Site-Specific Recombinase Genes in Three Shigella Subgroups and Nucleotide Sequences of a pinB Gene and an Invertible B Segment from Shigella boydii

AKIRA TOMINAGA, SHOICHI IKEMIZU, AND MASATOSHI ENOMOTO

Department of Biology, Faculty of Science, Okayama University, Okayama 700, Japan

Received 20 November 1990/Accepted 20 April 1991

Inversional switching systems in procaryotes are composed of an invertible DNA segment and a site-specific recombinase gene adjacent to or contained in the segment. Four related but functionally distinct systems have previously been characterized in detail: the Salmonella typhimurium H segment-hin gene (H-hin), phage Mu G-gin, phage PI C-cin, and Escherichia coli e14 P-pin. In this article we report the isolation and characterization of three new recombinase genes: pinB, pinD, and defective pinF from Shigella boydii, Shigella dysenteriae, and Shigella flexneri, respectively. The genes pinB and pinD were detected by the complementation of a hin mutation of Salmonella and were able to mediate inversion of the H, P, and C segments. pinB mediated H inversion as efficiently as the hin gene did and mediated C inversion with a frequency three orders of magnitude lower than that of the cin gene. pinD mediated inversion of H and P segments with frequencies ten times as high as those for the genes intrinsic to each segment and mediated C inversion with a frequency ten times lower than that for cin. Therefore, the pinB and pinD genes were inferred to be different from each other. The invertible B segment-pinB gene cloned from S. boydii is highly homologous to the G-gin in size, organization, and nucleotide sequence of open reading frames, but the S’ constant region outside the segment is quite different in size and predicted amino acid sequence. The B segment underwent inversion in the presence of hin, pin, or cin. The defective pinF gene is suggested to have the same origin as P-pin on e14 by the restriction map of the fragment cloned from a Pin+ transductant that was obtained in transduction from S. flexneri to E. coli Δpin.

Inversional switching systems in procaryotes, which mediate alternative expression of two sets of genes, are composed of an invertible DNA segment and a site-specific recombinase gene adjacent to or contained in the segment. To date, four systems have been characterized in detail: the Salmonella typhimurium H segment-hin gene (H-hin) (44, 47, 53) and phage Mu G-gin (26), phage PI C-cin (15, 18, 20), and Escherichia coli e14 P-pin (36, 38) systems. In S. typhimurium, H inversion switches expression of phase 2 flagellin and repressor genes, and in both phages inversion of the segments causes expression of another set of tail fiber genes. Any phenotypic change caused by P inversion is unknown. These systems are different not only in function but in structure. The sizes of the invertible segments are 1, 1.8, 3, and 4.2 kb in the H, P, G, and C segments, respectively. The positions and orientations of recombinase genes are also different from one another with respect to the invertible segment, although the G-gin and P-pin systems are similar in organization (13). Despite these differences, nucleotide sequences of inv sites where recombination takes place share high homology among the systems (14), and the recombinase genes are substitutable among them (24, 27, 31, 50). These findings together with the relationship of the resolution system in transposon Tn3 to the switching system (45) have led to the idea that these switching systems have evolved from some composite structures in which recombinase genes have a common origin and in which the genes of the associated set(s) have an origin distinct from one another (24, 25, 43). One of the approaches to developing the idea would be to detect and analyze more primitive switching systems which may have imperfect invertible segments or may consist only of a recombinase gene with the inv sites at its ends. Assuming that site-specific recombinases are usually substitutable for one another, it seems possible to detect a new recombinase gene from other organisms when inversion of a known segment can be used as a selection method. The present report deals with the detection and characterization of site-specific recombinase genes from strains of the genus Shigella. Genes from Shigella boydii and Shigella dysenteriae mediate inversion of the known invertible segments with different efficiencies. The recombinase gene and its adjacent invertible segment were cloned from S. boydii and sequenced; we find that they highly resemble the G segment-gin system of phage Mu in structure and nucleotide sequence.

MATERIALS AND METHODS

Bacterial strains, plagues, and tester plasmids. Four strains representing each of the four Shigella subgroups (40) were used: S. dysenteriae Sh16 (42), S. boydii C12, Shigella flexneri IID642, and Shigella sonnei IID969 (48). All strains except Sh16 were kindly supplied by the Medical School, Okayama University. S. typhimurium SL4273 (11) was used as a hin+ control. Strains EJ1449 (31) and SLA4266 (11) were used as DNA donors for cloning a hin gene and cloning a fragment which contained the fliB (phase 2 flagellin) and mutant hin genes, respectively. The main E. coli K-12 strains used were as follows. W3110 and EJ2518, a pin+ derivative of MC1061 [hsdR DelacI-A]X74 (5), were used as pin+ strains. K802 (hsdR) (52) was used as an indicator for PIC(+) and also as a parent of the Δpin ΔflfC (flagellin gene) strains. EJ2282 (ΔflfC purB51 Δpin recA56 hsdR), a host for

* Corresponding author.
recombinant plasmids, was constructed from K802 by the successive transduction of flic::Tn10 (10), purB51-Δpin-zcg2::Tn10 (9), and recA56-Δr1A::Tn10 (7), with Tn10 excision (28) after each transduction. Similarly, Ezj2517 (ΔflIC purB51 Δpin zcg2::Tn10 hasdR) was made from K802 by the transduction of flic::Tn10 followed by Tn10 excision and of purB51-Δpin-zcg2::Tn10.

