Phylogenetic Comparison of Retron Elements among the Myxobacteria: Evidence for Vertical Inheritance

SCOTT A. RICE AND BERT C. LAMPSON*

Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996

Received 25 July 1994/Accepted 15 October 1994

Twenty-eight myxobacterial strains, representing members from all three subgroups, were screened for the presence of retron elements, which are novel prokaryotic retroelements encoding reverse transcriptase. The presence of retrons was determined by assaying strains for a small satellite DNA produced by reverse transcription called multicopy, single-stranded DNA (msDNA). An msDNA-producing retron appeared to be absent from only one of the strains surveyed. DNA hybridization experiments revealed that retron elements similar to retron Mx162, first identified in *Myxococcus xanthus*, were found only among members of the *Myxococcus* subgroup; that is, each of the seven different genera which constitute this subgroup contained a Mx162 homolog. Another retron element also appeared to have a clustered distribution, being found exclusively within the *Nannocystis* subgroup of the myxobacteria. A retron element of the Mx162 type was cloned from *Melittangium lichenicola*, and its DNA sequence was compared with those of similar elements in *M. xanthus* and *Stigmatella aurantiaca*. Together, the degree of sequence diversity, the codon bias of the reverse transcriptase genes, and the clustered distribution of these retrons suggest a possible evolutionary scenario in which a common ancestor of the *Myxococcus* subgroup may have acquired this retroelement.

In 1984, Yee et al. (39) discovered a small, high-copy-number satellite DNA associated with the gram-negative soil bacterium Myxococcus xanthus. This DNA consisted of a single strand of DNA covalently linked by a 2'-5' phosphodiester bond to an internal guanine residue of a single-stranded RNA (6, 9). These molecules, termed multicopy, single-stranded DNA (msDNA), are encoded by a chromosomal gene called a retron (36). The original retron found in M. xanthus is designated Mx162 (16, 39), and a second unique retron element, designated Mx65, resides with Mx162 on the chromosome of M. xanthus (5). In addition to coding for the RNA and DNA regions of msDNA, the chromosomal gene also contains an open reading frame (ORF) that shows significant similarity to retroviral reverse transcriptases (RTs) (16). These genes associated with msDNA are the first prokaryotic RTs to be discovered (16, 20, 24). Other types of RT such as the newly discovered intron-encoded ORFs also appear to exist in prokaryotes (8). Synthesis of msDNA is postulated to occur by reverse transcription and is dependent on the production of a functional RT (16, 18, 24). Deletion of the retrons in M. xanthus does not affect growth or development of the bacterium (3, 16). Additionally, since the elements are not ubiquitous among bacterial strains, it is apparent that the elements are not required for survival, and despite the widespread nature of retrons, no function has been ascribed to them.

In addition to *M. xanthus* and the myxobacterium *Stigmatella aurantiaca* (9), retrons have been reported in 9 to 15% of all *Escherichia coli* strains tested (11, 20, 23, 35). Recently, we reported the presence of retron elements in several distantly related bacterial groups, suggesting that msDNA and RT genes are prevalent in the prokaryotes (29). As with *E. coli*, only a small percentage (5 to 15%) of strains of *Klebsiella pneumoniae*, *Proteus mirabilis*, *Salmonella* sp., *Rhizobium* sp., and *Bradyrhizobium* sp. tested contain a retron (29).

A number of characteristics distinguish the retron elements

of the myxobacteria from retrons found in the rest of the proteobacteria. First, unlike the case for other bacterial groups, in which retrons occur rarely, the elements appear to be nearly ubiquitous among natural *M. xanthus* strains (19). Second, DNA sequence analysis of retrons from *M. xanthus* and *S. aurantiaca* indicates that the codon bias of the RT gene is similar to the codon bias observed for other known myxobacterial genes (16). These features aroused speculation that bacterial RTs of the myxobacteria are ancient elements and that they evolved prior to the emergence of RTs and retroelements in eukaryotes (14). These intriguing evolutionary questions prompted a closer examination of the retroelements of the myxobacteria.

This report presents the findings of an extensive survey of the retron elements found among the myxobacteria. All but one of the 28 strains examined contained an msDNA-producing element. In addition, several different myxobacteria groups were found to contain a retron element similar to Mx162, a retron previously identified in *M. xanthus*. One of the retron elements, from the myxobacterium *Melittangium lichenicola*, was cloned, and its DNA sequence was compared with those of similar elements in *M. xanthus* and *S. aurantiaca*. Together, the degree of sequence diversity, the codon bias of the RT genes, and the distribution of these retrons suggest a possible evolutionary scenario in which a common ancester of the *Myxococcus* subgroup probably acquired this retroelement.

MATERIALS AND METHODS

Bacterial strains and plasmids. The myxobacterial strains used for this study and their sources are listed in Table 1. *E. coli* JM109 and *E. coli* DH5 α were used for cloning in conjunction with plasmids pB322 and pUC19 (38). *M. xanthus* DZF1 was grown at 30°C in CYE broth (28). *Chondromyces* and *Sorangium* strains were grown on VY/2 plates (0.5% baker's yeast, 0.1% CaCl₂ · 2H₂O, 1.5% agar) at 30°C (28). All other myxobacteria strains were grown at 30°C on MD1 plates (0.3% Difco Casitone, 0.07% CaCl₂ · 2H₂O, 0.2% MgSO₄ · 7H₂O, trace elements, 1.5% agar) (28). *E. coli* was grown at 37°C in modified Luria-Bertani medium (25). When necessary, tetracycline (12 µg/ml) or ampicillin (50 µg/ml) was added for selection of plasmids.

Detection of msDNA. Total RNA was purified by the guanidinium thiocyanate method (1) from freshly grown cells or frozen cell pellets. Purified RNA (this fraction also contains msDNA) was stored in ethanol at -80° C. msDNA was

^{*} Corresponding author. Phone: (615) 974-4015. Fax: (615) 974-4007. Electronic mail address: Lampson@UTKVX.UTK.EDU.

	<u>.</u>	msDNA]	Hybridizat	ion	D f					
Organism screened	Strain	production	Mx162	2 Mx65 Na e434		Reference	Location and/or strain source				
Subgroup Myxococcus											
Angiococcus disciformis	ATCC 33172	+	+	-	—	This study					
Archangium gephyra	ATCC 25201	+	+	-	—	This study					
Corallococcus (Myxococ- cus) coralloides	DK817	$+^{a}$	$+^{b}$	ND^{c}	ND	4	Yosemite, Calif.; Dale Kaiser				
	ATCC 25202	+	+	-	_	This study					
Cystobacter ferrugineus	Cb fe17	+	+	ND	ND	This study	Hans Reichenbach				
	Cb fe18	_	-	-	_	This study	Hans Reichenbach				
Cystobacter fuscus	Cb f17	+	+	ND	ND	This study	Hans Reichenbach				
Cystobacter violaceus	Cb vi4	+	+	ND	ND	This study	Hans Reichenbach				
Melittangium lichenicola	ATCC 25946	+	+	-	—	This study	Hans Reichenbach				
Myxococcus fulvus	ATCC 25199	+	+	-	_	This study					
Myxococcus macrosporus	ATCC 29619	+	+	+	_	This study					
Myxococcus stipitatus	ATCC 29611	+	+	-	_	This study					
Myxococcus virescens	ATCC 25203	+	+	+	_	This study					
Myxococcus xanthus	DZF1	+	+	+	_	19	David Zusman				
	19 other strains	+	+	+/-	_	19	Hans Reichenbach and Dale Kaiser				
Stigmatella aurantiaca	DW4	+	+	-	_	19	Minneapolis, Minn.; David White				
Subgroup Chondromyces											
Chondromyces apiculatus	Cm a15	+	-	-	_	This study					
	Cm a16	+	-	-	_	This study					
Chondromyces pedicalatus	Cm p5	+	-	-	_	This study	Montes Claros, Brazil; Hans Reichenbach				
Sorangium cellulosum	So ce5	+	-	-	-	This study	Cefalù, Sicily; Hans Reichenbach				
0	So ce7	+	-	-	_	This study	Delphi, Greece; Hans Reichenbach				
	So ce11	+	-	-	_	This study	Yucatan, Mexico; Hans Reichenbach				
Subgroup Nannocystis						2					
Nannocystis exedens	Na e1	+	-	-	+	This study	Hans Reichenbach				
2	Na e3	+	-	-	+	This study	Nebraska, Hans Reichenbach				
	Na e24	+	-	-	+	This study	Tenerife, Spain; Hans Reichenbach				
	Na e30	+	-	-	+	This study	Calpe, Spain; Hans Reichenbach				
	Na e39	+	-	-	+	This study	Cefalù, Sicily; Hans Reichenbach				
	Na e434	+	-	-	+	This study	Jerusalem, Israel; Hans Reichenbach				
	Na e465	+	-	-	+	This study	Hans Reichenbach				
	Na e585	+	-	-	+	This study	Pissouri, Cyprus; Hans Reichenbach				
	Na e619	+	-	-	+	This study	Mallorca, Spain; Hans Reichenbach				
	Na e642	+	_	_	+	This study	Bornholm, Denmark; Hans Reichenbach				

TABLE	1.	Survey	of	the	myxo	bacteri	a for	retron	elements
-------	----	--------	----	-----	------	---------	-------	--------	----------

^a Screened by ethidium bromide staining.

^b Probed by using the *msd* region only for the probe.

^c ND, not determined.

detected by the RT extension method (19, 21, 29). Labeled products were divided in half, with RNase A treatment of one set of samples, and separated on 4%acrylamide–8 M urea gels. Gels were dried and exposed to X-ray film at -80° C.

Detection of homologous retron elements. Chromosomal DNA was isolated from each myxobacterial strain and digested with restriction endonuclease PstI as described previously (29). Chromosomal DNA digests were separated on 0.7% agarose gels and transferred to nitrocellulose filters. Southern hybridizations were carried out at high stringency (50% formamide, 5× SSPE [1× SSPE is 0.18 M NaCl, 10 mM NaH₂PO₄, and 1 mM EDTA, pH 7.7], 5× Denhardt's solution, 0.3% sodium dodecyl sulfate) at 42°C. The Mx162 probe was prepared as a 1.3-kb XhoI-RsaI subfragment from plasmid pmsSB7 (29). This fragment contains part of the msd gene, which codes for the DNA portion of msDNA, and about two-thirds of the N-terminal region of the RT ORF. The Mx65 probe was prepared as a 1.2-kb EcoRI-BamHI fragment from plasmid pBPV-1 (5, 19). msDNA from *Nannocystis exedens* Na e434 was prepared for use as a probe as follows. msDNA was specifically labeled with ³²P by the RT extension method, RNase A treated, and then isolated by electrophoresis on a 4% acrylamide-8 M urea gel. The msDNA was cut out of the gel and recovered by elution. Gelpurified restriction fragments were labeled for probes with $\left[\alpha^{-32}P\right]dCTP$ by using the random hexamer labeling technique (7) and purified by using Sephadex G-50 columns.

