

# Caracterización de comunidades microbianas de baja abundancia provenientes de un consorcio con capacidad lignocelulolítica.

<sup>1</sup>Díaz-Rodríguez Carlos Andrés, <sup>2</sup> León-Robayo David, <sup>3</sup> Patiño-Suarez Cesar

<sup>1</sup>*Maestría en Ciencias Biológicas, Facultad de Ciencias, Universidad de los Andes. Colombia*

<sup>2</sup>*Ingeniería de Sistemas, Facultad de Ingeniería, Universidad de los Andes. Colombia*

<sup>3</sup>*Ingeniería de petróleos, Empresa Colombiana de Petróleo S.A. Colombia*

**Autor de correspondencia:** <sup>1</sup>Díaz-Rodríguez Carlos Andrés, Facultad de Ciencias, Universidad de los Andes. Colombia. Email: [c.diaz@uniandes.edu.co](mailto:c.diaz@uniandes.edu.co).

**Palabras clave:** MAGS, análisis de balance de flujo, consorcio lignocelulolítico, anotación funcional y taxonómica, secuencias de lectura larga.

## INTRODUCCIÓN

La transición energética desde fuentes no renovables hacia fuentes amigables con el medio ambiente es de vital importancia para todos los países en el mundo. Este cambio ha sido apalancado gracias a la implementación de una economía verde también llamada bioeconomía. Esta nueva economía, busca el aprovechamiento de los residuos vegetales, puesto que estos permiten la síntesis de gran variedad de productos tales como, combustibles de tercera generación (Arpia et al., 2021); compuestos base para reacciones de polimerización hacia productos químicos de gran complejidad (Li et al., 2020) compuestos para la industria alimenticia (Muscat et al., 2020); entre otras múltiples aplicaciones que todavía presentan algunos retos importantes (Singhvi et al., 2019). Específicamente para el caso colombiano, la implementación de este tipo de tecnologías impulsará la industria biotecnológica en el país, pues permite un mejor aprovechamiento de los residuos agroindustriales que, de acuerdo con Peñaranda González y colaboradores

(2017), no ha logrado superar el 17% de eficiencia. Actualmente existen diversas estrategias para el aprovechamiento de residuos, como por ejemplo procesos de gasificación para la producción de syngas (Ren et al., 2019); procesos de digestión anaerobia (Ahmed et al., 2019); pretratamiento térmico de residuos para la producción de biohidrógeno (Eskicioglu et al., 2017); así como diversos tipos de fermentación, tanto en estado sólido como fermentaciones sumergidas para la producción de compuestos de alto valor agregado (Cerdeira et al., 2019; Rahmati et al., 2020). Dentro de los métodos anteriormente descritos destacan las fermentaciones sumergidas para la conversión de biomasa vegetal. Dichas fermentaciones pueden ser llevadas a cabo por cultivos puros de organismos relativamente sencillos como bacterias (Ju et al., 2020; Sadhukhan et al., 2016) o por organismos más complejos como levaduras (Diethard et al., 2016; Li et al., 2007), además del uso de otras plataformas en las que se emplea la búsqueda de un organismo que presenta la

actividad deseada, la optimización de su producción y la posterior recuperación de la enzima, esto se logra comúnmente clonando su gen en una plataforma celular de expresión apropiada (Alcalde et al., 2006; Green et al., 2006; Tyson et al., 2005).

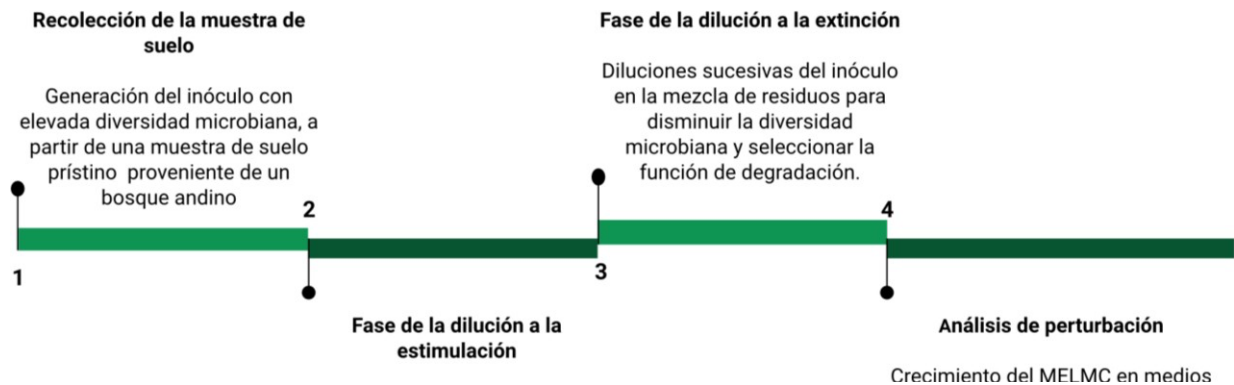
En este punto es importante resaltar que las fermentaciones realizadas por cultivos puros de microorganismos presentan limitaciones con respecto a los cultivos mixtos. Por ejemplo, los cultivos mixtos presentan una mayor resistencia a cambios bruscos en el medio de cultivo o contaminaciones con otros microorganismos del ambiente que afecten negativamente su crecimiento (Pandey et al., 2020). Por lo cual, el uso de cultivos mixtos de comunidades microbianas estables, capaces de degradar material vegetal complejo, ha sido ampliamente investigado en una gran variedad de sustratos. El resultado de dichos estudios indica que es posible obtener altos porcentajes en los procesos de despolimerización; por tal razón, es de vital importancia la caracterización de esas comunidades microbianas o consorcios, debido al gran potencial metabólico que presentan. Para investigar dicho potencial, la comunidad científica e industrial ha empleado una gran variedad de técnicas modernas como por ejemplo, el uso de análisis de metatranscriptómica con el fin de evaluar los perfiles de expresión de enzimas relacionadas con la degradación de material vegetal, y así poder generar cocteles enzimáticos que permitan la liberación de monómeros más sencillos (Simmons et al., 2014; Lopes, Ferreira Filho & Moreira, 2018) o el uso de técnicas más conservadas y ampliamente utilizadas como la anotación funcional de los metagenomas de los consorcios microbianos (Jia, Ng, Lu, Cai & Lee, 2018).

Dada la importancia que tiene el estudio de

comunidades de microorganismos también llamados consorcios microbianos, se han implementado diferentes estrategias para el ensamblaje de dichas comunidades. Por ejemplo, la selección de comunidades microbianas con una baja diversidad de especies, pero con funciones metabólicas específicas mediante el uso de medios selectivos (Jia et al., 2020) o el uso de herramientas de la biología sintética y la biología de sistemas para la creación de comunidades microbianas con funciones metabólicas definidas, que pueden ser aplicadas en áreas como la medicina o los biocombustibles (Song et al., 2014).

Específicamente, este trabajo se apoya en el artículo descrito por Díaz-García et al., (2020). En este artículo, los investigadores buscan la creación de un consorcio mínimo efectivo con potencial de degradación de material lignocelulósico, de ahora en adelante denominado MELMC por sus siglas en inglés, a partir de una muestra de suelo andino, utilizando una metodología para el ensamblaje de consorcios denominada *top-down*. Esta metodología, consiste en el uso de diluciones sucesivas para la creación de consorcios capaces de degradar materiales recalcitrantes como la biomasa vegetal (Kang et al., 2020).

