


RESEARCH ARTICLE

Open Access



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

Cecilio Valadez-Cano¹, Roberto Olivares-Hernández², Osbaldo Resendis-Antonio^{3,4}, Alexander DeLuna⁵ 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

¹Departamento 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



(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 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

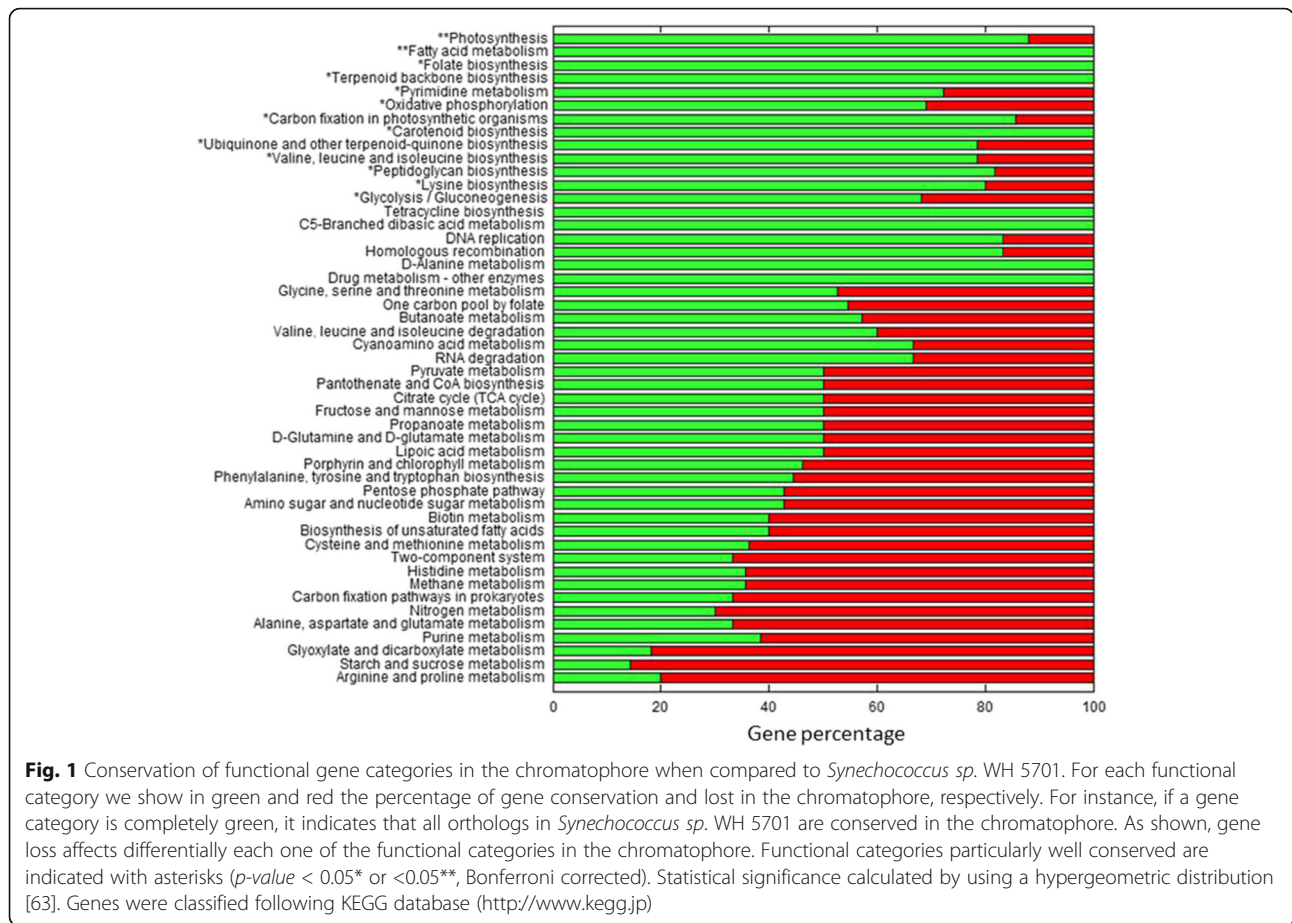
free-living relative of the chromatophore, it could be considered that 158 reactions were lost along genome reduction in the chromatophore (Table 1).

