JOURNAL OF BACTERIOLOGY, June 1997, p. 3594–3603 Vol. 179, No. 11
0021-9193/97/$04.00+0

The Plasmid R64 Thin Pilus Identified as a Type IV Pilus

SU-RYANG KIM AND TERUYA KOMANO*

Department of Biology, Tokyo Metropolitan University, Minamiohsawa, Hachioji, Tokyo 192-03, Japan

Received 14 November 1996/Accepted 2 April 1997

The entire nucleotide sequence of the *pil* region of the IncI1 plasmid R64 was determined. Analysis of the sequence indicated that 14 genes, designated *pilI* through *pilV*, are involved in the formation of the R64 thin pilus. Protein products of eight *pil* genes were identified by the maxicell procedure. The *pilN* product was shown to be a lipoprotein by an experiment using globomycin. A computer search revealed that several R64 *pil* genes have amino acid sequence homology with proteins involved in type IV pilus biogenesis, protein secretion, and transformation competence. The *pilS* and *pilV* products were suggested to be prepilins for the R64 thin pilus, and the *pilU* product appears to be a prepilin peptidase. These results suggest that the R64 thin pilus belongs to the type IV family, specifically group IVB, of pili. The requirement of the *pilR* and *pilU* genes for R64 liquid mating was demonstrated by constructing their frameshift mutations. Comparison of three type IVB pilus biogenesis systems, the *pil* system of R64, the toxin-coregulated pilus (*tcp*) system of *Vibrio cholerae*, and the bundle-forming pilus (*bfp*) system of enteropathogenic *Escherichia coli*, suggests that they have evolved from a common ancestral gene system.

Type IV pili are rod-like surface appendages produced by gram-negative bacteria such as *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae*, *Moraxella bovis*, *Myxococcus xanthus*, and *Vibrio cholerae*, as well as enteropathogenic and enterotoxigenic *Escherichia coli* (for reviews, see references 27, 28, and 35). They are flexible, with a diameter of 6 to 7 nm and a length of up to 20 μm. They are produced at the polar position of the bacterial cell. Many type IV pili play important roles in the attachment of bacterial pathogens to membranes of eukaryotic host cells, as do the other pili (13, 35). Type IV pili are also associated with the twitching motility of various bacteria and with the social motility of myxobacteria (40).

Type IV pili are composed of pilin subunits (35). Pilin molecules from various bacteria have amino acid sequence homology (see Fig. 4E). The type IV pilin family is usually divided into two groups. Group A consists of pilins from *P. aeruginosa*, *N. gonorrhoeae*, *M. bovis*, and so on. They are closely related in amino acid sequence and are produced from prepilin molecules through the cleavage of 6- to 7-amino-acid signal peptides. The N-terminal amino acid of type IVA mature pilins is phenylalanine and is N methylated. Group B pilins, including toxin-coregulated pilus (*tcp*) in *V. cholerae* (26) and bundle-forming pilus (*bfp*) in enteropathogenic *E. coli* (33, 34), are substantively different from type IVA pilins. Their signal peptides are longer than those of type IVA pilins. The N-terminal amino acid of type IVB mature pilins is methionine or leucine.

The C-terminal amino acid (glycine) of signal peptides and the 5th amino acid (glutamic acid) of mature pilins are completely conserved among type IVA and type IVB prepilins and related proteins (see Fig. 4E). A long hydrophobic segment is present at the N-terminal region of mature pilins. Cleavage of signal peptides from prepilins is carried out by a signal peptidase, specific for type IV prepilin, such as the PilD protein of *P. aeruginosa* (36). N methylation of mature pilins is also performed by the same signal peptidase protein.

Various extracellular protein secretion systems of gram-negative bacteria and DNA uptake systems of gram-positive bacteria are known to produce proteins containing extensive sequence homology to proteins required for type IV pilus biogenesis, including prepilins (12, 28).

Many gene products are required for the biogenesis of type IV pili. The gene organizations of the *tcp* system in *V. cholerae* (26) and the *bfp* system in enteropathogenic *E. coli* (33, 34) have been reported (see Fig. 6). In the *tcp* and *bfp* systems, 14 genes are involved in type IV pilus biogenesis. In the case of the *bfp* system and the longus (11) system in enteropathogenic and enterotoxigenic *E. coli*, respectively, the type IV pilus formation genes are encoded in large virulent plasmids. The virulent plasmid in enteropathogenic *E. coli* (EPEC) is called EPEC adherence factor (EAF).

Plasmid R64 is a conjugative plasmid belonging to the incompatibility group I1 (17, 18). The R64 transfer region is located within a 54.0-kb DNA segment, in which the *traABCD* genes are located at the left end and the *oriT* sequence is at the right end (Fig. 1A). R64 produces two types of sex pili, thin, flexible pili and thick, rigid pili (3, 4). Thin pili are required only for liquid mating, while thick pili are required for both surface and liquid mating. The leftmost 18.5-kb DNA segment of the R64 transfer region is responsible for R64 thin pilus formation (*pil* region), since *E. coli* cells harboring pKK641, containing the 18.5-kb segment together with the R64 *rep* segment (Fig. 1A), were shown to produce R64 thin pili (18, 19). *E. coli* cells harboring pKK641 are sensitive to the IncI1-specific phages Iα (5) and PR64FS (6), which utilize thin pili as receptors.

The leftmost 3.6-kb segment of pKK641 was sequenced and shown to contain the *traABCD* genes (15) (Fig. 1B). The *traBC* genes were inferred to be positive regulators of R64 transfer gene expression, since they were required both for thin-pilus formation and for conjugal transfer in liquid medium as well as on a solid surface.

The rightmost 4.1-kb segment of pKK641 was also sequenced and was shown to contain the *pilV* gene, the R64 shufflon, and a portion of the *rci* gene (20, 21). The C-terminal segment of the *pilV* gene is under the control of DNA rearrangement of the shufflon. The R64 shufflon consists of four DNA segments flanked and separated by seven 19-bp repeat sequences. The *rci* gene product promotes site-specific recom-

* Corresponding author. Mailing address: Department of Biology, Tokyo Metropolitan University, Minamiohsawa, Hachioji, Tokyo 192-03, Japan. Phone: 81-426-77-2568. Fax: 81-426-77-2559. E-mail: komano-teruya@c.metro-u.ac.jp.

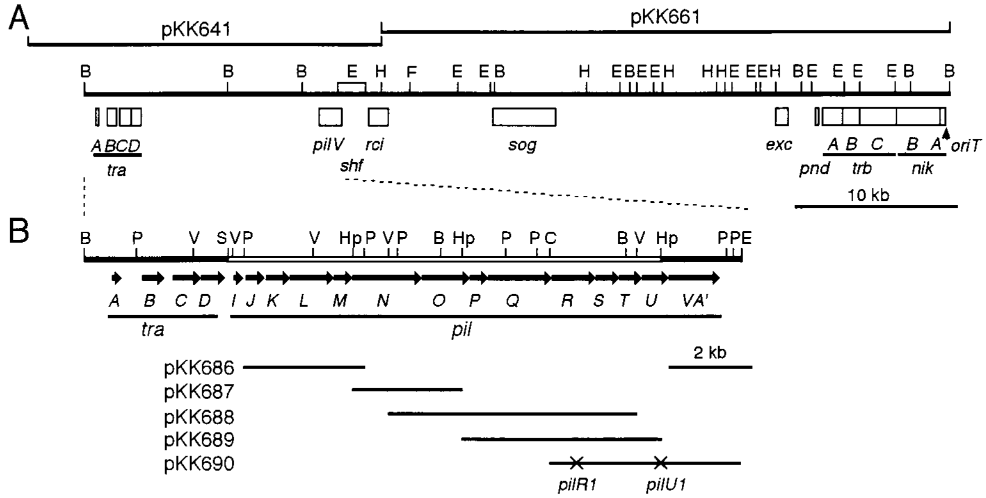

FIG. 1. (A) Organization of the 54-kb R64 transfer region. The horizontal bold line represents a restriction map. B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII. The open bar on the map shows the extent of movement of the *Eco*RI site due to DNA rearrangement of the shufflon. The open bars below the map represent the ORFs determined thus far. *tra*, *trb*, transfer; *pil*, thin-pilus formation; *shf*, shufflon; *rci*, shufflon-specific recombinase; *sog*, suppression of *dnaG*; *exc*, surface exclusion; *pnd*, plasmid stability; *nik*, *oriT*-specific nickase; *oriT*, origin of transfer. All genes are transcribed rightward except for *pnd* and *nikAB*, which are translated leftward. The DNA regions of pKK641 and pKK661 are indicated above the map. (B) Structure of the *pil* region. At the top is a restriction map of the 16.2-kb *Bgl*II-*Eco*RI segment. The DNA sequence newly determined in this study is represented by the open bar. pKK641 used in this study carries the shufflon of the fixed arrangement in which the *pilVA′* gene is expressed. C, *Cla*I; Hp, *Hpa*I; P, *Pst*I; S, *Sma*I; V, *Pvu*II. The arrows below the map indicate the coding sequences of various *tra* and *pil* genes. The solid lines indicate the DNA segments present in various plasmids. The crosses on pKK690 mark the locations of the *pilR1* and *pilU1* mutations.

bination between any two inverted repeat sequences, resulting in the inversion of four DNA segments independently or in groups. The shufflon is thought to function as a biological switch to select one of seven *pilV* genes, whereby the N-terminal segment remains constant while the C-terminal region is variable. The *pilV* gene product is a minor component of the R64 thin pilus. The seven variable C-terminal segments of the *pilV* gene determine the recipient specificity of R64 liquid mating (18, 19).

Recently, a DNA segment carrying the *pil* region of IncI1 plasmid ColIb-P9*drd-1*, closely related to R64, was cloned (18, 42). *E. coli* cells harboring this plasmid were shown to overproduce thin pili which showed a filamentous structure under an electron microscope (42). The purified thin pili consist of 19- and 48-kDa proteins, which were shown to be the products of the ColIb-P9 *pilS* and *pilVA′* genes, respectively.

In the present study, we have determined the nucleotide sequence of the remaining portion of the R64 *pil* region and have found that 14 genes are present in the same orientation within the R64 *pil* region. Protein products of eight genes were detected by the maxicell procedure. The deduced amino acid sequences of several *pil* genes suggest that the R64 thin pilus belongs to the type IV family of pili.

MATERIALS AND METHODS

Bacterial strains, phages, and plasmids. *E. coli* K-12 strains used were JM83 Δ(*lac-proAB*) *rpsL thi ara* φ80 d*lacZ*Δ*M15*, JM109 *recA1* Δ(*lac-proAB*) *endA1 gyrA96 thi hsdR17 supE44 relA1*/F′ *traD36 proAB lacI*^{q}*Z*Δ*M15* (38, 41), TN102 (17), and CSR603 *uvrA6 recA1 phr thr leuB6 proA2 thi argE3 lacY1 galK2 ara xyl mtl rpsL31 tsx supE33* (31).

Phages Iα (5) and PR64FS (6) were used as IncI1-specific phages. Phage vectors M13mp18 and M13mp19, and plasmid vectors pUC9 and pUC19, were used for sequencing and cloning, respectively (38, 41). The other plasmids used in this study are shown in Fig. 1.

Media. Luria-Bertani and H media were prepared as described by Miller (24). K and Hershey media were used for the maxicell experiment (31). The agar medium consisted of Luria-Bertani or H broth containing 1.5 or 1.2% agar, respectively. Antibiotics were added to liquid or solid medium at the following concentrations: ampicillin, 100 μg/ml; chloramphenicol, 25 μg/ml; kanamycin, 50 μg/ml; nalidixic acid, 20 μg/ml; and tetracycline, 12.5 μg/ml.

DNA manipulation and sequencing. The preparation of plasmid DNA, construction of plasmids, transformation, and other methods of DNA manipulation were performed as previously described (30). pKK686, pKK687, pKK688, pKK689, and pKK690 were constructed by inserting 3.0-kb *Pst*I, 2.8-kb *Hpa*I, 6.3-kb *Pvu*II, 5.0-kb *Hpa*I, and 4.1-kb *Cla*I-*Eco*RI fragments of pKK641 into pUC9 in the direction in which *pil* genes are expressed under the control of the *lac* promoter of the vector (see Fig. 1B).

