



UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ



FACULTAD DE CIENCIAS QUÍMICAS

POSGRADO EN CIENCIAS QUÍMICAS

Título del trabajo

Caracterización génica de *Chlorella vulgaris* para la biorremediación de aguas

Tesis que para obtener el grado de:

Maestra en Ciencias Químicas

Presenta:

Gómez de la Torre Alma Edith

Director de tesis:

Dra. Ruth Elena Soria Guerra

Codirector de Tesis:

Dr. Mishael Sánchez Pérez



REPOSITORIO INSTITUCIONAL



UASLP-Sistema de Bibliotecas
Repositorio Institucional Tesis digitales Restricciones de uso

DERECHOS RESERVADOS

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en este Trabajo Terminal está protegido por la Ley Federal de Derecho de Autor (LFDA) de los Estados Unidos Mexicanos.

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde se obtuvo, mencionando el autor o autores. Cualquier uso distinto o con fines de lucro, reproducción, edición o modificación será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



Caracterización génica de *Chlorella vulgaris* para la biorremediación de aguas © 2026 por Gómez de la Torre Alma Edith se distribuye bajo una licencia Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

Declaraciones Institucionales

Este proyecto se realizó en Laboratorio de Biotecnología de Células Vegetales y Bioinformática, adscrito a la Facultad de Ciencias Químicas de la Universidad Autónoma de San Luis Potosí, en el periodo comprendido entre enero 2024 y diciembre 2025, bajo la dirección de la Dra. Ruth Elena Soria Guerra y fue apoyado por el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT).

El programa de Maestría en Ciencias Químicas de la Universidad Autónoma de San Luis Potosí pertenece al Sistema Nacional de Posgrados (SNP) del CONAHCYT, registro 000519.

Número de la beca otorgada por CONAHCYT: **4017257**

Número de CVU: **1343823**

Los datos del trabajo titulado "Caracterización génica de *Chlorella vulgaris* para la biorremediación de aguas" se encuentran bajo el resguardo de la Facultad de Ciencias Químicas y pertenecen a la Universidad Autónoma de San Luis Potosí.

Aprobación de Tema de Tesis y Comité Tutorial

San Luis Potosí SLP a 22 de abril de 2024

Comité Académico

La presente es para que quede asentado que el tema de Tesis de maestría:

“Caracterización génica de *Chlorella vulgaris* para la biorremediación de aguas”

de la estudiante: Alma Edith Gómez de la Torre, que se llevará a cabo en el laboratorio de Biotecnología Molecular de la Facultad de Ciencias Químicas, es APROBADO.

Y su comité tutorial queda conformado por:

Directora: Dra. Ruth Elena Soria Guerra

Co director: Dr. Mishael Sánchez Pérez

Asesor: Dr. Alejandro Hernández Morales

Sin más por el momento, quedo de Uds.

ATENTAMENTE

Dra. Denisse Atenea de Loera Carrera
Coordinador del Posgrado



UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ



**FACULTAD DE CIENCIAS QUÍMICAS
POSGRADO EN CIENCIAS QUÍMICAS**

Título del trabajo

Caracterización génica de *Chlorella vulgaris* para la biorremediación de aguas

Tesis que para obtener el grado de:

Maestra en Ciencias Químicas

Presenta:

Gómez de la Torre Alma Edith

Director de tesis:

Dra. Ruth Elena Soria Guerra

Codirector de Tesis:

Dr. Mishael Sánchez Pérez

SINODALES:

Dra. Ruth Elena Soria Guerra _____

Dr. Mishael Sánchez Pérez _____

Dr. Alejandro Hernández Morales _____

Dra. Luzmila Martínez González _____

INTEGRANTES DEL COMITÉ TUTORIAL ACADÉMICO

Dra. Ruth Elena Soria Guerra: Director de tesis. Adscrita al Posgrado en Ciencias Químicas de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P.

Dr. Mishael Sánchez Pérez: Sinodal de tesis. Adscrito a la Universidad Autónoma del Estado de Morelos, Facultad de Ciencias del Deporte.

Dr. Alejandro Hernández Morales: Sinodal de tesis. Adscrito al Posgrado en Ciencias Químicas de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P.

Dra. Luzmila Martínez González: Sinodal de tesis. Asociada al Posgrado en Ciencias en Bioprocesos de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P.

Carta Cesión de Derechos

San Luis Potosí SLP a 8 de diciembre 2025

En la ciudad de San Luis Potosí el día *19* del mes de *noviembre* del año *2025*. El que suscribe *Alma Edith Gómez de la Torre* alumna del programa de posgrado *Maestría en Ciencias Químicas* adscrito a Facultad de Ciencias Químicas manifiesta que es autora intelectual del presente trabajo terminal, realizado bajo la dirección de: Dra. Ruth Elena Soria Guerra y Dr. Mishael Sánchez Pérez y cede los derechos del trabajo titulado “Caracterización génica de *Chlorella vulgaris* para la biorremediación de aguas” a la **Universidad Autónoma de San Luis Potosí**, para su difusión con fines académicos y de investigación.

Los usuarios de la información no deben reproducir de forma total o parcial texto, gráficas, imágenes o cualquier contenido del trabajo si el permiso expreso del o los autores. Éste, puede ser obtenido directamente con el autor o autores escribiendo a la siguiente dirección ruth.soria@uaslp.mx. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Alma Edith Gómez de la Torre

Carta de Análisis de Similitud

San Luis Potosí, S.L.P.a 20 de abril del 2026

L.B. Reyna Nayeli Ortiz Quintero
Biblioteca de Posgrado FCQ

Asunto: Reporte de porcentaje de similitud de tesis de grado

Por este medio me permito informarle el porcentaje de similitud obtenido mediante Ithenticate para la tesis titulada “Caracterización génica de Chlorella vulgaris para la biorremediación de aguas” presentada por la autora Gómez de la Torre Alma Edith. La tesis es requisito para obtener el grado de Maestría en el Posgrado en Ciencias Químicas. El análisis reveló un porcentaje de similitud de 17% excluyendo referencias y metodología.

Agradezco sinceramente su valioso tiempo y dedicación para llevar a cabo una exhaustiva revisión de la tesis. Quedo a su disposición para cualquier consulta o inquietud que pueda surgir en el proceso.

Sin más por el momento, le envío un cordial saludo.

ATENTAMENTE:

Dr. Fidel Martínez Gutiérrez
Coordinador Académico del Posgrado
en Ciencias Químicas

Agradecimientos:

Este trabajo representa el esfuerzo de una etapa importante en mi formación académica, la cual no habría sido posible sin el apoyo de personas importantes a quienes deseo agradecer.

Agradezco profundamente a mi Directora de tesis, a la Dra. Ruth Elena Soria Guerra, por su guía, confianza y apoyo durante el desarrollo de este trabajo, así como por compartir sus conocimientos y experiencia, y por su acompañamiento académico a lo largo de esta etapa de mi formación.

También quiero agradecer a mi coasesor, el Dr. Mishael Sánchez Pérez, por su valiosa orientación, disposición y apoyo constante a lo largo de este proyecto. Su acompañamiento fue fundamental para el desarrollo de este trabajo, al brindarme las herramientas y conocimientos necesarios en el área de bioinformática, así como por su disposición para compartir su experiencia y resolver dudas en cada etapa del proceso.

A mi comité evaluador, por sus valiosos comentarios y sugerencias que contribuyeron a mejorar esta investigación.

Expreso un profundo agradecimiento a mi familia, cuyo apoyo y amor incondicional han sido fundamentales a lo largo de este proceso. Su acompañamiento constante ha hecho más llevaderos los desafíos de este camino, ya que alcanzar esta meta no ha sido sencillo. Les expreso mi más sincero afecto y gratitud.

A mis compañeros y amigos, por su apoyo, motivación y compañía durante este proceso.

Por último, quiero expresar mi agradecimiento a CONAHCYT por la beca 4017257 otorgada para la realización de este proyecto.

Resumen

La biorremediación mediante el uso de microalgas representa una alternativa sostenible para el tratamiento de aguas contaminadas debido a su capacidad de asimilación de nitratos, fosfatos, amonio, materia orgánica e incluso metales pesados. En este contexto, el presente estudio tiene como objetivo caracterizar genéticamente a *Chlorella vulgaris* mediante datos de secuenciación masiva y anotación funcional del genoma para la selección de genes relacionados a la biorremediación. Se realizó un ensamblaje de *novo* y análisis funcional utilizando diversas herramientas bioinformáticas, incluyendo anotación con eggNOG-mapper y mapeo en KEGG (Kyoto Encyclopedia of Genes and Genomes).

De acuerdo con los resultados, se obtuvo un ensamblaje a nivel contigs con un tamaño total de 143.33 Mb utilizando la herramienta Unicycler. Asimismo el análisis funcional reveló una alta proporción de proteínas hipotéticas, lo que sugiere la presencia de múltiples secuencias aún no caracterizadas. En contraste con lo anterior, la evaluación mediante COG (Clusters of Orthologous Groups) mostró una prevalencia de genes con función desconocida, junto con categorías relacionadas con metabolismo, transcripción y señalización, lo que refleja un genoma altamente regulado con la capacidad de adaptarse a diferentes condiciones ambientales. De manera complementaria, se identificaron genes asociados a la asimilación de metales pesados, fármacos y compuestos inorgánicos (como nitrógeno y fósforo), dentro de los que se incluyen transportadores, enzimas y proteínas asociadas al estrés ambiental. Finalmente, el análisis de rutas metabólicas confirmó la capacidad de *C. vulgaris* para transformar diversos contaminantes, evidenciando su potencial en aplicaciones de procesos biorremediación.

Palabras clave: *Chlorella vulgaris*, Biorremediación, ensamblaje de *novo*, Anotación funcional, Rutas metabólicas

Abstract

Bioremediation using microalgae represents a sustainable alternative for the treatment of contaminated water due to their ability to assimilate nitrates, phosphates, ammonium, organic matter, and even heavy metals. In this context, the present study aims to genetically characterize *Chlorella vulgaris* using high-throughput sequencing data and functional genome annotation for the identification of genes related to bioremediation. A de novo assembly and functional analysis were performed using various bioinformatics tools, including annotation with eggNOG-mapper and mapping in KEGG (Kyoto Encyclopedia of Genes and Genomes).

According to the results, a contig-level assembly with a total size of 143.33 Mb was obtained using the Unicycler tool. Likewise, the functional analysis revealed a high proportion of hypothetical proteins, suggesting the presence of multiple sequences that are still not functionally characterized. In contrast, the evaluation using COG (Clusters of Orthologous Groups) showed a prevalence of genes with unknown function, along with categories related to metabolism, transcription, and signaling, reflecting a highly regulated genome capable of adapting to different environmental conditions. Additionally, genes associated with the assimilation of heavy metals, pharmaceuticals, and inorganic compounds (such as nitrogen and phosphorus) were identified, including transporters, enzymes, and proteins related to environmental stress. Finally, the analysis of metabolic pathways confirmed the ability of *C. vulgaris* to transform various contaminants, highlighting its potential in bioremediation processes.

Keywords: *Chlorella vulgaris*, Bioremediation, De novo assembly, Functional annotation, Metabolic pathways

ÍNDICE

1. INTRODUCCIÓN.....	1
2. ANTECEDENTES	2
2.1. Etapas del tratamiento de Aguas Residuales	2
2.2. Biorremediación	4
2.3. Microalgas	5
2.4. <i>Chlorella vulgaris</i>	6
2.5. Metabolismo del nitrógeno.....	9
2.6. Metabolismo del fósforo.....	10
2.7. Metabolismo de metales pesados.....	11
2.8. Metabolismo de Fármacos	14
2.9. Caracterización genética	19
2.10. Secuenciación masiva	20
2.11. Secuenciación por síntesis (illumina).....	21
2.12. Ensamblaje del genoma	22
2.13. Ensamblaje de novo	23
2.14. Control de Calidad	25
2.15. Anotación genómica	26
2. HIPÓTESIS.....	30
3. JUSTIFICACIÓN	30
4. OBJETIVO	30
5. OBJETIVOS ESPECÍFICOS.....	30
6. MATERIALES Y MÉTODOS.....	31
6.1. Cultivo de la microalga y extracción de ADN.....	31
6.2. Comprobación de la cantidad y calidad del ADN genómico	31
6.3. Secuenciación masiva de la microalga <i>C. vulgaris</i>	31
6.4. Ensamble del genoma	32
6.5. Anotación del genoma, anotación de ontología génica (GO) clasificación en categorías, componentes celulares (CC), función molecular (FM), y procesos biológicos (PC).....	32

6.6. Obtención de rutas metabólicas asociadas a la degradación de contaminantes	33
7. RESULTADOS Y DISCUSIÓN	34
7.1 Comprobación de la cantidad y calidad del ADN genómico	34
7.1.1. Cantidad y calidad del ADN genómico	34
7.1.2. Verificación de calidad	35
7.2 Ensamble del genoma	39
7.3 Anotación y Predicción de Genes	45
7.4 Rutas metabólicas identificadas en <i>C. vulgaris</i>	61
8. CONCLUSIONES.....	68
9. BIBLIOGRAFÍA.....	69
10. ANEXOS.....	77

ÍNDICE DE FIGURAS

Figura 1. Principales etapas del proceso de trato de Aguas residuales.....	3
Figura 2. Mecanismos de remoción de nitrógeno por microalgas en aguas residuales.	9
Figura 3. Metabolismo del fósforo en microalgas.....	11
Figura 4. Mecanismos de interacción metal-microalga.....	12
Figura 5. Procesos implicados en la remoción de fármacos por microalgas.	16
Figura 6. Proceso de secuenciación Illumina.....	22
Figura 7. Flujo de trabajo general del ensamblaje de novo de un genoma completo.	23
Figura 8. (A) K-mers con una longitud de 4 bases a partir de una secuencia. (B) Construcción del gráfico de bruijn a partir de los k-mers. (C) En el enfoque hamiltoniano, los k-mers (o secuencias) son nodos, mientras que son los bordes en el enfoque euleriano.....	25
Figura 9. Flujo de trabajo de anotación del genoma.....	29
Figura 10. Resultados de la extracción de ADN de <i>C. vulgaris</i>	35
Figura 11. Análisis de calidad realizado por el software FastQC para las lecturas crudas del genoma de <i>C. vulgaris</i>	36
Figura 12. Análisis de calidad realizado por el software FastQC para las lecturas filtradas del genoma de <i>C. vulgaris</i>	38
Figura 13. Mapa de calor del porcentaje de identidad promedio (ANI, %) entre genomas de <i>C. vulgaris</i> y microalgas relacionadas.....	42
Figura 14. Mapa de calor del porcentaje de cobertura de alineamiento entre genomas de <i>C. vulgaris</i> y microalgas relacionadas.....	44
Figura 15. Distribución de genes y proteínas anotados en el genoma de <i>C. vulgaris</i> LBCVB.....	46
Figura 16. Número de genes anotados por categoría funcional.....	47
Figura 17. Clasificación funcional de genes por categorías COG.	49
Figura A1. Distribución taxonómica de las secuencias asignadas a partir de las lecturas filtradas por Trimalore.	77

Figura A2. Evaluación de completitud del ensamblaje del genoma de <i>C. vulgaris</i> mediante BUSCO (Benchmarking Universal Single-Copy Orthologs).	78
Figura A3. Ruta de carcinogénesis química asociada a especies reactivas de oxígeno en <i>C. vulgaris</i> LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko05208).	79
Figura A4. Ruta de la fotosíntesis en <i>C. vulgaris</i> LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00195).	80
Figura A5. Ruta de Fijación de carbono mediante el ciclo de Calvin en <i>C. vulgaris</i> LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00710).	81
Figura A6. Ruta de Metabolismo de fármacos - citocromo P450 en <i>C. vulgaris</i> LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00982). ..	82
Figura A7. Ruta del Metabolismo de fármacos: otras enzimas en <i>C. vulgaris</i> LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00983). ..	83
Figura A8. Ruta del metabolismo de transportadores ABC en <i>C. vulgaris</i> LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko02010).	84
Figura A9. Ruta del Metabolismo Nitrógeno <i>C. vulgaris</i> LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00910).	85
Figura A10. Vía de las pentosas fosfato en <i>C. vulgaris</i> LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00030).	86
Figura A11. Ruta de Replicación del ADN en <i>C. vulgaris</i> LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko03030).	87

ÍNDICE DE TABLAS

Tabla 1. Tratamiento de aguas residuales mediante microalgas para remover nutrientes y compuestos orgánicos.	6
Tabla 2. Características genómicas de diversas microalgas.	8
Tabla 3. Asimilación de metales pesados por diferentes especies de microalgas.....	13
Tabla 4. Degradación de fármacos a través del metabolismo de microalgas.	15
Tabla 5. Evaluación de la calidad entre los ensamblajes realizados con ABySS, MEGAHIT, SPAdes y Unicycler mediante la herramienta QUAST.	40
Tabla 6. Genes y proteínas homólogas identificadas mediante BLASTp en microalgas verdes relacionados con metales pesados.....	50
Tabla 7. Genes y proteínas homólogas identificadas mediante BLASTp en microalgas verdes asociados a medicamentos.....	55
Tabla 8. Genes y proteínas homólogas identificadas mediante BLASTp en microalgas verdes vinculadas al transporte de N y P.	59
Tabla 9. Vías metabólicas identificadas en <i>C. vulgaris</i> LBCVB relacionadas con la respuesta a contaminantes en agua.....	62

1. INTRODUCCIÓN

La contaminación del agua se considera una de las problemáticas más apremiantes en la actualidad, debido a la presencia de contaminantes que son difíciles de manipular o eliminar como lo son los metales pesados, fármacos y compuesto inorgánicos (Nitrógeno y Fosforo) (Ahmad et al., 2021).

La biorremediación se fundamenta en el uso de microorganismos para la eliminación o biotransformación de contaminantes en fuente de energía o carbono a partir de aguas residuales con la producción consecuente de biomasa. Entre los microorganismos empleados, las microalgas destacan por su notable capacidad para remover nitratos, fosfatos, amonio y materia orgánica, lo que las convierte en una herramienta eficiente para la eliminación de contaminantes y motiva a la búsqueda de cepas con mayor resistencia a condiciones ambientales extremas (Ortiz-Villota et al., 2018).

La microalga *Chlorella vulgaris* ha demostrado una alta tasa de crecimiento y una rápida adaptación al crecimiento en aguas residuales (Asaad & Amer, 2024). Sin embargo, la optimización mediante ingeniería genética es actualmente un desafío debido principalmente a la falta de información y conocimientos limitados en genómica de microalgas (Wu et al., 2019). De esta manera, la caracterización genética aporta un panorama amplio y detallado de los componentes genéticos que conforman la biología de la microalgas que a su vez contribuyen a descifrar los mecanismos y genes involucrados en el procesos de biorremediación.

2. ANTECEDENTES

Hoy en día los cuerpos de agua en su gran mayoría están sujetos a una presión cada vez mayor debido al incremento en la demanda de este recurso hídrico, ya sea por el crecimiento demográfico, uso agrícola, domestico e industrial, que aunado al uso inadecuado de este recurso se ha generado un gran deterioro en su calidad (Escapa, 2017) (Ahmad et al., 2021). Los efectos de la contaminación del agua exigen varios riesgos tanto para la disponibilidad como para su calidad ya que no solo afecta la salud humana y aspectos socioeconómicos de la población sino también organismos acuáticos, terrestres, así como también al medio ambiente (Ortiz-Villota et al., 2018).

La contaminación del agua se define como la introducción de sustancias o elementos extraños en fuentes de aguas naturales, que alteran su calidad y que la hacen inapropiada tanto para el uso como para su consumo. Siendo los principales contaminantes xenobióticos, microplásticos, metales pesados, así como también una gran cantidad de compuestos de carbono, fosforo y nitrógeno (Ahmad et al., 2021). El tratamiento de aguas residuales presenta una de las problemáticas más apremiantes en la actualidad ya que no resultan adecuados para la completa eliminación de dichos contaminantes. Los sistemas de tratamiento de aguas convencionales consisten en tratamientos físicos, químicos o biológicos según su origen donde hay un pretratamiento, tratamiento primario, secundario o terciario según el orden de actuación sobre el agua residual (Escapa, 2017) (Ahmad et al., 2021).

2.1. Etapas del tratamiento de Aguas Residuales

El pretratamiento consiste en operaciones mecánicas y físicas para la separación de la materia por su naturaleza o por su tamaño que pueden ocasionar problemas en las etapas posteriores como: obstrucción de tuberías, bombas, depósitos de arena, roturas de equipos, etc. Generalmente esta etapa consta de métodos de tamizado y desbaste, que contribuyen a la eliminación de residuos de mediano y gran tamaño, adicionalmente también se eliminan las grasas y arenas (Gurunathan et al., 2017) (Escapa, 2017).

Por otra parte, el tratamiento primario se rige por la eliminación de sólidos en suspensión y coloides mediante procesos como la coagulación-floculación, en el cual las moléculas se aglutinan en partículas más grandes que permiten su posterior decantación, y busca reducir la DBO (demanda biológica de oxígeno) en un 30 a 40% (Escapa, 2017). Por consiguiente, los tratamientos secundarios tienen como objetivo degradar la materia orgánica biodegradable mediante el uso de diferentes sistemas como lodos activados y lechos bacterianos, donde diferentes tipos de bacterias se alimentan de materia orgánica y los nutrientes presentes en el agua (Escapa, 2017). No obstante, la mayoría de las plantas cuentan hasta con una segunda etapa, que generalmente se emplea únicamente para el tratamiento de aguas residuales municipales que no están empleados para manipular residuos industriales y por ende sustancias nocivas. Cabe destacar que dichas sustancias peligrosas son pretratadas en fuentes de generación industrial para reducir su concentración y por consiguiente se trasladan a plantas de tratamiento convencionales (Figura 1) (Gurunathan et al., 2017).

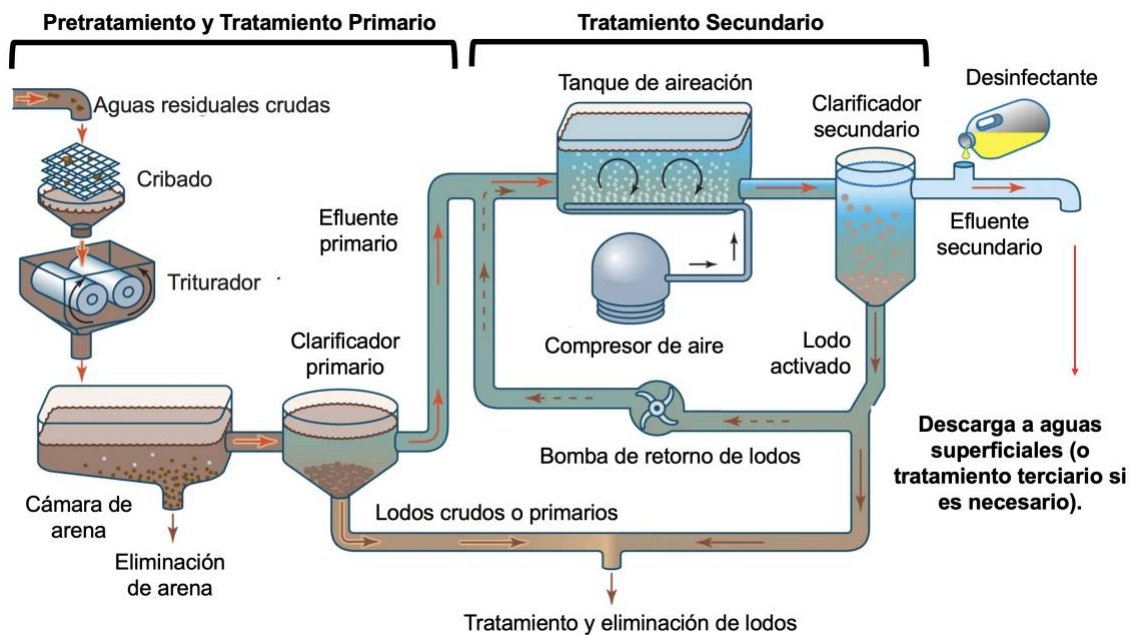


Figura 1. Principales etapas del proceso de trato de Aguas residuales.

A pesar del uso de estos tratamientos, la capacidad de degradación de las plantas convencionales es limitada, especialmente con los nutrientes inorgánicos (nitrógeno y fosforo), metales pesados y contaminantes emergentes que persisten aun en concentraciones elevadas incluso después de la segunda etapa de tratamiento y que eventualmente son liberadas al medio ambiente, siendo esta una de las principales causas de la eutrofización afectando nuestras reservas hídricas como ríos y mares teniendo un impacto directamente en el medio ambiente, ya que estos puede persistir en los ecosistemas un tiempo prolongado así como también acumularse en las cadenas tróficas provocando efectos irreversibles en los seres vivos (Gurunathan et al., 2017) (Ahmad et al., 2021).