Para lograr este resultado, Díaz-García et al., (2020) llevan a cabo la metodología descrita por la figura 1 tal y como se muestra a continuación:



**Figura 1.** Flujo de trabajo para la construcción del consorcio mínimo efectivo degradador dematerial lignocelulósico reportado en *Díaz-García et al.*, (2020).

Al observar dicha figura, es claro como las fases más importantes consisten en:

1. Una fase de estimulación, la cual consiste en seleccionar microorganismos con la capacidad metabólica por medio de la expresión de enzimas capaces para degradar material vegetal.
  2. Una fase de extinción que tiene por objetivo disminuir la diversidad microbiana presente en cada una de las muestras para obtener un MELMC compuesto de pocas especies.
  3. Y finalmente, una fase perturbación en donde se evalúa la capacidad de degradación de dicho consorcio para degradar un nuevo sustrato, empleándolo como única fuente de carbono.
- Además, para caracterizar la diversidad biológica en términos de especies, los investigadores utilizan el marcador

filogenético del gen ribosomal 16S, soportando los cambios ocurridos con ayuda de métricas de diversidad Alpha y realizan la asignación taxonómica de los microorganismos presentes empleando ASV's junto con técnicas estadísticas reportadas en el artículo. De esta forma, los investigadores demuestran que es posible el ensamblaje de un consorcio microbiano lignocelulolítico capaz de degradar, al menos en cierto porcentaje, una mezcla de tres residuos agroindustriales de gran importancia para Colombia: bagazo de caña de azúcar, cascarilla de arroz y rastrojo de maíz.

Finalmente, los investigadores realizan una secuenciación del metagenoma completo del consorcio MELMC a través de secuenciación de tercera generación, con el fin de entender con más profundidad el rol que tienen cada una de las especies bacteriana dentro de los procesos de degradación de material lignocelulósico, obteniendo datos de secuenciación con las siguientes características:

**Tabla 1.** Características de los datos secuenciación de metagenomas obtenidos por PacBio

Nombre de la muestra	Tamaño del archivo	Tamaño Del Inserto	Subread más largo	N50 del Subread más largo	Longitud media del read	Media del valor de calidad Q
Metagenom Jimenez 12D	11,2 GB	8200	10018	11185	6126	Q42

## OBJETIVOS

Este trabajo tiene como objetivo realizar la caracterización de la información proveniente de los datos de secuenciación metagenómica del consorcio mínimo efectivo degradador de lignocelulosa, reportado por Díaz-García et al., (2020) con el fin de debatir los posibles roles que tienen los distintos miembros dentro del consorcio en el proceso de degradación de lignocelulosa.

Como objetivos específicos se tienen:

1. Realizar la reconstrucción de los genomas bacterianos provenientes del consorcio mínimo efectivo degradador de lignocelulosa MELMC, reportados de mayor abundancia.
2. Elaborar la caracterización funcional de las especies reportadas de mayor abundancia, con el fin de develar su potencial metabólico en los procesos de degradación de polímeros vegetales complejos.
3. Reconstruir los modelos metabólicos a escala genómica de los microorganismos presentes en el consorcio reportados de mayor abundancia, con el fin de dilucidar su capacidad para crecer en diversos sustratos.

4. Estudiar la hipótesis planteada por Díaz-García et al., (2020) acerca de la capacidad de *Pseudomonas protegens* para crecer a una mayor tasa crecimiento específica cuando se utiliza glucosa como fuente de carbono.

## METODOLOGÍA

Para realizar la reconstrucción de los genomas a partir de metagenomas, se utilizará el metagenoma proveniente de la muestra proveniente de la fase de dilución a la extinción (transferencia  $10^{-7}$ ), de ahora en adelante denominada 12D. Los datos del análisis de *shotgun metagenomics* usados para este trabajo fueron proporcionados por el profesor Diego Jiménez, asociado al departamento de Ciencias Biológicas de la Universidad de los Andes. Estos datos de secuenciación fueron obtenidos gracias a una credencial con acceso a los datos almacenados en el servicio Next Cloud del Instituto Leibniz DSMZ Colección de microorganismos y cultivos celulares, instituto en donde fue llevada a cabo la secuenciación.

Una vez se identificaron los archivos de salida del ensamblador, se separaron los MAGs con el fin de categorizarlos individualmente y realizar su anotación funcional. También se procedió a usar el sistema de representación de grafos,

llamado Bandage (Ryan et al., 2015) con el fin de obtener una información más representativa de los datos que fueron generados.

### **Anotación taxonómica y funcional de los genomas ensamblados**

Para realizar el proceso de identificación taxonómica y funcional de los MAGs obtenidos durante el proceso de ensamblaje, se siguió el siguiente flujo de trabajo:

En primer lugar, para completar el proceso de ensamblaje de genomas a partir de metagenomas, se toma *12D.hifi.fastq.gz* y se realiza su ensamblaje usando Flye 2.8.1. Posteriormente, se toman los datos del metagenoma almacenados en *assembly.fasta* y se separan los MAGs asociados. Más adelante, se usa la plataforma RAST (Rapid Annotation using Subsystem Technology) para la anotación funcional de cada MAG; una vez se completa este proceso, se aíslan los genes ribosomales ARN 16S asociados a cada MAG; consecuentemente, se ejecuta BLASTn usando los datos 16S contra NCBI para la anotación taxonómica de cada MAG. Finalmente, se identifican los MAGs asociados a las especies de mayor abundancia reportados por Díaz-García et al. (2020)

### **Métricas de calidad**

Una vez se completa el proceso de anotación taxonómica y funcional de los genomas ensamblados, se usa la plataforma MiGA para obtener datos de completitud, calidad y contaminación en cada uno de los MAG y a partir de estos datos determinar cuáles son aquellos que cumplen con las características para el desarrollo del modelo metabólico.

### **Reconstrucción de los modelos metabólicos borrador**

De manera general, nos ayudará a obtener modelos a escala genómica curados de cada una de las especies o de especies filogenéticamente relacionadas, realizando el proceso de curación manual y el posterior análisis e implementación de restricciones tipo COBRA (Lewis *et al.*, 2012) para posteriormente realizar el análisis de balance de flujos (FBA, por sus siglas en inglés). Para lograr dicha labor, se utilizará la herramienta de creación automática de modelos metabólicos MODELSEED puesto que, como es discutido por Mendoza, Olivier, Molenaar & Teusink (2019) esta herramienta presenta una serie de características tales como: una interfaz amigable, tiempos cortos para la generación del modelo draft, buena documentación, entre otras características que la hacen adecuada para la creación de modelos metabólicos tipo borrador o draft, como son conocidos usualmente, y pueden ser exportados en formato SBML (Hucka et al., 2003) para ser alimentados a otras herramientas de curación manual en un refinamiento posterior.

