JOURNAL OF BACTERIOLOGY, Apr. 1995, p. 1892–1895 Vol. 177, No. 7
0021-9193/95/$04.00+0
Copyright © 1995, American Society for Microbiology

# Identification of the *hopG* Gene, a Component of *Escherichia coli* K-12 Type II Export System, and Its Conservation among Different Pathogenic *Escherichia coli* and *Shigella* Isolates

IGOR STOJILJKOVIC,[1]* ROLAND SCHÖNHERR,[1]† AND JOHANNES G. KUSTERS[2]

*Institut für Membranphysiologie/Mikrobiologie II, Universität Tübingen, Tübingen, Germany,[1] and Department of Bacteriology and Immunology, Faculty of Veterinary Medicine, University of Utrecht, Utrecht, The Netherlands[2]*

Received 14 November 1994/Accepted 27 January 1995

**The *Escherichia coli* K-12 gene coding for a component of a type II export system was identified and characterized. The HopG protein contains a typical prepilin peptidase cleavage site and has a high degree of homology with proteins PulG, OutG, and ExeG, which are components of type II secretion systems from *Klebsiella pneumoniae*, *Erwinia carotovora*, and *Aeromonas hydrophila*.**

The discovery of the pullulanase secretion system of *Klebsiella oxytoca* defined a new family of secretory systems that has a homolog in many gram-negative bacteria (7, 12, 20–22, 24, 25). These pullulanase-like, so-called type II export systems were shown to be responsible for the secretion of exotoxin A, elastase, alkaline phosphatase, and phospholipase C in *Pseudomonas aeruginosa* (3, 4, 30); the extracellular secretion of degradative enzymes in the plant pathogens *Erwinia chrysanthemi*, *Erwinia carotovora*, and *Xanthomonas campestris* (6, 11, 23); the uptake of DNA in *Bacillus subtilis* (16); and the formation of type IV pili in different gram-negative bacteria (9, 10, 15, 17, 18, 29). The analysis of nucleotide sequences in databases identified at least six *Escherichia coli* open reading frames (ORFs) of unknown function which possess some degree of homology with different components involved in production of type IV pili (10). Three ORFs of unknown function are found in the *thyA-recC* intergenic region (8), and the remaining three are scattered around the *E. coli* chromosome. A *pilQ*-like gene (*hopQ*, for homologous to *pil*) (13, 14) is located upstream of *aroK*, and a *pilC*-like gene (*hopC*) is located downstream of *guaC* (1). Finally, the so-called *hopD* gene, showing significant degrees of homology with different prepilin peptidases, is located close to the *bfr* gene at 72 min on the *E. coli* chromosome (2, 10, 32). In this communication, we describe a cluster of *E. coli hop* genes and study the distribution of *hopFGH*-like genes in different pathogenic strains of *E. coli.*

During a screen for genes involved in iron metabolism in the proximity of the *E. coli bfr* locus (28), an unknown ORF whose predicted product shows a similarity in amino acid sequence with proteins from different type II export systems was identified. The nucleotide sequence of the consecutive 0.5- and 0.7-kb *Nde*I fragments (Fig. 1) was determined (27), and analysis of the nucleotide sequence revealed the presence of three ORFs (Fig. 2). The second ORF (ORF2), encoding a 145-amino-acid protein, was located completely within the 0.5-kb *Nde*I fragment, while the first and the third ORFs were not completely sequenced. An amino acid comparison with the proteins in the SWISS-PROT database revealed that ORF2 has a very strong degree of homology with the proteins PulG, OutG and ExeG, which are part of type II secretion systems from *Klebsiella pneumoniae* (12), *E. carotovora* (23), and *Aeromonas hydrophila* (11), respectively. The overall homology between these proteins was around 76% (Fig. 3). In separate comparisons, the homology was highest with the *E. carotovora* OutG protein (73% identical amino acid residues), which is in accordance with the fact that *E. carotovora* is the closest relative of *E. coli* (19). The ORF2 gene was therefore named *hopG.*

The product of the first ORF (*hopF*) showed very high degrees of similarity to PulF from *K. pneumoniae* (12, 20, 21), ExeF from *A. hydrophila* (11), XcpS from *P. aeruginosa* (3, 4), and OutF from *E. carotovora* (23). Again, these proteins are parts of different secretory systems in gram-negative bacteria. The amino acid sequence LPLSTR...GPTL (Fig. 2) of the first ORF was identified as a leucine zipper motif, a protein motif found to be involved in protein-protein interactions (5). These protein domains of HopF and XcpS might be involved in a protein oligomerization process leading to the formation of a structure thought to be involved in protein transport across the outer membrane (25). The product of the third ORF (*hopH*)

