

# **Análisis bioinformático de las comunidades microbianas provenientes de una muestra de bosque andino empleadas en la construcción de un consorcio con actividad lignocelulolítica.**

<sup>1</sup>Díaz-Rodríguez Carlos Andrés, <sup>2</sup> León David, <sup>3</sup> Patiño 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>*Maestría en Biología Computacional, Facultad de Ingeniería, Universidad de los Andes.  
Colombia*

## **INTRODUCCIÓN**

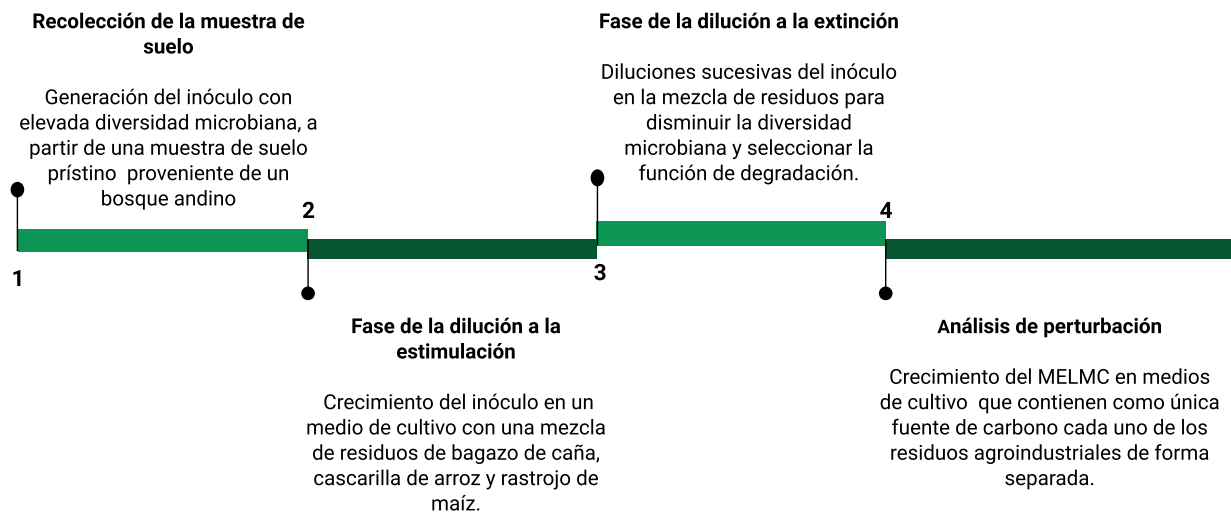
La transición energética desde fuentes no renovables hacia fuentes amigables con el medio ambiente es de vital importancia para todos los países en el mundo. Este cambio ha sido apalancado gracias a la implementación de una economía verde también llamada bioeconomía. Esta nueva economía, busca el aprovechamiento 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; Kohli et al., 2019); compuestos para la industria alimenticia (Mahro & Timm, 2007; 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 siguiente metodología



**Figura 1.** *Flujo de trabajo para la construcción del consorcio mínimo efectivo degradador de material 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 ASVs 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 última 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 bioinformática 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 la mayor cantidad de genomas bacterianos provenientes del consorcio mínimo efectivo degradador de lignocelulosa MELMC.
2. Elaborar la caracterización funcional de cada uno de los miembros del consorcio, 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 algunos de los microorganismos presentes en el consorcio, con el fin de dilucidar su capacidad para crecer en diversos sustratos.

## 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 LeibnizDSMZ Colección de microorganismos y cultivos celulares, instituto en donde fue llevada a cabo la secuenciación.

### Reconstrucción de genomas a partir de metagenomas

En la primera etapa, se utilizó el metagenoma proveniente de la transferencia en la que se obtiene el consorcio mínimo efectivo con actividad lignocelulolítica (MELMC). Puesto que dicha transferencia presentaba una diferencia significativa tanto en términos de degradación de material vegetal como en términos de diversidad de microorganismos con las demás muestras generadas en las demás fases del ensamblaje.