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. CO₂ 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



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

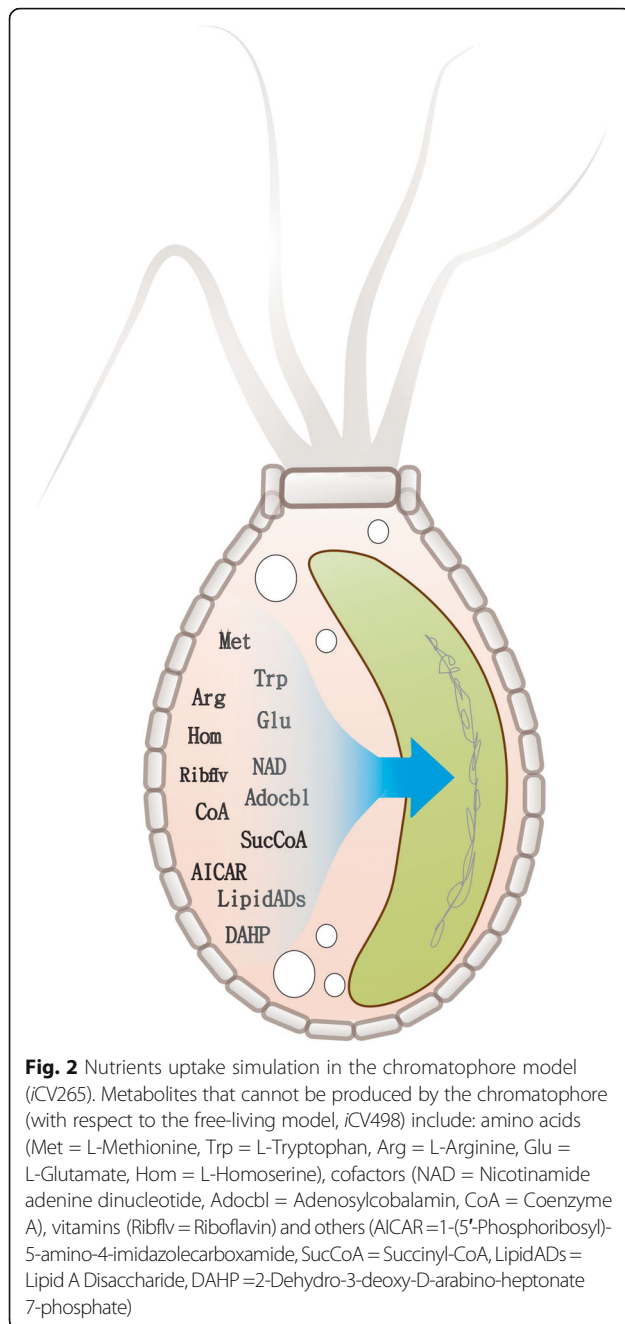
3-isopropylmalate dehydrogenase. In this case, we assumed that host encoded enzymes complement the pathway in the endosymbiont. Either by importing host encoded enzymes to the chromatophore or by exchanging intermediated metabolites between the symbionts. Similar situations have been proposed for other 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>i</i> CV498	<i>i</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



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

Robustness analysis of metabolic models

We assessed the robustness of the *iCV498* and the *iCV265* 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 *iCV498* showed 333 genes (66.86%) to be essential because its deletion decreases

the biomass production over 99% (Fig. 3). This result shows that *iCV498* 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 *iCV498* 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 *iJN678* when it is deleted (data not shown). In the *iCV265* model, all these 3 genes were lost. In addition, the non-essential gene in *iCV498* coding for an enzyme with arginase activity is the only one whose deletion decreases the growth rate in the *iCV265* chromatophore model. This suggests that genome reduction leading to *iCV265* caused metabolic restructuring because deletion of this enzyme with arginase activity in *iCV498* has no effect.

