# Molecular Analysis of the Escherichia coli phoP-phoQ Operon

MEGUMI KASAHARA, ATSUO NAKATA, AND HIDEO SHINAGAWA\*

Department of Experimental Chemotherapy, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565, Japan

Received 27 August 1991/Accepted 31 October 1991

The *phoP-phoQ* operon of *Salmonella typhimurium* is a member of the family of two-component regulatory systems and controls expression of the *phoN* gene that codes for nonspecific acid phosphatase and the genes involved in the pathogenicity of the bacterium. The *phoP-phoQ* operon of *Escherichia coli* was cloned on a plasmid vector by complementation of a *phoP* mutant, and the 4.1-kb nucleotide sequence, which includes the *phoP-phoQ* operon and its flanking regions, was determined. The *phoP-phoQ* operon was mapped at 25 min on the standard *E. coli* linkage map by hybridization with the Kohara mini set library of the *E. coli* chromosome (Y. Kohara, K. Akiyama, and K. Isono, Cell 50:495–508, 1987). The predicted *phoP* and *phoQ* gene products consist of 223 and 486 amino acids with estimated molecular masses of 25,534 and 55,297 Da, respectively, which correspond well with the sizes of the PhoP and PhoQ proteins identified by the maxicell method. The amino acid sequences of PhoP and PhoQ of *E. coli* were 93 and 86% identical, respectively, to those of *S. typhimurium*.

The phoP and phoQ genes of Salmonella typhimurium are members of the family of two-component regulatory systems; PhoP is a regulator protein and PhoQ is a sensor protein (17), and they regulate expression of the Salmonella virulence gene (pagC) (20) and the nonspecific acid phosphatase gene (phoN) (11). The phoN gene is induced by carbon, phosphate, nitrogen, and sulfur limitation (10), as well as by a low pH (17). The phoP and phoQ gene products have been shown to be involved in the transcriptional regulation of a number of the phoP-activated genes (pag) and the phoP-repressed genes (prg) (18). The phoP-phoQ operon has been mapped at 25 min on the Salmonella linkage map (11), and its nucleotide sequence has been reported previously (6, 17).

Escherichia coli has no phoN gene which is regulated by the phoP-phoQ operon (8), but a DNA sequence similar to that of the phoP gene was detected in E. coli (8) and several gram-negative species (6). Further, the phoN gene of S. typhimurium, when introduced into E. coli, is induced by phosphate starvation, as in S. typhimurium (8). These results suggest that functional phoP and phoQ genes exist in E. coli. We adopted E. coli, which is genetically better understood than S. typhimurium, for the study of mechanisms of signal transduction and physiological roles of the phoP-phoQ regulon. As a step toward these goals, we have cloned and determined the nucleotide sequence of the phoP-phoQ operon of E. coli.

## **MATERIALS AND METHODS**

**Bacterial strains.** E. coli K-12 and S. typhimurium strains used in this study are listed in Table 1.

Media. LB broth, LB agar, T broth, and T agar were as described by Miller (16). The media used for the routine preparation of M13 phage and for the maxicell method were as previously described (1). Ampicillin was added to LB broth or T broth at 100  $\mu$ g/ml. Chloramphenicol was added to LB agar at 20  $\mu$ g/ml. So that synthesis of acid phosphatase

by bacterial colonies could be examined, 5-bromo-4-chloro-3-indolyl phosphate (X-P) was added to agar plates to a final concentration of 40  $\mu$ g/ml.

**Recombinant DNA methods.** Standard methods for recombinant DNA manipulation as described by Sambrook et al. (22) were generally used.

**DNA sequencing.** The manipulation of M13mp18 phage (27) was as described by Messing et al. (15). A series of M13 phage derivatives with cloned DNA fragments with one end fixed and the other end formed by successive deletions at the 5' end was prepared as described by Henikoff (7). The DNA sequences were analyzed by the method of Sanger et al. (24).

Assay of chloramphenicol acetyltransferase. Chloramphenicol acetyltransferase in the extracts of sonicated cells was assayed by the method of Shaw (25). The specific activity was expressed as nanomoles of 5-thio-2-nitrobenzoate liberated per minute per unit of optical density of the cell culture at 450 nm.

Identification of the protein encoded by plasmids. The *phoP* and *phoQ* gene products were identified by the maxicell method, which allows isotope labeling of protein encoded by plasmids (23).

Southern hybridization. The DNA fragments of S. typhimurium and E. coli used as phoP probes were radiolabeled by the random primer method (4) and used to examine the presence of similar DNA sequences in the chromosomes by DNA-DNA hybridization as described by Southern (26).

**Enzymes and radioisotopes.** The restriction endonucleases, T4 DNA ligase, T4 DNA polymerase, Klenow fragment of DNA polymerase I, exonuclease III, mung bean nuclease, bacterial alkaline phosphatase, a random primer DNA labeling kit, and an M13 sequencing kit were obtained from Takara Shuzo (Kyoto, Japan). [ $\alpha$ -<sup>32</sup>P]dCTP (>400 Ci/mmol) and [<sup>35</sup>S]methionine were purchased from Amersham Japan (Tokyo, Japan).

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence data bases under the accession number D90393.

<sup>\*</sup> Corresponding author.

TABLE 1. Bacterial strains used in the study

Strain	Characteristic(s)	Reference or source
E. coli K-12		· · · · · · · · · · · · · · · · · · ·
ANCK10	F <sup>-</sup> leu lacY trp his argG rpsL ilv metA (or metB) thi	14
CSR603	recA1 uvrA6 phr-1	23
JC7623	thr-1 leu-6 thi-1 lacY1 galK2 ara-14 xyl-5 mtl-1 proA2 his-4 argE3 rspL-31 tsx-33 sup-37 recB21 recC22 sbcB15	19
JM103	Δ(pro-lac) supE thi (F' traD36 proAB lacI <sup>a</sup> Z M15)	15
S. typhimurium		
SJ10002	hsdR	K. Kutsukake
TA2362	phoP12	11

## **RESULTS AND DISCUSSION**

Cloning of the E. coli phoP gene and mapping of the gene on the chromosome. We cloned the E. coli phoP gene by functional complementation of an S. typhimurium phoP mutant with recombinant plasmids carrying E. coli chromosomal fragments. The gene library of E. coli K-12 strain ANCK10 was constructed by digesting the chromosomal DNA with EcoRV enzyme and ligating it with EcoRVdigested pBR322 (2). The gene library was amplified in an S. typhimurium hsdR strain, SJ10002. Recombinant plasmids that transformed white, X-P<sup>-</sup> colonies of TA2362 (phoP12) into blue colonies were selected. One such plasmid, pMK200, contained a 4.1-kb EcoRV fragment. The physical map of this fragment is shown in Fig. 1B, and it hybridized with the DNA fragment containing the phoP gene of S. typhimurium (data not shown).

To identify the location of the phoP gene on the E. coli

chromosome, hybridization analysis was performed with Kohara clones (12) E3G11, 14C1, E9G1, 15A8, E4C2, 7F9, 20E6, 3E11, 4D1, 2A3, and 7C10 of the *E. coli* genome library, which contain the chromosomal regions around 25 min, which is the locus of the *phoP* gene on the *S. typhimurium* linkage map. The 1.8-kb *Eco*RV-*ScaI* fragment of pMK200 and the 514-bp *Eco*RV fragment internal to the *phoP* gene of *S. typhimurium* were used as probes. Both probes hybridized with clone 7F9 (data not shown). These results suggest that the *phoP* gene is carried on pMK200 and located at 25 min on the *E. coli* linkage map as detailed in Fig. 1.

Nucleotide sequence of the phoP locus and identification of the phoP and phoQ genes. We have determined the DNA sequence of the 4.1-kb EcoRV chromosomal fragment carried on pMK200 which contains the E. coli phoP gene and its flanking regions. The entire nucleotide sequence is shown in Fig. 2. Two complete open reading frames (ORFs) whose deduced sequences are very similar to those of the S. typhimurium PhoP and PhoO proteins (17) are identified, and they appear to constitute an operon, as do the S. typhimurium phoP and phoQ genes. Therefore, we conclude that these ORFs are coding regions of the E. coli phoP and phoQ genes. The phoP coding region is located between nucleotides 1051 and 1719, begins with an ATG codon, ends with a TGA codon, and is preceded by a Shine-Dalgarno sequence, GGAG, at an appropriate position. The phoQ coding region is located between nucleotides 1722 and 3179, begins with an ATG codon which overlaps the stop codon of phoP by 1 nucleotide, and ends with a TAA codon. The putative -10sequence, cATAAT, is at nucleotides 1003 to 1008, and 12 nucleotides upstream of the -10 region, a direct repeat of hexanucleotide sequence GTTTA(T or C) is found, as with S. typhimurium (6). No sequence substantially similar to the consensus -35 sequence is found in this region. These regions were highly conserved between E. coli and S. typhimurium, while the other parts of the upstream region were not conserved. Functionally important sequences in

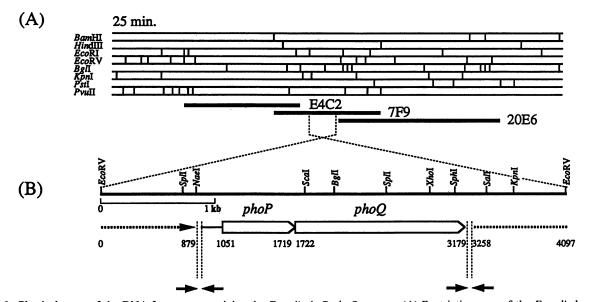


FIG. 1. Physical maps of the DNA fragment containing the *E. coli phoP-phoQ* operon. (A) Restriction map of the *E. coli* chromosome segment that contains the *phoP-phoQ* region carried by  $\lambda$  clones E4C2, 7F9, and 20E6, as described Kohara et al. (12). (B) Enlarged restriction map of the *phoP-phoQ* region. Arrows indicate the inverted repeats.

