Innate selfish characteristics of intronic homing enzyme in *Chlamydomonas* mitochondria assayed through *in vitro* and *in vivo* systems

Sayuri Kurokawa

A dissertation submitted to
Kochi University of Technology
in partial fulfillment of the requirements
for the degree of

Doctor of Engineering

Engineering Course
Department of Engineering
Graduate School of Engineering
Kochi University of Technology
Kochi, Japan

March 2006

General introduction

Homing enzyme is a kind of DNA sequence specific endonuclease of which recognition sequence is much longer than bacterial type II-restriction endonucleases. Homing enzymes are classified into four groups based on the conserved amino acid residues, LAGLIDADG, GIY-YIG, N-H-N or His-Cys box. LAGLIDADG family is the largest among them (for reviews see, Belfort and Perlmans, 1995; Belfort and Roberts, 1997; Chevalier and Stoddard, 2001). Most often they are coded in an ORF within the intron or found as a component of an intein, but some are encoded in a genome as a freestanding gene. Most probably homing enzymes detected in an intein or coded in an intronic ORF are originally freestanding ones, and translocated into the existent sites in the course of evolution. About 30 % of the group I introns contain the ORF for active homing enzyme, and the others contain vestigial sequences for homing enzymes (for a review see, Chevalier and Stoddard, 2001). This suggests that intronic homing enzyme easily loses its activity through accumulation of mutations. Most probably inactivation of homing enzymes is tolerated because they have no benefits to the host, and this instantly leads to that regular horizontal transmission of the gene to a new species before its functional deterioration is essential for its evolutionary long-term persistence. On the other hand, it has been shown that to make a copy of the homing enzyme gene, cleavage of the target site is essential. Therefore, the recognition property of a homing enzyme must be skillfully adapted to cleave the target sequence which is degenerated among a wide range of organisms.

So far, characteristics of homing enzymes have been analyzed *in vitro*, and none has been analyzed *in vivo*. This is mainly due to the difficulty to develop *in vivo* assay systems. However, my colleagues and I developed a practical method to introduce a point mutation into the intended site of the *Chlamydomonas reinhardtii* (*C. reinhardtii*) mitochondrial gene (Yamasaki *et al.,* 2005). Utilizing this method, I succeeded to assay the property of a kind of LAGLIDADG homing enzyme "I-*Csm*I" *in vivo*. I-*Csm*I is coded in the alpha-group I-intron located in the COB gene of *Chlamydomonas smithii* (*C. smithii*) mitochondrial genome (Colleaux *et al.,* 1990).

In Chapter 1, I showed the results obtained through *in vitro* assay. Basic

property of I-*Csm*I and the cleavage efficiency on variously mutated target sequences were assayed using a recombinant I-*Csm*I polypeptide and substrate DNA fragments. In Chapter 2, I showed the results obtained by *in vivo* assay. Cleavage ability of I-*Csm*I was monitored through alpha-intron invasion into the variously mutated target site in *C. reinhardtii* mitochondrial genome.

Contents

Chapter 1

Adaptation of intronic homing enzyme for successful horizontal transmission

Abstract

Group I introns are thought to be self-propagating mobile elements, and are distributed over a wide range of organisms through horizontal transmission. Intron invasion is Íinitiated through cleavage of a target DNA by a homing enzyme encoded in an ORF found within the intron. The intron is likely of no benefit to the host cell and is not maintained over time, leading to the accumulation of mutations after intron invasion. Therefore, regular invasional transmission of the intron to a new species at least once before its degeneration is likely essential for its evolutionary long-term existence. In many cases, the target is in a protein coding region which is well conserved among organisms, but contains ambiguity at the 3rd nucleotide position of the codon. Consequently, the homing enzyme might be adapted to overcome sequence polymorphisms at the target site.

To address whether codon degeneracy affects horizontal transmission, I investigated the recognition properties of a homing enzyme, I-*Csm*I, that is encoded in the intronic ORF of a group I intron located in the mitochondrial COB gene of the unicellular green alga *Chlamydomonas smithii*. I successfully expressed and purified three types of N-terminally truncated I-*Csm*I polypeptides, and assayed the efficiency of cleavage for 81 substrates containing single nucleotide substitutions. I found a slight but significant tendency that I-*Csm*I cleaves substrates containing a silent or tolerated amino acid change more efficiently than non-silent or non-tolerated ones. The published recognition properties of I-*Spom*I, I-*Sca*I and I-*Sce*II were reconsidered from this point of view, and I detected proficient adaptation of I-*Spom*I, I-*Sca*I and I-*Sce*II for target site sequence degeneracy. Based on the results described above, I propose that intronic homing enzymes are adapted to cleave sequences that might appear at the target region in various species, however such adaptation becomes less prominent in proportion to the time elapsed after intron invasion into a new host.

Introduction

Various molecular phylogenetic analyses suggest that group I introns in fungi and terrestrial/non-aquatic plants were horizontally transmitted multiple times in the course of evolution among distantly related species (Lambowitz, 1989; Belfort and Roberts, 1997; Cho *et al*., 1998). I have shown this is also the case for algal mitochondrial introns (Watanabe *et al*., 1998; Ehara *et al*., 2000). For reasons yet unknown, the distribution of group I introns is strongly biased, most commonly found in fungi [*e.g*., the COXI of *Podospora anserina* contains 15 group I introns (Cummings *et al*., 1990). About 30 % of group I introns contain an ORF that encodes a DNA sequence specific endonuclease (intronic homing enzyme). These intronic homing enzymes cleave a target sequence that is usually 16-30 base pairs (bp) long and non-palindromic (for a review, Chevalier and Stoddard, 2001). Cleavage of the chromosome initiates repair of the damaged DNA through homologous recombination. Consequently, after the repair, the donor intronic DNA is copied into the recipient chromosome. Thus, homing enzymes are essential for horizontal transmission of group I introns. Organelle introns are highly likely of no benefit to the host, *i.e*., they are thought to be selfish and parasitic elements that spread in populations. Therefore, when they integrate into the host genome, there is little or no selection for maintaining endonuclease function. Moreover, if there is any cost to the host cell for producing a functional endonuclease, then selection will work to fix the nonfunctional element. Therefore, regular horizontal transmission of an intron to a new species before its functional deterioration seems essential for its evolutionary long-term persistence. As an example, comprehensive analyses of the group I intron omega (also known as *Sc* LSU.1), which was first found in the *Saccharomyces cerevisiae* mitochondrial large subunit rRNA gene, clearly showed repeated horizontal transmissions, and the interval between the complete loss and reinvasion of the intron is estimated to be about 5.7 million years (Goddard and Burt, 1999). This leads to the hypothesis that intronic homing enzymes might be adapted to recognize variously degenerated target sequences among a wide range of organisms.

In addition to intronic homing enzymes, highly specific endonuclease activity is also detected among inteins, which are thought to be parasitic elements that exhibit horizontal transmission. Regular invasional transmission is likely essential for both

homing introns and inteins. In fact, for the target site of intein homing enzyme PI-*Sce*I, which is found in *Saccharomyces cerevisiae* vacuolar membrane H^+-ATPase, all of the nine nucleotides essential for the cleavage were mapped on the conserved codon first and second positions, and target sequence variations at codon third positions were tolerated for the endonuclease recognition (Gimble, 2001). On the other hand, the adaptations that permit efficient horizontal transfer of intronic homing enzymes have not been analyzed. To date, only three intronic homing enzymes that target a sequence within protein coding genes were investigated systematically for their recognition sequence ambiguity, *i.e*., I-*Spom*I that is encoded as an intronic ORF of a group I intron in the *Schizosaccharomyces pombe* COXI gene (Schafer *et al*., 1991; Pellenz *et al*., 2002), I-*Sca*I is in the COB gene of *Saccharomyces capensis* (Szczepanek and Lazowska, 1996; Monteilhet *et al*., 2000), and I-*Sce*II is in the COXI gene of *Saccharomyces cerevisiae* (Hanson *et al*., 1982; Sargueil *et al*., 1990; Wernette *et al*., 1992). To address the question, I investigated the recognition sequence of I-*Csm*I, including its degeneracy. I-*Csm*I is a homing enzyme encoded in the group I intron (named alpha or *Cs cob*.1) located in the apocytochrome b (COB) gene of the unicellular alga *Chlamydomonas smithii* (*C. smithii*) (Colleaux *et al*., 1990). This enzyme has the typical two LAGLIDADG motifs. The intronic ORF is probably translated as a fusion protein with the preceding exon, and the N-terminal peptide may be proteolytically removed to become an active form as seen in I-*Spom*I (Schafer *et al*., 1994). Enzyme activity of I-*Csm*I has been observed through artificial interspecific cell fusion between intron-bearing *C. smithii* and *C. reinhardtii* that lacks the intron in its COB gene (Remacle *et al*., 1990). However, systematic analysis of the target sequences and the homing enzyme's enzymatic properties have not been previously attempted. I overproduced several N-terminally truncated I-*Csm*I polypeptides in *E. coli*, and determined cleavable target sequences through an *in vitro* assay of substrates containing 81 different point mutations.

Based on the analyses of I-*Csm*I and these three intronic homing enzymes, I discuss the adaptation for successful horizontal transfer. Investigations performed for the intronic homing enzymes that have a recognition sequence in ribosomal RNA genes are less informative to answer our questions and are not considered in this paper.

Materials and Methods

Cloning and expression of WT and N-terminally truncated I-*Csm*I ORFs

The entire COB gene and the alpha-intron were amplified by PCR using total *Chlamydomonas smithii* (CC-1373) DNA as a template. I also used PCR to isolate the wild-type (WT) 374 amino acids I-*Csm*I ORF (*i.e.*, ORF(374)) and three N-terminally truncated ORFs, ORF(200), ORF(217) and ORF(237) (the number in parentheses indicates the amino acid encoded in the ORF). These four ORFs have different N-termini, however share the common WT stop codon. The two sets of primers used to amplify the original I-*Csm*I ORF(374) and ORF(237), contained *Xho*I sites at their tails. Forward primer containing an *Nde*I site, and reverse primers containing an *Fba*I site were used to amplify ORF(200) and ORF(217). After restriction enzyme digestion, the ORF(374) and ORF(237) PCR products were cloned into the *Xho*I site of pET19b (Novagen, CA, USA) in frame with a sequence encoding the 10-histidine tag, while ORF(200) and ORF(217) were cloned into the *Nde*I/*Bam*HI site of pET15b (Novagen, CA, USA) in-frame with a 6-histidine tag. The resulting plasmids were amplified in *E. coli* DH5/alpha, and *E. coli* BL21 CodonPlus (DE3) RIL (Stratagene, CA, USA) was for protein expression.

Expression and purification of whole or truncated I-*Csm*I polypeptides

Cultures containing whole or truncated ORFs were undertaken at 37 ˚C in 2.0 liters of LB broth containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol to an $OD_{600\ nm}$=0.6. Protein expression was induced by addition of isopropyl-thio-ß-D-galactopyranoside (0.1 mM final). The cells were incubated at 30 ˚C for an additional 4 hrs, collected by centrifugation, and resuspended in 40 ml of sonication buffer [50 mM HEPES (pH 7.0), 400 mM NaCl, 6 mM ß-mercaptoethanol, and 20 μg/ml lysozyme] and sonicated on ice. The lysate was centrifuged for 2 hrs at 10,000 g and the supernatant was loaded onto a Ni-NTA column (5 ml bed volume) (Qiagen, CA, USA) that was previously equilibrated with the wash buffer [50 mM HEPES (pH 7.0), 400 mM NaCl, 6 mM ß-mercaptoethanol, and 10 mM imidazole]. The column was washed with 50 ml of the wash buffer, and the protein was eluted with 100 ml of the elution buffer [50 mM HEPES (pH 7.0), 400 mM NaCl, 6 mM ß-mercaptoethanol, and 200 mM imidazole].

Homogeneity was assessed after staining with SDS-PAGE/comassie brilliant blue R-250. The products of ORF(200), ORF(217), ORF(237) and ORF(374) were named I-*Csm*I(200), I-*Csm*I(217), I-*Csm*I(237) and I-*Csm*I(374), respectively.

Reaction conditions to estimate the minimum target-site length

(Substrate DNA)

Chemically synthesized DNA fragments, which consist of 18, 20 or 24 nt symmetrically spanning the alpha-intron insertion point of the *C. reinhardtii* COB gene, were cloned into the *Eco*RV site of the pCITE-4a+ (Novagen, CA, USA). These plasmids were named pC-18nt, pC-20nt and pC-24nt (the number indicates the length of the inserted DNA fragment). The plasmids were first linearized by *Sca*I digestion, and then used as a substrate to determine the region encompassing the recognition sequence.

(Reaction conditions)

1.5 μg of linearized substrate described above was added to 50 μl of the reaction mixture containing [50 mM HEPES (pH 7.0), 0.01 % bovine serum albumin (BSA), 1 mM dithiothreitol (DTT), 25 mM NaCl, and 5 mM $MgCl_2$] and about 1 μg of recombinant homing enzyme I-*Csm*I(237). The reaction was carried out at 25 ˚C for 24 hrs and 10 μl was loaded onto a 0.8 % agarose gel to resolve the products.

Reaction to determine the cleavage point and its terminal shape

I determined the terminal shape of the substrate following the T4 DNA polymerase method by Nishioka *et al.* (1998). pC-24nt (2.0 μg) digested with I-*Csm*I(237) was recovered from a 0.8 % agarose gel by electro-elution and then treated with T4 DNA polymerase (Takara Bio, Kyoto, Japan) in the presence of 0.2 mM dNTPs. The DNA mixture was then treated with T4 DNA ligase (Takara Bio, Kyoto, Japan) for self-ligation and transformed into *E. coli*. Nucleotide sequence analysis of the plasmid was performed to determine the nature of cohesive termini generated by I-*Csm*I(200).

Reaction conditions used to investigate the effect of Na^+, divalent cations, pH and temperature

(Substrate DNA fragment)

A 1.8 kb DNA fragment, containing the entire COB gene of *C. reinhardtii* (CC-124) and

its flanking regions, was cloned into pT7-Blue2 vector (Novagen, CA, USA) and named pCOB1.8Kb. After linearization by *Not*I, the plasmid was used as a substrate for the reaction described below.

(Reaction mixture)

A 50 μl reaction mixture [25 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.01 % BSA, 50 mM Tris-HCl (pH 7.0)] was used, which contained 0.5 μg of linearized pCOB1.8Kb and 1.0 μg of I-C*sm*I(217), or 1.5 μg of I-C*sm*I(200) or I-*Csm*I(237). One of the parameters [*i.e.*, pH, NaCl concentration, species of divalent cations (5 mM), $MgCl_2$ concentration, or the temperature] in the reaction was altered to determine optimal conditions. Reagents used to make the buffers of specific pH value are as follows; MES for pH 6.0, HEPES for pH 7.0, Tris for pH 8.0 and 9.0, TAPS for pH 10.0. The reaction was incubated for 24 hrs with I-*Csm*I(237) and I-*Csm*I(200), and incubated for 6 hrs with I-*Csm*I(217), which reduced the formation of aggregates observed with this protein. The reaction products were resolved in a 0.8 % agarose gel, and stained with ethidium bromide. The relative quantities of the digested fragments were calculated using the NIH Image program version 1.61.

Assay of cleavable DNA sequences

A limited part of the *C. reinhardtii* COB gene, which is 104 nt long and containing the I-*Csm*I target sequence, was chemically synthesized and converted to double strand DNA. This double-stranded DNA fragment was used as a control to compare the cleavage efficiency of various substrates containing single mutations. Each one of the 27 nucleotides composing the target site was changed to the other three possible nucleotides utilizing PCR primers containing a specific mutation. These 81 DNA fragments, each containing single point mutations were used for a detailed analysis of substrate cleavage. 150 ng of each substrate was digested with 1 μg of I-*Csm*I(200) in the reaction mixture [50 mM HEPES (pH 7.0), 0.01 % BSA, 1 mM DTT, 25 mM NaCl, 5 mM $MgCl_2$] at 30 ˚C for 8 hrs. Electrophoresis of the samples was performed on a 3 % agarose gel, and stained by 10,000-fold diluted SYBR Green I dye (Molecular Probes, OR, USA) for 40 min (SDS/heat-treatment of samples before electrophoresis, described below, was omitted for a clearer image, without affecting the results). The image was developed using LAS-1000 image analyzer (Fuji Film Co., Tokyo, Japan). The cleavage ratio, *i.e.*, cleaved *vs.*

uncleaved fragments, was quantified by NIH Image and compared to WT substrate cleavage (*i.e.*, native *C. reinhardtii* COB sequence carrying substrate). The 81 substrates were grouped into 4 classes based on the following: i) The substrate much better than the control (the cleavage ratio of mutated substrate *vs.* control is more than 1.5) is denoted as +++; ii) The substrate as good as the control (*i.e.*, the ratio is between 1.2 and 0.8) is denoted as ++; iii) The substrate less efficiently cleaved (*i.e.*, the ratio is between 0.5 and 0.2) is denoted as +; iv) Scarcely cleaved substrate (*i.e.*, the ratio is below 0.1) is denoted as -.

Reaction conditions to measure the kinetic parameters

Linearized pCOB1.8Kb and a plasmid containing the N-terminally truncated homing enzyme, I-*Csm*I(200), was used to measure the kinetic parameters. 250 µl of reaction buffer [50 mM HEPES (pH 7.0), 0.01 % BSA, 1 mM DTT, 25 mM NaCl, and 5 mM $MgCl_2$] contained 1 µg of the recombinant protein and between 0.5 ng/µl and 10 ng/µl of substrate. 20 µl aliquots were removed at different time points from the reaction mixture, and terminated by the addition of 1 µl of 0.5 M EDTA and 1.25 µl of 10 % SDS, followed by heating the mixture to 50 ˚C for 5 min to completely denature the protein. Samples were electrophoresed on a 0.8 % agarose gel, then visualized by 10,000-fold diluted SYBR Green I dye. Relative intensities of the digested fragment were quantified using the Las-1000 and NIH Image. Km, Vmax and kcat were determined through a Hanes-Woolf's plot (Hanes, 1932).

Results

Activity of the N-terminal truncated I-*Csm*I polypeptides

Three N-terminally truncated I-*Csm*I polypeptides [I-*Csm*I(200), I-*Csm*I(217) and I-*Csm*I(237); the number in parentheses indicates the amino acid encoded in the ORF] were purified and yielded about 6 mg of protein per 1 g wet weight *E. coli*, while the entire I-*Csm*I ORF (*i.e.*, I-*Csm*I(374)), which contains the upstream COB exon, did not express even after several modified conditions were tested (Fig. 1-1). I assayed the enzyme activity of recombinant I-*Csm*I(200), I-*Csm*I(217) and I-*Csm*I(237) using linearized

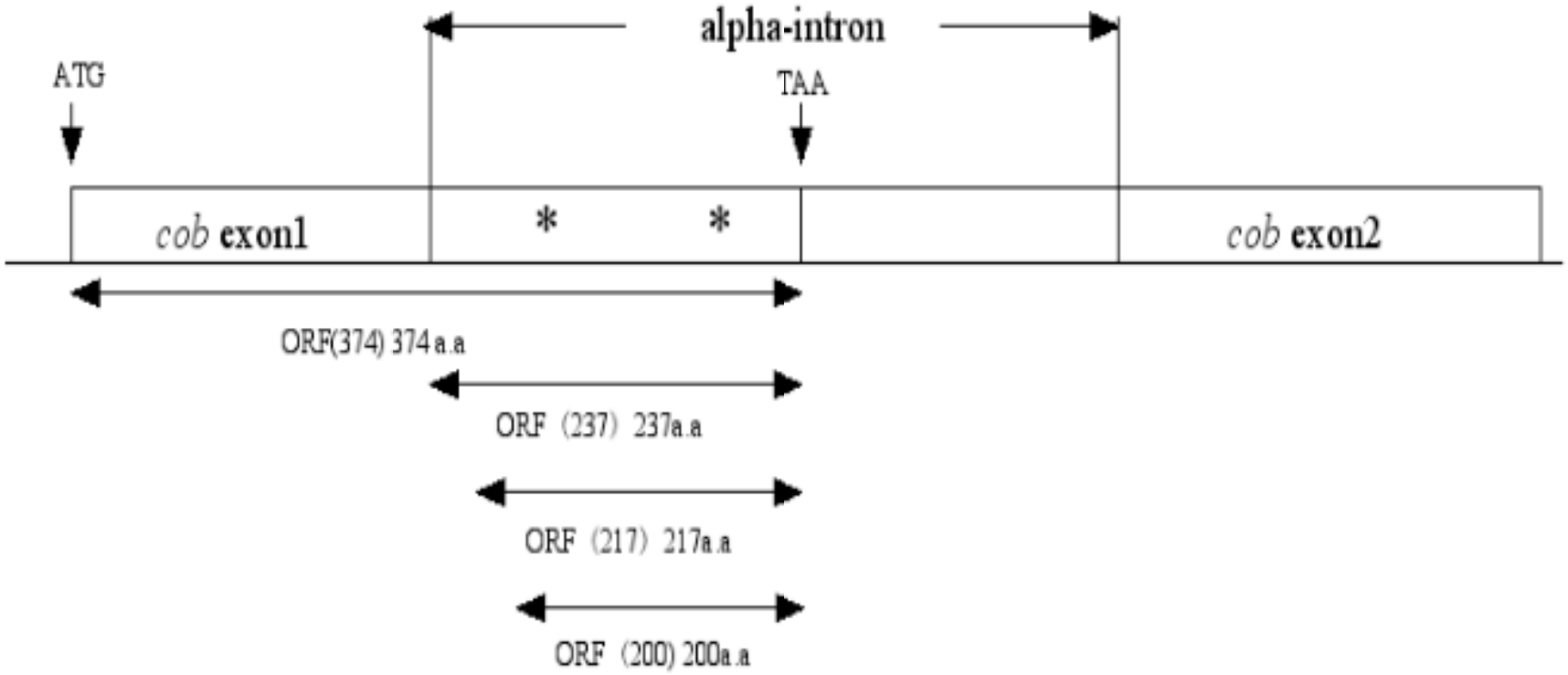


Fig. 1-1. Schematic of open reading frames that code whole I-*Csm*I **or N-terminally truncated** I-*Csm*I **polypeptides.**

I-*Csm*I is denoted as a fusion protein with the preceding apocytochrome b gene exon encoded polypeptide. Asterisks show the position of the LAGLIDADG motifs. a.a; amino acid residues

pCOB1.8Kb as a substrate. I-*Csm*I(200) is the smallest homing enzyme containing two LAGLIDADG motifs analyzed to date. It is even smaller than the type II restriction enzyme *Eco*RI (Newman *et al.,* 1981), which is a 277 amino acids homodimer that cleaves a symmetric 6 base restriction site. Recombinant proteins I-*Csm*I(200) and I-*Csm*I(237) cleaved the substrate at the expected target site, yielding two fragments of 1.2 kb and 3.7 kb in size. For I-*Csm*I(217), the quantity of protein was reduced from 1.5 to 1.0 μg and the incubation period was shortened from 24 to 6 hrs to reduce the amount of insoluble reaction products. Under these modified conditions, I-*Csm*I(217) exhibited sequence specific enzyme activity.

To determine the optimal conditions for enzyme activity, I tested the effect of Na^{+} and Mg^{2+} concentration, pH and temperature. The optimal pH for all three proteins was around 7.0 (Fig. 1-2A). The optimal Na^{+} and Mg^{2+} concentrations were 25 mM and 5 mM, respectively, for both I-*Csm*I(237) and I-*Csm*I(217) (Fig. 1-2B, 1-2C). In contrast, 75 mM Na^{+} and 10 mM Mg^{2+} were optimal for I-*Csm*I(200). The optimal reaction temperature was 35 ˚C for both I-*Csm*I(200) and I-*Csm*I(237), and 30 ˚C for I-*Csm*I(217) (Fig. 1-2D). A higher concentration of Mg^{2+} was progressively detrimental to all I-*Csm*I polypeptides. The presence of Mg^{2+} was essential for the endonuclease activity as a cofactor, while the same concentration of Mn^{2+} (5 mM) reduced the enzyme activity to 15 %, and no activity was observed with 5 mM of Zn^{2+}, Ca^{2+} or Co^{2+} (data not shown).

Kinetic parameters of I-*Csm*I(200)

I determined the kinetic parameters of I-*Csm*I(200) based on the data obtained by time course monitoring of the cleaved products in various concentrations of the linearized substrate pCOB1.8 Kb. The Km, Vmax, kcat were 2.5 X 10^{-9} M, 1.8 X 10^{-12} M/s and 4.7 X 10^{-4}/s, respectively. These parameters were similar to other representative intronic LAGLIDADG homing enzymes [*e.g.*, I-*Ceu*I (Turmel *et al.,* 1997), I-*Sce*IV (Wernette, 1998), I-*Dmo*I (Dalgaard *et al.,* 1994; Aagaard *et al.,* 1997)] that show characteristics of high affinity to the substrate DNA and slow turnover (Table 1-1).

