Intracellular Localization and Sequence Variation of the Dinoflagellate Minicircle DNA

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Biological Science

Satomi OWARI

TABLE OF CONTENTS

Abstract 1

Abbreviations 4

1. General Introduction 5

1.1. Plastids 5

1.1.1. Plastids have originated through endosymbioses 5

1.1.2. Evolution of plastid DNA gene composition 6

1.1.3. Maintenance of plastid DNA 6

1.2. Dinoflagellates 7

1.2.1. General characteristics of dinoflagellates 7

1.2.2. Phylogenetic relationship of dinoflagellates 8

1.2.3. Plastid diversity of dinoflagellates 9

1.3. Minicircle DNA of dinoflagellates 10

1.4. Purpose of this study 12

2. Intracellular Localization of Minicircle DNA in Dinoflagellates 13

2.1. Introduction 13

2.2. Materials 14

2.2.1. *Amphidinium massartii* 14

2.2.2. *Hemidinium nasutum* 15

2.3. Methods 16

2.3.1. Algal cultures 16

2.3.2. DNA extraction 17

2.3.3. Sequence analysis for *H. nasutum* minicircle DNAs 17

2.3.4. Southern blotting to determine the structure of *H. nasutum* minicircle DNAs 18
2.3.5. Southern hybridization to confirm the specificity of probes for *A. massartii* DNA-targeted whole-cell FISH 18
2.3.6. DNA-targeted whole-cell FISH of *A. massartii* and *H. nasutum*. 19
2.4. Results 21
2.4.1. Minicircle DNA sequences and minicircle size in *H. nasutum* 21
2.4.2. Localization of *Amphidinium massartii* minicircle DNAs by FISH 23
2.4.3. Localization of *Hemidinium nasutum* minicircle DNAs by FISH 25
2.5. Discussion 25
2.5.1. Technique of DNA-targeted whole-cell FISH 25
2.5.2. Localization of plastid minicircle DNAs 26
2.5.3. Conclusions and future perspectives 27
3. Sequence Variation of the Dinoflagellate *Hemidinium nasutum* Minicircle DNA 29
3.1. Introduction 29
3.2. Material and Methods 30
3.2.1. Algal culture 30
3.2.2. DNA and RNA extraction 30
3.2.3. Sequence analysis for *H. nasutum* minicircle DNAs 31
3.3. Results 32
3.3.1. DNA and cDNA sequences of the *psbA* coding region in *H. nasutum* 32

3.3.2. Sequence variation in minicircle DNA *psbA* coding regions......... 33
3.3.3. DNA and cDNA sequences of the 23S rRNA coding region in *H. nasutum* 34
3.3.4. Sequence variation in minicircle DNA 23S rRNA gene coding regions 34
3.3.5. Sequences of the non-coding regions of *H. nasutum* minicircle DNAs 35
3.4. Discussion 36
3.4.1. *psbA* coding sequences of *H. nasutum* minicircle DNA are highly variable. 36
3.4.2. Difference between DNA and cDNA in the proportion of possible functional *psbA* sequence 37
3.4.5. Sequence variations in the DNA non-coding region of minicircle DNA... 40
3.4.5. Conclusion and a future perspective... 41
4. General Discussion... 43
Acknowledgements 45
References 46
Tables 61
Figures... 64

Abstract

Photosynthetic dinoflagellates are known to have various types of plastids that have different origins and pigment compositions. The most typical dinoflagellate plastid contains peridinin as its major xanthophyll pigment and is a secondary plastid derived from a red alga. Interestingly, the plastid DNA of peridinin-containing dinoflagellates had been unknown for many years, but in 1999, approximately 2–6 kb of plasmid-like DNA molecules, "minicircle DNAs," each of which encoded a single plastid gene, were reported. Thus far, minicircle DNA has been confirmed in several marine dinoflagellates. However, despite that minicircle DNAs are known to encode plastid genes, there has been no convincing evidence that the minicircle DNAs are actually localized in the plastid.

The mRNA of *psbA* genes, which is encoded on minicircle DNAs in *Symbiodinium,* has been found in the plastids of this species (Takishita et al. 2003), although it has not been directly proven that minicircle DNA itself is localized in the plastid. On the other hand, it has also been reported that minicircle DNA in *Ceratium* is localized in the nucleus (Laatsch et al. 2004). Thus, clarifying the localization of minicircle DNA encoding plastid genes of dinoflagellates is a very important issue, and accordingly, one of the purposes of this study was to provide direct evidence for the subcellular localization of minicircle DNAs.

To accomplish this, I selected a marine dinoflagellate, *Amphidinium massartii*, and a freshwater species, *Hemidinium nasutum*. I performed

whole-cell fluorescence *in situ* hybridization (FISH) and used fluorescence microscopy and confocal laser microscopy to visualize the localization of the minicircle DNAs. I observed intracellular localization of the minicircle DNA probes (*psbA*, core sequence) and found that the minicircle DNA signals of *psbA* and the core sequence were scattered in the region of plastid autofluorescence and at the periphery of this region, but was not observed in the cell nucleus. From these results, it was strongly suggested that minicircle DNA was localized in the plastid.

In this study, I demonstrated the localization of minicircle DNA in the plastids of two species from distant dinoflagellate lineages. Thus, when considered in conjunction with indirect evidences of previous studies, it was clear that minicircle DNA was localized in the plastids of many dinoflagellate lineages. This supports the hypothesis that minicircle DNAs have evolved through fragmentation of the plastid genome. In addition, since the minicircle DNA signals of *A. massartii* and *H. nasutum* were scattered in the plastid, it was implied that minicircle DNA forms nucleoid-like structures in the plastid.

In addition, the localization study also indicated that there were many "aberrant" copies among the minicircle DNA sequences of *H. nasutum*. Thus, the secondary objective of the study was to clarify the extent of minicircle DNA sequence variation in *H. nasutum*. The partial sequences of the *psbA* and 23S rRNA coding regions were amplified using genomic PCR and RT-PCR, and non-coding sequences were amplified using genomic PCR. The mutation types of both genes included base substitutions and

insertions/deletions, which were similar to those reported as “aberrant” in previous studies, although the variant frequency was very high, and there were no identical sequences among them, except for several *psbA* cDNA sequences. As for the *psbA* cDNA sequences, three sequences were identical and the remaining eight were unique. The percentage of variants containing frameshift mutations was also very high, comprising 95 percent of DNA sequences and 64 percent of cDNA sequences.

For *H. nasutum*, in which minicircle DNA with the coding and non-coding regions had not been reported, the large number of sequence variations indicated that frequent base substitution or recombination must occur and that the DNA repair mechanism was not functioning properly. Furthermore, it was interesting that the percentage of *psbA* DNA sequences containing frameshift mutations was higher than that of *psbA* cDNA. Whether some mechanism to select or repair exists at the transcription or translational level is one of the interesting research themes that will be explored in future.

In this study, the plastid localization of dinoflagellate minicircle DNA was demonstrated for the first time. As such, the result is significant and contributes to the current understanding of dinoflagellate cells. In addition, the minicircle DNA sequences from *H. nasutum* revealed that the maintenance system for minicircle DNA is relatively unstable. Thus, the fragile maintenance system of minicircle DNA might have driven the losses and replacements of plastids within various dinoflagellate lineages.

Abbreviations

bp, base pair

°C, degrees Celsius

DAPI, 6′ -diamidino-2-phenylindole

DNA, deoxyribonucleic acid

cDNA, complementary DNA

DNase, deoxyribonuclease

FISH, fluorescent *in situ* hybridization

FITC, fluorescein isothiocyanate

kb, kilobases

RNA, ribonucleic acid

mRNA, messenger RNA

tRNA, transfer RNA

RNase, ribonuclease

RT-PCR, reverse transcription PCR

rRNA, ribosomal RNA

1. General Introduction

1.1. Plastids

1.1.1. Plastids have originated through endosymbioses

Plastids are organelles that are responsible for performing many important functions in plant and algal cells, including photosynthesis and the synthesis of fatty acids, amino acids, and photosynthetic pigments (Sandelius & Aronsson 2009). The plastids initially originated from a single endosymbiotic cyanobacterium that was incorporated into a eukaryotic host cell (Reyes-Prieto et al. 2007; Gould et al. 2008) and gave rise to the "primary plastids." The primary plastids are surrounded by two bounding membranes and are only found in three lineages: the glaucophytes, the red algae, and the clade containing green algae and land plants (Keeling 2010).

The plastids of other photosynthetic eukaryotes originated from a subsequent evolutionary process called secondary endosymbiosis, in which a primary algal cell was engulfed by another eukaryotic cell and incorporated into it as an organelle (Delwiche 1999; McFadden 2001; Archibald & Keeling 2002; Stoebe & Maier 2002; Palmer 2003; Keeling 2004; Archibald 2005; Gould et al. 2008), giving rise to "secondary plastids." The plastids in most algal lineages can be attributed to this process. The chlorarachniophytes and euglenophytes, for example, are known to have acquired their plastids from green algae, and the haptophytes, cryptomonads, heterokonts, dinoflagellates and apicomplexans have acquired their plastids from a red alga (Sanchez-Puerta et al. 2007; Keeling 2010).

1.1.2. Evolution of plastid DNA gene composition

As a result of plastid acquisition through endosymbiosis, the genome of the endosymbiotic cyanobacterium has been drastically reduced and has evolved into plastid DNA, which is usually circular and encodes approximately 100–200 genes for plastid proteins and ribosomal RNAs (rRNAs) (Palmer 1990; Reith & Munholland 1993; Simpson & Stern 2002; Ohta et al. 2003). The cyanobacterium genes that are no longer found in the plastid DNA are thought to have been lost or transferred to the nucleus (Stoebe & Kowallik 1999).

Indeed, most of the original plastid genes have moved to the host's nuclear genome in a process known as Endosymbiotic Gene Transfer (Archibald 2015; Timmis et al. 2004). Although many plastid genes have been transferred, some are specific to their respective lineages and universal plastid genes are also found (Green 2011). The functions of these shared genes are thought to be related to hydrophobicity (Adams & Palmer 2003) or redox control (Allen 2003). In addition, since the regular function of plastids is maintained with genes encoded by both plastid and nuclear genomes, complicated control systems are needed (Wise & Hoober 2007).

1.1.3. Maintenance of plastid DNA

Plastid DNA is relatively well conserved with respect to sequence and gene content (Palmer 1990) and, as such, is widely used for phylogenetic studies (Soltis et al. 1999). Notably, this conservation is maintained in spite

of constant DNA damage from radiation and exposure to the toxic reactive oxygen species produced by photosynthesis. This is because plastid DNA is maintained by the plastid DNA replication/recombination/repair (DNA-RRR) pathways (Day & Madesis 2007).

Plastid DNAs are associated with various proteins and form DNA-protein complexes, termed plastid nucleoids, which are also important to maintenance plastid DNAs (Sakai et al. 2004). These complexes contribute to the enzyme location, such as enzyme of DNA repair, DNA replication, recombination, transcription, and post-transcriptional control of gene expression (Krupinska et al. 2013).

1.2. Dinoflagellates

1.2.1. General characteristics of dinoflagellates

Dinoflagellates are unicellular protists that are characterized by a unique cell covering and two flagella that arise from the side and lie in surface grooves. Underneath their cell membranes are flattened vesicles (amphiesmal vesicles), which are filled with cellulose and form fortified thecal plates in the majority of lineages (thecate species) but are void of cellulose in other lineages (athecate or "naked" species). In addition, the cells contain a huge nucleus, called a dinokaryon, which contains the haploid nuclear genome that can be as large as 400 pg (cf. 3 pg in humans; Spector et al. 1981) and contains multiple copies of many genes (Shoguchi et al. 2013). The chromosomes of dinoflagellate genomes are also permanently condensed (Spector 1984) and do not use histones or form nucleosomes to package their

DNA (Gornik et al. 2012).

Approximately 4000 species, assigned to more than 550 genera, have been described, of which one half of the species and nearly three-quarters of the genera are known from the fossil record (Taylor 1990). The fossil species are known to have lived in marine environments, and their cysts are extremely resistant and can be preserved in ancient sediments, where they are of value in paleoecological studies (Fensome et al. 1999; Lee 2008). Alternatively, some of the living species are famous for causing "red tides," which can cause severe damage to fisheries, including aquaculture industries, and can also cause fish and shellfish poisoning. However, in comparison to marine species, the ecological and biological significance of freshwater species are poorly understood.

Of the living species, half have plastids and perform photosynthesis, although many of them also require other nutrients. Typical dinoflagellate plastids are derived from a red alga; however, several species have plastids derived from other algae, such as cryptomonads, haptophytes, diatoms, and green algae. The members that comprise other half of the dinoflagellate species have no plastid and are heterotrophic.

In addition, the morphology and lifestyles of dinoflagellates are known to be diverse (Taylor et al. 2007); some lineages are known to have evolved as parasites or mutualistic endosymbionts (zooxanthellae).

1.2.2. Phylogenetic relationship of dinoflagellates

Dinoflagellates are included in the infrakingdom Alveolata, along

with the apicomplexans and ciliates, based on an ultrastructural characteristics, molecular phylogenetic studies, and the presence of membranous sacs called alveolae (or amphiesmal vesicles in dinoflagellates) just beneath the cell membrane (Cavalier-Smith 1993, Gajadhar et al. 1991, Gould et al. 2008b, Wolters 1991). However, the phylogenetic relationships among orders within the dinoflagellates has remained unsettled due to the lack of statistical support for the phylogenetic backbone (Saldarriaga et al. 2004; Zhang et al. 2007), especially when relying on rapidly evolving plastid DNA (Zhang et al. 2000; Bachvaroff et al. 2006).

A few studies have provided morphological and ecological evidence that support phylogenetic relationships within the dinoflagellates (Moestrup and Daugbjerg 2007; Orr et al. 2012; Ramiro et al. 2007). Orr et al. (2012), for example, suggested that the thecate dinoflagellates originated from a single common origin, while Ramiro et al. (2007) noticed that freshwater species clustered into several monophyletic groups and that most freshwater species are not closely related to marine species. Thus, it became apparent that the marine-freshwater boundary has acted as a barrier to dinoflagellate diversification (Logares et al. 2007).

1.2.3. Plastid diversity of dinoflagellates

Most photosynthetic dinoflagellates possess peridinin-type plastids that originated from a red alga. Peridinin is a photosynthetic pigment unique to the dinoflagellates; however, a few dinoflagellate species have different types of plastids that seem to have originated from cryptomonads

(Schnepf & Elbra¨chter 1984), haptophytes (Tengs et al. 2000), diatoms (Dodge & Crawford 1969; Chesnick et al. 1997), and green algae (Watanabe et al. 1990). Those plastids are thought to have replaced the peridinin type plastid.

The half dinoflagellates species have no plastid and are heterotrophic. It is also believed that these dinoflagellates possessed peridinin-type plastids originally and lost the plastids (Puerta et al. 2006).

1.3. Minicircle DNA of dinoflagellates

In the peridinin-containing dinoflagellates examined so far, some plastid proteins and rRNAs are encoded on small circular DNA molecules, called minicircle DNAs (Zhang et al. 1999). The first minicircle DNA molecules were reported from a peridinin-containing dinoflagellate, *Heterocapsa triquetra* (Zhang et al. 1999). Since then, minicircle DNAs have been reported in several dinoflagellate genera, including *Heterocapsa* (Zhang et al. 2002), *Amphidinium* (Barbook & Howe 2000; Hiller 2001; Zhang et al. 2002), *Protoceratium* (Zhang et al. 2002), *Symbiodinium* (Moore et al. 2003), *Ceratium* (Laatsch et al. 2004), *Adenoides* (Nelson & Green 2005), and *Alexandrium* (Iida et al. 2009, 2010). Furthermore, no conventional plastid DNA has been reported from these dinoflagellates, despite the observation that a peridinin-containing species, *Lingulodinium polyedrum* (= *Gonyaulax polyedra*), seems to carry approximately 50–150 kb of circular DNA molecules that encode plastid genes (Wang & Morse 2006).

In contrast to conventional plastid DNA, dinoflagellate minicircle

DNAs are 2–6 kb circular DNA molecules, each of which is composed of unique non-coding region(s) and coding region(s) that encode single genes (or occasionally two or three genes) usually found in plastid DNA (Zhang et al. 1999; Barbrook & Howe 2000). So far, only about a dozen genes have been identified in minicircle DNA. Of these, all of can be found in the DNA of normal photosynthetic plastids (Bachvaroff et al. 2004) and encode for various proteins, including subunits of the two photosystems, the cytochrome b6f complex, the ATP synthase complex, rRNAs, and a small number of tRNAs (Howe et al. 2008; Mungpakdee et al. 2014). The most well-studied genes are *psbA* and the 23S rRNA gene (Howe et al. 2008). However, many other plastid genes have been shown to be encoded by the cell nuclear genome (Howe et al. 2008; Green 2011). Thus, peridinin-type plastids are thought to be particularly dependent on the expression of nuclear genes for their function (Dorrell & Howe 2015).

The non-coding regions of minicircle DNAs are also known to contain one or more core sequences, which are almost identical among the same-gene minicircle DNAs within species or cultured strains, but are almost entirely different between them (Zhang et al. 1999, 2002; Barbook & Howe 2000; Barbrook et al. 2001; Hiller 2001; Nelson & Green 2005). Thus, the core regions have been speculated to function as non-canonical promoters (Zhang et al. 2002; Moore et al. 2003; Howe et al. 2008; Nisbet et al. 2008) or replication origins (Zhang et al. 2002) or to anchor minicircle DNAs to the plastid membranes (Howe et al. 2008).

1.4. Purpose of this study

In a review of dinoflagellate minicircle DNAs, Howe et al. (2008) stated that the primary outstanding questions were "to settle the location of the minicircle DNAs, and to determine whether there is indeed an additional DNA species in the dinoflagellate chloroplast." It is, indeed, very important to understand the precise subcellular localization of minicircleDNAs to know how they evolved and how the genetic system for minicircle DNA functions in dinoflagellate cells.

Accordingly, the primary objective of this study was to directly demonstrate the subcellular localization of minicircle DNA. To confirm the result in multiple species, I selected two distantly related species of peridinin-containing dinoflagellates: the marine *Amphidinium massartii* and the freshwater *Hemidinium nasutum*. Then, using fluorescence *in situ* hybridization (FISH), I was able to demonstrate that the minicircle DNAs were indeed present in the plastids.