En consecuencia, surgen los procesos avanzados que buscan la degradación de compuestos específicos de los efluentes, dichos tratamientos se basan en osmosis inversa, destilación, coagulación adsorción mediate carbono activado, oxidación química, precipitación o nitrificación-desnitrificación, etc. Sin embargo, no es habitual encontrarlos como tratamientos terciarios debido a sus costos elevados y dificultad de operación (Escapa, 2017). Por ende, surge la necesidad de generar diferentes estrategias sostenibles y de bajo costo para la remediación de los efluentes y mejorar la calidad del agua. En este sentido, la biorremediación es una estrategia eficaz para tratar los altos niveles de nutrientes inorgánicos en aguas contaminadas (Ortiz-Villota et al., 2018).

2.2. Biorremediación

La biorremediación se fundamenta como el uso de microorganismos para la eliminación o biotransformación de contaminantes como fuente de energía o carbono a partir de aguas residuales con la producción consecuente de biomasa (Gurunathan et al., 2017) (Ortiz-Villota et al., 2018). Dentro de los microorganismos utilizados en la biorremediación se encuentran las microalgas, las cuales, por su capacidad de remoción significativa de nitratos, fosfatos, amonio y materia orgánica resultan ventajosas para la eliminación de dichos contaminantes (Ortiz-Villota et al., 2018). Además de que la biomasa producida representa una fuente potencial para la

producción de biocombustibles u otros compuestos de valor agregado (Ahmad et al., 2021).

2.3. Microalgas

Las microalgas son organismos fotosintéticos eucariotas y procariontes de estructura simple que crecen rápidamente debido a su carácter unicelular o multicelular (Ortiz-Villota et al., 2018), poseen un tamaño entre 5 y 50 μm y tienen la capacidad de remover materia inorgánica como amonio, nitritos, nitratos y fosfatos (Ramos & Pizarro, 2018), incluso se alimentan de iones metálicos tóxicos (Fe^{2+} , Cu^{2+} , Zn^{2+} , Mn^{2+}) mediante la acumulación, bioadsorción y biotransformación en biomasa (El-Sheekh et al., 2015). Algunas de las principales características de las microalgas incluyen la capacidad de transformar la energía solar y CO_2 en O_2 y biomasa; resistencia a condiciones extremas; tasas de crecimiento altas, incluso con materiales limitados; la capacidad de acumular grandes cantidades de ácidos grasos; además de que posee una notable persistencia a la adaptación, absorbiendo eficientemente fósforo y nitrógeno de aguas residuales, reteniendo el oxígeno disuelto y en consecuencia la eliminación de bacterias y patógenos (Kusuma et al., 2023). Simultáneamente, esta acumulación innata de nutrientes les permite sintetizar una amplia gama de compuestos bioactivos como: pigmentos, carbohidratos, proteínas y lípidos gracias a su particular fisiología y bioquímica, a su vez dichos compuestos bioactivos tiene un potencial significativo en aplicaciones comerciales (Asaad & Amer, 2024).

Los patrones de absorción de nutrientes pueden variar dependiendo de la presencia de nutrientes y de la especie de microalga debido a su diferente fisiología y morfología, ya que han demostrado eficiencia y efectividad variable para la eliminación de nitrógeno y fosforo, así como también de otros nutrientes (Asaad & Amer, 2024). Algunas de las especies más utilizadas en el tratamiento de aguas residuales son *Chlorella* sp., *Scenedesmus* sp. *Desmodesmus* sp. *Nitzschia* sp. *Cosmarium* sp, entre otras (Tabla 1) (Wang et al., 2016).

Tabla 1. Tratamiento de aguas residuales mediante microalgas para remover nutrientes y compuestos orgánicos.

Especie de Microalga	Efluente	Eficiencia total de Remoción
<i>Scenedesmus</i> sp.	Doméstico	La eliminación de N y P osciló entre el 90 y el 99 %
<i>Halochlorella rubescens</i>	Municipal	83.2% (NO ₃ ⁻ -N) y 73.2% (PO ₄ ³⁻)
<i>Chlorella vulgaris</i>	Industria cervecera	91.43% (DBO), 83.11% (DQO), 83.74% (NT), y 54.69% (PT)
<i>Scenedesmus obliquus</i>	Aguas residuales granja porcina	91,43 % (DBO), 83,11 % (DQO), 83,74 % (NT) y 54,69 % (PT)
<i>Chlorella sorokiniana</i> <i>Scenedesmus obliquus</i>	Aguas residuales industriales y municipales + aguas pluviales	63.6% (DQO), 4.2% (NT), y 82.7% (PT)
<i>Scenedesmus obliquus</i>	Aguas residuales municipales estimuladas	99.2% (NH ₄ ⁺), 91.2% (PO ₄), y 83.6% (DQO)
Consortio de <i>Chlorella sorokiniana</i> y lodos activados	Aguas residuales municipales sintéticas	81% (NH ₃), 39% (PO ₄ ³⁻), y 98% (DQO)
<i>Galdieria sulphuraria</i>	Aguas residuales tratamiento primario	6.26 mg L ⁻¹ d ⁻¹ (NH ₃), 1.41 mg L ⁻¹ d ⁻¹ (PO ₄ ³⁻), y 16.4 mg L ⁻¹ d ⁻¹ (DQO)

DBO, Demanda biológica de oxígeno; DQO, Demanda química de oxígeno; NH₃, Amoníaco; NH₄⁺, Amonio; PO₄³⁻, Fosfato; NT, Nitrógeno total; COT, Carbono orgánico total; PT, Fósforo total. Fuente: (Abdelfattah et al., 2022)

2.4. *Chlorella vulgaris*

La cepa *C. vulgaris* es una microalga verde perteneciente al orden de los *Chlorococcales*, de la familia *Oocytaceae*, del género *Chlorella*. Tiene una forma esférica con un tamaño que varía de 2 a 10 µm. Esta microalga unicelular dispone de un genoma de aproximadamente de 37.3 a 48.8 Mb (véase Tabla 2) que consta de 16 cromosomas que varían de 980kb a 4.0 Mb de tamaño (Noutoshi, 1998). Asimismo, su metabolismo puede ser autótrofo, en el cual requiere únicamente compuestos inorgánicos (CO₂, sales y luz), para sintetizar las moléculas necesarias para su crecimiento, o heterótrofo, que toma fuentes de carbono orgánico como azúcares, glicerol y ciertos acetatos o glutamatos. Por otra parte, también puede adoptar un metabolismo mixotrófico, es decir, combinar simultáneamente ambos tipos de fuente de energía para optimizar su crecimiento. En particular, este organismo presenta una organización celular similar a la de las plantas superiores, pues contiene pared celular,

mitocondrias y cloroplastos, siendo estos últimos esenciales para la fotosíntesis. Además de contener clorofila, posee una cantidad significativa de proteínas intracelulares, carbohidratos, lípidos, vitamina C, β -carotenos y vitaminas B (B1, B2, B6 y B12). Por ello, ha sido ampliamente utilizada en la industria alimentaria como suplemento, así como en la cosmética y en el tratamiento de aguas residual (Coronado-Reyes et al., 2020).

C. vulgaris se ha aplicado frecuentemente en el tratamiento de aguas residuales debido a su alta tasa de crecimiento y a su rápida adaptación (Asaad & Amer, 2024). Asaad y Amer (2024) mencionan que *C. vulgaris* puede emplearse eficientemente para el tratamiento de aguas con un tiempo de equilibrio óptimo para la adsorción de iones fosfatos y nitratos de 24h con un pH óptimo de 7 obteniendo una relación de biomasa y concentración de aniones del 80%. Asimismo, El-Sheekh & cols. (2016) reportan que *C. vulgaris* y *C. salina* han demostrado ser altamente eficientes en la reducción de nitratos, amonio, fosfatos, sulfatos, calcio, magnesio incluso metales pesados (Zn, Cu, Mn, Ni, Co, Fe y Cr) alcanzando una eficiencia de eliminación de estos últimos del 13.6 al 100%.

Tabla 2. Características genómicas de diversas microalgas.

	Tecnología de secuenciación	Tamaño del genoma (Mb)	Nivel	N50 (Mb)	GC (%)	Número de genes
<i>Chlorella vulgaris</i> (211/11P)¹	PacBio RSII; Illumina HiSeq	40.2	Scaffold	1.8	62	10,809
<i>Chlorella vulgaris</i> (PKVL7422)²	PacBio Sequel IIe	40.7	Contig	1	61.5	6,293
<i>Chlorella sorokiniana</i> (SLA-04)²	Oxford Nanopore MinION	38.6	Genoma Completo	3	66	-
<i>Chlorella variabilis</i> (NC64A)²	Sanger	46.2	Scaffold	0.0279	67	9,780
<i>Chlorella ohadii</i> (1)²	Illumina HiSeq; PacBio RS	57.1	Scaffold	0.0317	64	10,866
<i>Chlamydomonas reinhardtii</i> (CC-503 cw92 mt+)²	ABI 3739	111.1	Cromosoma	7.8	64	17,742
<i>Ostreococcus tauri</i> (RCC 4221)²	Illumina GAIIx	12.9	Cromosoma	0.7702	89	7.833
<i>Scenedesmus quadricauda</i> (LWG 002611)²	Ion Proton	65.4	Contig	0.0081	63	-

¹Cecchin et al., 2019

²National Center for Biotechnology Information [NCBI], n.d.

2.5. Metabolismo del nitrógeno

El nitrógeno (N) es un macroelemento esencial y de señalización para las microalgas que puede incorporarse tanto en forma orgánica como inorgánica siendo el orden de asimilación de este $\text{NH}_4^+ > \text{NO}_3^- > \text{NO}_2^-$ (amonio, nitrato y nitrito respectivamente) ya que se prefiere la forma más reducida de nitrógeno entre las fuentes de N inorgánico debido a que requiere un menor costo energético (Sanz-Luque et al., 2015) (Su, 2020). Sin embargo, el dinitrógeno su forma más abundante en la atmósfera o disuelta en agua solo puede asimilarse mediante una relación simbiótica con bacterias fijadoras. En las microalgas la asimilación de nitrato se realiza en dos pasos, uno de transporte y otros de reducción (Sanz-Luque et al., 2015).

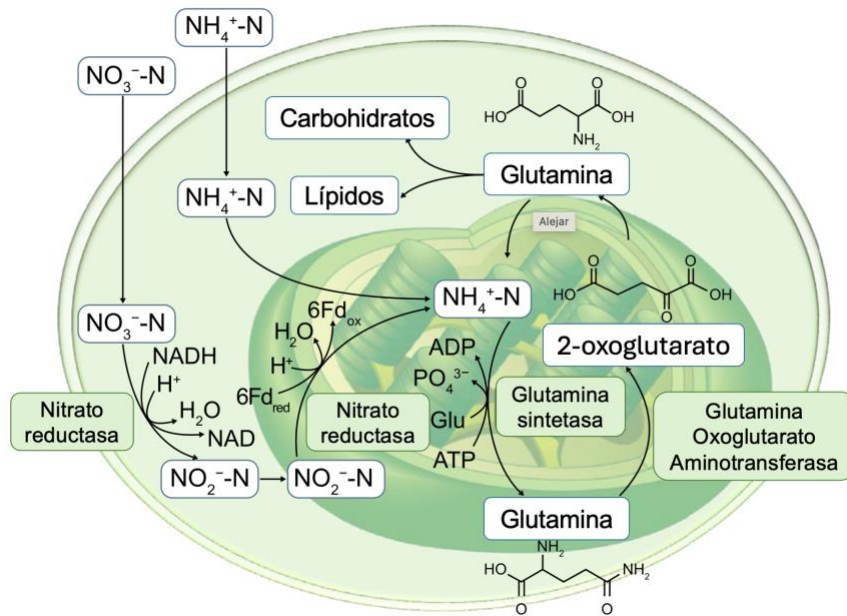


Figura 2. Mecanismos de remoción de nitrógeno por microalgas en aguas residuales (Abdelfattah et al., 2022).

En la primera etapa el nitrato se transporta al interior de la célula, luego una nitrato reductasa (NR) citosólica que cataliza la reducción del nitrato a nitrito que posteriormente se transporta al cloroplasto donde la enzima nitrito reductasa (NiR) cataliza su reducción a amonio (Sanz-Luque et al., 2015). La NR requiere Hierro y Molibdeno, y su actividad está estrechamente relacionada con factores abióticos como iluminación, la disponibilidad de N y la fuente de C. Por último, el amonio se incorpora

a los esqueletos carbonados originando glutamato, mediante el ciclo glutamina sintetasa/ glutamina oxoglutarato aminotransferasa o glutamato sintetasa (GS/GOGAT). Donde primero se incorpora amonio como grupo amida de la glutamina en una reacción que involucra glutamato y ATP catalizados por GS después el grupo amida se trasfiere de la forma reducida a α -oxoglutarato para originar dos moléculas de glutamato (Figura 2). Se ha reportado la interacción entre diferentes fuentes de nitrógeno que causan interferencia en la adsorción de N, aunque no hay una inhibición completa la adsorción de NO_3 puede reducirse parcialmente por la concentración ambiental de NH_4 , adicionalmente otros factores como la luz aumenta la inhibición de NO_3 por NH_4 (Su, 2020).

2.6. Metabolismo del fósforo

El fósforo (P) es un elemento fundamental en el crecimiento y metabolismo de las células eucariotas fotosintéticas ya que participa en varios procesos de vital importancia como en la síntesis de ácidos nucleicos (ADN y ARN), expresión genética, como ATP en el requerimiento energético en procesos metabólicos, así como también formando parte de los fosfolípidos, que es el principal componente de las membranas celulares. La asimilación del fósforo inorgánico (Pi) (PO_4^{3-} , HPO_4^{2-} y H_2PO_4^-) por las microalgas está dada principalmente por pH, ya que cuando estos poseen una carga más baja generalmente están más biodisponibles que los que tienen una carga mayor. Existen múltiples transportadores involucrados en el transporte Pi, que están fuertemente regulados por la disponibilidad de este, dado que existe una menor distribución de transportadores de alta afinidad de Pi con condiciones escasas de P y de forma contraria, una mayor disponibilidad de transportadores de Pi pero de abaja afinidad, se acumulan en condiciones ricas de P (Figura 3). Asimismo, la concentración de otros nutrientes como el N, e incluso la intensidad luminosa, también influyen en la tasa de captación máxima de Pi (Su, 2020).

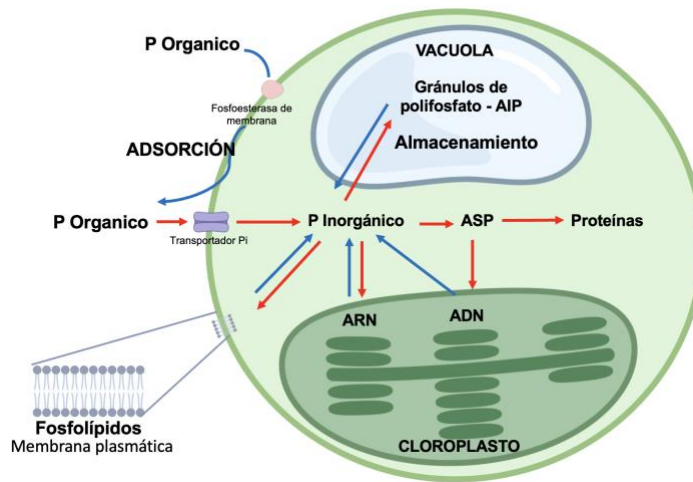


Figura 3. Metabolismo del fósforo en microalgas.

La flecha roja indica las condiciones de abundancia de Pi, mientras que la flecha azul señala las condiciones de deficiencia de Pi. ASP: polifosfato ácido soluble; AIP: polifosfato ácido insoluble (Bossia et al., 2024).

2.7. Metabolismo de metales pesados

El término metales pesados hace referencia a metales con una densidad superior a 5g/m^3 , dentro de los que se incluyen el Arsénico (As), Cadmio (Cd), Cromo (Cr), Cobre (Cu), Hierro (Fe), Plomo (Pb), Mercurio (Hg), Plata (Ag), Zinc (Zn) entre otros, los cuales son altamente utilizados en actividades humanas y que posteriormente son liberadas al medio ambiente a través de aguas residuales u otras vías. Dichos metales son de carácter no biodegradable por lo que tienden a persistir en el ambiente provocando una bioacumulación en las cadenas tróficas causando graves problemas ambientales y de salud para el ser humano. Dichos contaminantes, incluso en concentraciones muy bajas, son altamente tóxicos y cancerígeno (Leong & Chang, 2020).

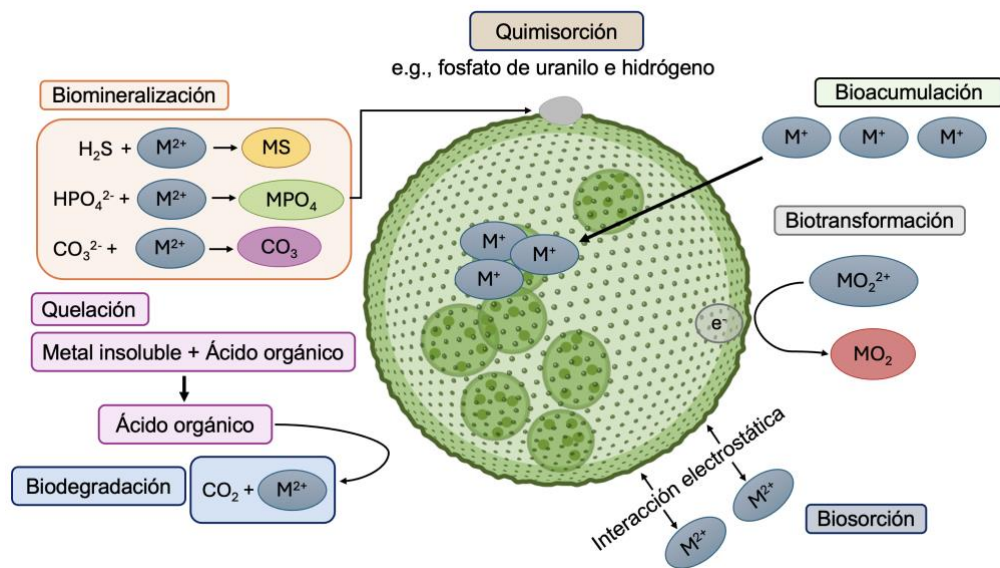


Figura 4. Mecanismos de interacción metal-microalga (Abdelfattah et al., 2022).

Las microalgas pueden absorber metales pesados como Boro (B), Cobalto (Co), Molibdeno (Mo), Magnesio (Mn) y Zinc (Zn) como oligoelementos para llevar a cabo procesos enzimáticos y metabólicos (véase Tabla 3), sin embargo, algunos metales pesados (As, Cd, Cr, Pb y Hg) pueden ser tóxicos para las microalgas. Por lo que generalmente diversas especies de microalgas tienen estrategias para contrarrestar la toxicidad de los metales pesados, como la inmovilización, exclusión y quelación, así como también antioxidantes o enzimas que reducen los metales pesados mediante reacciones redox (Tabla 4) (Gómez-Jacinto et al., 2015) (Leong & Chang, 2020). Debido a que la pared celular de las microalgas está compuesta principalmente de polisacáridos (celulosa y alginato), lípidos y proteínas, que proporcionan una gran cantidad de grupos funcionales (amino, carboxilo, hidroxilo, imidazol, fosfato, sulfonato, tiol entre otros) capaces de unirse a los metales pesados. Las microalgas pueden formar complejos organometálicos en el interior de vacuolas que contribuye a la regulación de iones metálicos en el citoplasma, lo que mitiga los efectos tóxicos de los metales (Leong & Chang, 2020). Adicionalmente, los metales pesados activan la biosíntesis de fitoquelatinas (proteínas/péptidos ricos en tioles) que reducen el estrés causado por estos (Gómez-Jacinto et al., 2015).

Tabla 3. Asimilación de metales pesados por diferentes especies de microalgas.

Cepa de microalga	Conc. Inicial del metal (mg/L)	Max. Sorción (mg/g)	Eficiencia de remoción (%)
As			
<i>Maugeotia genuflexa</i>	10	2.4	96
<i>Clamydomonas reinhardtii</i>	12	4.63	38.6
<i>Chlorella vulgaris</i>	12	3.89	32.4
<i>Scenedesmus almeriensis</i>	12	5.0	41.7
<i>Ulothrix cylindricum</i>	10	2.45	98
Cd			
<i>Parachlorella</i> sp.	100	96.2	--
<i>Scenedesmus</i>	200	48.4	60.5
<i>Chlorella</i> sp. Inmovilizada ^a	10	15.51	92.5
Cr (VI)			
<i>Chlorella vulgaris</i>	250	140	56
<i>Chlorella vulgaris</i>	147	63.2	43
<i>Scenedesmus quadricauda</i>	10	--	100
<i>Spirulina platensis</i>	2050	148.64	59.5
Pb			
<i>Chaetoceros</i> sp.	20	8	60
<i>Chlorella</i> sp.	20	10.4	78
<i>Clamydomonas reinhardtii</i> inmovilizada ^b	500	308.7	--
<i>Phormidium</i> sp.	10	2.305	92.2
Hg			
<i>Chlorella vulgaris</i>	5	17.49	72.9
<i>Chlorella</i> sp. DT	8	3.33	12.5
<i>Spirogyra</i> sp.	1	0.253	76

Fuente: Adaptado de Leong & Chang, 2020

^a *Chlorella* sp. Inmovilizada en un complejo de pellets de biocarbón derivados de hojas de jacinto de agua.

^b *Chlamydomonas reinhardtii* inmovilizada en perlas de alginato de calcio.

La asimilación de metales pesados se lleva a cabo mediante un proceso rápido de adsorción sobre la superficie de la célula (El-Sheekh et al., 2015), y puede producirse por diferentes vías: formando enlaces covalentes entre la pared celular ionizada y los metales pesados o por intercambio de iones con cationes de la pared celular. En cuanto al proceso de acumulación es más lento ya que los iones son transportados activamente a través de la membrana hasta el citoplasma, seguido de la difusión y posterior unión a proteínas y péptidos transportadores de metales, agentes reductores y fitoqueratinas) (Leong & Chang, 2020).

2.8. Metabolismo de Fármacos

Los fármacos son considerados contaminantes emergentes, los cuales pertenecen a un grupo diverso de compuestos orgánicos que contienen en su formulación principios activos, responsables de combatir infecciones, aliviar el dolor, regular el metabolismo, etc. Estas moléculas están diseñadas para mantener su estructura química durante largos periodos de tiempo, con el fin de ejecutar su función terapéutica dentro del organismo. En consecuencia, suelen ser persistentes, no biodegradables y, por lo tanto, bioacumulables en el medio ambiente. No obstante, aunque la mayoría de las concentraciones identificadas en el medio ambiente están en niveles de ng/L a mg/L, hay evidencia que demuestra que dichas concentraciones pueden inducir efectos ecológicos adversos en organismos objetivo y no objetivo, como alterar las comunidades microbianas, inhibir el crecimiento de microorganismos, reducir la actividad microbiana en el suelo, la tasa de desnitrificación de las bacterias, así como también fomentar el desarrollo de cepas resistentes a los antibióticos (Xiong et al., 2017).

Tabla 4. Degradación de fármacos a través del metabolismo de microalgas.

