plant biology

Plant Biology ISSN 1435-8603

RESEARCH PAPER

Comparative genomic studies suggest that the cyanobacterial endosymbionts of the amoeba *Paulinella chromatophora* possess an import apparatus for nuclear-encoded proteins

A. Bodył[1], P. Mackiewicz[2] & J. W. Stiller[3]

1 Department of Biodiversity and Evolutionary Taxonomy, Zoological Institute, University of Wrocław, Wrocław, Poland

2 Department of Genomics, Faculty of Biotechnology, University of Wrocław, Wrocław, Poland

3 Department of Biology, East Carolina University, Greenville, TX, USA

Keywords

Endosymbiosis; gene transfer; plastid; protein trafficking; Tic; Toc.

Correspondence

A. Bodył, Department of Biodiversity and Evolutionary Taxonomy, Zoological Institute University of Wrocław, ul. Przybyszewskiego 63/77,
51-148 Wrocław, Poland.
E-mail: bodyl@biol.uni.wroc.pl

Editor

M. Sugita

Received: 8 March 2009; Accepted: 5 August 2009

doi:10.1111/j.1438-8677.2009.00264.x

ABSTRACT

Plastids evolved from free-living cyanobacteria through a process of primary endosymbiosis. The most widely accepted hypothesis derives three ancient lineages of primary plastids, *i.e.* those of glaucophytes, red algae and green plants, from a single cyanobacterial endosymbiosis. This hypothesis was originally predicated on the assumption that transformations of endosymbionts into organelles must be exceptionally rare because of the difficulty in establishing efficient protein trafficking between a host cell and incipient organelle. It turns out, however, that highly integrated endosymbiotic associations are more common than once thought. Among them is the amoeba *Paulinella chromatophora*, which harbours independently acquired cyanobacterial endosymbionts functioning as plastids. Sequencing of the *Paulinella* endosymbiont genome revealed an absence of essential genes for protein trafficking, suggesting their residence in the host nucleus and import of protein products back into the endosymbiont. To investigate this hypothesis, we searched the *Paulinella* endosymbiont genome for homologues of higher plant translocon proteins that form the import apparatus in two-membrane envelopes of primary plastids. We found homologues of Toc12, Tic21 and Tic32, but genes for other key translocon proteins (*e.g.* Omp85/Toc75 and Tic20) were missing. We propose that these missing genes were transferred to the *Paulinella* nucleus and their products are imported and integrated into the endosymbiont envelope membranes, thereby creating an effective protein import apparatus. We further suggest that other bacterial/cyanobacterial endosymbionts found in protists, plants and animals could have evolved efficient protein import systems independently and, therefore, reached the status of true cellular organelles.

INTRODUCTION

Molecular phylogenetic analyses clearly demonstrate that plastids evolved from free-living cyanobacteria engulfed by phagotrophic protozoans (for reviews see Palmer 2003; Howe *et al.* 2008; Archibald 2009). These bacterial cells were probably capable of both photosynthesis and nitrogen fixation (Kneip *et al.* 2007; Deusch *et al.* 2008), providing clear selective advantages for their establishment and maintenance as permanent intracellular endosymbionts. In a next evolutionary stage, the endosymbionts underwent transformation into true cell organelles, or primary plastids, surrounded by a two-membrane envelope (Palmer 2003; Howe *et al.* 2008; Archibald 2009). This conversion involved two key processes: (i) massive gene transfer from the endosymbiont genome to the host nucleus, and (ii) origin of an import machinery to transport proteins encoded in the nucleus back into distinct plastid sub-compartments (Cavalier-Smith & Lee 1985). These sub-compartments include the outer and inner envelope membranes, the space between these two membranes, the stroma, as well as membranes and the lumen of the thylakoid system (Schünemann 2007; Inaba & Schnell 2008).

Primary plastids are characteristic for three main eukaryotic lineages: glaucophytes, red algae and green plants, including green algae and their higher plant descendants (Palmer 2003; Howe *et al.* 2008; Archibald 2009). Glaucophyte plastids are especially interesting because they retain several ancestral cyanobacterial characteristics, including (i) a bacterial cell wall (containing peptidoglycan) (Pfanzagl & Löffelhardt 1999), (ii) carboxysomes (Fathinejad *et al.* 2008) and (iii) unstacked thylakoids with phycobilisomes (Steiner *et al.* 2003). Although there is still some debate (see, *e.g.* Stiller *et al.* 2003; Howe *et al.* 2008; Kim & Graham 2008), it is widely accepted that all these plastids resulted from a single primary endosymbiosis (Palmer 2003; Archibald 2009). One of the key features adduced to favour the single origin of all primary plastids is the presence of homologous import machinery for nuclear-encoded proteins in their membranes (McFadden & van Dooren 2004; Steiner *et al.* 2005). This machinery is composed of two translocation systems called

Toc (the translocon of the outer chloroplast membrane) and Tic (the translocon of the inner chloroplast membrane) (for reviews see Soll & Schleiff 2004; Reumann *et al.* 2005; Inaba & Schnell 2008). Each of these translocons contains a central protein-conducting channel and associated receptor and regulatory proteins (see Fig. 1). Available data suggest that the Toc–Tic super-complex has a chimeric origin with some subunits derived from the endosymbiont and the others coming from the host. For example, it was shown that Toc75 has a cyanobacterial origin, whereas Tic110 was proposed to have evolved from a host protein (for details see Reumann *et al.* 2005; Löffelhardt *et al.* 2007; Kalanon & McFadden 2008). The common presence of Tic110, which is not found in cyanobacteria, is frequently considered to be strong evidence for a single origin of glaucophyte, red algal and green plant plastids (McFadden & van Dooren 2004; Steiner *et al.* 2005); however, this does not imply support for the idea that the evolutionary origin of plastid import machinery must have been a profoundly difficult and unique accomplishment (see, *e.g.* Cavalier-Smith 1992, 2008).

New support for separating these two arguments comes from the thecate, filose amoeba *Paulinella chromatophora* (Marin *et al.* 2005, 2007; Archibald 2006; Nowack *et al.* 2008). This amoeba harbours two cyanobacterial endosymbionts, which are highly integrated with the host cell. Interestingly, they divide synchronously with the *Paulinella* host, exchange metabolites with its cytosol and are incapable of growth outside the host cell (Kies 1974; Kies & Kremer 1979; Marin *et al.* 2005). As with glaucophyte plastids, the cyanobacterial endosymbionts of *P. chromatophora* retain a peptidoglycan wall (Kies 1974), which makes them osmotically stable structures. This feature could suggest, at first glance, a common cyanobacterial origin with glaucophyte plastids; however, *P. chromatophora* and glaucophytes belong to distinct eukaryotic super-groups [Rhizaria and Archaeplastida (Plantae), respectively] (Cavalier-Smith & Chao 2003; Rodríguez-Ezpeleta *et al.* 2005) and this has long been viewed as indicating that their endosymbionts/plastids have separate origins. In support of this hypothesis, phylogenetic trees based on the rDNA operon showed that the photosynthetic endosymbionts of *P. chromatophora* do not belong to the main primary plastid lineage but, rather, are strongly monophyletic with the cyanobacterial *Synechococcus* 'marine'/*Prochlorococcus/Cyanobium* clade (Marin *et al.* 2005, 2007). Finally, *P. chromatophora* has close heterotrophic and aplastidal relatives, *Paulinella ovalis* and *P. indentata*, which regularly engulf cyanobacteria belonging to the genus *Synechococcus* (Johnson *et al.* 1988; Hannah *et al.* 1996). Thus, all evidence indicates that *Paulinella* endosymbionts represent an independent lineage from classical primary plastids.