Para realizar el ensamblaje de las lecturas crudas, se utiliza la siguiente línea de comando:

```
#!/bin/bash
#SBATCH --mem=90G # Memoria Virtual/RAM, Default=2048
#SBATCH -t 48:00:00 # Walltime
#SBATCH -n 8

flye --pachio-hifi 12D.hifi.fastq.gz --out-dir resultados_12D -g 12000000000 --plasmids --meta
```

A continuación, se presenta una breve descripción asociada a cada uno de los flags utilizados

#### Memoria RAM 90GB

- Número de núcleos: 8
- Walltime 48h
- Se usa el flag para definir la característica de la secuencia: **--pachio-hifi**. Además, se implementó el flag **--resume** para retomar la ejecución del ensamblador en caso de que se presentaran fallas, también se usó el flag **--meta** para definir la información que se presentó de entrada como un metagenoma; por otra parte, se definió el tamaño de la muestra en bytes usando el flag **-g** el cual corresponde al tamaño total del metagenoma de cada una de las muestras.
- Finalmente, se usa el flag **-out-dir** para especificarle al sistema que cada output es un directorio único correspondiente a una de las dos muestras. En dicho directorio se retorna un conjunto de contigs que se usan para el análisis y la caracterización de las muestras, además de un archivo en el que se presenta un resumen de los resultados obtenidos, dicha sección es descrita más adelante.

Una vez se identificaron los archivos de salida del ensamblador, se separaron los contigs 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

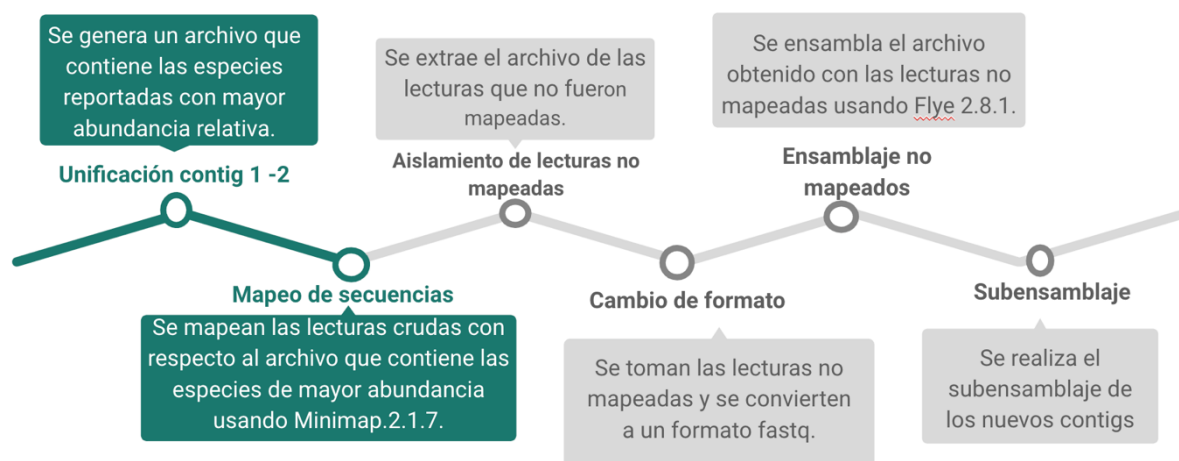
Para realizar el proceso de identificación taxonómica y funcional de los contigs obtenidos durante el proceso de ensamblaje, se siguió el siguiente flujo detallado de trabajo:

Ensamblaje de genomas a partir de metagenomas	Se toma <i>12D.hifi.fastq.gz</i> y se realiza el ensamblaje de genomas usando Flye 2.8.1
Aislamiento de contigs	Se toman los datos del metagenoma almacenados en <i>assemb</i> y <i>fasta</i> y se separa en los contigs asociados a dicho archivo
Anotación funcional	Se usa la plataforma RAST (Rapid Annotation using Subsystem Technology) para la anotación funcional de cada contig
Búsqueda y selección del gen ribosomal ARN 16S	Se aíslan los genes ribosomales ARN 16S asociados a cada contig.
Identificación Taxonómica	Se ejecuta un BLASTn usando los datos 16S contra NCBI para la anotación taxonómica de cada contig.
Identificación de especies de mayor abundancia	Se identifican los contigs asociados a las especies de mayor abundancia reportados por Díaz-García et al. (2020)