P1vir and P1cinC(+) were used for transduction. A P1cin mutant (49) from thermosensitive P1cmch100 (39) has a C segment fixed in either the C(+) or C(−) orientation. P1cinC(−) was used for the measurement of C inversion. To isolate a P1cinC(−) lysogen from S. dysenteriae that was resistant to P1C(−) (48), first a P1cinC(+) lysogen was made and then a C(−) variant was chosen by testing a number of colonies of P1C(+) lysogens for their inability to lyse indicator bacteria (K802) at 42°C. Muts (49) was used as probe for hybridization.

Plasmid pTY109, used for the measurement of H fragment inversion, is a pBR322 (2) derivative which carries a 6.75-kb SalI fragment [hin fjf(Off)] from strain SL2466 [fjf(On)]. First a plasmid carrying a fjf(On) fragment was isolated and transformed into strain Ezj2282 with a cin· plasmid (pS1730). Then, the fjf(Off) plasmid made by H inversion due to the cin gene was isolated after transformation into strain Ezj2282 (ΔflIC). Plasmid pTY101 is a derivative of the mini F plasmid pTN1105 (32) and carries the same SalI fragment as pTY109. Plasmid pPZ202 [lacZ(Off)], used for the measurement of P inversion, was constructed as follows. A 2.8-kb BamHI-BglII fragment (38) containing the P segment of e14 was subcloned from a P-pin plasmid (pHA201; Fig. 1C) into the BamHI site of pACYC184 (6). Then, a 3-kb PstI fragment containing the lacZ gene lacking P lac was inserted from plasmid pMC1871 (5) in frame into the PstI site of the major open reading frame (ORF) (38) of the cloned P segment.

Moreover, the promoter of a tetracycline resistance (Tet') gene on the vector, which can be used for lacZ expression when the P segment undergoes inversion, was deleted by BAL 31 digestion at a HindIII site in the promoter sequence to yield plasmid pPZ102 [lacZ(On)]. The P segment of pPZ102 was inverted by using another cin· plasmid (pAT101) (49) derived from pBR322 to yield pPZ202, which was determined to have the P segment fixed in the (+) orientation. The β-galactosidase activity expressed by pPZ202 was compared to that in strain MC1061 that was approximately 20 and 350 Miller units (30), respectively. The two plasmids differed in EcoRI digestion patterns because of the presence of one EcoRI site in the chloramphenicol acetyltransferase gene on the vector (6) and two EcoRI sites in the lacZ gene (5). pS1710, derived from pACYC184, carries a 3.3-kb HindIII fragment which contains a pin gene from an e14 variant (EJ1076) (9) in which a new HindIII site had been created in the P segment (unpublished data). pS1730 is a pACYC177 (6) derivative which carries a 3-kb BamHI-HindIII fragment (49) that contains the cin gene of phage P1. pS1743, a pACYC177 derivative, carries a 4.5-kb PstI fragment (47) which contains the cin gene from strain EJ1449.

Media. TLY broth, nutrient agar, and nutrient semisolid agar (NSS) were as described previously (9). Chloramphenicol (Sigma Chemical Co.), kanamycin (Meiji Seika Co.), tetracycline (Sigma), and ampicillin (Sigma) were used at concentrations of 12.5, 50, 25, and 50 mg/liter, respectively. X-Gal plates contained 40 mg of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (Sigma) per liter. Anti-flagellum serum against an e, n, x antigen was as previously described (31).

DNA manipulations. Total host DNA was prepared by the method described previously (1), except a solution of 1% sodium dodecyl sulfate (SDS) was used instead of the alkaline-SDS solution. Isolation of plasmid DNA and cloning procedures were described previously (29). Restriction endonucleases, T4 DNA ligase, and nucleases BAL 31 were purchased from Takara Shuzo Co. and used as recommended. Restriction fragments were examined by agarose gel electrophoresis with 1% agarose (Takara) in 40 mM Tris-acetate (pH 8.0)–1 mM EDTA buffer as previously described (29). Nucleotide sequences were determined by the dideoxy chain-termination method of Sanger et al. (41) using a sequencing kit (Takara) and [35S]dCTP (Amersham International; 400 Ci/mol). Fragments subcloned from plasmid pTSB916 into pUC118 and pUC119 (51) were dephosphorylated in opposite directions by using exonuclease III and mung-bean nuclease (Takara), and a series of deletion plasmids was used for sequencing. Nucleotide and amino acid sequences were analyzed by using DNASTAR sequence analysis software (Hitachi Software Engineering Co.). A restriction fragment, which was used as a probe and had been electroeluted from an agarose gel, and whole Mu DNA, prepared as described previously (49), were labeled by the random primed method (12) with digoxigenin-labeled dUTP and by using the nonradioactive DNA labeling kit (Boehringer Mannheim). For hybridization (46), total bacterial DNA which had been electrophoresed after digestion with several restriction enzymes was transferred to a nylon membrane filter (Amersham Hybond-N) by vacuum blotting (34) and was fixed by UV irradiation. Hybridization was done by the standard method (29) and detection of hybrid bands was performed by the digoxigenin-enzyme-linked immunosorbent assay Dig-ELISA method using the Detection kit (Boehringer Mannheim).

