Organization, developmental dynamics and evolution of the plastid nucleoid

Naoki Sato, Kimihiro Terasawa, Kazunori Miyajima, and Yukihiro Kabeya

Department of Molecular Biology, Faculty of Science, Saitama University, Saitama 338-8570 Japan

Corresponding author: Dr. Naoki Sato
Department of Molecular Biology,
Faculty of Science,
Saitama University
255 Shimo-Ohkubo, Sakura-Ku, Saitama,
Saitama Prefecture 338-8570 Japan
Telephone: +81-48-858-3623 (direct)
Fax: +81-48-858-3384
E-mail: naokisat@molbiol.saitama-u.ac.jp

Abstract

Plastid is a semi-autonomous organelle essential in photosynthesis and other metabolic activities of plants and algae. Plastid DNA is organized into particulate structure called nucleoid with various proteins and RNA, and the nucleoid is subject to dynamic changes during the development of plant cells. Characterization of the major DNA-binding proteins of nucleoids revealed essential differences in the two lineages of photosynthetic eukaryotes, namely, the nucleoids of green plants contain sulfite reductase as a major DNA-binding protein that represses the genomic activity, while prokaryotic DNA-binding protein HU is abundant in the plastid nucleoids of the rhodophyte lineage. In addition, current knowledge on the DNA-binding proteins, as well as the replication and transcription systems of plastids are reviewed from the comparative and evolutionary points of view. A revised hypothesis on the discontinuous evolution of plastid genomic machinery is presented: in spite of the cyanobacterial origin of plastids, the genomic machinery that make function the plastid genome is fundamentally different from the counterpart in cyanobacteria.

Key words: chloroplast evolution, plastid nucleoid, plastid genomic machinery, sulfite reductase, DNA polymerase, RNA polymerase, DNA-binding protein,

Introduction

The plastid is a semi-autonomous organelle that possesses its own genetic information. The genes identified in the plastid genome (plastome) encode mostly proteins involved in photosynthesis and gene expression (Sugiura, 1992). The genetic system of the plastid depends on both plastome and nuclear genome (Tayler, 1989; Goldschmidt-Clermont, 1998; Leon *et al.*, 1998). Plastid DNA is a circular molecule but, despite common belief, it is structurally variable, even within a single plant, with respect to multimeric forms (Deng *et al.*, 1989) and molecular organization (Lilly *et al.*, 2001). The plastid DNA is organized into a particulate structure called 'nucleoid' or 'plastid nucleus' (Kuroiwa 1991), which consists of various DNA-binding proteins, plastid DNA and uncharacterized RNA (Hansmann *et al.*, 1985; Nemoto *et al.*, 1988; Yurina *et al.*, 1995; Sakai, 2001). The plastid DNA or plastid nucleoid is visualized by fluorescence microscopy after staining with various DNA-specific fluorochromes, such as 4',6-diamidino-2-phenylindole (DAPI) (Coleman, 1979; Kuroiwa *et al.*, 1981). Figure 1 shows DAPI-stained nucleoids that are viewed with a fluorescence microscope. The nucleoids of the leaf chloroplasts in seed plants are small particles with a diameter of about 0.2 μm (Fig. 1 E 1 and 2), which are sometimes fused to form a larger particle or a filament.

It is widely accepted that plastids arose from a eubacteria-like endosymbiont, closely related to extant cyanobacteria (Gray, 1992; Martin *et al.*, 1998, 2002; Turner *et al.*, 1999). Figure 2 is a phylogenetic tree showing that a progenitor of plastids was derived from a common ancestor of extant cyanobacteria. This is based on most recent databases available in March 2003: The complete nucleotide sequences of the plastid DNA have been determined in more than 29 plants and algae (for recent sequences, see Turmel *et al.*, 1999; Ogihara *et al.*, 2000; Lemieux *et al.*, 2000; Schmitz-Linneweber *et al.*, 2001), while the genomic sequences

of more than 10 species of cyanobacteria have also been completely sequenced and most of these sequences are now publicly available, either in a finished form such as CyanoBase of the Kazusa DNA Research Institute (http://www.kazusa.or.jp/cyano/) or as draft sequences such as those in the DOE Joint Genome Institute (JGI, http://www.jgi.doe.gov/JGI_microbial/html/index.html). In addition, the complete nucleotide sequences of the nuclear genome of *A. thaliana* (Arabidopsis Genome Initiative, 2000) and *O. sativa* (Feng *et al.*, 2002; Sasaki *et al.*, 2002) have been determined. The genomic sequence of the malaria parasite *Plasmodium falciparum*, which has plastids as a result of secondary endosymbiosis, was also determined (Gardner *et al.*, 2002). The nuclear, mitochondrial (Ohta *et al.,* 1998) and plastid genomes of a unicellular rhodophyte *Cyanidioschyzon merolae* were completely sequenced recently in the Cyanidioschyzon Genome Project (CGP, collaboration of several Japanese laboratories including ours, leaded by Dr. T. Kuroiwa, http://merolae.biol.s.u-tokyo.ac.jp/; Kuroiwa, 1998; Matsuzaki *et al.*, 2003). These genomic data indicate that many genes in the original endosymbiont (or plastid progenitor) were transferred to the host nucleus or lost in various different lineages. However, many of the genes that were transferred to the nucleus still encode proteins that are targeted to the plastids (Abdallah *et al.*, 2000; Martin *et al.*, 2002). In addition, various proteins involved in the maintenance, replication and transcription of the plastid genome have been lost or functionally replaced by other proteins during the evolution of the photosynthetic eukaryotes. These data are being assembled to compare the genomic contents of cyanobacteria and plastids (Sato, 2002). In this context, the information of the nuclear genomes is also important in estimating the protein complements of the plastids.

In spite of rapid development in such genomic resources, biochemical, molecular biological, or functional characterization of the plastid nucleoids or ‘plastid genomic machinery’ has been retarded, because such studies are mostly limited to the genes encoded

by the plastid genome. Recent development in the study of plastid genomic machinery was motivated by the recognition of plastid nucleoids as a counterpart of nuclear chromosome (Kuroiwa 1991). Isolation, structural analysis, as well as biochemical characterization of transcription and replication of plastid nucleoids have just been started. This review summarizes recent advances in the study of plastid genomic machinery by integrating biochemical and molecular biological studies of the plastid nucleoids as well as bioinformatic studies of plant and cyanobacterial genomes. These data will be reconstructed in the framework of our hypothesis on the evolution of plastid genomic machinery, which highlights the fundamental differences in the genomic machineries of plastids and cyanobacteria. For reviews on the plastid genome and gene expression, see Mayfield *et al.* (1995), Sugiura *et al.* (1998), Rochaix (1992) and Mullet (1988). Recent reviews on the plastid nucleoids are: Sato *et al.* (1999), Sato (2001), and Sakai (2001).

II. Methodology of studying plastid genomic maehinery

'Plastid genomic machinery' is defined as an ensemble of DNA-binding proteins and transcription factors as well as various enzymes involved in replication and transcription, which make function the plastid genome. Plastid nucleoid is a structural entity representing the plastid genomic machinery, which is isolated as particles that are sedimented by centrifugation.

A. Biochemical characterization of plastid nucleoids

1. Isolation of plastid nucleoids and identification of nucleoid proteins

Nucleoids are prepared from isolated plastids by solubilizing the envelope and thylakoid membranes (Sato *et al.*, 1997; Sakai *et al.*, 1998b). Various nonionic detergents as well as various density gradient media have been utilized. However, the most important improvement was the use of so-called 'TAN' buffer invented by the Kuroiwa's group. 'TAN' was the name of a researcher who used it first in his laboratory. TAN buffer consists of 0.5 M sucrose, 1.2 mM spermidine, 1 mM EDTA, 1 mM PMSF, 20 mM TrisHCl (pH 7.5). Sucrose and spermidine are highly effective in stabilizing the compact structure of nucleoids. Low ionic strength and chelating of divalent cations with EDTA are essential in keeping nucleoids in a highly condensed state. Addition of NaCl at a concentration above 0.2 M or several millimolar $MgCl_2$ releases major DNA-binding proteins from the nucleoids. Spermidine is a member of natural polyamines which are present in the chloroplasts as well as in bacterial or eukaryotic cells, and are important in maintaining proper activity of gene expression (Matthews, 1993; Igarashi and Kashiwagi, 1999). However, spermidine should not be added in the buffer during the isolation of plastids by Percoll density gradient, since it binds to the negatively charged surface of Percoll. A simple and convenient method of preparing nucleoids from plastids consists of solubilization of plastids with Nonidet P-40, low speed centrifugation and filtration through a membrane filter to remove bulk contaminants (starch, fragments of cell nuclei, cuticles, and other insoluble materials), and sedimentation of nucleoids by high speed centrifugation (Sato *et al.*, 1997). Use of density gradient centrifugation through metrizamide (Nakano *et al.*, 1993) or sucrose (Sakai, 2001) is effective in purifying nucleoids to some extent, but the effectiveness of the density gradient is limited by the fact that the nucleoids are highly heterogeneous with respect to buoyant density.

Unfortunately, even the purest of currently available plastid preparations contains various hydrophobic proteins that are not believed to be true constituents of nucleoids. Although isolated nucleoid preparations look highly pure by inspection with a fluorescence

microscope after staining with a fluorochrome DAPI, various hydrophobic proteins originating from various cellular membranes stick to the nucleoid preparation as evidenced by microsequencing of representative protein bands of nucleoids. This might be a reason why exhaustive proteomic analysis of the plastid nucleoids has not been undertaken. Currently, important components are identified one by one by biochemical and molecular biological means. Various DNA-binding proteins are identified by 'reconstitution' experiments, in which a protein overexpressed in *E. coli* cells and highly purified from them is mixed with purified DNA to see the formation of particulate structures that look like nucleoids under microscopy (Sato *et al.*, 2001). Another method is to use DAPI as a fluorescent probe to monitor the condensation of DNA (Sekine *et al.*, 2002). This method relies on the changes in fluorescence quenching caused by condensation of DNA and fluorochrome. Effect of addition of nucleoid proteins, which are overexpressed in *E. coli* cells, on the transcription activity of nucleoids is a good criterion to identify functionally important components of the nucleoids, an approach that complements biochemical analysis.

Composition of proteins of nucleoids isolated from various different plants and organs has been reported. Changes in protein composition or transcription activity during the development of chloroplasts (Sato *et al.*, 1997; Nemoto *et al.*, 1991), chromoplasts (Hansmann *et al.*, 1985), proplastids (Sakai *et al.*, 1997, 1998a) and amyloplasts (Sakai *et al.*, 1999). Recently, we compared the nucleoids of etioplasts and chloroplasts in cucumber and pea (Hirama *et al.*, 2003). The protein composition is mostly identical in the etioplasts and chloroplasts of cucumber cotyledons, whereas the nucleoids of etioplasts and chloroplasts in pea leaves contained qualitatively and quantitatively different proteins. The etioplasts of cucumber cotyledons are large and contains highly developed prolamellar bodies, ready to change into chloroplasts soon after the start of illumination (about 3 – 6 hr), while the etioplasts of pea leaves are small and begin development long (more than 24 hr) after the start

of illumination. This suggests that the composition of nucleoids is determined not just by the state of greening but the state of global development of the plastid.

2. *In vitro* replication and transcription in isolated plastid nucleoids

Isolated plastid nucleoids retain a high activity of replication and transcription (Sakai, 2001). Since no other effective system of *in vitro* replication that reflects *in vivo* status is available, the nucleoid system provides a unique *in vitro* replication system to study the regulation of replication. Isolated nucleoids are, however, mostly used in analyzing transcription activity. Sakai (2001) suggested that the nucleoids showed a higher transcription activity than did 'transcriptionally active chromosome' (TAC). In parallel to this, a renewed method of preparation of TAC was also reported (Krause and Krupinska, 2000). TAC is purified by gel filtration chromatography as a soluble DNA-enzyme complex, and therefore potentially represents apparatus of transcription in a purer form, although it is not as condensed as isolated nucleoids. In a typical assay, transcription reaction with isolated nucleoids continues normally as long as 60 - 80 min (Sakai, 2001; Sekine *et al.*, 2002), which is much longer than the duration of net reaction usually obtained with soluble enzyme fraction from chloroplasts. The *in vitro* transcription reaction has been believed to be 'run-on' transcription, *i.e.*, the elongation of already initiated transcription. However, the long duration of transcription reaction obtained with nucleoid preparations might suggest that re-initiation takes place in the isolated nucleoids. A test for the initiation could be the effect of heparin, a polyanionic polysaccharide that inhibits transcription initiation. But the transcription activity of plastid nucleoids is enhanced 3 - 4 times by the addition of heparin (see below), and therefore, this is not a good test for judging the problem of the occurrence of re-initiation. It is not clear if the use of artificial template can give an answer, because efficiency of exogenously added

template is hard to estimate. The question of re-initiation must be addressed by different methods.

3. Microscopic observation of nucleoids within the cell

Various microscopic methods have been applied to examine nucleoids within the cells, such as electron microscopy and fluorescence microscopy. However, no method has been available to visualize plastid nucleoids within living cells until recently. As a consequence of our study on the localization of the DNA-binding protein called PEND (see below), we realized that a cbZIP-GFP fusion protein is useful in monitoring the localization of the plastid nucleoids within the cell (Fig. 3). The PEND protein is a DNA-binding protein that was discovered from the envelope membrane of developing pea chloroplasts (Sato *et al.*, 1993). The cbZIP domain is an N-terminal small domain of the PEND protein that binds to DNA as a dimer. It includes the N-terminal pre-sequence (15 amino acid residues) that is cleaved after import into the chloroplast. This pre-sequence itself is unable to target the cbZIP domain to the chloroplast, and a longer region including nearly the entire cbZIP domain is necessary for the chloroplast targeting. The cbZIP-GFP fusion protein, which is expressed from a chimeric plasmid construct driven by the 35S promoter (Fig. 3 A), is targeted to the chloroplast stroma, and subsequently binds to the chloroplast DNA. Consequently, the nucleoids emit bright green fluorescence upon blue excitation. This DNA construct can be introduced into leaves or protoplasts by particle gun or PEG-mediated transformation, and is allowed to express transiently. Alternatively, the DNA construct can be stably transformed into *A. thaliana* by the *Agrobacterium*-mediated method. Such transgenic lines are interesting materials to visualize plastid nucleoids in intact plants. Figure 3 shows localization and morphology of plastid nucleoids within various different tissues or organs of such transgenic plant. The

nucleoids are clearly seen as minute bright spots in the whole cotyledon of transgenic *A. thaliana* (Fig. 3 B). The nucleoids in developing leaf cells (Fig. 3 C), guard cells (Fig. 3 D1), root cells (Fig. 3 E), and leaf epidermal cells (Fig. 3 F1) are visualized clearly by this method. This method is particularly useful in examining plastid nucleoids in non-photosynthetic tissues, in which plastids are small. DAPI is a good fluorochrome to visualize DNA within the cell, but the plant tissue must be fixed with glutaraldelyde or formaldehyde to allow penetration of DAPI. In addition, DAPI stains all DNA within the cell, and specific staining of plastid DNA is not possible. The cbZIP-GFP fusion protein is particularly useful in observing dynamics of nucleoids in living cells.

B. Bioinformatic knowledge-finding

Comparative genomics and proteomics are powerful tools in organellar biochemistry. Proteomic analysis include preparation of subfractions of plastid, resolution of proteins in 1- or 2-dimensional gel electrophoresis or HPLC, mass spectrometric determination of the masses of molecular ion (or equivalent) and fragment ions, and/or determination of the amino-terminal sequences of isolated single polypeptides. These data are searched against the genomic sequence database to identify the gene that encodes the polypeptide being analyzed. Proteomics of various compartments of plastids has been presented, such as thylakoid lumen (Peltier *et al.*, 2000; Schubert *et al.*, 2002), and envelope membrane (Koo and Ohlrogge, 2002; Ferro *et al.*, 2002). These studies have provided information on many new polypeptides that compose each organellar compartment. Proteomics of nucleoids is being undertaken by the collaboration of various laboratories including ours.

III. Structural and compositional diversity of plastid nucleoids

A. Morphological diversity

Morphological diversity of plastid nucleoids in variously differentiated plant cells has been documented and some examples are shown in Figs. 1 and 2: A single small nucleoid is present in the center of proplastids in meristematic cells (Miyamura *et al.*, 1986), whereas nucleoids are replicated while attached to the envelope membranes (Fig. 1 E1, Fig. 3 C; Sato *et al.*, 1993). In mature leaf cells, only a small number of nucleoids are present within the stroma in close association with thylakoid membranes (Fig. 1 E2, Fig. 3 D1). Nucleoids in epidermal cell plastids and root cell plastids are also morphologically diverse (Fig. 3 E and F1). In addition, the replication cycle of plastid in seed plants is entirely uncoupled with the division cycle of the cell in which plastids reside: The nucleoids are extensively replicated in developing plastids of immature leaf cells. Then they are distributed to daughter plastids by segregation process mediated by envelope membrane. The division of algal plastids is synchronized with the cell cycle (Fig. 1 C), and is different from the division of plant plastids in this respect.

Phylogenetic diversity of plastid nucleoids has also been extensively studied. In the pioneering study by the Kuroiwa group, various different forms of plastid nucleoids were described (Kuroiwa *et al.*, 1981). The nucleoids of angiosperm chloroplasts are distributed throughout the stroma as small particles, while the nucleoids of brown algae form a ring-shaped structure. The chloroplast nucleoids of a green alga *Chlamydomonas reinhardtii* are also particulate and distributed over the stromal spaces (Fig. 1 D2). A mutant of *C. reinhardtii* is known, in which the nucleoids are aggregated as a large mass within the chloroplast (Misumi *et al.*, 1999). Although the partition of nucleoids during the chloroplast division is uneven in this mutant, each daughter chloroplast regains the original amount of

nucleoids by subsequent replication. Among the red algae, an interesting correlation between the morphology of nucleoids and cell complexity, and hence phylogeny, has been found. In the primitive unicellular red algae such as *Cyanidioschyzon merolae* (Fig. 1 C1) and *Cyanidium caldarium*, a single nucleoid is present at the center of the chloroplast. These algae divide by binary fission (Kuroiwa, 1998). Another unicellular species *Galdieria sulfuraria* divides by ternary fission and has multiple particulate nucleoids within the chloroplast. Multicellular thallous red algae have nucleoids scattered along the periphery of chloroplasts (Kuroiwa *et al.*, 1981). The morphological diversity of nucleoids is clearly identified, but we still do not know the mechanism underlying the morphological differences in nucleoids of various different organs and organisms, not even the physiological relevance of the morphology of nucleoids. The only case in which the localization of nucleoids is explained in molecular terms is that the PEND protein anchors nucleoids to the envelope membranes in developing pea chloroplast (Sato *et al.*, 1993; see below).