Cloning of a retron homologous to Mx162. *Melittangium lichenicola* chromosomal DNA was digested with *PstI* and separated by electrophoresis on a 0.7% agarose gel. *PstI* fragments in the range of 9 to 11 kb were recovered, purified by electroelution, and ligated into the dephosphorylated *PstI* site of pBR322. Tetracycline-resistant, ampicillin-sensitive colonies were screened by colony hybridization (25), using a 0.75-kb *Eco*RI-*Hin*dIII fragment of retron Mx162 as a probe. This probe hybridized to plasmid pNaeF4.15, containing a 10.8-kb fragment. From this positive clone, a 3.8-kb *Bam*HI fragment was subcloned into a dephosphorylated *Bam*HI site in pUC19 (designated pNaeBB4.0.A4). Overlapping fragments of pNaeBB4.0.A4 were subcloned into pUC19 for sequencing.

DNA sequencing. DNA sequencing of pNaeBB4.0.A4 subclones was performed by the dideoxy-chain termination method of Sanger et al. (30), using a Δ Taq version 2.0 sequencing kit from United States Biochemical. Overlapping clones or synthetic primers were used to ensure that sequencing runs extended through all cloning sites. The DNA sequence was determined for both strands. Single-stranded templates were prepared by alkali denaturation as described in the protocol for the sequencing kit (United States Biochemical). The nucleotide sequence of the DNA portion of the msDNA was determined directly by the method of Maxam and Gilbert (26). msDNA was isolated from freshly grown cells, treated with RNase A, gel purified on a 5% acrylamide gel, and labeled at the 5' end with [γ -³²P]ATP, using T4 polynucleotide kinase (New England Biolabs). Electrophoresis of sequencing reactions was performed by using 6% (wt/vol) acrylamide–8 M urea gels. Sequences were analyzed by using DNA Inspector (version IIe), the GenBank sequence database (version 7.3), and the package of the Genetics Computer Group of the University of Wisconsin (2).

Silent substitutions and codon bias. The degree of nucleotide sequence variation between two homologous genes is the result of the accumulation of nucleotide substitutions over a period of time since the two genes diverged from a common ancestor. For most genes, changes tend to occur at a uniform frequency or in a clock-like fashion (27, 37). One measure of gene divergence is the number of changes per unit time or the rate of change for that gene. To calculate this rate, two types of mutations, synonymous and nonsynonymous substitutions, are considered within protein coding regions. A nonsynonymous substitution is one that changes a codon and subsequently the encoded amino acid. Synonymous substitutions, or silent substitutions, change a particular codon but do not change the encoded amino acid; this is the result of degeneracy of the genetic code. While it has been proposed that mutations in bacteria occur at a regular, clocklike rate, mutations do not accumulate at the same rate for the two types of sites (22). Nonsynonymous substitutions may have detrimental effects and thus be eliminated from the observed population. However, since synonymous substitutions do not alter the amino acid sequence, these substitutions are presumed to be selectively neutral, and they accumulate in the population over time at relatively constant rates. The percentage of synonymous substitutions was based on the method of Li et al. (22). The rate is determined by dividing percent substitution by the evolutionary distance (time) between the two species, on the basis of 16S rRNA comparison.

Nucleotide sequence accession number. The sequence for retron ML162 was deposited into the GenBank database under accession number L36722.

RESULTS

Retron elements in the myxobacteria. On the basis of 16S rRNA sequence comparison, the myxobacteria, that is, the taxonomic order Myxococcales, fall into three major subgroups: Myxococcus, Chondromyces, and Nannocystis (34). In this study, 28 strains of myxobacteria, with representatives from each of the three major subgroups, were screened for the presence of retron elements by looking for the production of msDNA. The RT extension method was used to detect msDNA. Briefly, this technique uses Moloney murine leukemia virus RT and $\left[\alpha^{-32}P\right]dCTP$ to specifically radiolabel msDNA by extending the DNA strand, using the RNA strand as a template (19, 21). Extension continues to the branched guanine residue where reverse transcription stops, leaving an RNA arm at the 5' end. Subsequent treatment of the extended and labeled product with RNase A removes the 5' RNA arm, producing a faster-migrating species compared with the untreated sample. Figure 1 presents typical results from an RT extention assay. Strains So ce5, So ce7, So ce11, and Cm a16 (Table 1) all produced distinctly labeled bands in the size range of 145 to 245 nucleotides (Fig. 1, lanes 1 to 8). The presence of msDNA indicates the presence of the corresponding chromosomally encoded gene, the retron. The shift in mobility when the extended products were treated with RNase A (Fig. 1, lanes 2, 4, 6, and 8) is characteristic of msDNA, and the size difference reflects the length of the 5' RNA arm. The double bands present in some strains represent either the presence of more than one msDNA or heterogeneity of extension by the RT. No labeled bands were present for strain Cb fe18 (Fig. 1, lanes 9 and 10), indicating that retron elements are absent from this Cystobacter strain. We screened several species of myxobacteria which have not been previously studied for the presence of retron elements, including 3 Chondromyces, 4 Cystobacter, 3 Sorangium, 1 Melittangium lichenicola, 1 Myxococcus virescens, 1 Myxococcus fulvus, 1 Myxococcus macrosporus, 1 Myxococcus stipitatus, 1 Angiococcus disciformis, 1 Archangium gephyra, 1 Corallococcus (Myxococcus) coralloides, and 10 N. exedens strains. The positive and negative strains as determined by the RT extension method are summarized in Table 1. All of the strains tested positive for the production of msDNA except for Cystobacter ferrugineus Cb fe18. It appears that strains Cm p5, Cm a15, and Cm a16, contain two different msDNAs, while So ce5, So ce7, and So ce11 contain only a single msDNA (data not shown). In addition, we also screened 10 N. exedens strains by using the RT extension assay. All 10 strains tested positive (data not shown), with the msDNAs ranging in size from 120 to 180 bases. On the basis of the different sizes of msDNAs in the Nannocystis strains, it is apparent that some strains may also produce more than one species of msDNA. With the exception of one Cystobacter strain, retrons are ubiquitous among the myxobacteria reported here (Table 1).

Homology to known retrons. Identification of retron elements similar to the two elements Mx65 and Mx162, previously



FIG. 1. Detection of ³²P-labeled msDNA by the RT extension assay. The molecular weight standard (S) is radiolabeled pBR322 cut with *MspI*, and sizes are listed in base pairs. Strains are listed by genus and strain designation. Half of each labeling reaction was treated with RNase A, designated by a (+), prior to electrophoresis. Treated and untreated samples were electrophoresed in adjacent lanes. Note that RNase A-treated msDNAs (radiolabeled bands) migrate faster (lanes 2, 4, 6, and 8) than the untreated samples (lanes 1, 3, 5, and 7). Strains So ce17, So ce5, and Cm a16 are positive for the presence of msDNA, while Cb fe18 produced no labeled band and thus is considered to be negative for msDNA production.

identified in *M. xanthus*, was accomplished by DNA hybridization. Radiolabeled fragments derived from plasmid clones of these retrons (4–6) were hybridized to *PstI*-digested chromosomal DNA from each myxobacterial strain. In addition, msDNA isolated from *N. exedens* Na e434 was radiolabeled and also used as a hybridization probe.

All of the strains in the Myxococcus subgroup cross-hybridized with the Mx162 probe, indicating that these species share retrons that are similar in primary nucleotide sequence to Mx162 (Fig. 2, lanes 1 to 8). In contrast, the Mx65-based probe cross-hybridized with only three species of the genus Myxococcus (data not shown) (Table 1): M. xanthus, M. virescens, and M. macrosporus. Additionally, chromosomal DNA from strains in the subgroups Chondromyces and Nannocystis did not hybridize with either the Mx162 or Mx65 probe, indicating that these retrons are restricted to the Myxococcus subgroup (Table 1). Figure 3 presents a phylogenetic tree of the myxobacteria showing the clustered distribution of retron elements similar to Mx162 among the seven genera that make up the Myxococcus subgroup. Purified msDNA, from N. exedens Na e434, was also used as a hybridization probe. While the Myxococcus, Chondromyces, and Sorangium strains did not hybridize to this probe, all 10 Nannocystis strains surveyed hybridized to the Na e434 msDNA probe (Table 1). This finding suggests that members of the Nannocystis subgroup likely contain the same, or a very similar, retron.





FIG. 2. Southern hybridization of nine members of the *Myxococcus* subgroup with a labeled Mx162 probe. Chromosomes were digested with *Pst1*, except for *C. coralloides*, which was digested with *Bam*HI. Digested chromosomes were electrophoresed on a 0.7% agarose gel, transferred to nitrocellulose, and hybridized with a ³²P-labeled internal fragment of the RT gene from Mx162. Molecular weight standards are given in kilobase pairs. All strains tested reveal a single band which hybridized with the Mx162 probe, except *M. macrospons*, which shows two distinct bands. Weak but clearly visible hybridization signals appear for *M. lichenicola*, *A. gephyra*, and *A. disciformis* (lanes 6, 8, and 9).

An msDNA hybridization probe was also prepared from a member of the *Chondromyces* subgroup (*Sporangium* strain So ce11). However, because of difficulty in preparing this probe, only a small number of myxobacterial strains were screened with this probe (not shown). Nevertheless, chromosomal DNA from strains So ce7, So ce11, and Cm p5 (*Chondromyces*) hybridized strongly to the msDNA probe from So ce11. DNA from *Nannocystis* strain Na e1 and *M. xanthus* DZF1 gave a weak hybridization signal, and DNA from *Nannocystis* strain Na e434 and *Melittangium lichenicola* gave no hybridization signal (not shown).

Cloning and sequencing of a *Melittangium* **retron.** Because the chromosome digest from *Melittangium lichenicola* gave a particularly weak hybridization signal with the Mx162 probe (Fig. 2, lane 6), we were interested in cloning and sequencing the retron from *Melittangium lichenicola* to compare sequence diversity with similar retrons found in two other genera within the same myxobacterial subgroup. The objective was to be able to make some determination about the evolutionary history of the retron, i.e., to determine whether it is an ancient element or was recently introduced into these three species.

Maxam-and-Gilbert chemical sequencing of the DNA portion of the msDNA from *Melittangium lichenicola* (data not shown) revealed the DNA strand to have 162 bases, and it has been designated ML162. ML162 msDNA showed significant homology to Mx162 in the DNA region, with only 17 base differences, 10 of which are found in the central stem region which forms a stable hairpin structure due to complementary secondary folding of the DNA strand. The majority of these base differences are complementary mutations on opposite sides of the stem that help to maintain the stem-loop structure. In addition to the nucleotide differences in the DNA, 15 of the



10%

FIG. 3. Phylogenetic distribution of msDNA and Mx162 retrons in the myxobacteria. The phylogenetic tree of the myxobacteria, as determined by 16S rRNA sequencing, is abbreviated from that of Shimkets and Woese (34). Strains are listed as positive (+) or negative (-) for msDNA production and for Southern hybridization with a ³²P-labeled internal fragment of the RT gene from Mx162. ND indicates that the strain was not tested. It should be noted that *Desulfovibrio*, *Desulfosarcinia*, and *Bdellovibrio* strains are not myxobacteria and serve as out groups.

