Characterization of the Integrase Gene and Attachment Site for the Myxococcus xanthus Bacteriophage Mx9

Bryan Julien*

Kosan Biosciences, Inc., Hayward, California 94545

Received 9 September 2002/Accepted 5 August 2003

Bacteriophage Mx9 is a temperate phage that infects *Myxococcus xanthus*. It lysogenizes the bacteria by integrating into the bacterial chromosome by site-specific recombination at one of two sites, *attB1* or *attB2*. Integration at *attB1* results in deletion of DNA between the two *attB* sites. The *attB2* site lies within the 5' region of the *M. xanthus* tRNA^{Gly} gene. Mx9 integration requires a single protein, Int. Analysis of integration revealed that the phage attachment site (*attP*) is contained in the *int* gene and that upon integration, the 3' end of the *int* gene is altered. Plasmids containing fusions of the *pilA* or *mgl* promoter to *lacZ* integrated at either Mx9 *attB* site have higher levels of transcription than the same fusions integrated at the Mx8 *attB* site.

Mx9 is a general transducing phage that infects the gramnegative bacterium *Myxococcus xanthus* (9). The phage particle has a polyhedral head and a very short tail. Structurally, this phage resembles Mx8, which also infects *M. xanthus*.

The integrase gene and attachment site for Mx8 have been characterized (7, 8, 11). Integration of Mx8 by site-specific recombination requires a single phage protein, Int, and the phage attachment site, *attP*. Unlike the situation in most temperate bacteriophages, the Mx8 *attP* site is contained in the *int* gene, and upon insertion into the *M. xanthus* chromosome, the 3' end of the *int* gene is altered. This modified *int* gene produces a protein, IntX, with lower integrase specific activity (8).

Because no natural replicating plasmids have been identified for M. xanthus or for any other myxobacterium, phage attachment sites provide an efficient and stable alternative way to introduce new genes or add additional copies of existing genes to the cell. With the recent heterologous expression of the epothilone biosynthetic gene cluster in M. xanthus, the ability to engineer the host should prove to be valuable for further optimization of polyketide production (4). The Mx8 int gene and the attachment site can be used to integrate DNA into the chromosome, but expression of many genes is affected by insertion into the Mx8 attB sites; many developmental promoters, as well as two constitutive promoters, mgl and pilA, have reduced activity at the Mx8 sites (2, 6). Therefore, I set out to find another attachment site. Here I describe characterization of the int gene and attP from Mx9, as well as the sites of insertion in the M. xanthus chromosome.

MATERIALS AND METHODS

Bacteria, phage, and plasmids. DZ1 is a nonmotile strain of *M. xanthus* and was used for plating Mx9 and for characterization of the Mx9 attachment sites (12). DK816 is a natural *M. xanthus* isolate lysogenic for Mx9 (9). *M. xanthus* strains were grown in CYE medium (1) or 1% CTS (1% Casitone, 0.2% MgSO₄ · 7H₂O, 50 mM HEPES; pH 7.6). Phleomycin (Cayla) was used at a concentration of 30 μ g/ml. The Mx9 phage was reisolated from DK816 by growing a culture to the stationary phase, pelleting the cells, and plating dilutions of the supernatant onto DZ1. High-titer stocks of Mx9 were made by coring a

plaque and placing it in phage buffer (10 mM morpholine propanesulfonic acid [MOPS] [pH 7.6], 4 mM MgCl₂, 2 mM CaCl₂). The eluted phage were diluted and mixed with 0.5 ml of DZ1 in the early stationary phase. After the cells and phage were incubated at room temperature for 20 min, 2.5 ml of top agar was added, and the suspension was poured onto phage plates (1% BBL Trypticase, 0.1% MgSO₄ · 7H₂O, 1% agar, 10 mM MOPS; pH 7.6). The plates that exhibited confluent lysis after 2 days of incubation at 30°C were overlaid with 5 ml of phage buffer and incubated at 4°C overnight. The eluted phage were stored at 4°C. Phage stocks with concentrations greater than 1 × 10⁹ PFU/ml were obtained by this method. The plasmids used are described in Table 1.

Isolation of phage DNA. The phage from a high-titer stock were pelleted by centrifugation in an SS-34 rotor at 28,000 rpm for 3 h and then resuspended in TE (10 mM Tris [pH 7.6], 1 mM EDTA). The phage proteins were removed by extraction twice with phenol and twice with phenol-chloroform-isoamyl alcohol. The DNA was precipitated and resuspended in TE.

Isolation and sequence of the phage attachment site. To isolate the phage attachment site, phage DNA was partially cleaved with HinPI, and the fragments were ligated into pKOS35-93 cleaved with AccI. Plasmid pKOS35-93 is pBluescriptII SK+ with the kanamycin resistance from Tn5 ligated into the *SmaI* and EcoRI sites. One plasmid, pKOS35-117.9.7, integrated efficiently into the chromosome. The insert from this plasmid was sequenced.

Isolation of the bacterial attachment site. The bacterial attachment site (*attB*) was isolated by electroporating pKOS35-117.9.7 into DZ1, making chromosomal DNA, and then recovering the plasmid with flanking chromosomal DNA. Six kanamycin-resistant colonies were picked, and chromosomal DNA was prepared from each colony. The DNA was cleaved with either *PstI* or *XhoI*, ligated, and then transformed into *Escherichia coli*. Three colonies from each of the electroporations were picked, and the plasmids recovered were cleaved with *PstI* or *XhoI*. One plasmid from each preparation was sequenced by using either primer 183-66.3 (GAAGGAGGCACCATGCACGG) or primer 183-66.4 (CTCACTG AGAGTGAAGCCGC).

PCR amplification of Mx9 attB. Primers were designed to PCR amplify attB1 and attB2. Primers 183-99.4 (CGAGGTCCGGGACGCGCGCA) and 183-99.6 (TGCCAGGGCTTACGGCTTC) were used to amplify a 285-bp attB1 fragment, and primers 183-99.5 (TATCCCAGCAACCGCCGGAG) and 183-99.4 were used to amplify a 373-bp attB2 fragment. To amplify the native attB1 site, primers 183-99.6 and 249-179.7 (CAGCACGGGTGCAGCAAC) were used to amplify a 250-bp fragment. PCRs were performed by using chromosomal DNA from DZ1 and the FailSafe PCR system from Epicentre. The amplification conditions were 96°C for 2 min and then 30 cycles of 94°C for 30 s, 55°C for 1 min, and 72°C for 2 min.

Construction of a minimal integration plasmid. The *int* gene was PCR amplified from pKOS35-117.9.7 by using primers 111-74.4 (CCCAATTGGCTCA GGGCAGCGGCTCATT) and 111-82.5 (CCCCATGGCGCTCAGGGGTGCG TCGGACGCC). The PCR amplification conditions were those described above. The amplified fragment was ligated into the *Eco*RV site of pLitmus 28 (New England Biolabs) to create pKOS249-12. The *int* gene was removed from this plasmid by cleavage with *Eco*RI, and the DNA ends were made blunt with the Klenow fragment of DNA polymerase, followed by cleavage with *Pst*I, and the pHE24-2B (3) that was cleaved with *Pst*I, and the

^{*} Mailing address: Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545. Phone: (510) 731-5209. Fax: (510) 732-8401. E-mail: julien@kosan.com.

TABLE 1. Plasmids

Plasmid	Characteristics
pKOS35-117.9.9	Amp ^r Kan ^r ColE1, 4.6-kb fragment from Mx9 Amp ^r Co1E1, P _{T7A1} Mx8 <i>int attP</i> ⁻
pKOS139-47	Tc ^r , p15A, $P_{mgl} lacZ$, Mx8 attP
pKOS178-86	Tc ^r , p15A, $P_{nil4} lacZ$, Mx8 attP
pKOS178-177	Tc ^r , p15A, P _{pilA} lacZ, Mx9 int attP
pKOS178-188	Tc ^r , p15A, P _{mgl} lacZ, Mx9 int attP
pKOS249-31	Amp ^r Bleo ^r ColE1, P _{T7A1} Mx9 int attP

DNA ends were made blunt with the Klenow fragment of DNA polymerase I and cleaved with *NcoI*. The resulting plasmid, pKOS249-23, contained the *int* gene under control of the *E. coli* phage T7 A1 promoter that was engineered to contain two LacI binding sites to repress transcription. The bleomycin resistance gene was added to this plasmid by isolating the bleomycin resistance gene from pKOS183-112 as a *Bam*HI-to-*Hind*III fragment, and the DNA ends were made blunt with the Klenow fragment of DNA polymerase I; then the fragment was ligated with pKOS249-23, which was cleaved with *Xho*I, and the DNA ends were made blunt with the Klenow fragment of DNA polymerase I. This plasmid was designated pKOS249-31.