Detection of recombinase gene activity in transduction. E. coli K-12 strain EJ2282 (ΔflIC Δpin recA) with plasmid pTY109 was used as the recipient for transduction. Since the H segment of this plasmid is fixed in the “off” orientation by the hin mutation, this recipient strain is nonmotile. The presence of recombinase genes capable of complementing the hin mutation is detected by restoration of fliB expression. P1cinC(+) grown on each donor was mixed with 1 ml of broth culture of strain EJ2282 with pTY109 [Ap' hin fjf(Off)] at a multiplicity of infection of around 0.5 and was then incubated at 37°C for 20 min for adsorption. After centrifugation, about half of the mixture was streaked as lines on each of the two NSS plates either containing ampicillin or containing antiflagellum serum and ampicillin and then incubated at 30°C for 15 to 20 h.

Measurement of inversion rates of H, P, and C segments. In order to examine the rate of H inversion, a single colony from each strain transformed with pTY109 was cultured in broth for more than 100 generations by successive inoculation, and the plasmid isolated after 30, 50, or 100 generations (rough estimate) was transformed into EJ2282 (ΔflIC). The Ap' transformants obtained were tested for motility by stabbing them into NSS medium containing ampicillin.

P inversion was examined by a change in color on X-Gal plates of the transformants with plasmid pPZ202 (lacZ(OF)) in which expression of lacZ is controlled by the orientation of the P segment. A single colony from each strain transformed with pPZ202 [cm' lacZ(OF)] was cultured in broth at 37°C for about 30 generations. The culture was diluted and plated on X-Gal plates containing chloramphenicol, and the number of blue colonies was counted after overnight incubation at 37°C.

To examine the rate of C inversion, a P1cinC(−) lysogen from each strain was cultured in broth at 30°C, and the
occurrence of PIC(+) lysogens was measured at 42°C. After about 30 generations the number of bacteria capable of being infective centers because of the change to P1cinC(+) was counted by spotting aliquots of the dilutions on the lawn of indicator bacteria (K802) and incubating them overnight at 42°C. The number of viable bacteria of the same culture was counted at 30°C.

Cloning of the pin gene. Nonmotile E. coli (ΔfljC Δpin recA) containing the plasmid pTY109 [fljB(Off)] was used as the host for cloning the pin gene. BamHI fragments from bacterial DNA of each strain were inserted into the BamHI site of pBR322 and transformed into E. coli. Recombinant plasmids which had conferred motility on the recipient were isolated and digested with several restriction enzymes, and restriction maps were constructed.

Nucleotide sequence accession number. The nucleotide sequence data reported here will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases under accession number D00660.

RESULTS

Detection of functional equivalents of a S. typhimurium hin gene from Shigella spp. To search Shigella spp. for the presence of new switching systems, transduction by P1cinC(+) was carried out from four Shigella strains representing different subgroups into nonmotile E. coli K-12 strain EJ2282 (ΔfljC Δpin recA) with plasmid pTY109 [hin fliB (Off)]. The presence of recombinase genes capable of complementing the hin mutation was inferred from restoration of fliB expression (Table 1). S. dysenteriae, S. boydii, and E. coli W3110, used as a pin+ control, produced swarms (motile colonies) on the plain NSS medium but not on the medium containing anti-e,n,x serum, indicating that silent fliB (antigen e,n,x) is expressed on pTY109 by H inversion from the off to the on orientation and that its inversion is mediated by a hin+ equivalent abortively transduced from S. dysenteriae or S. boydii. Trail production (Table 1), which suggests the presence of cryptic flagellin genes in Shigella spp., will be described elsewhere.

