CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

Unidad Irapuato

Análisis evolutivo del organelo fotosintético de *Paulinella chromatophora*

Tesis que presenta:
M. en C. Cecilio Valadez Cano

Para Obtener el Grado de:
Doctor en Ciencias Biológicas

En la Especialidad:
Biotecnología de Plantas

Director de la Tesis:
Dr. Luis José Delaye Arredondo

Irapuato, Guanajuato **Septiembre, 2017**

Este trabajo titulado "**Análisis evolutivo del organelo fotosintético de *Paulinella chromatophora***" se llevó a cabo en el laboratorio de **Genómica Evolutiva** del departamento de **Ingeniería Genética** del Centro de Investigación y Estudios Avanzados de IPN unidad Irapuato bajo la asesoría del Dr. Luis José Delaye Arredondo. Así como en el laboratorio de **Simbiosis Microbiana y Evolución de Organelos** en la Universidad de Heinrich Heine, Düsseldorf, Alemania. A cargo de la Dra. Eva Nowack. Se agradece a CONACYT por las becas "nacionales y mixtas" otorgadas durante la duración de mi investigación.

Agradecimientos

Al Dr. Luis Delaye, por haberme aceptado como miembro de su laboratorio y su asesoría, además de la confianza depositada en mi preparación del proyecto.

A mi comité tutoral por sus valiosas sugerencias y aportaciones al proyecto.

Al Dr. Roberto Olivares Hernández, al Dr. Jorge Morales y Dra. Eva Nowack por permitirme realizar las estancias de investigación en sus laboratorios.

A Claudia Rivera Feregrino, Nelly Vazquez, Dr. Edmundo Lozoya y la Dra. Laila Pamela Partida por su valiosa ayuda y apoyo en el trabajo de laboratorio.

Al Consejo Nacional de Ciencia y Tecnología por las becas otorgadas (nacional y mixta).

Al CINVESTAV por permitirme ser parte de su plantel estudiantil.

A mis padres y hermanos cuyo apoyo incondicional fue esencial durante mi doctorado.

A Eridani Belmán Durán por su cariño, comprensión y apoyo incondicional.

A mis amigos del grupo de Genómica Evolutiva y compañeros del doctorado. Y en general a todo el personal que labora en CINVESTAV.

Contenido

Tabla de Figuras

1 Resumen

Se ha estimado que los organelos fotosintéticos de los organismos eucariontes, es decir, los plástidos, se originaron hace más de 1,500 millones de años (m.a.) a partir de un único evento de endosimbiosis primaria. En este evento de endosimbiosis primaria, un organismo heterótrofo eucarionte ingirió y retuvo una cianobacteria de vida libre. Dicho acontecimiento modificó profundamente la historia evolutiva de nuestro planeta. Debido a la antigüedad de este evento y a la integración genética y metabólica de los plástidos con sus hospederos, las etapas tempranas del proceso de oranelogénesis son difíciles de reconstruir. Una alternativa para comprender cómo se originan los plástidos es estudiar organismos que se encuentran en etapas tempranas de organelogénesis fotosintética. Tal es el caso del cromatóforo de *Paulinella chromatophora*.

P. chromatophora es un protista que tiene una nutrición autótrofa gracias a que contiene dos organelos fotosintéticos. El origen evolutivo de estos organelos fotosintéticos (también llamados cromatóforos) es independiente al de los plástidos del resto de organismos eucariontes. Los cromatóforos se originaron a partir de las alfa-cianobacterias, en tanto que los plástidos se originaron a partir de las beta-cianobacterias. El genoma de los cromatóforos se encuentra en etapas tempranas de organelogenésis. Es decir, que el grado de integración tanto genética y metabólica con su hospedero no es tan extremo como en el caso de los plástidos. En este proyecto se planteó estudiar tanto la antigüedad de los cromatóforos, así como los mecanismos evolutivos que participaron en la integración metabólica y genética del cromatóforo con su huésped.

Aunque se presume que el cromatóforo se originó recientemente (entre 60 y 200 m.a.), el intervalo propuesto se basa en asunciones indirectas sobre la antigüedad de otros sistemas simbióticos. Es por ello que una revisión de la edad de *P. chromatophora* es importante. El *artículo 1*: “How really ancient is *Paulinella chromatophora*?” trata sobre el cálculo de la edad aproximada de *P. chromatophora* (*i.e*. adquisición de la capacidad fotosintética del hospedero) mediante la determinación de su reloj molecular. Se encontró que el cromatóforo evolucionó aproximadamente entre 90 y 140 millones de años.

En el *artículo 2*: “Natural selection drove metabolic specialization of the chromatophore in *Paulinella chromatophora*” se estudió cómo cambió el metabolismo del chromatóforo al evolucionar en un plástido a partir de una cianobacteria de vida libre. Para este estudio, se utilizaron modelos estequiométricos de los metabolismos de una cianobacteria de vida libre y del cromatóforo. Con ello, se lograron predecir metabolitos que son esenciales para el cromatóforo (y posiblemente están siendo proporcionados por el hospedero) y se determinó que el metabolismo del cromatóforo es extremadamente frágil debido a la reducción genética experimentada. Finalmente concluimos que el metabolismo del cromatóforo está

particularmente adaptado para producir carbono reducido que es proporcionado al hospedero. Además, el análisis de evolución reductiva sugiere que la selección natural positiva determinó dichas capacidades metabólicas del cromatóforo.

Finalmente el *artículo 3*: "Massive protein import into the early evolutionary stage photosynthetic organelle of the amoeba *Paulinella chromatophora*" se enfocó en el estudio de la evolución del proteoma del organelo en etapas tempranas de organelogénesis. En dicho artículo se concluye que el hospedero contribuye en gran medida a la evolución del proteoma del cromatóforo mediante la importación de proteínas codificadas en el núcleo celular. Dichas proteínas importadas al interior del cromatóforo son de dos tipos diferentes: péptidos cortos de menos de 90 aminoácidos de largo y las segundas son proteínas con más de 268 aminoácidos. Se determinó que la mayoría de genes codificantes para proteínas que son transportadas al cromatóforo no tienen un origen alfa-cianobacterial y se identificó una extensión N-terminal que permite el transporte de proteínas de gran tamaño al cromatóforo. Dicha extensión permite la importación de proteínas de los plástidos cuando es expresada heterologamente en plantas de tabaco.

2 Abstract

It has been estimated that photosynthetic organelles of eukaryotic organisms originated more than 1,500 million year ago (my) from a single event of primary endosymbiosis. In this event, a heterotrophic organism effulged and retained a free-living cyanobacteria. This event, deeply modified the evolutionary history of our planet. Unfortunately, due to the antiquity of this event and the extreme genetic and metabolic integration of plastids to their hosts, the early steps of organellogenesis remain unknown. An alternative to this problem is to study organisms which are in early stages of photosynthetic organelogenesis. This is the case of the chromatophore of *Paulinella chromatophora*.

P. chromatophora is a protist endowed with autotrophic nutrition thanks to two photosynthetic organelles whose endosymbiotic origin is completely independent to that of plastids in other eukaryotic organisms. Chromatophores evolved from alpha-cyanobacterias while plastids from beta-cyanobacterias. The chromatophore's genome is in early stages of organellogenesis (*i.e*. the metabolic and genetic integration with its host is not as extreme as in plastids). In this project, I considered the use of computational tools to study the early evolution of the chromatophore of *P. chromatophora*.

It has been proposed that the chromatophore of *P. chromatophora* evolved relatively recently (60-200 my). However, this range is based on assumptions taken from other endosymbiotic systems and lacks actual information from organisms phylogenetically related to *P. chromatophora*. In this sense, a revision of the age of *P. chromatophora* is relevant. In *Article 1*: "How really ancient is *Paulinella chromatophora*?" we calculated the approximate age of *P. chromatophora* (acquisition of photosynthetic capacity by the host) by determining its molecular clock. We found that the chromatophore originated between 90 and 140 my.

In article 2: "Natural selection drove metabolic specialization of the chromatophore in *Paulinella chromatophora*" we investigated the evolutionary forces determining the functional specialization during the organelogenesis process (*i.e*. the process of metabolic integration and genetic reduction). We modelled *in silico* the metabolic capabilities of the chromatophore, predicted essential metabolites for the chromatophore (likely provided by the host) and determined its level of metabolic fragility. It was concluded that the metabolism of the chromatophore is specially adapted to produce reduced carbon which is provided to the host. In addition, the analysis of reductive evolution suggested that positive natural selection determined its metabolic capabilities.

Finally, in article 3: "Massive protein import into the early evolutionary stage photosynthetic organelle of the amoeba *Paulinella chromatophora*" we focused on the study of the evolution of the proteome of the photosynthetic organelles in early stages of organelogenesis. In this article,

we concluded that the host contributes greatly to the evolution of the proteome of the chromatophore by the import of proteins coded in the host-nucleus. These proteins imported to the chromatophore compartment are of two different types: Short peptides of less than 90 amino acids in length and larger proteins with more than 298 amino acids. It was determined that most of genes coding for proteins imported to the chromatophore do not have an alpha-cyanobacterial origin. Also, an N-terminal extension was identified which allows the transport of large proteins into the chromatophore. Heterologous expression of this N-terminal extension in tabacco plants allows the transport of proteins to plastids.

3 Introducción

3.1 Simbiosis

El término “simbiosis” hace referencia a “la asociación entre dos o más organismos de diferentes especies que puede durar durante todo el ciclo de vida de uno o todos los participantes” (Dimijian 2000).

Las relaciones simbióticas se encuentran ampliamente distribuidas en las distintas formas de vida existentes en el planeta. Una de las relaciones más íntimas que se pueden establecer entre los simbiontes es la relación endosimbiótica. Se puede describir al fenómeno de endosimbiosis como: “un organismo viviendo dentro de otro” (Wernegreen 2012). El tipo de interacción establecida por el endosimbionte con su hospedero se pueden clasificar en: mutualista (el endosimbionte y el hospedero se benefician); comensalista (uno de los participantes se beneficia sin causar perjuicio al otro simbionte); o parasítica (uno de los simbiontes se beneficia a expensas de perjudicar al otro) (Su et al. 2013; Dimijian 2000).

Los organismos endosimbiontes mutualistas pueden llevar a cabo distintas funciones que son benéficas para su hospedero. Por ejemplo: producción de energía, la fijación de nitrógeno, la producción de nutrientes que complementan la dieta de su hospedero entre otras (Wernegreen 2012). Esto le permite a los simbiontes (endosimbionte y su hospedero) explotar nuevos nichos ambientales (Valdivia et al. 2007). Un ejemplo son los metazoos que viven en las extremas condiciones de las fuentes hidrotermales que dependen de los nutrientes provistos por las bacterias endosimbiontes quimioautotróficas (Lane 2007).

El establecimiento de relaciones endosimbioticas ha tenido un profundo impacto en la evolución de la vida en nuestro planeta, principalmente en el origen de organelos citoplásmaticos de las células eucariontes: los plástidos y las mitocondrias.

3.2 Origen de los plástidos

La adquisición de la función fotosintética por los organismos eucariontes fue uno de los eventos más importantes de nuestro planeta debido a que permitió la colonización de nuevos nichos ecológicos, formando la base de la cadena alimenticia de muchos otros organismos (Reyes-Prieto et al. 2007).

Los organelos fotosintéticos (plástidos) de las células eucariontes se originaron a partir de un único evento de endosimbiosis primaria (eucarionte-procarionte) establecida cuando un protista unicelular retuvo a una cianobacteria de vida libre en su interior (Bhattacharya et al. 2004). La idea inicial sugiere que la cianobacteria fue ingerida a través de la fagocitosis como muchos de protistas no fotosintéticos lo hacen hoy en día. Sin embargo, en lugar de ser

digerida, la cianobacteria fue retenida en condiciones intracelulares (Reyes-Prieto et al. 2007) Figura 1.

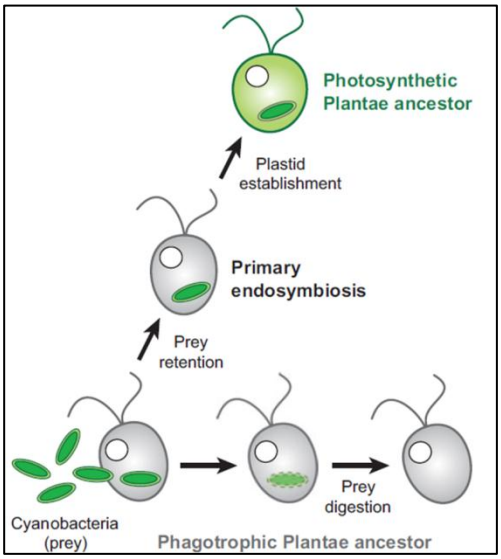

Figura 1. Endosimbiosis primaria que dio origen a los plástidos de los organismos fotosintéticos eucariontes. Tomada de (Reyes-Prieto et al. 2007).

El resultado de ésta endosimbiosis primaria fue el origen de un ancestro eucarionte fotosintético del supergrupo “Plantae” del cual se derivaron tres linajes distintos: Glaucophyta, Rhodophyta y Viridiplantae (Bhattacharya et al. 2007). A través de endosimbiosis secundarias y terciarias, establecidas entre organismos eucariontes fotótrofos y no fotótrofos, los plástidos (cuya principal característica es la presencia de dos o más membranas) se dispersaron a través de diversos organismos eucariontes (Figura 2) (Bhattacharya et al. 2007).

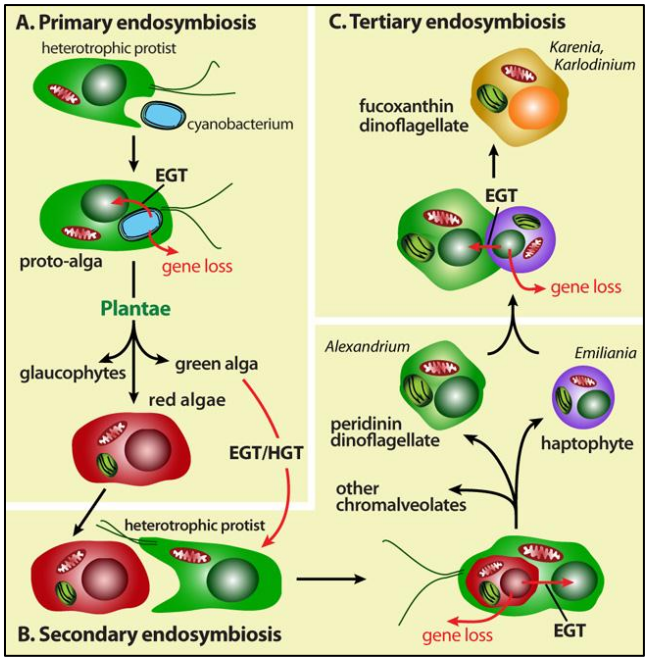

Figura 2. Sucesión de endosimbiosis primaria, secundaria y terciaria que permitieron la dispersión de los plástidos en los organismos eucariontes. Tomada de Chan & Bhattacharya 2010.

Existen múltiples evidencias que sugieren un origen monofilético de los plástidos en el supergrupo "Plantae". Entre ellas, se encuentran los análisis de filogénia molecular de genes tanto nucleares como de plástidos (Reyes-Prieto et al. 2007).

El origen de los organelos fotosintéticos en los organismos eucariontes fue un evento único y particularmente raro que se estima ocurrió hace más de 1,500 m.a. (Yoon et al. 2004). Debido a la antiguedad, el genoma de los plástidos se encuentra altamente reducido (rara vez excede los 200 kbp) y estrechamente integrado tanto genética como metabólicamente con su hospedero. Debido a las extensas transformaciones que ha sufrido el genoma de los plástidos, es difícil reconstruir su historia evolutiva temprana. Una alternativa es el estudio, por analogía, de organismos que se encuentren en las primeras etapas de organelogenesis. Tal es el caso del cromatóforo de *Paulinella chromatophora*.

3.3 El cromatóforo de *Paulinella* sp. como modelo de organelogénesis fotosintética

Las amoebas del género *Paulinella* pertenecen al phylum de los Cercozoa y se caracterizan por ser tecádas. Este género comprende especies tanto autótrofas como heterótrofas. Las especies autótrofas contienen dos cuerpos fotosintéticos alargados llamados cromatóforos (Bhattacharya et al. 2007) (Figura 3). Las especies heterótrofas se alimentan de cianobacterias por fagocitosis (Bhattacharya et al. 2012). En tanto que las especies fotosintéticas utilizan cromatóforos para fijar carbón. Las especies fotosintéticas del género *Paulinella* han sido reportadas en varias partes del mundo: Gran Bretaña, Suiza, Los Países Bajos, Austria, el Mar Báltico, Republica Checa, Ucrania, Estados Unidos, Canadá, Nueva Zelanda, Japón y Corea (Lhee et al. 2017).

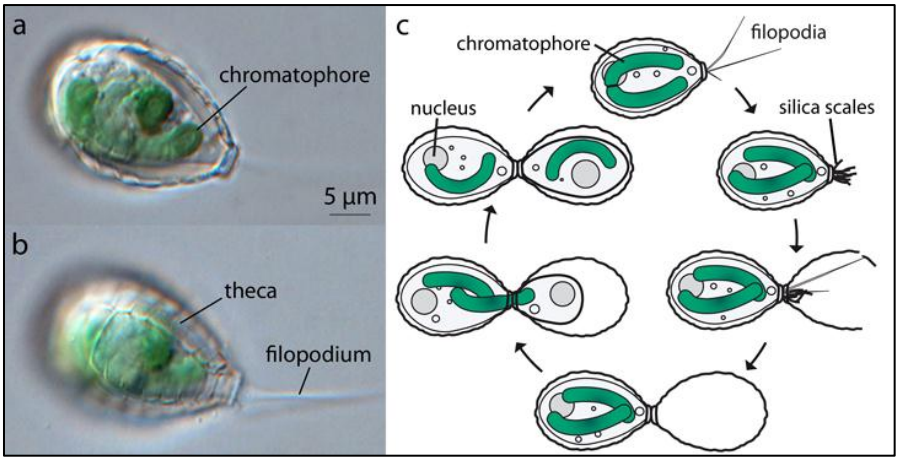

Figura 3. Micrografías de *Paulinella chromatophora* (a y b). Representación esquemática de la división celular de *P chromatophora* (c). Tomada de (Nowack 2014).

La división celular de las especies autótrofas de *Paulinella* comienza con la producción de una nueva teca vacía a partir de escamas de sílice que se producen en el citoplasma y son secretadas con ayuda de los filopodios. Posteriormente, se crea una nueva célula por mitosis. Esta nueva célula es transferida junto con uno de los cromatóforos a la nueva teca a través de una apertura localizada en el extremo superior (Nomura et al. 2014) (Figura 3).

Los análisis filogenéticos han mostrado que el origen de estos organelos fotosintéticos del género *Paulinella* es completamente independiente al de los plástidos del resto de los eucariontes (Marin et al. 2005; Kim & Park 2016). En la Figura 4 se muestra la posición filogenética de los cromatóforos de *P. chromatophora*. Se puede observar que los organelos fotosintéticos de *P*. *chromatophora* pertence al linaje de las alfa-cianobacterias en tanto que los plástidos pertenecen al linaje de las beta-cianobacterias. Las alfa-cianobacterias se caracterizan por tener RubisCo 1A y alfa-carboxisomas, mientras que las beta-cianobacterias contienen RubisCo 1B y beta-carboxisomas. Cabe destacar que los cromatóforos de las distintas especies fotosintéticas de *Paulinella* tienen un origen monofilético. Lo que indica que fue un único evento de endosimbiosis primaria el que dio origen a los organelos fotosintéticos en *Paulinella*; de forma similar a como sucedió con el origen de los plástidos en el resto de organismos eucariontes fotosintéticos (Yoon et al. 2009; Kim & Park 2016).

Hasta el día de hoy, 3 especies diferentes de *Paulinella* fotosintética han sido descritas: *Paulinella chromatophora* (Nowack et al. 2008); *Paulinella micropora* (Lhee et al. 2017; Yoon et al. 2009); y *Paulinella longichromatophora* (Kim & Park 2016). Siendo las dos primeras de agua dulce y la tercera una especie marina. La secuenciación del genoma del cromatóforo de *P. chromatophora* CCAC 0185 (Nowack et al. 2008) y *P. micropora* (Reyes-Prieto et al. 2010; Lhee et al. 2017) muestra que existen diferencias en los tamaños de los genomas (1.02 y ~0.97 Mpb, respectivamente). Además, 66 genes se han perdido diferencialmente entre ambas especies. Los análisis filogenéticos de 18S rADN nuclear y 16S rADN del cromatóforo sugieren que hubo una evolución divergente dentro de la población de *Paulinella* con nutrición autótrofa después del evento único de adquisición del cromatóforo (Lhee et al. 2017; Kim & Park 2016).

El tamaño del genoma de los cromatóforos de ambas especies de *Paulinella* representa aproximadamente 1/3 del tamaño del genoma de la cianobacteria *Synechococcus* sp. WH 5701. La cianobacteria WH 5701 es el organismo de vida libre filogenéticamente más cercano cuyo genoma ha sido secuenciado. Su genoma tiene un tamaño aproximado de 3 Mpb y 3346 genes codificantes de proteínas (Nowack et al. 2008). Como se puede apreciar, el cromatóforo evolucionó por pérdida genética. Esta reducción genética afectó genes que codifican para proteínas que son esenciales en una condición de vida libre, tales como aquellas proteínas que participan en la síntesis de aminoácidos y cofactores. Ello sugiere que la supervivencia y funcionalidad del cromatóforo depende de moléculas que son proporcionadas por el hospedero.

Además de la reducción genética experimentada por el cromatóforo, se ha encontrado evidencia de más 30 genes de origen endosimbiotico que han sido transferidos al núcleo del hospedero. La mayoría de estos genes codifican para proteínas pequeñas involucradas en la fotosíntesis y la foto-aclimatación, siendo el hospedero quien establece el control de la expresión de dichas proteínas (Nowack et al. 2011; Nakayama & Ishida 2009). Se ha encontrado evidencia bioquímica de que 3 proteínas que son codificadas en el núcleo del hospedero (codificadas por los genes *psaE*, *psaK1* y *psaK2*) y que tienen un origen evolutivo en el cromatóforo, forman parte del fotosistema I y son transportadas hasta el organelo fotosintético en donde ejercen su función (Nowack & Grossman 2012).

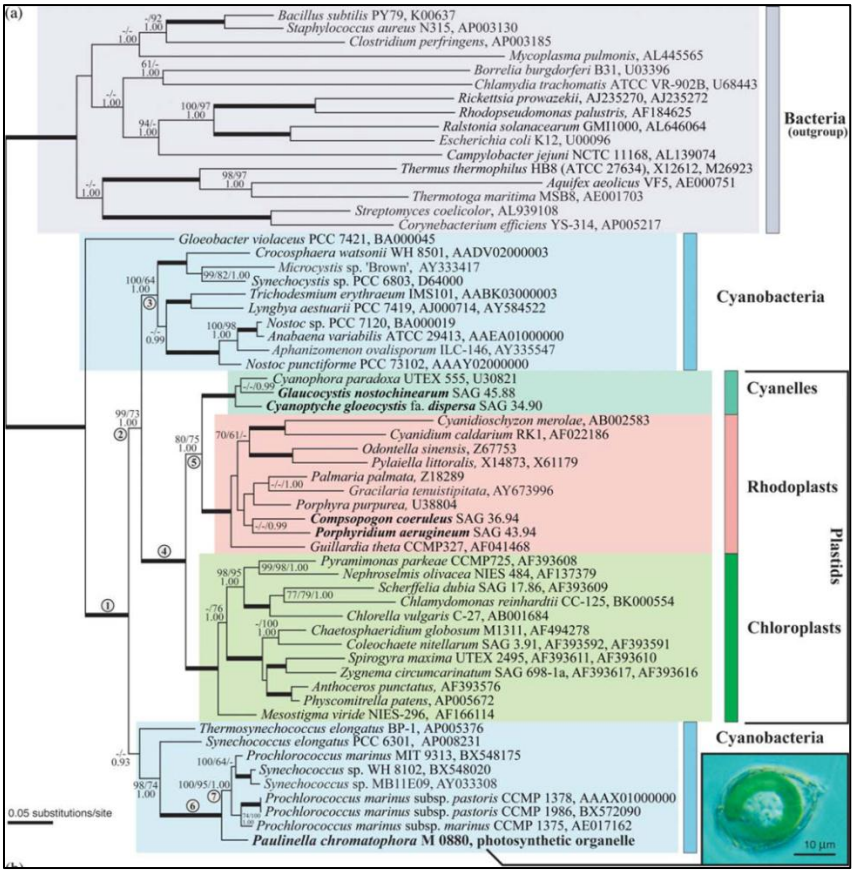

Figura 4. Árbol filogenético de los cromatóforos de *Paulinella chromatophora*, diversos plástidos y cianobacterias de vida libre. Tomada de (Marin et al. 2005).

El análisis de los cromatóforos de *Paulinella chromatophora* ha sido de vital importancia para el estudio de la organelogénesis fotosintética por su analogía con los plástidos del resto de organismos eucariontes fotosintéticos. De esta manera, la investigación realizada durante mi proyecto de doctorado contribuyó al conocimiento de la biología de *Paulinella chromatophora* y su cromatóforo, haciendo énfasis en la interacción genética y metabólica de los simbiontes y su impacto durante el proceso de organelogénesis.

4 Resumen y discusión de artículos publicados

4.1 Determinando la edad de *Paulinella chromatophora*

Como se ha mencionado previamente, se ha estimado que la endosimbiosis primaria a partir de la cual se originaron los plástidos, ocurrió hace más de 1,500 m.a. (Yoon et al. 2004). De aquí surge la importancia evolutiva del uso de *P. chromatophora* como modelo de estudio de la organelogénesis fotosintética pues se ha propuesto que sus cromatóforos tienen un origen mucho más reciente. Ello, permite indagar en los procesos evolutivos iniciales por analogía.

Inicialmente se había sugerido que el cromatóforo tiene una edad mínima de 60 m.a. (Nowack et al. 2008). Esto se basó en la propuesta de que en *Buchnera aphidicola* (una bacteria

endosimbionte de áfidos) un pseudogen necesita entre 40 y 60 m.a. para desintegrarse completamente. Dado que el genoma del cromatóforo aun contiene pseudogenes, se asumió, por analogía, una edad mínima de 60 m.a. para este organelo fotosintético. Por otro lado, (Berney & Pawlowski 2006) establecieron una edad máxima de los cromatóforos de 200 m.a. Esta estimación fue realizada calculando el tiempo de divergencia con un reloj molecular. Sin embargo, la especie contra la cual se comparó *P*. *chromatophora* no era el co-descendiente más cercano. Por lo tanto, la edad de 200 m.a. es una edad máxima, si el cálculo del reloj molecular fue realizado correctamente. Considerando las propuestas realizadas por los trabajos mencionados previamente, un intervalo de 60-200 m.a. ha sido sugerido recientemente (Nowack 2014). Sin embargo, no se había realizado ningún estudio directo para conocer el tiempo de divergencia de las *Paulinellas* heterótrofas de las fotosintéticas utilizando un reloj molecular. Por tal motivo, una revisión del tiempo de divergencia de *Paulinella chromatophora* fue llevada a cabo.

4.1.1 Reloj molecular

Las mutaciones son la fuente de variabilidad genética existente entre los individuos y especies y son originadas por errores en la replicación del ADN o daños que no han sido reparados apropiadamente (Moorjani et al. 2016). Una proporción relativamente grande de las mutaciones son neutrales (*i.e*. no tienen un efecto en la adecuación del individuo). De ésta manera, éstas nuevas mutaciones pueden dispersarse y fijarse en la población por un proceso aleatorio llamado deriva genética (Kimura 1984). Así entonces, la hipótesis del reloj molecular sostiene que las secuencias de ADN y de proteínas evolucionan a una tasa que es aproximadamente constante a lo largo del tiempo y proporcional a la tasa de mutación neutral. Por lo tanto, se asume que las diferencias genéticas entre cualesquiera dos especies son proporcionales al tiempo de divergencia entre estas (Ho 2008).

Para poder utilizar la distancia genética para determinar tiempos de divergencia, es necesario calibrar el reloj molecular. Para calibrar el reloj molecular se requiere evidencia del tiempo de divergencia entre especies derivada del registro fósil. Los fósiles nos permiten conocer el tiempo mínimo de existencia de un linaje determinado (Bromham & Penny 2003). Como se puede intuir, la exactitud de la estimación de tiempo de divergencia depende críticamente de los puntos de calibración utilizados. La estimación del tiempo de divergencia usando la información molecular junto con la información derivada del registro fósil, ha sido importante para determinar el tiempo de divergencia de distintos clados de organismos eucariontes. Por ejemplo, de acuerdo al trabajo publicado por (Douzery et al. 2004) los reinos eucariontes se diversificaron hace 950-1,259 m.a.

4.1.2 Contribución del *artículo 1* a la estimación de la edad de *Paulinella chromatophora*

La investigación presentada en el *artículo 1* "How really ancient is *Paulinella chromatophora*?" hace referencia a la estimación del tiempo de divergencia de *Paulinella chromatophora* del resto de las especies heterótrofas del genero de *Paulinella*. El trabajo se llevó a cabo usando secuencias del gen 18S ARNr de organismos pertenecientes al clado

Rhizaria y Stramenopila.

Para calibrar el reloj molecular en el *artículo 1* se utilizaron varias fuentes de información en los cuales se incluyen fósiles tanto químicos como morfológicos, así como de estimaciones previas de relojes moleculares. Las evidencias de calibración utilizadas fueron:

a) El origen de las diatomeas del género *Rhizosolenia* el cual es conocido con un alto índice de confiabilidad (91.5 ± 1.5 m.a.) (Damsté et al. 2004).

b) Un tiempo mínimo de divergencia de las diatomeas pennadas (80 m.a.) (Parfrey et al. 2011; Kooistra et al. 2007)

c) Un tiempo mínimo de divergencia para las diatomeas (133.9 m.a.)(Parfrey et al. 2011).

d) Un tiempo mínimo de divergencia de Euglyphidae de 40 m.a. (Barber et al. 2013).

e.1) Reloj molecular estimado para la divergencia de rhizaria de stramenopila (1232.5 m.a.) (Parfrey et al. 2011).

e.2) Reloj molecular estimado para la divergencia de rhizaria de stramenopila (754.1 m.a.) (Parfrey et al. 2011).

f) Evidencia disponible de los microfósiles en forma de vaso (VSM, por sus siglas en inglés *vase-shaped microfossils*) (742 m.a.) (Porter & Knoll 2000).

Las distintas evidencias se clasificaron en 4 esquemas diferentes de calibración: A (a, b, c, d, e.1); B (a, b, c, d, e.1, f); C (a, b, c, d, e.2); y D (a, b, c, d).

En la figura 5, se muestra el tiempo mínimo de divergencia entre *Paulinella chromatophora* de las *Paulinellas* heterótrofas (93.6 m.a.) bajo el esquema de calibración A (a, b, c, d, e.1).

