

Protein translocons in photosynthetic organelles of *Paulinella chromatophora*

Przemysław Gagat, Paweł Mackiewicz*

Department of Genomics, Faculty of Biotechnology, University of Wrocław, Fryderyka Joliot-Curie 14a, 50-383 Wrocław, Poland

Abstract

The rhizarian amoeba *Paulinella chromatophora* harbors two photosynthetic cyanobacterial endosymbionts (chromatophores), acquired independently of primary plastids of glaucophytes, red algae and green plants. These endosymbionts have lost many essential genes, and transferred substantial number of genes to the host nuclear genome via endosymbiotic gene transfer (EGT), including those involved in photosynthesis. This indicates that, similar to primary plastids, *Paulinella* endosymbionts must have evolved a transport system to import their EGT-derived proteins. This system involves vesicular trafficking to the outer chromatophore membrane and presumably a simplified Tic-like complex at the inner chromatophore membrane. Since both sequenced *Paulinella* strains have been shown to undergo differential plastid gene losses, they do not have to possess the same set of Toc and Tic homologs. We searched the genome of *Paulinella* FK01 strain for potential Toc and Tic homologs, and compared the results with the data obtained for *Paulinella* CCAC 0185 strain, and 72 cyanobacteria, eight Archaeplastida as well as some other bacteria. Our studies revealed that chromatophore genomes from both *Paulinella* strains encode the same set of translocons that could potentially create a simplified but fully-functional Tic-like complex at the inner chromatophore membranes. The common maintenance of the same set of translocon proteins in two *Paulinella* strains suggests a similar import mechanism and/or supports the proposed model of protein import. Moreover, we have discovered a new putative Tic component, Tic62, a redox sensor protein not identified in previous comparative studies of *Paulinella* translocons.

Keywords: *Paulinella chromatophora*; chromatophores; endosymbiosis; primary plastids; protein import; translocons

Introduction

The primary plastid endosymbiosis is one of the most important transitions in the evolution of life on our planet. It took place sometime before 1.5 billion years ago, when a phagotrophic eukaryote, the ancestor of glaucophytes, red algae and green plants, enslaved a β -cyanobacterium, which was transformed into a two-membrane photosynthetic primary plastid [1,2]. In order to become a true cell organelle, the prokaryote underwent a tremendous transformation that involved two key processes: (i) endosymbiotic gene transfer (EGT), i.e. gene transfer from the endosymbiont to the host nuclear genome, and (ii) origin of an import machinery in its envelope membranes for host genome-encoded proteins, including the products of EGT-derived genes [3–5].

Interestingly, primary plastid endosymbiosis was not a singular event in the history of our biosphere, but has happened at least twice, questioning thereby a paradigm that endosymbiont-to-organelle transformation is an exceptionally

rare evolutionary phenomenon [6]. For the second time, it took place in the case of *Paulinella chromatophora*, a testate filose amoeba, belonging to the supergroup Rhizaria, a lineage that is very distantly related to the primary plastid bearing lineages, i.e. glaucophytes, red algae and green plants, united in the supergroup Archaeplastida [7,8].

Paulinella chromatophora harbors two photosynthetic cyanobacterial endosymbionts (chromatophores), acquired independently of Archaeplastida primary plastids ~60 million years ago [8–11]. Similarly, to primary plastids, chromatophores are surrounded by a two-membrane envelope and are deeply integrated with their host cell. They divide synchronously with *Paulinella* cell (after being distributed to daughter cells), exchange metabolites with it and are incapable of independent life [9,12,13]. They are especially similar in their structure to glaucophyte primary plastids (cyanelles) because they still retain the peptidoglycan wall in the intermembrane space, a clear sign of their bacterial inheritance [12]. The intimacy between the *Paulinella* host and its endosymbiotic bodies is well emphasized by substantial reduction of the latter genome, which was sequenced for two *Paulinella* strains CCAC 0185 [11] and FK01 [14]. They were reduced in size and coding capacity by a factor

* Corresponding author. Email: pamac@smorfland.uni.wroc.pl

Handling Editor: Andrzej Bodyl

of three (to 1 Mb and ~900 genes), compared with their closest free-living relative, the cyanobacterium *Cyanobium gracile* PCC 6307 (~3 Mb and ~3300 genes) [11,14] (Fig. 1). The genome-size reduction involved loss of many genes, including those engaged in essential biosynthetic pathways and endosymbiotic gene transfer [11,14]. Transcriptome analyses have identified more than 30 EGT-derived genes in the *Paulinella* nuclear genome [14–16]; however, many more are expected; Nowack et al. [16] estimated their number between 40 and 125. Because the lost genes are of the utmost importance for cell functioning, and at least some of the transferred ones are transcriptionally regulated by the host and involved in photosynthesis or photosynthesis-related processes, *Paulinella* chromatophores are expected to have evolved a protein import route, or routes.

To check if chromatophore protein import system is similar to the primary plastid translocons at the outer and inner envelope membrane, Toc and Tic, respectively, Bodył et al. [17] searched for homologs of *Toc* and *Tic* genes in the completely sequenced chromatophore genome of *Paulinella* CCAC 0185 strain. They found that the *Paulinella* chromatophore genome encodes homologs to *Toc12*, *Toc64*, *Tic21* and *Tic32*, but lost those of *Toc75*, *Tic20*, *Tic55* and *Tic62*. They suggested that the missing genes, which are still present in the genomes of closely related cyanobacteria, were most probably relocated via the EGT to the host nuclear genome, and that their products are now imported into the photosynthetic bodies' membranes to create a Toc-Tic-like import machinery [17]. Mackiewicz and Bodył [18] also found that an EGT-derived gene encoding the subunit IV of PS I reaction center (*PsaE*) from *Paulinella* FK01 strain has a clear signal peptide predicted from an alternative initiation site [18]. Since signal peptides are involved in protein import into the endoplasmic reticulum, such a presequence in the case of *PsaE* indicates the possibility of endomembrane-dependent targeting to the outer chromatophore membrane. Subsequent studies confirmed the presence of a signal peptide-like presequence in another four of nine investigated EGT-derived proteins that evolved by modification of their N-terminal mature parts, but not the missing Toc and Tic components in the nuclear genome [16,19].