			GCT					CCZ			GCCI		CACO				CGG			GATG						TAT		90 AGCGC
	I	P		L 00	s	R	т 110		G	Q 12		A		Р 30	S	г I 140		ĸ	E 150		Al	4 V 160		Y	R 170		E	R 180
			CCG	CCF			ACCZ	GG:		GAT	CCT		CAAJ	AAT			GGT		TAAC	TAT		SCCC	ACA'		CCGC	TTA		E
	~			90		~~~~	200			21		~ ~ ~ ~	22			230		~~~	240			250			260			270
																I Q												ACTAC Y
	AT!	rgc		80 AC1	GTT	TGP	290 \TTG		TGC	30) GCG		CAAC		10 FATT	ICTG.	320 ATCGA		TGA	330 CCG1		GTC	340 IGGG		ATA	350 rCGC		TA	360 ACCAC
	I	A			F	D			A			N			L	I D		D			V I		Y	I			N	
			ACĂ					TGC			TGG			CACO		410 CCGCA PH	TAA			CCG						CGA		450 GAAT N
	•	n	-	60	•	•	470		2	480		Ũ	49		•••	500		•	510			์ 520		.,	530		Ŭ	540
																CCGGI ? V												CGCTG L
	ccr	בבח		50		000	560 TAT		2ጥጥ እ	570		ገልሞባ	58		יראי	590		<b>633</b> )	600 2008		AGCI	610		AAGI	620		TGI	630 CCAT
																S T												
			GGĂ					CAZ			AGTO			GA/			GAC			CGT			GCAI			ACC		720 CGAG
	L	L		Е 30	L	D	н 740		W	E 75(		L	а 76		P	ι Q 770		v	M 780		RY	G 790		Е	к 800		Y	Е 810
			GAA	AGA			TCG	CGG		GCG	CGTI		CGCC	GA			GCA		TATC	GAT		TGG	CGTT		CAGA			GAAA
								~~																				
				82				30			40			350		86			87			880			890			900
				CTG	AAA		ATG.	ACG		CTA	ACT		TTGO	STCO		86 TATCA I T	CGA		'TGA'	TGAG		AAAT				TCAG	бт <u>с</u>	
		A	R	CTG L 91	AAA K .0	A	ATG M 9	ACG T 20	PA	CTA N 9	ACT Y 30	I	TTGO G	GTCG R 940	A	TATCA	CGA: M	v	TGA D 96	TGAG E 0	LI	алат: < * 970	AAAC	стс	GTA 980		-	990
		A	R GCG	CTG L 91 ATG	AAA K 0 CTG	A	ATG M 9 GGC	ACG T 20 CTG	PA	SCTA N 9 ATTA	ACT Y 30 AGA	I	TTGO G FCCO	GTCG R 940 GCT1	A	TATCA I T 95 ATTTT	CGAT M 0 TTC	V ACTT	TGA D 960 TAC	TGAG E 0 CTCC <b>Ph</b>	l i CCTO <b>OP</b>	4AAT 570 50000	AAAC GCTG	GTT	980		AT G	990
F		A ATG	R GCG	CTG 1 91 ATG 100	AAA K O CTG	A TCC CAT	ATG M GGC 10: AAT	ACG T 20 CTG 10	P A CTTA	CTA 9 10:	ACT Y 30 AGA	I TTA:	TTGG G FCCG 10	GTCG R 940 GCTT )30	A TTTI	I T 95 ATTTT 104 ACAGG	CGAT M 0 TTC/ GAG/	V ACTT	TGA: D 960 TAC	TGAG E 0 CTCC <b>ph</b> 0 AATG	L I CCT( <b>oP</b> 1 CGC(	4AAT: 970 CCCC .060 STAC	AAAC GCTG FGGT	CTC	980 TAT 070	TTA/	ATG	990 990 7777A 1080 FGCG
Ē. S.		A ATG	R GCG	CTG 1 91 ATG 100	AAA K O CTG	A TCC	ATG M GGC 10: AAT	ACG T 20 CTG 10	P A CTTA	CTA 9 10:	ACT Y 30 AGA	I TTA:	TTGG G FCCG 10	GTCG R 940 GCTT )30	A TTTI	I T 95 ATTTT 104 ACAGG	CGAT M 0 TTC/	V ACTT	21050 2007 21050 21000 21050 2100 210	IGAG E CTCC <b>ph</b> AATG M	L I CCT( <b>oP</b> 1 CGC(	970 CCCC	AAAC GCTG FGGT	CTC	980 TAT 070	TTA/	ATG	990 990 7777A 1080 FGCG
		A ATG	R GCG CCA	CTG 1 91 ATG 100 TAA	AAA K CTG CTG 0 CCA	а <u>тсс</u> <u>сат</u> –1	ATG. 9: GGC 10: AAT 0 110	ACG T 20 CTG 10 CGC	P A CTTA GTTA	CTA 9 TTA 10: CAC	ACT 30 AGA 20 TAT 10	I TTA: TTT	ITGO G ICCO 10 AAT 11	GTCG R 940 GCTT )30 AATT	A TTTT?	I T 95 ATTTT 104 ACAGG 113	CGAT M 0 TTCJ 0 GAGJ SD	V ACTT AAAT	TGA D 960 TAC 1050 AAA M 1140	IGAG E CTCC <b>ph</b> AATG M - (1)	L P CCTC <b>oP</b> CGCC R V 	970 CCCCC 060 TAC 1 L .150	AAAC GCTG IGGI V -	CTC	980 TAT 070 TGA E - 160	rtaj Agao D -	ATG CAA N -	990 7777A 1080 7GCG A -
s.	- - -	A ATG CCC	R GCG CCA	CTG L 91 ATG 100 TAA	O CTG	A TCC CAT -1 CAC	ATG. M 9: GGC 10: AAT 0 11( CTT)	ACG T 20 CTG 10 CGC	P A CTTA GTTA GTTC	CTA 9 10: CAC 11: AGA	ACT 30 AGA 20 TAT 10 TTC	I TTA: TTTI AGGJ	ITGO G ICCO 10 AATA 11 ATGO	30 30 30 30 30 30 30 30 30 30 30 30 30 3	A TTTT TAAGJ	I T 95 ATTTT 104 ACAGG	CGAT M 0 TTC/ GAG/ SD	V ACTT AAAT	27GA D 960 27AC 1050 2AAA M 1140 2CGC	IGAG E CTCC <b>ph</b> AATG M - (1) AGAA	L H CCT( <b>oP</b> 1 CGC( R \ 	970 CCCCC .060 STAC / L .150 SCCA	AAAC GCTG IGGI V -	GTT 1 TGT V - 1 AGC	980 TAT 070 TGA E - 160 CGA	ITA AGAC D -	ATG CAA N -	990 <u>990</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1080</u> <u>1</u>
s.	- - -	A ATG CCC	R GCG CCA TTA	CTG 91 ATG 100 TAA 109 CGT R -	AAA K 0 CCTG 0 CCCA H H	A TCC CAT -1 CAC	ATG 99 GGC 10: AAT 0 11( CTT: L 1 -	ACG T 20 CTG 10 CGC 00 AAAA K	P A CTTA GTTA GTTC	SCTA N 9 TTTA 10: CAC 11: CACAC 11: CAGA 11: CAGA 11: CAGA 11: CAGA	ACT. Y 30 AGA 20 TAT 10 TTC: Q - (21	I TTA: TTT? AGG? D -	TTGG G TCCG 10 AATF 11 ATGC A S	GTCC R 940 030 030 030 030 030 030 030 030 030 0	A TTTT TAAGJ	I T 95 ATTTT 104 ACAGG 113 CAGG Q V 	CGA: M 0 TTCJ GAGJ SD 0 TTCGJ D -	V ACTT AAAAT AAAAT D A	TGA D 966 TACC 1050 CAAA M 1140 CGCC A -	TGAG E CTCC <b>ph</b> AATG - (1) AGAA E - (31)	L H CCTC OP I CGCC R V  I GATC D H 	44AT 970 CCCC 060 STAC 7 L 150 SCCAI 4 K - R	AAAC GCTG IGGI V -	CTC GTT 1 TTGT V - 1 AGC A -	GTA 980 TAT 070 TGA E - 160 CGA D -	ITA AGAC D -	- CAA' N - ITA' Y -	990 ITTA LOBO IGCG A - L170 ICTC L -
S. E. S.		A ATG CCCC ITTG L .1)	R GCG CCA TTA	CTG 91 100 100 TAA 109 CGT R - 118 CAT	AAA K 0 CTG 0 CCCA 0 CCAC H 1 0 0 AATAC	A TCC CAT -1 CAC	ATG M 99 GGC 100 AAT 0 0 110 CTT L 110 CTT L 110 SAT	ACG T 20 CTG 10 CGC 00 AAAA K - 90 AATT	P A CTTA GTTA GTTC V C  GCGA	SCTA N 9 TTTA 10: SCAC 11: SCAC 11: SCAC 11: SCAC 11: SCAC 11: SCAC 11: SCAC	ACT. Y 30 AGA 20 TAT' 10 TTC: 0 (21 00 TCGJ	I TTA: TTTI AGGJ D )	TTGG G S TCCC 10 AATF 11 ATGC A S 12 TCGG	940 GCTT 930 940 940 940 940 940 940 940 940 940 94	A TTTT TAAGJ H - GCCC	INTERPORT	CGA: M 0 0 0 GAGJ 5 D 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	V ACTT AAAAT D ATGA A ACGG	TGA D 966 TTACC 1050 CAAAA M 1140 CCGC2 A A - ( 1230 TCTC	TGAG E 0 CCTCC <b>ph</b> 0 AAATG M - (1) ) AGAA E - (31) ) STCA	L I CCCTC OP 1 CCCCCC R V  1 GATC D J J GATC D J 1 CCTGA	AAAAT ( * 970 CCCCC 060 STAC 7 L  150 SCCAJ A K - R 240 TTCC	AAAAC GCTG IGGI V - AAGA E - -	GGTT 1 TGT V - 1 LAGC A - 1 CTG	GTA 980 TAT 070 TGAJ E - 160 CGA D - 250 GCG	rtai D - rtai Y -	ATG CAA N - TTA Y -	990 TTTA LOBO FGCG A - L170 FCTC L -
s.	; ; ; ; ; ; ; ; ; ; ;	A ATG CCCC ITTG L .1)	R GCG CCA TTA L -	CTG 91 ATG 100 TAA 109 CGT R - 118 CAT. H	AAA K 0 CTG 0 CCCA 0 CCAC H 1 0 0 AATAC	A TCC CAT -1 CAC H -	ATG M 99 GGC 100 AAT 0 0 110 CTT L 110 CTT L 110 SAT	ACG T 20 CTG 10 CGC 00 AAAA K - 90 AATT	P A CTTA GTTA GTTC V C  GCGA	SCTA N 9 TTTA 10: SCAC 11: SCAC 11: SCAC 11: SCAC 11: SCAC 11: SCAC 11: SCAC	ACT. Y 30 AGA 20 TAT' 10 TTC: 0 (21 00 TCGJ	I TTTA: TTTT: AGG2 - .) ATC1 L -	TTGG G S TCCC 10 AATF 11 ATGC A S 12 TCGG	940 GCTT 930 940 940 940 940 940 940 940 940 940 94	A TTTT TAAGJ H - GCCC	1130 1220	CGA: M 0 0 0 GAGJ 5 D 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	V ACTT AAAAT D ATGA A ACGG	TGA D 966 TAC 1050 CAAA M 1140 CCGC2 A - 0 1230 TCTC L -	TGAG E 0 CCTCC <b>ph</b> 0 AAATG M - (1) ) AGAA E - (31) ) STCA	L I CCCTC OP 1 CCCCCC R V  1 GATC D J J GATC D J 1 CCTGA	AAAATI ( * 970 CCCCC 5TAC: 7 L	AAAAC GCTG IGGI V - AAGA E - -	GGTT 1 TGT V - 1 LAGC A - 1 CTG	GTA 980 TAT 070 TGAJ E - 160 CGA D - 250 GCG	rtai D - rtai Y -	ATG CAA N - TTA Y -	990 TTTA 1080 TGCG A - 1170 L TCC L -
S. E. S.	; ; ; ; ; ; ; ; ; ; ; ; ; ; ;	A ATG CCCC TTG L - .1) AAT( 1)	R GCG CCA TTA L - SAA	CTG 1 91 100 100 100 100 100 CGT R - 118 CAT 127	AAA K  CTG' 0 .CCAC H CCAC H I I I I I I I I 0 0 0 0 0 0 0 0 0 0 0	A TCC CAT -1 CAC H -	ATG. M 9 9 9 9 9 9 9 9 9 9 9 9 9 9	ACG T 20 CTG 10 CCGC 00 AAAA K - - 90 ATTC 1 30	P A CTTA GTTA GTTC V Q  GCGA A I 	SCTA N 9 TTA 10: CAC 11: CAGA 12: TTG: V - 12:	ACT: Y 30 AGA' 20 TAT' 10 10 TTC: 00 00 TCGJ 00 51 90	I TTTA: TTTT AGG2 D - .) ATC1 L - .)	TTGG G G ICCC 10 AATF 11 ATGC A S 12 TCGG G - 13	5TCC R 940 5CTJ 030 04ATJ 030 CTGG G - 210 SATT L - 000	A TTTT TTTT TTTT TTTT H - GCCP P -	TATCA I T 95 XTTTT 1044 ACAGG Q V  1222 GACGA GACGA C D E  1310	CGA: M 0 0 GAGJ 5 5 0 5 0 7 CGAGJ 0 5 0 0 7 0 0 0 0 7 0 0 0 7 0 0 0 0 7 7 0 0 7 7 7 0 7	V ACTT AAAAT D A CGG G -	TGA D 966 TACC 1050 AAAA 1140 CCGCJ A 1140 CCGCJ A 1230 TCTC L - (11320	TGAG E 0 CTCC <b>ph</b> 0 AATG (1) 0 CTCC (1) 0 CTCA (31) 0 CTCA (51) 0 CTCA (61)	L I CCTC OP CCCCC R V CCCC D J L L L L I CCTCA L L I 1 CTCA L	AAAAT, ( * 970 CCCCC 060 GTACC / L  150 GCCAA A K - R 240 XTTCC R  330	AAAAC GCTG IGGI V - AAGA E - GCCG R -	I I I I I I I I I I I I I I	GTA 9800 TAT 0700 TGAJ E - 1600 CGAZ D - 2500 GCGD R - 3400	AGAO D - TTAT Y - S -	ATG CAA' N - ITA' Y - CAAC N S	2260 990 TTTA 1080 FGCG A - 1170 FCTC L 260 CGAT D - 1350
S. E. S. E.	; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;	AATG CCCC TTG L - .1) AAT( ) .1) STT	R GCG CCA TTA SAA SAA	CTG 1 91 ATG 100 TAA 109 CGT - 118 CAT. H - 127 CTG L	AAAAK K CTG CTG CCCAC H CCAC H D CCAC I I I I I I I I CCCG C CCG C CCG C CCG C CCG C CCG C CCG C CCG C	A TCC CAT -1 CAC H - CCCG P I CCCG H - CCCG H - CCCG H - CAC H - CAC H - CAC H - CAC H - CAC H - CAC H - CAC - CAC H - C CAC H - C CAC H - C CAC H - C CAC H - C CAC H - C CAC H - C C C C C C C C C C C C C C C C C	ATG. M 9 9GGC 10: AAT 0 11: CTT: 11: GAT: D : 12: CTGG	ACG T 20 CTG 10 CGC 00 AAAA K - 90 AATT I 1 30 STA' V	P A CTTA GTTC GTTC V Q GCGA A I SCGAA I SCGAA I TTAA	SCTA N 9 TTTA 100 CCAC 110 CCAC 110 CCAC V - 120 V - 120 V - 120 V -	ACT: Y 30 AGA 20 TAT 20 TAT (21 00 0 0 0 0 0 0 0 0 0 0 0 0	I TTTA: TTTT: AGG3/ D - .) ATC: L - .) GTG4 E	TTGG G G TCCC 10 AATF 11 ATGC A S 122 G G - 13 AAAG S	STCC R 940 SCTI 030 AATI 220 CTGG G - 210 SATI L - 900 SCTG W	A TTTTI TAAGI H - GGCCA Q	I T 95 ATTTT 1044 ACAGG 2 V V V V V V V V V V V V V V V V V V	CGA: M 0 0 GAGJ 5 D 5 D 0 CAGGA D - 0 0 AGGA D - 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	V ACTT AAAAT D A CGG G - CCGA E	TTGA D 966 TTAC( 105( CAAAA) M 1140 CGC2 A A - ( 1230 TCTC L - ( 1320 ( 1320 V V	TGAG E 0 CTCC <b>ph</b> AATG M - (1) AAATG M - (31) ) STCA (61) ) ATTA L	L I CCTC OP CCCCC CP CCCCC C R \ CCCCC D J I CCCCC D J I CCCCC I I CCCCC I I CCCCCC I I CCCCCC	AAAAT; ( * 970 CCCCC 0060 STAC: 7 L  150 SCCAJ A K - R 2400 TTTCC  330 CCCGG	AAAAC GCTG TGGT V - AAAGA E - GCCG R - - GTGC A	GTT GTT TTGT V - 1 LAGC A - 1 CTGA W - 1 CTGA D	GTA 9800 TAT 0700 TGAJ E - 1600 CGA2 D - 2500 GCG27 R - 3400 TGAJ	TTAN AGAO D - TTAN Y - S - TAGO S -	CAAA N TTA' Y - CAAA N S CAAA	2000 990 FTTA 1080 FGCG A - 1170 FCTC L 260 CGAT D - 1350 SACT
S. E. S. E.		AATG CCCC TTG L - .1) AAT( ) .1) STT	R GCG CCA TTA L GAA SAA S -	CTG L 911AATG 1000TAA 1097CGT L 118CGTR - 118CCAT. H -	AAAA K 0 CCCG CCCAC CCCCAC CCCCAC CCCCAC CCCCCAC CCCCCC	A TCC CAT -1 CAC H - CAC H - CAC H - CAC H - CAC H - CAC H - CAC H - CAC H - CAC CAC H - CAC CAC CAC CAC CAC CAC CAC	ATG. 9 9: GGC 10: AAAT 0 110: CTT: 100: CTT:	ACG T 20 CTG 10 CCGC 00 AAAA K - 90 ATTC I I 30 GTA V -	P A CTTA GTTA GTTA GTTC V C GCGA A I  GCGA A I  TTAA L T T	SCTAN N 99 TTA 10: CCAC 11: CCAC 12: TTG V - 12: CCCG A -	ACT: Y 30 AGA 20 TAT' 10 TTC: Q - (21 00 0 - (51 90 CCC: R - (81	I TTTA: TTTT: AGGJ D D  .) ATCT L  .) GTGJ E  .)	TTGG G G TCCG 100 AAT7 111 ATGC A S 122 CGG G - 133 AAAG S G	5TCC R 940 5CTT 030 04ATT 030 CTGG G - 210 SATT L - 000 SCTG W -	A TTTT TAAGJ H - GGCCJ P - GGCCA Q -	TATCA I T 95 ATTITT 1044 ACAGG Q V  1222 GACGG Q V - - 1222 GACGG Q V - - - - - - - - - - - - - - - - - - -	CGAN M 0 0 CTTCJ 0 CGAGJ 5 SD 0 CTCGJ D 0 CTCGJ 0 0 CTCGJ 0 0 CTCGJ 0 0 CTCGJ 0 0 CTCGJ 0 0 CTCCJ 0 0 CTCCJ 0 0 CTTCJ 0 0 CTTCJ 0 0 CTTCJ 0 0 CTTCJ 0 0 CTTCJ 0 C CTTCJ 0 C CTTCJ 0 C CTTCJ 0 C CTTCJ 0 C CTTCJ 0 C CGAGJ 0 C CTTCJ 0 C CGAGJ 0 C CTTCJ 0 C CGAGJ 0 C C C C C C C C C C C C C C C C C C	V ACTT AAAAT D A CGG CGA E -	TGA D 966 TTACC 1050 CAAAI M 1140 CGC A A - ( 1230 TCTC L - ( 1320 M 1320 C AGT Y V - ( (	TGAG E 0 CTCC <b>ph</b> 0 AATG M - (1) 0 AATG (1) 0 STCA (31) (31) (31) (31) (31) (31) (31) (31)	L I CCTCC OP 1 CCCCCC C CCCCR C CCCCR 1 CCTCGA 1 CCTCGA 1 1 CCTCGA 1 1 CCTCGA 1 CCTCGA 5 S P 5 S P 5 S P 5 S P 5 S P 5 S CCTCC C OP 1 1 CCCCCCC OP 1 1 CCCCCCC OP 1 1 CCCCCCC OP 1 1 CCCCCCC C C CCCCCC C C C CCCCCC C C C C	AAAAT 5 * 970 5060 50	AAAAC GCTG IGGI V - AAGA E - SCCG R - STGC A -	CTCC GGTT TTGT V - 1 AGCC A - 1 CTGA W - 1 CTGA D -	GTA 980 TAT 070 TGA E - 160 CGA D - 250 GCGA R - 250 GCGA R - 340 TGAT D -	TTAJ P TAGO S TTAJ Y -	ATG CAA N TTA Y CAAO N S CAAO N S	2000 990 TTTA 1080 FGCG A - 1170 CGAT - 1260 CGAT - - 1350 SACT - -
S. E. S. E. S.	: : : : : : : : : : : : : : : : : : :	A ATG CCCC TTG L - 1) AAT( 1) STT '1)	R GCG CCA TTA L - GAA S - TCA S -	CTG 91 91 100 109 CGT 118 CGT 118 CGT 127 CTG 127 CTG 127 CTT 127 CTT 127 CTT 100 CGT CGT 100 CGT CGT 100 CGT CGT CGT CGT CGT CGT CGT CGT	AAA K 0 CTG CCCA 0 CCCA H 1 CCCCA I I I I I CCCCA C CCCCA C CCCCA I I I I	A TCC CAT - CAC H - CCCG P I X X ATT	ATG. M 9 9GGC 100: AAT 0 110: CTT: L 11: GAAT 0 110: CTT: L 11: GAAT 0 110: CTT: GAAG	ACG T 20 CTG 10 CGC 00 AAAA K - - 90 AATT 1 1 2 - - - - - - - - - - - - - - - - -	P A CTTA GTTC GTTC V C GTTC V C GTCA I TTAA L T TTAA L T GTGA	Illi 100 CCAC 111 CCAC 111 100 CCAC 112 122 122 CCG A - 133 TGG	ACT: Y 30 AGA: 20 TAT: 10 10 TTC; 00 TTC; 10 00 TCC; 00 0 CCC; 80 CCC; 80 CCC;	I TTTA: TTTTI AGGJ D - - - - - - - - - - - - - - - - - -	TTGG G G TCCG 10 AATA ATG A AAAAG G G 13 AAAAG G G 13 TGCA	GTCC R 940 GCTT 030 AATT 20 CTGC G - 210 SATT L - 000 SCTC W - 990 AGGC	A TTTT PAAGJ TTCAT H - GCCZA P - GCCZA Q -	I T 95 NTTTT 1044 NCAGG Q V  1133 NCAGG Q V  1222 GACGG D E  131( GACAG D K	CGAY M 0 0 CTTCJ 0 CAGA D - 0 CAGGA D - 0 CAGGA D - 0 CAGGA D - 0 CGAY C CGAY C CGAY C CGAY C CGAY C CGAY C CGAY C C C C C C C C C C C C C C C C C C C	V ACTT AAAAT AAAAT AAAAT AAAAAT AAAAAT AAAAAT AAAAAT AAAAAT AAAAAT AAAAAT AAAAAT AAAAAT	TGA D 966 TACC 1050 AAAA M 1140 CCGC2 A TCTC L - ( 1230 L 1320 A GTZ V - ( ( 1410 TAGC	TGAG E 0 CTCC 0 AATG (1) 0 AGGAA E (1) 0 AGGAA (31) 0 CTCA (31) 0 CTCA (31) 0 CTCA (31) 0 CTCA (31) 0 CTCA (31) 0 CTCCC (31) 0 CTCCC (31) 0 CTCCC (31) 0 CTCCC (31) 0 CTCCC (31) 0 CTCCC (31) 0 CTCCC (31) 0 CTCCCCCC (31) 0 CTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	L I CCTC OP CCCCC R V  I GATC D J - I CCTGA L I I CCTGA S J S S I CCTGG	AAAAT \$ 970 CCCCC 060 STAC' 7 L 150 SCCAI A K  150 SCCAI A K - R 330 SCCGC CCGC  330 SCCGC  - - - - - - - - - - - - -	AAAAC GCTG TGGT V - AAGA E - SCCCG R - STGC A - STGC A -	GGTT 1 CTGT V - 1 LAGC A - 1 CTG W - 1 CTG A - 1 CTG A - 1 CTG A - 1 1 CTG A - 1 1 CTG A - - 1 CTG A - - - - - - - - - - - - -	GTA 980 TAT 070 TGAJ E - 160 CGAT D - 250 GCGT R - 340 TGAJ D - 340 CAT 2 - - - - - - - - - - - - -	TTAJ AGAO D TTAJ Y - TAGO S - TTAJ Y -	TTATG CAAC N - TTAY - CAAC S C CAAC S C CAAC S C CAAC S C CAAC S C CAAC S C C C C	2000 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