B. DNA-binding proteins

Analysis of plastid nucleoids in seed plants revealed a striking flexibility of protein composition in various types of plastids, such as chloroplasts, etioplasts, proplastids, and chromoplasts (Hansmann *et al.*, 1985; Nemoto *et al.*, 1988, 1989; Sato *et al.*, 1993, 1997). In the plastids of vascular plants, various transcription factors have been known: sigma factors transcribe mainly photosynthetic genes together with the PEP (Isono *et al.*, 1997; Fujiwara *et al*., 2000; Allison, 2000; Kanamaru *et al.*, 2001); PTF1 is known as a *trans*-acting factor for the *psbD* light-responsive promoter (Baba *et al.*, 2001); CDF2 is known to regulate the *rrn* gene (Bligny *et al.*, 2000). The PEND protein is a DNA-binding protein that binds to the envelope membrane and anchors cpDNA with ‘TAAGAAGT’ motif (Sato *et al.*, 1993, 1997;

Sato and Ohta, 2001). Additionally, some sequence non-specific DNA-binding proteins are known to constitute the plastid nucleoids. The HC protein was identified by a database search as a basic copper-containing protein (Crevel, 1989; Sato, 2001), and bifunctional proteins such as sulfite reductase (abbreviated as 'SiR', Sato *et al.*, 2001) and the CND41 protein (Nakano *et al.*, 1997) have been described.

1. Prokaryotic DNA-binding proteins

Typical free-living prokaryotes have a whole set of major DNA-binding proteins such as HU/IHF, H-NS/StpA, Fis, and Dps (McLeod and Johnson, 2001; Azam and Ishihama, 1999; Azam *et al.*, 2000). Here, HU and IHF belong to a large protein family (Bonnefoy and Rouvière-Yaniv, 1991), while H-NS (Dame and Goosen, 2002) and StpA are members of another large protein family (Table 1). In many gamma proteobacteria such as *E. coli*, both HU and IHF are present, and each of these is a heterodimer consisting of alpha and beta subunits (Dame and Goosen, 2002; McLeod and Johnson, 2001). However, in many other bacteria, only a single copy or two copies of *hupA*-like gene is (are) present, and it is hard to classify the encoded protein(s) into HU or IHF (Fig. 4 A). Most cyanobacteria also contain a single HU. A single Dps is present in *E. coli*, while multiple Dps are present in some species of cyanobacteria such as *Anabaena* sp. PCC 7120 (Kaneko *et al.*, 2001). However, cyanobacteria lack H-NS/StpA and Fis (Table 1). There are a number of transcription factors, and several tens of response regulators in prokaryotes except in some species with very small genome (COMTRAF database: http:// comtraf.lab.nig.ac.jp/).

In bacterial nucleoids, the HU protein is known as the major DNA-binding protein. The name 'HU' seems to come from 'heat unstable' but this protein is in fact heat stable (Rouvière-Yaniv and Gros, 1975). No gene encoding HU has been detected in both plastid and

nuclear genomes of *A. thaliana*, but a homolog is present in the plastid genome of a primitive red alga *C. merolae* (Kobayashi *et al.,* 2002) and a cryptophyte *Guillardia theta* (formerly, *Cryptomonas* phi, Wang and Liu, 1991; Grasser *et al.*, 1997; Wu and Liu, 1997). Since the plastids of chromophytes are thought to originate from the plastids of rhodophyte-like endosymbiont, the plastomes of rhodophytes and chromophytes (brown algae, diatoms, cryptophytes etc.) are considered as members of a single lineage (Douglas *et al.*, 1991; Nelissen *et al.*, 1995; Ohta *et al.*, 1997; Moreira *et al.*, 2000). The HlpA protein (for HU-like protein) of *Guillardia theta* has been characterized and was shown to complement mutation in endogenous HU in *Bacillus subtilis* (Grasser *et al.*, 1997). All plastid-encoded HU-like proteins are similar to the cyanobacterial HU proteins. The *hlpA* gene has not been detected in any of the plastid genomes of green algae and land plants (Table 1), though ESTs encoding an HU homolog were found in *Chlamydomonas* and *Toxoplasma*. The gene for the HU homolog that must have been present in the endosymbiont genome was lost during the evolution of plastids in the chlorophyte lineage (see below, Sato, 2001). Phylogenetic analysis indicates that the HU-like proteins of the rhodophyte lineage (*Cyanidioschyzon*, *Toxoplasma*, and *Guillardia*) are sister to the cyanobacterial HU proteins, although the bootstrap value is not very high (Fig. 4 A). This is consistent with the hypothesis that the HU protein was transmitted from an ancestral cyanobacterium to the progenitor of plastids. The low bootstrap confidence level may be due to the very small size of the HU protein, as well as high evolution rates of the chloroplast genome. However, the monophyly of cyanobacterial HU and rhodophyte-chromophyte HU-like proteins is also supported by maximum parsimony analysis (results not shown). Interestingly, the HU-like protein of *Chlamydomonas* is not related to the cyanobacterial or rhodophytan HU proteins. This suggests that *Chlamydomonas* acquired the gene for the HU-like protein from an unknown prokaryote, and this observation is consistent with the non-plastidic localization of the *Chlamydomonas* HU-like protein

(unpublished data).

We performed immunoblot analysis with antibodies against the HU-like protein of *C. merolae*. HU and HU homolog were detected in the isolated nucleoids of *A. variabilis* M3 (cyanobacterium) and *C. merolae* but not in the nucleoids of *P. sativum* and *P. patens* (Sato *et al.*, 2003). A recent study indicated that the HU-like protein from *C. merolae* could also compact DNA (Kobayashi *et al.*, 2002), but the mode of compaction seems different with HU and SiR (Fig. 4 B and C; see below). There is a report of immunological detection of an HU-homolog in spinach chloroplasts (Briat *et al.*, 1984), and the presence of a histone-like protein in pea choroplast nucleoids (Yurina *et al.*, 1995). However, a recent computer survey in the complete genome of *A. thaliana* as well as all known plant and algal nuclear sequences failed to identify such an HU-homolog. Therefore, we can conclude that HU was lost from the chloroplast in the green lineage.

Other putative transcription factors with strong similarity to OmpR (Ycf27 and Ycf29), NtcA (Ycf28) and RbcR (Ycf30) were also found in the plastomes of the rhodophyte lineage (Sato, 2001 and Table 1). These transcription factors are known to bind to a specific target sequence and regulate expression of target genes in prokaryotic cell, but no homologs have been found in the nuclear and plastid genomes of green plants. A comparative analysis of the transcription factors encoded in the nuclear genome of *A. thaliana* was performed (Riechmann *et al.*, 2000), but no prokaryotic transcription factors were detected.

DnaB protein is encoded by many plastomes of the rhodophyte lineage. This protein acts as a DNA helicase during the replication or maybe in recombination in prokaryotes (Benkovic *et al.*, 2001;Yang *et al.*, 2002; Bujalowski, 2003). In contrast, DnaA protein, which is essential for bacterial replication initiation (Messer *et al.*, 2001; Erzberger *et al.*, 2002) , has never been shown to be encoded in the plastome or in the nuclear genome of plants and algae. Since the replication of plastome involves D-loops (Kolodner and Tewari,

1975), the mechanism of replication initiation must be fundamentally different from that of bacterial replication. In the complete genome of *A. thaliana*, several genes that encode DNA polymerases similar to bacterial DNA polymerase I have been detected, a fact which is consistent with the biochemical analysis of DNA synthesis activity in plastids (Sakai, 2001; Gaikwad *et al.*, 2002). This point will be discussed in detail below.

2. Structural or sequence-nonspecific DNA-binding proteins of plastid nucleoids of plants

Sulfite reductase (SiR), which is an enzyme that catalyzes assimilatory reduction of sulfite to sulfide, has been known as a stromal enzyme within the chloroplast (Aketagawa and Tamura, 1980). However, in the nucleoids of pea, SiR was found to be one of the major DNA-binding proteins (Sato *et al.*, 1997; Cannon *et al.*, 1999; Sato *et al.*, 2001). In addition, SiR is able to compact DNA as well as isolated nucleoids *in vitro* (Sato *et al.*, 2001, Sekine *et al.*, 2002). Sekine *et al.* (2002) suggested that SiR reversibly regulates the global transcription activity of plastid nucleoids through changes in DNA compaction. Therefore, DNA-binding is the second function of SiR in higher plant plastids. Cannon *et al.* (1999) also showed that SiR inhibits DNA synthesis *in vitro*. Interestingly, in yeast mitochondria, several enzymes of tricarboxilic acid cycle and amino acid metabolism are known to associate mitochondrial DNA and essential in maintaining the mitochondrial DNA (Zelenaya-Troitskaya *et al.*, 1995; Kaufman *et al.*, 2000). Since DNA binding is not known for bacterial SiR, we wondered if SiR from the lower plants or algae has the ability of compacting DNA. We performed immunoblot analysis of plastid nucleoids from pea, *P. patens* (moss) and *Cyanidioschyzon merolae* (red alga) as well as of nucleoids from a cyanobacterium *Anabaena variabilis* strain M3, with antibodies raised against maize SiR (Sato *et al.*, 2003). Sulfite reductase was

detected abundantly in the nucleoids of pea and *P. patens*, but only faintly in the nucleoids of *C. merolae* and *A. variabilis*. Further analysis of nucleoids of *C. merolae* suggested that the small amount of SiR that was detected in the nucleoids was easily washed out with a medium without spermidine, which is not the case for pea nucleoids (unpublished results). Sulfite reductase may have acquired the ability of tightly binding DNA during the evolution of land plants. Experiments with green algal nucleoids will be interesting in this respect.

Available evidence including these findings suggests, therefore, that SiR (eukaryotic major DNA-binding protein in plastid nucleoid) replaced HU (prokaryotic major DNA-binding protein) during the evolution of green plants, although the roles of these two proteins are not identical. Here, we present two results suggesting the functional differences in HU and SiR (Fig. 4). In the *in vitro* compaction experiments, HU forms large particles with DNA (Fig. 4 B), whereas SiR forms small tightly packed particles (Fig. 4 C). After staining with DAPI, the HU-DNA complexes emit bright fluorescence of DAPI, whereas the fluorescence intensity of the DAPI-stained SiR-DNA complexes is significantly attenuated (Sekine *et al.*, 2002). The HU protein supports or enhances the transcription activity of nucleoids in *C. merolae*, but it exhibited inhibitory effects on the transcription of nucleoids in pea, which normally do not contain HU protein (Fig. 4 D). SiR always lowers the transcription activity of nuleoids in *C. merolae* and pea (Fig. 4 E). The activation of transcription by HU was also reported in *E. coli* (Rouvière-Yaniv and Gros, 1975). These results suggest that the HU-DNA complex is bulky and keeps large space within itself that allows enzymes to function normally, while the SiR-DNA complex is tightly packed and is kept in inactive state.

Another well-studied DNA-binding protein of chloroplast is the CND41 protein, originally identified in tobacco (Nakano *et al.*, 1993, 1997). This protein binds to chloroplast DNA without specificity with respect to sequence. The DNA-binding domain is not similar to

known domains, but the CND41 protein is also similar to acidic proteinases such as subtilisin. The CND41 protein indeed exhibited proteinase activity with RubisCO as a substrate (Murakami *et al.*, 2000). But the natural substrate within the chloroplast remains to be identified. This protein is preferentially localized in non-photosynthetic cells such as cultured cells, and lowers the transcription activity of chloroplast *in vitro*. It is thought that the CND41 protein acts as a general repressor of the transcription in chloroplasts, but the relationship between the proteinase activity and the DNA-binding activity remains to be uncovered.

3. Sequence-specific DNA-binding proteins of plastid nucleoids

The PEND protein is a DNA-binding protein that binds to the envelope membrane and anchors cpDNA with the 'TAAGAAGT' motif (Sato *et al.*, 1993, 1997; Sato and Ohta, 2001). This protein was first discovered as a 130-kDa DNA-binding protein in the envelope membrane of developing pea chloroplasts (Sato *et al.*, 1993). A cDNA encoding a 70-kDa polypeptide was isolated (Sato *et al.*, 1998). Although the exact molecular structure of the 130-kDa protein has not been completely elucidated because of limitation of available analytical methods, the 130-kDa protein is believed to be a dimer of 70-kDa polypeptides, since in vitro-synthesized 70-kDa polypeptide yields a 130-140 kDa band with a low yield (unpublished results) and affinity-purified 130-kDa protein preparation contained significant amounts of a 70-kDa polypeptide (Sato *et al.*, 1998). However, we still did not identify the condition of dimerization or dissociation. The PEND protein is a special type of bZIP (basic region plus zipper region) protein that we named cbZIP (Sato and Ohta, 2001). A comparison of cDNA sequence and the N-terminal sequence of the isolated native PEND protein identified that there is a short (15 residues) pre-sequence, which is cleaved after import into chloroplasts. However, this short sequence is not enough for the import of this protein into

the chloroplast. A longer sequence including the N-terminal cbZIP region is necessary for its import (Terasawa *et al.*, 2003; Fig. 3). A *Brassica* protein known as GSBF1 (Waldmüller *et al.*, 1996) was turned out to be a homolog of the PEND protein. Recent studies from this laboratory suggest that the PEND protein is also present in various seed plants, such as *A. thaliana*, *Brassica napus* (2 copies), *Prunus yedoensis* (cherry), and cucumber. We identified the cDNA or genomic sequences of these homologs, and we also found similar targeting of the homologous proteins of pea, *A. thaliana* and *B. napus*. A series of GFP (green fluorescent protein)-fusion experiments suggested that the N-terminal region including the pre-sequence and cbZIP region is targeted to the chloroplast stroma, while the C-terminal transmembrane region is necessary for retention of the PEND protein in the envelope membrane (Fig. 3 A). The import into the envelope membrane was confirmed by *in vitro* import experiments. Although information concerning the role of the PEND protein is still limited, the protein is present only in developing chloroplasts, in which the nucleoids are actively replicated while being attached to the envelope membrane. The PEND protein might therefore be involved in the replication of chloroplast DNA, or repression of transcription during the amplification of chloroplast DNA.

The PTF1 protein was isolated as a bHTH DNA-binding protein that binds to the AAG-box, which is located upstream the *psbDC* promoter in the chloroplast genome (Baba *et al.*, 2001). This *cis*-element is involved in the light responsiveness of this promoter and the DNA-binding sites were identified by DNase I footprinting (Kim and Mullet, 1995). The importance of the AAG element in the light-dependent expression of the *psbDC* promoter has been demonstrated in transplastomic tobacco (Thum *et al.*, 2001).

The CDF2 was identified as a *trans*-acting factor that binds specifically to a promoter region of the chloroplast rRNA gene, and regulates the transcription from the two promoters depending on the developmental stages of spinach leaves, and is likely to be a key regulator

of the chloroplast development (Bligny *et al.*, 2000). Unfortunately, no molecular data is available for this putative protein.

PD3 was isolated as a DNA-binding protein with five AT-hooks and eight CxxC motifs (Sato *et al.*, 1995). This protein binds to the AT-rich sequences of the chloroplast DNA. This is a homolog of the ENBP1, which was originally isolated as a DNA-binding protein that has affinity for the promoter of *ENOD12* gene (Christiansen *et al.*, 1995).

C. Enzymes

1. DNA polymerase

The replication of the chloroplast have long been studied by several groups (Kolodner and Tewari, 1975; Meeker *et al.*, 1988; Heinhorst and Cannon, 1993; Kunnimalaiyaan and Nielsen, 1997; Mühlbauer *et al.*, 2002), but the purification of DNA polymerase was not successful until recently. DNA topoisomerase I (Nielsen and Tewari, 1988) and DNA primase (Nielsen *et al.*, 1991) were purified and partially characterized. DNA topoisomerase II was co-localized with the nucleoids at least at some stages of chloroplast development (Marrison and Leech, 1992). Sakai (2001) used isolated tobacco nucleoids to study the replication of plastid genome, and found that a 116-kDa polypeptide is responsible for the DNA synthesis activity. Curiously, they demonstrated that the DNA polymerase of mitochondria is very similar to the counterpart of chloroplasts, with respect to molecular mass, dependence on pH and salt concentration, and susceptibility to inhibitors. Both of these DNA polymerases are quite similar to DNA polymerase I of *E. coli*. They claimed that the mitochondrial DNA polymerase was the origin of the chloroplast enzyme, which must have been generated by gene duplication. It is known that the DNA polymerase of animal cells is

gamma-type enzyme, but no gamma-type polymerase has been detected in plants (Fliée *et al.*, 2002). This result is also supported by the analysis of the complete genome sequence of *A. thaliana*. Recently, a DNA polymerase of pea chloroplasts was purified and characterized (Gaikwad *et al.*, 2002). This enzyme is similar to the Class A polymerase such as DNA polymerase I of *E. coli*. A rice DNA polymerase (OsPolI-like) was also characterized by cDNA cloning and genomic analysis (Kimura *et al.*, 2002). This enzyme was detected in isolated chloroplast fraction, but no rigorous test was performed if it is only localized in the chloroplasts or present also in the mitochondria. The rice polymerase is a 100-kDa polypeptide, having a 3'-5' exonuclease domain and a polymerase domain (Fig. 5 A). The 5'-3' exonuclease domain was not present in the rice polymerase, although the authors claimed that the N-terminal domain of OsPolI-like corresponded to the 5'-3' exonuclease domain of *E. coli* DNA polymerase I, which is not substantiated by sequence comparison and motif analysis in our hands (Fig. 5 A). Similar DNA polymerases are also identified in the genome of *A. thaliana* (Arabidopsis Genome Initiative, 2000). There are two copies of putative polymerases, each having a 3'-5' exonuclease domain and a polymerase domain (Fig. 5A). Another putative protein having a helicase domain and a polymerase domain was also found. In *A. thaliana*, two putative genes that encode a 5'-3' exonuclease were found. Since no other gene that can encode DNA polymerase (except nuclear ones) was found, these two sets of DNA polymerases are likely to be organellar replication enzymes. In a recently sequenced genome of *C. merolae* (red alga), two putative Class A DNA polymerases were identified. One is similar to the plant enzyme, having a 3'-5' exonuclease domain and a polymerase domain (enzyme B), whereas another (enzyme A) is more similar to the bacterial DNA polymerase I with an additional 5'-3' exonuclease domain. However, we still do not know which one of the three enzymes is mitochondrial and which one is chloroplastic. Some of the polymerases are predicted to have a targeting sequence for mitochondria or for

chloroplasts as noted in Fig. 5 A, but the organellar targeting is currently the focus of intensive experimental studies in many laboratories.

Phylogenetic analysis (Fig. 5 B) indicates that these Class A polymerases of plants and algae are monophyletic but are not related to the Class A polymerases of cyanobacteria or alpha-proteobacteria. There was a misunderstanding on this point (Kimura *et al.*, 2002), but the proposed close relationship of cyanobacterial enzymes with plant enzymes was not substantiated by detailed analysis using new additional sequences (Fig. 5 B). This is a very important observation regarding the origin of the replication system in photosynthetic eukaryotes. In addition, no homologs of DNA polymerase gamma (as found in animals and fungi) or prokaryotic DNA polymerase III (as found in *E. coli* and cyanobacteria) have been detected in the genome of *A. thaliana* or *C. merolae*. Therefore, the origin and evolution of the replication system of eukaryotes must be very complicated with many exchanges or transfer of exogenous replication systems. This point will be discussed in the final part of this article.