76 predicted nucleotides that make up the RNA strand are different.

On the basis of hybridization of Mx162 DNA to *Melittangium lichenicola* (Fig. 2, lane 6), *PstI* chromosomal fragments of *Melittangium lichenicola* in the range of 9 to 11 kb were cloned into the *PstI* site of pBR322. The recombinant plasmids were transformed into *E. coli* and screened by colony hybridization. One positive clone (pNaeF4.15) was picked, and confirmation of the retron clone by Southern hybridization revealed a 10.8-kb *PstI* fragment that hybridized with the Mx162 probe. This primary clone was subsequently digested with *Bam*HI, and a 3.8-kb fragment was subcloned into the *Bam*HI site of pUC9 (pNaeBB4.0.A4). This clone was then further subcloned for sequencing.

Using the Maxam-and-Gilbert method-generated sequence of the msDNA (data not shown), we were able to locate the msd gene from the chromosomal clone and subsequently identify the predicted *msr* gene encoding the RNA strand of the msDNA. Two 17-bp inverted repeats were found; one (a2) is located just upstream of the predicted branched guanine residue which forms the 2',5' RNA-DNA linkage, and the other (a1) is just downstream of the msd gene (Fig. 4). These inverted repeats have been demonstrated to be required for the proper folding of the primary transcript to initiate the selfpriming reaction that begins msDNA synthesis (12, 18, 32). Analysis of the sequence in the vicinity of the msr and msd genes revealed five ORFs. The fifth ORF, beginning at position +1897, showed significant nucleotide identity (>70%) to the RTs found in the myxobacteria retrons Mx162 and Sa163 (13, 16). The predicted protein from ORF5 shows the seven conserved amino acid sequence domains, diagnostic for RT, including the highly conserved YADD sequence of domain 5 (40). Indeed, this ORF shows 79% nucleotide identity to the RT gene in retron Mx162. The RT ORF did not show significant nucleotide identity to the RT from Mx65. This is expected since Mx65 shows only 35% nucleotide identity to Mx162 (5, 15). The other four ORFs, ranging in size from 152 to 441 amino acids, showed no significant homology to any previously reported genes when searched through the Gen-Bank database. Comparison of the DNA sequence from a2 to the 3' end of the RT ORF with the Mx162 sequence revealed 77% nucleotide identity overall between the two retrons, with identity being highest in the msr and msd regions (83 to 90%) and lowest in the spacer region which is located between msd and the RT ORF (39.7%).

The G+C content of the myxobacteria is 67 to 72% (28). This high G+C content is reflected in the preferential usage of codons that are GC rich. The RT ORF of the retron found in Melittangium lichenicola, like most known genes from M. xanthus, shows a heavy bias toward codons that use G and C over A and T. For example, there are six possible codons for the amino acid leucine (CTA, CTC, CTG, CTT, TTA, and TTG), yet of the 48 leucine residues in the RT ORF of ML162, 98% use either CTC or CTG. There is a similarly strong preference observed among 20 known genes of *M. xanthus*, in which 88% of all leucine codons are either CTC or CTG (33). Highly statistically significant, nonrandom codon usage within the RT gene is likewise reflected in the large chi-square value for leucine codons of ML162 (not shown). This bias toward GCrich codons is pervasive throughout the RT ORF, and it is similar to the codon preference observed among other known genes of *M. xanthus* and consistent with an organism with a high-G+C genome. Such a GC codon bias has also been noted previously for the RT ORF of retron Mx162 (16).

Retron junctions. At present, the precise ends of retrons in the myxobacteria are unknown. Therefore, the ends of a retron

element can be defined only as unique DNA sequences which are associated with the genes encoding msDNA and RT. If two similar or isogenic chromosomes exist, in which one of the chromosomes lacks the retron element, then comparison of the corresponding region between the two chromosomes should reveal the junctions of the retron element. However, for M. xanthus, there are no known strains in which the Mx162 retron element is absent. Alternatively, a different strategy can be used to identify the ends of this genetic element by comparison with the ML162 retron from Melittangium lichenicola. Since the two retrons from Melittangium lichenicola and M. xanthus are very similar in DNA sequence, and if the chromosomal regions in which these elements have inserted are significantly different in the Melittangium and Myxococcus DNA sequences, then it could be possible to approximate where the 5' and 3' ends of the element lie. Figure 5 presents an alignment of flanking DNA sequence from M. xanthus and M. lichenicola around the 3' end of the RT ORFs of their respective retrons. In this 3' region, sequence identity rapidly degenerates from >70% to almost zero and extends for several hundred base pairs of sequence downstream from the RT ORF. This 3' junction appears to be very near the stop codon of the RT ORF (Fig. 5). Interestingly, at the 5' end of the retron, significant nucleotide identity is maintained over 300 bp upstream of the msr gene. This indicates that the retron could be inserted in a gene common to M. xanthus and Melittangium lichenicola or probably in some homologous location on each chromosome. Alternatively, the retron could have been transmitted vertically from a common ancestor of M. xanthus and Melittangium lichenicola, in which case the junction sequences would not be distinguishable from host sequences if the two species diverged at similar rates. A similar situation is observed with the DNA sequence flanking the retron Sa163 from S. aurantiaca (13), in which case significant nucleotide identity is maintained upstream of the msr gene but not downstream of the RT gene.

Sequence divergence. By calculating the rate of nucleotide change at synonymous sites for the RT gene encoded by the Mx162-type retrons, the divergence of this gene can be compared with that of other bacterial genes. Such comparisons may provide a hint about the origin and age of this RT gene found in the myxobacterial genome.

The average rate of silent substitution found for genes common to Salmonella species and E. coli is 0.335 substitutions per million years (31) (Table 2). The fastest and slowest rates of substitution between E. coli and Salmonella species are 0.63 substitutions per million years and 0.08 substitutions per million years, respectively. The Melittangium and Myxococcus myxobacteria are estimated to have diverged from a common ancestor about 90 million years ago. This estimate is based on comparison of their 16S rRNA sequences (34) and a universal substitution rate calculated for bacterial genomes (27). The percent divergences of the RT genes from retrons Mx162 and ML162 were compared by the method of Li et al. (22) (Table 2). ML162 and Mx162 show 46% divergence at silent sites, which translates into a rate of 0.26 silent substitutions per million years. This is close to the average rate of change (0.335)for genes reported in Salmonella species and E. coli, which presumably diverged about 140 million years ago (31). In fact, it can be argued that the degree of sequence variation observed among the Mx162 RT genes is more significant than what is reflected in the calculated rate of silent substitutions. This is because of the high G+C content of the myxobacterial genome, which results in strong codon bias in protein-coding regions. This in turn may reduce the rate at which substitutions at silent sites accumulate, by limiting the number of alternative codons that can be used. Thus, the amount of sequence vari-

