales ha incorporado un fragmento del ADN de la comunidad. Este conjunto de clones se llama biblioteca metagenómica o «metagenoteca». En esta metagenoteca estarán representados fragmentos de material genético de especies bacterianas conocidas pero, sobre todo, de especies desconocidas que no pueden ser cultivadas en el laboratorio. Se estima que sólo el 1% de la diversidad bacteriana del planeta es cultivable en el laboratorio. Esta metagenoteca se crece

en medios de selección apropiados para detectar las actividades biológicas de interés. Por ejemplo, para buscar celulasas se crece la biblioteca en un medio conteniendo celulosa cristalina (avicel) como única fuente de carbono y se seleccionan aquellos clones capaces de crecer y degradarla.

Utilizando este enfoque desarrollamos metagenotecas a partir de comunidades microbianas especializadas en la degradación de biomasa vegetal. Estas bibliotecas fueron principalmente utilizadas para buscar actividades lignocelulolíticas. Los clones identificados fueron evaluados en cepas de *E.coli* llamadas «etanológicas» porque son capaces de realizar la fermentación alcohólica de los azúcares liberados. Por último, para identificar los genes responsables de estas actividades, los fragmentos candidatos fueron secuenciados parcial o totalmente y analizados con herramientas bioinformáticas.

## MATERIALES Y MÉTODOS

### Obtención de muestras

La muestra de contenido ruminal fue obtenida de un novillo Holando de 480 kg alimentado en praderas y sacrificado en el Frigorífico Pando (Canelones, Uruguay). La muestra de lodos anaerobios fue obtenida de un digestor alimentado con contenido ruminal [sólidos totales (ST) 19,3% (w/w) y sólidos volátiles (SV) 17,3% (w/w)], grasa de trinchado de curtiembre (ST 25,1% y SV 17,6%) y lodos de purga de lagunas anaerobias de una aceitera (ST 25,6% y SV 21,7%) el cual opera en el Departamento de Reactores de Facultad de Ingeniería (Universidad de la República, Montevideo, Uruguay). Todas las muestras fueron mantenidas a 4 °C hasta su procesamiento.

### Obtención de ADN metagenómico

Para separar los microorganismos presentes en las muestras 500 µl de líquido ruminal o del lodo del digestor se diluyeron en 3 mL de PBS. Se agregaron 3 mL de Percoll (Sigma-Aldrich, St. Louis, EE.UU.) y se centrifugaron durante 20 minutos a 10.000 x g. Se separó la capa opalescente y se realizó la tinción de

Gram para verificar la presencia de células bacterianas. La fracción celular se resuspendió en 1 mL de búfer de lisis (NaCl 700 mM, Tris-HCl 50 mM pH = 8,0, EDTA 100 mM, SDS 4%). Se agregaron bolitas de zirconia (aproximadamente 100 µl) y se agitó horizontalmente durante 30 segundos. Se incubó a 70 °C durante 15 minutos mezclando cada 5 minutos y se centrifugó a 4 °C durante 5 minutos a 16000 x g. El sobrenadante fue tratado con 15 µl de proteinasa K (20 mg/ mL) durante 1 hora a 37 °C. Posteriormente se agregaron 160 µl de CTAB 10% en NaCl 0,7 M y se incubó 10 minutos a 65 °C. El ADN se extrajo con fenol y cloroformo, se precipitó con isopropanol y se resuspendió en 200 µl de agua ultrapura.

### **Construcción de las metagenotecas y selección de actividades enzimáticas de interés**

El ADN fue fraccionado por tamaño en geles de agarosa. Se seleccionaron aquellos fragmentos entre 35 y 45 kpb. Los fragmentos fueron clonados en el vector fosmídico pCC1FOS utilizando el kit CopyControl Fosmid Library Production (Epicentre Biotechnologies, Madison, EE.UU.) y siguiendo las instrucciones del fabricante. Se infectaron células de EPI300 y las transfectantes resistentes al cloranfenicol (12,5 µg/mL) se ordenaron en grupos de 10 o más clones en placas de 96 pocillos. Estos clones fueron evaluados en medio LB con el agregado de 0.2% carboximetil celulosa sal disódica (CMC) (Sigma-Aldrich) para detectar actividad celulasa y en LB con 0,5% de xilano de avena (Sigma-Aldrich) para detectar actividad xilanasas. Se incubaron las placas a 30 °C durante 7 días y se revelaron con Rojo Congo durante 20 minutos. La presencia de un halo amarillo alrededor de la colonia se tomó como evidencia de la presencia de actividad glicolítica. A su vez se evaluó la capacidad de los clones de crecer en medio mínimo M9 (MgSO<sub>4</sub> 2 mM, CaCl<sub>2</sub>

0,1 mM, Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O 12,8 g/l, KH<sub>2</sub>PO<sub>4</sub> 3 g/l, NaCl 0,5g/l, NH<sub>4</sub>Cl 1g/l, timina 50 µg/mL, leucina 200 µM) con 0,5% de avicel como única fuente de carbono.

### **Pretratamiento de bagazo de caña**

Bagazo de caña molido con un contenido de humedad del 54% fue obtenido de la empresa productora de alcoholes ALUR. Este sustrato fue sometido a dos tipos de pretratamiento: incubación con agitación durante 2 horas con ácido fosfórico 1% o con ácido sulfúrico 1%, seguido de 1 hora de autoclave a 121 °C. Luego de este

tratamiento, la fracción líquida fue neutralizada con NaOH hasta lograr pH= 6,5 y la fracción sólida fue lavada 3 veces con abundante agua destilada. La biomasa obtenida se mantuvo a 4 °C hasta su incorporación en placas de M9 en una concentración final de 1%.

### **Secuenciación y análisis bioinformático**

La secuenciación de los fósmidos seleccionados se realizó mediante la técnica de shotgun por MacroGen Inc. (Seúl, República de Corea). El análisis bioinformático de la secuencia obtenida se realizó utilizando la plataforma CAMERA 2.0.

### **Producción y detección de etanol**

Fue evaluada la capacidad de utilizar papel de filtro y producir etanol a partir del mismo en condiciones aerobias (matraces) y anaerobias (Fleaker, Corning, EE.UU.). Se sembraron cultivos semilla en M9-glucosa 1% con los antibióticos correspondientes. Estos fueron utilizados para inocular medio M9 con papel de filtro 1% (w/v). La producción de etanol fue seguida mediante cromatografía gaseosa durante 48 horas (Agilent GC 6890N, inyector 7683B; columna megabore HP-PLOT de 15 m x 0,53 mm y 40 µm film).

### **Determinación de glucosa y celobiosa**

La concentración de glucosa y celobiosa en el medio de cultivo fue determinada por HPLC utilizando un HPLC Waters (Milford, EE.UU.) detectados por índice de refracción. La separación se realizó en una columna Bio-Rad (Hercules, EEUU) Aminex HPPX87P (300 mm x 7,8 mm d.i.) a 85 °C usando agua ultrapura como fase móvil (0,5 mL/min).

## RESULTADOS

### Construcción de metagenotecas

Para lograr los objetivos propuestos se comenzó con la construcción de bibliotecas metagenómicas a partir de muestras de comunidades microbianas obtenidas de distintos hábitats en donde tiene lugar la degradación de biomasa vegetal. Concretamente, se seleccionaron muestras de rumen bovino y de lodo de digestores anaerobios alimentados con contenido ruminal. En ambos casos se logró separar la fracción bacteriana y purificar su ADN. Este ADN fue fragmentado y clonado en fósidos que fueron transfectedados en células de *E.coli* EPI300.

Se obtuvieron 30.000 clones para la biblioteca construida a partir de contenido ruminal bovino y 100.000 para la biblioteca obtenida a partir de lodos de digestor anaerobio. Cada uno de estos clones alberga un fragmento de ADN metagenómico de 42 kpb en promedio.

### Selección de las actividades de interés

Las metagenotecas fueron sembradas en medios conteniendo CMC o xilano de avena, con el fin de seleccionar actividades enzimáticas para celulasas o xilanasas.

También se las colocó en medio mínimo con avicel como única fuente de carbono. El avicel es una forma de celulosa microcristalina. Se presume que los clones capaces de crecer en avicel tendrán, además de hidrolasas, los sistemas de transporte necesarios para secretar a estas enzimas y para internalizar a los glúcidos liberados. Los clones positivos primariamente fueron confirmados y sus fósidos, reaislados, fueron utilizados para transformar células frescas. Cuando el

fenotipo observado se repetía en estas células, se tomó como evidencia que las funciones de interés efectivamente provienen del fragmento metagenómico introducido y no de causas espúreas. Para descartar los posibles clones duplicados que hayan surgido de seleccionar dos o más veces el mismo fósido, se realizaron perfiles de restricción de los fósidos con la enzima EcoRV. En caso de constatarse idénticos perfiles de restricción en dos o más clones, se eligió un clon representante de la clase para continuar el estudio descartándose los duplicados.

De 58 clones positivos que presentaron una o más actividades de interés, se seleccionaron 13 para continuar el estudio en virtud de la reproducibilidad de sus fenotipos en medios de selección (Cuadro 1).

### Capacidad de crecimiento en papel de filtro como única fuente de carbono

El papel de filtro está compuesto esencialmente por celulosa y su conversión en azúcares reductores es un ensayo comúnmente utilizado para determinar las actividades celulolíticas de enzimas purificadas. Se razonó que si los clones seleccionados producen enzimas celulolíticas, era posible que algunos de ellos fueran capaces de crecer en papel de filtro como única fuente de carbono.

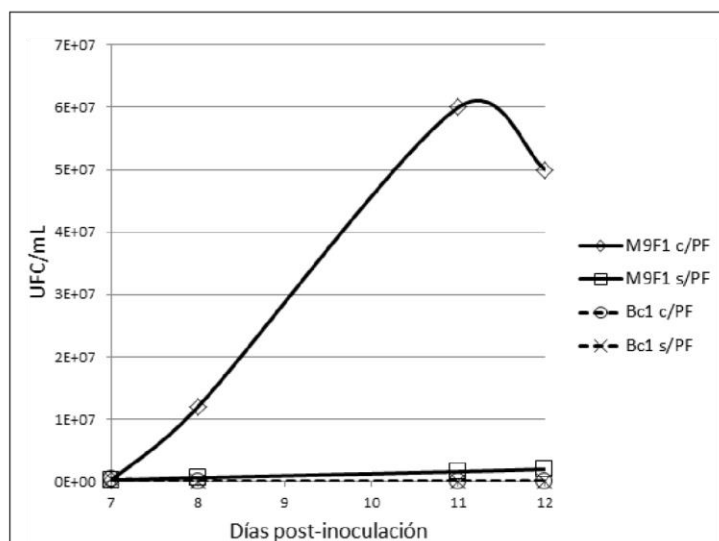
Para evaluar esta alternativa se utilizó medio mínimo M9 al que se le agregó una tira de papel de filtro (Whatman N°2) como única fuente de carbono. Como controles negativos, se utilizaron el mismo medio pero sin el papel de filtro y la cepa parental sin los fósidos seleccionados. La Figura 1 muestra un experimento tipo donde se aprecia crecimiento a partir de papel de filtro en presencia del fósido M9F1. La Figura 2 muestra los recuentos obtenidos en estas condiciones.

**Cuadro 1.** Colección inicial de clones con actividades de interés confirmadas seleccionados a partir de las metagenotecas crecidas en distintos medios

Clon	Origen	Características en medios de selección
Csd1	Rumen	Halo en CMC
Csd5	Rumen	Halo en CMC
Csd9	Rumen	Halo en CMC
Csd14	Rumen	Halo en CMC
Csd23	Rumen	Halo en CMC
Xil3	Rumen	Halo en xilano de avena
Xil4	Rumen	Halo en xilano de avena
Xil5	Rumen	Halo en xilano de avena
Lq5B10	Rumen	Crece en medio mínimo con avicel
BH8	Fermentador	Halo en CMC
EF9	Fermentador	Halo en CMC y xilano
M9F1	Fermentador	Crece en medio mínimo con avicel
M9F2	Fermentador	Crece en medio mínimo con avicel



**Figura 1.** Medio mínimo M9 conteniendo papel de filtro como única fuente de carbono a los 7 días postinoculación. El tubo de la izquierda fue inoculado con la cepa EPI300 conteniendo un fósido elegido al azar, como control negativo. El tubo de la derecha fue inoculado con la cepa EPI300 conteniendo el fósido M9F1.



**Figura 2.** Recuento de unidades formadoras de colonias (UFC), en presencia (c/PF) y ausencia (s/PF) de papel de filtro para EPI300 conteniendo M9F1 o un fósido control (Bc1).



**Figura 3.** Cultivos de 19 días en medio mínimo M9 conteniendo 1% (w/v) de papel de filtro. De izquierda a derecha: LY180, LY180 (M9F1) y EPI300 (M9F1).

Estos datos sugieren que el crecimiento de EPI300 (M9F1) en papel de filtro es lento y requiere de una fase lag relativamente extensa. Recién a partir de los 7 días post-inoculación comienza a observarse un aumento de la densidad celular. Por otra parte, si estos cultivos estaban creciendo a partir de la celulosa del papel, la integridad estructural de éste debería verse afectada con el tiempo. Para intentar observar esta transformación, se aumentó el volumen de los cultivos y se extendió el tiempo de incubación hasta 19 días. Luego de este período el papel de filtro incubado con EPI300 (M9F1) perdió casi totalmente su estructura macroscópica. Un efecto similar pero de menor intensidad se observó con la cepa etanologénica LY180 (M9F1) y no se observaron cambios en el papel en presencia de LY180 sin fósido alguno (Figura 3).

### Producción de etanol utilizando papel de filtro

Para determinar la capacidad de M9F1 para participar en la conversión de celulosa en etanol se utilizó la cepa etanologénica LY180. Esta cepa ha sido genéticamente modificada buscando maximizar la fermentación de hexosas y pentosas en etanol.

El fósido M9F1 fue transformado en LY180 y la cepa resultante fue crecida en condiciones anaerobias en medio mínimo M9 conteniendo papel de filtro como única fuente de carbono. La producción de etanol a las 48 horas post-inoculación se determinó mediante cromatografía de gases. Paralelamente, se utilizó el plásmido pLOI297 que porta los genes de piruvato decarboxilasa (*pdC*) y alcohol deshidrogenasa (*adhB*) de *Zymomonas mobilis* y que potencia la producción de etanol en cepas de *E.coli*. Este plásmido se utilizó para

transformar a la cepa EPI300 conteniendo el fósido M9F1 (Figura 4).

A continuación se estudió la cinética de producción de etanol conjuntamente con la liberación de glucosa y celobiosa, los productos principales de la degradación enzimática de la celulosa. Para ello los sobrenadantes de cultivos anaerobios conteniendo papel de filtro como única fuente de carbono fueron analizados por GC y HPLC. Como se observa en la Figura 5, se libera tanto glucosa como celobiosa al medio. Tal como se espera, la glucosa es rápidamente consumida lo que coincide con la aparición del etanol. La celobiosa es mantenida en niveles relativamente constantes a lo largo de todo el experimento.

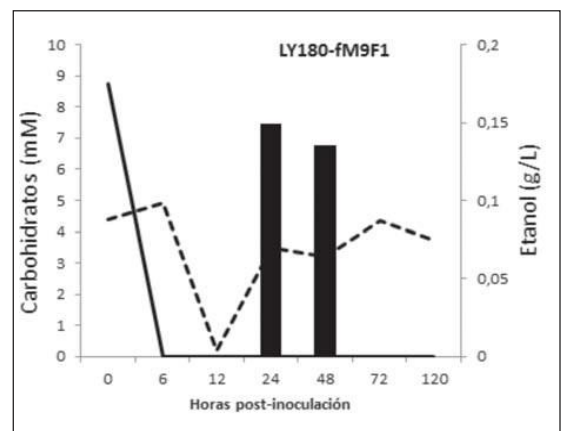
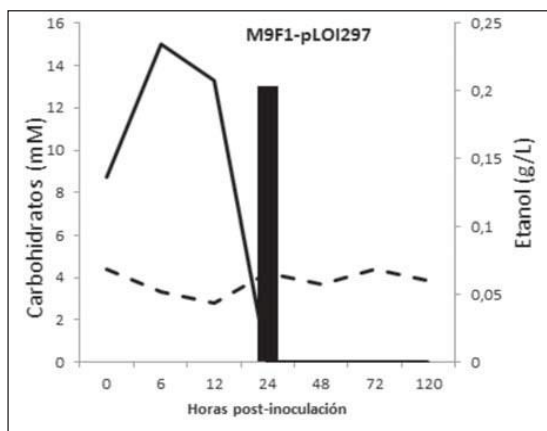
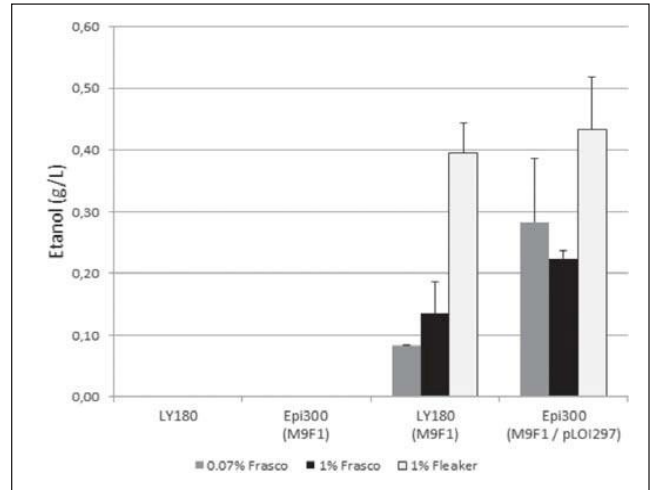
Para comprobar si la presencia de M9F1 favorece el crecimiento en celobiosa, se realizaron curvas de crecimiento en medio mínimo M9 conteniendo celobiosa como única fuente de carbono. En la Figura 6 se puede observar que M9F1 favorece el crecimiento en celobiosa con respecto a la cepa control sin M9F1.

### Utilización del bagacillo de caña de azúcar

Estos experimentos demuestran que M9F1 es necesario y suficiente para viabilizar el crecimiento de las cepas de *E.coli* estudiadas en papel de filtro como única fuente de carbono. Sin embargo, el papel de filtro no contiene lignina por lo que no constituye un sustrato lignocelulósico. Además, no representa una materia prima relevante desde el punto de vista del proceso industrial de una bio-refinería. Por estas razones, se realizaron también ensayos en un sustrato de mayor valor agroindustrial como es el bagazo de caña de azúcar.

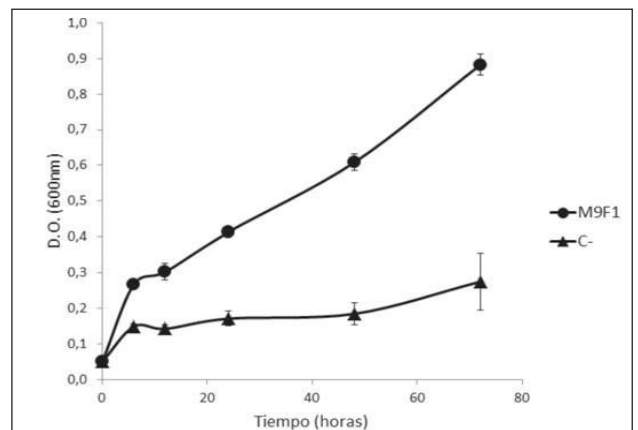
Bagazo de caña molido («bagacillo») conteniendo 54% de humedad fue provisto por la empresa productora de alcoholes ALUR (Artigas, Uruguay). Este sustrato fue sometido a pretratamiento con ácido fosfórico 1% o con ácido sulfúrico 1%,

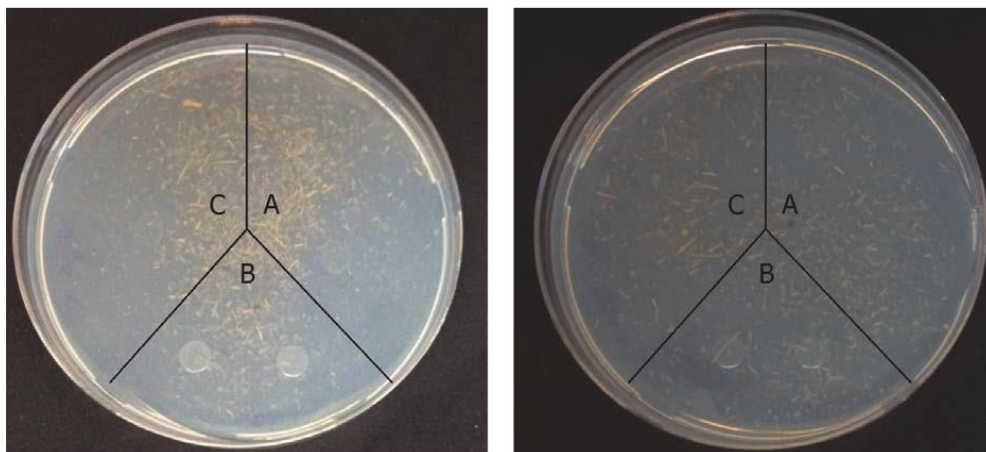
**Figura 4.** Producción de etanol a partir de papel de filtro (0,07% y 1%) a las 48h post-inoculación en dos condiciones anaerobias: frascos cerrados y Fleakers. Se muestran los promedios con la desviación estándar de tres determinaciones para la cepa LY180, la cepa EPI300 conteniendo el fósido M9F1, la cepa LY180 conteniendo el fósido M9F1 y la cepa Epi300 conteniendo el fósido M9F1 y el plásmido etanologénico pLOI297.



**Figura 5.** Concentración de etanol (barras), glucosa (línea continua) y celobiosa (línea punteada) a lo largo del tiempo para EPI300 conteniendo el fósido M9F1 y el plásmido etanologénico pLOI297 (M9F1pLOI297, panel izquierdo) y para LY180 conteniendo el fósido M9F1 (LY180-fM9F1, panel derecho). Ambos experimentos fueron llevados a cabo en anaerobiosis, en medio mínimo con papel de filtro como única fuente de carbono.

**Figura 6.** Curva de crecimiento de EPI300 transformada con M9F1 o con un fósido control (C-) en medio mínimo M9 conteniendo celobiosa como única fuente de carbono. Las barras de error corresponden a la desviación estándar de tres replicados.





**Figura 7.** Placas conteniendo 1% de bagacillo de caña de azúcar pretratado con 1% de ácido sulfúrico (izquierda) o 1% de ácido fosfórico (derecha). Cada placa fue sembrada con dos gotas de 40µL de cada uno de los siguientes cultivos: (A) EPI300, (B) Epi300 (M9F1) y (C) LY180.

seguido de 1 hora de autoclave a 121 °C. La fracción sólida lavada con abundante agua destilada se incorporó en placas de medio mínimo M9 y agar a una concentración del 1%.

Como se muestra en la Figura 7, M9F1 es suficiente para permitir el crecimiento de EPI300 en bagacillo de caña de azúcar pretratado como única fuente de carbono.

### **Secuenciación y análisis bioinformático del clon M9F1**

Con el propósito de obtener información acerca de los genes involucrados en estos fenotipos así como también determinar el microorganismo del que provienen, se secuenció el fósido M9F1 en su totalidad mediante la técnica de «*shotgun*».

Se realizaron 962 reacciones de secuenciación por el método de Sanger. Luego de eliminar las secuencias provenientes de los vectores utilizados para la secuenciación así como las lecturas de baja calidad, quedaron 862 lecturas que fueron ensambladas en tres contigs con una cobertura promedio de 8.62X. Uno de los contigs estaba totalmente contenido en el pCC1FOS. Para completar la secuencia del inserto de M9F1 fue necesario secuenciar 435 bases adicionales a partir de los extremos de los otros dos contigs para lograr ensamblarlos y obtener una única secuencia de 33996 nucleótidos.

Mediante el algoritmo Metagene se identificaron un total de 28 marcos de lectura abiertos (ORFs). Las posibles funciones de estos ORFs fueron predichas basándose en su homología con clústers de grupos de ortólogos (COGs) y con las familias de proteínas Pfam (Cuadro 2).

A partir de estas predicciones se puede identificar en M9F1 un posible operón integrado por genes relacionados con el metabolismo de polisacáridos en el que se identifican varias glicosiltransferasas. Esas enzimas normalmente participan de la síntesis de polisacáridos de la pared celular aunque es posible que este operón participe también en el transporte de monosacáridos hacia el interior celular. De hecho, se identifican genes que podrían codificar para proteínas de membrana y permeasas.

Otro de los ORFs que llaman la atención es el #28 que codifica para una proteína no caracterizada que presenta un dominio Ycel. Estos dominios son característicos tanto en proteínas del periplasma como en proteínas secretadas. También se los encuentra asociados a módulos de unión a carbohidratos (CBMs) en celulasas. Si este gen efectivamente codificara una celulasa funcional, la misma tendría muy baja homología con las familias de celulasas conocidas. Su caracterización minuciosa será objeto de futuros estudios.



**Cuadro 2.** Asignaciones de los ORFs detectados en M9F1 en función de sus homologías con clústers de grupos de ortólogos (COGs) y con familias de proteínas Pfam. Los ORFs que no presentan homologías con ninguna secuencia conocida se presentan únicamente con su número y tamaño en aminoácidos

	Tamaño (AA)	COG	e-value COG	Descripción COG	Pfam	e-value Pfam	Descripción Pfam
1	49						
2	80						
3	224				PF09560	1.20E-44	Spore_YunB
4	459	COG0008	1.00E-109	Glutamyl- and glutaminyl-tRNA synthetases	PF00749	6.60E-132	tRNA-synt_1c
5	231	COG1215	5.00E-06	Glycosyltransferases, probably involved in cell wall biogenesis			
6	698	COG1216	9.00E-05	Predicted glycosyltransferases	PF00535	6.30E-08	Glycos_transf_2
7	361	COG0381	3.00E-85	UDP-N-acetylglucosamine 2epimerase	PF02350	1.30E-108	Epimerase_2
8	354	COG1086	6.00E-84	Predicted nucleosidediphosphate sugar epimerases	PF02719	6.30E-113	Polysacc_synt_2
9	323	COG1091	5.00E-41	dTDP-4dehydrorhamnose reductase	PF04321	1.00E-12	RmLD_sub_bind
0	92						
1	265						
2	151						
3	661	COG0463	2.00E-12	Glycosyltransferases involved in cell wall biogenesis	PF00535	1.20E-29	Glycos_transf_2
4	107	COG0564	5.00E-13	Pseudouridylate synthases, 23S RNAspecific	PF00849	1.30E-07	PseudoU_synth_2
5	111	COG3077	3.00E-08	DNA-damage-inducible protein J	PF04221	3.60E-11	RelB
6	147	COG0456	2.00E-08	Acetyltransferases	PF00583	9.50E-21	Acetyltransf_1
7	236	COG1187	2.00E-54	16S rRNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases	PF01479	4.90E-06	S4

Continuación

<b>8</b>	<b>150</b>	<b>COG2707</b>	<b>7.00E-33</b>	<b>Predicted membrane protein</b>	<b>PF04284</b>	<b>8.90E-73</b>	<b>DUF441</b>
<b>9</b>	393	COG2807	3.00E-15	Cyanate permease	PF07690	5.60E-33	MFS_1
<b>10</b>	306	COG0583	1.00E-33	Transcriptional regulator	PF03466	1.70E-34	LysR_substrate
<b>11</b>	571	COG0507	2.00E-18	ATP-dependent exoDNase (exonuclease V), alpha subunit - helicase superfamily I member			
<b>12</b>	672						
<b>13</b>	390	COG1454	1.00E-124	Alcohol dehydrogenase, class IV	PF00465	9.70E-199	Fe-ADH
<b>14</b>	343	COG2267	2.00E-26	Lysophospholipase	PF00561	8.90E-06	Abhydrolase_1
<b>15</b>	393	COG3284	3.00E-16	Transcriptional activator of acetoin/glycerol metabolism	PF01590	3.00E-05	GAF
<b>16</b>	503	COG1012	1.00E-145	NAD-dependent aldehyde dehydrogenases	PF00171	7.40E-237	Aldedh
<b>17</b>	236	COG0518	5.00E-37	GMP synthase - Glutamine amidotransferase domain	PF00117	1.00E-12	GATase
<b>18</b>	194	COG2353	2.00E-42	Uncharacterized conserved protein	PF04264	6.90E-60	Ycel

Entre las posiciones 2543 y 3557 del inserto se detectó un posible gen que codifica para un ARNr 16S. Este gen permite trazar un árbol filogenético para M9F1 utilizando la base de datos Green Genes. Este análisis sugiere que el origen de M9F1 es una bacteria del filo Firmicutes. Posiblemente se trata de un Clostridiales relacionado con *Syntrophomonas*. Sin embargo, a excepción de esta región, el resto de M9F1 no presenta homología con ninguno de los genomas secuenciados de *Syntrophomonas*. Por lo tanto, es posible que el microorganismo no pertenezca a una especie caracterizada.

## CONCLUSIONES Y RECOMENDACIONES

Este Proyecto planteó como objetivo el desarrollo de un microorganismo capaz de producir etanol a

partir de biomasa vegetal. Además de ser capaz de degradar lignocelulosa y de fermentar etanol, el organismo ideal debe tener bajos requerimientos nutricionales y adaptarse bien tanto al laboratorio como al entorno industrial.

Una estrategia consiste en recurrir a genes y operones conocidos especializados en alguno de estos aspectos e introducirlos en un organismo que se destaque en las demás capacidades. Por ejemplo, mediante la introducción de genes de celulasas ( $\beta$ -glucosidasas, endo- y exo-glucanasas) de *Aspergillus aculeatus* y *Trichoderma reesei* en una cepa de *Saccharomyces cerevisiae* especializada en la producción de vinos se logró que esta última creciera en celulosa y produjera hasta 2,6% de etanol. La desventaja con este enfoque es que toma como punto de partida genes ya descritos, con sus virtudes y limitaciones. La mayor parte de estos genes provienen de organismos «cultivables» o que pueden ser crecidos en condiciones de laboratorio. Sin embargo, más del

99% de los microorganismos, que a su vez constituyen el mayor reservorio genético del planeta, no pueden ser cultivados en el laboratorio.

En este trabajo se prefirió una estrategia alternativa basada en la metagenómica funcional para la búsqueda de nuevas enzimas lignocelulolíticas. Una de las fortalezas de este enfoque consiste en identificar genes novedosos que guardan poca o ninguna similitud con otros ya conocidos (18). Se eligieron hábitats en los cuales es esperable encontrar comunidades microbianas especializadas en la degradación de lignocelulosa.

La estrategia de selección permitió identificar al menos 13 clones que demostraron algún fenotipo relevante frente a glucanos y xilanos. De aquí en adelante el estudio se centró en uno de ellos, el fósido M9F1, que también tiene la capacidad de sostener el crecimiento de la cepa parental de *E.coli* en un medio mínimo conteniendo papel de filtro como única fuente de carbono. La presencia de M9F1 parece promover la liberación de glucosa y celobiosa al medio de cultivo que son los productos esperados de la acción de las exo- y endoglucanasas, así como de las  $\beta$ -glucosidasas. De hecho, también se confirmó que M9F1 tiene la capacidad de crecer mejor en celobiosa que la cepa parental.

Para identificar a los genes responsables de estos fenotipos se procedió a secuenciar completamente el inserto metagenómico del fósido M9F1. Tal como se espera del ADN aislado de comunidades microbianas naturales, la secuencia de M9F1 guarda muy poca homología con cualquier secuencia conocida. Si bien el organismo de origen parece estar emparentado con el género *Syntrophomonas*, ninguno de los genes identificados presenta homología con genes de las especies de este género. En función de su similitud con *Syntrophomonas*, es posible que el fragmento M9F1 provenga de un microorganismo anaerobio, Gram-negativo, posiblemente con capacidad de esporular (ver también el ORF #3 del Cuadro 2) y con altas demandas nutricionales. Es posible que establezca estrechas relaciones simbióticas con otras especies presentes en su ambiente lo que podría hacer muy dificultoso su aislamiento o crecimiento en medios estériles.

Con respecto a los genes identificados en M9F1 ninguno de ellos muestra homología con ninguna

de las familias de glicosilhidrolasas descritas en la base de datos CAZy lo que podría subrayar el carácter novedoso de los genes contenidos en este fragmento. Llama la atención la presencia de un clúster de glicosiltransferasas entre los ORFs 5 y 13 (Cuadro 2) que podría jugar algún papel en la degradación de las paredes celulares. Sin embargo, demostrar fehacientemente si son éstos los genes responsables de los fenotipos observados dependerá de estudios que están actualmente en curso.

La introducción del fósido M9F1 en *E.coli* confirió a esta bacteria la capacidad de utilizar celulosa como única fuente de carbono. Sin embargo, para producir etanol en estas condiciones hicieron falta elementos genéticos adicionales. *E.coli* es capaz de fermentar etanol pero también puede producir ácido láctico reduciendo el rendimiento del cultivo. Además, *E.coli* tiene un fuerte sistema de represión por catabolito mediante el cual la presencia de glucosa bloquea tanto el transporte como la utilización de otros azúcares. La cepa LY180 contiene mutaciones en las vías de producción de lactato y en la represión por catabolito. Si bien esta cepa es incapaz de producir etanol en condiciones anaerobias utilizando papel de filtro como única fuente de carbono, sí puede hacerlo cuando se le introduce M9F1. Un resultado similar se observó cuando a la cepa parental EPI300 con el fósido M9F1 se le agregó también el plásmido pLOI297 que porta genes de *Z.mobilis* capaces de derivar el flujo de carbono desde el lactato hacia el etanol.

Desde un punto de vista práctico, M9F1 confiere a *E.coli* la capacidad de crecer en una materia prima abundante y de bajo costo como lo es el bagazo de caña de azúcar. El bagazo y la vinaza son los principales productos secundarios de la producción de bioetanol a partir de caña de azúcar. El bagazo es el residuo fibroso que constituye el 12-14% del peso de la caña y es normalmente quemado para la producción de vapor y energía eléctrica. Sin embargo, la conversión de bagazo en etanol permitiría aumentar su valor entre 2 y 6 veces mejorando la ecuación económica de la biorefinería. El crecimiento de *E.coli* en bagazo como única fuente de carbono es el primer paso para alcanzar este objetivo. Aún queda por determinar cuánto etanol es posible producir a partir de esta materia prima cuando se introduce M9F1 en cepas etanológicas como LY180.

Una estrategia alternativa en la que el potencial de M9F1 debe ser evaluado es formando parte de «consorcios simbióticos». En esta estrategia dos o más microorganismos cooperan para degradar y fermentar la materia prima. En estos consorcios existen microorganismos especializados en la sacarificación de la materia prima lignocelulósica y también microorganismos especializados en la fermentación de los azúcares liberados. La capacidad de M9F1 para conferir el crecimiento en bagazo abre las puertas a esta alternativa.

Para que M9F1 pueda utilizarse en el marco de un proceso industrial es necesario aumentar significativamente los rendimientos de etanol obtenidos en este trabajo. Se debe tener en cuenta que M9F1 es un fragmento de ADN natural. Como tal, no está optimizado para expresar los genes relevantes para la degradación de lignocelulosa. De hecho, existen otros genes en M9F1 que seguramente no participan de este proceso. Por lo tanto, un

primer paso para mejorar el rendimiento de la conversión de lignocelulosa en etanol es colocar los genes relevantes en un contexto genético adecuado para maximizar las funciones buscadas. De esta manera, se podrá mejorar su expresión y eliminar el material genético irrelevante.

En conclusión, además de generar un organismo capaz de convertir biomasa vegetal en etanol, el principal beneficio de este Proyecto ha sido el desarrollo y la validación de un proceso basado en la metagenómica funcional que permite obtener elementos genéticos novedosos de potencial aplicación agroindustrial. Este proceso puede emplearse para mejorar el organismo obtenido pero también para buscar funciones metabólicas que viabilicen otras bio-conversiones relevantes como ser grasas en biodiesel, glicerol en etanol, o vinazas ricas en proteínas en monómeros industriales como el estireno

## AGRADECIMIENTOS

Queremos agradecer muy especialmente al Dr. Lonnie Ingram y al Dr. Ismael Nieves de la Universidad de Florida en Gainesville (EE.UU.) por facilitarnos gentilmente la cepa LY180 y el plásmido pLOI297 así como también permitirnos utilizar sus laboratorios e instalaciones para llevar adelante algunos de los estudios que aquí se presentan. También queremos agradecer a la Lic. Daniella Senatore por su apoyo técnico en la construcción de las metagenotecas, a la Dra. Liliana Borzacconi y al M.Sc. Mauricio Passeggi por facilitarnos las muestras de lodos de los reactores anaerobios, a la Ing. Betiana Bouzas de Frigorífico Pando por la provisión de muestras de rumen y al Ing. Nikolai Guchin de ANCAP por proveernos bagacillo de caña de azúcar obtenido del ingenio azucarero de ALUR en Bella Unión (Artigas, Uruguay). Este trabajo fue financiado por el Proyecto INIA FPTA 2007-256. También contó con financiación de la Agencia Nacional de Investigación e Innovación (Proyecto FSE 2009-23) y de PEDECIBA-Biología.

## BIBLIOGRAFÍA

**AAKVIK, T.; LALE, R.; LILES, M.; VALLA, S.** 2011. Metagenomic Libraries for Functional Screening, p. 171-181. In F. de Bruijn (ed.),

Handbook of Molecular Microbial Ecology I, First ed. John Wiley & Sons, Inc.

**ALTERTHUM, F.; INGRAM, L. O.** 1989. Efficient ethanol production from glucose, lactose, and xylose by recombinant *Escherichia coli*. *Applied and Environmental Microbiology* **55**:1943-1948.

**CANTAREL, B. L.; COUTINHO, P. M.; RANCUREL, C.; BERNARD, T.; LOMBARD, V.; HENRISSAT, B.** 2009. The Carbohydrate-Active En *Zymes database* (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**:D233-8.

**DESANTIS, T. Z.; HUGENHOLTZ, P.; LARSEN, N.; ROJAS, M.; BRODIE, E. L.; KELLER, K.; HUBER, T.; DALEVI, D.; HU, P.; ANDERSEN, G. L.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069-72.

**GEDDES, C. C.; NIEVES, I. U.; INGRAM, L. O.** 2011. Advances in ethanol production. *Curr Opin Biotechnol* **22**:312-319.

**HAHN-HAGERDAL, B.; GALBE, M.; GORWA-GRAUSLUND, M. F.; LIDEN, G.; ZACCHI, G.** 2006. Bio-ethanol—the fuel of tomorrow from the residues of today. *Trends Biotechnol* **24**:549-56.

**HANDA, N.; TERADA, T.; DOI-KATAYAMA,**

- Y.; HIROTA, H.; TAME, J. R.; PARK, S. Y.; KURAMITSU, S.; SHIROUZU, M.; YOKOYAMA, S. 2005. Crystal structure of a novel polyisoprenoid-binding protein from *Thermus thermophilus* HB8. *Protein Sci* **14**:1004-10.
- HANDELSMAN, J. 2004. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiol. Mol. Biol. Rev.* **68**:669-685.
- IEA. 2012. Tracking clean energy progress. OECD/IEA, Paris.
- KHRAMTSOV, N., L. MCDADE, A. AMERIK, E. YU, K. DIVATIA, A. TIKHONOV, M. MINTO, G. KABONGO-MUBALAMATE, Z. MARKOVIC, M. RUIZ-MARTINEZ, AND S. HENCK. 2011. Industrial yeast strain engineered to ferment ethanol from lignocellulosic biomass. *Bioresour Technol* **102**:8310-3.
- KUMAR, R.; SINGH, S.; SINGH, O. 2008. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of Industrial Microbiology and Biotechnology* **35**:377-391.
- METAGENOMICS, C. O. 2007. The new science of metagenomics: Revealing the secrets of our microbial planet. The National Academies Press, Washington, DC.
- MILLER, E. N.; JARBOE, L. R.; YOMANO, L. P.; YORK, S. W.; SHANMUGAM, K. T.; INGRAM, L. O. 2009. Silencing of NADPH-dependent oxidoreductase genes (*yqhD* and *dkgA*) in furfural-resistant ethanologenic *Escherichia coli*. *Appl Environ Microbiol* **75**:4315-23.
- NICHOLS, N.; MONCEAUX, D.; DIEN, B.; BOTHAST, R. 2008. Production of ethanol from corn and sugarcane, p. 3-15. *In* J. D. Wall, C. S. Harwood, and A. Demain (ed.), *Bioenergy*. ASM Press, Washington, DC.
- OHTA, K.; BEALL, D. S.; MEJIA, J. P.; SHANMUGAM, K. T.; INGRAM, L. O. 1991. Metabolic engineering of *Klebsiella oxytoca* M5A1 for ethanol production from xylose and glucose. *Appl Environ Microbiol* **57**:2810-5.
- RAPPE, M. S.; GIOVANNONI, S. J. 2003. The uncultured microbial majority. *Annu Rev Microbiol* **57**:369-94.
- SIEBER, J. R.; SIMS, D. R.; HAN, C.; KIM, E.; LYKIDIS, A.; LAPIDUS, A. L.; MCDONNALD, E.; ROHLIN, L.; CULLEY, D. E.; GUNSALUS, R.; MCINERNEY, M. J. 2010. The genome of *Syntrophomonas wolfei*: new insights into syntrophic metabolism and biohydrogen production. *Environ Microbiol* **12**: 2289-301.
- SIMON, C.; DANIEL, R. 2009. Achievements and new knowledge unraveled by metagenomic approaches. *Appl Microbiol Biotechnol* **85**:265-76.
- STANCIK, L. M.; STANCIK, D. M.; SCHMIDT, B.; BARNHART, D. M.; YONCHEVA, Y. N.; SLONCZEWSKI, J. L. 2002. pH dependent expression of periplasmic proteins and amino acid catabolism in *Escherichia coli*. *J Bacteriol* **184**: 424658.
- SUN, S.; CHEN, J.; LI, W.; ALTINTAS, I.; LIN, A.; PELTIER, S.; STOCKS, K.; ALLEN, E. E.; ELLISMAN, M.; GRETHE, J.; WOOLEY, J. 2011. Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Res* **39**:D546-51.
- TANS, P.; KEELING, R. 2012, posting date. Trends in carbon dioxide. NOAA/ESRL (<http://www.esrl.noaa.gov/gmd/ccgg/trends/>). [Online.]
- TUCK, C. O.; PÉREZ, E.; HORVÁTH, I. T.; SHELDON, R. A.; POLIAKOFF, M. 2012. Valorization of Biomass: Deriving More Value from Waste. *Science* **337**:695-699.
- ZUROFF, T.; CURTIS, W. 2012. Developing symbiotic consortia for lignocellulosic biofuel production. *Applied Microbiology and Biotechnology* **93**:1423-1435.

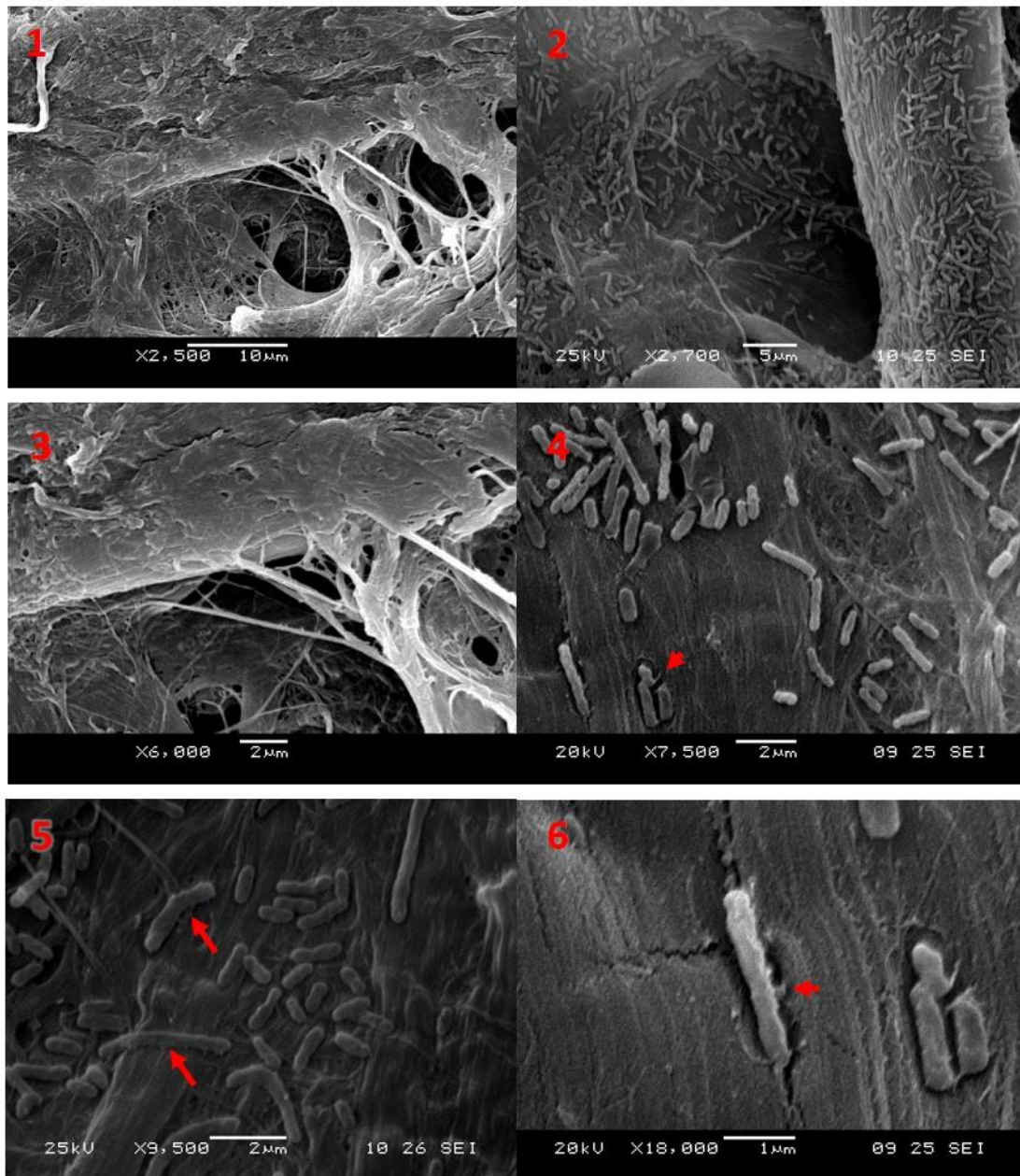
## **Ensayos complementarios: Microscopía electrónica de barrido.**

### **Metodología**

5 mL de medio mínimo M9-Cm12,5-Ara suplementado con una tira de papel de filtro estéril (200 x 40 mm) o 1% de bagazo de caña pretratado, fueron inoculados con el clon Epi300 (M9F1) y un clon control (cepa parental con un fósforo elegido al azar) e incubados a 37°C y 150 rpm. Luego de 10 días los sustratos se retiraron del medio y se secaron a temperatura ambiente. Las tiras de papel y el bagazo se fijaron con una solución de formaldehído al 2% (p/v) y se conservaron para su visualización por microscopía electrónica de barrido (MEB) utilizando el servicio brindado por Facultad de Ciencias, UdelaR.

### **Resultados**

Debido a que, durante la interacción entre el clon Epi300 (M9F1) y el papel de filtro. Se observaba la formación de un biofilm sobre las tiras de papel, se realizaron estudios de microscopía electrónica de barrido con el fin de indagar en las interacciones del microorganismo con las fibras de celulosa. La ausencia de halo en placas de CMC y el crecimiento en avicel permitían intuir de la expresión de al menos una exocelulasa con actividad en celulosa cristalina y de proteínas o módulos adicionales implicados en la unión a carbohidratos. Para esto, tiras de papel de filtro fueron incubadas con el clon Epi (M9F1) y un clon control y se observaron bajo MEB. En la Figura 8 se observa la interacción del microorganismo con las fibras de papel y la presencia de protuberancias que facilitan dicha interacción (119,232,233).



**Figura 8.** Imágenes de microscopía electrónica de barrido (SEM) mostrando la adherencia de las bacterias a las fibras de papel. La cepa Epi (M9F1) fue incubada durante 10 días en M9-PF a 30°C y 150 rpm. Las tiras de papel fueron retiradas del tubo, secadas a temperatura ambiente durante 12 horas y fijadas con formaldehído 2% hasta su visualización. Las imágenes 1 y 3 corresponden al control negativo. Las flechas señalan la presencia de protuberancias en la superficie bacteriana, las cuales facilitan la interacción con el sustrato.