In silico simulation of metabolic-gene losses 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 *iCV265* 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 *iCV265* model (we refer to them as Proposed Nutrients) (Fig. 2). In the second one, we randomly selected metabolites from the *iCV498* 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 *iCV265*; but differing in their in silico evolutionary history and gene content.

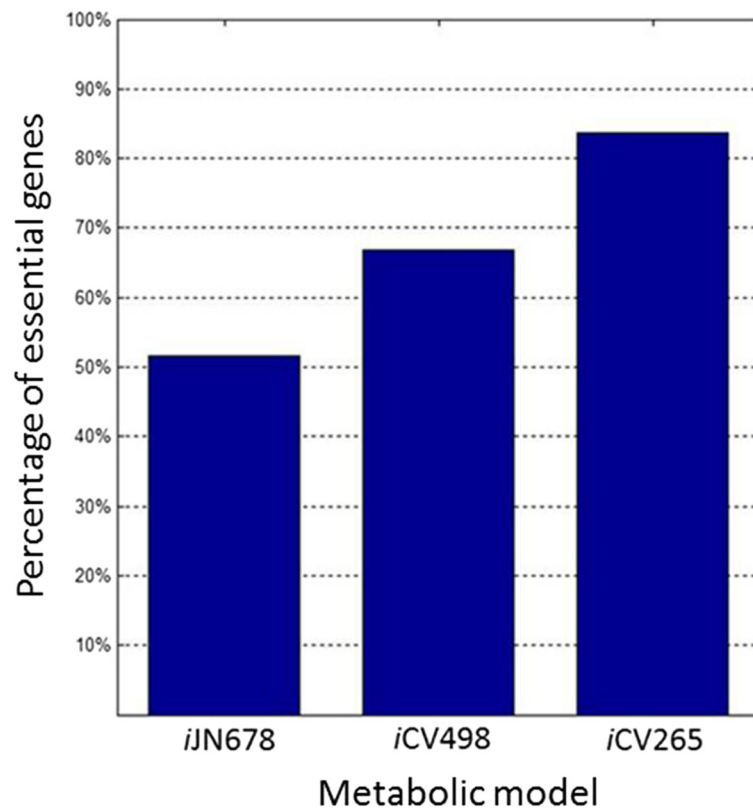


Fig. 3 Genetic robustness analysis of metabolic models of *Synechocystis* sp. PCC 6803 (*iJN678*) [26], *Synechococcus* sp. WH 5701 (*iCV498*) and the chromatophore of *Paulinella chromatophora* (*iCV265*). Percentage of essential genes (Y axis) for each metabolic model (X axis)

Simulations under the Proposed Nutrients conditions resulted in reduced metabolic networks with 295.1 (\pm 2.63) genes on average. In these reduced networks, of the 498 genes present in the free-living ancestor (model *iCV498*), 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 (\pm 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.