Frameshift mutations were introduced into *pilR* and *pilU* genes of pKK690 as described previously (17). pKK690 DNA was digested partially with *Rsa*I or completely with *Hpa*I, ligated with a 1.45-kb DNA cassette for tetracycline resistance from pUC7Tc, and used to transform *E. coli* JM83 cells. The DNA fragment for tetracycline resistance was removed from the resultant plasmid DNA by *Bam*HI digestion of DNA, followed by self-ligation. A 22-bp DNA sequence, AATTCCCCGGATCCGGGGAATT, remaining at the *Rsa*I site (position 12311 in Fig. 2) or at the *Hpa*I site (position 14378) gave rise to the *pilR1* and *pilU1* frameshift mutations, respectively. The *pilR1* and *pilU1* mutations in pKK690 were transferred into pKK641 by the in vivo gene replacement method (17).

The nucleotide sequence was determined by the dideoxy chain termination method (32). The sequences of both strands were determined with overlapping fragments. The resulting sequence was analyzed by using DNAsis (Hitachi) and the GenomeNet WWW server (Kyoto University).

Identification of protein products. Plasmids pKK686 through -689, and pKK690 and its insertion derivatives, were introduced into *E. coli* CSR603 by transformation, and the protein products were labeled with [^{35}S]methionine according to the maxicell procedure (31). For pKK687, globomycin (final concentration, 50 to 200 μg/ml) was added to a 0.5-ml culture of maxicells 20 min prior to the addition of [^{35}S]methionine. The labeled proteins were separated by sodium dodecyl sulfate-17.5% polyacrylamide gel electrophoresis (SDS-17.5% PAGE) and visualized by fluorography. Molecular-weight-standard proteins were stained with Coomassie brilliant blue.

Conjugal transfer and phage sensitivity. Liquid mating was performed as described previously (17). *E. coli* JM83 and TN102 cells were used as donor and recipient cells, respectively. A culture of log-phase donor cells was mixed with an overnight culture of recipient cells. The mixture was incubated for 90 min at 37°C. Sensitivity to phages Iα and PR64FS was determined as described previously (17).

Materials. Restriction enzymes, sequencing kits, and other materials for recombinant DNA experiments were commercially obtained and used as recommended by the manufacturers. [α-^{32}P]dCTP (800 Ci/mmol) and [^{35}S]methionine (1,000 Ci/mmol) were from Du Pont, NEN Research Products, and from ICN Biomedicals, respectively. Globomycin was a generous gift from M. Inukai.