Cepa de microalga	Fármaco	Tipo de Efluente	Mecanismo de remoción y Eficiencia		
			Bioadsorción	Bioacumulación	Biodegradación
<i>Chlorella Sorokiniana</i>	Diclofenaco	Sintético	--	--	67%
	Diclofenaco	Doméstico después del tratamiento en reactor anaeróbico	9%	--	40 -60%
	Ibuprofeno		0.6%	--	100%
	Paracetamol		0.1%	--	100%
	Trimetoprima		7.6%	--	60%
Carbamazapina	16.7%		--	30%	
					60 – 19.7%
<i>Chlorella vulgaris</i>	Diclofenaco	Sintético	--	--	71%
	Levofloxacina		0.36 - 0.56%	0.85 – 1.1 %	2.8-92.3%
<i>Chlorella sp.</i>	Cefalexina	Sintético	71.2 – 82.7 %	--	--
	Florfenico		1.4 %	1.3 %.	72 - 97%
<i>Chlorella pyrenoidosa</i>	Norgestrel	Sintético	--	--	60 – 100%
	Cefalexina		--	--	92.7%
	7-ACA		--	--	96%
<i>Chlamydomonas sp.</i>	Ciprofloxacina	Municipal modificado	5.5%	--	65%
	Sulfadiazina		0.55%	--	17%
<i>Chlamydomonas acidophila</i>	Eritromicina con cafeína	Sintético	--	--	99%
	Claritromicina		--	--	65 – 93%.
	Carbamazapina		--	--	50 – 64%
<i>Chlamydomonas mexicana</i>	Carbamazapina	Sintético	--	194 – 135 mg/g	35.21%
<i>Scenedesmus obliquus</i>	Progesterona	Sintético	--	--	99%
	Diclofenaco		--	--	99%
	Carbamazapina		--	226 – 104 mg/g	21 – 28%
<i>Picocystis sp.</i>	Diclofenaco	Sintético	0.17%	0.17%	24 – 73%

Fuente: Adaptado de Rempel et al., 2021

Como ya se ha mencionado, las microalgas se alimentan de nutrientes presentes en su entorno circundante, lo que les confiere una gran flexibilidad para crecer y sobrevivir en ambientes extremos. Por ello, la adaptabilidad de las microalgas mixotróficas las convierte en candidatas prometedoras, ya que se ha demostrado que pueden eliminar, transformar y absorber fármacos presentes en aguas residuales (Tabla 4). Los mecanismos de asimilación de fármacos en microalgas incluyen la bioadsorción, la bioacumulación y la biodegradación, tanto intracelular como extracelular. Asimismo, se citan otros mecanismos e interacciones, como la volatilización (extracción), biotransformación, bioprecipitación (biomineralización) y mediante procesos de oxidación-reducción (Figura 5) (Xiong et al., 2017).

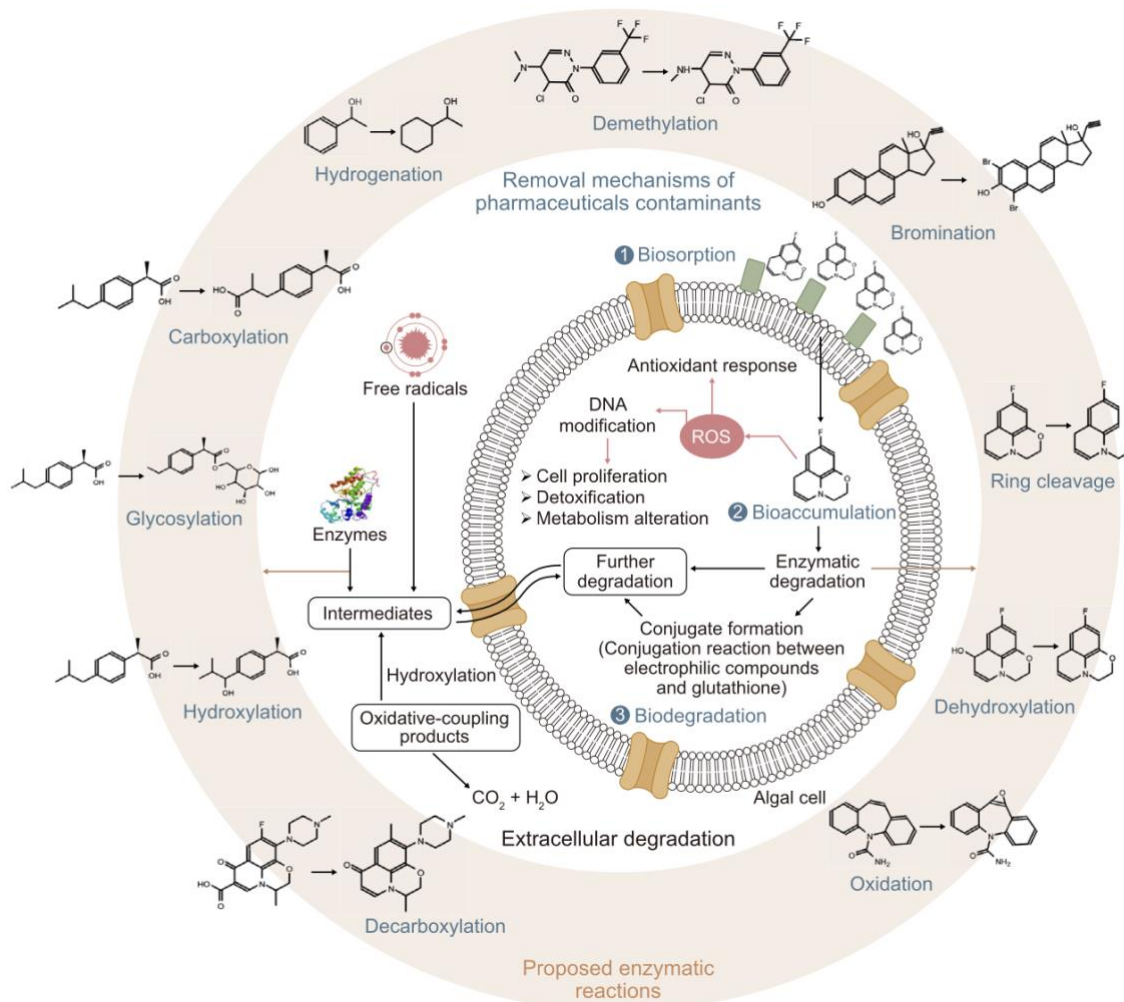


Figura 5. Procesos implicados en la remoción de fármacos por microalgas (Abdefattah et al., 2022).

Particularmente la bioadsorción de fármacos por las células de las microalgas ocurre de forma pasiva, ya que la pared celular de las algas presenta una carga negativa debido a los grupos funcionales dominantes, como el carboxilo (OH^-), sulfato (SO_4^{2-}), amino (NH_2^-) y carboxilato (COO^-) (Xiong et al., 2017). Estos actúan como sitios de unión y también como intercambiadores de iones que contribuyen a la acomplejación de iones metálicos y a la adsorción de sustancias orgánicas presentes en aguas contaminadas. En este sentido, ya que la adsorción es extracelular, el proceso de sorción varía significativamente dependiendo de la hidrofobicidad, la estructura y los grupos funcionales de los fármacos, así como también de la especie

de la microalga. Por otro lado, los contaminantes con grupos catiónicos son atraídos activamente hacia la superficie las microalgas mediante interacciones electrostáticas, resultando en una biosorción eficaz. Asimismo, la tasa de biosorción está influenciada por el oxígeno disuelto, la concentración del contaminante, el nivel inicial de pH y el tiempo de retención hidráulica (Kundu et al. 2024).

A diferencia de la bioadsorción, que se limita a la retención superficial de compuestos, la bioacumulación, es un proceso metabólico activo de internalización que involucra la absorción impulsada por energía, donde los contaminantes orgánicos son absorbidos por las microalgas en conjunto con los nutrientes presentes en el medio a través de la membrana celular. Cabe resaltar que la acumulación de fármacos en las células puede provocar la generación de especies reactivas de oxígeno (ROS). A las ROS se les asignan dos funciones celulares principales, en el que la primera, en niveles normales, las ROS actúan como moléculas de señalización esenciales en el control del metabolismo celular, que incluye la defensa contra patógenos, la muerte celular programada, así como también el comportamiento de los estomas. En el segundo caso, el exceso de estas, las fuertes propiedades oxidativas provocan daños graves a los componentes celulares (lípidos, proteínas y ADN), así como también un aumento en la tasa de mutagénesis, lo cual finalmente conduce a la muerte celular programada (Xiong et al., 2017).

Por otra parte, la biodegradación consiste en la eliminación completa o parcial de contaminantes de una fase acuosa, convirtiéndolos en intermediarios poco tóxicos o incluso no tóxicos. Las microalgas poseen un sistema enzimático complejo, mediante el cual catalizan la descomposición metabólica de los contaminantes durante la biodegradación, a través de varios procesos enzimáticos como la hidroxilación, la deshidrogenación, la glicosilación, la hidrogenación, la carboxilación, la desmetilación, la hidrólisis, entre otros (Xiong et al., 2017). En particular, la biodegradación de contaminantes mediante el uso de microalgas ocurre en tres etapas, en las que actúan diversas enzimas que facilitan las reacciones catalíticas. La primera fase se caracteriza por los ataques iniciales de la enzima citocromo P450, que aumenta la hidrofobicidad

de un compuesto mediante la adición o el enmascaramiento de un grupo hidroxilo, lo que generalmente implica reacciones de hidrólisis, oxidación o reducción (Abdelfattah et al., 2022). En la segunda etapa, enzimas como la glutatión-S transferasa (GST) catalizan la reacción de conjugación entre los compuestos electrofílicos y el glutatión, lo que protege a la célula del daño oxidativo. Por último, en la tercera etapa participan diversas enzimas (como la carboxilasa, la deshidrogenasa, las lacasas y la descarboxilasa, entre otras) para completar la degradación de los contaminantes. No obstante, la biodegradabilidad de los compuestos estará determinada por la complejidad de su estructura (Xiong et al., 2017) (Kundu et al. 2024).

Igualmente, las microalgas tienen la capacidad de realizar una degradación extracelular de fármacos mediante la liberación de varias sustancias poliméricas extracelulares incluyendo polisacáridos, proteínas, enzimas, sustituyentes (grupos metilo (CH₃) y acetilo (-COCH₃) unidos a polisacáridos), que pueden actuar como surfactantes y emulsificantes, aumentando la biodisponibilidad de los contaminantes y, a su vez, influyendo en la degradación intracelular a través de los subproductos formados. Por lo tanto, dichas sustancias poliméricas tienen la posibilidad de formar una biopelícula que actúa como un sistema digestivo externo, permitiendo metabolizar compuestos orgánicos, que se encuentran de forma disuelta, coloidal o sólida (Xiong et al., 2017)

A pesar de los beneficios que ofrece el empleo de microalgas en la degradación de contaminantes, aún persisten desafíos por delante, de los cuales las aguas residuales son un ambiente hostil para la mayoría de las de microalgas debido al desequilibrio nutricional, la falta de ciertos oligoelementos esenciales y presencia de compuestos tóxicos. Esto con lleva a la búsqueda de mejores cepas para condiciones extremas, especialmente para desequilibrios de nutrientes. Sin embargo, la optimización mediante ingeniería genética es actualmente un desafío debido principalmente a la información y conocimientos limitados en genómica de microalgas (Wu et al., 2019).

En este sentido, la caracterización génica proporciona una cobertura muy amplia acerca de la información genética sobre los sistemas biológicos de las microalgas (Azaman et al., 2020). Por ejemplo, Azaman et al. (2020) reportan la expresión genética diferencial de un total de 937 y 1124 genes regulados positiva y negativamente en muestras mixotróficas respectivamente. Adicionalmente, el análisis transcriptómico reveló que a condiciones mixotróficas causa una regulación positiva de los genes implicados en la producción de carotenoides, biosíntesis de ácidos grasos y acumulación de triacilglicerol en *Chlorella sorokiniana*. Por otro lado, un estudio de genómica comparada entre tres cepas de *Chlorella* reveló que las vías metabólicas implicadas en la asimilación de nitrógeno y fósforo se enriquecieron en la cepa *Chlorella sorokiniana* BD09 durante su crecimiento en aguas residuales. Adicionalmente, descubrieron que la variación estructural de los genes y el reordenamiento genómico podría contribuir a las capacidades diferenciales en la tolerancia en aguas residuales entre las cepas, como lo indica la variación del número de copias del gen, la reorganización de dominios de los ortólogos involucrados, así como también una inversión cromosómica de ~1Mb de longitud observada en *Chlorella sorokiniana* BD08 y *Dachan*. Además, especularon que identificaron una bacteria (*Microbacterium chocolatum*) asociada, que desempeña un posible papel en la sinergia de la eliminación de nutrientes en *Dachan* (Wu et al., 2019).

2.9. Caracterización genética

La caracterización genética permite evaluar la construcción genética de una especie y la relación que guarda esta con otras, así como también la expresión de ciertas características específicas a partir del entorno mediante técnicas de secuenciación, las cuales consisten en la determinación del orden de las bases nitrogenadas Adenina (A), Timina (T), Citosina (C) y Guanina (G) del ADN de un organismo a partir de diferentes técnicas como la secuenciación de Sanger, que consiste en un método enzimático que emplea desoxirribonucleótidos terminales, o la secuenciación automática empleando geles y capilares. Sin embargo, a comienzos del siglo XXI surgen nuevas técnicas basadas en la pirosecuenciación o la secuenciación en paralelo que emplea la

secuenciación de nueva generación (NGS, por sus siglas en inglés, Next Generation Sequencing) (Hernández et al., 2019).

2.10. Secuenciación masiva

La secuenciación masiva o HTS (por sus siglas en inglés, High-Throughput Sequencing) es una herramienta de NGS que permite el procesamiento masivo de genomas completos para la identificación rápida de genes de un organismo, es decir proporciona el catálogo completo de genes que una especie puede expresar a través de su ciclo de vida, reduciendo notablemente costos y tiempo en comparación con otras técnicas. Las NGS es aplicable a varias áreas de la biología, como la genómica, transcriptómica, epigenética, genómica de poblaciones, metagenómica entre otros (Aguilar-Bultet & Falquet, 2015). Entre las NGS que destacan están: la secuenciación por ligación (Sequencing by Oligonucleotide Ligation and Detection) del equipo SOLiD, secuenciación por síntesis y semiconducción por el equipo Ion Torrent y la secuenciación por síntesis del equipo Hiseq de la empresa Illumina (Reuter et al., 2015).

La secuenciación masiva consiste en 4 etapas principales: la extracción de ADN o ARN de la muestra o aislado, la preparación de bibliotecas o librerías, la secuenciación previamente mencionada y por último el análisis bioinformático e interpretación de resultados. Donde el primero consiste la obtención y purificación del ADN o ARN de interés, seguido de la creación de bibliotecas que consiste en la fragmentación de ADN o ARN en secuencias de un tamaño determinado por métodos bioquímicos (enzimáticos), físicos (nebulización o ultrasonidos) o la amplificación de fragmentos del ADN por PCR (reacción en cadena de la polimerasa). Por consiguiente, los fragmentos son marcados con índices para la posterior secuenciación de múltiples muestras y un adaptador de ADN complementario a uno existente en el secuenciador que permite adherir los fragmentos a un soporte para ser secuenciados. Por consiguiente, la etapa del análisis bioinformático, donde a partir de diferentes programas se puede realizar la generación y control de calidad, el alineamiento contra

bases de datos específicos, ensamblaje de *nov*, anotación del genoma, etc. para su posterior análisis e interpretación (Hernández et al., 2019).

2.11. Secuenciación por síntesis (illumina)

La secuenciación por síntesis consiste en la amplificación clonal de fragmentos de ADN ligados a un adaptador en la superficie de un portaobjetos de vidrio. Las bases se leen utilizando una estrategia de terminación cíclica reversible que se encarga de secuenciar la cadena plantilla un nucleótido a la vez, a través de rondas consecutivas de adición de bases, lavado, formación de imágenes y escisión. Adicionalmente esta técnica incorpora 3'-O-azidometil-dNTP marcados con fluorescencia para detener la reacción de polimerización permitiendo la eliminación de bases no incorporadas y la posterior lectura de los nucleótidos agregados. Por consiguiente, se eliminan los restos fluorescentes y el bloque 3' repitiendo el proceso nuevamente. En este tipo de secuenciación la tasa de error está por debajo del 1% siendo el error más común la sustitución (Figura 6) (Reuter et al., 2015).

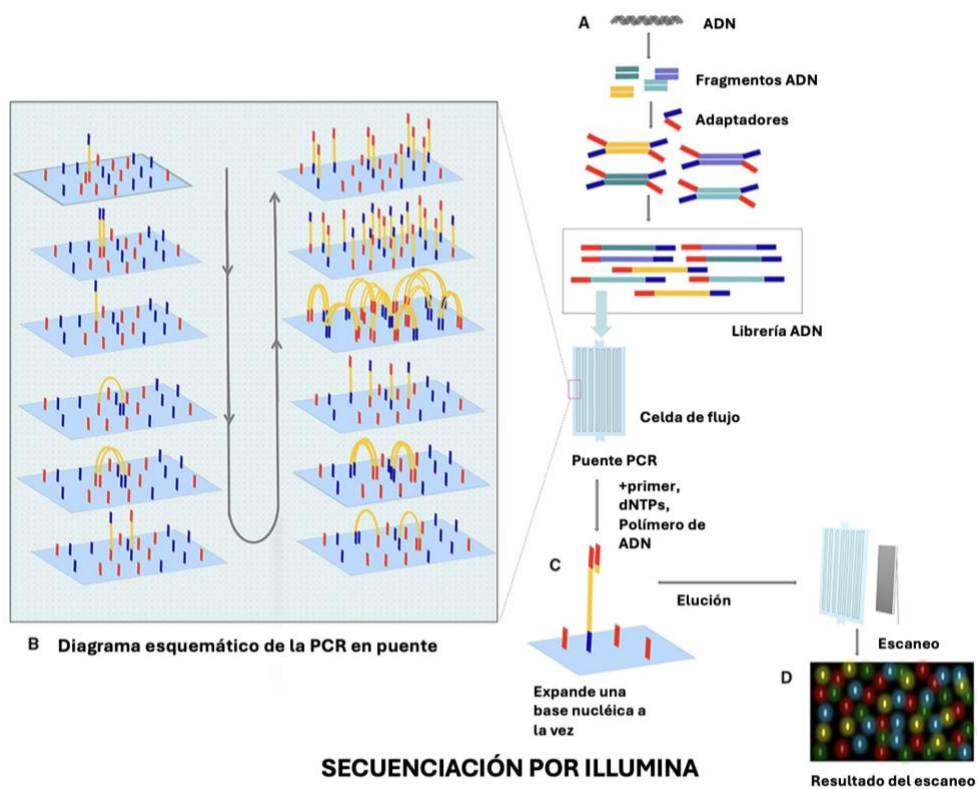


Figura 6. Proceso de secuenciación Illumina.

(A) Biblioteca de ADN. Se rompe el ADN del genoma para formar fragmentos de ADN, se añaden adaptadores en ambos extremos y se construye una biblioteca de ADN monocatenario (ssDNA). Después de agregar la biblioteca de ADN a la celda de flujo, los fragmentos de ADN se unirán a la superficie de la celda de flujo al pasar a través de ella. **(B)** PCR de puente. Cada fragmento de ADN se agrupa en su posición. Después de la amplificación por puente, cada grupo contiene muchas copias de la plantilla ssDNA. **(C)** SBS (Sequencing by Synthesis). La ADN polimerasa, los cebadores y cuatro dNTP con fluorescencia específica se añaden al sistema de reacción al mismo tiempo. El extremo 3' de estos dNTP está conectado con un grupo de azida, que bloquea la incorporación de la siguiente base, por lo que solo se puede extender una base a la vez. Después de lavar los dNTP y la enzima restantes con agua, se escanean las imágenes. Tras el escaneo, se agrega un reactivo químico, se corta el grupo de azida y se atenúa la fluorescencia, iniciando el siguiente ciclo. **(D)** Recepción de la señal fluorescente (Zhang et al., 2021).

2.12. Ensamblaje del genoma

El ensamblaje del genoma consiste en la construcción de una secuencia genómica desconocida a través de la detección de superposiciones entre un conjunto de lecturas genómicas redundantes (Di Genova et al., 2020). Proceso en el cual secuencias o *reads* (lecturas o secuencia de ADN obtenida de un secuenciador), que son fragmentos secuenciados continuos, se solapan en fragmentos más grandes llamados *contigs* que a su vez se ensamblan en fragmentos aún más prologados (*scaffolds*)

hasta la reconstrucción del genoma (Figura 7) (Hernández et al., 2019). Cabe destacar que el uso de lecturas pareadas (paired-end reads), que consisten en secuenciar los dos extremos de la misma molécula de ADN, proporciona información adicional sobre la distancia y orientación entre secuencias, mejorando significativamente la precisión del ensamble y facilitando la identificación de regiones repetidas (Basantani et al., 2017).

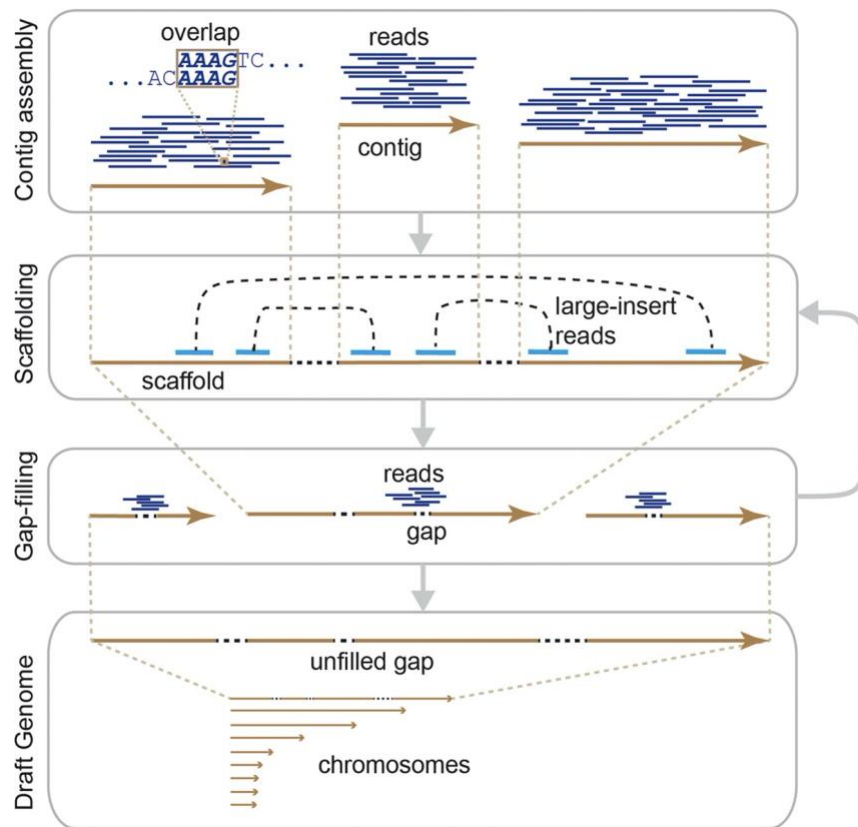


Figura 7. Flujo de trabajo general del ensamblaje de novo de un genoma completo.

Mediante la superposición de lecturas, se ensamblan *contigs* a partir de lecturas cortas antes de realizar el *scaffolding* con lecturas de inserción larga, y los huecos restantes se rellenan. Los pasos de *scaffolding* y de relleno de huecos pueden realizarse de manera iterativa hasta que no se ensamblen más *contigs* o no se resuelvan huecos adicionales antes de la finalización. A través de este procedimiento, se construye un genoma en borrador compuesto por cromosomas. Algunos huecos sin relleno pueden permanecer en el genoma en borrador (Sohn & Nam, 2016).

2.13. Ensamblaje de novo

En este aspecto, el ensamblaje *de novo* es una estrategia de ensamblaje del genoma en la cual se utiliza la información obtenida a partir de una secuenciación previa para

la reconstrucción del genoma en cuestión, sin conocimiento previo del mismo. Dicha herramienta es de mucha utilidad ya que permite determinar el tamaño del genoma, el contenido de G y C, además del contenido de regiones repetidas (Aguilar-Bultet & Falquet, 2015). Por otro lado, existe el ensamblaje híbrido que consiste en el armado del genoma utilizando secuencias cortas y secuencias largas (Di Genova et al., 2020). Una vez obtenido en el ensamblaje del genoma, es posible hacer una asignación a las funciones de este, mediante su anotación lo que sugiere una serie de procesos como la búsqueda de posibles genes, su traducción a proteínas, así como también la identificación de dichas proteínas mediante la comparación con bases de datos (Hernández et al., 2019).