To examine the possibility that the hin+ equivalents of the two strains are located on e14 (9) or an e14-like element on the chromosome (as in the case of P-pin), transduction was carried out by using the two strains as donors and strain EJ2517 (ΔfljC), which carries purB Δpin zcg2::Tn10 (in that order) as the recipient (9). Pur+ Tet+ transductants are expected to restore the ability for H inversion when a gene of interest is on the element. The donor S. dysenteriae yielded a number of Pur+ clones but none of them (60 clones) were Tet+ and the five clones transformed with pTY109 did not cause H inversion either. S. boydii yielded only five Pur+ clones by repeated transduction and all of them were Tet+ Hin. Transduction from S. sonnei and S. flexneri was also attempted. S. sonnei yielded 27 (45%) Pur+ Tet+ cotransductants among 60 Pur+ transductants, and all were Hin. S. flexneri yielded Pur+ Tet+ cotransductants at a frequency of 48% (39 of 81 Pur+ transductants); of the 10 cotransductants transformed with pTY109, 5 became motile due to H inversion, although a hin+ equivalent had not been detected in the previous transduction (Table 1). These results suggest that the hin+ equivalents of S. dysenteriae and S. boydii may be on e14 as in E. coli, that S. flexneri probably has defective pin gene at the same locus as E. coli, and that its defect can be repaired by recombination or another mechanism after transduction to E. coli Δpin. Henceforth, the defective gene of S. flexneri is designated pinF, and the pin gene of E. coli is designated pinE if necessary. Putative genes in S. boydii and S. dysenteriae are designated pinB and pinD, respectively.

Inversion rates of H, P, and C segments in S. boydii and S. dysenteriae. The genes pinB and pinD may be alike or different from each other in structure and function. To distinguish between pinB and pinD genes, inversion rates of H, P, and C segments were measured in S. boydii and S. dysenteriae and compared with the rate of the recombinase gene intrinsic to each segment.

The inversion rate of the H segment in S. dysenteriae was estimated to be 1.4 × 10−3, which is 10 times higher than those in S. boydii and in S. typhimurium, which was used as a control (Table 2).

TABLE 1. Occurrence of motile clones from a nonmotile E. coli strain with tester plasmid pTY109 in transduction from Shigella spp.*

<table>
<thead>
<tr>
<th>Donor</th>
<th>NSS medium containing:</th>
<th>No antisera</th>
<th>Anti-e,n,x serum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Swarm</td>
<td>Trail</td>
</tr>
<tr>
<td>S. dysenteriae</td>
<td></td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>S. boydii</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>S. flexneri</td>
<td></td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>S. sonnei</td>
<td></td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>E. coli W3110</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

* +, detection of swarms or trails; −, no detection. The number of swarms detected was 1 or 2 per one line streaked on NSS and that of trails was less than 10.

TABLE 2. Inversion rates of the plasmid-encoded H segment in S. dysenteriae and S. boydii

<table>
<thead>
<tr>
<th>Strain</th>
<th>No. of swarms/no. of Amp' transfectants testeda</th>
<th>Inversion ratea</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. dysenteriae</td>
<td>16/376</td>
<td>1.4 × 10−3</td>
</tr>
<tr>
<td>S. boydii</td>
<td>0/264</td>
<td>1.0 × 10−4</td>
</tr>
<tr>
<td>S. typhimurium</td>
<td>1/528</td>
<td>1.1 × 10−4</td>
</tr>
</tbody>
</table>

* a 30, 50, and 100 indicate number of generations.

a The inversion rate was expressed as a value (number of swarms per plasmid per generation) from H(−) to H(+) orientation during 30 generations.

TABLE 3. Inversion rates of the plasmid-encoded P segment in S. dysenteriae

<table>
<thead>
<tr>
<th>Strain</th>
<th>No. of blue colonies/no. of colonies examined</th>
<th>No. of generations</th>
<th>Inversion ratea</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. dysenteriae</td>
<td>1/199</td>
<td>30.9</td>
<td>1.6 × 10−4</td>
</tr>
<tr>
<td></td>
<td>2/182</td>
<td>30.8</td>
<td>3.6 × 10−4</td>
</tr>
<tr>
<td></td>
<td>1/116</td>
<td>30.1</td>
<td>2.9 × 10−4</td>
</tr>
<tr>
<td>E. coli EJ2518</td>
<td>1/1,090</td>
<td>30.0</td>
<td>3.1 × 10−5</td>
</tr>
</tbody>
</table>

* The inversion rate was expressed as a value (number of blue colonies per cell per generation) in inversion from the P(+) to the P(−) orientation. This will be reduced to at least 1/10 when expressed as a rate (number of blue colonies per plasmid per generation), assuming that 10 to 20 copies of the plasmid are maintained in a bacterium and one of them undergoes inversion.
The inversion rate of the P segment was measured only for *S. dysenteriae* and the *E. coli* pin+ strain EJ2518 (ΔlacZ) (Table 3), since the two types of *S. boydii* transformants obtained with pPZ202 [lacZ(Off)] or pPZ102 [lacZ(On)] were blue on the X-Gal plate and indistinguishable from each other. The rate of P inversion in *S. dysenteriae* was 1.6 × 10–4 to 3.6 × 10–4, which is ten times higher than the rate in the pin+ strain EJ2518. Plasmid DNA isolated from each blue colony was digested with EcoRI and electrophoresed to visualize inverted DNA fragments; the plasmids from all the blue colonies (total, 5) had digestion patterns identical to that of pPZ102 (data not shown). Although P inversion could not be tested in *S. boydii*, the activity of the pinB gene for P inversion was indicated by the change in the EcoRI digestion pattern of pPZ202, which had been propagated in the Δpin strain (EJ2282) with a pinB plasmid (pTSB916) (see below).