β-Galactosidase assays. Seed cultures of two isolates for each integration site were grown in 1% CTS (5 ml) to the mid-log to late log phase. To start an assay culture, 35 ml of 1% CTS was inoculated with 1 ml of a seed culture at an optical density at 600 nm (OD₆₀₀) of 0.073. β-Galactosidase assays were performed by removing an aliquot of cells and adding them to Z buffer to obtain a combined volume of 1 ml. The cells were lysed by adding 1 drop of 0.1% sodium dodecyl sulfate and 2 drops of chloroform and vortexing the sample for 5 s. The assay was initiated by adding 0.1 ml of *o*-nitrophenyl β-D-galactopyranoside (8 mg/ml) and mixing. The reactions were stopped by adding 0.5 ml of 1 M Na₂CO₃. The OD₆₀₀ of the cell culture and the OD₄₂₀ of the enzyme reaction mixtures were determined with a SpetraMax 250 plate reader. Miller units were determined as previously described (10).

Nucleotide sequence accession numbers. The Mx9 sequence has been deposited under GenBank accession number AY247757. The accession numbers for *attB1* and *attB2* are AY297770 and AY297771, respectively.

RESULTS

Identification of the Mx9 *int* gene and attachment site. To identify the *int* gene and attachment site, a library of 5- to 8-kb fragments of Mx9 was made, and a clone that was able to integrate into the *M. xanthus* chromosome was identified. The insert in this plasmid, pKOS35-117.9.7, was sequenced. Five complete open reading frames (ORFs) and one partial ORF were identified in the 4.6-kb fragment (Fig. 1). ORF 1 was the only reading frame whose product exhibited amino acid similarity with the products of other known integrase genes and therefore was designated *int*. The other ORFs resembled ORFs from Mx8; ORF 2, ORF 3, ORF 4, ORF 5, and ORF 6 showed similarity to P15, P14, P16, P17, and P18, respectively, from Mx8. Based on the degrees of similarity of these ORFs in Mx8 and Mx9, it appears that Mx8 and Mx9 are very similar phages.

Because the Mx8 attachment site is located within the Mx8 *int* gene, the Mx9 *int* gene was examined for sequences which



FIG. 1. Physical map of the *int* region of Mx9. The boxes represent putative ORFs. The cross-hatched box in *int* indicates the position of *attP*.

indicate that there is an attachment site. This analysis revealed a DNA segment within the *int* gene (nucleotides 1397 to 1428 [Fig. 2]) that exhibited sequence similarity to tRNA^{Gly} from various organisms. Since Mx8 integrates into the tRNA^{Asp} gene of *M. xanthus*, the sequence that showed similarity to tRNA^{Gly} was predicted to serve as the site of integration for Mx9.

To test this prediction, chromosomal DNA from six integrants containing pKOS35-117.9.7 were cleaved with restriction enzymes, ligated, and transformed into *E. coli* to recover the plasmid along with flanking chromosomal DNA. Sequencing performed with primers adjacent to the proposed attachment site revealed that the point of recombination was indeed that of the putative tRNA^{Gly.} Furthermore, the sequence of flanking chromosomal DNA showed that there were two *attB* sites. It appeared from the number of integrants at each site (three for *attB1* and three for *attB2*) that the two sites served equally well as the insertion site (Fig. 3).

Structure of the two *attB* **sites.** Figure 3 shows 360 bp from each of the *attB* sites. The two sites have a common 42-bp core sequence that is also found in the Mx9 *int* gene. In addition, there are 22 bp 5' to both *attB* sites that are identical at 21 positions. There is a putative inverted repeat that may play a role in integrase protein binding at the *attB* and *attP* sites (Fig. 3B). The site of integration within *attB2* lies in the 5' end of tRNA^{Gly} gene, as shown in Fig. 3B. However, the sequence of *attB1* does not contain a complete tRNA^{Gly} gene. Figure 4 shows the predicted folding of this segment of *attB2* into a corresponding tRNA.

Analysis of the *attR* and *attL* half-sequences for both *attB* sites revealed that the two *attR* sequences are identical, whereas the *attL* sequences differ. This is also the case with the two Mx8 *attB* sites (7). Plasmids containing the Mx8 *int* gene preferentially integrate at *attB1*, and this integration often is accompanied by a deletion between *attB1* and *attB2* (8).

To determine if the identical *attR* sites are due to the presence of two *attB* sites containing identical *attR* sites or due to deletion of the DNA between the two *attB* sites after integration into one of them, PCR analysis was performed either with primers 183-99.4 and 183-99.6 for *attB1* or with primers 183-99.4 and 183-99.5 for *attB2*. A PCR fragment was detected by using the primers specific for *attB2*, but no fragment was detected by using the primers specific for *attB1* (data not shown). This suggests that a deletion may occur upon integration of *attB1*, but to be certain that the lack of a PCR product was not due to a failure to PCR amplify the DNA fragment, further experiments were performed.

Next, the genomic sequence of *M. xanthus* strain DK1622, generated by Monsanto and available at The Institute for Genome Research web site, was examined for the two *attB* sites (www.TIGR.org). The *attB2* sequence was almost identical to the sequence identified previously (Fig. 3B), but only the first 178 bp of the *attB1* site shown in Fig. 3A was present before the sequence diverged. By using this sequence information for *attB1*, a primer was designed that was approximately 100 bp downstream from the point at which the sequence diverged (primer 249-179.7). By using this primer along with 183-99.6, the primer 5' to the *attB1* site, and DZ1 genomic DNA, a PCR product that was approximately 250 bp long was isolated and sequenced. This PCR product was identical to that obtained