Generalmente, los programas de ensamblaje de genomas utilizan dos clases de algoritmos: Superposición–Diseño–Consenso (Overlap–Layout–Consensus, OLC) y De Bruijn Graph (DBG). El primero busca identificar el solapamiento entre todas las lecturas, luego construye un grafo que representa dichas superposiciones y, finalmente, genera la secuencia consenso, es decir, una secuencia representativa altamente conservada entre las lecturas de ADN. El enfoque OLC es el más adecuado para el ensamblaje de lecturas largas y de baja cobertura, y es empleado por ensambladores como Arachne, Celera Assembler, CAP3 y PCAP, entre otros (Basantani et al., 2017). El enfoque DBG utiliza lecturas cortas para ensamblar el genoma. Como primer paso, descompone las lecturas en subsecuencias de longitud k , llamadas k -mers. Después, construye un grafo definiendo primero un prefijo (un k -mer sin el último nucleótido) y un sufijo (un k -mer sin el primer nucleótido). La superposición del conjunto total de prefijos y sufijos únicos forma los nodos del grafo, y los bordes se agregan en función de los k -mers que vinculan un sufijo y un prefijo dados. El ensamblaje de secuencias más largas se realiza encontrando un ciclo euleriano en el grafo, es decir, un recorrido que pasa por cada borde exactamente una vez, y luego colapsando la secuencia de los k -mers en dicho recorrido para reconstruir secuencias más extensas. Muchos ensambladores de genomas de lectura corta se han desarrollado en base al enfoque DBG, como Euler-USR, Velvet, ABySS,

SOAPdenovo y MEGAHIT, siendo estos tres últimos diseñados con un enfoque en la mejora de la eficiencia computacional y en el manejo de volúmenes de datos más extensos, como ocurre en el caso de los metagenomas, lo que permite obtener ensamblajes más precisos (Figura 8) (Dufault-Thompson & Jiang, 2022).

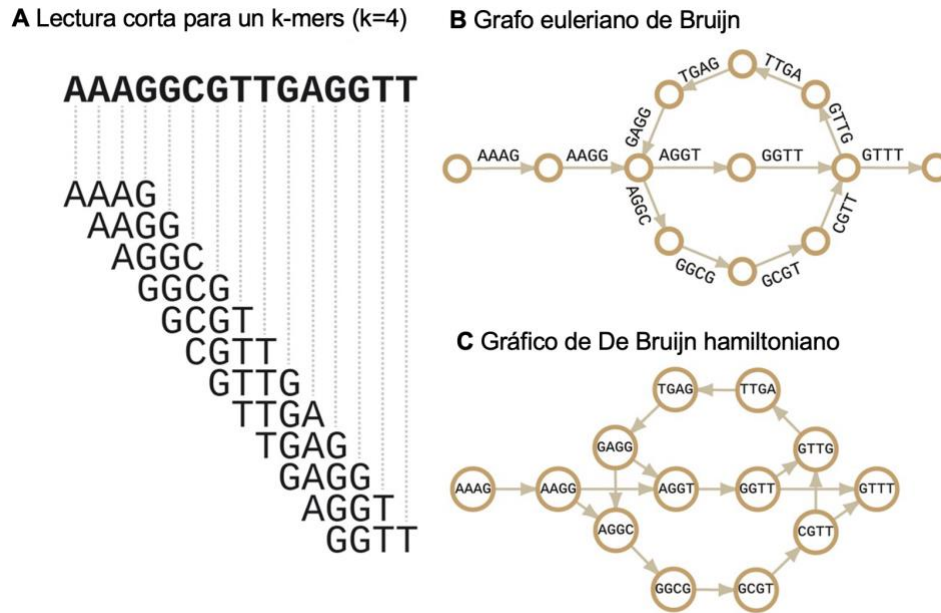


Figura 8. (A) K-mers con una longitud de 4 bases a partir de una secuencia. (B) Construcción del gráfico de bruijn a partir de los k-mers. (C) En el enfoque hamiltoniano, los k-mers (o secuencias) son nodos, mientras que son los bordes en el enfoque euleriano.

2.14. Control de Calidad

El flujo de trabajo en análisis genómicos implica un control de calidad como parte integral de su proceso de análisis, debido a que la mayoría de los secuenciadores pueden generar errores, contaminaciones o sesgos técnicos, que pueden afectar de forma significativa los análisis posteriores, así como su interpretación. Generalmente, los secuenciadores realizan un informe de control de calidad, sin embargo, habitualmente solo se centran en identificar problemas generados por el propio equipo. Por ello, existen varias herramientas de control de calidad de secuencias, como FastQc, Sickle y RNA-SeQC (Castellanos-Rodríguez et al., 2023) (Shin & Xu, 2016).

En el que FastQC permite visualizar la calidad de las lecturas, tanto en paired-end reads como en single-end. En el caso de doble extremo (paired-end), Sickle elimina simultáneamente el segmento de lectura inversa correspondiente y, al mismo tiempo filtrar el segmento de lectura directo con muchas bases de baja calidad. Por otra parte, RNA-SeQC calcula los indicadores de calidad en datos de RNA-seq para verificar si los datos generados cumplen con los estándares para un análisis confiable. Entre las métricas que reporta se encuentran la profundidad de secuenciación (es decir, la cobertura), la distribución de lecturas en regiones génicas (exones, intrones, genes) y el contenido de ARNr, entre otras (Castellanos-Rodríguez et al., 2023).

El paso de evaluación de calidad mide un conjunto de estadísticas en un archivo de lecturas secuenciadas para evaluar si su contenido coincide con las expectativas del experimento y si los datos son adecuados para análisis posteriores. Por consiguiente, los datos que superan el control de calidad se someten después a etapas adicionales de análisis, que pueden incluir el recorte de adaptadores, el filtrado de contaminantes y de lecturas de baja calidad, y finalmente el mapeo de las lecturas resultantes a un genoma o transcriptoma de referencia (Shin & Xu, 2016).

2.15. Anotación genómica

La anotación del genoma se refiere al proceso de identificar las posiciones y funciones correspondientes a todos los genes, regiones codificantes, elementos reguladores, etc., en el genoma de un organismo. Puede dividirse conceptualmente en dos etapas: anotación estructural y anotación funcional. La anotación estructural se ocupa de encontrar las posiciones y la estructura de genes (codificantes de proteínas y no codificantes) y otros sitios de relevancia biológica, como regiones reguladoras, pseudogenes y elementos repetitivos como transposones. Mientras que la anotación funcional busca determinar las funciones de toda la información biológica relevante de cada uno de los sitios anotados en la primera etapa (Ejigu & Jun, 2020).

Como ya se mencionó la identificación de genes codificantes y otros elementos reguladores son el centro de la anotación génica, por lo que resulta fundamental la

predicción genética como principio de la anotación estructural, un procesos complejo, especialmente para el ADN eucariota. Esta predicción se logra mediante el uso de diferentes enfoques, entre ellos: el método estadístico (o método *ab initio*), el método de similitud de secuencias, es decir, anotación basada en homología o por la combinación de ambos. El método *ab initio* se basa en modelos estadísticos, como el modelo oculto de Markov (HMM), para identificar los promotores, las regiones codificantes y no codificantes, así como las uniones intrón-exón de la secuencia del genoma mediante información intrínseca de la secuencia genómica (patrones, señales o características estadísticas). Este método implementa programas como AUGUSTUS, FGENESH, GENSCAN basados en HMM. Por otra parte, el enfoque de homología alinea la secuencia con etiquetas de secuencias expresadas (EST), ADN complementario (ADNc) o evidencia de proteínas, y utiliza las similitudes detectadas para la predicción de genes. Utilizado herramientas de alineamiento como BAST (por sus siglas en inglés, Basic Local alignment Search Tool) y TWINSKAN. En cambio, el otro grupo comprende a programas que combinan *ab initio* y el enfoque basados en homología para la predicción de genes. Dentro de las herramientas que comprenden este grupo son FGENESH-C basado en HMM y cDNA para las predicciones y GenomeScan que utiliza información extrínseca de alineamiento de proteínas BLAST (Ejigu & Jun, 2020).

Como siguiente paso es la asociación de información biológica con secuencias de genes o proteínas que fueron identificadas durante la anotación estructural, en el que su objetivo es asignar una descripción funcional a cada gen, lo que incluye su función molecular, componentes celulares en los que actúan y los procesos biológicos en los que están involucrados. La predicción de la función se puede lograr automáticamente y directamente utilizando herramientas de alineación local, como BLAST, donde se busca en una base de datos de proteínas alineaciones de alta puntuación. Después, la función se asigna a la secuencia de consulta desconocida basada en una secuencia conocida, siempre y cuando la puntuación de alineación sea la más alta, por encima de algún valor umbral especificado. En otras palabras, las

herramientas identifican las relaciones evolutivas en base a relaciones ortólogas y parálogas entre secuencias. Los ortólogos son genes que se han originado a partir de un solo gen ancestral en el último ancestro común de los genomas comparados, mientras que los parálogos son genes dentro del mismo genoma que han surgido de duplicaciones. Dentro de las bases de datos para anotaciones funcionales se encuentra la Ontología Génica (GO), un recurso ampliamente utilizado y considerado la base de conocimientos más completa para la asignación de funciones a genes y proteínas. GO clasifica la información en tres categorías principales: función molecular (la actividad de un producto genético a nivel molecular), componentes celulares (ubicación del producto genético) y procesos biológicos (actividad biológica en la que participa la función de un gen) (Ejigu & Jun, 2020). Además de GO, KEGG Orthology (KO), en conjunto con KEGG (por sus siglas en inglés, Kyoto Encyclopedia of Genes and Genomes), actúa como un vínculo entre los datos genómicos y la base de datos. Esto permite integrar los genes en rutas metabólicas y esquemas de señalización, facilitando la comprensión de las funciones y la utilidad a alto nivel de un sistema biológico, ya sea a nivel celular, de organismo o de ecosistema (Kanehisa Laboratories, s.f.). (Ejigu & Jun, 2020). Por otra parte, se encuentra COG (por sus siglas en inglés, Clusters of Orthologous Genes), que es una base de datos que realiza una clasificación filogenética de las proteínas codificadas en 2,103 genomas completos de bacterias y 193 de arqueas (National Center for Biotechnology Information, s. f.). Los COG se organizan en 17 categorías funcionales, representadas por una letra. Por ejemplo: J, traducción (incluye la estructura y biogénesis de los ribosomas); L, replicación, recombinación y reparación; K, transcripción; O, chaperonas moleculares y funciones relacionadas; M, estructura y biogénesis de la pared celular y membrana externa; N, secreción, motilidad y quimiotaxis; T, transducción de señales; P, transporte y metabolismo de iones inorgánicos; C, producción y conversión de energía; G, metabolismo y transporte de carbohidratos; E, metabolismo y transporte de aminoácidos; F, metabolismo y transporte de nucleótidos; H, metabolismo de coenzimas; I, metabolismo lipídico; D, división celular y partición cromosómica; R, predicción funcional general; y S, sin predicción funcional o

desconocida. Cabe resaltar que la asignación de cada letra a una función puede variar dependiendo de la bibliografía consultada (Figura 9) (Tatusov, 2000).

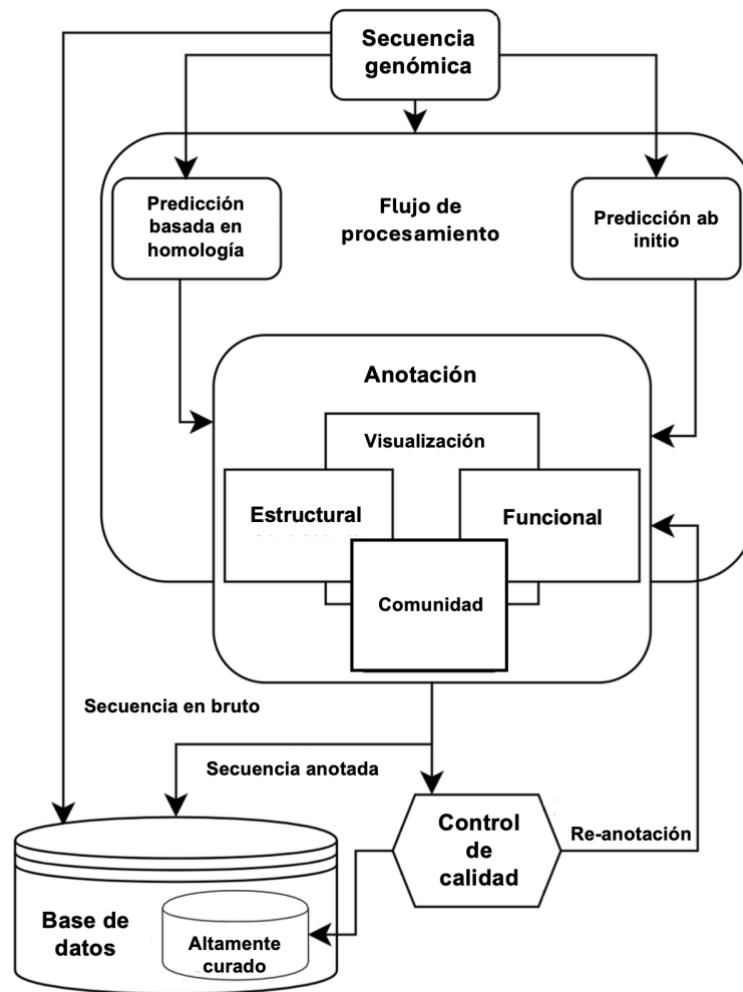


Figura 9. Flujo de trabajo de anotación del genoma (Ejigu & Jun, 2020).

2. HIPÓTESIS

Las técnicas de secuenciación masiva permiten la Identificación funcional de genes de la microalga *Chlorella vulgaris* útiles para la biorremediación de aguas.

3. JUSTIFICACIÓN

La biorremediación es un proceso biológico en el cual los microorganismos incorporan contaminantes en sus procesos biológicos para usarlos como energía. La microalga *Chlorella vulgaris* presenta cualidades de adaptación ambiental que la hacen un buen candidato para el análisis de biorremediación.

4. OBJETIVO

Realizar la caracterización genómica de *Chlorella vulgaris* mediante datos de secuenciación masiva y anotación funcional del genoma para la selección de genes relacionados a la biorremediación.

5. OBJETIVOS ESPECÍFICOS

- a) Obtener la secuenciación del genoma de *C. vulgaris* a partir de la secuenciación masiva.
- b) Establecer el ensamblaje de *novo* de *C. vulgaris* mediante estrategias bioinformáticas.
- c) Anotar y caracterizar genes funcionales del genoma.
- d) Identificar genes asociados a la degradación de contaminantes.
- e) Determinar las rutas más relevantes relacionadas con la asimilación de contaminantes.

6. MATERIALES Y MÉTODOS

6.1. Cultivo de la microalga y extracción de ADN

C. vulgaris fue cultivada en medio Bold a un volumen de 50 mL, y crecida a 25°C con agitación continua y mantenido en condiciones estériles, posteriormente la biomasa de la microalga se recuperó en la fase media exponencial (6 días) y se cosechó mediante centrifugación a 4000 rpm por 5 min, posteriormente, para la extracción del ADN se hizo uso del kit ZymoBIOMICS™ DNA Miniprep kit siguiendo las indicaciones del fabricante.

6.2. Comprobación de la cantidad y calidad del ADN genómico

Se cuantificó y verificó la calidad e integridad del ADN extraído mediante el uso de NanoDrop™2000 (ThermoFisher Scientific) y mediante un gel de agarosa de electroforesis al 1% usando Super Buffer 1x.

La cuantificación del ADN se realizó mediante espectrofotometría por medio de la lectura de absorbancias con una longitud de onda específica de 260 nm. La pureza del ADN se determinó mediante los parámetros 260/230 y 260/280.

6.3. Secuenciación masiva de la microalga *C. vulgaris*

El ADN se fragmentó mediante un ultrasonificador y las bibliotecas de secuenciación se generaron utilizando el kit NEB Next Ultra Dna library, posteriormente la secuenciación de estas bibliotecas se llevó a cabo mediante la tecnología de secuenciación por síntesis desarrollada por Illumina, utilizando un equipo HiSeq 2500, empleando lecturas pareadas (*paired-end reads*) de 100 nucleótidos.

6.4. Ensamble del genoma

Para el análisis de calidad y filtrado de las secuencias obtenidas, se utilizó el software FastQc y Trimgalore respectivamente.

Una vez realizado el análisis de calidad y filtrado de las lecturas (en inglés reads), se ejecutó el software KmerGenie este nos permitió calcular el mejor valor de K para optimizar el proceso de ensamble, una vez obtenido este valor se procedió a generar el ensamble *de novo* para ello se utilizaron cuatro softwares (ABYSS, MEGAHIT, SPAdes y Unicycler) los cuales permitieron generar ensamblajes a nivel *contig* y *scaffold*.

Posteriormente con el fin de evaluar la calidad del ensamble se utilizó el software QUAST el cual llevo a cabo el análisis de calidad del ensamble a nivel *contig* y/o *scaffold*. Por último, se utilizó la herramienta bioinformática Sealer para cerrar gaps presentes en el ensamble a nivel *scaffold*.

6.5. Anotación del genoma, anotación de ontología génica (GO) clasificación en categorías, componentes celulares (CC), función molecular (FM), y procesos biológicos (PC).

Una vez obtenido el ensamble a nivel *contig* y/o *scaffold*, se utilizaron distintos softwares como Augustus (3.2.1), GeneMark-ES y NCBI-blast (V2.2.29). Primeramente, se utilizó GeneMark-es para la predicción de genes *ab initio* utilizando *Chlamydomonas reinhardtii* como referencia, una vez obtenida las secuencias codificantes predichas se ejecutó en BLASTp con la opción habilitada de “análisis con bases de datos no redundantes (nr)” y con un e-value $\leq 1e^{-5}$, posteriormente se utilizó Augustus (3.21) con el fin de predecir el total de genes codificadores. Para la anotación funcional de genes se utilizó NCIB-blast (V2.2.29) junto con algunas bases de datos

como Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG) y Ontología génica(GO), esta última permitió clasificar los genes en varias categorías como, componente celular (CC), proceso biológico (BP) y función molecular (MF).

6.6. Obtención de rutas metabólicas asociadas a la degradación de contaminantes

La base de datos “Encyclopedia de Genes y Genomes de Kyoto” (KEGG) se utilizó para reconstruir las rutas metabólicas relacionadas a la degradación de contaminantes. Para identificar los grupos de genes ortólogos, las secuencias se presentaron a KEGG Automatic Annotation Server (KAAS) y se anotaron las secuencias con un identificador único de ortólogo (KO) de KEGG.

7. RESULTADOS Y DISCUSIÓN

7.1 Comprobación de la cantidad y calidad del ADN genómico

7.1.1. Cantidad y calidad del ADN genómico

Chlorella vulgaris fue cultivada hasta alcanzar la fase exponencial bajo condiciones axénicas. La biomasa se recuperó por centrifugación a los 6 días, posteriormente se utilizó para la extracción de ADN.

La extracción de ADN se realizó con el fin de obtener material genético de alta calidad para los análisis metagenómicos y trazar los mecanismos moleculares responsables de la asimilación de contaminantes. Se utilizó el kit ZymoBIOMICS™ DNA Miniprep kit siguiendo las instrucciones del fabricante con la finalidad de garantizar la calidad y pureza requeridos para los análisis posteriores. Para validar la extracción del genoma se cuantificó por espectrofotometría obteniendo una concentración de 49.4 ng/ μ L, con una relación de absorbancia de 260/280 de 1.988 las cuales cumplen con las especificaciones del laboratorio de secuenciación (Figura 10a). Por otra parte, se verificó la calidad e integridad del ADN extraído a partir de un gel de agarosa de electroforesis al 1% (Figura 10b), mostrando una banda clara y definida con un poco de barrido que se le puede atribuir a la fragmentación del genoma a consecuencia de la extracción. Sin embargo, se obtuvo un tamaño superior a 5Kb correspondiente al tamaño esperado, lo que indica que el ADN está intacto y libre de contaminantes, garantizando la calidad necesaria para el análisis de secuenciación por la tecnología Illumina.

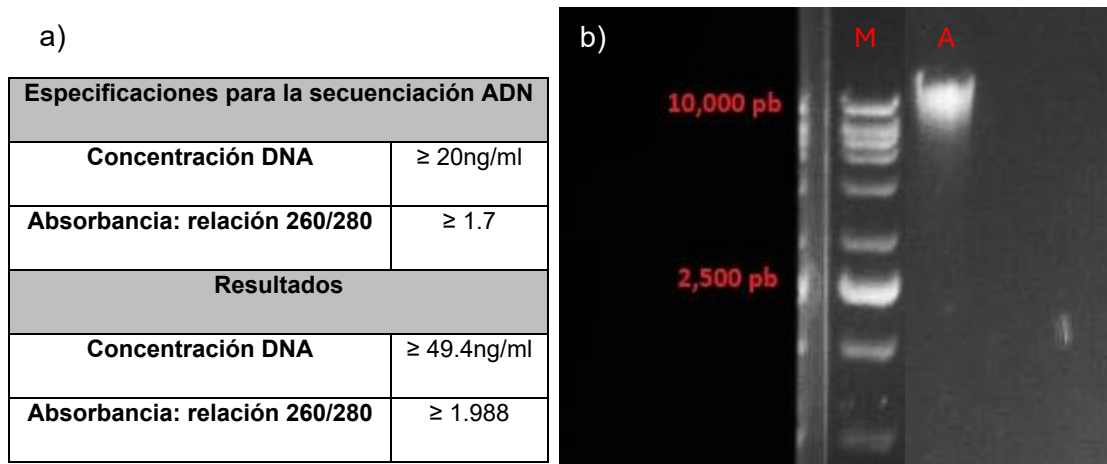


Figura 10. Resultados de la extracción de ADN de *C. vulgaris*.

a) Especificaciones de la extracción del ADN requeridas y resultados a partir de la extracción. b) Resultados de electroforesis en gel de agarosa al 1%, que muestra la integridad del ADN extraído de la microalga. Carril M, marcador de peso molecular. Carril A, muestra de ADN extraído.

7.1.2. Verificación de calidad

La construcción de las librerías (Kit NEB Next Ultra Dna library) y la secuenciación se llevó a cabo por Zymo Research (Orange, California, EUA) con la tecnología de secuenciación por síntesis en el equipo Illumina HiSeq 2500, empleando lecturas pareadas (paired-end reads) de 100 nucleótidos. Se recibieron 2 archivos de 1MB de salida en formato Fasta cada uno. Con el fin de determinar la calidad de las secuencias se realizó un control de calidad a cargo de la herramienta *FastQC* (Fast Quality Control), que usualmente es el primer paso en los procesos de análisis de datos de alto rendimiento. Este software se encarga de medir un conjunto de estadísticas en un archivo de lecturas secuenciadas para verificar si su contenido coincide con las expectativas del experimento y si estas son óptimas para los siguientes análisis (De Sena Brandine & Smith, 2021) (Figura 11a y Figura 11b).

Se desarrolló un análisis de calidad de las lecturas pareadas crudas a partir de la secuenciación y de las lecturas pareadas filtradas por Trimgalore, este último con la finalidad de obtener lecturas de mejor calidad. Para ello se utilizó el software *FastQC*. Donde se evaluaron los principales valores de interés como: La cantidad de lecturas,

bases totales de las lecturas, rango de longitud de las lecturas y el porcentaje de los nucleótidos guanina y citosina (GC), además de la calidad resultado de distintos módulos de FastQC.

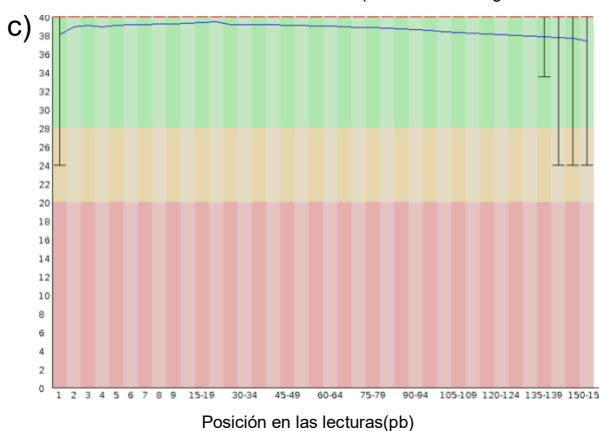
a)

Medida	Valor
Nombre del archivo	Zr12871_3_S547_LOO8_R1_002.fastqc.gz
Tipo de archivo	Llamandas de bases convencionales
Codificación	Sanger /Illumina 1.9
Secuencias totales	84849392
Bases totales	12,8 GBP
Secuencias marcadas como de mala calidad	0
Longitud de las %GC	151 59

b)

Medida	Valor
Nombre del archivo	Zr12871_3_S547_LOO8_R2_001.fastqc.gz
Tipo de archivo	Llamandas de bases convencionales
Codificación	Sanger /Illumina 1.9
Secuencias totales	84849392
Bases totales	12,8 GBP
Secuencias marcadas como de mala calidad	0
Longitud de las %GC	151 59

Puntuaciones de calidad en todas las bases (codificación Sanger/Illumina 1.9)



Puntuaciones de calidad en todas las bases (codificación Sanger/Illumina 1.9)

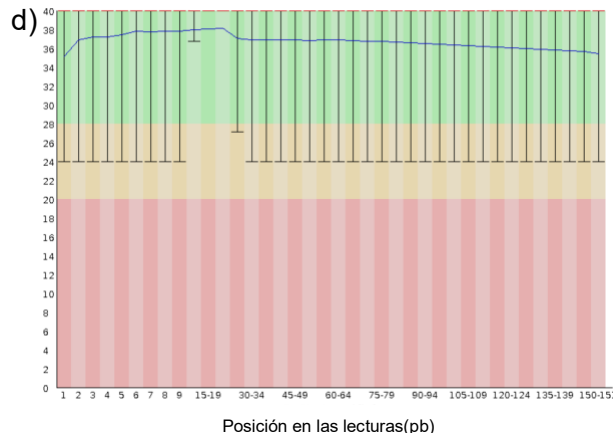


Figura 11. Análisis de calidad realizado por el software FastQC para las lecturas crudas del genoma de *C. vulgaris*. a) Parámetros básicos de calidad para las secuencias sentido crudas. b) Parámetros básicos de calidad para las secuencias anti-sentido crudas. c) Gráfico de calidad para las secuencias sentido crudas. d) Gráfico de calidad para las secuencias anti-sentido crudas.

