Nucleotide Sequence and Functional Analysis of the *meta*-Cleavage Pathway Involved in Biphenyl and Polychlorinated Biphenyl Degradation in *Pseudomonas* sp. Strain KKS102

YUTAKA KIKUCHI,¹* YUJI YASUKOCHI,¹ YUJI NAGATA,¹ MASAO FUKUDA,² and MASAMICHI TAKAGI¹

Department of Agricultural Chemistry, The University of Tokyo, Bunkyo-ku, Tokyo 113,¹ and Department of Bioengineering, Nagaoka University of Technology, Kamitomioka, Nagaoka, Niigata 940-21,² Japan

Received 15 February 1994/Accepted 10 May 1994

Pseudomonas sp. strain KKS102 is able to degrade biphenyl and polychlorinated biphenyls via the meta-cleavage pathway. We sequenced the upstream region of the bphA1A2A3BCD (open reading frame 1 [ORF1]) A4 and found four ORFs in this region. As the deduced amino acid sequences of the first, second, and third ORFs are homologous to the meta-cleavage enzymes from Pseudomonas sp. strain CF600 (V. Shingler, J. Powlowski, and U. Marklund, J. Bacteriol. 174:711–724, 1992), these ORFs have been named bphE, bphG, and bphF, respectively. The fourth ORF (ORF4) showed homology with ORF3 from Pseudomonas pseudoalcaligenes KF707 (K. Taira, J. Hirose, S. Hayashida, and K. Furukawa, J. Biol. Chem. 267:4844–4853, 1992), whose function is unknown. The functions of meta-cleavage enzymes (BphE, BphG, and BphF) were analyzed by using crude extracts of Escherichia coli which expressed the encoding genes. The results showed that bphE, bphG, and bphF encode 2-hydroxypenta-2,4-dienoate hydratase, acetaldehyde dehydrogenase (acylating), and 4-hydroxy-2-oxovalerate aldolase, respectively. The biphenyl and polychlorinated biphenyl degradation pathway of KKS102 is encoded by 12 genes in the order bphEGF (ORF4)A1A2A3BCD (ORF1)A4. The functions of ORF1 and ORF4 are unknown. The features of this bph gene cluster are discussed.

The aerobic degradation pathways of various aromatic compounds (benzene, toluene, xylene, phenol, naphthalene, biphenyls, polychlorinated biphenyls [PCBs], etc.) have been studied in many bacteria (28, 32). The initial conversion steps of these aromatic compounds are carried out by different enzymes, and the compounds are transformed to catecholic intermediates. These are then cleaved by dioxygenase and transformed to Krebs cycle intermediates. Bacterial aromatic ring cleavage pathways are classified into two groups, the *ortho*-cleavage pathway, catecholic compounds are transformed to the common intermediate 3-oxoadipate enol-lactone, which is further converted to succinate and acetyl coenzyme A (acetyl-CoA) (32). In the *meta*-cleavage pathway, they are transformed to pyruvate and a short-chain aldehyde (32).

Detailed study of the *meta*-cleavage pathway was carried out in TOL plasmid pWW0 to determine its gene organization, regulation of expression, and enzymatic functions (2, 3). The *meta*-cleavage operon is composed of 13 genes which encode enzymes for the conversion of methylbenzoate to pyruvate and acetaldehyde (10). Another study of the *meta*-cleavage pathway was carried out in plasmid NAH7, which encodes the enzymes required for the degradation of naphthalene via salicylate (25, 35). Plasmid NAH7 contains two operons, the *nah* and *sal* operons. The *meta*-cleavage pathway genes of the *sal* operon are similar to those of the TOL plasmid, and DNA sequences of the two operons are homologous (1, 12). Recently, another *meta*-cleavage pathway was characterized in *Pseudomonas* sp. strain CF600, which is able to grow on phenol, cresols, or 3,4-dimethylphenol (27). The *meta*-cleavage pathway of CF600 is composed of nine genes (*dmpQBCDEF* *GHI*) which encode enzymes for the conversion of catechol to acetyl-CoA (27). Polypeptide analysis and nucleotide sequence determination of these genes showed that the *meta*-cleavage pathway of CF600 is very closely related to that in the TOL plasmid (27). Moreover, the nucleotide sequence data showed a new operon-encoded *meta*-cleavage pathway enzyme that is able to transform acetaldehyde to acetyl-CoA (27).

A mixed culture which exhibited a capability to degrade highly chlorinated PCBs was isolated from soil in our laboratory (17). This mixed culture was composed of *Pseudomonas* sp. strain KKS102 and *Pseudomonas fluorescens* KKL101. These two strains had a symbiotic relationship, and KKS102 played a major role in degradation of biphenyl and PCBs (17).

We have already cloned from KKS102 and sequenced eight genes, bphA1A2A3BCD (open reading frame 1 [ORF1]A4) for biphenyl and PCB degradation (7, 16, 18). (The nucleotide sequence data for bphA1A2A3B have been submitted to the DDBJ, EMBL, and GenBank nucleotide sequence databases under accession number D17319.) These eight genes were found to be clustered (16). This bph gene cluster encodes enzymes for conversion of PCB to 2-hydroxypenta-2,4-dienoate and chlorobenzoic acid (16). The biphenyl and PCB catabolic pathway in KKS102 is shown in Fig. 1. In the first catabolic step, two atoms of oxygen are inserted at carbon positions 2 and 3 by biphenyl dioxygenase (encoded by the bphA region) to produce dihydrodiol. Biphenyl dioxygenase is a multicomponent enzyme which consists of a large and a small subunit of iron-sulfur proteins (the products of bphA1 and bphA2, respectively), ferredoxin (the product of bphA3), and ferredoxin reductase (the product of bphA4). 2,3-Dihydrodiol is converted to 2,3-dihydroxybiphenyl by 2,3-dihydrodiol dehydrogenase (the product of bphB). 2,3-Dihydroxybiphenyl is cleaved at the meta position by 2,3-dihydroxybiphenyl dioxygenase (the product of bphC) to yield 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate, which is transformed into corresponding

^{*} Corresponding author. Mailing address: Department of Agricultural Chemistry, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan. Phone: 03(3812)2111, ext. 5178. Fax: 03(3812)9246.



FIG. 1. Proposed catabolic pathway for degradation of biphenyl and PCBs in KKS102. Compounds: 1, biphenyl; 2, 2,3-dihydroxy-4-phenylhexa-4,6-diene (dihydrodiol compound); 3, 2,3-dihydroxybiphenyl; 4, 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate; 5, 2-hydroxypenta-2,4-dienoate; 6, 4-hydroxy-2-oxovalerate; 7, acetaldehyde; 8, acetyl-CoA. Genes encode the following enzymes: *bphA1A2*, iron-sulfur proteins; *bphA3*, ferredoxin; *bphA4*, ferredoxin reductase; *bphB*, dihydrodiol dehydrogenase; *bphC*, 2,3-dihydroxybiphenyl dioxygenase; *bphD*, 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase; *bphE*, 2-hydroxypenta-2,4-dienoate hydratase; *bphF*, 4-hydroxy-2-oxovalerate aldolase; *bphG*, acetaldehyde dehydrogenase (acylating). The chlorobenzoic acid, which is produced by the product of *bphD*, seems to be degraded by *P. fluorescens* KKL101 (15).

chlorobenzoic acid and 2-hydroxypenta-2,4-dienoate by 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (the product of *bphD*).

In this study, we describe nucleotide sequences and functions of *meta*-cleavage genes in KKS102. This work elucidates the gene organization and pathway for biphenyl and PCB degradation in KKS102. We also demonstrate the features of the *bph* gene cluster in KKS102 compared with the *bph* gene cluster in *Pseudomonas pseudoalcaligenes* KF707 (8, 29).

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture conditions. The bacterial strains and plasmids used in this study are listed in Table 1. *Pseudomonas* sp. strain CF600 (27) was a gift from

Victoria Shingler (University of Umeå, Umeå, Sweden). *Pseudomonas* sp. strain KKS102 was grown at 30°C in a mineral salts medium (pH 7.0) (17) containing Casamino Acids (0.5%) and biphenyl (0.1%). *Escherichia coli* MV1190 was grown in Luria broth (24) at 37°C. Ampicillin (50 μ g/ml) was added for selection of a plasmid encoding β -lactamase in *E. coli*.

