UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

CENTRO DE INVESTIGACIÓN EN DINÁMICA CELULAR

"ANÁLISIS GENÓMICO Y METAGENÓMICO PARA LA BÚSQUEDA DE PROTEÍNAS CON UTILIDAD EN DECONSTRUCCIÓN DE LIGNOCELULOSA"

Lic. Ramón Alberto Batista García

Director de Tesis: Dr. Jorge Luis Folch Mallol

Cuernavaca, Morelos, Enero de 2016.

A mi abuelo…

Al lector

De La Habana a México, así fue la travesía que acompañó el inicio del camino de hacerme Doctor en Ciencias. Llegado a la UAEM y con la emoción del colegial que aprendería Biología Molecular me vinculé al Post-grado de la extinta Facultad de Ciencias (ahora Centro de Investigación en Dinámica Celular), y específicamente al Laboratorio de Biología Molecular de Hongos del Centro de Investigación en Biotecnología.

Quisiera agradecer a todos los involucrados en mi formación académica y a la Universidad Autónoma de Estado de Morelos que me acogió para mis estudios doctorales; a México y su pueblo que permitieron mi sustento económico durante estos cinco años y dejaron contagiarme con su cultura milenaria; a los amigos de siempre y a los nuevos juntados en este período. Agradecer la sapiencia de tanta gente que acompañó mi crecimiento académico, pero en especial a Jordi por impulsar mi inquietud por el conocimiento, considerarme como un colega más y nunca dejar de atender mis ideas o aportaciones a nuestro grupo de trabajo.

La etapa del doctorado, en lo personal, me hace sentir satisfecho académicamente. Participar en congresos internacionales, intercambios con científicos de la talla de Claudio Scazzocchio con quién hice filogenias de madrugada entre tequilas, contribuir en nuevos proyectos para el laboratorio, impartir docencia de pregrado y participar en varias estancias nacionales e internacionales, matizaron estos años por México y el mundo.

Hoy me aproximo a una nueva etapa profesional…, con el compromiso de generar y divulgar conocimiento, pero sobretodo con el sabor al mundo de los extremófilos lignocelulolíticos y con la convicción de hacer de ellos mi línea de trabajo.

Finalmente agradecer a mi familia, y en especial a mis padres por inculcarme lo valioso del saber y la virtud del hombre dichoso, y a la vez acurrucar mis logros y criticar mis errores.

Resumen

La creciente demanda de combustibles fósiles y los pronósticos de la industria petrolera internacional demandan la necesidad de nuevos sistemas combustibles como el bioetanol. Actualmente, los desechos lignocelulósicos se definen como las principales materias primas para la producción de bioetanol y otros productos valorizables en los esquemas de biorrefinación. Sin embargo, la deconstrucción de la biomasa vegetal constituye un reto para la factibilidad y sostenibilidad de las biorrefinerías. La organizada estructura polimérica y composición química de la lignocelulosa define la recalcitrancia de este material, y a la vez se necesitan de múltiples proteínas para su degradación, entre ellas: proteínas amorfogénicas, celulasas, xilanasas, esterasas y peroxidasas. Los enfoques genómicos y metagenómicos complementados con estudios de ecología clásica, permiten el análisis integral de la biodiversidad y recursos genéticos de un ecosistema determinado. Atendiendo a los criterios anteriores, el objetivo general de este trabajo fue: “Analizar genes y/o proteínas con utilidad en la deconstrucción de lignocelulosa para su conversión en materias primas de interés para las biorrefinerías”.

Se identificaron las poblaciones bacterianas degradadoras de bagazo de caña de azúcar mediante una librería de ARNr16S obtenida de un metagenoma de bagazo de caña de azúcar, y se evidenciaron distantes relaciones filogenéticas con las secuencias de referencias. A la vez se realizó la caracterización bioquímica y estructural de una expansina de *Schizophyllum commune* con utilidad en la liberación de glucosa y N-acetilglucosamina a partir de celulosa cristalina y quitina. Para complementar los estudios anteriores, se estudiaron las actividades ligninolíticas de cuatro hongos, los cuales mostraron capacidad para colonizar diferentes sustratos lignocelulósicos. *Aspergillus caesiellus* H1 fue aislado de bagazo de caña de azúcar en fermentación; mientras *Cadophora* sp. TS2, *Emericellopsis* sp. TS 11 y *Pseudogymnoascus* sp. TS12 fueron aislados de la esponja marina *Steletta normani*. Estos ascomicetos mostraron novedosos perfiles de enzimas lignocelulolíticas respecto a termotolerancia, haloestabilidad y temperatura y pH óptimos.

Tabla de contenidos

Introducción General

La biotecnología ha experimentado un desarrollo vertiginoso en los últimos años, así lo evidencia el sinnúmero de productos biotecnológicos disponibles en el mercado internacional. El crecimiento de la biotecnología y otras áreas afines, sin dudas, se acompaña de la sistematización en el conocimiento en disciplinas como Microbiología, Bioquímica, Ecología, Biología Molecular y Celular, Ingeniería Genética y de Proteínas, entre otras. El conocimiento generado, estructurado y sistematizado durante las diferentes etapas establecidas en el desarrollo de la Biotecnología, han permitido la obtención de productos como yogurts, leches ácidas y vinos, hasta productos como anticuerpos monoclonales, enzimas recombinantes y lípidos, los cuales requieren para su obtención de una solidez cognitiva cualitativamente superior (Figura 1).

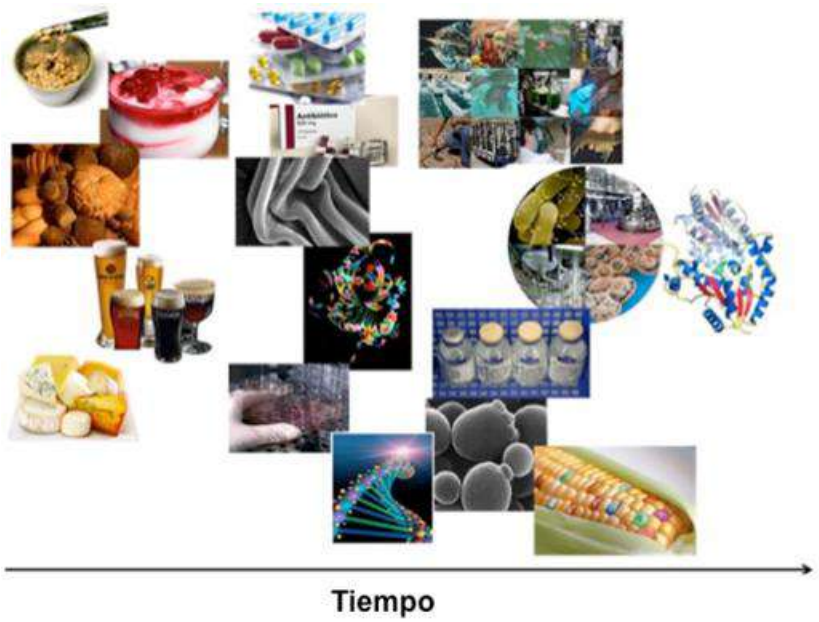


Figura 1. Desarrollo histórico de la Biotecnología considerando los productos que marcan sus principales etapas.

El curso histórico de la Biotecnología está influenciado por los postulados de Koch, la Microbiología Pasteuriana y el rápido crecimiento de las ciencias moleculares con aplicación a la Biología. Las prospecciones con interés biotecnológico han evolucionado desde el cultivo microbiano *in vitro* hasta el uso de técnicas

moleculares como genómica y metagenómica, entre otras. Aunque muchos grupos de investigación invierten esfuerzos y recursos en aproximaciones excluyentes al conocimiento, otros (como el nuestro) combinan ambos paradigmas para ponderar positivamente las tasas de éxito. Es por ello, que pudiéramos definir en una frase el ambiente metodológico de esta tesis: *"Con Pasteur y las ómicas"* (sin pretender ser reduccionistas en la sistematización del conocimiento científico).

Pero la doctrina esencial de la Biotecnología es la obtención de productos derivados de sistemas vivos. En esta consideración, es imprescindible pensar en los sistemas vivos y catalizadores biológicos que se involucran en convertir sustratos en productos. Desde tiempos remotos, donde el conocimiento inconsciente y empírico permitía la obtención de bioproductos, hasta la época más contemporánea después de la Revolución Francesa, los microorganismos han sido los *"vedette"* de la Biotecnología. La fisiología microbiana, y su versatilidad y plasticidad metabólicas, han realzado el carácter imprescindible del mundo microbiano para la evolución de la Biotecnología. Primero las bacterias, y luego los hongos filamentosos y levaduras, han sido los microorganismos más estudiados para la búsqueda de nuevos productos, mientras otros han ganado en interés como las microalgas. Sin embargo, los intereses metodológicos y conceptuales muestran inflexiones en las aproximaciones al conocimiento, y actualmente uno de los mayores intereses se centra en buscar nuevos y robustos catalizadores para mejorar la eficiencia de la conversión de sustratos en productos.

La prospección de proteínas en microorganismos, especialmente aquellas con actividad enzimática, constituye una actividad prioritaria para las aplicaciones biotecnológicas. El estudio de nuevas proteínas contribuye notablemente a una mejor comprensión de los nichos ecológicos microbianos, y resulta atractivo para la comprensión de su fisiología en diferentes ecosistemas y ambientes abióticos. Si para la Biotecnología, los biocatalizadores con novedosas propiedades bioquímicas y estructurales se definen como necesidades relevantes, entonces debemos diseñar estrategias adecuadas y redundantes para su búsqueda en los diferentes sistemas biológicos de interés. La Ecología Microbiana es una las disciplinas que permite el estudio sistemático de los microorganismos y/o sus

recursos genéticos, ambos contextualizados en un hábitat particular. De esta forma podemos acceder a la diversidad microbiana (entiéndase diversidad taxonómica y metabólica-funcional) a través de la caracterización de cultivos microbianos axénicos (enfoque Pasteuriano) y/o, del estudio directamente de sus materiales genéticos y proteínas (enfoque *ómico*). Las estrategias complementarias plantean el estudio de cultivos microbianos y del metagenoma asociado al ecosistemas de interés (Figura 2). En la ejecución de este proyecto doctoral fueron consideradas ambas estrategias experimentales: *(i)* cultivos microbianos (específicamente hongos filamentosos) aislados de ambientes extremos o con desbalance de nutrientes (bagazo de caña de azúcar y esponjas marinas en profundidades de 700m) y, *(ii)* estudios metagenómicos (bagazo de caña de azúcar).

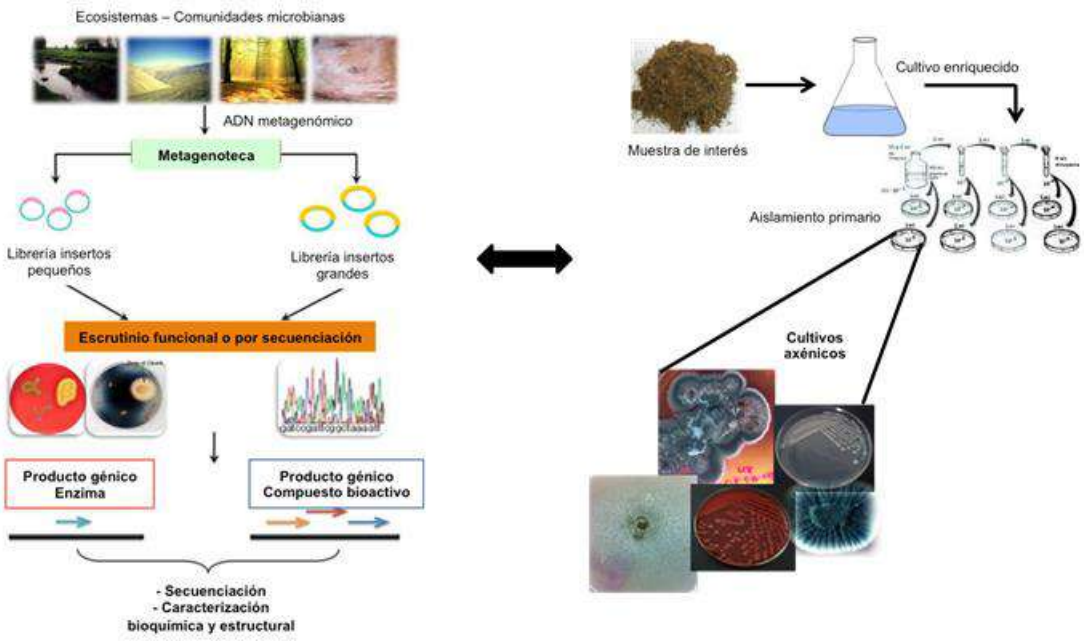


Figura 2. Enfoques ecológicos. Estudios metagenómicos *vs.* cultivos microbianos *in vitro*.

El mundo y sus problemas…, es tópico de editoriales responsables y amarillistas, libros, novelas, pintura y poesía…, pero también es tema de la ciencia porque jamás la actividad científica de un país o región puede desarrollarse

descontextualizada de la entropía y desarrollo sociales. En medio del estudio de fenómenos y procesos biológicos que contribuyan a generar conocimiento básico y aplicado para acercarnos a las soluciones que demanda la sociedad, se abren brechas a la motivación académica e intelectual, y a los nichos de investigación. De acuerdo a ello, este proyecto se ha desarrollado fundamentalmente entorno a la búsqueda de nuevas cepas fúngicas y proteínas con utilidad en la transformación de biomasa vegetal. El Laboratorio de Biología Molecular de Hongos perteneciente al Centro de Investigación en Biotecnología de la Universidad Autónoma del Estado de Morelos, se interesa en el estudio de la fisiología de hongos con potencialidades para la degradación de lignocelulosa.

La lignocelulosa es el polímero más sintetizado en la naturaleza, y se producen más de 200,000 millones de toneladas por año. La gestión integral de los residuos lignocelulósicos no se aborda eficientemente en la actualidad porque no existen tratamientos físicos, químicos y biológicos, que combinadamente permitan su aprovechamiento de manera eficiente y económicamente rentable, y su valorización como fuente de materia prima para diferentes procesos productivos. La producción de biocombustibles, específicamente de bioetanol, es quizás la aplicación más conocida de la lignocelulosa. Sin embargo, la biomasa vegetal puede relacionarse con muchísimas aplicaciones adicionales con utilidad en las industrias del papel, los cosméticos, farmacéutica, biomédica, alimenticia, textil, petroquímica y química, entre otras. Pero todas estas aplicaciones tienen un factor común, necesitan de la deconstrucción de la lignocelulosa. Este es el actual *“cuello de botella”*, y al que muchos (como nuestro laboratorio) apuestan creatividad, esfuerzo y dinero, pero sobre todo sapiencia académica e intelectual.

Las aplicaciones anteriores necesitan de azúcares fermentables, pero las fuentes más accesibles de estos azúcares son los alimentos, por ejemplo el maíz. La obtención de glucosa a partir de maíz o caña de azúcar sería un ejercicio productivo fácil y de bajo costo, considerando los contenidos de almidón y sacarosa en estos cultivos, respectivamente. ¿Pero podremos hacer competir la producción de alimentos con otras aplicaciones? Entonces llegan a esta discusión los interés transnacionales más inescrupulosos y la sensatez de la ética científica.

Producir alimentos para convertirlos en bioetanol por ejemplo, sería tentar una crisis de la cual la ciencia, olvidando su compromiso y carácter sociales, no pudiera escapar con muchos éxitos. De acuerdo a esta panorámica, la degradación de lignocelulosa, materiales lignocelulósicos, residuos agrícolas o biomasa vegetal como también se le define, se ha llevado las más serias discusiones en eventos científicos internacionales, y a la vez los titulares más amarillistas de la prensa.

El éxito de las estrategias actuales que se emplean para deconstruir lignocelulosa está confinado por la organizada estructura polimérica de este material, la cual le confiere la naturaleza de recalcitrante. La lignocelulosa (como se presentará más adelante, ver Capítulo I) se constituye fundamentalmente de lignina, hemicelulosa, celulosa y pectina (Figura 3). Estos polímeros se organizan tridimensionalmente en una red con regiones de alta cristalinidad, las cuales son de difícil acceso para las enzimas hidrolíticas y manifiestan bajos niveles de biodegradación. Estos polímeros pueden interactuar entre ellos para estabilizar químicamente la estructura, lo cual también justifica su baja biodegradabilidad. La estructura química y composición de la lignocelulosa demandan múltiples actividades enzimáticas para su degradación (celulasas, xilanasas, lacasas, esterasas, peroxidasas, oxigenasas, entre otras), y también de otras proteínas accesorias (expansinas, swolleninas y looseninas) para conferir cambios estructurales en sus regiones más cristalinas. Por tanto, la búsqueda de estas enzimas (especialmente celulasas y xilanasas) y otras proteínas, como las expansinas, que interactúen con estos polímeros para *"relajar"* su estructura química, resulta atractiva para las aplicaciones biotecnológicas relacionadas con la degradación de biomasa vegetal.

Las aplicaciones mencionadas anteriormente y el aprovechamiento integral de la lignocelulosa desde una perspectiva de compromiso social y sostenible ecológica y económicamente, se insertan en los esquemas de las biorrefinerías actuales. Las biorrefinerías o refinerías verdes son complejos industriales-biotecnológicos que combinan tratamientos físicos, químicos y biológicos para la obtención de productos y materias primas a partir de la biomasa vegetal. Uno de los aspectos más limitantes para el desarrollo y rentabilidad de las biorrefinerías es

precisamente los tratamientos biológicos (enzimáticos) con posibilidades de operación en condiciones industriales (altas temperaturas, pH extremos, alta salinidad). Una de las direcciones fundamentales de la investigación en este campo de estudio es la descripción de proteínas atractivas con utilidad en esquemas de biorrefinación.

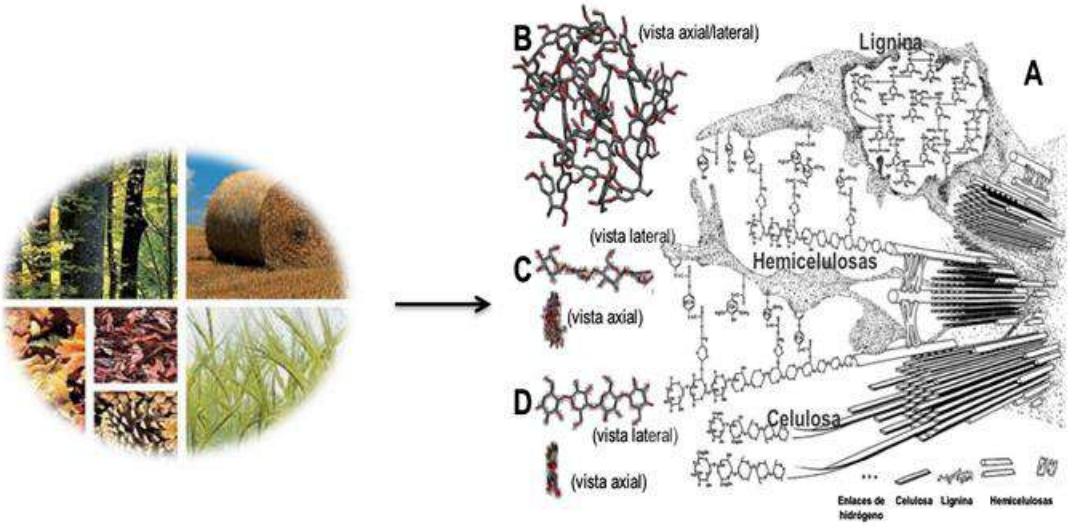


Figura 3. Estructura de la biomasa vegetal.

Atendiendo los argumentos anteriores, en este trabajo hemos estudiado las actividades celulasa y xilanasa de cuatro hongos aislados de ambientes extremos y/o con desbalance de nutrientes. *Aspergillus caesiellus* H1 fue aislado de una muestra de bagazo de caña de azúcar suplementada con 2M NaCl. El bagazo de caña de azúcar es un residuo lignocelulósico obtenido de la industria azucarera y resulta atractivo para el estudio de comunidades microbianas degradadoras de lignocelulosa. La caracterización de celulasas y xilanasas con potencialidades de hidrólisis en medios hipersalinos (2M NaCl) resulta de gran novedad y además se justifica si consideramos las necesidades de las industrias biotecnológicas. A la vez, se realizó el estudio del metagenoma asociado al bagazo para el análisis microbiano de diversidad procarionte en este ecosistema. Por otra parte, también se estudiaron las actividades ligninolíticas de tres hongos marinos asociados con la esponja *Stelleta normani*; *Cadophora* sp. TS2, *Emericellopsis* sp. TS11 y *Pseudogymnoascus* sp. TS12 que mostraron novedosos e interesantes perfiles

para actividades celulasas, xilanasas y fenoloxidasas en general. Interesados en otras actividades accesorias, también estudiamos una expansina de *Schizophyllum commune* y describimos sus peculiaridades estructurales y bioquímicas en un trabajo conjunto con el grupo de la Dra. Katiuska Arévalo en la Universidad de Nuevo León. Adicionalmente, con el interés de valorizar nuestra investigación con estudios de aplicación a nivel de laboratorio, estudiamos la potencialidad de nuestros hongos (cepas H1, TS2, TS11, TS12, entre otras cepas) para la remoción de compuestos fenólicos en aguas industriales, y para la degradación de hidrocarburos policíclicos aromáticos.

Por último, animados y con la responsabilidad de compartir nuestras experiencias y perspectivas sobre la aplicación de la metagenómica para el estudio de comunidades microbianas lignocelulolíticas, decidimos dedicar un apartado a una revisión analítica y crítica sobre este tema, considerando los antecedentes, perspectivas y utilidad de los enfoques metagenómicos utilizando sustratos ligninocelulósicos exclusivamente. Debemos señalar que a la fecha no se dispone de revisiones que contemplen estos aspectos en ecosistemas lignocelulósicos.

En adelante presentaremos un documento organizado por capítulos. En cada uno de ellos encontraremos diferentes enfoques ecológicos para el estudio de las actividades ligninolíticas de interés. Los diferentes capítulos le conducirán por historias que acompañadas de Pasteur y las *ómicas*, recrearán años de trabajo que llevan el sabor de los microorganismos lignocelulolíticos y extremófilos. *“Con Pasteur y las ómicas: de los lignocelulolíticos a los extremófilos”* define una estrategia conceptual general de varios capítulos relacionados entre ellos (ver mapa conceptual, Figura 4).

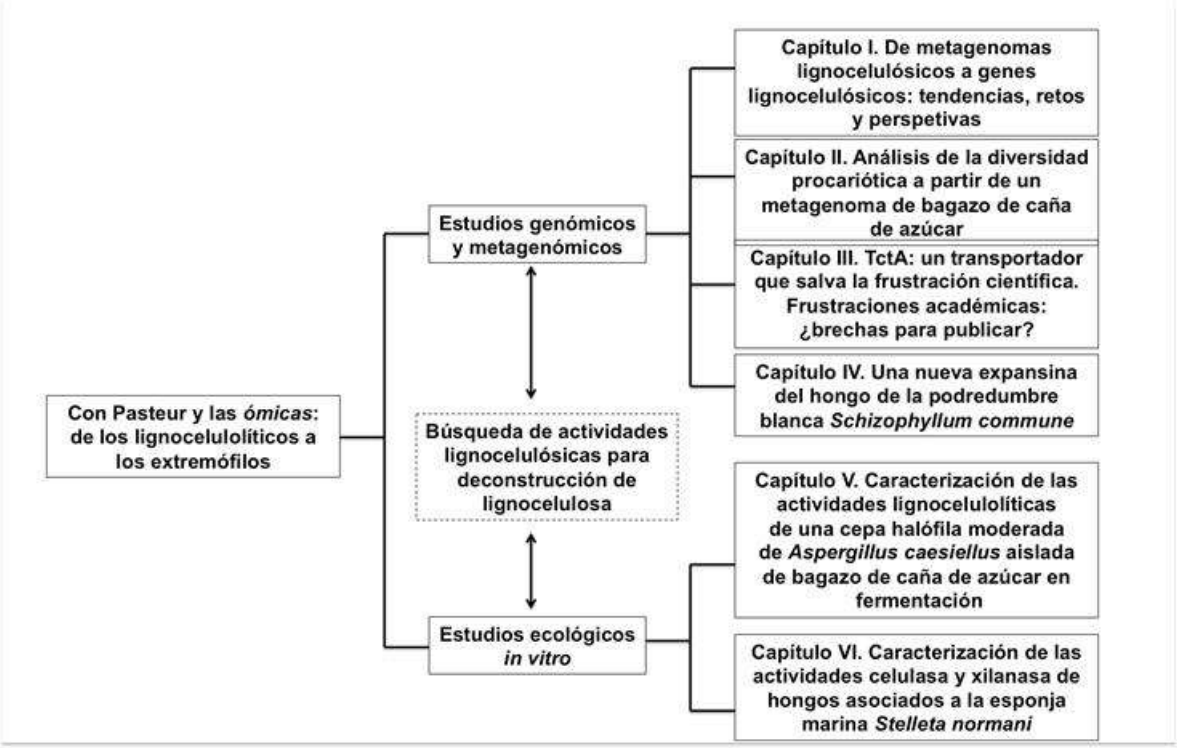


Figura 4. Mapa conceptual: “Con Pasteur y las *ómicas*: de los lignocelulósicos a los extremófilos”.

Justificación

La declinación, explotación irracional de yacimientos de combustible fósil, así como los pronósticos de la industria petrolera relacionados con reservas, producción y comercialización de petróleo y gas, demandan la búsqueda de alternativas rentables, ecológicas y sostenibles para la producción de combustibles alternativos. Los desechos lignocelulósicos se producen en grandes volúmenes (200,000 millones de tonelada por años) y no se dispone de tratamientos integrales (físicos, químicos y biológicos) eficientes para su explotación y aprovechamiento integral. La industria requiere que se desarrollen investigaciones que generen conocimientos con potencialidades de aplicación que tributen a la producción de biocombustibles (bioetanol) a partir de desechos lignocelulósicos y de materias primas valorizables para las biorrefinerías. Específicamente, las expansinas y actividades celulasas, xilanasas y fenoloxidasas en general, son requeridas para la deconstrucción de lignocelulosa y posteriores aplicaciones en refinerías verdes.

Hipótesis

Análisis genómicos y metagenómicos posibilitarían la prospección de nuevas proteínas (con o sin propiedades catalíticas) que contribuyan a la deconstrucción de la biomasa vegetal con potenciales aplicaciones en la industria de la biorrefinación.

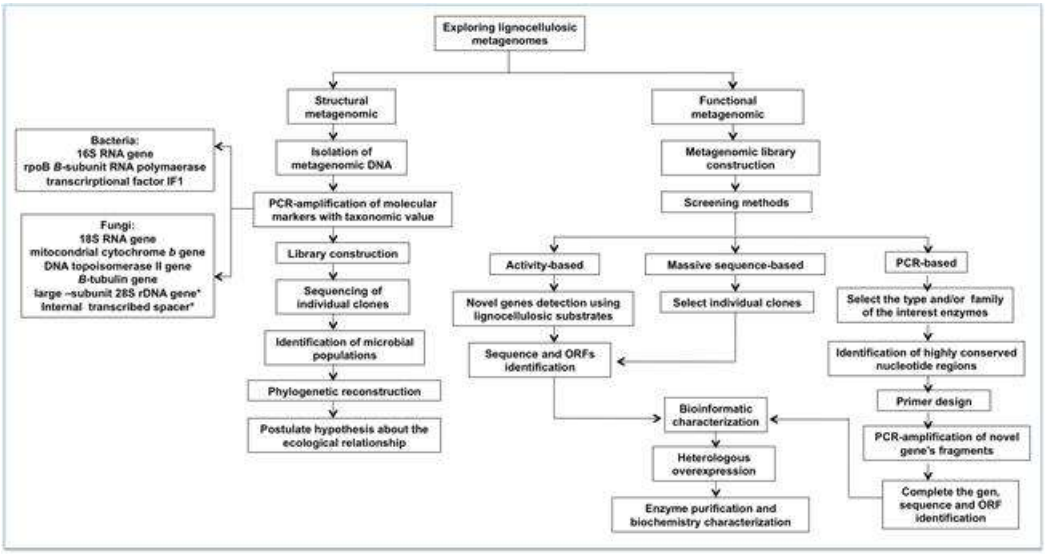
Objetivo General

Analizar genes y/o proteínas con utilidad en la deconstrucción de lignocelulosa para su conversión en materias primas de interés para las biorrefinerías.

Objetivos particulares

Analizar genomas y metagenotecas (funcional y estructuralmente) mediante herramientas bioinformáticas y ecológicas.

Caracterizar genes y/o proteínas derivados del análisis de genomas y metagenomas.



Capítulo I.

De metagenomas lignocelulósicos a genes lignocelulolíticos: tendencias, retos y perspectivas.

Los estudios metagenómicos se han popularizado en las caracterizaciones ecológicas (diversidad taxonómica y metabólica-fisiológica) a partir de la comprensión que sólo el 1% de los microorganismos pueden mantenerse en condiciones de laboratorio. La metagenómica estructural, y por otro lado la funcional, permiten la descripción de las comunidades microbianas asociadas a un ecosistema determinado y a la misma vez, el acceso a las proteínas de las diferentes poblaciones de microorganismos que lo habitan.

Aunque las metodologías para la construcción y caracterización de una librería metagenómica son generales, merece importancia discutir las especificidades para librerías obtenidas a partir de diferentes muestras ambientales (suelos, lodos activados, aguas dulces y de mar, entre otras), y para la búsqueda de diferentes enzimas. Atendiendo esta necesidad, y a la carencia de revisiones analíticas que aborden estos aspectos en ambientes exclusivamente lignocelulósicos, hemos escrito el siguiente trabajo: *"From lignocellulosic metagenomes to lignocellulosic genes: trends, challenges and prospects".*

En el trabajo anterior discutimos de manera crítica la importancia de los enfoques metagenómicos relacionados al inventario de biodiversidad de una muestra lignocelulósica, y a diferentes tipos de escrutinios para la descripción de celulasas, xilanasas, entre otras proteínas con utilidad en biorrefinerías. Así mismo, se analizan las peculiaridades de la extracción de ADN a partir de materiales como bagazos, y las dificultades metodológicas que intrínsicamente tienen los enfoques metagenómicos de estos residuos. En su sección final presentamos un análisis sobre la búsqueda de enzimas celulolíticas en otros ambientes lignocelulósicos relacionados con insectos. Esta revisión también recopila y analiza los pocos trabajos existentes relacionados con metagenómica estructural de bagazo de caña de azúcar. En esta reseña participan dos grupos de investigación con amplia trayectoria en metagenómica de ambientes lignocelulósicos (Dra. Paola Talia y Dr. Alan Dobson), los cuales son expertos en el tema y están adscritos al Instituto de Investigaciones Agropecuarias en Buenos Aires, Argentina, y a Environmental Research Institute de University College Cork en Cork, Irlanda, respectivamente.

Biotechnology for Biofuels

From lignocellulosic metagenomes to lignocellulosic genes: trends, challenges and prospects.

--Manuscript Draft--

Manuscript Number:	
Full Title:	From lignocellulosic metagenomes to lignocellulosic genes: trends, challenges and prospects.
Article Type:	Review
Abstract:	Lignocellulose is the most abundant biomass on earth and its possibilities to be used for obtaining vast amounts of compounds that are currently obtained from petrol or other fossil sources (natural gas, mineral carbon) have been scarcely exploited. The main reason being that lignocellulose is a complex mixture of polymers whose structural features hinder the access to the monosaccharides, phenolic compounds and acids that compose these polymers. Although microorganisms such as fungi and bacteria can decompose lignocellulose to its monomeric compounds and use them as carbon sources; and even that some of their enzymes and proteins involved in lignocellulose degradation are quite well studied, we are still lacking a comprehensive landscape of how the whole process occurs. It is also true that due to the limitations of culture-based methods to study lignocellulolytic organisms we may be missing some of the key elements that contribute to lignocellulose degradation. In this review we focus on metagenomic approaches to study lignocellulose degradation from structural and functional points of view, which may provide novel insights on this process and help to understand key elements in order to rationally design methods for the extraction of compounds in biomass that could make biorefineries more efficient.
Corresponding Author:	Jorge Luis Folch-Mallol, Ph.D. Universidad Autónoma del Estado de Morelos Cuernavaca, Morelos MEXICO
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universidad Autónoma del Estado de Morelos
Corresponding Author's Secondary Institution:	
First Author:	Ramón Alberto Batista-García
First Author Secondary Information:	
Order of Authors:	Ramón Alberto Batista-García
	María del Rayo Sánchez-Carbente, PhD
	Paola Talía, PhD
	Stephen A. Jackson, Ph.D
	Alan W. Dobson, PhD
	Jorge Luis Folch-Mallol, Ph.D.
Order of Authors Secondary Information:	

From lignocellulosic metagenomes to lignocellulosic genes: trends, challenges and prospects.

Ramón Alberto Batista-García[1-2], María del Rayo Sánchez-Carbente[1], Paola Talia[3], Stephen A. Jackson[4], Alan W. Dobson[4-5], Jorge Luis Folch-Mallol[1*].

[1]Centro de Investigación en Biotecnología, Universidad Autónoma del estado de Morelos, Cuernavaca, Morelos, México. [2]Facultad de Ciencias, Universidad Autónoma del estado de Morelos, Cuernavaca, Morelos, México. [3]Instituto de Biotecnología, CICVyA, CNIA, INTA Castelar, Argentina. [4]Marine Biotechnology Centre, Environmental Research Institute, University College Cork. Cork, Ireland. [5]School of Microbiology, University College Cork. Cork, Ireland.

* Corresponding author

[1]Centro de Investigación en Biotecnología

Universidad Autónoma del estado de Morelos

Avenida Universidad 1001

Col. Chamilpa

Cuernavaca 62209; Mexico

E-mail: jordi@uaem.mx

Abstract

Lignocellulose is the most abundant biomass on earth and its possibilities to be used for obtaining vast amounts of compounds that are currently obtained from petrol or other fossil sources (natural gas, mineral carbon) have been scarcely exploited. The main reason being that lignocellulose is a complex mixture of polymers whose structural features hinder the access to the monosaccharides, phenolic compounds and acids that compose these polymers. Although microorganisms such as fungi and bacteria can decompose lignocellulose to its monomeric compounds and use them as carbon sources; and even that some of their enzymes and proteins involved in lignocellulose degradation are quite well studied, we are still lacking a comprehensive landscape of how the whole process occurs. It is also true that due to the limitations of culture-based methods to study lignocellulolytic organisms we may be missing some of the key elements that contribute to lignocellulose degradation. In this review we focus on metagenomic approaches to study lignocellulose degradation from structural and functional points of view, which may provide novel insights on this process and help to understand key elements in order to rationally design methods for the extraction of compounds in biomass that could make biorefineries more efficient.

Keywords: metagenomic libraries, cellulases, xylanases, lignocellulosic materials, guts, lignocellulose degradation, biorefineries.

Introduction

The current energy crisis requires urgent solutions to satisfy the increasing demands for fossil fuels. Land and ocean oil reserves are rapidly diminishing and the oil industry´s forecasts predict a significant decrease in the levels of exploitable fossil fuels [1–5]. Because of this it is necessary to develop alternative energy generation systems to counteract these problems [6–13]. In this respect, biofuels represent a viable alternative source of renewable energy [14].

Bioethanol as an alternative energy source has for many years been of major interest to several laboratories around the world. It has many advantages when compared to traditional fuels, and is an excellent fuel for blending with gasolines of different octane ratings [15–17]. A major question however is, how to produce sufficient quantities of bioethanol on an industrial scale in an economically sustainable manner? A number of different applied and basic research strategies have been and continue to be undertaken in an attempt to achieve the requisite ethanol production levels to meet the ongoing demand for energy. One approach has been to focus on bioethanol production from sugars derived from food [18–22]; while others have preferred to focus on alcohol production from lignocellulosic wastes or from plant biomass to generate so called second generation bioethanol [13, 23–28].

Plant biomass is the most important, abundant, widespread material on earth. For many years, second generation bioethanol has been identified for its recognized potential as an ecological and friendly environmental source of mixed sugars for biofuel production [14]. Bioethanol production from lignocellulosic based material is currently quite challenging however, because of a dearth of cost-effective break-through technologies to facilitate the conversion of plant biomass into alcohol [29,

30]. Biomass saccharification is a very complex process, typically leading to quite low yields, and is often regarded as the critical conversion step, depending on the chemical composition of the biomass [30–34].

Lignocellulosic materials are predominantly composed of three polysaccharides: cellulose, hemicellulose and pectin, and a heterogeneous aromatic polymer: lignin [32, 35]. Cellulose is a homopolymer, the major reservoir of glucose in nature; while pectin, lignin and hemicellulose are complex heteropolymers of organic acidic, cyclic compounds and sugars. Different molecular interactions result in tight complexes being formed between these polymers. Hemicellulose, for example, has a β-(1→4)-linked backbone composed of xyloglucans, xylans, mannans and glucomannans, and β-(1→3, 1→4)-glucans [36]. Moreover several acids confer an even greater structural complexity given that they can esterify hemicellulose [37]. Thus the overall recalcitrant crystalline and amorphous structures of these polymers limit its conversion into fermentable sugars for the production of ethanol and others biofuels [30, 32].

Our understanding of how to efficiently transform vegetable biomass through technologically scalable, replicable and sustainable processes in currently incomplete. It will be necessary to acquire knowledge to efficiently deconstruct vegetable biomass and in this way obtain fermentable sugars for low-cost bioethanol production [38]. Physical, chemical and biological treatments will ultimately need to be used in integrated production systems for an optimal exploitation of our global lignocellulosic resources [33].

Biological methods are frequently used following chemical and physical treatments, with enzymes typically being used to release fermentable sugars and other

recoverable materials, which are useful for obtaining bioethanol and other biotechnological products [18]. Production schemes involving these steps are thus incorporated into the design of modern biorefineries, which are factories that fully take advantage of lignocellulosic waste to produce not only bioethanol, but also biogas, bioplastics, colorants, papers, dyes, resins, oils, amongst other final products are also produced [11, 32]. The search therefore for new and robust biocatalysts capable of hydrolyzing lignocellulosic substrates in industrial conditions is a real challenge for the operational need of biorefineries. In an attempt to obtain these types of robust biocatalysts research groups are currently focusing on exploring the genetic diversity of the microbial communities inhabiting biomass or lignocelullosic biomass rot.

Metagenomic based approaches

Microorganisms possess the ability to colonize a wide variety of natural and anthropogenic environments, including very specialized ecological niches and even extreme habitats; which is possible primarily due to the immense metabolic diversity and genetic adaptability of these microbes [39]. Prokaryotic organisms harbor the highest metabolic plasticity and are widely represented in all possible nutritional categories. Some studies estimate that approximately 4-6 x 10^{30} prokaryotes inhabit the earth [40], with around 2.6 x 10^{29} microorganisms being calculated to be present in soil and 1.2 x 10^{29} in the open oceans. In addition both oceanic and terrestrial subsurfaces have been estimated to contain 3.5 x 10^{30} and between 0.25 and 2.5 x 10^{30} microorganisms respectively [40].

Metagenomic based approaches have over the past few decades been developed in efforts to assess, analyze and exploit biodiversity in a wide variety of different environmental niches. Other *"omics"* based approaches such as proteomics, transcriptomics, together with metabolomics and microbiomics are also now being employed to analyze microbial metabolic and physiological biodiversity (Figure 1). Microbial isolates and *in vitro* studies of microbial metabolic diversity are really not representative of the biodiversity of an ecosystem and are always limited according to the culture media and environmental conditions used during incubation. As a result, *in vitro* recovered diversity is much lower than the total percentage of cultivable organisms. Conventional ecology (cultivation dependent methods) identifies phenotypes of interest in independent colonies through isolation of microorganisms and obtaining axenic cultures.

For this reason, ecological investigation of microbes through culture independent methods has become invaluable [41–44], with metagenomic approaches being employed to study the estimated < 1% of microorganisms, which are not cultivable under laboratory conditions [38, 45–52]. This percentage is uncertain and is an estimate, which varies between different ecosystems. While for terrestrial habitats it is recognized that more than 99% of bacteria cannot be cultured in the laboratory, studies in marine ecosystems suggest that as few as 0.001 – 0.1% of microbes are currently cultivable [41, 53]. These data highlight the relevance of metagenomic studies and show that difficulties in cultivating microorganisms do not enable the functional characterization of their proteins and the subsequent biotechnological applications for obtaining biofuels as third generation bioethanol.

There are two main areas in metagenomics: *(i)* structural and *(ii)* functional metagenomics (Figure 2). The first area studies the composition and structure of microbial communities and describes the major genera and species that inhabit an ecosystem [50]. It can also provide information about the role of microorganisms in biogeochemical cycles, propose ecological interactions of microbial communities and postulate hypotheses about evolutionary aspects in specific ecosystems [54]. Functional metagenomic explores genomes to study genes encoding new proteins and it is defined as a powerful tool for the recovery of novel biomolecules [50, 55]. At the same time it allows for direct cloning of large DNA fragments and simultaneous analysis of gigabases of DNA.

The best contribution to knowledge of the functional metagenomic approach has been to identify truly novel protein sequences representing even novel families. Some of these have no close relatives and even form deeply branched lineages. Indeed in the specific case of lignocellulosic enzymes, cellulases, xylanases,

esterases and lipases included, this remark is genuinely true [56]. The greatest biotechnology interest lies in the description of new biocatalysts with robust and resistant properties over a widespread range of environmental conditions.

Diverse metagenomes obtained from rivers, seas, oceans, lakes, soils, rumens, foods, fecal materials, sediments, insects guts and sludges have been described and analysed [57–69]. However, very few studies have been made on the metagenomics of bacteria from lignocellulosic rich ecosystems [70, 71]. It is possible that the structural complexity and their composition, limit the microbial populations that colonize them and limit the methods for the extraction of high quality DNA for further molecular applications as cloning. However, natural lignocellulosic materials represent one of the best options to study the lignocellulosic microbial communities especially unculturable populations.

It is clear that lignocellulosic environments provide enormous microbial diversity that remain largely unstudied. Sugarcane bagasse, wheat, corn or rice may be eligible as substrates for analysis of associated metagenomes. Therefore there is a high probability of finding truly novel cellulases, xylanases, ligninases and esterases/lipases from as yet uncharacterized microbial biomass, possibly even from new lineages.

For example, sugarcane bagasse which is a fibrous waste from sugar mills and is generally composed of 35 - 50% cellulose and 20 - 30% each of hemicellulose and lignin [72]. In addition it is one of the most recalcitrant wastes in agriculture with a calculated production of about 250 million metric tons per year [70]. The increasing interest in bagasse-based biorefineries using sugarcane bagasse as a raw material to obtain high-value associated products supports metagenomic derived-studies of

its non-cultivable microbial communities. A comprehensive knowledge of the non-cultivable microorganisms naturally inhabiting sugarcane bagasse may allow access to their metabolic resources such as proteins that define their ecological niches (lignocellulose degraders).

An alternative strategy to the Next-Generation Sequencing (NGS) approach is function-driven metagenomics, which is based on the use of screening procedures to discover enzymes and/or other functions of interest within environmental samples.

While a few methodological steps are involved in the construction of a metagenomic library from sample collection to the identification of positive clones for a specific lignocellulosic activity (Figure 3), it is often not an easy task to success depending on a number of key factors. We will now discuss some of them and share some of our experiences in the analysis of lignocellulosic metagenomes from sugarcane bagasse.

The extraction of high quality DNA is the first critical step in the construction of metagenomic libraries from lignocellulosic rich ecosystems. The fibrous nature of these materials however often constitutes a limitation in DNA extraction. Moreover, lignocellulosic materials such as sugarcane bagasse usually contain aliphatic acids such as acetic, formic, and levulinic acid together with furan derivatives such as furfural and hydroxymethylfurfural as well as phenolic compounds [72], that are common compounds derived from the production of molasses; that are concomitantly extracted with the metagenomic DNA. They are considered strong potential contaminants in DNA solutions given that they contribute to the denaturation of nucleic acids and moreover inhibit numerous enzymes and negatively interfere with DNA transformation [73, 74]. In addition the co-extracted

substances can cause a blackish colour in the crude DNA solution [75]. Furthermore, plant biomass may also contain derivatives of fertilizers, preservatives, stabilizers and other pollutants from industrial processes and these compounds can also affect the integrity and stability of isolated DNA. Other materials such as corn, rice and wheat straws, sawdust and agave fibers, usually have residues of organic acids, phenols, amines, amides, resins, oils, polycyclic and aromatic compounds and hydrocarbons, which again can affect the stability and structure of nucleic acids isolated from them.

The recalcitrant nature of these materials does not allow the development of a wide variety of microbial communities, this argument supports the very low yields of DNA obtained from the lignocellulosic samples. Another drawback is that there are not standardized methods to extract nucleic acids from these substrates so reproducibility of results is highly questionable (considering purity and yield). Additionally there are no commercially available kits to extract nucleic acids from these materials. It is often advisable to modify previous protocols described for the isolation of DNA [70, 71]. Considerations such as the sample´s granularity should always be considered to improve the yields of DNA.

In response to these problems it is advisable to enrich microbial populations of lignocellulosic substrates. Fermentation of these substrates with saline solutions and some additional carbon and energy sources to stimulate growth of microorganisms can be performed. Thus DNA yields can be improved and the pollutants present in the sample can be diluted. This strategy can also be used if ecological successions are required in the sample in order to specifically promote the growth of certain microbial groups. Sometimes it is favorable to enrich lignocellulosic microbial populations present in the starting sample to positively weigh the probability of

finding genes encoding cellulases, xylanases and lipases/esterases. Some authors report that metagenomic libraries containing DNA isolated from microbial communities, which are enriched with cellulose (as a major carbon and energy source) promote up four times more the probability to find genes encoding for glycosyl hydrolases compared with DNA libraries constructed with unenriched communities [76, 77]. Enrichments of anaerobic populations can also be made if we are interested in hydrolases of anaerobic organisms, which are very attractive for the biorefineries. It is also possible to make further screenings for sizes to eliminate microbial populations of specific sizes, resulting common procedure when we want to study only prokaryotic or eukaryotic enzymes. A similar procedure was used to study the bacterial community of the Sargasso Sea when the seawater samples were filtered to remove eukaryotic microorganisms [78]. Obviously when microbial populations are enriched, the structural studies lose meaning because they are intentionally altering population densities in natural biomass. However in functional studies these modifications can be attractive to researchers according to their perspectives.

After metagenomic DNA isolation and purification, the DNA should be fragmented and cloned into vectors. The fragment size depends on the metagenomic library design. Overall fragments of large sizes (> 20 kb) ensure greater success in functional screenings. The average size of genes is highly variable and large fragments allow for finding complete ORFs (open reading frames). If the intention is to find complete operons it is recommended to clone larger fragments (≥ 25 kb). When small fragments (< 3 kb) are cloned it is very probable to find incomplete genes, especially eukaryotic genes that have several introns and the standard sizes are larger [79, 80].

The choice of appropriate vectors for the construction of the library is also an important aspect to consider. It must be considered whether the vector allows the expression of the protein of interest. Over the course of metagenomic studies different vectors have been used for the construction of metagenomic libraries: plasmids, fosmids, cosmids, viruses and even bacterial artificial chromosomes are some of the vectors that have been used [61, 80–83].

In some cases the proper selection of vectors allows for increased success of the search. For example, using larger vectors made possible to clone DNA fragments of several tenths of kbs and increase the DNA titer of the library [83]. A higher DNA titer represents the metagenome better. Moreover, vectors for eukaryotic protein expression must also be considered if the main interest, for example, is to study lignocellulosic enzymes from filamentous fungi or insects. Often cellulases, xylanases and lipases/esterases that have been found, would be overexpressed in yeast or filamentous fungi for industrial use [84, 85]. This may be best used in screening for eukaryotic enzymes. This will improve in heterologous expression systems that functionally express recombinant enzymes.

In general the vectors should be selected considering the size of the fragments to be cloned, the presence of molecular markers for the transcription and expression of prokaryotic and eukaryotic genes, the presence of signals that ensure protein excretion if desired, among other factors. Cosmid and fosmid libraries show best results from the construction of libraries to the success of the screening methods. Our experience indicates that building libraries with viral vectors, like Lambda phage based vectors (phage libraries obtained) results in difficulties with screening methods to detect lignocellulosic enzymes. Moreover using plasmids with approximate sizes of 5 kb limits the DNA titer in the library [61]. The selection vector must consider the

interest and purpose of each library to be built. In general terms, libraries with inserts size between 2 – 10 kb can be constructed using plasmid or Lambda expression vectors [86]. Lambda phages, cosmids and fosmids as expression vector are used in libraries with inserts between 20 – 50 kb [87, 88], while artificial chromosome vectors are used in libraries with inserts up to 100 kb.

After obtaining and cloning the DNA into the appropriate vectors the cellular system for the construction of the library should be carefully considered. The easiest and most commonly chosen is *Escherichia coli* because there are many advantages that have ensured that it remains a valuable host for the efficient, cost-effective and high-level production of heterologous proteins [89]. However its limitations are many, greater when dealing with lignocellulolytic enzymes [14]. *E. coli* has long been the quintessential recombinant expression system but has greatly limited the search for lignocellulolytic enzymes from metagenomes. If we analyze the lignocellulolytic enzymes characterized to date in metagenomes the higher percentage belong to prokaryotic proteins [14]. The answer to this may be due to *E. coli* has being used in the construction of the library.

The expression of eukaryotic proteins in *E. coli* is often difficult [90, 91]. Thus, the probability of finding lignocellulolytic fungal enzymes is markedly reduced when bacterial systems are used as host of metagenomes. Many of these eukaryotic sequences have rare codons used by *E. coli*, on the other hand some of these need post-translational modifications which may not be made in *E. coli* [90, 91]. For example, the formation of disulfide bridges is difficult in the reducing conditions of *E. coli*'s cytoplasm affecting the folding of the protein and consequently its activity [92]. Glycosylations also are not performed in this system [90, 91] and cellulases and xylanases sometimes need these post-translational modifications [93, 94]. In other

examples, some esterases require other post-translational modification such as phosphorylation [95]. The absence of adequate export routes for fungal extracellular enzymes is another disadvantage for screening methods when using *E. coli*.

Moreover, the promoter sequences for transcription and essential sequences for translation of these genes are different in prokaryotes and eukaryotes [14]. Another important aspect is the intron editing mechanisms, which are absent in *E. coli* [96]. The majority of the fungal genes which encode cellulases, esterases and xylanases have between three and five introns (approximately 50-70 bp) and require post-transcripcional editing [97, 98]. All this discussion supports that activity-screens in practice are most suited to find prokaryotic enzymes when metagenomic libraries are constructed in *E. coli*. This does not mean that eukaryotic proteins cannot be found from these libraries, screening methods could also be used to allow this but the occurrence frequency is very low. Some alternative possibilities would be to use other bacterial systems to mitigate some of the above considerations. Some groups have used *Pseudomonas, Bacillus* and *Streptomyces* as hosts for the metagenomes [99–101]. But perhaps the best solution to overcome the limitations of bacterial systems is to build metagenomic libraries in fungal systems (yeast or filamentous fungal systems). We can also find limitations to proper expression of proteins from yeast. Some yeasts hyperglycosylate these proteins and do not recognize the signal peptides for protein secretion [102]. However, yeast and filamentous fungi make post-transcriptional and post-translational modifications, have an efficient secretion routes to export cellulases, xylanases and esterases and allow the folding of many of these proteins [103].

Structural and functional metagenomics for the discovery of lignocellulolytic microorganisms: general remarks

Distinctive molecular markers can be amplified to describe the composition of microbial communities from metagenomic DNA (Figure 2). For bacteria, fragments of the 16S rRNA gene are usually amplified and suffice for identification [104], but other molecular markers are helpful when 16S rDNA is not informative enough (*i.e.* rpoB, beta-subunit RNA polymerase, transcriptional factor IF1, ATP citrate lyase, citrate synthase, etc.) [105, 106]. For fungi many molecular markers with taxonomic value have been described to ensure proper identification. These include the mitochondrial cytochrome *b* gene, the DNA topoisomerase II gene (*TOP2*), the β-tubulin gene and different rRNA gene regions. Among the regions of the rRNA genes the most representative are the 5´end of the large-subunit 28S rDNA gene (D1-D2 region), the internal transcribed spacers 1 and 2 (ITS1 and ITS2) regions between the small- and large-subunit rRNA genes and some regions in the 18S RNA gene [107–111].

Libraries with fragments (not exceeding 1500 bp) of these genes (amplicons obtained from metagenomic DNA) are constructed and individual clones can be sequenced and analyzed [60] (Figure 2). Some studies prefer to analyze the relative amounts of microbial communities. Quantitative PCR is the proper technique for these approaches [112]. Description of the microbial populations of a lignocellulosic metagenome to analyze and to propose hypotheses from the ecological relationships can be established from the different groups identified. Phylogenies and evolutionary relationships that can be established between different groups, genera and microbial species identified in the metagenome can also direct methods to allow screening of certain groups of enzymes [70, 71].

But without doubt, the major interests for biotechnology is the detection of new attractive biocatalysts for industry and associated challenges. There is in particular much interest in studying the enzymes involved in the degradation of wood and

lignocellulosic wastes from lignocellulosic metagenomes [38]. Cellulases, xylanases, esterases, lipases and oxidases are in general defined as the top priorities [17]. Many methods of screening can be used to detect these enzymes (Figure 2). We should mention that great interest is to find lignocellulolytic fungal enzymes to clone and express them in yeast that will be used in the production of bioethanol.

During the screening, it should be clear to define which enzymes we want to find and from which microorganisms [17]. There are screening methods, such as PCR-based, that require this assumption in order for its success. The best work strategy integrates various screening methods [50].

One of the most widespread methods is activity-based [50]. These are based on the degradation of a substrate and usually employ a colour change or the presence of a halo around the positive clone screened [44]. In these methods there are no restrictions on the search other than the substrate, the success rate is high and it is correlated to the number of analyzed clones. Substrates must be carefully selected and should be broad spectrum. For example, tributyrin is an excellent substrate for esterase detection, while cellulose and xylan are good substrates for cellulases and xylanases detection [113–115]. In all these cases the halo presence around the clone is observed when the specific catalyst for such activity is present. Substrate mixtures can also be used for the same type of enzymes. The creativity of researchers may be an interesting aspect in this sense. Of course these methods have many disadvantages associated with the expression system and the type and origin of the enzyme prospected. It requires that transcription and translation processes are successful, that proteins are exported efficiently and correctly folded and finally that there is evidence of activity in the assay conditions. The detection of the activity also depends of several factors that often cannot be controlled. For

fungal lignocellulolytic enzymes the methods limit the search (when the library is constructed in *E. coli*) but still may be useful.

During activity-based screenings when the metagenomic libraries show high titers of clones (greater than 100 000) and there are not automated systems to facilitate the screenings, it is possible to inoculate the entire library in minimum medium with the substrates of interest and subsequently re-isolate the clones that have grown. These clones carry a gene from the metagenome. It is very easy to prove this statement retransforming *E. coli* with the isolated vector of the “positive clone” and a microbial population where all clones must show the same phenotype (phenotype found during the primary screening) must obtained. Finally the sequence of the fragment confirms the presence of the gene that confers the detected activity.

Other methods are PCR-based [50] (Figure 4). These methods allow the identification of highly conserved domains in a particular type of enzyme [116]. In these cases it is necessary to complete the gene in order to characterize the activity. It may not be easy to complete the gene sequence considering a complete metagenomic library but there are methods to do it (Genome Walker and 5' - 3' RACE kits) [117–119]. There are degrees of intrinsic restriction because primers employed will have a marked influence on the proteins that can be found and their homology with those already reported. Its disadvantage is the design of primers and subsequent amplification of the amplicons.

The design of primers is a critical step. Consensus and degenerate primers and primers combining degenerate and consensus regions in the same sequence (CODEHOPs primers: Consensus-degenerate hybrid oligonucleotide primers) may be used [116, 120, 121]. Degenerate primers increase the probability of finding

sequences that code for proteins with lower percentages of homology than those used to design the primers, while consensus primers will bias for the detection of sequences that code for proteins very closely related to those used for the design of primers (Figure 4).

Furthermore CODEHOPs primers are designed from amino acid sequences motifs highly conserved between members of a protein family and have proven to be highly effective in the identification and characterization of distantly related family members [121] (Figure 4). These primers are a perfect combination of consensus and degenerate regions in a pool of related primers. After multiply aligned proteins blocks of highly conserved amino acids are identified, they are used to design a set of primers containing all possible nucleotide sequences encoding 3-4 highly conserved amino acids within a 3 degenerate core. A longer 5´ non-degenerate clamp region contains the most probable nucleotide predicted for each flanking codon [120]. CODEHOPs primers have been used to detect new genes in plants, animal and bacterial species [122]. Moreover their application in the PCR-based screening of metagenomes could be of great use to identify new sequences of lignocellulolytic enzymes. In our experience the use of CODEHOPs primers has been successful in the search for xylanases (unpublished), however we have also found nonspecific products during screening for cellulases and also xylanases in metagenomes from activated sludge and sugarcane bagasse (unpublished).

PCR-based methods are peculiarly difficult in the search for enzymes such as cellulases, lipases/esterases and xylanases. These enzymes are grouped in more than 267 families between glycosidases, transferases, lyases and esterases (CAZY web site). Highly conserved amino acid regions are identified to design primers, which however sometimes show a low level of nucleotide conservation percentage

and this makes primer design difficult. The most useful and most used sequences are domains that are important for binding to polysaccharides because these enzymes share homology only in small regions of catalytic motifs. Moreover, it is impossible to consider all families or even all representatives of one family in the primers design.

This type of screening method is most appropriate when the prospecting of metagenomic libraries is based on a particular type of enzyme or particular enzyme family.

The screening based on colony hybridization is similar to the previous method. This can be used as probes with highly conserved carbohydrate-binding motifs used to locate colonies contain similar sequences by DNA-DNA hybridization. Usually many false positive clones for the activity of interest are detected. Another important type of screen is a sequencing-based method. Massive sequencing methods are used but there are major concerns with fragment assembly and its feasibility compared with other methods for libraries with more than 500,000 clones.

Stable-isotope probing is a powerful tool in microbial ecology and it is a useful method to search for lignocellulolytic enzymes in metagenomes. This method can describe the ecological niche of the microbial lignocellulosic communities and their interactions and even identify microorganisms involved in specific metabolic processes under conditions, which approach those occurring *in situ* [123]. Stable-isotope probing is also a technique with taxonomic value because it is a culture-independent procedure that allows the isolation of DNA from microorganisms involved in specific degradation process [124]. In stable-isotope probing techniques a substrate (cellulose if we search for cellulases) is enriched with a stable isotope

((13)CH(3)OH or (13)CH(4)) and later the DNA of the active microorganisms is collected by the selective recovery through density-gradient centrifugation [125]. Subsequently active microorganisms are identified by 16S RNA sequencing and subsequently their genes are studied through genomic approaches. This technique is more informative when is combined with microarrays and metagenomics data [125].

Scrutiny of cellulases, xylanases and esterases is a critical step to obtain novel lignocellulosic genes. A screening approach considering several of these strategies positively affected the success rates and the discovery of entirely new sequences and proteins with really very different structural characteristics to existing enzymes and consequently with novel catalytic properties. This review focuses on the utility of uncultured methods to detect and characterize cellulolytic system in environmental samples and insect gut.

Lygnocellulolytic enzimes from bacteria and fungi: structural and functional metagenomics approaches

Structural Metagenomics: Identification of lignocellulose-degrading microbia consortium grown on lignocellulose substrates

Sources of lignocellulose include agricultural wastes such as corn stover, bagasse, wood, grass, municipal waste and dedicated feedstocks crops such as miscanthus and switchgrass, could potentially provide energy via biofuels if systems for unlocking this energy were devised as robust, efficient and inexpensive [126, 127].

Bacterial and fungal communities are the more abundant and efficient organisms that participate in several stages of the lignocellullosic material decomposition from

terrestrial biomass. Those microorganisms have evolved remarkable physiological and functional diversity and therefore are the major reservoir of genetic source for potential biotechnological enzymes [128–134]. In particular, members of the *Gammaproteobacteria*, *Firmicutes* and *Bacteroidetes* have been proposed as candidates in lignocellulose biodegradation [71, 135, 136]. Also fungi like *Trichosporon* and *Coniochaeta* are considered as potential sources of hydrolytic enzymes [137].

Several metagenomes from different sources such as forest soils [138, 139], tropical peat swamp forest [140], switch grass-adapted compost community [141], biogas reactors [142, 143], yak rumen [144] and air-metagenome [145] have described the microbial communities diversity and their metabolic capabilities as natural biomass bioprocesors; where the prokaryotes are predominant, there are also fungi representatives such as *Aspergillus fumigatus* and *Sacharomyces cerevisae* [146]. Between these metagenomes it has been reported the presence of *Thermobifida fusca*, an important bacterial degrader of plant cell walls and commonly found in decaying organic matter. However and despite the importance of the use of lignocellulosic material as a source of biofuel and chemicals in biorefinery, few structural metagenomic studies coming from microbe communities growing specifically on lignocellullosic substrates have been reported (Table 1). The use of lignocellulose-enriched substrates is remarkable since we should expect to find the more efficient microbe population with a robust subset of lygnocellulolytic enzymes. Between these studies, Rattanachomsri et al. [70] and Wongwilaiwalin et al. [147], obtained metagenomes from sugarcane bagasse pile microbe community identifying mainly aerobic and facultative anaerobic bacteria according to the sequences generated with the 16S rDNA gene marker; and in the former, the presence of

cellulolytic and hemicellulolytic ascomycota -identified using the universal ITS (internal transcribed spacer). These findings reflect the aerobic and high temperature micro environmental conditions from the sugarcane bagasse decomposition and the selective pressure on the microbes for the utilization of plant biomass [70, 147]. Lignocellulolytic bacteria of the phyla *Actinobacteria* and *Furmicutes* particularly class *Bacilli* are present in this study and have been reported in previous soils and waste metagenomes with high lignocellulose content [148–151].

More recently, another metagenome study using poplar chips under anaerobic conditions show that the identified microbial community is different from those host-associate communities such as those of mammals or insects (Table 1) [133]. The more abundant microbial phyla are the *Firmicutes* and *Proteobacteria* representing the 45.9 and 32.3% of the metagenome binning, respectively; followed by *Bacteroidetes* with the 9.9%. Among the dominant members found in the phyla of *Protobacteria* is a bacterium similar to *Magnetospirillum*, which is hypotethized to play a role in the anaerobic breakdown of aromatic compounds [133]. The authors mention that the reads identifying fungi were low probably because their contribution to the consortium functioning is low. The comparison between these metagenomes obtained from highly enriched lignocellulose substrates vs. other metagenomes from other sources, revealed that the consortium isolated from such diverse environmental microbes communities shared similar composite phylum profiles comprising mainly Firmicutes, reflecting convergent adaptation of microcosm structure and differing at genus level [141]. The importance of metagenomic studies is not only the discovery of new lygnocellulolytic organism but also emphasize the importance microbe consortium to achieve highly efficient biomass degradation. In this way of thinking we can also take advantage to manipulate the tools that nature

has already forged to improve the microbe consortium to improve the biomass utilization.

Another issue to point out is that the majority of metagenomic analysis has used mainly 16S rRNA gene as a universal marker to isolate and characterize prokaryote populations and sometimes ITS marker to identify fungal populations; however this marker usually does no detect all the fungal phyla [152]. Moreover, in different metagenomes the characterization of fungi was either neglected because the studies do not consider the use of specific fungal markers or because the reads were too low to consider for further identification [133]. Also it has been hypotethized that fungi are not present in metagenomes analysis because conditions such as high temperature and relatively low oxygen limit their culture conditions [153].

Culture-independent approaches to identify and evaluate cellulolytic enzymes from bacteria and fungi.

Functional metagenomics studies allow the characterization of specific lignocellulolytic activities according to the interest of the researchers. The last decade has been enlightened with the characterization of new lignocellulolytic enzymes members that catalyse reactions in several steps of the lignocellulose-degrading pathways, however stronger efforts have to be made to complete the enzymatic lignocellulolytic degradation pathway from different microbes and in consequence might give light to understand their use to exploit the vast lignocellulosic plant biomass as potential source of biofuels. Previous reviews have focused on lignocellulolytic enzymes previously characterized from metagenomes, and few have focused their attention in those coming from microbe growth in lignocellulosic substrates [154, 155]. In this review, we will focus in the major

discoveries from the last five years from the biomass-degrading enzymes belonging to the GHases families. Cellulases and hemicellulases have been recognized as very useful biocatalysts because of their wide-ranging versatility in industrial applications, including food technology, textile production, biofuel formation, and paper production [42]. The composition of crystalline cellulose is quite homogenous in different types of plants, however hemicellulose and lignin are polymers with a diverse composition and/or linkages between monomers. This diversity generated high evolution pressure and as a consequence there is also an enzyme diversification to fit the necessity of the microbes that use lignocellulose substrates as carbohydrate sources [130]. Table 2 gives a summary of the lignocellulolytic enzymes recently characterized from wide ranges of metagenome studies. It is important to highlight that many other non-catalytic proteins such as expansins, swollenins and loosenin which induce weakening of the rigid cellulose structure, may contribute to lignocellulose degradation in ways that are not yet clearly understood [156–159].

Culture-independent approaches to identify and evaluate the cellulolytic enzymes from insects.

Some insects can secrete lignocellulolytic enzymes to allow them to use lignocellulosic substrates as energy sources [57, 160–162]. Therefore, insects represent a unique resource from which to search for novel and efficient cellulolytic enzymes. The termites are the most efficient decomposers of wood on earth [163], and the insects which have been most studied with respect to their cellulolytic systems. The cellulolytic activities in insects have been attributed to endogenous enzymes or/and enzymes from symbiotic microorganisms in their gut.

However, in some insects such as *Drosophila melanogaster*, *Anopheles gambiae* and *Bombyx mori* these types of enzymes are absent and are unable to degrade this component [164]. Some other insects, such as *Limnoria* (wood borers), however do not contain microorganisms in the digestive tracts, but produce endogenous enzymes necessary for lignocellulose degradation [165].

It is interesting to note that there is a controversy about the role of both endogenous and/or symbiotic enzymes for biomass bioconversion. Because most gut microorganisms are as yet unculturable, culture independent approaches and molecular analyses as with other ecosystems have been employed to assess their biodiversity and lignocellulolytic mechanisms (Table 3). Furthermore, because of the number of bacterial species estimated in the insect's gut is around several hundred [166], to date, relative few studies have been published [57, 167–170]. Thus, and probably because of these limitations, the most used culture independent methods in the identification of insect cellulases is the construction and screening of cDNA libraries, which have been useful to identify endogenous and symbiotic insect cellulases [171–175]. However, due to the advantages and availability of next generation sequencing techniques, which will allow the complete characterization and understanding of complex cellulolytic systems of insects, these approaches may be more frequently used in the near future.