During the course of the above investigation, I also observed unexpected sequence variation among copies of *psbA*-coding minicircle DNAs in *H. nasutum*. Because this curious observation could provide clues toward understanding the maintenance and evolutionary mechanisms of minicircle DNA sequences, the second objective of this study was to determine the extent of sequence variation in the minicircle DNA of *H. nasutum* and to obtain further insights regarding the diversity of minicircle DNA. Accordingly, I sampled many copies of *psbA* and the 23S rRNA gene in the minicircle DNA of *H. nasutum* to assess the variation in their sequences.

2. Intracellular Localization of Minicircle DNA in Dinoflagellates

2.1. Introduction

The minicircle DNAs of peridinin-containing dinoflagellates are plasmid-type DNAs that encode plastid genes but are structurally different from conventional known plastid DNA. Although minicircle DNAs have been sequenced from several peridinin-containing dinoflagellates, the subcellular localization of the molecules is still controversial, as no direct evidence for plastid localization of minicircle DNAs has been provided.

Several lines of indirect evidence have implied that minicircle DNAs are present in plastids. Sequence analyses, for example, have failed to recover targeting sequences, which would be necessary to deliver the gene products to the plastids if the minicircle DNAs were localized elsewhere (Howe et al. 2008). In addition, the mRNAs of *psbA* were detected in plastids of a dinoflagellate, *Symbiodinium* sp., using in situ hybridization (Takishita et al. 2003), and in *Amphidinium carterae* and *L. polyedrum*, PsbA protein synthesis was shown to be inhibited by chloramphenicol, which is known to block protein synthesis in the plastids of other organisms (Wang et al. 2005). On the other hand, Laatsch et al. (2004) suggested that minicircle DNAs are localized outside the plastids of *Ceratium horridum* on the basis of cell fractionation, Southern blotting, and *in situ* hybridization.

Because the subcellular localization of minicircle DNAs is

fundamental for understanding the origin and evolution of minicircle DNA and for studying the function and maintenance of minicircle DNAs, this thesis chapter is focused on determining the precise subcellular localization of dinoflagellate minicircle DNA. To accomplish this, I selected two suitable species of dinoflagellates, *Amphidinium massartii* and *Hemidinium nasutum*, which are distantly related, and performed DNA-targeted whole-cell fluorescence *in situ* hybridization (FISH) with specific probes.

2.2. Materials

Two dinoflagellate species, *Amphidinium massartii* and *Hemidinium nasutum*, were chosen because they were considered to be suitable for investigating the localization of minicircle DNA (see species descriptions below). Selection of these species also seemed reasonable, based on a variety of dinoflagellate phylogenetic trees that have divided the photosynthetic dinoflagellates into two main clades: the *Amphidinium* clade and a second clade, which includes the other genera (Moestrup and Daugbjerg, 2007; Orr et al. 2012). Therefore, I selected one species from each group.

2.2.1. *Amphidinium massartii*

Members of the marine genus *Amphidinium* are suitable for cell staining and *in situ* hybridization because their athecate cells allow buffers, dyes, and probes to permeate easily into their cells, and their flat cell shape allows the cells to attach to glass slides with a greater amount of surface area. Additionally, *Amphidinium* species are easily grown in culture and exhibit

high growth rates.

In terms of minicircle DNAs, *Amphidinium carterae* is one of the most studied peridinin-containing dinoflagellates, and many minicircle DNA sequences have been reported from this species (Howe et al. 2008). However, the species also has a highly reticulate plastid (Murray et al. 2004), which would inhibit the determination of minicircle DNA localization under a fluorescent microscope. Therefore, the present study uses *Amphidinium massartii*, which is the closest relative to *A. carterae* and contains a large central plastid with radially-arranged, thick lobes (Murray et al. 2004) that should aid in microscopic observation.

In addition, two complete minicircle DNA sequences have been sequenced from *A. massartii* (TM16): one encodes *psbA* and the other encodes 23S rRNA gene (Hayashi 2005). The *psbA* minicircle DNA was determined to be 3569 bp long, with a coding sequence of 1023 bp, and the 23S rRNA gene minicircle DNA was determined to 3922 bp long, although the exact region of the gene could not be determined, much like the 23S rRNA in the minicircle DNAs of other dinoflagellates. A 198-bp core region was also identified by comparing the non-coding regions of the two minicircle DNA sequences and was determined to be identical between them (Hayashi 2005).

2.2.2. *Hemidinium nasutum*

I chose *Hemidinium nasutum* as a representative of the non-*Amphidinium* clade. *H. nasutum* is also suitable for cell staining for the

same reasons as *A. massartii*. The cells of *H. nasutum* are also big (20–30 μm), athecate, and flat, and the thick lobes of the plastids are arranged parallel to the longitudinal axis, which was assumed would facilitate microscopic observation. In addition, *H. nasutum* lives in fresh water environments, and the minicircle DNAs of fresh water species have not been studied. In a preliminary experiment, I found that *H. nasutum* contained peridinin as a photosynthetic pigment (data not shown).

2.3. Methods

2.3.1. Algal cultures

A culture of the dinoflagellate *Amphidinium massartii* (TM16), which was originally established from a sample collected at Shirahama Beach (Wakayama prefecture, Japan) by Dr. M. Tamura (currently at Okinawa Institute of Science and Technology Graduate University, Japan), was kindly provided by Dr. T. Horiguchi (Hokkaido University, Japan). The alga was cultured in Daigo IMK medium (Wako Pure Chemical, Osaka, Japan) under 50 $\mu mol \cdot m^{-2} \cdot s^{-1}$ with a 14 h light/10 h dark cycle.

A cultured strain of the dinoflagellate *Hemidinium nasutum* (NIES471) was obtained from the National Institute for Environmental Studies, Japan. The alga was cultured in AF6/2 medium (AF6 medium [Kato 1982] diluted with distilled water to one-half) under 50 $\mu mol \cdot m^{-2} \cdot s^{-1}$ with a 14 h light/10 h dark cycle.

2.3.2. DNA extraction

A well-grown culture that was near the stationary phase was used for DNA extraction because the minicircle DNA copy number was expected to be high (Koumandou & Howe 2007). DNA extraction was performed according to the standard phenol extraction method of Koumandou and Howe (2007), with slight modifications. Briefly, cells were collected using centrifugation at 2500 × *g* for 5 min at 4 °C and then resuspended in TEN buffer (20 mM Tris-HCl, 0.5 mM EDTA, 0.1 M NaCl, pH 8) with 1% (w/v) sodium dodecyl sulfate and 500 $\mu g \cdot mL^{-1}$ Proteinase K. The mixture was incubated for 1 h at 50 °C. Then the DNA was extracted using the phenol extraction method, precipitated using ethanol, and finally, dissolved in distilled water.

2.3.3. Sequence analysis for *H. nasutum* minicircle DNAs

To obtain sequences of *psbA* and the 23S rRNA gene in *H. nasutum*, I used degenerate primers (Table 2.1) that were designed on the basis of dinoflagellate *psbA* and 23S rRNA gene sequences obtained from the NCBI public database. Polymerase chain reaction (PCR) amplification was performed using Ex taq polymerase (Takara Bio, Shiga, Japan) with the following conditions: 94 °C for 2 min; followed by 35 cycles of 94 °C for 30 s, 43 °C for 30 s, and 72 °C for 3 min; and a final extension at 72 °C for 7 min. Each amplified product was cloned into the pGEM T-easy Vector (Promega, Madison, WI, US) and sequenced using a Genetic Analyzer 3100 (Applied Biosystems, Foster City, CA, USA) and a Big Dye Terminator v3.1 Cycle

Sequencing Kit (Applied Biosystems).

To obtain full-length minicircle DNA sequences, outward-facing *psbA* and 23S rRNA gene primers (Table 2.1) were constructed on the basis of the obtained sequences (see above). PCR amplification was performed as described above with each annealing temperature shown in Table 2.1. Sequencing was performed as described above.

2.3.4. Southern blotting to determine the structure of *H. nasutum* minicircle DNAs

Genomic DNA of *H. nasutum* was electrophoresed on a 1% agarose gel in TAE buffer and neutral transferred to a Hybond-N+ nylon membrane (GE Healthcare GE Healthcare Bioscience, Piscataway, NJ, USA) by capillary blotting, according to the manufacturer's instructions.

Probes for the nuclear-encoded 28S rRNA gene, minicircle-encoded *psbA,* and the minicircle-encoded 23S rRNA gene were prepared using PCR from the plasmid clone with specific primers for each gene and subsequent gel purification (Table 2.1). Alk Phos Direct (GE healthcare, Piscataway, NJ, USA) was used for probe labeling and hybridization, and an LAS-1000 imaging system (Fujifilm, Tokyo, Japan) was used for fluorescence signal detection, according to the manufacturer's instructions.

2.3.5. Southern hybridization to confirm the specificity of probes for *A. massartii* DNA-targeted whole-cell FISH

Digoxigenin-labeled DNA probes for the nuclear-encoded 28S rRNA

gene, minicircle-encoded *psbA* and the core sequence were prepared from the plasmid clones using DNA-specific primers (Table 2.1) and a PCR DIG Probe Synthesis Kit (Roche Diagnostics, Mannheim, Germany), according to the manufacturer's instructions. For a negative control, human tissue-type plasminogen activator (*tPA*) probe was amplified using the control plasmid and primers from the kit.

For southern blotting, genomic DNA of *A. massartii* was electrophoresed on a 1% agarose gel in TAE buffer and neutral transferred to a Hybond-N+ nylon membrane (GE Healthcare) by capillary blotting, according to the manufacturer's instructions. Each of the probes was then hybridized and detected using a DIG Nucleic Acid Detection Kit (Roche Diagnostics), according to the manufacturer's instructions.

2.3.6. DNA-targeted whole-cell FISH of *A. massartii* and *H. nasutum*

Algal cells were collected using low-speed centrifugation (500 × *g*, 20 °C for 5 min) and fixed with 3% paraformaldehyde in PHEM-NaCl buffer (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 2 mM MgCl2, pH 6.9, 3% NaCl) for 15 min on ice. To dehydrate the cells and diminish chlorophyll autofluorescence in moderation, the cells were subject to an ethanol series of 35%, 70%, 85%, and 85% for 5 min each on ice, after which the cells were rehydrated using an ethanol series of 70% and 35% for 5 min each on ice and twice in 2× SSC (300 mM NaCl, 30 mM trisodium citrate, pH 7.0) for 5 min at room temperature. After rehydration, the cells were treated with 100 $\mu g \cdot mL^{-1}$ RNase A (Roche Diagnostics), settled onto coverslips coated with

0.1% (w/v) poly-L-lysine solution (Wako Pure Chemical), and incubated for 2 h at 37 °C in a moist chamber. The cells were permeated 3 times with PBS-T (0.2% [v/v] Tween-20 in PBS [13 mM NaCl, 7 mM Na_2HPO4, 3 mM NaH_2PO4, pH 7.2]) for 5 min at room temperature and overnight at 4 °C. Then the cells were washed 3 times in 2× SSC at room temperature.

For hybridization, the cells were incubated in a hybridization buffer (50% deionized formamide, 2× SSC, 50 mM sodium phosphate; pH 7) for 1.5 h at 37 °C in a moist chamber. After removing the buffer, fresh hybridization buffer that contained 10% dextran sulfate and denatured DIG-labeled DNA probe (1% [v/v], prepared as described above) was applied to the cells on each coverslip. Next, the coverslips on a glass slide were each covered with another coverslip to prevent evaporation the probe, incubated for 7 min at 85 °C, and then incubated overnight at 37 °C for hybridization. Hybridized cells were washed 3 times in 2× SSC containing 50% formamide for 5 min at 45 °C and then 3 times in 2× SSC for 5 min at room temperature to remove unhybridized DNA probe.

For detection, the cells were treated with blocking buffer (Roche Diagnostics) for 1 h at room temperature and then incubated with an anti-Digoxigenin antibody produced in a mouse (11333062910, 1:125 dilution in blocking buffer; Roche Diagnostics, Mannheim, Germany) for 1 h at 37 °C in a moist chamber. The cells were then washed with PBS-T, treated with an anti-mouse IgG antibody conjugated with fluorescein isothiocyanate (F0257, 1:50 dilution in blocking buffer; SIGMA Aldrich, St. Louis, MO, USA), and washed again with PBS-T. The immunostained cells were mounted with

Slow-Fade anti-fade reagent (Molecular Probes, Eugene, OR, USA) that contained 0.5 μg•mL^{-1} 4′,6-diamidino-2-phenylindole dihydrochloride (DAPI). The samples were observed under a DMRD epifluorescence microscope (Leica Microsystem, Wetzlar, Germany) equipped with a DP71 CCD camera (OLYMPUS, Tokyo, Japan). Images for the samples treated with the minicircle-encoded 23S rRNA gene, minicircle DNA *psbA*, minicircle DNA core sequence, and *tPA* probes were taken with the same exposure conditions (ISO 800), whereas those treated with the nuclear-encoded 28S rRNA gene probe were captured with a less sensitive exposure condition (ISO 400).

To identify the localization in detail, the samples of *A. massartii* were also observed under a LSM 710 confocal laser-scanning microscope (Carl Zeiss, Jena, Germany) using the Zen software.

2.4. Results

2.4.1. Minicircle DNA sequences and minicircle DNA size in *H. nasutum*

I performed PCR using degenerate primers for *psbA* and the 23S rRNA gene and sequenced the coding regions of each gene from five clones of each. Each cloned sequence was unique, with different base substitutions and/or insertions/deletions (*psbA*: Figure 3.1, #01-04, 09; 23S rRNA gene: Figure 3.2, #01-05).

If a gene is encoded within a minicircle DNA molecule, non-coding regions can be acquired using outward-facing primer sets. Hayashi (2005), for example, designed a primer set that was named “the outward primer set

for non-coding regions." One of the primers, Dino*pabAF*2o, was designed to bind 100 bp downstream of Hayashi's partial *psbA* sequences, and another, Dino*pabA*R2o, was designed to bind 109 bp upstream of the partial *psbA* sequences. Similarly, I designed the primers HN23sF1o and HN23sR1o to bind at locations 76 bp downstream and 84 bp downstream of the 23S rRNA partial gene sequences, respectively.

Using outward-facing primers, I was also able to amplify and sequence several non-coding regions, suggesting that the genes were encoded in minicircle DNA molecules. The size of the *psbA* minicircle DNA was about 2000 bp, and the size of the 23SrRNA minicircle DNA was about 2500 bp. In addition, I obtained two sequences for *psbA* and the 23S rRNA gene, each, and found that the sequences were not identical (*psbA*: Figure 3.3, #61pA, 62pA; 23S rRNA gene: Figure 3.3, #64rR, 65rR). There was also sequence variation in the areas where the sequences from the coding and non-coding regions overlapped However, the results suggested that the plastid genes of *H. nasutum* were encoded in DNA molecules with minicircle DNA structures, as has been observed in other peridinin-containing dinoflagellates.

I also conducted rough estimates of the sizes of the *psbA* and 23S rRNA gene minicircle DNAs using genomic Southern blotting analyses with probes for partial *psbA* and 23S rRNA gene sequences, of which the lengths were 404 bp and 374 bp, respectively. For a positive control, I used a nuclear 28S rRNA gene probe that gave a signal for the broad mass of nuclear genomic DNA observed at high molecular weight (Figure 2.1, lane 3). The *psbA* probe showed a strong broad band that ranged from 1500 bp to 2800 bp

(Figure 2.1, lane 1), and the 23S rRNA gene probe showed two broad bands that ranged from 1900 bp to 2100 bp and from 3000 bp to 3600 bp (Figure 2.1, lane 2). No high molecular weight signal was detected for either gene, and genomic Southern blotting analysis also suggested that the *psbA* and the 23S rRNA gene were encoded in minicircle DNAs. The broad bands are probably indicative of a mixture of linear monomers and relaxed monomeric circles, and many other types of variation are also possible in the same-gene minicircle DNAs. In chapter 3, I analyzed the sequence variation of *H. nasutum* minicircle DNA in detail.

2.4.2. Localization of *Amphidinium massartii* minicircle DNAs by FISH

To determine the localization of minicircle DNAs in the cell, I chose 932 bp of *psbA* and 169 bp of the minicircle DNA core region sequence of *A. massartii* to use as specific probes. I also constructed a nuclear-encoded 28S rRNA gene-specific probe (700 bp) to use as a positive control and a *tPA* human gene-specific probe (169 bp) to use as a negative control, and the specificity of each probe was confirmed using total genome Southern blotting (Figure 2.2).

When tested, the nuclear 28S rRNA gene probe gave a strong signal for the broad mass of nuclear genomic DNA at high molecular weight (Figure 2.2, lane 2). The *psbA* probe showed a weak band at about 3.5 kb and a strong band at about 5 kb (Figure 2.2, lane 3), which probably correspond to linear monomers and relaxed monomeric circles, respectively. The other minicircle-specific probe, the core region probe, gave a weak band at about 3

kb and several strong bands at about 4 kb, 5 kb, and 6 kb (Figure 2.4, lane 4). Thus, the probes constructed to detect the nuclear-encoded 28S rRNA gene, minicircle-encoded *psbA*, and minicircle DNA core region were sufficient to identify specific bands able to detect specific loci in *A. massartii*. In contrast, no signal was detected from the *tPA* probe, indicating that it was suitable to use as a negative control (lane 1 in Figure 2.2).

Using these probes, I employed FISH analysis on *A. massartii* cells. The *tPA* probe (negative control) was unable to detect any signal (Figure 2.3a), whereas the nuclear 28S rRNA gene probe (positive control) detected a signal, which appeared as a bright spot near the edge of the nucleus (Figure 2.3b).

Epifluorescence microscopy revealed that the probe for *psbA* (Figure 2.3c) was localized in spots that were scattered within the region of chlorophyll autofluorescence. The probe for the core region produced similar spotted signals, as well as clusters of signals with variable size and intensity throughout the region of chlorophyll autofluorescence (Figure 2.3d). None of the signals overlapped with DAPI staining, which was used to stain the nucleus, although a few spots of minicircle DNA signal were also observed in the area of pyrenoid, which is a dark area with no chlorophyll autofluorescence near the center of the cell (Figure 2.3c, d).

FISH observations using a confocal laser-scanning microscope enabled us to identify the localization of minicircle DNA in further detail. The core probe produced spotted signals that were actually located along the outer edge of the autofluorescence, between thylakoids and the cytoplasm or

pyrenoids (Figure 2.4), indicating that minicircle DNA was present at the periphery of the plastid. Similar results were obtained with the *psbA* probe (data not shown).