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases as accession no. D88588.

```
3601 CCCGTGACCCGTCTTCGCAGCCTGTACCACCCTTTCTGCCGTTCCCGGGATTCTTTCTGTCACTACTCGCTTGTGCCGTTTTTTCGCAACCGGTACGTGCAGTTACGACGGTACAGCTGC
3721 TTCTTTTCCCGTTCCGAAAGGAGTCTTTGATGCCACAGCAACATCCCGGACGACTACAGGTACTGGTCGTCGACACGCACTGCAAACGCAAGCTGTTTTCAACGAAAACACAGACAGATC
                              pilI M  P  Q  Q  H  P  G  R  L  Q  V  L  V  V  D  T  H  C  K  R  K  L  F  S  T  K  T  Q  T  D  P
3841 CCGATGAGCTTGCCCGACGCTTCTGCACGCCGGATAACTGCCTTGTGGTGGTACTATGCAACAACCGATTTCTGTTCCGCCTGGAACGTGCACCCGGTAGCCACTGCCGCTGGCGCAAGG
       D  E  L  A  R  R  F  C  T  P  D  N  C  L  V  V  V  L  C  N  N  R  F  L  F  R  L  E  R  A  P  G  S  H  C  R  W  R  K  G
3961 GTAGCCGCTCCCGCCATCAGCACCTGCAGGACTGGCTGTCGTAAAAACAACAAAAACAAAAGACCTCGTCAAAACAGGAGGGGACCATGTCGTTTTTATTTAAATTAACCATCGCGATTC
      S  R  S  R  H  Q  H  L  Q  D  W  L  S  *                              pilJ M  S  F  L  F  K  L  T  I  A  I  L
4081 TGATTGTCTTTGCAATCCTTTTTCCGATTTTAATTTCATCAGATAACAAACTGTCCGCACTTTTTTGGAAGATACCAAAAGATTTCTGGGTTTTCCTGATTATCATGTACGGGGCAATCA
      I  V  F  A  I  L  F  P  I  L  I  S  S  D  N  K  L  S  A  L  F  W  K  I  P  K  D  F  W  V  F  L  I  I  M  Y  G  A  I  I
4201 TCACGGTGGTGTTCATAGAAAGCAGTGAACAGGCACGCATCCGCCGCCAACAATATTTGCAACAGTACAAAATGCCAAATGAGATACATGAAAATGTGGTGATAGAAAACAGTATTATTT
      T  V  V  F  I  E  S  S  E  Q  A  R  I  R  R  Q  Q  Y  L  Q  Q  Y  K  M  P  N  E  I  H  E  N  V  V  I  E  N  S  I  I  S
4321 CAAGAGAAACAACAGATGACGAATCTGACATTAACACCCTGTACCTTGATTCATTTAAAAGTAAAAACAACTGCAAATCTTTTGAAACCAATTCAAGTTTACTGGATAAAAATGGTAAGC
       R  E  T  T  D  D  E  S  D  I  N  T  L  Y  L  D  S  F  K  S  K  N  N  C  K  S  F  E  T  N  S  S  L  L  D  K  N  G  K  P
4441 CTGTGGAAATAATGGTGTGTGATGTTCAGGAACTGCATTTTTTCGTTCCAAAGAATTAAACAGTTAATCCCCCTTACTCTGTTATCTTAGTCGCCGGCATCAACTAACATGAAAAATAAT
      V  E  I  M  V  C  D  V  Q  E  L  H  F  F  V  P  K  N  *                                           pilK M  K  N  N
4561 AACGTACCATTATATCTATGGTTACTTTGGTGTACGTTTTGCCTGTACATCACTACCCCGTGGCAATACTTCCGTTATCTATACACTGACCACATATCACTATCTTATATATTATATTTC
     N  V  P  L  Y  L  W  L  L  W  C  T  F  C  L  Y  I  T  T  P  W  Q  Y  F  R  Y  L  Y  T  D  H  I  S  L  S  Y  I  L  Y  F
4681 TGTATAGTGGTTATTATCTGCTTTTCCCCACTACTTTTCAGATGCGTGCTTCTTATCGTAAGGTCTGTATCTAAATACAACTTCACATTCTTTGCGAAAAAAAATGGCAGTAAAGAATCA
     C  I  V  V  I  I  C  F  S  P  L  L  F  R  C  V  L  L  I  V  R  S  V  S  K  Y  N  F  T  F  F  A  K  K  N  G  S  K  E  S
4801 AGCATACATCTTAGCCCCTATGCCGATACACGTTGCCCGGAGTCAATGAGGAAACACTGGGATGTGTTACTTTCTCTGATATCAAAAGCAATTAATAATGATGAAACCATTGTTATGCAT
     S  I  H  L  S  P  Y  A  D  T  R  C  P  E  S  M  R  K  H  W  D  V  L  L  S  L  I  S  K  A  I  N  N  D  E  T  I  V  M  H
4921 TCACATCTCTTTACTCCTGCAAGAATAAATAAACTTATAAAGAGACTGGAATCTTATGGTTTGAATTTCTCTGTTATTACATGCCCACGGAAAACAAGTCTGTATGAAAGATTATCAATT
     S  H  L  F  T  P  A  R  I  N  K  L  I  K  R  L  E  S  Y  G  L  N  F  S  V  I  T  C  P  R  K  T  S  L  Y  E  R  L  S  I
5041 CCTTTTGTATATTTCTTATTTCAGTGGTCTCTCCCGCATTTGCATTCGGACGGAATGACAGTCGTCATTCGTCCTCTCCAGCCGCCTCACACCAATTAACAAACACGTATATACGAACAC
     P  F  V  Y  F  L  F  Q  W  S  L  P  H  L  H  S  D  G  M  T  V  V  I  R  P  L  Q  P  P  H  T  N  *
5161 GGAGTCTCCGTTATGAAAAAAAACTATCTGACACTTATACTGCCGTCATTATTGTCCGGCTGTGCCGGAATTTCCGGCAGCAATAACGTGGCACCATCAGATACGCCTGCTCTGGTTTTT
        pilL M  K  K  N  Y  L  T  L  I  L  P  S  L  L  S  G  C  A  G  I  S  G  S  N  N  V  A  P  S  D  T  P  A  L  V  F
5281 GTTGACGGACAAATCAGTGAGAGTGCTGGCGTTCTGGCACTTACGCAACAACGTCTTTCTCCACCTAAAATAACTCCCCCCCCGAGAGTCAGCGTGCCAGAGACAGTAAAAACAGCACAT
     V  D  G  Q  I  S  E  S  A  G  V  L  A  L  T  Q  Q  R  L  S  P  P  K  I  T  P  P  P  R  V  S  V  P  E  T  V  K  T  A  H
5401 CCAGTCAGCGAATCCCGACCGTTACCTTCACCAACAATAAGTCAGAATTTTTCCCGGTTAATCATTACAGGCAAGCCTGGTAGCGCCCCTGTAACACTATCGCCGACGGCCCGTAACCGG
     P  V  S  E  S  R  P  L  P  S  P  T  I  S  Q  N  F  S  R  L  I  I  T  G  K  P  G  S  A  P  V  T  L  S  P  T  A  R  N  R
5521 ACCGTCAGTCAGTGGATAAAAACATTACTACCGGAAGGATGGAAATTCAGACAAGAGAATGCAGCCACGCCCAAGCTGAATACCAAACTGGTTTCATGGTCTGCCAATGATCAGTGGACC
     T  V  S  Q  W  I  K  T  L  L  P  E  G  W  K  F  R  Q  E  N  A  A  T  P  K  L  N  T  K  L  V  S  W  S  A  N  D  Q  W  T
5641 CGTTCTCTGAATCGGTTGCTGGAAGAACAACACCTATGGGGCCATCTGGACTGGGATAATAAAACGTTGACCGTCACAACATCAGCTGCTGCACCGGCACAGGACACGGCAACCACTATC
     R  S  L  N  R  L  L  E  E  Q  H  L  W  G  H  L  D  W  D  N  K  T  L  T  V  T  T  S  A  A  A  P  A  Q  D  T  A  T  T  I
5761 TCGCCAGCAGCCATGGTAACCACATCCACTTCCCCTGAATCTCCAACAACAGCAAATAGCCAGAACAAACCACGAAATCCATTCAGAGGAAACAGCGTATCTTCGTCAACCCCCGCCGCC
     S  P  A  A  M  V  T  T  S  T  S  P  E  S  P  T  T  A  N  S  Q  N  K  P  R  N  P  F  R  G  N  S  V  S  S  S  T  P  A  A
5881 ACCCCGACTCCAACCGGCTCCACAGTCAAGTCTATCCCGCTCATGACCGGAACGCCTGTTAAACCGGTATCTCAGGGTAAAGAGTGGCGTGCGCCGGCAGGCACCACATTAAGAGAGAAC
     T  P  T  P  T  G  S  T  V  K  S  I  P  L  M  T  G  T  P  V  K  P  V  S  Q  G  K  E  W  R  A  P  A  G  T  T  L  R  E  N
6001 ATTATCAAGTGGGCAGAAGAAACAAAATGCGAAAGCATGGCATCAACGCACTGGATGGTCATATGGCCACTTTCTGTAACCGATTACCGTCTGGATGCGCCTTTAATGTTCCGGGGATCA
     I  I  K  W  A  E  E  T  K  C  E  S  M  A  S  T  H  W  M  V  I  W  P  L  S  V  T  D  Y  R  L  D  A  P  L  M  F  R  G  S
6121 TTTGAATCAGCCATGGAGCAGGTATTTGAGTTATATCGGACTGCGCAAAAACCTTTATACGCACAGGCTTCTCGCATGCAGTGCATCATTACAGTTTCTGATACAAGGGATGGTCGCTAA
     F  E  S  A  M  E  Q  V  F  E  L  Y  R  T  A  Q  K  P  L  Y  A  Q  A  S  R  M  Q  C  I  I  T  V  S  D  T  R  D  G  R  *
                                                                                                                 pilM   M
6241 TGGGATGGCTGGTCATGGCCGTCGGCCTGACCATTCTGATTATCACTGGTAGCTACCAGAATCAGAAGATGTCTGAAACCACAAACGCCCAGCAATATGCAAGTGCGTCTGTATGGGCAA
       G  W  L  V  M  A  V  G  L  T  I  L  I  I  T  G  S  Y  Q  N  Q  K  M  S  E  T  T  N  A  Q  Q  Y  A  S  A  S  V  W  A  S
6361 GTCAGATACTGATGATTGCTAACAGAATCAATGATATCAGATACGTATCAGGCCAGCAGGACGGAGTAATTTCCTCAGATAAACTGGCTCTCCCTGTCACTCCGGACTCCCGAATTAAAC
       Q  I  L  M  I  A  N  R  I  N  D  I  R  Y  V  S  G  Q  Q  D  G  V  I  S  S  D  K  L  A  L  P  V  T  P  D  S  R  I  K  H
6481 ATCAGTTACAACAGGGGCGTCTGTGGGTATGGATGCCTGAACAACCAGGTCTTGTTGAAACTTTGCGTTCAAAAAGTCGTGGTTCTGCACTAATAGGGATCTTCCAAAACGGGCAACTCA
       Q  L  Q  Q  G  R  L  W  V  W  M  P  E  Q  P  G  L  V  E  T  L  R  S  K  S  R  G  S  A  L  I  G  I  F  Q  N  G  Q  L  T
6601 CCTGGCTATCAGGAACTGCAACCGGTCTGACACCCCCGGCAGGCATTACTGCTGGTTCTGTCGTCTACGTTAACTAATCTGGATATATTCATGAAAAAATCACACCAGCGTTCAATGAAG
       W  L  S  G  T  A  T  G  L  T  P  P  A  G  I  T  A  G  S  V  V  Y  V  N  *         pilN M  K  K  S  H  Q  R  S  M  K
6721 CTGGCGGTGCTCCCCTGCATGATCGCCGTCGCTCTTTCCATCTCCGGATGTACTTTCAGCGAAATCAACAAAATGCAGAAAAAAGCACAGGAAGACTCAGCACATGCACGGGAAAAGGTA
     L  A  V  L  P  C  M  I  A  V  A  L  S  I  S  G  C  T  F  S  E  I  N  K  M  Q  K  K  A  Q  E  D  S  A  H  A  R  E  K  V
6841 TCAGCCCTTTCGGCCCGTAAATCGCAGGCTCTTACCTGGCTCGATAATCAATGGATAAACCCTGTTCCGGTCGCTCAGGTATCAAGAGAGAAAAAACAAACAGCTCCGGCCTGCTACATC
     S  A  L  S  A  R  K  S  Q  A  L  T  W  L  D  N  Q  W  I  N  P  V  P  V  A  Q  V  S  R  E  K  K  Q  T  A  P  A  C  Y  I
6961 ACGCAGGCAAGAAAAGGAGAGATCACTCTGCAGGAACTGGGGCAACGTATTACTGCCGTATGCGGCATCCCTGTGATCATCACGCCTGACGCAGCCAATTCAACTCTTGAAGGAGGCGCT
     T  Q  A  R  K  G  E  I  T  L  Q  E  L  G  Q  R  I  T  A  V  C  G  I  P  V  I  I  T  P  D  A  A  N  S  T  L  E  G  G  A
7081 ACCCGCCAGATGACAGGAACACTACCAGCACCAGATGAAAATGGGCGTCTACCGTTAAGCAGCCTGGGCAGCACAACAATGACTACCTCCACTCAGCCATTAACGCTGAATAACCTCATG
     T  R  Q  M  T  G  T  L  P  A  P  D  E  N  G  R  L  P  L  S  S  L  G  S  T  T  M  T  T  S  T  Q  P  L  T  L  N  N  L  M
7201 TGGCAGGGAGATATCAATGGTCTTCTGGATCTGATGGCCAGCCGGAGTGGTCTGTACTGGCGCATGGATAATGGTCGGATTGTATTCTATCTGACTGAAACCAGAACGTATCCACTTCAT
     W  Q  G  D  I  N  G  L  L  D  L  M  A  S  R  S  G  L  Y  W  R  M  D  N  G  R  I  V  F  Y  L  T  E  T  R  T  Y  P  L  H
7321 ATGCTGAACACCAAAACCAGCAGCAGTTCCAGTGTCAGCTCTGGCTCAACAAGCACAATGGGGGCAACAGGAGGCCAGGATAACTCAGCATCCGGTGATGCAACGTCCTCTCAGAGCACA
     M  L  N  T  K  T  S  S  S  S  S  V  S  S  G  S  T  S  T  M  G  A  T  G  G  Q  D  N  S  A  S  G  D  A  T  S  S  Q  S  T
7441 ACCGTTGGTCAGGAATACGATCTGTATGAAGACATCCGGAAAACTATTGAAGCAATGCTGACACCAGAAAAAGGCCGTTACTGGTTATCTGCATCGAGCTCAACGCTGACTGTCACTGAT
     T  V  G  Q  E  Y  D  L  Y  E  D  I  R  K  T  I  E  A  M  L  T  P  E  K  G  R  Y  W  L  S  A  S  S  S  T  L  T  V  T  D
7561 ACTCCAGCTGTCCAGGAAGCCGTCGCACGATATGTGGACGAACAAAACAGTATTATGAACCGCCAGGTAGCCCTGAACGTACAGGTTCTGAGCGTCAGCAATACCAGAAACGAACAGTTC
     T  P  A  V  Q  E  A  V  A  R  Y  V  D  E  Q  N  S  I  M  N  R  Q  V  A  L  N  V  Q  V  L  S  V  S  N  T  R  N  E  Q  F
7681 GGTCTGGACTGGAACCTTGTTTATAAATCGCTACATTCCGCCGGAGCAACGTTGAACAATGCAAGCGGAGATTTTACAGGCGCTACATCTGCAGGCGTATCAATTCTGGATACGGCAACA
     G  L  D  W  N  L  V  Y  K  S  L  H  S  A  G  A  T  L  N  N  A  S  G  D  F  T  G  A  T  S  A  G  V  S  I  L  D  T  A  T
7801 GGGAATGCCGCCAAATTCAGCGGTTCCAGTCTTCTGATTAAAGCGCTGAGTGAACAGGGCGATGTCAGTGTTGTGACTTCACAAGAAAGCACTGTCACAAACCTGACGCCGGTACCTATC
     G  N  A  A  K  F  S  G  S  S  L  L  I  K  A  L  S  E  Q  G  D  V  S  V  V  T  S  Q  E  S  T  V  T  N  L  T  P  V  P  I
7921 CAGATGGCAGATCAGACGGTTTACGTCGCCCAGTCAGCAACAACAACGACTACGGATGTAGGAGCAACAACAACATTAACGCCGGGCATGATCACCACCGGATTCAATATGACCCTGCTG
     Q  M  A  D  Q  T  V  Y  V  A  Q  S  A  T  T  T  T  T  D  V  G  A  T  T  T  L  T  P  G  M  I  T  T  G  F  N  M  T  L  L
8041 CCTTTAATTCAGAAAACGGGCAATCTCCAGTTGCAGATGAATTTTAATCTGTCAGATCCCCCAACAATCCGTAGCTTTACGTCAAAAGACGGAAACAGTTACATCGAAATGCCGTATACC
     P  L  I  Q  K  T  G  N  L  Q  L  Q  M  N  F  N  L  S  D  P  P  T  I  R  S  F  T  S  K  D  G  N  S  Y  I  E  M  P  Y  T
8161 AAACTGCGTTCACTGAGCCAGAAGGTCAATCTGAAAGAAGGGCAATCACTTGTCGTTACTGGTTTCGATCAGAACAATACGACGACAAGTAAAGCCGGTACGTTTACGCCAGCAAATCCA
     K  L  R  S  L  S  Q  K  V  N  L  K  E  G  Q  S  L  V  V  T  G  F  D  Q  N  N  T  T  T  S  K  A  G  T  F  T  P  A  N  P
8281 TTATTTGGTGGTTCACAAACCGGGAAAAATGAACGCAGCACGCTTGTAATCATCATTACCCCGACTTTCCCGTCAGGAGGCAACAATGGCTGATGAAGATATCAACCCGGTTGTACTTCT
     L  F  G  G  S  Q  T  G  K  N  E  R  S  T  L  V  I  I  I  T  P  T  F  P  S  G  G  N  N  G  *
                                                                                         pilO M  A  D  E  D  I  N  P  V  V  L  L
8401 CGCAGATCCTAAAGTAAATCATCGCGTGTGGGCAGCCTGTCTTAAATGGAGTCCTGTCGTAAAAAAACAGCGAGTCCCTTCTCATCAGAAACACAAGCCCCATGTAAAATCCAGACGGCT
       A  D  P  K  V  N  H  R  V  W  A  A  C  L  K  W  S  P  V  V  K  K  Q  R  V  P  S  H  Q  K  H  K  P  H  V  K  S  R  R  L
8521 TACCTCTCTGAAAGTGACTGTAGGCAGCAGAAGTTCCCGTGGGAAGATCAGCCGCATTACCGGCACAGGAATACTGGCCAGGCCGGAGCGCAATCATTATTTTCCCTTGCTCTGGCTTT
       T  S  L  K  V  T  V  G  S  R  S  S  R  G  K  I  S  R  I  T  G  T  G  I  L  A  R  P  E  R  N  H  Y  F  S  L  A  L  A  F
8641 CTGCTCATGGGTGAGAAATGGATACGGCGTCTTCCGGTACAGCGATAAGGAGCTTTTATTCCTTGCCTCCATTAATGGCCAGCCTGCCGTTATGGCTGATCTGTCAGGAAATGATGCTGA
       C  S  W  V  R  N  G  Y  G  V  F  R  Y  S  D  K  E  L  L  F  L  A  S  I  N  G  Q  P  A  V  M  A  D  L  S  G  N  D  A  D
8761 TGTTGCTCAAAAGGTCAGTCTTTTTCTGACGATGAACGAAGAGCCCCCCGAAAAATGGCAGGTCGTATCTCCTCTGGAACATCCTGATAACTGGGAAAGTATTATTACCCGGTTATCATC
       V  A  Q  K  V  S  L  F  L  T  M  N  E  E  P  P  E  K  W  Q  V  V  S  P  L  E  H  P  D  N  W  E  S  I  I  T  R  L  S  S
8881 AGCAGATCTCCGCCGTTGCAAACTTACTGTCGGCAATCGCAGCAAATTCACACTGCCTGCCGTACTTTTTTTAGTAGCAGCCAGCGCAGGCACTGTATTCTGGATGACTCAGCCAGAGCC
       A  D  L  R  R  C  K  L  T  V  G  N  R  S  K  F  T  L  P  A  V  L  F  L  V  A  A  S  A  G  T  V  F  W  M  T  Q  P  E  P
9001 GGATGTCGTACCGACGCCCGAGGAGATAGCTGCACGGGCAAGATTGCAGTTTAAAAAGCCCGATCCACCGCCAGAACTCCCGCATCCATGGGCATCACAACCTGTAATCAGTGATTTTCT
       D  V  V  P  T  P  E  E  I  A  A  R  A  R  L  Q  F  K  K  P  D  P  P  P  E  L  P  H  P  W  A  S  Q  P  V  I  S  D  F  L
9121 GAAAGCCTGCGCGGACCTTCGAAAACCATCACCGGTCGCGCTAGAAGGCTGGAAACTTACCGGTGGAACATGTACTCCCGAGACATTCACACTGATTTACGAACGACAGCCAGGAGGAAC
       K  A  C  A  D  L  R  K  P  S  P  V  A  L  E  G  W  K  L  T  G  G  T  C  T  P  E  T  F  T  L  I  Y  E  R  Q  P  G  G  T
```

