

## Different Foreign Genes Incidentally Integrated into the Same Locus of the *Streptococcus suis* Genome

Tsutomu Sekizaki,\* Daisuke Takamatsu,† Makoto Osaki, and Yoshihiro Shimoji

National Institute of Animal Health, Tsukuba, Ibaraki, Japan

Received 27 July 2004/Accepted 22 October 2004

Some strains of *Streptococcus suis* possess a type II restriction-modification (RM) system, whose genes are thought to be inserted into the genome between *purH* and *purD* from a foreign source by illegitimate recombination. In this study, we characterized the *purHD* locus of the *S. suis* genomes of 28 serotype reference strains by DNA sequencing. Four strains contained the RM genes in the locus, as described before, whereas 11 strains possessed other genetic regions of seven classes. The genetic regions contained a single gene or multiple genes that were either unknown or similar to hypothetical genes of other bacteria. The mutually exclusive localization of the genetic regions with the atypical G+C contents indicated that these regions were also acquired from foreign sources. No transposable element or long-repeat sequence was found in the neighboring regions. An alignment of the nucleotide sequences, including the RM gene regions, suggested that the foreign regions were integrated by illegitimate recombination via short stretches of nucleotide identity. By using a thermosensitive suicide plasmid, the RM genes were experimentally introduced into an *S. suis* strain that did not contain any foreign genes in that locus. Integration of the plasmid into the *S. suis* genome did not occur in the *purHD* locus but occurred at various chromosomal loci, where there were 2 to 10 bp of nucleotide identity between the chromosome and the plasmid. These results suggest that various foreign genes described here were incidentally integrated into the same locus of the *S. suis* genome.

*Streptococcus suis* is a gram-positive pathogen that has been identified as the cause of a wide range of clinical disease syndromes in pigs, other animals, and humans (5, 10, 28). Currently, *S. suis* strains are classified into 35 capsular serotypes (16, 17, 20, 34), and the genomic heterogeneity and the phylogenetic diversity of the strains have been described (9, 19, 32). Previous findings suggest that the heterogeneity of this bacterium is due, at least in part, to the integration of foreign genes (37, 38, 43).

One example of the foreign genes identified is a cluster of the *SsuDATII* genes encoding a type II restriction-modification (RM) system, which was found in *S. suis* strain DAT1 (37, 38). The RM gene region, comprising 3,503 bp, was thought to be inserted in an intergenic space of a conserved purine biosynthetic gene cluster between *purH* and *purD* (37, 38). A sequence comparison between strains with and without the RM genes showed that the borders of the RM region could be clearly recognized as if the exogenous region had simply been inserted. The other example of the foreign genes is the *sly* gene, which encodes a cholesterol-binding cytolysin (12, 15, 22). With a few exceptions, all *S. suis* strains have either *sly* or a hypothetical gene, designated *orf102*, at the same locus in the genome, and either of the two genes is thought to have been acquired by *S. suis* from a foreign source and to have replaced the gene that had existed in the locus. In contrast to the borders of the RM gene region, however, the flanking sequences of *sly* and *orf102* show a relatively low degree of similarity

(approximately 66% nucleotide identity) to one another, so that the borders of the foreign genes are unclear (43). In both examples, sequences affecting the integration of foreign genes, such as long-repeat sequences or transposable elements, were not found, a result which suggests that these integration events had occurred by illegitimate recombination. However, the structural differences of the borders between the *purHD* and *sly* regions lead us to raise the question of whether these foreign genes were integrated via the same mechanism of gene acquisition.

In bacteria, a genetic conversion which joins DNA molecules at sites where they have no or a few identical base pairs is referred to as illegitimate (nonhomologous) recombination (2). Such events were often identified as intramolecular recombination events following double-strand breaks and usually occurred within or near small direct or inverted repeats. These events are exemplified by the formation of deletions in bacterial chromosomes (6) and plasmids (1, 3) and the formation of transducing phages after UV irradiation of lysogenic cells (14, 39). On the other hand, it was shown that naturally transformable bacteria incorporate a foreign DNA that has a region with homology to the resident genome on one side (11, 35). The homologous stretches serve as anchors for homologous recombination and subsequently create a novel joint on the other side, where a short (a few base pairs) stretch of sequence identity exists, by illegitimate recombination. Such foreign DNA acquisition, referred to as homology-facilitated (or homology-directed) illegitimate recombination, is thought to occur widely in many bacterial species (30). It is, therefore, possible that a similar mechanism for genetic conversion was employed to generate the genetic structures as observed in the *purHD* and *sly* regions of *S. suis*. However, the natural transformation has not yet been demonstrated for *S. suis*, and only two examples of foreign gene regions have so far been examined. Thus, fur-

\* Corresponding author. Mailing address: Molecular Bacteriology Section, National Institute of Animal Health, 3-1-5 Kannondai, Tsukuba, Ibaraki 305-0856, Japan. Phone: 81-29-838-7743. Fax: 81-29-838-7907. E-mail: sekizaki@affrc.go.jp.

† Present address: Division of Infectious Diseases, Veterans Affairs Medical Center and the University of California, San Francisco, CA 94121.

TABLE 1. *S. suis* strains used in this study

Strain	Sero-type	Source	Origin	Reference
DAT1	2	Diseased pig	Japan	38
227	2	Diseased pig	Japan	38
NCTC10237	1	Diseased pig	The Netherlands	34
NCTC10234	2	Diseased pig	The Netherlands	34
4961	3	Diseased pig	Denmark	34
6407	4	Diseased pig	Denmark	34
11538	5	Diseased pig	Denmark	34
2524	6	Diseased pig	Denmark	34
8074	7	Diseased pig	Denmark	34
14636	8	Diseased pig	Denmark	34
22083	9	Diseased pig	Denmark	16
4417	10	Diseased pig	Denmark	16
12814	11	Diseased pig	Denmark	16
8830	12	Diseased pig	Denmark	16
10581	13	Diseased pig	Denmark	16
13730	14	Diseased human	The Netherlands	16
NCTC10446	15	Diseased pig	The Netherlands	16
2726	16	Diseased pig	Denmark	16
93A	17	Clinically healthy pig	Canada	16
NT77	18	Clinically healthy pig	Canada	16
42A	19	Clinically healthy pig	Canada	16
86-5192	20	Diseased calf	United States	16
14A	21	Clinically healthy pig	Canada	16
88-1861	22	Diseased pig	Canada	16
89-2479	23	Diseased pig	Canada	17
88-5299A	24	Diseased pig	Canada	17
89-3576-3	25	Diseased pig	Canada	17
89-4109-1	26	Diseased pig	Canada	17
89-5259	27	Diseased pig	Canada	17
89-590	28	Diseased pig	Canada	17

ther examples are needed to analyze the mechanism of foreign gene acquisition in *S. suis*.

We report here that some serotype reference strains of *S. suis* which were previously shown to lack the RM system possess various genes between *purH* and *purD*. Nucleotide sequence comparisons among them and the experimental integrations of a plasmid carrying the RM genes into the *S. suis* genome suggest that illegitimate integrations of foreign genes in *S. suis* occurred randomly via a few base pairs of nucleotide identity between resident and foreign DNA.

MATERIALS AND METHODS

**Bacterial strains, media, and culture conditions.** The *S. suis* strains used in this study are listed in Table 1. These strains, except DAT1 and 227, are the reference strains of each serotype. The phylogenetic relationships among the strains used have already been described elsewhere (9, 19, 32, 43). The serotype reference strains of *S. suis* described hereafter are referred to by their serotypes. All of the *S. suis* strains were grown in Todd-Hewitt (TH) broth or agar medium (Difco Laboratories, Becton Dickinson, Sparks, Md.) at 28 or 37°C under 5% CO<sub>2</sub>. The *Escherichia coli* strain used was MC1061 (8), which was cultured in Luria-Bertani broth or agar medium (Difco Laboratories) at 37°C. When necessary, antibiotics were added to culture media at the following concentrations: 50 µg of ampicillin/ml, 10 µg of chloramphenicol (CAM)/ml, and 50 µg of spectinomycin (SPC)/ml for *E. coli*; and 5 µg of CAM/ml and 100 µg of SPC/ml for *S. suis*.

**DNA methods.** Restriction enzymes and DNA-modifying enzymes were purchased from Takara Shuzo Co., Ltd. (Tokyo, Japan) and used according to the manufacturer's recommendations. The vector plasmids used were pSET1 (41) and pSET4s (42). Minipreparations of recombinant plasmids from *E. coli* and transformations of *E. coli* were performed by standard procedures (36). Isolation of plasmids from *S. suis* and transformation of *S. suis* were carried out by using the methods described previously (40, 41). Genomic Southern hybridization was performed by using the procedures described previously (33). DNA fragments used for probes were labeled with digoxigenin by using a digoxigenin-PCR labeling mixture (Roche Diagnostics GmbH, Mannheim, Germany) according to

the manufacturer's instructions. The DNA fragments were amplified by PCR with the following primers: SsuDAT1I5' plus SsuDAT1I3' for SsuDAT1I genes; 16conF1 plus 16conR3 for serotype 16; 10conF1 plus 10conR2 for serotype 10; 12conF1 plus 12R1 for serotype 12; 13conF plus 13conR for serotype 13; 1conF plus 1conR for serotype 1; 6conF plus 6conR for serotype 6; and 20conF plus 20conR for serotype 20. PCR was performed by using a Perkin-Elmer thermal cycler model 2400 or 9600 (PE Biosystems Japan, Tokyo, Japan) and by using Takara *Ex Taq* polymerase and a Takara LA PCR kit, as described previously (38, 41).