Capítulo 3.

**LA INCORPORACIÓN DE UN FRAGMENTO  
DE ADN METAGENÓMICO EN LA CEPA DE  
*Escherichia coli* MS04 MEJORA LA  
PRODUCCIÓN DE ETANOL A PARTIR DE  
BIOMASA**

---



Loaces, I., Amarelle, V., Muñoz-Gutierrez, I., Fabiano, E., Martínez, A. & Noya, F. (2015) **Improved ethanol production from biomass by a rumen metagenomic DNA fragment expressed in *Escherichia coli* MS04 during fermentation.** Applied Microbiology and Biotechnology. Doi: 10.1007/s00253-015-6801-0.

## Resumen

A partir de fluido ruminal se construyó una librería metagenómica la cual fue evaluada en busca de enzimas celulolíticas. Uno de los fósmidos seleccionados, Csd4, confiere a *Escherichia coli* la habilidad de utilizar avicel, carboximetil-celulosa (CMC), papel de filtro, bagazo de caña de azúcar pre-tratado y xilano de avena como fuente de carbono. Csd4 fue secuenciado en su totalidad y se identificaron una endo-glucanasa, una xilosidasa/ arabinosidasa y una lacasa entre los 28 ORFs anotados. Los análisis filogenéticos mostraron que deriva de un organismo relacionado a *Prevotella ruminicola*, aunque no se encontraron homólogos en las bases de datos. La incorporación del fósido en la cepa etanológica de *E. coli* MS04 permitió la producción de etanol a partir de CMC, avicel, papel de filtro y bagazo de caña de azúcar. El agregado de una  $\beta$ -glucosidasa tuvo un efecto positivo en el crecimiento de la cepa en CMC. La producción de etanol a partir de bagazo de caña de azúcar, luego de la sacarificación con las celulasas comerciales de *Trichoderma reesei*, fue 3 veces mejor con MS04 (Csd4) respecto a MS04 con un fósido control.

## Objetivos

- ☞ Seleccionar, dentro de un metagenoma de rumen vacuno, clones con actividad celulolítica.
- ☞ Caracterizar el fenotipo y el genotipo del clon Epi (Csd4).
- ☞ Evaluar el potencial de la cepa MS04 complementada con el fósido Csd4 para producir etanol a partir de sustratos celulósicos.

## Principales resultados y conclusiones

El análisis funcional de 30.000 clones metagenómicos reveló la presencia de 45 clones con actividad endo-glucanasa positiva, visualizada por la formación de halos en placas de LB suplementado con CMC. Extractos celulares de 9 clones fueron, a su vez, capaces de hidrolizar el sustrato 4-metil-umberiferil-celobiósido (4-MUC), mostrando actividad exo-celulasa. Uno de ellos, denominado Epi (Csd4), presentó ambas actividades en células intactas y en el sobrenadante, por lo que fue seleccionado para su caracterización fenotípica y genotípica. Ensayos complementarios mostraron que la actividad 4-MUCasa de Epi (Csd4) es completamente inhibida en presencia de glucosa o xilosa, pero no cuando se adiciona avicel, CMC o celobiosa. Este clon fue capaz de crecer en medio mínimo suplementado con avicel, bagazo de caña de azúcar, papel de filtro o xilano de avena luego de 48 horas de incubación a 30°C. El fósido Csd4 (34.406 pb) fue secuenciado en su totalidad y sus ORFs anotados. La comparación de la secuencia con las bases de datos permitieron inferir que Csd4 deriva de un microorganismo relacionado a *P. ruminicola*, pero que dicho organismo aún no está secuenciado. Los ORFs inferidos en Csd4 fueron sometidos a análisis de BLASTp y se detectaron una endo-glucanasa perteneciente a la familia de glicosil-hidrolasas 5 (GH5), una xilosidasa/ arabinosidasa y una lacasa.

La complementación de la cepa MS04 con el fósido Csd4 le otorgó la capacidad de crecer y producir etanol en medio mínimo suplementado con CMC, avicel, papel de filtro y bagazo de caña de azúcar pre-tratado. A su vez, el agregado de  $\beta$ -glucosidasa exógena mejoró el rendimiento de la cepa cuando se usó CMC como fuente de carbono.

Estos resultados nos permiten concluir que Csd4 es suficiente y necesario para conferir a cepas de *E. coli* la habilidad de degradar y metabolizar sustratos celulósicos. Los genes codificados en este fósido se encuentran bajo la regulación nativa del microorganismo de origen. Su clonado y sobre-expresión permitiría, potencialmente, mejorar los rendimientos obtenidos.

## Contribución personal al trabajo

En este trabajo realicé la construcción de la librería metagenómica y el screening de la misma con la colaboración de V. Amarelle. Llevé a cabo los ensayos fenotípicos y de regulación del clon Epi (Csd4). Participé en los análisis filogenéticos y de asignación de funciones de los ORFs detectados. Los ensayos de fermentación los realicé en las instalaciones del Laboratorio de Ingeniería de Vías Metabólicas de Bacterias, Departamento de Ingeniería Celular y Biocatálisis, del Instituto de

Biotecnología de la Universidad Nacional Autónoma de México (IBT-UNAM), con la asesoría del Dr. Alfredo Martínez.

El trabajo presentado en este capítulo fue financiado parcialmente por un proyecto FPTA del Instituto Nacional de Investigación Agropecuaria (INIA), por un proyecto FSE de la Agencia Nacional de Investigación e Innovación (ANII) y una beca de movilidad de la Agencia Mexicana de Cooperación Internacional para el Desarrollo de la Secretaría de Relaciones Exteriores de México (AMEXCID-SRE) y por PEDECIBA.

# Improved ethanol production from biomass by a rumen metagenomic DNA fragment expressed in *Escherichia coli* MS04 during fermentation

Inés Loaces<sup>1</sup> · Vanesa Amarelle<sup>1</sup> · Iván Muñoz-Gutiérrez<sup>2,3</sup> ·  
Elena Fabiano<sup>1</sup> · Alfredo Martínez<sup>2</sup> · Francisco Noya<sup>1</sup>

Received: 4 March 2015 / Revised: 21 June 2015 / Accepted: 24 June 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** With the aim of improving current ethanologenic *Escherichia coli* strains, we screened a metagenomic library from bovine ruminal fluid for cellulolytic enzymes. We isolated one fosmid, termed Csd4, which was able to confer to *E. coli* the ability to grow on complex cellulosic material as the sole carbon source such as avicel, carboxymethyl cellulose, filter paper, pretreated sugarcane bagasse, and xylan. Glucanolytic activity obtained from *E. coli* transformed with Csd4 was maximal at 24 h of incubation and was inhibited when glucose or xylose were present in the media. The 34,406-bp DNA fragment of Csd4 was completely sequenced, and a putative endoglucanase, a xylosidase/arabinosidase, and a laccase gene were identified. Comparison analysis revealed that Csd4 derived from an organism closely related to *Prevotella ruminicola*, but no homologies were found with any of the genomes already sequenced. Csd4 was introduced into the ethanologenic *E. coli* MS04 strain and ethanol production from CMC, avicel, sugarcane bagasse, or filter paper was observed. Exogenously expressed  $\beta$ -glucosidase had a positive effect on cell growth in agreement with the fact that no putative  $\beta$ -glucosidase was found in Csd4. Ethanol production from sugarcane bagasse was improved threefold by

Csd4 after saccharification by commercial *Trichoderma reesei* cellulases underlining the ability of Csd4 to act as a saccharification enhancer to reduce the enzymatic load and time required for cellulose deconstruction.

**Keywords** Rumen metagenomics · CBP · Cellulolytic enzymes · Bioethanol · Biomass hydrolysis

## Introduction

The sustainable and competitive production of lignocellulosic ethanol depends on several factors, including the improvement of expression levels and performance of recombinant cellulolytic enzymes that are responsible for releasing the fermentable sugars from cellulose during the saccharification step (Lambertz et al. 2014). A major bottleneck for the competitive production of lignocellulosic ethanol is the high enzyme load during the saccharification step of pretreated lignocellulosic biomass (Gao et al. 2011; Lambertz et al. 2014). Despite recent efforts to reduce the production cost of these enzymes, their contribution to the total cost of the resulting biofuel remains significantly high (Klein-Marcuschamer et al. 2012; Tao et al. 2014). The efficient hydrolysis of lignocellulose is performed by the synergistic action of a multicomponent enzymatic system consisting of at least three major enzymatic groups: cellulases that degrade the main cellulose chain composed of  $\beta$ -1,4-linked units of glucose; hemicellulases that degrade the more complex hemicellulose structure mainly releasing xylose, arabinose, and glucose, but also mannose and galactose; and laccases and peroxidases that attack the recalcitrant polyphenolic structure of lignin (French et al. 2013; Horn et al. 2012; Shin et al. 2010).

The most widely used process configuration for biofuel production from starch and lignocellulose is named separate

✉ Francisco Noya  
fnoya@iibce.edu.uy

<sup>1</sup> Departamento de Bioquímica y Genómica Microbianas, Instituto de Investigaciones Biológicas Clemente Estable, Av. Italia 3318, Montevideo 11600, Uruguay

<sup>2</sup> Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, A.P. 510-3., Cuernavaca, Morelos 62270, México

<sup>3</sup> Present address: Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

hydrolysis and fermentation (SHF) because the hydrolysis and the fermentation steps take place in different reactors. Under this setup, conditions dictated for the enzymatic reaction differ from the conditions required by the fermentative microorganism (Sainz 2009). So far, no single natural microorganism can produce the ideal enzyme mix for biomass hydrolysis. Therefore, industrial enzymes consist of mixtures comprising recombinant proteins formulated for particular biotechnological applications. Currently, the commercial cellulase cocktails, derived from the fungi *Trichoderma reesei* and *Aspergillus*, work at pH 5 and 50 °C (Lambertz et al. 2014; Merino and Cherry 2007), but growth and fermentation conditions of industrial microorganism are pH 5–7 and 30–37 °C. Therefore, given better suited enzymes, a superior configuration in terms of cost is the simultaneous saccharification and fermentation (SSF) process, in which both steps take place in the same reactor (Margeot et al. 2009; Mingardon et al. 2011). In addition to reducing the number of reactors, the SSF process diminishes the end-product inhibition of cellulases (Olofsson et al. 2008).

An attractive alternative to both SHF and SSF is the consolidated bioprocessing (CBP) in which a single microorganism can degrade the cellulosic biomass and ferment the resulting sugars into biofuels such as ethanol (Lynd et al. 2005). A natural organism combining fast growth on cellulosic substrates and the ability to produce biofuels has not been isolated yet. The development of engineered microorganisms offers a feasible alternative (Ingram et al. 1999; Mingardon et al. 2011). *Escherichia coli* offers a good platform for the CBP because there are many genetic tools available to facilitate its genetic modification, can grow in mineral salt media without complex supplements, and can ferment a wide variety of fermentable sugars, such as hexoses, pentoses, and uronic acids (Muñoz-Gutiérrez and Martínez 2013; Orencio-Trejo et al. 2010). Several strategies involving expression of heterologous cellulases in fermentative ethanologenic strains are currently being pursued (Kojima et al. 2012; Linger et al. 2010; Muñoz-Gutiérrez et al. 2014; Zheng et al. 2012) including the development of ethanologenic *E. coli* strains engineered to heterologously express saccharolytic enzymes to produce ethanol directly from lignocellulosic biomass (Edwards et al. 2011; Ryu and Karim 2011).

Functional metagenomics is an experimental strategy that allows the isolation of enzymatic activities of interest that are present in a consortium of microorganisms (Aakvik et al. 2011). This approach has been extensively employed for mining new biocatalysts to improve many industrially relevant processes in biotechnology and pharmaceuticals (Leis et al. 2013). Several metagenome-derived cellulase genes have been identified in metagenomic libraries made from natural microbial communities such as cow rumen, termite gut, soil, and plant- or algae-associated microorganisms (Bao et al. 2011; Bhat et al. 2013; Duan et al. 2009; Martin et al. 2014;

Ni and Tokuda 2013; Voget et al. 2006). Functional metagenomics offers the opportunity to find saccharolytic enzymes that are more efficient at the growth condition of the fermentative microorganism. The heterologous expression of these novel saccharolytic enzymes in the fermentative microorganism can be used to decrease the commercial enzyme loading. Bovine rumen contains a huge consortium of microorganisms with the capacity to degrade the recalcitrant plant cell wall biomass (White et al. 2014) representing an important source for the discovery of lignocellulolytic enzymes (Brulc et al. 2009).

In this work, using a functional metagenomic approach, a cellulolytic clone from bovine rumen was identified and further characterized. The 34-kbp DNA sequence comprises 28 open reading frames including three for putative enzymes involved in complex carbohydrates degradation: an endoglucanase, a xylosidase/arabinoxylanase, and a laccase. The fragment was introduced into the ethanologenic *E. coli* MS04 strain (Fernández-Sandoval et al. 2012) and ethanol production from carboxymethyl cellulose (CMC), avicel, sugarcane bagasse, and xylan was evaluated. Complementation with a heterologous  $\beta$ -glucosidase was also assayed.

## Materials and methods

### Metagenomic library construction

Metagenomic DNA was isolated from ruminal fluid of cows fed on natural pastures as previously described (Rodríguez et al. 2015) and size fractionated by electrophoresis on agarose gels. The 30–50-kbp fraction was isolated and cloned using a Copy Control Fosmid Library Production Kit (Epicentre Biotechnologies, Madison, WI, USA) into the fosmid vector pCC1FOS (Epicentre Biotechnologies) following the manufacturer's instructions. EPI300-T1<sup>R</sup> cells were transformed and selected in Luria broth (LB) solid media supplemented with 12.5  $\mu$ g/mL chloramphenicol (Cm). The library was arranged in 96-well plates with LB-Cm liquid medium and stored with 25 % (v/v) glycerol at –80 °C.

### Library screening

For the identification of clones positive for cellulase activity, the metagenomic library was replicated onto LB-Cm agar plates containing 0.01 % (v/v) arabinose, to induce multiple copies of the fosmid, and 0.5 % (w/v) CMC sodium salt (DP: 400). The plates were incubated at 30 °C for 5 days. Then, the plates were washed with distilled water and stained with Congo red (Teather and Wood 1982). Clones expressing CMC activity were identified by the formation of clear orange halos around the colony. Fosmids were isolated from positive clones using QIAprep<sup>R</sup> Miniprep kit (Qiagen,

Limburg, Netherlands) from the primary screening and were re-introduced into fresh electro-competent *E. coli* EPI300-T1<sup>R</sup> cells to confirm their phenotypes. All the positive clones were further assessed for 1,4- $\beta$ -D-glucanase activity using the fluorogenic substrate 4-methylumbelliferyl  $\beta$ -D-cellobioside (4-MUC) as reporter substrate. Clones were grown overnight in 96-well plates with 200  $\mu$ l of liquid LB-Cm with 0.01 % of arabinose. The cells were lysed by freeze-thaw cycles, and the cell extracts were incubated with 0.01 % 4-MUC for 1 h at 37 °C. The plates were observed under UV light and positive clones for 4-MUCase were detected by blue light emission due to the release of 4-methylumbelliferone.

Sugarcane bagasse was obtained from ALUR (Montevideo, Uruguay). All remaining substrates were purchased from Sigma-Aldrich, Saint Louis, MO, USA.

### Characterization of cellulolytic positive clones

Filter paper, xylan, and sugarcane bagasse were evaluated as sole carbon sources in M9 minimal media (2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 1.28 % Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, 0.3 % KH<sub>2</sub>PO<sub>4</sub>, 0.05 % NaCl, 0.1 % NH<sub>4</sub>Cl, 0.005 % thymine, 200  $\mu$ M leucine) supplemented with 1 % (w/v) of each substrate. A clone named Epi300 (Csd4) was grown overnight in M9 media supplemented with 1 % of glucose at 37 °C and 150 rpm. Cells were washed and resuspended in minimum media. Five microliters of cells were placed on final media plate, and the growth was followed over a period of 5 days. Kinetic of glucanase activity was determined in 200  $\mu$ L of LB-Cm, containing 0.01 % (v/v) arabinose and 0.05 % 4-MUC, inoculated with Epi300 (Csd4) (initial OD<sub>620</sub> 0.05) and supplemented with 0.5 % (w/v) of either glucose, xylose, cellobiose, CMC, avicel PH-101, or not supplemented. Fluorescence was determined (excitation wavelength, 365 nm; emission wavelength, 445 nm) every 8 h for 60 h in 96-well plates using Varioskan Flash Multimode Reader (Thermo Scientific, Waltham, MA, USA).

### Sequencing and gene annotation

Fosmid Csd4 sequence was determined by a shotgun library construction followed by Sanger sequencing of individual clones performed and assembled by Macrogen Inc. (Seoul, South Korea). Primer walking was used to close the gaps between contigs. Bioinformatic tools were run on the WebMGA server (Wu et al. 2011). Open reading frame (ORF) prediction was done by MetaGene (Noguchi et al. 2006) and annotation by RPS Blast (NCBI, USA). Comparison with the rumen metagenome from JGI (Hess et al. 2011) was done using IMG/M (Markowitz et al. 2012) and Blast (Zhang et al. 2000).

The DNA sequence of Csd4 is available in the GenBank database under the accession number KP843855.

### Preparation of *Thermobifida fusca* $\beta$ -glucosidase BglC

*T. fusca*  $\beta$ -glucosidase BglC (Spiridonov and Wilson 2001) was expressed on the surface of MS04 (pAIDABglCRHis) cells and prepared as previously described (Muñoz-Gutiérrez et al. 2012). It was added at a final concentration of 10 U/g of substrate in the fermentation assay when indicated. For determination of activity, cells were suspended on 3.65 mL 50 mM phosphate buffer (pH 7) and incubated with 100 mM *p*-nitrophenyl- $\beta$ -D-glucopyranoside (100  $\mu$ L) (*p*NPG, Sigma-Aldrich) at 37 °C. The kinetics were followed by taking samples (750  $\mu$ L each) at regular time intervals. The reaction was terminated by adding 250  $\mu$ L of 2 M Na<sub>2</sub>CO<sub>3</sub> to each sample. The absorbance, at 400 nm, was measured after cells were centrifuged at 9300 $\times$ g for 2 min, and enzyme activity was calculated using the extinction coefficient of *p*-nitro-phenol. One pNPGase unit was defined as the amount of BglC attached to the cells that catalyzes the formation of 1  $\mu$ mol pNP per minute (Muñoz-Gutiérrez et al. 2012).

### Fermentation assays

Fermentations were performed using the ethanologenic strain *E. coli* MS04 (Fernández-Sandoval et al. 2012) carrying the Csd4 fosmid or a randomly picked fosmid as negative control, transferred by electroporation. During strain construction, cultures were maintained aerobically in LB with 50  $\mu$ g/mL kanamycin and 25  $\mu$ g/mL chloramphenicol.

Batch fermentations were conducted in Fleakers mini-fermenters (Beall et al. 1991) containing 200-mL minimal AM1 medium (Martínez et al. 2007) supplemented with 1 % (w/v) of carbon sources, 0.01 % (w/v) citric acid and 0.2 % (w/v) sodium acetate (Sigma-Aldrich, Saint Louis, MO, USA) (Fernández-Sandoval et al. 2012) at 37 °C and 150 rpm, without aeration. Automatic addition of 2 N KOH preserved a constant pH. Prior to fermentations, the cells were adapted to grow in mini-fermenters as described previously (Muñoz-Gutiérrez et al. 2012). The inoculum was prepared from an overnight culture in 200-mL AM1 medium supplemented with 2 % glucose without aeration, at 37 °C. When the cells reached an OD<sub>600</sub> of 0.5, they were centrifuged for 10 min at 4000 $\times$ g and seeded into the mini-fermenters at an initial OD<sub>600</sub> of 0.1.

Mixed fermentation assays were performed with glucose and CMC as carbon sources and compared with fermentation with glucose alone. Avicel, xylan from birch wood, CMC, filter paper, and sugarcane bagasse were evaluated as carbon sources for ethanol production. A mini-fermenter containing 1 % (w/v) of each substrate in AM1 medium was inoculated at an initial OD<sub>620</sub> of 0.5 and incubated for 48 h. AM1 with glucose or with no carbon sources were employed as controls. Effect of exogenous BglC (10 U/g of substrate) on ethanol production from 1 % (w/v) CMC was also assayed in

minimum media AM1. All experiments were conducted in triplicates.

### Saccharification and fermentation of sugarcane bagasse

Sugarcane bagasse was pretreated with 1 % (w/v) sulfuric acid for 1 h with continuous stirring at 150 rpm, and room temperature, followed by 1 h of autoclave at 121 °C. The solid fraction was separated by centrifugation, washed with distilled water three times until pH neutralization, and conserved at 4 °C until used. Ten grams of pretreated biomass was suspended in 100 mL of 50 mM sodium citrate pH 5.5, and cellulase from *T. reesei* (Sigma-Aldrich, Saint Louis, MO, USA) was added at 40 U/g of biomass and incubated at 45 °C for 24 h. After enzymatic treatment, the suspension was added to 200 mL of AM1 minimal media at a final concentration of 1 % (w/v), in the presence of BglC (10 U/g of substrate) and inoculated with MS04 or MS04 (Csd4) for 72 h at 37 °C, pH 7, and 150 rpm using mini-fleakers. Samples were centrifuged, and the cell-free culture broth was frozen until analysis.

### Quantitative determinations

Cell growth was determined spectrophotometrically as optical density at 600 nm (OD<sub>600</sub>). The concentration of glucose was measured with an enzymatic analyzer (Model 2700, YSI Inc., Yellow Springs, OH, USA). The ethanol was analyzed by gas chromatography using n-butanol (1 %) as internal standard (6850 Series GC System, Agilent, Santa Clara, CA, USA). Alternatively, ethanol and oxygen were monitored continuously by gas phase sensors BCP-EtOH and BCP-O<sub>2</sub>EC, respectively (BlueSens, Herten, Germany).

## Results

### Identification of *E. coli* Epi300 (Csd4)

A 30,000 clone metagenomic library was obtained from the ruminal fluid of a cow fed on natural pastures. Forty-five clones were positive for cellulase activity using CMC as substrate. These positive clones, identified by a halo of hydrolysis around the colony, were selected and confirmed by isolating the containing fosmid and re-introducing it into fresh *E. coli* EPI300-T1<sup>R</sup> cells. Restriction analysis with *NotI* showed that insert sizes were 40 kb on average. Only the fosmids that consistently reproduced the hydrolysis halo were selected for further analysis. Five of the 45 positive clones were also positive for 4-MUC activity in cellular extracts. One of these, which also displayed this activity in intact cells (data not shown), named Epi300 (Csd4), was selected for further characterization. Csd4 4-MUCase activity was also detected in the cell-free supernatant of a 24-h culture in LB media.

### Fosmid Csd4 sequence analysis

A shotgun library of Csd4 with fragment size of 3 kb on average was constructed. Two hundred ninety-four Sanger reads were assembled into two contigs resulting in a coverage of 8.05X. Primer walking was used for gap closing. Final assembly, after filtering out the vector sequence, produced a single contig of 34,406 bp corresponding to the insert of Csd4. Using MetaGene (Noguchi et al. 2006), 28 ORFs were predicted in Csd4. COGs were assigned by RPS Blast, and the closest homologs were identified by BlastP (Table 1). Most of the predicted protein sequences of these ORFs displayed high identity with proteins from uncultured organisms and from bacteroidetes mainly from genus *Prevotella* and *Bacteroides*. The Csd4 sequence was compared with the *Prevotella* genomes deposited in NCBI databases, including genomes from the Human Microbiome Project (Aagaard et al. 2013), and with members of the *Prevotella* group so far sequenced by the Hungate1000 Project (Creevey et al. 2014), but only local identities were identified (data not shown). However, regions of the Csd4 sequence showed more extensive similarities with scaffolds from the Joint Genome Institute (JGI) metagenome “Bovine rumen microbial communities from Illinois University” (Taxon Object ID 2061766007) (Hess et al. 2011). The scaffolds with higher scores (over 800 bits) were mainly associated with the order Bacteroidales and about half of them with the genus *Prevotella*. In fact, ORFs Csd4\_22 and Csd4\_23 showed high homology with putative ORFs from JGI rumen microbiome (Locus IDs HiSeq\_04030830 and HiSeq\_02855500, respectively) that are located in scaffolds associated with *Prevotella ruminicola* (HiSeq\_NODE\_2404729 and HiSeq\_NODE\_1861550, respectively). Also, the genomic context upstream of Csd4\_22 showed shared synteny to the one upstream of HiSeq\_04030830 suggesting a close phylogenetic relationship (Fig. 1). However, none of these scaffolds could be globally aligned with the sequence of Csd4. These observations suggest that Csd4 derives from a *Prevotella* whose genome has not been sequenced yet and that it is neither represented in the JGI rumen microbiome nor in the genomes currently sequenced by the Hungate1000 Project.

Three ORFs with predicted protein sequences putatively associated with lignocellulose degradation were identified in Csd4: Csd4\_16, Csd4\_22, and Csd4\_23 (Table 1).

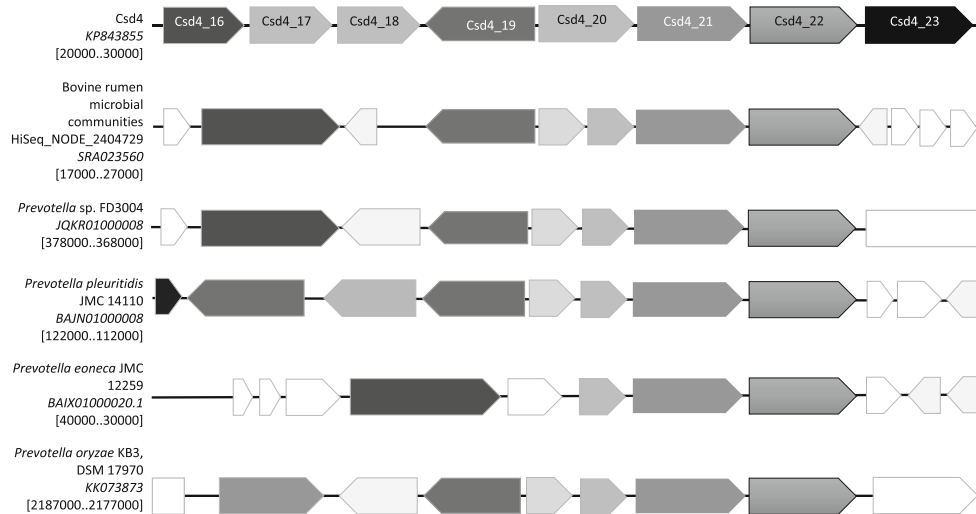
The ORF Csd4\_16 encoding a putative  $\beta$ -xylosidase/arabinoxidase belonging to CAZy glycoside hydrolase family 43 (GH43) (Cantarel et al. 2009) was detected between nucleotide positions 20131 and 21648. GH43 includes mostly enzymes that have  $\beta$ -1,4-xylosidase activity (EC 3.2.1.37). They are part of an array of hemicellulases that are involved in the final breakdown of plant cell wall. They hydrolyze  $\beta$ -1,4 glycosidic bonds between two xylose units in short xylooligosaccharides (Barker et al. 2010). Csd4\_16 showed

**Table 1** ORFs assigned to Csd4 sequence and their closest homologs

ORF <sup>a</sup>	Start	End	Strand	Length (aa)	COG <sup>b</sup>	Description [organisms] <sup>c,d</sup>	Identity (%) <sup>e</sup>	E-value <sup>e</sup>	Accession no. <sup>e</sup>
Csd4_1	88	903	+	271					
Csd4_2	1383	2162	+	259		Hypothetical protein [ <i>Prevotella buccalis</i> DNF00985]	56	6.00E-94	KGF40528.1
Csd4_3	2251	3588	-	445	COG0334	Glutamate dehydrogenase [ <i>Prevotella</i> ]	87	0.00E+00	WP_028902358.1
Csd4_4	3897	4559	+	220		Putative uncharacterized protein [ <i>Prevotella</i> sp. CAG:1185]	33	8.00E-30	WP_021963427.1
Csd4_5	4615	8073	+	1152		Hypothetical protein [ <i>Prevotella</i> sp. AGR2160]	37	3.00E-101	WP_028910338.1
Csd4_6	8535	11516	+	993	COG0574	Phosphoenolpyruvate synthase [ <i>Prevotella</i> sp. CAG:1124]	80	0.00E+00	WP_021948672.1
Csd4_7	12022	12651	+	209	COG0810	Cell envelope biogenesis protein TonB [ <i>Prevotella paludivivens</i> ]	51	7.00E-50	WP_018464825.1
Csd4_8	12691	13413	-	240		RDD family protein [ <i>Prevotella veroralis</i> ]	39	5.00E-45	WP_018910394.1
Csd4_9	13552	14469	+	305	COG0574	Membrane protein [ <i>Bacteroides acidifaciens</i> ]	61	7.00E-118	WP_024988293.1
Csd4_10	14459	15322	+	287		Membrane protein [ <i>Bacteroides</i> ]	27	5.00E-18	WP_007663060.1
Csd4_11	15335	15955	+	206		Hypothetical protein [ <i>Prevotella multisaccharivorax</i> ]	43	1.00E-19	WP_007575351.1
Csd4_12	16099	17364	+	421		Hypothetical protein [ <i>Bacteroides finegoldii</i> ]	31	3.00E-51	WP_007767102.1
Csd4_13	17426	18400	+	324	COG0714	Magnesium chelatase [ <i>Prevotella histicola</i> ]	71	2.00E-172	WP_025792001.1
Csd4_14	18478	19191	+	237	COG1721	Hypothetical protein [ <i>Bacteroides acidifaciens</i> ]	54	3.00E-81	WP_024988291.1
Csd4_15	19139	19792	+	217	COG1721	Uncharacterized protein [ <i>Bacteroides</i> sp. CAG:754]	66	1.00E-98	CDA86114.1
<b>Csd4_16</b>	20131	21648	+	505	COG3507	<b><math>\beta</math>-Xylosidase family GH43 [<i>Prevotella</i> sp. CAG:732]</b>	61	0.0	<b>WP_022328427.1</b>
Csd4_17	21830	22504	+	224	COG0580	Porin [ <i>Prevotella enoea</i> ]	63	4.00E-91	WP_025065050.1
Csd4_18	22749	23759	+	336	COG0191	Fructose-bisphosphate aldolase [ <i>Prevotella</i> ]	92	0.00E+00	WP_004341759.1
Csd4_19	23812	25335	-	507	COG0062	Putative YjeF-related sugar kinase [ <i>Prevotella</i> sp. CAG:732]	68	0.00E+00	WP_022328422.1
Csd4_20	25426	26562	+	378	COG0634	Adenylate kinase [ <i>Prevotella stercorea</i> CAG:629]	84	7.00E-111	WP_022429669.1
Csd4_21	26566	27729	+	387	COG0536	GTPase CgtA [ <i>Prevotella buccae</i> ]	87	0.00E+00	WP_004341745.1
<b>Csd4_22</b>	27726	28517	+	263	COG1496	<b>Polyphenol oxidase laccase [<i>Prevotella</i> sp. S7MS2]</b>	59	3.00E-112	<b>KG161157.1</b>
<b>Csd4_23</b>	28593	29741	+	382	COG2730	<b>Endoglucanase [uncultured microorganism]</b>	84	0.0	<b>ADB80112.1</b>
Csd4_24	29793	29999	-	373	COG0138	aICARFT/IMPase Hase bienzyme [ <i>Prevotella</i> sp. CAG:1185]	89	0.0	WP_021965082.1
Csd4_25	30011	31132	-	133	COG0335	50S ribosomal protein L19 [ <i>Prevotella maculosa</i> ]	79	4.00E-64	WP_008564746.1
Csd4_26	31317	31718	-	255	COG0737	Putative uncharacterized protein [ <i>Prevotella</i> sp. CAG:255]	53	5.00E-92	WP_021855259.1
Csd4_27	31995	32762	+	281	COG0737	5'-Nucleotidase [ <i>Prevotella</i> sp. CAG:592]	66	2.00E-141	WP_022316292.1
Csd4_28	32767	33612	+	223		Hypothetical protein [ <i>Prevotella histicola</i> ]	73	4.00E-121	WP_008822032.1

<sup>a</sup> ORFs identified by MetaGene<sup>b</sup> Assigned by RPS Blast<sup>c</sup> Closest homolog identified by BlastP 2.2.30+ using nonredundant protein sequences<sup>d</sup> Genes associated with lignocellulose degradation are shown in bold





**Fig. 1** Schematic representation of the Csd4 metagenomic fragment between positions 20,000 and 30,000 bp (*first row*). Approximate positions of putative ORFs and their direction of transcription are indicated. ORFs potentially related with lignocellulose degradation are indicated with *black boxes*. Below Csd4 are shown regions of rumen

microbiome scaffolds and genomes of organisms with similar genomic organization. *Boxes marked with similar shades of gray*, positions and orientations relative to the ones in Csd4 represent putative orthologs ORFs. Accession numbers are indicated in *italics* followed by approximate fragment positions between *square brackets*

homology with a xylosidase gene from *Prevotella* sp. CAG:732 (WP\_022328427.1) with an E-value undistinguishable from 0 and 61 % identity at the protein sequence.

The ORF Csd4\_22 deduced protein sequence has 59 % identity with a polyphenol oxidase laccase of *Prevotella* sp. S7MS2 (KGI61157.1) belonging to COG1496. This cluster also includes RL5, a confirmed multicopper polyphenol oxidase with laccase activity (EC 1.10.3.2) isolated from cow rumen (Beloqui et al. 2006). Csd4\_22 has a 43 % identity with RL5 (E-value  $1E-62$ ). Laccases are able to oxidize a wide variety of phenolic and nonphenolic compounds and are widely distributed among both prokaryotes and eukaryotes. It is worth noting that the screening that elicited Csd4 was not designed to detect either xylanases or laccases.

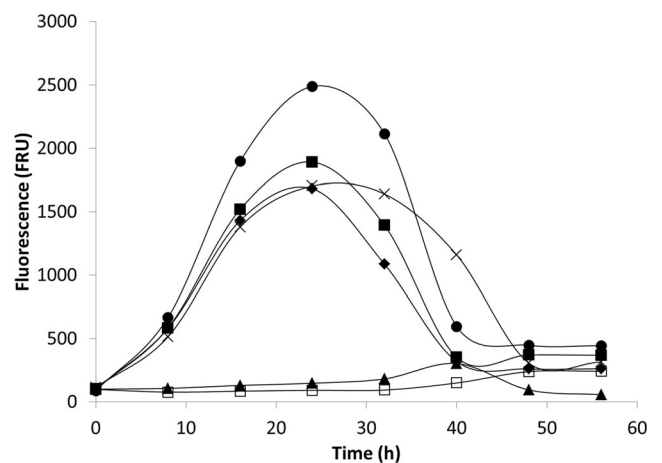
Next to Csd4\_22, the ORF Csd4\_23 was assigned into COG2730. This cluster has been clearly linked to endo- $\beta$ -1,4-glucanases (EC 3.2.1.4) belonging to GH5 family (Aspeborg et al. 2012; Gupta et al. 2011) that also includes endo- $\beta$ -1,4-xylanases (EC 3.2.1.8) and  $\beta$ -glucosidases (EC 3.2.1.21) among others. GH5 enzymes are abundant in different ecological niches, which have been evidenced by their frequent identification in various metagenomes studies (Duan and Feng 2010; Hess et al. 2011).

### Characterization of the Epi300 (Csd4) clone

The *E. coli* Epi300 (Csd4) clone was selected based on its ability to generate a halo of hydrolysis on LB-CMC plates after 5 days of incubation at 30 °C. To verify if Csd4 was able to confer to *E. coli* EPI300-T1<sup>R</sup>, the capacity to grow on

complex cellulosic substrates as the sole carbon source, avicel, filter paper, pretreated sugarcane bagasse or xylan were added to solid minimum media M9. Epi300 (Csd4) growth was evaluated after 48 h of incubation at 30 °C. The strain was able to form colonies on all of these substrates (data not shown). This allowed us to conclude that Csd4 cellulolytic genes were successfully expressed in *E. coli*.

To assess the regulation of Csd4 cellulolytic genes, enzymatic activity was determined in complete cells using 4-MUC as a reporter substrate. Fluorescence was measured during culture growth in rich medium (LB) supplemented with glucose, xylose, cellobiose, CMC, or avicel. As shown in Fig. 2,



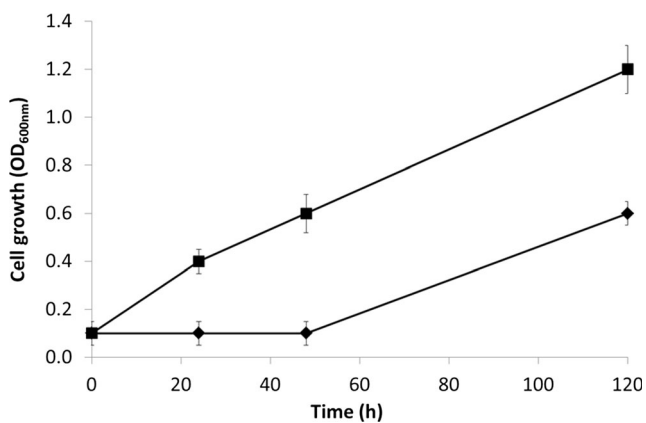
**Fig. 2** Characterization of Epi300 (Csd4). 4-MUC hydrolysis by Epi300 (Csd4) during growth in LB media supplemented with glucose (*white squares*), xylose (*black triangles*), cellobiose (*black squares*), CMC (*black circles*), avicel (*black triangles*), or with no supplement (*diamonds*). Initial OD<sub>620</sub> 0.05. FRU fluorescent relative units. Experiments were conducted in triplicate, and averages were plotted. *Error bars* indicate standard deviations

the enzymatic activity was maximal at 24 h. Interestingly, when glucose or xylose was added to the media, 4-MUC cleavage was abolished. In contrast, cellobiose, avicel, or CMC did not affect 4-MUCase activity. CMC is a highly specific substrate for endoacting cellulases (Dashtban et al. 2010), while 4-MUC is usually employed as exocellulase substrate (van Tilbeurgh et al. 1988). This allowed us to infer that ORF Csd4\_23 likely encode a by-functional endocellulase exocellulase.

### Enhancement of growth rate by exogenously produced BglC

At this point, we wanted to continue the study of Csd4 in an ethanogenic host; therefore, it was transformed into the ethanogenic strain *E. coli* MS04 (Fernández-Sandoval et al. 2012) resulting in strain *E. coli* MS04 (Csd4). All further experiments were performed on this host.

*E. coli* cannot readily utilize the cellobiose that is released by the action of endoglucanases and cellobiohydrolases. Thus, a  $\beta$ -glucosidase (EC 3.2.1.21) is necessary to split the cellobiose into two units of glucose that can be used by the cells. Since Csd4 does not carry a putative  $\beta$ -glucosidase, we wanted to determine if this enzyme could enhance the growth of MS04 (Csd4) in CMC as the sole carbon source. The  $\beta$ -glucosidase BglC from *T. fusca* was heterologously produced in *E. coli* and added exogenously to the culture media. The data showed that the addition of BglC had a positive effect on growth rate in minimum media with CMC as the sole carbon source (Fig. 3). No lag phase was detected when BglC was added and OD<sub>600</sub> of 0.6 was reached by the strain at 48 h. Conversely, when no BglC was added, a lag phase of 2 days was presented and the OD<sub>600</sub> of 0.6 was not reached until 120 h. Thus,  $\beta$ -glucosidase enhances the growth of MS04 (Csd4) on CMC.



**Fig. 3** MS04 (Csd4) grown in minimal media containing CMC with (squares) or without (diamonds) BglC at a concentration of 10 U/g of substrate. Experiments were conducted in triplicate, and averages were plotted. Error bars indicate standard deviations

### Ethanol production by MS04 (Csd4)

To evaluate the potential of *E. coli* MS04 (Csd4) for lignocellulosic ethanol production, we conducted fermentation assays of glucose in the presence and in the absence of CMC. When MS04 was carrying a control fosmid, glucose was depleted after 24 h of incubation (Fig. 4) and regardless of CMC being present as a supplemental carbon source, cell growth and ethanol production plateaued after glucose exhaustion. Maximal ethanol production was about 4 g/L. Furthermore, when Csd4 was present in MS04 cells, cell growth profile was not significantly altered. It also stopped after glucose exhaustion and reached similar levels than with the control fosmid. Again, no differences were seen on cell growth by the presence of CMC in the culture media. As expected, ethanol production from glucose stopped after all the glucose was used up and reached similar concentrations as the negative control strain. However, when the media was supplemented with CMC, ethanol concentration increased beyond the point when the glucose was fully consumed. It reached 5.3 g/L or 27 % more than the concentration obtained from glucose alone. Thus, Csd4 was able to enhance ethanol production by MS04 in the presence of glucose and CMC.

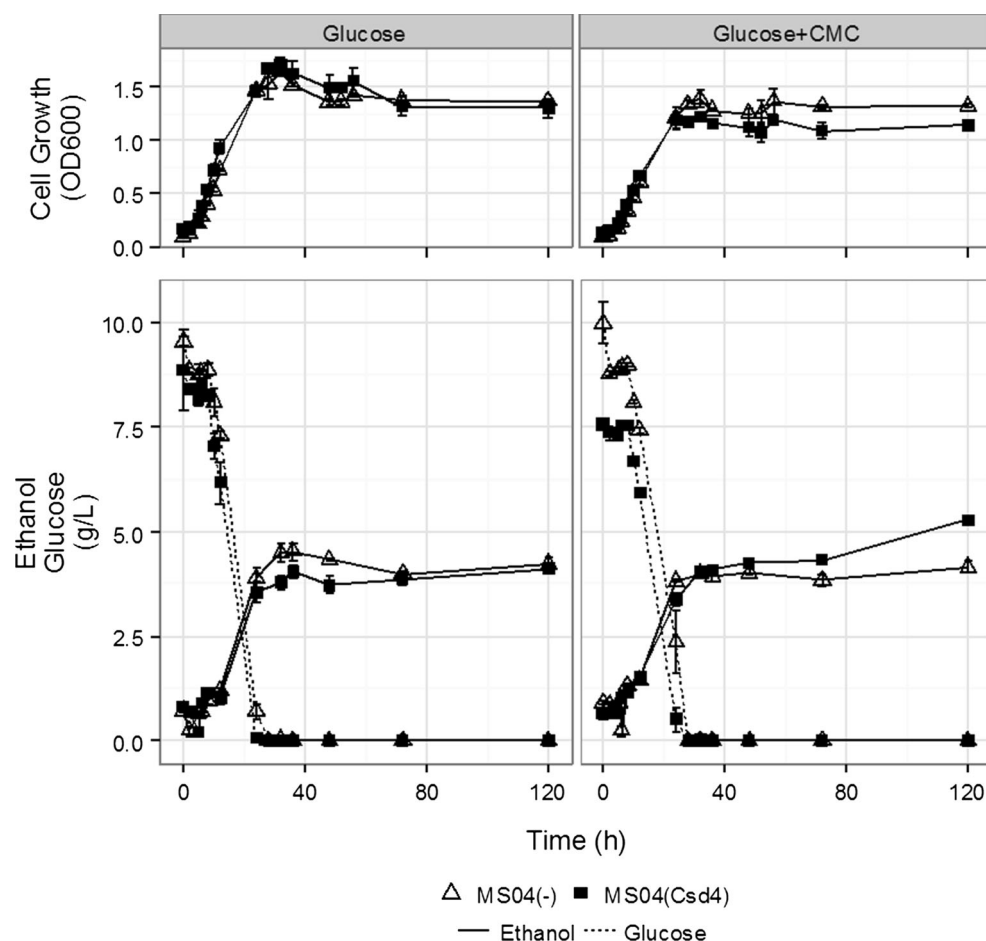
To assess if Csd4 was also able to sustain ethanol production by MS04 from CMC in the absence of glucose, fermentations were conducted in minimal media with 1 % CMC supplemented with exogenous BglC (Fig. 5). MS04, carrying a control fosmid, was not able to produce ethanol from CMC as the sole carbon source, while MS04 (Csd4) produced up to 3.7 g/L of ethanol by 48 h of incubation. These data confirmed that Csd4 supplemented with BglC was sufficient to confer the ability to ferment CMC into ethanol by MS04 in the absence of any other fermentable carbon source and without the addition of exogenous endocellulase or exocellulase. A theoretical maximum of 5.66 g of ethanol per liter could be produced from 10 g/L of CMC, meaning that MS04 (Csd4) supplemented with BglC reached a yield of 65.4 % of this maximum.

Finally, we also tested the ability of MS04 (Csd4) to convert other cellulosic substrates into ethanol. As shown in Table 2, in addition to CMC, MS04 (Csd4) was also able to convert avicel, filter paper, and sugarcane bagasse, but not birch wood xylan into ethanol. Neither BglC nor a xylosidase was added in these experiments, and ethanol levels were significantly lower than those obtained when BglC was present (Fig. 5).

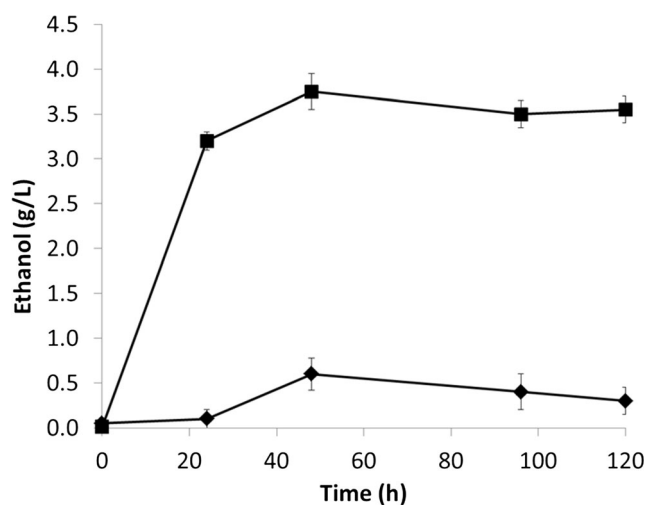
### Enhancement of sugarcane bagasse conversion by Csd4

*T. reesei* cellulases are the dominant enzymes used in industrial applications. They are mostly comprised of cellobiohydrolases that catalyze the release of glucose and cellobiose from the reducing and nonreducing ends of the

**Fig. 4** Cell growth, glucose consumption, and ethanol production by MS04 carrying a randomly picked control fosmid (white triangles) or Csd4 fosmid (black squares) in AM1 minimal media containing 1 % glucose (leftmost plots) or 1 % glucose+ 1 % CMC (rightmost plots), at 37 °C and 150 rpm without aeration. BglC was exogenously added to the culture media at a concentration of 10 U/g of substrate. Experiments were conducted in triplicate, and averages were plotted. Error bars indicate standard deviations



cellulose chain (Lynd et al. 2002). Endoglucanases, such as the one that is putatively encoded by the ORF Csd4\_23 in fosmid Csd4, complement the action of cellobiohydrolases



**Fig. 5** Ethanol production on mini-fermenters in minimal media containing CMC by MS04 carrying a randomly picked control fosmid (diamonds) or by MS04 (Csd4) (squares). BglC was exogenously added to the culture media at a concentration of 10 U/g of substrate. Initial OD<sub>620</sub> 0.1. Experiments were conducted in triplicate, and averages were plotted. Error bars indicate standard deviations

by decreasing the degree of polymerization of the cellulose. To evaluate if Csd4 could complement the action of *T. reesei* cellulases on an industrial feedstock, we mimicked a SHF process in which pretreated sugarcane bagasse was hydrolyzed by *T. reesei* cellulases for 24 h and then fermented by

**Table 2** Ethanol production in mini-fermenters by MS04 (Csd4) from different cellulosic substrates at 1 % (w/v) as the sole carbon source

Substrate	Ethanol (g/L) <sup>a</sup>
Glucose	1.87±0.48
CMC	0.81±0.01
Avicel	1.10±0.20
Filter paper	0.89±0.12
Sugarcane bagasse	0.84±0.14
Birchwood xylan	0.31±0.01
None <sup>b</sup>	0.25±0.08

<sup>a</sup> Ethanol was determined after 48 h of incubation at 37 °C and 150 rpm. pH was adjusted to 7.0 before autoclaving and remained uncontrolled during the fermentation. BglC was not added. Experiments were done in triplicate. Averages and standard deviations are shown

<sup>b</sup> Minimal media without the addition of any carbon source

MS04 carrying either Csd4 or a control fosmid. BglC was added to all the treatments. As shown in Table 3, in the absence of Csd4, the action of *T. reesei* cellulases produced a modest improvement on ethanol production. However, when Csd4 was present, the ethanol yield was over threefold higher. This indicates that Csd4 improves the conversion of hydrolyzed biomass into ethanol by MS04 under a SHF-like setup.