FIG. 2. Nucleotide and deduced amino acid sequences of the R64 *pil* region. The sequence is numbered from the leftmost *Bgl*II site shown in Fig. 1A. Putative promoter sequences and initiation codons for various *pil* genes are boxed. Predicted Shine-Dalgarno sequences are underlined. The signal peptides and hydrophobic domains of the *pil* proteins are also underlined.

```
9241  TATTGAAGGCTTTCTGGCAAGAAGCAAAGAAATCTTTAATGTTATTCCTGACTTTAACCTTAAAGATGGTGCACGCCTGGCTTCTGTAACCCGGCAACTGCCTTCCCTGCCCCGCAGGGA
       I  E  G  F  L  A  R  S  K  E  I  F  N  V  I  P  D  F  N  L  K  D  G  A  R  L  A  S  V  T  R  Q  L  P  S  L  P  R  R  D
9361  TGAGGCAGTCCCTACCCCCTCAGAACAACTGATGCGGGTGTTCACGTGGTTTCAGAAAAAACAGTTAACGCCGGCCATTAATGAAATAGCCATACCGGAACCATTACCAGGTAACGATGG
       E  A  V  P  T  P  S  E  Q  L  M  R  V  F  T  W  F  Q  K  K  Q  L  T  P  A  I  N  E  I  A  I  P  E  P  L  P  G  N  D  G
9481  TGAACCAGCCCCTGTTCAGAAATGGAAAGAATATCAGTTCAGTTTATCAACGCCTGTAAATCCTGATGAGTTGTTCCCTTTATTTCAGGATACCGGCGTTCGTCTGAGCAATATCCATTT
       E  P  A  P  V  Q  K  W  K  E  Y  Q  F  S  L  S  T  P  V  N  P  D  E  L  F  P  L  F  Q  D  T  G  V  R  L  S  N  I  H  F
9601  TGAATTAAACGGCGGTACTTTCAGTTATTCCTCAGAAGGTCATATTTATGCGTCCAGGTAAATTACTCATTATTCCATCCATATTTTTCTTCTCCGGTTTTTCATTTGCAACAACACAAC
       E  L  N  G  G  T  F  S  Y  S  S  E  G  H  I  Y  A  S  R  *
                                                     pilP  M  R  P  G  K  L  L  I  I  P  S  I  F  F  F  S  G  F  S  F  A  T  T  Q  P
9721  CTCTTGTCACAATAGGAGAACTTGAAGCACAACAAAACCGGAATATTCTTTTACAGGCAAAAGTTCAGGGCGCACAGCTACAGAAACAACTCGAAGAATCAGATGTTACTTCATCATCTG
       L  V  T  I  G  E  L  E  A  Q  Q  N  R  N  I  L  L  Q  A  K  V  Q  G  A  Q  L  Q  K  Q  L  E  E  S  D  V  T  S  S  S  E
9841  AAACTGTTTCAGGTTTCTCTGGTGTGACTGCCTCTTTGCCATCAGTTTCAGAGCAGCCGACATCTAAAAACAGAAAGGAATTGCCCGTTATTATGGAAATAAACGGCAAAGATAAACGTC
       T  V  S  G  F  S  G  V  T  A  S  L  P  S  V  S  E  Q  P  T  S  K  N  R  K  E  L  P  V  I  M  E  I  N  G  K  D  K  R  L
9961  TTAATGCCGTTCTTCGCATGGCTGATGGACGTCAGACAAGCGTCACCACTGGCAGCCAGCTTCCAGGCACATCCGTCACAGTGAAGTCGATCTCATTGTCTGGCGTCACGTTAAGTGATG
       N  A  V  L  R  M  A  D  G  R  Q  T  S  V  T  T  G  S  Q  L  P  G  T  S  V  T  V  K  S  I  S  L  S  G  V  T  L  S  D  G
10081 GCACCACCCTTACTTTCTGAGTAACGTGATATGGATAAATATAATCTGAAGGATGCTCTCCTTTTTATATCCAGGGGAGACACCCATGAAATTCTTATAGAAACAAACCAGCGGACAAGG
       T  T  L  T  F  *     pilQ  M  D  K  Y  N  L  K  D  A  L  L  F  I  S  R  G  D  T  H  E  I  L  I  E  T  N  Q  R  T  R
10201 CCTGACGTACAAAGTAATCTCCAGGAGCTGCTAAAATTATATCCGGATATCAACCCAAAAGTAGTCTCGTTGTCGGAGTTACAGGAGGCAGGTCAGGATGAGAAAAACAGAGGCCCGAAA
       P  D  V  Q  S  N  L  Q  E  L  L  K  L  Y  P  D  I  N  P  K  V  V  S  L  S  E  L  Q  E  A  G  Q  D  E  K  N  R  G  P  K
10321 GAACGAATTATTCATCTGAAAGACCTTGCTGATGTCAGTGAAAGTGAGAAGAAAGTTCTATCTTATTTTGAAACTGCCAGAAAACTTGGTGCTTCAGATATTCATTTCTTAATTTCAGAG
       E  R  I  I  H  L  K  D  L  A  D  V  S  E  S  E  K  K  V  L  S  Y  F  E  T  A  R  K  L  G  A  S  D  I  H  F  L  I  S  E
10441 TCTATTTTCAAAGTTCGCATGAGAATATTCGGTGAACTGCAGACTGTAGATGAAGATCAACCAGCACTGGGCTATTCACTTTGCGCAACAGCCATTCTGTCAATGGCTGATGTCACTGAA
       S  I  F  K  V  R  M  R  I  F  G  E  L  Q  T  V  D  E  D  Q  P  A  L  G  Y  S  L  C  A  T  A  I  L  S  M  A  D  V  T  E
10561 ACCAGTTTCTTCCCTCAAAGGGAGCAGGATGCGCGCCTGTCCCCACAATTAATGCGAAAAATAGGAATATTTGGTGCACGCTACAGTCACAGACCTACGGGAGACGGTCTGATTGCAGTT
       T  S  F  F  P  Q  R  E  Q  D  A  R  L  S  P  Q  L  M  R  K  I  G  I  F  G  A  R  Y  S  H  R  P  T  G  D  G  L  I  A  V
10681 ATGCGCCTGATACCCGATGACGGAGATAAAGTTCCCACGTTTAAACAACTGGGCTTTATCCCGGAACAAATCCGGTTGCTGAACATAATGCTGCGCCGCCCGGAAGGAAAAATAGTCCTG
       M  R  L  I  P  D  D  G  D  K  V  P  T  F  K  Q  L  G  F  I  P  E  Q  I  R  L  L  N  I  M  L  R  R  P  E  G  K  I  V  L
10801 TCAGGTCCAACCGGTTCCGGTAAATCGACAACGTTACGCAGCGCCTGCCGTGTATATCTTGACGATAATCAGGGGCGACACCTGCTAACAATCGAAGATCCGCTGGAAGGACAGATACTG
       S  G  P  T  G  S  G  K  S  T  T  L  R  S  A  C  R  V  Y  L  D  D  N  Q  G  R  H  L  L  T  I  E  D  P  L  E  G  Q  I  L
10921 GGAGCAACACAGACTCCGATTATCTGCGACAAATCCGACGAAGATGCCGTCAAACTGGCATGGAGTCGGGCTATTTCATCGGCAATGCGACTTGACCCTGATGCCATCATGGAAGGGGAA
       G  A  T  Q  T  P  I  I  C  D  K  S  D  E  D  A  V  K  L  A  W  S  R  A  I  S  S  A  M  R  L  D  P  D  A  I  M  E  G  E
11041 ATGCGCGATCTTATCTCGATGATGTCAACCACCTACGCGGCACAGACCGGGCACATAGTTCTGACGACATTACATACAAACTCAGCTCTGGGTATACCGGAGCGAATGATTACGATGGGT
       M  R  D  L  I  S  M  M  S  T  T  Y  A  A  Q  T  G  H  I  V  L  T  T  L  H  T  N  S  A  L  G  I  P  E  R  M  I  T  M  G
11161 ATGAATGCCGATCTTATTTGCGATGCGCAATTACTGATTGGGATGATCAGCCAGCGCCTTGTACCAACTCTGTGTCCCTCCTGCCGTATTCCATGGGAAAAGCGAGCACCTGAACTTAGC
       M  N  A  D  L  I  C  D  A  Q  L  L  I  G  M  I  S  Q  R  L  V  P  T  L  C  P  S  C  R  I  P  W  E  K  R  A  P  E  L  S
11281 GACGACGAGCGTGATTACCTTGAGCGGCACTGCAATAAAGATTCACTCTGCAGTACAGATAATATCTGGTTCCGAAATCATCACGGATGCTCTGAATGTAATCACGACGTAATCATTAAC
       D  D  E  R  D  Y  L  E  R  H  C  N  K  D  S  L  C  S  T  D  N  I  W  F  R  N  H  H  G  C  S  E  C  N  H  D  V  I  I  N
11401 GGCAGAAAGCGCGGTGAAATAGGTAAAGGCCTCACCGGCAGGACTGTAATTGCTGAGGTGATAGAACCTGATAACCGCTTATTTCAGATTCTGAAAACACGCGGCAAAGTTGCAGCACGT
       G  R  K  R  G  E  I  G  K  G  L  T  G  R  T  V  I  A  E  V  I  E  P  D  N  R  L  F  Q  I  L  K  T  R  G  K  V  A  A  R
11521 AAATACTGGCTGGAAAACATGAAAGGCATCAGTCGTGTAGAGCATCTTCTTCGCCGTATCAATGAAGGTCTGGTAGACCCTCTTGAAGCCGACAGAATCATCCCTCTGGACGAAGACGAG
       K  Y  W  L  E  N  M  K  G  I  S  R  V  E  H  L  L  R  R  I  N  E  G  L  V  D  P  L  E  A  D  R  I  I  P  L  D  E  D  E
11641 CGCCTCAGCATCGATGATGTCTAACAGGAGAACAACATGCGTGAAATGAATTTCAGTCAACGGCTGAGGCGTTTCATTGTCAGAAAAACGTTCAGCGCCCCCTACAGGGTGCAGTTTTAC
       R  L  S  I  D  D  V  *        pilR  M  R  E  M  N  F  S  Q  R  L  R  R  F  I  V  R  K  T  F  S  A  P  Y  R  V  Q  F  Y
11761 GAAGCCCTGCGTTTTCTGCTGGAAAATAAACAGCCATTAAAGACCGCACTGGAACAGATGCGGGATGCCTGGACAGACTTTGGACGAAAATGGCATCCCTTTGCTGAACTGGCCACAGAC
       E  A  L  R  F  L  L  E  N  K  Q  P  L  K  T  A  L  E  Q  M  R  D  A  W  T  D  F  G  R  K  W  H  P  F  A  E  L  A  T  D
11881 TGCATTGAGTCTCTTCGTGAAAACAGTGGCGAAAACTCTCTGGAATATACTCTTAGTTTATGGGTTCCTCAGGAAGAGGCCGCAGTAATCAGCGCCGGGATAAGAAGTGGTTCAATTGTG
       C  I  E  S  L  R  E  N  S  G  E  N  S  L  E  Y  T  L  S  L  W  V  P  Q  E  E  A  A  V  I  S  A  G  I  R  S  G  S  I  V
12001 GATGCACTGCAATTCGCCACCACTCTTACAGATGCAAAAGAACAAATCCATCAGGCTATATGGCAAATGGCTATTTATCCGGTGGGATTGTTAATCATGATGACGGGGACATTATATGTT
       D  A  L  Q  F  A  T  T  L  T  D  A  K  E  Q  I  H  Q  A  I  W  Q  M  A  I  Y  P  V  G  L  L  I  M  M  T  G  T  L  Y  V
12121 CTGAATACAGAGCTAATACCTGAACTGAGCAAAATAAGCTCCCCCGATTCATGGAGTGGTGCTCTTGGTTTTCTTTATGGATTATCTGTTTTTGTCGATAATTACGGCGCTATATGTGCC
       L  N  T  E  L  I  P  E  L  S  K  I  S  S  P  D  S  W  S  G  A  L  G  F  L  Y  G  L  S  V  F  V  D  N  Y  G  A  I  C  A
12241 GTTCTTTTTGCAGTGATCACAGGTTTAATATCATGGTCTCTGCCAAACTGGAAAAGCCCTGACAGTGTTCGTACTTTTGCCGACAAGATAATGCCATGGAGTATTTATCAGGATATACAG
       V  L  F  A  V  I  T  G  L  I  S  W  S  L  P  N  W  K  S  P  D  S  V  R  T  F  A  D  K  I  M  P  W  S  I  Y  Q  D  I  Q
12361 GGAGCAACATTTCTCCTGAATATGGCTGCACTTCTGAAAGCCAAAATGACGACGCTCAACTCGCTAAACATTCTTCAGGAGTTTGCATCACCATGGTTAAGCACACGTCTTGACAGCATT
       G  A  T  F  L  L  N  M  A  A  L  L  K  A  K  M  T  T  L  N  S  L  N  I  L  Q  E  F  A  S  P  W  L  S  T  R  L  D  S  I
12481 ATTTACCGCGTCCGCCAGGGGGATCATCTTGGACTGGCTTTGCGTCAGTGCGGTTATCAGTTCCCATCCAGGGAAGCAGCCAACTTCCTGTCTTTATTACAGGGAGATGGCGCAACAGAG
       I  Y  R  V  R  Q  G  D  H  L  G  L  A  L  R  Q  C  G  Y  Q  F  P  S  R  E  A  A  N  F  L  S  L  L  Q  G  D  G  A  T  E
12601 CTAATCAGCAATTATGGTCAACGATGGCTTAGTCAAACACTTCAACGTGTAAAGAAACGGGCAAACGTGATTCGTTTAATTATGCTGATATTTCTTGTCATGTCTCTCATGTTACTGGTT
       L  I  S  N  Y  G  Q  R  W  L  S  Q  T  L  Q  R  V  K  K  R  A  N  V  I  R  L  I  M  L  I  F  L  V  M  S  L  M  L  L  V
12721 TTTGCCATTATGGATATTCAGTCTATTAGTGACAACAGCATGGGTAACTTTTAAACAAAATGAGGATATTTATGTTAGTAGAAAATATTAATACCACTTTAACTGGAAATAATAAGAAGA
       F  A  I  M  D  I  Q  S  I  S  D  N  S  M  G  N  F  *            pilS  M  L  V  E  N  I  N  T  T  L  T  G  N  N  K  K  N
12841 ATGAGCCTCATGATAAAGGATGGGCAATTTTGGAGCAGGGAACAATAGCGCTCGTTGTTCTTTTTGTTATAGTCGTTGTTCTGGGTAGTTTGTATGCGTTGCGTACAAGAGCCAATGTGG
        E  P  H  D  K  G  W  A  I  L  E  Q  G  T  I  A  L  V  V  L  F  V  I  V  V  V  L  G  S  L  Y  A  L  R  T  R  A  N  V  A
12961 CAACTGAAACCGCAAATATACAAACAATTATCACAAGTGCCCAGAGCCTGCTAAAAGGTAGCGATGGATATACATTTACCAGCAGTGCCAAAATGACAGGTGCCCTCATTCAGATGGGTG
        T  E  T  A  N  I  Q  T  I  I  T  S  A  Q  S  L  L  K  G  S  D  G  Y  T  F  T  S  S  A  K  M  T  G  A  L  I  Q  M  G  V
13081 TCATTCCCTCTGGGATGACCGTACAGGGCGACAAAACATCTGGTACAGCAACGCTCTATAACGCATGGGGCGGTGCTGTAACTGTTGCCCCGGCTTCAACATCAGGGTTCAACAACGGTT
        I  P  S  G  M  T  V  Q  G  D  K  T  S  G  T  A  T  L  Y  N  A  W  G  G  A  V  T  V  A  P  A  S  T  S  G  F  N  N  G  F
13201 TTACCGTTACCTATGATAAAGTCCCACAGGATGCCTGTATCCAGATTGCCACACGAATCAGTAAAACTGGTTTAACTAACGGCATTACCCTGAACAATACTGCTCACAGTGATGGAAAAG
        T  V  T  Y  D  K  V  P  Q  D  A  C  I  Q  I  A  T  R  I  S  K  T  G  L  T  N  G  I  T  L  N  N  T  A  H  S  D  G  K  V
13321 TCACTACTGAAGAAGCCAGTACGCAATGCAAGGCAGATAATGGCAGCACTGGTACAAACAAACTTATCTTCACCATCAATGGTTAAGGAGTTTTAATGTCAGGGGGTACAGTACCCCTCA
        T  T  E  E  A  S  T  Q  C  K  A  D  N  G  S  T  G  T  N  K  L  I  F  T  I  N  G  *   pilT  M  S  G  G  T  V  P  L  R
13441 GACTTATTTCATTAATCCTGCTTTTTTCTCTTATTCCACCCGCCCGAGCTTCATACGTAGCTCCCAAACAGATCTGGACCAACAAGTGGGACAATTGCTTTGCCGCTGCCGGCGCGCGCT
        L  I  S  L  I  L  L  F  S  L  I  P  P  A  R  A  S  Y  V  A  P  K  Q  I  W  T  N  K  W  D  N  C  F  A  A  A  G  A  R  Y
13561 ATCAAATTGAACCATTATTATTGAAGGCTATTTCTGCCGGGGAATCTTCGTTGAGGCCGGGCGCAATTAATATAAATAAGGACAAAAAAACAGGTAAAGCCAGTAGCACGGATTATGGGT
        Q  I  E  P  L  L  L  K  A  I  S  A  G  E  S  S  L  R  P  G  A  I  N  I  N  K  D  K  K  T  G  K  A  S  S  T  D  Y  G  L
13681 TGATGCAAATAAACTCAACCCATATTCCAAAGCTGATAAATATGGGCGTAATAAAAAAGTCTGAGGATTTAATTACCAAACCATGCCTTAATATCCATATAGGAAGCTGGATACTGGCAA
        M  Q  I  N  S  T  H  I  P  K  L  I  N  M  G  V  I  K  K  S  E  D  L  I  T  K  P  C  L  N  I  H  I  G  S  W  I  L  A  R
13801 GACATTTTCAGATATGTGGCGTCAGCTGGAATTGCCTTGGGTCATACAATGCCGGATTCCGTAAGGACAGACATGAAACAAGAGAGCAATATGCCAATAAAATATGGAGAATTTATCGTG
        H  F  Q  I  C  G  V  S  W  N  C  L  G  S  Y  N  A  G  F  R  K  D  R  H  E  T  R  E  Q  Y  A  N  K  I  W  R  I  Y  R  D
13921 ACATGAAAGGGATATGTCTTCCCGGACAGGGAGGCAGGCAATGCAGACAATCGTAGGCATAAGTCTATCCCCTGTTTATATACTGCCATTGATGTTCACATTTTCCATTATTGGACGAAC
        M  K  G  I  C  L  P  G  Q  G  G  R  Q  C  R  Q  S  *
                                     pilU  M  Q  T  I  V  G  I  S  L  S  P  V  Y  I  L  P  L  M  F  T  F  S  I  I  G  R  T
14041 TCTGCTATTCAGAGTTCGTTATTTTTTATCTGATACGGGACATCTCTGGTATAAGACACACCCTGCTGTTCTGTCAGGAATATGGCTATACAGTATAGCTGTGGCGTTAATTATTTTATC
        L  L  F  R  V  R  Y  F  L  S  D  T  G  H  L  W  Y  K  T  H  P  A  V  L  S  G  I  W  L  Y  S  I  A  V  A  L  I  I  L  S
14161 ATCATCACCACTGCTCTACCGTATTCATGCTGTCCTTATTCTTTCGTTCATACTGCAAATGGCGGTAACGGATGCACTGACAGGTTTATTACCAGGAACATTCACCCGACGTTTTTTAAT
        S  S  P  L  L  Y  R  I  H  A  V  L  I  L  S  F  I  L  Q  M  A  V  T  D  A  L  T  G  L  L  P  G  T  F  T  R  R  F  L  I
14281 TGCAGGAATGCTATCACAAATCACGACTGATATATGGTGGTTCCGTACAACAGAATTTGCCACTGCTGCAATTGTCTTATTCTGCCTTCATAAATTAGTTAACCGCCATCGTCTTAATAT
        A  G  M  L  S  Q  I  T  T  D  I  W  W  F  R  T  T  E  F  A  T  A  A  I  V  L  F  C  L  H  K  L  V  N  R  H  R  L  N  I
14401 TGGCACTGGAGACCTGTGGCTTATTGCAGGTATTACCGCATGGTCAGGGTTATATAACGCCATATGGTGTGTGTTGCTGGGCACAGGAGGATTTGTCTTATGGCATTCAACATGGTGCAT
        G  T  G  D  L  W  L  I  A  G  I  T  A  W  S  G  L  Y  N  A  I  W  C  V  L  L  G  T  G  G  F  V  L  W  H  S  T  W  C  I
14521 AAAAGGGCATAAAGAAGGCCCTCTGGGTCCATGGCTATGTTTTGGCCATGTCCTTTTGCTGCTGGATAATCTTTATCAACCACTATGGGTAATCTGATGAAAAAATATGACCGAGGCTGG
        K  G  H  K  E  G  P  L  G  P  W  L  C  F  G  H  V  L  L  L  L  D  N  L  Y  Q  P  L  W  V  I  *
                                                                                              pilV  M  K  K  Y  D  R  G  W
14641 GCATCGCTGGAAACTGGCGCGGCTTTACTGATTGTTATGCTCCTGATTGCCTGGGGAGCGGGCATATGGCAGGACTATATTCAAACAAAAGGATGGCAAACTGAAGCACGTCTTGTCAGC
        A  S  L  E  T  G  A  A  L  L  I  V  M  L  L  I  A  W  G  A  G  I  W  Q  D  Y  I  Q  T  K  G  W  Q  T  E  A  R  L  V  S
```