On the basis of their bioinformatics analyses, Mackiewicz et al. [20] formulated a model for protein import into *Paulinella* chromatophores. According to the model, *Paulinella* EGT-derived proteins established endomembrane-dependent mechanism, vesicular trafficking, to pass the outer chromatophore membrane, while the inner membrane is crossed by a simplified Tic-like complex. All investigated EGT-derived proteins, presumed to follow this pathway, characterize low molecular weight and nearly neutral charge, which constitute good adaptations to their passage through the peptidoglycan wall still present in the chromatophores. Mackiewicz et al. [20] also identified additional putative elements of the system, for example chaperons in the intermembrane space and components of the molecular motor responsible for pulling the imported proteins into the chromatophore stroma. It should be noted that part of the model, concerning the vesicular trafficking to the outer chromatophore membrane, has recently been confirmed by Nowack et al. [21]. They used immunogold labeling electron microscopy

to show that *PsaE*, *PsaK1* and *PsaK2* from *Paulinella* CCAC 0185 strain are localized both in chromatophores and the endomembrane system, including the Golgi apparatus. They also indicated that these proteins are indeed PS I subunits as they associate with PS I components. Interestingly, Nowack et al. [21] revealed that the three proteins do not have typical signal peptides at their N-termini. In the case of *PsaE* some internal targeting information was suggested by Mackiewicz et al. [20], whereas, for *PsaK* proteins algorithms recognized uncleavable signal peptides in their N-terminal hydrophobic domains, suggesting that their N-termini might fulfill a signal peptide role in the protein import after all [19].

Both *Paulinella* strains have undergone differential plastid gene losses [14,22] and have even been suggested to represent two different species on the basis of some morphological and phylogenetic analyses [8]. Therefore it is interesting to check if they possess the same set of potential Toc and Tic homologs. The common maintenance of the same set of proteins would indicate a similar import mechanism and/or support the model proposed by Mackiewicz et al. [20]. Therefore, we searched *Paulinella* FK01 strain genome for Toc and Tic homologs, and compared the results with the data obtained for *Paulinella* CCAC 0185 strain, 72 cyanobacteria, eight Archaeplastida, as well as some other bacteria. We carried out more sensitive homology and conserved domain searches than in previous approaches, which is justified by a short length and relatively high divergence of these proteins [17].

Material and methods

We downloaded 867 and 841 sequences of chromatophore-encoded proteins for *Paulinella* CCAC 0185 strain from Genbank [23] and for *Paulinella* FK01 strain from Debashish Bhattacharya's Laboratory website (<http://dmlab.rutgers.edu/home/index.php>), respectively. We also acquired proteomes of 72 completely sequenced cyanobacteria, eight Archaeplastida and two reference bacteria for comparative studies from Genbank [23]. On the basis of the obtained sequences a local protein database was created. Sequences of 16S and 23S rRNA for phylogenetic studies were extracted from gbk files downloaded from GenBank (genome database) for appropriate organisms. This set included sequences, which were derived from two *Paulinella chromatophora* genomes, 72 completely sequenced cyanobacterial genomes, eight primary plastid genomes and five genomes representing different bacterial lineages as an outgroup.

To find potential Toc-Tic components, we searched the database using *Pisum sativum* translocons as query sequences with PsiBlast (*E*-value set to 0.01, word size to two, filtering for low complexity regions on, five iterations) to achieve better sensitivity than with standard BLAST or FASTA [24]. In the case of *Tic21* a homolog from *Arabidopsis thaliana* was used as *Tic21* is absent from *P. sativum* [25]. The obtained candidates were verified in terms of domain content by searching CDD database with *E*-value < 0.01 (Tab. 1) [26]. Alignments of *Toc12*, *Tic21*, *Tic32* and *Tic62* from *P. sativum* or *A. thaliana* and two strains of *P. chromatophora*, CCAC0185 and FK01, were performed in PSI-Coffee [27], slow and accurate algorithm dedicated to distantly related