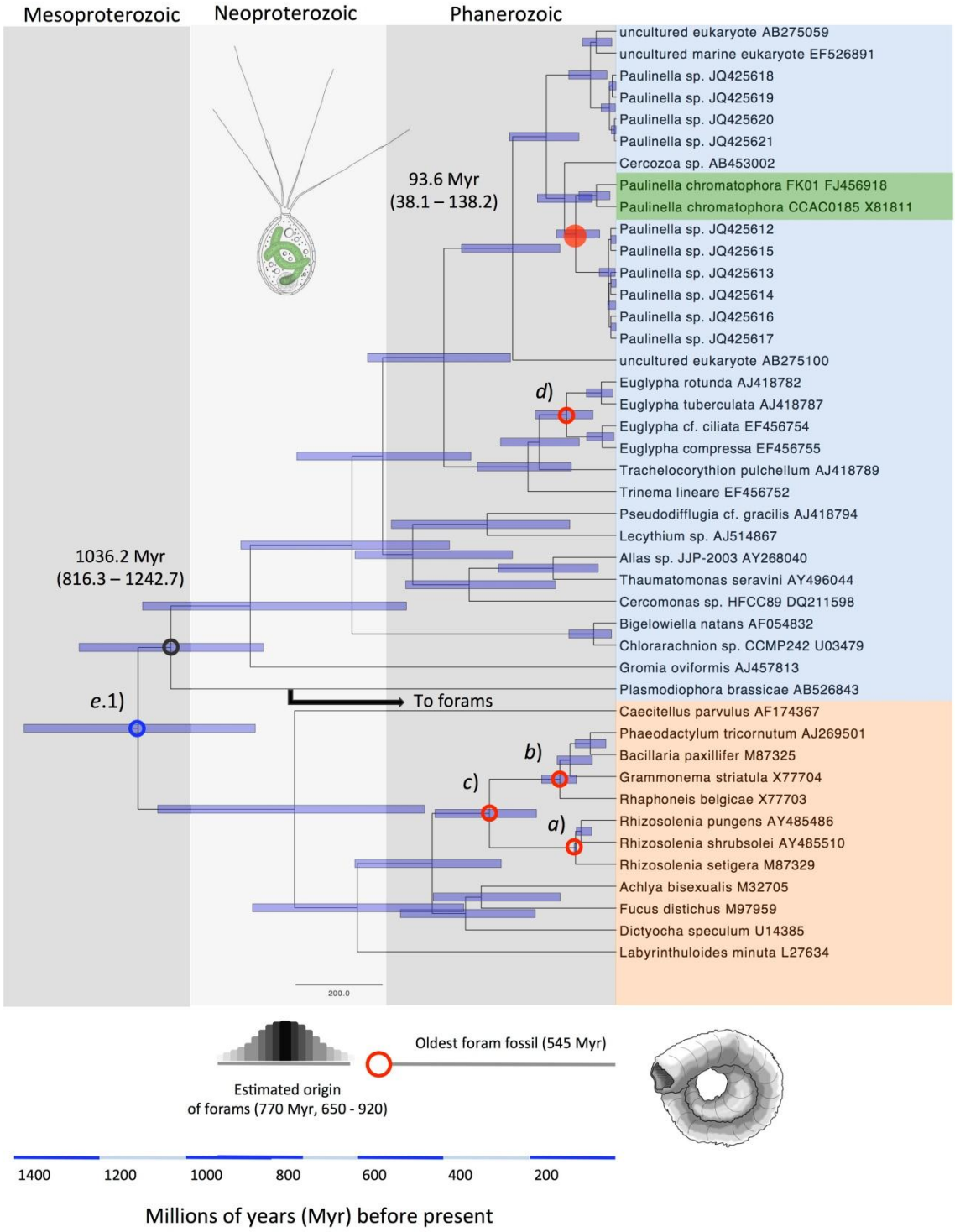

Figura 5. El origen de *Paulinella chromatophora* en el tiempo de acuerdo a su reloj molecular log-normal de su SSU ARNr. El árbol está calibrado bajo el esquema A (a, b, c, d, e.1). Los círculos rojos abiertos pequeños indican calibraciones basados en el registro fósil (a, b, c, y d); el círculo azul abierto pequeño indica la calibración basada en estimaciones previas (e.1); el círculo rojo completo representa la edad estimada para *P. chromatophora*.

Como ya se describió previamente, los cálculos de los tiempos de divergencia dependen críticamente de los puntos de calibración utilizados. Como es de esperarse, los distintos enfoques de calibración utilizados sugieren distintas edades para *P. chromatophora*: A) 93; B) 141.4; C) 55.8; y D) 47.9 m.a. La discusión e interpretación de los cálculos de relojes moleculares debe ser congruente con el conocimiento firmemente bien establecido del registro fósil. Por lo que la elección del (los) esquema(s) que mejor explican con mayor exactitud el tiempo de divergencia de *P. chromatophora* (A, B, C, D) permanecía por ser develado.

4.1.3 Contribución al *artículo 1*

Mi contribución al *artículo 1* consistió en proponer utilizar el registro fósil de los foraminíferos, el cual se encuentra detalladamente descrito, como marco de referencia para valorar las edades sugeridas para *P. chromatophora* por los distintos esquemas de calibración (A, B, C, D). Los foraminíferos tienen un registro fósil detallado cuya primera aparición en el registro fósil está fechado hace 550 m.a. (Culver 1991). Un análisis filogenético basado en 15 genes concatenados mostró que los foraminíferos son un grupo hermano de *Plasmodiophora brassicae* (Parfrey et al. 2011). Por lo tanto se concluyó que la edad calculada para *P. brassicae* debe ser al menos igual a la edad mínima propuesta para los foraminíferos. Inspeccionando los 4 esquemas de calibración, se pudo concluir que solamente los esquemas A y B estiman una edad aceptable para *P. brassicae* en términos de comparación con el origen de los foraminíferos (Figura 6).

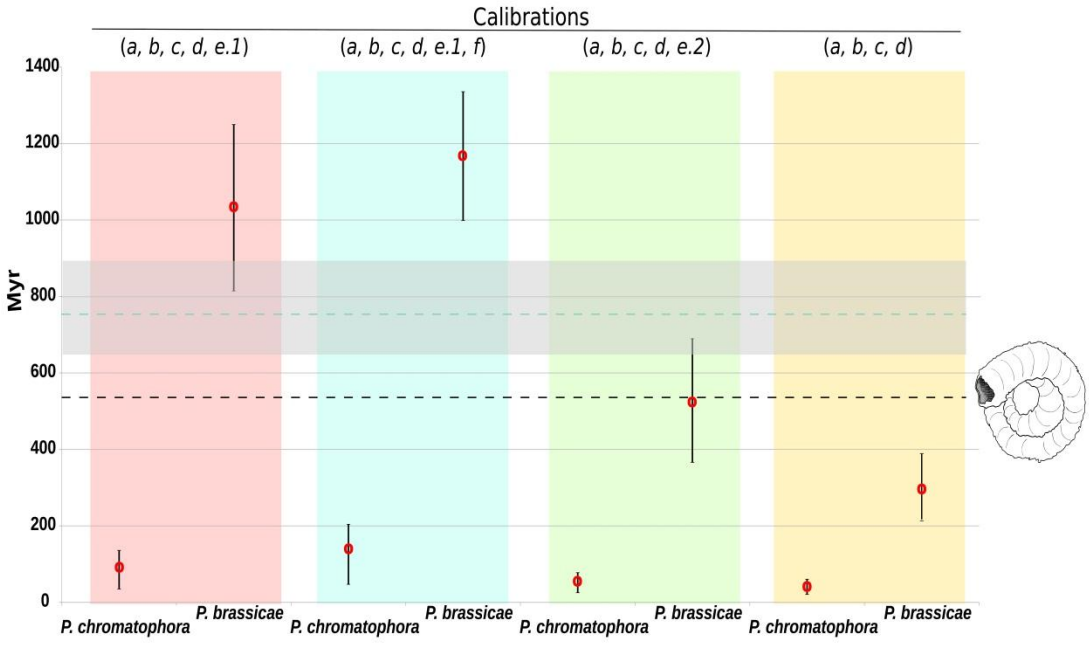

Figura 6. Concordancia entre los tiempos de divergencia calculados por los distintos esquemas de calibración y el registro fósil de los primeros foraminíferos. Solamente los esquemas de calibración que asumen un tiempo de divergencia entre Rhizaria de Stramenopila de ~1232 m.a. (e.1) son consistentes con el registro fósil y el tiempo de divergencia propuesto de los foraminíferos. La linea punteada negra representa la fecha del fósil foraminífero más antiguo (545 m.a.). La línea punteada azul representa el tiempo estimado de origen de los foraminíferos 770 m.a. (650-920).

Lo anterior fue crucial para seleccionar los esquemas de calibración que plantean una edad mínima de 93.6 m.a. y una máxima de 141.4 m.a. para *P. chromatophora*. El intervalo calculado es consistente con las ideas propuesta por (Nowack et al. 2008) y (Douzery et al. 2004) quienes han propuesto que los cromatóforos tienen una edad mínima de 60 m.a. y una máxima de 200 m.a. (Nowack 2014). A diferencia de las propuestas de Nowack 2008 y Douzery 2004, el análisis realizado en el *artículo 1*, se basó directamente en un estudio de reloj molecular en *P. chromatophora* y otros organismos filogenéticamente cercanos. Además, en el *artículo 1*, utilizando las mismas restricciones de calibración, se propone que las dos especies autótrofas de *Paulinella* (*Paulinella chromatophora* y *Paulinella micropora* FK01) divergieron hace 45.7-64.7 m.a.

4.2 Análisis evolutivo del proceso de organelogénesis del cromatóforo de *P. chromatophora*

Uno de los aspectos más importantes del proceso de organelogénesis, son las modificaciones genéticas, metabólicas y funcionales que son experimentadas por el organismo de vida libre en condiciones intracelulares hasta convertirse en un organelo. Para estudiar este proceso, en el *artículo 2* "Natural selection drove metabolic specialization of the chromatophore in *Paulinella chromatophora*" se utilizaron como modelos biológicos el cromatóforo de *P. chromatophora* y la cianobacteria de vida libre *Synechococcus* sp. WH 5701 la cual se consideró como un ancestro "hipotético" del cromatóforo (Figura 7).

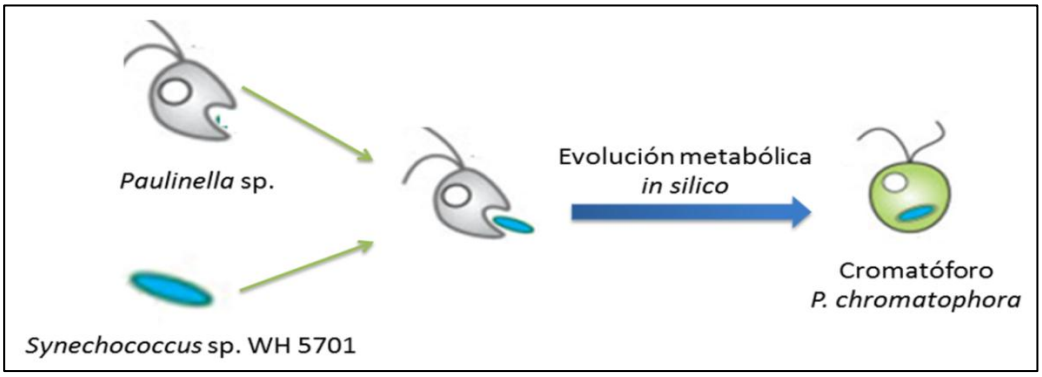

Figura 7. Representación esquemática del análisis evolutivo computacional del cromatóforo de *Paulinella chromatophora*, usando como ancestro hipotético la cianobacteria *Synechococcus* sp. WH 5701

4.2.1 La evolución reductiva del metabolismo del cromatóforo está determinada por la interacción metabólica con su hospedero

Debido a la estructura poblacional de los endosimbiontes en condiciones intracelulares (pequeño tamaño poblacional y experimentación de constantes cuellos de botellas) se ha determinado que la evolución de su genoma y sus capacidades funcionales está principalmente

determinada por la deriva genética y la selección natural purificadora. Debido a la interacción establecida con el hospedero, existe una relajación de la selección natural sobre secuencias codificantes que no son necesarias en las condiciones intracelulares (por ejemplo, síntesis de metabolitos que son ahora proporcionados por el hospedero) (McCutcheon & Moran 2011; Moran & Plague 2004). Como consecuencia, el genoma de los endosimbiontes acumula mutaciones ligeramente deletéreas por el proceso del "trinquete de Muller" (Moran 1996). El resultado es la disrupción de secuencias codificantes que finalmente son eliminadas del genoma (McCutcheon & Moran 2011).

Además de la deriva genética, la selección natural a nivel del holobionte (término referido al hospedero más el endosimbionte) impone una fuerte restricción en la evolución de las capacidades funcionales del endosimbionte por mecanismos como "parter fidelity feedback" (PFF). PFF asegura que modificaciones genéticas experimentadas por el endosimbionte, que afectan la adecuación del holobionte serán eliminadas por la selección natural (Shou 2015). De esta manera se puede explicar la prevalencia de mitocondrias funcionales en las células eucariontes (Shou 2015). Así pues, el genoma de los organismos endosimbiontes tiende a reducirse, mientras que se retienen solamente aquellos genes que son indispensables para su supervivencia y su funcionalidad en el holobionte.

Debido a la reducción genética experimentada y la dependencia nutricional establecida con su hospedero, el cultivo experimental de los endosimbiontes sin su hospedero es muy complicado sino imposible, en la actualidad. Una alternativa para su estudio, es el modelaje mediante el uso de herramientas computacionales. Particularmente para el estudio del metabolismo de organismos endosimbiontes, herramientas computacionales y enfoques estequiométricos han sido utilizados (Thomas et al. 2009; Pál et al. 2006; González-Domenech et al. 2012; Belda et al. 2012). Por ejemplo, el análisis de balance de flujos (Flux Balance Analysis, FBA) que es un método estequiométrico que permite predecir fenotipos celulares en condiciones ambientales definidas y pérdidas genéticas de interés (Orth et al. 2010).

Con la secuenciación de su genoma (Nowack et al. 2008), se pudo determinar que el cromatóforo conserva genes que participan en rutas biosintéticas de metabolitos esenciales para los organismos eucariontes (cofactores y aminoácidos), lo que sugiere que los metabolitos proporcionados por el cromatóforo a su hospedero pueden ser varios. Por otro lado, a pesar de la falta de evidencia experimental, se sugiere que el cromatóforo es incapaz de producir ciertos metabolitos que posiblemente están siendo proporcionados por el hospedero. Lo anterior señala una estrecha comunicación metabólica entre el hospedero y su endosimbionte fotosintético.

En este sentido, una primer parte del *artículo 2* estuvo enfocada en: *i*) analizar las rutas metabólicas que han sido conservadas por la selección natural purificadora en el cromatóforo; *ii*) la construcción y análisis de modelos metabólicos del cromatóforo (de aquí en adelante

referido como *i*CV265) y la cianobacteria de vida libre *Synechococcus* sp. WH 5701 (*i*CV498); y *iii*) la predicción de las capacidades metabólicas del cromatóforo y su interacción con el hospedero.

4.2.1.1 Contribución del *artículo 2* a la predicción de la interacción metabólica existente entre el cromatóforo y su hospedero

A pesar de que no existe evidencia experimental sobre la interacción metabólica establecida entre el cromatóforo y su hospedero, el análisis computacional nos permitió obtener una serie de metabolitos que no pueden ser producidos por el cromatóforo y que según nuestro modelo, son esenciales para la funcionalidad del cromatóforo y que por lo tanto es muy probable que estén siendo proporcionados por el hospedero. La comparación del contenido genético, el análisis funcional y simulación de la evolución reductiva (ver métodos en *artículo 2*) entre el cromatóforo y la cianobacteria de vida libre nos llevó a las siguientes conclusiones:

Existen 13 categorías metabólicas que han sido afectadas en menor proporción por la pérdida genética en el genoma del cromatóforo (rutas que han sido preferencialmente conservadas). La retención de éstas 13 categorías puede atribuirse a la selección natural purificadora a nivel del holobionte por lo que seguramente tienen un papel adaptativo (*i.e*. fotosíntesis y la ruta de la glucolisis/gluconeogénesis) (Figura 8).

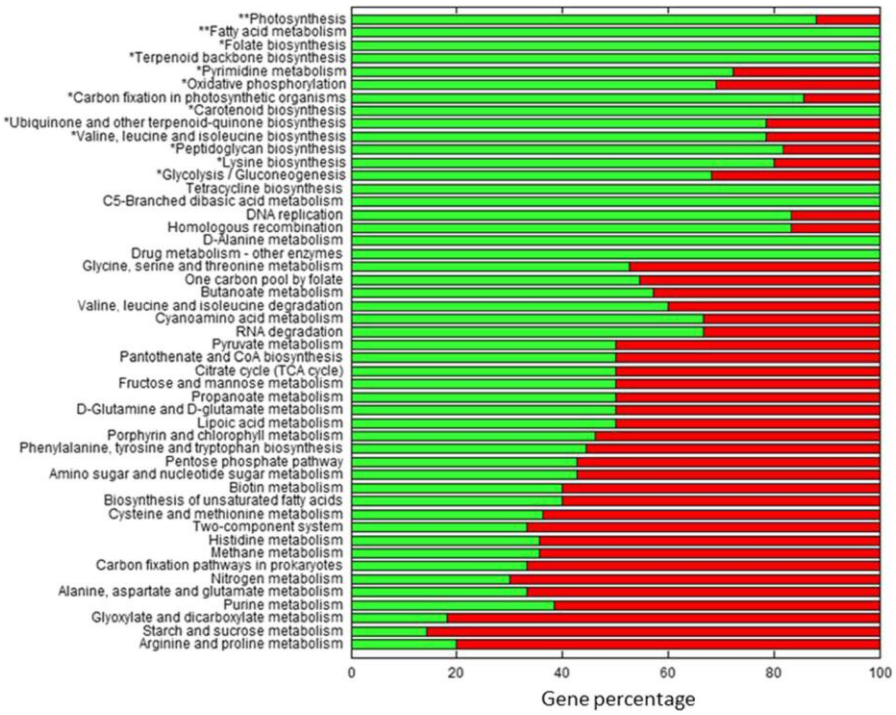

Figura 8. Conservación genética de categorías funcionales en el cromatóforo en comparación con *Synechococcus sp*. WH 5701. Para cada categoría funcional se muestra en verde y rojo el porcentaje de conservación y pérdida genética en el cromatóforo, respectivamente. Categorías funcionales bien conservadas son indicadas con asteriscos (p-valor < 0.05* o < 0.05**, corrección de Bonferroni).

La conservación de rutas metabólicas observada en el cromatóforo es análogo a lo

observado en otros organismos endosimbiontes como *Buchnera aphidicola* (endosimbionte de áfidos) (Canbäck et al. 2004). En *B. aphidicola* las rutas metabólicas que participan en la síntesis de aminoácidos que son proporcionados al hospedero se han conservado por la selección natural operando al nivel del holobionte (Shigenobu et al. 2000). De la misma manera, el análisis de selección realizados a 681 alineamientos de secuencias ortólogos entre *Paulinella chromatophora* y *Paulinella micropora* demostró que la mayoría de ellas tienen señales de selección natural purificadora (Reyes-Prieto et al. 2010). El resultado del análisis de conservación genética en el metabolismo del cromatóforo enfatiza el papel de la selección natural en el mantenimiento de las asociaciones mutualistas entre los endosimbiontes y su hospedero.

El metabolismo del cromatóforo se encuentra reducido. La construcción de los modelos metabólicos *i*CV265 y *i*CV498 demuestran que el metabolismo del cromatóforo se encuentra reducido en comparación con el de la cianobacteria de vida libre, además de que es altamente frágil a las pérdidas genéticas pues más del 83.77% de sus genes metabólicos son esenciales (mientras que en *i*CV498 el 66.86% son esenciales).

Los resultados del análisis de fragilidad del cromatóforo coinciden con observaciones previas que sugieren que el metabolismo de los organismos endosimbiontes es más frágil que el de sus relativos de vida libre (Thomas et al. 2009; González-Domenech et al. 2012; Belda et al. 2012). La diferencia de fragilidad metabólica existente entre el cromatóforo y *Synechococcus sp*. WH 5701 refleja la transición de vivir en vida libre a condiciones intracelulares más estables.

El cromatóforo requiere de metabolitos proporcionados por el hospedero que son esenciales para su funcionalidad. El modelo *i*CV265 predice que 13 metabolitos tienen que estar siendo proporcionados por el hospedero al cromatóforo (Figura 9).

El presente análisis enfatiza el poder predictivo de las herramientas computacionales en el estudio de los organismos endosimbiontes. A pesar de que no existe evidencia experimental sobre los metabolitos que el hospedero proporciona al cromatóforo, el modelo *i*CV265 sugiere que los metabolitos presentandos en la figura 9 son esenciales para la funcionalidad del cromatóforo. Para validar los metabólitos predichos por nuestro modelo, se llevó a cabo un análisis de evolución reductiva (ver métodos en *artículo 2*). Este análisis de evolución reductiva demuestra que los metabolitos predichos explican con aproximadamente el 80% de exactitud el contenido genético del modelo del cromatóforo. Este valor contrasta con el 59% predicho cuando se realiza en análisis de evolución reductiva con metabolitos al azar.

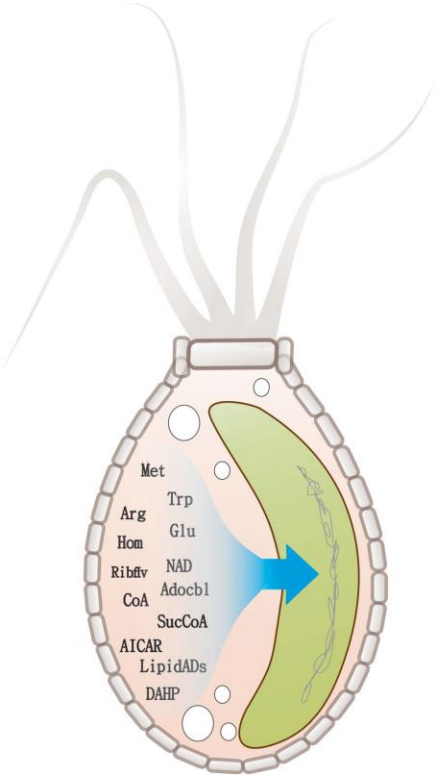

Figura 9. Simulación de metabolitos proporcionados por el hospedero al cromatóforo. El modelo *i*CV265 predice que los metabolitos que no pueden ser producidos por el cromatóforo, pero que son esenciales se incluyen: aminoácidos (Met = Metionina, Trp = Triptófano, Arg = Arginina, Glu = Glutamato, Hom = Homoserina), cofactores (NAD = nicotinamida adenina dinucleótido, Adocbl = Adenosilcobalamina, CoA = Coenzima A), vitaminas (Ribflv = Riboflavina) y otros (AICAR = 1-(5'-fosforibosil)-5-amino-4-imidazolcarboxamide), SucCoA =Succinil-CoA, LipidsADs = Lípido A Disacárido, DAHP = 2-Deshidro-3-desoxi-D-arabino-heptonato 7-fosfato).

El modelo *i*CV265 nos permitió establecer una tasa teórica de carbono reducido que es exportado por el cromatóforo a su hospedero. En el análisis de FBA, la función objetivo normalmente usada es la producción de biomasa (*i.e.* simulación de crecimiento celular). Sin embargo, muchas otras funciones objetivos pueden ser analizadas por este enfoque estequimétrico (Khannapho et al. 2008). Particularmente, el crecimiento de los organismos endosimbiontes se encuentra restringido por el hospedero, lo que hace obvio suponer que la funcionalidad metabólica de los organismos endosimbiontes mutualistas se enfoca en la producción de metabolitos proporcionados al hospedero. Por ejemplo, se ha determinado que *Chlorella* sp. (el endosimbionte fotosintético de *Paramecium bursaria*), proporciona el 57% del carbono fotosintéticamente asimilado a su hospedero (Johnson 2011).

De igual manera, la principal función del cromatóforo es la de proporcionar carbono reducido fotosintéticamente asimilado a su hospedero (Kies & Kremer 1979). Por tal motivo, en el presente trabajo se analizó por primera vez la producción de metabolitos proporcionados al hospedero como función objetivo. Aunque no existe evidencia experimental, se estableció una tasa teórica de la capacidad del endosimbionte de producir carbono reducido que está siendo exportado al hospedero.

4.2.2 La selección natural determinó la especialización funcional metabólica del cromatóforo de *P. chromatophora*

En previas investigaciones, una conclusión a la que se ha llegado, es que el metabolismo de los endosimbiontes mutualistas y los organelos de origen mutualista están especialmente adaptados para cumplir una función que beneficia al holobionte. (Wang et al. 2006) han sugerido que el metabolismo de los plástidos se caracteriza por tener: i) una longitud de trayectoria promedio más larga; ii) un diámetro mayor; iii) se encuentra centrado al Ciclo de Calvin; y iv) presenta una mejor organización modular. Todo esto cuando se compara con el metabolismo de una cianobacteria de vida libre. Dichas características se atribuyeron a la especialización funcional de los plástidos como productores de carbono reducido. Además, el metabolismo de organismos endosimbiontes mutualistas, como *Buchnera* (Russell et al. 2014) y *Blochmannia* (Zientz et al. 2004) están especialmente adaptados para sobre-producir aminoácidos que son exportados a su hospedero. Esta capacidad de sobre-producir metabolitos que son exportados fue consecuencia de reestructuraciones metabólicas causadas por la pérdida genética. Por ejemplo, la pérdida de la ruta biosintética de purina que permite producir histidina en tasas mayores que en organismos de vida libre (Thomas et al. 2009).

Aunque es lógico pensar que la selección natural a nivel del holobionte ha determinado las modificaciones metabólicas experimentadas por los organismos endosimbiontes mutualistas, pocos trabajos se han enfocado en estudiar la evolución de dicho fenómeno.

4.2.2.1 Contribución del *artículo 2* al análisis de especialización funcional del cromatóforo

Una segunda parte importante del *artículo 2* se basó en el análisis de especialización funcional que experimentó el cromatóforo. La comparación, análisis y simulaciones reductivas realizadas a los modelos *i*CV265 e *i*CV498 nos permitió llegar a las siguientes conclusiones:

El metabolismo del cromatóforo está especialmente adaptado para producir carbono reducido. El cálculo de la distribución del flujo de carbono mediante FBA nos permitió determinar que el metabolismo del cromatóforo es capaz de producir una mayor cantidad de carbono reducido que puede estar siendo exportado hacia el hospedero, en comparación con la cianobacteria de vida libre (Figura 10).

La capacidad del cromatóforo de producir una mayor tasa de carbono reducido que la cianobacteria de vida libre está determinada por la interacción metabólica establecida con el hospedero y la pérdida de rutas metabólicas, lo que permitió un redireccionamiento del flujo del carbono inorgánico hacía la producción de carbono reducido.

En análisis de evolución reductiva (ver métodos en *artículo 2*) sugiere que la probabilidad de que la tasa máxima de producción de carbono reducido en el cromatóforo sea consecuencia de un modelo aleatorio es menor del 5% lo que sugiere que la selección natural positiva es la causa de la especialización funcional del metabolismo del cromatóforo (Figura 11).

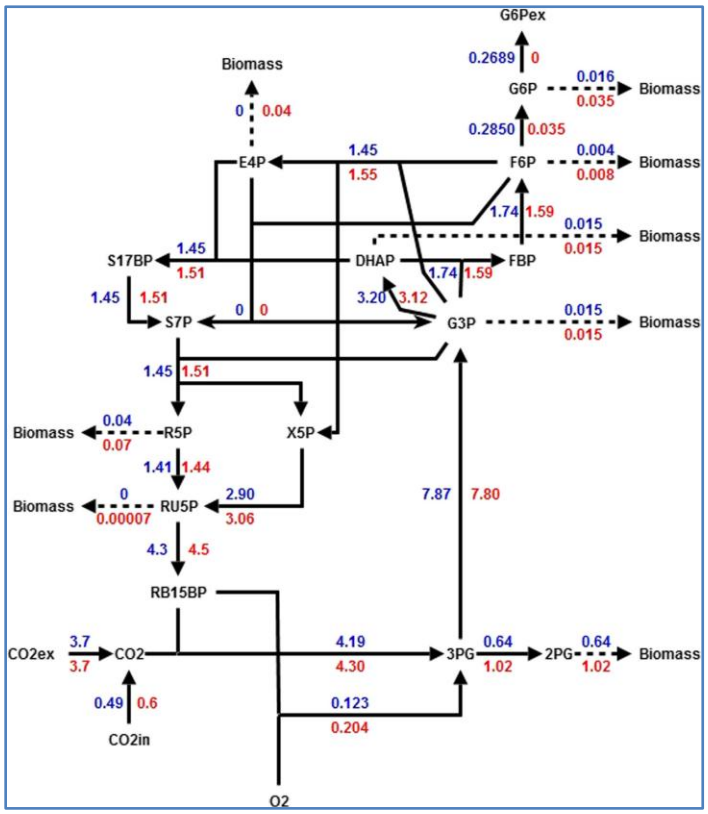

Figura 10. Distribución de los flujos metabólicos obtenidos con FBA del metabolismo central de los modelos del cromatóforo y *Synechoccocus sp*. WH 5701, azul y rojo, respectivamente.

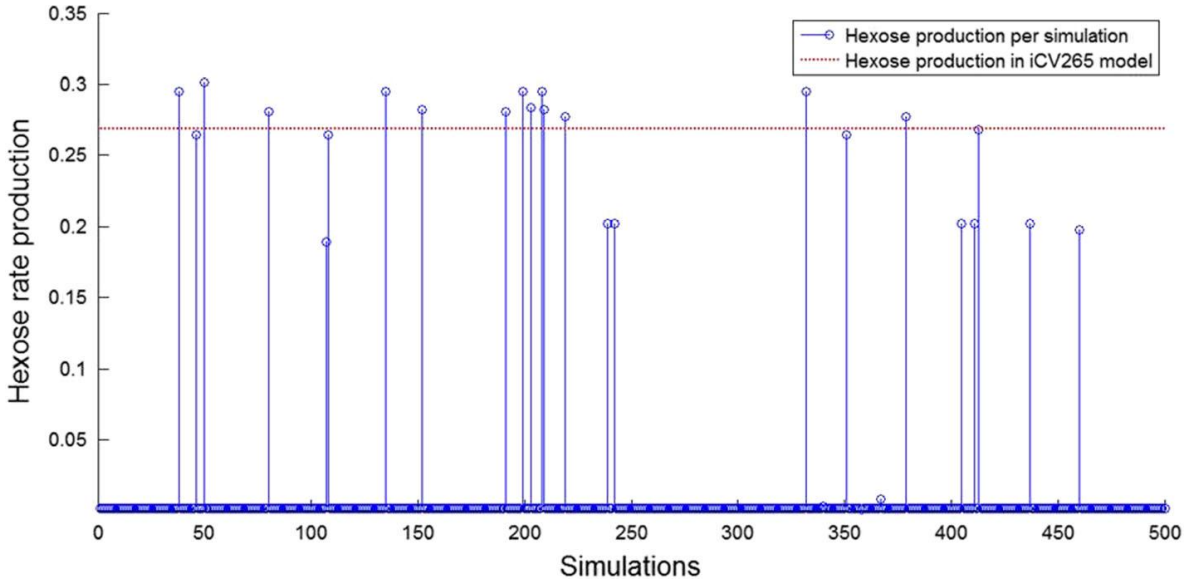

Figura 11. La tasa de exportación de hexosa (carbono reducido) en el modelo del cromatóforo se obtiene solamente en 2.6% de las simulaciones reductivas bajo un modelo de evolución reductiva aleatorio. La tasa de hexosa exportada (eje Y) en 500 simulaciones independientes (eje X). La línea punteada roja indica la tasa de hexosa exportada por el modelo del cromatóforo (*i*CV265).