2. Prokaryotic RNA polymerase and sigma factors

The prokaryote-type RNA polymerase (RNAP) consists of the core enzyme and a sigma factor. The prokaryotic (or multisubunit) core enzyme of chloroplast RNAP, also called PEP for plastid-encoded polymerase, was purified in maize (Hu and Bogorad, 1990; Hu *et al.*, 1991) and mustard (Pfannschmidt and Link, 1994, 1997; Pfannschmidt *et al.*, 2000), and was found to consist of four major subunits called alpha, beta, beta', and beta''. The beta' and beta'' subunits correspond to the N-terminal and C-terminal halves of the bacterial beta' subunit. They are encoded by separate genes on the chloroplast genome (*rpoC1* and *rpoC2*, respectively) as they are in cyanobacteria, except in a red alga *C. merolae*, in which the two

genes are secondarily fused to form a single beta' gene (Ohta *et al.*, 1999). The alpha subunit is encoded by the chloroplast genome in most algae and plants, while it is encoded by the nuclear genome in a moss *P. patens* (Kobayashi *et al.*, 2002).

The prokaryotic RNAP of mustard changes its subunit composition during the development of the plant (Pfannschmidt and Link, 1994, 1997; Satoh *et al.*, 1999). The RNAP in mature leaves consists of four subunits as described above, while the enzyme in developing (or etiolated) leaves contains a number of additional polypeptides that are to be characterized. In addition, the catalytic beta subunit is phospharylated in mustard (Pfannschmidt *et al.*, 2000; Ogrzewalla *et al.*, 2002). This makes the enzyme sensitive to rifampicin, a potent inhibitor of prokaryotic RNAP, which otherwise do not inhibit chloroplast RNAP. The phosphorylation of a sigma factor is regulated by the redox state within the chloroplast, which is mediated by glutathione (Baginsky *et al.*, 1999; Baena-Gonzalez *et al.*, 2001). This is particularly important in the chloroplast, in which transcription is activated when photosynthesis is actively performed, so as to supply the reaction center protein D1 of the photosystem II and other components that are rapidly turned over as a result of photoinhibition.

Recently, nuclear genes encoding plastid sigma factors have been identified in algae and land plants (Tanaka *et al.,* 1996; Tanaka *et al.*, 1997; Kestermann *et al.*, 1998; Tan and Troxler, 1999; Oikawa *et al.*, 2000). These sigma factor genes (*SIG*) are members of a gene family in each species (Allison, 2000). There are six *SIG* genes in *A. thaliana*. The *SIG* genes in plants appear to originate from a sigma factor gene of an ancestral cyanobacterium-like endosymbiont. In the moss *P. patens*, two genes, each encoding a plastid sigma factor, were identified and named *PpSig1* and *PpSig2*, respectively (Hara *et al*., 2001a; Hara *et al*., 2001b). It is known that the *SIG* genes in plants and cyanobacteria are subject to various types of regulation such as tissue-specific (Isono *et al.*, 1997; Tan and Troxler, 1999;

Beardslee *et al.*, 2002), plastid development-specific (Lahiri and Allison, 2000), light-dependent (Kanamaru *et al.*, 1999; Morikawa *et al.*, 1999) and circadian clock-controlled expression (Kanamaru *et al*., 1999; Morikawa *et al*., 1999). In addition, different sigma factors play distinct roles under the normal light condition (Oikawa *et al.*, 1998; Tan and Troxler, 1999; Beardslee *et al*., 2002). The *sig2* mutant of *A. thaliana* has pale-green leaves and exhibit lowered activity of synthesis of chlorophyll (Shirano *et al.*, 2000). The transcription of various tRNA genes is lowered in this mutant, and the decreased $tRNA^{Glu}$ content is a reason for the lowered synthesis of chlorophyll, since tRNA-linked glutamate is the substrate for the synthesis of aminolevulinic acid, which is, in turn, the key precursor to the synthesis of tetrapyrroles (Kanamaru *et al.*, 2001). The expression of the NEP (see below) is enhanced in the *sig2* mutant, probably by a feedback mechanism.

3. Phage-type RNA polymerase

Phage-type RNA polymerase (RPOT) is encoded in the nucleus of various eukaryotes, and is known to be targeted to mitochondria or plastids in vascular plants (Hess and Börner, 1999). The one transported to plastids is called NEP, nuclear-encoded RNA polymerase. In dicotyledons, three types of *RPOT* genes have been identified. The *RpoT;1* gene of *A. thaliana* (Hedtke *et al*., 1997) and the *RpoT1* or *RpoT-A* gene of *N. sylvestris* (Kobayashi *et al.*, 2001a; Hedtke *et al.*, 2002), with orthologs in the majority of eukaryotic organisms, encode the mitochondrial RNA polymerase. This type of RPOT has been identified in a number of land plants and algae (Cermakian *et al.*, 1996). The *RpoT;3* gene of *A. thaliana* encodes a plastid-targeted RPOT (NEP), which has been identified in only angiosperms (Hedtke *et al.*, 1997). In addition, the products of the *RpoT;2* gene of *A. thaliana* and the *RpoT2* or *RpoT-B* gene of *N. sylvestris* are found to be dually targeted to mitochondria and

plastids (Hedtke *et al.*, 2000; Kobayashi *et al.*, 2001b; Hedtke *et al.*, 2002). In the angiosperm plastids, the NEP and PEP play distinct roles in transcription (Hajdukiewicz *et al.*, 1997; Hess and Börner, 1999; Sakai *et al.*, 1998; Krause *et al.*, 2000). The class I genes (mainly photosystem genes and *rbcL*) are exclusively transcribed by the PEP, whereas the class II genes (genes encoding divergent but photosynthesis-related functions other than photosystems I and II) are transcribed by both PEP and NEP, and the class III genes (mainly housekeeping genes) are transcribed by the NEP. In addition, distinct promoters are recognized by the two RNA polymerases. There are two classes of promoters, named CT (consensus-type) and NC-II (non consensus-type) promoters, in various plastid genes in angiosperms. The CT promoters are similar to the bacterial sigma 70-type promoters and recognized by the PEP (Satoh *et al.*, 1999), while the NC-II promoters are similar to the mitochondrial promoters and recognized by the NEP (Kapoor *et al.*, 1997; Liere and Maliga, 1999).

The high similarity of sequences as well as the excellent conservation of the positions of introns in the flowering plant *RPOT* genes reported to date suggests that the plastid-type *RPOT* genes are generated from the mitochondrion-type *RPOT* gene by gene duplication. Analysis of RPOT sequences in lower land plants suggested that the RPOTs of flowering plants are clearly classified into two groups, each representing plastid- and mitochondrion-localized enzymes. The latter group also includes the so-called ‘dual targeting’ enzymes. The moss RPOTs form a sister group to the cluster of all flowering plant RPOTs (Kabeya *et al.*, 2002). Further phylogenetic analysis suggested that the gene duplication generating the plastid isozymes occurred in the vascular plant lineage after the separation of angiosperms from gymnosperms (Kabeya *et al.*, 2002; Kabeya *et al.*, 2003).

We previously reported that the two PpRPOTs are localized to only mitochondria (Kabeya *et al.*, 2002). However, Richter *et al.* (2002) reported that the identical two

PpRPOTs are dually targeted to mitochondria and plastids. Namely, both PpRPOTs are targeted to mitochondria when they are translated from the second methionine, whereas they are targeted to plastids when they are translated from the first one (Fig. 6 C). In these experiments, translation was forced to start from the AUG codon just downstream the translation leader sequence. In contrast, Kabeya *et al.* (2002) used the natural context of translation initiation with the native 5' leader sequences. The effect of 5' untranslated region has been confirmed in our recent experiments (Fig. 5 B; Kabeya and Sato, manuscript in preparation).

However, cytological analysis on the localization of the two RPOTs in the moss cells revealed different results. Immunoblot experiment of isolated mitochondria and chloroplasts using antibodies raised against the two RPOTs detected RPOT2 in the mitochondrial fraction. No signal was detected in the chloroplast fraction (Kabeya and Sato, manuscript in preparation). A RPOT-GFP fusion in stably transformed moss protonemata and gametophores was localized only in mitochondria. The transcription activity of isolated mitochondria was not inhibited by tagetitoxin, an inhibitor of prokaryotic RNA polymerase (Mathews and Durbin, 1990), while the transcription activity of isolated chloroplasts was totally abolished by the addition of tagetitoxin, indicating that NEP is not present in the moss chloroplasts (Fig. 6 A). Therefore, a plausible explanation is that the first methionine codon is not used as an initiator *in vivo*, although the amino acid sequence beginning from it has a property of plastid targeting sequence (Fig. 6 C). The dual targeting to mitochondria and chloroplasts is now known in various enzymes in angiosperms (Peeters and Small, 2001; Watanabe *et al.*, 2001). In all these dually targeted proteins known to date, the use of the first methionine results in targeting to chloroplasts, and the use of the second methionine results in targeting to mitochondria. In many cases, the first methionine is in fact used as the initiation codon *in vivo*. However, the dual targeting of the angiosperm RPOTs will have to be

re-examined with caution, because the translation efficiency of the first methionine codon *in vivo* has never been tested. This is especially important in the case of angiosperm RPOTs, in which the situation is complicated by the fact that the two others of the three isozymes are targeted to mitochondria and chloroplasts, respectively.

Currently, no transcription factors that act with the NEP have been reported. In yeast mitochondria, Mtf1 or mtTFB is necessary in determining the specificity of the RNA polymerase (Karlok *et al.*, 2002). Mtf1 is partially similar to the sigma factors of prokaryotic RNAP. Although no Mtf1 homolog has been found in plants, it is interesting that a maize sigma factor ZmSig2B is localized in mitochondria as well as chloroplasts (Beardslee *et al.*, 2002). This raises a possibility that some of the plant sigma factors may also function in the chloroplast as a co-factor of the NEP.

The physiological function of RPOT is not limited to transcription of house-keeping genes. In yeast mitochondria, the product of *Rpo41* gene encoding the RPOT plays a role as a primase in the replication of mitochondrial genome (Wang and Shadel, 1999). Involvement of RNA polymerase in replication was also shown in trypanosome mitochondria (Grams *et al.*, 2002). It is possible that the NEP also functions as the primase during the replication of the plastid genome. Currently, various tag-lines of *A. thaliana* that have a T-DNA insertion in one of the *RPOT* genes are being analyzed in various laboratories.

4. Post-transcriptional modification

Isolated plastid nucleoid contains significant amount of RNA (Nemoto *et al.*, 1988; Sakai, 2001; Yurina *et al.*, 1995), and is likely to be the site of post-transcriptional modification of RNA. The post-transcriptional modification in the plastids includes *cis*- and *trans*-splicing of group II introns, RNA editing, processing at the 3' and 5' ends, and polyadenylation (or

exaxtly, addition of GA-rich tail by polynucleotide phosphorylase. Lisitsky *et al.*, 1996). The components involved in the splicing is analyzed genetically in a green alga *Chlamydomonas reinhardtii* (Rivier *et al.*, 2001). The analysis of *in vitro* editing of the *psbL* and *ndhB* genes suggests that different protein factors are necessary for the editing of different sites, as well as common factors (Hirose and Sugiura, 2001). The RNA binding protein cp31 is also involved in the editing of both sites. Various RNA-binding proteins as well as specific nuclear-encoded factors are known to be involved in the mRNA processing and degradation (Drager *et al.*, 1999; Monde *et al.*, 2000; Rochaix, 2001; Drapier *et al.*, 2002). The processing at the 3'-untranslated region determines the stability of chloroplast RNA (Komine *et al.*, 2002). Various RNA-binding proteins, such as cp29, cp31 and cp33, are reported in the chloroplast in tobacco (Hirose *et al.*, 1994) and *A. thaliana* (Ohta *et al.*, 1995). These RNA-binding proteins contain two RRM (RNA-recognition motif) consisting of about 80 amino acid residues. RNA-binding proteins of another type (GRP) containing a single RRM and a C-terminal glycine-rich domain are present in the nucleus or cytoplasm and are involved in responses to various environmental stresses and circadian rhythm (Lorkovic and Barta, 2002). The GRP type proteins are also widely present in cyanobacteria, but detailed phylogenetic analysis indicated that the cyanobacterial GRP and plant GRP or chloroplast RRM proteins are not directly related (Maruyama *et al.*, 1999). Some of the chloroplast RNA-binding proteins are enriched in the nucleoids (unpublished results). However, the majority of the RNA-binding proteins are recovered in the soluble fraction after sedimentation of nucleoids. Nakamura *et al.* (2001) showed that the *psbA* transcript is preferentially bound to various cpRBPs and is stabilized as ribosome-free mRNP. The physiological roles of the chloroplast RNA-binding proteins are still obscure, although roles in RNA editing (Hirose and Sugiura, 2001) and processing (Yang *et al.*, 1996; Monde *et al.*, 2000) were suggested.

IV. Evolution of plastid genomic machinery

Plastids are descendants of an ancestral endosymbiont, which is thought to be similar to an ancestor of extant cyanobacteria. The plastid genome originates from the genome of endosymbiont, but the size of the plastid genome is more than ten times smaller than that of the genomes of present-day cyanobacteria. Sato (2001) put forward a hypothesis on the discontinuous evolution of plastid genetic machinery: most transcription regulators that had been present in the cyanobacterium-like ancestor were rapidly lost after the primary endosymbiosis, while plastids of land plants acquired transcription factors from the eukaryotic host. This evolutionary change of gene regulation in the plastid genome, which is also the basis of nuclear control of plastid gene expression, must have accompanied various changes in regulatory elements and interactions between components of transcriptional apparatus and regulators. The elucidation of this regulatory shift requires determination of the time and the phyla in which each change occurred.

A. Cyanobacterial-plastid lineage

The progenitor of plastid is estimated to be an ancestral cyanobacterium. All extant cyanobacteria belong to a single clade within the diversity of prokaryotes. There are various types of cyanobacteria that are different in morphology and photosynthetic or physiological properties (for reviews, see Rippka *et al.*, 1979; Fay and van Baalen, 1987; Bryant, 1994). Morphological traits include cellular organization such as unicellular, filamentous, or multicellular, as well as capability of differentiation of akinetes, hormogonia and heterocysts. Photosynthetic characteristics include major antenna pigments such as chlorophyll *b*, phycocyanin and phycoerythrin. Physiological properties include utilization of various

different nitrogen compounds (ammonium, nitrate, nitrite, urea and dinitrogen), as well as cell motility and phototaxis. Various lines of evidence suggest that the chlorophyll *b*-containing species in the genera *Prochloron*, *Prochlorococcus*, and *Prochlorothrix* are not monophyletic and belong to various different sub-lineages of cyanobacteria (Honda *et al.*, 1999). The presence of phycocyanins and phycoerythrins is not related to phylogeny either. Some species are known to possess both chlorophyll *b* and phycobiliproteins (Hess *et al.*, 1996). Apparently, none of the morphological and physiological traits listed above can be used as phylogenetic markers, except that the formation of heterocysts is limited to a small clade including *Anabaena/Nostoc*. Phylogenetic analysis with the 16S rRNA sequences suggests that the plastids are in a sister group to the cyanobacteria (Hess *et al.*, 2001; Ting *et al.*, 2002), although the main focus of these analyses was on the phylogenetic status of marine *Prochlorococcus* strains. Our phylogenetic analysis using conserved protein sequences in 21 chloroplasts and 8 cyanobacteria indicates that the situation is more complicated as imagined before (Fig. 2). The cyanobcteria are clearly separated into two major clades, one including various classical laboratory strains and another including more recently identified marine isolates. The plastid sequences are also split into two major clades, one representing chlorophyte lineage and another representing rhodophyte-chromophyte lineage plus glaucophytes. The two lineages of cyanobacteria and the two lineages of plastids seem converge at a point near the origins of both clades. This suggests that the progenitor of plastids was a common ancestor of the eight extant cyanobacteria analyzed here, but the date of branching was not long before the diversification of extant cyanobacteria. Similar results were also obtained by an entirely different approach, namely, whole genome comparison by homolog group clustering (Sato, 2002). The data of a presumptive very primitive cyanobacterium *Gloeobacter vialaceus*, which is currently being sequenced and annotated in Kazusa DNA Research Institute, will shed light on this issue. There are a number of papers

describing the single origin of plastid genomes (Nelissen *et al.*, 1995) and even single origin of photosynthetic eukaryotes (Moreira *et al.*, 2000; Baldauf *et al.*, 2000; van de Peer *et al.*, 2003). We must evaluate these results on the monophyly of all plastids with caution, because several recent papers cast doubt about the simplictic view of the monophyly of all plastids (Vogl *et al.*, 2003; Stiller *et al.*, 2003). In our analysis using 14 non-photosynthetic genes that are common in 21 plastids, 8 cyanobacteria and 4 bacteria (Fig. 2 B) suggests that the bacterial branch is located within the diversification of rhodophyte plastids. Essentially similar branching of bacteria between cyanobacteria and plastids was found in the 16S rRNA tree (Hess *et al.*, 2001). The green and red lineages diverge from the cyanobacteria in a cluster analysis (De Las Rivas *et al.*, 2002). The phylogenetic trees compatible with the monophyly of all plastids have been obtained by the mixed use of photosynthetic and house-keeping genes (Fig. 2A; Martin *et al.*, 2002), but the photosynthesis genes might have been subject to extensive horizontal gene transfer (Raymond *et al.*, 2002). Therefore, further study is needed to unequivocally identify the origin of cyanobacteria and plastids as well as the relationship between them.

Although the progenitor of plastid is still to be identified, the size and content of the genome of the progenitor is expected to be almost within the diversification of those of extant cyanobacteria. In other words, the progenitor is expected to have had the genomic system, which is similar to the one that the present-day cyanobacteria have in common. This is the basis for the evolution of genomic machinery.

B. Loss of prokaryotic DNA-binding proteins

As described in III B 1, the presence of prokaryotic DNA-binding proteins are limited to algae of the rhodophyte-chromophyte lineage, and the number of such proteins is also quite

limited (Table 1). Essentially the same is true for prokaryotic transcription factors and response regulators with DNA-binding activity. NtcA, RbcR, and OmpR homologs are only found encoded by the plastid genome of algae of the rhodophyte lineage, and no genes that encode such homologs have been detected in the nuclear genomes of plants and algae (Table 1). In a previous article (Sato, 2001), one of the authors pointed out that an explanation for this scarcity of prokaryotic DNA-binding proteins and transcription factors in plastids is that "the plastids within the cytoplasm of eukaryotic host cell do not suffer from extensively changing environmental stresses that are to be overcome by free-living prokaryotic cells by sophisticated network of transcriptional regulation." This statement is still valid in a general sense. The large reduction in the genome size and genome content with virtually no transcription factors (Moran, 2002) was first reported in a human parasite, *Mycoplasma genitalium* (Fraser *et al.*, 1995), and subsequently in various other parasitic pathogens, such as *Chlamydia, Chlamydophila, Rickettsia* and spirochetes. Two strains of *Buchnera*, intracellular symbionts present in the bacteriocytes of aphids, also provide a different example of genomic reduction in symbiotic prokaryotes (Shigenobu *et al.*, 2000; van Hom *et al.*, 2002), which began symbiosis about 200 Myr ago. In these parasites or symbionts, essential genes such as those encoding DNA and RNA polymerases, components of translation machinery, and basic metabolic enzymes are retained, with concomitant loss of nearly all of the regulatory proteins. The situation must have been similar in the progenitor of plastids. About 600 kbp seems the lower limit of the genome size of a bacterium, which is recognizable as an independent organism. *Mycoplasma* and *Buchnera* have such minimal sizes of genomes. Even though these orgasisms retain their own genomic machinery, such as DNA polymerase III, DnaA protein, RNA polymerase, sigma factors, and HU protein, many important enzymes in the biosynthesis of metabolites are lacking. These metabolites are provided by the host cell, and therefore, the symbiosis is essential in the maintenance of

parasite life. This can be characterized as ‘metabolic symbiosis’. Further reduction of the genome to about 150 – 190 kbp, which is the size of the plastid genomes, is critical to the transition from symbiosis or parasitism to endosymbiosis. In other words, this necessitates the loss of independence of the endosymbiont: beyond the metatolic symbiosis, essential genomic machinery is dependent on and provided by the host cell.