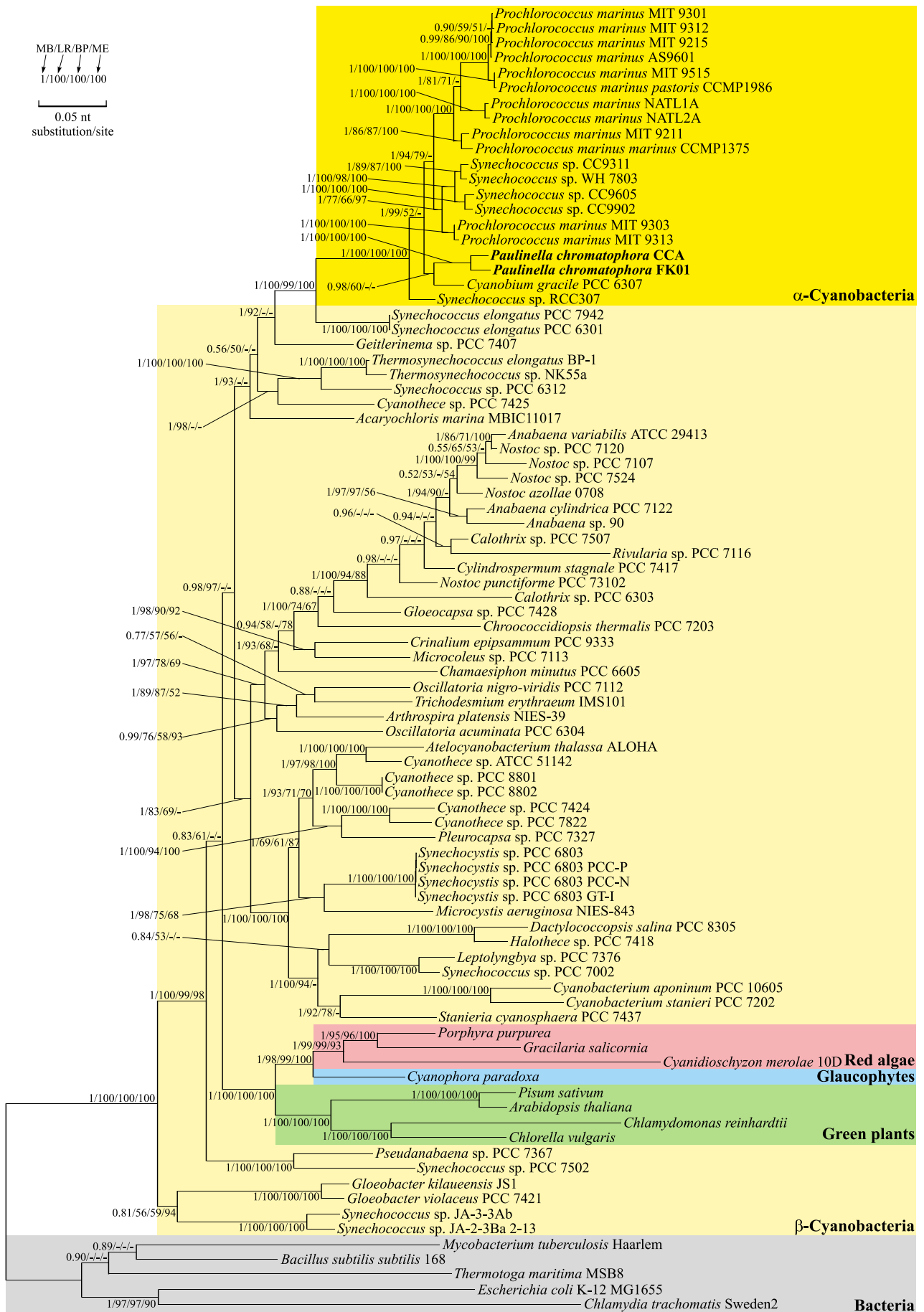


Fig. 1 MrBayes tree based on 16S and 23S rRNA alignment. An independent origin of *Paulinella* endosymbionts from primary plastids of red algae, glaucophytes and green plants is clearly visible. The former are descendants of α -cyanobacteria whereas the latter of β -cyanobacteria. Numbers at nodes, in the presented order, correspond to posterior probabilities estimated in MrBayes (MB), support values calculated by local rearrangements-expected likelihood weights method (LR), as well as bootstrap percentages calculated in TreeFinder (BP) and PAUP using minimum evolution method (ME). Values of the posterior probabilities and bootstrap percentages lower than 0.50 and 50%, respectively, were omitted or indicated by a dash “-”. The length of branches leading to Bacteria taxa was shortened by 50% in comparison to their original length because of very high substitution rate.

proteins using profile information. The alignments were visualized and edited in Jalview 2.4.0.b2 [28]. The N-terminal transmembrane β -barrel regions were predicted with BOC-TOPUS [29], whereas α -helical transmembrane regions with TopPred [30] and TMPred [31].

Phylogenetic trees for the concatenated alignment of 16S and 23S rRNA were inferred by maximum likelihood (ML) method in TreeFinder [32], Bayesian approach in MrBayes 3.2.1 [33] and minimum evolution (ME) method based on logDet/paralinear distance in PAUP [34]. In TreeFinder and MrBayes approaches, we used separate models of nucleotide substitutions for each type of rRNA. Two different GTR+ Γ (5) models for each data partition were applied in the ML method as suggested by the Propose Model module in Treefinder considering all criteria ($-\ln L$, AIC, AICc,

BIC, HQ). In MrBayes analyses, we assumed two separate mixed+I+ Γ (5) models for each rRNA partition to sample appropriate models across the substitution model space in the Bayesian MCMC analysis itself avoiding the need for a priori model testing [35]. In TreeFinder detailed search depth set to 2 was applied whereas in PAUP the final tree was searched from 10 starting trees obtained by stepwise and random sequence addition followed by the tree-bisection-reconnection (TBR) branch-swapping algorithm. To assess significance of particular branches, non-parametric bootstrap analyses were performed on 1000 replicates in these two methods. Additionally, we applied the local rearrangements-expected likelihood weights (LR-ELW) method in TreeFinder. In MrBayes analyses, two independent runs starting from random trees were applied, each using eight Markov chains. Trees were sampled every 100 generations for 10 000 000 generations. In the final analysis, we selected trees from the last 2 984 000 generations that reached the stationary phase and convergence (i.e. the standard deviation of split frequencies stabilized and was lower than the proposed threshold of 0.01). The temperature parameter for heating the chains was suitably adjusted to keep the proportions of successful state exchanges between chains close to the suggested range from 0.10 to 0.70.

Tab. 1 Domains characteristic of Toc and Tic proteins from *Pisum sativum* and *Arabidopsis thaliana*, which were used to predicted their homologs in *Paulinella*, cyanobacteria and bacteria proteomes.

Protein name	CDD domain	Additional information
Tic20	233225	Tic20, chloroplast protein import component, length = 267
Tic21	256942	DUF3611, protein of unknown function, length = 183
Tic22	233226	TIC22, chloroplast protein import component, length = 270
	252499	Tic22-like family, length = 268
Tic32	212492	Retinol dehydrogenase, length = 269
	235736	Oxidoreductase, length = 315
Tic55	239830	Tic55, a 55 kDa LLS1-related non-heme iron oxygenase, length = 134
Tic62	178748	Translocon at the inner envelope of chloroplast, length = 576
Toc12	197617	DnaJ molecular chaperone homology domain, length = 60
	249696	DnaJ domain, length = 63
Toc34	130064	GTP-binding protein (chloroplast envelope protein translocase, length = 313
	206652	Toc34 like, length = 248
Toc75	215598	Protein TOC75, length = 796
	130065	Chloroplast envelope protein translocase, IAP75 family, length = 718
Toc64	166363	Indole-3-acetamide amidohydrolase, length = 422
	181375	Amidase, length = 395
	238112	TPR, Tetratricopeptide repeat domain, length = 100
Toc159	233224	Chloroplast protein import component Toc86/159, length = 763