In termites, most reports have been on their endogenous cellulases, and some of them have been heterologous expressed [163, 176]. Nimchua et al. [177] identified 14 clones with cellulase and xylanse activities from a metagenomic fosmid library of *Microcerotermes* sp., a wood-feeding higher termite. The enzymatic activities have been characterized in three of these clones. Furthermore, the *Reticulitermes*

flaviceps gut has been analysed by metatranscriptomic analyses of endogenous and symbiotic cDNA libraries. These authors identified 171 candidate genes encoding lignocellulases, and subsequently performed a functional analysis of phenoloxidase activity [168]. The metagenomic analysis of hindgut microbiota of the higher termite *Nasutitermes ephratae* also revealed a large number of genes involved in the degradation of cellulose and hemicelluloses [57]. In their study, the authors identified genes involved in other relevant symbiotic functions, such as in H_2 production, reductive acetogenesis and N_2 fixation.

Todaka et al. [172] analysed the gut symbionts of four representative lower termite species and a well-feed cockroach by metatranscriptomics in order to compare the lignocellulose-degrading system in these species. These researchers obtained around 4000 clones as ESTs from the cDNA libraries, and thus identified cellulases and hemicellulases in more than 10% of the clones from each library. More recently, a metagenomic sequencing analysis of the gut symbionts of grasshopper (*Acridacinerea*) and cutworm (*Agrotisipsilon*) was performed through Illumina sequencing [170]. These authors compared the results with a previously reported metagenomic analysis of termite gut microbiome and conclude that the grasshopper could be a good candidate for the discovery of biocatalysts due to the high cellulolytic activities in the gut.

The ligninocellulolytic activities of the insects have a high potential in biotechnological production of energy, such as methane and ethanol from renewable plant material. For this reason, an increase in the number of endogenous and symbiotic cellulolytic enzymes is expected in the near future.

While the search for lignocellulolytic enzymes from the metagenome of lignocellulolytic substrates such as sugarcane bagasse may yield success, the likelihood of novelty is diminished in an ecological niche, which favours such proteins. Investigations into the metagenomes from more extreme environments, such as deep-sea seawater or marine invertebrate metagenomes may yield truly novel enzymes as the physico-chemical properties of extreme marine environments (e.g. temperature, pressure, salinity) differ greatly from the terrestrial environment.

Such enzymes may be found in isolates from the extreme environment (e.g. cellulase from seawater isolate, *Marinimicrobium* sp. – Zhao et al. [178]; xylanase from marine sediment isolate *Streptomyces* sp. – Liu et al. [179]).

Alternatively metagenomic cloning can be used, as functional screens are readily available, to identify functions of interest. For example, a cold active esterase was identified from a clone library generated from deep-sea seawater [180], while a glycoside hydrolase with several functions (beta-glucosidase, beta-fucosidase and beta-galactosidase activities) was identified from seawater derived metagenomic library [181].

Acknowledgments: This work was funded in part by grant CB-153789-Q CONACyT from the Mexican Government.

The authors declare they have no competing interests

References

1. Stern R: **The Iranian petroleum crisis and United States national security**. *Proc Natl Acad Sci U S A* 2007, **104**:377–382.

2. Murphy DJ, Hall CAS: **Energy return on investment, peak oil, and the end of economic growth**. *Ann N Y Acad Sci* 2011, **1219**:52–72.

3. Höök M, Davidsson S, Johansson S, Tang X: **Decline and depletion rates of oil production: a comprehensive investigation**. *Philos Trans A Math Phys Eng Sci* 2014, **372**.

4. Kumhof M, Muir D: **Oil and the world economy: some possible futures**. *Philos Trans A Math Phys Eng Sci* 2014, **372**.

5. Miller RG, Sorrell SR: **The future of oil supply**. *Philos Trans A Math Phys Eng Sci* 2014, **372**.

6. Chisti Y: **Biodiesel from microalgae beats bioethanol**. *Trends Biotechnol* 2008, **26**:126–131.

7. Song D, Hou L, Shi D: **[Exploitation and utilization of rich lipids-microalgae, as new lipids feedstock for biodiesel production--a review]**. *Sheng Wu Gong Cheng Xue Bao* 2008, **24**:341–348.

8. Rhodes CJ: **Oil from algae; salvation from peak oil?**. *Sci Prog* 2009, **92**(Pt 1):39–90.

9. Abdelaziz AEM, Leite GB, Hallenbeck PC: **Addressing the challenges for sustainable production of algal biofuels: II. Harvesting and conversion to biofuels**. *Environ Technol* 2013, **34**:1807–1836.

10. Rawat I, Bhola V, Kumar RR, Bux F: **Improving the feasibility of producing biofuels from microalgae using wastewater**. *Environ Technol* 2013, **34**:1765–1775.

11. Hariskos I, Posten C: **Biorefinery of microalgae - opportunities and constraints for different production scenarios**. *Biotechnol J* 2014.

12. Profeta MI, Romero JM, Jorge NL, Grand A, Hernández-Laguna A: **Theoretical study of the gas-phase thermolysis of 3-methyl-1,2,4,5-tetroxane**. *J Mol Model* 2014, **20**.

13. Zeng Y, Zhao S, Yang S, Ding S-Y: **Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels**. *Curr Opin Biotechnol* 2014, **27C**:38–45.

14. Li L-L, McCorkle SR, Monchy S, Taghavi S, van der Lelie D: **Bioprospecting metagenomes: glycosyl hydrolases for converting biomass**. *Biotechnol Biofuels* 2009, **2**.

15. Zhang R, He H, Zhang C, Shi X: **[Preparation of ethanol-diesel fuel blends and exhausts emission characteristics in diesel engine]**. *Huan Jing Ke Xue* 2003, **24**:1–6.

16. Hansen AC, Zhang Q, Lyne PWL: **Ethanol-diesel fuel blends -- a review**. *Bioresour Technol* 2005, **96**:277–285.

17. Li L-L, Taghavi S, McCorkle SM, Zhang Y-B, Blewitt MG, Brunecky R, Adney WS, Himmel ME, Brumm P, Drinkwater C, Mead DA, Tringe SG, Lelie D van der: **Bioprospecting metagenomics of decaying wood: mining for new glycoside hydrolases**. *Biotechnol Biofuels* 2011, **4**.

18. Torney F, Moeller L, Scarpa A, Wang K: **Genetic engineering approaches to improve bioethanol production from maize**. *Curr Opin Biotechnol* 2007, **18**:193–199.

19. Xu J, Thomsen MH, Thomsen AB: **Feasibility of hydrothermal pretreatment on maize silage for bioethanol production**. *Appl Biochem Biotechnol* 2010, **162**:33–42.

20. Yangcheng H, Jiang H, Blanco M, Jane J: **Characterization of normal and waxy corn starch for bioethanol production**. *J Agric Food Chem* 2013, **61**:379–386.

21. Magaña C, Núñez-Sánchez N, Fernández-Cabanás VM, García P, Serrano A, Pérez-Marín D, Pemán JM, Alcalde E: **Direct prediction of bioethanol yield in sugar beet pulp using near infrared spectroscopy**. *Bioresour Technol* 2011, **102**:9542–9549.

22. Lima MA, Gomez LD, Steele-King CG, Simister R, Bernardinelli OD, Carvalho MA, Rezende CA, Labate CA, Deazevedo ER, McQueen-Mason SJ, Polikarpov I: **Evaluating the composition and processing potential of novel sources of Brazilian biomass for sustainable biorenewables production**. *Biotechnol Biofuels* 2014, **7**.

23. Lee J: **Biological conversion of lignocellulosic biomass to ethanol**. *J Biotechnol* 1997, **56**:1–24.

24. Oleskowicz-Popiel P, Lisiecki P, Holm-Nielsen JB, Thomsen AB, Thomsen MH: **Ethanol production from maize silage as lignocellulosic biomass in anaerobically digested and wet-oxidized manure**. *Bioresour Technol* 2008, **99**:5327–5334.

25. Zhao X-Q, Zi L-H, Bai F-W, Lin H-L, Hao X-M, Yue G-J, Ho NWY: **Bioethanol from lignocellulosic biomass**. *Adv Biochem Eng Biotechnol* 2012, **128**:25–51.

26. Yue Z-B, Li W-W, Yu H-Q: **Application of rumen microorganisms for anaerobic bioconversion of lignocellulosic biomass**. *Bioresour Technol* 2013, **128**:738–744.

27. Sarks C, Jin M, Sato TK, Balan V, Dale BE: **Studying the rapid bioconversion of lignocellulosic sugars into ethanol using high cell density fermentations with cell recycle**. *Biotechnol Biofuels* 2014, **7**.

28. Zabed H, Faruq G, Sahu JN, Azirun MS, Hashim R, Boyce AN: **Bioethanol production from fermentable sugar juice**. *ScientificWorldJournal* 2014, **2014**.

29. M.E. H: *Biomass Recalcitrance - Deconstructing the Plant Cell Wall for Bioenergy.* Oxford: Blackwell Publishing; 2008.

30. Geng W, Huang T, Jin Y, Song J, Chang H-M, Jameel H: **Comparison of sodium carbonate-oxygen and sodium hydroxide-oxygen pretreatments on the chemical composition and enzymatic saccharification of wheat straw**. *Bioresour Technol* 2014, **161**:63–68.

31. Canilha L, Kumar Chandel A, dos Santos Milessi TS, Fernandes Antunes FA, da Costa Freitas WL, das Graças Almeida Felipe M, da Silva SS: **Bioconversion of sugarcane biomass into ethanol: an overview about composition, pretreatment methods, detoxification of hydrolysates, enzymatic saccharification, and ethanol fermentation**. *J Biomed Biotechnol* 2012, **2012**.

32. Liu S, Lu H, Hu R, Shupe A, Lin L, Liang B: **A sustainable woody biomass biorefinery**. *Biotechnol Adv* 2012, **30**:785–810.

33. Hongdan Z, Shaohua X, Shubin W: **Enhancement of enzymatic saccharification of sugarcane bagasse by liquid hot water pretreatment**. *Bioresour Technol* 2013, **143**:391–396.

34. Chen C, Ding S, Wang D, Li Z, Ye Q: **Simultaneous saccharification and fermentation of cassava to succinic acid by Escherichia coli NZN111**. *Bioresour Technol* 2014, **163C**:100–105.

35. Rezende CA, de Lima MA, Maziero P, Deazevedo ER, Garcia W, Polikarpov I: **Chemical and morphological characterization of sugarcane bagasse submitted to a delignification process for enhanced enzymatic digestibility**. *Biotechnol Biofuels* 2011, **4**.

36. Scheller HV, Ulvskov P: **Hemicelluloses**. *Annu Rev Plant Biol* 2010, **61**:263–289.

37. Brink J van den, Vries RP de: **Fungal enzyme sets for plant polysaccharide degradation**. *Appl Microbiol Biotechnol* 2011, **91**:1477–1492.

38. Gilbert J, Li L-L, Taghavi S, McCorkle SM, Tringe S, van der Lelie D: **Bioprospecting metagenomics for new glycoside hydrolases**. *Methods Mol Biol* 2012, **908**:141–151.

39. Leis B, Angelov A, Liebl W: **Screening and expression of genes from metagenomes**. *Adv Appl Microbiol* 2013, **83**:1–68.

40. Whitman WB, Coleman DC, Wiebe WJ: **Prokaryotes: the unseen majority**. *Proc Natl Acad Sci U S A* 1998, **95**:6578–6583.

41. Kennedy J, Marchesi JR, Dobson ADW: **Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments**. *Microb Cell Fact* 2008, **7**.

42. Steele HL, Jaeger K-E, Daniel R, Streit WR: **Advances in recovery of novel biocatalysts from metagenomes**. *J Mol Microbiol Biotechnol* 2009, **16**:25–37.

43. Kennedy J, Flemer B, Jackson SA, Lejon DPH, Morrissey JP, O'Gara F, Dobson ADW: **Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism**. *Mar Drugs* 2010, **8**:608–628.

44. Kennedy J, O'Leary ND, Kiran GS, Morrissey JP, O'Gara F, Selvin J, Dobson ADW: **Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems**. *J Appl Microbiol* 2011, **111**:787–799.

45. Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, Wittwer P: **Metagenomic gene discovery: past, present and future**. *Trends Biotechnol* 2005, **23**:321–329.

46. Li X, Qin L, Dai S, Jiang S, Liu Z: **[Marine microbial metagenomics: progress and prospect]**. *Wei Sheng Wu Xue Bao* 2007, **47**:548–553.

47. Banik JJ, Brady SF: **Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules**. *Curr Opin Microbiol* 2010, **13**:603–609.

48. Mocali S, Benedetti A: **Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology**. *Res Microbiol* 2010, **161**:497–505.

49. Lombard N, Prestat E, van Elsas JD, Simonet P: **Soil-specific limitations for access and analysis of soil microbial communities by metagenomics**. *FEMS Microbiol Ecol* 2011, **78**:31–49.

50. Simon C, Daniel R: **Metagenomic analyses: past and future trends**. *Appl Environ Microbiol* 2011, **77**:1153–1161.

51. Suenaga H: **Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities**. *Environ Microbiol* 2012, **14**:13–22.

52. Williamson SJ, Yooseph S: **From bacterial to microbial ecosystems (metagenomics)**. *Methods Mol Biol* 2012, **804**:35–55.

53. Amann RI, Ludwig W, Schleifer KH: **Phylogenetic identification and in situ detection of individual microbial cells without cultivation**. *Microbiol Rev* 1995, **59**:143–169.

54. Simon C, Daniel R: **Achievements and new knowledge unraveled by metagenomic approaches**. *Appl Microbiol Biotechnol* 2009, **85**:265–276.

55. Chistoserdovai L: **Functional metagenomics: recent advances and future challenges**. *Biotechnol Genet Eng Rev* 2010, **26**:335–352.

56. Tuffin M, Anderson D, Heath C, Cowan DA: **Metagenomic gene discovery: how far have we moved into novel sequence space?**. *Biotechnol J* 2009, **4**:1671–1683.

57. Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernandez M, Murillo C, Acosta LG, et al.: **Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite.** *Nature* 2007, **450**:560–565.

58. Rasheed Z, Rangwala H, Barbará D: **16S rRNA metagenome clustering and diversity estimation using locality sensitive hashing**. *BMC Syst Biol* 2013, **7 Suppl 4**.

59. Yasir M, Khan H, Azam SS, Telke A, Kim SW, Chung YR: **Cloning and functional characterization of endo-β-1,4-glucanase gene from metagenomic library of vermicompost**. *J Microbiol* 2013, **51**:329–335.

60. Akbari VG, Pandya RD, Singh SP: **Extraction of the metagenomic DNA and assessment of the bacterial diversity from the petroleum-polluted sites**. *Environ Monit Assess* 2014.

61. Batista-García RA, Sánchez-Reyes A, Millán-Pacheco C, González-Zuñiga VM, Juárez S, Folch-Mallol JL, Pastor N: **A novel TctA citrate transporter from an activated sludge metagenome: Structural and mechanistic predictions for the TTT family**. *Proteins* 2014.

62. Curson ARJ, Burns OJ, Voget S, Daniel R, Todd JD, McInnis K, Wexler M, Johnston AWB: **Screening of metagenomic and genomic libraries reveals three classes of bacterial enzymes that overcome the toxicity of acrylate**. *PLoS One* 2014, **9**.

63. Jadeja NB, More RP, Purohit HJ, Kapley A: **Metagenomic analysis of oxygenases from activated sludge**. *Bioresour Technol* 2014.

64. Jin H, Li B, Peng X, Chen L: **Metagenomic analyses reveal phylogenetic diversity of carboxypeptidase gene sequences in activated sludge of a wastewater treatment plant in Shanghai, China**. *Ann Microbiol* 2014, **64**:689–697.

65. Kanwar N, Scott HM, Norby B, Loneragan GH, Vinasco J, Cottell JL, Chalmers G, Chengappa MM, Bai J, Boerlin P: **Impact of treatment strategies on cephalosporin and tetracycline resistance gene quantities in the bovine fecal metagenome**. *Sci Rep* 2014, **4**.

66. Tan H, Mooij MJ, Barret M, Hegarty PM, Harington C, Dobson ADW, O'Gara F: **Identification of novel phytase genes from an agricultural soil-derived metagenome**. *J Microbiol Biotechnol* 2014, **24**:113–118.

67. Tseng C-H, Tang S-L: **Marine microbial metagenomics: from individual to the environment**. *Int J Mol Sci* 2014, **15**:8878–8892.

68. Yarden O: **Fungal association with sessile marine invertebrates**. *Front Microbiol* 2014, **5**.

69. Xu Z, Hansen MA, Hansen LH, Jacquiod S, Sørensen SJ: **Bioinformatic approaches reveal metagenomic characterization of soil microbial community**. *PLoS One* 2014, **9**.

70. Rattanachomsri U, Kanokratana P, Eurwilaichitr L, Igarashi Y, Champreda V: **Culture-independent phylogenetic analysis of the microbial community in industrial sugarcane bagasse feedstock piles**. *Biosci Biotechnol Biochem* 2011, **75**:232–239.

71. Kanokratana P, Mhuantong W, Laothanachareon T, Tangphatsornruang S, Eurwilaichitr L, Pootanakit K, Champreda V: **Phylogenetic Analysis and Metabolic Potential of Microbial Communities in an Industrial Bagasse Collection Site**. *Microb Ecol* 2013, **66**:322–334.

72. Chandel AK, Antunes FAF, Anjos V, Bell MJ V, Rodrigues LN, Polikarpov I, Azevedo ER de, Bernardinelli OD, Rosa CA, Pagnocca FC, Silva SS da: **Multi-scale structural and chemical analysis of sugarcane bagasse in the process of sequential acid–base pretreatment and ethanol production by Scheffersomyces shehatae and Saccharomyces cerevisiae**. *Biotechnol Biofuels* 2014, **7**.

73. Robe P, Nalin R, Capellano C, Vogel TM, Simonet P: **Extraction of DNA from soil**. *Eur J Soil Biol* 2003, **39**:183–190.

74. Lim HK, Chung EJ, Kim J-C, Choi GJ, Jang KS, Chung YR, Cho KY, Lee S-W: **Characterization of a Forest Soil Metagenome Clone That Confers Indirubin and Indigo Production on Escherichia coli**. *Appl Environ Microbiol* 2005, **71**:7768–7777.

75. Tebbe CC, Vahjen W: **Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast.** *Appl Environ Microbiol* 1993, **59**:2657–2665.

76. Grant S, Sorokin DY, Grant WD, Jones BE, Heaphy S: **A phylogenetic analysis of Wadi el Natrun soda lake cellulase enrichment cultures and identification of cellulase genes from these cultures**. *Extremophiles* 2004, **8**:421–429.

77. Talia P, Sede SM, Campos E, Rorig M, Principi D, Tosto D, Hopp HE, Grasso D, Cataldi A: **Biodiversity characterization of cellulolytic bacteria present on native Chaco soil by comparison of ribosomal RNA genes**. *Res Microbiol* 2012, **163**:221–232.

78. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers Y-H, Smith HO: **Environmental genome shotgun sequencing of the Sargasso Sea**. *Science* 2004, **304**:66–74.

79. Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM: **Cloning the Soil Metagenome: a Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms**. *Appl Environ Microbiol* 2000, **66**:2541–2547.

80. Cheng J, Pinnell L, Engel K, Neufeld JD, Charles TC: **Versatile broad-host-range cosmids for construction of high quality metagenomic libraries**. *J Microbiol Methods* 2014, **99**:27–34.

81. Lussier F-X, Chambenoit O, Côté A, Hupé J-F, Denis F, Juteau P, Beaudet R, Shareck F: **Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector**. *J Ind Microbiol Biotechnol* 2011, **38**:1321–1328.

82. Selvin J, Kennedy J, Lejon DPH, Kiran GS, Dobson ADW: **Isolation identification and biochemical characterization of a novel halo-tolerant lipase from the metagenome of the marine sponge Haliclona simulans**. *Microb Cell Fact* 2012, **11**.

83. Yoon MY, Lee K-M, Yoon Y, Go J, Park Y, Cho Y-J, Tannock GW, Yoon SS: **Functional screening of a metagenomic library reveals operons responsible for enhanced intestinal colonization by gut commensal microbes**. *Appl Environ Microbiol* 2013, **79**:3829–3838.

84. Nasr S, Soudi MR, Hatef Salmanian A, Ghadam P: **Partial Optimization of Endo-1, 4-B-Xylanase Production by Aureobasidium pullulans Using Agro-Industrial Residues**. *Iran J Basic Med Sci* 2013, **16**:1245–1253.

85. Nathan VK, Esther Rani M, Rathinasamy G, Dhiraviam KN, Jayavel S: **Process optimization and production kinetics for cellulase production by Trichoderma viride VKF3**. *Springerplus* 2014, **3**.

86. Jiang C, Wu B: **Molecular cloning and functional characterization of a novel decarboxylase from uncultured microorganisms**. *Biochem Biophys Res Commun* 2007, **357**:421–426.

87. Pearce DA, Newsham KK, Thorne MAS, Calvo-Bado L, Krsek M, Laskaris P, Hodson A, Wellington EM: **Metagenomic analysis of a southern maritime antarctic soil**. *Front Microbiol* 2012, **3**.

88. Sae-Lee R, Boonmee A: **Newly derived GH43 gene from compost metagenome showing dual xylanase and cellulase activities**. *Folia Microbiol (Praha)* 2014.

89. Hannig G, Makrides SC: **Strategies for optimizing heterologous protein expression in Escherichia coli**. *Trends Biotechnol* 1998, **16**:54–60.

90. Jana S, Deb JK: **Strategies for efficient production of heterologous proteins in Escherichia coli**. *Appl Microbiol Biotechnol* 2005, **67**:289–298.

91. Delic M, Göngrich R, Mattanovich D, Gasser B: **Engineering of Protein Folding and Secretion—Strategies to Overcome Bottlenecks for Efficient Production of Recombinant Proteins**. *Antioxid Redox Signal* 2014.

92. Zerbs S, Giuliani S, Collart F: **Small-Scale Expression of Proteins in E. coli**. In *Methods in Enzymology*. *Volume 536*. Elsevier; 2014:117–131.

93. Payne CM, Resch MG, Chen L, Crowley MF, Himmel ME, Taylor 2nd LE, Sandgren M, Ståhlberg J, Stals I, Tan Z, Beckham GT: **Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose**. *Proc Natl Acad Sci U S A* 2013, **110**:14646–14651.

94. Veeresh J, Chuan WJ: **Microbial cellulases: Engineering, production and applications**. *Renew Sustain Energy* 2014, **33**:188–203.

95. Xie X, Langlais P, Zhang X, Heckmann BL, Saarinen AM, Mandarino LJ, Liu J: **Identification of a novel phosphorylation site in adipose triglyceride lipase as a regulator of lipid droplet localization**. *Am J Physiol Endocrinol Metab* 2014.

96. Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C: *Human Evolutionary Genetics, Second Edition*. Garland Science; 2013:690.

97. Flipphi M, Fekete E, Ag N, Scazzocchio C, Karaffa L: **Spliceosome twin introns in fungal nuclear transcripts**. *Fungal Genet Biol FG B* 2013, **57**:48–57.

98. Xia D, Wei Y, Zhang G, Zhao Q, Zhang Y, Xiang Z, Lu C: **cDNA cloning, expression, and enzymatic activity of a novel endogenous cellulase from the beetle Batocera horsfieldi**. *Gene* 2013, **514**:62–68.

99. Eom T, Muslimov IA, Tsokas P, Berardi V, Zhong J, Sacktor TC, Tiedge H: **Neuronal BC RNAs cooperate with eIF4B to mediate activity-dependent translational control.** *J Cell Biol* 2014, **207**:237–252.

100. Fang C, Stiegeler E, Cook GM, Mascher T, Gebhard S: **Bacillus subtilis as a platform for molecular characterisation of regulatory mechanisms of Enterococcus faecalis resistance against cell wall antibiotics**. *PLoS One* 2014, **9**:1–9.

101. Flinspach K, Rückert C, Kalinowski J, Heide L, Apel AK: **Draft Genome Sequence of Streptomyces niveus NCIMB 11891, Producer of the Aminocoumarin Antibiotic Novobiocin.** *Genome Announc* 2014, **2**:e01146–13.

102. Cao L, Yu L, Guo Z, Shen A, Guo Y, Liang X: **N-Glycosylation Site Analysis of Proteins from Saccharomyces cerevisiae by Using Hydrophilic Interaction Liquid Chromatography-Based Enrichment, Parallel Deglycosylation, and Mass Spectrometry**. *J Proteome Res* 2014, **13**:1485–1493.

103. Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, Fischer R, Commandeur U: **Challenges and advances in the heterologous expression of cellulolytic enzymes : a review**. 2014:1–15.

104. Vinje H, Almøy T, Liland KH, Snipen L: **A systematic search for discriminating sites in the 16S ribosomal RNA gene**. *Microb Inform Exp* 2014, **4**.

105. Turova TP, Kovaleva OL, Gorlenko VM, Ivanovskii RN: **[Use of genes of carbon metabolism enzymes as molecular markers of Chlorobi Phylum representatives].** *Mikrobiologiia* 2014, **83**:72–82.

106. Khamis A, Colson P, Raoult D, Scola B La: **Usefulness of rpoB gene sequencing for identification of Afipia and Bosea species, including a strategy for choosing discriminative partial sequences.** *Appl Environ Microbiol* 2003, **69**:6740–6749.

107. Hinrikson HP, Hurst SF, De Aguirre L, Morrison CJ: **Molecular methods for the identification of Aspergillus species.** *Med Mycol* 2005, **43 Suppl 1**:S129–37.

108. Hinrikson HP, Hurst SF, Lott TJ, Warnock DW, Morrison CJ: **Assessment of ribosomal large-subunit D1-D2, internal transcribed spacer 1, and internal transcribed spacer 2 regions as targets for molecular identification of medically important Aspergillus species.** *J Clin Microbiol* 2005, **43**:2092–2103.

109. Wang L, Yokoyama K, Takahasi H, Kase N, Hanya Y, Yashiro K, Miyaji M, Nishimura K: **Identification of species in Aspergillus section Flavi based on sequencing of the mitochondrial cytochrome b gene.** *Int J Food Microbiol* 2001, **71**:75–86.

110. Yokoyama K, Wang L, Miyaji M, Nishimura K: **Identification, classification and phylogeny of the Aspergillus section Nigri inferred from mitochondrial cytochrome b gene.** *FEMS Microbiol Lett* 2001, **200**:241–246.

111. Spiess B, Seifarth W, Hummel M, Frank O, Fabarius A, Zheng C, Morz H, Hehlmann R, Buchheidt D: **DNA microarray-based detection and identification of**

fungal pathogens in clinical samples from neutropenic patients. *J Clin Microbiol* 2007, **45**:3743–3753.

112. Kim J, Lim J, Lee C: **Quantitative real-time PCR approaches for microbial community studies in wastewater treatment systems: applications and considerations**. *Biotechnol Adv* 2013, **31**:1358–1373.

113. JunGang L, KeGui Z, WenJun H: **Cloning and biochemical characterization of a novel lipolytic gene from activated sludge metagenome, and its gene product**. *Microb Cell Fact* 2010, **9**.

114. Alvarez TM, Paiva JH, Ruiz DM, Cairo JPLF, Pereira IO, Paixão DAA, de Almeida RF, Tonoli CCC, Ruller R, Santos CR, Squina FM, Murakami MT: **Structure and function of a novel cellulase 5 from sugarcane soil metagenome**. *PLoS One* 2013, **8**.

115. Verma D, Kawarabayasi Y, Miyazaki K, Satyanarayana T: **Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (Mxyl) retrieved from compost-soil metagenome**. *PLoS One* 2013, **8**.

116. Itoh N, Isotani K, Makino Y, Kato M, Kitayama K, Ishimota T: **PCR-based amplification and heterologous expression of Pseudomonas alcohol dehydrogenase genes from the soil metagenome for biocatalysis**. *Enzyme Microb Technol* 2014, **55**:140–150.

117. Bouyer DH, Stenos J, Crocquet-Valdes P, Moron CG, Popov VL, Zavala-Velazquez JE, Foil LD, Stothard DR, Azad AF, Walker DH: **Rickettsia felis: molecular characterization of a new member of the spotted fever group.** *Int J Syst Evol Microbiol* 2001, **51**:339–347.

118. Inokuma H, Brouqui P, Drancourt M, Raoult D: **Citrate Synthase Gene Sequence: a New Tool for Phylogenetic Analysis and Identification ofEhrlichia**. *J Clin Microbiol* 2001, **39**:3031–3039.

119. Suzuki K, Ojima T, Nishita K: **Purification and cDNA cloning of a cellulase from abalone Haliotis discus hannai**. *Eur J Biochem* 2003, **270**:771–778.

120. Rose TM, Henikoff JG, Henikoff S: **CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design**. *Nucleic Acids Res* 2003, **31**:3763–3766.

121. Rose TM: **CODEHOP-mediated PCR - a powerful technique for the identification and characterization of viral genomes**. *Virol J* 2005, **2**.

122. Morant M, Hehn A, Werck-Reichhart D: **Conservation and diversity of gene families explored using the CODEHOP strategy in higher plants**. *BMC Plant Biol* 2002, **2**.

123. Radajewski S, Ineson P, Parekh NR, Murrell JC: **Stable-isotope probing as a tool in microbial ecology**. *Nature* 2000, **403**:646–649.

124. Radajewski S, Webster G, Reay DS, Morris SA, Ineson P, Nedwell DB, Prosser JI, Murrell JC: **Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing**. *Microbiology* 2002, **148**(Pt 8):2331–2342.

125. Dumont MG, Murrell JC: **Stable isotope probing - linking microbial identity to function**. *Nat Rev Microbiol* 2005, **3**:499–504.

126. Blanch HW, Adams PD, Andrews-Cramer KM, Frommer WB, Simmons BA, Keasling JD: **Addressing the need for alternative transportation fuels: the Joint BioEnergy Institute.** *ACS Chem Biol* 2008, **3**:17–20.

127. Gomez LD, Steele-King CG, McQueen-Mason SJ: **Sustainable liquid biofuels from biomass: the writing's on the walls.** *New Phytol* 2008, **178**:473–485.

128. Kambourova M, Mandeva R, Fiume I, Maurelli L, Rossi M, Morana A: **Hydrolysis of xylan at high temperature by co-action of the xylanase from Anoxybacillus flavithermus BC and the beta-xylosidase/alpha-arabinosidase from Sulfolobus solfataricus Oalpha.** *J Appl Microbiol* 2007, **102**:1586–1593.

129. Yu H, Zeng G, Huang H, Xi X, Wang R, Huang D, Huang G, Li J: **Microbial community succession and lignocellulose degradation during agricultural waste composting.** *Biodegradation* 2007, **18**:793–802.

130. Dashtban M, Schraft H, Qin W: **Fungal bioconversion of lignocellulosic residues; Opportunities & perspectives**. *Int J Biol Sci* 2009, **5**:578–595.

131. Jeya M, Joo A-R, Lee K-M, Tiwari MK, Lee K-M, Kim S-H, Lee J-K: **Characterization of beta-glucosidase from a strain of Penicillium purpurogenum KJS506.** *Appl Microbiol Biotechnol* 2010, **86**:1473–1484.

132. Huang D-L, Zeng G-M, Feng C-L, Hu S, Lai C, Zhao M-H, Su F-F, Tang L, Liu H-L: **Changes of microbial population structure related to lignin degradation during lignocellulosic waste composting.** *Bioresour Technol* 2010, **101**:4062–4067.

133. Van der Lelie D, Taghavi S, McCorkle SM, Li LL, Malfatti S a., Monteleone D, Donohoe BS, Ding SY, Adney WS, Himmel ME, Tringe SG: **The metagenome of an anaerobic microbial community decomposing poplar wood chips**. *PLoS One* 2012, **7**.

134. Bugg T, Ahmad M, Hardiman EM, Rahmanpour R: **Pathways for degradation of lignin in bacteria and fungi**. 2011, **28**.

135. Berlemont R, Martiny AC: **Phylogenetic distribution of potential cellulases in bacteria.** *Appl Environ Microbiol* 2013, **79**:1545–1554.

136. Jiménez DJ, Dini-Andreote F, van Elsas JD: **Metataxonomic profiling and prediction of functional behaviour of wheat straw degrading microbial consortia.** *Biotechnol Biofuels* 2014, **7**:92.

137. Johnson EA: **Biotechnology of non-Saccharomyces yeasts-the basidiomycetes.** *Appl Microbiol Biotechnol* 2013, **97**:7563–7577.

138. Woo HL, Hazen TC, Simmons B a., DeAngelis KM: **Enzyme activities of aerobic lignocellulolytic bacteria isolated from wet tropical forest soils**. *Syst Appl Microbiol* 2014, **37**:60–67.

139. Jiménez DJ, Korenblum E, van Elsas JD: **Novel multispecies microbial consortia involved in lignocellulose and 5-hydroxymethylfurfural bioconversion.** *Appl Microbiol Biotechnol* 2014, **98**:2789–803.

140. Bunterngsook B, Kanokratana P, Thongaram T, Tanapongpipat S, Uengwetwanit T, Rachdawong S, Vichitsoonthonkul T, Eurwilaichitr L: **Identification and characterization of lipolytic enzymes from a peat-swamp forest soil metagenome.** *Biosci Biotechnol Biochem* 2010, **74**:1848–1854.

141. Allgaier M, Reddy A, Park JI, Ivanova N, D'Haeseleer P, Lowry S, Sapra R, Hazen TC, Simmons B a., Vandergheynst JS, Hugenholtz P: **Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community**. *PLoS One* 2010, **5**.

142. Schluter C, Lam KKY, Brumm J, Wu BW, Saunders M, Stevens TH, Bryan J, Conibear E: **Global analysis of yeast endosomal transport identifies the vps55/68 sorting complex.** *Mol Biol Cell* 2008, **19**:1282–1294.

143. Rademacher A, Zakrzewski M, Schluter A, Schonberg M, Szczepanowski R, Goesmann A, Puhler A, Klocke M: **Characterization of microbial biofilms in a thermophilic biogas system by high-throughput metagenome sequencing.** *FEMS Microbiol Ecol* 2012, **79**:785–799.

144. 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: **Metagenomic insights into the fibrolytic microbiome in yak rumen.** *PLoS One* 2012, **7**:e40430.

145. Cao C, Jiang W, Wang B, Fang J, Lang J, Tian G, Jiang J, Zhu TF: **Inhalable microorganisms in Beijing's PM2.5 and PM10 pollutants during a severe smog event**. *Environ Sci Technol* 2014, **48**:1499–1507.

146. Cao C, Jiang W, Wang B, Fang J, Lang J, Tian G, Jiang J, Zhu TF: **Inhalable Microorganisms in Beijing ' s PM 2.5 and PM 10 Pollutants during a Severe Smog Event**. *Enviromental Sci Technol* 2014, **48**:1499–1507.

147. Wongwilaiwalin S, Laothanachareon T, Mhuantong W, Tangphatsornruang S, Eurwilaichitr L, Igarashi Y, Champreda V: **Comparative metagenomic analysis of microcosm structures and lignocellulolytic enzyme systems of symbiotic biomass-degrading consortia**. *Appl Microbiol Biotechnol* 2013, **97**:8941–8954.

148. Kumar R, Singh S, Singh O V.: **Bioconversion of lignocellulosic biomass: Biochemical and molecular perspectives**. *J Ind Microbiol Biotechnol* 2008, **35**:377–391.

149. Stott MB, Crowe MA, Mountain BW, Smirnova A V, Hou S, Alam M, Dunfield PF: **Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand.** *Environ Microbiol* 2008, **10**:2030–2041.

150. Navarro-Noya YE, Jan-Roblero J, Gonzalez-Chavez M del C, Hernandez-Gama R, Hernandez-Rodriguez C: **Bacterial communities associated with the rhizosphere of pioneer plants (Bahia xylopoda and Viguiera linearis) growing on heavy metals-contaminated soils.** *Antonie Van Leeuwenhoek* 2010, **97**:335–349.

151. Hanreich A, Schimpf U, Zakrzewski M, Schlüter A, Benndorf D, Heyer R, Rapp E, Pühler A, Reichl U, Klocke M: **Metagenome and metaproteome analyses of microbial communities in mesophilic biogas-producing anaerobic batch fermentations indicate concerted plant carbohydrate degradation**. *Syst Appl Microbiol* 2013, **36**:330–338.

152. Prewitt ML, Diehl S V, Mcelroy TC, Diehl WJ: **Comparison of general fungal and basidiomycete-specific ITS primers for identification of wood decay fungi**. *For Prod J* 2008, **58**:66–71.

153. Wongwilaiwalin S, Rattanachomsri U, Laothanachareon T, Eurwilaichitr L, Igarashi Y, Champreda V: **Analysis of a thermophilic lignocellulose degrading microbial consortium and multi-species lignocellulolytic enzyme system**. *Enzyme Microb Technol* 2010, **47**:283–290.

154. Duan CJ, Feng JX: **Mining metagenomes for novel cellulase genes**. *Biotechnol Lett* 2010, **32**:1765–1775.

155. Van Dyk JS, Pletschke BI: **A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes--factors affecting enzymes, conversion and synergy.** *Biotechnol Adv* 2012, **30**:1458–1480.

156. Quiroz-Castañeda RE, Martínez-Anaya C, Cuervo-Soto LI, Segovia L, Folch-Mallol JL: **Loosenin, a novel protein with cellulose-disrupting activity from Bjerkandera adusta.** *Microb Cell Fact* 2011, **10**:8.

157. Yao S-M, Lin Q, Zhang G-L, Yang H, Deng X-F, Nie G, Zheng X-B, Liu Y-X: **[Soluble expression of A/H1N1 influenza virus HA with Drosophila S2 cell line and its bio-activity identification].** *Zhonghua shi yan he lin chuang bing du xue za zhi = Zhonghua shiyan he linchuang bingduxue zazhi = Chinese J Exp Clin Virol* 2013, **27**:360–362.

158. Bunterngsook B, Mhuantong W, Champreda V, Thamchaiphenet A, Eurwilaichitr L: **Identification of novel bacterial expansins and their synergistic actions on cellulose degradation**. *Bioresour Technol* 2014, **159**:64–71.

159. Gourlay K, Hu J, Arantes V, Andberg M, Saloheimo M, Penttilä M, Saddler J: **Swollenin aids in the amorphogenesis step during the enzymatic hydrolysis of pretreated biomass**. *Bioresour Technol* 2013, **142**:498–503.

160. Ohkuma M: **Symbioses of flagellates and prokaryotes in the gut of lower termites**. *Trends Microbiol* 2008, **16**(May):345–352.

161. Tokuda G, Watanabe H: **Hidden cellulases in termites: revision of an old hypothesis.** *Biol Lett* 2007, **3**(March):336–339.

162. Brune A: **Symbiotic digestion of lignocellulose in termite guts.** *Nat Rev Microbiol* 2014, **12**:168–80.

163. Watanabe H, Tokuda G: **Cellulolytic systems in insects.** *Annu Rev Entomol* 2010, **55**:609–632.

164. Kunieda T, Fujiyuki T, Kucharski R, Foret S, Ament S a., Toth a. L, Ohashi K, Takeuchi H, Kamikouchi a., Kage E, Morioka M, Beye M, Kubo T, Robinson GE, Maleszka R: **Carbohydrate metabolism genes and pathways in insects: Insights from the honey bee genome**. *Insect Mol Biol* 2006, **15**:563–576.

165. King AJ, Cragg SM, Li Y, Dymond J, Guille MJ, Bowles DJ, Bruce NC, Graham I a, McQueen-Mason SJ: **Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes.** *Proc Natl Acad Sci U S A* 2010, **107**:5345–5350.

166. Hongoh Y: **Diversity and genomes of uncultured microbial symbionts in the termite gut.** *Biosci Biotechnol Biochem* 2010, **74**:1145–1151.

167. Brennan Y, Callen WN, Christoffersen L, Dupree P, Goubet F, Healey S, Hernández M, Keller M, Li K, Palackal N, Sittenfeld A, Tamayo G, Wells S, Hazlewood GP, Mathur EJ, Short JM, Robertson DE, Steer B a.: **Unusual microbial xylanases from insect guts**. *Appl Environ Microbiol* 2004, **70**:3609–3617.

168. Tartar A, Wheeler MM, Zhou X, Coy MR, Boucias DG, Scharf ME: **Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite Reticulitermes flavipes.** *Biotechnol Biofuels* 2009, **2**(i):25.

169. Bastien G, Arnal G, Bozonnet S, Laguerre S, Ferreira F, Fauré R, Henrissat B, Lefèvre F, Robe P, Bouchez O, Noirot C, Dumon C, O'Donohue M: **Mining for hemicellulases in the fungus-growing termite Pseudacanthotermes militaris using functional metagenomics.** *Biotechnol Biofuels* 2013, **6**:78.

170. Shi W, Xie S, Chen X, Sun S, Zhou X, Liu L, Gao P, Kyrpides NC, No EG, Yuan JS: **Comparative Genomic Analysis of the Endosymbionts of Herbivorous Insects Reveals Eco-Environmental Adaptations: Biotechnology Applications**. *PLoS Genet* 2013, **9**.

171. Girard C, Jouanin L: **Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, Phaedon cochleariae**. *Insect Biochem Mol Biol* 1999, **29**:1129–1142.

172. Todaka N, Inoue T, Saita K, Ohkuma M, Nalepa C a., Lenz M, Kudo T, Moriya S: **Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach**. *PLoS One* 2010, **5**:1–10.

173. Scharf ME, Wu-Scharf D, Pittendrigh BR, Bennett GW: **Caste- and development-associated gene expression in a lower termite.** *Genome Biol* 2003, **4**:R62.

174. Lee SJ, Kim SR, Yoon HJ, Kim I, Lee KS, Je YH, Lee SM, Seo SJ, Dae Sohn H, Jin BR: **cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, Apriona germari**. *Comp Biochem Physiol - B Biochem Mol Biol* 2004, **139**:107–116.

175. Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, ffrench-Constant RH: **Pyrosequencing of the midgut transcriptome of the poplar leaf beetle Chrysomela tremulae reveals new gene families in Coleoptera**. *Insect Biochem Mol Biol* 2009, **39**:403–413.

176. Ni J, Tokuda G: **Lignocellulose-degrading enzymes from termites and their symbiotic microbiota**. *Biotechnol Adv* 2013, **31**:838–850.

177. Nimchua T, Thongaram T, Uengwetwanit T, Pongpattanakitshote S, Eurwilaichitr L: **Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes**. *J Microbiol Biotechnol* 2012, **22**:462–469.

178. Zhao K, Guo LZ, Lu WD: **Extracellular production of novel halotolerant, thermostable, and alkali-stable carboxymethyl cellulase by marine bacterium Marinimicrobium sp. LS-A18**. *Appl Biochem Biotechnol* 2012, **168**:550–567.

179. Liu Z, Zhao X, Bai F: **Production of xylanase by an alkaline-tolerant marine-derived Streptomyces viridochromogenes strain and improvement by ribosome engineering**. *Appl Microbiol Biotechnol* 2013, **97**:4361–4368.

180. Fu C, Hu Y, Xie F, Guo H, Ashforth EJ, Polyak SW, Zhu B, Zhang L: **Molecular cloning and characterization of a new cold-active esterase from a deep-sea metagenomic library**. *Appl Microbiol Biotechnol* 2011, **90**:961–970.

181. Wierzbicka-Woś A, Bartasun P, Cieśliński H, Kur J: **Cloning and characterization of a novel cold-active glycoside hydrolase family 1 enzyme with β-glucosidase, β-fucosidase and β-galactosidase activities.** *BMC Biotechnol* 2013, **13**:22.

182. Graham JE, Clark ME, Nadler DC, Huffer S, Chokhawala H a, Rowland SE, Blanch HW, Clark DS, Robb FT: **Identification and characterization of a**

multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nat Commun* 2011, **2**:375.

183. Alvarez TM, Goldbeck R, Santos CR Dos, Paixão D a a, Gonçalves T a., Franco Cairo JPL, Almeida RF, de Oliveira Pereira I, Jackson G, Cota J, Büchli F, Citadini AP, Ruller R, Polo CC, de Oliveira Neto M, Murakami MT, Squina FM: **Development and Biotechnological Application of a Novel Endoxylanase Family GH10 Identified from Sugarcane Soil Metagenome**. *PLoS One* 2013, **8**.

184. Weerachavangkul C, Laothanachareon T, Boonyapakron K, Wongwilaiwalin S, Nimchua T, Eurwilaichitr L, Pootanakit K, Igarashi Y, Champreda V: **Alkaliphilic endoxylanase from lignocellulolytic microbial consortium metagenome for biobleaching of eucalyptus pulp**. *J Microbiol Biotechnol* 2012, **22**:1636–1643.

185. Bao L, Huang Q, Chang L, Sun Q, Zhou J, Lu H: **Cloning and characterization of two beta-glucosidase/xylosidase enzymes from yak rumen metagenome.** *Appl Biochem Biotechnol* 2012, **166**:72–86.

186. Zhou J, Bao L, Chang L, Zhou Y, Lu H: **Biochemical and kinetic characterization of GH43 beta-D-xylosidase/alpha-L-arabinofuranosidase and GH30 alpha-L-arabinofuranosidase/beta-D -xylosidase from rumen metagenome.** *J Ind Microbiol Biotechnol* 2012, **39**:143–152.

187. Wong DWS, Chan VJ, McCormack AA: **Functional cloning and expression of a novel Endo-alpha-1,5-L-arabinanase from a metagenomic library.** *Protein Pept Lett* 2009, **16**:1435–1441.

188. Wong DWS, Chan VJ, Liao H, Zidwick MJ: **Cloning of a novel feruloyl esterase gene from rumen microbial metagenome and enzyme characterization in synergism with endoxylanases.** *J Ind Microbiol Biotechnol* 2013, **40**:287–295.

189. Rashamuse KJ, Visser DF, Hennessy F, Kemp J, Roux-van der Merwe MP, Badenhorst J, Ronneburg T, Francis-Pope R, Brady D: **Characterisation of two bifunctional cellulase-xylanase enzymes isolated from a bovine rumen metagenome library.** *Curr Microbiol* 2013, **66**:145–151.

190. Watanabe A, Yano K, Ikebukuro K, Karube I: **Cyanide hydrolysis in a cyanide-degrading bacterium, Pseudomonas stutzeri AK61, by cyanidase.** *Microbiology* 1998, **144 (Pt 6**:1677–1682.

191. Liu N, Yan X, Zhang M, Xie L, Wang Q, Huang Y, Zhou X, Wang S, Zhou Z: **Microbiome of fungus-growing termites: a new reservoir for lignocellulase genes.** *Appl Environ Microbiol* 2011, **77**:48–56.

192. Zhang M, Liu N, Qian C, Wang Q, Wang Q, Long Y, Huang Y, Zhou Z, Yan X: **Phylogenetic and Functional Analysis of Gut Microbiota of a Fungus-Growing Higher Termite: Bacteroidetes from Higher Termites Are a Rich Source of β-Glucosidase Genes**. *Microb Ecol* 2014, **68**:416–425.

1231

Table 1
Click here to download Table: Table 1.docx

Table 1. Phyla distribution in different lignocelulolytic substrates.

	Metagenome source	Sugarcane Bagasse pile (Rattanachomsri et al. [70])	Poplar chips (Van der Lelie et al. [133])	Sugarcane Bagasse pile (Wongwilaiwalin et al. [147])
	Phyla			
Bacteria	Acidobacteria	Yes (10.9%)	Yes (minor %)	
	Actinobacteria	Yes (minor %)	Yes (minor %)	Yes (minor %)
	Aquificae			
	Bactereoidetes	Yes (15.3%)	Yes (9.9%)	Yes (23.8%)
	Chlamydiae			
	Chlorofobi			
	Chloroflexi			
	Chrysiogenetes			
	Cyanobacteria		Yes (minor %)	
	Deferribacteres			
	Deinococcus-Thermus	Yes (minor %)		
	Disctyoglomi			
	Fibrobacteres			
	Firmicutes	Yes (35.5%)	Yes (45.9%)	Yes (50.1%)
	Fusobacteria			
	Gemmatimonadetes			
	Lentisphaerae			
	Nitrospirae			
	Planctomycetes	Yes (minor %)	Yes (minor %)	Yes (minor %)
	Proteobacteria	Yes (24.6%)	Yes (32.3%)	Yes (14.1%)
	Spirochaetes	Yes (minor %)		Yes (5.1%)
	Synergistetes			Yes (minor %)
	Thermodesulfobacteria			
	Thermotogae			Yes (minor %)
	Verrucomicrobia			
Fungi	Ascomycota	Yes	N.I.	ND
	Basidiomycota	ND	N. I.	ND
	Chytridiomycota	ND	N. I.	ND
	Glomeromycota	ND	N. I.	ND
	Zygomycota	ND	N.I.	ND

ND=not determined, NI=non identified

Table 2
Click here to download Table: Table 2.docx

Table 2. Hydrolytic enzymes of bacteria and fungi identified by culture-independent methods.

Culture-Independent Methods	Candidate glycoside hydrolase genes/clone	Substrate used	Number of Clones with Assayed and characterized Activity/Reference
Metagenome sequencing from switchgrass-adapted compost community (548 733 reads)	800 candidates (25 full length)	carboxymethyl-cellulose	1 Endoglucanase (GH9)/[141]
Pyrosequencing (1 283 902 reads)	37 candidate genes	Avicel	4 Endoglucanases (GH5)/[182]
Metagenomic library from sugarcane field land soil (26 900 clones of 1-8kb size)	1 candidate gene	Carboxymethyl-cellulose	1 Endoglucanase (GH5)/[114]
Metagenomic DNA from sugarcane bagasse compost	Several genes	Sugarcane bagasse and filter paper	1 Endoxylanase (GH10)/[183]
	1 candidate gene	Xylan beechwood	1 Endoxylanase (GH10)/[184]
Metagenomic DNA pyrosequencing (144 253 raw reads)	57 candidate genes	Carboxymethyl-cellulose, birchwood xylan or β-glucan	General presence of cellulases, and xilanases/[147]
Metagenome cosmid library of yak rumen	4000 candidate clones	4-nitrophenyl-β-d-glucopyranoside (pNPG), 4-nitrophenyl-β-d-xylopyranoside (pNPX)	1 β-glucosidase (GH3), 1 β-xylosidase (GH3)/[185]
		Fluorescent 4-methylumbelliferyl-β-d-xylopyranoside (MuX)	1 β-xylosidase (GH 43), 1 α-l-arabinofuranosidase (GH 30)/[186]
Metagenome DNA library of cow rumen	Not mentioned (refered to Wong et al. [187])	p-nitrophenyl ferulate	1 feruloyl esterase/[188]
Metagenome fosmid DNA library of bovine rumen	70 000 clones	Carboxymethyl-cellulose and birchwood xylan	2 Bifunctional Cellulase–Xylanase (GH5)/[189]

Table 3
Click here to download Table: Table 3.docx

Table 3. Hydrolytic enzymes of insect and their endosymbiont identified by culture-independent methods.

Order/Species	Culture-Independent Methods	Origin	Candidate genes/clone	Number of clones with assayed and characterized activity/Reference
Termite/*Nasutitermes aphratea*	454 pirosequencing	Hindgut Symbiont	33 symbiotic bacterial	NP/[57]
Termite/*Reticulitermes flavipes*	Metatranscriptomic cDNA library	Endogenous and Symbiont	171 candidate gene encoding lignocellulases	NP/[168]
Termite/*Reticulitermes speratus*	cDNA libraries	Endogenous	1 endoglucanase	1 enzyme/[190]
Termite/*Microcerotermes sp.*	Metagenomic fosmid libraries	Endogenous	2 cellulases 12 xylanases	3 enzymes/[177]
Termites/*Macrotermes annandalei*	Metagenomic fosmid libraries	Gut Symbiont	13 gene encoding cellulases	1 xylanase/[191]
Termite/*Reticulitermes flavipes*	cDNA libraries and macroarrays	Endogenous and Symbiont	4 cellulases	NP/[173]
Termite/*Pseudacanthotermes militaris*	Two Fosmid libraries	Gut symbiont	101 positive clones	6 enzymes candidates/[169]
Termites/fungus-growing termite, *Macrotermes annandalei*	Fosmid libraries 454pyrosequencing	Gut Symbiont	10 gene encoding putative β-glucosidase	3 enzymes/[192]
Mastotermitidae/*Mastotermes darwinenesis* Termopsidae/*Hodotermopsis sjoestedti* Kalotermitidae/*Neotermes koshunensis* Rhinotermitidae/*Reticulitermes speratus* Cryptocercidae/*Crytocercus puntulatus*	cDNA libraries	Symbiotic protists	74 cellulases/ 24 xylanases 89 cellulases/ 17 xylanases 90 cellulases/ 73 xylanases 74 cellulases/ 11 Xylanases 31 cellulases/ 14 xylanases	NP/[172]
Grasshopper/*Acrida cinerea* Cutworm/*Agrotis ípsilon* *Termite/Nasutitermes sp.*	Metagenomic sequencing. Illumina genome analyzer II	Gut Symbiont	31 (*A. cinerea*), 40 (*A. ípsilon*) and 52 (*Nasutitermes sp.*)	4 enzymes/[170]
Termite/*taxonomic identification was not conducted*	Genomic libraries	Gut Symbiont	1 xylanase 3 xylanases	4 enzymes/[167]
Coleoptera/*Apriona germari*	cDNA libraries	Endogenous	1 endoglucanase	1 enzyme/[174]
Coleoptera/*Phaedon cochleariae*	cDNAs gut library	Endogenous and Symbiont	7 genes encoding putative alpha-amylase, cysteine proteinase, trypsin, chymotrypsin, cellulase, pectinase and xylanase	NP/[171]
Limnoriidae/*Limnoria quadripunctata*	Transcriptome	Endogenous	4 cellulases	NP/[165]
Coleotera/*Chrysomela tremulae*	454 pyrosequencing	Endogenous	20 gene encoding cellulases	NP/[175]

NP = not provided by the authors

Figure 1
Click here to download high resolution image

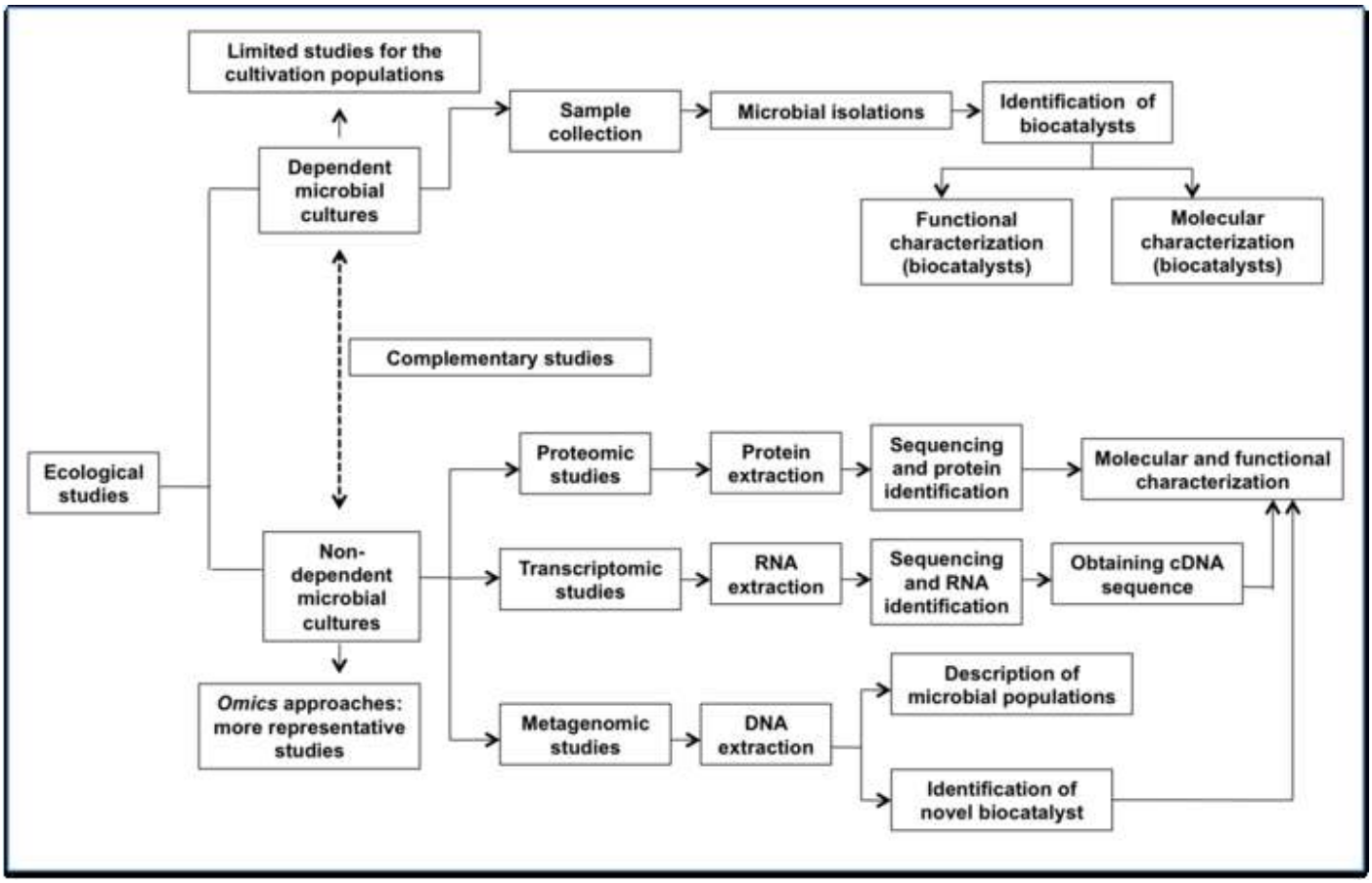


Figure 2
Click here to download high resolution image

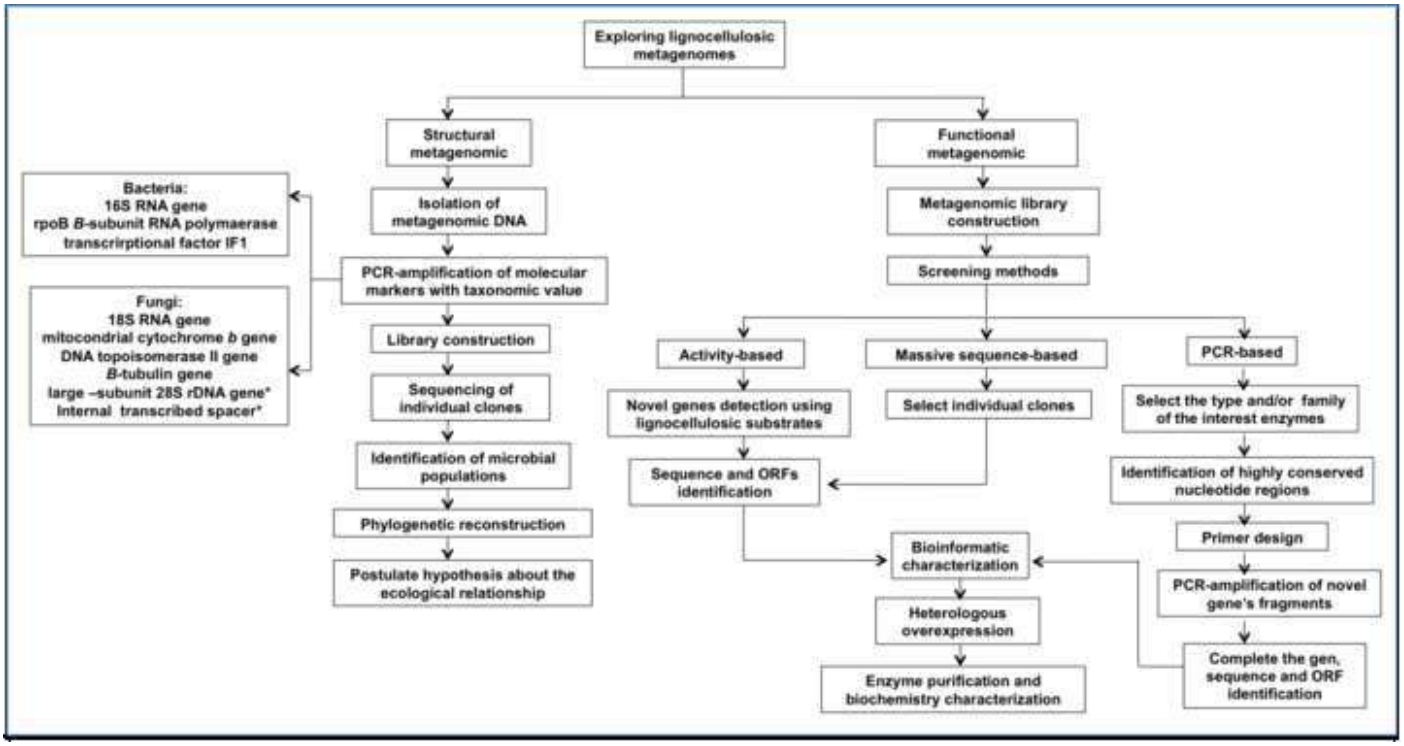


Figure 3
Click here to download high resolution image

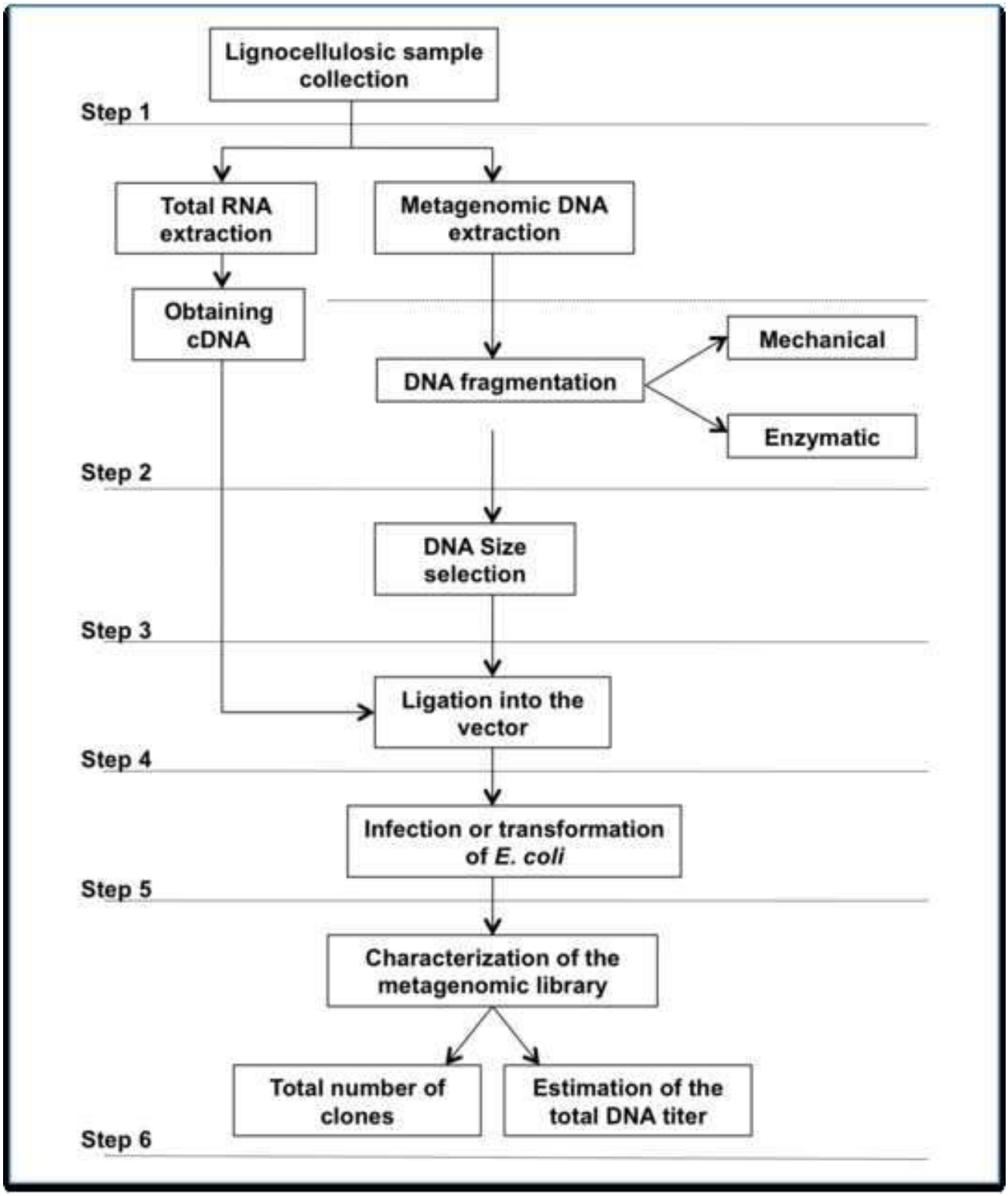
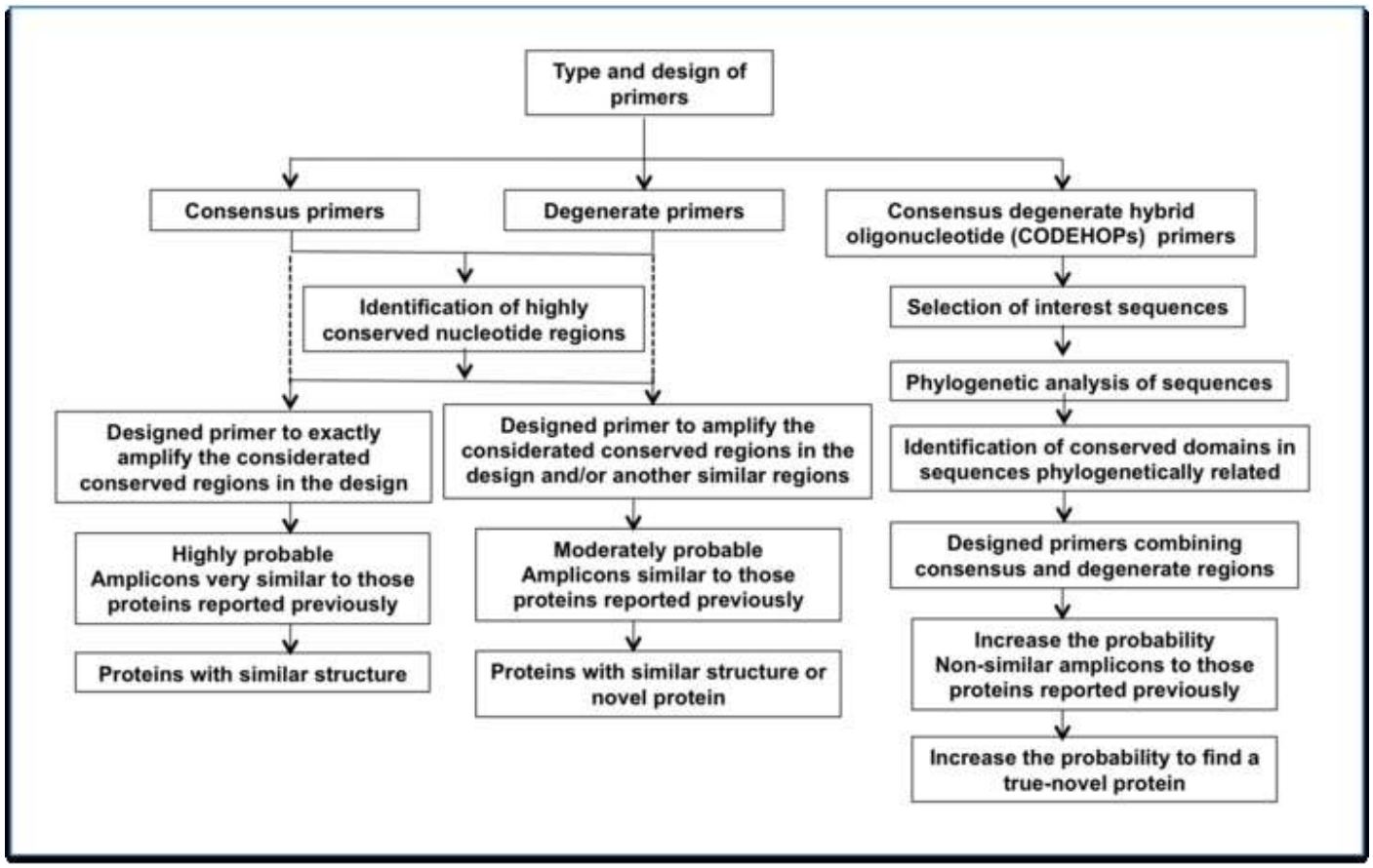
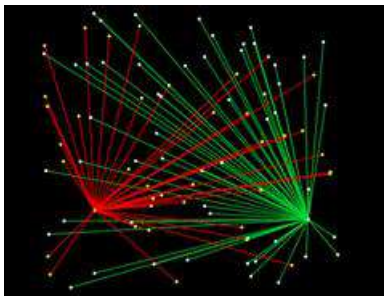


Figure 4
Click here to download high resolution image





Capítulo II.