VOL.	185.	2003
· OL.	100,	2000

GTGGCC	10 GCTCAG CGAGTC	GGGT CCCA	20 GCGTC CGCAG	GGACO	3 CCAC CGTG	0 TACC ATGC	CAAC GTTG	40 CCCT GGGA	CTCG GAGC	5 ACTTO FGAAO	50 GTGCI CACGʻ	AGTCO	60 CGTCG GCAGC	CCGC CGCG	70 CGGCCC GCCGGG	GCGT	rgcgi ACGC!	B0 ACTCC FGAGG	GTGG CACC	90 GGT CCA	GTC/ CAG'	AGTO TCAO	10(3CG? 2GC#) FCG1 AGCA	rggi ACCA	11 FACC	0 TGCT ACGA	AGGG TCCC	120 CGT GCA
V A	LR	G	A S	D	АТ	т	N	P :	S R	L	V	2 S	v	A A	G P	R	Α	т Р	W	G	v	S	A	S	W	Y	LL	G	R
ACAGCA TGTCG T A	130 AACGGG TTGCCC TGCCC	GGAG CCTC E	140 TACAT ATGTA Y I	CGTG# GCAC1 V	15 GTAG CATC S S	0 CGAC GCTC D	CGCG GCGC A	160 GCGAA CGCT A H	AGAA(TCTT(K K	17 GGGCC CCCGC G	70 CATCO STAGO H I	CAATO GTTAO P M	180 GGCAA CCGTT A	CTGC GACG T A	190 GGCGGA CCGCCI A E	GCGC CGCC R	2) GTTG(CAAC(L	0 CCGAC GCTG P T	GTCA CAGΊ S	210 CCA GGT P	ATC(FAG(I	GACO CTGC D	220 GTCA CAG V) AACG ITGC N	GCTC CGAG A	23 CTGG GACC L	0 CGCT GCGA A L	GGAG CCTC E	240 GTG CAC V
GCCCGC CGGGCC A R	250 GCTTGT CGAACA L V	GGCC CCGG A	260 CTCCA GAGGT L Q	GCAGO CGTCO Q	27 AAAG TTTC Q S	0 TGCC ACGC A	GACG CTGC T	280 CCGCC GGCGC P I	CATCO GTAGO PSS	29 GTCCG CAGGC S	90 3GCC(2CGG(G 1	GCACI CGTG2 R T	300 FTTCG AAAGC F	GCGC CGCG G A	310 GGTGGC CCACCG V A	GGAI CCTI	32 FGAC ACTGI D	20 FGGCT ACCGA WL	CATC GTAG I	330 ACTO TGAO T	GAG(CTC(E	GCCA CGGI A	340 AAGO TTCO K) CGCC GCGG R	TCG AGC L	35 STGT CACA V	0 GCCC CGGG C P	CGAC. GCTG' D	360 AAT TTA N
GAGCGO CTCGCO E R	370 CCGCCA GCCGGT R H	FCTT AGAA L	380 CGCCA GCGGT R H	TATGO ATACO M	39 AGGC TCCG E A	0 GCTC CGAC L	CTGG GACC W	400 GGCA1 CCGT2 G N	IGACO ACTGO 1 T	41 GGATO CCTAO D	10 GTGGA CACC V I	AGCTO FCGAO E L	420 CACGC GTGCG T	CGCG GCGC P R	430 CGTCGT GCAGCA V V	GAAC CTTC K	4 GGCGC CCGCC A	10 CACCT GTGGA H L	GGCG CCGC A	450 GGAG CCTC G	CTT(GAA(L	CTCA GAGI L	460 AAGO TTCC K) CCAG GGTC P	AGG TCC E	47 GGGC CCG G	0 CGCT GCGA P L	GAGC CTCG S	480 GCA CGT A
GCCACO CGGTGO	490 CGTCAA SCAGTT	FAAG ATTC	500 GTGCG CACGC	CTCTA GAGAI	51 CCGG GGCC	0 CAAG GTTC	GCGC CGCG	520 ATCAT TAGT	CAA GTT	53 GGCGG CGCC	30 GCGC2 CGCG3	AAATO PTTAO	540 CAACG GTTGC	GCGA	550 GTGGGG CACCCC	ccce	5 (GGTG2 CAC	50 AATCC' FTAGG	FTTC \AAG	570 GGC0 CCG0	GTG(CAC(CTCG GAGC	580 GACC) CGCG GCGC	AAA TTT	59 AAG TTC	0 AGGC TCCG	GAAG	600 GCC CGG
АТ	VN	к	VR	s	ΤG	к	R	II	ск	A	A (2 I	N	GΕ	W G	P	v	N P	F	G	v	L	D	R	Е	K	ΕA	K	A
GAGCGO CTCGCO E R	610 CCTCAC GGAGTG L T	GCTG CGAC L	620 ACGGC TGCCG T A	AGCGG TCGCC A	63 AGTG TCAC E C	0 CCGG GGCC R	GCG CGC A	640 GTGCT CACGI V I	CCCC AGGGG	65 GCAC1 CGTGA H	50 FTCC(AAGG(F I	GCGCG CGCGC R A	660 GGACC CCTGG D	GGCG CCGC R R	670 CCGCGA GGCGCT R E	GTTI CAAF F	68 CTC: AGAGA	30 FTCCA AAGGT FQ	GGTC CCAG V	690 TTTC AAAC F	CTGO GACO L	GGGC CCCG G	700 CAC GTO P) CGCC SCGG R	CCG GGC P	71 GCG CGC G	0 AAGA TTCT E E	GAAG CTTC K	720 GCG CGC A
CTCCTC GAGGAC L L	730 CAAGGA GTTCCT K E	AGAT FCTA D	740 GTGGA CACCT V D	CGTCG GCAGC V	75 AGGC TCCG E A	0 GCGC CGCC R	CACC STGG T	760 GTCAT CAGT V	TTTC AAAAC F	77 CCGGC GGCCC R	70 CGCAC GCGTC R S	GCAA1 CGTTA 5 N	780 IGGAC ACCTG G	GAGA CTCT R D	790 CACGAC GTGCTG T T	AAAG TTTC K	8 (SACGO TGCO T)0 GACG CTGC G R	CGAG GCTC E	810 CGTC GCAC R	CGCC GCGC R	GTGC CACG V	820 CGG GCC P) GTGC CACG V	CGG GCC P	83 ATG TAC D	0 AGTT TCAA E L	GTGG(CACC(W	840 CCC GGG P
	050				07	•																				0.5			
GTGCTC CACGAC V L	CTCGA GAGCT L D	IGCG. ACGC' A	860 ATGCA TACGT MQ	GGCCA CCGG'I A	87 GTCC CAGG S P	U GTCI CAGA S	GAC CTG D	880 CTCG7 GAGC <i>I</i> L \	TTTT AAAAC 7 F	89 CCCGA GGGCI P	90 AACGO TTGCO N 2	CGAAG GCTTC A K	900 GGTG CCAC G	AGAG TCTC E R	910 GCAGCG CGTCGC Q R	CGCA GCG1 A	92 GAC CTG D	20 ACGAA GCTT TK	GATG CTAC M	930 ACGO TGCO T	CGCC GCGC R	GTGC CACG V	940 TGC ACO L) CGCA CGT R	CTG GAC T	GCGC A	U FATC ATAG L S	CGCGC CGCGC CGCC	GCT CGA A
GTGCTC CACGAC V L GGTGTC CCACAC G V	SGAGCTI GGAGCTI L D 970 CGTGGTC GCACCAC V V	IGCG. ACGC A SGGC SGGC G	860 ATGCA TACGT MQ 980 FGGGA ACCCT WD	GGCCA CCGG'I A TTACA AATG'I Y	B7 CAGG S P 99 TCTG AGAC I C	U GTCI CAGA S 0 CCGC GGCG R	CGAC D D CACG TGC T	880 CTCG7 GAGC# L \ 1000 CAGG6 GTCCC Q 0	TTTTC AAAAC 7 F GCTGC GACC 5 C	89 GGGC1 P 101 CGGC1 GCCGA G	90 AACGO TTGCO N 2 LO TACCO ATGGO Y F	GAAG GCTTC A K 1 GAGAI CTCTA R D	900 GGGTG CCAC G 020 GTGC CACG V	AGAG TCTC E R AGTC TCAG	910 GCAGCG CGTCGC Q R 1030 TGGTGG ACCACC G G	CGCA GCG1 A CGCG GCGC A	GACI CTGI D 104 CGCC CGCC R	20 ACGAAO T K T K 0 CAGGAO TCCTO Q E	GATG M 1 GCGT CGCA R	930 ACGC TGCC T 050 CGGI GCCF R	CGCC R R CGCC C	STGC CACG V 1 CCCG GGGC P	940 TGC ACG L .060 CCI GGA A	GCA CGT R CGCG CGC	CTG GAC T ACA TGT D	95 CGCG A 107 AGC TCG K I	U TATC ATAG L S 0 GCAT CGTA R M	CGCGC GCGCC A 10 GTGGC CACCC W	GCT CGA A 080 GCC CGG A