The most important event in the loss of independence is the switching to a new replication system. The loss of DnaA protein and DNA polymerase III, which are both essential in bacterial replication, is common in all algal and plant plastids. In this regard, the organelles and parasitic bacteria are fundamentally different. The DNA polymerase, which are found encoded by the nuclear genome of various plants and algae, and which are similar to bacterial DNA polymerase I, seems a plausible candidate for the organellar replication enzyme as discussed above. In this context, it is clear that the mitochondrial DNA polymerase had already existed when plastid emerged. This mitochondrial polymerase could be recruited as a plastid polymerase (Fig. 7). The history of mitochondrial DNA polymerases is also complicated, because, as described above, the mitochondria of animals and fungi contain gamma-type polymerase, another distant member of the Class A family. The initial mitochondria must have had DNA polymerase III as do alpha-proteobacteria, but this must have been replaced by a host-encoded DNA polymerase. Whether this was a gamma-type polymerase or a PolI-type polymerase is still unknown. Nor is known the origin of gamma-type polymerase and the organellar PolI-type polymerases. In one scenario, the initial mitochondriate eukaryote possessed a PolI-type polymerase, which might be acquired from the complex bacterial fusions giving rise to the progenitor of eukaryotes (Gupta, 1999). This polymerase is replaced by the gamma-type polymerase in the ancestor of animals and fungi, but is retained in the plant-algal lineage. Then, cyanobacterial endosymbiosis engendered plastids, and the mitochondrial replication system is duplicated to provide a polymerase

adapted for plastid replication. It is reasonable that replacement to gamma-type became difficult after such gene duplication. Another scenario is that the first mitochondrial replicase was the gamma-type enzyme, but replaced by the PolI-type enzyme in the ancestor of photosynthetic eukaryotes. In this case, the origin of the PolI-type enzyme might be horizontal gene transfer or another hypothetical endosymbiosis which no longer left a trace of its invasion.

The changes in transcription system are more clearly traced from the existing evidence (Fig. 7). The prokaryotic RNA polymerase (core enzyme) of the cyanobacterial origin (PEP) remains to be encoded in the plastid genome until now in most of plants and algae. A clear evidence for the cyanobacterial origin of the plastid RNA polymerase is that both have split beta' subunit (namely, beta' and beta'') in contrast to all other bacteria. Sequence-based phylogenetic analysis also supports the cyanobacterial origin of the plastid RNA polymerases. The sigma subunit is encoded in the nuclear genome in all plants and algae analyzed, and this is taken as evidence for the nuclear control of chloroplast transcription. Other prokaryotic proteins that are retained in the plastid genome are ribosomal RNA, transfer RNA, ribosomal proteins, and initiation factor, as well as subunits of some enzymes involved in the metabolism in the chloroplasts such as AccD.

C. Gain of eukaryotic DNA-binding proteins and regulatory proteins

The plastids of land plants, however, have acquired a number of transcriptional regulators as well as phage-type RNA polymerase from the nucleus of eukaryotic host. Some of these are bifunctional proteins such as sulfite reductase and CND41. An estimation based on comparison of *A. thaliana*, *Synechocystis* and yeast genomes suggested that 1,500 out of 2,300 putative plastid proteins encoded in the cell nucleus originated from ancestral

eukaryotic host (Abdallah *et al.*, 2000), which is consistent with our current hypothesis. This situation is reasonable because the cells of land plants are subject to differentiation of cells and tissues, and new regulatory proteins are needed to assure coordinated expression of the plastome during the differentiation of each plant cell. The details will be explained in the following section.

D. Discontinuous evolution of plastid genomic machinery

A current version of the hypothesis of discontinuous evolution of plastid genomic machinery is shown in Fig. 7. In this figure, the evolutionary events are depicted from the bottom to the top. The initial endosymbiosis event is shown at the bottom. Here, the nature of DNA polymerase of the mitochondria in this ancestral eukaryote remains to be identified (see above). The initial endosymbiosis might be either a single event (Moreira *et al.*, 2000; Baldauf *et al.*, 2000), or parallel (multiple) events (Palmer, 2003), or maybe repeated several times. The repeated endosymbiosis has not been tested rigourously, but a secondary loss of plastid was postulated in tryponosomes (Hannaert *et al.*, 2003). Recently, Nozaki *et al.* (2003) extended the previous analysis by Baldauf *et al.* (2000) using the combined four protein sequences by adding the data of a rhodophyte *C. merolae* and showed that all green plants, rhodophytes and heterokonts as well as apicomplexa (Alveolata) and kinetoplastida (Discicristata) are monophyletic. The results suggest that the non-photosynthetic organisms such as trypanosomes might once have had plastids but lost them later. As an extension of this argument, even the organisms that have currently plastids might have lost plastids once and then acquired other ones, as in the case of secondary symbiosis. Therefore a detailed analysis will be necessary to discriminate a simple primary endosymbiosis from a repeated endosymbiosis. As this does not affect the essential point of our hypothesis, we do not discuss

this problem any further here.

One of the most important among the initial events during the establishment of the endosymbiosis was the loss of replication system from the endosymbiont. The cyanobacterial DNA polymerases as well as related proteins such as DnaA protein did not remain in any plastid-bearing organisms, and therefore, must have lost and replaced by the system that was provided by the host cell at the earliest moment of the endosymbiosis. The most probable scenario is the duplication of the nuclear-coded mitochondrial replication system. However, such radical change did not occur in the transcription and translation systems. No essential difference in the translation system is found between cyanobacteria and plastids (Sugiura *et al.*, 1998), though a number of ribosomal protein subunits of plastids are now encoded in the nuclear genome.

The transcription system can be defined as intermediate between the two. The RNA polymerase of the cyanobacterial origin remained until the present-day land plants and algae, but the sigma factors that function with the core enzyme are encoded in the nuclear genome in all the algae and plants analyzed to date. Most if not all of the transcription factors and structural proteins of the cyanobacterial nucleoid were lost during the establishment of endosymbiosis. We define this as 'Phase 1' of discontinuous evolution of plastid genomic machinery (Fig. 7). In this phase, some DNA-binding proteins remain in the algae of the rhodophyte lineage including the chromophytes resulting from the secondary endosymbiosis. In contrast, virtually no prokaryotic factors remained in the green lineage including green algae and land plants. At the same time, RNA processing systems were introduced into the plastids from the host cell. This must have been necessary to compensate for the loss of all or nearly all of the prokaryotic transcription regulation system. The RNA processing system is well studied in a green alga *Chlamydomonas*, but we still do not know exactly if this system is operative in all green algae or in all algae including red algae. Further functional genomic

studies on the *Cyanidioschyzon* plastid gene expression will shed light on this issue.

The second phase of the evolution of plastid genomic machinery consists of massive import of protein factors from the host cell. Among these proteins, sulfite reductase is noteworthy. As described in III-B-2, this enzyme is essential in the assimilation of sulfate through sulfite and sulfide in both cyanobacteria and plastids. However, sulfite reductase acquired the ability of binding to DNA in lower land plants at least in the mosses. The binding of sulfite reductase severely represses the transcription activity of the plastid nucleoids (Fig. 4). This general repression of the activity of plastid nucleoid is related to the ability of plants to develop non-photosynthetic plastids. The green algal plastids are always green and photosynthetically active except for mutants such as the *y-1* mutant of *Chlamydomonas*, whereas the lower land plants develop totally non-photosynthetic cells such as spores and rhizoid. Epidermal cells remain photosynthetic up to pteridophytes. In addition, vascular tissues are no longer photosynthetic in vascular plants. We suspect that the DNA-binding of sulfite reductase is involved in such total repression of plastid genome in non-photosynthetic plastids especially those in dormant cells such as spores and seeds.

During the evolution of flowering plants, especially angiosperms (Phase 3), many additional transcription factors were introduced from the host cell. The DNA-binding proteins, such as PEND, CND41, PTF1, and PD3 proteins, appeared in angiosperms, and are supposed to be the products of novel creation of proteins, since the origins of these proteins are hard to speculate. In addition, the mitochondrial RNA polymerase (RPOT) is duplicated and targeted to the chloroplast during the evolution of angiosperms. This is the origin of the NEP, the nuclear-encoded RNA polymerase of the chloroplast. The duplication of RPOT seemed to occurr many times during the evolution of plants, but the first successful invasion into the chloroplast was achieved in angiosperms. Then, another gene duplication occurred in dicots, and a special type of RPOT, so-called ‘dually targeting’ enzyme was produced. This

enzyme is potentially targeted to the chloroplast when translation begins from the first AUG codon, which is not the case *in vivo* in all cell types tested. A plausible explanation for this strange protein might be that this protein is in the process of evolution to produce active chloroplast enzyme.

V. Conclusion and prospects

The structure and function of plastid nucleoids are rather diverse depending on organisms and developmental stages. This was not evident ten years ago when people believed in the essential similarity of plastids and cyanobacteria. The hypothesis on the discontinuous evolution of plastid genomic machinery has gained additional support after its initial proposal (Sato, 2001). Especially, the emerging genomic data on the red alga *Cyanidioschyzon merolae* is giving invaluable information on a lineage other then the green lineage. The advance in further genome sequencing as well as comparative proteomics of nucleoids will shed more light on the evolution of plastid genomic machinery or nucleoids.

ACKNOWLEDGMENTS

The authors thank Dr. Masashi Tazawa for his initial advice in writing this article. They are also grateful to all the present and past members of the laboratory and many collaborators, especially Drs. Tsuneyoshi Kuroiwa and Toshiharu Hase. Many useful data were provided from the Cyanidioschyzon Genome Project. The work in the authors's laboratory was supported in part by Grants-in-Aid for Scientific Research (Nos. 10170202, 11151202, and 11440232) from the Japanese Society for the Promotion of Science and from the Ministry of

Education, Science and Culture, Japan.

REFERENCES

Abdallah, F., Salamini, F., and Leister, D. (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends Plant Sci.* **5**, 141-142.

Aketagawa, J., and Tamura, G. (1980) Ferredoxin-sulfite reductase drom spinach. *Agric. Biol. Chem.* **44**, 2371-2378.

Allison, L., A. (2000) The role of sigma factors in plastid transcription. *Biochimie* 82, 537-548.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815

Azam, T. A., and Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.* **274**, 33105-33113.

Azam, T. A., Hiraga, S., and Ishihama, A. (2000) Two types of localization of the DNA-binding proteins within the Escherichia coli nucleoid. *Genes to Cells* **5**, 613, 626.

Baba, K., Nakano, T., Yamagishi, K. and Yoshida, S. (2001) Involvement of a nuclear-encoded basic helix-loop-helix protein in transcription of the light-responsive promoter of *psbD*. *Plant Physiol*. **125**, 595-603.

Baena-González, E., Baginsky, S., Mulo, P., Summer, H., Aro, E.-M., and Link, G. (2001) Chloroplast transcription at different light intensities. Glutathione-mediated phosphorylation of the major RNA polymerase involved in redox-regulated organellar gene expression. Plant Physiol. 127, 1044-1052.

Baginsky, S., Tiller, K., Pfannschmidt, T., and Link, G. (1999) PTK, the chloroplast RNA polymerase-associated protein kinase from mustard (*Sinapis alba*), mediates redox control of plastid in vitro transcription. *Plant Mol. Biol.* **39**, 1013-1023.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I., and Doolittle, W. F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972-977.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marchall, M., and Sonnhammer, E. L. L. (2002) The Pfam protein families database. *Nucleic Acids Res.* **30**, 276-280.

Beardslee, T. A., Roy-Chowdhury, S., Jaiswal, P., Buhot, L., Lerbs-Mache, S, Stern, D. B., and Allison, L. A. (2002) A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts. *Plant J.* **31**, 199-209.

Benkovic, S. J., Valentine, A. M., and Salinas, F. (2001) Replisome-mediated DNA replication. *Annu. Rev. Biochem.* **70**, 181-208.

Bligny, M., Courtois, F., Thaminy, S., Chang, C. C., Lagrange, T., Baruah-Wolff, J., Stern, D., and Lerbs-Mache, S. (2000) Regulation of plastid rDNA transcription by interaction of CDF2 with two different RNA polymerase. *EMBO J.* **19**, 1851-1860.

Bonnefoy, E., and Rouvière-Yaniv, J. (1991) HU and IHF, two homologous histone-like proteins of Escherichia coli, form different protein-DNA complexes with short DNA fragments. *EMBO J.* **10**, 687-696.

Briat, J.-F., Letoffe, S., Mache, R., and Rouviere-Yaniv, J. (1984) Similarity between the bacterial histone-like protein HU and a protein from spinach chloroplasts. *FEBS Letters* **172**, 75-79

Bryant, D. A. (1994) "The Molecular Biology of Cyanobacteria." Kluwer Academic Publishers, Dordrecht.

Bujalowski, W. (2003) Expanding the physiological role of the hexameric DnaB helicase. *Trends Biochem. Sci.* **28**, 116-118.

Cannon, G. C., Ward, L. N., Case, C. I., and Heinhorst, S. (1999) The 68 kDa DNA compacting nucleoid protein from soybean chloroplasts inhibits DNA synthesis *in*

vitro. *Plant Mol Biol.* **39**, 835-845.

Cermakian, N., Ikeda, M., Cedergren, R., and Gray, M. W. (1996) Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Res*. **24**, 648-654.

Chiu, W.-I., Niwa, Y., Zeng, W., Hirose, T., Kobayashi, H., and Sheen, J. (1995) Engineered GFP as vital reporter in plants. *Current Biology* **6**, 325-330.

Christiansen, H., Hansen, A. C., Vijn, I., Pallisgaard, N., Larsen, K., Yang, W.-C., Bisseling, T., Marcker, K. A., and Jensen, E. Ø. (1996) A novel type of DNA-binding protein interacts with a conserved sequence in an early nodulin *ENOD12* promoter. *Plant Mol. Biol.* **32**, 809-821.

Coleman, A. W. (1979) Use of the fluorochrome 4',6-diamidino-2-phenylindole in genetic and developmental studies of chloroplast DNA. *J. Cell Biol.* **82**, 299-305.

Crevel, G., Laine, B., Sautiere, P., and Galleron, C. (1989) Isolation and characterization of DNA-binding proteins from the cyanobacterium *Synechococcus* sp. PCC 7002 (*Agmenellum quadruplicatum*) and from spinach chloroplasts. *Biochem. Biophys. Acta*., 1007, 36-43

Dame, R. T., and Goosen, N. (2002) HU: promoting or counteracting DNA compaction? *FEBS Letters* **529**, 151-156.

De Las Rivas, J., Lozano, J. J., and Ortiz, A. R. (2002) Comparative analysis of chloroplast genomes: Functional annotation, genome-based phylogeny, and deduced evolutionary patterns. *Genome Res*. **12**, 567-583.

Deng, X.-W., Wing, R. A., and Gruissem, W. (1989) The chloroplast genome exists in multimeric forms. *Proc. Natl. Acad. Sci. U. S. A*. **86**, 4156-4160.

Douglas, S. E., Murphy, C. A., Spencer, D. F., and Gray, M. W. (1991) Cryptomonad algae are evolutionary chimeras of two phylogenetically distinct unicellular eukaryotes.

Nature **350**, 148-151.

Drager, R. G., Higgs, D., Kindle, K. L., and Stern, D. B. (1999) 5' to 3' exoribonucleolytic activity is a normal component of chloroplast mRNA decay pathways. *Plant J.* **19**, 521-531.

Drapier, D., Girard-Bascou, J., Stern, D. B., and Wollman, F.-A. (2002) A dominant nuclear mutation in Chlamydomonas identifies a factor controlling chloroplast mRNA stability by acting on the coding region of the atpA transcript. *Plant J.* **31**, 687-697.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol*. **300**, 1005-1016.

Erzberger, J. P., Pirruccello, M. M., and Berger, J. M. (2002) The structure of bacterial DnaA: Implications for general mechanisms underlying DNA replication initiation. *EMBO J.* **21**, 4763-4773.

Fay, P., and van Baalen, C. (1987) "The Cyanobacteria." Elsevier, Amsterdam.

Ferro, M., Salvi, D., Rivière-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., Garin, G., Joyard, J., and Rolland, N. (2002) Integral membrane proteins of the chloroplast envelope: Identification and subcellular localization of new transporters. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11487-11492.

Fliée, J., Forterre, P., Sen-Lin, T., and Laurent, J. (2002) Evolution of DNA polymerase families: Evidence for multiple gene exchange between cellular and viral proteins. *J. Mol. Evol.* 54, 763-773.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu,

P.-C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison III, C. A. , and Venter, J. C. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397-403.

Fujiwara, M., Nagashima, A., Kanamaru, K., Tanaka, K., and Takahashi, H. (2000) Three new nuclear genes, *sigD*, *sigE*, *sigF*, encoding putative plastid RNA polymerase sigma factors in *Arabidopsis thaliana*. *FEBS Lett*. **481**, 47-52.

Gaikwad, A., Hop, D. V., and Mukherjee, S. K. (2002) A 70-kDa chloroplast DNA polymerase from pea (*Pisum sativum*) that shows high processivity and displays moderate fidelity. *Mol. Genet. Genomics* 267, 45-56.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Paln, A., Nelson, K. E., Bowman, S., Paulsen I. T., James, K., Elsen, J. A., Rutherford, K., *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum. Nature* **419**, 498-511.

Goldschmidt-Clermont, M. (1998) Coordination of nuclear and chloroplast gene expression in plant cells. *Int. Rev. Cytol.* **177**, 115-180.

Grams, J., Morris, J. C., Drew, M. E., Wang, Z., Englund, P. T., and Hajduk, S. L. (2002) A Trypanpsome mitochondrial RNA polymerase is required for transcription and replication. *J. Biol. Chem*. **277**, 16952-16959.

Grasser, K. D., Ritt, C., Krieg, M., Fernández, S., Alonso, J. C., and Grimm, R. (1997) The recombinant product of the *Cryptomonas* Φ plastid gene *hlpA* is an architectural HU-like protein that promotes the assembly of complex nucleoprotein structures. *Eur. J. Biochem.* **249**, 70-76.

Gray, M. W. (1992) The endosymbiont hypothesis revisited. *Int. Rev. Cytol*. **141**, 233-357.