The mean inversion rates of the C segment in *S. boydii* and *S. dysenteriae* were 2.6 × 10–3 and 4.1 × 10–3, respectively (Table 4). Since the ratio from C(−) to C(+) by the *cin* gene has been reported to be 3.4 × 10–3 (31), it was determined that *S. boydii* and *S. dysenteriae* can mediate C inversion with frequencies three orders and one order of magnitude lower than that of the *cin* gene, respectively. A similar experiment was performed with *S. sonnei*, since our earlier study had suggested that this strain can mediate C inversion (48). It was found that the supernatant from a broth culture of *S. sonnei* contains some kind of colicin factor, which is able to make a very small lysis spot which is indistinguishable from a PlcinC(+) plaque on the indicator strain K802. When the bacteria lysogenized with PlcinC(−) were spotted on the indicator that had previously been made resistant to this colicin and were tested by shifting up to 42°C, no lysis spot or plaque could be detected, showing that the previous finding may be ascribed to the misjudgment of lysis spots of the colicin of *S. sonnei*.

These results showed that *S. boydii* mediates H inversion with a frequency similar to that of the *hin* gene and mediates C inversion with a frequency 1,000 times lower than that of the *cin* gene, while *S. dysenteriae* mediates inversion of the H and P segments with frequencies 10 times higher than those of the genes intrinsic to each segment and mediates inversion of the C segment with a frequency 10 times lower than that of the *cin* gene. Assuming that pinB or pinD is the only recombinase gene in each strain, the pinD gene is clearly different in activity from pinB and also different from *hin*, *pinE*, and *cin* genes.

The pin gene and its adjacent invertible segment from *S. boydii* and an *E. coli* transductant. Recombinant plasmid pTSB916, which contains a 5.9-kb BamHI fragment from *S. boydii*, was cloned and the restriction map was constructed (Fig. 1A). When the plasmid was digested with BamHI and *MluI*, six bands (4, 3.8, 3.3, 1.5, 1.1, and 1.0 kb) appeared, the total length of which was 10.7 kb, exclusive of a 4.3-kb vector fragment; this is longer than the cloned fragment by 4.8 kb. We inferred that the extra 4.8-kb fragment (3.8 + 1.0 or 3.3 + 1.5) resulted from a putative invertible segment, which was cloned together with the pinB gene and could have alternative orientation. It is reasonable to think that a *MluI* site is located asymmetrically in the segment since the BamHI sites are fixed as the cloning site and one *MluI* site is variable. When a 2-kb *SalI* fragment or a 3.1-kb *AvaI* fragment (Fig. 1A) was deleted, the deletion plasmid produced the 1.0- or 3.3-kb fragment by digestion with BamHI and *MluI* in addition to the large fragment containing the vector. This indicates that the pinB gene, or part of it, is located on these fragments and that its defect results in fixed orientation of the invertible segment. Therefore, the pinB gene was mapped on the right side of the invertible segment that contained the *MluI* site (Fig. 1A). The invertible segment, designated the B segment, was found to be 3 kb by sequencing this region, as described below. The arrangement of the B segment and pinB resembles those of *P-pinE*, *C-cin*, and *G-gin*, unless a promoter site for each recombinase gene is taken into account. However, the size of the B segment was clearly different from those of the P and C segments and similar to that of the G segment (3 kb).

The B segment was found to undergo inversion in the presence of the known recombinase genes, *pinE*, *cin*, and *hin*. Plasmid pTSB917 was made from pTSB916 by deleting the *SalI* fragment that contained a part of pinB (Fig. 1A); this was confirmed later by sequencing pinB. When pTSB917 coexisted with a plasmid carrying one of the three genes *pinE* (pS710), *cin* (pS7130), and *hin* (pS7143) in Δpin bacteria (EJ2282) during approximately 30 generations, pTSB917 produced two sets of fragments resulting from two orientations of the B segment after digestion with *MluI* and *SalI* (Fig. 2). This result, together with the finding that the pinB gene can mediate H, P, and C inversions, indicates that the B-pinB system operates well and is indeed a member of the family of the inversive switching systems represented by H-hin.