DNA manipulation and transformation and the isolation of plasmids were carried out as described by Sambrook et al. (24). Overexpression plasmids were constructed from pAQN, which has the same structure as pAQI (pMB9 replicon, *lacI*^q, aqualysin I gene of *Thermus aquaticus* [*aqn*]; carries a gene encoding ampicillin resistance [Ap^r]) (30) except for in the aqualysin I-coding region. pAQN was digested with *Eco*RI and *Hind*III to replace the 1.8-kb aqualysin I-coding fragment with another

TABLE 1. Bacterial strains and plasmids used

Strain or plasmid	Relevant characteristics	Reference
Pseudomonas strains		
KKS102	Biphenyl and PCB degrader	17
CF600	Phenol and 3,4-dimethyl phenol degrader	27
<i>E. coli</i> MV1190	Δlac -proAB thi supE Δsrl -recA306::Tn10 F' traD36 proAB lacI $^{9}Z\Delta M15$	34
Plasmids		
pUC18	pMB9 replicon, Ap ^r	20
pKH20	pUC18 carrying 18-kb BamHI-BamHI fragment, Ap ^r	18
pUC119	pMB9 replicon, Ap ^r	33
pKH200	pUC119 carrying 6.4-kb SmaI-SmaI fragment of pKH20, Apr (Fig. 2)	16
pKH205	pUC119 carrying 4.1-kb <i>Hin</i> cII- <i>Eco</i> T22I fragment of pKH20, Ap ⁷ (Fig. 2)	This study
pAON	pMB9 replicon, <i>lacI</i> ^q , <i>aan</i> , Ap ^r	30
pAQN1	pMB9 replicon, <i>lac1</i> ^q , Ap ^r	This study
pKH402	pAQN carrying 3.4-kb <i>Eco</i> T22I- <i>Eco</i> T22I fragment in place of <i>aqn</i> , Ap ^r (Fig. 2)	This study
pKH403	pAON carrying 0.9-kb <i>Eco</i> T22I- <i>Kpn</i> I fragment in place of <i>aqn</i> , Ap ^r (Fig. 2)	This study
pKH404	pAQN carrying 1.1-kb NruI-NruI fragment in place of agn, Apr (Fig. 2)	This study
pKH405	pAQN carrying 1.7-kb SmaI-PvuII fragment in place of aqn, Apr (Fig. 2)	This study
рКН406	pAQN carrying 1.4-kb StuI-StuI fragment in place of aqn, Apr (Fig. 2)	This study



FIG. 2. Structures of plasmids. The directions of transcription by the *lac* promoter and *tac* promoter of the expression vectors are indicated by short arrows. The locations of the *bph* genes are also shown. See the legend to Fig. 1 for gene designations. B, *Bam*HI; Hc, *Hinc*II; EI, *Eco*T221; N, *Nru*I; K, *Kpn*I; S, *Sma*I; St, *Stu*I; Pv, *Pvu*II; P *lac*, *lac* promoter; Ptac, *tac* promoter.

DNA fragment. The resulting plasmid contained the origin of replication from pUC18 and the *lac1*⁴ gene and expressed a gene on an inserted DNA fragment under the control of the *tac* promoter. pAQN1 was constructed as a control vector by removing the aqualysin I gene and self-ligating. Structures of the pKH series of plasmids are shown in Fig. 2.

Chemicals. Coenzymes (NAD, NADH, and acetyl-CoA) and lactate dehydrogenase were obtained from Sigma Chemical Co. and Wako Pure Chemical Industries Ltd., Osaka, Japan. All other chemicals were of the highest purity commercially available. Solutions of 2-hydroxypenta-2,4-dienoate were synthesized from $D_{,L}$ -allylglycine as described by Collinsworth et al. (4). The UV spectrum of the aqueous solutions of 2-hydroxypenta-2,4-dienoate showed a λ_{max} at 265 nm.

To prepare L-(S)-4-hydroxy-2-oxovalerate, L-(S)-4-methyl-2oxobutyrolactone was synthesized from catechol by using crude extracts of phenol-grown *Pseudomonas* sp. strain CF600 as described by Shingler et al. (27). Solutions of L-(S)-4-hydroxy-2-oxovalerate were prepared by mild alkaline hydrolysis of L-(S)-4-methyl-2-oxobutyrolactone (5).

Nucleotide sequence determination. Unidirectional deletion mutants of pKH205 were constructed by using exonuclease III and mung bean nuclease as described by Henikoff (13), with a slight modification. DNA sequencing of the region upstream of the *bph* gene cluster for both strands was carried out with a *BcaBEST* dideoxy nucleotide sequencing kit (Takara Shuzo Co., Kyoto, Japan). The nucleotide sequences were analyzed with GENETYX software (version 21.0; Software Development Co. Ltd., Tokyo, Japan).

Analysis of plasmid-encoded polypeptides. pKH402 and pKH406 were introduced into *E. coli* MV1190, and the transformants were cultivated in 10 ml of Luria broth. The cell density was monitored by measuring the turbidity of the culture at 550 nm. When the turbidity reached a value of 0.6, isopropyl- β -D-thiogalactopyranoside (IPTG) was added to a final concentration of 5 mM and cultivation was continued another 3 h. Cells from 1 ml of the culture were harvested by centrifugation at 15,000 × g for 10 min and washed with an equal volume of phosphate-buffered saline (NaCl [8.0 g/liter], KCl [0.2 g/liter], Na₂HPO₄ [1.44 g/liter], KH₂PO₄ [0.24 g/liter] [pH 7.4]). The cells were boiled at 100°C for 5 min in the loading buffer (0.125 M Tris-HCl [pH 6.8], 2% sodium dodecyl sulfate [SDS], 5% 2-mercaptoethanol, 7.5% glycerol, 0.005% bromophenol blue), and then 5µl of each lysate generated by this treatment was loaded directly onto a SDS-12.5% polyacrylamide gel. SDS-polyacrylamide gel electrophoresis (PAGE) was done by the method of Laemmli (19). Electrophoresis was performed at 20 mA until the tracking dye reached the bottom of the gel. Proteins were visualized by staining with Coomassie brilliant blue R250.

Crude extract preparation. All procedures for crude extract preparations were performed at 4°C. *E. coli* harboring pKH402, -403, -404, -405, or -406 was grown in 10 ml of Luria broth and induced with IPTG by the method described above. KKS102 was grown in mineral salts medium containing Casamino Acids and biphenyl at 30°C for 2 days. The cells were harvested, washed twice with ice-cold 10 mM phosphate (Na₂HPO₄-KH₂PO₄) buffer (pH 7.5), and resuspended in 50 mM phosphate (Na₂HPO₄-KH₂PO₄) buffer (pH 7.5) containing 2 mM dithiothreitol. The suspended cells were broken by sonication and centrifuged at 110,000 × g for 1 h. The resulting supernatants were used as crude extracts. These were kept on ice before use. The protein concentrations of the crude extracts, Richmond, Calif.).

Enzyme activity assays. 2-Hydroxypenta-2,4-dienoate hydratase activity was measured by the method of Harayama et al. (11). One unit of activity was defined as the amount of enzyme required to cause a decrease in A_{265} of 1.0/min.

Assays of 4-hydroxy-2-oxovalerate aldolase activity and acetaldehyde dehydrogenase (acylating) activity were performed by the method of Shingler et al. (27). 4-Hydroxy-2-oxovalerate aldolase activity was measured by monitoring the oxidation of NADH (A_{340}) in the presence of excess lactate dehydrogenase. One unit of activity was defined as the amount of enzyme required to catalyze the oxidation of 1 µmol of NADH per min (27). Acetaldehyde dehydrogenase (acylating) activity was measured by monitoring the coenzyme A-stimulated reduction of NAD⁺ (A_{340}). One unit of activity was defined as the amount of enzyme required to reduce 1 µmol of NAD⁺ per min (27).

Nucleotide sequence accession number. The nucleotide sequence data in this report have been submitted to the DDBJ, EMBL, and GenBank nucleotide sequence databases under accession number D16407.

RESULTS

Nucleotide sequencing and sequence analysis of meta-cleavage enzymes. An approximately 4-kb upstream region of the bphA1 was sequenced. Computer analysis identified four ORFs in this region, whose nucleotide and predicted amino acid sequences are shown in Fig. 3. The SWISS-PROT amino acid sequence data bank was searched for sequences similar to the predicted amino acid sequences of the four ORFs. The predicted amino acid sequences of the first, second, and third ORFs showed homology with DmpE (42.9% identity), DmpF (78.3% identity), and DmpG (80.8% identity) from Pseudomonas sp. strain CF600 (27), respectively. The first, second, and third ORFs were named bphE, bphG, and bphF, respectively. The deduced amino acid sequence of bphE was also homologous to that of XylJ (40.3% identity) from the TOL plasmid (14). The fourth ORF (designated ORF4) was similar to ORF3 (66.4% identity) from P. pseudoalcaligenes KF707, whose function is unknown (29).

