JOURNAL OF BACTERIOLOGY, Apr. 1992, p. 2384–2387
0021-9193/92/072384-04$02.00/0

Vol. 174, No. 7

NOTES

Similarity between the *Myxococcus xanthus* and *Stigmatella aurantiaca* Reverse Transcriptase Genes Associated with Multicopy, Single-Stranded DNA

MEI-YIN HSU, CHUNYING XU, MASAYORI INOUYE,* AND SUMIKO INOUYE

Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey at Rutgers, Piscataway, New Jersey 08854

Received 14 November 1991/Accepted 28 January 1992

To determine the evolutional relationship of bacterial retroelements of *Myxococcus xanthus* and *Stigmatella aurantiaca*, the nucleotide sequence of 3,060 bases encompassing *msr*, *msd*, and the upstream region of *msd* (downstream of *msr*) of *S. aurantiaca* DW4 was determined and compared with the same region from *M. xanthus*. An open reading frame was found 92 bases upstream of *msd* which encoded a polypeptide of 480 amino acid residues having 73% identity with the reverse transcriptase of *M. xanthus*. Together with high homologies in *msr* (86%) and *msd* (81%) regions, the present data indicate that the reverse transcriptase genes as well as the retrons of *M. xanthus* (retron-Mx162) and *S. aurantiaca* (retron-Sa163) were derived from a common progenitor retron which possibly existed before the two myxobacterial species diverged.

Multicopy single-stranded DNA (msDNA) was originally found in *Myxococcus xanthus*, a gram-negative soil bacterium, as a satellite DNA consisting of a single-stranded DNA of 162 bases (21). A highly homologous msDNA was subsequently found in *Stigmatella aurantiaca* DW4 (16), another myxobacterium, and its entire primary structure of the branched RNA linked msDNA was determined (2, 3). msDNAs from *M. xanthus* and *S. aurantiaca* are designated msDNA-Mx162 and msDNA-Sa163, respectively (9, 10). msDNA-Sa163 consists of a 163-base single-stranded DNA (msDNA) and a 76-base RNA (msdRNA). The 5′ end of the DNA is attached to the 2′-OH group of the rG residue at position 19 of the RNA molecule by a 2′,5′-phosphodiester linkage. The 3′ ends of the DNA and RNA molecules are hybridized to each other by 8 bases. The proposed secondary structure of msDNA-Sa163 is shown in Fig. 1A (2, 3).

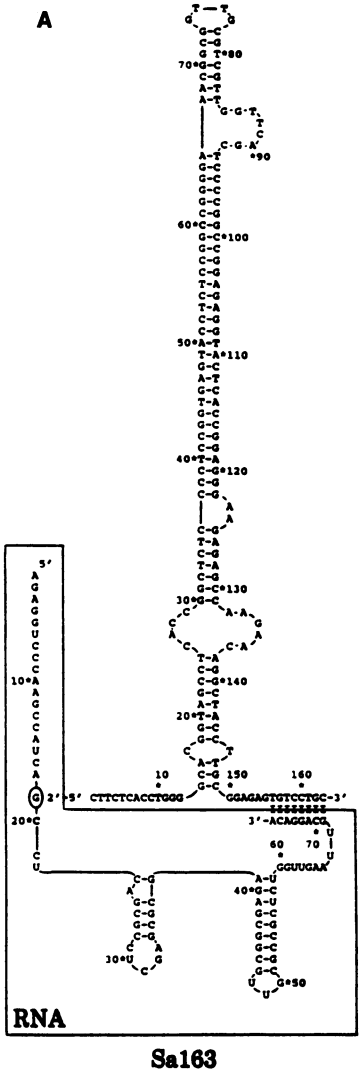
The structure of msDNA-Mx162 has also been determined and shows a high homology (>80%) to msDNA-Sa163 (6). msDNAs have been found in a variety of myxobacteria (4) and also in a minor population of natural isolates of *Escherichia coli* (for reviews, see references 9 and 10). In addition to msDNA-Mx162 and msDNA-Sa163, five other msDNAs have been identified: msDNA-Mx65 (5), msDNA-Ec67 (14), msDNA-Ec73 (18), msDNA-Ec86 (15), and msDNA-Ec107 (7). In all cases, msDNA is encoded from a single chromosomal locus, designated a retron, consisting of *msr*, *msd*, and the gene for reverse transcriptase (RT), which is essential for the biosynthesis of msDNA (for reviews, see references 9 and 10).

