

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ciencias y Humanidades



Diseño *in silico* de cebadores a partir de genes de resistencia a condiciones alcalinas para el tamizaje de microalgas alcalófilas y alcalino tolerantes

Trabajo de graduación en modalidad de tesis presentado por
Mateo Alvarado Rodas
para optar al grado académico de Licenciado en Biotecnología Molecular

Guatemala
2020

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ciencias y Humanidades



Diseño *in silico* de cebadores a partir de genes de resistencia a condiciones alcalinas para el tamizaje de microalgas alcalófilas y alcalino tolerantes

Trabajo de graduación en modalidad de tesis presentado por
Mateo Alvarado Rodas
para optar al grado académico de Licenciado en Biotecnología Molecular

Guatemala
2020

Vo.Bo.

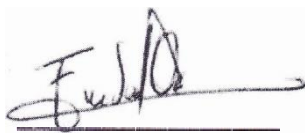
A handwritten signature in black ink, appearing to read 'A. Soto F.', written over a horizontal line.

MsC. Anna Yunuen Soto Fernández

Tribunal examinador:

A handwritten signature in black ink, appearing to read 'Paul Pennington', written over a horizontal line.

PhD. Pamela Marie Pennington Aycinena

A handwritten signature in black ink, appearing to read 'E. Solórzano', written over a horizontal line.

MsC. Elizabeth Solórzano Ortiz

A handwritten signature in black ink, appearing to read 'A. Soto F.', written over a horizontal line.

MsC. Anna Yunuen Soto Fernández

Fecha de aprobación del examen de graduación:

Guatemala, 4 de diciembre de 2020

ÍNDICE

Lista de figuras.....	i
Lista de cuadros	ii
Resumen.....	iii
Abstract	iii
I. Introducción.....	1
II. Antecedentes	2
III. Justificación	3
IV. Objetivos.....	4
V. Hipótesis.....	5
VI. Marco teórico.....	6
VII. Metodología.....	26
VIII. Resultados y discusión	34
IX. Conclusiones.....	50
X. Recomendaciones.....	51
XI. Bibliografía.....	52
XII. Anexos.....	63

Lista de figuras

- Figura 1: Árbol filogenético de las secuencias de la superfamilia KefB realizado con test de filogenia de bootstrapping de 1000 réplicas y con el modelo de sustitución de Jones-Taylor-Thornton. En rojo se encuentran los antiports TrAK_N, en azul los que presentan TraK_C, morado Usp o Usp_Like y verde Na_H_Exchanger_C (NaHC) (Construido en MegaX) 40
- Figura 2: Árbol filogenético de las secuencias de la superfamilia NhaP realizado con test de filogenia de bootstrapping de 1000 réplicas y con el modelo de sustitución de Jones-Taylor-Thornton . Como modelo de sustitución se utilizó Jones-Taylor-Thornton . En rojo se encuentran los antiports TrAK_N y en morado Usp o Usp_Like (construido con Mega X) 41
- Figura 3: Región seleccionada del alineamiento de las secuencias proteicas de la superfamilia KefB, correspondiente a las posiciones 78 a 102. Con las flechas se señalan los aminoácidos altamente conservados en las secuencias. (explorador de secuencias de Mega X) 42
- Figura 4: Región seleccionada del alineamiento de las secuencias proteicas de la superfamilia KefB, correspondiente a las posiciones 166 a 209. Con las flechas se señalan los aminoácidos altamente conservados en las secuencias. 43
- Figura 5: Región seleccionada del alineamiento de las secuencias proteicas de la superfamilia KefB, correspondiente a las posiciones 450 a 470. Con las flechas se señalan los aminoácidos altamente conservados en las secuencias. 44
- Figura 6: Regiones seleccionadas del alineamiento de las secuencias proteicas de la superfamilia NhaP: la sección 1 corresponde a los aminoácidos 76 a 91, la 2 a los de 248 a 278 y la 3 a los de 440 a 455. Con las flechas se señalan los aminoácidos altamente conservados en las secuencias 45
- Figura 7: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_7*. Las otras imágenes se encuentran en anexos 49
- Figura 8: Segmento del cuadro de información de especies relevantes con las indicaciones de la sección 1.1 de la metodología 63
- Figura 9: Segmento del cuadro de información de especies aisladas en Guatemala generado con las indicaciones de la sección 1.1 de la metodología 63
- Figura 10: Segmento del cuadro sobre las características de las secuencias descargadas generado con las indicaciones de la sección 1.2 de la metodología 63
- Figura 11: Segmento de un cuadro de secuencias alineadas para el género *Arthrospira* generado con las indicaciones de la sección 2.1 64
- Figura 12: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_8*. 64

Figura 13: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso <i>Synechococcus</i> _9.	65
Figura 14: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso <i>Synechococcus</i> _11.	65
Figura 15: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso <i>Synechococcus</i> _13.	66
Figura 16: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso <i>Synechococcus</i> _15.	66
Figura 17: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso <i>Synechococcus</i> _16.	67
Figura 18: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso <i>Synechococcus</i> _17.	67
Figura 19: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso <i>Synechococcus</i> _20.	68
Figura 20: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso <i>Synechococcus</i> _22.	68

Lista de cuadros

Cuadro 1: Resumen de información genómica y proteómica disponible en línea de especies de interés y aisladas en Guatemala. Los géneros con 5 o más genomas disponibles están resaltados en negrita.	35
Cuadro 2: Cebadores propuestos para agrupaciones similares de antiports de <i>Synechococcus</i> analizados. En negrita se resaltan los cebadores que dieron resultados.	47

Resumen

El cultivo de microalgas se perfila, a futuro, como un modo de producción de biomasa y metabolitos secundarios de alto valor agregado. Además, puede asociarse a la generación de emisiones de dióxido de carbono, haciendo sostenible al sistema. Actualmente, los cultivos a escala industrial son poco viables por la dificultad de controlar las condiciones óptimas de cultivo en volúmenes importantes, elevando así los costos de mantenimiento. Utilizar microalgas adaptadas a condiciones alcalinas de cultivo minimiza el problema, ya que los cultivos alcalinos son menos susceptibles a contaminación por otros microorganismos y son capaces de captar CO₂. La alcalinotolerancia y alcalofilia se deben principalmente a proteínas de membrana que regulan la entrada y salida de iones alcalinos y protones, garantizando la homeostasis a pH elevados. En esta investigación, se propuso el diseño de cebadores para identificar molecularmente especies alcalinotolerantes y alcalofílicas. Se realizó una revisión de las secuencias disponibles en bases de datos para identificar regiones adecuadas para cebadores en los genes asociados. Se realizó igualmente una descripción de las regiones de las proteínas y un análisis filogenético para completar la descripción de estas y la selección de las regiones. Como resultado, se determinó que la mayoría pertenecían a dos superfamilias y que poseían regiones conservadas reducidas, intercaladas con regiones variables y ricas en aminoácidos neutros. No se identificó una región adecuada para diseñar cebadores que identifiquen un número importante de especies, por lo que se diseñaron cebadores para el género *Synechococcus* por la información disponible y posibilidad de aislamiento en Guatemala.

Abstract

Microalgae have the potential to become an alternative source of biomass as well as high-end metabolites. By their photosynthetic property and rapid generation time, they are also capable of using anthropogenic CO₂, which can reduce the overall cost of the system. However, industrial size microalgae cultivation is not viable yet, as maintaining and monitoring those systems for optimal production is very expensive. An approach to overcome this limitation is the cultivation of alkalitolerant and alkaliphilic microalgae. Culture media at high pH is less susceptible to microbial contamination and is capable of fixing environmental CO₂, which reduces costs associated with production. The capacity to grow at high pH is due to cation-proton antiports, that extrude alkaline cations and harness free protons, allowing a intercellular neutral pH. In this work, we evaluated if it were possible to design primers using the sequences from alkalitolerant and alkaliphilic species. Firstly, a review of available sequences was made in order to classify and identify potential regions for PCR primer design for the genes associated with the resistance. A description of the sequences, as well as a phylogenetic analysis were conducted to better classify and complement the intended goal. We determined that most of the proteins belonged to two superfamilies, with a limited number of conserved regions interspersed with regions rich with neutrally-charged aminoacids. As a result, it was not possible to select a region that could used for primer design for multiples . Instead, we proposed multiple sets of primers for the identification of species of the genus *Synechococcus* due to previous isolation of the genus and relatively abundant information in free databases.

I. Introducción

Las microalgas se perfilan como microorganismos útiles para la producción de biomasa y productos de alto valor agregado. Para tener los rendimientos óptimos, la identificación de las especies adecuadas es indispensable. El trabajo realizado tuvo la finalidad de determinar si era posible diseñar un método general para detectar microalgas para cultivo industrial en pH elevados. Para ello, se diseñaron cebadores para la reacción de PCR para la detección de estas especies. Este trabajo se enfocó en el estudio de las secuencias de diferentes especies y las relaciones evolutivas entre estas. Se presentó como hipótesis que las secuencias proteicas derivadas de genes de resistencia podrían utilizarse para identificar las especies de interés. Se esperaba encontrar regiones conservadas que permitieran el diseño de los cebadores. Se concluyó que las secuencias proteicas derivadas de estos genes no presentaron regiones conservadas para el diseño de cebadores, ya que son altamente variables. Se concluyó también que el estudio de microalgas de interés debe concentrarse en una sola especie y no en varias simultáneamente.

II. Antecedentes

El estudio de microalgas en Guatemala representa un volumen reducido de publicaciones. La mayoría de artículos publicados en revistas indexadas estudian los florecimientos de microalgas nocivas en la costa y los lagos (Komárek *et al.*, 2013).

A nivel de tesis de licenciatura, las publicaciones desde el 2010 se concentran principalmente en el uso de microalgas como indicadores de calidad de agua (Morales Molina, 2016), caracterización y crecimiento para producción de lípidos (Raxón Matías, 2018; Ruiz, 2013; Estrada Palencia, 2013) y diseño de reactores (Lobo de León, 2018). También se han realizado trabajos centrados en la descripción morfológica, principalmente para fines didácticos y de uso interno en instituciones, así como un trabajo centrado en la caracterización molecular de microalgas. En general, las publicaciones se concentran en aplicaciones y campos bien establecidos del estudio de microalgas.

Hasta el día de hoy, hay pocos estudios de microalgas con enfoque molecular en Guatemala. No se ha realizado ningún estudio genómico o multi ómico de microalgas, así como estudios molecular distintos de la identificación de especies. Tampoco se han publicado estudios de cultivo de especies aisladas localmente ni evaluaciones de cultivo de microalgas a gran escala. Esto se debe principalmente por el costo y la falta de plataformas para continuarlos.

III. Justificación

El cultivo de microalgas se presenta como una alternativa biotecnológica para la producción de combustibles y productos finos, como pigmentos y fármacos. El gran número de especies, la diversidad metabólica y el uso de diferentes sustratos para su crecimiento las perfilan como una fuente de productos biotecnológicos para el sector energético, alimentario y farmacéutico. Además, pueden cultivarse sin competir con explotaciones agrícolas. Sin embargo, existen varias limitantes biológicas y fisicoquímicas que impiden el cultivo eficiente a gran escala. Entre estas limitantes se encuentran el requerimiento de luz, el control de las condiciones y la susceptibilidad a contaminación. Se requiere de luz en la mayoría de procesos, lo que condiciona los diseños de biorreactores para aumentar el contacto con la misma. Sumado a esto, la mayoría de sistemas requieren del control preciso de sus condiciones para producir rendimientos óptimos y cambios leves pueden minimizar la producción de biomasa. Finalmente, la mayoría de sistemas actuales son muy susceptibles a contaminaciones por depredadores y bacterias. Estos tres aspectos elevan los costos de producción y vuelven a la tecnología aún restrictiva en relación a costos.

Una solución posible para superar estas limitantes es el uso de algas extremófilas, en particular, alcalófilas y alcalino tolerantes. La alcalinidad, con un pH entre 8 y 11, vuelve al sistema más robusto: es menos susceptible a cambios en el ambiente, inhibe o mata a la mayoría de organismos contaminantes e incluso capta dióxido de carbono ambiental. Estas características se traducen en un sistema más estable y más económico de mantener, en especial cuando se trabaja a mayor escala. Por lo anterior, ha surgido una línea de investigación para el aislamiento, cultivo y optimización del uso de microalgas alcalófilas y alcalino tolerantes para la producción de biomasa. No obstante, hay pocas investigaciones que estudian los genes de resistencia a pH elevados en microalgas, ya que no priorizan el estudio de los factores responsables del crecimiento en medio básico. Este trabajo es un primer esfuerzo, al proponer un método de tamizaje específico para detectar la presencia de microalgas alcalófilas y alcalino tolerantes para fines industriales.

IV. Objetivos

A. General

Diseñar un conjunto de cebadores *in silico* basados en secuencias conservadas de genes de resistencia, para el tamizaje de especies de microalgas alcalino tolerantes y alcalófilas con facilidades de cultivo a gran escala.

B. Específicos

1. Realizar una descripción y análisis de secuencias y regiones conservadas de los genes de resistencia (ATPasa) obtenidos de las bases de datos
2. Realizar un agrupamiento inicial por un alineamiento múltiple de secuencias para identificar la filogenia de los genes de resistencia y sus genes parálogos.
3. Diseñar cebadores degenerados y específicos para detectar los genes de resistencia alcalina en diferentes especies de microalgas.
4. Generar una base de datos de cebadores para genes de resistencia de alcalino tolerancia y alcalófilia con potencial industrial presentes en microalgas.

V. Hipótesis

Se pueden diseñar cebadores a partir de genes de resistencia a condiciones alcalinas en microalgas, para identificar especies con potencial para cultivo a gran escala.

VI. Marco teórico

A. Taxonomía y clasificación de microalgas

1. Generalidades

Las microalgas son un conjunto heterogéneo de microorganismos fotosintéticos. Se caracterizan por ser parcial o completamente autótrofos, es decir que producen materia orgánica a partir de luz y materia inorgánica (principalmente, dióxido de carbono - CO₂ -). Pueden utilizar otras fuentes de carbono orgánico, según las circunstancias y las especies. Se han aislado de todos los ambientes, estando presentes en suelos y agua y tanto en condiciones templadas como en condiciones extremas (Andersen, 2005). Actualmente, se estima que hay de 200 000 a 800 000 especies de algas, pero solo se han descrito 35 000 (Cheng & Odgen, 2011).

La principal clasificación de las microalgas se basa en las características fisiológicas que presentan. El primero es el dominio al que pertenecen: procariotas o eucariotas. Las cianobacterias constituyen el grupo procariota de microalgas. Son bacterias fotosintéticas sin clasificación de Gram. Se caracterizan por presentar membranas internas plegadas en su citoplasma, lo que compartimentaliza el ambiente intracelular. Esta compartimentalización genera un gradiente de potencial entre ambientes y vuelve funcional la maquinaria fotosintética (Andersen, 2005). Se capta la energía solar mediante pigmentos fotosintéticos y, por el ciclo de Calvin, se reduce el dióxido de carbono y agua en oxígeno, carbono orgánico (usualmente azúcares). La mayoría de cianobacterias son organismos de vida libre, pudiendo también vivir de forma colonial o en simbiosis con otros organismos (Malcata *et al.*, 2018).

Las microalgas eucariotas constituyen el otro grupo de microalgas. Este conjunto es significativamente más heterogéneo. Los únicos aspectos que las unen son la presencia de núcleo y su capacidad fotosintética, mediada por el cloroplasto. Este organelo está compartimentalizado similarmente a una cianobacteria y en este se realiza el ciclo de Calvin. Todas presentan alguna variante del organelo (Malcata *et al.*, 2018).

Existen 4 grupos principales de microalgas eucariotas: diatomeas, dinoflagelados, algas verdes (chlorophyta) y haptophyta. Cada grupo presenta sus especificidades: algunas son de vida libre, mientras otras son coloniales; la mayoría son autótrofas, pero también existen mixotróficas; unas son bénticas y otras son planctónicas; entre otras características significativas (Bellinger & Sigeo, 2015).

2. Clasificación de microalgas según fines aplicados

Se distinguen otros grupos según los fines aplicados, además de agruparlas por características fisiológicas. Tradicionalmente, se dividen por toxicidad (Ebenezer *et al.*, 2012), y por presencia según la calidad del agua (Palmer, 1969). A nivel de cultivo, se distinguen las microalgas por su origen, marinas o de agua fresca, así como por la presencia de características extremófilas (Bellinger & Sigeo, 2015). Finalmente, también se agrupan por las moléculas que produzcan y lípidos (Neofotis *et al.*, 2016; Zhang *et al.*, 2018). Estas clasificaciones son de interés, ya que el modo de cultivo y los productos sintetizados están asociados a los costos y beneficios de la producción.

Las microalgas son de suma importancia ecológica, ya que son la base de las redes tróficas en los ecosistemas que ocupan. Por ser organismos autótrofos y de talla microscópica, son los principales responsables de convertir carbono inorgánico y otros elementos a compuestos orgánicos y disponibles para organismos superiores (Cembella, 2003). Eventualmente, el humano las consume, ya sea de forma directa o indirectamente por medio de pescados (García *et al.*, 2017).