SDS-PAGE analysis of polypeptide products. To identify the products of *bphE*, *bphG*, *bphF*, and ORF4, *E. coli* harboring pKH402 (containing *bphE*, *bphG*, *bphF*, and ORF4) or pKH406 (containing ORF4) was induced with IPTG, and total proteins were analyzed by SDS-PAGE (Fig. 4). Induced pro-

GTCGACCACGGCCAGCATCTGCGGGGGCCAACTGGTGTTCTTCGAGCAGGTGGCGAAAGCGCAGGATGCTGACCCGGTCAGGGATGCGCTC 90 HincII

AGA	.616	GAG	AGTO	CCC	GCG	AAC	TCG	CGI	ACA	GCG	TGG	STCI	CGI	ANCP	GGG	СТТС	TTCC	ATG	GCC	AGG	I'CAC	TCA		CGH	ACC	ACI	GII		180
~	AAT	GAA	TCGO	CAA	CAT	CGT		CAG	SCTC	-	CGG	CGG	GCC	GTCC	GGT	CTTG	GCGC	GAG	TAC	ATGO	CGGC	GCA	ATC	AAC	GAC	AGC	AAC	TCC	270
TGA		GGC	ACC/	ACC/	AGG TGT	CGT/		TCG	STCG	ACG				1110 2011	GAC	CGGA	GUUG GTTT	TGA		CAGO		TAGG		GGT		CG11	TGA	AGAG	360 450
GTO	CAT	CAA	TGT	GTC	AGC	CCG	CGC	AAG	TCG	CGG	AGI	GAP	GTC	GAGI	GAA	ACAT	AGGG	TTT	TCA	CGAC	GTGI	TTA	AAA	ATT	ATG	GCA	TGC	ATA	540
TAP	TCT	ATT	TTA	ГАТ	AAA	AAT	ACG	TAT	TTA	AAA	TCC	CAAT	TGC	GCAA	ACA	АТАТ	GTCA	AAA	AAG	ATTO	STCC	CAG	GGC	AAT	CAT	GCG	CCI	TCG	630
GTO	CTC	ATC	CAT	GCA	ATC	ACT	TGC	GTG	TCG	GCA	ACA	GCC	CT	гтсс	CTG	ATCG	GCGI	CGC	AAC	GACO	CGC	TCA	GTA	TCT	GGG	GAC	ACA	AGT	720
TT1	TGG	GTG	ССТО	GCA	TGA	GCGI	ATC	AGC	CAA	GCA	GGG	GCC	CA <u>GC</u>	<u>SC</u> TO	SACC	TTCG	C <u>TTG</u>	<u>C</u> AC	GAA	CTAT	rgc#	TGC	GAT	CAC	AAC	CGA	CGG	GCA	810
GTO	CGC	ACG	GAA	ATC	GCA	GAAI	ACC	ATG	SAAC	C <u>AA</u>	GGA		ATC	CTCC	SAAT	GAGC	AAAC	CAG	ATA	CGGI	TTTC	GAC	CAG	CGC	GGC	CAAG	TAP	AGC	900
															м	S	K P	D	Т	v	s	Т	s	Α	A	s	ĸ	A	
m cc		~~~~	~~~~	~~~~		~~~~	~~~	~~ >							(b)	phE)		NCC	TC N	TCC						TOT			000
TGC	.CGA n	UTT T.	6110 T.	V V	F	٨	A CGC	н	ICAC T	R R	v		v v		D P	V	RN		TGA	ונטנ ה	E E	K	חטטו	L	D	v	A	V V	990
'n	U	ŋ	Б	-	Ľ	~			•		•		•		•	•			-	U	2		D	2	5	•		-	
CGC	GGT	GCA	GGA	GAT	AAA	гасо	CGT	GCG	GGC	GCT	GAC	GGC	GGG	GACO	GACG	CCTG	TCGG	GCC	GCA	AGAT	rcga	GCI	GAC	стс	CGI	GAC	CGI	CCA	1080
A	v	Q	Е	I	N	т	v	R	A	L	т	A	G	R	R	L	s c	R	ĸ	I	G	L	т	s	v	т	v	Q	
AAA	GCA	GCT	CGGG	GGT	TGG	GCAG	GCC	GGA	ACTA	CGG	CAI	IGC I	GTI	rcgo	TGA	CATG	GCAC	GTA	CCG.	AGGO	STG7	AGA	GGI	CTC	:cc1	CAA	GGA	TGT	1170
ĸ	Q	Г	G	v	G	Q	Р	D	Y	G	м	Г	F.	A	D	м	AF	T	E	G	E	E	v	S	Г	ĸ	D	v	
GCI	אסדי	222	CAAC	GT	CGA	SGCO	CGA	GAT	AGC	стт	ידקי	татт	CGC	SCCO	таа	сстт	GAAG	GCG	ATC	ААТТ	rgac	GGT	GGC	AGA	TCI	ידדי	TCG	SCGC	1260
L	Q	P	K	v	E	A	E	I	A	F	v	F	G	R	N	L	E G	D	Q	L	Т	v	A	D	L	F	R	A	1200
CAT	CGA	GTT	CGC	GGT	TCC	GGCI	ААТ	САА	AAI	CGT	GGC	GATO	GCC	GGΛ1	TAGC	саас	TGGG	λτά	TCC	ATA	rcad	CGP	CAC	CAT	TGC	CGP	TAP	ACGC	1350
I	E	F	A	v	Ρ	A	I	к	I	v	G	s	R	I	A	N	WC) I	н	I	Т	D	Т	I	Α	D	N	A	
			~~~		~~~												~		~ ~ ~										1 4 4 0
CTC	GTC		CCTO L	GTA V	v	GCT0	999 0	5	TAD.	.GCC D	.GAF K	ACC P	ישפי ד.		D D	F		R	010	AGGC A	ی اور اور	M	v	.GAI M	GGA F	R	<i>۲</i> الال	G	1440
3	5	0	Б	1	•	D	0	0	•	•	K		5	C	U	•			¥		Ŭ	••	•		5		¥	U	
TGI	GCC	GGT	GTC	TTC	CGG	CGTO	GGG	TTC	GGC	CTG	cci	rggo	STTO	coco	CGCT	саас	GCAP	CGC	тст	GGC	rggo	GAF	GG1	CAT	GGC	TCG	GGG	CGG	1530
v	Ρ	v	s	s	G	v	G	s	A	с	L	G	s	P	L	N	A 1	L	W	L	A	к	v	м	A	R	A	G	
moo																													
100	TCC	GCT	GCG	TGC	TGG	CGA	CAC	TGI	GCI	CTC	cGG	STGO	GC	rggo	Secc	AATG	GTGC	ccg	TGG	CTG	GAGO	GCGP	ACG1	TAT	CGA	ATG1	GCC	GCAT	1620
R	STCC P	GCT L	GCG R	TGC' A	TGG( G	CGA( D	CAC T	TG1 V	GCI L	CTC S	ccco G	STGC A	CGC: L	rggo G	GGCC P	AATG M	GTGC V F	ccg v	tgg A	CTG( G	GAGO G	GCG7 D	V V	ATI F	CGP D	NTG1 V	rgco R	SCAT I	1620
R	P	GCT L	GCG' R TGG'	TGC' A TTC	TGG G GGT	CGAG D AACG	CAC T	TGI V AGI	IGCI L	S S	G G G	STGC A	CGC L	rggo G	GCC P ATG	AATG M ACCC	GTGC V F GCA <i>P</i>	CCG V V	TGG A CAA	CTGO G GGC	GAGO G IGCO	CGF D	V V	F F	CGA D	V V	R	I I CATC	1620 1710
R	P CGG	GCT L ACT	GCG R TGG	TGC' A TTC	TGG G GGT	CGAC D AACC	CAC T CGC	TGI V AGI	IGCI L ICTI F	S S CGC	сссо с с	A A A A G G F	CGC: L L L ATC	IGGO G GAC <i>I</i>	GGCC P AATG	AATG M ACCC	GTGC V P GCAA	CCG V ACT	TGG A CAA	CTGO G GGCI A	GAGO G TGCO A	D D CATC	V V CATC	F F GGI	CGA D D AGC	NTGI V CGGC	R R R R R R	I I CATC	1620 1710
R CGC	P P CGG G	GCT L ACT L	GCG' R TGG' G	TGC ⁴ A TTC S	TGG G GGT V	CGAC D AACC T	CAC T CGC A	TGI V AGI V	IGCI L ICTI F	S S CGC A	сссо с с к	STGC A A <u>GG</u> E	CGC: L L L L ATC	rggg g gac <i>i</i>	GCC P AATG M ( <i>bp</i>	AATG M ACCC T F <i>hG</i> )	GTGC V F GCA <i>I</i> K	CCG V ACT L	TGG A CAA K	CTGC G GGC A	GAGO G IGCO A	D D CATC I	V V CATC I	F F GGI G	CGA D CAGC S	V V CGGC G	R R R R R R R R R R R	I I CATC I	1620 1710
CGC A GGI	P CCGG G	GCT L ACT L	GCG R TGG G CTG	TGC' A TTC S ATG	TGG G GGT V ATC	CGAG D AACG T AAGJ	CAC T CGC A ATC	TGI V AGI V ATC	rgc1 L rcT1 F CAGC	CCTC S CCGC A	G G G K K G G C G G C G	GTGC A A A G G A I C G A I	CGC: L L L L L L L L L L L L L L L L L L L	IGGO G GAC <i>I</i> AGTO	GGCC P AATG M ( <i>bp</i> CGAA	AATG M ACCC T F <i>hG</i> ) ATGG	GTGC V F GCAA K GCGC	CCG V NACT L	TGG A CAA K GGT	CTGC G GGC A CGGJ	GAGO G TGCO A AATO	D D CATC I CGAC	V V CATC I	F F GGI G	CGA D AGC S	ATGI V CGGC G CGAC	R R R R R R R R R R R R R R R R R R R	I I CATC I CCTA	1620 1710 1800