**DNA sequencing and data analysis.** PCR products were directly sequenced by dye terminator chemistry by using an Applied Biosystems model 310 automated DNA sequencer (PE Biosystems). The sequencing was done initially with the primers used for PCR, followed by a series of primer walking reactions with primers designed from the sequenced regions. The sequence data obtained were assembled and analyzed with Sequencher software, version 3.1.1 (Hitachi Software Engineering Co., Ltd., Yokohama, Japan), and GENETYX-MAC software, version 12.1 (GENETYX Corp., Tokyo, Japan). The putative genes were identified on the basis of the adopted criteria that an open reading frame (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 deduced amino acid sequences obtained were searched for similarity with databases by using the BLAST network service available at the National Center for Biotechnology Information, Bethesda, Md. (<http://www.ncbi.nlm.nih.gov>).

**Experimental integration of plasmid. (i) Insertion of a gene cassette into the RM genes.** As mentioned above, the recombinant plasmids were constructed in *E. coli*. Because genes for the SsuDAT1I system overlap each other in a head-to-tail manner and the restriction gene must be accompanied by the cognate modification gene (38), the RM gene region with insertion of a gene cassette was constructed by allelic exchange on the genome as shown in Fig. 1. A 936-bp

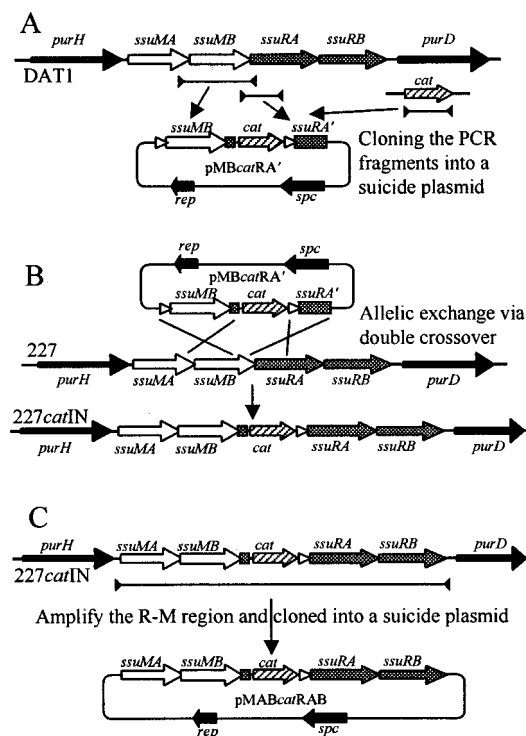


FIG. 1. Strategy for construction of a suicide plasmid carrying a *cat* gene cassette placed between the *ssuMB* and *ssuRA* genes of the *SsuDAT1I* system. (A) Construction of pMBcatRA'. Lines between the closed arrowheads are PCR fragments amplified from the above regions. The amplified fragments were fused and cloned into pSET4s. (B) Generation of a mutant RM gene containing the *cat* gene cassette by allelic exchange via double crossover on the chromosome. (C) Construction of pMABcatRAB. The line between the closed arrowheads is a PCR fragment amplified from the genomic DNA of *S. suis* 227*catIN*. The amplified fragment was cloned into pSET4s.

fragment containing the whole length of the *ssuMB* gene with a 120-bp upstream region and a 3-bp downstream region, which overlapped the first 23 nucleotides of the *ssuRA* gene, and a 594-bp fragment containing the first 443 nucleotides of the *ssuRA* gene, which overlapped the last 171 nucleotides of the *ssuMB* gene, were amplified separately from the genomic DNA of *S. suis* DAT1 with primers MBKI5' plus MBKI3' and RAKI2-5' plus RAKI3', respectively. The two PCR products were digested with EcoRI plus EcoT22I or EcoT22I plus BamHI, fused at the EcoT22I site, and cloned between the EcoRI-BamHI sites of pSET4s, generating pMBRA'. A CAM acetyltransferase (*cat*) gene cassette containing its promoter region but missing the transcriptional terminator was amplified from pSET1 with primers CT2 plus CT3, digested with EcoT22I, and cloned into a unique EcoT22I site of pMBRA', generating pMBcatRA' (Fig. 1A).

(ii) **Gene replacement by allelic exchange in *S. suis*.** One field isolate, strain 227, was selected for the experiment as it did not carry any cryptic plasmids and was shown to possess the same genetic organization with respect to the RM genes as strain DAT1 (38). Procedures for selection of mutants whose genes were replaced by an allelic exchange via double crossover were described previously (33, 42). Briefly, *S. suis* strain 227 was transformed with pMBcatRA', and the cells were grown at 28°C in the presence of CAM and SPC. At mid-logarithmic growth phase, the cells were diluted with TH broth containing CAM and grown at 28°C to early logarithmic phase. The cultures were then shifted to 37°C and incubated for 4 h. Subsequently, the cells were spread on TH agar containing CAM and incubated at 37°C. Temperature-resistant CAM-resistant (Cm<sup>r</sup>) colonies obtained were screened for loss of vector-mediated SPC resistance (Sp<sup>c</sup>) to detect putative mutants which had exchanged their wild-type allele for a genetic segment containing the *cat* gene as a consequence of homologous recombination via a double crossover. Finally, the genetic organization of the mutant alleles was examined by Southern hybridization, and the resulting mutant obtained was designated 227*cat*IN (Fig. 1B).

(iii) **Selection of integration mutants.** The DNA region containing the whole length of the RM genes with insertion of the *cat* cassette was amplified with primers SsuDAT115'B plus SsuDAT113'E from the genomic DNAs of 227*cat*IN. The fragment was digested with BamHI plus EcoRI and cloned into the corresponding sites of pSET4s to generate the plasmids designated pMABcatRAB (Fig. 1C). *S. suis* serotype 2 (NCTC10234) was transformed with pMABcatRAB and grown on TH agar containing CAM and SPC at 28°C for 40 h. The cells were harvested and suspended in saline solution. Aliquots of appropriately diluted bacterial suspension, which contained approximately 10<sup>7</sup> cells, were spread on prewarmed TH agar containing CAM and grown at 37°C for 18 h. Bacterial lawn that appeared on the plate was transferred to another TH agar plate containing CAM by replica plating and grown at 37°C for 18 h. Colonies that appeared on the plates were purified by two consecutive single-colony isolations on TH agar containing CAM at 37°C. Two hundred forty-three temperature-resistant Cm<sup>r</sup> mutants which did not possess an autonomously replicating plasmid (e.g., one that had lost its temperature sensitivity) were selected from 25 independent experiments and used in this study. Twelve of the mutants obtained from 3 of the 25 independent experiments were designated Int1 to Int12. For estimation of the integration frequency, the bacterial suspension of *S. suis* serotype 2 carrying pMABcatRAB was spread on prewarmed TH agar containing CAM plus SPC and grown at 37°C for 24 h. The numbers of temperature-resistant colonies that appeared on the agar were used for the calculation.

(iv) **Nucleotide sequence determination of the junction between the integrated plasmid and genome.** The genomic DNA of the temperature-resistant Cm<sup>r</sup> mutants was digested with EcoT22I plus PstI, which cut pMABcatRAB at three sites and generated cohesive ends compatible with each other. The digested DNA was self-ligated and used as templates for inverse PCRs with two sets of primers, each of which amplified either end of the junction regions independently. The primers, which were designed from the sequence of pMABcatRAB, were as follows: TS2 plus oriE1-50 and SsuMB3' plus TS1 for mutants Int1, Int5, Int6, and Int7; TS2 plus PC1 and SsuMB3' plus TS1 for mutants Int2, Int3, Int9, Int11, and Int12; MA1 plus MA2 and MA3 plus dam-in2 for mutant Int4; SsuRA3' plus PC1 and SsuMB3' plus RA3 for mutant Int10. The DNA fragments amplified by the inverse PCR were directly sequenced. On the basis of the sequences of junction regions, 11 primer sets were designed to amplify the target regions where the plasmid was integrated, from the genomic DNA of *S. suis* serotype 2. The primers used for the amplifications of the target regions of mutants Int1, Int2, Int3, Int4, Int5, Int6, Int7, Int9, Int10, Int11 and Int12 were 1-1FJ plus 1-1Rend, 2-1Fend plus 2-1Rend, 3-1Fend plus 3-1Rend, 3-2Fend plus 3-2Rend, 5FJ plus 5Rend, 9Fend plus 9Rend, 21Fend plus 21Rend, 2-61Fend plus 2-61Rend, 2-69FJ plus 2-69RJ, S-1Fend plus S-1RJ, and S-3Fend plus S-3Rend, respectively.