FIG. 2—*Continued.*

RESULTS

Nucleotide sequence of the *pil* region in plasmid R64*drd-11*. pKK641 contained all the information necessary for the formation of the R64 thin pilus as described in the Introduction. Hence, to examine the gene organization of the R64 *pil* region, the nucleotide sequence of a 10.7-kb *Sma*I-*Hpa*I segment (Fig. 1B) of pKK641 was determined by the dideoxy chain termination method (32). Figure 2 shows an 11,160-bp nucleotide sequence of the R64 *pil* region. A 10,734-bp sequence from the *Sma*I site at position 3644 to the *Hpa*I site at position 14378 was newly determined in this study.

On examination of the sequence, 13 open reading frames (ORFs) were newly identified and tentatively designated as genes *pilI* through *pilU* (Fig. 2). The *pilV* gene, which was previously shown to be required for R64 thin-pilus formation, sequentially follows the *pilU* gene (20). The gene organization of the R64 *pil* region, deduced from the nucleotide sequence, is shown schematically in Fig. 1B. These genes are situated side by side in the same orientation as the *traABCD* genes. All genes start with an ATG initiation codon, preceded by potential Shine-Dalgarno sequences. Most genes stop with a TAA termination codon. Intergenic spaces between neighboring genes vary from 49 to −16 nucleotides (that is, a 49-nucleotide space between genes and a 16-nucleotide overlap). Since no promoter or terminator structure was found in the coding and intergenic regions of the *pilI* through *pilV* genes, these 14 genes may form a single long operon. The overlapping nature of the neighboring genes suggests the presence of translational coupling between these genes.

The termination codon of the preceding *traD* gene is located 267 bp upstream from the *pilI* gene. In this intergenic region between the *traD* and *pilI* genes, sequences CTGTCA (positions 3657 to 3662) and TTTTTT (positions 3679 to 3684) may function as −35 and −10 promoter recognition sequences for σ70 of *E. coli* RNA polymerase. In the shufflon region, two sets of 19-bp repeat sequences, which function as the crossover sites of DNA inversion of the shufflon, are located in the inverted orientation with 8-bp spaces. These 19-bp inverted repeat sequences may also function as transcription terminator signals for the *pil* operon.

Identification of the *pil* gene products. To identify the *pil* gene products, the maxicell procedure was performed (31). Various segments of pKK641 DNA were subcloned into the high-copy-number vector pUC9 to generate pKK686 (containing the *pilJ* to *pilM* genes), pKK687 (*pilN*), pKK688 (*pilO* to *pilS*), pKK689 (*pilP* to *pilT*), and pKK690 (*pilR* to *pilV*) (Fig. 1B). These plasmids were introduced into *E. coli* CSR603 cells by transformation, and the protein products were analyzed by the maxicell procedure. Maxicells harboring these plasmids produced various *pil* proteins (Fig. 3). The identification of *pil* proteins is described below.

Properties of R64 *pil* gene products. The number of amino acids, molecular mass, pI, and cellular location for each of the 14 *pil* gene products were predicted from the deduced amino acid sequences and are summarized in Table 1.

The *pilI* gene encodes an 84-amino-acid basic protein, and the *pilJ* gene encodes a 150-amino-acid acidic protein. The *pilJ* product was detected as an 18-kDa protein by the maxicell procedure (Fig. 3, lane 2). A computer search of PilI and PilJ amino acid sequences failed to detect any significant sequence similarity with known proteins in the database.

The *pilK* gene product. The *pilK* gene encodes a 196-amino-acid basic protein. The *pilK* product could not be detected by the maxicell procedure. The amino acid sequence of the PilK protein has 27% identity with that of the R64 TraD protein (15) (Fig. 4A). The R64 *traD* gene is located 1.7 kb upstream from the *pilK* gene within the R64 transfer region (Fig. 1B). It was shown that the frameshift mutation within the *traD* coding sequence did not affect transfer frequency in liquid and surface mating of R64 (15). No known protein containing an amino acid sequence with significant similarity to that of the PilK or TraD protein was found.