R CGC A GGI G	FTCC P CCGG G T T	GCT L ACT L GAC D	GCG R TGG G CTG L I	TGC A TTC S ATG	TGG G GGT V ATC	CGAC D AACC T AAGI	CAC T CGC A ATC I	TGI V AGI V ATC I	rgci L TCTI F CAGC S	CCTC S CCGC A A CACG T	G G K K G G C C A	AGGA E CGA1 D	CGC: L L ATC * TTC/ S	rggo g gac <i>i</i> gaci v	GGCC P AATG M ( <i>bp</i> CGAA E	AATG M ACCC T R hG) ATGG M G	GTGC V F GCAA K GCGC	CCG V ACT L CAT M	TGG A CAA K GGT V	CTGC G GGC A CGG7 G	G G IGCO A AATO I	CATC I CGAC	CGT V ATC I CGC7 A	F F GGG1 G AGCC A	CGA D AGC S TCC S	ATGI V CGGC G CGAC D	R R R R R R R R R R R R R R R R R R R	I I CATC I CCTA L	1620 1710 1800
R CGC A GGI G	G T C G G T T	GCT L ACT L GAC D	GCG R TGG G CTG L I	TGC A TTC S ATG	TGG G GGT V ATC	CGAC D AACC T AAGZ K	CAC T CGC A ATC I	TGI V AGI V ATC I	IGCI L ICTI F CAGO S	CTC S CGC A A CACC T	G G K K G G C C A	AGGZ E CGAT D	EGC: L L L L ATC *	rggo g gac <i>i</i> gaci v	GGCC P AATG M ( <i>bp</i> CGAA E	AATG M ACCC T R hG) ATGG M G	GTGC V F GCAA K GCGC	CCG V AACT L CAT M	TGG A CAA K GGT V	GGC GGC A CGG G	GAGO G IGCO A AATO I	CATC I CGAC	V ATC I CGC7 A	F F GGI G AGCC A	CGA D AGC S TCC S	ATGT V CGGC G CGAC D	R R R R R R R R R R R R R R R R R R R	I I CATC I CCTA L	1620 1710 1800
R CGC A GGI G GCC	G CGG G TACC T	GCT L ACT L GAC D	GCG R TGG G CTG L I CAG	TGC A TTC S ATG M GCG	TGG G GGT V ATC I	CGAC D AACC T AAGZ K	CAC T CGC A ATC I GTC	TGI V AGI V ATC I GCC	F CAGC S CACC	CTC S CCGC A CACC T	G G G C G C G C C C C C C C C C C C C C	E GAT CGAT	L L AATO * TTC/ S GGGC	rggo g gac <i>i</i> agto v rgc1	GGCC P AATG M ( <i>bp</i> CGAA E	AATG M ACCC T R hG) ATGG M G GGCC	GTGC V F GCAA K GCGC A TGAC	CCG V ACT L CAT M	TGG A CAA K GGT V ACT	CTGC G GGC A CGGA G GCCC	GAGO G IGCO A AATO I GGT/	CATC I CGAC D	CGT V EATC I CGCZ A	F F G G A G C G C G A C G A C G A C G A C C	CCGP D TAGC S CTCC S	CGAC D CGAC	CGCC R CAAC N CGGC G	ECAT I EATC I ECTA L EGTG	1620 1710 1800 1890
R CGC A GGI G GCC A	GTCC P G G T T GCGC R	GCT L ACT L GAC D GCG A	GCG R TGG G CTG L I CAG Q	TGC A TTTC S ATG M GCG A	TGGG G GGT V ATC I I TTG L	CGAG D AAACO T AAAGA K	CAC T CGC A ATC I GTC V	TGI V AGI V ATC I GCC A	CGCT L TCTT F CAGC S CACC T	CCTC S CCGC A CACG T CACC T	G G G K G G C C A C C A C C A	AGGA A E CGAT D CGAC E	L L L L L L L L L L L L L L L L L L L	rggc g gac <i>i</i> Agtc V rgc1 A	GGCC P AATG M ( <i>bp</i> CGAA E TGAA E	AATG M ACCC T F <i>hG</i> ) ATGG M G GGCC G L	GTGC V F GCAA K GCGC A TGAC	CCG V AACT L CAT M CGCG R	TGG A CAA K GGT V ACT L	CTGC G GGCC A G G G CGGJ P	G G IGCC A AAATC I GGT/ V	GCGA D L CATC I CGAC D F	ACGT V I A CGCZ A	F F G G AGCC A CGAC D	CCGP D CAGC S CTCC S CATT I	NTGI V G G CGAC D TGAC D	R R CAAC N CGGC G CATC I	I I CATC I CCTA L CGTG V	1620 1710 1800 1890
R CGC A GGI G GCC A	FTCC P G G T T SCCCC R	EGCT L EACT L EGAC D EGCG A	GCG' R TGG' G CTG/ L I CAG	TGC A TTC S ATG M GCG A	TGG G GGT V ATC I I TTG L	CGAC D AACC T AAG K SGCC G S	CAC T CGC A ATC I GTC V	TGI V AGI V ATC I GCCC A	CGCI L TCTI F CAGC S CACC T	CCTC S CCGC A T CACC T CACC T	G G G G G G G G G G G G G G G G G G G	A A A CGA1 D CGAC E	2002: L L ATC * * S S G G T C C	IGGO G GGAC <i>I</i> AGTC V IGC1 A	GGCC P NATG M ( <i>bp</i> CGAA E TGAA E	AATG M ACCC T F hG) ATGG M G GGCC G L	GTGC V F GCAA K GCGC A TGAC	CCCG V V AACT L CCAT M CCCAT R	TGG A CAA K GGT V ACT L	CTGG G GGC A CGGG G G CGC P	GAGO G IGCO A AAATO I GGT/ V	CATC D CATC I CGAC D ATTC F	V ZATC I CGC7 A CGCC7 A	F F GGGI G AGCC A CGAC D	CCGA D CAGC S CTCC S CATT I	ATGI V CGGC G CGAC D IGAC D	R R R R R R R R R R R R R R R R R R R	I I CATC I CCTA L CCTA V	1620 1710 1800 1890
R CGC A GGT G GCC A TTC	FTCC P CCGG G CACC T F SCGCC R R CGAC	GCT L GACT L GAC D CGCG A	GCG R TGG G CTG L L I CAG Q Z XACC T	TGC A TTTC S ATG M GCG A A GCG A	TGG( G GGT) V ATC, I I I CGCG A	CGAC D AAACC T AAAG K S GGCC G GGC	CAC T CGC A ATC I GTC V GCG A	TGI V AGI V ATC I GCC A SCAC H	CGCI L TCTI F CAGC S CACC T CGT/ V	CCTC S CCGC A CACG T CACC T CACC T	CCGC G CC <u>A</u> K GCCC A CCAC H SAA'	AGGA E CGA1 D CGAC E TGA ⁷ D	L L L AATC * TTC/ S G G TTCC A	IGGC G G G A G A G C C T G C T C C T C C T	GGCC P MATG ( <i>bp</i> CGAA E CGAA E CGAA E	AATG M ACCC T R hG ATGG M G GGCC G I GGCC CGCC R F	GTGC V F GCAA K GCGC A TGAC T TGAC T T CATC/	CCCG AACT L CCAT M CCCAT R R CCCAA K	TGG A CAA K GGT V ACT L ACC P	CTGG G GGCC A CGGG P CGGC A	GAGC G IGCC A AATC I SGT/ V GATC	GCGA D CATC I CGAC D ATTC F GCGG R	ACGT V CATC I CGC7 A CGCC A CGCC V	F F GGGI G AGCC A CGAC D SATC I	CCGA D TAGC S TTCC S CATT I CCGAG	ATGI V CGGC G CGAC D TGAC D L	CGCC R CAAC N CGGC G CATC I I TACC	CATC I CATC I CCTA L CGTG V GCCC P	1620 1710 1800 1890 1980