Gupta, R. S., Mukhtar, T., and Singh, B. (1999) Evolutionary relationships among photosynthetic prokaryotes (*Heliobacterium chlorum, Chloroflexus aurantiacus*,

cyanobacteria, *Chlorobium tepidum* and proteobacteria): implications regarding the origin of photosynthesis. *Mol. Microbiol.* **32**, 893-906.

Hajdukiewicz, P. T. J., Allison, L. A., and Maliga, P. (1997) The two RNA polymerases encoded by the nuclear and plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* **16**, 4041-4048.

Hansmann, P., Falk, H., Ronai, K., and Sitte, P. (1985) Structure, composition, and distribution of plastid nucleoids in *Narcissus pseudonarcissus*. *Planta* **164**, 459-472.

Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J.-P., Rigden, D. J., Michels, P. A. M., and Opperdoes, F. R. (2003) Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1067-1071.

Hara, K., Sugita, M., and Aoki, S. (2001a) Cloning and characterization of the cDNA for a plastid σ factor from the moss *Physcomitrella patens*. *Biochem. Biophys. Acta*. **1517**, 302-306.

Hara, K., Morita, M., Takahashi, R., Sugita, M., Kato, S., and Aoki, S. (2001b) Characterization of two genes, *Sig1* and *Sig2*, encoding distinct plastid σ factors in the moss *Physcomitrella patens*: phylogenetic relationships to plastid σ factors in higher plants. *FEBS Lett*. **499**, 87-91.

Hedtke, B., Börner, T., and Weihe, A. (1997) Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis. Science* **277**, 809-811.

Hedtke, B., Meixner, M., Gillandt, S., Richter, E., Börner, T., and Weihe, A. (1999) Green fluorescent protein as a marker to investigate targeting of organellae RNA polymerases of higher plants in vivo. *Plant J.* **17**, 557-561.

Hedtke, B., Börner, T., and Weihe, A. (2000) One RNA polymerase serving two genomes. *EMBO Rep.* **1**, 435-440.

Hedtke, B., Legen, J., Weihe, A., Herrman, R. G., and Börner, T. (2002) Six active

phage-type RNA polymerase genes in *Nicotiana tabacum*. *Plant J*. **30**, 625-637.

Heinhorst, S., and Cannon, G. C. (1993) DNA replication in chloroplasts. *J. Cell Sci.* **104**, 1-9.

Hess, W. R., and Börner, T. (1999) Organellar RNA polymerases of higher plants. *Int. Rev. Cytol*. **190**, 1-59.

Hess, W. R., Partensky, F., Staay, G. W. M. V. D., Garcia-Fernandez, J. M., Borner, T., and Vaulot, D. (1996) Coexistence of phycoerythrin and a chlorophyll a/b antenna in a marine prokaryote. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11126-11130.

Hess, W., R., Rocap, G., Ting, C. S., Larimer, F., Stilwagen, S., Lamerdin, J., and Chisholm, S. W. (2001) The photosynthetic apparatus of Prochlorococcus: Insights through comparative genomics. *Photosynth. Res.* **70**, 53-71.

Hirama, T., Sekine, K., and Sato, N. (2003) Analysis of sulfite reductase in the plastid nucleoid during greening. *Plant Cell Physiol.* **44**, s58.

Hirose, T., and Sugiura, M. (2001) Involvement of a site-specific *trans*-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast *in vitro* RNA editing system. *EMBO J.* **20**, 1144-1152.

Hirose, T., Sugita, M., and Sugiura, M. (1994) Characterization of a cDNA encoding a novel type of RNA-binding protein in tobacco: its expression and nucleic acid-binding properties. *Mol. Gen. Genet*. **244**, 360-366.

Honda, D., Yokota, A., and Sugiyama, J. (1999) Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. *J. Mol. Evol.* **48**, 723-739.

Hu, J., and Bogorad, L. (1990) Maize chloroplast RNA polymerase: The 180-, 120-, and 38-kilodalton polypeptides are encoded in chloroplast genes. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 1531-1535.

Hu, J., Troxler, R. F., and Bogorad, L. (1991) Maize chloroplast RNA polymerase: The 78-kilodalton polypeptide is encoded by the plastid *rpoC1* gene. *Nucl. Acids Res.* **19**, 3431-3434.

Igarashi, K., and Kashiwagi, K. (1999) Polyamine transport in bacteria and yeast. *Biochem. J.* **344**, 633-642.

Iratni, R., Diederich, L., Harrak, H., Bligny, M., and Lerbs-Mache, S. (1997) Organ-specific transcription of the rrn operon in spinach plastids. *J. Biol. Chem.* **272**, 13676-13682.

Isono, K., Shimizu, M., Yoshimoto, K., Niwa, Y., Satoh, K., Yokota, A., and Kobayashi, H. (1997) Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains for σ^{70} factors of bacterial RNA polymerases in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A*. **94**, 14948-14953.

Kabeya, Y., Hashimoto, K., and Sato, N. (2002) Identification and characterization of two phage-type RNA polymerase cDNAs in the moss *physcomitrella patens*: Implication of recent evolution of nuclear-encoded RNA polymerase of plastids in plants. *Plant Cell Physiol*. **43**, 245-255.

Kabeya, Y., Sekine, K., and Sato, N. (2003) Evolution of organellar transcription machinery in bryophytes and vascular plants. *In* "New Frontiers in Bryology" (Wood, A. J., Ed.). Kluwer Academic Publishers, Dordrecht (in press).

Kanamaru, K., Fujiwara, M., Seki, M., Katagiri, T., Nakamura, M., Mochizuki, N., Shinozaki K., Tanaka, K., and Takahashi, H. (1999) Plastidic RNA polymerase σ factors in *Arabidopsis*. *Plant Cell Physiol*. **41**, 832-842

Kanamaru, K., Nagashima, A., Fujiwara, M., Shimada, H., Shirano, Y., Nakabayashi, K., Shibata, D., Tanaka, K., and Takahashi, H. (2001) An Arabidopsis sigma factor (SIG2)-dependent expression of plastid-encoded tRNAs in chloroplasts. *Plant Cell Physiol*. **42**, 1034-1043

Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T., et al. 2001. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* **8**, 205-13.

Kapoor, S., Suzuki, J. Y., and Sugiura, M. (1997) Identification and functional significance of a new class of non-consensus-type plastid promoters. *Plant J.* **11**, 327-337.

Karlok, M. A., Jang, S.-H., and Jaehning, J. A. (2002) Mutation in the yeast mitochondrial RNA polymerase specificity factor, Mtf1, verify an essential role in promoter utilization. *J. Biol. Chem.* **277**, 28143-28149.

Kaufman, B. A., Newman, S. M., Hallberg, R. L., Slaughter, C. A., Perlman, P. S., and Butow, R. A. (2000) *In organello* formaldehyde crosslinking of proteins to mtDNA: Identification of bifunctional proteins. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7772-7777.

Kestermann, M., Neukirchen, S., Kloppstech, K., and Link, G. (1998) Sequence and expression characteristics of a nuclear-encoded chloroplast sigma factor from mustard (*Sinapis alba*). *Nucleic Acids Res.*, **26**, 2747-2753

Kim, M., and Mullet, J. E. (1995) Identification of a sequence-specific DNA binding factor required for transcription of the barley chloroplast blue light-responsive *psbD-psbC* promoter. Plant Cell 7, 1445-1457.

Kimura, S., Uchiyama, Y., Kasai, N., Namekawa, S., Saotome, A., Ueda, T., Ando, T., Ishibashi, T., Oshige, M., Furukawa, T., Yamamoto, T., Hashimoto, J., and Sakaguchi, K. (2002) A novel DNA polymerase homologous to *Escherichia coli* DNA polymerase I from a higher plant, rice (*Oryza sativa*). Nucleic Acids Res. 30, 1585-1592.

Kobayashi, T., Takahara, M., Miyagishima, S., Kuroiwa, H., Sasaki, N., Ohta, N., Matsuzaki, M. and Kuroiwa, T. (2002) Detection and localization of a chloroplast-encoded HU-like protein that organizes chloroplast nucleoids. *Plant Cell* **14**, 1579-1589.

Kobayashi, Y., Dokiya, Y., Sugiura, M., Niwa, Y., and Sugita, M. (2001a) Genomic organization and organ-specific expression of a nuclear gene encoding phage-type RNA polymerase in *Nicotiana sylvestris*. *Gene* **279**, 33-40.

Kobayashi, Y., Dokiya, Y., and Sugita, M. (2001b) Dual targeting phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts. *Biochem. Biophys. Res. Commun*. **289**, 1106-1113.

Kobayashi, Y., Sugiura, C., and Sugita, M. (2002) The moss *Physcomitrella patens* chloroplast *rpoA* gene is present in the nuclear genome. MOSS2002 Abstract Book, p57.

Kolodner, R.D., and Tewari, K.K. (1975). Chloroplast DNA from higher plants replicates by both the Cairns and the rolling circle mechanism. *Nature* **256**, 708-711.

Komine, Y., Kikis, E., Schuster, G., and Stern, D. (2002) Evidence for *in vivo* modulation of chloroplast RNA stability by 3'-UTR homopolymeric tails in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U. S. A*. **99**, 4085-4090.

Koo, A. J. K., and Ohlrogge, J. B. (2002) The predicted candidates of Arabidopsis plastid inner envelope membrane proteins and their expression profiles. *Plant Physiol.* **130**, 823-836.

Krause, K., and Krupinska, K. (2000) Molecular and functional properties of a highly purified transcriptionally active chromosomes from spinach chloroplasts. *Physiol. Plant.* **109**, 186-195.

Krause, M., Maier, R. M., Kofer, W., Krupinska, K., and Herrmann, R. G. (2000) Disruption of plastid-encoded RNA polymerase genes in tobacco: expression of only a distinct set of genes is not based on selective transcription of the plastid chromosome. *Mol. Gen. Genet*., 263, 1022-1030

Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001) MEGA2: molecular evolutionary

genetics analysis software. *Bioinformatics* **17**, 1244-1245.

Kunnimalaiyaan, M., and Nielsen, B. L. (1997) Fine mapping of replication origins (*ori*A and *ori*B) in *Nicotiana tabacum* chloroplast DNA. *Nucl. Acids Res.* **25**, 3681-3686.

Kuroiwa, T. (1991) The replication, differentiation, and inheritance of plastids with emphasis on the concept of organelle nuclei. *Int. Rev. Cytol.* **128**, 1-62.

Kuroiwa, T. (1998) The primitive red algae *Cyanidium caldarium* and *Cyanidioschyzon merolae* as model system for investigating the dividing apparatus of mitochondria and plastids. *BioEssays* **20**, 344-354.

Kuroiwa, T., Suzuki, T., Ogawa, K., and Kawano, S. (1981) The chloroplast nucleus: distribution, number, size, and shape, and a model for the multiplication of the chloroplast genome during chloroplast development. *Plant Cell Physiol.* **22**, 381-396.

Lahiri, S. D., and Allison, L. A. (2000) Complementary expression of two plastid-localized sigma-like factors in maize. *Plant Physiol*., 123, 883-894.

Lemieux, C., Otis, C., and Turmel, M. (2000) Ancestral chloroplast genome in *Mesostigma viride* reveals and early branch of green plant evolution. *Nature* **403**, 649-652.

Leon, P., Aroyo, A., and Mackenzie, S. (1998) Nuclear control of plastid and mitochondrial development in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 453-480.

Liere, K., and Maliga, P. (1999) In vitro characterization of the tobacco *rpoB* promoter reveals a core sequence motif conserved between phage-type plastid and plant mitochondrial promoters. *EMBO J*., **18**, 249-257.

Lilly, J. W., Havey, M. J., Jackson, S. A., Jiang, J. (2002) Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. *Plant Cell* **13**, 245-254.

Lisitsky, I., Klaff, P., and Schuster, G. (1996) Addition of destabilizing poly(A)-rich sequences to endonuclease cleavage sites during the degradation of chloroplast mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13398-13403.

Lorkovic, Z. J., and Barta, A. (2002) Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant Arabidopsis thaliana. *Nucleic Acids Res.* **30**, 623-635.

Marrison, J. L., and Leech, R. M. (1992) Co-immunolocatization of topoisomerase II and chloroplast DNA in developing, dividing and mature wheat chloroplasts. *Plant J.* **2**, 783-790.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M., and Kowallik, K. V. (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162-165.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and Penny, D. (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12246-12251.

Maruyama, K., Sato, N. and Ohta, N. (1999) Conservation of structure and cold-regulation of RNA-binding proteins in cyanobacteria: probable convergent evolution with eukaryotic glycine-rich RNA-binding proteins. *Nucleic Acids Res.* **27**, 2029-2036.

Matsuzaki, M., *et al.* (2003) Genome sequence of the primitive photosynthetic eukaryote *Cyanidioschyzon merolae*. Submitted.

Mathews, D. E., and Durbin, R. D. (1990) Tagetitoxin inhibits RNA synthesis directed by RNA polymerases from chloroplasts and *Escherichia coli*. *J. Biol. Chem*., 265, 493-498.

Matthews, H. R. (1993) Polyamines, chromatin structure and transcription. *Bioessays* **15**, 561-566.

Mayfield, S. P., Yohn, C. B., Cohen, A. and Danon, A. (1995) Regulation of chloroplast gene expression. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**, 147-166.

McLeod, S. M., and Johnson, R. C. (2001) Control of transcription by nucleoid proteins. *Curr. Opinion Microbiol.* **4**, 152-159.

Meeker, R., Nielsen, B., and Tewari, K.K. (1988). Localization of replication origins in pea chloroplast DNA. *Mol. Cell. Biol.* **8**, 1216-1223.Nielsen, B. L., and Tewari, K. K. (1988) Pea chloroplast topoisomerase I: purification, characterization, and role in replication. *Plant Mol Biol.* **11**, 3-14.

Messer, W., Blaesing, F., Jakimowicz, D., Krause, M., Majka, J., Nardmann, J., Schaper, S., Seitz, H., Speck, C., Weigel, C., Wegrzyn, G., Welzeck, M., Zakrzewska-Czerwinska, J. (2001) Bacterial replication initiator DnaA. Rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading. *Biochimie* **83**, 5-12.

Misumi, O., Suzuki, L., Nishimura, Y., Sakai, A., Kawano, S., Kuroiwa, H., and Kuroiwa, T. (1999) Isolation and phenotypic characterization of Chlamydomonas reinhardtii mutants defective in chloroplast DNA segregation. *Protoplasma* **209**, 273-282.

Miyamura, S., Nagata, T., and Kuroiwa, T. (1986) Quantitative fluorescence microscopy on dynamic changes of plastid nucleoids during wheat development. *Protoplasma* **133**, 66-72.

Monde, R. A., Schuster, G., and Stern, D. B. (2000) Processing and degradation of chloroplast mRNA. *Biochimie* **82**, 573-582.

Moran, N. A. (2002) Microbial minimalism: Genome reduction in bacterial pathogens. *Cell* **108**, 583-586.

Moreira, D., Le Guyader, H., and Philippe, H. (2000) The origin of red algae and the evolution of chloroplasts. *Nature* **405**, 69-72.

Morikawa, K., Ito, S., Tsunoyama, Y., Nakahira, Y., Shirano, T., and Toyoshima, Y. (1999) Circadian-regulated expression of a nuclear-encoded plastid sigma factor gene (sigA) in wheat seedling. *FEBS Lett.*, 451, 275-278.

Mühlbauer, S. K., Lössl, A., Tzekova, L., Zou, Z., and Koop, H.-U. (2002) Functional analysis of plastid DNA replication origins in tobacco by targeted inactivation. *Plant J.* **32**, 175-184.

Mullet, J. E. (1988) Chloroplast development and gene expression. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39**, 475-502.

Murakami, S., Kondo, Y., Nakano, T., and Sato, F. (2000) Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells. *FEBS Lett.*, 468, 15-18.

Nakamura, T., Ohta, M., Sugita, M., and Sugiura, M. (2001) Chloroplast ribonucleoproteins function as a stabilizing factor of ribosome-free mRNAs in the stroma. *J. Biol. Chem.* **276**, 147-152.

Nakano, T., Sato, F., and Yamada, Y. (1993) Analysis of nucleoid-proteins in tobacco chloroplasts. *Plant Cell Physiol.* **34**, 873-880.

Nakano, T., Murakami, S., Shoji, T., Yoshida, S., Yamada, Y., and Sato, F. (1997) A novel protein with DNA binding activity from tobacco chloroplast nucleoids. *Plant Cell* **9**, 1673-1682.

Nelissen, B., Van de Peer, Y., Wilmotte, A., and De Wachter, R. (1995) An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Mol. Biol. Evol.* **12**, 1166-1173.

Nemoto, Y., Kawano, S., Nakamura, S., Mita, T., Nagata, T., and Kuroiwa, T. (1988) Studies on plastid-nuclei (nucleoids) in *Nicotiana tabacum* L. I. Isolation of proplastid-nuclei from cultured cells and identification of proplastid-nuclear proteins. *Plant Cell Physiol.* **29**, 167-177.

Nemoto, Y., Kawano, S., Kondoh, K., Nagata, T., and Kuroiwa, T. (1989) Studies on plastid-nuclei (nucleoids) in Nicotiana tabacum L. III. Isolation of chloroplast-nuclei

from mesophyll protoplasts and identification of chloroplast DNA-binding proteins. *Plant Cell Physiol.* **31**, 767-776.

Nielsen, B. L., and Tewari, K. K. (1988) Pea chloroplast topoisomerase I: purification, characterization, and role in replication. *Plant Mol Biol.* **11**, 3-14.

Nielsen, B. L., Rajasekhar, V. K., and Tewari, K. K. (1991) Pea chloroplast DNA primase: characterization and role in initiation of replication. *Plant Mol. Biol.* **16**, 1019-1034.

Nozaki, H., Matsuzaki, H., Takahara, M., Misumi, O., Kuroiwa, H., Hasegawa, M., Shin-I, T. Kohara, Y., Ogasawara, N., and Kuroiwa, T. (2003) The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. J. Mol. Evol. **56**, 485-497.

Ogihara, Y., Isono, K., Kojima, T., Endo, A., Hanaoka, M., Shiina, T., Terachi, T., Utsugi, S., Murata, M., Nori, N., Takumi, S., Ikeo, K., Gojobori, T., Murai, R., Murai, K., Matsuoka, Y., Ohnishi, Y., Tajiri, H., and Tsunewaki, K. (2000) Chinese spring wheat (*Triticum aestivum* L.) chloroplast genome: complete sequence and contig clones. *Plant Mol. Biol. Rep.* **18**, 243-253.

Ogrzewalla, K., Piotrowski, M., Reinbothe, S., and Link, G. (2002) The plastid transcription kinase from mustard (*Sinapis alba* L.). A nuclear-encoded CK2-type chloroplast enzyme with redox-sensitive function. *Eur. J. Biochem.* **269**, 3329-3337.

Ohta, M., Sugita, M., and Sugiura, M. (1995) Three types of nuclear genes encoding chloroplast RNA-binding proteins (cp29, cp31 and cp33) are present in *Arabidopsis thaliana*: presence of cp31 in chloroplasts and its homologue in nuclei/cytoplasms. *Plant Mol. Biol.* **27**, 529-539.

