Evidence for Horizontal Transfer of *Ssu*DAT1I Restriction-Modification Genes to the *Streptococcus suis* Genome

TSUTOMU SEKIZAKI,[1]* YOSHIKO OTANI,[2] MAKOTO OSAKI,[1] DAISUKE TAKAMATSU,[1]
AND YOSHIHIRO SHIMOJI[1]

*National Institute of Animal Health, Tsukuba, Ibaraki 305-0856,[1] and Kenhoku Livestock
Hygiene Service Center, Mito, Ibaraki 310-0002,[2] Japan*

Different strains of *Streptococcus suis* serotypes 1 and 2 isolated from pigs either contained a restriction-modification (R-M) system or lacked it. The R-M system was an isoschizomer of *Streptococcus pneumoniae* *Dpn*II, which recognizes nucleotide sequence 5′-GATC-3′. The nucleotide sequencing of the genes encoding the R-M system in *S. suis* DAT1, designated *Ssu*DAT1I, showed that the *Ssu*DAT1I gene region contained two methyltransferase genes, designated *ssuMA* and *ssuMB*, as does the *Dpn*II system. The deduced amino acid sequences of M.*Ssu*MA and M.*Ssu*MB showed 70 and 90% identity to M.*Dpn*II and M.*Dpn*A, respectively. However, the *Ssu*DAT1I system contained two isoschizomeric restriction endonuclease genes, designated *ssuRA* and *ssuRB*. The deduced amino acid sequence of R.*Ssu*RA was 49% identical to that of R.*Dpn*II, and R.*Ssu*RB was 72% identical to R.*Lla*DCHI of *Lactococcus lactis* subsp. *cremoris* DCH-4. The four *Ssu*DAT1I genes overlapped and were bounded by purine biosynthetic gene clusters in the following gene order: *purF-purM-purN-purH-ssuMA-ssuMB-ssuRA-ssuRB-purD-purE*. The G+C content of the *Ssu*DAT1I gene region (34.1%) was lower than that of the *pur* region (48.9%), suggesting horizontal transfer of the *Ssu*DAT1I system. No transposable element or long-repeat sequence was found in the flanking regions. The *Ssu*DAT1I genes were functional by themselves, as they were individually expressed in *Escherichia coli*. Comparison of the sequences between strains with and without the R-M system showed that only the region from 53 bp upstream of *ssuMA* to 5 bp downstream of *ssuRB* was inserted in the intergenic sequence between *purH* and *purD* and that the insertion target site was not the recognition site of *Ssu*DAT1I. No notable substitutions or insertions could be found, and the structures were conserved among all the strains. These results suggest that the *Ssu*DAT1I system could have been integrated into the *S. suis* chromosome by an illegitimate recombination mechanism.

More than 3,000 restriction-modification (R-M) systems have been identified in a wide variety of microorganisms, where they are thought to protect the host from invasion by foreign DNA. Only a minority of R-M systems have been sequenced (6, 37, 56). Among them, some type II R-M systems, which recognize 4-bp palindromic sequence 5′-GATC-3′, involve a variety of isoschizomers. Three classes can be distinguished by their manners of DNA cleavage and susceptibilities to DNA methylation. The first class of isoschizomers, represented by *Sau*3AI, which was described for *Staphylococcus aureus*, is prevented from digesting host DNA by a cognate 5-methylcytosine methyltransferase and is not influenced by the modification of N^6-methyladenine (48, 56). The second class, represented by *Dpn*I, which was described for *Streptococcus pneumoniae*, is unique among restriction endonucleases in that it cleaves only at methylated DNA sequence 5′-GmeATC-3′, and thus the cells producing *Dpn*I do not carry the corresponding methyltransferase gene (22, 23). The third class, including *Mbo*I, *Dpn*II, and *Lla*DCHI, which were described for *Moraxella bovis*, *S. pneumoniae*, and *Lactococcus lactis*, respectively, contains restriction endonucleases complementary to the second class, in that they cleave at the same sequence, 5′-GATC-3′, but only when it is unmethyl-

ated (11, 22, 23, 30, 55). Bacterial cells expressing the third class of restriction endonucleases also contain an N^6-methyladenine methyltransferase, which methylates adenine in the 5′-GATC-3′ sequences of the host DNA. The N^6-methyladenine methyltransferase is functionally identical to the Dam methylase of *Escherichia coli*, and the evolutionary relatedness of these methylases has been discussed (29, 35).

Except for the second class, the type II R-M systems described above are generally composed of two structural genes, one for an endonuclease and a second for a methyltransferase, as is typical of other type II R-M systems (6, 56). However, two of these, *Dpn*II and *Lla*DCHI, have genes that encode two methyltransferases (8, 24, 30). A study of well-characterized *Dpn*II and a comparison of it with the *Lla*DCHI system indicated that in addition to the conventional double-stranded DNA methyltransferase, a single-stranded DNA methyltransferase was encoded, and this activity potentially facilitated the uptake of intact DNA via conjugation in the respective host bacteria. Another property of interest is that the two methyltransferases of the *Dpn*II system, M.*Dpn*II and M.*Dpn*A, were highly homologous to those of the *Lla*DCHI system, M.*Lla*A and M.LlaB, respectively (75 and 88% identity). Thus the two M systems are thought to have the same origin, whereas the two restriction endonucleases showed relatively low homology (only 31% identity), suggesting the complexity of the evolutionary origin of restriction endonuclease genes (30).

Many genes for R-M systems have been found to be located on transferable elements such as plasmids and bacteriophages,

* Corresponding author. Mailing address: Laboratory of Molecular Bacteriology, National Institute of Animal Health, 3-1-1 Kannondai, Tsukuba, Ibaraki 305-0856, Japan. Phone: 81 (298) 38-7743. Fax: 81 (298) 38-7907. E-mail: sekizaki@niah.affrc.go.jp.

or, in some cases, genes encoding proteins involved in DNA mobility, such as transposases, integrases, and invertases, are found in the vicinity of R-M systems (5, 7, 19, 26, 38, 46, 51, 54, 56). These genetic structures may facilitate the transfer of R-M systems and have led to the speculation that R-M genes could migrate among microorganisms of different genera. The acquisition or loss of an R-M system followed by a double-strand breakage may play a role in illegitimate recombination when the two genetic segments have a few base pairs of homology (21). The above findings, together with recent studies on the complete sequences of bacterial genomes, have led to a proposal that R-M systems are likely to be mobile genetic elements that mediate large-scale chromosomal conversions and that such genome plasticity may play a role in genome dynamics (20). However, these studies have thus far been conducted with only a few strains or with few different species (3, 17, 20, 47, 58). While the results of the above studies imply horizontal transfer of R-M systems, the precise locations of R-M systems in the chromosome and/or the genetic structures flanking the R-M systems have yet to be studied extensively. Therefore, further studies will be needed in order to make a more generalized statement relative to the mobility of R-M systems.

Streptococcus suis is a gram-positive, facultatively anaerobic coccus that has been implicated as the cause of a wide range of clinical disease syndromes in swine and other domestic animals. The disease syndromes caused by *S. suis* in swine include arthritis, meningitis, pneumonia, septicemia, endocarditis, polyserositis, abortions, and abscesses (9). *S. suis* has also been implicated in human disease (28). *S. suis* strains are currently classified into 35 serotypes on the basis of their capsular polysaccharide antigens (12, 13, 36). Among them, strains frequently isolated from diseased pigs and exhibiting a high degree of virulence mostly belong to serotype 2 (9, 15, 16). Although several genes of *S. suis* have been cloned and characterized (34, 41, 43–45, 49), the R-M system(s) of this bacterial species has not been described.

We found that some strains of serotypes 1 and 2 of *S. suis* possessed an R-M system that mimicked the *Dpn*II and *Lla*D-CHI systems, whereas the reference strain of serotype 2, NCTC10234, as well as other strains lacked this R-M system. This report describes the genetic organization of the R-M system and its flanking regions. The R-M system was found to contain two genes encoding methyltransferases similar to those of the *Dpn*II and *Lla*DCHI systems and two genes encoding restriction endonucleases, one similar to R.*Dpn*II and the other highly homologous to R.*Lla*DCHI. Comparison of the DNA sequences of the R-M system and its flanking regions among different strains containing the R-M system and those lacking the system indicated a horizontal transfer of the R-M system, which may have occurred by a unique mechanism of gene transfer.

MATERIALS AND METHODS

Bacterial strains, vectors, enzymes, media, and culture conditions. *S. suis* strains of serotypes 1 and 2 used are listed in Table 1. Strain NCTC10234, a reference strain of serotype 2, was purchased from the National Collection of Type Cultures, Central Public Health Laboratory, London, England. Strain DAT1 was previously described (49). Other *S. suis* strains were independently isolated from diseases pigs in Japan and stocked in our laboratory. *E. coli* strains used were SCS110 (*thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44 rpsL* Δ(*lac-proAB*) [F′ *traD36 proAB lacI*�q*Z*ΔM15]) (Stratagene, La Jolla, Calif.),

TABLE 1. Strains of *S. suis* used and their R-M profiles

Strains[a]	Serotype
Strains expressing the R-M phenotype	
DAT1, 193, 195, 196, 197, 198, 199, 200, 202, 205, 220, 221, 222, 227, 228, 229, 230, 233, 234, 235, 236, 238, 239, 243, 244, 247	2
211, 212	1
Strains with the null phenotype	
NCTC10234, DAT2, 194, 207, 209, 210, 213, 223, 226, 246	2
203, 204	1

[a] Strains except NCTC10234 were isolated from diseased pigs in Japan from 1987 to 1996.

C600 (F⁻ *thr-1 thi-1 leuB6 lacY1 tonA21 supE44*) (39), XL-1 blue MRF′ [Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F′ *proAB lacI*ᵍ*Z*ΔM15 Tn*10*]) (Stratagene), XLOLR {Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 *endA1 thi-1 recA1 gyrA96 relA1 lac* [F′ *proAB lacI*ᵍ*Z*ΔM15 Tn*10*] Suʳλ⁻} (Stratagene), and DH5α (F⁻ *endA1 recA1 hsdR17* (rₖ⁻ mₖ⁺) *supE44 thi-1 gyrA relA1* ϕ80*lacZ*ΔM15 Δ(*lacZYA-argF*) *deoR*⁺) (39). The plasmid vectors used for gene cloning were pUC19 (59), pCR2.1 (Invitrogen, Groningen, The Netherlands), and pHSG576 (50). Phagemid vector λZAP Express (Stratagene) was used for the construction of the genomic library. All restriction endonucleases and other enzymes were purchased from Takara Shuzo Co. Ltd. (Tokyo, Japan) except *Dpn*I, which was purchased from Boehringer GmbH (Mannheim, Germany), and used according to the manufacturers' recommendations. *S. suis* strains were grown in Todd-Hewitt (TH) broth or agar medium (Difco Laboratories, Detroit, Mich.) at 37°C under 5% CO₂ for 18 h. *E. coli* strains were cultured in Luria-Bertani broth or agar medium (Difco Laboratories) supplemented with, when necessary, ampicillin (50 μg/ml), kanamycin (25 μg/ml), and chloramphenicol (25 μg/ml) at 37°C for 18 h.

Analysis of DNA methylation. Genomic DNA was isolated from *S. suis* and *E. coli* grown on agar medium by standard procedures as described previously (34, 42). Methylation of DNA was determined directly by testing the susceptibility of genomic DNA to the restriction endonucleases *Dpn*I and *Mbo*I in vitro. Restriction digests were analyzed by agarose gel electrophoresis by standard procedures (39).

Restriction endonuclease assay in crude extracts. The presence of restriction endonucleases could be detected in crude extracts of *S. suis* and *E. coli*. Extracts of *S. suis* were prepared according to the procedures described by Muckerman et al. (31). Briefly, 5-ml cultures of *S. suis* were grown in TH broth to an optical density at 600 nm of approximately 0.4. After centrifugation at 20,000 × *g* for 2 min at 4°C, the cell pellet was suspended in 0.1 ml of 10 mM Tris-HCl (pH 7.6)–50 mM NaCl. Then 2.5 μl of Triton X-100 was added, and the mixture was incubated at 30°C for 15 min for cell lysis. After incubation, the cells were removed by centrifugation at 20,000 × *g* for 5 min at 4°C, and the extracts were directly used for the assay. Extracts of *E. coli* were prepared as described by Schleif (40). Briefly, bacterial cells collected from 1 ml of an overnight shaking culture in Luria-Bertani broth supplemented with appropriate antibiotics were suspended in 0.1 ml of 10 mM Tris-HCl (pH 7.6)–10 mM 2-mercaptoethanol in Eppendorf tubes. The cell suspensions were then sonicated on ice with a Sonifier 250 (Branson Ultrasonics Corp., Danbury, Conn.) with a microtip. Sonication was carried out twice for 10 s with a 1-min interval and with a 50% duty cycle and an output of 7. After sonication, the cell debris was removed by centrifugation at 20,000 × *g* for 15 min at 4°C, and the supernatants were directly used for the assay. Enzyme reactions were performed by standard procedures for restriction endonuclease digests with a high-salt universal buffer (39).