FIG. 2. Nucleotide sequence of the EcoRV fragment containing the *E. coli phoP-phoQ* operon and its flanking regions and the deduced amino acid sequences. The amino acid sequences of the PhoP and PhoQ proteins of *E. coli* and *S. typhimurium* were compared. SD and dots mark the ribosome-binding site. A putative -10 sequence is indicated. Hexanucleotide repeated sequences are indicated with wavy lines. The two convergent arrows indicate inverted repeat sequences that may serve as transcriptional terminators. Two hydrophobic transmembrane stretches in PhoQ are indicated with double underlines.  $\Im$ , extra serine residue in the *S. typhimurium* PhoQ protein. Asterisks indicate termination codons. The amino acid sequences of *S. typhimurium* PhoP and PhoQ are as described by Miller et al. (17).

the regulatory regions are highly conserved (13). Therefore, the repeated hexanucleotide sequences and the putative -10 region may be very important for the *phoP* promoter. Sequences that may form stem-loop structures are located

upstream of phoP and downstream of the TAA translational termination codon of phoQ. They may function as transcriptional terminators or regulatory elements of the downstream genes.

s.	P F 	1450 FTCAG Q -	GTTG <b>#</b>	TCTC			E -	ATTA L -	TCT/ S 1	I N	ATGA D	CGA		GAT	CAAZ	L -	GACO T -	CGCG A -		GAAT.	аса		-	IGG. E -	153 AAAC T -	
E. S.	(131) TTGA L I  (161)	racgc. R 1		TGGC		TGG	156 TCAG S -	CAAA	GATT		TAAT	GCT		ACT	CTAT	159 FCC P -		IGCG A -		TGC	GGG		GCC	ATA T -		
s.	DV	163 TACTG L	ATGGO	ACG			к -	AATT	CAGO	GCAC		TCC		AGA E	AGTO	I -	TACO	T T			GCC		GCT			
s.	GAAT E L  (221)	rgcgc R	TGATO * M	17 58888 K H		ACT L	R	CTTT	TTT? F	P -	GCTC L	TCG		CGG	GTAC		TTTC	TGT		CAAC	GGC A		GGT			G
E. S.	L	1810 TTCGC <u>S L</u> 	TTGCC A (31)	TACO Y O	<u>- 1</u>		<u> </u>	CTGA <u>L I</u> - V	TCGO G -	<u> </u>	TAGC <u>S</u> (41)	GTC V	<u>s</u>	TTC( <u>F</u> 1	GATA D P	к ' -	ACT7 T 1 	ACGT F 	R -	GCT L - (51)	GTT L -	R -	rgg( G -		s -	A N -
E. S.		190 GTTCT F Y 	ATACO		А К 	GTG	GGAA E -	N N 	ATAJ K -	AGTT L I	GCAT H S (71)	GTC V -	E -	TTA( L   - ·		E 1 - ·	AAT# N 1 - 1	ATCG D	K M	GCA Q - (81)	AAG S -	P -	CACO T -		Т -	C L -
E. S.	I	1990 TTATG Y D 	ATGAG E - (91)	AACG N G T -	Q K	L	L -	TGGG WA - T	Q Q -	R - (	TGAC D N 101)	GTG V I	P 1	TGG( W 1 	CTGA L M - I	41 C·	AAGA K M - S	TGA I I -	Q -	AGCC' P - (111)	TGA D E	W -	GCT( L -		S T	A N -
E. S.		2080 TTTTC: F H 	ATGAA E (121)	ATTO I E 	A T		v -	AACG ND DA	ATAC T -	CCAG S - (	L T 131)	TTG L -	L -	AGT( S ( - 1	GGAG G D E -		CATI H S	CGA I A	Q -	AGCA. Q E (141)	ACA Q K	L -	GCA Q K	Е	<b>v</b>	R -
E. S.		217( AGATG D D 	ATGAC D - (151)	GACG	E -		Т -	CACTO H S	CGGI V -	а - (	AGTA V 161)	AAC N -	V I	TAC( Y 1 	CCGG P	A !	ACAT T S - A	CGC R	м -	rGCC: P - (171)	AAA K Q	L -	AACO T -	I I	<b>v</b> -	G V -
E. S.		2260 GATA D T 	CCATT I (181)	PCCGG PV - I	E -	L _	к -	AGTTO S S R -	CCT Y -	м - (	GGTC <u>V</u> 191)	TGG. W -	AGC S	<u>W 1</u>	гтта <u>F 1</u> - V	<u>, ,</u>	TATO Y V	TGC L	S A	AGC A - (201)	CAA N	<u>L</u>	GCTO L	GTT) L	<u>v</u>	Ă
E. S.	тсссо 	SCTGC LLL	W (211)	GTCG V A I -		CTG	GTGG W -	S L	raco _ R	P P -	LATC	GAA E	GCC A	CTGO L 1	GCAA A K	1 1	GAAG E V	TCC R	GCGA E -	LACT	GGA E -	AGAJ E	ACA: H	N	R 1	G E
E. S.		CTCA L N	P	GCCA A 1 E -	CAAC	GCG: R	AGAA E	CTGA( L T	CCAG S -	STCT L -		CGA R 1 Q	AAC N	CTGI L 1	AACC N R	۲ ۱	rtgi L L	TAA K	s -	TGA	ACG R -	Е	ACG1 R	Y	DJ	A K

The phoP and phoQ genes could code for proteins of 223 and 486 amino acids with predicted molecular weights of 25,534 and 55,297, respectively. The deduced amino acid sequences of PhoP and PhoQ of S. typhimurium (17) and those of E. coli are 93 and 86% identical, respectively. There are 15 amino acid differences between the PhoP sequence of E. coli and that of S. typhimurium, while the PhoQ sequence of E. coli differs from that of S. typhimurium in 70 amino acids. Met-2 of PhoP and Ser-330 of PhoQ of S. typhimurium were absent in the corresponding positions in PhoP and PhoQ of E. coli (Fig. 2). The E. coli PhoQ protein has two putative transmembrane sequences in the N-terminal regions as predicted by the hydropathy profile, a phenomenon which is often found in the sensor proteins belonging to members of the family of the two-component regulatory systems (21). The features of the E. coli phoP-phoQ region mentioned above are very similar to those of the S. typhimurium phoP-phoQ regulon (17). Two ORFs were found, one upstream of the phoP gene and the other downstream of the phoQ gene, both of which extend further into unsequenced