Se propone el papel de la selección natural actuando como un "ingeniero metabólico" durante el proceso de especialización funcional de los endosimbiontes mutualistas. Esto considerando algunas de las estrategias para el redireccionamiento metabólico en las prácticas de ingeniería metabólica (pérdidas genéticas, suministro de precursores y reducción de flujos en rutas metabólicas competidoras (Pickens et al. 2011)) para la producción de un metabolito de particular interés. De la misma manera, el metabolismo del cromatóforo experimentó pérdidas genéticas, tuvo acceso a metabolitos precursores proporcionados por el hospedero y perdió rutas metabólicas completas. Además, el desarrollo obligatorio intracelular del cromatóforo sugiere que existe un cambio en su funcionalidad para la producción de metabolitos que tienen un impacto positivo en el holobionte, en lugar de respuestas a las presiones selectivas a las que están expuestos lo organismos de vida libre (Burgard et al. 2003).

4.2.3 Contribución al *artículo 2*

Para el *artículo 2*, llevé a cabo el planteamiento del proyecto, construcción de los modelos *i*CV265 e *i*CV498, las simulaciones evolutivas, interpretación de resultados y la escritura del manuscrito.

4.3 Contribución de *Paulinella chromatophora* al estudio de la evolución del proteoma durante la organelogénesis fotosintética

El proteoma de los plástidos contiene alrededor de 2,000 proteínas, de los cuales solamente una pequeña fracción son codificados en su genoma (los genomas de los plástidos codifican entre 50 y 200 proteínas). La mayoría de éstas proteínas presentes en los plástidos son codificados en el núcleo de la célula (ya sean de origen endosimbiótico o no) y son importados al plástido mediante un sofisticado aparato de importación de proteínas (TIC-TOC por sus siglas en inglés: *translocon complex of the inner and outer chloroplast membranes*) (Nakayama & Archibald 2012). El transporte de proteínas a través de TIC-TOC está facilitado por extensiones N-terminal llamadas "péptidos de transporte de cloroplasto" (cTPs, por sus siglas en inglés) de una longitud de 50-70 aminoácidos (Garg & Gould 2016). Una segunda alternativa al transporte de proteínas a los plástidos se realiza a través del sistema endomembranoso. El transporte se realiza co-traduccionalmente a través del retículo endosplasmático mediante la identificación de una péptido señal (SPs por sus siglas en inglés) (Villarejo et al. 2005).

La interacción "proteómica" establecida con el hospedero determinó la evolución reductiva que experimentó el endosimbionte. De igual manera, su regulación funcional pasó a estar determinada por el hospedero.

Debido a lo antiguo de la endosimbiosis primaria por la cual se originaron y lo complejo del sistema de transporte (sistema TIC-TOC) de los plástidos, es difícil indagar en los procesos iniciales que permitieron la interacción proteómica entre los simbiontes. Por tal motivo fue interesante el estudio del proteoma del organelo fotosintético en etapas iniciales de *P. chromatophora* el cual se llevó a cabo en el *artículo 3* "Massive protein import into the early evolutionary stage photosynthetic organelle of the amoeba *Paulinella chromatophora*".

4.3.1 Contribución del *artículo 3* al estudio del proteoma del cromatóforo de *Paulinella chromatophora*.

El proteoma de *P. chromatophora* fue caracterizado mediante espectrometría de masas. Algunas de las contribuciones realizadas en el *artículo 3* al estudio del proteoma fueron:

- En el proteoma del cromatóforo fueron identificados 641 proteínas (283 de ellas con un alto grado de confianza (*i.e*. se identificaron con al menos dos diferentes péptidos y cuando menos en dos réplicas)) de las cuales 422 son codificadas por el cromatóforo (238 con alto grado de confianza) lo que corresponde aproximadamente el 49% (27% con alta confianza) del total de 867 proteínas codificadas por el genoma del cromatóforo.
- Las 219 proteínas (45 con alta confidencia) restantes que fueron identificadas en el cromatóforo son codificadas por el núcleo de *P. chromatophora*.
- Se encontraron dos tipos de proteínas codificadas en el núcleo que son importadas al cromatóforo: el primer grupo incluye péptidos cortos de menos de 90 aminoácidos de largo y el segundo son proteínas con más de 268 aminoácidos.
- La función de las proteínas codificadas en el núcleo y que son importadas en el cromatóforo incluyen enzimas que involucradas en el metabolismo primario y llenan vacíos de rutas metabólicas que se encuentran incompletas en el cromatóforo.
- Las proteínas que con más de 268 aminoácidos conservan una extensión N-terminal de aproximadamente 200 aminoácidos (variando de 97-221 aminoácidos) que fue interpretado como un péptido de transporte del cromatóforo (crTP, por sus siglas en inglés “chromatophore transit peptide”). En éstas crTPs se pueden observar “motivos” altamente conservados.
- El desarrolló una herramienta computacional para identificar los motivos conservados en el transcriptoma de *P. chromatophora* permitió identificar secuencias del hospedero que posiblemente codifican para proteínas que pueden estar siendo importados al cromatóforo pero que no fueron identificados en el análisis experimental.
- De las 433 proteínas codificadas por el hospedero y que son candidatos a ser importados al cromatóforo (identificadas en análisis proteómico y la búsqueda *in silico* de motivos conservados), solamente 17 tienen un origen α-cianobacterial y posiblemente se originaron por EGT (de sus siglas en inglés “*endosimbiotic gene transfer*”), 26 de ellas tienen un origen bacteriano y el resto de ellas o bien tienen un origen en el hospedero o este es desconocido.
- La expresión heterologa de un gen con localización en el cromatóforo y motivos de transporte conservados en la planta de tabaco, permitió determinar que el crTP también permite la importación de proteínas a los plástidos de la planta de tabaco.

4.3.2 Contribución al *artículo 3*

En el análisis del proteoma, solamente 238 proteínas fueron identificadas de un total de 867 proteínas que son codificadas por el genoma del cromatóforo, lo que sugiere una eficiencia

del 27.5%. Dada la baja eficiencia del análisis proteómico, era altamente probable que mucha información se estuviera perdiendo. Por tal motivo, la búsqueda de proteínas codificadas por el hospedero e importadas en el cromatóforo se basó en la identificación de la extensión N-terminal previamente mencionada. La secuencia de la extensión N-terminal identificada no se encuentra muy bien conservada. En este sentido, mi contribución al *artículo 3* fue la propuesta de definir motivos que se encuentran altamente conservados en la extensión N-terminal de las proteínas con más de 268 aminoácidos y su utilización para la identificación *in silico* de proteínas que no fueron identificadas en el análisis proteómico. Lo anterior permitió incrementar el número de posibles proteínas codificadas por el hospedero e importadas al endosimbionte de 24 (encontradas en el análisis proteómico) a 291 (proteoma mas predichas *in silico*).

Además de contribuir al análisis de los datos y la interpretación metabólica de enzimas codificadas por el hospedero pero que son importadas en el cromatóforo.

4.4 Validación experimental de predicciones *in silico*

Si bien el trabajo realizado durante mi proyecto de doctorado fue totalmente teórico empleando herramientas computacionales para analizar la información disponible. El proteoma (en el *artículo 3*) y el transcriptoma disponible de *P. chromatophora* (Zhang et al. 2017), nos da una idea del poder predictivo que pueden tener el uso de las herramientas computacionales en el estudio de los organismos endosimbiontes.

En el *artículo 2*, en el modelo *i*CV265 (modelo del cromatóforo) se hace referencia a 44 reacciones enzimáticas las cuales son esenciales para el cromatóforo, pero que no son codificadas por el cromatóforo (revisar "*i*CV265 table", en *artículo 2*). De ésta manera entonces, el modelo predice que el hospedero tiene que estar codificando para dichas enzimas complementando así el metabolismo del cromatóforo.

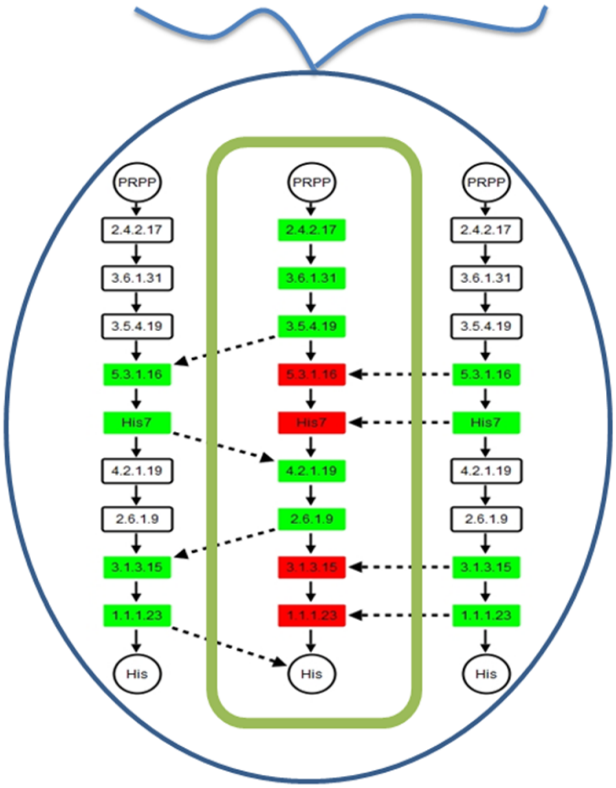

Figura 12. Complementación metabólica entre el cromatóforo y su hospedero. Dos posibilidades: intercambio activo de metabolitos (izquierda) e importación de enzimas al compartimento del cromatóforo (derecha).

Ese es el caso de la síntesis de histidina en el cromatóforo. En la ruta de la biosíntesis de la histidina, 4 reacciones se han perdido en el cromatóforo. Sin embargo, el modelo *i*CV265 predice que son esenciales y por lo tanto dichas enzimas deben estar siendo codificadas por el hospedero para complementar la ruta. La complementación puede ser ya sea por un transporte activo de metabolitos entre el cromatóforo y el citoplasma del hospedero (izquierda en Figura 12) o la importación de enzimas codificadas por el hospedero al interior del cromatóforo (derecha en Figura 12). En cualquiera de los dos casos, el modelo predice que las enzimas que complementan el metabolismo del cromatóforo, son codificadas por el hospedero.

Posteriormente, el acceso a la información del transcriptoma (Zhang et al. 2017) y el proteoma publicado en el *artículo 3* permitió determinar que de las 44 reacciones predichas como "codificadas por el hospedero" en el modelo *i*CV265, el hospedero codifica para enzimas que pueden llevar a cabo 34 de dichas reacciones. De las anteriormente mencionadas, 12 enzimas tienen la extensión N-terminal que permite el transporte de proteínas al interior del cromatóforo. Finalmente, de las proteínas codificadas por el hospedero y que complementan el metabolismo del cromatóforo, 6 enzimas fueron encontrados físicamente en el compartimento del cromatóforo.

5 Conclusiones

El análisis de organismos que tienen una relación simbiótica obligada, como es el caso de los organismos endosimbiontes puede estar fuertemente restringido por su incapacidad de ser cultivados sin su hospedero. En este sentido, las herramientas computacionales ofrecen una buena alternativa para el estudio tanto evolutivo como funcional de dichos organismos (Holman et al. 2009).

El uso de herramientas computacionales durante mi proyecto de investigación me permitieron indagar en varios aspectos evolutivos del organelo fotosintético de *Paulinella chromatophora* a pesar de no tener acceso físico al organismo. De igual manera, durante la estancia realizada con la Dra. Eva Nowack en el laboratorio de "Simbiosis Microbiana y Evolución de Organelos" pude constatar que algunas de las predicciones realizadas en el *artículo 2*, con respecto a enzimas codificadas por el hospedero que son esenciales y complementan el metabolismo del cromatóforo, realmente están siendo importadas en el interior del cromatóforo (revisar *artículo 3*). Lo anterior enfatiza el poder predictivo de las herramientas computacionales.

Dado la similitud de eventos de endosimbiosis primaria entre los plástidos y el cromatóforo, es altamente probable que las conclusiones del presente proyecto hayan ocurrido también en los plástidos durante las etapas iniciales de la organelogenesis. Por ejemplo, la interacción metabólica, genética y proteómica establecida con el hospedero condicionó la evolución reductiva experimentada por el endosimbionte fotosintético. Además, dado la importancia de la complementación nutricional del hospedero por el endosimbionte, es muy probable que la selección natural positiva al nivel del holobionte hayan determinado las reestructuraciones metabólicas que permitieron la especialización funcional de los plástidos como productores de fotosintatos.

6 Artículos

6.1 Artículo 1. How really ancient is *Paulinella chromatophora*.

PLOS | CURRENTS TREE OF LIFE

How Really Ancient Is Paulinella Chromatophora?

March 15, 2016 · *Tree of Life*

Citation

Delaye L, Valadez-Cano C, Pérez-Zamorano B. How Really Ancient Is Paulinella Chromatophora?. PLOS Currents Tree of Life. 2016 Mar 15 . Edition 1. doi: 10.1371/currents.tol.e68a099364bb1a1e129a17b4e06b0c6b.

Tweet

Authors

Luis Delaye

Ingeniería Genética, CINVESTAV Irapuato, Irapuato, Guanajuato, México.

Cecilio Valadez-Cano

Ingeniería Genética, CINVESTAV - Irapuato; Irapuato, Guanajuato, México.

Bernardo Pérez-Zamorano

Ingeniería Genética, CINVESTAV - Irapuato; Irapuato, Guanajuato, Mexico.

Abstract

The ancestor of Paulinella chromatophora established a symbiotic relationship with cyanobacteria related to the Prochloroccocus/Synechococcus clade. This event has been described as a second primary endosymbiosis leading to a plastid in the making. Based on the rate of pseudogene disintegration in the endosymbiotic bacteria Buchnera aphidicola, it was suggested that the chromatophore in P. chromatophora has a minimum age of ~60 Myr. Here we revisit this estimation by using a lognormal relaxed molecular clock on the 18S rRNA of P. chromatophora. Our time estimates show that depending on the assumptions made to calibrate the molecular clock, P. chromatophora diverged from heterotrophic Paulinella spp. ~ 90 to 140 Myr ago, thus establishing a maximum date for the origin of the chromatophore.

Funding Statement

BPZ wishes to thank CINVESTAV Irapuato for funding and support. The authors also declare that no competing interests exist.

Introduction

Mitochondria and plastids evolved from freeliving bacteria by symbiogenesis more than one billion years

ago[1]. Both events boosted the evolution of eukaryotes by expanding their metabolic abilities. Primary endosymbiosis leading to organelles (i.e., mitochondria and plastids) was thought to be unique in the history of life until the recent discovery of an independent primary endosymbiosis in *Paulinella chromatophora*[2]. This thecate filose amoeba hosts in its cytoplasm photosynthetic organelles of cyanobacterial origin, called chromatophores. Phylogenetic analysis of 16S rRNA showed that the chromatophores originated from marine αcyanobacteria from the *Prochlorococcus*/*Synechococcus* clade. In contrast, "classical" plastids of plants and algae evolved from an ancient unknown lineage of cyanobacteria[3].

The genome of the cyanobacteria *Synechococcus* WH5701, one of the closest known free-living relatives of the chromatophore with a sequenced genome, has 2917 protein-coding genes[4]. In contrast, the chromatophores of *P. chromatophora* strains FK01 and M0880/a contain between 841 and 867 protein-coding genes respectively[5,6]. Remaining genes in the chromatophore suggest a strong metabolic interdependence with the amoebal nucleocytoplasm. The discovery that proteins of the chromatophore photosynthetic apparatus are encoded in the host genome and imported back into the cyanobacterial-derived compartment, reinforces the suggestion that the chromatophore is a bona fide primary organelle[7].

It is likely that the origin of the chromatophore is one or two orders of magnitude more recent than the establishment of the primary plastids of plants and algae. But, how ancient is the chromatophore? Since there is no fossil record of *P. chromatophora* or its close relatives, it is difficult to have a precise answer to this question. However, an initial guess suggested the chromatophore has a minimum age of 60 Myr. This was based on the proposal that in *Buchnera aphidicola* (Aphid's endosymbiotic bacteria) a pseudogene needs between 40-60 Myr to disintegrate completely[8]. Since the genome of the chromatophore has pseudogenes, the same tempo was extrapolated for the chromatophore[6].

Surprisingly, the suggestion that the chromatophore in *P. chromatophora* has a minimum age of 60 Myr underwent a change in part of the subsequent literature indicating that *P. chromatophora* has ~60 Myr. Clearly, this is an assertion that requires clarification and further analysis. The recent identification by single cell genomics of non-photosynthetic close relatives of *P. chromatophora*[9] and novel attempts to calibrate the origin of major eukaryotic groups[10,11] offer an opportunity to revisit the origin in time of this extraordinary symbiosis.

Methods

Taxon sampling. Based on previous published phylogenetic analyses we retrieved a sample of 18S rRNA sequences from rhizaria and stramenopila from SILVA database[12]. These included sequences from: i) the phylogenetic analysis whereby close relatives of *P. cromatophora* were identified[9]; ii) the phylogenetic analyses describing divergence times of major eukaryotic groups[10,11]; and iii) sequences from phylogenetic analyses of euglyphids[13] and diatoms[14]. We did not include sequences from foraminifera due to their high rate of evolution reflected in extreme long branches. Sequences were aligned with MUSCLE[15]. The final alignment contained 43 sequences and 1128 aligned positions without gaps. The phylogenetic tree reconstructed from these sequences (see below) is in general terms congruent with published phylogenetic analyses.

Molecular clock analyses. BEAUTi was used to prepare xml files. Time trees were inferred by using a lognormal relaxed molecular clock as implemented in BEAST 2[16]. To select the model of evolution we

used jModelTest[17]. The Tamura and Nei 1993 model (TrN) plus gamma distribution and an estimated proportion of invariant sites were rated best by the bayesian information criterion (BIC). Therefore, all the analyses were conducted under this model of evolution. In addition, we used a Yule tree prior for all analyses.

To calibrate the clock, we relied upon several sources of information (Fig 2). These included organismal as well as chemical fossils. We also used previous published estimates of the time divergence between rhizaria and stramenoplia. These sources of information were organized into four different calibration schemes. All four schemes used information from non-controversial fossil record (evidences: a, b, c and d). However, the schemes differ on the use of soft evidence from previous molecular clock analyses (evidence: e.1 and e.2); and on the use of evidence from vaseshaped microfossils (VSM) described by[18]. VSM were originally assigned to rhizaria. However, a posterior molecular clock analysis favored a reinterpretation of VSM as members of amoebozoa[10]. Thus, the four calibration schemes were: (a, b, c, d, e.1); (a, b, c, d, e.1, f); (a, b, c, d, e.2); (a, b, c, d).

Priors on the age of nodes were adjusted to reflect confidence on time divergence. For instance, we assigned a strong prior on the origin of rhizosolenid diatoms because the origin of this lineage has been dated to 91.5 ± 1.5 Myr based on chemical fossils[14]. The 95% probability distribution on this case is between 91.5 to 97.0 Myr. Other calibrations received less strong priors, as indicated by the larger number of years contained along the 2.5% to 9.5% quantiles of the shape column from Table 1. Priors used for the origin of pennate diatoms and for the origins of earliest diatoms were based on those used by[11] and represent minimal divergence times. The prior used for the origin of Eugliphydae was based on the supposition that the group is much older than the fossil evidence as suggested by[19]. This prior also represents a minimal divergence time. Finally, to assign a prior to the divergence in time of rhizaria from stramenopila we first averaged BEAST time estimates published by[11] and then we parametrized a normal distribution to include the variability of these estimates. We used this normal distribution as a prior. We constructed a second prior for the same divergence of rhizaria from stramenopila but now based on the inference published by[10]. All sequences within each prior were restricted to be monophyletic. The root of the tree was determined by making rhizaria monophyletic.

We evaluated each calibration scheme by looking at convergence and ESS > 200 in Tracer after running chains of length 10^{10} and sampling each 10^4 generations. We discarded 10% of generations as burnin. Finally, we conducted all analyses without sequences (i.e., sampling from priors) to determine the impact of priors and to test whether sequences are informative on estimated divergence dates. Trees were obtained with TreeAnnotator and visualised with FigTree. We provide BEAUTi xml files of the four calibration schemes for BEAST 2 analyses here.

Results and discussion

In Fig. 1 we show an estimate of the divergence in time of *P. chromatophora* by using a lognormal relaxed clock on 18S rRNA. To calibrate this tree we relied on: a) the origin of rhizosolenid diatoms, which is known with high confidence (91.5 ± 1.5 Myr)[14]; b) a minimal time divergence of pennate diatoms (80 Myr)[11,20]; c) a minimal time divergence for diatoms (133.9 Myr)[11,21]; d) a minimal time divergence of Euglyphidae 40 Myr[19]; and e.1) a time estimation of the divergence of rhizaria from stramenopila ~1232 Myr[11]. This gave us an estimation for the origin of *P. chromatophora* of 93.6 Myr (38.1 – 138.2).

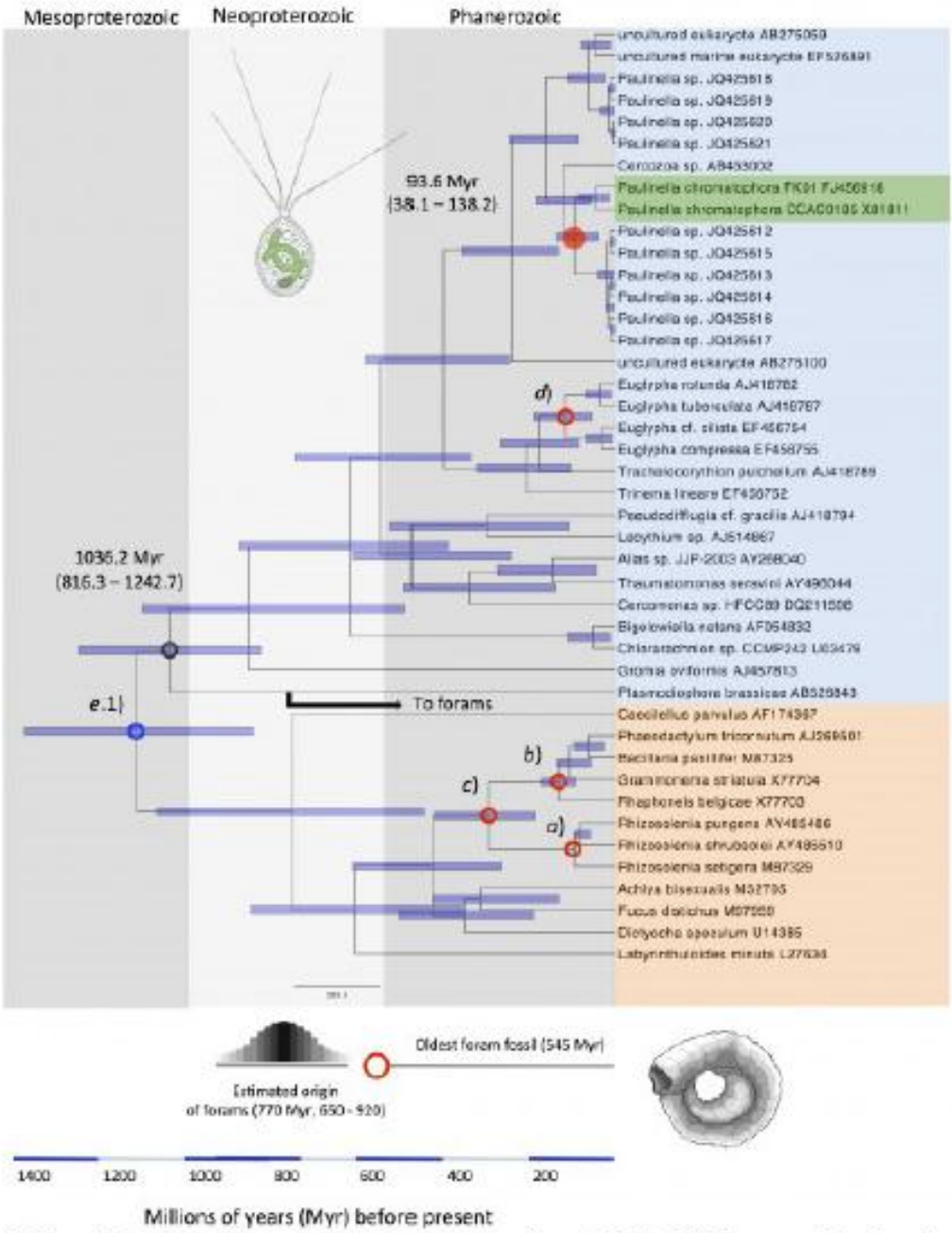

Fig. 1: The origin of P. chromatophora in time according to SSU rRNA lognormal molecular clock.

Tree calibrated under the scheme (a, b, c, d, e.1). Small open red circles indicate calibrations based on fossil record (a, b, c, and d); small open blue circle indicate calibration based on previous estimation (e.1); red full circle, estimated age of *P. chromatophora*. Notice that the divergence in time of *Plasmodiophorabrassicae* (black open circle) is consistent with the proposed origin in time of forams. Species names in orange: stramenopila; species name in blue and green: rhizaria.

Evidence	Min	Distribution	Shape	Ref.
(*a*) Rhizosolenid diatoms Chemical fossil	91.5	γ (2, 1)	93.2 (92 – 97)	[14]
(*b*) Pennate diatoms Fossil	80.0	γ (3, 10)	107 (86 – 152)	[20]
(*c*) Earliest diatoms Fossil	133.9	γ (2, 100)	302 (158 – 691)	[21]
(*d*) Euglyphidae Silicious plates from the Middle Eocene match to modern species *Euglypha* and *Scuttglypha*	40.0	γ (2, 200)	376 (88 – 989)	[19]
(*e.1*) Divergence of rhizaria from stramenopila Molecular clock estimate	1232.5	Normal (0, 100)	1230 (1040 – 1430)	[11]
(*e.2*) Divergence of rhizaria from stramenopila Molecular clock estimate	754.1	Normal (0, 100)	754.5 (559 – 950)	[10]
(*f*) Vase-shape microfossils (VSM) Eukaryotic microfossils of controversial affiliation: could belong to filose euglyphids or to lobose arcellinids	742.0	γ (2, 40)	809 (752 – 965)	[17]

Fig. 2: Calibration constrains.

Evidence: used to derive priors for calibration constraints; Min: Minimal divergence time (Offset parameter in BEAUTi) in Myr; Distribution: Used to model each prior together with their respective parameter values (alpha α and beta β for gamma γ and mean and sigma σ for normal distributions); Shape: shows the median for each distribution and the values containing the 2.5% and 97.5% quantiles. Stronger priors span few years between the 2.5% and 97.5% quantiles.

However other calibration schemes are possible. For instance, we can use the report of vase-shaped microfossils (VSM) of ~742 to 770 Myr (calibration f) that were described tentatively as members of Euglyphida[18] together with calibrations a, b, c, d and e.1, to get an estimate of 141.4 Myr (48.7 – 210.4) for the origin of *P. chromatophora*. Alternatively, we can follow the suggestions that VSMs belong to amoebozoa and that rhizaria diverged from stramenopila 754.1 Myr (639.8 – 903.3)[10] together with calibration points a, b, c, d and e.2 to get an estimate of 55.8 Myr (25.4 – 78.6) for the origin of *P. chromatophora*. Finally, we can estimate the divergence of *P. chromatophora* based only on calibration points a, b, c and d; thus avoiding controversial VSM fossils (calibration f) and the soft information provided from previous estimates of the divergence between rhizaria and stramenopila (calibrations e.1 and e.2). By this, we get an estimate of 47.9 Myr (28.8 – 64.9).

Which of the four time estimates is closer to the true divergence time of *P. chromatophora*? Foraminifera offer a clue. Foraminifera belong to rhizaria and as well as diatoms have a detailed fossil record. However, their 18S sequences are too divergent to include them in our analysis. The first appearance of forams in the fossil record is dated at 545 Myr[22]. Recent molecular time estimates of forams suggest this group originated 770 Myr (650 – 920)[23]. A phylogenetic analysis based on 15 concatenated genes shows that forams are a sister group of *Plasmodiophorabrassicae* with the exclusion of *Gromiaoviformis* as an outgroup[11]. Therefore, the origin of the lineage leading to *P. brassicae* in our tree has to be at least as old as the proposed origin of forams. By looking at our time estimates, it happens that only the two proposals that are based on a time of ~1232 Myr for the origin of rhizaria (i.e., that uses calibration e.1) result in a date for the origin of *P. brassicae* that is congruent with the estimated origin of forams (Fig. 3).

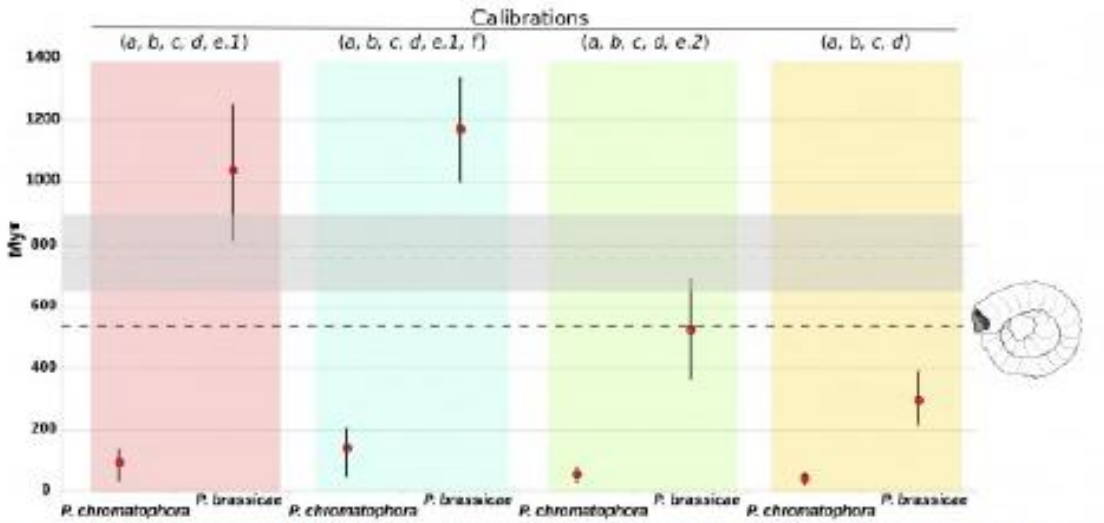

Fig. 3: Concordance between different time divergence inferences and earliest forams in the fossil record.

Only calibration schemes assuming a time divergence of rhizaria from stramenopila of ~1232 Myr (e.1) are consistent with the fossil record and proposed time divergence of forams. The horizontal black broken line represents the date of the oldest foram fossil at 545 Myr. The blue broken line represents the estimated origin of forams 770 Myr (650 – 920).