## Discussion

The microbiota inhabiting the bovine rumen has been extensively studied. High-throughput sequencing information from metagenomic libraries both from ruminal fluid and from fiber-attached microorganisms has been published (Brulc et al. 2009; Hess et al. 2011; Ross et al. 2012). Besides, many new hydrolytic enzymes active on lignocellulose have been identified by metagenomic approaches and further characterized (Chang et al. 2011; Cheema et al. 2012; Dai et al. 2012; Ferrer et al. 2012b; Gong et al. 2012, 2013; Gruninger et al. 2014; Ko et al. 2013; Nguyen et al. 2012; Rashamuse et al. 2013; Wang et al. 2013; Zhao et al. 2010; Zhou et al. 2012). However, despite all these efforts, with over  $10^{10}$  microorganisms per gram of content, the diversity of the rumen microbiota is still not fully characterized. In this study, using a functional metagenomic approach, we were able to identify 45 positive clones in CMC plates out of 30,000 clones that carried about 1.2 Gbp of metagenomic DNA information. This represents about one potential glucanolytic gene every 26,666 kbp. These odds are similar to the ones previously reported (Hess et al. 2011) where about one carbohydrate-active gene was identified every 16,940 kbp of sequenced metagenomic DNA.

According to our analysis, the parental organism from which Csd4 derives is a *Prevotella* closely related to *Prevotella ruminicola*. At the nucleotide level, Csd4 has only partial sequence identity to the sequenced *Prevotella* genomes and to the rumen microbiome, suggesting that the parental

organism has not been yet fully sequenced. This is not unexpected considering that *Prevotellaceae* family is the most abundant and diverse in the rumen microbiota (Creevey et al. 2014; Kim et al. 2011). Although lacking global synteny, regions of Csd4 are conserved in other members of the family (Fig. 1). This observation underlines the high level of genome plasticity and transfer events among *Prevotellas* (Purushe et al. 2010).

Three of the 28 ORFs predicted in Csd4 sequence encode proteins related to complex carbohydrate degradation: a xylosidase (belongs to the GH43), an endoglucanase (from GH5) and a polyphenol oxidase (Table 1). The importance of enzymes belonging to the families GH5 and GH43 in lignocellulosic biomass degradation in the rumen has been reported (Bao et al. 2011; Ferrer et al. 2012a; Ko et al. 2013). These ORFs can explain the ability of *E. coli* MS04 to ferment CMC, avicel, filter paper, or sugarcane bagasse after the introduction of Csd4 (Table 2). Although Csd4 was able to sustain fermentation from cellulosic substrates, continuous cell growth was not observed (Fig. 4). This might indicate that the energy requirements for cell survival are barely covered in these conditions. However, there is plenty of room for improvement. For example, in Csd4, all enzymes are under the control of native promoters, the fosmid is a single-copy extrachromosomal element, and it is expressing high levels of resistance markers. Also,  $\beta$ -glucosidase addition had a positive effect in strain growth in minimum media with CMC (Fig. 3) but no  $\beta$ -glucosidase gene is present in Csd4. Elimination of superfluous ORFs from Csd4, better tuned regulation of gene expression, addition of a  $\beta$ -glucosidase gene, and integration into the host genome are just a few of the strategies that can be undertaken.

Development of a cellulolytic *E. coli* has been considered for some time (Lambertz et al. 2014) but encountered difficulties mostly because of the low protein secretion ability of the host (French et al. 2013). However, the construction of an efficient cellulolytic system in *E. coli* is still a promising strategy for developing a CBP microorganism able to produce high-value compounds directly from lignocellulosic biomass. Even before reaching this goal, the expression of cellulolytic enzymes by fermentative organisms can improve the saccharification yield of commercial enzymes alone and help at reducing the time and cost of the industrial process. Fosmid Csd4 was able to promote the production of ethanol by *E. coli* MS04 both from glucose in conjunction with CMC (Fig. 4), and from CMC alone (Fig. 5) with a yield of 65 % of the theoretical maximum. This value is interesting taking into account what was expressed before about Csd4 ORFs expression. But, even more interesting was its ability to potentiate the action of *T. reesei* cellulases. Ethanol production from a lignocellulosic feedstock was improved threefold by Csd4 after saccharification by commercial *T. reesei* cellulases (Table 3).

**Table 3** Ethanol production in mini-fermenters from pretreated sugarcane bagasse in the presence of BglC

Strain	Ethanol (g/L)	
	No saccharification	<i>T. reesei</i> cellulase saccharification
MS04 (–)	0.78±0.12	0.96±0.35
MS04 (Csd4)	1.00±0.14	3.11±0.28

Saccharification was conducted with 40 U of *T. reesei* cellulases per gram of bagasse for 24 h at 45 °C and pH of 5.0. Ethanol was determined after 48 h of incubation with MS04 strains at 37 °C, pH 7.0, and 150 rpm. BglC (10 U/g) was added in all treatments. Experiments were done in triplicate. Averages and standard deviations are shown

While CBP might be considered the ultimate goal, any improvement in the production of cellulolytic enzymes by ethanologenic *E. coli* can be considered a success, as this would reduce the required loading levels of exogenously produced cellulases for simultaneous saccharification and fermentation or co-fermentation process setups.

**Acknowledgments** We are very grateful to Nikolai Guchin (ANCAP, Uruguay) for providing valuable materials and helpful advice. Special thanks to Betiana Bouzas (Frigorífico Pando, Uruguay) for providing access to bovine rumen samples. This work was supported by grants #FSE\_2009\_23 and #FSE\_2011\_6383 from Agencia Nacional de Investigación e Innovación (ANII, Uruguay) and grant #FPTA-256 from Instituto Nacional de Investigación Agropecuaria (INIA, Uruguay). IL and FN were also supported by the Programa de Desarrollo de las Ciencias Básicas (PEDECIBA, Uruguay), by Sistema Nacional de Investigadores (SNI, Uruguay), and by the bilateral international cooperation (Fondo Conjunto de Cooperación Uruguay-México) of Secretaría de Relaciones Exteriores, México, and Agencia Mexicana de Cooperación Internacional para el Desarrollo (SRE-AMEXCID, México).

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Aagaard K, Petrosino J, Keitel W, Watson M, Katancik J, Garcia N, Patel S, Cutting M, Madden T, Hamilton H, Harris E, Gevers D, Simone G, McInnes P, Versalovic J (2013) The human microbiome project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB J* 27(3):1012–1022. doi:10.1096/fj.12-220806
- Aakvik T, Lale R, Liles M, Valla S (2011) Metagenomic libraries for functional screening. In: *Handbook of molecular microbial ecology* I. John Wiley & Sons, Inc., pp 171–181
- Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* 12:186
- Bao L, Huang Q, Chang L, Zhou J, Lu H (2011) Screening and characterization of a cellulase with endocellulase and exocellulase activity from yak rumen metagenome. *J Mol Catalysis B: Enzymatic* 73(1–4):104–110. doi:10.1016/j.molcatb.2011.08.006
- Barker I, Petersen L, Reilly J (2010) Mechanism of xylobiose hydrolysis by GH43 b-xylosidase. *J Phys Chem* 114:15389–15393
- Beall DS, Ohta K, Ingram LO (1991) Parametric studies of ethanol production from xylose and other sugars by recombinant *Escherichia coli*. *Biotechnol Bioeng* 38(3):296–303. doi:10.1002/bit.260380311
- Beloqui A, Pita M, Polaina J, Martínez-Arias A, Golyshina OV, Zumarraga M, Yakimov MM, García-Arellano H, Alcalde M, Fernandez VM, Elborough K, Andreu JM, Ballesteros A, Plou FJ, Timmis KN, Ferrer M, Golyshin PN (2006) Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: biochemical properties, structural analysis, and phylogenetic relationships. *J Biol Chem* 281(32):22933–22942. doi:10.1074/jbc.M600577200
- Bhat A, Riyaz-Ul-Hassan S, Ahmad N, Srivastava N, Johri S (2013) Isolation of cold-active, acidic endocellulase from Ladakh soil by functional metagenomics. *Extremophiles* 17(2):229–239
- Bruce JM, Antonopoulos DA, Miller ME, Wilson MK, Yannarell AC, Dinsdale EA, Edwards RE, Frank ED, Emerson JB, Wacklin P, Coutinho PM, Henrissat B, Nelson KE, White BA (2009) Genocentric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci U S A* 106(6):1948–1953. doi:10.1073/pnas.08061911105
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The carbohydrate-active enzymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res* 37(Database issue):D233–D238. doi:10.1093/nar/gkn663
- Chang L, Ding M, Bao L, Chen Y, Zhou J, Lu H (2011) Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. *Appl Microbiol Biotechnol* 90(6):1933–1942. doi:10.1007/s00253-011-3182-x
- Cheema TA, Jirajaroenrat K, Sirinarumit T, Rakshit SK (2012) Isolation of a gene encoding a cellulolytic enzyme from swamp buffalo rumen metagenomes and its cloning and expression in *Escherichia coli*. *Anim Biotechnol* 23(4):261–277. doi:10.1080/10495398.2012.722156
- Creevey CJ, Kelly WJ, Henderson G, Leahy SC (2014) Determining the culturability of the rumen bacterial microbiome. *Microb Biotechnol* 7(5):467–479. doi:10.1111/1751-7915.12141
- Dai X, Zhu Y, Luo Y, Song L, Liu D, Liu L, Chen F, Wang M, Li J, Zeng X, Dong Z, Hu S, Li L, Xu J, Huang L, Dong X (2012) Metagenomic insights into the fibrolytic microbiome in yak rumen. *PLoS ONE* 7(7), e40430. doi:10.1371/journal.pone.0040430
- Dashtban M, Maki M, Leung KT, Mao C, Qin W (2010) Cellulase activities in biomass conversion: Measurement methods and comparison. *Crit Rev Biotechnol* 30(4):302–309
- Duan CJ, Feng JX (2010) Mining metagenomes for novel cellulase genes. *Biotechnol Lett* 32(12):1765–1775. doi:10.1007/s10529-010-0356-z
- Duan CJ, Xian L, Zhao GC, Feng Y, Pang H, Bai XL, Tang JL, Ma QS, Feng JX (2009) Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J Appl Microbiol* 107(1):245–256. doi:10.1111/j.1365-2672.2009.04202.x
- Edwards MC, Henriksen ED, Yomano LP, Gardner BC, Sharma LN, Ingram LO, Doran Peterson J (2011) Addition of genes for cellobiase and pectinolytic activity in *Escherichia coli* for fuel ethanol production from pectin-rich lignocellulosic biomass. *Appl Environ Microbiol* 77(15):5184–5191. doi:10.1128/aem.05700-11
- Fernández-Sandoval MT, Huerta-Beristain G, Trujillo-Martínez B, Bustos P, González V, Bolívar F, Gosset G, Martínez A (2012) Laboratory metabolic evolution improves acetate tolerance and growth on acetate of ethanologenic *Escherichia coli* under non-aerated conditions in glucose-mineral medium. *Appl Microbiol Biotechnol* 96(5):1291–1300. doi:10.1007/s00253-012-4177-y
- Ferrer M, Ghazi A, Beloqui A, Vieites JM, López-Cortés N, Marín-Navarro J, Nechitaylo TY, Guazzaroni M-E, Polaina J, Waliczek A (2012a) Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. *PLoS ONE* 7(6), e38134
- Ferrer M, Ghazi A, Beloqui A, Vieites JM, López-Cortés N, Marín-Navarro J, Nechitaylo TY, Guazzaroni ME, Polaina J, Waliczek A, Chernikova TN, Reva ON, Golyshina OV, Golyshin PN (2012b) Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. *PLoS ONE* 7(6), e38134. doi:10.1371/journal.pone.0038134
- French CE, Barnard DK, Fletcher E, Kane SD, Lakhundi SS, Liu C-K, Elfick A (2013) Synthetic biology for biomass conversion. In: Suib SL (ed) *New and future developments in catalysis*. Elsevier, Amsterdam, pp 115–140
- Gao D, Uppugundla N, Chundawat S, Yu X, Hermanson S, Gowda K, Brumm P, Mead D, Balan V, Dale B (2011) Hemicellulases and

- auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnol Biofuels* 4(1):5
- Gong X, Gruninger RJ, Qi M, Paterson L, Forster RJ, Teather RM, McAllister TA (2012) Cloning and identification of novel hydrolase genes from a dairy cow rumen metagenomic library and characterization of a cellulase gene. *BMC Res Notes* 5:566. doi:10.1186/1756-0500-5-566
- Gong X, Gruninger RJ, Forster RJ, Teather RM, McAllister TA (2013) Biochemical analysis of a highly specific, ph stable xylanase gene identified from a bovine rumen-derived metagenomic library. *Appl Microbiol Biotechnol* 97(6):2423–2431. doi:10.1007/s00253-012-4088-y
- Gruninger RJ, Gong X, Forster RJ, McAllister TA (2014) Biochemical and kinetic characterization of the multifunctional  $\beta$ -glucosidase/ $\beta$ -xylosidase/ $\alpha$ -arabinosidase, Bgxa1. *Appl Microbiol Biotechnol* 98(7):3003–3012. doi:10.1007/s00253-013-5191-4
- Gupta V, Natarajan C, Kumar K, Prasanna R (2011) Identification and characterization of endoglucanases for fungicidal activity in *Anabaena laxa* (cyanobacteria). *J Appl Phycol* 23(1):73–81. doi:10.1007/s10811-010-9539-1
- Hess M, Sczyrba A, Egan R, Kim TW, Chokhwalala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331(6016):463–467. doi:10.1126/science.1200387
- Horn S, Vaaje-Kolstad G, Westereng B, Eijsink VG (2012) Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels* 5(1):45
- Ingram LO, Aldrich HC, Borges ACC, Causey TB, Martínez A, Morales F, Saleh A, Underwood SA, Yomano LP, York SW, Zaldivar J, Zhou S (1999) Enteric bacterial catalysts for fuel ethanol production. *Biotechnol Prog* 15(5):855–866. doi:10.1021/bp9901062
- Kim M, Morrison M, Yu Z (2011) Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol Ecol* 76(1):49–63. doi:10.1111/j.1574-6941.2010.01029.x
- Klein-Marcuschamer D, Oleskowicz-Popiel P, Simmons BA, Blanch HW (2012) The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol Bioeng* 109(4):1083–1087. doi:10.1002/bit.24370
- Ko KC, Lee JH, Han Y, Choi JH, Song JJ (2013) A novel multifunctional cellulolytic enzyme screened from metagenomic resources representing ruminal bacteria. *Biochem Biophys Res Commun* 441(3):567–572. doi:10.1016/j.bbrc.2013.10.120
- Kojima M, Akahoshi T, Okamoto K, Yanase H (2012) Expression and surface display of *Cellulomonas* endoglucanase in the ethanologenic bacterium *Zymobacter palmae*. *Appl Microbiol Biotechnol* 96(4):1093–1104. doi:10.1007/s00253-012-4424-2
- Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, Fischer R, Commandeur U (2014) Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnol Biofuels* 7(1):135
- Leis B, Angelov A, Liebl W (2013) Screening and expression of genes from metagenomes. *Adv Appl Microbiol* 83:1–68. doi:10.1016/B978-0-12-407678-5.00001-5
- Linger JG, Adney WS, Darzins A (2010) Heterologous expression and extracellular secretion of cellulolytic enzymes by *Zymomonas mobilis*. *Appl Environ Microbiol* 76(19):6360–6369. doi:10.1128/AEM.00230-10
- Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66(3):506–577. doi:10.1128/mmbr.66.3.506-577.2002
- Lynd LR, Zyl WHV, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: An update. *Curr Opin Biotechnol* 16(5):577–583. doi:10.1016/j.copbio.2005.08.009
- Margeot A, Hahn-Hägerdal B, Edlund M, Slade R, Monot F (2009) New improvements for lignocellulosic ethanol. *Curr Opin Biotechnol* 20(3):372–380. doi:10.1016/j.copbio.2009.05.009
- Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 40(D1):D115–D122. doi:10.1093/nar/gkr1044
- Martin M, Biver S, Steels S, Barbeyron T, Jam M, Portetelle D, Michel G, Vandenberg M (2014) Identification and characterization of a halotolerant, cold-active marine endo- $\beta$ -1,4-glucanase by using functional metagenomics of seaweed-associated microbiota. *Appl Environ Microbiol* 80(16):4958–4967
- Martínez A, Grabar TB, Shanmugam KT, Yomano LP, York SW, Ingram LO (2007) Low salt medium for lactate and ethanol production by recombinant *Escherichia coli* B. *Biotechnol Lett* 29(3):397–404. doi:10.1007/s10529-006-9252-y
- Merino S, Cherry J (2007) Progress and challenges in enzyme development for biomass utilization. In: Olsson L (ed) *Biofuels. Advances in biochemical engineering/biotechnology*, vol 108. Springer, Berlin, pp 95–120
- Mingardon F, Chanal A, Tardif C, Fierobe H-P (2011) The issue of secretion in heterologous expression of *Clostridium cellulolyticum* cellulase-encoding genes in *Clostridium acetobutylicum* ATCC 824. *Appl Environ Microbiol* 77(9):2831–2838. doi:10.1128/aem.03012-10
- Muñoz-Gutiérrez I, Martínez A (2013) Polysaccharide hydrolysis with engineered *Escherichia coli* for the production of biocommodities. *J Ind Microbiol Biotechnol* 40(5):401–410. doi:10.1007/s10295-013-1245-y
- Muñoz-Gutiérrez I, Oropeza R, Gosset G, Martínez A (2012) Cell surface display of a  $\beta$ -glucosidase employing the type V secretion system on ethanologenic *Escherichia coli* for the fermentation of cellobiose to ethanol. *J Ind Microbiol Biotechnol* 39(8):1141–1152. doi:10.1007/s10295-012-1122-0
- Muñoz-Gutiérrez I, Moss-Acosta C, Trujillo-Martínez B, Gosset G, Martínez A (2014) Ag43-mediated display of a thermostable  $\beta$ -glucosidase in *Escherichia coli* and its use for simultaneous saccharification and fermentation at high temperatures. *Microb Cell Fact* 13(1):106
- Nguyen NH, Maruset L, Uengwetwanit T, Mhuantong W, Harnpicharnchai P, Champreda V, Tanapongpipat S, Jirajaroenrat K, Rakshit SK, Eurwilaichitr L, Pongpattanakitsote S (2012) Identification and characterization of a cellulase-encoding gene from the buffalo rumen metagenomic library. *Biosci Biotechnol Biochem* 76(6):1075–1084. doi:10.1271/bbb.110786
- Ni J, Tokuda G (2013) Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol Adv* 31(6):838–850. doi:10.1016/j.biotechadv.2013.04.005
- Noguchi H, Park J, Takagi T (2006) Metagenome: Prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* 34(19):5623–5630. doi:10.1093/nar/gkl723
- Olofsson K, Bertilsson M, Lidén G (2008) A short review on ssf - an interesting process option for ethanol production from lignocellulosic feedstocks. *Biotechnol Biofuels* 1(1):7
- Orencio-Trejo M, Utrilla J, Fernández-Sandoval MT, Huerta-Beristain G, Gosset G, Martínez A (2010) Engineering the *Escherichia coli* fermentative metabolism. In: Wittmann C, Krull R (eds) *Biosystems engineering II. Advances in biochemical engineering / biotechnology*, vol 121. Springer, Berlin, pp 71–107
- Purushe J, Fouts D, Morrison M, White B, Mackie R, Coutinho P, Henrissat B, Nelson K (2010) Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: Insights into their environmental niche. *Microb Ecol* 60(4):721–729. doi:10.1007/s00248-010-9692-8
- Rashamuse KJ, Visser DF, Hennessy F, Kemp J, Roux-van der Merwe MP, Badenhorst J, Ronneburg T, Francis-Pope R, Brady D (2013) Characterisation of two bifunctional cellulase-xylanase enzymes

- isolated from a bovine rumen metagenome library. *Curr Microbiol* 66(2):145–151. doi:10.1007/s00284-012-0251-z
- Rodríguez MC, Loaces I, Amarelle V, Senatore D, Iriarte A, Fabiano E, Noya F (2015) Est10: A novel alkaline esterase isolated from bovine rumen belonging to the new family XV of lipolytic enzymes. *PLoS ONE* 10(5), e0126651. doi:10.1371/journal.pone.0126651
- Ross E, Moate P, Bath C, Davidson S, Sawbridge T, Guthridge K, Cocks B, Hayes B (2012) High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* 13(1):53
- Ryu S, Karim M (2011) A whole cell biocatalyst for cellulosic ethanol production from dilute acid-pretreated corn stover hydrolyzates. *Appl Microbiol Biotechnol* 91(3):529–542. doi:10.1007/s00253-011-3261-z
- Sainz M (2009) Commercial cellulosic ethanol: The role of plant-expressed enzymes. *Vitro Cell Dev Biol Plant* 45(3):314–329. doi:10.1007/s11627-009-9210-1
- Shin H-D, McClendon S, Vo T, Chen RR (2010) *Escherichia coli* binary culture engineered for direct fermentation of hemicellulose to a bio-fuel. *Appl Environ Microbiol* 76(24):8150–8159. doi:10.1128/aem.00908-10
- Spiridonov NA, Wilson DB (2001) Cloning and biochemical characterization of Bglc, a  $\beta$ -glucosidase from the cellulolytic actinomycete *Thermobifida fusca*. *Curr Microbiol* 42(4):295–301. doi:10.1007/s002840110220
- Tao L, Schell D, Davis R, Tan E, Elander R, Bratis A (2014) NREL 2012 achievement of ethanol cost targets: Biochemical ethanol fermentation via dilute-acid pretreatment and enzymatic hydrolysis of corn stover. In: Technical Report NREL/TP-5100-61563, NREL, U.S. Department of Energy, Golden, Colorado
- Teather RM, Wood PJ (1982) Use of Congo Red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol* 43(4):777–780
- van Tilbeurgh H, Loontjens FG, De Bruyne CK, Claeysens M (1988) Fluorogenic and chromogenic glycosides as substrates and ligands of carbohydrases. *Methods Enzymol* 160:45–59. doi:10.1016/0076-6879(88)60106-6
- Voget S, Steele HL, Streit WR (2006) Characterization of a metagenome-derived halotolerant cellulase. *J Biotechnol* 126(1):26–36. doi:10.1016/j.jbiotec.2006.02.011
- Wang L, Hatem A, Catalyurek UV, Morrison M, Yu Z (2013) Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS ONE* 8(11), e78507. doi:10.1371/journal.pone.0078507
- White BA, Lamed R, Bayer EA, Flint HJ (2014) Biomass utilization by gut microbiomes. *Annu Rev Microbiol* 68(1):279–296. doi:10.1146/annurev-micro-092412-155618
- Wu S, Zhu Z, Fu L, Niu B, Li W (2011) WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12(1):444
- Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7(1-2):203–214. doi:10.1089/10665270050081478
- Zhao S, Wang J, Bu D, Liu K, Zhu Y, Dong Z, Yu Z (2010) Novel glycoside hydrolases identified by screening a chinese holstein dairy cow rumen-derived metagenome library. *Appl Environ Microbiol* 76(19):6701–6705. doi:10.1128/aem.00361-10
- Zheng Z, Chen T, Zhao M, Wang Z, Zhao X (2012) Engineering *Escherichia coli* for succinate production from hemicellulose via consolidated bioprocessing. *Microb Cell Fact* 11(1):37
- Zhou J, Bao L, Chang L, Liu Z, You C, Lu H (2012)  $\beta$ -xylosidase activity of a GH3 glucosidase/xylosidase from yak rumen metagenome promotes the enzymatic degradation of hemicellulosic xylans. *Lett Appl Microbiol* 54(2):79–87. doi:10.1111/j.1472-765X.2011.03175.x

**Capítulo 4.**  
**CARACTERIZACIÓN DE UNA CELULOSA**  
**NOVEDOSA CON POTENCIAL PARA SER**  
**UTILIZADA EN LA PRODUCCIÓN INDUSTRIAL**  
**DE BIOETANOL**

---



Loaces, I., Bottini, G., Moyna, G., Fabiano, E., Martínez, A. & Noya, F. (2015) **Ethanol production from *E. coli* MS04 expressing a multifunctional halotolerant glucanase from cow rumen**. *Enzyme and Microbial Technology*

## Resumen

La celulasa codificada por el ORF Csd4\_23 en el fósido Csd4, fue amplificada, purificada y caracterizada bioquímicamente. Esta enzima, denominada EndoG, mostró su máxima actividad a pH 5.0 y 50°C. Resultó ser termo-estable y tolerante a sales, cationes divalentes, detergentes, furfural, líquidos iónicos y etanol. EndoG no fue inhibida por los productos finales de la degradación de celulosa: glucosa o celobiosa. Ensayos de especificidad de sustrato mostraron que EndoG posee actividad sobre 4-metilumbeliferil- $\beta$ -D-cellobioside (4-MUC), *p*-nitrofenil- $\beta$ -D-cellobiosido (pNPC), carboximetil-celulosa (CMC), celulosa tratada con ácido fosfórico (PASC), avicel, lichenano y bagazo de caña de azúcar. La expresión de EndoG en la cepa MS04 le confiere la capacidad de producir etanol a partir de CMC.

## Objetivos

- ☞ Caracterizar las propiedades bioquímicas de la celulasa codificada en el fósido Csd4, denominada EndoG, obtenida del metagenoma de rumen.
- ☞ Determinar el potencial de EndoG para ser utilizada en condiciones industriales.
- ☞ Expresar EndoG en una cepa etanológica y evaluar la producción de etanol a partir de CMC.

## Principales resultados y conclusiones

Los resultados obtenidos en esta etapa confirmaron la hipótesis de que el rumen bovino puede ser una fuente de enzimas novedosas con potencial para ser utilizadas a nivel industrial. EndoG, una enzima seleccionada a partir del metagenoma de rumen construido previamente, fue asignada a la familia de glicosil-hidrolasas 5 (GH5) y de sus homólogos más cercanos solo uno fue caracterizado. EndoG resultó ser activa en un rango considerable de pH y temperatura, permanece activa luego de ser incubada 1 hora a 55°C y mantiene 81% de su actividad máxima cuando se agrega 4 M de NaCl. A su

vez, no se detectó inhibición por glucosa o celobiosa, cuando se adicionan a una concentración final de 10 mM. La enzima tolera hasta 2 M de acetato, 50 mM de furfural y cantidades residuales de líquidos iónicos, todos ellos inhibidores generados usualmente en los pre-tratamientos químicos de la biomasa lignocelulósica. Esta celulasa no es afectada por el agregado de  $\text{Fe}^{+2}$ ,  $\text{Ca}^{+2}$ ,  $\text{Cd}^{+2}$ ,  $\text{K}^+$  or  $\text{Ni}^{+2}$ , pero es estimulada en presencia de  $\text{Mn}^{+2}$  o  $\text{Mg}^{+2}$ . A su vez, butanol, Triton X100, PMSF, EDTA o Tween 20 reducen ligeramente la actividad enzimática, mientras que tolueno, glicerol, metanol o etanol no tienen efecto.

La enzima recombinante fue clonada en el vector inducible pTrc99A e incorporada a la cepa etanologénica MS04. Ensayos en mini-fermentadores mostraron que, en co-cultivo con la cepa MS04/pTrc99A-BglC, se obtienen 3,35 g/L de etanol a partir de 7,5 g/L de CMC en 12 horas.

Estos resultados nos permiten concluir que EndoG es una celulasa multifuncional con potencial para ser utilizada en condiciones industriales de producción de etanol, debido, principalmente, a la diversidad de condiciones en las que permanece activa.

## **Contribución personal al trabajo**

En el trabajo presentado en este capítulo realicé los análisis bioinformáticos. Me encargué de la expresión heteróloga de EndoG y de los ensayos de caracterización funcional. El grupo del Dr. Moyna nos facilitó los líquidos iónicos y colaboró con el análisis de los resultados obtenidos con estos materiales.

Los ensayos de fermentación los realicé en las instalaciones del Laboratorio de Ingeniería de Vías Metabólicas de Bacterias, Departamento de Ingeniería Celular y Biocatálisis, del IBT-UNAM, con la asesoría del Dr. Alfredo Martínez, gracias a una beca de movilidad de la AMEXID-SRE y una beca de Pasantías en el Exterior de PEDECIBA.

El trabajo presentado en este capítulo fue financiado parcialmente por un proyecto FSE de la ANII y PEDECIBA.

# Ethanol production from *E. coli* MS04 expressing a multifunctional halotolerant glucanase isolated from cow rumen

Inés Loaces<sup>1\*</sup>, Gualberto Bottini<sup>2</sup>, Guillermo Moyna<sup>2</sup>, Elena Fabiano<sup>1</sup>, Alfredo Martínez<sup>3</sup> and Francisco Noya<sup>1</sup>

<sup>1</sup>Departamento de Bioquímica y Genómica Microbianas, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay

<sup>2</sup>Departamento de Química del Litoral, CENUR Litoral Norte, Universidad de la República, Paysandú, Uruguay

<sup>3</sup>Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, A.P: 510-3. Cuernavaca, Morelos, 62270 México

\* Corresponding author

Mailing address: Departamento Bioquímica y Genómica Microbianas, Instituto de Investigaciones Biológicas Clemente Estable, Av. Italia 3318, Montevideo, Uruguay. CP: 11600. Phone: (598) 24871616. Fax: (598) 2487-5461. E-mail: [iloaces@iibce.edu.uy](mailto:iloaces@iibce.edu.uy).

## Abstract

Cellulosic materials represent a sustainable source of chemicals and fuels because of their low cost and ample supply. The current limitations on bioconversion of lignocellulosic biomass include poor enzyme stability and inhibition by secondary or final products. Here, we report the biochemical characterization of novel, rumen metagenome-derived cellulase, named EndoG. The deduced amino acid sequence was assigned to glycoside hydrolase family 5, and showed similarity to non-characterized proteins. The 1146pb ORF encoding EndoG was over-expressed in *Escherichia coli* and the corresponding protein purified. EndoG displayed a wide range of pH activity with a maximum at pH 5.0 and at least 65% activity at pH between 4.5 and 7.5. The enzyme was highly stable at 55°C for 1 hour, and retained 81% activity with 4 M NaCl. EndoG was also active in the presence of diverse divalent cations, detergents, EDTA, acetate, furfural, imidazolium ionic liquids and ethanol. Glucose or cellobiose had no effect in EndoG performance. EndoG behaved as a multifunctional glucanase displaying activity on 4-methylumbelliferyl- $\beta$ -D-cellobioside, *p*-nitrophenyl- $\beta$ -D-cellobioside, carboxymethylcellulose, phosphoric acid swollen cellulose, Avicel, xylan, lichenan and sugar cane bagasse. When expressed in the ethanologenic *E.coli* strain MS04, EndoG was able to convert avicel

and CMC into ethanol in a single step. This highly stable, salt tolerant cellulase is a suitable candidate for industrial applications.

## **Keywords**

Metagenome, Cellulolytic enzyme, Cellulose saccharification, Halotolerant cellulase.

## **Introduction**

Cellulosic materials are a sustainable feedstock for chemical and fuel production because of their relative low cost and plentiful supply. The main impediment to more widespread utilization of this resource is the absence of low-cost technology for overcoming the recalcitrance of cellulosic biomass [1]. To date, enzymatic hydrolysis is the most convenient and eco-friendly method for cellulose hydrolysis resulting in appreciable sugar yields [2] but the industry demands better cellulases with higher catalytic efficiency on insoluble cellulosic substrates, wider temperature range, increased stability, and higher tolerance to end-product inhibition [3].

Based on their catalytic action modes, there are three major types of cellulase activities:  $\beta$ -1,4-endoglucanase (EC 3.2.1.4), which cut internal sites on crystalline cellulose surfaces; exoglucanase, including cellobiohydrolase (EC 3.2.1.91) and cellodextrinase (EC 3.2.1.74), which remove cellobiose on the reducing or non-reducing ends of cellulose chains and  $\beta$ -glucosidase (EC 3.2.1.21), which hydrolyze soluble cellodextrins and cellobiose to glucose [4].

The enzymatic hydrolysis may take place in a separate step followed by fermentation, called separate hydrolysis and fermentation (SHF) process, or it may take place together with the fermentation in a simultaneous saccharification and fermentation of sugars (SSF) process. The ultimate objective would be a one-step consolidated bioprocess (CBP) of lignocellulose to bioethanol, in which all of these steps occur in a single reactor, where a single microorganism or microbial consortium converts pretreated biomass into products without added enzymes [5].

In this work, we cloned and expressed in *Escherichia coli* a metagenomic-derived bovine rumen cellulase. The recombinant enzyme was fully characterized. This novel cellulase was thermostable and tolerant to high concentrations of inorganic salts, imidazolium ionic liquids (ILs), inhibitors (such as furfural or acetate) and final products (glucose and cellobiose). We were also able to express it in an ethanologenic *E.coli* to convert cellulose into ethanol in a single step.

## Materials and methods

### Cloning, expression of *endoG* and protein purification

*endoG* was initially identified in one clone of a bovine rumen metagenomic library [6]. The clone containing *endoG* was sequenced and the predicted gene was amplified by polymerase chain reaction (PCR) with sense (5'-AGGCACCATGGCTATGAAAAAGATTCTTTATTTCTTTGC-3') and antisense (5'-ATATTAAGCTTTTTATCAAGCCCTTTGGCACCC-3') primers. NcoI and HindIII sites were introduced in the sense and antisense primers, respectively (shown in bold). PCR was performed using Pfu polymerase (Thermo-Scientific, USA). PCR fragments were cloned into pET28a (Novagen, USA) digested with NcoI and HindIII, creating a C-terminal 6xHis Tag fusion. Recombinant EndoG-6xHis, rEndoG, was expressed in *Escherichia coli* BL21 (DE3) pLysS (Novagen, USA) in 2xYT medium (16g/L tryptone, 10g/L yeast extract and 5g/L NaCl, pH 7.0) induced during 4 hours at 30°C in the presence of 0.1 mM isopropyl thio-β-D-galactoside (IPTG). Cells were lysed by sonication and rEndoG was purified using a Ni-NTA resin (Life Technologies, USA). Purity and integrity were checked by SDS-PAGE. Protein concentration was determined by the Bicinchoninic Acid Protein Assay Kit (Sigma Aldrich, USA) using bovine serum albumin as standard.

### Sequence and phylogenetic analyses

Signal peptide prediction was done using SignalP 4.1 server [7].

Closest homologs to EndoG were identified using NCBI BLASTP [8]. Multiple sequences alignments were computed with MEGA 6 [9] with the Muscle method. Phylogenetic trees were constructed using the Maximum likelihood method implemented in MEGA 6.

### Enzymatic assays

Substrates 4-methylumbelliferyl-β-D-cellobioside (4-MUC) and p-nitrophenyl-β-D-cellobioside (pNPC) (Sigma Aldrich, USA) were used for enzymatic characterization. Exoglucanase activity was determined as previously described [10] using 0.5 μg of rEndoG, and 2 mM of 4-MUC or pNPC as substrate. Reactions were carried out at 45°C for 15 min. The release of 4-methylumbelliferone was followed by fluorescence (excitation 365 nm, emission 445 nm). The pNPC quantification was made spectrophotometrically at 410 nm ( $\epsilon=1.6 \text{ mM}^{-1}\text{cm}^{-1}$ ).

### pH and temperature effects on rEndoG activity and enzyme thermostability

The effect of pH on the activity of rEndoG was evaluated by carrying out enzymatic reactions in a mix of 0.05 M acetic acid, 0.05 M MES and 0.1 M Tris to ensure constant ionic strength [11]. pH

was varied between 3.5 and 11.0 as needed. The effect of temperature on rEndoG activity was determined between 20 and 80°C in 50 mM sodium phosphate pH 5.5. For the thermostability assay, rEndoG was pre-incubated at 50, 60, 70 or 80°C for 10–60 min in the absence of substrates. The residual activity was determined at pH 5.5 and 45°C.

### **Effect of metals and potential inhibitors on EndoG activity**

The effect of metal ions on the activity of rEndoG were determined at pH 5.5 and 45°C with the following metal salts at final concentrations of 5 mM: MgCl<sub>2</sub>, CoCl<sub>2</sub>, CaCl<sub>2</sub>, NiCl<sub>2</sub>, KCl, CuSO<sub>4</sub>, MnCl<sub>2</sub>, AgNO<sub>3</sub>, ZnCl<sub>2</sub>, HgCl<sub>2</sub>, FeCl<sub>2</sub>, and Cd(OAc)<sub>2</sub>. Analogously, acetone, methanol, propanol, butanol, toluene, Tween 20, glycerol, SDS and Triton X100 were evaluated at a final concentration of 5% (v/v), EDTA at 50 mM, and phenylmethanesulfonyl fluoride (PMSF) at 5 mM.

To determine the effect of glucose (up to 1 M), cellobiose (up to 10 mM), acetate (up to 2 M), ethanol (up to 15%), furfural (up to 150 mM) or NaCl (up to 4M) on rEndoG activity the enzymatic reactions were carried out at pH 5.0 and 45°C using 2mM 4-MUC as substrate.

The effect of 0-15% (w/v) imidazolium ILs 1-*n*-butyl-3-methylimidazolium chloride ([C<sub>4</sub>mim]Cl), 1-*n*-butyl-3-methylimidazolium bromide ([C<sub>4</sub>mim]Br), 1-ethyl-3-methylimidazolium bromide ([C<sub>2</sub>mim]Br), and 1-ethyl-3-methylimidazolium acetate ([C<sub>2</sub>mim][OAc]) on the activity of rEndoG was evaluated at 45°C and pH 5.0 using pNPC as substrate.

When 4-MUC was used as a substrate, activity was reported as relative to the control reaction without any additive or treatment. In all cases, reactions were stopped by adding 100 µL of 1 M NaCO<sub>3</sub>. All the assays were performed in triplicate.

### **Determination of substrate specificity**

EndoG specific activity against 1% (w/v) Carboxymethyl-cellulose (CMC sodium salt, DP 400, DS 0.7, Sigma-Aldrich), lichenan (Megazyme, Ireland), xylan from birchwood (Sigma Aldrich), avicel (PH-101, Sigma Aldrich), pre-treated sugar bagasse or PASC (prepared according to [12]) was assayed at 45°C and pH 5.0. Reaction products were quantified by determining the concentration of reducing sugars by the method of the dinitrosalicylic acid (DNS) [13]. One unit (U) of glucanase activity was defined as the amount of enzyme that released 1 µmol of reducing sugars per min. Specific activity was defined as the units of enzyme per mg of protein.

### **Nucleotide sequence accession number**

The sequence of EndoG is available in the GenBank protein database under the accession number AKF17200.

## Conversion of cellulose into ethanol

*endoG* was cloned into pTrc99A [14] between NcoI and HindIII sites, and transformed into ethanologenic *E. coli* strain MS04 [6, 15]. The resulting strain was induced with 10  $\mu$ M IPTG and grown up to 3.5 OD<sub>600</sub>. Spontaneous lysis occurred at higher densities. To provide  $\beta$ -glucosidase activity, MS04 cells expressing BglC [16] were similarly induced and grown up to 5.5 OD<sub>600</sub>. Both strains were combined at a ratio 1:2 (EndoG:BglC) to obtain an initial 5.1 OD<sub>600</sub> in minimal salt medium AM1 and mixed with 7.5 g/L CMC or avicel. Fermentation was conducted at 45°C. Samples were taken every 2 h, cells and solids were removed by centrifugation and ethanol was analyzed by gas chromatography using n-butanol as the internal standard (6850 Series GC System, Agilent, DE, USA). All assays were performed in triplicates.

## Results

### Origin and phylogeny of EndoG

We have previously isolated a collection of cellulolytic enzymes from a cow rumen metagenomic fosmid library based on the formation of halos of hydrolysis around the host colonies on agar plates supplemented with CMC [6]. One of these genes, named *endoG*, encodes a protein of 382 amino acids. From sequence comparisons EndoG can be classified as a glycoside hydrolase family 5 (GH5) [17]. A putative 20 amino acid long signal peptide was predicted from sequence analysis [7], with a putative cleavage site between position 22 and 23 (VTA-CT D=0.590 D-cutoff=0.570). BLAST analysis showed that EndoG shares the highest homology at the amino acid sequence level with different cellodextrinases and cellobiosidases from uncultured microorganisms and with GH5 proteins from *Prevotella* (Table 1 and Fig. S1).

We also compared EndoG with 297 ORFs annotated as “cellulases” from the whole rumen metagenome (DPI\_rumen on MG\_RAST database, ID: 4491686.3) [18]. EndoG showed 75% of identity with one hit (ADE83642.1) belonging to a strain of *Prevotella*. In the same way, EndoG showed high homology with a putative ORF (Locus ID HiSeq\_02855500) from Joint Genome Institute (JGI) metagenome “Bovine rumen microbial communities from Illinois University” (Taxon Object ID 2061766007) that is located in scaffolds associated with *Prevotella ruminicola* (HiSeq\_NODE\_1861550) [19].

### Expression and purification of EndoG

Recombinant C-terminal 6xHis-tagged rEndoG was heterologously expressed in *E. coli* BL21 (DE3) pLysS. Both supernatant and cell lysate showed activity against 4-MUC. The purified rEndoG

migrated as a single band in SDS-PAGE with an apparent molecular weight of about 45 kDa, which was consistent with the theoretical value of 44.0 kDa (data not shown).

### **Enzymatic properties of rEndoG**

The effects of pH and temperature on rEndoG activity were evaluated against pNPC and 4-MUC substrates, respectively. Its optimal pH was 5.0, with around 60% of its activity at pH between 4.0 and 7.5. EndoG reaction rate was maximal at 50°C displaying over 70% of this activity between 30°C and 55°C at pH 5.0 (Fig. 1A and 1B).

To assay the thermostability of rEndoG, the enzyme was pre-incubated with no substrates at 50°C, 55°C or 60°C. Aliquots were taken every 10 min during 1 h and their activities determined under standard conditions. The analysis showed that rEndoG retained over 90% of its activity after 1 h at 50°C or after 50 min at 55°C. After 1 h at 60°C rEndoG conserved 70% of its maximal activity (Fig. 1C).

The effects of metal ions, chelating agents and chemicals on the activity of rEndoG were measured.  $Zn^{2+}$ ,  $Co^{2+}$  and  $Ag^{2+}$  negatively affected its activity, while  $Hg^{2+}$  and  $Cu^{2+}$  completely abolished it. On the other hand,  $Mn^{2+}$  and  $Mg^{2+}$  stimulated 4-MUC cleavage (Table 2).

### **EndoG tolerance to industrial inhibitors**

To evaluate the robustness of EndoG as an industrial cellulase, its activity was measured in the presence of diverse compounds formed in the conversion of lignocellulose into ethanol.

rEndoG activity was measured in the presence of glucose or cellobiose, two products of cellulose hydrolysis. None affected enzyme activity when added up to 1 M in standard assay conditions using 4-MUC as substrate (data not shown). Similarly, rEndoG activity was only moderately affected by acetate, commonly released from hemicellulose during pretreatment [15], and ethanol, produced during fermentation. 67% of the maximal activity was retained in the presence of 2 M of acetate (Fig. 2A) and 82% in 15% (v/v) of ethanol (Fig. 2B). Likewise, in the presence of furfural, an aldehyde commonly produced from xylose during acid pretreatment [20], enzyme activity was mostly unaffected at concentrations commonly reached under standard conditions (7-15 mM) [20, 21]. At higher concentrations (115 mM), a significant inhibition was seen but these furfural levels are only reached under harsh pretreatment [22] (Fig. 2C). rEndoG was also very tolerant up to 4 M NaCl (Fig. 2D).

The enzyme was severely inhibited by the presence of SDS. Butanol, Triton X100, PMSF, EDTA or Tween 20 reduced the enzyme activity only slightly. However, the presence of propanol, acetone, toluene, glycerol or methanol had almost no effect on enzyme performance (Fig. 3A).

Given the use of ILs in lignocellulose pretreatment schemes [23], we assessed the effect of imidazolium ILs on EndoG activity. The enzyme retained 89% of its activity in the presence of



[C<sub>4</sub>mim]Br in concentrations of up to 5% (w/v), and lost 75% of its activity at 15% of this IL. In contrast, EndoG was more sensitive to [C<sub>4</sub>mim]Cl or [C<sub>2</sub>mim]Br, retaining approximately 60% of its maximal activity in the presence of 2.5% (w/v) of either of these imidazolium salts, but only 18 or 35%, respectively, when the concentration of the ILs reached 15%. The most detrimental effect on the activity of the enzyme was observed with [C<sub>2</sub>mim][OAc]. This IL caused the activity of EndoG to drop to 44 and 9% at concentrations of 2.5 and 15% (w/v), respectively (Fig. 3B).

### **Substrate specificity**

Diverse lignocellulosic substrates were used to evaluate the specificity of EndoG: 1,4- $\beta$ -D-glucan (CMC, Avicel and PASC),  $\beta$ -1,3/4-glucan (lichenan), 1,4- $\beta$ -D-xylan, and pretreated sugar cane bagasse. The results are shown in Table 3. The enzyme activity was maximal against bagasse. On the other end, the lowest activity was obtained with PASC, a form of amorphous cellulose.

### **Consolidated bioprocessing**

In order to determine the potential of EndoG in a putative CBP process, the expression plasmid pTrc99A-EndoG was transformed into the ethanologenic strain MS04. The resulting strain was co-cultivated with MS04/pTrc99A-BglC, a  $\beta$ -glucosidase expressing strain, and evaluated for CMC and Avicel fermentation in the presence of IPTG. After incubation at 45°C, the maximum ethanol concentration obtained from CMC was 3.3 g/L and 2.0 g/L from Avicel, representing 48% of its theoretical maximum (Figure 6).

### **Discussion**

EndoG shows significant degrees of homology to metagenomic fragments previously found in other rumen microbiomes and to proteins assigned to the genus *Prevotella* (Table 1 and Figure S1). The presence of a signal peptide in its sequence is consistent with the fact that the parental clone displayed extracellular halo on CMC plates [6] and 4-MUCase activity was detected in culture supernatants of IPTG-induced BL21 (DE3) pLyS cells carrying pET28a-EndoG.

EndoG is a multifunctional enzyme since it showed exoglucanase activity on Avicel, pNPC and 4-MUC; endoglucanase activity on CMC; and xylanase activity on birch wood xylan (Table 3). In addition, it was also able to break down lichenan, PASC and sugar cane bagasse. One of EndoG closest homologs, CelEx-BR12 (74% identity with EndoG), was also classified as a multifunctional enzyme because of these activities [24].

EndoG was active and stable over a broad temperature range. It displayed over 80% of its maximal activity at 37°C, which makes feasible its utilization in SSF and CBP processes employing mesophilic microorganisms (Figure 1). Saccharolytic enzymes that are active at lower temperatures reduce costs associated to liquor heating and cooling, and losses due to ethanol evaporation [25]. Yield is also compromised due to end-product inhibition of cellulases [26]. However, EndoG was mostly unaffected by the presence of glucose, cellobiose, or ethanol (data not shown and Figure 2B). This enzyme also appeared tolerant to solvents used in organosolvation of lignin such as methanol, ethanol or acetone [27] (Figure 3A). It was also resistant to common pretreatment derivatives such as acetic acid and furfural (Figure 2A and Figure 2C), as well as to other byproducts, chelating agents, or surfactants that can accompany the pretreated biomass [26]. Only SDS, a potent protein denaturalizing agent, affected its performance (Figure 3A). However, EndoG might not be easily incorporated into processes involving alkaline pretreatment [21]. Although, the enzyme was tolerant to the presence of  $K^+$  or  $Ca^{2+}$  (Table 2), its activity was almost abolished at alkaline pH (Figure 1A).