GTGCTC CACGAC V L GGTGTC CCACAC G V AGTGG7 TCACCZ S G	SOUCCEGA GGAGCT L D 970 CGTGGT GCACCA V V 1090 CCGCCCC AGCGGG R P	IGCG. ACGC' A GGGC' CCCG. G CAAA(GTTT' K	860 ATGCA TACGT M Q 980 IGGGA ACCCT W D 1100 CCCGCG GGGCGG P A	GGCCA CCGGI A ITACA AATGI Y CGTCI GCAGA V	87 GTCC CAGG S P 99 TCTG AGAC I C 111 GGTA CCAT W Y	U GTCI CAGA S 0 CCGC GGCC R 0 CGGG GCCC G	CACG CACG CACG CACG T CACG CTCC CGAG	880 CTCGJ GAGCH L \ 1000 CAGGC GTCCC Q C 1120 CGTCF GCAGJ R E	CTTTC AAAAC F GCTGC GACC GACC C CACC C CACC C C T T	89 CCCGA GGCT P 101 CGGCT G CCCGA G 113 CCCGG GCCGC A	40 4ACGO TTGCO N 4 10 TACCO 4TGGO Y F 30 GCGAO CGCTO A 7	CGAAG GCTTC A K 1 GAGAI CTCTA CTCTA CTCTA CTCTA CACTG GTGAC C L	900 GGTG CCAC G 020 GTGC CACG V .140 GCACA GTGT H	AGAG TCTC E R AGTC TCAG Q S GGAA CCTT R K	910 GCAGCG CGTCGC Q R 1030 TGGTGG ACCACC G G 1150 GGCGGG CCGCCC A G	CGCA GCGI A CGCC GCGC A CTGC GACG C	92 AGACH CTGT D 104 CGCCC R 116 CGCCC R 116 CGACC CTGC D	20 ACGAA(CGCTT) T K AO CAGGA(CTCCT) Q E CGCT(GCGA(P L	GATG M 1 GCGT CGCA R 1 CGTC GCAG V	930 ACGC TGCC T 050 CGGJ GCCF R 170 ATCF TAGJ I	CGCC R FGCC C C ACGC C TCC K	STGC V 1 CCCG SGGC P 1 CTCG SAGC L	940 TGC ACC L .060 CCI :GGA A .180 TGC ACC V	CGCA R CGCGT CGCG CGCC C C TGG ACCC L	GAC T ACA TGT D GGC CCG G	95 GCGC A 107 AGC TCG K 119 AGC ATG TAC H	U TATC ATAG L S GCAT CGTA R M CGGC GCGG A A	CGCGC GCGCC A 10 GTGGC CACCC W 12 CGTCC ACAGC V	GCT CGA A 080 GCC CGG A 200 GAC CTG D
GTGCTC CACGAC V L GGTGTC CCACAC G V AGTGG7 TCACC2 S G ACCACC TGGTCC T T	CTCTGA GAGCTI L D 970 CGTGGTG GCACCA V V 1090 TCGCCCC AGCGGGG R P 1210 GGACGAC GCACGAC CTGCTC D D	rgcg A GggC CCCG G CAAAA K CGTG GCAC V	860 ATGCA TACGT M Q 980 IGGGA ACCCT W D 1100 CCCGC GGGCG GGGCG P A 1220 IACACC ATGTG Y T	GGCCA CCGGI A TTACA AATGI Y CGTCI GCACA CGTCG GCACC CGTGG H	GTCC CAGG S P 99 TCTG AGAC I C 111 GGTA CCAT W Y 123 TCGA AGCT L D	0 GTCT CAGA S 0 CCGGC GGCCC G CCGGG GCCCC G CCGAG GCCCC E	CACC D CACCG CACCG CTGC CGAG L CGAG CCTCC CCTG. D	880 CTCG7 GAGC/ L \ 1000 CAGGC GTCCC Q C 1120 CGTC/ GCAG7 R E 1240 TACTC ATGAC Y C	CACACC CGACC CGACC CGACC CGTGC I T CGCCGC CGGCC CGGCC CGCCGC CGCCGC CGCCGC	89 CCCGA G I101 CGGCI GCCGA G I13 CCCGA G CCCGA I25 CCCCG A	JO AACGO TGCO N I </td <td>CGAAG GCTTC A K 1 GAGAT CTCTA CTCTA STGAC T L 1 CACTG C L 1 CACTG C N</td> <td>900 GGTG G 020 GTGC CACACG V 140 GCACA GTGT H 260 CAAGT TTCA K</td> <td>AGAG TCTC E R AGTC TCAG Q S GGAA CCTT R K TGTC ACAG L S</td> <td>910 GCAGCG QGTCGC Q R 1030 GGCGGG ACCACC G G 1150 GGCGGGG CCGCCC A G 1270 GCTGAA CGACTT L K</td> <td>CGCCA GCGT A CGCCC A CTGC GACCC C CCGG A</td> <td>933GACI PCTGY D 104 CCCCC R 116 CCCCCC CCTGC D 128 CCCCCC P</td> <td>20 ACGAAC T K 10 CAGGAC TCCTC Q E 30 CCGCCC GCGAC P L 30 CCGCCC GCGGC P P</td> <td>GATG TAC M 1 GCGT CGCA R 1 CGCA CGCA CGCA V 1 ACCA CGGT P</td> <td>930 ACGC TGCC T 050 CGGT R 170 ATCZ I 290 CCCTZ GGAI P</td> <td>CGCC R CGCC C C C C C C C C C C C C C C</td> <td>STGC CACG V 1 CCCCG SGGC P 1 1 CCCCG SAGC L 1 CACC STGG H</td> <td>940 TGC GACG L 060 CCT GGA A 180 TGC CACG V 300 CAGG TCC Q</td> <td>) GCGA GCGT R CGCG C 1 CGCG C 1 CTGG GACC C 1 CTGG GACC GCTC G (</td> <td>CTG GAC T ACA TGT D GGC G G GAA CTT G</td> <td>95 CGCC GCCC A CCCC A CTCCG K I 119 ATGC TTCC H I 1310 (GTGZ CCC S S [</td> <td>U TATCO ATAGO CGTAC CGTAC CGTAC CGGC CGGC CGGC CGGC C</td> <td>CGCGC GCGCG A 10 GTGGC CACCC W 12 CGTCC CACCC V 12 CGTCC CACCC C CGCCCG C CGCCCG C G</td> <td>GCT CGA A 080 GCC CGG A 200 GAC CTG D 320 CCT GAC CCT GGA P</td>	CGAAG GCTTC A K 1 GAGAT CTCTA CTCTA STGAC T L 1 CACTG C L 1 CACTG C N	900 GGTG G 020 GTGC CACACG V 140 GCACA GTGT H 260 CAAGT TTCA K	AGAG TCTC E R AGTC TCAG Q S GGAA CCTT R K TGTC ACAG L S	910 GCAGCG QGTCGC Q R 1030 GGCGGG ACCACC G G 1150 GGCGGGG CCGCCC A G 1270 GCTGAA CGACTT L K	CGCCA GCGT A CGCCC A CTGC GACCC C CCGG A	933GACI PCTGY D 104 CCCCC R 116 CCCCCC CCTGC D 128 CCCCCC P	20 ACGAAC T K 10 CAGGAC TCCTC Q E 30 CCGCCC GCGAC P L 30 CCGCCC GCGGC P P	GATG TAC M 1 GCGT CGCA R 1 CGCA CGCA CGCA V 1 ACCA CGGT P	930 ACGC TGCC T 050 CGGT R 170 ATCZ I 290 CCCTZ GGAI P	CGCC R CGCC C C C C C C C C C C C C C C	STGC CACG V 1 CCCCG SGGC P 1 1 CCCCG SAGC L 1 CACC STGG H	940 TGC GACG L 060 CCT GGA A 180 TGC CACG V 300 CAGG TCC Q) GCGA GCGT R CGCG C 1 CGCG C 1 CTGG GACC C 1 CTGG GACC GCTC G (CTG GAC T ACA TGT D GGC G G GAA CTT G	95 CGCC GCCC A CCCC A CTCCG K I 119 ATGC TTCC H I 1310 (GTGZ CCC S S [U TATCO ATAGO CGTAC CGTAC CGTAC CGGC CGGC CGGC CGGC C	CGCGC GCGCG A 10 GTGGC CACCC W 12 CGTCC CACCC V 12 CGTCC CACCC C CGCCCG C CGCCCG C G	GCT CGA A 080 GCC CGG A 200 GAC CTG D 320 CCT GAC CCT GGA P