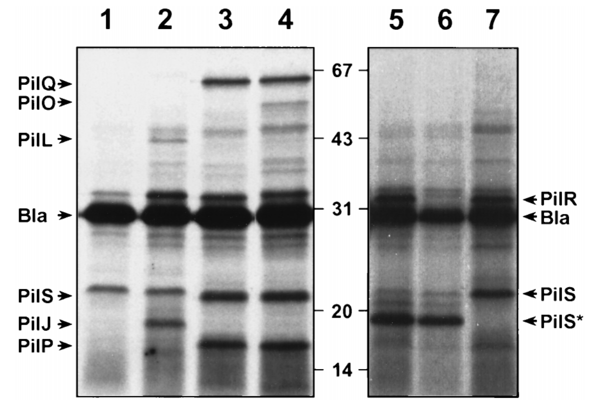

FIG. 3. Detection of *pil* gene products by the maxicell procedure. *E. coli* CSR603 cells harboring each plasmid were labeled with [^{35}S]methionine by the maxicell procedure (31), and each sample was analyzed by SDS-PAGE and then by autoradiography. Lanes: 1, pUC9; 2, pKK686; 3, pKK689; 4, pKK688; 5, pKK690; 6, pKK690 *pilR1*; 7, pKK690 *pilU1*. Identified protein bands are indicated on the left and right. PilS* represents the mature form of the PilS protein. Numbers between lanes 4 and 5 show the sizes (in kilodaltons) of marker proteins.

The *pilL* and *pilM* gene products. The *pilL* gene encodes a 355-amino-acid basic protein, and the *pilM* gene encodes a 145-amino-acid basic protein. The amino acid sequence of the N-terminal region of the *pilL* product reveals features indicative of signal peptides of bacterial lipoproteins (Fig. 2). A faint 43-kDa protein band was detected as the *pilL* product by the maxicell experiment (Fig. 3, lane 2). The observed molecular mass (43 kDa) is significantly higher than the calculated value (39 kDa), suggesting a modification of the *pilL* product as a lipoprotein. No protein containing sequence similarity to the PilL and PilM proteins could be found in the database.

The *pilN* gene product is a lipoprotein. The *pilN* gene encodes a 560-amino-acid neutral protein. Since the amino acid sequence of the N-terminal region of the *pilN* product was also predicted to encode a lipoprotein, this possibility was tested by using globomycin (14). Globomycin is known to inhibit signal peptidase II, which cleaves the signal peptide of bacterial lipoproteins. Maxicells harboring pKK687 produced a protein with an apparent molecular mass of 67 kDa, which was higher than the calculated value of 60 kDa (Fig. 5, lane 2). Proteins with decreased mobility appeared upon the addition of globomycin at 50 to 200 μg/ml in the maxicell culture (Fig. 5, lanes 3 to 5), suggesting the accumulation of a prePilN protein with a higher molecular mass. Thus, it is likely that the *pilN* gene encodes a lipoprotein that is processed in the same manner as *E. coli* lipoproteins. The N-terminal 26-amino-acid signal peptide of the *pilN* product is believed to be removed by the action of signal peptidase II (Fig. 2). The mature PilN protein may be integrated into the bacterial outer membrane.

The amino acid sequence of the PilN protein has 25% identity to the BfpB sequence (Fig. 4B). The *bfpB* gene is a member of a gene cluster required for the formation of bundle-forming pili (*bfp*) in enteropathogenic *E. coli* (33, 34). The BfpB protein is also a lipoprotein. Both PilN and BfpB proteins contain serine-rich segments (Fig. 4B). The PilN amino acid sequence

TABLE 1. Properties of R64 *pil* gene products

Gene	No. of amino acids	Molecular mass		Predicted pI	Predicted location[b]	Comment
		Calculated (Da)	Found (kDa)[a]			
pilI	84	9,953	NT	10.5	C	
pilJ	150	17,476	18	5.3	IM	
pilK	196	23,184	ND	9.4	IM	
pilL	355	38,603	43	10.0	IM/OM	Lipoprotein
pilM	145	15,716	ND	9.9	P/OM	
pilN	560	59,848	70	7.8	IM/OM	Prelipoprotein
			67			Lipoprotein
pilO	431	48,162	50	9.2	IM	
pilP	150	16,003	16	9.8	IM	
pilQ	517	58,280	58	5.6	C	Nucleotide-binding protein
pilR	365	41,415	34	7.9	IM	Integral membrane protein
pilS	204	21,265	22	8.2	IM	Prepilin
			19			Pilin
pilT	186	20,758	ND	10.2	IM	
pilU	218	24,600	ND	9.5	IM	Prepilin peptidase
pilVA'	474	49,527	ND	9.5	P/OM	Prepilin

[a] Summarized from Fig. 3 and 5. ND, not detected; NT, not tested.
[b] C, cytoplasm; IM, inner membrane; OM, outer membrane; P, periplasm.

also has a slight similarity to those of the TcpC and PulD proteins, which are related to the formation of the type IV pilus.

The *pilO* and *pilP* gene products. The *pilO* gene encodes a 431-amino-acid basic protein, and the *pilP* gene encodes a 150-amino-acid basic protein. The products of the *pilO* and *pilP* genes were identified as 50- and 16-kDa proteins, respectively (Fig. 3, lane 4). No protein similar to either the PilO or the PilP protein could be found in the database.

The *pilQ* gene product contains a nucleotide-binding motif. The *pilQ* gene encodes a 517-amino-acid acidic protein. A 58-kDa protein was identified as the *pilQ* product (Fig. 3, lanes 3 and 4). Since the PilQ protein is rich in hydrophilic amino acids and does not carry a hydrophobic segment, it may be located in cytoplasm. The PilQ protein has amino acid sequence identity with many bacterial proteins related to the biogenesis of type IV pili, the protein secretion system, and the DNA uptake system (12, 28) (Fig. 4C). All these proteins carry a nucleotide-binding motif (39). Their amino acid sequences are highly conserved in Walker boxes A and B, Asp boxes, and CXXC sequences. These proteins may function in the energy-fueling steps in the biogenesis and function of type IV pili and related systems.

The *pilR* gene product. The product of the *pilR* gene is a 365-amino-acid neutral protein. Maxicells harboring $pilR^+$ plasmids produced a 34-kDa protein (Fig. 3, lanes 5 and 7). Although its size was smaller than the calculated value (41 kDa), the 34-kDa protein was shown to be the *pilR* product, since pKK690 *pilR1* did not produce the 34-kDa protein (Fig. 3, lane 6). The PilR protein may be an integral inner membrane protein, since it carries three putative transmembrane domains (Fig. 2). The PilR protein has amino acid sequence similarity with many bacterial proteins related to biogenesis of the type IV pilus, including the BfpE protein (Fig. 4D). The BfpE protein was also reported to migrate faster in SDS-PAGE than expected (33, 34).

The *pilS* gene product is a prepilin. The *pilS* gene encodes a 204-amino-acid basic protein. Maxicells harboring $pilS^+$ plasmids pKK688 and pKK689 produced a 22-kDa protein (Fig. 3, lanes 4 and 3, respectively), while those harboring $pilS^+$ plasmid pKK690 produced a 19-kDa protein (Fig. 3, lane 5). It is likely that the *pilS* product is the 22-kDa protein and is processed to the 19-kDa protein in the presence of the *pilU* product, since pKK690 *pilU1* produced the 22-kDa protein and not the 19-kDa protein (compare Fig. 3, lanes 5 and 7). The *pilS* gene appears to encode the major pilin, since the purified R64 thin pilus was shown to contain a 19-kDa protein as a major component (18). From the similarity of the *pilS* product to the type IV prepilins (Fig. 4E), the *pilS* product was assumed to lose the N-terminal 23-amino-acid signal peptide through the putative type IV prepilin peptidase activity of the *pilU* product. The signal peptide of the PilS protein is unusually long, as is that of the TcpA protein of *V. cholerae*. The PilS 23rd and 28th amino acid residues are glycine and glutamic acid, respectively, which are completely conserved among all type IV prepilins and related proteins (Fig. 4E).

The *pilT* gene product. The *pilT* product is a 186-amino-acid basic protein that has sequence similarity to the *bfpH* gene and gene X (gene 19) of IncF plasmids, including F, R1, and R100 (2) (Fig. 4F). The identity between the amino acid sequences of R64 PilT and the F 19-kDa protein (product of gene X) is as high as 43%. In IncF plasmids, gene X is located in the "leading region" and is the first to enter the recipient cells during conjugation. Recently, gene X was shown to be required for the efficient transfer of R1 (2). The introduction of two stop codons into the coding region of gene X resulted in a 10-fold reduction in R1 transfer frequency. It is interesting that the defective phenotype of the gene X mutation in R1 transfer was complemented in *trans* by the R64 *pilT* gene (2).

PilT-homologs are also present in *Shigella flexneri* (IpgF), *Salmonella typhi* (IagB), and *E. coli* (ORF138 and Slt) (Fig. 4F). In *S. flexneri* and *S. typhi*, the IpgF and IagB proteins are involved in the invasion of the eukaryotic host cells by the bacterial cells (1, 25). The *E. coli slt* gene encodes a 70-kDa soluble lytic transglycosylase (7, 8). X-ray crystallography revealed that the C-terminal domain of Slt protein has a three-dimensional structure similar to those of hen egg and T4 lysozymes (37). It is possible that PilT homologs carry peptidoglycan-lytic activity.

The *pilU* gene product is a prepilin peptidase. The *pilU* product is a 218-amino-acid basic protein. In the presence of the *pilU* gene, the size of the PilS protein was reduced from 22 to 19 kDa, suggesting that the *pilU* gene encodes a prepilin peptidase for the *pilS* product, as described above. In fact, the amino acid sequence of the PilU protein shows similarity to those of type IV prepilin peptidases, including BfpP, TcpJ,

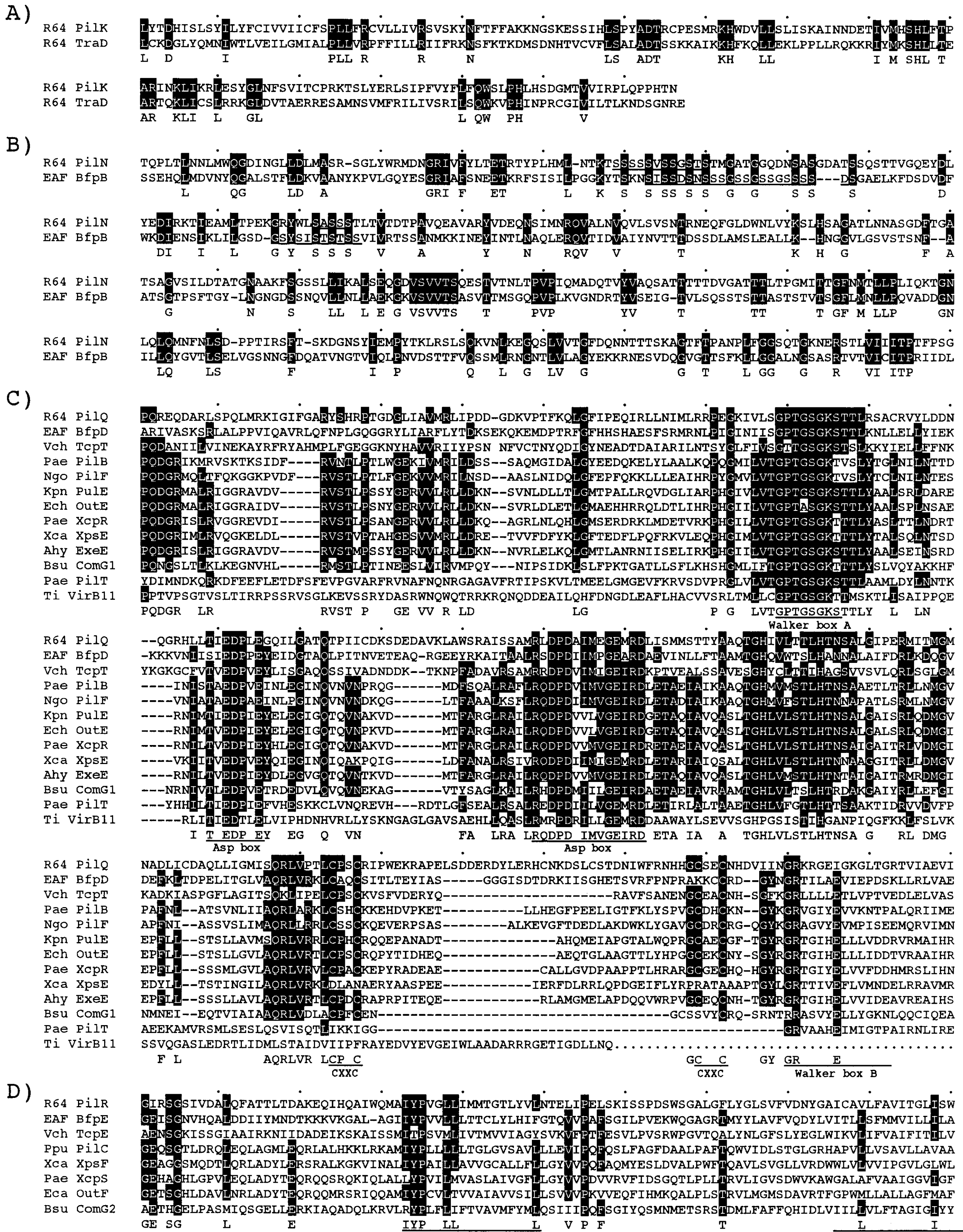
A)
R64 PilK
R64 TraD
B)
R64 PilN
EAF BfpB
C)
R64 PilQ
EAF BfpD
Vch TcpT
Pae PilB
Ngo PilF
Kpn PulE
Ech OutE
Pae XcpR
Xca XpsE
Ahy ExeE
Bsu ComG1
Pae PilT
Ti VirB11
Walker box A
Asp box
Asp box
CXXC
CXXC
Walker box B
D)
R64 PilR
EAF BfpE
Vch TcpE
Ppu PilC
Xca XpsF
Pae XcpS
Eca OutF
Bsu ComG2

FIG. 4.