Results and discussion

We found that both *Paulinella* chromatophore genomes encode the same set of sequence homologs to Toc12, Toc34, Toc64, Toc159, Tic21, Tic32 and Tic62 but lost those of Toc75, Tic20, Tic22 and Tic55 (Fig. 2). Interestingly, the missing genes are still present in the α -cyanobacterial genomes (*Cyanobium gracile*, *Prochlorococcus marinus* and *Synechococcus* CC9311, CC9605, CC9902 and WH 7803) closely related to *Paulinella* (Fig. 1). This indicates that they might have been lost during chromatophore genome reduction or transferred to the nucleus. An interesting case is Tic22 homolog, which is absent not only from *Paulinella* but also from all α -cyanobacteria grouped with *Paulinella* except for *Synechococcus* RCC307 (Fig. 1). The latter α -cyanobacterium is placed at the basal position to the clade indicating that the gene encoding Tic22 must have been lost after the divergence of this basal lineage from the other α -cyanobacteria. It should be mentioned that Tic22 homolog is present in all other, more than 50 studied cyanobacterial taxa (Fig. 2). Independent losses concern also the gene for Tic62 homolog in the cyanobacterium *Atelocyanobacterium thalassa*, Tic21 in the land plant *Pisum sativum* as well as Tic55 in the cyanobacterium *Synechococcus* sp. JA-3-3Ab and

Species name	Toc/Tic components										
	Tic32	Toc12	Tic21	Tic62	Tic55	Toc75	Tic20	Tic22	Toc159	Toc34	Toc64
<i>Paulinella chromatophora</i> CCAC 0185											
<i>Paulinella chromatophora</i> FK01											
<i>Acaryochloris marina</i> MBIC11017											
<i>Anabaena cylindrica</i> PCC 7122											
<i>Anabaena</i> sp. 90											
<i>Anabaena variabilis</i> ATCC 29413											
<i>Arthrospira platensis</i> NIES39											
<i>Calothrix</i> sp. PCC 6303											
<i>Calothrix</i> sp. PCC 7507											
<i>Atelocyanobacterium thalassa</i> ALOHA											
<i>Chamaesiphon minutus</i> PCC 6605											
<i>Chroococcidiopsis thermalis</i> PCC 7203											
<i>Crinalium epipsammum</i> PCC 9333											
<i>Cyanobacterium aponinum</i> PCC 10605											
<i>Cyanobacterium stanieri</i> PCC 7202											
<i>Cyanobium gracile</i> PCC 6307											
<i>Cyanothece</i> sp. ATCC 51142											
<i>Cyanothece</i> sp. PCC 7424											
<i>Cyanothece</i> sp. PCC 7425											
<i>Cyanothece</i> sp. PCC 7822											
<i>Cyanothece</i> sp. PCC 8801											
<i>Cyanothece</i> sp. PCC 8802											
<i>Cylindrospermum stagnale</i> PCC 7417											
<i>Dactylococcopsis salina</i> PCC 8305											
<i>Geitlerinema</i> sp. PCC 7407											
<i>Gloeobacter kilaeensis</i> JS1											
<i>Gloeobacter violaceus</i> PCC 7421											
<i>Gloeocapsa</i> sp. PCC 7428											
<i>Halothece</i> sp. PCC 7418											
<i>Leptolyngbya</i> sp. PCC 7376											
<i>Microcoleus</i> sp. PCC 7113											
<i>Microcystis aeruginosa</i> NIES843											
<i>Nostoc azollae</i> 0708											
<i>Nostoc punctiforme</i> PCC 73102											
<i>Nostoc</i> sp. PCC 7107											
<i>Nostoc</i> sp. PCC 7120											
<i>Nostoc</i> sp. PCC 7524											
<i>Oscillatoria acuminata</i> PCC 6304											
<i>Oscillatoria nigroviridis</i> PCC 7112											
<i>Pleurocapsa</i> sp. PCC 7327											
<i>Prochlorococcus marinus</i> AS9601											
<i>Prochlorococcus marinus</i> MIT 9211											
<i>Prochlorococcus marinus</i> MIT 9215											
<i>Prochlorococcus marinus</i> MIT 9301											
<i>Prochlorococcus marinus</i> MIT 9303											
<i>Prochlorococcus marinus</i> MIT 9312											
<i>Prochlorococcus marinus</i> MIT 9313											
<i>Prochlorococcus marinus</i> MIT 9515											
<i>Prochlorococcus marinus</i> NATL1A											
<i>Prochlorococcus marinus</i> NATL2A											
<i>Prochlorococcus marinus marinus</i> CCMP1375											
<i>Prochlorococcus marinus pastoris</i> CCMP1986											
<i>Pseudanabaena</i> sp. PCC 7367											
<i>Rivularia</i> sp. PCC 7116											
<i>Stanieria cyanosphaera</i> PCC 7437											
<i>Synechococcus elongatus</i> PCC 6301											
<i>Synechococcus elongatus</i> PCC 7942											
<i>Synechococcus</i> sp. CC9311											
<i>Synechococcus</i> sp. CC9605											
<i>Synechococcus</i> sp. CC9902											
<i>Synechococcus</i> sp. JA23B'a(213)											
<i>Synechococcus</i> sp. JA33Ab											
<i>Synechococcus</i> sp. PCC 6312											
<i>Synechococcus</i> sp. PCC 7002											
<i>Synechococcus</i> sp. PCC 7502											
<i>Synechococcus</i> sp. RCC307											
<i>Synechococcus</i> sp. WH 7803											
<i>Synechocystis</i> sp. PCC 6803											
<i>Synechocystis</i> sp. PCC 6803 GT1											
<i>Synechocystis</i> sp. PCC 6803 PCCN											
<i>Synechocystis</i> sp. PCC 6803 PCCP											
<i>Thermosynechococcus elongatus</i> BP1											
<i>Thermosynechococcus</i> sp. NK55a											
<i>Trichodesmium erythraeum</i> IMS101											
<i>Arabidopsis thaliana</i>											
<i>Chlamydomonas reinhardtii</i>											
<i>Chlorella variabilis</i>											
<i>Chondrus crispus</i>											
<i>Cyanidioschyzon merolae</i>											
<i>Cyanophora paradoxa</i>											
<i>Physcomitrella patens</i>											
<i>Pisum sativum</i>											
<i>Bacillus subtilis</i> 168											
<i>Escherichia coli</i> K12 MG1655											

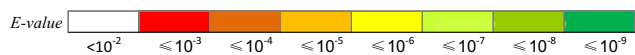


Fig. 2 Distribution of Toc and Tic homologs in *Paulinella chromatophora*, cyanobacteria, Archaeplastida and two bacteria. The homologs were found by PsiBlast searches and verified for domain content by searching conserved domain database (CDD). Only best hits with *E*-values obtained from CDD equal or lower than 0.01 are indicated. The color scheme corresponds to *E*-value range. Please note that the genomes of *Paulinella* chromatophores encode homologs of Toc12, Toc34, Toc159, Tic21, Tic32 and Tic62 but those of, Toc75, Tic20, Tic22, Tic55 and Toc64 were lost. In the case of Toc64 significant homologous sequences were found; however, they did not contain a tetratricopeptide domain, responsible for Hsp90 docking, and therefore we left this column blank.

the red alga *Cyanidioshyzon merolae*. The gene for Tic21 is also absent from two studied *Gleobacter* species, which often take basal position to the rest cyanobacteria in phylogenetic trees [36–38], suggesting that lack of this gene might have characterised the ancestor of all cyanobacteria.