**Oligonucleotides used in PCRs.** The following oligonucleotides were purchased from Hokkaido System Science Co., Ltd. (Sapporo, Japan) and were used

as PCR primers (shown in the 5'-to-3' direction): purH, GCGCTAGCTATTT TGACCAATA; purD, GCAAAGTCTTTTGACCACTCTA; purN-F, GTGTT TGATCAGGCAACGGCTCCA; purH-R2, GTCTTGGTCACGGACAGAA CCACCT; purD-F2, GAGCAGGTCTTTGTTGCTCCTGGAA; purE, GCCC CTGGCATTGTCACGATA; MBKI5', CAAGCCCACTAGTTGAAGAATT GTA; MBKI3', CTATGAAAAGTCTTTTGTGATCT; RAKI2-5', ATGCA TGCTTCAACTCAAGAAGGAGA; RAKI3', CCGTAAACATAGTCTGTGA AGTTT; RAKI5', AACTAGATTGGATAAGGAAGATGA; CT2, ATGCATC ACCGAACCTAGAGCTTGATG; CT3, ATGCATATTATAAAGCCAGTCA TTA; SsuDAT115'B, CGGATCCAAGTATACACCCAGCTGGGAAG; SsuDAT113'E, GGAATTCCTTGATTATCTAAACAAATCATGC; SsuDAT115', AAGTATAGCACCCAGCTGGAGAAG; SsuDAT113', CTTGATTAT CTAACAAATCATGC; SsuMB3', CACATCTGGATACCTGTGCAC; SsuRA3', GTAA TCTGGCGTTCGATTAG; SsuRB5', TGGTGAAGTGTGGCAAAGGT; MA1, CCCCTCTCAACAAACGGTTTCAAA; MA2, TCCGATGTGGAGCGAGCA GCACGCA; MA3, CTGGGCTTGAAGAATTAGAAAGCAT; RA3, TACAGG ACTATGTTTACGGTGTAGA; oriE1-50, GCTCACTATTAGCCACCCAG GCT; SP3, ACTAGTGTTCGTGAATACATGTTATA; SP4, ACTAGTGTTC TAAAATCTGAT; PC1, ACTAGTTATCTACACGACGGGG; LZ3, AGATCTC GGTGATGACGGTGAAGAAC; TS1, AGATCTATTAATCGCAACATCAA ACC; TS2, ACTAGTTATCGGCATAATCGTT; Rafuse3', CCGTAAACATAG CCTGTAAAGTTT; dam-in2, CAGTGGTATGGCCTTAGAA; dam-in4, TGG TCGTAAAAATAGAAATGGGT; dam-in7, GGTCTAAGTTGAACAGAGAA; 16conF1, CAAGGAGTGTGCTAAATAACCTTCT; 16conR3, AACTTGCCTTA CTCTCACATTA; 10conF1, CAGATAGTGTGAATGAAGTTAAGTT; 10conR2, CAATCTACCATTGGTGCAGATAC; 12conF1, CG AGAAATAC AGTATTAACAGTCC; 12R1, CAATAGCCGATACCATCATGAACAAAGT; 13conF, CGCTTCGTAATGGGGAGGCGAAAGA; 13conR, GCAAGTATTTT GCCTTGTTAGAGA; 1conF, CTTTTGTGAGAAGGAGGAGTTAGA; 1conR, AGAAAAAAGCACCTTAAGGTGCG; 6conF, GGAGTGGATTGCA GTATGTTTGGGA; 6conR, AGAGGGAGGAGAATAACTAGCAAT; 20conF, GTCTAGTCCGTTAAGCTGTTCTCA; 20conR, GTCTCTACTGCTAGCCT CCTCT; 1-1FJ, GCCATCTACCCATCCAAATAA; 1-1Rend, GTAGTGGG GCCTATCGAATTTGGTA; 2-1Fend, TTCAAGTATCGCAGGCTTACTCTCT; 2-1Rend, CAATCTCTCAGGCTGGACACCTT; 3-1Fend, AGTCTGCAATTT GGCGACCAATCTT; 3-1Rend, AGCCAAACACGGAGGAAGTTTATCA; 3-2Fend, CCTCTATGCC GCGTCTGAGCTATT; 3-2Rend, CTTCAGCAGC ATGGCTGCTGTGAT; 5FJ, CTGAG GTGTAGTCTCCCTTCAAT; 5Rend, ATTGCGATTTGGATGCGGCCGCTT; 9Fend, GGACCTGTACTACCCCT ACCTGGAA; 9Rend, TGATTATTTAGAAACTGGTCTGAT; 21Fend, CAAC CTCTCTTTAGAGTATGCTGT; 21Rend, GAGAAAAAAGCATACTTTGAT GATA; 2-61Fend, GCATGTATTGCCTCTTTCATTCGAT; 2-61Rend, TGGCG GCAGG TCTGTAGGAACAAT; 2-69FJ, AGTCAATTCGAGCTGAGAATCT GAC; 2-69RJ, GAAGATCGTGTGAAGAAGCAGGCTT; S-1Fend, TTCTACA AGTGTGATGGCAGCTCA; S-1RJ, GAGCTGAATAAGTATGCTGATC GAA; S-3Fend, AGTGCAGTAGTTGATTGCTCTAGCA; and S-3Rend, TGCC CTAGACAAATCTTATAGGATT.

**Nucleotide sequence accession numbers.** The nucleotide sequences for the *purHD* regions in serotypes 1, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 24, 25, 27, and 28 determined in this study have been deposited in the DDBJ/EMBL/GenBank database under accession nos. AB183829, AB183830, AB183831, AB183832, AB183833, AB183834, AB183835, AB183836, AB183837, AB183838, AB183839, AB183840, AB183841, AB183842, AB183843, AB183844, AB183845, AB183846, and AB183847, respectively. Sequences for the *purHD* regions in serotypes 2, 3, 7, 23, and 26, previously published elsewhere (38, 39), were extended in this study and were assigned accession nos. AB045618, AB058942, AB058943, AB058944, and AB058945, respectively. The nucleotide sequence data of the genomic regions in serotype 2, which were used for the targets of the integration mutants Int1, Int2, Int3, Int4, Int5, Int6, Int7, Int9, Int10, Int11, and Int12, have been deposited in the database under accession nos. AB183474, AB183475, AB183476, AB183477, AB183478, AB183479, AB183480, AB183481, AB183482, AB183483, and AB183484, respectively.

## RESULTS

**Genes and gene products of the extra regions between *purH* and *purD*.** With a set of PCR primers, *purH* plus *purD*, which were designed from the internal sequences of *purH* and *purD*, respectively, we were able to amplify the *purHD* regions, ranging from 4.5 to 1.0 kb, of the reference strains excluding serotypes 17, 18, 19, and 21 (data not shown). In the serotypes 17, 18, 19, and 21, however, each of the *purH* and *purD* gene

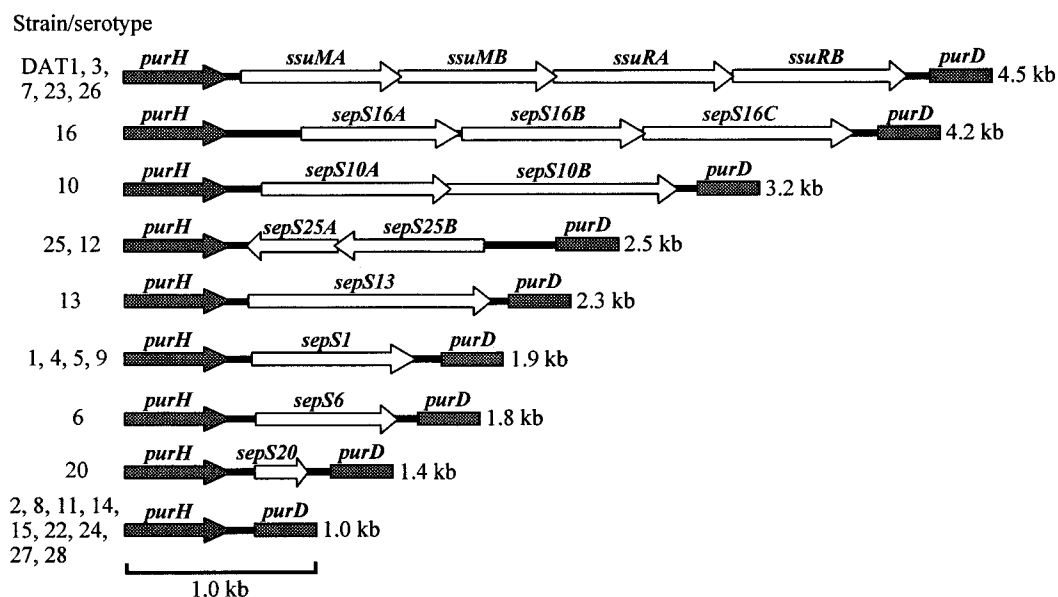


FIG. 2. Schematic representations of *purHD* regions. Genetic organizations of representative serotypes (serotypes 16, 10, 25, 13, 1, 6, 20, and 2) are depicted here, and the corresponding serotypes that showed identical or very similar genetic structures are listed on the left. The structures of RM genes were taken from previous reports (37, 38). Sizes of the DNA fragments amplified by PCR are shown on the right. Open arrows, genes identified in these regions; filled arrows, *purH*; filled boxes, *purD*.

regions could be amplified with specific primers, i.e., *purN-F* plus *purH-R2* and *purD-F2* plus *purE*, respectively (data not shown). Therefore, the *purH* gene was thought to be separated from the *purD* gene in these serotypes, and the four serotypes above have not been examined further. The amplified *purHD* fragments above were directly sequenced. The putative genes and their organization found in the sequenced fragments are summarized in Fig. 2. All of the genes identified in this study were located on the same DNA strand as the *pur* genes except for those found in the serotypes 12 and 25, which were located on the complementary strand. No transposable element or long-repeat sequence was found in the sequenced regions. The average G+C content of each sequence, especially that of the coding region (27 to 35%), was clearly lower than that of conserved *pur* genes (49%) (38) and of the total genome of *S. suis* (39 to 41%) (24).