In this report, we determined the nucleotide sequence of the RT gene of *S. aurantiaca* and found that the *S. aurantiaca* RT consists of 480 amino acid residues and has 73% identity with the *M. xanthus* RT. Its evolutional implication will be discussed. The RT gene is the first gene sequenced in *S. aurantiaca*.

To determine the nucleotide sequence of the region upstream of *msd* (downstream of *msr*), the 3.6-kb *Bam*HI fragment (from B^e to B^f; Fig. 1B) from pSTA1 was subcloned into the unique *Bam*HI site of pUC9 (19) and the resulting plasmid was designated pSTA5 (Fig. 1B). In pSTA1, the 13.3-kb *Eco*RI fragment from *S. aurantiaca* (2) was cloned into the *Eco*RI site of pACYC184 (1). The 3.6-kb *Bam*HI fragment was digested with *Xho*I, *Sma*I, and *Rsa*I, and the fragments thus generated were subcloned into pUC9. The DNA sequence was determined by the chain termination method (17). The nucleotide sequence of 3,060 bases encompassing *msr*, *msd*, and the gene for RT was thus determined as shown in Fig. 2. A long open reading frame starting from an initiation codon (ATG) was found 92 bases upstream of *msd*. The open reading frame codes for a polypeptide of 480 amino acid residues, and a ribosome-binding sequence (AGG) is found 7 bases upstream of the initiation codon (residues 753 to 755). In Fig. 2, *msd*, the gene for msDNA, and *msr*, the gene for msdRNA, are boxed, and inverted repeat sequences consisting of two 33-base sequences immediately upstream of the branched G residue (residue 458) and downstream of the 5′ end of msDNA are indicated by a1 and a2. These inverted repeat sequences have been shown to be essential in order to form a stem structure in the primary transcript placing the branched G at the end of the stem structure for the msDNA priming reaction (8).

When the nucleotide sequence of *S. aurantiaca* from residues 122 to 3060 is compared with the corresponding nucleotide sequence of *M. xanthus* (12), homologies of *msr* and *msd* are 86 and 81%, respectively (6). There are 18 base

* Corresponding author.



substitutions, 16 base insertions, and 1 base deletion within the 92-base sequence between the 5′ end of msDNA and the initiation codon (residues 763 to 765) in *S. aurantiaca*. Within the 1,440-base open reading frame (residues 763 to 2202), there are 281 base substitutions, 17 base insertions, and 32 base deletions. The homology of this region is 77% between *S. aurantiaca* and *M. xanthus*. The homology at the region of the 3′-end-encoding sequence (residues 2203 to 2205) drops to 50% as a result of many deletion mutations. As expected from the homology of the DNA sequence, the RT amino acid sequence deduced from the DNA sequence is significantly similar to that of RT-Mx162. *M. xanthus* has been shown to contain two independent retrons, retron-Mx162 and retron-Mx65, responsible for two different species of msDNA, msDNA-Mx162 (6) and msDNA-Mx65 (5). Each retron contains a unique region coding for its own RT, RT-Mx162 for retron-Mx162 (12) and RT-Mx65 for retron-Mx65 (11). Figure 3 shows the amino acid sequence of the *S. aurantiaca* RT together with those of RT-Mx162 and -Mx65. There are 351 identical amino acid residues out of 480 residues between RT-Sa163 and RT-Mx162 (73% identity). When the 239-residue sequence encompassing only the RT domain (reference 20 and see below) is compared with RT-Mx162, the identity increases to 80%. This high identity between the two RTs is unique among all the other known bacterial RTs, which do not share more than 40% identities. It is also interesting that the identity between RT-Sa163 and RT-Mx65 is only 35%, as is the identity between RT-Mx162 and RT-Mx65. These results indicate that retron-Sa163 and retron-Mx162 were possibly derived from a common progenitor retron which was most likely acquired before the two myxobacterial species diverged. Supporting this is the fact that all 20 independent, natural isolates of *M. xanthus* from various locations in the world contain retrons which have greater than 80% nucleotide sequence homologies to retron-Mx162 (13).

