Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 942–945, February 1990
Biochemistry

Two independent retrons with highly diverse reverse transcriptases in *Myxococcus xanthus*

(multicopy single-stranded DNA/2′,5′-phosphodiester/codon usage/myxobacteria)

SUMIKO INOUYE, PETER J. HERZER, AND MASAYORI INOUYE

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

Communicated by Russell F. Doolittle, November 27, 1989 (received for review September 29, 1989)

ABSTRACT **A reverse transcriptase (RT) was recently found in *Myxococcus xanthus*, a Gram-negative soil bacterium. This RT has been shown to be associated with a chromosomal region designated a retron responsible for the synthesis of a peculiar extrachromosomal DNA called msDNA (multicopy single-stranded DNA). We demonstrate that *M. xanthus* contains two independent, unlinked retrons, one for the synthesis of msDNA-Mx162 and the other for msDNA-Mx65. The structural analysis of the retron for msDNA-Mx65 revealed that the coding regions for msdRNA (*msr*) and msDNA (*msd*), and an open reading frame (ORF) downstream of *msr* are arranged in the same manner as found for the Mx162 retron. The ORF encodes a polypeptide of 427 amino acid residues. The amino-terminal domain (residues 1–138) shows no striking similarity to these proteins presently available in the data bases including the msDNA-Mx162 ORF, while the sequence from residues 139–394 can be aligned with various known RT sequences and has 47% identity with the RT domain of the msDNA-Mx162 ORF. On the basis of these findings, possible origins of two highly diverse retrons on the *M. xanthus* chromosome are discussed.**

msDNA (multicopy single-stranded DNA) is an unusual satellite DNA originally found in *Myxococcus xanthus*, a Gram-negative bacterium with a complex life cycle, which forms fruiting bodies (1). The major msDNA of *M. xanthus* (msDNA-Mx162) consists of 162 bases of single-stranded DNA, the 5′ end of which is linked to the 20th guanosine residue of a 77-base RNA molecule (msdRNA) by a unique 2′,5′-phosphodiester bond (2). The msDNA molecule has been shown to be synthesized by a reverse transcriptase (RT) using a precursor RNA as a primer for initiating msDNA synthesis as well as a template to form the branched RNA-linked msDNA (2–4). The gene for this RT has been demonstrated to be located immediately downstream of the msdRNA coding region and encodes a polypeptide of 485 amino acid residues. msDNA-synthesizing systems similar to that of *M. xanthus* were also found in *Escherichia coli* (5, 6). Temin (7) has proposed to designate these systems "retrons," implying that the retron may be an ancestor of retroviruses. Recently, we found that *M. xanthus* contains another msDNA (msDNA-Mx65; previously designated mrDNA) consisting of a 65-base single-stranded DNA and a 49-base branched RNA (8). Although there is no sequence homology in either the DNA or RNA between msDNA-Mx162 and msDNA-Mx65, they share remarkable similarities in their secondary structures, including stem–loop structures in DNA and RNA, RNA and DNA hybrid formation at their 3′ ends, and the unique 2′,5′-phosphodiester linkage (8). In this paper, we characterized the retron for msDNA-Mx65 and compared it with the retron for msDNA-Mx162, which has been cloned (8).* It was found that the Mx65 retron is highly diverse and independent from the Mx162 retron and that RT associated with the Mx65 retron has only 47% identity with RT for the Mx162 retron.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Materials. The clone of the 9.0-kilobase (kb) *Pst* I fragment containing the Mx65 retron was obtained (8). Restriction enzymes were purchased from New England Biolabs and Boehringer Mannheim.

A deletion mutant strain of the Mx162 retron, ΔmsSX, was previously isolated (9). A deletion mutant of the Mx65 retron (Δms65) was constructed as described for ΔmsSX (9).

Methods. The DNA sequence was determined by the chain-termination method (10) using synthetic oligonucleotides as primer.