Although the 95% confidence intervals are rater large, the estimates of 93.6 and 141.4 Myr are consistent with the original suggestion that the chromatophore in *P. chromatophora* has a minimum age of 60 Myr[5]; and with a more recent proposal of a maximum age of 200 Myr for the divergence of *P. chromatophora* from Euglypha rotunda[24].

If our estimates are correct, the two strains of *P. chromatophora* diverged from each other about (45.7 – 64.7 Myr). The genomes of the chromatophores in these two strains differ in about 33 genes[5]. This give us a rate of ~ 0.25 to 0.36 gene inactivations per million year since the divergence of the two strains. These show that genome reduction in the chromatophore is a process that continues slowly however stately. Whether these genes are been lost by genetic drift or natural selection is still a mater of research. Further refinements on the time of origin of *P. chromatophora* and their chromatophores are expected as our understanding of the evolution of rhizaria and the eukaryotic tree of life improves.

Acknowledgements

We thank Professor Adrian Reyes-Prieto, Mtra. Diana Rossetti and an anonymous reviewer for useful suggestions to the manuscript. We also thank Sofía Delaye for providing drawings of protists.

References

1. Sagan, L. (1967). On the origin of mitosing cells. J Theor Biol. 14: 255-274.

2. Marin, B., Nowack, E.C.M, and Melkonian, M. (2005). A plastid in the making: evidence for a second primary endosymbiosis. Protist. 156: 425-432.

3. Criscuolo, A., Gribaldo, S. (2011) Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. Mol Biol Evol. 28:3019-3032.

4. Qiu, H., Yang, E.C., Bhattacharya, D., Yoon, H.S. (2012) Ancient Gene Paralogy May Mislead Inference of Plastid Phylogeny. Mol Biol Evol. 29:3333-3343.

5. Reyes-Prieto, A., Yoon, H.S., Moustafa, A., Yang, E.C., Andersen, R.A., Boo, S.M., Nakayama, T., Ishida, K., Bhattacharya, D. (2010) Differential gene retention in plastids of common recent origin. Mol Biol Evol. 27:1530-1537.

6. Nowack, E.C.M., Melkonian, M. and Glöckner G. (2008) Chromatophore genome sequence of Paulinella sheds light on acquisition of photosynthesis by eukaryotes. Curr. Biol. 18: 410-418.

7. Nowack, E.C., Grossman, A.R. (2012). Trafficking of protein into the recently established photosynthetic organelles of Paulinella chromatophora. Proc Natl Acad Sci U S A. 109: 5340–5345. doi: 10.1073/pnas.1118800109.

8. Gómez-Valero, L., Latorre, A., Silva, F.J. (2004). The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont Buchnera aphidicola. Mol Biol Evol. 21:2172-2181.

9. Bhattacharya, D., Price, D.C., Yoon, H.S., Yang, E.C., Poulton, N.J., Andersen, R.A., Das, S.P. (2012). Single cell genome analysis supports a link between phagotrophy and primary plastid endosymbiosis. Sci. Rep. 2:356. doi: 10.1038/srep00356.

10. Berney, C. and Pawlowski, J. (2006). A molecular time-scale for eukaryote evolution recalibrated with

the continuous microfossil record. Proc. R. Soc. B. 273: 1867-1872.

11. Parfrey, L.W., Lahr, D.J., Knoll, A.H., Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. U S A. 108: 13624-13629. doi: 10.1073/pnas.1110633108.

12. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl. Acids Res. 41 (D1): D590-D596.

13. Heger, T.J., Mitchell, E.A., Todorov, M., Golemansky, V., Lara, E., Leander, B.S., Pawlowski, J. (2010). Molecular phylogeny of euglyphid testate amoebae (Cercozoa: Euglyphida) suggests transitions between marine supralittoral and freshwater/terrestrial environments are infrequent. Mol Phylogenet Evol. 55: 113-122.

14. Damsté, J.S., Muyzer, G., Abbas, B., Rampen, S.W., Massé, G., Allard, W.G., Belt, S.T., Robert, J.M., Rowland, S.J., Moldowan, J.M., Barbanti, S.M., Fago, F.J., Denisevich, P., Dahl, J., Trindade, L.A., Schouten, S. (2004). The rise of the rhizosolenid diatoms. Science. 304: 584-587.

15. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792-1797.

16. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H, et al. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol 10(4): e1003537. doi:10.1371/journal.pcbi.1003537.

17. Darriba, D., Taboada, G.L., Doallo, R., Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nature Methods. 9: 772.

18. Porter, S.M., Knoll, A.H. (2000). Testate amoebae in the Neoproterozoic Era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. Paleobiology. 26: 360-385.

19. Barber, A., Siver, P.A., Karis, W. (2013). Euglyphid testate amoebae (rhizaria: euglyphida) from an arctic eocene waterbody: evidence of evolutionary stasis in plate morphology for over 40 million years. Protist. 164: 541-555.

20. Kooistra, W., Gersonde, R., Medlin, L., Mann, D.G. (2007). In The Evolution of Primary Producers in the Sea, The origin and evolution of the diatoms: Their adaptation to a planktonic existence, eds Falkowski PG, Knoll AH (Elsevier, Burlington, MA), pp 201–249.

21. Harwood, D.M., Nikolaev, V.A., Winter, D.M. (2007). Cretaceous records of diatom evolution, radiation, and expansion. Paleontol Soc Papers. 13:33–59.

22. Culver, S.J. (1991). Early Cambrian Foraminifera from West Africa. Science. 254:689-691.

23. Groussin, M., Pawlowski, J., Yang, Z. (2011). Bayesian relaxed clock estimation of divergence times in foraminifera. Mol Phylogenet Evol. 61: 157-166.

24. Nowack, E.C. (2014). Paulinella chromatophora – rethinking the transition from endosymbiont to organelle. Acta Soc Bot Pol. 83: 387-397.

6.2 Artículo 2. Natural selection drove metabolic specialization of the chromatophore in *Paulinella chromatophora.*

Valadez-Cano et al. BMC Evolutionary Biology (2017) 17:99
DOI 10.1186/s12862-017-0947-6

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Natural selection drove metabolic specialization of the chromatophore in *Paulinella chromatophora*

Cecilio Valadez-Cano[1], Roberto Olivares-Hernández[2], Osbaldo Resendis-Antonio[3,4], Alexander DeLuna[5] and Luis Delaye[1*]

Abstract

Background: Genome degradation of host-restricted mutualistic endosymbionts has been attributed to inactivating mutations and genetic drift while genes coding for host-relevant functions are conserved by purifying selection. Unlike their free-living relatives, the metabolism of mutualistic endosymbionts and endosymbiont-originated organelles is specialized in the production of metabolites which are released to the host. This specialization suggests that natural selection crafted these metabolic adaptations. In this work, we analyzed the evolution of the metabolism of the chromatophore of *Paulinella chromatophora* by in silico modeling. We asked whether genome reduction is driven by metabolic engineering strategies resulted from the interaction with the host. As its widely known, the loss of enzyme coding genes leads to metabolic network restructuring sometimes improving the production rates. In this case, the production rate of reduced-carbon in the metabolism of the chromatophore.

Results: We reconstructed the metabolic networks of the chromatophore of *P. chromatophora* CCAC 0185 and a close free-living relative, the cyanobacterium *Synechococcus sp.* WH 5701. We found that the evolution of free-living to host-restricted lifestyle rendered a fragile metabolic network where >80% of genes in the chromatophore are essential for metabolic functionality. Despite the lack of experimental information, the metabolic reconstruction of the chromatophore suggests that the host provides several metabolites to the endosymbiont. By using these metabolites as intracellular conditions, in silico simulations of genome evolution by gene lose recover with 77% accuracy the actual metabolic gene content of the chromatophore. Also, the metabolic model of the chromatophore allowed us to predict by flux balance analysis a maximum rate of reduced-carbon released by the endosymbiont to the host. By inspecting the central metabolism of the chromatophore and the free-living cyanobacteria we found that by improvements in the gluconeogenic pathway the metabolism of the endosymbiont uses more efficiently the carbon source for reduced-carbon production. In addition, our in silico simulations of the evolutionary process leading to the reduced metabolic network of the chromatophore showed that the predicted rate of released reduced-carbon is obtained in less than 5% of the times under a process guided by random gene deletion and genetic drift. We interpret previous findings as evidence that natural selection at holobiont level shaped the rate at which reduced-carbon is exported to the host. Finally, our model also predicts that the ABC phosphate transporter (pstSACB) which is conserved in the genome of the chromatophore of *P. chromatophora* strain CCAC 0185 is a necessary component to release reduced-carbon molecules to the host.

(Continued on next page)

*Correspondence: luis.delaye@cinvestav.mx
[1]Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Irapuato, Km. 9.6 Libramiento Norte Carr. Irapuato-León, 36821 Guanajuato, Irapuato, Mexico
Full list of author information is available at the end of the article

BioMed Central

© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

(Continued from previous page)

Conclusion: Our evolutionary analysis suggests that in the case of *Paulinella chromatophora* natural selection at the holobiont level played a prominent role in shaping the metabolic specialization of the chromatophore. We propose that natural selection acted as a "metabolic engineer" by favoring metabolic restructurings that led to an increased release of reduced-carbon to the host.

Keywords: Endosymbiont, Metabolic evolution, Adaptation, Metabolic integration

Background

Paulinella chromatophora is an amoeba dispensed with phototrophic nutrition that contains blue-green photosynthetic organelles of cyanobacterial origin termed chromatophores [1, 2]. These novel organelles have a monophyletic origin in different strains of photosynthetic *Paulinella* that have been described [3] and were acquired through a primary endosymbiotic event about ~90 to 140 Mya [2–6].

Chromatophore genome sequencing from two strains of *P. chromatophora* (FK 01 [7] and CCAC 0185 [5]), revealed a size of 0.977 and 1.02 Mbp, respectively. This represents about 1/3 of the genome size of *Synechococcus sp.* WH 5701, the closest free-living relative cyanobacterium with a sequenced genome. *Synechococcus sp.* WH 5701 has a genome of ~3 Mbp and 3346 protein-coding genes [5]. It indicates that the chromatophore evolved by genome reduction. However, genome reduction in *P. chromatophora* is not as extreme as in plastids which rarely exceed 200 Kbp [2].

Chromatophores are genetically integrated with their host. More than 30 nuclear encoded genes of chromatophore origin have been identified [7, 8]. And some of the protein products coded by these genes are imported back into the chromatophore and participate in the photosynthetic apparatus [9]. Accordingly, chromatophores have been described as plastids in the making.

P. chromatophora nutrition relies on the reduced-carbon photosynthetically assimilated by the chromatophore [10]. This endosymbiotic-nutrient dependency has been observed in other organisms such as aphids and tsetse flies housing prokaryotic endosymbionts [11]. Particularly for aphids, host essential amino acids are provided by an endosymbiotic bacterium called *Buchnera aphidicola* [12]. Sequencing of the genome of *B. aphidicola* revealed a high degree of genetic degradation, while genes necessary for the syntrophic relationship with its host have been retained [12].

Prokaryotic endosymbionts evolve small genomes mainly by the combined action of genetic drift and negative selection [13–16]. In host-restricted conditions, the endosymbiont experiences a lack of recombination and horizontal gene transfer, as well as recurrent population bottlenecks lowering its effective population size (N_e) and a concomitant relaxation of natural selection [15–17]. The combined action of these factors allows the accumulation of slightly deleterious mutations through a process called Muller's ratchet [14, 17]. As a consequence, many genes become pseudogenes and are subsequently lost. In addition, selection at holobiont level by mechanisms like "partner fidelity feedback" have been proposed to promote the evolution of mutualistic interactions [18].

Something that should be considered is that, differing from free-living relatives, the metabolism of mutualistic endosymbionts is specialized in the production of metabolites that are released to their host as nutrients [19, 20]. This metabolic specialization is the consequence of metabolic restructuring caused by gene loss and genome reduction [20]. Resulting reduced genomes code for fewer genes, however, they are more integrated to the host. The extreme cases are organelles of endosymbiotic origin such as chloroplasts [21]. Therefore, if mutualistic endosymbionts show metabolic adaptations to provide nutrients to their hosts [19, 20], natural selection must have participated in the evolution of these systems.

During early stages of organellogenesis, the cyanobacteria that evolved into the chromatophore, had access to metabolites provided by the host. It is likely that the availability these metabolites render of some metabolic routes dispensable in the endosymbiont. The loss of these biosynthetic pathways in the endosymbiont led to restructurings and changes in the remaining metabolic fluxes. Taking into consideration all these modifications experienced by the chromatophore and the nutrient dependency of the holobiont for the photosynthetic function of the chromatophore, we made the analogy of natural selection acting as a "metabolic engineer" directing the strategies for the metabolic specialization of the chromatophore. In general, the objective of metabolic engineering is the directed improvement of metabolic capabilities through the deletion of metabolic genes or the introduction of new ones [22]. By using these strategies, microorganisms have been engineered for the improvement of the yield and the production and consumption rates of desired metabolites. For instance, for the of 1-butanol production in cyanobacteria [23], many more examples can be found elsewhere [24, 25].

In this work, we reconstructed the genome based metabolic models of the chromatophore of *Paulinella*

chromatophora and the cyanobacteria *Synechococcus sp.* WH 5701. We inquired into the metabolic capabilities of the chromatophore; the possible metabolic interaction of the chromatophore with its host; and in silico simulate the process of metabolic evolution experienced by the chromatophore in host-restricted conditions.

Results

Differential gene retention of functional categories in the chromatophore genome

Our first objective was to determine to what extent genetic loss affected functional metabolic categories in the chromatophore (i.e. which functional gene categories were preferentially preserved) when compared to the genome of *Synechococcus sp.* WH 5701. We compared against *Synechococcus sp.* WH 5701 because is the closest free-living cyanobacterium with a sequenced genome and it is likely to be similar in gene content to the ancestor of the chromatophore. To assess the statistical significance we used a hypergeometric distribution.

As is shown in Fig. 1, genes belonging to 13 functional categories have been less affected by genome erosion. In particular, photosynthesis and fatty acid biosynthesis categories are less affected. Retention of these 13 functional categories in the chromatophore can be attributed to a host-level selection protecting from gene loss. Conserved genes very likely play an adaptive role in the holobiont.

In silico metabolic reconstruction of the chromatophore of *P. chromatophora* and *Synechococcus sp.* WH 5701

To better understand the role in the symbiosis played by remaining genes in the chromatophore, we reconstructed two metabolic models. One for the chromatophore of *P. chromatophora* CCAC 0185 [5] and the other for *Synechococcus sp.* WH 5701, the closest free-living cyanobacterium with a sequenced genome. The rationale behind this is to use *Synechococcus sp.* WH 5701 as a proxy of the ancestral cyanobacterium that evolved into the chromatophore.

Metabolic model reconstruction of the free-living cyanobacterium *Synechococcus sp.* WH 5701 was done by identifying orthologs to those protein-coding genes reported in the metabolic model of *Synechocystis sp.* PCC 6803 (*i*JN678) [26]. The resulting metabolic model of the free-living organism (*i*CV498) comprised 743 metabolic reactions with 698 metabolites and 498 protein-coding genes. Metabolic model reconstruction of the chromatophore was done by identifying those genes in the genome of the chromatophore of *P. chromatophora* CCAC 0185 that are orthologous to the free-living metabolic model (*i*CV498). The metabolic model of the chromatophore (*i*CV265) comprised 627 reactions, 615 metabolites and 265 protein-coding genes. Because *Synechococcus sp.* WH 5701 is a close free-living relative of the chromatophore, it could be considered that 158 reactions were lost along genome reduction in the chromatophore (Table 1).

By using the biomass equation of the cyanobacterium *Synechocystis sp.* PCC 6803 [26], we tested the functionality of the *i*CV498 and the *i*CV265 metabolic models with Flux Balance Analysis (FBA). Biomass production was set as objective function. In silico growth was simulated under autotrophic conditions. CO_2 and photons uptake were set to 3.7 mmol × gDW^{-1} × h^{-1} and 100 mmol × gDW^{-1} × h^{-1} respectively and set as constraining metabolites as in [26].

In model *i*CV498, almost every metabolic pathway for biomass production is complete. The exceptions were 9 reactions for which no orthologous exist in *Synechococcus sp.* WH 5701 when compared to *i*JN678 (see model *i*CV498 in Additional file 1). These reactions had to be added to the *i*CV498 model in order to produce all the components necessary for the biomass equation. In this way, *i*CV498 showed an in silico growth rate of 0.0884 h^{-1} which is identical to the in silico growth reported for *Synechocystis sp.* PCC 6803 metabolic model under autotrophic conditions [26].

Under these conditions, the metabolic model of the chromatophore (*i*CV265) did not show in silico growth. This was obviously due to the reduced metabolic capabilities caused by the genomic reduction process experimented by the photosynthetic endosymbiont. Genome reduction has affected the metabolic capabilities of the chromatophore in two ways: a) some biosynthetic pathways were completely lost; while b) some other were partially lost.

For example, in *Synechocystis sp.* PCC 6803 riboflavin is synthesized by four genes that perform six reactions by using Guanosine 5′-triphosphate (G5P) and D-Ribulose 5-phosphate (R5P) as precursors metabolites [26]. All these genes for riboflavin biosynthesis were lost in the chromatophore. In this case, we assumed that the host provides riboflavin to the chromatophore. The possible explanation for this loss is that riboflavin is the main precursor for flavin mononucleotide (riboflavin 5′-monophosphate, FMN) and flavin adenine dinucleotide, two main compounds that work as coenzymes for many of the enzymes such as oxidoreductases including NADH dehydrogenase as well as in biological blue-light photo receptors. This observation is concomitant with the loss in some functional gene categories; as in oxidative phosphorylation (Fig. 1). As the hypothesis is that the metabolic network must preserve its functionality, whenever we found a similar situation, exchange reactions were added to the metabolic model to simulate the incorporation of riboflavin and other metabolites as additional nutrients from the host. These metabolites included amino acids, cofactors, vitamins and other

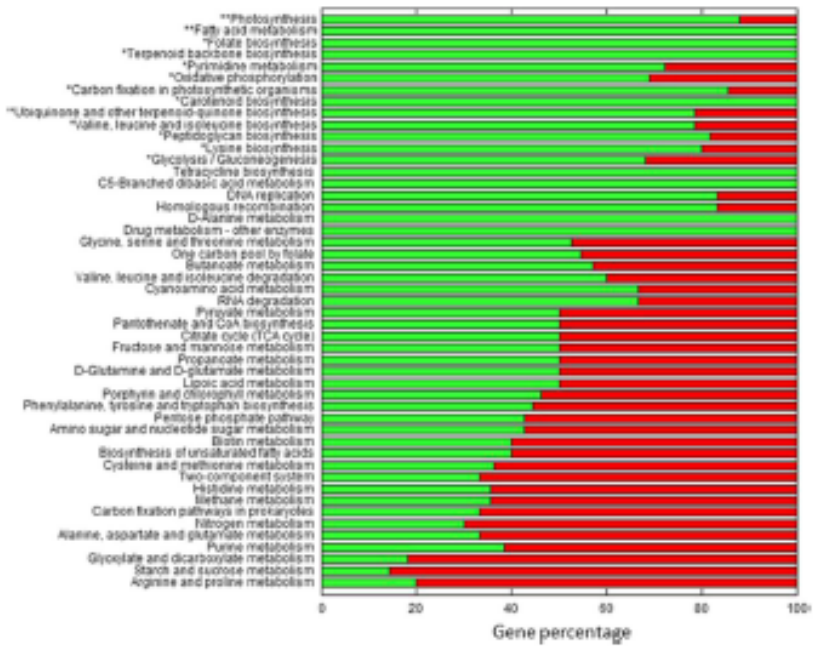

Fig. 1 Conservation of functional gene categories in the chromatophore when compared to *Synechococcus sp.* WH 5701. For each functional category we show in green and red the percentage of gene conservation and lost in the chromatophore, respectively. For instance, if a gene category is completely green, it indicates that all orthologs in *Synechococcus sp.* WH 5701 are conserved in the chromatophore. As shown, gene loss affects differentially each one of the functional categories in the chromatophore. Functional categories particularly well conserved are indicated with asterisks (*p*-value < 0.05* or <0.05**, Bonferroni corrected). Statistical significance calculated by using a hypergeometric distribution [63]. Genes were classified following KEGG database (http://www.kegg.jp)

molecules which are essential for the biomass equation but cannot be produced by the chromatophore (Fig. 2).

Some other biosynthetic pathways are truncated in the chromatophore because single gene coding enzymes were lost. For example, in the biosynthetic pathway of leucine, most gene coding enzymes are present in the chromatophore except for the gene coding for 3-isopropylmalate dehydrogenase. In this case, we assumed that host encoded enzymes complement the pathway in the endosymbiont. Either by importing host encoded enzymes to the chromatophore or by exchanging intermediated metabolites between the symbionts. Similar situations have been proposed for other host-endosymbiont systems [12]. For this reason, we assumed that the production of these metabolites is shared between the host and the endosymbiont (see model *i*CV265 in Additional file 2).

Table 1 Characteristics of metabolic models of *Synechococcus sp.* WH 5701 (*i*CV498) and the chromatophore (*i*CV265)

	Metabolic model	
	*i*CV498	*i*CV265
Genes	498	265
Metabolites	698	615
Intracellular metabolites	661	578
Extracellular metabolites	37	37
Reactions	743	627
Enzymatic reactions	624	478
Transport reactions	82	70
Exchange reactions	37	37

In addition, some reactions in the chromatophore model *i*CV265 for which no orthologous genes exist with the free-living model *i*CV498 but are essential for in silico growth were assumed to be present (see model *i*CV265 in Additional file 2).

Finally, chromatophores lost the ability to store photosynthates as well as the capacity to synthesize sucrose [5]. Because of that, glycogen was removed from the biomass equation in *i*CV265. Under these conditions, in silico growth of the *i*CV265 model was 0.1568 h^{-1}. This is an unrealistic rate because growth of the chromatophores is restricted to host division which is much lower than growth rate reported for

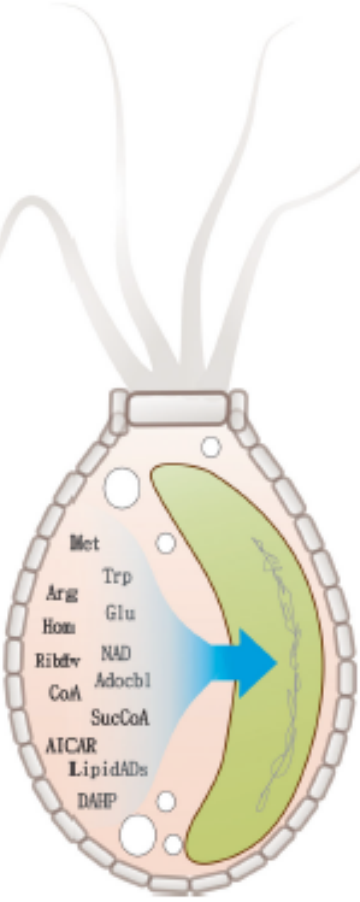

Fig. 2 Nutrients uptake simulation in the chromatophore model (*i*CV265). Metabolites that cannot be produced by the chromatophore (with respect to the free-living model, *i*CV498) include: amino acids (Met = L-Methionine, Trp = L-Tryptophan, Arg = L-Arginine, Glu = L-Glutamate, Hom = L-Homoserine), cofactors (NAD = Nicotinamide adenine dinucleotide, Adocbl = Adenosylcobalamin, CoA = Coenzyme A), vitamins (Ribflv = Riboflavin) and others (AICAR = 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide, SucCoA = Succinyl-CoA, LipidADs = Lipid A Disaccharide, DAHP = 2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate)

free-living cyanobacteria and even other photosynthetic eukaryotes [27].

Robustness analysis of metabolic models

We assessed the robustness of the *i*CV498 and the *i*CV265 models to single gene deletions. Genetic robustness was defined as the capacity of the models to maintain its metabolic capabilities (in silico biomass production) after a genetic deletion. Under phototrophic conditions, model *i*CV498 showed 333 genes (66.86%) to be essential because its deletion decreases the biomass production over 99% (Fig. 3). This result shows that *i*CV498 is less robust than the metabolic model of *Synechocystis sp.* PCC 6803 where 51.6% of the genes are essential under these same conditions [26]. In addition, there is a decreasing robustness in the model of the chromatophore where 222 of the 265 genes (83.77%) are essential (Fig. 3). This indicates that the genomic reduction experimented by the chromatophore rendered its metabolic network fragile. The same result has been observed for other metabolic networks of endosymbionts [20, 28, 29].

Interestingly, we found that there are 3 non-essential genes in the metabolic model *i*CV498 whose single deletion decreases in silico growth rate. These include genes encoding the enzymes acetyl-CoA synthetase, malic enzyme (NAD) and fumarase. Of these three, the last enzyme is the only one decreasing the in silico growth rate in *i*JN678 when it is deleted (data not shown). In the *i*CV265 model, all these 3 genes were lost. In addition, the non-essential gene in *i*CV498 coding for an enzyme with arginase activity is the only one whose deletion decreases the growth rate in the *i*CV265 chromatophore model. This suggests that genome reduction leading to *i*CV265 caused metabolic restructuring because deletion of this enzyme with arginase activity in *i*CV498 has no effect.

In silico simulation of metabolic-gene loses in the chromatophore of *P. chromatophora*

Based on the metabolic network of the free living *Synechococcus sp.* WH 5701, we simulated in silico the gene loss. We evaluated the impact of intracellular conditions (metabolite availability) on the evolution of the chromatophore. In particular, we asked whether the set of metabolites predicted to be provided by the host in the *i*CV265 model (Fig. 2) determined actual gen content of the chromatophore after genome reduction.

Two in silico intracellular conditions were evaluated. In the first one, we simulated genetic reduction under in silico intracellular conditions where available nutrients were those predicted in the *i*CV265 model (we refer to them as Proposed Nutrients) (Fig. 2). In the second one, we randomly selected metabolites from the *i*CV498 model (the same number as in the first condition) and assigned as available nutrients under intracellular conditions (we refer to them as Randomized Nutrients; see Additional file 3: Table S1). The algorithm to simulate genome reduction is explained in detail in the methods section.

This algorithm allowed us to obtain in silico evolved chromatophores whose metabolic capabilities regarding the biomass production are equivalent to those of *i*CV265; but differing in their in silico evolutionary history and gene content.

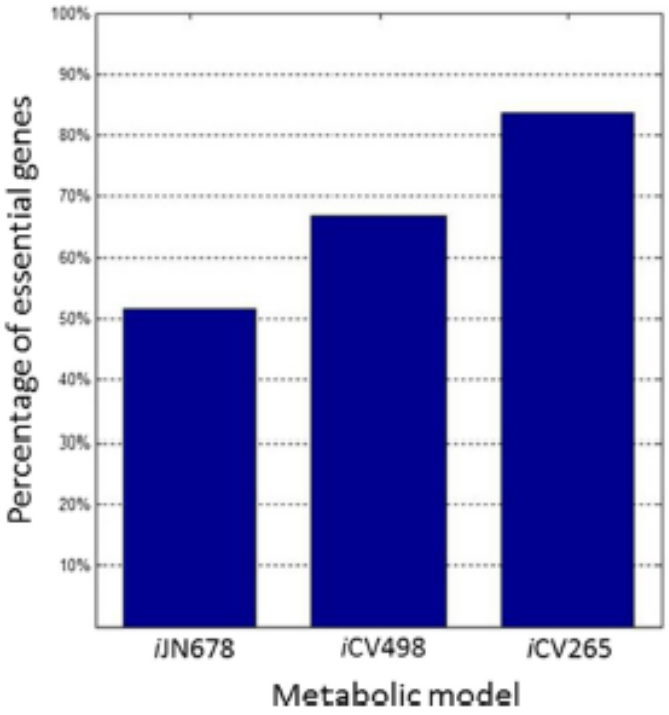

Fig. 3 Genetic robustness analysis of metabolic models of *Synechocystis* sp. PCC 6803 (*i*JN678) [26], *Synechococcus* sp. WH 5701 (*i*CV498) and the chromatophore of *Paulinella chromatophora* (*i*CV265). Percentage of essential genes (Y axis) for each metabolic model (X axis)

Simulations under the Proposed Nutrients conditions resulted in reduced metabolic networks with 295.1 (± 2.63) genes on average. In these reduced networks, of the 498 genes present in the free-living ancestor (model *i*CV498), 52.2% are strictly conserved in the 500 simulations and 26.7% are dispensable in all of them. In Randomized Nutrients simulations, reduced networks have an average size of 326.8 (± 5.26) genes and 54% and 16.2% are conserved and dispensable in the 500 simulations, respectively.

As is shown in Fig. 4, the proportion of: i) essential genes (predicted to be essential in 500 simulations); ii) variable genes (predicted to be conserved in 1 to 499 simulations); and iii) dispensable genes (predicted to be lost in 500 simulations), varies between metabolic pathways. These proportions also vary between treatments (i.e., Proposed Nutrients or Randomized Nutrients). Surprisingly, the most extreme case is that of the genes participating in photosynthetic activity. In Proposed Nutrients 77.6%, 18.3% and 4.1% are predicted to be essential, variable and dispensable, respectively. While for Randomized Nutrients none was predicted to be essential nor dispensable because 100% of them were variable.

Genetic concordance was evaluated between these simulated minimal networks and the *real* chromatophore model (*i*CV265). This was done by measuring sensitivity and specificity as in [30]. In Fig. 5, we show the fraction of true-positives and false-positives for every cutoff (1 to 500). True-positive and false-positive for every cutoff (1 to 500) form a curve whose area under the curve represents the probability that a gene conserved in *i*CV265 is present in more simulations than a gene which has been lost.

The area under the curve shows the contribution of the nutrients available in intracellular conditions explaining the evolutionary history experimented by the chromatophore. Accordingly, the accuracy obtained under the Proposed Nutrients condition was 77.4%, while that of the Randomized Nutrients was 59.8%. The difference between the areas under the curve from both conditions is statistically significant (p-value < 0.001, Chi-square test of homogeneity).