Essential target region

Digestion was not observed using pC-18nt and pC-20nt, while almost complete cleavage

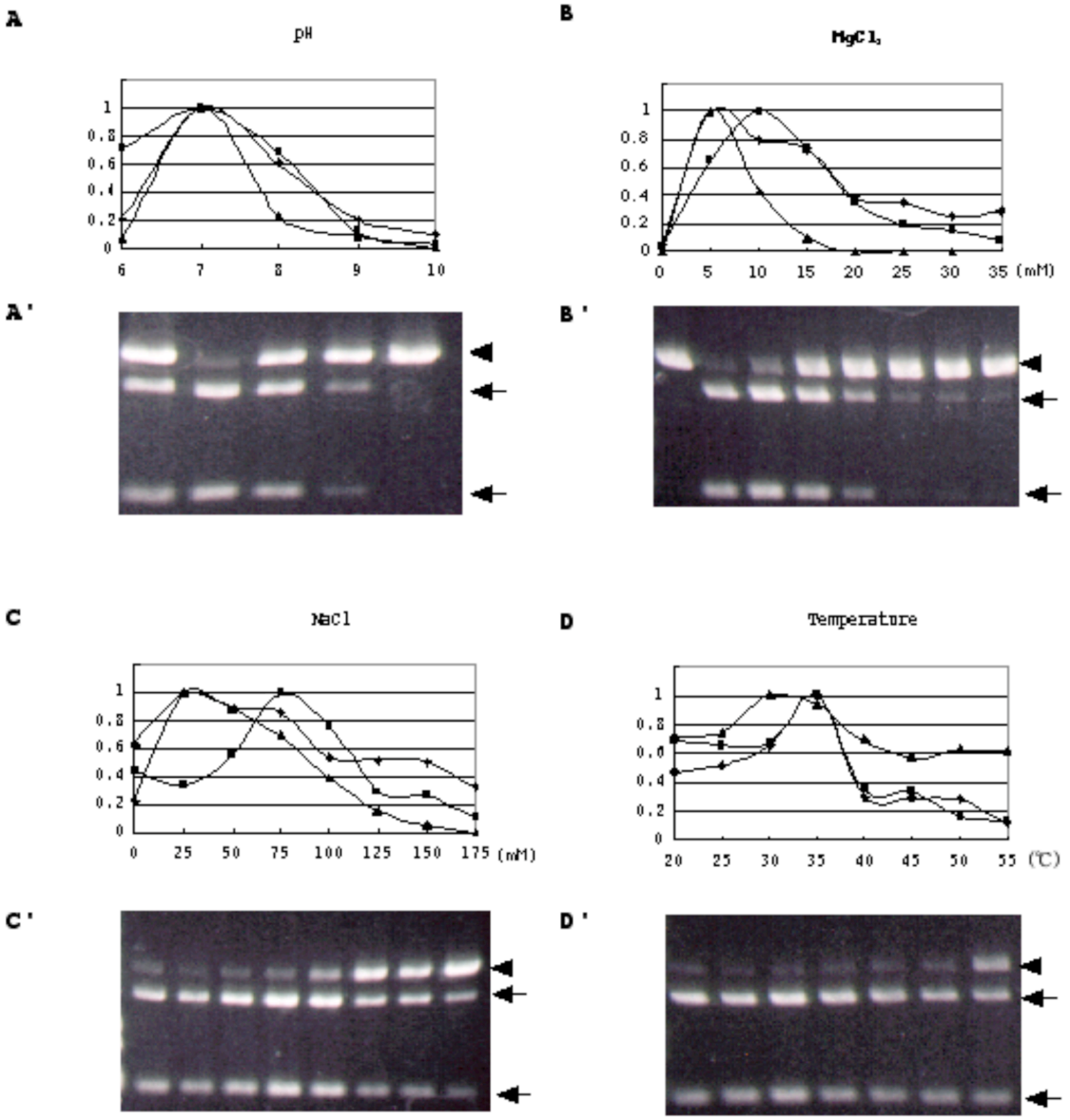


Fig. 1-2. Effects of pH, Mg2+, Na+ and temperature on the substrate cleavage reaction using recombinant homing enzyme I-*Csm*I **polypeptides**.
The conditions used to assay enzyme cleavage were as described in Materials and Methods. —◆— , reaction with recombinant protein I-*Csm*I(237); —▲— , I-*Csm*I(217); —■— , I-*Csm*I(200). Vertical axis of each graph (A-D) shows relative activity. The electrophoresis patterns of substrate cleavage by I-*Csm*I(200) are shown in (A') - (D'). Each lane in the agarose gel corresponds to a specific condition denoted in the axis of abscissa shown above the graph. An arrowhead denotes the position of the original substrate, while arrows show the cleaved substrates.

Table 1-1. Kinetic properties of intronic LAGLIDADG endonucleases.

	I-*Csm*I	I-*Ceu*I	I-*Sce*IV	I-*Dmo*I
Km	2.5X10-9M	0.9X10-9M	0.14-0.77X10-9M	4X10-9M
Vmax	1.8X10-12M/s	n.d. 1)	0.9-1.5X10-10M/s	n.d.
kcat	4.7X10-4/s	3.7X10-5/s	3-6X10-4/s	8.3X10-3/s
No. of motif per peptide	two	one	two	two

1) n.d.; not determined

was observed for pC-24nt by I-*Csm*I(200). This suggests that the recognition region of I-*Csm*I(200) resides between 12 nt upstream (+) and 12 nt downstream (-) of the intron insertion site, while 10 nt upstream and 10 nt downstream is insufficient for recognition.

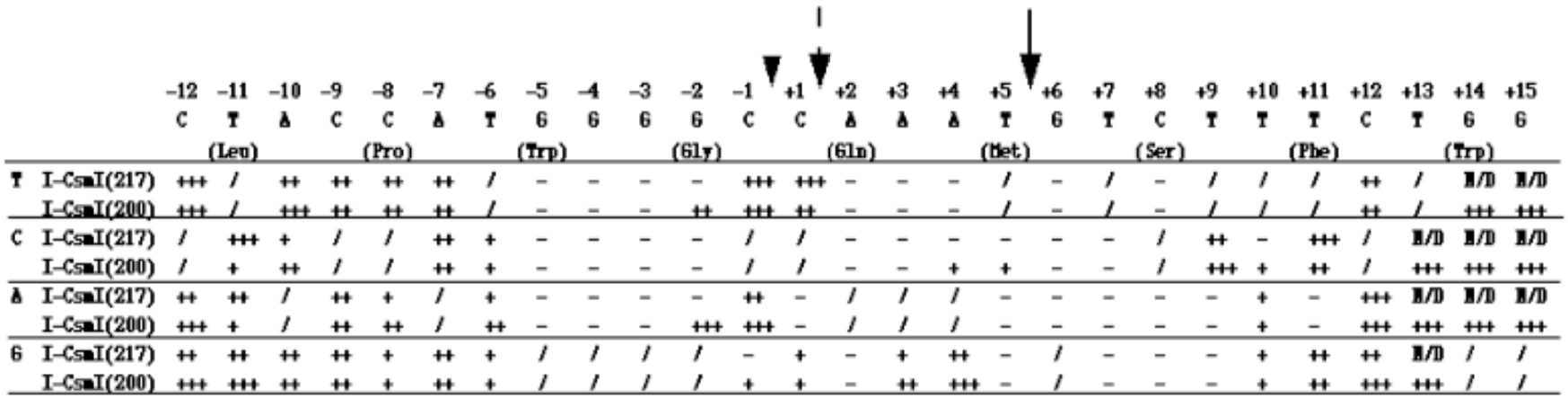
Cleavage point and mutational analysis of cleavable sequences

The precise cleavage site on each strand was determined through DNA sequencing of the substrate whose termini were blunt-ended by T4 DNA polymerase treatment. It became clear that cleavage occurs 5 nt downstream of the intron insertion site on the coding strand and one nt downstream of the insertion site on the non-coding strand, creating 3'-overhangs of 4 nt (Fig. 1-3). This terminal overhang is typical for DNA cleaved by LAGLIDADG homing enzymes.

81 variants (104 bp each) containing single nt substitutions between -12 and +15 were assayed to discern the critical nucleotides involved in recognition. Positions -5 through -3, +2 and +6 through +8 are strictly recognized by I-*Csm*I(200) and I-*Csm*I(217), as the original bases are essential for cleavage, while any substitution was permitted for positions -12 through -6 and +12 through +15 [some examples of cleaved pattern are shown in Fig. 1-4]. The majority of substitutions that blocked substrate cleavage were between -5 and +11 in relation to the intron insertion site. Therefore, the span of critical bases are not centered at the intron insertion site, but are spread almost symmetrically with respect to the cleavage points of coding and non-coding strands. A summary of substrate cleavability is classified into 4 groups (+++, ++, + and -; see Materials and Methods for details) and shown in Fig. 1-3. As a result of cleavage with I-*Csm*I (200), 26 % (8 %), 21 % (26 %), 15 % (14 %) and 38 % (51 %) kinds of substrates were classified into four classes, +++, ++, + and -, respectively [the results of I-*Csm*I(217) are shown in parentheses]. I-*Csm*I(200) and I-*Csm*I(217) showed almost identical sequence recognition properties (Fig. 1-3). A prominent difference in cleavage efficiency was observed for only two substitutions, the original G at position - 2 for A and T. I-*Csm*I(217) did not cleave these mutated substrates, whereas I-*Csm*I(200) cleaved both, with the G to A mutation the most efficient of the two (Fig. 1-3).

Correlation between the type of amino acid substitution and cleavage efficiency

I analyzed whether there is any correlation between the type of amino acid substitution



		−12	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	+15
		C	T	A	C	C	A	T	G	G	G	G	C	C	A	A	A	T	G	T	C	T	T	T	C	T	G	G
			(Leu)			(Pro)			(Trp)			(Gly)			(Gln)			(Met)			(Ser)			(Phe)			(Trp)	
T	I-CsmI(217)	+++	/	++	++	++	++	/	−	−	−	−	+++	+++	−	−	−	/	−	/	−	/	/	/	++	/	N/D	N/D
	I-CsmI(200)	+++	/	+++	++	++	++	/	−	−	−	++	+++	++	−	−	−	/	−	/	−	/	/	/	++	/	+++	+++
C	I-CsmI(217)	/	+++	+	/	/	++	+	−	−	−	−	/	/	−	−	−	−	−	−	/	++	−	+++	/	N/D	N/D	N/D
	I-CsmI(200)	/	+	++	/	/	++	+	−	−	−	−	/	/	−	−	+	+	−	−	/	+++	+	++	/	+++	+++	+++
A	I-CsmI(217)	++	++	/	++	+	/	+	−	−	−	−	++	−	/	/	/	−	−	−	−	−	+	−	+++	N/D	N/D	N/D
	I-CsmI(200)	+++	+	/	++	++	/	++	−	−	−	+++	+++	−	/	/	/	−	−	−	−	−	+	−	+++	+++	+++	+++
G	I-CsmI(217)	++	++	++	++	+	++	+	/	/	/	/	−	+	−	+	++	−	/	−	−	−	+	++	++	N/D	/	/
	I-CsmI(200)	+++	+++	++	++	+	++	+	/	/	/	/	+	+	−	++	+++	−	/	−	−	−	+	++	+++	+++	/	/

Fig. 1-3. Mutational analyses of the recognition efficiency by recombinant homing enzymes I-*Csm*I**(200) and** I-*Csm*I**(217).** The coding sequence of *C. reinhardtii* COB gene and the assigned amino acids are shown on top. The three possible base substitutions for each position are indicated to the left side. An arrowhead indicates the intron insertion site. An arrow with a dotted line shows the cleavage site of the non-coding strand, while an arrow with solid line denotes the cleavage site for the coding strand. The numbering is in relation to the intron insertion site. "+++", substrate cleavage above the wild-type levels (more than 150 %); "++", cleavage almost the same or slightly less than the wild-type levels (120 - 80%); "+", cleavage below the wild-type levels (50 -20%); "-" almost no cleavage (less than 10%); "/", position of the wild-type nucleotide.

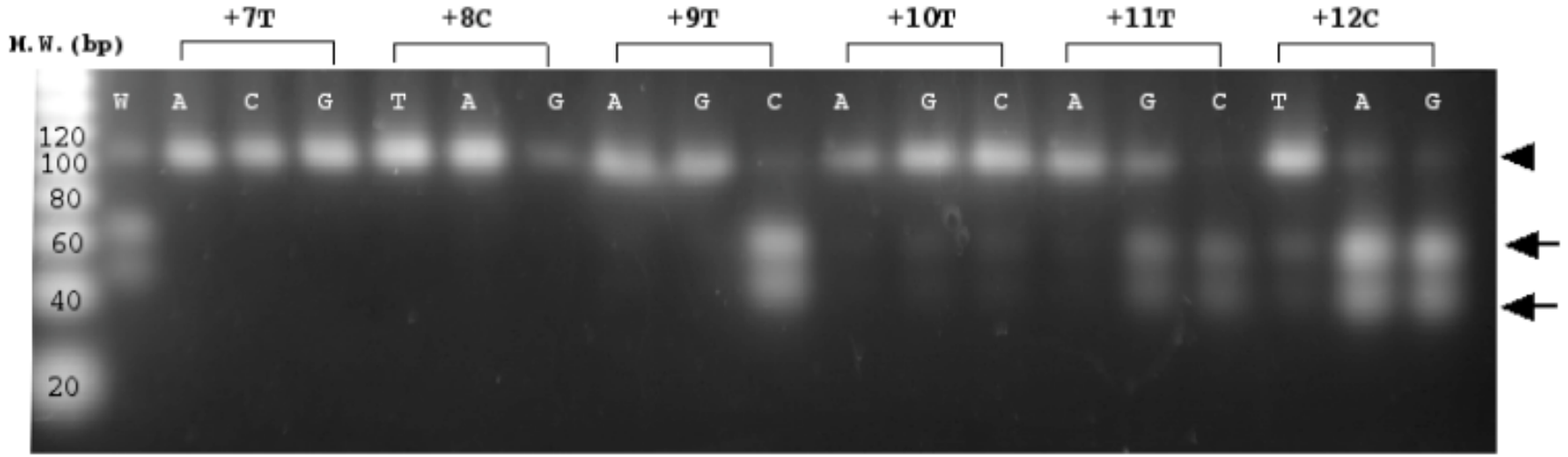


Fig. 1-4. Cleavage pattern of linearized substrates containing single base substitutions by I-*Csm*I**(200)**.
The numbering is in relation to the intron insertion site, with "+" indicating upstream, followed by the nucleotide that is the original base at the given position, while the nucleotide denoted below shows the base after substitution. M.W., 20 bp molecular weight marker ladder. An arrowhead indicates the position of substrate DNA (104 base pairs), while arrows indicate the positions of cleaved substrate (60 and 44 bp). W; substrate DNA containing the *Chlamydomonas reinhardtii* wild type COB sequence.

induced by single nt substitution (silent/tolerated change, or non-silent/non-tolerated change) and how efficiently the substrates are cleaved by two kinds of N-terminal truncated I-*Csm*I polypeptides. A survey of Genbank registered sequences of various organisms showed the target DNA sequences of I-*Csm*I, I-*Spom*I, I-*Sce*II and I-*Sca*I correlate to the amino acid sequences WGQMS(F/H), TGWT(A/V)PPL, FGHPEV and W(G/A)TVI, respectively. Therefore, F/H, A/V and G/A amino acid changes at the specific sites were functionally tolerated in this investigation. 48 substrates containing single nt substitutions at the core recognition region (between -5 and +11) were analyzed from this point of view.

Substrates containing a silent or tolerated amino acid change

Seven of 48 substrates contained a silent amino acid change. However, two of seven such substrates [containing TCT(Ser) changed to TCA and TCG(Ser), mutation position +9 in Fig. 1-3] were not cleaved at all by I-*Csm*I(217) and I-*Csm*I(200), and additionally the substrate contains the change GGC(Gly) to GGG(Gly) (position -1) was not cut by I-*Csm*I(217) even though these silent changes must be tolerated in nature. On the other hand, three silent substrates [TCT(Ser) to TCC(Ser), position +9; GGC(Gly) to GGT/GGA(Gly), position -1] were cut efficiently by the two I-*Csm*I polypeptides. Additionally, CAA(Gln) to CAG(Gln) (position +3) was efficiently cut by I-*Csm*I(200).

Substrates containing a non-silent or non-tolerated amino acid change

41 of 48 substitutions caused non-silent/non-tolerated amino acid changes. Showing an adaptation to the possible target DNA sequences, I-*Csm*I polypeptides only slightly cleaved most of them (Table 1-2). Such property is also prominently detected in I-*Spom*I and I-*Sca*I. However, TAA(Stop) instead of CAA(Gln) (position +1), TGC(Cys) and TCC(Ser) instead of TTC(Phe) (position +11) were efficiently cleaved by the both I-*Csm*I enzymes, even though these codons are not observed at these positions in nature. In contrast, none of the non-silent/non-tolerated substitutions were cleaved efficiently by I-*Sca*I (Table 1-2).

Table 1-2. Type of amino acid substitution contained in the substrate and the cleavage efficiency .

Type of substitution	Homing endonuclease	Efficiently1) cleaved	Moderately2) cleaved	Not or scarcely3) cleaved
Silent amino acid changes	I-*Spom*I	66.7%(4/6)	33.3%(2/6)	0.0%(0/6)
	I-*Sca*I	14.3%(1/7)	85.7%(6/7)	0.0%(0/7)
	I-*Sce*II	100.0%(7/7)	0.0%(0/7)	0.0%(0/7)
	I-*Csm*I(217)	42.9%(3/7)	14.3%(1/7)	42.9%(3/7)
	I-*Csm*I(200)	57.1%(4/7)	14.3%(1/7)	28.6%(2/7)
Synonymous amino acid changes4)	I-*Spom*I	42.9%(3/7)	14.3%(1/7)	42.9%(3/7)
	I-*Sca*I	0.0%(0/2)	50.0%(1/2)	50.0%(1/2)
	I-*Sce*II	0.0%(0/2)	50.0%(1/2)	50.0%(1/2)
	I-*Csm*I(217)	0.0%(0/9)	11.1%(1/9)	88.9%(8/9)
	I-*Csm*I(200)	11.1%(1/9)	11.1%(1/9)	77.8%(7/9)
Non-silent and non-synonymous amino acid changes	I-*Spom*I	0.0%(0/12)	25.0%(3/12)	75.0%(9/12)
	I-*Sca*I	0.0%(0/20)	30.0%(6/20)	70.0%(14/20)
	I-*Sce*II	30.0%(9/30)	43.3%(13/30)	26.7%(8/30)
	I-*Csm*I(217)	10.3%(3/32)	10.3%(3/32)	81.3%(26/32)
	I-*Csm*I(200)	15.6%(5/32)	15.6%(5/32)	68.8%(22/32)

1) Efficiently cleaved: efficiency more than 80% of the wild type substrate for I-*Spom*I and I-*Csm*I, while more than 78% for I-*Sce*II; for I-*Sca*I, efficiency of originally described as "mutant cleaved as well as the wild-type".
2) Moderately cleaved: 80-30% of the wild type substrate for I-*Spom*I and I-*Csm*I, while 60-42% for I-*Sce*II; for I-*Sca*I, efficiency of originally described as "reduced cleavage".
3) Not or scarcely cleaved: less than 30% of the wild type substrate for I-*Spom*I and I-*Csm*I, while 33 % for I-*Sce*II; and for I-*Sca*I, efficiency of originally described as "no cleavage".
4) Adopted synonymous amino acid changes are as follows: Ala/Gly, Arg/Lys, Asn/Gln, Asp/Glu, Ser/Thr, Ile/Leu/Met/Val, Phe/Trp/Tyr.

Discussion

The original I-*Csm*I ORF is fused with the preceding exon, which is not rare for group I intronic ORFs. The entire ORF of I-*Spom*I also extends into the upstream exon of the COXI gene, and it has been reported that the N-terminal truncated polypeptide, including the two LAGLIDADG motifs, has similar sequence specificity to that detected using mitochondrial extracts (Pellenz *et al.*, 2002). Considering the above, I tried to overproduce three kinds of N-terminally truncated recombinant I-*Csm*I polypeptides that retain the two LAGLIDADG motifs instead of the entire I-*Csm*I (374 amino acids) (Fig. 1-1), because I failed to express the whole I-*Csm*I ORF for reasons that are unclear. I found that all of the N-terminal truncated I-*Csm*I polypeptides retain the specificity to cleave the target site, and the kinetic parameters of I-*Csm*I(200) are very similar to that reported for representative intronic homing enzymes of LAGLIDADG motifs (Table 1-1). The optimal conditions of selected factors were also very similar to other homing enzymes, with the exception of the preferred pH. I-*Csm*I displayed its highest activity at pH 7.0, which is very close to the reported physiological pH value of 7.5 in yeast mitochondria (Wernette *et al.,* 1990), while most of the LAGLIDADG enzymes show their highest activity at an alkaline pH between 8.5 and 9.5 [*e.g.*, optimal pH is 9.2 for I-*Ani*I (Geese *et al.,* 2003)], and between 8.5 and 9.0 for the recombinant I-*Sca*I (Monteilhet *et al.,* 2000). Having a host pH that is lower than the optimum pH observed for many homing enzymes may act to reduce endonuclease activity and prevent overdigestion of the genomic DNA.

I-*Csm*I(200)'s optimal conditions for Na^{+} and Mg^{2+} are clearly shifted to a concentration higher than that of I-*Csm*I(217) and I-*Csm*I (237) (Fig. 1-2B, 1-2C). This suggests that the three dimensional conformation of this enzyme is different from the others possibly because of the recessed N-terminal region, and may explain the differences in cleavage activity between I-*Csm*I(200) and I-*Csm*I(217). I-*Csm*I(200) seems to tolerate a higher degree of sequence ambiguity than I-*Csm*I(217) at position -2, because I-*Csm*I(200) can efficiently cleave the mutated substrates of -2A and -2T (instead of the original -2G), while I-*Csm*I(217) only tolerates the original base -2G (Fig. 1-3).

Cleavage of a target DNA is an essential step for lateral transfer of an intron. Therefore, if a homing enzyme shows very stringent recognition of the target core

sequence, this step could be a bottleneck for horizontal transmission of an intron. The target site of I-*Csm*I corresponds to the amino acid sequence of Trp-Gly-Gln-Met-Ser-(Phe/His). This is a highly conserved region in COB genes among a wide range of organisms. Our systematic induction of a point mutation and the cleavage assay showed a clear tendency that I-*Csm*I polypeptides efficiently cleave silent change containing substrates than non-synonymous/non-tolerated change containing ones (Table 1-2).

It is obvious that stop codons are never tolerated at the internal regions of a gene. However, our systematic induction of a point mutation introduced stop codons, *i.e.*, TGA and TAG stop codons from TGG(Trp), and TAA stop codon from CAA(Gln). The substrate DNA that contains TGA or TAG was not cleaved, while the substrate containing a TAA stop codon was efficiently cleaved by the both I-*Csm*I polypeptides (Fig. 1-3). Moreover, substrates including a codon that highly likely appears in nature were not cleaved [*e.g.*, TCA/TCG(Ser) from TCT(Ser), and three Ile codons AT(T/C/A) from ATG(Met)]. The above instances indicate that the recognition property of I-*Csm*I is not skillfully adapted to recognize target sequences that are highly likely to appear in nature.

It is possible that the recognition property of I-*Spom*I, I-*Sca*I and I-*Sce*II are adapted to recognize multiple possible target sequences, because these homing enzymes cleaved substrates containing various kinds of silent/tolerated amino acid changes efficiently, and none of them were remained uncleaved (Table 1-2).

Considering the above, I propose that homing enzymes are adapted to recognize diverse target sequences to facilitate horizontal transmission to a new species, as evidently seen with I-*Spom*I, I-*Sca*I and I-*Sce*II. However, immediately after a successful invasion, mutations begin to accumulate that lead to a loss of further adaptation, because homing enzyme activity is only essential for intron invasion and thereafter it is useless to the cell. Invasion of I-*Csm*I might be evolutionarily older than the other three homing enzymes compared in this study, because I-*Csm*I showed the least adapted properties among the four. Actually, remnants of homing enzyme ORFs that include frame shifts or stop codons within the ORF are frequently found (*e.g.*, Watanabe *et al.,* 1998). Comprehensive analysis of omega homing enzyme and its associated group I intron revealed that it is more common to find an inactive intron/ORF combination than it is to find an active intron/ORF combination or an intron-less allele (Goddard and Burt, 1999).

It has been proved that some of intronic homing enzymes are bifunctional. They work not only as an endonuclease but also as a maturase to preserve splicing. The bifunctional activity of I-*Spom*I (Schafer *et al.*, 1994), I-*Sca*I (Szczepanek and Lazowska, 1996), and I-*Ani*I (Ho *et al.,* 1997) has been observed. I-*Csm*I could also be a bifunctional protein that acts as a maturase, which may also preserve its endonuclease activity for horizontal transmission. These bifunctional enzymes are recognized as intermediates, and may likely lose their endonuclease activity over time, retaining only their maturase activity (Watanabe *et al.,* 1998; Geese *et al.,* 2003).

Chapter 2

Recognition property of a mitochondrial homing enzyme in *Chlamydomonas* assayed through intron invasion

Abstract

Target sequence cleavage is the essential, initial step for successful intron invasion into an intronless allele. DNA cleavage at a specific site is performed by an exonuclease, termed a homing enzyme, which is encoded by an open reading frame within the intron. Thus far, the recognition properties of homing enzymes have only been analyzed *in vitro*, using purified, recombinant homing enzyme and various mutated DNA substrates, but it is unclear whether the homing enzyme behaves similarly *in vivo*. To answer this question, I determined the recognition properties of I-*Csm*I *in vivo*, and compared them to our previous *in vitro* results. I-*Csm*I is a homing enzyme encoded by the ORF of the alpha-group I-intron, located in the mitochondrial COB gene of the green alga *Chlamydomonas reinhardtii*. The *in vivo* recognition properties of the enzyme were determined as the frequency of intron invasion into a mutated target site. For this purpose, I utilized hybrid diploid cells developed by crossing alpha-intron-plus *C. smithii* to intron-minus *C. reinhardtii* containing mutated target sequences. The intron invasion frequency was much higher than expected when compared to the cleavage frequency of the respective mutated substrates from the previous *in vitro* experiments. Even the substrates that had very little cleavage in the *in vitro* experiment were efficiently invaded *in vivo*, and were accompanied by a large degree of coconversion of the flanking genetic markers. Considering the relative ease of the homing enzyme invading into various mutated target sequences encoding strongly conserved amino acid residues among a wide range of organisms, I propose that the principle bottleneck for lateral intron transmission is not the sequence specificity of the homing enzyme, but instead is limited by the rare occurrence of inter-specific cell fusion.

Introduction

Molecular phylogenetic analyses have demonstrated that group I introns are mobile elements that can be laterally transmitted to a wide range of organisms (Lambowitz, 1989; Belfort and Roberts, 1997; Cho *et al*., 1998; Watanabe *et al.,* 1998). They are thought to be parasitic genetic elements that have persisted for an evolutionarily long time (Goddard and Burt, 1999). Site-specific double-stranded DNA cleavage is the essential initial step for regular intron invasion. Without cleavage of the target DNA sequence, the intron cannot invade the intronless genome. DNA cleavage is performed by a sequence specific exonuclease (termed a homing enzyme) encoded in an open reading frame (ORF) found within the intron. Cleavage of the target DNA triggers a series of reactions resulting in the transmission of the intervening sequence unidirectionally from the intron-containing DNA strand to the cleaved, intron-lacking recipient DNA strand. *In vitro* analysis has shown that the homing enzyme recognizes a non-palindromic 16 to 30 base pair (bp) sequence encoding highly conserved amino acid residues (for reviews, Belfort and Perlman, 1995; Chevalier and Stoddard, 2001). On the other hand, the homing enzyme also recognizes DNA sequences with a varying degree of degeneracy, and is adapted to cleave these sequences in nature; *i.e.,* most of the nucleotide differences are at the codon 3rd positions in the target sequence.