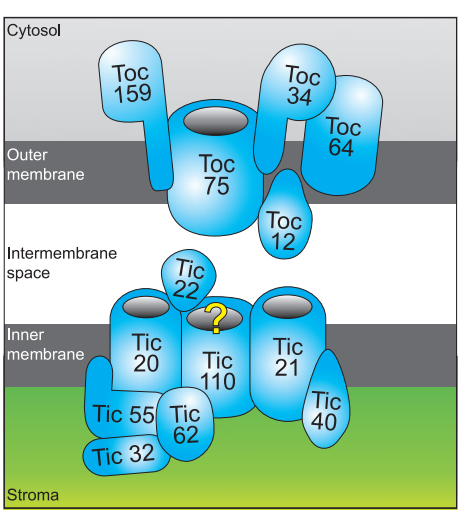

Fig. 1. Structure of the import apparatus of higher plant plastids. These plastids possess two translocons: the Toc and Tic systems, with individual proteins playing key roles as protein-conducting channels. There is only one such protein, Toc75, in the outer membrane whereas at least two, Tic20 and Tic21, occur in the inner membrane. The exact function of Tic110 is still under debate; some research groups suggest that it represents an additional channel protein (with a main transport function), whereas others argue that it functions as a scaffold protein. The remaining proteins in Toc and Tic translocons fulfill receptor (Toc12, Toc34, Toc64, Toc159, Tic22), chaperone (Tic40) and regulatory (Tic32, Tic55, Tic62) functions. The numbers on respective proteins/subunits indicate their predicted molecular weights in kDa. All Toc and Tic subunits are encoded in the host nucleus and imported into the plastid, as indicated by their blue colour.

It has been hotly debated whether the photosynthetic bodies of *P. chromatophora* should be considered obligate endosymbionts or whether they have already attained the status of true cellular organelles (Bhattacharya & Archibald 2006; Theissen & Martin 2006; Yoon *et al.* 2006; Bodył *et al.* 2007). New light on this problem is shed by the completely sequenced genome of *Paulinella* endosymbionts (Nowack *et al.* 2008). In contrast to previous suppositions (see Yoon *et al.* 2006), these studies demonstrate that the genome has undergone a drastic reduction, losing ~75% of its original gene content. Among the sequences lost were many with known and essential functions. For example, genes encoding pathways for the biosynthesis of several amino acids (*e.g.* glutamine, arginine, methionine) and cofactors (*e.g.* nicotinamide adenine dinucleotides, riboflavine, biotin, coenzyme A) are completely absent from the *Paulinella* endosymbiont genome (Nowack *et al.* 2008). More importantly, individual genes have been lost from otherwise intact and essential endosymbiont biosynthetic pathways and sub-cellular structures. Good examples are *hemD* (encoding the uroporphyrinogen III synthase of the chlorophyll-heme-phycocyanobilin pathway) and *sulA* (encoding a cell-division inhibitor that blocks polymerisation of FtsZ) (Nowack *et al.* 2008). Another surprising feature of this genome is a paucity of genes encoding solute channels/membrane transporters.

The absence of essential genes encoding uroporphyrinogen III synthase, SulA and solute channels/membrane transporters from the endosymbiont strongly suggests that they now reside in the *Paulinella* nucleus. In fact, Nakayama & Ishida (2009) recently found empirical evidence for such transfer in the case of *psaE*, which encodes the cyanobacterial photosystem I reaction centre subunit IV. These new findings argue that a number of essential proteins could be encoded by genes in the *Paulinella* cell nucleus and trafficked to the endosymbiont. They also beg the further question, what kind of import machinery do these proteins use? To address this issue, we searched for homologues of plastidic Toc and Tic subunits, representing potential components of a transloca-

tion machinery, in the completely sequenced genome of *Paulinella* endosymbionts and in 33 completely sequenced genomes of free-living cyanobacteria.

MATERIAL AND METHODS

Searches for Toc and Tic homologues

We applied the search35 program (Pearson 2000), which performs a highly sensitive search based on a rigorous Smith–Waterman alignment (Smith & Waterman 1981), to find potential homologues of Toc and Tic proteins in the inferred proteomes of the *Paulinella* endosymbiont, 33 completely sequenced cyanobacteria, as well as in *Bacillus subtilis*, *Escherichia coli* and *Streptomyces coelicolor* used as outgroups in phylogenetic analyses (see below). We queried with sequences of all well characterised Toc and Tic proteins from *Pisum sativum* and *Arabidopsis thaliana*. Statistical parameters of the searches were estimated from shuffled copies of each library sequence (random sequences with the length and amino acid composition of the original sequence), which provides greater statistical accuracy. The best hits with E-values equal or lower than 0.01 are presented in Fig. 2.

Identification of domains and sequence motifs in Toc and Tic proteins

Protein domains and motifs were identified in the InterPro database (http://www.ebi.ac.uk/interpro) using all available applications/databases, including BlastProDom, FPrintScan, HMMPIR, HMMPfam, HMMSmart, HMMTigr, ProfileScan, ScanRegExp, patternScan, SuperFamily, SignalPHMM, TMHMM, HMMPanther and Gene3D. In prediction of transmembrane α-helices, we used 14 different algorithms: ConPredII (with separate models for prokaryotes and eukaryotes), DAS, HMMTOP, MEMSAT3, PHDhtmm, PHOBIUS,

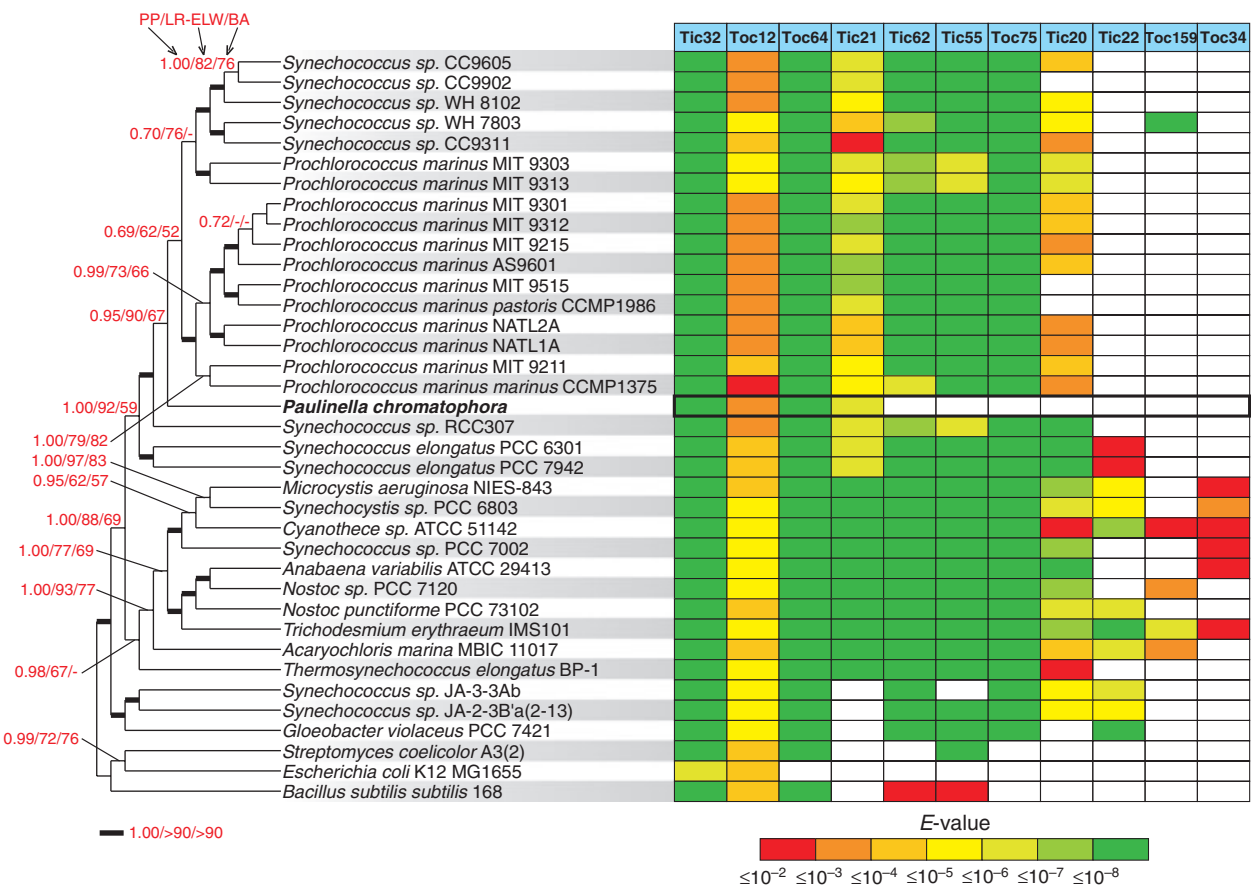

Fig. 2. Phylogenetic position of the cyanobacterial endosymbionts of *Paulinella chromatophora* and distribution of Toc and Tic homologues in cyanobacteria, including the *Paulinella* endosymbiont. The left panel presents Bayesian phylogenetic analyses of 16S and 23S rDNA sequences from completely sequenced cyanobacterial species and strains and the photosynthetic endosymbionts of *P. chromatophora* (see Material and Methods). Numbers at nodes correspond to (i) posterior probabilities calculated in the MrBayes program (PP), (ii) support values obtained in TreeFinder by the local rearrangement method (LR-ELW), and (iii) the maximum likelihood bootstrap analysis (BA). Posterior probabilities and bootstrap percentages ≤0.50% or 50%, respectively, were omitted or indicated by a dash '-'. Note that maximum likelihood and Bayesian methods produced the same tree topology. In the right panel, results are presented of searches for homologues to Toc and Tic proteins from *Pisum sativum* and *Arabidopsis thalina* in the *Bacillus*, *Escherichia* and *Streptomyces* outgroup bacteria, 33 cyanobacteria and *Paulinella* endosymbionts. Only the best hits with E-values equal to or lower than 0.01 are indicated. Please note that the genome of *Paulinella* endosymbionts encodes homologues of Toc12, Toc64, Tic21 and Tic32, but those of Toc75, Tic20, Tic55 and Tic62 are absent.