Modeling selection and drift to explain metabolic evolution of the chromatophore

We are interested in understanding the role played by natural selection during the evolution of the metabolic capabilities of the chromatophore. Chromatophores provide the host with reduced-carbon, probably a hexose. This in analogy to the origin of plastids. It has been proposed that during the early stages of plastid evolution,

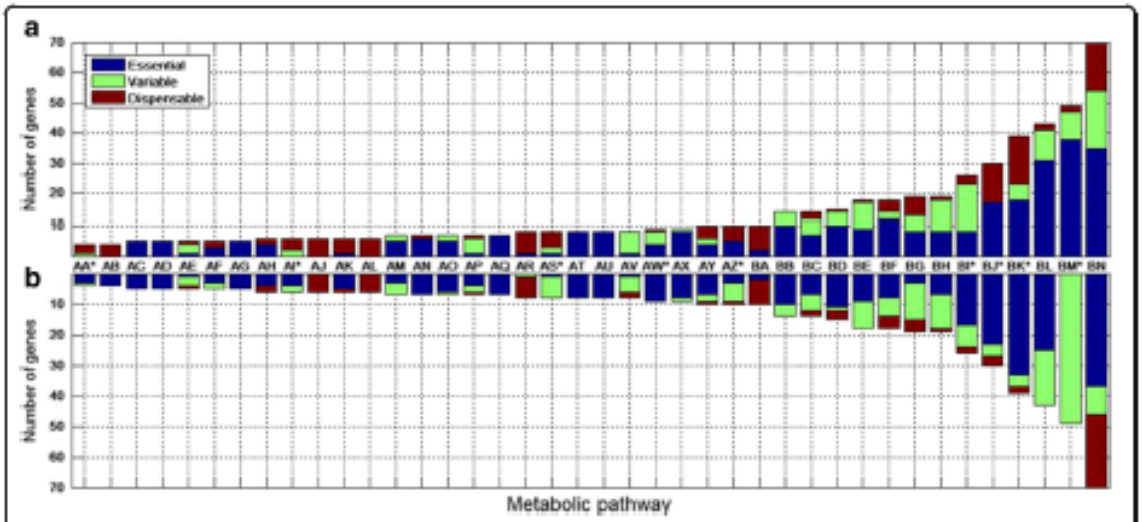

Fig. 4 Variation in the proportion of genes classified as essential, variable and dispensable in different metabolic pathways according to two different sets of available nutrients for the chromatophore. **a** Set of Proposed Nutrients; **b** set of Randomized Nutrients. AA to BN metabolic pathways: AA* = Citrate cycle (TCA cycle); AB = Lipopolysaccharide biosynthesis; AC = Carotenoid Biosynthesis; AD = Folate biosynthesis; AE = Glycerolipid metabolism; AF = Hydrogen production; AG = Steroid biosynthesis; AH = Aminosugars metabolism; AI* = Nicotinate and nicotinamide metabolism; AJ = Nucleotide sugars metabolism; AK = Riboflavin metabolism; AL = Thiamine metabolism; AM = Carbon fixation; AN = Glutamate metabolism; AO = Lysine metabolism; AP = Nitrogen metabolism; AQ = Terpenoid backbone biosynthesis; AR = Fructose and mannose metabolism; AS* = Pantothenate and CoA biosynthesis; AT = Peptidoglycan biosynthesis; AU = Ubiquinone and other pterpenoids biosynthesis; AV = Urea cycle and metabolism of amino groups; AW* = Alanine, aspartate and glutamate metabolism; AX = Valine leucine and isoleucine biosynthesis; AY = Histidine metabolism; AZ* = Pentose phosphate pathway; BA = Starch and sucrose metabolism; BB = Fatty acid biosynthesis; BC = Glyoxylate and dicarboxylate metabolism; BD = Sulfur Cysteine and methionine metabolism; BE = Arginine and proline metabolism; BF = Pyrimidine metabolism; BG = Glycolysis/Gluconeogenesis; BH = Pyruvate metabolism; BI* = Phenylalanine tyrosine and tryptophan biosynthesis; BJ* = Purine metabolism; BK* = Porphyrin and chlorophyll metabolism; BL = Oxidative phosphorylation; BM* = Photosynthesis and BN = Transport. * Metabolic pathways with a difference in categorical genes with *p-value* < 0.05

the photosynthetic endosymbiont exported reduced-carbon in the form of an hexose-phosphate through an hexose phosphate transporter of bacterial origin (non-cyanobacterial) [31, 32].

To study how the potential rate of carbon exportation evolved, a hexose export reaction was added to the metabolic models. This reaction was defined as objective function. To ensure biomass components production, the biomass reaction was fixed to 0.0884 h^{-1} which, as stated previously, is the growth rate of *Synechocystis sp.* PCC 6803 metabolic model under autotrophic growth conditions [26].

Under these conditions, there is no exportation of reduced-carbon in the *i*CV498 model. However, in the *i*CV265 chromatophore model, the potential rate of hexose exported without affecting the in silico growth rate was 0.2689 mmol × gDW^{-1} × h^{-1}. In Fig. 6, we show the fluxes calculated with FBA of the central metabolism of the models of the chromatophore and the free-living cyanobacteria in conditions previously mentioned.

Fluxes calculated for production of metabolites precursors used to produce biomass components are produced in less quantity in the chromatophore's model (Fig. 6). This is a consequence of the loss of metabolic capabilities in the metabolism of the chromatophore which allow the redirection of carbon through the gluconeogenic pathway for the production of hexoses as metabolic objective, instead of being used in the production of biomass components.

To analyze the efficiency of the metabolic networks in terms of hexose production at overall metabolism, we calculated the yields. The yields are parameters that measure the efficiency of the metabolic network and allow the comparison across different microorganisms. For instance, the yield of the ethanol production is higher in *Saccharomyces cerevisiae* compared to *Zymomonas mobilis*, this was the result of the specialization of the microorganism to produce specific metabolites [33]. Therefore, we calculated the yields of carbon, energy and reducing equivalents (extracellular CO_2, ATP and NADPH) required to produce hexose (Table 2). These results show that the model of the chromatophore is more efficient for producing hexose from the external carbon than the free-living cyanobacteria. It means that metabolic restructurings experienced by the chromatophore

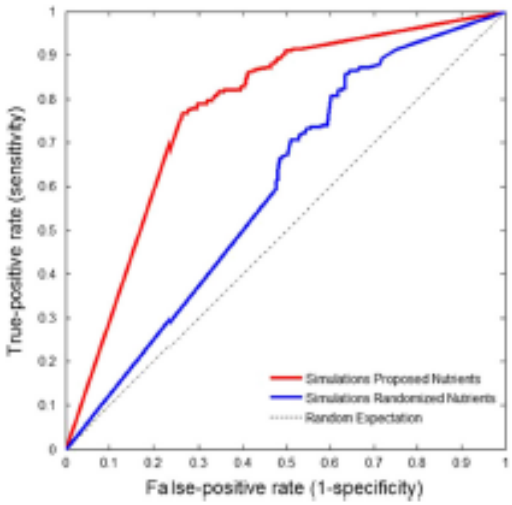

Fig. 5 Genetic concordance between in silico evolved chromatophores and the chromatophore model (*i*CV265). The area under the curve indicates the predictive accuracy for evolved chromatophores (i.e. simulations) obtained under Proposed and Randomized Nutrients conditions. Proposed Nutrients: simulation with metabolites proposed as nutrients in the *i*CV265 model (Fig. 2) and Randomized Nutrients: simulations with randomized metabolites assigned as nutrients (Additional file 3: Table S1). Area under the curve: Proposed Nutrients = 0.7742 (*p*-value = 2.853E-26); Randomized Nutrients = 0.5987 (*p*-value = 1.42E-4)

rendered its metabolism more efficient to produce hexose which can be provided to the host.

The yields suggest that the loss of some metabolic capabilities in the ancestral cyanobacterium caused a redirection of fixed CO_2 causing changes in metabolic fluxes and consequently increasing the rate of reduced-carbon exported to the extracellular compartment.

We then inquired about the evolutionary forces that determined genetic conservation and metabolic functionality in chromatophores. Specifically, we wanted to infer if these metabolic capabilities of the *i*CV265 model (the potential rate of hexose exportation of 0.2689 mmol × $gDW^{-1} \times h^{-1}$) could have been possible under a random model of evolution or were the consequence of natural selection for metabolic specialization of the chromatophore and its positive impact at the holobiont. To tests this, we simulated the metabolic reduction process with hexose export and biomass production as evolutionary restrictions in a random model where hexose exported must be greater than zero. It means that every gene affecting the in silico growth rate of 0.0884 h^{-1} and impairing hexose export was considered as essential, while the hexose export rate could always vary while being greater than zero (purifying selection restriction for hexose export).

Minimal networks obtained in silico under these conditions were variable in size and gene content. Of 500 simulations, only 175 (66.03%) of the genes conserved in model *i*CV265 are conserved in all the simulations. Conversely, there are 45 genes predicted to be essential in all these 500 simulations which are not conserved in model *i*CV265. The metabolic networks from these 500 simulations are different in gene content and show different hexose export rates however they are equivalent in biomass production (Additional file 3: Figure S1).

As shown in Fig. 7, after in silico metabolic reduction, hexose export rate in minimal networks obtained under these conditions tend to be minimal and close to zero (hexose export rate could not reach zero because of the restriction we imposed). On the other hand, only 2.6% of simulations have a potential rate of hexose exported equal or higher than the metabolic model of the chromatophore (0.2689 mmol × $gDW^{-1} \times h^{-1}$). This suggests that the probability of obtaining a potential rate of hexose exported similar to that of *i*CV265 under a random model is less than 5%. We got a similar result by varying the growth rate constraint of 0.0884 h^{-1} under plausible biological values (see Additional file 3: Figure S2).

Although variable, our simulations evolved metabolic networks that have approximately the same number of reactions than *i*CV265. The average number of reactions with non-zero fluxes in the reduced metabolic models of the 500 simulations is 416.15 ± 3.91. This is slightly less than the number of reactions with non-zero fluxes in the *i*CV265 model (442 reactions). This shows that the small percentage of simulations (2.6%) showing a potential rate of hexose exported equal or greater than that of the chromatophore (0.2689 mmol × $gDW^{-1} \times h^{-1}$) is not due simply to smaller size of the simulated metabolic networks.

These in silico experiments suggest that the potential rate of hexose exported in model *i*CV265 is unlikely to be the outcome of only genetic drift and purifying selection (i.e., less than 5% of the simulated networks export hexose at a rate comparable to that of *i*CV265). This suggests that the potential rate of hexose exported was the result of a process of functional specialization in which the increasing rate of hexose exportation was favored by natural selection due to the positive impact at the holobiont level.

Interestingly all these 2.6% of in silico evolved chromatophores have in common the conservation of a phosphate transporter via ABC system which is also present in the chromatophore model (*i*CV265). Conservation of this phosphate transporter allows the simulated network to get the phosphate necessary to be able to export fixed carbon. Without this transporter most of fixed carbon is oxidized in the pentose phosphate pathway releasing only a small amount to the extracellular compartment (data not shown).

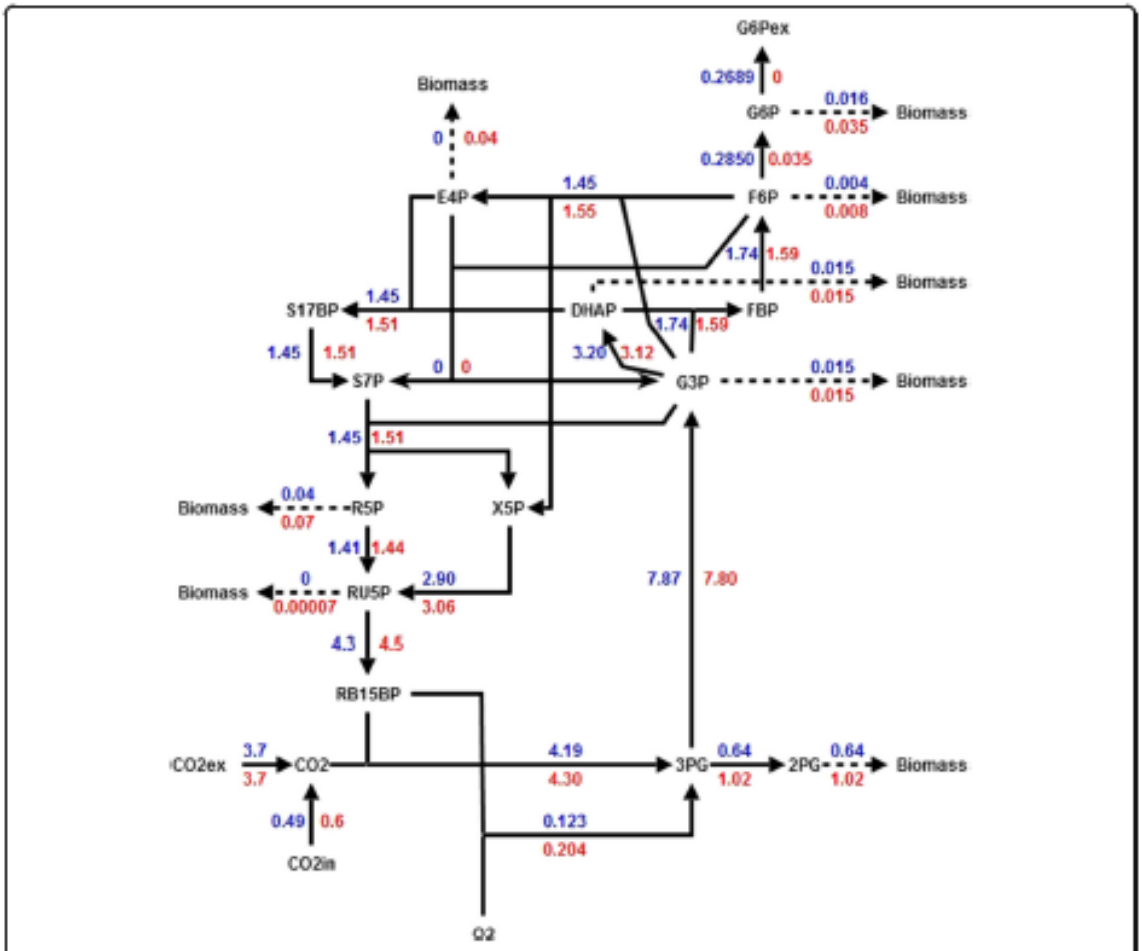

Fig. 6 Flux distribution obtained with FBA of the central metabolism of *i*CV265 (*blue*) and *i*CV498 (*red*)

Metabolic integration of the chromatophore to its host

In our previous simulations, we assumed that nutrients (Fig. 2) were available simultaneously for the chromatophore since the beginning of the evolutionary process at the onset of the endosymbiosis. However, it is likely that this has not been the case and transporters for these nutrients were gained (or lost) sequentially. For instance, metabolic transport activity in the chromatophore is reduced due to loss of most transporters in comparison to free-living cyanobacteria [5]. And it has been reported that a large percentage of solute transporters in plastids from Plantae have host and bacterial (non-cyanobacterial) origin [31, 34].

Therefore, we simulated the evolutionary acquisition of transporters and its consequences in gene loss and the capability of the chromatophore to export fixed carbon to its host.

For every simulation, we used *i*CV498 as a free-living ancestor of the chromatophore under nutrient-rich conditions (Fig. 2). However, in this experiment, the model *i*CV498 did not have access to all nutrients since the beginning of the simulation. Instead, we randomly assigned a transport allowing the uptake of the respective nutrient. We then randomly deleted one gene at a time from *i*CV498. If the deleted gene affected the growth rate (0.0884 h^{-1}) or impaired hexose

Table 2 Yields ($Y_{P/S}$) analysis of extracellular CO_2, ATP and NADPH consumed in hexose production for both models

	Metabolic model	
	*i*CV265	*i*CV498
	mmol Hexose/ mmol	mmol Hexose/ mmol
CO_2	0.069	0.009
ATP	0.023	0.002
NADPH	0.036	0.004

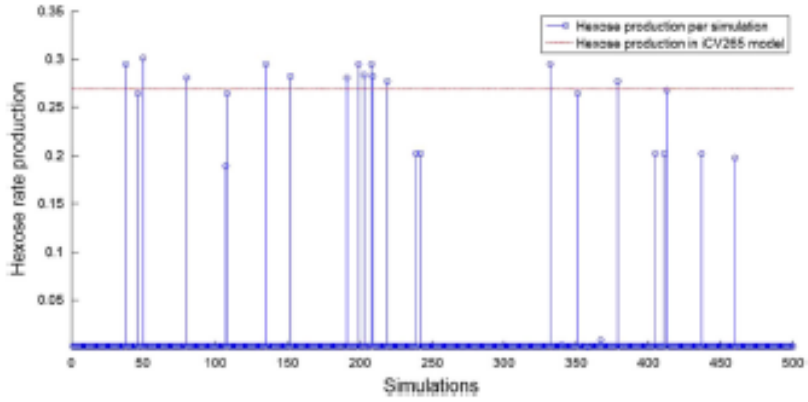

Fig. 7 Hexose exportation rate of the chromatophore model (*i*CV265) is achieved in only 2.6% of the simulations under a random model of evolution. Hexose exported rate (Y axis) for 500 independent simulations (X axis). Red-dotted line indicates the hexose export rate in the chromatophore model (*i*CV265)

exportation, we considered this gene as essential and we restored it to the model. In this way we analyzed the selective impact caused by gene loss due to the addition of a single transporter and the concomitant relaxation of natural selection for retention of specific biosynthetic pathways. Once we analyzed every gene in the model, we randomly assigned a second transport and then we repeated the gene loss simulation mentioned above. Simulation stops when in silico chromatophore has access to the 13 nutrients (Fig. 2) and all genes have been evaluated for their essentiality.

As shown in Fig. 8, after the incorporation of the 13 transporters, the probability of getting a potential rate of hexose exported equal or higher than the metabolic model of the chromatophore (0.2689 mmol × gDW^{-1} × h^{-1}) is less than 5% in 500 simulations, in agreement with our previous result (Fig. 8).

During the process of metabolic integration, it is noted that the maximum rate of hexose exportation becomes greater with every metabolite obtained as nutrient. However, by inspecting the frequency distribution of simulations with different potential rates of hexose exported, it is obvious that as metabolic integration advances, the probability of getting the maximum rate of hexose exportation decreases (i.e. the frequency of networks with large export rate becomes smaller).

These changes in frequency distributions during the process of metabolic integration can be interpreted in terms of the functional specialization of the chromatophore. As metabolic integration proceeds (with the addition of more transporters), the chromatophore increased its capacity to provide fixed carbon to its host. However, continued gene loss led to a simplified metabolic network and a smaller fraction of in silico evolved chromatophores can export as much fixed carbon as *i*CV265. The evolutionary landscape becomes smaller as evolution proceeds.

The metabolism of the chromatophore is specially adapted to produce carbon for its host

As shown above, the potential rate of hexose exported in the chromatophore model (*i*CV265) is highly dependent on phosphate consumption. In addition, the growth rate of the chromatophore is coupled to the host's growth rate. As shown above, the potential rate of hexose exported is unlikely to be the outcome of a random evolution.

To test the impact of these two restrictions, we analyzed the metabolic properties of models *i*CV265 and *i*CV498 in potential rate of hexose exported under growth rate and phosphate uptake restrictions. As shown in Fig. 9, the potential rate of hexose exported in the *i*CV498 model is robust with respect to growth rate and phosphate uptake (i.e. a given growth rate can sustain the hexose rate exportation with different rates of phosphate consumption). This contrasts with the chromatophore model (*i*CV265) where, for a given growth rate, only a specific consumption of phosphate is necessary to sustain hexose release. In addition, the capacity of hexose export in the *i*CV265 model for a determined growth rate restriction is greater than the *i*CV498 model.

Discussion

In this work, we show that the metabolic network of the chromatophore of *P. chromatophora* is different to the metabolic network of its free-living relative *Synechococcus sp.* WH 5701. We suggest that these differences evolved by natural selection. Gene-loss and carbon flux redirection guided by natural selection led to metabolic

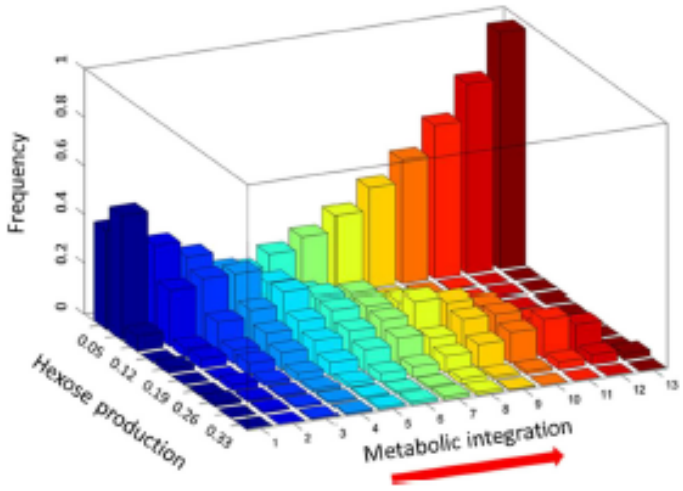

Fig. 8 Metabolic integration with the host determines the rate of hexose exported by in silico chromatophores as well as the frequency of simulations that provide higher rates of reduced-carbon to the host. X axis, hexose export rate; Y axis, frequency of simulations; Z axis, transporters added to the model (metabolic integration)

specialization of the chromatophore as a reduced-carbon provider.

Purifying selection and the maintenance of the symbiosis

Our analysis showed that some metabolic pathways have been preferentially conserved in the chromatophore (Fig. 1). These preserved metabolic pathways (i.e. photosynthesis, carbon fixation, and gluconeogenesis) very likely play a prominent role in the symbiosis. This pattern is analogous to the one observed in many other endosymbionts e.g. *Buchnera aphidicola* [35]. In this later case, biosynthetic pathways producing essential amino acids for the host [12, 35] are preserved by host-level natural selection.

As mentioned above, differential conservation of gene category functions suggests that purifying selection is preserving relevant symbiotic functions. Accordingly, estimation of the rate of nucleotide substitution in 681 DNA alignments of protein-coding genes orthologous between chromatophores of two different strains of *P. chromatophora* (CCAC 0185 [5] and FK 01 [7]) showed that most of them have signals of purifying selection [7].

It has been suggested that host-level selection prevents the fixation of deleterious mutations in endosymbionts thus lowering the chances of a mutational meltdown resulting in extinction [36, 37]. And, of course, this prevents the consequent replacement of non-functional endosymbionts [38]. In addition, selective pressure to maintain functional proteins increases with the time of host-endosymbiont interaction [36] and combined with very strong bottlenecks may help to reduce the accumulation of deleterious mutations. This has been proposed to explain mitochondrial genome evolution [39].

Metabolic integration of the chromatophore to its host

Comparison of the metabolic models of the chromatophore and the cyanobacterium *Synechococcus sp.* WH 5701 allowed us to inquire into the evolution of the metabolic interaction of the chromatophores with its host. For example, several metabolic pathways in the chromatophore are incomplete. It is likely that the host supplies these metabolites as nutrients to the chromatophore. Metabolic pathway sharing is a hallmark of endosymbiotic organisms. For example *Wolbachia*, which are endosymbionts of many animal species, show a degraded genome [40, 41] whose limited metabolic capabilities are complemented by its host [42]. In turn, the endosymbiont provides the host with nutrients such as riboflavin, positively impacting host fitness [42]. Equally remarkable is the likely coupled production of some metabolites between the chromatophore and its host. As mentioned above, this collaboration in metabolite biosynthesis has been observed in other symbiotic systems [43–46].

Fragility of a reduced metabolic network

To study the metabolic capabilities of the chromatophores we used FBA. This stoichiometric approach can predict cellular phenotypes in specific environmental conditions. Generally, biomass production is fixed as objective function. In absence of biomass composition, the use of a biomass equation from a related organism is a valid starting point for metabolic analysis [47–49]. In

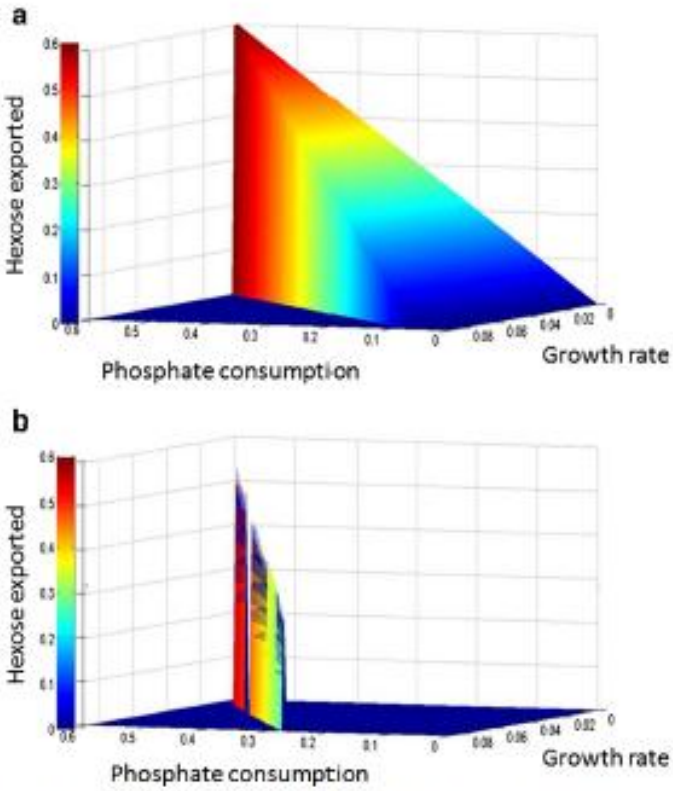

Fig. 9 The metabolic model of the chromatophore (*i*CV265) is specially adapted to produce higher rates of reduced-carbon under phosphate and growth restrictions in comparison with *Synechococcus sp.* WH 5701 (*i*CV498). **a** Hexose export rate in the free-living model *i*CV498. **b** Hexose export rate in the chromatophore model *i*CV265. Y axis, growth rate restriction. X axis, phosphate uptake restriction. Z axis, potential rate of hexose exported. Colors in axis Z are only for drawing purposes and do not represent an extra variable

this way, FBA has been used to infer the metabolic capabilities of different organisms whose cultivation and experimental management is challenging or not yet possible, as in the case of endosymbionts. For example, biomass composition and the metabolic model of *Escherichia coli* were used for metabolic analysis of *Buchnera aphidicola* [20, 30], *Sodalis glossinidius* [29], and *Blattabacterium cuenoti* [28]. In the same way, we used the biomass composition and stoichiometric model of *Synechocystis sp.* PCC 6803 as a starting point to model the metabolism of the chromatophore and *Synechococcus sp.* WH 5701 [26].

We found that the metabolism of the chromatophore is highly fragile to gene deletions. Approximately 84% of the genes in the model are essential when singly deleted in comparison with ~67% of the genes in *Synechococcus sp.* WH 5701. A similar difference in metabolic fragility was found by [20] when comparing the models of *B. aphidicola* and its free-living relative *Escherichia coli* where 84% and 19% of genes were essential, respectively. In the same way, the metabolic network of two strains of *Blattabacterium cuenoti* (Bge and Pam), the obligated primary endosymbiont of cockroaches, were shown to be highly fragile to single gene deletion. It was found that 76.1% and 79.6% were essential genes, respectively [28]. Finally, in *Sodalis glossinidius* (the secondary non-obligated endosymbiont in early stages of tsetse flies), 44.54% metabolic genes were found to be essential, compared with its ancestral network where only 25.48% are predicted to be essential [29].

Our robustness analyses of the *i*CV265 and the *i*CV498 models agree with the generalization that metabolic networks of endosymbionts are more fragile than their free-living counterparts. This metabolic fragility of endosymbionts contrasts with theoretical estimations that suggest that, in general, metabolic systems are robust and complex [50]. However, the metabolic systems of endosymbionts are considered more robust [28] than minimalist metabolic networks [51]. The difference in metabolic fragility of the chromatophore

when compared to *Synechococcus sp.* WH 5701 reflects the transition from a free-living style to a more stable condition inside *Paulinella chromatophora*.

Metabolic environment as a determinant of gene content

It has been shown that retention of metabolic genes in endosymbionts is determined by the metabolic requirements and molecular environment of the host [52, 53]. With the use of FBA and the metabolic model of *Synechococcus sp.* WH 5701 as a proxy of the ancestor of the chromatophore, we evaluated the impact of the host-metabolic environment in the reduction of the metabolic system of the endosymbiont. The proposed host-metabolic environment (Proposed Nutrients) predicted with 77.42% of accuracy the actual gene content of the chromatophore. This is in contrast with the 59.8% of accuracy obtained when using a randomly set of host-provided metabolites (Randomized Nutrients). This emphasizes the contribution of the intracellular metabolic environment to the evolution of the metabolism in the chromatophore.

Similar reductive simulations have been used to predict the set of essential genes of pathogens located in certain environmental niches (like the bloodstream) within the human body [52]. In the same way, reductive evolution simulations using *E. coli* as free-living ancestor predicts with 80% of accuracy the metabolic gene content of *B. aphidicola* and *Wigglesworthia glossindia* [30].

Inspection of the proportion of dispensable, variable, and essential genes by in silico reductive simulations (i.e. Proposed Nutrients and Randomized Nutrients) predicts differential gene retention patterns between different metabolic pathways. For example, in Randomized Nutrients simulations, photosynthesis pathway (which is the *raison d'être* of the symbiosis) 100% of genes are predicted as "variable" (none of the genes are predicted to be retained in the 500 simulations) while in Proposed Nutrients ~78% are essential. This means that under Randomized Nutrients, photosynthesis function could be useful but not essential and could have been lost in the chromatophore by chance. Clearly, the set of metabolites comprising Randomized Nutrients cannot account for the metabolic gene content of extant chromatophores.

Maximization of biomass production is regularly used as objective function in FBA analysis. It allows predicting the distribution of fluxes through a metabolic network [54]. The maximization of biomass function is used as a proxy of evolutionary fitness. However, many other objective functions can be used [54, 55]. For instance, it was estimated that *Chlorella* (the photosynthetic endosymbiont of *Paramecium bursaria*), releases 57% of its photosynthates to its host [56]. This means that most carbon photosynthetically assimilated is destined to symbiotic interaction instead of biomass production of *Chlorella* itself. In the same way, *P. chromatophora* has phototrophic nutrition. It depends on carbon assimilates which derivate from the endosymbiotic cyanobacterium whose inorganic carbon rate assimilation is the same as a free-living cyanobacteria [10]. But unlike its free-living relatives, its growth rate is restricted by *P. chromatophora*. Considering the above metabolic analysis of the chromatophore, which predict an in silico growth rate of 0.1568 h^{-1}, it is difficult to consider the biomass as the only objective function in chromatophores. Taking into consideration that chromatophores provide the host with reduced-carbon, a reaction simulating hexose export to extracellular compartment was added. This reaction was defined as objective function. And to ensure biomass components production, biomass reaction was fixed to 0.0884 h^{-1} which is the growth rate of a free-living relative cyanobacterium. Interestingly, under these conditions the metabolic model of the chromatophore predicts a potential rate of hexose exportation of 0.2689 mmol × gDW^{-1} × h^{-1}. As far as we know, this is the first metabolic reductive evolutionary analysis where metabolic functionality (i.e. hexose export) of the endosymbiont is explored as objective function, differing from previous analyses where biomass is set as objective function of mutual endosymbionts as *B. aphidicola* [20, 30], *S. glossinidius* [29] and *B. cuenoti* [28].

ABC phosphate transporter is an essential component of the chromatophore

All simulations showing a hexose exportation rate equivalent to that of the chromatophore model (*i*CV265) share the ABC phosphate transporter. This P_i-dependency in the chromatophore agrees with that observed in isolated spinach chloroplasts [57]. It has been shown that photosynthesis declines dramatically (less than 10% of the maximum rate) in chloroplast in the absence of P_i in the reaction medium. Also, carbon export from the chloroplast is inhibited [58], with up to 60% of ^{14}C fixed being retained in the chloroplast [57]. As mentioned above, this observation agrees with the more than 95% of simulations which predict that lack of ABC phosphate transport favors carbon retention in the chromatophore instead of being released to the host. Therefore, we predict that lack of ABC transporter in the genome of the chromatophore of *Paulinella* FK01 is compensated by a phosphate transporter coded in the host [7].

