This is an electronic version of an article published in Bioscience, Biotechnology, and Biochemistry [Volume 77, No. 4, pp. 782-788, 2013]. Published version is available online at: https://www.jstage.jst.go.jp/article/bbb/77/4/77_120919/_article. Bioscience, Biotechnology, and Biochemistry is available online at: www.tandfonline.com/Article DOI; 10.1271/bbb.120919

The *Apl*I Restriction-Modification System in an Edible Cyanobacterium, *Arthrospira* (*Spirulina*) *platensis* NIES-39, Recognizes the Nucleotide Sequence 5'-CTGCAG-3'

Hideaki SHIRAISHI[†] and Yosuke TABUSE[1]
Graduate School of Biostudies, Kyoto University, Yoshida-Konoecho, Sakyo-ku, Kyoto 606-8501, Japan

Received November 30, 2012; Accepted January 18, 2013

[†] To whom correspondence should be addressed.
Tel: +81-75-753-3998;
E-mail: siraisi@kuchem.kyoto-u.ac.jp

[1] Present address: ORBIS Inc., 2-1-4 Hiratsuka, Shinagawa-ku, Tokyo 142-0051, Japan

Abbreviation: R/M, restriction and modification

Abstract

Degradation of foreign DNAs by restriction enzymes in an edible cyanobacterium *Arthrospira platensis* is a potential barrier for gene-transfer experiments in this economically valuable organism. We overproduced in *Escherichia coli* the proteins involved in a putative restriction-modification system of *A. platensis* NIES-39. The protein produced from the putative type II restriction enzyme gene NIES39_K04640 exhibited an endonuclease activity that cleaved DNA within the sequence 5'-CTGCAG-3' between the A at the fifth position and the G at the sixth position. We designated this enzyme *Apl*I. The protein from the adjacent gene NIES39_K04650, which encodes a putative DNA (cytosine-5-)-methyltransferase, rendered DNA molecules resistant to *Apl*I by modifying the C at the fourth position (but not the C at the first position) in the recognition sequence. This modification enzyme, M.*Apl*I, should be useful for converting DNA molecules into *Apl*I-resistant forms for use in gene-transfer experiments. Summary of restriction enzymes in various *Arthrospira* strains is also presented in this paper.

Key words: *Arthrospira platensis*; *Spirulina*; restriction enzyme; DNA methyltransferase; edible cyanobacteria

Introduction

Arthrospira platensis, formerly known as *Spirulina platensis*, is an edible cyanobacterium that has traditionally been consumed as food by people of Chad and Niger along the shores of Lake Chad, which is characterized by high levels of carbonate/bicarbonate and high pH (for reviews of this organism, see refs. 1 and 2). In the 1970s, modern commercial cultivation of this filamentous cyanobacterium began, because it was revealed to contain high levels of proteins, vitamins and minerals that make it suitable for human consumption. Thereafter, it has been cultivated and used worldwide as a source of food and food additives as well as animal and fish feed. As an edible photosynthetic microorganism that can be propagated outdoors as a monoalgal culture,[3] this cyanobacterium has the potential to be used in an economically advantageous manner for the production of vast amounts of nutrients and/or useful compounds. These efforts would be aided by genetic engineering of this organism.

One obstacle that limits the use of this cyanobacterium in genetic engineering is that all strains of this organism analyzed thus far contain multiple restriction enzymes that degrade foreign DNA molecules.[4-6] In accord with these observations, recent determinations of the nucleotide sequences of the whole genomes of *A. platensis* NIES-39, *A. platensis* C1 (PCC9438) and *Arthrospira* sp. PCC 8005 revealed that these strains also contain multiple genes encoding putative restriction endonucleases.[7-9] Therefore, in order to apply bioengineering techniques to this species, it will be important to characterize its restriction and modification (R/M) systems. The resulting knowledge will help provide means for evading degradation of DNA molecules introduced into cells during gene-transfer experiments; specifically, it will allow design of DNA molecules devoid of susceptible sequences, as well as modification of DNA by cognate methylation enzymes to prevent recognition and cleavage at restriction sites.

Biochemical investigation of restriction enzymes in the 1980s and 1990s showed that *Arthrospira* strains contain different sets of restriction enzymes depending on isolates.[4-6] Restriction enzymes found in *Arthrospira* strains in those period were isoschizomers of nine restriction enzymes: *Bsi*WI, *Tth*111I, *Hae*III, *Pvu*I, *Pvu*II, *Hin*dIII, *Sna*BI, *Hgi*CI and *Hgi*DI (Table 1). The activity(s) of one to four members of these nine isoschizomers were detected in each strain. Results of recent whole genome sequencing of *A. platensis* NIES-39, *A. platensis* C1 and *Arthrospira* sp. PCC 8005 revealed that the genomes of these strains contain many more putative restriction enzyme genes than expected from previous biochemical studies.[7-9] As shown in Table 1, the genome of *A. platensis* NIES-39 contains genes encoding putative proteins having significant homology with *Bsi*WI, *Sna*BI, *Hgi*CI and *Hgi*DI, which correspond

to four of the nine isoschizomers that had been detected in *Arthrospira* strains by biochemical studies. In the case of *A. plantensis* C1 and *Arthrospira* sp. PCC 8005, they contain putative genes for seven and six of these nine isoschizomers, respectively. In addition to these genes, genes for putative restriction enzymes that had not been found in previous biochemical studies were also found in the genome sequences. For example, in the case of *A. platensis* NIES-39, six additional genes for putative type II restriction enzymes were detected in the genome (Table 1). Thus, for developing bioengineering systems utilizing these strains, it is important to characterize these newly identified putative R/M systems as well as those identified by previous studies.

In this paper, we report the characterization of one of the putative R/M systems in *A. platensis* NIES-39 consisting of a protein encoded by the NIES39_K04640 gene, which has homology with type II restriction endoncleases of *Bsu*BI/*Pst*I family, and a protein encoded by an adjacent gene NIES39_K04650, which has homology to DNA (cytosine-5-)-methyltransferases. Although the restriction enzyme of this putative R/M system share similarity with restriction enzymes in *Bsu*BI/*Pst*I family, the cognate modification enzyme is strikingly different from the enzymes of well-characterized R/M systems in this family; namely, whereas *Bsu*BI and *Pst*I are accompanied by N-6 adenine-specific methyltransferases,[10] this putative restriction enzyme is accompanied by a C-5 cytosine-specific methyltranferase. In addition to the difference of the cognate modification enzyme, the amino acid sequence of this putative restriction enzyme is less conserved than the sequences of *Bsu*I and *Pst*I. Therefore, this R/M system as a whole appears to be evolutionarily distant from well-characterized R/M systems in the *Bsu*BI/*Pst*I family. It is thus of interest to examine whether this R/M system retains the same sequence specificity as other members of the *Bsu*BI/*Pst*I restriction enzyme family. We show here that the restriction enzyme of this R/M system is isoschizomeric to *Bsu*BI and *Pst*I, and the cognate modification enzyme recognizes a common target sequence with M.*Bsu*BI and M.*Pst*I, despite it belongs to a different class of DNA methyltransferase.