So far, the analysis of homing enzyme properties has been limited to *in vitro* assay systems using recombinant enzymes and artificial substrates. Unfortunately, no procedure is available to address the differences between the *in vitro* and *in vivo* behaviors of the homing enzymes. Therefore, the efficiency of homing enzyme cleavage *in vivo* remains unknown. It is also unknown why most homing enzymes are active *in vitro* around pH 9.0 (Monteilhet *et al*., 2000; Geese *et al*., 2003), which is an alkaline environment compared to the physical pH value of pH 7.5 in yeast mitochondria (Wernette *et al*., 1990). The lack of answers to these basic questions is probably due to the difficulty in developing strains bearing the intended mutations in the target region. The unicellular green alga *Chlamydomonas smithii* has a 1075-bp long group I intron (alpha-intron or *Cs* COB.1) located in the apocytochrome b (COB) gene of the mitochondrial genome (Colleaux *et al*., 1990). The alpha-intron contains an ORF encoding a homing enzyme, I-*Csm*I, which has the two typical LAGLIDADG motifs.

Using two types of N-terminally truncated recombinant I-*Csm*I polypeptides, I previously performed systematic *in vitro* assays to determine the cleavage efficiency of various mutated target sequences (Kurokawa *et al*., 2005). Furthermore, I recently developed a practical method for mitochondrial transformation of *Chlamydomonas reinhardtii* (Randolph-Anderson *et al*., 1993; Yamasaki *et al*., 2005). Using this technique, I developed various *C. reinhardtii* strains containing different point mutations at the I-*Csm*I target site. These strains have been very useful for probing the recognition properties of I-*Csm*I *in vivo*. The two species are interfertile and have very similar mitochondrial genome organization and DNA sequences, except that *C. reinhardtii* lacks the alpha-intron located within the COB gene. Therefore, the *in vivo* recognition properties of I-*Csm*I were investigated by analyzing the invasion of the alpha-intron into the *C. reinhardtii* strains containing the various mutated target sites within their mitochondrial genomes. The analysis is complicated because *Chlamydomonas* mitochondrial DNA is transmitted to the meiotic progeny from the paternal (mt^-) parent but not from the maternal (mt^+) parent (Boynton *et al*., 1987; Matagne *et al*., 1988). Maternal mitochondrial genomes are subjected to degradation in the course of meiosis (Beckers *et al.,* 1991) preceding the fusion of mt^+ and mt^- mitochondria (Nakamura *et al*., 2003). However, in the diploid cells, which are rare zygotes and account for about 5 % of the mated cells, inheritance of the mitochondrial genome is no longer strictly uniparental. This means that the analysis of diploid cells is essential when investigating the sequences susceptible to I-*Csm*I cleavage *in vivo*, as measured by intron invasion into the intronless *C. reinhardtii* mitochondrial genome. In the hybrid diploid cell, fusion of *C. smithii* and *C. reinhardtii* mitochondria occurs without elimination of mt^+ mitochondrial genomes, and the diploid cell divides mitotically, escaping the maturation steps of meiosis (Gillham, 1978).

Materials and Methods

Algal strains and culture conditions

Wild type *Chlamydomonas smithii* CC-1373, mating type plus (mt^+), *Chlamydomonas reinhardtii* CC-124, mating type minus (mt^-), and the respiratory deficient strain CC-2654

dum-1 (mt^-) were obtained from the *Chlamydomonas* Genetic Center (Dept. of Botany, Duke Univ. NC 27706, USA). The strains were cultured in Tris-acetate-phosphate (TAP) medium (Gormman and Levine, 1965) at 25 ˚C under constant cool white fluorescent light (84 μmol m^{-2} s^{-1}) with vigorous shaking, unless indicated otherwise.

Developing *C. reinhardtii* strains containing various single point mutations

The 5.0 kb DNA construct, which contains a wild-type *C. reinhardtii* COB gene and upstream region was used as starting material (Yamasaki *et al*., 2005). I-*Csm*I's target sequence in the 5.0 kb DNA construct was modified using the mega primer method (Sambrook and Russell, 2001). In total, 83 types of single point mutation-containing 5.0 kb DNA constructs were prepared and ligated into the *E. coli* vector pT7-Blue (Novagen, Milwaukee, USA). The respiratory deficient *C. reinhardtii* strain *dum*-1, which lacks the entire COB gene and the upstream region, was transformed using a mutated 5.0 kb DNA construct to compensate the deleted region as described elsewhere (Yamasaki *et al*., 2005). After biolistic bombardment of *dum*-1, the plates were kept in the dark for 3-4 weeks until respiratory-active colonies appeared. Single colonies were re-streaked, isolated, and sequenced to verify whether they contained the correct mutation.

Isolation of hybrid diploid cells by crossing *C. smithii* and *C. reinhardtii*

I conferred spectinomycin resistance (spc R) on mt^+-*C. smithii* by introducing plasmid pEX-50-AAD (Takahashi *et al*., 1996) into the plastid, while zeocin (or bleomycin) resistance (ble R) was conferred to each one of mt^--*C. reinhardtii* transformants by re-introducing pSP124S (Lumbreras *et al*., 1998) into the nucleus.

The mating and isolation of rare diploid cells were performed as follows: The *C. smithii* (mt^+, spc R) and single point mutation-carrying *C. rinhardtii* (mt^-, ble R) were grown at 25 ˚C under constant light on NH_4Cl deprived TAP agar plates for one week. Then, the cells were suspended into gamete-inducing medium (Martin and Goodenough, 1975; Harris, 1989) for about 3 h at 25 ˚C under constant white fluorescent light (120 μmol m^{-2} s^{-1}). Two kinds of gametes were mixed for 30 min and allowed to mate, and spread on TAP plates containing 100 μg/ml spectinomycin and 10 μg/ml zeocin (Invitrogen, California, USA). Colonies that appeared after 5-7 days were chosen, and two rounds of single-colony isolation was performed using plates containing

spectinomycin and zeocin. The diploidy of these cells was confirmed by the existence of mating type-specific genes within the cell, *fus1* (mt^+ specific) and *mid1* (mt^- specific), using the PCR method of Zamora *et al.* (2004).

PCR product analysis to indicate the presence of alpha-intron in the target region

Each diploid cell was grown to late log phase in 3 ml of TAP medium. Then, total DNA was prepared using DNeasy Plant Mini Kit (Qiagen, California, USA) after disrupting the cells with glass beads. PCR was performed using a set of primers that flank the I-*Csm*I target sequence (the location of primer set A is shown in Fig. 2-1). With this primer set, a 3 kb PCR product is expected when the alpha-intron is inserted at the target site, and a 2 kb PCR product is expected for the intronless target site. At least 10 hybrid diploids were analyzed for each type of cross.

Coconversion frequency of the diagnostic restriction enzyme sites associated with intron invasion

To probe the range of coconversion using genetic markers, I digested the PCR product with diagnostic restriction enzymes, which exist only in the *C. smithii* mitochondrial genome. Following the confirmation of intron invasion, two types of PCR were performed to amplify the upstream and downstream regions flanking the alpha-intron.

A 1 kb PCR fragment was produced from the following set of primers: one that anneals at the 3'-terminus of the alpha-intron, and one that anneals to the left arm of the genome (primer set B in Fig. 2-1). This PCR product contains a diagnostic *Nhe*I restriction enzyme site (hereafter *Nhe*I*) that is located 500 bp upstream of the I-*Csm*I target site. To generate the 5.5 kb PCR fragment, one primer hybridizes to the alpha-intron's 5' end and the other to the N-terminus of the ND2 gene (primer set C in Fig. 2-1). This product contains diagnostic *Nco*I and *Hpa*I restriction enzyme sites (hereafter *Nco*I* and *Hpa*I*), which are located 1.4 kb and 5.0 kb downstream of the forward primer site, respectively (Fig. 2-1). PCR products were purified using a PCR purification kit (Qiagen, California, USA) and digested with the appropriate restriction enzymes.

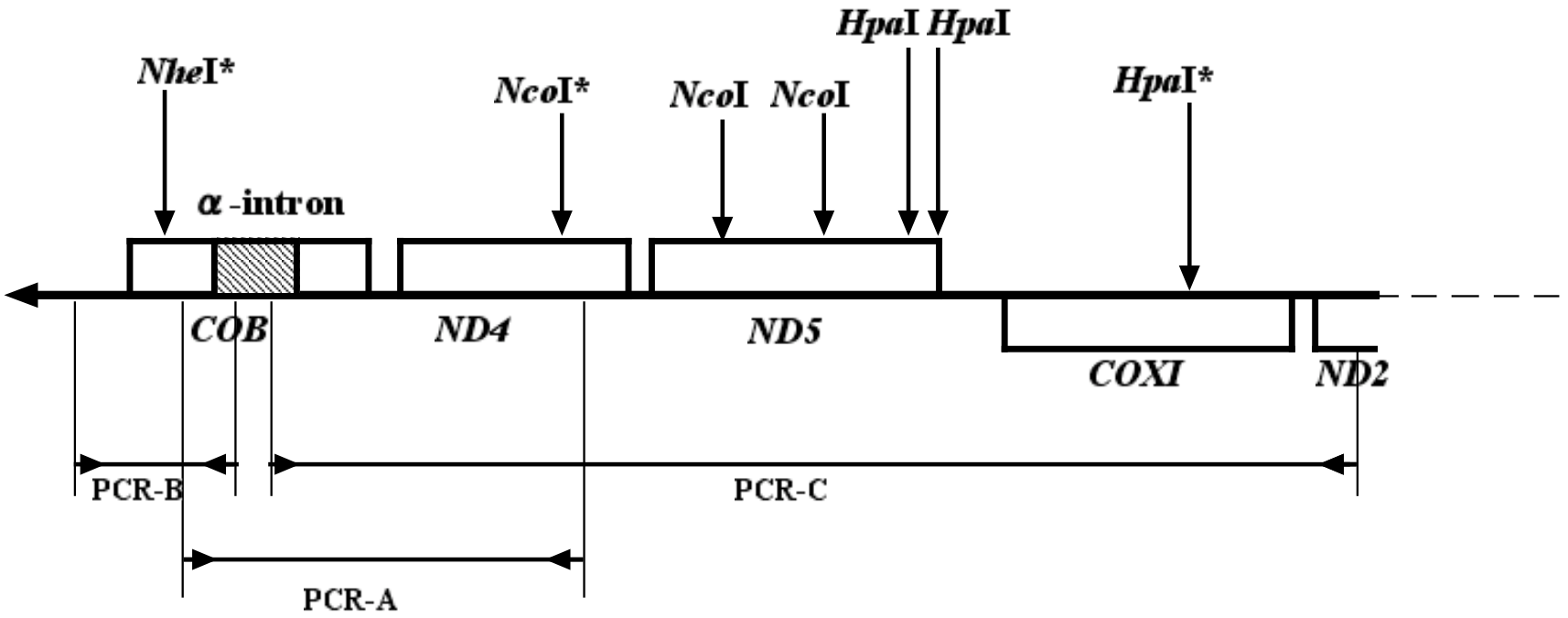


Fig. 2-1. Partial restriction map of the linear mitochondrial genome of *Chlamydomonas smithii* along with the designation of amplified regions by PCR method.
Asterisk designates unique restriction site that is lacking in the *C. reinhardtii* allelic genes.

Calculation of the frequencies of intron invasion and coconversion of the flanking genetic markers

In this study, the intron invasion frequency is defined as the ratio of intron-containing diploid cells to the total number of examined diploid cells. A diploid cell that is heteroplasmic for the presence of the intron was counted as both an intron-containing and intronless diploid cell. The same approach was used to calculate the coconversion frequency of the flanking genetic markers when I observed similar heteroplasmy.

Results

Creation of *C. reinhardtii* strains containing single point mutations at the I-*Csm*I target site

I prepared 33 types of DNA constructs and used them to introduce various single point mutations into the I-*Csm*I target region of the *C. reinhardtii* mitochondrial COB gene (Table 2-1). These mutations correspond to the six highly conserved amino acid residues at the target site, Trp-Gly-Gln-Met-Ser-Phe. I attempted to transform respiratory deficient *dum*-1 strains two or three times by biolistic-bombardment with each type of DNA construct. As a result, I obtained 14 types of respiratory active strains that contain the desired point mutation. However, I was not able to obtain the other 19 types of DNA constructs using the same approach. It became clear that our success was related to the type of mutation present in the 5.0 kb DNA construct. 17 of the 19 unsuccessful transformation constructs contained non-synonymous/non-silent mutations. On the other hand, nine of the 14 successfully transformed constructs contained silent mutations, four had synonymous ones (one Met to Leu mutation and three Met to Ile mutations at different positions), and only one construct had non-synonymous/non-silent mutations [TGG(Trp)/TGT(Cys)]. These results suggest that our inability to obtain transformants was caused by the inactive COB peptides produced from the mutated COB gene of the 5.0 kb DNA constructs. Supporting the above results, some of the transformants containing silent mutations showed slower growth, *i.e.,* four of the five randomly selected transformants containing GGC/GGG(Gly), CAA/CAG(Gln), CCA/CCC(Pro) or TCT/TCC(Ser) mutations showed about 20 % slower growth than the

Table 2-1. List of mutations introduced into 5.0 kb-DNA construct and resulting assigned amino acid changes.

	Pro			Trp			Gly			Gln			Met			Ser		
	#-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9
	C	C	A	T	G	G	G	G	C	C	A	A	A	T	G	T	C	T
A				×			×		○	×				×	○			×
				Arg			Ser		Gly	Lys				Lys	Ile			Ser
T			○		×	○		×	○		×	×	○		○			
			Pro		Leu	Cys		Val	Gly		Leu	His	Leu		Ile			
G			○	×					○	×	×	○	×	×				○
			Pro	Gly					Gly	Glu	Arg	Gln	Val	Arg				Ser
C			○		×	×	×				×			×	○*			○
			Pro		Ser	Cys	Arg				Pro			Thr	Ile			Ser

*: very limitted mating ability (see text), #: the numbering is in relation to the intron insertion site, ○: transformant was obtained, ×: transformants was not obtained.

wild-type, with the exception of the GGC/GGA(Gly) mutation (data not shown).

Zeocin resistance was conferred to each type of successfully transformed *C. reinhardtii* (mt^-) strain by introducing the BLE gene into the nucleus, followed by mating to *C. smithii* (mt^+, spc R) to obtain chimeric diploid cells. More than 10 hybrid diploids were developed from each possible mating combination, except for the *C. reinhardtii* strain, which contains the mutation ATG/ATC(Ile). A cross between this strain and *C. smithii* resulted in very limited mating because of unknown reasons, and I was unable to analyze those diploid cells. All together, I successfully obtained 13 types of hybrid diploid cells as a result of crosses between the intron-containing *C. smithii* and various mutated intronless *C. reinhardtii* strains.

Considering the possibility of spontaneous antibiotic resistance, I confirmed the diploidy of the double-antibiotic resistant cells by checking for the presence of mating type specific genes by PCR (Fig. 2-2). Among 134 of 183 randomly chosen zeocin and spectinomycin resistant cells, three were actually false-positive diploid cells. The mt^+ specific gene, *fus1*, could not be detected by PCR. Therefore, these cells were likely mt^- haploid, not diploid. It is likely that the spectinomycin resistance of these mt^- cells is a result of a spontaneous mutation within a chloroplast ribosomal protein or its rRNA genes (Harris *et al*., 1989). Such pseudo-diploids were not analyzed further.

Intron invasion into the mutated target site of *C. reinhardtii* mitochondrial genome

Alpha-intron invasion into the *C. reinhardtii* target site was analyzed using a set of primers flanking the I-*Csm*I target site (Fig. 2-1). With this primer set, the expected PCR product is 3 kb long when the target site contains the alpha-intron and 2 kb long when the intron is missing (Fig. 2-3). The intron invasion frequency was calculated as a ratio of the two PCR products and is summarized in Table 2-2.

The +6 base, relative to the intron insertion site, was the most critical base among the mutated sites investigated in this study: Two types of mutations positioned at +6, ATG(Met) to ATA(Ile) and ATT(Ile), resulted in a low invasion frequency, 6 % and 35 % respectively. Position +6 is near the cleavage point of the coding strand, located between bases +5 and +6. This result is consistent with our previous *in vitro* assay, which showed that the bases located around the coding strand cleavage point are critical for intron invasion (Kurokawa *et al*., 2005). Another striking feature of the *in vivo* assay

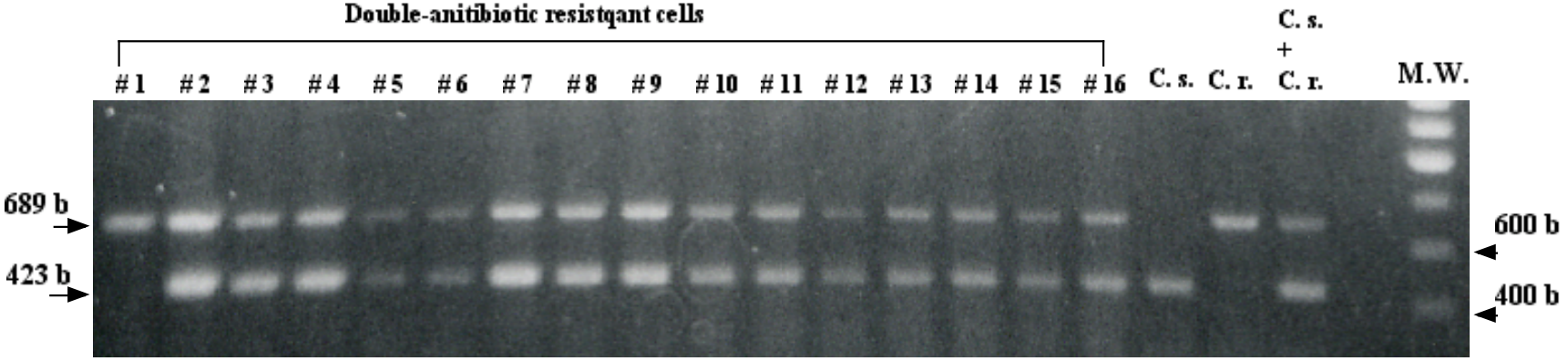


Fig. 2-2 Some examplified results of mating-type specific PCR.

689 b PCR product is yielded from mt^- specific gene *mid1*. While 423 b product is from mt^+ specific gene *fus1*. Diploid cells (# 2-16) resulted in 689 b and 423 b PCR products.. The double-antibiotic resistant cell (# 1) resulted in only 689 b product. Therefore, the cell used for #1 is likely mt- haploid, not diploid.

C. s. : wild- type *C. smithii* (mt^+), C. r.: wild-type *C. reinhardtii* (mt^-), M.W., 200 bp molecular weight marker ladder.

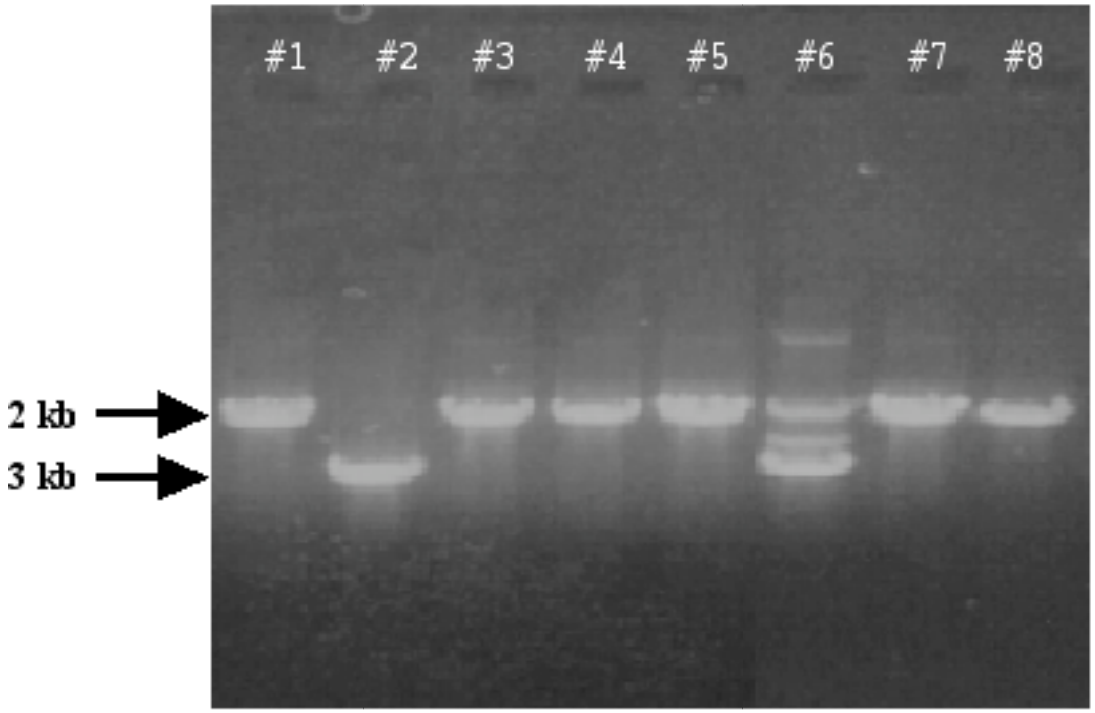


Fig. 2-3. Some exemplified results of intron invasion assay.

Hybrid diploid cells developed by cross between *C. smithii* and *C. reinhardtii* bearing single point mutation in the target sequence were assayed for the existence of alpha intron. 3 kb PCR product was yielded from diploid cells, #1, #3, #5, #7 and #8, showing the intron invasion. While 2 kb product was from #2, indicating the lack of alpha intron. Diploid #6 resulted in 2 kb and 3 kb PCR products along with minor other bands, reflecting the heteroplasmy for existence of the intron.

Table 2-2. Intron invasion ratio in each type of mutation.

Position of mutation	Type of mutation	Result of *in vitro* cleavage	Type of PCR product and intron invasion ratio			Specific restriction sites and coconversion frequency			
		I-*Csm*I(200)	3kb	2kb	2&3kb	*Nhe*I*	*Nco*I*	*Hpa*I*	No. of diploid
-7	**CCA(Pro)/CCT(Pro)**	++[(1)]	10	2	0	-	+	+	1
						-/+	-/+	-	1
						+	-	-	1
						+	+	-	1
						+	+	+	6
				83% [(2)]		82% [(3)]	82%	70%	
-7	**CCA(Pro)/CCG(Pro)**	++	15	0	0	-	-	-	1
						-	-/+	-	1
						-	+	-	1
						-/+	-	-	1
						-/+	-/+	-	2
						-/+	+	-	1
						-/+	+	+	1
						+	-/+	-	1
						+	-/+	-/+	1
						+	+	-	1
						+	+	-/+	3
						+	+	+	1
				100%		60%	65%	32%	
-7	**CCA(Pro)/CCC(Pro)**	++	10	1	0	-	-/+	-/+	1
						-/+	-	-	1
						-/+	-	-/+	1
						-/+	+	-/+	1
						-/+	+	+	1
						+	-	-	1
						+	-/+	-/+	1
						+	+	-/+	1
						+	+	+	2
				91%		64%	58%	53%	
-4	**TGG(Trp)/TGT(Cys)**	-	4	9	0	-	-	-	2
						+	-/+	-	1
						+	+	-	1
				31%		50%	40%	0%	
-1	**GGC(Gly)/GGA(Gly)**	+++	10	0	0	-	-/+	-/+	1
						-	+	+	2
						+	+	-	2
						+	+	-/+	1
						+	+	+	4
				100%		70%	91%	67%	
-1	**GGC(Gly)/GGT(Gly)**	+++	13	0	0	-	-	-	2
						-	+	-	1
						+	-	-	3
						+	+	-	1
						+	+	-/+	3
						+	+	+	3
				100%		77%	62%	38%	

continued

-1	**GGC(Gly)/GGG(Gly)**	+	11	0	1	-	-	-	2
						-	-/+	-	1
						-/+	-	-	1
						-/+	-/+	-/+	1
						-/+	+	+	1
						+	-	-	3
						+	-/+	-/+	1
						+	+	+	2
				92%		60%	40%	36%	
+3	**CAA(Gln)/CAG(Gln)**	++	9	0	1	-/+	-	-	2
						-/+	-/+	-	1
						-/+	-/+	-/+	1
						-/+	+	-/+	1
						+	-/+	-	2
						+	-	-	2
						+	+	+	1
				91%		67%	43%	25%	
+4	**ATG(Met)/TTG(Leu)**	-	7	8	0	+	+	+	7
				47%		100%	100%	100%	
+6	**ATG(Met)/ATA(Ile)**	-	0	14	1	+	-/+	-	1
				6%		100%	50%	0%	
+6	**ATG(Met)/ATT(Ile)**	-	6	13	2	-/+	-	-	1
						-/+	-/+	-	1
						+	-	-	4
						+	-/+	-/+	1
						+	+	+	1
				35%		80%	30%	22%	
+9	**TCT(Ser)/TCG(Ser)**	-	5	5	1	-/+	-	-	1
						+	-	-	2
						+	+	-/+	1
						+	+	+	2
				50%		86%	50%	43%	
+9	**TCT(Ser)/TCC(Ser)**	+++	11	0	1	-/+	-/+	-	1
						-/+	-	-/+	2
						-	+	+	1
						+	-	-	1
						+	+	+	7
				92%		73%	69%	71%	
	***C. reinhardtii* Wild** × ***C. smithii* Wild**	++	10	0	0	-	+	-	1
						-	+	+	2
						+	-/+	-	1
						+	+	-	2
						+	+	+	4
				100%		70%	91%	60%	

(1) '+++' designates about 80% to complete digestion of the substrate DNA by *in vitro* cleavage assay (Kurokawa *et al*., 2005), '++' about 50%-60% cleavage, '+' about 5%-25%, '-' almost no digestion. (2) Intron invasion ratio (see Materials and Methods for the definition). (3) Coconversion frequency of the genetic marker (see Materials and Methods for the definition). *Nhe*I*, *Nco*I*, *Hpa*I* show the diagnostic restriction sites in *Chlamydomonas smithii* mitochondrial genome.

is that all of the mutated sequences that are cleaved moderately or efficiently *in vitro* [*i.e.,* the cleavage frequency of the substrate is more than 50 % (designated as '++' or '+++' in Table 2-2)] also showed a high frequency of alpha-intron invasion (83-100 %). On the other hand, one apparent inconsistency between the *in vitro* and *in vivo* data concerned the mutation GGC/GGG(Gly) at position -1. This mutant target site showed only a low amount of cleavage in the *in vitro* experiment, while here it was efficiently invaded *in vivo* at a frequency as high as 92 %. Intron invasion ratio was 100 % when wild-type *C. smithii* and *C. reinhardtii* were crossed (Table 2-2).