PRED-TMR, SOSUI, TMHMM, TMPRED, TUPS, waveTM and WHAT. Hydropathy profiles were made according to the Kyte–Doolittle scale. Detailed descriptions of methods and results of all searches, along with relevant references, are provided in Supporting Information (see Appendix S1).

Phylogenetic analyses

16S and 23S rDNA sequences from the *Paulinella* endosymbiont, cyanobacterial species/strains and the *Bacillus*, *Escherichia* and *Streptomyces* outgroup bacteria were downloaded from Genbank (http://www.ncbi.nlm.nih.gov). These three bacteria were chosen as outgroups for three reasons: (i) they represent three major eubacterial lineages: Firmicutes, Proteobacteria and Actinobacteria; (ii) their genomes are well annotated; and (iii) they are free-living bacteria, meaning their genomes were not subjected to extensive genome reduction (*e.g.* due to transition to a parasitic or endosymbiotic lifestyle).

The sequences were aligned using R-Coffee (Wilm *et al.* 2008), which is specifically designed for RNA and utilises structural information about the molecule. Sites suitable for further phylogenetic analyses were extracted from the alignment using Gblocks 0.91b (Castresana 2000) and assuming less stringent criteria; this allows: (i) smaller final blocks of alignment, (ii) gap positions within the final blocks and (iii) less strict flanking positions.

Phylogenetic trees were inferred from concatenated alignments of 16S and 23S rDNA sequences (in sum 4185 sites) using two approaches: maximum likelihood in Treefinder (Jobb *et al.* 2004) and Bayesian inference in MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003). In both cases we used separate models of nucleotide substitutions for each type of rDNA. GTR+Γ (five rate categories) models were applied in the ML approach as suggested by the Propose Model module in Treefinder (Jobb *et al.* 2004), whereas GTR+I+Γ (five rate categories) models were used in the Bayesian inference as indicated by MrAIC 1.4.2 (Nylander 2004), which uses the PhyML program (Guindon & Gascuel 2003) to calculate model parameters.

Bayesian analyses included two simultaneous, independent runs starting from random trees, each using four Markov chains. Trees were sampled every 100 out of 50 million generations. In the final analysis, we selected trees from the last 19 million generations that reached stationarity with an average standard deviation of split frequencies well below 0.01. Edge support of the tree was assessed in TreeFinder by bootstrap analysis with ML, using 1000 replicates and the Local Rearrangements Expected–Likelihood Weights method (LR-ELW).

RESULTS AND DISCUSSION

Structure and evolution of Toc and Tic translocons in classical plastids

The Toc and Tic translocons in modern green algal and higher plant plastids have multi-subunit structures (Fig. 1) (Soll & Schleiff 2004; Reumann *et al.* 2005; Inaba & Schnell 2008). The core Toc translocon is composed of three proteins: Toc34, Toc75 and Toc159. Toc34 and Toc159 function as receptors for imported proteins, whereas Toc75 forms a protein-conducting channel. Toc64 represents an additional receptor protein, but it is loosely associated with the core Toc complex. Toc12, Toc64 and Tic22 constitute an inter-envelope space complex designated to deliver imported proteins to the core Tic complex. In addition to Tic22, the core Tic complex involves four additional proteins: Tic20, Tic21, Tic40 and Tic110. Tic20 and Tic21 appear to represent translocation/insertion pores; Tic40 functions as a co-chaperone. The role of Tic110 is controversial and remains under debate; it has been suggested to function either as a protein-conducting channel or as a scaffold protein (see Inaba *et al.* 2003; Bédard & Jarvis 2005; Kalanon & McFadden 2008; Balsera *et al.* 2009). Tic55, Tic62 and Tic32 are proposed to be involved in redox regulation of the protein import process.

It appears clear that these multi-subunit translocons must have evolved gradually from much simpler transport systems (for a recent discussion see Cavalier-Smith 2006a). Such initial Toc and Tic translocons could have been composed of only protein-conducting channels. In support of this gradualistic hypothesis, the Toc75 pore not only has a channel but also a receptor domain (Ertel *et al.* 2005). Moreover, it was proven experimentally that higher plant plastids containing Toc75, but with Toc receptors inactivated by either thermolysin (Chen *et al.* 2000) or T-DNA insertion (Hust & Gutensohn 2006), are still able to import plastid proteins. It is even possible that the ancestral primary plastid import machinery contained only the Toc75 pore in the outer envelope membrane. Its function could have been insertion of solute channels/membrane transporters into the outer membrane and import of other channels/transporters into the periplasmic space, from where they could be inserted into the inner plastid membrane (Cavalier-Smith 2006a; Bodył *et al.* 2007). In accordance with this hypothesis, the outer membrane of modern higher plant plastids contains 'free' Toc75 proteins mediating insertion of outer envelope proteins (Tu *et al.* 2004).

Cyanobacterial genomes encode homologues of key Toc and Tic subunits: pre-adaptation for evolution of an import apparatus?

Available data clearly demonstrate that the cyanobacterial endosymbionts of *P. chromatophora* were acquired independently from glaucophyte, red algal and green plant plastids (Marin *et al.* 2005, 2007). Thus, if they have some sort of protein import apparatus, as strongly suggested by the absence of essential genes from the *Paulinella* endosymbiont genome (see Yoon *et al.* 2006; Bodył *et al.* 2007; Nowack *et al.* 2008), this machinery would have evolved independently of the Toc–Tic super-complex operating in classical primary plastids. Here, however, emerge two key questions: (i) is the establishment of protein trafficking between host and endosymbiont as difficult as has been commonly argued, and (ii) what kind of import machinery for nuclear-encoded proteins is most likely to have evolved in the *Paulinella* endosymbiosis?

Our broad searches for Toc and Tic homologues in 33 completely sequenced genomes from free-living cyanobacterial species, and particularly those closely related to *Paulinella* endosymbionts, revealed genes for all key Toc and Tic proteins, including Toc12, Toc64, Toc75, Tic20, Tic21, Tic32, Tic55 and Tic62 (Fig. 2). Similar to other cyanobacteria, we also found significant hits to homologues of *toc* and *tic* genes

in *Synechococcus* sp. WH 5701 (the incomplete genome is still under construction and, therefore, is not shown in Fig. 2). This cyanobacterium is a member of the *Cyanobium* clade that clusters with the *Synechococcus* 'marine'/*Prochlorococcus* group in rRNA operon trees, with *Paulinella* endosymbionts branching just outside this clade (Marin *et al.* 2007). We emphasise that all the cyanobacteria investigated have homologues of Toc75 (Fig. 2). In a freshly acquired cyanobacterial endosymbiont, the opening of a Toc75 channel would be oriented toward the inter-membrane space and, therefore, it could not participate in the import of nuclear-encoded proteins (Steiner *et al.* 2005). However, such import would quickly become possible after the *toc75* gene was transferred to the host nucleus, resulting in insertion of its protein product into the outer cyanobacterial membrane in a reverse orientation (Steiner *et al.* 2005). Cyanobacteria also contain homologues of Tic20 and Tic21 (Fig. 2). Both these proteins were suggested to constitute protein-conducting channels in the inner membrane of primary plastids (Chen *et al.* 2002; Teng *et al.* 2006).