**Figura 2.** Flujo de trabajo para la anotación taxonómica y funcional de los posibles genomas ensamblados.

## Validación de los ensamblajes

Para corroborar que el proceso de ensamblaje de los posibles genomas fue realizado correctamente, y los contigs asociados a los posibles genomas sí poseen validez en términos biológicos, se realizó un proceso de mapeo de las secuencias crudas contra los contigs de las especies reportadas con mayor abundancia, para subensamblar las lecturas no mapeadas y determinar si es posible la generación de nuevos contigs. Dicho proceso se describe a continuación.



**Figura 3.** Flujo de trabajo para la anotación taxonómica y funcional de los posibles genomas ensamblados.

### Reconstrucción de los modelos metabólicos borrado

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 borrado 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

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.

Luego del proceso de ensamblaje, se revisaron los archivos de salida. Estos archivos fueron categorizados por el ensamblador de la siguiente manera: **00-assembly**, directorio donde se encuentra un borrador del ensamblaje en formato fasta; **10-consensus**, directorio donde se encuentran trozos del ensamblaje en formato fasta; **20-repeat**, directorio donde se encuentran las distintas representaciones gráficas de los contigs circulares representativos del proceso de ensamblaje del metagenoma; **30-contigger**, directorio donde se encuentran las distintas representaciones gráficas de los contigs circulares preliminares y los ensamblajes sin pulir; **40-**

**polishing**, directorio donde se encuentran los contigs filtrados en formato fasta y las representaciones gráficas de los contigs circulares con bordes pulidos; **22-plasmids**, directorio donde se encontrarían secuencias de plásmidos de ser encontradas.

Además de esto, también se presenta los siguientes archivos:

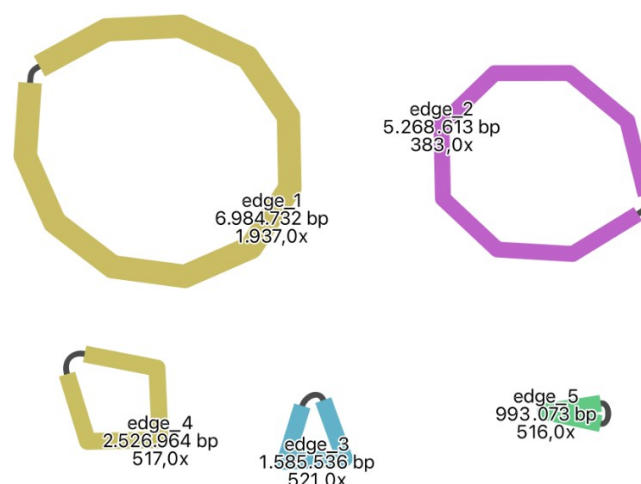
1. **Assembly.fasta**: archivo que contiene el ensamblaje final en formato FASTA.
2. **Assembly\_info.txt**: archivo en el que se encuentra una tabla donde se observa el número de contigs identificados, así como su respectiva cobertura y su tamaño.
3. **Assembly\_graph.gfa**: archivo que contiene representación gráfica del contig circular final representativo del ensamblaje
4. **Flye.log**: archivo que contiene la información del proceso de ensamblaje del metagenoma y su ejecución por pasos.
5. **Assembly\_graph.gv**: archivo correspondiente a los metadatos de la representación gráfica del contig circular ensamblado.

Una vez se identificaron los archivos de salida del ensamblador, se separaron los contigs 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 metagenomas, en los cuales se observa su longitud en pares de bases y su cobertura; así como su respectiva representación gráfica del contig circular ensamblado en el cual se denota la longitud y la profundidad de la lectura.