2.4.3. Localization of *Hemidinium nasutum* minicircle DNAs by FISH

To determine the localization of minicircle DNAs in the cells of *H. nasutum*, I chose 404 bp of the *H. nasutum psbA* sequence to use as a specific probe. I also constructed a nuclear-encoded 28S rRNA gene-specific probe (314 bp) as a positive control. Using these probes, I employed FISH analysis on *H. nasutum* cells, using the *psbA* gene as a minicircle DNA probe, the 28S rRNA gene as a positive control, and the *tPA* human gene as a negative control.

The *tPA* probe (negative control) did not give any signal (Figure 2.5a), whereas the nuclear 28S rRNA gene probe (positive control) produced a signal that appeared as a bright spot near the edge of the nucleus (Figure 2.5b). Epifluorescence microscopy revealed that the *psbA* probe (Figure 2.5c) was localized along the outer edge of the region of chlorophyll autofluorescence, between thylakoids and the cytoplasm, again indicating that the minicircle DNA was present at the periphery of the plastid. None of the signals overlapped with the DAPI staining.

2.5. Discussion

2.5.1. Technique of DNA-targeted whole-cell FISH

Using FISH and minicircle DNA-specific probes produced distinct

spotted signals within the *A. massartii* plastid. These signals are not likely to have been nonspecific signals or artifacts, since the probes were highly specific, as was confirmed using total-genome Southern blotting. In addition, my *in situ* hybridization procedure worked well for *A. massartii* cells, as indicated by the positive and negative control probes (Figure 2.3a). In fact, the single dot signal that was observed at the periphery of the nucleus when using the positive control (the nuclear 28S rRNA gene probe) was the expected positive result, since the same 28S rRNA gene localization was also shown using FISH in another dinoflagellate, *Prorocentrum micans* (Geraud *et al.* 1991). Therefore, we concluded that the FISH results obtained using minicircle DNA-specific probes reliably indicated the localization of minicircle DNAs in *A. massartii* cells.

2.5.2. Localization of plastid minicircle DNAs

FISH observations with both epifluorescence and confocal laser-scanning microscopy indicated a spotted distribution of minicircle DNAs throughout the plastids. However, confocal laser-scanning microscopy further revealed that the spots of minicircle DNAs were actually present at the periphery of the plastids. This distribution pattern also indicates that the copies of minicircle DNAs are not evenly distributed, but rather form nucleoid-like aggregations in the plastids that may even be anchored to the plastid envelope.

The present study provides the first direct evidence for plastid localization of minicircle DNAs in two different lineages of

peridinin-containing dinoflagellates: *A. massartii* and *H. nasutum*. With several circumstantial lines of evidence suggesting that the plastid localization of minicircle DNAs occurs in wide-range of dinoflagellate lineages, the fragmentation of plastid DNA into minicircle DNAs would had to have occurred within the plastid in the peridinin-containing dinoflagellate lineage. However, we still cannot rule out the possibility that minicircle DNAs are present in the nucleus of *C. horridum* (Laatsch et al. 2004), since the species was not used in the present study.

2.5.3. Conclusions and future perspectives

The plastid localization of dinoflagellate minicircle DNAs was directly demonstrated in the present study for the first time, and the results of the study suggest that minicircle DNAs are likely to have evolved from a conventional plastid DNA. Accordingly, it is now reasonable to investigate minicircle DNA in the context of "molecular evolution of the plastid DNA.". Thus, the present study has opened a door for progress in understanding the evolution and cell biology of dinoflagellate plastids.

However, from the results of this study, it remains unclear how copies of minicircle DNA are packed into the aggregations or how the aggregations are anchored to the envelope membrane. In higher plant plastids, the DNA-binding PEND protein is known to be involved in anchoring plastid DNA to the plastid envelope (Sato et al. 1998; Sato & Ohta 2001). Thus, it would be interesting to investigate whether any DNA-binding protein like PEND is found in peridinin-type plastids and also whether the core region or

stem-loop structures of the non-coding sequences are involved.

3. Sequence Variation of the Dinoflagellate *Hemidinium nasutum* Minicircle DNA

3.1. Introduction

In previous reports on dinoflagellate minicircle DNAs, the sequences of minicircle DNA coding regions have usually been observed to be identical within cultured strains. For example, only one sequence has been reported for the *psbA* minicircle DNA of *Heterocapsa triquetra* (CCMP 449), *H. pygmaea* (CCMP 1490), *H. niei* (CCMP 447), *H. rotundata* (NEPCC D680), *P. reticulatum* (NEPCC D535), *Amphidinium carterae* (CCMP 1314), *A. operculatum* (CCAP 1102/6), *Adenoides eludens* (NEPCC D683), and *A. massartii* (TM16) (Zhang et al. 1999, 2002; Barbrook & Howe 2000; Hayashi 2005; Nelson & Green 2005). However, several sequence variations have also been reported.

In the 23S rRNA gene minicircle DNAs of *H. triquetra* and *Symbiodinium* species, for example, 25–300 bp deletions in coding regions have been reported (Zhang et al. 1999; Santos et al. 2002, 2003; Barbrook et al. 2006), and in the *psbA* and *psbD* minicircle DNAs of *Alexandrium* species, variant minicircle DNAs with ~100 bp insertions/deletions in coding regions have also been found (Iida et al. 2009, 2010). Furthermore, several "jumbled" minicircle DNAs, which encode more than two imperfect genes, (Zhang et al. 2001) and "empty" minicircle DNAs, which contain short (<300 bp) segments of genes or no gene segments, (Barbrook et al. 2001; Hiller 2001; Green 2004; Barbrook et al. 2006) have also been reported. These "aberrant" minicircle

DNAs, however, are usually minorities compared to “normal” minicircle DNAs (Howe et al. 2008).

In the previous chapter, the comparison of cloned partial minicircle DNA sequences revealed that *H. nasutum psbA* and 23S rRNA genes contained extensive insertions/deletions and base substitutions, suggesting that there may be huge sequence variation among the minicircle DNAs of the species. I surmised that, by sequencing addition clones, I would likely identify additional sequence variants of both *psbA* and 23S rRNA gene minicircle DNAs, and I also hoped to identify which sequence represented the “normal” minicircle DNA sequence for each gene. Therefore, the purpose of this study was to determine the extent of variation in the DNA sequences of the *psbA* and 23S rRNA gene minicircle DNAs in *H. nasutum*.

3.2. Material and Methods

3.2.1. Algal culture

A cultured strain of the dinoflagellate *Hemidinium nasutum* (NIES471) was obtained from the National Institute for Environmental Studies, Japan. The alga was cultured in AF6/2 medium (Kato 1982) under 50 $\mu mol \cdot m^{-2} \cdot s^{-1}$ with a 14 h light/10 h dark cycle.

3.2.2. DNA and RNA extraction

A well-grown culture that was near the stationary phase was used for DNA extraction, because the copy number of minicircle DNA was expected to be high (Koumandou & Howe 2007). DNA extraction was performed

according to the standard phenol extraction method of Koumandou and Howe (2007), with slight modifications. Briefly, cells were collected using centrifugation at 2500 × g for 5 min at 4 °C and then re-suspended in TEN buffer (20 mM Tris-HCl, 0.5 mM EDTA, 0.1 M NaCl, pH 8) with 1% (w/v) sodium dodecyl sulfate and 500 μg•mL−1 Proteinase K. The mixture was incubated for 1 h at 50 °C. Then the DNA was extracted using the phenol extraction method, precipitated using ethanol, and finally, dissolved in distilled water.

Total RNA was extracted using the RNeasy Plast Mini Kit (QIAGEN, Tokyo, Japan), according to the manufacturer's instructions. The obtained total RNA was used for reverse transcription (RT) PCR assays. Complementary DNA (cDNA) was amplified using a SuperScript III First-Strand System (Thermo Fisher Scientific Inc. Wilmington, DE, USA) with random primers, following the manufacturer's procedure.

3.2.3. Sequence analysis for *H. nasutum* minicircle DNAs

To obtain sequences of *psbA* and the 23S rRNA gene in *H. nasutum*, I used degenerate primers (Table 3.1) designed on the basis of dinoflagellate *psbA* and 23S rRNA gene sequences obtained from the NCBI public database. PCR amplification was performed using Ex taq polymerase (Takara Bio, Shiga, Japan) with the following conditions: 94 °C for 2 min; followed by 35 cycles of 94 °C for 30 s, 43 °C for 30 s, and 72°C for 3 min; and a final extension at 72 °C for 7 min. Each amplified product was cloned into the pGEM T-easy Vector (Promega, Madison, WI, US) and sequenced using a

Genetic Analyzer 3100 (Applied Biosystems, Foster City, CA, USA) and a Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). To obtain full-length minicircle DNA sequences, outward-facing primers were made on the basis of the obtained *psbA* and 23S rRNA gene sequences (Table 3.1). PCR amplification was performed as described above with each annealing temperature shown in Table 3.1. Sequencing was performed as described above.

The *psbA* cDNA fragments were amplified using PCR with the cDNA made as described above. PCR amplification was performed as described above with each annealing temperature shown in Table 3.1. Sequencing was performed as described above.

Alignments were performed using MEGA 6.0 (Tamura et al. 2013).

3.3. Results

3.3.1. DNA and cDNA sequences of the *psbA* coding region in *H. nasutum*

For the *psbA* coding region of *H. nasutum*, 55 and 11 partial sequences were amplified using genomic PCR and RT-PCR, respectively. Of the 55 sequences obtained using the genomic PCR, 16 and 7 sequences were obtained with an inward-facing degenerated primer set (Dino*psbA*F1i, Dino*psbA*R1i) (Table 3.1) and an outward-facing degenerated primer set (Dino*psbA*F2o, Dino*psbA*R2o) (Table 3.1), respectively, and were used to develop outward-facing primers. Based on the alignment of these 23 sequences, outward-facing primers, which included two forward primers (HN*psbA*F2, HN*psbA*F3) and one reverse primer (HN*psbA*R2) (Table 3.1),

were designed to amplify nearly whole lengths of the minicircle DNA molecules in a single PCR. Twelve of the 55 sequences were obtained using one outward-facing primer set (HN*psbA*F2, HN*psbA*R2) (Table 3.1), although two of the sequences (#20 and #21) were truncated because only a single strand was sequenced, and the other 20 sequences were obtained using the second specific primer set (HN*psbA*F3, HN*psbA*R2) (Table 3.1), although sequence #40 was truncated because only a single strand was sequenced.

In the 55 *psbA* coding sequences, no identical sequence was found; each sequence had at least one unique base substitution or insertion/deletion (Figure 3.1). In addition, only two of the 55 sequences (#9 and #18, 3.6%) appeared to encode functional amino acid sequences, and both of those were only partial sequences. All other sequences contained at least one frameshift mutation.

In contrast, only 11 cDNA sequences were obtained using RT-PCR with a *psbA* primer set (HN*psbA*F4, HN*psbA*R3) (Table 3.1). Three of the 11 sequences (#81, #88, and #90) were identical, whereas the other 8 sequences were unique. Of the 11 sequences, four (#81, #82, #88, and #90, 36.4%) appeared to encode the functional PsbA amino acid sequences, and the others had at least one frameshift mutation (Figure 3.1).

3.3.2. Sequence variation in minicircle DNA *psbA* coding regions

Within an alignment of 66 DNA and cDNA sequences, many base substitutions and insertions/deletions were observed (Figure 3.1). The

mutations were not evenly distributed, and 15 variable regions were observed to contain more mutations than the other regions. Other mutations that were observed outside of these variable regions seemed to be distributed randomly in the alignment (Fig 3.1). In the alignment, sequences that share the same variation state were different among sites (Fig. 3.1), suggesting that the distribution of variation exhibits a mosaic pattern among the copies of same-gene minicircle DNAs *H. nasutum*.

3.3.3. DNA and cDNA sequences of the 23S rRNA coding region in *H. nasutum*

For the 23S rRNA gene coding region, 17 and 5 partial sequences were amplified using genomic PCR and RT-PCR, respectively. Of the 17 sequences obtained using genomic PCR, 10 and 7 sequences were obtained with an inward-facing degenerated primer set (Dino23SF1, Dino23SR1i) (Table 3.1) and an outward-facing specific primer set (HN23SF1o, HN23SR1o) (Table 3.1), respectively, and, the five cDNA sequences were obtained using RT-PCR and an inward-facing specific primer set (HN23SF2, HN23SR2) (Table 3.1). In addition, among 17 DNA sequences and 5 cDNA sequences, no identical sequences were found; each sequence had at least one unique base substitution or insertion/deletion (Figure 3.2).

3.3.4. Sequence variation in minicircle DNA 23S rRNA gene coding regions

Within an alignment of the 22 DNA and cDNA sequences, many base substitutions and insertions/deletions were observed, and no identical sequences were found, as was observed for the *psbA* variants. However, in

contrast to the *psbA* alignment, variations in the 23S rRNA gene alignment appeared to be distributed randomly. In addition, 11 variable regions were identified, and In the sequence alignment, sequences that share the same variation state were different among sites, as was observed in the variation of *psbA* coding region sequences.

3.3.5. Sequences of the non-coding regions of *H. nasutum* minicircle DNAs

For the non-coding regions of *H. nasutum* minicircle DNAs, 8 sequences were amplified using genomic PCR. Of the 8 sequences, 2 sequences (Figure 3.3, pA #51 and pA #54) were obtained with an outward-facing degenerated primer set of *psbA* (Dino*psbA*F2o, Dino*psbA*R2o) (Table 3.1). One sequence (Figure 3.3 pA#18+) was obtained with the specific primer set of *psbA* (HN*psbA*F2, HN*psbA*R2) (Table 3.1), which was expected to encode a functional protein, and the remaining sequences (Figure 3.3, rR, #01, 02, 03, 05, and 07) were obtained with an outward-facing specific primer set of the 23S rRNA gene (HN23SF1o, HN23SR1o) (Table 3.1).

The sequences of non-coding regions from both *psbA* and 23S rRNA gene minicircle DNAs were all unique, in agreement with results for the *psbA* and 23S rRNA coding regions (Figure 3.2). Total lengths varied from 1656 bp to 2186 bp, and a shared core sequence of approximately 150 bp was identified. Interestingly, the non-coding regions other than the core sequence (non-core non-coding regions) exhibited sequence variation that was different from those seen in the coding region and the core sequence. Almost

the entire sequence of the non-core non-coding regions consisted of ~15 bp repeats. Sequence variations were also observed in the core regions, and the core sequence also exhibited a stem-loop structure of 128 bp.

3.4. Discussion

3.4.1. *psbA* coding sequences of *H. nasutum* minicircle DNA are highly variable.

Minicircle DNAs have been reported for 16 dinoflagellate species (Zhang et al. 1999, 2002; Barbrook & Howe 2000; Moore et al. 2003; Laatsch et al. 2004; Hayashi 2005; Nelson & Green 2005; Barbrook et al. 2006b). In many of these species, only a single sequence of minicircle DNA for each gene has been reported, and no sequence variation has been observed (Howe et al. 2008). There are a few reports of so-called “aberrant” minicircle DNAs, which encode no gene or only a small fragment of a gene (Zhang et al. 1999, 2001; Barbrook et al. 2001; Hiller 2001; Green 2004; Nisbet et al. 2004; Barbrook et al. 2006), and several sequence variations have also been reported in the dinoflagellate *Alexandrium tamarense* (Iida et al. 2009; Iida et al. 2010).

In one strain of *A. tamarense* (OFAT0105-8), about half of the *psbA* minicircle DNAs had abnormal sequences, in which a fragment of *psbA* was missing from the coding region but was present in the non-coding region, probably without losing the genetic information for *psbA* (Iida et al. 2009). In addition, Iida et al. (2009) obtained 35 complete *psbA* sequences from *A. tamarense* and found that the sequences could be classified as either normal *psbA* sequences or one of four variant type sequences that had frameshifts

mutations, whereas sequences of the non-coding region were categorized into 8 types. However, in *A. tamarense*, the percentage of the variant minicircle DNAs were low (48%), when compared with the percentage of "normal" minicircle DNAs (52%) (Iida et al. 2009), as has been reported in other species (Howe et al. 2008).

In contrast, the proportion of variant minicircle DNA sequences in *H. nasutum* was surprisingly high, with at least 96% of the 55 *psbA* sequences encoding non-functional proteins. This high percentage of sequence variation suggests that a high level of misreplication occurs during the minicircle DNA replication of *H. nasutum* and that the restoration system of minicircle DNAs might be weak or even absent. Thus, even though the proportion of functional genes appears to be very low, it is a mystery how *H. nasutum* maintains any functional plastid genes at all.

3.4.2. Difference between DNA and cDNA in the proportion of possible functional *psbA* sequence

This study showed that the minicircle DNA sequences of *H. nasutum* were highly variable and that the majority of the minicircle DNA coding sequences contained frameshift mutations. It was, therefore, difficult to determine which DNA and cDNA sequences represented the "normal" minicircle DNA sequence. The proportion of *psbA* DNA sequences with frameshift mutations was higher than those of *psbA* cDNA sequences, of which 4 of 11 sequences (36%) encoded functional proteins. In addition, 3 of the 4 functional cDNA sequences were identical, implying that not all of the

minicircle DNA *psbA* genes were transcribed or that some of the transcripts were corrected at the mRNA level. Since the same large deletions, frameshifts, and variation patterns were found in both DNA and cDNA sequences, it is unlikely that the abnormal genes were omitted from transcription, and it appears that there is no mechanism for selecting "normal" genes at the transcription level.

One possible explanation for the difference in the proportion of "functional" sequences the DNA and cDNA is that because of the high mutation rate, the regulatory mechanisms of sequence transcription are also likely to malfunction in some minicircle DNA copies and the proportion of abnormal sequence transcription is reduced. Although mutation occurs equally in the DNA of all minicircle DNAs, if the proportion of abnormal genes was very high, like in the case of *psbA* in *H. nasutum*, abnormal genes would have a greater probability of acquiring mutations in their regulatory regions.

Another possibility is the presence of RNA editing. In a previous study on the variant minicircle DNAs of *A. tamarense*, a possible case of RNA editing was reported (Iida et al. 2009). The results of the present study cannot rule out the possibility that an RNA editing system also restores the genetic information of minicircle DNAs in *H. nasutum*, but there is also no concrete evidence.

3.4.3 Recombination may play a role in creating sequence variation

It has been reported that "aberrant" minicircle DNAs are created

from minicircle DNAs with normal genes *via* replication slippage (Zhang et al. 2001; Santos et al. 2002; Iida et al. 2009, 2010) and recombination (Zhang et al. 1999, 2001; Barbrook et al. 2001; Hiller 2001; Nisbet et al. 2004; Iida et al. 2009, 2010). In *H. nasutum*, the distribution of sequence variation in the alignment appears to be mosaic, in that one sequence mutation (base substitution, insertion/deletion, etc.) in a sequence is shared by several minicircle DNA copies, whereas another sequence mutation in the same sequence is exclusively shared by different copies. I presume that this situation could result from frequent recombination.