PCR amplifications. Synthetic oligonucleotide primers specifically designed for amplification of the R-M gene region in this study are listed in Table 2. *Ex Taq* polymerase (Takara) was used for the amplification according to the manufacturer's instructions except that the concentration of MgCl₂ was 3 mM. DNA amplification was carried out in a Perkin-Elmer thermal cycler, model 2400 or 9600 (PE Biosystems Japan, Tokyo, Japan), and the program consisted of incubation for 1 min at 96°C, 30 cycles of 20 s at 96°C, 20 s at 60°C, and 3 min at 68°C, and final incubation for 5 min at 68°C.

An LA PCR kit (Takara) was used for inverse PCR according to the manufacturer's instructions except that the concentration of MgCl₂ was 3 mM. The rationale and the protocol for inverse PCR were essentially the same as those described elsewhere (32). Genomic DNA of *S. suis* strain DAT1 was digested with *Bam*HI and self-ligated using a DNA ligation kit (Takara). The ligation mixture was used as a template for the inverse PCR.

TABLE 2. Oligonucleotide primers

Primer	Sequence (5′–3′)[b]	Location or description
dam-F	YRAARTGGRCWGGKGGHAAR[a]	Degenerate primer
dam-R	CGDCCRWAYGGHACRTTRAA[a]	Degenerate primer
dam-in1	TTTCCCCCTGCTCTGCAATT	*purH* region near *Bam*HI site
dam-in2	CAGTGGGTATGGCCTTAGAA	*ssuMB* region
purH3′	CAGCTGCAGGTATCAAGGCTAT	3′ end of *purH* gene
ssuMB3′	CACATCTGGATACTTGTCAC	5′ end of *ssuRA* gene
ssuRA5′Ec	GGAATTCGCTTCAACTCAAGAAGGAGA	3′ end of *ssuMB* gene with *Eco*RI site
ssuRA3′Pst	AACTGCAGGTAATCTGGCGTTCGATTAG	5′ end of *ssuRB* gene with *Pst*I site
ssuRB5′Ec	GGAATTCTGGTGAAGGTTGGCAAAGGT	3′ end of *ssuRA* gene with *Eco*RI site
ssuRB3′Pst	AACTGCAGCCTGCTCAGACTCTAACAAC	5′ end of *purD* gene with *Pst*I site
ssuRB3′	CCTGCTCAGACTCTAACAAC	5′ end of *purD*

[a] Y, C or T; R, A or G; W, A or T; K, G or T; H, A or C or T; D, A or G or T.
[b] Unique restriction cleavage sites introduced in the oligonucleotides are underlined.

Construction of genomic library and cloning procedures. Genomic DNA of *S. suis* DAT1 was partially digested with *Sau*3AI and size fractionated by standard procedures (39). DNA fragments of approximately 5 to 7 kb were ligated with *Bam*HI-digested dephosphorylated λZAP Express vector (Stratagene) and subjected to in vitro packaging with a GigaPack III kit (Stratagene) according to the instructions of the manufacturer. The library was screened via plaque hybridization using standard procedures (39). Phages from the hybridizing plaques were purified. The phagemids were rescued by helper phage EXASSIST (Stratagene) and used to infect *E. coli* XLOLR to create plasmid subclones by following the instructions of the manufacturer.

Construction of subclones was essentially accomplished by digestion and self-ligation procedures using a restriction cleavage site in a cloned fragment and the multicloning sites of the vector plasmid. The plasmids were digested with an appropriate restriction endonuclease and self-ligated to create subclones in which a part of the original insert fragment had been deleted. The combinations of restriction enzymes and the subclones constructed were as follows. For pDAM4, digestion with *Kpn*I, *Sac*I, *Pst*I, *Cla*I, and *Spe*I generated pDAM4K, pDAM4Sa, pDAM4P, pDAM4C, and pDAM4Sp, respectively; for pDAM7, digestion with *Ava*I generated pDAM7Av; for pDAM9, digestion with *Kpn*I, *Sac*I, *Pst*I, *Cla*I, and *Spe*I generated pDAM9K, pDAM9Sa, pDAM9P, pDAM9C, and pDAM9Sp, respectively; for pDAM10, digestion with *Pst*I generated pDAM10P. Some of the subclones were modified further by the same procedures. pDAM4C and pDAM9C digested with *Pst*I generated pDAM4CP and pDAM9CP, respectively. pDAM4Sa digested with *Eco*RI generated pDAM4SaE. A 1.5-kb *Hinc*II fragment of pDAM9K was subcloned in pUC19; the recombinant plasmid was designated pDAM9KHc. These subclones were used for sequencing the DNA of the cloned fragments using universal primers designed for the different vector plasmids.

Hybridization techniques. Genomic Southern hybridization and plaque hybridization were performed as described previously (39) with a DIG DNA labeling and detection kit (Boehringer GmbH). Restriction endonuclease fragments were separated on agarose gels and transferred to positively charged nylon membranes (Nytran Plus; Schleicher & Schuell, Dassel, Germany) by using the vacuum transfer method with VacuGene (Pharmacia-LKB Biotechnology AB, Uppsala, Sweden) without depurination in accordance with the instructions of the manufacturer. Plaque hybridization was carried out on a positively charged nylon membrane (MAGNA LIFT; Micron Separations Inc., Westboro, Mass.). Prehybridization and hybridization were carried out at 63°C for 2 and 16 h, respectively. After hybridization, the sheets were washed twice at room temperature for 5 min in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate [pH 7.2]) containing 0.1% sodium dodecyl sulfate (SDS), followed by two 15-min washes at 63°C in 0.1× SSC containing 0.1% SDS. DNA fragments cloned in plasmids were purified from agarose gels after electrophoresis and labeled with digoxigenin by essentially the same procedures described previously (42).

DNA sequencing and analysis. The sequencing of cloned DNA fragments and various PCR products was carried out by dye terminator chemistry with specifically designed primers on an Applied Biosystems model 373A automated DNA sequencer (PE Biosystems). The sequences were assembled and analyzed with Sequencer, version 2 (Hitachi Software Engineering Co., Ltd., Yokohama, Japan), and Genetyx-Mac, version 10.1 (Software Developing Company, Tokyo, Japan). Sequences were searched against current DNA and protein databases by BLAST program network services available at the National Center for Biotechnology Information (Bethesda, Md.) (http://www.ncbi.nlm.nih.gov).

Pulsed-field gel electrophoresis (PFGE) analysis. Preparation of genomic DNAs and their restriction digests was performed as described previously (33) with some modifications. Briefly, *S. suis* strains grown on TH agar were harvested and suspended in TES (10 mM Tris-HCl [pH 7.5], 5 mM EDTA, 0.15 M NaCl) to an optical density at 600 nm of 0.5. The cell suspensions were mixed with an equal volume of molten 3% low-melting-point preparative-grade agarose (Bio-Rad Laboratories, Richmond, Calif.) to obtain agarose plugs. Cells were lysed by submerging the plugs in lysis solution (*N*-acetylmuramidase [Seigagaku Kogyo, Tokyo, Japan] [30 µ/ml] and lysozyme [5 mg/ml] in TES) at 37°C for 1 h with gentle shaking. The lysis solution was replaced with ESP solution (0.5 M EDTA [pH 9.5], 1% [wt/vol] laurylsarcosine, 1 mg of proteinase K per ml), and the plugs were incubated at 50°C for 20 h. The plugs were washed several times with TE buffer (10 mM Tris-HCl, 2 mM EDTA [pH 8.0]) and stored at 4°C until use. The plugs were cut into small pieces, equilibrated with the recommended restriction buffer, placed in fresh restriction buffer supplemented with 0.02% [wt/vol] bovine serum albumin and 5 U of restriction enzyme *Sma*I, and then incubated at 37°C for 1 h. Contour-clamped homogeneous electric field (CHEF) electrophoresis was done in a Bio-Rad CHEF DRII system. Agarose gels (1.0% [wt/vol]) were electrophoresed in 0.5× TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA [pH 8.0]) for 20 h at 200 V and 10°C. The pulse times were ramped from 5 to 50 s.

Nucleotide sequence accession numbers. The sequence of the entire R-M region and its flanking regions of *S. suis* strain DAT1 obtained in this study has been deposited in the DDBJ/EMBL/GenBank database under accession no. AB045609. The DDBJ/EMBL/GenBank accession numbers of the genetic regions flanking the R-M genes of the strains tested are as follows: 205, AB045610 and AB045611; 220, AB045612 and AB045613; 211, AB045614 and AB045615; 212, AB045616 and AB045617. The DNA sequences for the *purH-purD* loci of strains NCTC10234, 213, 246, 203, and 204 have been deposited in the DDBJ/EMBL/GenBank database with accession no. AB045618, AB045619, AB045620, AB045621, and AB045622, respectively.

RESULTS

Restriction enzyme phenotypes of *S. suis*. The R-M phenotypes of *S. suis* strains were first examined by testing the susceptibility of their DNA to *Dpn*I and *Mbo*I. Cleavage by *Dpn*I and not by *Mbo*I indicates methylation of adenine at the 5′-GATC-3′ sequence and, hence, an *Mbo*I (or *Dpn*II) methylase phenotype. The converse indicates the absence of methylation. Total DNA prepared from 40 strains was incubated with either *Dpn*I or *Mbo*I. The results obtained are summarized in Table 1, and representative cleavage patterns of several strains are shown in Fig. 1A. Genomic DNA isolated from 12 of 40 strains was digested to small fragments by *Mbo*I; thus these bacteria had the unmethylated phenotype. On the other hand, the DNA from the remaining 28 strains could be digested only by *Dpn*I, indicating that their DNAs were protected by methylation from *Mbo*I digestion.
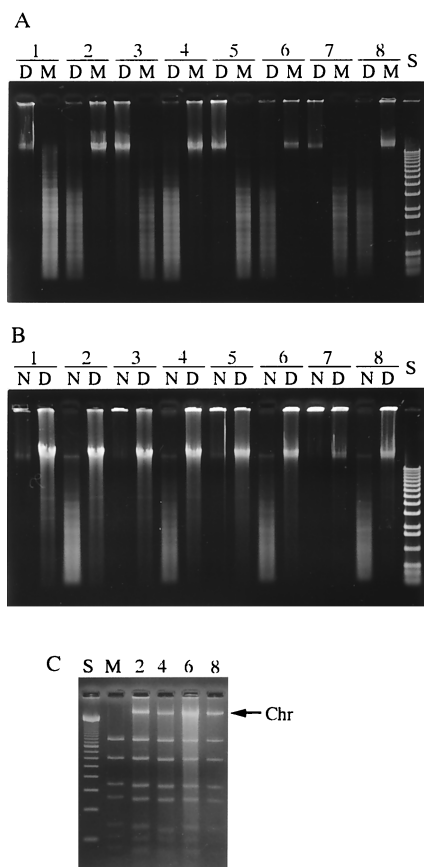
FIG. 1. (A) Methylation of 5′-GATC-3′ sequence in the DNA of various strains of *S. suis*. Total DNAs from the following strains were treated with two complementary restriction endonucleases and analyzed by agarose gel electrophoresis: Lane 1, NCTC10234; lane 2, DAT1; lane 3, 204; lane 4, 205; lane 5, 207; lane 6, 211; lane 7, 213; lane 8, 220. D, digested with *Dpn*I (specific for methylated sequence); M, digested with *Mbo*I (specific for unmethylated sequence). S, 1-kb ladder DNA size standards (GIBCO/BRL). (B) DNA cleavage activities of *S. suis* crude extracts using the total DNAs of *S. suis* NCTC10234 (N, unmethylated DNA) and DAT1 (D, methylated DNA) as substrates. Lane numbers for the strains from which the crude extracts were prepared are the same as in panel A. S, 1-kb ladder size standards (GIBCO/BRL). (C) Restriction cleavage activities of the crude extracts using unmethylated pUC19 as the substrate. The lane numbers of the strains from which the crude extracts were prepared match those in panel B. M, pUC19 digested with *Mbo*I; S, 100-bp ladder size standards (GIBCO/BRL); Chr, contaminating chromosomal DNA.