Ohta, N., Sato, N., Nozaki, H., and Kuroiwa, T. (1997) Analysis of the cluster of ribosomal protein genes in the plastid genome of a unicellular red alga *Cyanidioschyzon merolae*: translocation of the *str* cluster as an early event in the rhodophyte-chromophyte lineage

of plastid evolution. *J. Mol. Evol.* **45**, 688-695.

Ohta, N., Sato, N., and Kuroiwa, T. (1998) Structure and organization of the mitochondrial genome of the unicellular red alga *Cyanidioscyzon merolae* deduced from the complete nucleotide sequence. *Nucleic Acids Res.* **26**, 5190-5198.

Ohta, N., Sato, N., and Kuroiwa, T. (1999) The organellar genomes of *Cyanidioschyzon merolae*. *In* Enigmatic Microorganisms and Life in Extreme Environments. (Seckbach, J. Ed.) pp. 139-149. Kluwer Academic Publishers, Dordrecht.

Oikawa, K., Tanaka, K., and Takahashi, H. (1998) Two types of differentially photo-regulated nuclear genes that encode sigma factors for chloroplast RNA polymerase in the red alga Cyanidium caldarium strain RK-1. *Gene* **210**, 277-285.

Oikawa, K., Fujiwara, M., Nakazato, E., Tanaka, K., and Takahashi, H. (2000) Characterization of two plastid □factors, SigA and SigB, that mainly function in matured chloroplasts in *Nicotiana tabacum*. *Gene* **261**, 221-228.

Palmer, J. D. (2003) The symbiotic birth and spread of plastids: How many times and whodunit? *J. Phycol.* **39**, 4-11.

Peeters, N., and Small, I. (2001) Dual targeting to mitochondria and chloroplasts. *Biochem. Biophys. Acta*., 1541, 54-63.

Peltier, J.-B., Friso, G., Kalume, D. E., Roepstorff, P., Nilsson, F., Adamska, I., and van Wijk, K. J. (2000) Proteomics of the chloroplast: Systematic identification and targeting analysis of luminal and peripheral thylakoid proteins. *Plant Cell* **12**, 319-341.

Pfannschmidt, T., and Link, G. (1994) Separation of two classes of plastid DNA-dependent RNA-polymerases that are differentially expressed in mustard (*Sinapis alba* L.) seedlings. *Plant Mol. Biol.* 25, 69-81.

Pfannschmidt, T., and Link, G. (1997) The A and B forms of plastid DNA-dependent RNA polymerase from mustard (*Sinapis alba* L.) transcribe the same genes in a different

developmental context. *Mol. Gen. Genet.* **257**, 35-44.

Pfannschmidt, T., Ogrzewalla, K., Baginsky, S., Sickmann, A., Meyer, H. E. and Link, G. (2000) The multisubunit chloroplast RNA polymerase A from mustard (*Sinapis alba* L.). Integration of prokaryotic core into a larger complex with organelle-specific functions. *Eur. J. Biochem.* 267, 253-261.

Raymond, J., Zhaxybayeva, O., Gogarten, J. P., Gerdes, S. Y., and Blankenship, R. E. (2002) Whole-genome analysis of photosynthetic prokaryotes. *Science* **298**, 1616-1620.

Richter, U., Kiessling, J., Hedtke, B., Decker, E., Reski, R., Börner, T., and Weihe, A. (2002) Two *RpoT* genes of *Physcomitrella patens* encode phage-type RNA polymerases with dual targeting to mitochondria and plastids. *Gene*, 290, 95-105.

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, D., Sherman, B. K., and Yu, G.-L. (2000) *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. Science 290, 2105-2109.

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., and Stanier, R. Y. (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* **111**, 1-61.

Rivier, C., Goldschmidt-Clermont, M., and Rochaix, J. D. (2001) Identification of an RNA-protein complex involved in chloroplast group II intron trans-splicing in *Chlamydomonas reinhardtii*. *EMBO J.* **20**, 1765-1773.

Rochaix, J.-D. (1992) Post-transcriptional steps in the expression of chloroplast genes. *Annu. Rev. Cell Biol.* **8**, 1-28.

Rochaix, J.-D. (2001) Posttranscriptional control of chloroplast gene expression. From RNA to photosynthetic complex. *Plant Physiol.* **125**, 142-144.

Rouvière-Yaniv, J., and Gros, F. (1975) Characterization of a novel, low-molecular-weight DNA-binding protein from *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 3428-3432.

Sakai, A. (2001) In vitro transcription / DNA synthesis system using isolated organelle-nuclei: Application to the analysis on the mechanisms regulating the function of organelle genomes. J. Plant Res. 114, 199-211.

Sakai, A., Saito, C., Inada, N., and Kuroiwa, T. (1997) Transcriptional activities of the chloroplast-nuclei and proplastid-nuclei isolated from tobacco exhibit different sensitivities to tagetitoxin: implicatikon of the presence of distinct RNA polymerase. Plant Cell Physiol. 39, 923-934.

Sakai, A., Suzuki, T., Miyamura, Y., Kawano, S., Nagata, T., and Kuroiwa, T. (1998a) Comparative analysis of plastid gene expression in tobacco chloroplasts and proplastids: relationship between transcription and transcript accumulation. *Plant Cell Physiol.* **39**, 581-589.

Sakai, A., Suzuki, T., Miyazawa, Y., and Kuroiwa, T. (1998b) Simultaneous isolation of cell-nuclei, plastid-nuclei and Mitochondrial-nuclei from cultured tobacco cells; comparative analysis of their transcriptional activities in invitro. *Plant Sci.*, **133**, 17-31

Sakai, A. Miyazawa, Y., Suzuki, T., Sasaki, N., Kawano, S., Kuroiwa, T. (1999) Plastid gene expression during amyloplast formation in cultured tobacco cells. *J. Plant Physiol.* **154**, 71-78.

Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K., Chiden, Y., Hayashi, M., Okamoto, M., *et al.* (2002) *Nature* **420**, 312-316.

Sato, N. (2000) SISEQ: Manipulation of multiple sequence and large database files for

common platforms. *Bioinformatics* **16**, 180-181.

Sato, N. (2001) Was the evolution of plastid genetic machinery discontinuous? *Trends Plant Sci.* **6**, 151-156.

Sato, N. (2002) Comparative analysis of the genomes of cyanobacteria and plants. *Genome Informatics* **13**, 173-182.

Sato, N., and Ohta, N. (2001) DNA-binding specificity and dimerization of the DNA-binding domain of the PEND protein in the chloroplast envelope membrane. *Nucleic Acids. Res.* **29**, 2244-2250.

Sato, N., Albrieux, C. Joyard, J., Douce, R., and Kuroiwa, T. (1993) Detection and characterization of a plastid envelope DNA-binding protein which may anchor plastid nucleoids. *EMBO J.* **12**, 555-561.

Sato, N., Misumi, O., Joyard, J., and Douce, R. (1995) DNA-binding proteins mediate interaction of nucleoids with envelope membrane in developing plastids. *In* Photosynthesis: from Light to Biosphere. (Mathis, P., Ed.) Vol. III. pp. 635-638. Kluwer Academic Publishers, Dordrecht.

Sato, N., Misumi, O., Shinada, Y., Sasaki, M., and Yoine, M. (1997) Dynamics of localization and protein composition of plastid nucleoid in light-grown pea seedlings. *Protoplasma* **200**, 163-173.

Sato, N., Ohshima, K., Watanabe, A., nishiyama, Y., Joyard, J., and Douce, R. (1998) Molecular characterization of the PEND protein, a novel bZIP protein present in the envelope membrane that is the site of nucleoid replication in developing plastids. *Plant Cell* **10**, 859-872.

Sato, N., Rolland, N., Block, M. A. and Joyard, J. (1999) Do plastid envelope membranes play a role in the expression of the plastid genome? *Biochimie* **81**, 619-629.

Sato, N., Nakayama, M., and Hase, T. (2001) The 70-kDa major DNA-compacting protein of

the chloroplast nucleoid is sulfite reductase. *FEBS Lett*., 487, 347-350.

Sato, N., Sekine, K., Kabeya, Y., Ehira, S., Onuma, M., and Ohta, N. (2003) Discontinuous evolution of plastid genomic machinery: Radical replacement of major DNA-binding proteins. In "(Endo)Symbiosis and Eukaryotic Organells" (Sugiura, M. ed.), Logos Verlag, Berlin (in press).

Satoh, J., Baba, K., Nakahira, Y., Tsunoyama, Y., Shiina, T., and Toyoshima, Y. (1999) Developmental stage-specific multi-subunit plastid RNA polymerases (PEP) in wheat. *Plant J*., 18, 407-415

Sekine, K., Hase, T., and Sato, N. (2002) Reversible DNA conpaction by sulfite reductase regulates transcriptional activity of chloroplast nucleoids. *J. Biol. Chem*., **277**, 24399-24404.

Schmitz-Linneweber, C., Maier, R. M., Alcaraz, J.-P., Cottet, A., Herrmann, R. G., and Mache, R. (2001) The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. *Plant Mol. Biol.* **45**, 307-315.

Schubert, M., Petersson, U. A., Haas, B. J., Funk, C., Schröder, W. P., and Kieselbach, T. (2002) Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. *J. Biol. Chem.* **277**, 8354-8365.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**, 81-86.

Shirano, Y., Shimada, H., Kanamaru, K., Fujiwara, M., Tanaka, K., Takahashi, H., Unno, K., Sato, S., Tabata, S., Hayashi, H., Miyake, C., Yokota, A., and Shibata, D. (2000) Chloroplast development in *Arabidopsis thaliana* requires the nuclear-encoded transcription factor Sigma B. *FEBS Letters* **485**, 178-182.

Stiller, J. W., Reel, D. C., and Johnson, J. C. (2003) A single origin of plastids revisited:

convergent evolution in organellar genome content. *J. Phycol.* **39**, 95-105.

Sugiura, M. (1992) The chloroplast genome. *Plant Mol. Biol.* **19**, 149-168.

Sugiura, M., Hirose, T., and Sugita, M. (1998) Evolution and mechanism of translation in chloroplasts. *Annu. Rev. Genet.* **32**, 437-459.

Tan, S., and Troxler, R. (1999) Characterization of two chloroplast RNA polymerse sigma factors from *Zea mays*: photoregulation and differential expression. *Proc. Natl. Acad. Sci. U. S. A*., **96**, 5316-5321

Tanaka, K., Oikawa, K., Ohta, N., Kuroiwa, H., Kuroiwa, T., and Takahashi, H. (1996) Nuclear encoding of a chloroplast RNA polymerase sigma subunit in a red alga. *Science*, 272, 1932-1935

Tanaka, K., Tozawa, Y., Mochizuki, N., Shinozaki, K., Nagatani, A., Wakasa, K., and Takahashi, H. (1997) Characterization of three cDNA species encoding plastid RNA polymerase sigma factors in Arabidopsis thaliana: Evidence for the sigma factor heterogeneity in higher plant plastids. *FEBS Lett*., 413, 309-313

Taylor, W. C. (1989) Regulatory interactions between nuclear and plastid genomes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 211-233.

Terasawa, K., and Sato, N. (2003) Analysis of intracellular localization of the PEND protein. Plant Biology 2003, Abstract book.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Thum, K. E., Kim, M., Morishige, D. T., Eibl, C., Koop, H.-U., and Mullet, J. E. (2001) Analysis of barley chloroplast *psbD* light-responsive promoter elements in transplastomic tobacco. *Plant Mol. Biol.* **47**, 353-366.

Ting, C. S., Rocap, G., King, J., and Chisholm, S. W. (2002) Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol.* **10**, 134-142.

Turmel, M., Otis, C., and Lemieux, C. (1999) The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: Insights into the architecture of ancestral chloroplast genomes. *Proc. Natl. Acad. Sci. USA* **96**, 10248-10253.

Turner, S., Pryer, K. M., Miao, V. P. W., and Palmer, J. D. (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J. Eukaryot. Microbiol.* **46**, 327-338.

Van de Peer, Y., Baldauf, S. L., Doolittle, W. F., and Meyer, A. (2003) An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. *J. Mol. Evol.* **51**, 565-576.

van Hom, R. C., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernandez, J. M., Jimenez, L., Postigo, M., Silva, F. J., Tamames, J., Viguera, E., Latorre, A., Valencia, A., Moran, F., and Moya, A. (2002) Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 581-586.

Vogl, C., Badger, J., Kearney, P., Li, M., Clegg, M., and Jiang, T. (2003) Probabilistic analysis indicates discordant gene trees in chloroplast evolution. *J. Mol. Evol.* **56**, 330-340.

Waldmüller, S., Müller, U. and Link, G. (1996) GSBF1, a seedling-specific bZIP DNA-binding protein with preference for a 'split' G-box-related element in *Brassica napus* RbcS promoters. Plant Mol. Biol. 32, 631-639.

Wang, S and Liu, X.-Q. (1991) The plastid genome of Cryptomonas Φ encodes and hsp70-like protein, a histone-like protein, and an acyl carrier protein. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10787-10787

Wang, Y., and Shadel, G. (1999) Stability of the mitochondrial genome requires an amino-terminal domain of yeast mitochondrial RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 8046-8051

Watanabe, N., Che, F. S., Iwano, M., Takayama, S., Yoshida, S., and Isogai, A. (2001) Dual targeting of spinach protoporphyrinogen oxidase II to mitochondria and chloroplasts by alternative use of two in-frame initiation codons. *J. Biol. Chem.* **276**, 20474-2048.

Wu, H., and Liu, X.-Q. (1997) DNA binding and bending by a chloroplast-encoded HU-like protein overexpressed in *Escherichia coli*. *Plant Mol Biol.* **34**, 339-343.

Yang, J., Schuster, G., and Stern, D. B. (1996) CSP41, a sequence-specific chloroplast mRNA binding protein, is an endoribonuclease. *Plant Cell* **8**, 1409-1420.

Yang, S., Yu, X., VanLoock, M. S., Jezewska, M. J., Bujalowski, W., and Egelman, E. H. (2002) Flexibility of the rings: structural asymmetry in the DnaB hexameric helicase. *J. Mol. Biol.* **321**, 839-849.

Yurina, N. P., Belkina, G. G., Karapetyan, N. V., and Odintsova, M. S. (1995) Nucleoids of pea chloroplasts: microscopic and chemical characterization. Occurrence of histone-like proteins. *Biochem. Mol. Biol. Int.* **36**, 145-154.

Zelenaya-Troitskaya, O., Perlman, P. S., and Butow, R. A. (1995) An enzyme is yeast mitochondria that catalyzes a step in branched-chain amino acid biosynthesis also functions in mitochondrial DNA stability. *EMBO J.* **14**, 3268-3276.

Figure legends

Figure 1. Fluorescence micrographs showing the localization of nucleoids within the plastids. The cells were stained with DAPI and then photographed with a fluorescence microscope with a UV filter cube.

A1 (Nomarski differential interference image) and A2 (fluorescence), *Anabaena variabilis* M3 (filamentous cyanobacterium); B, fluorescence image of *Synechococcus* sp. PCC 7942 (unicellular cyanobacterium); C1 (Nomarski) and C2 (fluorescence), *Cyanidioschyzon merolae* 10D (red alga); D1 (Nomarski) and D2 (fluorescence), *Chlamydomonas reinhardtii* 5D (green alga); E1 and E2, etioplast and chloroplast of pea (flowering plant); F1 (Nomarski) and F2 (fluorescence), *Physcomitrella patens* (moss). Large arrowhead, nucleus; small arrowhead, plastid nucleoid; MtN, mitochondrial nucleoid. Scale is identical in all images.

Figure 2. Phylogeny of plastids and cyanobacteria.

A. A Neighbor-Joining tree of combined protein sequences that are conserved in eight cyanobacteria and 21 plastids. Each number indicates bootstrap confidence level based on 1,000 replications. The following 27 genes were used in the analysis: *rps2, rps3, rps7, rps8, rps11, rps12, rps19, rpl2, rpl14, rpl16, rpoC1, rpoC2, psaA, psaB, psaC, ycf4, psbA, psbB, psbC, rbcL, ccsA, petA, petB, petD, atpA, atpB, atpH*. Each class of protein sequences was individually aligned by the Clustal X program (Thompson *et al.*, 1994), and then the alignments were combined. The calculation was done with the MEGA2 program for Windows (Kumar *et al.*, 2001), but identical or essentially similar results were obtained with the Philip package for UNIX (protdist and neighbor commands) as well as the PAUP program for Mac OS. All sequence manipulation was done with the SISEQ package (Sato, 2001).

B. A condensed Neighbor-Joining tree of cyanobacteria, plastids and bacteria, using 14

non-photosynthesis genes. The genes included were: *rps2, rps3, rps7, rps8, rps11, rps12, rps19, rpl2, rpl14, rpl16, rpoC1, rpoC2, atpA,* and *atpB.*

Database sources (GenBank accession numbers are indicated except otherwise stated). Chloroplasts: *Anthoceros formosae* (hornwort, AB086179); *Arabidopsis thaliana* (angiosperm, AP000423); *Atropa belladonna* (angiosperm, NC_004561); *Chaetosphaeridium globosum* (coleochaete, NC_004115); *Chlorella vulgaris* C-27 (green alga, AB001684); *Cyanidioschyzon merolae* 10D (red alga, CGP); *Cyanidium caldarium* RK1 (red alga, AF022186); *Cyanophora paradoxa* (glaucophyte, CPU30821); *Guillardia theta* (cryptophyte, NC_000926); maize (*Zea mays*, angiosperm, ZMA86563); *Marchantia polymorpha* (liverwort, X04465); *Mesostigma viride* (prasinophyte, AF166114); *Nephroselmis olivacea* (prasinophyte, NC_000927); *Odontella sinensis* (diatom, NC_001713); *Oenothera elata* subsp. Hookeri (angiosperm, OEL271079); pine (*Pinus thunbergii*, gymnosperm, D17510); *Porphyra purpurea* (red alga, U38804); *Psilotum nudum* (pteridophyte, AP004638); rice (*Oryza sativa*, angiosperm, NC_001320); tobacco (*Nicotiana tabacum*, angiosperm, NC_001879); wheat (*Triticum aestivum*, angiosperm, NC_002762). Cyanobacteria: *Anabaena* sp. PCC 7120, BA000019; *Nostoc punctiforme*, JGI; *Synechocystis* sp. PCC 6803, AB001339; *Thermosynechococcus* sp. BP-1, BA000039; *Prochlorococcus marinus* MED4, JGI; *Prochlorococcus marinus* MIT9313, JGI; *Synechococcus* sp. WH8102, JGI; *Trichodesmium erythraeum*, JGI. Other bacteria: *Escherichia coli* K-12, U00096; *Bacillus subtilis* 168, AL009126; *Rhodopseudomonas palustris*, JGI; *Chlorobium tepidum* TLS, AE006470.