The 4.5-kb fragments obtained from strain DAT1 and serotypes 3, 7, 23, and 26 contained genes encoding SsuDAT1I or its isoschizomers, as has been previously described (37), whereas no putative gene was found in the intergenic space of the 1.0-kb fragments. The 4.2-kb fragment obtained from serotype 16 contained three putative genes, designated *sepS16A* (for *suis* extra gene in *pur* region of serotype 16), *sepS16B*, and *sepS16C*, the latter two overlapped each other in a head-to-tail manner. The 3.2-kb fragment obtained from serotype 10 contained two putative genes, designated *sepS10A* and *sepS10B*, which overlapped each other in a head-to-tail manner. The 2.5-kb fragment obtained from serotype 25 contained two overlapping genes, designated *sepS25A* and *sepS25B*. The 2.5-kb fragment obtained from serotype 12 was almost identical to that of serotype 25 (97.2% nucleotide identity); however, an ORF corresponding to *sepS25B* in serotype 12 contains an 8-bp duplication that causes a frameshift mutation, resulting in a 180-nucleotide truncation at the 5' terminus. On

the basis of the 16S rRNA sequences (9, 19, 43), serotype 12 is not related to serotype 25. The 1.9-kb fragments obtained from serotypes 1, 4, 5, and 9 were very similar to one another (95.9 to 99.9% nucleotide identity). They each contained a single gene; that which was found in the fragment of serotype 1 was designated *sepS1*. Although serotypes 4 and 5 are closely related in the 16S rRNA-based tree, they are genetically distant from serotypes 1 and 9, which are also unrelated to each other (9, 19, 43). The 2.3-, 1.8-, and 1.4-kb fragments obtained from serotypes 13, 6, and 20, respectively, each contained a single gene, designated *sepS13*, *sepS6*, and *sepS20*, respectively. Thus, the genetic regions containing the novel genes were categorized into seven classes.

The deduced translational products of the genes found in the extra genetic regions are summarized in Table 2. The deduced SepS16A, SepS16B, and SepS16C proteins showed amino acid similarity with *Staphylococcus aureus* hypothetical proteins. As with *S. suis*, the genes showing similarity on the *S. aureus* genome are also contiguous, although they are not positioned between the *S. aureus purH* and *purD* genes. The deduced SepS10A protein showed a low degree of amino acid similarity to a *Staphylococcus epidermidis* hypothetical protein. The deduced SepS25A and SepS25B proteins showed amino acid similarity to *Streptococcus pyogenes* hypothetical proteins Spy0970 and Spy0968, whose genes were also located in tandem but were separated from the *pur* genes on the *S. pyogenes* genome. The rest of proteins encoded in the extra regions showed no significant similarity to proteins in the database. The extra regions of the seven classes, together with the RM genes, showed no similarity with one another at either the DNA or deduced amino acid sequence levels.

**Distribution of the extra genes among *S. suis*.** The internal regions of the PCR fragments which contained the putative gene(s) were amplified with specific primers from the *purHD*

TABLE 2. Predicted gene products and the similarity to amino acid sequences in the database

Gene	Product size		Similarity <sup>a</sup>		Description of closest relative <sup>c</sup>
	No. of amino acids	kDa	BLAST E value	Amino acid identity <sup>b</sup>	
<i>sepS16A</i>	275	33.0	9e-53	39/275	Hypothetical protein of <i>S. aureus</i> [AC025591]
<i>sepS16B</i>	315	36.1	3e-46	35/314	Hypothetical protein of <i>S. aureus</i> [AC025591]
<i>sepS16C</i>	363	42.1	7e-37	28/379	Hypothetical protein of <i>S. aureus</i> [AC025591]
<i>sepS10A</i>	319	36.4	1e-14	26/276	Hypothetical protein of <i>S. epidermidis</i> [AF270077]
<i>sepS10B</i>	390	46.0			No similarity
<i>sepS25A</i>	150	16.9	9e-18	37/127	Hypothetical protein (Spy0970) of <i>S. pyogenes</i> SF370 [AE006544]
<i>sepS25B</i>	311	36.8	8e-67	47/263	Hypothetical protein (Spy0968) of <i>S. pyogenes</i> SF370 [AE006544]
<i>sepS13</i>	418	48.5			No similarity
<i>sepS1</i>	283	32.6			No similarity
<i>sepS6</i>	248	28.9			No similarity
<i>sepS20</i>	93	10.5			No similarity

<sup>a</sup> Values do not appear in the respective sections for the putative proteins that showed no similarity to the proteins in the database.

<sup>b</sup> Percent amino acid identity/number of amino acids evaluated.

<sup>c</sup> Numbers in brackets are accession numbers from the DDBJ/EMBL/GenBank database.

fragments obtained above. The fragments were labeled and used as hybridization probes against genomic DNAs from *S. suis* strains that had been digested with HindIII and separated by agarose gel electrophoresis. A DNA probe obtained from serotype 16 showed a hybridizing fragment of 18 kb only in the digested DNA of serotype 16, whereas no hybridizing fragments appeared in those of other strains (data not shown). A DNA probe prepared from serotype 1 showed a hybridizing fragment of either 2.8 or 9.5 kb in the digested DNAs of serotypes 1, 4, 5, and 9, whereas no hybridizing fragments appeared in those of other strains (data not shown). Similarly, probes derived from the internal regions of gene segments between *purH* and *purD* in other serotypes hybridized only with DNA from serotypes, yielding a similar class of PCR amplicon between *purH* and *purD* as the strain from which the probe was derived.

**Nucleotide sequence comparison of the *purHD* regions.** The nucleotide sequences determined were aligned bidirectionally from the *purH* and *purD* sides. As shown in Fig. 3, the genetic regions expressed to yield PurH and PurD were almost completely identical to one another in all of the strains examined. The intergenic sequences of serotypes 22, 11, 24, 2, 14, 15, 8, 27, and 28, which did not possess any genes between *purH* and *purD*, were very similar, except that a 7-bp deletion was noted in serotype 22 (Fig. 3, lower alignment). Although the RM gene regions (Fig. 3, upper alignment, lines 1 to 5 from the top) have been thought simply to be inserted 11 bp downstream of the stop codon of *purH*, as described before (38), the alignment including other sequences determined in this study revealed different putative borders which discriminate the extra genetic regions from the conserved *pur* genes. As shown in the lower alignment of Fig. 3, serotypes 25 and 12 showed nucleotide mismatches further upstream from the putative insertion site of the RM genes, whereas serotypes 20, 16, 9, 1, 4, 5, 10, 13, 25, and 12 showed nucleotide mismatches further downstream. These results indicated that the extra regions of these serotypes were not simply inserted between *purH* and *purD*, but that either of these genes was integrated in this position and replaced the intergenic sequences. Moreover, serotypes 20, 16, 9, 1, 4, and 5 showed regions of 37 to 40 bp homologous to the corresponding regions of DAT1 and serotypes 3, 7, 23, and 26

(5' regions of the upper alignment in Fig. 3). The homologous regions are apparently parts of the extra regions and a few nucleotide mismatches common to serotypes 26, 20, 16, 9, and 1 also appear within the homologous regions, indicating that either of these extra genes that had existed at this position was further replaced by another gene.

One-to-one comparison of the sequences more obviously showed the borders of the junction regions where the recombination events to incorporate the extra regions may have occurred; some representatives are shown in Fig. 4. Next to the highly conserved *pur* regions, sequences with a low degree of similarity, where a few base pairs of identity appeared contiguously, were seen at one or both ends of the border regions. The relatively low degree of sequence similarity shown in the junction regions resembles that observed in the flanking regions of *sly* and *orf102* (43). The putative borders appeared at the same positions in a few comparisons (e.g., a comparison among serotypes 22, 25, and 10); however, most of them appeared at different positions from one another. Moreover, the comparison between serotype 26 (or other strains possessing the RM genes) and serotype 22 revealed that serotype 26 (and other strains possessing the RM genes) had small imperfect direct repeats consisting of 7 bp at the putative borders. A very similar 7-bp sequence was also present in serotype 22 (Fig. 4), a result which suggested that the extra region of serotype 26 (or other strains possessing the RM genes) had been deleted at these short repeats.