Crude extracts of the different strains were prepared, and their restriction endonuclease activities were assayed by digestion of the genomic DNA of strain NCTC10234, as an unmethylated DNA substrate, and that of DAT1, as a methylated DNA substrate. Restriction endonuclease activity was only detected in strains that exhibited the methylated DNA phenotype (Fig. 1B, lanes 2, 4, 6, and 8), whereas crude extracts of the strains that showed the unmethylated DNA phenotype digested neither methylated nor unmethylated DNA and were designated null phenotypes (Fig. 1B, lanes 1, 3, 5, and 7). These results were confirmed further by digestion of *E. coli* genomic DNA. *E. coli* K-12 strains are known to possess an orphan methylase called Dam, with a methylation property that is the same as that of M.*Mbo*I. Genomic DNA of *E. coli*

C600, with a Dam⁺ phenotype, and of SCS110, a Dam-deficient mutant, was digested with crude extracts of *S. suis*; a restriction endonuclease activity that could digest only unmethylated SCS110 DNA was seen in extracts of the strains which showed the methylated phenotype, but no endonuclease activity was seen in extracts of the strains which showed the null phenotype (data not shown). To examine the enzymatic specificity of extracts of strains that showed restriction endonuclease activity, we compared the restriction endonuclease cleavage patterns of vector plasmid pUC19 prepared from Dam-deficient mutant SCS110. Representative cleavage patterns are shown in Fig. 1C. Although chromosomal contamination derived from *S. suis* strains appeared, the cleavage patterns of the crude extract were essentially identical to that of *Mbo*I digestion. Therefore, these strains were considered to possess an R-M system similar to the *Mbo*I (or *Dpn*II) system. The R-M system of *S. suis* DAT1 was designated *Ssu*DAT1I. The R-M phenotype did not correlate with the capsular serotype, nor did the virulence phenotype of the strains correlate with the R-M phenotype, since the strains were isolated from diseased pigs.

Presence and distribution of the R-M genes in *S. suis* strains. Because the R-M phenotype found in 28 strains of *S. suis* was the same as that of the *Mbo*I system, degenerate primers, designated dam-F and dam-R, were designed on the basis of the nucleotide sequences of the related methylase genes that are found in the database (Table 2). PCR amplification was conducted with genomic DNA of strain DAT1 as a template, and the amplified fragment was cloned into pCR2.1 and sequenced. The nucleotide sequence of the cloned fragment was highly homologous to that of M.*Dpn*II (71.8% identity) (data not shown). The fragment was then labeled with digoxigenin and used as a hybridization probe against genomic DNA from several *S. suis* strains that had been digested with either *Bam*HI, *Eco*RI, or *Hin*dIII and separated by agarose gel electrophoresis. Strains expressing the R-M phenotype provided hybridizing DNA fragments of approximately 6.5, 25, and 10 kb, respectively, whereas no hybridizing fragments appeared in strains which exhibited the null phenotype (data not shown). These results suggested that strains expressing the R-M phenotype possessed a single copy of the gene and that strains with the null phenotype lacked the corresponding gene.

Cloning, sequencing, and genetic organization of the *Ssu*DAT1I gene region. After plaque hybridization of the *S. suis* DAT1 genomic library with the labeled probe described above, 10 positive plaques were identified and the phagemids were rescued to create plasmid subclones as described in Materials and Methods. Five plasmid subclones, designated pDAM3, pDAM4, pDAM7, pDAM9, and pDAM10, were obtained (Fig. 2). Restriction digests of the five subclones showed that pDAM3 and pDAM4 were identical, whereas pDAM7 and pDAM10 contained the same insert in opposite orientations relative to the vector plasmid (Fig. 2). These subcloned fragments were sequenced, and the deduced translation products of several open reading frames (ORFs) were compared to amino acid sequences in databases using the BLAST program. The five subclones contained one or both of the methyltransferase genes, designated *ssuMA* and *ssuMB*. The region upstream of the methyltransferase genes contained four ORFs associated with purine biosynthesis (described in detail below).
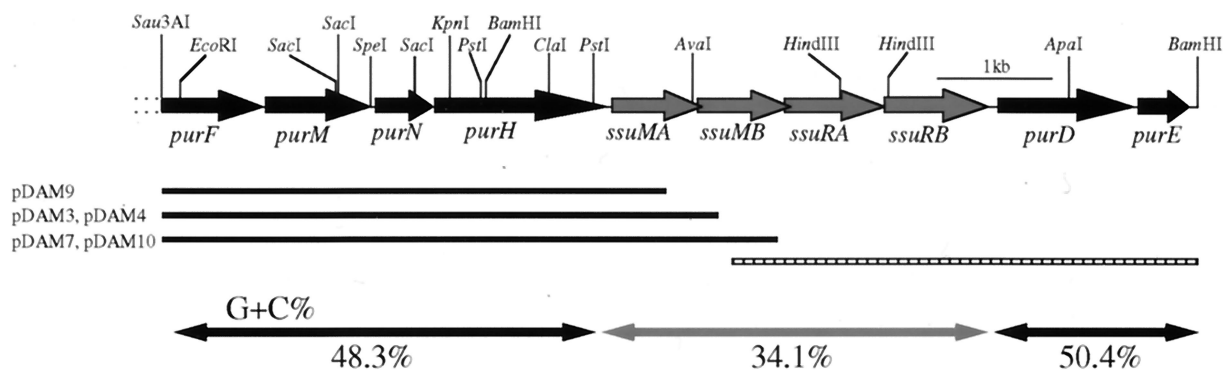
FIG. 2. Physical and genetic map of the 9,618-bp sequence of *S. suis* DAT1 with the putative genes indicated. Shaded and solid arrows, ORFs of the *SsuDAT1I* genes and *pur* genes, respectively. Horizontal solid lines, positions of the cloned fragments of plasmids pDAM3, pDAM4, pDAM7, pDAM9, and pDAM10; striped bar, region amplified by an inverse PCR; shaded and solid double-headed arrows, average G+C content of the *SsuDAT1I* region and that of the *pur* region, respectively.

An inverse PCR was employed to amplify a fragment containing the downstream DNA region using synthetic oligonucleotides dam-in1 and dam-in2. An amplified fragment of approximately 4.5 kb was then directly sequenced. The entire DNA sequence of the *SsuDAT1I* genes and the flanking regions was 9,618 bp, and it included 10 putative ORFs (Fig. 2). These were identified on the basis of the adopted criterion that an ORF consists of at least 40 codons preceded by a potential Shine-Dalgarno sequence at an appropriate distance (6 to 15 bp) from one of the commonly used initiation codons (AUG, UUG, and GUG). The results of the database analysis of the deduced amino acid sequences are summarized in Table 3. As expected, a gene encoding a restriction endonuclease appeared just downstream of *ssuMB*. However, unexpectedly, downstream of this gene there was an additional restriction endonuclease gene. These two genes were designated *ssuRA* and

ssuRB (Fig. 2). Other ORFs found in the sequenced region encoded products that were highly homologous to enzymes involved in purine biosynthetic pathways (Table 3). Genes identified in the entire sequenced region were in the order *purF-purM-purN-purH-ssuMA-ssuMB-ssuRA-ssuRB-purD-purE*, as shown in Fig. 2. All of these genes were encoded on the same DNA strand. No transposable element or long-repeat sequence was found in the 9,618-bp sequence. The average G+C content of the purine biosynthetic genes was 48.9%, which was higher than that of the total genome of *S. suis* (39 to 41%) (18), and the average G+C content of the *SsuDAT1I* region was much lower (34.1%) (Fig. 2). The codon usage pattern for the *SsuDAT1I* system genes was somewhat different from that of the purine biosynthetic gene clusters. For example, the most frequently used codon for leucine in the *SsuDAT1I* system was TTA (45% of the total leucine codons in *SsuDAT1I* genes),

TABLE 3. Predicted gene products and the homology exhibited by their potential protein products to amino acid sequences in the database[a]

Predicted gene	Coding position		Product size		Predicted pI	Homology		Description and function of closest relative[b]
	Start	Stop	aa[c]	kDa		BLAST score	aa identity[d]	
purF	<1	940	>312	>34.6	4.67	e-124	73/300	Phosphoribosylpyrophosphate amidotransferase precursor of *L. lactis* (EC 2.4.2.14) [U64311]
purM	995	2017	340	36.4	4.54	e-128	69/339	Phosphoribosylformylglycineamide cycloligase of *L. lactis* subsp. *cremoris* (EC 6.3.3.1) [AF016634]
purN	2014	2565	183	20.5	4.88	8e-58	71/150	Phosphoribosylglycineamide formyltransferase homolog of *S. pyogenes* (EC 22.1.2.2) [U70775]
purH	2575	4122	515	56.6	4.83	e-169	58/513	Phosphoribosylamidoimidazole carboxyamide formyltransferase of *B. subtilis* (EC 6.3.2.6) [J02732]
ssuMA	4187	5017	276	31.8	4.93	e-114	71/275	M.*Lla*DCHI A of *L. lactis* [U16027]
						e-110	70/274	M.*Dpn*II 1 of *S. pneumoniae* [M14339]
ssuMB	5007	5822	271	31.2	9.50	e-145	90/266	M.*Dpn*A of *S. pneumoniae* [M14339]
						e-139	86/263	M.*Lla*DCHI B of *L. lactis* [U16027]
ssuRA	5800	6726	308	36.0	5.20	3e-78	49/299	R.*Dpn*II of *S. pneumoniae* [M14339]
ssuRB	6726	7631	301	34.4	5.17	e-126	72/300	R.*Lla*DCHI of *L. lactis* [U16027]
purD	7752	9014	420	45.3	4.48	e-125	54/419	Phosphoribosylamineglycine ligase of *L. lactis* (EC 6.3.4.13) [AJ000883]
purE	9040	9528	162	16.8	6.36	4e-41	61/139	Phosphoribosylamidoimidazole carboxylase catalytic subunit of *B. subtilis* (EC 4.1.1.21) [J02732]

[a] Proteins relating to purine biosynthesis have been named according to their closest relative, probable function, and *E. coli* gene designations (60). Coding region positions are numbered with respect to the entire 9,618-bp sequence determined (AB045609) and from the first base of the start codon to the last base of the stop codon. The Genetyx-Mac program (Software Developing Company) was used to make predictions of coding regions, molecular masses, and isoelectric points.
[b] Numbers in blackets are accession numbers in the DDBJ/EMBL/GenBank database.
[c] aa, amino acids.
[d] Percent identity/number of amino acids evaluated.

although TTA was the least-used codon for *pur* genes (8.4% of the total leucine codons in *pur* genes).

Genetic structure and protein analysis of the *Ssu*DAT1I system. The complete DNA sequence and the deduced amino acid sequences of the *Ssu*DAT1I region are shown in Fig. 3. The four ORFs in the *Ssu*DAT1I gene region overlapped each other in a head-to-tail manner (Fig. 2 and 3). Only *ssuRB* had a typical Shine-Dalgarno sequence (Fig. 3). Interestingly, the competence-regulated sequence, the so-called combox, was present at the translation initiation site of the *ssuMB* gene. This sequence was originally found in several competence-regulated genes and was recently shown to be present in the methyltransferase gene of the *Dpn*II system (25). The DNA sequence of the combox, including the first start codon, the second start codon, and a typical Shine-Dalgarno sequence prior to the second start codon, was completely identical to that described for the *Dpn*II system (Fig. 3). A palindromic sequence, typical of an rho-independent terminator, could not be found at the end of the *ssuRB* gene.

Homology searches showed that the deduced protein encoded by *ssuMA* was 71 and 70% identical to M.*Lla*A and M.*Dpn*II, respectively (Table 3). Alignments of the sequences of these three proteins revealed significant similarities among them (Fig. 4A). Thus, it was concluded that *ssuMA* encoded an N^6-methyladenine methyltransferase, which was designated M.*Ssu*MA. The deduced protein encoded by *ssuMB* was 90 and 86% identical to M.*Dpn*A and M.*Lla*B, respectively (Table 3). Alignment of these three proteins highlighted striking similarities among them (Fig. 4B). Thus it was concluded that *ssuMB* encoded a second N^6-methyladenine methyltransferase, which was designated M.*Ssu*MB. The similarities among M.*Ssu*MB, M.*Dpn*A, and M.*Lla*B suggested that M.*Ssu*MB methylates single-stranded DNA, as described for M.*Dpn*A (8).

The deduced amino acid sequence encoded by *ssuRA* was similar to that of R.*Dpn*II, whereas the sequence encoded by *ssuRB* was highly homologous to that of R.*Lla*DCHI (Table 3). Both R.*Dpn*II and R.*Lla*DCHI are endonucleases which specifically recognize sequence 5′- ↓ GATC-3′ and cleave prior to the 5′ guanine as indicated by the arrow. Thus, both *ssuRA* and *ssuRB* encode restriction endonucleases, which we named R.*Ssu*RA and R.*Ssu*RB, respectively. All currently characterized type II endonucleases displaying significant similarity are isoschizomers of each other, indicating that R.*Ssu*RA and R.*Ssu*RB likely recognize 5′-GATC-3′ sites and cleave them as R.*Dpn*II and R.*Lla*DCHI do. Alignment of the amino acid sequences of these four endonucleases revealed similarities at their centers and near the C-terminal ends of the proteins (Fig. 4C). The conservation of these residues suggests that they may constitute the catalytically active sites of the enzymes. R.*Ssu*RA and R.*Ssu*RB have a limited homology (31%) with each other, and they do not have any significant homology with the GATC-specific type II endonucleases of other classes, such as *Sau*3AI and *Dpn*I.