Figure 3. Direct visualization of plastid nucleoids with a Green Fluorescent Protein (GFP) derivative. A DNA construct encoding the DNA-binding domain called ‘cbZIP’ of the PEND protein fused to a synthetic GFP (Chiu *et al.*, 1995) was stably transformed into *Arabidopsis*

thaliana. In this line, the cbZIP-GFP is synthesized within the cytoplasm, and then targeted to the plastids. Then, the protein binds to the chloroplast DNA and localizes to the plastid nucleoids. The localization of plastid nucleoids is directly visualized under a fluorescence microscope. A. a schematic diagram showing the DNA construct, which was inserted into a pBI101-derived plasmid and then introduced into *Arabidopsis* plant via *Agrobacterium* infection. B. fluorescence image of a three-day-old cotyledon (low magnification); C, fluorescence image of a leaf cell of a 3-day-old plant; D1 (fluorescence) and D2 (Nomarski), a pair of guard cells of a 10-day-old leaf; E, fluorescence image of a 3-day-old root cell; F1 (fluorescence) and F2 (Nomarski), an epidermal cell of a 10-day-old leaf.

Figure 4. DNA compacting activity of HU protein and sulfite reductase.

A. Phylogenetic tree of bacterial and plastid-encoded HU proteins. This tree was constructed by the Neighbor-Joining method using the PAM distance matrix with the MEGA2 program. Each number on the branch indicates bootstrap confidence level in percentage. Bar indicate 0.1 substitutions per site. The following data were used. Cyanobacteria: *Synechococcus* sp. WH8102, Joint Genome Initiative (JGI) web site; *Prochlorococcus marinus* MED4, JGI web site; *Prochlorococcus marinus* MIT9313, JGI web site; *Synechocystis* sp. PCC 6803 (GenBank AB001339: sll1712); *Anabaena* sp. PCC 7120 (SwissProt S78244); *Thermosynechococcus* sp. BP-1 (GenBank BA000039). Bacteria: *Escherichia coli* A and B subunits (SwissProt P02342 and P02341, respectively); *Bacillus subtilis* (SwissProt P08821); *Mesorhizobium meliloti* (SwissProt P02344); *Rickettsia prowazekii* (GenBank AJ235269); *Clostridium pasteurianum* (SwissProt P05385). Algae and Apicomplexa: *Toxoplasma gondii* (EST, GenBank W06256); *Cyanidioschyzon merolae* (GenBank AB071964); *Guillardia theta* (SwissProt P29214); *Chlamydomonas reinhardtii* (EST, GenBank AV387318).

B and C. Formation of DNA-protein complex as visualized by fluorescence microscopy.

Recombinant *C. merolae* HU (B) or maize sulfite reductase (SiR) were mixed with pea chloroplast DNA and allowed to form complex overnight. The complex was stained with DAPI and then examined by fluorescence microscope. Bar indicates 10 μm.
D and E. Effects of HU (D) and SiR (E) on the transcription activity of isolated plastid nucleoids in *C. merolae* (Cme) and pea. Relative activity with respect to the activity without addition of DNA-binding protein is indicated.

Figure 5. Schematic diagram of putative organellar DNA polymerases in plants.
A. Structural comparison of prokaryotic DNA polymerase and putative organellar DNA polymerases of *Arabidopsis thaliana*, rice and *Cyanidioschyzon merolae*. Black rectangle, putative transit sequence (TargetP prediction of targeting to mitochondrion or chloroplast; Emanuelsson *et al.*, 2000); shaded rectangle, 5'-3' exonuclease domain; striped rectangle, 3'-5' exonuclease domain; hatched rectangle, polymerase domain; horizontally striped part, DEAD-box helicase domain. All domain identification was based on similarity search using the hmmpfam program with the Pfam database (Bateman *et al.*, 2002) version 7.7.
B. Phylogenetic relationship of putative plant organellar DNA polymerases and bacterial DNA polymerases I. This is a Neighbor-Joining tree constructed by the MEGA2 program using the amino acid sequences as follows: *Anabaena* sp. PCC 7120 (alr1254); *Nostoc punctiforme* (contig 429, gene 46), JGI; *Synechocystis* sp. PCC 6803 (slr0707), SwissProt Q55971; *Thermosynechococcus* sp. BP-1 (tll1339); *Prochlorococcus marinus* MED4 (gene 1552), JGI; *Prochlorococcus marinus* MIT9313 (gene 3388), JGI; *Synechococcus* sp. WH8102 (gene 1957), JGI; *Thermus aquaticus*, SwissProt P19821; *Helicobacter pylori*, SwissProt P56105; *Lactococcus lactis*, SwissProt Q9CDS1; *Streptococcus pneumoniae*, SwissProt P13252; *Bacillus subtilis*, SwissProt O34996; *Escherichia coli*, SwissProt P00582;

Rickettsia felis, SwissProt Q9RAA9; *Sinorhizobium meliloti*, GenBank AL591782; *Rhodopseudomonas palustris* (gene 579), JGI; *Mycobacterium tuberculosis*, SwissProt Q07700; *Cyanidioschyzon merolae* A and B, Cyanidioschyzon Genome Project (Matsuzaki *et al.*, 2003; to be published); *Rhodothermus marinus*, GenBank AF121780; *Arabidopsis thaliana*, At1g50840, At3g20540, At4g32700, MIPS database; rice, OsPolI-like, GenBank AB047689; *Caenorhabditis elegans*, W03A3.2. GenBank. Essentially similar topology of tree was obtained by Maximum Parsimony (with MEGA2 and Phylip) and quartet puzzling (with Tree-puzzle). Each number indicates the bootstrap confidence level obtained from 1,000 replicates. Note that the DNA polymerases of cyanobacteria and plastids are divergent.

Figure 6. Absence of NEP in the chloroplasts of *Physcomitrella patens*.

A. Tagetitoxin sensitivity of the transcription activity in the chloroplasts and mitochondria. Tagetitoxin is a potent inhibitor for prokaryotic and plastid-encoded RNA polymerases.

B. Effect of 5' untranslated sequence on the translation and targeting of RPOT1 protein in *Physcomitrella patens.* Various chimeric constructs containing the 5' untranslated sequence of RPOT1 and GFP (hatched part) with or without translation leader sequence of pea RbcS 3A gene (black part) were introduced into moss protoplasts by the PEG-mediated transformation. After 24 hr, the cells were examined under a fluorescence microscope. Note that the polypeptide translated from the first methionine was targeted to chloroplasts, while the polypeptide translated from the second methionine was targeted to mitochondria. However, the first methionine is not efficiently used as the initiator in the natural context.

C. Two possible initiator methionines at the 5' end of the 'dually targeted' *RPOT* genes in *P. patens*, *A. thaliana* and *Nicotiana sylvestris*. In the two RPOT genes of *P. patens*, the first methionine is not used as the initiation codon in the native context. In *A. thaliana* and *N. sylvestris*, one of the three RPOTs has been claimed to be 'dually targeted', but the only data

that support dual targeting were based on the GFP-fusion experiments with forced translation constructs.

Figure 7. A current view on the discontinuous evolution of plastid genomic maehinery. The creation of initial photosynthetic eukaryote involved an ancestral mitochondriate eukaryote and an ancestor of cyanobacteria (bottom). The nature of DNA polymerase of the mitochondria in this ancestral eukaryote remains to be identified (see text). The initial endosymbiosis might be either a single event, or parallel (multiple) events, or maybe repeated several times (see text). The chromosomal and plastid DNA are shown but mitochondrial DNA is not shown for simplicity. For details see text.

Table 1. Survey on the representative prokaryotic DNA-binding proteins and transcription factors in plants and algae

Protein	Eu-bacteria	Cyano-bacteria	*Buchnera* sp.	Rhodophytes and chromophytes		Green algae		Land plants	
				cp	nuc	cp	nuc	cp	nuc
HU, IHF	1-4	1	1	0-1	0	0	1	0	0
H-NS,StpA	0-2	0	1	0	0	0	0	0	0
Fis	0-1	0	1	0	0	0	0	0	0
Dps	0-1	1-2	0	0	0	0	0	0	0
DnaB	1	1	1	1	0	0	0	0	0
DnaA	1	1	1	0	0	0	0	0	0
OmpR homologues	some	some	0	1-2	0	0	0	0	0
NtcA	0-1	1	0	1	0	0	0	0	0
CbbR, RbcR	0-2	1-3	0	1	0	0	0	0	0

Each number indicates the copy number of genes encoding respective protein. Cp, chloroplast DNA; nuc, nuclear DNA. The data for the nuclear genomes of rhotophytes and chromophytes are taken from the nuclear genome of a rhodophyte *Cyanidioschyzon merolae* (Matsuzaki *et al.*, 2003; Cyanidioschyzon Genome Project). The data for the nuclear genomes of green algae are based on the EST data of *Chlamydomonas reinhardtii*. The superfamilies, such as HU and IHF, and H-NS and StpA, are shown as a single category.

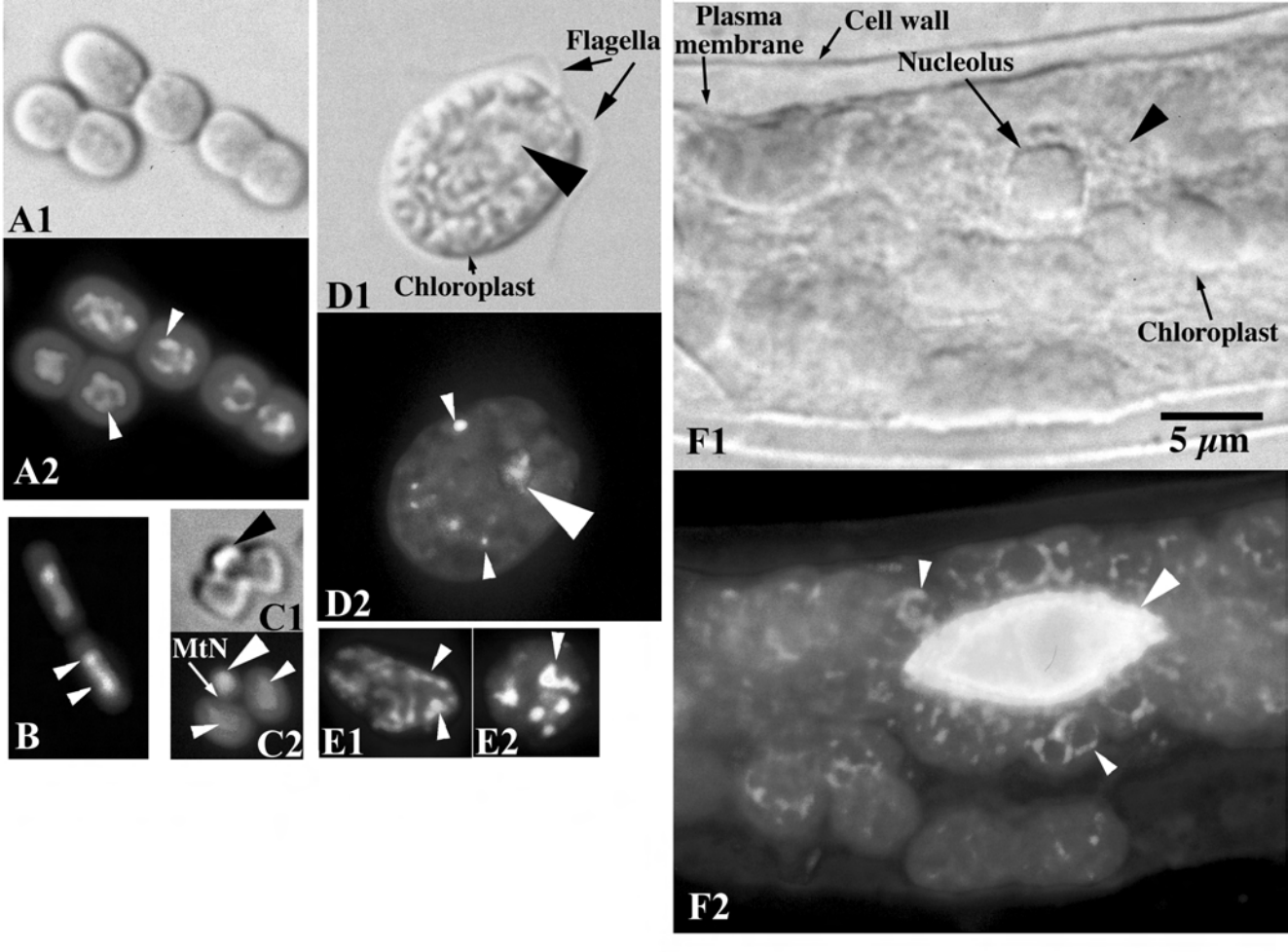

A1
A2
B
C1
MtN
C2
D1
Flagella
Chloroplast
D2
E1
E2
Plasma membrane
Cell wall
Nucleolus
Chloroplast
F1
5 μm
F2

Fig. 1

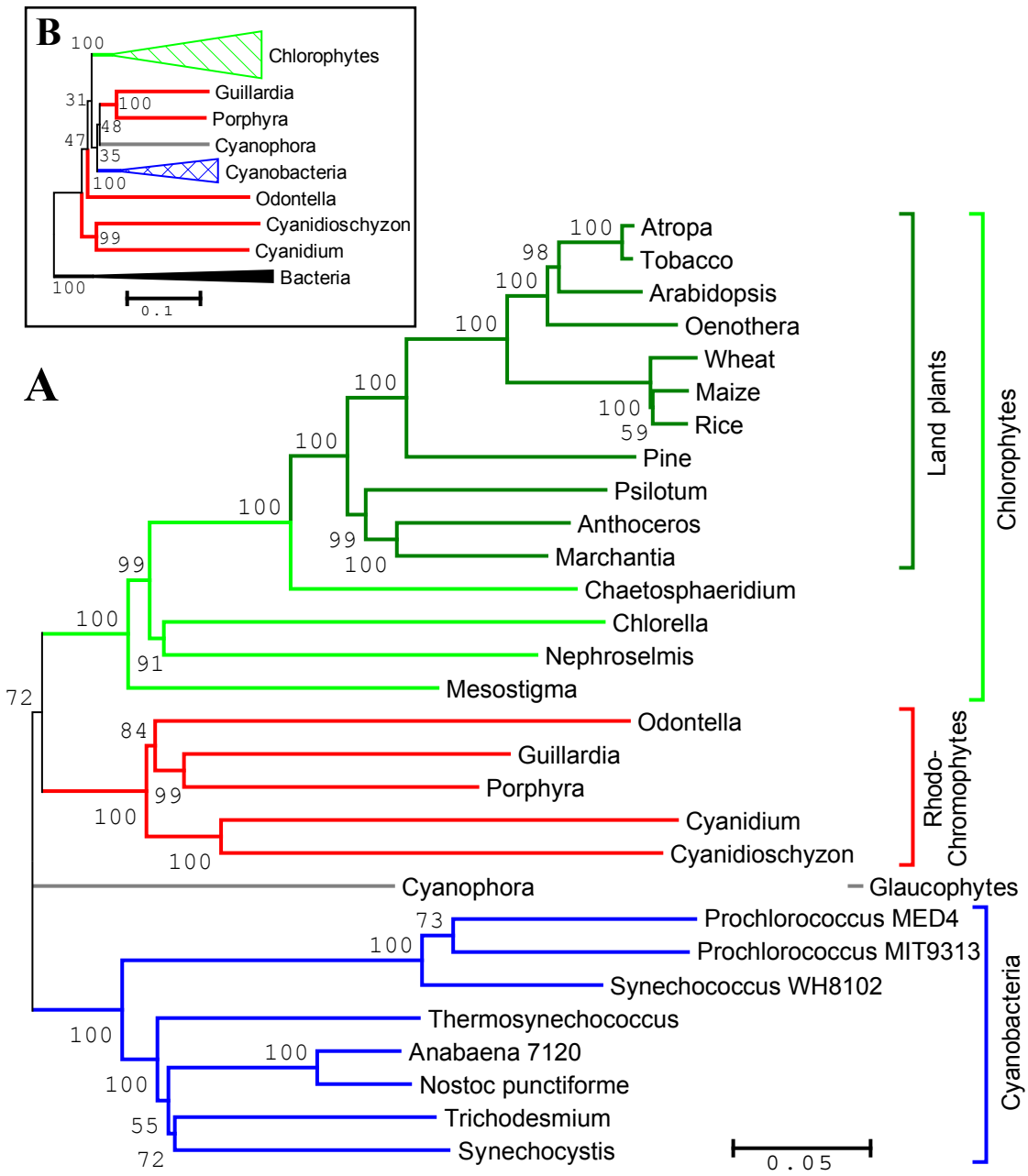

B
Chlorophytes
Guillardia
Porphyra
Cyanophora
Cyanobacteria
Odontella
Cyanidioschyzon
Cyanidium
Bacteria
0.1
A
Atropa
Tobacco
Arabidopsis
Oenothera
Wheat
Maize
Rice
Pine
Psilotum
Anthoceros
Marchantia
Chaetosphaeridium
Chlorella
Nephroselmis
Mesostigma
Odontella
Guillardia
Porphyra
Cyanidium
Cyanidioschyzon
Cyanophora
Prochlorococcus MED4
Prochlorococcus MIT9313
Synechococcus WH8102
Thermosynechococcus
Anabaena 7120
Nostoc punctiforme
Trichodesmium
Synechocystis
0.05
Land plants
Chlorophytes
Rhodo-
Chromophytes
Glaucophytes
Cyanobacteria

Fig. 2

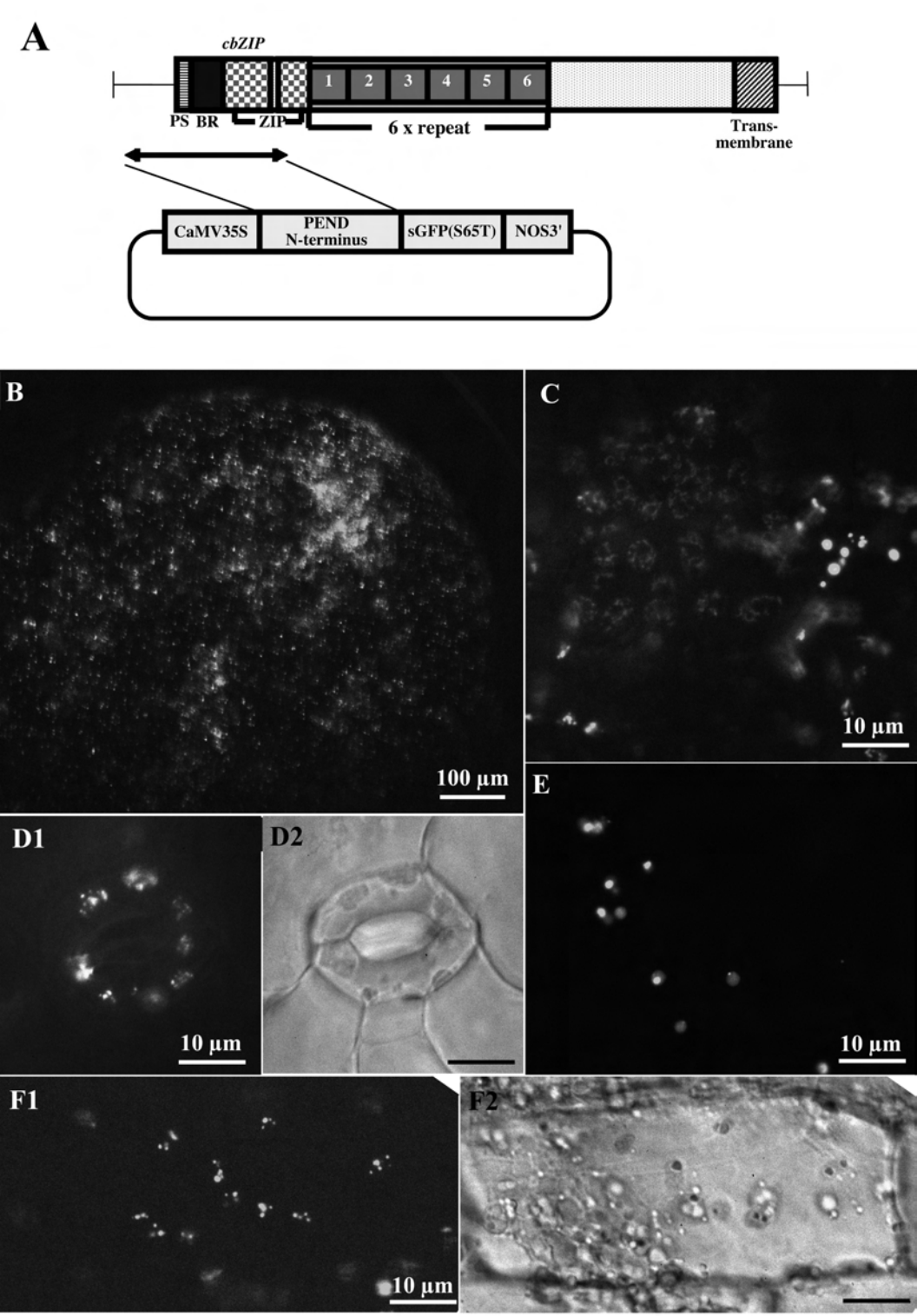

A
cbZIP
1
2
3
4
5
6
PS BR
ZIP
6 x repeat
Trans-
membrane
CaMV35S
PEND
N-terminus
sGFP(S65T)
NOS3'
B
100 µm
C
10 µm
D1
10 µm
D2
E
10 µm
F1
10 µm
F2