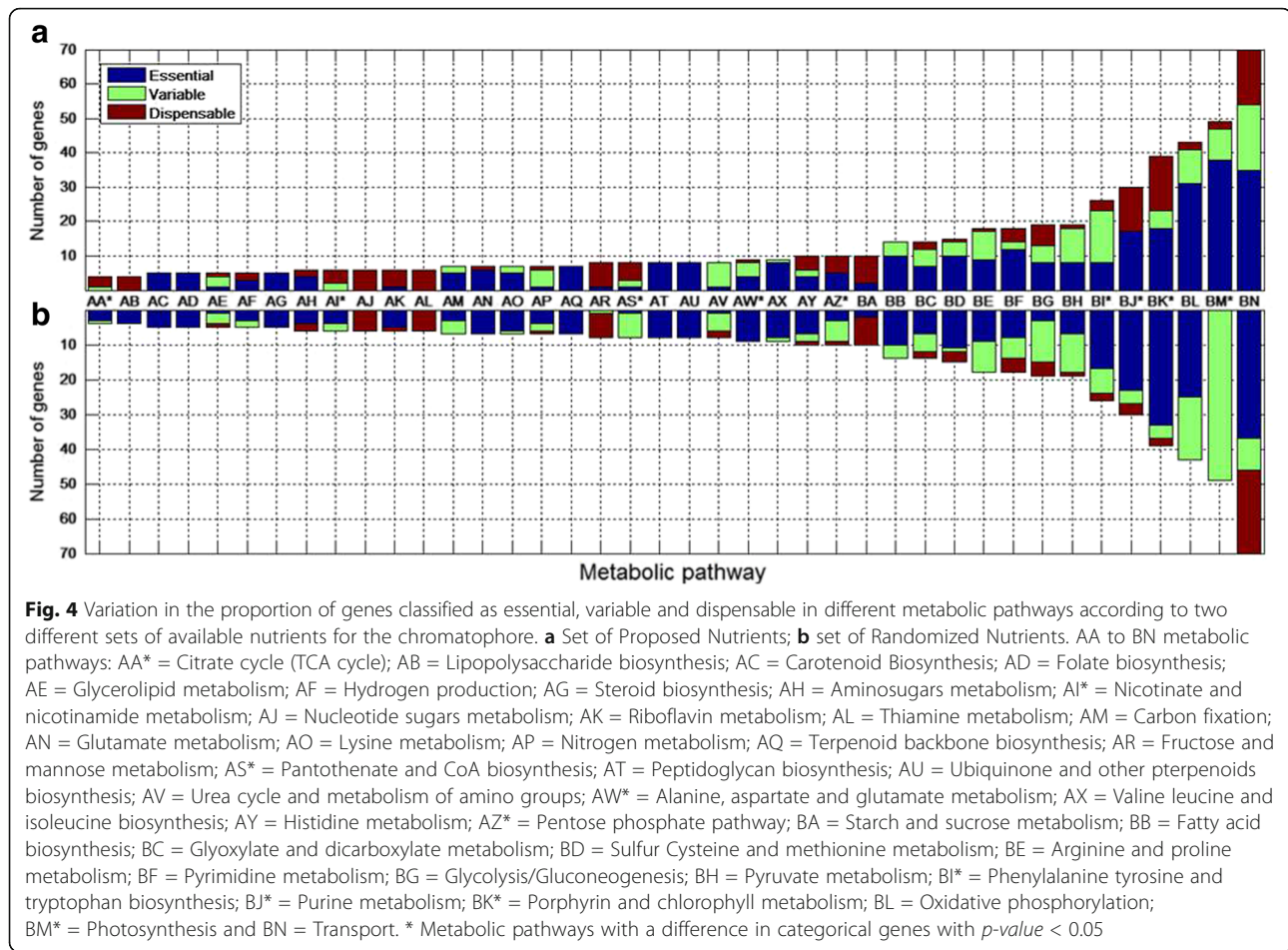
Genetic concordance was evaluated between these simulated minimal networks and the *real* chromatophore

model (*iCV265*). 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 *iCV265* 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,



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].

