JOURNAL OF BACTERIOLOGY, Feb. 1996, p. 894–898 Vol. 178, No. 3
0021-9193/96/$04.00+0

Gene Integration in the *Escherichia coli* Chromosome Mediated by Tn*21* Integrase (Int21)

M. VICTORIA FRANCIA AND JUAN M. GARCÍA LOBO*

Departamento de Biología Molecular, Facultad de Medicina, Universidad de Cantabria, Santander, Spain

Received 31 July 1995/Accepted 14 November 1995

A replication-thermosensitive, pSC101-derived plasmid containing the *int* gene and RHS-2 from the integron in Tn*21* and a kanamycin resistance marker has been constructed and used to obtain Tn*21* integrase (Int21)-mediated plasmid integration in the *Escherichia coli* chromosome. Colonies carrying an integrated plasmid were obtained after growth at 42°C. Southern hybridization and PCR experiments indicated that they contained the plasmid specifically integrated through the RHS into different positions in the *E. coli* chromosome. Nucleotide sequence determination of the plasmid-chromosome junctions showed that integration sites in the chromosome were pentanucleotides with the sequence described for Int21 secondary sites.

Tn*21* integrase (Int21) is a site-specific recombinase belonging to the family of λ integrases. The gene coding for this enzyme is found in the 5′ conserved regions of integrons, which are a kind of genetic element involved in the spread of antibiotic resistance genes and present in many bacterial genera (16). Int21 has proved able to catalyze the site-specific joining of two plasmid molecules. For the reaction to occur, each partner must contain a sequence called the 59-bp element, or RHS, which is the target for integrase action and at which 3′ end the recombination reaction occurs (6, 9, 10). Furthermore, Tn*21* integrase can also fuse a plasmid containing an RHS to a second plasmid without a specific site. In this case secondary sites are used as targets in the second molecule and the reaction is at least a thousand times less efficient than the reaction between two primary sites (4, 13). Secondary sites have been characterized, and they consist of a loosely conserved pentanucleotide with the sequence GWTMW. Only the first and third positions are absolutely conserved. In most cases described until now, the Tn*21* integrase reaction has been detected after the conjugal transfer of the recombinants.

Site-specific recombinases are valuable tools for biotechnologists because they allow the insertion of genes into controlled positions of the genome (7). Since Int21 has proven useful in catalyzing site-specific recombination between plasmids at a high frequency, we decided to investigate the possible use of Int21 for site-specific integration into the bacterial chromosome.

With this aim in mind, we constructed plasmid pSU739, with a thermosensitive replicon in *E. coli* derived from plasmid pSC101 (11). Plasmid pJU739 contains the kanamycin resistance gene from Tn*5* and the *int* gene and a recombination hot spot (RHS-2) from the transposon Tn*21* (10). A map and the scheme for the construction of this plasmid are shown in Fig. 1. The sequence of Tn*21* RHS-2 is shown in Fig. 2. Plasmid pSU739 was introduced into *E. coli recA* strain DH5α (Gibco BRL), where it was stably maintained as long as the growing temperature was maintained at 30°C. The use of a *recA* strain and the absence of sequence homology between the plasmid and the *E. coli* chromosome should guarantee a low level of Int21-independent plasmid integration. Growth of this strain at 42°C on agar containing kanamycin (50 μg/ml) allowed the selection of cells carrying the plasmid integrated in the chromosome. Since the plasmid cannot replicate and be maintained in an autonomous state at 42°C, we assumed that Int21-mediated recombination should be the major cause of plasmid integration. The frequency of Int21-mediated integration of plasmid pSU739 into the chromosome, calculated as the number of colonies growing on kanamycin agar at 42°C divided by the number of colonies that grow at 30°C, was 6.5×10^{-6}. The frequency of integration of the thermosensitive plasmid pSU300, the parent of pSU735, under the same condition was less than 10^{-9}.