Because no homolog to the crucial outer membrane translocon Toc75 was found in the two chromatophore genomes and the *Paulinella* nuclear genome [16], the discovery of homologs to its two receptors, Toc34 and Toc159, was much unexpected (Fig. 2). However, we do not consider these proteins to be components of protein import machinery without Toc75, especially taking into account their relatively low *E*-values, in both PsiBlast and CDD analyses (Fig. 2). They may be distantly related to the plastid Toc components but they do not seem to be their functional homologs. Interestingly, quite significant *E*-values obtained for some cyanobacterial Toc34/Toc159 homologs suggest that these proteins might be of cyanobacterial origin. This contrasts with the common prevailing idea that they are derived from an ancient eukaryotic GTPase, Toc159 as the result of Toc34 duplication [39–41]. A thorough phylogenetic investigation is, however, necessary to clarify the issue, which should be at present considered only a hypothesis.

We also found sequences significantly similar to *Pisum sativum* Toc64 in both *Paulinella* chromatophore genomes and bacterial genomes (Fig. 2). However, they should not be considered its true functional homologs because all of them did not have an important tetratricopeptide (TPR) domain, responsible for Hsp90 docking, and therefore cannot fulfil Toc64 function. The gene for Toc64 encoding the TPR domain is also absent from glaucophytes, rhodophytes and the green alga *Chlamydomonas reinhardtii* but present in the other green alga *Chlorella variabilis*, some prasinophytes and land plants (Fig. 2 and data not shown). It indicates that the gene is a late acquisition in the green lineage.

In contrast to Toc64, homologs of Toc12, Tic21 and Tic32 in both *Paulinella* strains contain all the essential domains present in plant counterparts (Fig. 3). *Paulinella* Tic21 with its four well predicted α -helical transmembrane regions is suggested to create a protein conducting channel at the inner chromatophore membrane and can be involved in protein import. Such function of plant Tic21 was recently reported by Kikuchi et al. [42] and Hirabayashi et al. [43], and Lv et al. [44] proved functional compatibility of plant and cyanobacterial Tic21 by showing that the knockout mutant of *Arabidopsis tic21* was rescued by *Synechocystis* ortholog. Tic32 homologs, which possess dehydrogenase and calmodulin-binding domains, probably regulate the import via redox sensing [20]. Interestingly, Toc12 homologs, instead of N-terminal transmembrane β -barrel domains typical of plant proteins, comprise C-terminal transmembrane

α -helical regions. Toc12, is one of the four chromatophore-encoded Hsp40 (DnaJ) proteins and a putative component of molecular motor responsible for pulling imported proteins into the chromatophore stroma [20].

The new discovery of the undertaken analyses was identification of potential homologs to Tic62 in *Paulinella*. In *Pisum sativum* Tic62 consists of two characteristic regions: the N-terminal part with a highly conserved NAD(P)-binding domain and C-terminal part with binding sites for ferredoxin-NAD(P)-oxido-reductase (FNR; Fig. 3) [45]. Similar N-terminal organization was shown for *Paulinella* Tic62 homologs, but the C-terminal region with FNR interacting repeats is missing from both *Paulinella* proteins (Fig. 3). However, the C-terminal region, which apparently evolved only in vascular plants, is not necessary for binding to the Tic translocon/inner membrane. The region responsible for mediating binding to the Tic complex is localized in the central part of *Pisum sativum* Tic62 [46], whereas the C-terminal region allows for specific and strong binding of ferredoxin-NAD(P)-oxido-reductase molecules, especially when Tic62 is present at the stroma lamellae of the thylakoid membrane [47]. The reversible interaction with Tic complex/inner membrane is supposed to be mediated by some hydrophobic contacts [48]. One or two transmembrane domains were proposed in *Pisum sativum* Tic62 by K uchler et al. [45]. In agreement with that, we predicted two such domains using TopPred [30] and TMPred [31]. Interestingly, we also identified one potential transmembrane domains in each of *Paulinella* Tic62 proteins using these two programs (Fig. 3).

We suggest that *Paulinella* Tic62 homolog is a new component of the simplified but fully-functional Tic-like translocon at the inner chromatophore membrane, composed, beside Tic62, of Tic21, Tic32 and Toc12 (Hsp40; Fig. 4). The fact that these proteins possess all the essential domains to fulfil the task further strengthens the reliability of a Tic complex at the inner chromatophore membrane. Because Tic55, the third redox sensing protein [49–51], has not been discovered in the chromatophore genome in previous [17] and present study, as well as in the *Paulinella* nuclear genome [16], we suggest Tic32 and Tic62 to perform the task of redox sensing on their own or with some unknown partner.