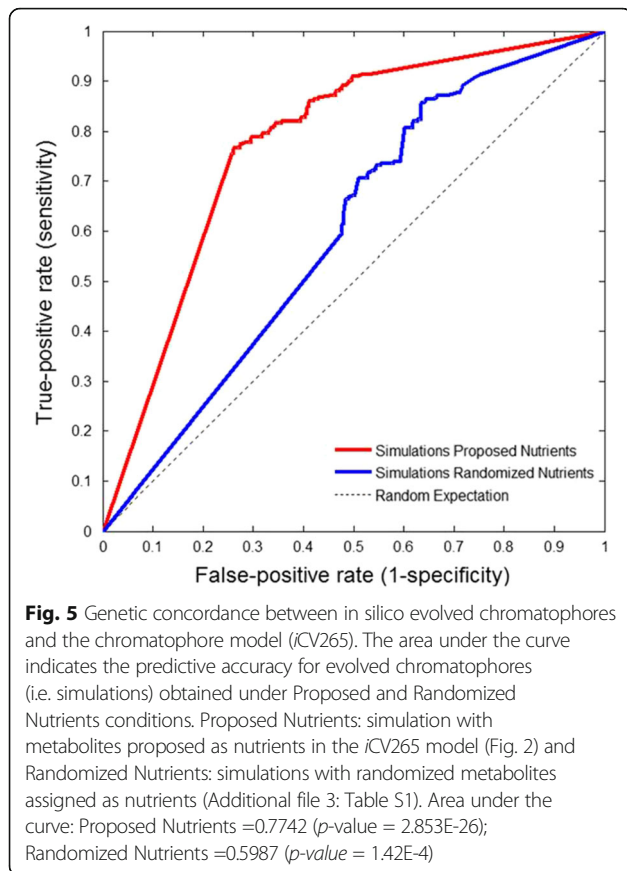
To study how the potential rate of carbon exportation evolved, a hexose export reaction was added to the metabolic models. This reaction was defined as objective function. To ensure biomass components production, the biomass reaction was fixed to 0.0884 h^{-1} which, as stated previously, is the growth 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 *iCV498* model. However, in the *iCV265* chromatophore model, the potential rate of hexose exported without affecting the in silico growth rate was $0.2689 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$. 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



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 *iCV265* model (the potential rate of hexose exportation of $0.2689 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$) 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 test 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).

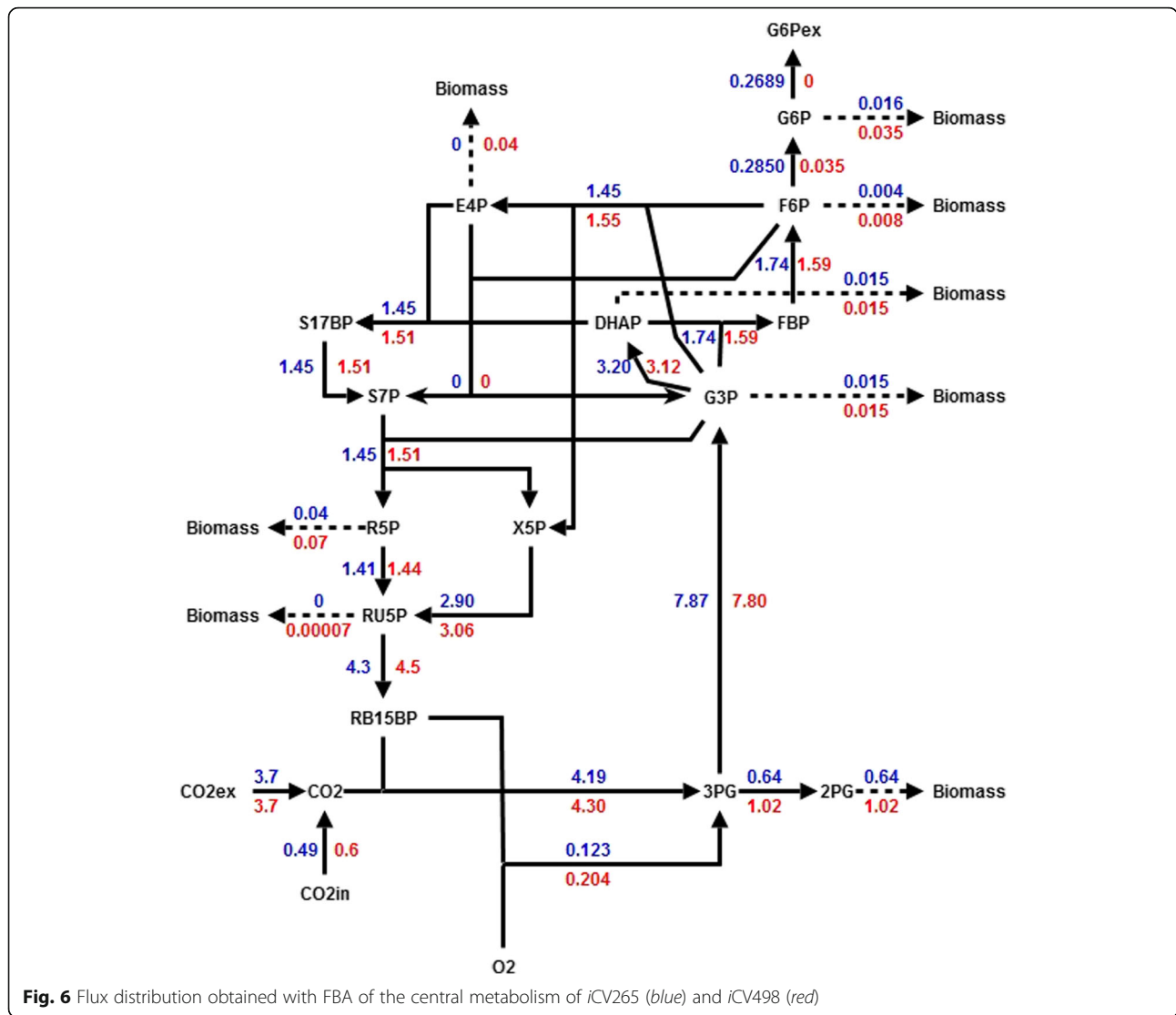
Minimal networks obtained in silico under these conditions were variable in size and gene content. Of 500 simulations, only 175 (66.03%) of the genes conserved in model *iCV265* are conserved in all the simulations. Conversely, there are 45 genes predicted to be essential in all these 500 simulations which are not conserved in model *iCV265*. 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 ($0.2689 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$). This suggests that the probability of obtaining a potential rate of hexose exported similar to that of *iCV265* 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 *iCV265*. 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 *iCV265* 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 ($0.2689 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$) 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 *iCV265* 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 *iCV265*). 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 (*iCV265*). 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).