The role of natural selection on the evolution of the metabolism of the chromatophore

Inspection of FBA calculated central metabolic fluxes in the chromatophore and in the free-living cyanobacteria showed that the endosymbiont is better at producing hexose. This is likely a host related adaptation. To

investigate whether this and other characteristics of the metabolic model of the chromatophore evolved by natural selection, we simulated in silico reductive evolution with a null model not including positive selection. As a proxy of genome reduction by purifying selection and random genetic drift, we submitted the metabolic model of *Synechococcus sp.* WH 5701 with the following algorithm: a) first, we simulated host-level purifying selection by requiring that the rate of hexose exportation of the model must be always greater than 0 and biomass is produced at 0.0884 mmol × gDW^{-1} × h^{-1}; b) next, we performed rounds of single gene deletion until no more genes could be deleted; c) finally, we repeated this process 500 times. By this, we obtained a population of 500 reduced metabolic networks all of them capable of producing 0.0884 mmol × gDW^{-1} × h^{-1} of biomass, but differing in hexose rate exportation. Differences in rates of hexose exportation were due to contingency-dependent loss of alternative pathways [30]. With this experiment, we could determine if the potential rate of hexose exported in *i*CV265 (0.2689 mmol × gDW^{-1} × h^{-1}) is easily obtained by host-level purifying selection (hexose exportation >0) and contingency-dependent evolution on random gene deletion. Our evolutionary reductive analyses showed that <5% of simulations were predicted to export hexose at a similar rate as the model *i*CV265. This suggests that metabolic functionality of *i*CV265 is unlikely to be determined by genetic drift alone. Therefore, we conclude that natural selection at holobiont level may have contributed to shape metabolic functionality of the chromatophore.

Natural selection as metabolic engineer

According to the above mentioned, we consider suitable to make the analogy of natural selection as metabolic engineer. Metabolic engineering can be defined as "the directed improvement of product formation or cellular properties through the modification of specific biochemical reactions or introduction of new ones" [22]. One of the objectives of metabolic engineers is to redirect the flux of mass through the metabolism of organisms towards a desired metabolic product. Some genetic strategies to redirect metabolic flux toward production of a desired metabolite include: increasing the precursor supply; altering the regulation (overexpressing) genes; increasing the efficiency of bottleneck enzymes; reducing flux toward unwanted byproducts; or eliminating competing pathways by gene-deletion [59]. It has been proposed that cellular metabolism of free-living microorganisms is primed, through natural selection, for the maximum responsiveness to the history of selective pressures rather than for the overproduction of specific chemical compound [60]. In host-restricted conditions this responsiveness to free-living selective pressures are no longer needed. Instead, new biological objectives are defined now related to holobiont survival.

For instance, it was proposed that the chloroplast metabolic network has improved photosynthetic properties in comparison to free-living cyanobacteria [21]. For example, the metabolic network in chloroplast has: i) a longer average path length; ii) a larger diameter; iii) is Calvin Cycle-centered; iv) and presents better modular organization when compared with the network of free-living cyanobacteria [21]. In a similar way, the metabolism of the chromatophore (*i*CV265) seems to be tailored for the exportation of reduced-carbon; that is, when comparing the export of reduced-carbon between the *i*CV265 and the *i*CV498 models (with phosphate as restrictive nutrient) we found that *i*CV265 shows higher rates of hexose exported than the free-living *i*CV498 model at the cost of increased consumption of phosphate (Fig. 9).

The evolutionary mechanism outlined above applies when the host benefits from the endosymbiont. In particular, mechanisms such as "partner fidelity feedback" (PFF) promote cooperation between symbionts. PFF requires individuals to be "associated for an extended series of exchanges that last long enough that a feedback operates" [18]. Similar mechanisms likely operated in other symbiotic systems. For example, *Buchnera* [61] and *Blochmannia* [62] overproduce essential amino acids (EAAs) to its host. This overproduction of EAAs was consequence of metabolic restructurings due to metabolic-gene losses. For example, the truncation of the purine biosynthesis pathway which allows the endosymbiont to produce histidine at higher rates than free-living relatives [20]. Reductive evolutionary simulations carried out by [20] showed that this truncation is an improbable evolutionary event under conditions tested.

Conclusion

Our main objective was to better understand the metabolic changes experienced by the free-living cyanobacteria to become a chromatophore. In addition, we assessed the evolutionary forces driving organellogenesis. We found evidence that certain metabolic pathways are preferentially conserved in the chromatophore. We also found that the pattern of metabolic gene loss strongly depends on the availability of nutrients from its host. The high fragility of the chromatophore network reflects the transition to a more stable environment and, consequently, its simplification. The chromatophore is specialized in producing reduced-carbon which could be released to the host. This specialization was consequence of metabolic restructurings which could not be possible in free-living conditions. We interpret this specialization as consequence of natural selection acting as a metabolic engineer which modifies intrinsic

metabolic properties of the endosymbiont impacting positively at the holobiont level. Our in silico simulations allowed us to determine that metabolic specialization of the chromatophore is an unlikely result of purifying host-level selection and genetic drift alone. In this way, computational analysis of biological systems allows to obtain new insights on the evolutionary forces shaping metabolic evolution of mutualistic endosymbionts.

Methods

Differential gene retention of functional categories in the chromatophore genome

To identify metabolic pathways preferentially conserved in the chromatophore we carried out a statistical analysis using the program GeneMerge [63]. First, we classified each of the genes in both genomes (the chromatophore of *Paulinella chromatophora* CCAC 0185 [5] and *Synechococcus sp.* WH 5701) according to the functional categories of KEGG orthology (http://www.genome.jp/kegg/ko.html). Then we carried out the statistical analysis with GeneMerge. GeneMerge is a program written in Perl which allows the identification of overrepresented functions or categories in a sample by using a hypergeometric distribution [63].

Metabolic reconstruction of the *i*CV498 and the *i*CV265 models

A draft metabolic model was initially reconstructed by identifying orthologous genes between *Synechococcus sp.* WH 5701 and the metabolic model of *Synechocystis sp.* PCC 6803 (*i*JN678) [26]. Because this draft metabolic network had many inconsistencies we performed a manual refinement. This consisted in reviewing literature and databases to fill gaps in the model. We followed recommendations of [64].

The metabolic network of the endosymbiont was reconstructed by identifying orthologs between the chromatophore and *Synechococcus sp.* WH 5701. *Synechococcus sp.* WH 5701 is the closest free-living relative of the chromatophore with a sequenced genome [5].

The metabolic capabilities of both organisms were tested with Flux Balance Analysis [65]. FBA is an optimization algorithm based on lineal programming provided in the Matlab COBRA toolbox [66]. FBA determines the flux distribution of all reactions in the model by maximizing an objective function [30].

The functionality of metabolic models is evaluated by their capacity to produce every metabolite that is necessary for in silico growth. For this, the biomass equation of *Synechocystis sp.* PCC 6803 was assigned as objective function in both models. In silico growth was simulated under autotrophic conditions with CO_2 and photons uptake set to 3.7 mmol × gDW^{-1} × h^{-1} and 100 mmol × gDW^{-1} × h^{-1}, respectively. These were restrictive metabolites in the systems. Nutrient assignment for metabolic functionality of the chromatophore was based on the literature [5] and metabolite requirements predicted by the model for in silico biomass production.

Network robustness analysis

In both models, robustness to gene deletions was analyzed by using the function singleGeneDeletion of the COBRA toolbox. If deletion of a single gene decreases the biomass production over 99%, compared with wild type, this gene was consider as essential for biomass production.

Simulation of metabolic reductive evolution in the chromatophore

To simulate genome reduction, we used the metabolic model of *Synechococcus sp.* WH 5701 (*i*CV498) as a proxy of the free-living ancestor of the chromatophore (Fig. 2). Genetic loss was simulated under Proposed Nutrients and Randomized Nutrients intracellular conditions. All nutrients were available simultaneously since the beginning of the simulations. The algorithm starts by randomly deleting a gene from the *i*CV498 model (i.e., setting its flux to zero) and then evaluating the impact of this deletion in the metabolic functionality by using FBA. If in silico growth rate in this network (lacking a gene) was equal to or above the growth rate of a free-living cyanobacteria (≥ 0.0884 h^{-1}), then this gene was considered as non-essential and permanently removed. In contrast, if the growth rate was below 0.0884 h^{-1} then this gene was considered as essential and retained in the model. This process was repeated until each of the genes in the model was evaluated. The whole process is initiated 500 times which results in a population of 500 reduced metabolic networks.

Genetic concordance between the 500 reduced metabolic networks and chromatophore model (*i*CV265) was analyzed as in [30]. In each of the 500 simulations, a binary variable was assigned for each gene in *i*CV498 depending on whether the gene is predicted to be conserved or not among the 500 simulations. This allowed us to determine the number of occurrences that a gene is predicted as essential in the 500 simulated reduced networks.

Measures of sensitivity and specificity were obtained calculating the fraction of true-positives (fraction of genes predicted to be conserved by the simulations and present in *i*CV265) and false-positives (fraction of genes predicted by the simulations and not present in *i*CV265) for every cutoff (minimal fraction of simulated genomes in which a gene must be present to be predicted as

conserved in *i*CV265). Figure 5 plots true-positive and false-positive (1-specificity) predictions for every cutoff (1 to 500) to form a ROC curve. The area under the curve represents how well the simulations recover gene content in *i*CV265. The area under the curve was empirically calculated as in [67].

Simulation of metabolic integration of the chromatophore with its host

We performed this analysis by using the same algorithm used in the simulation of reductive evolution. However, this analysis was performed only in Proposed Nutrient conditions (Fig. 2). In addition, a reaction simulating hexose export from the chromatophore to the host was defined as objective function and the growth rate equation (biomass equation) was fixed to 0.0884 h^{-1}. Also, in this simulation, a non-essential gene was defined as one whose deletion does not affect the growth rate (0.0884 h^{-1}) and the hexose export. Specifically, the rate of hexose export could vary while being always greater than zero. Otherwise, the gene was defined as essential.

In this analysis the model does not have access to all 13 nutrients at the same time from the beginning of the simulation. Instead, we randomly allow the model to have access to one of the 13 Proposed Nutrients (Fig. 2) and subsequently applied our algorithm of reductive evolution. Once we evaluated the impact of singly deleting each one of the genes, we randomly allowed the model to have access to a second nutrient and newly applied our algorithm of reductive evolution. The analysis stops when *i*CV498 has access to all 13 nutrients and all genes have been tested for essentiality.

Additional files

Additional file 1: Metabolic model of *Synechococcus sp.* WH 5701 (*i*CV498). (XLSX 187 kb)

Additional file 2: Metabolic model of the chromatophore (*i*CV265). (XLSX 121 kb)

Additional file 3: Figure S1 and S2 and Tables S1. (DOCX 149 kb)

Acknowledgements
Not applicable.

Funding
CV thanks to CONACYT for the doctoral fellowship granted.

Availability of data and material
Data and material will be available from the corresponding author.

Authors' contributions
CV performed the model construction, evolutionary simulations and wrote the manuscript. RO and OR provided support in mathematical modelling of metabolic reconstructions. AL and LD contributed to interpretation and discussion of the data. CV and LD designed the study. All authors approved the final version of the manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent of publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Irapuato, Km. 9.6 Libramiento Norte Carr. Irapuato-León, 36821 Guanajuato, Irapuato, Mexico. [2]Departamento de Procesos y Tecnología, Universidad Autónoma Metropolitana-Cuajimalpa, Av. Vasco de Quiroga 4871, Santa Fe, Del. Cuajimalpa, C.P. 05348 Ciudad de Mexico, México, Mexico. [3]Human Systems Biology Laboratory, Coordinación de la Investigación Científica-Red de Apoyo a la Investigación (RAI), UNAM, México City, Mexico. [4]Instituto Nacional de Medicina Genómica (INMEGEN), 14610 México City, Mexico. [5]Unidad de Genómica Avanzada (Langebio), Centro de Investigación y de Estudios Avanzados del IPN, Guanajuato, Irapuato, Mexico.

Received: 4 January 2017 **Accepted:** 28 March 2017
Published online: 14 April 2017

References

1. Nakayama T, Archibald JM. Evolving a photosynthetic organelle. BMC Biol. 2012;10:35–7.
2. Nowack ECM. Paulinella chromatophora - Rethinking the transition from endosymbiont to organelle. Acta Soc Bot Pol. 2014;83:387–97.
3. Yoon HS, Nakayama T, Reyes-Prieto A, Andersen RA, Boo SM, Ishida KI, et al. A single origin of the photosynthetic organelle in different Paulinella lineages. BMC Evol Biol. 2009;9:98.
4. Marin B, Nowack ECM, Melkonian M. A plastid in the making: Evidence for a second primary endosymbiosis. Protist. 2005;156:425–32.
5. Nowack ECM, Melkonian M, Glöckner G. Chromatophore Genome Sequence of Paulinella Sheds Light on Acquisition of Photosynthesis by Eukaryotes. Curr Biol. 2008;18:410–8.
6. Delaye L, Valadez-Cano C, Pérez-Zamorano B. How Really Ancient Is Paulinella Chromatophora? PLoS Curr Tree Life. 2016:1–12. Edition 1. doi:10.1371/currents.tol.e68a099364bb1a1e129a17b4e06b0c8b.
7. Reyes-Prieto A, Yoon HS, Moustafa A, Yang EC, Andersen RA, Boo SM, et al. Differential gene retention in plastids of common recent origin. Mol Biol Evol. 2010;27:1530–7.
8. Nowack ECM, Vogel H, Groth M, Grossman AR, Melkonian M, Glöckner G. Endosymbiotic gene transfer and transcriptional regulation of transferred genes in Paulinella chromatophora. Mol Biol Evol. 2011;28:407–22.
9. Nowack ECM, Grossman AR. Trafficking of protein into the recently established photosynthetic organelles of Paulinella chromatophora. Proc Natl Acad Sci U S A [Internet]. 2012;109:5340–5. Available from: http://www.pnas.org/content/109/14/5340.
10. Kies L, Kremer BP. Function of cyanelles in the thecamoeba Paulinella chromatophora. Naturwissenschaften. 1979;66:578–9.
11. Moya A, Peretó J, Gil R, Latorre A. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet [Internet]. 2008;9:218–29. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18268509.
12. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature. 2000;407:81–6.
13. Lee MC, Marx CJ. Repeated, selection-driven genome reduction of accessory genes in experimental populations. PLoS Genet. 2012;8:2–9.
14. McCutcheon JP, Moran N. a. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. [Internet]. Nat Publ Group. 2011;10:13–26. Available from: http://dx.doi.org/10.1038/nrmicro2670.
15. Marais GAB, Calteau A, Tenaillon O. Mutation rate and genome reduction in endosymbiotic and free-living bacteria. Genetica. 2008;134:205–10.

16. Kuo C, Moran N. a, Ochman H. The consequences of genetic drift for bacterial genome complexity The consequences of genetic drift for bacterial genome complexity. Genome Res. 2009;1450–4.
17. Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc Natl Acad Sci U S A. 1996;93:2873–8.
18. Shou W. Acknowledging selection at sub-organismal levels resolves controversy on pro-cooperation mechanisms. elife. 2015;4:1–19.
19. Bennett GM, Moran NA. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proc Natl Acad Sci [Internet]. 2015;112:10169–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25733367.
20. Thomas GH, Zucker J, Macdonald SJ, Sorokin A, Goryanin I, Douglas AE. A fragile metabolic network adapted for cooperation in the symbiotic bacterium Buchnera aphidicola. BMC Syst Biol. 2009;3:24.
21. Wang Z, Zhu X-G, Chen Y, Li Y, Hou J, Li Y, et al. Exploring photosynthesis evolution by comparative analysis of metabolic networks between chloroplasts and photosynthetic bacteria. BMC Genomics [Internet]. 2006;7:100. Available from: http://www.biomedcentral.com/1471-2164/7/100.
22. Stephanopoulos G. Metabolic Fluxes and Metabolic Engineering. Metab Eng. 1999;1(1):1–11.
23. Lan EI, Liao JC. Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. Metab Eng [Internet]. 2011;13:353–63. [cited 2017 13]. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1096717611000474.
24. Lee JW, Kim TY, Jang YS, Choi S, Lee SY. Systems metabolic engineering for chemicals and materials [Internet]. Trends Biotechnol Elsevier; 2011. p. 370–378. [cited 2017 Mar 13]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21561673.
25. Lee SY, Mattanovich D, Villaverde A, et al. Microb Cell Fact [Internet]. 2012;11:156. Available from: http://www.microbialcellfactories.com/content/11/1/156.
26. Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 2012;109:2678–83. Available from: http://www.pnas.org/content/109/7/2678.abstract.
27. Lürling M, Eshetu F, Faassen EJ, Kosten S, Huszar VLM. Comparison of cyanobacterial and green algal growth rates at different temperatures. Freshw Biol. 2013;58:552–9.
28. González-Domenech CM, Belda E, Patiño-Navarrete R, Moya A, Peretó J, Latorre A. Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two Blattabacterium cuenoti strains, primary endosymbionts of cockroaches. BMC Microbiol [Internet] BioMed Central Ltd.; 2012;12(Suppl 1):S5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3287516&tool=pmcentrez&rendertype=abstract.
29. Belda E, Silva FJ, Peretó J, Moya A. Metabolic networks of Sodalis glossinidius: A systems biology approach to reductive evolution. PLoS One. 2012;7(1):e30652. doi:10.1371/journal.pone.0030652.
30. Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. Chance and necessity in the evolution of minimal metabolic networks. Nature. 2006;440:667–70.
31. Karkar S, Facchinelli F, Price DC, Weber APM, Bhattacharya D. Metabolic connectivity as a driver of host and endosymbiont integration. Proc Natl Acad Sci U S A [Internet]. 2015;2014213175. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4547263/.
32. Facchinelli F, Colleoni C, Ball SG, Weber APM. Chlamydia, cyanobiont, or host: Who was on top in the ménage à trois? Trends Plant Sci. 2013;18:673–9.
33. Villadsen J, Nielsen J, Lidén G. Bioreaction Engineering Principles. New York: Kluwer Academic/Plenum Publishers; 2003. Available from: http://www.springerlink.com/index/10.1007/978-1-4419-9688-6.
34. Tyra HM, Linka M, Weber APM, Bhattacharya D. Host origin of plastid solute transporters in the first photosynthetic eukaryotes. Genome Biol [Internet]. 2007;8:R212. Available from: http://genomebiology.com/2007/8/10/R212.
35. Canbäck B, Tamas I, Andersson SGE. A phylogenomic study of endosymbiotic bacteria. Mol Biol Evol. 2004;21:1110–22.
36. Allen JM, Light JE, Perotti MA, Braig HR, Reed DL. Mutational meltdown in primary endosymbionts: Selection limits Muller's Ratchet. PLoS One. 2009;4(3):e4969. doi:10.1371/journal.pone.0004969.
37. Lynch M, Gabriel W. Mutation Load and the Survival of Small Populations. Evolution (N Y). 1990;44:1725–37.
38. Andersson SG. e, Kurland CG. Reductive evolution of resident genomes. Trends Microbiol. 1998;6:263–8.
39. Bergstrom CT, Pritchard J. Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics. 1998;149:2135–46.
40. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, et al. The Wolbachia genome of Brugia malayi: Endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 2005;3:0599–614.
41. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, et al. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: A streamlined genome overrun by mobile genetic elements. PLoS Biol. 2004;2:327–41.
42. Moriyama M, Nikoh N, Hosokawa T, Fukatsu T. Riboflavin Provisioning Underlies Wolbachia's Fitness Contribution to Its Insect Host. MBio. 2015;6:1–8.
43. Wilson ACC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, et al. Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium Buchnera aphidicola. Insect Mol Biol. 2010;19:249–58.
44. Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, et al. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell [Internet]. Elsevier. 2013;153:1567–78. Available from: http://dx.doi.org/10.1016/j.cell.2013.05.040.
45. Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, et al. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Mol Biol Evol. 2014;31:857–71.
46. Hansen AK, Moran NA. The impact of microbial symbionts on host plant utilization by herbivorous insects. Mol Ecol. 2014;23:1473–96.
47. Oberhardt MA, Puchalka J, Fryer KE, VAP MDS, Papin JA. Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. J Bacteriol. 2008;190:2790–803.
48. Ates O, Oner ET, Arga KY. Genome-scale reconstruction of metabolic network for a halophilic extremophile, Chromohalobacter salexigens DSM 3043. BMC Syst Biol [Internet] BioMed Central Ltd. 2011;5:12. Available from: http://www.biomedcentral.com/1752-0509/5/12.
49. Puchalka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA, Martins dos Santos VA. Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. PLoS Comput Biol. 2008;4(10):e1000210. doi:10.1371/journal.pcbi.1000210.
50. Wagner A. Robustness and evolvability: A paradox resolved. Proc R Soc Biol Sci. 2008;275:91–100.
51. Gabaldón T, Peretó J, Montero F, Gil R, Latorre A, Moya A. Structural analyses of a hypothetical minimal metabolism. Philos Trans R Soc Lond B Biol Sci [Internet]. 2007;362:1751–62. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2442391&tool=pmcentrez&rendertype=abstract.
52. Ding T, Case KA, Ornelo MA, Roland HA, Metz ZP, Diao X, et al. Predicting Essential Metabolic Genome Content of Niche-Specific Enterobacterial Human Pathogens during Simulation of Host Environments. PLoS One [Internet]. 2016;11:e0149423. Available from: http://dx.plos.org/10.1371/journal.pone.0149423.
53. Mendonça AG, Alves RJ, Pereira-Leal JB. Loss of genetic redundancy in reductive genome evolution. PLoS Comput Biol. 2011;7(2):e1001082. doi:10.1371/journal.pcbi.1001082.
54. Khannapho C, Zhao H, Bonde BK, Kierzek AM, Avignone-Rossa CA, Bushell ME. Selection of objective function in genome scale flux balance analysis for process feed development in antibiotic production. Metab Eng. 2008;10:227–33.
55. Patil KR, Rocha I, Forster J, Nielsen J. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics [Internet]. 2005;6:308. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=16375763&retmode=ref&cmd=prlinks.
56. Johnson MD. The acquisition of phototrophy: Adaptive strategies of hosting endosymbionts and organelles. Photosynth Res. 2011;107:117–32.
57. Giersch C, Robinson SP. Regulation of photosynthetic carbon metabolism during phosphate limitation of photosynthesis in isolate spinach chloroplast. Photosynth Res. 1987;14:211–27.
58. Heldt HW, Chon CJ, Maronde D. Role of orthophosphate and other factors in the regulation of starch formation in leaves and isolated chloroplasts. Plant Physiol. 1977;59:1146–55.
59. Pickens LB, Tang Y, Chooi Y-H. Metabolic Engineering for the Production of Natural Products. Annu Rev Chem Biomol Eng. 2011;2:211.
60. Burgard AP, Pharkya P, Maranas CD. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. Biotechnol Bioeng. 2003;84:647–57.

61. Russell CW, Poliakov A, Haribal M, Jander G, van Wijk KJ, Douglas AE. Matching the supply of bacterial nutrients to the nutritional demand of the animal host. Proc R Soc B Biol Sci [Internet]. 2014;281:20141163. Available from: http://rspb.royalsocietypublishing.org/content/281/1791/20141163.short?rss=1
62. Zientz E, Dandekar T, Gross R. Metabolic Interdependence of Obligate Intracellular Bacteria and Their Insect Hosts. Microbiol Mol Biol Rev. 2004;68:745–70.
63. Castillo-Davis CI, Hartl DL. GeneMerge - Post-genomic analysis, data mining, and hypothesis testing. Bioinformatics. 2003;19:891–2.
64. Thiele I, Palsson B. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 2010;5:93–121.
65. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat Biotech [Internet]. Nat Publ Group. 2010;28:245–8. Available from: http://dx.doi.org/10.1038/nbt.1614%5Cnhttp://www.nature.com/nbt/journal/v28/n3/abs/nbt.1614.html.
66. Schellenberger J, Thiele I, Orth JD. Quantitative prediction of cellular metabolism with constraint- based models: the COBRA Toolbox v2.0. Nat Protoc. 2012;6:1290–307.
67. Hanley JA, BJ MN. The Meaning and Use of the Area under a Receiver Operating (ROC) Curve Characteristic. Radiology [Internet]. 1982;143:29–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7063747.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

6.3 Artículo 3. Massive protein import into the early evolutionary stage photosynthetic organelle of the amoeba *Paulinella chromatophora*

Please cite this article in press as: Singer et al., Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba *Paulinella chromatophora*, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.08.010

Current Biology
Article

CellPress

Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba *Paulinella chromatophora*

Anna Singer,[1] Gereon Poschmann,[2] Cornelia Mühlich,[3] Cecilio Valadez-Cano,[1,4] Sebastian Hänsch,[5] Vanessa Hüren,[1] Stefan A. Rensing,[3,6] Kai Stühler,[2,7] and Eva C.M. Nowack[1,8,*]

[1]Department of Biology, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany
[2]Molecular Proteomics Laboratory, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany
[3]Faculty of Biology, University of Marburg, 35043 Marburg, Germany
[4]Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Irapuato, 36821 Irapuato, Guanajuato, México
[5]Center for Advanced Imaging, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany
[6]BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany
[7]Institute for Molecular Medicine, University Hospital Düsseldorf, 40225 Düsseldorf, Germany
[8]Lead Contact
*Correspondence: e.nowack@uni-duesseldorf.de
http://dx.doi.org/10.1016/j.cub.2017.08.010

SUMMARY

The endosymbiotic acquisition of mitochondria and plastids more than 1 Ga ago profoundly impacted eukaryote evolution. At the heart of understanding organelle evolution is the re-arrangement of the endosymbiont proteome into a host-controlled organellar proteome. However, early stages in this process as well as the timing of events that underlie organelle integration remain poorly understood. The amoeba *Paulinella chromatophora* contains cyanobacterium-derived photosynthetic organelles, termed "chromatophores," that were acquired more recently (around 100 Ma ago). To explore the re-arrangement of an organellar proteome during its integration into a eukaryotic host cell, here we characterized the chromatophore proteome by protein mass spectrometry. Apparently, genetic control over the chromatophore has shifted substantially to the nucleus. Two classes of nuclear-encoded proteins—which differ in protein length—are imported into the chromatophore, most likely through independent pathways. Long imported proteins carry a putative, conserved N-terminal targeting signal, and many specifically fill gaps in chromatophore-encoded metabolic pathways or processes. Surprisingly, upon heterologous expression in a plant cell, the putative chromatophore targeting signal conferred chloroplast localization. This finding suggests common features in the protein import pathways of chromatophores and plastids, two organelles that evolved independently and more than 1 Ga apart from each other. By combining experimental data with in silico predictions, we provide a comprehensive catalog of almost 450 nuclear-encoded, chromatophore-targeted proteins. Interestingly, most imported proteins seem to derive from ancestral host genes, suggesting that the re-targeting of nuclear-encoded proteins that resulted from endosymbiotic gene transfers plays only a minor role at the onset of chromatophore integration.

INTRODUCTION

Eukaryotes co-opted photosynthetic carbon fixation more than 1 Ga ago from prokaryotes by endosymbiotic uptake of a cyanobacterium and stably integrating it as a photosynthetic organelle (plastid) [1, 2]. At the heart of understanding organelle evolution and the physiology of the resulting chimeric organism is the re-arrangement of the endosymbiont proteome as a consequence of loss of endosymbiont genes and import of nuclear-encoded proteins. But the ancient origin of plastids (and mitochondria) makes it difficult to infer the nature of the initial interactions between host and endosymbiont and the timing of events that underpin organellogenesis. The thecate amoeba *Paulinella chromatophora* (Cercozoa, Rhizaria) offers the exciting opportunity to explore interactions of a eukaryotic host cell with a photosynthetic organelle at an earlier stage of integration [3]. *P. chromatophora* harbors photosynthetic organelles termed "chromatophores" that originated only 90–140 Ma ago [4] from an α-cyanobacterium through an endosymbiotic event that is independent of the one that gave rise to plastids in the Archaeplastida [5, 6]. The chromatophore is surrounded by two envelope membranes and a peptidoglycan layer [7]. Its genome (1 Mb) is reduced to approximately one-third of the ancestral cyanobacterial genome [8], and >70 genes of α-cyanobacterial origin, which most likely result from endosymbiotic gene transfers (EGTs) from the chromatophore, are now expressed from the nuclear genome [9–11]. Overall, nuclear- and chromatophore-encoded functions are highly complementary, with many genes for proteins involved in the same metabolic pathway or cellular process being distributed over both chromatophore and nuclear genome

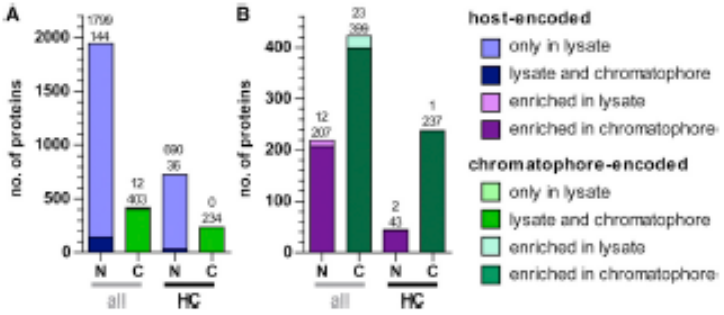

Figure 1. Proteins Identified by MS Analysis
Numbers of proteins identified in (A) whole-cell lysates and (B) chromatophores. Nuclear-encoded proteins ("N") are in blue or violet color; chromatophore-encode proteins ("C") are in light and dark green color. For both groups, numbers for all proteins identified ("all") and high-confidence proteins ("HC") are displayed separately. A color code indicates identification of proteins exclusively in the whole-cell lysates or both samples (in A) and enrichment of the protein in the chromatophore or whole-cell lysate (in B). Total numbers of proteins is given above each bar. See also Figures S1–S3 and Tables S1, S2, and S3.

[9]. In addition to genes of eukaryotic and EGT origin, many genes apparently originating from various bacterial lineages through horizontal gene transfer (HGT) have predicted functions that would compensate for lost chromatophore functions [9].

We recently found that two EGT-derived small photosystem I (PSI) subunits, PsaE and PsaK, are synthesized in the amoeba cytoplasm and are imported into the chromatophore. Surprisingly, despite the lack of an N-terminal signal peptide (SP; which usually confers co-translational protein import into the endoplasmic reticulum [ER]) as well as transmembrane domains (which could confer post-translational insertion into the ER membrane through the GET pathway [12]), PsaE seems to traffic on the way to the chromatophore through the Golgi [13]. Compounds of a hypothetical basic protein translocator, most notably an ortholog of TIC21 that forms a protein-conducting channel in the inner plastid membrane, have been identified in the chromatophore genome [14, 15]. A homolog of TOC75, the main import channel of the chloroplast outer membrane has not been found in *P. chromatophora*. Hence, the mechanism through which proteins are imported into the chromatophore remains unclear. Also, whether the chromatophore has evolved the capacity to import, besides short proteins, also higher-molecular-weight metabolic enzymes or metabolic pathways are rather connected through extensive metabolite shuttling between chromatophore and host is currently unknown.