All these data indicate that any cyanobacterial endosymbiont would contain all the essential elements required to create a Toc–Tic-based import system. Interestingly, it was recently demonstrated that Toc159 could have evolved from the thylakoid SRP54 protein, which hydrolyses GTP and constitutes the plastid signal recognition particle (SRP) (Hernández Torres *et al.* 2007). SRPs target signal peptide-containing proteins to cellular membranes and are universal in all domains of life (Grudnik *et al.* 2009). This means that even the Toc159 receptor, regarded previously as evidence of a host cell contribution to the plastid import apparatus, could also be derived from the cyanobacterial ancestor of primary plastids. Thus, given that all necessary Toc and Tic components undoubtedly were present at the time of its engulfment, it is reasonable to postulate that the cyanobacterial endosymbiont of *P. chromatophora* could have evolved a comparable import system to that of classical primary plastids in one or both of its envelope membranes.

Absence of Omp85/Toc75 homologue in the genome of *Paulinella* endosymbionts and its implications

Searches of the completely sequenced genome of *Paulinella* endosymbionts revealed homologues of only Toc12, Toc64, Tic21 and Tic32; in contrast, Toc75, Tic20, Tic55 and Tic62 were not found, although they all exist in closely related cyanobacteria (Figs 2 and 3). Among these proteins, especially intriguing is the absence of a Toc75 homologue.

Cyanobacteria are Gram-negative bacteria and their envelope is composed of two membranes, the plasmalemma and the outer membrane (Cavalier-Smith 2006b). The outer cyanobacterial membrane contains the Omp85 protein, from which Toc75 apparently evolved (Ertel *et al.* 2005; Löffelhardt *et al.* 2007). Omp85 plays an essential function in the biogenesis of the outer negibacterial membrane by inserting solute channels/membrane transporters (Voulhoux & Tommassen 2004; Gentle *et al.* 2005); it is found in all sampled cyanobacteria and its loss during evolution of distinct groups of bacteria is highly improbable (Cavalier-Smith 2006b). Thus, the absence of the homologue of Omp85/Toc75 from the genome of *Paulinella* endosymbionts demands a specific explanation.

There are two alternative explanations for this intriguing absence: either the Omp85/Toc75 homologue is no longer needed by the endosymbiont, and has been lost completely, or it has been transferred to the *Paulinella* nucleus and the protein is transported back to the endosymbiont outer membrane. Which of these scenarios appears more likely? Two membranes currently surround the endosymbionts of *P. chromatophora* (Kies 1974). After their initial engulfment, however, the original endosymbionts were likely surrounded by three envelope membranes: the two cyanobacterial membranes and a phagosomal membrane of the host (for a detailed discussion see Cavalier-Smith 2000). Thus, it could be speculated that the outer negibacterial membrane was lost, negating the need for an Omp85/Toc75 homologue. In this case, the current outer membrane of *Paulinella* endosymbionts would represent the host phagosomal membrane. Phagosome-derived outermost membranes are characteristic of numerous algal lineages with eukaryotic (secondarily) derived plastids with greater than two envelope membranes, such as euglenids, cryptophytes, heterokonts and dinoflagellates (Archibald 2009; Bolte *et al.* 2009).

It should be noted, however, that *Paulinella* endosymbionts still biosynthesise and maintain a peptidoglycan wall (Kies 1974). This suggests that the eukaryotic phagosomal membrane was the one that was lost, rather than the outer cyanobacterial membrane. Interestingly, it was originally suggested that glaucophyte plastids had lost the outer negibacterial membrane (see, for example, Kies 1988), but the presence of an Omp85/Toc75 homologue (Steiner *et al.* 2005) and galactolipids (Koike *et al.* 2007) indicate their outer membrane is derived from the original outer cyanobacterial membrane (for other arguments see Cavalier-Smith 1987). Additional support for this view comes from recent detailed analyses of isolated *Cyanophora paradoxa* plastid membranes (Koike *et al.* 2007). Considering these data, it is most reasonable that the outer membrane of *Paulinella* endosymbionts, likewise, corresponds to the outer negibacterial membrane. This also implies that its biogenesis requires the presence of an Omp85/Toc75 homologue. Therefore, we suggest that, rather than having been lost completely, *omp85/toc75* and genes for other missing Toc and Tic homologues (*i.e. tic20*, *tic55*, and *tic62*), were transferred to the *P. chromatophora* nucleus, and that their protein products are imported back into the endosymbionts (or rather plastids) to form an import apparatus (Figs 3 and 4).

Finally, it also is possible that the outer membrane of *Paulinella* endosymbionts contains an Omp85/Toc75 protein, but is chimeric in nature with features of both bacterial outer and eukaryotic phagosomal membranes. It is straightforward to envision a scenario by which uncontrolled division of endosymbionts resulted in disruption of symbiosomes (*i.e.* modified phagosomes) and their escape into the host cytosol. During such escapes, the outer bacterial membrane could acquire lipids and proteins specific for the host phagosomal membrane, thereby gaining characteristics of an endomembrane system-like membrane, while retaining others from its bacterial ancestry (see also Kilian & Kroth 2003).

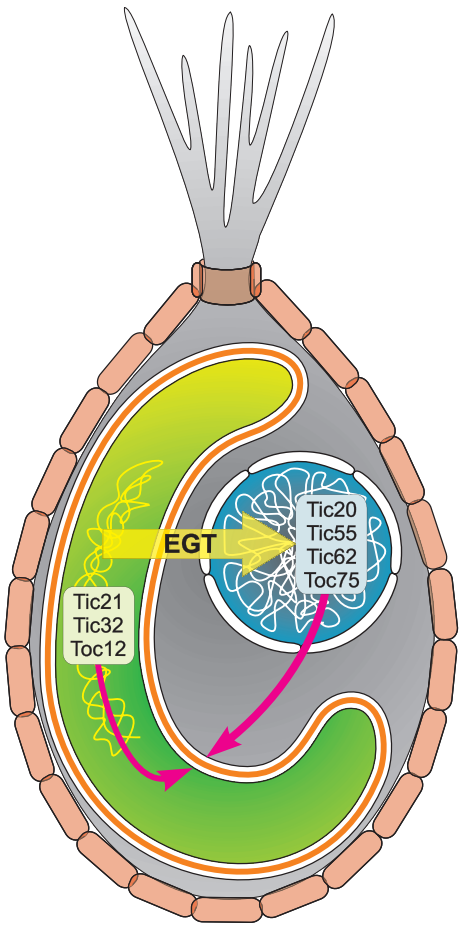

Fig. 3. Possible distribution of the genes encoding Toc and Tic proteins in the endosymbiont and the host genome of *Paulinella chromatophora*. In contrast to free-living cyanobacteria, *Paulinella* endosymbionts encode homologues of only three Toc and Tic proteins: Toc12, Tic21 and Tic32 (see Fig. 2). Because Toc75 plays a key role in biogenesis of the outer bacterial membrane, and Tic20 is responsible for transport of amino acids, we suggest that the genes for these key proteins were transferred *via* endosymbiotic gene transfer (EGT) to the host nucleus. It is also possible that other genes (*e.g.* those encoding redox sensing Tic55 and Tic62 proteins) were transferred to the *Paulinella* nucleus. This scenario implies that Toc75, Tic20, Tic55 and Tic62 are targeted to the proper endosymbiont membrane from the host cytosol, while Toc12, Tic21 and Tic32 are delivered to envelope membranes from the endosymbiont matrix, creating an import apparatus for nuclear-encoded proteins (see Fig. 4).

Structure of the hypothetical import apparatus of *Paulinella* endosymbionts

Translocation across the outer membrane

Based on our bioinformatics results, it is useful at this stage to speculate on the structure of a hypothetical import apparatus (Fig. 4). Given that genes encoding Toc12, Tic21 and Tic32 remain under strong purifying selection in the *Paulinella* endosymbiont genome, it is reasonable to postulate that translocons analogous to those in primary plastids have evolved. Probably the simplest mechanisms would be based on the homologue of Omp85/Toc75 that, alone, would be sufficient to establish an insertion and/or translocation system for nuclear-encoded proteins in the outer endosymbiont membrane. As discussed above, the cyanobacterial form of this protein has both receptor and channel domains. Thus, a *Paulinella* Omp85/Toc75 protein would be able to both recognise and translocate endosymbiont/plastid-destined proteins (Fig. 4B).