		_	530		254	-	~~~		50			560			70			80			590			260	-	<b>m</b> 2 <i>C</i>	26	
E.														P L			L	Q	S	т	L	R	S	L	R	s		
s.	-	-	- (2	 71)	-	-	-	-	-	-	-	-,	- 281)		-	-	-	-	-	-		- 291)		-	-	N	-	-
		2	620		263	0		264	40		2	650		260	50		26	70		26	680			2690	0		27	00
Е.								GGT	AAT	GCT				AGCC														
s.	-			- к			-					-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
			(3)	)1)									311)									321)						
	TGCG		710 CGGGi	ACAT	272 IGCT			273 CGA0		GCA		740 GGT(		275 CCAC	60 FGCT				CACO		770 AGC			278( CAA)		GTA'	27: FCA	
E. S.		G	G	г L				Е		H	P	v	Α	P L	L	D		L	т	s	Ά		N					
5.		/s	3 (3:	31)									341)						•			351)						
			800					282							0		285				360			2870			28	
Е.									s	Ρ	Е	I	s	TTTG1 F V														
s.	-				-		-		-	-	-	-	_ 371)		-				-		-		-	-	-	-	-	-
				, 1																		-						
		TAA				TTG	CCT		GTT?	IGTO	CGA		TCT	293 GCAAG	GCA	AAC	CGA	CGA	GCAT	CTC		TATT	rgt(		CGA			rG
E. S.	D -		A (		Y -	с -	L -	E -	F	v -				AR 					н -						E	D -	D	G -
			(39	)1)								(	401)								(	411)						
			980		299			300							0		303							3050			300	
Е.		G	II	Ъ	S			Е	v	I	F	D	R	GGTC <i>I</i> G Q	R	v												
s.	-	-	- (42		-	-	-	s	L	v	-		- 131)		-	A	-	-	-	-	- ;	- 441)	-	-	-	-	-	-
			•									•									•							
			070		308			309				100		311			31:				130			314			31	
E.														GCCGO A G														
s.	-	-			-	-						-	I	- s							-	-	-					-
			(4	51)								,	461)								•	471)						
						•		~ ~ ~									~ ~ .			~ ~					•			
		CCA		CTG		GAA	AGA		ATA	AAT?			ATAC	320 TTCA0	0 GCA			10 ГАА				ATA		323( GGT			324 FAC	
E. S.	R	CCA Q		ICTGO 5 A	CGCC P	GAA K	AGA D	TGAI E	ATAI	AAT7																		
	R	CCA Q	GCAT'	FCTGO 5 A ? T	CGCC P	GAA K	AGA D	TGAI E	ATAI	AAT7																		
	R -	CCA Q - 3	GCAT H 2 - 1 (4) 250	FCTG( 5 A ? T 31)	CGCC P Q 326	GAA K - 0	AGA D E	TGAI E - 327	* * 70		ATGO 32	280	ATAC	TTCAG	GCA	TTA	CGT	TAA 00	GCAI	33	STT.	ATA	ATC	GGT	TGC	AGA!	333	CA 30
	R -	CCA Q - 33 GTG	GCAT H - ] (4) 250 GATG	TCTG( 5 A 9 T 31) CTTAJ	GCC P Q 326 ACAT	GAA K - 0 GGA	AGA D E	TGAI E - 327 CCAI	ATA * * 70	CACI	ATGO 32 FCT	280 TAA	ATAC CTGG	TTCA	GCA	TCT	330 TGA	DO ACG	GCAT	33 33	STT.	GAAJ	ACG	GGT 3320	TGC	AGA:	333 STT	CA 30 AA
	R - GCCT	CCA Q - 3: GTG S 3:	GCAT H 2 (4) 250 GATG SD 340	ICTGO 5 A 7 T 81) CTTAJ	326 326 ACAT M 335	GAA K GGA E	AGA D E ATA Y	TGAJ E - 327 CCAJ Q 336	ATAI	CACI T	ATGO 32 ICT: L 33	280 7880 7880 7880	ATAC CTGG W	329 CCCGP P D 338	O ATTT F	TTA	330 IGAJ 8	DO ACG R	GCA1 TCAC H	33 760 W 34	STT STT SCA Q	GAAJ K	ATC ACG R	3320 CCCC P 3410		AGA: GGT( V	333 3777 L 342	CA 30 AA K 20
	R - GCCT AACG	3: GTG 3: S 3: CGG	GCAT H (4) (4) 250 GATG( 50 340 CTTT	CTTA	326 2 326 ACAT M 335 ATTT	GAA K GGA GGA E 0 TAT	AGA D E ATA Y TGA	IGA/ E - 327 CCA/ Q 336 CCCC	ATAI * * 70 ACTO L 50 GATO	CACI T	32 FCT: L 33 FCC2	280 7AA0 N 370	ATAC CTGG W CGAG	329 CCCGA P D	O ATTT F	TCT	33( IGA) E 339	DO ACG R 90 GAT	GCA1 TCAC H GGAP	33 33 33 34 34	STT. S10 SCA Q 100 CGA	GAAJ K AGTI	ACG	3320 CCCC P 3410 CAG		AGA: GGT( V ACT(	333 3777 1 342 3670	CA 30 AA K 20 CA
	R - GCCT AACG	32 	GCAT' H 2 (4) 250 GATGO SD 340 CTTTI F 1	CTTA	326 9 0 326 ACAT M 335 F	GAA K - GGA E 0 TAT I	AGA D E ATA Y TGA	1GA/ E - 327 CCA/ Q 336 CCCC	ATAI * * ACTO L SO GATO I	CACI T	32 FCT: L FCC2 P	280 FAAC N 370 AGAC D	ATAC CTGG W CGAG	329 CCCGA P D 338 TTGGC L A	GCA 0 ATTT F 0 GGGG G	TCT	330 IGAI E 339 GGC0 A	DO ACG R SAT M	GCA1 TCAC H GGAP	33 TGC W 34 AGC S	STT. SCAC Q IOO CGAI E	GAAJ K AGTI	ACG R IGA	3320 CCCC P 3410 CAG		AGA: GGT( V ACT(	333 STT L 342 SGT V	CA 30 AA K 20 CA S
	R - GCCT AACG R GTCA	GTG GTG GTG G G G G G G G G G G G G G G	GCAT' H 2 (4) 250 GATG( 50 340 CTTT F 1 430 GGAT(	TCTGG A T 31) CTTAJ AATAJ N GGCAJ	3260 9 0 3260 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	GAA GGA GGA TAT	AGA D E ATA Y TGA D GGT	327 CCAP Q 336 CCCC P 345 CAGO	ATAI * * ACTO L SATO I SATO I SATO	CACT T S CTCT S	ATGO 32 FCT: L 33 FCC2 P 34 CCC0	280 TAAC N 370 AGAC D 160	CTGG W CGAG E CGAA	TTCAG 329 CCCGA P D 338 TTGGG L A 347 AGCTA	GCA	TCT L TCT L TCT	330 IGA 339 GGC0 A 348 ICT0	DO ACG R GAT M SO CGG	GCAT TCAC H GGAP E IGAA	33 750 W 34 AGC S 34 ACC	310 3CA Q 100 CGA 190 CAA	GAAJ K AGTI V CTGO	ATC ACG R IGA D	3320 CCCC P 3410 CAG S 3500 ATT		AGA GGT V ACTO L GGT	333 3777 L 342 3670 V 351 ACAO	CA BO AA K 20 CA S LO GG
	R - GCCT AACG R GTCA	GTG GTG GTG G G G G G G G G G G G G G G	GCAT H (41 250 GATG SD 340 CTTT F F 430 GGAT C D ( D ( )	CTTAI	CGCC P Q 3260 CAT M 3350 XTTT F 3440 XATG W	GAA K GGA E 0 TAT I 0 GCA Q	AGA D E ATA Y TGA D GGT V	327 CCAF Q 336 CCCC P 345 CAGC S	ATAI * * % % % % % % % % % % % % % % % % %	CACT T S CTCT S GGGC	32 FCT: L 33 FCCI P 34 CCCC	280 FAAGAG D 160 F	CTGG W CGAG E CGAA E	329 CCCGA P D 338 TTGGC L A 347 AGCTA S Y	GCA 0 TTT F 0 GGGG G 0 CGA D	TCT L TCT L TCA: H	330 IGAJ E 339 GGC0 A 348 ICT0 L	DO ACG R BO GAT G	GCAT TCAC H GGAF E IGAA	33 CTGC W 34 LAGC S 34 LACC T	310 GCA Q 1000 CGAJ E 190 CAA N	GAAJ K AGTI V CTGC	ATC R R D STC. S	3320 CCCC P 3410 CAG 3500 ATTI L	D GGT V D TCG R D ACT L	AGA: GGT( V ACT( L GGT/ V	333 3777 L 342 36370 V 351 ACAO Q	CA BO AAA K 20 CA S LO GG A
	R - GCCT AACG R GTCA H CAGT	CCA Q - 33 GTG G G G G G A A G A A	GCAT H (44) 250 GATG( SD 340 CTTT2 F H 430 GGAT( D ( 520 CCAC	TCTGO A T 31) CTTAJ AATAJ N SGCAJ G K	3260 9 0 3260 3350 3350 4 4 4 4 4 4 4 4 4 5 3 5 3 6 4 4 0 4 7 5 3 5 3 5 3 5 4 4 0 4 5 5 6 4 5 6 5 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 7 7 7	GAA K - 0 GGA E 0 TAT I 0 GCA Q 0 GCC	AGA D E ATA Y TGA D GGT V GAC	TGAN E - 327 CCAN Q 336 CCCC P 345 CAGC S 354 CCCCC	ATAJ * 70 ACTC L 50 SATC I 50 CCAC H 10 CCGCC	CACT T S CGGC G GCTC	32 ICT: I 33 PCCA P 34 CCCCC P 35 GATC	280 IAAG N 370 AGAG D 160 STTC F 550 GCGJ	CTGG W CGAG E CGAA E	329 CCCGJ P D 338 TTGGC L A 347 AGCTJ S Y 356 TTCCG	CGCA 00 ATTT F 00 CGGG G 0 ACGA D 0 0 CGA D 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	TTA TCT L TCT L TCA H	330 IGAJ E 339 GGC0 A 348 ICT0 L 357 ACC0	DO ACG R BO GAT( G G 70 GGA(	GCAT TCAC H GGAA E TGAA E	33 33 33 33 34 34 34 34 35 35 35 35	310 3CA Q 100 CGA 190 CGA 190 CAA N 580 CAA	GAAA K AGTT V CTGC W	ACG R IGA D STC S IGA	3320 CCCC P 3410 CCAG S 3500 ATT7 L 3590 TCT0	) GGT V ) TCG R ) ACT L ) GAT	AGA GGT( V ACT( L GGT/ V SAT1	333 3777 L 342 36070 V 351 ACA0 Q 360 TTC1	CA 30 AAA K 20 CA S CA S CA S CA S CA S CA S CA S CA
	R - GCCT AACG R GTCA H CAGT	CCA Q - 33 GTG G G G G G A A G A A	GCAT H (44) 250 GATG( SD 340 CTTT2 F H 430 GGAT( D ( 520 CCAC	TCTGO A T 31) CTTAJ AATAJ N SGCAJ G K	3260 9 0 3260 3350 3350 4 4 4 4 4 4 4 4 4 5 3 5 3 6 4 4 0 4 7 5 3 5 3 5 3 5 4 4 0 4 5 5 6 4 5 6 5 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 9 0 7 7 7 7	GAA K - 0 GGA E 0 TAT I 0 GCA Q 0 GCC	AGA D E ATA Y TGA D GGT( V GAC(	TGAN E - 327 CCAN Q 336 CCCC P 345 CAGC S 354 CCCCC	ATAJ * 70 ACTC L 50 SATC I 50 CCAC H 10 CCGCC	CACT T S CGGC G GCTC	32 ICT: I 33 PCCA P 34 CCCCC P 35 GATC	280 IAAG N 370 AGAG D 160 STTC F 550 GCGJ	CTGG W CGAG E CGAA E	329 CCCGJ P D 338 TTGGC L A 347 AGCTJ S Y 356	CGCA 00 ATTT F 00 CGGG G 0 ACGA D 0 0 CGA D 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	TTA TCT L TCT L TCA H	330 IGAJ E 339 GGC0 A 348 ICT0 L 357 ACC0	DO ACG R BO GAT( G G 70 GGA(	GCAT TCAC H GGAA E TGAA E	33 33 33 33 34 34 34 34 35 35 35 35	310 3CA Q 100 CGA 190 CGA 190 CAA N 580 CAA	GAAA K AGTT V CTGC W	ACG R IGA D STC S IGA	3320 CCCC P 3410 CCAG S 3500 ATT7 L 3590 TCT0	) GGT V ) TCG R ) ACT L ) GAT	AGA GGT( V ACT( L GGT/ V SAT1	333 3777 L 342 36070 V 351 ACA0 Q 360 TTC1	CA 30 AAA K 20 CA S CA S CA S CA S CA S CA S CA S CA
	R - GCCT AACG R GTCA H CAGT V	CCCA Q - 3: GTG GTG G 3: CCGG G 3: CCGG G 3: CCGG Q 3: 2 CCA S 3: 2 CCA S 3: 2 CCA S 3: 2 CCA S 3: 2 CCA S 3: 2 CCA S 3: 2 CCA S 3: 2 CCA S 3: 2 CCA S 3: 3: 2 CCA S 3: 3: 3: 3: 3: 3: 3: 3: 3: 3: 3: 3: 3:	GCAT H 2 250 GATGO SD 340 CTTT2 F 1 430 GGATC D C 520 CCAC H V 610	TCTGG T T T T T T T T T T T T T	CGCC P Q 3266 ACAT M 3355 ATTT F 3444 AATG W 3536 W 3536 S 3626	GAA K - 0 GGA E 0 TAT I 0 GCA Q 0 0 GCCA P 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	AGA D E ATA Y TGA D GGT V GGT C V GAC T	TGAI E - 327 CCAP Q 336 CCCC P 345 CCAGC S 354 CGCC A 363	ATAI * * 70 ACTC L 50 GATC I 50 CCAC H 0 CCAC H 0 CCAC A 0 2 CCAC	CACI T S CGGC G G G CTC L	32 FCT: 1 33 FCC2 P 34 CCCC P 35 SATC M 36	280 IAAG N 370 AGAG D 460 F 550 GCGJ R 540	CTGG W CGAG E CGAA E ACCG P	329 CCCGJ P D 338 TTGGC L A 347 AGCTJ S Y 356 TTCCC F R 365	CGCA 00 ATTT F 00 CGGG D 00 CGA D 00 CGA D 00 CGA 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	TTA TCT L TCT H ACT/L	330 FIGAJ E 339 GGCC A 348 FICTC L 355 ACCC P 366	DO ACG' R BO GAT( M 30 CGG' G 70 G G G G G G G G G G G G G G G G G	GCAT TCAC H GGAF E TGAA E	33 7760 W 34 AAGC S 34 AACC T 35 CGT R 36	310 3CA Q 100 CGAJ E 190 CAAC N 580 TAT I	GAAA K AGTT V CTGC W IGAT D	ACG R IGA D STC S IGA D	3320 CCCC P 3410 CCAG S 3500 ATT7 L 3590 TCT0 L 3680	D GGT( V ) TCGJ R ) ACT( L ) GAT( M	AGA GGT( V ACT( L GGT/ V SAT7 I	3333 3777 1 342 3577 V 3517 ACAC Q 351 ACAC Q 351 ACAC 2 355 365 365 365 365 365 365 365	CA 30 AAA K 20 CA S CA S CA S CA S CA S CA S CA S CA
	R - GCCT AACG R GTCA H CAGT V TTTC	CCA Q - 3: GTG S 3: CGGG G 3: CCA Q 3: CCA Q 3: CCA N 3: TGTI	GCAT: H : - 1 (44) 250 GATG( SD 340 CTTTI F 1 430 D ( 520 CCACT H V 610 ACCCC	TCTG( A T D1) CTTAJ AATAJ N N GGCAJ GGCCJ H GGCGC	CGCC P Q 3260 ACAT M 33350 ACAT F 33440 AATG W 3530 W 3530 W 3530 CGGG	GAA K - 0 GGGA E 0 TAT I 0 GCCA Q 0 GCCC P 0 CGT	AGA D E ATA Y TGA D GGT V GGT V GAC T T CGGC	TGAI E - 327 CCAI Q 336 CCCC P 345 CAGC S 345 CAGC S 354 CGCC A 363 CCCC	ATAI * * 0 ACTO L 50 SATO H 0 CCAO A 0 CCAO A 0 SCAD	CACT T S CCGGC G G G G CCTC L	32 ICT: J 33 ICC: P 34 CCCC P 35 SATC M 36 CGAJ	280 IAAG N 370 AGAG D 160 F 550 GCGJ R 540 ICCAG	CTGG W CGAG E CGAA E ACCG P	329 CCCGA P D 338 TTGGC L A 347 AGCTA S Y 356 TTCCG F R	CGCA 00 ATTT F 00 CGGG G 0 CGA D 0 CGA E 0 0 CGTT	TTA TCT L TCT L TCA H ACTI L TTAT	330 IGAJ E 339 GGC0 A 348 ICT0 L 355 ACC0 P 366 CAT2	DO ACG'R BO GATO GGG G G G G G G G G G G G G G G G G G	GCAT TCAC H GGAF E TGAA E CTGG W GGGT	33 7760 W 34 1AGC S 34 1ACC T 35 7 CG1 R 36 ACC	310 32CA Q 100 CGAJ E 190 CAA N 580 TAT I 570 CGGJ	GAAM K AGTT V CTGC W IGAT D	ACG R D STC S S CGA	3320 CCCC P 33410 CAG <sup>2</sup> S 3500 ATTI L 3590 L 3590 L 3590 C ATTI L 3590 C CCCC S	D GGT V D CCG R D CCTG	AGA GGT( V ACT( L SGT/ V SAT1 I SCG/	333 3777 L 342 347 342 3670 V 3511 ACAO Q 350 TCT S 369 AGTO	CA 30 AAA K 20 CA S CA S CA S CA S CA S CA S CA S CA