GTGCTC CACGAC V L GGTGTC CCACAC G V AGTGG7 TCACCA S G ACCACC TGGTGC T T GGCTCA CTGAGT D S	CTCGA GAGCTJ L D 970 CTGGTGTG CACCA V V 1090 CGCGCGC R P 1210 GGACGA CCTGCT D D 1330 AGGACGA CCTGCC G R	rgccg. ACGC' A GGGC' CCCG. G CAAAA STTT' K CGTG' V CAAC. STTG' N	860 ATGCAA TACGT' M Q 980 TGGGA' ACCCT. W D 1100 CCCGCC GGGCGG P A 1220 TACACC ATGTG Y T 1340 ACCTAA TGGAT(T Y	GGCCA CCCGT A TTACA AATGT Y CGTCT GCAGA V CGTCT GCAGA H CCGTCG GCCAC GCCAC G	GTCC (GTCC (CAGG S P 99 TCTG AGACC I C 1111 GGTA (CCAT W Y 123 TCGA AGCT L D 1355 AAGG TTCC E G	0 GTC1 CAGA S 0 CCGCC GCCCC G CCGCC G CCGAC GCCCC C C C	CACC CACGG D CACGG TGC T CGACG CTGC CCTG. D CACC. T TGG T	8800 CTCGT GGAGZ L V 1000 CAGGG GTCCC Q C 1120 CGTCZ GCAGT R F 1240 TACTC ATGAC Y C 1360 ATGCZ TACGT M F	CGG4 CGG4 CGG4 CGG4 CGG4 CGG4 CGG4 CGG4	898 GGCCGAGA GGGCT P 101 CGGCT GCCGA G 113 CGCGG A 125 CGCCG A 125 CGCCG A 137 CGGCC A 137 CGGCC A 137 CGGCC A 137 CGGCC A 137 CGGCC A CGGCC A CGGCC A CGGCC A CGGCC A CGGCC A CGGCC A CGGCC A CGGCC A CGGCC A CGGCC A CGGCC A CGGCC A CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C CGGCCA C C CGGCCA C C CGGCCA C C CGGCCA C C CGGCCA C C CGGCCA C C CGGCCA C C CGGCCA C C CGGCCA C C CGGCCA C C C C	June June <td>CGAAG GGTTC A K 1 EAGAT CTCTA CACTO ETGAC C D 1 TTAAC CAATTO C N 1 ITTAC CAATTO C N 1 ITTAC CAAACO C L L</td> <td>900 GGGTG CCAC G CACG CACG V 140 CCACA CGTGT H 260 CAAGT TTCA K 380 CAGCC GTCG Q</td> <td>AGAG TCTCC E R AGTCC TCAG Q S GGAA CCTT R K TGTCCA ACAG L S ATCA TAGTC H H</td> <td>910 GCAGCG CGTCGC Q R 1030 TGGTGG ACCACC G G 1150 GGCGGG CCGCCC A G 1270 GCTGAA CGACTT L K 1390 CCGGGC GCCCG R A</td> <td>CGCA GCGT A CGCCC GCCCC GACG C CCCGG A GAGA CTCT R</td> <td>GGCCG GGCCG GGCGCG GGCCG GGCCG GGCCG GGCCG GGCCG GGCCG GGCCG GGCCG A A</td> <td>20 CCGAAK GCTT K 10 CAGGACK CAGGACK CCGCTT W E</td> <td>GCTAC M 1 CCGT CCGA R 1 CCGA CCGA CCGA A</td> <td>930 ACGC TGCCT 050 CGGT GCCZ R 170 ATCZ TAGT I 290 CCTF GGAT P 410 CGTG GCAC R</td> <td>CGCC GCGC R CGCC C C C C C C C C C C C C</td> <td>GTGCC CACG V 1 CCCCG GGGCC P 1 CTCG GGGC CTCG GAGC H 1 CACC GTGG H 1 CTAC CACC CTGG H L</td> <td>940 TGC ACC L 0600 CCT GGA A 1800 TGC ACC V 3000 CACG TCC Q 4200 CAA GTT P</td> <td>GGGG R R GGGG GGCG C C C GGAG GGAG C C C C C C C</td> <td>CTG GAC T ACA TGT D GGC CCG G G G AGC TTG C CTT G C TTCG E</td> <td>955 :CGCC :GCCG :GCCG :GCCG :GCCG :GCCG :GCCG :TCCG K I 1119 :TCCG K I 1119 :TCCG K I :TCG :TCCG</td> <td>0 TATCC ATAGG GCATU CGGAT CCGGC CGGC GCCG GCCG GCCG GCCG GC</td> <td>CGCGG GCGCCG A 10 GTGGC CACCC W 12 GGCCC W 12 CGCCCG G G CGCCCG R</td> <td>GCT CGA A 0800 GCC CGG A 2000 GAC CTG GAC CTG GAC CTG GAC CTG GAC CTG SGA P 4400 NAC</td>	CGAAG GGTTC A K 1 EAGAT CTCTA CACTO ETGAC C D 1 TTAAC CAATTO C N 1 ITTAC CAATTO C N 1 ITTAC CAAACO C L L	900 GGGTG CCAC G CACG CACG V 140 CCACA CGTGT H 260 CAAGT TTCA K 380 CAGCC GTCG Q	AGAG TCTCC E R AGTCC TCAG Q S GGAA CCTT R K TGTCCA ACAG L S ATCA TAGTC H H	910 GCAGCG CGTCGC Q R 1030 TGGTGG ACCACC G G 1150 GGCGGG CCGCCC A G 1270 GCTGAA CGACTT L K 1390 CCGGGC GCCCG R A	CGCA GCGT A CGCCC GCCCC GACG C CCCGG A GAGA CTCT R	GGCCG GGCCG GGCGCG GGCCG GGCCG GGCCG GGCCG GGCCG GGCCG GGCCG GGCCG A A	20 CCGAAK GCTT K 10 CAGGACK CAGGACK CCGCTT W E	GCTAC M 1 CCGT CCGA R 1 CCGA CCGA CCGA A	930 ACGC TGCCT 050 CGGT GCCZ R 170 ATCZ TAGT I 290 CCTF GGAT P 410 CGTG GCAC R	CGCC GCGC R CGCC C C C C C C C C C C C C	GTGCC CACG V 1 CCCCG GGGCC P 1 CTCG GGGC CTCG GAGC H 1 CACC GTGG H 1 CTAC CACC CTGG H L	940 TGC ACC L 0600 CCT GGA A 1800 TGC ACC V 3000 CACG TCC Q 4200 CAA GTT P	GGGG R R GGGG GGCG C C C GGAG GGAG C C C C C C C	CTG GAC T ACA TGT D GGC CCG G G G AGC TTG C CTT G C TTCG E	955 :CGCC :GCCG :GCCG :GCCG :GCCG :GCCG :GCCG :TCCG K I 1119 :TCCG K I 1119 :TCCG K I :TCG :TCCG	0 TATCC ATAGG GCATU CGGAT CCGGC CGGC GCCG GCCG GCCG GCCG GC	CGCGG GCGCCG A 10 GTGGC CACCC W 12 GGCCC W 12 CGCCCG G G CGCCCG R	GCT CGA A 0800 GCC CGG A 2000 GAC CTG GAC CTG GAC CTG GAC CTG GAC CTG SGA P 4400 NAC
GTGCTC CACGAC V L GGTGTC CCACAC G V AGTGGT S G ACCACC TGGTGC T T GACTCA CTGAGT D S TTGGCCC AACCGC L A	CCTCGA GAGCT: L D 970 CCTGGTG CCTGCCC V V 1090 CCCCCCCC D D 1330 CCCCCCCC G R 1450 CCCCCCCC G G	IGCG. ACGC' A GGGC' CCCG. G CAAAC STTT' K CGTG' SCAC V V CAAC STTG' N TATAC I	Note of the second seco	GGCCA CCGGT A TTACA AATGT Y CGTCT GCACA CGTCG GCCACA G GCCAC G GCCACA G GCCGCC G GCCGCC G P	SFOC CAGG SP 999 TCTG AGACC IC CONT AGACCAT U 1111 GGTA CCAT W Y 123 TCGA AGAC TCGA TCCA TCCA TCCA TCCA TCCA	0 GTCI CAGA S 0 CCGCC GGCCC G 0 CCGAG GCCCC G 0 CCGAG GCCCC G 0 CCGAG GCCCC G 0 CCGAG GCCCC S	CGAC CACTG D CACG TGC T CGAC C CGAC C CCTG C CACC C CCTG C C C C C C C C C C C C C C C	8800 CTCG7 GAGCI L \ 1000 CAGGC GTCCC Q C 1120 CCGTC2 R F 1240 TACTC ATGAC Y C 1360 ATGC2 X C 1360 ATGC2 K E K E	CGGTT CGCCCC CGCCCC CGCCCC CGCCCCCC	899 GGCCGA GGCT P 101 GGCT GCCGA G 113 GCCGG A 125 GCCGC A 137 VTTGG C CCGCC A 149 GCGGCC A	90 AACGO N 2 I 0 I 0 I 0 I 0 I 0 GCGAC GCGAC GCGAC 0 GGGAGA 1 I 0 I 0 GCTCC 1 GCTCC 1 GCTCC 1 GCTCC 1 GCTCC 1 GCTC 1	CGAAG GCTTC GCTTC A K I BAGAT CTCTA CTCTA CTCTA CTCTA CACTC C CACTC C CACTC C C C	900 GGCGAC G 020 GGTGCCAC CACGG V 140 CCACA H 260 CAGGT H 260 CAGGC GTCGG GTCG GTCG GTCG GTCG GTCG G 500 CTCAGCAC S	AGAG TCTCC E R AGTC TCAG Q S GGAA CCTT R K TGTCC ACAG L S ATCAC H H H H TGAG ACTCC V S	910 GCAGCG CGTCGC Q R 1030 TGGTGG ACCACC G G 1150 GGCGGGC CCGCCC A G 1270 GCTGAA CGGCTT L K 1390 CCGGGCCG R A 1510 CACGGC T A	CGCA GCGT A CGCCC GCCCC GACC C CCCGG A GAGA CTCT R SAAG