**Integration of a plasmid carrying RM genes.** We examined whether the *purHD* region is a hot spot for illegitimate recombination by using *SsuDATII* genes as the model. The *SSuDATII* genes with an insertion of the *cat* gene were cloned into a suicide plasmid, and the recombinant plasmid was introduced into *S. suis* serotype 2, for which the *purHD* region did not possess any foreign genes (38). Since temperature-resistant and Cm<sup>r</sup> mutants obtained from the bacterial culture were expected to carry the RM genes being integrated into the genome, we selected 243 such mutants. All of the mutants were still Spc<sup>r</sup>, indicating that the mutants contained the vector region as well as the RM gene region. Genomic DNAs were isolated from the mutants and used for PCR with primers *purN-F* plus *purE*, which encompassed the *purHD* region. All

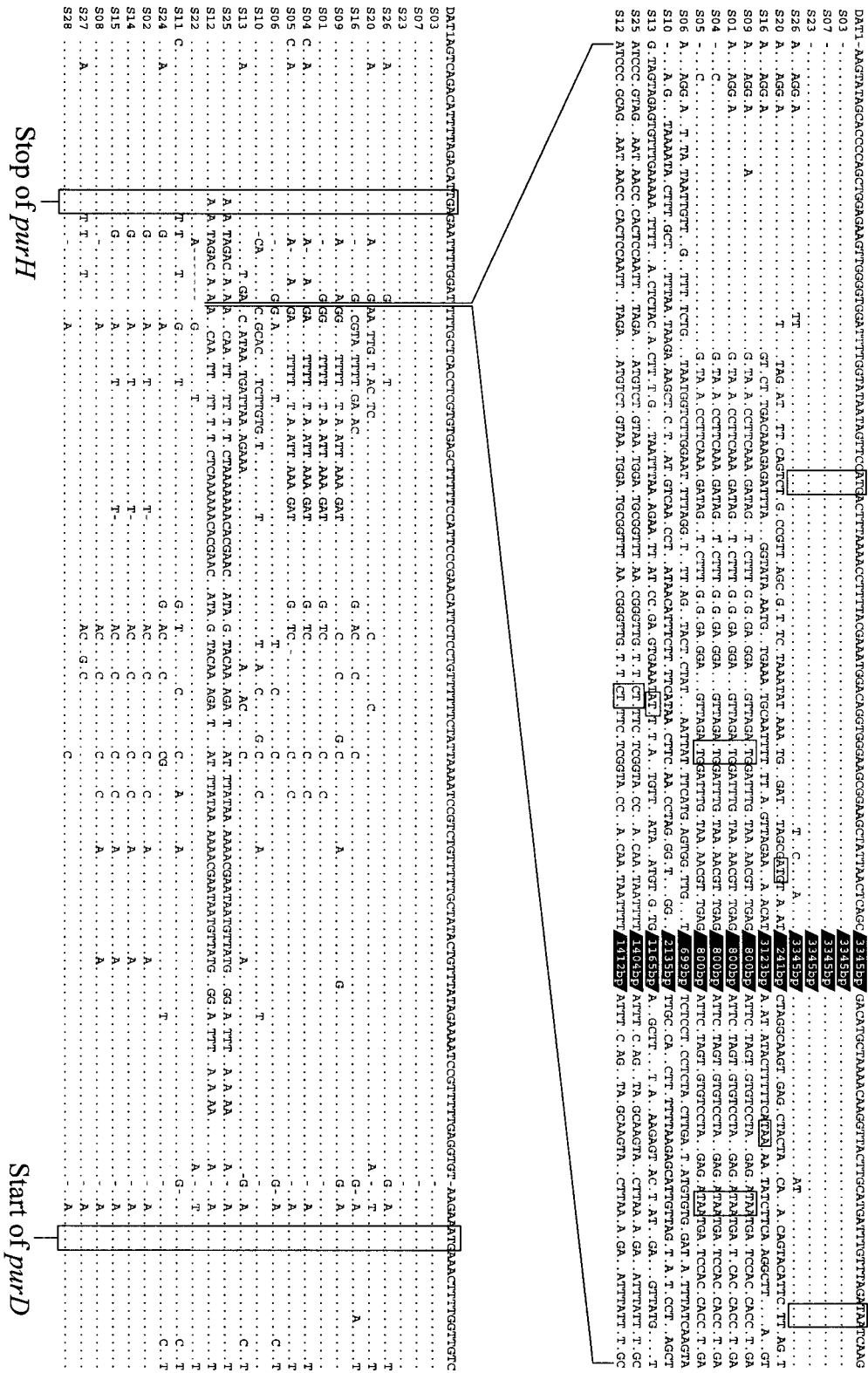


FIG. 3. Alignment of the partial sequences of the *purH*D regions from *S. suis* reference strains and DAT1. The sequences are aligned bidirectionally from the *purH* and *purD* regions (lower sequences), and the extra portions are shown in the upper alignment and are connected by the lines with the lower sequences. Strains are indicated to the left of the sequences; reference strains are referred to as S plus the serotype number. Numbers of nucleotides not appearing in the upper alignment are indicated in white type in black parallelograms. 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 the sequences.

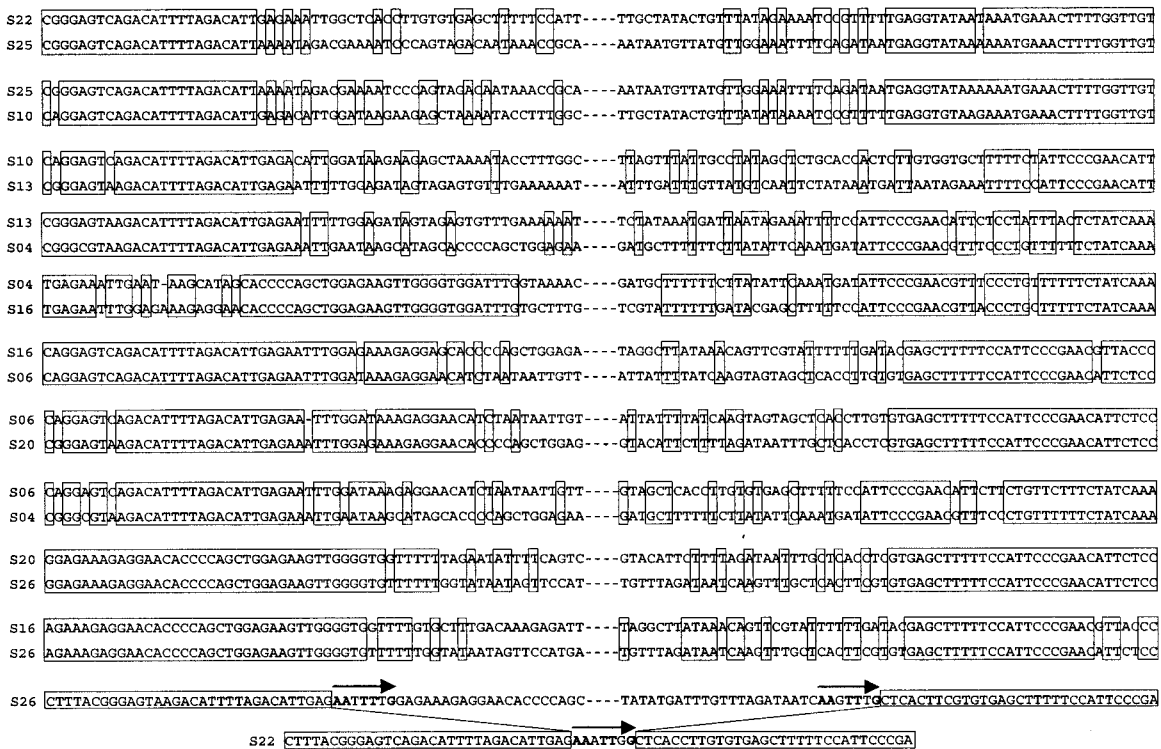


FIG. 4. One-to-one sequence comparisons of the border regions of some serotypes. Serotypes compared are indicated to the left of the sequences as S plus serotype number. Boxes, identical nucleotides; arrows, short direct repeats.

of the samples produced amplicons of the same size (4.0 kb) as that of the parent strain (data not shown), indicating that the genetic organization of the amplified regions was not affected by the integration, i.e., none of the integration events had occurred at the *purHD* locus.

**Structures of integrated plasmids.** To analyze the genetic structures of the integrated plasmids, we further examined 12 temperature-resistant *Cm<sup>r</sup>* mutants, named Int1 to Int12. The genomic DNAs of the mutants were digested with *Stu*I, which cut the pMAB*cat*RAB at one site (see Fig. 6). The digested DNAs were separated by agarose gel electrophoresis and examined by Southern hybridization with an SsuDATII probe. The digested DNAs of *S. suis* strain 227, which possesses the RM genes, and its derivative 227*cat*IN provided one and two fragments that hybridized with the probe, respectively, whereas those of the mutants, except Int8, provided two hybridizing fragments of different sizes (Fig. 5), demonstrating that pMAB*cat*RAB was integrated into the *S. suis* genome at different positions. The digested DNA of the mutant Int8 provided three hybridizing fragments, one of which was the same size as the whole length of pMAB*cat*RAB, implying that the plasmid was duplicated in tandem.

The 12 mutants were then analyzed by PCR by using various combinations of primers designed from the sequence of pMAB*cat*RAB. The strategy of using PCRs for mapping the structure of the integrated plasmids assumed a Campbell-type recombination event. If the plasmid integrates by a single recombination event, then PCR amplicons would be obtained from internal segments of the integrated plasmid whereas plasmid-derived priming sites at the junction of the recombination

event would be separated and would not yield amplicons. As shown in Fig. 6, the genomic DNAs of the mutants Int1, Int2, Int3, Int5, Int6, Int7, Int9, Int11, and Int12 produced PCR fragments for all of the primer sets except those encompassing the replication region (*rep*) of the plasmid, indicating that the whole plasmid was integrated into the host genome via the *rep* region. Similarly, the genomic DNAs of the mutants Int4 and Int10 did not produce a PCR fragment with the primer sets encompassing *ssuMA* and *ssuRA*, respectively, indicating that these mutants formed cointegrates via these regions. However, the genomic DNA of the mutant Int8 produced PCR fragments with all of the primer sets examined. This result, together with the above results, indicates that at least two copies of the plasmid were present in tandem in the genome due to tandem duplication.