An interesting argument for functioning of a simplified Tic system provided Kikuchi et al. [52]. They suggested that the primordial Archaeplastida Tic complex was probably based on Tic20. Similarly to Tic21 and mitochondrial translocon at the inner membrane Tim23/Tim17, Tic20 contains four well predicted α -helical transmembrane domains capable of forming a protein conducting channel, and therefore together with Tic21 and Tim23/Tim17 represent a good example of parallel evolution of import functions [42]. Since a Tic20-based system might have functioned

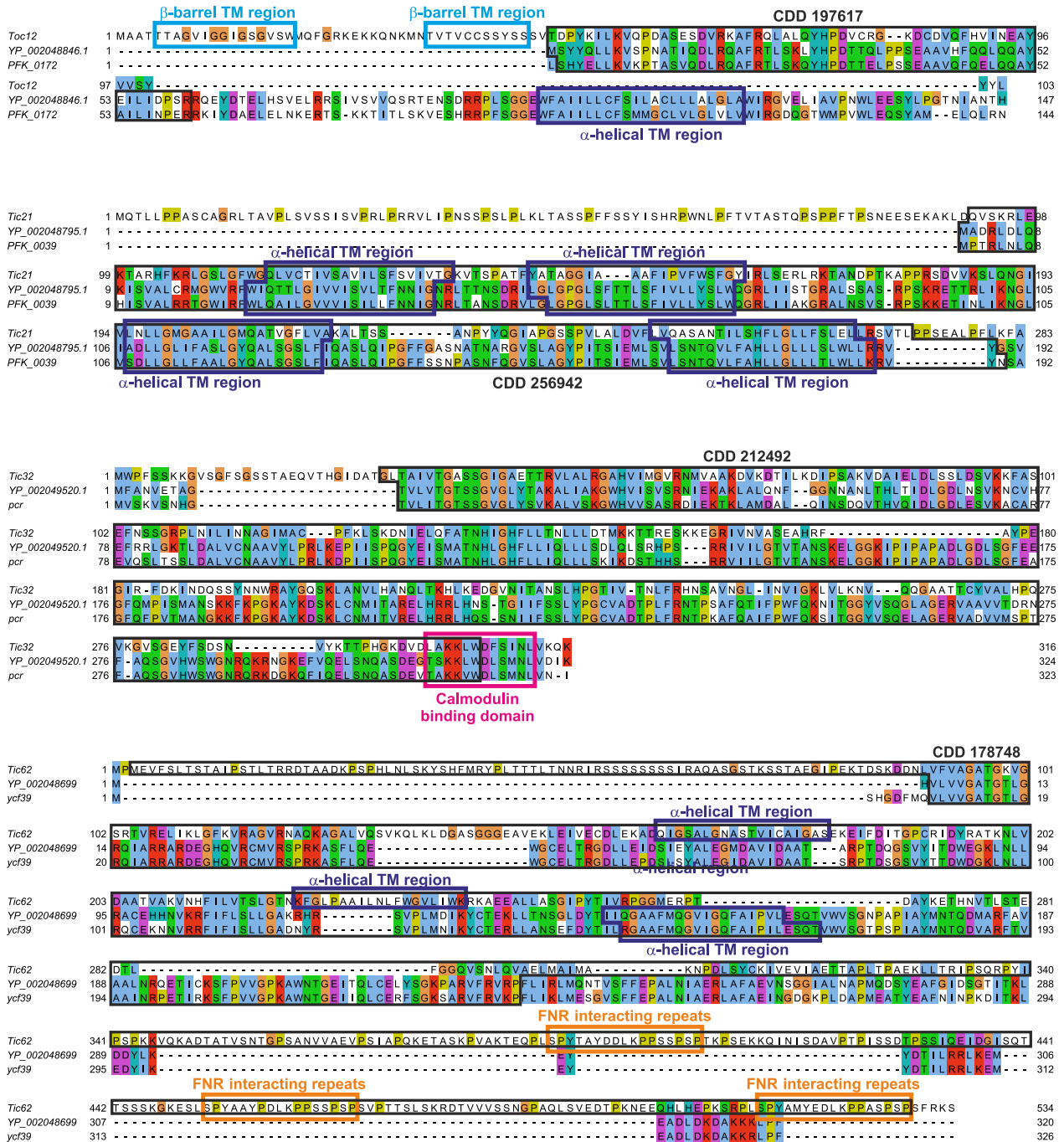


Fig. 3 Alignment of Tic12, Tic21, Tic32 and Tic62 from *Pisum sativum* and two strains of *Paulinella chromatophora* CCAC0185 and FK01. The range of domains characteristic of a given protein was indicated according to CDD database prediction. A characteristic motif of calmodulin binding domain in Tic32 was also marked [20], as well as NADH(P) binding domain (CDD 257784) and FNR-interacting repeats in Tic62 [46]. The N-terminal transmembrane β -barrel regions were predicted with BOCTOPUS [29], α -helical transmembrane regions with TopPred [30] and TMpred [31]. The average range of the regions was calculated based on the two latter predictions.

in the ancestor of Archaeplastida, it is easy to imagine the possibility of analogous system based on Tic21 in the case of *Paulinella*. In support of this protein import via Tic21 was experimentally proved in plants [42,43]. Although additional core translocon subunits of eukaryotic origin, such as Tic214, Tic100 and Tic56 were added to Tic20 in the 'green' lineage, they are missing from basal Archaeplastida, glaucophytes and red algae [52]. This suggests that a simple

Tic apparatus is responsible for protein import into primary plastids of glaucophytes and red algae and probably *Paulinella* chromatophora as well. Kikuchi et al. [52] also showed that eukaryote-derived Tic110, previously considered the main translocation pore, is not a component of the complex consisting of Tic20, Tic214, Tic100 and Tic56, but plays a role of a scaffold for stromal molecular chaperones at a later stage during protein import [49,50].

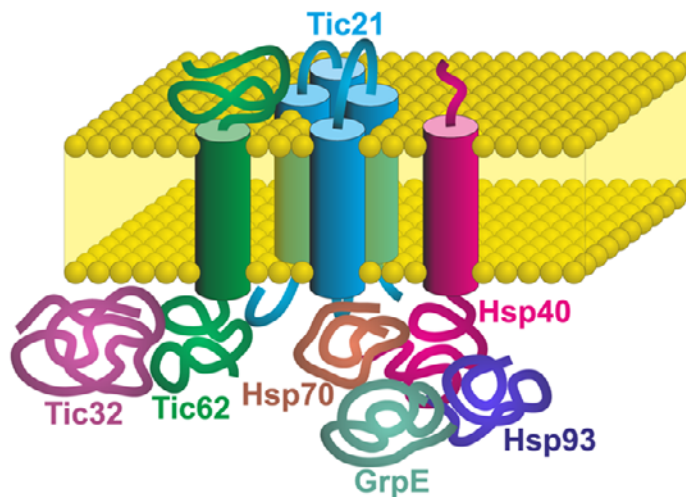


Fig. 4 Model for a protein translocon in inner membrane of *Paulinella* photosynthetic chromatophores. The translocon is a simplified Tic-like apparatus and is responsible for final import step of nuclear-encoded proteins across the inner chromatophore membrane. It is composed of Tic21 (protein-conducting channel), Tic32 and Tic62 (calcium and redox-sensing regulatory proteins) as well as a molecular motor responsible for pulling imported proteins into the organelle stroma. The latter could consist of Hsp93, Hsp70, Hsp40 (Toc12) and GrpE.