EndoG was strongly inhibited by Hg (II), a metal that interacts with sulfur-containing amino acid (EndoG has 3 Cys and 6 Met) [28], and Cu (II), a potent redox agent. Similarly to others cellulases, EndoG activity was stimulated in the presence of Mg (II), Mn (II), Ca (II) and Co (II) [10, 24] (Table 2).

Interestingly, EndoG was stable and active in the presence of 4 M of NaCl (Figure 2D). Usually, halotolerant enzymes are derived from extreme environments [29], which is not the case of EndoG. It has been suggested that halotolerance in cellulases is correlated to their ILs tolerance [30]. These neoteric solvents, have been shown to dissolve both cellulose and lignin efficiently [31]. Consequently, ILs have been proposed as pretreatment agents and as potential reaction media for the enzymatic hydrolysis of cellulose [32]. Although the activity of EndoG was affected considerably at high concentrations of imidazolium salts, it was tolerant to lower concentrations similar to that observed for other cellulases [23, 32] (Figure 4B). Therefore, the enzyme could be employed in the presence of trace concentrations of ILs resulting from IL-pretreatment processes [23].

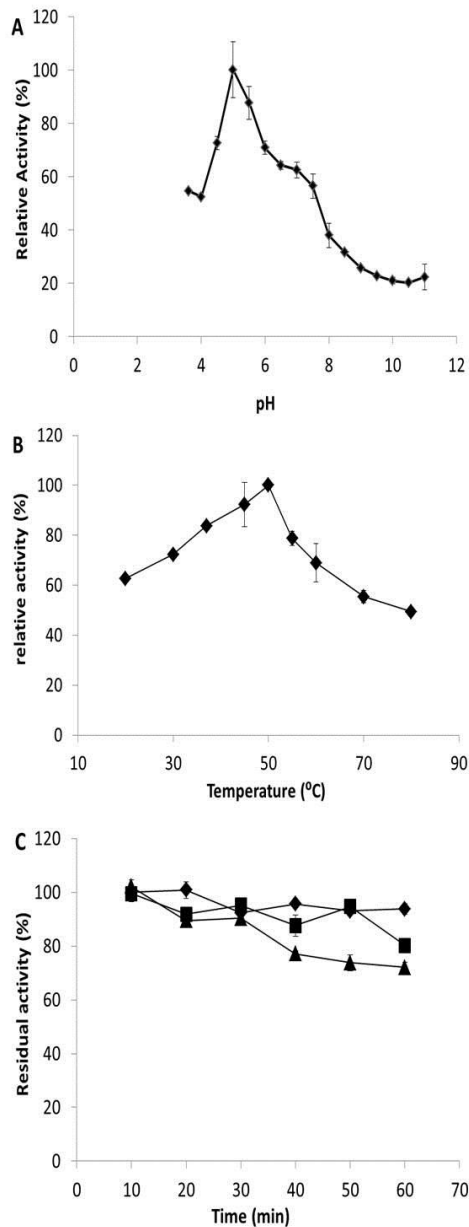
Remarkably, the incorporation of EndoG into ethanologenic MS04 strain made possible ethanol production from CMC or Avicel (Figure 4) with very promising yields. Taken together the data showed that EndoG is a novel, robust, multifunctional cellulase with potential to be expressed directly by the fermentative microorganism thus reducing exogenous enzyme loads and shortening the time of biomass saccharification.

### **Acknowledgements**

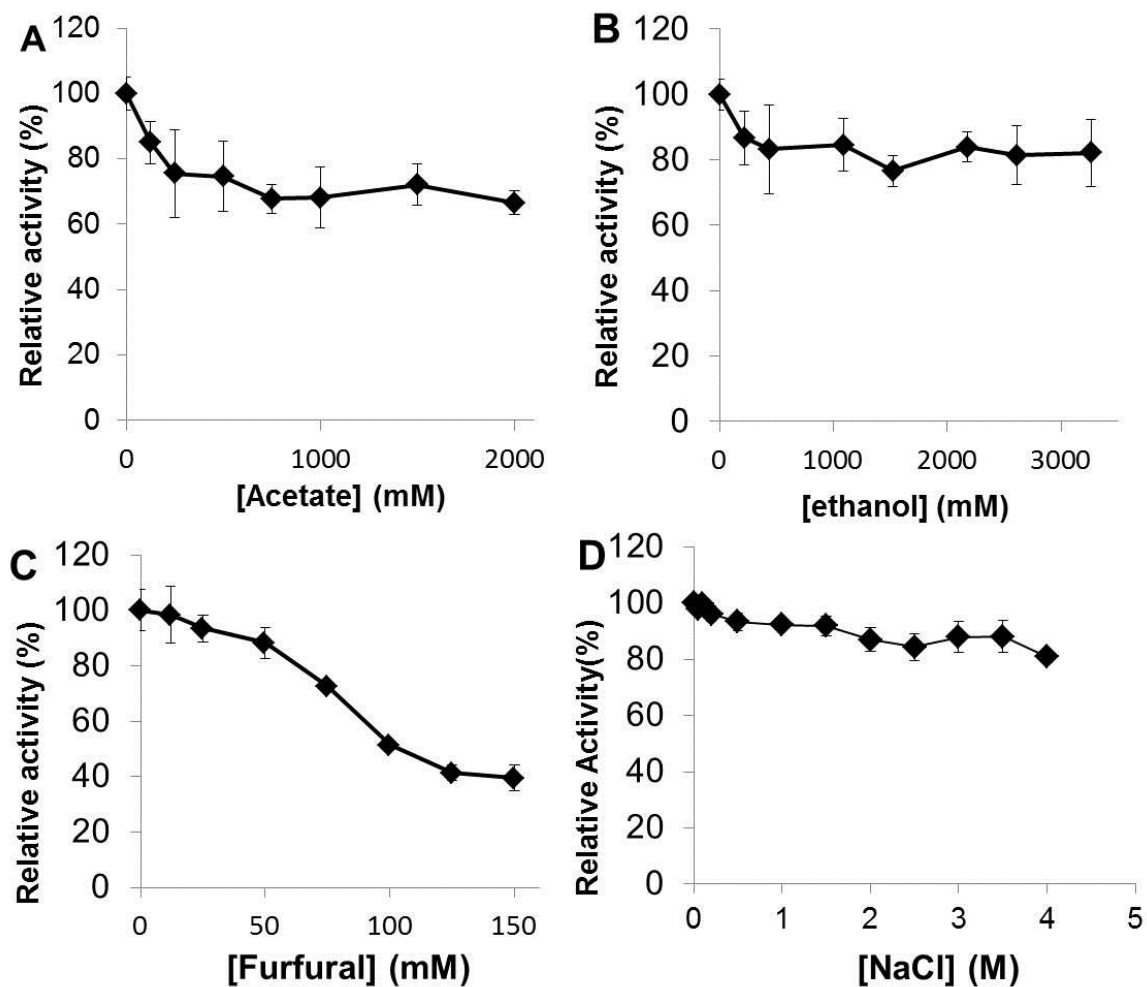
We are very grateful to Nikolai Guchin (ANCAP, Uruguay) for providing valuable materials and to Beatriz Alvarez (School of Science, UdelaR, Uruguay) for helpful advice. This work was supported by grants #FSE\_2009\_23 and #FSE\_2011\_6383 from ANII (Agencia Nacional de

Investigación e Innovación, Uruguay), and grant #FPTA-256 from INIA (Instituto Nacional de Investigación Agropecuaria, Uruguay). IL and FN were also supported by PEDECIBA (Uruguay), by SNI (Uruguay), and by the bilateral international cooperation of SRE-AMEXCID (Mexico).

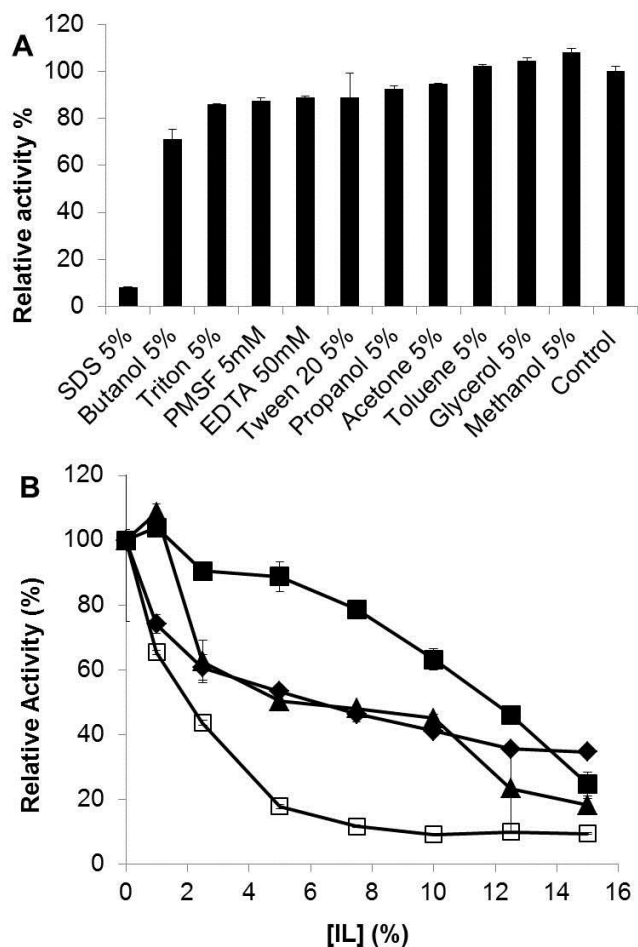
## Figures



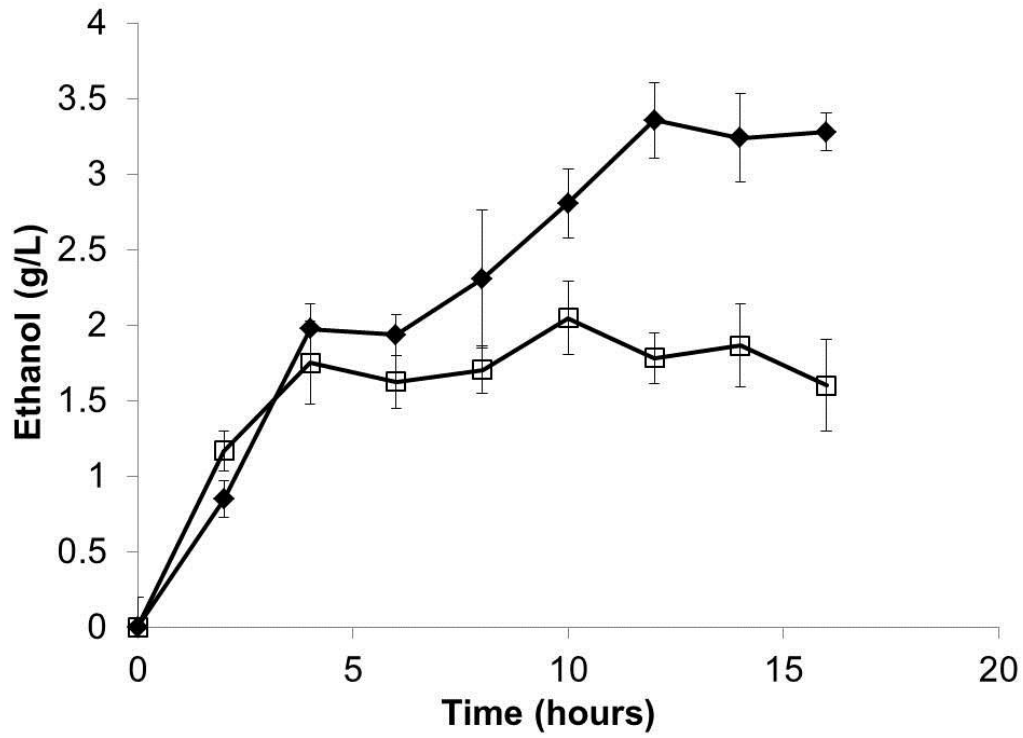
**Fig. 1** Enzymatic properties of recombinant EndoG. Effect of pH (A) and incubation temperature (B) on rEndoG activity using 150 nM pNPC or 2 mM 4-MUC as substrates, respectively. (C) rEndoG thermostability. EndoG was pre-incubated at 50°C (◆), 55°C (■) and 60°C (▲) in the absence of substrate. Residual glucanase activity was assayed at standard conditions with 4-MUC. All data points represent the mean  $\pm$  SD of experiments performed in triplicate.



**Fig. 2** Activity of rEndoG in presence of common inhibitors. The effects of acetate (A), ethanol (B), furfural (C), and NaCl (D) on rEndoG activity were measured at standard reaction conditions using 2 mM 4-MUC as substrate. All data points represent the mean  $\pm$  SD of experiments performed in triplicate.



**Fig. 3** (A) Tolerance of rEndoG to solvents, detergents and inhibitors. All reactions were performed with 2 mM 4-MUC as substrate under standard conditions. Control reactions were identical except that water was added instead of the additive. (B) Activity of rEndoG in the presence of imidazolium ILs: [C4mim]Br (■), [C4mim]Cl (◆), [C2mim]Br, (▲) and [C2mim][OAc] (□). rEndoG activity was measured at standard reaction conditions using 0.5 mM pNPC as substrate. All data points represent the mean  $\pm$  SD of experiments performed in triplicate.



**Fig. 4** Ethanol production from 7.5 g/L CMC (◆) or avicel (□) by co-culture of MS04 cells expressing EndoG and BglC in 200 mL mini-fermenters incubated at 45°C and 150 rpm. Data points represent the mean  $\pm$  SD of experiments performed in triplicate.

## Tables

**Table 1.** EndoG related proteins.

Related Proteins (Accession number) <sup>a</sup>	Origin	Identity (%)	Max score	Query cover (%)	E-value	Reference
Cellodextrinase (ADB80112.1)	Buffalo rumen uncultured microorganism	82	639	95	0	{Duan, 2009 #22}
Cellobiosidase (ACA61171.1)	Buffalo rumen uncultured microorganism	77	614	97	0	{Duan, 2009 #22}
Hypothetical protein GH5 (WP_028911423.1)	<i>Prevotella</i> sp.	76	598	95	0	Unpublished
Cellodextrinase (ADB80113.1)	Uncultured microorganism	74	598	98	0	Unpublished
CelEx-BR12 (AGK74985.1)	Cow rumen uncultured microorganism	74	595	98	0	{Ko, 2013 #16}
Hypothetical protein GH5 (WP_013065043.1)	<i>Prevotella ruminicola</i>	74	593	98	0	{Purushe, 2010 #36}

<sup>a</sup> Comparison made with non-redundant database using BLASTp at NCBI.

**Table 2.** Effect of metal ions on enzymatic activity of EndoG

Metallic ion	Relative activity (%)
Hg	3 ± 2
Cu <sup>+2</sup>	5 ± 1
Zn <sup>+2</sup>	23 ± 1
Co <sup>+2</sup>	29 ± 1
Ag <sup>+</sup>	54 ± 1
Fe <sup>+2</sup>	72 ± 2
Ca <sup>+2</sup>	74 ± 2
Cd <sup>+2</sup>	76 ± 3
K <sup>+</sup>	81 ± 4
Ni <sup>+2</sup>	83 ± 5
Mn <sup>+2</sup>	136 ± 2
Mg <sup>+2</sup>	174 ± 5
Control	100 ± 2

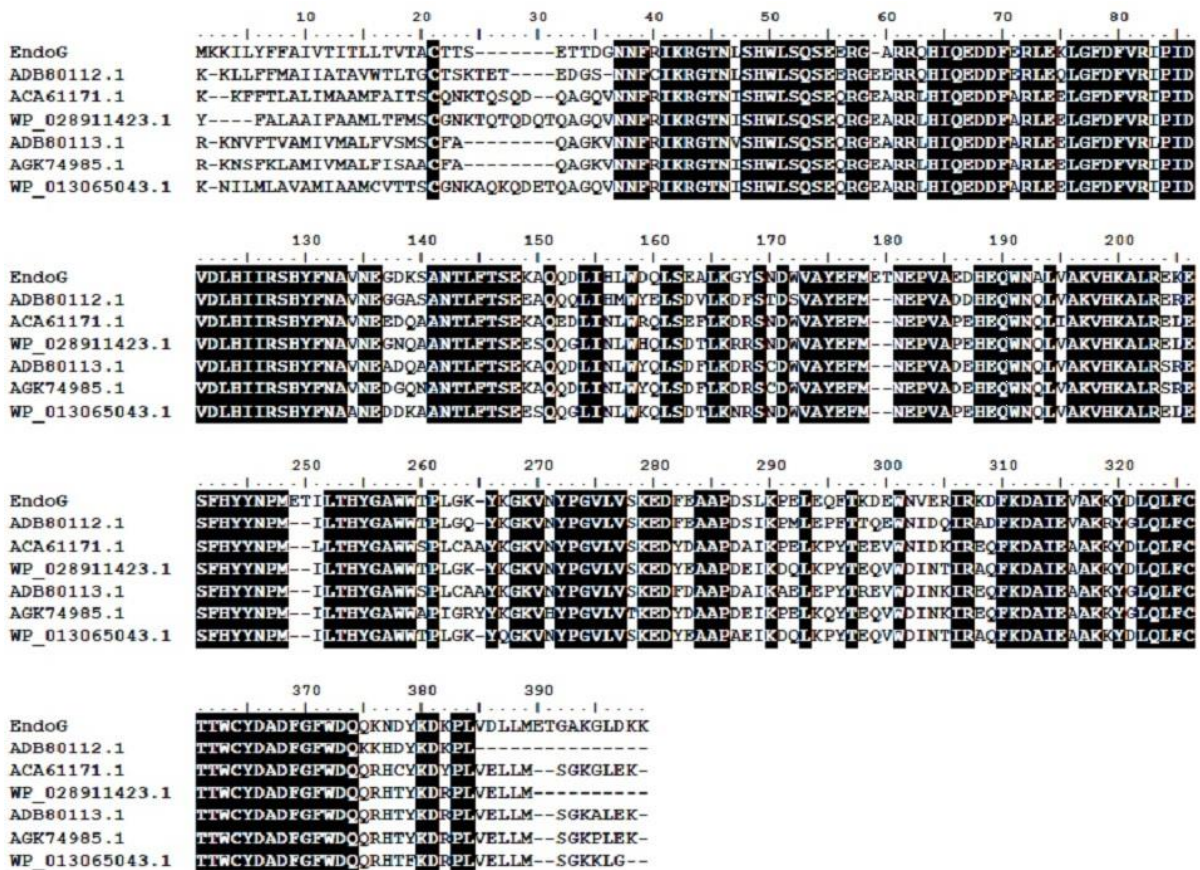
EndoG activity was evaluated in the presence of 5 mM of each cation. All reactions were performed with 2 mM 4-MUC as substrate under standard conditions. Control reactions were identical except that water was added instead of the additive. All data points represent the mean ± SD of experiments performed in triplicate.

**Table 3.** Substrate specificity of EndoG.

Substrate	Activity (U mg <sup>-1</sup> )
Bagasse	35.5 ± 0.3
Xylan	18.5 ± 2.2
Avicel	13.9 ± 0.5
Lichenan	12.3 ± 0.3
CMC	9.2 ± 0.9
PASC	8.5 ± 0.3



## Supplementary Figures



**Fig. S1** Comparison of amino acid sequence of EndoG and their closest homologs. GenBank accession numbers are shown. Regions of identity among sequences are shown as black columns. Signal peptide sequence is indicated by underlines. Alignments were performed using Mega version 6.0.

## References

- [1] L.R. Lynd, P.J. Weimer, W.H. van Zyl, I.S. Pretorius, Microbial Cellulose Utilization: Fundamentals and Biotechnology, *Microbiology and Molecular Biology Reviews*, 66 (2002) 506-577.
- [2] A.K. Chandel, G. Chandrasekhar, M.B. Silva, S. Silverio da Silva, The realm of cellulases in biorefinery development, *Critical reviews in biotechnology*, 32 (2012) 187-202.
- [3] C. Lambertz, M. Garvey, J. Klinger, D. Heesel, H. Klose, R. Fischer, U. Commandeur, Challenges and advances in the heterologous expression of cellulolytic enzymes: a review, *Biotechnology for Biofuels*, 7 (2014) 135.
- [4] B. Henrissat, A classification of glycosyl hydrolases based sequence similarities amino acid, *Biochem Journal*, 280 (1991) 309-316.
- [5] R. den Haan, E. van Rensburg, S.H. Rose, J.F. Görgens, W.H. van Zyl, Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing, *Current opinion in biotechnology*, 33 (2015) 32-38.
- [6] I. Loaces, V. Amarelle, I. Muñoz-Gutierrez, E. Fabiano, A. Martinez, F. Noya, Improved ethanol production from biomass by a rumen metagenomic DNA fragment expressed in *Escherichia coli* MS04 during fermentation, *Appl. Microbiol. Biotechnol.*, In press (2015).
- [7] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nat Meth*, 8 (2011) 785-786.
- [8] Z. Zhang, S. Schwartz, L. Wagner, W. Miller, A Greedy Algorithm for Aligning DNA Sequences, *Journal of Computational Biology*, 7 (2000) 203-214.
- [9] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0, *Molecular biology and evolution*, 30 (2013) 2725-2729.
- [10] L. Bao, Q. Huang, L. Chang, J. Zhou, H. Lu, Screening and characterization of a cellulase with endocellulase and exocellulase activity from yak rumen metagenome, *Journal of Molecular Catalysis B: Enzymatic*, 73 (2011) 104-110.
- [11] K.J. Ellis, J.F. Morrison, Buffers of constant ionic strength for studying pH-dependent processes, *Methods in enzymology*, 87 (1981) 405-426.
- [12] S. Hetzler, D. Bröker, A. Steinbüchel, Saccharification of Cellulose by Recombinant *Rhodococcus opacus* PD630 Strains, *Appl. Environ. Microbiol.*, 79 (2013) 5159-5166.
- [13] G.L. Miller, Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar, *Anal. Chem.*, 31 (1959) 426-428.

- [14] E. Amann, B. Ochs, K.-j. Abel, Tightly regulated *tuc* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli* Gene, 69 (1988) 301-315.
- [15] M.T. Fernandez-Sandoval, G. Huerta-Beristain, B. Trujillo-Martinez, P. Bustos, V. Gonzalez, F. Bolivar, G. Gosset, A. Martinez, Laboratory metabolic evolution improves acetate tolerance and growth on acetate of ethanologenic *Escherichia coli* under non-aerated conditions in glucose-mineral medium, *Appl Microbiol Biotechnol*, 96 (2012) 1291-1300.
- [16] I. Munoz-Gutierrez, R. Oropeza, G. Gosset, A. Martinez, Cell surface display of a beta-glucosidase employing the type V secretion system on ethanologenic *Escherichia coli* for the fermentation of cellobiose to ethanol, *J Ind Microbiol Biotechnol*, 39 (2012) 1141-1152.
- [17] H. Aspeborg, P. Coutinho, Y. Wang, H. Brumer, B. Henrissat, Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5), *BMC Evolutionary Biology*, 12 (2012) 186.
- [18] E.M. Ross, P.J. Moate, C.R. Bath, S.E. Davidson, T.I. Sawbridge, K.M. Guthridge, B.G. Cocks, B.J. Hayes, High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing, *BMC genetics*, 13 (2012) 53.
- [19] M. Hess, A. Sczyrba, R. Egan, T.W. Kim, H. Chokhawala, G. Schroth, S. Luo, D.S. Clark, F. Chen, T. Zhang, R.I. Mackie, L.A. Pennacchio, S.G. Tringe, A. Visel, T. Woyke, Z. Wang, E.M. Rubin, Metagenomic discovery of biomass-degrading genes and genomes from cow rumen, *Science*, 331 (2011) 463-467.
- [20] A. Martinez, M.E. Rodriguez, M.L. Wells, S.W. York, J.F. Preston, L.O. Ingram, Detoxification of dilute acid hydrolysates of lignocellulose with lime, *Biotechnology progress*, 17 (2001) 287-293.
- [21] P. Alvira, E. Tomás-Pejó, M. Ballesteros, M. Negro, Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review, *Bioresource technology*, 101 (2010) 4851-4861.
- [22] L. Qin, Z.-H. Liu, B.-Z. Li, B.E. Dale, Y.-J. Yuan, Mass balance and transformation of corn stover by pretreatment with different dilute organic acids, *Bioresource technology*, 112 (2012) 319-326.
- [23] R. Wahlström, A. Suurnäkki, Enzymatic hydrolysis of lignocellulosic polysaccharides in the presence of ionic liquids, *Green Chemistry*, (2015).
- [24] K.C. Ko, J.H. Lee, Y. Han, J.H. Choi, J.J. Song, A novel multifunctional cellulolytic enzyme screened from metagenomic resources representing ruminal bacteria, *Biochemical and biophysical research communications*, 441 (2013) 567-572.
- [25] A.L. Huston, *Biotechnological aspects of cold-adapted enzymes, Psychrophiles: from biodiversity to biotechnology*, Springer2008, pp. 347-363.
- [26] Y. Sun, J. Cheng, Hydrolysis of lignocellulosic materials for ethanol production: a review, *Bioresour Technol*, 83 (2002) 1-11.

- [27] X. Zhao, K. Cheng, D. Liu, Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis, *Appl Microbiol Biotechnol*, 82 (2009) 815-827.
- [28] A. Tejirian, F. Xu, Inhibition of Cellulase-Catalyzed Lignocellulosic Hydrolysis by Iron and Oxidative Metal Ions and Complexes, *Applied and environmental microbiology*, 76 (2010) 7673-7682.
- [29] X. Li, H.-Y. Yu, Purification and characterization of an organic-solvent-tolerant cellulase from a halotolerant isolate, *Bacillus* sp. L1, *J Ind Microbiol Biotechnol*, 39 (2012) 1117-1124.
- [30] N. Ilmberger, D. Meske, J. Juergensen, M. Schulte, P. Barthen, U. Rabausch, A. Angelov, M. Mientus, W. Liebl, R. Schmitz, W. Streit, Metagenomic cellulases highly tolerant towards the presence of ionic liquids—linking thermostability and halotolerance, *Appl Microbiol Biotechnol*, 95 (2012) 135-146.
- [31] S.S. Tan, D.R. MacFarlane, J. Upfal, L.A. Edye, W.O. Doherty, A.F. Patti, J.M. Pringle, J.L. Scott, Extraction of lignin from lignocellulose at atmospheric pressure using alkylbenzenesulfonate ionic liquid, *Green Chemistry*, 11 (2009) 339-345.
- [32] M.d.C. Portillo, A. Saadeddin, Recent trends in ionic liquid (IL) tolerant enzymes and microorganisms for biomass conversion, *Critical reviews in biotechnology*, (2014) 1-8.
- [33] C.J. Duan, L. Xian, G.C. Zhao, Y. Feng, H. Pang, X.L. Bai, J.L. Tang, Q.S. Ma, J.X. Feng, Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens, *Journal of Applied Microbiology*, 107 (2009) 245-256.
- [34] J. Purushe, D. Fouts, M. Morrison, B. White, R. Mackie, P. Coutinho, B. Henrissat, K. Nelson, Comparative Genome Analysis of *Prevotella ruminicola* and *Prevotella bryantii*: Insights into Their Environmental Niche, *Microb Ecol*, 60 (2010) 721-729.

**Capítulo 5.**  
**PRODUCCIÓN DE ETANOL A PARTIR DE**  
**GLICEROL CRUDO DERIVADO DE LA**  
**INDUSTRIA DEL BIODIESEL**

---

Loaces, I., Rodríguez, C., Amarelle, V., Fabiano, E. & Noya, F. **Ethanol production using *E. coli* from biodiesel derived glycerol.** En preparación

## Resumen

Actualmente en nuestro país se producen 90.000 toneladas anuales de biodiesel. Como residuo de esta industria se genera "glicerina cruda", una mezcla de glicerol, metanol, NaOH o KOH, cenizas y otras materias orgánicas, en una relación de 10% (p/p). Este desecho puede convertirse en productos con valor agregado lo cual repercutiría favorablemente en toda la cadena de producción de biocombustibles. En esta parte del trabajo se evaluó la metagenoteca obtenida del reactor anaerobio en medio mínimo suplementado con glicerina cruda y se seleccionó un clon en base a su rápido crecimiento en este medio. Éste fue complementado con el plásmido pLOI297 y su fósido se transfirió a la cepa productora de etanol LY180. Se realizaron ensayos de fermentación de glicerol puro y crudo, evaluando diversas condiciones de fermentación.

## Objetivos

- ☞ Seleccionar, dentro del metagenoma obtenido de lodos de un reactor anaerobio, clones que promuevan el crecimiento de *Escherichia coli* a partir de glicerol crudo derivado de biodiesel.
- ☞ Evaluar el potencial de las cepas Epi (G1, pLOI297) y LY180 (G1) para producir etanol a partir de glicerol puro y crudo.

## Principales resultados y conclusiones

El screening del metagenoma construido a partir de lodos de un reactor anaerobio en placas de M9 suplementada con 4% (v/v) de glicerol crudo permitió la detección de un clon, denominado Epi (G1), capaz de crecer a las 16 horas de incubación a 30°C. Su fósido fue secuenciado en su totalidad, 41312 pb, y no mostró homología con ninguno de los genomas depositados en la base de datos del NCBI. Los 38 ORFs predichos fueron sometidos a análisis de comparación, y resultó que la mayoría presentan similitud de secuencia con proteínas no caracterizadas, pertenecientes a los phylum Bacteroidetes (*Bacteroides*, *Hymenobacter*, *Elizabethkingia*, *Alistipes sp.* y *Prevotella sp.*) y Firmicutes (*Desulfitobacterium*).

El clon Epi (G1) fue complementado con el plásmido pLOI297 (que posee los genes alcohol dehidrogenasa (*adh*) y piruvato decarboxilasa (*pdh*) de *Zymomonas mobilis*) (234). Se realizaron ensayos de fermentación a partir de glicerol puro, suplementado con triptona y extracto de levadura, y de glicerol crudo obteniéndose un máximo de 10,4 g/L y 32 g/L de etanol, respectivamente, a las 24 horas de incubación, aproximadamente.

Por otro lado, el fósido G1 fue incorporado a la cepa etanológica LY180 (235). Esta cepa es incapaz de utilizar glicerol puro como fuente de carbono en condiciones anaerobias, pero es capaz de producir 75 g/L de etanol a partir de 50% de glicerol crudo previamente neutralizado, en 12 horas a 37°C.

Se determinó que hay una clara relación dosis-respuesta entre la producción de etanol y la cantidad de sustrato añadido al medio, acompañado, en el caso de LY180 (G1), de la producción de H<sub>2</sub>, y en el caso de Epi (G1, pLOI297) de ácido acético y ácido propiónico.

## **Contribución personal al trabajo**

La primer actividad que realicé fue poner a punto los medios de cultivo necesarios para realizar los screening en medio sólido y líquido y para los ensayos de fermentación, tanto con glicerol puro, glicerina cruda y con glicerina cruda "neutralizada". Luego realicé la selección de clones en medio sólido y la confirmación mediante curvas de crecimiento en medio líquido.

F. Noya se encargó del ensamblado del fósido G1 luego de su secuenciación; mientras que la mutagénesis la realizó C. Rodriguez. Realicé en los análisis bioinformáticos del fósido, la anotación de los ORFs y los estudios filogenéticos del ORF seleccionado.

Los ensayos de fermentación los realicé en las instalaciones del IIBCE, por lo que fue necesario implementar en el laboratorio un sistema que permitiera el control de las condiciones de fermentación y la detección de etanol en tiempo real.

El trabajo presentado en este capítulo fue financiado por un proyecto FSE de la ANII y PEDECIBA.

## Ethanol production using *E. coli* from biodiesel derived glycerol

Inés Loaces<sup>1</sup>, Cecilia Rodriguez<sup>1</sup>, Vanesa Amarelle<sup>1</sup>, Elena Fabiano<sup>1</sup>, Francisco Noya<sup>1\*</sup>

Departamento de Bioquímica y Genómica Microbianas<sup>1</sup>, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay

\* Corresponding author: Mailing address: Departamento Bioquímica y Genómica Microbianas, Instituto de Investigaciones Biológicas Clemente Estable, Av. Italia 3318, Montevideo, Uruguay. CP: 11600. Phone: (598) 24871616. Fax: (598) 2487-5461. E-mail: fnoya@iibce.edu.uy

### Abstract

Biodiesel is a renewable, efficient, environmentally friendly and biodegradable fuel made from vegetable oil and animal fats. As a result of the successful biodiesel and oleochemicals manufacturing taking place worldwide 2 million tonnes of glycerol consistently arrive at market every year. Crude glycerol represents a great opportunities for new applications. Because of its reduced state, compared to sugars, it can be used to significantly increase the product yield of chemicals, such as succinate, ethanol, and propanediols. In this study a functional metagenomic library from an anaerobic reactor was evaluated looking for genetics elements that promote the use and fermentation of crude glycerol by *Escherichia coli*. One clone was selected based on its fast growth rate on crude glycerol. This fosmid, named G1, was fully sequenced (41 Kpb) and no obvious gene involved in glycerol uptake was detected. Ethanologenic strain LY180 and Epi300 (pLOI297) were complemented with G1 and evaluated for ethanol production using crude and pure glycerol as raw materia, in several growth conditions. A maximum of 10.4 g/L and 32 g/L were obtained when Epi300 (G1, pLOI297) was employed for pure and crude glycerol fermentation, respectively. While, using LY180 (G1) the ethanol production increased to 75 g/L when crude glycerol was utilized as the sole carbon source.

### Keywords

Crude glycerol; biodiesel; *Escherichia coli*; functional metagenomics

### Introduction

The worldwide energy demand has been growing exponentially, while the reserves of fossil fuels have not been increasing. It is extensively accepted that the dependence on fossil fuels as primary energy source contributes to global climate change, environmental damage and health problems (Varrone, Giussani et al. 2012).



Biodiesel is one of the best choices of alternative fuels. It is a renewable, efficient, environmentally friendly and biodegradable fuel made from vegetable oil and animal fats. It is currently the only alternative fuel that has the potential to completely displace its petroleum counterpart, as 100% biodiesel can be used in conventional diesel engines with little to no major modifications to fuel system and little impact on engine performance (Johnson and Taconi 2007). Oils and fats generated in the food industry and restaurants can be recovered and used as feedstock for biodiesel (Lai 2014). It is produced by a transesterification reaction of a triglyceride and an alcohol, a process that generates large amounts of glycerol as a byproduct, about 10% (w/w). As a result of the successful biodiesel and oleochemicals manufacturing taking place worldwide since more than a decade, 2 million tons of glycerol consistently reach the market every year (Ciriminna, Pina et al. 2014). As the worldwide production of biodiesel continues to increase, the development of effective uses for glycerol may prove essential for the economics and competitiveness of the biodiesel industry. For that reason, more attention is being paid to the utilization of crude glycerol from biodiesel production in order to defray the production cost of biodiesel and to promote biodiesel industrialization on a large scale (Yang, Hanna et al. 2012).

The reduced state of glycerol, compared to sugars, such as glucose, xylose, etc., promises to significantly increase the product yield of chemicals, whose production from these sugars is limited by the availability of reducing equivalents (Dharmadi, Murarka et al. 2006, Clomburg and Gonzalez 2013). Various works evaluated the utilization of glycerol to produce diverse value-added chemicals like 1,3-propanediol, hydrogen or aminoacids (Papanikolaou, Fick et al. 2004, Sarma, Brar et al. 2012, Meiswinkel, Rittmann et al. 2013). Industrial production of ethanol is dominated by the conversion of sugars from starch or sucrose to ethanol by the yeast *Saccharomyces cerevisiae* (Gray, Zhao et al. 2006). Furthermore, the discovery that *E. coli* (Gonzalez, Murarka et al. 2008) ferment glycerol in an ethanol-1,2-PDO-dependent manner opened the possibility to use these organisms for the direct conversion of glycerol into ethanol. Glycerol uptake and ethanol production metabolic route were modified on *E. coli* strains (Hu and Wood 2010, Matthew S. Wong 2014), but no new genes or alternative strategies were evaluated for this purpose.

In this study metagenomic was used to look for genes that contribute to crude glycerol utilization as a carbon source by *Escherichia coli*. Ethanologenic strain LY180 and Epi300 complemented with plasmid pLOI297 were evaluated for ethanol production from crude and pure glycerol in several growth conditions.

## **Materials and methods**

### **Metagenomic library construction**

Metagenomic DNA was isolated from the bacteria fraction of an anaerobic reactor fed with ruminal digesta (total solids (TS) 19,3% (w/w); volatile solids (VS) 17,3% (w/w)); tannery fats (TS 25,1%, VS

16,6%) and inoculated with an activated sludge purge of oil-wastewater treatment (TS 25,6% y VS 21,7%). This reactor operates at the Departamento de Reactores of Facultad de Ingeniería, Montevideo, Uruguay. DNA fraction of 30-50Kbp was selected and cloned using Copy Control Fosmid Library Production Kit (Epicentre Biotechnologies, Madison, WI, US). Epi300-T1<sup>R</sup> *E. coli* were transformed and selected in Luria broth (LB) (Bertani 1951) medium supplemented with chloramphenicol (Cm) 12.5µg/mL, placed in 96 well plates and stored at -80°C.

### **Screening of metagenomic library**

Crude glycerol was obtained as a sub-product of biodiesel industry (ALUR, Montevideo, Uruguay). The impurities of the industrial feedstock were mainly composed of ash (4.7% wt/wt), methanol (4% wt/wt), non-glycerol organic materials (13.9% wt/wt) and water (2.8% wt/wt). Minimum media M9 (MgSO<sub>4</sub> 2 mM, CaCl<sub>2</sub> 0,1 mM, Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O 12,8 g/L, KH<sub>2</sub>PO<sub>4</sub> 3 g/L, NaCl 0,5 g/l, NH<sub>4</sub>Cl 1 g/L, thymine 50 µg/mL, leucine 200 µM) supplemented with 4% (v/v) crude glycerol was used for agar plate screening. Metagenomic library was placed in M9- 1% (w/v) glucose plates, grow overnight and then transposed to screening media. Positive clones were those able to grow after 16 hours of incubation at 30°C. Growth curves in liquid media were made with 4% (v/v) of crude neutralized glycerol. A 50% (v/v) crude glycerol dilution was neutralized using 1% HCl until pH 7.0, and then was filtrated using filter paper and the liquid fraction was sterilized by autoclaving.

### **G1 sequence and in silico analysis**

The DNA of G1 fosmid was sequenced by Ion Torrent PMG, using the Ion PGM™ Sequencing Kit (Life Technologies, Invitrogen, CA, US) in Ion 314™ Chip (Life Technologies). The library construction was done using Ion Xpress™ Plus commercial Kit (fragmentation of 200pb) (Life Technologies) and Ion PGM™ Template OT2 200 Kit (Life Technologies) was used for the template preparation. *E.coli* DNA contamination was filtered by Mirabait using K12 strain genome as template. The reads obtained were assembled using Mira 4.0 rc5, with a coverage of 40X (7500 reads of 235bp on average were random selected). Manual trimming was done with Gap5 program. Open reading frames (ORFs) were predicted from the assembled contigs using MetaGene (Noguchi, Park et al. 2006) on WebMGA on-line server. The translated protein sequences of the predicted genes were subjected to BLASTp comparison with the non-redundant protein database (NR, <http://blast.ncbi.nlm.nih.gov/>) and the CDD domain database (including Cluster of Ortholog Genes classification, COG, <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

ORF G1\_23 sequence was compared with protein database of NCBI using BLASTP algorithm (Zhang, Schwartz et al. 2000). Complementary analysis was made with HMMER (Finn, Clements et al. 2011) using profile Hidden Markov Model (HMMs). The prediction of transmembrane motifs was performed by TMHMM2.0 server (Krogh, Larsson et al. 2001), and protein subcellular location prediction was

done by PSORTb v.3.0 (Nancy, Wagner et al. 2010). The prediction of signal peptide was carried out using SignalP 4.1 (Nielsen, Engelbrecht et al. 1997).

Closest homologs to ORF G1\_23 were selected and multiple sequences alignments were computed with MEGA 6 (Tamura, Stecher et al. 2013) by Muscle method. Phylogenetic tree was constructed using the Maximum likelihood method implemented in MEGA 6 (Tamura, Stecher et al. 2013).

### **Gene identification by mutagenesis**

G1 fosmid was *in vitro* mutagenized using EZ-Tn5 Insertion Kit (Epicentre Biotechnologies, Madison, WI, US) according to the manufacturer's instructions and transformed into fresh electrocompetent EPI300-T1<sup>R</sup> cells. The Tn5-mutants were selected in LB-Cm plates supplemented with kanamycin 50µg/mL. Transformants were transferred to M9- 4% (v/v) crude glycerol plates, incubated at 30°C during 16 h and colonies smaller than wild type Epi300 (G1) were selected. The single insertion of the transposon was determinate by restriction analysis. Flanking DNA was sequenced by conventional Sanger method (Macrogen, Seoul, Korea).

### **Ethanol production assay**

For ethanol production Epi300 (G1) was complemented with pLOI297 plasmid (Ingram, Conway et al. 1991) and the selected metagenomic fosmid (G1) was transferred by electroporation to ethanologenic strain LY180 (Jarboe, Grabar et al. 2007). Competent, electro-competent cell, pure plasmid and fosmid were obtained with classical methods. All the strains were conserved at -80°C.

Batch fermentations were carried out in mineral M9, for Epi300 T1<sup>R</sup>, or NBS medium (Zhang, Shanmugam et al. 2010) for LY180, employing pure glycerol or neutralized crude glycerol as carbon source. Fermentation was performed in Fleakers mini-fermenters containing 200 mL of mineral medium, without aeration, 37°C, and 150 rpm. pH was adjusted to 7.0 before autoclaving and remained un-controlled during the fermentation. To develop the inoculum, a 5mL LB liquid overnight culture containing the ethanologenic strain was seeded into a mini-fermenter with 20 g/L of glucose as carbon source. When the cells reached an OD<sub>620</sub> of almost 2, they were centrifuged for 10 min at 4,000 g and inoculated, at an initial OD<sub>620</sub> of 1, into mini-fermenters containing 4, 10, 20 or 50% (v/v) of pure or neutralized crude glycerol. When pure glycerol was used the media was supplemented with tryptone and yeast extract at 1% final concentration. When more than 20% of carbon source was added, the media was supplemented with 1 mM HCl-betaine. For fermentation assay with initial OD<sub>620</sub> of 8 the inoculum was developed in 1 L Wheaton flask.

Kinetic of ethanol production was determined in 1.2 L fermenter containing 1 L of mineral NBS medium, 37°C, and 150 rpm. pH was adjusted to 7.0 before autoclaving and remained un-controlled

during the fermentation. Inoculum development proceeded just as discussed earlier. N<sub>2</sub> was added to fermenter atmosphere before incubation to displace oxygen.

## **Analyses**

Growth was determined spectrophotometrically by optical density OD at 620 nm (OD<sub>620nm</sub>), during 40 hours in 96 well plates using Varioskan Flash Multimode Reader (Thermo Scientific, MA, US).

Ethanol and oxygen was measured by BCP-EtOH/ BCP-O<sub>2</sub>EC detector (BlueSens, Herten, Germany). H<sub>2</sub> and CO<sub>2</sub> were determined by gas chromatography (310C Gas Chromatograph, SRI, California, US) with argon as gas carrier.

For detection and quantification of volatile fatty acids (VFA) an HPLC separation was performed. Samples were centrifuged (15000 rpm, 15 min, 5°C) and supernatants (100 µL) were injected in a Waters modular HPLC system (Waters Associates, Milford, MA.). The HPLC system consisted in a binary HPLC pump (Waters 1525), an autosampler (Waters 717 plus), a Rezex ROA-Organic Acid H+ (8%) (300\*7.8 mm) column (Phenomenex, USA) and a photodiode array detector (Waters 2998) linked to Empower 2 (Waters) chromatography data software. The temperature of the column was set at 30°C. The mobile phase used was: H<sub>2</sub>SO<sub>4</sub> 0.015M and was pumped at 0.3 mL/min by isocratic mode. Quantitative analysis was performed against a standard solution containing 5mg/mL of each VFA.

All assays were made for triplicate. Tukey test was used for statistical analysis, with p value of 0.05, using Infostat software.

## **Results and Discussion**

### **Metagenomic library screening**

A 100,000 clones metagenomic library was made from the purge sludge of an anaerobic digester. *NotI* restriction pattern of fosmids confirmed that they were 30-40 Kbp long. Because *E. coli* is capable to growth aerobically with glycerol as carbon source (Martínez-Gómez, Flores et al. 2012), the functional screening was made in crude glycerol without neutralization, looking for clones with a faster growth than Epi300 (after 24-48 hours of incubation all clones were able to growth). With this approach, Epi300 (G1) clone was selected (data no show). The utilization of neutralized crude glycerol as a carbon source was evaluated in liquid media (Fig. 1). Epi 300 (G1) confirmed that it was able to growth in minimum media with 4% of glycerol as sole carbon source.

## G1 sequence analysis

G1 fosmid (41.312pb, including pCC1FOS vector) was completely sequenced. A total of 114,683 reads were obtained, with a mean read length of 206 pb. Its nucleotide sequence did not show homology with any reference Genome in the NCBI database (refseq\_genomic, 11797300 sequences as of January 2015). 38 ORFs were predicted in G1 sequence (Table 1), mostly with homology to non-characterized proteins, belonging to Phylum Bacteroidetes (*Bacteroides*, *Hymenobacter*, *Elizabethkingia*, *Alistipes* sp. and *Prevotella* sp.) and Firmicutes (*Desulfitobacterium*).

Tn5-induced *in vitro* mutagenesis of G1 fosmid was used to identify the gene responsible for the fast growth rate in glycerol as carbon source. Loss of phenotype mutants were isolated by screening M9-crude glycerol plates and the single insertion of the transposon was verified by restriction analysis. Three mutant clones were selected (data not shown), and the mutation flanking regions were sequenced and mapped inside G1 sequence (Fig. 2). Two of the mutations (mt9 and mt34) interrupted the sequence of ORF G1\_23, while mutant 40 (mt40) was at the 3' end of this ORF (Fig. 2).

BLASTp results of ORF G1\_23 (Fig. 3a, Table 2) showed high sequence identity (SI) with hypothetical proteins. Those which showed higher SI are [WP\\_031540335.1](#), from *Bacteroidetes* (E-value: 1e-28, SI: 38%) and [WP\\_028913135.1](#), from *Prevotella* sp. MA2016 strain (E-value: 1e-19, SI: 37%). Other matches were the putative protein named PRU\_1480 ([ADE82526.1](#)) from *Prevotella ruminicola*, showed 33 % of SI (E-value: 8e-28), a hypothetical protein, [WP\\_033149883.1](#) (E-value: 3e-26, SI: 31%) from *Prevotella* sp. RM4 and one from *P. ruminicola* ([WP\\_041385930.1](#)) (E-value: 4e-28, SI: 32%). According to PhyloPythiaS, G1 was related to *Alistipes* at Genera level. No conserved domains or homology with any Pfam or COG could be established for this ORF. The prediction of a putative signal peptide on ORF G1\_23 was made using SignalP and PSORTb V3.0 and it appeared to encompass the first 20 residues. The putative excision site was predicted between residues 23 and 24 (data not shown). The prediction of transmembrane helices using TMHMM 2.0 server of G1\_23 assigned 0.4% of probability for a transmembrane motive on first 25 amino acids. These results suggest that ORF G1\_23 could be secreted to the extracellular space.

## Ethanol production using Epi300 strains

To evaluate the potential of ethanol production from glycerol Epi300 (G1) was complemented with pLOI297. This plasmid carries the *pet* operon (Alterthum and Ingram 1989) (alcohol dehydrogenase (*adh*) and pyruvate decarboxylase (*pdc*) genes from *Z. mobilis*) and confers to *E. coli* the ability to produce ethanol as principal fermentative product (Senthilkumar and Gunasekaran 2005).

To verify the fermentative utilization of glycerol by *E. coli*, ethanol and oxygen were followed for 50 h in 1 L fermenter with M9 media supplemented with 10% pure glycerol. Because it has been previously reported that *E. coli* requires a rich-media to effectively ferment glycerol (Dharmadi, Murarka et al. 2006), tryptone and yeast extract were added to the grow media. Kinetic of ethanol production was

evaluated using Epi300 (carrying a negative fosmid (C-) plus pLOI297) and Epi300 (G1, pLOI297) in anaerobic conditions (Fig. 4a). A lag phase of nine hours was observed for both strains, after that ethanol increased during 9 hours for Epi300 (C-, pLOI297) and 13 hours for Epi300 (G1, pLOI297) until a maximum value of 3.4 and 5.6 g/L ethanol was reached for Epi300 (C-, pLOI297) and Epi300 (G1, pLOI297), respectively.