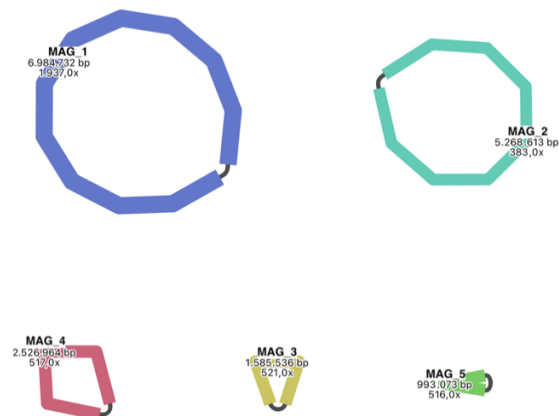
### **RESULTADOS Y DISCUSIÓN**

Una vez que los datos fueron descargados, se cargaron al clúster de la Universidad de los Andes y posteriormente se revisaron las herramientas disponibles para realizar el ensamblaje de los datos. Se utilizó el ensamblador Flye en su versión 2.8.1. La elección de dicho ensamblador se basa en sus características funcionales, ya que dicho programa se especializa en el manejo de datos de secuenciación producidos por herramientas de tercera generación, tales como PacBio y Oxford Nanopore. Una vez

se revisó la documentación de la herramienta, se alimenta el programa con las lecturas demultiplexadas obtenidas en el proceso de secuenciación por PacBio.

Una vez se identificaron los archivos de salida del ensamblador, se separaron los MAGs con el fin de categorizarlos individualmente y realizar su anotación funcional. También se procedió a usar el sistema de representación de grafos, llamado Bandage (Ryan et al., 2015) con el fin de obtener una información más representativa de los datos que fueron generados.

Tras completar el ensamblaje de ambas muestras descritas anteriormente, se obtuvieron las siguientes tablas representativas de cada uno de los genomas, en las cuales se observa su longitud en pares de bases y su cobertura; así como su respectiva representación gráfica del MAG circular ensamblado en el cual se denota la longitud y la profundidad de la lectura.



**Figura 2.** MAGs circulares obtenidos del ensamblaje de la muestra 12D. Graficados por Bandage (Ryan et al., 2015)

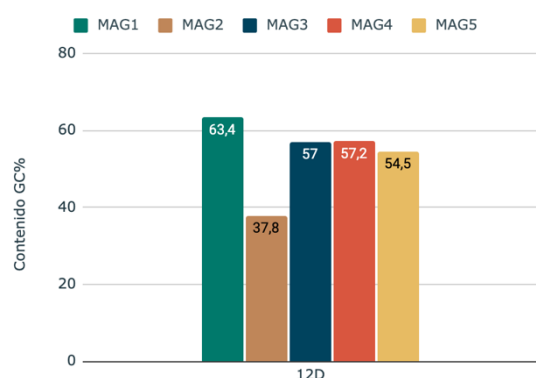
En la figura anterior, se observa claramente como todos los MAGs generados durante el proceso de ensamblaje son cerrados y poseen forma circular parcial o totalmente. Resulta evidente además, la diferencia en tamaños asociadas a cada reconstrucción por separado tal y como se muestra en la tabla 2.

**Tabla 2.** Resultados ensamblaje del metagenoma de la muestra 12D

Nombre de la secuencia	Longitud (pb)	Cobertura	Circular
MAG_1	6984732	991	Sí
MAG_2	5268613	403	Sí
MAG_3	1585536	535	Sí
MAG_4	2526964	531	Sí
MAG_5	993073	528	Sí

Como se puede observar, en general se tienen muy buenas profundidades de secuenciación. De hecho, la profundidad del MAG 1, representado por el edge\_1 se acerca a 2000x; lo que lo define como datos con un gran número de lecturas únicas, que en este nivel de cobertura pueden poseer la capacidad de identificar la presencia de mutaciones de baja ocurrencia. Por otra parte, los otros MAGs poseen una profundidad de tres a cinco veces la recomendada para este tipo de análisis de datos (Sims et al., 2014).

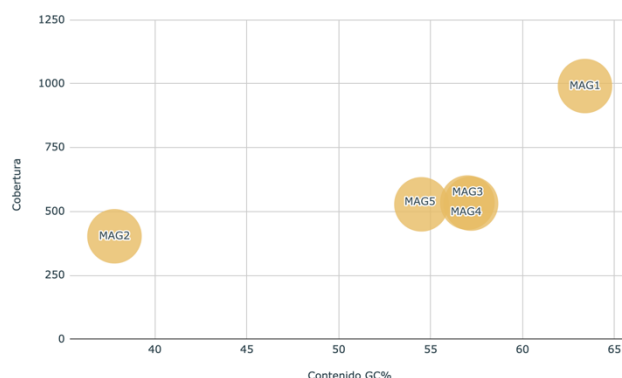
Posteriormente se utilizó la herramienta DFAST (DDBJ Fast Annotation and Submission Tool) para obtener el contenido de guanina y citosina de cada uno de los MAGs, como se puede observar a continuación:



**Figura 3.** *Contenido GC de los MAGs. Obtenido a través de DFAST*

Como se puede observar, el MAG 1 presenta un contenido más alto de guanina y citosina con respecto a los demás. De la misma manera se puede observar que el contenido GC mas bajo lo presenta el MAG 2, con un 37,8%. Por otra parte, los MAGs 3 y 4 tienen un contenido GC idéntico y finalmente, el promedio del contenido GC presente en la muestra es de 53,98..

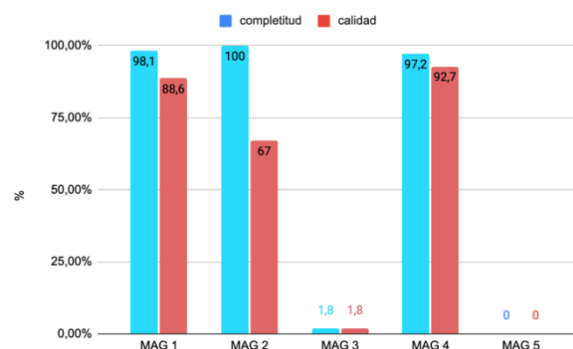
Posteriormente, se usaron los datos obtenidos, en conjunto con la información que se conocía sobre la profundidad de las lecturas para realizar el grafico mostrado en la figura 4..



**Figura 4.** *Contenido de guanina y citosina vs cobertura de los MAGs.*

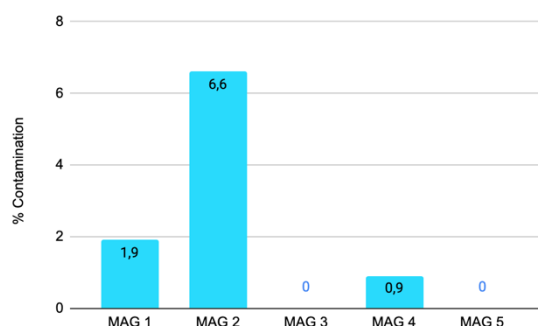
Como se puede notar, hay una gran diferencia tanto en cobertura como en contenido de guanina y citosina, entre el MAG 1 y el MAG 2; a diferencia del MAGs 3 y 4 que poseen una cobertura y un contenido GC idéntico. Por otra parte, el MAG 5 presenta una cobertura similar al MAG 3 y 4; sin embargo, su contenido GC difiere, por lo que es necesario caracterizar más a fondo dicha diferencia.