Conclusion

Our studies reveal that chromatophore genomes from both *Paulinella* strains encode the same set of translocons that could potentially create a simplified but fully-functional Tic-like translocon at the inner chromatophore membranes. Moreover, we have discovered a new putative Tic component, namely Tic62, a redox sensor protein not identified

in previous comparative studies of *Paulinella* translocons [17]. The common maintenance of the same set of Toc/Tic proteins indicates a similar import mechanism in the two investigated *Paulinella* strains and supports the proposed model. We also suggest a possibility that Toc34 and Toc159, GTPases of presumed eukaryotic origin [39–41], might in fact be derived from cyanobacterial proteins.

Acknowledgments

We are very grateful to the Reviewers and the Editor for their excellent comments and insightful remarks which significantly improved the paper. This work was supported by Polish National Science Centre grant No. 2011/01/N/NZ8/00150 and Wrocław Centre for Networking and Supercomputing grant No. 307 to PG and PM.

Authors' contributions

The following declarations about authors' contributions to the research have been made: conducting analyses and writing the manuscript: PG, PM.

Competing interests

No competing interests have been declared.

References

1. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol.* 2004;21:809–818. <http://10.1093/molbev/msh075>
2. Deusch O, Landan G, Roettger M, Gruenheit N, Kowalik KV, Allen JF, et al. Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol Biol Evol.* 2008;25:748–761. <http://10.1093/molbev/msn022>
3. Cavalier-Smith T, Lee JJ. Protozoa as hosts for endosymbioses and the conversion of symbionts into organelles. *J Eukaryot Microbiol.* 1985;32:376–379. <http://10.1111/j.1550-7408.1985.tb04031.x>
4. Bodył A, Mackiewicz P, Stiller JW. Early steps in plastid evolution: current ideas and controversies. *Bioessays.* 2009;31:1219–1232. <http://10.1002/bies.200900073>
5. Gross J, Bhattacharya D. Mitochondrial and plastid evolution in eukaryotes: an outsiders' perspective. *Nat Rev Genet.* 2009;10:495–505. <http://10.1038/nrg2610>
6. Bodył A, Mackiewicz P, Gagat P. Organelle evolution: *Paulinella* breaks a paradigm. *Curr Biol.* 2012;22:R304–306. <http://10.1016/j.cub.2012.03.020>
7. Bhattacharya D, Helmchen T, Melkonian M. Molecular evolutionary analyses of nuclear-encoded small subunit ribosomal RNA identify an independent rhizopod lineage containing the Euglyphina and the Chlorarachniophyta. *J Eukaryot Microbiol.* 1995;42:65–69. <http://10.1111/j.1550-7408.1995.tb01541.x>
8. Yoon HS, Nakayama T, Reyes-Prieto A, Andersen RA, Boo SM, Ishida K, et al. A single origin of the photosynthetic organelle in different *Paulinella* lineages. *BMC Evol Biol.* 2009;9:98. <http://10.1186/1471-2148-9-98>
9. Marin B, Nowack EC, Melkonian M. A plastid in the making: evidence for a second primary endosymbiosis. *Protist.* 2005;156:425–432. <http://10.1016/j.protis.2005.09.001>
10. Yoon HS, Reyes-Prieto A, Melkonian M, Bhattacharya D. Minimal plastid genome evolution in the *Paulinella* endosymbiont. *Curr Biol.* 2006;16:R670–672. <http://10.1016/j.cub.2006.08.018>
11. Nowack EC, Melkonian M, Glockner G. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr Biol.* 2008;18:410–418. <http://10.1016/j.cub.2008.02.051>
12. Kies L. Electron microscopical investigations on *Paulinella chromatophora* Lauterborn, a thecamoeba containing blue-green endosymbionts (Cyanelles). *Protoplasma.* 1974;80:69–89. <http://10.1007/BF01666352>