3. Diversidad intra especies de microalgas

Las microalgas presentan un alto grado de diversidad intraespecie. Es común que se reporten diferentes variedades de una misma especie con funciones metabólicas adicionales o expresadas diferencialmente. Por ejemplo, se han reportado especies de *Chlamydomonas reinhardtii*, una de las microalgas más estudiadas, que fermentan azúcares para producir hidrógeno (Xu *et al.*, 2019) al igual que variedades productoras de lípidos (Fedorov *et al.*, 2005). Otros ejemplos de diversidad son la tolerancia a condiciones extremas y producción variable de lípidos (Neofotis *et al.*, 2016) y otros metabolitos secundarios (Bajhaiya *et al.*, 2015).

Esta diversidad aparente se explica por la adaptabilidad de las microalgas. Los tiempos de generación rápidos, usualmente de un día, permiten a una misma especie adaptarse a diferentes condiciones ambientales y presiones selectivas si se mantiene en un ambiente determinado. Como consecuencia, es común encontrar en la literatura las mismas especies reportadas y aisladas en todo el mundo con metabolismos distintos (Minhas *et al.*, 2016; Neofotis *et al.*, 2016).

La diversidad también se debe a la dificultad de identificar especies a nivel morfológico e incluso molecular. Ciertas familias, como las diatomeas y microalgas de morfología esférica, no presentan características distintivas evidentes entre sí, lo que ha resultado en diferentes especies como una, incluso si presentan un metabolismo distinto (Zou *et al.*, 2016; Bellinger, 2015). El desarrollo de métodos moleculares ha reducido este problema, pero incluso a nivel de género se observa una diversidad significativa. Estudios genómicos recientes de *Nannochloropsis* oleaginosas han revelado que los genes accesorios conforman la mayoría genoma. Los asociados a la síntesis de lípidos están también entre los menos conservados (Wang *et al.*, 2014). Aunque no se tengan tantos estudios de esta envergadura, es posible afirmar que haya más especies en la misma situación.

B. Estudio y cultivo de microalgas

1. Antecedentes

Tradicionalmente, el estudio de microalgas se ha concentrado en algas nocivas, por el impacto negativo de las cianotoxinas en la salud humana (Ebenezer *et al.*, 2012). A raíz de esto, la mayoría de métodos desarrollados se han concentrado en estas especies y géneros. En paralelo, se han utilizado diatomeas para evaluar la calidad de agua y las metodologías desarrolladas son una práctica bien establecida y utilizada (Bellinger, 2015).

No fue hasta finales de la década de los 1970 que se empezó a explorar y explotar el potencial biotecnológico de las microalgas. En esos años empezaron los primeros esfuerzos de aislamiento e identificación de especies de microalgas de alto potencial biotecnológico. Gracias a estos programas, se identificaron a más de 30 000 especies y variedades nuevas de interés (Cheng & Odgen, 2011). Los géneros más comunes

identificados fueron *Botryococcus*, *Chaetoceros*, *Chlorella*, *Dunaliella*, *Haematococcus*, *Isochrysis*, *Nannochloropsis*, *Scenedesmus* (Li *et al.*, 2015).

2. Situación actual

Se ha propuesto, en los últimos años, el cultivo de microalgas como una fuente de energía alternativa y verde a los combustibles fósiles. Esto se debe a la capacidad de ciertas especies de sintetizar ácidos grasos en cantidades relativamente elevadas y de forma no competitiva con cultivos alimentarios (Larkum *et al.*, 2012). Como resultado, la mayoría de la investigación actual se enfoca en microalgas productoras de ácidos grasos y en su genética. Se incluyen en esta categoría las investigaciones dirigidas a la optimización de medios y extracción de ácidos grasos (Bajhaiya *et al.*, 2016). En conjunto a lo anterior, la investigación de microalgas se ha expandido para abarcar otras aplicaciones. En un análisis de publicaciones sobre microalgas realizado en 2019 en Europa, otros términos asociados son “ambiente”, “alimentos”, “químicos” y “pigmentos”. Esto indica un cambio en la tendencia de investigación, pero sigue siendo incipiente (Rumin *et al.*, 2020).

3. Limitaciones técnicas del cultivo de microalgas a gran escala

No se cultivan microalgas a gran escala aún, pese al potencial y a los esfuerzos, ya que se no ha encontrado una cepa o un sistema poco afectado por la variabilidad de la luz y nutrientes sin que se consuman los productos sintetizados. Para que la producción sea rentable, es necesario mantener un equilibrio de condiciones bastante estricto (Wang *et al.*, 2013). A gran escala, un sistema de microalgas es también susceptible a la competencia de bacterias y a contaminación por forrajeros. La contaminación de estos puede sustituir el cultivo de microalgas o consumirlas, respectivamente. Nuevamente, el control de ambas condiciones aumenta los costos de mantenimiento (Mutanda *et al.*, 2011). El único tipo de sistema que se mantiene productivo y sin contaminantes suele ser el cultivo de algas extremófilas y, en particular, alcalófilas y alcalino tolerantes.

C. Algas alcalófilas y alcalino tolerantes

1. Características generales

Las microalgas alcalófilas y alcalino tolerantes son un subgrupo heterogéneo de microalgas que sobreviven, e incluso prosperan, en condiciones de pH mayores a 7.5. Se deben distinguir dos grupos: las microalgas alcalino tolerantes y las alcalófilas. Las primeras se caracterizan por tener un pH ideal de crecimiento de 8.5 aproximadamente y tolerar hasta 9.0 de pH (Tu *et al.*, 2018; Taylor *et al.*, 2012). El segundo grupo requiere de pH mayor a 9 para crecer, como *Arthrospira platensis* y *Arthrospira máxima* (Gimmler & Degenhart, 2001; Vadlamani *et al.*, 2017). Debe mencionarse que las microalgas alcalófilas suelen ser también halotolerantes y halófilas, dado que los ambientes alcalinos suelen ser ricos en sales (Liu *et al.*, 2019). Existen dos mecanismos principales de adaptación al pH elevado. El primero consiste en la síntesis de solutos accesorios, principalmente glicerol, sacarosa y trehalosa. Estos se mantienen en el citoplasma y contribuyen a la regulación de la presión osmótica (Chi *et al.*, 2014). El segundo mecanismo es la presencia de ATPasas especializadas.

2. ATPasas: características generales y genes de resistencia a alcalinidad

Las ATPasas son un conjunto de proteínas de membrana presentes en todos los seres vivos. Pertenecen a la familia de transportadores membrenarios, ya que regulan la entrada y salida de moléculas e iones del ambiente celular. Es una forma de transporte activo, pues consume ATP para realizarse. Fueron de primero descubiertas en animales en los años 60 y posteriormente en todos los dominios de la vida (Popova & Balnokin, 2013). Este conjunto de proteínas está dividido en diferentes superfamilias, como CPA1, CPA2 y LmrB para citar algunas (Resch *et al.*, 2011). Por ello y su diversidad, estas superfamilias son altamente diversas, lo que ha resultado en secuencias variables y con aminoácidos puntuales conservados (Saier, 1999, Resch, 2011).

Existen tres tipos: uniporters que transportan en una dirección un solo compuesto; simportes, que transportan dos simultáneamente y finalmente antiportes, que transportan dos compuestos en sentido contrario. Por estas características, son esenciales para la homeostasis. En microalgas, los más estudiados son los antiportes de

sodio-hidrógeno a partir de resultados extrapolado de investigaciones en plantas (Gimmler & Degenhart, 2001; Taylor *et al.*, 2012).

Las ATPasas responsables de la resistencia a pH elevado han sido menos estudiadas, incluso si se sabe que se diferencian del resto. En efecto, en condiciones básicas, hay una mayor concentración de protones dentro del citoplasma que en el ambiente. Asimismo, en medios ricos en sales, los cationes presentes en el ambiente tienden a ingresar al citoplasma, por la diferencia de gradiente. Como resultado, estas microalgas se encuentran en un ambiente pobre en protones, diferente de los ambientes ácidos en los que viven la mayoría de seres vivos (Kumari & Rathore, 2020). El sistema de ATPasas tolerantes cumple como función reingresar protones al citoplasma, aunque no se conozca precisamente el mecanismo de acción (Gimmler & Degenhart, 2001). Para las especies halófilas, esto usualmente está acoplado a la expulsión de sodio del mismo, haciendo de estas ATPasas antiports (Gimmler & Degenhart, 2001; Kumari & Rathore, 2020).

3. Utilización de resistencia a medio básico desde un punto de vista de producción y económico

El cultivo de microalgas en condiciones básicas presenta ventajas que lo hace interesante para realizarlo a gran escala. El pH elevado tiende a inhibir el crecimiento microbiano y la presencia de depredadores: la descendencia de depredadores como *Daphnia* no sobrevive a pH mayores a 8.5 y a pH mayores de 9.5 solo sobreviven unas cuantas especies (Mutanda *et al.*, 2011; Tu *et al.*, 2018). El sistema es menos susceptible a los cambios en el ambiente e incluso es capaz de captar dióxido de carbono y almacenarlo bajo la forma de bicarbonato y carbonato. Esta última característica reduce la necesidad de introducir dióxido de carbono en el medio y, en conjunto a lo mencionado antes, reduce los costos de producción y manutención de este tipo de sistema (Vadlamani *et al.*, 2017; Kuo *et al.*, 2018).

D. Genes ortólogos y parálogos

1. Utilización de resistencia a medio básico desde un punto de vista de producción y económico

Como parte del material genético, los genes están sujetos a eventos evolutivos. Esto resulta en un aumento de diversidad de proteínas transcritas, tanto en composición como en funcionalidad. Se distinguen varios tipos de relaciones evolutivas derivadas de estos eventos. Para fines de este trabajo, se describen dos: genes ortólogos y genes parálogos (Gogarten & Olendzenski, 1999).

Los genes ortólogos son genes de función similar en diferentes especies. Surgen por un evento de especiación, porque surgen con la aparición de nuevas especies. Al ser genes funcionales, conservan características y regiones similares, según su importancia. Estas características los hacen de suma utilidad en bioinformática y genómica. La comparación de secuencias entre especies conocidas y desconocidas permite identificar y anotar secuencias nuevas. Esto, a su vez, facilita el estudio de genomas de especies menos investigadas, permite el estudio de familias de genes y descubrimientos de nuevas biomoléculas, entre otras aplicaciones (Li *et al.*, 2003).

Los genes parálogos son genes duplicados y presentes en un mismo genoma. Surgen por eventos de duplicación, que es un error en la replicación del material genético y conduce a la repetición y relocalización de la secuencia original. Sucede a nivel intraespecies. Estos genes pierden su función original variablemente, porque pueden cambiar de función, ser complementarios al gen original o expresarse selectivamente según el tejido (Hoffmann & Palmgren, 2016). Es necesario diferenciar los genes parálogos de los ortólogos, para realizar un análisis comparativo adecuado de los últimos. Al momento de escritura de este trabajo, se tiene una comprensión limitada del origen y funcionalidad de la mayoría de genes parálogos (Gogarten & Olendzenski, 1999).

2. Generalidades de la familia de ATPasas

La familia de genes de ATPasas transportadoras de membrana presenta una diversidad importante. Su importancia en la homeostasis, sumada a la diversidad de ambientes y a su presencia en todos los dominios de la vida, han resultado en una familia extensa. Existen diferentes tipos de ATPasas interespecies e intraespecies, con funciones distintas y especificidades al tejido (Armesto *et al.*, 2015; Liu *et al.*, 2016; Galarza-Muñoz *et al.*, 2011). Sin embargo, debe tenerse en mente que las ATPasas responsables de la alcalino tolerancia y alcalofilia comparten pocas características con

la mayoría de proteínas de esta familia (Popova & Balnokin, 2013), por lo que los resultados actuales no necesariamente aplican a estas proteínas.

E. Identificación de microalgas

1. Generalidades y biosprospección tradicional

La biosprospección es la exploración de biodiversidad con el fin de encontrar moléculas pequeñas, macromoléculas y genes para fines comerciales (Beattie *et al.*, 2012). Tradicionalmente, se asocia con la industria farmacéutica por el descubrimiento de metabolitos secundarios, aunque se ha expandido el alcance a otros sectores biotecnológicos. En estas ramas, el esfuerzo se concentra en la exploración y descubrimiento de microorganismos que presentan características o bioproductos de interés, como enzimas (Gençkal & Tari, 2006), genes de resistencias a condiciones extrema de cultivo (Varshney *et al.*, 2015) y rutas metabólicas (Kajikawa *et al.*, 2015).

Existen dos tipos de biosprospección. El primero recopila y utiliza los conocimientos tradicionales de un área y comunidad determinadas. Se centra en la investigación y explotación de especies ya conocidas. Este acercamiento está asociado a problemas legales y morales, pues el conocimiento utilizado pertenece al acervo de la comunidad local y generalmente se realiza poca o nula compensación económica (Abdelgawad, 2012). El segundo acercamiento es la biosprospección orientada a la genómica.

2. Biosprospección orientada a la genómica

El segundo tipo de biosprospección se funda en el estudio genómico de los organismos presentes en un ambiente. Se concentra principalmente en los genes, proteínas y rutas metabólicas que les permitan adaptarse a las condiciones ambientales (Kajikawa *et al.*, 2015; Elhag *et al.*, 2017). Por lo general, estos ambientes presentan presiones ecológicas específicas como concentraciones elevadas de ciertos compuestos, condiciones fisicoquímicas extremas y poca disponibilidad de micro y macronutrientes (Guisado *et al.*, 2015; Gençkal & Tari, 2006; Varshney *et al.*, 2015). Las condiciones pueden asemejarse a procesos productivos, ya que estos se optimizan para la producción

y no para el crecimiento de organismos. Como consecuencia, presentan características similares a las presiones ecológicas antes citadas y se les consideran hostiles a la mayoría de organismos vivos. Cuando se investigan estos ambientes, es más probable que se encuentren microorganismos adaptados a condiciones de estrés (Gençkal & Tari, 2006; Pascoal *et al.*, 2020).

La bioprospección centrada en microorganismos resulta con alto grado de aplicabilidad. Primeramente, la identificación de genes, proteínas y rutas metabólicas expande la cantidad de recursos para ingeniería genética y metabólica. El ejemplo más notorio es en biorremediación, en donde se han descubierto microorganismos que metabolizan hidrocarburos de cadena larga (Smits *et al.* 1999). Segundamente, la bioprospección facilita la identificación de organismos mejor adaptados. Estos pueden integrarse directamente al sistema productivo, como en el caso de las microalgas para producción de lípidos (Larkum *et al.*, 2012) y microalgas extremófilas (Varshney *et al.*, 2015). Ambas ventajas se traducen, en última instancia, en ahorro de costos.

Sumado a lo anterior, este acercamiento no conlleva a problemáticas de conocimiento y propiedad intelectual, diferenciándose del primer tipo, ya que no se utiliza conocimiento local. De hecho, todos los ambientes son reservorios potenciales de genes y rutas metabólicas, como en una diversidad de río (Minhas *et al.*, 2016), lagos, plantas de tratamiento de aguas y arrozales (Han & Gu, 2013).

3. Consideraciones de la biosprospección orientada a la genómica

Pese a estas ventajas, se deben tomar en cuenta ciertas limitaciones y consideraciones. Por una parte, se deben identificar a los microorganismos de interés, en especial si no han sido previamente aislados o no son cultivables. Para ello, son necesarios recursos bioinformáticos y ómicos para identificar las secuencias de interés, al igual que tamizajes a gran escala (Bajhaiya *et al.*, 2016). Esto depende, a su vez, de la cantidad de recursos disponibles. Por ello, es más fácil continuar líneas de investigación con mayor información y recursos. El estudio de organismos raros requiere de trabajo previo de tamizaje, secuenciación, generación de librerías y análisis (Kajikawa *et al.*, 2015), así como un número mayor de genomas secuenciados y anotados (Cretoiu *et al.*, 2015).

Estas limitantes se reducen a medida que se realizan más estudios multi ómicos, ya que facilitan el proceso de identificación, acortan el tiempo y permiten la descripción del metabolismo (Mutanda *et al.*, 2011; Lauritano *et al.*, 2019). No obstante, la accesibilidad y costo de estas tecnologías es todavía restrictiva, obligando a utilizar otros acercamientos.

Por otra parte, si el organismo es cultivable, se debe evaluar sus condiciones de crecimiento y el rendimiento de los productos de interés. Esto debe realizarse *in vitro* e *in vivo*, trabajando la composición del medio (Bajhaiya *et al.*, 2016), la escala del proceso (de los Cobos-Vasconcelos *et al.*, 2016), si el sistema es cerrado o abierto y la interacción entre estos factores (Mutanda *et al.*, 2011). Por lo general, el proceso debe emular ciertas condiciones del ambiente de origen y evaluar más de una cepa aislada a la vez. Ambos aspectos elevan los costos y el alcance del esfuerzo.

4. Antecedentes en los estudios de microalgas enfocados en identificación

Tradicionalmente, se estudiaban las microalgas tóxicas por ser una amenaza de salud pública (Ebenezer *et al.*, 2012). Posteriormente, surgió la necesidad de distinguir microalgas de otra forma que no fuese morfológica, en especial para diatomeas. Se elaboraron diferentes métodos moleculares, inspirados del método de identificación molecular por “barcoding” en animales (Zou *et al.*, 2016; Hadi *et al.*, 2016). Sin embargo, no son aplicables al conjunto de todas las especies por su heterogeneidad y semejanza a plantas superiores.