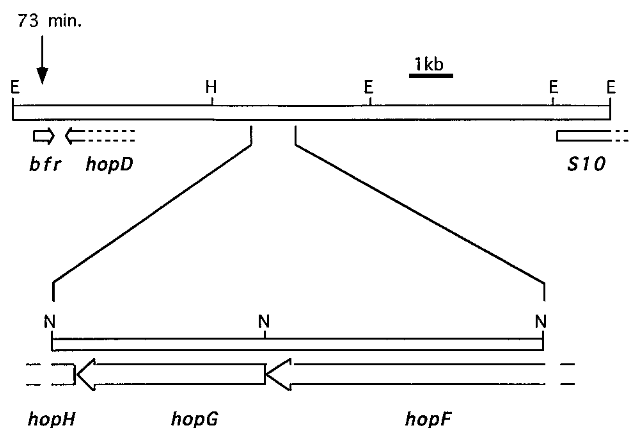


FIG. 1. Restriction map of the *Hin*dIII-*Eco*RI 4-kb DNA insert carrying the *hopF*, *hopG*, and *hopH* genes. The *bfr* and *hopD* genes are located approximately 4 kb proximal to the *Hin*dIII restriction site (2, 32). The *rhe* ribosomal operon is located distal to the right *Eco*RI restriction site. The *hop* genes, located at 73 min on the *E. coli* chromosomal map, are transcribed in the counterclockwise direction. E, *Eco*RI; H, *Hin*dIII; N, *Nde*I.

* Corresponding author. Present address: Department of Microbiology and Immunology, Oregon Health Sciences University, Portland, OR 97201. Phone: (503) 494-6841. Fax: (503) 494-6862.