Several colonies growing on kanamycin agar at 42°C were analyzed. They did not apparently contain plasmids, as judged by agarose gel electrophoresis of cleared lysates. Chromosomal DNA was prepared basically as described in reference 14 and analyzed by Southern hybridization after digestion with restriction endonuclease *Hae*II. The 1.3-kb *Hae*II fragment from plasmid pSU739 that contained RHS-2 (Fig. 1) was labelled with digoxigenin (Boehringer Mannheim) in a randomly primed reaction with Klenow DNA polymerase. Then it was used to probe chromosomal DNA from the colonies carrying integrated pSU739. As a general rule, two bands of variable size were seen in each chromosome (Fig. 3), indicating that the 1.3-kb *Hae*II band from pSU739 had been interrupted, probably as result of the use of RHS-2 for site-specific integration into the *E. coli* chromosome. The variable sizes of the bands that hybridized in each independent integration indicated that there was not a unique site in the *E. coli* chromosome for Int21-mediated integration.

Since the previous experiment did not exclude the possible use of a recombination site in the pSU739 1.3-kb *Hae*II band other than the RHS, a PCR experiment was carried out to further localize the crossover point. Two primers (GATGG ATCACCAAGGTAG and CCTGATAGTTTGGCTGTGA GC) were synthesized from sequences taken at both sides of RHS-2 (Fig. 1 and 2). They were able to amplify a 130-bp region when plasmid pSU739 was used as the template in a PCR. The same primers were used to amplify chromosomal DNA from the colonies carrying the integrated plasmid. None of these chromosomal DNAs supported the amplification of the 130-bp band. A control amplification with a different pair of primers (Fig. 1) was positive in all cases (data not shown). This result further localizes the crossover point to the 130-bp

* Corresponding author. Mailing address: Cardenal Herrera Oria s/n, 39011-Santander, Spain. Phone: 34 42 201948. Fax: 34 42 201945. Electronic mail address: JMGLOBO@MEDI.UNICAN.ES.