RESULTS AND DISCUSSION

DNA Sequencing of the Mx65 Retron. The coding region for msDNA-Mx65 was previously cloned from a 9.0-kb *Pst* I fragment (8) and its restriction map is shown in Fig. 1. Previously, the 283-base-pair (bp) *Alu* I fragment [*Alu* I(A) to *Alu* I(B) in Fig. 1] was sequenced, which contains the RNA coding region (msdRNA) as well as the msDNA coding region as shown in Fig. 1. It should be noted that on the basis of the restriction map of the msDNA-Mx162 fragment (1) as well as Southern blot hybridization experiments with msDNA-Mx162 and msDNA-Mx65 as probes (unpublished results), the msDNA-Mx65 coding region is separated from the msDNA-Mx162 coding region by at least 10 kb. These two regions thus appear to be unlinked on the *M. xanthus* chromosome.

Fig. 2 shows the DNA sequence of 1640 nucleotides flanking the msDNA-Mx65 coding region. As in the case of msDNA-Mx162, the long ORF exists downstream of the msdRNA coding region (see also Fig. 1). The ORF starts from the initiation codon (residues 279–281) 28 bases upstream of the 5′ end of msDNA [in the case of msDNA-Mx162, 77 bases (4)] and encodes a polypeptide of 427 amino acid residues. In Fig. 2, the msDNA coding region [*msd*(Mx65) originally designated *mrd*] is indicated by the closed box on the lower strand and the orientation is from right to left. Similarly, the msdRNA sequence [*msr*(Mx65) originally designated *mrr*] is indicated by the closed box on the upper strand and the orientation is from left to right. The *msd* and *msr* regions overlap by 6 bases. Arrows with letters a1 and a2 also indicate an inverted repeat, which comprises a 14-base sequence immediately upstream of the branched G residue (residues 132–145; Fig. 2, sequence a2) and another 14-base sequence upstream of *msd* (residues 265–252; sequence a1). Although

Abbreviations: msDNA, multicopy single-stranded DNA; RT, reverse transcriptase; ORF, open reading frame.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. M30609).

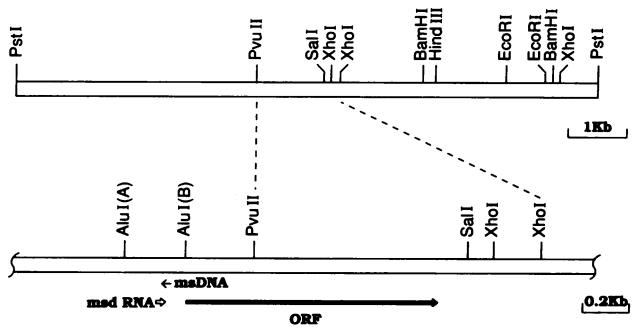

FIG. 1. Restriction map of the 9.0-kb *Pst* I fragment containing the Mx65 retron. The location and orientation of the genes encoding msDNA and msdRNA are indicated by a thin arrow and an open arrow, respectively. The large solid arrow represents an open reading frame (ORF) and its orientation. Only two *Alu* I sites (A and B) are shown. The DNA sequence between *Alu* I(A) and *Alu* I(B) was determined by Dhundale *et al.* (8).

this inverted repeat is much shorter than that found in the Mx162 retron and has three mismatches between sequences a1 and a2, it is considered to be essential to form a secondary structure in a long primary transcript that serves as the primer as well as the template for msDNA synthesis (2, 15).