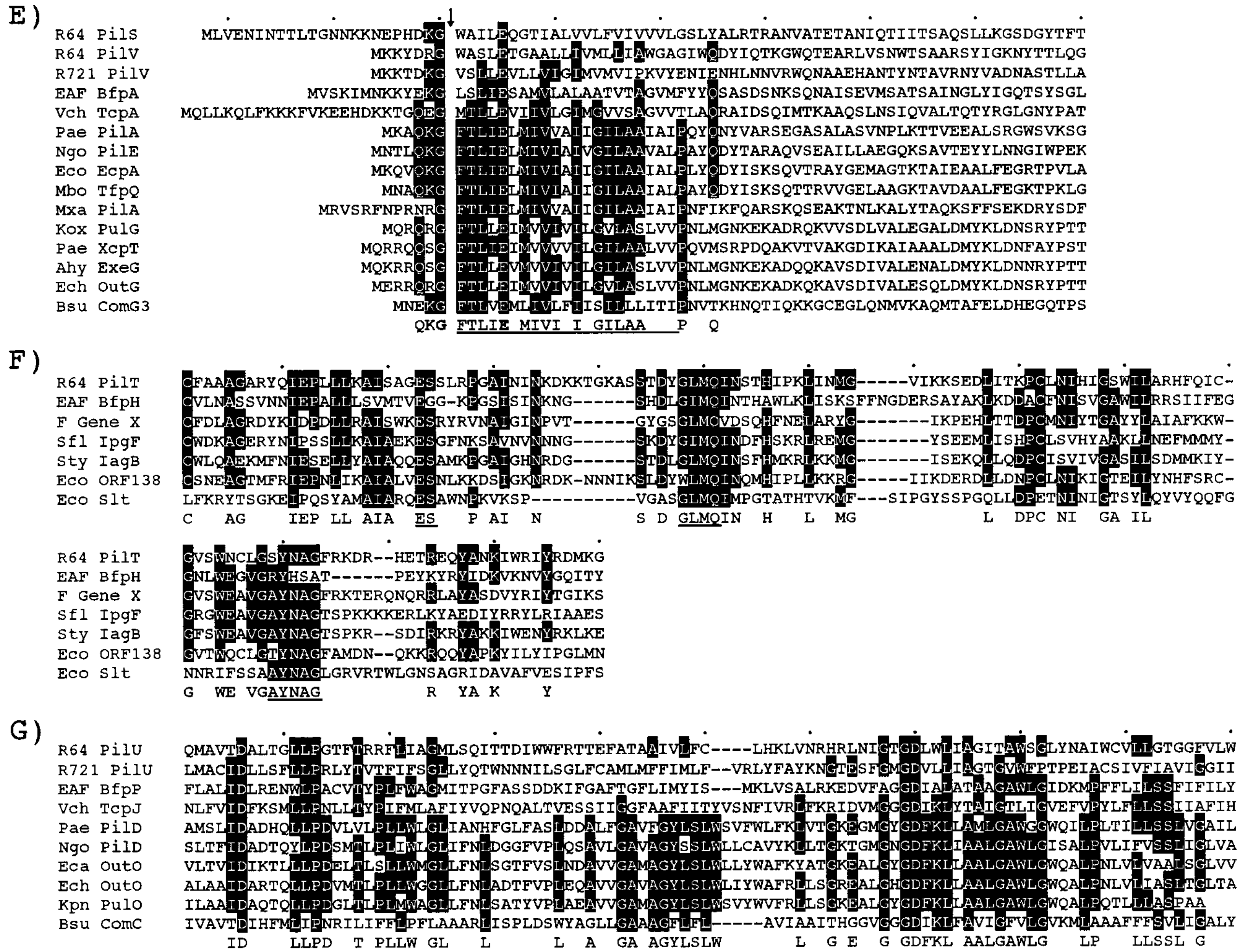

FIG. 4. Alignment of the deduced amino acid sequences of the Pil proteins with those of known proteins. Conserved amino acid residues are printed in white on black. The consensus sequence is given below each alignment. Gaps, marked by dashes, were introduced to reveal the maximal similarity among the sequences. (A) Alignment of amino acid sequences of R64 PilK and R64 TraD (GenBank accession no. D14607). (B) Alignment of sequences of R64 PilN and EAF BfpB (Z68186). Polyserine segments are underlined. (C) Alignment of sequences of R64 PilQ, EAF BfpD (Z68186), *V. cholerae* (Vch) TcpT (X64098), *P. aeruginosa* (Pae) PilB (M32066), *N. gonorrhoeae* (Ngo) PilF (L10291), *Klebsiella pneumoniae* (Kpn) PulE (M32613), *Erwinia chrysanthemi* (Ech) OutE (L02214), *P. aeruginosa* XcpR (X62666), *Xanthomonas campestris* (Xca) XpsE (X59079), *Aeromonas hydrophila* (Ahy) ExeE (X66504), *Bacillus subtilis* (Bsu) ComG1 (M29691), *P. aeruginosa* PilT (M55524), and pTiA6 (Ti) VirB11 (J03216). Walker boxes A and B, Asp boxes, and CXXC sequences are indicated. (D) Alignment of sequences of R64 PilR, EAF BfpE (Z68186), *V. cholerae* TcpE (X64098), *Pseudomonas putida* (Ppu) PilC (X74276), *X. campestris* XpsF (X59079), *P. aeruginosa* XcpS (X62666), *Erwinia carotovora* (Eca) OutF (X70049), and *B. subtilis* ComG2 (M29691). Hydrophobic domains are underlined. (E) Alignment of sequences of R64 PilS, R64 PilV (X05022), R721 PilV (X62169), EAF BfpA (Z68186), *V. cholerae* TcpA (M33514), *P. aeruginosa* PilA (M14849), *N. gonorrhoeae* PilE (L10291), *Eikenella corrodens* (Eco) EcpA (Z12609), *Moraxella bovis* (Mbo) TfpQ (M59712), *M. xanthus* (Mxa) PilA (L39904), *Klebsiella oxytoca* (Kox) PulG (M32613), *P. aeruginosa* XcpT (X62666), *A. hydrophila* ExeG (X66504), *E. chrysanthemi* OutG (L02214), and *B. subtilis* ComG3 (M29691). The cleavage site of type IV pilins is shown by a downward arrow. The conserved glycine and glutamic acid residues are indicated by boldface. (F) Alignment of sequences of R64 PilT, EAF BfpH (Z68186), F gene X (M97768), *S. flexneri* (Sfl) IpgF (L04309), *S. typhi* (Sty) IagB (X80892), *E. coli* (Eco) ORF138 (U28375), and *E. coli* Slt (U14003). Three motifs conserved in putative lytic transglycosylases are underlined. (G) Alignment of sequences of R64 PilU, R721 PilU (X62169), EAF BfpP (Z68186), *V. cholerae* TcpJ (M74708), *P. aeruginosa* PilD (M32066), *N. gonorrhoeae* PilD (L11715), *E. carotovora* OutO (X70049), *E. chrysanthemi* OutO (L02214), *K. pneumoniae* PulO (M32613), and *B. subtilis* ComC (M30805).

PilD, and others (Fig. 4G). It is also similar to the R721 PilU protein (16). Since the R721 *pilU* gene is located just upstream of the R721 *pilV* gene and the deduced amino acid sequence is 23% identical to that of the R64 *pilU* gene, the R721 *pilU* gene may also encode a prepilin peptidase. However, the amino acid sequences of R64 and R721 PilU proteins are greatly divergent from those of the type IV prepilin peptidases of other bacteria.

The *pilV* gene product. The DNA rearrangement of the shufflon converts the C-terminal segments of the *pilV* genes, creating seven different *pilV* genes (18–20). The constant region of the PilV proteins consists of 361 amino acids, while the number of amino acids in the PilV variable region fluctuates between 69 and 113. The N-terminal amino acid sequence of the PilV proteins bears a structure similar to that of type IV prepilin (Fig. 4E). The conserved glycine and glutamic acid appear at the 7th and 12th amino acids of the PilV protein, respectively. It is most likely that the N-terminal 6-amino-acid peptide of the *pilV* product is cut off by the *pilU* function. The prediction that the *pilV* product is a minor pilin is supported by the previous finding that the PilV protein is a minor component of the R64 thin pilus (18, 42).

Effects of *pilR* and *pilU* mutations on liquid mating and sensitivity to phages. The effects of the *pilR1* and *pilU1* mutations on thin-pilus formation were measured by the transfer frequency in liquid medium and sensitivity to IncI1-specific

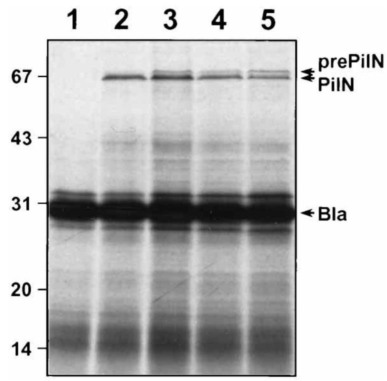

FIG. 5. Effects of globomycin on the production of PilN protein. *E. coli* CSR603 cells harboring pKK687 were labeled with [^{35}S]methionine in the absence or presence of globomycin by the maxicell procedure (31). Labeled proteins were analyzed as described in the legend to Fig. 3. Lanes: 1, pUC9; 2, pKK687; 3 to 5, pKK687 with 50, 100, and 200 μg of globomycin/ml, respectively. PilN and its precursor protein bands are indicated on the right side. Numbers on the left show the sizes (in kilodaltons) of marker proteins.

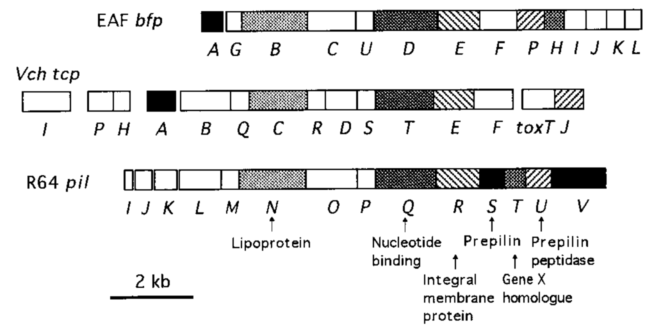

FIG. 6. Comparison of gene organization among various type IVB pilus-forming systems. The gene organizations of EAF *bfp*, *V. cholerae* (Vch) *tcp*, and R64 *pil* systems are compared. All genes except *tcpI* are transcribed rightward. Genes with similar functions are indicated.

phages. For liquid mating, the pKK641-pKK661 system, in which DNA rearrangement of the shufflon does not occur because of lack of the *rci* activity, was used (17). *E. coli* donor cells harboring both pKK641 and pKK661 transmitted pKK661 carrying *oriT* to recipient cells upon conjugation (Table 2). From donor cells harboring pKK641 *pilR1* or pKK641 *pilU1* together with pKK661, however, transfer frequency was negligible. In contrast, the *pilR1* and *pilU1* mutations did not affect the transfer frequency on solid surfaces (data not shown). *E. coli* cells harboring pKK661 and either pKK641 *pilR1* or pKK641 *pilU1* were resistant to phages Iα and PR64FS, while cells harboring wild-type pKK641 and pKK661 were sensitive to them (Table 2). These results indicate that *pilR* and *pilU* genes are involved in thin-pilus formation and, consequently, in liquid mating.