Coconversion of genetic markers associated with intron invasion

Next, I determined whether there is a correlation between the coconversion of the target site-flanking regions and intron invasion. For this purpose, I took advantage of the presence of restriction enzyme sites for *Nhe*I, *Nco*I and *Hpa*I. The *C. smithii* mitochondrial genome contains at least one unique restriction site for each of these restriction enzymes, while they do not exist in the *C. reinhardtii* genome in the region surrounding the alpha-intron insertion site. *Nhe*I* is located about 500 bases upstream of the target site between the alpha-intron and the left-arm of the *C. smithii* mitochondrial genome, while *Nco*I* and *Hpa*I* are located downstream of the alpha-intron, about 1.4 kb and 5.0 kb from the intron insertion site, respectively (Fig. 2-1). For the intron-invaded mitochondrial genome, the upstream and downstream regions surrounding the alpha-intron were amplified by PCR. The upstream 1 kb PCR product was incubated with the restriction enzyme *Nhe*I, while the 5.5 kb downstream PCR product was incubated with *Nco*I or *Hpa*I (Fig. 2-4). Restriction enzyme analysis indicated a high frerquency of coconversion for *Nhe*I* and *N*coI*, irrespective of the type of mutation at the target site, while the coconversion of *Hpa*I*, which is the most downstream marker, showed a significantly lower frequency of coconversion (Table 2-2). This suggests that exonucleolytic degradation progresses bi-directionally away from the cleavage point before repair synthesis begins in the mitochondrial matrix. Interestingly, when the intron invasion frequency was very low, coconversion of the farthest genetic marker *Hpa*I* was not detected [*i.e.,* the *Hpa*I* coconversion ratio was null for ATG(Met)/ATA(Ile) (6 % intron invasion frequency) and TGG(Trp)/TGT(Cys) (31 % intron invasion frequency)].

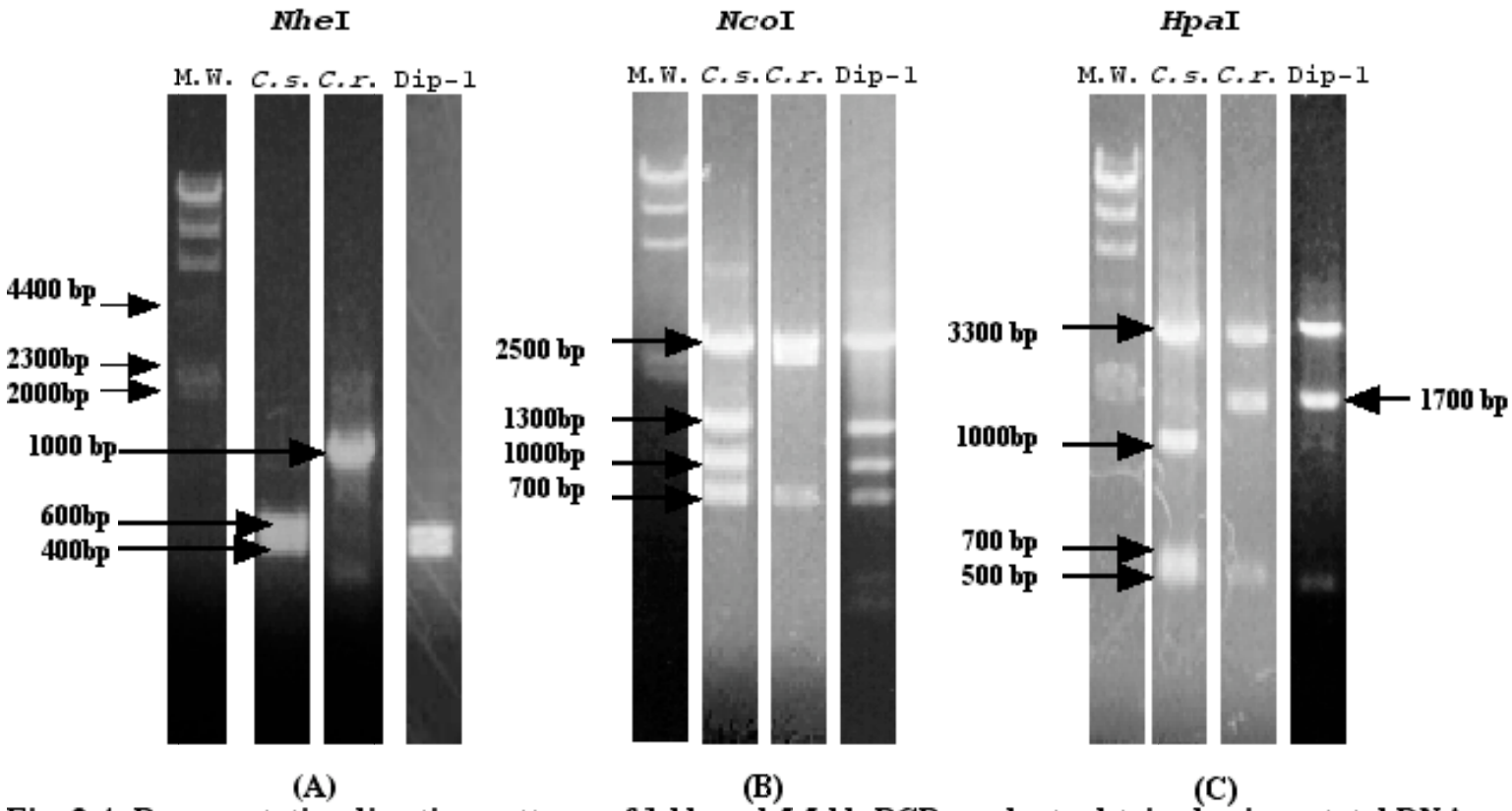


Fig. 2-4. Representative digestion patterns of 1 kb and 5.5 kb PCR products obtained using a total DNA prepared from various hybrid *Chlamydomonas* diploid cells.

(A)Restriction patterns of 1 kb PCR products by restriction enzyme *Nhe*I; (B) Restriction patterns of 5.5 kb PCR products by restriction enzyme *Nco*I; (C)Restriction patterns of 5.5 kb PCR products by restriction enzyme *Hpa*I. One of the hybrid cells, Dip-1 showed NheI*/+, NcoI*/+, and HpaI*/- (see text and Table 2 for detail).
M.W., Molecular weight marker; C.s.; wild-type *Chlamydomonas smithii*, C.r.; wild-type *Chlamydomonas reinhardtii*, Dip-1; hybrid diploid of *C. reinhardtii* and *C. smithii*.

Discussion

The native I-*Csm*I ORF is fused to its preceding exon and the mature N-terminus of I-*Csm*I has not yet been identified because of the difficulties in isolating this very low expression-level enzyme. Therefore, the enzymes used in the previous *in vitro* assay were recombinant polypeptides containing different N-termini [I-*Csm*I(200) and I-*Csm*I(217)] (Kurokawa *et al*., 2005). I-*Csm*I(200) is composed of 200 amino acid residues, while I-*Csm*I(217) is 217 amino acid residues. These recombinant I-*Csm*Is were demonstrated that their recognition specificity is tuned to efficiently cleave the most likely encountered target sequences in nature (*e.g.,* most of the tolerated variations occur at the codon third positions in the target site). Therefore, it is likely that our previous *in vitro* data on recognition properties of I-*Csm*I reflect the basic characteristics of I-*Csm*I. It is also likely that apparent inconsistency in substrate cleavage and intron invasion frequencies is a result of the differences in the method of assay, not the nature of the participating homing enzyme, whether recombinant or native. The assembly of some accessory proteins with I-*Csm*I in the mitochondrial matrix might explain this inconsistency. I have no solid data to exclude this possibility, however the interpretation described below seems reasonable enough to explain the inconsistency without involving any accessory elements.

The *in vitro* cleavage assay was performed in 50 μl of a reaction mixture containing 25 mM NaCl, 5 mM $MgCl_2$, 50 mM Hepes (pH 7.0), 1 mM dithiothreitol, 0.01 % (w/v) bovine serum albumin, and 1.0 μg of purified recombinant I-*Csm*I(200) or I-*Csm*I(217) (Kurokawa *et al*., 2005). After incubating the reaction at 30 ˚C for 6-8 h, the cleaved and uncleaved substrates were separated on an agarose gel and quantified. It is possible that substrates having a low amount of cleavage *in vitro* could remain uncleaved even after an infinitely long incubation time, because I-*Csm*I may lose its cleavage activity before binding to the target region. The substrate DNA trapped by the inactive enzyme could remain uncleaved. Actually, at high concentrations of both the enzyme and substrate, I observed the accumulation of high molecular weight aggregates composed of substrate DNA and homing enzyme in our previous *in vitro* assay (Kurokawa *et al*., 2005). On the other hand, newly translated I-*Csm*I might be continuously supplied to the mitochondrial matrix, and inactive I-*Csm*I bound to the

target region might be removed from the DNA when the mitochondrial genome replicates. This scenario may explain how a target sequence that cleaves at only 50 % was almost entirely invaded by the alpha-intron *in vivo*. For example, substrate DNA carrying the mutations CCA/CCG(Pro) or CCA/CCC(Pro) were cleaved at a frequency of 50 % (as designated '++' in Table 2-2) *in vitro*, but was invaded by the alpha-intron with a frequency of 100 % or 91 %, respectively. On the other hand, four kinds of transformants containing TGG(Trp)/TGT(Cys) [almost no cleavage *in vitro*], ATG(Met)/TTG(Leu) [almost no cleavage *in vitro*], ATG(Met)/ATT(Ile) [almost no cleavage *in vitro*], or TCT(Ser)/TCG(Ser) [5-25 % cleavage *in vitro*] mutations, showed an intermediate intron invasion frequency *in vivo* (31-50 %). It is likely that these target sequences were cleaved inefficiently even in the context of the mitochondrial matrix, because of the very limited recognition by I-*Csm*I.

Table 2-2 shows coconversion of genetic markers from *C. smithii* to *C. reinhardtii* in associated with alpha-intron invasion. The extent of coconversion was often as far as 5.0 kb from the cleavage site. This result coincides with earlier results obtained for wild-type *C. reinhardtii* and *C. smithii* (Remacle and Matagne, 1993). The large region of coconversion in *Chlamydomonas* mitochondria is in contrast to the shorter region of coconversion observed for the omega intron in *Saccharomyces cerevisiae*. In the case of the omega group I intron, found in the mitochondrial large subunit of the rRNA gene, coconversion region between omega-plus and omega-minus *S. cerevisiae* strains is limited to a few hundred bases long (Jacquier and Dujon, 1985; Muscarella and Vogt, 1989). The region of coconversion appears to depend on the size of the gap produced after cleavage of the target sequence by the homing enzyme. Therefore, the implicated exonucleases (Huang *et al*., 1999) (*e.g.,* DNA exonuclease A-like enzyme) might be much more active in *C. reinhardtii* than in *S. cerevisiae*. Unfortunately, no other report is available describing the amount of gene conversion that occurs in association with intron invasion. If the wide range of coconversion observed in *Chlamydomonas* is common for many other organisms, then the methods used to determine the relationship of organisms based on the comparison of organella DNA sequences may be flawed. This effect may be amplified when the organellar genome being analyzed contains many group I introns, as in fungi.

In this study, I performed crosses between *C. smithii* (mt^+, $alpha^+$, *Nhe*I*, *Nco*I*,

*Hpa*I*) and *C. reinhardtii* (mt^-, $alpha^-$) to probe the frequency of intron invasion. Therefore, the mitochondrial genome in the hybrid diploid, which contains all of the mt^+ genetic markers (*Nhe*I*, *Nco*I*, *Hpa*I* and the alpha-intron), has two possibilities concerning its origin; if the *C. smithii* mitochondrial genome or the *C. reinhardtii* genome were subjected to a larger degree of coconversion accompanying alpha-intron invasion. Remacle and Matagne (1993) have shown that successful cleavage of mt^- mitochondrial genomes is essential in developing a diploid cell that is composed of only mt^+ type mitochondrial genomes by crossing alpha-plus mt^+-*C. smithii* with alpha-minus mt^--*C. reinhardtii*. Considering this, the 5.5 kb PCR product containing the physical markers, *Nhe*I*, *Nco*I*, *Hpa*I* and alpha-intron is likely the result of a large amount of coconversion associated with intron invasion, but not a result of the remaining *C. smithii* mt^+ mitochondrial genome itself. Therefore, the corresponding PCR products were considered results of intron invasion in this study.

In diploids that are a result of artificial somatic cell fusion between wild-type $alpha^+$-*C. smithii* and $alpha^-$-*C. reinhardtii,* the alpha-intron is transmitted to *C. reinhardtii* as efficiently as that observed in sexually developed diploid cells (Remacle and Matagne, 1993). This demonstrates that the expression level of the homing enzyme in the somatic cell is not very different from that in the zygote. Considering the limited variation in the target sequence and the ability to recognize degenerated target sequences, which was systematically investigated *in vitro*, the principle bottleneck for lateral intron transmission is not likely related to the presence of a cleavable target sequence in the genome. Furthermore, in this investigation I showed that inter-specific transmission of a group I intron is easier than expected from the results of the *in vitro* cleavage assay. Also, the tendency to exclude the intron-invaded mitochondrial genomes was very weak, judging from the heteroplasmy of intron-plus and minus molecules even after two-rounds of single-colony isolation. All together, it is plausible that the bottleneck of intron invasion across species is the rare occurrence of inter-specific cell fusion. The discovery of a common group I intron sequence in both a plant-parasitic fungus and its host (Nishida and Sugiyama, 1995) supports the direct cell to cell transmission model of group I introns.

Conclusion

Enzymatic characterization of intronic endonuclease I-*Csm*I, which is a member of LAGLIDADG homing enzymes, was carried out using a recombinant enzyme produced in *E. coli* and variously mutated synthetic substrates. *In vitro* assay using the synthetic double strand DNA fragment revealed that divalent cation such as Mg^{2+} is essential, while it does not require ATP. Kinetic analyses of the enzyme showed that it has very high affinity to the substrate molecule and quite slow turn over ratio. These characteristics were similar to other representative LAGLIDADG homing enzymes. The cleaved double strand DNA was 3'-overhangs of four nucleotides. Analyses of cleavage efficiency for variously mutated substrates uncovered that the recognition region of I-*Csm*I reside between –5 and +11 nucleotide in relation to the intron insertion site. Generally intronic homing enzymes show ambiguity for their recognition nucleotide, however it is limited to mutation that might be tolerated in nature. Contrastingly, I-*Csm*I cleaved target sequences that contain highly likely non-tolerable ones in nature. This is probably due to I-*Csm*I is a bi-functional enzyme. It works not only as a homing enzyme but also as a protein to assist the folding of intron into the correct three dimensional structure.

The base positioned at +6 in relation to the intron insertion site was recognized most strictly *in vivo* among the mutated sites investigated in this study: Two types of mutations positioned at +6, ATG(Met) to ATA(Ile) and to ATT(Ile), resulted in very limited invasion ratio, 6 % and 35 % respectively. Position +6 is very close to the cleavage point of the coding strand, which is between +5 and +6. This is consistent with the result of *in vitro* assay, which showed the prominent tendency that mutations located around the coding strand cleavage point are severely recognized. On the other hand, types of mutated sequences that are cleaved moderately or efficiently *in vitro*, *i.e.*, the substrates of which cleavage ratio is more than 50 %, were invaded by alpha-intron with very high frequency (83-100 %) *in vivo*.

Considering the limited variation in the target sequence and the prominent ability to recognize the degenerated target sequence, which was systematically investigated *in vitro*, the principle bottleneck for the lateral intron transmission is most

probably not the absence of the cleavable target sequence in a genome. Furthermore, this investigation showed that direct cell to cell transmission of group I intron is easier than expected form the *in vitro* cleavage assay. All together, the bottleneck of intron invasion across species is plausibly the opportunity of interspecific cell fusion.

Acknowledgement

I am very grateful my advisor Professor Takeshi Ohama. Without his generous and appropriate advices I would not have been accomplished this work. I thank Professor Shigeyuki Yokoyama (Tokyo University, RIKEN), Yoshitaka Bessho (RIKEN) and Hitoshi Miyasaka (Kansai Electric Power Company Inc.) for cooperation. I also thank Professors Yoshihiro Matsuda (Kobe University) and Tatsuaki Saito (Okayama University of Science) for advices. I am indebted to B.A. Yoshihiro Adachi, T. Komatsu, C. Nakagoshi and S. Minakuchi (Kochi University of Technology) for their technical support in determining the I-*Csm*I cleavage points and making point mutation bearing DNA-constructs. This work was supported by grants from the Science Research Promotion Fund and the Sasagawa Scientific Research Grant from The Japan Science Society.

Finally, I am deeply indebted to my family and friends for their support and encouragement through this endeavor.

References

Aagaard, C., Awayez, M.J. and Garrett, R.A. 1997. Profile of the DNA recognition site of the archaeal homing endonuclease I-*Dmo*I. Nucl. Acids Res. 25: 1523-1530.

Beckers, M.C., Munaut, C., Minet, A. and Matagne, R.F. 1991. The fate of mitochondrial DNAs of mt^+ and mt^- origin in gametes and zygotes of *Chlamydomonas*. Curr. Genet. 20: 239-243.

Belfort, M. and Perlman, P.S. 1995. Mechanisms of intron mobility. J. Biol. Chem. 270: 30237-30240.

Belfort, M. and Roberts, R.J. 1997. Homing endonucleases: keeping the house in order. Nucl. Acids Res. 25: 3379-3388.

Boynton, J.E., Harris, E.H., Burkhart, B.D., Lamerson, P.M. and Gillham, N.W. 1987. Transmission of mitochondrial and chloroplast genomes in crosses of *Chlamydomonas*. Proc. Natl. Acad. Sci. USA 84: 2391-2395.

Chevalier, B.S. and Stoddard, B.L. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucl. Acids Res. 29: 3757-3774.

Cho, Y., Qiu, Y.L., Kuhlman, P. and Palmer, J.D. 1998. Explosive invasion of plant mitochondria by a group I intron. Proc. Natl. Acad. Sci. USA 95: 14244-14249.

Colleaux, L., Michel-Wolwertz, M.R., Matagne, R.F. and Dujon, B. 1990. The apocytochrome *b* gene of *Chlamydomonas smithii* contains a mobile intron related to both *Saccharomyces* and *Neurospora* introns. Mol. Gen. Gnet. 223: 288-296.

Cummings, D.J., McNally, K.L., Domenico, J.M. and Matsuura, E.T. 1990. The complete

DNA sequence of the mitochondrial genome of *Podospora anserina*. Curr. Genet. 17: 375-402.

Dalgaard, J.Z., Garrett, R.A. and Belfort, M. 1994. Purification and characterization of two forms of I-*Dmo*I, a thermophilic site-specific endonuclease encoded by an archaeal intron. J. Biol. Chem. 269: 28885-28892.

Ehara, M., Watanabe, K.I. and Ohama, T. 2000. Distribution of cognates of group II introns detected in mitochondrial *cox*1 genes of a diatom and a haptophyte. Gene 256: 157-167.

Geese, W.J., Kwon, Y.K., Wen, X. and Waring, R.B. 2003. *In vitro* analysis of the relationship between endonuclease and maturase activities in the bi-functional group I intron-encoded protein, I-*Ani*I. Eur. J. Biochem. 270: 1543-1554.

Gillham, N.W. 1978. Organelle heredity. Raven Press, New York.

Gimble, F.S. 2001. Degeneration of a homing endonuclease and its target sequence in a wild yeast strain. Nucl. Acids Res. 29: 4215-4223.

Goddard, M.R. and Burt, A. 1999. Recurrent invasion and extinction of a selfish gene. Proc. Natl. Acad. Sci. USA 96: 13880-13885.

Gormman, D.S. and Levine, R.P. 1965. Cytochlome F and plastocyanin: Their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci .USA 54: 1665-1669.

Hanes, C.S. 1932. Studies on plant amylases. I. The effect of starch concentrations upon the velocity of hydrolysis by the amylase of germinated barley. Biochem. J. 26: 1406-1421.

Hanson, D.K., Lamb, M.R., Mahler, H.R. and Perlman, P.S. 1982. Evidence for translated

intervening sequences in the mitochondrial genome of *Saccharomyces cerevisiae*. J. Biol. Chem. 257: 3218-3224.

Harris, E.H. 1989. The *Chlamydomonas* Sourcebook. Academic Press, New York.

Harris, E.H., Burkhart, B.D., Gillham, N.W. and Boynton, J.E. 1989. Antibiotic resistance mutations in the chloroplast 16S and 23S rRNA genes of *Chlamydomonas reinhardtii*: correlation of genetic and physical maps of the chloroplast genome. Genetics 123: 281-292.

Ho, Y., Kim, S.J. and Waring, R.B. 1997. A protein encoded by a group I intron in *Aspergillus nidulans* directly assists RNA splicing and is a DNA endonuclease. Proc. Natl. Acad. Sci. USA 94: 8994-8999.

Huang, Y.J., Parker, M.M. and Belfort, M. 1999. Role of exonucleolytic degradation in group I intron homing in phage T4. Genetics 153: 1501-1512.

Jacquier, A. and Dujon, B. 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41: 383-394.

Kurokawa, S., Bessho, Y., Higashijima, K., Shirouzu, M., Yokoyama, S., Watanabe, K.I. and Ohama, T. 2005. Adaptation of intronic homing endonuclease for successful horizontal transmission. FEBS J. 272: 2487-2496.

Lambowitz, A.M. 1989. Infectious introns. Cell 56: 323-326.

Lumbreras, V., Stevens, D.R. and Purton, S. 1998. Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. Plant J. 14: 441-448.

Martin, N.C. and Goodenough, U.W. 1975. Gametic differentiation in Chlamydomonas reinhardtii. I. Production of gametes and their fine structure. J. Cell Biol. 67: 587-605.

Matagne, R.F., Rongvaux, D. and Loppes, R. 1988. Transmission of mitochondrial DNA in cross involving diploid gametes homozygous or heterozygous for the mating-type locus in *Chlamydomonas*. Mol. Gen. Genet. 214: 257-262.

Monteilhet, C., Dziadkowiec, D., Szczepanek, T. and Lazowska, J. 2000. Purification and characterization of the DNA cleavage and recognition site of I-*Sca*I mitochondrial group I intron encoded endonuclease produced in *Escherichia coli*. Nucl. Acids Res. 28: 1245-1251.

Muscarella, D.E. and Vogt, V.M. 1989. A mobile group I intron in the nuclear rDNA of Physarum polycephalum. Cell 56: 443-454.

Nakamura, S., Aoyama, H. and van, Woesik, R. 2003. Strict paternal transmission of mitochondrial DNA of *Chlamydomonas* species is explained by selection against maternal nucleoids. Protoplasma 221: 205-210.

Newman, A.K., Rubin, R.A., Kim, S.H. and Modrich, P. 1981. DNA sequences of structural genes for *Eco* RI DNA restriction and modification enzymes. J. Biol. Chem. 256: 2131-2139.

Nishida, H. and Sugiyama, J. 1995. A common group I intron between a plant parasitic fungus and its host. Mol. Biol. Evol. 12: 883-886.

Nishioka, M., Fujiwara, S., Takagi, M. and Imanaka, T. 1998. Characterization of two intein homing endonucleases encoded in the DNA polymerase gene of *Pyrococcus kodakaraensis* strain KOD1. Nucl. Acids Res. 26: 4409-4412.

Pellenz, S., Harington, A., Dujon, B., Wolf, K. and Schafer, B. 2002. Characterization of the I-*Spom* I endonuclease from fission yeast: insights into the evolution of a group I intron-encoded homing endonuclease. J. Mol. Evol. 55: 302-313.

Randolph-Anderson, B.L., Boynton, J.E., Gillham, N.W., Harris, E.H., Johnson, A.M.,

Dorthu, M.P. and Matagne, R.F. 1993. Further characterization of the respiratory deficient *dum-1* mutation of *Chlamydomonas reinhardtii* and its use as a recipient for mitochondrial transformation. Mol. Gen. Genet. 236: 235-244.

Remacle, C., Bovie, C., Michel-Wolwertz, M.R., Loppes, R. and Matagne, F. 1990. Mitochondrial genome transmission in *Chlamydomonas* diploids obtained by cross and artificial fusions: Role of the mating type and of a 1kb intron. Mol. Gen. Genet. 223: 180-184.

Remacle, C. and Matagne, R.F. 1993. Transmission, recombination and conversion of mitochondrial markers in relation to the mobility of a group I intron in *Chlamydomonas*. Curr. Genet. 23: 518-525.

Sambrook, J. and Russell, D.W. 2001. Molecular Cloning a laboratory manual Third Edition. Cold Spring Harbor: Cold Spring Hatbor Laboratory Press.

Sargueil, B., Hatat, D., Delahodde, A. and Jacq, C. 1990. *In vivo* and *in vitro* analyses of an intron-encoded DNA endonuclease from yeast mitochondria. Recognition site by site-directed mutagenesis. Nucl. Acids Res. 18: 5659-5665.

Schafer, B., Merlos-Lange, A.M., Anderl, C., Welser, F., Zimmer, M. and Wolf, K. 1991. The mitochondrial genome of fission yeast: inability of all introns to splice autocatalytically, and construction and characterization of an intronless genome. Mol. Gen. Genet. 225: 158-167.