Sequence Comparison Between Mx65-RT and Mx162-RT. In the case of msDNA-Mx162, an ORF of 485 residues exists in a chromosomal locus, organized in a manner identical to that shown in Fig. 2, which encodes a RT. The Mx162-RT consists of at least two domains with the RT domain assigned to the sequence from residues 170–441 on the basis of sequence similarity with retroviral RTs and RT from *E. coli* retrons. A similar RT domain can be found in the sequence from residues 139–394 in the Mx65-RT and its alignment with other RT sequences is shown in Fig. 3. It also contains the Tyr-Xaa-Asp-Asp sequence, the highly conserved sequence in all known RTs (boxed in Figs. 2 and 3). Although the Mx65-RT is substantially different from the Mx162-RT, within the 256-residue sequence there are 122 identical residues between the two RTs (47% identity). When compared with *E. coli* RTs, 52 residues (20%) are shared with all bacterial RTs (Fig. 3, open and solid circles). There are a total of 75 and 76 residues identical with the *E. coli* Cl-I-RT (Ec67) and the *E. coli* B-RT (Ec86) (29% and 30%, respectively). When compared with the human immunodeficiency virus RT (11), there are a total of 32 identical residues (13%) among which 15 residues (solid circles) are shared with all the RTs listed in Fig. 3. As noted previously, the region surrounding the Tyr-Xaa-Asp-Asp sequence shows particularly high similarity among all RTs (4). The sequence comparison shown in Fig. 3 clearly demonstrates that the Mx65-RT is more closely related to the Mx162-RT than to *E. coli* RTs. This can also be seen from the sequence from residues 246–256, which is deleted in all the other RTs except for the Mx162-RT.

The amino-terminal domain from residues 1–138 showed no striking similarity to any other known proteins presently available in the data bases. It is also rather surprising to find that there is little similarity between this domain and the amino-terminal domain (169 residues) of the Mx162-RT except for the sequence of 25 residues from residues 114–138 in the Mx65-RT. This sequence can be aligned with the Mx162-RT sequence from residues 145–169 having 11 identical residues (44% identity). It was previously pointed out that the codon usage in the Mx162-RT gene is very similar to those of other genomic genes so far characterized for *M. xanthus* using high GC codons (4). As shown in Table 1, the codon usage of the Mx65-RT gene is also very similar to that of the Mx162-RT gene, suggesting that the Mx65-RT gene is also as old as the other genomic genes of *M. xanthus*.