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

Nombre de la secuencia	Longitud (pb)	Cobertura	Circular
Contig_1	6984732	991	Sí
Contig_2	5268613	403	Sí
Contig_3	1585536	535	Sí
Contig_4	2526964	531	Sí
Contig_5	993073	528	Sí

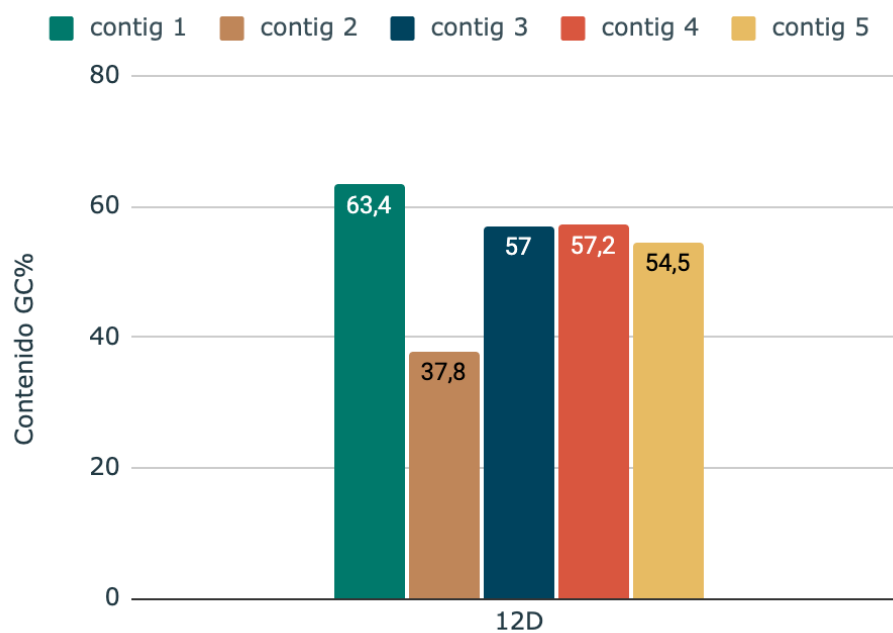


**Figura 4.** Contigs circulares obtenidos del ensamblaje de la muestra 12D. Graficados por Bandage (Ryan et al., 2015)



Como se puede observar, en general se tienen muy buenas profundidades de secuenciación. De hecho, la profundidad del contig 1, representado en 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 contigs poseen una profundidad de tres a cinco veces la recomendada para este tipo de análisis de datos (Sims et al., 2014).

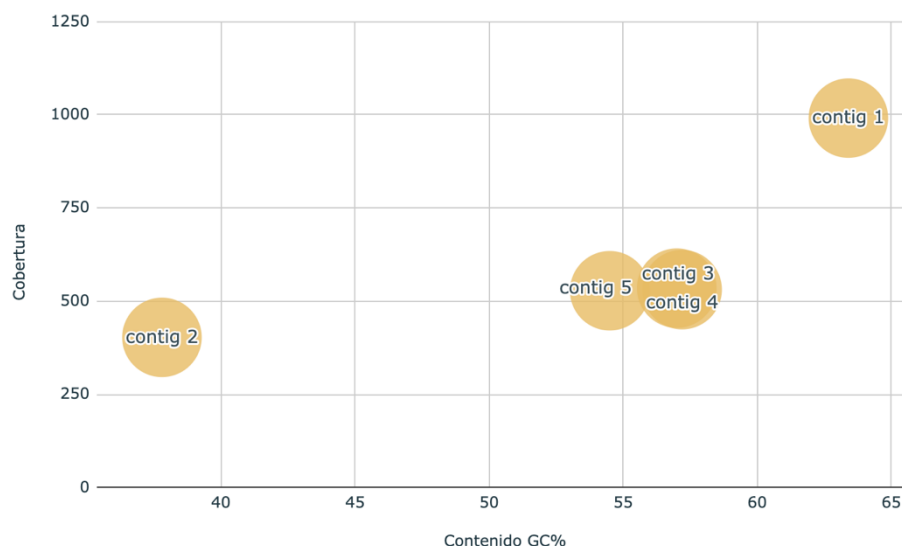
Posteriormente se utilizó la herramienta DFAST (DDBJ Fast Annotation and Submission Tool) para obtener el contenido de guanina y citosina de cada uno de los contigs, como se puede observar a continuación.



**Figura 5.** *Contenido GC de los contigs. Obtenido a través de DFAST*

Como se puede observar, el contig 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 contig 2, con un 37,8%. Por otra parte, los contigs 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 conocíamos sobre la profundidad de las lecturas para realizar el siguiente gráfico.