Results and Discussion

Expression in E. coli of the proteins for a putative R/M system from A. platensis NIES-39

In order to characterize the biochemical properties of the proteins encoded by the NIES39_K04640 and NIES39_K04650 genes (Fig. 1A), they were produced as recombinant proteins with a hexahistidine tag in *Escherichia coli*. The proteins were then purified using a nickel-chelating column that binds the tag. The molecular masses of the recovered proteins were

40 kDa and 41 kDa, respectively, as estimated from the mobility on an SDS-polyacrylamide gel (Fig. 1). These values were consistent with the expected molecular weights calculated from the amino acid sequences of the recombinant proteins (39.3 kDa and 40.1 kD for the recombinant proteins for NIES39_K04640 and NIES39_K04650, respectively).

The recombinant protein form the NIES39_K04640 gene has an endonuclease activity

The recombinant protein from the NIES39_K04640 gene (K04640 protein) was assayed for endonuclease activity using the DNA of λ phage as a substrate. Since optimal reaction conditions for this protein was unknown, enzymatic activity was preliminarily examined using several different buffers that had been used for the reactions of other restriction enzymes. Endonuclease activity of the K04640 protein was detected in many different buffers, but the most efficient cleavage was observed in a buffer containing 20 mM Tris-HCl (pH 8.5), 10 mM $MgCl_2$, 1 mM dithiothreitol, and 100 mM KCl ("buffer K" supplied from Takara Bio Co.). This buffer had a relatively basic pH compared with others whose pH were 7.5-7.9. To determine whether the weakly alkaline pH of the buffer promoted the efficient cleavage, we examined the activity of the K04640 protein under various pH conditions. Although the activity of the K04640 protein was detectable at pH 7.5, it was more active under basic conditions with optimal pH at around 9.5 (Figs. 2A and 2B). It is of interest that the optimal pH roughly corresponds to the conditions of the natural habitat of *A. platensis*. It is not conceivable, however, that the optimal pH conditions of this enzyme are directly related to the normal conditions in healthy cells, since such alkaline conditions would deteriorate many other enzymatic reactions in the cell. One possible speculation is that the high pH optimum of this enzyme is a result of some uncommon event in which selection pressure was applied to unhealthy cells whose intracellular pH conditions were temporarily affected by environmental conditions.

In Fig.2A, it is also shown that complete digestion of λ phage DNA by the K04640 protein resulted in the formation of DNA fragments with discrete lengths. This result indicates that the K04640 protein cleaved DNA at specific positions.

We also examined the effect of monovalent cations (Fig. 2C). The endonuclease activity was enhanced in the presence of moderate concentration (around 50 mM) of K^+ or Na^+. In contrast, NH_4^+ appeared to have negative effect under these conditions. In the presence of high concentration of NH_4Cl, the reaction buffer was slightly acidified (pH 8.04 at 200 mM NH_4Cl), but the reduction of the enzyme activity was greater than that expected from the change of pH.

K04640 protein, or AplI, cleaves DNA at 5'-CTGCAG-3' sequences

Examination of the DNA fragments generated by the complete digestion of λ phage DNA

(Fig. 2A) suggested that their lengths were the same as those generated by *Pst*I digestion. It was further confirmed by performing electrophoresis of *Pst*I and K04640 protein digests of λ phage DNA on adjacent lanes on an agarose gel. As shown in Fig. 3A, *Pst*I and the K04640 protein produced the same set of digestion products, indicating that the K04640 protein had the same sequence specificity as *Pst*I whose recognition sequence was 5'-CTGCAG-3'.[11] In order to determine the exact cleavage site in the recognition sequence, a ^{32}P-labeled double-stranded DNA having the recognition sequence was synthesized and cleaved with the K04640 protein, and the cleavage site was determined by electrophoresis on a sequencing gel. As shown in Fig. 3B, the K04640 protein cut the DNA at the phosphodiester bond between A and G in the recognition sequence 5'-CTGCAG-3' to cleave it into 5'-CTGCA and G-3'. Thus, the cleavage site was also the same as *Pst*I.[11] We designated this *Pst*I isoschizomer of *A. platensis* NIES39 *Apl*I, according to the nomenclature for restriction enzymes.[12]

The *Apl*I belongs to the *Bsu*BI/*Pst*I restriction endonuclease family (Pfam accession number: PF06616),[13] sharing 37 and 30% identical amino acids with *Bsu*BI and *Pst*I, respectively. The number of the identical amino acids is, however, less than that between *Bsu*BI and *Pst*I that share 46% of the amino acids.[10] Nonetheless, our study showed that *Apl*I still retained the same recognition sequence and cleavage site as *Bsu*BI and *Pst*I.

The recombinant protein from the NIES39_K04650 gene has a modification activity that renders DNA resistant to AplI

The protein encoded by the NIES39_K04650 gene that resides adjacent to the *Apl*I gene (NIES39_K04640) shares high sequence similarity with DNA (cytosine-5-)-methyltransferases including the signature sequence of the active site of C-5 cytosine-specific DNA methylases (PROSITE accession number: PS00094) at amino acid positions 85-97.[14] To examine if this protein renders DNA molecules resistant to *Apl*I, modification activity of the recombinant protein (K04650 protein) from the NIES39_K04650 gene was assayed using λ phage DNA as a substrate. As shown in Fig. 4A, the DNA pretreated with the K04650 protein in the presence of a methyl donor, *S*-adenosyl methionine, became resistant to both *Apl*I and its isoschizomer, *Pst*I. The modified DNA was susceptible to *Hin*dIII and *Eco*RI, indicating that the modification was specific to the restriction sites for *Apl*I/*Pst*I.