Different initial concentrations of pure glycerol were used to evaluate the performance of these strains. A clear dose-response dependence was observed (Fig. 4b), reaching a maximum of ethanol of 10.4 g/L when 50% of pure glycerol was fermented using Epi300 (G1, pLOI297). Statistically significant differences were observed between the responses of the strain Epi300 with or without G1 fosmid, in all crude glycerol concentrations evaluated.

The Epi300 (G1-Tn5) mutants obtained (mt9, mt34 and mt40) were complemented with pLOI297 plasmid and evaluated for ethanol production too. When 10% of pure glycerol was employed as carbon source, all the mutants showed similar levels of ethanol production (Fig. 4c) significantly lower ( $p > 0.05$ ) than Epi (G1, pLOI297).

The aim of our work was to ferment biodiesel-derived crude glycerol for ethanol production. Because of that, and after verifying the capacity pure glycerol utilization by Epi300 (G1, pLOI297), neutralized crude glycerol was employed as carbon source.

Different initial concentrations of crude glycerol were employed as carbon source in minimum media. A marked increase in ethanol production was obtained compared with pure glycerol (Fig. 4 and 5). A larger crude glycerol concentration determined an increase in ethanol production. 32 g/L ethanol was obtained when 50% of crude glycerol was fermented (Fig. 5a), more than 300% of improvement respect to the same initial concentration of pure glycerol. Statistical difference between Epi300 (G1, pLOI297) and Epi300 (pLOI297) were determined when 4 or 50% of crude glycerol was employed as carbon source (Fig 5a). The performance of the Epi300 (G1-Tn5) mutants was also evaluated, and showed no statistical difference with parental strain (Fig. 5b) when crude glycerol was used as carbon source.

In fermentations in which crude glycerol was substituted by crude glycerol no addition of extra nutrients was needed. In both, pure and crude glycerol Epi300 (G1, pLOI297) fermentations, ethanol was produced as the main fermentation product, but acetic and propionic acid were detected in order 0.1-1.12 mg/mL (data no showed). According to Gonzalez et al. (2008) (Gonzalez, Murarka et al. 2008) ethanol and acetate are the primary fermentation products, when glycerol is metabolized through the glycerol dehydrogenase (encoded by *gldA*) and dihydroxyacetone kinase (encoded by *dhaKLM*) pathway in *E. coli*. Ethanol levels obtained using Epi300 (G1, pLOI297) were higher in all crude glycerol concentrations used compared to pure glycerol containing media. These results might be explained by the presence of some carbon or electron sources in the crude glycerol which contributed to ethanol production principally at more than 20% crude glycerol. Matter of organic non-glycerin

(MONG) such as fatty acid contained in the crude glycerol might influence on cell growth and ethanol production (Lee, Kim et al. 2012).

### **Ethanol production using LY180 strain**

Strains carrying multi-copy plasmids are generally not used in industry due to their instability (Hahn-Hägerdal, Karhumaa et al. 2007). Presence of plasmids requires auxotrophic or antibiotic resistance markers to be retained in the cells, both of which are not applicable in industrial media. So, genetic stability is a prerequisite when designing strains for industrial use. LY180 strain ( $\Delta$ frdBC::(*Zm frg cel*YEc)  $\Delta$ ldhA::(*Zm frg casA B*Ko) *adhE*::(*Zm frg estZ*Pp FRT)  $\Delta$ ackA::FRT *rrlE*::(*pdc adhA adhB* FRT)  $\Delta$ mgA::FRT) successfully produces ethanol as main fermentation product from sugars, tolerates furfural and high ethanol concentrations (Alterthum and Ingram 1989, Ingram, Conway et al. 1991, Yomano, York et al. 1998, Miller, Jarboe et al. 2009).

Therefore, the ethanol production by LY180 complemented with G1 fosmid, G1-Tn5 mutants and parental strain was evaluated. Using mini-fermenters system with 10% neutralized crude glycerol (Fig. 6a)  $7.8 \pm 0.5$  g/L ethanol was produced by LY180 (G1) strain, significantly more than parental strain or LY180 complemented with mutants G1-mt9 and G1-mt34. CO<sub>2</sub> and H<sub>2</sub> trace amounts were detected, this is consistent with the fact that during glycerol fermentation hydrogen is generated by the disproportionation of formate into CO<sub>2</sub> and H<sub>2</sub>, a reaction catalyzed by the enzymatic complex FHL (Sawers, Blokesch et al. 2004). This metabolic pathway is not modified in LY180 strains. Murarka et al. 2008 (Murarka, Dharmadi et al. 2008), propose that hydrogen recycling takes place during glycerol fermentation and that this gas is used as electron donor in the conversion of fumarate to succinate.

The effect of increased crude glycerol concentration in ethanol production was also evaluated with LY180. 4, 10, 20 and 50% of neutralized crude glycerol was added to 200 mL of media and incubated for 24 hours at 37°C and 150 rpm (Fig. 6b). Increasing of crude glycerol concentrations had a positive effect in ethanol production, reaching a maximum of  $51.18 \pm 3.79$  and  $60.20 \pm 7.56$  g/L when LY180 or LY180 (G1) were used, respectively.

Kinetic of ethanol production was evaluated. Ethanol and oxygen were measured along time during fermentation of NBS minimum media supplemented with 50% neutralized crude glycerol in 1 L fermenter. LY180 and LY180 (G1) were compared in microaerophilic conditions (Fig. 6c), showing a fast increase in ethanol level in the first 12 hours, and then a constant but slower rate of production in both strains, reaching 73 g/L and 63 g/L of ethanol at 50 hours of fermentation, for LY180 (G1) and LY180, respectively.

Initial OD increase from 1 to 8, had a positive effect in ethanol production, enhancing from 51.0 to 67.2 and 60.2 to 75.2 g/L when LY180 and LY180 (G1) were employed, respectively, in 50% crude glycerol (data not shown).

Fermentation of pure glycerol was evaluated too. Neither LY180 nor LY180 (G1) could produce ethanol from this substrate using the same minimal medium without additives or supplemented with tryptone and yeast extract.

## Conclusions

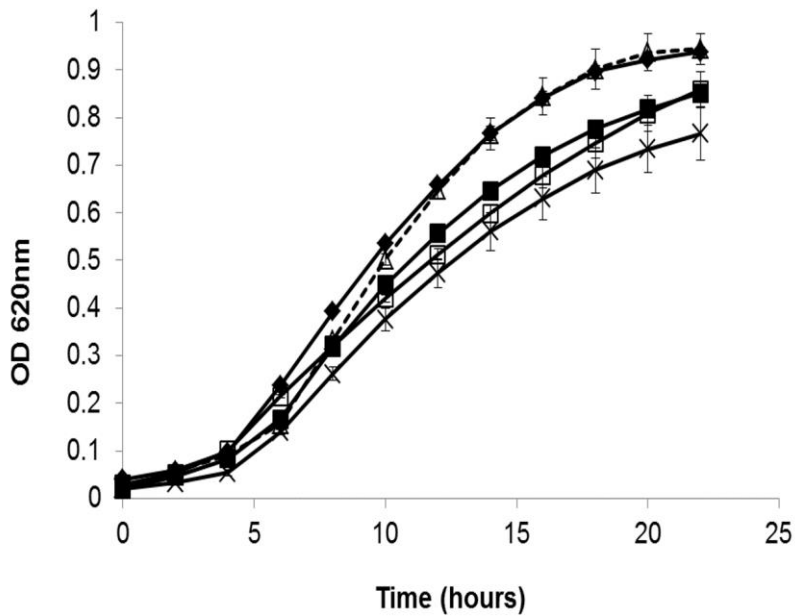
The synthesis of ethanol from glycerol is the only pathway that can perform the two major functions that determine the fermentative metabolism: attaining redox balance and generation of ATP via substrate-level phosphorylation. The operation of the ethanologenic pathway results in the net production of 1 ATP per glycerol metabolized, which favorably compares with the ATP yield obtained by fermentation of other sugars.

The fermentative conversion of glycerol to ethanol has previously been reported by metabolically engineered *E. coli* strains. Durnin et al. (Durnin, Clomburg et al. 2009) reported the production of more than 20 g/L of ethanol accompanied by hydrogen or formate production from 60 g/L of crude glycerol by a *E. coli* strain with dihydroxyacetone kinase (*dhaKLM*) and glycerol dehydrogenase (*gldA*) genes under control of  $P_{LtetO-1}$ . Similar results were obtained from 50 g/L of crude glycerol after 96 hours of fermentation by *E. coli* MG1655 (pZSKLMgldA) strain (Cintolesi, Clomburg et al. 2012). An *E. coli* strain carrying a plasmid that overexpressed glycerol uptake genes *dhaKLM*, *gldA*, and *glpK*, as well as the ethanol pathway gene *adhE* produced 41.9 g/L of ethanol 72 hours from 110 g/L glycerol (Matthew S. Wong 2014).

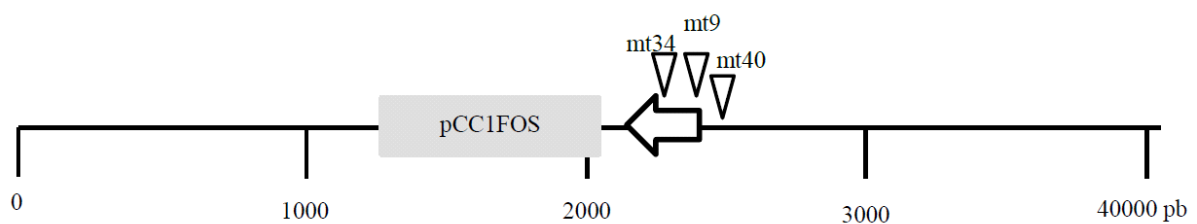
In our work we used two *E. coli* strains: one carrying a plasmid that overexpressed the *Z. mobilis* homoethanol pathway (*adh* and *pdh* genes), and other that have the pET operon integrated within *rrlE*, a 23S ribosomal RNA subunit, ensuring high expression levels. Fumarate reductase (*frdBC*), lactate dehydrogenase (*ldhA*), propionate kinase (*acaK*) and native aldehyde/alcohol dehydrogenase (*adhE*) are deleted in this strain, ensuring that only ethanol and fumarate (or CO<sub>2</sub> and H<sub>2</sub>) can be produced by fermentative assimilation of carbon sources. But not glycerol uptake genes were incorporated or overexpressed in these strains. Metagenomic was used looking for genetics resources for glycerol uptake and/ or metabolism. No-neutralized crude glycerol was employed for agar plates screening and G1 clone (in which no obvious gene involved in glycerol uptake was identified) was selected based on its fast growth rate. Is possible that its advantage was related to tolerance to same inhibitor present in glycerin which is eliminated during neutralize and filtration steps. Further works using ORF G1\_23 and its encoded protein, could explain its implication in the observed advantage in crude glycerol utilization.



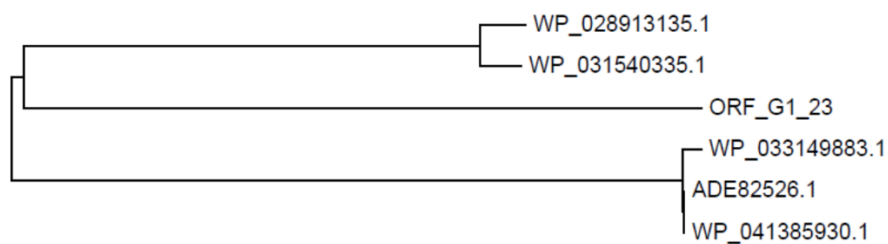
## Figures



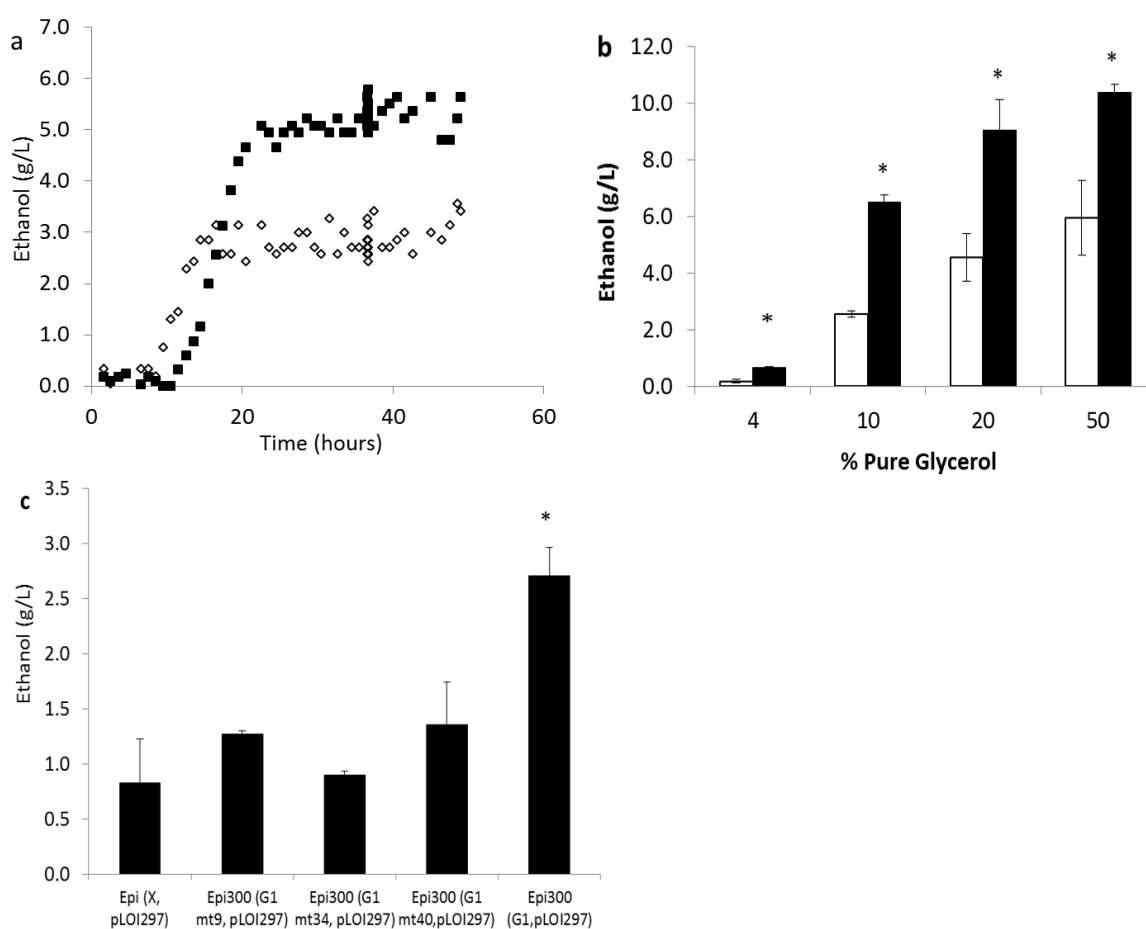
**Fig. 1** Growth (as OD<sub>620nm</sub>) in M9 4% (v/v) crude glycerol of Epi300 (G1) (♦), selected Tn5 mutants: mt9 (x), mt34 (■), mt40 (▲) and Epi300 negative control (□). Incubation at 37°C.



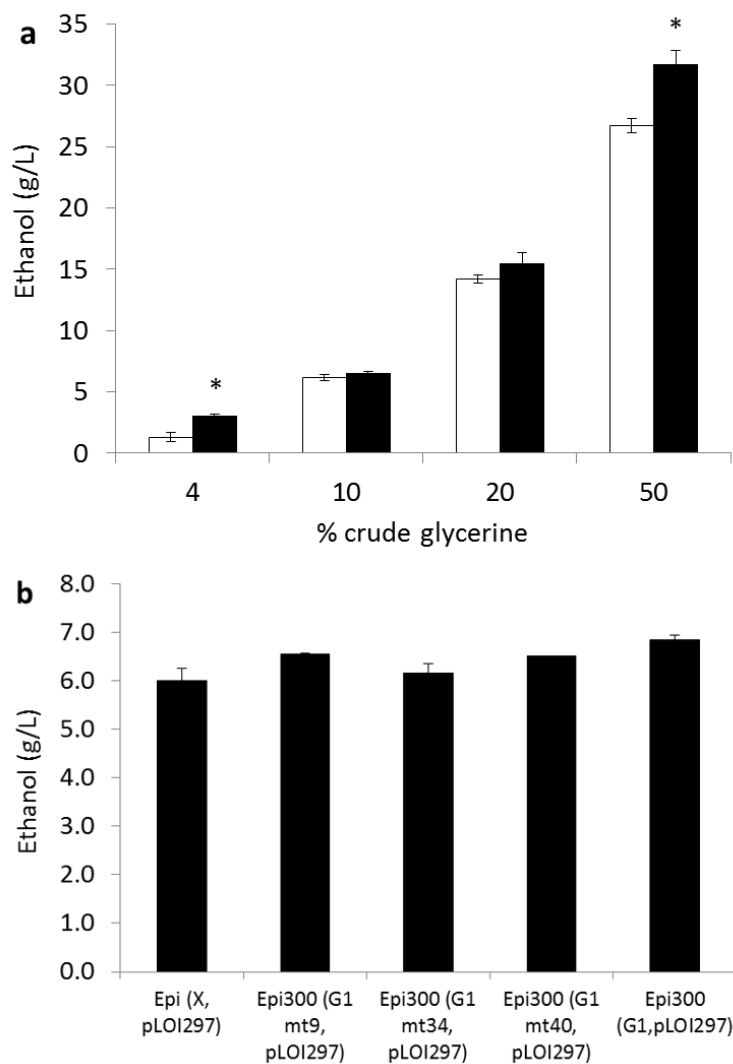
**Fig. 2** Relative position of transposon Tn5 insertion sites in G1 fosmid (41,312pb). The triangles indicates the insertion of each mutant in G1 sequence. Black triangles indicate mutations in ORF G1\_23, while white triangles indicate mutations outside this ORF. The ORF G1\_23 is indicate by an arrow and the pCC1FOS (12510-20519 pb) is showe in grey box.



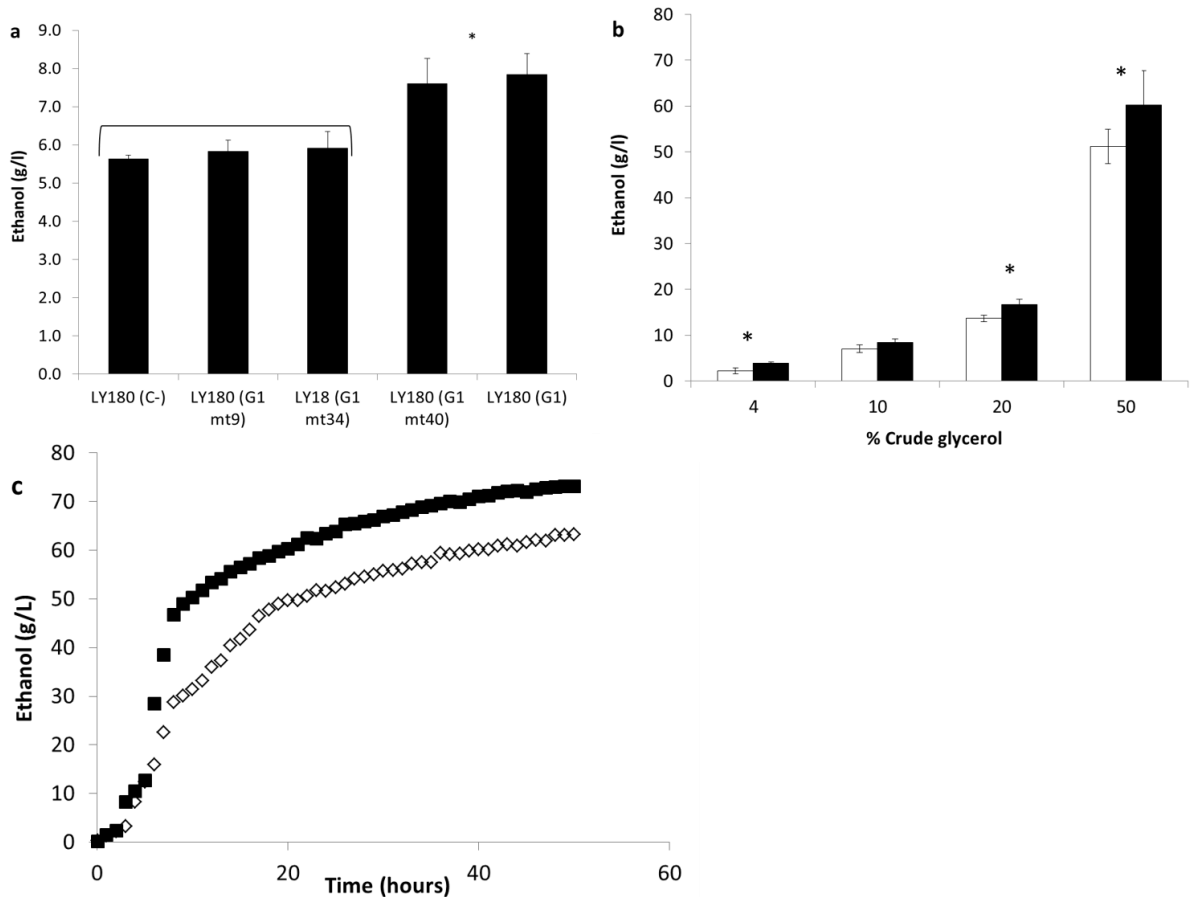
**Fig. 3** Phylogenetic relationship of ORF G1\_23 and its closest homologs identified using NCBI BLASTP. Multiple sequences alignments were computed with MEGA 6 with the Muscle method. Phylogenetic tree was constructed using the Maximum likelihood method implemented in MEGA 6.



**Fig. 4** Ethanol production in M9 minimum media with pure glycerol as carbon source. (a) Kinetic of ethanol production from 10% (v/v) crude glycerol during 50 hours, incubation at 37°C, 150 rpm by Epi300 (C-) (◇), Epi300(G1) (■); (b) ethanol production with different concentrations of pure glycerol added to media using Epi300 (negative control, pLOI297) (empty bars) or Epi300 (G1, pLOI297) (full bars); (c) Ethanol production from 10% (v/v) pure glycerol after 24 hours of incubation at 37°C, 150 rpm using Epi300 (pLOI297), Epi300 (G1, pLOI297) and selected Tn5 mutants. \* is present when statistical significant difference were found between strain with and without G1 fosmid.



**Fig. 5** Ethanol production in M9 minimum media with crude glycerol as carbon source by Epi300 strains. (a) ethanol production with different concentrations of crude glycerol added to media using Epi300 (negative control, pLOI297) (empty bars) or Epi300 (G1, pLOI297) (full bars); (b) Ethanol production from 10% (v/v) pure glycerol after 24 hours of incubation at 37°C, 150 rpm using Epi300 (pLOI297), Epi300 (G1, pLOI297) and selected Tn5 mutants. \* is present when statistical significant difference were found between strain with and without G1 fosmid.



**Fig. 6** Ethanol production in NBS minimum media with crude glycerol as carbon source. (a) Ethanol production from 10% (v/v) crude glycerol after 24 hours of incubation at 37°C, 150 rpm using ethanologenic strain LY180, LY180 (G1) and selected Tn5 mutants; (b) ethanol production using LY180 (G1) with different concentration of liquid fraction of crude glycerol added to media after neutralized using LY180 negative control (empty bars) or LY180 (G1) (full bars); (c) ethanol production kinetic in microaerobic conditions along 60 hours in 1L fermenter using 50% (v/v) crude glycerol-NBS by LY180 (◇) or LY180-G1 (■). \* is present when statistical significant difference were found between strain with and without G1 fosmid.

## Tables

**Table 1.** ORFs assigned to G1 sequence

#ORF	start	end	strand	length (aa)	Closest homolog using BLASTP (b)		E-value	Identity (%)	Gene Bank ID
					Description [Organisms]				
G1_1	1	255	-	85	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [ <i>Alistipes sp.</i> CAG:831]		1.00E-18	67%	<a href="#">WP_021928565.1</a>
G1_2	249	932	-	227	outer membrane protein beta-barrel domain protein [ <i>Alistipes sp.</i> CAG:831]		6.00E-38	39%	<a href="#">WP_021928564.1</a>
G1_3	929	1345	-	138	Lipoprotein [ <i>Bacteroidales bacterium</i> CF50]		2.00E-31	46%	<a href="#">AGY53676.1</a>
G1_4	1464	2723	-	419	putative uncharacterized protein [ <i>Bacteroides sp.</i> CAG:770]		2.00E-114	45%	<a href="#">WP_022285662.1</a>
G1_5	2720	3589	-	289	GTP pyrophosphokinase [ <i>Bacteroidales bacterium</i> CF50]		8.00E-78	47%	<a href="#">AGY53675.1</a>
G1_6	3568	4956	-	462	relA/SpoT family protein [ <i>Bacteroides sp.</i> CAG:545]		0	66%	<a href="#">WP_022017717.1</a>
G1_7	4966	5211	-	81	ribonuclease BN [ <i>Alistipes sp.</i> CAG:514]		4.00E-11	44%	<a href="#">WP_021839834.1</a>
G1_8	5232	5954	-	240	ribonuclease BN [ <i>Bacteroides sp.</i> CAG:545]		2.00E-45	37%	<a href="#">WP_022017716.1</a>
G1_9	5959	6954	-	331	NAD-dependent DNA ligase LigA [ <i>Alistipes sp.</i> CAG:831]		3.00E-144	65%	<a href="#">WP_021928558.1</a>
G1_10	6869	7987	-	372	NAD-dependent DNA ligase LigA [ <i>Alistipes sp.</i> CAG:831]		2.00E-153	64%	<a href="#">WP_021928558.1</a>
G1_11	8079	9311	+	410	6-phosphofructokinase [ <i>Alistipes timonensis</i> ]		0	67%	<a href="#">WP_010260254.1</a>
G1_12	9376	9615	+	79	alpha-amylase [ <i>Hymenobacter norwichensis</i> ]		0.007	34%	<a href="#">WP_022825614.1</a>
G1_13	9882	10928	+	348	alpha-amylase [ <i>Bacteroides barnesiae</i> ]		9.00E-97	48%	<a href="#">WP_018709950.1</a>
G1_14	11163	11564	+	133	MULTISPECIES: RNA polymerase subunit sigma-24 [ <i>Desulfitobacterium</i> ]		7.00E-04	40%	<a href="#">WP_014794121.1</a>
G1_15	11561	11779	+	72	RNA polymerase sigma factor, sigma-70 family [ <i>Gordonibacter pamelaee</i> ]		0.002	35%	<a href="#">WP_015538795.1</a>
G1_16	11837	12070	-	77	ADP-ribosylglycohydrolase family protein [ <i>Prevotella ruminicola</i> ]		1.00E-04	41%	<a href="#">WP_013063668.1</a>
G1_17	12034	12333	-	99	hypothetical protein [ <i>Prevotella sp.</i> FD3004]		1.00E-19	57%	<a href="#">WP_036911212.1</a>
G1_18	12926	13585	-	219	chloramphenicol acetyltransferase				
G1_19	13804	14151	+	115	resolvase				
G1_20	15544	16299	+	251	replication protein				
G1_21	17093	18052	+	319	plasmid-partitioning protein SopA				
G1_22	20459	20749	+	96	hypothetical protein BRDCF_p399 [ <i>Bacteroidales bacterium</i> CF50]		1.00E-22	58%	<a href="#">AGY53026.1</a>
<b>G1_23</b>	<b>21400</b>	<b>22734</b>	-	<b>444</b>	<b>hypothetical protein [<i>Prevotella ruminicola</i>]</b>		<b>8.00E-28</b>	<b>33%</b>	<a href="#">WP_013064512.1</a>
G1_24	23124	24230	+	368	transcriptional regulator [ <i>Prevotella fusca</i> ]		1.00E-44	30%	<a href="#">WP_025077699.1</a>

**Table 1, Cont.** ORFs assigned to G1 sequence

#ORF	start	end	strand	length (aa)	Closest homolog using BLASTP (b)		E-value	Identity (%)	Gene Bank ID
					Description [Organisms]				
G1_25	24227	24547	-	106	Phosphonate metabolism protein PhnP [uncultured bacterium]		2.00E-34	56%	<a href="#">EKD31793.1</a>
G1_26	24493	24999	-	168	beta-lactamase domain-containing protein [ <i>Bacteroides</i> sp. CAG:545]		2.00E-62	66%	<a href="#">WP_022015881.1</a>
G1_27	25161	26423	-	420	Fe-S oxidoreductase [ <i>Alistipes indistinctus</i> ]		5.00E-179	59%	<a href="#">WP_009134945.1</a>
G1_28	26416	26763	-	115	dihydroneopterin aldolase [ <i>Alistipes</i> sp. CAG:435]		2.00E-32	46%	<a href="#">WP_022324543.1</a>
G1_29	26872	27588	-	238	uroporphyrin-III C/tetrapyrrole methyltransferase [ <i>Alistipes</i> sp. CAG:435]		3.00E-93	57%	<a href="#">WP_022324546.1</a>
G1_30	27530	29476	-	648	hypothetical protein BRDCF_p1264 [ <i>Bacteroidales bacterium</i> CF50]		9.00E-179	44%	<a href="#">AGY53891.1</a>
G1_31	29632	31986	-	784	hypothetical protein BRDCF_p1264 [ <i>Bacteroidales bacterium</i> CF50]		4.00E-55	25%	<a href="#">AGY53891.1</a>
G1_32	32050	33078	+	342	tRNA threonylcarbamoyladenine modification protein TsaD [ <i>Alistipes</i> ]		0	73%	<a href="#">WP_004328979.1</a>
G1_33	33176	34621	+	481	oxidoreductase [ <i>Bacteroides</i> sp. CAG:770]		0	62%	<a href="#">WP_022285915.1</a>
G1_34	35079	36149	+	356	putative PD-(D/E)XK nuclease family protein [ <i>Elizabethkingia anophelis</i> ]		4.00E-24	28%	<a href="#">CDN76750.1</a>
G1_35	36306	36833	-	175	peptidyl-prolyl cis-trans isomerase [ <i>Alistipes</i> sp. CAG:831]		4.00E-72	70%	<a href="#">WP_021927072.1</a>
G1_36	36830	37420	-	196	uncharacterized protein [ <i>Alistipes</i> sp. CAG:831]		1.00E-66	54%	<a href="#">WP_021927077.1</a>
G1_37	37503	37754	-	83	50S ribosomal protein L31 [ <i>Alistipes</i> sp. CAG:831]		4.00E-47	87%	<a href="#">WP_021927126.1</a>
G1_38	37823	38455	-	210	hypothetical protein [ <i>Bacteroides fragilis</i> ]		4.00E-28	36%	<a href="#">WP_032568461.1</a>
G1_39	38961	39902	+	313	phosphomannomutase [ <i>Bacteroides</i> sp. CAG:770]		5.00E-137	63%	<a href="#">AGY53846.1</a>
G1_40	39937	40383	+	148	hypothetical protein BRDCF_p1220 [ <i>Bacteroidales bacterium</i> CF50]		1.00E-43	52%	<a href="#">AGY53847.1</a>
G1_41	40395	41036	+	213	hypothetical protein BRDCF_p1220 [ <i>Bacteroidales bacterium</i> CF50]		1.00E-71	47%	<a href="#">AGY53847.1</a>
G1_42	41040	41312	-	90	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [ <i>Alistipes</i> sp. CAG:831]		1.00E-40	74%	<a href="#">WP_021928565.1</a>

<sup>(a)</sup> ORFs identified by MetaGene.

<sup>(b)</sup> Assigned by RPS BLAST.

<sup>(c)</sup> Closest homolog identified by BLASTP 2.2.30+ using non redundant protein sequences.

<sup>(d)</sup> Gene interrupted by Tn5 is showed in bold. Genes belong to pCC1FOS are in red.

**Table 2.** Closest homologs of ORF G1\_23 identified using NCBI BLASTP

Description [Origin]	E-value	Identity (%)	GenBank ID	Reference
MULTISPECIES: hypothetical protein [ <i>Bacteroides</i> ]	1.00E-28	38%	<a href="#">WP_031540335.1</a>	
hypothetical protein [ <i>Prevotella ruminicola</i> ]	4.00E-28	32%	<a href="#">WP_041385930.1</a>	
hypothetical protein PRU_1480 [ <i>Prevotella ruminicola</i> 23]	1.00E-27	33%	<a href="#">ADE82526.1</a>	(Purushe, Fouts et al. 2010)
hypothetical protein [ <i>Prevotella sp.</i> RM4]	4.00E-26	31%	<a href="#">WP_033149883.1</a>	
hypothetical protein [ <i>Prevotella sp.</i> MA2016]	1.00E-19	37%	<a href="#">WP_028913135.1</a>	

## References

- Alterthum, F. and L. O. Ingram (1989). "Efficient Ethanol Production from Glucose, Lactose, and Xylose by Recombinant *Escherichia coli*." Appl Environmental Microbiology **55**(8): 1943-1948.
- Bertani, G. (1951). "STUDIES ON LYSOGENESIS I: The Mode of Phage Liberation by Lysogenic *Escherichia coli*." Journal of bacteriology **62**(3): 293.
- Cintolesi, A., J. M. Clomburg, V. Rigou, K. Zygorakis and R. Gonzalez (2012). "Quantitative analysis of the fermentative metabolism of glycerol in *Escherichia coli*." Biotechnol Bioeng **109**(1): 187-198.
- Ciriminna, R., C. D. Pina, M. Rossi and M. Pagliaro (2014). "Understanding the glycerol market." European Journal of Lipid Science and Technology **116**(10): 1432-1439.
- Clomburg, J. M. and R. Gonzalez (2013). "Anaerobic fermentation of glycerol: a platform for renewable fuels and chemicals." Trends Biotechnol **31**(1): 20-28.
- Dharmadi, Y., A. Murarka and R. Gonzalez (2006). "Anaerobic fermentation of glycerol by *Escherichia coli*: a new platform for metabolic engineering." Biotechnol Bioeng **94**(5): 821-829.
- Durnin, G., J. Clomburg, Z. Yeates, P. J. Alvarez, K. Zygorakis, P. Campbell and R. Gonzalez (2009). "Understanding and harnessing the microaerobic metabolism of glycerol in *Escherichia coli*." Biotechnol Bioeng **103**(1): 148-161.
- Finn, R. D., J. Clements and S. R. Eddy (2011). "HMMER web server: interactive sequence similarity searching." Nucleic acids research: gkr367.
- Gonzalez, R., A. Murarka, Y. Dharmadi and S. S. Yazdani (2008). "A new model for the anaerobic fermentation of glycerol in enteric bacteria: trunk and auxiliary pathways in *Escherichia coli*." Metab Eng **10**(5): 234-245.
- Gray, K. A., L. Zhao and M. Emptage (2006). "Bioethanol." Current opinion in chemical biology **10**(2): 141-146.
- Hahn-Hägerdal, B., K. Karhumaa, M. Jeppsson and M. Gorwa-Grauslund (2007). Metabolic Engineering for Pentose Utilization, Springer.
- Hu, H. and T. K. Wood (2010). "An evolved *Escherichia coli* strain for producing hydrogen and ethanol from glycerol." Biochem Biophys Res Commun **391**(1): 1033-1038.
- Ingram, L. O., T. Conway and F. Alterthum (1991). Ethanol production by *Escherichia coli* strains co-expressing *Zymomonas* PDC and ADH genes, US Patent 5000000.
- Jarboe, L., T. Grabar, L. Yomano, K. Shanmugan and L. Ingram (2007). Development of ethanologenic bacteria. Biofuels, Springer: 237-261.
- Johnson, D. T. and K. A. Taconi (2007). "The glycerin glut: Options for the value-added conversion of crude glycerol resulting from biodiesel production." Environmental Progress **26**(4): 338-348.
- Krogh, A., B. Larsson, G. Von Heijne and E. L. Sonnhammer (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." Journal of molecular biology **305**(3): 567-580.
- Lai, E. (2014). "Biodiesel: Environmental Friendly Alternative to Petrodiesel." J Pet Environ Biotechnol **5**: e122.
- Lee, S. J., S. B. Kim, S. W. Kang, S. O. Han, C. Park and S. W. Kim (2012). "Effect of crude glycerol-derived inhibitors on ethanol production by *Enterobacter aerogenes*." Bioprocess Biosyst Eng **35**(1-2): 85-92.
- Martínez-Gómez, K., N. Flores, H. M. Castañeda, G. Martínez-Batallar, G. Hernández-Chávez, O. T. Ramírez, G. Gosset, S. Encarnación and F. Bolívar (2012). "New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol." Microb Cell Fact **11**: 46.
- Matthew S. Wong, M. L., Ryan W. Black, Thao Q. Le, Sharon Puthli, Paul Campbell, Daniel J. Monticello (2014). "Microaerobic Conversion of Glycerol to Ethanol in *Escherichia coli*." Appl Environmental Microbiology **80**(10): 3276-3282.
- Meiswinkel, T. M., D. Rittmann, S. N. Lindner and V. F. Wendisch (2013). "Crude glycerol-based production of amino acids and putrescine by *Corynebacterium glutamicum*." Bioresour Technol **145**: 254-258.
- Miller, E. N., L. R. Jarboe, L. P. Yomano, S. W. York, K. T. Shanmugam and L. O. Ingram (2009). "Silencing of NADPH-dependent oxidoreductase genes (*yqhD* and *dkgA*) in furfural-resistant ethanologenic *Escherichia coli*." Appl Environ Microbiol **75**(13): 4315-4323.



- Murarka, A., Y. Dharmadi, S. S. Yazdani and R. Gonzalez (2008). "Fermentative utilization of glycerol by *Escherichia coli* and its implications for the production of fuels and chemicals." *Appl Environ Microbiol* **74**(4): 1124-1135.
- Nancy, Y. Y., J. R. Wagner, M. R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S. C. Sahinalp, M. Ester and L. J. Foster (2010). "PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes." *Bioinformatics* **26**(13): 1608-1615.
- Nielsen, H., J. Engelbrecht, S. Brunak and G. von Heijne (1997). "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." *Protein engineering* **10**(1): 1-6.
- Noguchi, H., J. Park and T. Takagi (2006). "MetaGene: prokaryotic gene finding from environmental genome shotgun sequences." *Nucleic Acids Res* **34**(19): 5623-5630.
- Papanikolaou, S., M. Fick and G. Aggelis (2004). "The effect of raw glycerol concentration on the production of 1,3-propanediol by *Clostridium butyricum*." *Journal of Chemical Technology & Biotechnology* **79**(11): 1189-1196.
- Purushe, J., D. E. Fouts, M. Morrison, B. A. White, R. I. Mackie, P. M. Coutinho, B. Henrissat and K. E. Nelson (2010). "Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their environmental niche." *Microb Ecol* **60**(4): 721-729.
- Sarma, S. J., S. K. Brar, E. B. Sydney, Y. Le Bihan, G. Buelna and C. R. Soccol (2012). "Microbial hydrogen production by bioconversion of crude glycerol: A review." *International Journal of Hydrogen Energy* **37**(8): 6473-6490.
- Sawers, R., M. Blokesch and A. Böck (2004). "Anaerobic formate and hydrogen metabolism." *EcoSal-Escherichia coli and Salmonella: Cellular and Molecular Biology*.
- Senthilkumar, V. and P. Gunasekaran (2005). "Bioethanol production from cellulosic substrates: Engineered bacteria and process integration challenges." *Journal of Scientific & Industrial Research* **64**: 845-853.
- Tamura, K., G. Stecher, D. Peterson, A. Filipinski and S. Kumar (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." *Mol Biol Evol* **30**(12): 2725-2729.
- Varrone, C., B. Giussani, G. Izzo, G. Massini, A. Marone, A. Signorini and A. Wang (2012). "Statistical optimization of biohydrogen and ethanol production from crude glycerol by microbial mixed culture." *International Journal of Hydrogen Energy* **37**(21): 16479-16488.
- Yang, F., M. A. Hanna and R. Sun (2012). "Value-added uses for crude glycerol--a byproduct of biodiesel production." *Biotechnol Biofuels* **5**: 13.
- Yomano, L., S. York and L. Ingram (1998). "Isolation and characterization of ethanol-tolerant mutants of *Escherichia coli* KO11 for fuel ethanol production." *Journal of Industrial Microbiology and Biotechnology* **20**(2): 132-138.
- Zhang, X., K. T. Shanmugam and L. O. Ingram (2010). "Fermentation of Glycerol to Succinate by Metabolically Engineered Strains of *Escherichia coli*." *Applied and Environmental Microbiology* **76**(8): 2397-2401.
- Zhang, Z., S. Schwartz, L. Wagner and W. Miller (2000). "A Greedy Algorithm for Aligning DNA Sequences." *Journal of Computational Biology* **7**(1-2): 203-214.

# **DISCUSIÓN GENERAL, CONCLUSIONES Y PERSPECTIVAS**

---

## Discusión general y conclusiones

La actual demanda de procesos amigables con el medio ambiente ha propiciado el desarrollo de alternativas renovables para la obtención de productos químicos y combustibles (144,236). La búsqueda de catalizadores biológicos capaces de satisfacer las necesidades actuales de la industria acompaña esta tendencia, propiciando la utilización de microorganismos y enzimas en diversos procesos.

Los biocombustibles son la alternativa actual para los carburantes derivados del petróleo, habiéndose producido en 2012 1.470.000 barriles de bioetanol y 430.000 de biodiesel por día a nivel mundial (<http://www.eia.gov/>, 1 barril son 159 litros). Pero su producción, basada principalmente en la utilización de los azúcares derivados de la caña de azúcar o el maíz y los aceites de la soja o la palma, compite directamente con la producción de alimentos (26), causa altos niveles de erosión y consume altos niveles de fertilizante nitrogenado (6). Una de las materias primas alternativas para la producción de bioetanol es la lignocelulosa, la cual es ubicua, abundante y su obtención no representa impedimentos para la producción de alimentos (23). Sin embargo, la lignocelulosa es un material recalcitrante que requiere altos costos de capital y manejo para su procesamiento (86). El desarrollo de pretratamientos más eficientes, que generen menor cantidad de inhibidores y contaminantes ambientales (77,237), cepas etanológicas capaces de fermentar tanto pentosas como hexosas (32) y enzimas robustas, aptas para la sacarificación de celulosa en condiciones industriales y con menor costo de producción (238) son necesarios para viabilizar la producción de etanol de segunda generación a nivel industrial.

Debido a que la mayoría de los organismos no pueden ser cultivados en el laboratorio, los enfoques independientes de cultivo, como la metagenómica, se han aplicado desde hace más de una década para acceder y caracterizar la diversidad filogenética en diversas comunidades microbianas, así como su potencial metabólico y las funciones ecológicas de las mismas (133). Es por esto que la metagenómica es considerada una fuente potencial a catalizadores novedosos (146,147).

Aplicando esta metodología, en este trabajo de tesis se detectaron 49 clones con actividad positiva para celulasa o xilanasas en una biblioteca obtenida a partir de líquido ruminal bovino, un ambiente rico en enzimas novedosas capaces de hidrolizar biomasa lignocelulósica (153,174,175), a las cuales es posible acceder mediante aproximaciones de metagenómica funcional (122) utilizando *Escherichia coli* como hospedero.

De los 49 clones, ocho fueron seleccionados para profundizar en su caracterización fenotípica y sus fósidos fueron secuenciados. Al ser evaluados en carboximetil-celulosa, 4-MUC, xilano, avicel, papel de filtro y bagazo de caña de azúcar, 6 mostraron actividad en al menos dos sustratos, lo cual permite inferir que son capaces de expresar enzimas multifuncionales o varias enzimas involucradas en la degradación de biomasa. La expresión del fósido Csd4 en la cepa etanológica MS04 mostró que

es posible la generación de etanol a partir de CMC y que su complementación con  $\beta$ -glucosidasa mejora tanto el crecimiento como la producción de etanol en medio mínimo.

Mediante el análisis de la secuencia de los fósmidos de estos clones se puede inferir que éstos provendrían de bacterias del género *Bacteroidetes*, aunque no existe identidad con los genomas depositados en las bases de datos. Esto confirma que la metagenómica es una herramienta adecuada para acceder a genomas de organismos no cultivados previamente (134). Fortaleciendo este concepto, se corroboró que los más de 10 genes que codifican para diferentes enzimas implicadas en la degradación de celulosa o hemicelulosa identificados, mostraron homología con glicosil hidrolasas no caracterizadas previamente. Pero de las 133 familias GH descritas hasta el momento, sólo se obtuvieron representantes de las familias GH 5, 26, 43, 45 y 130.

Los sesgos introducidos con el método de lisis celular, la extracción de ADN y al utilizar a *E. coli* como hospedero, seguramente fueron los determinantes de la baja diversidad de familias GH obtenidas. A su vez, el haber construido la biblioteca metagenómica únicamente a partir de líquido ruminal condicionó el acceso a la diversidad filogenética y funcional, ya que la microbiota del rumen asociada a las fibras de alimento (es decir, a la biomasa lignocelulósica) y la presente en la fracción líquida son diferentes (175).

El aplicar diversos métodos de lisis celular y emplear múltiples hospederos aseguraría un mayor acceso a la diversidad genética presente en el rumen bovino, pero esto representaría un aumento significativo en la cantidad de trabajo, principalmente en las etapas de selección y caracterización de los clones.

Por su parte, a partir de un digestor anaerobio, habitado por una comunidad filogenéticamente más alejada de *E. coli*, se obtuvieron un menor número de clones positivos. De todos modos, uno de los fósmidos obtenidos, denominado M9F1, resultó ser especialmente interesante debido a las actividades conferidas al hospedero y la nula identidad de secuencia mostrada con glucosidasas depositadas en las bases de datos. La complementación de la cepa de *E. coli* etanologénica LY180 con este fósmino permitió la obtención de etanol a partir de papel de filtro en medio mínimo.

El ORF codificante para la endoglucanasa presente en el fósmino Csd4, fue clonado en el vector pET28a y sobreexpresado en la cepa BL21. La enzima codificada, EndoG, resultó ser una celulasa multifuncional con cualidades interesantes que la hacen prometedora para su utilización en procesos industriales, como la tolerancia a altas concentraciones de sal, diversos solventes orgánicos y líquidos iónicos. Esta enzima, que no es inhibida por los productos finales de la degradación de celulosa, es estable a altas temperaturas y es activa en un rango relativamente amplio de pH y temperatura. A su vez, la expresión de EndoG en la cepa MS04 permitió la producción de etanol a partir de CMC o avicel en un proceso de un único paso, lo cual representa el principal objetivo de esta tesis de doctorado.

La aplicación de biorrefinerías se ha propuesto como un medio para aumentar la viabilidad económica de la industria de los biocombustibles (6,30). En su forma "convencional" una biorrefinería haría uso de una fracción de la materia prima (por ejemplo, una porción de azúcares o aceites) para obtener un co-producto de mayor valor, junto con el biocombustible. Un modelo económicamente más viable para una biorrefinería, sin embargo, debe considerar el uso de subproductos o de residuos generados durante la producción del biocombustible (5). Por ejemplo, las fracciones ricas en glicerol generadas por la industria del biodiesel tienen el potencial de ser utilizado como materia prima para la producción de otros productos (239).

Con el fin de identificar elementos genéticos que promuevan la utilización de glicerina cruda por *E. coli* se evaluó la biblioteca del digestor anaerobio en medio mínimo sólido suplementado con este sub-producto generado por ALUR. Es así como se seleccionó el fósmino G1, que confiere a *E. coli* la habilidad de crecer más rápido en este medio. A su vez, fue posible la obtención de etanol a partir de la glicerina cruda neutralizada utilizando la cepa Epi300-T1<sup>R</sup> complementada con el operón pET y la cepa etanológica LY180. El análisis de la secuencia del fósmino G1 no reveló la presencia de genes implicados en la metabolización del glicerol, pero la mutagénesis *in vitro* del mismo mostró que el ORF G1\_23 podría estar implicado en el fenotipo observado. La proteína codificada por este ORF posee menos de 40% de identidad de secuencia con proteínas hipotéticas, por lo cual no es posible asignar una función.