Selection of the mutants by CAM did not allow us to estimate the integration frequency, as many false-positive colonies

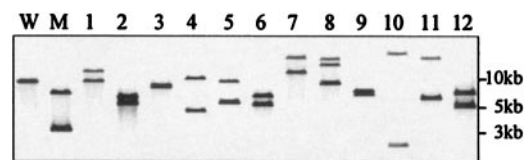


FIG. 5. Genomic Southern hybridization of temperature-resistant *Cm<sup>r</sup>* *S. suis* mutants. Genomic DNAs were digested with *Stu*I, separated by agarose gel electrophoresis, and probed with labeled SsuDATII genes. Molecular sizes are indicated on the right. Lane numbers correspond to the numbers in the designations of mutants Int1 to Int12. W, wild-type strain 227; M, mutant strain 227*cat*IN.

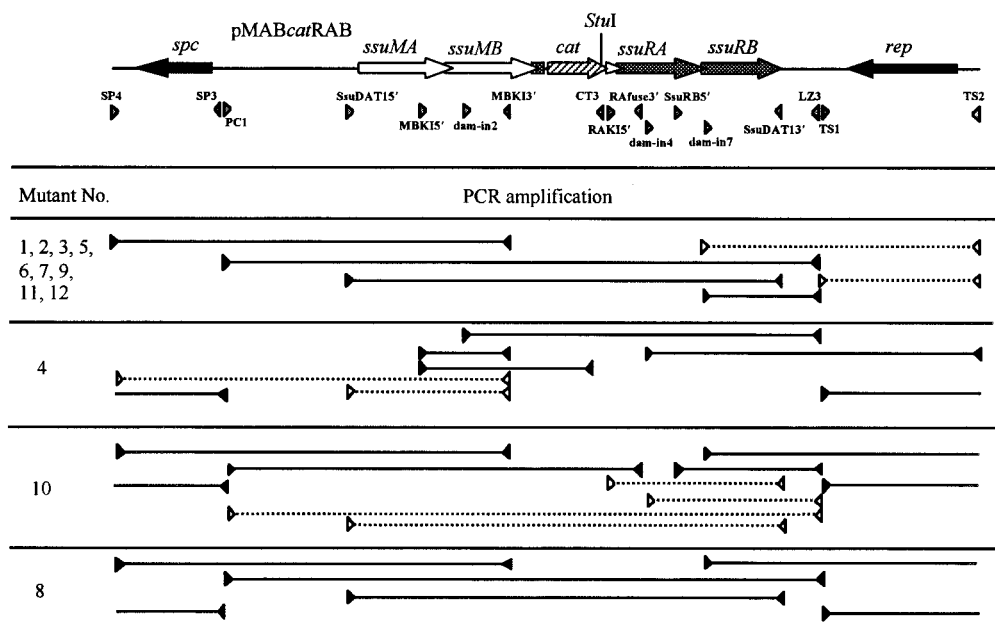


FIG. 6. Schematic representations of the results of systematic PCR analyses of temperature-resistant  $Cm^r$  mutants generated by integration of pMABcatRAB into the genome. Positions of primers relative to the genetic organization of pMABcatRAB are indicated at the top by gray arrowheads. Solid lines between the filled arrowheads, regions amplified by PCR; dashed lines between the open arrowheads, regions not amplified by PCR.

appeared on the selective agar. However, the results described above indicated that almost all of the mutants were generated by Campbell-type recombination. Therefore, we selected the temperature-resistant mutants on TH agar containing CAM and SPC in order to calculate the integration frequency. The frequency was estimated to be  $5.6 \times 10^{-6}$  to  $1.5 \times 10^{-7}$ , depending on the experiment.

**Nucleotide sequences of junctions between plasmid and chromosome.** As the integration sites of the plasmid in the mutants except Int8 could be estimated by PCR analyses described above, the junction regions were then amplified by inverse PCR and directly sequenced. The sequence information enabled us to amplify the target in the original host chromosome of serotype 2, and the amplified segments were directly sequenced. None of the integration target sites were localized in *purHD*; rather, all were in the vicinity of or within

other miscellaneous genes, which were apparently different from purine biosynthetic genes (Table 3).

As shown in Fig. 7, the sequences of the junction points between the plasmid and the host chromosome revealed small regions of identity between the two molecules (Fig. 7). These ranged from 2 to 10 nucleotides in length. Additional short regions of identity were also present around the junction points, as was seen in the border regions in the *purHD* locus. There was no consensus sequence discernible among the junction points, although six of them (Int1, Int3, Int5, Int9, Int11, and Int12) contained the tetranucleotide TTGG. The recognition sequence of SsuDAT1I ( $5'$ -GATC- $3'$ ) was seen only in the junction point of Int4. Of the 11 mutants examined, 9 of them, i.e., Int2, Int3, Int4, Int5, Int7, Int9, Int10, Int11, and Int12, showed complete identity in the junction sequences. However, the other two, Int1 and Int6, showed one or two nucleotide

TABLE 3. Predicted proteins which exhibited similarity to the putative translational products of the target regions in the mutants

Mutant	Description or function of closest relative <sup>a</sup>	Similarity	
		BLAST E value	Amino acid identity <sup>b</sup>
Int1	Putative prolyl-tRNA synthetase of <i>Streptococcus pneumoniae</i> [AE007339; AAK74442] (7)	5e-43	81/107/119
Int2	Surface immunogenic protein of <i>Streptococcus agalactiae</i> (7) [AF151362; AAG18478]	6e-23	52/98/118
Int3	Putative glutathione reductase of <i>S. pneumoniae</i> [AE008446; AAK99496]	1e-121	71/305/305
Int4	Putative glycosyl hydrolase family 3 of <i>S. agalactiae</i> [AE014223; AAM99592]	0.0	74/594/596
Int5	Putative <i>lacI</i> family regulator YvdE of <i>Bacillus subtilis</i> [Z99121; CAB15468]	3e-23	42/144/146
Int6	Putative 23S rRNA methyltransferase of <i>S. pneumoniae</i> [AE008555; AAL00715]	2e-63	50/244/249
Int7	Extracellular protein of <i>S. suis</i> [AY341262; AAQ19848]	5e-82	84/194/195
Int9	Putative ABC transporter permease protein of <i>Lactococcus lactis</i> [AE006342; AAK05202]	3e-77	41/403/397
Int10	Putative <i>N</i> -acetylneuramic acid synthetase NeuB of <i>S. agalactiae</i> [AF355776; AEAAK43615]	1e-158	80/338/338
Int11	Putative ABC transporter permease of <i>S. pneumoniae</i> [AE008562; AAL00778]	1e-113	82/233/234
Int12	Putative ribosomal protein L11 methyltransferase of <i>S. agalactiae</i> [AE014280; AAN00829]	1e-133	85/275/275

<sup>a</sup> Numbers in brackets are accession numbers and protein identification numbers from the GenBank/DBJ/EMBL database.

<sup>b</sup> Percent amino acid identity/number of amino acids evaluated/number of amino acids queried.