Organization of purine biosynthetic genes. The genetic organization of the region flanking the *Ssu*DAT1I genes revealed that the purine biosynthetic genes were clustered in this bacterium, as has been observed in other gram-positive bacteria such as *Bacillus subtilis* (4). The gene order, excluding the *Ssu*DAT1I genes, is identical to that found in the genome sequence of *Streptococcus pyogenes* (B. A. Roe, S. P. Linn,

L. Son, X. Yuan, S. Clifton, M. McShan, and J. Ferretti, Streptococcal Genome Sequencing Project, University of Oklahoma, Norman [http://www.genome.ou.edu/strep.html]) and similar to the gene order of the *B. subtilis pur* operon (4). All the *pur* genes found in *S. suis* except *purF*, of which the 5′ end was truncated, were preceded by typical Shine-Dalgarno sequences. These *pur* genes were separated by intergenic sequences of 13 to 25 bp, with the exception of *purM* and *purN*, which overlapped (Table 3). However, the intergenic sequences between *purH* and *purD*, which were interrupted by the *Ssu*DAT1I genes, were longer than those found among *pur* genes, i.e., 64 and 120 bp (Table 3).

Expression of methylation and restriction genes in *E. coli*. Plasmid pDAM4C, used for DNA sequencing, was digested with *Pst*I (one *Pst*I site is in the cloned insert and the other is in the multicloning sites of the vector), and the 1.2-kb fragment containing *ssuMA* was subcloned into pUC19 to create pSUMA1. The translational direction of *ssuMA* was opposite in orientation to that of the *lacZ* promoter on the vector. Introduction of pSUMA1 into *E. coli* SCS110 and restriction cleavage analysis of the pSUMA1 recovered from SCS110 using *Dpn*I and *Mbo*I showed that SCS110 expressed the methyltransferase (Fig. 5A).

The DNA fragment from approximately 100 bp upstream of *ssuMA* to approximately 100 bp downstream of *ssuMB* was amplified by PCR using primers purH3′ and ssuMB3′ (Fig. 3). The amplified 2.2-kb fragment was ligated to pCR2.1, and the ligation mixture was used to transform *E. coli* DH5α. Fifteen clones were picked. The translational direction of *ssuMA* and *ssuMB* was opposite to the orientation of the *lacZ* promoter in 14 clones. All 14 clones contained two or more nucleotide substitutions caused by misreading during the PCR amplification, resulting in undesired amino acid changes. Clone pSUMAB14 contained only one nucleotide substitution in the ORF of *ssuMB* (A at nucleotide 667 of *ssuMB* was replaced by G), which converted Ile$_{223}$ to a similar amino acid, Val$_{223}$, and this subclone was used for further study. pSUMAB14 was digested with *Ava*I, filled in with Klenow enzyme, self-ligated, and introduced into *E. coli* DH5α to create a plasmid carrying only *ssuMB* and designated pSUMB1. pSUMAB14 and pSUMB1 were then introduced into *E. coli* SCS110, and the plasmids recovered from SCS110 were digested with either *Dpn*I or *Mbo*I. Both plasmids could be cleaved by *Dpn*I but not by *Mbo*I (Fig. 5A), indicating that the methylation genes were functional in *E. coli*.