CTTC K	930 GACA CCTG7 D 104 CCGCC CCGCC CCCCCC D 1228 CCCCC D 228 CCCCC D 1228 CCCCC P 140 CGCCT CCCAA A 1522 GGTGT CCAA A V	20 CGCAA4 CGCAA4 CGCAA4 CGCAA4 CGCAA4 CGCAA4 CGCAA4 CGCCAA CGCCAA CGCCAA CGCCAA CGCCAA CGCCAA CGCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCCAA4 CGCCAA4	GATG TAC M 1 GCGT GCGT CGCA R 1 CGTC CGAC A 1 1 GCTC CGAC A 1 1 CGTC CGAC A 1	9300 ACGCC T GCC7 R 1700 ATC2 R 1700 ATC2 R 2900 CCT7 GGGAT P 4100 CGTG GGCAC R 5300 CCTCCG GAGCC L	CGCC R R CGCC C C C C C C C C C C C C C	Grades V 1 1 1 1 1 1 1 1 1 1 1 1 1	940 TGC GACG L 060 CCT GGA A 180 TGC ACG V 300 CACG V 300 CACG TCC Q 420 CAA GTT P 540 GCG CGC G	GCGT GCGT R GCGC C GCCC GCCC GCCC GCCC G	CTG GAC T TACA TGT TGT GGC G G G G G G G G CTT G C TCG E TGC L I	955 CGCC: GCCC: GCCC: GCCC: TCCG TCCG K I 1119 TACCG TACCG AATGC CACCS S I 1430 TACCG AATGC CACCS S I 1430 TACCG AATGC CACCS CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	0 TATCC ATAG ATAG GCAT CGGCAT CGGC GCCGGC GCCGGC CGCCGC CACCGC CCCGTC GCCACC CCCGTC GCCACC CCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCGTC CCCCCC CCCCC CCCCCC CCCCCC CCCCCC	GGCGG GCGCGC A 110 GGCGC GGCGC GGCGC GGCCC R 12 GGCCC R R 15 GGCCC R R 15 GGCCC R R	GCT GCT A 080 GCC CGG A 200 GAC CTG D 320 CTT GAC F P 440 N 560 STG CAC V
GTGCTC CACGAC V L GGTGTC CCACAA G V AGTGG7 TCACCA S G ACCACC TGGTCC TGGTCC TGGTCC TGGACTCA CTGAGT D S TTGGCCC AACCGG L A	CTCGA GAGCTJ L D 970 CGTGGTG GCACCA V V 1090 CGCGGCG R P 1210 GGCGGGG R P 1210 GGCGGGG CTGCTC D D 1330 GGACGAC CCTGCCG G R 1450 CGGGGGG G G 1570	IGCG. ACGC'A GGGC'GCCG. G CAAAA STTT'K K CGTG'SCAC. V V CAAC. STTG'N N TATAC I	AGC ANGE AND	GGCCA CCGGT A TTACA AATGT Y CGTCT GCAGA V GCACC CGTCG GCCACC G GCCACC G GCCACC G GCCACC G GCCACC G GCCACC G CGCCAC G CGCCAC G CCGCCAC H	STOCK CAGG SP 99 TCTGGAAGAC I C 1111 GGTAA CCCAT W Y 1233 AGCT L D 1355 AAGGC TTCCC TCGA TCCA TCCA TCCA TCCA TC	U GTCT GTCT CAGA S O CCGCC G G CCGCC G CCGAG G G CCGAG G CCGAG G CCGAG G CCGAG G CCGAG CCCC C CCGAG CCCC C CCGAC S CCCCC C CCGCC C C CCGCC C C CCGCC C C CCGCC C C CCGCC C C C C C C C C C C C C C C C C C C C	CGAC CACTG D CACG GTGC T CGAC CTGC CCTG. D CACC. D CACC. T CTGG T CCTGG T V V	8800 GTGGG GAGCI L V 10000 CAGGC GTCCC Q C 11200 CGTCZ GCAGT R F 12400 TACGCA TACGA Y C 13600 ATGCA Y C 13600 ATGCA TACGA K F 1480 AAGGA TTCCTT K E 16000	TTTTC AAAAC 7 F GCTGC GGACC GGACC GGCCG CGTGC CGGCC CGCC CGGCC CGCCC CGCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC	899 CCCCA CCCCA CCCCA P 101 CCCCA G CCCCA C CCCCA C CCCCCA C CCCCCA C CCCCCA C CCCCCA C CCCCCA C CCCCCA C CCCCCA C CCCCA C CCCCA C CCCCA P 101 CCCCA C CCCCA P 101 CCCCA C CCCCA P 101 CCCCA C C CCCCA P 101 CCCCA C C CCCCA C C CCCCA C C CCCCA C C CCCCA C C CCCCA C C CCCCA C C CCCCA C C CCCCA C C C CCCA C C C CCCA C	40 AACGO N 10 PACCO ATGGO ATGGO Y 10 CCCCC G GGAGA CCCCC G CCCCC G CCCCC G CCCCC G CCCCC G CCCCC G CCCCCC G CCCCCC G CCCCCC G CCCCCC G CCCCCC GAACD CCCCCC G CCCCCC G CCCCCC CCCCCC CCCCCC CCCCCC CCCCCC CCCCCC CCCCCC CCCCCC CCCCCCCCC CCCCCCC CCCCCC CCCCCCCCCC CCCCCCCCCCCCCCCCCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CGAAG GGTTC GGTTC A K 1 EAGAT CTCTA C D CACTO CTGAC C CACTO CTGAC C CACTO CACT	900 GGGTG GGGTG CCAC GCACG CACGG V V 140 CCACA CGTGT H 260 CAAGT K 380 CCACG CGTCG Q 500 TTCAC GTCG S C	AGAG TCTCC E R AGTC TCAG Q S GGAA CCTTF R K TGTCC ACAG L S ATCAC TAGTC H H TGAGG ACTCC V S	910 GCAGCG CGTCGC Q R 1030 TGGTGG ACCACC G G 1150 GGCGGG CCGCCC A G 1150 GCCGGCC A G 1270 SCTGAA CGACTT L K 1390 CCGGGC GGCCCG R A 1510 CACGGCCG T A 1510 CACGGCCG T A 1630	CGC2 GCG1 A CGCC GGCC C CGACG C C CCCGG A GAGA C TCT R SAAG C TTC K	Gaca Gaca Gaca CCTG7 D 104 CGCC CCTGC CCCC CCCCC CCCCC CCCCCC CCCCCC CCCCCC	10 CGGAAG SCCTT T K 10 CGGCAG CGCTT Q E 10 CGCCAG P L 10 CGCCAG P P 0 0 CGCCAG P P 0 0 CGCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCAG 0 CCCCCCAG 0 CCCCCCAG 0 CCCCCCAG 0 CCCCCCAG 0 CCCCCCAG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCG 0 CCCCCCCCCGCCG 0 CCCCCCCCCC	GATG TAC M 1 1 GCGT GCAG V 1 1 CGTC CCAG V 1 1 CGTC CCAG P 1 1 CCTC A 1 1 CGTC CCAG CCAG CCAG CCAG CCAG CCAG CCAG	9300 ACGC T CGCT 0500 CCGGT R 1700 ATC2 T ATC2 T ATC2 T ATC2 F 4100 CCT2 GGAT P 4100 CCGT2 GGAC R S5300 CTCCG GAGCC L	CGCC GCGC R CGCCGC C C C C C C C C C C C	Trace	940 TGC ACG L 060 CCT CGA A 180 TGC CACG V 300 CACG V 300 CACG CCA CCA CCA CCA CCA CCA CCA CCA CC	GGCA GGCA CGCG CCGC CCGC CCGC CCGC GGCC CCGCC CCGCC	CTG GAC T TGACA TGT CCG G G G G G G G CCG G G G A G CTT G CCG CCG G CCG CCG CCG CCG CCG C	955 CGCC: CGCC: CGCC: CGCC: CGCC: A TGCG K I 1119 A TGCG K I 119 A TGCG K I 1310 GTGZ S I 1310 GTGZ S I 1430 L I 1550 CTACG CT	U TATC TATC: ATAG CGCAT CGTAT CGGCCG CGGCCG CGCCGG CGCCGG CGCCGG CGCCGC	CGCGC GCGCCA A 11 GTGGC CACCC W 12 GTGCC V 13 GGCCC C CGGA CGGA CGGA CGGA CGGA CGGA	GCTA A 0800 GCCC CCGG A 2000 GACC CTG D 3200 CCTG D 3200 CCTG P 4400 STG CAC V V