.1	BamHI																		-					2000	
+1	GGATCCC	GATC	AGGTC	CTCGC	''I'GG	JAGGT	C'I'IC	TCCO	GCGA	CGAC	AGCG	CCTACC	CGCG	CCT (GCCI	"I'CG4	ACAAC	AGCC OF	RF1	TCGA	ICAAC	CGCC	AGG(5
+101	CACGGCI	CCACO	TTCC	CTGGA	. CGT	GCCG	CAGG.	ACCT	GCCO	G CCC	GCCT	GGACCO	CCIG	CGC (GCCTA	CGCC	CCCGC	ftg al	IG CT	GGAC	GCCA	ACGTO	GCCC	CCA	2
+201	CGCGAAG	KCCGG	CGCCC	CGCCA	. CGG	AGCC	CCGC	GCCC	CGCG	C ACC	GTGG	AGGTGO	CGCA	CGC (GTGGC	TGCC	GAGGO	CTTCC	CTCC	AGGI	'GCAC	TCCC	GCCGC	CGACO	3
+301	CTGCCGG	CCAC	CACCI	rgege	CAT	CGCG	CCCA	TCGA	CCTG	t aca	ACCT	GCTCT"	CGCG	CTG (CGCAC	cca	CCGCI	rcgaz	AGAA	GGCC	cccc	GCGC	GCTC	GCGC1	r
+401	TCGAACI	GGTC	20000	GCGCC	CCG	CCGC	GCCIV	GGTG	CTGG	A GCC	CTGG	GAGCA	GTGC'	TGG 2	AGTGC	CACC	GCGC	CGTC PVU	STAC	ACGO	GAGO	AGCO	cccc	3CGG1	C
+501	GGTGCGC	ACCT	rcgg	CCGTC	AGC	GCCT	CGCO	GCGC	TCGC	G CGC	CTCT	IGCCC	ACGCO	GAA (GAGCO	FGCA	ACGTC	GCAGO	CTGT	TGGG	TCCC	iacco	TGC	CGGTC	3
+601	TTCTGGG	TCAT	CGACO	CTGGG	CGC	GGCC.	ACGC'	TGAC	GCTG	G GGC	CTCAC	CGGCT	GACGO	GAG I	AGCGG	CTG	GTCC2	AGCGC	CCGC	CGCC		ACGI	GCTC	GATG	2
+701	CGCGCGA	CGTG	CCGGI	AAGGC	CTC	GCGG	AACA	GCTG	CGCA	a coo	CCTG	CGCAA	GACG	GCC (CCTC	GCCI	rtcg/	AGTO	SCTG	GCCA	AGGA	CGCC	XGGC'I	rcaco	2
+801	GAAGGAC	CACG	rgcga	CGCCG	CGC	TCCA	GCTG	GAGT	GCCT	G CG#	AGGCC	GCGTC	TCTT	CGA (GTGG	CGCC	CCCC	CACCI	TACC	GCCC	GCGI	GAGO	TGAT	rgeéd	2
+901	ACCCCG	TGGA	GAGO	GCCGC	GCT	GCGC	TACG	GCAA	CGAG	c ca	GAGGC	CCGCG	CCAC	cgc (CTGCI	GGGG	CGACO	GCGC	жсс	GGGC	TCGC	GCGI	AGTO	GAAGO	2
+1001	TCACGCA	GGTG	CACGA	ACCTG	GTC	.GGCG.	AAGG	CACG	CGCA	T CCA	AGGGA	GAGGT	GTGG	ACC (GCGAC	iscoc	GTGCC	GCAGO	CTTC	TTCC	CCAC	CTTC	St <u>y</u> I, CACCA	/ <u>Nco</u> . ATGG/	I A
+1101	CCTGGAA	.99000	CGTC	GAAGG	ACG	CCAG	CTGC	GGCTV	GCCC	g cac	CTTCC	GCCGC!	CCGG	CAT (secce	AAG	sccco	CTGCC	GAGC	ACAT	GCTC	GCGC	TGCC	3CCT(G
+1201	GCGTACO	CCCC	cccc	CGCGC	CGA	GGAA	GAGG	CGCT	CCGG	C AG4	ACGCC	CGAGG	CCGC/	<u>Alu</u> AAG (I CTCAT	CCGC	CGCGC	GAGAC	CGCG	CGCC	TACO	STGCC	CCG	GAC	2
+1301	CGGCCAC	GGGCC	CTGGI	AGCAG	GTG	TATC	GCTG	GACG	GCAA	G GTC	3GTGG	CCCTCI	CGTG	GGG (caca	GCCI	ragge	GACI	rcgc	GTCA	ACCAC	CGCC	TCTC	GTT	2
+1401	GACACGO	ATAC	CGAAC	GCCCG	GAC	CGCG	TATT	TCAG	CCGT	с тас	GAGAA	ACTGA	CGCIN	GAC	GCTA	CATO	CGACO	GCGGC	CTC	GACI	CIG	TGTA		ACTC	ł
																a	2				Ľ	sı	->	•	
+1501	CTCAACO	GCCCC	BCCC	GCCGA	CGG	GCCC	GCCG	GAAG	GATO	G TGC	GCGCC	GACG/	AAAT	CGA :	ATAGO	GAGI	IGGTO		GAGA	GGTC	TGGA	CCGC	TAG		
																	~~~~								
		_ <u>_Xh</u>	<u>0I</u>								0	RF2													_
+1598	CTCAACO	<u>Xh</u> CCTCC	<u>oI</u> BAGCO	TAGG	AAC	CCCC	CTGC GACG	CCCC	TTCT AAGA	G GTI C CAZ	0 Ig <b>aa</b> a Actitt	RF2 IG CTC AC GAG	GAC	ACT (	TCCC	CAA	GTAC	CCTO	CAAG	TTGO	CTCI	CTCC	CTCC	CCGA	3
+1598 +1696	CTCAACO	<u>Xh</u> GGAG	<u>ol</u> BAGCO CTCGO	STAGG CATCC	AAC TTG	GGCGG	CTGC GACG CAAC	GCCG CGGC. GACG	TTCT AAGA CAAC	G GTT C CAZ	O IGAAA ACTITT	RF2 IG CTC AC GAC	GAC	ACT ( TGA (	CTCCC BAGGC BAGGC	CAAC GTTC	GGTAC	GGAC	CAAG	TTGC AACC	CTCI CGAGZ	CTCC	GAG	CCGAC BGCTK	
+1598 +1696	CTCAACO GAGTTGC CACTACO GTGATGC	<u>Xh</u> GGAG GTCG	<u>oI</u> BAGCO CTCGO BGGCO CCCGO	STAGG CATCO SGGAA	AAC TTG GCG CGC	GGCGG GAAC	CTGC GACG CAAC GTTG	GCCG CGGC GACG CTGC	TTCT AAGA CAAC GTTG	G GTT C CAZ C GCC G CGC	O FGAAA ACTITT COTTT COTTT CAAA	RF2 IG CTY AC GAO ICCCAO AGGGTO	CTG	ACT ( TGA ( ACC ( TGG (	CTCCO BAGGO BTAGT CATCA	CAAC CTTC CCTC	GGGGI CCATC CGGGI GCCCT		GTTC CAAG GAGA	TTGO AACO GCCO CGGO	GAGI GAGI GTGI	CTCC GAGC GGCT	CTCC GAGC ACCC	CCGAC BGCTC BTGCC CACGC	
+1598 +1696 +1796	CTCAACG GAGTTGC CACTACG GTGATGC CCAGGTC GGTCCAC	<u>Xh</u> GGAGG GGAGG CAGCC AGCTC		STAGG CATCO SGGAA CCCTT SGTGC CCACG	AAC TTG GCG CGC CGC	CCCCC CCCCC CCTCC CCTCC CCTCC	CTGC GACG CAAC GTTG GCCT CGGA	GCCG CGGC GACG CTGC CTGC	TTCT AAGA CAAC GTTG CGAC GCTG	G GTT C CAZ C GCC G CGC C GCT G CGZ	O IGAAA ACTITI OGTITI OGAAA ICGGA	RF2 TG CTV AC GAO FCCCAO AGGGTV T TCG	GAC	ACT ( TGA ( ACC ( TGG ( TCCG	CTCCC BAGGC BTAGT CATCA	CAAC CTTC CCTC CCCC	GGTAC CCATK CGGGJ GGCCT	CCTC CGGAC AGGGC FCCCC	GAGA	TTGC AACC GCCC CGGC	GETCI GAGI GTGI CACI	CTCC GAGC AGCCI CCCG TCCC	CTCC GAGC ACCC ATGGC	CCGAC GGCTC GGCTC GTGCC CACGC	G G
+1598 +1696 +1796	CTCAACG GAGTTGC CACTACG GTGATGC GGTCCAC	<u>Xh</u> CCTCC GGAGC GTCGC CAGCTC AGCTC TCGAC		STAGG CATCO SGGAA CCCTT SGTGC CCACG	AAC TTG GCG CGC CGC	GGCGG GGAAC CTTG CCTG	CTGC GACG CAAC GTTG GCCT CGGA	GCCG CGGC. GACG CTGC CTGC GGGA	TTCT AAGA CAAC GTTG CGAC GCTG	G GTI C CAA C GCC G CCC C GCI G CCI	O IGAAA ACTIT SCAAA ICGGA	RF2 IG CTX AC GAC ICCCAC AGGGTX I TCG	GAC	ACT ( TGA ( TGG ( TGG (	CTCCC BAGGC STAGI CATCA	CAAC CTTC CCAC	GGGGI CCATC	CCTC CGGAC AGGGC PCCCC	BAGA BAGA	TTGC AACC CCCC CCCC	CTCI CAGZ CACI	AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI AGGCI	CTCC GAGC ACCC ATGGC	CCGAC GGCTY GGCTY GTGCC CACGC	2 2 3 AG
+1598 +1696 +1796 +1897	CTCAACC GAGTTGC GAGTTGC GTGATGC GGTCCAC GGTCCAC C QGTCCAC C QGTCCAC	<u>Xh</u> GCCTCC GGAGC CAGCCC CAGCTC CAGCTC TCGAC TCGAC TCGAC		STAGG CATCO SGGAA CCTT CTG	AAC TTG CCC CCC CCC	CCCCC CCCCCC CCTCC CCTCC CCTCC CCCCCCCC	CTGO GACG CAAC GTTG GCCT CGGA al	GCCG CGGC CTGC CTGC CGGA	TTCT AAGA GTTG CGAC GCTG		O TGAAA ACTITT SCAAA ACCGGA ACCCT	RF2 TG CTX AC GAA TCCCAA ACGGTX T TCG A CG CCC	CTG	ACT ( TGA ( TGG ( TCCG'	GCT	CCT	GAA	GTC	ACG	TTGC AACC CGGC CGGC CGGC	GCT	CCCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GCTCC GAGC ACCC ATGGC AGTA/	CCGAC GGCTK GGCTK CACGC ACTC/	AG AAC
+1598 +1696 +1796 +1897	CTCAACG GAGTTGC GTGATGC GTGATGC GGTCCAC GGTCCAC CCAGGTC GGTCCAC CCAGGTC GGTCCAC CCAGGTC GGTCCAC CCAGGTC GGTCCAC CCAGGTC GTGATGC GTGATGC GTGATGC	<u>Xh</u> GCCTCC GGAGC CAGCTC CAGCTC TCGAC TCGAC TCGAC A CCC A		STAGG CATCO SGGAA CCCTT GGTGC CCACG CTG L	AAC TTG CGC CGC CGC CGC CGC CGC CGC CGC CG	CGCCGC CCGCC CCTCG CCTCG CCTCG CCCTG CCCTG CCCTG CCCTG CCCTG CCCTG CCCTG CCCTG CCCTG CCCTG CCCCGC	CTGC GACG GTTG GTTG GCCT CGGA al TTC F	GCCC GACG CTGC CTGC GGGA L GTC V	TTCT AAGA CAAC GTTG CGAC GCTG CCC P	G GTI C CAF C GCC G CGC G CGC G CGC GCC C A	O TGAAA ACTITT SCAAA TCGGA AGCCT BCC C A	RF2 IG CTV AC GAC ICCCAC AGGGTV I TCG' A CG CCC P P P	CTG CCCG CCCCG CCTCC CCTCC	ACT ( TGA ( ACC ( TGG ( TGG ( TCCG' CGT R	GCT A	CCT CCT	GAA	GTC V	ACG T	TTGC AACC CGGC CGGC CGGC CGGC CGGC CGGC	GCTCI CGAGZ CACI CCCCC CCCC CCT A	CTCC CAACC CCCCZ ACCCCZ ACCCC A	GCT ACCO ACCO ACCO ACCO ACCT A	CCGAC GGCTC GGCTC CACCC P CCCC P	AG AAC N
+1598 +1696 +1796 +1897 +1978	CTCAACC GAGTTGC GTGATGC GTGATGC GGTCCAC GGTCCAC ACC M T ACC GTC T V	<u>Xh</u> CCTCC CGGAGC CAGCTC CAGCTC A CCAGCTC A CCAGCTC A CCCC A CCCC A	oI BAGCO CTCGC CCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC GCCCGC AAG K AAG K	STAGG CATCO SGGAA CCCTT CCACG CTG L CCGC R	AAC TTG CGC CGC CGC CGC CGC CGC CGC CGC CG	CCTCC CCTCC CCTCC CCTCC CCTCC CCCTCC CCCCC S CCCCC CCCCC CCCCC CCCCC CCCCCC	CTGC GACG GTTG GCCT CGGA TTC F GCG A	GCCG CGGC CTGC CTGC GGGA GGCA L GTC V AAG K	TTCT AAGA GTTG GTTG GCTG CCC P GCC A	G GTI C CA2 C GCC G CCC G CCC A GCC (C A	COTT COTT COTT COCC C C C C C C C C C C	RF2 TG CTX AC GAG T TCCCA AGGGTX T TCC A CG CCC P P AC GCC D A	CTCC CTCC CTCC CAG Q CTCC L	ACT ( TGA ( TGG ( TCCG' CGT R CTC L	GCT ACG T	CCT CCCT P CCCC R	GAA E TGG W	GTC V AGGGC IGAGC GTC V AAG K	ACG GCC ACC ACC ACC ACC ACC A	TTGC AACC CGGC CGGC CGGC CCC P ATC I	GCTCI CGAGZ CCACI CCACI CCACI CCACI A CCC T	CCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC	GCT CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCGAC GCTC CACGC ACTC/ CCC P GCC G	AG AAC N GGC G
+1598 +1696 +1796 +1897 +1978 +2059	CTCAACC GAGTTGC GGTGATGC GTGATGC GGTCACC GGTCCAC CACGGTC ACC GTC ACC GTC ACC GAC	ACCTOC CGGAGC CAGCC CAGCT CAGCT TCGAC A CCC A CCC A CCC A CCC A CCC A CCC A CCC A CCC CAGCT C CAGCT C CAGCT C CAGCT C CAGCT C CAGCT C CAGCA C CAGCA C CAGCA C CAGCA C CAGCA C CAGCA C CAGCA C CAGCA C CAGCA C CAGCA C CAGCA C CAGCA C CAGCA C C CAGCA C C CAGCA C C C C	<u>oI</u> <u>33AGCC</u> <u>3GCCC</u> <u>3GCC</u> <u>3GCC</u> <u>3GCC</u> <u>4</u> AAG K AAG K RF3 <u>TGG</u>	STAGG CATCC SGGAA CCCTT GGTG CTG L CGC R GTG	AAC COC COC COC COC COC CAC	CCCCCC CCTTG CCTTG CCTG CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCC	CTGC GACG GTTG GTTG CGGA TTC F GCG A CAG	GCCG GGCC CTGC CTGC GGGA L GTC V AAG K CTC	TTCT AAGA GTTG CGAC GCTG CCC P GCC A GTC	GCC 2	COTT COTT COTT COTT COAAA COCT COCC CA CCC CA CAC G H CAAA G	RF2 TG CTX AC GAX TCCCCAC ACCOGTX T TCCC T TCCC A CC CCC P P P AC CCC D A GC CCC	CTC CTC CTCC CTCC CTCC CTCC CTC L CTC	ACT ( TGA ( TGG ( TGG ( TGC ( TCCG ⁷ R CTC L GCG	CTCCC BAGGC STAGT CATCA CCATCA ACG T GAG	CAAC COTTO COCO COCO P COCO R COC R GAA	GOTAC CCATC CCCATC CCCCATC GCCCATC GCCCATC GCCCATC GCCCATC GCCCATC GCCCATC GCCCATC GCCCATC GCCCATC GCCCATC GCCATC GCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCATC CCCCATC CCCCATC CCCCATC CCCCATC CCCC	GCCTC CGGAC AGCOCC ICAGC IGAGC GTC V AAG K CII GAC	CAAG BAGA TTCT CCAGC ACG T GCC A TTC	TTGC	GCTCT CGAGA GCTGA GCCCCC GCCCCC A A CCC T TCG	AGAG AGAG TCCG2 TCCG2 TCCG2 TCCG2 A GAG E CTG	CTCC GAGC TGGG GCT A GCC A GCC A AAG	CCCCA CCCCC P CCCCC CCCCC CCCCC CCCCC CCCCCCC	AG AAC N GGC G AAG
+1598 +1696 +1796 +1897 +1978 +2059	CTCAACC GAGTTGC GTGATGC GTGATGC GGTCCAC GGTCCAC ATG ACC M T ACC GTC T V ACC GAC T D	Zin GCCCC GGAGC CAGCC A A CCCCA A C C C C C	oL BAGCC CTCGC CTCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCGC CCCCGC CCCGC CCCCGC CCCGC CCCGC CCCGC CCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCGC CCCCCGC CCCCCGC CCCCCC CCCCCCC CCCCCC CCCCCC CCCCCC	CTG CCCC CCCC CCCCC CCCC CCCCC CCCCC CCCCC CCCC	AAC TTG GCG CCTT GAG E GAG E GAG E CAC H	COGCO CCTCG CCTCG CCTCG CCCTCG CCCTCG CCCCCG CCCCCG A CCCCCCG A CCCCCG A	CTGC GACG GTTG GCCT CGGA TTC F GCG A CAG Q	GCCG GGC GGC CTGC CCCT GGGA V AAG K CTC L	TTCT AAGA CGAC GTTG CGAC GCC A GCC A GCC A GTC V	G GTT C CAA C GCC G GCC G CGC A GCC C A GCC C A GCC C A	COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTT COTT COTT COTT COTT COTT COTT COTT COTT COTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTT COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTTC COTT	RF2 TG CTX AC GAC TCCCAC AGGGTC T TCG ² A CG CCC P P AC GCC D A AC GCC G A GC GCC G A	CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	ACT ( TGA ( TGG ( TGG ( TGCC ⁴ R CTCC ⁴ L GCG A	GAG GAG GAG GCT A ACG CA CA CA CA CA CA CA CA CA CA CA CA CA	CAAC COTIC CCCA CCCT P CCCC R GAA E	GAA E GAA E TGG W Hin GTC V	GCCTC CGGAC AGGGC TCCCCC UGAGC V AAG K <u>cIII</u> GAC D	CAGE CAGE CCAGE ACG T GCC A TTC F	TTGC AACC GGGC CAGCC CAGCC P ATC I TCC S	GCTCI CACI CACI CACI CCCCC CCCC CCCC CCCC	GCC AGAG GCC A GCC A GAG E CTG L	CTCC GAGC TGGC KGTAA GCC A GCC A AAG K	CCCCA CCCC P GGC GAG E	AG AAC N GGC G AAG K
+1598 +1696 +1796 +1897 +1978 +2059 +2140	CTC-AACG GAGT-TGC GTGATGC GGTGATGC GGTC-CAC GGTC-CAC C CACT ACC T ACC T C ACC C C C C C C	Zin GCCCC GGCAGCC CAGCC CAGCC TCCGAC A CCCC A C C C C C C C C C C C C C	oL BAGCC CCCCCC CCCCCC CCCCCC AAAG K AAAG K RF3 TGG W GCCC A	TAGG CATCC CACCTT CTG CTG CTG CCCC R GTG V TGG W	AAC CCC CCC CCC CCC CCC CAC H AAG K	CCTGC CCTGC CCTGC CCTGC CCTGC CCGC CCGC	CTGC GACG CAAC GTTG GCCT F F GCG A CAG Q AAG K	GCCG GACG CTGC CTGC CCCTY GGA L GTC V V AAG K CTC C L AAG K	TTCT AAGA GTTG CGAC GCTG CCC P GCC A GTC V AAG K	GCC C C GCC C GCC C GCC C GCC C GCC C GCC C GCC C GCC C A CCC C C C A CCC C C C C C GCC C C C C C C C C C C C	CITITI COTTITI CAAAA AGCCT AGCCC C A CAC G H CAC G K K K CAA G K CAA G E	RP2 IG CIV AC GAA ICCCAA AGGOTY IT TCC: A CC CCC P P AC CCC D A AC CCC D A CC CCC CC CCCC CC CCC CC CCCC CC CCC CC CCCC CC CCC CC CCCC CC CCCC CC CCC CC CCC CC CCCC CC CCCCC CC CC CCCC CC CCCC CC CCCCCCCC	CCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ACT ( TGA ( ACC ( TGG ( TGG ( TGG ( R CGT L GCG A CGC R	CACCA CATCA CATCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACA CACACACA CACACA CACACA CACACACACACACACACACACACACACACACACACACACA	CCAAG CCTTC CCCT P CCCT P CCCC R CCCT R CCCC R CCCC R CCCC R CCCC R CCCC R	GGTAC CCATK CCGGGJJ GCCCT CCCCT GAA E TGG GAA E TGG W Hin GTC V CTG L	GTC V AAGG GTC V AAG K CII GAC D GAG E	ACG ACG ACG ACG ACG A CCAGC A T T T C C C C C C C C C C C C C C C	TTGC AACC CCCC CAGCC CAGCC P ATCC S CAG Q	GCTCI COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE COACE	GCC AGAG TCGC A GAG E CTG L CAC H	GCT ACCC TGGC GCT A GCC A AAG K GAG E	CCGAG GCCC CACTC/ CCCC P GCC G GAG E GCG A	AAC N GGC G AAG K TGG W