No fue hasta la década de los años 70 que la bioprospección en microalgas despegó (Cheng & Odgen, 2011). Coincide con la investigación del potencial bioenergético y biotecnológico. Actualmente, los esfuerzos de bioprospección se concentran en la identificación, muestreo, proliferación y escalamiento de microalgas productoras de lípidos (Yang *et al.*, 2009; González-Davis *et al.*, 2012). Como tal, es común encontrar en la literatura investigaciones en las que aíslan, seleccionen y evalúen organismos aislados de forma local para producción de biocombustibles (Minhas *et al.*, 2016) y para estudiar la cinética de crecimiento. En menor escala, hay estudio de bioprospección y estudios moleculares de microalgas para forraje y otros tipos de alimento, pero sólo representan una fracción en comparación al estudio de producción

de lípidos (Rumin *et al.*, 2020). Más recientemente, ha surgido un interés por evaluar microalgas de florecimientos algales, por su tiempo de generación y uso de contaminantes antropogénicos (Pandey *et al.*, 2019).

5. Aspectos generales de las bases de datos

Las bases de datos son un conjunto virtual de datos ordenados que comparten una temática o propósito. En ciencias biológicas, se utilizan para organizar y vincular información de proteínas, secuencias nucleotídicas, agrupamientos taxonómicos, agrupamiento de especies en un bioma o artículos científicos, para citar unos ejemplos (Sayers *et al.*, 2019). Usualmente, la base de datos contiene un número importante de entradas, como en el caso de organización taxonómica, o un volumen importante de información por entrada, como en el caso de genomas (Buneman *et al.*, 2008). Esta organización es indispensable para fines biotecnológicos, ya que facilita el análisis de información generada previamente (Sharma *et al.*, 2013).

Un indicador de calidad de bases de datos es el grado de curación. Una base de datos curada es aquella con datos ordenados y sistematizados para facilitar la búsqueda y uso de los mismos. Resulta de la organización de información, revisión de fuentes y establecimiento de parámetros de clasificación. Una base de datos debe, idealmente, utilizar parámetros coherentes entre sí, ser fácil de leer para el usuario, ser entendible para los motores de búsqueda y presentar un formato útil para realizar análisis (Buneman *et al.*, 2008). En ciencias biológicas, es común encontrar herramientas de análisis integradas a las bases de datos, como BLAST en las bases de datos de NCBI (Johnson *et al.*, 2008).

Disponer de bases de datos óptimas es necesario para realizar ciertos tipos de investigación. A nivel de artículos, facilitan la sistematización de artículos para revisiones bibliográficas (Rumin *et al.*, 2020; Lu, 2011). A nivel taxonómico, facilitan estudios ecológicos y de biomasa. A nivel genómico y proteómico, las secuencias disponibles y debidamente anotadas permiten la comparación entre secuencias, determinación de filogenias y diseño de cebadores (Roumpeka *et al.*, 2017). Por ello, son conjuntos necesarios para un esfuerzo de identificación y bioprospección (Sharma *et al.*, 2013).

6. Situación actual de los recursos bioinformático de microalgas

Actualmente, existe una diversidad de bases de datos de microalgas, pero de alcance reducido. Las principales disponibles sirven para clasificar taxonómicamente a los microorganismos, como Algaebase y Algae DataBase (<http://algaedb.dpu.edu.in>) (Sharma *et al.*, 2013). Se enfocan en ordenar las publicaciones de descripciones de las especies, con poca información molecular. Existen otras bases de datos, como JGI y pico-PLAZA (Lauritano *et al.*, 2019) que cuentan adicionalmente con genes y genomas. Dichas bases tienden a estar curadas (Sharma *et al.*, 2013). Similarmente, se cuenta con bases de datos de regiones como de la subunidad 16S del ARN ribosomal, como *PhytoREF* para la identificación molecular y exclusiva de microalgas (Decelle *et al.*, 2015). También se cuenta con subconjuntos de secuencias de microalgas en bases de datos de talla mayor (Hadi *et al.*, 2016)

Adicionalmente, hay bases de datos que se enfocan en aspectos moleculares más específicos, como rutas metabólicas, interacciones proteicas (Sahoo *et al.*, 2020) y coexpresión de genes (Aoki *et al.*, 2016). Es de notar, sin embargo, que tienden a ser muy recientes, con 10 años o menos, y a enfocarse en producción de biocombustibles. La microalga *Chlamydomonas reinhardtii* es la microalga más estudiada, lo que se refleja por la disponibilidad de un gran número de bases de datos (May *et al.*, 2009). Esta especie, de gran interés biotecnológico, comparte pocas características con un gran número de especies. Como resultado, la información contenida en sus bases de datos tiene un uso limitado para la mayoría de especies.

En contraste, no se dispone de suficiente información o recursos bioinformáticos curados de ciertos grupos como protistas, lo que contribuye a menos investigaciones a nivel biotecnológico de las especies afectadas (Pascoal *et al.*, 2020). Usualmente, se dispone de información bruta, bajo la forma de genomas completos y reconocimiento de genes ortólogos, en particular en GenBank y *Genomes* de NCBI (Sayers *et al.*, 2019), pero requieren de mayor trabajo para que tengan aplicaciones subsecuentes.

F. Herramientas bioinformáticas

1. Alineamiento de secuencias: generalidades y tipos

El alineamiento de secuencias es una técnica bioinformática de comparación de secuencias nucleotídicas o proteicas. Consiste en comparar dos o más secuencias entre sí para observar la variabilidad entre estas según diferentes parámetros. Se utiliza con cuatro fines:

- Observar regiones conservadas
- Determinar posibles motivos en común
- Determinar si dos secuencias evolucionaron a partir de una secuencia ancestral
- Encontrar secuencias similares a la analizada en bases de datos

(Likic, 2008)

Para realizar un alineamiento, se requiere de un algoritmo de alineación. En resumen, estos consisten de dos pasos: primero, se establecen correspondencias entre elementos (aminoácidos y nucleótidos) y segundo, asignan un punteo a cada uno de estos. La correspondencia consiste en aparear cada elemento de cada secuencia hasta que todos lo estén. Si es necesario, se insertan espacios (gaps) entre elementos para realizarlo. El punteo funge como indicador de la calidad de la correspondencia (Huson *et al.*, 2010). Se basa en asignarle un valor a cada relación posible. Se otorga un valor mayor a una coincidencia, un valor menor a una coincidencia parcial y una penalización cuando hay espacios. En proteínas, se atribuye también un valor a cada tipo de relación según el parecido de las moléculas en talla, carga y estructura. Al conjunto de relaciones se les conoce como reglas de sustitución y se utilizan matrices para organizarlas. Estas reglas están sujetas al modelo que se utilice, al algoritmo y al tipo de elemento, por lo que las matrices se optimizan para realizar un alineamiento más útil (Likic, 2008).

Existen tres procedimientos de alineamiento de secuencias: alineamiento global, alineamiento local y alineamiento múltiple de secuencias. La principal diferencia entre estos es el conjunto de elementos que se comparan a la o las secuencias de interés. Solamente el tercer tipo, el alineamiento múltiple de secuencias, analiza más de dos de largo variable y entre sí, por lo que se describe únicamente este. Se utiliza cuando se estudian más de dos secuencias de interés (Likic, 2008), como es el caso de

superfamilias de proteínas y de genes con funciones similares. Existen dos tipos de algoritmo de alineamiento: progresivo e iterativo. En este trabajo, se empleó un algoritmo progresivo, realizado en tres etapas: construcción de la matriz de apareamiento de pares, la elaboración de secuencias de borrador y optimización del alineamiento por comparación de secuencias borrador. En cambio, los algoritmos iterativos

Alinear secuencias es necesario para el estudio y caracterización de estas, porque facilita la identificación de genes de interés y funciones de proteínas (Nalbantoğlu, 2014) y, por extensión, el diseño de mejores experimentos. Generalmente, de los alineamientos se obtienen secuencias consenso, los cuales son una combinación de las secuencias alineadas para observar las regiones y motivos conservados. Por lo anterior, es un paso obligatorio en los estudios de bioprospección y análisis filogenético.

G. Análisis filogenético

La filogenia es el estudio de la relación entre organismos por medio de características heredables, como el ADN. Permite retrazar el origen evolutivo de diferentes organismos. Tradicionalmente, se ha utilizado para realizar clasificaciones taxonómicas, pero también tiene aplicaciones para el estudio y comprensión de proteínas y genes (Hug *et al.*, 2016). Los resultados suelen expresarse como árboles filogenéticos que son un tipo de representación gráfica de la distancia y relaciones evolutivas hipotéticas entre las secuencias analizadas.

Es importante resaltar que los análisis filogenéticos siempre son posteriores a los datos considerados. Ningún árbol filogenético explica los eventos ni los fenómenos que dieron lugar a la distribución observada. Solamente presenta una hipótesis evolutiva a una distribución razonable de lo observado y pueda discutirse en base a ello. Sumado a esto, analizar un número elevado de secuencias es muy demandante en recursos informáticos, en especial si son secuencias mayores de 100 elementos (Huson *et al.*, 2010). Ambas condicionantes hacen la búsqueda del árbol real impráctica y larga, por lo que se utilizan modelos y algoritmos para reducir el tiempo, conduciéndose así una

búsqueda heurística de la mejor distribución posible (Gotoh *et al.*, 2014). Un tipo de árbol filogenético ampliamente utilizado son los basados en comparación de distancias.

1. Árboles basados en comparación de distancias:

Los árboles filogenéticos basados en la comparación de distancias emplean una matriz de sustitución para establecer las posibles relaciones entre las secuencias. En este acercamiento, la matriz de sustitución es el conjunto de relaciones entre los elementos considerados (aminoácidos o nucleótidos), es decir los elementos de las secuencias del alineamiento múltiple. Según estos, se cuantifica el grado de diferencia entre cada dos secuencias y se atribuye un punteo. El conjunto de puntajes entre secuencias sirve de base para realizar el agrupamiento (Gascuel & Steel, 2006). Por ello, son árboles no enraizados, es decir que no requieren de una secuencia de referencia para comparar. La principal ventaja es su velocidad y volumen de trabajo. Es significativamente más rápido que otros análisis filogenéticos, ya que procesa más de 1000 secuencias simultáneamente. Por ello, se ha usado para estudiar filos y taxones, y así determinar condiciones de crecimiento en la identificación de especies (Finkel *et al.*, 2016). Tradicionalmente, se ha utilizado en estudios filogenéticos de microalgas por la diversidad que presentan y la rapidez del método (Liu *et al.*, 2019; Li *et al.*, 2015; Kumari & Rathore, 2020)

Pese a estas ventajas, este método es considerablemente impreciso. Su forma de cálculo y los agrupamientos formados tienden a subestimar el grado de variación entre secuencias. Esto permite dar estimaciones conservadoras, pero se pierde información en el proceso. Por ello, se utiliza para visualizar preliminarmente un número importante de datos, pero no para datos con relaciones más claras y mejor estudiados. (Huson *et al.*, 2010). El método más conocido es el de *uniéndose de vecinos* (neighbor-joining). Dicho método minimiza la subestimación al substraer la distancia promedio entre agrupamientos, lo que contribuye a obtener estimaciones conservadoras. La calidad de los resultados también dependerá del modelo de sustitución utilizado y, en ocurrencia, el uso del método estadístico de bootstrapping (Hall, 2018).

2. Modelo de sustitución y método de bootstrapping

Para estimar más acertadamente la variación de secuencias emparentadas se debe tener una idea de cómo estas variaron originalmente. Existen diferentes juegos de premisas con los cuales modelar la estimación, pero actualmente se recurren a una combinación entre estas y evidencia empírica. La evidencia, derivada de alineamientos de varias familias de proteínas, es de suma importancia, ya que se fundamenta en tendencias y selecciones preferidas y de esto se deducen las posibles razones de mutación. Así, ciertos tipos de variación puntuales son más aceptadas que otras según la región considerada y se construyen matrices de sustitución más descriptivas de la realidad (Wang *et al.*, 2008). Uno de los modelos más utilizados es el de Jones-Taylor-Thornton (JTT), propuesto en los años 1970 y mejorado en 1992. Se elaboró en base a alineamientos de diferentes familias de proteínas y por ello es útil para grupos de secuencias variables. Permite comparar agrupaciones de las que no se disponga información, lo que contrarresta la estimación conservadora del modelo (Whelan & Goldman, 2001)

En paralelo al modelo de sustitución, se utilizan métodos estadísticos para verificar la calidad de los resultados obtenidos. En árboles de uniéndose de vecinos, es corriente utilizar el método de bootstrapping. Es un método estadístico de remuestreo que determina indirectamente la calidad de las bifurcaciones. Se parte del hecho que un árbol filogenético es una distribución de un conjunto y este se va subdividiendo en ramas. Cada rama se bifurca en un nodo. Entonces, para un árbol dado, existe un conjunto de nodos que dan una distribución de ramas. Lo que se pretende al realizar bootstrapping es comparar el árbol obtenido con otros árboles posibles, derivados del muestreo, y determinar la frecuencia de aparición de una bifurcación dada (Huson *et al.*, 2010). Se recomienda utilizar 1000 remuestreos para obtener un resultado adecuado y este se expresa mediante porcentajes de frecuencia asociados a las ramas. Cualquier porcentaje menor a 70% no es tomado seriamente, pues no se tiene certeza que la bifurcación de las ramas describa un fenómeno biológico (Hall, 2018). Tanto la matriz de sustitución como la técnica de bootstrapping permiten estimar la confianza en los resultados obtenidos. Debe resaltar que no permiten obtener el resultado, pero confieren un mayor rigor al análisis de datos.

3. Uso de técnicas moleculares

Tradicionalmente, se definen cinco grupos de sustancias para realizar análisis moleculares de microalgas: toxinas, proteínas, carbohidratos, ARN, ADN. Las secuencias genéticas y proteicas se utilizan más frecuentemente porque contienen una mayor cantidad de información. Pueden utilizarse para identificación taxonómica, caracterización de funciones y generación de productos de interés como enzimas, entre otras aplicaciones. Además, se utilizan porque son macromoléculas relativamente más estables, con el ADN la macromolécula más estudiada (Ebenezer *et al.*, 2012). La mayoría de técnicas aplicadas a microalgas derivan de otros organismos, en particular en proveniencia de plantas (Gimmler *et al.*, 2001).

4. Reacción en cadena de polimerasa (PCR)

La reacción en cadena de polimerasa (PCR por sus siglas en inglés) es una técnica de replicación *in vitro* de ADN. Emula, en condiciones de laboratorio, la replicación del material genético durante varios ciclos alternados de temperatura. Necesita de polimerasas, enzimas replicadoras de ADN, nucleótidos, ciertos iones y cebadores específicos que son secuencias oligonucleotídicas (Domingues, 2017). Los cebadores, en particular, son de importancia para la replicación del material genético, por lo que su diseño es esencial para la aplicación deseada (Elkins, 2015).

Las aplicaciones de la PCR son diversas. Se utiliza, por ejemplo, para clonar genes, cuantificar poblaciones, identificar mutaciones puntuales, identificar especies (Domingues, 2017). Esto se debe a la universalidad del ADN y la inclusión de un aspecto cuantitativo a la PCR con un diseño experimental adecuado. Para este trabajo, se expone sobre el diseño de cebadores y los tipos de PCR utilizados para identificación de especies.

5. Diseño de cebadores

El diseño de cebadores consiste en especificar las características del cebador a utilizar para una PCR satisfactoria. Se deben considerar dos aspectos para ello: las

generalidades y las características específicas de la variante de la PCR (Elkins, 2015;Thornton & Basu, 2011; Kumar & Chordia, 2015).

Las generalidades del diseño contemplan el largo y composición del cebador, lo que impacta directamente en sus propiedades fisicoquímicas. Un cebador, por lo general, debe ser de 18 a 22 pares de bases de largo. En este rango se combinan la especificidad del cebador al sitio de interés sin producir productos inespecíficos o estructuras secundarias. Se busca también que la secuencia no dé cebadores autocomplementarios o complementarios entre sí. Finalmente, la temperatura de anillamiento, un parámetro que determina el grado de separación del cebador y de unión a la región de interés, debe estar entre 58 y 62°C. Por estas consideraciones, se evita tener más de cuatro bases idénticas consecutivas o un porcentaje mayor al 50% de citosinas y guaninas en la composición del cebador (Elkins, 2015;Thornton & Basu, 2011; Kumar & Chordia, 2015). Los cebadores diseñados para variante de la PCR debe contemplar las consideraciones específicas de esta, como en el caso de la PCR degenerada.

6. PCR degenerada

La PCR degenerada es una variante de la reacción utilizada para identificar secuencias hipotéticas a partir de un grupo de secuencias ortólogas, sean proteicas o nucleotídicas (Smits *et al.*, 1999). Se utiliza principalmente para la detección de genes de interés en nuevos ambientes y genera secuencias de largo reducido. No permite la amplificación de los genes completos, sino solamente de secuencias conservadas, haciéndolo un método útil para exploración e identificación inicial (Cretoiu *et al.*, 2015).

En esta variante, se utilizan cebadores degenerados. Se caracterizan por tener una secuencia de alta afinidad a las regiones conservadas, pero no específica a ninguna, ya que se deben amplificar regiones con posibles mutaciones. Por ello, a la hora de diseñarlos, se priorizan las regiones con menos variación. Si se parte de una secuencia proteica, se busca mantener un grado de degeneración menor de 128. Resulta del producto de total de posibilidades de codones por aminoácido. Serina, leucina y arginina presentan seis codones, por lo que presentan un grado de degeneración de 16, mientras

que metionina y triptófano, como uno solo, tiene un valor de 1. Por ello, se prefieren regiones con aminoácidos con 1 o 2 grados de degeneración. Adicionalmente, se considera la afinidad de pares non WC para elegir el residuo más termodinámicamente favorable (Lang & Orgogozo, 2012).