FIG. 2. Nucleotide sequence of the Mx9 *int* gene and the deduced amino acid sequence. The amino acids are indicated by single letters under the DNA sequence. Stop codons are indicated by an asterisk. The sequence in **boldface** type is the Mx9 *attP* sequence. The arrows indicate inverted repeats.

from the DK1622 genomic sequence (Fig. 3C). Analysis of the sequence revealed that only 16 bp of the 42-bp core *att* site was present in the native *attB1* site.

Final proof that a deletion does occur between attB1 and attB2 is shown in Fig. 5. By using primers 183-99.4 and 183-99.5, the primers that amplify the attB2 site, PCR amplification

was performed with genomic DNA from the wild-type strain or strains harboring a plasmid integrated at either *attB1* or *attB2*. By using chromosomal DNA from DZ1, a strain with no plasmid integrated at either *attB* site, a 372-bp PCR product containing the *attB2* site was detected (Fig. 5, lane 2). Two strains that had insertions at *attB2* (Fig. 5, lanes 5 and 6) did not Α.

GTGAGC CACTCG	10 TGACCTCAAC ACTGGAGTTG	20 GGTTTGTTGG CCAAACAACC	30 GTGGGGAGCO CACCCCTCGO	40 GGGACAGCGGA CCCTGTCGCCT	50 CCACATGGTG GGTGTACCAC	60 CCAGGGCTTA GGTCCCGAAI	70 ACGGCTTCGC# IGCCGAAGCG1	80 ACACGGGGCTG IGIGCCCCGAC	90 GGCGATGCTG CCGCTACGAC	100 AACGGAGCGI TTGCCTCGCA	110 CCCATGTCCA AGGGTACAGGT	120 CGCG GCGC
ATGCCG TACGGC	130 CCTGGCTTGC GGACCGAACG	140 ACATAGGGAT TGTATCCCTA	150 TCGAAACCTC AGCTTTGGAC	160 CGACCCC GAGC CTGGGGC TCG	170 TTGGGAAGCT AACCCTTCGA	180 CGTGCTCTAC GCACGAGATG	190 CAACTGAGCI GTTGACTCGA	200 IACCACCGCAG ATGGTGGCGTC	210 GCGAAGCAGG CGCTTCGTCC	220 GCGCAAAGTA CGCGTTTCAT	230 CGGGCCGCCC GCCCGGCGGG	240 FGTG ACAC
GCTTGT CGAACA	250 CAACGGGAAG GTTGCCCTTC	260 TGAGGTGCTA ACTCCACGAT	270 CTCCGTCTCC GAGGCAGAGO	280 STCGACGGTGA GAGCTGCCACT	290 GCTGGTACGA CGACCATGCI	300 GTCCTGGAAG CAGGACCTTC	310 STTGGACTCGC CAACCTGAGCC	320 CGGTTGCGCGC GCCAACGCGCG	330 GTCCCGGACC CAGGGCCTGG	340 TCGAAGAGGT AGCTTCTCCA	350 AGACGCCTGG ATCTGCGGACC	360 CTCG GAGC
в.												
CGAGCC GCTCGG	10 GGGGACGGGA CCCCTGCCCT	20 GCGGCGGGGAC CGCCGCCCTG	30 CGGCTTCGCG GCCGAAGCGC	40 SCCGTTTACAG SGGCAAATGTC	50 CATCCTTGCT GTAGGAACGA	60 GCAAGACGCC CGTTCTGCGG	70 CCGAGGCCCG GGCTCCGGGC	80 Saaaagacgaa Ctttttctgctt	90 GGCCGGCAGT CCGGCCGTCA	100 CCCGAGTTTC GGGCTCAAAG	110 CTCAAGGACT GAGTTCCTGA	120 ACCG TGGC
GCCTTC CGGAAG	130 ATGGGT <u>GAGC</u> TACCCACTCG	140 GGCGGAAGGG CCGCCTTCCC	150 <u>ATTCGAACCC</u> TAAGCTTGGG	160 <u>TCGACCCCGA</u> GAGCTGGGGG CT	170 GCTTGGGAAG CGAACCCTTC	180 CTCGTGCTCT GAGCACGAGA	190 ACCAACTGAG	200 CTACCACCGC CGATGGTGGCG	210 AGGCGAAGCA TCCGCTTCGT	220 GGGCGCAAAG CCCGCGTTTC	230 TACGGGCCGCG ATGCCCGGCG	240 CCTG GGAC
TGGCTT ACCGAA	250 GTCAACGGGA CAGTTGCCCT	260 AGTGAGGTGC TCACTCCACG	270 TACTCCGTCT ATGAGGCAGF	280 CCTCGACGGT GGAGCTGCCA	290 GAGCTGGTAC CTCGACCATG	300 GAGTCCTGGA CTCAGGACCT	310 AGTTGGACTC TCAACCTGAG	320 CGCGGTTGCGC CGCCAACGCG	330 GCGTCCCGGA CGCAGGGCCT	340 CCTCGAAGAG GGAGCTTCTC	350 GTAGACGCCT CATCTGCGGA	360 GGCT CCGA
c.												
TGCCAG ACGGTC	10 GGCTTACGGC CCGAATGCCG	20 TTCGCACACG AAGCGTGTGC	30 GGGCTGGGCG CCCGACCCGC	40 SATGCTGAACG STACGACTTGC	50 GAGCGTCCCA CTCGCAGGGT	60 TGTCCACGCG ACAGGTGCGC	70 SATGCCGCCTG TACGGCGGAC	80 GCTTGCACAT CCGAACGTGTA	90 AGGGATTCGA TCCCTAAGCT	100 AACCTCGACC TTGGAGCTGG	110 CCGAGCTTGG6 GGCTCGAACCO	120 GAAG CTTC
CTCGGC GAGCCG	130 CTCGACCCGT GAGCTGGGCA	140 CCAGGCGTTA GGTCCGCAAT	150 TCAGCCGTTC AGTCGGCAAG	160 GCAAACCCTT CGTTTGGGAA	170 ACTTCGCCTT TGAAGCGGAA	180 GGGGGATTCCG CCCCTAAGGC		200 CCTGTCCATCC GACAGGTAGG	210 GTCGCAGCGG CAGCGTCGCC	220 GTAGCAGGGA CATCGTCCCT	230 GTCTCAGGGG CAGAGTCCCC	240 GGTT CCAA

FIG. 3. (A) Nucleotide sequence of the reconstituted Mx9 attB1 site. (B) Nucleotide sequence of the Mx9 attB2 site. The arrows indicate an inverted repeat in attB2. (C) Nucleotide sequence of the native Mx9 attB1. Boldface type indicates the core att site. The underlined nucleotides encode tRNA^{Gly}.

produce the 372-bp band and should not have amplified *attB2* due to the presence of a plasmid integrated at that site. If a deletion did occur between *attB1* and *attB2*, there should have been no detectable amplification of *attB2* when a plasmid integrated at *attB1*. The results showed that no *attB2* PCR product was detected, indicating that there was a deletion of DNA between *attB1* and *attB2* when integration occurred at *attB1* (Fig. 5, lanes 3 and 4).