	R - GCCT AACG R GTCA H CAGT V TTTC S	CCCA Q - 33: GTG G G G G AA( N S G AA( N S G AA( N S S CCGG G S S S S S S S S S S S S S S	GCAT: H 2 250 GATGGATGGATG SD 340 CTTTI F 1 430 GGGATC D C 520 CCACT H 5 610 ACCCCC P C 700	TCTG( A A T T T T T T T T T T T T T	CGCC P Q 3326 ACAT M 3353 ATTT F 3444 AATGA W 3533 ATGA W 3533 ATGA CGGG G 33710	GAA K - O GGA E 0 TAT I 0 GCC Q 0 0 GCC V 0 0 CCT V 0 0 0 CCT V 0	AGA' D ATA' Y TGA' D GGT' V GAC' T CGGG	TGAN E 327 CCAH Q 3366 CCCCC P 345 CCAG S 354 CGCC A 363 CCCCC P 372	ATAI * * 70 ACT( L 50 STAT( 1 50 CCA( H 60 CCA( A 50 CCA( H 60 CC	CACT T S CGGC G G G G CTC L	ATGO 32 ICT 1 33 ICCI P 34 CCCO P 34 CCCO P 35 SATO M 36 CGAI D 37	280 N 370 AGAGA D 460 550 550 550 GCGJ R 540 CCAC Q 730	CGAG W CGAG E CGAA E CGAA E CGAA E CGAA E CGAA CCG CGAA E CGAA E CGAA CCG CGAG E CCGA E CCGA E CCGA E CCGA E CCGA E CCGA E C CCGA E CCGA E CCGA E C C CCGA E C CCGA E C CCGA E C CCGA E C C C C C C C C C C C C C C C C C C	329 CCCGJ P D 338 TTGGCT/ S Y 356 TTCCC F R 365 GACGT D V 374	CGCA 00 NTTT F 00 CGGG 0 0 CGA 0 CGA 0 CGA CGA CGA 0 CGA 0 CGA 0 0 CGA 0 0 CGA 0 0 0 0 0 0 0 0 0 0 0 0 0	TTA TCT L TCT L TCA H ACT L L TAT L	330 IGAJ E 339 GGC0 A 348 ICT 2357 ACC0 P 366 CAT7 I 375	TAA DO ACG R DO SAT( M SO CGG TCA( Q SO SO SO SO SO SO SO SO SO SO	GCAT TCAC H GGAF E TGAA E CTGG W G GGGI G	33 33 77 8 34 34 35 34 35 7 35 7 35 7 35 7 35 7 37	310 3CA Q 100 CGA 190 CAA N 580 TAT I 570 CGGJ G G 260	GAAA K AGTT V CTGO W IGAT D ACGI R	ATC ACG R IGA D STC S S TGA C G R R R	3320 CCCC P 3410 CAG S 3500 ATTJ L 3590 TCTC L 3680 R 3770	TGC: GGT( V ) CTCGZ ) ACT( M ) SAT( M ) CTG( W	AGA GGTC V ACTC L GGTJ V SATT I SATT I SCGJ R	333 3777 1 342 351 351 360 2 360 2 360 2 360 2 360 2 360 2 360 2 360 2 378	CA 30 AAA K 20 CA S CA S CA S CA S CA S CA S CA S CA
	R - GCCT AACG R GTCA H CAGT V TTTC S GCGA	CCCA Q - 33: GTG G G G G G A A A A A A A A A A A A A	GCAT: H 2 (4) (4) (250 (5) (5) (5) (5) (5) (5) (5) (7) (7) (7) (7) (7) (7) (7) (7	TCTGG A T T T T T T T T T T T T T	3260 P Q 3260 ACAT M 33350 ATTT F 3440 W 3530 W 3530 W 3530 W 3530 CGGG G G 3710 CGAA	GAA K - 0 GGGA E 0 0 TAT I 0 GCCA Q 0 0 GCCA Q 0 0 CGT V 0 0 CGT V 0 0 CGT	AGA' D ATA( Y TGA( D GGT( V GGC( G G GCA( G GCA(	TGAN E 327 CCAH Q 3366 CCCCC P 345 CCAG S 354 CCCCC A 363 CCCCC P 372 CCCCC P 372 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ATAI * * 70 ACTC L 50 50 50 50 50 50 50 50 50 50 50 50 50	CACT T S CGGC G G CTC L F CTC L	32 ICT: J 33 ICC: P 34 CCC: P 35 SATC M 36 CGAI D 37 ICC:	280 N 370 AGAG D 460 F 550 GCGJ R 540 CCAC Q 30 GGAJ	CTGG W CGAG E CGAA E CGAA E STACCG Y	3225 CCCGJ P D 3386 TTGGC L A 347 AGCTA S Y 356 F R 365 GACGT D V	CGCA	TTA TCT L TCT L TCA H ACTI L TAT( I CGA	330 IGAJ E 339 GGC0 A 348 ICT I 357 ACC0 P 366 CATT I 375 ICC0	IAA OO ACG' R OO GAT O GAT O GAT O CGG' O CGG' O CGG' O CGG' O CGG' O CGG' O CGG' O CGG' O CGG' O CGG' O CGG' O CGG' O CGG' N CGG' O CGG' CGG' O CGG'	GCAT TCAC H GGAF E TGAF E CTGG W 3GGT G	333 TGC W 344 AGC S 344 AGC S 344 AGC T 355 CGI R 366 CACC T 377 GCC	310 310 35CA Q 100 CGAJ 570 CGGJ G G 60 CATC	GAAA K AGTT V CTGC W IGAT D ACGT R CATC	ACG R IGA D STC S S CGA	3320 CCCC P 3410 CCAG S 3500 ATTI L 3590 TCTC L 3680 TCTC R 3770 TGAJ	TGC: GGT( V ) TCG; R ) TCG; N ) CTG( W ) CTG( W ) AGA(	AGA GGT( V ACT( L GGT/ V SATT I SCG/ R SCG/ R	333 5777 1 342 5670 V 3511 ACA0 Q 3501 S 369 V 378 6600 V 378 6600	CA 30 AAA K 20 CCA S LO GG GG GG GG GG GG GG GG GG G
	R - GCCT AACG R GTCA H CAGT V TTTC S GCGA	CCA Q - 33 GTG GTG G G 34 CCA Q Q 35 GAA N 36 GAA N 36 TGTJ V 37 AAAA K	GCAT: H :: GATGC GATGC GATGC GATGC GATGC CCATTI F : H : S10 GGATCC CCACT F : H : S10 GGATCC CCACT F : CCACT F : CCACT F : CCATCC CCACT F : CCACT F : CCACT F : CCACT F : CCACT F : CCACT F : CCACT F	TCTGG A T T T T T T T T T T T T T	CGCC P Q 3266 ACAT M 3350 ATTT F 3440 AATGA W 3530 ATGAA SCGGG G 3711 CGAA K	GAA K - OGGAJ E OTAT'I OGCA( Q OGCC( P OCGT( V O ACA( Q	AGA' D ATA( Y TGA( D GGT( V GAC( T CGG( G G CGG( H	TGAN E 327 CCAP Q 336 CCCC P 345 CCCC A 354 CCCCC A 354 CCCCC P 354 CCCCC A 353 CCCCC P 372 CCCCCC C C C	ATAI * * * * * * * * * * * * * * * * * * *	CACT T S CGGC G G CTC L F CTC L ACAT H	ATGO 32 TCT: L 33 TCC2 P 34 CCC20 P 35 SATO D 37 TCC20 P 35 SATO D 37 TCC20 P	280 IAAG N 370 AGAA D 460 STTC F 550 SCGJ R 540 CCAC Q 730 SGAJ D	CTGG W CGAG E CGAA E CGAA E STAC Y CTG L	3299 CCCCGP P D 3388 TTGGC L A 3477 AGCTTS S Y 356 TTCCC F R 3665 GACGT D V 374 TTACP L Q	CGCA 00 NTTT F 00 CGGG 0 0 CGA 0 0 CGA F 0 0 CGA F 0 0 V	TTA TCT L TCT L TCT H ACT L TAT L CGA? D	CGT 330 IGAJ E 339 SGCC A 348 ICTC L 357 ACCC P 366 CAT I 375 CCC P	IAA ACG R JO GAT G G G G G G G G G G G G G G G G G G	GCAT TCAC H GGAF E TGAF E CTGG W 3GGT G	33 33 33 34 34 34 35 35 36 34 35 36 34 35 36 37 37 37 37 37 37 37 37	310 3CA Q 100 CGAJ E 190 CAA N 570 CGGJ G 60 CATC I	ATAI K AGTI V CTGO W TGAI D ACGI R ACGI I	ATC ACG R IGA D STC S S CGA D	GGT 332( CCCC P 341( CCAG S 350( ATT) L 359( TCTC L 359( CCCC S 350( CCAG S 350( CCCC P 350( CCCC P 350( CCCC P 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S 350( CCCC S S S 350( CCCC S S S 350( CCCC S S S 350( CCCC S S S S S S S S S S S S S S S S S	TGC: GGT( V ) CTGC ) SAT( V ) SAT( V ) SAT( V ) SAT( V ) SAT( V ) SAT( V )	AGA GGT( V ACT( L GGT/ V SATT I SCG/ R SCG/ R	333 342 342 342 351 351 351 351 351 351 351 351	CA 30 AAA K 20 CA S LOGA S LOGGA 00 TF F 00 GG G 00 CCP
	R GCCT AACG R GTCA H CAGT V TTTCC S GCGA E CTGG	CCA Q - 33 GTG GTG G G 33 CCA Q Q 33 CCA Q Q 33 CCA N Q 33 TGTI V V 33 TGTI V V 33 TGA C CA S G AAA N S TGA S	GCAT: H :: GATG GATG GATG SD 340 CTTTJ F 1 430 GGATC D C 520 520 520 F 1 610 ACCCCC P C 700 GCTGC L C 790 TATTC	TCTGG A A T T T T T T T T T T T T T	CGCC P Q 3266 MCAT M 3355 MTTF S 3444 W 3533 S CGGG G 3710 CGAA K 3800 ATAT	GAA K - 0 GGGA E 0 0 0 CGT4 Q 0 0 CGT4 Q 0 0 CGT4 Q 0 0 0 0 0 0 0 0 0 0 0 0 0	AGA' D ATA Y TGA D GGT V GACC T CGGC G G GCA H	TGAN E 327 CCAR Q 336 CCCC P 345 CAGC S 354 CGCC A 363 CCCC P 372 CTGC C 381 AGGZ	ATAI * * 70 ACTO L 50 SATO I 50 CCAO H 60 CCCAO H 60 CCCCA P 10 CCCCCA P 10 CCCCC P 10 CCCCCC P 10 CCCCC P 10 CCCCC P 10 CCCCC P 10 CCCCCC P 10	CACT T S CGGC G G CGGC G G CTC L H H CCCC	32 FCT L 33 FCCT P 34 CCCC P 35 SATC M 36 CGA1 D 37 CCC P 38 SGA2	280 TAAC N 370 AGAC D 460 F 550 GGCGJ R 550 GGCGJ R 730 GGAT D 320 TGAJ	ATAC CTGG W CGAG E CGAA E ACCCG Y STAC Y STAC Y	TTCAC 329 CCCGG P D 338 TTGGC L A 347 AGCTZ S Y 356 GACCT GACCT C R 365 GACCT D V 374 TTACA L Q 383 TACGC	CGCA 00 NTTT F 00 CGGG 0 0 CGGG 0 0 CGGT F 0 0 CGGT V 0 0 0 0 0 0 0 0 0 0 0 0 0	TTA TCT L TCT L TCA H ACTI L TATC I CGA D	33( IGAJ E 33( IGAJ E 33( CA 2 33( L 33( L 33( P 36( CAT) I 375 C C P 38( AAA)	TAA OO ACG' R O SAT' M O SAT' M O SGG' CGG	GCAT TCAC H GGAF E TGAF E SGGT G CGAA E GATC	333 7760 W 344 AGC S 340 T 35 CGT R 36 CGT T 37 CGT T 37 CGC T 37 CGC T 37 CGC T 37 CGC T 37 CGC T 37 CGC T 33 CGC T 33 CGC T CGC T S S S S CGC S S S S S S S S S S S S S S	310 3CA Q 100 CCAA S 360 CCAA G G G G G G G G G G G G G G G G G	GAAL K AGTT V CTGO W IGAT D ACGT R CATC I I TTCO	ACG R IGA D STC S S CGA CGT	GGT 332( CCCC P 341( CAG' S 350( ATT) L 350( ATT) L 350( CAG' S 350( CAG' S 350( CAG' S 350( CAG' S 350( CCCC S S 360( CCCC S S S 350( CCCC S S S 350( CCCC S S S 360( CCCC S S S S S S S S S S S S S S S S S	TGC: GGT( V ) CTCG M ) CTG( W ) CTG( W ) CTG( CTG)	AGA GGTC V ACTC L GGT/ V SATT I SCG7 R R SCTC L	333 5777 1 342 5777 342 5777 342 5777 3511 5777 3657 5778 5778 5778 5778 5778 5778 5777 5778 57777 5777 5777 5777 5777 5777 5777 5777 5777 5777 5777	CA BOAAK 202AS LOGGA 00TFF 900GG 00CCP 700C
	R GCCT AACG R GTCA H CAGT V TTTCC S GCGA E CTGG	CCA Q - 33 GTG GTG G G 33 CCA Q Q 33 CCA Q Q 33 CCA N Q 33 TGTI V V 33 TGTI V V 33 TGA C CA S G AAA N S TGA S	GCAT: H :: GATG GATG GATG SD 340 CTTTJ F 1 430 GGATC D C 520 520 520 F 1 610 ACCCCC P C 700 GCTGC L C 790 TATTC	TCTGG A A T T T T T T T T T T T T T	CGCC P Q 3266 MCAT M 3355 MTTF S 3444 W 3533 S CGGG G 3710 CGAA K 3800 ATAT	GAA K - 0 GGGA E 0 0 0 CGT4 Q 0 0 CGT4 Q 0 0 CGT4 Q 0 0 0 0 0 0 0 0 0 0 0 0 0	AGA' D ATA Y TGA D GGT V GACC T CGGC G G GCA H	TGAN E 327 CCAR Q 336 CCCC P 345 CAGC S 354 CGCC A 363 CCCC P 372 CTGC C 381 AGGZ	ATAI * * 70 ACTO L 50 SATO I 50 CCAO H 60 CCCAO H 60 CCCCA P 10 CCCCCA P 10 CCCCC P 10 CCCCCC P 10 CCCCC P 10 CCCCC P 10 CCCCC P 10 CCCCCC P 10 CCCCCC P 10 CCCCCC P 10 CCCCCC P 10 CCCCC P 10 CCCCC P 10 CCCCCC P 10 CCCCCC P 10 CC	CACT T S CGGC G G CGGC G G CTC L H H CCCC	32 FCT L 33 FCCT P 34 CCCC P 35 SATC M 36 CGA1 D 37 CCC P 38 SGA2	280 TAAC N 370 AGAC D 460 F 550 GGCGJ R 550 GGCGJ R 730 GGAT D 320 TGAJ	ATAC CTGG W CGAG E CGAA E ACCCG Y STAC Y STAC Y	325 CCCCGJ P D 338 TTGGCL A 347 AGCTTA S Y 356 TTCCCF R 356 GACGT D V 374 TTACA L Q 383	CGCA 00 NTTT F 00 CGGG 0 0 CGGG 0 0 CGGT F 0 0 CGGT V 0 0 0 0 0 0 0 0 0 0 0 0 0	TTA TCT L TCT L TCA H ACTI L TATC I CGA D	33( IGAJ E 33( IGAJ E 33( CA 2 33( L 33( L 33( P 36( CAT) I 375 C C P 38( AAA)	TAA OO ACG' R O SAT' M O SAT' M O SGG' CGG	GCAT TCAC H GGAF E TGAF E SGGT G CGAA E GATC	333 7760 W 344 AGC S 340 T 35 CGT R 36 CGT T 37 CGC T 37 CGC T 37 CGC T 37 CGC T 37 CGC T 37 CGC T 37 CGC T 33 CGC T 33 CGC T CGC T S S S S CGC S S S S S S S S S S S S S S	310 3CA Q 100 CCAA S 360 CCAA G G G G G G G G G G G G G G G G G	GAAL K AGTT V CTGO W IGAT D ACGT R CATC I I TTCO	ACG R IGA D STC S S CGA CGT	GGT 332( CCCC P 341( CAG' S 350( ATT) L 350( ATT) L 350( CAG' S 350( CAG' S 350( CAG' S 350( CAG' S 350( CCCC S S 360( CCCC S S S 350( CCCC S S S 350( CCCC S S S 360( CCCC S S S S S S S S S S S S S S S S S	TGC: GGT( V ) CTCG M ) CTG( W ) CTG( W ) CTG( CTG)	AGA GGTC V ACTC L GGT/ V SATT I SCG7 R R SCTC L	333 5777 1 342 5777 342 5777 342 5777 3511 5777 3657 5778 5778 5778 5778 5778 5778 5777 5778 57777 5777 5777 5777 5777 5777 5777 5777 5777 5777 5777	CA BOAAK 202AS LOGGA 00TFF 900GG 00CCP 700C