7. La PCR como método de detección e identificación molecular.

Se reconoce a la PCR como un método de detección preciso. Esto se debe a la especificidad propia del ADN. No es una técnica de alto rendimiento, como la secuenciación de siguiente generación, pero es altamente preciso y si se combina con otras características de las regiones de interés, permiten la detección de especies, genes y rutas metabólicas en un número importante de especies (Smits *et al.*, 1999; Zou *et al.*, 2016). Una técnica derivada es la PCR de grupo específico.

8. PCR de grupo específico (Group-Specific PCR primers)

La PCR de grupo específico es una variante de la reacción que utiliza regiones altamente conservadas intercaladas con regiones altamente variables. Se utiliza para identificar una característica puntual en los organismos de interés y, por ello, es ampliamente utilizada en identificación de especies (Hadi., *et al.*, 2016).

Se diseñan los cebadores en base a regiones altamente conservadas, ya que la variación de la secuencia se encuentra entre estos. Como tal, no se contemplan consideraciones adicionales en el diseño. El esfuerzo de diseño se concentra, en su lugar, en la identificación y caracterización de la región a amplificar. Esta región debe estar altamente conservada, pero debe presentar las suficientes variaciones como para distinguir entre dos organismos de interés. Por ello, se busca amplificar la región con estas características y no por características funcionales (Ananthkrishnan *et al.*, 2013).

Se requiere de las secuencias de los genes de interés para seleccionar estas regiones y diseñar los cebadores. Posteriormente, se secuencian los amplicones generados por este método para identificar los organismos de una muestra. Se requiere, entonces, de un estudio previo de las regiones a amplificar para que esta técnica sea útil. Este acercamiento al diseño de cebadores ha sido utilizado para identificar miembros de una familia, como en el caso de *Chlorella* (Wan *et al.*, 2011), identificar genes de

oxidación de metano mediada por nitritos en ambientes anaerobio (Gaby & Buckley, 2012), y para la técnica de identificación por la subunidad 16s (Zou *et al.*, 2016).

VII. Metodología

A. Descripción y análisis de motivos y regiones conservadas de los genes de resistencia (ATPasa) obtenidos de las bases de datos.

1. Revisión de artículos científicos y trabajos publicados

Se realizó una revisión bibliográfica de investigaciones que reportaron especies alcalófilas y alcalino tolerantes según la metodología de Finkel *et al.* (2016). Se seleccionaron los términos “Alkaliphilic”, “microalgae”, “ATPase”, “alkali tolerant” “High pH”, “bicarbonate”, “alkalinity”, “alkaliphiles”, “protein membrane”, “cyanobacteria” para obtener investigaciones asociadas a las microalgas de interés. Se eligieron dichos términos ya que fueron los más relevantes o repetidos obtenidos de la investigación del marco teórico. Los resultados quedaron registrados en el siguiente formato:

Términos utilizados	total de resultados	artículos obtenidos	Especies

Posteriormente, se ordenaron los resultados del cuadro anterior en el siguiente cuadro, para obtener información relevante de las especies citadas:

Especies/género/cepa	Ambiente de origen de la especie	Temperatura	Rango de pH y variabilidad del mismo en el tiempo	Características del medio	Referencia bibliográfica

Finalmente, se realizó una revisión de los trabajos publicados en Guatemala sobre identificación y uso de microalgas. Se registraron los datos en el siguiente formato:

Especie	Tipo de trabajo	Título	Año	Comentarios adicionales

2. Investigación en las bases de datos

Para la obtención de las secuencias de genes y aminoácidos de ATPasas, se utilizó las siguientes bases de datos:

- Uniprot (<https://www.uniprot.org>) para las secuencias proteicas específicas.
- Genomes de NCBI (<https://www.ncbi.nlm.nih.gov/guide/genomes-maps/>) para obtener las secuencias nucleotídicas de las proteínas encontradas en Uniprot.
- European Nucleotide Archive: <https://www.ebi.ac.uk/ena>) por las mismas razones que Genomes.

Se utilizó los resultados de la revisión de artículos científicos y trabajos publicados para buscar las secuencias. Se tomaron las especies, genes citados y presencia en Guatemala. Los resultados de la búsqueda fueron guardados y sistematizados según en el siguiente cuadro:

Gen /proteína	Vínculo (URL)	Tamaño de la secuencia	Descripción dada por la base de datos	Número de la secuencia (referencia)	Referencia bibliográfica (si hay)

3. Clasificación preliminar

Se clasificaron las secuencias, de forma preliminar, con los siguientes parámetros:

- Tipo de secuencia (nucleotídica o peptídica)
- Largo de secuencia
- Clasificación taxonómica, principalmente por familia y género
- Variedades (aplicable únicamente a especies con alta diversidad y que estén secuenciadas)
- Con presencia confirmada previamente en trabajos de identificación local
- Función descrita en la base de datos
- Grado de curado de la secuencia

Finalmente, para estimar la disponibilidad de información para la comparación entre especies, se realizó el siguiente cuadro, mostrado en resultados:

Género	Total de genomas anotados disponibles	Tipo de organismo	Reportado en Guatemala

4. Alineamiento de secuencias nucleotídicas y proteicas

Se realizó un alineamiento de secuencias múltiple para determinar, preliminarmente, las relaciones entre secuencias y especies. Para ello, se compiló de primero las secuencias FASTA de un mismo género en un único archivo para su alineamiento. Se realizó el alineamiento en MegaX v.10.1.8 (Kumar *et al.*, 2018) con el algoritmo progresivo MUSCLE (Edgar *et al.*, 2004) con los parámetros por defecto, es decir una penalización de 400.00 a los gaps, con iteraciones de 2048 de memoria máxima y 16 iteraciones. Se utilizó el algoritmo UPGMA para realizar el agrupamiento (clustering). Dichos parámetros se mantuvieron en el resto del trabajo

B. Análisis filogenético y agrupamiento de proteínas expresadas por genes de resistencia por alineamiento múltiple de secuencias.

1. Exploración y optimización del análisis múltiple de secuencias

Se separaron los archivos alineados previamente en diferentes archivos de formato multifasta. Se partió del supuesto que secuencias de un mismo género y de largo similar cumplen la misma función, ya que se supuso que las secuencias depositadas pertenecían a especies emparentadas. Este acercamiento también facilitó ordenar las secuencias obtenidas en la sección A.1. Se realizó nuevamente el alineamiento de cada nuevo archivo y se separaron visualmente las secuencias en grupos según las regiones conservadas y variables. Si se observaron gaps mayores a 5 aminoácidos y con 50% de aminoácidos conservados, se separaron las secuencias distintas en un archivo diferente. Se realizó este proceso con cada archivo fasta hasta que compartieran 70% o más de aminoácidos comunes. La función de observación entre aminoácidos está incluida en el explorador de secuencias de MegaX. Se nombraron los archivos con el formato “género #número de grupo”. El número de grupo dependió del total de secuencias y sus largos por archivo, con agrupaciones de secuencias más cortas con número más pequeños.

2. Establecimiento de secuencias consenso y análisis de dominios

Para facilitar el análisis y condensar la cantidad de información a analizar, se elaboraron secuencias consenso de cada archivo multifasta. Para las proteínas, se utilizó

EMBOSS Cons (https://www.ebi.ac.uk/Tools/msa/emboss_cons/), con los parámetros de pluralidad, setcase e identidad por defecto. Se utilizó la matriz de sustitución BLOSUM62 para establecer las correlaciones. Estas secuencias fueron analizadas con la herramienta de NCBI Batch Web-CD Search Tool (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) para la identificación de regiones conservadas (Shennan & Lu, 2020). Se utilizaron los parámetros automáticos de búsqueda, utilizando el conjunto de bases de datos disponibles y con 0.01 como valor E esperado de predicción. Posteriormente, se agruparon las secuencias según los dominios observados y se alinearon para estudiar las secuencias más detalladamente. Se anotaron las secuencias altamente conservadas, ya sea porque el explorador de secuencias señalaba aminoácido conservado o por un grado mayor a 70% de conservación en todas las secuencias.

3. Análisis filogenético y reclasificación de secuencias

Se realizó el análisis filogenético para relacionar las secuencias alineadas y agrupadas en la sección anterior. Se construyó un árbol con el modelo de uniéndose de vecinos (neighbor-joining) asociado a test de *bootstrapping* con 1000 réplicas (Hall, 2018). Como modelo de sustitución se utilizó el de Jones-Taylor-Thornton. El resto de parámetros se mantuvieron por defecto. Se evaluó la calidad de los agrupamientos con el porcentaje de *bootstrapping* asociado a cada separación, tomando en cuenta aquellos mayores a 70%.

C. Diseño de cebadores degenerados y específicos para genes de resistencia alcalina en microalgas

1. Diseño de cebadores a partir de proteínas

Para diseñar los cebadores degenerados, se siguió la metodología de Lang y Orgozoso (2012). Se agruparon las secuencias consenso más emparentadas y se alinearon para resaltar regiones altamente conservadas. Se eligió una región con 7 a 8 aminoácidos los menos degenerados posibles y se comprobaron si presentaban un grado de degeneración menor a 128. Si fue menor, se procedió a diseñar el cebador con la

traducción más adecuada. De no identificarse ninguna región que cumpliera con este grado o que el porcentaje de *bootstrapping* no fuese mayor a 70%, se omitió este paso y se prosiguió con la sección 3.2

2. Elección de especies para diseño de cebadores específicos

En caso no se pudiera realizar la sección anterior, se eligieron especies de alto interés para diseñar cebadores específicos. Las especies se eligieron por abundancia de información genómica disponible, aplicación industrial posible y que haya sido aislada en Guatemala. Para el diseño de cebadores, se descargaron las secuencias genéticas de las proteínas estudiadas en la sección 2 de los archivos europeos de nucleótidos (<https://www.ebi.ac.uk/ena>) referenciados en sus entradas respectivas en Uniprot. Se utilizaron los grupos armados en la sección 2.1 para separar las secuencias. Se alinearon los codones en Mega X con el algoritmo MUSCLE, eliminando los gaps para realizarlo. De estos alineamientos se construyeron las secuencias consenso mediante CONS (<https://www.bioinformatics.nl/cgi-bin/emboss/cons>) que utiliza el mismo algoritmo que EMBOSS. Se utilizó DNAFULL como matriz de sustitución para realizarlo. Se diseñaron los cebadores a partir de las secuencias consenso obtenidas en Primer3Plus (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Untergasser *et al.*, 2007) y se analizaron los parámetros de estabilidad con IDT Oligoanalyzer 3.1.

D. Construcción de base de datos de microalgas alcalino tolerantes y alcalófilas con potencial industrial

1. Información utilizada para generar la base de datos

Se utilizaron los cebadores diseñados en la sección anterior como elementos de la base de datos. Se anotaron con las características de los amplicones que generen, como se describió en la sección anterior. Adicionalmente, se utilizaron los datos del cuadro realizado en la sección 1.2 para detallar la descripción de las secuencias.

2. Elementos utilizados para anotar las entradas de la base de datos

Se anotaron las secuencias conforme a características de los amplicones (Zeisel *et al.*, 2013; Wang *et al.*, 2012). Los datos se presentaron bajo el formato del cuadro a continuación:

Nombre de la secuencia consenso	Superfamilia a la que pertenece	Largo de la secuencia consenso	Nombre de los cebadores	Secuencia de los cebadores	Largo del amplicón (de haber)

3. Validación de las entradas de la base de datos

Para validar la utilidad de las entradas, se compararon las entradas con las siguientes consideraciones (Zeisel *et al.*, 2013):

- *Largo de los amplicones conforme al método*: del largo adecuado según la PCR recomendada en resultados y discusión. Se utilizó un simulador de gel de electroforesis en el proceso (<http://insilico.ehu.es/PCR/index2.php?mo=Synechococcus>). Se compararon los cebadores con 12 genomas de *Synechococcus* con un máximo de dos discordancias y una en el extremo 3'.
- *Amplificación de secuencias no objetivo*: si el cebador amplifica a un organismo no específico, se evaluó si dicho organismo se puede aislar del mismo ambiente que las microalgas de interés. Se utilizó BLASTn para realizarlo (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=

BlastSearch&LINK_LOC=blasthome). Todos los parámetros utilizados fueron los por defecto.

Se rediseñaron los cebadores si más de la mitad no cumplían con estos requerimiento

VIII. Resultados y discusión

Para el caso de la resistencia a condiciones alcalinas, se dispone de un volumen importante de información. En el Cuadro 1 se presenta información proteómica y genómica de 28 géneros, para un total de 207 secuencias aminoacídicas. Debe resaltarse, en primer lugar, que la mayoría de secuencias son procarióticas, cuatro secuencias pertenecientes a organismos eucariotas, o 2% del total. Se puede inferir que el esfuerzo de investigación y secuenciación de microalgas alcalinotolerantes y alcalófilas se ha concentrado en cianobacterias. Esto puede deberse a la diferencia de uso entre microalgas, ya que los géneros considerados oleaginosos, como *Chlorella* y *Nannochloropsis*, crecen óptimamente a pH levemente ácidos (Mutanda *et al.*, 2011). Se ha priorizado el rendimiento sobre las condiciones de crecimiento, por lo que la disponibilidad de secuencias de microalgas eucariotas alcalófilas y alcalinotolerantes es menor.

Cuadro 1: Resumen de información genómica y proteómica disponible en línea de especies de interés y aisladas en Guatemala. Los géneros con 5 o más genomas disponibles están resaltados en negrita

Género	Genomas secuenciados disponibles	Tipo de organismo	Aislado en cuerpos de agua dulce guatemaltecos
Anabaena	12	Procariota	No
Aphanothece	1	Procariota	No
Arthrospira	6	Procariota	No
Calothrix	7	Procariota	No
Chlamydomonas	1	Eucariota	No
Chlorella	1	Eucariota	Sí
Cyanobacterium	5	Procariota	No
Cyanobacteria bacterium	1	Procariota	No
Cyanothece	2	Procariota	No
Cylindrospermopsis	5	Procariota	No
Euhalothece	2	Procariota	No
Fistulifera solaris	1	Eucariota	No
Halothece	1	Procariota	No

Género	Genomas secuenciados disponibles	Tipo de organismo	Aislado en cuerpos de agua dulce guatemaltecos
Leptolyngbya	22	Procariota	No
Merismopedia	1	Procariota	No
Microcystis	50	Procariota	Sí
Nannochloropsis	1	Eucariota	No
Nodularia	1	Procariota	No
Nostoc	14	Procariota	No
Phaeodactylum	1	Procariota	No
Phormidium	1	Procariota	No
Pleurocapsa	1	Procariota	No
Pseudanabaena	1	Procariota	No
Synechococcus	61	Procariota	Sí
Synechocystis	2	Procariota	No
Thalassiosira pseudonana	1	Procariota	No
Oscillatoria	4	Procariota	Sí
Snowella	1	Procariota	No

Entre los géneros obtenidos, se encuentran las dos tendencias principales de investigación. El primero, con ejemplos como *Microcystis*, *Cylindrospermopsis* y *Anabaena* constituye el grupo de especies cianotóxicas y por ello se han secuenciado (Lines & Beardall, 2018). El segundo grupo, con *Synechococcus*, *Leptolyngba* y *Arthrospira* se han estudiado para optimizar sus condiciones de crecimiento y estudiar propiedades como captación de dióxido de carbono y abundancia relativa en ambientes alcalinos (Klanchui *et al.*, 2017; López-Archilla *et al.*, 2004). Es notable también la sobrerrepresentación de ciertos géneros, en particular *Synechococcus* y *Microcystis*, abarcando más del 50% del total de las secuencias. Debe mencionarse que la abundancia de datos no se traduce necesariamente en una aplicación posible o segura para cultivos a gran escala, en especial para el primer grupo. Adicionalmente, un volumen importante de investigaciones de cianobacterias solo se concentra en el aspecto fisiológico de las cianobacterias aisladas. Consecuentemente, la información disponible requiere de análisis más profundos para obtener resultados útiles. Incluso con lo anterior, este acervo de información puede facilitar la fase de pre investigación. Se ha mostrado que la diversidad de cianobacterias es similar entre cuerpos de agua altamente alcalinos, sin importar su región en el mundo, ya que las presiones ambientales son muy similares (Zorz *et al.*, 2019). Entonces, disponer de esta información, junto a los genomas, facilita la tarea de elección de especies o, por lo menos, brinda una referencia inicial para la elección de especies cultivables a gran escala.