A 12.9-kb BamHI fragment containing a pin gene from one
of the Pin(+) transductants (EJ2553), which was derived from transduction from S. flexneri to an E. coli ΔpinE strain (EJ2517), was cloned and named pTSF213. Restriction maps of pTSF213 and pH201 (pHA201 carries a 8.0-kb BamHI fragment [3]) with wild-type P-pinE from strain EJ350 (9) were constructed (Fig. 1B and C). The locations of pin and inverted repeats (IRs) were inferred from the similarity to the restriction map of the P-pinE fragment reported previously (3, 38). Maps of pTSF213 and pH201 were identical with respect to a 4-kb BamHI-HindIII fragment containing P-pin but different in the 3'-flanking region. The ΔpinE strain (EJ2517) is supposed to have a deletion including the whole P segment and at least part of pinE, since the deletion was made from EJ350 (zcg3::Tn10) at the time of the excision of Tn10, which had been inserted between purB and P-pinE and another Tn10 (zcg2) which was inserted outside ΔpinE (3, 9). Therefore, S. flexneri is very likely to have an intact invertible segment identical to the P segment of E. coli in addition to a mutant pinF gene which is to be active after transduction to the ΔpinE strain.

Nucleotide sequences of the pinB gene and the B segment.

The nucleotide sequence of the 4.8-kb BamHI-MluI fragment at the B(+) orientation (Fig. 1A) was determined with both strands (Fig. 3A). Three complete and two incomplete open reading frames were revealed (Fig. 3B). The location of pinB was determined on the 3' side of the fragment from a predicted amino acid sequence which is highly homologous to those reported for several site-specific recombinases (15, 36, 53). It starts with GTG instead of ATG as in the gin gene (36), is preceded by a Shine-Dalgarno sequence and a

FIG. 2. Analysis of the invertible B segment by agarose gel electrophoresis. Plasmid pTSB917 carrying the B segment was propagated in Δpin strain EJ2282 with a plasmid carrying one of the three genes pinE, cin, and hin for about 30 generations in the presence of an antibiotic corresponding to a resistance marker encoded on each plasmid. The plasmids isolated were digested with Sall and MluI and electrophoresed. Lane 1, pTSB916 (Ap') carrying the B-pinB fragment; lane 2, pTSB917 (Ap') with the B segment fixed in the (+) orientation; lane 3, pTSB918 (Ap') with the B segment fixed in the (−) orientation; lane 4, pTSB917 and coresident pSiT70 (Cm' pinE); lane 5, pTSB917 and pSiT730 (Km' cin); lane 6, pTSB917 and pSiT743 (Km' hin). The two sets of fragments resulting from B inversion are shown by arrows on the right side of the panel. The length of each set is 8.2 kb.

SHIGELLA SITE-SPECIFIC RECOMBINASE GENES

Prigibow box, consists of 585 bp (position 3796 to 4380) and encodes a 195-amino-acid polypeptide with a molecular mass of 21,931 Da. It terminates with TAA followed by inverted repeats which were inferred to form a Rhö-independent transcriptional terminator, since they are G+C rich and are followed by poly(T)₉. The ORF can be truncated by Sall or AvaI, which explains why the deletion plasmids produced by these enzymes show the fixed orientation of the invertible segment. The presence of the B segment was inferred by the presence of two 38-bp IRs with only 3 bp differing between them. They contain a 26-bp inv consensus sequence (14), but 3 bp in the right IR sequence (IRR) and 4 bp in the left IR sequence (IRL) differ from the consensus. The occurrence of B inversion by recombination between the two IRs was verified by sequencing about 50 bp of the 3'-flanking region of IRL and 50 bp of the 5'-flanking region of IRR in the B(−) orientation. The B segment consists of 3,033 bp exclusive of the two IRs and contains the two complete internal ORFs, B175 and B177, which are transcribable in opposite directions. B175, which was transcribed toward pinB in the B(+) orientation, starts with ATG, consists of 525 bp (position 1679 to 2203, 175 amino acids), and terminates with TAA, and B177, which starts with ATG, consists of 531 bp (2770 to 2238, 177 amino acids), and terminates with TGA. These ORFs are followed by a common potential stem-loop structure, which would function as a transcriptional terminator for both ORFs. The other two ORFs are incomplete; one is on the 5' side of the cloned fragment and contains a 1,674-bp sequence which encodes 558 amino acids. The IRL within the sequence can separate it into a constant region (Bc; 1 to 654) and a variable region (Bv; 655 to 1674), provided that the crossover site is located within the central dinucleotide of the inv site (19, 21, 37). Another variable region (Bv'; 3727 to 2774), which may take the place of Bv when the B segment undergoes inversion, starts within the IRR present upstream of pinB, but no start codon or SD sequence was detected. The two variable regions, Bv and Bv', encode 340 and 318 amino acids, respectively. Neither a promoter sequence nor a start codon was observed in the constant region, and thus they may be present beyond the extreme left end of the fragment.