Determination of the modification site by the K04650 protein, or M.AplI

The recognition sequence of *Apl*I, 5'-CTGCAG-3', contains two cytosine residues at the first and fourth positions in the sequence. Since the K04650 protein belongs to the family of DNA (cytosine-5-)-methyltransferases, it is most likely that the C5 position of either one or both of these cytosines is the target of this protein. In order to examine the modification sites by the

K04650 protein, a DNA molecule that contained a set of restriction sites shown in Fig. 4B was prepared. This molecule contained a *Pvu*II site that partly overlaps an *Apl*I site. The cognate modification enzyme of *Pvu*II is a (cytosine-N4-)-methyl transferase that modifies the N4 position of C at the fourth position in *Pvu*II sites,[15,16] but the restriction enzyme *Pvu*II itself is sensitive to 5-methyl modification at the position as well as the N4-methyl modification.[17] Thus, if the K04650 protein modifies the C5 position of the first C residue in the *Apl*I site, the modified DNA becomes resistant to *Pvu*II. If the modified DNA is susceptible to *Pvu*II, the DNA does not contain modification there. Another restriction enzyme that we employed, *Hpy*CH4V, recognizes the central four-nucleotide sequence in the *Apl*I site. Therefore, if the modified DNA is resistant to this enzyme, it positively indicates that the DNA has modification(s) in the region of the central four-nucleotide sequence. This DNA molecule also contains a restriction site for *Hin*dIII that resides outside of the *Apl*I site and works as a control. As shown in the bottom panel of Fig.4B, the DNA molecule modified by the K04650 protein was resistant to *Apl*I and *Hpy*4V, but susceptible to digestion by *Pvu*II as well as *Hin*dIII. Thus the K04560 protein modified the nucleotides in the central region of the *Apl*I site, but not the C residue at the first position in the *Apl*I site. These results indicated that the K04650 protein, or M.*Apl*I, modified only the C residue at the fourth position in the recognition sequence of *Apl*I.

In contrast to the similarity of the restriction enzyme *Apl*I with *Bsu*BI/*Pst*I, the cognate modification enzyme M.*Apl*I shares similarity with neither M.*Bsu*BI nor M.*Pst*I, because M.*Apl*I is an enzyme of DNA (cytosine-5-)-methyltransferase family whereas M.*Bsu*BI and M.*Pst*I are DNA (adenine-N6-)-methyltransferases as mentioned in the Introduction.[10] This striking difference indicates that M.*Apl*I evolved independently from the M.*Bsu*BI and M.*Pst*I to recognize the same recognition sequence as these modification enzymes.

We showed in this study that the proteins produced from the NIES39_K04640 and NIES39_K04650 genes of *A. platensis* NIES-39 have restriction and modification activities, respectively, that recognize the nucleotide sequence 5'-CTGCAG-3'. In the case of an other filamentous cyanobacterium *Anabaena* sp. PCC 7120, characterization of its R/M systems helped establish an efficient gene-transfer procedure; it has three different type II restriction enzymes that are isoschizomeric to *Ava*I, *Ava*II and *Ava*III, and methylation of DNA by the all three cognate modification enzymes resulted in drastic improvement of gene-transfer efficiency.[18] Although further study on other R/M systems is apparently needed for *A. platensis* NIES-39, the information on the recognition sequence of *Apl*I and the production of the recombinant M.*Apl*I in this study would be helpful for establishing an efficient gene-transfer procedure for this strain in the future, as is the case with *Anabaena* sp. PCC 7120.

Experimental

Bacterial strains. *A. platensis* NIES-39 was obtained from the Microbial Culture Collection at the National Institute for Environmental Studies (NIES Collection; Tsukuba, Japan). *E. coli* KRX harboring a T7 RNA polymerase gene that is controlled by the *rha*BAD promoter was purchased from Promega.

Cell Culture. *A. platensis* NIES-39 was cultured at 30˚C under a photoperiod of 12h light-12h dark in a medium slightly modified from SOT medium.[19,20] For preparing 1 liter of the medium, 16.8 g of $NaHCO_3$, 0.5 g of K_2HPO_4, 2.5 g of $NaNO_3$ and 1 g of K_2SO_4 were dissolved in 970 ml of deionized water and sterilized by autoclaving. After cooling, 20 ml of 50x macroelements and 1 ml of 1000x microelements that had been individually prepared and autoclaved were added; the 50x macroelements contained in 100 ml 5 g of NaCl, 0.4 g of Na_2-EDTA, 50 mg of $FeSO_4$-$7H_2O$, 1 g of $MgSO_4$-$7H_2O$, 0.2 g of $CaCl_2$-$2H_2O$ and 14.3 mg of H_3BO_3, and the 1000x microelements contained in 100 ml 218 mg of $MnSO_4$-$5H_2O$, 22.2 mg of $ZnSO_4$-$7H_2O$, 7.9 mg of $CuSO_4$-$5H_2O$ and 2.1 mg of Na_2MoO_4-$2H_2O$.

Preparation of genomic DNA. Genomic DNA of *A. platensis* NIES-39 was prepared by combining the methods for isolating nucleic acids from cyanobacteria[21] and for enriching DNA from a mixture of DNA and RNA by differential precipitation with isopropanol.[22, 23] In brief, exponential-phase cells of *A. platensis* NIES-39 (wet weight, 1 g) were washed with 5 ml of a buffer containing 50 mM Tris-HCl (pH 8.0), 50 mM NaCl and 5 mM EDTA. The cells were suspended in 2 ml of a saturated solution of NaI and incubated for 20 minutes at 37˚C. After washing twice with distilled water, cells were suspended in 5 ml of 50 mM Tris-Cl (pH8.5), 50 mM NaCl, 5 mM EDTA, and then incubated with lysozyme (1 mg/ml) for 45 minutes at 37˚C. Sodium dodesyl sulfate and proteinase K were then added to make final concentrations of 1% (w/v) and 50 μg/ml, respectively. After incubating for 1 hour at 37˚C, the solution was repeatedly extracted with phenol until no interface was visible. To the one volume of the purified nucleic acid solution, 2/3 volume of 5M NaCl and one volume of isopropanol were added and mixed. Precipitated DNA was recovered by centrifugation and dissolved in TE (10 mM Tris-HCl, 0.2 mM EDTA, pH 8.0). The precipitation was repeated once more. In this second precipitation, a clump of DNA that appeared in the solution was recovered by entangling around a plastic tip of a micropipetter and transferred into a 70% ethanol solution. The DNA recovered from the 70% ethanol by centrifugation was dissolved in TE and stored at -25˚C.