In the alignment of *H. nasutum psbA* coding region sequences, I identified 15 variable regions, in which many mutations were accumulated in short site reagion. Accordingly, I wondered whether these variable regions corresponded to regions that were less important for the function of the PsbA protein. The PsbA protein is structurally well conserved within cyanobacteria and all photosynthetic eukaryotes (Golden 2004), especially in regards to hydropathy profiles (Iida et al. 2008). In my preliminary comparison between the 15 variable regions in the *psbA* sequences and the hydrophobic regions of the PsbA protein, no specific relationship was found. Therefore, it is possible that no functional constraint is involved in creating those variable regions in the *psbA* gene sequences of *H. nasutum* and that those variable regions may just be the vicinity of recombination borders. However, further investigation is needed to determine why those variable regions are present.

All of the 23S rRNA gene coding sequences amplified from DNA and cDNA were unique, and I could not discern which sequence was the functional version, since the gene is not a protein-coding gene. However, as in the *psbA* coding sequences, I also identified several variable regions in the 23S rRNA gene sequences. For a ribosomal RNA, secondary structure (base pairs, double-helices, loop, bulges, and single strands) is important for function (Petrov et al. 2013). By comparing the sequences with an alignment of *Escherichia coli* 23S rRNA sequences (Petrov et al. 2013), I found that the variable regions of *H. nasutum* 23S rRNA gene sequences seemed to have no relation to the secondary structure.

3.4.5. Sequence variations in the DNA non-coding region of minicircle DNA

Like the coding region of *psbA* minicircle DNA, the three non-coding region sequences of *psbA* minicircle DNAs and five non-coding region sequences of 23S rRNA gene minicircle DNAs were all unique. Nucleotide substitutions and short insertions/deletions are also found in the "core sequence," which is usually identical among the minicircle DNAs of individual strains (Zhang et al. 1999, 2002; Barbrook & Howe 2000; Hayashi 2005; Nelson & Green 2005). The high variability of minicircle DNA sequences in *H. nasutum* is, therefore, not restricted to coding regions, but is characteristic of the whole minicircle DNA molecule.

In addition, the results revealed that the non-coding regions of *H. nasutum* minicircle DNA contained many short tandem repeat sequences, which could be grouped into three types (TCGGTGCTGTTTCAT, TTTCAGGTTTG,

and GGCATATTCTTATT), and that the repeats were present throughout the whole non-coding region, except for the "core" sequences. The presence of the repeat sequences in the non-coding regions of minicircle DNA has also been reported from two dinoflagellate species: *A. tamarense* (Iida et al. 2009) and *Symbiodinium* sp. (Moore et al. 2003). However, those repeat sequences were different from those in *H. nasutum*. In *A. tamarense*, for example, only a few large (~400 bp) tandem repeats were reported in the "non-core non-coding" region, and in *Symbiodinium* sp., many inverted repeats were found in "non-core non-coding" region. It is strange that the minicircle DNA non-coding sequences of different species exhibit different modes of repeat sequence variation, such as many short tandem repeats, a few long tandem repeats, and many short inverted repeats. This makes me wonder whether the mechanisms involved in the evolution of "non-core non-coding" regions of minicircle DNA is different among species. In addition, there was no short tandem repeat observed in the "non-core non-coding" region in *H. nasutum* like in those of other species, which further supports the idea that different modes of evolution would have worked between the core and "non-core non-coding" regions within single minicircle DNA molecules.

3.4.5. Conclusion and a future perspective

The minicircle DNA sequence analysis of *H. nasutum* has provided an opportunity to capture new characteristics of the minicircle DNA of dinoflagellates. The existence of extensive sequence variation implies that base substitution and recombination occur often in the minicircle DNA of *H.*

nasutum and that the DNA repair and decomposition mechanisms may not function normally. In the future, it will be interesting to explore how *H. nasutum* maintains the function of its plastid.

4. General Discussion

The location of dinoflagellate minicircle DNAs has been controversial, even though they are known to encode plastid genes (Takishita et al. 2003; Laatsch et al. 2004). This has made it difficult to understand the maintenance and function of the peridinin-containing plastids of dinoflagellates cells (Howe et al. 2008) and the evolution of minicircle DNA in the dinoflagellate lineage (Howe et al. 2008). In the present study, I have demonstrated, for the first time, that the minicircle DNA of dinoflagellates is localized in the plastids. Now it is reasonable to hypothesize that the minicircle DNA has evolved from a canonical plastid DNA, as Howe et al. (2008) presumed, since it has been known that the plastids of peridinin-containing dinoflagellates originated from a red algal endosymbiont (Keeling 2010).

It appears that at least ten genes are maintained in the form of minicircle DNAs in dinoflagellates, so the control systems for the minicircle DNA, such as replication, repair, degradation, and gene expression, must also have been maintained. Although previous studies have provided limited information regarding the replication and transcription of minicircle DNA (Leung & Wong 2009; Dang & Green 2010; Barbrook et al. 2012), unknown maintenance mechanisms that are different from those in the ordinary plastids may exist and elucidating such mechanisms will be important for understanding the maintenance of functional peridinin-containing plastids in dinoflagellate cells.

Considering the maintenance of minicircle DNA in plastids, the great diversity of minicircle DNA sequences in *H. nasutum*, which has been uncovered by the present study, is puzzling. My data showed that in *H. nasutum*, most of the "coding" sequences of *psbA* and other minicircle DNA copies were non-functional, due to the accumulation of mutations. Thus, it is very strange that *H. nasutum* is able to maintain such healthy-looking plastids, although the further confirmation of plastid activity is needed. One possibility is that *H. nasutum* has evolved a novel means of maintaining its plastids without the involvement of minicircle DNA, so that the functional constraint on minicircle DNAs has loosened. This makes us to be curious to know the maintenance system of minicircle DNA in *H. nasutum*.

Another possibility is that the maintenance system for dinoflagellate minicircle DNAs is unstable and that, in *H. nasutum*, the system is now breaking down. In such a scenario, the high sequence variation observed in *H. nasutum* minicircle DNAs would indicate an intermediate state in the process of losing the function of minicircle DNAs and possibly the plastid itself. This could also provide an explanation for the fact that many dinoflagellate lineages appear to have lost or replaced their plastids (Keeling 2010). Thus, the fragile maintenance system of minicircle DNA might have driven the loss and replacement of plastids in a variety of dinoflagellate lineages.

Acknowledgements

I would like to thank:

My supervisor Dr. Ken-ichiro Ishida for offering me this research project and for editing my paper.

The students and staff of University of Tsukuba and Tokyo University and all of the members of my laboratory, both past and present, for their friendship and help.

The members of Chitose Bio Evolution Pte. Ltd. for their kind support.

References

Adams, K., & Palmer, J. D. (2003). Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution*, *29*(3), 380–395.

Allen, J. F. (2003). The function of genomes in bioenergetic organelles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1429), 19–38.

Archibald, J. M. (2005). Jumping genes and shrinking genomes – probing the evolution of eukaryotic photosynthesis with genomics. *International Union of Biochemistry and Molecular Biology: Life*, *57*(8), 539–547.

Archibald, J. M. (2015). Genomic perspectives on the birth and spread of plastids. *Proceedings of the National Academy of Sciences*, *112*(33), 10147–10153.

Archibald, J. M., & Keeling, P. J. (2002). Recycled plastids: A "green movement" in eukaryotic evolution. *Trends in Genetics*, *18*(11), 577–584.

Bachvaroff, T. R., Concepcion, G. T., Rogers, C. R., Herman, E. M., & Delwiche, C. F. (2004). Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist*, *155*(1), 65–78.

Bachvaroff, T. R., Sanchez-Puerta, M. V., & Delwiche, C. F. (2006). Rate variation as a function of gene origin in plastid-derived genes of peridinin-containing dinoflagellates. *Journal of Molecular Evolution*, *62*(1), 42–52.

Barbrook, A. C., Dorrell, R. G., Burrows, J., Plenderleith, L. J., Nisbet, R. E. R., & Howe, C. J. (2012). Polyuridylylation and processing of transcripts from multiple gene minicircles in chloroplasts of the dinoflagellate *Amphidinium carterae. Plant Molecular Biology*, *79*(4-5), 347–357.

Barbrook, A C., & Howe, C. J. (2000). Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*. *Molecular & General Genetics*, *263*(1), 152–8.

Barbrook, A. C., Santucci, N., Plenderleith, L. J., Hiller, R. G., & Howe, C. J. (2006). Comparative analysis of dinoflagellate chloroplast genomes reveals rRNA and tRNA genes. *BMC Genomics*, *7*(297), doi: 10.1186/1471-2164-7-297.

Barbrook, A. C., Symington, H., Nisbet, R. E. R., Larkum, A., & Howe, C. J. (2001). Organisation and expression of the plastid genome of the dinoflagellate *Amphidinium operculatum*. *Molecular Genetics and Genomics*, *266*(4), 632–638.

Barbrook, A. C., Visram, S., Douglas, A. E., & Howe, C. J. (2006b). Molecular diversity of dinoflagellate symbionts of cnidaria: the psbA minicircle of symbiodinium. *Protist*, *157*(2), 159–171.

Cavalier-Smith, T. (1993). Kingdom protozoa and its 18 phyla. *Microbiological reviews 57*(4), 953–994.

Chesnick, J. M., Kooistra, W., Wellbrock, U., & Medlin, L. K. (1997). Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *Journal of Eukaryotic Microbiology*, *44*(4), 314–320.

Dang, Y., & Green, B. R. (2010). Long transcripts from dinoflagellate chloroplast minicircles suggest "rolling circle" transcription. *Journal of Biological Chemistry*, *285*(8), 5196–5203.

Day, A., & Madesis, P. (2007). DNA replication, recombination, and repair in plastids. In *Cell and Molecular Biology of Plastids* (pp. 65–119), Springer.

Delwiche, C. F. (1999). Tracing the thread of plastid diversity through the tapestry of life. *The American Naturalist*, *154*(s4), S164–S177.

Dodge, J. D., & Crawford, R. M. (1969). Observations on the fine structure of the eyespot and associated organelles in the dinoflagellate *Glenodinium foliaceum*. *Journal of Cell Science*, *5*(2), 479–493.

Dorrell, R., & Howe, C. (2015). Integration of plastids with their hosts: Lessons learned from dinoflagellates. *Proceedings of the National Academy of Sciences*, *112*(33), 10247–10254.

Fensome, R. A., Saldarriaga, J. F., & Taylor, "Max" F. J. R. (1999). Dinoflagellate phylogeny revisited: reconciling morphological and molecular based phylogenies. *Grana*, *38*(2-3), 66–80.

Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Carmona, E. V., & Sogin, M. L. (1991). Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Molecular and Biochemical Parasitology*, *45*, 147–154.

Géraud, M.-L., Herzog, M., & Soyer-Gobillard, M.-O. (1991). Nucleolar localization of rRNA coding sequences in *Prorocentrum micans* Ehr.(dinomastigote, kingdom Protoctist) by in situ hybridization. *BioSystems*, *26*(2), 61–74.

Golden, S. S. (2004). Light-responsive gene expression and the biochemistry of the photosystem II reaction center. In *The Molecular Biology of Cyanobacteria* (pp. 693–714), Springer.

Gornik, S. G., Ford, K. L., Mulhern, T. D., Bacic, A., McFadden, G. I., & Waller, R. F. (2012). Loss of nucleosomal DNA condensation coincides

with appearance of a novel nuclear protein in dinoflagellates. *Current Biology*, *22*(24), 2303–2312.

Gould, S. B., Tham, W.-H., Cowman, A. F., McFadden, G. I., & Waller, R. F. (2008). Alveolins, a new family of cortical proteins that define the protist infrakingdom Alveolata. *Molecular Biology and Evolution*, *25*(6), 1219–1230.

Gould, S. B., Waller, R. F., & McFadden, G. I. (2008b). Plastid Evolution. *Annual Review of Plant Biology*, *59*(1), 491–517.

Green, B. R. (2004). The chloroplast genome of dinoflagellates–a reduced instruction set? *Protist*, *155*(1), 23–31.

Green, B. R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *The Plant Journal*, *66*(1), 34–44.

Hayashi, A. (2005). Diversity and Evolution of Minicircle DNA in the Dinoflagellate *Amphidinium massartii*. masters thesis. Kanazawa University. Japan. Retrieved from Divisions of Biological Sciences.

Hiller, R. G. (2001). Empty minicircles and petB/atpA and psbD/psbE (cytb 559 α) genes in tandem in *Amphidinium carterae* plastid DNA. *FEBS Letters*, *505*(3), 449–452.

Howe, C. J., Nisbet, R. E. R., & Barbrook, A. C. (2008). The remarkable chloroplast genome of dinoflagellates. *Journal of Experimental Botany*, *59*(5), 1035–1045.

Iida, S., Kobiyama, A., Ogata, T., & Murakami, A. (2008). The D1 and D2 proteins of dinoflagellates: unusually accumulated mutations which influence on PSII photoreaction. *Photosynthesis Research*, *98*(1-3), 415–425.

Iida, S., Kobiyama, A., Ogata, T., & Murakami, A. (2009). Identification of transcribed and persistent variants of the psbA gene carried by plastid minicircles in a dinoflagellate. *Current Genetics*, *55*(5), 583–91.

Iida, S., Kobiyama, A., Ogata, T., & Murakami, A. (2010). Differential DNA rearrangements of plastid genes, psbA and psbD, in two species of the dinoflagellate Alexandrium. *Plant and Cell Physiology*, *51*(11), 1869–1877.

Kato, S. (1982). Laboratory culture and morphology of *Colacium vesiculosum* Ehrb.(Euglenophyceae). *Japanese Journal of Phycology*, *30*, 63–67.

Keeling, P. J. (2004). Diversity and Evolutionary History of Plastids and Their Hosts the Tree of Eukaryotes. *American Journal of Botany*, *91*(10), 1481–1493.

Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1541), 729–748.

Koumandou, V. L., & Howe, C. J. (2007). The Copy Number of Chloroplast Gene Minicircles Changes Dramatically with Growth Phase in the Dinoflagellate *Amphidinium operculatum*. *Protist*, *158*(1), 89–103.

Krupinska, K., Melonek, J., & Krause, K. (2013). New insights into plastid nucleoid structure and functionality. *Planta*, *237*(3), 653–664.

Laatsch, T., Zauner, S., Stoebe-Maier, B., Kowallik, K. V., & Maier, U. G. (2004). Plastid-Derived Single Gene Minicircles of the Dinoflagellate *Ceratium horridum* Are Localized in the Nucleus. *Molecular Biology and Evolution*, *21*(7), 1318–1322.

Lee, R. E. (2008). Phycology. *Cambridge University Press.*

Leung, S. K., & Wong, J. T. Y. (2009). The replication of plastid minicircles involves rolling circle intermediates. *Nucleic Acids Research*, *37*(6), 1991–2002.

Logares, R., Shalchian-Tabrizi, K., Boltovskoy, A., & Rengefors, K. (2007). Extensive dinoflagellate phylogenies indicate infrequent marine–freshwater transitions. *Molecular Phylogenetics and Evolution*, *45*(3), 887–903.

McFadden, G. I. (2001). Primary and secondary endosymbiosis and the origin of plastids. *Journal of Phycology*, *37*(6), 951–959.

Moestrup, Ø., & Daugbjerg, N. (2007). On dinoflagellate phylogeny and classification. In *Unraveling the Algae: The Past, Present, and Future of Algal Systematics* (pp. 215–230), CRC Press.

Moore, R. B., Ferguson, K. M., Loh, W. K. W., Hoegh-Guldberg, O., & Carter, D. A. (2003). Highly organized structure in the non-coding region of the psbA minicircle from clade C Symbiodinium. *International Journal of Systematic and Evolutionary Microbiology*, *53*(6), 1725–1734.

Mungpakdee, S., Shinzato, C., Takeuchi, T., Kawashima, T., Koyanagi, R., Hisata, K., … Shoguchi, E. (2014). Massive Gene Transfer and Extensive RNA Editing of a Symbiotic Dinoflagellate Plastid Genome. *Genome Biology and Evolution*, *6*(6), 1408–1422.

Murray, S., Flø Jørgensen, M., Daugbjerg, N., & Rhodes, L. (2004). *Amphidinium* revisited. II. Resolving species boundaries in the *Amphidinium operculatum* species complex (Dinophyceae), including the descriptions of *Amphidinium trulla* sp. nov. and *Amphidinium gibbosum*. comb. nov. *Journal of Phycology*, *40*(2), 366–382.

Nelson, M. J., & Green, B. R. (2005). Double hairpin elements and tandem repeats in the non-coding region of *Adenoides eludens* chloroplast gene minicircles. *Gene*, *358*, 102–110.

Nisbet, R. E. R., Hiller, R. G., Barry, E. R., Skene, P., Barbrook, A. C., & Howe, C. J. (2008). Transcript Analysis of Dinoflagellate Plastid Gene Minicircles. *Protist*, *159*(1), 31–39.

Ohta, N., Matsuzaki, M., Misumi, O., Miyagishima, S. Y., Nozaki, H., Tanaka, K., & Shin, I. T. (2003). Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. *DNA Research*, *10*(2), 67–77.

Orr, R. J. S., Murray, S. A., Stüken, A., Rhodes, L., & Jakobsen, K. S. (2012). When naked became armored: an eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. *PLoS ONE*, *7*(11), e50004.

Palmer, J. D. (1990). Contrasting modes and tempos of genome evolution in land plant organelles. *Trends in Genetics*, *6*, 115–120.

Palmer, J. D. (2003). The symbiotic birth and spread of plastids: how many times and whodunit? *Journal of Phycology*, *39*(1), 4–11.

Petrov, A. S., Bernier, C. R., Hershkovits, E., Xue, Y., Waterbury, C. C., Hsiao, C., … Harvey, S. C. (2013). Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Research*, *41*(15), 7522–7535.

Reith, M., & Munholland, J. (1993). A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. *The Plant Cell*, *5*(4), 465–475.

Reyes-Prieto, A., Weber, A. P. M., & Bhattacharya, D. (2007). The origin and establishment of the plastid in algae and plants. *Annual Review of Genetics*, *41*, 147–168.