	R GCCT AACG R GTCA H CAGT V TTTC S GCGA E CTGG G	CCA Q GTG GTG GTG G GG G G G G A CCA C G G G G G G G G A CCA C G G G G	GCAT: H :: GATG GATG GATG GATG GATG CTTT F 1 430 GGATC D C 520 520 520 610 ACCCCAC F C 610 ACCCCCAC F C 700 GCTG GCTG GCTG C ACCCCAC F C 1 1 3880	CTTAI AATAI CTTAI CTTAI AATAI SGCAA S K CGGCAA S K CGGCAA S G CAAAT 2 M CTTTI	CGCC P Q 3266 ACAT M 3350 ATTT F 3344 ATGA W 3530 ATTT CGAA CGGG G G G G G G G G G G G G G G G	GAA K - 0 GGGA E 0 0 CGTAT I 0 0 CGTAT V 0 0 0 CGTA V 0 0 0 0 0 0 0 0 0 0 0 0 0	AGA' D ATA( Y TGA( D GGT( C GGC T CGGG G G CA( H GCC) P	TGAN E 3277 CCAN Q 3366 CCCCG P 3455 CCAGC S 3545 CCAGC S 3545 CCAGC S 3545 CCAGC C 3545 CCAGC S 3545 CCAGC S 3545 CCAG S 3545 CCAG S 3545 CCAG S 3545 CCAG S 355 CCAG S 3565 CCCCC S 3575 CCAG S 3565 CCCG S 3575 CCAG S 3575 CCAG S 3575 CCAG S 3575 CCAG S 3575 CCAG S 3575 CCAG S 3575 CCAG S 3575 CCAG S 3575 CCAG S 3575 CCAG S 3575 CCAG S 3575 CCGC S 3575 CCG S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3575 CCGC S 3775 CCGC S 3775 CCGC S 3775 CCGC S 3775 CCGC S 3775 CCGC S 3775 CCGC S 3775 CCGC S 3775 CCC S 3775 CCCC S 3775 CCCC S 3775 CCCC S 3775 CCCCC S 3775 CCCCC S 3775 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ATAI * * 10 CCCA 10 CCCCA 10 CCCA 10 CCCA 10 CCCA 10 CCCA 10 CCCA 10	CACT T S CGGC G G CTC L CTC L CTC L CCCC P	ATGO 32 ICT: 1 33 ICC: P 34 CCC: P 34 CCC: P 35 SATO D 37 TCC: P 38 SCA: H 33	280 IAAA N 370 AGAGA D 460 F 550 GCGJ R 550 GCGJ R 540 CCAC Q 30 GGAT D 320 TGAJ E 910	ATAC CTGG W CGAG E CGAA E ACCCG P STAC Y STAC Y STAC G	TTCAC 329 CCCGG P D 338 TTGGC L A 347 AGCTZ S Y 356 GACCT CCC GACCT CCC 347 AGCTZ S Y 356 CCCGZ A 347 A 356 CCCGZ A 347 A 357 CCCGZ A 347 A 356 CCCGZ A 347 A 356 CCCGZ A 347 A 356 CCCGZ A 347 A 356 CCCGZ A 347 A 356 CCCGZ A 356 CCCGZ A 356 CCCGZ A 357 CCCGZ A 356 CCCGZ A 356 CCCGZ A 356 CCCZ A 356 CCCZ A 356 CCCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 356 CCZ A 357 CCC A 357 CCC A 357 CCC A 357 CCC A 374 A A A A A A A A A A A A A	CGCA 00 VITT F 00 CGGG CGGG CGGT CGGT V 00 CGGT L 10 10 10 10 10 10 10 10 10 10	TTA L TCT L TCA H ACTI L TATC I CGA E	CGT 330 IGAJ E 339 360 A 348 ICTC P 360 CAT I 375 CCC P 388 AAAA N 393	TAA ACG'R 90 GACG'R 90 GACG'R 90 GGAC D GGAC D GGAC D GGAC CAC CAC CAC CAC CAC CAC CAC CAC CA	GCAT TCAC H GGAF E CTGG W GGATC GATC M	33 TGC W 34 AAGC S 34 AAGC S 35 CGI T 35 CGI T 37 GCC A 38 SAAC N 33 SAAC N 33 SAAC N 33	310 360 37 37 310 37 37 37 37 37 37 37 37 37 37 37 37 37	GAAJ K AGTT V CTGC W CTGCT D ACGT R CATC I S	ATC R IGA D STC S S CGA D CGT V	GGT 332(C CCCC P 331(C CAGT 350(C ATTT L 359(C R 377(C R 377(C) R 3366( G G G 3395(	TGC: CGGT V CTCG: CTCG: CTGC M CTGC M CTGC M CTGC CTGC CTT F 0	AGA GGTC V ACTC L GGTJ V SATT I SCGJ R SCTC L TCGG R	333 342 342 342 342 342 342 342	CA 30 AAA K 20 CCA CCA 20 CCA 20 CCA 20 CCA 20 CCA 20 CCA 20 CCA C
	R GCCT AACG R GTCA H CAGT V TTTC S GCGA E CTGG G CAAAA	CCCA Q GTG GTG GTG S GAC Q 34 CCCA Q 35 GAA V 33 GAA V 33 TGT V 33 TGA C GA S C G G A A A A A C C G G G G G G G G G G	GCAT: H :: GATG( GATG( SD GATG( SD GATG( SD GCATG( D CCACT H V 610 CCACT	CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI SGCAI SGCAI SGCAI SGCAI SGCAI CTTAI CTTAI CAAAT	CGCC P Q 3326 ACAT M 3355 ATTT F 3344 ATGA 3353 ATTT S 3353 ATTT S 3353 ATTT S 3353 ATTT S 3353 ATTT S 3354 A CGAA S CGAA S CGAA S CGAA S CGAA S C C C C C C C C C C C C C C C C C C	GAA K - 0 GGGA C C C C C C C C C C C C C	AGA' D ATA' Y TGA' D GGT( C GGC T GGCC H GCCC P TGG	TGAN E 3327 CCAM Q 3366 CCCC P 3455 CAGC S 3545 CAGC S 3545 CAGC S 3545 CAGC S 3545 CAGC S 3545 CAGC S 3545 CAGC S 3545 CCAM P 372 CCAM P 3366 CCAM P 3376 CCAM P 3366 CCAM P 3366 CCAM P 3376 CCAM P 3366 CCAM P 3366 CCAM P 3366 CCAM P 3366 CCAM P 3376 CCAM P 3376 CCAM P 3376 CCAM P 3376 CCAM P 3377 CCAM P 3376 CCAM P 3377 CCAM P 3377 CCAM P 3377 CCAM P 3377 CCAM P 3377 CCAM P 3377 CCAM P 3377 CCAM P 3377 CCAM P 377 CCAM P 377 CCAM C C 3377 CCAM C C 3377 CCAM CCAM C C 3377 CCAM C C 3377 CCAM C C C 3377 CCAM C C C C 3377 CCAM C C C C C C 3377 CCAM C C C C C C C C C C C C C C C C C	ATAI ATAI ATAI ACTO L 50 CCAO H CCCAO H CCCAO H CCCAO H CCCAO H CCCCAO H CCCCAO H CCCCAO H CCCCAO H CCCCO CCCAO H CCCCO CCAO CCCO CCAO CCCO CCAO CCAO CCAO CCCO CCO	CACT T S CGGC G G G CTC L F CTC L H CCCC P	ATGO 32 ICT: J 33 ICC2 P 34 ICC2 P 34 ICC2 P 35 SATC D 37 ICC2 P 38 SCA H 39 ITA	2800 <b>IAAG</b> N 3700 AGAG D 3700 C 5500 5500 5500 5500 5500 5600 CCAC Q 3300 5600 7300 5000 7300 5000 7300 7300 7300 7300 7300 7300 7300 7300 7300 7300 7500	CTGG W CGAG E CGAA E CGAA E STACCG Y STACCG Y STACCG G G G G G G G G G G	322 CCCGJ P D 338 TTGGC L A 347 AGCTY S Y 356 TTCCC F R 365 GACGT D V 374 TTACA L Q 383 TACGQ Y A	CGCA 00 11 11 11 10 10 10 10 10 10	TTA TCT L TCT L TCT L TCA: H ACT/ L CGA: D CGGA: E ACT/	CGT 330( IGA) 332 332 332 332 332 1 335 1 375 2 375 2 38- N 395 395 395 395 38- N 395 395 395 395 395 395 395 395	TAA DO ACG R DO GACG R D CGG CG CG CG CG CG CG CG CG C	GCAT H GGAA E TGAA E CTGG W GGGT G GATC M CAAC	CCCC 33 CTGC W 34 JAGC S 34 JAGC S 34 JAGC S 34 JAGC T 35 CCCC T 37 CCCC A 36 CCCC A 36 CCCCC A 36 CCCCC A 36 CCCCCC A 36 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	310 3CA Q 100 CGAJ E 190 CAAC N 580 CAAC I 570 C C 570 C C C C C C C C C C C C C C C C C C C	GAAL K AGTI V TGAI D ACGI R CATC I TTCC S CAGO	ACG R IGA D STC S S CGA CGA CGA CGA	GGT 3320 CCCC P 3410 S 3500 ATTI L 3590 TCC R 3770 TCC R 3770 TCC R 3866 G G 3950 G 3950 CCCC P	TGC: GGT( V ) CGGT( R ) CTG( ) CTG( ) CTG( ) CTG( ) CTG( ) CTG( ) CTG( ) CTG( ) CTG( ) CGT( ) CGGT( ) CGT( ) CGGGT( ) CGGGT( ) CGT( ) CGGGT( ) CGGGT( ) CGGGT( ) CGGGT( ) CGGGT( ) CGGGT( ) CGGGT( ) CGGGT( ) CGGGGT( ) CGGGT( ) CGGGT( ) CGGGGT( ) CGGGGT( ) CGGGGT( ) CGGGGT( ) CGGGGT( ) CGGGGT( ) CGGGGT( ) CGGGGT( ) CGGGGT( ) CGGGGGT( ) CGGGGGT( ) CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	AGA GGTC V ACTCC L GGT/ V SATT I SCGJ R SCTCC L TCGC R	333 342 342 342 351 1 342 351 342 351 351 351 351 351 351 351 351	CA 30 AAA K 20 CA S CA S C
	R GCCT AACG R GTCA H CAGT V TTTC S GCGA E CTGG G CAAAA N	CCA Q - 3; GTG GTG S 3; GTG G 3; GGA Q 3; GGA N 3; GGA N 3; TGA; D 3; TGA; D 3; S 3; CCA S C C C S S C C C S C C C C C C C C	GCAT: H : - 1 (41 250 GATG( 50 GATG( 50 CTTT) F 1 430 CTTT) F 1 430 CTTT) F 1 430 CTTT F 1 430 CTTTT F 1 430 CTTTT F 1 430 CTTTT F 1 430 CTTTT F 1 430 CTTTTT F 1 430 CTTTTT F 1 430 CTTTTT F 1 430 CTTTTT F 7 F 7 F 7 F 7 F 7 F 7 F 7 F 7	CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI SGCAI SGCAI SGCAI SGCAI CTTAI CTTAI CTTAI SGCAI CAAAT	CGCC P Q 3326 NCAT M 3335 NTGAT F 3344 NTGA S 3353 NTGA X 3353 NTGA X 3380 NTAT I 3899 NTAT I 3398	GAA K - OGGA E OTAT I OGCA Q OCGT V OCGT V OCGT V O OCGT V O OCGT V O OCGT S O O CGT O O O CGT O O O O O O O O O O O O O O O O O O O	AGA' D E ATAG Y TGAC D GGTC V GGCC G G GCCC G G GCCC G G GCCC G G G G G G G G G	TGAN E 327 CCAN Q 3366 CCCC0 P 345 CCAC P 345 CCAC C A 363 CCCC0 A 363 CCCC0 A 363 CCCC0 A 363 CCCC0 A 372 CCAC A 366 CCCC0 A 372 CCC0 A 366 CCCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 A 372 CCC0 C C 372 CCC0 C C 372 CCC0 C C 372 CCC0 C C 372 CCC0 C 372 CCC0 C 3372 CCC0 C C 3372 CCC0 C 3390 CCC0 C C 3390 CCC0 C C 3390 CCC0 C C 3390 CCC0 C C 3390 CCC0 C C 3390 CCC0 C C 3390 CCC0 C C 3390 CCC0 C C C C C C C C C C C C C C C C	ATAI * 70 ACTC L 50 50 50 50 50 50 50 50 50 50	CACT T S CCGGC G G CCTC L CCTC L CCTC L CCTC L CCTC L CCTC L CCTC C CCTC C CCTC C CCTC C CCTC C CCTC C CCTC C CCTC C CCTC C CCTC C CCTC C CCCC C C C C C C C C C C C C C C C C	ATGO 32 FCT L 33 FCCJ P 34 CCCO P 34 CCCO P 35 SATO P 36 SATO P 33 SATO P 33 SATO P 33 SATO P 34 SATO P SATO P SATO P SATO P SATO P SATO P SATO SATO P SATO SATO SATO SATO SATO SATO SATO SATO	280 280 N 370 AGAG D 370 AGAG F 550 550 F 550 F 550 F 550 F 550 F 550 SCGJ R 550 SCGJ R 320 TGAI E 910 O TGAI CAGO V 30 SCGJ V 30 SCG SCGJ V SCG SCG SCG SCG SCG SCG SCG SCG SCG SCG	ATAC CTGG W CGAG E CGAA P CCGG P STAC C G STAC G G CTG G C CTG G C CTG G C CTG G	3225 CCCGJ P D 338 TTGGC L A 347 AGCTA S Y 356 TTCCC F R 365 GACGT D V 374 TTACA L Q 3833 74 CAACC Y A 392 CAACC Q R 401	CGCA 00 NTTT F 00 CGGG CGGG CGGG CGGT CGTT 00 CGTT 00 CGTT 00 CGTT 00 CGTT 00 CGTT 00 CGTT 00 CGTT 00 CGGG 00 CGTT 00	TTA TCT' L TCT' L TCA: H ACT' L TAT( L CGA: D CGGA:	CGT 333( 1GAJ E 339: 339: 348: 1CT( L 357: 366: 2AT: 1 375: 1CC( P 384: N 39: 39: 39: 39: 39: 39: 39: 39:	IAA OO ACG' R OO G G G G G G G G G G G G G	GCAT TCAC H GGAP E TGAP E CTGG W GGATC G M CCAAC N	333 2756 W 344 340 35 27 35 27 35 26 27 37 37 36 27 37 37 37 37 37 37 37 37 37 37 37 37 37	310 GCA Q 0 CGA E 190 CGA CGA S 0 CGA C CGA C CCA C C CCA C C C C C C C C	GAAAA K AGTT V CTGC W FGAT D CTGC R CATC S CAGC S	ATC R IGA D STC S S CGA D CGT V CGA D	GGT 3320 CCCC P 3410 CAG S 3500 ATT L 3550 CATT L 3550 C CATT L 3680 C TCC C C S 3770 C TCG C S 3860 G G G G G G G G G G G G G G G G G G G	TGC: GGT( V) CTCGJ R DCTC E OCTC F 0 GGA D 0 0 CTT F	AGA GGT( V ACT( L GGT/ V SAT1 I GCG7 R SCT( L TCG( R TCG7 V	3333 342 342 342 35TT 2 342 35TT 3 3 3 3 3 3 3 3 3 3 3 3 3	CA 30 AAA K 20 CA S CA S CA CA S CA CA S CA CA S CA CA CA CA CA CA CA CA CA CA