† Present address: Max-Planck-Gesellschaft z.F.d.W, AG Moleculare und zelluläre Biophysik, 07747 Jena, Germany.

```
NdeI
CATATGAAGCAGCAACTGCCGCTGAGTACACGCATTCTTTTAGGTCTGAGCGACACGTTGCA
  HisMetLysGlnGlnLeuProLeuSerThrArgIleLeuLeuGlyLeuSerAspThrLeuGl
    ...hopF'

    ACGTACCGGCCCGACATTATTAGCGACAGTGTTTATTGTCGCTGTAGGTTTCTGGCTCTG
    nArgThrGlyProThrLeuLeuAlaThrValPheIleValAlaValGlyPheTrpLeuTr

    GTTAAAACGCGGCAATAACCGCCACCGTTTTCATGCCATGTTGCTGCGCGTTGCGCTCAT
    pLeuLysArgGlyAsnAsnArgHisArgPheHisAlaMetLeuLeuArgValAlaLeuIl

    CGGCCCGCTGATTTGCGCCATTAACAGCGCACGCTATCTCCGACCTTTAAGTATTTTGCA
    eGlyProLeuIleCysAlaIleAsnSerAlaArgTyrLeuArgProLeuSerIleLeuGl

    ATCCAGCGGCGTCCCTCTGCTGGATGGGATGAATTTGTCCACCGAAAGCCTCAACAACCT
    nSerSerGlyValProLeuLeuAspGlyMetAsnLeuSerThrGluSerLeuAsnAsnLe

    CGAAATTCGCCAGCGTCTGGCAAATGCGGCAGAGAACGTTCGCCAGGGTAACAGCATTCA
    uGluIleArgGlnArgLeuAlaAsnAlaAlaGluAsnValArgGlnGlyAsnSerIleHi

    TCTTTCGCTGGAACAAACCGCAATTTTCCCGCCGATGATGCTCTACATGGTGGCCTCTGG
    sLeuSerLeuGluGlnThrAlaIlePheProProMetMetLeuTyrMetValAlaSerGl

    CGAAAAAAGCGGGCAGCTCGGCACATTAATGGTCAGAGCCGCAGATAACCAGGAGACACT
    yGluLysSerGlyGlnLeuGlyThrLeuMetValArgAlaAlaAspAsnGlnGluThrLe

    CCAACAAAATCGGATCGCCTTAACGCTCTCCATCTTCGAGCCAGCACTCATTATTACGAT
    uGlnGlnAsnArgIleAlaLeuThrLeuSerIlePheGluProAlaLeuIleIleThrMe

    GGCACTGATCGTCCTGTTTATTGTCGTGTCGGTACTCCAACCTCTTCTTCAACTTAACTC
    tAlaLeuIleValLeuPheIleValValSerValLeuGlnProLeuLeuGlnLeuAsnSe
                     NdeI
  AATGATTAATTAAGGAAAGCATATGCGCGCAACGGATAAGCAACGCGGTTTTACATTACT
  rGlyLysHisSTOP            METArgAlaThrAspLysGlnArgGlyPheThrLeuLe
                           hopG

  GGAAATTATGGTGGTGATCGTCATTATCGGCGTACTTGCCAGCCTGGTGGTTCCTAACCT
    uGluIleMetValValIleValIleIleGlyValLeuAlaSerLeuValValProAsnLe

    AATGGGCAATAAAGAAAAGGCGGATAAGCAAAAAGCCGTCAGCGATATTGTGGCGCTGGA
    uMetGlyAsnLysGluLysAlaAspLysGlnLysAlaValSerAspIleValAlaLeuGl

    AAACGCCCTTGATATGTACAAACTCGACAATCATCACTATCCAACCACAAATCAGGGGCT
    uAsnAlaLeuAspMetTyrLysLeuAspAsnHisHisTyrProThrThrAsnGlnGlyLe

    TGAATCCTTAGTAGAAGCGCCGACACTGCCACCGCTGGCCGCAAACTATAACAAGGAAGG
    uGluSerLeuValGluAlaProThrLeuProProLeuAlaAlaAsnTyrAsnLysGluGl

    TTATATCAAGCGTCTGCCTGCCGATCCCTGGGGCAATGATTATGTCCTCGTTAATCCTGG
    yTyrIleLysArgLeuProAlaAspProTrpGlyAsnAspTyrValLeuValAsnProGl

    TGAACATGGTGCATACGATCTGCTTTCAGCAGGGCCCGATGGTGAAATGGGAACCGAGGA
    yGluHisGlyAlaTyrAspLeuLeuSerAlaGlyProAspGlyGluMetGlyThrGluAs

    CGACATCACCAACTGGGGTTTGAGCAAGAAGAAAAAGTAAGTCAGCGATGAATCAGCAAC
  pAspIleThrAsnTrpGlyLeuSerLysLysLysLysSTOP           METAsnGlnGlnA
                                                     hopH'

    GCGGGTTTACCCTGCTGGAGATGATGCTGGTCGTGGCGCTGGTCGCTATCACGGCAAGCG
    rgGlyPheThrLeuLeuGluMetMetLeuValValAlaLeuValAlaIleThrAlaSerV

             NdeI
  TGGTGCTCTTCACATATG
   alValLeuPheThrTyr
```

FIG. 2. Nucleotide sequence of the 1.14-kb DNA fragment carrying the *hopF′GH′* genes. The putative leucine zipper motif in the *hopF* product starts at the first Leu residue shown in the first line of the HopF amino acid sequence and finishes at the first Leu shown in the second line of the HopF amino acid sequence. The conserved prepilin peptidase recognition sequence (GFTLLE) begins at Gly 9 in HopG and at Gly 6 in HopH′.

was homologous with the *pulH* gene product, a component of the pullulanase export apparatus.

Both HopG and HopH possessed potential cleavage sites (GFTLLE) for a type IV prepilin peptidase in their N termini (10, 20, 29). Interestingly, this sequence is not only conserved in pullulanase PulGHIJ and type IV pilus GHIJ homologs, but also in the secretory proteins of many other gram-negative bacteria and in some gram-positive bacteria (7, 10, 20, 29).