Para continuar con el análisis de las caracterizaciones, se usa la plataforma MiGA para obtener los datos de completitud y calidad que se observan a continuación.



**Figura 5.** *Análisis de completitud y calidad para cada MAG ensamblado.*

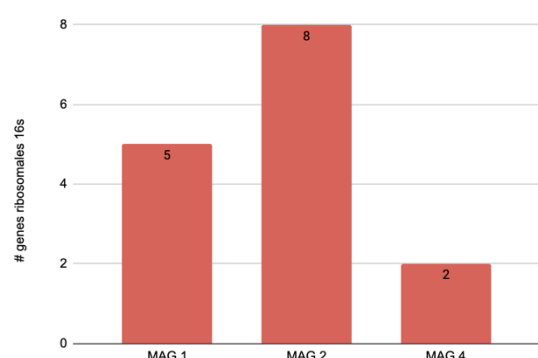
De la misma manera, se usó MiGA para obtener los datos del porcentaje de contaminación en cada MAG.



**Figura 6.** Porcentaje de contaminación asociado a cada MAG.

Teniendo en cuenta los datos obtenidos en la figura 7 y en la figura 8, se decide remover de próximos análisis a los MAGs 3 y 5 debido a su bajo nivel de completitud y de calidad, además nos encontramos con que la plataforma MiGA no obtuvo niveles de contaminación ya que los datos no eran óptimos para este tipo de análisis.

Una vez se tuvieron identificados los MAGs que se usarían en el análisis, se procede a extraer los genes ribosomales 16S, de los cuales se obtuvo la siguiente cantidad.



**Figura 7.** Cantidad genes ribosomales 16S asociados a cada MAGS.

con el fin de caracterizar funcionalmente cada uno de los MAGs obtenidos en el proceso de ensamblaje, se realiza un alineamiento de dichas secuencias contra la base de datos del NCBI, y se obtienen los microorganismos posiblemente asociados a cada uno de estos MAGs, tal y como se muestra en la tabla 3.

Como se observa en la tabla 3, los alineamientos generados poseen un porcentaje de identidad mayor al 98%, encontrando una diversidad en términos de microorganismos bastante similar a la obtenida por Díaz-García et al., (2020) mediante el análisis de diversidad alfa de los datos secuenciados gen ribosomal de ARN 16S en donde, tanto la especie *Pseudomonas protegens* como la especie *Paenibacillus sp.*, se encontraban presentes en mayor abundancia relativa. Además, resulta interesante observar que, gracias a este enfoque, una especie de *Ochrobactrum sp* pudo ser caracterizada, lo que nos muestra la capacidad que tiene este tipo de enfoque de secuenciación metagenómico en dilucidar la denominada “materia oscura” refiriéndose a todas las especies microbianas que no han podido ser caracterizadas mediante técnicas de cultivo tradicionales. A continuación se muestra en forma detallada dicha anotación funcional realizada en la plataforma RAST.

**Tabla 3.** Alineamiento BLAST de las secuencias del gen ribosomal 16S de cada uno de los Contigs ensamblados

Contig	Probable microorganismo	% Identidad	Valor E	Código de acceso
1	<i>Pseudomonas protegens</i>	100	0.0	<a href="#">CP043179.1</a>
2	<i>Paenibacillus sp.</i>	98.52	0.0	<a href="#">KF011594.1</a>
4	<i>Ochrobactrum sp.</i>	99.38	0.0	<a href="#">KF479631.1</a>

Luego de extraer los genes ribosomales 16S,



**Tabla 4.** Resumen de las características más importante obtenidas en la anotación funcional usando la plataforma RAST de cada uno de los contigs ensamblados

Categoría	MAG 1	MAG 2	MAG 4
Cofactores, Vitaminas, Grupos prostéticos, Pigmentos	226	131	110
Pared celular y cápside	44	52	24
Virulencia, Patogenicida, Defensa	68	42	35
Metabolismo de Potasio	13	5	3
Fotosíntesis	0	0	0
Miscellaneous	48	8	14
Fagos, Profagos, Elementos transponibles, Plásmidos	6	5	5
Transporte de membrana	119	11	31
Adquisición y Metabolismo de Hierro	42	19	0
Metabolismo de ARN	55	57	26
Nucleosidos y Nucleótidos	109	101	63
Metabolismo de proteínas	225	142	147
División y Ciclo Celular	0	3	0
Motilidad y quemotaxis	72	0	0
Regulación and Señalización celular	70	12	10
Metabolismo Secundario	4	5	4
Ácidos grasos, lípidos, isoprenoides	151	50	25
Metabolismo de Nitrógeno	17	5	0
Dormancia y Esporulación	1	32	1
Respiración	117	45	78
Respuesta a Estrés	108	37	44
Metabolismo de Compuestos Aromáticos	126	5	7
Aminoácidos y derivados	531	249	160
Metabolismo del azúfre	33	14	3
Metabolismo del fósforo	41	32	20
Carbohidratos	299	198	108

De acuerdo con los resultados obtenidos en la anotación funcional usando la plataforma RAST de cada uno de los MAGs ensamblados, el MAG 1 presenta el mayor número lecturas asociadas a aminoácidos y derivados de estos, seguido por la categoría de cofactores, vitaminas y pigmentos; para el MAG 2 se tienen una mayor cantidad de

proteínas asociadas a la esporulación y dormancia lo cual tiene sentido, pues previamente algunas especies de *Paenibacillus* presentan esporas (Redondo, Jaminson & Fernández, 2005).

**Tabla 5.** Anotación funcional usando la plataforma CAZy de cada uno de los MAGs ensamblados

<b>Categoría</b>	<b>MAG 1</b>	<b>MAG 2</b>	<b>MAG 4</b>
Carbohydrate binding modules (CBM)	9	10	4
Glycosyl hydrolases (GH)	46	90	23
Glycosyl Transferases(GT)	80	29	27
GH associated with GT	1	0	1
CBM associated with GH	6	26	0
Auxiliary activities associated with CBM	2	1	0
Auxiliaryactivities (AA)	6	2	4
Carbohydrate esterases (CE)	6	9	5
Polysaccharide Lyase (PL)	3	1	1

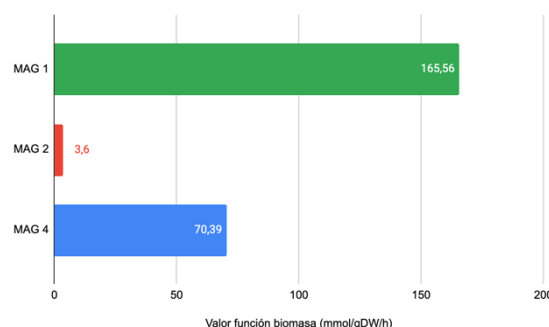
En la tabla anterior se observa como el MAG 1, presenta una mayor cantidad de anotaciones a enzimas relacionadas con glicosil transferasas y glicosil hidrolasas, potencial similar al del MAG 2 pero se puede observar una diferencia notable en la distribución de dichas anotaciones; por

ejemplo, el MAG 2 presenta casi el doble de anotaciones de glicosil hidrolasas y por su parte, el MAG 1 posee más del doble de anotaciones relacionadas con glicosil transferasas.