A nivel de identificación de regiones conservadas, se obtuvo dos grupos de proteínas transportadoras por dominio, NhaP y KefB, usualmente situadas entre los aminoácidos 10 y 350 en las secuencias. Ambos pertenecen a la superfamilia de transportadores de cationes/protones CPA1, responsable de la extrusión de diferentes iones alcalinos y alcalinotérreos y de la incorporación de protones (Resch *et al.*, 2011). Se observó, preliminarmente, que ambos conjuntos son heterogéneos, con secuencias entre 400 a 800 aminoácidos. Se esperaba este resultado, ya que esta familia de transportadores es altamente diversa. Entre estos, aproximadamente un tercio presentaron dominios adicionales, usualmente TrA-K, Usp y los intercambiadores de sodio e hidrógeno tipo C (NaHC). Tanto TrA-K como NaHC están involucradas en la función de antiport, pero los más notorios son Usp y NaHC, pues median la respuesta celular en caso de un estrés ambiental. Se ha supuesto que se modulan conforme a señales de estrés intracelular, posiblemente por altas concentraciones de cationes

alcalinos (Jia *et al.*, 2018). Al momento de esta investigación, no se ha vinculado causalmente el estrés salino a estos antiports en cianobacterias, principalmente porque no se ha investigado, pero es evidente que hay una función mayor de este subconjunto en la homeostasis bajo condiciones anómalas. Pese a estas observaciones, no se pudo inferir otros dominios o funciones del resto de las secuencias. De hecho, 50% de las secuencias consenso de proteínas KefB y 66% de las secuencias de proteínas NhaP (sin símbolos asociados en las imágenes 1 y 2) están constituidas únicamente por el dominio del transportador, pudiendo representar hasta 90% de la secuencia. El análisis bioinformático de este trabajo no permitió dilucidar características más específicas del funcionamiento de estas proteínas, porque el algoritmo utilizado, RPS-BLAST, compara elementos específicos de las secuencias con las bases de datos utilizadas. No se toman en cuenta secuencias de aminoácidos seguidos a menos que sean los mismos que los de la secuencia comparada. Se puede suponer que aminoácidos específicos están conservados al ser de importancia en la cohesión y estructura de la proteína (Resch *et al.*, 2011). Sumado a esto, las características específicas de la proteína *in vivo* resultan difíciles de estudiar y predecir *in vitro* e *in silico*, ya que sus funciones dependen del estado fisiológico de la cianobacteria y de la interacción con otros sistemas intracelulares. Experimentalmente, se ha observado que antiports, teóricamente capaces de trasladar a ambos lados diferentes iones, demuestran preferencias a ciertos iones según el pH (Dawut *et al.*, 2018). Por ello, en una misma célula, diferentes *antiports* operan en simultáneo según el pH, los nutrientes disponibles, la forma de la célula y la concentración de iones tóxicos (como litio) en el ambiente (Wei *et al.*, 2007). Esto se ha corroborado en diferentes especies de microalgas y para las superfamilias NhaP y KefB (Jangir *et al.*, 2019).

Posteriormente, se evaluó la relación en cada grupo de proteínas mediante un árbol filogenético construido con el método de uniéndose de vecinos (Neighbor-Joining), - ver Figuras 1 y 2. . De forma general, se observa que los antiports con regiones adicionales (NaHC, TrkA_N, TrkA_C y Usp) están agrupados - tanto en el árbol KefB como en el NhaP - con las dominios TrAK_N (rojo), TraK_C (azul), Usp o Usp_Like (morado) y Na_H_Exchanger_C (NaHC; verde). En el árbol filogenético de la superfamilia NhaP (Figura 2), se observa un ordenamiento similar de genes ortólogos en las regiones Usp y TrA_K. Así, *Anabaena* está más emparentado con *Nostoc*, luego con *Arthrospira*, posteriormente con *Synechococcus* y finalmente con *Euhalothece*. No

obstante, el porcentaje de *bootstrapping* no alcanza 70% para la mayoría de separaciones de ramas en los dos árboles por lo que la mayoría de relaciones presentadas son entonces de baja confianza. El criterio se satisface sobre todo entre especies de un mismo orden, como el caso entre *Anabaena*, *Nostoc* y *Calothrix* que figuran en el orden *Nostocales*. La falta de confianza se debe, nuevamente, a la alta diversidad en estas superfamilias de proteínas. Como se ha mencionado antes, los antiports catión-protón son altamente variables y los aminoácidos conservados pueden ser un fragmento largo de la secuencia (Resch *et al.*, 2011). El método de *bootstrapping* no considera factores evolutivos, ya que realiza una estimación de la distribución estadística según la comparación de sección de secuencias aleatoriamente remuestreadas (Hall, 2018). En el caso presente, el alto grado de variabilidad se observó por los porcentajes reducidos de la estructura proporciona poca información relevante, como en la rama con proteínas con la región TrAK_N – ver Figura 2 –. Debe resaltar que se observan algunas excepciones, por ejemplo en el grado de cercanía entre *Cyanothece* 5 y *Calothrix* 16 en la Figura 1, *Arthrospira* 4, *Calothrix* 8, *Nostoc* 8 y *Anabaena* 5 en la Figura 2. Esta última relación es particularmente interesante, ya que implica especies de dos órdenes distintos, presentan un antiport de composición similar y probablemente de función similar, aunque debe probarse empíricamente para confirmar esta observación.

Figura 1: Árbol filogenético de las secuencias de la superfamilia KefB realizado con test de filogenia de bootstrapping de 1000 réplicas y con el modelo de sustitución de Jones-Taylor-Thornton. En rojo se encuentran los antiports TrAK_N, en azul los que presentan TraK_C, morado Usp o Usp_Like y verde Na_H_Exchanger_C (NaHC)

(Construido en MegaX)

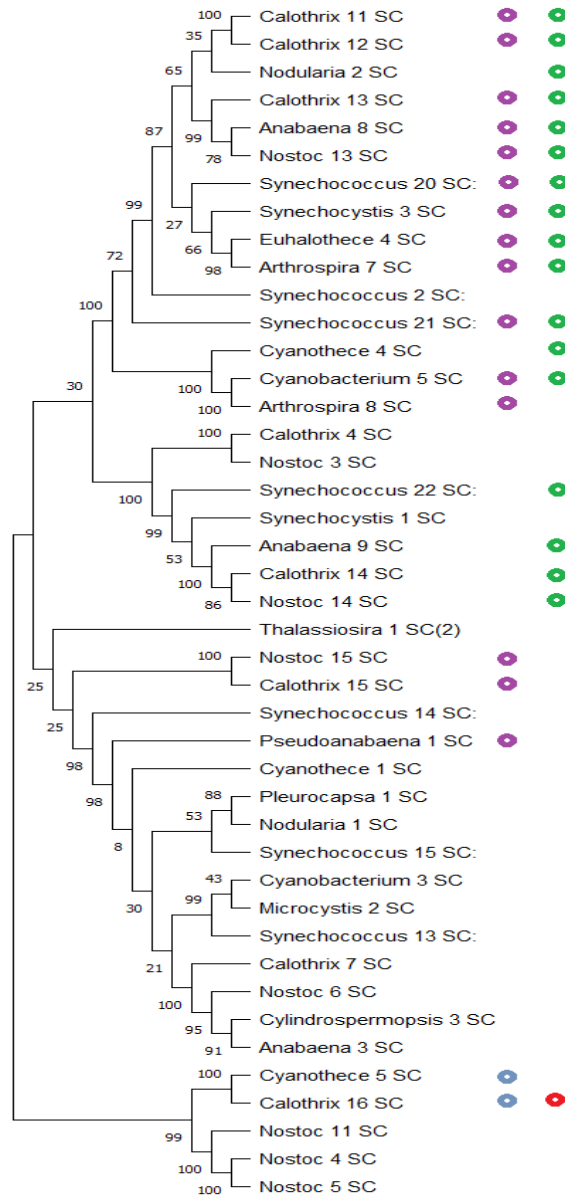
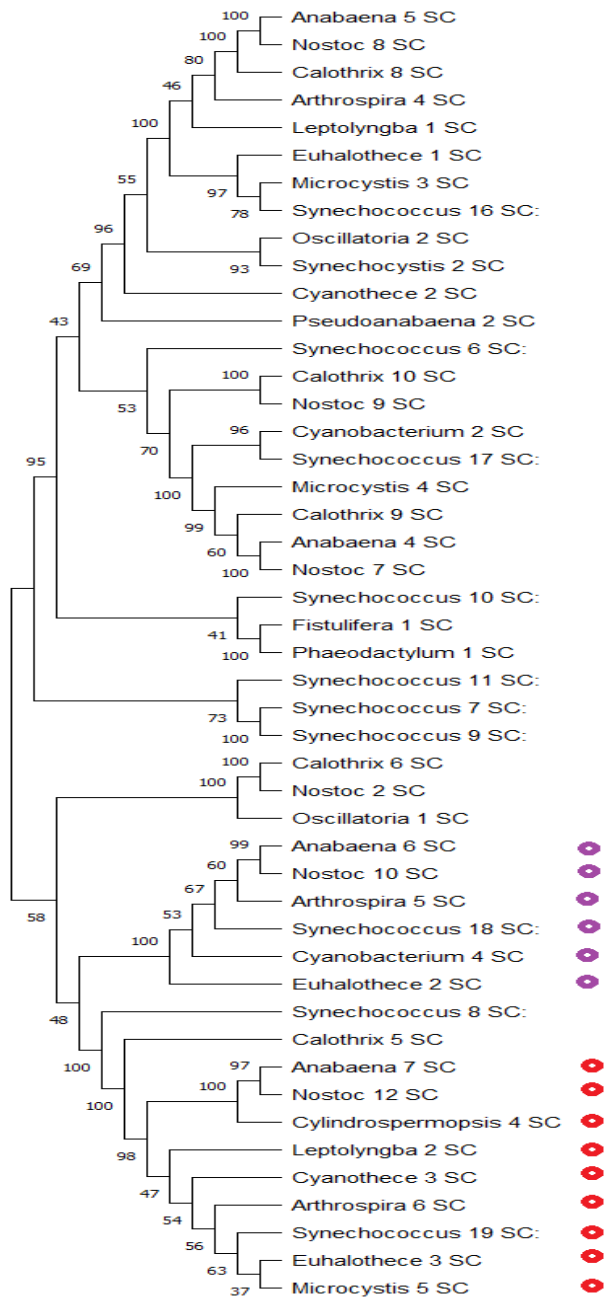


Figura 2: Árbol filogenético de las secuencias de la superfamilia NhaP realizado con test de filogenia de bootstrapping de 1000 réplicas y con el modelo de sustitución de Jones-Taylor-Thornton . Como modelo de sustitución se utilizó Jones-Taylor-Thornton. En rojo se encuentran los antiports TrAK_N y en morado Usp o Usp_Like (construido con Mega X)



los aminoácidos apolares constituyen los plegamientos intermembranarios que interactúan con las cadenas alifáticas de los ácidos grasos. En cambio, los aminoácidos cargados (puntuales o un motivo) están en la superficie polar de pasaje (la cavidad). Las regiones apolares pueden ser altamente variables, ya que por sus propiedades son intercambiables entre sí. Consecuentemente, estas proteínas pueden presentar regiones altamente variables intercaladas y que no brindan información específica sobre esta. En contraste, pueden resaltarse ciertas regiones altamente conservadas en todo un conjunto, como el motivo EGESXXNDX en la segunda región de KefB en la Figura 5. Se encuentra en pleno dominio del transportador y está presente independientemente de regiones adicionales y talla de la proteína y no es exclusiva a cianobacterias, como se resalta en revisiones anteriores (Resch *et al.*, 2011). Con las herramientas de este trabajo, no se puede determinar la función específica y la literatura disponible se enfoca en aminoácidos puntuales. Solo se puede suponer que tiene un rol específico en el transporte de cationes en esta familia y se ha conservado por el origen monofilético de las cianobacterias se conservó dicha región.

Estas observaciones respaldan el porcentaje bajo de *bootstrapping* obtenido en el análisis filogenético, ya que las proteínas son altamente variables. Sumado a esto, las regiones más variables, las alifáticas, presentan los codones más degenerados según Lang y Orgogozo (2012), por lo que los intentos de diseño de cebadores degenerados no satisface un grado de degeneración menor a 128. En otros organismos, como en plantas, se han diseñado exitosamente cebadores para este propósito (Olías *et al.*, 2009), pero fue gracias al grado de conservación más elevado entre estas, lo que no resulta posible en cianobacterias. Por lo tanto, no se puede diseñar un PCR de grupo específico para cianobacterias a partir de esta característica. Pese a esta imposibilidad, se ha evidenciado en esta discusión que el estudio de cianobacterias alcalófilas y alcalinotolerantes debe realizarse de especies en especie, incluso si se encuentran múltiples especies en un mismo ambiente ecológico. Este acercamiento ha cobrado interés recientemente, como lo evidencia un estudio en *A. platensis* que buscó caracterizar los antiports sodio hidrógeno y sus niveles de expresión de los transportadores en función del pH (Jangir *et al.*, 2019). Esto puede extrapolarse a otras especies, como *Synechococcus* al disponerse de una cantidad importante de información y antecedente de aislamiento en Guatemala (Morales-Molina, 2016). Por ello, se proponen los siguientes cebadores preliminarmente para una PCR en tiempo real de

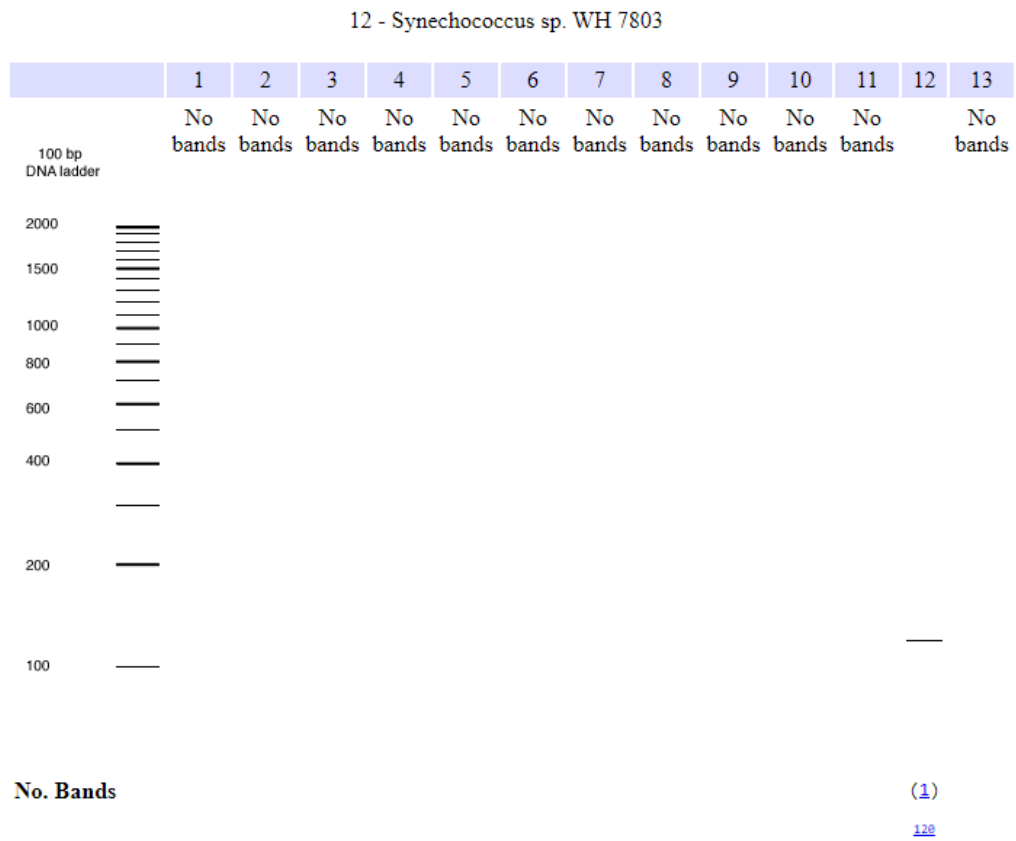
ADNc y así medir la expresión de los genes de antiports. Todos producen amplicones no mayores a 150 pares de bases.

Cuadro 2: Cebadores propuestos para agrupaciones similares de antiports de *Synechococcus* analizados. En negrita se resaltan los cebadores que dieron resultados

Número de la secuencia consenso	Superfamilia a la que pertenece	Largo de la secuencia proteica consenso	Nombre de los cebadores	Secuencia de los cebadores (de 5' a 3')	Largo del amplicón (de haber) y cepa de <i>Synechococcus</i> asociada
Synechococcus 7	NhaP	422	Synec_7-F	AGTGCGAAT CGGGCTTCA A	120 (WH 7803)
			Synec_7-R	GAAACTGAC GCACCACAT CG	
Synechococcus 8	NhaP	438	Synec_8-F	GCTGTTTGCG ATCCGTCTG	Sin bandas
			Synec_8-R	TCAGGATCGT CAGAAACACC A	
Synechococcus 9	NhaP	445	Synec_9-F	AGTTTCATCT GCTGGACGGG	Sin bandas
			Synec_9-R	CGCAAAGCCC TGAATCAACA	
Synechococcus 11	NhaP	445	Synec_11-F	GATTTGGTTA TGTCGCCCGC	Sin bandas
			Synec_11-R	TGTGTGCAGC CGTAATCCAT	
Synechococcus 13	KefB	466	Synec_13-F	TGAGAGTCCT GAAGCCTCGA	Sin bandas

			Synec_13-R	GGGAGATTAA TACGGGCGCA	
Synechococcus 14	KefB	471	Synec_14-F	GTGTTTGCA GGGGTTGGT TC	132 (PCC 7002)
			Synec_14-R	CGCCTGGGG TTCCTTAAAG A	
Synechococcus 15	KefB	506	Synec_15-F	CCCTCACGG CAACCAGTA TT	95 (PCC 7002)
			Synec_15-R	CGCGGCACC AACGATAAT TT	
Synechococcus 16	NhaP	536	Synec_16-F	CTGTTGGTG ATTGTGGGG CT	107 (PCC 7002)
			Synec_16-R	GCTTCAAAC AAGAGCGGT GG	
Synechococcus 17	NhaP	551	Synec_17-F	TCGGTTTATC CTCTGCTGT CG	98 (PCC 7002)
			Synec_17-R	AGAGAGCCT TTGACATTG CC	
Synechococcus 20	KefB	709	Synec_20-F	GCCGACCTA GAAAATGCC CT	141 (PCC 7002)
			Synec_20-R	GTAACGGGT GGAGTGCAG AA	
Synechococcus 22	KefB	718	Synec_22-F	GTCTGTGGC AGCAAGATC CT	141 (PCC 6301) 141 (PCC 7942)
			Synec_22-R	GACAGGCTT CAACCAGAC GA	

Figura 7: Imagen de la PCR *in silico* del par de cebadores asociados a la secuencia consenso *Synechococcus_7*. Las otras imágenes se encuentran en anexos



IX. Conclusiones

En conclusión, no se pudo diseñar cebadores a partir de genes de resistencia a condiciones alcalinas para identificar más de una especie de interés, ya que el grado de variabilidad no lo permite.