We used the T7 gene *10* expression system in order to investigate whether the ORFs identified by nucleotide sequence determination code for proteins (31). As can be seen from Fig. 4, a strong protein band of approximately 16 to 17 kDa was expressed from the plasmid pSUKSEH4. The apparent molecular mass of the expressed protein is in accordance with the deduced molecular mass of the protein encoded by the *hopG* gene (16.09 kDa). In addition, two protein bands (one of 21 kDa and the other of 42 kDa [Fig. 4, lane 1]) of lower intensities were also detected in the T7 expression experiment. The expression of the plasmid pSUSKEH4 carrying the same insert in the inverted orientation to the T7 promoter did not reveal any protein bands (Fig. 4, lane 3).

```
HOPG        MRATDK-------------QRGFTLLEIMVVIVIIGVLASLVVPNLMGNK     37
PULG        MQ-----------------RQRGFTLLEIMVVIVILGVLASLVVPNLMGNK    34
OUTG        MQQSQRGCGQNSYGQSGYRQRGFTLLEIMVVIVILGVLASLVVPNLMGNK     50
EXEG        MQKR--------------RQSGFTLLEVMVVIVILGILASLVVPNLMGNK     36
            *.                 *.******.******.*.*************
HOPG        EKADKQKAVSDIVALENALDMYKLDNHHYPTTNQGLESLVEAPTLPPLAA     87
PULG        EKADRQKVVSDLVALEGALDMYKLDNSRYPTTEQGLQALVSAPSAEPHAR     84
OUTG        EKADRQKAVSDIVSLESALDMYKLDNNRYPSTEQGLKALVTKPTVQPEPR    100
EXEG        EKADQQKAVSDIVALENALDMYKLDNNRYPTTEQGLDALVNKPTAAPEPR     86
            ****.**.***.*.**.*********  .**.*.***..**. *.  * .

HOPG        NYNKEGYIKRLPADPWGNDYVLVNPGEHGAYDLLSAGPDGEMGTEDDITN    137
PULG        NYPEGGYIRRLPQDPWGSDYQLLSPGQHGQVDIFSLGPDGVPESNDDIGN    134
OUTG        NYPADGYIRRLPQDPWGTDYQLLNPGQHGKLDIFSLGPDGMPGTEDDIGN    150
EXEG        SYRDGGYIKRLPQDPWGNPYQMLSPGQFGKIDIFSMGLDGEAGTDDDIGN    136
            .*   .***.***.****. *  ...**. *   *..* * **   ...***.*

HOPG        WGLSKKKK   145
PULG        WTIGK-K-   140
OUTG        WNLDK-K-   156
EXEG        WNLKDFQ-   143
            *.. . .

Identity  : 84 ( 53.2%)
Similarity: 37 ( 23.4%)
```

FIG. 3. Amino acid comparison between the *E. coli* HopG, *K. pneumoniae* PulG (12, 20), *E. carotovora* OutG (23), and *A. hydrophila* ExeG (11) proteins. Asterisks indicate identical amino acids, and dots indicate similar amino acids. The amino acid sequence recognized by prepilin peptidase is in boldface type. The comparison was performed with the program CLUSTAL from the PC/GENE software package (Intelligenetics Inc.).

In order to investigate the distribution of *hopFGH*-like sequences in different members of the family *Enterobacteriaceae*, Southern hybridization (26) was carried out with the *hopF* and *hopGH* genes as probes (Genius digoxigenin labelling and detection kit; Boehringer Mannheim). As can be seen in Fig. 5, several strains of *E. coli* gave strong hybridization signals (high stringency conditions [26]) with both probes. The most prominent signals were found in DNA from enteropathogenic *E. coli* (EPEC) and enteroinvasive and enterotoxigenic *E. coli*, while the DNA from the enterohemorrhagic *E. coli* strain did not contain sequences which would hybridize with the *hop* probes. Interestingly, the *hopGH* gene probe did not detect any