GGT GGT GGT GGT F	FTCC P G CCGG G T T SCCGC R R CCGAC D	GCT L FACT L GAC D CGCG A CGCC A	GCG' R TGG' G CTG/ L I CAGG Q 2 :ACC T	TGC A TTCC S ATG M GCG A A GCG A A GT S	TGG G GGT V ATC I I TTG L G GCG A	CGAC D T AAAG K G G G G G G G G G G G G G	CAC T CGC A ATC I GTC V GCC A	TGI V AGI V ATC I GCC A GCCAC H	CGCT L F CAGC S CACC T CGT/ V	CCTC S CCGC A CACG T CACC T CACC T K	CCGC G CC <u>A</u> K CCAC H SAA' N	A A A CGA1 D CGAC E TGA7 D	L L ATC S G G TGC A	IGGC G G AGTC V IGC I A CCTC L	GGCC P MATG ( <i>bp</i> CGAA E TGAA E SGCTG L	AATG M ACCC T F hG) ATGG M G GGCC G I GGCC R F	GTGC V F GCAA K GCGC A TGAC T TGAC T CATC I H	CCCG V AACT L CCAT M CCCAT R R ACAA K	TGG A CAA K GGT V ACT L ACC P	CTGG G GGC A CGGG G G CGC A	G G G G G A A A A A C G G A T G A T G A T G A T G G T G G G G	GCGA D L CATC I CGAC D ATTC F GCGG R	ACGI V I CGCZ A CGCCZ A CGCCC A	F G G AGCCC A CGAC D SATC I	CCGA D CAGC S CTCC S CAT1 I CCGAC D	ATGI V GGGC G CGAC D IGAC D CCTT L	CGCC R CAAC N CGGC G CATC I TACC T	I I CCTA L CCTA L CCTA V SGCCC P	1620 1710 1800 1890 1980
R CGC A GGT G GCC A TTC F GCC	FTCC P CCGG G CACC T SCGC R CGAC D SGCC	EGCT L EACT L EGAC D EGCG A EGCG A ECGCC A ECGCC	GCG' R TGG' G CTGJ L I CAGG Q Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z	TGC A TTCC S ATG M GCG A AGT S CCC	TGG G GGT V ATC. I I TTG C GCG A TAC	CGAC D AAACC T AAAGJ K SGCC G G G G G G G C TGTJ	CAC T CGC A ATC GTC V GCG A ATC	TGI V AGI V ATC I GCCC A SCAC H	CGC1 L CCT1 F CAGC S CACC T CGT7 V SGTC	CCCC S CCCC A CACC T CACC T CACC T K SGGTC	CCGC G CC <u>A</u> K CCAC H SGAA' N CAAC	AGGA AGGA E CGA1 D CGAC E TGA D	L L ATC TTC/ S G G TTCC A	rggc g gac/ agtc V lgc1 a cctc L	GGCC P MATG M ( <i>bp</i> CGAA E I GGAA E CCAC	AATG M ACCCC T F hG ATGG G G G G G G G G G C G C TGC	GTGC V F GCAA K GCGC A TGAC T GATCA E ATCA E ATCA	CCCG V AACT L CCAT M CCCAT R R ACAA K CCCT	TGG A CAA K GGT V ACT L ACC P GAA	CTGC G G G CGGC A CGGC A CGGC	GAGC G IGCC A AATC I SGT/ V GATC M GAAC	CATC I CGAC D ATTC F GCGG R R CATC	ACG1 V CATC I CGC7 A CGCC A CGCC A CGCC V	F F G G AGCC A CGAC D SATC I CACC	CCGA D CAGC S CTCC S CATT I CCGAC D	ATGI V CGGC G CGAC D IGAC D CCTT L	CGCC R CAAC N CGGC G CATC I TACC T	CAT I CATC I CCTA L CCTA L CGTG V SCCCC P CCAG	1620 1710 1800 1890 1980 2070
R CGC A GGT G GCC A TTC F GCC A	FTCC P G CCGGG T GCGGC R CCGAC D SGCCC A	GCT L GACT L GGAC D CGCG A CGCC A CGCC A L CGCC I	GCG' R TGG' G CTG/ L I CAGG Q / CAGG Q / CAGG R T S ACC T S GGCC G G	TGC A S ATG M GCG A AGT S CCC	TGG G GGT V ATC I I TTG C GCG A TAC Y	CGAC D T AAAG K GGCC G G C TGT C	CAC T CGC A ATC I GTC V GCG A ATC I	TGI V AGI V ATC I GCC A CCC P	CGCT L F CAGC S CACC T V CGT/ V SGTC V	CCCCC S CCCCC A CACCC T CACCC T AAAAC K SGGTC V	CCAC G CCAC A CCAC H SAA' N CAAC	AGGA E CGA1 D CGAC E TGA' D CGGC G	L L AATC S G TGC A CGA/ E	IGGC G G AGTC V IGCT A CCTC L AGAC D	GGCC P MATG M ( <i>bp</i> CGAA E TGAA E SCTG L CCAC	AATG M ACCC T F hG) ATGG GGCC G I GGCC G I C CTGC L P	GTGC V E GCAA K GCGC A TGAC T TGAC T CATCA H CCCGC	CCCG V AACT L CCAT M CCCAT R ACAA K K CGCT L	TGG A CAA K GGT V ACT L ACC P GAA N	CTGC G GGC? A CGGC G G CGCC A CGTC V	GAAC G G C C C C C C C C C C C C C C C C	CATC I CGAC D ATTC F GCGG R CATC M	ACGT V I CGCC A CGCC A CGCC A CGCC V V CGCC V	F F G G G A G C G A C C A C C A C C A C C A C C A C C A C C C G I C C G I C C G I C C G I C C G I C C G I C C C C	CCGA D CAGC S CTCC S CATT I CCGAC D CTGC C	ATGI V GGGC G CGAC D TGAC D CCT? L G	R R R R R R R R R R R R R R R R R R R	CATC I CATC I CCTA L CGTG V GGCCC P CCAG Q	1620 1710 1800 1890 1980 2070
GGT GGT GGT GGT A TTC F GCCC A	FTCC P G CACC T T GCGC R CGAC D GGCC A	GCT L GACT L GGCC D GCCC A CGCCC A SATC I	GCG' R TGG' G CTGJ L I CAGG Q Z X ACC T : ACC T : GGCC G I	TGC A S ATG M GCG A A GCG A A GCC P	TGG G GGT V ATC I I TTG L G G CG CG A TAC Y	CGAC D T AAAG/ K C C C C	CAC T CGC A ATC I GTC V GCG A ATC I	TGI V AGI V ATC I GCCC A SCAC H	CGCTI F CAGC S CACC T CGT/ V SGTC V	CCTC S CCGC A CACG T CACC T CACC T AAAAC K SGGTC V	GGCC G K GGCCC A CCAC H SAAC N	AGGZZ E CGAN D CGAC E TGA' D CGGC G	L L A TTC/ S G G TTCC A CGA/ E	IGGC G GAC7 AGTC V IGC1 A CCTC L AGAC D	GGCC P MATG ( <i>bp</i> CGAA E IGAA E GCTG L CCAC H	AATG M ACCC T F hG) ATGG GGCC G I GGCC G I SCGCC R F CTGC L P	GTGC V E GCAA K GCGC A TGAC T TGAC T C A C C C G C C G C A	CCCG V V AACT L CCAT M CCCAT M CCCAT R ACAA K CCCT L	TGG A CAA K GGT V ACT L ACC P GAA N	CTGC G GGC? A CGGG P CCGC A CGTC V	GAGC G IGCC A AAATC I GATC M GAAC N	CATC I CGAC D ATTC F GCGC R CATC M	ACGI V I CGC7 A CGC7 A CGC7 V V SGT0 V	F F G G A G C G A C G A C G A C G A C G A C G A C G A C G T	CCGA D S CTCC S CTCC S CAT1 I CCGAC D CTGC C	V CGGC G CGAC D CGAC D CCT C CGG G	R R R R R R R R R R R R R R R R R R R	I CATC I CCTA L CCTA L CCTA V GCCCG P CCAG Q	1620 1710 1800 1890 1980 2070