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].

Table 2 Yields ($Y_{P/S}$) analysis of extracellular CO₂, ATP and NADPH consumed in hexose production for both models

	Metabolic model	
	<i>iCV265</i>	<i>iCV498</i>
	mmol Hexose/ mmol	mmol Hexose/ mmol
CO ₂	0.069	0.009
ATP	0.023	0.002
NADPH	0.036	0.004

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 *iCV498* as a free-living ancestor of the chromatophore under nutrient-rich conditions (Fig. 2). However, in this experiment, the model *iCV498* 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 *iCV498*. If the deleted gene affected the growth rate (0.0884 h⁻¹) or impaired hexose

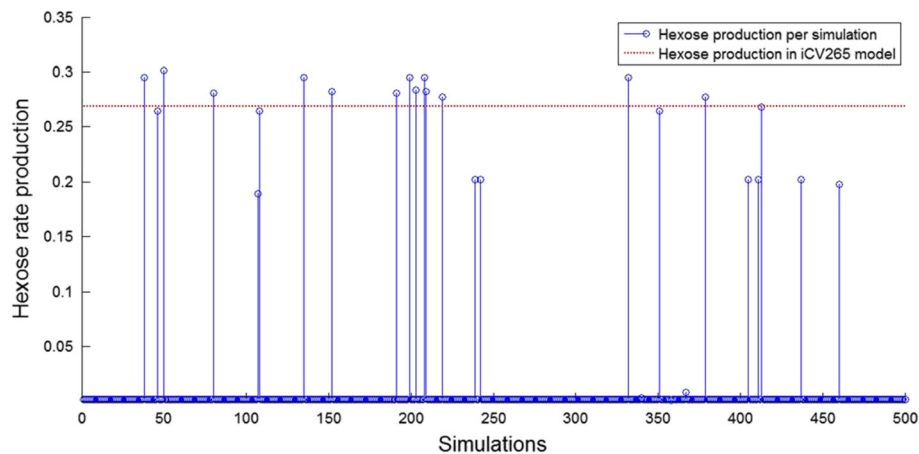


Fig. 7 Hexose exportation rate of the chromatophore model (*iCV265*) is achieved in only 2.6% of the simulations under a random model of evolution. Hexose exported rate (Y axis) for 500 independent simulations (X axis). Red-dotted line indicates the hexose export rate in the chromatophore model (*iCV265*)