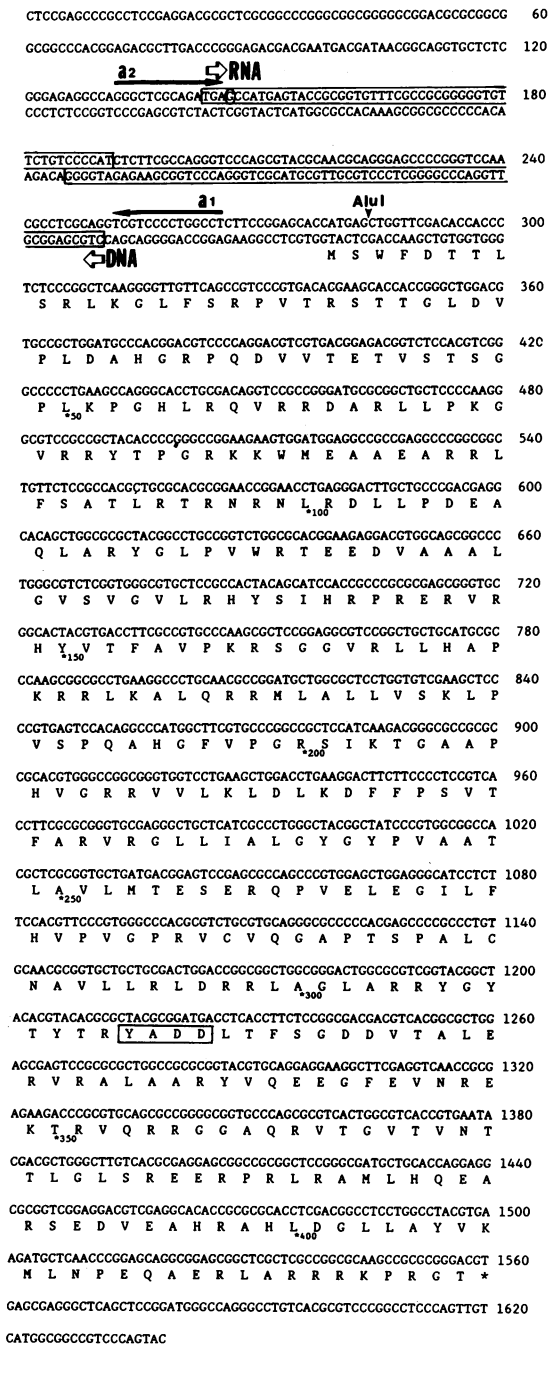

FIG. 2. Nucleotide sequence of the chromosomal region encompassing the Mx65-msDNA and msdRNA coding regions and an ORF region downstream of *msr*. The sequence covers from the *Alu* I(A) site (see Fig. 1) to 78 bp downstream of the ORF. The complementary strand is only shown from bases 121–300. The boxed region of the upper strand (positions 143–191) and the boxed region of the lower strand (positions 186–250) correspond to the sequences of msdRNA and msDNA, respectively. The starting sites for DNA and RNA and the 5′ to 3′ orientation are indicated by open arrows. The msdRNA and msDNA regions overlap at their 3′ ends by 6 bases. The circled G residue at position 206 represents the branched guanosine of RNA linked to the 5′ end of the DNA strand in msDNA. Long solid arrows labeled a1 and a2 represent inverted repeat sequences proposed to be important in the secondary structure of the primary RNA transcript involved in the synthesis of msDNA (8). The ORF begins with the initiation codon at base 279. The YXDD amino acid sequence highly conserved among known RT proteins is boxed. Numbers on the right-hand column enumerate the nucleotide bases, and numbers with asterisks enumerate amino acids (single-letter code). Arrowhead labeled *Alu* I indicates *Alu* I(B) in Fig. 1. The DNA sequence was determined by the chain-termination method (10) using synthetic oligonucleotides as primers.

```
HIV    177 VKLKPGMDGP KVKQ..WPLT E..EKIKALV EICTEMEKEG KISKIGPENP 222

Ec86    37 LRLLTYTADF KYRIYTVEKK GDEKRMRTIY QPSREL.KAL QGWV...LRN  82
Ec67    32 NVLYRIGSDN QYTQFTIPKK G..KGVRTIS APTDRL.KDI QRRICDLLSD  78
Mx162  170 AFHREVDTAT HYVSWTIPKR D..GSKRTIT SPKPEL.KAA QR...WVLS. 212
Mx65   139 SIHRPRERVR HYVTFAVPKR S..GGVRLLH APKRRL.KAL QR...RMLA. 181
             xx       xox    xox    x  o     ox  o ●x  ox     o

HIV    223 YNTPVFAIKK KDSTKWRKLV DFRELNKRTQ DFWEVQLGIP HPAGLKKK.K 271