In contrast to *M. xanthus*, only 13% of *E. coli* natural isolates contain retrons, and these retrons show substantial diversities in their primary sequences and sizes of msDNAs, msdRNAs, and RTs (for reviews, see references 9 and 10).

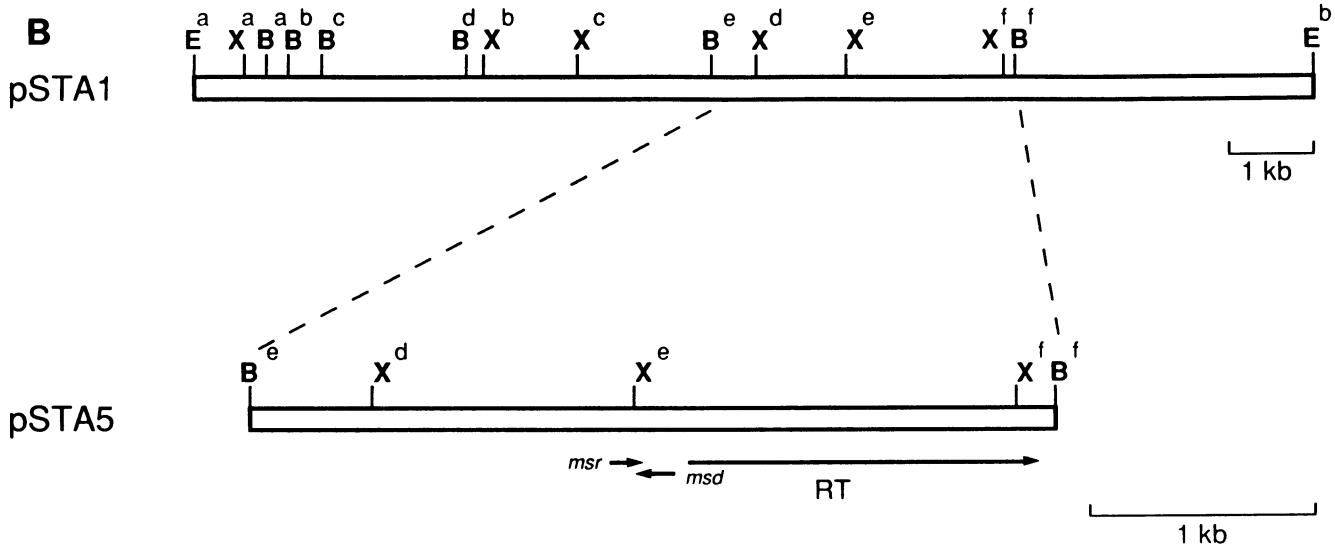


FIG. 1. (A) Proposed structure of msDNA-Sa163. The sequence in the box is RNA. The branched rG is circled. (B) Restriction maps of pSTA1 and pSTA5. The locations and the orientations of *msr*, *msd*, and the RT gene are indicated with arrows. E, B, and X, *Eco*RI, *Bam*HI, and *Xho*I sites, respectively.