13. Kies L, Kremer BP. Function of cyanelles in the tecamoeba *Paulinella chromatophora*. *Naturewissenschaften*. 1979;66:578–579.
14. 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;27:1530–1537. <http://10.1093/molbev/msq032>
15. Nakayama T, Ishida KI. Another acquisition of a primary photosynthetic organelle is underway in *Paulinella chromatophora*. *Curr Biol*. 2009;19:R284–R285. <http://dx.doi.org/10.1016/j.cub.2009.02.043>
16. 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;28:407–422. <http://10.1093/molbev/msq209>
17. Bodył A, Mackiewicz P, Stiller JW. Comparative genomic studies suggest that the cyanobacterial endosymbionts of the amoeba *Paulinella chromatophora* possess an import apparatus for nuclear-encoded proteins. *Plant Biol*. 2010;12:639–649. <http://10.1111/j.1438-8677.2009.00264.x>
18. Mackiewicz P, Bodył A. A hypothesis for import of the nuclear-encoded PsaE protein of *Paulinella chromatophora* (Cercozoa, Rhizaria) into its cyanobacterial endosymbionts/plastids via the endomembrane system. *J Phycol* 2010;46:847–859. <http://10.1111/j.1529-8817.2010.00876.x>
19. Mackiewicz P, Bodył A, Gagat P. Possible import routes of proteins into the cyanobacterial endosymbionts/plastids of *Paulinella chromatophora*. *Theory Biosci*. 2012;131:1–18. <http://10.1007/s12064-011-0147-7>
20. Mackiewicz P, Bodył A, Gagat P. Protein import into the photosynthetic organelles of and its implications for primary plastid endosymbiosis. *Symbiosis*. 2012;58:99–107. <http://10.1007/s13199-012-0202-2>
21. Nowack EC, Grossman AR. Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proc Natl Acad Sci USA*. 2012;109:5340–5345. <http://10.1073/pnas.1118800109>
22. Qiu H, Yang EC, Bhattacharya D, Yoon HS. Ancient gene paralogy may mislead inference of plastid phylogeny. *Mol Biol Evol*. 2012;29:3333–3343. <http://10.1093/molbev/mss137>
23. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Res*. 2013;41:D36–D42. <http://10.1093/nar/gks1195>
24. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–3402. <http://10.1093/nar/25.17.3389>
25. Kalanon M, McFadden GI. The chloroplast protein translocation complexes of *Chlamydomonas reinhardtii*: a bioinformatic comparison of Toc and Tic components in plants, green algae and red algae. *Genetics*. 2008;179:95–112. <http://10.1534/genetics.107.085704>
26. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, et al. CDD: a conserved domain database for protein classification. *Nucleic Acids Res*. 2005;33:D192–196. <http://10.1093/nar/gki069>
27. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*. 2011;39:W13–17. <http://10.1093/nar/gkr245>
28. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25:1189–1191. <http://10.1093/bioinformatics/btp033>
29. Hayat S, Elofsson A. BOCTOPUS: improved topology prediction of transmembrane β barrel proteins. *Bioinformatics*. 2012;28:516–522. <http://10.1093/bioinformatics/btr710>
30. Claros MG, von Heijne G. TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci*. 1994;10:685–686. <http://10.1093/bioinformatics/10.6.685>
31. Hofmann. TMbase – a database of membrane spanning proteins segments. *Biol Chem Hoppe Seyler*. 1993;374.
32. Jobb G, von Haeseler A, Strimmer K. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol*. 2004;4:18. <http://10.1186/1471-2148-4-18>
33. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–542. <http://10.1093/sysbio/sys029>
34. Rogers JS, Swofford DL. A fast method for approximating maximum likelihoods of phylogenetic trees from nucleotide sequences. *Syst Biol*. 1998;47:77–89. <http://10.1080/106351598261049>
35. Huelsenbeck JP, Larget B, Alfaro ME. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol Biol Evol*. 2004;21:1123–1133. <http://10.1093/molbev/msh123>
36. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA*. 2013;110:1053–1058. <http://10.1073/pnas.1217107110>
37. Criscuolo A, Gribaldo S. Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. *Mol Biol Evol*. 2011;28:3019–3032. <http://10.1093/molbev/msr108>
38. Ochoa de Alda JAG, Esteban R, Diago ML, Houmard J. The plastid ancestor originated among one of the major cyanobacterial lineages. *Nat Commun*. 2014;5. <http://10.1038/ncomms5937>
39. Reumann S, Inoue K, Keegstra K. Evolution of the general protein import pathway of plastids (review). *Mol Membr Biol*. 2005;22:73–86. <http://10.1080/09687860500041916>
40. Gross J, Bhattacharya D. Reevaluating the evolution of the Toc and Tic protein translocons. *Trends Plant Sci*. 2009;14:13–20. [http://S1360-1385\(08\)00286-0](http://S1360-1385(08)00286-0)
41. Hernandez Torres J, Maldonado MA, Chomilier J. Tandem duplications of a degenerated GTP-binding domain at the origin of GTPase receptors Toc159 and thylakoidal SRP. *Biochem Biophys Res Commun*. 2007;364:325–331. <http://10.1016/j.bbrc.2007.10.006>
42. Kikuchi S, Oishi M, Hirabayashi Y, Lee DW, Hwang I, Nakai M. A 1-megadalton translocation complex containing Tic20 and Tic21 mediates chloroplast protein import at the inner envelope membrane. *Plant Cell*. 2009;21:1781–1797. <http://10.1105/tpc.108.063552>
43. Hirabayashi Y, Kikuchi S, Oishi M, Nakai M. In vivo studies on the roles of two closely related *Arabidopsis* Tic20 proteins, AtTic20-I and AtTic20-IV. *Plant Cell Physiol*. 2011;52:469–478. <http://10.1093/pcp/pcr010>
44. Lv HX, Guo GQ, Yang ZN. Translocons on the inner and outer envelopes of chloroplasts share similar evolutionary origin in *Arabidopsis thaliana*. *J Evol Biol*. 2009;22:1418–1428. <http://10.1111/j.1420-9101.2009.01755.x>
45. Kuchler M, Decker S, Hormann F, Soll J, Heins L. Protein import into chloroplasts involves redox-regulated proteins. *EMBO J*. 2002;21:6136–6145. <http://10.1016/j.bbamcr.2010.01.015>
46. Stengel A, Benz P, Balsera M, Soll J, Bolter B. TIC62 redox-regulated translocon composition and dynamics. *J Biol Chem*. 2008;283:6656–6667. <http://10.1074/jbc.M706719200>
47. Benz JP, Stengel A, Lintala M, Lee YH, Weber A, Philippar K, et al. *Arabidopsis* Tic62 and ferredoxin-NADP(H) oxidoreductase form light-regulated complexes that are integrated into the chloroplast redox poise. *Plant Cell*. 2009;21:3965–3983. <http://10.1105/tpc.109.069815>
48. Balsera M, Stengel A, Soll J, Bolter B. Tic62: a protein family from metabolism to protein translocation. *BMC Evol Biol*. 2007;7:43. <http://10.1186/1471-2148-7-43>
49. Li HM, Chiu CC. Protein transport into chloroplasts. *Annu Rev Plant Biol*. 2010;61:157–180. <http://10.1146/annurev-arplant-042809-112222>
50. Shi LX, Theg SM. The chloroplast protein import system: from algae to trees. *Biochim Biophys Acta*. 2013;1833:314–331. <http://10.1016/j.bbamcr.2012.10.002>
51. Stengel A, Benz JP, Soll J, Bolter B. Redox-regulation of protein import into chloroplasts and mitochondria: similarities and differences. *Plant Signal Behav*. 2010;5:105–109. <http://10.4161/psb.5.2.10525>
52. Kikuchi S, Bedard J, Hirano M, Hirabayashi Y, Oishi M, Imai N, et al. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science*. 2013;339:571–574. <http://10.1126/science.1229262>