To explore the proteome composition of an early-evolutionary-stage photosynthetic organelle, we set out to identify chromatophore-targeted proteins by protein mass spectrometry (MS). In particular, our study aimed to (1) test whether, in addition to short EGT-derived proteins, also host- or HGT-derived metabolic enzymes contribute to the chromatophore proteome, (2) deduce metabolic pathways that localize to the chromatophore, (3) identify common sequence motifs within chromatophore-targeted proteins that might serve as chromatophore targeting signal, and (4) develop an in silico approach to predict chromatophore-targeted proteins.

RESULTS

The *P. chromatophora* Proteome Dataset

To examine the chromatophore proteome composition, chromatophores were isolated from *P. chromatophora* CCAC0185 to high purity (Figures S1A and S1B). Proteins extracted from isolated chromatophores and whole-cell lysates showed distinct banding patterns in SDS-PAGE. The prominent two bands at ~17 kDa that represent the chromatophore-encoded phycobiliproteins [13] appear strongly enriched in isolated chromatophores (Figure S1C). MS analysis of three independent chromatophore isolations and whole-cell lysates yielded a highly reproducible pattern of normalized ion intensities between replicates (Figure S1D).

Overall, we identified 2,452 *P. chromatophora* proteins (Figure 1; Table S1). 2,358 of these were identified in the whole-cell lysates and 641 in the chromatophores. 960 and 283 proteins were identified with at least two different peptides and in at least two replicates in whole-cell lysates or chromatophores, respectively. These proteins are regarded as "high-confidence (HC) proteins"; the remaining proteins are regarded as "low-confidence (LC) proteins." Marker proteins for cytoskeleton, ER, mitochondrion, and peroxisome were absent in chromatophores (Table S2). Also, of the 100 nuclear-encoded proteins that were most abundant in the whole-cell lysates (Table S3, green section), only two proteins (annotated as β-carbonic anhydrase and polyubiquitin) were identified in the chromatophores, although with strongly diminished average ion intensities as compared to whole-cell lysates (>200-fold and 4.5-fold lower signal intensity for the carbonic anhydrase and polyubiquitin, respectively). At least for polyubiquitin, a specific interaction with chromatophore outer membrane proteins seems possible (see Discussion). Together, these results indicate high purity of the chromatophore samples with virtually no unspecific contamination from other cell compartments.

Chromatophore-Encoded Proteins

Of the 641 proteins (283 HC proteins) identified in the chromatophore fraction, 422 (238 HC proteins) were chromatophore encoded (Figure 1B). This corresponds to a coverage of 49% (27% for HC proteins) of the total 867 chromatophore-encoded proteins. In accord with its photosynthetic function, the 20 proteins yielding highest average ion intensities in chromatophore samples were chromatophore-encoded subunits of the cyanobacterial light-harvesting antennae, the carboxysome, the photosynthetic apparatus (PsaC, PsaD, PsaF, and PsbO), the F-type ATPase, metabolic enzymes involved in carbon fixation via the Calvin cycle, and the molecular chaperones GroEL and DnaK (Table S3, yellow section). Of the 422 chromatophore-encoded proteins identified in chromatophores, 399 proteins (or 237 of 238 HC proteins) appeared enriched in chromatophores

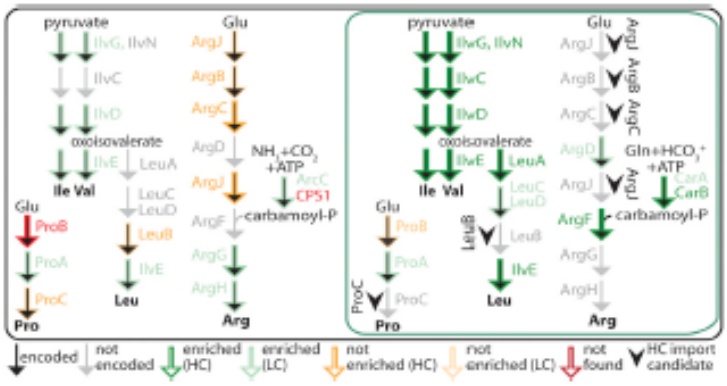

Figure 2. Long Import Candidates Fill Gaps in Chromatophore-Encoded Metabolic Pathways

Chromatophore-encoded (within green rectangle) and nuclear-encoded genes underlying the proline, leucine, and arginine biosynthetic pathway are displayed as simple black arrows. Grey arrows and letters indicate that the respective gene is missing from the chromatophore genome or is not found in nuclear transcriptome data. Colorful arrows surrounding the black arrows and letters represent the corresponding gene products and indicate whether the protein was found enriched (green), not enriched (orange), or not found at all (red) in chromatophores (within green rectangle) or whole-cell lysate (outside of green rectangle) by MS. Dark colors represent HC proteins, and pale colors represent LC proteins. Black arrowheads represent HC import candidates. See also Figure S2 and Table S3.

as compared to whole-cell lysates (i.e., average ion intensities in chromatophores [Int_C] – average ion intensities in lysates [Int_L] > 0; Figure 1B). The 23 chromatophore-encoded proteins that appear depleted in chromatophores most likely reflect rather the semiquantitative nature of MS analyses than protein export (all of the depleted proteins are of low abundance, i.e., 3.6%–0.1% of the average protein abundance found across all proteins identified within chromatophores). In line with this interpretation, several of these depleted proteins have functions that clearly localize to the chromatophore (e.g., an ATP synthase subunit, 50S ribosomal protein, and plastocyanin precursor).

Two Classes of Proteins Appear to Be Imported into the Chromatophore

The remaining 219 proteins (45 HC proteins) identified in the chromatophores were nuclear-encoded. 207 of these 219 proteins (43 of 45 nuclear-encoded HC proteins) were enriched in chromatophores as compared to whole-cell lysates and were regarded as import candidates (Figure 1B; Table S3). Examining the import candidates revealed two groups of proteins. Considering only the 145 proteins for which full-length coding sequence (CDS) information was available, the first group comprises 85 short proteins of <90 amino acid length (with 10 HC import candidates); the second group, with a broader variance in lengths, comprises 60 proteins with a length of >268 amino acids (with 24 HC import candidates) (Figure S2). Retrieval of the full-length mRNA sequence was confirmed by the presence of a spliced leader (SL) sequence at the 5′ end (see [9]) or an in-frame stop codon upstream of the start methionine.

Notably, among the short LC import candidates were the nuclear-encoded PSI subunits PsaE and PsaK, which have been demonstrated before to be part of PSI [13]. Other short import candidates were four EGT-derived high-light-inducible (HLI) photoprotective proteins [10, 11] and two proteins annotated as cold-shock proteins. However, most of the short import candidates do not have homologs in other organisms (67 proteins) or are conserved proteins of unknown function (10 proteins). Aligning short import candidates with homologous sequences from other species confirmed the previous notion that short chromatophore-targeted proteins lack N-terminal pre-sequences [13] (Figure S3A). Interestingly, many of the short import candidates (39 proteins) exhibit conspicuous cysteine motifs (CxxC or CxxxxC) and/or stretches of positively charged amino acids that are characteristic of antimicrobial peptides (AMPs) (Figure S3B).

Among the 24 long HC import candidates, 18 have specific annotated functions (Table S3, blue section). Most of these proteins represent enzymes involved in the primary metabolism, and several specifically fill gaps in chromatophore-encoded metabolic pathways (Figure 2). Examples are the pyrroline-5-carboxylate reductase (ProC) and the 3-isopropylmalate dehydrogenase (LeuB) that catalyze the only enzymatic reactions in proline and leucine biosynthesis, respectively, for which the corresponding gene is missing from the chromatophore genome (Figure 2). The arginine biosynthetic pathway is encoded by a complementary patchwork of nuclear and chromatophore genes. Enzymes that perform the first three and the fifth step of the pathway (the bifunctional glutamate/ornithine acetyltransferase ArgJ, the N-acetylglutamate kinase ArgB, and the N-acetyl-gamma-glutamyl-phosphate reductase ArgC) were found to be imported into the chromatophore; enzymes that catalyze steps four and six (ArgD and ArgF) as well as the generation of carbamoyl phosphate (CarA and CarB) are chromatophore-encoded and were also identified in the chromatophore. Enzymes that catalyze the last two steps of the pathway (ArgG and ArgH) are nuclear-encoded and carry a predicted mitochondrial transit peptide (mTP) [13]. Both proteins were identified in whole-cell lysates, but not in chromatophores, in congruence with a putative mitochondrial localization.

Identification of a Putative N-Terminal Import Signal in Long Import Candidates

In contrast to the short import candidates, all long HC import candidates showed conspicuous N-terminal extensions compared to orthologous sequences from other species (Figure 3A). Aligning the N termini of these 24 proteins revealed a conserved protein sequence of approximately 200 amino acids length (varying from 97 to 221 amino acids) that we interpret in analogy to mitochondrial and plastid transit peptides (mTPs and cTPs) as a putative chromatophore transit peptide (crTP)

Please cite this article in press as: Singer et al., Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba *Paulinella chromatophora*, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.08.010

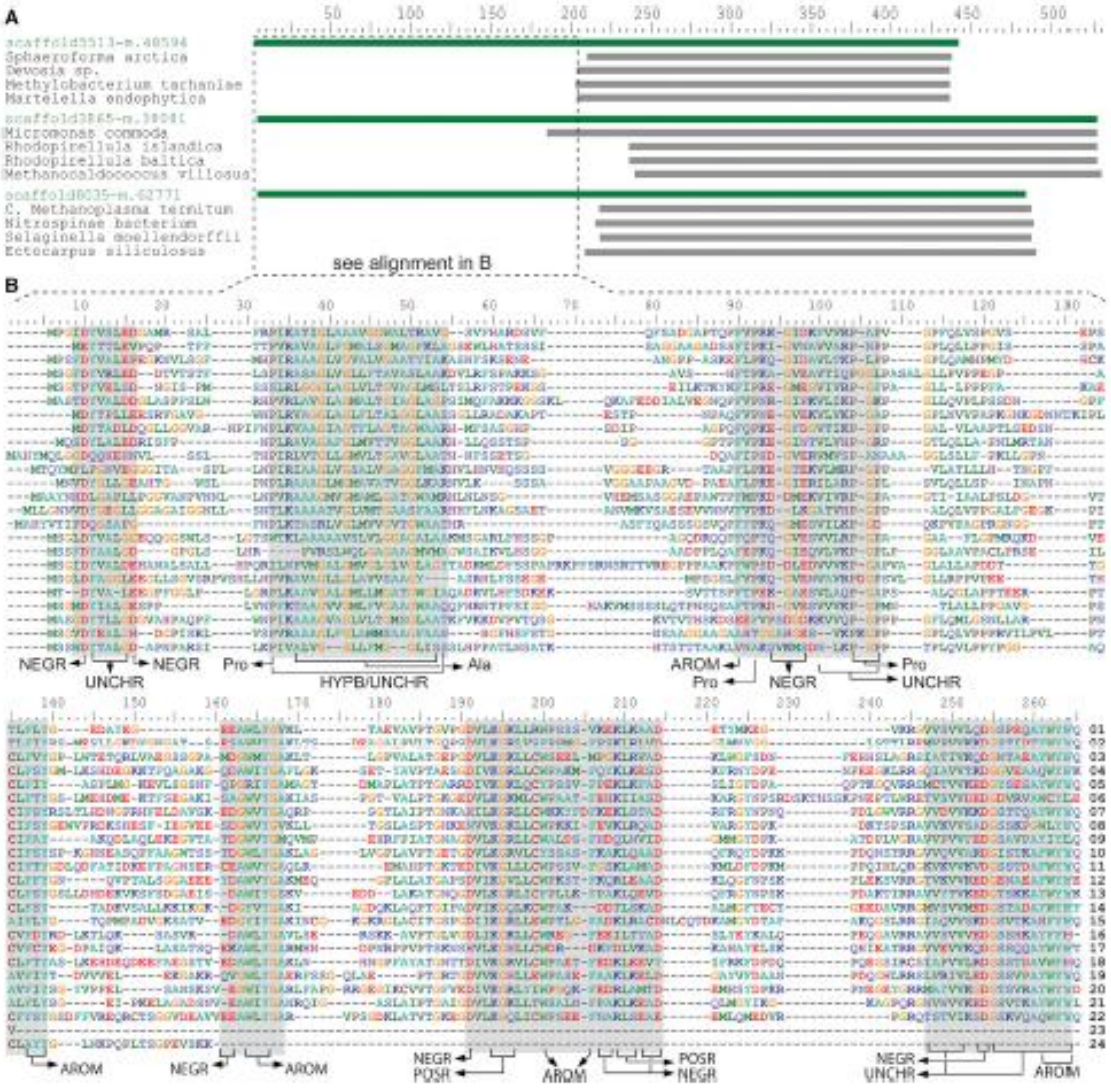

Figure 3. Long Import Candidates Contain a Putative crTP

(A) Schematic amino acid alignments for three long import candidates (annotated as flavin reductase [scaffold5513-m.48594], N-acetylglutamate kinase [scaffold3865-m.38081], and pyrroline-5-carboxylate reductase [scaffold8035-m.62771]) with orthologous protein sequences from other species (selected among the 5 best BlastP hits against the nr database at NCBI). *P. chromatophora* import candidates show conspicuous N-terminal extensions (represented in green color) as compared to sequences from other species.

(B) Alignment of crTPs from the 24 long HC import candidates. Conserved motifs used for the HMM prediction approach are highlighted in gray. Features identified as having strong predictive value in distinguishing imported and not imported proteins by the SVM approach and correlating to conserved sequence motives in the alignment are indicated underneath the alignment.

For abbreviations, see Table S7. For identity of sequences 01–24, see Table S3. See also Figures S2 and S4.

(Figure 3B). Also, 26 out of 36 LC import candidates carried the N-terminal crTP. The remaining 10 LC import candidates that lack a crTP might represent proteins that specifically associate with the outer chromatophore membrane, proteins imported through a crTP-independent pathway, or contaminants (8 out of 10 proteins were identified with one or two peptides only). Analyzing the 24 crTPs from HC import candidates revealed that crTPs show no significant sequence similarities to known protein sequences in the NCBI nr protein database (blastp; E value cutoff of 1.0E−3), and no protein family membership,

domains, repeats, or Gene Ontology (GO) terms were found by InterProScan [16]. The net charge of the crTPs at pH 7 varies between +9.0 and −13.7. Known targeting sequences (SPs, mTPs, or cTPs) were not predicted by TargetP [17]. Only for 2 out of 24 proteins (scaffold1147-m.16381 and scaffold2581-m.28831), a transmembrane domain was predicted within the crTP. However, all 24 sequences have a predicted propensity to form hydrophobic or amphiphilic α helices (Figures S4A and S4B) [18]. Finally, we noted that, in chromatophore samples, MS-identified tryptic peptides originated not exclusively from the conserved part of import candidates but also from their crTPs (Figure S4C). This finding suggests that the crTP is either not cleaved off after protein translocation or that the cleaved crTP is relatively stable.

In Silico Prediction of Chromatophore-Targeted Proteins

To computationally identify additional (long) import candidates within the 60,973 translated nuclear *P. chromatophora* transcripts, we generated support vector machine (SVM) models for the long import candidates. In addition to that, a hidden Markov model (HMM) approach was applied (see STAR Methods for details). We used a combination of the SVM (with a cutoff of ≥0.663) and the HMM approach with ≥4, ≥3, or ≥2 identified sequence motifs (SVM&HMM_4, SVM&HMM_3, or SVM&HMM_2). These three approaches, respectively, yielded a total of 478, 498, or 519 predicted chromatophore-targeted proteins >250 amino acids. To investigate specificity and sensitivity of the three different methods, we compared in silico predictions with MS-inferred protein localizations. Of the 33 MS-identified long import candidates, which were not used as training or test data for the SVM and HMM approaches, SVM&HMM_4 and SVM&HMM_3 failed to identify 11 and SVM&HMM_2 failed to identify 10 proteins. These 10 false-negative proteins were the LC import candidates lacking crTPs (see above). The remaining protein that was missed by SVM&HMM_4 and SVM&HMM_3 but identified by SVM&HMM_2 showed an unusually divergent crTP. Of the 670 full-length proteins that were found exclusively in whole-cell lysates by MS, all three methods predicted the same 9 proteins as imported. Manual inspection of these sequences revealed the conserved crTP sequence at their N termini, suggesting that these proteins are imported into the chromatophore but were missed by the MS analysis (all 9 proteins were identified with ≤3 peptides). Thus, because SVM&HMM_2 showed a slightly increased sensitivity over SVM&HMM_3 and SVM&HMM_4, whereas the specificity of the 3 models was the same (Figure 4A), for further analyses, SVM&HMM_2 predictions were used.

Out of the 519 proteins predicted with SVM&HMM_2, 220 proteins were predicted only by SVM, 113 only by HMM, and 186 by both approaches (Table S4). 5, 15, and 50 of these proteins were identified as import candidates by MS, respectively. The N termini of all proteins predicted with both models, and 105 out of 113 proteins predicted by the HMM_2 model, aligned well with the known crTP sequences, whereas the N termini of the remaining 8 proteins predicted by HMM_2 alone did not. Of the proteins identified by SVM alone, only three (all of them MS-identified import candidates) contained an N terminus that was easily alignable to the known crTPs, although the crTP sequence in these three proteins was quite divergent and much shorter than usual (e.g., Figure 3B, sequences 23 and 24). Among the remaining 225 predicted import candidates that lack an easily alignable crTP sequence, there were only two MS-identified import candidates. Thus, whether these proteins are imported into the chromatophore based on similar physicochemical properties of their N termini with the conserved crTP sequence or represent false-positive predictions is currently unclear and will have to be tested experimentally. Therefore, at this point, we consider only the 291 in-silico-predicted proteins with alignable crTP as confidently predicted import candidates (Table S5).

Reconstruction of Chromatophore-Localized Metabolic Pathways and Processes

Many of these 291 confidently predicted import candidates seem to fill further gaps in chromatophore-localized biosynthetic pathways/processes (Figures 4B–4D). For some nuclear-encoded proteins that would be expected to localize to the chromatophore (based on the presence/absence pattern of genes involved in the same pathway on the chromatophore genome), no predictions were obtained, owing to the absence of 5′ full-length transcripts. One example is the uroporphyrinogen III synthase (HemD) involved in chlorophyll biosynthesis. However, manual inspection of the corresponding transcript revealed the presence of a truncated crTP. In this way, complete chromatophore-localized biosynthetic pathways for further amino acids, co-factors, and nucleotides were reconstructed (Figure 4C). Further examples of functions provided by predicted import candidates include the ligation of DNA nicks during DNA replication through a DNA ligase of bacterial origin [9] (Figure 4D) and the detoxification of methylglyoxal (Figure 4B). In the pathways analyzed, most nuclear-encoded enzymes for which a chromatophore-encoded variant is present did not feature a crTP (Figures 4B–4D). For a few enzymes, however, a nuclear isoform carrying a crTP was observed despite the presence of a chromatophore gene for the same enzyme (e.g., GlbA and HemF; Figures 4B and 4C).

Taken together, experimental evidence and in silico predictions identified 433 putatively chromatophore-targeted proteins (Table S5). 226 of these proteins are of unknown function, the remaining 207 involved in diverse cellular processes (Figure 4E). The largest functional group (47 of 116 proteins that could be assigned to a specific functional class) contains proteins involved in genetic information processing (Figure 4E); in particular, we identified 22 zinc finger-type transcription factors, several DNA and RNA helicases, and one poly-A polymerase. Also, many metabolic enzymes seem to be imported into the chromatophore with 13 proteins involved in amino acid, 13 in carbohydrate, two in each lipid and co-factor/vitamin, and 9 in nucleotide metabolism. 9 chromatophore-targeted proteins function in photosynthesis and light acclimation, and further 9 proteins respond to oxidative stress. Finally, 7 proteins might be involved in protein transport and folding (Figure 4E); among them is an ortholog of the chloroplast protein translocator component TIC32. Overall, chromatophore-targeted proteins showed a pronounced over-representation of plastid or sub-plastidial compartment-related GO terms (Figure 4F).

Phylogenetic Origin of Imported Proteins

Interestingly, only 17 of the 433 (experimentally determined + predicted) import candidates are of α-cyanobacterial origin

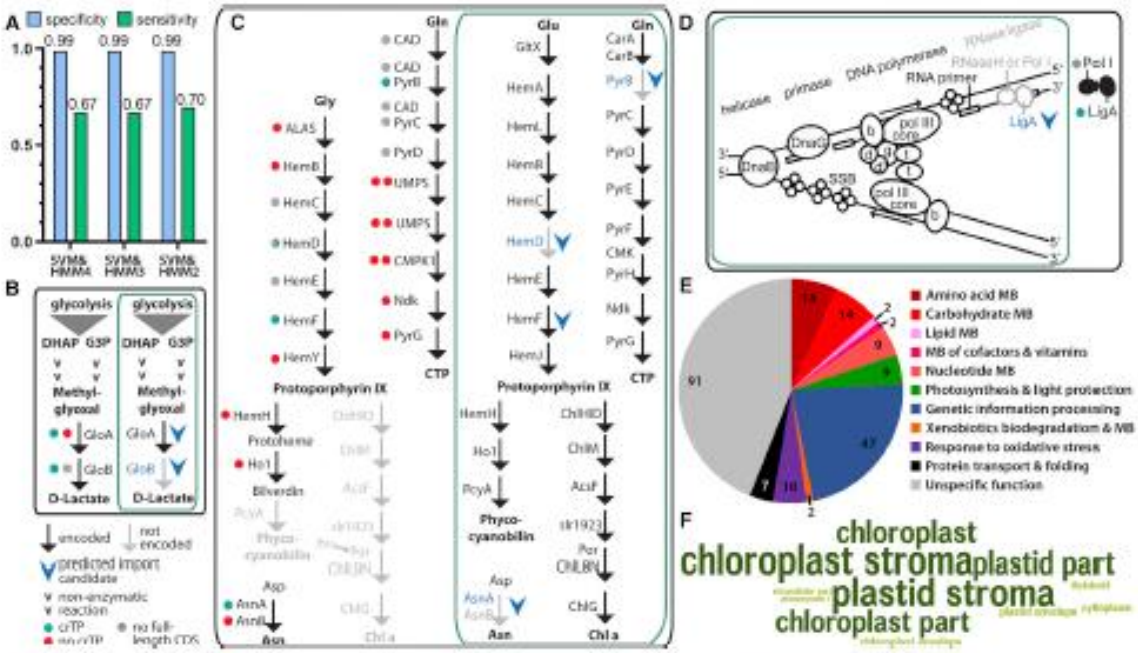

Figure 4. Functional Analyses of Experimentally Determined and In-Silico-Predicted Import Candidates

(A) Sensitivity and specificity of the three prediction methods SVM&HMM_4, SVM&HMM_3, and SVM&HMM_2 as evaluated by comparison of in-silico-predicted imported and not imported proteins with experimentally determined protein localizations (see main text).

(B–D) Reconstruction of metabolic pathways with predicted import candidates. Candidates are involved in (B) detoxification (e.g., degradation of methylglyoxal by lactoylglutathione lyase [GloA] and hydroxyacylglutathione hydrolase [GloB]); (C) fill gaps in co-factor (e.g., chlorophyll a [Chl a] and phycocyanobilin), nucleotide (e.g., cytidine triphosphate [CTP]), and amino acid (e.g., asparagine [Asn]) biosynthetic pathways; and (D) participate in genetic information processing (e.g., ligation of DNA nicks through the DNA ligase A [LigA]) in the chromatophore. Localization of underlying genes is represented as in Figure 2. Blue arrowheads and letters represent in-silico-predicted import candidates. The circles adjacent to the nuclear-encoded enzymes indicate targeting predictions: green, presence of crTP; red, absence of crTP; gray, absence of full-length CDS; and gray/green, manual identification of truncated crTP (see main text). Multiple circles indicate presence of multiple nuclear-encoded protein isoforms.

(E) Functional classes of import candidates (for details, see Table S5). MB, metabolism.

(F) Word-cloud visualization of GO bias analysis (for cellular component) of import candidates. All import candidates (experimentally determined + in silico predicted; see Table S5) were analyzed against the background of all nuclear-encoded *P. chromatophora* proteins.

See also Tables S4, S5, S6, and S7.

and most likely arose through EGT (Table S5). Most of these EGT proteins are involved in photosynthesis and light acclimation. Additional 26 import candidates are of bacterial origin and were most likely acquired through HGT [9–11] (see Table S5). Thus, the remaining 390 import candidates are of host or uncertain origin.

It has been described before that many EGT/HGT genes in *P. chromatophora* experienced gene family expansions probably after the transfer event [9]. One example is the nuclear-encoded cysteine synthase A (CysK) of α-cyanobacterial origin that diversified into three nuclear isoforms. Two of these isoforms were identified by unique peptides in our MS analysis; the chromatophore-encoded CysK (PCC0298) was not experimentally identified. Intriguingly, subcellular localization of the nuclear-encoded CysK isoforms correlated with features in the protein sequences (Figure 5). Whereas scaffold4337-m.41170 (which was enriched in chromatophores) acquired a crTP after gene duplication, scaffold2205-m.25816 (which was exclusively found in whole-cell lysates and thus most likely remains in the cytoplasm) shows no N-terminal extension but acquired a C-terminal Rossmann fold. The Rossmann fold is a protein structural motif that is found in many nucleotide-binding proteins but is neither found in the chromatophore nor other cyanobacterial CysK proteins. Its specific function in the cytoplasmic CysK isoform remains elusive.

Subcellular Localization of Chromatophore-Targeted Proteins upon Expression in Plant Cells

To test whether the *P. chromatophora* crTP or cryptic targeting signals within short import candidates would show any targeting capacity in the heterologous context of a plant cell, we expressed recombinant fluorescent fusion proteins transiently in tobacco (Figure 6). For this purpose, (1) the crTP derived from scaffold5513-m.48594 (annotated as flavin reductase; Figure 3B, sequence 5) and (2) the short import candidate PsaE (scaffold17097-m.100929) were fused to the N terminus of the yellow fluorescent protein (YFP), transiently expressed in *Nicotiana*

Figure 5. crTP Acquisition in the EGT-Derived Protein CysK

Schematic amino acid alignment (left) and proteomic data and gene localization (right) for cyanobacterium-derived cysteine synthase A isoforms in *P. chromatophora*. Abbreviations: chr, chromatophore; Int_C and Int_L, average ion intensity in chromatophores and whole-cell lysates; Loc, protein localization determined by MS; neg, negative value; nuc, nucleus; pos, positive value; Rep_C and Rep_L, identified in x chromatophore/lysate replicates. See also Table S3.

benthamiana plants, and protein localization was analyzed by fluorescence microscopy of isolated protoplasts. Whereas PsaE::YFP as well as YFP, which was expressed alone as a control, showed a clear cytoplasmic localization (Figures 6B and 6C), crTP::YFP localized to the plastid (Figure 6D). Within the plastid, crTP::YFP fluorescence was observed over the envelope as well as the stroma as indicated by the complementary pattern of YFP signal and chlorophyll autofluorescence of the thylakoid grana stacks. Thus, the N-terminal crTP surprisingly seems to direct YFP to the chloroplast and mediate its translocation across the chloroplast envelop membranes. To test whether the plastid-targeting capacity found in the flavin reductase crTP was a general crTP feature, we expressed two further crTPs in the tobacco system. We chose one "typical" crTP (derived from scaffold2991-m.31974) and one unusually short crTP that was only predicted by the SVM, but not the HMM approach (derived from scaffold4793-m.44132; Figure 3B, sequences 12 and 24, respectively). Expression of both constructs in tobacco resulted in plastid-localized YFP fluorescence (Figure S5), suggesting that the plastid-targeting capacity is a general feature of crTPs.

DISCUSSION

The Chromatophore Proteome Dataset

The aim of this study was to determine the proteome of the chromatophore, a unique early-stage primary photosynthetic organelle. Based on the coverage of chromatophore-encoded proteins in our MS analysis, we extrapolate that our MS data cover roughly half of the total chromatophore proteome (and more than one quarter with HIC). Purity of the chromatophore isolates was confirmed by the absence of marker proteins for other cell compartments and nearly all of the 100 most highly abundant nuclear-encoded proteins (Table S3, green section). Because the chromatophores take up a considerable space within *P. chromatophora* cells (see Figure S1B), a moderate enrichment of a specific protein in chromatophores compared to whole-cell lysates (as indicated by $Int_C - Int_L > 0$) was considered sufficient to classify a nuclear-encoded protein that was identified in chromatophores as true import candidate. Using this criterion, we identified two classes of chromatophore-targeted proteins, the "short" and "long" import candidates.

Short Import Candidates

For the short import candidates, no common sequences features could be identified that might function as chromatophore import signal. Nevertheless, the set of proteins in this group appears to be not random but contains mainly two classes of short proteins, for which chromatophore localization is plausible based on their predicted cellular function: (1) there are several photosynthesis-related and photoprotective proteins (PsaE, PsaK, and 4 HLI proteins), and (2) the largest group of short import candidates consists of AMP-like proteins. AMPs are effectors of the innate immunity in eukaryotes that kill potentially harmful bacteria through formation of transmembrane pores, membrane rupture, or self-translocation across bacterial membranes followed by the attack of intracellular targets [19]. In members of the amoebozoa, similar proteins were found to be discharged into phagosomes mediating the lysis of prey bacteria [20]. However, recently AMP-like proteins are increasingly recognized to play a role in endosymbiosis [21]. These "symbiotic AMPs" have been described from phylogenetically unrelated endosymbiotic systems (plants and insects). In both systems, endosymbionts are confined to specific symbiont-housing organs. Many symbiotic AMPs are selectively expressed in these symbiotic organs and targeted to the endosymbionts, where they can inhibit cell division, induce morphological and physiological transformations, or are involved in confining the symbiont to the symbiotic organ [22–24]. Hence, the AMP-like short import candidates might derive from short proteins that were originally involved in cell defense or lysis of prey bacteria in the phagotrophic *P. chromatophora* ancestor [3] and evolved novel functions involved in the control of chromatophore growth, division, physiology, and/or metabolite transport [21].

Long Import Candidates

The identified long import candidates provide the first experimental evidence that, through the import of nuclear-encoded metabolic enzymes, chimeric metabolic pathways are completed within the chromatophore. The crTP, found in all HIC and most LC long import candidates, indicates that a common targeting and import mechanism is used by long proteins to enter the chromatophore. The 10 LC import candidates without crTP might be imported into the chromatophore through a crTP-independent pathway or co-purify with the chromatophore due to protein/protein interactions with chromatophore outer membrane proteins. In plastids, most outer envelope proteins lack cTPs as well [25]. Also, some of the nuclear-encoded proteins that were identified, but not enriched, in chromatophores might specifically interact with the chromatophore. One example is polyubiquitin. In plastids and mitochondria, ubiquitination of outer-membrane proteins is involved in the control of the outer-membrane proteome composition [26, 27]. A similar function of ubiquitin is conceivable for the chromatophore.