Sakai, A., Takano, H., & Kuroiwa, T. (2004). Organelle nuclei in higher plants: structure, composition, function, and evolution. *International Review of Cytology*, *238*, 59–118.

Saldarriaga, J. F., Taylor, "Max" F. J. R., Cavalier-Smith, T., Menden-Deuer, S., & Keeling, P. J. (2004). Molecular data and the evolutionary history of dinoflagellates. *European Journal of Protistology*, *40*(1), 85–111.

Sanchez-Puerta, M. V., Lippmeier, J. C., Apt, K. E., & Delwiche, C. F. (2007). Plastid genes in a non-photosynthetic dinoflagellate. *Protist*, *158*(1), 105–117.

Sandelius, A. S., & Aronsson, H. (2009). The Chloroplast: Interactions with the Environment, vol. 43. *Springer.*

Santos, S. R., Gutierrez-Rodriguez, C., & Coffroth, M. A. (2003). Phylogenetic identification of symbiotic dinoflagellates via length heteroplasmy in domain V of chloroplast large subunit (cp23S)—ribosomal DNA sequences. *Marine Biotechnology*, *5*(2), 130–140.

Santos, S. R., Taylor, D. J., Kinzie, R. A, Sakai, K., & Coffroth, M. A. (2002). Evolution of length variation and heteroplasmy in the chloroplast rDNA of symbiotic dinoflagellates (Symbiodinium, Dinophyta) and a novel insertion in the universal core region of the large subunit rDNA. *Phycologia*, *41*(4), 311–318.

Sato, N., Ohshima, K., Watanabe, A., Ohta, N., Nishiyama, Y., Joyard, J., & 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. *The Plant Cell*, *10*(5), 859–872.

Sato, N., & Ohta, N. (2001). DNA-binding specificity and dimerization of the DNA-binding domain of the PEND protein in the chloroplast envelope membrane. *Nucleic Acids Research*, *29*(11), 2244–2250.

Schnepf, E., & Deichgräber, G. (1984). "Myzocytosis," a kind of endocytosis with implications to compartmentation in endosymbioses: observations on *Paulsenella (Dinophyta). Naturwissenschaften*, *71*(4), 218–219.

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., & Satoh, N. (2013). Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology*, *23*(15), 1399–1408.

Simpson, C. L., & Stern, D. B. (2002). The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications. *Plant Physiology*, *129*(3), 957–966.

Soltis, P. S., Soltis, D. E., & Chase, M. W. (1999). Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature*, *402*(November), 402–404.

Spector, D. L. (1984). Dinoflagellate nuclei. In *Dinoflagellates* (pp. 107–147). Academic.

Spector, D. L., Vasconcelos, A. C., & Triemer, R. E. (1981). DNA duplication and chromosome structure in the dinoflagellates. *Protoplasma*, *105*(3-4), 185–194.

Stoebe, B., & Maier, U.-G. (2002). One, two, three: nature's tool box for building plastids Review article. *Evolution*, *219*(3), 123–130.

Stoebe, B., & Kowallik, K. V. (1999). Gene-cluster analysis in chloroplast genomics. *Trends in Genetics*, *15*(9), 344–7.

Takishita, K., Ishikura, M., Koike, K., & Maruyama, T. (2003). Comparison of phylogenies based on nuclear-encoded SSU rDNA and plastid-encoded psbA in the symbiotic dinoflagellate genus *Symbiodinium*. *Phycologia*, *42*(3), 285–291.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729.

Taylor, F. J. R. (1990). Phylum Dinoflagellata. In *Handbook of Protoctista* (pp. 419–437). Jones and Bartlett Publishers.

Taylor, F. J. R., Hoppenrath, M., & Saldarriaga, J. F. (2007). Dinoflagellate diversity and distribution. In *Protist Diversity and Geographical Distribution* (pp. 173–184). Springer.

Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C. F., & Jakobsen, K. S. (2000). Phylogenetic analyses indicate that the 19′ hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Molecular Biology and Evolution*, *17*(5), 718–729.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics, 5*(2), 123–135.

Wang, Y., Jensen, L., Højrup, P., & Morse, D. (2005). Synthesis and degradation of dinoflagellate plastid-encoded psbA proteins are light-regulated, not circadian-regulated. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(8), 2844–2849.

Wang, Y., & Morse, D. (2006). Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate *Lingulodinium*. *Nucleic Acids Research*, *34*(2), 613–9.

Watanabe, M. M., Suda, S., Inouya, I., Sawaguchi, T., & Chihara, M. (1990). *Lepidodinium viride* gen. et sp. nov. (Gymnodinaiales, Dinophyta), a green dinoflagellate with a chlorophyll *A*- and *B*-containing endosymbiont. *Journal of Phycology*, *26*(4), 741–751.

Wise, R. R., & Hoober, J. K. (2007). The Structure and Function of Plastids, vol. 23. *Springer Science & Business Media*.

Wolters, J. (1991). The troublesome parasites–molecular and morphological evidence that Apicomplexa belong to the dinoflagellate-ciliate clade. *Bio Systems*, *25*(1-2), 75–83.

Zhang, H., Bhattacharya, D., & Lin, S. (2007). A three-gene dinoflagellate phylogeny suggests monophyly of Prorocentrales and a basal position for *Amphidinium* and *Heterocapsa*. *Journal of Molecular Evolution*, *65*(4), 463–474.

Zhang, Z., Cavalier-Smith, T., & Green, B. R. (2001). A family of selfish minicircular chromosomes with jumbled chloroplast gene fragments from a dinoflagellate. *Molecular Biology and Evolution*, *18*(8), 1558–1565.

Zhang, Z., Cavalier-Smith, T., & Green, B. R. (2002). Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins. *Molecular Biology and Evolution*, *19*(4), 489–500.

Zhang, Z., Green, B. R., & Cavalier-Smith, T. (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature*, *400*(6740), 155–159.

Zhang, Z., Green, B. R., & Cavalier-Smith, T. (2000). Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. *Journal of Molecular Evolution*, *51*(1), 26–40.

Tables

Table 2.1. Primers used in this study

Primer Name	Primer Sequence (5′ –3′)	Application	Annealing Tempreture
Primers for acquisition *psbA* sequences			
Dino*psbA*F1i†	YTWTAYATHGGWTGGTTYGG	degenerate *psbA* inward forward	43 °C
Dino*psbA*R1i†	TCRTGCATWACYTCNATNCC	degenerate *psbA* inward reverse	43 °C
DinopabAF2o†	TGGCCWGTWATHGGWATHTGG	degenerate *psbA* outward forward	43 °C
DinopabAR2o†	CKRATWCCATCDATATCDACTG	degenerate *psbA* outward reverse	43 °C
Dino23SF1i†	TMRTRTTYATCAASCGACTGT	degenerate 23S rRNA gene inward forward	64 °C
Dino23SR1i†	AAGCCGACATCGAGGTGCMAA	degenerate 23S rRNA gene inward reverse	64 °C
HN23SF1o	TTACTGCAGGGATAACAGGCT	23S rRNA gene outward forward	50 °C
HN23SR1o	AATACTGGTCACATCTCAGTTC	23S rRNA gene outward reverse	50 °C
Primers for probe synthesis for *H. nasutum*			
HNlsuF1	CTGGCGATGAGGGATGAACCTA	nuclear LSU rRNA gene forward	
28-1483R†	GCTACTACCACCAAGATCTGC	nuclear LSU rRNA gene reverse	
HN*psbA*F1	AGCTGCAGTCGCTTATATTGCAGC	minicircle *psbA* coding region forward	
HN*psbA*R1	CACTGATTCCAAGTGGCATACCATC	minicircle *psbA* coding region reverse	
HN23SF2	CCTGTGAAGCTTTATAGGCAATGAGAGTA	minicircle 23S rRNA coding region forward	
HN23SR2	AAGCCGACATCGAGGTGCCAA	minicircle 23S rRNA coding region reverse	
Primers for probe synthesis for *A. massartii*			
AMlsuF1	GCACCAGCAACCAACTGATC	nuclear LSU rRNA gene forward	
28-1483R‡	GCTACTACCACCAAGATCTGC	nuclear LSU rRNA gene reverse	
AM*psbA*R1	GCACGGTTGAGGATATCAGC	minicircle *psbA* coding region forward	
AM*psbA*F1	TGGGGTTCTTTCGTTCAAAC	minicircle *psbA* coding region reverse	
AMcoreF1	GGAAATAACCCCTAGACTTTAACGG	minicircle non-coding core region forward	
AMcoreR1	TAGTTTAACGTCATTGCGGACCAGA	minicircle non-coding core region reverse	

† from Hayashi 2005

‡ from Murray et al. 2004

Primers unique to this study show no reference.

Table 3.1. Primers used in this study

Primer Name	Primer Sequence (5′ -3′)	Application	Annealing Tempreture
Primers for acquisition *psbA* sequences			
Dino*psbA*F1i†	YTWTAYATHGGWTGGTTYGG	degenerate *psbA* inward forward	43 °C
Dino*psbA*R1i†	TCRTGCATWACYTCNATNCC	degenerate *psbA* inward reverse	43 °C
Dino*pabA*F2o†	TGGCCWGTWATHGGWATHTGG	degenerate *psbA* outward forward	43 °C
Dino*pabA*R2o†	CKRATWCCATCDATATCDACTG	degenerate *psbA* outward reverse	43 °C
HN*psbA*F2	TGGAGGTTCATTATTCAGTGCAAT	minicircle *psbA* coding region forward	64 °C
HN*psbA*F3	GCAGAACACAACATTTTGATGCAC	minicircle *psbA* coding region forward	64 °C
HN*psbA*F4	AGCTGCAGTCGCTTATATTGCAGC	minicircle *psbA* cDNA coding region forward	50 °C
HN*psbA*R2	TAGCATGAAATTAAATGTACCACTGATT	minicircle *psbA* coding region reverse	64 °C
HN*psbA*R3	GAGCATTTCGTTCATGCATCACTTCC	minicircle *psbA* cDNA coding region reverse	50 °C
Primers for acquisition 23S rRNA sequences			
Dino23SF1i†	TMRTRTTYATCAASCGACTGT	degenerate 23S rRNA gene inward forward	43 °C
Dino23SR1i†	AAGCCGACATCGAGGTGCMAA	degenerate 23S rRNA gene inward reverse	43 °C
HN23SF1o	TTACTGCAGGGATAACAGGCT	23S rRNA gene outward forward	50 °C
HN23SR1o	AATACTGGTCACATCTCAGTTC	23S rRNA gene outward reverse	50 °C
HN23SF2	CCTGTGAAGCTTTATAGGCAATGAGAGTA	23S rRNA cDNA coding region forward	50 °C
HN23SR2	AAGCCGACATCGAGGTGCCAA	23S rRNA cDNA coding region reverse	50 °C

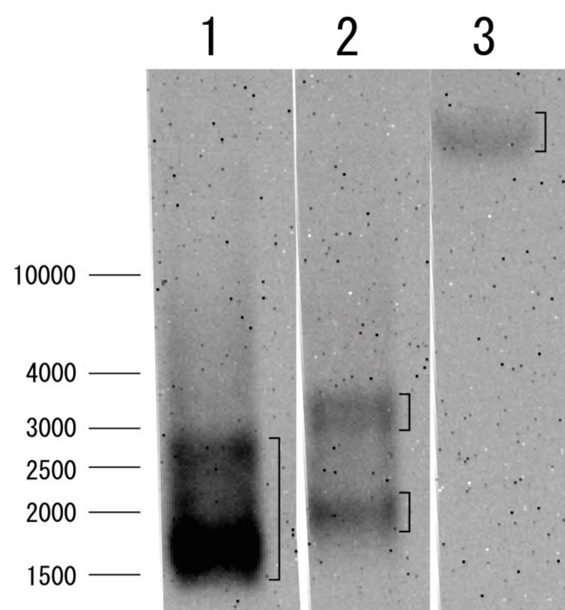
† from Hayashi 2005

Primers unique to this study show no reference.

Figures

Figure 2.1

Total genome Southern blotting for *H. nasutum*. Lane 1: Total DNA hybridized with the minicircle-encoded *psbA* probe, showing strong, broad bands ranging from 1500 bp to 2800 bp. Lane 2: Total DNA hybridized with the minicircle-encoded 23S rRNA probe, showing two broad bands, which ranged from 1900 bp to 2100 bp and 3000 bp to 3600 bp. Lane 3: Total DNA hybridized with the nuclear-encoded 28S rRNA gene probe, which specifically hybridizes with nuclear DNA.



1
2
3
10000
4000
3000
2500
2000
1500

Figure 2.2

Total genome Southern blotting for *A. massartii*. Lane 1: *tPA*-specific probe (negative control). Lane 2: Nuclear-encoded 28S rRNA gene probe, which hybridized with nuclear DNA (indicated by a bracket). Lane 3: Minicircle-encoded *psbA* probe showing a weak band at ~3.5 kb and a strong band at ~5 kb, which are probably linear monomers and relaxed monomeric circles, respectively (arrowheads). Lane 4: Minicircle-encoded core region probe that gave a weak band at ~3 kb and several strong bands at ~4 kb, 5 kb, and 6 kb (arrowheads), which indicates the presence of several minicircle DNAs with different genes.

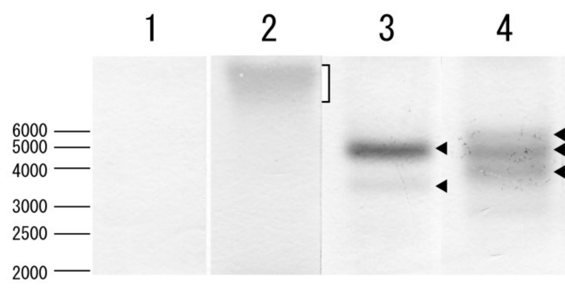

1
2
3
4
6000
5000
4000
3000
2500
2000

Figure 2.3

Fluorescence *in situ* hybridization (FISH) images with specific probe for controls and minicircle DNAs in *A. massartii*. (a) Unhybridized probe specific for *tPA* could not detect any signal as a control; (b) Localization of a nuclear gene detected by a probe specific for the 28S rRNA gene in the nucleus as a control; (c) Localization of the coding region of minicircle DNAs detected by a specific probe for *psbA* in the plastid; (d) Localization of a core region detected by a specific probe for the core sequence in the plastid. (1) FITC images (green); (2) DAPI stained images (blue); (3) Chlorophyll autofluorescence images (red); (4) Merged images of the FITC, DAPI, and chlorophyll images. (5) Phase contrast (PC) images. Scale bar, 5 μm.

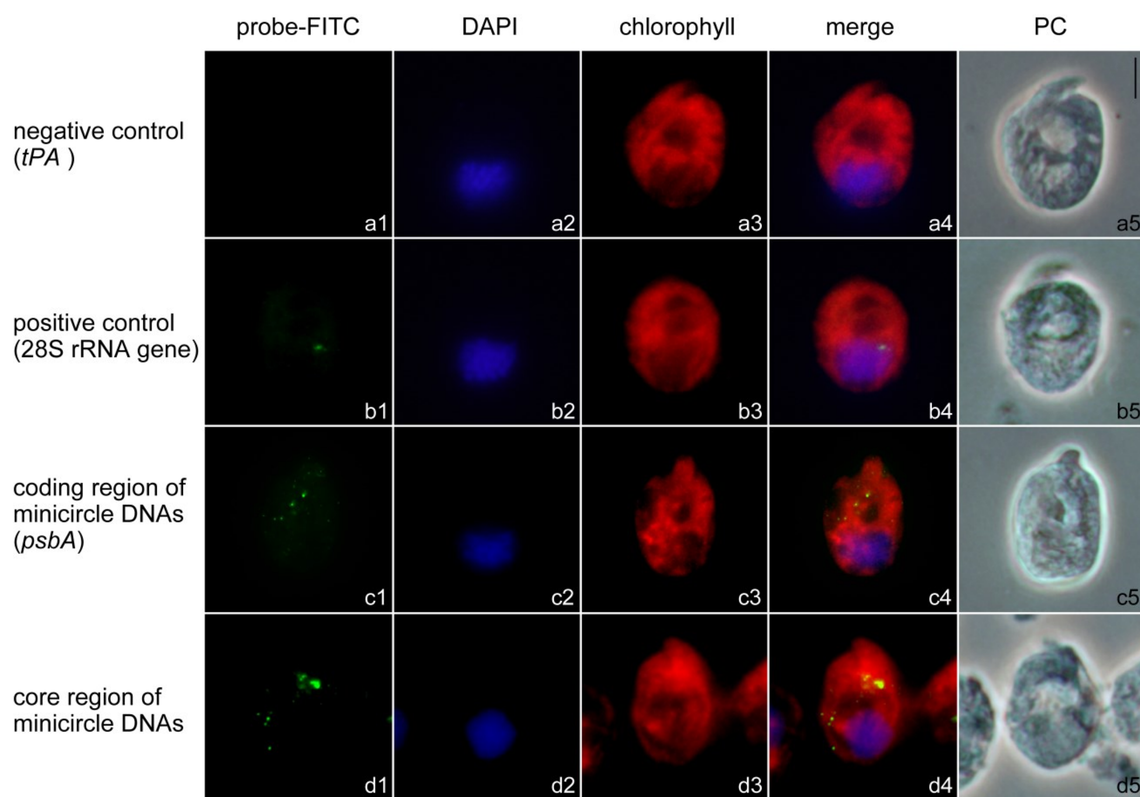

probe-FITC
DAPI
chlorophyll
merge
PC
negative control
(tPA)
positive control
(28S rRNA gene)
coding region of
minicircle DNAs
(psbA)
core region of
minicircle DNAs
a1
a2
a3
a4
a5
b1
b2
b3
b4
b5
c1
c2
c3
c4
c5
d1
d2
d3
d4
d5

Figure 2.4

Localization of minicircles using FISH images from confocal laser-scanning microscopy. Consecutive serial sections of minicircle DNA FISH images obtained using a core region probe for minicircle DNA. The cell images were acquired in 1-µm intervals from the center to the ventral side. FITC images (green), DAPI stained images (blue), and chlorophyll autofluorescence images (red) are merged. Scale bar, 5 µm.