Overproduction of recombinant proteins in E. coli. The coding regions for

NIES39_K04640 and NIES39_K04650 genes were individually amplified from the genomic DNA of *A. platensis* NIES-39 by PCR with iProof high-fidelity DNA polymerase (Bio-Rad) according to the manufacturer's instruction. The primers used in the PCR were designed to introduce *Nco*I and *Xho*I sites to the N-terminal side and C-terminal side, respectively, of the protein-coding sequences retrieved from the CyanoBase.[24] After digesting the amplified DNA fragments with *Nco*I and *Xho*I, they were inserted between the *Nco*I and *Xho*I sites of a T7 promoter-based expression vector, pET28b(+) (Novagen). Nucleotide sequences of the constructed plasmids were determined to confirm that the cloned DNA fragments had the same nucleotide sequences as published ones.[7] The plasmids, pET-K04640 and pET-K04650, thus constructed were individually introduced into *E. coli* KRX. The *E. coli* KRX cells harboring the expression plasmids were grown at 37˚C in LB medium[25] supplemented with 0.1% (w/v) glucose and kanamycin (25 μg/ml). Production of recombinant proteins was then induced by adding rhamnose and isopropyl β-D-1-thiogalactopyranoside to final concentrations of 0.1% (w/v) and 1 mM, respectively, when the optical density of the cultures at 600 nm reached at about 0.4. The cells were harvested at 2 hours after the induction.

Preparation of recombinant proteins. Unless otherwise stated, all of the following steps were carried out at 0-4˚C. The *E. coli* cells containing recombinant proteins were disrupted with a sonicator (Ohtake Works, model 5202PZT) in 20 mM sodium phosphate (pH 7.8), 0.5 M NaCl. After centrifugation for 10 min at 10,000 *g*, supernatants were applied onto Ni-NTA agarose columns (Qiagen). After washing the columns with the same buffer, recombinant proteins were eluted by stepwise washing with 20 mM sodium phosphate (pH 6.8), 0.5 M NaCl containing 0, 10, 50, 100 and 150 mM imidazole. Enzymatically active proteins for NIES39_K04640 and NIES39_K04650 eluted from the columns with the buffer containing 150 mM and 50 mM imidazole, respectively. The protein fractions were dialyzed against 10 mM Tris-HCl (pH 7.5), 50 mM KCl, 0.5 mM dithiothreitol for 16 hours and then against the same buffer containing 50% glycerol for 8 hours. The protein solutions were stored at -25˚C. Protein concentrations were determined by the Bradford method, with bovine γ-globulin as a standard.[26]

Enzymatic reaction with K04640 and K04650 proteins. Unless otherwise stated, reactions by the K04640 protein were carried out at 37˚C for 2 hours in 10 μl of a reaction mixture containing the K04640 protein (1 μg), 20 mM Tris-HCl (pH 8.5), 10 mM $MgCl_2$, 50 mM KCl, and 1 mM DTT, using the DNA (0.5 μg) of bacteriophage λ*c*I857*S*am7 (Takara Bio) as a substrate. In the experiments in Fig. 2, buffer or salt was changed from this solution as indicated in the figure. For complete digestion of DNA with the K04640 protein, incubation was prolonged to 16 hours. The reaction was terminated by adding an equal volume of 20 mM

Tris-HCl (pH 7.5), 6 mM EDTA, 0.1% SDS, 5% glycerol and 0.005% bromophenol blue. The reaction products were electrophoresed on 1% agarose gels followed by staining with ethidium bromide. Stained DNA was visualized and analyzed with a LAS-3000UVmini lumino-image analyzer (Fujifilm).

Reactions by the K04650 protein were performed at 37°C for 2 hours in a mixture containing the protein (80 ng/μl), DNA substrate (50-70 ng/μl), 50 mM Tris-HCl (pH 8.0), 50 mM NaCl, 10 mM EDTA and 133 μM *S*-adenosylmethionine. After recovering DNA with phenol extraction and ethanol precipitation, DNA was digested with the K04640 protein or other restriction enzymes to analyze modifications in the DNA. The digestion products were analyzed by electrophoresis on either 1% gels of UltraPure Agarose (Invitrogen) or 4% gels of NuSieve 3:1 Agarose (Takara Bio) depending on the size of the substrate.

Determination of the cleavage site by the K04640 protein. Five picomoles of a synthetic oligonucleotide , 5'-AGCTTGGATGCCAGCTGCAGATCGACA-3', containing the recognition sequence of the K04640 protein as underlined in the nucleotide sequence was labeled with [γ-^{32}P]ATP (PerkinElmer) and T4 polynucleotide kinase (Takara Bio). After the labeling reaction, enzyme was inactivated by heating for 15 minutes at 70°C, and the labeled oligonucleotide was mixed with 10 pmoles of a complementary oligonucleotide. The resulting double-stranded DNA substrate was digested with the K04640 protein and electrophoresed on a 20% polyacrylamide (1/20 bisacrylamide) gel containing 8.3M urea, along with a sequencing ladder prepared by mixing 5'-[^{32}P]-labeled synthetic oligonucleotides that had the same nucleotide sequence as the radiolabeled strand of the substrate but had successive deletions in the 3'-end region over the recognition sequence of the K04640 protein. Labeled DNA molecules on the gel was visualized and analyzed with the image analyzer FLA-7000 (Fujifilm).

Preparation of a DNA fragment for determining the modification site by the K04650 protein. A double-stranded oligonucleotide with overhangs compatible to *Hin*dIII site and *Pst*I site was prepared by annealing two oligonucleotides having the sequences, 5'-AGCTTGGATGCCAGCTGCA-3' and 5'-GCTGGCATCCA-3'. This molecule was inserted between the *Hin*dIII and *Pst*I sites of pUC118[27] to construct pUC118-Pv. This plasmid had a newly introduced *Pvu*II site that partly overlaps a *Pst*I site (see Fig. 4B for the nucleotide sequence of the relevant region). A DNA fragment including the multiple cloning site sequence region of pUC118-Pv was prepared by PCR with two primers that hybridized to the pUC118-Pv at the regions flanking the multiple cloning site sequence. The nucleotide sequences of the PCR primers used were 5'-GGCTCGTATGTTGTGTGGAATT-3' and 5'-GCAAGGCGATTAAGTTGGGTAA-3', which were designed to hybridize to pUC plasmids

in the "reverse" and "universal" orientations, respectively.[28] The PCR product was purified by electrophoresis on a 1.5% gel of low melting point agarose (SeaPlaque GTG agarose, Takara Bio).