Análisis de la diversidad procariótica a partir de un metagenoma de bagazo de caña de azúcar.

Como fue discutido en el capítulo anterior, los estudios metagenómicos estructurales permiten informar los perfiles de biodiversidad que se encuentran en un hábitat determinado. Son muy pocos los análisis sobre biodiversidad microbiana considerando técnicas independientes de cultivo en materiales ricos en lignocelulosa. Sólo existe un grupo tailandés con un par de antecedentes en esta área. Sin embargo, en Latinoamérica (primera región del mundo en producción de azúcar y por tanto de bagazos) no se dispone de información al respecto. Considerando que México es de los principales productores de bagazo de caña de azúcar en la región, y nuestro interés en describir las poblaciones microbianas asociadas a este material, realizamos el estudio de la composición estructural de las poblaciones bacterianas creciendo en una muestra de bagazo en fermentación: *"Prokaryotic diversity from the culture-independent taxonomic analysis of a sugarcane bagasse metagenome"*.

Este estudio permitió identificar los principales phyla, géneros y especies de bacterias (a través de librerías de RNA16S) relacionadas con la degradación del bagazo de caña de azúcar, y a la vez proponer sus posibles relaciones y nichos ecológicos. Adicionalmente, demostró que existe una elevada convergencia taxonómica entre las poblaciones descritas en este trabajo y los filotipos relacionados con procesos de degradación de biomasa vegetal en bagazos tailandeses. Debemos señalar que la mayoría de las secuencias encontradas en nuestro estudio muestran relaciones filogenéticas distantes con la secuencias de referencias consideradas en las reconstrucciones filogenéticas. Este hecho demuestra que el bagazo de caña es un sustrato ideal para describir poblaciones bacterianas lignocelulolíticas con nuevos filotipos, incluso especies. Derivado de esta consideración, resulta interesante el estudio de la fisiología de estos organismos para posteriores usos en biorrefinerías y degradación de lignocelulosa. Sus potencialidades metabólicas deben divergir de las descritas usualmente, de acuerdo a las relaciones filogenéticas que manifiestan con especies de referencia. En general fue poca la biodiversidad encontrada, este hecho se relaciona con la composición química de este material y con su desbalance de nutrientes y humedad.

Ecology and Evolution Open Access

Prokaryotic diversity from the culture-independent taxonomic analysis of a sugarcane bagasse metagenome

Journal:	*Ecology and Evolution*
Manuscript ID:	Draft
Wiley - Manuscript type:	Original Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Batista-García, Ramón Alberto; Universidad Autónoma Estado Morelos, Centro de Investigación en Dinámica Celular-Centro de Investigación en Biotecnología Casasanero, Rocío; Universidad Autónoma Estado Morelos, Centro de Investigación en Biotecnología Alvarez-Castillo, Alberto; Instituto Tecnológico de Zacatepec, Biotecnología Jackson, Stephen A.; University College Cork, Environmental Research Institute Dobson, Alan; University College Cork, Environmental Research Institute; University College Cork, School of Microbiology Folch-Mallol, Jorge Luis; Universidad Autónoma Estado Morelos, Centro de Investigación en Biotecnología
Search Terms:	Molecular Ecology, Bacteria, Ecosystems
Abstract:	Aim: We aim to analyse for the first time in Meso America the prokaryotic diversity profile of metagenome isolated from sugar cane bagasse. Location: Morelos, Mexico. Methods: Metagenomic DNA from fermented sugar cane bagasse was isolated and a 16S rDNA metagenome library was constructed. Microbial taxonomic study, and subsequently, a phylogenetic analysis were conducted. Results: The biodiversity present in the bagasse samples was explored and we identified novel bacterial groups inhabiting this lignocellulosic rich substrate. The most abundant phyla were Proteobacteria and Acidobacteria, while Candidate division TM7 was the minor phyla. Overall, the prokaryotic microbial diversity revealed the presence of 13 families and 17 genera, with Burkholderiaceae and Burkholderia being the major family and genus observed, respectively. We compared the bacterial biodiversity found in the Mesoamerican sugarcane bagasse with other Thai bagasse, concluding that the bacteria inhabiting both bagasses show a closely phylogenetic relationship. Since we were also able to detect some known bacterial genera, we could propose some specific metabolic pathways that may be important in lignocellulose degradation.

Main conclusions: Microbial communities inhabiting natural lignocellulosic substrates have to date received little attention, subsequently there is little information about the main bacterial genera present in these substrates resulting in a lack of a comprehensive knowledge about the microbial process involved in lignocellulosic degradation. This study provides some insight into the unique microbial community structure in sugarcane bagasse from Meso America (the most important sugar producing region in the world) and is the first such study to characterize these bagasse inhabiting populations.

Keywords: Bacterial communities; cellulolytic environments; metagenomic approaches; structural metagenomic; sugarcane bagasse; 16S rDNA

Original Article

Prokaryotic diversity from the culture-independent taxonomic analysis of a sugarcane bagasse metagenome

Ramón Alberto Batista-García[1-2]*, Rocío Casasanero[2], Alberto Alvárez-Castillo[3], Stephen A. Jackson[4], Alan D. W. Dobson[4-5] and Jorge Luis Folch-Mallol[2]

[1]Centro de Investigación en Dinámica Celular, Universidad Autónoma del Estado de Morelos (UAEM). Ave. Universidad 1001. Col. Chamilpa 62209. Cuernavaca, Mor. Mexico. [2]Centro de Investigación en Biotecnología, UAEM. Ave. Universidad 1001. Col. Chamilpa 62209. Cuernavaca, Mor. Mexico. [3]Instituto Tecnológico de Zacatepec. Calzada Tecnológico 27. Zacatepec 62780, Morelos. Mexico. [4]Environmental Research Institute, University College Cork (UCC). Lee road. Cork. Ireland. [5]School of Microbiology, UCC. Western road. Cork. Ireland.

*__Corresponding author__

Ramón Alberto Batista-García

Ave. Universidad 1001. Col. Chamilpa 62209. Cuernavaca, Mor., Mexico.

Tel. +(52) 777-3297057

rbatista25@yahoo.com, rabg@uaem.mx

Short running head: Prokaryotic diversity from a sugarcane bagasse metagenome

Total words: 6,905

ABSTRACT

Aim: We aim to analyse for the first time in Meso America the prokaryotic diversity profile of metagenome isolated from sugar cane bagasse.

Location: Morelos, Mexico.

Methods: Metagenomic DNA from fermented sugar cane bagasse was isolated and a 16S rDNA metagenome library was constructed. Microbial taxonomic study, and subsequently, a phylogenetic analysis were conducted.

Results: The biodiversity present in the bagasse samples was explored and we identified novel bacterial groups inhabiting this lignocellulosic rich substrate. The most abundant phyla were *Proteobacteria* and *Acidobacteria*, while Candidate division TM7 was the minor phyla. Overall, the prokaryotic microbial diversity revealed the presence of 13 families and 17 genera, with *Burkholderiaceae* and *Burkholderia* being the major family and genus observed, respectively. We compared the bacterial biodiversity found in the Mesoamerican sugarcane bagasse with other Thai bagasse, concluding that the bacteria inhabiting both bagasses show a closely phylogenetic relationship. Since we were also able to detect some known bacterial genera, we could propose some specific metabolic pathways that may be important in lignocellulose degradation.

Main conclusions: Microbial communities inhabiting natural lignocellulosic substrates have to date received little attention, subsequently there is little information about the main bacterial genera present in these substrates resulting in a lack of a comprehensive knowledge about the microbial process involved in lignocellulosic degradation. This study provides some insight into the unique microbial community structure in sugarcane bagasse from Meso America (the most important sugar producing region in the world) and is the first such study to characterize these bagasse inhabiting populations.

Keywords: Bacterial communities; cellulolytic environments; metagenomic approaches; structural metagenomic; sugarcane bagasse; 16S rDNA

INTRODUCTION

Plant biomass is the most widespread source of carbon on earth, and its use is yet under exploited. However, more recently there has been an increased interest in the integrated management of biomass, and especially of sugarcane bagasse (Guimarães et al., 2012). Sugarcane bagasse is predominantly composed of cellulose, hemicellulose, pectin and lignin (Rezende et al., 2011). Recalcitrant crystalline structures of these polymers limit bagasse conversion into fermentable sugars for its further use (Geng et al., 2014).

There is an increased interest in the use of cultured-independent metagenomic based taxonomic analysis of microbial communities present in the lignocellulosic rich bagasse. A more comprehensive knowledge of the sugarcane bagasse microbiome will provide further insights into the microbial community structure and metabolic potential and help define their potential role as lignocellulose degraders, thereby increasing our understanding of the natural degradation processes that occur in lignocellulolytic environments. Moreover, it may ultimately help facilitate a better and more optimal use of bagasse as an industrial resource for biofuels and other products.

Bacteria are the most abundant microorganisms on earth. They colonize a huge variety of environments, including very specialized ecological niches such as lignocellulosic habitats; which is largely possible due to the immense metabolic diversity and genetic adaptability of microbes (Leis et al., 2013). Given that only an estimated <1% of microorganisms are cultivable (Simon & Daniel, 2011), metagenomic based tools have been successfully employed to conduct comprehensive ecological studies of bacterial communities in a wide variety of ecosystems. Structural metagenomics allows both the composition and structure of microbial communities to be assessed by describing the major genera and species that inhabit an ecosystem (Simon & Daniel, 2011). It can also provide information about the ecological interactions between members of different microbial communities (Simon & Daniel, 2009).

While a wide range of quite diverse metagenomes have to date been extensively analysed (Curson et al., 2010; Rasheed et al., 2013; Patel et al., 2014; Xu et al., 2014) very few metagenomic studies have been performed of lignocellulosic based substrates (Rattanachomsri et al., 2011; Kanokratana et al., 2013). Sugarcane bagasse piles are an excellent cellulolytic environment and represent a unique habitat in which to study lignocellulosic microbial communities. The prokaryotic populations colonizing sugarcane bagasse in particular and other cellulosic materials in general have surprisingly been quite poorly studied to date and consequently there is lack of a comprehensive understanding of their microbial ecology. On other hand, there is not any study analysing

the microbial population in sugar cane bagasse in Meso America being very important to compare these bacterial communities with those described in other ecogeographical ecosystems.

The objective of our work was to analyse the prokaryotic diversity of a sugarcane bagasse metagenome, and subsequently, the structural composition of the bacterial communities was obtained by employing cultured-independent methods. The V3–V6 region of the 16S rDNA gene was used to analyse the bacterial populations, and a sequence dataset was analysed with strong algorithms using two taxonomic classifiers. Our groups are interested in biological pre treatments of lignocellulosic materials for biorefinery purposes and by exploring the biodiversity profiles, this work provides new insights to better understand the microbial populations present in sugarcane bagasse piles and thus may provide insights into new strategies for more efficient lignocellulose degradation.

MATERIALS AND METHODS

Bagasse sampling

Decaying humid bagasse sampling was performed on April 2008 in a Sugar Mill (N 18°39'15.9", W 99°11'10.1") (Fig. 1a). Bagasse samples were collected from large open-air piles, which had been standing for five months prior to the time of collection.

Three triplicate samples (1 kg) from three different piles were taken. These were collected from different positions in each pile, namely from the surface, the core and the bottom of the pile. The samples were then mixed to homogeneity prior to DNA extraction as shown in Fig. 1b. Subsequently, a composted sample for further experiments was generated, and consequently a representative sample was obtained (Fig. 1b). The bagasse samples were placed in sterile plastic bags and stored on dry ice. After that a composted sample was fermented in the solid state (30 days at 28^{o}C). The bagasse fermentation allowed to enrich the microbial communities involved in the lignocellulose degradation (object of study in this work).

Chemical bagasse characterization

The lignin, cellulose and hemicellulose composition was determined according to (Dominguez-Dominguez et al., 2012). The bagasse was ground and dried at 70^{o}C for 24 h. To determine the cellulose composition, 15 mL of $C_2H_4O_2$ (80%) and 1.5 mL of concentrated HNO_3 were added to 1 g of milled bagasse. The sample was treated at reflux for 20 min and filtered thereafter. The residue was washed with absolute ethanol, baked dried at 100^{o}C and finally weighted (material A). It was later burned at 540^{o}C, cooled to room temperature in a desiccator and weighted (material B). The cellulose percentage was calculated using the following equation:

$$\%\ cellulose = \frac{material\ A - material\ B}{sample\ weight} * 100$$

To determine the lignin composition, 70 mL of H_2SO_4 (1.25%) was added to 1 g of milled bagasse. The sample was treated at reflux for 120 min, filtered thereafter and then washed with water. Subsequently 30 mL of H_2SO_4 (72%) was added and the material was agitated at 100 rpm for 4 h. The solids were then recovered, filtrated, washed, dried (100 °C) and weighted (material C). The material was cremated at 540°C, cooled to room temperature in a desiccator and weighted (material D). The lignin percentage was calculated using the following equation:

$$\%\ lignin = \frac{material\ C - material\ D}{sample\ weight} * 100$$

The percentage of hemicellulose was calculated considering the cellulose and lignin compositions:

$$\%\ hemicellulose = 100\% - (\%\ cellulose + \%\ lignin)$$

The residual reducing sugars were calculated using the 3,5-dinitrosalicylic acid method (Miller, 1959). The total reducing sugars were expressed per g of bagasse.

Metagenomic DNA extraction from bagasse

10 g of bagasse (after fermentation, see 2.1) was suspended in 100 mL of lysis buffer (Brady, 2007). The mixture was frozen on dry ice and later defrosted at 80°C to promote cell lysis. This procedure was repeated three times. Bagasse fibres and suspended particles were removed by centrifugation (12000 rpm) and subsequent filtration (0.45 µm). The resulting solution was treated with phenol-chloroform-isoamyl alcohol (25:24:1) to remove humic acids and other co-extracted contaminating substances. Finally it was treated with chloroform to remove phenol traces. DNA was precipitated with ethanol, recovered in 50 µL of water, analysed by electrophoresis and quantified using a Nanodrop ND-1000. Finally, an additional purification step was performed using the Fermentas DNA extraction kit. DNA was extracted from three independent samples.

PCR amplicon library preparation for sequencing

A 16S rRNA PCR amplicon gene library of the full-length V3–V6 region (1300 bp) was prepared using metagenomic DNA as a template. The PCR primers and conditions were described by (Felske et al., 1996). Three individual PCR reactions were performed.

Amplicons were purified and ligated to pGEM-T Easy vector. Clones derived from electrotransformed *Escherichia coli* DH5α were selected on Luria-Bertani agar plates containing isopropyl-β-D-thiogalactopyranoside/5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (3 and 15 µg/mL, respectively) and ampicillin (100 µg/mL). Representative colonies were randomly selected and subsequently, sequenced (for both extremes) in the Langebio's Genomic Services Unit (Mexico). Sequences were assembled in the DNA Baser edit program version 4.16.0.

Sequencing data analysis

All sequences were processed by Ribosomal Database Project (RDP) pipeline and NGS analysis pipeline of the SILVA rRNA gene database project (SILVAngs 1.1) (Quast et al., 2013).

The sequencing dataset was quality-filtered using the RDP (Release 11.2, 2014) and SILVAngs (Release 115, 2014) pipelines. During initial processing reads with more than 2% of ambiguities or 2% of homopolymers, shorter than 100 bases, sequences of low alignment quality (50 alignment identity, 40 alignment score) and identical reads were excluded from downstream analysis. Primer portions were trimmed, and the orientation of reads and chimera formation were checked.

Unique sequences were aligned using the secondary-structure INFERNAL aligner procedure in the RDP pipeline (Nawrocki & Eddy, 2007) and the SILVA Incremental Aligner (Pruesse et al., 2012). Furthermore, (OTUs: operational taxonomic units) sequences were clustered by complete-linkage clustering. The RDP Classifier allowed the assignment of taxa using naïve Bayesian algorithm with a confidence threshold of 95% (Wang et al., 2007; Cole et al., 2013); while SILVA Classifier classified each OTU using cd-hit-est (version 3.1.2) (Li & Godzik, 2006) running in *accurate mode*, ignoring overhangs, and applying identity criteria of 1.00 and 0.98, respectively. The classification was performed by a local nucleotide BLAST search against the non-redundant version of the SILVA SSU Ref dataset (release 115) using blastn (version 2.2.28+) with default settings (Camacho et al., 2009). According with the function "[(% sequence identity + % alignment coverage)/2]", reads without any BLAST hits or reads with weak BLAST hits showing values lower than 93, remain unclassified (Ondov et al., 2011). Subsequently, phylum and full-taxonomic fingerprints were constructed (Ionescu et al., 2012; Klindworth et al., 2013). The taxonomic nomenclature considered is in accordance with the Bergey's Manual of Systematic Bacteriology (Garrity & Holt, 2012). Rarefaction curves, Shannon – Weaver index, the Chao1 richness estimator and the abundance-base coverage estimator were obtained using RDP tools.

Phylogenetic trees were prepared using the server Phylogeny.fr. MUSCLE and ClustalW for the multiple alignments, Gblocks for the alignment curation and neighbour-joining method were used in the analysis (Dereeper et al., 2008).

A network visualization showing OTUs interaction was generated to compare the prokaryotic diversity derived of our bagasse sample with a dataset obtained from the Thai bagasse (GenBank accession numbers HM362440-HM362617) (Rattanachomsri et al., 2011). The network map was prepared in Qiime (Quantitative Insights Into Microbial Ecology) version 1.8.0 (Caporaso et al., 2010) and was visualised in Cytoscape program version 3.1.1 (Lynch et al., 2013). Silva and RDP classifiers were used to prepare the network files in Qiime.

The 16S rRNA gene dataset was deposited in GenBank under accession numbers: KM882649-KM882821.

RESULTS

Chemical sugarcane bagasse composition

While the residual reducing sugar analysis showed 0.18 mg/g bagasse, the chemical composition revealed that the bagasse was composed of 53.73% cellulose, 26.37% lignin and 19.92% hemicellulose. This data confirms the recalcitrant nature of this waste product.

Metagenomic DNA isolated from sugarcane bagasse

For metagenomic analysis, the quantity of DNA isolated is important to ensure adequate representation of the microbial communities in the sample. Isolation and purification of metagenomic DNA from lignocellulosic rich substrates has a number of drawbacks including the fact that there are no standardized methods available for nucleic acid extraction. In this study an average of 0.5 µg metagenomic DNA per 1 g of bagasse was isolated. Crude DNA solutions prepared from the bagasse were extremely dark and viscous suggesting the presence of contaminants. The isolation of DNA resulted in co-extraction of other undesirable compounds. Repeated DNA purification steps were therefore undertaken to ensure the removal contaminants, with the DNA being purified until complete decolouration was observed. Overall DNA quality was assessed by reading absorbance (A), with a ratio A260/A280=1.79 being obtained, confirming that DNA purity was suitable for metagenomic analyses.

Prokaryotic microbial diversity

A library of 500 clones with amplicons derived from 16S rDNA PCR was obtained, and 200 clones were subsequently analysed. Following quality filtering, 177 reads with an average length of 1305 bp were analysed. During the quality control we identified and discarded: five sequences that showed more than 2% of

ambiguities, nine that were shorter than 100 bases, four that revealed low alignment quality and five that were classified as duplicated reads.

Three bacterial phyla (*Acidobacteria*, Candidate division TM7 and *Proteobacteria*) were observed. Comparison of OTUs against RDP and SILVA shows that the *Proteobacteria* and *Acidobacteria* were the largely predominant bacterial lineages constituting 60 and 34% respectively (Fig. 2a). According to the SILVA classifier the classification resulted in 100% of the sequences being classified (phylum level) in more than 40 OTUs, and subsequently the full-taxonomic fingerprint was obtained (Fig. 2b). Overall, the prokaryotic diversity revealed the presence of 13 families and 17 genera, with *Burkholderiaceae* and *Burkholderia* being the major family and genus respectively. On the other hand, the taxonomic analysis based on the RDP classifier also resulted in 100% of the sequences being classified at the phylum level. However, this classifier (with similar confidence threshold as that of the SILVA classifier) identified 48 hits as unclassified sequenced at the division-class-genera-species levels; with a large number of acidobacterial sequences (78.7%) not being classified below the phylum level. Similarly one sequence from the phylum *Proteobacteria* could not be classified below the phylum level. Although all sequences were assigned to phyla, 27.1% of reads were not classified at genus level. Five genera (*Acidobacteriaceae* family) belonging to the phylum *Acidobacteria* were identified (Fig. 3a), while the phylum *Proteobacteria* showed the following distribution: *Betaproteobacteria* (35%), *Alphaproteobacteria* (18%), *Gammaproteobacteria* (5%) and *Deltaproteobacteria* (2%) (Fig. 3b). Taxonomic classification revealed that the best-represented genus in the class *Alphaproteobacteria* is *Hyphomicrobium*, with an additional four genera being identified (Fig. 3c). Nine reads which classified in the *Caulobactereaceae* and *Beijerinckiaceae* families were related to uncultured bacteria. *Achromobacter* and *Burkholderia*, and *Haliangium* genera represent *Betaproteobacteria* and *Deltaproteobacteria* classes respectively (Fig. 3c and Fig. 3e). Sixty-two sequences were grouped in the genus *Bulkholderia* (Fig. 2b and Fig. 3d). Nine reads were classified as *Gammaproteobacteria* and related to different genera. Three sequences belonging to the *Xanthomonadaceae* family also showed similarity with uncultured bacteria (Fig. 3f).

Rarefaction curves were obtained at different genetic distance levels in order to compare the richness of the genetic diversity as a function of number of sequences (Fig. 4). A typical trend was observed at higher genetic distance level (0.01, 0.03, 0.05, 0.15 and 0.28). The curve at a genetic distance level of 0.28 (phylum level) showed a rapid asymptotic behaviour according to the low phylum diversity found in the bagasse.

The number of OTUs in the sample was determined with Shannon and Chao1 diversity indexes being calculated (Table 1). Identification of unique phylotypes using RDP and SILVA pipelines showed 42 OTUs from the

metagenomic dataset analysis. Chao1 species richness estimates predict 59.21 OTUs at ≥98% sequence identity for the sugarcane bagasse sample suggesting that 29% of the diversity was not sampled, while on the other hand, Chao1 phylum and class richness show that 100% of the phylum and class level diversities were sampled.

DISCUSSION

It is clear that lignocellulosic rich environments contain microbial diversity that remains as yet largely understudied. The increasing demand of biorefineries to use bagasse as a raw material to obtain products of associated high-value supports our approach of attempting to gain a better understanding of the overall structures of the microbial communities present in the bagasse (Lima et al., 2014). Thus a more comprehensive knowledge of the microorganisms present within bagasse may allow greater potential access to their metabolic capabilities which are important in allowing them inhabit this particular lignocellulose rich ecosystem. With this in mind this study focused on analysing the prokaryotic diversity of a sugarcane bagasse metagenome.

The sugarcane bagasse composition was similar to levels previously reported (Rattanachomsri et al., 2011; Rezende et al., 2011; Siqueira & Rocas, 2013). The bagasse polymeric composition suggests that quite unique microbial diversity is likely to be present in this lignocellulosic rich environment. Analysis of these bacterial population should provide valuable insights into not only overall population structures but also potentially of methods or approaches to control the microbial processes within these populations for optimal utilization of the bagasse for bio-fuel applications.

Regarding the metagenomic DNA extraction, our yield was low compared with previously described from soils (Yun et al., 2004; Amorim et al., 2008). As previously stated bagasse is a very recalcitrant lignocellulosic waste, primarily due to its chemical structure and composition (Chandel et al., 2014). This recalcitrance is likely not to allow the development of a wide variety and amount of microbial communities (Silva et al., 2009); which may explain the low yields of DNA obtained from the samples. Furthermore, its fibrous nature also constitutes a limitation for DNA extraction, while the granularity of a sample is also known to have a marked effect on the DNA yield (Brady, 2007).

The three phyla identified in the bagasse (*Acidobacteria*, Candidate division TM7 and *Proteobacteria)* have previously been reported to be dominant in lignocellulosic rich environments (Ghio et al., 2012; Huang et al., 2012; Talia et al., 2012). Our results confirm that these phyla are homogeneously represented in lignocellulosic eco-habitats. At the same time, 48 sequences were unclassified at the division-class-genera-species levels, these results strongly suggest that bagasse is an interesting lignocellulosic habitat in which to identify new strains,

species and even genera of bacteria. A notable proportion of sequence reads without classification to phylum level or below the phylum level has previously been found in structural metagenomic studies from cellulolytic and non-cellulolytic ecosystems (Kanokratana et al., 2013; Xu et al., 2014).

A large number of environmental metagenomes have to date been structurally analysed (Rasheed et al., 2013; Curson et al., 2014; Kanwar et al., 2014; Patel et al., 2014; Xu et al., 2014). However, the microbial composition presents in lignocellulosic environments remains largely under-studied. Only two other studies have to date undertaken a phylogenetic analysis of the complex microbial community structure in industrial bagasse, and both reported higher levels of diversity than had previously been anticipated with the prevalence of *Proteobacteria* (Rattanachomsri et al., 2011; Kanokratana et al., 2013). The low microbial diversity present in our sample is more than likely related to the structural complexity and recalcitrant nature of bagasse, resulting in a somewhat limited number of genera being capable of colonizing this ecosystem. Notwithstanding this, clear differences were observed in the microbial communities in our bagasse samples when comparing community structures reported for other lignocellulosic and non-lignocellulosic environments (Fig. 5). Previous studies are consistent that *Proteobacteria* and *Bacteroides* are the main phyla identified in lignocellulosic environments (Huang et al., 2012; Pope et al., 2012; Talia et al., 2012; Gruninger et al., 2014).

Regarding the rarefaction analysis, gradual increases observed in the curves at different levels (strain, species, genus) support the taxonomic analysis proposed by the RDP and SILVA´s classifier algorithms. Rarefaction curves for the bagasse samples showed some levelling off indicating that the library was representative and that the estimations of microbial diversity were likely to be accurate (Jackson et al., 2012).

On other hand, results derived from Shannon and Chao1 diversity indexes are comparable with other metagenomic studies which report effective sampling values close to 70%, while contrasting with those published results by authors that only recovered sequences representing between 50-60% of the metagenomic biodiversity (Rattanachomsri et al., 2011; Jackson et al., 2012; Kanokratana et al., 2013).

Finally, we compared our results with those obtained by the Thai group since these are the only ones currently available. OTUs-level comparisons using a network map showed that there is a close relationship between the identified OTUs in both sets of samples (Fig. 6). The network map did not reveal a limited degree of shared OTUs between both bagasse samples. This suggests that the microbiome of the bagasses was very similar in these ecogeographic habitats. In addition while a high convergence in network interaction was apparent, it should be noted that no relationships between some branches of the networks were observed. This visualization

is in accordance with the 16S rDNA-based taxonomic classification obtained for the two sugarcane bagasse samples.

Linking phylogenetic taxonomy to functional capability

The composition of microbial communities within a given ecosystem is known to reflect the metabolic roles of these bacteria (Jackson et al., 2012). All the genera identified contain at least one species with the potential to degrade lignocellulose. Bagasse is known to contain very low levels of available nitrogen and accessible organic matter. Its enriched composition of lignin, cellulose and hemicellulose together with the low free water content and other factors (acidic pH), largely limits the ability of many bacteria to colonize the bagasse resulting in somewhat limited microbial diversity coupled with quite long natural degradation processes (Goldbeck et al., 2014). There was a predominance of aerobic, microaerophilic and facultative anaerobic genera, while sequences related to anaerobic genera were not observed.

The phylum *Acidobacteria* comprises two classes: *Acidobacteria* and *Holophagae.* In this study we only found sequences related with the *Acidobacteria* class. Species of the Subdivision 1 (*Acidobacteria* class) can grow under conditions of nutrient deprivation and over quite a broad pH range (Männistö et al., 2012). Sugarcane bagasse is a moderately acidic and low-nutrient environment and therefore not surprisingly, perhaps around 34% of the analysed sequences were identified as *Acidobacterium, Terroglobus, Edaphobacter, Granulicella* and *Telmatobacter* (all genera belonging to the *Acidobacteria* class). Representatives of the phylum *Acidobacteria* have previously been reported to possess a high potential to degrade plant biomass (xylan, cellulose, hemicellulose, pectin). Indeed, their metabolic plasticity helps to ensure their survival in low-nitrogen environments such as bagasse (Ward et al., 2009). This explains the high percentage of *Acidobacteria,* which we observed in the sugarcane bagasse samples.

Acidobacteria also possess the capacity to produce extracellular cellulose, with cellulose synthesis being implicated in the molecular mechanisms involved in survival under stressful conditions, such as those likely to be encountered in lignocellulosic ecosystems. In addition the ability to synthesise extracellular cellulose suggests that at least some members of the *Acidobacteria* possess the ability to survive repeated cycles of rehydration and drying (Ward et al., 2009). In addition cellulose derived from acidobacterial (cellulose synthesis and/or bagasse degradation) is believed to promote colonization due to enhanced adherence to the substrate, biofilm-formation, thereby providing a loose network for nutrient and water retention. Thus these physiological aspects of extracellular cellulose production are likely to be important for survival of the *Acidobacteria* in the bagasse.

A phylogenetic reconstruction was conducted involving 60 bagasse derived sequences from the class *Acidobacteria* (Fig. 7). A 16S rDNA sequence subset of members of all species of this class was collected from NCBI. Overall, the phylogenic analysis reveals that the majority of the 16S rDNA reads are not grouped with previously described *Acidobacteria*. The RR168 sequence shows no phylogenetic relationship with the rest of the analysed sequences (Fig. 7). Only three sequences were directly grouped with species of the genera *Telmatobacter, Acidobacterium* and *Edaphobacter*. These results indicate the existence of a substantial number of new ribotypes from the 16S rDNA sequences obtained in this study. They may represent novel genera, species and even strains given that the aforementioned phylogenetic reconstruction shows a distant relationship with the previously reported genera of the *Acidobacteria* class.

Betaproteobacteria was the most represented class in the phylum *Proteobacteria*, being *Burkholderia* the most representative genus. *Burkholderia* comprises more than 70 species worldwide distributed (Mathew et al., 2014), and several species have been described for their lignocellulolytic potentials (Fujii et al., 2012). One of the most distinctive physiological characteristics of *Burkholderia* is their ability to produce siderophores with the primary role of iron acquisition/chelation (Vandamme et al., 2007). Siderophore-mediated nutrient acquisition has been identified as an important survival factor by *Burkholderia*, particularly in metal-poor environments (Thomas, 2007). Siderophore production is a likely strategy employed by *Burkholderia* to successfully colonize low-metal environments such as bagasse, to not only allow metal acquisition but to also promote mineral dissolution from fibrous materials within the bagasse (Shirvani & Nourbakhsh, 2010). Thus siderophore production by *Burkholderia* could play an important role in bagasse degradation by facilitating Fe and other metal uptake by the degrader microbial populations (e.g. *Acidobacteria* and *Burkholderia* species) under Fe^{2+}-limiting conditions.

In addition there are other important links between microbial siderophore production and lignocellulose. Different species of *Burkholderia* have been reported to produce catecholate and hydroxamate (de Los Santos-Villalobos et al., 2012; Kvitko et al., 2012). Some of these siderophores have been implicated in the redox speciation of Fe^{2+}, by facilitating the reduction of Fe^{3+}. The reaction between H_2O_2 and reduced Fe generates oxygen radicals, which are known to play an important role in the depolymerization of lignin, hemicellulose and cellulose (Xu & Goodell, 2001; Arantes & Milagres, 2007). While *Burkholderia* are known to protect themselves from H_2O_2 through the production of catalases/peroxidases (Lefebre et al., 2005).

It has been observed that siderophores are capable of decreasing lignocellulosic viscosity and promote its transformation (Milagres et al., 2002). Thus siderophore production by *Burkholderia* sp. within the bagasse

would clearly facilitate an important survival strategy both from a nutrient uptake and a lignocellulose biotransformation perspective.

Other minority taxonomic classes and subdivisions were found to represent approximately 29% of the analysed sequences. Cellulolytic activities have previously been described in some of the identified genera within this population (Wadell & Bang, 2008; Song et al., 2013). For example a read classified as *Methylovirgula* sp. was found in the 16S rDNA library. This genus contains only one described species to date namely *Methylovirgula ligni,* an obligate acidophile with a pH optimum between 4.5 and 5 (Vorob'ev et al., 2009), consistent with the moderate pH of bagasse.

A radial phylogenetic visualization shows the relationships of the sequences classified as *Proteobacteria* (60% of the total) (Fig. 8). 16S rDNA sequences (37 in total) belonging to type species of the genera and classes identified in the full-taxonomic fingerprint were collected in the NCBI. The phylogenic analyses revealed that only four 16S rDNA bagasse sequences were directly grouped with species, *Burkholderia caribiensis, Pseudomonas aeruginosa, Methylovirgula ligni* and *Beijerinckia derxii*. The phylogenetic visualization for the bagasse sequences identified as *Proteobacteria* demonstrates once again that there is a clear predominance of bacteria showing a truly distant relationship with the type species.

To the best of our knowledge, this is the first study in Mexico and indeed in Central or South America describing the complex composition of the microbial populations inhabiting sugarcane bagasse feedstock piles from sugar mills. Sugar mills in Mexico and in Central and South America, represent an important economic activity and bagasse generation is around 350 million metric tons per year on a worldwide basis (http://www.fas.usda.gov). These production levels generate serious pollution problems and demand the need for proper waste management strategies (Kiatkittipong et al., 2009). These reasons support this type of study, whereby exploration of microbial diversity can shed further light on not only the likely metabolic activities of the microbial populations present, but of their potential in facilitating further degradation of this quite recalcitrant biomass. Additionally, this work shows that bacterial populations inhabiting two different eco-geographically ecosystems are related phylogenetically.

ACKNOWLEDGMENTS

This work was funded by grant CB-153789-Q CONACyT and "Proyecto Biomasa" (Acuerdo 13/2007) of UAM-Cuajimalpa. RB-G and R-C received a scholarship from CONACyT. Additionally RB-G received an EMBO fellowship: ASTF503-2013 to perform part of this work at UCC, in Ireland. We are grateful to Dr.

Champreda (National Center for Genetic Engineering and Biotechnology, Thailand) who shared with us his dataset derived from the metagenome analysis from bagasse and Jamie Fitzgerald for his help with the network visualization.

REFERENCES

Amorim J.H., Macena T.N.S., Lacerda G. V, Rezende R.P., Dias J.C.T., Brendel M., & Cascardo J.C.M. (2008) An improved extraction protocol for metagenomic DNA from a soil of the Brazilian Atlantic Rainforest. *Genetics and molecular research: GMR*, **7**, 1226–1232.

Arantes V. & Milagres A.M.F. (2007) The effect of a catecholate chelator as a redox agent in Fenton-based reactions on degradation of lignin-model substrates and on COD removal from effluent of an ECF kraft pulp mill. *Journal of Hazardous Materials*, **141**, 273–279.

Brady S.F. (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nature Protocols*, **2**, 1297–1305.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., & Madden T.L. (2009) BLAST+: architecture and applications. *BMC bioinformatics*, **10**, .

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.I., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., & Knight R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Chandel A.K., Antunes F.A.F., Anjos V., Bell M.J. V, Rodrigues L.N., Polikarpov I., Azevedo E.R. de, Bernardinelli O.D., Rosa C.A., Pagnocca F.C., & Silva S.S. da (2014) Multi-scale structural and chemical analysis of sugarcane bagasse in the process of sequential acid–base pretreatment and ethanol production by Scheffersomyces shehatae and Saccharomyces cerevisiae. *Biotechnology for Biofuels*, **7**, .

Cole J.R., Wang Q., Fish J.A., Chai B., McGarrell D.M., Sun Y., Brown C.T., Porras-Alfaro A., Kuske C.R., & Tiedje J.M. (2013) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, .

Curson A.R.J., Burns O.J., Voget S., Daniel R., Todd J.D., McInnis K., Wexler M., & Johnston A.W.B. (2014) Screening of metagenomic and genomic libraries reveals three classes of bacterial enzymes that overcome the toxicity of acrylate. *PloS one*, **9**, .

Curson A.R.J., Sullivan M.J., Todd J.D., & Johnston A.W.B. (2010) Identification of genes for dimethyl sulfide production in bacteria in the gut of Atlantic Herring (Clupea harengus). *The ISME journal*, **4**, 144–146.

Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.-F., Guindon S., Lefort V., Lescot M., Claverie J.-M., & Gascuel O. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, **36**, W465–469.

Dominguez-Dominguez M.M., Álvarez-Castillo A., Granados-Baeza M., & Hernández-Campos F. (2012) Estudio d ela cinética de pretratamiento e hidrólisis ácida del bagazo de caña de azúcar. *Revista Iberoamericana de Polímeros*, **13**, 200–211.

Felske A., Engelen B., Nübel U., & Backhaus H. (1996) Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. *Applied and Environmental Microbiology*, **62**, 4162–4167.

Fujii K., Ikeda K., & Yoshida S. (2012) Isolation and characterization of aerobic microorganisms with cellulolytic activity in the gut of endogeic earthworms. *International Microbiology: The Official Journal of the Spanish Society for Microbiology*, **15**, 121–130.

Garrity G.M. & Holt G.J. (2012) *Bergey´s Manual of Systematic Bacteriology.* New York.

Geng W., Huang T., Jin Y., Song J., Chang H.-M., & Jameel H. (2014) Comparison of sodium carbonate-oxygen and sodium hydroxide-oxygen pretreatments on the chemical composition and enzymatic saccharification of wheat straw. *Bioresource technology*, **161**, 63–68.

Ghio S., Lorenzo G.S. Di, Lia V., Talia P., Cataldi A., Grasso D., & Campos E. (2012) Isolation of Paenibacillus sp. and Variovorax sp. strains from decaying woods and characterization of their potential for cellulose deconstruction. *International Journal of Biochemistry and Molecular Biology*, **3**, 352–364.

Goldbeck R., Damasio A.R.L., Goncalves T.A., Machado C.B., Paixao D.A.A., Wolf L.D., Mandelli F., Rocha G.J.M., Ruller R., & Squina F.M. (2014) Development of hemicellulolytic enzyme mixtures for plant biomass deconstruction on target biotechnological applications. *Applied microbiology and biotechnology*, **98**, 8513–8525.

Gruninger R.J., Sensen C.W., McAllister T.A., & Forster R.J. (2014) Diversity of Rumen Bacteria in Canadian Cervids. *PLoS ONE*, **9**, .

Guimarães K.A., Alves L., Sacramento T.M., & Frédéric L. (2012) Application of succinylated sugarcane bagasse as adsorbent to remove methylene blue and gentian violet from aqueous solutions – Kinetic and equilibrium studies. *Dyes and Pigments*, **92**, 967–974.

Huang S., Sheng P., & Zhang H. (2012) Isolation and Identification of Cellulolytic Bacteria from the Gut of Holotrichia parallela Larvae (Coleoptera: Scarabaeidae). *International Journal of Molecular Sciences*, **13**, 2563–2577.

Ionescu D., Siebert C., Polerecky L., Munwes Y.Y., Lott C., Häusler S., Bižić-Ionescu M., Quast C., Peplies J., Glöckner F.O., Ramette A., Rödiger T., Dittmar T., Oren A., Geyer S., Stärk H.-J., Sauter M., Licha T., Laronne J.B., & de Beer D. (2012) Microbial and chemical characterization of underwater fresh water springs in the Dead Sea. *PloS One*, **7**, .

Jackson S.A., Kennedy J., Morrissey J.P., O'Gara F., & Dobson A.D.W. (2012) Pyrosequencing reveals diverse and distinct sponge-specific microbial communities in sponges from a single geographical location in Irish waters. *Microbial Ecology*, **64**, 105–116.

Kanokratana P., Mhuantong W., Laothanachareon T., Tangphatsornruang S., Eurwilaichitr L., Pootanakit K., & Champreda V. (2013) Phylogenetic analysis and metabolic potential of microbial communities in an industrial bagasse collection site. *Microbial Ecology*, **66**, 322–334.

Kanwar N., Scott H.M., Norby B., Loneragan G.H., Vinasco J., Cottell J.L., Chalmers G., Chengappa M.M., Bai J., & Boerlin P. (2014) Impact of treatment strategies on cephalosporin and tetracycline resistance gene quantities in the bovine fecal metagenome. *Scientific reports*, **4**, .

Kiatkittipong W., Wongsuchoto P., & Pavasant P. (2009) Life cycle assessment of bagasse waste management options. *Waste Management (New York, N.Y.)*, **29**, 1628–1633.

Klindworth A., Pruesse E., Schweer T., Peplies J., Quast C., Horn M., & Glöckner F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, **41**, .

Kvitko B.H., Goodyear A., Propst K.L., Dow S.W., & Schweizer H.P. (2012) Burkholderia pseudomallei known siderophores and hemin uptake are dispensable for lethal murine melioidosis. *PLoS neglected tropical diseases*, **6**, .

Lefebre M.D., Flannagan R.S., & Valvano M.A. (2005) A minor catalase/peroxidase from Burkholderia cenocepacia is required for normal aconitase activity. *Microbiology (Reading, England)*, **151**, 1975–1985.

Leis B., Angelov A., & Liebl W. (2013) Screening and expression of genes from metagenomes. *Advances in applied microbiology*, **83**, 1–68.

Li W. & Godzik A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, **22**, 1658–1659.

Lima M.A., Gomez L.D., Steele-King C.G., Simister R., Bernardinelli O.D., Carvalho M.A., Rezende C.A., Labate C.A., Deazevedo E.R., McQueen-Mason S.J., & Polikarpov I. (2014) Evaluating the composition and processing potential of novel sources of Brazilian biomass for sustainable biorenewables production. *Biotechnology for biofuels*, **7**, .

De Los Santos-Villalobos S., Barrera-Galicia G.C., Miranda-Salcedo M.A., & Peña-Cabriales J.J. (2012) Burkholderia cepacia XXVI siderophore with biocontrol capacity against Colletotrichum gloeosporioides. *World Journal of Microbiology & Biotechnology*, **28**, 2615–2623.

Lynch S.C., Eskalen A., Zambino P.J., Mayorquin J.S., & Wang D.H. (2013) Identification and pathogenicity of Botryosphaeriaceae species associated with coast live oak (Quercus agrifolia) decline in southern California. *Mycologia*, **105**, 125–140.

Männistö M.K., Rawat S., Starovoytov V., & Häggblom M.M. (2012) Granulicella arctica sp. nov., Granulicella mallensis sp. nov., Granulicella tundricola sp. nov. and Granulicella sapmiensis sp. nov., novel acidobacteria from tundra soil. *International Journal of Systematic and Evolutionary Microbiology*, **62**, 2097–2106.

Mathew A., Eberl L., & Carlier A.L. (2014) A novel siderophore-independent strategy of iron uptake in the genus Burkholderia. *Molecular Microbiology*, **91**, 805–820.

Milagres A., Arantes V., Medeiros c., & A.Machuca (2002) Production of metal chelating compounds by white and brown-rot fungi and their comparative abilities for pulp bleaching. *Enxyme and Microbial Technology*, **30**, 562–565.

Miller G.L. (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, **31**, 426–428.

Nawrocki E.P. & Eddy S.R. (2007) Query-dependent banding (QDB) for faster RNA similarity searches. *PLoS computational biology*, **3**, .

Ondov B.D., Bergman N.H., & Phillippy A.M. (2011) Interactive metagenomic visualization in a Web browser. *BMC bioinformatics*, **12**, .

Patel D.D., Patel A.K., Parmar N.R., Shah T.M., Patel J.B., Pandya P.R., & Joshi C.G. (2014) Microbial and Carbohydrate Active Enzyme profile of buffalo rumen metagenome and their alteration in response to variation in the diet. *Gene*, **545**, 88–94.

Pope P.B., Mackenzie A.K., Gregor I., Smith W., Sundset M.A., McHardy A.C., Morrison M., & Eijsink V.G.H. (2012) Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PloS One*, **7**, .

Pruesse E., Peplies J., & Glöckner F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics (Oxford, England)*, **28**, 1823–1829.

Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., & Glöckner F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, **41**, D590–596.

Rasheed Z., Rangwala H., & Barbará D. (2013) 16S rRNA metagenome clustering and diversity estimation using locality sensitive hashing. *BMC systems biology*, **7 Suppl 4**, .

Rattanachomsri U., Kanokratana P., Eurwilaichitr L., Igarashi Y., & Champreda V. (2011) Culture-independent phylogenetic analysis of the microbial community in industrial sugarcane bagasse feedstock piles. *Bioscience, Biotechnology, and Biochemistry*, **75**, 232–239.

Rezende C.A., de Lima M.A., Maziero P., Deazevedo E.R., Garcia W., & Polikarpov I. (2011) Chemical and morphological characterization of sugarcane bagasse submitted to a delignification process for enhanced enzymatic digestibility. *Biotechnology for biofuels*, **4**, .

Shirvani M. & Nourbakhsh F. (2010) Desferrioxamine-B adsorption to and iron dissolution from palygorskite and sepiolite. *Applied Clay Science*, **48**, 393–397.

Silva C.F., Azevedo R.S., Braga C., da Silva R., Dias E.S., & Schwan R.F. (2009) Microbial diversity in a bagasse-based compost prepared for the production of Agaricus brasiliensis. *Brazilian Journal of Microbiology: [publication of the Brazilian Society for Microbiology]*, **40**, 590–600.

Simon C. & Daniel R. (2009) Achievements and new knowledge unraveled by metagenomic approaches. *Applied microbiology and biotechnology*, **85**, 265–276.

Simon C. & Daniel R. (2011) Metagenomic analyses: past and future trends. *Applied and environmental microbiology*, **77**, 1153–1161.

Siqueira J.F.J. & Rocas I.N. (2013) Microbiology and treatment of acute apical abscesses. *Clinical microbiology reviews*, **26**, 255–273.

Song N., Cai H.-Y., Yan Z.-S., & Jiang H.-L. (2013) Cellulose degradation by one mesophilic strain Caulobacter sp. FMC1 under both aerobic and anaerobic conditions. *Bioresource Technology*, **131**, 281–287.

Talia P., Sede S.M., Campos E., Rorig M., Principi D., Tosto D., Hopp H.E., Grasso D., & Cataldi A. (2012) Biodiversity characterization of cellulolytic bacteria present on native Chaco soil by comparison of ribosomal RNA genes. *Research in Microbiology*, **163**, 221–232.

Thomas M.S. (2007) Iron acquisition mechanisms of the Burkholderia cepacia complex. *Biometals: An International Journal on the Role of Metal Ions in Biology, Biochemistry, and Medicine*, **20**, 431–452.

Vandamme P., Opelt K., Knöchel N., Berg C., Schönmann S., De Brandt E., Eberl L., Falsen E., & Berg G. (2007) Burkholderia bryophila sp. nov. and Burkholderia megapolitana sp. nov., moss-associated species with antifungal and plant-growth-promoting properties. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 2228–2235.

Vorob'ev A. V, de Boer W., Folman L.B., Bodelier P.L.E., Doronina N. V, Suzina N.E., Trotsenko Y.A., & Dedysh S.N. (2009) Methylovirgula ligni gen. nov., sp. nov., an obligately acidophilic, facultatively methylotrophic bacterium with a highly divergent mxaF gene. *International Journal of Systematic and Evolutionary Microbiology*, **59**, 2538–2545.

Wadell E. & Bang S. (2008) Exploring the phylogenetic diversity of the homestake gold mine, lead, SD. *Journal of Industrial Microbiology and Biotechnology*, **36**, 585–598.

Wang Q., Garrity G.M., Tiedje J.M., & Cole J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.

Ward N.L., Challacombe J.F., Janssen P.H., Henrissat B., Coutinho P.M., Wu M., Xie G., Haft D.H., Sait M., Badger J., Barabote R.D., Bradley B., Brettin T.S., Brinkac L.M., Bruce D., Creasy T., Daugherty S.C.,

Davidsen T.M., DeBoy R.T., Detter J.C., Dodson R.J., Durkin A.S., Ganapathy A., Gwinn-Giglio M., Han C.S., Khouri H., Kiss H., Kothari S.P., Madupu R., Nelson K.E., Nelson W.C., Paulsen I., Penn K., Ren Q., Rosovitz M.J., Selengut J.D., Shrivastava S., Sullivan S.A., Tapia R., Thompson L.S., Watkins K.L., Yang Q., Yu C., Zafar N., Zhou L., & Kuske C.R. (2009) Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Applied and Environmental Microbiology*, **75**, 2046–2056.

Xu G. & Goodell B. (2001) Mechanisms of wood degradation by brown-rot fungi: chelator-mediated cellulose degradation and binding of iron by cellulose. *Journal of Biotechnology*, **87**, 43–57.

Xu Z., Hansen M.A., Hansen L.H., Jacquiod S., & Sørensen S.J. (2014) Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PloS One*, **9**.

Yun J., Kang S., Park S., Yoon H., Kim M.-J., Heu S., & Ryu S. (2004) Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Applied and Environmental Microbiology*, **70**, 7229–7235.

BIOSKETCHES

Batista-García is interested to study the microbial composition of lignocellulosic rich eco-habitats.

The Research team is interested to conduct metagenomic studies, both, structural and functional.

Author contribution: R.A.B.G and J.L.F.M conceived the ideas. A.A.C, R.C and R.A.B.G collected the data. All authors analysed the data. R.A.B.G, S.A.J, A.D.W.D and J.L.F.M wrote the manuscript.

560 **Tables**

561 **Table 1.** Analysis of 16S rRNA sequences from sugarcane bagasse. Chao1 species richness and Shannon

562 diversity indexes were calculated at different genetic distance levels.

563

Genetic distance	Richness	Chao1 richness	Shannon index
0.02	42	61.37	2.88
0.03	35	50.00	2.70
0.05	17	36.00	2.47
0.15	10	10.00	0.77
0.3	3	3.00	0.61

564

565

566

567

568

569

Figure Legends

Fig. 1 Sugarcane bagasse sampling. (a) Location of the collection site at Zacatepec municipality (Morelos State), Mexico. Cartesian coordinates were estimated by measuring satellite instruments (GPS: global positioning system, Model eTrex 20). (b) General scheme of sampling for one pile. Three samples in triplicates (R1, R2 and R3) of sugarcane bagasse were taken at different levels (surface, core and bottom) of the column of each sampled pile. A compost sample was obtained pooling R1, R2 and R3 from each sampling level (CSS: surface compost sample, CSC: core compost sample and CSB: bottom compost sample). A final compost sample (CSF) for further experiments was obtained pooling CSS, CSC and CSB.

Fig. 2 Taxonomic fingerprint. (a) Phylum fingerprint. (b) Full-taxonomic fingerprint. Fingerprints were derived from SILVA pipeline analysis.

Fig. 3 Relative taxonomic representation. (a) Classified genera in the phylum *Acidobacteria* and its relative abundance. (b) Relative abundance of *Proteobacterial* classes. (c, d, e and f) Classified genera and relative abundance of *Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria* and *Gammaproteobacteria* respectively. All values are shown according to the total sequences analysed. This classification was obtained from SILVA.

Fig. 4 Sample-based rarefaction curves of 16S rDNA sequences amplified from sugarcane bagasse metagenome. Rarefaction curves were calculated in RDP pipeline. Assignment of OTUs at different genetic distance levels were obtained: 0.01 (strain), 0.03 (species), 0.05 (genus), 0.15 (class) and 0.28 (phylum).

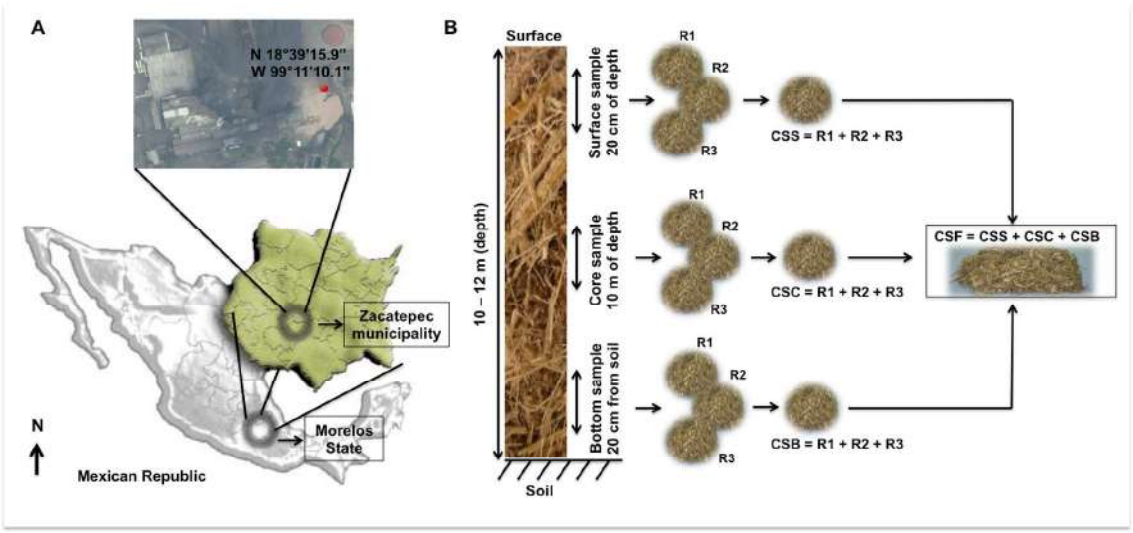
Fig. 5 Culture-independent taxonomic analysis reveals significant differences in the microbial composition derived from cellulolytic (box enclosed with solid lines) and non-cellulolytic (box enclosed with dashed lines) environmental metagenomic sequence sets.

Fig. 6 OTU network map showing OTU interaction between a sugarcane bagasse of this study (red network) and Thai sugarcane bagasse (green network) (Rattanachomsri et al., 2011).

Fig. 7 Radial phylogenetic reconstruction for the phylum *Acidobacteria*. The relationship between 16s rDNA gene dataset derived from sugarcane bagasse (only reads classified in *Acidobacteria* phylum) and 19 sequences of 16S rDNA belonging at 19 species of the class are showed.

Fig. 8 Radial phylogenetic reconstruction for the phylum *Proteobacteria*. The relationship between 16S rDNA gene dataset derived from sugarcane bagasse (only reads classified in the phylum *Proteobacteria*) and 43 sequences of the 16S rDNA belonging to 43 type species and/or genera are shown.