Fig. 3

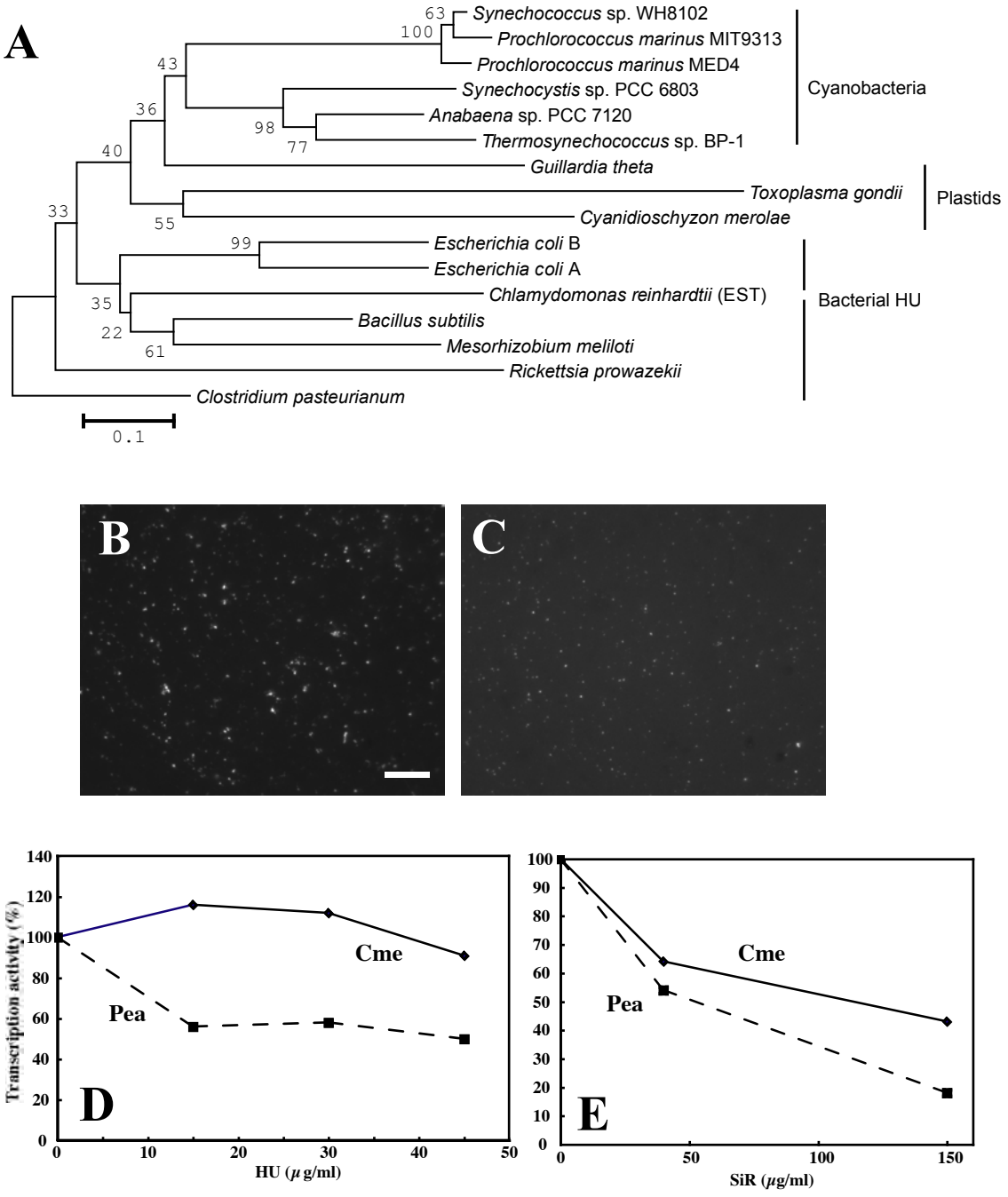

A
63
100
43
36
98
77
40
33
55
99
35
22
61
Synechococcus sp. WH8102
Prochlorococcus marinus MIT9313
Prochlorococcus marinus MED4
Synechocystis sp. PCC 6803
Anabaena sp. PCC 7120
Thermosynechococcus sp. BP-1
Guillardia theta
Toxoplasma gondii
Cyanidioschyzon merolae
Escherichia coli B
Escherichia coli A
Chlamydomonas reinhardtii (EST)
Bacillus subtilis
Mesorhizobium meliloti
Rickettsia prowazekii
Clostridium pasteurianum
0.1
Cyanobacteria
Plastids
Bacterial HU
B
C
D
Transcription activity (%)
Cme
Pea
HU (μg/ml)
E
Cme
Pea
SiR (μg/ml)

Fig. 4

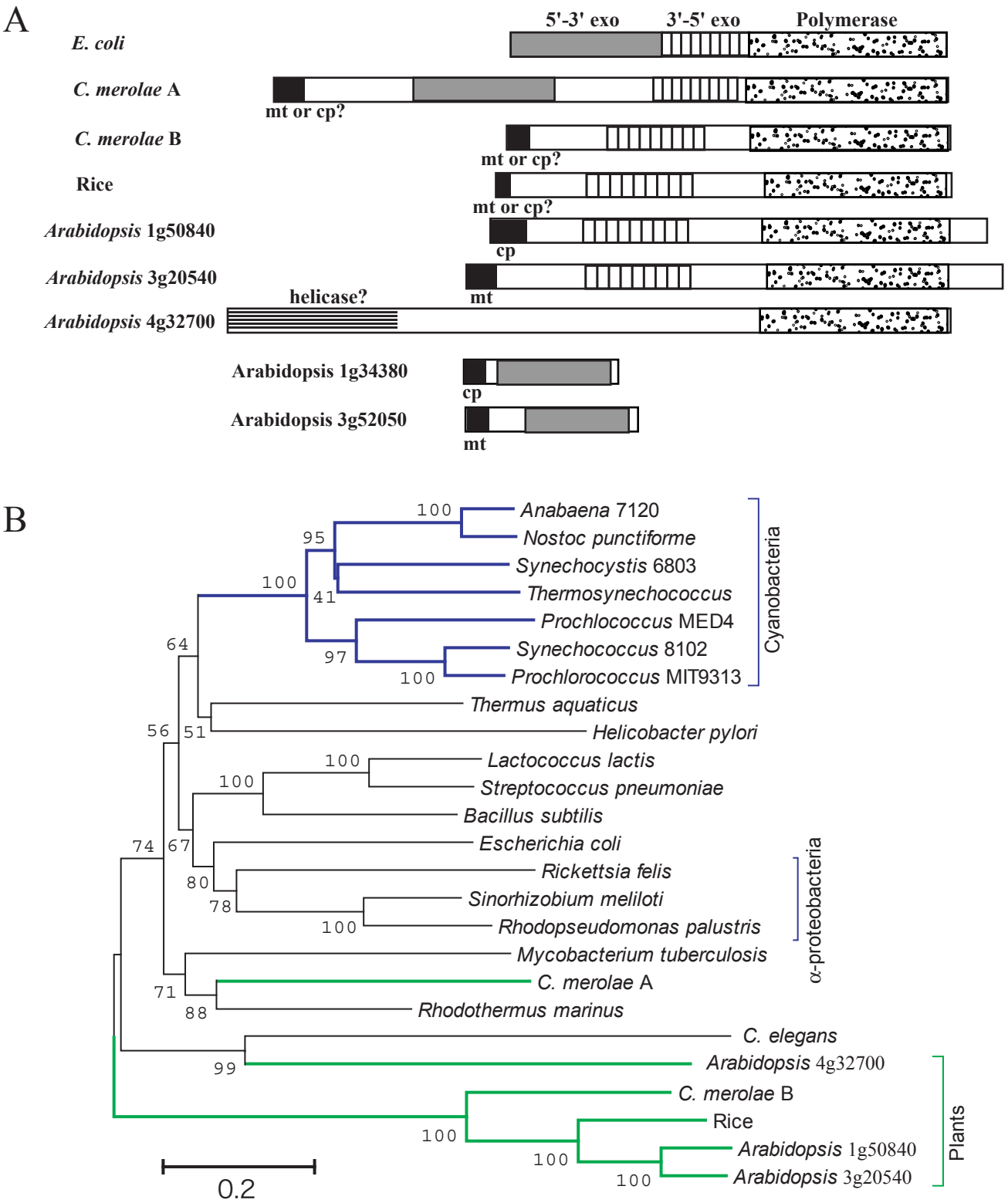
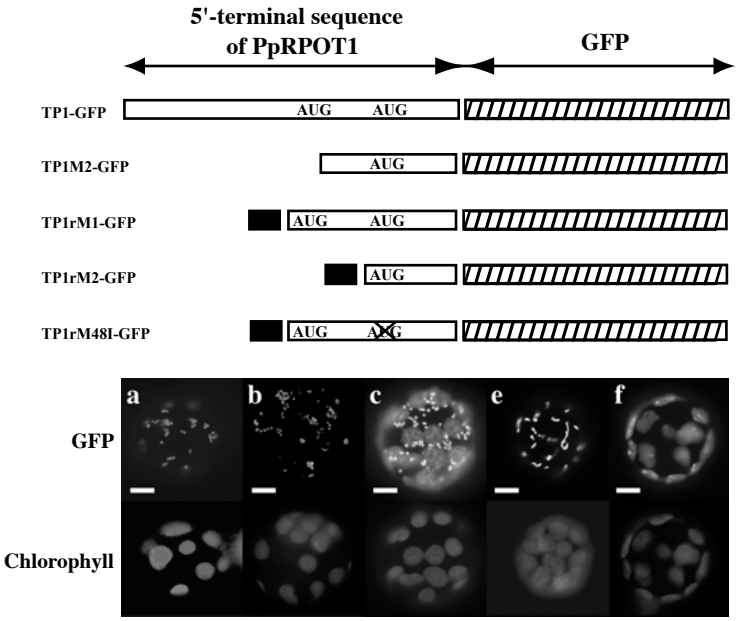

A
5'-3' exo
3'-5' exo
Polymerase
E. coli
C. merolae A
mt or cp?
C. merolae B
mt or cp?
Rice
mt or cp?
Arabidopsis 1g50840
cp
Arabidopsis 3g20540
mt
helicase?
Arabidopsis 4g32700
Arabidopsis 1g34380
cp
Arabidopsis 3g52050
mt
B
Anabaena 7120
Nostoc punctiforme
Synechocystis 6803
Thermosynechococcus
Prochlococcus MED4
Synechococcus 8102
Prochlorococcus MIT9313
Cyanobacteria
Thermus aquaticus
Helicobacter pylori
Lactococcus lactis
Streptococcus pneumoniae
Bacillus subtilis
Escherichia coli
Rickettsia felis
Sinorhizobium meliloti
Rhodopseudomonas palustris
α-proteobacteria
Mycobacterium tuberculosis
C. merolae A
Rhodothermus marinus
C. elegans
Arabidopsis 4g32700
C. merolae B
Rice
Arabidopsis 1g50840
Arabidopsis 3g20540
Plants
100
95
100
41
64
97
100
56
51
100
100
74
67
80
78
100
71
88
99
100
100
100
0.2

Fig. 5

A

Inhibitor	Incorporation (%)	
	mitochondria	plastids
None	100	100
Tagetitoxin	150.6	0.3

B



C

```
PpRPOT1 MVAIGVLEPISAVGRTGRRISDVKLVGLRSHQDCAPTFNLGGRVRGGMWRAAVRQLSRQPREGLRG
PpRPOT2 -----------MPAEVCWTKGILSTTACIFPEHVKQVLLTGYPVAGMWRSAAQQLARQKLHGVRS

NsRpoT-B --------MSSTKTPISLTIKLNQFTD-----KPTGLDINPYHNSPIMWRNIIKQLSSRTPQKLLF
AtRpoT;2 --------MSSAQTPLFLANQTKVFDHLIPLHKPFISSPNPVSQSFPMWRNIAKQAISRSAARLNV
```

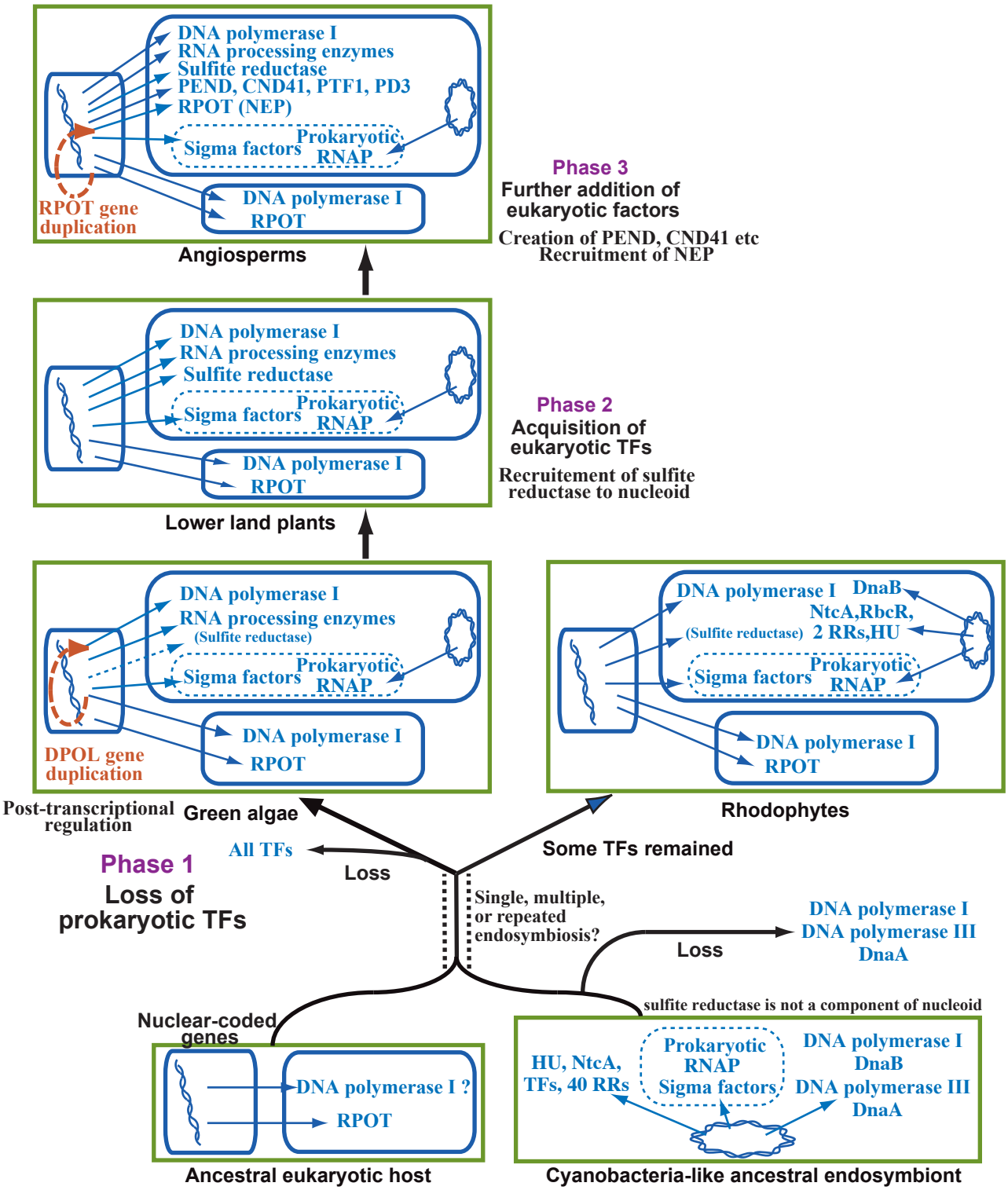

DNA polymerase I
RNA processing enzymes
Sulfite reductase
PEND, CND41, PTF1, PD3
RPOT (NEP)
Sigma factors
Prokaryotic RNAP
DNA polymerase I RPOT
RPOT gene duplication
Angiosperms
Phase 3
Further addition of eukaryotic factors
Creation of PEND, CND41 etc
Recruitment of NEP
DNA polymerase I
RNA processing enzymes
Sulfite reductase
Sigma factors
Prokaryotic RNAP
DNA polymerase I RPOT
Lower land plants
Phase 2
Acquisition of eukaryotic TFs
Recruitement of sulfite reductase to nucleoid
DNA polymerase I
RNA processing enzymes
(Sulfite reductase)
Sigma factors
Prokaryotic RNAP
DNA polymerase I RPOT
DPOL gene duplication
DNA polymerase I
DnaB
NtcA,RbcR,
(Sulfite reductase)
2 RRs,HU
Sigma factors
Prokaryotic RNAP
DNA polymerase I RPOT
Rhodophytes
Post-transcriptional regulation
Green algae
All TFs
Loss
Some TFs remained
Phase 1
Loss of prokaryotic TFs
Single, multiple, or repeated endosymbiosis?
Loss
DNA polymerase I
DNA polymerase III
DnaA
sulfite reductase is not a component of nucleoid
Nuclear-coded genes
DNA polymerase I ?
RPOT
Ancestral eukaryotic host
HU, NtcA, TFs, 40 RRs
Prokaryotic RNAP
Sigma factors
DNA polymerase I
DnaB
DNA polymerase III
DnaA
Cyanobacteria-like ancestral endosymbiont

Fig. 7