Schafer, B., Wilde, B., Massardo, D.R., Manna, F., Del Giudice, L. and Wolf, K. 1994. A mitochondrial group-I intron in fission yeast encodes a maturase and is mobile in crosses. Curr. Genet. 25: 336-341.

Szczepanek, T. and Lazowska, J. 1996. Replacement of two non-adjacent amino acids in the *S. cerevisiae* bi2 intron-encoded RNA maturase is sufficient to gain a homing-endonuclease activity. EMBO J. 15: 3758-3767.

Takahashi, Y., Utsumi, K., Yamamoto, Y., Hatano, A. and Satoh, K. 1996. Genetic engineering of the processing site of D1 precursor protein of photosystem II reaction center in *Chlamydomonas reinhardtii*. Plant. Cell Physiol. 37: 161-168.

Turmel, M., Otis, C., Cote, V. and Lemieux, C. 1997. Evolutionarily conserved and functionally important residues in the I-*Cue*I homing endonuclease. Nucl. Acids Res. 25: 2610-2619.

Watanabe, K.I., Ehara, M., Inagaki, Y. and Ohama, T. 1998. Distinctive origins of group I introns found in the COXI genes of three green algae. Gene 213: 1-7.

Wernette, C.M. 1998. Structure and activity of the mitochondrial intron-encoded endonuclease, I-*Sce*IV. Biochem. Biophys. Res. Commun. 248: 127-133.

Wernette, C.M., Saldahna, R., Perlman, P.S. and Butow, R.A. 1990. Purification of a site-specific endonuclease, I-*Sce* II, encoded by intron 4 alpha of the mitochondrial *cox*I gene of *Saccharomyces cerevisiae*. J. Biol. Chem. 265: 18976-18982.

Wernette, C., Saldanha, R., Smith, D., Ming, D., Perlman, P.S. and Butow, R.A. 1992. Complex recognition site for the group I intron-encoded endonuclease I-*Sce*II. Mol. Cell Biol. 12: 716-723.

Yamasaki, T., Kurokawa, S., Watanabe, K.I., Ikuta, K. and Ohama, T. 2005. Shared molecular characteristics of successfully transformed mitochondrial genomes in *Chlamydomonas reinhardtii*. Plant Mol. Biol. 58: 515-527.

Zamora, I., Feldman, J.L. and Marshall, W.F. 2004. PCR-based assay for mating type and diploidy in *Chlamydomonas*. Biotechniques 37: 534-536.

Vice-theses

(1) Adaptation of intronic homing endonuclease for successful horizontal transmission (2005, FEBS J.)

(2) Shared molecular characteristics of successfully transformed mitochondrial genomes in *Chlamydomonas reinhardtii* (2005, Plant Mol. Biol.)

the FEBS Journal

Adaptation of intronic homing endonuclease for successful horizontal transmission

Sayuri Kurokawa[1], Yoshitaka Bessho[2], Kyoko Higashijima[2], Mikako Shirouzu[2,3], Shigeyuki Yokoyama[2,3,4], Kazuo I. Watanabe[5] and Takeshi Ohama[1]

1 Graduate School of Engineering, Department of Environmental Systems Engineering, Kochi University of Technology (KUT), Kochi, Japan
2 RIKEN Genomic Sciences Center, Tsurumi, Yokohama, Japan
3 RIKEN Harima Institute at SPring-8, Mikazuki-cho, Sayo, Hyogo, Japan
4 Graduate School of Science, University of Tokyo, Bunkyo, Tokyo, Japan
5 Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX, USA

Keywords
I-*Csm*I; *Chlamydomonas smithii*; horizontal transmission; group I intron; selfish element

Correspondence
T. Ohama, Department of Environmental Systems Engineering, Kochi University of Technology (KUT), Tosayamada, Kochi 782–8502, Japan
Fax: +81 887 572520
Tel: +81 887 572512
E-mail: ohama.takeshi@kochi-tech.ac.jp

(Received 10 January 2005, revised 26 February 2005, accepted 18 March 2005)

doi:10.1111/j.1742-4658.2005.04669.x

Group I introns are thought to be self-propagating mobile elements, and are distributed over a wide range of organisms through horizontal transmission. Intron invasion is initiated through cleavage of a target DNA by a homing endonuclease encoded in an open reading frame (ORF) found within the intron. The intron is likely of no benefit to the host cell and is not maintained over time, leading to the accumulation of mutations after intron invasion. Therefore, regular invasional transmission of the intron to a new species at least once before its degeneration is likely essential for its evolutionary long-term existence. In many cases, the target is in a protein-coding region which is well conserved among organisms, but contains ambiguity at the third nucleotide position of the codon. Consequently, the homing endonuclease might be adapted to overcome sequence polymorphisms at the target site. To address whether codon degeneracy affects horizontal transmission, we investigated the recognition properties of a homing enzyme, I-*Csm*I, that is encoded in the intronic ORF of a group I intron located in the mitochondrial COB gene of the unicellular green alga *Chlamydomonas smithii*. We successfully expressed and purified three types of N-terminally truncated I-*Csm*I polypeptides, and assayed the efficiency of cleavage for 81 substrates containing single nucleotide substitutions. We found a slight but significant tendency that I-*Csm*I cleaves substrates containing a silent or tolerated amino acid change more efficiently than non-silent or nontolerated ones. The published recognition properties of I-*Spom*I, I-*Sca*I, and I-*Sce*II were reconsidered from this point of view, and we detected proficient adaptation of I-*Spom*I, I-*Sca*I, and I-*Sce*II for target site sequence degeneracy. Based on the results described above, we propose that intronic homing enzymes are adapted to cleave sequences that might appear at the target region in various species, however, such adaptation becomes less prominent in proportion to the time elapsed after intron invasion into a new host.

Various molecular phylogenetic analyses suggest that group I introns in fungi and terrestrial/nonaquatic plants were horizontally transmitted multiple times in the course of evolution among distantly related species [1–3]. We have shown this is also the case for algal mitochondrial introns [4,5]. For reasons yet unknown,

Abbreviations
cob, apocytochrome *b*; nt, nucleotide(s); ORF, open reading frame.

the distribution of group I introns is strongly biased, most commonly found in fungi (e.g. the *cox-1* of *Podospora anserina* contains 15 group I introns [6]). About half of group I introns contain an open reading frame (ORF) that encodes a DNA sequence specific endonuclease (intronic homing enzyme). These intronic homing enzymes cleave a target sequence that is usually 16–30 base pairs (bp) long and nonpalindromic (reviewed in [7]). Cleavage of the chromosome initiates repair of the damaged DNA through homologous recombination. Consequently, after the repair, the donor intronic DNA is copied into the recipient chromosome. Thus, homing endonucleases are essential for horizontal transmission of group I introns. Organelle introns are highly likely of no benefit to the host, i.e. they are thought to be selfish and parasitic elements that spread in populations. Therefore, when they integrate into the host genome, there is little or no selection for maintaining endonuclease function. Moreover, if there is any cost to the host cell for producing a functional endonuclease, then selection will work to fix the nonfunctional element. Therefore, regular horizontal transmission of an intron to a new species before its functional deterioration seems essential for its evolutionary long-term persistence. As an example, comprehensive analyses of the group I intron omega (also known as *Sc* LSU.1), which was first found in the *Saccharomyces cerevisiae* mitochondrial large subunit rRNA gene, clearly showed repeated horizontal transmissions, and the interval between the complete loss and reinvasion of the intron is estimated to be about 5.7 million years [8]. This leads to the hypothesis that intronic homing enzymes might be adapted to recognize variously degenerated target sequences among a wide range of organisms.

In addition to intronic homing enzymes, highly specific endonuclease activity is also detected among inteins, which are thought to be parasitic elements that exhibit horizontal transmission. Regular invasional transmission is likely essential for both homing introns and inteins. In fact, for the target site of intein homing endonuclease PI-*Sce*I, which is found in *Saccharomyces cerevisiae* vacuolar membrane H^+-ATPase, all of the nine nucleotides essential for the cleavage were mapped on the conserved codon first and second positions, and target sequence variations at codon third positions were tolerated for the endonuclease recognition [9]. On the other hand, the adaptations that permit efficient horizontal transfer of intronic homing enzymes have not been analyzed. To date, only three intronic homing enzymes that target a sequence within protein coding genes were investigated systematically for their recognition sequence ambiguity, i.e. I-*Spom*I that is encoded as an intronic ORF of a group I intron in the *Schizosaccharomyces pombe* COXI gene [10,11], I-*Sca*I is in the COB gene of *Saccharomyces capensis* [12,13], and I-*Sce*II is in the COXI gene of *Saccharomyces cerevisiae* [14–16]. To address the question, we investigated the recognition sequence of I-*Csm*I, including its degeneracy. I-*Csm*I is a homing enzyme encoded in the group I intron (named alpha or *Cs cob*.1) located in the apocytochrome b (COB) gene of the unicellular alga *C. smithii* [17]. This enzyme has the typical two LAGLIDADG motifs. The intronic ORF is probably translated as a fusion protein with the preceding exon, and the N-terminal peptide may be proteolytically removed to become an active form as seen in I-*Spom*I [18]. Endonuclease activity of I-*Csm*I has been observed through artificial interspecific cell fusion between intron-bearing *C. smithii* and *C. reinhardtii* that lacks the intron in its COB gene [19]. However, systematic analysis of the target sequences and the homing endonuclease's enzymatic properties have not been previously attempted. We overproduced several N-terminally truncated I-*Csm*I polypeptides in *Escherichia coli*, and determined cleavable target sequences through an *in vitro* assay of substrates containing 81 different point mutations.

Based on the analyses of I-*Csm*I and these three intronic homing enzymes, we discuss the adaptation for successful horizontal transfer. Investigations performed for the intronic homing enzymes that have a recognition sequence in ribosomal RNA genes are less informative to answer our questions and are not considered in this paper.

Results

Activity of the N-terminal truncated I-CsmI polypeptides

Three N-terminally truncated I-*Csm*I polypeptides [I-*Csm*I(200), I-*Csm*I(217), and I-*Csm*I(237); the number in parentheses indicates the amino acid encoded in the ORF] were purified and yielded about 6 mg of protein per 1 g wet weight *E. coli*, while the entire I-*Csm*I ORF (i.e. I-*Csm*I(374)), which contains the upstream COB exon, did not express even after several modified conditions were tested (Fig. 1). We assayed the endonuclease activity of recombinant I-*Csm*I(200), I-*Csm*I(217), and I-*Csm*I(237) using linearized pCOB1.8Kb as a substrate. I-*Csm*I(200) is the smallest homing endonuclease containing two LAGLIDADG motifs analyzed to date. It is even smaller than the type II restriction enzyme *Eco*RI [20], which is a 277 amino acid homodimer that cleaves a symmetric six

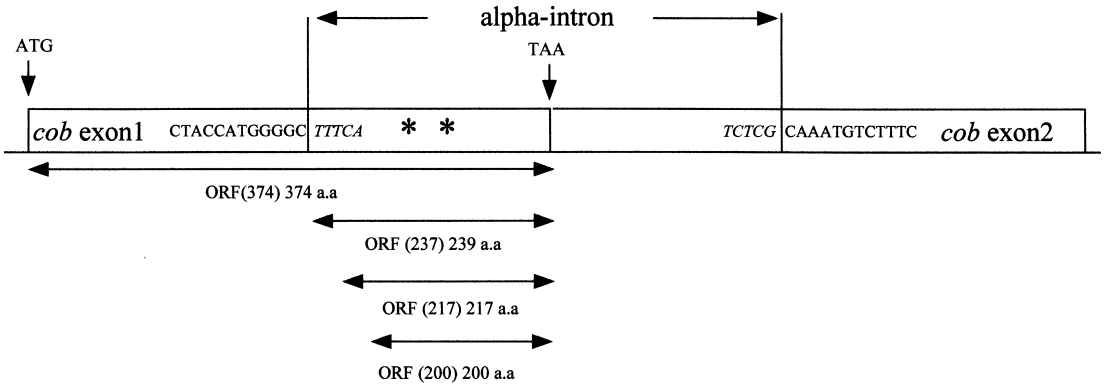


Fig. 1. Schematic of open reading frames that encode whole I-*Csm*I or N-terminally truncated I-*Csm*I polypeptides. I-*Csm*I is denoted as a fusion protein with the preceding apocytochrome *b* gene exon encoded polypeptide. The target sequence of I-*Csm*I and the bordering intron sequences are shown in upright and italicized characters, respectively. Asterisks show the position of the LAGLIDADG motifs. a.a, amino acid residues.

base restriction site. Recombinant proteins I-*Csm*I(200) and I-*Csm*I(237) cleaved the substrate at the expected target site, yielding two fragments of 1.2 kb and 3.7 kb in size. For I-*Csm*I(217), the quantity of protein was reduced from 1.5 to 1.0 μg and the incubation period was shortened from 24 to 6 h to reduce the amount of insoluble reaction products. Under these modified conditions, I-*Csm*I(217) exhibited sequence specific endonuclease activity.

To determine the optimal conditions for endonuclease activity, we tested the effect of Na^+ and Mg^{2+} concentration, pH, and temperature. The optimal pH for all three proteins was around 7.0 (Fig. 2A). The optimal Na^+ and Mg^{2+} concentrations were 25 mM and 5 mM, respectively, for both I-*Csm*I(237) and I-*Csm*I(217) (Fig. 2B,C). In contrast, 75 mM Na^+ and 10 mM Mg^{2+} were optimal for I-*Csm*I(200). The optimal reaction temperature was 35 °C for both I-*Csm*I(200) and I-*Csm*I(237), and 30 °C for I-CsmI(217) (Fig. 2D) A higher concentration of Mg^{2+} was progressively detrimental to all I-*Csm*I polypeptides. The presence of Mg^{2+} was essential for the endonuclease activity as a cofactor, while the same concentration of Mn^{2+} (5 mM) reduced the enzyme activity to 15%, and no activity was observed with 5 mM of Zn^{2+}, Ca^{2+}, or Co^{2+} (data not shown).

Kinetic parameters of I-CsmI(200)

We determined the kinetic parameters of I-*Csm*I(200) based on the data obtained by time course monitoring of the cleaved products in various concentrations of the linearized substrate pCOB1.8Kb. The K_m, V_{max}, k_{cat} were 2.5×10^{-9} M, 1.8×10^{-12} M·s^{-1}, and 4.7×10^{-4} s^{-1}, respectively. These parameters were similar to other representative intronic LAGLIDADG homing endonucleases (e.g. I-*Ceu*I [21], I-*Sce*IV [22], I-*Dmo*I [23,24]) that show characteristics of high affinity to the substrate DNA and slow turnover (Table 1).

Essential target region

Digestion was not observed using pC-18nt and pC-20nt, while almost complete cleavage was observed for pC-24nt by I-*Csm*I(200). This suggests that the recognition region of I-*Csm*I(200) resides between 12 nt upstream (+) and 12 nt downstream (–) of the intron insertion site, while 10 nt upstream and 10 nt downstream is insufficient for recognition.

Cleavage point and mutational analysis of cleavable sequences

The precise cleavage site on each strand was determined through DNA sequencing of the substrate whose termini were blunt-ended by T4 DNA polymerase treatment. It became clear that cleavage occurs five nt downstream of the intron insertion site on the coding strand and one nt downstream of the insertion site on the noncoding strand, creating 3′ overhangs of four nt (Fig. 3). This terminal overhang is typical for DNA cleaved by LAGLIDADG homing enzymes.

Eighty-one variants (104 bp each) containing single nt substitutions between −12 and +15 were assayed to discern the critical nucleotides involved in recognition. Positions −5 through −3, +2, and +6 through +8 are strictly recognized by I-*Csm*I(200) and I-*Csm*I(217), as the original bases are essential for cleavage, while any substitution was permitted for positions −12 through −6 and +12 through +15 (some examples of cleaved

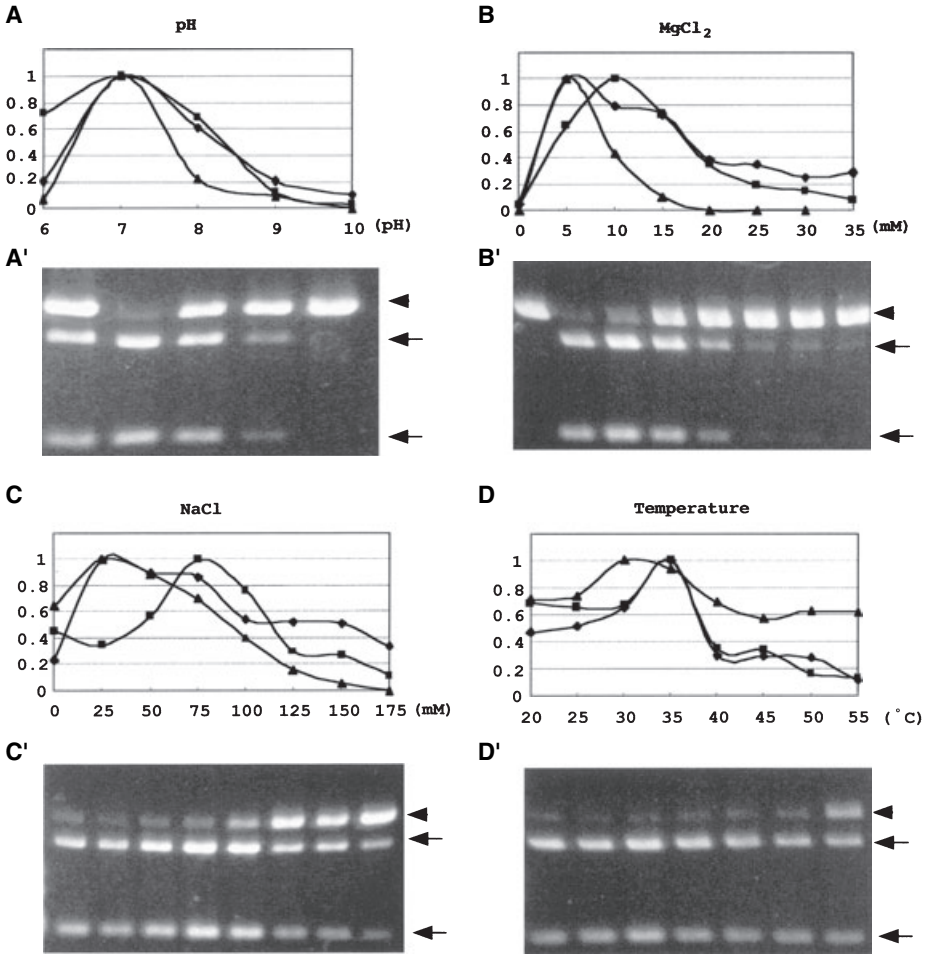


Fig. 2. Effects of pH, Mg^{2+}, Na^{+} and temperature on the substrate cleavage reaction using recombinant homing enzyme I-*Csm*I polypeptides. The conditions used to assay enzyme cleavage were as described in Experimental procedures. ◆, reaction with recombinant protein I-*Csm*I(237); ▲, I-*Csm*I(217); ■, I-*Csm*I(200). Vertical axis of each graph (A–D) shows relative activity. The electrophoresis patterns of substrate cleavage by I-*Csm*I(200) are shown in (A′–D′). Each lane in the agarose gel corresponds to a specific condition denoted in the axis of abscissa shown above the graph. An arrowhead denotes the position of the original substrate, while arrows show the cleaved substrates.

Table 1. Kinetic properties of intronic LAGLIDADG endonucleases. n.d., Not determined.

	I-*Csm*I	I-*Ceu*I	I-*Sce*IV	I-*Dmo*I
K_m	2.5×10^{-9} M	0.9×10^{-9} M	$0.14–0.77 \times 10^{-9}$ M	4×10^{-9} M
V_{max}	1.8×10^{-12} M·s^{-1}	n.d.	$0.9–1.5 \times 10^{-10}$ M·s^{-1}	n.d.
k_{cat}	4.7×10^{-4} s^{-1}	3.7×10^{-5} s^{-1}	$3–6 \times 10^{-4}$ s^{-1}	8.3×10^{-3} s^{-1}
Number of motif per peptide	Two	One	Two	Two

pattern are shown in Fig. 4). The majority of substitutions that blocked substrate cleavage were between −5 and +11 in relation to the intron insertion site. Therefore, the span of critical bases are not centered at the intron insertion site, but are spread almost symmetrically with respect to the cleavage points of coding and noncoding strands. A summary of substrate cleavability is classified into four groups (+++, ++, +, and –; see Experimental procedures for details) and shown in Fig. 3. As a result of cleavage with I-*Csm*I(200), 26% (8%), 21% (26%), 15% (14%), and 38% (51%) kinds of substrates were classified into four classes, +++, ++, +, and –, respectively [the results of I-*Csm*I(217) are shown in parentheses]. I-*Csm*I(200)

		-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	+15
		C	T	A	C	C	A	T	G	G	G	G	C	C	A	A	A	T	G	T	C	T	T	T	C	T	G	G
		(Leu)				(Pro)			(Trp)			(Gly)			(Gln)			(Met)			(Ser)			(Phe)			(Trp)	
T	I-CsmI(217)	+++	/	++	++	++	++	/	–	–	–	–	+++	+++	–	–	–	/	–	/	–	/	/	/	++	/	N/D	N/D
	I-CsmI(200)	+++	/	+++	++	++	++	/	–	–	–	++	+++	++	–	–	–	/	–	/	–	/	/	/	++	/	+++	+++
C	I-CsmI(217)	/	+++	+	/	/	++	+	–	–	–	–	/	/	–	–	–	–	–	–	/	++	–	+++	/	N/D	N/D	N/D
	I-CsmI(200)	/	+	++	/	/	++	+	–	–	–	–	/	/	–	–	+	+	–	–	/	+++	+	++	/	+++	+++	+++
A	I-CsmI(217)	++	++	/	++	+	/	+	–	–	–	–	++	–	/	/	/	–	–	–	–	–	+	–	+++	N/D	N/D	N/D
	I-CsmI(200)	+++	+	/	++	++	/	++	–	–	–	+++	+++	–	/	/	/	–	–	–	–	–	+	–	+++	+++	+++	+++
G	I-CsmI(217)	++	++	++	++	+	++	+	/	/	/	/	–	+	–	+	++	–	/	–	–	–	+	++	++	N/D	/	/
	I-CsmI(200)	+++	+++	++	++	+	++	+	/	/	/	/	+	+	–	++	+++	–	/	–	–	–	+	++	+++	+++	/	/

Fig. 3. Mutational analyses of the recognition efficiency by recombinant homing enzymes I-*Csm*I(200) and I-*Csm*I(217). The coding sequence of *C. reinhardtii cob* and the assigned amino acids are shown on top. Bases corresponding to the codon third position are shown with underline. The three possible base substitutions for each position are indicated to the left side. An arrowhead indicates the intron insertion site. An arrow with a dotted line shows the cleavage site of the noncoding strand, while an arrow with solid line denotes the cleavage site for the coding strand. The numbering is in relation to the intron insertion site. '+ + +', substrate cleavage above the wild-type levels (more than 150%); '+ +', cleavage almost the same or slightly less than the wild-type levels (120–80%); '+', cleavage below the wild-type levels (50–20%); '–' almost no cleavage (less than 10%); '/', position of the wild-type nucleotide. N/D; not determined.

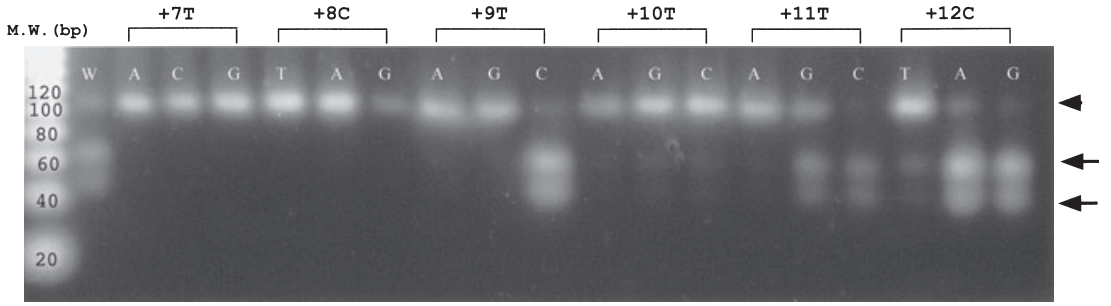


Fig. 4. Cleavage pattern of linearized substrates containing single base substitutions by I-*Csm*I(200). The numbering is in relation to the intron insertion site, with '+' indicating upstream, followed by the nucleotide that is the original base at the given position, while the nucleotide denoted below shows the base after substitution. M.W., 20 bp molecular mass marker ladder. An arrowhead indicates the position of substrate DNA (104 bp), while arrows indicate the positions of cleaved substrate (60 and 44 bp). W; substrate DNA containing the *Chlamydomonas reinhardtii* wild-type *cob* sequence.

and I-*Csm*I(217) showed almost identical sequence recognition properties (Fig. 3). A prominent difference in cleavage efficiency was observed for only two substitutions, the original G at position −2 for A and T. I-*Csm*I(217) did not cleave these mutated substrates, whereas I-*Csm*I(200) cleaved both, with the G to A mutation the most efficient of the two (Fig. 3).

Correlation between the type of amino acid substitution and cleavage efficiency

We analyzed whether there is any correlation between the type of amino acid substitution induced by single nt substitution (silent/tolerated change, or nonsilent/nontolerated change) and how efficiently the substrates are cleaved by two kinds of N-terminal truncated I-*Csm*I polypeptides. A survey of GenBank registered sequences of various organisms showed the target DNA sequences of I-*Csm*I, I-*Spom*I, I-*Sce*II, and I-*Sca*I correlate to the amino acid sequences YGQMS(F/H), TGWT(A/V)PPL, FGHPEV, and W(G/A)TVI, respectively. Therefore, F/H, A/V, and G/A amino acid changes at the specific sites were functionally tolerated in this investigation.

Forty-eight substrates containing single nt substitutions at the core recognition region (between −5 and +11) were analyzed from this point of view.