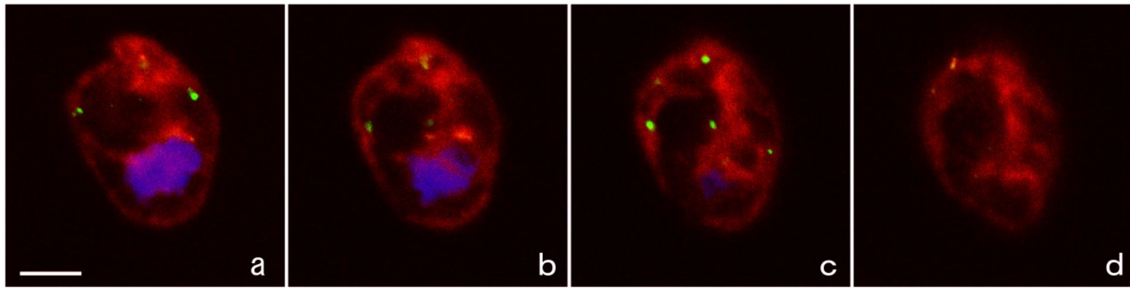

a
b
c
d

Figure 2.5

Fluorescence *in situ* hybridization (FISH) images with specific probes for controls and minicircle DNA in *H. nasutum*. (a) Unhybridized probe specific for *tPA* could not detect any signal as a control. (b) Localization of a nuclear gene, detected using a probe specific for the 28S rRNA gene as a control. (c) Localization of the coding region of a minicircle DNA, detected using a probe specific for *psbA*. (1) FITC images (green); (2) DAPI stained images (blue); (3) Chlorophyll autofluorescence images (red); (4) Merged images of FITC, DAPI, and chlorophyll images. (5) Phase contrast (PC) images. Scale bar, 10 μm.

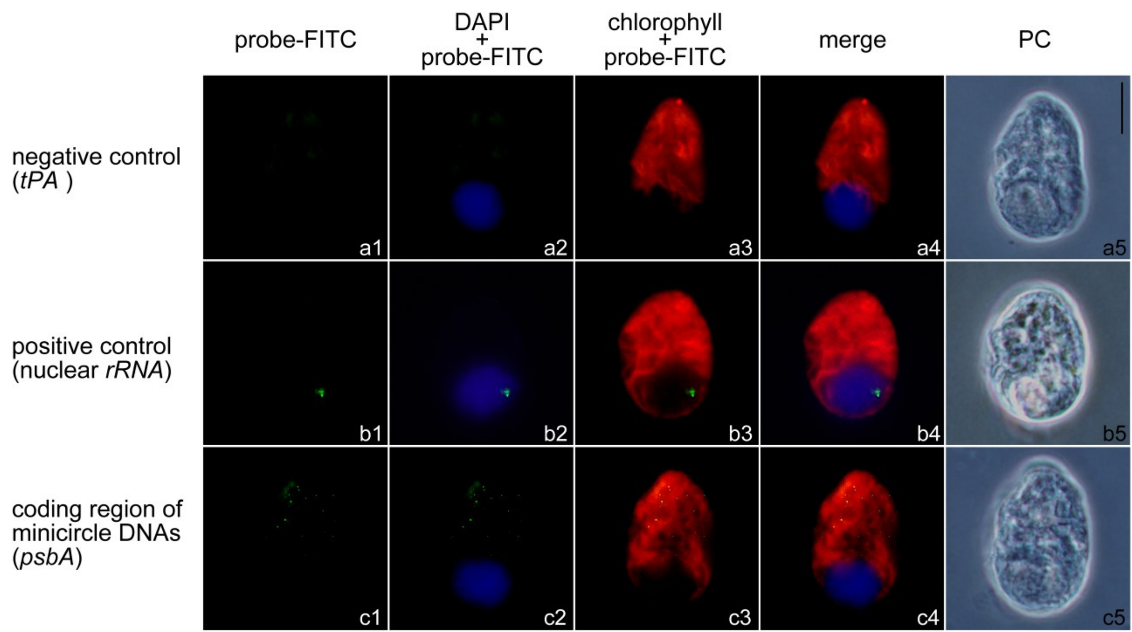

probe-FITC
DAPI + probe-FITC
chlorophyll + probe-FITC
merge
PC
negative control (tPA)
positive control (nuclear rRNA)
coding region of minicircle DNAs (psbA)
a1
a2
a3
a4
a5
b1
b2
b3
b4
b5
c1
c2
c3
c4
c5

Figure 3.1

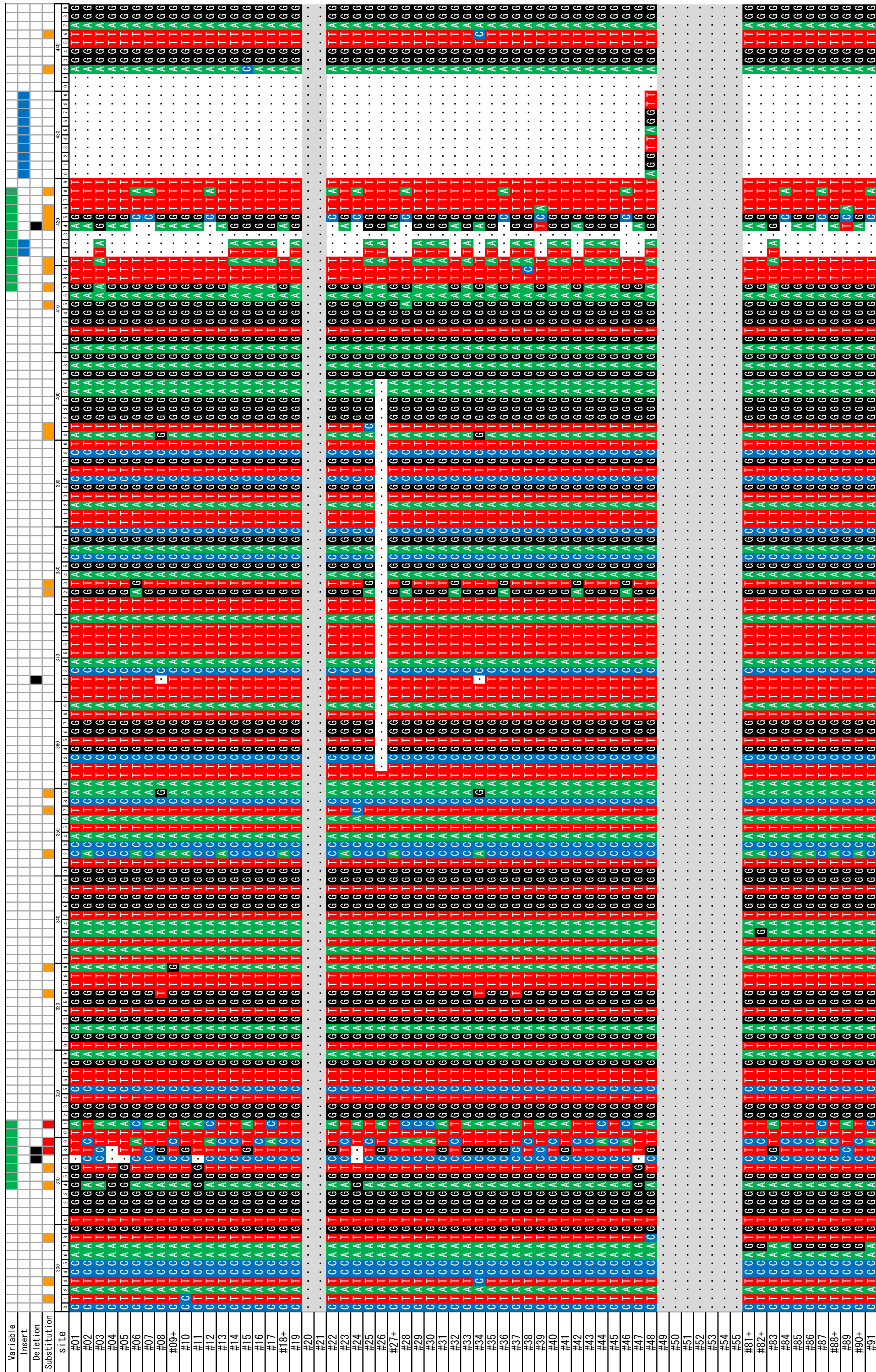
Comparison of DNA and cDNA sequence variation of *psbA*. Fifty-five partial *psbA* sequences (#01-#55) from DNA and 11 *psbA* partial sequence (#81-#91) from cDNA. Substitution sites are orange (1 type) or red (2 or 3 type), deletion sites are black, and insertion sites are blue (1 type) or dark blue (2 or 3 type). Variable regions are green, and gray backgrounds indicate missing sequences. The "+" symbol indicates, no frameshift mutations. Forward primers and reverse primer of the sequences are as follows: #01 - #16, Dino*psbA*F1i and Dino*psbA*R1i; #17 - #27, #48, HN*psbA*F2 and HN*psbA*R2; #28 - #47, HN*psbA*F3 and HN*psbA*R2; #49 - #55, Dino*psbA*F2o and Dino*psbA*R2o; #81 - #91, HN*psbA*F4 and HN*psbA*R3.

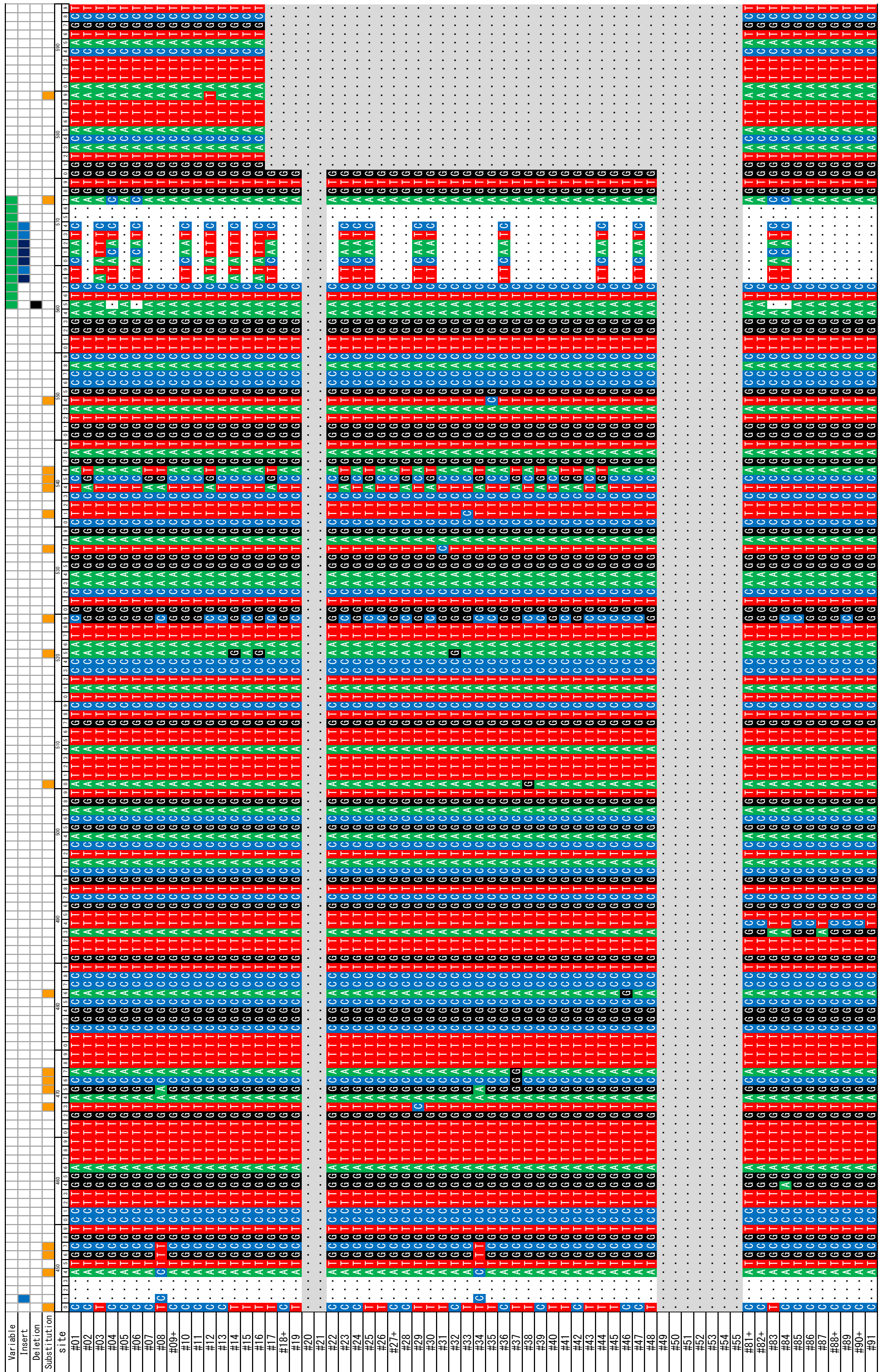
Variable
Insert
Deletion
Substitution
site

#01
#02
#03
#04
#05
#06
#07
#08
#09+
#10
#11
#12
#13
#14
#15
#16
#17
#18+
#19
#20
#21
#22
#23
#24
#25
#26
#27+
#28
#29
#30
#31
#32
#33
#34
#35
#36
#37
#38
#39
#40
#41
#42
#43
#44
#45
#46
#47
#48
#49
#50
#51
#52
#53
#54
#55
#81+
#82+
#83
#84
#85
#86
#87
#88+
#89
#90+
#91

Variable
Insert
Deletion
Substitution

site	150–299
#01	TGCAGCATTTATTCTAGCACCTCCAGCGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCGAATGCTATCGGTGTTCATTT
#02	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#03	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTT..........TCAGTCCAAGTTCAAATGCAATAGGTGTTCATTT
#04	......................CCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCGAATGCTATCGGTGTTCATTT
#05	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCGAATGCTATCGGTGTTCATTT
#06	TGCAGCATTTATTCTAGCACCTCCAATGGACATTGATGGAATCCGAGAACCAGTAGCTGAGAGTCTTCTGTATGGGAAT......AATATGATGAGTGGTGCAGGA............GTTCCAAGTTCAAATGCAATAGGTGTTCATTT
#07	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAATGGTTATAATATGATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#08	TGCAGCATTTATTCTAGCACCTCCAGTGGATATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#09+	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#10	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AACATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCTATCGGTGTTCATTT
#11	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCGAATGCTATCGGTGTTCATTT
#12	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGAAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTT..........TCAGTCCAAGTTCAAATGCAATAGGTGTTCATTT
#13	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCAAATGCAATAGGTGTTCATTT
#14	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTG............ATTCCAAGTTCAAATGCTATCGGTGTTCATTT
#15	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCAAATGCAATAGGTGTTCATTT
#16	TGCAGCATTTATTCTAGCACCTCCAGTGGACGTTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTG............ATTCCAAGTTCAAATGCTATCGGTGTTCATTT
#17	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTT..........TCAGTCCAAGTTCAAATGCAATAGGTGTTCATTT
#18+	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#19	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTG............ATTCCAAGTTCAAATGCTATCGGTGTTCATTT
#20	....
#21	....
#22	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTAGGGGAAT......AATATGATGAGTGGTGCAGGA............GTTCCAAGTTCAAATGCAATAGGTGTTCATTT
#23	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#24	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCGAATGCAATAGGTGTTCATTT
#25	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#26	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTT..........TCAGTCCAAGTTCAAATGCAATAGGTGTTCATTT
#27+	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#28	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCAAATGCTTTCGGTGTTCATTT
#29	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTT..........TCAGTCCAAGTTCAAATGCAATAGGTGTTCATTT
#30	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTATATGGGAAT......AATATCATGAGTGGTGCAGTT..........TCAGTCCAAGTTCAAATGCAATAGGTGTTCATTT
#31	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTTAGTTAC..AGTCATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#32	TGCAGCATTTATTCTAGCACCTTCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#33	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTT..........TCAGTCCAAGTTCAAATGCAATAGGTGTTCATTT
#34	TGCAGCATTTATTCTAGCACCTCCAGTGGATATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#35	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#36	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCAAATGCAATAGGTGTTCATTT
#37	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#38	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTG............ATTCCAAGTTCAAATGCTATCGGTGTTCATTT
#39	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCTATCGGTGTTCATTT
#40	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTG............ATTCCAAGTTCAAATGCTATAGGTGTTCATTT
#41	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTTAGTTAC..AGTCATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#42	TGCAGCATTTATTCTAGCACCTTCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#43	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTG............ATTCCAAGTTCAAATGCTATCGGTGTTCATTT
#44	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTT..........TCAGTCCAAGTTCAAATGCAATAGGTGTTCATTT
#45	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTG............ATTCCAAGTTCAAATGCTATCGGTGTTCATTT
#46	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATGATGAGTGGTGCAGGA............GTTCCAAGTTCAAATGCAATAGGTGTTCATTT
#47	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCGAATGCTATCGGTGTTCATTT
#48	TGCAGCATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#49	TGCAGCATTTATTCTAGCACCTC...
#50	TGCAGCATTTATTCTAGCACCTC...
#51	TGCAGCATTTATTCTAGCACCTC...
#52	TGCAGCATTTATTCTAGCACCTC...
#53	TGCAGCATTTATTCTAACACCTC...
#54	TGCAGCATTTATTCTAGCACCTC...
#55	TGCAGCATTTATTCTAGCACCTC...
#81+	......TTTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#82+	......TTTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#83	......TTTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTT..........TCAGTCCAAGTTCAAATGCAATAGGTGTTCATTT
#84	......TTTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTTAGTTAC..AGTCATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#85	......TTTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGGA............GTTCCAAGTTCAAATGCAATAGGTGTTCATTT
#86	......TTTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAGT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#87	......ATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#88+	......TTTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#89	......ATTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCTATCGGTGTTCATTT
#90+	......TTTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATCATGAGTGGTGCAGTC............ATTCCAAGTTCAAATGCAATAGGTGTTCATTT
#91	......TTTTATTCTAGCACCTCCAGTGGACATTGATGGAATCCGAGAACCAGTAGCTGGGAGTCTTCTGTATGGGAAT......AATATGATGAGTGGTGCAGGA............GTTCCAAGTTCAAATGCAATAGGTGTTCATTT





site	600–749
Variable	
Insert	
Deletion	
Substitution	
#01	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATAGTTC
#02	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#03	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACC......GTAGGAGATTCTTC
#04	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTTACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#05	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCCATGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATAGTTC
#06	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACAGTAG..GAAGGAGATTCTTC
#07	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACACA....GTAGGAGATTCTTC
#08	AGTATTTCAAGCAGAACACAACATTTTGGTGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATTCTTC
#09+	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#10	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCCTG.ATC..TGGTCACATCATCATTATTGGCTGAAACA......GTAGGAGATTCTTC
#11	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCCATGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATAGTTC
#12	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCA..
#13	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#14	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATTCTTC
#15	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCA..
#16	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATTCTTC
#17	....GTAAGAATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#18+	....GCATGAATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#19	....GCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATTCTTC
#20	....GTAAGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#21	....GCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#22	....GTATGGATC..TGGTCACATCA......................................
#23	....GCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#24	....GCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATTCTTC
#25	....GCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#26	....GCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACACA....GTAGGAGATTCTTC
#27+	....GCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATAGTTC
#28	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#29	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATAACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#30	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#31	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACACA....GTAGGAGATTCTTC
#32	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATTCTTC
#33	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATTCTTC
#34	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATTCTTC
#35	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGGTCACTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#36	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTTCGGGTTCA...
#37	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGT..C
#38	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATC..TGGTCACATCATCATTATGGGCTGAAACA......GTAGGAGATAGTTC
#39	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTTCGGGTTCA...
#40	....
#41	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#42	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATTCTTC
#43	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGACATGCATGGATC..TGGTCACATCATCATTATGGGCTGAAACA......GTAGGAGATAGTTC
#44	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#45	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGACATGCATGGATC..TGGTCACATCATCATTATGGGCTGAAACA......GTAGGAGATAGTTC
#46	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACAGTAG..GAAGGAGATTCTTC
#47	..................................CCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACA......GTAGGAGATAGTTC
#48	....GCATGGGTCACTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#49	....
#50	....
#51	....
#52	....
#53	....
#54	....
#55	....
#81+	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#82+	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#83	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTCTAGCTGAAACC......GTAGGAGATTCTTC
#84	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCA..
#85	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#86	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAAAGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGACTCTTC
#87	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTTCGGGTTCA...
#88+	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#89	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTTCGGGTTCA...
#90+	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCATTATTCAGTGCAATGCATGGATCGTTGGTCACATCATCACTTTTAGCTGAAACA......GTAGGAGATTCTTC
#91	AGTATTTCAAGCAGAACACAACATTTTGATGCACCCATTCCACATCTTTGGTGTGGCAGGTGTATTTGGAGGTTCA..