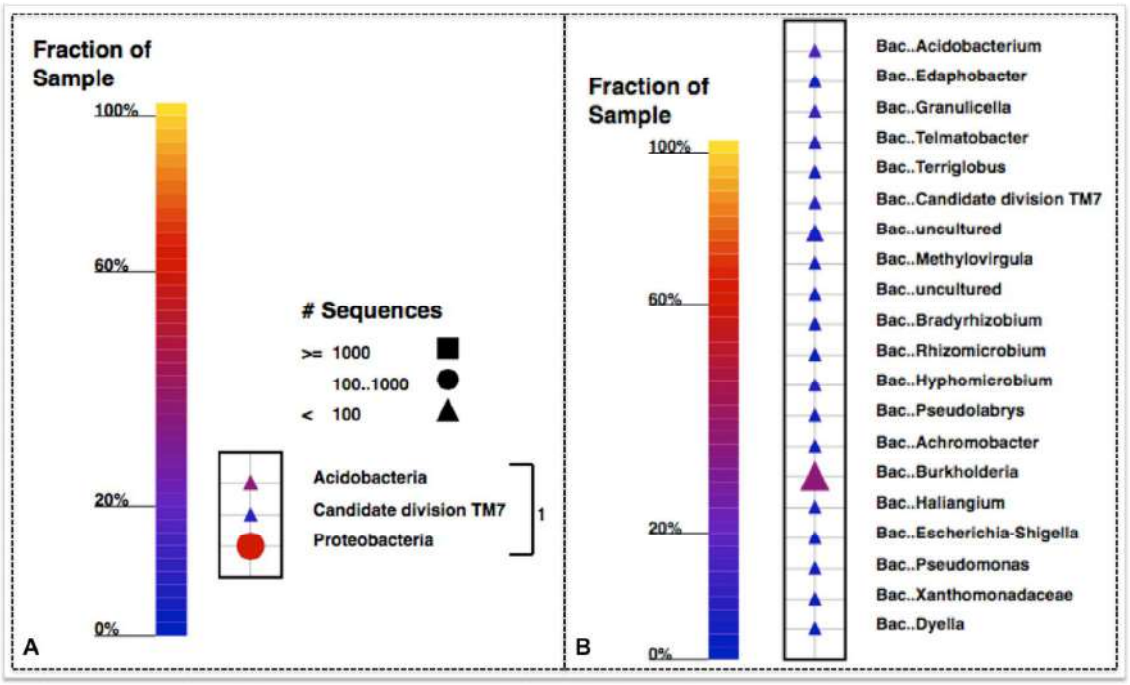
600 Fig. 1



601

602

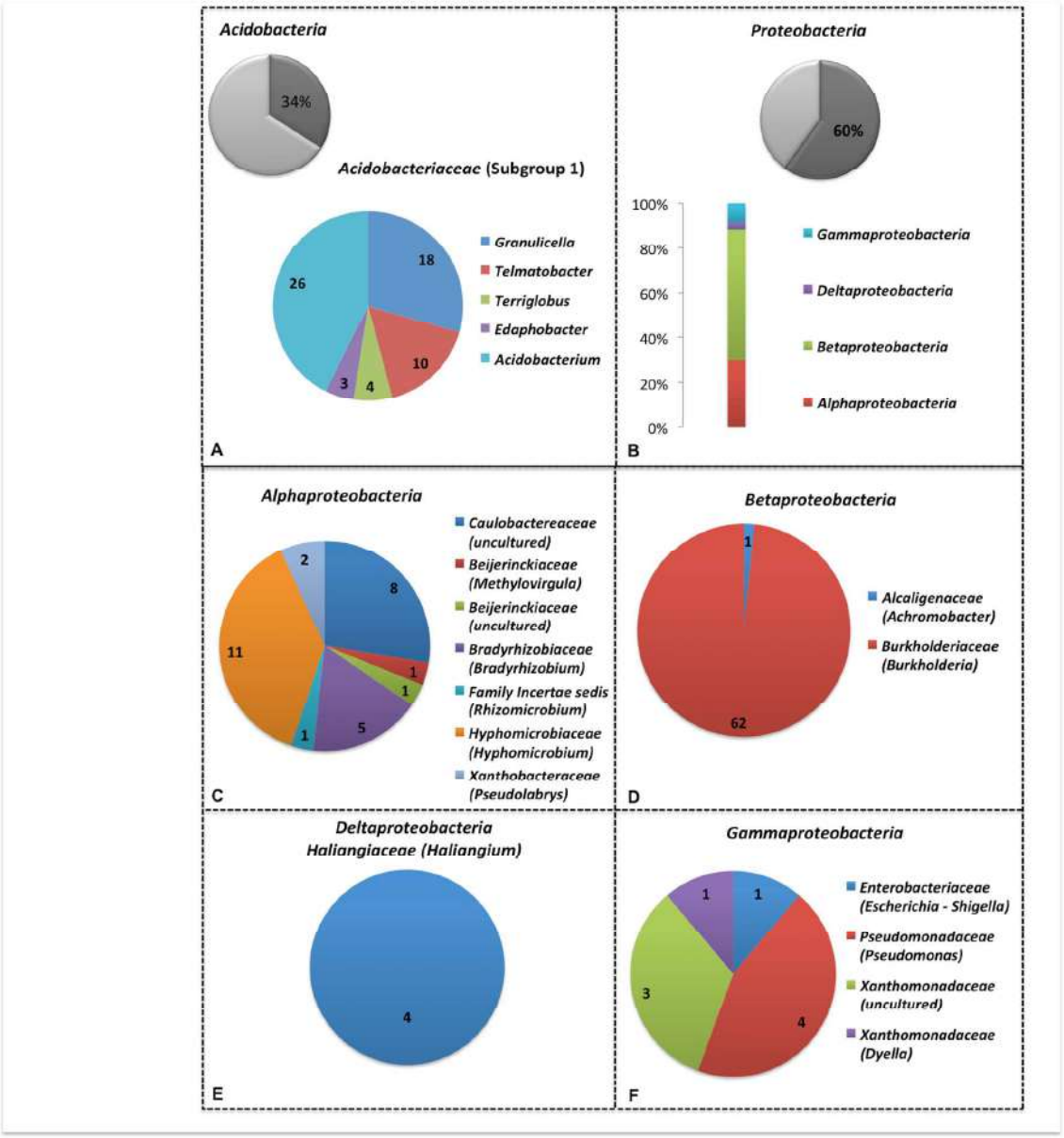
603 Fig. 2



604

605

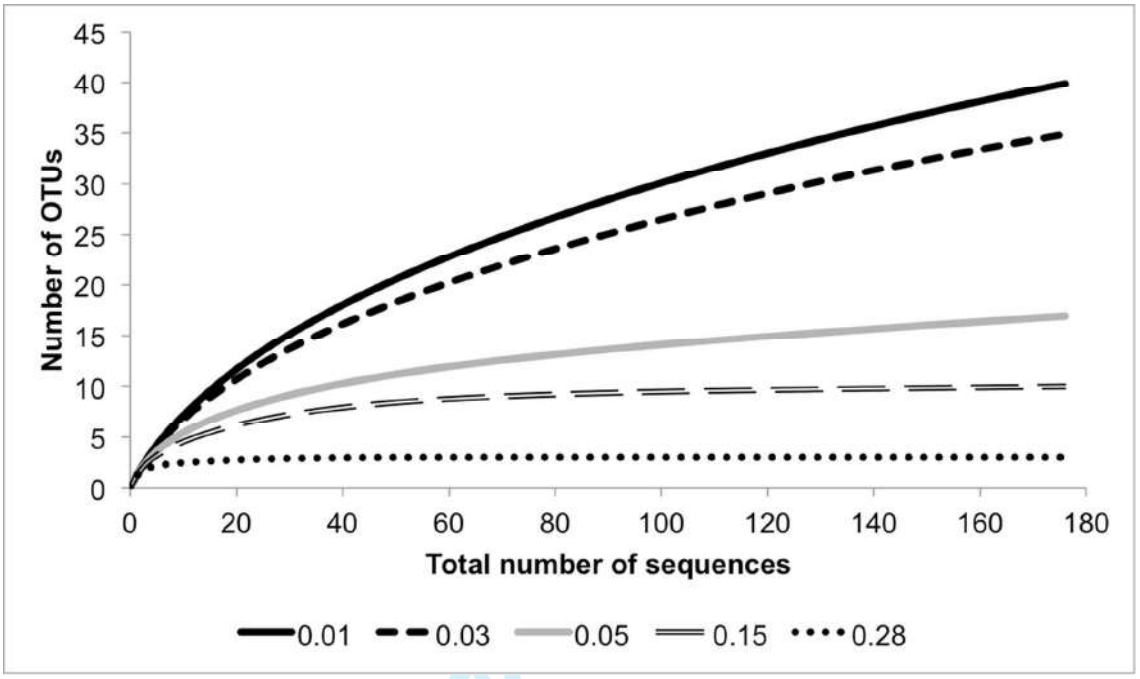
606 Fig. 3



607

608

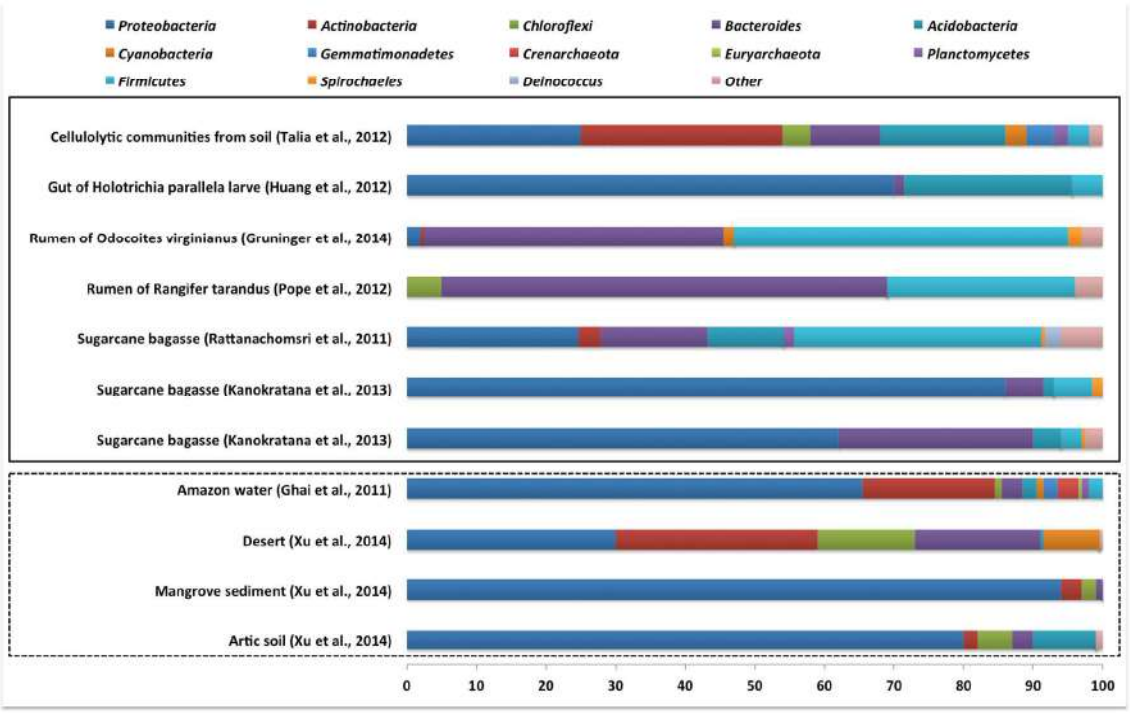
609 Fig. 4



610

611

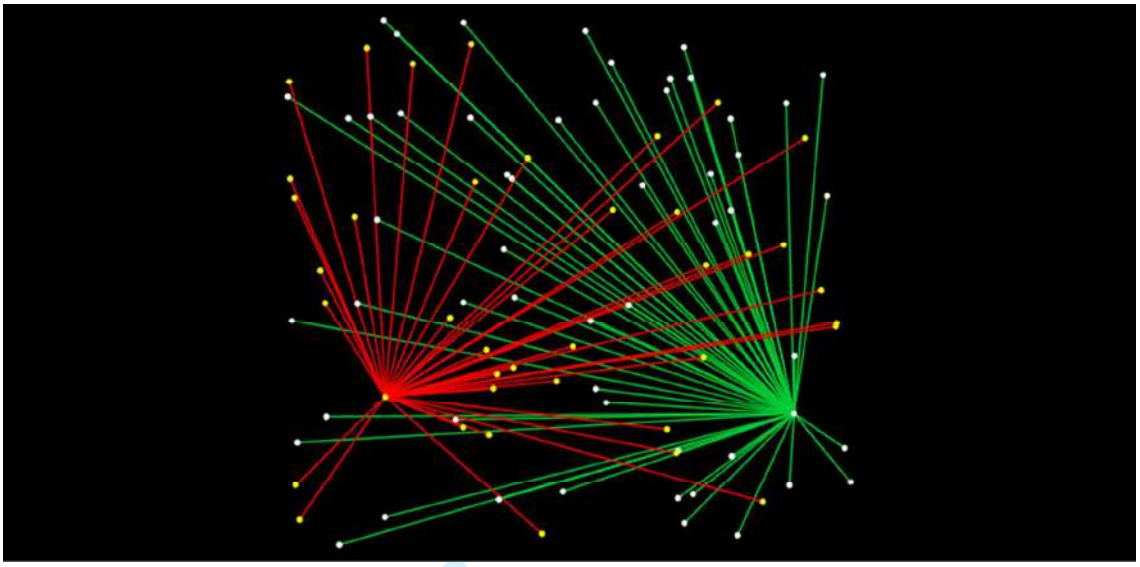
612 Fig. 5



613

614

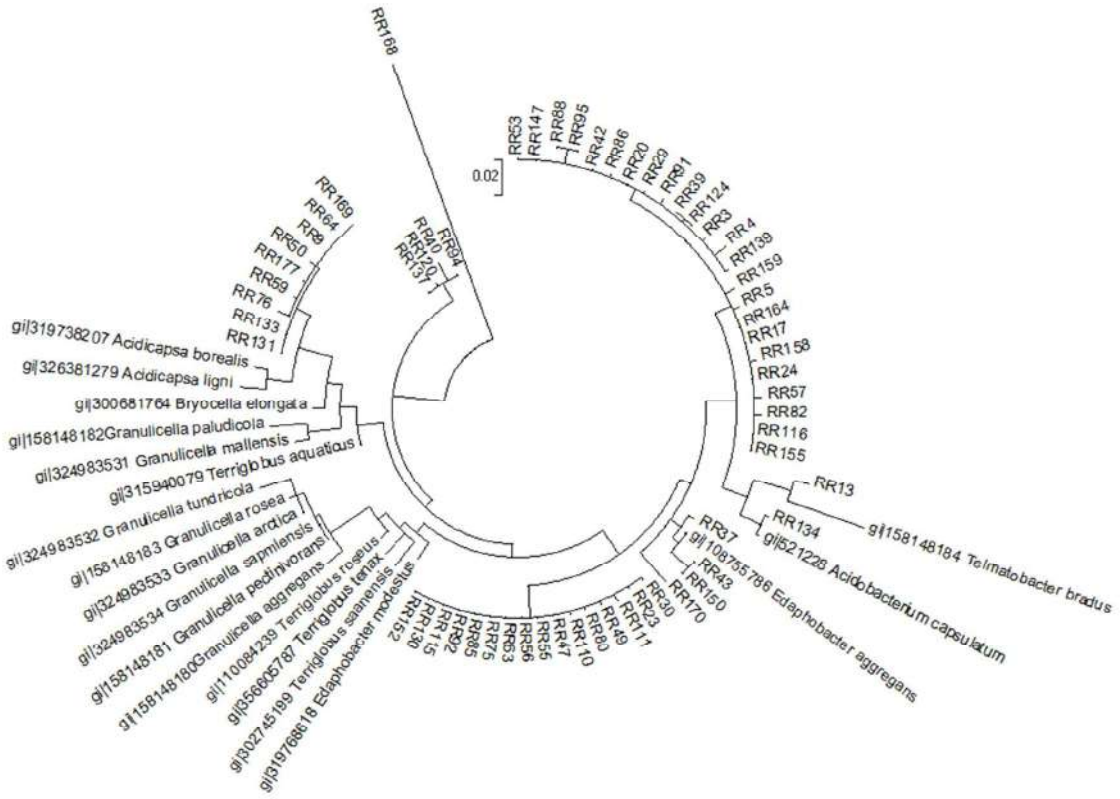
615 Fig. 6



616

617

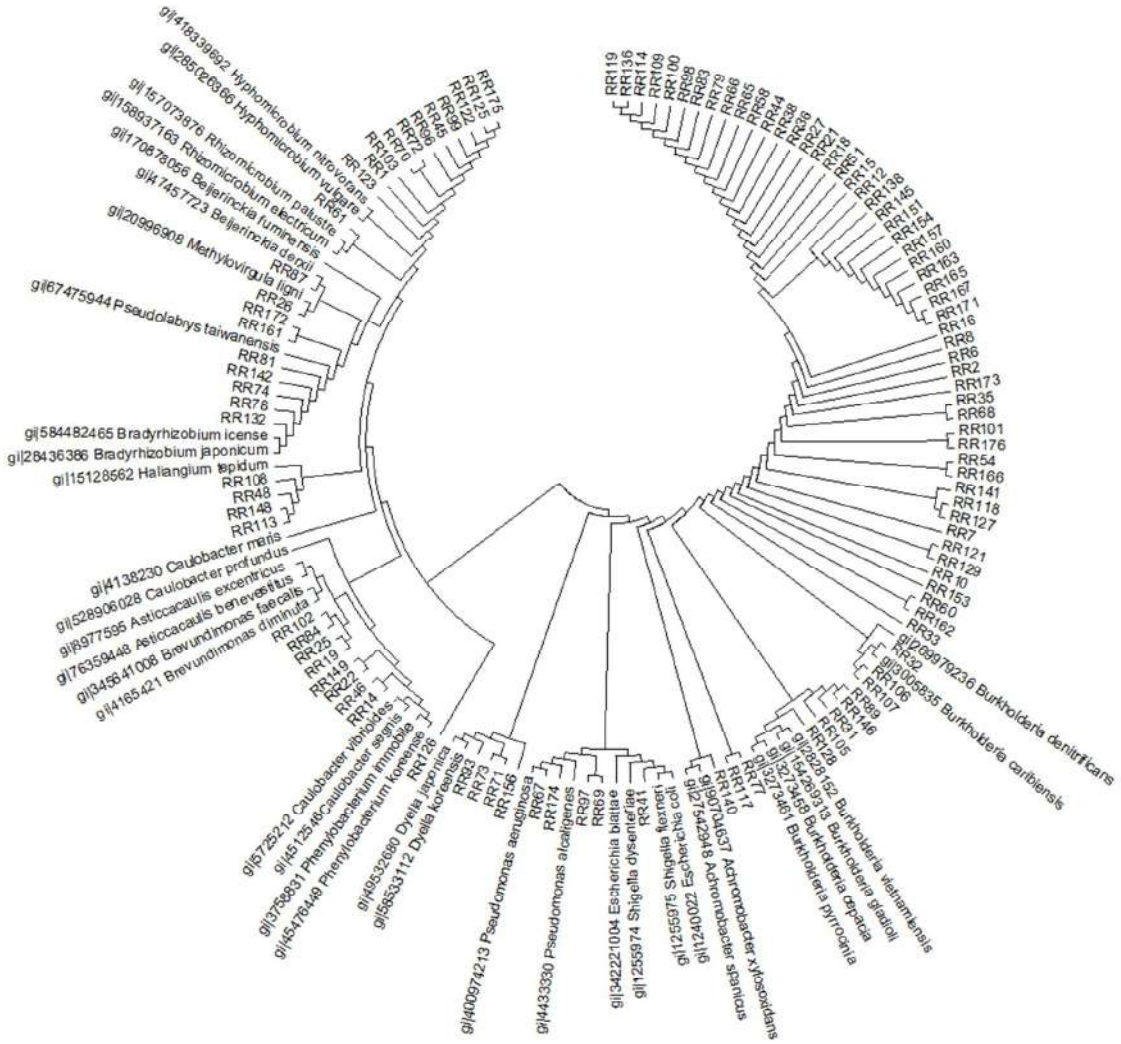
618 Fig. 7



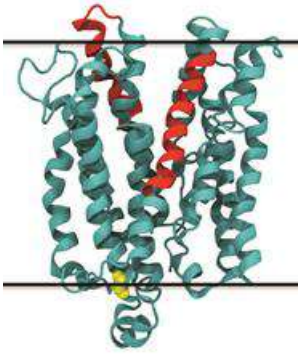
619

620

621 Fig. 8


gi|418339892 Hyphomicrobium nitrovorans
gi|285026366 Hyphomicrobium vulgare
gi|157073876 Rhizomicrobium palustre
gi|158937163 Rhizomicrobium electricum
gi|170878056 Beijerinckia fluminensis
gi|47457723 Beijerinckia derxii
gi|20996909 Methylovirgula ligni
gi|67475944 Pseudolabrys taiwanensis
gi|584482465 Bradyrhizobium icense
gi|28436386 Bradyrhizobium japonicum
gi|15128562 Halianglium tepidum
gi|4138230 Caulobacter meris
gi|5289060 28 Caulobacter profundus
gi|8977595 Asticcacaulis excentricus
gi|76359448 Asticcacaulis benevestitus
gi|34564100 8 Brevundimonas faecalis
gi|4165421 Brevundimonas diminuta
gi|57252 12 Caulobacter vibrioides
gi|45125 46 Caulobacter segnis
gi|37588 31 Phenylobacterium immobile
gi|45478449 Phenylobacterium koreense
gi|49532680 Dyella japonica
gi|56533112 Dyella koreensis
gi|400974213 Pseudomonas aeruginosa
gi|4433330 Pseudomonas alcaligenes
gi|342221004 Escherichia blattae
gi|1255974 Shigella dysenteriae
gi|12559 75 Shigella flexneri
gi|12400 22 Escherichia coli
gi|12754 2948 Achromobacter xylosoxidans
gi|5907 04637 Achromobacter silanicus
gi|3273461 Burkholderia pyrrocinia
gi|154269313 Burkholderia cepacia
gi|6281 52 Burkholderia vietnamiensis
gi|3005835 Burkholderia caribiensis
gi|209979236 Burkholderia denitrificans

622



Capítulo III.

TctA: un transportador que salva la frustración científica.

Frustraciones académicas: ¿brechas para publicar?

Resulta prácticamente imposible narrar una tesis doctoral sin recordar las frustraciones. Durante el ejercicio de un doctorado muchas son las veces que los experimentos se resisten a dejar conocer lo que puede ser llamado como *"la caja negra de la vida"*. Sin embargo, considero que también hemos de generar adaptaciones a estos eventos, que sin dudas sacuden la versatilidad de nuestra fisiología y respuestas para mantener la homeostasis.

Así fue la historia de TctA… Una vez que construímos nuestras librerías con metagenomas aislados de bagazo de caña de azúcar y lodos activados, comenzamos las prospecciones funcionales de enzimas lignocelulolíticas. Nuestras librerías fueron analizadas para la detección de actividades celulasas y xilanasas mediante escrutinios basados en actividad usando celulosa, xilano y goma arábica. La goma arábica también permitió analizar otras actividades como esterasas. Luego de la secuenciación de una clona crecida sobre goma arábica como única fuente de carbono, detectamos que el probable gen que permitía su crecimiento estaba incompleto. Sin embargo, pudimos detectar hacía el extremo 3´ del fragmento un marco abierto de lectura con homología de secuencia a transportadores de ácidos tricarboxílicos. Aunque la frustración fue inminente, decidimos estudiar la secuencia anterior, lo cual nos permitió la publicación del primer artículo del periodo doctoral: *"A novel TctA citrate transporter from activated sludge metagenome: Structural and mechanistic predictions for the TTT family"*. Este trabajo lleva el refrán de *"buscando peras se encuentran manzanas"*, pero contribuyó con la predicción del mecanismo de acción de estos transportadores.

Como vemos los escrutinios basados en actividad pueden resultar difíciles y sin éxito, más cuando no contamos con herramientas de robótica automatizada para su ejecución. Por otro lado, muchos son los factores que ponderan negativamente las tasas de éxito de este tipo de escrutinio (Ver Capítulo I).

TctA fue la secuencia que permitió incorporar a nuestro proyecto doctoral aspectos de modelación y dinámicas moleculares, docking de proteína y análisis *in silico* de estructuras.

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A novel TctA citrate transporter from an activated sludge metagenome: Structural and mechanistic predictions for the TTT family

Ramón Alberto Batista-García,[1] Ayixon Sánchez-Reyes,[1] César Millán-Pacheco,[1] Víctor Manuel González-Zuñiga,[2] Soledad Juárez,[2] Jorge Luis Folch-Mallol,[3] and Nina Pastor[1]*

[1] Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Cuernavaca, 62209 Morelos, México

[2] Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, 62210 Morelos, México

[3] Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos, Cuernavaca, 62209 Morelos, México

ABSTRACT

We isolated a putative citrate transporter of the tripartite tricarboxylate transporter (TTT) class from a metagenomic library of activated sludge from a sewage treatment plant. The transporter, dubbed TctA_ar, shares ~50% sequence identity with TctA of *Comamonas testosteroni* (TctA_ct) and other β-*Proteobacteria*, and contains two 20-amino acid repeat signature sequences, considered a hallmark of this particular transporter class. The structures for both TctA_ar and TctA_ct were modeled with I-TASSER and two possible structures for this transporter family were proposed. Docking assays with citrate resulted in the corresponding sets of proposed critical residues for function. These models suggest functions for the 20-amino acid repeats in the context of the two different architectures. This constitutes the first attempt at structure modeling of the TTT family, to the best of our knowledge, and could aid functional understanding of this little-studied family.

Proteins 2014; 00:000–000.

Key words: metagenomic; TctA; molecular modeling; docking; gene phylogeny; citrate.

INTRODUCTION

There are at least 10^6–10^8 species of microorganisms in the world,[1] and approximately 99% of the microorganisms that inhabit ecosystems cannot be cultivated in the laboratory.[2,3] Metagenomic tools provide a new avenue to study these microbial populations and their genes, in the prospection for novel functions. Screening methods are critical in the search, isolation, and characterization of new proteins from metagenomics constructions. Among the most used currently are function-based and sequence-based screenings.[4]

Looking for a novel esterase that degrades organophosphates in a metagenomic library derived from activated sludge, we isolated a clone with three complete open reading frames (ORFs). Two of these code for a putative thioesterase, most likely responsible for the parathion degradation, and a dihydratase, which will be described elsewhere. The remaining ORF codes for a putative citrate transporter of the tripartite tricarboxylate transporter (TTT) class.[5] Given that these transporters can be recruited to transport aromatic substrates,[6] these three ORFs reside in the same clone, and they may be functionally coupled, we analyzed and modeled the putative transporter.

Additional Supporting Information may be found in the online version of this article.

Abbreviations: TctA_ar, novel TctA protein from activated sludge metagenome; TctA_ct, TctA protein from *Comamonas testosteroni* E6; TTT, tripartite tricarboxylate transporter

Grant sponsor: CONACyT; Grant number: CB-153789-Q.

Ramón Alberto Batista-García and Ayixon Sánchez-Reyes have contributed equally to this work.

*Correspondence to: N. Pastor, Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Col. Chamilpa, Cuernavaca, 62209 Morelos, México. E-mail: nina@uaem.mx

Received 9 October 2013; Revised 14 January 2014; Accepted 28 January 2014
Published online 3 February 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24529

The TctABC transporter system includes three proteins: TctA, TctB, and TctC.[7] All these proteins belong to the TTT family (TC 2.A.80).[5] TctB and TctC are a small protein with four transmembrane helices and a periplasmic tricarboxylate-binding receptor, respectively, and do not seem to be essential for function. TctAs are large proteins, with 12 putative transmembrane α-helical segments. They have 500 amino acids on average, ranging between 416 and 637 amino acids in length. A short repeated sequence, $(Hy)_6$-G-$(Hy)_3$-G*-$(Hy)_3$-G*-$(Hy)_2$-P*G*-Hy, where Hy is a hydrophobic residue and * is a highly conserved residue, was presented[5] as evidence of an ancient duplication event.

Crystallized structures of this family of proteins do not exist to date, and functional studies have only been carried out for TctA proteins of the genera *Salmonella*,[7,8] *Comamonas*,[9] *Corynebacterium*,[10,11] and *Bordetella*.[12] The proposed transport mechanism is the symport of citrate with sodium, although divalent cations also play an important role in binding citrate to TctC,[10] eventually bringing citrate to TctA. To the best of our knowledge, the only reported mutation for this family was isolated in ciprofloxacin-stressed *Salmonella*; it involves the mutation of G109 to serine and eliminates citrate transport.[8]

Citrate transport can be achieved with a large variety of secondary transporters, both symporters (reviewed in Ref. 10) and antiporters (reviewed in Ref. 13), as can be seen in the Transporter Classification Database (TCDB).[14] Currently, there are crystal structures for various members of the MFS, IT, APC, and MOP superfamilies. Some superfamilies, such as MFS, can harbor antiporters and symporters with similar transporter architectures (reviewed in Ref. 15), and similar architectures can be accommodated by sequences with very low similarity, as reported recently in the case of two citrate transporters from TC classes 2.A.11 and 2.A.47.[16] TctA proteins display less than 15% identity with any of the crystallized transporters deposited in the Protein Data Bank (PDB).[17] Therefore, any exercise in structure modeling must consist of threading the sequence of the protein over plausible structural templates. We propose two possible structures for this novel protein, and a list of critical residues for function, that can be assayed by site-directed mutagenesis in future studies.

MATERIALS AND METHODS

Library construction

DNA extraction from duplicate samples of activated sludge, collected from an industrial and urban sewage treatment plant of Cuernavaca, Morelos, México, was performed using the UltraClean®Microbial DNA Isolation Kit, MO BIO Laboratories. DNA was characterized using agarose gel electrophoresis and nanodrop readings. We amplified fragments of the genes coding for the 16S and fungal 18S RNA with PCR.[18,19]

We fragmented the isolated DNA mechanically by nebulization with nitrogen. The result was confirmed using agarose gel electrophoresis and quantitation in nanodrop. Nebulized DNA was repaired with T4 DNA polymerase to generate blunt ends, using standard procedures. DNA was ligated to vector pJET1.2 and subsequently transformed by electroporation of *E. coli* DH5α. The library was amplified and preserved in glycerol (20%) at −70°C.

Selection and characterization of clones

The functional scrutiny consisted in testing for esterase activity.[20] Plasmid extraction was performed for the positive clones using GeneGet Plasmid Miniprep Kit|Thermo Scientific. The size of the inserts was estimated with a *BglII* digestion, and the inserts were sequenced.

Bioinformatic analysis

We used the SnapGene Viewer 2.1 program and the site of NCBI to predict ORFs. Protein–protein BLASTs were carried out to ascertain sequence homology with the nonredundant protein set at the NCBI site.

Sequence alignment between functionally studied TctA proteins and our clone was performed with Clustal Omega Multialign Software and ESPript 2.2 servers. The binary comparison scores are expressed as percent identity.

Phylogenetic analysis was performed online with the server Phylogeny.fr. BLAST searches on the nonredundant protein collection at the National Center for Biotechnology Information (NCBI) were done, using as queries our clone and the TctAs of *Salmonella, Comamonas, Bordetella*, and *Corynebacterium*. A nonredundant set was constructed from the highest 100 hits for each search, and the TctAs reported in Ref. 5, complemented with five TctA homologs of archaebacteria.

Molecular modeling

The predicted amino acid sequence for the putative TctA protein (TctA_ar hereon), and TctA from *Comamonas testosteroni* E6 (TctA_ct hereon), were submitted to the I-TASSER server[21] without constraints. Selected templates that were not transporters were forbidden in subsequent modeling attempts. We generated models with the same template for both TctA_ar and TctA_ct.

Each model was inserted into a dimyristoylphosphatidylcholine (DMPC) bilayer with 25% cholesterol and solvated with water and 0.15*M* KCl, using the CHARMM-GUI server. The systems were structurally relaxed by energy minimizations for 10,000 conjugate gradient steps and a short molecular dynamics run (25,000 steps at 30°C), using the CHARMM36 force field in NAMD.[22] This procedure eliminates steric clashes and regularizes

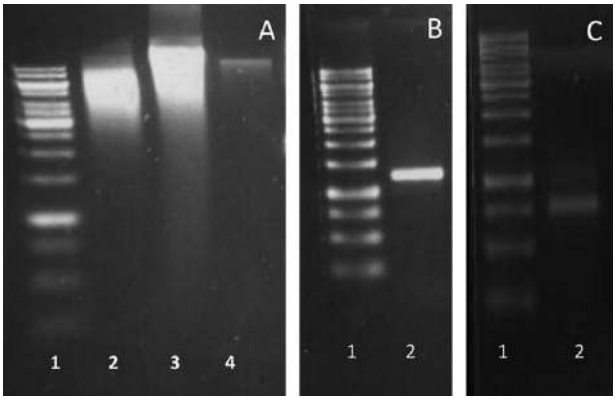


Figure 1

(A) Metagenomic DNA isolation from activated sludge samples. Lane 1: 1 kb ladder, lanes 2 and 3: activated sludge metagenomic DNA, lane 4: soil metagenomic DNA. (B) PCR to amplify approximately 1300 bp of the gene coding for 16S RNA. Lane 1: 1 kb ladder, lane 2: band corresponding to gene fragments coding for 16S RNA. (C) PCR to amplify approximately 750 bp of the gene coding for the 18S RNA. Lane 1: 1 kb ladder, lane 2: band corresponding to gene fragments coding for 18S RNA.

the stereochemistry of the models, but does not alter the trace of the main chain of the protein.

We docked citrate to the relaxed models using AUTODOCK/VINA[23] as a functional criterion. We used the default AUTODOCK/VINA parameters, allowing for flexibility in citrate but not in the proteins. Structural alignments between the model structure of TctA_ar and TctA_ct and the closest structural homolog detected by I-TASSER were used to identify putative relevant functional residues.

RESULTS AND DISCUSSION

Characterization of metagenomic DNA library

We isolated 10 μg of DNA from the sample on average. Because the kit we used is recommended for extraction of soil metagenomic DNA, we also isolated DNA from a natural soil sample, as a positive control. For the control sample, we obtained a defined band of high molecular weight (Fig. 1A).

We ascertained DNA concentration and quality by reading absorbance of the sludge and soil samples at 260 and 280 nm in nanodrop. For the sludge the ratio (A260/280) was 1.8, and for the soil sample the ratio (A260/280) was 2.01, confirming that DNA purity is suitable for metagenomic construction. As another quality criterion, we performed PCR to amplify fragments of the genes coding for 16S and fungal 18S RNA [Fig. 1(B,C)]. These PCRs served to confirm the presence of genomic material of bacteria and fungi in our samples. It is beyond the interest of this work to describe the structure of the microbial communities present in the activated sludge sample; hence, the PCRs are just taken as qualitative criteria of the purity of the metagenomic DNA, and presence of bacterial and fungal organisms in the activated sludge samples.

For the construction of the metagenomic DNA library, fragments between 6 and 10 kb are the target, and were achieved with nebulization at 8 psi for 5 s. These fragments were ligated in pJET1.2/blunt and transformed in *E. coli* DH5α, to obtain a metagenomic library of 20,000 clones. The title of DNA was approximately 90 mb, and we estimated that 93% of the clones have an insert with an average size of approximately 8.5 kb.

Bioinformatic analysis of the insert

A clone was identified positive for methyl parathion degradation. Restriction analysis with *BglII* defined the presence of an insert of about 6 kb. We found a complete ORF (1311 bp) that encodes a protein of 45.1 kDa theoretical molecular weight (463 amino acids) with homology to TctA proteins, part of the TTT family; the gene was named *tctA_ar*. TctA consist of between 416 and 637 amino acids,[5] which brackets the size of TctA_ar. The sequence determined in this study was deposited in the NCBI database under Accession No. KF570143.

To produce an educated guess regarding the origin of our cloned transporter, we selected a set of 119 TctA sequences that include those described in Ref. 5, and five archaeal TctAs (see Methods). With this set, we built a phylogenetic tree using Phylogeny.fr software (Supporting Information Fig. S1). It clearly shows that TctA_ar is not archaeal, and probably originated in a β-*Proteobacteria*, given its nearest neighbors in the tree. Given the possibility of horizontal transfer amongst bacteria, yeast, and plants, it is currently impossible to determine the host of the protein, and should definitely be considered a clone from a soil sample. This search selected a TctA homolog in a plant (gi 224156094, from *Populus trichocarpa*), the first report of a eukaryotic TctA.

The amino acid sequence of TctA_ar exhibited ~50% identity with the TctA protein of *C. testosteroni*, and with other TctA proteins of β-*Proteobacteria*, as summarized in Table I. Given that the TctA of *Comamonas* has been functionally characterized, we chose it over that of *Verminephrobacter* as a control protein for our modeling efforts.

Conserved regions in the TctA sequences from Table I were identified with Clustal. The alignment in Figure 2 shows high conservation in the first half of the protein, reaching 85% pairwise identity. Two repeated 20-amino acid signatures $((Hy)_6\text{-}G\text{-}(Hy)_3\text{-}G^*\text{-}(Hy)_3\text{-}G^*\text{-}(Hy)_2\text{-}P^*G^*\text{-}Hy)^5$ are located between amino acids L20 and G40 in the N-terminal repeat, and between D251 and

Table I
Pairwise Identity Matrix Between the Closest Neighbor and Functionally Characterized TctAs and TctA_ar

	1	2	3	4	5	6
1. *Verminephrobacter eiseniae* (gi 121610934)	100.00	88.47	73.20	54.13	40.60	37.85
2. *Comamonas testosteroni* (gi 264676989)	88.47	100.00	72.20	51.61	39.00	37.85
3. *Bordetella holmesii* (gi 491160357)	73.20	72.20	100.00	47.58	40.80	39.00
4. TctA_ar (KF570143)	54.13	51.61	47.58	100.00	28.18	25.06
5. *Salmonella enterica* (gi 446304178)	40.60	39.00	40.80	28.18	100.00	35.06
6. *Corynebacterium glutamicum* (gi 489959448)	37.85	37.85	39.00	25.06	35.06	100.00

R271 in the C-terminal repeat (boxed in full lines in Fig. 2; TctA_ar numbering). A highly conserved region, located between amino acids 61 and 111 of these sequences, is also apparent (boxed in broken black lines in Fig. 2). In the absence of directed mutagenesis studies of these transporters, the functional or structural relevance of these sequence signatures is unknown. The only reported mutant for this family, G109S, in *Salmonella enterica*[8] is marked with an asterisk.

Structural modeling and functional inferences

Given the negligible sequence similarity of TctAs to transporter structures deposited in the PDB, we chose I-TASSER[21] as a modeling tool, in view of its success in Critical Assessment of protein Structure Prediction (CASP). To have a measure of the robustness and possible sense of our structural predictions, we generated models for both TctA_ar and TctA_ct. We reasoned that the ~50% sequence identity shared by these two proteins should be reflected in the conservation of structural features.

Both TctA_ar and TctA_ct are predicted to be mostly α helical by I-TASSER, with many long and broken helices, in agreement with published hydropathy plots.[5] In our first attempt at modeling of TctA_ar, I-TASSER selected as templates HEAT repeat proteins. Given the information in Table I and Supporting Information Figure S1 that strongly suggests TctA_ar to be a citrate transporter, we discarded these HEAT repeat templates as biologically irrelevant. The following rounds of modeling denied I-TASSER the possibility of choosing previously identified templates and their homologs with up to 50% sequence identity. Subsequent modeling attempts selected a variety of transporters, from different families. These are collected and summarized in Table II.

TctA_ar and TctA_ct selected different templates. To have both proteins referred to the same template, we included some of the transporter templates already chosen by one protein in the forbidden list, so I-TASSER was forced to try new templates. To speed up this process, we also specified a particular template (4K1C for TctA_ct). We obtained six possible structures for each TctA sequence, based on the following PDB templates: 4F35, 1KPL, 4K1C, 3VVN, 3DH4, and 4IKV, representing all the superfamilies in Table II. We used TctA-specific features to select the most plausible models from this set. For each template, we located the 20-amino acid repeat on the structures of both proteins. Successful templates placed these repeats at equivalent positions in both modeled proteins. Furthermore, we required that the location of G107 (the equivalent of G109 in *S. enterica*) be close by in the models of both proteins. Only two templates complied with these conditions, 3VVN (MATE, MOP superfamily[24] and 4K1C (CDF superfamily[25]), as shown in Figure 3. The overall quality of the models, as indicated by their C-scores and TM-scores is as follows: TctA_ar with the 3VVN template (−3.22, 0.35 ± 0.13), TctA_ar with the 4K1C template (−2.76, 0.40 ± 0.13), TctA_ct with the 3VVN template (−1.68, 0.51 ± 0.15), and TctA_ct with the 4K1C template (−2.81, 0.39 ± 0.13). As expected from the low sequence similarity to the deposited structures in the PDB, C-scores are low, and TM-scores are also low. The MATE transporters have 12 transmembrane helices, in a 6 + 6 topology, whereas the CDF transporters have a 5 + 5 topology. The two templates provide different solutions for the ~70 amino acid length difference between TctA_ar and TctA_ct: in the context of 3VVN, insertions are spread throughout the whole structure; in the context of 4K1C, most of the difference is pushed to the C-terminus, concentrated in a helical hairpin that lies outside the core of the transporter (see Fig. 4). Although both MATE and CDF transporters described to date are antiporters, the exact mechanism for translocation has not been elucidated, and there is no *a priori* reason to forbid symport in these architectures.

A possible role of the 20-amino acid repeat can be proposed from these structures. In those derived from 3VVN, this segment lies in two symmetry-related helices (helices 1 and 7, as envisioned in Ref. 5) that line the ample binding pocket of the transporter. In the 4K1C-derived structures, the first repeat lies in the MRb helix, apposed to helix 6, where the other repeat lies; glycine-rich helices can slide easily against other helices and may therefore be involved in the transport cycle conformational switch. Furthermore, as MRb helices are not conserved in the CDF superfamily (Supporting Information

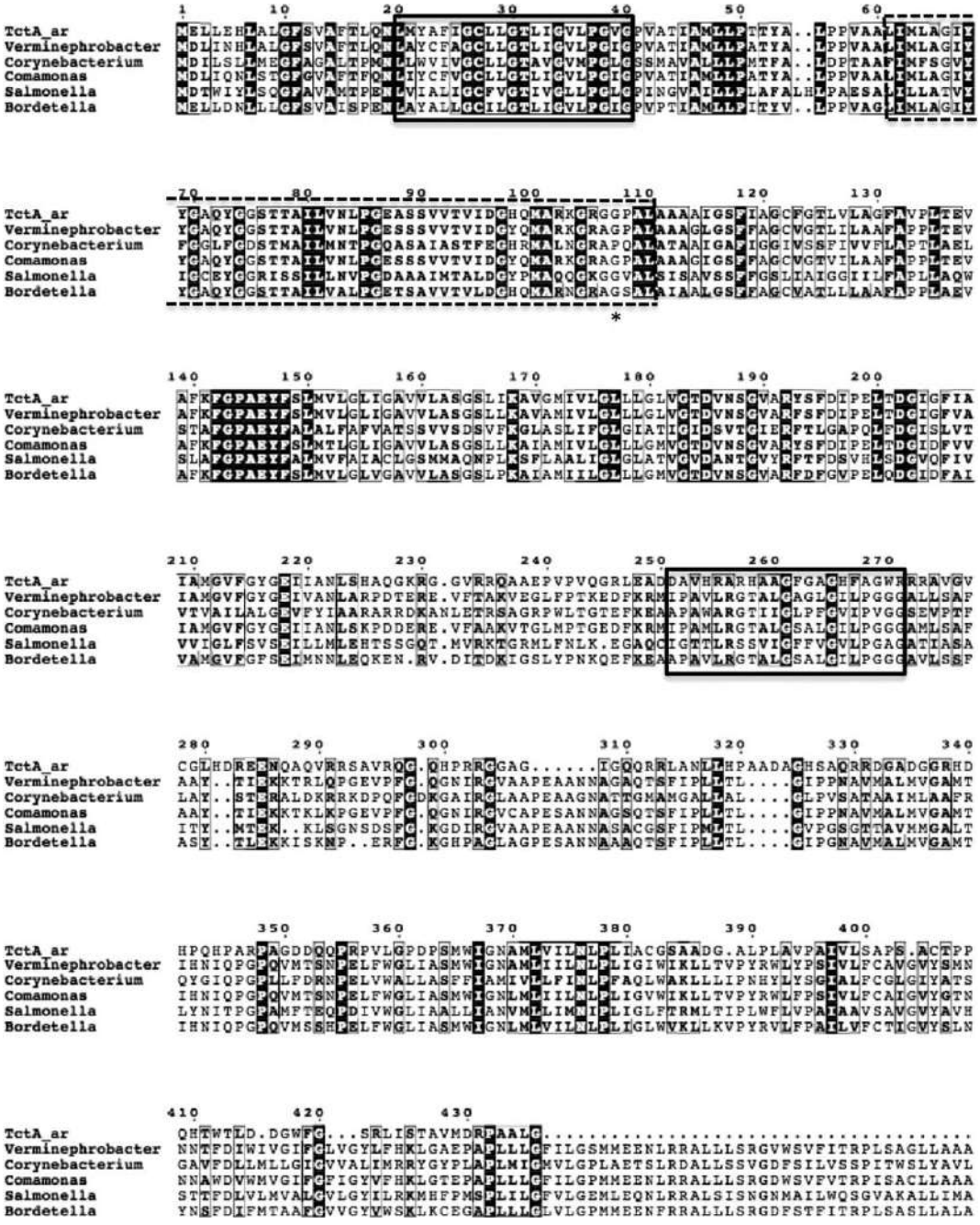


Figure 2

Conserved sequence blocks in the multiple sequence alignment between TctA_ar and functionally characterized TctAs. Boxes indicate conserved sequences, and similar sequences are indicated by darker background. The thick full line boxes indicate the repeated 20-amino acid blocks. The thick broken line box indicates a highly conserved region in the N-terminal repeat of TctA proteins. The asterisk marks the mutated G that eliminates transporter function in *Salmonella*.

Fig. 2 in Ref. 25), this model could also accommodate TctA sequences with truncated N-termini. With regard to G107, in 3VVN-derived structures it lies at the beginning of helix 3, facing the connecting loop between the two repeats, and could be involved in the translocation cycle. In 4K1C-derived structures, it lies also at the beginning of helix 3, facing the loop between helices 8 and 9, at the opposite end of the acidic helix. Its functional role in this template is harder to glean.

The TM scores to the templates range from 0.856 in the best case (TctA_ar in the 3VVN template) to 0.700 in the worst case (TctA_ct in the 4K1C template), with

Table II
Transporter Templates Selected by I-TASSER for TctA_ar and TctA_ct

PDB ID	Family	Transported species	TC number	Superfamily
4F35	DASS	Divalent anion:Na^+ symport	2.A.47.5.2	IT
3Q17, 3ORG, 1KPL, 3ND0	CLC	Cl^-:H^+ antiport	2.A.49.5	–
4K1C,	VCX	Ca^{2+}:H^+ antiport	2.A.19	CDF
4KPP, 4KJS	CAX			
3V5U	NCX	Ca^{2+}:Na^+ antiport		
3VVN, 3MKT, 3HUK	MATE	Substrate:H^+ antiport Substrate:Na^+ antiport	2.A.66.1	MOP
2XQ2, 3DH4	SSS	Galactose:Na^+ symport	2.A.21.3	APC
2JLN	NCS1	Benzyl-hydantoin:cation symport	2.A.39.3	APC
3QE7	NCS2	Nucleobase:cation symport	2.A.40.1.1	APC
3GIA	APC	Amino acid:H^+ symport	2.A.3.6.3	APC
2WIT	BCCT	Glycine-betaine:Na^+ symport	2.A.15.1	APC
4IU8	NNP	Nitrate:H^+ symport; nitrate/nitrite antiport	2.A.1.8.10	MFS
1PW4	OPA	Glycerol-phosphate:phosphate antiport	2.A.1.4.3	MFS
4APS, 4IKV, 2XUT, 4LEP	POT	Peptide:H^+ symport	2.A.17.1	MFS
4J05	PHS	Phosphate:H^+ symport	2.A.1.9	MFS
2GFP	DHA1	Grug:H^+ antiport	2.A.1.2	MFS
4GBY	SP	Xylose:H^+ symport	2.A.1.1	MFS

coverages ranging from ~70% to ~90%. TM scores above 0.5 are good indicators of structural compatibility, and our models are clearly above this threshold. We stress that this does not mean that the models are correct; it only suggests their plausibility. Furthermore, TM scores above 0.5 are not common when the trial sequence cannot be successfully threaded into a template, an expected outcome for protein sequences that have negligible homology to reported structures in the PDB; we found many instances of TM scores below 0.3.

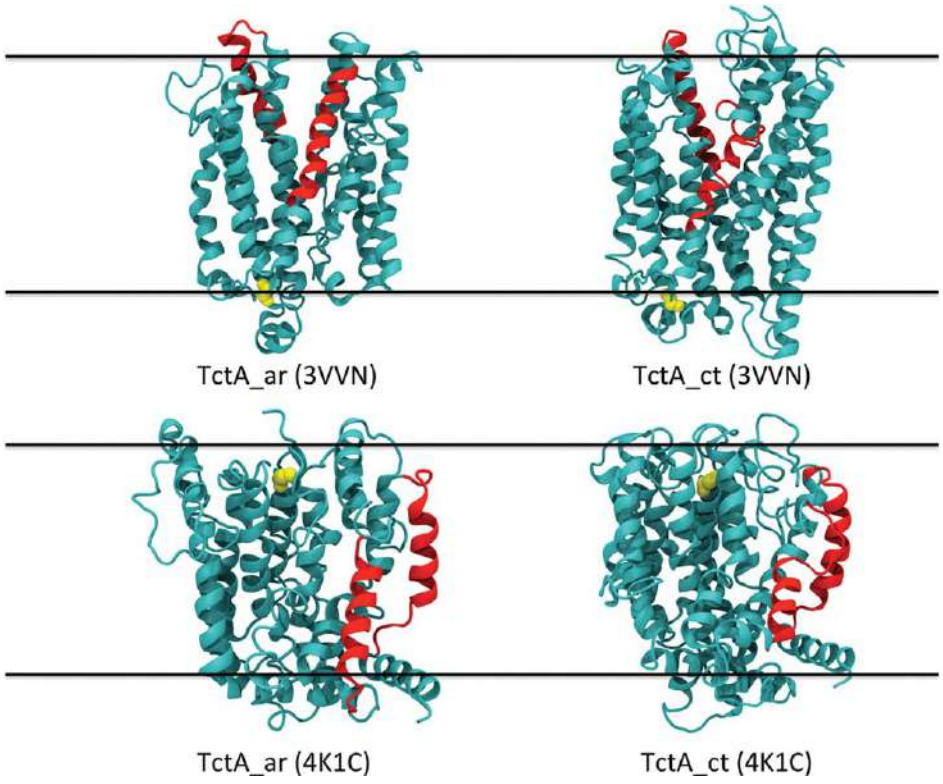


Figure 3
Model structures for TctA_ar and TctA_ct derived from 3VVN (top) and 4K1C (bottom). The structures are shown in cyan ribbons, highlighting in red ribbons the 20-amino acid repeat. G107 is shown as yellow spheres. The horizontal lines mark the approximate location of a lipid bilayer. The extracellular/vacuolar space lies at the top of the figure.

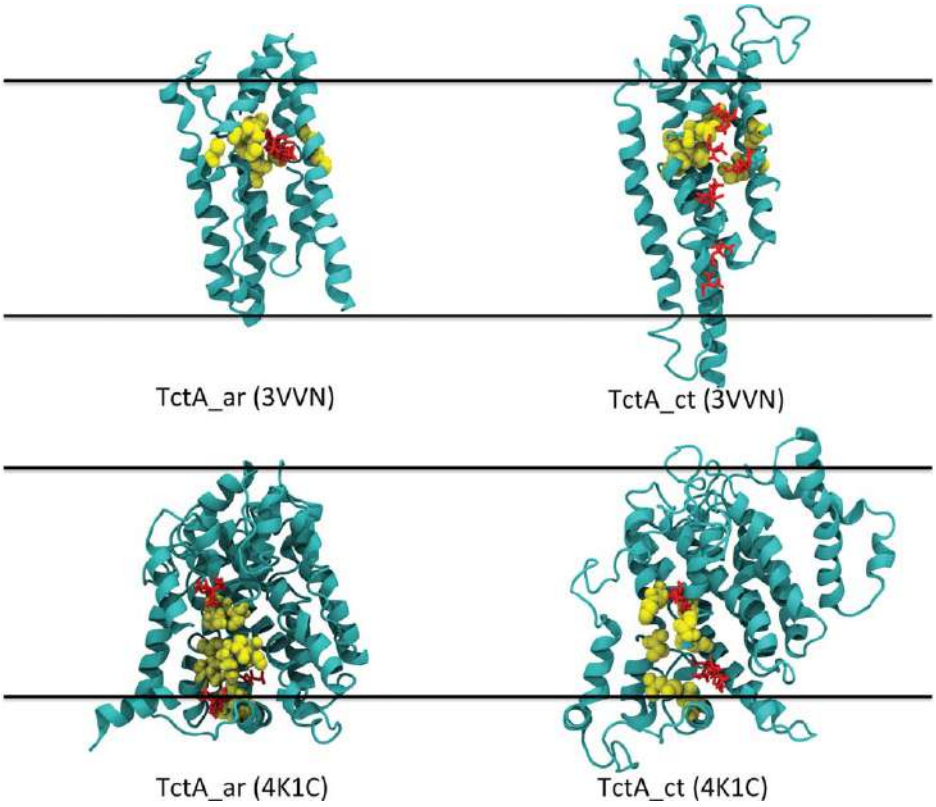


Figure 4

Citrate binding sites for TctA_ar and TctA_ct derived from 3VVN (top) and 4K1C (bottom). The models for each transporter, based on the same template, were superimposed and are presented in the same orientation, which is approximately rotated 90° with respect to that shown in Figure 3. For clarity, roughly one of the repeats has been cut away to show the citrate binding sites. The transporters (cyan ribbons) are immersed in a bilayer (horizontal lines), with the bound citrate molecules (red sticks), and the substrate binding residues (yellow spheres) translated from the template. The extracellular/vacuolar space lies at the top of the figure. The acidic helix in 4K1C-derived structures points toward the viewer and lies at the center bottom of the structures. The extra helical hairpin at the C-terminus of TctA_ct lies to the right of the main structure. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Docking of citrate in the TctA models

Many regions in the models display loops inserted in helical regions, and we did not attempt to fix those, as this lies beyond the scope of this work. By necessity, the models are still at a low-resolution phase, and require experimental input to be improved. Nonetheless, the models can be useful to inquire about the function of the proteins. Given that we are proposing these structures to be citrate transporters, we tested their ability to bind citrate, using AUTODOCK/VINA[23] with the standard default parameters, allowing for flexibility in citrate. We restricted the docking area to the surfaces not in contact with the DMPC bilayer and analyzed the default 10 poses returned by the program. The results are shown in Figure 4.

To guide the analysis, we identified equivalent binding site residues from the templates and their homologs in both proteins, by performing structural alignments with STAMP[26] in VMD.[27] 3VVN is a MATE transporter from *Pyrococcus furiosus*, and it uses protons as cotransported solutes. Given that TctAs are proposed to use sodium in cotransport, we used the NorM MATE homolog from *Vibrio cholerae* (3MKT and 3MKU[28] to identify the sodium-binding site. This architecture presents a large substrate-binding site; crystal structures of other NorM homologs have been crystallized with cationic drugs and peptides, binding at different depths and locations.[24,28,29]. Table III summarizes the sodium binding pocket residues in *V. cholerae* NorM (3MKU[28] and the structurally equivalent residues in TctA_ar and TctA_ct. The counterion for sodium is D371, and it is conserved in our models at the appropriate location. All of these residues lie in the C-terminal half of the transporter, where sequence variability is greater. TctA_ar binds citrate at only one place, adjacent to the putative sodium-binding site (Fig. 4, upper left), whereas TctA_ct binds citrate at that location and at various others (Fig. 4, upper right). As these binding sites are found along a line perpendicular to the membrane plane, and along the dyad axis of the transporter, we find them suggestive of a possible translocation pathway, which can also be tested experimentally.

Table III
Ligand Binding Residues in Templates and Model Transporters

3MKU	TctA_ar	TctA_ct	4K1C	TctA_ar	TctA_ct
E255	A259	A263	G102	G86	V82
F259	G263	L274	V105	S89	P85
S285	R291	Q315	E106	S90	G86
F288	A294	F318	G298	A328	T323
Y367	V358	V413	A301	R331	P327
D371	D362	D415	E302	D332	N329
Y398	G388	L445	E83	I67	M63
F429	L414	R470	E230	R235	V213
L433	G417	A474	D234	E239	G217

4K1C is a calcium/proton antiporter that belongs to a family that can also use sodium as cotransported ion.[25,30–32]. The available crystal structures indicate at least two possible binding sites for calcium, which in our case would correspond to citrate. These two sites, also listed in Table III, are occupied, as seen in Figure 4 by the close proximity of citrate molecules to the extrapolated binding site residues. The same architecture has been shown to transport organic and inorganic ions, as happens in the APC family,[33] so this proposal is reasonable. At each of the binding sites there is at least one positive side chain, which binds citrate directly.

From the data presented above, it is currently impossible to choose one template over the other. Table III constitutes a set of direct functional hypotheses, susceptible to exploration with site-directed mutagenesis. As the critical residues are not shared between the two architectures, it may be possible to validate or discard the models in a straightforward approach.

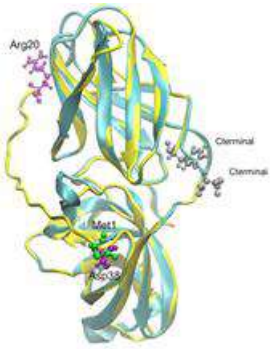
CONCLUSIONS

The modeling attempt presented here rationalizes the 20-amino acid repeats that are a hallmark of TctA transporters, and also suggests why the G109S mutation in TctA of *S. enterica* abrogates function. Furthermore, we suggest two disjoint sets of putative citrate and/or sodium-binding residues, which are presented as concrete functional hypotheses to be tested by mutagenesis in future studies. After experimental validation of any of the two proposed architectures, this would be a relevant starting point for understanding substrate specificity and the transport mechanism for this understudied transporter family.

REFERENCES

1. Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. PNAS 2002;99:10494–10499.
2. Daniel R. The soil metagenome: a rich resource for the discovery of novel natural products. Curr Opin Biotech 2004;15:199–204.
3. Schloss R, Patrick D, Handelsman J. Biotechnological prospects from metagenomics. Curr Opin Biotech 2003;14:303–310.
4. Simon C, Daniel R. Metagenomic analyses: past and future trends. Appl Environ Microbiol 2011;77:1153–1161.
5. Winnen B, Hvorup RN, Saier MH. The tripartite tricarboxylate transporter (TTT) family. Res Microbiol 2003;154:457–465.
6. Hosaka M, Kamimura N, Toribami S, Mori K, Fukuda M, Masai E. A novel tripartite aromatic acids transporter is essential for terephthalate uptake in Comamonas sp. strain E6. Appl Environ Microbiol 2013;79:6148–6155.
7. Widenhorn KA, Somers JM, Kay WW. Expression of the divergent tricarboxylate transport operon (tctI) of Salmonella typhimurium. J Bacteriol 1988;170:3223–3227.
8. Ricci V, Loman N, Pallen M, Ivens A, Fookes M, Langridge GC, Wain J, Piddock LJ. The TCA cycle is not required for selection or survival of multidrug-resistant Salmonella. J Antimicrob Chemother 2012;67:589–599.
9. Ma YF, Zhang Y, Zhang JY Chen DW, Zhu Y, Sheng H, Wang SY, Jiang CY, Zhao GP, Liu SJ. The complete genome of Comamonas testosteroni reveals its genetic adaptations to changing environments. Appl Environ Microbiol 2009;75:6812–6819.
10. Brocker M, Schaffer S, Mack C, Bott M. Citrate utilization by Corynebacterium glutamicum is controlled by the CitAB two-component system through positive regulation of the citrate transport genes citH and tctCBA. J Bacteriol 2009;191:3869–3880.
11. Polen T, Schluesener D, Poetsch A, Bott M, Wendisch VF. Characterization of citrate utilization in Corynebacterium glutamicum by transcriptome and proteome analysis. FEMS Microbiol Lett 2007;73:109–119.
12. Antoine R, Jacob-Dubuisson F, Drobecq H, Willery E, Lesjean S, Locht C. Overrepresentation of a gene family encoding extracytoplasmic solute receptors in Bordetella. J Bacteriol 2003;185:1470–1474.
13. Magalhaes JV. How a microbial drug transporter became essential for crop cultivation on acid soils: aluminium tolerance conferred by the multidrug and toxic compound extrusion (MATE) family. Ann Bot 2012;106:199–203.
14. Saier MH, Yen MR, Noto K, Tamang DG, Elkan C. The transporter classification database: recent advances. Nucl Acids Res 2009;3:274–278.
15. Shi Y. Common folds and transport mechanisms of secondary active transporters. Ann Rev Bioph 2013;42:51–72.
16. Kebbel F, Kurz M, Arheit M, Grutter MG, Stahlberg H. Structure and substrate-induced conformational changes of the secondary citrate/sodium symporter CitS revealed by electron crystallography. Structure 2013;21:1243–1250.
17. Berman HM, Westbrook J, Feng Z. The protein data bank. Nucl Acid Res 2000;28:235–242.
18. Felske A, Engelen B, Nubel U, Backhaus. Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. Appl Environ Microbial 1996;62:4162–4167.
19. Bomeman J, Jack R. PCR primers that amplify fungul rRNA genes from environmental samples. Appl Environ Microbial 2000;66:4356–4360.
20. Cho CM, Mulchandani A, Chen W. Bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of organophosphate nerve agents. Appl Environ Microbiol 2002;68:2026–2030.
21. Zhang Y, Skolnick J. Segment assembly, structure alignment and iterative simulation in protein structure prediction. BMC Biol 2013;11:44.
22. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–1802.
23. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, #efficient loptimization, and multithreading. J Comput Chem 2010;31:455–461.
24. Tanaka Y, Hipolito CJ, Maturana AD, Ito K, Kuroda T, Higuchi T, Katoh T, Kato HE, Hattori M, Kumazaki K, Tsukazaki T, Ishitani R, Suga H, Nureki O. Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. Nature 2013;496:247–251.
25. Waight AB, Pedersen BP, Schlessinger A, Bonomi M, Chau BH, Roe-Zurz Z, Risenmay AJ, Sali A, Stroud RM. Structural basis for

alternating access of a eukaryotic calcium/proton exchanger. Nature 2013;99:107–110.
26. Rusell RB, Barton GH. Multiple protein sequence alignment from tertiary structure comparison. Prot Struct Funct Genet 1992;67:309–323.
27. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996: 33–38.
28. He X, Szewczyk P, Karyakin A, Evin M, Hong WX, Zhang Q, Chang G. Structure of a cation-bound multidrug and toxic compound extrusion transporter. Nature 2010;467:991–994.
29. Lu M, Symersky J, Radchenko M, Koide A, Guo Y, Nie R, Koide S. Structures of a Na+-coupled, substrate-bound MATE multidrug transporter. PNAS 2013;110:2099–2104.
30. Abramson J, Paz A, Philipson KD. It's all in the symmetry. Science 2012;335:669–670.
31. Liao J, Li H, Zeng W, Cao H, Cui Z. Structural insight into the ion-exchange mechanism of the sodium/calcium exchanger. Science 2012;335:686–690.
32. Nishizawa T, Kita S, Maturana AD, Furaya N, Hirata K, Kasuya G, Ogasawara S, Dohmae N, Iwamoto T, Ishitani R, Nureki O. Structural basis for the counter-transport mechanism of a H+/Ca2+ exchanger. Science 2013;341:168–172.
33. Wong FH, Chen JS, Reddy V, Day JL, Shlykov MA, Wakabayashi ST, Saier ST. The amino acid-polyamine-organocation superfamily. J Mol Microbiol Biotech 2012;22:105–113.



Capítulo IV.

Una nueva expansina del hongo de la podredumbre blanca *Schizophyllum commune*.

Ya hemos mencionado como las prospecciones metagenómicas permiten el estudio de nuevas poblaciones bacterianas (Capítulos I y II) y el acceso a nueva secuencias de interés (Capítulos I y III). Pero adicionalmente las prospecciones genómicas también resultan de utilidad en el marco de la caracterización de nuevas proteínas de interés para la deconstrucción de lignocelulosa. En este caso, interesados por proteínas (sin actividad enzimática) que en su interacción con polímeros de estructuras cristalinas (especialmente con celulosa) permitan una "relajación" structural que favorezca posteriormente su hidrólisis, analizamos secuencias de expansinas en basidiomicetos. Finalmente, en el marco del proyecto doctoral de Omar Eduardo Tovar Herrera del grupo de la Dra. Katiuska Arévalo adscrita al Instituto de Biotecnología de la Universidad Autónoma de Nueva León, participamos en la caracterización bioquímica y estructural de la única expansina codificada en el genoma de *Schizophyllum commune*: *"A novel expansin protein from the white-rot fungus Schizophyllum commune"*.

Las expansinas son proteínas amorfogénicas que "relajan" las zonas cristalinas de la celulosa y otros polímeros (como la quitina) y favorecen la hidrólisis de estos polímeros por sus respectivas enzimas. Esta fue la primera expansina en hongos basidiomicetos reportada y se encontró por primera vez en estas proteínas su efecto cooperativo para la degradación de quitina. Estos elementos realzan su posible uso en biorrefinerías que funcionen no sólo con lignocelulosa, sino también con residuos de la industria camaronera por ejemplo. Por su parte, el análisis estructural también demostró algunas novedades para las expansinas fúngicas estudiadas anteriormente; mientras que los estudios filogenéticos agrupan esta proteína en clados distantes filogenéticamente de expansinas bacterianas y de plantas, y en el mismo clado de expansinas de ascomicetos, aportando las primeras evidencias preliminaries de una nueva familia de expansinas.

Este estudio ha servido de antecedente y nos ha permitido plantear el proyecto *"Canonical expansin and loosenin proteins in basidiomycetes"*, el cual pretende estudiar proteínas amorfogénicas en aproximadamente 16 genomas de basidiomicetos para generar conocimiento relacionados con la estructura de estas proteínas y con posibles aplicaciones en esquemas de biorrefinerías.

RESEARCH ARTICLE

A Novel Expansin Protein from the White-Rot Fungus *Schizophyllum commune*

Omar Eduardo Tovar-Herrera[1☯], Ramón Alberto Batista-García[2,3☯], María del Rayo Sánchez-Carbente[2], María Magdalena Iracheta-Cárdenas[1], Katiushka Arévalo-Niño[1]*, Jorge Luis Folch-Mallol[2]*

1 Instituto de Biotecnología. Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Nuevo León, México, **2** Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México, **3** Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México

☯ These authors contributed equally to this work.
* jordi@uaem.mx (JLFM); katiushka.arevalonn@uanl.edu.mx (KAN)

Citation: Tovar-Herrera OE, Batista-García RA, Sánchez-Carbente MdR, Iracheta-Cárdenas MM, Arévalo-Niño K, Folch-Mallol JL (2015) A Novel Expansin Protein from the White-Rot Fungus *Schizophyllum commune*. PLoS ONE 10(3): e0122296. doi:10.1371/journal.pone.0122296

Academic Editor: Nikolas Nikolaidis, California State University Fullerton, UNITED STATES

Received: November 21, 2014

Accepted: February 17, 2015

Published: March 24, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by grant CB-153789-Q and PAICYT-UANL CN-1008-11 from the National Council for Science and Technology (CONACyT) from the Mexican government: http://www.conacyt.gob.mx. RAB-G and OET-H received a scholarship from CONACyT. The in silico analyses were performed during RAB-G's research stay funded by the Europe Molecular Biology Organization (EMBO: ASTF503-2013) in the Environmental Research Institute, University College Cork, Cork

Abstract

A novel expansin protein (ScExlx1) was found, cloned and expressed from the Basidiomycete fungus *Schizophylum commune*. This protein showed the canonical features of plant expansins. ScExlx1 showed the ability to form "bubbles" in cotton fibers, reduce the size of avicel particles and enhance reducing sugar liberation from cotton fibers pretreated with the protein and then treated with cellulases. ScExlx1 was able to bind cellulose, birchwood xylan and chitin and this property was not affected by different sodium chloride concentrations. A novel property of ScExlx1 is its capacity to enhance reducing sugars (N-acetyl glucosamine) liberation from pretreated chitin and further added with chitinase, which has not been reported for any expansin or expansin-like protein. To the best of our knowledge, this is the first report of a *bona fide* fungal expansin found in a basidiomycete and we could express the bioactive protein in *Pichia pastoris*.

Introduction

Expansins are non-enzymatic proteins that induce extensibility and stress relaxation of plant cell walls, acting as loosening agents [1,2]. Also, they are implicated in cell enlargement and other developmental events requiring cell wall loosening, such as fruit softening, seed germination and organ abscission [3].

Expansins *sensu stricto* are in the range of 225–275 amino acids and consist of two compact domains (D1 and D2) attached by a short linker region (~4 aa) [1,4,5]. D1 is distantly related to the catalytic domain of glycoside hydrolase family-45 (GH45) and D2 is distantly related to group-2 grass pollen allergens [2]. These proteins form a long shallow groove with highly conserved polar and aromatic residues suitably positioned along both domains, to potentially bind plant cell wall polysaccharides [3,6]. Furthermore, cellulose-active proteins with distant homology to either D1 or D2 only have been reported (swollenin, loosenin, cerato-platanin and

(Ireland). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Jorge Luis Folch-Mallol is currently an academic editor for PLOS ONE but this does not alter the authors' adherence to PLOS ONE Editorial policies and criteria.

group-1 allergens of grass pollen) [7–10]. Nomenclature for classifying these proteins has been established and they were designated as expansin-related proteins [5].

Many microbial expansin proteins have been reported, including BsExlx1 from *Bacillus subtilis* [4], PcExl1 from the plant pathogenic bacteria *Pectobacterium carotovorum* [6] and HcExlx2 from the marine bacteria *Hahella chejuensis* [11], with BsExlx1 being the best well characterized expansin from a non-plant source and one of the bacterial expansins crystalized to this date [4,6,11,12]. In addition, a protein designated as Asper-EXP from the plant pathogenic fungus *Aspergillus niger* was recently reported [12]. All these proteins have been demonstrated to bind and act on cellulosic networks and some of them have shown to act synergistically with cellulases and xylanases [6,11,13,14]. Although plant and bacterial expansins have been characterized in terms of polysaccharide binding profiles, filter paper weakening activity, cell wall loosening and synergistic/enhancing effect with glycosyl hydrolases as mentioned above, it is worth noting that there are no published data regarding to the effect of an expansin protein over chitin polysaccharides and no canonical expansin proteins from basidiomycete fungi have been previously reported, although expansin-related proteins have.

In this work, we identified the first true member of fungal expansin in a Basidiomycete (ScExlx1). This protein was obtained from one of the most common white-rot fungi, *Schizophyllum commune*. Also, this is the first report where an expansin protein showed a chitin hydrolysis enhancing effect in addition to being active on cellulose. Finally, a bioactive expansin from an eukaryotic non-plant source was expressed in *Pichia pastoris* for the first time.

Materials and Methods

Strains and growth conditions

A fungal strain of *Schizophyllum commune* RVAN10 isolated from the northeast region of Nuevo León State in Mexico was used for this work. Mycelium was grown on YPD medium (yeast extract-peptone-dextrose) for its propagation and storage. *S. commune* was growth (6 days and 28°C) on mineral base media [8] supplemented with 2% wheat straw as sole carbon source.

Escherichia coli DH5-α was used for the construction and propagation of recombinant plasmids by incubating each transformant in LB medium (Luria-Bertani, Difco, #Cat. 240230) supplemented with the appropriate antibiotics when necessary (ampicillin 100 μg/ml for pJET-ScExlx1; zeocin 25 μg/ml for pPicZαA-ScExlx1) at 37°C for 24 h. For the heterologous expression of ScExlx1 in *Pichia pastoris* all media and protocols are described in the *Pichia* expression manual (Invitrogen, #Cat. K1740-01).

Sequence analysis: alignment, phylogenies and modeling

Proteins sequences were aligned with MUSCLE as implemented by Geneious software (Version 7.1.5. Biomatters, Ltd.) using default parameters. N-linked and O-linked glycosylation sites, disulfide bonds and the presence of signal peptide were predicted using the Hirst [15], DiANNA 1.1 [16] and the SignalP 4.1 web server [17], respectively.

Cladogram visualization of the phylogenetic trees was performed in order to describe the relationships of the expansin derived from the *S. commune´*s genomic analysis with other expansins and expansins-like proteins. For phylogenetic reconstruction, sequences from previously characterized/annotated in NCBI expansins/expansins-like were selected: five sequences from bacteria [4,6,11,12], four from plants [3]; gi_91806950, gi_332658441, gi_114794319 and gi_332661523) and five from fungi (*A. niger*, Asper-EXP [12] and other four sequences which showed the best hits during BLAST analysis using ScExlx1 sequence as query). At the same time, a radial visualization of the phylogenetic divergence was constructed in order to describe

the relationship between EXLX1, EXLX2 subfamilies and ScExlx1 (in both subfamilies only those expansins experimentally evaluated were considered). Additionally, other expansins not characterized to this date and the Asper-Exp protein [12] were considered. Phylogenetic trees were prepared using the server Phylogeny.fr (http://www.phylogeny.fr/), this platform considers various bioinformatics algorithms to construct a robust phylogenetic tree from a set of sequences [18,19]; for the generation of phylogenetic trees, MUSCLE was used for the multiple alignments, Gblocks for the automatic alignment curation (in order to eliminate poorly aligned positions, not allowing smaller final blocks and less strict flanking positions), BioNJ for tree building and TreeDyn for tree drawing [19]. The Neighbor-Joining (NJ) method was used to estimate the phylogenetic tree; the aligned sequences were bootstrapped 1000 times and the Jones-Thornton-Taylor (JTT) model was used to estimate distances for amino acids [19]. The parameters used during the MUSCLE alignment were those recommended by the Phylogeny.fr platform (custom mode with 16 as the maximum number of iterations).

The amino acid sequence for the ScExlx1 expansin was submitted to the I-TASSER server [20] without constraints in order to get a three-dimensional model of the protein. A second modeling was performed using PDB 3D30 as a template. The visualization and structural alignment were obtained in VMD (Visual Molecular Dynamic).

Cloning and heterologous expression of ScExlx1

Isolation of total RNA was performed on 6-day-old culture of *S. commune* on wheat straw medium using the Trizol method (Invitrogen, #Cat. 15596–026). First-strand cDNA synthesis was performed using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific, #Cat. K1631) following the manufacturer´s instructions. The amplification of the full-length *ScExlx1* cDNA was performed using specific primers designed from *S. commune* H4-8 genome (http://genome.jgi.doe.gov; protein ID: 2642684): Forward primer ScExlx1F (5′-ggtaccgtccaccacaccaccgcgaa-3′) and Reverse primer ScExlx1R (5′-tctagaccgaactgcgacccgccctcg-3′). In both primers, sequences for restriction sites (underlined sequences) were added for *Kpn*I and *Xba*I recognition. The 654 bp PCR fragment was purified and cloned in the pJET vector (Thermo Scientific, #Cat. K1231) resulting in pJET-ScExlx1. The *ScExlx1* cDNA was further sequenced using the pJET primers.

The *ScExlx1* cDNA fragment cloned into pJET vector was digested with *Kpn*I and *Xba*I and purified with GeneJET Gel extraction kit (Thermo Scientific, #Cat. K0691). In parallel, pPICZαA was digested using the same restriction enzymes, and *ScExlx1* was ligated at the corresponding sites into pPICZαA in frame with both the yeast α-secretion factor and C-terminal His_6 tag encoding sequences. pPicZαA-ScExlx1 was linearized with restriction enzyme *Sac*I and used for transformation of *P. pastoris* X-33 by electroporation and selected on YPD plates containing zeocin (100 μg/ml) (Invitrogen, #Cat. R250-01). The pPicZαA vector without insert was also transformed into X-33 and this strain was used as a negative control.