R CGC A GGT G G CC A TTC F GCC A GCC	FTCC P G CACC T SCGC R CGAC A SGCC A	GCT L GACT L GACT D GCCG A CGCC A CGCC A CGCC I CATT	GCG' R TGG' G CTGJ L I CAGG Q A CCG G CCC.	TGC A TTCC S ATG M GCCG A A GCCC P ATG	TGG G G TTG I C G C G C G C G TAC Y G TG	CGAC D AAACC T AAAG/ K S GGCC G C TGT/ C C	CAC T CGC A ATC I GTC V GCC A ATC I GCC	TGI V AGI V ATC I GCC A CCC H CCC P	CGC1 L F CAGC S CAGC T CGT7 V SGTC V	CCTC S CCGC A CACC T CACC T CACC T CACC K SGTC V	G G G G G G G G G G G G G G G G G G G	STGC A A CGAI D CGAC E TGA' D CGGC G G	L L L TTC/ S G G TCC/ A CGA/ E	G G G AGTC V IGCT A CCTC L A AGAC D GGTC	GGCC P MATG ( <i>bp</i> CGAA E IGAA E GCTG L CCAC H	AATG M ACCC T F hG) ATGG M G GGCC G I CCGCC R F CCTGC L P TACC	GTGC V F GCAA K GCGC A T GCGC A T GCCGC A A GCCGA	CCCG V V AACT L CCAT M CCCAT M CCCAT R ACAAA K CCCT L	TGG A CAA K GGT V ACT L ACC P GAA N CGT	CTGC G G CGG2 A CGG2 A CGG2 V CGCC	GADC G G G A AATC I G G G G A G G A C A G C A G C	CATC I CCGAC D ATTC F CATC M CATC	ACGT V EATC I CGC7 A CGC7 A CGC7 A CGC7 V SGT7 V CTC0	CATI F GGGI G CGAC D GATC T CACC	CCGA D S CTCC S CTCC S CTCC D CCGAC C C C C C C C C C C C C C C C C C	ATGI V CGGC G CGAC D CGAC D CGGT G G G G	CGCC R CAAC N CGGC G T T CGGC G CGCC	GCAT I CATC I CCTA L CCTA L CCTA V GCCCC P CCCAG Q CCCCGC	1620 1710 1800 1890 1980 2070 2160
R CGC A GGT G GCC A TTC F GCC A GCC A	TCC P G CCGG G T GCGC R CGAC D GGCC A CACC T	GCT L GACT L GACT D GGCC A CGCC A SATC I SATT I	GCG' R TGG' G CTGJ L I CAG( Q P SACC T SACC T SACC T CCC. P	TGC A TTCC S ATG M GCG A A GCG A CCC P ATG M	TGG G GGT. V ATC. I TTG C GCG TAC Y GTG V	CGAC D AACC T AAGJ K S GGCC G C C S C C S C C S C C S C C S C C S C S C S C S C S C S C S C S C S C S C S C S C S C S C S C S C S C S S C S C S S C S C S S C S C S S C S C S S C S S C S S C S S S C S S S C S S S S S S S S S S S S S S S S S S S S	CAC T CGC A T CGC A C C C C C C C C C C C C C C C C C	TGI V AGI V ATC I GCCC A CCCC P CCCC V	CGCT L F CAGC S CACC T V CGGTC V S GGTC V S	CCTC S CCGC A CACC T T CACC T T AAAAC K SGGTC V SGCGC R	G G G G G G G G G G G G G G G G G G G	AGGA E CGA1 D CGAC E TGA G G G G G G G G G G G G G G G G G G	L L L L L L L L L L L L L L L L L L L	G G G AGTC V IGCT A CCTC L A GGTC V	GGCC P MATG M ( <i>bp</i> CGAA E CCAA H CCAC H GCCAC Q	AATG M ACCC T F hG) ATGG M G GGCC G I C GGCC G I C C TACC Y C	GTGC V F GCAA K GCGC A TGAC T TGAC T CATCA A CCCGC A A GCCGA	CCCG V V AACT L CCAT M CCCAT R K CCCAT R K CCCAT L L AAAAT I	TGG A CAA K GGT V ACT L ACC P GAA N CGT V	CTGC G GGC? A CGG2 G G CGCC A CGTC V CGCC A	GAT( N CAAC CAG CAG CAG	CATC I CGAC D ATTC F GCGC R CATC M CATC I	ACGI V I CATC I CGC7 A CGC7 A CGC7 V V SGT7 V SGT7 V SGT7 S	CATT F GGGI G AGCC A CGAC D SATC T CACC S	CCGA D CAGC S CTCC S CATT I CCGAC C C C C C C C C C C C C C C C C C	ATGI V CGGC G CGAC D TGAC D CGGT G G G G G SAGC S	R R R R R R R R R R R R R R R R R R R	GCAT I CATC I CCTA L CCTA L CGTG V GGCCC P CCAG Q CCGGC G	1620 1710 1800 1890 1980 2070 2160
R CGC A GGT G GCC A TTC F GCCC A GCCC A	TCC P G CACC T GCGAC R CGAC A CACC T GGGC	GCT L GACT L GACT D GGCG A CGCC A SATC I SATC	GCG ¹ R TGG ² G CTG <i>i</i> L I CAGG Q <i>i</i> CAGG C G C CCC R I CCCC	TGC A TTCC S ATG M GCC A A GCC M GCC	TGGG G GGT, V ATC. I TTGG CG GCG TAC Y GTGG V AAC	CGAC D AAACC T AAAGJ K SGCC G G C C S G C S C S C S C S C S C	CAC T CGC A ATC I GTC V GCC A ATC I GCC A GAC	TGI V AGI V ATC I GCC A C C C C C C C C C C C C C C C C C	CGC1 L F CAGC S CACC T CGT7 V SGTC V SGTC S	CCCC S CCCC A CACCC T CACCC T AAAAC K GGTC V SCCCC R	CCAC G CCAC A CCAC H SAAC N CCAC V CCAC	A A A A A A A A A A A A A A A A A A A	L L L L L L L L L L L L L L L L L L L	G G G AGTC V IGC1 A CCTC D SGTC V	GGCC P M (bp CGAA E TGAA E TGAA E CCAC H GCAG Q	AATG M ACCC T F hG) ATGG M G GGCC G L F CTGC C C C C C C C C C C C C C C C C C	GTGC V F GCAA GCGC K GCGCC T TGAC T TGAC T C A C CCGCC A S C CCGCC T C C C C C C C C C C C C C C C	CCCG V V AACT L CCAT M CCCAT K CCCAT K CCCAT K CCCAT N ACAAA K L AAAAT I	TGG A CAA K GGT V ACT L ACC P GAA N CGT V GGT	CTGC G G CGGC P CGGC A CGGC V CGGC A GGGG	GATC GATC GATC GATC M GAAC N CAGC S CGGG	CATC I CGAC D ATTC F GCGG R CATC I I CCATC	ACGT V I CGC7 A CGC7 A CGC7 V V CCC7 S	CATI F GGGI G CGAC D GATC T CACC T CACC S	CGA D CAGC S CTCC S CAT1 I CGAC C C CAAC K	ATGI V CGGC G CGAC D CGAC D CCTT G CGGT G G G G G S AAC	R R R R R R R R R R R R R R R R R R R	GCAT I CATC I CCTA L CGTG V GGCCC P CCAG Q CCAG G G CCATC	1620 1710 1800 1890 1980 2070 2160 2250
R CGCC A GGT G GCCC A TTC F GCCC A GCCC A CCCC P	TCC P G CCGG G T CACC T CGAC D GGCC A CACC T GGGC G	GCT L GACT L GGCG D GGCG A CGCCG A CGCCG A CGCCG A CATC I CATT I CACT	GCG ^T R TGG ^T G CTG ^T CAGG Q CAGC T CAGC T CCCC R CCCC R	TGC A S ATG M GCG A A GCC P ATG M GCC A	TGG( G GGT, V ATC, I TTG( L GCG A TAC Y GTG V AAC	CGAC D AAACC T AAAGJ K S GGCC G G C S GGCC G C S C S C S C S C	CAC T CGC A ATC I GTC V GCC A ATC I GCC A GAC D	TGI V AGI V ATC I GCC H CCC H CCC C H CCC C C C C C C C C	CGC1 L F CAGC S CACC T CGT/ V SGTC S STTC F	CACC A CACC T CACC T AAAAC K SGTC V SGCGC R CACC	GGCC G G G G G G G G C A C C A C C A C C A C C A C C A C C A C C A C C A C C C C C C C C C C C C C C C C C C C C	E CGA1 D CGAC E TGA G G G G G G G G G G G G G G G G G G	L L L L L L L L L L L L L L L L L L L	G G G AGTC V IGCT A AGAC D SGTC V S CTCC S	GGCC P ( <i>bp</i> CGAA E IGAA E CCAC L CCAC H SCAG Q