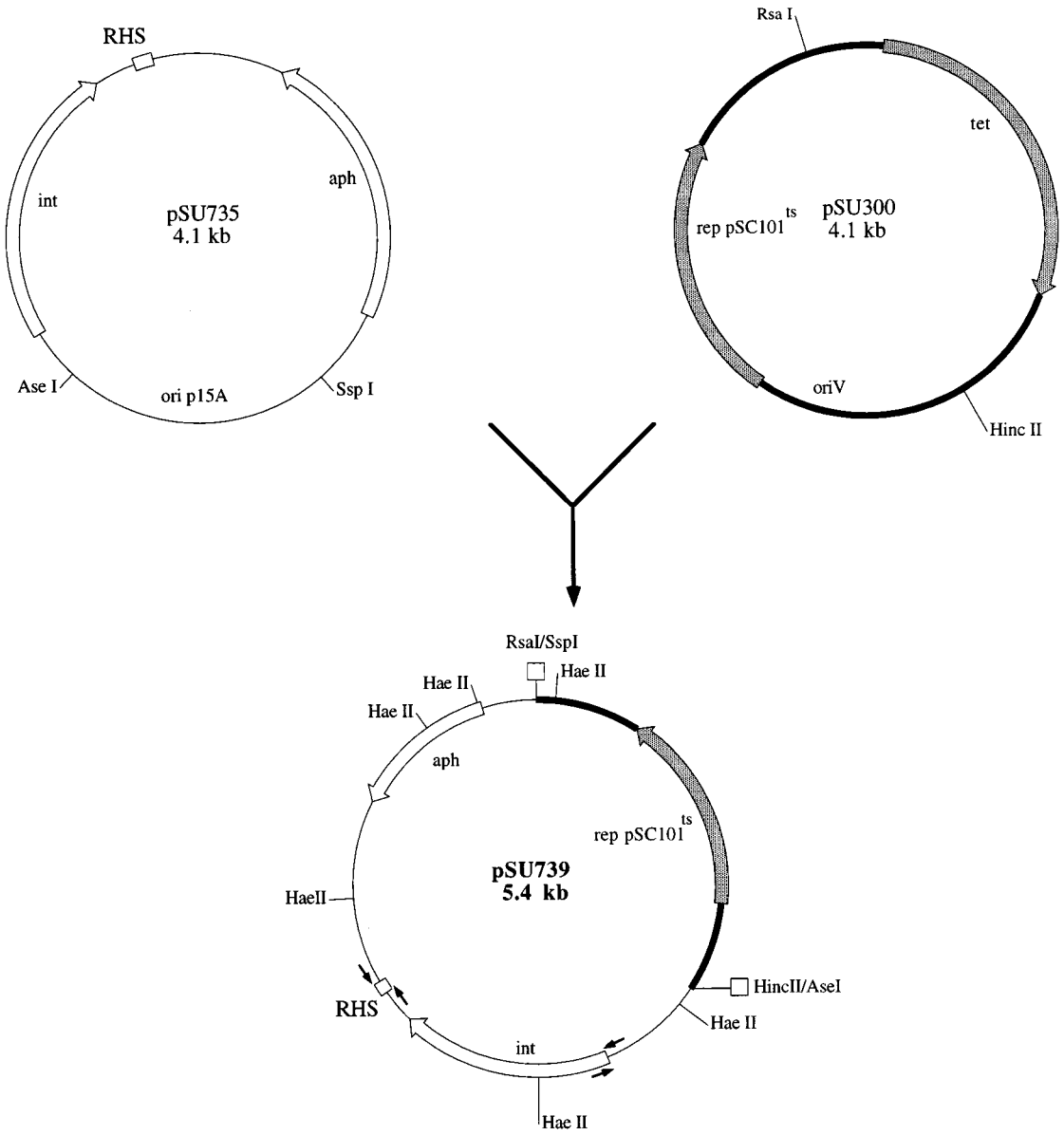

FIG. 1. Restriction and genetic map of plasmid pSU739. The replication region of the thermosensitive pSC101 derivative pSU300 (11), obtained as a *Hin*cII-*Rsa*I fragment, was ligated to a *Ssp*I-*Ase*I fragment from plasmid pSU735 that contained *int* and RHS-2 from Tn*21* and the kanamycin resistance gene from Tn*5*. The small arrows in the pSU739 map around the RHS indicate the position of the oligonucleotides used for PCR and sequencing. The small arrows around the 5′ end of *int* show the positions of the primers used for the control PCRs.

fragment containing RHS-2 but still does not prove that recombination occurred precisely at the RHS.

To determine the exact position of the recombination site and to map the integration points in the *E. coli* chromosome, the nucleotide sequences of 12 pairs of plasmid-chromosome junctions were determined with the primers used in the PCR experiments. Sequences were determined by using total chromosomal DNA as a template with the *fmol* DNA sequencing kit from Promega. The sequences obtained confirmed that the integration in the chromosome had occurred site specifically at RHS-2 of plasmid pSU739. We could also determine the exact position and the sequence around each integration site in the *E. coli* chromosome. These sequences are shown in Fig. 4. In all cases the integration occurred next to a G in a pentanucleotide with the consensus sequence defined for Int21 secondary sites (4). Interestingly, a second consensus pentanucleotide was found in 10 of the 12 described sequences at a distance from 3 to 7 bp from the 5′ side of the integration site but in the other DNA strand. This second pentanucleotide was not observed at this frequency in 50 previously determined Int21 secondary sites in multicopy plasmids (4, 13). This difference in the structures of target sites could be due to differences in the structures of the substrates, such as the degree of supercoiling or the presence of accessory proteins.

RHSs are palindromic sequences able to adopt a hairpin-type secondary structure. Integration at the RHS always happens asymmetrically at one of the ends that would be located at the base of the hairpin stem. We looked for the presence of palindromic sequences around the chromosome integration sites to determine the possible role of some secondary struc-