Southern blots. To confirm that the B segment and the pinB gene are derived from genomic DNA of S. boydii and to determine whether the sequence of the B segment is homologous to that of the G segment of Mu, a hybridization assay was carried out. S. boydii total DNA and plasmid pTSB916 DNA were digested with BamHI and Sall and were hybridized with the 5.9-kb BamHI fragment containing either the B segment and the pinB gene in pTSB916 or the whole Mu genome. Three inherent plasmids of S. boydii were not detected by these probes (data not shown). In the case of the B-pinB probe, the BamHI cut showed only one band (5.9 kb) in the genomic digest (data not shown), and the BamHI and Sall cut showed two bands, of 1.7 and 4.2 kb (Fig. 4, lane 2). These results are consistent with the existence of the B segment and pinB gene in the genome. With the Mu probe, the BamHI and Sall cut showed only one band, of 4.2 kb, which contained the B segment and the 5'-terminal half of the pinB gene in pTSB916 (Fig. 4, lane 3) and showed one weak band of approximately 23 kb in addition to the 4.2-kb band in the genomic digest (Fig. 4, lane 4). These results demonstrate the homology between the B segment and the G segment and suggest that a fragment of at least 1.7 kb which lies downstream of the B segment is nonhomologous to the Mu genome and that a DNA structure other than B-pinB is
FIG. 3. Nucleotide sequences and genetic organization of the B segment at the (+) orientation and the pinB gene. (A) The sequence of the 4,810-bp BamHI-MluI fragment is numbered from the BamHI site. IRL and IRR, which flank the B segment, are boxed. The inv sequences are marked by broken lines. The six site in pinB is indicated by < to >. The deduced amino acid sequences of five potential ORFs are shown: B175, B177, and PinB are complete ORFs and Bc and Bv are connected in the (+) orientation to create Bv'. The translation start and stop sites of the three complete ORFs are indicated by arrows and asterisks, respectively. The Bc start site seems to be located beyond the 5' end of the fragment. Bv and Bv' are considered to start from the glycine in IRL and IRR (13). The Shine-Dalgarno (SD) sequence and the consensus promoter sequence (=) are underlined. Palindromic sequences which may serve as transcriptional terminators are indicated by converging arrows. (B) Diagrammatic representation of ORFs. ORFs are indicated by open arrows and the number of amino acids in each ORF is shown in parentheses.

Inversion was that 10 times higher than that for the recombinase genes intrinsic to each segment. The frequency of C inversion for pinB and pinD was three orders and one order of magnitude lower than that for the cin gene, respectively. The difference in activity among these recombinase genes suggests that pinB and pinD are different kinds of genes and that pinD is also different from any of the known recombinase genes, such as hin, pin, and cin. Phenotypes of the pinB and pinD genes are unknown, as is the case for the pin gene on E. coli e14. The transduction of these two genes into the mutant pin locus of E. coli was negative, suggesting that these genes are not present on e14, which would be homol-
The presence of a defective pinF gene was suggested by the occurrence of Pin+ transductants among Pur+ Tet+ cotransductants in the transduction from *S. flexneri* to *E. coli* Δpin. The Δpin locus, flanked by purB and zcg2::Tn10, seems to be deleted with a whole P segment and at least part of the pin gene. Therefore, the Pin+ transductants would result from quadruple crossovers with the homologous regions of *S. flexneri*; two events at both extremes of the purB-Tet1 (zcg2::Tn10) segment and two events within a remnant of the recipient pin gene, its corresponding part of the donor containing a mutation. If this were the case, however, the frequency of quadruple crossovers would be much lower than that observed (50% of Pur+ Tet+ cotransductants were also Pin+). Another plausible possibility is that pinF is inactive because of the insertion of some kind of insertion element, as in the *hylF* operon (35), which would be precisely excised at a fairly high frequency when the fragment containing mutant pinF is introduced to an *E. coli* background and integrated at a homologous region. The occurrence of Pin+ transductants among Pur+ Tet+ cotransductants with the high efficiency and the likeness in restriction maps between the P-pinE fragment and the fragment isolated from a Pin+ transductant (Fig. 1) suggest that a defective pinF gene which is similar to that of *E. coli* in structure and location exists on *S. flexneri* e14. In the family *Enterobacteriaceae*, the genera *Escherichia* and *Shigella* are very closely related (33). Our data suggest that *S. flexneri* may be the most closely related to *E. coli* of the four *Shigella* strains tested.

The 4.8-kb fragment containing the pinB gene and the invertible B segment was cloned from *S. boydii* and sequenced. The pinB gene, which consists of 585 bp (195 amino acids), has its putative promoter (−10 sequence) within the IRR, although a −35 sequence could not be detected and its transcriptional direction is toward the outside of the IRR. The B segment, flanked by the two IRs, contains the two complete ORFs, B175 and B177, and the two variable regions, Bv and Bv', which may be transcribed with the 5' constant region (Bc) outside the B segment in the B(+) and B(−) orientations, respectively, and may result in the ORFs Bcv and Bcv', whose 5' coding region is not contained in the cloned fragment. Such a transcription system has been previously reported for the G, C, and P segments (18, 26, 38). The organization of B-pinB is similar to that of G-gin in the size of the invertible segment and in the transcriptional direction of the recombinase gene.