1. Se determinó por medio del análisis de las secuencias proteicas de los antiports que estas pertenecen a las superfamilias de transportadores KefB y NhaP. Presentan pocas regiones altamente conservados, así como aminoácidos puntuales conservados y además de regiones variables, pero con características polares similares
2. El análisis filogenético de los agrupamientos de secuencias alineadas reveló que los antiports están distantemente emparentados, incluso si ciertas características de resistencia parecen presentar un origen monofilético. La información analizada no permitió establecer una distribución comprensible de las secuencias más allá de ciertos órdenes.
3. No se pudo diseñar cebadores degenerados y específicos para diferentes especies de microalgas por la variabilidad alta entre secuencias y abundancia de regiones de aminoácidos altamente degenerados. En contraste, se diseñaron cebadores para el género *Synechococcus* por ser aislable, presentar potencial industrial y por contar con recursos bioinformáticos vastos.
4. Se generó una base de datos de cebadores para microalgas del género *Synechococcus* para con las secuencias proteicas analizadas en este trabajo, ya que se ha reportado en cuerpos de agua dulce de Guatemala, contar con cepas estudiadas y adquiribles y por presentar potencial industrial.

X. Recomendaciones

Se recomienda realizar el estudio de una sola especie alcalófila o alcalino tolerante, preferiblemente del género *Synechococcus* por las facilidades de trabajo que este tiene, para determinar expresión de los genes de resistencia según las condiciones ambientales. De quererse aislar localmente, se recomienda que se haga en un cuerpo de agua dulce y de condiciones alcalinas. Asimismo, es recomendable estudiar más a fondo las características de los antiports de la especie aislada, ya que facilita la optimización del medio y otros procesos para un escalamiento a futuro.

XI. Bibliografía

Abdelgawad, B. W. (2012). The Bt brinjal case: *The first legal action against monsanto and its Indian collaborators for biopiracy*. *Biotechnology Law Report*, 31(2), 136-139.

Ananthkrishnan, G., Choudhary, N., Roy, A., Sengoda, V. G., Postnikova, E., Hartung, J. S., ... & Brlansky, R. H. (2013). *Development of primers and probes for genus and species specific detection of 'Candidatus Liberibacter species' by real-time PCR*. *Plant disease*, 97(9), 1235-1243.

Andersen, R. A. (Ed.). (2005). *Algal culturing techniques*. Elsevier.

Aoki, Y., Okamura, Y., Ohta, H., Kinoshita, K., & Obayashi, T. (2016). *ALCOdb: gene coexpression database for microalgae*. *Plant and Cell Physiology*, 57(1), e3-e3.

Armesto, P., Infante, C., Cousin, X., Ponce, M., & Manchado, M. (2015). *Molecular and functional characterization of seven Na⁺/K⁺-ATPase β subunit paralogs in Senegalese sole (*Solea senegalensis* Kaup, 1858)*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 182, 14-26.

Bajhaiya, A. K., Dean, A. P., Driver, T., Trivedi, D. K., Rattray, N. J., Allwood, J. W. & Pittman, J. K. (2016). *High-throughput metabolic screening of microalgae genetic variation in response to nutrient limitation*. *Metabolomics*, 12(1), 9.

Bellinger, E. G., & Sigeo, D. C. (2015). *Freshwater algae: identification, enumeration and use as bioindicators*. John Wiley & Sons.

Buneman, P., Cheney, J., Tan, W. C., & Vansummeren, S. (2008, June). *Curated databases*. In *Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (pp. 1-12).

Cembella, A. D. (2003). *Chemical ecology of eukaryotic microalgae in marine ecosystems*. *Phycologia*, 42(4), 420-447.

- Cheng, K. C., & Ogden, K. L. (2011). *Algal biofuels: the research*. American Institute of Chemical Engineers (AIChE).
- Chi, Z., Elloy, F., Xie, Y., Hu, Y., & Chen, S. (2014). *Selection of microalgae and cyanobacteria strains for bicarbonate-based integrated carbon capture and algae production system*. *Applied biochemistry and biotechnology*, 172(1), 447-457.
- Cretoiu, M. S., Berini, F., Kielak, A. M., Marinelli, F., & van Elsas, J. D. (2015). *A novel salt-tolerant chitobiosidase discovered by genetic screening of a metagenomic library derived from chitin-amended agricultural soil*. *Applied microbiology and biotechnology*, 99(19), 8199-8215.
- Dawut, K., Sirisattha, S., Hibino, T., Kageyama, H., & Waditee-Sirisattha, R. (2018). *Functional characterization of the NhaA Na⁺/H⁺ antiporter from the green picoalga *Ostreococcus tauri**. *Archives of biochemistry and biophysics*, 649, 37-46.
- de los Cobos-Vasconcelos, D., García-Cruz, E. L., Franco-Morgado, M., & González-Sánchez, A. (2016). *Short-term evaluation of the photosynthetic activity of an alkaliphilic microalgae consortium in a novel tubular closed photobioreactor*. *Journal of applied phycology*, 28(2), 795-802.
- Decelle, J., Romac, S., Stern, R. F., Bendif, E. M., Zingone, A., Audic, S. & Gourvil, P. (2015). *Phyto REF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy*. *Molecular ecology resources*, 15(6), 1435-1445.
- Domingues, L. (Ed.). (2017). *PCR: methods and protocols*. Springer New York.
- Ebenezer, V., Medlin, L. K., & Ki, J. S. (2012). *Molecular detection, quantification, and diversity evaluation of microalgae*. *Marine Biotechnology*, 14(2), 129-142.
- Edgar, R. C. (2004). *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. *Nucleic acids research*, 32(5), 1792-1797.

- Elhag, O., Zhou, D., Song, Q., Soomro, A. A., Cai, M., Zheng, L. & Zhang, J. (2017). *Screening, expression, purification and functional characterization of novel antimicrobial peptide genes from Hermetia illucens (L.)*. PloS one, 12(1), e0169582.
- Elkins, K. M. (2015). *Primer design for PCR reactions in forensic biology*. In *PCR Primer Design* (pp. 17-30). Humana Press, New York, NY.
- Estrada Palencia, E. G. (2013). *Evaluación del potencial uso de biomasa de las diferentes cepas de microalgas del Atlántico de Guatemala en diferentes áreas de la industria* (Doctoral dissertation, Universidad de San Carlos de Guatemala).
- Finkel, Z. V., Follows, M. J., Liefer, J. D., Brown, C. M., Benner, I., & Irwin, A. J. (2016). *Phylogenetic diversity in the macromolecular composition of microalgae*. PLoS One, 11(5), e0155977.
- Gaby, J. C., & Buckley, D. H. (2012). *A comprehensive evaluation of PCR primers to amplify the nifH gene of nitrogenase*. PloS one, 7(7), e42149.
- Galarza-Muñoz, G., Soto-Morales, S. I., Holmgren, M., & Rosenthal, J. J. (2011). *Physiological adaptation of an Antarctic Na⁺/K⁺-ATPase to the cold*. Journal of Experimental Biology, 214(13), 2164-2174.
- García, J. L., de Vicente, M., & Galán, B. (2017). *Microalgae, old sustainable food and fashion nutraceuticals*. Microbial Biotechnology, 10(5), 1017-1024.
- Gascuel, O., & Steel, M. (2006). *Neighbor-joining revealed*. Molecular biology and evolution, 23(11), 1997-2000.
- Gençkal, H., & Tari, C. (2006). *Alkaline protease production from alkalophilic Bacillus sp. isolated from natural habitats*. Enzyme and Microbial Technology, 39(4), 703-710.
- Gimmler, H., & Degenhard, B. (2001). *Alkaliphilic and alkali-tolerant algae*. In *Algal adaptation to environmental stresses* (pp. 291-321). Springer, Berlin, Heidelberg.
- Gogarten, J. P., & Olendzenski, L. (1999). *Orthologs, paralogs and genome comparisons*. Current opinion in genetics & development, 9(6), 630-636.

- González-Davis, O., Ponce-Rivas, E., Sánchez-Saavedra, M. D. P., Muñoz-Márquez, M. E., & Gerwick, W. H. (2012). *Bioprospection of microalgae and cyanobacteria as biocontrol agents against Vibrio campbellii and their use in white shrimp Litopenaeus vannamei culture*. Journal of the World Aquaculture Society, 43(3), 387-399.
- Gotoh, O. (2014). *Heuristic Alignment Methods*. In *Multiple Sequence Alignment Methods* (pp. 29-43). Humana Press, Totowa, NJ.
- Guisado, I. M., Purswani, J., Gonzalez-Lopez, J., & Pozo, C. (2015). *Physiological and genetic screening methods for the isolation of methyl tert-butyl ether-degrading bacteria for bioremediation purposes*. International Biodeterioration & Biodegradation, 97, 67-74.
- Hadi, S. I., Santana, H., Brunale, P. P., Gomes, T. G., Oliveira, M. D., Matthiensen, A. & Brasil, B. S. (2016). *DNA barcoding green microalgae isolated from neotropical inland waters*. PloS one, 11(2), e0149284.
- Hall, B. G. (2018). *Phylogenetic trees made easy: A how to manual (No. 576.88 H174p)*. Sinauer.
- Han, P., & Gu, J. D. (2013). *A newly designed degenerate PCR primer based on pmoA gene for detection of nitrite-dependent anaerobic methane-oxidizing bacteria from different ecological niches*. Applied microbiology and biotechnology, 97(23), 10155-10162.
- Hoffmann, R. D., & Palmgren, M. (2016). *Purifying selection acts on coding and non-coding sequences of paralogous genes in Arabidopsis thaliana*. BMC genomics, 17(1), 1-13.
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J. & Suzuki, Y. (2016). *A new view of the tree of life*. Nature microbiology, 1(5), 1-6.
- Huson, D. H., Rupp, R., & Scornavacca, C. (2010). *Phylogenetic networks: concepts, algorithms and applications*. Cambridge University Press.
- Jangir, M. M., Vani, B., & Chowdhury, S. (2019). *Analysis of seven putative Na⁺/H⁺ antiporters of Arthrospira platensis NIES-39 using transcription profiling and in silico*

studies: an indication towards alkaline pH acclimation. *Physiology and Molecular Biology of Plants*, 25(5), 1175-1183.

Jia, Q., Zheng, C., Sun, S., Amjad, H., Liang, K., & Lin, W. (2018). *The role of plant cation/proton antiporter gene family in salt tolerance*. *Biologia plantarum*, 62(4), 617-629.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). *NCBI BLAST: a better web interface*. *Nucleic acids research*, 36(suppl_2), W5-W9.

Kajikawa, M., Kinohira, S., Ando, A., Shimoyama, M., Kato, M., & Fukuzawa, H. (2015). *Accumulation of squalene in a microalga Chlamydomonas reinhardtii by genetic modification of squalene synthase and squalene epoxidase genes*. *PLoS One*, 10(3), e0120446.

Klanchui, A., Cheevadhanarak, S., Prommeenate, P., & Meechai, A. (2017). *Exploring components of the CO₂-concentrating mechanism in alkaliphilic cyanobacteria through genome-based analysis*. *Computational and structural biotechnology journal*, 15, 340-350.

Komárek, J., Zapomělová, E., Šmarda, J., Kopecký, J., Rejmánková, E., Woodhouse, J. & Komarkova, J. (2013). *Polyphasic evaluation of Limnoraphis robusta, a water-bloom forming cyanobacterium from Lake Atitlán, Guatemala, with a description of Limnoraphis gen. nov.* *Fottea*, 13(1), 39-52.

Kumar, A., & Chordia, N. (2015). *In silico PCR primer designing and validation*. In *PCR primer design* (pp. 143-151). Humana Press, New York, NY.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). *MEGA X: molecular evolutionary genetics analysis across computing platforms*. *Molecular biology and evolution*, 35(6), 1547-1549.

Kumari, J., & Rathore, M. S. (2020). *Na⁺/K⁺-ATPase a Primary Membrane Transporter: An Overview and Recent Advances with Special Reference to Algae*. *Journal of membrane biology*.

- Kuo, C. M., Jian, J. F., Sun, Y. L., Lin, T. H., Yang, Y. C., Zhang, W. X. & Lin, C. S. (2018). *An efficient Photobioreactors/Raceway circulating system combined with alkaline-CO₂ capturing medium for microalgal cultivation*. *Bioresource technology*, 266, 398-406.
- Lang, M., & Orgogozo, V. (2012). *Identification of homologous gene sequences by PCR with degenerate primers*. In *Molecular Methods for Evolutionary Genetics* (pp. 245-256). Humana Press.
- Larkum, A. W., Ross, I. L., Kruse, O., & Hankamer, B. (2012). *Selection, breeding and engineering of microalgae for bioenergy and biofuel production*. *Trends in biotechnology*, 30(4), 198-205.
- Lauritano, C., Ferrante, M. I., & Rogato, A. (2019). *Marine natural products from microalgae: An-omics overview*. *Marine drugs*, 17(5), 269.
- Li, L., Cui, J., Liu, Q., Ding, Y., & Liu, J. (2015). *Screening and phylogenetic analysis of lipid-rich microalgae*. *Algal Research*, 11, 381-386.
- Likic, V. (2008). *The Needleman-Wunsch algorithm for sequence alignment*. Lecture given at the 7th Melbourne Bioinformatics Course, Bi021 Molecular Science and Biotechnology Institute, University of Melbourne, 1-46.
- Lines, T., & Beardall, J. (2018). *Carbon acquisition characteristics of six microalgal species isolated from a subtropical reservoir: potential implications for species succession*. *Journal of phycology*, 54(5), 599-607.
- Liu, C., Liu, J., Hu, S., Wang, X., Wang, X., & Guan, Q. (2019). *Isolation and identification of a halophilic and alkaliphilic microalgal strain*. *PeerJ*, 7, e7189
- Liu, J., Liu, J., Chen, A., Ji, M., Chen, J., Yang, X., ... & Xu, G. (2016). *Analysis of tomato plasma membrane H⁺-ATPase gene family suggests a mycorrhiza-mediated regulatory mechanism conserved in diverse plant species*. *Mycorrhiza*, 26(7), 645-656.
- Lobo de León, J.A. (2018). *Diseño y optimización de un medio de cultivo para el crecimiento de *Arthrospira platensis* en condiciones de laboratorio (tesis de licenciatura no publicada)*. Universidad del Valle de Guatemala, Guatemala.

- Lopez-Archilla, A. I., Moreira, D., López-García, P., & Guerrero, C. (2004). *Phytoplankton diversity and cyanobacterial dominance in a hypereutrophic shallow lake with biologically produced alkaline pH*. *Extremophiles*, 8(2), 109-115.
- Lu, Z. (2011). *PubMed and beyond: a survey of web tools for searching biomedical literature*. Database, 2011.
- Malcata, F. X., Pinto, I. S., & Guedes, A. C. (2018). *Marine Macro-and Microalgae: An Overview*. CRC Press.
- May, P., Christian, J. O., Kempa, S., & Walther, D. (2009). *ChlamyCyc: an integrative systems biology database and web-portal for Chlamydomonas reinhardtii*. *Bmc Genomics*, 10(1), 209.
- Minhas, A. K., Hodgson, P., Barrow, C. J., Sashidhar, B., & Adholeya, A. (2016). *The isolation and identification of new microalgal strains producing oil and carotenoid simultaneously with biofuel potential*. *Bioresource technology*, 211, 556-565.
- Morales Molina, S.E. (2016). *Diversidad de macroinvertebrados acuáticos y microalgas bentónicas como indicadores de la calidad de agua del río Pansalic, reserva forestal protectora de manantiales cordillera Alux* (Doctoral dissertation, Universidad de San Carlos de Guatemala).
- Mutanda, T., Ramesh, D., Karthikeyan, S., Kumari, S., Anandraj, A., & Bux, F. (2011). *Bioprospecting for hyper-lipid producing microalgal strains for sustainable biofuel production*. *Bioresource technology*, 102(1), 57-70.
- Nalbantoğlu, Ö. U. (2014). *Dynamic programming*. In *Multiple Sequence Alignment Methods* (pp. 3-27). Humana Press, Totowa, NJ.
- Neofotis, P., Huang, A., Sury, K., Chang, W., Joseph, F., Gabr, A., ... & Polle, J. E. (2016). *Characterization and classification of highly productive microalgae strains discovered for biofuel and bioproduct generation*. *Algal Research*, 15, 164-178.
- Olias, R., Eljakaoui, Z., Li, J. U. N., de Morales, P. A., Marín-Manzano, M. C., Pardo, J. M., & Belver, A. (2009). *The plasma membrane Na⁺/H⁺ antiporter SOS1 is essential*

for salt tolerance in tomato and affects the partitioning of Na⁺ between plant organs. Plant, cell & environment, 32(7), 904-916.