We should also consider a model in which nuclear-encoded proteins are imported into *Paulinella* endosymbionts *via* the endomembrane (EM) system (Fig. 4C). EM-mediated import occurs across outermost membranes in eukaryotic alga-derived (secondary) plastids, where nuclear-encoded proteins carry signal peptides and are targeted to the plastid through the ER and/or Golgi apparatus (Bolte *et al.* 2009). EM-mediated targeting of plastid proteins was also recently described for several proteins (*e.g.* α-amylase and α-carbonic anhydrase) in primary plastids of higher plants (Chen *et al.* 2004; Villarejo *et al.* 2005).

It also is possible that an Omp85/Toc75-independent channel exists in the outer membrane of *Paulinella* endosymbionts (Fig. 4D). Such an alternative translocation pore could have evolved from a pre-existing channel in the outer negibacterial membrane, or from a channel re-located from another cellular compartment. A good candidate for the latter would be the mitochondrion, with its Tom and Tim translocons that resemble the plastid Toc–Tic super-complex (Cavalier-Smith 2006a; Neupert & Herrmann 2007). It is possible that the Tom40 or Tim22 channel was re-located to the outer membrane of *Paulinella* endosymbionts, thereby creating a functional translocation pore. In support of this view, the outer membrane of higher plant plastids contains the OEP16 channel, probably engaged in the import of protochlorophyllide oxidoreductase A (Reinbothe *et al.* 2004; Pollmann *et al.* 2007). Available data indicate that this channel evolved from the mitochondrial Tim22 pore, which was re-located to the outer plastid membrane in the green plant lineage (Cavalier-Smith 2006a).

Finally, we should seriously consider a model in which protein translocation across the outer membrane of *Paulinella* endosymbionts proceeds *via* several distinct pathways (Fig. 4 A). For example, some proteins could be imported *via* the endomembrane system (Fig. 4C), whereas others are translocated through a membrane channel, such as Omp85/Toc75 or Tim22 (Fig. 4B and D).

Translocation across the inner membrane

A good candidate for a protein-conducting channel in the inner membrane of *Paulinella* endosymbionts is the homologue of Tic21 encoded by its genome (Figs 2 and 3). As with its higher plant counterpart, *Paulinella* Tic21 contains four well-predicted α-helical transmembrane domains (see Figs S4 and S5; Table S3 in the online Appendix S1), suggesting that it is capable of forming a translocation pore for imported proteins. It was proposed that Tic21 from higher plants functions as a protein-conducting channel (Teng *et al.* 2006) and/or an iron permease (Duy *et al.* 2007). It is possible that the original activity of the *Paulinella* Tic21 homologue was iron transport but it could well be pre-adapted to function as a protein-conducting pore (Fig. 4B–D). Dual functions of Toc and Tic components were recently postulated as a key factor in evolution of the protein import apparatus of primary plastids (Gross & Bhattacharya 2009).

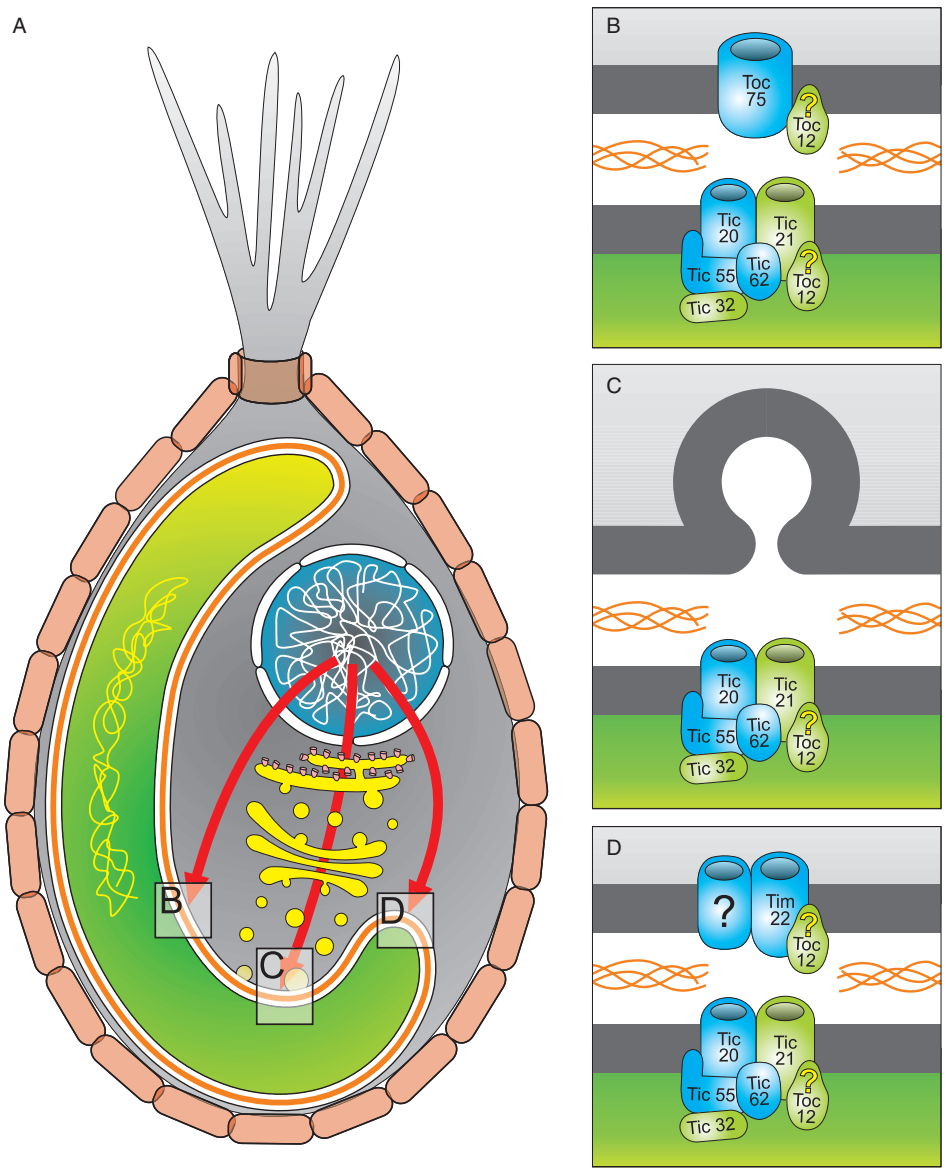

Fig. 4. Hypothetical import mechanism for nuclear-encoded proteins into the cyanobacterial endosymbionts of *Paulinella chromatophora*. As with classical primary plastids, *Paulinella* endosymbionts are surrounded by two membranes (A). Thus, considering the universal presence of key Toc and Tic homologues in a wide variety of cyanobacteria including those closely related to *Paulinella* endosymbionts (see Fig. 2), we hypothesize that a Toc–Tic-like import apparatus could has evolved (plate and scenario B). All proposed scenarios assume that the inner membrane of *Paulinella* endosymbionts contains a Tic-like translocon, because the inner membrane of their cyanobacterial ancestor contained Tic20 and Tic21, which are pre-adapted to form protein-conducting channels. However, protein translocation across the outer membrane could have proceeded *via* three distinct pathways. First, the key role of Omp85/-Toc75 in biogenesis of the outer negibacterial membrane suggests that a Toc75 channel is responsible for this import step (B). Second, it is possible that nuclear-encoded proteins are targeted to *Paulinella* endosymbionts in Golgi apparatus- and/or ER-derived vesicles that fuse with the outer membrane and liberate imported proteins into the inter-membrane space (C). Third, a model should also be considered in which proteins cross the outer envelope membrane thorough a Toc75-independent pore, such as a mitochondrial channel (*e.g.* Tim22) or other unknown channel (D). It is also possible that two or all of these pathways co-exist in the outer membrane of *Paulinella* endosymbionts and are used by distinct groups of nuclear-encoded proteins (A). The location of a Toc12 homologue is unclear: it might be anchored in the outer, inner or in both envelope membranes. The endosymbiont import machinery could be regulated by Tic32, Tic55 and Tic62 homologues. Proteins in green are encoded in the endosymbiont genome, whereas those in blue are presumably localized in the host nucleus.

Another candidate for a protein-conducting channel in the inner membrane of *Paulinella* endosymbionts is Tic20 (Figs 3 and 4B–D). Its higher plant counterpart folds into four α-helical transmembrane domains (Chen *et al.* 2002) and belongs to a pre-protein and amino acid transporter (or PRAT) family, along with bacterial channels for branched amino acids and mitochondrial Tim17, 23 and 22 channels for pre-proteins (Rassow *et al.* 1999). Therefore, we hypothesise that the Tic20 gene was transferred to the *Paulinella* nucleus and its protein product is imported and

inserted into the inner endosymbiont membrane (Figs 3 and 4B–D).