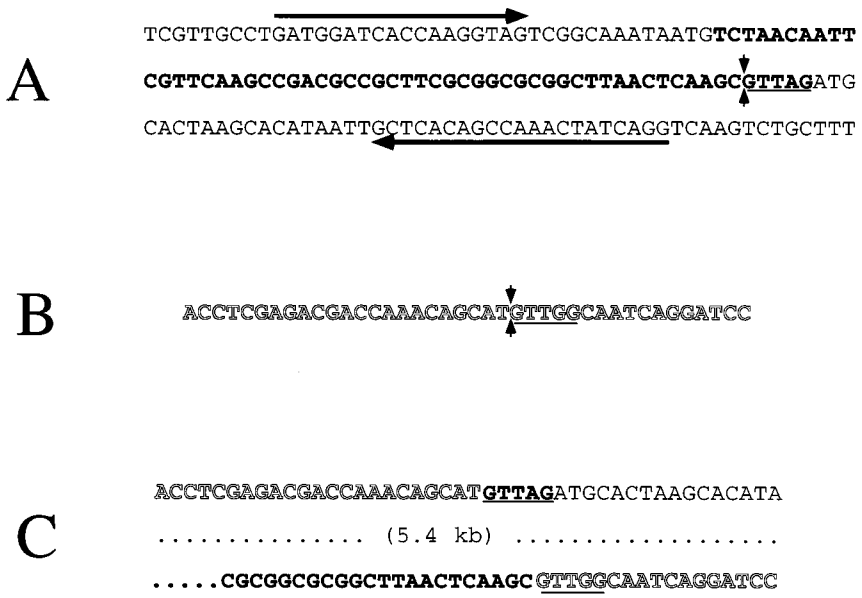

FIG. 2. Diagram showing the Int21-mediated recombination reaction at a secondary site in the *E. coli* chromosome. (A) Nucleotide sequence of plasmid pSU739 around RHS-2. The sequence of RHS-2 is in boldface. The vertical arrows indicate the recombination crossover point. The arrows on the sequence mark the oligonucleotides used for PCR and sequencing. The terminal pentanucleotide of the RHS is underlined. (B) Sequence of the *E. coli* chromosome at integration site 19. (C) Sequence of the chromosome with pSU739 integration site 19 is shown. Note that after integration, the terminal pentanucleotide of the RHS is replaced with the pentanucleotide in the integration site.

ture in the election of secondary sites. Only the sequence around integration site 50 (Fig. 4) proved able to form an stable stem-loop structure. This insertion occurred in a repetitive extragenic palindromic (REP) sequence (15) that formed a cluster with at least two more REP copies that were almost identical to a cluster of REP units found among *phn* genes (2) at min 93.05 in the *E. coli* chromosome. The integration site in the REP sequence was not in the base of the stem, as it is in the RHS. It was instead found in the variable loop in the middle of a REP sequence. All these data argued against the importance of the secondary structure in Int21 integration at secondary sites.

To map the integration sites to the *E. coli* chromosome, we searched the bacterial section of GenBank by using the FINDPATTERNS program of the Genetics Computer Group package (5) for identities with the sequences at each integration point. Seven of the 11 different sequences were already present in the data bank (EMBL release 41.0 plus updates), and we could determine their positions on the *E. coli* map. The locations of these sequences are shown in Fig. 4. It was surprising to find that most of the insertion points were clustered around *E. coli oriC*. This finding could simply reflect the higher copy number of the region around *oriC* in fast-dividing *E. coli* cells. Alternatively, it could be the result of a higher degree of accessibility of this chromosomal region for Int21 binding.

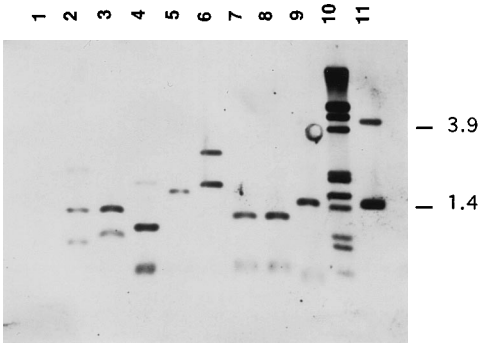

FIG. 3. Southern blot showing that plasmid pSU739 integrated at different positions in the *E. coli* chromosome. Aliquots (3 μg each) of *Hae*II-digested chromosomal DNA from different isolates carrying integrated plasmid pSU739 were hybridized with a probe containing RHS-2 from pSU739. The different bands observed with each isolate indicated that many different sites were used for integration in the *E. coli* chromosome (lanes 2 to 9). Lane 1 contains the same digestion of chromosomal DNA from DH5α without any plasmid (negative control). Lane 10 contains the digoxigenin-labelled molecular size marker III from Boehringer Mannheim. Lane 11 contains the pSU739 plasmid digested with *Hin*dIII, as a positive control.