Several attempts to clone the two restriction endonuclease genes into *E. coli* were unsuccessful. Thus, *ssuRA* and *ssuRB* were individually amplified by PCR using synthetic primers containing unique restriction cleavage sites, i.e., ssuRA5′Ec plus ssuRA3′Pst and ssuRB5′Ec plus ssuRB3′Pst, respectively. The positions of the primers are indicated in Fig. 3. The unique restriction sites introduced in the primers permitted their cloning into low-copy-number vector pHSG576 in the orientation opposite to that of the *lacZ* promoter. Four and five clones containing *ssuRA* and *ssuRB*, respectively, were obtained and were designated pSURA1~pSURA4 and pSURB1~pSURB5, respectively. These plasmids were then introduced into *E. coli* SCS110 carrying pSUMAB14. Crude extracts of these clones were prepared by sonication and examined for their DNA cleavage activities using unmethylated pUC19 DNA. All showed
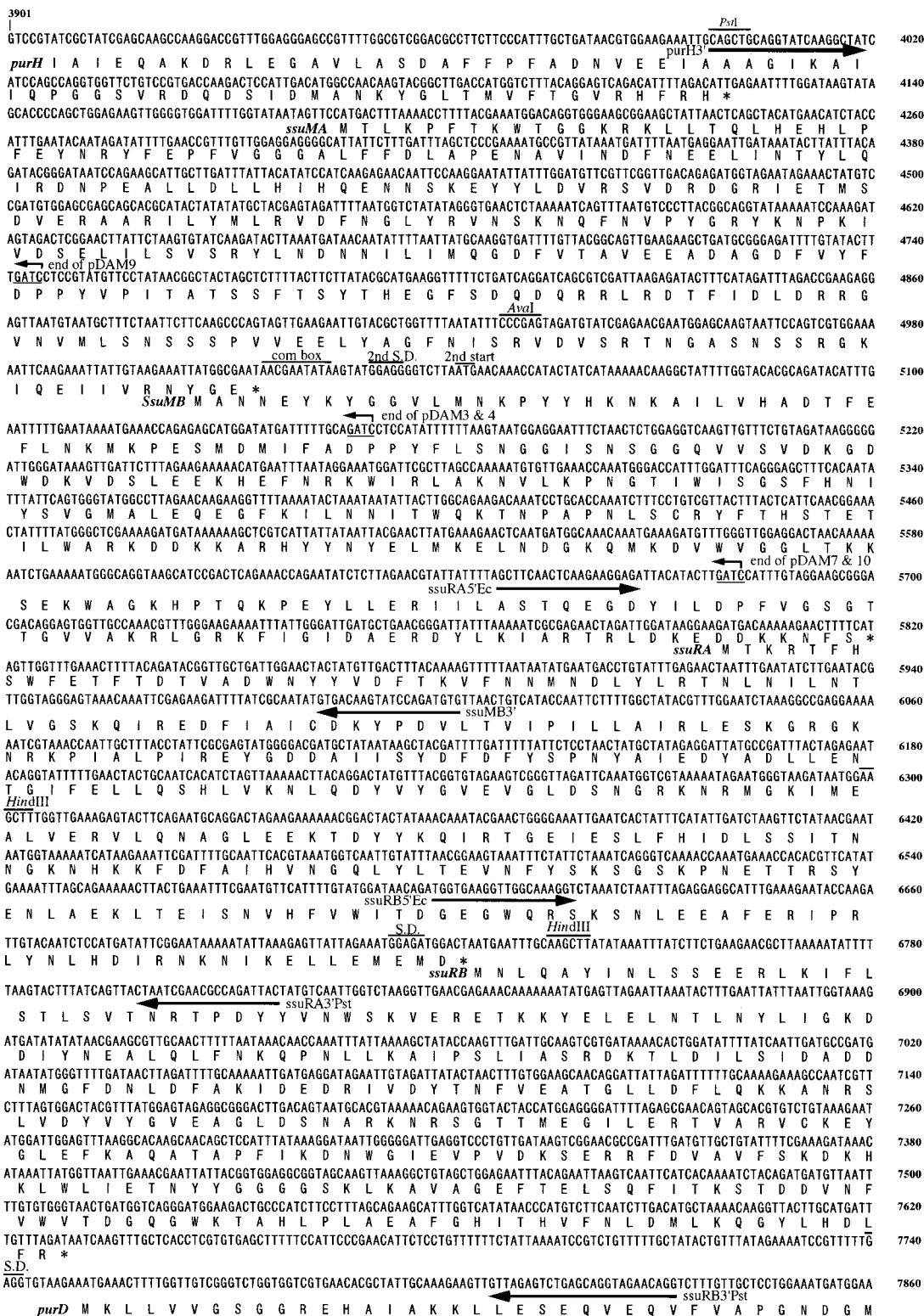
```
3901
|
GTCCGTATCGCTATCGAGCAAGCCAAGGACCGTTTGGAGGGGAGCCGTTTTGGCGTCGGACGCCTTCTTCCCATTTGCTGATAACGTGGAAGAAATTGCAGCTGCAGGTATCAAGGCTATC    4020
purH  I  A  I  E  Q  A  K  D  R  L  E  G  A  V  L  A  S  D  A  F  F  P  F  A  D  N  V  E  E  I  A  A  A  G  I  K  A  I
                                                                                                          purH3'
ATCCAGCCAGGTGGTTCTGTCCGTGACCAAGACTCCATTGACATGGCCAACAAGTACGGCTTGACCATGGTCTTTACAGGAGTCAGACATTTTAGACATTGAGAATTTTGGATAAGTATA    4140
I  Q  P  G  G  S  V  R  D  Q  D  S  I  D  M  A  N  K  Y  G  L  T  M  V  F  T  G  V  R  H  F  R  H  *
GCACCCCAGCTGGAGAAGTTGGGGTGGATTTTGGTATAATAGTTCCATGACTTTAAAACCTTTTACGAAATGGACAGGTGGGAAGCGGAAGCTATTAACTCAGCTACATGAACATCTACC    4260
                                    ssuMA  M  T  L  K  P  F  T  K  W  T  G  G  K  R  K  L  L  T  Q  L  H  E  H  L  P
ATTTGAATACAATAGATATTTTGAACCGTTTGTTGGAGGAGGGGCATTATTCTTTGATTTAGCTCCCGAAAATGCCGTTATAAATGATTTTAATGAGGAATTGATAAATACTTATTTACA    4380
F  E  Y  N  R  Y  F  E  P  F  V  G  G  G  A  L  F  F  D  L  A  P  E  N  A  V  I  N  D  F  N  E  E  L  I  N  T  Y  L  Q
GATACGGGATAATCCAGAAGCATTGCTTGATTTATTACATATCCATCAAGAGAACAATTCCAAGGAATATTATTTGGATGTTCGTTCGGTTGACAGAGATGGTAGAATAGAAACTATGTC    4500
I  R  D  N  P  E  A  L  L  D  L  L  H  I  H  Q  E  N  N  S  K  E  Y  Y  L  D  V  R  S  V  D  R  D  G  R  I  E  T  M  S
CGATGTGGAGCGAGCAGCACGCATACTATATATGCTACGAGTAGATTTTAATGGTCTATATAGGGTGAACTCTAAAAATCAGTTTAATGTCCCTTACGGCAGGTATAAAAATCCAAAGAT    4620
D  V  E  R  A  A  R  I  L  Y  M  L  R  V  D  F  N  G  L  Y  R  V  N  S  K  N  Q  F  N  V  P  Y  G  R  Y  K  N  P  K  I
AGTAGACTCGGAACTTATTCTAAGTGTATCAAGATACTTAAATGATAACAATATTTTAATTATGCAAGGTGATTTTGTTACGGCAGTTGAAGAAGCTGATGCGGGAGATTTTGTATACTT    4740
V  D  S  E  L  I  L  S  V  S  R  Y  L  N  D  N  N  I  L  I  M  Q  G  D  F  V  T  A  V  E  E  A  D  A  G  D  F  V  Y  F
    ← end of pDAM9
TGATCCTCCGTATGTTCCTATAACGGCTACTAGCTCTTTTACTTCTTATACGCATGAAGGTTTTTCTGATCAGGATCAGCGTCGATTAAGAGATACTTTCATAGATTTAGACCGAAGAGG    4860
D  P  P  Y  V  P  I  T  A  T  S  S  F  T  S  Y  T  H  E  G  F  S  D  Q  D  Q  R  R  L  R  D  T  F  I  D  L  D  R  R  G
                                                                          Aval
AGTTAATGTAATGCTTTCTAATTCTTCAAGCCCAGTAGTTGAAGAATTGTACGCTGGTTTTAATATTTCCCGAGTAGATGTATCGAACGAATGGAGCAAGTAATTCCAGTCGTGGAAA    4980
V  N  V  M  L  S  N  S  S  S  P  V  V  E  E  L  Y  A  G  F  N  I  S  R  V  D  V  S  R  T  N  G  A  S  N  S  S  R  G  K
                                    com box      2nd S.D.     2nd start
AATTCAAGAAATTATTGTAAGAAATTATGGCGAATAACGAATATAAGTATGGAAGGGGTCTTAATGAACAAACCATACTATCATAAAAACAAGGCTATTTTGGTACACGCAGATACATTTG    5100
I  Q  E  I  I  V  R  N  Y  G  E  *
                          SsuMB  M  A  N  N  E  Y  K  Y  G  G  V  L  M  N  K  P  Y  Y  H  K  N  K  A  I  L  V  H  A  D  T  F  E
AATTTTTGAATAAAATGAAACCAGAGAGCATGGATATGATTTTTGCAGATCCTCCATATTTTTTAAGTAATGGAGGAATTTCTAACTCTGGAGGTCAAGTTGTTTCTGTAGATAAGGGGG    5220
F  L  N  K  M  K  P  E  S  M  D  M  I  F  A  D  P  P  Y  F  L  S  N  G  G  I  S  N  S  G  G  Q  V  V  S  V  D  K  G  D
                                                      ← end of pDAM3 & 4
ATTGGGATAAAGTTGATTCTTTAGAAGAAAAACATGAATTTAATAGGAAATGGATTCGCTTAGCCAAAATGTGTTGAAACCAAATGGGACCATTTGGATTTCAGGGAGCTTTCACAATA    5340
W  D  K  V  D  S  L  E  E  K  H  E  F  N  R  K  W  I  R  L  A  K  N  V  L  K  P  N  G  T  I  W  I  S  G  S  F  H  N  I
TTTATTCAGTGGGTATGGCCTTAGAACAAGAAGGTTTTAAAAATACTAAATAATATTACTTGGCAGAAGACAAATCCTGCACCAAATCTTTCCTGTCGTTACTTTACTCATTCAACGGAAA    5460
Y  S  V  G  M  A  L  E  Q  E  G  F  K  I  L  N  N  I  T  W  Q  K  T  N  P  A  P  N  L  S  C  R  Y  F  T  H  S  T  E  T
CTATTTTATGGGCTCGAAAAGATGATAAAAAAGCTCGTCATTATTATAATTACGAACTTATGAAAGAACTCAATGATGGCAAACAAATGAAAGATGTTTGGGTTGGAGGACTAACAAAA    5580
I  L  W  A  R  K  D  D  K  K  A  R  H  Y  Y  N  Y  E  L  M  K  E  L  N  D  G  K  Q  M  K  D  V  W  V  G  G  L  T  K  K
                                                                                        ← end of pDAM7 & 10
AATCTGAAAAATGGGCAGGTAAGCATCCGACTCAGAAACCAGAATATCTCTTAGAACGTATTATTTTAGCTTCAACTCAAGAAGGAGATTACATACTTGATCCATTTGTAGGAAGCGGGA    5700
    S  E  K  W  A  G  K  H  P  T  Q  K  P  E  Y  L  L  E  R  I  I  L  A  S  T  Q  E  G  D  Y  I  L  D  P  F  V  G  S  G  T
                                                                                ssuRA5'Ec
CGACAGGAGTGGTTGCCAAACGTTTGGGAAGAAAATTTATTGGGATTGATGCTGAACGGGATTATTTAAAAATCGCGAGAACTAGATTGGATAAGGAAGATGACAAAAAGAACTTTTCAT    5820
T  G  V  V  A  K  R  L  G  R  K  F  I  G  I  D  A  E  R  D  Y  L  K  I  A  R  T  R  L  D  K  E  D  D  K  K  N  F  S  *
                                                                                                      ssuRA  M  T  K  R  T  F  H
AGTTGGTTTGAAACTTTTACAGATACGGTTGCTGATTGGAACTACTATGTTGACTTTACAAAAGTTTTTAATAATATGAATGACCTGTATTTGAGAACTAATTTGAATATCTTGAATACG    5940
S  W  F  E  T  F  T  D  T  V  A  D  W  N  Y  Y  V  D  F  T  K  V  F  N  N  M  N  D  L  Y  L  R  T  N  L  N  I  L  N  T
TTGGTAGGGAGTAAACAAATTCGAGAAGATTTTATCGCAATATGTGACAAGTATCCAGATGTGTTAACGTGTCATACCAATTCTTTTGGCTATACGTTTGGAATCTAAAGGCCGAGGAAAA    6060
L  V  G  S  K  Q  I  R  E  D  F  I  A  I  C  D  K  Y  P  D  V  L  T  V  I  P  I  L  L  A  I  R  L  E  S  K  G  R  G  K
                                        ssuMB3'
AATCGTAAACCAATTGCTTTACCTATTCGCGAGTATGGGGACGATGCTATAATAAGCTACGATTTTGATTTTTATTCTCCTAACTATGCTATAGAGGATTATGCCGATTTACTAGAGAAT    6180
N  R  K  P  I  A  L  P  I  R  E  Y  G  D  D  A  I  I  S  Y  D  F  D  F  Y  S  P  N  Y  A  I  E  D  Y  A  D  L  L  E  N
ACAGGTATTTTTGAACTACTGCAATCACATCTAGTTAAAAACTTACAGGACTATGTTTACGGTGTAGAAGTCGGGTTAGATTCAAATGGTCGTAAAAATAGAATGGGTAAGATAATGGAA    6300
T  G  I  F  E  L  L  Q  S  H  L  V  K  N  L  Q  D  Y  V  Y  G  V  E  V  G  L  D  S  N  G  R  K  N  R  M  G  K  I  M  E
HindIII
GCTTTGGTTGAAAGAGTACTTCAGAATGCAGGACTAGAAGAAAAAACGGACTACTATAAACAAATACGAACTGGGGAAATTGAATCACTATTTCATATTGATCTAAGTTCTATAACGAAT    6420
A  L  V  E  R  V  L  Q  N  A  G  L  E  E  K  T  D  Y  Y  K  Q  I  R  T  G  E  I  E  S  L  F  H  I  D  L  S  S  I  T  N
AATGGTAAAAATCATAAGAAATTCGATTTTGCAATTCACGTAAATGGTCAATTGTATTTAACGGAAGTAAATTTCTATTCTAAATCAGGGTCAAAACCAAATGAAACCACACGTTCATAT    6540
N  G  K  N  H  K  K  F  D  F  A  I  H  V  N  G  Q  L  Y  L  T  E  V  N  F  Y  S  K  S  G  S  K  P  N  E  T  T  R  S  Y
GAAAAATTTAGCAGAAAAACTTACTGAAATTTCGAATGTTCATTTTGTATGGATAACAGATGGTGAAGGTTGGCAAAGGTCTAAATCTAATTTAGAGGAGGCATTTGAAAGAATACCAAGA    6660
E  N  L  A  E  K  L  T  E  I  S  N  V  H  F  V  W  I  T  D  G  E  G  W  Q  R  S  K  S  N  L  E  E  A  F  E  R  I  P  R
                                                                              ssuRB5'Ec
                                                                          S.D.          HindIII
TTGTACAATCTCCATGATATTCGGAATAAAAATATTAAAGAGTTATTAGAAATGGAGATGGACTAATGAATTTGCAAGCTTATATAAATTTATCTTCTGAAGAACGCTTAAAAATATTTT    6780
L  Y  N  L  H  D  I  R  N  K  N  I  K  E  L  L  E  M  E  M  D  *
                                                            ssuRB  M  N  L  Q  A  Y  I  N  L  S  S  E  E  R  L  K  I  F  L
TAAGTACTTTATCAGTTACTAATCGAACGCCAGATTACTATGTCAATTGGTCTAAGGTTGAACGAGAAACAAAAAAATATGAGTTAGAATTAAATACTTTGAATTATTTAATTGGTAAAG    6900
    S  T  L  S  V  T  N  R  T  P  D  Y  Y  V  N  W  S  K  V  E  R  E  T  K  K  Y  E  L  E  L  N  T  L  N  Y  L  I  G  K  D
                                    ssuRA3'Pst
ATGATATATATAACGAAGCGTTGCAACTTTTTAATAAACAACCAAATTTATTAAAAGCTATACCAAGTTTGATTGCAAGTCGTGATAAAACACTGGATATTTTATCAATTGATCGCCGATG    7020
    D  I  Y  N  E  A  L  Q  L  F  N  K  Q  P  N  L  L  K  A  I  P  S  L  I  A  S  R  D  K  T  L  D  I  L  S  I  D  A  D  D
ATAATATGGGTTTTGATAACTTAGATTTTGCAAAAATTGATGAGGATAGAATTGTAGATTATACTAACTTTGTGGAAGCAACAGGATTATTAGATTTTTTGCAAAAGAAAGCCAATCGTT    7140
N  M  G  F  D  N  L  D  F  A  K  I  D  E  D  R  I  V  D  Y  T  N  F  V  E  A  T  G  L  L  D  F  L  Q  K  K  A  N  R  S
CTTTAGTGGACTACGTTTATGGAGTAGAGGCGGGACTTGACAGTAATGCACGTAAAAACAGAAGTGGTACTACCATGGAGGGGATTTTAGAGCGAACAGTAGCCGTGTCTGTAAAGAAT    7260
L  V  D  Y  V  Y  G  V  E  A  G  L  D  S  N  A  R  K  N  R  S  G  T  T  M  E  G  I  L  E  R  T  V  A  R  V  C  K  E  Y
ATGGATTGGAGTTTAAGGCACAAGCAACAGCTCCATTTATAAAGGATAATTGGGGGATTGAGGTCCCTGTTGATAAGTCGGAACGCCGATTTGATGTTGCTGTATTTTCGAAAGATAAAC    7380
G  L  E  F  K  A  Q  A  T  A  P  F  I  K  D  N  W  G  I  E  V  P  V  D  K  S  E  R  R  F  D  V  A  V  F  S  K  D  K  H
ATAAATTATGGTTAATTGAAACGAATTATTACGGTGGAGGCGGTAGCAAGTTAAAGGCTGTAGCTGGAGAATTTACAGAATTAAGTCAATTCATCACAAAATCTACAGATGATGTTAATT    7500
K  L  W  L  I  E  T  N  Y  Y  G  G  G  G  S  K  L  K  A  V  A  G  E  F  T  E  L  S  Q  F  I  T  K  S  T  D  D  V  N  F
TTGTGTGGGTAACTGATGGTCAGGGATGGAAGACTGCCCATCTTCCTTTAGCAGAAGCATTTGGTCATATAACCCATGTCTTCAATCTTGACATGCTAAAACAAGGTTACTTGCATGATT    7620
V  W  V  T  D  G  Q  G  W  K  T  A  H  L  P  L  A  E  A  F  G  H  I  T  H  V  F  N  L  D  M  L  K  Q  G  Y  L  H  D  L
TGTTTAGATAATCAAGTTTGCTCACCTCGTGTGAGCTTTTTCCATTCCCGAACATTCTCCTGTTTTTTCTATTAAAATCCGTCTGTTTTTGCTATACTGTTTATAGAAAATCCGTTTTTG    7740
F  R  *
S.D.
AGGGTGTAAGAAATGAAACTTTTGGTTGTCGGGGTCTGGTGGTCGTGAACACGCTATTGCAAAGAAGTTGTTAGAGTCTGAGCAGGTAGAACAGGTCTTTGTTGCTCCTGGAAATGATGGAA    7860
                                                                          ssuRB3'Pst
purD  M  K  L  L  V  V  G  S  G  G  R  E  H  A  I  A  K  K  L  L  E  S  E  Q  V  E  Q  V  F  V  A  P  G  N  D  G  M
```