Other Toc- and Tic-like proteins in the Paulinella import apparatus

The *Paulinella* endosymbiont genome encodes a homologue of Toc12 (Figs 2 and 3), which contains two domains: (i) a DnaJ domain providing a binding site for HSP70s and (ii) a transmembrane α-helical domain functioning as a membrane anchor (see Figs S1 and S2; Table S1 in the online Appendix S1). Thus, it resembles its higher plant homologue (see Figs S1 and S2; Table S1 in the online Appendix S1) and it is possible that *Paulinella* Toc12 is anchored in the inner and/or the outer endosymbiont membrane (Fig. 4B–D), operating as a molecular motor that pulls imported proteins into the endosymbiont matrix by binding HSP70s (for details on functioning of such a molecular motor see Soll & Schleiff 2004; Reumann *et al.* 2005; Inaba & Schnell 2008).

A homologue of Tic32 was also found in the *Paulinella* endosymbiont genome (Figs 2 and 3). It shows clear similarity along its whole sequence to its higher plant counterparts, possessing dehydrogenase/reductase (SDR), NAD(P)-binding and calmodulin (CaM)-binding domains (see Fig. S6; Table S4 in the online Appendix S1). In addition, the sequence of *Paulinella* Tic32 possesses all conserved motifs and residues associated with dehydrogenase/reductase activity (Fig. S6). Finally, the sequence motif characteristic of the CaM-binding domain is more conserved in *Paulinella* Tic32 (Fig. S6) than in free-living cyanobacteria, such as *Nostoc* sp. PCC 7120 (Kalanon & McFadden 2008 and data not shown). Since the higher plant Tic32 combines with Tic55 and Tic62 to form a redox state-sensitive regulation complex (Soll & Schleiff 2004; Reumann *et al.* 2005; Inaba & Schnell 2008), we hypothesise that genes encoding Tic55 and Tic62 homologues were transferred to the *Paulinella* nucleus and that these proteins are imported into the endosymbiont matrix to create such a complex (Figs 3 and 4B–D).

Our homology searches of the *Paulinella* endosymbiont genome also identified a sequence similar to the higher plant Toc64 receptor (Fig. 2). However, in contrast to its higher plant counterpart (Qbadou *et al.* 2007), the *Paulinella* homologue does not contain the tetratrico-peptide (TPR) domain at the C-terminus (see Fig. S3; Table S2 in the online Appendix S1). Since these four TPR motifs form the docking site for HSP90-bound plastid proteins (Qbadou *et al.* 2007), we assume that the *Paulinella* Toc64 does not fulfill functions related to protein translocation across endosymbiont membranes (see also Fig. S3), and thus we did not include it in our hypothesised import machinery (Fig. 4B–D).

Possible functions of the hypothetical import apparatus of *Paulinella* endosymbionts

It is reasonable to postulate that one of the functions of the hypothetical import apparatus of *Paulinella* endosymbionts is insertion of channels/transporters into the outer and inner envelope membranes (Bodył *et al.* 2007). Toc75 and its OEP80 homologue were suggested to operate as insertion machinery for membrane proteins residing in the outer membrane of higher plant plastids (Tu *et al.* 2004; Patel *et al.* 2008). Moreover, it was hypothesised that the first stage in the evolution of both mitochondria and primary plastids was development of insertion systems for solute channels/membrane transporters in their envelope membranes (Cavalier-Smith 2006a). It is possible that the cyanobacterial endosymbionts of *P. chromatophora* reside at this proposed initial stage in evolution of a protein import apparatus for endosymbiont-derived organelles; however, it appears more likely that they already import some proteins into the matrix. As noted in our introduction, the *Paulinella* endosymbiont genome does not encode uroporphyrinogen III synthase and SulA, and these proteins probably are imported into the endosymbiont matrix. Very strong support for this hypothesis comes from the demonstration that PsaE, which needs to be targeted to the endosymbiont thylakoids, is encoded in the nuclear genome of *P. chromatophora* (Nakayama & Ishida 2009).

The principle of minimising endosymbiotic events and number of successful transformations of endosymbionts into organelles

Recent evolutionary models in the field of plastid evolution have been dominated by the principle of minimising endosymbiotic events. This was based on the assumption that, because of the complexity of efficient protein trafficking between a host cell and incipient organelle, transformation of an endosymbiont into a plastid was a very difficult process, and one that would be nearly impossible to repeat (see, *e.g.* Cavalier-Smith 1992, 2008). Largely based on this principle, the kingdom Plantae (or Archaeplastida) was created for glaucophytes, red algae and green plants (Cavalier-Smith 1998; Adl *et al.* 2005). Whereas this assemblage of algae with primary plastids has received some support as a monophyletic group from other data (Rodríguez-Ezpeleta *et al.* 2005; Reyes-Prieto & Bhattacharya 2007; but also see Nozaki *et al.* 2007 and Kim & Graham 2008 for alternative results), acceptance of a single origin of 'archaeplastids' should not be recycled as a circular argument in favour of minimising all endosymbioses when constructing evolutionary scenarios.

The case of *P. chromatophora* establishes that the transformation of a cyanobacterium into an integrated organelle was not a unique evolutionary event. Moreover, our analyses show that all cyanobacteria possess homologues of key Toc and Tic proteins (Fig. 2), indicating that they are pre-adapted to create a Toc–Tic-like import system for nuclear-encoded proteins (Figs 3 and 4). Most previous evolutionary scenarios for the origin of a plastid import apparatus focused on stroma-targeted proteins with N-terminal targeting signals. It is very likely, however, that such evolution started from membrane transporters (inserted into the outer and/or inner plastid membranes), which were devoid of N-terminal transit peptides (Cavalier-Smith 2006a; Bodył *et al.* 2007). Thus, it is possible that at least some other cyanobacterial endosymbionts [see, *e.g.* those residing in the diatoms *Pinnularia* spp. (Schmid 2003) and *Rhopalodia* spp. (Kneip *et al.* 2008)] import solute channels/membrane transporters and, therefore, have also passed an accepted threshold to be considered true cellular organelles.

PERSPECTIVES

Extensive experimental work is needed to demonstrate the fate of the essential genes missing from *Paulinella* endos-

ymbionts, and to explain how this newborn plastid imports missing proteins needed for its biogenesis. Similar analyses to those reported by Nakayama & Ishida (2009) should be carried out to demonstrate that other missing but essential genes, especially those engaged in a potential translocation system, were transferred from the *Paulinella* endosymbiont to the nuclear genome. Determination of the full complement of potential players in plastid protein-trafficking in *P. chromatophora*, however, will require concerted transcriptomic and genomic analyses of the host cell nucleus. Further empirical investigations will then be needed to verify the functions of prospective transport proteins identified through computational analyses. Clearly there is much innovative and detailed work ahead before a rigorous understanding the mechanics of this endosymbiotic association is achieved.

Paulinella chromatophora is one of several exciting new models for investigating the process by which endosymbionts are harnessed as permanent organelles. Yoon *et al.* (2009) recently demonstrated that post-endosymbiotic speciation has occurred in *P. chromatophora,* with substantial genetic divergence among M0880/a and FK01 photosynthetic strains, opening additional avenues of comparative and experimental research on how the endosymbiosis proceeded in these independently evolving lines. With respect to our results, it will be interesting to compare the complement of potential Toc and Tic components in the respective endosymbiont genomes.

Given the evolutionary weight that has been placed on translocon similarities among primary plastid lineages, the possibility that similar complexes have evolved in *P. chromatophora* also has broader implications for understanding how the three ancient plastid lineages came into being. As we have outlined, there are diverse possibilities for evolving effective protein trafficking between host and endosymbiont. If *P. chromatophora* has adopted alternative mechanisms for protein import from those of classical primary plastids, it would tend to support the idea that similarities between Toc–Tic transport systems found in the plastids of green plants, red algae and glaucophytes result from their single origin. Alternatively, if comparable Toc–Tic translocons have evolved very early in the *Paulinella* endosymbiosis, it would indicate the presence of strong selection that drives parallel evolution of protein transport mechanisms using the same pre-existing components. Finally, we hope the results and hypotheses put forward in this study will help to stimulate further comparative and empirical research into these 'missing links' in the process of primary plastid endosymbiosis.