Palmer, C. M. (1969). *A composite rating of algae tolerating organic pollution 2.* Journal of Phycology, 5(1), 78-82.

Pandey, M. K., Dasgupta, C. N., Mishra, S., Srivastava, M., Gupta, V. K., Suseela, M. R., & Ramteke, P. W. (2019). *Bioprospecting microalgae from natural algal bloom for sustainable biomass and biodiesel production.* Applied microbiology and biotechnology, 103(13), 5447-5458.

Pascoal, F., Magalhães, C., & Costa, R. (2020). *The Link Between the Ecology of the Prokaryotic Rare Biosphere and Its Biotechnological Potential.* Frontiers in Microbiology, 11, 231.

Popova, L. G., and Yu V. Balnokin. *Na⁺-ATPases of halotolerant microalgae.* Russian journal of plant physiology 60.4 (2013): 472-482.

Raxón Matias, Y. B. (2018). *Evaluación del rendimiento y caracterización fisicoquímica a escala laboratorio de ácidos grasos de interés alimentario de la microalga (Chlorella vulgaris) procedente de Amatitlán* (Doctoral dissertation, Universidad de San Carlos de Guatemala).

Roumpeka, D. D., Wallace, R. J., Escalettes, F., Fotheringham, I., & Watson, M. (2017). *A review of bioinformatics tools for bio-prospecting from metagenomic sequence data.* Frontiers in genetics, 8, 23.

Ruiz Morales, R. A. (2013). *Aislamiento, purificación y desarrollo de biomasa de Odontella aurtica; Cocsinodiscus granii; Nitzschia sp; Bacillaria sp; Scenedesmus vacuolatus. Microalgas provenientes del mar Atlántico de Guatemala.*

Rumin, J., Nicolau, E., Junior, R. G. D. O., Fuentes-Grünwald, C., Flynn, K. J., & Picot, L. (2020). *A bibliometric analysis of microalgae research in the world, Europe, and the European Atlantic area.* Marine drugs, 18(2), 79.

- Sahoo, S., Mahapatra, S. R., Parida, B. K., Narang, P. K., Rath, S., Misra, N., & Suar, M. (2020). *dEMBF v2. 0: An Updated Database of Enzymes for Microalgal Biofuel Feedstock*. *Plant and Cell Physiology*, 61(5), 1019-1024.
- Saier, M. H., Eng, B. H., Fard, S., Garg, J., Haggerty, D. A., Hutchinson, W. J. & Young, G. B. (1999). *Phylogenetic characterization of novel transport protein families revealed by genome analyses*. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, 1422(1), 1–56. doi:10.1016/s0304-4157(98)00023-9
- Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K. & Holmes, J. B. (2019). *Database resources of the national center for biotechnology information*. *Nucleic acids research*, 47(Database issue), D23.
- Sharma, P., Patil, P., Rao, N., Swamy, K. V., Khetmalas, M. B., & Tandon, G. D. (2013). *Algal Database—Bioprospecting indigenous algae for industrial application*.
- Thornton, B., & Basu, C. (2011). *Real-time PCR (qPCR) primer design using free online software*. *Biochemistry and Molecular Biology Education*, 39(2), 145-154.
- Tu, Z., Liu, L., Lin, W., Xie, Z., & Luo, J. (2018). *Potential of using sodium bicarbonate as external carbon source to cultivate microalga in non-sterile condition*. *Bioresource technology*, 266, 109-115.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. (2007). *Primer3Plus, an enhanced web interface to Primer3*. *Nucleic acids research*, 35(suppl_2), W71-W74.
- Vadlamani, A., Viamajala, S., Pendyala, B., & Varanasi, S. (2017). *Cultivation of microalgae at extreme alkaline pH conditions: A novel approach for biofuel production*. *ACS Sustainable Chemistry & Engineering*, 5(8), 7284-7294.
- Varshney, P., Mikulic, P., Vonshak, A., Beardall, J., & Wangikar, P. P. (2015). *Extremophilic micro-algae and their potential contribution in biotechnology*. *Bioresource technology*, 184, 363-372.

- Wan, M., Rosenberg, J. N., Faruq, J., Betenbaugh, M. J., & Xia, J. (2011). *An improved colony PCR procedure for genetic screening of Chlorella and related microalgae*. *Biotechnology letters*, 33(8), 1615-1619.
- Wang, D., Ning, K., Li, J., Hu, J., Han, D., Wang, H. & Chang, X. (2014). *Nannochloropsis genomes reveal evolution of microalgal oleaginous traits*. *PLoS Genet*, 10(1), e1004094.
- Wang, H. C., Li, K., Susko, E., & Roger, A. J. (2008). *A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny*. *BMC evolutionary biology*, 8(1), 331.
- Wang, H., Zhang, W., Chen, L., Wang, J., & Liu, T. (2013). *The contamination and control of biological pollutants in mass cultivation of microalgae*. *Bioresource technology*, 128, 745-750.
- Wang, X., Spandidos, A., Wang, H., & Seed, B. (2012). *PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update*. *Nucleic acids research*, 40(D1), D1144-D1149.
- Wei, Y., Liu, J., Ma, Y., & Krulwich, T. A. (2007). *Three putative cation/proton antiporters from the soda lake alkaliphile Alkalimonas amylolytica N10 complement an alkali-sensitive Escherichia coli mutant*. *Microbiology (Reading, England)*, 153(Pt 7), 2168.
- Whelan, S., & Goldman, N. (2001). *A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach*. *Molecular biology and evolution*, 18(5), 691-699.
- Yang, H. L., Lu, C. K., Chen, S. F., Chen, Y. M., & Chen, Y. M. (2009). *Isolation and characterization of Taiwanese heterotrophic microalgae: screening of strains for docosahexaenoic acid (DHA) production*. *Marine biotechnology*, 12(2), 173-185.
- Zeisel, A., Yitzhaky, A., Bossel Ben-Moshe, N., & Domany, E. (2013). *An accessible database for mouse and human whole transcriptome qPCR primers*. *Bioinformatics*, 29(10), 1355-1356.

Zhang, D., Wen, S., Wu, X., & Cong, W. (2018). *Effect of culture condition on the growth, biochemical composition and EPA production of alkaliphilic Nitzschia plea isolated in the Southeast of China*. *Bioprocess and biosystems engineering*, 41(6), 831-839.

Zorz, J. K., Sharp, C., Kleiner, M., Gordon, P. M., Pon, R. T., Dong, X., & Strous, M. (2019). *A shared core microbiome in soda lakes separated by large distances*. *Nature communications*, 10(1), 1-10.

Zou, S., Fei, C., Wang, C., Gao, Z., Bao, Y., He, M., & Wang, C. (2016). *How DNA barcoding can be more effective in microalgae identification: a case of cryptic diversity revelation in Scenedesmus (Chlorophyceae)*. *Scientific reports*, 6, 36822.

XII. Anexos

Figura 8: Segmento del cuadro de información de especies relevantes con las indicaciones de la sección 1.1 de la metodología

Especies/género/cepa	Ambiente de origen de la especie	Temperatura de crecimiento	Rango de pH y variabilidad del mismo en el tiempo	Características del medio	Referencia bibliográfica
1 <i>Chlorella sorokiniana</i> str. SLA-0	Soap Lake (WA, USA)	No reportado	10	Altamente alcalino y rico en bicarbonato	Vadlamani, A., Penyalá, B., Viamajala, S., & Varanasi, S. (2018). High productivity cultivator
2 <i>Chlorella vulgaris</i>	Cuerpo de agua salino	No reportado	9.6 - 10	Altamente alcalino y rico en bicarbonato	Chowdhury, R., Keen, P. L., & Tao, W. (2019). Fatty acid profile and energy efficiency of biodi
3 <i>Fistulifera</i> sp.	Lake Okeechobee, Florida	25°C	9-dic	Altamente alcalino y rico en bicarbonato	Berthold, D. E., Rosa, N. D. L., Engene, N., Jayachandran, K., Gantar, M., Laughinghouse, H. D., de los Cobos-Vasconcelos, D., García-Cruz, E. L., Franco-Morgado, M., & González-Sánchez, A.
4 <i>Halospirulina</i> sp.	Lago Texcoco, México	24°C	9.5	Altamente alcalino y rico en bicarbonato	de los Cobos-Vasconcelos, D., García-Cruz, E. L., Franco-Morgado, M., & González-Sánchez, A.
5 <i>Picochlorum</i> sp.	Lago Texcoco, México	24°C	9.5	Altamente alcalino y rico en bicarbonato	de los Cobos-Vasconcelos, D., García-Cruz, E. L., Franco-Morgado, M., & González-Sánchez, A.
6 <i>Microcoleus chthonoplastes</i>	Khilganta soda lake, Chita oblast, Russia	20°C	9.5 - 9.8	Altamente alcalino y rico en bicarbonato	Kupriyanova, E. V., Markelova, A. G., Lebedeva, N. V., Gerasimenko, L. M., Zavarzin, G. A., & P
7 <i>Nannochloris</i> sp.	Songnen Plain of Northeast China	No reportado	9.0 - 10.5	Tierra alcalina y salina	Liu, C., Liu, J., Hu, S., Wang, X., Wang, X., & Guan, Q. (2019). Isolation and identification of a h
8 <i>Nitzschia plea</i>	Agua superficial de Shenzhen, China	15 - 40 °C	>10	Altamente alcalino y rico en bicarbonato	Zhang, D., Wen, S., Wu, X., & Cong, W. (2018). Effect of culture condition on the growth, bioc
9 <i>Synechocystis aquatilis</i>	lake Lonar, MS, India	No reportado	9.5<	Lagunas alcalina y rica en sodio	Bilejezlerim, S. A. A. E. V., & Uyarlanmas, D. (2010). Application of Plackett-Burman desig
10 <i>Thalassiosira pseudonana</i>	No reportado	23°C	8.5 - 10	No reportado	Piiparinen, J., Barth, D., Eriksen, N. T., Teir, S., Spilling, K., & Wiebe, M. G. (2018). Microalgal t
11 <i>Phaeodactylum tricornutum</i>	No reportado	23°C	8.5 - 10	No reportado	Piiparinen, J., Barth, D., Eriksen, N. T., Teir, S., Spilling, K., & Wiebe, M. G. (2018). Microalgal t
12 <i>Chlamydomonas</i> sp.	No reportado	23°C	8.5 - 10	No reportado	Piiparinen, J., Barth, D., Eriksen, N. T., Teir, S., Spilling, K., & Wiebe, M. G. (2018). Microalgal t
13 <i>Microcoleus chthonoplastes</i>	Khilganta soda lake, Russia	21°C	9.8	Conformador de estromatolitos	Kupriyanova, E., Villarejo, A., Markelova, A., Gerasimenko, L., Zavarzin, G., Samuelsen, G., ...
14 <i>Picochlorum</i> sp.	Texcoco Soda Lake	20 - 25°C	>9	Lagunas alcalina y con alta concentración de sodi	Granada-Moreno, C. L., Aburto-Medina, A., de los Cobos Vasconcelos, D., & González-Sánche
15 <i>Scenedesmus</i> sp.	Texcoco Soda Lake	20 - 25°C	>9	Lagunas alcalina y con alta concentración de sodi	Granada-Moreno, C. L., Aburto-Medina, A., de los Cobos Vasconcelos, D., & González-Sánche
16 <i>Chlorella sorokiniana</i>	Cuerpos de agua dulce húngaros	24°C	>7.5	Aguas residuales y eutrofizadas	Figler, A., Dobronoki, D., Márton, K., Nagy, S. A., & Bácsi, I. (2019). Salt tolerance and desalini
17 <i>Chlorella vulgaris</i>	Cuerpos de agua dulce húngaros	24°C	>7.5	Aguas residuales y eutrofizadas	Figler, A., Dobronoki, D., Márton, K., Nagy, S. A., & Bácsi, I. (2019). Salt tolerance and desalini
18 <i>Chlorococcum</i> sp.	Cuerpos de agua dulce húngaros	24°C	>7.5	Aguas residuales y eutrofizadas	Figler, A., Dobronoki, D., Márton, K., Nagy, S. A., & Bácsi, I. (2019). Salt tolerance and desalini

Figura 9: Segmento del cuadro de información de especies aisladas en Guatemala generado con las indicaciones de la sección 1.1 de la metodología

Especie	Tipo de trabajo	Título	Año	Comentarios adicionales	Referencia
<i>Nitzschia</i> sp.	Guía de Microalgas del estanque (de la UVG)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Navicula</i> sp.	Guía de Microalgas del estanque (de la UVG)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Navicula</i> sp.	Guía de Microalgas del estanque (de la UVG)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Scenedesmus</i> sp.	Guía de Microalgas del estanque (de la UVG)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Scenedesmus quadricauda</i>	Guía de Microalgas del estanque (de la UVG)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Selenastrum</i> sp.	Guía de Microalgas del estanque (de la UVG)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Cosmarium</i> sp.	Guía de Microalgas del estanque (de la UVG)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Synedra</i> sp.	Guía de Microalgas del estanque (de la UVG)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Synechococcus</i> sp.	Guía de Microalgas del estanque (de la UVG?)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Roicosphenia curvata</i>	Guía de Microalgas del estanque (de la UVG?)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Closterium</i> sp.	Guía de Microalgas del estanque (de la UVG?)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar
<i>Roicosphenia curvata</i>	Guía de Microalgas del estanque (de la UVG?)	Guía de Algas del Agua del Estanque	S.F.	Guía incompleta	Sin publicar

Figura 10: Segmento del cuadro sobre las características de las secuencias descargadas generado con las indicaciones de la sección 1.2 de la metodología

Tipo de secuencia (nucleotídica o peptídica)	Función descrita en la base de datos	Entrada en UNIPROT	Nombre en base de datos	Nombre del gen	Largo de secuencia	Grado de curado de la secuencia	Procarionta o Eucariota	Con presencia confirmada previamente en trabajos de identificación local	Clasificación taxonómica, principalmente por familia y género	Varietades (aplicable únicamente a especies con alta diversidad y que estén secuenciadas)
Proteica	Na ⁺ /H ⁺ Exchanger domain-containing protein	A0A433USR2	A0A433USR2_ANAVA	DSM107003_22760	517	Predicho o Homologado	Procarionta	No	<i>Trichormus variabilis</i>	SAG 1403-4b
Proteica	Na ⁺ /H ⁺ antiporter	A0A124KTE2	A0A124KTE2_ANAVA	NIES23_51040	470	Predicho o Homologado	Procarionta	No	<i>Trichormus variabilis</i>	NIES-23
Proteica	Na ⁺ /H ⁺ antiporter	A0A124KFT1	A0A124KFT1_ANAVA	NIES23_05150	517	Predicho o Homologado	Procarionta	No	<i>Trichormus variabilis</i>	NIES-23
Proteica	Na ⁺ /H ⁺ antiporter	A0A433US56	A0A433US56_ANAVA	DSM107003_23180	523	Predicho o Homologado	Procarionta	No	<i>Trichormus variabilis</i>	SAG 1403-4b
Proteica	Na ⁺ /H ⁺ antiporter	A0A0K2M4R9	A0A0K2M4R9_9NOST	AA650_22650	628	Predicho o Homologado	Procarionta	No	<i>Anabaena</i> sp.	WA102
Proteica	Na ⁺ /H ⁺ antiporter	A0A124KRV9	A0A124KRV9_ANAVA	NIES23_44530	712	Predicho o Homologado	Procarionta	No	<i>Trichormus variabilis</i>	NIES-23
Proteica	Na ⁺ /H ⁺ -exchanging protein	A0A124KMQ4	A0A124KMQ4_ANAVA	NIES23_30330	520	Predicho o Homologado	Procarionta	No	<i>Trichormus variabilis</i>	NIES-23

Figura 11: Segmento de un cuadro de secuencias alineadas para el género *Arthrospira* generado con las indicaciones de la sección 2.1

1	>tr K1WGT6 K1WGT6_ARTPT_Sodium/hydrogen_exchanger_OS_Arthrospira_platensis_C1_OX_459495_
255	-----MLMVITIISGIGAQVLADYLKVPVPAIVFLLIFGILVGNLILGLVHPNLLGDGLEVIVSLSVALILFEGGSLLELRALGAVS
2	>tr A0A4Y7UM2 A0A4Y7UM2_9CYAN_Sodium_hydrogen_exchanger_OS_Arthrospira_sp_SRM16_OX_1
642	MDTSFEITMLMVITIISGIGAQVLADYLKVPVPAIVFLLIFGILVGNLILGLVHPNLLGDGLEVIVSLSVALILFEGGSLLELRALGA
3	>tr D5A0G4 D5A0G4_ARTPN_Putative_Na+/H+_antiporter_OS_Arthrospira_platensis_(strain_NIES-39_
642	MDTSFEITMLMVITIISGIGAQVLADYLKVPVPAIVFLLVFGILVGNLILGLVHPNLLGDGLEVIVSLSVALILFEGGSLLELRALG
4	>tr B5VWQ2 B5VWQ2_LIMMA_Sodium/hydrogen_exchanger_OS_Arthrospira_maxima_CS-328_OX_513
642	MDTSFEITMLMVITIISGIGAQVLADYLKVPVPAIVFLLIFGILVGNLILGLVHPNLLGDGLEVIVSLSVALILFEGGSLLELRALGA
5	>tr A0A328V5X0 A0A328V5X0_9CYAN_Sodium;proton_antiporter_OS_Arthrospira_sp_O9.13F_OX_1982
642	MDTSFEITMLMVITIISGIGAQVLADYLKVPVPAIVFLLIFGILVGNLILGLVHPNLLGDGLEVIVSLSVALILFEGGSLLELRALGA
1	>tr D4ZP58 D4ZP58_ARTPN_Putative_Na+/H+_antiporter_OS_Arthrospira_platensis_(strain_NIES-39_
255	-----MEQFLNLIPESPIVPFTILILVSLTIPPIFERLRPLGLVGLLFAGVVLGPNMGVQLSPDSEIMKLLSDIGI
2	>tr K1WQ56 K1WQ56_ARTPT_Sodium/hydrogen_exchanger_OS_Arthrospira_platensis_C1_OX_459495_
255	-----MEQFLNLIPESPIVPFTILILVSLTIPPIFERLRPLGLVGLLFAGVVLGPNMGVQLSPDSEIMKFLSDIGI
3	>tr B5W087 B5W087_LIMMA_Sodium/hydrogen_exchanger_OS_Arthrospira_maxima_CS-328_OX_51304
255	-----MEQFLNLIPESPIVPFTILILVSLTIPPIFERLRPLGLVGLLFAGVVLGPNMGVQLSPDSEIMKFLSDIGI
4	>tr A0A5Q4H7N5 A0A5Q4H7N5_9CYAN_Sodium;proton_antiporter_OS_Arthrospira_sp_PLM2.Bin9_OX
255	-----MEQFLNLIPESPIVPFTILILVSLTIPPIFERLRPLGLVGLLFAGVVLGPNMGVQLSPDSEIMKLLSDIGI
5	>tr A0A328V4E6 A0A328V4E6_9CYAN_Sodium;proton_antiporter_OS_Arthrospira_sp_O9.13F_OX_1982
255	-----MEQFLNLIPESPIVPFTILILVSLTIPPIFERLRPLGLVGLLFAGVVLGPNMGVQLSPDSEIMKLLSDIGI
6	>tr D4ZNN3 D4ZNN3_ARTPN_Na+/H+_antiporter_OS_Arthrospira_platensis_(strain_NIES-39_/_IAM_M-
722	MGNLGYLINTDAGSLSMITLTTPLAQILSQIQDPVAVFLIIMGMMLIAPLVFERIKLPGIVGLILAGVIIGPHGIGLLARDNT

Figura 12: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_8*.