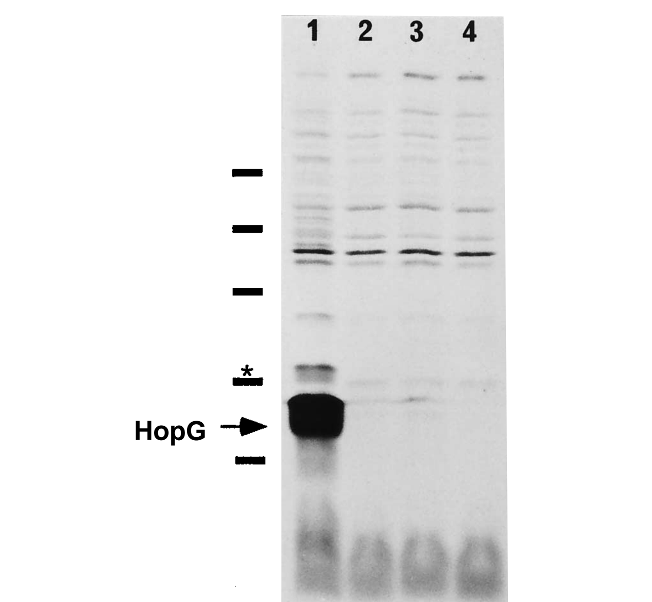


FIG. 4. Polyacrylamide gel electrophoresis of proteins expressed in vivo from the T7 promoter on plasmids pSUKSEH4 (lane 1), pSUKS1 (lane 2), pSUSKEH4 (lane 3), and pSUSK1 (lane 4) in *E. coli* BL21. A sodium dodecyl sulfate–15% polyacrylamide gel was used for the separation of plasmid-encoded products. The positions of protein size markers (14K, 20K, 26.6K, 39K, and 55.5K) are shown to the left of the gel.

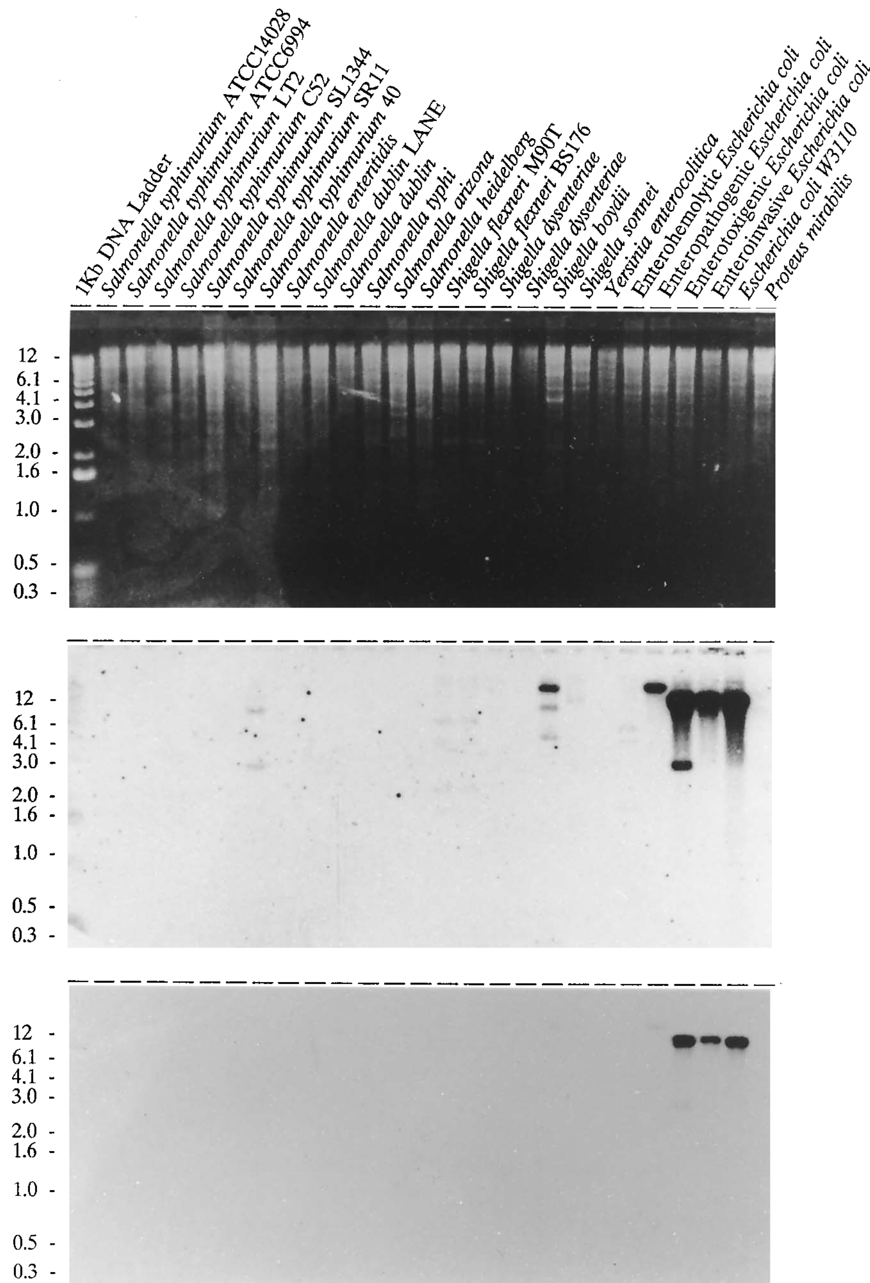


FIG. 5. The distribution of *hopF*- and *hopGH*-like sequences in different members of the *Enterobacteriaceae*. The top picture is an ethidium bromide-stained agarose gel of *Eco*RI-digested chromosomal DNA. The middle picture is the blot hybridized with the *hopF* probe. The bottom picture is the blot hybridized with the *hopGH* probe. Sizes of DNA fragments (in kilobases) are indicated in the ladders at the left (Gibco BRL).