BLAST Searches. The BLASTP search for sequences similar to the amino acid sequences of R/M system proteins were carried out using either CyanoBase Similarity Search (http://genome.microbedb.jp/blast/blast_search/cyanobase/genes) [24] for *A. platensis* NIES-39 or NCBI/BLAST (http://blast.ncbi.nlm.nih.gov) for *A. platensis* C1 (taxid: 459495) and *A. platensis* sp. PCC 8005 (taxid: 376219). Amino acid sequences of the following proteins were used as query sequences for the BLASTP search (shown in parentheses are UniProt accession numbers[29]): *Bsi*WI (E5Q8V3), *Tth*111I (G3JXC7), *Hae*III (O68584), *Afa*22MI (Q762K1), *Pvu*II (P23657), *Hin*dIII (P43870), *Sna*BI (Q6SA25), *Hgi*CI (P25258), *Hgi*DI (P24599), *Pst*I (P00640), *Eco*29kI, Q46944. *Bsu*BI (P33562), *Nsp*I (Q9ZHP2), *Eco*KMrr (P24202), *Bsa*WI (Q6UQ65), *Age*I (Q9KHV6), *Bpu*10I α subunit (O52851), *Bpu*10I β subunit (O52852), *Bbv*CI subunit 1 (Q5D6Y5), *Bbv*CI subunit 2 (Q5D6Y4) and M.*Afa*22MI (Q762K2). The sequences whose E-values were $< e^{-35}$ were judged to be homologs. When E-value of the most similar sequence was > 0.01, it was judged that homolog was absent in the genome.

Acknowledgement

We thank Tan Inoue for critical reading of the manuscript. This work was supported financially in part by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1) "*Spirulina platensis* (*Arthrospira*): Physiology, Cell-Biology and Biotechnology," ed. Vonshak A, Taylor & Francis, London. 1997.

2) Habib MAB, Pariv M, Huntington TC, and Hasan MR. "A review on culture, production and use of spirulina as food for humans and feeds for domestic animals and fish," FAO Fisheries and Aquaculture Circular No. 1034. Food and Agriculture Organisation (FAO) of the United Nations, Rome, Italy (2008).

3) Vonshak A, Boussiba S, Abellovich A, and Richmond A, *Biotech. Bioeng.*, **25**, 341-349 (1983).

4) Kawamura M, Sakakibara M, Watanabe T, Kita K, Hiraoka N, Obayashi A, Takagi M, and Yano K, *Nucleic Acids Res.*, **14**, 1985-1989 (1986).

5) Kumano M and Sakakibara M, *Baiosaiensu To Indasutori* (in Japanese), **46**, 3512-3515 (1988).

6) Tragut V, Xiao J, Bylina EJ, and Borthakur D, *J. Appl. Phycol.*, **7**, 561-564, (1995).

7) Fujisawa T, Narikawa R, Okamoto S, Ehira S, Yoshimura H, Suzuki I, Masuda T, Mochimaru M, Takaichi S, Awai K, Sekine M, Horikawa H, Yashiro I, Omata S, Takarada H, Katano Y, Kosugi H, Tanikawa S, Ohmori K, Sato N, Ikeuchi M, Fujita N, and Ohmori M, *DNA Res.*, **17**, 85-103 (2010).

8) Cheevadhanarak S, Paithoonrangsarid K, Prommeenate P, Kaewngam W, Musigkain A, Tragoonrung S, Tabata S, Kaneko T, Chaijaruwanich J, Sangsrakru D, Tangphatsornruang S, Chanprasert J, Tongsima S, Kusonmano K, Jeamton W, Dulsawat S, Klanchui A, Vorapreeda T, Chumchua V, Khannapho C, Thammarongtham C, Plengvidhya V, Subudhi S, Hongsthong A, Ruengjitchatchawalya M, Meechai A, Senachak J, and Tanticharoen M, *Stand. Genomic Sci.*, **6**, 43-53 (2012).

9) Janssen PJ, Morin N, Mergeay M, Leroy B, Wattiez R, Vallaeys T, Waleron K, Waleron M, Wilmotte A, Quillardet P, Tandeau de Marsac N, Talla E, Zang C-C, and Leys N, *J. Bacteriol.*, **192**, 2465-2466 (2010).

10) Xu GL, Kapfer W, Walter J, and Trautner TA, *Nucleic Acids Res.*, **20**, 6517-6523 (1992).

11) Brown NL and Smith M, *FEBS Lett.*, **65**, 284-287 (1976).

12) Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev SK, Dryden DTF, Dybvig K, Firman K, Gromova ES, Gumport RI, Halford SE, Hattman S, Heitman J, Hornby DP, Janulaitis A, Jeltsch A, Josephsen J, Kiss A, Klaenhammer TR, Kobayashi I, Kong H, Krüger DH, Lacks S, Marinus MG, Miyahara M, Morgan RD, Murray NE, Nagaraja V, Piekarowicz A, Pingoud A, Raleigh E, Rao DN, Reich N, Repin VE, Selker EU, Shaw PC, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Van Etten JL, Vitor JMB, Wilson GG, and Xu SY, *Nucleic Acids Res.*, **31**, 1805-1812 (2003).

13) Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR and Bateman A, *Nucleic Acids*

Res., **38**(Database issue), D211-222 (2010).
14) Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, and Hulo N, *Nucleic Acids Res*., **38**(Database issue), D161-D166 (2010).
15) Blumenthal RM, Gregory SA, and Cooperider JS, *J. Bacteriol*., **164**, 501-509 (1985).
16) Butkus V, Klimasauskas S, Petrauskiene L, Maneliene Z, Lebionka A, and Janulaitis A, *Biochim. Biophys. Acta*, **909**, 201-207 (1987).
17) Nelson M, and McClelland M, *Nucleic Acids Res*. 17(Suppl), r389-r415 (1989).
18) Elhai J, Vepritskiy A, Muro-Pastor AM, Flores E, and Wolk CP, *J. Bacteriol.,* **179**, 1998-2005 (1997).
19) Ogawa T, and Terui G, *J. Ferment. Technol*., **48**, 361-367 (1970).
20) Hamana K, Miyagawa K, and Matsuzaki S, *Biochem. Biophys. Res. Comm*., **112**, 606-613 (1983).
21) Williams JGK, *Methods Enzymol*., **167**, 766-778 (1988).
22) Marmur J, *J. Mol. Biol.,* **3**, 208-218 (1961).
23) Wu X, Zarka A, and Boussiba S, *Plant Mol. Biol. Rep*., **18**, 385-392 (2000).
24) Nakao M, Okamoto S, Kohara M, Fujishiro T, Fujisawa T, Sato S, Tabata S, Kaneko T, and Nakamura Y, *Nucleic Acids Res*., **38**(Database issue), D379-D381 (2010).
25) Sambrook J and Russell DW, "Molecular Cloning, a Laboratory Manual" 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001).
26) Bradford MM, *Anal. Biochem*., **72**, 248–254 (1976).
27) Vieira J and Messing J, *Methods Enzymol*., **153**, 3-11 (1987).
28) Vieira V and Messing J, *Gene*, **19**, 259-268 (1982).
29) The UniProt Consortium, *Nucleic Acis Res*., **40** (D1), D71-D75 (2012).
30) Scheldeman P, Baurain D, Bouhy R, Scott M, Mühling M, Whitton BA, Belay A, and Wilmotte A, *FEMS Microbiol. Lett*., **172**, 213-222 (1999)