Ten randomly chosen Zeocin-resistant *P. pastoris* transformants were then screened for protein expression in 12.5 ml of BMGY (buffered complex medium containing glycerol) at 28°C in an orbital shaker (220 rpm) for 18 h to an OD_{600} of 2–6, and expression was induced by transferring cells into 50 ml of BMMY (buffered complex medium containing methanol) and growing them for a another 3 days. Each day the medium was supplemented with methanol at 0.5% (v/v). The supernatant was then analyzed by SDS-PAGE to determine which transformant had the best secretion yield.

Protein purification and quantification. SDS-PAGE and Western Blot

Culture supernatant was concentrated using Vivaspin centrifugal units (Sartorius, #Cat. VS2001) with a 10 kDa cut-off at 7,000 rpm and 4°C. The concentrated supernatant was loaded on a HisTrap excel Nickel column (GE Healtcare, #Cat. 17-3712-05) connected to a peristaltic pump and previously equilibrated with phosphate buffer (20 mM NaH_2PO_4, 20 mM Imidazole, 0.5 M NaCl pH 7.4). SDS-PAGE and Western Blot analyzed fractions were pooled, buffer exchanged and dialyzed using Vivaspin centrifugal units coupled to diafiltration cups (Sartorius, #Cat. VSA005) against acetate buffer (50 mM, pH 5).

Total protein concentrations of crude supernatant or purified fractions were determined by Bradford assay [21] using Bovine Serum Albumin (BSA) as calibration standard. Molecular mass estimation of recombinant ScExlx1 was done loading 5 μg of protein into 12% SDS-polyacrilamyde gel. Protein bands were visualized by Coomassie Blue R-250 (Sigma-Aldrich) staining and PageRuler Plus Pre-stained Protein Ladder (Thermo Scientific, #Cat. 26619) was used for molecular mass estimation.

For Western blot analysis, purified ScExlx1 was run on a 12% SDS-PAGE and blotted onto a nitrocellulose membrane (Bio-Rad) using a wet tank blotting system (Bio-Rad). After transference, the membrane was washed three times with phosphate buffer containing 0.1% Tween 20 (PBST) pH 8. The membrane was blocked with PBST plus skimmed milk (3%) for 20 min and washed with PBST. c-Myc (9E10) (Santa Cruz Biotechnology, #Cat. Sc-40) and anti-His antibodies (Roche, #Cat. 11922416001) were used for immunodetection (dilution 1:5000) and signal detection was visualized using an anti-mouse alkaline phosphatase conjugate (Sigma, #Cat. A3562) (dilution 1:10000) by incubating 30 min in PBST plus skimmed milk. The membrane was washed three times with PBST and 1 ml of Fast Red TR/Naphthol AS-MX (Sigma, #Cat. F4648) was added for detection of alkaline phosphatase. Deglycosylation was performed using PNGase F (New England Bio-labs, #Cat. P07045) in order to remove ScExlx1 N-linked glycans, according to the manufacturer´s instructions.

Disruptive activity of ScExlx1 on cotton fibers and avicel

Cotton fibers were mercerized according to [7]; briefly, 1 mg of cotton fibers were incubated with 25% NaOH for 15 min at 4°C, and washed several times with distilled water until a pH ~7 was reached. Cotton fibers were suspended in 1 ml of sodium acetate 50 mM pH 5, containing 20 μg of ScExlx1. After incubation for 72 h at 25°C the amount of reducing sugars from supernatant was analyzed by DNS (3,5-Dinitrosalicylic acid) method as described by [22] and the fibers were washed with acetate buffer, sonicated for 1 min and visualized by light microscopy. Avicel PH-101 (1 mg) (Fluka, #Cat. 11365) was treated using 20 μg of ScExlx1 in 1 ml of sodium acetate and incubated as mentioned above. After incubation, microtubes were centrifuged at 13,000 rpm for 5 min and washed 3 times with acetate buffer. Afterwards, avicel was observed by light microscopy. Acetate buffer or proteins from mock supernatant were used as control treatments for cotton fibers and avicel experiments. For each experiment, a total of three replicates were made.

Salt effect on binding to polysaccharides

Polysaccharides binding profile of ScExlx1 was carried out as mentioned by Chen et al. [23] with some modifications. 2.5 mg of Avicel or chitin from shrimp shells (Sigma, #Cat. C7170) were incubated with ScExlx1 (40 μg) in sodium phosphate buffer 50 mM pH 7.4 containing 0 to 500 mM NaCl for 15 min under agitation. After incubation samples were centrifuged at 13,000 rpm for 5 min at room temperature and unbound protein was measured by densitometry with Quantity One software (Bio-Rad). Bound protein quantity was determined by

subtracting the amount of protein used for the experiment minus the amount of protein detected in supernatants. For each experiment, a total of three replicates were made.

Enzymatic hydrolysis

Along with the disruptive activity of ScExlx1 experiments, an independent assay was performed under the same conditions, but adding a cellulase cocktail from *Trichoderma reesei* (Sigma, #Cat. C-2730) in order to determine if ScExlx1 treatment increased the amount of released reducing sugars during the experiments. Briefly, 0.25 U of cellulase was added to 1 mg of cotton fibers after 72 h incubation with ScExlx1 and temperature increased to 50°C for 3 h. Aliquots of 50 μl were taken at 0, 5, 10, 20, 40, 60 and 180 min. In addition to cellulolytic experiments with ScExlx1 treated cotton fibers plus cellulase, a similar experiment was performed in order to determine the effect of a chitinolytic enzyme when acting on ScExlx1 treated chitin. Briefly, chitin (5 mg) was incubated with ScExlx1 (50 or 100 μg) in phosphate buffer 100 mM pH 7.4 for 24 h at 25°C. After incubation, chitinase from *Streptomyces griseus* (0.25 U; Sigma, #Cat. C6137) was added in 1 mL reaction and incubated for 2 h at 37°C. The reactions were centrifuged at 13,000 rpm for 5 min at room temperature and 500 μl of supernatant were mixed with 500 μl of DNS and boiled for 5 min. The amount of reducing sugars released was determined by DNS method using N-acetylglucosamine (NAG) (Sigma, #Cat. A8625) as a calibration standard (1, 2, 3, 4, 5, 10, 15 and 20 μmol/ml). Phosphate buffer 100 mM pH 7.4 and BSA (50 μg) were used as controls.

Statistical analysis

For statistical treatment of experimental data, the arithmetic mean and the standard deviation were calculated. Simple classification ANOVA tests were applied to determine significant differences between the different cases. Firstly, the assumptions of ANOVA were revised: analysis of homogeneity of variance (Hartley-Cochran-Bartlett test) and normal distribution (Kolmogorov-Smirnov and Lilliefors tests) were performed [24]. Subsequently ANOVAs were conducted to demonstrate the similarities or differences between the data of the population of samples. Finally, a post hoc analysis that defines the order of the differences found in the ANOVAs was developed. The Fisher LSD, Tukey HSD and Duncan tests were considered for the post hoc analyses [24,25]. The use of these three tests ensures greater statistical robustness of the proposed analysis. Differences were considered to be significant if $p \leq 0.05$. All statistical calculations were performed in SPSS software (Version 20).

Results

ScExlx1 sequence analysis

A search in the *S. commune* H4-8 v3.0´s genome (http://genome.jgi.doe.gov/Schco3/Schco3.home.html) to find putative expansin and expansin-like proteins was made using a *Clavibacter michiganensis* expansin sequence as query (gi|WP_012038166). Only one protein (protein ID: 2642684) of 239 amino acids with both classic domains for canonical expansins (Domain 1 from amino acid 20 to 118; Domain 2 from amino acid 123 to 239) was found. The nucleotide sequence analysis of ScExlx1 (654 bp) from *S. commune* RVAN10 (GenBank accession number KP698384) showed a 97.2% identity with the ScExlx1 available sequence from *S. commune* H4-8. The amino acid differences between them resulted in 6 amino acids changes at positions: 31 (Ser/Pro), 34 (Asn/Thr), 127 (Val/Ala), 129 (Asp/Tyr), 143 (Asp/Glu) and 225 (Ile/Val). Four of these amino acid changes (34, 127, 143 and 225) were conservative, while only changes at positions 31 and 129 were non-conservative (polar uncharged/non polar, hydrophobic; acidic

polar/polar uncharged, respectively). When comparing global amino acid changes with the other two versions of *S. commune* genomes (*S. commune* Leonen D v1.0 and Tattone D v1.0, available in http://genome.jgi.doe.gov/programs/fungi/index.jsf), ten amino acid changes were detected. Changes at positions 30 (Thr/Ser), 44 (His/Arg), 170 (Ile/Val), 216 (Ser/Thr) and 221 (Thr/Ser) were conservative; while other five changes occurred as follows: 25 (Met/Thr), non polar hydrophobic/polar uncharged, 41 (Lys/Gln), polar basic/polar uncharged, 128 and 182 (both Asn/Asp) polar uncharged/polar acid, and 223 (Pro/Ser) non polar hydrophobic/polar uncharged. This comparative analysis between ScExlx1 and the available *S. commune* putative expansin sequences showed no discrepancies between amino acids responsible for binding to polysaccharides neither the "active" site of the protein suggested by [26]. A signal peptide (signal cleavage from Met-1 to Ala-19, score 0.811) from the ScExlx1 sequence was predicted; and we found one putative N-glycosylation site (Asn-54, score 0.99), while no O-glycosylation sites were detected.

A notably feature of plant expansins from both EXPA and EXPB families is the highly conserved formation of three disulfide bonds in D1 [3]. Similarly, putative disulfide bonds formation between positions 60–86 (score 0.011), 89–107 (score 0.0104) and 110–144 (score 0.0237) were predicted in ScExlx1. Two of the three putative disulfide bonds are located in D1 and the third is between D1 and D2, indicating a similarity between EXPB1 from *Zea mays* and ScExlx1 but was absent in BsExlx1 [3,4], being this lack of disulfide bonds a common feature in bacterial expansins.

In addition, an alignment including sequences from expansins: PcExl1 [6] and two previously crystalized expansins EXPB1 [3] and BsExlx1 [4] revealed a 29.3/42.6% identity/similarity with ScExlx1 and showed that nine of the ten amino acids that form the shallow groove which potentially serves as a polysaccharide-binding site in D1, are strictly conserved between these proteins (including the most conserved residues Thr-12 and Asp-82). Similarly, the three aromatic amino acids (Trp-125, Trp-126 and Tyr-157) forming a planar platform in BsExlx1, which makes D2 resemble to type A CBM (Carbohydrate binding module) are substituted in ScExlx1 by aromatic polar/uncharged amino acids (equivalents to Tyr-161, Tyr-162 and Tyr-195). Furthermore, the three classic motifs from plant expansins (TWYG, GGACG and HFD) were conserved in ScExlx1, presenting slight modifications (T**T**YG, GGAC**S** and H**L**D) (Fig. 1).

Regarding the phylogenetic analysis, two clusters can be observed in the cladogram (Fig. 2A); one that groups the expansin proteins from plants (including members of Expansins

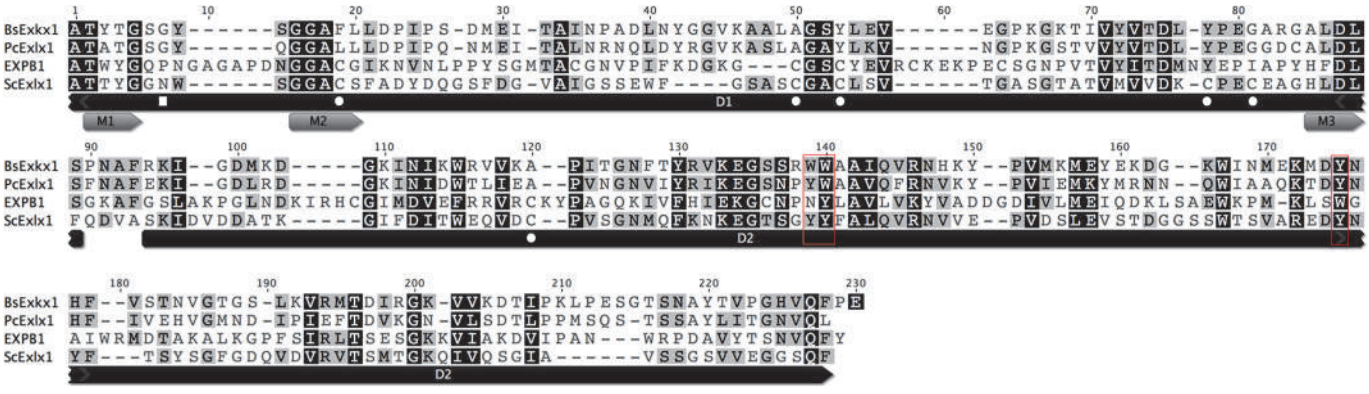


Fig 1. Protein alignment of BsExlx1, PcExl1, EXPB1 and ScExlx1. Darker background indicates high amino acid conservation among the sequences. M1, M2 and M3 indicate the three classic motifs of plant expansins. The boxed amino acids are important residues involved in binding and creep activity in BsExlx1. White square indicates a putative N-glycosylation site in ScExlx1 and white circles indicate the predicted residues to form disulfide bonds.

doi:10.1371/journal.pone.0122296.g001

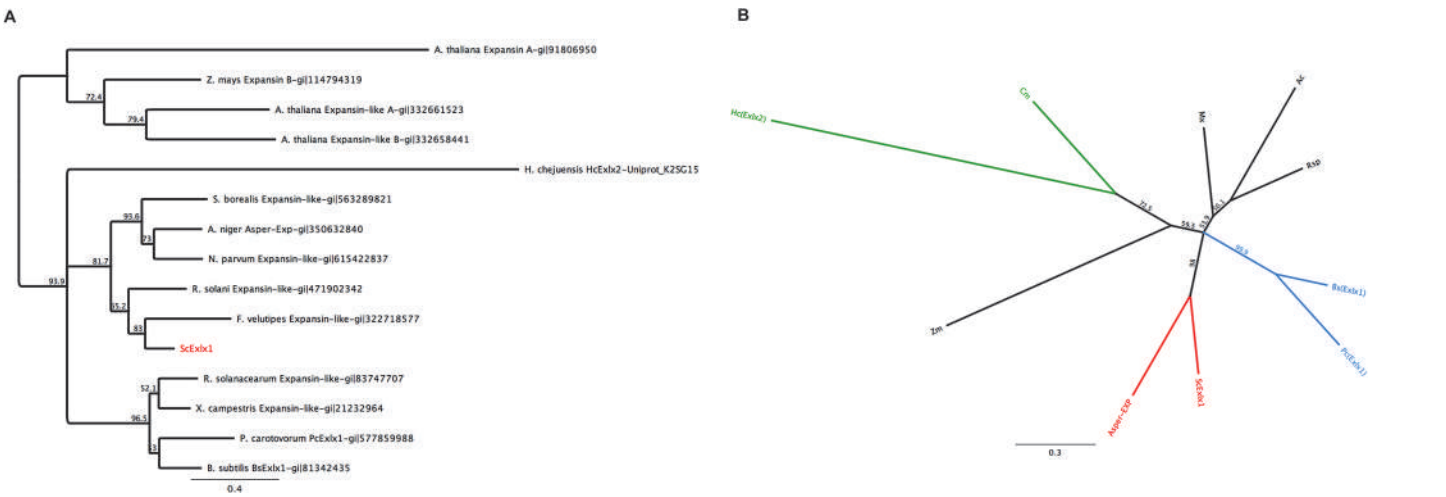


Fig 2. Phylogeny of ScExlx1. A) Phylogenetic relation of ScExlx1 with expansin and expansin-like proteins from plants, bacteria and fungi. Branch lengths represent the amount of genetic change between a node and its descendant. B) Phylogenetic divergence between expansin subfamilies. Ac. *Acidovorax citrulli* (gi|120612050); Mx. *Myxococcus xanthus* (gi|108762346); Rsp. *Roseiflexus* sp. (gi|148655687); Cm. *Clavibacter michiganensis* (gi|148272660). Hc. *Hahella chejuensis* (Uniprot: K2SG15) Pc. *Pectobacterium carotovorum* (gi|577859988); Bs. *Bacillus subtilis* (PDB 3D30). Asper-EXP. *Aspergillus niger* (gi|350632840).

doi:10.1371/journal.pone.0122296.g002

from type A, B, expansin-like A and expansin-like B) which was used to root the tree, and a second cluster that groups previously experimentally characterized expansins [4,6,11–13] and other fungal proteins annotated as expansins in the NCBI database. The phylogenetic analysis showed that ScExlx1 is grouped directly with other fungal expansins (Ascomycota and Basidiomycota) and forming a node with a not experimentally evaluated expansin from the basidiomycete *Flammulina velutipes* (an endophytic fungus). A second sub-branch with prokaryotic expansin-like proteins is observed in Fig. 2A. It is important to note that the HcExlx2 expansin is forming a separate branch, showing no direct phylogenetic relationship with the rest of the microbial expansins.

On the other hand, a radial visualization was obtained to describe the relationship between distinct Exlx subfamilies. Lee et al. [11] proposed that Exlx expansins are divided in two different subfamilies when they found that HcExlx2 is not grouped with BsExlx1. In this order, they define a new expansin subfamily named Exlx2, but this is still controversial. The radial phylogenetic visualization revealed a markedly phylogenetic divergence between ScExlx1 with Exlx1 and Exlx2 subfamilies. This result is not surprising because ScExlx1 is a fungal protein while Exlx1 and Exlx2 are bacterial expansins. It is worth noting that the only fungal expansin previously described (Asper-Exp) groups together with ScExlx1 (Fig. 2B).

In our first attempt to model ScExlx1, I-TASSER selected as templates expansins from the PDB: 3D30 (BsExlx1 from *Bacillus subtilis*), 2HCZ (EXPA from *Zea mays*) and, 4JCW and 4JJC (both Exlx from *Clavibacter michiganensis*). A fifth PDB was chosen to model ScExlx1, 1N10 (a crystal structure from *Phleum pratense*, a major timothy grass pollen allergen). A three-dimensional model beginning in Arg-20 (the signal peptide was removed before modeling) with TM-score of 0.73±0.10 and C-score of 0.14 was obtained from I-TASSER. In this attempt PDB 3D30 was identified by I-TASSER as the major template. In order to obtain a more accurate three-dimensional model, we submitted a new modeling round using PDB 3D30 as a template and obtained a definitive model (Fig. 3A) with TM-score of 0.8±0.10 and C-score of 0.49. These score values confirm a high confidence in the quality of the obtained model and

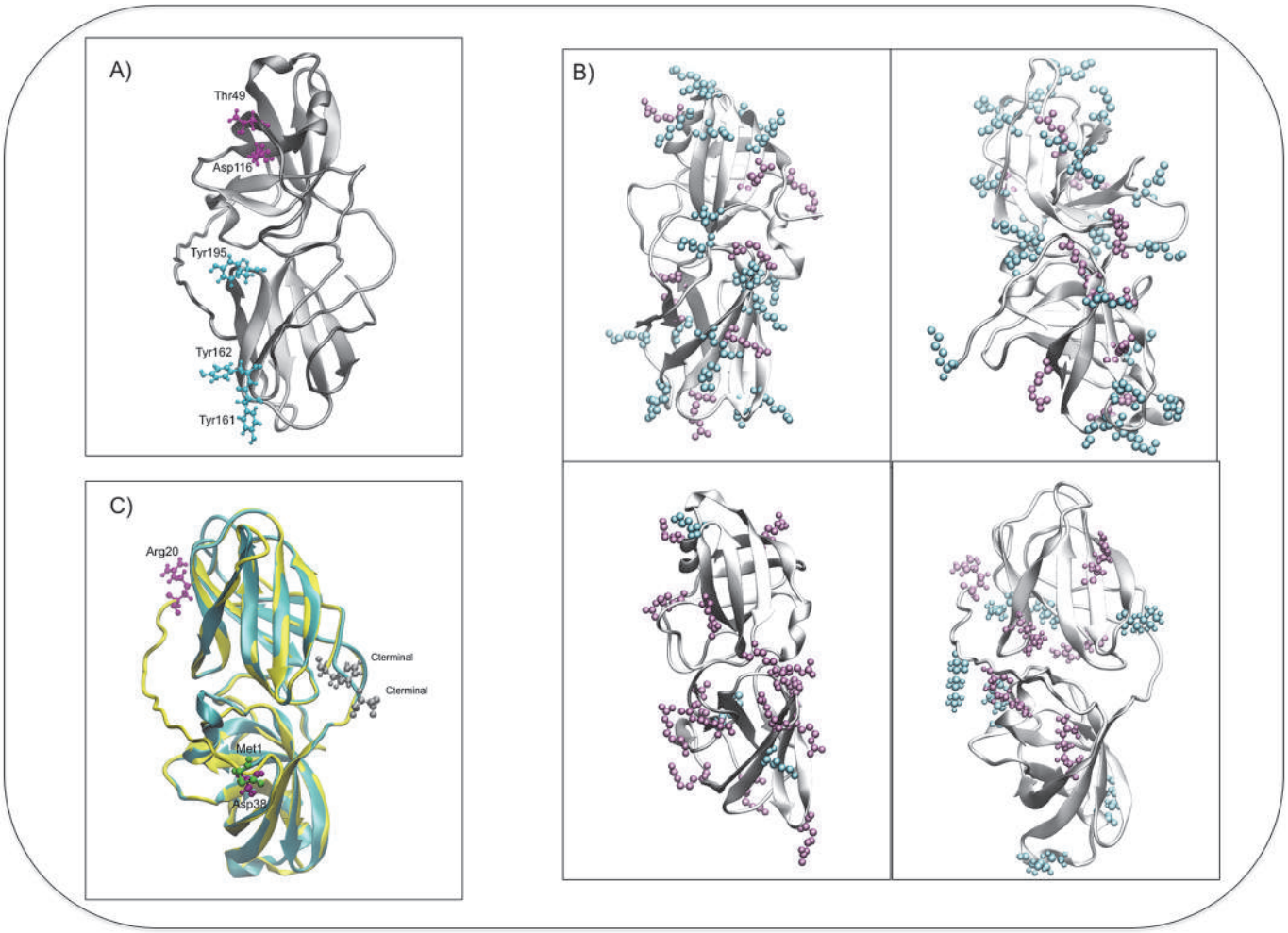


Fig 3. Structure of ScExlx1 and comparison with previously crystallized expansins. A) Three-dimensional model proposed for ScExlx1. Most conserved amino acids between plant and microbial expansins in D1 are depicted in magenta (Thr-49 and Asp-116). Sugar-binding residues in D2 are showed in cyan (Tyr-160, Tyr-161 and Tyr-195). B) Three-dimensional models showing the positive charged amino acids (Arg+Lys) between different EXLX proteins reported previously and ScExlx1 (Lysine is depicted in cyan and Arginine is depicted in magenta). Top-left, PDB: 3D30. Top-right, PDB: 2HCZ. Bottom-left, PDB: 4JCW. Bottom-right, ScExlx1. C) BsExlx1 (cyan model) and ScExlx1 (yellow model) superimposed, showing an N-terminal extension in ScExlx1 that it is absent in BsExlx1. Amino acids depicted in silver, C-terminal in both proteins. Amino acid in green, Met-1 of BsExlx1. Amino acids in purple, Arg-20 and Asp-38 depicting the N-terminal extension in ScExlx1).

doi:10.1371/journal.pone.0122296.g003

additionally suggest the expansin activity of the amino acidic sequence ID: 2642684 from *S. commune*.

As it has been described for PcExlx1, we also found a marked difference in the number of positively charged residues (Arg+Lys) in BsExlx1 compared with ScExlx1. Fig. 3B shows the amount of these residues in the crystallized expansins: BsExlx1 (27 residues), EXPB (26 residues) and *Clavibacter* Exlx (19 residues). ScExlx1 revealed only 13 (Arg+Lys), even less than PcExlx1 [16]. This characteristic supports the acidic properties of ScExlx1 (pI 4.6).

A three-dimensional superposition considering the models derived from BsExlx1 (PDB 3D30) and ScExlx1 was visualized in VMD software previous structural alignment of both models (Fig. 3C). We found an important difference in the structural alignment analysis because an extension of 37 amino acids located in the N-terminal end of ScExlx1 was observed. The function of this extension is not clear to date, it has been described in expansins from

plants [3] and fungi (Asper-Exp) [12]. Due to this extension, the N-terminal end of ScExlx1 and Asper-Exp suggest a structural difference between fungal and bacterial expansins.

Production of recombinant ScExlx1

The ScExlx1 gene product was cloned in frame with sequences encoding the yeast α-factor secretion peptide and a $(His)_6$ tag situated in C-terminus. The recombinant gene was expressed under methanol inducible AOX promoter from *P. pastoris*. The protein was purified and detection was carried out by SDS-PAGE in which three bands were observed with approximate molecular masses of ~24, ~28 (predicted size) and ~30 kDa (Fig. 4A). Confirmation that the three bands are ScExlx1 related was made by western blot analysis detecting the presence of the *myc* epitope present in the recombinant ScExlx1 protein (Fig. 4B).

Preliminary data suggest that the ~30 kDa band is a glycosylated form of the ScExlx1 protein, since bioinformatic analysis showed a putative glycosylation site (Asn-54) and upon treatment with PNGaseF this band disappears (data not shown).

Effect of ScExlx1 on cotton fibers and avicel

The evaluation of the effect of ScExlx1 on mercerized cotton fibers and avicel was carried out, considering that expansins and expansin-related proteins from different sources have been demonstrated to act on cellulose [4,6,12].

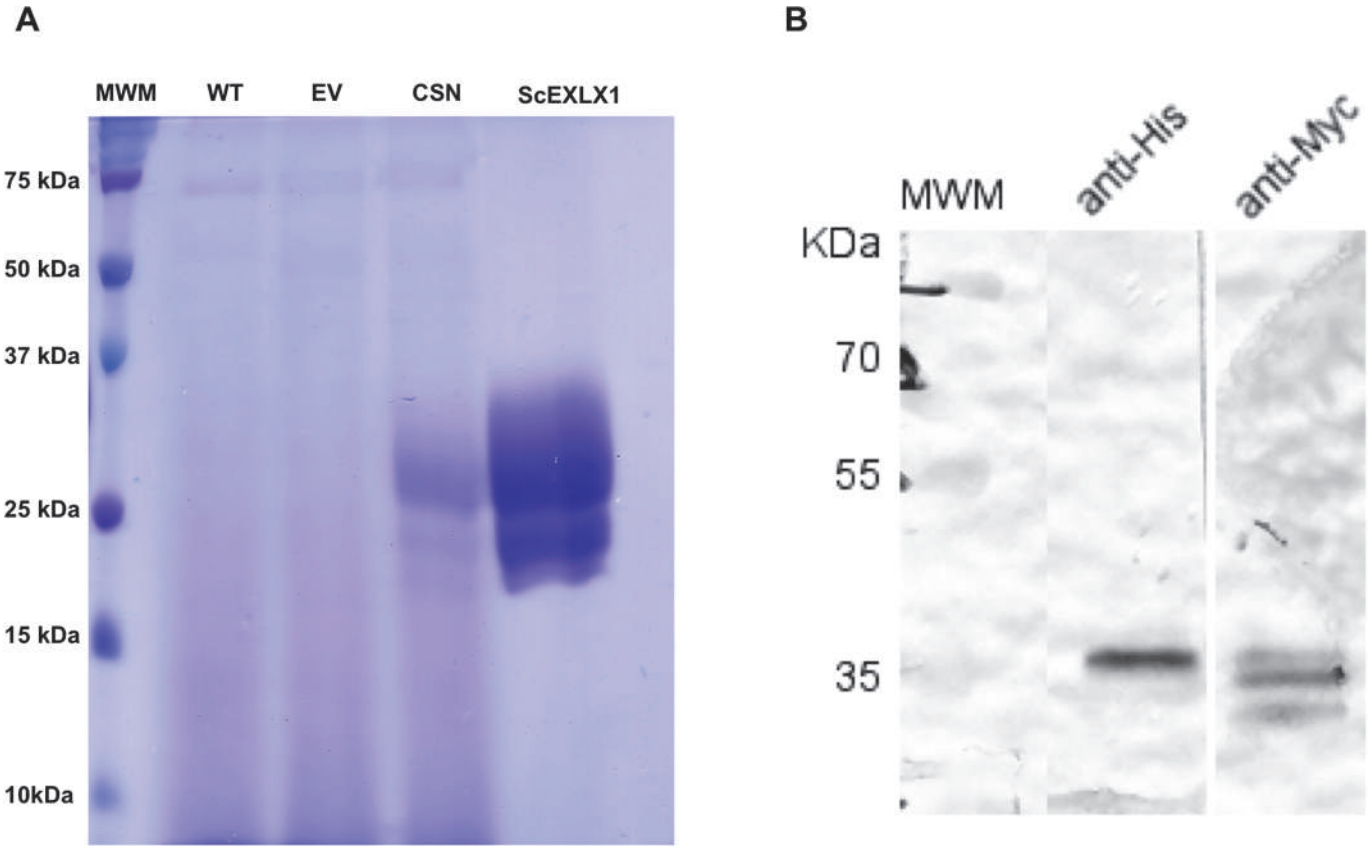


Fig 4. Purification and western blot analysis of recombinant ScExlx1. A) SDS-PAGE of control and recombinant strains. Supernatant from wild type *P. pastoris* X-33 (WT). Supernatant from *P. pastoris* X-33 transformed with pPICZαA empty vector (EV). Supernatant from *P. pastoris* transformed with pPICZαA-ScExlx1 (CSN). Purified ScExlx1. B) Western blot analysis of recombinant ScExlx1.

doi:10.1371/journal.pone.0122296.g004

Fibers treated with buffer or mock supernatant proteins showed no visual changes when observed by light microscopy, while fibers incubated with ScExlx1 exhibited a "bubble" effect, similar to that reported by previous works when treating cotton fibers with expansin-related proteins (Fig. 5A and 5B) [7,8]. Furthermore, when avicel was incubated with 20 μg of ScExlx1 for 72 h at 25°C a reduction in particle size was observed when compared with untreated controls (Fig. 5C and 5D). These results confirm that ScExlx1 is a functional protein and acts on crystalline cellulose.

Binding of ScExlx1 to polysaccharides

The binding capacities of bacterial expansins (on avicel, birchwood xylan and peptidoglycan) and the salt influence in the expansin-substrate interaction have been evaluated [4,6]. In this context, we evaluated the binding capacities of ScExlx1 to avicel and birchwood xylan at different salt concentrations, and we included chitin in the binding experiments given that some expansin-related proteins have been shown to bind this polymer [8,23].

ScExlx1 exhibited the ability to bind to avicel, xylan from birchwood and chitin. Nevertheless, increments in salt concentration showed no effect in binding capability of ScExlx1 to avicel under the conditions used in this study, (Fig. 6A and S1 Table), a different behavior to that reported for the bacterial expansin from *Pectobacterium carotovorum* (PcExlx1) and *Bacillus subtilis* (BsExlx1) [4,6]. In a similar manner, binding of ScExlx1 to chitin was not significantly affected when higher salt concentrations were used for the experiments (Fig. 6B and S1 Table).

Despite previous evidence of salt effect in decreasing up to 90% the binding capability of bacterial expansin from *B. subtilis* (BsExlx1) to xylan [27], no apparent effect was observed when increasing amounts of NaCl were assayed for this polysaccharide (Fig. 6C).

Enzymatic hydrolysis

ScExlx1 and cellulase activity. We aimed to determine if ScExlx1 produced an enhancement of cellulase activity when in combination with a *T. reesei* cellulase cocktail. A total of 6.9 glucose μmol were produced when cellulose hydrolysis reactions were carried out with ScExlx1 treated cotton fibers, while 5.6 glucose μmol were liberated from untreated control in a 3 h experiment (Fig. 7). The same behavior was observed when the experiment was performed for 48 h, given that 11.1 glucose μmol were produced from ScExlx1 treated cotton fibers and only 8.9 glucose μmol from untreated control (S1 Fig.). In addition, when ScExlx1 and cellulases were incubated together, no synergism between ScExlx1 and cellulase occurred (data not shown), and no reducing sugars were detected when incubating cotton fibers with ScExlx1 only.

ScExlx1 and chitinase activity. Expansin-related proteins have shown binding capacities to chitin polysaccharide [8,23]. Interestingly, no reports about the effect of an expansin or expansin-like proteins on chitin have been reported. For this reason, pretreatment of chitin with ScExlx1 before addition of chitinase from *Streptomyces griseus* was performed. This experiment resulted in releasing of 2-fold NAG amount when hydrolyzing ScExlx1 pretreated chitin when compared with buffer and BSA treated chitin (Fig. 8). Doubling ScExlx1 quantity (100 μg) during chitin pretreatment resulted in 0.2 increment of NAG μmol liberated when comparing with original concentration of ScExlx1 used for pretreatment (50 μg).

These results indicate that ScExlx1 is capable of modifying the chitin polymer, an interesting point of view given that fungal cell wall is composed mainly by this polysaccharide. Probably, expansins from fungi could play an important role during cell wall remodeling of these organisms, although more evidence to verify this function for fungal expansins is needed.

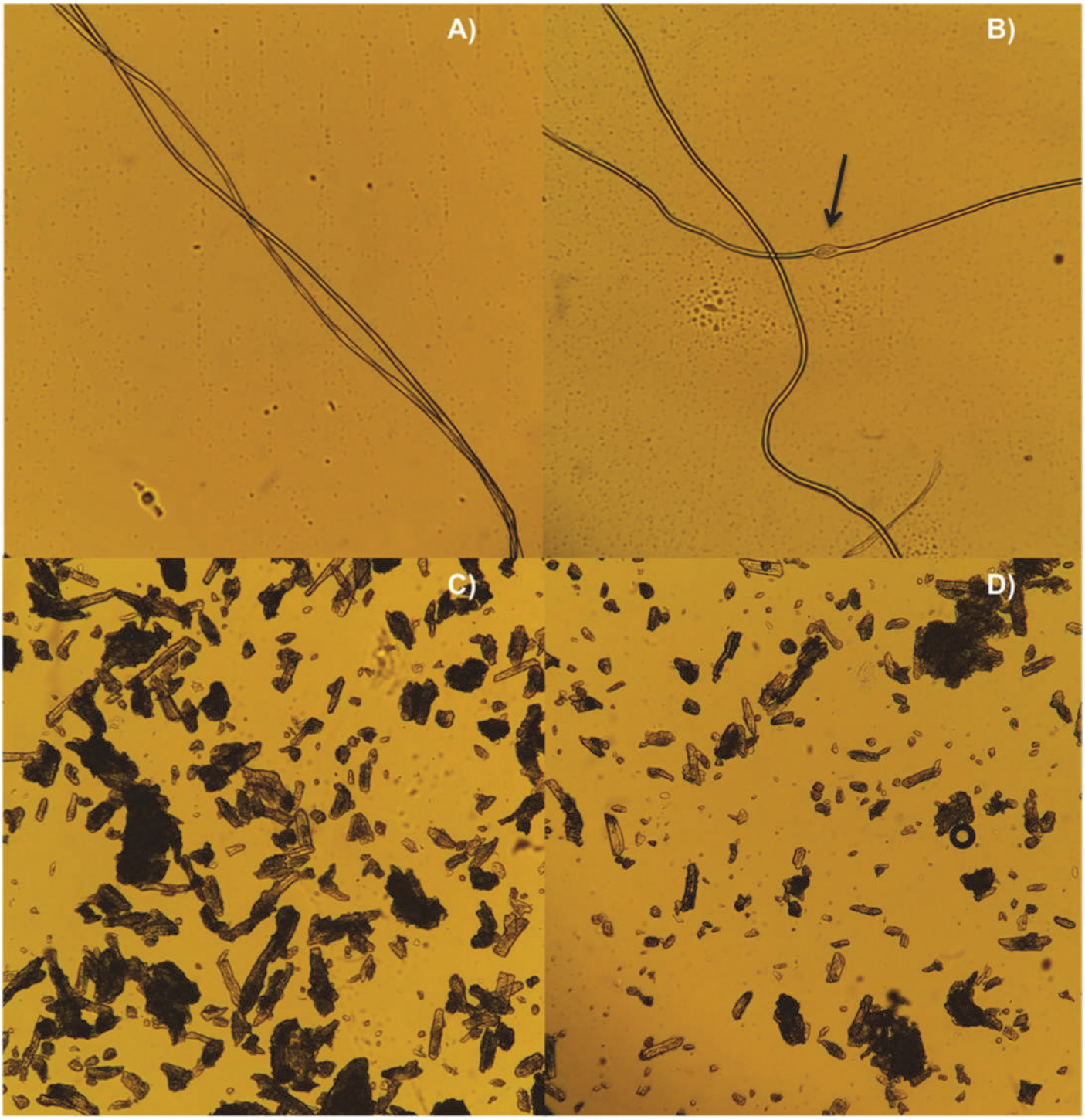


Fig 5. Disrupting activity of ScExlx1 on cotton fibers and avicel. Light microscopy (10X) of cotton fibers and avicel incubated with ScExlx1 or proteins from mock supernatant for 72 h at 25°C. A) Proteins from mock supernatant acting on cotton fibers. B) "Bubble" effect on cotton fibers generated by ScExlx1. C) Avicel incubated with proteins from mock supernatant. D) Reduction in avicel size particle mediated by ScExlx1.

doi:10.1371/journal.pone.0122296.g005

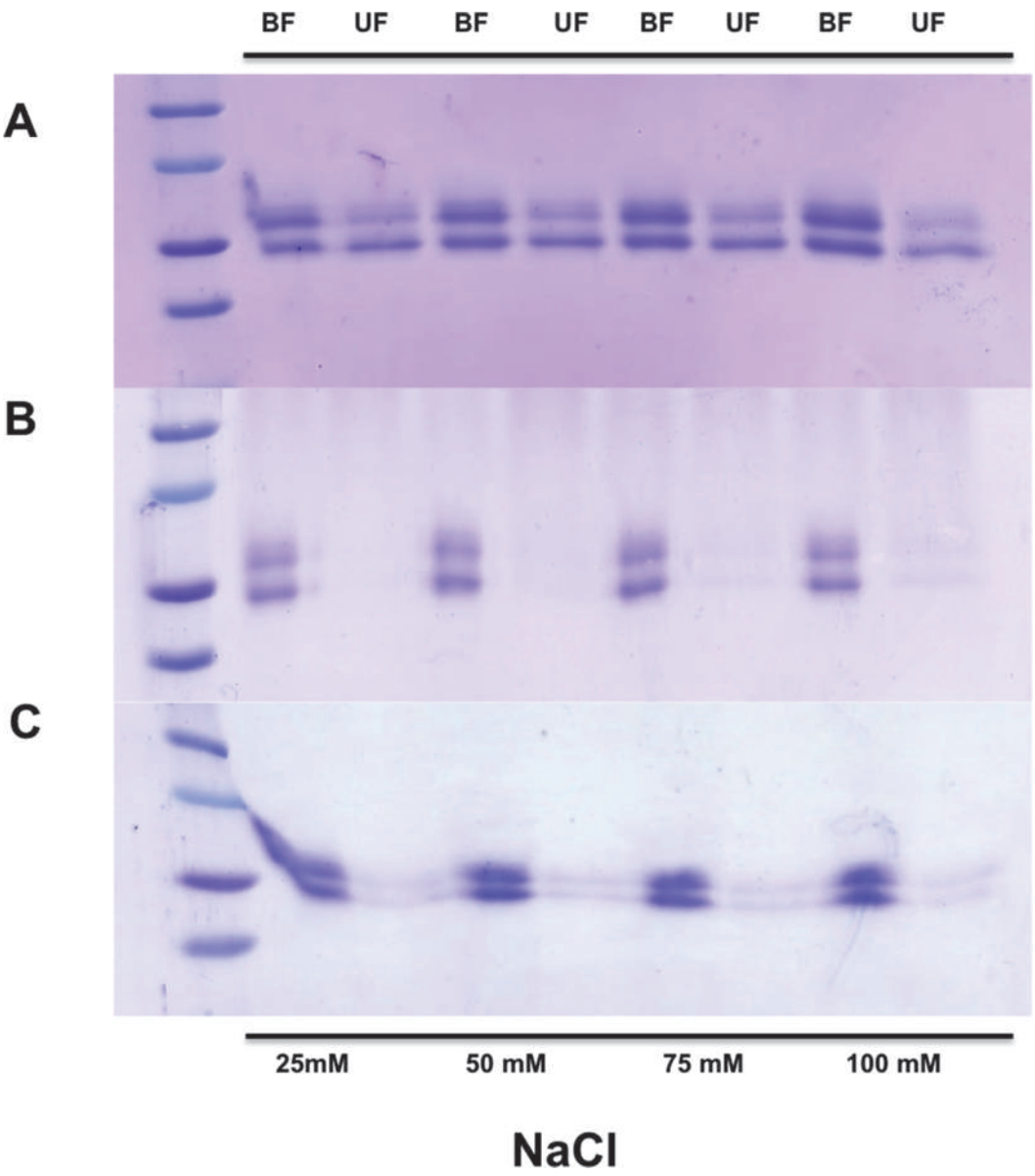


Fig 6. Neutral salt addition effect on ScExlx1 binding to plant cell wall polysaccharides and chitin. Nacl effect on ScExlx1 binding to: A) avicel; B) chitin; C) xylan.

doi:10.1371/journal.pone.0122296.g006

Discussion

Schizophyllum commune is one of the most commonly found fungi and can be isolated from nearly all continents in the world, except for Antarctica [28]. Although *S. commune* has been found causing illness in humans, animals and trees, its ecological niche is to adopt a saprobic lifestyle by causing white rot [28–31]. After *S. commune* genome was sequenced, its enormous potential as a microbial protein factory for plant cell wall material deconstruction was revealed, and in this context we studied the protein ScExlx1 discovered trough *S. commune* genome analysis.

Evaluation of ScExlx1 effect on cotton fibers and avicel was studied. Previous reports have described the effect of expansin-like proteins over filter paper, avicel, PASC (Phosphoric

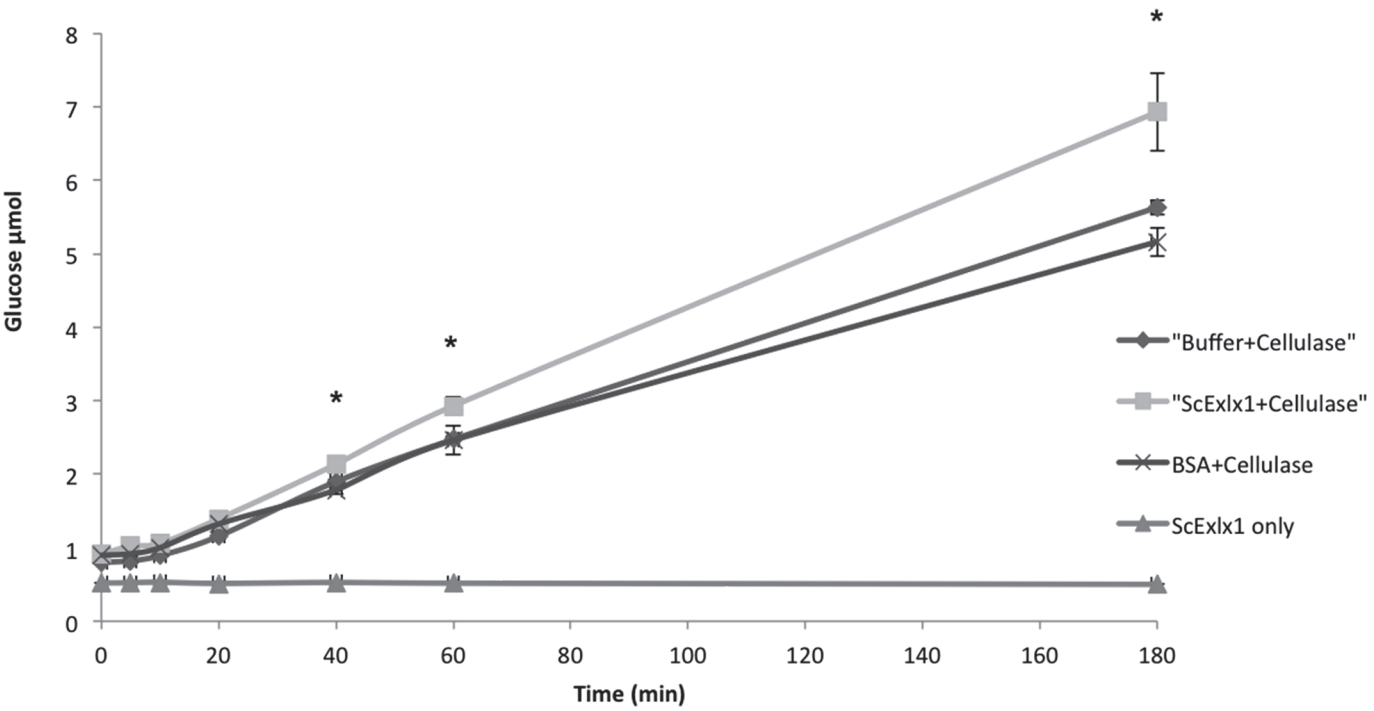


Fig 7. Effect of ScExlx1 on the enzymatic hydrolysis of cellulose. Mercerized cotton fibers (1 mg) were incubated with 20 µg of ScExlx1, 20 µg of BSA or sodium acetate buffer (pH 5) for 72 h at 25°C. After incubation, temperature was raised to 50°C and cellulase cocktail from *T. reesei* was added (0.25 U) in a 3 h experiment. Reducing sugars were quantified by DNS method and compared with a glucose standard curve. Experiments were performed in triplicate, and the data points and error bars indicate means ± standard deviations. *Statistical differences between treatments ($p<0.05$) at each point.

doi:10.1371/journal.pone.0122296.g007

acid-swollen cellulose) and cotton fibers [7,8,32]. As a member of the expansin superfamily, ScExlx1 caused a modification in crystalline cellulose anatomy, proposing that this effect can serve to enhance cellulose hydrolysis by *S. commune* when living as a white rot. In the same way, the "bubble" effect generated by loosenin [8], swollenin [7] and ScExlx1 (in this work), have been proposed to aid in the amorphogenesis step during lignocellulose degradation with hydrolytic enzymes.

An interesting finding was that sequence analysis showed a non-polar to slightly polar change in the amino acids responsible for polysaccharide binding at domain 2 (Fig. 1). This fact, did not alter the ability of ScExlx1 to bind cellulose and birchwood xylan as previously reported for plant pathogenic (PcExl1), soil (BsExlx1) and marine bacterial (HcExlx2) expansins [4,6,11]. Also, some researchers have studied the effect of NaCl over protein-protein and protein-ligand interactions [4,6,33,34]. Binding of ScExlx1 to avicel was not affected by NaCl addition even at concentrations of 0.5 M, as showed by densitometry and statistical analysis (See Fig. 6 and S1 Table), proposing that salt interaction with the ScExlx1-cellulose complex is different to that reported for other microbial expansins [4,6,11], where salt addition affects positively (PcExl1) and negatively (BsExlx1) the binding capacity of the expansin protein. With regard to the binding capability on birchwood xylan, a similar behavior to that observed with avicel was shown. It is possible that hydrophobic interactions are more important in the binding of ScExlx1 to its substrates than electrostatic bonds. Perhaps higher NaCl concentrations need to be evaluated in order to detect any effect (>0.5 M). Another possibility is that using lower substrate concentrations could allow binding changes easier to see. Additionally, HcExlx2 has the ability to bind in a stronger way to xylan from oat spelts than to cellulose [13], suggesting that bacterial expansins may have different targets in the plant cell wall architecture,

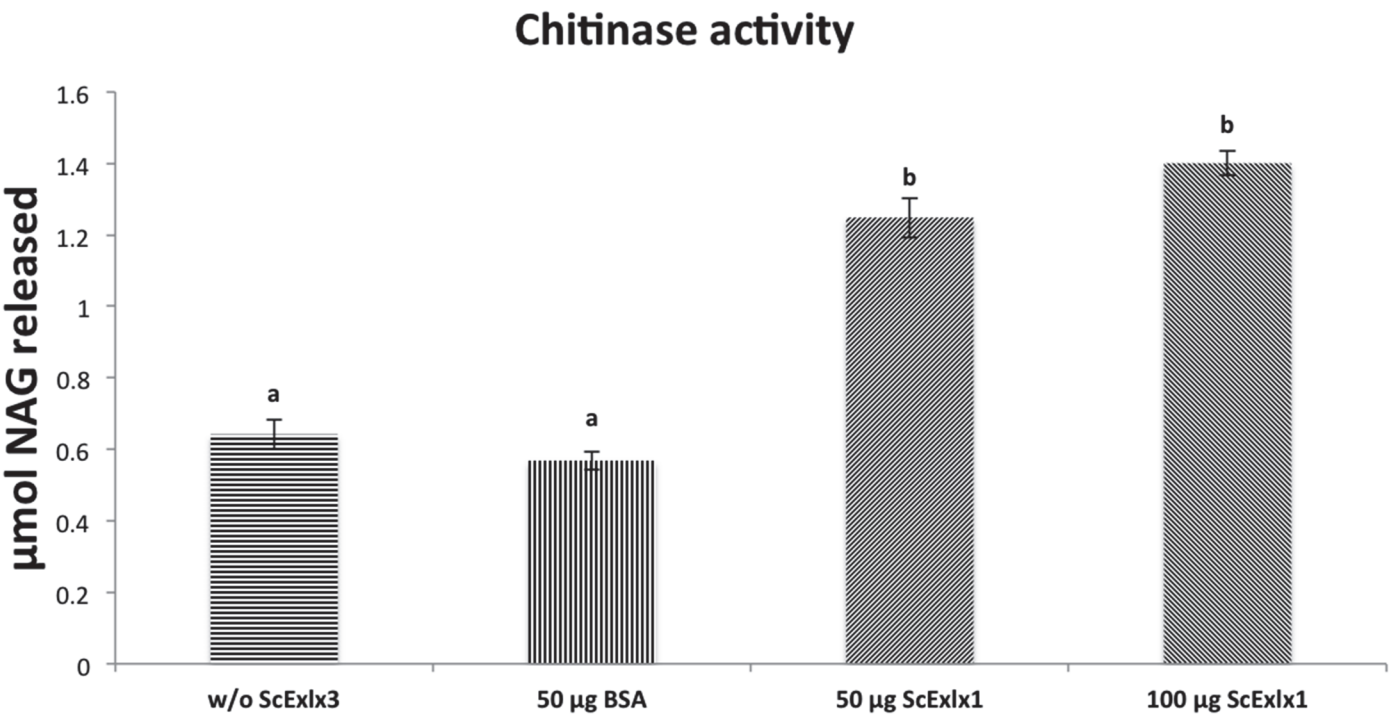


Fig 8. ScExlx1 is a chitin active protein that enhances chitin hydrolysis. Chitin from shrimp shells (5 mg) was incubated with 50 and 100 µg of ScExlx1, 50 µg of BSA, or sodium phosphate buffer 100 mM pH 7 for 24 h at 25°C. After incubation, temperature was increased to 37°C and chitinase from *S. griseus* (0.25 U) was added. After 2 hours of incubation, released N-acetylglucosamine was measured by DNS method and compared with a standard curve. Experiments were performed in triplicate, and the data points and error bars indicate means ± standard deviations. Letters indicate statistical differences in each treatment ($p<0.05$).

doi:10.1371/journal.pone.0122296.g008

since BsExlx1 binds better to whole plant cell wall and PcExl1 binds better to cellulose [4,6,35]. Besides these two plant cell wall components, we evaluated the binding capability of ScExlx1 to the fungal cell wall polysaccharide chitin, insomuch as proteins like swollenin, an expansin-like protein, loosenin containing only D1 from canonical expansins, ZmEXPB (a plant expansin), BsExlx1 and HcExlx2 have displayed better binding capacities to chitin than to cellulose. In this context, densitometry experiments showed only a small increment in bound protein quantity when comparing 0 M and 0.5 M of NaCl under the experimental conditions evaluated (S1 Table). In general, these results suggest that different interactions can be displayed by fungal and bacterial expansins depending on the polysaccharide matrix, and we can also suggest (but not conclude) that ScExlx1-polysaccharide interaction is not of ionic nature. Nonetheless, although in the case of crystalline cellulose and xylan different behaviors with regard to binding properties in the presence of NaCl are exhibited depending of the expansin evaluated, we would have expected an ionic character interaction at least between ScExlx1-chitin and ScExlx1-xylan, since amino acids Tyr-161, Tyr-162 and Tyr-195 on D2 that have been demonstrated to participate in polysaccharide binding [27] in microbial expansins could serve to contribute in the hydrogen bonding through its side chain with both polysaccharides. Perhaps, different experimental conditions need to be evaluated in order to detect a more pronounced effect. It is worth noting that in all cases both, the putative glycosylated and deglycosylated forms are able to bind the polysaccharides.

It is interesting to note that in the biotechnological field, this salt-tolerant characteristic of ScExlx1 could be useful in several ways, for example: 1) the need of novel proteins tailored for the ionic liquid process technology used for the production of advanced cellulosic biofuels, and

2) addition of acetate have shown to increase the hydrolytic capacity of some chitin deacetylases [36], so ScExlx1 could help to deacetylate the chitin polysaccharide in the presence of this enzymes in an halophile environment in order to produce chitosan, a biotechnological product with high added value. This salt-tolerant property makes ScExlx1 an ideal candidate for further research in a structure-function basis and likely industrial applications.

Regarding to enzymatic hydrolysis, expansins and expansin-like proteins have been evaluated with the aim of enhancing lignocellulose hydrolysis and reducing enzyme loadings [6,8,11,12,23,32,37–41]. ScExlx1 displayed no synergistic effect when used together cellulase cocktail from *T*. *reesei* (data not shown) what is consistent with results obtained by Olarte et al. [6] and opposite to results reported by Lee et al. [11,12]. Similarly, a slightly but significant increment in cellulose hydrolysis was detected when ScExlx1 pre-treated cotton fibers were exposed to cellulase cocktail either after 3 or 48 h. With the data obtained in this work, we demonstrated that ScExlx1 exhibits a cellulolytic enhancing effect when used previously with cellulase at low enzyme loadings.

Finally, this is the first report that shows the effect of an expansin protein on the polysaccharide chitin. This fact is very significant by two main reasons: 1) the enhancing effect of chitin hydrolysis by chitinase can be exploited in several ways as this polymer has gained huge scientific interest due to it is numerous biotechnological and medical applications [42,43]; 2) there is the possibility that ScExlx1 can be acting as a fungal cell wall remodeling factor, allowing the fungus to grow in a similar manner than that of plant expansins. Notwithstanding, it remains unclear the effect generated over chitin polymer by ScExlx1; experiments to obtain this information and analysis of physiological role of ScExlx1 in *S*. *commune* are being carried out.

Supporting Information

S1 Fig. Effect of ScExlx1 on the enzymatic hydrolysis of cellulose in a 48 h experiment. Mercerized cotton fibers (1 mg) were incubated with 20 μg of ScExlx1, 20 μg of BSA or sodium acetate buffer (pH 5) for 72 h at 25°C. After incubation, temperature was raised to 50°C and cellulase cocktail from *T*. *reesei* was added (0.25 U) in a 48 h experiment. Reducing sugars were quantified by DNS method and compared with a glucose standard curve. Experiments were performed in triplicate, and the data points and error bars indicate means ± standard deviations.
(TIF)

S1 Table. Effect of NaCl over binding to polysaccharides. Experiments were performed in triplicate, and different letters indicate different statistical orders.
(DOCX)

Acknowledgments

We are very grateful to Laura Ines Cuervo Soto for her technical support during the experiment performance. We are also grateful to Dr. Marcela Ayala for providing the PNGase-F used in de-glycosylation experiments.

Author Contributions

Conceived and designed the experiments: OETH RABG MRSC MMIC KAN JLFM. Performed the experiments: OETH RABG MRSC. Analyzed the data: OETH RABG MRSC MMIC KAN JLFM. Contributed reagents/materials/analysis tools: OETH RABG MRSC MMIC KAN JLFM. Wrote the paper: OETH RABG KAN JLFM.

References

1. Sampedro J, Cosgrove DJ. The expansin superfamily. Genome Biol. 2005 Jan; 6(12):242. PMID: 16356276
2. Cosgrove DJ. Loosening of plant cell walls by expansins. Nature. 2000 Sep 21; 407(6802):321–6. PMID: 11014181
3. Yennawar NH, Li L-C, Dudzinski DM, Tabuchi A, Cosgrove DJ. Crystal structure and activities of EXPB1 (Zea m 1), a beta-expansin and group-1 pollen allergen from maize. Proc Natl Acad Sci U S A. 2006 Oct 3; 103(40):14664–71. PMID: 16984999
4. Kerff F, Amoroso A, Herman R, Sauvage E, Petrella S, Filée P, et al. Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. Proc Natl Acad Sci U S A. 2008 Nov 4; 105(44):16876–81. doi: 10.1073/pnas.0809382105 PMID: 18971341
5. Kende H, Bradford K, Brummell D, Cho H-T, Cosgrove D, Fleming A, et al. Nomenclature for members of the expansin superfamily of genes and proteins. Plant Mol Biol. 2004 May; 55(3):311–4. PMID: 15604683
6. Olarte-Lozano M, Mendoza-Nuñez M, Pastor N, Segovia L, Folch-Mallol J, Martínez-Anaya C. PcExl1 a novel acid expansin-like protein from the plant pathogen *Pectobacterium carotovorum*, binds cell walls differently to BsEXLX1. PLoS One. 2014 Jan; 9(4):e95638. doi: 10.1371/journal.pone.0095638 PMID: 24755657
7. Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, Nyyssönen 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 Sep 20; 269(17):4202–11. PMID: 12199698
8. Quiroz-Castañeda RE, Martínez-Anaya C, Cuervo-Soto LI, Segovia L, Folch-Mallol JL. Loosenin, a novel protein with cellulose-disrupting activity from *Bjerkandera adusta*. Microb Cell Fact. 2011 Jan; 10:8. doi: 10.1186/1475-2859-10-8 PMID: 21314954
9. Cosgrove DJ, Bedinger P, Durachko DM. Group I allergens of grass pollen as cell wall-loosening agents. Proc Natl Acad Sci U S A. 1997 Jun 10; 94(12):6559–64. PMID: 9177257
10. Pazzagli L, Pantera B, Carresi L, Zoppi C, Pertinhez TA, Spisni A, et al. Cerato-platanin, the first member of a new fungal protein family: cloning, expression, and characterization. Cell Biochem Biophys. 2006; 44:512–21. PMID: 16679539
11. Lee HJ, Lee S, Ko H-J, Kim KH, Choi I-G. An expansin-like protein from *Hahella chejuensis* binds cellulose and enhances cellulase activity. Mol Cells. 2010 Apr; 29(4):379–85. doi: 10.1007/s10059-010-0033-z PMID: 20213317
12. Georgelis N, Nikolaidis N, Cosgrove DJ. Biochemical analysis of expansin-like proteins from microbes. Carbohydr Polym. Elsevier Ltd.; 2014 Jan 16; 100:17–23. doi: 10.1016/j.carbpol.2013.04.094 PMID: 24188833
13. Lee HJ, Kim IJ, Kim JF, Choi I-G, Kim KH. An expansin from the marine bacterium *Hahella chejuensis* acts synergistically with xylanase and enhances xylan hydrolysis. Bioresour Technol. Elsevier Ltd; 2013 Dec; 149:516–9. doi: 10.1016/j.biortech.2013.09.086 PMID: 24140897
14. Kim IJ, Ko H-J, Kim T-W, Nam KH, Choi I-G, Kim KH. Binding characteristics of a bacterial expansin (BsEXLX1) for various types of pretreated lignocellulose. Appl Microbiol Biotechnol. 2013 Jun; 97 (12):5381–8. doi: 10.1007/s00253-012-4412-6 PMID: 23053073
15. Hamby SE, Hirst JD. Prediction of glycosylation sites using random forests. BMC Bioinformatics. 2008; 9:500. doi: 10.1186/1471-2105-9-500 PMID: 19038042
16. Ferrè F, Clote P. DiANNA: A web server for disulfide connectivity prediction. Nucleic Acids Res. 2005; 33.
17. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 2011. p. 785–6.
18. Dereeper A, Audic S, Claverie J-M, Blanc G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol. 2010; 10:8. doi: 10.1186/1471-2148-10-8 PMID: 20067610
19. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008; 36.
20. Zhang Y, Skolnick J. Segment assembly, structure alignment and iterative simulation in protein structure prediction. 2013;2–5.
21. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976 May 7; 72:248–54. PMID: 942051
22. Quiroz-Castañeda RE, Balcázar-López E, Dantán-González E, Martinez A, Folch-Mallol J, Martínez Anaya C. Characterization of cellulolytic activities of *Bjerkandera adusta* and *Pycnoporus sanguineus* on solid wheat straw medium. Electron J Biotechnol. 2009 Oct 15; 12(4).

23. Chen X, Ishida N, Todaka N, Nakamura R, Maruyama J, Takahashi H, et al. Promotion of efficient Saccharification of crystalline cellulose by *Aspergillus fumigatus* Swo1. Appl Environ Microbiol. 2010 Apr; 76(8):2556–61. doi: 10.1128/AEM.02499-09 PMID: 20173066

24. Hoaglin D, Welsch R. The Hat Matrix in Regression and ANOVA. Am Stat. 1978; 32:17–22.

25. Simpson FR. Statistical tables for biological, agricultural and medical research. Eugen Rev. 1939; 30:298–9.

26. Pastor N, Sonia D, Ernesto P, Segovia L, Mart C. Electrostatic analysis of bacterial expansins. 2014; (November):1–9.

27. Georgelis N, Tabuchi A, Nikolaidis N, Cosgrove DJ. Structure-function analysis of the bacterial expansin EXLX1. J Biol Chem. 2011 May 13; 286(19):16814–23. doi: 10.1074/jbc.M111.225037 PMID: 21454649

28. Ohm R a, de Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE, et al. Genome sequence of the model mushroom *Schizophyllum commune*. Nat Biotechnol. Nature Publishing Group; 2010 Sep; 28(9):957–63. doi: 10.1038/nbt.1643 PMID: 20622885

29. Kano R, Oomae S, Nakano Y, Minami T, Sukikara M, Nakayama T, et al. First Report on *Schizophyllum commune* from a Dog. 2002; 40(9):3535–7. PMID: 12202614

30. Tanaka H, Takizawa K, Baba O, Maeda T, Fukushima K, Shinya K, et al. Basidiomycosis: *Schizophyllum commune* osteomyelitis in a dog. J Vet Med Sci. 2008 Dec; 70(11):1257–9. PMID: 19057147

31. Premamalini T, Ambujavalli BT, Anitha S, Somu L, Kindo AJ. S*chizophyllum commune* a causative agent of fungal sinusitis: a case report. Case Rep Infect Dis. 2011 Jan; 2011:821259. doi: 10.1155/2011/821259 PMID: 22567479

32. Jäger G, Girfoglio M, Dollo F, Rinaldi R, Bongard H, Commandeur U, et al. How recombinant swollenin from *Kluyveromyces lactis* affects cellulosic substrates and accelerates their hydrolysis. Biotechnol Biofuels. BioMed Central Ltd; 2011 Jan; 4(1):33. doi: 10.1186/1754-6834-4-33 PMID: 21943248

33. Von Hippel PH, Schleich T, Timasheff SN, Fasman GD. Effects of neutral salts on the structure and conformational stability of macromolecules in solution. Structure and Stability of Biological Macromolecules. 1969. p. 417–574.

34. Dumetz AC, Snellinger-O'brien AM, Kaler EW, Lenhoff AM. Patterns of protein protein interactions in salt solutions and implications for protein crystallization. Protein Sci. 2007; 16:1867–77. PMID: 17766383

35. Kim IJ, Ko H-J, Kim T-W, Choi I-G, Kim KH. Characteristics of the binding of a bacterial expansin (BsEXLX1) to microcrystalline cellulose. Biotechnol Bioeng. 2013 Feb; 110(2):401–7. doi: 10.1002/bit.24719 PMID: 22949138

36. Zhao Y, Park R-D, Muzzarelli R a a. Chitin deacetylases: properties and applications. Mar Drugs. 2010 Jan; 8(1):24–46. doi: 10.3390/md8010024 PMID: 20161969

37. Gourlay K, Hu J, Arantes V, Andberg M, Saloheimo M, Penttilä M, et al. Swollenin aids in the amorphogenesis step during the enzymatic hydrolysis of pretreated biomass. Bioresour Technol. Elsevier Ltd; 2013; 142:498–503.

38. Yao Q, Sun T-T, Liu W-F, Chen G-J. Gene Cloning and Heterologous Expression of a Novel Endoglucanase, Swollenin, from *Trichoderma pseudokoningii* S38. Biosci Biotechnol Biochem. 2008; 72 (11):2799–805. PMID: 18997434

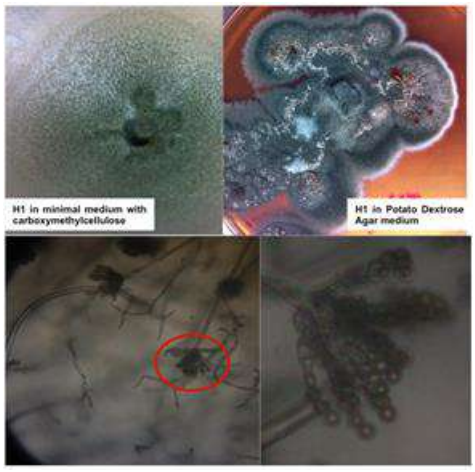
39. Wang M, Cai J, Huang L, Lv Z, Zhang Y, Xu Z. High-level expression and efficient purification of bioactive swollenin in *Aspergillus oryzae*. Appl Biochem Biotechnol. 2010 Nov; 162(7):2027–36. doi: 10.1007/s12010-010-8978-0 PMID: 20446057

40. Kang K, Wang S, Lai G, Liu G, Xing M. Characterization of a novel swollenin from *Penicillium oxalicum* in facilitating enzymatic saccharification of cellulose. BMC Biotechnol. 2013 May 20; 13(1):42.

41. Kim ES, Lee HJ, Bang W-G, Choi I-G, Kim KH. Functional characterization of a bacterial expansin from *Bacillus subtilis* for enhanced enzymatic hydrolysis of cellulose. Biotechnol Bioeng. 2009; 102:1342–53. doi: 10.1002/bit.22193 PMID: 19058186

42. Park BK, Kim M-M. Applications of chitin and its derivatives in biological medicine. Int J Mol Sci. 2010 Jan; 11(12):5152–64. doi: 10.3390/ijms11125152 PMID: 21614199

43. Dutta PK, Dutta J, Tripathi VS. Chitin and chitosan: Chemistry, properties and applications. 2004; 63 (January):20–31.



Capítulo V.