signal in EPEC DNA, while the *hopF* gene probe hybridized strongly to the EPEC DNA. The *hopF* probe gave also a strong signal with *Shigella boydii* DNA but not with that of other *Shigella* species. The hybridization of the *hopGH* probe was confined to *E. coli* strains (Fig. 5).

These data indicate that the *hopFGH* gene cluster, together with the *hopD* gene, might be a part of an as-yet undescribed large *E. coli* export or type IV pilus-encoding operon. This newly discovered secretion system, found in different pathogenic *E. coli* isolates and some *Shigella* isolates, might be a new virulence factor of these bacteria.

We are indebted to V. Braun and K. Hantke for their hospitality, support, and suggestions. We thank Reneé Tsolis and Andreas F. Baümler for carefully reading the manuscript.

J.G.K. was supported by a fellowship of the Royal Netherlands Academy of Arts and Sciences. We acknowledge the support of the Deutsche Forschungsgemeinschaft (SFB 323, project B1).

## REFERENCES

1. **Andrews, S. C., and J. R. Guest.** 1988. Nucleotide sequence of the gene encoding the GMP reductase of *Escherichia coli* K12. Biochem. J. **255:**34–43.
2. **Andrews, S. C., P. M. Harrison, and J. R. Guest.** 1989. Cloning, sequencing, and mapping of the bacterioferritin gene (*bfr*) of *Escherichia coli* K-12. J. Bacteriol. **171:**3940–3947.