Ec86    83 ILDKL..... S.SSPFSIGF E.KHQSILNN ATPHIGANFI LNIDLEDFFP 125
Ec67    79 CRDEIFAIRK I.SNNYSFGF E.RGKSIILN AYKHRGKQII LNIDLKDFFE 126
Mx162  213 ...NVV.ERL P.VHGAAHGF V.AGRSILTN ALAHQGADVV VKVDLKDFFP 256
Mx65   182 ...LLV.SKL P.VSPQAHGF V.PGRSIKTG AAPHVGRRVV LKLDLKDFFP 225
             x   x x x   xxoo x  xxoo x  o  o o  xx  x o●xooox

HIV    272 SVTVLDVGDA YFSVPLDEDF RKYT..... .AFTIP.SIN NETPGIRYQY 313

Ec86   126 SLTANKVFG. VFHS.L.... ..........  .GYN.RLISS VLTKICCYK. 156
Ec67   127 SFNFGRVRG. YFLS..NQDF ..........  .LLN.PVVAT TLAKAACYN. 161
Mx162  257 SVTWRRVKGL LRKGGLREGT STLLSLLSTE APREAVQFRG KLLHVAKGP. 305
Mx65   226 SVTFARVRGL LIALGYGYPV AATLAVLMTE SERQPVELEG ILFHVPVGP. 274
             ●xx  x● ox x   x          x  x xx   x  x   x  o xx  xx

HIV    314 NVLPQGWKGS PAIFQS...S MTKILEPFKK QNPDIVIYQY MDDLYVGS.D 359

Ec86   157 NLLPQGAPSS PKLANLICSK LDYRIQGYAG SR.GLIYTRY ADDLTL.SAQ 204
Ec67   162 GTLPQGSPCS PIISNLICNI MDMRLAKLAK KY.GCTYSRY ADDITI.STN 209
Mx162  306 RALPQGAPTS PGITNALCLK LDKRLSALAK RL.GFTYTRY ADDLTF.SWT 353
Mx65   275 RVCVQGAPTS PALCNAVLLR LDRRLAGLAR RY.GYTYTRY ADDLTF.S.. 320
             x   ●●xox● ●   ox  x  xo ox  xo  x  o xoxo● o●●xox ●

HIV    360 LEIGQHRTKI EELRQH.... .....LLRWG LTTPDKK... HQKEPPFLWM 397

Ec86   205 ......SMKK V....VKARD FLFSIIPSEG LVINSKKTCI SGPRSQRKVT 244
Ec67   210 KNTFPLEMAT VQPEGVVLGK VLVKEIENSG FEINDSKTRL TYKTSRQEVT 259
Mx162  354 KAKQPKPRRT QRPPVAVLLS RVQEVVEAEG FRVHPDKTRV ARKGTRQRVT 403
Mx65   321 .GDDVTALER VR...ALAAR Y....VQEEG FEVNREKTRV QRRGGAQRVT 362
                      x   x          x  x● x x   ●oxx  x x  xxoo

HIV    398 GYELHPDKWT VQPIVLPEKD SWTVNDIQKL VGKLNWAS 435

Ec86   245 GLVISQEKVG .....IGREK YKEIRAKIHH IFCGKSSE 277
Ec67   260 GLTVNR.... ..IVNIDRCY YKKTRALAHA LYRTGEYK 291
Mx162  404 GLVVNAAGKD APAARVPRDV VRQLRAAIHN RKKGKPGR 441
Mx65   363 GVTVNT.... ..TLGLSREE RPRLRAMLHQ EARSEDVE 394
             ● xx             o      xoo  o
```