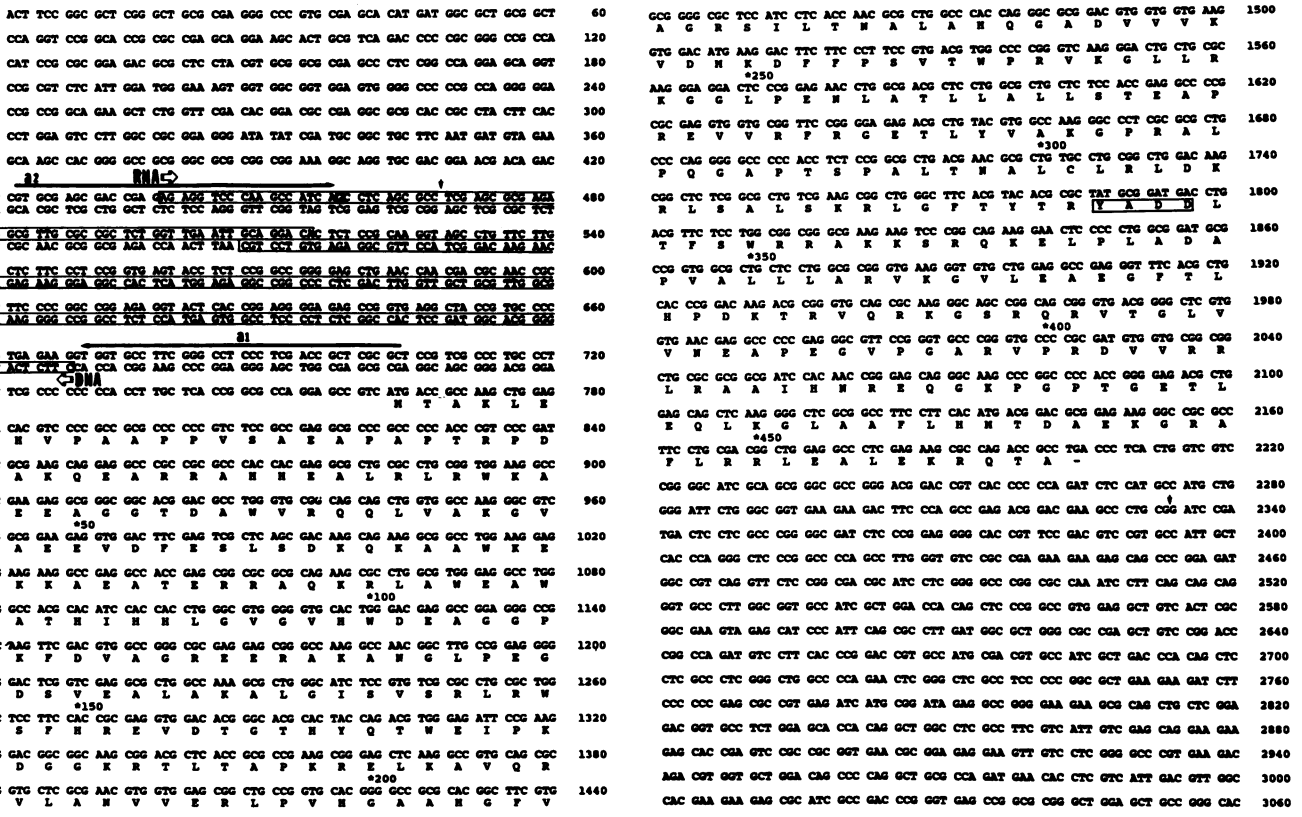


FIG. 2. Nucleotide sequence of 3,060 bases encompassing *msr*, *msd*, and the RT gene of *S. aurantiaca*. Base 1 corresponds to 468 bases upstream of the X^e site in Fig. 1B, and base 3060 corresponds to 727 bases downstream of the B^f site in Fig. 1B. The sequence from base 421 to base 720 which contains *msr* and *msd* is shown double stranded. The boxed regions of the upper strand (bases 440 to 540) and the lower strand (bases 508 to 670) correspond to the sequences of msdRNA and msDNA, respectively (3). The starting sites for msDNA and msdRNA are indicated by open arrows. The circled G at the position 458 is the branched rG of msdRNA linked to the 5′ end of msDNA. Long solid arrows labeled with a1 and a2 represent inverted repeated sequences proposed to form the secondary structure in the primary RNA transcript which serves to prime msDNA synthesis (6). Amino acids are indicated by single letters. The YXDD sequence highly conserved among known RTs is boxed. X^e and B^f sites are indicated by arrows. Numbers on the right-hand side and numbers with asterisks represent numbers for bases and amino acids, respectively.