TABLE 2. Effects of *pilR* and *pilU* mutations on liquid mating and sensitivity to phages

Plasmids	Frequency of transfer (%)[a]	Sensitivity to phage[b]	
		Iα	PR64FS
pKK661, pKK641	0.51	S	S
pKK661, pKK641 *pilR1*	0.001	R	R
pKK661, pKK641 *pilU1*	0.002	R	R

[a] Frequency of transfer of pKK661 from donor cells harboring the indicated plasmids in liquid mating is indicated as percent of transconjugants to donor cells.

[b] Sensitivity to phages Iα and PR64FS of *E. coli* cells harboring the indicated plasmids. S, sensitive; R, resistant.

DISCUSSION

The R64 thin pilus is a type IV pilus. The present results strongly suggest that the R64 thin pilus belongs to the type IV family of pili, since (i) the PilS and PilV proteins carry structures homologous to those of type IV prepilins, (ii) the PilU protein carries a structure homologous to that of type IV prepilin peptidase, and (iii) the PilN, PilQ, PilR, and PilT proteins carry structures homologous to those of proteins related to type IV pilus biogenesis. The purified thin pili of R64 or ColIb-P9 were shown to consist of 19- and 48-kDa proteins, which might be the products of *pilS* and *pilVA′* genes, respectively (18, 19, 42). The reduction of the size of the PilS protein in the presence of the *pilU* gene suggests that the *pilU* product functions as a prepilin peptidase. The N-terminal amino acid sequences of the R64 PilS and PilV proteins have a slight similarity to each other, but they are not so similar to those of type IVA pilins (Fig. 4E). In addition, the N-terminal amino acids of putative mature pilins of R64 *pilS* and *pilV* products are tryptophan. These results indicate that the R64 thin pilus belongs to the type IVB group.

Gene organization of the R64 *pil* region. The gene organization of the R64 *pil* region was compared to those of the *bfp* system of enteropathogenic *E. coli* and the *tcp* system of *V. cholerae* (26, 33, 34) (Fig. 6). The organizations of these type IVB pilus biogenesis systems reveal some similarity to each other. The relative positions of the PilN, PilQ, and PilR homolog genes are conserved among the three systems. Prepilin peptidase genes are located at the downstream positions of these systems. PilT homolog genes are also located at downstream positions, while the relative positions of prepilin peptidase genes and PilT homolog genes are different between the R64 *pil* and EAF *bfp* systems. The locations of prepilin genes are different among these systems. The prepilin genes are located at upstream positions in the *bfp* and *tcp* systems, while prepilin genes are located in the downstream region of the R64 *pil* system. In the R64 *pil* region, it is reasonable that the *pilV* gene is located at the 3′-most position of the *pil* operon, since its C-terminal segments are under the control of the shufflon. Expression of the *bfp*, *tcp*, and *pil* systems is positively controlled by the *bfpT*, *toxRT*, and *traBC* genes, respectively (9, 10, 15). Most of these regulatory genes are located outside the type IV pilus biogenesis gene cluster.

The similarity of amino acid sequences among various genes, as well as the conservation of gene organization, indicates that these type IV pilin biogenesis systems have evolved from a common ancestral gene system.

Possible function of the R64 thin pilus in conjugation. To our knowledge, the R64 thin pilus is the first example of a type IV pilus that is involved in bacterial conjugation. The involvement of at least *pilR* and *pilU* genes in liquid mating was demonstrated by the present study. Many pathogenic bacteria use type IV pili to attach themselves to eukaryotic host cells. R64 thin pili may be used to attach donor cells to recipient cells at the first step of liquid mating. R64 thin pili contain PilV proteins, the C-terminal segments of which are under the control of the shufflon DNA rearrangement. Thus, seven different kinds of R64 thin pili with different PilV proteins may be produced and used to attach donor cells to different kinds of recipient cells to determine recipient specificity in liquid mating of R64 (18, 19). For example, thin pili with PilVA′, PilVC, or PilVC′ protein are used to recognize *E. coli* K-12 recipient cells, and those with PilVA′ or PilVB′ protein are used for *Salmonella typhimurium* LT2. In R64 surface mating, however, this step can be skipped, since thin pili are not required for solid-surface mating (17, 18). R64 thick pili are required for both solid-surface and liquid mating.

In many pili, an adhesin was shown to be located at the tip (13, 22, 23, 29). R64 PilV proteins may also function as an adhesin. The speculation that the PilV protein is localized to the tip of the R64 thin pilus awaits further investigation.

ACKNOWLEDGMENTS

We are grateful to M. Inukai for globomycin. We thank Kathy Takayama for critical reading of the manuscript.

This work was supported in part by a grant from the Ministry of Education, Science, and Culture of Japan.

REFERENCES

1. **Allaoui, A., R. Menard, P. J. Sansonetti, and C. Parsot.** 1993. Characterization of the *Shigella flexneri ipgD* and *ipgF* genes, which are located in the proximal part of the *mxi* locus. Infect. Immun. **61:**1707–1714.
2. **Bayer, M., R. Eferl, G. Zellnig, K. Teferle, A. Dijkstra, G. Koraimann, and G. Högenauer.** 1995. Gene *19* of plasmid R1 is required for both efficient conjugative DNA transfer and bacteriophage R17 infection. J. Bacteriol. **177:**4279–4288.
3. **Bladley, D. E.** 1983. Derepressed plasmids of incompatibility group I determine two different morphological forms of pilus. Plasmid **9:**331–334.
4. **Bladley, D. E.** 1984. Characterization and function of thick and thin conjugative pili determined by transfer-derepressed plasmids of incompatibility group I_1, I_2, I_5, B, K and Z. J. Gen. Microbiol. **130:**1489–1502.
5. **Coetzee, J. N., D. E. Bladley, and R. W. Hedges.** 1982. Phages Iα and I2-2: IncI plasmid-dependent bacteriophages. J. Gen. Microbiol. **128:**2797–2804.
6. **Coetzee, J. N., F. A. Sirgel, and G. Lecatsas.** 1980. Properties of a filamentous phage which absorbs to pili coded by plasmids of IncI complex. J. Gen. Microbiol. **117:**547–551.
7. **Dijkstra, B. W., and W. Keck.** 1996. Peptidoglycan as a barrier to transenvelope transport. J. Bacteriol. **178:**5555–5562.
8. **Dijkstra, B. W., and A. M. Thunnissen.** 1994. 'Holy' proteins. II. The soluble lytic transglycosylase. Curr. Opin. Struct. Biol. **4:**810–813.
9. **DiRita, V. J., C. Parsot, G. Jander, and J. J. Mekalanos.** 1991. Regulatory cascade controls virulence in *Vibrio cholerae*. Proc. Natl. Acad. Sci. USA **88:**5403–5407.
10. **Giron, J. A., A. S. Y. Ho, and G. K. Schoolnik.** 1991. An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. Science **254:**710–713.
11. **Giron, J. A., M. M. Lavine, and J. B. Kaper.** 1994. Longus, a long pilus ultrastructure produced by human enterotoxigenic *Escherichia coli*. Mol. Microbiol. **12:**71–82.
12. **Hobbs, M., and J. S. Mattick.** 1993. Common components in the assembly of type 4 fimbriae, DNA transfer systems, filamentous phage and protein secretion apparatus: a general system for the formation of surface-associated protein complexes. Mol. Microbiol. **10:**233–243.
13. **Hultgren, S. J., S. Abraham, M. Caparon, P. Falk, J. W. St. Gemme III, and S. Normark.** 1993. Pilus and nonpilus bacterial adhesins: assembly and function in cell recognition. Cell **73:**887–901.
14. **Inukai, M., M. Takeuchi, K. Shimizu, and M. Arai.** 1978. Mechanism of action of globomycin. J. Antibiot. **31:**1203–1205.
15. **Kim, S.-R., N. Funayama, and T. Komano.** 1993. Nucleotide sequence and characterization of the *traABCD* region of IncI1 plasmid R64. J. Bacteriol. **175:**5035–5042.
16. **Kim, S.-R., and T. Komano.** 1992. Nucleotide sequence of the R721 shufflon. J. Bacteriol. **174:**7053–7058.
17. **Komano, T., N. Funayama, S.-R. Kim, and T. Nisioka.** 1990. Transfer region of IncI1 plasmid R64 and role of shufflon in R64 transfer. J. Bacteriol. **172:**2230–2235.
18. **Komano, T., S.-R. Kim, and T. Yoshida.** 1995. Mating variation by DNA inversions of shufflon in plasmid R64. Adv. Biophys. **31:**181–193.
19. **Komano, T., S.-R. Kim, T. Yoshida, and T. Nisioka.** 1994. DNA rearrangement of the shufflon determines recipient specificity in liquid mating of IncI1 plasmid R64. J. Mol. Biol. **243:**6–9.
20. **Komano, T., A. Kubo, and T. Nisioka.** 1987. Shufflon: multi-inversion of four contiguous DNA segments of plasmid R64 creates seven different open reading frames. Nucleic Acids Res. **15:**1165–1172.
21. **Kubo, A., A. Kusukawa, and T. Komano.** 1988. Nucleotide sequence of the *rci* gene encoding shufflon-specific DNA recombinase in the IncI1 plasmid R64: homology to the site-specific recombinases of integrase family. Mol. Gen. Genet. **213:**31–35.
22. **Kuehn, M. J., J. Heuser, S. Normark, and S. J. Hultgren.** 1992. P pili in uropathogenic *E. coli* are composite fibers with distinct fimbrillar adhesive tips. Nature (London) **356:**252–255.
23. **Lindberg, F., B. Lund, L. Johansson, and S. Normark.** 1987. Localization of the receptor-binding protein adhesin at the tip of the bacterial pilus. Nature (London) **328:**84–87.
24. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Miras, I., D. Hermant, N. Arricau, and M. Y. Popoff.** 1995. Nucleotide sequence of *iagA* and *iagB* genes involved in invasion of HeLa cells by *Salmonella enterica* subsp. *enterica* ser. Typhi. Res. Microbiol. **146:**17–20.
26. **Ogierman, M. A., S. Zabihi, L. Mourtzios, and P. A. Manning.** 1993. Genetic organization and sequence of the promoter-distal region of the *tcp* gene cluster of *Vibrio cholerae*. Gene **126:**51–60.
27. **Ottow, J. C.** 1975. Ecology, physiology, and genetics of fimbriae and pili. Annu. Rev. Microbiol. **29:**79–108.
28. **Pugsley, A. P.** 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol. Rev. **57:**50–108.
29. **Rudel, T., I. Scheuerpflug, and T. F. Meyer.** 1995. *Neisseria* PilC protein identified as type-4 pilus tip-located adhesin. Nature (London) **373:**357–359.
30. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Sancar, A., A. M. Hack, and W. D. Rupp.** 1979. Simple method for identification of plasmid-coded proteins. J. Bacteriol. **137:**692–693.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
33. **Sohel, I., J. L. Puente, S. W. Ramer, D. Bieber, C.-Y. Wu, and G. K. Schoolnik.** 1996. Enteropathogenic *Escherichia coli*: identification of a gene cluster coding for bundle-forming pilus morphogenesis. J. Bacteriol. **178:**2613–2628.
34. **Stone, K. D., H.-Z. Zhang, L. K. Carlson, and M. S. Donnenberg.** 1996. A cluster of fourteen genes from enteropathogenic *Escherichia coli* is sufficient for the biogenesis of a type IV pilus. Mol. Microbiol. **20:**325–337.
35. **Strom, M. S., and S. Lory.** 1993. Structure-function and biogenesis of the type IV pili. Annu. Rev. Microbiol. **47:**565–596.
36. **Strom, M. S., D. N. Nunn, and S. Lory.** 1993. A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family. Proc. Natl. Acad. Sci. USA **90:**2404–2408.
37. **Thunnissen, A. M. W. H., A. J. Dijkstra, K. H. Kalk, H. J. Rozeboom, H. Engel, W. Keck, and B. W. Dijkstra.** 1994. Doughnut-shaped structure of a bacterial muramidase revealed by X-ray crystallography. Nature (London) **367:**750–764.
38. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153:**3–11.
39. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945–951.
40. **Wu, S. S., and D. Kaiser.** 1995. Genetic and functional evidence that type IV pili are required for social gliding motility in *Myxococcus xanthus*. Mol. Microbiol. **18:**547–558.
41. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33:**103–109.
42. **Yoshida, T., and T. Komano.** Unpublished data.