Substrates containing a silent or tolerated amino acid change

Seven of 48 substrates contained a silent amino acid change. However, two of seven such substrates [containing TCT(Ser) changed to TCA and TCG(Ser), mutation position +9 in Fig. 3] were not cleaved at all by I-*Csm*I(217) and I-*Csm*I(200), and additionally the substrate contains the change GGC(Gly) to GGG(Gly) (position −1) was not cut by I-*Csm*I(217) even though these silent changes must be tolerated in nature. On the other hand, three silent substrates [TCT(Ser) to

TCC(Ser), position +9; GGC(Gly) to GGT/GGA (Gly), position −1] were cut efficiently by the two I-*Csm*I polypeptides. Additionally, CAA (Gln) to CAG (Gln) (position +3) was efficiently cut by I-*Csm*I(200).

Substrates containing a nonsilent or nontolerated amino acid change

Forty-one of 48 substitutions caused nonsilent/nontolerated amino acid changes. Showing an adaptation to the possible target DNA sequences, I-*Csm*I polypeptides only slightly cleaved most of them (Table 2). Such property is also prominently detected in I-*Spom*I and I-*Sca*I. However, TAA(Stop) instead of CAA(Gln) (position +1), TGC(Cys) and TCC(Ser) instead of TTC(Phe) (position +11) were efficiently cleaved by the both I-*Csm*I enzymes, even though these codons are not observed at these positions in nature. In contrast, none of the nonsilent/nontolerated substitutions were cleaved efficiently by I-*Sca*I (Table 2).

Discussion

The original I-*Csm*I ORF is fused with the preceding exon, which is not rare for group I intronic ORFs. The entire ORF of I-*Spom*I also extends into the upstream exon of the COXI gene, and it has been reported that the N-terminal truncated polypeptide, including the two LAGLIDADG motifs, has similar sequence specificity to that detected using mitochondrial extracts [11]. Considering the above, we tried to overproduce three kinds of N-terminally truncated recombinant I-*Csm*I polypeptides that retain the two LAGLIDADG motifs instead of the entire I-*Csm*I (374 amino acid) (Fig. 1), because we failed to express the whole I-*Csm*I ORF for reasons that are unclear. We found that all of the N-terminal truncated I-*Csm*I polypeptides retain the specificity to cleave the target site, and the kinetic parameters of I-*Csm*I(200) are very similar to that reported for representative intronic homing enzymes of LAGLIDADG motifs (Table 1). The optimal conditions of selected factors were also very similar to other homing enzymes, with the exception of the preferred pH. I-*Csm*I displayed its highest activity at pH 7.0, which is very close to the reported physiological pH value of 7.5 in yeast mitochondria [25], while most of the LAGLIDADG enzymes show their highest activity at an alkaline pH between 8.5 and 9.5 (e.g. optimal pH is 2.9 for I-AniI [26], and between 8.5 and 9.0 for the recombinant I-*Sca*I [13]). Having a host pH that is lower than the optimum pH observed for many homing enzymes may act to reduce endonuclease activity and prevent overdigestion of the genomic DNA.

I-*Csm*I(200)'s optimal conditions for Na^+ and Mg^{2+} are clearly shifted to a concentration higher than that of I-*Csm*I(217) and I-*Csm*I (237) (Fig. 2B,C). This suggests that the three-dimensional conformation of this enzyme is different from the others possibly because of the recessed N-terminal region, and may explain the differences in cleavage activity between I-*Csm*I(200) and I-*Csm*I(217). I-*Csm*I(200) seems to tolerate a higher degree of sequence ambiguity than I-*Csm*I(217) at position −2, because I-*Csm*I(200) can efficiently cleave the mutated substrates of −2 A and −2T (instead of the original −2G), while I-*Csm*I(217) only tolerates the original base −2G (Fig. 3).

Cleavage of a target DNA is an essential step for lateral transfer of an intron. Therefore, if a homing enzyme shows very stringent recognition of the target

Table 2. Type of amino acid substitution contained in the substrate and the cleavage efficiency. Efficiently cleaved: efficiency more than 80% of the wild type substrate for I-*Spom*I and I-*Csm*I, while more than 78% for I-*Sce*II; for I-*Sca*I, efficiency of originally described as 'mutant cleaved as well as the wild type'. Moderately cleaved: 80–30% of the wild-type substrate for I-*Spom*I and I-*Csm*I, while 60–42% for I-*Sce*II; for I-*Sca*I, efficiency of originally described as 'reduced cleavage'. Not or scarcely cleaved: less than 30% of the wild-type substrate for I-*Spom*I and I-*Csm*I, while 33% for I-*Sce*II; and for I-*Sca*I, efficiency of originally described as 'no cleavage'.

Type of substitution	Homing endonuclease	Efficiently cleaved %	Moderately cleaved %	Not or scarcely cleaved %
Silent or tolerated amino acid changes	I-*Spom*I	67 (4/6)	33 (2/6)	0 (0/6)
	I-*Sca*I	13 (1/8)	88 (7/8)	0 (0/8)
	I-*Sce*II	100 (7/7)	0 (0/7)	0 (0/7)
	I-*Csm*I(217)	43 (3/7)	14 (1/7)	43 (3/7)
	I-*Csm*I(200)	57 (4/7)	14 (1/7)	29 (2/7)
Non-silent or non-tolerated amino acid changes	I-*Spom*I	16 (3/19)	21 (4/19)	63 (12/19)
	I-*Sca*I	0 (0/22)	32 (7/22)	68 (15/22)
	I-*Sce*II	28 (9/32)	44 (14/32)	28 (9/32)
	I-*Csm*I(217)	7 (3/41)	10 (4/41)	83 (34/41)
	I-*Csm*I(200)	15 (6/41)	15 (6/41)	71 (29/41)

core sequence, this step could be a bottleneck for horizontal transmission of an intron. The target site of I-*Csm*I corresponds to the amino acid sequence of Trp-Gly-Gln-Met-Ser-(Phe/His). This is a highly conserved region in COB genes among a wide range of organisms. Our systematic induction of a point mutation and the cleavage assay showed a clear tendency that I-*Csm*I polypeptides efficiently cleave silent change containing substrates than nonsynonymous/nontolerated change containing ones (Table 2).

It is obvious that stop codons are never tolerated at the internal regions of a gene. However, our systematic induction of a point mutation introduced stop codons, i.e. TGA and TAG stop codons from TGG(Trp), and TAA stop codon from CAA(Gln). The substrate DNA that contains TGA or TAG was not cleaved, while the substrate containing a TAA stop codon was efficiently cleaved by the both I-*Csm*I polypeptides (Fig. 3). Moreover, substrates including a codon that highly likely appears in nature were not cleaved [e.g. TCA/TCG(Ser) from TCT(Ser), and three Ile codons AT(T/C/A) from ATG(Met)]. The above instances indicate that the recognition property of I-*Csm*I is not skillfully adapted to recognize target sequences that are highly likely to appear in nature.

It is possible that the recognition property of I-*Spom*I, I-*Sca*I, and I-*Sce*II are adapted to recognize multiple possible target sequences, because these homing enzymes cleaved substrates containing various kinds of silent/tolerated amino acid changes efficiently, and none of them were remained uncleaved (Table 2).

Considering the above, we propose that homing enzymes are adapted to recognize diverse target sequences to facilitate horizontal transmission to a new species, as evidently seen with I-*Spom*I, I-*Sca*I, and I-*Sce*II. However, immediately after a successful invasion, mutations begin to accumulate that lead to a loss of further adaptation, because homing endonuclease activity is only essential for intron invasion and thereafter it is useless to the cell. Invasion of I-*Csm*I might be evolutionarily older than the other three homing enzymes compared in this study, because I-*Csm*I showed the least adapted properties among the four. Actually, remnants of homing endonuclease ORFs that include frame shifts or stop codons within the ORF are frequently found (e.g. [4]). Comprehensive analysis of omega homing endonuclease and its associated group I intron revealed that it is more common to find an inactive intron/ORF combination than it is to find an active intron/ORF combination or an intron-less allele [8].

It has been proved that some of intronic homing enzymes are bifunctional. They work not only as an endonuclease but also as a maturase to preserve splicing. The bifunctional activity of I-*Spom*I [18], I-*Sca*I [12], and I-*Ani*I [27] has been observed. I-*Csm*I could also be a bifunctional protein that acts as a maturase, which may also preserve its endonuclease activity for horizontal transmission. These bifunctional enzymes are recognized as intermediates, and may likely lose their endonuclease activity over time, retaining only their maturase activity [4,26].

Experimental procedures

Cloning and expression of wild type and N-terminally truncated I-CsmI ORFs

The entire COB gene and the alpha intron were amplified by PCR using total *C. smithii* (CC-1373) DNA as a template. We also used PCR to isolate the wild-type 374 amino acid I-*Csm*I ORF (i.e. ORF(374)) and three N-terminally truncated ORFs, ORF(200), ORF(217) and ORF(237) (the number in parentheses indicates the amino acid encoded in the ORF). These four ORFs have different N-termini, however, share the common wild-type stop codon. The two sets of primers used to amplify the original I-*Csm*I ORF(374) and ORF(237), contained *Xho*I sites at their tails. Forward primer containing an *Nde*I site, and reverse primers containing an *Fba*I site were used to amplify ORF(200) and ORF(217). After restriction enzyme digestion, the ORF(374) and ORF(237) PCR products were cloned into the *Xho*I site of pET19b (Novagen, CA, USA) in frame with a sequence encoding the 10-histidine tag, while ORF(200) and ORF(217) were cloned into the *Nde*I/*Bam*HI site of pET15b (Novagen, CA, USA) in-frame with a His_6 tag. The resulting plasmids were amplified in *E. coli* DH5α and *E. coli* BL21 CodonPlus (DE3) RIL (Stratagene, CA, USA) was for protein expression.

Expression and purification of whole or truncated I-CsmI polypeptides

Cultures containing whole or truncated ORFs were undertaken at 37 °C in 2.0 L of LB broth containing 100 $\mu g \cdot mL^{-1}$ ampicillin and 34 $\mu g \cdot mL^{-1}$ chloramphenicol until $D_{600} = 0.6$. Protein expression was induced by addition of isopropyl thio-β-D-galactoside (0.1 mM final). The cells were incubated at 30 °C for an additional 4 h, collected by centrifugation, and resuspended in 40 mL of sonication buffer [50 mM Hepes (pH 7.0), 400 mM NaCl, 6 mM 2-mercaptoethanol, and 20 $\mu g \cdot mL^{-1}$ lysozyme] and sonicated on ice. The lysate was centrifuged for 2 h at 10 000 ***g*** and the supernatant was loaded onto a Ni-NTA column (5 mL bed volume) (Qiagen, CA, USA) that was previously equilibrated with the wash buffer [50 mM Hepes (pH 7.0), 400 mM NaCl, 6 mM 2-mercaptoethanol, and 10 mM imidazole]. The column was washed with 50 mL of the

wash buffer, and the protein was eluted with 100 mL of the elution buffer [50 mM Hepes (pH 7.0), 400 mM NaCl, 6 mM 2-mercaptoethanol, and 200 mM imidazole]. Homogeneity was assessed after staining with SDS/PAGE/Coomassie brilliant blue R-250. The products of ORF(200), ORF(217), ORF(237), and ORF(374) were named I-*Csm*I(200), I-*Csm*I(217), I-*Csm*I(237) and I-*Csm*I(374), respectively.

Reaction conditions to estimate the minimum target-site length

Substrate DNA

Chemically synthesized DNA fragments, which consist of 18, 20, or 24 nt symmetrically spanning the alpha intron insertion point of the *C. reinhardtii* COB gene, were cloned into the *Eco*RV site of the pCITE-4a + (Novagen, CA, USA). These plasmids were named pC-18nt, pC-20nt, and pC-24nt (the number indicates the length of the inserted DNA fragment). The plasmids were first linearized by *Sca*I digestion, and then used as a substrate to determine the region encompassing the recognition sequence.

Reaction conditions

Linearized substrate (1.5 μg) described above was added to 50 μL of the reaction mixture containing [50 mM Hepes (pH 7.0), 0.01% bovine serum albumin, 1 mM dithiothreitol, 25 mM NaCl, and 5 mM $MgCl_2$] and about 1 μg of recombinant homing enzyme I-*Csm*I(237). The reaction was carried out at 25 °C for 24 h and 10 μL was loaded onto an 0.8% agarose gel to resolve the products.

Reaction to determine the cleavage point and its terminal shape

We determined the terminal shape of the substrate following the T4 DNA polymerase method by Nishioka *et al.* [28]. pC-24nt (2.0 μg) digested with I-*Csm*I(237) was recovered from an 0.8% agarose gel by electro-elution and then treated with T4 DNA polymerase (Takara Bio, Kyoto, Japan) in the presence of 0.2 mM dNTPs. The DNA mixture was then treated with T4 DNA ligase (Takara Bio) for self-ligation and transformed into *E. coli*. Nucleotide sequence analysis of the plasmid was performed to determine the nature of cohesive termini generated by I-*Csm*I(200).

Reaction conditions used to investigate the effect of Na^+, divalent cations, pH, and temperature

Substrate DNA fragment

A 1.8 kb DNA fragment, containing the entire COB gene of *C. reinhardtii* (CC-124) and its flanking regions, was cloned into pT7-Blue2 vector (Novagen, CA, USA) and named pCOB1.8Kb. After linearization by *Not*I, the plasmid was used as a substrate for the reaction described below.

Reaction mixture

A 50 μL reaction mixture [25 mM NaCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, 0.01% (v/v) bovine serum albumin, 50 mM Tris/HCl (pH 7.0)] was used, which contained 0.5 μg of linearized pCOB1.8Kb and 1.0 μg of I-C*sm*I(217), or 1.5 μg of I-C*sm*I(200) or I-*Csm*I(237). One of the parameters [i.e. pH, NaCl concentration, species of divalent cations (5 mM), $MgCl_2$ concentration, or the temperature] in the reaction was altered to determine optimal conditions. Reagents used to make the buffers of specific pH value are as follows; Mes for pH 6.0, Hepes for pH 7.0, Tris for pH 8.0 and 9.0, TAPS for pH 10.0. The reaction was incubated for 24 h with I-*Csm*I(237) and I-*Csm*I(200), and incubated for 6 h with I-*Csm*I(217), which reduced the formation of aggregates observed with this protein. The reaction products were resolved in an 0.8% agarose gel, and stained with ethidium bromide. The relative quantities of the digested fragments were calculated using the NIH IMAGE program version 1.61.

Assay of cleavable DNA sequences

A limited part of the *C. reinhardtii* COB gene, which is 104 nt long and containing the I-*Csm*I target sequence, was chemically synthesized and converted to double strand DNA. This double-stranded DNA fragment was used as a control to compare the cleavage efficiency of various substrates containing single mutations. Each one of the 27 nucleotides composing the target site was changed to the other three possible nucleotides utilizing PCR primers containing a specific mutation. These 81 DNA fragments, each containing single point mutations were used for a detailed analysis of substrate cleavage. One hundred and fifty nanograms of each substrate was digested with 1 μg of I-*Csm*I(200) in the reaction mixture [50 mM Hepes (pH 7.0), 0.01% (v/v) bovine serum albumin, 1 mM dithiothreitol, 25 mM NaCl, 5 mM $MgCl_2$] at 30 °C for 8 h. Electrophoresis of the samples was performed on a 3% agarose gel, and stained by 10 000-fold diluted SYBR Green I dye (Molecular Probes, OR, USA) for 40 min (SDS/heat-treatment of samples before electrophoresis, described below, was omitted for a clearer image, without affecting the results). The image was developed using LAS-1000 image analyzer (Fuji Film Co., Tokyo, Japan). The cleavage ratio, i.e. cleaved vs. uncleaved fragments, was quantified by NIH Image and compared to wild-type substrate cleavage (i.e. native *C. reinhardtii cob* sequence carrying substrate). The 81 substrates were grouped into four classes based on the following: (a) The substrate much better than the control

(the cleavage ratio of mutated substrate vs. control is more than 1.5) is denoted as + + +; (b) The substrate as good as the control (i.e. the ratio is between 1.2 and 0.8) is denoted as + +; (c) The substrate less efficiently cleaved (i.e. the ratio is between 0.5 and 0.2) is denoted as +; (d) Scarcely cleaved substrate (i.e. the ratio is below 0.1) is denoted as –.

Reaction conditions to measure the kinetic parameters

Linearized pCOB1.8Kb and a plasmid containing the N-terminally truncated homing endonuclease, I-*Csm*I(200), was used to measure the kinetic parameters. Two hundred and fifty microliters of reaction buffer [50 mM Hepes (pH 7.0), 0.01% (v/v) bovine serum albumin, 1 mM dithiothreitol, 25 mM NaCl, and 5 mM $MgCl_2$] contained 1 μg of the recombinant protein and between 0.5 $ng{\cdot}\mu L^{-1}$ and 10 $ng{\cdot}\mu L^{-1}$ of substrate. Twenty-microliter aliquots were removed at different time points from the reaction mixture, and terminated by the addition of 1 μL of 0.5 M EDTA and 1.25 μL of 10% sodium dodecyl sulfate, followed by heating the mixture to 50 °C for 5 min to completely denature the protein. Samples were electrophoresed on an 0.8% agarose gel, then visualized by 10 000-fold diluted SYBR Green I dye. Relative intensities of the digested fragment were quantified using the Las-1000 and NIH IMAGE. K_m, V_{max} and k_{cat} were determined through a Hanes–Woolf plot [29].

Acknowledgements

We thank Professors Yoshihiro Matsuda (Kobe University) and Tatsuaki Saito (Okayama University of Science) for advice, and B.Eng. Yoshihiro Adachi (Kochi University Tech) for his technical support in determining the I-*Csm*I cleavage points and Ms. Mariya Takeuchi for her encouragement. This work was supported by the Sasagawa Scientific Research Grant, and the Regional Science Promotion Program.

References

1 Lambowitz AM (1989) Infectious introns. *Cell* **56**, 323–326.
2 Belfort M & Roberts RJ (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res* **25**, 3379–3388.
3 Cho Y, Qiu YL, Kuhlman P & Palmer JD (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc Natl Acad Sci USA* **95**, 14244–14249.
4 Watanabe KI, Ehara M, Inagaki Y & Ohama T (1998) Distinctive origins of group I introns found in the COXI genes of three green algae. *Gene* **213**, 1–7.
5 Ehara M, Watanabe KI & Ohama T (2000) Distribution of cognates of group II introns detected in mitochondrial *cox*1 genes of a diatom and a haptophyte. *Gene* **256**, 157–167.
6 Cummings DJ, McNally KL, Domenico JM & Matsuura ET (1990) The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. *Curr Genet* **17**, 375–402.
7 Chevalier BS & Stoddard BL (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res* **29**, 3757–3774.
8 Goddard MR & Burt A (1999) Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci USA* **96**, 13880–13885.
9 Gimble FS (2001) Degeneration of a homing endonuclease and its target sequence in a wild yeast strain. *Nucleic Acids Res* **29**, 4215–4223.
10 Schafer B, Merlos-Lange AM, Anderl C, Welser F, Zimmer M & Wolf K (1991) The mitochondrial genome of fission yeast: inability of all introns to splice autocatalytically, and construction and characterization of an intronless genome. *Mol Gen Genet* **225**, 158–167.
11 Pellenz S, Harington A, Dujon B, Wolf K & Schafer B (2002) Characterization of the I-*Spom* I endonuclease from fission yeast: insights into the evolution of a group I intron-encoded homing endonuclease. *J Mol Evol* **55**, 302–313.
12 Szczepanek T & Lazowska J (1996) Replacement of two non-adjacent amino acids in the *S. cerevisiae* bi2 intron-encoded RNA maturase is sufficient to gain a homing-endonuclease activity. *EMBO J* **15**, 3758–3767.
13 Monteilhet C, Dziadkowiec D, Szczepanek T & Lazowska J (2000) Purification and characterization of the DNA cleavage and recognition site of I-*Sca*I mitochondrial group I intron encoded endonuclease produced in *Escherichia coli*. *Nucleic Acids Res* **28**, 1245–1251.
14 Hanson DK, Lamb MR, Mahler HR & Perlman PS (1982) Evidence for translated intervening sequences in the mitochondrial genome of *Saccharomyces cerevisiae*. *J Biol Chem* **257**, 3218–3224.
15 Sargueil B, Hatat D, Delahodde A & Jacq C (1990) *In vivo* and *in vitro* analyses of an intron-encoded DNA endonuclease from yeast mitochondria. Recognition site by site-directed mutagenesis. *Nucleic Acids Res* **18**, 5659–5665.
16 Wernette C, Saldanha R, Smith D, Ming D, Perlman PS & Butow RA (1992) Complex recognition site for the group I intron-encoded endonuclease I-*Sce*II. *Mol Cell Biol* **12**, 716–723.
17 Colleaux L, Michel-Wolwertz MR, Matagne RF & Dujon B (1990) The apocytochrome *b* gene of

Chlamydomonas smithii contains a mobile intron related to both *Saccharomyces* and *Neurospora* introns. *Mol Gen Gnet* **223**, 288–296.
18 Schafer B, Wilde B, Massardo DR, Manna F, Del Giudice L & Wolf K (1994) A mitochondrial group-I intron in fission yeast encodes a maturase and is mobile in crosses. *Curr Genet* **25**, 336–341.
19 Remacle C, Bovie C, Michel-Wolwertz MR, Loppes R & Matagne RF (1990) Mitochondriral genome transmission in *Chlamydomonas* diploids obtained by sexual crosses and artificial fusions: role of the mating type and of a 1 kb intron. *Mol Gen Genet* **223**, 180–184.
20 Newman AK, Rubin RA, Kim SH & Modrich P (1981) DNA sequences of structural genes for *Eco* RI DNA restriction and modification enzymes. *J Biol Chem* **256**, 2131–2139.
21 Turmel M, Otis C, Cote V & Lemieux C (1997) Evolutionarily conserved and functionally important residues in the I-*Cue*I homing endonuclease. *Nucleic Acids Res* **25**, 2610–2619.
22 Wernette CM (1998) Structure and activity of the mitochondrial intron-encoded endonuclease, I-*Sce*IV. *Biochem Biophys Res Commun* **248**, 127–133.
23 Dalgaard JZ, Garrett RA & Belfort M (1994) Purification and characterization of two forms of I-*Dmo*I, a thermophilic site-specific endonuclease encoded by an archaeal intron. *J Biol Chem* **269**, 28885–28892.
24 Aagaard C, Awayez MJ & Garrett RA (1997) Profile of the DNA recognition site of the archaeal homing endonuclease I-*Dmo*I. *Nucleic Acids Res* **25**, 1523–1530.
25 Wernette CM, Saldahna R, Perlman PS & Butow RA (1990) Purification of a site-specific endonuclease, I-*Sce* II, encoded by intron 4 alpha of the mitochondrial *cox*I gene of *Saccharomyces cerevisiae*. *J Biol Chem* **265**, 18976–18982.
26 Geese WJ, Kwon YK, Wen X & Waring RB (2003) *In vitro* analysis of the relationship between endonuclease and maturase activities in the bi-functional group I intron-encoded protein, I-*Ani*I. *Eur J Biochem* **270**, 1543–1554.
27 Ho Y, Kim SJ & Waring RB (1997) A protein encoded by a group I intron in *Aspergillus nidulans* directly assists RNA splicing and is a DNA endonuclease. *Proc Natl Acad Sci USA* **94**, 8994–8999.
28 Nishioka M, Fujiwara S, Takagi M & Imanaka T (1998) Characterization of two intein homing endonucleases encoded in the DNA polymerase gene of *Pyrococcus kodakaraensis* strain KOD1. *Nucleic Acids Res* **26**, 4409–4412.
29 Hanes CS (1932) Studies on plant amylases. I. The effect of starch concentrations upon the velocity of hydrolysis by the amylase of germinated barley. *Biochem J* **26**, 1406–1421.

Plant Molecular Biology (2005) 58:515–527
DOI 10.1007/s11103-005-7081-3

Shared molecular characteristics of successfully transformed mitochondrial genomes in *Chlamydomonas reinhardtii*

Tomohito Yamasaki[1], Sayuri Kurokawa[1], Kazuo I. Watanabe[2], Kyosuke Ikuta[3] and Takeshi Ohama[1,*]

[1]*Graduate School of Engineering, Department of Environmental Systems Engineering Kochi University of Technology (KUT), 782-8502, Tosayamada, Kochi, Japan (*author for correspondence; e-mail ohama.takeshi@kochi-tech.ac.jp);* [2]*Department of Chemistry and Biochemistry University of Texas, Austin, TX, 78712, USA;* [3]*Department of Biology Osaka Kyoiku University, 582-8582, Kashiwara, Osaka, Japan*

Received 15 March 2005; accepted in revised form 10 May 2005

Key words: Chlamydomonas reinhardtii, homologous recombination, mitochondrial transformation, respiratory deficient strains

Abstract

Three types of respiratory deficient mitochondrial strains have been reported in *Chlamydomonas reinhardtii*: a deficiency due to (i) two base substitutions causing an amino acid change in the apocytochrome *b* (COB) gene (i.e., strain named *dum-15*), (ii) one base deletion in the COXI gene (*dum-19*), or (iii) a large deletion extending from the left terminus of the genome to somewhere in the COB gene (*dum-1*, *-14*, and *-16*). We found that these respiratory deficient strains of *C. reinhardtii* can be divided into two groups: strains that are constantly transformable and those could not be transformed in our experiments. All transformable mitochondrial strains were limited to the type that has a large deletion in the left arm of the genome. For these mitochondria, transformation was successful not only with purified intact mitochondrial genomes but also with DNA-constructs containing the compensating regions. In comparison, mitochondria of all the non-transformable strains have both of their genome termini intact, leading us to speculate that mitochondria lacking their left genome terminus have unstable genomes and might have a higher potential for recombination. Analysis of mitochondrial gene organization in the resulting respiratory active transformants was performed by DNA sequencing and restriction enzyme digestion. Such analysis showed that homologous recombination occurred at various regions between the mitochondrial genome and the artificial DNA-constructs. Further analysis by Southern hybridization showed that the wild-type genome rapidly replaces the respiratory deficient monomer and dimer mitochondrial genomes, while the *E. coli* vector region of the artificial DNA-construct likely does not remain in the mitochondria.