En el proceso de secuenciación del genoma de *C. vulgaris* se generaron un total de 169 698 784 de lecturas pareadas con un tamaño de 151pb y el contenido de GC fue del 59%. En la parte inferior de la figura 11(c y d) se representan los gráficos de calidad por base para las secuencias, sentido y anti-sentido. Donde estos gráficos representan en el eje las abscisas cada una de las bases de una secuencia y en el eje de las ordenadas los valores de calidad phred (Q). El valor Q representa la probabilidad

de que una base se llame incorrectamente en la secuenciación, donde $Q=30$ corresponde nominalmente a una tasa de error del 0.1% (Liao et al., 2017). Estos valores dividen el gráfico horizontalmente en 3 franjas: verde, amarilla y roja, donde la primera representa valores de calidad altos (valores mayores a 30), la segunda valores de calidad intermedios (valores menores a 30 y mayor a 20) y la última valores de calidad bajos (valores menores a 20) (Cortese et al., 2021). En la Figura 11c se observa que las lecturas sentido-crudas mostraron una distribución completa sobre la franja verde obteniendo un valor Q superior a 30. Mientras que las lecturas anti-sentido mostraron una distribución menos homogénea indicando una mayor variabilidad de las bases en diferentes posiciones. Sin embargo, se obtuvo un valor Q superior a 20 lo que indica que se obtuvo un valor intermedio en la calidad (Figura 11d).

Las lecturas filtradas mostraron una disminución en el número total de lecturas, así como también una disminución en la longitud de las secuencias con un tamaño entre 20 y 135 pb. Las lecturas sentido-filtradas mantuvieron la misma calidad que las lecturas sentido crudas en todas las bases. Por otra parte, las lecturas anti-sentido filtradas presentaron una mejoría en la calidad de las primeras bases en comparación con lo observado en las lecturas anti-sentido crudas. Sin embargo, se conservó un valor Q intermedio (Figura 12).

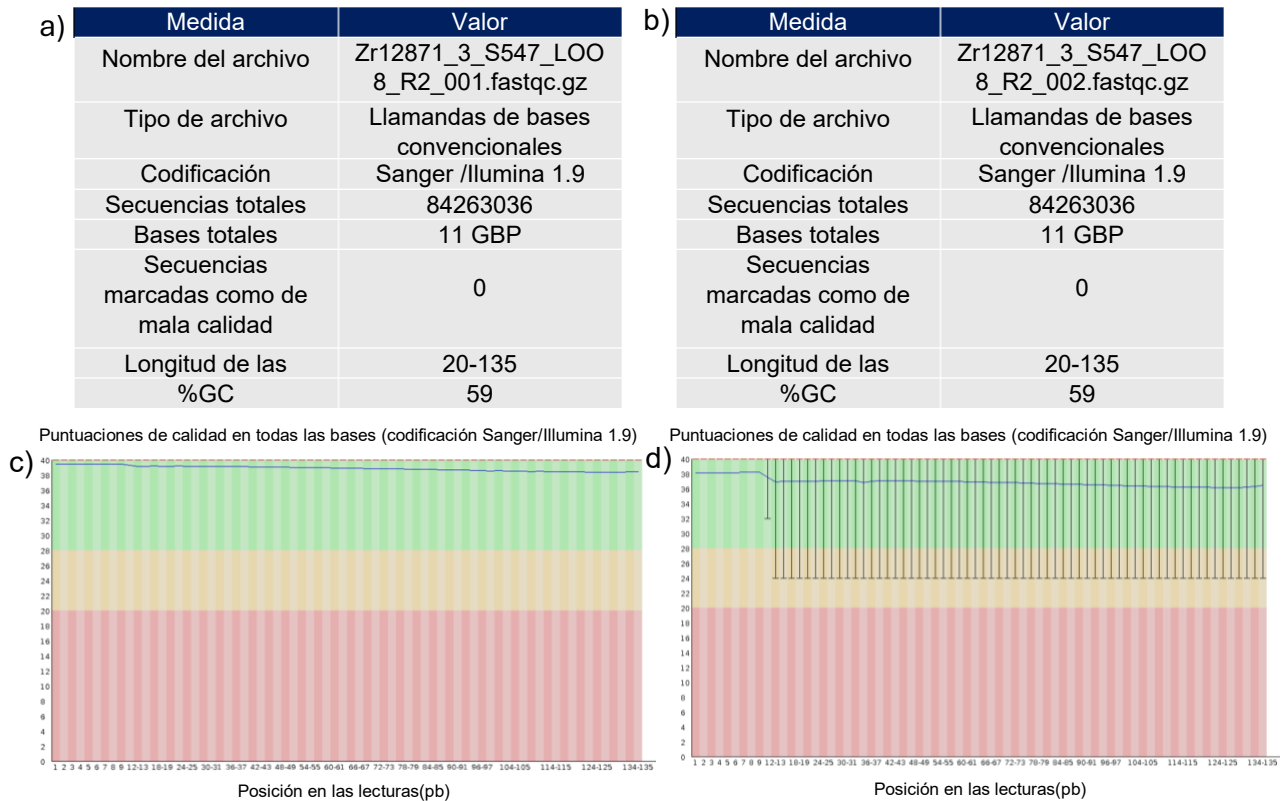


Figura 12. Análisis de calidad realizado por el software FastQC para las lecturas filtradas del genoma de *C. vulgaris*. a) Parámetros básicos de calidad para las secuencias sentido filtradas. b) Parámetros básicos de calidad para las secuencias anti-sentido filtradas. c) Gráfico de calidad para las secuencias sentido filtradas. d) Gráfico de calidad para las secuencias anti-sentido filtradas.

Por consiguiente, se utilizó la herramienta Kraken para asignar etiquetas taxonómicas a las lecturas obtenidas tras el filtrado con Trimgalore y, de esta manera, estimar la composición taxonómica de la muestra en cuestión. De un total de 84,263,036 lecturas procesadas, solo el 83.3% pudo clasificarse, de las cuales 50,671,528 (72.83%) estaban taxonómicamente afiliadas al clado *Viridiplantae*, donde el 70.73% (49,206,966 lecturas) pertenece al filo *Chlorophyta* y, por consiguiente, el 70.61% (49,124,981 lecturas) corresponde a la especie *C. reinhardtii*. Adicionalmente, también se clasificaron 18,254,181 (26.24%) lecturas pertenecientes a bacterias, entre las cuales predominó el género *Pseudomonas* con el 11.82% (8,226,888 lecturas) (Figura A1). En esta dirección, se ha observado que la interacción sinérgica entre microalgas y bacterias puede ser fundamental en la eliminación de contaminantes, en base a un complejo intercambio nutricional de metabolitos entre dichos

microorganismos (Nagarajan et al., 2022). No obstante, el presente estudio se centra en los genes de *C. vulgaris* para comprender el papel que pueden desempeñar las microalgas dentro de esta comunidad microbiana.

7.2 Ensamble del genoma

Para el ensamblaje de *novo* del genoma de *C. vulgaris* se utilizaron cuatro ensambladores (ABYSS, MEGAHIT, SPAdes y Unicycler), con el objetivo de obtener un mejor ensamblaje en términos de contigüidad y completitud funcional. Una vez realizados los ensamblajes, los mismos fueron evaluados usando la herramienta QUAST (*Quality Assessment Tool*), la cual permite evaluar y comparar la calidad de ensamblajes genómicos utilizando diversas métricas (número de *contigs*, mayor *contig*, Longitud total, N50, L50, etc.). Los resultados estadísticos de los ensamblajes, obtenidos mediante QUAST, se presentan la Tabla 5.

A partir del ensamblaje con las diferentes herramientas, se observa que Unicycler logró ensamblar el genoma con 12,466 contigs, mientras que ABYSS, MEGAHIT y SPAdes, lo hicieron con 31,809, 27,018 y 25,998 contigs, respectivamente. Esto señala que Unicycler generó contigs mucho más largos, lo que resultó en un genoma menos fragmentado y más contiguo. Asimismo, con MEGAHIT se obtuvo el contig más largo, de 1,413,394 pb seguido por Unicycler (909,350), en comparación ABYSS (479,042) y SPAdes (630,416), lo que señala que MEGAHIT y Unicycler generaron fragmentos individuales más largos, lo que sugiere un ensamblaje de mayor calidad. Sin embargo, MEGAHIT (27,018 contigs) generó más *contigs* que Unicycler (12,466 contigs), lo que evidencia una mayor fragmentación del genoma por parte de MEGAHIT. Otra métrica relevante es el valor de N50, que describe la longitud de contig mínima de los contigs más largos que se necesitan para cubrir el 50% del genoma. En este caso, el valor más alto de N50 se obtuvo con Unicycler, con un N50 de 24,335 pb, lo que sugiere que se necesitan menos fragmentos para cubrir la mitad del genoma. Por lo tanto, a partir de la evaluación realizada por QUAST, el mejor ensamblaje está dado por la herramienta de ensamblaje Unicycler.

Tabla 5. Evaluación de la calidad entre los ensamblajes realizados con ABySS, MEGAHIT, SPAdes y Unicycler mediante la herramienta QUAST.

Ensamble	ABySS	MEGAHIT	SPAdes	Unicycler
# contigs (>= 0 bp)	361781	56409	124170	17177
# contigs (>= 1000 bp)	19197	17229	17646	11627
# contigs (>= 5000 bp)	7090	7195	7544	6316
# contigs (>= 10000 bp)	3285	3902	3923	3748
# contigs (>= 25000 bp)	680	1141	963	1182
# contigs (>= 50000 bp)	140	311	257	342
Longitud total (>= 0 bp)	189686651	173152234	172872623	144323546
Longitud total (>= 1000 bp)	128418324	156863471	151156539	142743515
Longitud total (>= 5000 bp)	97515425	133366296	126927293	129192635
Longitud total (>= 10000 bp)	70856074	109911876	100957344	110743465
Longitud total (>= 25000 bp)	31385964	66773538	55694566	70630264
Longitud total (>= 50000 bp)	13419153	38615642	32071122	42045565
# contigs	31809	27018	25998	12466
Contig más grande	479042	1413394	630416	909350
Longitud total	137196880	163606752	156989699	143330750
GC (%)	64.12	64.91	65	64.48
N50	10497	18485	15793	24335
N90	1770	2470	2546	5051
auN	25184	71624.3	53560.9	86254.9
L50	3065	1841	2127	1224
L90	15662	11067	11491	6278
# N's por 100 kbp	149.57	0	0.04	0
# predicción genes (único)	38166	52684	51786	40415
# genes predichos (>= 0 bp)	160394 + 0	226889 + 0	220883 + 0	197729 + 0
# genes predichos (>= 300 bp)	40729 + 0 part	72082 + 0	72737 + 0	59188 + 0 part
# genes predichos (>= 1500bp)	2763 + 0 part	6428 + 0	6502 + 0 part	6268 + 0 part
# genes predichos (>= 3000 bp)	220 + 0 part	589 + 0 part	621 + 0 part	620 + 0 part

Al comparar el ensamblaje realizado (en adelante, *C. vulgaris* LBCVB) con los genomas de referencia de *C. vulgaris* disponibles en NCBI, se percibe que el genoma ensamblado presenta un tamaño aproximadamente de tres a cuatro veces más,

alcanzando los 143.33 Mb. En cambio, los genomas reportados, como el de *C. vulgaris* 211/11P (40.18 Mb), presentan tamaños entre 37.34 y 48.87 Mb. Asimismo, en contraste con el análisis de Kraken, el tamaño del genoma puede reflejar la presencia de organismos asociados.

Adicionalmente, este resultado puede deberse a diversos factores, entre ellos la presencia de regiones repetidas, las cuales ejercen una gran influencia en la calidad y la contiguidad del ensamblaje, ya que afectan la capacidad del software para reconstruir correctamente el genoma (Schatz et al., 2010). De manera consistente, las microalgas verdes como *C. reinhardtii* y *Volvox carteri* se caracterizan por una alta densidad genómica de repeticiones en tándem en las regiones intrónicas y en regiones de secuencias codificantes (CDS) (Zhao et al., 2013).

Para evaluar aún más la calidad y la integridad del genoma, se realizó un análisis Benchmarking Universal Single-Copy Orthologs (BUSCO) con base en 425 genes ortólogos de referencia del linaje *Chlorophyta*, por lo que este análisis identificó información completa, es decir, completitud para 289 (65.65 %) ortólogos e información fragmentada para 51 (12.00 %) y 95 (22.35 %) faltantes, mientras que se identificaron 27 duplicados, lo que sugiere una calidad intermedia del genoma ensamblado de *novο* para *C. vulgaris* (Figura A2).

En este sentido, se llevó a cabo un análisis de identidad promedio de nucleótidos (ANI, Average Nucleotide Identity) utilizando la herramienta pyANI, mediante el método ANIm, con el objetivo de evaluar la similitud genómica entre *C. vulgaris* y otras especies de microalgas verdes. Los resultados se muestran en la Figura 13, en forma de un mapa de calor con agrupamiento jerárquico, donde los valores representan los porcentajes de identidad promedio entre cada par de genomas. Los bloques en la diagonal principal del gráfico indican grupos de genomas coherentes entre sí, delimitando una especie biológica u otro taxón, lo que constituye una buena aproximación para la diferenciación de especies (Interpreting The Graphical Output — Pyani 0.2.9 Documentation, s. f.).

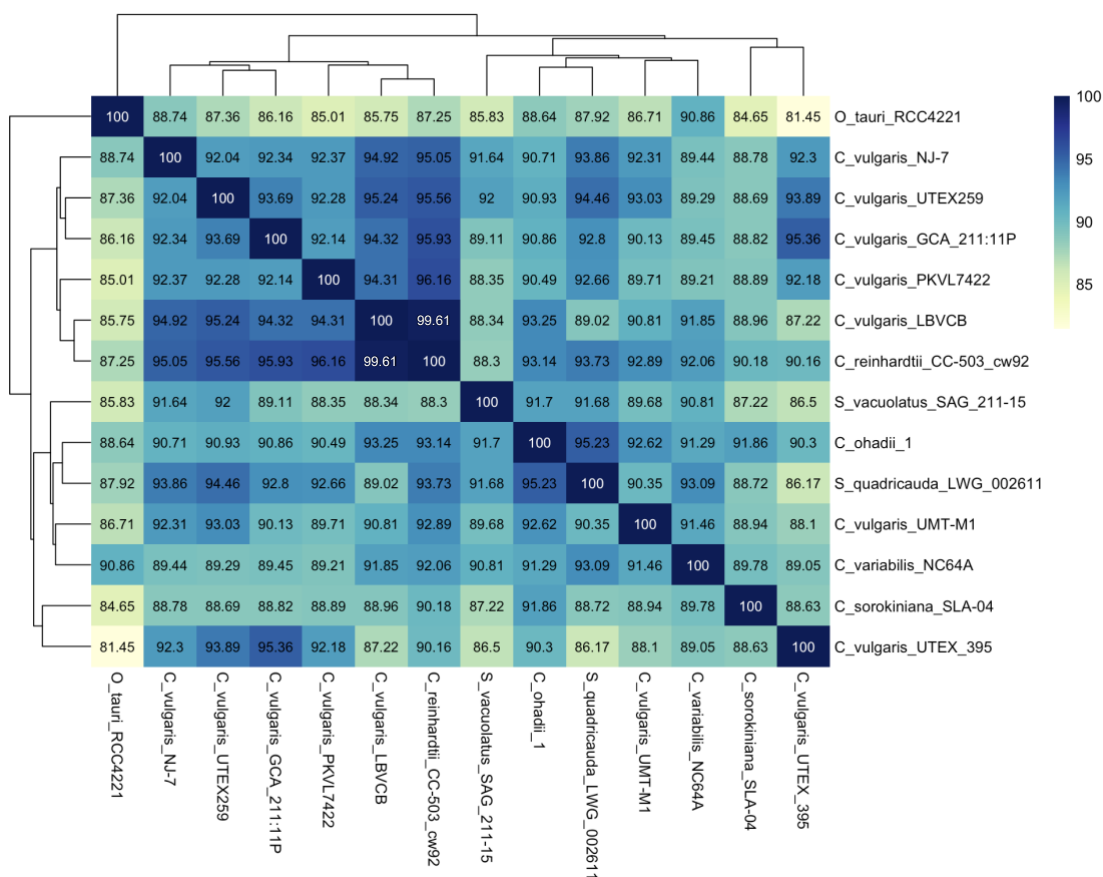


Figura 13. Mapa de calor del porcentaje de identidad promedio (ANI, %) entre genomas de *C. vulgaris* y microalgas relacionadas.

Los valores dentro de cada recuadro representan el porcentaje de identidad nucleotídica promedio entre pares de genomas, mientras que el dendrograma muestra la relación filogenómica entre las cepas analizadas. La escala de colores indica el nivel de similitud genómica con base en el umbral propuesto para la delimitación de especies: tonos azul marino (≥ 95 –100 %) corresponden a genomas pertenecientes a la misma especie, mientras que tonos amarillos (< 85 %) reflejan una divergencia significativa, es decir, especies distintas. (*Interpreting The Graphical Output — Pyani 0.2.9 Documentation*, s. f.).

Asimismo, se observa en el dendrograma asociado al mapa de calor una separación clara entre los genomas pertenecientes al género *Chlorella* y los de otros géneros (*Ostreococcus*, *Scenedesmus*, *Chlamydomonas* y algunas especies del género *Chlorella*). Por otro lado, los valores de ANI dentro del clado de *C. vulgaris* fueron superiores al 94% en comparación con el genoma de estudio (*C. vulgaris* LBCVB), lo que indica una alta similitud genómica y sugiere una relación filogenética muy cercana, posiblemente correspondiente a variantes o subespecies altamente conservadas.

En particular, *C. reinhardtii* mostró un porcentaje relativamente alto (99.61 %) en comparación con las demás cepas de *C. vulgaris*. Esto podría deberse a la conservación de genes esenciales entre ambas especies, que aunque taxonómicamente lejanas, pertenecen al mismo filo *Chlorophyta* dentro de las algas verdes unicelulares. Además, dicho resultado también puede explicarse por la similitud en el tamaño de sus genomas: *C. vulgaris* LBCVB (143.33 Mb) y *C. reinhardtii* (111.1 Mb). En contraste, los valores de ANI entre *C. vulgaris* LBCVB y otras especies cercanas, como *C. sorokiniana* y *C. variabilis*, fueron de aproximadamente 89 %, lo cual indica que pueden pertenecer al mismo género, pero no a la misma especie. Por lo tanto, se deduce que *C. vulgaris* LBCVB puede clasificarse dentro del especie *C. vulgaris* por su alta similitud genómica.

Además del porcentaje de identidad nucleotídica, se evaluó la cobertura de alineamiento entre los genomas comparados para complementar la evaluación de la identidad nucleotídica promedio (ANI). La cobertura se refiere a la proporción del genoma que se alinea a un genoma de referencia durante la comparación entre dos organismos (Interpreting The Graphical Output — Pyani 0.2.9 Documentation, s. f.). En este caso, los valores de cobertura obtenidos a partir del método ANIm (integrado en pyANI) se representan en la Figura 14 mediante un mapa de calor con agrupamiento jerárquico.

Los resultados muestran que los genomas pertenecientes al género *C. vulgaris* presentan altos porcentajes de cobertura entre sí, superiores al 87%, lo que refleja una alta conservación genómica entre las cepas. Por lo tanto, se puede decir que este patrón es comparable con los valores de porcentaje de identidad observados anteriormente, lo que indica que comparten la mayoría de las regiones genómicas alineadas entre sí, característica típica de organismos que pertenecen a la misma especie. Asimismo, las coberturas entre los genomas de *C. vulgaris* LBCVB y los de otros géneros, como *Ostreococcus*, *Scenedesmus* y algunas especies del género *Chlorella*, son menores al 1%, lo que refleja una variación genómica significativa; es decir, son taxonómicamente diferentes a nivel de género y especie. De

igual manera, *C. reinhardtii* presentó un valor alto de cobertura (84.59%), lo que sugiere que una gran parte de su genoma pudo alinearse con *C. vulgaris* LBVCB, sin embargo, de forma opuesta, se obtuvo una cobertura menor (65.12%), lo que puede reflejar un alineamiento asimétrico entre este par de genomas (*Interpreting The Graphical Output — Pyani 0.2.9 Documentation*, s. f.).

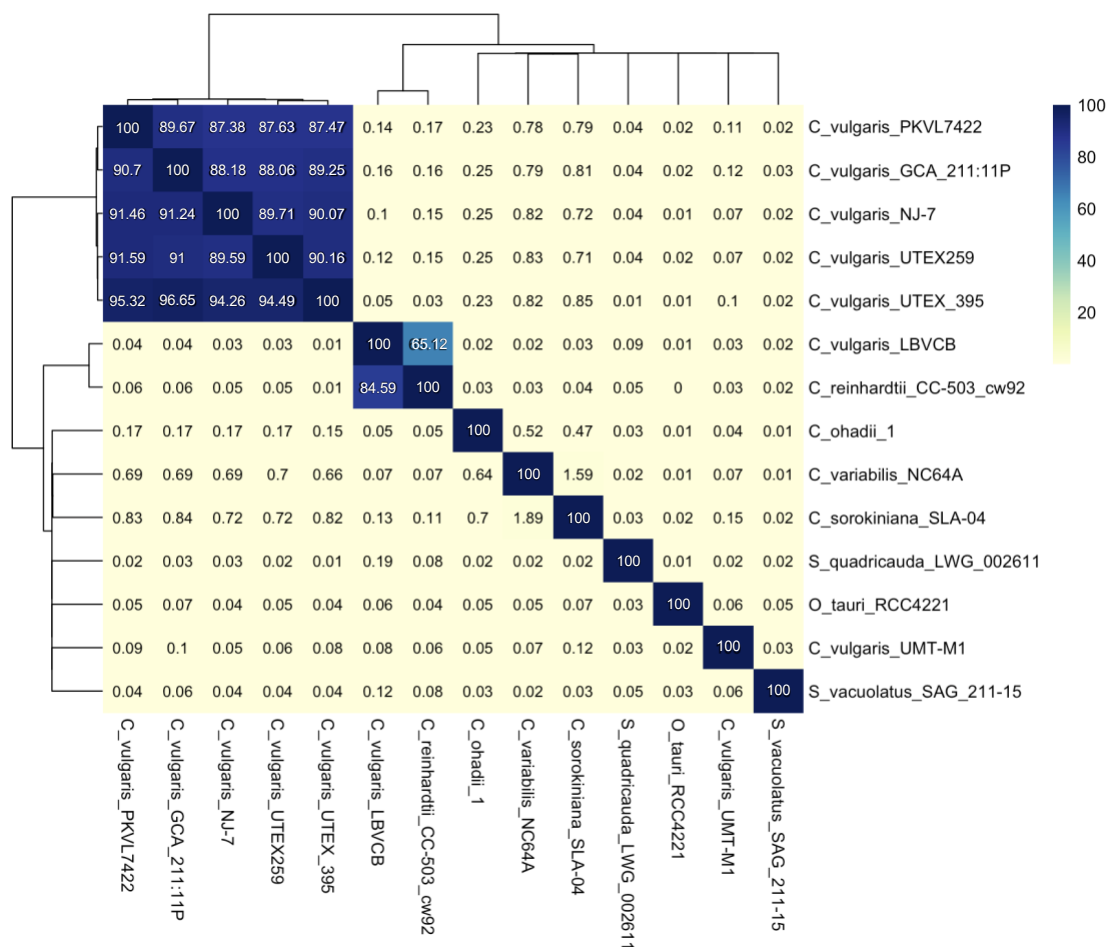


Figura 14. Mapa de calor del porcentaje de cobertura de alineamiento entre genomas de *C. vulgaris* y microalgas relacionadas.