CAAG K	AATG M ACCC T F hG) ATGG M G GGCC G I CTGC C C C C C C C C C C C C C C C C C	GTGC V E GCAA K GCGC T TGAC T TGAC T CATC A A GCCG A A GCCG A E	CCCG V V L CCAT M CCCAT M CCCAT K CCCAT K CCCAT I V	TGG A CAA K GGT V ACT L ACT P GAA N CGT V GGT V	CTGC G G CGGC P CGCC A CGCC A CGCC A CGCC A CGCC A GGGG	GAAG G G A AAATC I GAAT M GAAA N CAG G G G G G G G G G	CATC I CGAC D ATTC F GCGC R CATC I CATC I CCATC I	ACGT V I CGC7 A CGC7 A CGC7 V SGT0 V SGT0 S SCC0 A	CATI F G G G G G G G G G G G G G G G G G G	CGA D SAGC S CTCC S CATT I CGAC C C CAAC K GGGC G	ATGI V CGGC G CGAC D CGGC G G G G G G G G G G G CAAC K	R R R R R R R R R R R R R R R R R R R	GCAT I CATC I CCTA L CGTG V GGCCC P CCAG Q CCAG G CATC I	1620 1710 1800 1890 1980 2070 2160 2250
R CGCC A GGT G GCCC A TTC F GCCC A GCCC A CCCC P	G G G CACC T G CGAC R CGAC D G G G G G G G	GCT L SACT L GAC D CGCC D CGCC A CGCC A CGCC A CGCC I I SACT I CATT I CACT	GCG ^T R G CTG ^J CAGG Q A CCCC R CCCC R	TGC A S ATG M GCG A A GCC P ATG M GCC A	TGG G GGT. V ATC. I TTG C GCG A TAC Y GTG V AAC N	CGAQ D T T AAAG2 K C C C C C C C C C C C C C C C C C C	CAC T CGC A ATC I GTC V GCC A ATC I GCC A GAC D	TGI V AGI V ATC I GCC A C GCC C P C GCC V C GAC E	CGC1 L F CAGC S CAGC T CAGC T CAGC T CAGC T CAGC S S T C C T C S S T C C T C T C S S C A C C T T S S S S S S S S S S S S S S S S	CACC A CACC T CACC T AAAAC K GGTC V GCGCC T	G G G G G G G G G G G G G G G G G G G	AGGA E CGA1 D CGA0 E TGA7 D CGG0 G G G G G G G G G G G G G G G G G	L L L TTC/ S G TGC A CGA/ E CAAC T	G G G AGTC V IGCT A CCTC D SGTC V S	GGCC P MATG M ( <i>bp</i> CGAA E IGAA E GCTG L CCAC H GCAG Q CAAG Q K	AATG M ACCCC T F hG) ATGG GGCC G I SCGCC C F CTGC L P TACCC Y C GCCP A 1	GTGC V E GCAA K GCGC A TTGAC T CATCA A CATCA A CATCA E TTCGA	CCCG V V AACT L CCAT M CCCAT M CCCAT K CCCAT K CCCAT N ACAAA K L AAAAT I V	TGG A CAA K GGT V ACT L ACC P GAA N CGT V GGT V	CTGG G G CGG2 A CGG2 A CGG2 A CGG2 A G G G	GAGG G IGCC A AAATC I GAATC M GAAC N CAGG G G	CGAC I CGAC D ATTC F CGAC R CATC I CATC I CGCC A	ACG1 V I CGC7 A CGC7 A CGC7 A CGC7 V CCC7 C CCC7 A	CATI F G G G G G G G G G G G G G G G G G G	CCGA D S CTCC S CATT I CCGAC C C CAAC K GGGGC G	ATGI V CGGC G CGAC D CGAC D CCCT G G G G G G G G G CAAC K	R R R R R R R R R R R R R R R R R R R	GCAT I CATC I CCTA L CGTG V GCCCAG Q CCAG G CATC I	1620 1710 1800 1890 1980 2070 2160 2250
R CGCC A GGT G GCCC A GCCC A GCCC A GCCC A CCCC P ATC	TCCG P G CACC T SCCGC R CCGAC R CCGAC T CCGAC T G G G CCGTC	CGCT L CGACT L CGACT D CGCCG A CGCCG A CGCCG A CGCCG T L CATT I CATG	GCG ^C R G CTGJ L I CAGG Q CAGG G CCCC R CCCC R AACC	TGC A S ATG M GCG A A GCC P ATG M GCCC A CCG	TGG G GGT. V ATC. I TTG C GCG A TAC Y GTG V AAC N GCC	CGAG D T AAAG2 K SGCC G C C SGCC A C SGCC A A T C SGCC A A T C SGCC A C S SGCC C S S S S S S S S S S S S S S	CAC T CGC A ATC I GTC V GCC A A TC I GCC A CCG	TGI V AGI V ATC I GCC A C C C C C C C C C C C C C C C C C	CGC1 L F CAGC S CAGC T CGT7 V SGTC S STTC F	CCCCC S CCCCC A CACCC T CACCC T AAAAC K SGGTC V SGCGCC R T SATC	CGGC G K G G G C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C C C C C C C C C C C C C C C C C C C	E CGA1 D CGA0 E TGA' D CGGC G G G G G G G G G G G G G G G G	L L L L L L L L L L L L L L L L L L L	G G G G G G G G G G G G G G G G G G G	GGCCC P MATG GGAA E GGCAG L CCAC H GCCAC H GCCAC K GGTC	AATG M ACCC T F hG ATGG M G GGCC C G C C G C C C C C C C C C C C	GTGC V F GCAA K GCGC A TGAC T TGAC T C A C C C C C C C C C C C C C C C C	CCCG V V AACT L CCAT M CCCAT M CCCAT K CCCAT K CCCAT K L AAAAAT L V V TGAG	TGG A CAA K GGT V ACT L ACC P GAA N CGT V CGA	CTGG G G CGG2 A CGG3 P CGGC A CGG7 A CGG7 A GGG6 G	GAAC G A AAATC I GAATC M GAAC N CAGC G CCGCC	CCATC I CCATC D ATTC F CCATC I CCATC I CCATC I CCATC	V I CGC7 A CGC7 A CGC7 A CGC7 V V CTC7 S CCC7 C A CCC7 C CCC7 A	CATI F G AGCC A CGAC D CACC T CACC T CACC S CAAC K CCCC	CCGA D CAGC S CTCC S CTCC S CTCC C S CCAAC C C C C C C C C C C C C C C	ATGI V CGGC G CGAC D TGAC D CCCT G G G G G G G G CAAC K G G CAAC	R R R R R R R R R R R R R R R R R R R	GCAT I CATC I CCTA L CGTG V GCCCAG Q CCAG G CCAG G CATC I SGAC	1620 1710 1800 1890 1980 2070 2160 2250 2340
CGCC A GGT G G G G G G G G G C C A C C C C A C C C C	TCCGG G CCGGG T CCGAC T CCGAC T CCGAC G CCGTC V	GCT L ACT L GGAC D GGCG A CGCC A CGCC A CGCC A CGCC I CATT I CATT I CCTG L	GCG ^C R G CTG L I CAG Q CAG G CCC R CCC R R AACC N	TGC A TTTC S ATG M GCCG A A GCCC A GCCC A CCCG P	TGGG G GGT. V ATC. I TTG GCG A GTG V AAC N GCC A	CGAQ D T AAAGJ K SGCC G G C S G C S C S C S C S C S C S	CAC T CGC A ATC I GTC V GCC A A TC GCC A GAC D CCG P	TGI V AGI GCCCA GCCCA H GCCCCA P GCCCCA P GCCCCC P	CGCI L F CAGC S CAGC S CACC T V SGTC S STTC F SGCTC L	CCTC S CCGC A CACC T CACC T AAAAC K GGTC V SGCGC R CACC T CACC I	CGCC G K G G G C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C C C C C C C C C C C C C C C C C C C	E CGA1 D CGA2 E TGA' D CGGC G G G G G G G G G G