Luego, el MAG que posee la mayor cantidad de anotaciones es el MAG 2 con 168 anotaciones, seguido por el MAG 1 con 159. finalizando con el MAG 4 que posee 65 anotaciones.

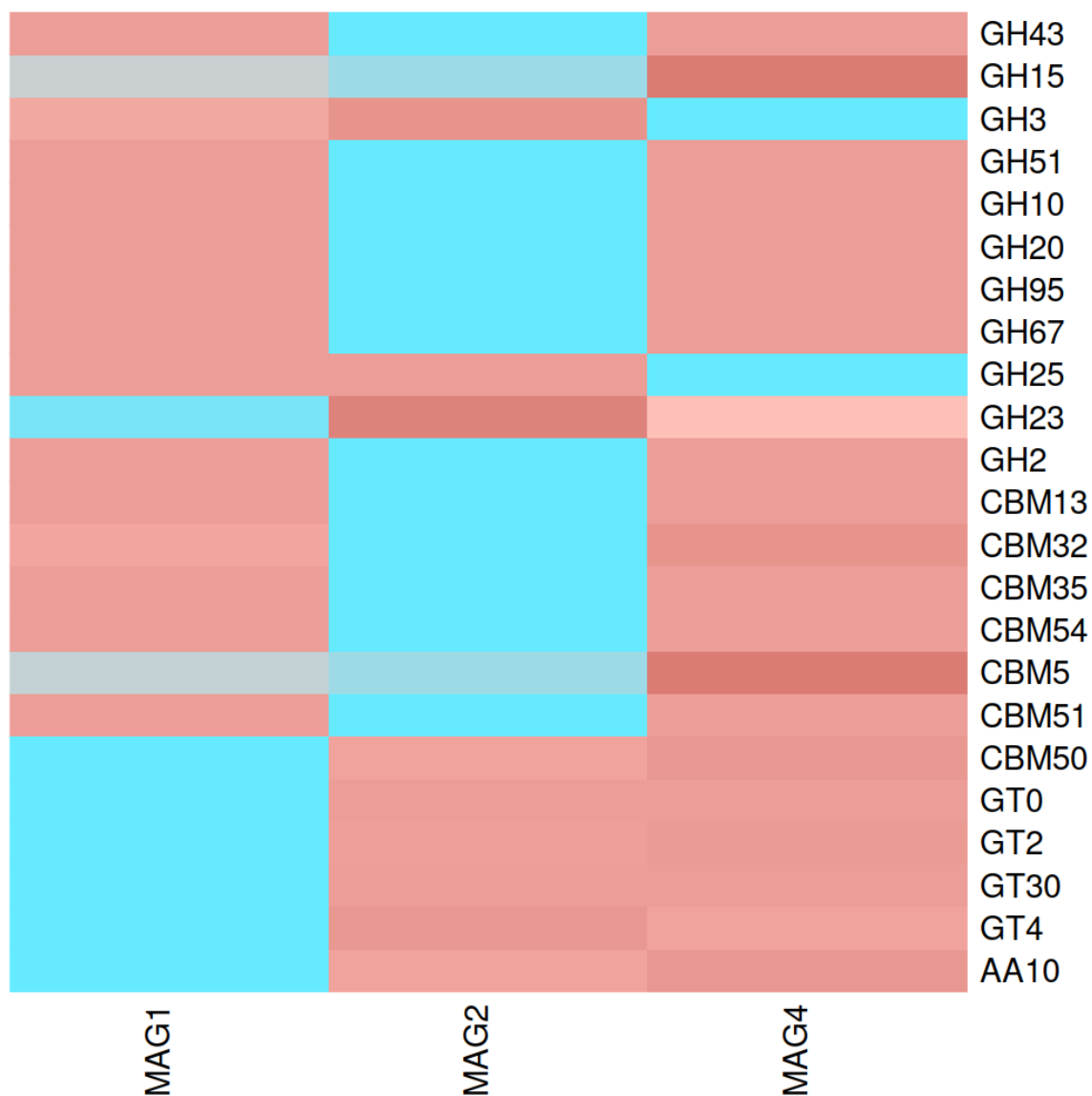
Un aspecto interesante de la tabla anterior es el hecho de tener un número tan dispar en las diferentes enzimas anotadas a lo largo de todos los MAGs; particularmente, los MAGs 1 y 2 asociados a *Pseudomonas protegens* y *Paenibacillus sp*, presentan un posible sinergismo enzimático mientras que el rol de *Ochrobactrum* no es muy claro puesto que la cantidad de enzimas asociadas a la degradación de lignocelulosa.

Una vez se han anotado funcionalmente y caracterizado taxonómicamente los MAGs que se creen son posibles genomas de las especies presentes en el consorcio mínimo efectivo MELMC, se realiza la ejecución de un análisis de variabilidad de flujo FBA por sus siglas en inglés, con las reconstrucciones automáticas generadas para modelos metabólicos a escala genómica desarrollada por la plataforma MODELSEED, con el fin de determinar si cada uno de estos posibles genomas, tienen la capacidad de generar de generar biomasa celular a partir de los precursores esenciales dentro de los que se encuentran ácidos nucleicos, proteínas y demás compuestos presentes en la célula. La clasificación taxonómica y funcional de dichos MAGs obedece al desarrollo de los objetivos propuestos y por lo tanto fue realizada obteniendo las siguientes conclusiones: la especie encontrada en la primera fase de ensamblaje de metagenomas corresponde a *Ochrobactrum sp* para el MAG 4. Dicha clasificación se realizó usando el marcador filogenético del gen ribosomal 16S.



**Figura 8.** Valor función biomasa (mmol/gDW/h).

De la figura 8 resulta interesante ver como la anotación realizada previamente permite la generación de biomasa por cada uno de los MAGs evaluados. Aunque ese valor es bastante elevado para las condiciones de crecimiento normales que tendría una célula, permiten brindar una descripción de forma cualitativa del estado de la reconstrucción del metabolismo celular de cada especie ensamblada, una posible estrategia para refinar el valor de dicho valor de biomasa, comúnmente denominada función objetivo, es la curación de las posibles patologías de red asociadas a errores de la anotación o con ciclos termodinámicamente inviables como se muestra en gran variedad de artículos especializados en modelos a escala genómica, comúnmente denominados GEMM, por sus siglas en inglés.



**Figura 9.** *Heatmap con la cantidad de genes asociados a cada familia cazy*

Para la elaboración del mapa de calor se consideran el número de genes asociadas a cada una de las familias cazy obtenidas en la anotación funcional de cada uno de los MAGS obtenidos en el proceso.

Basados en la anotación funcional realizada con la base de datos dbCAN, 11 glicosil hidrolasas, abreviadas de ahora en adelante como GH. Dentro de las familias predichas se encuentran familias relacionadas con la degradación de hemicelulosa; específicamente se encontró que el MAG2

presentaba un elevado número de genes asociados a las familias denotadas como GH10 conocidas como xilanasas, así como familias GH3 y GH43 involucradas en la degradación y arabinosa, los cuales hacen parte de dichos polímeros vegetales. Además, se predicen funcionalmente las enzimas glicosil transferasas denotadas de ahora en adelante como GT, en donde se destacan las enzimas GT2 y GT4 involucradas en los procesos de (de)polimerización de Sucrosa.