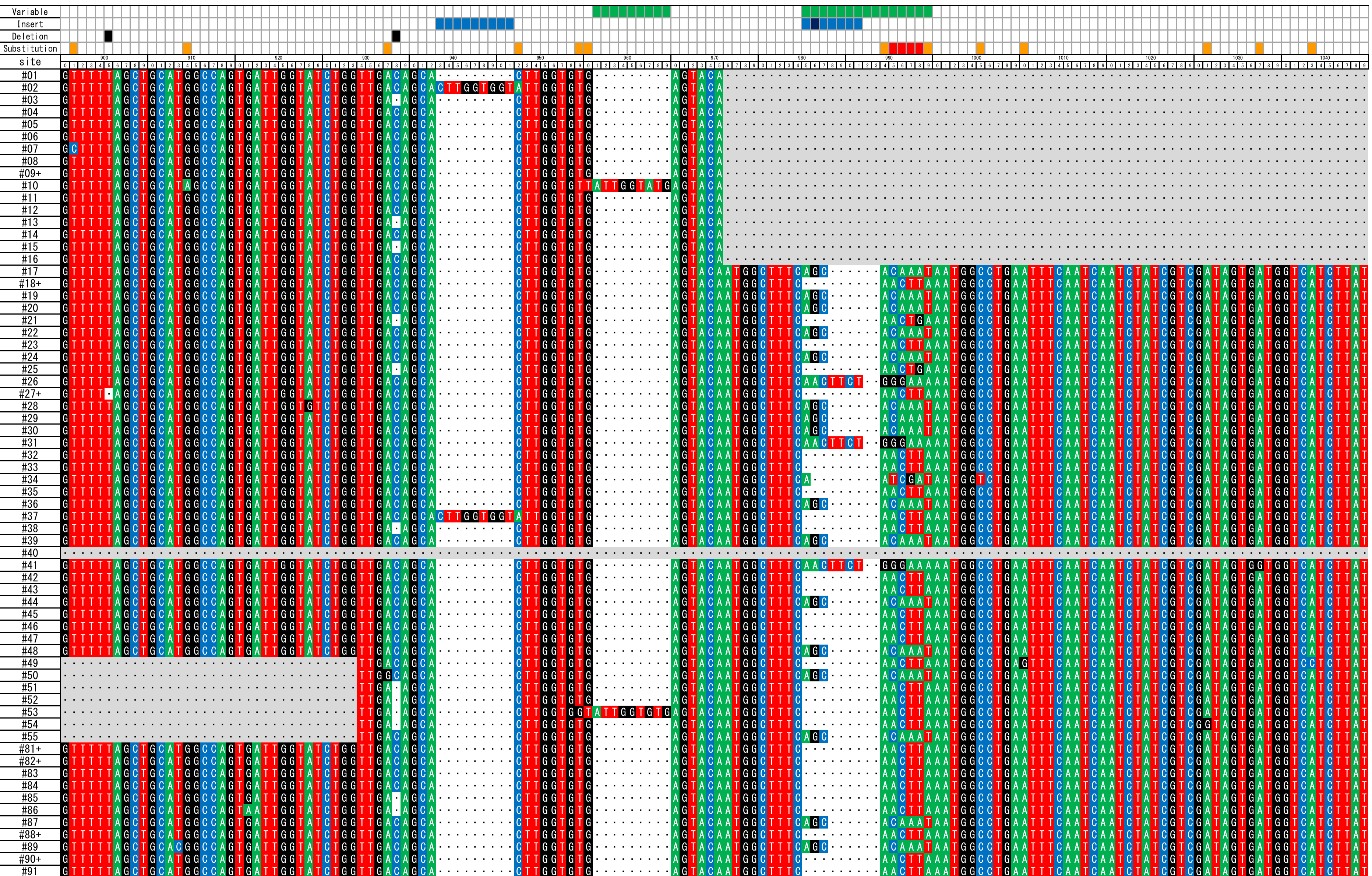


Figure 3.2

Comparison of DNA and cDNA sequence variation of the 23S rRNA gene. Seventeen partial 23S rRNA gene sequences (#01-#17) from DNA and 5 23S rRNA gene partial sequence (#81-#85) from cDNA. Substitution sites are orange (1 type) or red (2 or 3 type), deletion sites are black, and insertion sites are blue (1 type) or dark blue (2 or 3 type). Variable regions are green, and gray backgrounds indicate missing sequences. Forward primers and reverse primer of the sequences are as follows: #01 - #10, Dino23SF1i and Dino23SR1I; #11 - #17, HN23SF1o and HN23SR1; #81 - #85, HN23SF2 and HN23SR2.

Variable
Insert
Delition
Substitution
site
#01
#02
#03
#04
#05
#06
#07
#08
#09
#10
#11
#12
#13
#14
#15
#16
#17
#81
#82
#83
#84
#85

Variable
Insert
Delition
Substitution
site
#01
#02
#03
#04
#05
#06
#07
#08
#09
#10
#11
#12
#13
#14
#15
#16
#17
#81
#82
#83
#84
#85

Variable
Insert
Delition
Substitution
site
#01
#02
#03
#04
#05
#06
#07
#08
#09
#10
#11
#12
#13
#14
#15
#16
#17
#81
#82
#83
#84
#85

Variable
Insert
Delition
Substitution
site
#01
#02
#03
#04
#05
#06
#07
#08
#09
#10
#11
#12
#13
#14
#15
#16
#17
#81
#82
#83
#84
#85

Variable / Insert / Delition / Substitution / site	1200–1349
#01	CCCTGTGAAGCTTTATAGGCAATGAGAGTATTGGTCCTGCTTCTG··AATGCAGTTAACCTTCTTCTCAATTGAATGGAC·TATCGCTCCATTCACAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#02	CCCTGTGAAGCTTTATAGGCAATGAGAGTATTGGTCCTGCTGATTA·GATGCAGTAAACCTTCTTCTCAATTGAATGGA··TATAGCTCCATTCCAAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#03	CCCTGTGAAGCTTTATAGGCAATGAGAGTAATGGTCCTGCTTCTT·CGATGCAGTAAACCTTCTTCTCAATTGAATGGAC·TATCGCTCCATTCACAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#04	CCCTGTGAAGCTTTATAGGGAATGAGAGTATTGGTCCTGCAACTCTAACAGCAGTAAACCTTCTTCTCAATTGAATGGAC·TATCGCTCCATTCACAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#05	CCCTGTGAAGCTTTATAGGCAATGAGAGTATTGGTCCTGCTGATTA·GATGCAGTAAACCTTCTTCTCAATTGAATGGA··TATAGCTCCATTCCAAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#06	CCCTGTGAAGCTTTATAGGCTATGAGAGTATTGGTCCTGCTTCTTT·GATGCAGTAAACCTTCTTCTCACTTGAATGGA··CTGATATCCATTCCCAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#07	CCCTGTGAAGCTCTATAGGCAATGAGAGTATTGGTCCTGCTTCTTC·GATGCAGTAAACCTTCTTCTCAATTGAATGGAC··ATAGCTCCATTCCAAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGGTTATTTCATTGGG
#08	CCCTGTGAAGCTTTATAGGCAATGAGAGTATTGGTCCTGCAACTCTAACAGCAGTTAACCTTCTTCTCAATTGAATGGAC·TGAGATCCCATTCACAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#09	CCCTGTGAAGCTTTATAGGCTATGAGAGTATTGGTCCTGCTTCTTT·GATGCAGTAAACCTTCTTCTCAATTGAATGGACACTAAGATCCATTCACAAGATTGATGCAACAAGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#10	CCCTGTGAAGCTTTATAGGCAATGAGAGTATTGGTCCTGCAACTCTAACAGCAGTTAACCTTCTTCTCAATTGAATGGAC·TGAGATCCCATTCACAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#11	··
#12	··
#13	··
#14	··
#15	··
#16	··
#17	··
#81	CCCTGTGAAGCTTTATAGGCAATGAGAGAATTGGTCCTGCTTCTTC·GATGCAGTAAACCTTCTTCTCAATTGAATGGAC·TTTGA·TCCATTCACAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#82	CCCTGTGAAGCTTTATAGGCTATGAGAGTATTGGTCCTGCTTCTTC·GATGCAGTAAACCTTCTTCTCAATTGAATGGAC·TTTGAATCCATTCGCAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTTATTGGG
#83	CCCTGTGAAGCTTTATAGGCTATGAGAGTATTGGTCCTGCTTCTTT·GATGCAGTAAACCTTCTTCTCACTTGAATGGAC·T·GATATCCATTCCCAAGATTGACGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#84	CCCTGTGAAGCTTTATAGGCTATGAGAGTATTGGTCCTGCTTCTTAATTTGCAGTTAACCTTCTTCTCAATTGAATGGACAAAGACCTCCATTCCCAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG
#85	CCCTGTGAAGCTTTATAGGGAATGAGAGTATTGGTCCTGCAACTCTAACAGCAGTAAACCTTCTTCTCAATTGAATGGAC·TATCGCTCCATTCACAAGATTGATGCAGCATGGATGATTGCATCAATGGTTCATGATTATTTCATTGGG

Variable / Insert / Delition / Substitution / site	1350–1499
#01	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCGTTAGGTCTCACAAGTGAATAGATAGAAGGTTCAGCAAATTCGACTCAAA·GTATTGAAATGCTATCGAATTGACCTTCAGACGAATGTATCAGATAAAA
#02	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCGTTAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACTCAA··GTATTGAAATGCTATCGAATTGACCTTCTCACGAATGTATCAGATAAAA
#03	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCAACAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACCTCGC·GTATTGAAATGCTATCGAATTGACCTTCTCACGAATGTATCATATAAAA
#04	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATGACACCAACAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACCGCTA·GTATTGAAATGCTATCGAATTGACCTTCTCACGAATGTATCAGATAAAA
#05	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCGTTAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACTCAA··GTATTGAAATGCTATCGAATTGACCTTCTCACGAATGTATCATATAAAA
#06	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCGTTAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACTTTG·GTATTGAAATGCTATCGAATTGACCTTCAGACGAATGTATCAGATAAAA
#07	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCAACAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACCGCTA·GTATTGAAATGCTATCGAATTGACTTTCTCACGAATGTATCAGATAAAA
#08	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCAACAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCGACCTAAC·GTATTGAAATGCTATCGAATTGACCTTCTCACGAATGTATCAGATAAAA
#09	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCAACAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAAACGAAA·GTATTGAAATGCTATCGAATTGACCTTCTCACGAATGTATCAGATAAAA
#10	GCGATAACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCGTTAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCGAACCTAACGTATTGAAATGCTATCGAATTGACCTTCAGACGAGTGTATCAGATAAAA
#11	··
#12	··
#13	··
#14	··
#15	··
#16	··
#17	··
#81	GCGATGACCTTATTATTTAACCTTCAGGTGTACAGAGCTAATATGACACCAACAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACCGCTA·GTATTGAAATGCTATCGAATTGACCTTCTACCGAATGTATCAGATAAAA
#82	GCGATGACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATGGCACCAACAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACCTCGC·GTATTGAAATGCTGTCGAATTGACCTTCTCACGAATGTATCAGATAAAA
#83	GCGATGACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCAACAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACCTTTG·GTATTGAAATGCTGTCGAATTGACCTTCAGACGAATGTATCAGATAAAA
#84	GCGATGACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATTCCACCGTTAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACCTTCA·GTATTGAAATGCTGTCGAATTGACCTTCTTACGAATGTATCAGATAAAA
#85	GCGATGACCTTATTATTTGACCTTCAGGTGTACAAAGCTAATATGACACCAACAGGTCTCACAAGTGAATGGATAGAAGGTTCAGCAAATTCAACCGCTA·GTATTGAAATGCTGTCGAATTGACCTTCTCACGAATGTATCAGATAAAA

Variable / Insert / Delition / Substitution	
site	1500–1649
#01	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#02	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#03	GTTACTGCAGGGATAACAGGCTTCATATTTGTATGCAGCTCACATGTTTGCAAGAAA...
#04	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#05	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#06	GTTACTGCAGGGATAACAGGCCTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#07	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#08	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#09	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#10	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#11	......................TCATATTTGCATGCAGCTCACATGTTTGCAAGAAATTGGCACCTCGATGTCGGCTTGCCTGATGGAGAGCAGAAGGAATGCTCATAGCTGGGAGTGCTCATCCCAATATTGGTACGTGAGCTGGGTTC
#12	......................TCATATTTGCATGCAGCTCACATGTTTGCAAGAAATTGGCACCTCGATGTCGGCTTGCCTGATGGAGAGCAGAAGGAATGCTTATAGCTGGGAGTGCTCATCCCAATATTGGTACGTGAGCTGGGTTC
#13	......................TCATATTTGCATGCAGCTCACATGTTTGCAAGAAATTGGCACCTCGATGTCGGCTTGCCGATTGGAGAGCAGAAGGAATGCTCATAGCTGGGAGTGCTCATCCCAATATTGGTACGTGAGCTGGGTTC
#14	......................TCATATTTGCATGCAGCTCACATGTTTGCAAGAAATTGGCACCTCGATGTCGGCTTGCCGATTTCAGAGCAGAAGGAATGCTCATAGCTGGGAGTGCTCATCCCAATATTGGTACGTGAGCTGGGTTC
#15	......................TCATATTTGCATGCAGCTCACATGTTTGCAAGAAATTGGCACCTCGATGTCGGCTTGCCTGATGGAGAGCAGAAGGAATGCTCTGCTGTGGGAGTGCTCATCCCAATATTGGTACGTGAGCTGGGTTC
#16	......................TCATATTTGCATGCAGCTCACATGTTTGCAAGAAATTGGCACCTCGATGTCGGCTTGCCGATTTCAGAGCAGAAGGAATGCTCATAGCTGGGAGTGCTCATCCCAATATTGGTACGTGAGCTGGGTTC
#17	......................TCATATTTGCATGCAGCTCACATGTTTGCAAGAAATTGGCACCTCGATGTCGGCTTGCCTGATTGAGAGCAGAAGGAATGCTCATAGCTGGGAGTGCTCATCCCAATATTGGTACGTGAGCTGGGTTC
#81	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#82	GTTACTGCAAGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#83	GTTACTGCA.GGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#84	GTTACTGCA.GGATAAACGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...
#85	GTTACTGCAGGGATAACAGGCTTCATATTTGCATGCAGCTCACATGTTTGCAAGAAA...