	R GCCT AACG R GTCA H CAGT V TTTC S GCGA CTGG G CAAAA N CTCG	CCA Q - 3: GTG S S S S S S S S S S S S S S S S S S	GCAT: H :: GATG( GATG( SD 340 CTTTI F 1 430 GGATG( 520 CCACT H V 610 GGATG( CCCACT H V 610 GGCTG( CCCACT H V 610 GCCTG( CCCACT H V 610 GCCTG( CCCACT H V 610 GCCTG( CCCACT H V 610 GCCTG( CCCACT H V 700 GCCTG( CCCACT H V 700 GCCTG( CCCACT H V 700 GCCTG( CCCACT H V 700 GCCTG( CCCACT H V 700 GCCTG( CCCACT H V 700 GCCTG( CCCACT H V 700 GCCTG( CCCACT H V 700 GCCTG( CCCACT H V 700 GCCTG( CCCACT H V 700 GCCCACT H V 700 GCCCACT H V 700 GCCCACT H V 700 GCCACT H V 700 CCCACT H V 700 CCCACT H V 700 CCCCACT H V 700 CCCACT H V CCCACT H V CCC	CTTAN AATAM CTTAN CTTAN CTTAN CTTAN CTTAN CTTAN CTTAN CTTAN CCCCCCCCCC	CGCC P Q 3326 NCAT M 3335 NTTT F 33444 W 3530 W 3530 W 3530 W 3530 W 3530 W 3530 W 3530 W 3530 W 33500 W 33530 W 33530 W 33530 W 33500 W 33530 W 33500 W 33500 W 33500 W 33500 W 33500 W 33500 W 33500 W 33500 W 33500 W 3300 W 3 W 3	GAA K - 0 GGA E 0 TAT I 0 GCA Q 0 GCCA V 0 GCCA V 0 CCT V 0 0 CCT V 0 0 TTAC F 0 0 TTAC F 0 0 TTAC F	AGA' D ATA' Y TGA' D GGT( D GGT( G G CGGC G G G CGGC G G CGC G G T CGGC G G T CGGC G G T T G G T	TGAN E 327 CCAN Q 336 CCCC P 345 CCCC S 354 CCCCC A 354 CCCCC A 354 CCCCC A 354 CCCCC A 354 CCCCCC A 354 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCC A 356 CCCCCCC A 356 CCCCCCC A 356 CCCCCCC A 356 CCCCCCC A 356 CCCCCCC A 356 CCCCCCC A 356 CCCCCCCC A 356 CCCCCCCC A 357 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ATAI * 70 CCAC 10 50 50 50 50 50 50 50 50 50 5	CACI T S CGGC G SCTC L SCTC L CCTC P CGAT D	322 ICT: I 333 ICT: I 335 ICC: P 340 CCCC P 35 SATO P 35 SATO P 35 SATO P 35 SATO P 35 SATO P 35 SATO P 35 SATO P 40 SATO SATO P SATO P SATO SATO SATO SATO SATO SATO SATO SATO	2800 IRAA N 370 AGA D 370 SGCGJ R 550 SGCGJ R 550 SGCGJ R 540 SGCGJ D 320 SGCGJ D 320 SGCGJ D 320 SGCGJ O SGCG O SGCG O SGCG O SGCG SGCG S S S S S S S S S S S S S S S	CTGG W CGAG E CGAA E CCGA CCG CCG CCTG G CCTG G CCTG G CCTG C CCTG C CCTG C CCTG C CCTG C CCTG C CCTG C CCTG C CCA C CCA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C C CAA C	329 CCCGG P D 338 TTGGC L A 347 AGCTY S Y 356 F R 365 GACGT F R 365 GACGT D V 374 TTACA L Q 383 TACGO Y A 392 CAACC Q R	CGCA 00 NTTT F 10 CGGG CGGG 0 CGGG 0 CGGT CGGT 0 CGGTT CGGT CGGT CGGT CGGT CGG	TTA TCT L TCT L TCA H TCA H TCA H CGA CGA CGA L GGT	CGT 333( IGAJ E 339( GA 348( ICT( I 357( ACCC) P 366( CAT( I 375( CCC) P 384( A A 375( CCC) P 386( A A A A A A A A A A A A A	TAA DO ACG R DO SAT CG CG CG CG CG CG CG CG CG CG	GCAT TCAC H GGAA E GGAA G GATC M CCAAC N CCAAC	333 TGC W 344 AGC 34C AGC T 35 GC T 37 GCC A 38 GCC A 32 GCC A 32 GCC A 32 GCC A 34 A 34 GCC A 34 A 34 A 34 A 34 A 34 A 34 A 34 A	310 3CA Q Q 100 CGA E 190 CGA S 80 FAT S 60 CATC I 50 CGGJ G CATC I 50 CGGJ CAAC Y 0 20 CAAC Y 0 20 CGA CAAC Y 0 20 CGA S CAAC Q 0 20 CGA S CAAC Q 0 20 CGA S CAAC N 5 CAAC Q 0 20 CGA S CAAC N 5 CAAC S CAAC N 5 CAAC N 5 CAAC S CAAC N 5 CAAC N 5 CAAC S CAAC N 5 CAAC N 5 CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S CAAC S C S	GAAL GAAL K AGTI V CTGO W IGAI D CACGI I CACGI S CAGO S CAGO S	ATC R GGA D CGA D CGA D CGA CCA	GGT 3320 CCCC P 3410 CAG S 3500 AATTI L 3590 TCT L 3590 TCT R 3590 TCT R 3680 G G G G G G G G G G G G G G G G G G G	TGC: GGT( V ) GGT( R ) GGA CTG( W ) CTG( W ) CTG( C TGC) C TGC C T C C C T C C C C C C C C C C C	AGA GGT(V V ACT(L GGT/J V SATT I GGT/S R SCG/J R SCG/J R SCG/J TCG <sup>1</sup> R TCG <sup>1</sup> R TCG <sup>1</sup> R	3333 3423 3433	CA 30 AAA K 20 CA CA 20 CA CA CA CA CA CA CA CA CA CA
	R GCCT AACG R GTCA H CAGT V TTTC S GCGA CTGG G CAAAA N CTCG	CCA Q Q - 33: CGGG G G 34 CCA Q 35: CGGG N 36 TGT: V 37 TGT: V 37 TGT: V 37 TGT: CGC T AAAAC K 33 CGGC Q 34 CCA S C S C	GCAT: H :: - 1 (41) 250 GATG( 50) 340 CTTTI F 1 430 CTTTI F 1 430 CCACC F 0 610 CCACC F 0 610 CCACC F 0 700 CCACC F 0 790 CCACC F 1 1 380 GCGGGC R 1 970 FCAT F 1 970	CTTAN AATAM CTTAN CTTAN CTTAN CTTAN CTTAN CTTAN CTTAN CTTAN CCCCCCCCCC	CGCC P Q 3326 ACAT M 3335 ATGAT F 3344 ATGA 3353 ATGA X 3353 ATGA X 3353 ATGA X 3362 CGGA X X 3380 ATAT I 3398 CGGA D	GAA K - OGGA E OTAT I OGCA Q OGCA Q OGCA Q O OCC V O ACAA Q O O TTAG S O TTAG S O TGT V	AGA' D ATA' Y TGA' D GGT( D GGT( G G CGGC G G G CGGC G G CGC G G T CGGC G G T CGGC G G T T G G T	TGAN E 327 CCAN Q 336 CCCCC P 3345 CCCCC A 363 CCCCC A 354 CCCCC A 354 CCCCC A 354 CCCCC A 354 CCCCC A 354 CCCCCC A 354 CCCCCC A 355 CCCCCCCCCC	ATAI * * 70 CCCC 50 CCAC 1 50 CCAC 1 50 CCAC 1 50 CCCC 1 50 CCCC 1 50 CCCC 5 5 CCCC 5 5 5 5 5 5 5 5 5 5 5	CACI T S CGGC G SCTC L SCTC L CCTC P CGAT D	ATGO 32 ICT: J 33 ICT: J 33 ICC: P 34 ICC: P 34 ICC: D 37 ICC: D 37 ICC: P 36 ICC: ICC: P 36 ICC: P I	280 IAAC N 370 AGAC D 370 AGAC D 350 350 350 350 350 350 350 350 350 350	CTGG W CGAG E CGAA E CCGA CCG CCG CCTG G CCTG G CCTG G CCTG C CCTG C CCTG C CCTG C CCTG C CCTG C CCTG C CCTG C CCA C CCA C C CAA C C CAA C C CAA C C CAA C C CAA C C CAA C C C CAA C	3225 3225 3256 3256 3267 3387 3374 3356 3356 3356 3356 3356 3356 3356 3356 3356 3356 3356 3457 247 347 347 347 347 347 347 347 3	CGCA 00 VTTT F 10 CGGG CGGG CGGG CGGTT CGGT CGGT CGGT CGGT CGGT CGGT CGGT CGGT CGGT	TTA TCT L TCT L TCA H TCA H TCA H CGA CGA CGA L GGT	CGT 333( IGAJ E 339( GA 348( ICT( I 357( ACCC) P 366( CAT( I 375( CCC) P 384( A A 375( CCC) P 386( A A A A A A A A A A A A A	TAA DO ACG R DO SAT CG CG CG CG CG CG CG CG CG CG	GCAT TCAC H GGAA E CTGG W GATC GATC M CCAAC N CCAAC	333 TGC W 344 AGC 34C AGC T 35 GC T 37 GCC A 38 GCC A 32 GCC A 32 GCC A 32 GCC A 34 A 34 GCC A 34 A 34 A 34 A 34 A 34 A 34 A 3 3 3 3	310 3CA Q Q 100 CGA E 190 CGA S 80 FAT S 60 CATC I 50 CGGJ G CATC I 50 CGGJ CAAC Y 0 20 CAAC Y 0 20 CGA CAAC Y 0 20 CGA S CAAC Q 0 20 CGA S CAAC Q 0 20 CGA S CAAC N 5 CAAC Q 0 20 CGA S CAAC N 5 CAAC S CAAC N 5 CAAC N 5 C C C C C C C C C C C C C C C C C	GAAL GAAL K AGTI V CTGO W IGAI D CACGI I CACGI S CAGO S CAGO S	ATC R GGA D CGA D CGA D CGA CCA	GGT 3320 CCCC P 3410 CAG S 3500 AATTI L 3590 TCT L 3590 TCT R 3590 TCT R 3680 G G G G G G G G G G G G G G G G G G G	TGC: GGT( V ) GGT( R ) GGA CTG( W ) CTG( W ) CTG( C TGC) C TGC C T C C C T C C C C C C C C C C C	AGA GGT(V V ACT(L GGT/J V SATT I GGT/S R SCG/J R SCG/J R SCG/J TCG <sup>1</sup> R TCG <sup>1</sup> R TCG <sup>1</sup> R	3333 3423 3433	CA 30 AAA K 20 CA CA 20 CA CA CA CA CA CA CA CA CA CA
	R GCCT AACG R GTCA H CAGT V TTTC S GCGA E CTGG G CAAAA N CTCG R AGCA	CCA Q - 3: GTG GTG G G G G G G G G G G G G G G G	GCAT: H :: GATG( GATG( SD 340 CTTTI F 1 430 GGATG( 520 CCACT H 5 610 GGCTG( L ( 700 GCCCCC I 1 380 GGCTG( C I 1 380 GCCGCC F ( F ) 790 F ) F ( 790 F ) F ( F ) F ) F ) F ) F ) F ) F ) F ) F ) F )	CTTAI AATAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI CTTAI	CGCC P Q 3326 ACAT M 3335 ATTT F 3344 ATTT F 3344 ATTT F 3344 ATTT S 3353 CGGA CGGA ATTT I 3389 TGAT I 3398 A CGTT A Q 407 A	GAA K - O GGGA E O TAT I O GGCA Q O CGTA Q O CGTA Q O CGTA Q O CGTA V O D ACAA Q O CGTA V O D TAT I O O CGTA V O O CGTA V O O CGTA O O O C C O O C O O C O O O O O O O O	AGA' D ATA( Y TGA( D GGT( C GGC G G G G G G G G G C C G G G C C G G C C G G C C G G C C C G C C C G C C C C G C	TGAN E 327 CCAM Q 336 CCCC P 3345 CCCC S 354 CCCC C S 354 CCCC C 354 CCCC C 354 CCCCC P 372 CCCC A 353 CCCCC P 372 CCCM C 354 CCCCC P 354 CCCCC P 354 CCCCC P 354 CCCCC P 354 CCCCC P 354 CCCCC P 354 CCCCCC P 354 CCCCCC P 354 CCCCCC P 354 CCCCCC P 354 CCCCCC P 354 CCCCCC P 354 CCCCCC P 354 CCCCCCCCCCC P 354 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ATAI * 70 ACT( L 50 50 50 50 50 1 50 50 50 1 50 50 50 50 50 1 50 50 50 50 50 50 50 50 50 50	CACT T S CCGGC G G CCTC L CCTC L CCTC L CCTC L CCTC C L CCTC C L CCTC C CCCC C C CCCCC C C CCCCC C C CCCCCC	ATGO 32 TCT: 1 33 TCC: P 34 CCC: P 35 CCC: P 36 CCC: P CCC: CCC: P CCC: CCC: P CCC: P CCC: CCC: P CCC: CCC: P CCC: CCC: P CCC: P CCC: CCC: P CCC: CCC: P CCC: P CCC: CCC: P CCC: CCC: P CCC: P CCC: CCC: P CCC: P CCC: CCC: P CCC: CCC: P CCC: CCC: P CCC: P CCC: CCC: P CCC: CCC: P CCC: P CCC: CCC: P CCC: CCC: P CCC: P CCC: CCC: P CCC: C	280 1770 1770 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1600 1770 1600 1770 1600 1770 1600 1770 1600 1770 1600 1770 1600 1770 1600 1770 1600 1770 1600 17700 1770 1770 1770 1770 1770 1770 1770 1770 1770 1770	CTGG W CGAG E CGAA E CCCG C CCCG G CCCG G CCCG G CCCG G CCCG G CCCG G CCCG G CCCG G CCCG G CCCG G CCCG G CCCG G CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C C CCCG C	329 CCCCGJ P D 338 TTGGCL A 347 AGCTY S Y 356 TTCCC F R 365 S GACGT D V 374 TTCCC F R 365 S CACCC Q R 400 CCACC Q R 400 CCACC Q R 400 CCACC C Q R 400 CCACC C	CGCA 00 NTTT F 10 CGGG 0 CGGG 0 CGGT CGGT 0 CGGT 0 CGGTT CGGT CGGT CGGT CGGT CGGT CGGT CGGT CGGT CGGT	TTA TCT L TCT L TCA H TCA H TCA H CGA CGA CGA L GGT	CGT 333( IGAJ E 339( GA 348( ICT( I 357( ACCC) P 366( CAT( I 375( CCC) P 384( A A 375( CCC) P 386( A A A A A A A A A A A A A	TAA DO ACG R DO SAT CG CG CG CG CG CG CG CG CG CG	GCAT TCAC H GGAA E CTGG W GATC GATC M CCAAC N CCAAC	333 TGC W 344 AGC 34C AGC T 35 GC T 37 GCC A 38 GCC A 32 GCC A 32 GCC A 32 GCC A 34 A 34 GCC A 34 A 34 A 34 A 34 A 34 A 34 A 3 3 3 3	310 3CA Q Q 100 CGA E 190 CGA S 80 FAT S 60 CATC I 50 CGGJ G CATC I 50 CGGJ CAAC Y 0 20 CAAC Y 0 20 CGA CAAC Y 0 20 CGA S CAAC Q 0 20 CGA S CAAC Q 0 20 CGA S CAAC N 5 CAAC Q 0 20 CGA S CAAC N 5 CAAC S CAAC N 5 CAAC N 5 C C C C C C C C C C C C C C C C C	GAAL GAAL K AGTI V CTGO W IGAI D CACGI I CACGI S CAGO S CAGO S	ATC R GGA D CGA D CGA D CGA CCA	GGT 3320 CCCC P 3410 CAG S 3500 AATTI L 3590 TCT L 3590 TCT R 3590 TCT R 3680 G G G G G G G G G G G G G G G G G G G	TGC: GGT( V ) GGT( R ) GGAACT( W ) CTG( W ) CTG( W ) CTG( C TGC) C TGC C TGC C TGC C C T C C C C C C C C	AGA GGT(V V ACT(L GGT/J V SATT I GGT/S R SCG/J R SCG/J R SCG/J TCG <sup>1</sup> R TCG <sup>1</sup> R TCG <sup>1</sup> R	3333 3423 3433	CA 30 AAA K 20 CA CA 20 CA CA CA CA CA CA CA CA CA CA