With this work we have demonstrated the possibility of Int21 mediating plasmid integration into the chromosome of *E. coli*. The analysis of the recombination sites showed that the integration reaction mediated by Tn*21* integrase was site specific. The absence of a single hot spot for integration revealed the absence of sequences similar to integron RHS in the *E. coli* chromosome. However, recombination was still possible and occurred at multiple secondary sites clustered around the replication origin. On the other hand, the construction of plasmid pSU739 indicates that the block of DNA containing the *int* gene and RHS-2 constitutes an integrative cassette for *E. coli*. It could be easily included in any plasmid to make an integration vector. The efficiency and specificity of the integration reaction mediated by this kind of vector would be greatly increased by the integration of an RHS in the *E. coli* chromosome, as has been described for the integration of plasmids containing other site-specific recombinases (8). A site-specific recombination system based on Int21 requires only the integrase and the recombination site and should have the same properties and uses as those systems based on Cre and FLP integrases (7). Since integrons are broadly distributed in bacteria, such an integration system could be useful for bacterial species that lack integration vectors based on phage *att* sites.

The integration at secondary sites represents a further advantage in terms of the stability of the integrated genes. The

		min.	*locus*
Ins.5	CACCTACTGATTCATCAAAACAATAGTTATGGAGTATTTATGAGTAAGGA	86.18	*corA*
Ins.19	ACCTCGAGACGACCAAACAGCATGTTGGCAATCAGGATCC	86.9	*hemG*
Ins.24	CGACAATCACATTACCTAAATCAAAGATATAGAGCATTTTTGCCTCCTT	87.97	ECOUW87
Ins.27	AGTAGGCCATGCTGAATAAACTGAAGTTATCCAGATAGTTCGCCAGCTC	63.83	*metC*
Ins.49	CAGCCGTCCGGGGGATTAACCCTGAGATAATGACTGATGGAACTCATTG		
Ins.50	GGAATTTGTAGGCCTGATAAGACGCGTTA		
Ins.54	AGCAGGAACTGGAAAnTAACCAGCAGTTAAGTCAGCGTCTGnTT		
Ins.59	CCGAAAACGGCGGTTAACGTGATTGGTTATACCGACAGCACGGGTGG	79.2	ECOUW76
Ins.63	ACGAGCGACGTCCTGTGGCGATAACACGCCGGGAATGT		
Ins.64/72	ACGATAATGCCCCCACCATCCGCCAGTTAAACAGCACATCTTCTTCCTG	89.67	ECOUW89
Ins.73	GACGTTTACGCCTGTGATCACTATGTTTATCCGCAATATCGGCTGGGG	85.83	*aslB*

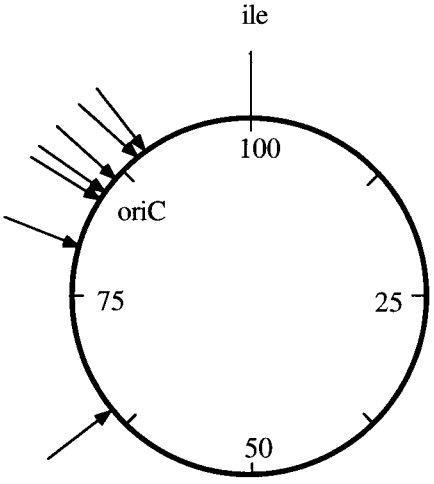

FIG. 4. Nucleotide sequences of the sites of Int21-mediated pSU739 integration into the *E. coli* chromosome and locations of the integration sites on the *E. coli* chromosome map. Map positions have been taken from the *E. coli* database. In cases in which the target sequence was not identified as an individual gene but as part of a large sequence entry in the database, the map location was extrapolated from the position of the match. Ins., insertion.

integrated plasmid is flanked by an RHS and a secondary site. Accordingly, for the plasmid to again become autonomous, the excision reaction should occur between these two sites. Excision mediated by Int21 and involving two RHSs seems to be a process less efficient than integration (3). The frequency of excision between an RHS and a secondary site should be lowered more, becoming very inefficient. As a matter of fact, we were unable to detect this excision even in the presence of elevated integrase amounts (data not shown). Accordingly, integration at secondary sites in the chromosome contributes to insert stability by reducing the frequency of the excision reaction.