Variable / Insert / Delition / Substitution	
site	1650–1799
#01	....
#02	....
#03	....
#04	....
#05	....
#06	....
#07	....
#08	....
#09	....
#10	....
#11	AGAACGATGTGAATCAGTTTGGTCTATATCTTTAAGTGTTTGCA...TTGTTGTTGGTCGAATTTAGTACGCAAGGACCAATACGGAATGCTAAACCTGATTGTGGCATGTGAGTGGTATCACTCTTTCTAGATAAGTGATTATAAGACC
#12	AGAACGATGTGAATCAGTTTGGTCTATATCTTTAAGTGTTTGCTTCAGAGCTGTTGGTCGAATTTAGTACGCAAGGACCAATACGGAATGCTAAACCTGATTGTGGCATGTGAGTGGGATCACTCTTTCTAGATAAGTGATTATAAGACT
#13	AGAACGATGTGAATCAGTTTGGTCTATATCTTTAAGTGTTTGCATCAGAGCTGTTGGTCTCATTTAGTACGCAAGGACCAATACGGAATGCTAAACCTGATTGTGACATGTGAGTGGTATCACTCTTTCTAGATAAGTGATTATAAGACT
#14	AGAACGATGTGAATCAGTTTGGTCTATATCTTTAAGTGTTTGCATGA...TTGTTGGTCGAATTTAGTACGCAAGGACCAATACGGAATGCTAAACCTGATTGTGGCATGTGAGTGGTATCACTCTTTCTAGATAAGTGATTATAAGACT
#15	AGAACGATGTGAATCAGTTCGATCGATGTCTTTAAGTGTTTGCATCA..GTTGTTGGTCGAATGTAGTACG.AAGGACCAATACGGAATGCTAAACCTGATTGTGGCATGTGAGTGGGATCACTCTTTCTAGATAAGTGATTATAAGACT
#16	AGAACGATGTGAATCAGTTTGGTCTATATCTTTAAGTGTTTGCATGA...TTGTTGGTCGAATTTAGTACGCAAGGACCAATACGGAATGCTAAACCTGATTGTGGCATGTGAGTGGTATCACTCTTTCTAGATAAGTGATTATAAGACT
#17	AG..
#81	....
#82	....
#83	....
#84	....
#85	....

site	1800										1810										1820										1830										1840			
	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3
Variable																																												
Insert																																												
Delition																																												
Substitution																																												
#01	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#02	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#03	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#04	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#05	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#06	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#07	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#08	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#09	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#10	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#11	G	G	C	A	C	C	T	T	A	A	G	A	T	T	G	A	G	T	T	G	G	T	T	T	C	A	G	G	·	T	T	T	C	T	C	G	T	A	T	T	C	A	G	G
#12	G	G	C	A	C	C	T	T	A	A	G	A	T	T	C	·	·	·	·	G	G	T	T	T	C	A	G	G	·	T	T	T	C	T	C	G	T	A	T	T	C	A	G	G
#13	G	G	C	A	C	C	T	T	A	A	G	A	T	T	T	A	·	·	·	G	G	T	T	T	C	A	G	G	·	T	T	T	C	T	C	G	T	A	T	T	C	A	G	G
#14	G	G	C	A	C	C	T	T	A	A	G	A	T	T	G	A	G	T	T	G	G	T	T	T	C	A	G	G	G	T	·	·	·	·	·	·	·	·	T	T	C	A	G	G
#15	G	G	C	A	C	C	T	T	A	A	G	A	T	T	T	A	·	·	·	G	G	T	T	T	C	A	G	G	·	T	T	T	C	T	C	G	T	A	T	T	C	A	G	G
#16	G	G	C	A	C	C	T	T	A	A	G	A	T	T	G	A	G	T	T	G	G	T	T	T	C	A	G	G	G	T	·	·	·	·	·	·	·	·	T	T	C	A	G	G
#17	·	·	·	·	·	·	T	T	A	A	G	A	T	T	G	A	G	T	T	G	G	T	T	T	C	A	G	G	·	T	T	T	C	T	C	A	T	A	T	T	C	A	G	G
#81	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#82	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#83	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#84	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
#85	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·

Figure 3.3

Complete alignment of the non-coding regions of psbA and 23S rRNA minicircles. The yellow background indicates the tandem repeat sequence TCGGTGCTGTTTCAT. The pink background indicates the tandem repeat sequence TTTCAGGTTTG. The sky blue background indicates the tandem repeat sequence GGCATATTCTTATT. Core sequences are indicated with a purple background. pA, sequences from the *psbA* minicircle non-coding region; rR, sequences from the 23S rRNA gene minicircle non-coding region. Sequence numbers are common to Figure 3.1 and 3.2

site (0–149)	sequence
pA#51	TTTCA·GTGGTTTTGTAGG ...
pA#54	TTTCA·GTGGTTTTGTAGG ...
pA#18+	TTTCA·GTGGTTTTGTAGGTTTCAGGTTTGTCGGTGCTGTTTCATTTTCTCGTAT ... TCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCAT ... TTT
rR#01	TTGGGCATATTCTTATTGGGATTGGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTG ... TCGGTGCTGTTTCATTTTCAGGTTTG ... TCGGTGCTGTTTCATTTTCTCGTTT
rR#02	TTGGGCATATTCTTATTGGGC ... TTGTCGGTGCTGTTTCATTTTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCAT ... TTT
rR#03	TTGGGCA ...
rR#05	TTGGGCATATTCTTATTGGG ... ATTGGGCTTGTCGGTGCTGTTTCAT ... TTT
rR#07	...

site (150–299)	sequence
pA#51	... TTTCAGGTTTGTCGGTGCTGTTTCATTTTGGAGGTT ...
pA#54	...
pA#18+	CTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGAGCTGTTTCATTTACAGGTTTGTCGGTCGGTGCTGTTTCATTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTGGGCATA
rR#01	CTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTGGGAGG ... TTTCAGGTTTGTCGGTCGGTGCTGTTTCATTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATTT ...
rR#02	CTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTGTCGGTCGGTGCTGTTTCAT ... TCGTTTCAGGTTTGTCGGTGCTGTTTCATTTTCTCGTT ...
rR#03	... TATTTGATGTGTCGGAGTTG ...
rR#05	CTCGTATTCAGGTTTGGGCATATTCTTATTGGGATTG ... GGCTTGTCGGTCGGTGCTGTTTCATTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCAT ... TT ...
rR#07	... TTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATTTTCTCGTT ...

site (300–449)	sequence
pA#51	...
pA#54	...
pA#18+	TTCTTATTGGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTGGGCATATTCTTATTGGGATTGGGCTTGTCGGTGCTGTTCCATTTTCAGGTTTGGTCGGTGCTGTTTCATTTTCAGGTTTGTCGGTGCTGTTTCATTTTCTCGTGTTCAGG
rR#01	...
rR#02	...
rR#03	...
rR#05	...
rR#07	...

site (450–599)	sequence
pA#51	... TCTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTGGTAGATTC ... AGGTTTGTCGGTGCTGTTTCATTTTCAGGTTTG ...
pA#54	... TTTCAGGTTTG ...
pA#18+	TTTGGGCATATTCTTATTGGGATTGGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATATTCTTATT ... GGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTG ...
rR#01	... TCTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATTTTCAGTTTCTTTGGCTTTGTCGGTGCTGTTTCATTTTCAGGTTTGTCGGTGCTGTT
rR#02	... TCTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCAT ... TTTCAGGTTTG ...
rR#03	... CACATATCAGAATGAATTGGTGGAGG ... TTTCAGGTTTG ...
rR#05	... TCTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTG ... TCGGTGCTGTTTCATTTTCAGGTTTGTCGGTGCTGTT
rR#07	... TCTCGTATTCAGGTTTGGGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATTTTCAGTTTCT ... TTGGC ...

site (600–749)	sequence
pA#51	... TCGGTGCTGTTTCAT ... TTTCAGGTTTGGGCATATTCTTATT ... GGGATTGGGCTTATCGGTGCTGTTTCATTTTCAGGTTTG ...
pA#54	... TCGGTGCTGTTTCATTT ... TCTCGTATTCAGGTTTGGGCATATTCTTATT ... GGGC ...
pA#18+	... TCGGTGCTGTTTCATTT ... TCTCGTATTCAGGTTTGGGCATATTCTTATT ... GGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTG ...
rR#01	TCATTTTGAATTGTCGGTGCTGTTTCATTTTGAATTGTCGGTGCTGTTTCATTTGAATTGTCGGTGCTGTTTCATTTTGAATTGTCGGTGCTGTTTCATTTTCTTCAGGTTTCAGGTTTGTCGGTGCTGTTTCATTTT··GAATTGTCG
rR#02	... TCGGTGCTGTTTCATTT ... TCTCGTATTCAGGTTTGGGCATATTCTTATT ... GGGCTTGTCGGTGCTGTTTCGTTTTCAGGTTTGTCG
rR#03	... TCGGTGCTGTTTCAT ... TTTCAGGTTTGTCGGTGCTGTTTCATTTT··GAATTGTCG
rR#05	TCATTTTGAATTGTCGGTGCTGTTTCATTT ... TCTCGTATTCAGGTTTGGGCATATTCTTATT ... GGGCTGG ...
rR#07	... TTTGTCGGTGCTGTTTCATTT ... TCTCGTATTCAGGTTTGGGCATATTCTTATT ... GGGATTG ...

site (750–899)	sequence
pA#51	... TCGGTGCTGTTTCATTTTCAGGTTTGTCGGTGCTGTTTCATTTT ... CTCGTATTCAGGT·TTGGGCATATTCTTATTGGGCTTGT
pA#54	... TTGTCGGTGCTGTTTCATTTT ... CTCGTATTCAGGT·TTGGGCATATTCTTATTGGGCTGG·
pA#18+	... GGCATATTCTTATTGGGCTTGTCGGTGCTGTTTCATTTTCAGGTTTGTCGGTGCTGTTTCATTTT ... CTCGTATTCAGGT·TTGGGCATATTCTTATTGGGCTGG·
rR#01	GTGCTGTTTCATTTTCTTCAGGTTTCAGGTTTGTCGGTGCTGTTTCATTTTCAGGTTTGTCGGTGCTGTTTCATTTTCAGGTTTGTCGGTGCTGTTTCATTTT·CTCGTTTCTCGTATTCAGGT·TTGGGCATATTCTTATTGGGCTGG·
rR#02	GTGCTGTTTCAT ... TTTCAGGTTTGTCGGTGCTGTTTCATTTTCAGGTTTGTCGGTGTTGTTTCATTTTCAGGTTTGTCGGTGCTGTTTCATTTT·CTCGTTTCTCGTATTCAGGTGCAGGGCATATTCTTATTGGGC ...
rR#03	GTGCTGTTTCATTTT ... GAATTGTCGGTGCTGTTTCATTTT··GAATTGTCGGTGCTGTTTCATTTT··GAATTGTCGGTGCTGTTTCATTTGGAGGTTTCTCGTATTCAGGT·TTGGGCATATTCTTATTGGGCTGG·
rR#05	... TTCACATTCACTTCGTTTCGCTCGCTTGAAGG ... TTT·CTCGTTTCTCGTATTCAGGT·TTGGGCATATTCTTATTGGGC ...
rR#07	... GGCTTGTCGGTGCTGTTTCATTTTCAGGTTTGTCGGTGCTGTTTCATTTT·CTCGTTTCTCGTATTCAGGT·TTCGGCATATTCTTATTGGGC ...

site	900	910	920	930	940	950	960	970	980	990	1000	1010	1020	1030	1040
pA#51	CGGTGCTGTT	TCATTTTCTC	GTTTCTCGTA	TTCAGGGTTT	CTTTGGCTTT	CAGTCGAC·A	TTTTCTCGAA	GAAACTTTGT	ATTTGTGAAC	AGGTTTTGGA	CCGATTTAGG	GCTTGCGCAG	GTGTGATGTG	CTGGTGTGAT	GCGCATGCGC
pA#54	··········	··········	··········	TGCAGGGTTT	CTTTGGCTTT	CAGTCGATAA	AACGGCCGAA	GAAACTTTGC	TTGTAACATC	AGGTTTTGGA	CCGATTTTCG	GCTTGCGCAG	GTGTGCTCGT	GT·····GAT	GGGCCTGCGC
pA#18+	··········	··········	··········	TGCAGGGTTT	CTTTGGCTTT	CAGTCGATAA	AACGGCCGAA	GAAACTTTGC	TTGTAACATC	AGGTTTTGGA	CCGATTTTCG	GCTTGCGCAG	GTGTGCTCGT	GT·····GAT	GGGCCTGCGC
rR#01	··········	··········	··········	TGCAGGGTTT	CTTTGGCTTT	CAGTCGATAT	TTAGGCCGAA	GAAACTTTGC	TTGTAACAAC	AGGTTTTGGA	CCGATTTAGG	GCTTGCGCAG	GTGTGCTCGT	GT·····GAT	GGGCCTGCGC
rR#02	··········	··········	··········	··AAGGGTTT	CTTTGGCTTT	CAGTCGAC·A	TTTTCTCGAA	GAAACTTTGT	ATTTAGCATC	AGGTTTTGGA	CCGATTTTCG	GCTTGCGCAG	GTGTGCTCGT	GT·····GAT	GCGCATGCGC
rR#03	··········	··········	··········	TGCAGGGTTT	CTTTGGCTTT	CAGTCGATCA	TTCGTCCGAA	GAAACTTTGC	TTGTAACAAC	AGGTTTTGGA	CCGATTTAGG	GCTTGCGCAG	GTGTGCTCGT	GT·····GAT	GGGCCTGCGC
rR#05	··········	··········	··········	···AGGGTTT	CTTTGGCTTT	CAGTCGATAA	AACGGCCGAA	GAAACTTTGC	TTGTAACAAC	AGGTTTTGGA	CCTCTTTTCG	GCTTGCGCAG	GTGTAATGTG	CC·······T	GCTCATGCGC
rR#07	··········	··········	··········	···AGGGTTT	CTTTGGCTTT	CAGTCGATAT	TTCGTCCGAA	GAAACTTTGC	TTGTAACAAC	AGGTTTTGGA	CCTATTTTCG	GCTTGCGCAG	GTGTAATGTG	CC·······T	GCTCATGCGC

site	1050	1060	1070	1080	1090	1100	1110	1120	1130	1140	1150	1160	1170	1180	1190
pA#51	AAGCCTTGTT	TGGGCCTAAT	TCGTAATTGA	TGAACAGAAG	TTTCTTCGGA	GAATTTTGG·	ATTTATCAAT	AGCTGTTGG·	·CTGGGTCAG	GGCATATTCG	TATTCTTCTT	TGTTCAAGAA	ATGTTATAAC	·AGTTGGGCA	CTTTGTGTTC
pA#54	TAGCCGAAAA	ACAGCCTAAT	TCGTAATTGA	TTCACAGAAG	TTTCTTCGGG	GA·TTTCGG·	GTTT······	··CTCGTAT·	·TCAGGTTTG	GGCATATTCT	TATT······	··········	··········	·····GGGCT	TGAGGTTTCA
pA#18+	TAGCCGAAAA	ACAGCCTAAT	TCGTAATTGA	TGAACAGAAG	TTTCTTCGGG	GA·TTTCAG·	GTTT······	··CTCGTAT·	·TCAGGTTTG	GGCATATTCT	TATT······	··········	··········	··········	··········
rR#01	TAGCCGAAAA	ACAGCCTAAT	TCGTAATTGA	TGAACAGAAG	TTTCTTCGGA	G·········	··········	··········	····GTTTAG	GGCATATTCG	TATTCGTAAA	TTCTAGTTTT	AT········	···CTTTGGG	TGGTTTTGTA
rR#02	AAGCCTTGTT	TGGGCCGTTT	TCGTAATTGA	TGAACAGAAG	TTTCTTCGGA	CGATTTTAGA	AT·TTCGTAT	TC········	··········	··········	·····GTAAA	TTCTAGTTTT	ATCTTTGGT·	··········	TGGTTTTGTA
rR#03	TAGCCGAAAA	ACAGCCTAAT	TGGTAATTGA	TGAACAGAAG	TTTCTTCGGA	GAATTTTGG·	ATTTTCGTAT	TC········	··········	··········	·····GTAAA	TTCTAGTTTT	ATCTTTGGTT	TATCTTTGGG	TGGTTTTGTA
rR#05	AAGCCTTGTT	TCAGCCGTTT	TCGTAATTGA	TTCACAGAAG	TTTCTTCGGA	GAATTTTGG·	ATTTTCGTAT	TC········	··········	··········	·····GTGAA	TTCTAGTTTT	AT········	···CTTTGGG	TGGTTTTGTA
rR#07	AAGCCTTGTT	TCAGCCTAAT	TCGTAATTGA	TGAACAGAAG	TTTCTTCGGA	GAATTTTGG·	ATTTTCGTAT	TC········	··········	··········	·····GTAAA	TTCTCGTTTT	AT········	···CTTTGGG	TGGTTTTGTA

site	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300	1310	1320	1330	1340
pA#51	AGCATATTCA	ATCGGTGGTG	GTTTCAGGTT	TGTCGGTGCT	GTTTCATTTT	CAGGTTTGTC	GGTGCTGTTT	CATTTTCAGT	TTCT·TAGGT	TTGTCGGTGC	TGTTTCATTT	TCAGGTTTGT	CGGTGCTGTT	TCATTTTCTC	GTTTCTCGTA
pA#54	GGTTTGTCGG	TGCTGTTTCA	TTTTCAGGTT	TGTCGGTGCT	GTTTCATTTT	CAGGTTTGTC	GGTGCTGTTT	CATTTTCAGT	TTCTTTGGCT	TTGTCGGTGC	TGTTTCATTT	TCAGGTTTGT	CGGTGCTGTT	TCATTTTCTC	GTTTCTCGTA
pA#18+	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··GGGCTTGT	CGGTGCTGTT	TCATTTTCTC	GTTTCTCGTA
rR#01	GGTTNTGG··	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········
rR#02	GGTTTTAG··	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········
rR#03	GGTTTT·G··	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········
rR#05	GGTTTTGG··	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········
rR#07	GGTTTT·G··	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········	··········

site	1350	1360	1370	1380	1390	1400	1410	1420	1430	1440	1450	1460	1470	1480	1490
pA#51	TTCAGGTTTG	GGCATATTCT	TATTGGGCTT	TATAGGTTTT	GGGTCGTTTT	CATTCACTTC	GTTTCGCTCG	CTTCAAGGTT	TTATTCGTCG	TATTCTTATT	CGTAGGTAGA	TATTCTTATT	CGTTGCATCC	AATCAATCAG	CAGGCCACTT
pA#54	TTCAGGTTTG	GGCATATTCT	TATTGGGCTT	TATAGGTTTT	GGGTCGTTTT	CATTCACTTC	GTTTCGCTCA	CTTGAAGGTT	TTATTCGTCG	TATTCATAAT	CGGAGGTAGA	TATTCTTATT	CGT···ATAG	CATCAATCAG	CAGGCCACTT
pA#18+	TTCAGGTTTG	GGCATATTCT	TATTGGGCTT	TATAGGTTTT	GGGTCGTTTT	CATTCACTTC	GTTTCGCTCG	CTTCAAGGTT	TTATTCGTCG	TATTCATAAT	CTGAGGTAGA	TATTCTTATT	CGT···ATAG	CATCAATCAG	CAGGCCACTT
rR#01	··········	··········	··········	··········	··········	··········	··········	··········	·········G	GCATATTCGT	ATTCTTATTT	GATCAAGA··	··········	··········	··········
rR#02	··········	··········	··········	··TTGGTTTG	GGGTCGTTTT	CATTCACTTC	GTTTCGCTCG	CTTGAATAAT	TTAGGGTTTG	GCATATTCGT	ATTCTTCTTC	GATCAAGG··	··········	··········	··········
rR#03	··········	··········	··········	··········	··········	··········	··········	··········	········TG	GCATATTCGT	ATTCTTATTT	GATCAAGA··	··········	··········	··········
rR#05	··········	··········	··········	··········	··········	··········	··········	··········	·········T	GCATATTCGT	ATTCTTATTT	GATCAAGA··	··········	··········	··········
rR#07	··········	··········	··········	··········	··········	··········	··········	··········	········TG	GCATATTCGT	ATTCTTATTT	GATCAAGA··	··········	··········	··········

site	1500	1510
pA#51	TCAACCAACT	TC·
pA#54	TCAACCAACT	TC·
pA#18+	TCAACCAACT	TC·
rR#01	··········	···
rR#02	··········	···
rR#03	··········	···
rR#05	··········	···
rR#07	··········	···