+1598 +1696 +1796 +1897 +1978 +2059 +2140 +2221	CTCAACC GAGTTGC GTGATGC GTGATGC GGTCAC CACTACC GGTCAC T ACC GTC T ACC GTC T C ACC GAC T C ACC GAC T C ACC GAC C ACC GAC C C ACC GAC C C ACC ACC C C C C C C C C C C C C C	<u>ZY</u> CCTCC GGAGC GTCGA CAGCTC TCGAC nod CCCC A CCC A CCC A CCC A C C CAC A C C C C C C C C C C C C C C C C C C C C	oL BAGCC CTCGC CTCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCGCC CCCCCC CCCCCC CCCCCC CCCCCC CCCCCC	CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG	AAC CGC CGC CGC CGC CGC CAC E GAG E GAG E CAC H AAG K AAC	COCCO COCCO COCTO COCTO COCTO COCTO COCTO COCTO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCCO COCTO COCTO COCTO COCCO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCCO COCCO COCCO COCCO COCCO COCCO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO	CTGC GACG GTTG GCCTT CGGA TTC F GCG A CAG CAG K CAG CTG	GCCG GACG CTGC CTGC GGGA L GTC K CTC L L CTC L L CTC K K CTC K K CTC C C CCCT K	TTCT AAGA GTTG CGAC GCTG CCCC P GCCC A GTC V AAG K GTC K	GCC 2 GCC 2 GCC 2 GCC 2 GCC 2 GCC 2 A GCC 2 A GCC 2 A GCC 2 A GCC 2 A GCC 2 A GCC 2 A GCC 2 A GCC 2 A C GCC 2 A C GCC 2 C C C C C C C C C C C C C	O COADA ACTITI COTITI COADA ACCAR ACCCC A ACCCC A ACCCC A ACCCC A ACCAC A ACCAC A AAAA G K AAAA G K AAAA G K AAAAA G K AAAAA G K AAAAAAA G K AAAAAAAAAA	RP2 TG CTX AC GAC AC GAC ACCONT TT TOG ⁷ A CG CCC P P AC GCC A C CG CCC A C CG CCC A C CC CCC C C CCC C C CCC C CCC C C CCC C C CCC C C CCC C C CCC C CCCC C CCC C CCCC C CCCC C CCCC C CCCCC C CCCC C CCCC C C CCCC C CCCC C CCCCC	CAG CAG CAG CAG CAG C CAG C CAG C C C C	ACT ( TGA ( TGG ( TCCCG ⁴ ) TCCCG ⁴ R CTC L GCG R CGC R GAA	CTCCCC SAGGC STAGT CACC CACC A ACG T GAG E CGC R GCC	CAAC COTTY COCCCI CCCT P CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCCC	GAA CCATC CCCCT GAA E TGG W Hin GTC V CTG L CTG	CCCC	BAGA CCAGC T CCAGC T GCC A GCC A T T C C C C C C C C C C C C C C C C	TTGC AACC CGGC CAGCC P ATC I TCC S CAG Q AAG	CCCCC CCCCCC CCCCCC CCCCCC CCCCCC CCCCC CCCC	CCCG2 CCCG2 TCCGC TCCGC TCCGC CCCG2 CCCG CCCG	CTCC CACCC CTGGC CTGGC CTG CTG CTG	CCGA GCCC P CCC P CCC P CCC C C C C C C C C	AAC N GGC G AAG K TOG W CAC
+1598 +1696 +1796 +1897 +1978 +2059 +2140 +2221	CTCAACC GAGTACC GTGATGC GTGATGC CCACGGC GGTCCAC Z ORF5 ATG ACC M T ACC GTC T V ACC GTC T V ACC GAC T D GAG AAC E K AAG GCC K A	<u>Xh</u> CCTCC GGAGC GTCGA GTCGA CCAGC TCGAC A CCCA A CCCA A CCA A CAA E CAA T CAA T CAA T	oL BAGCC CCCCCC CCCCCC CCCCCC CCCCCC CCCCCC	CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG	AAC CCC CCC CCC CCC CCC CAC CAC CAC CAC	COCCOC COCTO COCTO COCTO COCTO COCTO COCTO COCTO S S COCTO S S COCTO S S COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO COCTO	CTGC GAC GTTG GTTG CGGA TTC F GCG A A CAG Q AAG A K CTG L	GCCG GCCGC CTGC CCTGC GGGA CCCCT GGC CCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCCT CCCCCT CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCC	TTCT AAGA GTTG CGAC GCCC P GCCC A GTC V AAG K GTC V	GCC (CA) GCC	O RAAAA RACTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COTTITI COCO A COC CO CO CO CO CO CO CO CO C	RP2 TG CTV AC GAA TCCCAAGOTY TTCC A CG CCC P P AC GCC CG CCC P P AC GCC A GCC CG CCC CG CCCC CG CCCC CG CCC CG CCCC CG CCCC CG CCCC CG CCC CG CCC	CTCC CCCCCC CCCCCC CCCCCC CCCCC CCCCC CCCC	ACT ( TGA ( TGG ( TCCCG' R CTC L GCG A CCCC R CCC R GAA E	CTCCCC BAGGC CATC2 CATC2 CATC2 CACG CACG R GAGG R GCCC A	CAAA COTTO COCCT P CCCT P CCCC R GAA E GCG A GCG A GCG G GCG G	GAA CCATC CCATC CCATC CCCCT CCCCT CCTCG CCTCG CCTCG CCTCG CCTCG CCTCG	CCCC CCCC CCCCC CCCCC CCCC CCCC CCCC CCCC	ACG TTC CCAGC ACG T GCC A TTC F CGC R GAC D Sty/I	TTGC GCCC CGGC CAGCC P ATC I TCC S CAG Q AAG K	GCTCI CACT CACT A ACC T CG S GCC A T TCG S GCC A T TCG S GCC A T TCG S	CACC AGAC CACC A CACC A CACC A CACC A CACC A CACC A CACC A CACC A CACC A CACC A CACC A CACC A CACC A CACC	GCTCC GAG GCT A GCC A AAG K GAG E CTG L	CCCGAC STGCC CACTC/ CCCC P GGC G GAG E GAG E GAG E GAG E CCG A CCCC	AAC N GGC G AAG K TOG W CAC H
+1598 +1696 +1796 +1897 +1978 +2059 +2140 +2221 +2302	CTCLACC GACTACC GTGATGC GTCLCC GGTCCCC C GGTCCCC ATC ACC M T ACC GT C C C C C C C C C C C C C C C C C	XD CCTCC GGAGC KOTCGG CAGCC TCGAG CAGCC A CGC A CGC A C CGC A C C CGC A C C CGC A C C CGC C A C C CGC C C CGC C C CGC C C C	eL 33AGCC 34GCC 35GCC 35GCC 35GCC 35GCC 4 AAG K AAG K AAG K TGG W GCC A CAC H CAC CAC CAC CCCCC CCCCC CCCCCC CCCCCC CCCCCC CCCCCC	CTG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CC	AAC CCCTT CCCTT CCCTT CCCTT CCCTT CCCTT CCCTT CCCTT CCCTT CCCTT CCCC CCTT CCCTT CCCTT CCCTT CCCC CCTT CCCCC CCTT CCCCCC	CCCCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CTGCC GACC GTTG GCCTT CCGGA TTC F GCG . A CAG CAG CAG CTG L AAC N	GCCG GCCG GCCG GCCG GCCG CCCCT GCCG CCCCT C GCC GC	TTCTT AAGA GTTG GCC CCC P GCCC A GTC V AAG K GTC V CTG L	C C CAR C C CAR C C CAR C C CC C C C C C C C C A C C C C A C C C C A C C C C A C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C	CONTINUES OF CONTI	RF2 TG CTX AC GAC ACGAC AGGOOTY T TOG ² A CG CCC CG CCC P P AC GCC CG CCC A CG CCC A CCC A CG CCC A CCC A CCC A CCC A CCC A CCC A CCC A CCC CCC A CCC A CCC CCC A CCC CCC A CCC CCC A CCC CCC A CCC CCC A CCC CCC A CCC CCC A CCC CCC CCC CCC CCC A CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCCCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCCC CCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCCC CCCCCC	CAG CAG CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCCCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC CCTCC	ACT ( ACC ( ACC ( TGG ( TGG ( TCCG' R CTC L GCG A CGC R GAA E GCG A	CTCCCC CACC CACC CACC CACC CACC CACC CACC CACC A CACC CACC CACC CACC A CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CACC CAC	CCAAC COTTO CCCT P CCCC R GAA E GCCG G GCC G GAC D	GAA CCATC CCATC CCACC CCCC CCCC CCCC CCC	CCCC CCCCC CCCCC CCCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCCC CCCCC CCCCC CCCCCC	ACG T CCAGC T GCC A T T C GCC A T T T C GCC R GAC GAC C C C C C C C C C C C C C C C C	TTGC CGGC CGGC CAGCC P ATC I TCC S CAG Q AAG K GCG A	GCT GCT A CCCCC GCT A A CCC A CCC A CCC CCC CCCC A CCCC A CCCC A CCCCC A CCCCC A CCCCCC A CCCCCC A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCC A A CCCCCCC A A CCCCCCC A A CCCCCCC A A CCCCCCCC	CTCGC GCCC A GCCC A GAG CTG CTG CAC H GAC D GCC G C C C C C C C C C C C C C	CTCCC CACCO CTCCC A AAG CTCC CTCC L CTCC L	CCGAA SGCTX PTGCC CACTC/ ACTC/ CCC P GGC G GAG E GAG E CG A CCC S	AAC N GGC G G AAG K TGG W CAC H GTG V
+1598 +1696 +1796 +1897 +1978 +2059 +2140 +2221 +2302 +2383	CTCLACC GACTACC GTCLACC GTCLACC GGTCLAC GGTCLAC CACC ACC T ACC T ACC CACC T ACC CACC T ACC CACC CACC T ACC CACC CACC T C CACC ACC	XA CCTCC GGAGG CAGCT CCAGC CAGCT TCGA A CCCA A CCCA A CCCA A CCCA A CCCA A C CCCA A C CACA C C CACA C C CCCA C C CCCACC C CACCC C CACCC C CACCC C CACCC C CACCC C CACCC C CACCC C CACCC C CACCC C CACCC C C CACCC C C CACCC C C CACCC C C C CACCC C C C C C C C C C C C C C C C C C C	eL sasecc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc cocosc c	CTG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CC	AAC TTG GCG CCCT CCCT CCCT CCCT CCCT CCCT CCCC CCCC R CCCC R CCCC R CCCC CCCT CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCCCC	COCCO COCCOCC COCCOCC COCCOCC COCCOCC COCCOC	CTGC GACG GTTG GCCTT CCGGA TTTC F GCG A CTG CAG CTG CTG CTG CTG CTG CTG CTG CTG CTG CT	GCCG' GCCGC GCCGC GCCCTGC GCCCTGC GCCCTGC K CTCC C CCCC G GCCC G GCCC G GCCC G GCCCC G GCCCC G GCCCC G GCCCCCC	TTCTT AAGA GTTG GCTG CCCC P GCC A GTC V AAG K GTG V CTG L CCCC	C C C C C C C C C C C C C C C C C C C	C C C C C C C C C C C C C C C C C C C	RP2 TG CTX AC GAA AGGGTY TF TGG' A CG CCC P P AC GCC A CG CCC P P AC GCC GC CCC CG CCC	CAG CAG CAG CAG CTCC CTCC CTCC CTCC CTCC	ACT ( TGA ( TGG (	CTCCCC BAGGC STAOT CATC2 ICAGC GCT A ACG T GAG E GCC A GAG E AAC	CAAC COTTA COCCT P CCCC R GAA E GCG A GCC G G G G G CCCT P CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCC R CCCCCC	CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG	CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCCC CCCCC CCCCC CCCCCC	BAGA PTTC PTCT CCAGC T GCC A GCC A TTC F CGC R GAC D CGC R GAC L GAC TTC TTC TTC CGC R GCC C A CGC T T CCAGC T T CCAGC T T CCAGC T T CCAGC T T CCAGC T T CCAGC T T CCAGC T T CCAGC T T CCAGC T T CCAGC T T CGC R T CGC R T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T CGC T T T CGC T T T CGC T T T CGC T T T CGC T T T CGC T T T CGC T T T CGC T T T CGC T T T T T CGC T T T T T T T T T T T T T	TTGC AACC CGGGC CAGCC P ATC I TCC S CAG Q AAG K GCG A AGC	GCTCI GCT A ACCC A ACCC T TCG S GCCC A TTCC F CTG L ATCC	CCCCC	CTCC CA CTG CTG CTG CTG CTG CTG L CTG L AAG	CCCAA STGCC STGCC STGCC STGCC STGCC CCC P GGC G GGC G GGC A GAG E GGC A CCC CCC CCC CCC CCC CCC	AAC N GGC G AAG K TOG W CAC H GTG V ACG
+1598 +1696 +1796 +1897 +1978 +2059 +2140 +2221 +2302 +2383	CTCLACC GACTACC GTGATCCAC GGTCCAC GGTCCAC C CACTACC ACC T ACC GTC T ACC GAC AAC C C C C C C C C C C C C C C	<u>Xh</u> CCTCC GGAGC CAGCC CAGCC TCGAC TCGAC A CGCC A C CGCC A C C C C C C C C C	⊘I 33AGCC 33GGCCC 3GGCCCCCCCCCCCCCCCCCCCCC	CTIG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCAC	AAC TTG GCG CCC CCC CCC CCC CCC CC	CCCCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CTGC GACG GTTG GCCTTG CCGGA al TTC F GCG A CAG Q AAG A K CTG L L AAC N N TTC F	GCCCF GGCA CTGC CCCTT GGCA K CTC C CCCT K CTC C G GCC C G GCC C G GCC C G GCC C CAC C H	TTCT TAAGA CAAC GTTG GCTG CCCC P GCCC A GTCC V AAG K GTCC V AAG K CTCG C CCC R	C C CAP C C CAP C C CAP C C CAP C C C C CAP C C C CAP C C C CAP	O VAAAA ACTITI COTITI COAAA AAACCT AAACCA AAAAG K AAAAG K AAAAG K AAAAG K AAAAG K AAAAG K AAAAG K AAAAG K AAAAAG K AAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG K AAAAAAG C C C C C C C C C C C C C	RP2 TG CTX AC GAC ACGOTY T TCG7 A CG CCC P P AC GCC CG CCC P P AC GCC CG CCC CG CCCC CG CCC CG CCCC CG CCC CG CCC CG CCCC CG CCC CG CCC CG CCC C	CAG CTCC CCAG CCTCC CCCC CCCC CCCC CCCC	ACT ( ACC ( TGG ( TGG ( TGG ( TGG ( R CTC L CTC L GCG A CCC R GAA E GCG A CCC S	CTCCCC SAGGC SAGC SATO SATO SATO SATO SAGC CGC R GAG CGC R GAG CGC R GAG CGC R GAG C CGC R GAG C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C C C C A C A C A C C C C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A	CAAC COTTA COTTA COCCT P COCCT P COCCT P COCCT R GAA E GCG A GAC C D TAC Y	GAA E CCATC CCCCT GAA E TGG GAA E TGG GTC CTG L CTG L CTG L CTG L CTG L CTG CTG V	CCCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	STTC CAAG TCT GCC ACG T GCC A TTC F CGC R GAC D Styl AAG K TGG W	TTGC CCCC P ATCC I TCCC CAGC Q AAGC K GCG A AGC S	GCTCI GCTCI CACI GCCCCC GCTCA A ACCC T TCG S GCCCA T TCG F CTG L ATCC I ATCCI I	CTCC GCC A GCC A GAG E CTG L CAC H GAC D GAC G CCC Q CCC P	CTCC GAC TGGAC TGGAC TGGAC TGGAC A AAG CTG L AAG K CTG L AAG K	CCCC P CCCC P CCCC C CCC C C C C C C C	AG AG AAC N GGC G AAG K TGG W CAC H GTG V ACG T
+1598 +1696 +1796 +1897 +1978 +2059 +2140 +2221 +2302 +2383 +2464	CTCLACC GACTACC GGTCLACC GGTCLACC GGTCLAC CACTACC ACC ACC T ACC T ACC CACC T ACC CACC ACC	XA CCTCC GGAGG CAGCC CAGCC TCGAG A CAGC A CCCG A CCCG A CCCG A CCCG C CCCG C CCCG C CCCG C CCCG C CCCG C CCCG C CCCG C CCCCCC	eL shacecose shacecose shacecose and and and and and and and and and	TAGG CCTG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCACG CCCCCC CCACG CCCCCCCC	AAC TTG GCG CCGC CCGC CCGC CCGC CAC H AAC CCGC R TTC F ATC I	CCCCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CTGC GACG GTTC GCGA TTC F GCG . A CAG . Q AAG . K CTG . L AAC . N TTC . F TTC . F	GCCC GCCC GCCC GCCC GCCC GCCC CCCC CCC	TTCTT AAGA GTTG GCTG CCCC P GCCC A GTCC V AAG K GTCG V CTCG L CCCC R AAG K	C C C C C C C C C C C C C C C C C C C	CONTINUES OF CONTI	RP2 TG CTX AC GAC AGGGTY T TCC: A CG CCC P P CG CCC P P AC GCC CG CCC A CG CCC A A CG CCC A A C CCC A A C CCC A A C CCC A A C CCC A A C CCC CCC A A C CCC A A C CCC A A C CCC A A C C CCC A A C C C A A C C C C A A C C C C C A A C C C C C C C C C C C C C	CAG CAG CAG CTCC CTCC CTCC CTCC CTCC CTC	ACT ( TGA ( TGG ( TGC ( TGC ( TGC ( TCCC) CGT R CTC L GCG A CGC R GAAA E GCG A TCC S GCG A	CTCCCC CAGC CAGC CAGC CAGC CAGC CAGC CAGC Q CAGC Q	CAAC COTIN COTIN COCCI P CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCCC R CCCC R CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC	GAA E TGG W Hin GTC CTG L CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG	CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCCC CCCCC CCCCC CCCCCC	BAGA PTTC PTCT CCAGC T GCC A T GCC A T GCC A T GCC A T GCC A T GCC A T GCC C A C GCC C A C GCC C A C C C C C C C C C C C C C	TTGC AACC CCCC P ATC I TCC S CAG Q AAG C CAG Q AAG C CAG Q CAG C C C C C C C C C C C C C C C C C C	GCTCH GCTCH GCTCH ACCCCCC ACCCCCC ACCCCCC ACCCCCC ACCCCCCCC	CTCC GCC A GCC A GAG E CTG CAC D GAC D GAC C P GCC V	CTCC CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER CALCER	CCCAA TGCC TGCC CCC P GCC G GAG CCC CCC CCC CCC CCC	AAC N GGC G G AAG K TGG W CAC H GTG V ACG T CCGG R