Table 1. Type II Restriction Enzymes in *Arthrospira* (*Spirulina*) Strains

Strain[a]	Origin	Type II restriction enzymes[b]										References
		*Bsi*WI	*Tth*111I	*Hae*III	*Pvu*I	*Pvu*II	*Hin*dIII	*Sna*BI	*Hgi*CI	*Hgi*DI	Others	
S. platensis siamese[c]	Ethiopia	+	+	+								4
S. platensis UTEX2340[c]	Chad	+	+		+	+						5
S. maxima UTEX2342[c]	Chad	+	+		+	+						5
Spirulina sp. K[c]	Kenya	+										5
S. platensis pacifica[c]	San Diego		+		+	+	+					6
Spirulina sp. M1[c]	Mexico							+	+	+		5
Spirulina sp. M2[c]	Mexico							+		+		5
Spirulina sp. J[c]	Jamaica							+		+		5
A. platensis NIES-39[d]	Chad	+	-	-	-[e]	-	-	+	+	+	6[f]	7, this study
A. platensis C1[d]	Chad	+	+	-	-[e]	+	+	+	+	+	N.A.[g]	8, this study
Arthrospira sp. PCC 8005[d]	India[h]	+	-	-	-[e]	+	+	+	+	+	N.A.[g]	9, this study

[a]Names of the strains are shown as appeared in original articles.

[b]Names of isoschizomers are shown. Blanks in this column indicate that there is no evidence of the corresponding enzymes; they do not necessarily mean absence of the enzymes.

[c]For these strains, enzyme activities detected in cellular extracts are indicated by +.

[d]For these strains, presence and absence of putative homologs in their genomes are shown by + and -, respectively. The genes detected in the genome of *A. platensis* NIES-39 are as follows (shown in parentheses are E-values): *Bsi*WI, NIES39_A08930 ($4e^{-98}$); *Sna*BI, NIES39_B00210 ($3e^{-75}$); *Hgi*CI, NIES39_J02970 ($1e^{-65}$); *Hgi*DI, NIES39_A03530 ($5e^{-87}$). Of these genes, NIES39_A08930 and NIES39_B00210 were not mentioned as putative restriction enzyme genes in ref. 7. The genes detected in the genome of *A. platensis* C1 are as follows (the numbers with "ZP_" at the beginnings are NCBI RefSeq IDs): *Bsi*WI, ZP_17051558 (=SPLC1_S050850) ($3e^{-126}$); *Tth*111I, ZP_17054749 (=SPLC1_S370190) ($4e^{-51}$); *Pvu*II, ZP_17054363 ($6e^{-78}$); *Hin*dIII, ZP_17051500 ($4e^{-37}$); *Sna*BI, ZP_17054873 (=SPLC1_S380190) ($2e^{-55}$); *Hgi*CI, ZP_17055080 ($5e^{-82}$); *Hgi*DI, ZP_17053038 ($2e^{-104}$). The genes detected in the genome of *Arthrospira* sp. PCC 8005 are as follows: *Bsi*WI, ZP_09785144 ($3e^{-126}$); *Pvu*II, ZP_09782544 ($2e^{-70}$); *Hin*dIII, ZP_09779362 ($4e^{-37}$); *Sna*BI, ZP_09781835 ($2e^{-55}$); *Hgi*CI, ZP_09780171 ($5e^{-82}$); *Hgi*DI, ZP_09784109 ($1e^{-106}$).

[e]These are based on the results of BLASTP search with the amino acid sequences of the proteins of *Afa*22MI R/M system (*Afa*22MI is isoschizomeric to *Pvu*I). Only solitary methylase genes encoding M.*Afa*22MI homologs are detected in the genomes of *A. platensis* NIES-39, *A. platenisis* C1, and *Arthrospira* sp. PCC8005. The corresponding methylase genes in these strains are NIES39_J00480, ZP_17056754 and ZP_09780102, respectively.

[f]These are putative restriction enzymes encoded by the following genes (shown in parentheses are names of structurally related restriction enzymes and E-values): NIES39_K04640 (*Pst*I, $1e^{-39}$), NIES39_A05820 (*Eco*29kI, $5e^{-61}$), NIES39_K04040 (*Bpu*10I α subunit, $1e^{-12}$; *Bpu*10I β subunit, $2e^{-21}$; *Bbv*CI subunit 1, $2e^{-21}$; *Bbv*CI subunit 2, $2e^{-11}$), NIES39_L01700 (*Nsp*I, $2e^{-73}$), NIES39_M00570

(*Eco*KMrr, $1e^{-17}$), NIES39_O02590 (*Age*I, $2e^{-94}$). In the genomes of both *A. platensis* C1 and *Arthrospira* sp. PCC 8005, genes with significant homology to NIES39_L01700, NIES39_M00570 and NIES39_O02590 are detected, but those related to NIES39_K04640, NIES39_A05820 and NIES39_K04040 are not found.

[g]Not available. The numbers of putative type II restriction enzyme genes in these strains are not given in refs. 8 and 9.

[h]Exact origin of this strain is unknown, but its sequence of internally transcribed spacer (ITS) was reported to be identical to that of an Indian strain, *Arthrospira indica* (MCRC isolate straight).[30)]

Figure legends

Fig. 1. The genes, plasmids and proteins used in this study.