Dado el bajo número de clones positivos obtenidos en esta parte del trabajo, surge cuestionar si la metodología empleada fue la adecuada. La selección de la muestra ambiental a partir de la cual se contruye una biblioteca metagenómica es un factor clave para la obtención de clones positivos para la actividad de interés. Aunque se cree que esta comunidad está especializada en la degradación de desechos industriales, en presencia de posibles inhibidores que puedan generarse en el proceso, no existe evidencia que permita suponer que en este ambiente existirían microorganismos especializados en la utilización de glicerol o tolerantes a los inhibidores presentes en la glicerina cruda (metanol, hidróxido de sodio, aluminio, entre otros).

Una vez más parece evidente que la comunidad microbiana presente en el reactor anaerobio es filogenéticamente distante a *E. coli*, lo cual condiciona la expresión correcta de las enzimas codificadas en su metagenoma. La utilización de un microorganismo filogenéticamente más relacionado a esta comunidad como hospedero seguramente aumentaría la detección de clones positivos.

El método de screening tal vez no fue el adecuado teniendo en cuenta que el objetivo final era la utilización de este sustrato en medio líquido y en condiciones microaerofílicas para la producción de etanol. Una alternativa podría ser realizar la selección de clones en estas condiciones (medio líquido, glicerina neutralizada, incubación sin oxígeno). Otra opción sería el enriquecimiento previo de la

biblioteca en clones capaces de crecer en estas condiciones para luego proceder con la selección en placas. Esta estrategia suele ser más efectiva si el enriquecimiento se hace con la comunidad de partida (en este caso el lodo del reactor anaerobio) y luego se procede a la extracción y clonado del ADN metagenómico (240). Aunque más compleja, otra alternativa sería evaluar directamente la producción de etanol a partir de la glicerina cruda a cada uno de los clones. Para esto es necesario transferir todos los fósmidos a una cepa etanológica o construir directamente la biblioteca en un hospedero con esta característica.

A nivel industrial se considera que los títulos de etanol deben superar los 40 g/L para considerarse un proceso viable. En este trabajo de Tesis, los rendimientos obtenidos a partir de celulosa están lejos de los esperados, pero se logró triplicar los rendimientos de etanol luego de la sacarificación con celulasas comerciales cuando la cepa etanológica contiene al fósrido Csd4 (ver capítulo 3). A nivel de laboratorio los rendimientos obtenidos son similares a varios trabajos reportados en la literatura. La expresión de una endoglucanasa (Cel5A), una exoglucanasa (Cel9E) y una  $\beta$ -glucosidasa en la cepa de *E. coli* LY01 permitió la producción de 3.59 g/L a partir de 10 g/L de PASC (110). Con mejores resultados, Vasan y colaboradores (108) reportaron la expresión de una endoglucanasa de *Enterobacter cloacae* en *Zymomonas mobilis*, la cual produce 5% y 4% de etanol a partir de carboximetil-celulosa y bagaso pretratado con NaOH, respectivamente. Por otro lado, los intentos de conseguir cepas celulolíticas de la levadura *Saccharomyces cerevisiae*, el microorganismo más utilizado para la producción de etanol, no ha logrado alcanzar los rendimientos esperados. Por ejemplo, la co-expresión de varias celulasas y un transportador de celodextrinas en *S. cerevisiae* determina un máximo de 4,3 g/L de etanol (241). Cuando *S. cerevisiae* expresa los genes codificantes para una endoglucanasa de *Trichoderma reesei* (EGI) y una  $\beta$ -glucosidasa de *Saccharomycopsis fibuligera* (BGL1), produce 1 g/L de etanol a partir de 10 g/L de PASC (242).

## Perspectivas

En esta tesis se identificaron genes novedosos que codifican para celulasas y hemicelulasas no caracterizadas previamente, surge como primer actividad adicional clonar, purificar y caracterizar bioquímicamente las enzimas identificadas en los fósridos Csd6, Csd8, Csd9 y Xil3.

A su vez, en dos de los fósridos secuenciados, M9F1 y Csd23, no se detectaron glicosil hidrolasas, por lo cual sería muy interesante identificar el o los genes responsables de los fenotipos observados presentes en estos fósridos. Ésto se intentó mediante mutagénesis *in vitro* y posterior evaluación de los mutantes obtenidos, pero esta estrategia no arrojó resultados positivos. Una alternativa es el sub-clonado de los ORFs que se presumen implicados en los fenotipos, o sea, las glicosiltransferasas en M9F1 y los que codifican para proteínas hipotéticas en Csd23, y la posterior evaluación de los clones obtenidos. La limitante en este caso es que si el fenotipo se debe a la co-

expresión de múltiples enzimas no se obtendrían clones capaces de reproducir el fenotipo original. Otra opción es fragmentar los fósmidos, clonar los fragmentos de ADN obtenidos en otro vector y transferirlos a nuevos hospederos para evaluar si es posible obtener los fenotipos originales conferidos por los fósmidos parentales.

En esta tesis, fue posible la producción de etanol a partir de CMC o avicel mediante la cepa MS04/pTrc99A-EndoG, pero estos no son sustratos aptos para la producción de etanol a nivel industrial, por lo que resta evaluar la producción de etanol a partir de desechos agropecuarios, tales como bagazo de caña o rastrojo de trigo o arroz, o cultivos energéticos como Switchgrass o caña común. A su vez, la complementación con las otras enzimas detectadas, ya sean celulasas o xilanasas, permitirá determinar si existe sinergia entre estas enzimas. El desarrollo de un operón que confiera la capacidad de degradar lignocelulosa surge como una estrategia prometedora para la producción de biocombustibles y productos con alto valor agregado a partir de biomasa vegetal. Para que su implementación sea exitosa, es necesario no solo seleccionar los genes codificantes para celulasas y hemicelulasas si no también los promotores implicados en su expresión y el vector, el cual idealmente debería ser compatible con diversos microorganismos.

Por otro lado, el fósrido G1 permitió la producción de etanol a partir de glicerina cruda. Falta confirmar si el ORF G1\_23 es el responsable de la capacidad conferida a *E. coli* o si se trata de otro ORF y en qué proceso está implicado. A su vez, una vez confirmado el gen responsable, sería interesante evaluar si su sobre-expresión mejora los rendimientos de etanol obtenidos o si es posible su expresión en cepas que produzcan otros productos con mayor valor agregado, como por ejemplo 1,3-propanodiol, ácido D-láctico, succinato, ácido propiónico y 3-poli-hidroxibutirato, los cuales mejorarían la ecuación económica de la producción de biodiesel.

## **REFERENCIAS BIBLIOGRÁFICAS**

---



1. Naik SN, Goud V V., Rout PK, Dalai AK. Production of first and second generation biofuels: A comprehensive review. *Renew Sustain Energy Rev.* 2010;14(2):578–97.
2. Otero J, Panagiotou G, Olsson L. Fueling Industrial Biotechnology Growth with Bioethanol. In: Scheper T, Olsson L, editors. *Advances in Biochemical Engineering / Biotechnology.* 2007.
3. Jyoti S, Kaur S, Bittencourt E, Le Y, Buelna G, Ricardo C. Microbial hydrogen production by bioconversion of crude glycerol : A review. *Int J Hydrogen Energy.* Elsevier Ltd; 2012;37(8):6473–90.
4. Moser BR. Biodiesel production, properties, and feedstocks. *Biofuels.* Springer; 2011. p. 285–347.
5. Yazdani SS, Gonzalez R. Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. *Curr Opin Biotechnol.* 2007;18(3):213–9.
6. Balat M, Balat H, Oz C. Progress in bioethanol processing. *Prog Energy Combust Sci.* 2008;34:551–73.
7. Lynd LR, Weimer PJ, Zyl WH Van, Isak S, Pretorius IS. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiol Mol Biol Rev.* 2002;66(3):506–739.
8. Caspeta L, Nielsen J. Economic and environmental impacts of microbial biodiesel. *Nat Biotechnol.* 2013;31(9):789–93.
9. Songstad D, Lakshmanan P, Chen J, Gibbons W, Hughes S, Nelson R. Historical Perspective of Biofuels: Learning from the Past to Rediscover the Future. In: Tomes D, Lakshmanan P, Songstad D, editors. *Biofuels.* Springer; 2011.
10. Álvarez C. Biocombustibles: desarrollo histórico-tecnológico, mercados actuales y comercio internacional. *Econ Inf.* 2009;(359):63–89.
11. Aspinall N, Mills L, Strahan D, Boyle R, Cuming V, Stopforth K, et al. *Global Trends in Renewable Energy.* Frankfurt; 2014.
12. Lins C. *Renewables 2014: Global Status Report - Key Findings.* 2014.
13. Dirección Nacional de Energía. *Balance energético nacional 2013.* Montevideo; 2013.
14. Dirección Nacional de Energía. *Balance energético nacional 2012.* Montevideo; 2012.
15. Bittencourt G, Reig Lorenzi N. La industria de Biocombustibles en Uruguay, situación actual y perspectivas. Documentos de Trabajo No 11/09. Montevideo: Departamento de Economía, Facultad de Ciencias Sociales, Universidad de la república.; 2009.
16. Olivet B. *INFORME : MEDIO AMBIENTE Y ENERGÍA EN URUGUAY.* Montevideo; 2014.
17. ALUR. *Actualización Biodiesel.* 2014.
18. Bittencourt G, López A, Reig Lorenzi N, Rodríguez R, Rozemberg R, Saslavsky D, et al. *La industria de Biocombustibles en el Mercosur.* Montevideo; 2009.

19. Cremonez P, Feroldi M, Feiden A, Gustavo J, José D, Dieter J, et al. Current scenario and prospects of use of liquid biofuels in South America. *Renew Sustain Energy Rev. Elsevier*; 2015;43:352–62.
20. Hill J, Nelson E, Tilman D, Polasky S, Tiffany D. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc Natl Acad Sci U S A*. 2006;103(30):11206–10.
21. Richard TL. Challenges in scaling up biofuels infrastructure. *Science (80- ) [Internet]*. 2010;329(5993):793–6.
22. Scharlemann JPW, Laurance WF. Environmental science. How green are biofuels? *Science*. 2008;319:43–4.
23. Tilman D, Socolow R, Foley J a, Hill J, Larson E, Lynd L, et al. Beneficial Biofuels—The Food, Energy, and Environment Trilemma. *Science*. 2009;325(5938):4–5.
24. Stephen JD, Mabee WE, Saddler JN. Will second-generation ethanol be able to compete with first-generation ethanol? Opportunities for cost reduction. *Biofuels, Bioprod Biorefining. John Wiley & Sons, Ltd.*; 2012 Mar 1;6(2):159–76.
25. Zah R, Böni H, Gauch M, Hischier R, Lehmann M, Wäger P. Ökobilanz von Energieprodukten: Ökologische Bewertung von Biotreibstoffen. Schlussbericht. Abteilung Technologie und Gesellschaft, Empa im Auftrag des Bundesamtes für Energie, des Bundesamtes für Umwelt und des Bundesamtes für Landwirtschaft, Bern. 2007.
26. Naylor RL, Liska AJ, Burke MB, Falcon WP, Gaskell JC, Rozelle SD, et al. The Ripple Effect: Biofuels, Food Security, and the Environment. *Environ Sci Policy Sustain Dev*. 2007;49(9):30–43.
27. Havlík P, Schneider U a., Schmid E, Böttcher H, Fritz S, Skalský R, et al. Global land-use implications of first and second generation biofuel targets. *Energy Policy*. 2011;39(10):5690–702.
28. Sarkar N, Ghosh SK, Bannerjee S, Aikat K. Bioethanol production from agricultural wastes : An overview. *Renew Energy. Elsevier Ltd*; 2012;37(1):19–27.
29. Humbrid D, Davis R, Tao L, Kinchin C, Hsu D, Aden A. Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol. NREL Technial Rep. 2011.
30. Menon V, Rao M. Trends in bioconversion of lignocellulose : Biofuels, platform chemicals & biorefinery concept. *Prog Energy Combust Sci. Elsevier Ltd*; 2012;38(4):522–50.
31. Kamm B, Kamm M. Principles of biorefineries. *Appl Microbiol Biotechnol*. 2004;64(2):137–45.
32. Geddes CC, Nieves IU, Ingram LO. Advances in ethanol production. *Curr Opin Biotechnol [Internet]. Elsevier Ltd*; 2011;22(3):312–9.
33. Johnson DT, Taconi KA. The Glycerin Glut : Options for the Value-Added Conversion of Crude Glycerol Resulting from Biodiesel Production. *wiInterScience*. 2009;26(4):338–48.

34. Meiswinkel TM, Rittmann D, Lindner SN, Wendisch VF. Crude glycerol-based production of amino acids and putrescine by *Corynebacterium glutamicum*. *Bioresour Technol*. Elsevier Ltd; 2013;145:254–8. 35. Walker JM. *Biofuels : Methods and Protocols*. 2009.
36. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* [Internet]. *Biotechnology for Biofuels*; 2013;6(1):41.
37. Scheller HV, Ulvskov P. Hemicelluloses. *Plant Biol*. 2010;61(1):263.
38. Gírio FM, Fonseca C, Carvalheiro F, Duarte LC, Marques S. Bioresource Technology Hemicelluloses for fuel ethanol : A review. *Bioresour Technol*. Elsevier Ltd; 2010;101(13):4775–800.
39. Shallom D, Shoham Y. Microbial hemicellulases. *Curr Opin Microbiol*. 2003;6(3):219–28.
40. Gellerstedt G, Henriksson G. Lignins: Major Sources, Structure and Properties. In: Belgacem MN, Gandini A, editors. *Monomers, Polymers and Composites*. 2008th ed. Elsevier; 2008. p. 39–66.
41. Liu CJ, Miao YC, Zhang KW. Sequestration and transport of lignin monomeric precursors. *Molecules*. 2011;16(1):710–27.
42. Buxton D, Redfearn D. Plant Limitations to Fiber Digestion and Utilization. *New Dev forage Sci Contrib to Enhanc fiber Util by Rumin*. 1997;127(5):810–3.
43. Grabber JH. How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. *Crop Sci*. 2005;45(3):820–31.
44. Glazer AN, Nikaido H. *Microbial biotechnology: fundamentals of applied microbiology*. Cambridge University Press; 2007.
45. Brown Jr RM, Saxena IM. Cellulose biosynthesis: a model for understanding the assembly of biopolymers. *Plant Physiol Biochem*. Elsevier; 2000;38(1):57–67.
46. Festucci-Buselli R a., Otoni WC, Joshi CP. Structure, organization, and functions of cellulose synthase complexes in higher plants. *Brazilian J Plant Physiol*. 2007;19(1):1–13.
47. Atalla RH, Hackney JM, Uhlin I, Thompson NS. Hemicelluloses as structure regulators in the aggregation of native cellulose. *Int J Biol Macromol*. 1993;15(2):109–12.
48. Hilden L, Johansson G. Recent developments on cellulases and carbohydrate-binding modules with cellulose affinity. *Biotechnol Lett*. Springer; 2004;26(22):1683–93.
49. Koyama M, Helbert W, Imai T, Sugiyama J, Henrissat B. Parallel-up structure evidences the molecular directionality during biosynthesis of bacterial cellulose. *Proc Natl Acad Sci U S A*. 1997;94(17):9091–5.
50. Sampedro J, Cosgrove DJ. The expansin superfamily. *Genome Biol*. 2005;6(12):242.
51. Quiroz-Castañeda RE, Martínez-Anaya C, Cuervo-Soto LI, Segovia L, Folch-Mallo J. Loosenin, a novel protein with cellulose-disrupting activity from *Bjerkandera adusta*. *Microb Cell Fact*. BioMed Central Ltd; 2011;10(1):8.

52. Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, Nyssönen E, et al. Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *Eur J Biochem.* 2002;269(17):4202–11.
53. Henrissat B. A classification of glycosyl hydrolases based sequence similarities amino acid. *Biochemistry.* 1991;280:309–16.
54. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014;42(D1):490–5.
55. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J.* 2004 Sep 15;382(Pt 3):769–81.
56. Levit MN, Shkrob AM. Lignin and ligninase. *Bioorg Khim.* 1992;18(3):309–45.
57. Kennedy J, O’Leary ND, Kiran GS, Morrissey JP, O’Gara F, Selvin J, et al. Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. *J Appl Microbiol.* 2011;111:787–99.
58. Kostylev M, Wilson D. A distinct model of synergism between a processive endocellulase (TfCel9A) and an exocellulase (TfCel48A) from *Thermobifida fusca*. *Appl Environ Microbiol.* 2014;80(1):339–44.
59. Horn SJ, Sørli M, Vårum KM, Våljamäe P, Eijsink VGH. Measuring processivity. *Methods in Enzymology* [Internet]. California: Elsevier Inc.; 2012. p. 69–95.
60. Merino S, Cherry J. Progress and challenges in enzyme development for biomass utilization. *Biofuels.* Springer; 2007;95–120.
61. Viikari L, Vehmaanperä J, Koivula a. Lignocellulosic ethanol: From science to industry. *Biomass and Bioenergy* 2012;46:13–24.
62. Forsberg Z, Vaaje-kolstad G, Westereng B, Bunsæ AC, Stenstrøm Y, Mackenzie A, et al. Cleavage of cellulose by a cbm33 protein. *Protein Sci.* 2011;20(9):1479–83.
63. Lamed R, Setter E, Kenig R, Bayer E. Cellulosome: a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. *Biotechnol Bioeng.* 1983;13:163–81.
64. Doi RH, Kosugi A. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat Rev Microbiol.* 2004;2(7):541–51.
65. Bayer E a, Belaich J-P, Shoham Y, Lamed R. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol.* 2004;58:521–54.
66. Yaniv O, Frolow F, Levy-Assraf M, Lamed R, Bayer E a. Interactions between family 3 carbohydrate binding modules (CBMs) and cellulosomal linker peptides. [Internet]. 1st ed. *Methods in enzymology.* Elsevier Inc.; 2012. 247-59 .
67. Fernandes AC, Fontes CMG a, Gilbert HJ, Hazlewood GP, Fernandes TH, Ferreira LM a. Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the Presence of Xylan-Binding Domains in Enzyme Complexes. *Biochem J.* 1999;342:105–10.

68. Kosugi A, Murashima K, Doi RH. Xylanase and Acetyl Xylan Esterase Activities of XynA, a Key Subunit of the Clostridium cellulovorans Cellulosome for Xylan Degradation. *Appl Environ Microbiol.* 2002;68(12):6399–402.
69. Zverlov V V., Fuchs KP, Schwarz WH. Chi18A, the endochitinase in the cellulosome of the thermophilic, cellulolytic bacterium Clostridium thermocellum. *Appl Environ Microbiol.* 2002;68(6):3176–9.
70. Tamaru Y, Doi RH. Pectate lyase A, an enzymatic subunit of the Clostridium cellulovorans cellulosome. *Proc Natl Acad Sci U S A.* 2001;98(7):4125–9.
71. Bayer EA, Morag E, Lamed R. The cellulosome--a treasure-trove for biotechnology. *Trends Biotechnol. ENGLAND;* 1994 Sep;12(9):379–86.
72. Fierobe HP, Mechaly A, Tardif C, Belaich A, Lamed R, Shoham Y, et al. 4 Domain Structure and Engineering-Designer Nanosomes: Selective Engineering of Dockerin-containing Enzymes into Chimeric Scaffoldins to Form Defined Nanoreactors. *Spec Publ R Soc Chem. London: The Society,* 1980-; 2002;275:113–26.
73. Vazana Y, Morais S, Barak Y, Lamed R, Bayer E a. Designer cellulosomes for enhanced hydrolysis of cellulosic substrates. [Internet]. 1st ed. *Methods in enzymology.* Elsevier Inc.; 2012. 429-52 p.
74. Mingardon F, Chanal A, Tardif C, Bayer E a, Fierobe H-P. Exploration of new geometries in cellulosome-like chimeras. *Appl Environ Microbiol.* 2007;73(22):7138–49.
75. Carvalho AL, Dias FM V, Prates J a M, Nagy T, Gilbert HJ, Davies GJ, et al. Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc Natl Acad Sci U S A.* 2003;100(24):13809–14.
76. Mood SH, Golfeshan AH, Tabatabaei M, Jouzani GS, Najafi GH, Gholami M, et al. Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. *Renew Sustain Energy Rev. Elsevier;* 2013;27:77–93.
77. Galbe M, Zacchi G. Pretreatment: the key to efficient utilization of lignocellulosic materials. *Biomass and Bioenergy. Elsevier;* 2012;46:70–8.
78. Schuster BG, Chinn MS. Consolidated Bioprocessing of Lignocellulosic Feedstocks for Ethanol Fuel Production. *BioEnergy Res.* 2012;6(2):416–35.
79. Hiden A, Inoue H, Tsukahara K, Fujimoto S, Minowa T, Inoue S, et al. Wet disk milling pretreatment without sulfuric acid for enzymatic hydrolysis of rice straw. *Bioresour Technol. Elsevier;* 2009;100(10):2706–11.
80. Lamsal B, Yoo J, Brijwani K, Alavi S. Extrusion as a thermo-mechanical pre-treatment for lignocellulosic ethanol. biomass and bioenergy. *Elsevier;* 2010;34(12):1703–10.
81. Chang K-L, Thitikorn-amorn J, Hsieh J-F, Ou B-M, Chen S-H, Ratanakhanokchai K, et al. Enhanced enzymatic conversion with freeze pretreatment of rice straw. *Biomass and Bioenergy. Elsevier;* 2011;35(1):90–5.
82. Tadesse H, Luque R. Advances on biomass pretreatment using ionic liquids: an overview. *Energy Environ Sci. Royal Society of Chemistry;* 2011;4(10):3913–29.

83. Portillo M del C, Saadeddin A. Recent trends in ionic liquid (IL) tolerant enzymes and microorganisms for biomass conversion. *Crit Rev Biotechnol*. Informa Healthcare USA, Inc. New York; 2014;(0):1–8.
84. Mesa L, González E, Cara C, González M, Castro E, Mussatto SI. The effect of organosolv pretreatment variables on enzymatic hydrolysis of sugarcane bagasse. *Chem Eng J*. Elsevier B.V.; 2011;168(3):1157–62.
85. Klein-Marcuschamer D, Oleskowicz-Popiel P, Simmons B a., Blanch HW. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol Bioeng*. 2012;109(4):1083–7.
86. Limayem A, Ricke SC. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. *Prog Energy Combust Sci*. Elsevier Ltd; 2012;38(4):449–67.
87. Juturu V, Wu JC. Microbial cellulases: Engineering, production and applications. *Renew Sustain Energy Rev*. 2014;33:188–203.
88. Bettiga M, Bengtsson O, Hahn-Hägerdal B, Gorwa-Grauslund MF. Arabinose and xylose fermentation by recombinant *Saccharomyces cerevisiae* expressing a fungal pentose utilization pathway. *Microb Cell Fact*. 2009;8:40.
89. Fernández-Sandoval M. T, Huerta-Beristain G, Trujillo-Martinez B, Bustos P, González V, Bolívar F, et al. Laboratory metabolic evolution improves acetate tolerance and growth on acetate of ethanologenic *Escherichia coli* under non-aerated conditions in glucose-mineral medium. *Appl Microbiol Biotechnol*. 2012;96(5):1291–300.
90. Ingram LO, Conway T, Clark DP, Sewell GW, Preston JF. Genetic engineering of ethanol production in *Escherichia coli*. *Appl Environ Microbiol*. 1987 Oct;53(10):2420–5.
91. Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, Ingram LO. Silencing of NADPH-dependent oxidoreductase genes (*yqhD* and *dkgA*) in furfural-resistant ethanologenic *Escherichia coli*. *Appl Environ Microbiol*. 2009;75(13):4315–23.
92. Dien BS, Cotta M a, Jeffries TW. Bacteria engineered for fuel ethanol production: current status. *Appl Microbiol Biotechnol* 2003;63(3):258–66.
93. Maki M, Leung KT, Qin W. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *Int J Biol Sci*. 2009;5(5):500–16.
94. Sainz MB. Commercial cellulosic ethanol: The role of plant-expressed enzymes. *Vitr Cell Dev Biol - Plant*. 2009;45(3):314–29.
95. Hasunuma T, Kondo A. Consolidated bioprocessing and simultaneous saccharification and fermentation of lignocellulose to ethanol with thermotolerant yeast strains. *Process Biochem* [Internet]. Elsevier Ltd; 2012;47(9):1287–94.
96. Ballesteros M, Oliva JM, Negro MJ, Manzanares P, Ballesteros I. Ethanol from lignocellulosic materials by a simultaneous saccharification and fermentation process (SFS) with *Kluyveromyces marxianus* CECT 10875. *Process Biochem*. 2004;39(12):1843–8.

97. Kitagawa T, Tokuhiko K, Sugiyama H, Kohda K, Isono N, Hisamatsu M, et al. Construction of a  $\beta$ -glucosidase expression system using the multistress-tolerant yeast *Issatchenkia orientalis*. *Appl Microbiol Biotechnol*. Springer; 2010;87(5):1841–53.
98. Oberoi HS, Babbar N, Sandhu SK, Dhaliwal SS, Kaur U, Chadha BS, et al. Ethanol production from alkali-treated rice straw via simultaneous saccharification and fermentation using newly isolated thermotolerant *Pichia kudriavzevii* HOP-1. *J Ind Microbiol Biotechnol*. 2012;39(4):557–66.
99. Lau MW, Gunawan C, Balan V, Dale BE. Comparing the fermentation performance of *Escherichia coli* KO11, *Saccharomyces cerevisiae* 424A(LNH-ST) and *Zymomonas mobilis* AX101 for cellulosic ethanol production. *Biotechnol Biofuels*. 2010;3:11.
100. Olofsson K, Palmqvist B, Lidén G. Improving simultaneous saccharification and co-fermentation of pretreated wheat straw using both enzyme and substrate feeding. *Biotechnol Biofuels*. BioMed Central Ltd., Floor 6 London WC 1 X 8 HL UK; 2010;3:17.
101. Zhang J, Lynd LR. Ethanol production from paper sludge by simultaneous saccharification and co-fermentation using recombinant xylose-fermenting microorganisms. *Biotechnol Bioeng*. 2010;107(2):235–44.
102. Lynd LR, van Zyl WH, McBride JE, Laser M. Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol*. 2005 ;16(5):577–83.
103. Demain AL, Newcomb M, Wu JHD, Demain AL, Newcomb M, Wu JHD. Cellulase, Clostridia, and Ethanol. 2005;69(1):124–54.
104. Tyurin M V, Tyurin M V, Desai SG, Desai SG, Lynd LR, Lynd LR. Electrotransformation of. *Microbiology*. 2004;70(2):883–90.
105. Jennert KCB, Tardif C, Young DI. Gene transfer to *Clostridium cellulolyticum* ATCC35319. *Microbiol*. 2000;146:3071–80.
106. Chung D, Cha M, Guss AM, Westpheling J. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proc Natl Acad Sci*. 2014;111(24):8931–6.
107. Linger JG, Adney WS, Darzins A. Heterologous expression and extracellular secretion of cellulolytic enzymes by *Zymomonas mobilis*. *Appl Environ Microbiol*. 2010;76(19):6360–9.
108. Vasan PT, Piriya PS, Prabhu DIG, Vennison SJ. Cellulosic ethanol production by *Zymomonas mobilis* harboring an endoglucanase gene from *Enterobacter cloacae*. *Bioresour Technol*. Elsevier Ltd; 2011;102(3):2585–9.
109. Yamada R, Hasunuma T, Kondo A. Endowing non-cellulolytic microorganisms with cellulolytic activity aiming for consolidated bioprocessing. *Biotechnol Adv*. Elsevier Inc.; 2013.
110. Ryu S, Karim MN. A whole cell biocatalyst for cellulosic ethanol production from dilute acid-pretreated corn stover hydrolyzates. *Appl Microbiol Biotechnol*. 2011;91(3):529–42.
111. Muñoz-Gutiérrez I, Oropeza R, Gosset G, Martinez A. Cell surface display of a  $\beta$ -glucosidase employing the type V secretion system on ethanologenic *Escherichia coli* for the fermentation of cellobiose to ethanol. *J Ind Microbiol Biotechnol*. 2012;39(8):1141–52.

112. Lu Y, Percival Zhang Y-H, Lynd LR. Enzyme–microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *PNAS*. 2006;103(44):16165–9.
113. Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, Wittwer P. Metagenomic gene discovery: Past, present and future. *Trends Biotechnol*. 2005;23(6):321–9.
114. Singh B, Gautam SK, Verma V, Kumar M, Singh B. Metagenomics in animal gastrointestinal ecosystem: Potential biotechnological prospects. *Anaerobe*. 2008;14(3):138–44.
115. Cerqueira V, Hollenbach E, Maboni F, Camargo FO, Peralba M do Carmo R., Bento F. Bioprospection and selection of bacteria isolated from environments contaminated with petrochemical residues for application in bioremediation. *World J Microbiol Biotechnol*. Springer Netherlands; 2012;28(3):1203–22.
116. Mutanda T, Ramesh D, Karthikeyan S, Kumari S, Anandraj A, Bux F. Bioprospecting for hyper-lipid producing microalgal strains for sustainable biofuel production. *Bioresour Technol*. 2011 Jan;102(1):57–70.
117. Acharya S, Chaudhary A. Bioprospecting thermophiles for cellulase production: A review. *Brazilian J Microbiol*. 2012;43(3):844–56.
118. Lo YC, Saratale GD, Chen WM, Bai M Der, Chang JS. Isolation of cellulose-hydrolytic bacteria and applications of the cellulolytic enzymes for cellulosic biohydrogen production. *Enzyme Microb Technol*. 2009;44(6-7):417–25.
119. Dantur KI, Enrique R, Welin B, Castagnaro AP. Isolation of cellulolytic bacteria from the intestine of *Diatraea saccharalis* larvae and evaluation of their capacity to degrade sugarcane biomass. *AMB Express*. 2015;5(1).
120. Rameshwar Tiwari SM. Running Head: Bioprospecting for Biomass Hydrolysing Fungi Bioprospecting for Superior Biomass Hydrolysing Fungi from Diverse Habitats. *J Biodiversity, Bioprospecting Dev*. 2015;02(02):1–7.
121. Ferrés I, Amarelle V, Noya F, Fabiano E. Identification of Antarctic culturable bacteria able to produce diverse enzymes of potential biotechnological interest. *Adv Polar Sci*. 2015;26(1):71–9.
122. Duan C-J, Feng J-X. Mining metagenomes for novel cellulase genes. *Biotechnol Lett*. 2010 Dec [cited 2013 Dec 13];32(12):1765–75.
123. Schmeisser C, Steele H, Streit WR. Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol*. 2007;75:955–62.
124. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol*. 1998;5(10):R245–9.
125. Suenaga H. Targeted metagenomics: A high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ Microbiol*. 2012;14:13–22.
126. Handelsman J, Tiedje J. *THE NEW SCIENCE OF METAGENOMICS: Revealing the Secrets of Our Microbial Planet* [Internet]. National Academies Press. 2007.



127. Scholz MB, Lo CC, Chain PSG. Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. *Curr Opin Biotechnol.* 2012;23(Table 1):9–15.
128. Schloss PD, Handelsman J. Biotechnological prospects from metagenomics. *Curr Opin Biotechnol.* 2003;14:303–10.
129. Krause L, Diaz NN, Goesmann A, Kelley S, Nattkemper TW, Rohwer F, et al. Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res.* 2008;36(7):2230–9.
130. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen J a, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science.* 2004;304(5667):66–74.
131. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature.* Nature Publishing Group; 2004;428(6978):37–43.
132. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust E V. Untangling Genomes from Metagenomes: Revealing an Uncultured Class of Marine Euryarchaeota. *Science.* 2012;335(6068):587–90.
133. Leis B, Angelov A, Liebl W. Screening and expression of genes from metagenomes. [Internet]. *Advances in applied microbiology.* Elsevier; 2013. 1-68.
134. Rondon M, August P, Bettermann A, Brady S, Grossman T, Lilles M, et al. Cloning the metagenome: a stragy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol.* 2000;66(6):2541–7.
135. Terrón-González L, Medina C, Limón-Mortés MC, Santero E. Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries. *Sci Rep.* 2013;3:1107.
136. Martinez A, Kolvek SJ, Lai C, Yip T, Hopke J, Brown KA, et al. Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. 2004;70(4):2452–63.
137. Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl Environ Microbiol.* 2010;76(5):1633–41.
138. Gabor EM, Alkema WBL, Janssen DB. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol.* 2004;6(9):879–86.
139. Dashtban M, Maki M, Leung KT, Mao C, Qin W. Cellulase activities in biomass conversion: measurement methods and comparison. *Crit Rev Biotechnol.* 2010;30(4):302–9.
140. Teather RM, Wood PJ. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol.* 1982 Apr;43(4):777–80.
141. Ghose T, others. Measurement of cellulase activities. *Pure Appl Chem.* 1987;59(2):257–68.

142. Wong KK, Tan LU, Saddler JN. Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. *Microbiol Rev.* 1988;52(3):305–17.
143. Morais S, Barak Y, Caspi J, Hadar Y, Lamed R, Shoham Y, et al. Contribution of a xylan-binding module to the degradation of a complex cellulosic substrate by designer cellulosomes. *Appl Environ Microbiol.* 2010;76(12):3787–96.
144. Lorenz P, Eck J. Metagenomics and industrial applications. *Nat Rev Microbiol.* 2005;3:510–6.
145. Dewan SS. *Global Markets for Enzymes in Industrial Applications.* Wellesley, MA; 2014.
146. Lorenz P, Schleper C. Metagenome - A challenging source of enzyme discovery. *J Mol Catal B Enzym.* 2002;19-20:13–9.
147. Steele HL, Jaeger KE, Daniel R, Streit WR. Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microbiol Biotechnol.* 2008;16:25–37.
148. Nyssönen M, Tran HM, Karaoz U, Weihe C, Hadi MZ, Martiny JBH, et al. Coupled high-throughput functional screening and next generation sequencing for identification of plant polymer decomposing enzymes in metagenomic libraries. *Front Microbiol.* 2013;4:1–14.
149. Healy FG, Ray RM, Aldrich HC, Wilkie a. C, Ingram LO, Shanmugam KT. Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol.* 1995;43:667–74.
150. Voget S, Steele HL, Streit WR. Characterization of a metagenome-derived halotolerant cellulase. *J Biotechnol.* 2006;126:26–36.
151. Feng Y, Duan C, Pang H, Mo X. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. 2007;319–28.
152. Bao L, Huang Q, Chang L, Zhou J, Lu H. Enzymatic Screening and characterization of a cellulase with endocellulase and exocellulase activity from yak rumen metagenome. *Journal Mol Catal B, Enzym. Elsevier B.V.;* 2011;73(1-4):104–10.
153. Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science.* 2011;331(6016):463–7.
154. Ferrer M, Ghazi A, Beloqui A, Vieites JM, López-Cortés N, Marín-Navarro J, et al. Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. *PLoS One.* 2012;7(6):e38134.
155. Cottrell MT, Moore J a, David L, Kirchman DL. Chitinases from Uncultured Marine Microorganisms Chitinases from Uncultured Marine Microorganisms. 1999;65(6):2553–7.
156. Sharma S, Khan FG, Qazi GN. Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. *Appl Microbiol Biotechnol. Springer;* 2010;86(6):1821–8.
157. Henne A, Schmitz R a, Bömeke M, Daniel R, Gottschalk G. Screening of Environmental DNA Libraries for the Presence of Genes Conferring Lipolytic Activity on *Escherichia coli*

- Screening of Environmental DNA Libraries for the Presence of Genes Conferring Lipolytic Activity on *Escherichia coli*. *Appl Environ Microbiol*. 2000;66(7):3113–6.
158. Rodríguez MC, Loaces I, Amarelle V, Senatore D, Iriarte A, Fabiano E, et al. Est10: A Novel Alkaline Esterase Isolated from Bovine Rumen Belonging to the New Family XV of Lipolytic Enzymes. *PLoS One*. 2015;10(5):e0126651.
  159. Ferrés I, Amarelle V, Noya F, Fabiano E. Construction and screening of a functional metagenomic library to identify novel enzymes produced by Antarctic bacteria. *Adv Polar Sci*. 2015;26(1):96–101.
  160. Neveu J, Regeard C, DuBow MS. Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. *Appl Microbiol Biotechnol*. Springer; 2011;91(3):635–44.
  161. Van Hellemond EW, Janssen DB, Fraaije MW. Discovery of a novel styrene monooxygenase originating from the metagenome. *Appl Environ Microbiol*. 2007;73(18):5832–9.
  162. Banik J, Brady SF. Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr Opin Microbiol*. 2010;13(5):603–9.
  163. Ferrer M, Beloqui A, Timmis KN, Golyshin PN. Metagenomics for mining new genetic resources of microbial communities. *J Mol Microbiol Biotechnol*. 2008;16:109–23.
  164. Simon C, Daniel R. Metagenomic analyses: Past and future trends. *Appl Environ Microbiol*. 2011;77(4):1153–61.
  165. Kirk O, Borchert TV, Fuglsang CC. Industrial enzyme applications. *Curr Opin Biotechnol*. 2002;13(4):345–51.
  166. Bhat MK. Cellulases and related enzymes in biotechnology. *Biotechnol Adv*. 2000;18(5):355–83.
  167. Percival Zhang Y-H, Himmel ME, Mielenz JR. Outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv*. 2006;24(5):452–81.
  168. Juturu V, Wu JC. Microbial xylanases: Engineering, production and industrial applications. *Biotechnol Adv*. 2012;30(6):1219–27.
  169. Ferrer M, Golyshina O, Beloqui A, Golyshin PN. Mining enzymes from extreme environments. *Curr Opin Microbiol*. 2007;10(3):207–14.
  170. Rooks DJ, McDonald JE, McCarthy AJ. Metagenomic approaches to the discovery of cellulases. 1st ed. *Methods in enzymology*. Elsevier Inc.; 2012. 375-94.
  171. Xiong X, Yin X, Pei X, Jin P, Zhang A, Li Y, et al. Retrieval of glycoside hydrolase family 9 cellulase genes from environmental DNA by metagenomic gene specific multi-primer PCR. *Biotechnol Lett*. Springer; 2012;34(5):875–82.
  172. Elifantz H, Waidner L a., Michelou VK, Cottrell MT, Kirchman DL. Diversity and abundance of glycosyl hydrolase family 5 in the North Atlantic Ocean. *FEMS Microbiol Ecol*. 2008;63(3):316–27.

173. Cobucci-Ponzano B, Perugini G, Strazzulli A, Rossi M, Moracci M. Thermophilic glycosynthases for oligosaccharides synthesis. 1st ed. *Methods in enzymology*. Elsevier Inc.; 2012; 273-300.
174. Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM, et al. High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* [Internet]. *BMC Genetics*; 2012;13(1):53. Available from: *BMC Genetics*
175. Brulc JM, Antonopoulos DA, Berg ME, Wilson MK, Yannarell AC, Dinsdale EA, et al. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *PNAS*. 2009;106(6):1948–53.
176. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*. 2007;450(7169):560–5.
177. Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, Pinto-Tomás A a., et al. An insect herbivore microbiome with high plant Biomass-degrading capacity. *PLoS Genet*. 2010;6(9).
178. Beloqui A, Nechitaylo TY, López-Cortés N, Ghazi A, Guazzaroni ME, Polaina J, et al. Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. *Appl Environ Microbiol*. 2010;76(17):5934–46.
179. King AJ, Cragg SM, Li Y, Dymond J, Guille MJ, Bowles DJ, et al. Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proc Natl Acad Sci U S A*. 2010;107(12):5345–50.
180. Wilson DB. Microbial diversity of cellulose hydrolysis. *Curr Opin Microbiol*. Elsevier Ltd; 2011;14(3):259–63.
181. Russell JB, Muck RE, Weimer PJ. Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. *FEMS Microbiol Ecol*. 2009;67(2):183–97.
182. Duan C, Xian L, Zhao G, Feng Y, Pang H, Bai X, et al. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J Appl Microbiol*. 2009;107:245–56.
183. Wang L, Hatem A, Catalyurek U V, Morrison M, Yu Z. Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS One*. 2013;8(11):e78507.
184. Zhao S, Wang J, Bu D, Liu K, Zhu Y, Dong Z, et al. Novel glycoside hydrolases identified by screening a chinese holstein dairy cow rumen-derived metagenome library. *Appl Environ Microbiol*. 2010;76(19):6701–5.
185. Chang L, Ding M, Bao L, Chen Y, Zhou J, Lu H. Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. *Appl Microbiol Biotechnol*. 2011;90(6):1933–42.
186. Passeggi M, Benzo M, Borzacconi L. Codigestión de Contenido Ruminal, Residuo de Descarne y Lodo biológico Secundario. Evaluación en Reactores Semi- Continuos de Mezcla

- Completa y de Percolación. X Oficina e Seminário Latino Americano de Digestão Anaeróbia. Ouro Preto; 2011. p. 7.
187. Larue R, Yu Z, Parisi V a., Egan AR, Morrison M. Novel microbial diversity adherent to plant biomass in the herbivore gastrointestinal tract, as revealed by ribosomal intergenic spacer analysis and rrs gene sequencing. *Environ Microbiol.* 2005;7:530–43.
  188. Cappuccino JG, Sherman N, Microbiology A. *A laboratory manual.* The Benjamin/Cumming Pud. comp INC California; 1983.
  189. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning.* Cold spring harbor laboratory press New York; 1989.
  190. Loaces I, Amarelle V, Fabiano E, Muñoz-Gutiérrez I, Martinez A, Noya F. Improved ethanol production from biomass by a rumen metagenomic DNA fragment expressed in *Escherichia coli* MS04 during fermentation. *Appl Microbiol Biotechnol.* 2015;
  191. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* [Internet]. BioMed Central Ltd; 2011;12(1):444. Available from: <http://www.biomedcentral.com/1471-2164/12/444>
  192. Noguchi H, Park J, Takagi T. MetaGene: Prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* 2006;34(19):5623–30.
  193. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725–9.
  194. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 2012;40(W1):445–51.
  195. Syed MH, Karpinets T V., Parang M, Leuze MR, Park BH, Hyatt D, et al. BESC knowledgebase public portal. *Bioinformatics.* 2012;28(5):750–1.
  196. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc Handb.* 2005;571–607.
  197. Deng W, Xi D, Mao H, Wanapat M. The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. *Mol Biol Rep. Springer;* 2008;35(2):265–74.
  198. Shedova EN, Berezina O V, Lunina N a, Zverlov V V, Schwartz WH, Velikodvorskaia G a. Cloning and characterization of a large metagenomic DNA fragment containing glycosyl-hydrolase genes. *Mol Gen Mikrobiol Virusol.* 2009;24(1):11–5.
  199. Ferrer M, Golyshina O V, Chernikova TN, Khachane AN, Reyes-Duarte D, Santos VAP, et al. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol. Wiley Online Library;* 2005;7(12):1996–2010.
  200. Palackal N, Lyon CS, Zaidi S, Luginbühl P, Dupree P, Goubet F, et al. A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Appl Microbiol Biotechnol. Springer;* 2007;74(1):113–24.

201. Ko K-C, Lee JH, Han Y, Choi JH, Song JJ. A novel multifunctional cellulolytic enzyme screened from metagenomic resources representing ruminal bacteria. *Biochem Biophys Res Commun.* Elsevier Inc.; 2013;441(3):567–72.
202. Fan Z, Werkman JR, Yuan L. Engineering of a multifunctional hemicellulase. *Biotechnol Lett.* 2009;31(5):751–7.
203. Angelov A, Mientus M, Liebl S, Liebl W. A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. *Syst Appl Microbiol.* 2009;32(3):177–85.
204. Tang YQ, Ji P, Hayashi J, Koike Y, Wu XL, Kida K. Characteristic microbial community of a dry thermophilic methanogenic digester: Its long-term stability and change with feeding. *Appl Microbiol Biotechnol.* 2011;91(5):1447–61.
205. Li T, Mazéas L, Sghir A, Leblon G, Bouchez T. Insights into networks of functional microbes catalysing methanization of cellulose under mesophilic conditions. *Environ Microbiol.* Wiley Online Library; 2009;11(4):889–904.
206. Warren RL, Freeman JD, Levesque RC, Smailus DE, Flibotte S, Holt R a. Transcription of foreign DNA in *Escherichia coli*. *Genome Res.* 2008;18(11):1798–805.
207. Sørensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol.* 2005;115(2):113–28.
208. Lilie H, Schwarz E, Rudolph R. Advanced in refolding of proteins produced in *E. Coli*. *Curr Opin Biotechnol.* 1998;9:497–501.
209. Mergulhão FJM, Summers DK, Monteiro G a. Recombinant protein secretion in *Escherichia coli*. *Biotechnol Adv.* 2005;23(3):177–202.
210. Van Tilbeurgh H, Loontjens FG, De Bruyne CK, Claeysens M. Fluorogenic and chromogenic glycosides as substrates and ligands of carbohydrases. *Methods Enzymol.* Elsevier; 1988;160:45–59.
211. Sun Y, Cheng J. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour Technol.* 2002 May;83(1):1–11.
212. Goberna M, Insam H, Franke-Whittle IH. Effect of biowaste sludge maturation on the diversity of thermophilic bacteria and archaea in an anaerobic reactor. *Appl Environ Microbiol.* 2009;75(8):2566–72.
213. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. *Science.* 2008;320(5883):1647–51.
214. Beloqui A, Pita M, Polaina J, Martínez-Arias A, Golyshina O V, Zumárraga M, et al. Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: biochemical properties, structural analysis, and phylogenetic relationships. *J Biol Chem.* 2006;281(32):22933–42.
215. Mello L V, Chen X, Rigden DJ. Mining metagenomic data for novel domains: BACON, a new carbohydrate-binding module. *FEBS Lett [Internet]. Federation of European Biochemical Societies;* 2010 Jun 3 [cited 2013 Dec 13];584(11):2421–6.

216. Cristina Ravanal M, Callegari E, Eyzaguirre J. Novel bifunctional  $\alpha$ -L-arabinofuranosidase/xylobiohydrolase (ABF3) from *Penicillium purpurogenum*. *Appl Environ Microbiol.* 2010;76(15):5247–53.
217. Yoshida S, Hespen CW, Beverly RL, Mackie RI, Cann IKO. Domain analysis of a modular  $\alpha$ -L-arabinofuranosidase with a unique carbohydrate binding strategy from the fiber-degrading bacterium *Fibrobacter succinogenes* S85. *J Bacteriol.* 2010;192(20):5424–36.
218. Anantharaman V, Aravind L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol.* 2003;4(2):R11.
219. Buist G, Steen A, Kok J, Kuipers OP. LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol.* 2008;68(4):838–47.
220. Spaink HP. Specific recognition of bacteria by plant LysM domain receptor kinases. *Trends Microbiol.* Elsevier; 2004;12(5):201–4.
221. Naumoff DG.  $\beta$ -Fructosidase superfamily: Homology with some  $\alpha$ -L-arabinases and  $\beta$ -D-xylosidases. *Proteins Struct Funct Bioinforma.* Wiley Online Library; 2001;42(1):66–76.
222. Senoura T, Ito S, Taguchi H, Higa M, Hamada S, Matsui H, et al. New microbial mannan catabolic pathway that involves a novel mannosylglucose phosphorylase. *Biochem Biophys Res Commun.* Elsevier Inc.; 2011;408(4):701–6.
223. Kawahara R, Saburi W, Odaka R, Taguchi H, Ito S, Mori H, et al. Metabolic mechanism of mannan in a ruminal bacterium, *Ruminococcus albus*, involving two mannoside phosphorylases and cellobiose 2-epimerase: Discovery of a new carbohydrate phosphorylase, B-1,4-mannooligosaccharide phosphorylase. *J Biol Chem.* 2012;287(50):42389–99.
224. Chiku K, Nihira T, Suzuki E, Nishimoto M, Kitaoka M, Ohtsubo K, et al. Discovery of Two  $\beta$ -1,2-Mannoside Phosphorylases Showing Different Chain-Length Specificities from *Thermoanaerobacter* sp. X-514. *PLoS One.* 2014;9(12):e114882.
225. Ni J, Tokuda G. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol Adv.* Elsevier Inc.; 2013 Apr 23;
226. Dowd SE, Callaway TR, Wolcott RD, Sun Y, McKeehan T, Hagevoort RG, et al. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol.* 2008;8:125.
227. Callaway TR, Dowd SE, Edrington TS, Anderson RC, Krueger N, Bauer N, et al. Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing. *J Anim Sci.* American Society of Animal Science; 2010;88(12):3977–83.
228. Creevey CJ, Kelly WJ, Henderson G, Leahy SC. Determining the culturability of the rumen bacterial microbiome. *Microb Biotechnol.* 2014;7(5):467–79.
229. Teeling H, Waldmann J, Lombardot T, Bauer M, Glöckner FO. TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics.* 2004;5:163.

230. Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO. Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ Microbiol.* 2004;6(9):938–47.
231. Gao F, Zhang CT. GC-Profile: A web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. *Nucleic Acids Res.* 2006;34(WEB. SERV. ISS.):686–91.
232. Krause DO, Denman SE, Mackie RI, Morrison M, Rae AL, Attwood GT, et al. Opportunities to improve fiber degradation in the rumen: Microbiology, ecology, and genomics. *FEMS Microbiol Rev.* 2003;27(5):663–93.
233. Miron J, Ben-Ghedalia D, Morrison M. Invited review: adhesion mechanisms of rumen cellulolytic bacteria. *J Dairy Sci. Elsevier;* 2001;84(6):1294–309.
234. Ingram LO, Conway T, Alterthum F. ethanol production by *Escherichia coli* strains co-expressing *Zymomonas pdc* and *adh* genes. US; 5000000, 1991.
235. Miller EN, Turner PC, Jarboe LR, Ingram LO. Genetic changes that increase 5-hydroxymethyl furfural resistance in ethanol-producing *Escherichia coli* LY180. *Biotechnol Lett.* 2010;32(5):661–7.
236. Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, et al. How biotech can transform biofuels. *Nat Biotechnol.* 2008 Feb;26(2):169–72.
237. Alvira P, Tomás-Pejó E, Ballesteros M, Negro MJ. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresour Technol. Elsevier Ltd;* 2010;101(13):4851–61.
238. Chandel AK, Chandrasekhar G, Silva MB, Silvério da Silva S. The realm of cellulases in biorefinery development. *Crit Rev Biotechnol [Internet].* 2012 Sep [cited 2013 Aug 21];32(3):187–202.
239. Posada JA, Rincón LE, Cardona CA. Bioresource Technology Design and analysis of biorefineries based on raw glycerol : Addressing the glycerol problem. *Bioresour Technol. Elsevier Ltd;* 2012;111:282–93.
240. Taupp M, Mewis K, Hallam SJ. The art and design of functional metagenomic screens. *Curr Opin Biotechnol. Elsevier Ltd;* 2011;22(3):465–72.
241. Yamada R, Nakatani Y, Ogino C, Kondo A. Efficient direct ethanol production from cellulose by cellulase- and cellodextrin transporter-co-expressing *Saccharomyces cerevisiae*. *AMB Express.* 2013;3(1):34.
242. Haan R Den, McBride JE, La C, Lynd LR, Zyl WH Van. Functional expression of cellobiohydrolases in *Saccharomyces cerevisiae* towards one-step conversion of cellulose to ethanol. 2007;40:1291–9.



## **ANEXOS**

---

## Anexo 1.

### COMPOSICIÓN DE SOLUCIONES Y MEDIOS DE CULTIVO

#### Soluciones

☞ Amortiguador fosfato salino (PBS, phosphate buffered saline):

NaCl	8.0 g/L
KCl	0.2 g/L
Na <sub>2</sub> HPO <sub>4</sub>	1.44 g/L
KH <sub>2</sub> PO <sub>4</sub>	0.24 g/L

Ajustar el pH a 7.4 con HCl. Esterilizar por autoclave.

☞ Solución de lisis:

NaCl	700 mM,
Tris-HCl	50 mM pH=8.0,
EDTA	100 mM,
SDS	140 mM

☞ Solución para gradiente de Percoll:

Tris	100mM
EDTA	100mM, pH 8
NaCl	1.5M
CTAB	1%

☞ Amortiguador TAE (Tris-acetato) para corrida electroforética:

Tris-acetato	0,04 M
EDTA	0,001M, pH 8

#### Mini geles para SDS-PAGE:

☞ Gel concentrador:

1,25 mL amortiguador concentrador

Tris-base 0,5 M, SDS 14 mM, pH 6,8

3,4 mL solución acrilamida 30%/ bisacrilamida 0,8%

5 µL TEMED 10%

50 µL persulfato de amonio 10%

4,2 mL H<sub>2</sub>O destilada

☞ Gel separador (12%):

2,5 mL amortiguador separador

Tris-base 1,5 M, SDS 14 mM, pH 8,8

4,04 mL solución acrilamida 30%/ bisacrilamida 0,8%

5  $\mu$ L TEMED 10%  
100  $\mu$ L Persulfato de amonio 10%  
3,56 mL H<sub>2</sub>O destilada

☞ Azul de Coomassie :

Azul de Coomassie 180 mg  
Metanol 95% 90 mL  
Ácido acético glacial 20 mL  
H<sub>2</sub>O destilada 90 mL

☞ Solución decolorante :

Metanol 95% 50 mL  
Ácido acético glacial 16 mL  
H<sub>2</sub>O destilada 130 mL

☞ Solución decolorante :

Metanol 95% 50 mL  
Ácido acético glacial 16 mL  
H<sub>2</sub>O destilada 130 mL

☞ Buffer de carga (2x):

Tris-base 0,5 M 12 mL  
Glicerol 10 mL  
SDS 20% 10 mL  
Azul de bromofenol 5 mL  
H<sub>2</sub>O c.s.p. 45 mL

Soluciones para purificación de proteínas recombinantes:

☞ Amortiguador 2x His-Tag:

Na<sub>2</sub>H<sub>2</sub>PO<sub>4</sub> 100 mM

NaCl            600 mM

Ajustar el pH a 8.

☞ Solución de unión:

Imidazol        5 mM

NaCl            300 mM

Na<sub>2</sub>H<sub>2</sub>PO<sub>4</sub>      50 mM

☞ Solución de lavado:

Imidazol        20, 50, o 100 mM

NaCl            300 mM

Na<sub>2</sub>H<sub>2</sub>PO<sub>4</sub>      50 mM

☞ Solución elución:

Imidazol        250 mM

NaCl            300 mM

Na<sub>2</sub>H<sub>2</sub>PO<sub>4</sub>      50 mM

Glicerol        10% (v/v)

## Medios de cultivo

☞ Luria Bertani (LB)

Triptona            10 g/L

Extracto de levadura 5 g/L

NaCl                10 g/L

Para preparar medio sólido agregar 16g/L de agar.

☞ Medio mínimo M9

MgSO<sub>4</sub>            2 mM

CaCl<sub>2</sub>            0,1 mM

Na <sub>2</sub> HPO <sub>4</sub> ·7H <sub>2</sub> O	12,8 g/L
KH <sub>2</sub> PO <sub>4</sub>	3,0 g/L
NaCl	0,5 g/L
NH <sub>4</sub> Cl	1,0 g/L
Timina	50 µg/mL
Leucina	200 µM

☞ 2xYT

Triptona	16 g/L
Extracto de levadura	10 g/L
NaCl	5 g/L

Ajustar el pH a 7, esterilizar por autoclave.

## Anexo 2.

### ABREVIATURAS

ADN	Ácido desoxirribonucleico
CTAB	Bromuro de hexadeciltrimetilamonio
Cm <sub>12,5</sub>	Cloranfenicol 12,5 µg/mL
CMC	Carboximetil-celulosa
DO <sub>620 nm</sub>	Densidad óptica a 620 nm
4-MUC	4-metilumbeliferil-β-D-celobiósido
Ara	Arabinosa 0,01% (p/v)
Km <sub>50</sub>	Kanamicina 50 µg/mL
Amp <sub>50</sub>	Ampicilina 50 µg/mL
IPTG	Isopropil-β-D-galactopiranosido
X-Gal <sub>20</sub>	5-bromo-4-cloro-3-indolil-β-D-galactopiranosido 20 µg/mL
SDS-PAGE	electroforesis en gel de poliacrilamida con dodecilsulfato sódico
pNPC	p-nitrofenil-celobiósido
DTT	Ditiotreitol
PMSF	Fluoruro de fenilmetilsulfonilo
EDTA	Ácido etildiaminotetraacético

### Anexo 3. TABLAS DE ORFs ASIGNADOS A LOS NUEVE FÓSMIDOS SECUENCIADOS

**Tabla 1.** Asignación de ORFs del fósido Csd4

ORF <sup>(a)</sup>	Posición relativa			Tamaño (aa)	Descripción [Organismo] <sup>(c,d)</sup>	Closest homolog using BLASTP <sup>(b)</sup>		
	Inicio	Fin	Hebra			E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	Gene Bank ID <sup>(c)</sup>
Csd4_1	88	903	+	271				
Csd4_2	1383	2162	+	259	hypothetical protein [ <i>Prevotella buccalis</i> DNF00985]	6.00E-94	56	<a href="#">KGF40528.1</a>
Csd4_3	2251	3588	-	445	glutamate dehydrogenase [ <i>Prevotella</i> ]	0.00E+00	87	<a href="#">WP_028902358.1</a>
Csd4_4	3897	4559	+	220	putative uncharacterized protein [ <i>Prevotella</i> sp. CAG:1185]	8.00E-30	33	<a href="#">WP_021963427.1</a>
Csd4_5	4615	8073	+	1152	hypothetical protein [ <i>Prevotella</i> sp. AGR2160]	3.00E-101	37	<a href="#">WP_028910338.1</a>
Csd4_6	8535	11516	+	993	phosphoenolpyruvate synthase [ <i>Prevotella</i> sp. CAG:1124]	0.00E+00	80	<a href="#">WP_021948672.1</a>
Csd4_7	12022	12651	+	209	cell envelope biogenesis protein TonB [ <i>Prevotella paludivivens</i> ]	7.00E-50	51	<a href="#">WP_018464825.1</a>
Csd4_8	12691	13413	-	240	RDD family protein [ <i>Prevotella veroralis</i> ]	5.00E-45	39	<a href="#">WP_018910394.1</a>
Csd4_9	13552	14469	+	305	membrane protein [ <i>Bacteroides acidifaciens</i> ]	7.00E-118	61	<a href="#">WP_024988293.1</a>
Csd4_10	14459	15322	+	287	membrane protein [ <i>Bacteroides</i> ]	5.00E-18	27	<a href="#">WP_007663060.1</a>
Csd4_11	15335	15955	+	206	hypothetical protein [ <i>Prevotella multisaccharivorax</i> ]	1.00E-19	43	<a href="#">WP_007575351.1</a>
Csd4_12	16099	17364	+	421	hypothetical protein [ <i>Bacteroides fingoldii</i> ]	3.00E-51	31	<a href="#">WP_007767102.1</a>
Csd4_13	17426	18400	+	324	magnesium chelatase [ <i>Prevotella histicola</i> ]	2.00E-172	71	<a href="#">WP_025792001.1</a>
Csd4_14	18478	19191	+	237	hypothetical protein [ <i>Bacteroides acidifaciens</i> ]	3.00E-81	54	<a href="#">WP_024988291.1</a>
Csd4_15	19139	19792	+	217	uncharacterized protein [ <i>Bacteroides</i> sp. CAG:754]	1.00E-98	66	<a href="#">CDA86114.1</a>
<b>Csd4_16</b>	<b>20131</b>	<b>21648</b>	+	<b>505</b>	<b>glycosyl hydrolase family 43 [<i>Prevotella</i> sp. CAG:732]</b>	<b>0.0</b>	<b>61</b>	<b><a href="#">WP_022328427.1</a></b>
Csd4_17	21830	22504	+	224	porin [ <i>Prevotella enoeca</i> ]	4.00E-91	63	<a href="#">WP_025065050.1</a>
Csd4_18	22749	23759	+	336	fructose-bisphosphate aldolase [ <i>Prevotella</i> ]	0.00E+00	92	<a href="#">WP_004341759.1</a>
Csd4_19	23812	25335	-	507	putative YjeF-related sugar kinase [ <i>Prevotella</i> sp. CAG:732]	0.00E+00	68	<a href="#">WP_022328422.1</a>
Csd4_20	25426	26562	+	378	adenylate kinase [ <i>Prevotella stercorea</i> CAG:629]	7.00E-111	84	<a href="#">WP_022429669.1</a>
Csd4_21	26566	27729	+	387	GTPase CgtA [ <i>Prevotella buccae</i> ]	0.00E+00	87	<a href="#">WP_004341745.1</a>
<b>Csd4_22</b>	<b>27726</b>	<b>28517</b>	+	<b>263</b>	<b>polyphenol oxidase [<i>Prevotella</i> sp. S7MS2]</b>	<b>3.00E-112</b>	<b>59</b>	<b><a href="#">KGI61157.1</a></b>
<b>Csd4_23</b>	<b>28593</b>	<b>29741</b>	+	<b>382</b>	<b>cellodextrinase [uncultured microorganism]</b>	<b>0.0</b>	<b>84</b>	<b><a href="#">ADB80112.1</a></b>
Csd4_24	29793	29999	-	373	aICARFT/IMPCHase bienzyme [ <i>Prevotella</i> sp. CAG:1185]	0.0	89	<a href="#">WP_021965082.1</a>
Csd4_25	30011	31132	-	133	50S ribosomal protein L19 [ <i>Prevotella maculosa</i> ]	4.00E-64	79	<a href="#">WP_008564746.1</a>
Csd4_26	31317	31718	-	255	putative uncharacterized protein [ <i>Prevotella</i> sp. CAG:255]	5.00E-92	53	<a href="#">WP_021855259.1</a>
Csd4_27	31995	32762	+	281	5'-nucleotidase [ <i>Prevotella</i> sp. CAG:592]	2.00E-141	66	<a href="#">WP_022316292.1</a>
Csd4_28	32767	33612	+	223	hypothetical protein [ <i>Prevotella histicola</i> ]	4.00E-121	73	<a href="#">WP_008822032.1</a>

<sup>(a)</sup> ORFs identificados por MetaGene.

<sup>(b)</sup> Asignados por RPS Blast.

<sup>(c)</sup> Homólogos más cercanos identificados por BlastP 2.2.30+ usando secuencias de proteínas no redundantes.

<sup>(d)</sup> Los genes asociados a la degradación de lignocelulosa o de particular interés se muestran en negritas.

**Tabla 2.** Asignación de ORFs del fósido Csd6

ORF <sup>(a)</sup>	Posición relativa			Tamaño (aa)	Descripción [Organismo] <sup>(c,d)</sup>	Closest homolog using BLASTP <sup>(b)</sup>		
	inicio	fin	Hebra			E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	Gene Bank ID <sup>(c)</sup>
Csd6_c1.1	229	975	+	248	nitroreductase [ <i>Prevotella</i> sp. FD3004]	6.00E-78	54	<a href="#">WP_036912851.1</a>
Csd6_c1.2	965	1138	-	57	hypothetical protein [ <i>Clostridium hiranonis</i> ]	7.00E-14	58	<a href="#">WP_006440287.1</a>
Csd6_c1.4	1992	2570	+	192	putative GDSL lipase [uncultured microorganism]	5.00E-101	74	<a href="#">ACA61157.1</a>
Csd6_c1.5	2574	2951	-	125	endoribonuclease L-PSP [ <i>Bacteroides</i> sp. CAG:545]	1.00E-67	82	<a href="#">WP_022016360.1</a>
Csd6_c1.6	2963	3586	-	207	nitroreductase family protein [ <i>Prevotella</i> sp. CAG:5226]	2.00E-53	46	<a href="#">WP_022101816.1</a>
Csd6_c1.7	3653	4303	+	216	pyrophosphatase PpaX [ <i>Alistipes</i> sp. CAG:435]	2.00E-65	53	<a href="#">WP_022323411.1</a>
Csd6_c1.8	4317	5078	-	253	molecular chaperone DnaJ [ <i>Butyrivibrio</i> sp. WCD2001]	3.00E-66	46	<a href="#">WP_027202747.1</a>
Csd6_c1.9	5130	5678	+	182	hypothetical protein [ <i>Butyrivibrio</i> sp. WCD2001]	1.00E-46	50	<a href="#">WP_026498890.1</a>
Csd6_c1.12	6432	7259	-	275	dihydroorotate oxidase [ <i>Prevotella ruminicola</i> 23]	1.00E-147	73	<a href="#">ADE82360.1</a>
Csd6_c1.13	7162	7863	-	233	dihydropyrimidine dehydrogenase [ <i>Prevotella brevis</i> ]	4.00E-104	77	<a href="#">WP_027455612.1</a>
Csd6_c1.14	7888	8577	-	229	transport protein [ <i>Prevotella</i> sp. CAG:1058]	4.00E-88	57	<a href="#">WP_021852386.1</a>
Csd6_c1.15	8807	10099	+	430	phosphotransferase enzyme family protein [ <i>Bacteroides</i> sp. CAG:709]	0	69	<a href="#">WP_022148809.1</a>
Csd6_c1.16	10102	10842	+	246	Mannose-1-phosphate guanyltransferase [ <i>Alistipes</i> sp. CAG:435]	1.00E-101	59	<a href="#">WP_022324822.1</a>
Csd6_c1.17	11127	11939	-	270	membrane protein [ <i>Ruminococcus gnavus</i> ]	4.00E-124	64	<a href="#">WP_039959598.1</a>
Csd6_c1.19	12263	12607	-	114	hydrogenase nickel incorporation protein HypA [ <i>Shuttleworthia</i> sp. MSX8B]	5.00E-45	58	<a href="#">WP_044959080.1</a>
Csd6_c1.20	12749	15481	-	910	pyridine nucleotide-disulfide oxidoreductase [ <i>Clostridium</i> sp. CAG:349]	0	67	<a href="#">WP_022320736.1</a>
Csd6_c1.21	15603	16274	-	223	hydrogenase accessory protein HypB [ <i>Firmicutes bacterium</i> CAG:313]	1.00E-118	72	<a href="#">WP_022330185.1</a>
Csd6_c1.22	16368	17291	+	307	hypothetical protein [ <i>Alistipes</i> sp. CAG:514]	3.00E-151	68	<a href="#">WP_021839425.1</a>
Csd6_c1.24	18410	19678	+	422	endo-1,4-beta-mannosidase [ <i>Bacteroides dorei</i> ]	5.00E-161	54	<a href="#">WP_007831873.1</a>
<b>Csd6_c1.25</b>	<b>19691</b>	<b>20569</b>	<b>+</b>	<b>292</b>	<b>cellulase [uncultured bacterium]</b>	<b>6.00E-94</b>	<b>52</b>	<b><a href="#">ADU86906.1</a></b>
Csd6_c1.26	20529	21215	+	228	beta(1,4)-glucan glucanohydrolase [uncultured bacterium]	2.00E-90	62	<a href="#">AFJ44729.1</a>
Csd6_c1.27	21215	21748	+	178	putative beta-mannanase [unidentified microorganism]	7.00E-81	74	<a href="#">CAJ19138.1</a>
Csd6_c2.1	2	724	-	241	hypothetical protein [ <i>Bacteroides</i> sp. CAG:1060]	5.00E-104	79	<a href="#">WP_021844028.1</a>
Csd6_c2.2	797	1660	+	287	dihydropteroate synthase [ <i>Bacteroides</i> sp. CAG:1060]	9.00E-87	52	<a href="#">WP_021843902.1</a>
Csd6_c2.3	2315	3328	+	337	thiamine-monophosphate kinase [ <i>Prolixibacter bellariivorans</i> ]	1.00E-52	41	<a href="#">WP_025866090.1</a>
Csd6_c2.4	3333	4373	+	346	phosphomethylpyrimidine kinase [ <i>Prevotella</i> sp. 10(H)]	1.00E-99	50	<a href="#">WP_029903426.1</a>
Csd6_c2.5	5003	5773	+	256	hydroxyethylthiazole kinase [ <i>Brachyspira murdochii</i> ]	2.00E-71	50	<a href="#">WP_013112591.1</a>
Csd6_c2.6	6753	7163	+	137	thiamine-phosphate pyrophosphorylase [ <i>Prolixibacter bellariivorans</i> ]	5.00E-50	66	<a href="#">WP_025866088.1</a>
Csd6_c3.2	456	2303	+	615	hypothetical protein [ <i>Bacteroides</i> sp. CAG:1060]	0	81	<a href="#">WP_021844069.1</a>
Csd6_c3.3	2230	2895	+	221	alanine racemase [ <i>Alistipes</i> sp. CAG:435]	9.00E-111	76	<a href="#">WP_022323604.1</a>
Csd6_c3.4	2840	3232	+	130	alanine racemase [ <i>Alistipes</i> sp. CAG:514]	1.00E-59	77	<a href="#">WP_021838703.1</a>
Csd6_c3.5	3229	4632	+	467	peptidase C1-like family [ <i>Bacteroides</i> sp. CAG:1060]	0	73	<a href="#">WP_021842871.1</a>
Csd6_c4.1	2	523	-	174	MULTISPECIES: hypothetical protein [ <i>Alistipes</i> ]	2.00E-38	45	<a href="#">WP_015547801.1</a>
Csd6_c4.2	634	1293	+	219	tENA/THI-4 family protein [ <i>Bacteroides</i> sp. CAG:709]	9.00E-55	43	<a href="#">WP_022147260.1</a>
Csd6_c4.3	1528	1803	-	91	rubrerythrin [ <i>Ruminococcaceae bacterium</i> AB4001]	6.00E-52	87	<a href="#">WP_037278365.1</a>



**Tabla 3.** Asignación de ORFs del fósido Csd8

ORF <sup>(a)</sup>	Posición relativa				Descripción [Organismo] <sup>(c, d)</sup>	Homólogo más cercano por BLASTP <sup>(b)</sup>		
	Inicio	Fin	Hebra	Tamaño (aa)		E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	GenBank ID <sup>(c)</sup>
Csd8_c2.1	2	160	+	52	hypothetical protein [Alistipes sp. CAG:514]	5.00E-11	65	<a href="#">WP_021839595.1</a>
Csd8_c2.2	1385	2299	-	304	hypothetical protein [Bacteroides sp. CAG:1060]	9.00E-143	70	<a href="#">WP_021842836.1</a>
Csd8_c2.3	2360	2821	+	153	hypothetical protein [Bacteroides sp. CAG:1060]	1.00E-50	54	<a href="#">WP_021842835.1</a>
Csd8_c2.4	2802	3377	-	191	hypothetical protein [Bacteroides sp. CAG:1060]	5.00E-99	74	<a href="#">WP_021843596.1</a>
Csd8_c2.5	3456	4385	+	309	sLEI family protein [Bacteroides sp. CAG:1060]	7.00E-102	59	<a href="#">WP_021843594.1</a>
Csd8_c2.6	4385	5914	+	509	protease [Bacteroides sp. CAG:545]	0	57	<a href="#">WP_022016849.1</a>
Csd8_c2.7	5920	6204	+	94	probable collagenase [Bacteroides sp. CAG:1060]	3.00E-15	41	<a href="#">WP_021843592.1</a>
Csd8_c2.8	6191	6949	-	252	hypothetical protein [Bacteroides sp. CAG:1060]	1.00E-91	53	<a href="#">WP_021843591.1</a>
Csd8_c2.9	7530	8384	-	284	protein-(Glutamine-N5) methyltransferase [Bacteroides sp. CAG:545]	4.00E-90	51	<a href="#">WP_022016851.1</a>
Csd8_c2.10	8469	9485	-	338	d-alanine-D-alanine ligase [Bacteroides sp. CAG:1060]	0	73	<a href="#">WP_021843598.1</a>
Csd8_c2.12	11989	14301	+	770	dNA topoisomerase [Bacteroides sp. CAG:1060]	0	72	<a href="#">WP_021843600.1</a>
Csd8_c2.13	14312	14812	+	166	hypothetical protein [Bacteroides sp. CAG:1060]	1.00E-92	76	<a href="#">WP_021843601.1</a>
Csd8_c2.14	14796	15425	-	209	4-diphosphocytidyl-2-C-methyl-D-erythritol synthase [Bacteroides sp. CAG:1060]	4.00E-93	64	<a href="#">WP_021843602.1</a>
Csd8_c2.15	15550	15768	+	72	putative rod shape-determining protein [Bacteroides sp. CAG:709]	2.00E-27	84	<a href="#">WP_022147119.1</a>
Csd8_c2.16	15765	16580	+	271	putative rod shape-determining protein [Alistipes sp. CAG:435]	2.00E-159	82	<a href="#">WP_022323730.1</a>
<b>Csd8_c5.2</b>	<b>815</b>	<b>2104</b>	<b>-</b>	<b>429</b>	<b>glycosyl hydrolase family 43 [Bacteroides sp. CAG:633]</b>	<b>0</b>	<b>63</b>	<b><a href="#">WP_022162901.1</a></b>
Csd8_c5.4	2499	3398	+	299	nLP/P60 protein [Alistipes sp. CAG:435]	1.00E-63	36	<a href="#">WP_022323007.1</a>
Csd8_c5.5	3426	3806	+	126	hypothetical protein [Bacteroides sp. D20]	1.00E-29	49	<a href="#">WP_009037653.1</a>
Csd8_c5.6	3799	4497	+	232	fHA domain protein [Prevotella copri CAG:164]	3.00E-21	30	<a href="#">WP_022121752.1</a>
Csd8_c5.7	4605	5096	+	164	kinase domain protein [Bacteroides sp. D20]	2.00E-62	60	<a href="#">EFA20480.1</a>
Csd8_c8.2	184	1245	+	353	kinase domain protein [Bacteroides sp. D20]	1.00E-63	62	<a href="#">EFA20477.1</a>
Csd8_c8.3	1284	1835	+	183	hypothetical protein [Bacteroides sp. D20]	2.00E-66	62	<a href="#">WP_034529372.1</a>
Csd8_c8.5	4098	4400	+	100	hypothetical protein [Hallella seregens]	1.00E-20	47	<a href="#">WP_027953067.1</a>
Csd8_c8.6	4522	5622	+	366	hypothetical protein [Alistipes sp. CAG:514]	9.00E-148	60	<a href="#">WP_021839010.1</a>
Csd8_c8.7	5543	5980	+	145	hypothetical protein [Bacteroides sp. CAG:709]	9.00E-58	77	<a href="#">WP_022147836.1</a>
Csd8_c8.8	5950	6687	+	245	3-oxoacyl-[acyl-carrier-protein] reductase [Bacteroides sp. CAG:1060]	8.00E-164	93	<a href="#">WP_021843309.1</a>
Csd8_c8.9	6692	7123	+	144	comF family protein [Bacteroides sp. CAG:1060]	7.00E-47	58	<a href="#">WP_021843310.1</a>
Csd8_c11.1	352	846	+	164	transposase DDE domain protein [Escherichia coli 2362-75]	6.00E-115	100	<a href="#">EFR18021.1</a>
Csd8_c11.2	874	1290	+	138	transposase [Escherichia coli]	2.00E-95	100	<a href="#">WP_032161000.1</a>

**Tabla 3, continuación.** Asignación de ORFs del fósido Csd8

ORF <sup>(a)</sup>	Posición relativa		Homólogo más cercano por BLASTP <sup>(b)</sup>					
	Inicio	Fin	Hebra	Tamaño (aa)	Descripción [Organismo] <sup>(c, d)</sup>	E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	GenBank ID <sup>(c)</sup>
Csd8_c16.1	662	1021	-	119	ribosomal protein L31 [Escherichia coli 910096-2]	9.00E-62	89	ESE27907.1
Csd8_r42.1	49	1965	+	638	serine/threonine protein kinase [Prevotella sp. CAG:474]	9.00E-180	47	WP_022310834.1
Csd8_r42.3	4010	4690	+	226	caspase [Bacteroides sp. D20]	8.00E-88	57	WP_009037639.1
Csd8_rc42.4	4684	5217	+	177	FHA domain protein [Bacteroides sp. D20]	2.00E-27	36	EFA20471.1
Csd8_c42.5	5214	6131	+	305	hypothetical protein HMPREF0969_01883 [Bacteroides sp. D20]	6.00E-98	55	EFA20470.1
Csd8_c42.6	6178	7257	+	359	predicted protein [Bacteroides sp. D20]	2.00E-26	33	WP_009037640.1
Csd8_c42.7	7457	7912	+	151	hypothetical protein [Bacteroides uniformis]	1.00E-46	49	WP_035458133.1
Csd8_c42.8	8041	8202	+	54	hypothetical protein [Bacteroides sp. D20]	1.00E-18	74	WP_034529372.1

<sup>(a)</sup> ORFs identificados por MetaGene.

<sup>(b)</sup> Asignados por RPS Blast.

<sup>(c)</sup> Homólogos más cercanos identificados por BlastP 2.2.30+ usando secuencias de proteínas no redundantes.

<sup>(d)</sup> Los genes asociados a la degradación de lignocelulosa o de particular interés se muestran en negritas.

**Tabla 4.** Asignación de ORFs del fósrido Csd9

ORF <sup>(a)</sup>	Posición relativa				Descripción [Organismo] <sup>(c,d)</sup>	Homólogo más cercano por BLASTP <sup>(b)</sup>		
	Inicio	Fin	Hebra	Tamaño (aa)		E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	GenBank ID <sup>(c)</sup>
Csd9_c1.1	2	706	+	234	TonB-dependent receptor [ <i>Prevotella ruminicola</i> ]	3.00E-136	87	<a href="#">WP_013064632.1</a>
Csd9_c1.2	744	2465	+	573	putative lipoprotein	0	90	<a href="#">AIF25919.1</a>
Csd9_c1.3	2501	4387	+	628	putative SusD2	0	98	<a href="#">AIF25920.1</a>
<b>Csd9_c1.5</b>	<b>4796</b>	<b>6040</b>	+	<b>414</b>	<b>putative cellulase</b>	<b>0</b>	<b>98</b>	<b><a href="#">AIF25922.1</a></b>
Csd9_c1.6	6102	7310	-	402	putative endoglycosidase precursor protein	0	99	<a href="#">AIF25923.1</a>
Csd9_c1.8	8259	8762	+	167	putative transcriptional regulator	5.00E-105	94	<a href="#">AIF25924.1</a>
Csd9_c1.10	9131	11575	-	814	putative spore coat protein CotH	0	99	<a href="#">AIF25925.1</a>
Csd9_c1.11	11556	12269	-	237	putative lipoprotein	3.00E-171	98	<a href="#">AIF25934.1</a>
Csd9_c1.12	12266	13423	-	385	putative hypothetical protein PRU_2234	0	99	<a href="#">AIF25926.1</a>
Csd9_c1.13	13420	14412	-	330	putative D-alanine-D-alanine ligase	0	100	<a href="#">AIF25927.1</a>
Csd9_c1.14	14420	15448	-	342	putative ribosomal large subunit pseudouridine synthase, RluA family	0	100	<a href="#">AIF25928.1</a>
Csd9_c1.15	15445	16107	-	220	putative PASTA domain-containing protein	1.00E-160	100	<a href="#">AIF25929.1</a>
Csd9_c1.16	16568	17134	+	188	putative translation elongation factor P	3.00E-133	100	<a href="#">AIF25931.1</a>
Csd9_c1.17	17134	18216	+	360	putative group 2 family glycosyltransferase	0	98	<a href="#">AIF25932.1</a>
Csd9_c1.18	18241	18927	+	228	hypothetical protein [ <i>Prevotella sp.</i> FD3004]	8.00E-137	83	<a href="#">WP_036911624.1</a>
Csd9_c1.19	18936	20474	-	512	cell surface protein [ <i>Prevotella melaninogenica</i> ]	2.00E-48	28	<a href="#">WP_013265240.1</a>
Csd9_c1.20	20587	22224	-	545	hypothetical protein [ <i>Prevotella ruminicola</i> ]	0	61	<a href="#">WP_041386176.1</a>
Csd9_c1.21	22997	23317	+	106	hypothetical protein [ <i>Prevotella sp.</i> Sc00028]	3.00E-66	93	<a href="#">AGH14007.1</a>
Csd9_c1.22	23401	23880	+	159	MULTISPECIES: UDP-2,3-diacetylglucosamine hydrolase [ <i>Prevotella</i> ]	5.00E-91	95	<a href="#">WP_028906659.1</a>
Csd9_c1.23	23840	24103	+	87	hypothetical protein [ <i>Prevotella sp.</i> Sc00028]	1.00E-52	95	<a href="#">AGH14006.1</a>
Csd9_c1.24	24117	25172	+	351	2-dehydro-3-deoxygluconokinase [ <i>Prevotella sp.</i> Sc00028]	0	90	<a href="#">AGH14005.1</a>
Csd9_c1.25	25247	25765	-	172	hypothetical protein [ <i>Prevotella sp.</i> FD3004]	4.00E-63	68	<a href="#">WP_036913509.1</a>
Csd9_c1.26	25942	27078	-	378	acetate kinase [ <i>Prevotella sp.</i> MA2016]	0	92	<a href="#">WP_028912032.1</a>
Csd9_c1.27	27163	28173	-	336	phosphate acetyltransferase [ <i>Prevotella sp.</i> Sc00028]	0	90	<a href="#">AGH14002.1</a>
Csd9_c1.28	28207	28995	-	262	membrane protein [ <i>Prevotella sp.</i> FD3004]	3.00E-152	81	<a href="#">WP_036911614.1</a>
Csd9_c1.29	29008	29790	-	260	MULTISPECIES: dihydropteroate synthase [ <i>Prevotella</i> ]	2.00E-137	72	<a href="#">WP_028904079.1</a>
Csd9_c1.30	29953	30939	+	328	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase [ <i>Prevotella ruminicola</i> ]	0	86	<a href="#">WP_028906649.1</a>
Csd9_c1.31	30944	31315	+	123	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase [ <i>Prevotella sp.</i> ]	2.00E-54	74	<a href="#">WP_036911612.1</a>

**Tabla 4, continuación.** Asignación de ORFs del fósido Csd9

ORF <sup>(a)</sup>	Posición relativa				Homólogo más cercano por BLASTP <sup>(b)</sup>			
	Inicio	Fin	Hebra	Tamaño (aa)	Descripción [Organismo] <sup>(c,d)</sup>	E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	GenBank ID <sup>(c)</sup>
Csd9_c1.34	33318	34394	+	358	hypothetical protein [ <i>Prevotella sp.</i> FD3004]	4.00E-137	57	<a href="#">WP_036911609.1</a>
Csd9_c1.32	31333	32343	+	336	MULTISPECIES: hypothetical protein [ <i>Prevotella</i> ]	6.00E-150	75	<a href="#">WP_033147334.1</a>
Csd9_c1.35	34664	36088	-	474	aldo/keto reductase [ <i>Prevotella sp.</i> MA2016]	0	86	<a href="#">WP_028911009.1</a>
Csd9_c1.36	36094	36345	-	83	ferredoxin [ <i>Prevotella ruminicola</i> ]	2.00E-38	94	<a href="#">WP_028906450.1</a>

<sup>(a)</sup> ORFs identificados por MetaGene.

<sup>(b)</sup> Asignados por RPS Blast.

<sup>(c)</sup> Homólogos más cercanos identificados por BlastP 2.2.30+ usando secuencias de proteínas no redundantes.

<sup>(d)</sup> Los genes asociados a la degradación de lignocelulosa o de particular interés se muestran en negritas.

**Tabla 6.** Marcos abiertos de lectura asignados al fósido Csd18

ORF <sup>(a)</sup>	Posición relativa				Descripción [Organismo] <sup>(c,d)</sup>	Homólogo más cercano por BLASTP <sup>(b)</sup>		
	Inicio	Fin	Hebra	Tamaño (aa)		E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	Gene Bank ID <sup>(c)</sup>
Csd18_c1.3	3655	4065	+	136	ATPase AAA [ <i>Bacteroides barnesiae</i> ]	6.00E-54	72	<a href="#">WP_018709390.1</a>
Csd18_c1.4	4119	5321	+	400	ATPase AAA [ <i>Bacteroides barnesiae</i> ]	0	64	<a href="#">WP_018709390.1</a>
Csd18_c1.5	5341	5784	+	147	pyridoxamine 5-phosphate oxidase [ <i>Alistipes</i> sp. AL-1]	2.00E-37	50	<a href="#">WP_032134625.1</a>
Csd18_c1.6	5808	6629	-	273	hypothetical protein [ <i>Prevotella</i> sp. Sc00028]	5.00E-90	51	<a href="#">AGH13984.1</a>
Csd18_c1.7	6852	7595	-	247	MULTISPECIES: hypothetical protein [ <i>Bacteroidales</i> ]	1.00E-34	37	<a href="#">WP_007850563.1</a>
<b>Csd18_c1.8</b>	<b>8805</b>	<b>9732</b>	-	<b>309</b>	<b>putative carbohydrate-active enzyme [uncultured organism]</b>	<b>0</b>	<b>91</b>	<b><a href="#">ADX05725.1</a></b>
Csd18_c1.9	9900	10769	-	289	butanol dehydrogenase [ <i>Prevotella</i> sp. MA2016]	4.00E-57	79	<a href="#">WP_028911299.1</a>
Csd18_c1.10	10694	11080	-	128	ATP-dependent helicase [ <i>Bacteroides pyogenes</i> ]	0	41	<a href="#">WP_027326181.1</a>
Csd18_c1.11	11126	14332	-	1068	MULTISPECIES: 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Bacteroides</i> ]	5.00E-154	78	<a href="#">WP_005777498.1</a>
Csd18_c1.12	14436	15254	+	272	dEAD/DEAH box helicase domain protein [ <i>Bacteroides</i> sp. CAG:702]	0	71	<a href="#">WP_022040023.1</a>
Csd18_c1.13	15415	16641	-	408	permease YjgP/YjgQ family [ <i>Bacteroides</i> sp. CAG:633]	6.00E-107	60	<a href="#">WP_022162475.1</a>
Csd18_c1.14	16652	17437	-	261	queuine tRNA-ribosyltransferase [ <i>Bacteroides</i> sp. CAG:530]	0	83	<a href="#">WP_022129645.1</a>
Csd18_c1.15	17776	18813	-	345	lon protease [ <i>Bacteroides</i> sp. CAG:462]	0	71	<a href="#">WP_022383975.1</a>
Csd18_c1.16	18920	21094	-	724	diacylglycerol kinase [ <i>Prevotella copri</i> ]	1.00E-37	45	<a href="#">WP_006849117.1</a>
Csd18_c1.17	22232	22639	-	135	diacylglycerol kinase [ <i>Prevotella</i> sp. CAG:1185]	2.00E-42	55	<a href="#">WP_021964078.1</a>
Csd18_c1.18	22696	23130	-	144	2-amino-3-ketobutyrate CoA ligase [ <i>Bacteroides graminisolvens</i> ]	3.00E-99	79	<a href="#">WP_024996973.1</a>
Csd18_c1.19	23274	23813	+	179	2-amino-3-ketobutyrate CoA ligase [ <i>Bacteroides graminisolvens</i> ]	3.00E-99	79	<a href="#">WP_024996973.1</a>
Csd18_c1.20	25264	26565	+	433	arginine decarboxylase [ <i>Clostridium nexile</i> ]	0	75	<a href="#">WP_004613076.1</a>
Csd18_c1.21	26578	27441	+	287	spermidine synthase [ <i>Clostridium</i> sp. CAG:510]	1.00E-167	77	<a href="#">WP_022124011.1</a>
Csd18_c1.22	27445	28302	+	285	agmatinase [ <i>Robinsoniella</i> sp. KNHs210]	4.00E-151	71	<a href="#">WP_027293336.1</a>
Csd18_c1.23	28505	28690	+	61	saccharopine dehydrogenase [ <i>Firmicutes bacterium</i> CAG:646]	6.00E-22	77	<a href="#">WP_022009922.1</a>
Csd18_c1.24	28925	29767	+	280	carboxynorspermidine dehydrogenase [butyrate-producing bacterium SS3/4]	0	89	<a href="#">WP_015553675.1</a>
Csd18_c1.25	29779	30930	+	383	carboxynorspermidine decarboxylase [ <i>Clostridium</i> sp. CAG:510]	0	68	<a href="#">WP_022124008.1</a>
Csd18_c1.26	31440	31895	-	151	hypothetical protein [ <i>Alistipes</i> sp. CAG:268]	4.00E-18	45	<a href="#">WP_022308780.1</a>
Csd18_c1.27	34227	35228	-	333	toxin HipA [ <i>Prevotella dentasini</i> ]	0	80	<a href="#">WP_025002844.1</a>
Csd18_c1.28	35231	35539	-	102	phosphatidylinositol kinase [ <i>Prevotella</i> sp. RM4]	5.00E-56	79	<a href="#">WP_033150366.1</a>
Csd18_c1.29	35526	35747	-	73	transcriptional regulator [ <i>Prevotella dentasini</i> ]	3.00E-33	81	<a href="#">WP_025002842.1</a>

**Tabla 7.** Marcos abiertos de lectura asignados al fósido Xil3

Posición relativa					Homólogo más cercano por BLASTP <sup>(b)</sup>			
ORF <sup>(a)</sup>	Inicio	Fin	Hebra	Tamaño (aa)	Descripción [Organismo] <sup>(c,d)</sup>	E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	GenBank ID <sup>(c)</sup>
Xil3_c1.1	3	1331	-	443	putative outer membrane protein [ <i>Alistipes sp.</i> CAG:29]	2.00E-165	57	<a href="#">WP_022332711.1</a>
Xil3_c1.2	1234	1872	-	212	putative outer membrane protein [ <i>Alistipes sp.</i> CAG:29]	1.00E-47	56	<a href="#">WP_022332711.1</a>
Xil3_c1.3	1890	3725	-	611	ragB/SusD domain protein [ <i>Alistipes sp.</i> CAG:29]	0	60	<a href="#">WP_022332712.1</a>
Xil3_c1.4	3729	7016	-	1095	putative outer membrane protein [ <i>Alistipes sp.</i> CAG:29]	0	65	<a href="#">WP_022332713.1</a>
<b>Xil3_c1.5</b>	<b>7029</b>	<b>8507</b>	-	<b>492</b>	<b>glycosyl hydrolase family 26 [<i>Alistipes sp.</i> CAG:29]</b>	<b>1.00E-164</b>	<b>63</b>	<b><a href="#">WP_022332714.1</a></b>
Xil3_c1.6	8511	9557	-	348	hypothetical protein [ <i>Alistipes sp.</i> CAG:29]	3.00E-52	38	<a href="#">WP_022332715.1</a>
Xil3_c1.7	9577	10992	-	471	putative lipoprotein [ <i>Alistipes sp.</i> CAG:53]	0	56	<a href="#">WP_022061518.1</a>
<b>Xil3_c1.8</b>	<b>11127</b>	<b>12152</b>	-	<b>341</b>	<b>glycosidase [<i>Alistipes timonensis</i>]</b>	<b>0</b>	<b>73</b>	<b><a href="#">WP_010262638.1</a></b>
Xil3_c1.9	12155	13315	-	386	sodium transporter [ <i>Alistipes shahii</i> ]	0	73	<a href="#">WP_044053978.1</a>
Xil3_c1.10	13882	14742	+	286	AraC family transcriptional regulator [ <i>Alistipes timonensis</i> ]	3.00E-140	68	<a href="#">WP_010262631.1</a>
Xil3_c1.11	14749	15729	-	326	endoribonuclease L-PSP [ <i>Bacteroides fragilis</i> ]	4.00E-112	53	<a href="#">WP_032531152.1</a>
Xil3_c1.12	15934	16563	+	209	membrane protein [ <i>Bacteroides uniformis</i> ]	1.00E-27	38	<a href="#">WP_044468041.1</a>
Xil3_c1.13	16554	17726	+	390	membrane protein [ <i>Parabacteroides distasonis</i> ]	7.00E-99	43	<a href="#">WP_036611773.1</a>
Xil3_c1.14	18009	19193	+	394	MULTISPECIES: cytochrome C-binding protein [ <i>Bacteroides</i> ]	5.00E-180	59	<a href="#">WP_031535594.1</a>
<b>Xil3_c1.15</b>	<b>19194</b>	<b>20984</b>	-	<b>596</b>	<b>putative carbohydrate-active enzyme [uncultured organism]</b>	<b>0</b>	<b>86</b>	<b><a href="#">ADX05729.1</a></b>
Xil3_c1.16	20977	22446	-	489	cytochrome C-binding protein [ <i>Bacteroidaceae bacterium MS4</i> ]	3.00E-173	52	<a href="#">WP_042368667.1</a>
Xil3_c1.17	22421	23488	-	355	hypothetical protein [ <i>Bacteroides uniformis</i> ]	2.00E-72	44	<a href="#">WP_016273963.1</a>
Xil3_c1.18	23568	24038	-	156	putative lipoprotein [ <i>Prevotella sp.</i> CAG:604]	1.00E-26	42	<a href="#">WP_022111191.1</a>
Xil3_c1.19	24035	25486	-	483	cytochrome C-binding protein [ <i>Prevotella ruminicola</i> ]	2.00E-175	51	<a href="#">WP_041386827.1</a>
Xil3_c1.20	25483	26943	-	486	hypothetical protein PRU_2221 [ <i>Prevotella ruminicola</i> 23]	8.00E-88	36	<a href="#">ADE81782.1</a>
Xil3_c1.21	27005	28468	-	487	MULTISPECIES: hypothetical protein [ <i>Bacteroides</i> ]	5.00E-92	39	<a href="#">WP_031535636.1</a>
Xil3_c1.22	28491	30275	-	594	starch-binding protein [ <i>Bacteroides graminisolvens</i> ]	0	60	<a href="#">WP_024995911.1</a>
Xil3_c1.23	30293	33919	-	1208	collagen-binding protein [ <i>Bacteroides graminisolvens</i> ]	0	60	<a href="#">WP_024995910.1</a>
Xil3_c1.24	33989	35329	-	446	sugar transporter [ <i>Prevotella sp.</i> Sc00066]	0	71	<a href="#">AGH14131.1</a>
<b>Xil3_c1.25</b>	<b>35389</b>	<b>36006</b>	-	<b>205</b>	<b>glycosidase [uncultured microorganism]</b>	<b>8.00E-135</b>	<b>92</b>	<b><a href="#">ADB80102.1</a></b>

<sup>(a)</sup> ORFs identificados por MetaGene.

<sup>(b)</sup> Asignados por RPS Blast.

<sup>(c)</sup> Homólogos más cercanos identificados por BlastP 2.2.30+ usando secuencias de proteínas no redundantes.

<sup>(d)</sup> Los genes asociados a la degradación de lignocelulosa o de particular interés se muestran en negritas.