+2626	GTG	CTG	GTG	AAG	GTG	GAC	CTG	AAG	GAC	TTC	TTC	CCC	TCG	GTG	AAC	TGG	CGC	CGG	GTG	AAG	GGC	CTG	TTG	CGC	AAG	GGC	GGC
	v	L	v	ĸ	v	D	L	K	D	F	F	Ρ	S	v	N	W	R	R	v	К	G	L	L	R	K	G	G
+2707	CTG	AAG	GAG	AAC	ACC	TCC	ACG	CTG	CTG	GCG	CTG	ATG	TCC	ACG	GAG	GCC	CCG	CGC	GAG	CGC	ATG	TCC	TTC	CGG	OGC	AAG	ACG
	$\mathbf{L}$	K	Е	N	Т	S	Т	L	L	A	L	М	S	Т	Е	А	Ρ	R	E	R	М	S	F	R	G	К	т
	000	~~~	000	000		000	000	100	000	000	000	~~~	000	000	000	100	maa	Sina.	<u> </u>	2000	200		000	CILC	maa	CTTC	000
+2700	CIG	CAC	GIG	GCG	AAG	GGG	CCC	AGG	GUU	CTC	CCG	CAG	GGC	GCG		ACC	TCC	CCG	GGC	ATC	ACC	AAC	GCG N	CIG.	100	T	D
	ь	н	v	А	ĸ	G	Р	ĸ	А	Ц	Р	Ŷ	G	А	P	. <b>T</b> .	5	Р	G	Т	1	14	А	ы	C	Pvul	II I
+2869	CTG	GAC	AAG	CGG	CTG	TCG	GCG	CTG	TCG	CGC	AAG	CIG	GGC	TTC	ACG	TAC	ACG	CGC	TAC	GCG	GAC	GAC	CTG	ACC	TTC	AGC	TGG
	L	D	к	R	L	S	A	L	S	R	К	L	G	F	т	Y	т	R	Y	А	D	D	L	т	F	S	W
							StyI																				
+2950	ACG	AAG	GCG	AAG	GCG	ccc	AAG	GCC	CGC	CGC	GCG	CAG	GGA	GCG	CCG	GTG	GCG	GTG	CTG	CTC	GCG	CGA	GTG	AAG	GAC	ATC	GTG
	т	К	А	К	А	Ρ	К	А	R	R	А	Q	G	А	Ρ	v	А	v	L	L	А	R	V	K	D	I	V
+3031	GAG	GCG	GAA	GGC	TTC	ACG	GTC	CAC	CCG	GAC	AAG	ACG	CGC	GTC	GCC	CGC	AAG	GGC	AGC	CGT	CAG	CGC	GTC	ACG	GGG	CTC	GTG
	E	А	Е	G	F	Т	v	н	Ρ	D	K	Т	R	V	A	R	К	G	S	R	Q	R	v	т	G	L	v
+3112	GTG	AAC	GAG	GCC	AAG	GAC	GGC	ACG	CCC	GCC	GCC	CGG	GTC	ccc	CGC	GAC	GTG	GTG	CGC	CGC	CTC	CGC	GCG	GCC	ATC	CAC	AAC
	v	N	E	A	К	D	G	Т	P	A	A	R	v	Р	R	D	v	v	R	R	L	R	A	A	I	Н	N
+3193	CGC	CTG	AAG	GGC	AAG	CCC	GGC	CGC	GAA	GGC	GAA	TCG	CTG	GAG	CAG	CTC	AAG	GGC	ATG	GCG	GCG	TTC	ATC	CAC	ATG	ACG	GAC
	R	L	K	G	к	Ρ	G	R	Ε	G	Е	S	L	Е	Q	L	К	G	М	А	А	F	Ι	Н	М	Т	D
+3274	000	~~~	<b>N</b> N C	000	000	አመረግ	mad	cma	<b>C</b> M m	CNC	CILC	000	220	CTPC	<b>C h h</b>	000	000	100	ccc	000	CAC	ccc	መልሮ	cocr	ምሮእ		א ידעי־ע־
+J2/4	000	GAG	AAG	GGC	CGC D	M	V	t	GAI	CAG	T	GCG N	AAG	T	GAA	SCC N	GCC N	ACG m	D	SCC N	CAG	acc x	1AG	000.		5.000	JUIN
	Р	£	v	G	r	<b>t</b> 41	t	ц	D	¥	ц	А	r	ц	5	A	A	1	£	A	¥	A					
+3358	መድርሳ	2001	አረጥሮ	NCCN	2000		າດຕາມ	ባለ ርግሞ	്നവ		r a	гъссо		ላ አረታምሳ	Race:	י ביתה	00222	$\gamma \alpha \gamma \gamma \gamma \gamma$	~								