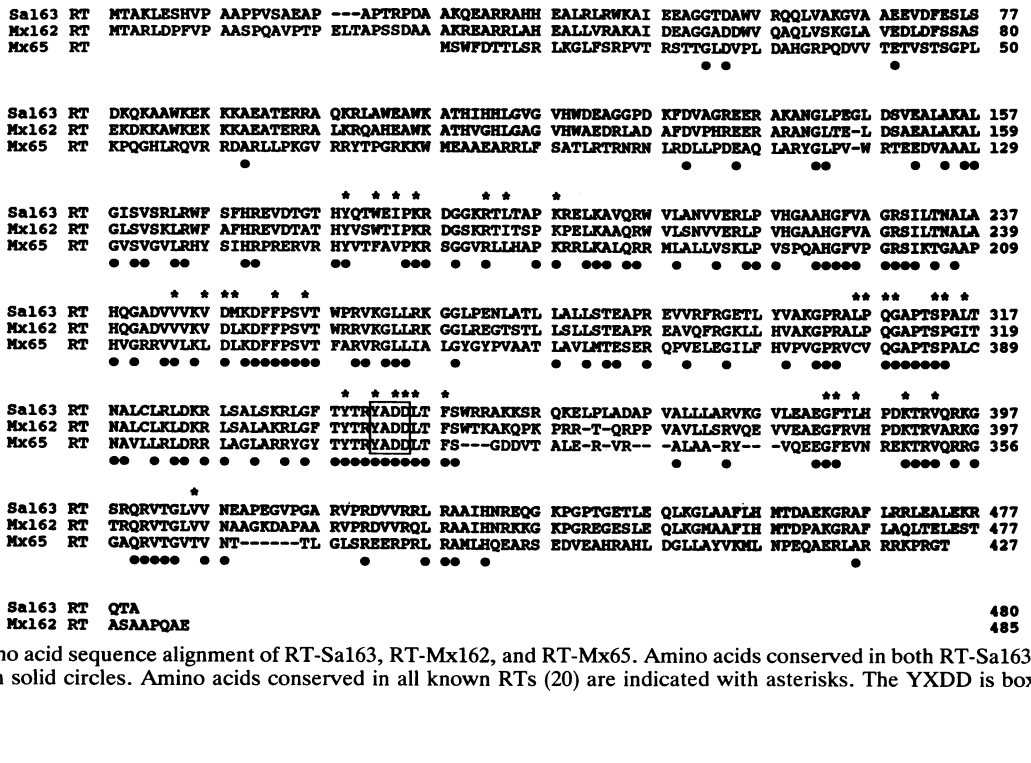


FIG. 3. Amino acid sequence alignment of RT-Sa163, RT-Mx162, and RT-Mx65. Amino acids conserved in both RT-Sa163 and RT-Mx162 are marked with solid circles. Amino acids conserved in all known RTs (20) are indicated with asterisks. The YXDD is boxed.

Therefore, retrons are considered to be acquired into the *E. coli* genome after this bacterial species was established.

From the sequence alignment of 82 bacterial and eukaryotic RTs, the RT structure can be divided into 7 subdomains and contains 17 conserved amino acid residues plus 25 functionally homologous residues (20). In this alignment, 14 out of 17 identical and 18 out of 25 functionally homologous residues are found in RT-Sa163. These identical and homologous residues are marked with asterisks in Fig. 3. The GenBank-EMBL accession number for the RT-Sa163 gene is M86352.

We thank Susan Eagle for the critical reading of the manuscript.