Por otro lado, la especie *Pseudomonas protegenes* (asociada al MAG1) presentaba un elevado número de genes asociados a las familias denotadas como CBM (módulos de unión a carbohidratos) las cuales en su mayoría estaban asociadas a dominios proteicos de unión a diversos carbohidratos. Otra familia de enzimas con elevado número de genes asociados a este MAG fueron la familia de enzimas con actividad auxiliar denotada como AA, siendo la AA10, una monooxigenasa dependiente de cobre recientemente catalogada como LPMO, la enzima más abundante.

El proceso de anotación funcional se acorde con los resultados de anotación realizado por Díaz-García et al., (2020) puesto que resalta el potencial asociado a *Paenibacillus sp* (asociada al MAG2 en este trabajo) como miembro clave dentro del consorcio mínimo debido a su elevado potencial de expresión de enzimas involucradas en procesos de despolimerización de material lignocelulósico. Además, este enfoque amplió la anotación funcional de cazymes realizadas por Díaz-García et al., (2020) puesto que describe no solamente algunos de los genes asociados a las familias GH sino que además describe las familias CBM, GT y AA, las cuales son fundamentales en este tipo de procesos de depolimerización.

Por otro lado, de acuerdo con las hipótesis planteadas por Díaz-García et al.,(2020) *Paenibacillus sp* podría tener una afinidad más alta por sustratos específicos como compuestos derivados de la celulosa. Además, en el mismo artículo se plantea el posible rol de la especie *Pseudomonas* en los procesos de depolimerización y procesos catabólicos, conduciéndolos a tener una afinidad más alta por sustratos aromáticos

provenientes de compuestos derivados de la lignina y usar como fuente de carbono a la glucosa. Este último planteamiento pudo ser, al menos de forma parcial, descrito por los modelos metabólicos tipo borrador asociados a los MAGS, en donde se observa claramente la mayor velocidad de crecimiento específica por parte del MAG1 con respecto a los demás. Es importante aclarar que estos resultados deben ser tomados como una descripción cualitativa del comportamiento de los miembros del consorcio y no una descripción cuantitativa. Esto debido a que los modelos aún no han sido sometidos a un proceso de curación adecuado y por ende presentan un porcentaje importante de patologías en su red metabólica que produzcan sobreestimaciones de las funciones celulares y por ende, alejen al modelo de las condiciones de crecimiento in vivo. Además, una limitante importante de los resultados anteriormente presentados es la imposibilidad de recrear las condiciones de crecimiento de las especies de forma individual en otro tipo de sustratos distintos a la glucosa, y por ende evaluar las velocidades de crecimiento específicas de los MAGS con mayor capacidad de crecimiento en sustratos complejos tal y como se reporta gracias a su producción de enzimas como se observa en la sección de anotación funcional.

Dentro de las perspectivas futuras de este trabajo, se encuentra realizar un proceso de binning para agrupar la diversidad de contigs lineales (reportados en la sección de anexos de este documento) que no pudieron ser caracterizados taxonómicamente, para luego obtener secuencias mucho más largas que permitan realizar una correcta clasificación; además, se llevará a cabo un

proceso de curación de los modelos metabólicos que sirvan como insumo para la creación de un modelo de comunidad, que permita describir de forma adecuada las interacciones ecológicas presentes en los miembros del consorcio. Una vez establecido el modelo de comunidad, se podría implementar un análisis de velocidades de crecimiento ante diversas fuentes de carbono distintas de glucosa y que se encuentren presentes en los residuos lignocelulósicos como por ejemplo la xilosa.

## MATERIAL SUPLEMENTARIO:

Todos los resultados de los ensamblajes se pueden observar en el siguiente hipervínculo:

<https://github.com/davidarturolr/BCOM4102/tree/master/Proyecto>

## REFERENCIAS

- Arpia, A., Chen, W., Lam, S., Rousset, P., & de Luna, M. (2021). Sustainable biofuel and bioenergy production from biomass waste residues using microwave-assisted heating: A comprehensive review. *Chemical Engineering Journal*, 403, 126233. doi: 10.1016/j.cej.2020.12623
- Li, H., Liang, Y., Li, P., & He, C. (2020). Conversion of biomass lignin to high-value polyurethane: A review. *Journal Of Bioresources And Bioproducts*, 5(3), 163-179. doi: 10.1016/j.jobab.2020.07.002.
- Muscat, A., de Olde, E., de Boer, I., & Ripoll-Bosch, R. (2020). The battle for biomass: A systematic review of food-feed-fuel competition. *Global Food Security*, 25, 100330. doi: 10.1016/j.gfs.2019.100330.
- Singhvi, M., & Gokhale, D. (2019). Lignocellulosic biomass: Hurdles and challenges in its valorization. *Applied Microbiology And Biotechnology*, 103(23-24), 9305-9320. doi.org/10.1007/s00253-019-10212-7
- Peñaranda Gonzalez, L., Montenegro Gómez, S., & Giraldo Abad, P. (2017). Aprovechamiento de residuos agroindustriales en Colombia. *Revista De Investigación Agraria Y Ambiental*, 8(2), 141-150. doi: 10.22490/21456453.2040
- Ren, J., Cao, J., Zhao, X., Yang, F., & Wei, X. (2019). Recent advances in syngas production from biomass catalytic gasification: A critical review on reactors, catalysts, catalytic mechanisms and mathematical models. *Renewable And Sustainable Energy Reviews*, 116, 109426. doi: 10.1016/j.rser.2019.109426
- Eskicioglu, C., Monlau, F., Barakat, A., Ferrer, I., Kaparaju, P., Trably, E., & Carrère, H. (2017). Assessment of hydrothermal pretreatment of various lignocellulosic biomass with CO<sub>2</sub> catalyst for enhanced methane and hydrogen production. *Water Research*, 120, 32-42. doi: 10.1016/j.watres.2017.04.068
- Rahmati, S., Doherty, W., Dubal, D., Atanda, L., Moghaddam, L., & Sonar, P. et al. (2020). Pretreatment and fermentation of lignocellulosic biomass: reaction mechanisms and process engineering. *Reaction Chemistry & Engineering*, 5(11), 2017-2047. doi: 10.1039/d0re00241k
- Ju, J., Wang, D., Heo, S., Kim, M., Seo, J., & Kim, Y. et al. (2020). Enhancement of 1,3-propanediol production from industrial by-product by *Lactobacillus reuteri* CH53. *Microbial Cell Factories*, 19(1). doi: 10.1186/s12934-019-1275-x
- Sadhukhan, S., Villa, R., & Sarkar, U. (2016). Microbial production of succinic acid using crude and purified glycerol from a *Crotalaria juncea* based biorefinery. *Biotechnology Reports*, 10, 84-93. doi: 10.1016/j.btre.2016.03.008