Integration results in alteration of the carboxy terminus of the Mx9 Int protein. Because *attP* lies within the *int* gene, integration into the chromosome alters the 3' end of the *int* gene. In the 1,160 bp of *attR* that has been sequenced, no stop codon has been identified (data not shown). Thus, 70 amino acids should be removed from Int, and more than 389 amino acids should be added to the Int protein that is synthesized after integration into the chromosome. These additional amino acids presumably reduce the enzymatic activity of Int because the IntX protein of Mx8 has lost 112 residues and contains 13 added amino acids and is less active in site-specific recombination (8).

Mx9 Int is the only phage protein required for integration. To determine whether *int* is necessary and sufficient for integration, the *int* gene was PCR amplified and ligated into an *E*. *coli* expression vector that uses an engineered phage T7 A1 promoter. When plasmid pKOS249-31 was electroporated into DZ1, it integrated efficiently into the chromosome; approximately 1×10^4 colonies per µg of DNA were obtained. Thus, the Mx9 *int* gene is the only phage-encoded protein required for integrative recombination into the bacterial chromosome.

Transcription from the *pilA* and the *mgl* promoters integrated at the two Mx9 attB sites. Because the interest in Mx9 integration was to find a phage attachment site on the M. xanthus chromosome that supports efficient expression of genes from a variety of promoters, fusions of lacZ to the mgl or pilA promoters were constructed, and transcription from these promoters at the two Mx9 attB locations, the Mx8 attB location, and the native chromosomal location was analyzed. Figure 6A shows the levels of expression of the *pilA* promoter (P_{pilA}) at the four different locations. Surprisingly, there was little transcription when the P_{pilA} plasmid was integrated by homologous recombination at the pilA location (pKOS178-86). This suggests that there may be a deletion in the *pilA* promoter region that eliminates activation of the *pilA* promoter in DZ1 since there was no expression in several isolates that were examined. As observed previously, little transcription from P_{pilA} is seen when it is integrated at the Mx8 attB site



FIG. 4. Predicted cloverleaf secondary structure for $tRNA^{Gly}$ from *M. xanthus*. The bases that are within the core *attB* sequence are outlined.

(pKOS178-86 plus pKOS139-29). However, the Mx9 sites showed high levels of transcription from P_{piL4} (pKOS178-177), and the levels were fairly similar at the two sites, although *attB2* had high variability of expression in the two isolates examined. In addition, the regulation at the two sites was similar; transcription from P_{piL4} increased during the late log and stationary phases.

The results of transcription from the *mgl* promoter (P_{mgl}) are shown in Fig. 6B. Transcription from P_{mgl} at the two Mx9 *attB* (pKOS178-188) sites was better than transcription at the Mx8 site (pKOS139-47 plus pKOS139-29), but it was not as good when the promoter was integrated by homologous recombination at the chromosomal *mgl* location (pKOS139-47). However, the lower level of expression at the two Mx9 sites may have been vector dependent. When a plasmid that contained only the *attP* site was used and integrated by supplying



FIG. 5. Agarose gel containing PCR-amplified DNA fragments. Lane 1, 100-bp ladder from New England Biolabs; lane 2, PCR amplification for detection of *attB2* in wild-type strain DZ1; lanes 3 and 4, PCR amplification for detection of *attB2* in two independent isolates that contain a plasmid integrated at *attB1*; lanes 5 and 6, PCR amplification for detection of *attB2* in two independent isolates that contain a plasmid integrated at *attB1*; lanes 5 and 6, PCR amplification for detection of *attB2* in two independent isolates that contain a plasmid integrated at *attB2*.



FIG. 6. (A) *lacZ* gene transcribed from the *pilA* promoter integrated at either the *pilA* chromosomal location, Mx9 *attB1* or *attB2*, or the Mx8 *attB* sites. (B) *lacZ* gene transcribed from the *mgl* promoter integrated at either the *mgl* chromosomal location, Mx9 *attB1* or *attB2*, or the Mx8 *attB* sites.

the *int* gene in *trans*, P_{mgl} functioned just as well at both Mx9 sites as it did at the chromosomal *mgl* location (data not shown). The data from these studies indicate that the Mx9 *attB* sites are good sites for expression of foreign or native genes.

DISCUSSION

The Mx9 *int* gene and attachment site were identified, along with the site of integration into the *M. xanthus* chromosome. The analysis revealed remarkable similarity to the *int* gene and attachment site in the myxophage Mx8 (7, 8, 11). Both phages contain *attP* within the *int* gene and integrate within a tRNA gene. They both have two *attB* sites, and it appears that adjacent chromosomal DNA is deleted when integration occurs at one of the sites. For both, Int is the only phage-encoded protein needed for integration.

One difference between the Mx8 and Mx9 phage integration systems is the length of the core sequences. The core sequence for Mx8 integration is smaller, composed of 29 bp. The *attB2* site has two nucleotides that differ at one end, which may account for the preference of Mx8 for inserting at attB1. The att core region of Mx9 is 42 bp long, but only one of the two integration sites, attB2, contains all 42 bases. The attB1 site contains only 16 bases of the core sequence. The lack of a complete core sequence in *attB1* may explain why there is always a deletion between attB1 and attB2 when integration occurs at attB1. The Int protein may bind to the inverted repeat within the 42-bp core. Binding of the λ Int protein to its att sites has been demonstrated (5). Since attB1 contains one-half of the inverted repeat, only one-half of the necessary protein complex can form; however, once it has assembled, it may interact with the complementary half of proteins formed from attB2 to allow integration. This should result in a looping out of the DNA between attB1 and attB2 and its subsequent loss upon integration of DNA.

In PCRs to detect *attB1* with primers 183-99.4 and 183-99.6, the conditions were such that if the distance between *attB1* and *attB2* was less than 2 kb, then a PCR product should have been detected. Since no product was observed, the results suggest that the distance between the two sites is greater than 2 kb. Analysis of the DK1622 sequence showed that the two *attB* sites are 6.7 kb apart. Analysis of this sequence revealed two ORFs that have sequence similarity to transposase genes, suggesting the presence of a transposon. The product of another ORF that was identified exhibited high levels of sequence similarity to proteins whose functions are unknown. Clearly, the ORFs between the two *attB* sites are not critical for growth under laboratory conditions since strains with integrations at *attB1* have no visible growth defects.

ACKNOWLEDGMENT

I thank Monsanto for use of the M. xanthus sequence.

REFERENCES

- Campos, J. M., and D. R. Zusman. 1975. Regulation of development in Myxococcus xanthus: effect of 3':5'-cyclic AMP, ADP, and nutrition. Proc. Natl. Acad. Sci. USA 72:518–522.
- Fisseha, M., M. Gloudemans, R. E. Gill, and L. Kroos. 1996. Characterization of the regulatory region of a cell interaction-dependent gene in *Myxococcus xanthus*. J. Bacteriol. 178:2539–2550.
- Julien, B., and R. Calendar. 1995. Purification and characterization of the bacteriophage P4 delta protein. J. Bacteriol. 177:3743–3751.
- Julien, B., and S. Shah. 2002. Heterologous expression of the epothilone biosynthetic genes in *Myxococcus xanthus*. Antimicrob. Agents Chemother. 46:2772–2778.
- Landy, A. 1989. Dynamic, structural, and regulatory aspects of lambda sitespecific recombination. Annu. Rev. Biochem. 58:913–949.
- Li, S. F., and L. J. Shimkets. 1988. Site-specific integration and expression of a developmental promoter in *Myxococcus xanthus*. J. Bacteriol. 170:5552– 5556.
- Magrini, V., C. Creighton, and P. Youderian. 1999. Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: genetic elements required for integration. J. Bacteriol. 181:4050–4061.
- Magrini, V., M. L. Storms, and P. Youderian. 1999. Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: regulation of integrase activity by reversible, covalent modification. J. Bacteriol. 181:4062– 4070.
- Martin, S., E. Sodergren, T. Masuda, and D. Kaiser. 1978. Systematic isolation of transducing phages for *Myxococcus xanthus*. Virology 88:44–53.
- Miller, J. H. 1992. A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Tojo, N., K. Sanmiya, H. Sugawara, S. Inouye, and T. Komano. 1996. Integration of bacteriophage Mx8 into the *Myxococcus xanthus* chromosome causes a structural alteration at the C-terminal region of the IntP protein. J. Bacteriol. 178:4004–4011.
- Zusman, D. R., D. M. Krotoski, and M. Cumsky. 1978. Chromosome replication in *Myxococcus xanthus*. J. Bacteriol. 133:122–129.