ACKNOWLEDGEMENTS

We are very grateful to two anonymous reviewers for their excellent comments and suggestions. A.B. is supported by funds from Wrocław University, grant BS/1018/2008.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Comparision of Toc and Tic homologs found in the genome of *Paulinella chromatophora* endosymbionts with their higher plant counterparts.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

REFERENCES

Adl S.M., Simpson A.G., Farmer M.A., Andersen R.A., Anderson O.R., Barta J.R., Bowser S.S., Brugerolle G., Fensome R.A., Fredericq S., James T.Y., Karpov S., Kugrens P., Krug J., Lane C.E., Lewis L.A., Lodge J., Lynn D.H., Mann D.G., McCourt R.M., Mendoza L., Moestrup O., Mozley-Standridge S.E., Nerad T.A., Shearer C.A., Smirnov A.V., Spiegel F.W., Taylor M.F. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology*, **52**, 399–451.

Archibald J.M. (2006) Endosymbiosis: double-take on plastid origins. *Current Biology*, **16**, R690–R692.

Archibald J.M. (2009) The puzzle of plastid evolution. *Current Biology*, **19**, 81–88.

Balsera M., Goetze T.A., Kovács-Bogdán E., Schürmann P., Wagner R., Buchanan B.B., Soll J., Bölter B. (2009) Characterization of Tic110, a channel-forming protein at the inner envelope membrane of chloroplasts, unveils a response to Ca^{2+} and a stromal regulatory disulfide bridge. *Journal of Biological Chemistry*, **284**, 2603–2616.

Bédard J., Jarvis P. (2005) Recognition and envelope translocation of chloroplast preproteins. *Journal of Experimental Botany*, **56**, 2287–2320.

Bhattacharya D., Archibald J.M. (2006) The difference between endosymbionts and organelles: response to Theissen and Martin. *Current Biology*, **16**, 1017–1018.

Bodył A., Mackiewicz P., Stiller J.W. (2007) The intracellular cyanobacteria of *Paulinella chromatophora*: endosymbionts or organelles? *Trends in Microbiology*, **15**, 295–296.

Bolte K., Bullmann L., Hempel F., Bozarth A., Zauner S., Maier U.G. (2009) Protein targeting into secondary plastids. *Journal of Eukaryotic Microbiology*, **56**, 9–15.

Castresana J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540–552.

Cavalier-Smith T. (1987) The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. *Annals of the New York Academy of Sciences*, **503**, 55–71.

Cavalier-Smith T. (1992) The number of symbiotic origins of organelles. *Biosystems*, **28**, 91–106.

Cavalier-Smith T. (1998) A revised six-kingdom system of life. *Biological Reviews*, **73**, 203–266.

Cavalier-Smith T. (2000) Membrane heredity and early chloroplast evolution. *Trends in Plant Science*, **5**, 174–182.

Cavalier-Smith T. (2006a) Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium. *Proceedings. Biological Sciences*, **273**, 1943–1952.

Cavalier-Smith T. (2006b) Rooting the tree of life by transition analyses. *Biology Direct*, **1**, 19.

Cavalier-Smith T. (2008) Evolution and relationships of algae: major branches of the tree of life. In: Brodie J., Lewis J. (Eds), *Unravelling the Algae: the past, present, and future of Algal systematics*. CRC Press, Boca Raton USA: pp. 21–55.

Cavalier-Smith T., Chao E.E. (2003) Phylogeny and classification of phylum Cercozoa (Protozoa). *Protist*, **154**, 341–358.

Cavalier-Smith T., Lee J.J. (1985) Protozoa as hosts for endosymbioses and the conversion of symbionts into organelles. *Journal of Protozoology*, **32**, 376–379.

Chen K., Chen X., Schnell D.J. (2000) Initial binding of preproteins involving the Toc159 receptor can be bypassed during protein import into chloroplasts. *Plant Physiology*, **122**, 813–822.

Chen X., Smith M.D., Fitzpatrick L., Schnell D.J. (2002) *In vivo* analysis of the role of atTic20 in protein import into chloroplasts. *Plant Cell*, **14**, 641–654.

Chen M.H., Huang L.F., Li H.M., Chen Y.R., Yu S.M. (2004) Signal peptide-dependent targeting of a rice alpha-amylase and cargo proteins to plastids and extracellular compartments of plant cells. *Plant Physiology*, **135**, 1367–1377.

Deusch O., Landan G., Roettger M., Gruenheit N., Kowallik K.V., Allen J.F., Martin W., Dagan T. (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Molecular Biology and Evolution*, **25**, 748–761.

Duy D., Wanner G., Meda A.R., von Wirén N., Soll J., Philippar K. (2007) PIC1, an ancient permease in *Arabidopsis* chloroplasts, mediates iron transport. *Plant Cell*, **19**, 986–1006.

Ertel F., Mirus O., Bredemeier R., Moslavac S., Becker T., Schleiff E. (2005) The evolutionarily related beta-barrel polypeptide transporters from *Pisum sativum* and *Nostoc* PCC7120 contain two distinct functional domains. *Journal of Biological Chemistry*, **280**, 28281–28289.

Fathinejad S., Steiner J.M., Reipert S., Marchetti M., Allmaier G., Burey S.C., Ohnishi N., Fukuzawa H., Löffelhardt W., Bohnert H.J. (2008) A carboxysomal carbon-concentrating mechanism in the cyanelles of the 'coelacanth' of the algal world, *Cyanophora paradoxa*? *Physiologia Plantarum*, **133**, 27–32.

Gentle I.E., Burri L., Lithgow T. (2005) Molecular architecture and function of the Omp85 family of proteins. *Molecular Microbiology*, **58**, 1216–1225.

Gross J., Bhattacharya D. (2009) Revaluating the evolution of the Toc and Tic protein translocons. *Trends in Plant Science*, **14**, 13–20.

Grudnik P., Bange G., Sinning I. (2009) Protein targeting by the signal recognition particle. *Biological Chemistry*, **390**, 775–782.

Guindon S., Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Hannah F., Rogerson A., Anderson O.R. (1996) A description of *Paulinella indentata* n. sp. (Filosea: Euglyphina) from subtidal coastal benthic sediments. *Journal of Eukaryotic Microbiology*, **43**, 1–3.

Hernández Torres J., Maldonado M.A., Chomilier J. (2007) Tandem duplications of a degenerated GTP-binding domain at the origin of GTPase receptors Toc159 and thylakoidal SRP. *Biochemical and Biophysical Research Communications*, **364**, 325–331.

Howe C.J., Barbrook A.C., Nisbet R.E., Lockhart P.J., Larkum A.W. (2008) The origin of plastids. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, **363**, 2675–2685.

Hust B., Gutensohn M. (2006) Deletion of core components of the plastid protein import machinery causes differential arrest of embryo development in *Arabidopsis thaliana*. *Plant Biology*, **8**, 18–30.

Inaba T., Schnell D.J. (2008) Protein trafficking to plastids: one theme, many variations. *Biochemical Journal*, **413**, 15–28.

Inaba T., Li M., Alvarez-Huerta M., Kessler F., Schnell D.J. (2003) atTic110 functions as a scaffold for coordinating the stromal events of protein import into chloroplasts. *Journal of Biological Chemistry*, **278**, 38617–38627.

Jobb G., von Haeseler A., Strimmer K. (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology*, **4**, 18.

Johnson P.W., Hargraves P.E., Sieburth J.M. (1988) Ultrastructure and ecology of *Calycomonas ovalis* Wulff, 1919, (Chrysophyceae) and its redescription as a testate rhizopod, *Paulinella ovalis* n. comb. (Filosea: Euglyphina). *Journal of Protozoology*, **35**, 618–626.

Kalanon M., McFadden G.I. (2008) The chloroplast protein translocation complexes of *Chlamydomonas reinhardtii*: a bioinformatic comparison of Toc and Tic components in plants, green algae and red algae. *Genetics*, **179**, 95–112.

Kies L. (1974) Elektronenmikroskopische Untersuchungen an *Paulinella chromatophora* Lauterborn, einer Thekamöbe mit blaugrünen Endosymbionten (Cyanellen). *Protoplasma*, **80**, 69–89.

Kies L. (1988) The effect of penicillin on the morphology and ultrastructure of *Cyanophora*, *Gloeochaete* and *Glaucocystis* (Glaucocystophyceae) and their cyanelles. *Endocytobiosis & Cell Research*, **5**, 361–372.