- Jia, Y., Niu, C., Lu, Z., Zhang, X., Chai, L., & Shi, J. et al. (2020). A Bottom-Up Approach To Develop a Synthetic Microbial Community Model: Application for Efficient Reduced-Salt Broad Bean Paste Fermentation. *Applied And Environmental Microbiology*, 86(12). doi: 10.1128/aem.00306-20
- Puentes-Téllez, P., & Falcao Salles, J. (2018). Construction of Effective Minimal Active Microbial Consortia for Lignocellulose Degradation. *Microbial Ecology*, 76(2), 419-429. doi: 10.1007/s00248-017-1141-5.
- Jiménez, D., de Lima Brossi, M., Schückel, J., Kračun, S., Willats, W., & van Elsas, J. (2016). Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. *Applied Microbiology And Biotechnology*, 100(24), 10463-10477. doi: 10.1007/s00253-016-7713-3.
- Jiménez, D., Wang, Y., Chaib de Mares, M., Cortes-Tolalpa, L., Mertens, J., & Hector, R. et al. (2019). Defining the eco-enzymological role of the fungal strain *Coniochaeta* sp. 2T2.1 in a tripartite lignocellulolytic microbial consortium. *FEMS Microbiology Ecology*, 96(1). doi: 10.1093/femsec/fiz186
- Jiménez, D., Chaib De Mares, M., & Salles, J. (2018). Temporal Expression Dynamics of Plant Biomass-Degrading Enzymes by a Synthetic Bacterial Consortium Growing on Sugarcane Bagasse. *Frontiers In Microbiology*, 9. doi: 10.3389/fmicb.2018.00299
- Serrano-Gamboa, J., Rojas-Herrera, R., González-Burgos, A., Folch-Mallol, J., Jiménez, D., & Sánchez-González, M. (2019). Degradation profile of nixtamalized maize pericarp by the action of the microbial consortium PM-06. *AMB Express*, 9(1). doi: 10.1186/s13568-019-0812-7.
- Minty, J., Singer, M., Scholz, S., Bae, C., Ahn, J., & Foster, C. et al. (2013). Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proceedings Of The National Academy Of Sciences*, 110(36), 14592-14597. doi: 10.1073/pnas.1218447110.
- Gilmore, S., Lankiewicz, T., Wilken, S., Brown, J., Sexton, J., & Henske, J. et al. (2019). Top-Down Enrichment Guides in Formation of Synthetic Microbial Consortia for Biomass Degradation. *ACS Synthetic Biology*, 8(9), 2174-2185. doi: 10.1021/acssynbio.9b00271.
- Sangwan, N., Xia, F., & Gilbert, J. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, 4(1). doi.org/10.1186/s40168-016-0154-5
- Chen, L., Anantharaman, K., Shaiber, A., Eren, A., & Banfield, J. (2020). Accurate and complete genomes from metagenomes. *Genome Research*, 30(3), 315-333. doi.org/10.1101/gr.258640.119
- White, R., Bottos, E., Roy Chowdhury, T., Zucker, J., Brislawn, C., & Nicora, C. et al. (2016). MolecuLo Long-Read Sequencing Facilitates Assembly and Genomic Binning from Complex Soil Metagenomes. *Msystems*, 1(3). doi.org/10.1128/msystems.00045-16.
- Driscoll, C., Otten, T., Brown, N., & Dreher, T. (2017). Towards long-read metagenomics: complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture. *Standards In Genomic Sciences*, 12(1). doi.org/10.1186/s40793-017-0224-8
- Anantharaman, K., Hausmann, B., Jungbluth, S., Kantor, R., Lavy, A., & Warren, L. et al. (2018). Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *The ISME*

Journal, 12(7), 1715-1728.  
doi.org/10.1038/s41396-018-0078-0

Ward, L., Shih, P., & Fischer, W. (2018). MetaPOAP: presence or absence of metabolic pathways in metagenome-assembled genomes. *Bioinformatics*. doi.org/10.1093/bioinformatics/bty510

Momeni, B., Xie, L., & Shou, W. (2017). Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *Elife*, 6. doi.org/10.7554/elife.25051

Gore, J., Youk, H., & van Oudenaarden, A. (2009). Snowdrift game dynamics and facultative cheating in yeast. *Nature*, 459(7244), 253-256. doi.org/10.1038/nature07921

Cao, X., Hamilton, J., & Venturelli, O. (2018). Understanding and Engineering Distributed Biochemical Pathways in Microbial Communities. *Biochemistry*, 58(2), 94-107. doi: 10.1021/acs.biochem.8b01006

Volkova, S., Matos, M., Mattanovich, M., & Marín de Mas, I. (2020). Metabolic Modelling as a Framework for Metabolomics Data Integration and Analysis. *Metabolites*, 10(8), 303. doi: 10.3390/metabo10080303

Lewis, N., Nagarajan, H., & Palsson, B. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4), 291-305. doi: 10.1038/nrmicro2737.

Merino, M., Andrews, B., & Asenjo, J. (2014). Stoichiometric model and flux balance analysis for a mixed culture of *Leptospirillum ferriphilum* and *Ferroplasma acidiphilum*. *Biotechnology Progress*, 31(2), 307-315. doi: 10.1002/btpr.2028.

Hunt, K., Jennings, R., Inskeep, W., & Carlson, R. (2016). Stoichiometric modelling of assimilatory and dissimilatory biomass utilization in a microbial community. *Environmental Microbiology*, 18(12), 4946-4960. doi: 10.1111/1462-2920.13444.

Zhuang, K., Ma, E., Lovley, D., & Mahadevan, R. (2012). The design of long-term effective uranium bioremediation strategy using a community metabolic model. *Biotechnology And Bioengineering*, 109(10), 2475-2483. doi: 10.1002/bit.24528.

Hamilton, J., Calixto Contreras, M., & Reed, J. (2015). Thermodynamics and H<sub>2</sub> Transfer in a Methanogenic, Syntrophic Community. *PLOS Computational Biology*, 11(7), e1004364. doi: 10.1371/journal.pcbi.1004364.

Wolfsberg, E., Long, C., & Antoniewicz, M. (2018). Metabolism in dense microbial colonies: <sup>13</sup>C metabolic flux analysis of *E. coli* grown on agar identifies two distinct cell populations with acetate cross-feeding. *Metabolic Engineering*, 49, 242-247. doi: 10.1016/j.ymben.2018.08.013.

Reed, J., Patel, T., Chen, K., Joyce, A., Applebee, M., & Herring, C. et al. (2006). Systems approach to refining genome annotation. *Proceedings Of The National Academy Of Sciences*, 103(46), 17480-17484. doi.org/10.1073/pnas.0603364103

Burgard, A., & Maranas, C. (2003). Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnology And Bioengineering*, 82(6), 670-677. doi.org/10.1002/bit.10617

Segre, D., Vitkup, D., & Church, G. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings Of The National Academy Of Sciences*, 99(23), 15112-15117.