**Figura 6.** Contenido de guanina y citosina vs cobertura de los contigs.

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

Luego de caracterizar funcionalmente cada uno de los contigs obtenidos en el proceso de ensamblaje, extraer los genes ribosomales 16S y realizar un alineamiento de dichas secuencias contra la base de datos del NCBI, se obtienen los microorganismos posiblemente asociados a cada uno de estos contigs, tal y como se muestra a continuación:

**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>
3	<i>Ochrobactrum sp.</i>	99.38	0.0	<a href="#">KF479631.1</a>
4	<i>Ochrobactrum sp.</i>	99.38	0.0	<a href="#">KF479631.1</a>

\*En el contig 5 no se logró aislar la secuencia del gen ribosomal ARN 16S

Como se observa en la tabla anterior, 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.

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

<b>Categoría</b>	<b>Contig 1</b>	<b>Contig 2</b>	<b>Contig 3</b>	<b>Contig 4</b>	<b>Contig 5</b>
Cofactores, Vitaminas, Grupos prostéticos, Pigmentos	226	131	18	110	1
Pared celular y cápside	44	52	2	24	0
Virulencia, Patogenicida, Defensa	68	42	3	35	2
Metabolismo de Potasio	13	5	4	3	0
Fotosíntesis	0	0	0	0	0
Miscellaneous	48	8	4	14	1
Fagos, Profagos, Elementos transponibles, Plásmidos	6	5	2	5	0
Transporte de membrana	119	11	15	31	6
Adquisición y Metabolismo de Hierro	42	19	18	0	7
Metabolismo o de ARN	55	57	4	26	2
Nucleosidos y Nucleótidos	109	101	11	63	5
Metabolismo de proteínas	225	142	24	147	9
División y Ciclo Celular	0	3	0	0	0
Motilidad y quemotaxis	72	0	12	0	1
Regulación and Señalización celular	70	12	1	10	5
Metabolismo Secundario	4	5	0	4	0
Ácidos grasos, lípidos, isoprenoides	151	50	32	25	13
Metabolismo de Nitrógeno	17	5	3	0	16
Dormancia y Esporulación	1	32	0	1	0
Respiración	117	45	12	78	18
Respuesta a Estrés	108	37	7	44	0
Metabolismo de Compuestos Aromáticos	126	5	20	7	9
Aminoácidos y derivados	531	249	90	160	38
Metabolismo del azufre	33	14	0	3	2
Metabolismo del fósforo	41	32	5	20	2
Carbohidratos	299	198	59	108	41

De acuerdo con los resultados obtenidos en la anotación funcional usando la plataforma RAST de cada uno de los contigs ensamblados, el contig 1 está relacionado a aminoácidos y derivados y presenta el mayor número de estos, seguido por la categoría de cofactores, vitaminas, grupos prostéticos y pigmentos, para el contig 2 en una preponderancia relacionada al contig 1 pero con variaciones en metabolismo de proteínas. Para el contig 3, la consideración de mayor valoración corresponde a aminoácidos y derivados, carbohidratos, junto a ácidos grasos. En el contig 5, carbohidratos es la mayor de las valoraciones.

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

<b>Categoría</b>	<b>Contig 1</b>	<b>Contig 2</b>	<b>Contig 3</b>	<b>Contig 4</b>	<b>Contig 5</b>
Carbohydrate binding modules (CBM)	9	10	0	4	0
Glycosyl hydrolases (GH)	46	90	7	23	0
Glycosyl Transferases (GT)	80	29	13	27	4
GH associated with GT	1	0	1	1	0
CBM associated with GH	6	26	0	0	0
Auxiliary activities associated with CBM	2	1	0	0	0
Auxiliary activities (AA)	6	2	3	4	1
Carbohydrate esterases (CE)	6	9	0	5	3
Polysaccharide Lyase (PL)	3	1	0	1	0

En la tabla anterior se observa como el contig 1, presenta una mayor cantidad de anotaciones a enzimas relacionadas con glicosil transferasas y glicosil hidrolasas, potencial similar al del contig 2 pero se puede observar una diferencia notable en la distribución de dichas anotaciones; por ejemplo el contig 2 presenta casi el doble de anotaciones de glicosil hidrolasas y por su parte, el contig 1 posee más del doble de anotaciones relacionadas con glicosil transferasas. También se puede observar que el único contig con sin presencia de esterases de carbohidratos, y de la misma manera, junto con el contig 4 y 5 muestran ausencia de anotaciones a enzimas relacionadas con actividades auxiliares asociadas con CBM (coal-bed methane).