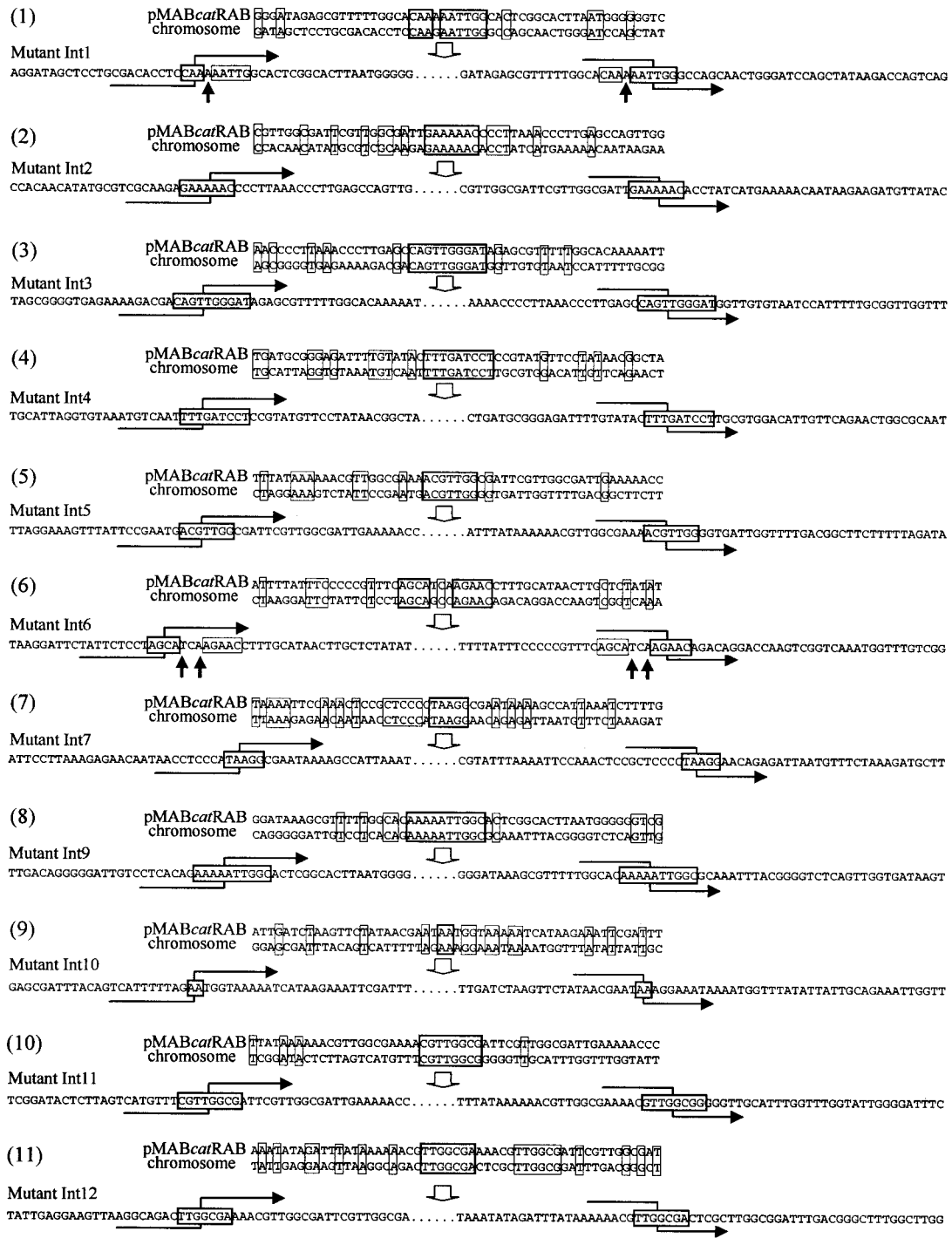


FIG. 7. Sequences of the junction points between pMABcatRAB and the *S. suis* genome. The sequences of pMABcatRAB and the corresponding *S. suis* genomic sequences are shown in the upper part of each panel. The sequence of the mutant is shown in the lower part of each panel. Stretches of sequence identity that are presumed to be the junction points are boxed with thick lines. Other positions with nucleotide identity are boxed with thin lines. Hooked arrows intercrossing the junction points of the mutants indicate possible positions of crossover, showing the recombination flow starting from the chromosome and shifting to the plasmid (left) and from the plasmid to the chromosome (right). Nucleotide mismatches found in the junction regions of Int1 and Int6 are indicated by upward-pointing arrows.

mismatches in the junction points when the plasmid and chromosome were compared. The two mutants had nucleotides that originated from the plasmid in both the junction points (Fig. 7, indicated by upward-pointing arrows), which indicated that strand crossover had taken place at different positions.

DISCUSSION

We have demonstrated here the presence of novel extra genetic regions in the *purHD* locus of some serotype reference strains of *S. suis*. The extra genetic regions, which were cate-

gorized into seven classes, showed no similarity with those of other classes and thus were unique to the strain(s) of each group. The RM genes are also present uniquely in the *purH-purD* regions of some strains (37), and constitute an eighth class of sequences there; it is thus evident that the eight classes of extra regions are mutually exclusive. The conserved genetic organization of the *pur* gene cluster (38), as has been observed with other gram-positive bacteria (4, 13, 44), together with the low G+C contents of the extra genetic regions compared with the *pur* genes, support the notion that the novel genetic regions found in this study were also transferred from foreign sources. Moreover, the fact that some of the phylogenetically unrelated serotypes share common genes indicates that the foreign genes have also spread among the *S. suis* strains, as has been described for the RM genes and *sly* regions (37, 38, 43).

Nearly half of the deduced translational products encoded by the novel foreign genes showed no similarity to proteins in the database; however, it is noteworthy that genes of other bacteria encoding the proteins which showed similarity to those of serotypes 16 and 25 are also contiguous in the genomes of the counterparts, suggesting that the *sepS16ABC* and *sepS25AB* gene clusters were both incorporated in a block. Moreover, the deduced SepS16ABC proteins showed similarity to three hypothetical proteins described for at least two *S. aureus* strains; however, the three proteins of *S. suis* showed no similarity to any proteins of other *S. aureus* strains, especially of those whose complete genome sequences have been determined (our unpublished observation), suggesting that the *S. aureus* genes encoding the hypothetical proteins were also acquired from foreign sources. As discussed below, these extra genetic regions were not only inserted independently into the genome of different strains but also replaced by a newcomer. This finding suggests that some of the foreign genes were successively transferred from one bacterial genome to the other.

The RM gene region has been considered simply to be inserted between *purH* and *purD* on the basis of the comparison between only two classes of *purHD* regions, i.e., one possessing the RM genes and the other lacking the genes, such as serotype 2 (37, 38). In the present study, however, we obtained information on seven additional classes of foreign genes that showed striking sequence variation even in the borders between the exogenous regions and the *pur* genes. Genetic borders of each exogenous region as estimated from the alignment of all of these sequences indicated that these genes were incorporated into different nucleotide positions of the same locus. Furthermore, some of them were likely to have been substituted for predecessors that had been inserted previously. One-to-one comparison of the sequences clearly showed differences in the border positions where the recombination may have occurred. Nucleotides with a few base pairs of identity scattered throughout the junction regions were observed next to the highly conserved *pur* genes. The identical sequences ranged from 1 to at least 6 nucleotides in length, which is reminiscent of illegitimate recombination sites described in other studies (1, 11, 23, 26, 29–31, 35, 39). The low degree of sequence similarity in the junction regions was also observed in the flanking sequences of *sly* and *orf102* in *S. suis* (43). From these observations, together with the fact that no vestige of sequences affecting their integration, such as long-repeat sequences or remnants of transposable elements, were found in

the close vicinity, we can consider that the foreign genes that have thus far been found in *S. suis* were integrated by the same mechanism by using illegitimate recombination.

Once a foreign gene is integrated in the genome, it can be transferred among *S. suis* by homologous recombination en bloc along with the conserved flanking regions (37, 43) or by means of homology-facilitated illegitimate recombination (11, 30, 35), and accumulation of the foreign genes may contribute to the genetic heterogeneity in *S. suis*. However, in the original insertion, the foreign genes found in this study must have been integrated by illegitimate recombination. Therefore, one can pose the following question: why are there so many different foreign genes at the same chromosomal locus? One possible explanation is that the intergenic space of *purHD* involves a hot spot for the type of recombination that is induced by these foreign genes. Indeed, at least one class of the foreign genes, RM genes, is proposed to comprise mobile genetic elements that can migrate among bacteria of different genera (25). The R enzyme can elicit a double-strand breakage, and this breakage may induce illegitimate recombination when the two genetic segments have a few base pairs of nucleotide identity (27).

To assess the validity of the above explanation, we constructed a thermosensitive suicide plasmid that carried the *SsuDATII* genes together with a *cat* gene, and we obtained *S. suis* mutants in which the plasmid and the host genome formed cointegrates. None of the 243 mutants examined formed cointegrates via the *purHD* region. If the integration events occurred uniformly all over the genome, one should be able to find a certain integration site among 500 mutants by a PCR spanning 4 kb, since the genome size of this bacterium is estimated to be 1,800 to 2,000 kb (13, 44). On the other hand, if a particular region, e.g., the *purH-purD* region, is the hot spot for the recombination and if the integration events occur five times more frequently there than at other regions, one should be able to find the integration site among 100 mutants by PCR. From this rationale, our results indicated that even if the *purHD* region was the hot spot for recombination, the frequency of the event was not high (less than 2.1 times higher than at other regions). Therefore, the *purHD* region was not likely to be the hot spot for recombination.

Although a limited number of integration mutants were sequenced, no mutant that had obviously integrated the plasmid via mobile genetic elements, such as a transposon, insertion sequence, integron, or intron, was obtained. This finding implies that the integration had occurred by illegitimate recombination irrespective of the mobile genetic elements. Integration sites localized on the plasmid pMAB*cat*RAB appeared clearly nonrandom, as 9 out of 12 mutants formed cointegrates via the *rep* region of the plasmid. The *rep* region of the plasmid usually forms a single strand as a replication intermediate so that this region is likely to form a Holliday junction between the plasmid and the genome, leading to the recombination event, or, alternatively, the *rep* region of the plasmid shares more sequence similarity with the genome than other segments of the plasmid. However, the plasmid was integrated into the genome of the 12 sequenced mutants at apparently different regions. Although we cannot rule out the possibility that there was something different about the events that occurred during the evolution of the different serotypes and the experimental integration described in this study, our results imply that the

*purHD* region did not show any preference for integrating particular foreign genes, i.e., the foreign genes found in this study were incidentally integrated into the same locus of the genome among the different strains by illegitimate recombination, presumably via a few base pairs of nucleotide identity.

The structures of the eight classes of foreign genes identified here may have represented the results of several consecutive recombination events, and therefore it is not possible to define the extra genetic segments as putative plasmids, phages, or conjugative transposons or to predict the exact limits of these inserted elements. Nevertheless, the present results together with previous findings (37, 38, 43) suggest that foreign gene acquisitions may have occurred by the same mechanism at various chromosomal loci. The fact that different foreign genes were incidentally integrated into the same locus of the *S. suis* genome and/or exchanged among the species by legitimate or illegitimate recombination leads us to propose a hypothesis to explain the presence of various foreign genes in the same locus, i.e., a high rate of gene acquisition in this bacterium. Actually, the integration frequency determined in this study was a little higher than those observed elsewhere (18, 21). However, the frequency may have been overestimated by elimination of plasmid-cured cells caused by postsegregational killing (25). If the foreign gene did not induce postsegregational killing, the illegitimate recombination may have occurred randomly at a frequency as low as  $10^{-7}$  to  $10^{-9}$  (18, 21). Therefore, it is conceivable that *S. suis* has the ability to take a foreign DNA into the genome through some mechanism, such as natural competence. *S. suis* is not known to enter a state of competence in vitro; however, it is possible that some isolates may, under certain growth conditions, be competent and/or that the ancestors of the *S. suis* populations were competent.