This work was supported by a grant from the National Institute of General Medical Sciences (GM44012).

REFERENCES

1. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. J. Bacteriol. **134:** 1141–1156.
2. **Furuichi, T., A. Dhundale, M. Inouye, and S. Inouye.** 1987. Branched RNA covalently linked to the 5′ end of a single-stranded DNA in *Stigmatella aurantiaca*: structure of msDNA. Cell **48:**47–53.
3. **Furuichi, T., S. Inouye, and M. Inouye.** 1987. Biosynthesis and structure of stable branched RNA covalently linked to the 5′ end of multicopy single-stranded DNA of *Stigmatella aurantiaca*. Cell **48:**55–62.
4. **Dhundale, A., T. Furuichi, S. Inouye, and M. Inouye.** 1985. Distribution of multicopy single-stranded DNA among myxobacteria and related species. J. Bacteriol. **164:**914–917.
5. **Dhundale, A., M. Inouye, and S. Inouye.** 1988. A new species of multicopy single-stranded DNA from *Myxococcus xanthus* with conserved structural features. J. Biol. Chem. **263:**9055–9058.
6. **Dhundale, A., B. Lampson, T. Furuichi, M. Inouye, and S. Inouye.** 1987. Structure of msDNA from *Myxococcus xanthus*: evidence for a long, self-annealing RNA precursor for the covalently linked, branched RNA. Cell **51:**1105–1112.
7. **Herzer, P. J., T. Kawaguchi, S. Inouye, and M. Inouye.** Direct integration of a retron (retron-Ec-107) into the *Escherichia coli* genome by replacing a palindromic 34-bp intergenic sequence. Mol. Microbiol., in press.
8. **Hsu, M.-Y., M. Inouye, and S. Inouye.** 1989. Structure requirements of the RNA precursor for the biosynthesis of the branched RNA-linked multicopy single-stranded DNA of Myxococcus Xanthus. J. Biol. Chem. **264:**6214–6219.
9. **Inouye, M., and S. Inouye.** 1991. Retroelements in bacteria. Trends Biochem. Sci. **16:**18–21.
10. **Inouye, M., and S. Inouye.** 1991. msDNA and bacterial reverse transcriptase. Annu. Rev. Microbiol. **45:**163–186.
11. **Inouye, S., P. J. Herzer, and M. Inouye.** 1990. Two independent retrons with highly diverse reverse transcriptase in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. USA **87:**942–945.
12. **Inouye, S., M.-Y. Hsu, S. Eagle, and M. Inouye.** 1988. Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. Cell **56:**709–717.
13. **Lampson, B. C., M. Inouye, and S. Inouye.** 1991. Survey of multicopy single-stranded DNAs and reverse transcriptase genes among natural isolates of *Myxococcus xanthus*. J. Bacteriol. **173:**5363–5370.
14. **Lampson, B. C., J. Sun, M.-Y. Hsu, J. Vallejo-Ramirez, S. Inouye, and M. Inouye.** 1989. Reverse transcriptase in a clinical strain of *E. coli*: its requirement for the production of branched RNA-linked msDNA. Science **243:**1033–1038.
15. **Lim, D., and W. K. Maas.** 1989. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. Cell **56:**891–904.
16. **Qualls, G. T., K. Stephens, and D. White.** 1979. Morphogenetic movements and multicellular development in the fruiting mxyobacterium, *Stigmatella aurantiaca*. Dev. Biol. **66:**270–274.
17. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
18. **Sun, J., M. Inouye, and S. Inouye.** 1991. Association of a retroelement with a P4-like cryptic prophage (retronphage ϕR73) integrated into the selenocystyl tRNA gene of *Escherichia coli*. J. Bacteriol. **173:**4171–4181.
19. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.
20. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:**3353–3362.
21. **Yee, T., T. Furuichi, S. Inouye, and M. Inouye.** 1984. Multicopy single-stranded DNA isolated from a Gram-negative bacterium, *Myxococcus xanthus*. Cell **38:**203–209.