Caracterización de las actividades lignocelulolíticas de una cepa halófila moderada de *Aspergillus caesiellus* aislada de bagazo de caña de azúcar en fermentación.

En los capítulos anteriores hemos analizado cómo las herramientas metagenómicas y genómicas resultan útiles para la descripción de comunidades microbianas involucradas en la degradación de lignocelulosa, y a la vez, cómo pueden ser usadas para el estudio de nuevas proteínas. Sin embargo, no sólo los métodos *ómicos* son considerados en la búsqueda de proteínas con potenciales usos en biorrefinerías. También las prospecciones *in vitro* de cultivos son una poderosa herramienta ecológica para el estudio de novedosos perfiles enzimáticos en el contexto de un sistema microbiano con posibles aplicaciones posteriores en procesos biotecnológicos. Una de las ventajas que aún avala este tipo de enfoque (Ecología Microbiana Clásica) sobre los estudios metagenómicos y genómicos, es la caracterización de las actividades enzimáticas en un escenario fisiológico determinado, además de los inconvenientes analizados para los escrutinios de metagenomas basados en actividad y por PCR (ver Capítulo I).

Considerando los elementos anteriores, y que el bagazo de caña de azúcar es un sustrato exclusivamente lignocelulósico, estudiamos las comunidades microbianas celulolíticas presentes en una muestra de bagazo en fermentación suplementada con 2M NaCl. Las enzimas con potencialidades hidrolíticas en condiciones hipersalinas resultan muy atractivas para las biorrefinerías de acuerdo a los márgenes de salinidad en que estas operan. Así, la búsqueda de celulasas, xilanasas y en general, de enzimas ligninolíticas a partir de organismos halófilos se convierte en un área de interés para la deconstrucción de lignocelulosa.

Se realizó el estudio de una cepa halófila lignocelulolítica de *Aspergillus caesiellus* y se caracterizó su crecimiento en condiciones salinas e hipersalinas, así como su posibilidad de colonizar diferentes sustratos lignocelulósicos naturales y los perfiles de enzimas ligninolíticas expresados diferencialmente en ellos. El estudio de hongos halófilos ha sido menos extensivo comparado con los relacionados con bacterias halófilas. Este es el primer reporte que informa una cepa halófila de *Aspergillus caesiellus* con posibles usos en deconstrucción de lignocelulosa: *"Characterization of lignocellulolytic activities from a moderate halophile strain of Aspergillus caesiellus isolated from a sugarcane bagasse fermentation"*. Este trabajo demuestra que la caracterización de aislamientos microbianos y con ello

los enfoques de la Microbiología Pasteuriana, no pasan de moda aún con las ventajas de los métodos *ómicos*.

Aspergillus caesiellus H1 se ha convertido en una *"cepa modelo"* de gran interés para nuestro laboratorio. En su primera caracterización se evidenció que existe una expresión diferencial de celulasas y xilanasas dependiendo del sustrato lignocelulósico que coloniza, siendo la paja de trigo el residuo agrícola que mejor coloniza y que más estimula la producción de enzimas como celulasas, xilanasas, peroxidasas y esterasas. De acuerdo con estos resultados, y con nuestro interés de acceder a sus recursos genéticos, actualmente estamos estudiando el transcriptoma de este ascomiceto crecido sobre paja de trigo a diferentes concentraciones de NaCl. El análisis de su transcriptoma nos permitirá por un lado, caracterizar los niveles diferenciales de transcritos de los genes codificando enzimas lignocelulósicas en diferentes concentraciones de NaCl, y por otro, describir posibles mecanismos fisiológicos relacionado con la homeostasis de este hongo en condiciones hipersalinas.

OPEN ACCESS Freely available online

PLOS ONE

Characterization of Lignocellulolytic Activities from a Moderate Halophile Strain of *Aspergillus caesiellus* Isolated from a Sugarcane Bagasse Fermentation

Ramón Alberto Batista-García[1,2☯], Edgar Balcázar-López[1,2☯], Estefan Miranda-Miranda[3], Ayixón Sánchez-Reyes[1,2], Laura Cuervo-Soto[1,2], Denise Aceves-Zamudio[2], Karina Atriztán-Hernández[4], Catalina Morales-Herrera[2], Rocío Rodríguez-Hernández[2], Jorge Folch-Mallol[2]*

1 Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, Mexico, **2** Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México, **3** Centro Nacional de Investigación Disciplinaria en Parasitología Veterinaria, Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Cuernavaca, Morelos, Mexico, **4** Facultad de Ciencias Biológicas, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, Mexico

Abstract

A moderate halophile and thermotolerant fungal strain was isolated from a sugarcane bagasse fermentation in the presence of 2 M NaCl that was set in the laboratory. This strain was identified by polyphasic criteria as *Aspergillus caesiellus*. The fungus showed an optimal growth rate in media containing 1 M NaCl at 28°C and could grow in media added with up to 2 M NaCl. This strain was able to grow at 37 and 42°C, with or without NaCl. *A. caesiellus* H1 produced cellulases, xylanases, manganese peroxidase (MnP) and esterases. No laccase activity was detected in the conditions we tested. The cellulase activity was thermostable, halostable, and no differential expression of cellulases was observed in media with different salt concentrations. However, differential band patterns for cellulase and xylanase activities were detected in zymograms when the fungus was grown in different lignocellulosic substrates such as wheat straw, maize stover, agave fibres, sugarcane bagasse and sawdust. Optimal temperature and pH were similar to other cellulases previously described. These results support the potential of this fungus to degrade lignocellulosic materials and its possible use in biotechnological applications.

Citation: Batista-García RA, Balcázar-López E, Miranda-Miranda E, Sánchez-Reyes A, Cuervo-Soto L, et al. (2014) Characterization of Lignocellulolytic Activities from a Moderate Halophile Strain of *Aspergillus caesiellus* Isolated from a Sugarcane Bagasse Fermentation. PLoS ONE 9(8): e105893. doi:10.1371/journal.pone.0105893

Editor: Melanie R. Mormile, Missouri University of Science and Technology, United States of America

Received April 20, 2014; **Accepted** July 25, 2014; **Published** August 27, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: The work was partially funded by grant CB-153789-Q from the National Council for Science and Technology (CONACyT) from the Mexican government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Jorge L. Folch-Mallol is currently an Academic Editor for PLOS ONE but this does not alter the authors' adherence to PLOS ONE Editorial policies and criteria.

* Email: jordi@uaem.mx

☯ These authors contributed equally to this work.

Introduction

For many years, extremophile microorganisms were described exclusively in the Archae and Eubacteria domains [1]. These peculiar organisms have been found and studied in environments with very extreme conditions such as temperature, pressure, pH, salinity and radiation, among others. Physiological and metabolic studies on extremophile microorganisms derived in the development of biotechnological applications with incidence in the agricultural, pharmacological and environmental industries [2–6]. The description of microbial diversity in extremophile ecosystems has been of great interest for a long time until our days. However, the report of eukaryotic extremophiles took many years.

In particular, environments with high concentrations of salt, called saline and hypersaline environments, have been attractive ecosystems for studying the structures of microbial communities that inhabit them [7–9], since sodium chloride has been considered an effective microbicide. For example, treating food with salt is an effective method of preserving meat since ancient times [10]. The isolation and characterization of microorganisms in habitats with over 1 M NaCl is important for the identification of metabolites and/or robust proteins with potential industrial applications, and to understand the cellular physiology, molecular biology and biochemistry that support the survival of these organisms under extreme conditions [11–13].

Numerous studies have described halophilic bacterial genera in hypersaline ecosystems [5], [8], [14–21]. However the first report of a fungus isolated from a solar saltern appeared in the year 2000 [22]. Since then, many studies on biodiversity and physiology have reported the characterization of halophilic fungi present in saline and hypersaline ecosystems. Many species of Ascomycetes, and some Basidiomycetes, have been described with the ability to grow in these environments [22], [23]. Currently, the presence of fungi

in the structure of microbial communities that compose some ecosystems with extreme salinity conditions has been described [24]. Until 2009, only 10 orders of fungi known to tolerate low water activity (a_w) due to high salt concentrations have been reported [22].

Among the most studied halophilic fungal genera are *Cladosporium* [22], [23], [25], *Wallemia*, *Scopulariopsis*, *Alternaria* [24], [26] and some species of *Aspergillus* and *Penicillium* [23]. Particularly in the *Aspergillus* genera, *A. niger, A. sydowwi, A. flavus, A. tubingensis* and *A. versicolor* have been isolated and described as halotorelant and halophile fungi as part of the hypersaline environments [22]. The study of these fungi has allowed the characterization of halophilic enzymes with interesting properties [27]. For example, cellulases and xylanases have been described from halophilic microorganisms, filamentous fungi included, with interesting biochemical properties such as activity in the presence of high salt concentrations, at acidic pH, or in ionic liquids [11], [13], [28–33]. In general, halophilic microorganisms can be isolated from the sea, saline or hypersaline lakes, solar salterns, and salted foods, among other habitats with high concentrations of salt.

It was assumed for a long time that extremophile microorganisms were strictly growing in extreme conditions; however this assumption was proved to be false because there are many examples of extremophiles isolated from non-extreme environments [34–36]. In this way, some reports demonstrate that halophilic and/or halotolerant microorganisms are not restricted to saline or hypersaline habitats and can be found almost everywhere in non-saline environments [37–46]. The general principles of microbial ecology, the physiological plasticity and metabolic versatility of the microorganisms do not impede the isolation of halophile and/or halotolerant organisms from non-saline environments.

The isolation of halophile and/or halotolerant fungi able to grow in lignocellulosic materials is a very interesting source for the search of industrially useful enzymes, potentially capable of producing fermentable sugars from agricultural wastes. One of the major limitations of using enzymes in industry is obtaining robust biocatalysts capable of operating under rigorous industrial conditions. Because of this demand, the study of halophilic and halotolerant microorganisms is an appropriate strategy for the characterization of enzymes such as cellulases, esterases, lipases and xylanases with potential application in biorefineries. Sugarcane bagasse is considered a recalcitrant, high cellulose content agroindustrial waste. The microbial communities growing on sugarcane bagasse, especially filamentous fungi must have efficient enzymes to obtain energy and carbon from lignocellulose degradation. The aims of this work were to characterize a filamentous fungus isolated from sugarcane bagasse with the potential to grow in concentrations of sodium chloride higher than 1 M, and to analyse the lignocellulolytic activities it produces.

Materials and Methods

Isolation and preservation of microorganisms

Three grams of non-sterile sugarcane bagasse (as the main carbon source and from where the microorganisms were isolated) were mixed with 2.5 g of sterile soil (as a source of organic matter) into 1000 mL Erlenmeyer flasks containing 250 mL of Vogels medium as described in [47] and a final concentration of 2 M NaCl. Fermentation was performed at 25°C and 150 rpm for 30 days. Primary isolation of cellulolytic microorganisms was done taking one mL of the sugar bagasse culture and serial dilutions were performed up to 10^{-10}. Two hundred µL of each dilution were inoculated in Petri plates containing Vogels medium supplemented with 2% carboxymethylcellulose (CMC, Sigma Catalogue No. C5678) and 0.5 M NaCl. The cultures were incubated for 10 days at 30°C and were observed daily for secondary isolation of cellulolytic bacteria and/or fungi. A fungus was isolated in Potato Dextrose Agar (PDA) and Saboraud agar media (both from DIFCO). This fungus was stored at 4°C in saline solution (0.5% NaCl) supplemented with glycerol (20%). Isolates were performed in triplicate to achieve representativeness of the obtained microbial populations.

Identification of a fungal moderate halophilic strain

Mycelium from a fungus growing on PDA plates was collected after 10-days for genomic DNA isolation according to a previously reported method [48]. For identification, we analysed molecular markers previously described to be distinctive to filamentous fungi. A fragment of the 18S ribosomal DNA was amplified by PCR using primers nu-SSU-0817 and nu-SSU-1536 as described by Bornenman and Hartin [49]. Also, regions of the 28S large subunit RNA gene (D1–D2) and internal transcribed spacers 1 (ITS1 region) were amplified. These regions have been particularly useful for the molecular identification of fungi. The primers and conditions used for these PCR reactions have been previously described by Peterson [50] and Hinrikson *et al.* [51], respectively.

Amplicons were analysed in 1% agarose gel electrophoresis in 1x TBE buffer [52] and purified from the agarose gels using a commercial gel extraction kit (Fermentas Catalogue No. K0513) and sequenced in both directions using the same primers used for the amplification. Sanger sequencing was performed at the Sequencing Unit of the Instituto de Biotecnología of the Universidad Nacional Autónoma de México.

The sequences were analysed using the website of the National Centre for Biotechnology Information (NCBI) (www.ncbi.nih.gov). BLAST search was performed to determine similar sequences. Phylogeny studies allowed defining the relationship of these sequences with those obtained in the BLAST analysis. Phylogenetic analysis was performed online with the server Phylogeny.fr (www.phylogeny.fr). This platform considers various bioinformatics programs to reconstruct a robust phylogenetic tree from a set of sequences. These tools allowed identification of the fungal strain considering molecular criteria.

Micromorphological identification was performed to complete a polyphasic vision for the microbial identification. Size, shape and grouping of conidiophores, morphological aspects of the colony, and the fungal hyphae were analysed. The aspect of the colony on different media (PDA, Saboraud agar and Malt extract agar (MEA, from DIFCO) was also considered.

Growth Rate Determination

Specific growth rate of the fungal colony (expressed as mm/day) was determined as follows. Plugs of 7 mm of diameter obtained from a fungal pre-culture in Vogel's medium supplemented with 2% CMC were inoculated in the same medium at different temperatures (28, 37, and 42°C) or supplemented with NaCl (0, 0.5, 1, 1.5, 2 or 3 M, final concentration). The diameter of the colony was measured every 24 hours for 10 days. Experiments were performed in triplicate for subsequent statistical analysis of data.

Solid-state Fermentation

Solid-state fermentations were performed using the following autoclaved substrates: agave fibre (*Agave fourcroydes*), sugarcane bagasse (*Saccharum officinarum*), maize stover (*Zea mays*), wheat straw (*Triticum aestivum*), and pine sawdust (*Pinus sylvestris*).

Erlenmeyer flasks of 500 mL including 2 g of each substrate were inoculated with two plugs of 7 mm of diameter of the fungal strain previously grown on PDA plates. Humidity in the system was maintained by adding 20% (w/v) of distilled water to the solids. Fermentation was allowed to take place at 28°C for 10 days. Subsequently, soluble fermentation products were collected in 10 mL in 60 mM citrate buffer pH 5.

Enzymatic activity determinations

Cellulase activity was determined qualitatively and quantitatively. For qualitative determinations, seven-day cultures of the fungus grown on agar Vogel's medium supplemented with 2% CMC and NaCl (0, 0.5, 1, 1.5, 2 M, final concentration) at 28°C were set up. Petri dishes with the colonies were then inundated with approximately 15 mL of Congo red (1% diluted in distilled water) for 10–15 minutes. Subsequently, these dishes were washed three times with approximately 15 mL of a 1 M NaCl solution. Discoloration halos around the colony indicated the degradation of cellulose due to the production of cellulases [53], [54]. Determinations were performed in triplicate.

For quantitative determination of cellulase activity, plugs of 7 mm of diameter from pre-cultures of the fungus grown in Vogels medium with 2% CMC were inoculated in 500 mL Erlenmeyer flasks with 100 mL of the same medium plus NaCl (0, 0.5, 1, 1.5, 2 M). The flasks were incubated for 9 days at 28°C and 150 rpm. Enzymatic activity and protein concentration were determined every 24 h from 2 mL of the supernatants of these liquid cultures. Enzymatic activity was assessed by the production of reducing sugars from polymeric substrates using the 3,5-dinitrosalysilic acid (DNS) assay described by Miller [55]. Protein concentration was determined by the Lowry method [56]. Enzymatic specific activity is expressed as IU/mg protein.

For cellulases activity measurements, CMC (2%) dissolved in 50 mM citrate buffer pH 5 was used as a substrate. The enzymatic reaction contained 200 µL of supernatant, 300 µL of 50 mM citrate buffer pH 5 and 500 µL of substrate solution. The reaction mixtures were incubated at 50°C for 30 minutes. Briefly, 50 µL aliquots were taken every 5 minutes (after adding the supernatant to the reaction mixture) up to 45 minutes and then mixed with 50 µL of a DNS solution, boiled for 5 minutes and immediately cooled on ice for 5 minutes. Finally 500 µL of water were added and absorbance measured at λ 540 nm in a spectrophotometer (BioMate, ThermoSpectronic). Reducing sugars concentration was extrapolated from a glucose standard curve ranging from 0.1 to 2 mg/mL; concentration values were plotted against time, and the slope was calculated to determine the velocity of the reaction. Concentration of released reducing sugars *vs*. time was used to calculate enzymatic activities, where 1 IU is defined as 1 µmol of glucose equivalent released per minute, under the assayed conditions. For specific activity calculation, protein concentration in mg/mL was determined against a bovine serum albumin (BSA) standard curve [57]. Xylanases activity was measured in a similar way using 2% oat xylan (Sigma Catalogue No. X0627) as substrate. The absorbance readings were compared against a standard curve of xylose (0.1 to 2 mg/mL).

MnP activity was determined spectrophotometrically (λ 270 nm) by following the formation of Mn^{3+}−malonate complex at pH 4.5 in 50 mM sodium malonate buffer with 0.5 mM $MnSO_4$ [58]. To start the reaction, H_2O_2 was added to a final concentration of 0.1 mM [59]. The reaction was followed for 30 seconds at room temperature. ΔAbs min^{-1} was converted to UL^{-1} using the malonate extinction coefficient of 11 590 $M^{-1}cm^{-1}$ [58]. MnP specific activity is expressed as IU/mg protein.

Laccase activity was monitored by oxidation of 1 mM of 2,2′-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS) in acetate buffer (pH 4.5), measuring formation of the cation radical ($\varepsilon 436 = 2.9 \times 104\ M^{-1}\ cm^{-1}$). The reaction mixture volume was 1 mL and was incubated for 30 minutes at room temperature. One unit of laccase activity is defined as the amount of enzyme catalysing the oxidation of 1 µmol of substrate per minute. ΔAbs min^{-1} was determined at λ 420 nm using spectrophotometer (BioMate, ThermoSpectronic).

Lipase/esterase and esterase activities were assessed using 2-naphthyl acetate (Sigma Catalogue No. N6875) and 4-nitrophenyl acetate (Sigma Catalogue No. N8130) as substrates, respectively. A stock solution (250 µM) of 2-naphthyl acetate was prepared in phosphate buffer saline (PBS) pH 6.5, while a stock solution of 4-nitrophenyl acetate (300 µM) was prepared in potassium phosphate buffer pH 6. The reaction volume was 1 mL and it was followed for 30 minutes at room temperature. The determination of these activities was measured spectrophotometrically at λ 538 nm for detection of 2-naphtol and at λ 410 nm for 4-nitrophenol products of enzymatic hydrolysis of lipase/esterase and esterase respectively, using the molar extinction coefficient for each substrate (23 598 $M^{-1}cm^{-1}$ for 2-naphthol and 17 700 $M^{-1}cm^{-1}$ for 4-nitrophenol). Lipase/esterase and esterase specific activities are expressed as IU/mg protein.

To determine all activities we plotted the absorbance values *vs* time, and then calculate the value of the slope of the best fitted line to the experimental points. Previously, we determined the protein concentration as described above. Finally we calculated the volumetric enzymatic and specific enzymatic activities (express as IU/mg protein).

Triplicate independent assays were performed and three readings for each sample were taken in all cases. A BioMate, ThermoSpectronic spectrophotometer was used for all the measurements.

Zymograms

Zymograms were performed to identify cellulase and xylanase activities from the supernatants of the fungal culture on Vogel's medium supplemented with 2% CMC and NaCl (0, 0.5, 1, 1.5, 2 M), or from soluble products recovered from the solid-state fermentations. Zymograms were performed as described in Quiroz-Castañeda *et al*. [57]. Briefly, 50 µg of protein were loaded per lane in the gels for the experiments in media containing NaCl, or 30 µg from the samples of the solid-state fermentation experiments. For cellulases, the gels were embedded with 2% CMC, while 2% oat xylan was used in the gels for xylanases. *Trichoderma viride* cellulase (Sigma Catalogue No. 1794), and xylanase (Sigma Catalogue No. 3876) were included as controls. Once the native-PAGE electrophoresis was carried out, gels were incubated with a 1% Congo red solution (in water) for 30 minutes at room temperature and then washed 3 times with a 1 M NaCl solution. The cellulase activity developed as clear bands. The molecular weight of the bands was estimated against a protein marker (Fermentas Catalogue No. 26612).

Optimal temperature and pH of cellulase activity

Enzyme reactions were performed as described earlier at different incubation temperatures (20, 30, 40, 50, 60, 70 and 80°C) in 50 mM sodium citrate buffer, pH 5. Different pH conditions ranging from 3 to 8 were tested at 50°C in citrate or phosphate buffer depending of the pH tested. All measurements were determined in triplicate.

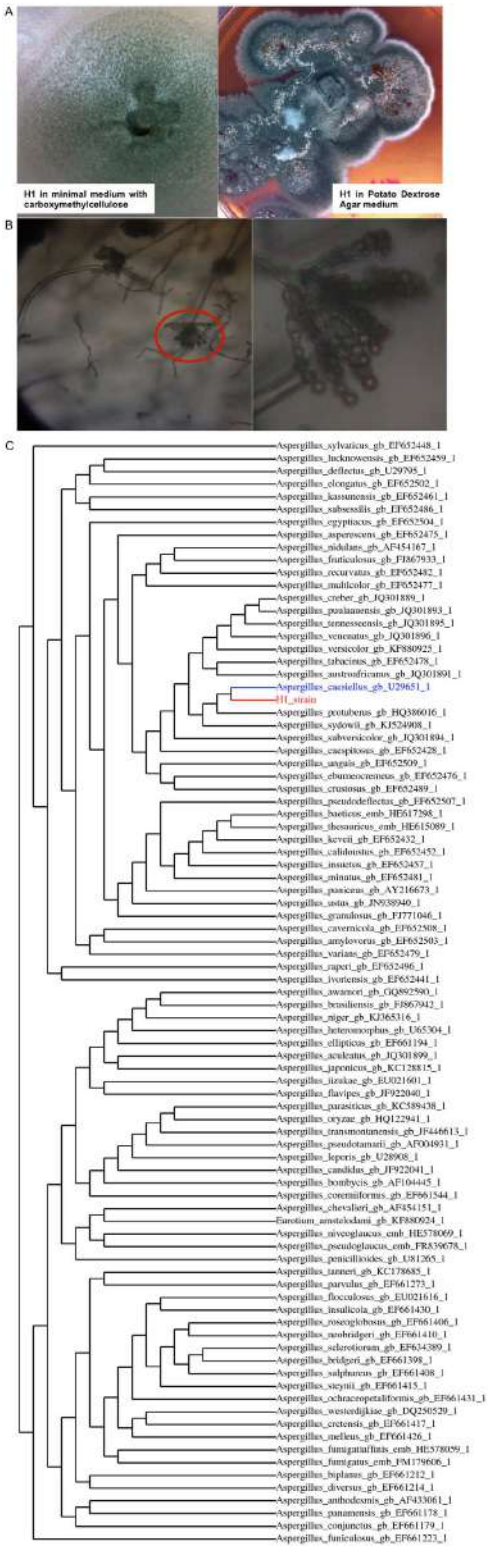


Figure 1. (A) *A. caesiellus* grown in Vogel's medium supplemented with CMC (2%) with 0.5 M NaCl and PDA medium. (B) Microcultures of *A. caesiellus* in Saboraud Agar medium. (C) Molecular phylogeny for the D1–D2 domain of 28S rDNA gene.
doi:10.1371/journal.pone.0105893.g001

Thermal-stability of the cellulase activity

Cellulases thermostability was determined by incubating 200 µL of supernatant for 60 minutes at the following temperatures: 30, 40, 50, 60 and 70°C. Subsequently, the samples were cooled on ice for 5 minutes and the residual activity was determined as described before (at 50°C and pH 5). All measurements were performed in triplicate.

Statistical calculations

For statistical treatment of experimental data, the arithmetic mean and the standard deviation were calculated. Simple classification ANOVA tests were applied to determine significant differences between the different cases. Firstly, the assumptions of ANOVA were revised: analysis of homogeneity of variance (Hartley-Cochran-Bartlett test) and normal distribution (Kolmogorov-Smirnov and Lilliefors tests) were performed [60]. Subsequently ANOVAs were conducted to demonstrate the similarities or differences between the data of the population of samples. Finally, a post hoc analysis that defines the order of the differences found in the ANOVAs was developed. The Fisher LSD, Tukey HSD and Duncan tests were considered for the post hoc analyses [60], [61]. The use of these three tests ensures greater statistical robustness of the proposed analysis. All statistical calculations were performed in STATISTICA (last version for computers).

Results and Discussion

Isolation of microorganisms

No bacterial or yeast growth was detected in the primary isolation of microorganisms, whereas a well-represented filamentous fungus was isolated in all serial dilutions up to 10^{-5}. This fungal strain was named H1 and showed the capacity to grow in 0.5 M NaCl and CMC as a carbon resource. During the secondary isolation in selective culture media (Sabouraud Agar or PDA added with 0.5 M NaCl) the fungus grew as a dark green colony on the aerial mycelium and brownish-grey in the vegetative mycelium (Figure 1). The production of metabolites secreted as dark brown and bright yellow pigments was evident. The selection forces we used to isolate microorganisms able to degrade lignocellulosic material under high salinity conditions were a complex source of carbon and energy (sugarcane bagasse) and 2 M NaCl. High NaCl concentrations inhibit the growth of most microorganisms; this ratio may explain the very low representation of the microbial population obtained during primary isolation taking in account that it came from a non-saline environment [1], [62–64].

Identification of H1

Macroscopic observations of the colony and micromorphological characteristics (hyphae, conidiophores and conidial shape) suggested that the isolate belonged to the genus *Aspergillus* (Figure 1).

Light green grainy colonies were observed in Vogel's medium with 2% CMC and 0.5 M NaCl, while medium dark green colonies were observed in PDA (Figure 1A). Microscopic examination of the culture mycelia showed columnar conidiophores of approximately 350 µm in length with vesicles 20–35 µm in diameter and pale conidia with globose and subglobose and

smooth spores. Spinose or roughened conidia were never observed. The conidia length was between 5 and 6 μm while the conidia width was approximately 3 μm (Figure 1B). These observations coincide with the characteristics described for *Aspergillus caesiellus* in the CBS-KNAW Fungal Biodiversity Centre (www.cbs.knaw.nl).

To confirm the identity of this strain, molecular markers for H1 strain were analysed. We chose to amplify a fragment of the 18S rDNA, the D1–D2 domain (28S rDNA) and the ITS1 region; the determined sequences were annotated in the GenBank with the accession numbers KJ476140, KJ476141 and KJ476142, respectively. BLAST DNA sequences analyses showed similarity with various species of *Aspergilli* demonstrating that H1 belongs to the *Aspergilli* genus.

Many molecular markers with taxonomic value have been described in the genus *Aspergillus*. These include the mitochondrial cytochrome *b* gene [65–67], a putative aflatoxin pathway regulatory gen (*aflR*) [68], the DNA topoisomerase II gene (*TOP2*) [69], the β-tubulin gene [70] and different rDNA gene regions [71]. Among the regions of the rDNA genes the most representative are the 5′end of the large-subunit 28S rDNA gene (D1–D2 region) [50] and the internal transcribed spacers 1 and 2 (ITS1 and ITS2) regions between the small- and large-subunit rDNA genes [51], [71], [72]. It has been reported that for the *Aspergilli*, the ITS1 and D1–D2 regions are the most variable and therefore the most useful for molecular identification of this genus species [51].

To define the species related to H1 strain, phylogenetic reconstructions based on the amplified sequences for each of the molecular markers were studied. Sequences with more than 80% identity and 70% coverage obtained through Blastn analysis were considered to reconstruct the phylogenies. The phylogenetic analysis of the 18S rDNA fragment was inconclusive for species identification. In this case, the H1 sequence does not group with any particular species of *Aspergillus,* although the closest relative to H1 was *Aspergillus versicolor* (Figure S1). On the other hand, the analysis of molecular phylogenies for sequences of the ITS regions were also not conclusive because H1 strain was grouped individually on a single branch of the phylogenetic tree (Figure S1). In this case, the most closely related species was *Aspergillus niger* (Figure S1).

Phylogeny analysis obtained for the sequences of the D1–D2 domain allowed us to propose the identity of strain H1 as *Aspergillus caesiellus*. Numerous studies show this region as very useful for identification filamentous fungi, not only for *Aspergillus* species [73–79]. In this case, our sequence was grouped only with a reference strain of *A. caesiellus* (*A. caesiellus* NRRL a-14879 annotated in the CBS-KNAW Fungal Biodiversity Centre). In this phylogeny it was found that *Aspergillus sydowii* and *Aspergillus protuberus* were the closest species to H1 (Figure 1C).

A. sydowii and *A. protuberus* are classified in the *Aspergillus* section *Versicolores* which was established as the *A. versicolor* group by Thom and Churche [80] and it was after reconsidered by Thom and Raper [81]. *A. protuberus* does not produce soluble pigments or exudates on MEA and its colonies show reverse light pinkish yellow to pinkish yellow colour. Moreover it presents conidia with finely roughened walls that can also be ellipsoidal to pyriform [82]. *A. sydowii* either produce soluble pigments or exudates on MEA and exhibit colonies with unpigmented reverse to brownish pink colour [82]. *A. sydowii* isolates growing specifically on Saboraud media appear as dark green colonies with a white fringe (www.thunderhouse4-yuri.blogspot.ie) and excrete purple pigments [83]. Its colonies on PDA are blue-green colour, reverse reddish and often with reddish exudates. Red-brown colour in PDA and Czapeks agar media is a very distinctive characteristic of *A. sydowii* [83], [84]. Besides, the micromorphological characterization of this species confirms spherical, and very echinulate or spinose (rough, jagged texture) conidia [83], [84].

Strain H1 shows significant differences with all the above descriptions because the presence of abundant exudates and dark soluble pigments (non-reddish and non-purple) on Saboraud agar, MEA and PDA media was detected and microculture conidia with the previously described characteristics for *A. sydowii* and *A. protuberus* were not observed. Additionally the aspect of the H1′s colony does not coincide with those described for the previous two species. Cultural and morphological criteria support the molecular identification proposal for strain H1 because they coincide phylogenetically with H1 and the rooted closest species. Systematics and taxonomy of fungi, especially for the genus *Aspergillus* spp., continue up to day using morpho-culture characters for the distinction between related species [82], [85–88]. On the other hand, *A. protuberus* cannot grow at 37°C [82], while H1 showed thermotolerant behaviour (see next section). In our case, it was critical to evaluate the colony texture, colour on the nutrient media, excretions of exudates and soluble pigments and the characteristics of the conidia to conclude the taxonomic identification. Thus, attending molecular and micromorphological criteria H1 strain was identified as *Aspergillus caesiellus*. We consider that the identification of the strain H1 is robust and conclusive according to the current criteria for identification of filamentous fungi. The polyphasic approach for the identification of microorganisms ensures better taxonomic conclusion according to several studies [89], [90].

A. caesiellus has been poorly studied. There are very few reports about the biology and physiology of this filamentous fungus. Some of these studies have reported the potential of this fungus to produce keratinase [91], amylases [92], and invertases [93]. However, little is known about the biology of this species; besides the original report, only few studies describe the isolation of *A. caesiellus* from different sources including air and dust, [94], a marine sponge [95], and chicken litter [91]. *A. caesiellus* has also been reported as a pathogen fungus or opportunistic pathogen [96], [97].

To the best of our knowledge, this is the first report describing the isolation of *A. caesiellus* from samples of lignocellulosic material. This strain is able to grow in different NaCl concentrations and tolerant up to 2 M NaCl. Phylogenetically related species such as *A. versicolor* and *A. sydowii* have been described as halotolerant and halophilic by some authors [22], [98–101]. These species were related to H1 strain in the constructed molecular phylogenies, a finding that also supports its molecular identification. This study demonstrates the flexibility and plasticity of microbial physiology and the possibility to isolate halotolerant and/or halophilic microorganisms from non-hypersaline environments. The assumptions of microbial ecology are general and valid but not absolute in these terms.

A. caesiellus H1 is a moderate halophile and thermotolerant fungus

The growth curves of *A. caesiellus* H1 showed statistically significant greater growth rates in the presence of NaCl as compared to those without NaCl for the three temperatures tested (28, 37 and 42°C) (Figure 2). H1 growth was inhibited only in 3 M NaCl. The higher specific growth rate was obtained in cultures of H1 at 37°C and 1 M NaCl. Similar specific growth rates were observed under conditions of 0.5 and 1 M NaCl in experiments at 28°C, while at 42°C, the higher specific growth rate is achieved in the culture conditions of 1.5 M NaCl (Table 1). These results

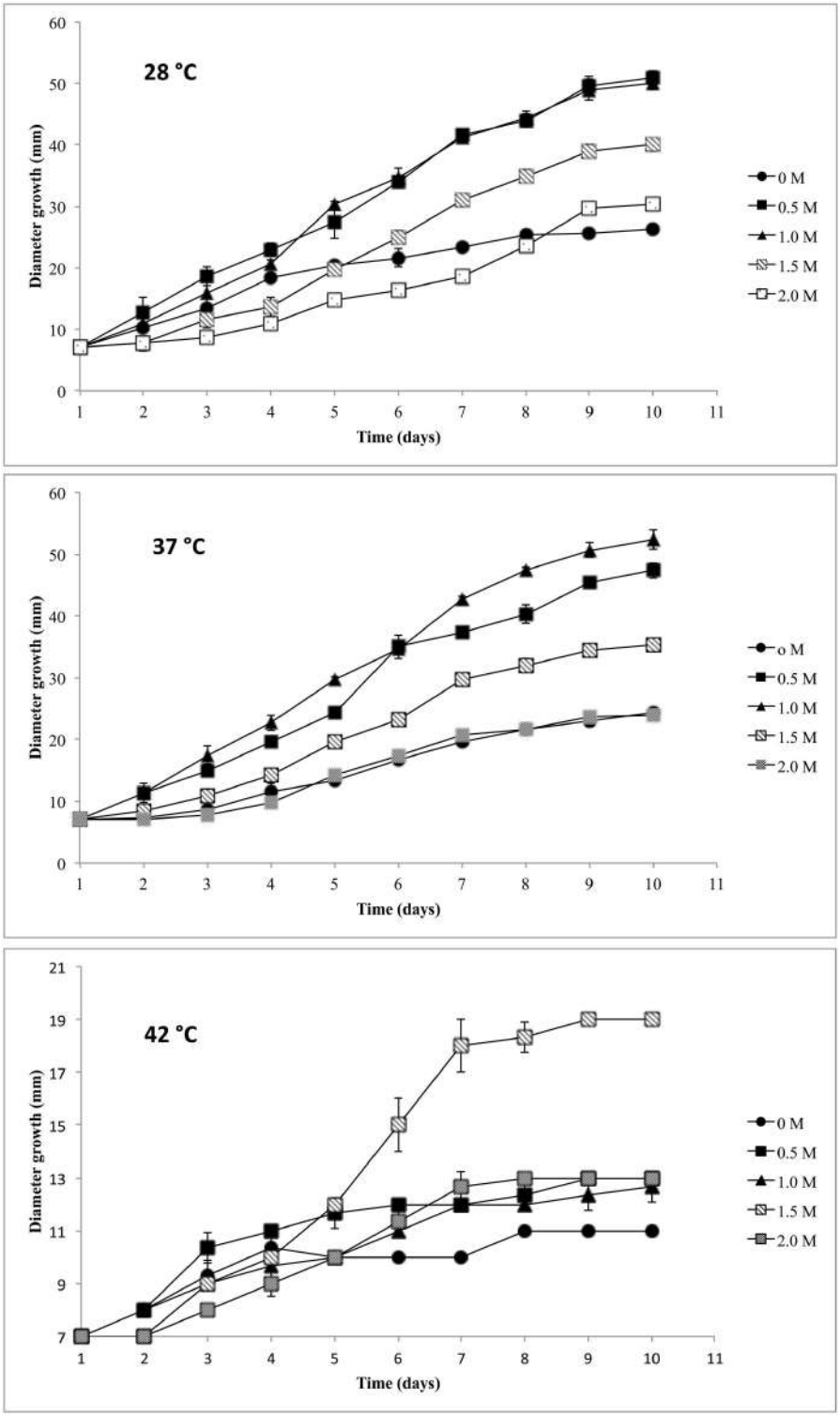


Figure 2. Growth curves of H1 at different temperatures and NaCl concentrations.
doi:10.1371/journal.pone.0105893.g002

Table 1. Specific growth rate (mm/day) of the strain H1 at different temperatures and NaCl concentrations.

NaCl (M)	Growth rate 28°C	Growth rate 37°C	Growth rate 42°C
0	2.17 ± 0.03^{d}	2.17 ± 0.07^{e}	0.39 ± 0.03^{d}
0.5	5.12 ± 0.22^{a}	4.80 ± 0.02^{b}	0.62 ± 0.03^{c}
1.0	5.22 ± 0.02^{a}	5.40 ± 0.10^{a}	0.63 ± 0.07^{c}
1.5	4.18 ± 0.14^{b}	3.59 ± 0.07^{c}	1.62 ± 0.04^{a}
2.0	2.81 ± 0.05^{c}	$2.28\pm0.10^{d\ e}$	0.80 ± 0.01^{b}

Different letters indicate different statistical orders.
doi:10.1371/journal.pone.0105893.t001

show that the H1 is also a thermotolerant strain. Thermotolerant organisms can grow in a range of temperatures above 40°C [102], [103].

There is still much controversy about the definition of extremophile microorganisms. However, one of the most accepted, defines extremophile organisms as those showing optimal growth parameters in different environmental conditions than those normal for humans [104]. In the particular case of halophile microorganisms, there are also disputes about which is the best definition. Certainly it is even difficult to define the boundaries between the concepts: halotolerant and halophile. The establishment of the difference is harder in fungi, even up to date the limits between halotolerant and halophilic strains is not outspoken [46]. However, these definitions are more than fifty years old. Larsen [105] and Kushner and Kamekura [106] introduced three important definitions: moderate halophiles, extreme halophiles and halotolerant.

Halotolerance and halophilia are very clearly defined for Bacteria and Archae. Prokaryotes are classified as obligated halophiles when they require NaCl for optimal growth. Bacteria can be grouped in two categories according with the NaCl concentration required: extremely halophilic or moderately halophilic. In general, definitions mark off extremely halophilic bacteria when they require NaCl concentrations of 2.5–5.2 M (15–30%) for their optimal growth rate, while moderate halophiles grow optimally in media containing 0.5–2.5 M (3–15%) NaCl [107–110]. Some authors have defined for convenience halophilic bacteria as microorganisms that form colonies on agar plates with 20% added NaCl [44]. General definitions as that of Madigan *et al*. [104] classify halophile organisms as those requiring high concentrations of NaCl (1.5 M) to growth optimally. In contrast, halotolerant microorganisms are defined as those showing optimal growth parameters in the absence of salt, but can survive in not-common NaCl concentrations [111], [112].

For fungi, the term halophile was introduced for first time in 1975 to describe xerophilic species inhabiting food that showed superior growth in media with NaCl as the controlling solute [113]. Gunde-Cimerman *et al*. [22], [24] proposed as halophilic or xerophilic fungi those growing well at a_w of $\leq$0.85, corresponding to 17% NaCl (3 M) or 50% glucose added to their growth medium. In the same review [22] the authors noted that these fungi have different halophilic characteristics when they are compared with the majority of halophilic prokaryotes. Plemenitaš *et al*. [114] described that most halophilic fungi do not require salt for viability because they can show adaptive properties to grow in any salinity range, from freshwater to saturated NaCl concentrations.

Thus, there are several criteria to classify microorganisms attending their growth and survival in different NaCl concentrations. We consider that *A. caesiellus* H1 can be classified as a moderate halophile fungus because, although it does not exhibit the best growth rate at NaCl concentration higher than 17%, it has greater specific growth rates in the presence of high salt concentrations than those in the absence of NaCl. However, the H1 strain was not isolated from a hypersaline environment and is capable of growing without NaCl in the medium. Our proposed classification attends the halotolerance and halophilic concepts reviewed by different authors. This shows that moderate extremophile microorganisms can live in environments that are not extreme (see Introduction).

This study is the first report of a moderate halophile strain of *A. caesiellus*. Other species have been described in the genus as halophile/halotolerant, for example, *A. versicolor, A. sydowii, A. flavus*, etc. Some species, such as *A. versicolor and A. sydowii* have been found as part of the fungal microbiota of hypersaline environments [22], [99], [100], [115–117].

The physiology and metabolism of microorganisms with the ability to grow in high concentrations of NaCl ensure a number of applications of biotechnological interest [118–121]. In particular, the study of the ability of *A. caesiellus* H1 for degradation of lignocellulosic material became an attraction for our research group. H1 is a strain isolated from sugarcane bagasse, able to grow at a range of NaCl between 0.5 and 2 M and uses cellulose as the sole source of carbon and energy. These qualities point to *A. caesiellus* H1 as a good candidate to study some of its lignocellulolytic enzymes.

Cellulases activity

Cellulase enzymatic activity was observed at its highest level within the sixth day of culture in 1.5 M NaCl, using CMC as only carbon source at 28°C (Figure 3). From the fifth day on, cultures of H1 in the presence of varying salt concentrations showed higher cellulase specific activity compared to cultures in the absence of NaCl with exception of the 2 M condition, which from day 7 on anyway showed similar activities to the control culture (without NaCl) (Figure 3). To the best of our knowledge this is the first report of cellulase activity for *A. caesiellus*.

Zymograms of supernatants of *A. caesiellus* H1 cultures in CMC as a carbon source showed two isoforms for cellulases (of around 50 and 35 kDa), regardless of the NaCl concentration (Figure 4). The 50 kDa activity band diminishes at high salt concentrations (1.5 and 2 M), since the amount of protein loaded was the same for all the lanes. We were not able to detect any bands in glucose grown cultures in this experiment suggesting that cellulase expression is repressed by glucose (Figure 4, lane 1). These results indicate that cellulases from *A. caesiellus* H1 are capable of functioning at concentrations up to 2 M NaCl, since growth is supported in this condition with CMC as a sole carbon source.

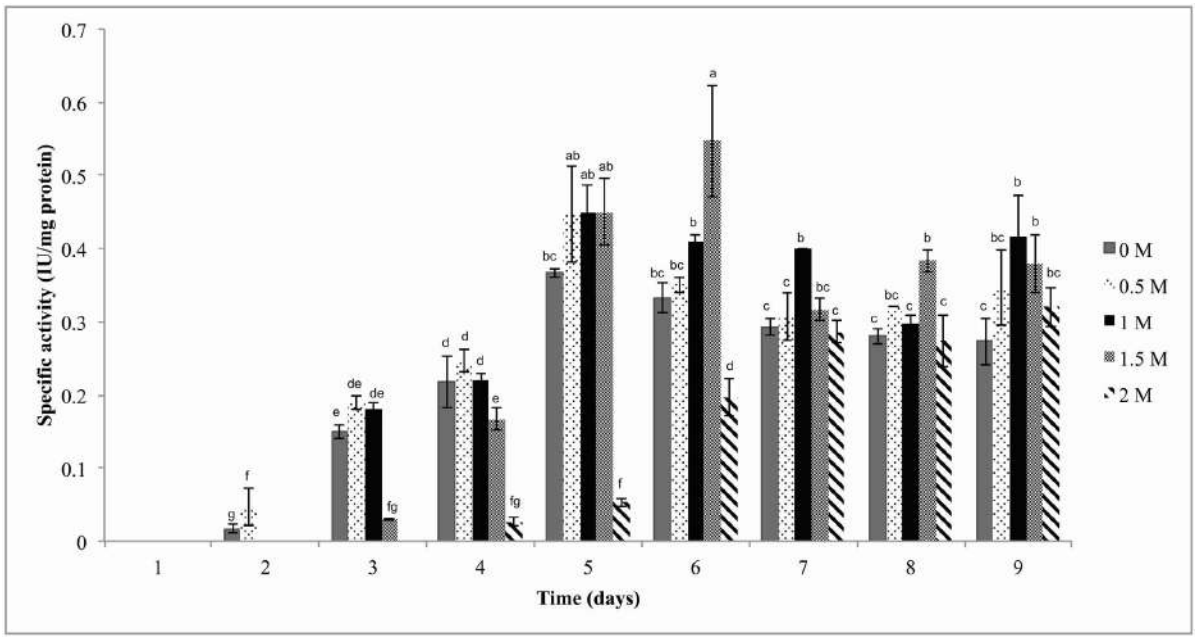


Figure 3. Cellulase activity of H1 cultures on CMC (2%) as the sole carbon source at different NaCl concentrations.
doi:10.1371/journal.pone.0105893.g003

Solid-state fermentation of natural substrates by H1

We were interested in testing this fungal strain for enzyme production of lignocellulosic substrates in similar conditions to those found in the field, so solid state fermentations were set up in the absence of NaCl. Wheat straw, maize stover and agave fibres (in that order) were the substrates where the best cellulase and xylanase activities were obtained during solid-state fermentation. Xylanase production was favoured over cellulase production when these substrates were used, and it was higher in wheat straw and maize stover (Figure 5). When a PAGE native gel with embedded xylan was stained with Congo red, four bands of approximately 10, 12, 15 and 20 kDa with xylanase activity were observed for the maize stover fermentation (Figure 6A, lane 3). The 12, 15 and 20 kDa bands also are induced, although in different proportions, when wheat straw and agave fivers were used as substrates, being the 15 kDa band very abundant in the agave fibres fermentation (Figure 6A, lanes 2 and 5). In the sugar cane bagasse fermentation only the 15 kDa band was observed (Figure 6A, lane 4). Again, no xylanase activity was observed in the zymogram when glucose was used as the carbon source Figure 6A, lane 6), indicating a repression by this sugar for xylanases expression. Distinct band patterns were observed for different substrates suggesting a differential expression of xylanases according to the substrate used.

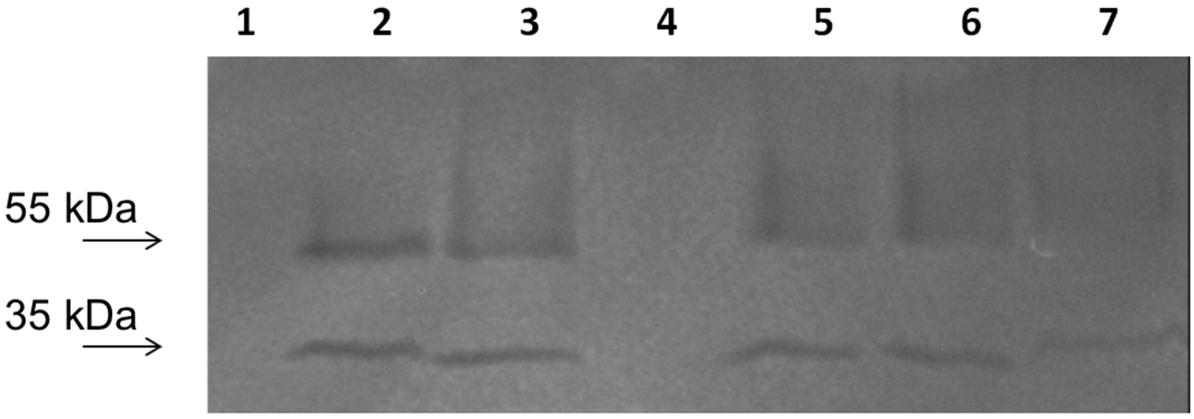


Figure 4. Zymogram for detection of extracellular cellulases from liquid cultures of strain H1 at different NaCl concentrations in **CMC (2%) as the sole carbon source.** Lane 1: Control culture in 2% glucose without NaCl. Lane 2: Culture without NaCl. Lane 3: Culture with 0.5 M NaCl. Lane 4: No sample charge. Lane 5: Culture with 1 M NaCl. Lane 6: Culture with 1.5 M NaCl. Lane 7: Culture with 2 M NaCl. Arrows indicate the position were the molecular weight markers migrated.
doi:10.1371/journal.pone.0105893.g004

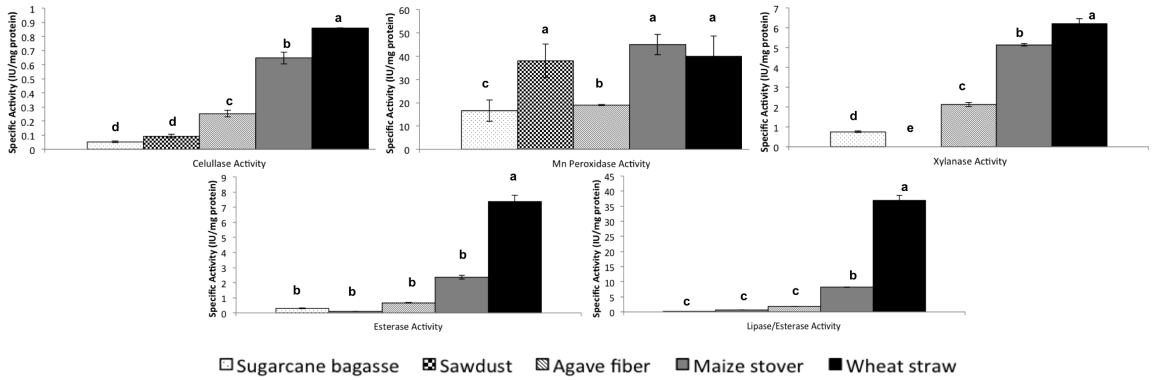


Figure 5. Enzymatic activities from solid-state fermentation of H1 in different substrates. Different letters indicate different statistical orders.
doi:10.1371/journal.pone.0105893.g005

In a native gel with cellulose stained with Congo red, we observed three bands with cellulase activity for wheat straw and maize stover cultures, while only two bands were evident in sugarcane bagasse and agave fibres (Figure 6B). Note that in cultures with glucose as a carbon source, only a low basal expression of cellulases is observed (Figure 6B), suggesting again that glucose represses the expression of most of the isoforms of these enzymes. Sawdust proved the worst substrate for induction of these enzymes (Figure 5).

A. caesiellus H1 also showed strong Mn peroxidase activity (Figure 5). No statistically significant differences in the Mn peroxidase activity measured in the cultures grown on wheat straw, maize stover and sawdust were observed. In the latter substrates the best Mn peroxidase activity was obtained (Figure 5). Cultures of the fungus in agave fibres and sugarcane bagasse produced lower Mn peroxidase activity with statistically significant differences between them (Figure 5).

Fungal laccases have been widely studied for their industrial applications. However, no laccase activity for H1 strain was detected in any of the conditions we tested.

Wheat straw, maize stover and agave fibres were the substrates where the higher esterase and lipase/esterase activity were determined (Figure 5). This fungus showed a very high lipase/esterase activity with 2-naphthyl acetate as a substrate in wheat straw and maize stover (Figure 5). In general, the substrates where the best enzymatic activities were observed were wheat straw and maize stover. These results support the potential of this fungus to degrade lignocellulosic material and its potential biotechnological applications. Experiments are under way to test H1 for lignocellulosic substrates degradation in the presence of NaCl.

Optimum temperature and thermo-stability of cellulases activity

No statistically significant differences in the cellulase activity of the fungus at 50, 60 and 70°C were observed (Figure 7A). When fungal supernatant was incubated for one hour at 60°C, it kept 40% of the cellulases activity (residual activity). This result suggests that *A. caesiellus* cellulases are thermostable. One of the important factors for the industrial application of cellulolytic enzymes is its

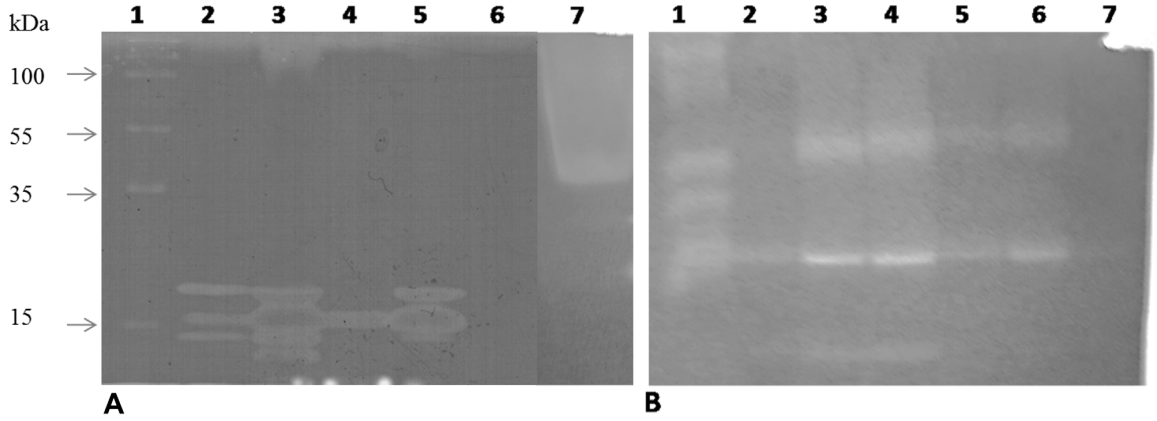


Figure 6. Zymograms for detection of extracellular xylanases and cellulases from solid-state cultures of H1 in different natural substrates. (A) Native gel for xylanases. Line 1: Molecular Weight marker. Line 2: Culture in wheat straw. Line 3: Culture in. maize stover Line 4: Culture in. sugarcane bagasse Line 5: Culture in. agave fibre Line 6: Culture in glucose. Line 7: Positive control. Xylanasas from *Trichoderma viridae.* (B) Native gel for cellulases. Line 1: Positive control. Cellulases from *Trichoderma viridae.* Line 2: Culture in glucose. Line 3: Culture in wheat straw. Line 4: Culture in maize stover. Line 5: Culture in sugarcane bagasse. Line 6: Culture in agave fibre. Line 7: Culture in sawdust. The arrows indicate bands with activity.
doi:10.1371/journal.pone.0105893.g006

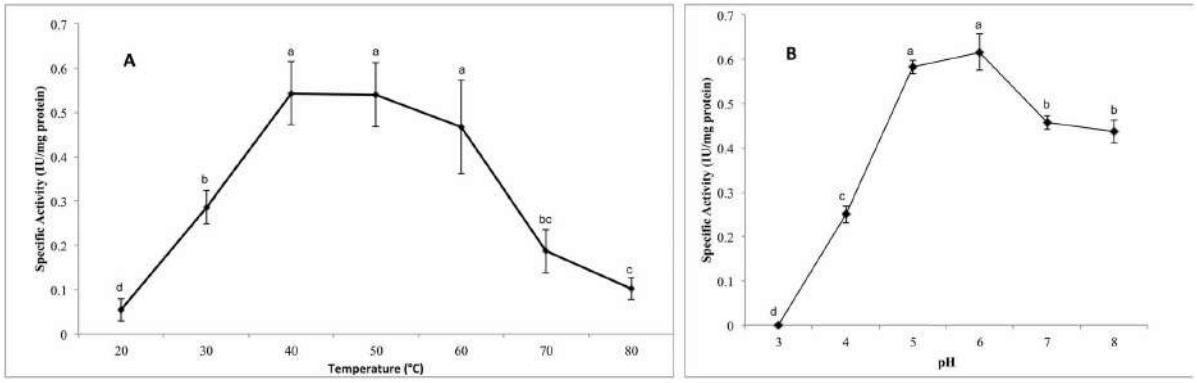


Figure 7. (A) Optimal temperature of celullase activity of H1. (B) Optimal pH of celullase activity of H1. Different letters indicate different statistical orders.
doi:10.1371/journal.pone.0105893.g007

sturdiness [57], [122–125]. Our results are comparable with those obtained by Liu *et al*. [95] and Raddadi *et al*. [126], and different than those reported by Pang *et al*. [127], since they lost the cellulase activity when supernatant was heated for one hour at 50°C. In our case, the residual activity decreases to 22% (residual activity) when the supernatants were exposed for one hour at 70°C. A similar result was reported by Narra *et al*. [128], who found a dramatic decrease of the cellulolytic activity at this temperature.

Cellulase Ph Optimum

The best cellulase activity was observed at pH values between 5 and 6 (Figure 7B). This result is consistent for other cellulases [95], [128]. However, some cellulases show optimal pH in the neutral or basic range [127], [129].

In this work we have described the isolation and characterization of a moderate halophile *A. caesiellus* strain that produces thermostable cellulases and other biotechnological interesting activities. Our results show that this strain has a great potential for lignocellulose degradation and could be used for biorefinery applications.

Supporting Information

Figure S1 Phylogenies for molecular taxonomic identification of moderate halophile strain H1. (A) Molecular phylogeny considering the sequence of the fragment of the 18S ribosomal DNA. (B) Molecular phylogeny considering the sequence of the regions of the ITS1 region.
(TIF)

Acknowledgments

We are grateful to Dr. Claudia Martínez-Anaya for her critical review of this manuscript and Mss. Aspasia Giannika for her help in the image editing.

Author Contributions

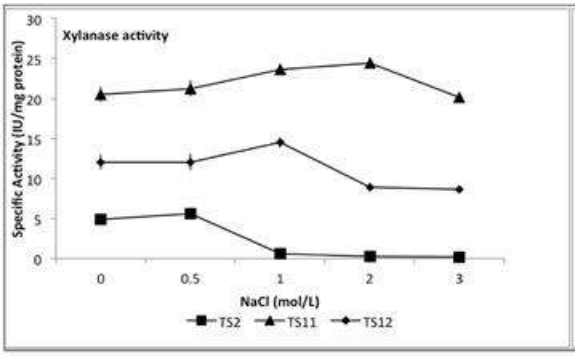
Conceived and designed the experiments: JF-M EM-M RB-G EB-L. Performed the experiments: RB-G EB-L LC-S DA-Z KA-H CM-H AS-R. Analyzed the data: JF-M AS-R RB-G EB-L. Contributed reagents/materials/analysis tools: RR-H. Contributed to the writing of the manuscript: RB-G EB-L JF-M.

References

1. Oren A (2002) Molecular ecology of extremely halophilic Archaea and Bacteria. FEMS Microbiology Ecology 39, 1–7.
2. Nicolaus B, Lama L, Manca MC, Gambacorta R (1999) Extremophiles: Polysaccharides and enzymes degrading polysaccharides (INIST-CNRS, Cote INIST: L 27265, 35400008510647.0040: Research signpost, Trivandrum, INDE).
3. Conte F (2008) Engineering of glycoside hydrolases of biotechnological interest from extremophilic organisms. Doctoral thesis. Universitá di Napoli Federico II, Italy.
4. Marques E, Uesugi C (2013) Avaliação de bactérias extremófilas facultativas na produção de fitomassa do híbrido "urograndis" de eucalipto, a partir de sementes. Revista Árvore 37, 41–47.
5. Bhattacharya A, Pletscheke B (2014) Review of the enzymatic machinery of *Halothermothrix orenii* with special reference to industrial applications. Enzyme and Microbial Technology 55, 159–169.
6. Díaz-Tena E, Rodríguez-Ezquerro A, López de Lacalle Marcaide LN, Gurtubay Bustinduy L, Elías-Sáenz A (2014) A sustainable process for material removal on pure copper by use of extremophile bacteria. Journal of Cleaner Production. In Press, Corrected Proof: dx.doi.org/10.1016/j.jclepro.2014.01.061.
7. Gomes J, Steiner W (2004) The Biocatalytic Potential of Extremophiles and Extremozymes. Food Technol. Biotechnol. 42, 223–235.
8. Mesbah NM, Wiegel J (2009) *Natronovirga wadinatrunensis* gen. nov., sp. nov. and *Natranaerobius trueperi* sp. nov., halophilic, alkalithermophilic microorganisms from soda lakes of the Wadi An Natrun, Egypt. Int. J. Syst. Evol. Microbiol. 59, 2042–2048.
9. Hamamura N, Itai T, Liu Y, Reysenbach AL, Damdinsuren N, et al. (2014) Identification of anaerobic arsenite-oxidizing and arsenate-reducing bacteria associated with an alkaline saline lake in Khovsgol, Mongolia: Arsenic transformation in a Mongolian saline lake. Environmental Microbiology Reports. DOI:10.1111/1758-2229.12144.
10. Gunde-Cimerman N, Zalar P, de Hoog S, Plemenitaš A (2000) Hypersaline waters in salterns – natural ecological niches for halophilic black yeasts. FEMS Microbiology Ecology 32, 235–240.
11. Zhang T, Datta S, Eichler J, Ivanova N, Axen SD, et al. (2011) Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance. Green Chem. 13, 2083–2090.
12. Zhang G, Li S, Xue Y, Mao L, Ma Y (2012) Effects of salts on activity of halophilic cellulase with glucomannanase activity isolated from alkaliphilic and halophilic Bacillus sp. BG-CS10. Extremophiles 16, 35–43.
13. Gunny AN, Arbain D, Edwin Gumba R, Jong BC, Jamal P (2014) Potential halophilic cellulases for in situ enzymatic saccharification of ionic liquids pretreated lignocelluloses. Bioresource Technology 155, 177–181.
14. Antón J, Oren A, Benlloch S, Rodríguez-Valera F, Amann R, et al. (2002) *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds. Int. J. Syst. Evol. Microbiol. 52, 485–491.
15. Simmons CW, Claypool JT, Marshall MN, Jabusch LK, Reddy AP, et al. (2014) Characterization of bacterial communities in solarized soil amended with lignocellulosic organic matter. Applied Soil Ecology 73, 97–104.
16. Sorokin DY, Kolganova TV (2013) Bacterial chitin utilization at halophilic conditions. Extremophiles 18, 243–248.