Abbreviations: aad, aminoglycoside adenine transferase; *cob*, apocytochrome b; *cox1*, subunit 1 of cytochrome c oxidase; DIG, digoxigenin; *nd*, NADH dehydrogenase; non-WT, non-wild type; PCR, polymerase chain reaction; *rtl*, reverse transcriptase-like protein

Introduction

Artificial stable incorporation of foreign DNA into the mitochondrial genome has previously been demonstrated only in two species, the yeast *Saccharomyces cerevisiae* (Johnston *et al.*, 1988) and the unicellular green alga *Chlamydomonas reinhardtii* (Randolph-Anderson *et al.*, 1993). However, in the latter case, successful mitochondrial transformation has been demonstrated only

by introducing the entire wild-type mitochondrial genome into a respiratory deficient mutant *dum-1* strain (Randolph-Anderson *et al*., 1993), and rescue of strains containing other respiratory deficiencies (e.g., *dum-14*, *-15*, *-16*, and *-19*) were not attempted.

Curiously, no follow-up paper has been published, in contrast to the many papers that have been published since the first report of successful *C. reinhardtii* chloroplast transformation in 1988 (Boynton *et al*., 1988). Transformation of the *C. reinhardtii* chloroplast seems much easier than the mitochondria, because only a single large chloroplast exists in the cell (for a review, Harris, 2001). In addition to this, another significant reason for underutilization of this mitochondrial transformation system is probably due to the lack of a convenient transformation vector that can be amplified in *E. coli*. Therefore, we developed three kinds of artificial DNA-constructs to transform respiratory deficient *C. reinhardtii* strains. Various combinations of the DNA-constructs and the *dum* strains were tried to find out the most efficient set.

The mitochondrial genome of *C. reinhardtii* is a linear 15.8 kb molecule (Gray and Boer, 1988). A 0.5 kb inverted repeat sequence is present at the right and left termini of the linear mitochondrial DNA. It remains unclear what DNA configuration and mechanisms are involved in DNA replication. However, the replication origin (*ori*) of the *Chlamydomonas* mitochondrial genome is thought to be located in this inverted repeat sequence, because A + T rich clusters and numerous short direct and inverted repeats are detected in this region (Gray and Boer, 1988; Nedelcu and Lee, 1998). Both strands of mitochondrial DNA encode genetic information, and each strand may be organized into a single transcriptional unit, suggested by the mapping of two separate promoters in the gap between the *nd5* and *cox1*, one pointing in each direction.

We discovered that only a limited number of respiratory deficient strains are competent for mitochondrial transformation, and mitochondrial genome transitions were analyzed by Southern hybridization and restriction enzyme digestion. Here, we describe the common characteristics of the competent mitochondrial genomes, and suggest a plausible reason for their successful transformation.

Experimental procedures

Respiratory deficient strains

Respiratory deficient C. reinhardtii *strains used in this study are as follows*

dum-1 CC-2654 was obtained from the *Chlamydomonas* Genetic Center (c/o Dr. Elizabeth Harris, Department of Botany, Duke University, Durham, NC 27706, USA) and is a subclone of the original *dum-1* CC-2255 (mt^-) strain; wall-less strain *cw-15* CC-277 (mt^+) was also obtained from *Chlamydomonas* Genetic Center; *dum-14* (mt^-), *dum-15* (mt^-), *dum-16* (mt^-), and *dum-19* (mt^-) strains were kindly provided by Dr. R.F. Matagne (University of Liege, Belgium). The gene organization of wild-type mitochondria and respiratory deficient strains are shown schematically in Figure 1.

Preparation of recipient cells

Respiratory deficient *Chlamydomonas* strains were cultured under constant cool white fluorescent light (84 μmol m^{-2} s^{-1}) in Tris–acetate–phosphate (TAP) medium (Gorman and Levine, 1965) at 25 °C with vigorous shaking. About 7 ml of mid-log phase cultures ($2–3 \times 10^6$ cells/ml) were collected on a 9 cm diameter membrane (Hybond-N^+, Amersham, Burkinghamshire, England) by filtration, and transferred to a TAP-1% agar-plate containing 50 μg/ml ampicillin to reduce the possibility of bacterial contamination.

Conditions for biolistic bombardment

We used a PDS-1000/He particle delivery system (Bio-Rad, California, USA) to transform the mitochondria and chloroplast, following the standard procedures designed for bacteria (Smith *et al*., 1992). We tested gold particles of 40 nm (G-40, E-Y Laboratories, Inc. California, USA), 100 nm (EM. GC 100, British Biocell International, England), 200 nm (EM. GC 200, British Biocell International), 600 nm (Bio-Rad catalogue #165–2262), and also a mixture of 100 and 600 nm gold particles. Gold particle mixtures were prepared as follows: 3 mg of 600 nm gold particles and the sediment of the 100 nm gold particles (prepared from 1 ml colloidal suspension) were suspended together in 50 μl of 50% glycerol. Five μl of donor DNA (1 μg/μl), 50 μl of 2.5 M $CaCl_2$,

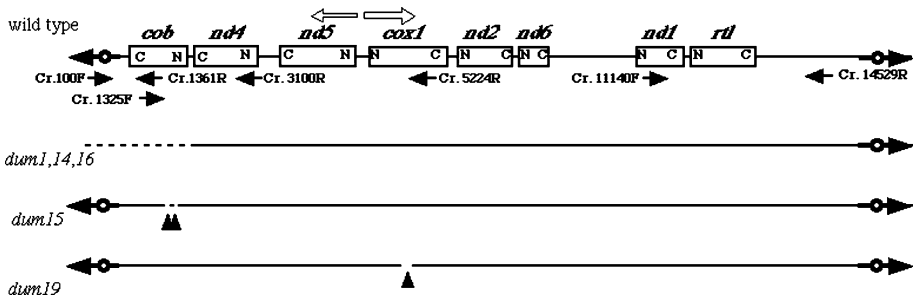


Figure 1. The relative gene organization and positions of mutations or deleted regions of respiratory deficient mitochondrial strains. Horizontal thick arrows indicate the long terminal inverted repeat, while open arrows designate the direction of transcription. Vertical arrow-heads identify the location of mutations. Open-circle indicate the expected replication origin (see text). C; C-terminus of the gene product. N; N-terminus of the gene product. Relative position and the direction of PCR primers are shown by short arrows.

and 20 μl of freshly prepared 0.1 M spermidine solutions (S-0266, Sigma, St. Louis, USA) were added to coat the gold particles with the DNA. The *Chlamydomonas* strains were bombarded with the gold particles using He at a pressure of 1100 pounds per square inch (1 psi = 6895 kPa) in a pressure reduction chamber at 25 inches Hg. After bombardment, respiratory competent transformants were selectively grown on a TAP-1% agar-plate and incubated under limited lighting condition [a 2:22 hour light (17 μmol m^{-2} s^{-1}): dark regimen] at 25 °C for 4 weeks. The exposure to limited light stimulates the growth of transformed strains.

Organization of transformed mitochondrial genomes

Culture growth conditions and total DNA preparation

Four weeks after biolistic-bombardment, individual colonies were inoculated into 20 ml (to prepare DNA for Southern hybridization) or 5 ml (to prepare DNA for PCR) of TAP medium. The cultures were grown for an additional 2 weeks under limited lighting (a 2:22 hour light (17 μmol m^{-2} s^{-1}): dark regimen) with shaking (Phase I culture). About 50 μl of the Phase I culture was inoculated into 20 ml of new TAP-medium (alternatively, 12 μl was inoculated into 5 ml of culture). The resulting culture (Phase II culture) was grown under the same conditions as the Phase I culture. This operation was repeated once more to obtain the Phase III culture.

The CTAB method (Murray and Thompson, 1980) was applied to separate total DNA from 20 ml cultures, and the Plant DNeasy mini kit (Qiagen, GmbH, Germany) was used for DNA isolation from the 5 ml cultures after disrupting the cells with glass beads.

Southern hybridization

One μg of intact DNA or *Sac*I-digested DNA were electrophoresed on a 0.8% agarose gel. The gel was transferred onto a nylon membrane and hybridized with alkali-unstable digoxigenin (DIG)-labeled probes, which were prepared following the manufacturer's recommendations (Roche, Mannheim, Germany). Four kinds of DIG-probes were used to analyze the mitochondrial genome. A 108 bp COB-probe, hybridizing to the middle of the COB gene (bases 1215–1323; numbering follows Gray and Boer 1988; Genbank acc. no. NC_001638), a 173 bp ND4-probe hybridizing to the middle part of the ND4 gene (2613–2786), and T7-promoter region and U19 primer region probes (120 bp each) were used to determine whether the pT7Blue-2 vector is retained in the transformants.

All hybridizations described above were performed at 44 °C for 18 hours. The membrane was washed once at 44 °C in a solution containing 2XSSC-0.1% SDS, followed by two consecutive washes in 0.1XSSC-0.1% SDS at 44 °C. The image was detected using a LAS-1000 (Fuji film Co., Tokyo, Japan).

Artificial DNA-constructs used for respiratory deficient mitochondrial transformation

5.0-kb construct

This construct consists of three PCR fragments, or elements, all of which were amplified from the total DNA of wild-type *C. reinhardtii* (CC-277,

cw-15 mt$^+$). The first element corresponds to bases 1–1780 of the *C. reinhardtii* mitochondrial genome, which contains the left inverted repeat sequence, the COB gene, and the C-terminal 67 bp of the NADH-cytochrome *c* reductase subunit 4 (*nd4*) gene [hereafter, C(*nd4*); Figures 1 and 2A]. The second element corresponds to bases 4780–6620, which carries the N-terminal 180 bp of the ND5 gene [hereafter, N(*nd5*)], along with the two expected promoters (one in each direction) and the entire COXI gene. The final element corresponds to bases 14400–15758, and contains the right inverted repeat sequence of the genome, expected region for DNA replication (*ori*), and modules of the fragmented rRNA genes. These three elements were assembled into one continuous DNA fragment (totally about 5.0-kb), whose gene organization is: left terminus-*cob*-C(*nd4*)/(*nd5*)N-two promoters-*cox1*/right terminus (Figure 2A), and ligated into the *Eco*RV site of the *E. coli* pT7Blue-2 T-vector (Novagen, Milwaukee, USA). This plasmid is designated p5.0 kb. The plasmid p5.0 kb was digested with *Not*I or *Bam*HI to linearize it, when it is required. This construct has potential to compensate the deficiencies of all *dum* strains used in this study, because this bears the left arm to the entire *cob* region and also the wild type COXI gene.

3.8-kb construct

This construct shares the first and the third elements of the 5.0-kb construct, while the middle element is 1.2-kb shorter than that for p5.0 kb, is comprised of bases 4780–5400 (Figure 2B), and contains the N-terminal region of the ND5 gene and the N-terminal 358 bp sequence of the COXI gene [hereafter N(*cox1*)]. These fragments were assembled as follows: left terminus-*cob*-C(*nd4*)/(*nd5*)N-two promoters-N(*cox1*)/right terminus, and ligated into the *Eco*RV site of the pT7Blue-2

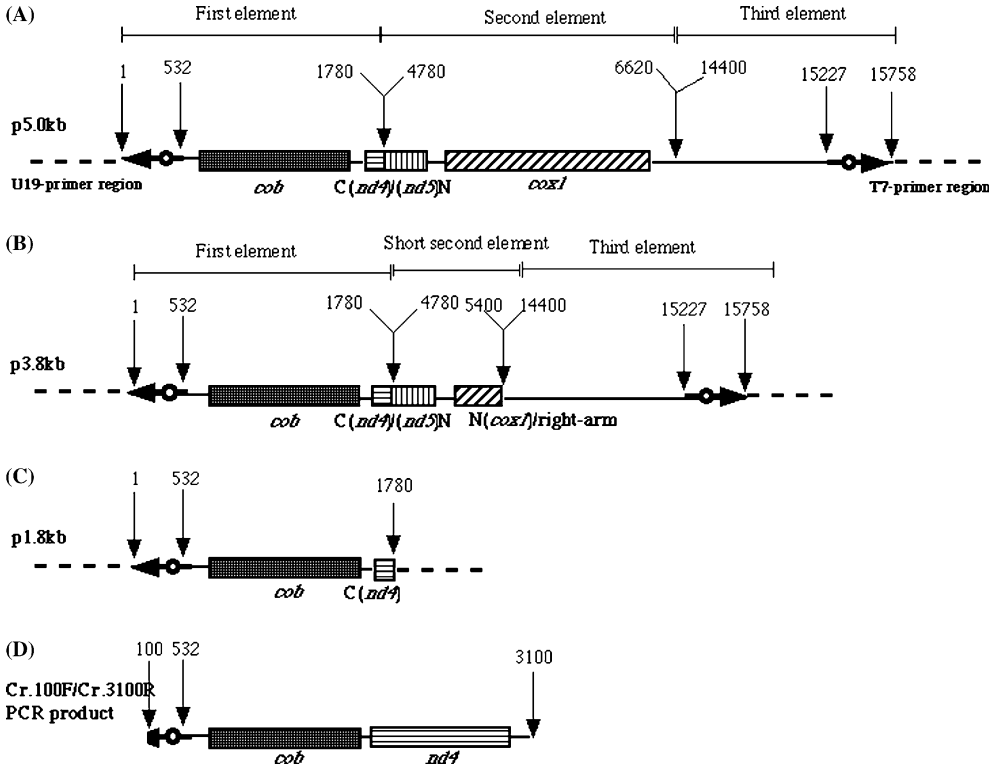


Figure 2. The gene organization of the DNA-constructs. Horizontal arrows show the repeated sequence of the mitochondrial genome at the termini, while vertical arrows show the boundary between the elements. The number refers to the nucleotide position from the mitochondrial sequence of *C. reinhardtii* (GenBank accession number NC_001638). (A) 5.0-kb DNA-construct, p5.0 kb. (B) 3.8-kb DNA-construct, p3.8 kb. (C) 1.8-kb DNA-construct, p1.8 kb. (D) PCR product obtained using a primer set, Cr.100F and Cr.3100R. Open-circles indicate the expected replication origin (see text). Dotted line shows pT7Blue-2 vector region.

Table 1. Transformation of *Chlamydomonas reinhardtii dum-1* using various sizes of gold particles.

Transformation of the chloroplast by pEX-50-AADA					
Gold particle size	600 + 100 nm	600 nm	200 nm	100 nm	40 nm
No. of colonies per plate	156, 103	79, 79	22, 22	5, 3	0, 0
Transformation of the mitochondria by p5.0 kb					
Gold particle size	600 + 100 nm	600 nm	200 nm	100 nm	40 nm
No. of colonies per plate	14, 11, 10, 9	9, 6, 5, 4	N/D[a]	0, 0, 0, 0	0, 0, 0, 0

[a] not determined.

T-vector. This plasmid is designated p3.8 kb. This construct has potential to compensate the deficiencies of *dum-1, -14, -15*, and *-16*, but not *dum-19*.

1.8-kb construct

This construct contains bases 1–1780 of the mitochondrial genome (i.e., the first element of the 5.0-kb construct itself; Figure 2C). This construct was ligated into the *Eco*RV site of the pT7Blue-2 T-vector, and is referred to as p1.8 kb. Basic potency of this construct is same to the p3.8 kb.

Cr.100F/Cr.3100R-RCR product

This PCR product (3.0 kb long) was obtained using a primer set (Cr.100F and Cr.3100R) and the *cw-15* mitochondrial genome as a template. This contains bases 100–3100 of the mitochondrial genome (Figure 2D), i.e., most part of the left terminal inverted repeat region (left arm; bases 1–545), whole COB and ND4 genes.

Results

Requirements for successful transformation of C. reinhardtii *mitochondria*

We attempted to transform five respiratory deficient strains of *C. reinhardtii* (*dum-1*, *-14*, *-15*, *-16*, and *-19*) (Figure 1) with DNA-constructs that contain compensatory regions (Figure 2). The strains *dum-1*, *-14*, and *-16* contain large deletions in the left terminus of their mitochondrial genome and showed no significant difference in the frequency of mitochondrial transformation (only the data for *dum-1* is shown in Table 1). Colonies active in respiration began to appear after 2 weeks of incubation in the dark with the latest appearing 4 weeks after bombardment (an example of respiration active transformants is shown in Figure 3). In contrast, no transformants were obtained for *dum-15* even after 20 attempts (4 plates per attempt), and only one actively respirating colony was yielded for *dum-19* through 22 attempts. However, DNA sequencing of the COXI gene revealed that the colony resulted from non-canonical back mutation accompanied by complex base changes, GGT/TG/AGC to GGT/GCA/AGC, which rescued the frame shift mutation in the *cox1* of *dum-19*. Consequently, no transformants due to homologous recombination were obtained from the *dum-15* and *dum-19* strains.

On the other hand, chloroplast transformation for all of these five strains was repetitively successful using the plasmid pEX-50-AAD (Takahashi *et al.*, 1996), which contains the bacterial aminoglycoside adenine transferase (*aadA*) gene conferring spectinomycin resistance (Goldschmidt-Clermont, 1991; Takahashi *et al.*, 1996). Moreover, the frequency of the chloroplast transformation was

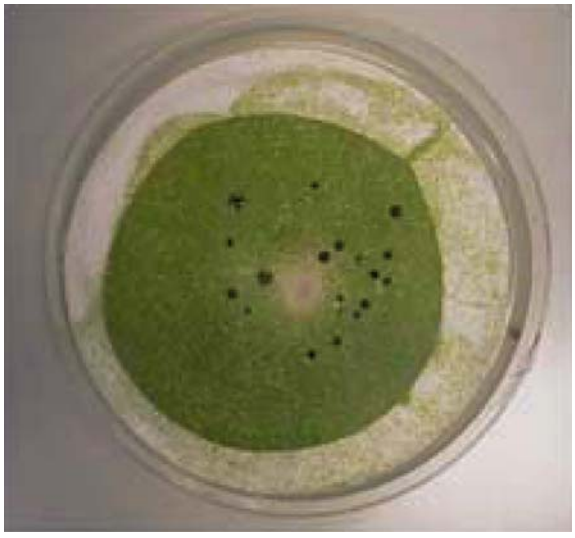


Figure 3. Appeared respiratory active transformants after 4 weeks of biolistic bombardment.

almost the same among the strains, i.e., 100–150 colonies per plate (a frequency of about 1×10^{-5} per cell) (data for *dum-1* is shown in Table 1).

Variable transformation efficiency depends on the gold particle size

Randolph-Anderson *et al.* (1995) suggest that 600 nm gold micro-projectiles give the highest transformation frequency for *C. reinhardtii* chloroplast. However, these particles are apparently too large to penetrate the mitochondria, which have a much smaller average diameter, estimated at 200–300 nm (Sager and Palade, 1957), therefore, we tried 40 and 100 nm gold particles for mitochondrial bombardment. Preliminary tests on chloroplast bombardment using 40, 100, or 200 nm gold particles resulted in a much lower frequency of transformation than the 600 nm particles, and no mitochondrial transformants were observed when bombarded with the 40 or 100 nm particles (Table 1). The highest frequency for chloroplast transformation was observed using a mixture of 100 and 600 nm particles, so we tested this mixture for mitochondrial transformation as well. The mixed projectiles increased the frequency about two-fold for mitochondrial and chloroplast transformations compared to the bombardment with the 600 nm particles alone (Table 1).

Variable transformation efficiency depends on the DNA-constructs

The closed-circular plasmid p5.0 kb (Figure 2A) showed the highest transformation efficiency (11 ± 5.1 transformants per 14–21 $\times 10^6$ cells, ca. 0.5–0.8 $\times 10^{-6}$ transformants per cell) (see Experimental procedures and Table 2), similar to that achieved by Randolph-Anderson *et al.* (1993) using the entire 15.8-kb mitochondrial genome (0.7×10^{-6} transformants per cell).

The transformation efficiencies using closed-circular p3.8 kb (Figure 2B) and p1.8 kb (Figure 2C) were about 50% and 20% of that for p5.0 kb, respectively (Table 2). Considering that mitochondrial genome in the *dum-1* mutant, with deletion ranging in size from ca. 1.7 to 1.5 kb (Randolph-Anderson *et al.*, 1993), then, the homologous region between plasmid p5.0 kb and the *dum-1* mitochondrial genome is 3.3–3.5 kb long, compared to 2.1–2.3 kb for p3.8 kb and 0.1–0.3 kb for p1.8 kb. The comparative ratio of the transformation efficiency between p5.0 kb and p3.8 kb was about 2:1 (Table 2), which is close to the differences in the length of the two construct's regions of homology (3.3–3.5 kb vs. 2.1–2.3 kb) with *dum-1*. Far reduced transformation efficiency of the p1.8 kb than the other two constructs is also in consistent with the difference of homologous regions (Table 2).

The closed-circular plasmid p5.0 kb, showed about 5 times higher transformation efficiency than the linear form prepared by digestion with *Not*I or *Bam*HI (choice of a restriction enzyme had no significant effect, Table 2). Even attempting transformation using the PCR product (3 kb long), corresponding to bases 100–3100 and containing the most part of left arm, the entire COB and ND4 genes (Figure 2D), resulted in transformants (Table 2), Sequencing of the PCR product transformed colonies indicated the homologous recombination most probably at the ND4 gene (also see below).

Table 2. Transformation of *Chlamydomonas reinhardtii dum-1* using various donor DNA-constructs and mixed gold particles of 100 and 600 nm.

DNA Construct	c.c-p5.0 kb[a]	c.c-p3.8 kb[b]	linear-p5.0 kb[c]	3kb-PCR product[d]	c.c-p1.8 kb
Experiment 1	18, 16, 16, 13	11, 9, 9, 7	6, 3, 1, 1	3, 2, 1, 1	1, 0, 1, 0
Experiment 2	11, 6, 2, 3	6, 3, 3, 2	3, 1, 1, 1	N/D[e]	N/D
Experiment 3	14, 11, 10, 9	13, 3, 3, 1	3, 2, 2, 1	N/D	N/D
Average (colonies per plate)	10.8 ± 5.1	5.8 ± 3.9	2.1 ± 1.5	N/D	N/D

[a] Closed circular form of 5.0 kb construct ligated into the pT7Blue-2 T-vector.
[b] Closed circular form of 3.8 kb construct ligated into the pT7Blue-2 T-vector.
[c] p5.0 kb linearized by *Bam*HI was used for experiment 1 and 2, while *Not*I was used for experiment 3.
[d] PCR product obtained using a primer set, Cr.100F and Cr.3100R.
[e] not determined.

Analysis of the mixed population of mitochondrial genomes in transformed cells

Preliminary characterization of transformants showed a mixed population of mitochondrial genomes. Therefore, we investigated the transition of heterogeneous populations of mitochondrial genomes associated with successive mitotic cell divisions as described below.

Analysis of the Phase I culture of dum-14/ *p5.0 kb transformants*

A mixed population of mitochondrial genomes was detected from some *dum-14*/p5.0 kb transformants of the Phase I culture (see Experimental procedures), and their representative gene organization was investigated. The mitochondrial gene organization was analyzed for *dum-14* transformants with restored respiration activity by the plasmid p5.0 kb. Total DNA was isolated from 14 *dum-14* transformants of the Phase I culture. PCR amplification was performed using the seven primers shown in Figure 1 [each primer name is composed of the number showing its 5′-base position and direction of extension, forward (F) or reverse (R)].

PCR amplification performed using the primer set, Cr.1325F/Cr.5224R, produced two kinds of PCR products 3.9 kb and 0.9 kb in length (Figure 4). DNA sequencing and restriction enzyme digestion of the 3.9 kb fragment showed that it contained a wild type segment, *cob-nd4*, which did not originally exist in the respiratory deficient *dum-1*, *-14*, *-16* strains (Figure 1). Further analysis of other PCR products by similar methods and by Southern hybridization (see below) suggested that gene organization of this mitochondrial genome segment is: left arm-*cob-nd4-nd5-cox1-nd2-nd6-nd1-rtl*-right arm, i.e., wild-type (Figure 5A). This is a result of homologous recombination between the first element of plasmid p5.0 kb and the *dum-14* mitochondrial genome. On the other hand, DNA sequencing of the 0.9 kb fragment reveled that this fragment contains the non-wild type (non-WT) segment (Figure 5B), *cob*-C(*nd4*)/(*nd5*)N-*cox1* ("/", "N", and "C" represent a direct fusion of truncated genes or elements, intact N-terminus of the C-terminus truncated gene, and intact C-terminus of the N-terminus truncated gene, respectively), which was derived from the plasmid p5.0 kb DNA.

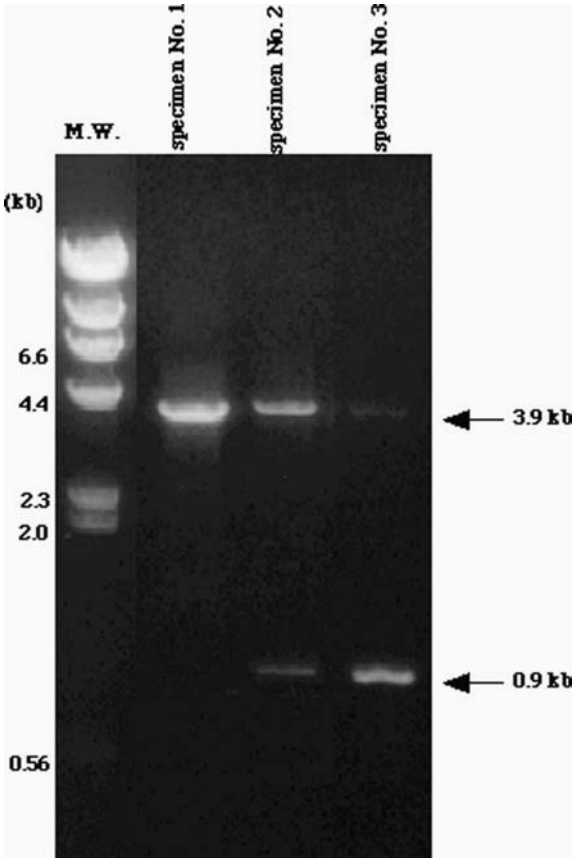


Figure 4. Mitochondrial genome analysis of transformants yielded by p5.0 kb. For PCR amplification, total DNA prepared from the Phase I culture of specimen Nos. 1, 2, and 3, and the primer set, Cr.1325F and Cr.5224R were used. A single band of 3.9-kb long was detected from specimen, No. 1, while 3.9-kb and 0.9-kb products were detected from Nos. 2 and 3. The 3.9-kb band was more intense than the 0.9-kb band for specimen No. 2, while the 0.9-kb band is more intense than the 3.9-kb band in the case of No. 3.