FIG. 3. Amino acid sequence alignment (single-letter code) of the msDNA-Mx65 ORF with three known bacterial RTs (Ec86[6], Ec67[5], and Mx162[4]) and human immunodeficiency virus (HIV) RT (11). Sequence alignment was carried out according to Devereux *et al.* (12): ●, amino acid residues shared by all five proteins; ○, amino acid residues shared by four bacterial RTs; x, amino acid residues shared by Mx65- and Mx162-RTs. Numbers on the right and left indicate the amino acid position from the amino terminus for each RT. The YXDD consensus sequences are boxed.

Analysis of Deletion Mutations. Next, to examine the possible functions of these genes, deletion mutations of the Mx65 retron with and without an Mx162 retron deletion mutation were constructed. Fig. 4 shows the DNA pattern of polyacrylamide gel electrophoresis of the DNA preparations from these mutants. It appears that there is no effect on msDNA-Mx162 production by the deletion of the Mx65 retron (lane 2) and that, similarly, the deletion of the Mx162 retron has no effect on msDNA-Mx65 production (lane 4). A double deletion (lane 3) was also constructed. None of these deletion strains showed any effect on cellular growth, cellular mobility, fruiting body formation, spore formation, or germination.

It was interesting to test whether RTs are able to functionally complement each other. We have previously demonstrated that a deletion mutation at the region 100 bp upstream of *msd* (Mx162) and an insertion mutation at a site 500 bp upstream of *msd* (Mx162) caused a significant reduction in msDNA production (9). When these mutations were combined with the deletion mutation of the Mx65 retron (δms65), no effect on the production of msDNA-Mx162 was observed (data not shown). This result indicates that the reduced production of msDNA-Mx162 in the mutants appears to be due to mutation of the Mx162-RTs but not to complementation by Mx65-RT. The present data suggest that Mx65-RT is

Table 1. Codon usage of the Mx65- and Mx162-RT genes

		Mx65		Mx162*	
aa	Codon	*n*	%	*n*	%†
Ala	GCU	1	2	4	5
	GCC	17	42	25	34
	GCA	3	7	2	3
	GCG	20	49	43	58
Arg	CGU	3	5	3	7
	CGC	26	42	25	56
	CGA	5	8	1	2
	CGG	27	44	15	33
	AGA	0	0	0	0
	AGG	1	2	1	2
Asn	AAU	1	17	1	17
	AAC	5	83	5	83
Asp	GAU	3	18	0	0
	GAC	14	82	24	100
Cys	UGU	0	0	0	0
	UGC	2	100	1	100
Gln	CAA	1	9	0	0
	CAG	11	92	16	100
Glu	GAA	2	9	5	17
	GAG	21	91	25	83
Gly	GGU	1	3	0	0
	GGC	23	72	28	90
	GGA	3	9	1	3
	GGG	5	16	2	7
His	CAU	2	17	0	0
	CAC	10	83	15	100
Ile	AUU	0	0	2	29
	AUC	4	100	5	71
	AUA	0	0	0	0
Leu	UUA	0	0	0	0
	UUG	3	5	2	4
	CUU	0	0	0	0
	CUC	15	28	12	25
	CUA	0	0	0	0
	CUG	36	67	35	71
Lys	AAA	0	0	0	0
	AAG	15	100	38	100
Met	AUG	6	—	3	—
Phe	UUU	0	0	0	0
	UUC	11	100	14	100
Pro	CCU	0	0	1	4
	CCC	15	58	10	37
	CCA	3	11	0	0
	CCG	8	31	16	59
Ser	UCU	0	0	1	4
	UCC	8	42	12	46
	UCA	1	5	1	4
	UCG	4	21	7	27
	AGU	1	5	0	0
	AGC	5	27	5	19
Thr	ACU	1	3	0	0
	ACC	10	36	7	24
	ACA	1	3	0	0
	ACG	16	58	22	76
Trp	UGG	3	—	9	—
Tyr	UAU	1	8	0	0
	UAC	11	92	3	100
Val	GUU	1	2	0	0
	GUC	16	38	7	18
	GUA	0	0	0	0
	GUG	25	60	31	82
Total		427		485	
	G or C in the 3rd base, %		92		96

*From ref. 4.
†Calculated for each amino acid.

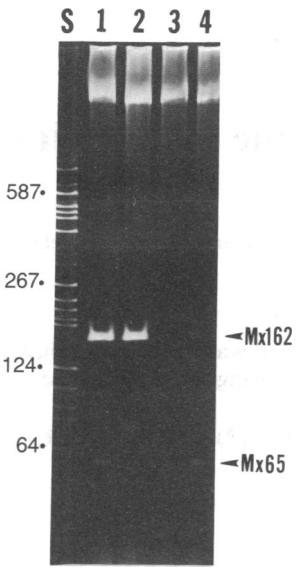

FIG. 4. Deletion mutations at the *msr–msd* region for the Mx65 retron in *M. xanthus* DZF1. Strain ΔmsSX is a deletion mutant of the Mx162 retron isolated from strain DZF1 (9). A deletion mutation of the Mx65 retron (Δms65) was constructed as described for ΔmsSX (9), in which the entire *msd–msr* region and the RT gene was replaced with a gene for streptomycin resistance. Total DNA was prepared from DZF1, Δms65, Δms65/ΔmsSX, and ΔmsSX and treated with RNase A (1). Samples were electrophoresed on a 5% polyacrylamide gel. Lanes: S, pBR322 digested by *Msp* I as a molecular weight marker; 1, DNA from DZF1; 2, Δms65; 3, Δms65/ΔmsSX; 4, ΔmsSX. Numbers on the left indicate sizes of DNA fragments in bp. Positions of msDNA-Mx162 and -Mx65 treated with RNase A are also indicated by arrows.

probably unable to complement the Mx162-RT responsible for synthesizing the longer msDNA-Mx162 molecule.