FIG. 3. Coding sequence of the 3,960-bp region comprising a part of the *purH* gene, the complete *Ssu*DAT1I genes, and a part of the *purD* gene. The deduced amino acid is indicated underneath the first nucleotide of each codon in the coding sequence. Arrows underneath the nucleotide sequence, positions of the primers used for the amplification and cloning of the restriction and modification genes; bent arrows, terminal ends of the cloned fragments of plasmids pDAM3, pDAM4, pDAM7, pDAM9, and pDAM10; lines above the sequence, sequences resembling a combox, the second Shine-Dalgarno sequence (S.D.), and the second start codon in the *ssuMB* gene. The coding sequence shown corresponds to nucleotides 3901 to 7860 of the sequence with accession no. AB045609.
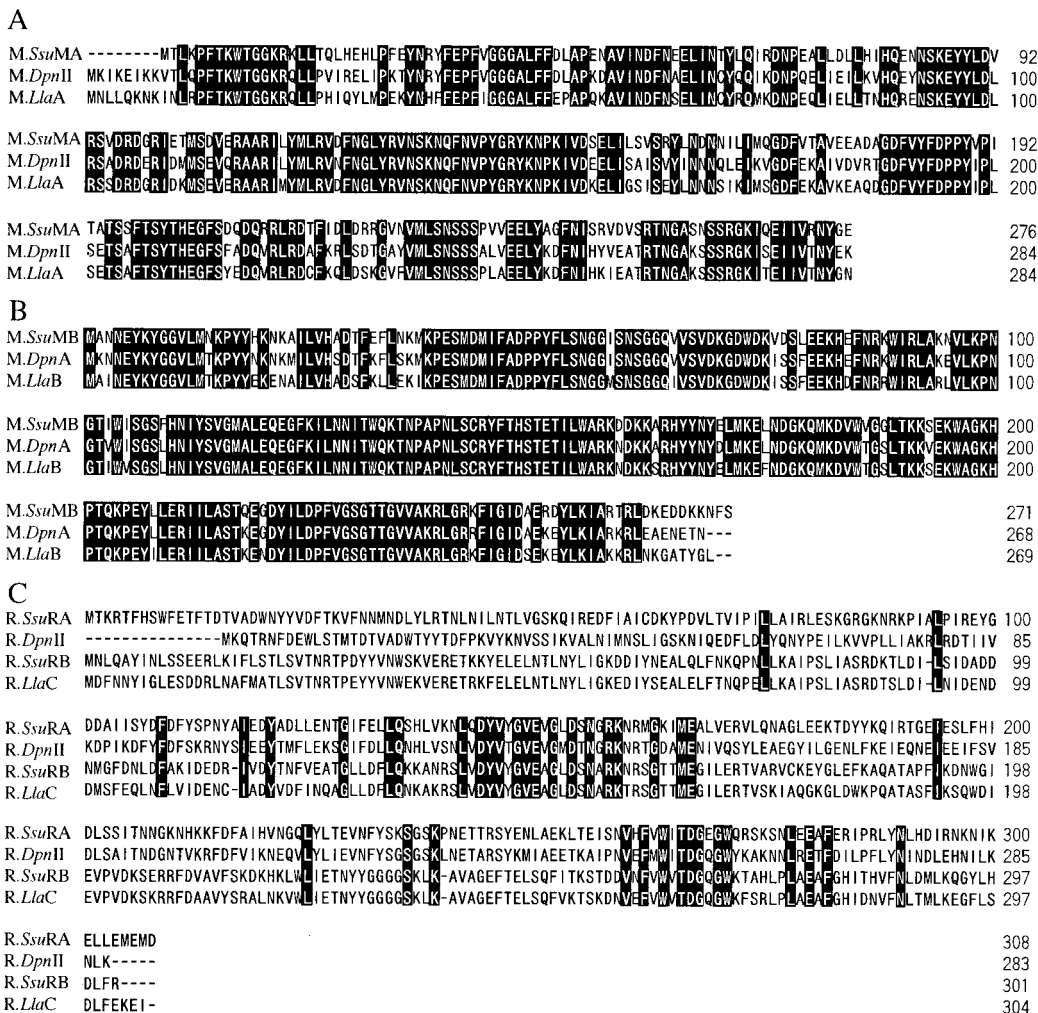
A

```
M.SsuMA   --------MTLKPFTKWTGGKRKLLTQLHEHLPFEYNRYFEPFVGGGALFFDLAPENAVINDFNEELINTYLQIRDNPEALLDLLHIHQENNSKEYYLDV  92
M.DpnII   MKIKEIKKVTLCPFTKWTGGKRQLLPVIRELIPKTYNRYFEPFVGGGALFFDLAPXDAVINDFNAELINQYCQIKDNPQELIEILKVHQEYNSKEYYLDL 100
M.LlaA    MNLLQKNKINLRPFTKWTGGKRQLLPHIQYLMPEKYNHFFEPFIGGGALFFEPAPQKAVINDFNSELINQYRQMKDNPEQLIELLTNHQRENSKEYYLDL 100

M.SsuMA   RSVDRDGRIETMSDVERAARILYMLRVDFNGLYRVNSKNQFNVPYGRYKNPKIVDSELILSVSRYLNDNNILIMQGDFVTAVEEADAGDFVYFDPPYVPI 192
M.DpnII   RSADRDERIDMMSEVCRAARILYMLRVNFNGLYRVNSKNQFNVPYGRYKNPKIVDEELISAISVYINNNQLEIKVGDFEKAIVDVRTGDFVYFDPPYIPL 200
M.LlaA    RSSDRDGRIDKMSEVERAARIMYMLRVDFNGLYRVNSKNQFNVPYGRYKNPKIVDKELIGSISEYLNNSIKIMSGDFEKAVKEAQDGDFVYFDPPYIPL 200

M.SsuMA   TATSSFTSYTHEGFSDQDQRRLRDTFILDLRRFGVNVMLSNSSSPVVEELYAGFNISRVDVSRTNGASNSSRGKIQEIIVRNYGE              276
M.DpnII   SETSAFTSYTHEGFSFADQVRLRDAFXRLSDTGAYVMLSNSSSALVEELYKDFNIHYVEATRTNGAKSSSRGKISEIIVTNYEK              284
M.LlaA    SETSAFTSYTHEGFSYEDQVRLRDCFKQLDSKGVFVMLSNSSSPLAEELYKDFNIHKIEATRTNGAKSSSRGKITEIIVTNYGN              284
```

B

```
M.SsuMB   MANNEYKYGGVLMNKPYYHKNKAILVHADTFEFLNKMKPESMDMIFADPPYFLSNGGISNSGGQVVSVDKGDWDKVDSLEEKHEFNRKWIRLAKNVLKPN 100
M.DpnA    MKNNEYKYGGVLMTKPYYNKNKMILVHSDTFKFLSKMKPESMDMIFADPPYFLSNGGISNSGGQVVSVDKGDWDKISSFEEKHEFNRKWIRLAKEVLKPN 100
M.LlaB    MAINEYKYGGVLMTKPYYEKENALLVHADSFKLLEKIKPESMDMIFADPPYFLSNGGISNSGGQIVSVDKGDWDKISSFEEKHDFNRRWIRLARLVLKPN 100

M.SsuMB   GTIWISGSFHNIYSVGMALEQEGFKILNNITWQKTNPAPNLSCRYFTHSTETIILWARKDDKKARHYYNYELMKELNDGKQMKDVWVGGLTKKSEKWAGKH 200
M.DpnA    GTIWISGSLHNIYSVGMALEQEGFKILNNITWQKTNPAPNLSCRYFTHSTETIILWARKNDKKARHYYNYCLMKELNDGKQMKDVWTGSLTKKVEKWAGKH 200
M.LlaB    GTIWWSGSLHNIYSVGMALEQEGFKILNNITWQKTNPAPNLSCRYFTHSTETIILWARKNDKKSRHYYNYELMKEFNDGKQMKDVWTGSLTKKSEKWAGKH 200

M.SsuMB   PTQKPEYLLERIILASTQEGDYILDPFVGSGTTGVVAKRLGRKFIGIDAERDYLKIARTRLDKEDDKKNFS                             271
M.DpnA    PTQKPEYLLERIILASTKEGDYILDPFVGSGTTGVVAKRLGRRFIGIDAEKEYLKIARKRLEAENETN---                            268
M.LlaB    PTQKPEYILERIILASTKEVDYILDPFVGSGTTGVVAKRLGRKFIGIDSEKEYLKIAKKRLNKGATYGL--                             269
```

C

```
R.SsuRA   MTKRTFHSWFETFTDTVADWNYYVDFTKVFNNMNDLYLRTNLNILNTLVGSKQIREDFIAICDKYPDVLTVIPILLAIRLESKGRGKNRKPIALPIREYG 100
R.DpnII   ---------------MKQTRNFDEWLSTMTDTVADWTYYTDFPKVYKNVSSIKVALNIMNSLIGSKNIQEDFLDLYQNYPEILKVVPLLIAKRLRDTIIV  85
R.SsuRB   MNLQAYINLSSEERLKIFLSTLSVTNRTPDYYVNWSKVERETKKYELELNTLNYLIGKDDIYNEALQLFNKQPNLLKAIPSLIASRDKTLDI-LSIDADD  99
R.LlaC    MDFNNYIGLESSDDRLNAFMATLSVTNRTPEYYVNWEKVERETRKFELELNTLNYLIGKEDIYSEALELFTNQPELLKAIPSLIASRDTSLDI-LNIDEND 99

R.SsuRA   DDAIISYDFDFYSPNYAIEDYADLLENTGIFELLQSHLVKNLQDYVYGVEVGLDSNGRKNRMGKIMEALVERVLQNAGLEEKTDYYKQIRTGEIESLFHI 200
R.DpnII   KDPIKDFYFDFSKRNYSIEEYTMFLEKSGIFDLLQNHLVSNLVDYVTGVEVGMDTNGRKNRTGDAMENIVQSYLEAEGYILGENLFKEIEQNEIEEIFSV 185
R.SsuRB   NMGFDNLDFAKIDEDR-IVDYTNFVEATGLLDFLQKKANRSLVDYVYGVEAGLDSNARKNRSGTTMEGILERTVARVCKEYGLEFKAQATAPFIKDNWGI 198
R.LlaC    DMSFEQLNFLVIDENC-IADYVDFINQAGLLDFLQNKAKRSLVDYVYGVEAGLDSNARKTRSGTTMEGILERTVSKIAQGKGLDWKPQATASFIKSQWDI 198

R.SsuRA   DLSSITNNGKNHKKFDFAIHVNGQLYLTEVNFYSKSGSKPNETTRSYENLAEKLTEISNVLFVWITDGEGWQRSKSNLEEAFERIPRLYNLHDIRNKNIK 300
R.DpnII   DLSAITNDGNTVKRFDFVIKNEQVLYLIEVNFYSGSGSKLNETARSYKMIAEETKAIPNVEFMWITDGQGWYKAKNNLRETFDILPFLYNINDLEHNILK 285
R.SsuRB   EVPVDKSERRFDVAVFSKDKHKLWLIETNYYGGGGSKLK-AVAGEFTELSQFITKSTDDVLFVWVTDGQGWXTAHLPLAEAFGHITHVFNLDMLKQGYLH 297
R.LlaC    EVPVDKSKRRFDAAVYSRALNKVWLIETNYYGGGGSKLK-AVAGEFTELSQFVKTSKDNVEFVWMTDGQGWKFSRLPLAEAFGHIDNVFNLTMLKEGFLS 297

R.SsuRA   ELLEMEMD                                                                                             308
R.DpnII   NLK-----                                                                                             283
R.SsuRB   DLFR----                                                                                             301
R.LlaC    DLFEKEI-                                                                                             304
```

FIG. 4. Sequence alignment of M.*Ssu*MA, M.*Dpn*II, and M.*Lla*A (A); M.*Ssu*MB, M.*Dpn*A, and M.*Lla*B (B); and R.*Ssu*RA, R.*Dpn*II, R.*Ssu*RB, and R.*Lla*C (*Lla*DCHI) (C). Black background, amino acids identical among all the sequences aligned; dashes, gaps in the aligned sequences. DDBJ/EMBL/GenBank accession numbers of the sequences are given in Table 3.

restriction endonuclease activity indistinguishable from that of *Mbo*I (Fig. 5B).

Comparison of the genetic region encoding the *Ssu*DAT1I R-M system with the corresponding region of strains lacking the R-M system. To confirm whether the genetic region encoding the *Ssu*DAT1I system was conserved among other *S. suis* strains and to compare the corresponding regions of strains carrying the R-M system and strains with the null phenotype, we performed PCR using primers purH3′ and ssuRB3′ to amplify the entire R-M region with a part of the flanking regions. When genomic DNA isolated from strains carrying the R-M system was used, a 3.8-kb fragment was amplified from all the strains tested (data not shown). On the other hand, when genomic DNA from the strains with the null phenotype was used, a 323-bp fragment was amplified from all the strains tested (data not shown). These results indicated that the genetic organization of the R-M system was conserved among the strains tested and that the entire R-M system was missing in the strains with the null phenotype. Direct sequencing of the amplified fragments from four strains having the R-M genes

and comparison among them showed that only a few nucleotide substitutions were present in this region (Fig. 6), indicating that the R-M region was conserved among the strains. Direct sequence determination of the 323-bp fragments amplified from five strains with the null phenotype and comparison with those of the R-M regions revealed the differences in DNA sequence between them. The 323-bp DNA sequences obtained from the five strains with the null phenotype were completely conserved. The unique DNA sequence of the R-M region comprising 3,503 bp, which extended from 53 bp upstream of the *ssuMA* gene to 5 bp downstream of the *ssuRB* gene, was inserted in an intergenic sequence between *purH* and *purD*. The insertion target site, 5′-GAT ↓ T(T/G)-3′, as indicated by the arrow, was highly conserved among the strains tested (Fig. 6). The sequence where the R-M genes were inserted was not the recognition site of *Ssu*DAT1I. Comparison of the 323-bp DNA regions of both groups revealed several mismatches present in upstream and downstream sequences (Fig. 6). No other notable substitutions or insertions could be found in these regions. The only notable feature found in the junction
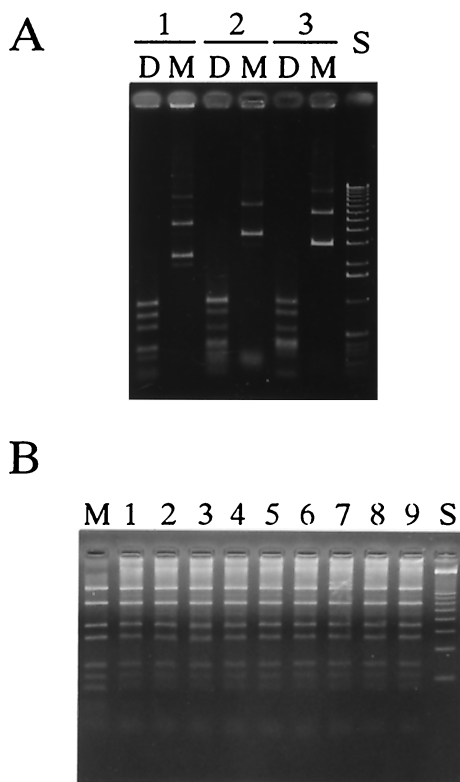
A