In spite of its low frequency, Int21-mediated recombination at secondary sites is a process that happens in nature, as has been recently proven for the *aadB* gene in plasmid pIE723; this gene is integrated at a secondary site (12). These rare events can be observed because of the strong selective pressure introduced by the use of antibiotics.

The loose conservation of sequences observed in chromosomal integration sites makes almost any chromosomal gene a potential target for Int21-mediated integration and provides a way for association of chromosomal genes with preexisting integrons. Chromosomal genes could then become incorporated in the integron either by Int21-mediated site-specific recombination or by general recombination with another integron (12). This mechanism can explain the described similitude between integron and chromosomal genes (1) and the presence of truncated genes in integrons. Integrons could then be seen as systems for generalized gene spread rather than specialized vehicles for the dissemination of antibiotic resistance genes.

This work was supported by grant PB93-0368 from the CICYT, Spain, and by contract BIO2CT920520 from the European Commission.

We acknowledge F. de la Cruz and Pilar Avila for helpful discussion and critical reading of the manuscript.

REFERENCES

1. **Bunny, K. L., R. M. Hall, and H. W. Stokes.** 1995. New mobile gene cassettes containing an aminoglycoside resistance gene, *aacA7*, and a chloramphenicol resistance gene, *catB3*, in an integron in pBWH301. Antimicrob. Agents Chemother. **39:**686–693.
2. **Chen, C., Q. Ye, Z. Zhu, B. L. Wanner, and C. T. Walsh.** 1990. Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the *phn* (*psiD*) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli*. B. J. Biol. Chem. **265:**4461–4471.
3. **Collis, C. M., and R. M. Hall.** 1992. Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. J. Bacteriol. **174:**1574–1585.
4. **Francia, M. V., F. De La Cruz, and J. M. García Lobo.** 1993. Secondary sites for integration mediated by the Tn*21* integrase. Mol. Microbiol. **10:**823–828.
5. **Genetics Computer Group.** 1994. Program manual for the Wisconsin package, version 8. Genetics Computer Group, Madison, Wis.
6. **Hall, R. M., D. E. Brookes, and H. W. Stokes.** 1991. Site specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point. Mol. Microbiol. **5:**1941–1959.
7. **Kilby, N. J., M. R. Snaith, and J. A. H. Murray.** 1993. Site-specific recombinases: tools for genome engineering. Trends Genet. **9:**413–421.

8. **Leslie, N. R., and D. J. Sherratt.** 1995. Site-specific recombination in the replication terminus region of *E. coli*: functional replacement of *dif*. EMBO J. **14:**1561–1570.
9. **Martínez, E., and F. de la Cruz.** 1988. Transposon Tn*21* encodes a Rec-A independent site-specific integration system. Mol. Gen. Genet. **211:**320–325.
10. **Martínez, E., and F. de la Cruz.** 1990. Genetic elements involved in Tn*21* site specific integration, a novel mechanism for the dissemination of antibiotic resistance genes. EMBO J. **9:**1275–1281.
11. **Mendiola, M. V., and F. de la Cruz.** 1989. Specificity of insertion of IS*91*, an insertion sequence present in haemolysin plasmids of *Escherichia coli*. Mol. Microbiol. **3:**979–984.
12. **Recchia, G. D., and R. H. Hall.** 1995. Plasmid evolution by acquisition of mobile gene cassettes: plasmid pIE723 contains the aadB gene cassette precisely inserted at a secondary site in the IncQ plasmid RSF1010. Mol. Microbiol. **15:**179–187.
13. **Recchia, G. D., H. W. Stokes, and R. H. Hall.** 1994. Characterization of specific and secondary recombination sites recognized by the integron DNA integrase. Nucleic Acids Res. **22:**2071–2078.
14. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions, p. 137. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Stern, M. J., G. F. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37:**1015–1026.
16. **Stokes, H. W., and R. M. Hall.** 1989. A novel family of potentially mobile DNA elements encoding site-specific integration functions: integrons. Mol. Microbiol. **3:**1669–1683.