When the predicted amino acid sequence of PinB was compared with those of the four recombinases reported previously by several groups (14) (Fig. 5), the highest homology was observed for Gin (193 amino acids): 78.9% (154 of 195) of the amino acids are identical in all 5 sequences. The homology of PinB to PinE (184 amino acids), Cin (186 amino acids), and Hin (190 amino acids) is 72.8% (142 amino acids identical), 68.7% (134 amino acids identical), and 66.7% (130 amino acids identical), respectively. All these values are more than 80% when equivalent amino acids are also taken into account. The number of amino acids that are conserved as a common sequence in the five recombinases including PinB is 104, corresponding to 53 to 56% of each sequence. The amino acids that are conserved in the four recombinases but not in PinE at positions 160 (Pro→Asp) and 175 (Leu→Ile) in the C-terminal region. This region has been reported to be important for target recognition (37) and partial DNA binding (4), and hence these amino acid changes in PinB may cause the rather low frequency of segment inversion in *S. boydii*. The enhancer six, which is required for efficient inversion of the invertible segments (16, 21, 23), is present in all the recombinase genes reported to date (17). The six site was also found in the 5' coding region of pinB (Fig. 3), and its 90-bp sequence is completely identical to that of gin (23). When a 50-bp sequence located between the IRR and the pinB gene was compared with the corresponding regions of the four recombinases, 80% (40 of 50 nucleotides) and 52% (26 of 50) homology with those regions of gin and pinE, respectively, were observed, indicating that the three genes are closely related with respect not only to the coding regions but also to their upstream regions.

The ORFs on the B segment were also found to be highly homologous to those of the G segment in their organization and amino acid sequences. B175 and B177 on the B segment correspond to the two complete genes *U* (175 amino acids) and *U'* (177 amino acids) on the G segment (22), respec-
tively. Likewise, Bv and Bv' on the B segment correspond to the two variable regions Sv (327 amino acids) and S'v (311 amino acids), respectively, on the G segment; Sv and S'v and the constant S′ region (Sc) compose the genes S and S′ (22). The homology in predicted amino acid sequences of the two corresponding genes or ORFs between the B and G segments is as follows: B175 versus U, 90% (158 identical amino acids of 175); B177 versus U', 98% (173 of 177); Bv versus Sv, 62% (211 of 340); and Bv' versus S'v, 96% (306 of 318). On the other hand, the homology between the two constant regions outside the segments was less than 30% when the perfect 177-amino-acid sequence of Sc was compared with several 177-amino-acid sequences arbitrarily chosen from Bc (218 amino acids). The amino acid sequence from Bv was found to have fairly high homology with that of the C-terminal region of the gene 19 (328 amino acids) (25) on the C segment of phage P1: 61% (208 of 340) of the amino acids are identical. Each of the four ORFs on the B segment was compared with four ORFs on the P segment of the e14 element (38). The maximum amino acid homology between the ORFs of the B and P segments, which were chosen in every possible combination, is 25 to 30%, indicating that there is no homology between the B and P segments at the amino acid sequence level.

The IRL and IRR of the B segment are 38-bp sequences (in which 3 bp are different) and contain the 26-bp inv consensus sequence (14), which is separated by a 5′-AA dinucleotide into two imperfect 12-bp IRs (14). When this sequence of the B segment was compared with those of other invertible segments, 23 (88.5%) of 26 bp for IRL and 24 (92.3%) of 26 bp for IRR were found to be identical to those of the G segment. Moreover, instead of the AA dinucleotide where recombination occurs (19, 21, 37), a 5′ GA was observed in both IRs of the B segment, as reported for inv of the G segment (36). The homology with inv sequences from the H, C, and P segments is 53.9 to 80.8% for IRL and 53.9 to 84.6% for IRR (14).

Taken together, these homology tests show that B-pinB of S. boydii is similar to G-gin of phage Mu in every way but that the 5′ constant region outside the B segment is quite different in size and predicted amino acid sequence. Phage has not been detected in S. boydii, though its culture fluid was tested on the indicator strains E. coli C and K12 at 0.02 (28) before and after its induction by mitomycin or UV irradiation. Mu and its related phage D108 are known to be noninducible and inducible, respectively (8). Thus, if an intact prophage like Mu or D108 were present in S. boydii, it would be detected by such tests. Although the hybridization assay suggested that the sequence homologous to the Mu genome is present in S. boydii, it remains unknown whether the B-pinB structure is part of a defective prophage into which a Mu-like phage has degenerated or whether it is a prototype structure to be integrated into a prophage to result in an intact phage, such as Mu. Assuming that B-pinB is part of some prophage, the products of the ORFs on the B segment and its flanking region would be tail fiber proteins concerned with host recognition, similar to those on the G segment of Mu and on the gene 19 product of P1.

REFERENCES

inversion in bacteriophage Mu DNA is stimulated by a site within the invertase gene and a host factor. Cell 41:771-780.