Luego, el contig que posee la mayor cantidad de anotaciones es el contig 2 con 168 anotaciones, seguido por el contig 1 con 159. De la misma manera, los contigs con menor cantidad de anotaciones son, el contig 5 con 8 anotaciones, seguido por el contig 3 con 24 anotaciones y finalizando con el contig 4 que posee 65 anotaciones.

**Tabla 6.** Potencial actividad degradadora de plásticos de los microorganismos más abundantes presentes en el consorcio mínimo efectivo degradador de lignocelulosa

Microorganismo	GenbankID	Nombre del gen	Nombre de la proteína	Plástico capaz de degradar
<b><i>Paenibacillus</i></b>				
<i>Paenibacillus</i> <i>sp. DK1</i>	MK045309	<i>alkB</i>	Alkane monooxygenase AlkB	PE
<i>Paenibacillus</i> <i>amylolyticus</i>	AB093482.1	<i>PlaA</i>	PLA depolymerase	PLA
<b><i>Pseudomonas</i></b>				
<i>Pseudomonas</i> <i>protegens</i>	-	-	Lipasa	PUR
<i>Pseudomonas</i> <i>sp.</i>	AB008494.1	<i>pvaB</i>	Oxidized polyvinyl alcohol hydrolase	PVA
<i>Pseudomonas</i> <i>fluorescens</i>	U10470	<i>phaZ</i>	Poly(3-hydroxyoctanoate) depolymerase	PHO

Tal y como se observa en la tabla anterior, los generos *Pseudomonas* y *Paenibacillus*, cuentan dentro de sus categorías con especies con alto potencial para la degradación de diversos plásticos.

**Tabla 7.** Anotación funcional con CAZyemzymes database

<b>Categoría</b>	<b>Contig 1</b>	<b>Contig 2</b>	<b>Contig 3</b>	<b>Contig 4</b>	<b>Contig 5</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, GH73 GH28, 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	GH23, GH1 GH24, GH2 GH140 GH103	GH26, GH25, GH6 GH28, GH36, GH3 GH65, GH0, GH19, GH23, GH4 GH32, GH51 GH102, GH103 GH13_11	-
Glycosyl Transferases (GT)	GT4, GT13, GT2 GT87, GT1, GT51	GT1, GT26, GT28 GT2, GT4, GT5 GT51	GT20, GT25 GT26, GT4 GT30, GT51 GT2	GT13, GT19, GT2 GT25, GT32, GT4 GT28, GT51, GT83	GT4 GT28 GT107
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	AA3, AA4	AA1, AA3_2	AA7
Polysaccharide Lyase	PL5_1, PL7_1, PL7_2	PL9_2	-	PL0	

Al ejecutar 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, se observa el número de reacciones y genes asociadas a cada uno de los contigs y el valor de la función de crecimiento máxima de biomasa, tal y como se muestra a continuación:

**Tabla 8.** *Función valor biomasa*

ID del modelo	Especie	Reacciones	Genes	Valor función biomasa (mmol/gDW/h)
Contig_1	<i>Pseudomonas</i>	1458	1082	165.56
Contig_2	<i>Paenibacillus</i>	1250	769	3.60
Contig_3	<i>Ochrobactrum</i>	730	248	20.49
Contig_4	<i>Ochrobactrum</i>	1074	556	70.39
Contig_5	<i>Rhizobium</i>	583	157	23.91