Los valores dentro de cada recuadro representan el porcentaje de cobertura entre pares de genomas, mientras que el dendrograma muestra la relación filogenómica entre las cepas analizadas. La escala de colores indica el nivel de conservación genómica con base en el umbral propuesto para la delimitación de especies: tonos azul marino (> 50 %) corresponden a genomas pertenecientes a la misma especie, mientras que tonos amarillos (< 50 %) reflejan una divergencia significativa, es decir, especies distintas. (*Interpreting The Graphical Output — Pyani 0.2.9 Documentation*, s. f.).

7.3 Anotación y Predicción de Genes

Tras obtener el ensamblaje genómico de *C. vulgaris LBCVB*, se procedió a realizar la etapa de anotación y predicción de genes, con el objetivo de identificar genes funcionales así como caracterizar sus posibles funciones. Para ello se empleó el pipeline de Funannotate, el cual es un software que integra múltiples herramientas de predicción, anotación y comparación genómica, incluyendo para este caso programas de predicción *ab initio* como GeneMark-ES/ET, Augustus y NCBI-BLAST. Adicionalmente, se utilizan Diamond y Exonerate para alinear evidencia de proteínas con bases de datos (por defecto, UniProtKB/SwissProt) y mapear proteínas al genoma (Gene Prediction — Funannotate 1.8.16 Documentation, s. f.).

La anotación del genoma identificó un total de 47,825 genes, de los cuales 47,192 son codificantes, es decir, aquellas proteínas con función predicha basada en bases de datos, de acuerdo con la Figura 15. El número de genes codificantes para proteínas es significativamente mayor en comparación con el genoma de referencia de *C. vulgaris 211/11P*, donde se predijeron 10,724 genes (Cecchin et al., 2019). En concordancia con lo anterior, más del 83.92 % de los genes codificantes identificados en *C. vulgaris LBCVB* no tuvieron un equivalente detectado en el genoma de referencia *C. vulgaris 211/11P* al realizar una comparación por BLASTp con valor $E \leq 1e-5$.

Por otro lado, se detectaron 47,139 proteínas hipotéticas, las cuales son proteínas que se predice que se expresarán a partir de un marco de lectura (ORF), pero para las cuales no hay evidencia experimental de traducción (Ijaq et al., 2015). En este sentido, solo se identificaron 2,687 con función asignada.

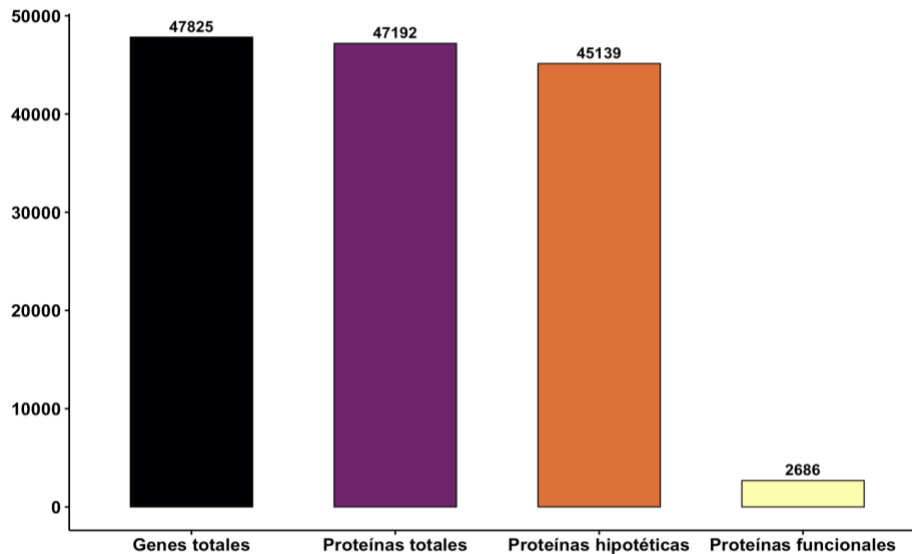


Figura 15. Distribución de genes y proteínas anotados en el genoma de *C. vulgaris* LBCVB.

Se evaluó la anotación funcional del genoma basada en ontologías mediante la herramienta EggNOG-mapper. El reporte informó que, de un total de 41,025 genes, el 92.39 % (37,904) está asociado a términos de grupos de genes ortólogos (COG), mientras que el 9.95 % (4,081) está asociado a términos de ontología génica (GO) y el 33.62 % (13,795) presenta rutas metabólicas asociadas (Figura 16). En este sentido, la mayoría de los genes fueron anotados en la categoría COG, esto se debe principalmente a que esta base de datos asigna funciones a nivel de genes ortólogos, los cuales son genes que se han originado a partir de un solo gen ancestral, pero que se encuentran en distintas especies, lo que permite clasificar un mayor número de proteínas, incluso algunas con función desconocida, como las proteínas hipotéticas. Por otro lado, las bases de datos GO y KEGG son más estrictas, ya que requieren evidencia funcional o asignación a rutas metabólicas específicas, por lo que generalmente anotan un menor número de genes.

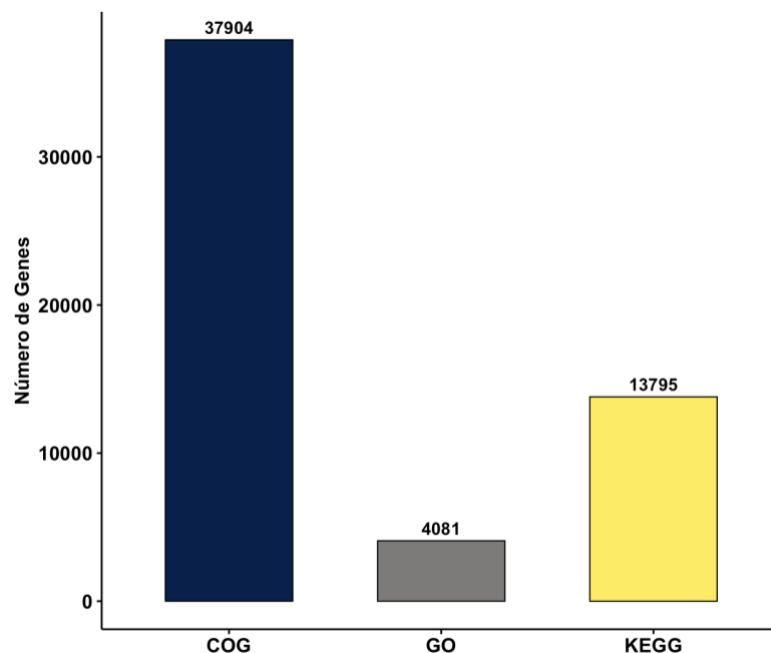


Figura 16. Número de genes anotados por categoría funcional. Ontología Génica (GO); COG (Grupos de genes ortólogos); KEGG (Enciclopedia de Kioto de Genes y Genomas).

Como parte de la anotación funcional, las proteínas predichas se clasificaron de acuerdo con el sistema COG, el cual agrupa genes ortólogos en categorías funcionales bien definidas. Del total de 41,025 genes, el grupo más grande pertenece a genes con “Función desconocida” (8315, 20.27%), lo que sugiere que gran parte de los genes sí tiene anotación, sin embargo, no se conoce su función biológica. En contraste, el 7.76% (3285) se clasificó como genes implicados en la “Transducción de señal”, seguido de genes “No asignados” (es decir, que no recibieron ninguna anotación) (3121, 7.61%), “Transporte y metabolismo de aminoácidos” (3005, 7.32%), “Transcripción” (2994, 7.30%), así como también el “transporte y metabolismo de carbohidratos” (2445, 5.96%).

Por consiguiente, en una proporción similar de 5.59% a 4.31% (2295 a 1769), se clasificaron las funciones de “Replicación, recombinación y reparación”, “Modificación postraduccional”, “Producción y conservación de energía”, “Transporte y metabolismo de iones inorgánicos”, “Producción y conversión de energía”, “Traducción/ Biogénesis ribosomal”, “Transporte y metabolismo de lípidos” y

“Biogénesis de la membrana”. Por último, las siguientes categorías contienen el menor número de genes asignados: “Biosíntesis, transporte y catabolismo de metabolitos secundarios” (1363, 3.32%), “Transporte y metabolismo de coenzimas” (1294, 3.15%), “Tráfico intracelular, secreción y transporte vesicular” (1198, 2.92%), “Transporte y metabolismo de nucleótidos” (856, 2.09%), “Control del ciclo celular” (627, 1.53%), “Procesamiento y modificación del ARN” (590, 1.44%), “Mecanismos de defensa” (559, 1.36%), “Motilidad celular” (383, 0.93%), “Citoesqueleto” (320, 0.78%), “Estructura y dinámica de la cromatina” (290, 0.71%), “Estructura nuclear” (78, 0.19%) y “Estructuras extracelulares” (31, 0.08%).

Li et al. (2016) en *C. sorokiniana* reportan una proporción similar en cuanto a algunas categorías, aunque realizaron la asignación a nivel contigs. En ambos casos, las categorías más abundantes en común son “Replicación, recombinación y reparación” (1251, 19.22 %), “Transcripción” (1000, 8.07 %), “Transporte y metabolismo de aminoácidos” (836, 9.34 %), “Modificación postraduccional” (829, 9.26 %) y “Transducción de señal” (842, 9.42 %) . Adicionalmente, en la categoría “Predicción de la función general” (2380, 19.22 %) mostraron la mayor proporción de contigs asignados, mientras que en este estudio no se clasificaron genes en esa categoría, sin embargo, esto puede indicar una relación en algunos grupos de genes funcionales principales en el genoma del clado *Chlorella*.

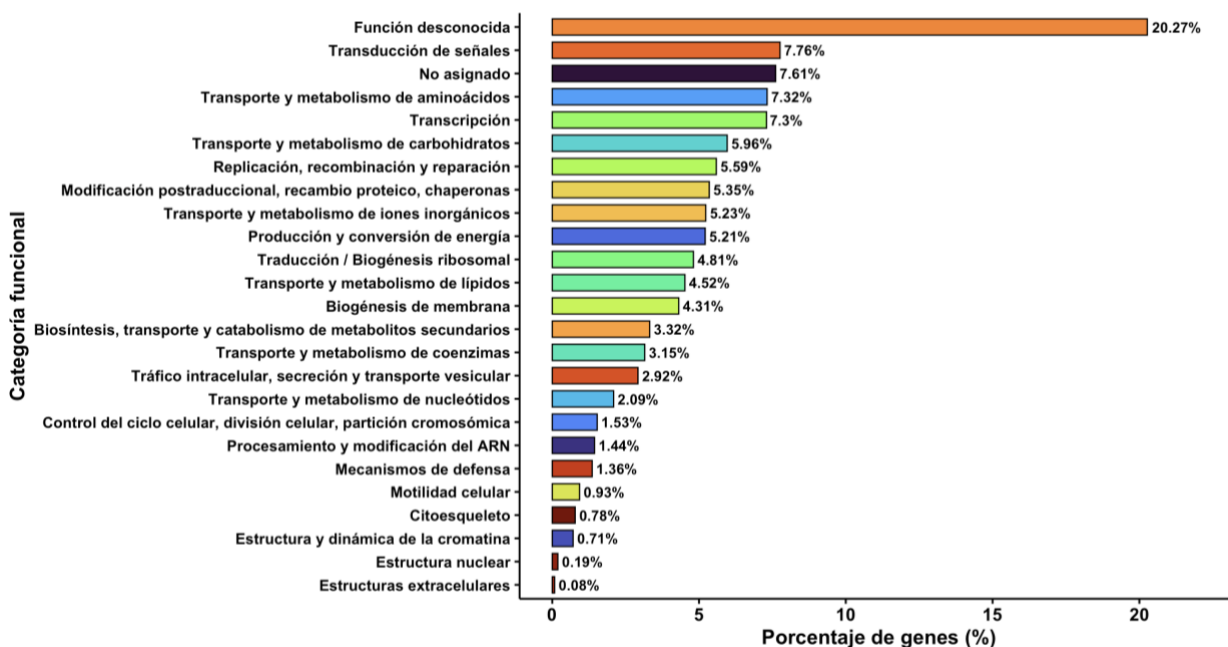


Figura 17. Clasificación funcional de genes por categorías COG. Distribución porcentual de los genes anotados en cada categoría funcional COG.

El análisis de anotación funcional mediante las categorías COG y GO permitió identificar distintos grupos de genes implicados en diversos procesos metabólicos, moleculares y estructurales, así como su organización funcional en categorías biológicas como transporte, transcripción, señalización, etc. Asimismo, este análisis se complementó mediante la herramienta BLASTp (Basic Local Alignment Search Tool), empleando bases de datos personalizadas construidas a partir de proteínas previamente reportadas en procesos de asimilación y transformación de contaminantes en otras microalgas del filo *Chlorophyta* (Género *Chlorella* y *C. reinhardtii*). Para ello, se utilizó como referencia la base de datos UniProt, lo que permitió la identificación específica de genes potencialmente involucrados en la asimilación de contaminantes en agua.

Tabla 6. Genes y proteínas homólogas identificadas mediante BLASTp en microalgas verdes relacionados con metales pesados.

<i>C. reinhardtii</i>						
ID	Identidad (%)	Cobertura (%)	Valor E	Proteína	Gen	Función
	100	100	0			Reparación del ADN (en procesos como la replicación, recombinación y reparación por escisión de bases). Unión de iones metálicos.
	100	100	0			Desintoxicación de iones metálicos (Cu o Cd). Biosíntesis de fitoquelatinas.
	100	100	0			Transporte, detoxificación y tolerancia a iones metálicos.
	100	100	0			Transportador de iones metálicos divalentes (Mn ²⁺ Fe ²⁺ y Cd ²⁺).
	100	100	0			Trasnsporte de iones de hierro.
	100	100	0			Biosíntesis de aminoácidos y aminoácidos aromáticos.
	100	100	0			Metilación
<i>Chlorella</i>						
	92.42	100	0			Cataliza la fijación de CO ₂ en el ciclo de Calvin, transformando ribulosa-1,5-bisfosfato (RuBP) en dos moléculas de 3-fosfoglicerato, primer paso en la síntesis de azúcares durante la fotosíntesis.
	95.146	100	.05E-152			Detoxificación de cianuro y metabolismo del azufre.
	98.883	94	.71E-114			Síntesis de ATP durante la fotosíntesis, al permitir el flujo de protones a través de la membrana tilacoidal.
<i>Chlorella</i>						

ID	Identidad (%)	Cobertura (%)	Valor E	Proteína	Gen	Función
	93.141	86	0			Desintoxicación de iones metálicos (Cu o Cd). Biosíntesis de fitoquelatinas.
	83.79	99	0			Unión de clorofila. Transporte de electrones en el fotosistema II.
	95.59	100	0			Actividad del transportador transmembrana de iones de Cd, Mg y Fe.
	87.50	82	0			En el proceso de traducción, se une el ARNr.
	86.64	100	0			Metilación. Unión de iones zinc.
<i>C. vulgaris</i>						
	68.263	98	0			En la metilación del arsenito (As(III)), un proceso de desintoxicación que convierte esta forma tóxica de arsénico en una forma menos peligrosa, el monometilarsénico (MMA(III)).
	69.10	97	0			Biosíntesis de aminoácidos y aminoácidos aromáticos.
	68.23	91	0			Reparación del ADN (en procesos como la replicación, recombinación y reparación por escisión de bases). Unión de iones metálicos.
	61.48	84	0			Desintoxicación de iones metálicos (Cu o Cd). Biosíntesis de fitoquelatinas.

Fuente: Elaboración propia a partir del BLASTp y información obtenida de Uniprot (UniProt, s. f.).

En la Tabla 6 se presentan los resultados del análisis BLASTp realizado contra la base de datos de proteínas relacionadas con la asimilación de metales pesados. Primeramente, se seleccionaron los alineamientos con las proteínas de *C. reinhardtii*, que mostraron coincidencias del 100% de identidad, 100% de cobertura y un valor E menor a 1×10^{-5} , con el fin de realizar una asignación funcional confiable. Bajo estas condiciones se identificaron siete proteínas de *C. reinhardtii*. Entre las más relevantes se encuentran *Ynf1p*, involucrada en la reparación del ADN bajo estrés metálico, una glutatión transferasa asociada a la detoxificación de iones como Cu y Cd, además de participar en la biosíntesis de fitoquelatinas, y *Ynf12p*, la cual transfiere grupos metilo ($-CH_3$) de un donador a un aceptor que puede ser ADN, proteínas u otras moléculas. De igual manera, se identificaron algunos transportadores, como *Ynf17p*, el transportador de iones Ca^{2+} y el transportador *Ynf18p*.

De manera complementaria, se realizó el mismo análisis con las proteínas de microalgas del género *Chlorella*. En este caso, se eligieron aquellas con un porcentaje mayor al 95% de identidad y cobertura, y con un valor E menor a 1×10^{-5} . Bajo estos criterios se detectaron diversos componentes, entre ellos *Ynf1p*, la cual cataliza dos reacciones: *Ynf1p* (primera etapa en la fijación de dióxido de carbono) y la fragmentación oxidativa del sustrato de pentosa durante la fotorrespiración. Asimismo, se identificó *Ynf12p*, localizada en el cloroplasto y encargada de la síntesis de ATP durante la fotosíntesis, *Ynf17p*, asociado al centro de reacción y vinculado con la generación de ATP a partir de ADP en presencia de un gradiente de protones a través de la membrana, y el

, perteneciente a la subunidad pequeña e involucrado en el proceso de traducción. También se registró el y una , la cual contribuye a la detoxificación de cianuro y participa en el metabolismo del azufre.

Posteriormente, se aplicó el mismo procedimiento para la alineación de proteínas correspondientes a *C. vulgaris*, utilizando como criterios una identidad y cobertura mayores al 60% y un valor $E < 1 \times 10^{-5}$. Si bien no se encontraron genes asociados a estos procesos, sí se detectaron proteínas que participan en la detoxificación y el metabolismo de compuestos metálicos. Entre los polipéptidos se identificaron una , involucrada en la metilación del arsénico (As(III)), un proceso de desintoxicación que convierte esta forma tóxica en compuestos menos peligrosos, como el monometilarsénico. Asimismo, se registró la enzima , que participa en la biosíntesis de aminoácidos y otros compuestos aromáticos. Adicionalmente, se

En conjunto, los resultados sugieren que la presencia de genes como en más de un género indica una conservación de los mecanismos de homeostasis y detoxificación en el filo *Chlorophyta*. Cabe destacar que, de acuerdo con los resultados obtenidos, una parte importante de los genes y proteínas identificados corresponde a transportadores, lo cual es de esperarse, ya que representan la primera línea de defensa frente al desequilibrio en la homeostasis de metales a nivel celular y subcelular. Cuando las reservas metálicas disminuyen, estos transportadores permiten la captación y distribución de cofactores necesarios para evitar síntomas de deficiencia, por el contrario, cuando los metales se acumulan por encima de la capacidad celular, facilitan su expulsión para prevenir

efectos tóxicos (Blaby-Haas & Merchant, 2012). Por lo tanto, los resultados demuestran que *C. vulgaris* podría presentar una mayor disposición para el procesamiento principal de Fe, Cd y As asociado con la presencia de genes como

En contraste, Qian et al. (2009) reportan que el Cu (0.5 y 1.5 mM) y el Cd (1.0 y 2.0 mM) pueden provocar una disminución significativa en el crecimiento de *C. vulgaris*, además, pueden aumentar la formación de ROS y de clorofila de forma sinérgica. Asimismo, observaron que ambos metales inhiben de forma independiente la expresión de los genes *psbA* y *rbcL*, mientras que estimulan de forma adaptativa la expresión del gen *psbB*, el cual forma parte del operón *psaA/B* del genoma del cloroplasto y codifica las apoproteínas P700 clorofila A2, relacionadas con el Fotosistema I (PSI).

Bajo las mismas condiciones metodológicas, se realizó la identificación de proteínas que podrían intervenir en la asimilación, transformación o degradación de compuestos farmacéuticos presentes en el agua. Por consiguiente, el BLASTp procesado contra la base de datos de proteínas correlacionadas con estos contaminantes, se extrajeron las coincidencias con proteínas de *C. reinhardtii*, las cuales presentaron alineaciones con 100% de identidad, 100% de cobertura y un valor E menor a 1×10^{-5} , con el objetivo de realizar una anotación funcional precisa. Del análisis se caracterizaron 5 proteínas validadas por Uniprot, codificadas por los genes

De igual manera se encontró un transportador implicado en la detoxificación.

participa como catalizador en la hidrólisis de pirofosfato inorgánico (PPi) en dos fosfatos así como también en mantener el equilibrio de fosfato en el cloroplasto, mientras que se relaciona con la respuesta al estrés oxidativo. Por otro lado, interviene en la síntesis y modificación de tRNA y en la traducción mitocondrial. En cuanto al está involucrado en el paso de protones y la síntesis de ATP en la mitocondria, en cambio, participa en la síntesis de clorofila.

Tabla 7. Genes y proteínas homólogas identificadas mediante BLASTp en microalgas verdes asociados a medicamentos.

<i>C. reinhardtii</i>						
ID	Identidad (%)	Cobertura (%)	Valor E	Proteína	Gen	Función
	100	100	0			Cataliza la hidrólisis de pirofosfato inorgánico (PPi) en dos fosfatos inorgánicos. Equilibrio de fosfato en el cloroplasto.
	100	100	0			Respuesta al estrés oxidativo.
	100	100	0			Enzima dependiente de ATP que participa en la síntesis/modificación de tRNA y en la traducción mitocondrial.
	100	100	0			Subunidad de la ATP sintasa mitocondrial involucrada en el paso de protones y la síntesis de ATP.
	100	100	0			Oxigenasa dependiente de hierro que participa en la síntesis de clorofila.
	100	100	0			Transportador transmembrana involucrado en la exportación y desintoxicación de xenobióticos.
<i>Chlorella</i>						
	96.59	100	0		-	Proceso metabólico de compuestos que contiene fosfatos.

Chlorella						
ID	Identidad (%)	Cobertura (%)	Valor E	Proteína	Gen	Función
	93.36	95	0			Componente del fotosistema II involucrado en la transferencia de electrones y la producción de oxígeno.
	94.92	100	0			Transportador transmembrana de xenobióticos para su desintoxicación.
	95.27	93	0			Enzima que fija CO ₂ en el ciclo de Calvin y participa en el primer paso de la síntesis de azúcares.
	98.95	100	0			Biosíntesis del grupo hemo y clorofila.
C. vulgaris						
	94.52	100	0			Enzima que fija CO ₂ en el ciclo de Calvin y participa en el primer paso de la síntesis de azúcares.
	85.80	99	0			Componente del fotosistema II involucrado en la transferencia de electrones y la producción de oxígeno.
	80.52	83	0			Biosíntesis de ácidos grasos y acetyl-CoA.

Fuente: Elaboración propia a partir del BLASTp y información obtenida de Uniprot (UniProt, s. f.).

En este mismo sentido, el análisis comparativo con los polipéptidos de microalgas de género *Chlorella*, y tras aplicar un cribado con un margen mínimo del 70 % de identidad y cobertura, así como también un valor E menor a 1×10^{-5} , se obtuvo como resultado los genes verificados psbB y RbcL, además de proteínas como la enzima difosfatasa inorgánica y un transportados de xenobióticos. De igual manera se percibió una tendencia similar en la evaluación con proteínas de la misma especie ya que de forma análoga se identificaron , además de con los mismos criterios de filtrado del análisis

anterior(es decir, alineamiento con identidad y cobertura del 70% asimismo un valor E menor a 1×10^{-5}).

, que codifica para la subunidad CP47, está involucrado en la transferencia de electrones y la producción de oxígeno en el fotosistema II (PSII). Por otro lado, el corresponde a la gran subunidad de RuBisCO, implicada en el proceso de fotorrespiración. La exposición a largo plazo de microalgas a altas concentraciones de productos farmacéuticos y de cuidado personal generalmente resulta en una disminución en el contenido de proteínas y clorofilas debido a la síntesis de ROS y a alteraciones en la regulación de los genes fotosintéticos (*rbcL*, *psbA*, *psbB* y *psbC*). Sin embargo, la respuesta positiva o negativa a la interacción con compuestos farmacéuticos es específica de cada cepa (Bai et al., 2025). Como en el caso de la cianobacteria *Microcystis aeruginosa*, la presencia de los antibióticos espiramicina y amoxicilina aumentó significativamente la expresión de los genes relacionados con la fotosíntesis (*psbA*, *psbB*, *rbcL* y *mcyB*) (Xiong et al., 2017).