B

FIG. 5. (A) Methylation of plasmid DNAs containing the cloned methylation gene(s) of the *Ssu*DAT1I system. The plasmids recovered from *E. coli* SCS110 were digested with restriction endonucleases and analyzed by agarose gel electrophoresis. Lane 1, pSUMA1; lane 2, pSUMAB14; lane 3, pSUMB1. D, digested with *Dpn*I; M, digested with *Mbo*I; S, 1-kb ladder size standards (GIBCO/BRL). (B) DNA cleavage activities of crude extracts prepared from *E. coli* strains carrying the following plasmids using unmethylated pUC19 as the substrate: lane 1, pSURA1; lane 2, pSURA2; lane 3, pSURA3; lane 4, pSURA4; lane 5, pSURB1; lane 6, pSURB2; lane 7, pSURB3; lane 8, pSURB4; lane 9, pSURB5. M, pUC19 digested with *Mbo*I; S, 100-bp ladder size standards (GIBCO/BRL).
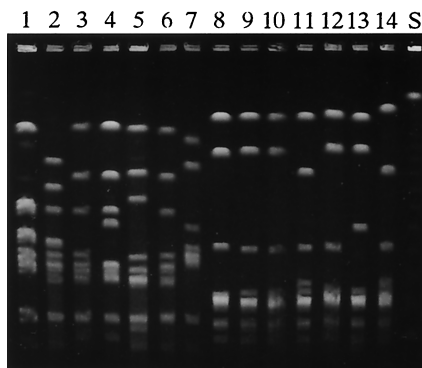
FIG. 7. PFGE of the *S. suis* genomic DNA digested with *Sma*I. Electrophoresis was carried out with a CHEF system for 20 h at 200 V and 10°C with pulse time ramping from 5 to 50 s in a 1% agarose gel. Strains carrying the R-M system were as follows: DAT1 (lane 1), 205 (lane 2), 220 (lane 3), 227 (lane 4), 243 (lane 5), 222 (lane 6), and 236 (lane 7). Strains with the null phenotype were as follows: NCTC10234 (lane 8), 204 (lane 9), 207 (lane 10), 209 (lane 11), 210 (lane 12), 213 (lane 13), and 246 (lane 14). S, λ DNA concatemers as size standards.

region was that a 3-bp sequence at the 5′ end of the 3,503-bp sequence, AAG (Fig. 6, left end) was also present at the 3′ end of the sequence (Fig. 6, right end). The DNA sequences of the region flanking the R-M system showed no sequence homology to published sequences flanking the *Dpn*II and *Lla*DCHI systems (24, 30).

PFGE analysis. Genomic DNAs of *S. suis* strains were digested with *Sma*I and separated by PFGE. Representative cleavage patterns of several strains are shown in Fig. 7. The PFGE patterns of the strains having the R-M system were distinguishable from those obtained from the strains with the null phenotype. The PFGE patterns obtained from the strains with the null phenotype were similar to each other, and some of them were indistinguishable, suggesting that different isolates of the same strain were included. However, striking variations were seen among the PFGE patterns of the strains
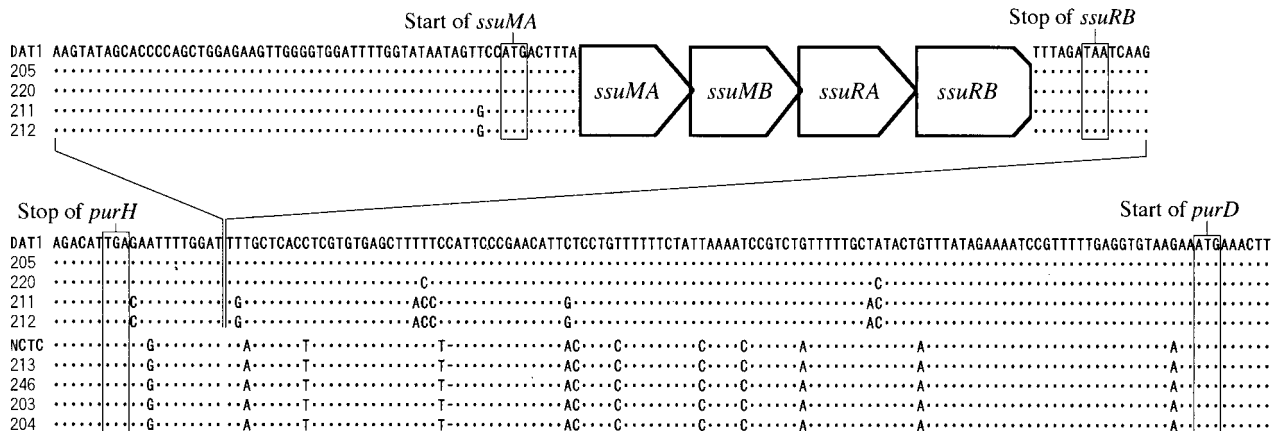
FIG. 6. Alignment of the sequences from the 3′ end of *purH* to the 5′ end of *purD*. Only the sequences that flank the *Ssu*DAT1I genes are shown. The 3,503-bp sequences from 53 bp upstream of *ssuMA* to 5 bp downstream of *ssuRB* (top) are connected by the lines with the lower sequences flanking the *Ssu*DAT1I region, which are aligned with the corresponding sequences of the strains lacking the R-M system. Strains are indicated to the left of the sequences (NCTC, NCTC10234). Dots, nucleotides identical to those of the aligned sequence of DAT1; dashes, gaps in the aligned sequences; boxes, stop and start codons of the genes appearing in these sequences.

having the R-M system, demonstrating that the strains used were independent.

DISCUSSION

We have demonstrated here the presence of an R-M system in *S. suis*. Cloning and DNA sequencing of the *Ssu*DAT1I gene region revealed that the genetic organization was unique for all type II R-M systems described to date, i.e., the region contained two restriction endonuclease genes and two methyltransferase genes. Other R-M systems comprising two methyltransferase genes and one restriction endonuclease gene are known and are isoschizomers of *Ssu*DAT1I, i.e., *Dpn*II, *Lla*DCHI, and *Mbo*I. In *Dpn*II and *Lla*DCHI, the order of the genes is similar to that of the *Ssu*DAT1I system. The gene encoding a conventional methyltransferase for double-stranded DNA is upstream of the gene encoding a second methyltransferase for single-stranded DNA. The restriction endonuclease gene is downstream and the genes overlap, as seen in the *Ssu*DAT1I genes (8, 24, 30). In *Mbo*I, the order of the genes is different. The gene encoding R.*Mbo*I is flanked by two methyltransferase genes, which are separated by 2-bp intergenic spaces (52). These differences suggest that the genetic rearrangement among the R-M genes could have occurred during the evolution of these R-M genes. R.*Ssu*RA was similar to R.*Dpn*II, and R.*Ssu*RB showed extensive homology with R.*Lla*DCHI, whereas R.*Ssu*RA (or R.*Dpn*II) and R.*Ssu*RB (or R.*Lla*DCHI) showed limited homology. The results suggest that an extra restriction endonuclease gene was acquired from some other bacterium and that this resulted in a composite genetic structure, i.e., the *Ssu*DAT1I system. Alternatively, the *Ssu*DAT1I system could be the prototype and a deletion of one restriction endonuclease gene resulted in the genetic structure represented by *Dpn*II and *Lla*DCHI.

The amino acid sequences of the methyltransferases share many conserved structural motifs, and, on the basis of the presence of these motifs in an unknown protein, one can predict its function (6, 56). In this study, striking similarities among the methyltransferases indicated that M.*Ssu*MA and M.*Ssu*MB are highly related to the methyltransferases of the *Dpn*II and *Lla*DCHI systems. On the other hand, different restriction endonucleases usually do not have extensive homology at the level of the amino acid sequence, even when they recognize the same DNA sequence (6, 56). Nevertheless, some exceptions are known; for example, the isospecific pairs *Eco*RI/*Rsr*I (1), *Pst*I/*Bsu*BI (57), *Xma*I/*Crf*9I (56), *Ngo*I/*Hae*II (47), and *Ban*I/*Hgi*CI (10) show some degree of homology. While R.*Dpn*II and R.*Lla*DCHI have limited homology, they were found experimentally to have the same recognition sequence (30). R.*Ssu*RA and R.*Ssu*RB had significant similarities to one of these restriction endonucleases. This indicates that the specificity and other enzymatic properties are the same, although differences in enzymatic properties between *Dpn*II and *Lla*DCHI have not been studied. To date, one example of an R-M system containing two restriction endonuclease genes has been reported. In that case, a gene encoding R.*Eco*47I, which recognizes 5′-GGWCC-3′ (W is A or T), was located in the vicinity of another R-M system, *Eco*47II, which recognizes similar but less specific sequence 5′-GGNCC-3′, where N is any nucleotide (46). In the *Ssu*DAT1I system, R.*Ssu*RA

and R.*Ssu*RB, whose genes were individually cloned and expressed in *E. coli*, could digest unmethylated pUC19 in the same way as *Mbo*I. Therefore, their sequence recognition specificities appeared identical.

We could not clone the *Ssu*DAT1I genes directly in *E. coli* from the phage library or from a PCR-derived fragment. However, the isoschizomeric *Lla*DCHI and *Mbo*I R-M genes have been cloned in *E. coli* on multicopy vectors (30, 52). *E. coli* possesses orphan methylase Dam, which has the same sequence specificity as the *Ssu*DAT1I system and which protects the host DNA from attack by the corresponding restriction endonucleases. Recently, the Dam methylase was found to be involved in the regulation of gene expression, and hence not all the recognition sequences in the host chromosomal DNA were methylated (14). If the enzymatic activity of *Ssu*DAT1I is stronger than those of the other isoschizomers, the cloning of the genes will inhibit the growth of the *E. coli* host cells. Indeed, the dual genes encoding the restriction endonucleases in the *Ssu*DAT1I system might be expected to express twice as much of the restriction endonuclease activity due in essence to a gene dosage effect (53).

The *Ssu*DAT1I system was not found in several *S. suis* strains, including NCTC10234. Moreover, the identity of the DNA sequences flanking this system, in conjunction with the differences in G+C content and codon usage between *Ssu*DAT1I and the flanking regions, supports the notion that this system did not arise in this organism. The *Ssu*DAT1I genes were not located on transferable elements such as plasmids and bacteriophages. Comparison of the corresponding DNA regions of strains with and without the R-M system revealed the precise location of the R-M genes. The *Ssu*DAT1I genes have been inserted into a 125-bp intergenic region separating *purH* and *purD*. This may ensure constitutive expression of the R-M genes. In most bacteria, purine biosynthetic genes are part of an essential operon. The overall genetic organization of the *pur* operon, except for the *Ssu*DAT1I genes, was identical to that found in *S. pyogenes* (Streptococcal Genome Sequencing Project, University of Oklahoma, Norman [http://www.genome.ou.edu/strep.html]) and was similar to that reported in *B. subtilis* (4). Because the *pur* regions of NCTC10234 and other *S. suis* strains lack the R-M genes and because the sequences flanking the R-M sequences in *S. suis* strains carrying the system are virtually identical, it appears that the latter *S. suis* strains have only recently acquired the sequence.

A natural competence for genetic transformation has not yet been demonstrated in *S. suis*. However, the combox (25), a competence-regulated sequence, was found in the initiation region of the *ssuMB* gene. This might support a hypothesis that the *Ssu*DAT1I system is an exogenous element and that *S. suis* acquired the *Ssu*DAT1I genes via a transformation event. However, the competence of some bacteria is coordinately expressed in response to the growth phase and culture conditions (2, 27). Further studies may be needed to determine whether *S. suis* has the potential to be competent for genetic transformation.

Recently, studies of a novel R-M system, *Hpy*188I, found in *Helicobacter pylori* strain J188 and comparison of its DNA sequence with those of other strains revealed that the R-M system has been horizontally transferred from other bacteria (58). The *Hpy*188I genes are flanked by 92-bp long direct re-

peats, suggesting that a transposition event involving the R-M system had taken place. In the region flanking SsuDAT1I, we could not find any long-repeat sequences or any notable nucleotide substitutions. The genetic structure was conserved among the S. suis strains tested. However, from the results of the PFGE analysis, we could not rule out the possibility that the strains harboring the R-M system were clonal. The structure of the genetic region flanking the SsuDAT1I system and stable maintenance of this genetic structure in the different strains of S. suis lead us to propose another hypothesis to explain the genetic conversion, i.e., illegitimate recombination. It is plausible that the SsuDAT1I genes were initially transferred on a transposon, followed by excision of the transposon, since no functions associated with DNA mobility, such as transposase, integrase, and invertase functions, were found in the vicinity of the SsuDAT1I genes. However, this hypothesis does not explain the facts that the conserved genetic structure of the R-M system was found in different strains and that no transposable element could be found in any of the strains tested. This indicated that excision of a transposable element did not occur following passage through the strains tested. It is also possible that the SsuDAT1I system was acquired by an ancestor of the S. suis DAT1 strain via illegitimate recombination from distant sources and that the R-M genes have been transferred among strains of S. suis en bloc along with the conserved flanking pur genes, perhaps by transformation, and incorporated into the recipient chromosome by homologous recombination.