17. Muruga B, Anyango B (2013) A Survey of Extremophilic Bacteria in Lake Magadi, Kenya. Columbia International Publishing American Journal of Molecular and Cellular Biology 2, 14–26.
18. Challacombe JF, Majid S, Deole R, Brettin TS, Bruce D, et al. (2013) Complete genome sequence of *Halorhodospira halophila* SL1. Stand Genomic Sci 8, 206–214.
19. Mesbah NM, Hedrick DB, Peacock AD, Rohde M, Wiegel J (2007) *Natranaerobius thermophilus* gen. nov., sp. nov., a halophilic, alkalithermophilic bacterium from soda lakes of the Wadi An Natrun, Egypt, and proposal of Natranaerobiaceae fam. nov. and Natranaerobiales ord. nov. Int. J. Syst. Evol. Microbiol. 57, 2507–2512.
20. Leela A (2013) Catalytic properties and industrial potential of halophilic extracellular protease from salt – pan isolates. PhD Thesis. Bharathidassan University.
21. Roldán-Carrillo T, Castorena-Cortés G, Reyes-Avila J, Zapata-Peñasco I, Olguín-Lora P (2013) Effect of porous media types on oil recovery by indigenous microorganisms from a Mexican oil field. Journal of Chemical Technology & Biotechnology 88, 1023–1029.
22. Gunde-Cimerman N, Ramos J, Plemenitaš A (2009) Halotolerant and halophilic fungi. Mycological Research 113, 1231–1241.
23. Butinar L, Zalar P, Frisvad JC, Gunde-Cimerman N (2005) The genus *Eurotium* - members of indigenous fungal community in hypersaline waters of salterns. FEMS Microbiol. Ecol. 51, 155–166.
24. Gunde-Cimerman N, Butinar L, Sonjak S, Turk M, Uršič V, et al. (2005) Halotolerant and Halophilic Fungi from Coastal Environments in the Arctics. In Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya, N. Gunde-Cimerman, A. Oren, and A. Plemenitaš, eds. (Springer Netherlands), 397–423.
25. Zalarc P, Hoog GS de, Schroers HJ, Crous J, Groenewald JZ, et al. (2007) Phylogeny and ecology of the ubiquitous saprobe *Cladosporium shpaerospermum*, with descriptions of seven new species from hypersaline environments. Studies in Mycology 58, 157–183.
26. Zalar P, Sybren de Hoog G, Schroers HJ, Frank JM, Gunde-Cimerman N (2005) Taxonomy and phylogeny of the xerophilic genus *Wallemia* (Wallemiomycetes and Wallemiales, cl. et ord. nov.). Antonie Van Leeuwenhoek 87, 311–328.
27. DasSarma S, Aurora P (2012) Halophiles. In eLS, John Wiley & Sons, Ltd, Chichester. United Kindom. DOI:10(9780470015902), a0000394.
28. Molitoris HP, Buchalo AS, Kurchenko I, Nevo E, Rawal BS, et al (2000) Physiological diversity of the first filamentous fungi isolated from the hypersaline dead sea. Fungal Diversity 5, 55–70.
29. Kutateladze NZ (2009) Selection of Microscopic Fungi - Pectinase Producers. Bulletin of the Georgian National Academy of Science 3.
30. Li X, Wang HL, Li T, Yu HY (2012) Purification and characterization of an organic solvent-tolerant alkaline cellulase from a halophilic isolate of *Thalassobacillus*. Biotechnol. Lett. 34, 1531–1536.
31. Xue DS, Chen HY, Lin DQ, Guan YX, Yao SJ (2012) Optimization of a natural medium for cellulase by a marine *Aspergillus niger* using response surface methodology. Appl. Biochem. Biotechnol. 167, 1963–1972.
32. Ren W, Zhang F, Yang X, Tang S, Ming H, et al. (2013) Purification and properties of a SDS-resistant xylanase from halophilic *Streptomonospora* sp. YIM 90494. Cellulose 20, 1947–1955.
33. Liu X, Huang Z, Zhang X, Shao Z, Liu Z (2014) Cloning, expression and characterization of a novel cold-active and halophilic xylanase from *Zunongwangia profunda*. Extremophiles 18, 441–450.
34. Onishi H, Fuchi H, Konomi K, Hidaka O, Kamekura M (1980) Isolation and distribution of a variety of halophilic bacteria and their clas- sification by salt-response. Agric Biol Chem 44, 1253–1258.
35. Horikoshi K (1999) Alkaliphiles: some applications of their products for biotechnology. Microbiol Mol Biol Rev 63, 735–750.
36. Saiz-Jimenez C, Laiz L (2000) Occurrence of halotolerant/halophilic bacterial communities in deteriorated monuments. Int Biodeterior Biodegrad 46, 319–326.
37. Abdel-Fafez S (1981) Halophilic fungi of desert soils in Saudi Arabia. Mycopathologia 75, 75–80.
38. Moubasher A, Abdel-Hafez S, Bagy M, Abdel-Satar M (1990) Halophilic and halotolerant fungi in cultivated desert and salt marsh soils from Egypt. Acta Mycologica 26, 65–81.
39. Nielsen P, Fritze D, Priest FG (1995) Phenetic diversity of alkaliphilic *Bacillus* strains: proposal for nine new species. Microbiology 141, 1745–1761.
40. Castillo G, Demoulin V (1997) NaCl salinity and temperature ef- fects on growth of three wood-rotting basidiomycetes from a Papua New Guinea coastal forest. Mycol Res 101, 341–344.
41. Arahal DR, Marquez MC, Volcani BE, Schleifer KH, Ventosa A (1999) *Bacillus marismortui* sp. nov., a new moderately halophilic species from the Dead Sea. Int J Syst Bacteriol 49, 521–530.
42. Piñar G, Ramos C, Rölleke S, Schabereiter-Gurtner C, Vybiral D, et al. (2001) Detection of indigenous *Halobacillus* populations in damaged ancient wall paintings and building materials: molecular monitoring and cultivation. Appl Environ Microbiol 67, 4891–4895.
43. Santos SX, Carvalho CC, Bonfá MR, Silva R, Gomes E (2004) Screening for pectinolytic activity of wood-rotting Basidio- mycetes and characterization of the enzymes. Folia Microbiol 49, 46–52.
44. Echigo A, Hino M, Fukushima T, Mizuki T, Kamekura M, et al. (2005) Endospore of halophilic bacteria of the family Bacillaceae isolated from non-saline Japanese soil may be transported by Kosa event (Asian dust storm). Saline Systems 1, 1–13.
45. Abrahão MC, Gugliotta AM, Da Silva R, Fujieda RJ, Boscolo M, et al. (2008) Ligninolytic activity from newly iso-lated basidiomycete strains and effect of these enzymes on the azo dye orange II decolourisation. Annals Microbiol 58, 427–432.
46. Arakaki R, Monteiro D, Boscolo R, Gomes E (2013) Halotolerance, ligninase production and herbicide degradation ability of basidiomicetes strains. Brazilian Journal of Microbiology 44, 1207–1214.
47. Vogel HJ (1956) A convenient growth medium for *Neurospora* (Medium N). Microbial Genet Bull 13, 42–43.
48. Kuhad RC, Kapoor RK, Lal R (2004) Improving the yield and quality of DNA isolated from white-rot fungi. Folia Microbiol 49, 112–116.
49. Borneman J, Hartin RJ (2000) PCR Primers that amplify fungal rRNA genes from environmental samples. Appl Environ Microbiol 66, 4356–4360.
50. Peterson SW (2000) Integration of modern taxonomic methods for Penicillium and Aspergillus. Phylogenetic relationships in Aspergillus based on rDNA sequence analysis. (Amsterdam: Harwood Academic Publishers.), 323–355.
51. Hinrikson HP, Hurst SF, Lott TJ, Warnock DW, Morrison CJ (2005) Assessment of ribosomal large-subunit D1–D2, internal transcribed spacer 1, and internal transcribed spacer 2 regions as targets for molecular identification of medically important *Aspergillus* species. J. Clin. Microbiol. 43, 2092–2103.
52. Sambrook J, Russell DW (2001) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press).
53. 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, 777–780.
54. Onsori H, Zamani MR, Motallebi M, Zarghami N (2005) Identification of over producer strain of endo-β-1,4-glucanase in *Aspergillus* Species: Characterization of crude carboxymethyl cellulase. African Journal of Bacteriology 4, 26–30.
55. Miller GL (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal. Chem. 31, 426–428.
56. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein Measurement with the Folin Phenol Reagent. J. Biol. Chem. 193, 265–275.
57. Quiroz-Castañeda RE, Balcázar-López E, Dantán-González E, Martinez A, Folch-Mallol J, et al. (2009) Characterization of cellulolytic activities of *Bjerkandera adusta* and *Pycnoporus sanguineus* on solid wheat straw medium. Electronic Journal of Biotechnology 12.
58. Järvinen J, Taskila S, Isomäki R, Ojamo H (2012) Screening of white-rot fungi manganese peroxidases: a comparison between the specific activities of the enzyme from different native producers. AMB Express 2, 62.
59. Wariishi H, Valli K, Gold MH (1992) Manganese(II) oxidation by manganese peroxidase from the basidiomycete *Phanerochaete chrysosporium*. Kinetic mechanism and role of chelators. J. Biol. Chem. 267, 23688–23695.
60. Hoaglin DC, Welsch RE (1978) The Hat Matrix in Regression and ANOVA. The American Statistician 32, 17–22.
61. Fisher RA, Yates F (1949) Statistical tables for biological, agricultural and medical research. Edinburgh, Oliver and Boyd Ltd. Ed. 3, 112.
62. Evans S, Hansen RW, Stone HM, Schneegurt MA (2013) Isolation and Characterization of Halotolerant Soil Fungi from the Great Salt Plains of Oklahoma (USA). Cryptogamie, Mycologie 34, 329–341.
63. Konte T, Plemenitaš A (2013) The HOG signal transduction pathway in the halophilic fungus *Wallemia ichthyophaga*: identification and characterization of MAP kinases WiHog1A and WiHog1B. Extremophiles 17, 623–636.
64. Shtarkman YM, Koçer ZA, Edgar R, Veerapaneni RS, D'Elia T, et al. (2013) Subglacial Lake Vostok (Antarctica) Accretion Ice Contains a Diverse Set of Sequences from Aquatic, Marine and Sediment-Inhabiting Bacteria and Eukarya. PLoS ONE 8, e67221.
65. Wang L, Yokoyama K, Takahasi H, Kase N, Hanya Y, et al. (2001) Identification of species in *Aspergillus* section Flavi based on sequencing of the mitochondrial cytochrome b gene. International Journal of Food Microbiology 71, 75–86.
66. Yokoyama K, Wang L, Miyaji M, Nishimura K (2001) Identification, classification and phylogeny of the *Aspergillus* section Nigri inferred from mitochondrial cytochrome b gene. FEMS Microbiology Letters 200, 241–246.
67. Spiess B, Buchheidt D (2013) Protocol of a LightCycler PCR Assay for Detection and Quantification of Aspergillus fumigatus DNA in Clinical Samples of Neutropenic Patients. In Laboratory Protocols in Fungal Biology, V.K. Gupta, M.G. Tuohy, M. Ayyachamy, K.M. Turner, and A. O'Donovan, eds. (Springer New York), 315–320.
68. Chang PK, Bhatnagar D, Cleveland TE, Bennett JW (1995) Sequence variability in homologs of the aflatoxin pathway gene *aflR* distinguishes species in *Aspergillus* section Flavi. Appl. Environ. Microbiol. 61, 40–43.
69. Kanbe T, Yamaki K, Kikuchi A (2002) Identification of the Pathogenic *Aspergillus* Species by Nested PCR Using a Mixture of Specific Primers to DNA Topoisomerase II Gene. Microbiology and Immunology 46, 841–848.
70. Geiser DM, Pitt JI, Taylor JW (1998) Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. Proc. Natl. Acad. Sci. USA 95, 388–393.

71. Iwen PC, Hinrischs SH, Rupp ME (2002) Utilization of the internal transcribed spacer regions as molecular targets to detect and identify human fungal pathogens. Med. Mycol. 40, 87–109.
72. White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics (In M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols-a guide to methods and applications. Academic Press, San Diego, Calif).
73. Guarro J (2012) Taxonomía y biología de los hongos causantes de infección en humanos. Enfermedades Infecciosas y Microbiología Clínica 30, 33–39.
74. Kwiatkowski NP, Babiker WM, Merz WG, Carroll KC, Zhang SX (2012) Evaluation of Nucleic Acid Sequencing of the D1/D2 Region of the Large Subunit of the 28S rDNA and the Internal Transcribed Spacer Region Using SmartGene IDNS Software for Identification of Filamentous Fungi in a Clinical Laboratory. The Journal of Molecular Diagnostics 14, 393–401.
75. Massire C, Buelow DR, Zhang SX, Lovari R, Matthews HE, et al. (2013) PCR Followed by Electrospray Ionization Mass Spectrometry for Broad-Range Identification of Fungal Pathogens. J. Clin. Microbiol. 51, 959–966.
76. Odero V, Galán-Sánchez F, García-Agudo L, García-Tapia AM, Guerrero-Lozano I, et al. (2013) Fungemia por *Trichosporon asahii* en paciente con neoplasia hematológica. Revista Iberoamericana de Micología. http://dx.doi.org/10.1016/j.riam.2013.09.001.
77. Santos DW, Padovan AC, Melo AS, Gonçalves SS, Azevedo VR, et al. (2013) Molecular Identification of Melanised Non-Sporulating Moulds: A Useful Tool for Studying the Epidemiology of Phaeohyphomycosis. Mycopathologia 175, 445–454.
78. Alam MZ, Alam Q, Jiman-Fatani A, Kamal MA, Abuzenadah AM, et al. (2014) *Candida* identification: a journey from conventional to molecular methods in medical mycology. World Journal Microbiol Biotechnol 1–15.
79. Kim MJ, Park PW, Ahn JY, Kim KH, Seo JY, et al. (2014) Fatal Pulmonary Mucormycosis Caused by *Rhizopus microsporus* in a Patient with Diabetes. Annals of Laboratory Medicine 34, 76.
80. Thom G, Church M (1926) The *Aspergilli*. Baltimore, MD: Williams and Wilkins Co.
81. Thom G, Raper K (1945) Manual of the *Aspergilli*. Boltimore, MD: Williams and Willkins Co.
82. Jurjevic Z, Stephen W, Bruce W (2012) *Aspergillus* section Versiolores: nine new species and multilocus DNA sequence based phylogeny. IMA Fungus: The Global Mycological Journal 3, 59–79.
83. Panchagnula B, Terlir R (2013) Morphological and molecular characterization of a lipolytic marine fungal isolate *Aspergillus sydowii* strain BTSS1005. International Journal of Biology, Pharmacy and Allied Sciences 2, 1488–1500.
84. Bindiya P, Ramana T (2012) Optimization of lipase production from an indigenously isolated marine *Aspergillus sydowii* of Bay of Bengal. Journal of Biochemical Technology 3, 203–211.
85. Klich M (2002) Identification of common *Aspergillus* species. Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands.
86. Sugui J, Peterson S, Clark L, Nardone G, Folio L, et al. (2012) *Aspergillus tanneri* sp. nov., a new pathogen that causes invasive disease refractory to antifungal therapy. J. Clin. Microbiol 50, 3309–3317.
87. Krimitzas A, Pyrri I, Kouvelis V, Kapsanaki-Gotsi E, Typas M (2013) A phylogenetic analysis of Greek isolates of *Aspergillus* species based on morphology and nuclear and mitocondrial gene sequences. BioMed Research International. Doi:10.1155/2013/260395.
88. Barton R (2013) Laboratory diagnosis of invasive aspergillosis: from diagnosis to prediction of outcome. Scientifica. Doi:10.1155/2013/459405.
89. Passarini MR, Santos C, Lima N, Berlinck RS, Sette LD (2013) Filamentous fungi from the Atlantic marine sponge *Dragmacidon reticulatum*. Arch Microbiol 195, 99–111.
90. Krishnamurthi S (2014) The Study of Prokaryotic diversity of a landfull environment using a polyphasic taxonomic approach. INFLIBNET. http://hdl.handle.net/10603/14893.
91. Marcondes NR, Taira CL, Vandresen DC, Svidzinski TI, Kadowaki MK, et al. (2008) New Feather-Degrading Filamentous Fungi. Microb Ecol 56, 13–17.
92. De Paris LD, Scheufele FB, Teixeira T, Guerreiro TL, Hasan SD (2010) Estudo do crescimento de *A. casiellus* em farelo de soja convencional para produção de enzimas. Estudos Tecnológicos Em Engenharia 6.
93. Novaki L, Hasan S, Kadowaki M, Andrade D (2010) Produção de invertase por fermentação em estado sólido a partir de farelo de soja. ENGEVISTA 12.
94. Lugauskas A, Krikštaponis A (2004) Filamentous Fungi Isolated in Hospitals and Some Medical Institutions in Lithuania. Indoor and Built Environment 13, 101–108.
95. Liu D, Zhang R, Yang X, Wu H, Xu D, et al. (2011) Thermostable cellulase production of *Aspergillus fumigatus* Z5 under solid-state fermentation and its application in degradation of agricultural wastes. International Biodeterioration & Biodegradation 65, 717–725.
96. Denning DW (1998) Invasive aspergillosis. Clinical Infectious Diseases 26, 584–589.
97. Browman J, Panizza B, Gandhi M (2007) Sphenoid sinus fungal balls. Annals of Otology, Rhinology and Laryngology 116, 514–519.
98. Abdel-Hafez SI (1981) Halophilic fungi of desert soils in Saudi Arabia. Mycopathologia 75, 75–80.
99. Butinar L, Frisvad JC, Gunde-Cimerman N (2011) Hypersaline waters – a potential source of foodborne toxigenic aspergilli and penicillia. FEMS Microbiology Ecology 77, 186–199.
100. Nayak S, Gonsalves V, Nazareth S (2012) Isolation and salt tolerance of halophilic fungi from mangroves and solar salterns in Goa - India. IJMS 41, 164–172.
101. Oren A, Gunde-Cimerman N (2012) Fungal Life in the Dead Sea. In Biology of Marine Fungi, C. Raghukumar, ed. (Springer Berlin Heidelberg), 115–132.
102. Kumar M, Yadav AN, Tiwari R, Prasanna R, Saxena AK (2014) Deciphering the diversity of culturable thermotolerant bacteria from Manikaran hot springs. Ann Microbiol 1–11.
103. Maller A, Silva TM, Damásio AR, Hirata IY, Jorge JA, et al. (2013) Functional properties of a manganese-activated exo-polygalacturonase produced by a thermotolerant fungus *Aspergillus niveus*. Folia Microbiol 58, 615–621.
104. Madigan M, Martinko J, Stahl D, Clark D (2012) Brock Biology of Microorganisms (San Francisco, USA: (13 Eth) Pearson Education).
105. Larsen H (1962) Halophilism. The Bacteria 4, 297–342.
106. Kushner DJ, Kamekura M (1988) Physiology of halophilic eubacteria. In *Halophilic bacteria Volume I*. Boca Raton, CRC Press 109–140.
107. González C, Gutierrez C, Ramirez C (1978) *Halobacterium vallismortis* sp. nov. an amylolytic and carbohydrate-metabolizing, extremely halophilic bacterium. Can J Microbiol 24, 710–715.
108. Hao MV, Kocur M, Komagata K (1984) *Marinococcus* gen nov., a new genus for motile cocci with meso-diaminopimelic acid in the cell wall; and *Marinococcus albus* sp. nov. and *Marinococcus halophilus* (Novitsky and Kushner) comb. nov. J Gen Appl Microbiol 30, 449–459.
109. Zvyagintseva IS, Tarasov AL (1987) Extreme halophilic bacteria from saline soils. Mikrobiologiya 56, 839–844.
110. Bouchotroch S, Quesada E, del Moral A, Llamas I, Bejar V (2001) *Halomonas maura* sp. nov., a novel moderately halophilic, exopolysaccharide-producing bacterium. Int J Syst Evol Microbiol 51, 1625–32.
111. Kivisto A, Karp M (2011) Halophilic anaerobic fermentative bacteria. Review. J Biotechnol 152, 114–124.
112. Yoon JH, Kang KH, Park YH (2003) *Halobacillus salinus* sp. nov., iso-lated from a salt lake on the coast of the East Sea in Korea. Int J Syst Evol Microbiol 53, 687–93.
113. Pitt J, Hocking A (1985) Fungi and food spoilage, First edition. Academic Press, Sydney.
114. Plemenitaš A, Vaupotic T, Lenassi M, Kogej T, Gunde-Cimerman N (2008) Adaptation of extremely yeast Horteae werneckii to increased osmolarity: a molecular perspective at a glance. Studies in Mycology 61, 67–75.
115. Dave G, Modi H (2013) Phytase producing microbial species associated with rhizosphere of mangroves in an Arid Coastal Region of Dholara. Academia Journal of Biotechnology 1, 027–035.
116. Dhanasekaran P (2014) Anti-infective potential of halophilic actinomycetes isolated from saltpan. Malaya Journal of Biosciences 1, 37–40.
117. Trincone A (2013) Biocatalytic Processes Using Marine Biocatalysts: Ten Cases in Point. Current Organic Chemistry 17, 1058–1066.
118. Margesin R, Schinner F (2001) Potential of halotolerant and halophilic microorganisms for biotechnology. Extremophiles 5, 73–83.
119. Babavalian H, Amoozegar MA, Pourbabaee AA, Moghaddam MM, Shakeri F (2013) Isolation and identification of moderately halophilic bacteria producing hydrolytic enzymes from the largest hypersaline playa in Iran. Microbiology 82, 466–474.
120. Bonfá MR, Grossman MJ, Piubeli F, Mellado E, Durrant LR (2013) Phenol degradation by halophilic bacteria isolated from hypersaline environments. Biodegradation 24, 699–709.
121. Oren A, Bardavid RE, Mana L (2014) Perchlorate and halophilic prokaryotes: implications for possible halophilic life on Mars. Extremophiles 18, 75–80.
122. Parawira W (2012) Enzyme research and applications in biotechnological intensification of biogas production. Critical Reviews in Biotechnology 32, 172–186.
123. Bhat MK, Bhat S (1997) Cellulose degrading enzymes and their potential industrial applications. Biotechnology Advances 15, 583–620.
124. Valderrama B, Ayala M, Vazquez-Duhalt R (2002) Suicide Inactivation of Peroxidases and the Challenge of Engineering More Robust Enzymes. Chemistry & Biology 9, 555–565.
125. Jiang Y, Liu X, Chen Y, Zhou L, He Y, et al. (2014) Pickering emulsion stabilized by lipase-containing periodic mesoporous organosilica particles: A robust biocatalyst system for biodiesel production. Bioresource Technology 153, 278–283.
126. Raddadi N, Cherif A, Daffonchio D, Fava F (2013) Halo-alkalitolerant and thermostable cellulases with improved tolerance to ionic liquids and organic solvents from *Paenibacillus tarimensis* isolated from the Chott El Fejej, Sahara desert, Tunisia. Bioresource Technology 150, 121–128.
127. Pang H, Zhang P, Duan CJ, Mo XC, Tang JL, et al. (2009) Identification of Cellulase Genes from the Metagenomes of Compost Soils and Functional Characterization of One Novel Endoglucanase. Curr Microbiol 58, 404–408.
128. Narra M, Dixit G, Divecha J, Kumar K, Madamwar D, et al. (2014) Production, purification and characterization of a novel GH 12 family endoglucanase from *Aspergillus terreus* and its application in enzymatic degradation of delignified rice straw. International Biodeterioration & Biodegradation 88, 150–161.
129. Li X, Yu HY (2013) Characterization of a halostable endoglucanase with organic solvent-tolerant property from *Haloarcula* sp. G10. International Journal of Biological Macromolecules 62, 101–106.



Capítulo VI.

Caracterización de las actividades celulasa y xilanasa de hongos asociados a la esponja marina *Stelleta normani*.

Los ambientes marinos se han definido como fuente extraordinaria de biocatalizadores y cepas microbianas con utilidad para aplicaciones biotecnológicas. En particular existe un interés creciente por estudiar la microbiota cultivable asociada a las esponjas marinas viviendo en profundades mayores a 500m. De acuerdo a estos criterios estudiamos la comunidad de hongos filamentosos asociados a *Stelleta normani*, y sus actividades celulasas y xilanasas sobre diferentes sustratos: *"Characterization of cellulose and xylanase activities froom deep see sponge associated fungi isolated from Stelletta normani"*.

Tres hongos halotolerantes, psicrotolerantes y celulolíticos fueron estudiados: *Cadophora* sp. TS2, *Emericellopsis* sp. TS11 y *Pseudogymnoascus* sp. TS12. Este constituye el primer reporte de la microbiota fúngica asociada con esta esponja, y de estas actividades lignolíticas en estos géneros de hongos. Los ascomicetos anteriores demostraron capacidad de crecer sobre sustratos ligocelulósicos (paja de trigo y rastrojo de maíz) y expresaron enzimas extracelulares útiles para la deconstrucción de lignocelulosa. Del mismo modo evidenciaron potencialidades para la liberación de azúcares fermentables a partir de celulosa cristalina (fibras de algodón). Curiosamente, los perfiles enzimáticos (celulasas y xilanasas) de TS2, TS11 y TS12 mostraron características interesantes en términos de haloestabilidad, termoestabilidad y temperatura óptima. Los aspectos anteriores señalan que las enzimas lignocelulíticas de estos hongos resultan atractivas para aplicaciones en biorrefinerías de lignocelulosa.

Los ecosistemas marinos profundos son hábitats de peculiar interés debido a sus características abióticas (oxígeno disuelto, presión, temperatura, composición de nutrientes). Los hongos viviendo en estas condiciones se definen como candidatos interesantes para la biotecnología de acuerdo a la plasticidad metabólica que deben presentar.

Este trabajo, junto al anterior (Capítulo V), demuestra la utilidad del estudio y caracterización de aislamientos microbianos. Estos dos capítulos (V y VI) reflejan que también las prospecciones ecológicas clásicas resultan pertinentes para la descripción de actividades enzimáticas (con novedos perfiles) relacionadas con la deconstrucción de lignocelulosa.

PLOS ONE

Characterization of Cellulase and Xylanase Activities from Deep Sea Sponge Associated Fungi Isolated from Stelletta normani

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Article
Full Title:	Characterization of Cellulase and Xylanase Activities from Deep Sea Sponge Associated Fungi Isolated from Stelletta normani
Short Title:	Cellulases and Xylanases from Sponge Associated Fungi from S. normani
Corresponding Author:	Jorge Luis Folch-Mallol, Ph.D. Universidad Autónoma del estado de Morelos Cuernavaca , Morelos MEXICO
Keywords:	Stelletta normani; Cadophora; Emericellopsis; Pseudogymnoascus; cellulases; xylanases; lignocellulosic materials
Abstract:	While bacteria associated with marine sponges have been extensively studied, much less information is available about sponge-derived fungi. Culture dependent approaches were employed to study fungi associated with the deep sea sponge Stelletta normani sampled at a depth of 751 metres. Three halotolerant fungal strains were isolated from the sponge and molecular based taxonomic approaches identified these strains as Cadophora sp. TS2, Emericellopsis sp. TS11 and Pseudogymnoascus sp. TS 12. TS2, TS11 and TS12 displayed psychrotolerance and halotolerant growth on cellulose and xylan as sole carbon sources, with optimal growth rates at 20ºC. They produced cellulase and xylanase activities, which displayed optimal temperature and pH values of between 50-70ºC and pH 5-8 respectively, together with good thermostability and halotolerance. In solid-state fermentations TS2, TS11 and TS12 produced cellulases, xylanases and peroxidase/phenol oxidases when grown on corn stover and wheat straw. This is the first time that cellulase, xylanase and peroxidase/phenol oxidase activities have been reported in these three fungal genera isolated from a marine sponge. Given the biochemical characteristics of these ligninolytic enzymes it is likely that they may prove useful in future biomass conversion strategies involving lignocellulosic materials.
Order of Authors:	Ramón Alberto Batista-García
	Thomas Sutton
	Stephen A. Jackson
	Omar Eduardo Tovar-Herrera
	Edgar Balcazar-López
	María del Rayo Sánchez-Carbente
	Ayixon Sánchez-Reyes
	Alan D.W. Dobson
	Jorge Luis Folch-Mallol, Ph.D.
Opposed Reviewers:	
Additional Information:	
Question	**Response**
Financial Disclosure Please describe all sources of funding that have supported your work. A complete funding statement should do the	This work was supported in part by the PharmaSea project (www.pharma-sea.eu) funded by the EU Seventh Framework Programme (contract no. 312184); by the Beaufort Marine Research Award, part of the Sea Change Strategy and the Strategy for Science Technology and Innovation (2006–2012), with the support of The Marine Institute under the Marine Research Sub-Programme of the National Development Plan 2007–2013; by the Department of Agriculture, Fisheries and Food (DAFF11/F/009 MabS), by the National Council for Science and Technology (CONACyT)-Mexico:

following: Include **grant numbers and the URLs** of any funder's website. Use the full name, not acronyms, of funding institutions, and use initials to identify authors who received the funding. **Describe the role** of any sponsors or funders in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. If they had no role in any of the above, include this sentence at the end of your statement: "*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*" If the study was **unfunded**, provide a statement that clearly indicates this, for example: "*The author(s) received no specific funding for this work.*" * typeset	Project CB-153789; and PRODEP-SEP Project UAEMOR-PTC-333. RAB-G acknowledges the Support of the Europe Molecular Biology Organization (EMBO: ASTF503-2013) for a research visit to University College Cork (Cork, Ireland). RABG together with OETH, EBL and ASR also received a sholarship from the National Council for Science and Technology (CONACyT) in Mexico.
Competing Interests You are responsible for recognizing and disclosing on behalf of all authors any competing interest that could be perceived to bias their work, acknowledging all financial support and any other relevant financial or non-financial competing interests. Do any authors of this manuscript have competing interests (as described in the PLOS Policy on Declaration and Evaluation of Competing Interests)? **If yes**, please provide details about any and all competing interests in the box below. Your response should begin with this statement: *I have read the journal's policy and the authors of this manuscript have the following competing interests:* **If no** authors have any competing interests to declare, please enter this statement in the box: "*The authors have declared that no competing interests exist.*"	I have read the journal's policies and the authors of this manuscript have the following competing interests: JLFM is currently an academic editor for PLoS ONE, but this does not alter the authors, adherence to PLoS ONE editorial policies and criteria.

* typeset	
Ethics Statement You must provide an ethics statement if your study involved human participants, specimens or tissue samples, or vertebrate animals, embryos or tissues. All information entered here should **also be included in the Methods section** of your manuscript. Please write "N/A" if your study does not require an ethics statement. **Human Subject Research (involved human participants and/or tissue)** All research involving human participants must have been approved by the authors' Institutional Review Board (IRB) or an equivalent committee, and all clinical investigation must have been conducted according to the principles expressed in the Declaration of Helsinki. Informed consent, written or oral, should also have been obtained from the participants. If no consent was given, the reason must be explained (e.g. the data were analyzed anonymously) and reported. The form of consent (written/oral), or reason for lack of consent, should be indicated in the Methods section of your manuscript. Please enter the name of the IRB or Ethics Committee that approved this study in the space below. Include the approval number and/or a statement indicating approval of this research. **Animal Research (involved vertebrate animals, embryos or tissues)** All animal work must have been conducted according to relevant national and international guidelines. If your study involved non-human primates, you must provide details regarding animal welfare and steps taken to ameliorate suffering; this is in accordance with the recommendations of the Weatherall report, "The use of non-human primates in research." The relevant guidelines followed and the committee that approved the study should be identified in the ethics statement.	N/A

If anesthesia, euthanasia or any kind of animal sacrifice is part of the study, please include briefly in your statement which substances and/or methods were applied. Please enter the name of your Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board, and indicate whether they approved this research or granted a formal waiver of ethical approval. Also include an approval number if one was obtained. **Field Permit** Please indicate the name of the institution or the relevant body that granted permission.	
Data Availability PLOS journals require authors to make all data underlying the findings described in their manuscript fully available, without restriction and from the time of publication, with only rare exceptions to address legal and ethical concerns (see the PLOS Data Policy and FAQ for further details). When submitting a manuscript, authors must provide a Data Availability Statement that describes where the data underlying their manuscript can be found. Your answers to the following constitute your statement about data availability and will be included with the article in the event of publication. **Please note that simply stating 'data available on request from the author' is not acceptable.** ***If, however, your data are only available upon request from the author(s), you must answer "No" to the first question below, and explain your exceptional situation in the text box provided.*** Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?	Yes - all data are fully available without restriction
Please describe where your data may be found, writing in full sentences. **Your answers should be entered into the box below and will be published in the form you provide them, if your manuscript is accepted.** If you are copying our sample text below, please ensure you replace any instances of XXX with the appropriate details.	All relevant data are within the paper and its Supporting Information files.

If your data are all contained within the paper and/or Supporting Information files, please state this in your answer below. For example, “All relevant data are within the paper and its Supporting Information files.” If your data are held or will be held in a public repository, include URLs, accession numbers or DOIs. For example, “All XXX files are available from the XXX database (accession number(s) XXX, XXX)." If this information will only be available after acceptance, please indicate this by ticking the box below. If neither of these applies but you are able to provide details of access elsewhere, with or without limitations, please do so in the box below. For example: “Data are available from the XXX Institutional Data Access / Ethics Committee for researchers who meet the criteria for access to confidential data.” “Data are from the XXX study whose authors may be contacted at XXX.” * typeset	
Additional data availability information:	Tick here if the URLs/accession numbers/DOIs will be available only after acceptance of the manuscript for publication so that we can ensure their inclusion before publication.

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
CENTRO DE INVESTIGACIÓN EN BIOTECNOLOGÍA

Cuernavaca, Mor, 18 August 2015

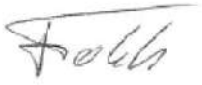
Dear Editor

Here we re submit a new version of a manuscript entitled "**Characterization of cellulase and xylanase activities from deep sea sponge associated fungi isolated from *Stelletta normani*".** The previous version was rejected, but Dr. Raffaella Balestrini was willing to see an improved version of the manuscript, so here we present it. This paper describes the isolation and cultivation of three marine derived fungi from the deep-sea sponge *Stelletta normani* sampled at a depth of 751 metres. Molecular based taxonomic approaches identified the three strains as *Cadophora* sp. TS2, *Emericellopsis* sp. TS11 and *Pseudogymnoascus* sp. TS 12. They display psycrophilic and halotolerant growth on cellulose and xylan as sole carbon sources, with optimal growth rates at 20ºC. They also produced cellulase and xylanases activities with optimal temperature and pH values of between 50-70ºC and pH 5-8 respectively, together with good thermostability and halotolerance. In solid-state fermentations on corn stover and wheat straw they also produce cellulases, xylanases as well as peroxidase/phenol oxidases.

This is the first time that cellulase, xylanase and peroxidase/phenol oxidase activities have been reported in these three fungal genera isolated from a marine sponge. We feel that the biochemical characteristics of the ligninolytic enzymes from these fungi may be useful in future biomass conversion strategies involving lignocellulosic materials.

We have attended all the reviewers' comments to the previous version (a file with the answers to the reviewers comments is uploaded) and made all the necessary changes to the previous version so we hope that this version meets the standards to be published in PloS ONE.

Sincerely,



Jorge Luis Folch-Mallol, PhD

Characterization of Cellulase and Xylanase Activities from Deep Sea Sponge Associated Fungi Isolated from *Stelletta normani*

Ramón Alberto Batista-García[1-2], Thomas Sutton[3], Stephen A. Jackson[3-4], Omar Eduardo Tovar-Herrera[5], Edgar Balcázar-López[1-2], María del Rayo Sánchez-Carbente[2], Ayixon Sánchez Reyes[1-2], Alan D. W. Dobson[3-4]*, Jorge Luis Folch-Mallol[2*]

[1]Centro de Investigación en Dinámica Celular, Universidad Autónoma del Estado de Morelos. Cuernavaca, Morelos, Mexico. [2]Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos. Cuernavaca, Morelos, Mexico. [3]School of Microbiology, University College Cork. Cork, Ireland. [4]Marine Biotechnology Centre, Environmental Research Institute, University College Cork. Cork, Ireland. [5]Instituto de Biotecnología. Facultad de Ciencias Biológicas. Universidad Autónoma de Nuevo León. San Nicolás de los Garza, Nuevo León, Mexico.

*Corresponding authors

Alan D. W. Dobson, PhD.

Tel. +353 (0) 21 4902743

Fax. +353 (0) 21 4903101

Email: a.dobson@ucc.ie

Jorge L. Folch-Mallol, PhD.

Tel. + 52 (777) 3297057

Fax. + 52 (777) 3297030

Email: jordi@uaem.mx

Abstract

While bacteria associated with marine sponges have been extensively studied, much less information is available about sponge-derived fungi. Culture dependent approaches were employed to study fungi associated with the deep sea sponge *Stelletta normani* sampled at a depth of 751 metres. Three halotolerant fungal strains were isolated from the sponge and molecular based taxonomic approaches identified these strains as *Cadophora* sp. TS2, *Emericellopsis* sp. TS11 and *Pseudogymnoascus* sp. TS 12. TS2, TS11 and TS12 displayed psychrotolerance and halotolerant growth on cellulose and xylan as sole carbon sources, with optimal growth rates at 20ºC. They produced cellulase and xylanase activities, which displayed optimal temperature and pH values of between 50-70ºC and pH 5-8 respectively, together with good thermostability and halotolerance. In solid-state fermentations TS2, TS11 and TS12 produced cellulases, xylanases and peroxidase/phenol oxidases when grown on corn stover and wheat straw. This is the first time that cellulase, xylanase and peroxidase/phenol oxidase activities have been reported in these three fungal genera isolated from a marine sponge. Given the biochemical characteristics of these ligninolytic enzymes it is likely that they may prove useful in future biomass conversion strategies involving lignocellulosic materials.

Keywords: *Stelletta normani*; *Cadophora; Emericellopsis; Pseudogymnoascus*; cellulases; xylanases; lignocellulosic materials

1. Introduction

Marine sponges (phylum *Porifera*) are the most primitive pluricellular animals [1] and are widely distributed in polar, tropical and temperate oceans growing at various depths [2]. Their origins date back to the Precambrian period and they have diversified into more than 9,000 species [3]. Their ability to filter up to 50,000 liters of seawater per litre of sponge per day is one of their main ecological roles [4].

Marine sponges often contain microbial communities consisting of symbiotic bacteria, archaea and unicellular eukaryotes in their tissues, which can in some instances constitute up to 35% of the total sponge biomass. The study of microbial communities associated with marine sponges has gained much attention in recent years [5–7]. This has highlighted the significant impact that the associated microbiota have on sponge physiology primarily through their provision of biochemically mediated defence mechanisms, and their role in carbon, sulphur and nitrogen cycling [3,7–9]. To date more than 32 bacterial phyla and candidate phyla, 2 major archaeal lineages and a number of fungal species have been identified from ecological studies of marine sponges [3,8].

Microbial populations associated with different marine sponges from locations geographically distributed from around the world have been studied including the Great Barrier Reef [9], the Red Sea [10], the Mediterranean [11], the northern Atlantic [12], South America [13], China [14] and the Indo Pacific [3]. Bacterial populations have been extensively described, while fungal communities have to date not received the same level of attention. There is a deficit of information regarding the diversity, kinds of interactions, eco-geographical distributions, seasonal successions, host specificity and potential

physiological roles of fungi associated with marine sponges [3,7]. Neither is it known how the sponge-associated fungi establish relationships with other sponge microorganisms (bacteria, algae or other fungi), or with the sponge host; or indeed whether the wide diversity observed with sponge-associated fungi represents symbiosis and/or parasitism [15]. Most studies to date have focused on marine-derived fungi associated with sponges, as a potential novel source of bioactive metabolites with biotechnological applications, including anti-tumor, antibacterial, antiviral, toxin inhibitors, and anti-inflammatory metabolites amongst others [3,7,13] (For review see [16]).

Although there is still much to be done, the few works conducted report a large diversity of fungi in marine sponges [15,17–19]; with *Eurotiales, Capnoidales, Pleosporales* and *Hypocreales* orders being identified as commonly associated with sponge species [3]. It is also known that the genera *Aspergillus* and *Penicilium* can be found as ubiquitous marine-derived fungi associated with sponges at different depths. Other genera which are frequently found include: *Alternaria, Acremonium, Beauveria, Cladosporium, Curvularia, Eurotium, Fusarium, Gymnascella, Paecilomyces, Petriella, Pichia, Spicellum* and *Trichoderma* [3].

Marine derived fungi play an important role in detritus processing in various marine ecosystems, and may be important in degrading ligninocellulose in marine environments, with lignocellulose degradation being reported in over 30 phylogenetically diverse marine fungal strains [15,20,21]. Indeed marine derived fungi are well recognized as a good source of enzymes of potential industrial interest (cellulases, xylanases, phenol oxidases, laccases, etc.) with strains exhibiting hydrolytic and oxidative activities being reported; for review see [22].

These fungal strains have predominantly been isolated from seawater, sediments, mangrove detritus and to a lesser extent marine sponges [17,23,24]. The growing demand for new and robust cellulases, xylanases and phenol oxidases for biotechnological purposes (*i.e.*, for the biofuel production) justifies the isolation and characterization of new fungal strains with lignocellulolytic properties, especially those living in deep sea. While there have been reports of cellulases being produced from marine derived fungi such as *Chaetomium indicum* and mangroves isolates such as *Hypoxlon oceanicum*, *Julella avicenniae*, *Lignincola laevis*, *Savoryella lignicola* and *Trematosphaeria mangrovei* [25,26], there are limited examples of cellulases being described from fungi isolated from marine sponges. One such report involves members of the phyla Ascomycota and Basidomycotina isolated from *Haliclona simulans*, which produced cellulase activities that were higher than the cellulase overproducing mutant *Hypocrea jecorina* QM9414 [27]. Xylanases from marine derived fungi have also been reported. These have been isolated from soft corals [28], marine sediments [29] and from shallow water marine sponges in Antarctica [30]. Many of these possess interesting biochemical characteristics with potential utility in biotechnological applications.

With this in mind we focused on the characterization of the lignocellulolityc activity of fungi isolated from the deep sea sponge *Stelletta normani* which had previously been studied with respect to its resident microbiota [7]. *Stelletta* is a genus of demosponge belonging to the family *Ancorinidae* and comprises more than 130 different species. Bacterial diversity has been described for a number of other *Stelletta* species including *S. kallitetilla* [31]*, S. maori* [6], and *S. pudica* [31]; but there are no reports to date on the fungal communities associated with

the genus *Stelletta* or with *S. normani* in particular. Given that *S. normani* had been collected at a depth of 751 metres we reasoned that any fungi cultivated from the sponge are likely to possess novel biochemistry, enabling them to survive at the extremes of temperature, light and pressure present at these depths [32].

In this work we describe the isolation and characterization of three lignocellulolityc-halotolerant fungal strains from *S. normani,* which were identified as *Cadophora* sp. TS2, *Emericellopsis* sp. TS11 and *Pseudogymnoascus* sp. TS12. The strains displayed psychrotolerance and halotolerant growth on cellulose and xylan, with optimal growth rates at 20ºC. They displayed thermostable and halotolerant cellulase and xylanase activities at optimal temperature and pH values of 50-70ºC and pH 5-8 respectively, while also successfully colonized maize stover and wheat straw. Given the biochemical characteristics of these enzymes it is likely that they may prove useful in future biomass conversion strategies involving lignocellulosic materials.

2. Materials and methods

2.1 Sponge Sampling

Specific permission was not required to obtain the marine sponge samples used in this study, as they were collected in Irish territorial water, by an Irish research vessel, funded by the Irish government. The sponge samples do not involve endangered or protected sponge species. The sponge samples (*Stelletta normani*: Class *Demospongiae*, Order *Astrophorida*, Family *Ancorinidae*) were collected on June 2013, from a depth of 751 m from Irish waters in the North Atlantic Ocean (53.9861;-12.61) with the remotely operated vehicle (R.O.V) *Holland I* on board the *R.V. Explorer.* The tissue samples were obtained *in situ* by excision of a piece (1-5 g) of one sponge and the species was identified by Bernard Picton (Ulster Museum) and Christine Morrow (Queens University Belfast). Upon retrieval, the sponge samples were washed with sterile artificial seawater (ASW) (33.3 g/L Instant Ocean, Aquatic Eco-Systems, Inc., Apopka, FL, USA) [7]. Samples were then placed in sterile plastic Ziploc bags and stored on dry ice for transport and subsequently used for fungal isolation.

2.2 Fungal isolation

Ten grams of sponge were macerated in 5 mL of sterile ASW, placed in a tube with sterile glass beads and vortexed. Primary isolation of fungi was performed by taking 1 mL of the macerated material with serial dilutions using sterile ASW being performed up to 10^{-5}. One hundred µL of each dilution was inoculated on Petri plates containing either Malt extract agar-ASW or Potato dextrose agar-ASW (DIFCO) and cultures were incubated for 20 days at 20ºC, as previously described [17]. The cultures were observed daily and pure cultures were

obtained. The fungi were stored at 4ºC in saline solution (0.5% NaCl) supplemented with glycerol (20%). Isolations were performed in triplicate.

2.3 Taxonomic identification of fungal strains

Fungal mycelium from the strains TS2, TS11 and TS12 growing on Malt extract agar-ASW plates was collected following 10-days growth for genomic DNA isolation as previously described [33]. Four molecular markers (fragments of: 18S ribosomal DNA [34], 28S large sub-unit RNA gene (D1-D2) [35] and internal transcribed spacers 1 and 2 (ITS1 and ITS2 regions) [35]) previously described to be distinctive for the accurate molecular taxonomic identification of filamentous fungi were analysed. These gene fragments were amplified by PCR using the primers and conditions, as previously described [34,35].

Amplicons were analysed by electrophoresis (1% agarose gel) and purified using a commercial gel extraction kit (Thermo Catalogue No. K0513). Sanger sequencing was performed with the same primers used for the PCR reactions. The sequences obtained were deposited in National Centre for Biotechnology Information (NCBI) under accession numbers KR336667 to KR336677. The sequences were analysed by BlastN using the NCBI website (www.ncbi.nih.gov) and phylogenetic analysis was performed online with the server Phylogeny.fr (www.phylogeny.fr) [36,37]. From the different sequences retrieved from the BLAST hits, one of each species was taken as a reference strain to construct the phylogenetic trees.

2.4 Growth rate determination

The specific growth rate of the three fungi (TS2, TS11 and TS12) (expressed as mm/day) was determined by inoculating plugs (7 mm in diameter) obtained from fungal pre-cultures grown in Vogel's medium [38] supplemented with either 2%

191 carboxymethycellulose (CMC) (Sigma) or 2% xylan (Sigma) in the same media
192 and incubating them at different temperatures (4, 20, 30, and 40ºC). The growth
193 rate determination in saline conditions (NaCl) was determined by adding NaCl
194 to the growth medium (0.5, 1.0, 1.5 and 2.0 M, final concentration). Cultures
195 were also grown without added NaCl. The diameter of the colony was
196 measured every 24 h for 15 days. Experiments were performed in triplicate for
197 subsequent statistical analysis of data.

198 **2.5 Fungal liquid cultures with CMC and xylan as sole carbon source at**
199 **different NaCl concentrations. Cellulase and xylanase activity**
200 **determinations.**

201 Cellulase and xylanase activities were measured both qualitatively and
202 quantitatively. For qualitative determinations, 10-day cultures of the fungi (TS2,
203 TS11 and TS12) grown on agar Vogel's medium supplemented with 2% CMC
204 or 2% xylan and NaCl (0, 0.5, 1.0, 1.5 and 2.0 M, final concentration) at 30ºC
205 were flooded with 15 mL Congo red (1% *v/v* diluted in distilled water) for 15 min
206 and washed 3 times with 20 mL of NaCl solution (1M). Cellulase and xylanase
207 activities were observed as discoloration halos around the fungal colonies [39].
208 Determinations were performed in triplicate.

209 For quantitative determination of both activities (cellulase and xylanase), 500
210 mL Erlenmeyer flasks with 100 mL Vogel´s medium (with 2% CMC or 2% xylan
211 added as a carbon source) supplemented with NaCl (0, 0.5, 1.0, 1.5 and 2.0 M,
212 final concentration) were inoculated with plugs (7 mm in diameter) from pre-
213 cultures of the fungi (TS2, TS11 and TS12) grown in the same solid media, and
214 incubated for 10 days at 30ºC and shaking at 150 rpm. Supernatants were

recovered by centrifugation at 10,300 g for 10 min. Protein concentration was determined every 24 h using the Lowry method [40].

Enzymatic activities were also calculated every 24 h employing the 3,5-dinitrosalysilic acid (DNS) assay [41] and were expressed as IU/mg protein. The analytical procedures and volumes employed in the reaction mixture to determine both activities were as previously described [42]. Briefly, for cellulases and xylanases activity measurements, CMC and oat xylan (both 2%) dissolved in citrate buffer (50 mM and pH 5) were used as substrates. In each case, the reaction mixture contained supernatant from the liquid media (200 µL), 50 mM citrate buffer pH 5 (300 µL) and substrate solution (500 µL). The reaction was incubated at 50ºC for 30 min and was monitored every 5 min. In summary, 50 µL aliquots were taken, mixed with DNS solution (50 µL), boiled for 5 min and then cooled on ice. The absorbance was measured at λ 540 nm in a spectrophotometer (BioMate, ThermoSpectronic). Glucose or xylose standard curves (ranging from 0.1 to 2.0 mg/mL) were used to extrapolate the reducing sugar concentrations and the slopes were calculated to determine the velocity of the reaction. The concentration of released reducing sugars *vs.* time was used to calculate enzymatic activities. One international unit (IU) was defined as 1 µmol of glucose or xylose equivalent released per minute, under the assay conditions. Triplicate independent assays were performed and three readings for each sample were taken in all cases.

The supernatants recovered from these liquid cultures were used to study the influence of the temperature and pH on cellulase and xylanase activities, and both, thermostability and halotolerance.

2.6 Optimal temperature and pH of cellulase and xylanase activities

240 Enzymatic reactions were performed as described earlier at different incubation
241 temperatures (1, 10, 20, 30, 40, 50, 60, 70 and 80ºC) in 50 mM sodium citrate
242 buffer, pH 5. Different pH conditions ranging from 2 to 10 were tested at the
243 optimal temperature in each case in citrate (2 to 6) or phosphate (7 to 10) buffer
244 depending on the pH being tested. All measurements were determined in
245 triplicate.

246 **2.7 Thermal-stability of the cellulase and xylanase activities**

247 An aliquot (500 µL) of each supernatant from liquid cultures was incubated at
248 30, 40, 50, 60, 70 and 80ºC for 1 h. Subsequently, supernatants were cooled on
249 ice for 5 min. Following this, the enzymatic activities from the heat treated
250 supernatants, were determined using the optimal conditions (temperature and
251 pH) for each activity obtained as described in the previous section. The residual
252 activities expressed in percentages were reported. All measurements were
253 performed in triplicate.

254 **2.8 Cellulase and xylanase activities in salinity conditions**

255 Cellulase and xylanase activities in different salinity conditions were determined
256 as previously described (employing the optimal temperature and pH for each
257 enzyme) with the addition of NaCl to a final concentration of 0.5, 1.0, 2.0 and
258 3.0 M in the reaction mixture. All measurements were performed in triplicate.

259 **2.9 Solid-state fermentation on maize stover and wheat straw**

260 Solid-state fermentations were performed using the following autoclaved
261 substrates: maize stover (*Zea mays*) and wheat straw (*Triticum aestivum*).
262 These substrates were selected because large quantities are produced
263 worldwide and they have also previously been shown to be very useful for
264 lignocellulolytic enzymes production [42]. Erlenmeyer flasks of 500 mL including

5 g of each substrate were inoculated with two plugs (7 mm in diameter) of each fungal strain (TS2, TS11 and TS12); which had previously been grown on Malt extract agar-ASW plates. Humidity in the system was maintained by adding 2.5 mL (50% w/v) Vogel´s solution to the solid substrates. Fermentation was allowed to take place at 30ºC for 12 days. Subsequently, the cultures (substrates with fungal growth) were collected, washed with 5 mL (50 mM citrate buffer pH 6) and the soluble fermentation products were recovered by filtration used Whatman® filter paper and centrifugation at 10,300 g for 20 min. The soluble products recovered from the solid-state fermentations were used to determine cellulase, xylanase, peroxidase and phenol oxidase activities. Also, saccharification experiments on cotton fibres were conducted, and zymograms for cellulases and xylanases were performed.

2.10 Enzymatic activities from solid-state fermentations

Cellulase and xylanase activities were determined as mentioned before (Review Cellulase and xylanase activity determinations section).

Peroxidase and phenol oxidase activities were tested according to the previously reported method using 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid (ABTS) as substrate for both complexes [43,44]. The ABTS concentration in the mixture was 2 mM and assays were performed in a 300 µL final volume. For phenol oxidase activity, reactions containing supernatants (50 µL) from fungal cultures grown under solid-state fermentation, ABTS (10 µL), and 50 mM acetate buffer pH 5 (240 µL) were used. Hydrogen peroxide (0.3%) was used for peroxidase determinations. In this case, 50 µL of supernatant, 10 µL of ABTS, 237 µL of acetate buffer and 3 µL of H_2O_2 were employed in the reactions. Enzyme assays were performed following published microplate (96

290 well) protocols [43,45]. Each reaction was incubated at room temperature for 5
291 min and then, the oxidation rate of ABTS to $ABTS^{+}$ released was measured at
292 436 nm [43]. The ABTS molar extinction coefficient used was 73,000 mM^{-1} cm^{-1}
293 and the calculations were performed as previously described [43,44]. The
294 volumetric activities were obtained in IU defined as μmoles of $ABTS^{+\cdot}$ formed
295 from ABTS min^{-1} (U) per mL^{-1}, and expressed as specific activities (IU/mg
296 protein) considering the protein concentration in each sample.
297 Triplicate independent assays were performed and three readings for each
298 sample were taken in all cases.

299 **2.11 Zymograms**

300 Zymograms were performed to identify cellulase and xylanase isoforms from
301 soluble products recovered from the solid-state fermentations from the three
302 strains (TS2, TS11 and TS12). Soluble products from the fermentations were
303 obtained as previously mentioned (see solid-state fermentation section).
304 Zymograms were performed as described in [42,46]. Briefly, 20 μg of protein
305 (without 2-mercaptoethanol and prior boiling) were loaded per lane in 10%
306 polyacrylamide gel. Gels did not contain SDS, but SDS was added at 0.05% to
307 the running buffer and after the run the gels were washed three times (40 min
308 each) in PCA buffer (50 mM KH_2PO_4, 50 mM citric acid pH 5.2) in order to
309 remove SDS. For cellulases, 2% CMC was added and co-polimerized with the
310 gel, while 2% oat xylan was used in the gel for xylanases. Once the
311 electrophoresis was carried out, gels were washed to remove the SDS and then
312 incubated with a 1% Congo red solution (in water) for 30 min at room
313 temperature and then washed 3 times with a 1 M NaCl solution. The cellulase
314 and xylanase activities developed as clear bands, as the substrate is degraded

315 resulting in loss of dye binding. The molecular weight of the bands was
316 estimated against a protein marker (Fermentas Catalogue No. 26612).

317 **2.12 Release of fermentable sugars from cotton fibres**

318 The saccharification potential of the fungi was evaluated, with each strain being
319 grown in Vogel´s medium supplemented with CMC or xylan, and the
320 supernatants recovered by centrifugation at 10,300 g for 10 min. Briefly, 5 mg of
321 cotton fibres (pharmaceutical-grade) was treated with 25% NaOH for 15 min at
322 4ºC, and subsequently washed five times with sterile distilled water to remove
323 the alkali [47]. The same amount of protein (20 µg in total) from different
324 cultures was incubated with the cotton fibres at the optimal temperature and pH
325 in each case: *(i)* supernatants from fungal cultures grown in Vogel´s medium
326 supplemented with 2% CMC, *(ii)* supernatants from fungal cultures grown in
327 Vogel´s medium supplemented 2% xylan and *(iii)* mix of proteins (1:1) from both
328 of the previous cultures *(i)* and *(ii)*. Additionally a negative control treatment was
329 performed (cotton fibres incubated with phosphate buffer 0.1 M pH 5), and
330 aliquots of 50 µL were taken at 1, 2 and 3 h following treatment. The release of
331 fermentable sugars was determined using the DNS method as previously
332 described [41].

333 **2.13 Statistical calculations**

334 For statistical treatment of experimental data, the arithmetic mean and the
335 standard deviations were calculated. Simple classification ANOVA (variance
336 analysis) tests were applied to determine significant differences between the
337 different cases. Firstly, the assumptions of ANOVA were revised: analysis of
338 homogeneity of variance (Hartley-Cochran-Bartlett test) and normal distribution
339 (Kolmogorov-Smirnov and Lilliefors tests) were performed. Subsequently

340 ANOVAs were conducted to demonstrate the similarities or differences between
341 the data of the population of samples. Finally, a post hoc analysis that defines
342 the order of the differences found in the ANOVAs was developed. The Fisher
343 LSD, Tukey HSD and Duncan tests were performed for the post hoc analyses.
344 The use of these three tests ensures greater statistical robustness of the
345 proposed analysis. All statistical calculations were performed in Statistica v12.6
346 (https://support.software.dell.com/statistica/download-new-releases).
347

3. Results and Discussion

3.1 Isolation and identification of fungal strains

Fourteen different strains were isolated from the *Stelletta normani* sponge samples. In all dilutions tested (serial dilutions up to 10^{-5}), TS2, TS11 and TS12 isolates were the most abundant and the preliminary screening for cellulase and xylanase activities showed them to be the most promising, as judged by the strong activity they showed. These criteria lead us to keep working with these species. Genomic DNA was isolated from the strains to amplify and sequence the molecular markers ITS1, ITS2 and D1-D2 regions (from 28S ribosomal subunit) and 18S rDNA from each of the fungal strains. This allowed us to taxonomically identify each of the three fungal strains. This phylogenetic analysis allowed us to propose that the TS2 isolate belongs to the *Cadophora* genus, with TS2 grouping directly with other *Cadophora* species (Fig. 1), being most phylogenetically related to *Cadophora fastigiata* and *Cadophora malorum* (Fig. 1 and S1 Figure). *Cadophora* spp. have been isolated from diverse regions around the globe (United States, South Africa, Uruguay, Spain, Sweden and Canada) [48]. They have also previously been reported in Antarctica, where extreme weather conditions including UV radiation and high salt concentrations are thought to influence their growth [49,50]. To the best of our knowledge, this is the first time that a *Cadophora* species has been isolated from a marine sponge, which may help provide new insights into the distribution and various ecological niches of this fungal genus.

Fig. 1 Phylogenetic reconstruction for the strain TS2. (A) Phylogeny for D1-D2 region. (B) Phylogeny for ITS2. (C) Phylogeny for ITS1. D1-D2, ITS2 and ITS1

372 sequences were annotated in the NCBI under accession numbers KR336670,
373 KR336675 and KR336672 respectively.

374 **S1 Figure** TS2 18S rDNA gene phylogeny. Sequence annotated under
375 accession number KR336667.

376 Phylogenetic analysis suggests that TS11 belongs to the *Emericellopsis* genus,
377 with the phylogenetic trees employing different molecular markers grouping
378 TS11 with different *Emericellopsis* species (Fig. 2 and S2 Figure). According to
379 these amplified sequences (ITS1, ITS2 and 18S rDNA) the closest related
380 species to TS11 were *Emericellopsis alkalina, Emericellopsis maritima* and
381 *Emericellopsis pallida* (Fig. 2 and S2 Figure). Although *Emericellopsis* spp.
382 have been reported to share high homology with *Acremonium* spp. when
383 comparing ITS and β-tubulin genes [51], the ITS2 primers used here were able
384 to resolve the placement of TS11 to the *Emericellopsis* genus (Fig. 2).

385 **Fig. 2** Phylogenetic reconstruction for the strain TS11. (A) Phylogeny for ITS1.
386 (B) Phylogeny for ITS2. ITS1 and ITS2 sequences were annotated in the NCBI
387 under accession numbers KR336673 and KR336676 respectively.

388 **S2 Figure** TS11 18S rDNA gene phylogeny. Sequence annotated under
389 accession number KR336668.

390 *Emericellopsis* species have previously been isolated from a number of
391 locations throughout the world and from several sources (skin from reptiles,
392 agricultural and forest soils, peat, rhizomes, prairies, freshwater, estuarine and
393 marine-mud sediments) with isolates from terrestrial and marine origins
394 commonly forming different clades [51,52]. Moreover *Emericellopsis* spp. have
395 also been isolated from different marine macroalgae and sponges and have
396 previously been studied due to their ability to produce non-ribosomal peptide

antibiotics [3,53–57]. They have also recently been reported to influence the community structure in photosynthetic microbial mats, by degrading the top photoautotrophic layer of these intertidal hypersaline mats [58].

Regarding strain TS12, the phylogenetic analysis with the four molecular markers employed (ITS1, ITS2, D1-D2 region and 18S rDNA), revealed that it belongs to the poorly defined *Pseudogymnoascus* genus (anamorph representatives of the *Geomyces* genus) (Fig. 3 and S3 Figure); with *Geomyces pannorum, Pseudogymnoascus destructans* and *Pseudogymnoascus verrucosus* being the closest related species to TS12 (Fig. 3 and S3 Figure). *Pseudogymnoascus* spp. have previously been isolated from Antarctica with *Pseudogymnoascus* sp. being isolated from Antarctic soils [59] and *Pseudogymnoascus* sp. F09-T18-1 which has been shown to produce novel nitroasterric acid derivatives being isolated from the Antarctic sponge *Hymeniacidon* sp. [60].

Fig. 3 Phylogenetic reconstruction for the strain TS12. (A) Phylogeny for D1-D2 region. (B) Phylogeny for ITS1. (C) Phylogeny for ITS2. D1-D2, ITS1 and ITS2 sequences were annotated in the NCBI under accession numbers KR336671, KR336674 and KR336677 respectively.

S3 Figure TS12 18S rDNA gene phylogeny. Sequence annotated under accession number KR336669.

A number of molecular markers have previously successfully been used for the taxonomic identification of fungal genera and species, including: the 18S rDNA gene [34], the mitochondrial cytochrome *b* gene [61], a putative toxin pathway regulatory gene (i.e. *aflR*) [62], the DNA topoisomerase II gene [63], the β-tubulin gene [64], the ITS regions between the small- and large-subunit rDNA

genes [65], and the variable regions D1-D2 from 28S rDNA [35,65]. In particular the D1-D2 region from 28S rDNA and both, ITS1 and ITS2 have previously been successfully employed for fungal taxonomy purposes [35,42,66], as was the case here where they allowed the identification of our three fungal strains TS2 (*Cadophora*), TS11 (*Emericellopsis*) and TS12 (*Pseudogymnoascus*) to the genus level.

The role of these fungi in the deep sea sponge *S. normani* is not clear, but given that sponges are filter feeders, and that then they are likely to be exposed to pollutants that may be present in the seawater, and may accumulate impurities from phytoplankton, or other suspended matter; it is tempting to speculate as others have, that some fungi in the sponge may produce degradative/hydrolytic enzymes to acquire nutrients from these materials [67]. In addition as previously mentioned, marine derived fungi are believed to play an important role in detritus processing and in lignocellulose degradation, so these fungi may also play a role in ligninolytic processes. Indeed a number of marine-derived fungi such as *Aspergillus sclerotiorum*, *Cladosporium cladosporioides* and *Mucor racemosus*, isolated from cnidarian samples in Sao Paulo Brazil, have been shown to produce high levels of lignin peroxidase, manganese peroxidase and laccase activity; all of which are important in fungal lignin degradation [23]. While *Marasmiellus* sp. and *Tinctoporellus* sp. isolated from the Brazilian sponges *Amphimedon viridis* and *Dragmacidon reticulata* have also been reported to be good producers of laccases [24]. With this in mind we focused our attention on the ligninolytic ability of the three fungal strains *Cadophora, Emericellopsis* and *Pseudogymnoascus*, particularly given that ligninolytic enzymes from marine-derived fungi are likely to find utility in

447 biotechnological applications with alkaline pH, high salinity, low temperature,
448 oligotrophic conditions, low water potential and hydrostatic pressure, such as
449 amongst others- the treatment of coloured industrial effluent and bioremediation
450 in high salt concentration environments, biofuel production, deinking, laundry
451 and in the food industry [3,22,68].

452 **3.2 Growth rate determination**

453 The growth rate of the three fungal strains under different temperature
454 conditions (4, 20, 30 and 40ºC) and salinity (0.5, 1.0, 1.5 M, final concentration)
455 on different carbon source (cellulose or xylan) was evaluated (Fig. 4).

456 **Fig. 4** Growth rates for TS2, TS11 and TS12 at different temperatures and
457 NaCl concentrations. Different letters (as superscripts) mean statistically
458 significant differences derived from the ANOVA analysis.

459 The fungi were able to grow on cellulose in the absence of NaCl at 4 and 30ºC
460 with optimal growth being observed at 20ºC; TS2 (4.56 ± 0.13 mm/day), TS11
461 (7.11 ± 0.11 mm/day) and TS12 (5.85 ±0.15 mm/day) (Fig. 4). Optimal growth
462 rates were also observed at 20ºC when the fungi were grown on cellulose in the
463 presence of 0.5 and at 1.0 M NaCl in the case of both TS11 and TS12
464 suggesting that these fungi are psychrotolerant in nature, when grown on
465 cellulose.

466 The ability of these three fungi to use cellulose as a sole carbon source may
467 reflect their habitats, with *Emericellopsis* being associated with marine algae
468 species were cellulose is found in algal cell walls. With respect to
469 *Pseudogymnoascus*, which is a member of the *Pseudoeurotiaceae* fungal
470 family; many members of which grow saprotrophically on woody tissues and

rotting vegetation and it is widely believed that they can degrade cellulosic substrates [69], while *Cadophora* is a natural wood-decay fungus [48,51,70].

With respect to fungal growth with xylan as a sole carbon source in the absence of NaCl, similar patterns were observed to those observed with cellulose as a carbon source. Although no growth was detected in any of the strains at 40ºC, growth was observed between 4 and 30ºC with optimal growth being observed at 20ºC; TS2 (4.12 ± 0.14 mm/day), TS11 (7.43 ± 0.06 mm/day) and TS12 (7.90 ±0.08 mm/day) (Fig. 4). Regarding the potential effect of salt on fungal growth, while no growth was observed at 1.5 M NaCl, the addition of 0.5 M NaCl to any of the fungi growing on either cellulose or xylan as a sole carbon source did not significantly affect ($p \leq 0.05$) their growth at any temperature (Fig. 4). However, increasing the salt concentration to 1.0 M had a noticeable negative effect (~2 to 5 and ~1.7 to 6 mm/day less than control treatment in all cases, on cellulose and xylan respectively) on the daily growth rates of TS2, TS11 and TS12 (Fig. 4). Interestingly, no growth was observed when TS11 and TS12 were grown at 40ºC and 1.0 M NaCl, while limited growth was observed in TS2 (Fig. 4), confirming the psychrotolerant nature of these fungi, and suggesting that while these fungi may exhibit halotolerance they are not thermotolerant under the conditions evaluated here. While strains TS2, TS11 and TS12 can be classified as halotolerant they are not halophilic given that their optimal growth parameters do not correspond to hypersaline conditions [71]. Given that these fungi were clearly capable of utilizing both cellulose and xylan, which together with lignin are the mainly components of plant cell walls, we decided to study their cellulolytic and xylanolytic production potential.

3.3 Cellulolytic and xylanolytic potential

An initial Congo Red based plate assay with TS2, TS11 and TS12, indicated that they possessed both cellulolytic and xylanolytic activity (data not shown). Once the cellulolytic and xylanolytic potential of the strains was confirmed, a quantitative assay to monitor enzyme production in the three strains was performed. Maximum enzymatic activity for both enzymes was observed at day 9 of fermentation: cellulase activity (TS2: 8.11 ± 1.12, TS11: 3.89 ± 0.41, and TS12: 7.09 ± 0.66) and xylanase activity (TS2: 4.15 ± 0.61, TS11: 11.52 ± 1.28, and TS12: 9.2 ± 0.97) (specific activities measured at 30ºC and pH 5, and expressed as IU/mg protein). For this reason, the supernatants from day 9 were subsequently used to determine the optimum temperature and pH for xylanase and cellulase activity and to determine the halotolerance and thermostability of the enzymes.

3.4 Effect of temperature and pH

The optimal pH range for fungal cellulases and xylanases is generally in the pH range from 4 to 6 [72–75]. Thus we assessed the effect of the temperature on cellulase and xylanase activity at pH 5. The optimum temperature for both cellulase and xylanase activity in each of the three fungi is shown in Fig. 5. In the case of TS11 and TS12 optimal xylanase activity was observed at 50ºC. In contrast optimal xylanase activity in TS2 was observed at 30ºC, with no activity being observed above 50ºC. These activities are similar to those previously described for fungal xylanases. Optimal cellulase activity in TS11 and TS12 was observed at 60ºC, while optimal activity in that TS2 was observed at 70ºC. The cellulase activity observed in TS2 at 70ºC was significantly higher ($p \leq 0.05$), than the activity observed in the other two fungal isolates (Fig. 5). The optimal temperature observed for the cellulases of these three species is unusual in two

521 ways: firstly, most of the fungal cellulases show a slightly lower optimal
522 temperatures (50ºC) [75] and secondly, it is worth noting that in TS2 the optimal
523 cellulase temperature is markedly higher than the environment in which the
524 fungus was isolated.

525 **Fig. 5** Optimal temperature for both cellulase and xylanase activities.

526 The optimal pH for both cellulase and xylanase activity in each fungal strain was
527 then assessed (Fig. 6). The optimal pH for cellulase and xylanase activity in
528 both TS11 and TS12 was between pH 5-6. While the optimal pH for cellulose
529 activity in TS2 was also pH 6 the optimal pH for xylanase activity was more
530 alkaline (pH 8), this being also somewhat unusual (Fig. 6). Xylanases active at
531 high pH are of great interest in the pulp and paper industry, since they can
532 reduce the consumption of chlorine chemicals [76].

533 **Fig. 6** Optimal pH for both cellulase and xylanase activities.

534 **3.5 Thermostability of cellulases and xylanases**

535 The thermostability of the fungal enzymes was subsequently evaluated (Fig. 7).
536 TS11 produced the most thermostable cellulase activity retaining up to 28.38%
537 of residual activity after incubating at 80ºC for 1 h, while TS12 and TS2 retained
538 lower levels of cellulase activity of 15.90% and 6.79% respectively.

539 **Fig. 7** Thermostability of cellulase and xylanase activities.

540 In marked contrast while TS11 again produced the most thermostable xylanase
541 activity retaining over 50.66% activity after 1 h at 70ºC and up to 11.61% activity
542 at 80ºC, TS12 showed no xylanase activity after incubation of 50ºC for 1 h,
543 while TS2 displayed 28.45% residual activity at that temperature, but no activity
544 at 60ºC. Thus TS11 produces the most thermostable enzymes and may prove a
545 suitable candidate in relation to genetic engineering, with the aim of further

546 improving the observed thermostability. It is intriguing that this species produces
547 thermostable enzymes given that it was isolated from a sponge at a depth of
548 751 m, where temperatures are around 3ºC. However other extremophile fungi
549 have previously been isolated from non-extreme habitats [42].