FIG. 2-Continued.

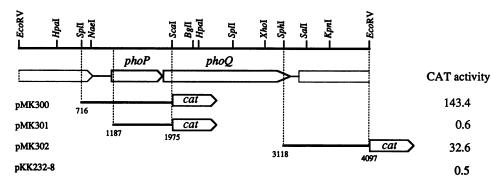


FIG. 3. Assay of the promoter activity for the *phoP* and *phoQ* genes. Bold lines indicate the chromosomal fragments of the *phoP-phoQ* region in the operon fusion plasmids. Arrows indicate the direction of the *cat* gene orientation. Chloramphenicol acetyltransferase (CAT) activity in extracts of sonicated cells of ANCK10 carrying plasmids with the operon fusions was assayed.

regions. Neither flanking ORF appears to be transcriptionally coupled to the phoP-phoQ operon. Analysis of promoter activity in the cloned phoP-phoQ

region by operon fusions. To locate the promoter of the phoP and phoQ genes, we constructed a series of operon fusion plasmids. DNA fragments containing nucleotides 716 to 1975, 1187 to 1975, or 3118 to 4097 were recloned into a promoter cloning vector, pKK232-8 (3), to construct fusions with the cat gene, and the resultant plasmids (pMK300, pMK301, and pMK302) were introduced into strain ANCK10. The promoter activities of the cloned DNA fragments in the fusion plasmids were measured by monitoring the chloramphenicol acetyltransferase in the extracts of the cells carrying these plasmids (Fig. 3). ANCK10 carrying pMK300 or pMK302 showed high levels of the enzyme activity, while ANCK10 carrying pMK301 showed no activity. This result suggests that a common promoter for phoPand phoQ is located upstream of the phoP gene and that another promoter for the gene distal to phoQ is located downstream of the phoQ gene. From these results, we conclude that the phoP and phoQ genes constitute a single operon. These results are also consistent with the hypothesis that the inverted sequences found upstream and downstream of the *phoP-phoQ* operon function as terminators.

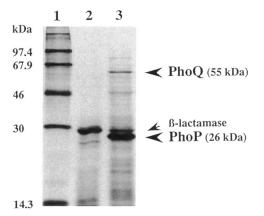


FIG. 4. Identification of the products of phoP and phoQ by the maxicell method. An autoradiogram of <sup>35</sup>S-labeled proteins produced in maxicells with  $phoP^+$ - $phoQ^+$  plasmid, which carries nucleotides 909 to 3232 of the chromosomal fragment (lane 3), and with pUC18 (lane 2) is shown. Isotope-labeled protein size markers (Amersham Japan, Tokyo) were run on lane 1.

Identification of the phoP and phoQ gene products. To identify the phoP and phoQ gene products, the plasmidencoded proteins were labeled with [ $^{35}S$ ]methionine by the maxicell method. Two proteins with approximate molecular masses of 26 and 55 kDa were found in the maxicells carrying the phoP<sup>+</sup>-phoQ<sup>+</sup> plasmids (Fig. 4, lane 3). The 46-kDa protein found in the maxicells carrying the phoP<sup>+</sup>phoQ<sup>+</sup> plasmids may be the degradation product of PhoQ. In the maxicells carrying the phoP<sup>+</sup> plasmids, only the 26-kDa protein was detected (data not shown). These proteins were not detected in the cells carrying the vector plasmid, pUC18 (Fig. 4, lane 2). The molecular masses of these products agreed very well with those predicted from the nucleotide sequences of the phoP and phoQ genes.

The phoP-phoQ operon of S. typhimurium has been shown to be necessary for Salmonella virulence (5). E. coli K-12 has no such virulence phenotype and lacks the phoN gene, but it has the phoP-phoQ operon. We recently constructed a  $\Delta(phoP-phoQ)$  strain, compared the whole-cell proteins of the strain with those of the wild-type strain by two-dimensional gel electrophoresis, and found that expression of at least 50 protein species changed as a result of this deletion mutation (9). Therefore, the phoP and phoQ genes of E. coli appear to regulate as many genes as their counterparts in S. typhimurium (18). Further work will be necessary to understand the physiological roles of the phoP-phoQ regulon in E. coli.

#### ACKNOWLEDGMENTS

We thank B. N. Ames and K. Kutsukake for the gifts of the bacterial strains. We are grateful to E. A. Groisman and coworkers for agreeing to publish our two related papers in the same issue.

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

#### REFERENCES

- 1. Amemura, M., K. Makino, H. Shinagawa, A. Kobayashi, and A. Nakata. 1985. Nucleotide sequence of the genes involved in phosphate transport and regulation of the phosphate regulon in *Escherichia coli*. J. Mol. Biol. 184:241–250.
- Bolivar, F., R. L. Rodriguez, P. J. Green, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2:95–113.
- 3. Brosius, J. 1984. Plasmid vectors for the selection of promoters. Gene 27:151-160.
- 4. Feinberg, P. I., and B. Vogelstein. 1984. A technique for radiolabeling DNA restricting endonuclease fragments to high specific activity: addendum. Anal. Biochem. 137:266-267.

- Fields, P. I., E. A. Groisman, and F. Heffron. 1989. A Salmonella locus that controls resistance to microbicidal proteins from phagocytic cells. Science 243:1059–1062.
- Groisman, E. A., E. Chiao, C. J. Lipps, and F. Heffron. 1989. Salmonella typhimurium phoP virulence gene is a transcriptional regulator. Proc. Natl. Acad. Sci. USA 86:7077-7081.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351– 359.
- 8. Kasahara, M., A. Nakata, and H. Shinagawa. 1991. Molecular analysis of the *Salmonella typhimurium phoN* gene, which encodes nonspecific acid phosphatase. J. Bacteriol. 173:6760–6765.
- 9. Kasahara, M., T. Ohara, A. Nakata, and H. Shinagawa. Unpublished data.
- 10. Kier, L. D., R. M. Weppelman, and B. N. Ames. 1977. Regulation of two phosphatases and a cyclic phosphodiesterase of *Salmonella typhimurium*. J. Bacteriol. 130:420-428.
- 11. Kier, L. D., R. M. Weppelman, and B. N. Ames. 1979. Regulation of nonspecific acid phosphatase in *Salmonella: phoN* and *phoP* genes. J. Bacteriol. 138:155–161.
- 12. Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell 50:495-508.
- Lee, T.-Y., K. Makino, H. Shinagawa, M. Amemura, and A. Nakata. 1989. Phosphate regulon in members of the family *Enterobacteriaceae*: comparison of the *phoB-phoR* operons of *Escherichia coli*, *Shigella dysenteriae*, and *Klebsiella pneumoniae*. J. Bacteriol. 171:6593-6599.
- Makino, K., H. Shinagawa, and A. Nakata. 1984. Cloning and characterization of the alkaline phosphatase positive regulatory gene (*phoM*) of *Escherichia coli*. Mol. Gen. Genet. 195:381–390.
- Messing, K., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. 9:309–321.
- 16. Miller, J. H. 1972. Experiments in molecular genetics. Cold

Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

- Miller, S. I., A. M. Kukral, and J. J. Mekalanos. 1989. A two-component regulatory system (*phoP phoQ*) controls Salmonella typhimurium virulence. Proc. Natl. Acad. Sci. USA 86: 5054–5058.
- Miller, S. I., and J. J. Mekalanos. 1990. Constitutive expression of the PhoP regulon attenuates *Salmonella* virulence and survival within macrophages. J. Bacteriol. 172:2485–2490.
- 19. Oishi, M., and S. D. Cosloy. 1972. The genetic and biochemical basis of the transformability of *Escherichia coli* K12. Biochem. Biophys. Res. Commun. 49:1568–1572.
- Pulkkinen, W. S., and S. I. Miller. 1991. A Salmonella typhimurium virulence protein is similar to a Yersinia enterocolitica invasion protein and a bacteriophage lambda outer membrane protein. J. Bacteriol. 173:86–93.
- Ronson, C. W., B. T. Nixon, and F. M. Ausubel. 1987. Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell 49:579–581.
- 22. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Sancar, A., R. P. Wharton, S. Seltzer, B. M. Kacinski, N. D. Clarke, and W. D. Rupp. 1981. Identification of the *uvrA* gene product. J. Mol. Biol. 148:45–62.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Shaw, W. V. 1975. Chloramphenicol acetyltransferase from chloramphenicol resistant bacteria. Methods Enzymol. 43:737– 743.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.
- 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.