**Tabla 8.** Asignación de ORFs del fósido Csd23

ORF <sup>(a)</sup>	Posición relativa			Tamaño (aa)	Descripción [Organismo] <sup>(c,d)</sup>	Homólogo más cercano por BLASTP <sup>(b)</sup>		
	Inicio	Fin	Hebra			E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	GenBank ID <sup>(c)</sup>
Csd23_1	2	121	-	40	hypothetical protein GOS_2224636 [marine metagenome]	3	46	<a href="#">ECZ20017.1</a>
Csd23_2	140	2101	-	653	DNA topoisomerase IV subunit B [ <i>Owenweeksia hongkongensis</i> ]	0	63	<a href="#">WP_014203663.1</a>
Csd23_3	2175	3488	-	437	aspartate kinase [ <i>Cytophagales bacterium</i> B6]	2.00E-124	46	<a href="#">WP_022829496.1</a>
Csd23_4	3460	4485	-	341	tRNA threonylcarbamoyladenosine modification protein Tsad [ <i>Marinilabilia salmonicolor</i> ]	1.00E-168	67	<a href="#">WP_010662731.1</a>
Csd23_5	4538	9016	+	1492	hypothetical protein [ <i>Flavobacteriales bacterium</i> BRH_c54]	3.00E-131	25	<a href="#">WP_045639617.1</a>
Csd23_6	9016	12288	+	1090	ATP-dependent helicase [ <i>Flavobacterium reichenbachii</i> ]	9.00E-136	30	<a href="#">WP_035681601.1</a>
Csd23_7	12349	14166	-	605	aminoacyl-histidine dipeptidase [ <i>Butyricimonas virosa</i> ]	0	56	<a href="#">WP_027202408.1</a>
Csd23_8	14171	14818	-	215	ribulose-phosphate 3-epimerase [ <i>Elizabethkingia</i> sp. BM10]	2.00E-110	73	<a href="#">WP_045179381.1</a>
Csd23_9	14905	15768	-	287	RNA polymerase sigma factor rpoD [ <i>Empedobacter brevis</i> ]	6.00E-177	84	<a href="#">WP_019975176.1</a>
Csd23_10	15839	18103	-	754	polynucleotide phosphorylase [ <i>Bacteroidales bacterium</i> ph8]	0	66	<a href="#">WP_019131579.1</a>
Csd23_11	18231	18500	-	89	MULTISPECIES: 30S ribosomal protein S15 [ <i>Odoribacter</i> ]	6.00E-37	71	<a href="#">WP_009135403.1</a>
Csd23_12	18620	19024	-	134	hypothetical protein [ <i>Bacteroides barnesiae</i> ]	2.00E-58	73	<a href="#">WP_025892370.1</a>
Csd23_13	19037	20605	-	522	putative Na(+)/H(+) antiporter [ <i>Methanocella arvoryzae</i> ]	3.00E-98	44	<a href="#">WP_012034436.1</a>
Csd23_14	20618	22351	-	577	hypothetical protein [Bacteroidetes oral taxon 274]	2.00E-30	25	<a href="#">WP_040310866.1</a>
Csd23_15	28086	28751	-	221	putative ATP-dependent DNA helicase [ <i>Geofilum rubicundum</i> JCM 15548]	6.00E-73	52	<a href="#">GAO30149.1</a>
Csd23_16	29409	29858	+	149	DNA-binding protein [ <i>Zunongwangia profunda</i> ]	7.00E-19	39	<a href="#">WP_013072140.1</a>
Csd23_17	29893	30945	+	350	exopolyphosphatase [ <i>Gillisia</i> sp. JM1]	6.00E-90	47	<a href="#">WP_026838988.1</a>
Csd23_18	30939	31463	+	174	peptidylprolyl isomerase [ <i>Anaerophaga thermohalophila</i> ]	1.00E-30	38	<a href="#">WP_010418430.1</a>
Csd23_19	31466	32797	+	443	70 kDa peptidylprolyl isomerase [ <i>Microscilla marina</i> ATCC 23134]	8.00E-43	28	<a href="#">EAY30224.1</a>
Csd23_20	32849	33760	+	303	peptidylprolyl isomerase [ <i>Runella limosa</i> ]	8.00E-47	38	<a href="#">WP_028525026.1</a>
Csd23_21	33762	35078	+	438	peptidylprolyl isomerase [ <i>Runella limosa</i> ]	8.00E-47	38	<a href="#">WP_028525026.1</a>
Csd23_22	35131	36570	-	479	hypothetical protein [ <i>Flavobacterium</i> sp. MS220-5C]	1.00E-140	43	<a href="#">WP_017843386.1</a>
Csd23_23	36574	36795	-	73	hypothetical protein [ <i>Draconibacterium</i> sp. JN14CK-3]	4.00E-10	41	<a href="#">WP_045026565.1</a>

<sup>(a)</sup> ORFs identificados por MetaGene.

<sup>(b)</sup> Asignados por RPS Blast.

<sup>(c)</sup> Homólogos más cercanos identificados por BlastP 2.2.30+ usando secuencias de proteínas no redundantes.

<sup>(d)</sup> Los genes asociados a la degradación de lignocelulosa o de particular interés se muestran en negritas.

**Tabla 9.** Marcos abiertos de lectura asignados al fósido M9F1

ORF <sup>(a)</sup>	Posición relativa		Hebra	Tamaño (aa)	Homólogo más cercano por BLASTP <sup>(b)</sup>		E-value <sup>(c)</sup>	Identidad (%) <sup>(c)</sup>	GenBank ID <sup>(c)</sup>
	Inicio	Fin			Descripción [Organismo] <sup>(c,d)</sup>				
M9F1_1	4569	5243	-	224		hypothetical protein [ <i>Syntrophomonas wolfei</i> ]	3.00E-80	57	<a href="#">WP_011640352.1</a>
M9F1_2	5442	6821	+	459		glutamyl-tRNA synthetase [ <i>Syntrophomonas wolfei</i> ]	0	56	<a href="#">WP_011640351.1</a>
<b>M9F1_3</b>	<b>7101</b>	<b>7796</b>	-	<b>231</b>		<b>MULTISPECIES: glycosyltransferase family 2 [<i>Pelosinus</i>]</b>	<b>2.00E-82</b>	<b>54</b>	<b><a href="#">WP_007930628.1</a></b>
<b>M9F1_4</b>	<b>7941</b>	<b>10037</b>	-	<b>698</b>		<b>glycosyl transferase family 2 [<i>Desulfosporosinus sp. Tol-M</i>]</b>	<b>0</b>	<b>72</b>	<b><a href="#">WP_034614419.1</a></b>
<b>M9F1_5</b>	<b>10055</b>	<b>11140</b>	-	<b>361</b>		<b>UDP-N-acetylglucosamine 2-epimerase [<i>Desulfosporosinus youngiae</i>]</b>	<b>7.00E-168</b>	<b>64</b>	<b><a href="#">WP_007778681.1</a></b>
<b>M9F1_6</b>	<b>11137</b>	<b>12201</b>	-	<b>354</b>		<b>polysaccharide biosynthesis protein [<i>Acetonema longum</i>]</b>	<b>0</b>	<b>70</b>	<b><a href="#">WP_004573041.1</a></b>
<b>M9F1_7</b>	<b>12202</b>	<b>13173</b>	-	<b>323</b>		<b>dTDP-4-dehydrorhamnose reductase [<i>Geobacter daltonii</i>]</b>	<b>3.00E-110</b>	<b>56</b>	<b><a href="#">WP_012646545.1</a></b>
M9F1_8	13880	14677	+	265		hypothetical PaecuDRAFT_3010 [ <i>Paenibacillus curdlanolyticus</i> YK9]	9.00E-14	31	<a href="#">EFM10051.1</a>
M9F1_9	14785	15240	+	151		BclB [ <i>Desulfosporosinus youngiae</i> ]	9.00E-33	55	<a href="#">WP_007778663.1</a>
M9F1_10	15418	17403	+	661		glycosyl transferase family protein [ <i>Paenibacillus sp. FSL H8-237</i> ]	1.00E-168	44	<a href="#">ETT67520.1</a>
M9F1_11	17333	17656	+	107		tRNA pseudouridine32 synthase [ <i>Acinetobacter indicus</i> ]	5.00E-70	100	<a href="#">WP_016658238.1</a>
M9F1_12	17768	18103	+	111		XRE family transcriptional regulator [ <i>Acinetobacter sp. MDS7A</i> ]	4.00E-53	79	<a href="#">WP_019837413.1</a>
M9F1_13	18196	18639	+	147		acetyltransferase [ <i>Acinetobacter sp. MDS7A</i> ]	5.00E-88	84	<a href="#">WP_026438150.1</a>
M9F1_14	18661	19371	-	236		MFS transporter [ <i>Acinetobacter baumannii</i> ]	2.00E-136	82	<a href="#">WP_034702289.1</a>
M9F1_15	19523	19975	+	150		UPF0756 membrane protein [ <i>Acinetobacter sp. CIP 53.82</i> ]	2.00E-95	99	<a href="#">ENW90744.1</a>
M9F1_16	20019	21200	-	393		Major Facilitator Superfamily; pfam07690 [ <i>Acinetobacter sp. CIP 53.82</i> ]	0	100	<a href="#">WP_005177440.1</a>
M9F1_17	21374	22294	-	306		LysR family transcriptional regulator [ <i>Acinetobacter indicus</i> ]	0	99	<a href="#">WP_016658241.1</a>
M9F1_18	22548	24263	-	571		hypothetical protein [ <i>Acinetobacter sp. CIP 53.82</i> ]	0	100	<a href="#">WP_005177435.1</a>
M9F1_19	24351	26369	-	672		hypothetical protein [ <i>Acinetobacter indicus</i> ]	0	99	<a href="#">WP_016658243.1</a>
M9F1_20	26824	27996	+	390		alcohol dehydrogenase [ <i>Acinetobacter indicus</i> ]	0	99	<a href="#">WP_016658244.1</a>
M9F1_21	28207	29238	+	343		alpha/beta hydrolase [ <i>Acinetobacter indicus</i> ]	0	99	<a href="#">WP_035363077.1</a>
M9F1_22	29296	30477	-	393		transcriptional regulator [ <i>Acinetobacter sp. MDS7A</i> ]	0	99	<a href="#">WP_019837401.1</a>
M9F1_23	30723	32234	+	503		aldehyde dehydrogenase [ <i>Acinetobacter indicus</i> ]	0	100	<a href="#">WP_016658247.1</a>
M9F1_24	32472	33182	+	236		GMP synthase (glutamine-hydrolysing) [ <i>Acinetobacter indicus</i> ]	2.00E-167	99	<a href="#">WP_016658248.1</a>
<b>M9F1_25</b>	<b>33268</b>	<b>33852</b>	-	<b>194</b>		<b>YceI-like domain protein [<i>Acinetobacter sp. WC-743</i>]</b>	<b>8.00E-100</b>	<b>77</b>	<b><a href="#">WP_009584260.1</a></b>

<sup>(a)</sup> ORFs identificados por MetaGene.

<sup>(b)</sup> Asignados por RPS Blast.

<sup>(c)</sup> Homólogos más cercanos identificados por BlastP 2.2.30+ usando secuencias de proteínas no redundantes.

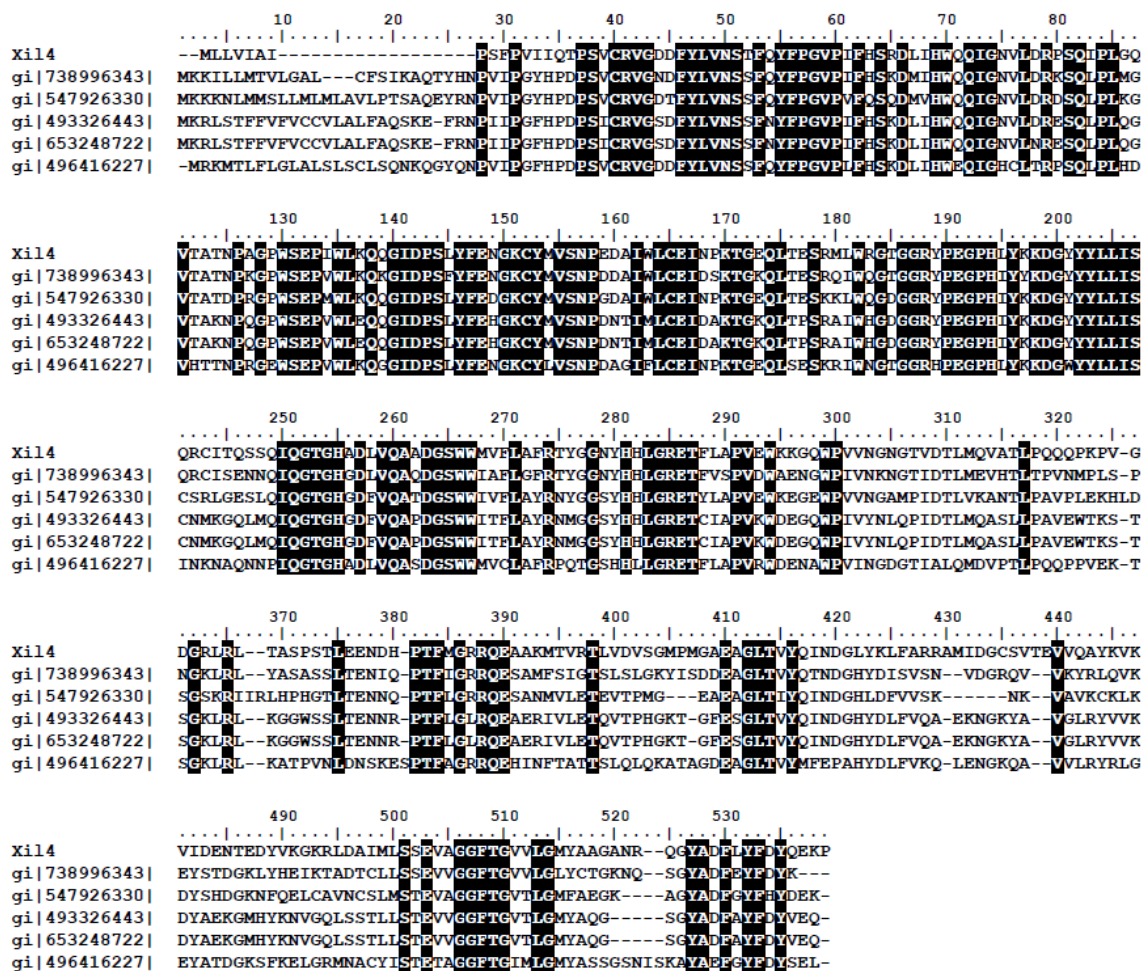
<sup>(d)</sup> El posible operón involucrado en el metabolismo de polisacáridos y el ORF con el dominio YceI se muestran en negritas.



Anexo 2.

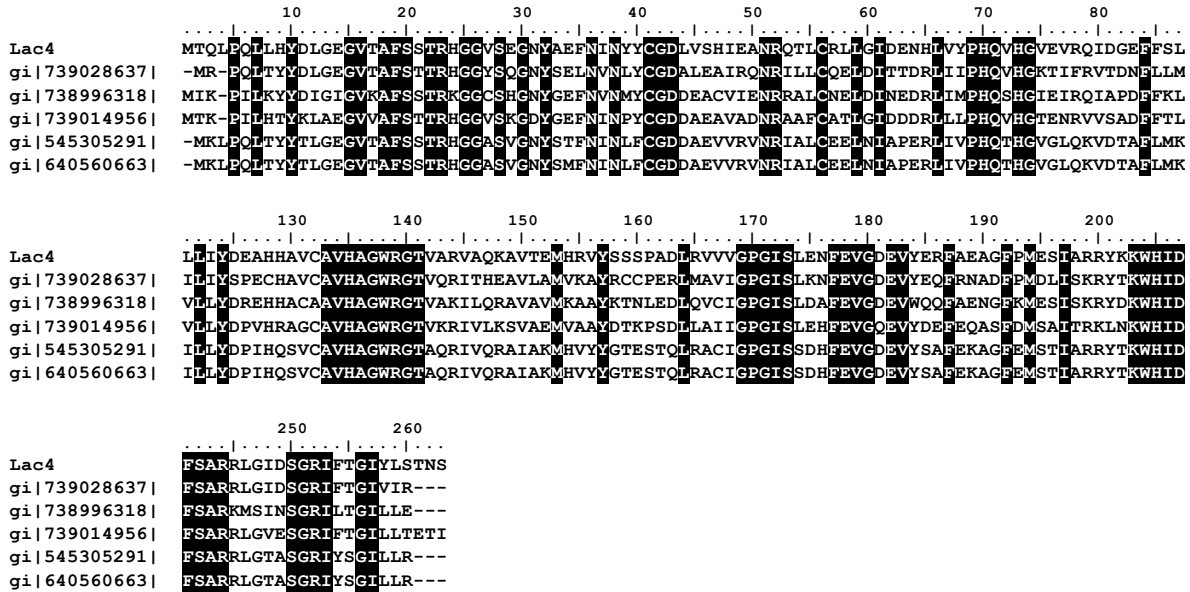
ALINEAMIENTO DE LA SECUENCIA DE AMINOÁCIDOS IDENTIFICADOS COMO RESPONSABLES DE LA ACTIVIDAD CELULOLÍTICA O XILANOLÍTICA EN LOS CLONES SELECCIONADOS

Figura 1. Comparación de la secuencia aminoacídica de Xil4 y sus homólogos más cercanos. Se muestran los números de acceso del GenBank. Las regiones de identidad entre las secuencias se muestran en columnas negras



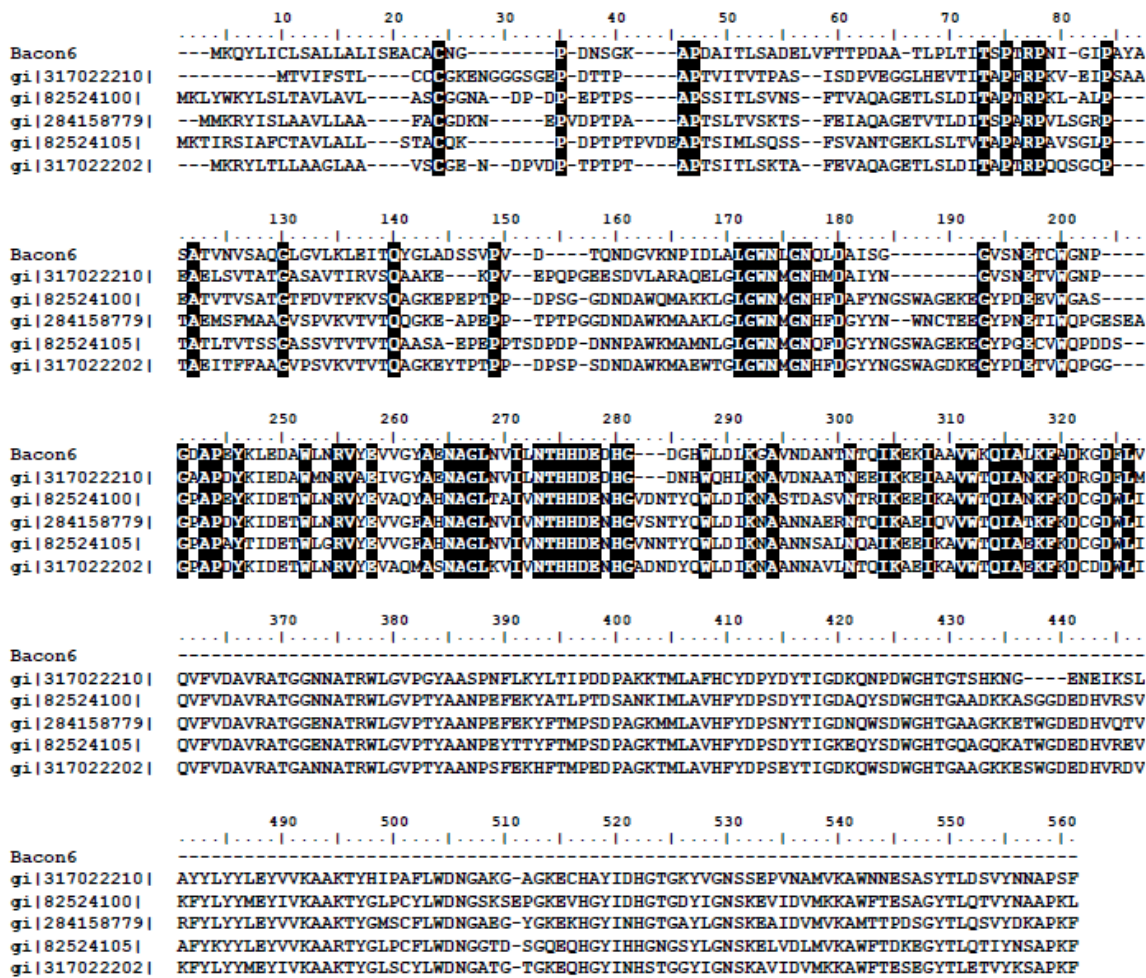
ID GenBank	ID Proteína	Descripción [Origen]	E-value	Identidad (%)	Referencia
gi 738996343	<a href="#">WP_036878541.1</a>	alfa-N-arabinofuranosidasa [ <i>Prevotella oryzae</i> ]	0	63	
gi 547926330	<a href="#">WP_022328427.1</a>	glicosil hidrolasa 43 [ <i>Prevotella sp. CAG:732</i> ]	0	61	
gi 493326443	<a href="#">WP_006283687.1</a>	xilosidasa/arabinosidasa [ <i>Prevotella bryantii</i> ]	0	59	
gi 653248722	<a href="#">WP_027454165.1</a>	alfa-N-arabinofuranosidasa [ <i>Prevotella bryantii</i> ]	0	59	
gi 496416227	<a href="#">WP_009125074.1</a>	glicosil hidrolasa 43 [ <i>Bacteroides fluxus</i> ]	0	58	

**Figura 2.** Comparación de la secuencia aminoacídica de Lac4 y sus homólogos más cercanos. Se muestran los números de acceso del GenBank. Las regiones de identidad entre las secuencias se muestran en columnas negras



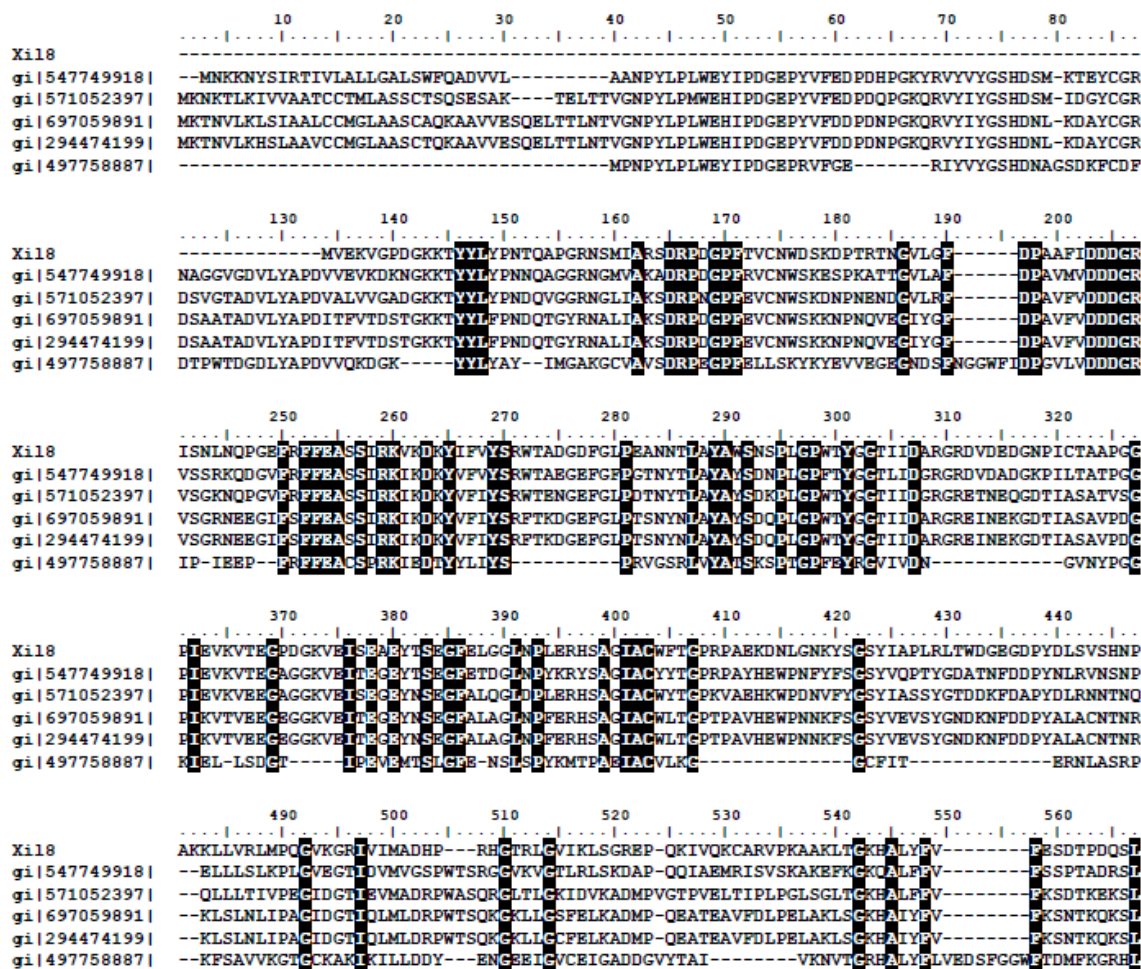
ID GenBank	ID Proteína	Descripción [Origen]	E-value	Identidad (%)	Referencia
gi 739028637	<a href="#">WP_036909807.1</a>	polifenol oxidasa [ <i>Prevotella sp.</i> FD3004]	2.00E-113	62	
gi 738996318	<a href="#">WP_036878516.1</a>	polifenol oxidasa [ <i>Prevotella oryzae</i> ]	1.00E-112	60	
gi 739014956	<a href="#">WP_036896329.1</a>	polifenol oxidasa [ <i>Prevotella sp.</i> S7 MS 2]	3.00E-112	59	
gi 545305291	<a href="#">WP_021584574.1</a>	Proteína YfiH [ <i>Prevotella pleuritidis</i> ]	6.00E-112	60	
gi 640560663	<a href="#">WP_024992025.1</a>	polifenol oxidasa [ <i>Prevotella pleuritidis</i> ]	8.00E-112	60	

**Figura 3.** Comparación de la secuencia aminoacídica de Bacon6 y sus homólogos más cercanos. Se muestran los números de acceso del GenBank. Las regiones de identidad entre las secuencias se muestran en columnas negras



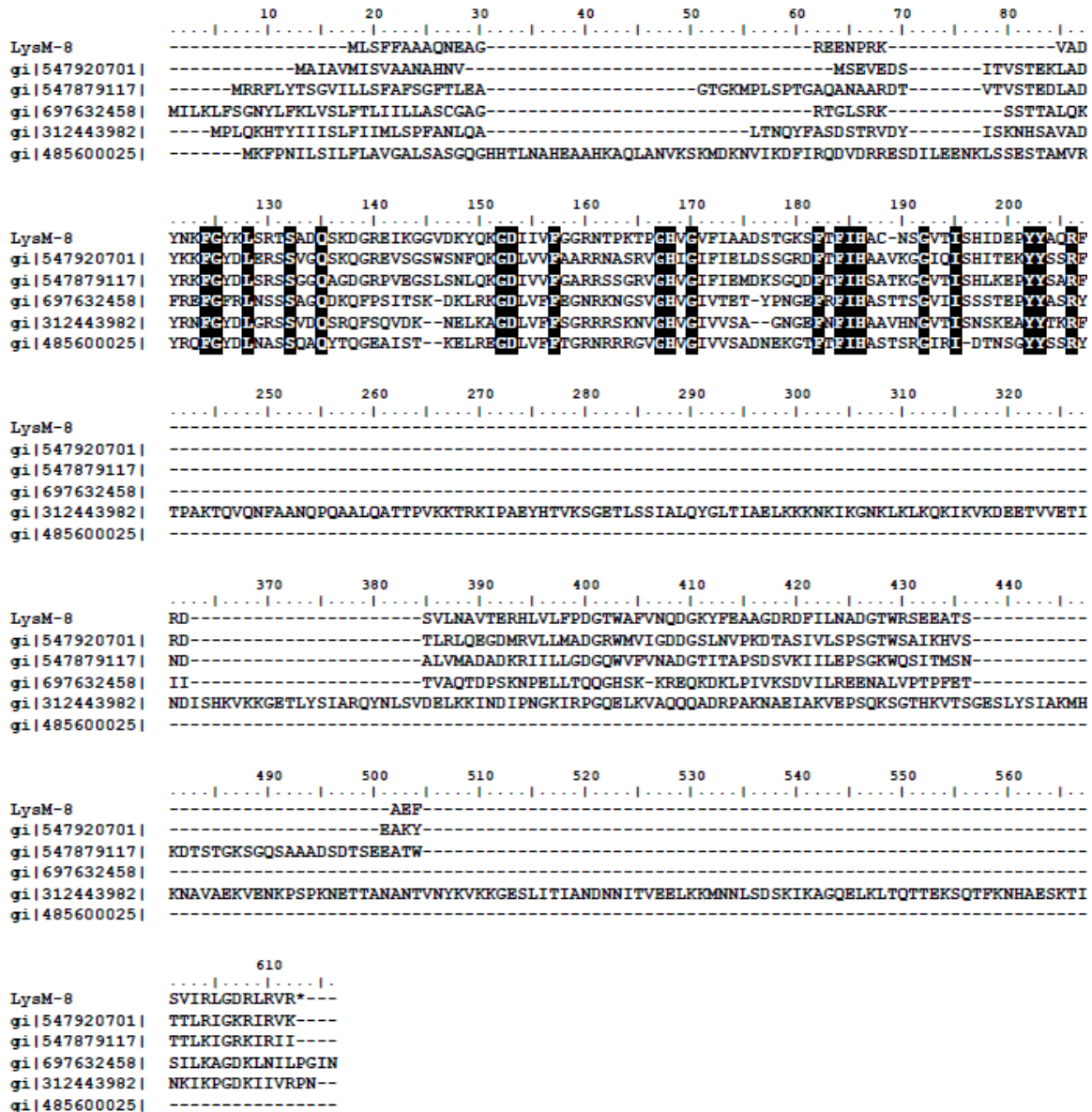
ID GenBank	ID Proteína	Descripción [Origen]	E-value	Identidad (%)	Referencia
gi 317022210	<a href="#">ADU86906.1</a>	<a href="#">celulasa [uncultured bacterium]</a>	6.00E-94	52	Dai, 2012
gi 82524100	<a href="#">CAJ19135.1</a>	<a href="#">celulasa/endoglucanasa [unidentified microorganism]</a>	5.00E-77	47	Ferrer, 2005
gi 284158779	<a href="#">ADB80100.1</a>	<a href="#">endoglucanasa [uncultured microorganism]</a>	2.00E-66	46	Duan, 2011
gi 82524105	<a href="#">CAJ19139.1</a>	<a href="#">celulasa [unidentified microorganism]</a>	4.00E-66	43	Ferrer, 2005
gi 317022202	<a href="#">ADU86902.1</a>	<a href="#">celulasa [uncultured bacterium]</a>	8.00E-65	47	Dai, 2012

**Figura 4.** Comparación de la secuencia de aminoácidos de Xil 8 y sus homólogos más cercanos. Se muestran los números de acceso del GenBank. Las regiones de identidad entre las secuencias se muestran en columnas negras



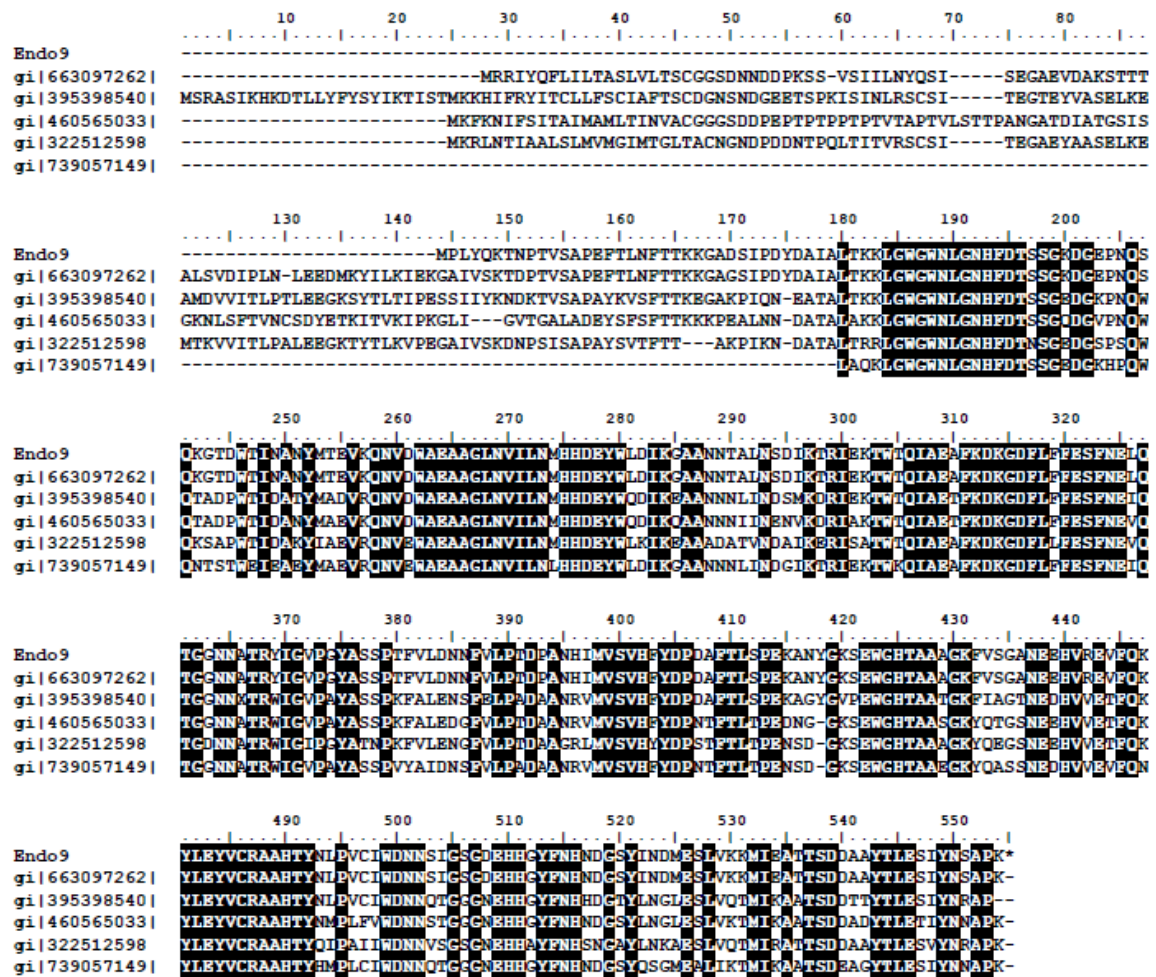
ID GenBank	ID Proteína	Descripción [Origen]	E-value	Identidad (%)	Referencia
gi 547749918	<a href="#">WP_022162901.1</a>	<a href="#">GH 43 [Bacteroides sp. CAG:633]</a>	0	63	
gi 571052397	<a href="#">AHF26057.1</a>	<a href="#">GH 43 [uncultured bacterium Contigcl_1748]</a>	1.00E-172	62	Wang, 2013
gi 697059891	<a href="#">WP_033149510.1</a>	<a href="#">GH 43 [Prevotella sp. RM4]</a>	3.00E-172	58	
gi 294474199	<a href="#">ADE83588.1</a>	<a href="#">GH 43 [Prevotella ruminicola 23]</a>	3.00E-172	58	Purushe, 2010
gi 497758887	<a href="#">WP_010073071.1</a>	<a href="#">Xilano 1,4-beta-xilosidasa [Clostridium cellulovorans]</a>	5.00E-48	33	

**Figura 5.** Comparación de la secuencia de aminoácidos de LysM 8 y sus homólogos más cercanos. Se muestran los números de acceso del GenBank. Las regiones de identidad entre las secuencias se muestran en columnas negras



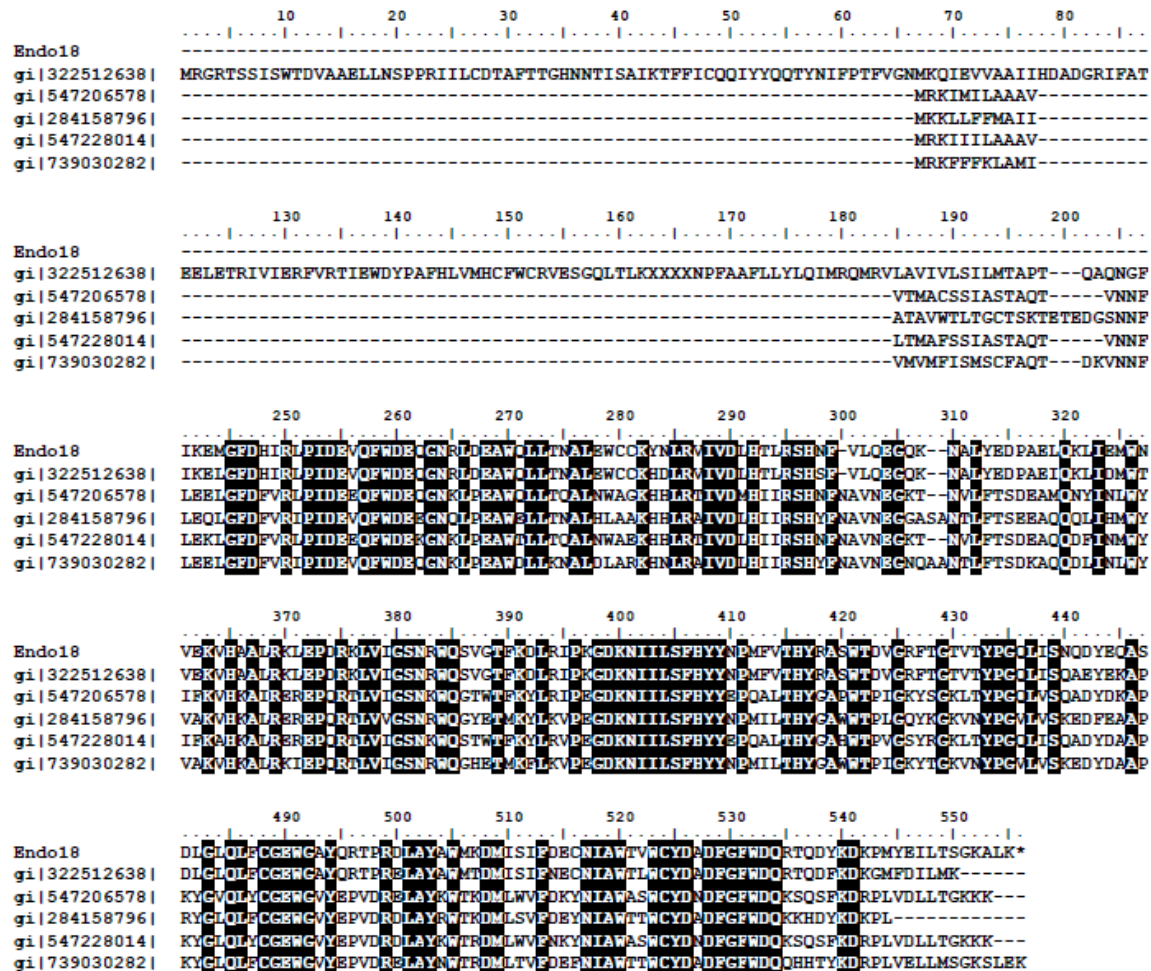
ID GenBank	ID Proteína	Descripción [Origen]	E-value	Identidad (%)	Referencia
gi 547920701	<a href="#">WP_022323007.1</a>	Proteína <a href="#">nLP/P60 [Alistipes sp. CAG:435]</a>	7.00E-65	37	
gi 547879117	<a href="#">WP_022284630.1</a>	Proteína <a href="#">nLP/P60 [Bacteroides sp. CAG:770]</a>	2.00E-61	35	
gi 697632458	<a href="#">CEA14920.1</a>	Proteína <a href="#">nLP/P60 [Porphyromonadaceae bacterium ]</a>	1.00E-32	33	
gi 312443982	<a href="#">ADQ80338.1</a>	Proteína <a href="#">nLP/P60 [Paludibacter propionigenes WB4]</a>	7.00E-27	33	Gronow, 2011
gi 485600025	<a href="#">AGK84828.1</a>	Proteína <a href="#">metaSSY_00740 [uncultured bacterium BAC25G1]</a>	3.00E-26	43	Yoon, 2013

**Figura 6.** Comparación de la secuencia de aminoácidos de Endo9 y sus homólogos más cercanos. Se muestran los números de acceso del GenBank. Las regiones de identidad entre las secuencias se muestran en columnas negras



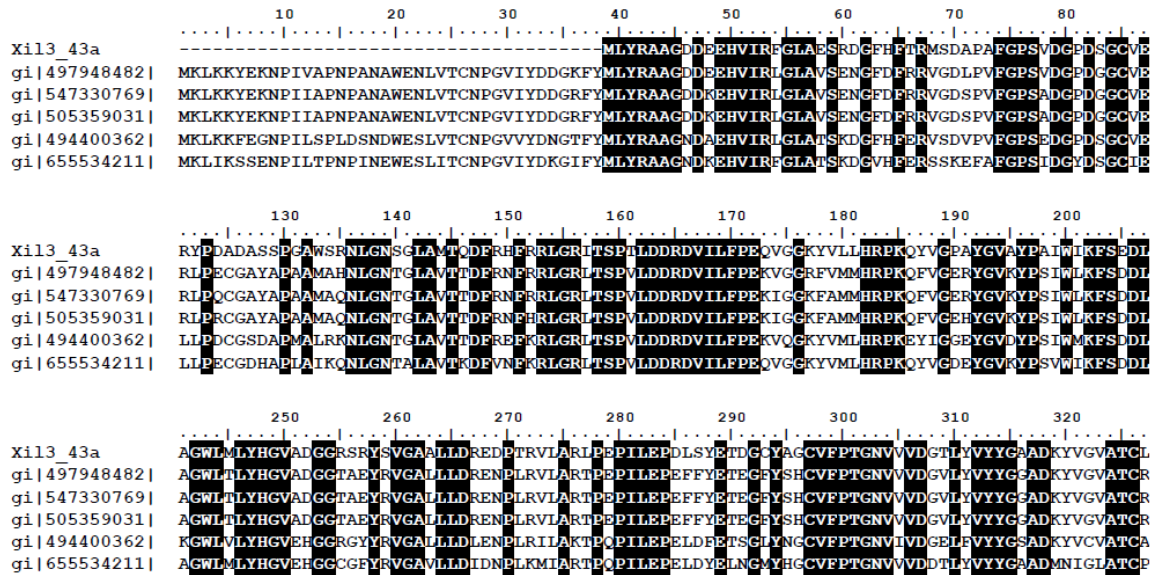
ID GenBank	ID Proteína	Descripción [Origen]	E-value	Identidad (%)	Referencia
gi 663097262	<a href="#">AIF25922.1</a>	<a href="#">celulasa putativa [uncultured bacterium Ad_087_C16]</a>	0	99	
gi 395398540	<a href="#">AFN57695.1</a>	<a href="#">GH5 [uncultured bacterium r_05]</a>	0	78	Ferrer, 2012
gi 460565033	<a href="#">AGH14017.1</a>	<a href="#">endoglucanasa [Prevotella sp. Sc00028]</a>	0	77	Rosewarne, 2014
gi 322512598	<a href="#">ADX05705.1</a>	<a href="#">CAZYma putativa [uncultured organism]</a>	0	73	Hess, 2011
gi 739057149	<a href="#">WP_036929107.1</a>	<a href="#">proteína hipotética [Prevotella sp. MA2016]</a>	0	78	

**Figura 7.** Comparación de la secuencia de aminoácidos de Endo18 y sus homólogos más cercanos. Se muestran los números de acceso del GenBank. Las regiones de identidad entre las secuencias se muestran en columnas negras



ID GenBank	ID Proteína	Descripción [Origen]	E-value	Identidad (%)	Referencia
gi 322512638	<a href="#">ADX05725.1</a>	<a href="#">CAZYma putativa [uncultured organism]</a>	0	91	Hess, 2011
gi 547206578	<a href="#">WP_021947505.1</a>	<a href="#">celodextrinasa [Prevotella sp. CAG:1124]</a>	1.00E-130	57	
gi 284158796	<a href="#">ADB80112.1</a>	<a href="#">celodextrinasa [uncultured microorganism]</a>	1.00E-129	60	Duan, 2011
gi 547228014	<a href="#">WP_021964992.1</a>	<a href="#">celodextrinasa [Prevotella sp. CAG:1185]</a>	8.00E-129	57	
gi 739030282	<a href="#">WP_036911420.1</a>	<a href="#">proteína hipotética [Prevotella sp. FD3004]</a>	2.00E-128	58	

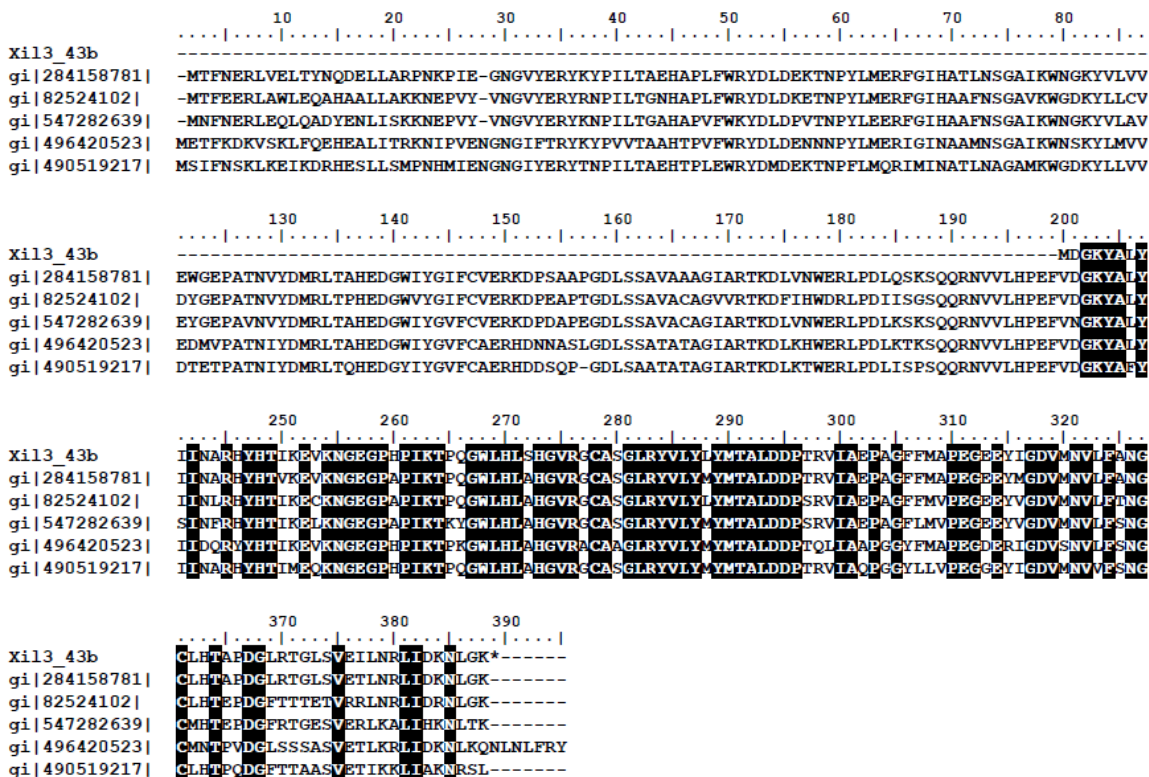
**Figura 8.** Comparación de la secuencia de aminoácidos de Xil3\_43a y sus homólogos más cercanos. Se muestran los números de acceso del GenBank. Las regiones de identidad entre las secuencias se muestran en columnas negras



ID GenBank	ID Proteína	Descripción [Origen]	E-value	Identidad (%)	Referencia
gi 497948482	<a href="#">WP_010262638.1</a>	<a href="#">glicosidasa [Alistipes timonensis]</a>	2.00E-159	74	
gi 547330769	<a href="#">WP_022061522.1</a>	<a href="#">glicosilasa putativa [Alistipes sp. CAG:53]</a>	2.00E-158	73	
gi 505359031	<a href="#">WP_015546133.1</a>	<a href="#">glicosilasa putativa [Alistipes shahii]</a>	3.00E-157	72	
gi 494400362	<a href="#">WP_007211644.1</a>	<a href="#">glicosidasa [Bacteroides cellulosilyticus]</a>	4.00E-152	69	
gi 655534211	<a href="#">WP_028908377.1</a>	<a href="#">glicosidasa [Prevotella ruminicola]</a>	2.00E-149	69	



**Figura 9.** Comparación de la secuencia de aminoácidos de Xil3\_43b y sus homólogos más cercanos. Se muestran los números de acceso del GenBank. Las regiones de identidad entre las secuencias se muestran en columnas negras



ID GenBank	ID Proteína	Descripción [Origen]	E-value	Identidad (%)	Referencia
gi 284158781	<a href="#">ADB80102.1</a>	<a href="#">glicosidasa [uncultured microorganism]</a>	1.00E-129	95	Duan, 2011
gi 82524102	<a href="#">CAJ19137.1</a>	proteína <a href="#">hipotética [unidentified microorganism]</a>	1.00E-112	85	Ferrer, 2005
gi 547282639	<a href="#">WP_022015988.1</a>	<a href="#">glicosidasa [Bacteroides sp. CAG:545]</a>	4.00E-108	81	
gi 496420523	<a href="#">WP_009129370.1</a>	proteína <a href="#">hipotética [Bacteroides oleiciplenus]</a>	9.00E-108	80	
gi 490519217	<a href="#">WP_004384722.1</a>	<a href="#">glicosidasa [Prevotella veroralis]</a>	2.00E-107	80	