(A) Schematic representation of the genomic region in *A. platensis* NIES-39 around a putative type II restriction enzyme gene, NIES39_K04640, and a putative DNA (cytosine-5-)-methyltransferase gene, NIES39_K04650. Numbers in parentheses show the genomic positions of the open reading frames for these genes. (B) Structures of the plasmids for overproducing the proteins for NIES39_K04640 and NIES39_K04650 in *E. coli*. Amino acid sequences of the N-terminal and C-terminal regions of the encoded proteins are also shown below the structure of the plasmids. Shown in italics are amino acid sequences that are derived from synthetic oligonucleotides used for constructing the plasmids and that are not encoded by the genomic sequences. (C) Extracts of *E. coli* KRX cells harboring pET28b(+) vector and those harboring pET-K04640 were electrophoresed on a 12% SDS-polyacrylamide gel, along with the recombinant K04640 protein (3 μg) purified from the cell extracts. (D) Extracts of *E. coli* KRX cells harboring pET28b(+) and those harboring pET-K04650 were electrophoresed on a 12% SDS-polyacrylamide gel, along with the recombinant K04650 protein (3 μg) purified from the cell extracts. Positions of molecular weight markers are shown on the right of the panels in D and E.

Fig. 2. Effect of pH and monovalent cation on the activity of K04640 protein

(A) Effect of pH on the reaction by K04640 protein. DNA of λ phage was incubated at 37˚C for 2 hours with the K04640 protein under various pH conditions and electrophoresed on a 1% agarose gel. The Tris-HCl and Glycine-KOH buffers with the indicated pH values were used in the reactions. Also electrophoresed were λ phage DNA that was not treated with enzymes ("No enzyme"), λ phage DNA digested with *Hin*dIII ("HindIII"), and λ phage DNA that was completely digested with the K04640 protein by incubating for 16 hours ("Complete digestion"). (B) Relative activity of the K04640 protein under various pH conditions. After performing experiments described in A, the intensity of the bands shorter than 7kb was quantitated and divided by the intensity of the whole bands on the same lane. The obtained values give rough estimate of the enzyme activity under the conditions used. The relative values relative to a maximal value were then calculated and plotted as "relative activity". Reactions in the Tris-HCl and Glycine-NaOH buffers are shown in open circles and closed circles, respectively. (C) Effect of monovalent cations on the activity of K04640 protein. DNA of λ phage was incubated with the K04640 protein in the presence of various concentrations of KCl (open circles), NaCl (closed circles) or NH_4Cl (triangles). Relative activity for each reaction was calculated as described above.

Fig. 3. Recognition sequence and the cleavage site by the K04640 protein

(A) Digestion pattern of λ phage DNA by *Pst*I and the K04640 protein. The λ phage DNA was completely digested with *Pst*I or the K04640 protein and electrophoresed on a 1% agarose gel. Intact λ phage DNA was also electrophoresed ("No Enzyme"). (B) Determination of the cleavage site by the K04640 protein. A double-stranded DNA that is radiolabeled with ^{32}P at the 5' end of one strand was cleaved with the K04640 protein. The digestion products was electrophoresed on an 20% polyacrylamide gel containing 8M urea, along with a sequencing ladder (lanes G, A, T and C) prepared by end-labeling a series of synthetic oligonucleotides with ^{32}P.

Fig. 4. Modification of DNA by the K04650 protein

(A) *S*-adenosylmethionine-dependent modification of DNA by the K04650 protein. The λ phage DNA was incubated with the K04650 protein in the absence ("K04650, No SAM") or presence ("K04650+SAM") of *S*-adenosylmethionine. Aliquots of each DNA were then digested with *Apl*I (K04640 protein), *Pst*I, *Eco*RI or *Hin*dIII. The reaction products were electrophoresed on a 1% agarose gel. The λ phage DNA that was not treated by any enzyme ("λ DNA") and the DNA that was treated with the K04650 protein but not with nucleases ("No Nuclease") was also electrophoresed. (B) Determination of the modification site by the K04650 protein. A DNA fragment containing the nucleotide sequence shown in the top panel was reacted with the K04650 protein. In the nucleotide sequence, the four C residues that are possible targets of modification by the K04650 protein are shown in white letters on black background. The modified fragment was then digested with *Apl*I (K04640 protein), *Pvu*II, *Hpy*CH4V, *Pst*I, or *Hin*dIII. The reaction products were electrophoresed on a 4% NuSieve 3:1 Agarose gel.