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 ($0.2689 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$) 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

iCV265. 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 (*iCV265*) 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 *iCV265* and *iCV498* 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 *iCV498* 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 (*iCV265*) 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 *iCV265* model for a determined growth rate restriction is greater than the *iCV498* 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

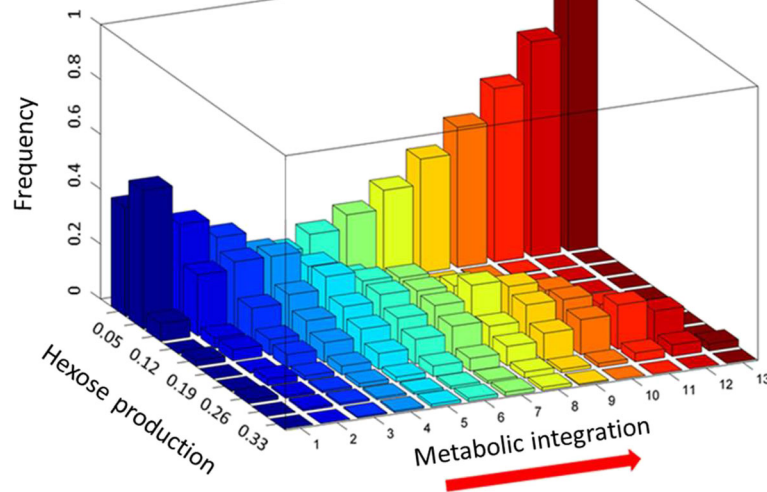


Fig. 8 Metabolic integration with the host determines the rate of hexose exported by in silico chromatophores as well as the frequency of simulations that provide higher rates of reduced-carbon to the host. X axis, hexose export rate; Y axis, frequency of simulations; Z axis, transporters added to the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