En este contexto, Wang et al. (2019) reportaron en *C. vulgaris* que la combinación de eritromicina y enrofloxacin incrementó de forma sinérgica la abundancia de los transcritos de *psaB* y *chlB*, siendo este aumento significativamente mayor que el observado en los tratamientos individuales. Asimismo, el análisis detallado mostró que, en el caso de *psbC*, el aumento de transcritos se debió principalmente a la enrofloxacin aplicada de forma individual.

De igual manera, el estudio de Xiong et al. (2016) describe que, en *C. vulgaris*, el aumento de la salinidad promovió significativamente la bioacumulación de levofloxacin y su posterior degradación en productos más simples y menos tóxicos por el sistema enzimático intracelular. Se observó un aumento significativo en la cantidad de una proteína similar a la ferroxidasa bajo el estrés salino, esta proteína dirige electrones hacia la hidroxilación catalizada por el citocromo P450, lo que mejoró la degradación de contaminantes orgánicos dentro de la célula.

Nuevamente, utilizando la misma estrategia de análisis, se comparó la base de datos exclusiva de componentes inorgánicos (P y N) con el proteoma anotado por

EggNOG-mapper. El resultado fue filtrado bajo los criterios de identidad y cobertura del 100 % y un valor E menor a 1×10^{-5} para el caso de *C. reinhardtii*. Como era de esperarse, se obtuvo una mayor cantidad de alineamientos asociados a proteínas con genes validados, dado que *C. reinhardtii* es considerado un organismo modelo. Entre los principales genes identificados, en términos funcionales, se encontraron diversos transportadores y enzimas catalíticas.

Entre ellos destacan los *NRT2*, pertenecientes a la familia AMT1 (González-Ballester et al., 2004), así como las proteínas transportadoras *NRT2*, que corresponden a la familia NNP (por sus siglas en inglés, Nitrate Nitrite Porter), la cual forma parte, a su vez, de la superfamilia MFS (Major Facilitator Superfamily). Los polipéptidos NRT2 constituyen un sistema transportador biespecífico de alta afinidad para nitrato/nitrito. En particular, NRT2.1 y NRT2.2 forman parte de un sistema de dos componentes que requiere una segunda proteína, NAR2, para ser completamente funcional en *C. reinhardtii* y en otras algas como *Chlorella NC64A* (Sanz-Luque et al., 2015). Por otro lado, *NAR2* está implicado en el transporte de molibdato dependiente de energía, el cual es fundamental para una asimilación adecuada de nitrato. Igualmente, se detectó el *NAR2*, el cual cataliza la reacción reductora de amonio y α -oxoglutarato para producir glutamato, especialmente bajo ciertas condiciones de estrés. *NAR2* codifican enzimas clave en el metabolismo del nitrógeno: *NAR2* cataliza la síntesis de glutamina a partir de glutamato y amonio para su reasimilación en el citoplasma, mientras que *NAR2* cataliza la síntesis de glutamina en el cloroplasto. Por último, AS (Arilsulfatasa) está implicada en el metabolismo de compuestos.

Para los alineamientos con proteínas de algas del género *Chlorella*, se identificaron los genes *NAR2* y *AS*, los cuales pertenecen a la microalga *Nannochloris sp.* (desiccata). Además, se encontraron otras proteínas como la subunidad alfa del coatómero y una glutamina

sintetasa. El umbral de selección empleado fue del 70 % de identidad y cobertura, con un valor E menor a 1×10^{-5} . Complementariamente, en el análisis comparativo de secuencias de *C. vulgaris* no se asignaron genes específicos, sin embargo, sí se identificaron diversas proteínas, entre ellas: *psbA*, involucrada en el proceso biosintético de polifosfatos, una *psbA*, asociada al metabolismo de ciertos aminoácidos; la enzima fosfoglicerato mutasa, así como *psbA*, previamente asociadas a los genes *psbA* presentes en *Nannochloris* sp. (desiccata). La presencia de estas proteínas en *C. vulgaris* y su correspondencia con genes identificados en *Nannochloris* sp. sugiere un grado de conservación funcional dentro del grupo de microalgas verdes (*Chlorellales*).

Tabla 8. Genes y proteínas homólogas identificadas mediante BLASTp en microalgas verdes vinculadas al transporte de N y P.

<i>C. reinhardtii</i>						
ID	Identidad (%)	Cobertura (%)	Valor E	Proteína	Gen	Función
	99.80		0			Transporte de amonio.
	98.86	100	0			Transporte de amonio.
	94.60	100	0			Participa en la degradación del glutamato mediante la actividad de la enzima glutamato deshidrogenasa (NAD+).
	100	100	0			Transportador de molibdato de alta afinidad, a través de un proceso dependiente de la energía.
	100	100	0			Cataliza la síntesis de glutamina a partir de glutamato y amonio (Asimilación y reciclaje del nitrógeno).
	100	96	0			Transportador transmembrana y asimilación de nitratos.

C. reinhardtii						
ID	Identidad (%)	Cobertura (%)	Valor E	Proteína	Gen	Función
	100	100	0			Transportador transmembrana y asimilación de nitratos.
	100	100	0			Cataliza la síntesis de glutamina y participa en la reasimilación de amonio en el cloroplasto.
	90.14	95	0			Enzima implicada en el metabolismo de compuestos fenólicos.
Chlorella						
	86.09	97	0			Transporte de proteínas biosintéticas desde el ER, a través de Golgi hasta la red trans Golgi.
	76.04	99	0			La IP6K convierte IP6 en IP7 y participa en la regulación de diversos procesos celulares.
	71.20	98	0			Asimilación de amoniaco. Biosíntesis del glutamato.
	71.34	100	0			Metabolismo del nitrógeno y del glutamato.
C. vulgaris						
	78.75	78	0			Proceso biosintético de polifosfato.
	78.09	78	0			Proceso biosintético de la glutamina.
	73.17	99	0			Proceso metabólico de asparagina, glutamina y oxalacetato.
	72.91	99	0			La IP6K convierte IP6 en IP7 y participa en la regulación de diversos procesos celulares.
	70.09	94	0			Cataliza la conversión entre intermediarios de la glucólisis.

Fuente: Elaboración propia a partir del BLASTp y información obtenida de Uniprot (UniProt, s. f.).

7.4 Rutas metabólicas identificadas en *C. vulgaris*

En esta sección se presentan las rutas metabólicas encontradas a partir de la anotación funcional del genoma de *C. vulgaris LBCVB*, con énfasis en aquellas relacionadas a la degradación de los agentes contaminantes estudiados. Se utilizó la base de datos Enciclopedia de Kyoto de Genes y Genomas (KEGG) para la reconstrucción de las rutas a partir de la anotación realizada por eggNOG-mapper, asignando cada gen a su correspondiente identificador único ortólogo (KO). Se evaluó la presencia de genes clave en rutas de interés relacionados a metales pesados, fármacos y compuestos inorgánicos (N y P).

Se identificaron un total de 9 rutas metabólicas en *C. vulgaris LBCVB*, 5 relacionadas con el metabolismo de metales pesados, 5 asociadas a fármacos y 3 implicadas en la asimilación de P y N, las cuales, al mismo tiempo, muestran una fuerte correlación funcional entre sí. En la Tabla 9 se reporta el conjunto de rutas esenciales para llevar a cabo el proceso de biorremediación en *C. vulgaris LBCVB*, que incluyen: Carcinogénesis química (especies reactivas de oxígeno), Fotosíntesis, Fijación de carbono mediante el ciclo de Calvin, Metabolismo de fármacos (citocromo P450), Metabolismo de fármacos (otras enzimas), Transportadores ABC, Metabolismo del nitrógeno, Vía de las pentosas fosfato y Replicación del ADN.

Tabla 9. Vías metabólicas identificadas en *C. vulgaris* LBCVB relacionadas con la respuesta a contaminantes en agua.

No	ID	Nombre	Genes extraídos	Contaminante	Figura
1	05208	Carcinogénesis química: Especies reactivas de oxígeno		Metales pesados	A3
2	00195	Fotosíntesis		Fármacos; Metales pesados	A4
3	00710	Fijación de carbono mediante el ciclo de Calvin		Fármacos; Metales pesados	A5
4	00982	Metabolismo de fármacos - citocromo P450		Fármacos	A6
5	00983	Metabolismo de fármacos: otras enzimas		Fármacos	A7

No	ID	Nombre	Genes extraídos	Contaminante	Figura
6	02010	Transportadores ABC		Metales pesados; Fármacos; Compuestos inorganicos	A8
7	00910	Metabolismo Nitrogeno		Nitrogeno	A9
8	00030	Vía de las pentosas fosfato		Fósforo	A10
9	03030	Replicación del ADN		Metales pesados	A11

Nota: Los genes marcados en color rojo son aquellos obtenidos a partir del análisis BLASTp.

Diversos estudios señalan que la fotosíntesis es una de los procesos más susceptibles a alteraciones durante la fitorremediación de diversos contaminantes mediante microalgas, debido a su estrecha relación con el estrés celular, el crecimiento y disponibilidad de energía. Una gran parte de los genes que codifican subunidades de los complejos de membrana y transportadores de electrones asociados a la fase lumínica de la fotosíntesis se encuentra en el genoma nuclear o en el cloroplasto de *C. vulgaris* (Cecchin et al., 2019). Del fotosistema II, se identificaron

mientras que en el fotosistema I, se detectaron la mayoría de los genes .

Cabe destacar que en *C. reinhardtii*, estudios como Mehra et al. (2024) han reportado la presencia de subunidades adicionales en el fotosistema II como: psbA , psbC, psbD además de psbB, lo que podría reflejar una mayor complejidad estructural y funcional en comparación con *C. vulgaris* LBCVB. Para el complejo citocromo b6/f solo se identificaron las subunidades centrales , mientras que en el conjunto proteico encargado del transporte de electrones en el proceso fotosintético,

. Por lo tanto, el transporte de electrones también puede verse afectado bajo ciertas condiciones de estrés. Ghariani et al. (2025) reportan que el complejo citocromo b6/f, encargado del acoplamiento entre los fotosistemas y de la generación del gradiente de protones en la fotosíntesis, presenta una regulación negativa del gen *petB* que codifica el citocromo b (N-terminal)/b6 al exponer a *Chlamydomonas sp.* a Ben-Ag (bentonita intercambiada con plata).

A su vez, la fotosíntesis está estrechamente relacionada con el ciclo de Calvin, ya que el ATP y el NADPH producidos durante el proceso fotosintético son necesarios

para impulsar la síntesis de gliceraldehído-3-fosfato (G3P), el cual se utiliza en la producción de azúcares, lípidos y aminoácidos. Se recuperó un conjunto completo de genes que codifican para las enzimas del ciclo de Calvin en *C. vulgaris LBCVB*. Entre ellos se identificó principalmente el gen *rbcL*, que codifica la subunidad grande de la RuBisCO, encargada de la fijación de CO₂ en la fotosíntesis (Figura A4). El *rbcL* es altamente conservado en organismos fotosintéticos, por lo que ha sido ampliamente utilizado como marcador molecular en diversos estudios (Bai et al., 2025) lo que contribuye a corroborar que la información recuperada independientemente del tamaño del ensamble pertenece a un organismo fotosintético.

El metabolismo del nitrógeno es una de las rutas mejor descritas, sobre todo en *C. reinhardtii*. Para el caso de *C. vulgaris LBCVB*, entre los genes involucrados en esta ruta se encuentran varios sistemas para captar nitrato y nitrito extracelular, como

, enzimas reductoras como

utilizados para la incorporación de NH₄⁺ en glutamina y glutamato. Asimismo, se recuperaron genes relacionados con la fijación de nitrógeno

a NH₄⁺, además de otras proteínas involucradas,

, entre otras (Figura a9).

La vía de carcinogénesis química: especies reactivas de oxígeno implica genes relacionados con la generación, detoxificación y señalización celular frente al estrés oxidativo. En *C. vulgaris LBCVB* se identificaron múltiples elementos que componen esta ruta, incluyendo los complejos mitocondriales , enzimas antioxidantes como , además de enzimas detoxificadoras (oxidorreductasas) como

y elementos cinasa claves en la señalización

Esta es

una de las vías principales que a menudo son inducidas por la exposición a metales

pesados y fármacos. Dependiendo del tipo de contaminante, las microalgas pueden generar radicales hidroxilo, anión superóxido, peróxido de hidrógeno y oxígeno singlete. Cuando la producción de ROS excede la capacidad celular para manejarlos, la célula se deteriora. Sin embargo, generalmente los contaminantes como los fármacos no exhiben una inhibición letal de las actividades de la microalga en sus concentraciones ambientales relativas. *C. reinhardtii* utiliza el oxígeno singlete como molécula de señalización para regular la fotoprotección en cambio, *C. sorokiniana* experimenta una remodelación tilacoidal con regulación a la baja de las enzimas involucradas en la eliminación de ROS (Bai et al., 2025).

En cambio, las vías metabólicas de fármacos (citocromo P450 y otras enzimas) exhiben la presencia de monooxigenasas y diversas deshidrogenasas encargadas de la oxidación inicial de xenobióticos. La segunda vía reúne enzimas de conjugación y detoxificación, carboxilesterasas y catalasas. En conjunto, estos genes reflejan un amplio sistema de biotransformación que permite a las microalgas modificar, neutralizar o eliminar compuestos tóxicos, como el fármaco azatioprina, relacionado estructuralmente con 6-mercaptopurina, y medicamentos como fluoruracilo. Estudios realizados sugieren que microalgas verdes como *N. oculata*, al ajustar las concentraciones de N y C en el medio, resulta en una alta capacidad de remoción del fármaco diclofenaco (alrededor del 68%), lo que indica que la absorción celular es mecanismo principal (Sánchez-Sandoval et al., 2021).

La ruta KEGG 02010 agrupa a los transportadores ABC, los cuales constituyen un conjunto amplio de proteínas de membrana impulsadas por ATP mediante la importación y exportación de una gran variedad de sustratos. En *C. vulgaris* *LBCVB* se identificó una gran diversidad de genes ABC, incluidos los sistemas transportadores

de iones minerales y orgánicos, así como también exportadores de compuestos tóxicos y xenobióticos (entre otros) (Figura A8).

En este sentido, en la Tabla 9 también se observa la vía de las pentosas fosfato, conformada principalmente por enzimas clave como además de transaldolasas y transquetolasas, esenciales en la producción de NADPH, indispensable para el transporte, síntesis de lípidos y la asimilación de nitrógeno (Figura a10).

Por último, el proceso de replicación del ADN, que alberga una gran variedad de enzimas y proteínas encargadas de duplicar el material genético durante la división celular, también se ve afectado por el exceso de ROS inducido por la presencia de contaminantes, lo que causa un daño severo a los componentes celulares como lípidos, proteínas y ADN. Esto provoca un aumento en la tasa de mutagénesis, que puede conducir a la muerte celular programada (Xiong et al., 2017).

8. CONCLUSIONES

- ◇ La extracción de ADN genómico de *C. vulgaris* fue eficiente, obteniendo ADN de alta pureza ($A_{260}/A_{280}=1.98$) y concentración (49.4 ng/mL).
- ◇ Se logró el ensamblaje del genoma a nivel de contigs (143.33 Mb) utilizando la herramienta Unicycler.
- ◇ El análisis genómico reveló una gran proporción de proteínas hipotéticas, lo que sugiere la presencia de numerosas secuencias aún no caracterizadas funcionalmente.
- ◇ El análisis COG mostró una alta proporción de genes con función desconocida, junto con funciones clave relacionadas con el metabolismo, la transcripción y la señalización, lo que sugiere que el genoma de *C. vulgaris* LBCVB está altamente regulado permitiéndole adaptarse a diferentes condiciones.
- ◇ Se identificaron genes involucrados en la asimilación de metales pesados, fármacos y compuestos inorgánicos (N y P). Entre las proteínas con mayor similitud se encontraron transportadores, enzimas y proteínas asociadas al estrés ambiental, de forma directa o indirecta.
- ◇ Se determinaron las rutas metabólicas más relevantes, lo cual confirma la capacidad del organismo para asimilar y transformar diversos contaminantes, y por ende su potencial para los procesos de biorremediación.

9. BIBLIOGRAFÍA

- Abdelfattah, A., Ali, S. S., Ramadan, H., El-Aswar, E. I., Eltawab, R., Ho, S., Elsamahy, T., Li, S., El-Sheekh, M. M., Schagerl, M., Kornaros, M., & Sun, J. (2022). Microalgae-based wastewater treatment: Mechanisms, challenges, recent advances, and future prospects. *Environmental Science And Ecotechnology*, 13, 100205. <https://doi.org/10.1016/j.ese.2022.100205>
- Ahmad, I., Abdullah, N., Koji, I., Yuzir, A., & Mohamad, S. (2021). Potential of Microalgae in Bioremediation of Wastewater. *BULLETIN OF CHEMICAL REACTION ENGINEERING AND CATALYSIS*, 16(2), 413-429. <https://doi.org/10.9767/bcrec.16.2.10616.413-429>
- Asaad, A. A., & Amer, A. S. (2024). Evaluation of *Chlorella vulgaris* biosorption capacity for phosphate and nitrate removal from wastewater. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-023-50748-3>
- Azaman, S. N. A., Wong, D. C. J., Tan, S. W., Yusoff, F. M., Nagao, N., & Yeap, S. K. (2020). De novo transcriptome analysis of *Chlorella sorokiniana*: effect of glucose assimilation, and moderate light intensity. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-74410-4>
- Bai, R. G., Nair, S. C., Joller-Vahter, L., & Kikas, T. (2025). Microalgae in Mitigating Industrial Pollution: Bioremediation Strategies and Biomagnification Potential. *Biomass*, 5(4), 61. <https://doi.org/10.3390/biomass5040061>
- Basantani, M. K., Gupta, D., Mehrotra, R., Mehrotra, S., Vaish, S., & Singh, A. (2017). An update on bioinformatics resources for plant genomics research. *Current Plant Biology*, 11-12, 33-40. <https://doi.org/10.1016/j.cpb.2017.12.002>
- Bossa, R., Di Colandrea, M., Salbitani, G., & Carfagna, S. (2024). Phosphorous Utilization in Microalgae: Physiological Aspects and Applied Implications. *Plants*, 13(15), 2127. <https://doi.org/10.3390/plants13152127>

- Castellanos-Rodríguez, Ó., Expósito, R. R., & Touriño, J. (2023). SeQual-Stream: approaching stream processing to quality control of NGS datasets. *BMC Bioinformatics*, 24(1). <https://doi.org/10.1186/s12859-023-05530-7>
- Cecchin, M., Marcolungo, L., Rossato, M., Girolomoni, L., Cosentino, E., Cuine, S., Li-Beisson, Y., Delledonne, M., & Ballottari, M. (2019). *Chlorella vulgaris* genome assembly and annotation reveals the molecular basis for metabolic acclimation to high light conditions. *The Plant Journal*, 100(6), 1289-1305. <https://doi.org/10.1111/tpj.14508>
- Coronado-Reyes, J. A., Salazar-Torres, J. A., Juárez-Campos, B., & González-Hernández, J. C. (2020). *Chlorella vulgaris*, a microalgae important to be used in Biotechnology: a review. *Food Science And Technology*, 42. <https://doi.org/10.1590/fst.37320>
- Cortese, I. J., Castrillo, M. L., Zapata, P. D., & Laczeski, M. E. (2021). EFECTO DEL FILTRADO DE SECUENCIAS EN EL ENSAMBLADO DEL GENOMA DE *Bacillus altitudinis* AISLADO DE *Ilex paraguariensis*. *Acta Biológica Colombiana*, 26(2), 170-177. <https://doi.org/10.15446/abc.v26n2.86406>
- De Sena Brandine, G., & Smith, A. D. (2021). Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Research*, 8, 1874. <https://doi.org/10.12688/f1000research.21142.2>
- Di Genova, A., Buena-Atienza, E., Ossowski, S., & Sagot, M. (2020). Efficient hybrid de novo assembly of human genomes with WENGAN. *Nature Biotechnology*, 39(4), 422-430. <https://doi.org/10.1038/s41587-020-00747-w>
- Dufault-Thompson, K., & Jiang, X. (2022). Applications of de Bruijn graphs in microbiome research. *iMeta*, 1(1). <https://doi.org/10.1002/imt2.4>

- Ejigu, G. F., & Jung, J. (2020). Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. *Biology*, 9(9), 295. <https://doi.org/10.3390/biology9090295>
- El-Sheekh, M. M., Farghl, A. A., Galal, H. R., & Bayoumi, H. S. (2015). Bioremediation of different types of polluted water using microalgae. *Rendiconti Lincei. Scienze Fisiche E Naturali*, 27(2), 401-410. <https://doi.org/10.1007/s12210-015-0495-1>
- Escapa, C. (2017). *Biorrelación de aguas contaminadas con nutrientes y fármacos mediante microalgas = Bioremediation of contaminated waters with nutrients and pharmaceuticals by microalgae*. [Tesis Doctoral, Universidad de León] <https://doi.org/10.18002/10612/6773>
- Ghariani, O., Elleuch, J., Ferretti, A. M., Econdi, S., Bisio, C., Michaud, P., Fendri, I., Guidotti, M., & Abdelkafi, S. (2025). Toxicological Effects of Silver-Modified Bentonite Nanocomposites on Microalgae: Impact on Cell Growth, Antioxidant Enzymes, and Gene Expression. *Nanomaterials*, 15(8), 629. <https://doi.org/10.3390/nano15080629>
- Gómez-Jacinto, V., García-Barrera, T., Gómez-Ariza, J. L., Garbayo-Nores, I., & Vilchez-Lobato, C. (2015). Elucidation of the defence mechanism in microalgae *Chlorella sorokiniana* under mercury exposure. Identification of Hg-phytochelatins. *Chemico-Biological Interactions*, 238, 82-90. <https://doi.org/10.1016/j.cbi.2015.06.013>
- Gurunathan, B., Selvakumari, I. A. E., Aiswarya, R., & Renganthan, S. (2017). Bioremediation of Industrial and Municipal Wastewater Using Microalgae. *En Energy, environment, and sustainability* (pp. 331-357). https://doi.org/10.1007/978-981-10-7485-1_16

- Hernández, M., Quijada, N. M., Rodríguez-Lázaro, D., & Eiros, J. M. (2019). Aplicación de la secuenciación masiva y la bioinformática al diagnóstico microbiológico clínico. *Revista Argentina de Microbiología*, 52(2), 150-161. <https://doi.org/10.1016/j.ram.2019.06.003>
- Ijaq, J., Chandrasekharan, M., Poddar, R., Bethi, N., & Sundararajan, V. S. (2015). Annotation and curation of uncharacterized proteins- challenges. *Frontiers In Genetics*, 6, 119. <https://doi.org/10.3389/fgene.2015.00119>
- Interpreting the Graphical Output — pyani 0.2.9 documentation. (s. f.). https://pyani.readthedocs.io/en/latest/interpreting_plots.html?utm_source=chatgpt.com
- Kanehisa Laboratories. (s.f.).: Kyoto Encyclopedia Of Genes And Genomes. <https://www.genome.jp/kegg/>
- Kundu, P., Dutta, N., & Bhattacharya, S. (2024). Application of microalgae in wastewater treatment with special reference to emerging contaminants: a step towards sustainability. *Frontiers In Analytical Science*, 4. <https://doi.org/10.3389/frans.2024.1513153>
- Kusuma, H. S., Illiyanasafa, N., Jaya, D. E. C., Darmokoesoemo, H., & Putra, N. R. (2023). Utilization of the microalga *Chlorella vulgaris* for mercury bioremediation from wastewater and biomass production. *Sustainable Chemistry And Pharmacy*, 37, 101346. <https://doi.org/10.1016/j.scp.2023.101346>
- Leong, Y. K., & Chang, J. (2020). Bioremediation of heavy metals using microalgae: Recent advances and mechanisms. *Bioresource Technology*, 303, 122886. <https://doi.org/10.1016/j.biortech.2020.122886>