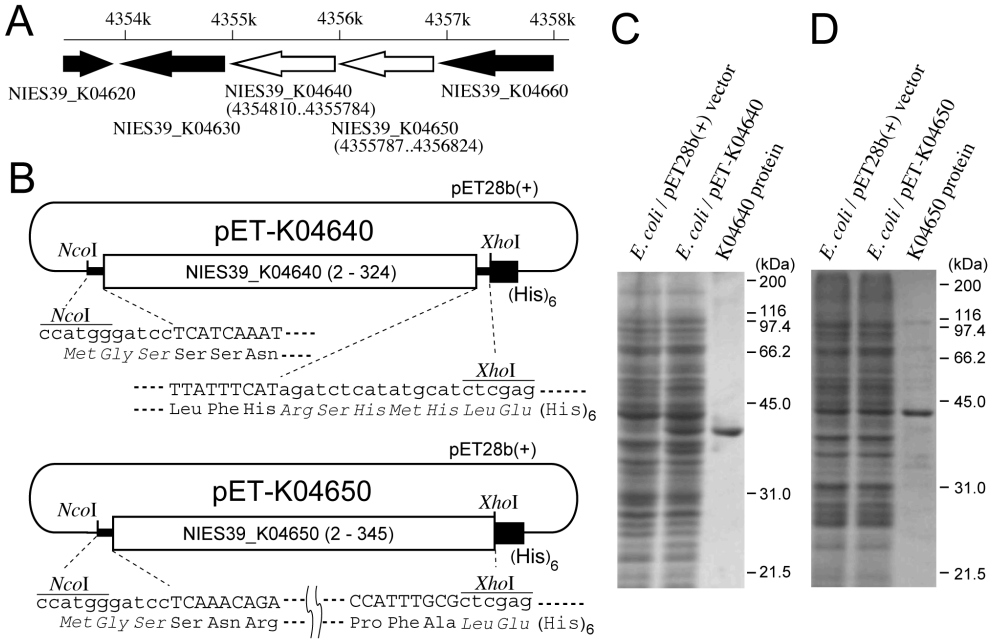

Fig.1 Shiraishi & Tabuse

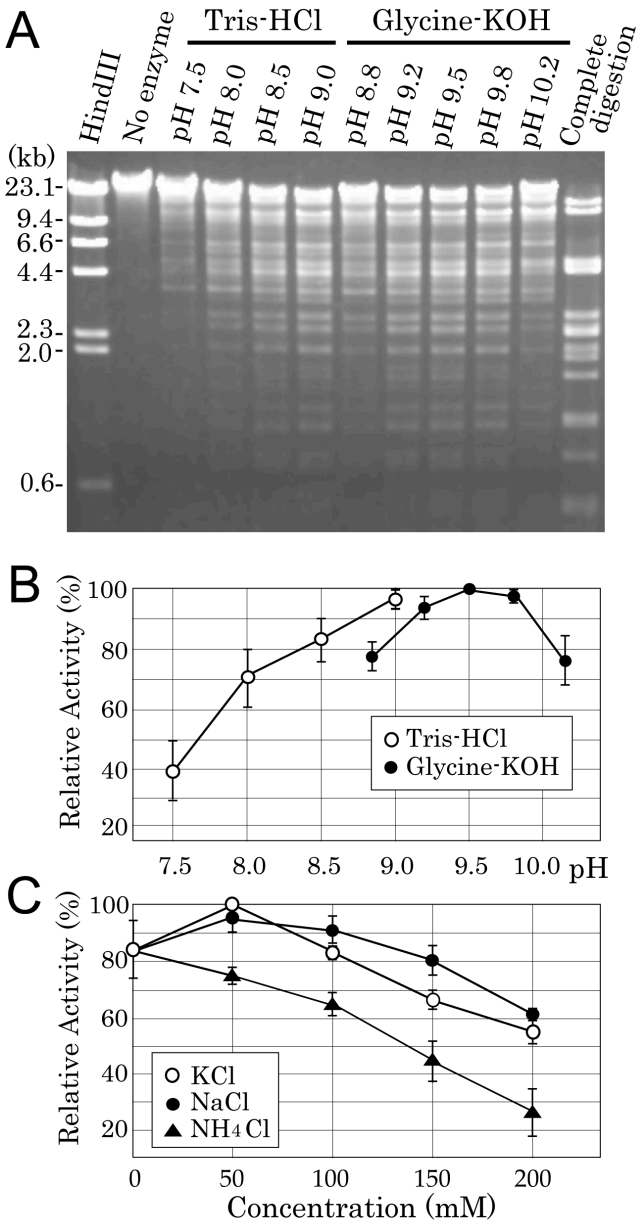

Fig. 2　Shiraishi & Tabuse

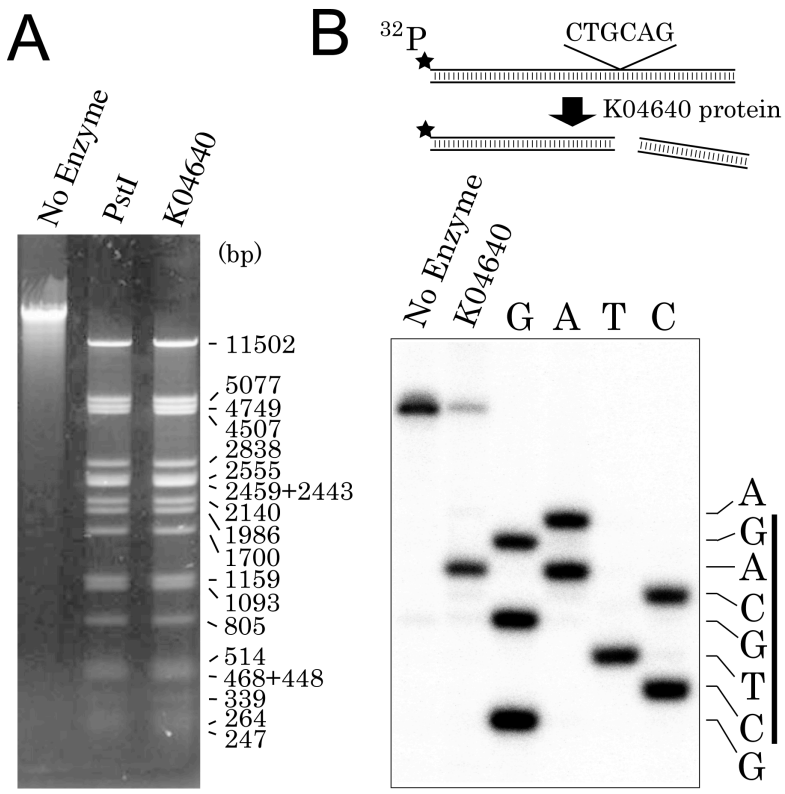

Fig. 3　Shiraishi & Tabuse

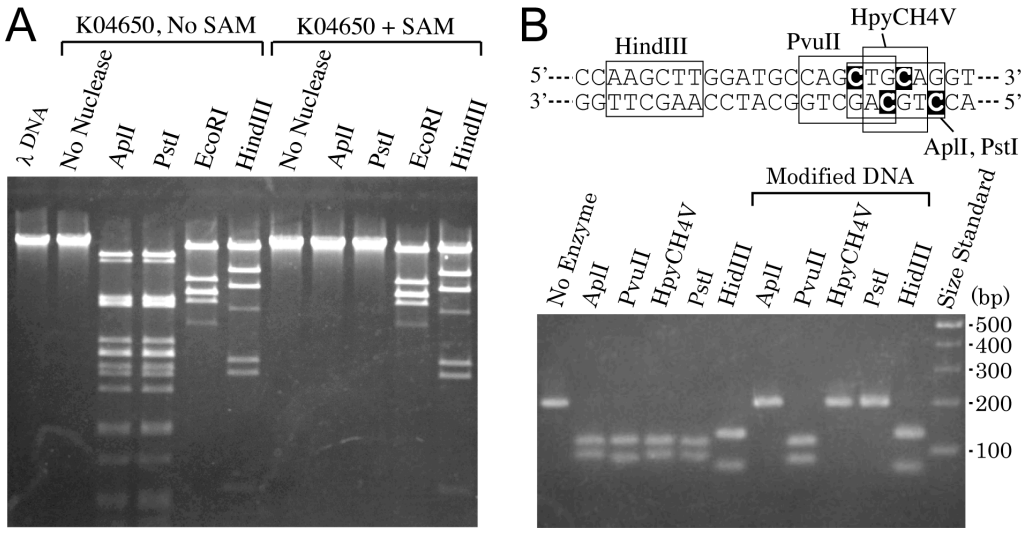

Fig. 4　Shiraishi & Tabuse