CONCLUSIONS

In the present paper, we demonstrate that there are two independent, unlinked retrons in the *M. xanthus* genome. The msdRNA coding region (*msr*), the msDNA coding region (*msd*), inverted repeats found in their regions, and an ORF of the Mx65 retron were arranged in a manner very similar to the Mx162 retron previously reported (4). The inverted repeats support the formation of a stable secondary structure of a primary transcript from this region, which is essential for the synthesis of msDNA by serving not only as a primer but also as a template. However, in spite of the similarity in the arrangement of the retron's elements, there is little nucleotide sequence similarity in these elements between the two retrons except for the RT domains in the ORFs downstream of *msr*.

Only clear sequence similarities were found within a part of the ORF, which can be aligned with the RT domain of the Mx162 ORF as well as other bacterial RTs and retroviral RTs (see Fig. 3). It is interesting to note that msDNA-Mx162 from *M. xanthus* is highly conserved among more than 20 independent *M. xanthus* strains isolated from all over the world (ref. 13; B. Lampson, M.I., and S.I., unpublished data). In addition, the codon usage of the Mx162-RT was very similar to other *M. xanthus* genes (4). Thus, it was assumed that the Mx162-RT is as old as other essential genomic genes. It should be noted that *Stigmatella aurantiaca*, another species of myxobacteria contains msDNA-Sa163, which was shown to be highly homologous to msDNA-Mx162 (14), but does not contain a msDNA homologous to msDNA-Mx65 (unpublished results). Therefore, it is tempting to assume that an ancestral myxobacterium already contained a retron related to the Mx162 (or Sal63) retron. On the other hand, it is not certain when and how the Mx65 retron was acquired into the *M. xanthus* genome. Since the Mx65-RT is more closely related to the Mx162-RT than other *E. coli* RTs, it is likely that the Mx65 retron diverged from an ancestral retron common for both the Mx65 and Mx162 retrons. Examination of the existence of msDNA-Mx65 in various independent natural isolates of *M. xanthus* may provide some insight into the question of the origin of the Mx65 retron.

We are grateful to Dr. Bert Lampson for his critical reading of the manuscript. The present work was supported by a grant from the National Institutes of Health (GM26843) and a grant from Takara Shuzo Co., Ltd.

1. Yee, T., Furuichi, T., Inouye, S. & Inouye, M. (1984) *Cell* **38,** 203–209.
2. Dhundale, A., Lampson, B. C., Furuichi, T., Inouye, M. & Inouye, S. (1987) *Cell* **51,** 1105–1112.
3. Lampson, B. C., Inouye, M. & Inouye, S. (1989) *Cell* **56,** 701–707.
4. Inouye, S., Hsu, M.-Y., Eagle, S. & Inouye, M. (1989) *Cell* **56,** 709–717.
5. Lampson, B. C., Sun, J., Hsu, M.-Y., Vallejo-Ramirez, J., Inouye, S. & Inouye, M. (1989) *Science* **243,** 1033–1038.
6. Lim, D. & Maas, W. K. (1989) *Cell* **56,** 891–904.
7. Temin, H. M. (1989) *Nature (London)* **339,** 254–255.
8. Dhundale, A., Inouye, M. & Inouye, S. (1988) *J. Biol. Chem.* **263,** 9055–9058.
9. Dhundale, A., Furuichi, T., Inouye, M. & Inouye, S. (1988) *J. Bacteriol.* **170,** 5620–5624.
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
11. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* **313,** 277–283.
12. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
13. Dhundale, A., Furuichi, T., Inouye, S. & Inouye, M. (1985) *J. Bacteriol.* **164,** 914–917.
14. Furuichi, T., Inouye, S. & Inouye, M. (1987) *Cell* **48,** 55–62.
15. Hsu, M.-Y., Inouye, S. & Inouye, M. (1989) *J. Biol. Chem.* **264,** 6214–6219.