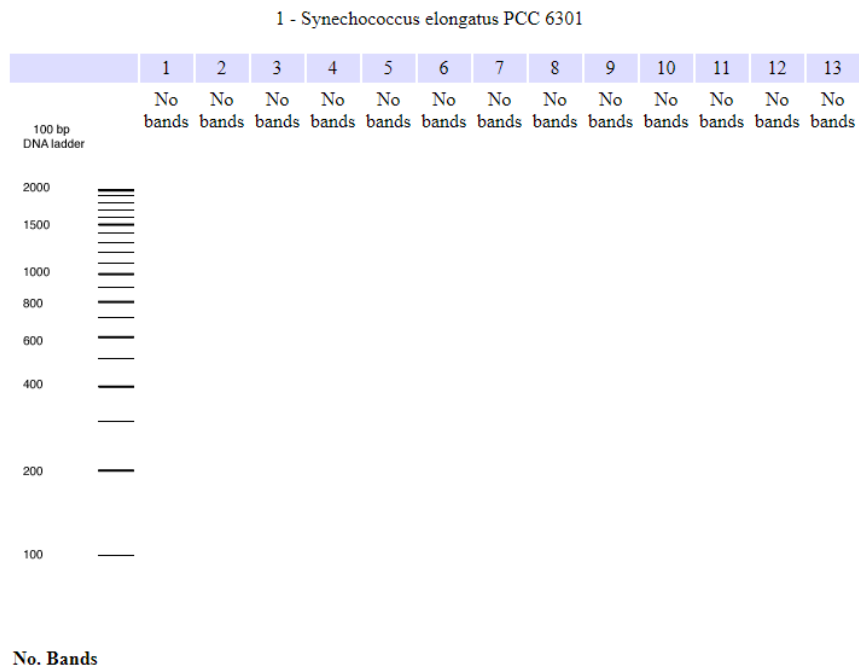


Figura 13: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_9*.

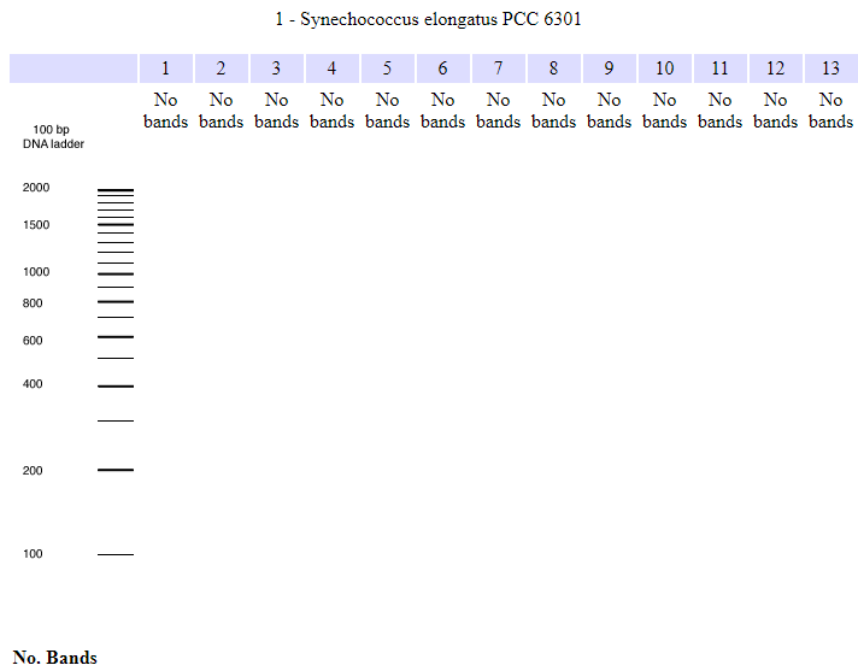


Figura 14: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_11*.

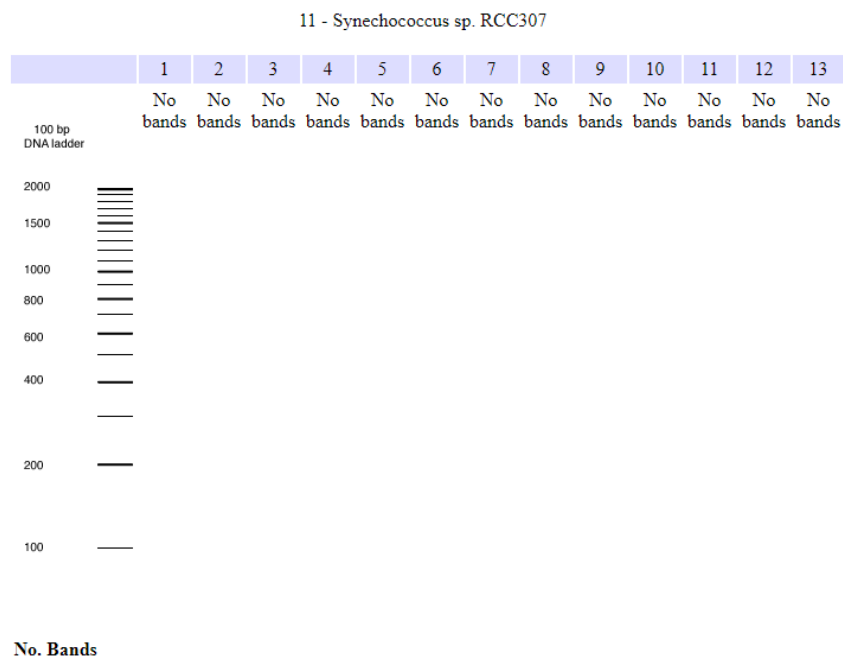


Figura 15: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_13*.

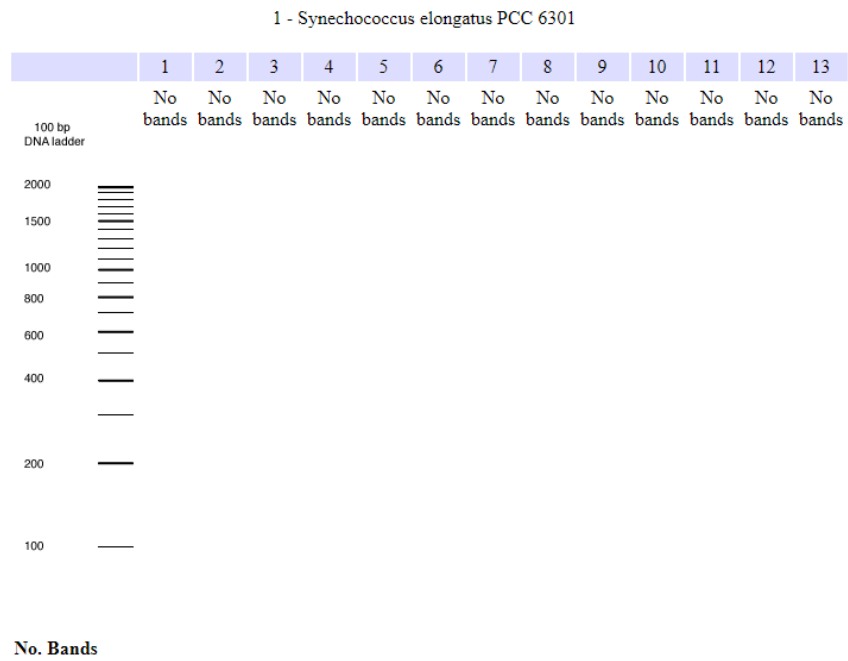


Figura 16: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_15*.

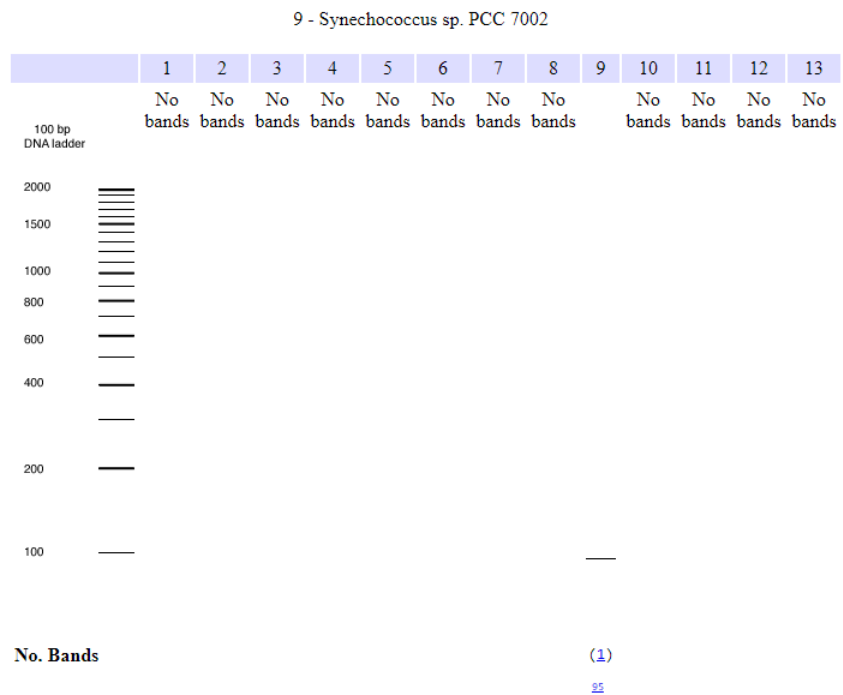


Figura 17: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_16*.

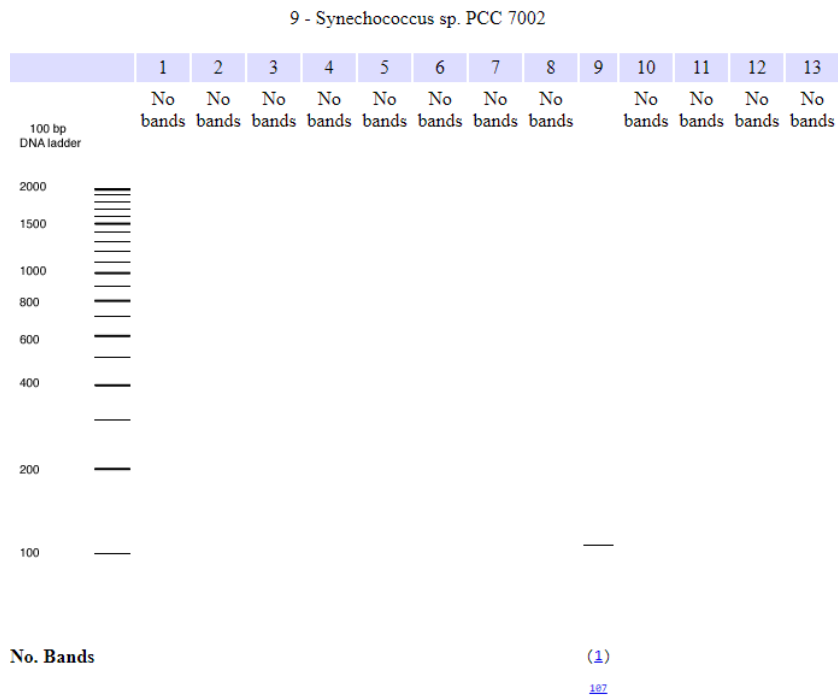


Figura 18 : Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_17*.

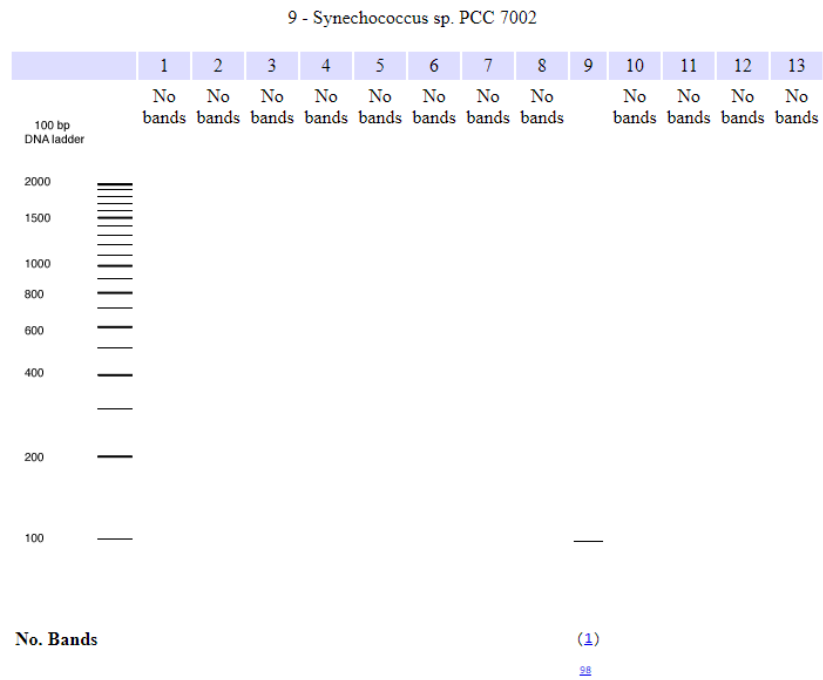


Figura 19: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_20*.

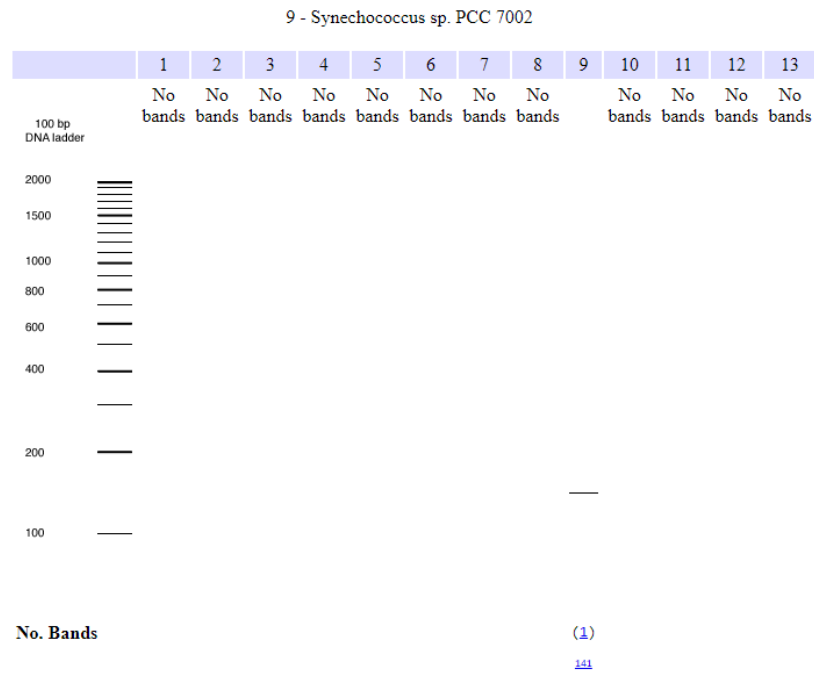


Figura 20: Imagen de la PCR in silico del par de cebadores asociados a la secuencia consenso *Synechococcus_22*.