FIG. 4. Complete nucleotide sequence of retron ML162 from *Melittangium lichenicola* (GenBank accession number L36722). Potential ORFs are labeled, and the ATG start codons are in boldface. The msDNA coding region for ML162 starts at nucleotide +1662 and includes the RNA (*msr*) and DNA (*msd*) coding regions for msDNA (boxed), the branched guanine residue at position +1597 (circled), and the a2 and a1 inverted repeats at +1662 and +1841 (arrows). The RT ORF (ORF5) begins at +1896 and ends at position +3342.

ation observed between the RT genes of *Myxococcus* and *Melittangium* species, although relatively small, is probably evolutionarily significant, having accumulated changes at the same rate as most other genes since the divergence of the genera *Myxococcus* and *Melittangium* from a common ancestor.

### DISCUSSION

It is evident from findings presented here and previously (4, 19, 29) that the family of retrons found in the myxobacteria are unique compared with retrons found in other bacteria and that the origin and acquisition of these retrons in the myxobacteria may be different from those of the other proteobacteria. The first indication is based on the distribution of retrons in the myxobacteria such as *E. coli*. Retrons are ubiquitous, or nearly so, in the myxobacteria,

being absent in only one Cystobacter strain out of 28 myxobacterial strains tested in this study. In contrast, only about 10% of all E. coli, Klebsiella, Salmonella, Proteus, Rhizobium, and Bradyrhizobium strains tested have been demonstrated to contain retrons (11, 29). The distribution of a particular retron type within the myxobacteria appears to cluster within a phylogenetic subgroup (Fig. 3). For example, elements very similar to retron Mx162 (originally discovered in M. xanthus) were found in five different species of Myxococcus and three species of *Cystobacter* as well as in *Melittangium lichenicola*, *C. coralloides*, Archangium gephyra, Angiococcus disciformis, and previously S. aurantiaca (Table 1) (9). On the basis of 16S rRNA analysis (34), all of these bacteria form a phylogenetically related cluster or subgroup within the myxobacteria (Fig. 3). Likewise, all 10 strains of the Nannocystis subgroup surveyed appear to share the same novel retron element (Table 1).