A 2.3 kb PCR product was amplified from one of the 14 examined *dum-14* transformants using the primer set: Cr.1325F/Cr.14529R (Figure 1). DNA sequencing reveled that this PCR product has a non-WT gene organization, *cob*-C(*nd4*)/(*nd5*)N-N(*cox1*)/right arm (Figure 5C), which likely is a result of homologous recombination between the 3rd element of plasmid p5.0 kb (Figure 2A) and the *dum-14* mitochondrial genome. No PCR product is expected when the wild-type genome is used as a template, because the two PCR primers (Cr.1325F and Cr.14529R) are positioned beyond the amplification limit of ExTaq polymerase (expected fragment size:

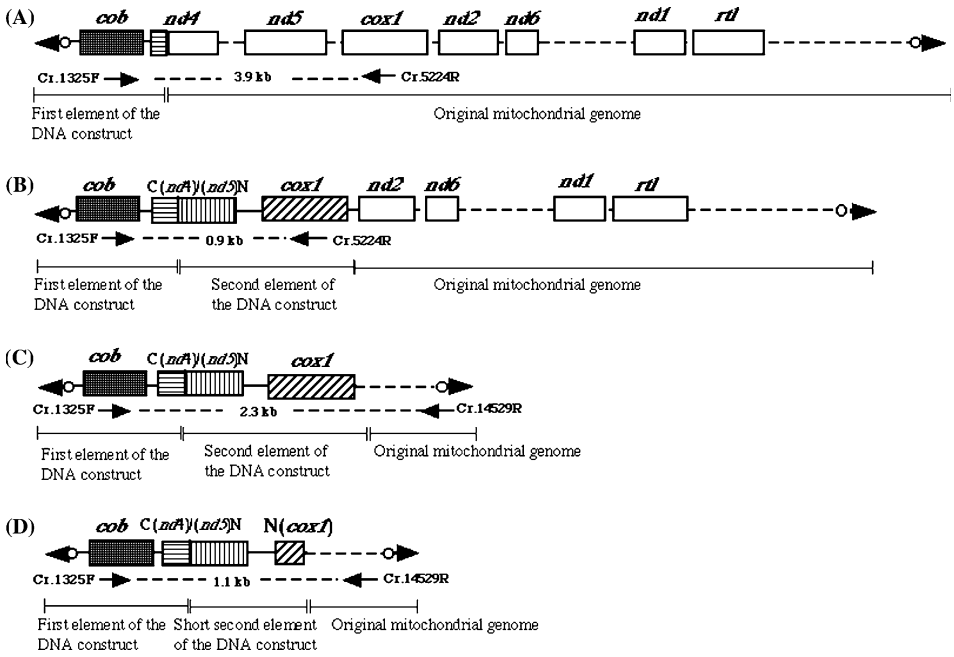


Figure 5. Structures of the recombinant mitochondrial genomes produced by biolistic bombardment. (A) Wild-type genome resulting from recombination with the first element of the DNA-construct p5.0 kb, p3.8 kb, or p1.8 kb. (B) Non wild-type genome resulting from recombination with the second element of the DNA-constructs p5.0 kb or the short second element of the p3.8 kb. (C) Non wild-type genome resulting from recombination with the third element of the p5.0 kb construct. (D) Non wild-type genome resulting from recombination with the third element of p3.8 kb. The relative position of each PCR primer used to analyze the gene organization is shown by an arrow. Genes or elements probably originated from *dum* mitochondrial genome are shown by dotted lies or plain boxes.

13 kb, ExTaq limit: ~8 kb; Takara Bio, Kyoto, Japan) (Figure 1).

Altogether, all of the examined *dum-14*/p5.0 kb transformants had their mutant mitochondrial genomes restored to wild type. In addition to the wild-type, four of the 14 transformants contained a non-WT genome that has the gene organization of: left arm-C(*cob*)N-C(*nd4*)/(*nd5*)N-*cox1-nd2-nd6-nd1-rtl*-right arm (Figure 5B), and one transformant has another non-WT gene organization, left arm-*cob*-C(*nd4*)/(*nd5*)N-N(*cox1*)/right arm (Figure 5C). Probably nine of the examined 14 transformants contained only the wild type, because no additional products besides the 3.9 kb fragment were detected by the PCR method, even after 40 cycles.

No PCR product was amplified using a primer that hybridizes to the ND4 gene and one that hybridizes to the U19 primer region of pT7Blue-2 vector (see Figure 2A). This shows that vector DNA is not maintained in the mitochondrial, and was also supported by the Southern hybridization result (see below).

(Analysis of the Phase I culture of *dum-1*/p3.8 kb transformants): Additionally, we detected three types of mitochondrial genomes in the mixed population of cells derived from transformants with p3.8 kb (Figure 2B). In all of the examined eight *dum-1* transformants, part of the wild-type specific gene organization, left arm-*cob-nd4-nd5-coxI* (Figure 5A), was confirmed by DNA sequencing of the Cr.1325F/Cr.5224R PCR product. In addition to this type, we detected a non-WT gene organization, left arm-*cob*-C(*nd4*)/(*nd5*)N-*cox1*, from all of the analyzed specimens (Figure 5C), and another non-WT organization, left arm-*cob*-C(*nd4*)/(*nd5*)N-N(*cox1*)/right arm (Figure 5D), from half of the specimens transformed with the plasmid p3.8 kb.

Analysis of the Phase II and III cultures

We performed similar PCR analysis using DNA prepared from the Phase II and III cultures (see Experimental procedures) of p5.0 kb transformants. Three of the five transformants that contained the non-WT segment, C(*nd4*)/(*nd5*)N, in the

Phase I culture DNA did not contain the corresponding PCR band in this phase. Furthermore, the PCR band corresponding to C(*nd4*)/(*nd5*)N-N(*cox1*)/right arm was also not detected from all cultures of p3.8 kb transformants. Using DNA prepared from the Phase III cultures, no non-WT mitochondrial genomes were detected in the culture by the PCR method using several sets of primers.

Analysis of mitochondrial genome transitions by Southern hybridization

We performed Southern hybridization to estimate the comparative quantities of wild-type vs. non-WT genomes detected by PCR amplification. As expected, no clear band was detected from untransformed *dum-1* and *-14* strains using a probe for the COB gene (Figure 6A and C). This is because most of the COB gene is absent in these *dum* mitochondrial genomes. On the other hand, two bands were detected using a probe for the ND4 gene irrespective of whether the genome was intact or digested by *Sca*I (Figure 6B and D). The lower band corresponds to the expected size of the left arm and most of the COB gene lacking *dum* mitochondrial genome, while the upper band corresponds to mitochondrial genome dimers that are fused head-to-head (Figure 6B, and D).

In contrast, a membrane blotted with DNAs prepared from the transformants (intact or *Sca*I digested) showed a single band in a position is similar to that for CC-124 wild-type DNA, while the band corresponding to the dimer mitochondrial genome was not detected (Figure 6A and C). We could not detect any clear specific band that corresponds to non-WT mitochondrial genomes even from the Phase I DNA using probes for the COB or ND4 genes (Figure 6). This is probably due to the lower sensitivity of the Southern hybridization method in relation to the PCR method.

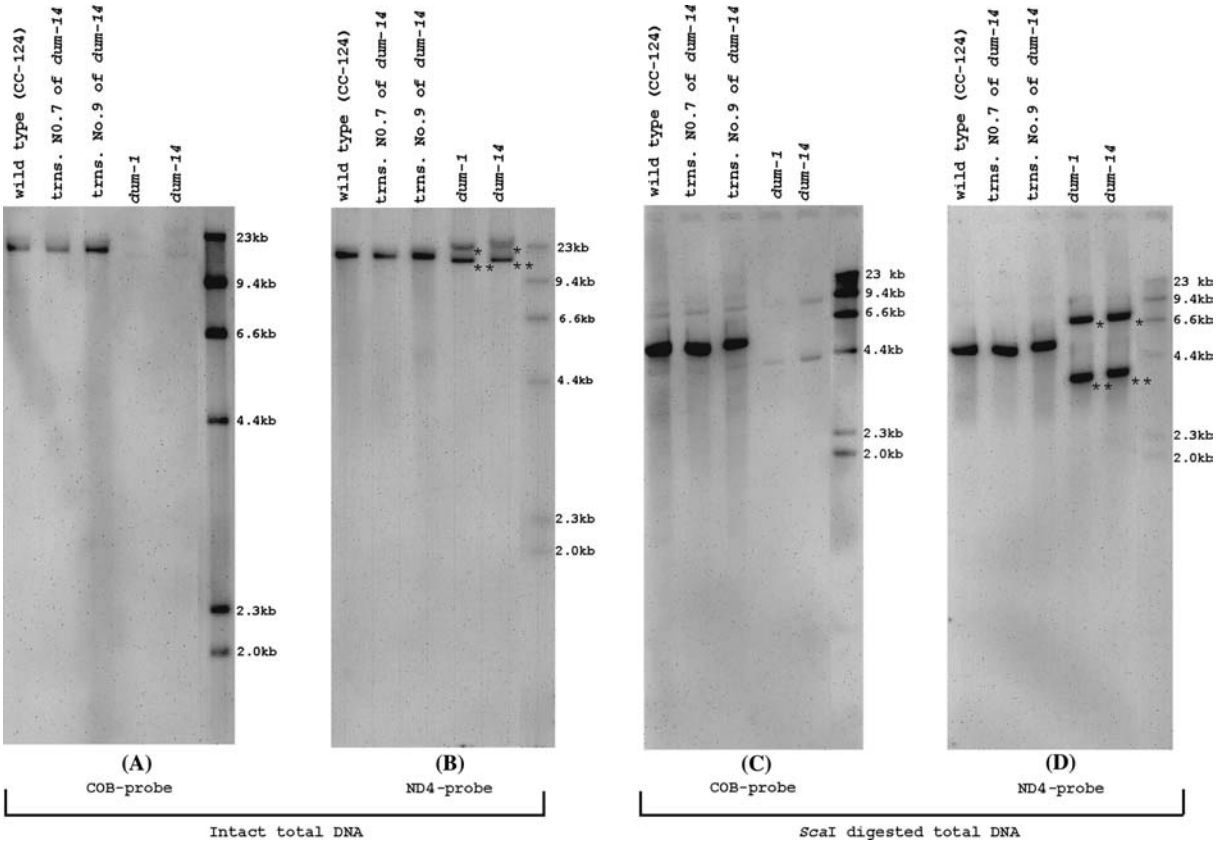


Figure 6. Mitochondrial genome analysis by Southern hybridization. *: band corresponding to head-to-head fused mitochondrial genome; **: band corresponding to the mitochondrial DNA that has a deletion of its left-arm and most of the COB gene. trans.: transformant.

No clear band was detected by hybridization using T7- or U19-probes for intact DNA blotted membranes (data not shown). This suggests that plasmid vector-region is not retained, even in the transformants of the Phase I culture.

Discussion

The ability to transform the *C. reinhardtii* mitochondrial genome using cloned DNA opens up the possibility of reverse genetic studies. Transformation of a DNA construct containing changes in the NADH dehydrogenase subunit genes (i.e., *nd1, nd2, nd4*, *nd5*, and *nd6*) would be extremely valuable to study the functions of the respiratory complex I (Cardol *et al*., 2002), because another species of which mitochondria are transformable, *Saccharomyces cerevisiae*, lacks this complex but instead has rotenone-insensitive NADH-quinone oxidoreductases (Seo *et al*., 1998). The COB gene of *C. reinhardtii* contains a target sequence of the intronic homing enzyme I-*Csm*I encoded in the ORF of a *C. smithii* group I intron named 'alpha' (Colleaux *et al*., 1990; Kurokawa *et al*., 2005). Mitochondrial transformation will also allow directed mutagenesis of the target sequence to investigate the recognition properties of the I-*Csm*I *in vivo*, because *C. reinhardtii* and *C. smithii* are interfertile and intron-invasion into the target site has been observed through mating. Moreover, the ability to transform both chloroplast (Boynton *et al*., 1988) and mitochondria provides a unique opportunity to explore the genetic interaction between these two organelles.

Effect of particle size on DNA delivery

Each series of experiments were performed using the same procedures, although we observed some variation in transformation efficiency between multiple plates even in a single experiment (Table 2), possibly because of poor distribution of the DNA-coated particles on each round plastic-disk (macrocarrier-disk). Irrespective of this, the overall transformation efficiency was consistently dependent on the key conditions described below.

Direct delivery of the donor DNA into the mitochondrial matrix appears essential for successful transformation, because it is likely that mitochondria cannot transport free cytosolic DNA into the matrix. We believe that this may also be true for delivering DNA into chloroplasts. In our experiments, the highest frequency of *C. reinhardtii* mitochondrial transformation was limited to about one-tenth of that for the chloroplast transformation (Table 1). This ratio may reflect the smaller volume of the mitochondria compared to the chloroplast.

Biolistic bombardment using the 40 or 100 nm gold particles was not successful for transforming the mitochondria, and using the 100 nm only slightly successful for the chloroplast, while the 600 nm particles were efficient for both organelles (Table 1). This might reflect the difference of physical energy among three kinds of particles, with only the 600 nm particles effective at 1100 psi, while the smaller particles may require a higher gas pressure to penetrate the *C. reinhardtii* cell wall. When the 100 and 600 nm gold particles were mixed, we achieved about two-fold higher transformation efficiency than using the 600 nm particles alone for the mitochondria as well as the chloroplast transformation (Table 1). This increase in efficiency may be cooperative, with the 600 nm particles perforating the cell wall, allowing some of the 100 nm particles to enter the cell and collide with the organelles, delivering the DNA into the organelle matrices. If the size of the mitochondrial cross section is constantly 200–300 nm in diameter, as previously reported by Randolph-Anderson *et al*. (1993), then even the 100 nm gold particles may not be small enough to enter the matrix without significant damage to the mitochondria. A recent report demonstrated that the mitochondria undergo a dynamic morphological change during the cell cycle, resulting in a fluctuating diameter of approximately 200 nm to 2.0 μm (Ehara *et al*., 1995). In our experiments the host cells were not synchronized, as they were cultured under continuous exposure to light preceding the bombardment, and it is possible that only the larger sizes of mitochondria survived the collision with the 100 nm gold particles.

It is known that the original 600 nm gold particles are contaminated up to 15% with 100–200 nm size particles (Randolph-Anderson *et al*., 1995), potentially reducing the differences in the observed transformation efficiencies between the "600 nm" particles and the 600 nm/100 nm particle mixture. Additionally, the closed-circular

DNA-construct may be more resistant to cellular degradation, in comparison to the linear form, and may explain the reproducibly higher transformation efficiency.

Characteristics of the successfully transformed strains

The five respiratory deficient mutants used in this study originally had an identical nuclear genetic background, because they were all developed from the same wild-type parent strain, *C. reinhardtii* 137C (Matagne *et al*., 1989). Respiratory mutants were obtained after growing a culture in medium containing acriflavine or ethidium bromide as a mutagen, which enhances mutations in organelle genomes predominantly over nuclear genomes. Moreover, it has been shown that respiratory deficiency in the mutants used is due to alternations in the mitochondrial genome (Matagne *et al*., 1989; Dorthu *et al*., 1992; Colin *et al*., 1995). All the transformable strains (*dum-1*, *-14*, and *-16*) have deletions in their mitochondrial genomes, ranging from the left terminus to somewhere within the neighboring COB gene (Dorthu *et al*., 1992). On the other hand, the strains that could not be transformed do not share this deletion in their genome: *dum-15* carries two base substitutions in the COB gene, which results in an amino acid change, serine (UCU) to tyrosine (UAC) (Colin *et al*., 1995), and *dum-19* has a one base deletion in the COXI gene, which introduces a UGA stop codon into the middle of the gene (Colin *et al*., 1995). This suggests that determinant of the successful mitochondrial transformation is highly likely the large deletion at the left terminus.

The mitochondrial genomes carrying deletions in their left termini are known to be quite unstable, possibly because of the absence of a terminal inverted repeat region (Randolph-Anderson *et al*., 1993). Such a wounded genome might induce the expression of required enzymes for homologous recombination in the mitochondria. This might explain why only the terminus-deleted strains are successful in transformation by homologous recombination between the delivered DNA and the mitochondrial genome. However in the case of the chloroplast, proteins that are required for homologous recombination seem to be abundant even when the entire genome is intact. This is supported by our data that shows wild-type *C. reinhardtii* and also these five respiratory deficient strains were easily transformed with pEX-50-AAD, and the construct was integrated into the chloroplast genome by homologous recombination (data not shown).

Conversion of a left arm deleted genome into a wild-type mitochondrial genome

How the linear *Chlamydomonas* mitochondrial genome replicates is not known, however two models have been proposed, a recombination-mediated model and a reverse-transcriptase model (Vahrenholz *et al*., 1993). A replication origin-like (OL) structure has been suggested in the terminal inverted sequences based on similarities to the human mitochondrial OL sequence (Nedelcu and Lee, 1998). Unfortunately, we do not have solid experimental data that proves whether the circular 5.0-kb or 3.8-kb DNA-constructs, which has OL-like sequences in their right and left termini, are able to replicate in the mitochondrial matrix or not. However, we could not detect the vector region of the p5.0 kb molecules within the *dum-14* transformants by Southern hybridization (Figure 6) or by PCR, therefore homologous recombination might occur immediately after introduction of the DNA-construct.

To restore the wild-type genome, homologous recombination between the mitochondrial genome and somewhere within the first element of the structure in Figure 5, followed by successful elimination of the *E. coli* vector region appears to be the simplest mechanism. The appearance of non-WT genomes detected by PCR must be the result of homologous recombination occurring in the other regions of the DNA-constructs. For example, the genome structure shown in Figure 5B is probably the result of homologous recombination with the second element of p5.0 kb or the shorter second element of p3.8 kb, while that in Figure 5C is the result of recombination with the third element of p5.0 kb. The structure of Figure 5D is probably the result of recombination between the *dum* genome and the third element of p3.8 kb. It is possible that even these non-WT genomes can compensate the respiratory deficiency of the *dum* mutants, because they have a promoter to transcribe the COB gene and a ribosomal binding signal for its translation. Probably even these non-WT mitochondrial genomes can be converted to

the wild-type genome through a secondary homologous recombination with the newly replicated wild-type genomes.

Most of the *dum* mutants obtained to date have deficient activities in respiratory complex III or IV, and very rarely in the complex I, because almost all of the *dum* mutants result from deletion or point mutation of the COB gene (subunit of complex III) or COXI gene (subunit of complex IV), respectively (Remacle and Matagne, 1998). In addition, a mixed population of mitochondrial genomes in the original *dum*-1 mutant eventually stabilized, with fragments never exceeding 1.7 kb in deletion size, which is the maximum permissible size to avoid the loss of ND4 (one of the subunits of complex I) (Randolph-Anderson *et al*., 1993). These results indicate that complex I activity is significant, if not essential, for survival of *C. reinhardtii*. On the other hand, homologous recombination within the second or third elements of the donor DNA-constructs leads to the loss of active ND4 and ND5 genes in the mitochondrial genome and acquisition of the defective C(*nd4*)/(*nd5*)N-gene from the donor construct. Therefore, the negative effect on formation of active complex I seems serious in such a recombinant. In contrast, mitochondrial DNA that recombined within the first element of the DNA-construct retains active ND4 and 5 genes and also restores the COB gene. Therefore, this recombinant DNA molecule, i.e., a wild-type genome, will be selected strongly over the non-WT genes. This is the most probable reason for the high frequency of the wild-type mitochondrial genome observed in the Phase I culture, despite the very short homologous region (100–200 bp) for recombination with *dum* mitochondrial genomes. We observed rapid disappearance of head-to-head ligated dimer and other non-WT mitochondrial genomes from the culture. This may be due to the faster division of the cells that have a larger number of the wild-type mitochondrial genome.

Acknowledgements

We sincerely thank Drs. C. Remacle and R. F. Matagne, (University of Liege, Belgium) for providing the respiratory deficient strains used in this study, and also to Dr. Y. Takahashi (Okayama University, Japan) for the plasmid used for chloroplast transformation. B. Eng. Y. Hayashi (KUT) provided technical assistance to isolate genomic DNAs. We also would like to thank Dr. Boynton J. and Ms. Mariya Takeuchi for encouraging us to perform this study. This work was supported by the Science Research Promotion Fund provided by the Promotion and Mutual Aid Corporation for Private School of Japan and Regional Science Promotion Program.

References

Boynton, J.E., Gillham, N.W., Harris, E.H., Hosler, J.P., Johnson, A.M., Jones, A.R., Randolph-Anderson, B.L., Robertson, D., Klein, T.M., Shark, K.B. and Sanford, J.C. 1988. Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. Science 240: 1534–1538.

Cardol, P., Matagne, R.F. and Remacle, C. 2002. Impact of mutations affecting ND mitochondria-encoded subunits on the activity and assembly of complex I in *Chlamydomonas*. Implication for the structural organization of the enzyme. J. Mol. Biol. 319: 1211–1221.

Colin, M., Dorthu, M-P., Duby, F., Remacle, C., Dinant, M-R., Duyckaerts, C. and Sluse Matagne, F. R.F. 1995. Mutations affecting the mitochondrial genes encoding the cytochrome oxidase subunit I and apocytochrome b of *Chlamydomonas reinhardtii*. Mol. Gen. Genet. 249: 179–184.

Colleaux, L., Michel-Wolwertz, M.R., Matagne, R.F. and Dujon, B. 1990. The apocytochrome *b* gene of *Chlamydomonas smithii* contains a mobile intron related to both *Saccharomyces* and *Neurospora* introns. Mol. Gen. Genet. 223: 288–296.

Dorthu, M-P., Remy, S., Michel-Wolwertz, M-R., Colleaux, L., Breyer, D., Beckers, M-C., Englebert, S., Duyckaerts, C., Sluse, F-E. and Matagne, R.F. 1992. Biochemical, genetic and molecular characterization of new respiratory- deficient mutants in *Chlamydomonas reinhardtii*. Plant Mol. Biol. 18: 759–772.

Ehara, T., Osafune, T. and Hase, E. 1995. Behavior of mitochondria in synchronized cells of *Chlamydomonas reinhardtii* (Chlorophyta). J. Cell Sci. 108: 499–507.

Goldschmidt-Clermont, M. 1991. Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker of site-directed transformation of *Chlamydomonas*. Nucl. Acids Res. 19: 4083–4089.

Gorman, D.S. and Levine, R.P. 1965. Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA 54: 1665–1669.

Gray, M.W. and Boer, P.H. 1988. Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 319: 135–147.

Harris, E. 2001. *Chlamydomonas* as a model organism. Plant Mol. Biol. 52: 363–406.

Johnston, S.A., Anziano, P.Q., Shark, K., Sanford, J.C. and Butow, R.A. 1988. Mitochondrial transformation in yeast by bombardment with microprojectiles. Science 240: 1538–1541.

Kurokawa, S., Bessho, Y., Higashijima, K., Shirouzu, M., Yokoyama, S., Watanabe, I.K. and Ohama, T. 2005.

Adaptation of intronic homing endonuclease for successful horizontal transmission. FEBS J. 272: 2487–2496.

Matagne, R.F., Michel-Wolwertz, M-R., Munaut, C., Duyckaerts, C. and Sluse, F. 1989. Induction and characterization of mitochondrial DNA mutants in *Chlamydomonas reinhardtii*. J. Cell Biol. 108: 1221–1226.

Murray, M.G. and Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. Nucl. Acids Res. 8: 4321–43215.

Nedelcu, A. and Lee, R.W. 1998. Short repetitive sequences in green algal mitochondrial genomes: potential roles in mitochondrial genome evolution. Mol. Biol. Evol. 15: 690–701.

Randolph-Anderson, B.L., Boynton, J.E., Dawson, J., Dunder, E., Eskes, R., Gillham, N.W., Johnson, A., Perlman, P.S., Suttie, J. and Heiser W.C. 1995. Sub-micron gold particles are superior to larger particles for efficient biolistic transformation of organelles and some cell types. US/EG Bulletin 2015 Bio-Rad.

Randolph-Anderson, B.L., Boynton, J.E., Gillham, N.W., Harris, E.H., Johnson, A.M., Dorthu, M.P. and Matagne, R.F. 1993. Further characterization of the respiratory deficient *dum-1* mutation of *Chlamydomonas reinhardtii* and its use as a recipient for mitochondrial transformation. Mol. Gen. Genet. 236: 235–244.

Remacle, C. and Matagne, R.F. 1998. Mitochondrial genetics. In: J.-D. Rochaix, M. Goldschmidt-Clermont and S. Merchant (Eds.), The Molecular Biology of Chloroplasts and Mitochondria in *Chlamydomonas*, Kluwer Academic Publishers, Dordrecht, pp. 661–674.

Sager, R. and Palade, G.E. 1957. Structure and development of the chloroplast in *Chlamydomonas*. I. The normal green cell. J. Biophys. Biochem. Cytol. 3: 463–487.

Seo, B.B., Kitajima-Ihara, T., Chan, E.K., Scheffler, I.E., Matsuno-Yagi, A. and Yagi, T. 1998. Molecular remedy of complex I defects: rotenone-insensitive internal NADH-quinone oxidoreductase of *Saccharomyces cerevisiae* mitochondria restores the NADH oxidase activity of complex I-deficient mammalian cells. Proc. Natl. Acad. Sci .USA 95: 9167–9171.

Smith, F.D., Harpending, P.R. and Sanford, J.C. 1992. Biolistic transformation of prokaryotes: factors that affect biolistic transformation of very small cells. J. Gen. Microbiol. 138: 239–248.

Takahashi, Y., Utsumi, K., Yamamoto, Y., Hatano, A. and Satoh, K. 1996. Genetic engineering of the processing site of D1 precursor protein of photosystem II reaction center in *Chlamydomonas reinhardtii*. Plant Cell Physiol. 37: 161–168.

Vahrenholz, C., Riemen, G., Pratje, E., Dujon, B. and Michaelis, G 1993. Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. Curr. Genet. 24: 241–247.