- doi.org/10.1073/pnas.232349399  
Schellenberger, J., Lewis, N., & Palsson, B. (2011). Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models. *Biophysical Journal*, 100(3), 544-553. doi.org/10.1016/j.bpj.2010.12.3707
- Shen, Y., Liu, J., Estiu, G., Isin, B., Ahn, Y., & Lee, D. et al. (2010). Blueprint for antimicrobial hit discovery targeting metabolic networks. *Proceedings Of The National Academy Of Sciences*, 107(3), 1082-1087. doi.org/10.1073/pnas.0909181107
- Edwards, J., & Palsson, B. (2000). The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings Of The National Academy Of Sciences*, 97(10), 5528-5533. doi.org/10.1073/pnas.97.10.5528
- Simmons, C., Reddy, A., D'haeseleer, P., Khudyakov, J., Billis, K., & Pati, A. et al. (2014). Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnology For Biofuels*, 7(1). doi: 10.1186/s13068-014-0180-0
- Lopes, A., Ferreira Filho, E., & Moreira, L. (2018). An update on enzymatic cocktails for lignocellulose breakdown. *Journal Of Applied Microbiology*, 125(3), 632-645. doi.org/10.1111/jam.13923
- Lawson, C., Harcombe, W., Hatzepichler, R., Lindemann, S., Löffler, F., & O'Malley, M. et al. (2019). Common principles and best practices for engineering microbiomes. *Nature Reviews Microbiology*, 17(12), 725-741. doi: 10.1038/s41579-019-0255-9
- Díaz-García, L., Huang, S., Spröer, C., Sierra-Ramírez, R., Bunk, B., Overmann, J., & Jiménez, D. (2020). Dilution-to-Stimulation/Extinction Method: a Combination Enrichment Strategy To Develop a Minimal and Versatile Lignocellulolytic Bacterial Consortium. *Applied And Environmental Microbiology*, 87(2). doi.org/10.1128/aem.02427-20.
- Kang, D., Jacquioud, S., Herschend, J., Wei, S., Nesme, J., & Sørensen, S. (2020). Construction of Simplified Microbial Consortia to Degrade Recalcitrant Materials Based on Enrichment and Dilution-to-Extinction Cultures. *Frontiers In Microbiology*, 10. doi: 10.3389/fmicb.2019.03010
- Lewis, N., Nagarajan, H., & Palsson, B. (2012). Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4), 291-305. doi: 10.1038/nrmicro2737.
- Seaver, S., Liu, F., Zhang, Q., Jeffries, J., Faria, J., & Edirisinghe, J. et al. (2020). The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Research*, 49(D1), D575-D588. doi: 10.1093/nar/gkaa746
- Mendoza, S., Olivier, B., Molenaar, D., & Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biology*, 20(1). doi: 10.1186/s13059-019-1769-1
- Hucka, M., Finney, A., Sauro, H., Bolouri, H., Doyle, J., & Kitano, H. et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4), 524-531. doi: 10.1093/bioinformatics/btg015
- Tiwari, K., Kananathan, S., Roberts, M., Meyer, J., Sharif Shohan, M., & Xavier, A. et al. (2021). Reproducibility in systems biology modelling. *Molecular Systems Biology*, 17(2). doi: 10.15252/msb.20209982

Lieven, C., Beber, M., Olivier, B., Bergmann, F., Ataman, M., & Babaei, P. et al. (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, 38(3), 272-276. doi: 10.1038/s41587-020-0446-y

*FEMS Microbiology Ecology*, 96(1). doi: 10.1093/femsec/fiz186

García-Jiménez, B., García, J., & Nogales, J. (2018). FLYCOP: metabolic modeling-based analysis and engineering microbial communities. *Bioinformatics*, 34(17), i954-i963. doi: 10.1093/bioinformatics/bty561.

Roell, G., Zha, J., Carr, R., Koffas, M., Fong, S., & Tang, Y. (2019). Engineering microbial consortia by division of labor. *Microbial Cell Factories*, 18(1). doi: 10.1186/s12934-019-1083-3

Bernstein, H. C., & Carlson, R. P. (2012). Microbial Consortia Engineering for Cellular Factories: in vitro to in silico systems. *Computational and structural biotechnology journal*, 3, e201210017. doi.org/10.5936/csbj.201210017.

Kolmogorov, M., Bickhart, D., Behsaz, B., Gurevich, A., Rayko, M., & Shin, S. et al. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nature Methods*, 17(11), 1103-1110. doi.org/10.1038/s41592-020-00971-x

Ryan R. Wick, Mark B. Schultz, Justin Zobel, Kathryn E. Holt. (2015). Bandage: interactive visualization of de novo genome assemblies, *Bioinformatics*, doi.org/10.1093/bioinformatics/btv383

Gan, Z., & Zhang, H. (2019). PMBD: a Comprehensive Plastics Microbial Biodegradation Database. Database, 2019. doi: 10.1093/database/baz119

Jiménez, D., Wang, Y., Chaib de Mares, M., Cortes-Tolalpa, L., Mertens, J., & Hector, R. et al. (2019). Defining the eco-enzymological role of the fungal strain *Coniochaeta* sp. 2T2.1 in a tripartite lignocellulolytic microbial consortium.

<b>Categoría</b>	<b>MAG 1</b>	<b>MAG 2</b>	<b>MAG 4</b>
Carbohydrate binding modules (CBM)	CBM50, CBM32	CBM32, CBM34, CBM54, CBM9	CBM50
Glycosyl hydrolases (GH)	GH19, GH13_30 GH24, GH65, GH28, GH3, GH23, GH13_3 GH13_11, GH73GH28, GH6	GH10, GH101, GH109, GH127 GH130, GH30_3 GH13_41 GH0 GH136, GH16_3 GH15, GH18 GH2, GH20 GH3 GH23 GH26, GH27 GH31, GH35, GH36, GH42	GH26, GH25, GH6 GH28, GH36, GH3 GH65, GH0, GH19, GH23, GH4 GH32, GH51 GH102, GH103GH13_11
Glycosyl Transferases (GT)	GT4, GT13, GT2 GT87, GT1, GT51	GT1, GT26, GT28GT2, GT4, GT5 GT51	GT13, GT19, GT2 GT25, GT32, GT4 GT28, GT51, GT83
CBM associated with GH	CBM48+GH13_9CB M48+GH13_11 CBM50+GH23	CBM32+GH16_24 GH81+CBM32 GH13+CBM20+ CBM34+CBM48	-
Auxiliary activities associated with CBM	AA10+CBM73	AA10+CBM5	-
Auxiliary activities (AA)	AA10, AA3, AA6, AA7, AA3_2	AA1_3, AA7	AA1, AA3_2
Polysaccharide Lyase	PL7_1, PL7_2	PL9_2	PL0

Nombre de la secuencia	Longitud (pb)	Cobertura	Circular
contig_5	968722	523	N
contig_2	800267	506	N
contig_10	794134	497	N
contig_18	387216	451	N
scaffold_31	332917	402	N
contig_44	195721	453	N
contig_4	169932	461	N
contig_80	167840	462	N
contig_53	139717	523	N
contig_38	121626	490	N
contig_37	106763	487	N
contig_260	85703	455	N
contig_289	75854	416	N
scaffold_96	70106	440	N
contig_146	62951	537	N
contig_391	52066	397	N
contig_99	36558	430	N
contig_384	31796	475	N
contig_345	17562	358	N

**Nota:** Esta tabla posee la información de los contigs cuya longitud supera 10000pb. La totalidad de los contigs puede ser observada en el material suplementario.