REFERENCES

1. **Aiken, C., and R. I. Gumport.** 1988. Restriction endonuclease RsrI from Rhodobacter sphaeroides, an isoschizomer of EcoRI: purification and properties. Nucleic Acids Res. **16:**7901–7916.
2. **Alloing, G., B. Martin, C. Granadel, and J.-P. Claverys.** 1998. Development of competence in Streptococcus pneumoniae: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. Mol. Microbiol. **29:**75–83.
3. **Alm, R. A., L.-S. L. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust.** 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. Nature **397:**176–180.
4. **Anagnostopoulos, C., P. J. Piggot, and J. A. Hoch.** 1993. The genetic map of Bacillus subtilis, p. 425–461. In A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), Bacillus subtilis and other gram-positive bacteria. American Society for Microbiology, Washington, D.C.
5. **Anton, B. P., D. F. Heiter, J. S. Benner, E. J. Hess, L. Greenough, L. S. Moran, B. E. Slatko, and J. E. Brooks.** 1997. Cloning and characterization of the BglII restriction-modification system reveals a possible evolutionary footprint. Gene **187:**19–27.
6. **Bickle, T. A., and D. H. Kruger.** 1993. Biology of DNA restriction. Microbiol. Rev. **57:**434–450.
7. **Brassard, S., H. Paquet, and P. H. Roy.** 1995. A transposon-like sequence adjacent to the AccI restriction-modification operon. Gene **157:**69–72.
8. **Cerritelli, S., S. S. Springhorn, and S. A. Lacks.** 1989. DpnA, a methylase for single-strand DNA in the DpnII restriction system, and its biological function. Proc. Natl. Acad. Sci. USA **86:**9223–9227.
9. **Clifton-Hadley, F. A.** 1983. Streptococcus suis type 2 infections. Br. Vet. J. **139:**1–5.
10. **Erdmann, D., A. Dusterhoft, and M. Kroger.** 1991. Cloning and molecular characterization of the HgiCI restriction/modification system from Herpetosiphon giganteus Hpg9 reveals high similarity to BanI. Eur. J. Biochem. **202:**1247–1256.
11. **Gelinas, R. E., P. A. Myers, and R. J. Roberts.** 1977. Two sequence-specific endonucleases from Moraxella bovis. J. Mol. Biol. **114:**169–179.
12. **Gottschalk, M., R. Higgins, M. Jacques, M. Beaudoin, and J. Henrichsen.** 1991. Characterization of six new capsular types (23 through 28) of Streptococcus suis. J. Clin. Microbiol. **29:**2590–2594.
13. **Gottschalk, M., R. Higgins, M. Jacques, K. R. Mittal, and J. Henrichsen.** 1989. Description of 14 new capsular types of Streptococcus suis. J. Clin. Microbiol. **27:**2633–2636.
14. **Heithoff, D. M., R. L. Sinsheimer, D. A. Low, and M. J. Mahan.** 1999. An essential role for DNA adenine methylation in bacterial virulence. Science **284:**967–970.
15. **Higgins, R., and M. Gottschalk.** 1996. Distribution of Streptococcus suis capsular types in 1995. Can. Vet. J. **37:**242.
16. **Higgins, R., M. Gottschalk, M. Beaudoin, and S. A. Rawluk.** 1992. Distribution of Streptococcus suis capsular types in Quebec and western Canada. Can. Vet. J. **33:**27–30.
17. **Jeltsch, A., and A. Pingoud.** 1996. Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. J. Mol. Evol. **42:**91–96.
18. **Kilpper-Balz, R., and K. H. Schleifer.** 1987. Streptococcus suis sp. nov., nom. rev. Int. J. Sys. Bacteriol. **37:**160–162.
19. **Kita, K., J. Tsuda, T. Kato, K. Okamoto, H. Yanase, and M. Tanaka.** 1999. Evidence of horizontal transfer of the EcoO1091 restriction-modification gene to Escherichia coli chromosomal DNA. J. Bacteriol. **181:**6822–6827.
20. **Kobayashi, I., A. Nobusato, N. Kobayashi-Takahashi, and I. Uchiyama.** 1999. Shaping the genome—restriction-modification systems as mobile genetic elements. Curr. Opin. Genet. Dev. **9:**649–656.
21. **Kusano, K., K. Sakagami, T. Yokochi, T. Naito, Y. Tokunaga, E. Ueda, and I. Kobayashi.** 1997. A new type of illegitimate recombination is dependent on restriction and homologous interaction. J. Bacteriol. **179:**5380–5390.
22. **Lacks, S., and B. Greenberg.** 1975. A deoxyribonuclease of Diplococcus pneumoniae specific for methylated DNA. J. Biol. Chem. **250:**4060–4066.
23. **Lacks, S., and B. Greenberg.** 1977. Complementary specificity of restriction endonucleases of Diplococcus pneumoniae with respect to DNA methylation. J. Mol. Biol. **114:**153–168.
24. **Lacks, S. A., B. M. Mannarelli, S. S. Springhorn, and B. Greenberg.** 1986. Genetic basis of the complementary DpnI and DpnII restriction systems of S. pneumoniae: an intercellular cassette mechanism. Cell **46:**993–1000.
25. **Lacks, S. A., S. Ayalew, A. G. de la Campa, and B. Greenberg.** 2000. Regulation of competence for genetic transformation in Streptococcus pneumoniae: expression of dpnA, a late competence gene encoding a DNA methyltransferase of the DpnII restriction system. Mol. Microbiol. **35:**1089–1098.
26. **Lee, K. F., P. C. Shaw, S. J. Picone, G. G. Wilson, and K. D. Lunnen.** 1998. Sequence comparison of the EcoHK31I and EaeI restriction-modification systems suggests an intergenic transfer of genetic material. Biol. Chem. **379:**437–441.
27. **Lorenz, M. G., and W. Wackernagel.** 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. **58:**563–602.
28. **Lutticken, R., N. Temme, G. Hahn, and E. W. Bartelheimer.** 1986. Meningitis caused by Streptococcus suis: case report and review of the literature. Infection **14:**181–185.
29. **Mannarelli, B. M., T. S. Balganesh, B. Greenberg, S. S. Springhorn, and S. A. Lacks.** 1985. Nucleotide sequence of the DpnII DNA methylase gene of Streptococcus pneumoniae and its relationship to the dam gene of Escherichia coli. Proc. Natl. Acad. Sci. USA **82:**4468–4472.
30. **Moineau, S., S. A. Walker, E. R. Vedamuthu, and P. A. Vandenbergh.** 1995. Cloning and sequencing of LlaDCHI restriction/modification genes from Lactococcus lactis and relatedness of this system to the Streptococcus pneumoniae DpnII system. Appl. Environ. Microbiol. **61:**2193–2202.
31. **Muckerman, C. C., S. S. Springhorn, B. Greenberg, and S. A. Lacks.** 1982. Transformation of restriction endonuclease phenotype in Streptococcus pneumoniae. J. Bacteriol. **152:**183–190.
32. **Ochman, H., A. S. Gerberi, and L. Hartl.** 1988. Genetic applications of an inverse polymerase chain reaction. Genetics **120:**621–623.
33. **Ogata, K., R. I. Aminov, T. Nagamine, M. Sugiura, K. Tajima, M. Mitsumori, T. Sekizaki, H. Kudo, H. Minato, and Y. Benno.** 1997. Construction of a Fibrobacter succinogenes genome map and demonstration of diversity at the genomic level. Curr. Microbiol. **35:**22–27.
34. **Osaki, M., D. Takamatsu, N. Tsuji, and T. Sekizaki.** 2000. Cloning and characterization of the gene encoding O-acetylserine lyase from Streptococcus suis. Curr. Microbiol. **40:**67–71.
35. **Palmer, B. R., and M. G. Marinus.** 1994. The dam and dcm strains of Escherichia coli—a review. Gene **143:**1–12.
36. **Perch, B., K. B. Pedersen, and J. Henrichsen.** 1983. Serology of capsulated streptococci pathogenic for pigs: six new serotypes of Streptococcus suis. J. Clin. Microbiol. **17:**993–996.
37. **Roberts, R. J., and D. Macelis.** 2000. REBASE-restriction enzymes and methylases. Nucleic Acids Res. **28:**306–307.
38. **Rochepeau, P., L. B. Selinger, and M. F. Hynes.** 1997. Transposon-like structure of a new plasmid-encoded restriction-modification system in Rhizobium leguminosarum VF39SM. Mol. Gen. Genet. **256:**387–396.

39. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
40. **Schleif, R.** 1980. Assaying of organisms for the presence of restriction endonucleases. Methods Enzymol. **65:**19–23.
41. **Segers, R. P., T. Kenter, L. A. de Haan, and A. A. Jacobs.** 1998. Characterisation of the gene encoding suilysin from Streptococcus suis and expression in field strains. FEMS Microbiol. Lett. **167:**255–261.
42. **Sekizaki, T., H. Ito, T. Asawa, and I. Nonomura.** 1993. DNA sequence of type 1 fimbrin, Fpul1, gene from a chicken pathogenic Escherichia coli serotype O78. J. Vet. Med. Sci. **55:**395–400.
43. **Serhir, B., D. Dugourd, M. Jacques, R. Higgins, and J. Harel.** 1997. Cloning and characterization of a dextranase gene (dexS) from Streptococcus suis. Gene **190:**257–261.
44. **Smith, H. E., U. Vecht, A. L. J. Gielkens, and M. A. Smits.** 1992. Cloning and nucleotide sequence of the gene encoding the 136-kilodalton surface protein (muramidase-released protein) of *Streptococcus suis* type 2. Infect. Immun. **60:**2361–2367.
45. **Smith, H. E., M. Damman, J. van der Velde, F. Wagenaar, H. J. Wisselink, N. Stockhofe-Zurwieden, and M. A. Smits.** 1999. Identification and characterization of the *cps* locus of *Streptococcus suis* serotype 2: the capsule protects against phagocytosis and is an important virulence factor. Infect. Immun. **67:**1750–1756.
46. **Stankevicius, K., P. Povilionis, A. Lubys, S. Menkevicius, and A. Janulaitis.** 1995. Cloning and characterization of the unusual restriction-modification system comprising two restriction endonucleases and one methyltransferase. Gene **157:**49–53.
47. **Stein, D. C., J. S. Gunn, and A. Piekarowicz.** 1998. Sequence similarities between the genes encoding the S.NgoI and HaeII restriction/modification systems. Biol. Chem. **379:**575–578.
48. **Sussenbach, J. S., C. H. Monfoort, R. Schiphof, and E. E. Stobberingh.** 1976. A restriction endonuclease from Staphylococcus aureus. Nucleic Acids Res. **3:**3193–3202.
49. **Takamatsu, D., M. Osaki, and T. Sekizaki.** 2000. Sequence analysis of a small cryptic plasmid isolated from Streptococcus suis serotype 2. Curr. Microbiol. **40:**61–66.
50. **Takeshita, S., M. Sato, M. Toba, W. Masahashi, and T. Hashimoto-Gotoh.**
1987. High-copy-number and low-copy-number plasmid vectors for lacZ α-complementation and chloramphenicol- or kanamycin-resistance selection. Gene **61:**63–74.
51. **Twomey, D. P., L. L. McKay, and D. J. O'Sullivan.** 1998. Molecular characterization of the *Lactococcus lactis Lla*KR2I restriction-modification system and effect of an IS*982* element positioned between the restriction and modification genes. J. Bacteriol. **180:**5844–5854.
52. **Ueno, T., H. Ito, F. Kimizuka, H. Kotani, and K. Nakajima.** 1993. Gene structure and expression of the MboI restriction-modification system. Nucleic Acids Res. **21:**2309–2313.
53. **Uhlin, B. E., and K. Nordstrom.** 1977. R plasmid gene dosage effects in Escherichia coli K12: copy mutants of the R plasmid R1drd-19. Plasmid **1:**1–7.
54. **Vaisvila, R., G. Vilkaitis, and A. Janulaitis.** 1995. Identification of a gene encoding a DNA invertase-like enzyme adjacent to the PaeR7I restriction-modification system. Gene **157:**81–84.
55. **Vovis, G. F., and S. Lacks.** 1977. Complementary action of restriction enzymes Endo R.DpnI and Endo R.DpnII on bacteriophage f1 DNA. J. Mol. Biol. **115:**525–538.
56. **Wilson, G. G., and N. E. Murray.** 1991. Restriction and modification systems. Annu. Rev. Genet. **25:**585–627.
57. **Xu, G.-L., W. Kapfer, J. Walter, and T. A. Trautner.** 1992. BsuBI—an isospecific restriction and modification system of PstI: characterization of the BsuBI genes and enzymes. Nucleic Acids Res. **20:**6517–6523.
58. **Xu, Q., S. Stickel, R. J. Roberts, M. J. Blaser, and R. D. Morgan.** 2000. Purification of the novel endonuclease, Hpy188I, and cloning of its restriction-modification genes reveal evidence of its horizontal transfer to the Helicobacter pylori genome. J. Biol. Chem. **275:**17086–17093.
59. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
60. **Zalkin, H., and P. Nygaard.** 1996. Biosynthesis of purine nucleotides, p. 561–579. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 1. ASM Press, Washington, D.C.