550 **3.6 Halotolerance of cellulolytic and xylanolytic activities**

551 The halotolerance of both cellulases and xylanases activities was evaluated by
552 the addition of different salt (NaCl) concentrations to the reaction mixture (Fig.
553 8). For cellulolytic halotolerance, 0.5 M NaCl had no significant effect ($p \leq 0.05$)
554 on the enzymatic activity when comparing to no addition of salt. However when
555 the NaCl concentration was increased to 1 M a significant decrease in
556 cellulolytic activity was observed for the three fungal strains (from 7.19 and 7.66
557 IU/mg protein for TS11 and TS12 to 4.05 and 2.60 IU/mg protein, respectively),
558 with the most marked decrease occurring in TS2 (decrease from 14.06 to 2.48
559 IU/mg protein). Further increases in NaCl concentrations to 2 and 3 M showed
560 similar effects (Fig. 8).

561 **Fig. 8** Halotolerance for both cellulase and xylanase activities.

562 In the case of xylanolytic activity, no statistical difference ($p \leq 0.05$) in relation to
563 enzymatic activity in any of the fungal strains was observed at either 0 or 0.5 M
564 NaCl concentrations. However, increases in xylanolytic activity in TS11 and
565 TS12 from 21.23 to 26.63 IU/mg protein and from 12.07 to 14.50 IU/mg protein
566 respectively, were observed in the presence of 1 M NaCl. Conversely,
567 xylanolytic activity in TS2 decreased from 5.63 IU/mg protein to almost non-
568 detectable levels, when the salt concentration was above 1 M. Moreover,
569 xylanolytic activity in TS11 continued to be high when the NaCl concentration

570 increased to 2 M, and decreased thereafter at 3 M; which was in marked
571 contrast to the effect exhibited in TS2 and TS12 (Fig. 8).
572 The three fungal strains isolated in this work have previously been reported to
573 produce cellulases [70,77], with *Cadophora malorum* isolated from an
574 expedition hut on Ross Island, Antarctica exhibiting strong cellulase activity
575 (>100 IU/mg protein) at psychrophilic temperatures (4 and 15ºC) [77].
576 *Emericellopsis* spp. have also been shown to produce cellulase activity on solid
577 media [78] while cellulases are thought to play a role in the infection of bats by
578 *Pseudogymnoascus* spp. [58]. However this is the first report of cellulase
579 activity and indeed of xylanase activity in these three fungi isolated from a
580 marine sponge.
581 **3.7 Saccharification of cotton fibres with enzyme crude extracts**
582 Given that the three fungi exhibited cellulolytic and xylanolytic activities we
583 evaluated their saccharification potential with mercerized cotton fibres (Fig. 9).
584 Supernatants of TS11 grown with xylan as carbon source (in which both
585 cellulases and xylanases can be produced, see below) exhibited the best
586 saccharification of the mercerized cotton fibres, with 1.75 µmol of glucose being
587 released after 3 h incubation; which was 2.4 and 2.0 times more glucose
588 released than TS12 and TS2 supernatants, respectively (Fig. 9). While cotton
589 fibres have no hemicellulose, the cellulase secretion or the presence of
590 xylanases with cellulase activity in the supernatants can justify the fermentable
591 sugars observed when cotton fibres were incubated with supernatants of fungal
592 cultures grown on xylan. When cellulose was used as a carbon source, TS2
593 exhibited the best potential, releasing 5.37 µmol of glucose after 3 h incubation;
594 which is 1.8 and 1.4 times more than TS12 and TS11, respectively (Fig. 9). The

595 fibre saccharification using supernatants from all fungi grown in presence of
596 cellulose was higher than those obtained when cotton was incubated with
597 supernatants collected of fungal cultures grown with xylan as carbon source.
598 These results are consistent with the chemical composition of the fibers. No
599 increase in saccharification of the cotton fibres was observed when
600 supernatants from both cellulose and xylan cultures were mixed together. The
601 degree of saccharification obtained with supernatants from the cellulose
602 cultured fungi is comparable to those obtained by our group using commercial
603 cellulases preparations on natural substrates as wheat straw [79],
604 demonstrating that these fungi clearly display interesting biotechnological
605 potential to prepare such cellulases cocktails.

606 **Fig. 9** Saccharification of cotton fibres. SCCX: Supernatants Colleted from
607 Cultures grown with Xylan. SCCC: Supernatants Colleted from Cultures grown
608 with Cellulose.

609 **3.8 Solid-state fermentation**

610 Thus due to the potential utility of these three fungi in biomass conversion we
611 assessed the production of cellulase, xylanase, peroxidase/phenol oxidase
612 activities in cultures grown on natural lignocellulosic materials (wheat straw and
613 corn stover). The latter enzymes are involved in lignin degradation, thus making
614 cellulose and hemicellulose readily available to cellulases and xylanases. The
615 production of these tripartite activities would allow fungal colonizing of
616 lignocellulosic wastes such as wheat straw and corn stover. The three fungi
617 colonized both substrates, with slightly higher overall enzyme titres being
618 produced on corn stover than on wheat straw (Table 1). Higher levels of
619 xylanase activity than cellulase activity were observed on both substrates.

Xylanase activity was measured in all three fungi on wheat straw, with the highest levels observed in TS12 (2.33 ± 0.20 IU/mg protein). Cellulase activity levels were the highest levels again being observed in TS12 (0.76 ± 0.11 IU/mg protein) on wheat straw.

Table 1. Total proteins and enzymatic activities (expressed in IU/mg protein) in supernantants collected from the solid-state fermentations.

Strain	Fermentation on wheat straw				Fermentation on corn stover			
	Prot	Activities from supernatants			Prot	Activities from corn stover		
		Cell	Xyl	Pox		Cell	Xyl	Pox
TS2	0.86 ± 0.05	0.38 ± 0.10	1.87 ± 0.21	127.82 ± 3.79	2.53 ± 0.31	0.47 ± 0.09	0.75 ± 0.11	114.17 ± 3.77
TS11	1.23 ± 0.19	0.67 ± 0.13	1.83 ± 0.37	99.84 ± 3.88	2.38 ± 0.30	0.42 ± 0.05	1.16 ± 0.14	104.95 ± 4.09
TS12	0.99 ± 0.10	0.76 ± 0.11	2.33 ± 0.20	118.37 ± 4.11	1.40 ± 0.19	0.43 ± 0.07	0.80 ± 0.08	114.25 ± 3.04

Prot (Total protein), Cell (cellulase), Xyl (xylanase) and Pox (phenol oxidase).

Regarding peroxidase/phenol oxidase activity (Pox), with the method employed allowing the assessment of both peroxidase and phenol oxidase enzymatic activity [44]; Pox activity was observed in all cultures on both substrates with the higher overall levels again being observed in wheat straw, and the highest Pox levels being observed in TS2 (127.82 ± 3.79 IU/mg protein). This is to the best of our knowledge the first time that peroxidase/phenol oxidase activity has been reported in these three fungal genera. This together with the fact that TS2, TS11 and TS12 displayed cellulase and xylanase activities on these agricultural waste materials, suggesting that they may be good candidates for further evaluation in fungal mediated lignocellulosic biomass conversion strategies; particularly for biorefinery related biotechnological applications.

3.9 Cellulolytic and xylanolytic zymograms

Although a number of enzyme activities (cellulases, xylanases, peroxidase/phenol oxidase) were detected in the supernatants obtained from the fungi (TS2, TS11 and TS12) grown under solid-state fermentation, our

643 primary focus was to further characterize the cellulase and xylanase activities.
644 To this end we used zymographic methods to monitor the production of
645 cellulase and xylanase isoforms, (Fig. 10) to corroborate the activities reported
646 in Table 1.

647 **Fig. 10** Zymogram for detection of cellulases and xylanases in different natural
648 substrates. (A) Cellulase activity bands, 1) Molecular weight marker, 2) Strain
649 TS2 wheat straw supernatant, 3) Strain TS2 corn stover supernatant, 4) Strain
650 TS11 wheat straw supernatant, 5) Strain TS11 corn stover supernatant, 6)
651 Strain TS12 wheat straw supernatant, 7) Strain TS12 corn stover supernatant.
652 (B) Xylanase activity bands, 1) Molecular weight marker, 2) Strain TS2 wheat
653 straw supernatant, 3) Strain TS2 corn stover supernatant, 4) Strain TS11 wheat
654 straw supernatant, 5) Strain TS11 corn stover supernatant, 6) Strain TS12
655 wheat straw supernatant, 7) Strain TS12 corn stover supernatant.

656 Three bands of approximately 35, 40 and 120 kDa corresponding to cellulase
657 activity were observed in TS2 grown in both wheat straw and corn stover (lanes
658 2 and 3, Fig. 10A). In contrast, two major bands were observed when TS11 was
659 grown on each substrate: one band of 40 kDa for both substrates, and another
660 of 50 kDa when the strain was grown on wheat straw and 55 kDa corresponding
661 to cellulase activity, when TS11 was grown on corn stover (lanes 4 and 5, Fig.
662 10A). Three bands of approximately 80, 100 and 130 kDa were observed when
663 TS12 was grown on wheat straw, while two bands (130 and 100 kDa) were
664 visualized when TS12 was grown on corn stover (lanes 6 and 7, Fig. 10A). Thus
665 it is clear that cellulase production in TS11 and TS12 is affected by the
666 substrate. This has previously been reported in both *Trichoderma* sp. and in

667 *Bjerkandera adusta* where different cellulolytic profiles are produced when the
668 fungi are grown on various natural substrates [72,80].
669 With respect to xylanase activity one predominant band of approximately 10
670 kDa was observed when TS2 was grown on both substrates (lanes 2 and 3, Fig.
671 10B). A major band of around 100 kDa was observed when TS11 was grown on
672 both substrates, with other smaller bands also being observed (lanes 4 and 5,
673 Fig. 10B). A specific band (approximately 10 kDa) of xylanase can also be
674 observed in the lane 4 (Fig. 10B), indicating substrate dependent xylanolytic
675 production. Low molecular weight xylanases were mainly produced by TS12
676 (lanes 6 and 7, Fig. 10B), being observed 1 and 2 bands respectively. TS12
677 could also produce different xylanase isoforms when it is grown on both
678 substrates, being this fact previously seen in *Pycnoporus sanguineus* [72]. The
679 low (≅12 kDa) and medium (≅35 kDa) molecular weight cellulases and
680 xylanases, which are produced by these three fungi (Fig. 10B) are particularly
681 attractive from the perspective of biotechnological processes for biomass
682 conversion. Because of the tightly polymer packaging present in the plant cell
683 wall it is necessary to find new low molecular weight enzymes to increase the
684 efficiency of the biomass degradation [81]. On other hand, low molecular weight
685 enzymes frequently show interesting biochemical properties such as
686 thermotolerance and high stability [82], and are typically good candidates for
687 overexpression and protein engineering based approaches to improve their
688 physical and biochemical properties.
689 Thus while it is clear that the precise role of fungi associated with marine
690 sponges has yet to be fully elucidated, nonetheless these fungi appear to be a
691 good source of potential novel biocatalysts with unique biochemical and

692 physical properties. It is clear that these fungal sponge communities represent
693 an extraordinary resource with respect to biocatalytic potential, and in particular
694 as demonstrated here, with respect to lignocellulose degradation which is
695 continuing to attract much interest due to the on-going need for improved
696 biomass conversion strategies.
697

698 **Acknowledgments**

699 We are grateful to Rocío Rodríguez for her support in conservation of the fungi.
700 We also thank Dr Louise Allcock, Chief Scientist aboard survey CE10004 of
701 Celtic Explorer to the Irish continental margin for facilitating the collection of
702 samples, and thank the captain and crew of Celtic Explorer and participating
703 scientists for their assistance.

704

References

1. Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Hacker J, et al. Molecular evidence for a uniform microbial community in sponges from different oceans. Appl Environ Microbiol. United States; 2002;68: 4431–40.

2. Worheide G, Sole-Cava AM, Hooper JNA. Biodiversity, molecular ecology and phylogeography of marine sponges: patterns, implications and outlooks. Integr Comp Biol. England; 2005;45: 377–85.

3. Suryanarayanan TS. The diversity and importance of fungi associated with marine sponges. Bot Mar. 2012;55: 553–64.

4. Weisz JB, Lindquist N, Martens CS. Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? Oecologia. Germany; 2008;155: 367–76.

5. Hentschel U, Schmid M, Wagner M, Fieseler L, Gernert C, Hacker J. Isolation and phylogenetic analysis of bacteria with antimicrobial activities from the Mediterranean sponges Aplysina aerophoba and Aplysina cavernicola. FEMS Microbiol Ecol. 2001;35: 305–12.

6. Schmitt S, Tsai P, Bell J, Fromont J, Ilan M, Lindquist N, et al. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. ISME J. England; 2012;6: 564–76.

7. Kennedy J, Flemer B, Jackson S a., Morrissey JP, O'Gara F, Dobson ADW. Evidence of a putative deep sea specific microbiome in marine sponges. PLoS One. 2014;9.

8. Taylor MW, Radax R, Steger D, Wagner M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. Microbiol Mol Biol Rev. United States; 2007;71: 295–347.

9. Webster NS, Taylor MW, Behnam F, Lücker S, Rattei T, Whalan S, et al. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. Environ Microbiol [Internet]. 2010;12: 2070–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21966903

10. Lee OO, Wang Y, Yang J, Lafi FF, Al-Suwailem A, Qian P-Y. Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. ISME J [Internet]. 2011;5: 650–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21085196

11. Haber M, Ilan M. Diversity and antibacterial activity of bacteria cultured from Mediterranean Axinella spp. sponges. J Appl Microbiol. England; 2014;116: 519–32.

12. Jackson SA, Kennedy J, Morrissey JP, O'Gara F, Dobson ADW. Pyrosequencing reveals diverse and distinct sponge-specific microbial communities in sponges from a single geographical location in Irish waters. Microb Ecol [Internet]. 2012;64: 105–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22281804

13. Trindade-Silva AE, Rua C, Silva GGZ, Dutilh BE, Moreira APB, Edwards RA, et al. Taxonomic and functional microbial signatures of the endemic marine sponge Arenosclera brasiliensis. PLoS One. United States; 2012;7: e39905.

14. Zhang D, Sun W, Feng G, Zhang F, Anbuchezhian R, Li Z, et al. Phylogenetic diversity of sulphate-reducing Desulfovibrio associated with three South China Sea sponges. Lett Appl Microbiol. 2015;

15. Richards TA, Jones MDM, Leonard G, Bass D. Marine fungi: their ecology and molecular diversity. Ann Rev Mar Sci. United States; 2012;4: 495–522.

16. Rateb ME, Ebel R. Secondary metabolites of fungi from marine habitats. Nat Prod Rep. England; 2011;28: 290–344.

17. Baker PW, Kennedy J, Dobson ADW, Marchesi JR. Phylogenetic diversity and antimicrobial activities of fungi associated with haliclona simulans isolated from irish coastal waters. Mar Biotechnol. 2009;11: 540–7.

18. Wang G, Li Q, Zhu P. Phylogenetic diversity of culturable fungi associated with the Hawaiian Sponges Suberites zeteki and Gelliodes fibrosa. Antonie Van Leeuwenhoek. Netherlands; 2008;93: 163–74.

19. He L, Liu F, Karuppiah V, Ren Y, Li Z. Comparisons of the fungal and protistan communities among different marine sponge holobionts by pyrosequencing. Microb Ecol. United States; 2014;67: 951–61.

20. Bucher VVC, Hyde KD, Pointing SB, Reddy CA. Production of wood decay enzymes , mass loss and lignin solubilization in wood by marine ascomycetes and their anamorphs. Microb Ecol. 2004;48: 331–7.

21. Bianchi T. The role of terrestrially derived organic carbon in the coastal ocean: A changing paradigm and the priming effect. PNAS. 2011;108: 19473–81.

22. Bonugli-Santos RC, Durrant LR, Sette LD. The production of ligninolytic enzymes by marine-derived basidiomycetes and their biotechnological potential in the biodegradation of recalcitrant pollutants and the treatment of textile effluents. Water Air Soil Pollut. 2012;223: 2333–45.

23. Bonugli-Santos RC, Durrant LR, da Silva M, Sette LD. Production of laccase, manganese peroxidase and lignin peroxidase by Brazilian marine-derived fungi. Enzyme Microb Technol. 2010;46: 32–7.

24. Bonugli-santos RC, Durrant LR, Sette LD. Laccase activity and putative laccase genes in marine-derived basidiomycetes. Fungal Biol [Internet]. Elsevier Ltd; 2010;114: 863–72. Available from: http://dx.doi.org/10.1016/j.funbio.2010.08.003

25. Pointing SB, Buswell J a., Jones EBG, Vrijmoed LLP. Extracellular cellulolytic enzyme profiles of five lignicolous mangrove fungi. Mycol Res. 1999;103: 696–700.

779 26. Burtseva Y V., Verigina NS, Sova V V., Pivkin M V., Zvyagintseva TN. Filamentous marine
780 fungi as producers of O-glycosylhydrolases: ??-1,3-glucanase from Chaetomium
781 indicum. Mar Biotechnol. 2003;5: 349–59.

782 27. Baker PW, Kennedy J, Morrissey J, O'Gara F, Dobson ADW, Marchesi JR. Endoglucanase
783 activities and growth of marine-derived fungi isolated from the sponge Haliclona
784 simulans. J Appl Microbiol. England; 2010;108: 1668–75.

785 28. El-Bondkly AMA. Molecular identification using ITS sequences and genome shuffling to
786 improve 2-deoxyglucose tolerance and xylanase activity of marine-derived fungus,
787 Aspergillus sp. NRCF5. Appl Biochem Biotechnol. United States; 2012;167: 2160–73.

788 29. Hou YH, Wang TH, Long H, Zhu HY. Novel cold-adaptive Penicillium strain FS010
789 secreting thermo-labile xylanase isolated from Yellow Sea. Acta Biochim Biophys Sin
790 (Shanghai). 2006;38: 142–9.

791 30. Del-Cid A, Ubilla P, Ravanal MC, Medina E, Vaca I, Levicán G, et al. Cold-active xylanase
792 produced by fungi associated with Antarctic marine sponges. Appl Biochem Biotechnol.
793 2014;172: 524–32.

794 31. Steindler L, Huchon D, Avni A, Ilan M. 16S rRNA phylogeny of sponge-associated
795 cyanobacteria. Appl Environ Microbiol. United States; 2005;71: 4127–31.

796 32. Kennedy J, O'Leary ND, Kiran GS, Morrissey JP, O'Gara F, Selvin J, et al. Functional
797 metagenomic strategies for the discovery of novel enzymes and biosurfactants with
798 biotechnological applications from marine ecosystems. J Appl Microbiol [Internet].
799 2011;111: 787–99. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21777355

800 33. Kuhad RC, Kapoor RK, Lal R. Improving the yield and quality of DNA isolated from white-
801 rot fungi. Folia Microbiol (Praha). Czech Republic; 2004;49: 112–6.

802 34. Borneman J, Hartin RJ. PCR primers that amplify fungal rRNA genes from environmental
803 samples. Appl Environ Microbiol. UNITED STATES; 2000;66: 4356–60.

804 35. Hinrikson HP, Hurst SF, Lott TJ, Warnock DW, Morrison CJ. Assessment of ribosomal
805 large-subunit D1-D2, internal transcribed spacer 1, and internal transcribed spacer 2
806 regions as targets for molecular identification of medically important Aspergillus
807 species. J Clin Microbiol. United States; 2005;43: 2092–103.

808 36. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr:
809 robust phylogenetic analysis for the non-specialist. Nucleic Acids Res [Internet].
810 2008;36: W465–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18424797

811 37. Dereeper A, Audic S, Claverie J-M, Blanc G. BLAST-EXPLORER helps you building
812 datasets for phylogenetic analysis. BMC Evol Biol. England; 2010;10: 8.

813 38. Vogel HJ. A convenient growth medium for Neurospora (Medium N). Microb Genet Bull.
814 1956;13: 42–3.

39. Teather RM, Wood PJ. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol. UNITED STATES; 1982;43: 777–80.

40. Lowry O, Rosebrough N, Lewis A, Randall R. The folin by oliver. Readings [Internet]. 1951;193: 265–75. Available from: http://www.life.illinois.edu/biochem/355/articles/LowryJBC193_265.pdf

41. Miller GL. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal Chem [Internet]. 1959;31: 426–8. Available from: http://pubs.acs.org/doi/pdf/10.1021/ac60147a030

42. Batista-García RA, Balcázar-López E, Miranda-Miranda E, Sánchez-Reyes A, Cuervo-Soto L, Aceves-Zamudio D, et al. Characterization of Lignocellulolytic Activities from a Moderate Halophile Strain of Aspergillus caesiellus Isolated from a Sugarcane Bagasse Fermentation. PLoS One [Internet]. 2014;9: e105893. Available from: http://dx.plos.org/10.1371/journal.pone.0105893

43. Floch C, Alarcon-Gutiérrez E, Criquet S. ABTS assay of phenol oxidase activity in soil. J Microbiol Methods. 2007;71: 319–24.

44. Bach CE, Warnock DD, Van Horn DJ, Weintraub MN, Sinsabaugh RL, Allison SD, et al. Measuring phenol oxidase and peroxidase activities with pyrogallol, l-DOPA, and ABTS: Effect of assay conditions and soil type. Soil Biol Biochem. 2013;67: 183–91.

45. Allison SD, Jastrow JD. Activities of extracellular enzymes in physically isolated fractions of restored grassland soils. Soil Biol Biochem. 2006;38: 3245–56.

46. Quiroz-Castañeda RE, Balcázar-López E, Dantán-González E, Martinez A, Folch-Mallol J, Martínez-Anaya C. Characterization of cellulolytic activities of Bjerkandera adusta and Pycnoporus sanguineus on solid wheat straw medium. Electron J Biotechnol. 2009;12.

47. Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, Nyyssonen E, et al. Swollenin, a Trichoderma reesei protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. Eur J Biochem. Germany; 2002;269: 4202–11.

48. Travadon R, Lawrence DP, Rooney-Latham S, Gubler WD, Wilcox WF, Rolshausen PE, et al. Cadophora species associated with wood-decay of grapevine in North America. Fungal Biol [Internet]. Elsevier Ltd; 2015;119: 53–66. Available from: http://linkinghub.elsevier.com/retrieve/pii/S187861461400172X

49. Blanchette R a., Held BW, Arenz BE, Jurgens J a., Baltes NJ, Duncan SM, et al. An Antarctic hot spot for fungi at Shackleton's historic hut on Cape Royds. Microb Ecol. 2010;60: 29–38.

50. Gonçalves VN, Vaz ABM, Rosa C a., Rosa LH. Diversity and distribution of fungal communities in lakes of Antarctica. FEMS Microbiol Ecol. 2012;82: 459–71.

51. Sigler L, Zuccaro A, Summerbell RC, Mitchell J, Paré J a. Acremonium exuviarum sp. nov., a lizard-associated fungus with affinity to Emericellopsis. Stud Mycol. 2004;50: 409–13.

52. Grum-Grzhimaylo A, Georgieva M. Are alkalitolerant fungi of the Emericellopsis lineage (Bionectriaceae) of marine origin? IMA Fungus [Internet]. 2013;4: 213–28. Available from: http://www.imafungus.org/Issue/42/17.pdf

53. Wiese J, Ohlendorf B, Blumel M, Schmaljohann R, Imhoff JF. Phylogenetic identification of fungi isolated from the marine sponge Tethya aurantium and identification of their secondary metabolites. Mar Drugs. Switzerland; 2011;9: 561–85.

54. Proksch P, Ebel R, Edrada R, Riebe F, Liu H, Diesel A, et al. Sponge-associated fungi and their bioactive compounds: The Suberites case. Bot Mar. 2008;51: 209–18.

55. Johnson LJ a N, Miller AN, McCleery R a., McClanahan R, Kath J a., Lueschow S, et al. Psychrophilic and psychrotolerant fungi on bats and the presence of Geomyces spp. on bat wings prior to the arrival of white nose syndrome. Appl Environ Microbiol. 2013;79: 5465–71.

56. Zukal J, Bandouchova H, Bartonicka T, Berkova H, Brack V, Brichta J, et al. White-nose syndrome fungus: A generalist pathogen of hibernating bats. PLoS One. 2014;9.

57. Raap J, Erkelens K, Ogrel A, Skladnev D a., Brückner H. Fungal biosynthesis of non-ribosomal peptide antibiotics and α, α-dialkylated amino acid constituents. J Pept Sci. 2005;11: 331–8.

58. Carreira C, Staal M, Falkoski D, de Vries R, Middelboe M, Brussaard CPD. Disruption photoautotrophic intertidal mats by filamentous fungi. Environ Microbiol. 2015; 1–33.

59. Godinho VM, Gonçalves VN, Santiago IF, Figueredo HM, Vitoreli G a., Schaefer CEGR, et al. Diversity and bioprospection of fungal community present in oligotrophic soil of continental Antarctica. Extremophiles [Internet]. 2015; Available from: http://link.springer.com/10.1007/s00792-015-0741-6

60. Figueroa L, Jimenez C, Rodriguez J, Areche C, Chavez R, Henriquez M, et al. 3-Nitroasterric Acid Derivatives from an Antarctic Sponge-Derived Pseudogymnoascus sp. Fungus. J Nat Prod. 2015;

61. Yokoyama K, Wang L, Miyaji M, Nishimura K. Identification, classification and phylogeny of the Aspergillus section Nigri inferred from mitochondrial cytochrome b gene. FEMS Microbiol Lett. Netherlands; 2001;200: 241–6.

62. Chang PK, Bhatnagar D, Cleveland TE, Bennett JW. Sequence variability in homologs of the aflatoxin pathway gene aflR distinguishes species in Aspergillus section Flavi. Appl Environ Microbiol. 1995;61: 40–3.

63. Kanbe T, Yamaki K, Kikuchi A. Identification of the pathogenic Aspergillus species by nested PCR using a mixture of specific primers to DNA topoisomerase II gene. Microbiol Immunol. Japan; 2002;46: 841–8.

64. Geiser DM, Pitt JI, Taylor JW. Cryptic speciation and recombination in the aflatoxin-producing fungus Aspergillus flavus. Proc Natl Acad Sci U S A. 1998;95: 388–93.

65. Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol. 2010;10: 189.

66. Jagielski T, Kosim K, Skora M, Macura AB, Bielecki J. Identification of Scopulariopsis species by partial 28S rRNA gene sequence analysis. Pol J Microbiol. Poland; 2013;62: 303–6.

67. Wang G. Diversity and biotechnological potential of the sponge-associated microbial consortia. J Ind Microbiol Biotechnol. 2006;33: 545–51.

68. Gunny AAN, Arbain D, Edwin Gumba R, Jong BC, Jamal P. Potential halophilic cellulases for in situ enzymatic saccharification of ionic liquids pretreated lignocelluloses. Bioresour Technol [Internet]. Elsevier Ltd; 2014;155: 177–81. Available from: http://dx.doi.org/10.1016/j.biortech.2013.12.101

69. Raudabaugh DB, Miller AN. Nutritional Capability of and Substrate Suitability for Pseudogymnoascus destructans, the Causal Agent of Bat White-Nose Syndrome. PLoS One. 2013;8: 1–9.

70. Reynolds HT, Barton H a. Comparison of the white-nose syndrome agent Pseudogymnoascus destructans to cave-dwelling relatives suggests reduced saprotrophic enzyme activity. PLoS One. 2014;9: 16–9.

71. Bouchotroch S, Quesada E, Del Moral a., Llamas I, Béjar V. Halomonas maura sp. nov., a novel moderately halophilic, exopolysaccharide-producing bacterium. Int J Syst Evol Microbiol. 2001;51: 1625–32.

72. Quiroz-Castañeda RE, Pérez-Mejía N, Martínez-Anaya C, Acosta-Urdapilleta L, Folch-Mallol J. Evaluation of different lignocellulosic substrates for the production of cellulases and xylanases by the basidiomycete fungi Bjerkandera adusta and Pycnoporus sanguineus. Biodegradation. 2011;22: 565–72.

73. Martínez-Anaya C, Balcázar-López E, Dantán-González E, Folch-Mallol JL. Celulasas fúngicas : Aspectos biológicos y aplicaciones en la industria energética. Rev Latinoam Microbiol. 2008;50: 119–31.

74. Saranraj P, Stella D, Reetha D. Microbial cellulases and its applictions: a review. Int J Biochem Biotech Sci. 2012;1: 1–12.

75. Baldrian P, Valášková V. Degradation of cellulose by basidiomycetous fungi. FEMS Microbiol Rev. 2008;32: 501–21.

76. Annamalai N, Thavasi R, Jayalakshmi S, Balasubramanian T. Thermostable and alkaline tolerant xylanase production by Bacillus subtilis isolated form marine environment. Indian J Biotechnol. 2009;8: 291–7.

77. Duncan SM, Farrell RL, Thwaites JM, Held BW, Arenz BE, Jurgens J a., et al. Endoglucanase-producing fungi isolated from Cape Evans historic expedition hut on Ross Island, Antarctica. Environ Microbiol. 2006;8: 1212–9.

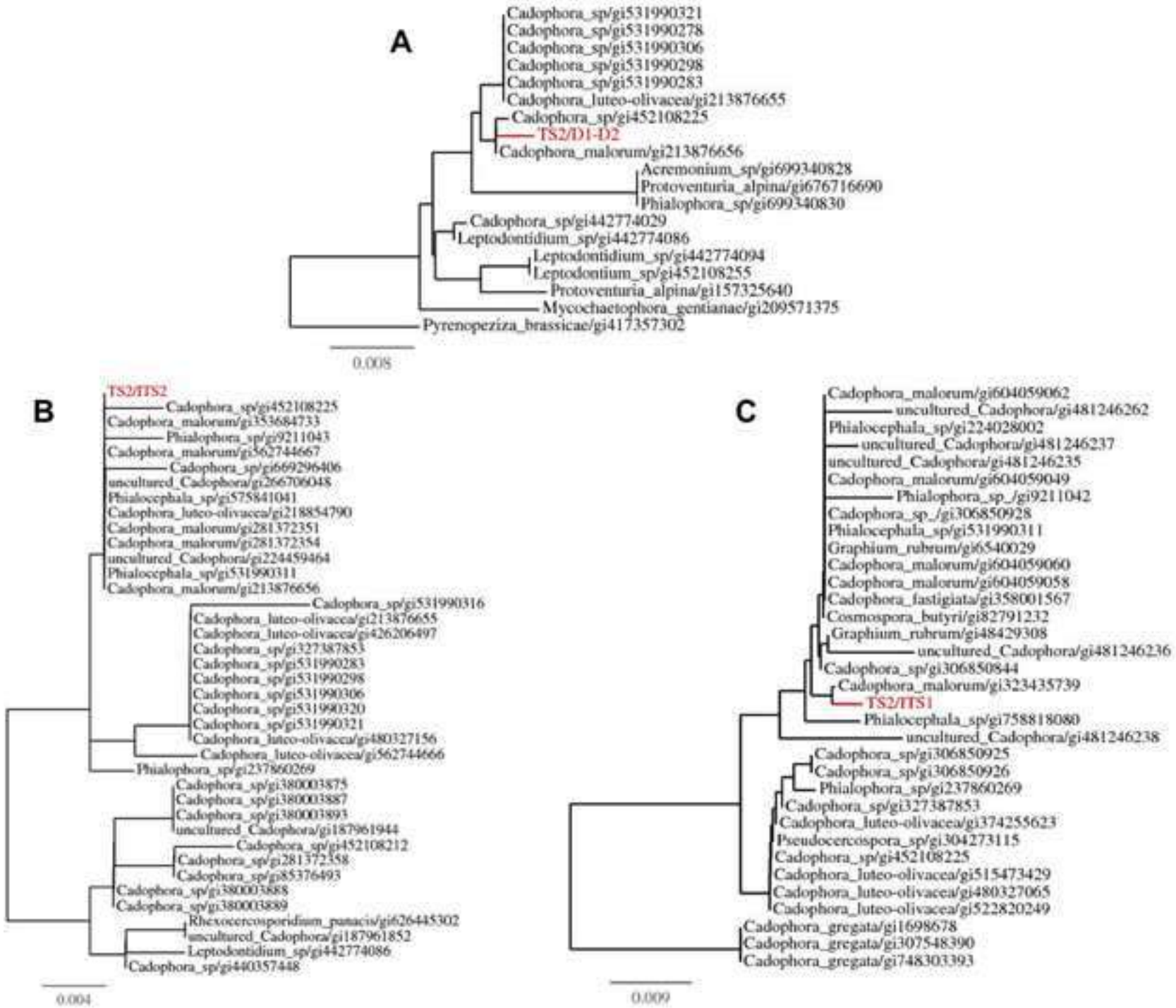
78. Zuccaro A, Summerbell RC, Gams W, Schroers HJ, Mitchell JI. A new Acremonium species associated with Fucus spp., and its affinity with a phylogenetically distinct marine Emericellopsis clade. Stud Mycol. 2004;50: 283–97.

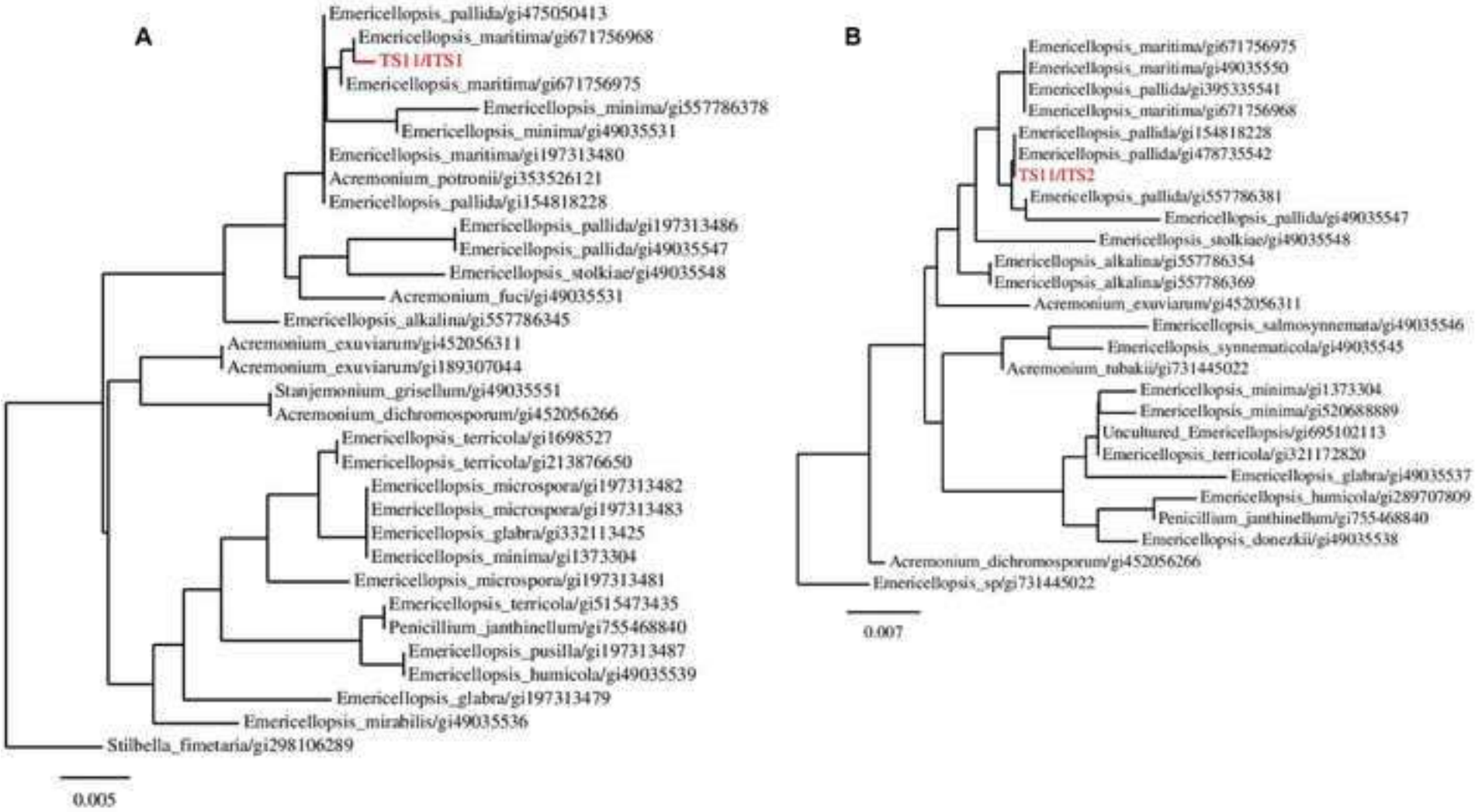
79. Molina C, Sánchez A, Serafin-Muñoz A, Folch-Mallol J. Optimization of enzymatic saccharification of wheat straw in a micro-scale system by response surface methodology. Rev Mex Ing Química. 2014;13: 765–8.

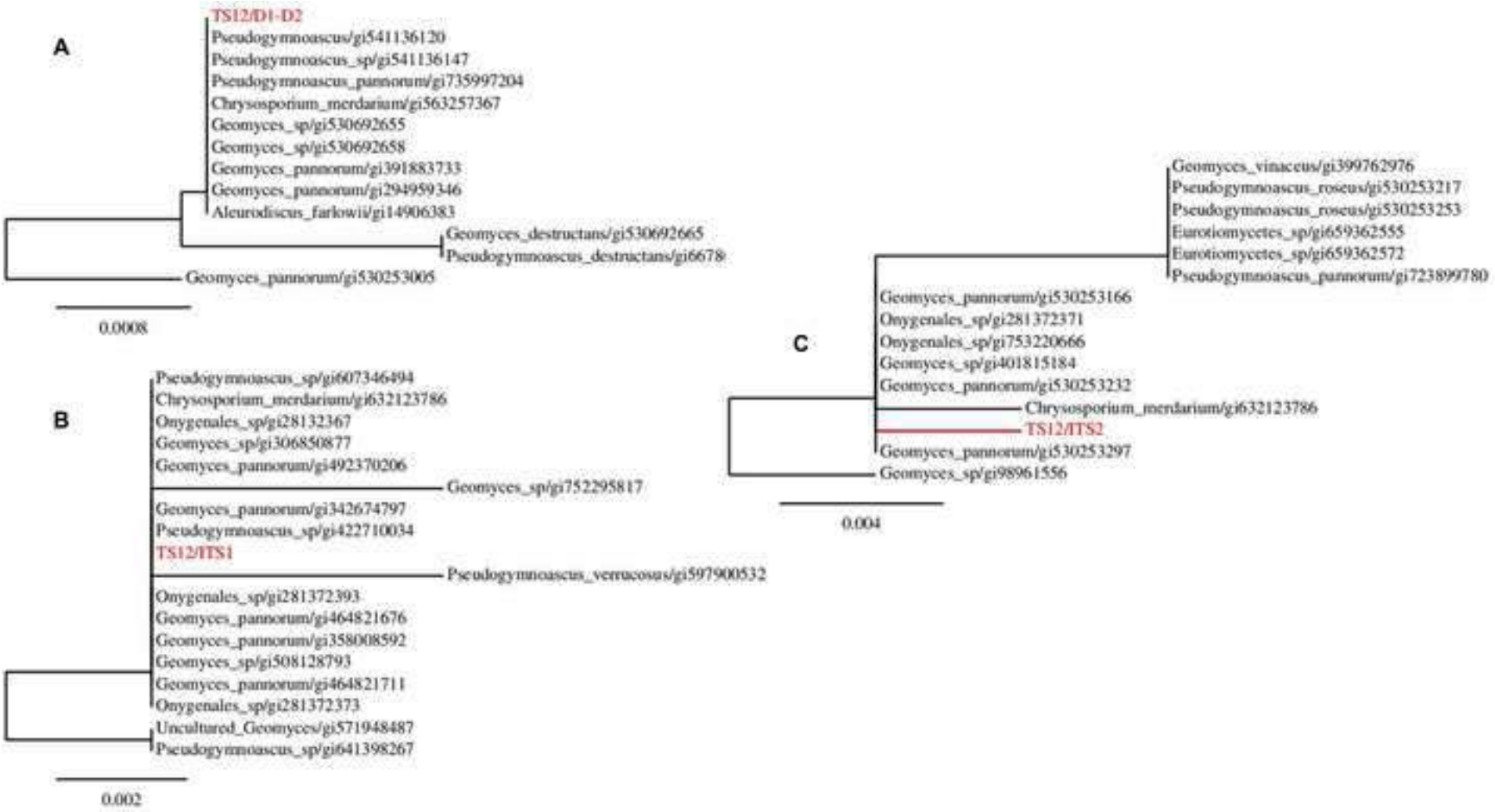
80. Dojnov B, Grujic M, Vujcic Z. Reliable simultaneous zymographic method of characterization of cellulolytic enzymes from fungal cellulase complex. Electrophoresis. 2015;doi: 10.10.

81. Arantes V, Saddler JN. Access to cellulose limits the efficiency of enzymatic hydrolysis : the role of amorphogenesis. 2010; 1–11.

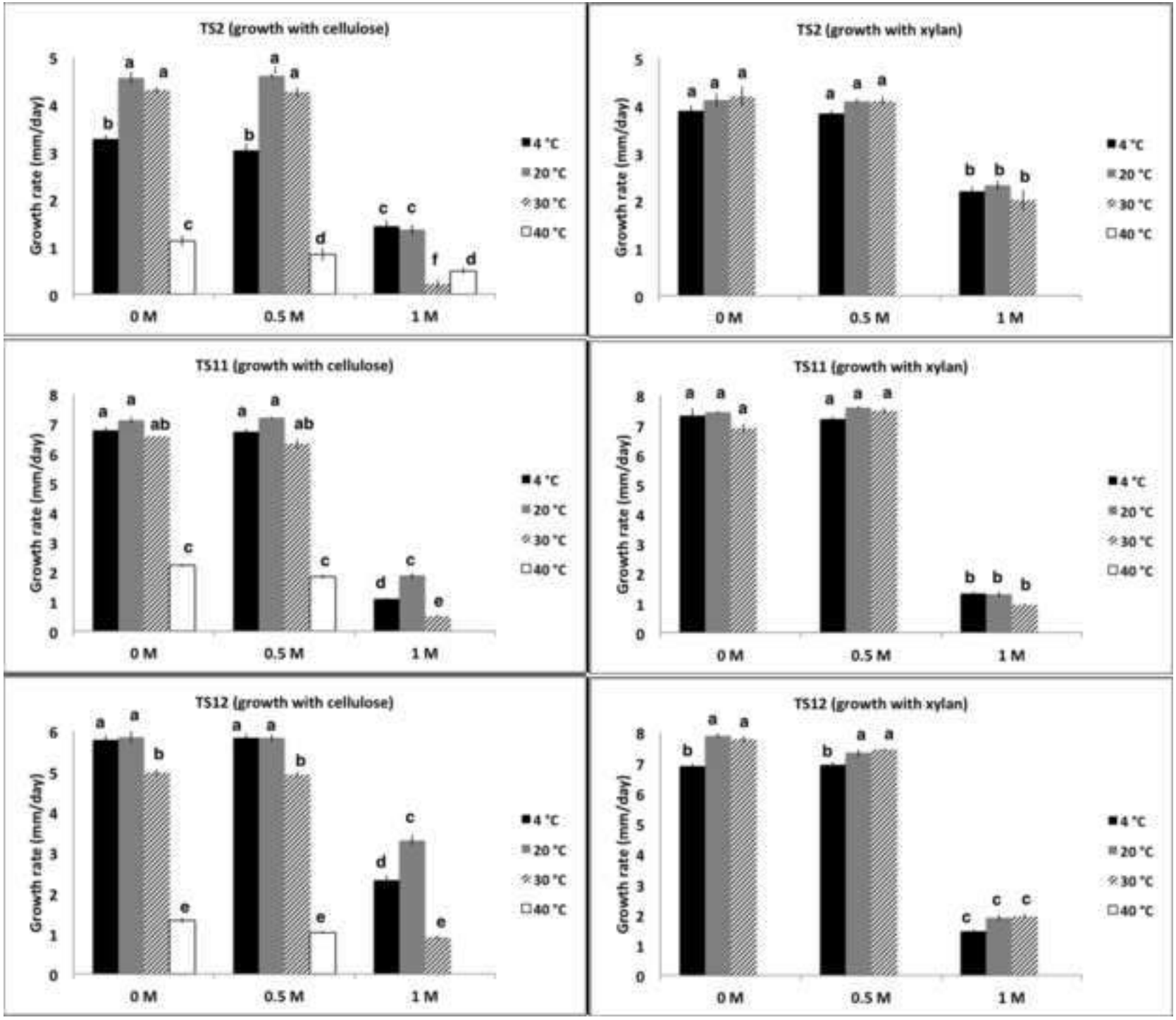
82. Fan X, Mattey M. Small enzymes with esterase activities from two thermophilic fungi , Emericella nidulans and Talaromyces emersonii. 1999; 1071–6.



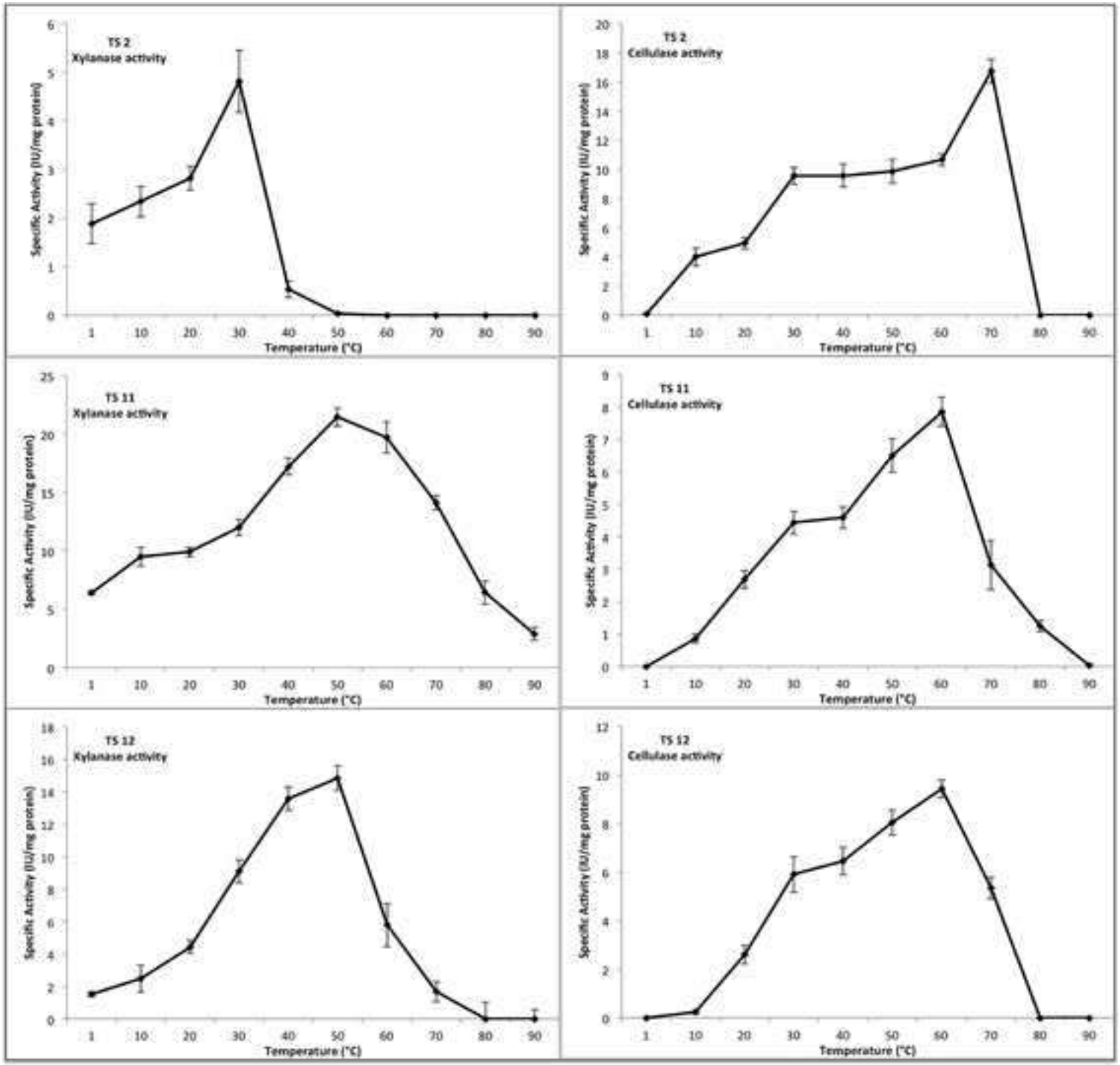




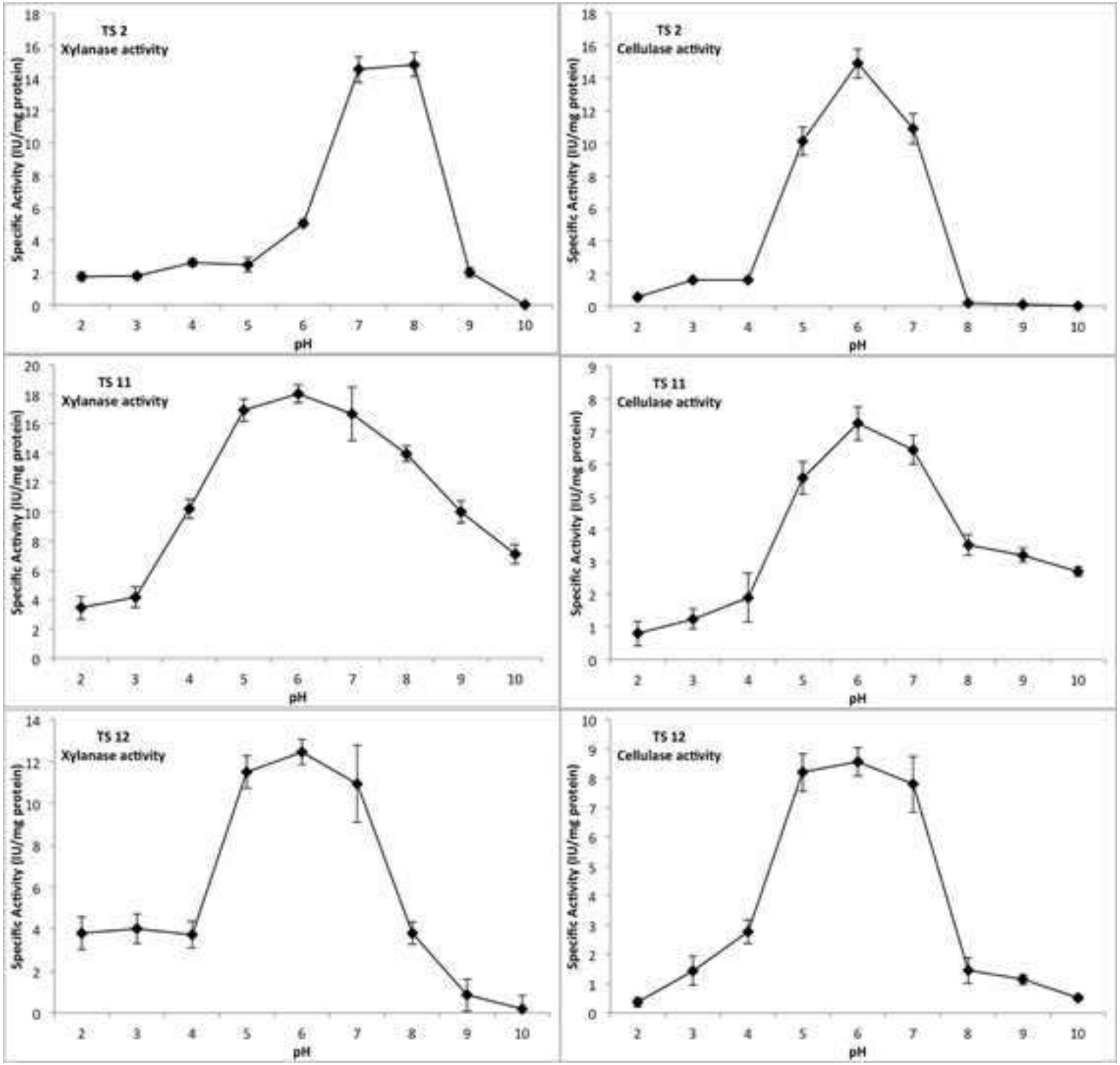
ick here to download Figure: Fig 4.tif

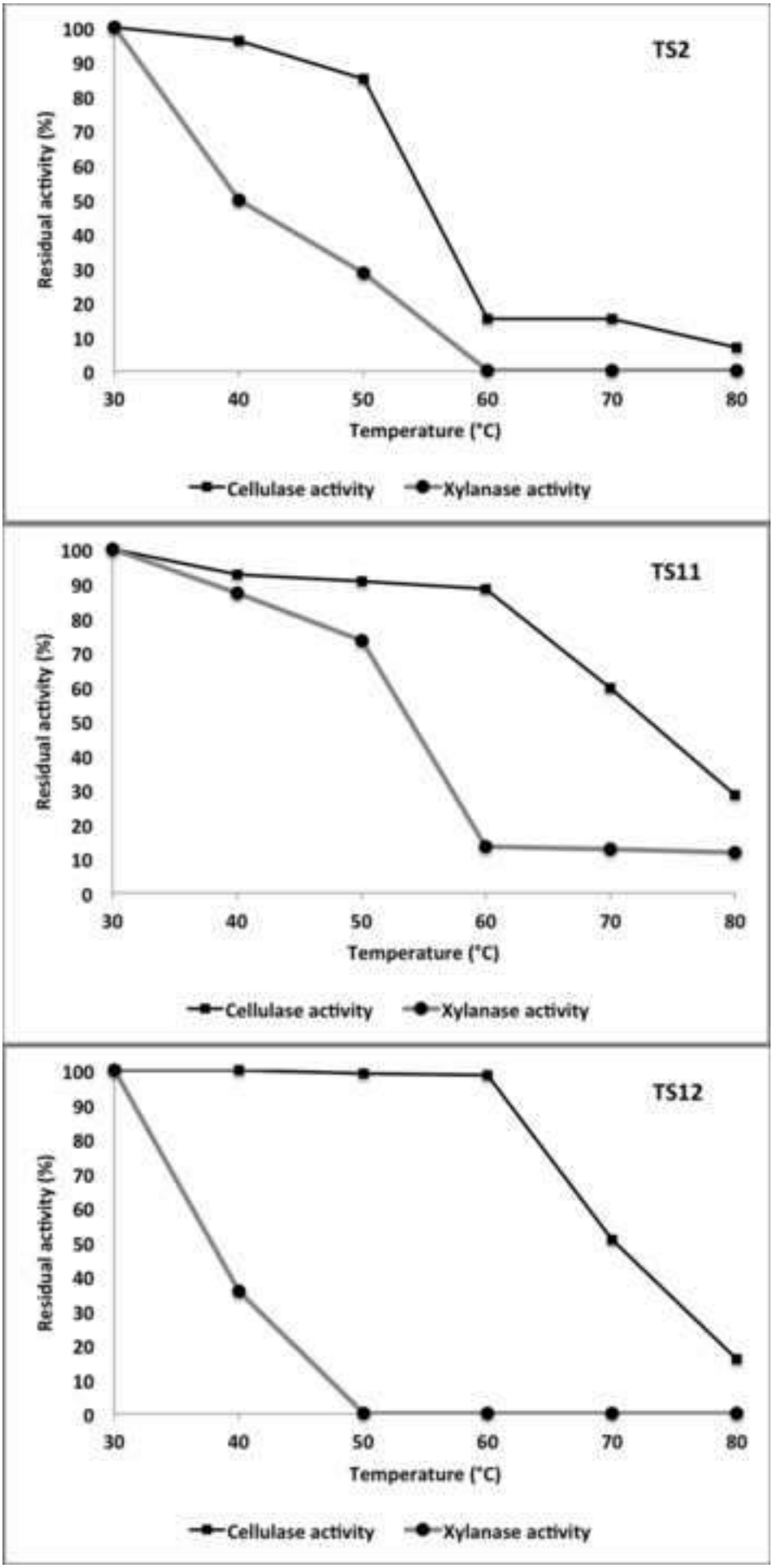


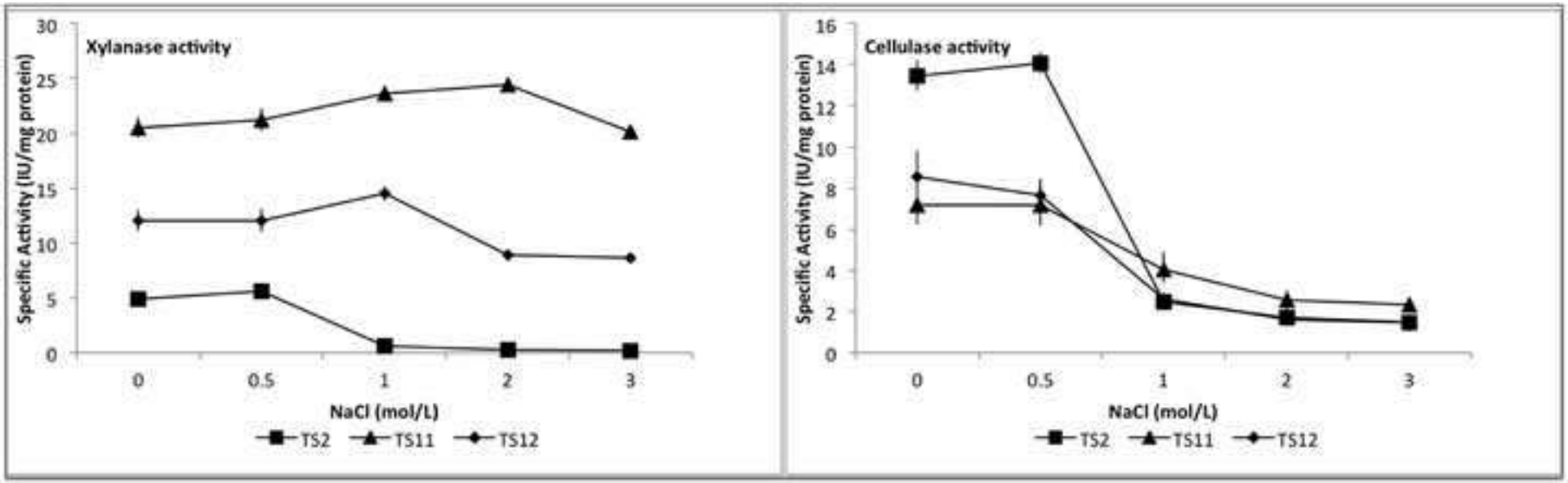
ick here to download Figure: Fig 5.tif



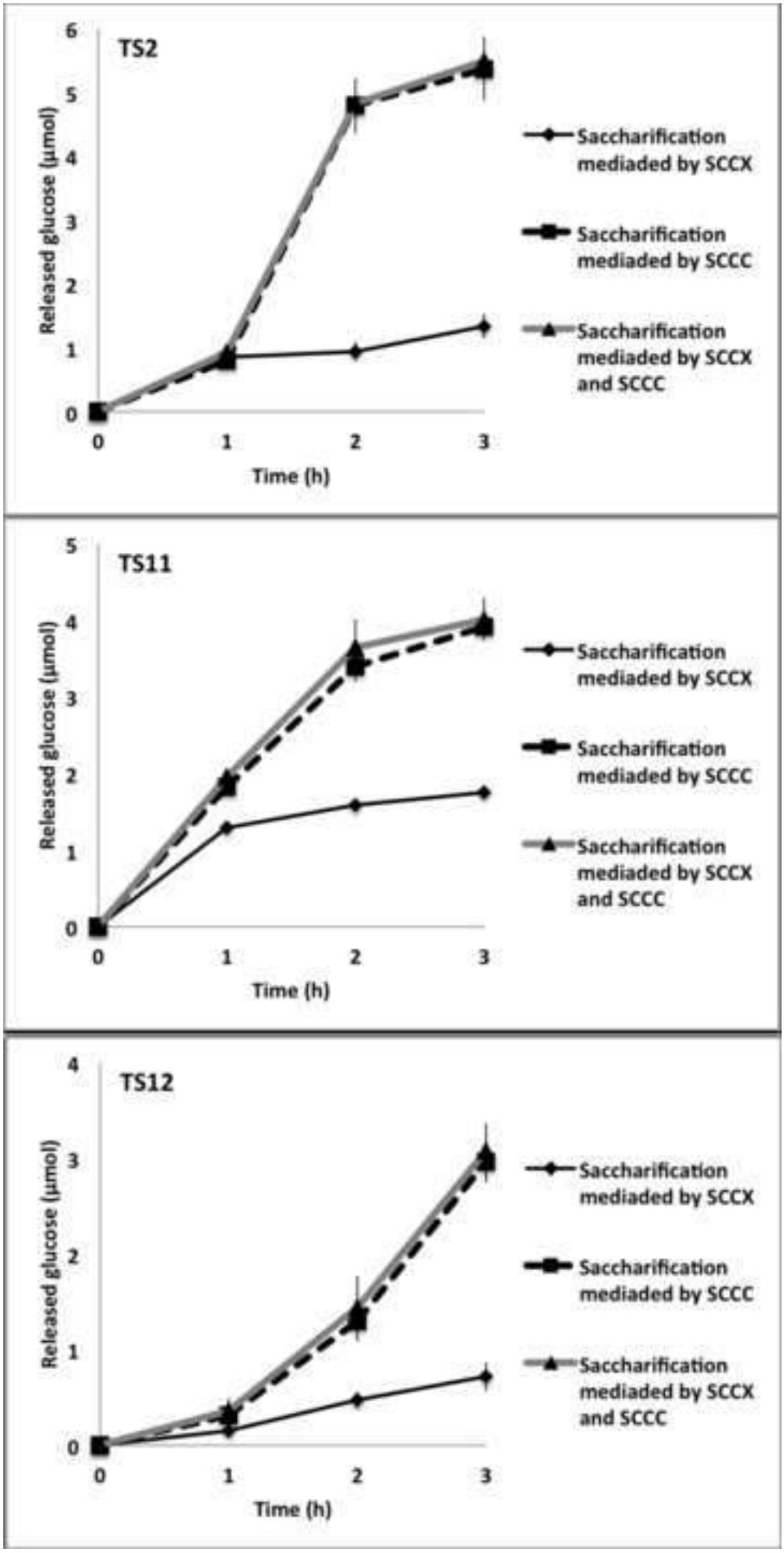
ick here to download Figure: Fig 6.tif



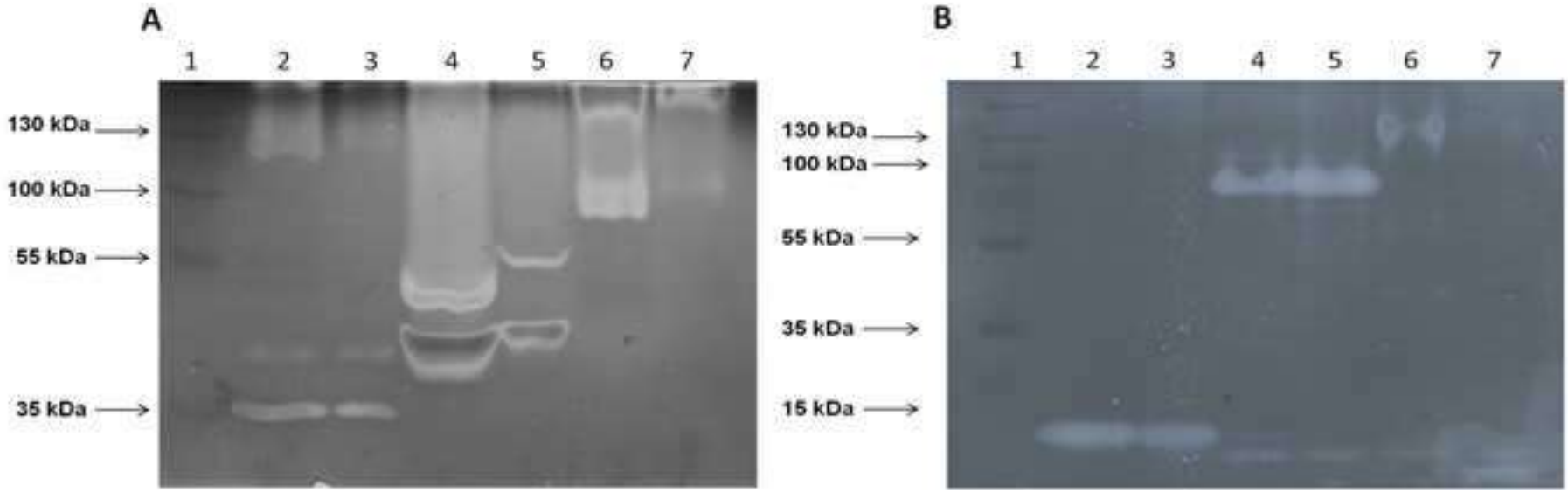




ick here to download Figure: Fig 9.tif



ick here to download Figure: Fig 10.tif



Conclusiones Generales

Las prospecciones de genomas a través de métodos ecológicos convencionales (Microbiología Pateuriana) y enfoques *ómicos*, resulta de utilidad para la descripción de nuevas proteínas con utilidad biotecnológica. En este trabajo ambas metodologías permitieron el estudio de expanisnas, celulasas, xilanasas y fenoloxidasas en general.

Las prospecciones metagenómicas permitieron describir las comunidades bacterianas colonizando una muestra de bagazo de caña de azúcar en fermentación. Las relaciones filogenéticas de las secuencias (RNAr16S) obtenidas de la librería metagenómica con secuencias de referencia mostraron relaciones evolutivas distantes, demostrando que los sustratos lignocelulósicos constituyen habitats ideales para la descripción de nuevos filotipos microbianos involucrados en la degradación de lignocelulosa.

Los análisis funcionales de metagenomas ambientales presentan numerosas desventajas que ponderan negativamente sus tasa de éxito.

La expansina de *Schizophyllum commune* favorece la liberación de glucosa y N-acetilglucosamina a partir de celulosa cristalina y quitina, aspecto que evidencia sus usos para decontrucción de polímeros con alta cristalinidad.

La cepa halófila de *Aspergillus caesiellus* H1 demostró colonizar diferentes sustratos naturales, y expresar en ellos perfiles proteicos diferenciales de actividades enzimáticas (celulasa, xilanasa, fenoloxidasa, esterasa) relacionadas con deconstrucción de biomasa.

Los hongos asociados a la esponja marina *Stelleta normani* muestran novedosos perfiles enzimáticos con potencialidades de aplicación en la degradación de materiales lignocelulósicos. *Cadophora* sp. TS2, *Emericellopsis* sp. TS11 y *Pseudogymnoascus* sp. TS12 se describieron como hongos lignocelulolíticos con actividades enzimáticas halotolerantes, termotolerante, psicrotolerante; y además fueron capaces de colonizar sustratos lignocelulolíticos naturales como paja de trigo y rastrojo de maíz.