© *1990 Oxford University Press*

Nucleic Acids Research, Vol. 18, No. 7 **1899**

Complete sequence of IS*629*

Sachiko Matsutani[+] and Eiichi Ohtsubo[*]
Institute of Applied Microbiology, University of Tokyo, Yayoi 1-1-1 Bunkyo-ku, Tokyo 113, Japan

Submitted February 26, 1990

EMBL accession no. X51586

The complete nucleotide sequence of IS*629*, originally isolated from the *Shigella sonnei* chromosome (1), was determined. IS*629* was 1310 base pairs (bp) in length and had 25-bp imperfect inverted repeats at its termini. IS*629* contained two large open reading frames named *orfA* and *orfB*: *orfA* (bp position 55 to 381) encodes 108 amino acids; *orfB* (bp position 378 to 1268) encodes 296 amino acids. A possible promoter and the Shine-Dalgarno (SD) sequence were located upstream of the initiation codon ATG for *orfA*. No such sequences were found in the upstream region of *orfB*. The IS*629* sequence showed partial homology with those of IS*3*, IS*600* (1), and IS2, and showed almost perfect homology with the IS*3411* sequence (2).

ACKNOWLEDGEMENT

This research was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

1. Matsutani,S., Ohtsubo,H., Maeda,Y. and Ohtsubo,E. (1987) *J. Mol. Biol.* **196**, 445–455.
2. Ishiguro,N. and Sato,G. (1988) *J. Bacteriol.* **170**, 1902–1906.

```
   .1 -35                     -10           SD
5'-TGAACCGCCCCGGGAATCCTGGAGACTAAACTTCCTGAGAAAGAGGTAAACAGGATGACTAAAAATACTCGTTTTTCCCC
                                                          M  T  K  N  T  R  F  S  P
                    100.
   CGAAGTCCGTCAACGGGCAGTCCGTATGGTTCTGGAAAGTCAGGGCGAATATGACTCACAATGGGCGACAATTTGTTCCA
     E  V  R  Q  R  A  V  R  M  V  L  E  S  Q  G  E  Y  D  S  Q  W  A  T  I  C  S  I
                                        200.
   TTGCTCCAAAGATTGGCTGTACGCCGGAGACTCTGCGTGTCCGGGTTCGCCAGCATGAGCGGGATACCGGGGGCGGTGAT
      A  P  K  I  G  C  T  P  E  T  L  R  V  R  V  R  Q  H  E  R  D  T  G  G  G  D
                                                            300.
   GGAGGGCTCACCACCGCTGAACGTCAGCGTCTGAAAGAGCTGGAGCGTGAAAATCGTGAACTGCGCCGCAGTAACGATAT
    G  G  L  T  T  A  E  R  Q  R  L  K  E  L  E  R  E  N  R  E  L  R  R  S  N  D  I
                                                                                400.
   CCTTCGCCAGGCTTCCGCTTATTTTGCGAAGGCGGAGTTTGACCGCCTCTGGAAAAAATGATGCCACTGCTGGATAAGCT
     L  R  Q  A  S  A  Y  F  A  K  A  E  F  D  R  L  W  K  K  *
                                                             M  M  P  L  L  D  K  L

   GCGTGAGCAGTACGGGGTCGGACCGCTATGCAGCGAACTGCATATTGCCCCGTCAACGTATTACCACTGTCAGCAACAGC
     R  E  Q  Y  G  V  G  P  L  C  S  E  L  H  I  A  P  S  T  Y  Y  H  C  Q  Q  Q  R
                    500.
   GACATCATCCGGATAAACGCAGTGCCCGTGCGCAGCGCGATGACTGGCTGAAGAAAGAGATACAGCGCGTATACGATGAA
      H  H  P  D  K  R  S  A  R  A  Q  R  D  D  W  L  K  K  E  I  Q  R  V  Y  D  E
                                        600.
   AATCACAAGGTATACGGTGTGCGTAAAGTCTGGCGTCAGTTGTTACGGGAAGGTATCAGAGTGGCCAGATGCACTGTGGC
    N  H  K  V  Y  G  V  R  K  V  W  R  Q  L  L  R  E  G  I  R  V  A  R  C  T  V  A
                                                            700.
   ACGTCTCATGGCGGTTATGGGACTTGCCGGTGTTCTCCGGGGTAAAAAGGTCCGTACGACCATCAGCCGGAAAGCCGTTG
     R  L  M  A  V  M  G  L  A  G  V  L  R  G  K  K  V  R  T  T  I  S  R  K  A  V  A
                                                                                800.
   CCGCAGGCGACCGCGTAAACCGTCAGTTCGTGGCAGAACGACCTGACCAGCTGTGGGTGGCTGATTTTACTTACGTCAGC
      A  G  D  R  V  N  R  Q  F  V  A  E  R  P  D  Q  L  W  V  A  D  F  T  Y  V  S

   ACATGGCAGGGCTTCGTCTATGTGGCGTTCATCATTGATGTGTTTGCCGGATACATCGTGGGGTGGCGGGTCTCATCGTC
    T  W  Q  G  F  V  Y  V  A  F  I  I  D  V  F  A  G  Y  I  V  G  W  R  V  S  S  S
                    900.
   CATGGAAACGACATTCGTGCTGGATGCACTGGAGCAGGCGTTATGGGCCCGTCGACCGTCCGGCACGGTCCATCACAGTG
     M  E  T  T  F  V  L  D  A  L  E  Q  A  L  W  A  R  R  P  S  G  T  V  H  H  S  D
                                       1000.
   ATAAAGGTTCTCAGTATGTATCGCTGGCCTACACACAGCGGCTTAAGGAAGCCGGATTACTGGCATCAACAGGAAGTACA
      K  G  S  Q  Y  V  S  L  A  Y  T  Q  R  L  K  E  A  G  L  L  A  S  T  G  S  T
                                                           1100.
   GGCGACTCGTATGACAACGCGATGGCGGAGAGCATCAATGGTCTTTACAAAGCGAAGGTAATACACCGTAAGAGCTGGAA
    G  D  S  Y  D  N  A  M  A  E  S  I  N  G  L  Y  K  A  K  V  I  H  R  K  S  W  K
                                                                               1200.
   AAACCGTGCAGAAGTGGAACTGGCCACACTCACGTGGGTGGACTGGTATAACAATCGACGATTGCTGGAAAGGCTGGGCC
     N  R  A  E  V  E  L  A  T  L  T  W  V  D  W  Y  N  N  R  R  L  L  E  R  L  G  H

   ATACTCCTCCGGCAGAAGCAGAAAAAGCTTATTATGCTTCCATCGGAAACGATGATCTGGCAGCCTGAGTTCACAGATAA
      T  P  P  A  E  A  E  K  A  Y  Y  A  S  I  G  N  D  D  L  A  A  *
                   1300.     1310.
   AACACTCTCCAGGAAACCCGGGGCGGTTCA-3'
```

Figure 1. The entire nucleotide sequence of IS*629*. Terminal inverted repeat sequences are shown by boldface letters. The possible promoter sequence (−10 and −35 regions) and the Shine-Dalgarno sequence (SD) are underlined. Putative amino acid sequences of the two open reading frames (*orfA* and *orfB*) are shown underneath the nucleotide sequence.

---

[*] To whom correspondence should be addressed

[+] Present address: National Institute of Hygienic Science, Kamiyoga 1-18-1, Setagaya-ku, Tokyo 158, Japan