**Tabla 9.** *Sub- Ensamblaje de las lecturas no mapeadas a los contig existentes, utilizando el ensamblador FLYE*

Nombre de la secuencia	Longitud (pb)	Cobertura	Circular
Contig_1	6984708	1	No
Contig_2	5268610	1	No
Contig_10	1369245	1	No
Contig_5	1342506	2	No
Contig_12	993075	2	Sí
Contig_7	929018	1	No
Contig_3	402284	2	No
Contig_21	243118	2	No
Contig_19	236896	1	No
Contig_20	211193	1	No
Contig_4	94178	1	No
Contig_18	6282	2	No
Contig_8	4152	4	No
Contig_17	1006	1	No

## MATERIAL SUPLEMENTARIO:

- Todos los resultados de los ensamblajes se pueden observar en el siguiente hipervínculo: <https://github.com/davidarturolr/BCOM4102/tree/master/Proyecto>

## REFERENCIAS

1. 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). <https://doi.org/10.1128/aem.02427-20>
2. 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. <https://doi.org/10.1038/s41592-020-00971-x>
3. Singhvi, M., & Gokhale, D. (2019). Lignocellulosic biomass: Hurdles and challenges in its valorization. *Applied Microbiology And Biotechnology*, 103(23-24), 9305-9320. <https://doi.org/10.1007/s00253-019-10212-7>
4. Lopes, A., Ferreira Filho, E., & Moreira, L. (2018). An update on enzymatic cocktails for lignocellulose breakdown. *Journal Of Applied Microbiology*, 125(3), 632-645. <https://doi.org/10.1111/jam.13923>
5. Zhu, N., Yang, J., Ji, L., Liu, J., Yang, Y., & Yuan, H. (2016). Metagenomic and metaproteomic analyses of a corn stover-adapted microbial consortium EMSD5 reveal its taxonomic and enzymatic basis for degrading lignocellulose. *Biotechnology For Biofuels*, 9(1). <https://doi.org/10.1186/s13068-016-0658-z>
6. Koren, S., Walenz, B., Berlin, K., Miller, J., Bergman, N., & Phillippy, A. (2017). Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. *Genome Research*, 27(5), 722-736. <https://doi.org/10.1101/gr.215087.116>
7. Dong, X., & Strous, M. (2019). An Integrated Pipeline for Annotation and Visualization of Metagenomic Contigs. *Frontiers In Genetics*, 10. <https://doi.org/10.3389/fgene.2019.00999>
8. Tamames, J., Cobo-Simón, M., & Puente-Sánchez, F. (2019). Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes. *BMC Genomics*, 20(1). <https://doi.org/10.1186/s12864-019-6289-6>.
9. Shah, N., Tang, H., Doak, T., & Ye, Y. (2010). Comparing bacterial communities inferred from 16s rrna gene sequencing and shotgun metagenomics. *Biocomputing 2011*, 165-176. [https://doi.org/10.1142/9789814335058\\_0018](https://doi.org/10.1142/9789814335058_0018)
10. Jiménez, D., Chaves-Moreno, D., & van Elsas, J. (2015). Unveiling the metabolic potential of two soil-derived microbial consortia selected on wheat straw. *Scientific Reports*, 5(1). <https://doi.org/10.1038/srep13845>
11. 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. [https://doi: 10.22490/21456453.2040](https://doi.org/10.22490/21456453.2040)

12. 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. [https://doi: 10.1016/j.rser.2019.109426](https://doi.org/10.1016/j.rser.2019.109426)
13. 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. [https://doi: 10.1016/j.cej.2020.126233](https://doi.org/10.1016/j.cej.2020.126233)
14. Mahro, B., & Timm, M. (2007). Potential of Biowaste from the Food Industry as a Biomass Resource. *Engineering In Life Sciences*, 7(5), 457-468. [https://doi: 10.1002/elsc.200620206](https://doi.org/10.1002/elsc.200620206)
15. 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. [https:// doi: 10.1016/j.gfs.2019.100330](https://doi.org/10.1016/j.gfs.2019.100330)
16. 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. [https:// doi: 10.1016/j.jobab.2020.07.002](https://doi.org/10.1016/j.jobab.2020.07.002)
17. Cerda, A., Artola, A., Barrena, R., Font, X., Gea, T., & Sánchez, A. (2019). Innovative Production of Bioproducts From Organic Waste Through Solid-State Fermentation. *Frontiers In Sustainable Food Systems*, 3. [https://doi:10.3389/fsufs.2019.00063](https://doi.org/10.3389/fsufs.2019.00063)
18. 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. [https://doi: 10.1016/j.watres.2017.04.068](https://doi.org/10.1016/j.watres.2017.04.068)
19. 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. [https://doi: 10.1039/d0re00241k](https://doi.org/10.1039/d0re00241k)
20. 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). [https://doi: 10.1186/s12934-019-1275-x](https://doi.org/10.1186/s12934-019-1275-x)
21. 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. [https://doi: 10.1016/j.btre.2016.03.008](https://doi.org/10.1016/j.btre.2016.03.008)