Origin of the crTP and Import Mechanism

Because crTPs do not show significant sequence similarity to protein sequences in public databases or any conserved domains, the origin of the crTP remains elusive. The fact that the crTP shows conserved sequence motifs excludes the possibility

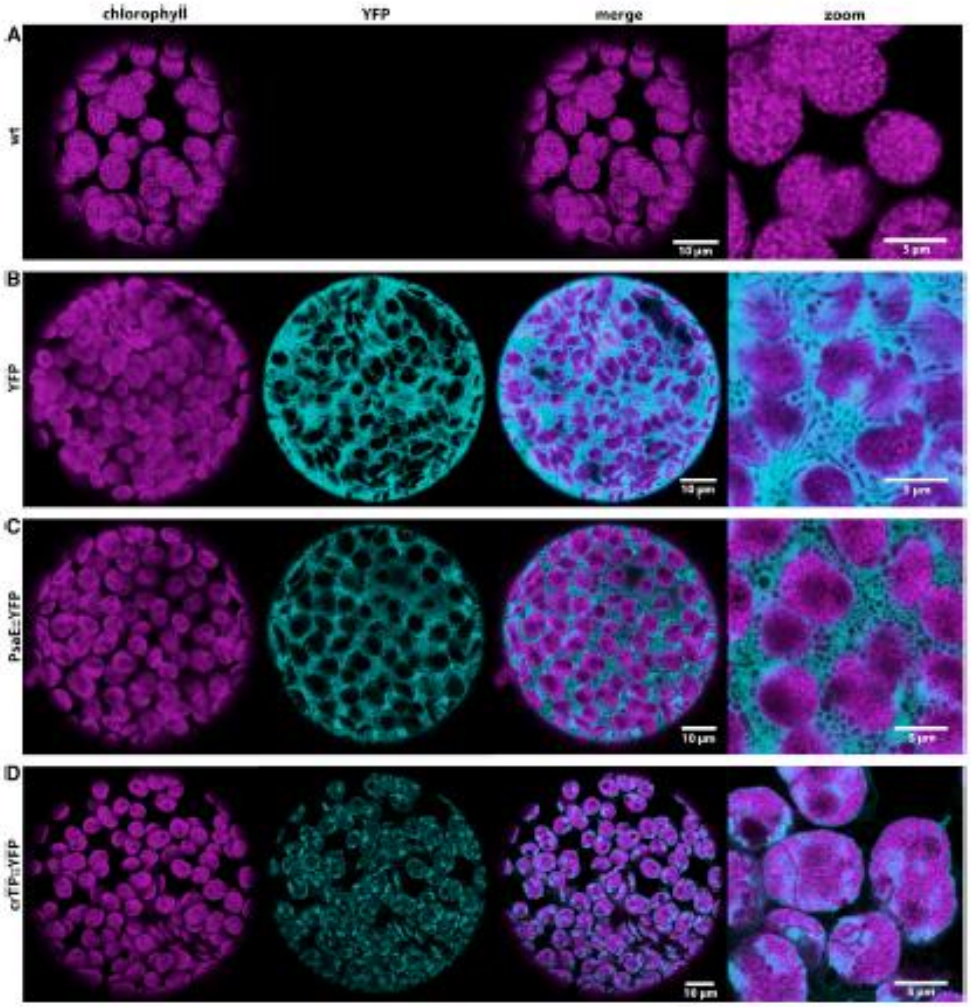

Figure 6. *P. chromatophora* crTP Targets Protein to the Chloroplast in Tobacco

(A) Wild-type *N. benthamiana*.

(B–D) YFP alone (B) and PsaE (C) as well as the crTP (D, from scaf5513-m.48594) fused to the N terminus of YFP were transiently expressed in *N. benthamiana* for 72 hr. Protoplasts were isolated and examined by confocal fluorescence microscopy. Shown are chlorophyll autofluorescence (chlorophyll) in magenta, YFP fluorescence (YFP) in cyan, and an overlay of the two fluorescence pictures (merge). Heterologous expression was performed three times independently for each construct. One representative protoplast for each construct is shown.

See also Figure S5.

that it evolved from random stretches of DNA that are transcribed and translated, as has been proposed for other organellar transit peptides [28]. Interestingly, cTPs and mTPs have recently been suggested to derive from AMPs [29]. This proposal is based on the facts that AMPs and TPs have structural similarities, both have the capacity to interact with bacterial membranes, and antimicrobial activity has been reported for some TPs. According to this proposal, an ingested organelle precursor might have resisted host AMPs through uptake of the AMPs using a specific transporter and their intracellular proteolytic degradation. Re-arrangements in the host genome led then to the fusion of AMP genes with other nuclear genes, resulting in gene products that are competent for import into the evolving organelle [29]. The *P. chromatophora* crTP does not show sequence similarity to the putative symbiotic AMPs identified in this study. However, AMPs are a very diverse class of short proteins that are defined by their physiological activity. Thus, experimental work will be required to test whether or not crTPs show antimicrobial activity.

So far, the interpretation of the crTP as a chromatophore-targeting sequence is based on correlation of chromatophore localization of nuclear-encoded proteins with occurrence of the

conserved N-terminal crTP sequence. Which role the crTP plays for import into the chromatophore, whether it is cleaved off after protein translocation, whether cytosolic factors, components of the endomembrane system, or molecular chaperones within the chromatophore are involved in the import process, is currently unknown. Also, the nature of the transport complex that presumably enables translocation across the chromatophore envelope membranes remains to be determined. Nevertheless, although protein localization studies in heterologous systems have to be interpreted with caution [30], the finding that N-terminal crTPs lead to import of crTP::YFP constructs into tobacco chloroplasts corroborates our interpretation of this sequence as a chromatophore transit peptide and suggests common components (e.g., the use of common components in the protein translocons that derive from conserved cyanobacterial proteins, such as the TIC21 ortholog identified in the chromatophore genome [14]) or principles (e.g., membrane contact through the formation of amphiphilic helices in the transit peptides [31–33]) in chromatophore and chloroplast protein import pathways. However, because in *P. chromatophora* no ortholog of the cyanobacterium-derived plastid outer-membrane translocon pore TOC75 was identified and crTPs are noticeably distinct in length and sequence features from cTPs, it is not entirely clear whether the observed chloroplast-targeting capacity of the crTP is indeed a result of TIC/TOC-mediated import [34] or an alternative import pathway, e.g., involving the secretory pathway [35]. Further experimental work will be required to understand the molecular basis for crTP-mediated protein targeting into chloroplasts. The targeting mechanism for short import candidates, such as PsaE, into the chromatophore seems to be either not functional in planta or disturbed by the fusion of the much larger YFP to the C terminus of PsaE.

In Silico Prediction of the Chromatophore Proteome

Through developing in silico predictions of chromatophore-targeted proteins, we generated a comprehensive map of putatively chromatophore-localized proteins. Predicted and experimentally determined import candidates contribute to chromatophore-localized processes, such as biosynthesis of amino acids, co-factors and nucleotides, genetic information processing, detoxification, and protection against reactive oxygen species. The finding that a large group of import candidates encodes proteins annotated as transcription factors suggests that, similar to the situation in mitochondria and plastids, nuclear-encoded proteins evolved into factors able to exert transcriptional control over chromatophore genes. Furthermore, nuclear factors, such as the identified TIC32 ortholog, might be involved in the control of protein import into the chromatophore. However, identity, composition, and way of functioning of a putative chromatophore protein translocon will have to be determined. The clear over-representation of plastid-associated GO terms in the chromatophore-targeted proteins underlines the strongly convergent evolution between plastids and the chromatophore.

The Chromatophore Proteome Re-arrangement Process

Interestingly, only few of the import candidates identified were of EGT origin, more were of HGT origin, but most seemed to derive from the ancestral host cell [9–11]. Thus, the re-targeting of EGT-derived proteins appears to play only a minor role at the onset of chromatophore integration, and it is initially mainly the import of host proteins that replaces functions lost from the chromatophore genome or adds new functions. The finding of HGT-derived, chromatophore-targeted proteins supports a previously proposed scenario in which, during a mixotrophic evolutionary state, the reduction of the endosymbiont genome drives the fixation of horizontally acquired compensatory genes from prey bacteria [9].

We found several instances where both a nuclear and a chromatophore gene encode the same enzymatic function, with both gene products being localized to the chromatophore. This configuration would allow one copy of the gene to acquire deleterious mutations and eventually be erased and has been postulated as gene transfer intermediate [36]. The nuclear-encoded cyanobacterial CysK proteins in *P. chromatophora*, with one isoform carrying a crTP and the others lacking crTPs (Figure 5), illustrates that transferred genes not always attain the targeting information necessary to be relocated into the organelle; instead, gene transfers or subsequent genomic re-arrangements can entail acquisition of other functional domains.

Elements of Parallel Evolution in Other Endosymbiotic Associations

Interestingly, chromatophore integration parallels, in many aspects, integration of diverse non-photosynthetic bacterial endosymbionts that are found in various eukaryotes [37–39]. Although extensive EGT did not occur in the systems analyzed, many endosymbionts underwent pronounced reductive genome evolution, sometimes to levels close to mitochondria and plastids [40]. Similar to the situation in *P. chromatophora*, genes lost from these endosymbiont genomes seem to be functionally compensated for by proteins expressed from the host genome (e.g., [41, 42]). Some of these nuclear genes apparently derive from bacteria other than the endosymbiont via HGT (e.g., [43, 44]). An initial MS-based study suggested that, at least in the *Buchnera*/aphid symbiosis, large-scale protein import into *Buchnera* cells does not occur [45]. However, recent studies found specific short host-encoded proteins to localize to bacterial endosymbionts [23, 24, 46], including a 21-kDa protein that is imported into *Buchnera* cells in aphids and most likely derives from a bacterium via HGT [47]. Extent of protein import—in particular the ability to import larger enzymes—as well as putative protein import routes remain unknown. Nevertheless, these findings, in combination with our results, suggest that the functional replacement of symbiont-encoded proteins through the import of host-encoded proteins (of diverse phylogenetic origins) might be a phenomenon much more widespread than currently recognized and is not necessarily tied to the occurrence of EGT.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.08.010.

AUTHOR CONTRIBUTIONS

E.C.M.N. conceived the study, analyzed the data, and wrote the paper. A.S. performed most of the experimental work and analyzed the data. G.P. and K.S. performed MS analyses. C.M. and S.A.R. implemented the SVM and HMM in silico approaches and performed predictions of chromatophore-targeted proteins. C.V.-C. helped with data analysis. S.H. helped with confocal laser microscopy. V.H. performed experiments leading to Figure S5.

ACKNOWLEDGMENTS

This study was supported by Deutsche Forschungsgemeinschaft grants NO 1090/1-1 (to E.C.M.N.), CRC 1208, project Z01 (to K.S.), and project Z02 (to S.H.) and Consejo Nacional de Ciencia y Tecnología (México) grant "Beca Mixta" (to C.V.-C.). We thank Fabio Facchinelli for advice and provision of plasmids and bacterial strains for the protein localization assay in tobacco and Andreas Weber for helpful comments on the manuscript. The authors acknowledge scientific and technical assistance of the Center for Advance Imaging (CAI) Düsseldorf.

Received: May 29, 2017
Revised: July 12, 2017
Accepted: August 3, 2017
Published: September 7, 2017

REFERENCES

1. Falkowski, P.G., Katz, M.E., Knoll, A.H., Quigg, A., Raven, J.A., Schofield, O., and Taylor, F.J.R. (2004). The evolution of modern eukaryotic phytoplankton. Science *305*, 354–360.
2. Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. *21*, 809–818.
3. Nowack, E.C.M. (2014). *Paulinella chromatophora* - rethinking the transition from endosymbiont to organelle. Acta Soc. Bot. Pol. *83*, 387–397.
4. Delaye, L., Valadez Cano, C., and Pérez Zamorano, B. (2016). How really ancient is *Paulinella chromatophora*? PLoS Curr. *8*, ecurrents.tol.e68a099364bb1a1e129a17b4e06b0c6b.
5. Marin, B., Nowack, E.C.M., Glöckner, G., and Melkonian, M. (2007). The ancestor of the *Paulinella* chromatophore obtained a carboxysomal operon by horizontal gene transfer from a Nitrococcus-like gamma-proteobacterium. BMC Evol. Biol. *7*, 85.
6. Marin, B., Nowack, E.C.M., and Melkonian, M. (2005). A plastid in the making: evidence for a second primary endosymbiosis. Protist *156*, 425–432.
7. Kies, L. (1974). [Electron microscopical investigations on Paulinella chromatophora Lauterborn, a thecamoeba containing blue-green endosymbionts (Cyanelles) (author's transl)]. Protoplasma *80*, 69–89.
8. Nowack, E.C.M., Melkonian, M., and Glöckner, G. (2008). Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. Curr. Biol. *18*, 410–418.
9. Nowack, E.C.M., Price, D.C., Bhattacharya, D., Singer, A., Melkonian, M., and Grossman, A.R. (2016). Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. Proc. Natl. Acad. Sci. USA *113*, 12214–12219.
10. Nowack, E.C.M., Vogel, H., Groth, M., Grossman, A.R., Melkonian, M., and Glöckner, G. (2011). Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. Mol. Biol. Evol. *28*, 407–422.
11. Zhang, R., Nowack, E.C.M., Price, D.C., Bhattacharya, D., and Grossman, A.R. (2017). Impact of light intensity and quality on chromatophore and nuclear gene expression in *Paulinella chromatophora*, an amoeba with nascent photosynthetic organelles. Plant J. *90*, 221–234.
12. Mariappan, M., Matoja, A., Dobosz, M., Bove, E., Hegde, R.S., and Keenan, R.J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. Nature *477*, 61–66.
13. Nowack, E.C.M., and Grossman, A.R. (2012). Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. Proc. Natl. Acad. Sci. USA *109*, 5340–5345.
14. Bodyl, A., Mackiewicz, P., and Stiller, J.W. (2010). Comparative genomic studies suggest that the cyanobacterial endosymbionts of the amoeba *Paulinella chromatophora* possess an import apparatus for nuclear-encoded proteins. Plant Biol (Stuttg) *12*, 639–649.
15. Gagat, P., and Mackiewicz, P. (2014). Protein translocons in photosynthetic organelles of *Paulinella chromatophora*. Acta Soc. Bot. Pol. *83*, 399–407.
16. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.
17. Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. *2*, 953–971.
18. Petersen, B., Petersen, T.N., Andersen, P., Nielsen, M., and Lundegaard, C. (2009). A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct. Biol. *9*, 51.
19. Brogden, K.A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat. Rev. Microbiol. *3*, 238–250.
20. Andrä, J., Herbst, R., and Leippe, M. (2003). Amoebapores, archaic effector peptides of protozoan origin, are discharged into phagosomes and kill bacteria by permeabilizing their membranes. Dev. Comp. Immunol. *27*, 291–304.
21. Mergaert, P., Kikuchi, Y., Shigenobu, S., and Nowack, E.C.M. (2017). Metabolic integration of bacterial endosymbionts through antimicrobial peptides. Trends Microbiol. *25*, p703–p712.
22. Czernic, P., Gully, D., Cartieaux, F., Moulin, L., Guefrachi, I., Patrel, D., Pierre, O., Fardoux, J., Chaintreuil, C., Nguyen, P., et al. (2015). Convergent evolution of endosymbiont differentiation in dalbergioid and inverted repeat-lacking clade legumes mediated by nodule-specific cysteine-rich peptides. Plant Physiol. *169*, 1254–1265.
23. Login, F.H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D., and Heddi, A. (2011). Antimicrobial peptides keep insect endosymbionts under control. Science *334*, 362–365.
24. Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H., et al. (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. Science *327*, 1122–1126.
25. Hofmann, N.R., and Theg, S.M. (2005). Chloroplast outer membrane protein targeting and insertion. Trends Plant Sci. *10*, 450–457.
26. Karbowski, M., and Youle, R.J. (2011). Regulating mitochondrial outer membrane proteins by ubiquitination and proteasomal degradation. Curr. Opin. Cell Biol. *23*, 476–482.

27. Ling, Q., Huang, W., Baldwin, A., and Jarvis, P. (2012). Chloroplast biogenesis is regulated by direct action of the ubiquitin-proteasome system. Science *338*, 655–659.

28. Baker, A., and Schatz, G. (1987). Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. Proc. Natl. Acad. Sci. USA *84*, 3117–3121.

29. Wollman, F.A. (2016). An antimicrobial origin of transit peptides accounts for early endosymbiotic events. Traffic *17*, 1322–1328.

30. Fuss, J., Liegmann, O., Krause, K., and Rensing, S.A. (2013). Green targeting predictor and ambiguous targeting predictor 2: the pitfalls of plant protein targeting prediction and of transient protein expression in heterologous systems. New Phytol. *200*, 1022–1033.

31. Bruce, B.D. (2001). The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. Biochim. Biophys. Acta *1541*, 2–21.

32. Moberg, P., Nilsson, S., Ståhl, A., Eriksson, A.C., Glaser, E., and Måler, L. (2004). NMR solution structure of the mitochondrial F1beta presequence from *Nicotiana plumbaginifolia*. J. Mol. Biol. *336*, 1129–1140.

33. von Heijne, G. (1986). Mitochondrial targeting sequences may form amphiphilic helices. EMBO J. *5*, 1335–1342.

34. Schleiff, E., and Becker, T. (2011). Common ground for protein translocation: access control for mitochondria and chloroplasts. Nat. Rev. Mol. Cell Biol. *12*, 48–59.

35. Villarejo, A., Burén, S., Larsson, S., Déjardin, A., Monné, M., Rudhe, C., Karlsson, J., Jansson, S., Lerouge, P., Rolland, N., et al. (2005). Evidence for a protein transported through the secretory pathway *en route* to the higher plant chloroplast. Nat. Cell Biol. *7*, 1224–1231.

36. Doolittle, W.F. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. *14*, 307–311.

37. Dubilier, N., Bergin, C., and Lott, C. (2008). Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat. Rev. Microbiol. *6*, 725–740.

38. Moya, A., Peretó, J., Gil, R., and Latorre, A. (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat. Rev. Genet. *9*, 218–229.

39. Nowack, E.C.M., and Melkonian, M. (2010). Endosymbiotic associations within protists. Philos. Trans. R. Soc. Lond. B Biol. Sci. *365*, 699–712.

40. McCutcheon, J.P., and Moran, N.A. (2011). Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. *10*, 13–26.

41. Hansen, A.K., and Moran, N.A. (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc. Natl. Acad. Sci. USA *108*, 2849–2854.

42. McCutcheon, J.P., and von Dohlen, C.D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. *21*, 1366–1372.

43. Alves, J.M.P., Klein, C.C., da Silva, F.M., Costa-Martins, A.G., Serrano, M.G., Buck, G.A., Vasconcelos, A.T.R., Sagot, M.F., Teixeira, M.M.G., Motta, M.C.M., and Camargo, E.P. (2013). Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. BMC Evol. Biol. *13*, 190.

44. Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R.P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A.C.C., et al. (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell *153*, 1567–1578.

45. Poliakov, A., Russell, C.W., Ponnala, L., Hoops, H.J., Sun, Q., Douglas, A.E., and van Wijk, K.J. (2011). Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. Mol. Cell. Proteomics *10*, M110.007039.

46. Ribeiro, C.W., Baldacci-Cresp, F., Pierre, O., Larrousse, M., Benyamina, S., Lambert, A., Hopkins, J., Castella, C., Cazareth, J., Alloing, G., et al. (2017). Regulation of differentiation of nitrogen-fixing bacteria by microsymbiont targeting of plant thioredoxin s1. Curr. Biol. *27*, 250–256.

47. Nakabachi, A., Ishida, K., Hongoh, Y., Ohkuma, M., and Miyagishima, S.Y. (2014). Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. Curr. Biol. *24*, R640–R641.

48. Gautier, R., Douguet, D., Antonny, B., and Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific α-helical properties. Bioinformatics *24*, 2101–2102.

49. McFadden, G.I., and Melkonian, M. (1986). Use of Hepes buffer for microalgal culture media and fixation for electron microscopy. Phycologia *25*, 551–557.

50. Poschmann, G., Seyfarth, K., Besong Agbo, D., Klafki, H.W., Rozman, J., Wurst, W., Wiltfang, J., Meyer, H.E., Klingenspor, M., and Stühler, K. (2014). High-fat diet induced isoform changes of the Parkinson's disease protein DJ-1. J. Proteome Res. *13*, 2339–2351.

51. Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics *2008*, 619832.

52. Widiez, T., Symeonidi, A., Luo, C., Lam, E., Lawton, M., and Rensing, S.A. (2014). The chromatin landscape of the moss *Physcomitrella patens* and its dynamics during development and drought stress. Plant J. *79*, 67–81.

53. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B *57*, 289–300.

54. Grefen, C., Donald, N., Hashimoto, K., Kudla, J., Schumacher, K., and Blatt, M.R. (2010). A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. Plant J. *64*, 355–365.

55. Waadt, R., and Kudla, J. (2008). In planta visualization of protein interactions using bimolecular fluorescence complementation (BiFC). CSH Protoc. *2008*, pdb.prot4995.

7 Bibliografía

Barber A, Siver PA, Karis W. 2013. Euglyphid Testate Amoebae (Rhizaria: Euglyphida) from an Arctic Eocene Waterbody: Evidence of Evolutionary Stasis in Plate Morphology For Over 40 Million Years. Protist. 164:541–555. doi: 10.1016/j.protis.2013.05.001.

Belda E, Silva FJ, Peretó J, Moya A. 2012. Metabolic networks of Sodalis glossinidius: A systems biology approach to reductive evolution. PLoS One. 7. doi: 10.1371/journal.pone.0030652.

Berney C, Pawlowski J. 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proc. R. Soc. B Biol. Sci. 273:1867–1872. doi: 10.1098/rspb.2006.3537.

Bhattacharya D et al. 2012. Single cell genome analysis supports a link between phagotrophy and primary plastid endosymbiosis. Sci. Rep. 2:1–8. doi: 10.1038/srep00356.

Bhattacharya D, Archibald JM, Weber APM, Reyes-Prieto A. 2007. How do endosymbionts become organelles? Understanding early events in plastid evolution. BioEssays. 29:1239–1246. doi: 10.1002/bies.20671.

Bhattacharya D, Yoon HS, Hackett JD. 2004. Photosynthetic eukaryotes unite: Endosymbiosis connects the dots. BioEssays. 26:50–60. doi: 10.1002/bies.10376.

Bromham L, Penny D. 2003. The modern molecular clock. Nat. Rev. Genet. 4:216–224. doi: 10.1038/nrg1020.

Burgard AP, Pharkya P, Maranas CD. 2003. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. Biotechnol. Bioeng. 84:647–657. doi: 10.1002/bit.10803.

Canbäck B, Tamas I, Andersson SGE. 2004. A phylogenomic study of endosymbiotic bacteria. Mol. Biol. Evol. 21:1110–1122. doi: 10.1093/molbev/msh122.

Culver SJ. 1991. Early Cambrian Foraminifera from West Africa. Science (80-.). 254:689–691.

Damsté JSS, Moldowan JM, Barbanti SM, Trindade LAF, Schouten S. 2004. The rise of the rhizosolenid diatoms. Science (80-.). 584:584–587. doi: 10.1126/science.1096806.

Dimijian GG. 2000. Evolving together: the biology of symbiosis, part 1. BUMC Proc. 13:217–226.

Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. 101:15386–15391. doi: 10.1073/pnas.0403984101.

Garg SG, Gould SB. 2016. The Role of Charge in Protein Targeting Evolution. Trends Cell Biol. 26:894–905. doi: 10.1016/j.tcb.2016.07.001.

González-Domenech CM et al. 2012. Metabolic stasis in an ancient symbiosis: genome-scale

metabolic networks from two Blattabacterium cuenoti strains, primary endosymbionts of cockroaches. BMC Microbiol. 12 Suppl 1:S5. doi: 10.1186/1471-2180-12-S1-S5.

Holman AG, Davis PJ, Foster JM, Carlow CKS, Kumar S. 2009. Computational prediction of essential genes in an unculturable endosymbiotic bacterium , Wolbachia of Brugia malayi. BMC Microbiol. 9:1–14. doi: 10.1186/1471-2180-9-243.

Johnson MD. 2011. The acquisition of phototrophy: Adaptive strategies of hosting endosymbionts and organelles. Photosynth. Res. 107:117–132. doi: 10.1007/s11120-010-9546-8.

Khannapho C et al. 2008. Selection of objective function in genome scale flux balance analysis for process feed development in antibiotic production. Metab. Eng. 10:227–233. doi: 10.1016/j.ymben.2008.06.003.

Kies L, Kremer BP. 1979. Function of cyanelles in the thecamoeba Paulinella chromatophora. Naturwissenschaften. 66:578–579. doi: 10.1007/BF00368819.

Kim S, Park MG. 2016. Paulinella longichromatophora sp. nov., a New Marine Photosynthetic Testate Amoeba Containing a Chromatophore. Protist. 167:1–12. doi: 10.1016/j.protis.2015.11.003.

Kimura M. 1984. The Neutral Theory of Molecular Evolution. (Cambridge Univ Press. New York). doi: 10.1038/scientificamerican1179-98.

Kooistra WHCF, Gersonde R, Medlin LK, Mann DG. 2007. *The Origin and Evolution of the Diatoms : Their Adaptation to a Planktonic Existence*. Elsevier Inc. doi: 10.1016/B978-0-12-370518-1.50012-6.

Lane CE. 2007. Bacterial Endosymbionts: Genome Reduction in a Hot Spot. Curr. Biol. 17:508–510. doi: 10.1016/j.cub.2007.04.032.

Lhee D et al. 2017. Diversity of the Photosynthetic Paulinella Species, with the Description of Paulinella micropora sp. nov. and the Chromatophore Genome Sequence for strain KR01. Protist. 168:155–170. doi: 10.1016/j.protis.2017.01.003.

Marin B, Nowack ECM, Melkonian M. 2005. A plastid in the making: Evidence for a second primary endosymbiosis. Protist. 156:425–432. doi: 10.1016/j.protis.2005.09.001.

McCutcheon JP, Moran N a. 2011. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10:13–26. doi: 10.1038/nrmicro2670.

Moorjani P, Amorim CEG, Arndt PF, Przeworski M. 2016. Variation in the molecular clock of primates. Proc. Natl. Acad. Sci. 113:10607–10612. doi: 10.1073/pnas.1600374113.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93:2873–2878. doi: 10.1073/pnas.93.7.2873.

Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr. Opin. Genet. Dev. 14:627–633. doi: 10.1016/j.gde.2004.09.003.

Nakayama T, Archibald JM. 2012. Evolving a photosynthetic organelle. BMC Biol. 10:35–37. doi: 10.1186/1741-7007-10-35.

Nakayama T, Ishida K ichiro. 2009. Another acquisition of a primary photosynthetic organelle is underway in Paulinella chromatophora. Curr. Biol. 19:284–285. doi: 10.1016/j.cub.2009.02.043.

Nomura M, Nakayama T, Ishida KI. 2014. Detailed process of shell construction in the photosynthetic testate amoeba Paulinella chromatophora (euglyphid, rhizaria). J. Eukaryot. Microbiol. 61:317–321. doi: 10.1111/jeu.12102.

Nowack ECM et al. 2011. Endosymbiotic gene transfer and transcriptional regulation of transferred genes in Paulinella chromatophora. Mol. Biol. Evol. 28:407–422. doi: 10.1093/molbev/msq209.

Nowack ECM. 2014. Paulinella chromatophora - Rethinking the transition from endosymbiont to organelle. Acta Soc. Bot. Pol. 83:387–397. doi: 10.5586/asbp.2014.049.

Nowack ECM, Grossman AR. 2012. Trafficking of protein into the recently established photosynthetic organelles of Paulinella chromatophora. Proc. Natl. Acad. Sci. U. S. A. 109:5340–5345. doi: 10.1073/pnas.1118800109.

Nowack ECM, Melkonian M, Glöckner G. 2008. Chromatophore Genome Sequence of Paulinella Sheds Light on Acquisition of Photosynthesis by Eukaryotes. Curr. Biol. 18:410–418. doi: 10.1016/j.cub.2008.02.051.

Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? Nat Biotech. 28:245–248. doi: 10.1038/nbt.1614.

Pál C et al. 2006. Chance and necessity in the evolution of minimal metabolic networks. Nature. 440:667–670. doi: 10.1038/nature04568.

Parfrey L, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. 108:13624–13629. doi: 10.1073/pnas.1110633108.

Pickens LB, Tang Y, Chooi Y-H. 2011. Metabolic Engineering for the Production of Natural Products. Annu. Rev. Chem. Biomol. Eng. 2:211. doi: 10.1146/annurev-chembioeng-061010-114209.

Porter SM, Knoll AH. 2000. Testate Amoebae in the Neoproterozoic Era: Evidence from Vase-shaped Microfossils in the Chuar Group, Grand Canyon vase-shaped microfossils in the Chuar Group, Grand Canyon. 26:360–385. doi: 10.1666/0094-8373(2000)026<0360.

Reyes-Prieto A et al. 2010. Differential gene retention in plastids of common recent origin. Mol. Biol. Evol. 27:1530–1537. doi: 10.1093/molbev/msq032.

Reyes-Prieto A, Weber APM, Bhattacharya D. 2007. The origin and establishment of the plastid in algae and plants. Annu. Rev. Genet. 41:147–168. doi: 10.1146/annurev.genet.41.110306.130134.

Russell CW et al. 2014. Matching the supply of bacterial nutrients to the nutritional demand of the animal host. Proc. R. Soc. B Biol. Sci. 281:20141163. doi: 10.1098/rspb.2014.1163.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature. 407:81–86. doi: 10.1038/35024074.

Shou W. 2015. Acknowledging selection at sub-organismal levels resolves controversy on pro-cooperation mechanisms. Elife. 4. doi: 10.7554/eLife.10106.

Su Q, Zhou X, Zhang Y. 2013. Symbiont-mediated functions in insect hosts. Commun. Integr. Biol. doi: 10.4161/cib.23804.

Thomas GH et al. 2009. A fragile metabolic network adapted for cooperation in the symbiotic bacterium Buchnera aphidicola. BMC Syst. Biol. 3. doi: 10.1186/1752-0509-3-24.

Valdivia RH et al. 2007. Endosymbiosis: the evil within. Curr. Biol. 17:R408-10. doi: 10.1016/j.cub.2007.04.001.

Villarejo A et al. 2005. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. Nat. Cell Biol. 7:1224–1231. doi: 10.1038/ncb1330.

Wang Z et al. 2006. Exploring photosynthesis evolution by comparative analysis of metabolic networks between chloroplasts and photosynthetic bacteria. BMC Genomics. 7:100. doi: 10.1186/1471-2164-7-100.

Wernegreen JJ. 2012. Endosymbiosis. Curr. Biol. 22:555–561. doi: 10.1016/j.cub.2012.06.010.

Yoon HS et al. 2009. A single origin of the photosynthetic organelle in different Paulinella lineages. BMC Evol. Biol. 9. doi: 10.1186/1471-2148-9-98.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. Mol. Biol. Evol. 21:809–818. doi: 10.1093/molbev/msh075.

Zhang R et al. 2017. Impact of light intensity and quality on chromatophore and nuclear gene expression in Paulinella chromatophora , an amoeba with nascent photosynthetic organelles. Plant J. 90:221–234. doi: 10.1111/tpj.13488.

Zientz E, Dandekar T, Gross R. 2004. Metabolic Interdependence of Obligate Intracellular Bacteria and Their Insect Hosts. Microbiol. Mol. Biol. Rev. 68:745–770. doi: 10.1128/MMBR.68.4.745.