Whichever mechanisms are employed for transporting the DNA into the cytoplasm, the present results indicate that a seemingly large number of integration events have occurred in the *purHD* region, perhaps by an intrinsic recombination mechanism of the host bacteria. Most of the foreign genes found in the *purHD* locus are functionally unknown, and not all of the genes may contribute directly to the host survival. Integrated genes that are not deleterious will become fixed within the bacterial population and may contribute to the evolution of this bacterium.

#### ACKNOWLEDGMENTS

We thank Toshio Fujisawa for preparing photographs and Sachio Manabe and Mitoyo Takahashi for technical assistance.

This work was supported in part by a Grant-in-Aid for the Pioneer Research Project to T.S. from the Ministry of Agriculture, Forestry, and Fisheries (MAFF), Japan, and by a Grant-in-Aid for the Zoonosis Research Project from MAFF.

#### REFERENCES

- Albertini, A. M., M. Hofer, M. P. Calos, and J. H. Miller. 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* **29**:319–328.
- Allgood, N. D., and T. J. Shilhavy. 1988. Illegitimate recombination in bacteria, p. 309–330. In R. Kucherlapati and G. R. Smith (ed.), *Genetic recombination*. American Society for Microbiology, Washington, D.C.
- Allgood, N. D., and T. J. Shilhavy. 1991. *Escherichia coli xonA (sbcB)* mutants enhance illegitimate recombination. *Genetics* **127**:671–680.
- 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.
- Arends, J. P., and H. C. Zanen. 1988. Meningitis caused by *Streptococcus suis* in humans. *Rev. Infect. Dis.* **10**:131–137.
- Brake, A. J., A. V. Fowler, I. Zabin, J. Kania, and B. Müller-Hill. 1978.  $\beta$ -Galactosidase chimeras: primary structure of a *lac* repressor- $\beta$ -galactosidase protein. *Proc. Natl. Acad. Sci. USA* **75**:4824–4827.
- Brodeur, B. R., M. Boyer, I. Charlebois, J. Hamel, F. Couture, C. R. Rioux, and D. Martin. 2000. Identification of group B streptococcal Sip protein, which elicits cross-protective immunity. *Infect. Immun.* **68**:5610–5618.
- Casadaban, M. J., and S. N. Cohen. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* **138**:179–207.
- Chatellier, S., J. Harel, Y. Zhang, M. Gottschalk, R. Higgins, L. A. Devriese, and R. Brousseau. 1998. Phylogenetic diversity of *Streptococcus suis* strains of various serotypes as revealed by 16S rRNA gene sequence comparison. *Int. J. Sys. Bacteriol.* **48**:581–589.
- Clifton-Hadley, F. A. 1983. *Streptococcus suis* type 2 infections. *Br. Vet. J.* **139**:1–5.
- de Vries, J., and W. Wackernagel. 2002. Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc. Natl. Acad. Sci. USA* **99**:2094–2099.
- Feder, I., M. M. Chengappa, B. Fenwick, M. Rider, and J. Staats. 1994. Partial characterization of *Streptococcus suis* type 2 hemolysin. *J. Clin. Microbiol.* **32**:1256–1260.
- Ferretti, J. J., W. M. McShan, D. Ajdic, D. J. Savic, G. Savic, K. Lyon, C. Primeaux, S. Sezate, A. N. Suvorov, S. Kenton, H. S. Lai, S. P. Lin, Y. Qian, H. G. Jia, F. Z. Najjar, Q. Ren, H. Zhu, L. Song, J. White, X. Yuan, S. W. Clifton, B. A. Roe, and R. McLaughlin. 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl. Acad. Sci. USA* **98**:4658–4663.
- Franklin, N. C. 1971. Illegitimate recombination, p. 175–194. In A. D. Hershey (ed.), *The bacteriophage lambda*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Gottschalk, M., A. Lebrun, H. Wisselink, J. D. Dubreuil, H. Smith, and U. Vecht. 1998. Production of virulence-related proteins by Canadian strains of *Streptococcus suis* capsular type 2. *Can. J. Vet. Res.* **62**:75–79.
- 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.
- 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.
- Hanada, K., T. Ukita, Y. Kohno, K. Saito, J. Kato, and H. Ikeda. 1997. RecQ DNA helicase is a suppressor of illegitimate recombination in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **94**:3860–3865.
- Harel, J., R. Higgins, M. Gottschalk, and M. Bigras-Poulin. 1994. Genomic relatedness among reference strains of different *Streptococcus suis* serotypes. *Can. J. Vet. Res.* **58**:259–262.
- Higgins, R., M. Gottschalk, M. Boudreau, A. Lebrun, and J. Henrichsen. 1995. Description of six new capsular types (29–34) of *Streptococcus suis*. *J. Vet. Diagn. Investig.* **7**:405–406.
- Ikeda, H., K. Moriya, and T. Matsumoto. 1981. In vitro study of illegitimate recombination: involvement of DNA gyrase. *Cold Spring Harbor Symp. Quant. Biol.* **45**:399–408.
- Jacobs, A. A. C., P. L. W. Loeffen, A. J. G. van den Berg, and P. K. Storm. 1994. Identification, purification, and characterization of a thiol-activated hemolysin (suilysin) of *Streptococcus suis*. *Infect. Immun.* **62**:1742–1748.
- Jones, I. M., S. B. Primrose, and S. D. Ehrlich. 1982. Recombination between short direct repeats in a *recA* host. *Mol. Gen. Genet.* **188**:486–489.
- Kilpper-Bälz, R., and K. H. Schleifer. 1987. *Streptococcus suis* sp. nov., nom. rev. *Int. J. Sys. Bacteriol.* **37**:160–162.
- Kobayashi, I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.* **29**:3742–3756.
- Kumagai, M., and H. Ikeda. 1991. Molecular analysis of the recombination junctions of  $\lambda$ bio transducing phages. *Mol. Gen. Genet.* **230**:60–64.
- 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.
- 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.
- Marvo, S. L., S. R. King, and S. R. Jaskunas. 1983. Role of short regions of homology in intermolecular illegitimate recombination events. *Proc. Natl. Acad. Sci. USA* **80**:2452–2456.
- Meier, P., and W. Wackernagel. 2003. Mechanisms of homology-facilitated illegitimate recombination for foreign DNA acquisition in transformable *Pseudomonas stutzeri*. *Mol. Microbiol.* **48**:1107–1118.
- Meima, R., B. J. Haijema, H. Dijkstra, G.-J. Haan, G. Venema, and S. Bron. 1997. Role of enzymes of homologous recombination in illegitimate plasmid recombination in *Bacillus subtilis*. *J. Bacteriol.* **179**:1219–1229.
- Okumabua, O., J. Staats, and M. M. Chengappa. 1995. Detection of genomic heterogeneity in *Streptococcus suis* isolates by DNA restriction

- fragment length polymorphisms of rRNA genes (ribotyping). *J. Clin. Microbiol.* **33**:968–972.
33. Osaki, M., D. Takamatsu, Y. Shimoji, and T. Sekizaki. 2002. Characterization of *Streptococcus suis* genes encoding proteins homologous to sortase of gram-positive bacteria. *J. Bacteriol.* **184**:971–982.
  34. Perch, B., K. B. Pedersen, and J. Henriksen. 1983. Serology of capsulated streptococci pathogenic for pigs: six new serotypes of *Streptococcus suis*. *J. Clin. Microbiol.* **17**:993–996.
  35. Prudhomme, M., V. Libante, and J.-P. Claverys. 2002. Homologous recombination at the border: insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA* **99**:2100–2105.
  36. 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.
  37. Sekizaki, T., M. Osaki, D. Takamatsu, and Y. Shimoji. 2001. Distribution of the *Ssu*DATII restriction-modification system among different serotypes of *Streptococcus suis*. *J. Bacteriol.* **183**:5436–5440.
  38. Sekizaki, T., Y. Otani, M. Osaki, D. Takamatsu, and Y. Shimoji. 2001. Evidence for horizontal transfer of *Ssu*DATII restriction-modification genes to the *Streptococcus suis* genome. *J. Bacteriol.* **183**:500–511.
  39. Shimizu, H., H. Yamaguchi, Y. Ashizawa, Y. Kohno, M. Asami, J. Kato, and H. Ikeda. 1997. Short-homology-independent illegitimate recombination in *Escherichia coli*: distinct mechanism from short-homology-dependent illegitimate recombination. *J. Mol. Biol.* **266**:297–305.
  40. 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.
  41. Takamatsu, D., M. Osaki, and T. Sekizaki. 2001. Construction and characterization of *Streptococcus suis*-*Escherichia coli* shuttle cloning vectors. *Plasmid* **45**:101–113.
  42. Takamatsu, D., M. Osaki, and T. Sekizaki. 2001. Thermosensitive suicide vectors for gene replacement in *Streptococcus suis*. *Plasmid* **46**:140–148.
  43. Takamatsu, D., M. Osaki, and T. Sekizaki. 2002. Evidence for lateral transfer of the *suilysin* gene region of *Streptococcus suis*. *J. Bacteriol.* **184**:2050–2057.
  44. Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**:498–506.