22. Li, P., Anumanthan, A., Gao, X., Ilangoan, K., Suzara, V., Düzgüneş, N., & Renugopalakrishnan, V. (2007). Expression of Recombinant Proteins in *Pichia Pastoris*. *Applied Biochemistry And Biotechnology*, 142(2), 105-124. <https://doi:10.1007/s12010-007-0003-x>
23. Diethard, M., Gasser, B., Egermeier, M., Marx, H., & Sauer, M. (2016). Industrial Microorganisms: *Saccharomyces cerevisiae* and other Yeasts. *Industrial Biotechnology*, 673-686. <https://doi: 10.1002/9783527807796.ch18>
24. Pandey, S., Singh, N., Yadav, T., Bansal, A., Thanki, A., Yadav, M., & Nayak, J. (2020). Metabolic Products of Mixed Culture Fermentation. *Engineering Of Microbial Biosynthetic Pathways*, 75-92. [https://doi: 10.1007/978-981-15-2604-6\\_5](https://doi: 10.1007/978-981-15-2604-6_5)
25. 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). <https://doi: 10.1186/s13068-014-0180-0>
26. Jia, Y., Ng, S., Lu, H., Cai, M., & Lee, P. (2018). Genome-centric metatranscriptomes and ecological roles of the active microbial populations during cellulosic biomass anaerobic digestion. *Biotechnology For Biofuels*, 11(1). <https://doi: 10.1186/s13068-018-1121-0>
27. Alcalde M, Ferrer M, Plou FJ, Ballesteros A. Environmental biocatalysis: from remediation with enzymes to novel green processes. *Trends Biotechnol.* 2006 Jun;24(6):281-7. <https://doi: 10.1016/j.tibtech.2006.04.002>.
28. Green BD, Keller M. Capturing the uncultivated majority. *Curr Opin Biotechnol.* 2006 Jun;17(3):236-40. <https://doi: 10.1016/j.copbio.2006.05.004>.
29. Tyson GW, Banfield JF. Cultivating the uncultivated: a community genomics perspective. *Trends Microbiol.* 2005 Sep;13(9):411-5. <https://doi: 10.1016/j.tim.2005.07.003>.
30. Lawson, C., Harcombe, W., Hatzenpichler, 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
31. 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.
32. Kang, D., Jacquiod, 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

33. Bohra, V., Dafale, N., Hathi, Z., & Purohit, H. (2019). Genomic annotation and validation of bacterial consortium NDMC-1 for enhanced degradation of sugarcane bagasse. *Annals Of Microbiology*, 69(7), 695-711. doi: 10.1007/s13213-019-01462-x
34. Song, H., Ding, M., Jia, X., Ma, Q., & Yuan, Y. (2014). Synthetic microbial consortia: from systematic analysis to construction and applications. *Chem. Soc. Rev.*, 43(20), 6954-6981. doi: 10.1039/c4cs00114a
35. 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
36. Sims D, Sudbery I, Illott NE, Heger A, Ponting CP. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*. doi: 10.1038/nrg3642.
37. Zhang, W., Jia, B. & Wei, C. (2019). PaSS: a sequencing simulator for PacBio sequencing. *BMC Bioinformatics* 20, 352. <https://doi.org/10.1186/s12859-019-2901-7>
38. Ryan R. Wick, Mark B. Schultz, Justin Zobel, Kathryn E. Holt. (2015). Bandage: interactive visualization of *de novo* genome assemblies, *Bioinformatics*, <https://doi.org/10.1>