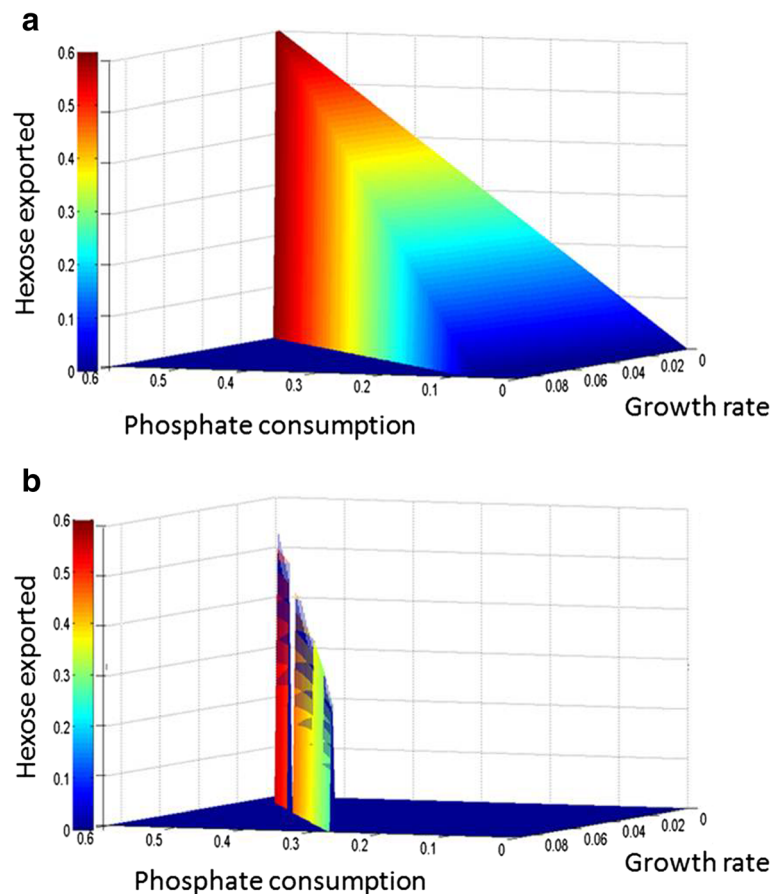


Fig. 9 The metabolic model of the chromatophore (*iCV265*) is specially adapted to produce higher rates of reduced-carbon under phosphate and growth restrictions in comparison with *Synechococcus sp.* WH 5701 (*iCV498*). **a** Hexose export rate in the free-living model *iCV498*. **b** Hexose export rate in the chromatophore model *iCV265*. 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 *iCV265* and the *iCV498* 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 glossinidia* [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 derive 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 $0.2689 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$. 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 $0.0884 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$; 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 $0.0884 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$ 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 *iCV265* ($0.2689 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$) 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 *iCV265*. This suggests that metabolic functionality of *iCV265* 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 (*iCV265*) seems to be tailored for the exportation of reduced-carbon; that is, when comparing the export of reduced-carbon between the *iCV265* and the *iCV498* models (with phosphate as restrictive nutrient) we found that *iCV265* shows higher rates of hexose exported than the free-living *iCV498* 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 *iCV498* and the *iCV265* 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 (*iJN678*) [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 CO₂ and

photons uptake set to $3.7 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$ and $100 \text{ mmol} \times \text{gDW}^{-1} \times \text{h}^{-1}$, 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 considered 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 (*iCV498*) 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 *iCV498* 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 ($\geq 0.0884 \text{ h}^{-1}$), 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 (*iCV265*) was analyzed as in [30]. In each of the 500 simulations, a binary variable was assigned for each gene in *iCV498* 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 *iCV265*) and false-positives (fraction of genes predicted by the simulations and not present in *iCV265*) for every cutoff (minimal fraction of simulated genomes in which a gene must be present to be predicted as

conserved in *iCV265*). 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 *iCV265*. 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 *iCV498* 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 (*iCV498*). (XLSX 187 kb)

Additional file 2: Metabolic model of the chromatophore (*iCV265*). (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

¹Departamento 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. ²Departamento 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. ³Human Systems Biology Laboratory, Coordinación de la Investigación Científica-Red de Apoyo a la Investigación (RAI), UNAM, México City, Mexico. ⁴Instituto Nacional de Medicina Genómica (INMEGEN), 14610 México City, Mexico. ⁵Unidad 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

- Nakayama T, Archibald JM. Evolving a photosynthetic organelle. *BMC Biol.* 2012;10:35–7.
- Nowack ECM. Paulinella chromatophora - Rethinking the transition from endosymbiont to organelle. *Acta Soc Bot Pol.* 2014;83:387–97.
- 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.
- Marin B, Nowack ECM, Melkonian M. A plastid in the making: Evidence for a second primary endosymbiosis. *Protist.* 2005;156:425–32.
- 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.
- 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.e68a099364bb1a1e129a17b4e06b0c6b.
- 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.
- 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.
- 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>.
- Kies L, Kremer BP. Function of cyanelles in the thecamoeba Paulinella chromatophora. *Naturwissenschaften.* 1979;66:578–9.
- 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>.
- 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.
- Lee MC, Marx CJ. Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genet.* 2012;8:2–9.
- 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>.
- 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/25713367>.
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. Lüring 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 glossiniidius*: 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;201421375 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 pipiensis* 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, *Acyrtosiphon 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, Puchałka 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. Puchałka J, Oberhardt MA, Godinho M, Bielecka A, Regenhart 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, Omolo MA, Reiland 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> <http://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) Curvel 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