Kies L., Kremer B.P. (1979) Function of cyanelles in the thecamoeba *Paulinella chromatophora*. *Naturwissenschaften*, **66**, 578–579.

Kilian O., Kroth P.G. (2003) Evolution of protein targeting into "complex" plastids: the "secretory transport hypothesis". *Plant Biology*, **5**, 350–358.

Kim E., Graham L.E. (2008) EEF2 analysis challenges the monophyly of Archaeplastida and Chromalveolata. *PLoS ONE*, **3**, 2621.

Kneip C., Lockhart P., Voss C., Maier U.G. (2007) Nitrogen fixation in eukaryotes – new models for symbiosis. *BMC Evolutionary Biology*, **7**, 55.

Kneip C., Voss C., Lockhart P.J., Maier U.G. (2008) The cyanobacterial endosymbiont of the unicellular algae *Rhopalodia gibba* shows reductive genome evolution. *BMC Evolutionary Biology*, **8**, 30.

Koike H., Ikeda Y., Yusa F., Kashino Y., Satoh K. (2007) Isolation and characterization of outer and inner envelope membranes of cyanelles from a glaucocystophyte, *Cyanophora paradoxa*. *Photosynthesis Research*, **93**, 45–53.

Löffelhardt W., von Haeseler A., Schleiff E. (2007) The β-barrel shaped polypeptide transporter, an old concept for precursor protein transfer across membranes. *Symbiosis*, **44**, 33–42.

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

Marin B., Nowack E.C., Glöckner G., Melkonian M. (2007) The ancestor of the *Paulinella* chromatophore obtained a carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like gamma-proteobacterium. *BMC Evolutionary Biology*, **7**, 85.

McFadden G.I., van Dooren G.G. (2004) Red algal genome affirms a common origin of all plastids. *Current Biology*, **14**, 514–516.

Nakayama T., Ishida K. (2009) Another acquisition of a primary photosynthetic organelle is underway in *Paulinella chromatophora*. *Current Biology*, **19**, 284–285.

Neupert W., Herrmann J.M. (2007) Translocation of proteins into mitochondria. *Annual Review of Biochemistry*, **76**, 723–749.

Nowack E.C., Melkonian M., Glöckner G. (2008) Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Current Biology*, **18**, 410–418.

Nozaki H., Iseki M., Hasegawa M., Misawa K., Nakada T., Sasaki N., Watanabe M. (2007) Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Molecular Biology and Evolution*, **24**, 1592–1595.

Nylander J.A.A. (2004) *MrAIC.pl. Program distributed by the author*. Evolutionary Biology Centre, Uppsala University, Sweden.

Palmer J.D. (2003) The symbiotic birth and spread of plastids: how many times and whodunit? *Journal of Phycology*, **39**, 4–11.

Patel R., Hsu S.C., Bédard J., Inoue K., Jarvis P. (2008) The Omp85-related chloroplast outer envelope protein OEP80 is essential for viability in *Arabidopsis*. *Plant Physiology*, **148**, 235–245.

Pearson W.R. (2000) Flexible sequence similarity searching with the FASTA3 program package. *Methods in Molecular Biology*, **132**, 185–219.

Pfanzagl B., Löffelhardt W. (1999) *In vitro* synthesis of peptidoglycan precursors modified with N-acetylputrescine by *Cyanophora paradoxa* cyanelle envelope membranes. *Journal of Bacteriology*, **181**, 2643–2647.

Pollmann S., Springer A., Buhr F., Lahroussi A., Samol I., Bonneville J.M., Tichtinsky G., von Wettstein D., Reinbothe C., Reinbothe S. (2007) A plant porphyria related to defects in plastid import of protochlorophyllide oxidoreductase A. *Proceedings of the National Academy of Sciences of USA*, **104**, 2019–2023.

Qbadou S., Becker T., Bionda T., Reger K., Ruprecht M., Soll J., Schleiff E. (2007) Toc64 – a preprotein-receptor at the outer membrane with bipartite function. *Journal of Molecular Biology*, **367**, 1330–1346.

Rassow J., Dekker P.J., van Wilpe S., Meijer M., Soll J. (1999) The preprotein translocase of the mitochondrial inner membrane: function and evolution. *Journal of Molecular Biology*, **286**, 105–120.

Reinbothe S., Quigley F., Springer A., Schemenewitz A., Reinbothe C. (2004) The outer plastid envelope protein Oep16: role as precursor translocase in import of protochlorophyllide oxidoreductase A. *Proceedings of the National Academy of Sciences of USA*, **101**, 2203–2208.

Reumann S., Inoue K., Keegstra K. (2005) Evolution of the general protein import pathway of plastids (review). *Molecular Membrane Biology*, **22**, 73–86.

Reyes-Prieto A., Bhattacharya D. (2007) Phylogeny of nuclear-encoded plastid-targeted proteins supports an early divergence of glaucophytes within Plantae. *Molecular Biology and Evolution*, **24**, 2358–2361.

Rodríguez-Ezpeleta N., Brinkmann H., Burey S.C., Roure B., Burger G., Löffelhardt W., Bohnert H.J., Philippe H., Lang B.F. (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Current Biology*, **15**, 1325–1330.

Ronquist F., Huelsenbeck J.P. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Schmid A.M. (2003) Endobacteria in the diatom *Pinnularia*. I. "Scattered ct-nucleoids" explained: DAPI-DNA-complexes stem from exoplastidial bacteria boring into the chloroplasts. *Journal Phycology*, **39**, 122–138.

Schünemann D. (2007) Mechanisms of protein import into thylakoids of chloroplasts. *Biological Chemistry*, **388**, 907–915.

Smith T.F., Waterman M.S. (1981) Identification of common molecular subsequences. *Journal of Molecular Biology*, **147**, 195–197.

Soll J., Schleiff E. (2004) Protein import into chloroplasts. *Nature Reviews Molecular Cell Biology*, **5**, 198–208.

Steiner J.M., Pompe J.A., Löffelhardt W. (2003) Characterization of *apcC*, the nuclear gene for the phycobilisome core linker polypeptide $L_c^{7.8}$ from the glaucocystophyte alga *Cyanophora paradoxa*. Import of the precursor into isolated cyanelles and integration of the mature protein into intact phycobilisomes. *Current Genetics*, **44**, 132–137.

Steiner J.M., Yusa F., Pompe J.A., Löffelhardt W. (2005) Homologous protein import machineries in chloroplasts and cyanelles. *Plant Journal*, **44**, 646–652.

Stiller J.W., Reel D.C., Johnson J.C. (2003) The case for a single plastid origin revisited: convergent evolution in organellar genome content. *Journal of Phycology*, **39**, 95–105.

Teng Y.S., Su Y.S., Chen L.J., Lee Y.J., Hwang I., Li H.M. (2006) Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. *Plant Cell*, **18**, 2247–2257.

Theissen U., Martin W. (2006) The difference between organelles and endosymbionts. *Current Biology*, **16**, 1016–1017.

Tu S.L., Chen L.J., Smith M.D., Su Y.S., Schnell D.J., Li H.M. (2004) Import pathways of chloroplast interior proteins and the outer-membrane protein OEP14 converge at Toc75. *Plant Cell*, **16**, 2078–2088.

Villarejo A., Burén S., Larsson S., Déjardin A., Monné M., Rudhe C., Karlsson J., Jansson S., Lerouge P., Rolland N., von Heijne G., Grebe M., Bako L., Samuelsson G. (2005) Evidence for a protein transported through the secretory pathway *en route* to the higher plant chloroplast. *Nature Cell Biology*, **7**, 1224–1231.

Voulhoux R., Tommassen J. (2004) Omp85, an evolutionarily conserved bacterial protein involved in outer-membrane-protein assembly. *Research in Microbiology*, **155**, 129–135.

Wilm A., Higgins D.G., Notredame C. (2008) R-Coffee: a method for multiple alignment of non-coding RNA. *Nucleic Acids Research*, **36**, 52.

Yoon H.S., Reyes-Prieto A., Melkonian M., Bhattacharya D. (2006) Minimal plastid genome evolution in the *Paulinella* endosymbiont. *Current Biology*, **16**, 670–672.

Yoon H.S., Nakayama T., Reyes-Prieto A., Andersen R., Boo S.M., Ishida K., Bhattacharya D. (2009) A single origin of the photosynthetic organelle in different *Paulinella* lineages. *BMC Evolutionary Biology*, **9**, 98.