3. **Bally, M., G. Ball, A. Badere, and A. Lazdunski.** 1991. Protein secretion in *Pseudomonas aeruginosa*: the *xcpA* gene encodes an integral inner membrane protein homologous to *Klebsiella pneumoniae* secretion function protein PulO. J. Bacteriol. **173:**479–486.
4. **Bally, M., A. Filloux, M. Akrim, G. Ball, A. Lazdunski, and J. Tommassen.** 1992. Protein secretion in *Pseudomonas aeruginosa*: characterization of seven xcp genes and processing of secretory apparatus components by prepilin peptidase. Mol. Microbiol. **6:**1121–1132.
5. **Branden, C., and J. Tooze.** 1991. Introduction to protein structure, p. 144. Garland Publishing, Inc., New York.
6. **Dums, F., J. M. Dow, and M. J. Daniels.** 1991. Structural characterization of protein secretion genes of the bacterial phytopathogen *Xanthomonas campestris* pathovar *campestris*: relatedness to secretion systems of other Gram-negative bacteria. Mol. Gen. Genet. **229:**357–364.
7. **Dupuy, B., M.-K. Taha, O. Possot, C. Marchal, and A. P. Pugsley.** 1992. PulO, a component of the pullulanase secretion pathway of *Klebsiella oxytoca*, correctly and efficiently process gonococcal type IV prepilin in *Escherichia coli*. Mol. Microbiol. **6:**1887–1894.
8. **Finch, P. W., R. E. Wilson, K. Brown, I. D. Hickson, A. E. Tomkinson, and P. T. Emmerson.** 1986. Complete nucleotide sequence of the *Escherichia coli recC* gene and of the *thyA-recC* intergenic region. Nucleic Acids Res. **14:**4437–4451.
9. **Hobbs, M., B. P. Dalrymple, P. T. Cox, S. P. Livingstone, S. F. Delaney, and J. Mattick.** 1991. Organization of the fimbrial gene region of *Bacterioides nodosus*: class I and class II strains. Mol. Microbiol. **5:**543–560.
10. **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 structure-associated protein complexes. Mol. Microbiol. **10:**233–243.
11. **Jiang, B., and S. P. Howard.** 1992. The *Aeromonas hydrophila exeE* gene, required for protein secretion and normal outer membrane biogenesis, is a member of the general secretion pathway. Mol. Microbiol. **6:**1351–1361.
12. **Katsugari, N., N. Takizawa, and Y. Murooka.** 1987. Entire nucleotide sequence of the pullulanase gene of *Klebsiella aerogenes* W70. J. Bacteriol. **169:**2301–2306.
13. **Løbner-Olesen, A., and M. G. Marinus.** 1992. Identification of the gene (*aroK*) encoding shikimic acid kinase I of *Escherichia coli*. J. Bacteriol. **174:**525–529.
14. **Martin, A., M. Hobbs, P. D. Free, Y. Jeske, and J. S. Mattick.** 1993. Characterization of pilQ, a new gene required for the biogenesis of type 4 fimbriae in *Pseudomonas aeruginosa*. Mol. Microbiol. **9:**857–868.
15. **Meyer, T. F., E. Billyard, R. Haas, S. Storzbach, and M. So.** 1984. Pilus genes of *Neisseria gonorrhoeae*: chromosomal organization and DNA sequence. Proc. Natl. Acad. Sci. USA **81:**6110–6114.
16. **Mohan, S., J. Aghion, N. Guillen, and D. Dubnau.** 1989. Molecular cloning and characterization of *comC*, a late competence gene of *Bacillus subtilis*. J. Bacteriol. **171:**6043–6051.
17. **Nunn, D., S. Bergman, and S. Lory.** 1992. Products of three accessory genes, *pilB*, *pilC*, and *pilD*, are required for biogenesis of *Pseudomonas aeruginosa* pili. J. Bacteriol. **172:**2911–2919.
18. **Nunn, D. N., and S. Lory.** 1991. Product of the *Pseudomonas aeruginosa* gene *pilD* is a prepilin leader peptidase. Proc. Natl. Acad. Sci. USA **88:**32821–32825.
19. **Olsen, G. J., C. R. Woese, and R. Overbeek.** 1994. The winds of (evolutionary) change: breathing new life into microbiology. J. Bacteriol. **176:**1–6.
20. **Pugsley, A. P.** 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol. Rev. **57:**50–108.
21. **Pugsley, A. P., C. d'Enfert, I. Reyss, and M. G. Kornacker.** 1990. Genetics of extracellular protein secretion by Gram-negative bacteria. Annu. Rev. Genet. **24:**67–90.
22. **Pugsley, A. P., and O. Possot.** 1993. The general secretory pathway of *Klebsiella oxytoca*: no evidence for relocalization or assembly of pilin-like pulG protein into a multiprotein complex. Mol. Microbiol. **10:**665–674.
23. **Reeves, P. J., D. Whitcombe, S. Wharam, M. Gibson, G. Allison, N. Bunce, D. P. Barallon, V. Mulholland, S. Stevens, D. Walker, and G. P. C. Salmond.** 1993. Molecular cloning and characterization of 13 *out* genes from *Erwinia carotovora* subspecies *carotovora*: genes encoding members of a general secretion pathway (GSP) widespread in Gram-negative bacteria. Mol. Microbiol. **6:**443–456.
24. **Russel, M.** 1994. Phage assembly: a paradigm for bacterial virulence factor export? Science **265:**612–614.
25. **Salmond, G. P. C., and P. J. Reeves.** 1993. Membrane traffic wardens and protein secretion in Gram-negative bacteria. Trends Biochem. Sci. **18:**8–12.
26. **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.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
28. **Stojiljkovic, I., A. J. Bäumler, and K. Hantke.** 1994. Fur regulon of gram-negative bacteria: characterization of new iron-regulated *Escherichia coli* genes by a Fur titration assay. J. Mol. Biol. **236:**531–545.
29. **Strom, M. S., and S. Lory.** 1993. Structure-function and biogenesis of the type IV pili. Annu. Rev. Microbiol. **47:**565–596.
30. **Strom, M. S., D. Nunn, and S. Lory.** 1991. Multiple roles of the pilus biogenesis protein PilD: involvement of PilD in excretion of enzymes from *Pseudomonas aeruginosa*. J. Bacteriol. **173:**1175–1180.
31. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82:**1074–1078.
32. **Whitchurch, C. B., M. Hobbs, S. P. Livingstone, V. Krishnapillai, and J. S. Mattick.** 1991. Characterization of a *Pseudomonas aeruginosa* twitching motility gene and evidence for a specialized protein export system wide spread in bacteria. Gene **101:**33–44.