FIG. 5. DNA sequence alignment of the 3' ends of ML162 and Mx162. The alignment was created by using the Bestfit program of the Genetics Computer Group software (2). Stop codons for each RT-coding sequence are boxed. The sequences shown begin at positions +3277 and +3268 for ML162 and Mx162, respectively, and end at positions +3440 and +3435. A close match between the two sequences appears to end around the stop codon for their respective RT ORFs.

 TABLE 2. Rates of change at silent sites for RT and other bacterial genes

Organisms compared	Time of divergence (million yr ago)	Gene	% Diver- gence at silent sites (corrected)	Rate of substitu- tion at silent sites/million yr
Melittangium sp Myxococcus sp.	90	RT	46	0.26
E. coli-Salmonella sp.	140	<i>trpA</i>	177	0.63
1	140	ptsH	23	0.08
	140	-	94 (avg)	0.335 (avg)

This clustered distribution of the Mx162 retron within the *Myxococcus* subgroup implies a vertical rather than a horizontal transmission of this particular element. In other words, an evolutionary scenario can be envisioned in which the genome of a common ancestor acquired an Mx162-type retron element, with subsequent inheritance of this element by all the different genera during the divergence and evolution of this subgroup.

Other evidence also tends to support this scenario. For example, a homolog of retron Mx162 found in *Melittangium lichenicola*, retron ML162, was cloned, and its DNA sequence was compared with sequences of two previously characterized examples of this retron type (Mx162 and Sa163). Overall nucleotide identity between retrons Mx162 and ML162 is 77%. In addition, the silent site substitution rate for the RT ORFs from these two retrons is calculated to be 0.26 substitutions per million years (Table 2). These figures are consistent with the view that these two genes have diverged for at least as long as the genera *Myxococcus* and *Melittangium* shared a common ancestor, for about 90 million years ago.

In contrast, when the retron elements found in other proteobacteria such as E. coli are compared with the findings presented here for the myxobacteria, it appears unlikely that the Mx162 retron is a recent acquisition by the myxobacterial genome. For example, in E. coli, the retron element Ec107 is widely distributed among ECOR strains, a reference collection of natural isolates (10, 11, 17). However, its occurrence is sporadic, and this element is found in distantly related phylogenetic branches of the ECOR collection. Also, DNA sequence determination of several individual examples of the Ec107 element revealed little or, in some cases, no nucleotide sequence diversity among these retrons (17). In addition, codon usage within the RT ORFs of most E. coli retrons is atypical for native E. coli genes (16). All of these features appear to mark the Ec107 retron as a recent addition to the genome of certain lineages of E. coli, with the likely horizontal spread to other strains.

The features described here for the Mx162 class of retron found in the myxobacteria do not resemble those of the E. coli retrons and appear to be inconsistent with a recent inheritance in the genome. Indeed, an analysis of the codon usage for the RT ORF from the retron of Melittangium lichenicola revealed a heavy bias toward codons that use predominantly G and C, and this usage is typical of other known myxobacterial genes (16, 33). Although recent acquisition from a GC-rich organism cannot be discounted, this characteristic does suggests that the retron-encoded RT gene has resided in the myxobacterial genome for millions of years, at least since the speciation of the Myxococcus subgroup. However, such inheritance may not be true for other myxobacterial retrons. Retron Mx65 was found in only two species other than M. xanthus and may have been introduced into or lost from the Myxococcus genome at a later time.

Evidence continues to indicate that, as in the eukaryotes, RT-encoding genetic elements of diverse types are probably widespread and a nascent characteristic of the prokarvotic world. This idea is supported by the prevalence of diverse retron elements, among a wide variety of bacteria, as documented here and in a previous report (29). Additional evidence comes from the recent discovery of RT-encoding group II introns in several groups of the eubacteria (8). Many questions about the retron elements need to be answered, including their origin in the bacterial genome, potential mobility, and relationship with other classes of retroelements. On the basis of the prevalence of RT genes, has the process of reverse transcription played a role in the generation of genetic variation or otherwise influenced the evolution of the bacterial genome? Indeed, the chromosomes of the myxobacteria appear to have a particular predilection for retron elements and may be a good place to look for clues to answer these questions. Because retrons can be deleted from the chromosome of *M. xanthus* without affecting their growth or development in the laboratory, no function has been linked to these genetic elements. It cannot be discounted, however, that these RTencoding elements have a function which has provided a selective advantage for the myxobacteria. This could be an alternative explanation for why retron elements have been retained by the myxobacterial genome for millions of years.

### ACKNOWLEDGMENTS

We thank Larry Shimkets and Hans Reichenbach for providing many of the myxobacteria strains used for this study. We also thank Tom Whittam for critical reading of the manuscript and help in generating the synonymous substitution rates and codon bias statistics.

#### REFERENCES

- Chomcyzynski, P., and N. Sacchi. 1987. Single-step method for RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162:156–159.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for VAX. Nucleic Acids Res. 12:387–395.
- Dhundale, A., T. Furuichi, M. Inouye, and S. Inouye. 1988. Mutations that affect production of branched RNA-linked msDNA in *Myxococcus xanthus*. J. Bacteriol. 170:5620–5624.
- 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.
- Dhundale, A., M. Inouye, and S. Inouye. 1988. A new species of multicopy single-stranded DNA from *Myxococcus xanthus* with conserved features. J. Biol. Chem. 263:9055–9058.
- Dhundale, A., B. Lampson, T. Furuichi, M. Inouye, and S. Inouye. 1987. Structure of msDNA from *Myxococcus xanthus*: evidence for a long, selfannealing RNA precursor for the covalently linked, branched RNA. Cell 51: 1105–1112.
- Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6–13.
- Ferat, J.-L., and F. Michel. 1993. Group II self-splicing introns in bacteria. Nature (London) 364:358–361.
- 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.
- Herzer, P. J., S. Inouye, and M. Inouye. 1992. Retron-Ec107 is inserted into the *Escherichia coli* genome by replacing a palindromic 34 bp intergenic sequence. Mol. Microbiol. 6:345–354.
- Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam. 1990. Phylogenetic distribution of branched RNA-linked multicopy, single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. 173:6175–6181.
- Hsu, M.-Y., M. Inouye, and S. Inouye. 1989. Structural requirements of the RNA precurser for the biosynthesis of the branched RNA-linked multicopy, single-stranded DNA of *Myxococcus xanthus*. J. Biol. Chem. 284:6214–6219.
- Hsu, M.-Y., C. Xu, M. Inouye, and S. Inouye. 1992. Similarity between the *Myxococcus xanthus* and *Stigmatella aurantiaca* reverse transcriptase genes associated with multicopy, single-stranded DNA. J. Bacteriol. 174:2384– 2387.
- Inouye, M., and S. Inouye. 1992. Retrons and multicopy single-stranded DNA. J. Bacteriol. 174:2419–2424.

- Inouye, S., P. J. Herzer, and M. Inouye. 1990. Two independent retrons with highly diverse reverse transcriptases in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. USA 87:942–945.
- Inouye, S., M.-Y. Hsu, S. Eagle, and M. Inouye. 1989. Reverse transcriptase associated with the branched RNA-linked msDNA in *Myxococcus xanthus*. Cell 56:709–717.
- Kawaguchi, T., P. J. Herzer, M. Inouye, and S. Inouye. 1992. Sequence diversity of the 1.3 kb retron (retron-Ec107) among three distinct phylogenetic groups of *Escherichia coli*. Mol. Microbiol. 6:355–361.
- Lampson, B. C., M. Inouye, and S. Inouye. 1989. Reverse transcriptase with concomitant ribonuclease H activity in the cell-free synthesis of branched RNA-linked msDNA of *Myxococcus xanthus*. Cell 56:701–707.
- Lampson, B. C., M. Inouye, and S. Inouye. 1991. Survey of multicopy singlestranded DNAs and reverse transcriptase genes among natural isolates of *Myxococcus xanthus*. J. Bacteriol. 173:5363–5370.
- Lampson, B. C., J. Sun, M.-Y. Hsu, J. Vallejo-Ramirez, S. Inouye, and M. Inouye. 1989. Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. Science 243:1033–1038.
- Lampson, B. C., M. Viswanathan, M. Inouye, and S. Inouye. 1990. Reverse transcriptase from *Escherichia coli* exists as a complex with msDNA and is able to synthesize double-stranded DNA. J. Biol. Chem. 265:8490–8496.
- Li, W.-H, C.-I. Wu, and C.-C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2:150–174.
- Lim, D., T. A. T. Gomes, and W. K. Maas. 1990. Distribution of msDNA among serotypes of enteropathogenic *Escherichia coli* strains. Mol. Microbiol. 4:1711–1714.
- Lim, D., and W. Mass. 1989. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. Cell 56:891– 904.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499–560.

- Ochman, H., and A. C. Wilson. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. 26:74–86.
- Reichenbach, H., and M. Dworkin. 1992. The myxobacteria, p. 3416–3487. In A. Balows (ed.), The prokaryotes. Springer-Verlag, New York.
- Rice, S., J. Bieber, J.-Y. Chun, G. Stacey, and B. C. Lampson. 1993. Diversity of retron elements among a population of rhizobia and other gram-negative bacteria. J. Bacteriol. 175:4250–4254.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
- Sharp, P. M. 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. J. Mol. Evol. 33:23–33.
- Shimamoto, T., M.-Y. Hsu, S. Inouye, and M. Inouye. 1993. Reverse transcriptases from bacterial retrons require specific secondary structures at the 5'-end of the template for the cDNA priming reaction. J. Biol. Chem. 268:2684–2692.
- Shimkets, L. 1993. The myxobacterial genome, p. 85–107. *In* M. Dworkin and D. Kaiser (ed.), Myxobacteria II. American Society for Microbiology, Washington, D.C.
- Shimkets, L., and C. R. Woese. 1992. A phylogenetic analysis of the myxobacteria: basis for their classification. Proc. Natl. Acad. Sci. USA 89:9459– 9463.
- Sun, J., M. Inouye, and S. Inouye. 1991. Association of a retroelement with a P4-like cryptic prophage (retronphage f73) integrated into the selenocystyl tRNA gene in *Escherichia coli*. J. Bacteriol. 173:4171–4181.
- 36. Temin, H. M. 1989. Retrons in bacteria. Nature (London) 339:254-255.
- Wilson, C. A., S. S. Carlson, and T. J. White. 1977. Biochemical evolution. Annu. Rev. Biochem. 46:573–639.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.
- Yee, T., T. Furuichi, S. Inouye, and M. Inouye. 1984. Multicopy singlestranded DNA isolated from a gram-negative bacterium, *Myxococcus xanthus*. Cell 38:203–209.
- Xiong, Y., and T. Eickbush. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9:3353–3362.