- Liao, P., Satten, G. A., & Hu, Y. (2017). PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. *Genetic Epidemiology*, 41(5), 375-387. <https://doi.org/10.1002/gepi.22048>
- Li, L., Zhang, G., & Wang, Q. (2016b). De novo transcriptomic analysis of *Chlorella sorokiniana* reveals differential genes expression in photosynthetic carbon fixation and lipid production. *BMC Microbiology*, 16(1), 223. <https://doi.org/10.1186/s12866-016-0839-8>
- Mehra, H. S., Wang, X., Russell, B. P., Kulkarni, N., Ferrari, N., Larson, B., & Vinyard, D. J. (2024). Assembly and Repair of Photosystem II in *Chlamydomonas reinhardtii*. *Plants*, 13(6), 811. <https://doi.org/10.3390/plants13060811>
- Nagarajan, D., Lee, D., Varjani, S., Lam, S. S., Allakhverdiev, S. I., & Chang, J. (2022). Microalgae-based wastewater treatment – Microalgae-bacteria consortia, multi-omics approaches and algal stress response. *The Science Of The Total Environment*, 845, 157110. <https://doi.org/10.1016/j.scitotenv.2022.157110>
- National Center for Biotechnology Information. (s.f.). *Genome Data Viewer*. <https://www.ncbi.nlm.nih.gov/datasets/genome/>
- National Center for Biotechnology Information (NCBI). (s.f.). *Cog*. <https://www.ncbi.nlm.nih.gov/research/cog>
- Noutoshi, Y. (1998). Molecular anatomy of a small chromosome in the green alga *Chlorella vulgaris*. *Nucleic Acids Research*, 26(17), 3900-3907. <https://doi.org/10.1093/nar/26.17.3900>
- Ortiz-Villota, M. T., Romero-Morales, M. A., & Meza-Rodríguez, L. D. (2018). La biorremediación con microalgas (*Spirulina máxima*, *Spirulina platensis* y *Chlorella vulgaris*) como alternativa para tratar la eutrofización de la laguna de

- Ubaque, Colombia. *Revista de Investigación Desarrollo E Innovación*, 9(1), 163-176. <https://doi.org/10.19053/20278306.v9.n1.2018.8153>
- Qian, H., Li, J., Sun, L., Chen, W., Sheng, G. D., Liu, W., & Fu, Z. (2009). Combined effect of copper and cadmium on *Chlorella vulgaris* growth and photosynthesis-related gene transcription. *Aquatic Toxicology*, 94(1), 56-61. <https://doi.org/10.1016/j.aquatox.2009.05.014>
- Ramos, R., & Pizarro, R. (2018). Crecimiento y capacidad de biorremediación de *Chlorella vulgaris* (*Trebouxiophyceae*, *Chlorophyta*) cultivada en aguas residuales generadas en el cultivo del pez dorado *Seriola lalandi* (Perciformes: Carangidae). *Revista de Biología Marina y Oceanografía*, 53(1), 75-86. <https://doi.org/10.4067/s0718-19572018000100075>
- Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-ThroughPut Sequencing Technologies. *Molecular Cell*, 58(4), 586-597. <https://doi.org/10.1016/j.molcel.2015.05.004>
- Sánchez-Sandoval, D., González-Ortega, O., Navarro-Martínez, M. F., Castro-Tapia, J. M., De la Cruz, R. F. G., & Soria-Guerra, R. E. (2021). Photodegradation and Removal of Diclofenac by the Green Alga *Nannochloropsis oculata*. *Phyton*, 90(5), 1519-1533. <https://doi.org/10.32604/phyton.2021.015087>
- Sanz-Luque, E., Chamizo-Ampudia, A., Llamas, A., Galvan, A., & Fernandez, E. (2015). Understanding nitrate assimilation and its regulation in microalgae. *Frontiers In Plant Science*, 6. <https://doi.org/10.3389/fpls.2015.00899>
- Shi, H., & Xu, X. (2016). Learning the Sequences Quality Control of Bioinformatics Analysis Method. *International Conference On Education, E-learning And*

- Management Technology (EEMT 2016)*. <https://doi.org/10.2991/iceemt-16.2016.90>
- Sohn, J., & Nam, J. (2016). The present and future of the novel whole-genome assembly. *Briefings in Bioinformatics*, bbw096. <https://doi.org/10.1093/bib/bbw096>
- Su, Y. (2020). Revisiting carbon, nitrogen, and phosphorus metabolisms in microalgae for wastewater treatment. *The Science Of The Total Environment*, 762, 144590. <https://doi.org/10.1016/j.scitotenv.2020.144590>
- Tatusov, R. L. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1), 33-36. <https://doi.org/10.1093/nar/28.1.33>
- UniProt. (s. f.). UniProt. <https://www.uniprot.org/>
- Wang, Y., Ho, S., Cheng, C., Guo, W., Nagarajan, D., Ren, N., Lee, D., & Chang, J. (2016). Perspectives on the feasibility of using microalgae for industrial wastewater treatment. *Bioresource Technology*, 222, 485-497. <https://doi.org/10.1016/j.biortech.2016.09.106>
- Wang, G., Zhang, Q., Li, J., Chen, X., Lang, Q., & Kuang, S. (2019). Combined effects of erythromycin and enrofloxacin on antioxidant enzymes and photosynthesis-related gene transcription in *Chlorella vulgaris*. *Aquatic Toxicology*, 212, 138-145. <https://doi.org/10.1016/j.aquatox.2019.05.004>
- Wu, T., Li, L., Jiang, X., Yang, Y., Song, Y., Chen, L., Xu, X., Shen, Y., & Gu, Y. (2019). Sequencing and comparative analysis of three *Chlorella* genomes provide insights into strain-specific adaptation to wastewater. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-45511-6>

- Xiong, J., Kurade, M. B., & Jeon, B. (2016). Biodegradation of levofloxacin by an acclimated freshwater microalga, *Chlorella vulgaris*. *Chemical Engineering Journal*, 313, 1251-1257. <https://doi.org/10.1016/j.cej.2016.11.017>
- Xiong, J., Kurade, M. B., & Jeon, B. (2017). Can Microalgae Remove Pharmaceutical Contaminants from Water? *Trends In Biotechnology*, 36(1), 30-44. <https://doi.org/10.1016/j.tibtech.2017.09.003>
- Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y., Hao, H., Yi, W., Li, M., & Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers In Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.766364>

10. ANEXOS

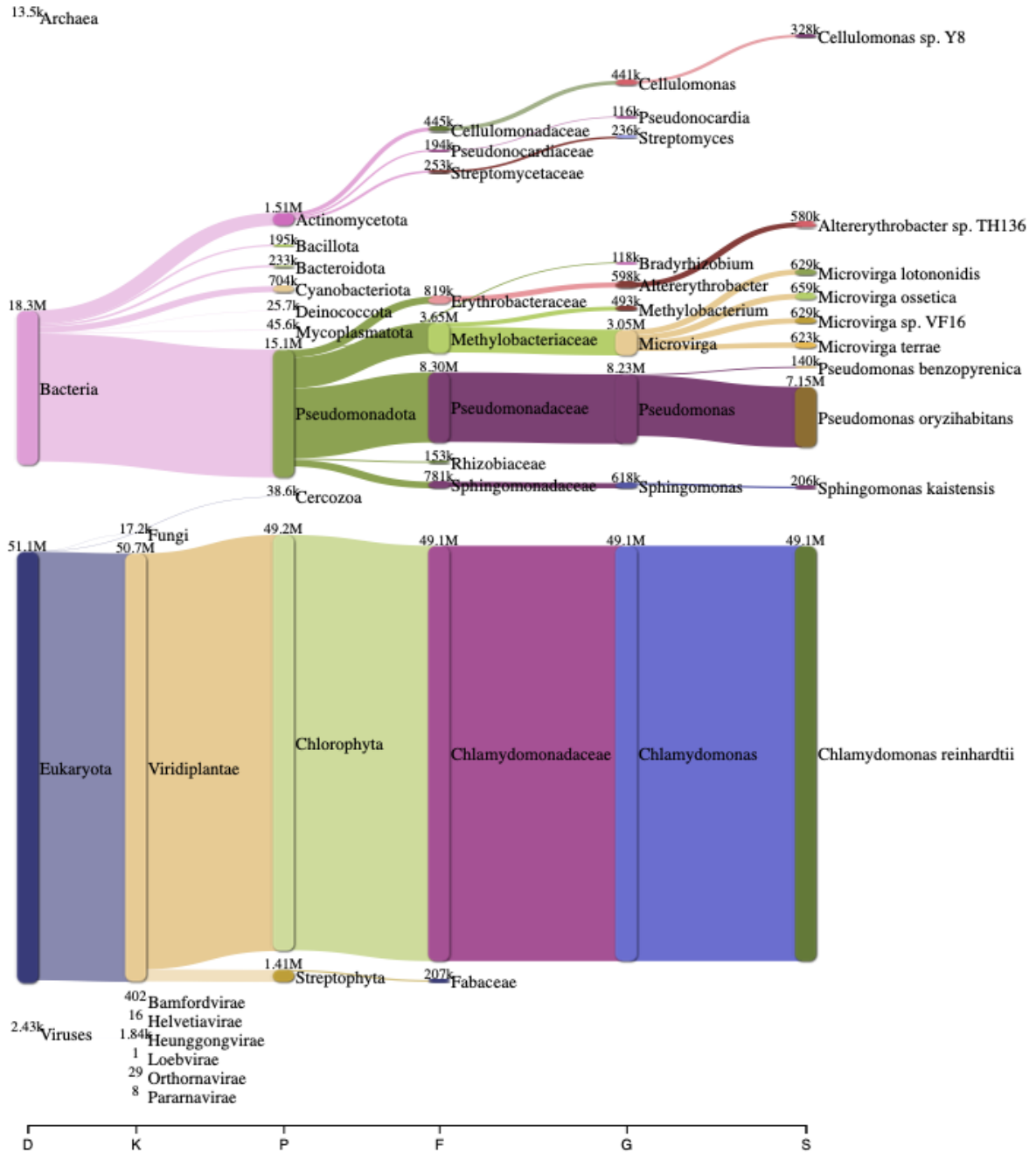


Figura A1. Distribución taxonómica de las secuencias asignadas a partir de las lecturas filtradas por Trimgalore.

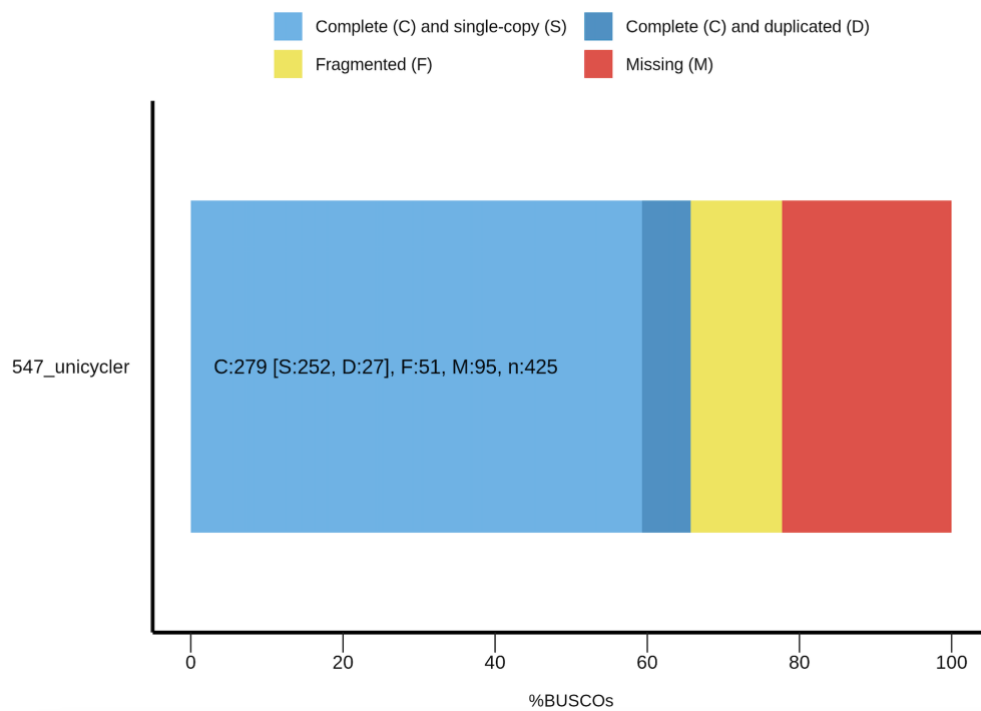


Figura A2. Evaluación de completitud del ensamblaje del genoma de *C. vulgaris* mediante BUSCO (Benchmarking Universal Single-Copy Orthologs).



Figura A3. Ruta de carcinogénesis química asociada a especies reactivas de oxígeno en *C. vulgaris* LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko05208).

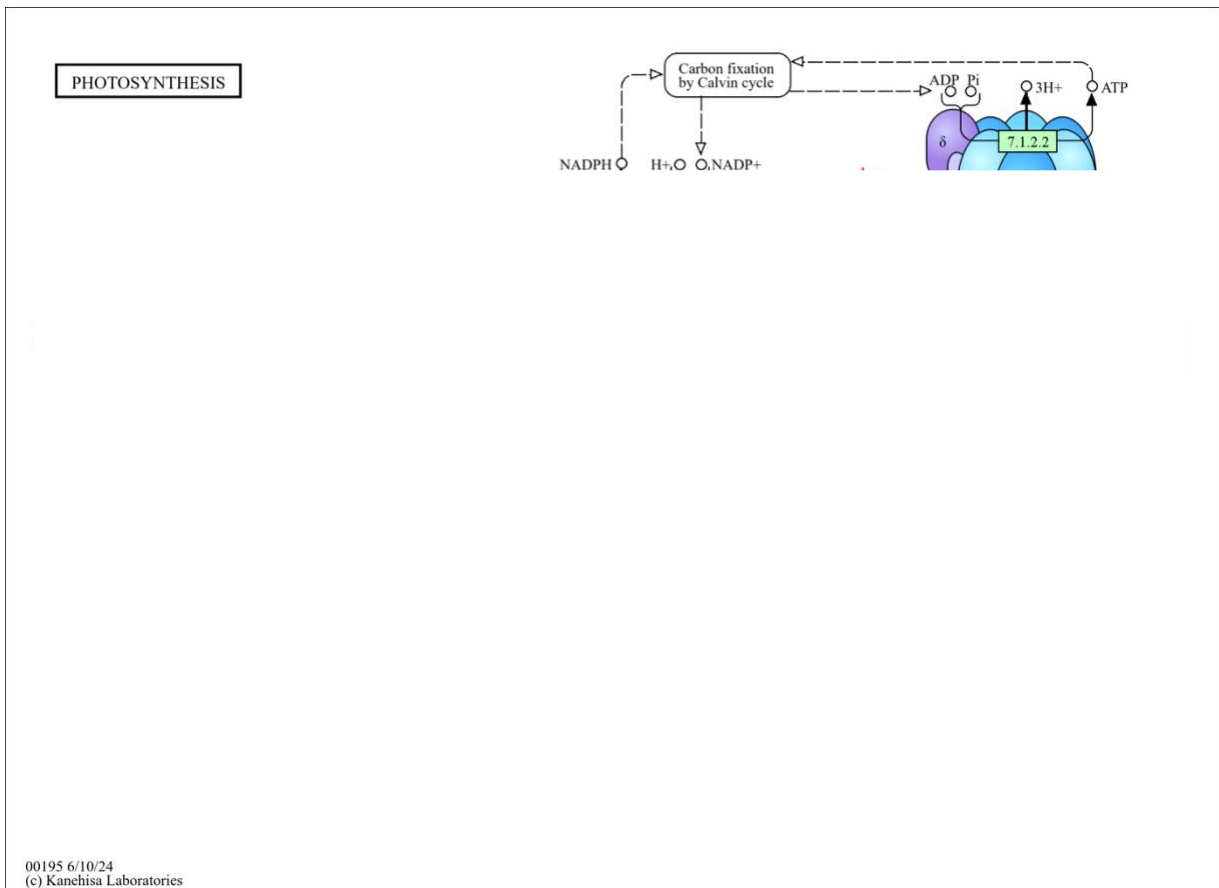


Figura A4. Ruta de la fotosíntesis en *C. vulgaris* LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00195).



Figura A5. Ruta de Fijación de carbono mediante el ciclo de Calvin en *C. vulgaris* LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00710).

DRUG METABOLISM - CYTOCHROME P450

Cyclophosphamide & Ifosfamide

00982 10/9/19
(c) Kanehisa Laboratories

Figura A6. Ruta de Metabolismo de fármacos - citocromo P450 en *C. vulgaris* LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00982).

CARBON FIXATION BY CALVIN CYCLE

00710 6/10/24
(c) Kanehisa Laboratories

Figura A7. Ruta del Metabolismo de fármacos: otras enzimas en *C. vulgaris* LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00983).



Figura A8. Ruta del metabolismo de transportadores ABC en *C. vulgaris* LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko02010).



Figura A9. Ruta del Metabolismo Nitrógeno *C. vulgaris* LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00910).

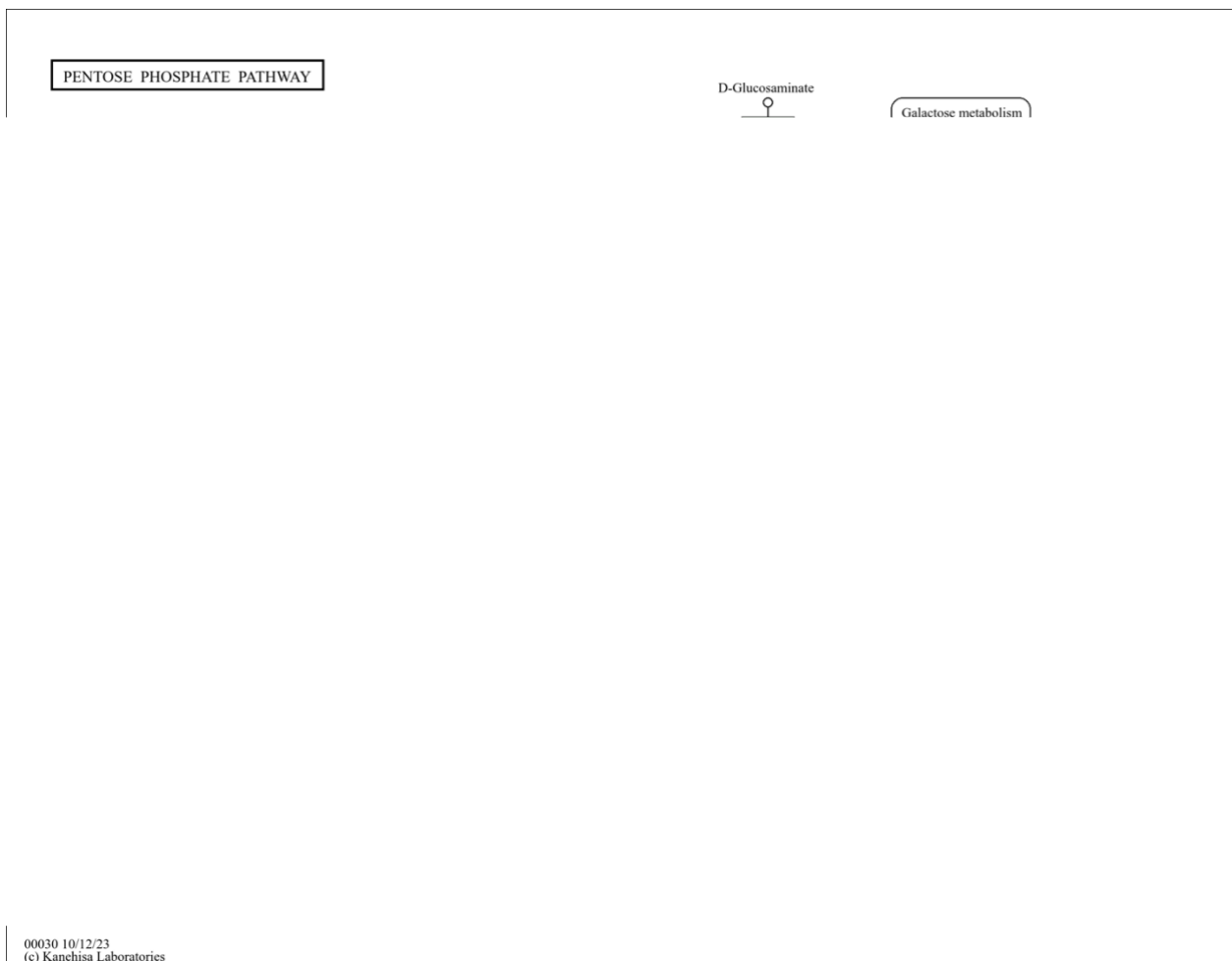


Figura 10. Vía de las pentosas fosfato en *C. vulgaris* LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko00030).

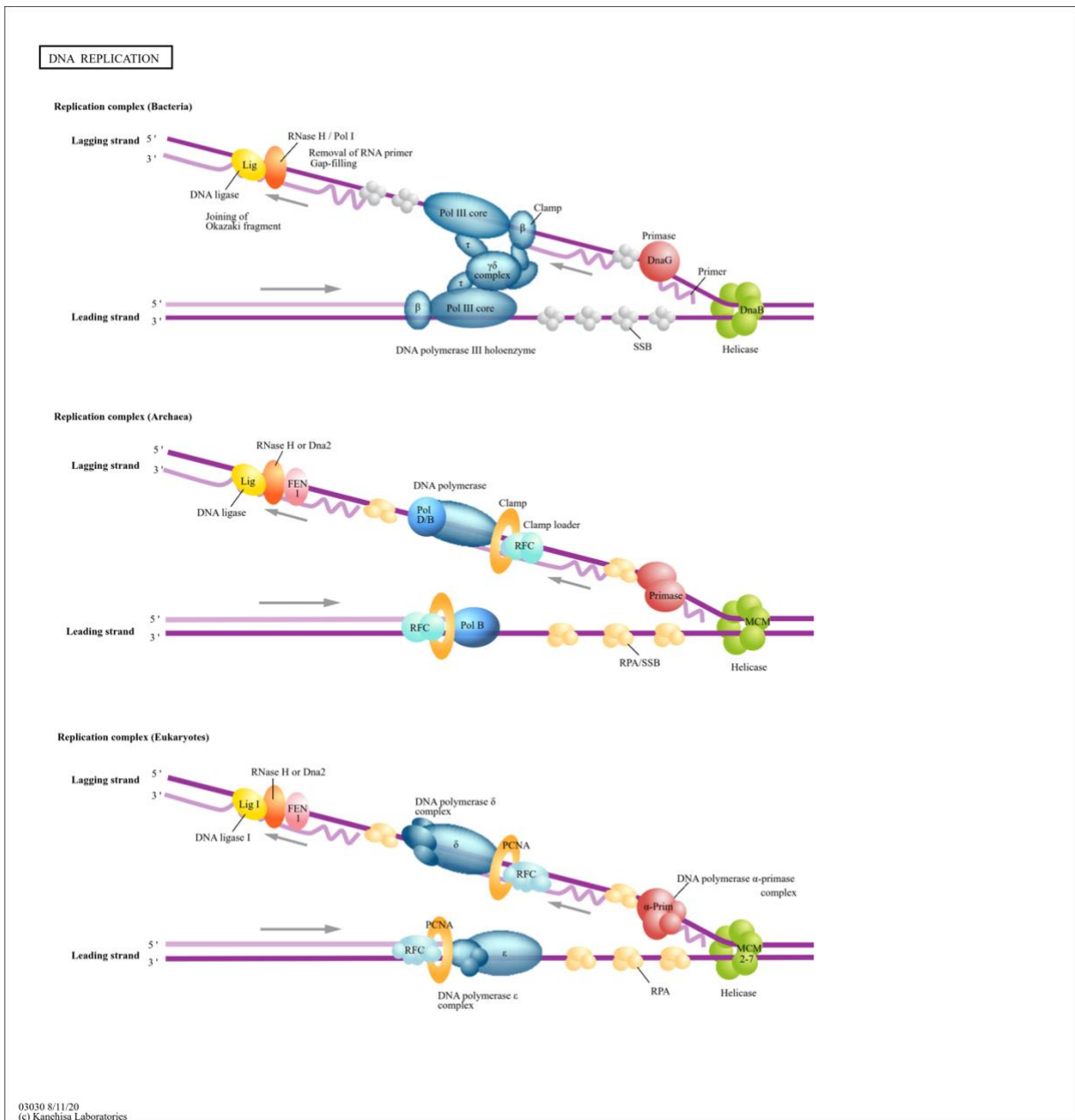


Figura A11. Ruta de Replicación del ADN en *C. vulgaris* LBCVB. Los recuadros en verde indican genes presentes en el genoma (KEGG: ko03030).