G G G G G G	L L L L L L L L L L L L L L L L L L L	G G G G G G G G G G G C C C C C C C C C	GGCCC P MATG GGAA E GGAG CGAA E CCAC H GCAG Q CAAG K CAAG K C V	AATG M ACCC T F hG ATGG M G GGCC C G C C C C C C C C C C C C C C	GTGC V F GCAA K GCGC A TGAC T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA T C CATCA C CATCA C CATCA C CATCA CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA C CATCA CATCA CATCA C CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CATCA CA	CCCG V V AACT L CCAT M CCCAT M CCCAT N ACAAA K CCCAT L AACAAA K L AACAAA K CCCCA R V CCCAT C ACAAA K CCCAT L L CCAT K CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT N CCCAT C CCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C CCCAT C C CCCAT C C CCCAT C C CCCAT C C CCCAT C C CCCAT C C C C	TGG A CAA K GGT L ACC P GAA N CGT V CGA D	CTGG G G CGGC A CGGC A CGGC A CGGC A G G G TTTC F	GAAC G AAATC I GGATC M GAAC N CAGG G CGGC A	CATC D CATC D CGAC F CATC M CATC I CGCC A CGAC D	V CATC I CGCA A CGCC A CGCC A CGCC A CGCC A CACC T	CATI F G AGCCA A CCACC T CACC T CACC S CAAC K CCCCCA	CGA D CAGC S CTCC S CATT C C C C C C C C C C C C C C C C C	V G G C G C G C G C G C G C C G C C G C C G C C C C C C C C C C C C C C C C C C C C	R R R R R R R R R R R R R R R R R R R	GCAT I CATC I CCTA L CCTA L CCTA L CCTA CCTA CCTA CCT	1620 1710 1800 1890 1980 2070 2160 2250 2340
R CGCC A GGT G G G G G C C A G C C C C A C C C C C A C C C C	TCCGG G CCCGG G CACCC T SCCGC D CCCAC T CCCAC T SCCGCC G CCCCC V	GCT L GACT L GGAC D CGCG A CGCG A CGCC A CGCC A CGCC A CGCC A CGCC A CGCC C A CGCC C A CGCC C C C	GCG ^C R G CTG ^C CAG Q CAG Q CAG G CCC R R CCCC R R	TGC A TTTC S ATG M GCCG A A GCCC A GCCC A CCCG P	TGGG G GGT I I TTGG L GGCG A GTG V AAC N GCCC A	CGAQ D T AAAGJ K C G G G G G G G G G G G G G G G G G G	CAC T CGC A ATC GTC V GCC A ATC I GCC A CCG P	TGI V AGI V ATC I GCCCA H CCCCA P C C C C C C C C C C C C C C C C	CGCI L F CAGC S CACC T CGT/ V CTCC S GGTC L	CCTC S CCGC A CACC T CACC T AAAAC K GGTC V SCGCC T CACC T SATC	CGGC G K G G CAC H SAAC N CAAC V CGAZ E CATC M	AGGZ A E CGA1 D CGAC E TGA7 D CGGC G G G G G G G G G G G G G G G G	L L L TTC/ S G G TTC/ A CGAA K CAAC K T CAAC T	IGGC G G G G G G G C C C C C C C C C C C	GGCCC P AATG M (bp CGAA E TGAA E CCAC H GCAC Q CAAG C CAAG K SGTC V	AATG M ACCCC T F hG) ATGG G G G G G G G G G C T G C C T G C C C C	GTGC V E GCAA K GCGC A TGAC T CATC/ A CATC/ A CCGC C E CCCC L	CCCG V AACT L CCAT M CCCAT M CCCAT M CCCAT M CCCAT N CCCAT N CCCAT V CCCAT V V CCCAT V V CCCAT V V V V V V V V V V V V V V V V V V V	TGG A CAA K GGT V ACT L ACC P GAA N CGT V CGA D	CTGG G G G CGGC A CGGC A CGGC A CGGC A CGGC A CGGC A CGGC A CGGC A CGGC A CGGC A CGGC A CGGC A CGGC A CGGC C A CGGC C C C	GAAC G A AAATC I GAAC M GAAC S CAGG G CGCC A	CATC I CGAC D ATTC F GCGGC R CATC I CCATC D CCATC D	V CATC I CGCZ A CGCC A CGCC A CGCC C CCCC C CCCC C CCCCC	FATI F G AGCC A CGAC T CACC T CACC S CAAC K CGCC A	CGA D S TCC S CATT I CGAC C C C C C C C C C C C C C C C C C	V GGGC G CGAC D CGAC D CCCT G G G G G G G G G G G G G G G G G	R R R R R R R R R R R R R R R R R R R	CATC I CATC I CCTA L CCTA L CCTA CCTA CCTA CCCAG Q CCCAG C CCAG C CCATC I CCCAG C CCATC I CCTA	1620 1710 1800 1890 1980 2070 2160 2250 2340
R CGCC A GGT G G G CCC A C G CCC A C CCC A C C C C C	TCCGP G CCGGG T CACCC T CCGAC D CCGAC T CCGAC T CCGAC C T CCGAC C T CCGAC C T CCGCC C C CCGCC C C CCGCC C C C C	CGCT L CGACT L CGAC D CGCC A CGCC A CGCC A CACT L CCTG L CCTG C CCTG	GCG ^C R G CTG ² CCTG ² CCGG ² CCCC R CCCC R CCCC R CCCC R	TGC A TTTC S ATG M GCCG A A GCCC P ATG M GCCCG P CCCG P	TGGG G GGT. V ATC. I I TTGG L GCG A GTG V A A C C C A GCCC A	CGAQ D AAACQ T AAAGA K SGCCG G G G G G G G G C C C C C C C C C	CAC T CGC A ATC I GTC V GCC A GCC A GCC G CCG P GCA A	TGI V AGI GCC A GCC A C C C C C C C C C C C C C C	CGCT L F CAGC S CACC T V CTCC S GGTC V S GGTC L GCTC C C C C C C C C C C C C C C C C C	CCTC S CCGC A CACC T CACC T AAAAC K GGTC V GCGCC T GGCCT A	CGAC R CAAC N CAAC N CCAC CAAC N CCAAC N CCAAC CAAC CAAC CAAC CAAC CAAC CAAC CAAC CAAC CAAC CAAC CAAC CAAC CAAC N CCAAC N CCAAC N CCAAC N CCAAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N N CCAC N CCAC N CCAC N N CCAC N CCAC N CCAC N CCAC N N CCAC N CCAC N CCAC N CCAC N CCAC N CCAC N N CCAC N N CCAC N N CCAC N N CCAC N N CCAC N N CCAC N N CCAC N N N CCAC N N N N	AGGA E E CGA1 D CGAC E TGA D CGGC G G G G G G G G G G G G G G G CGGC A A A CCGAC T T R T G CGGC A C G C G C C G C C C C C C C C	L L L L L L L L L L L L L L L L L L L	IGGCI G G G G G G G C C C C C C C C C C	GGCCC P AATG M (bp CGAA E TGAA E GGCTG L CCACC H GCCACC K CAAG K CAAG CAAG C CAAG C CAAG C CAAG C C CAAG C C C C	AATG M ACCCC T F hG) ATGG G G G G G G G G G G C T G C C T G C C T A I I T A C T C C C C C C C C C C C C C C C C	GTGC V E GCAA K GCGCC A TGAC A TGAC A CATCA A CCCCC C L TTGAA	CCCG V V AACT L CCAT M CCCAT M CCCAT M CCCAT N CCCAT N CCCAT V CCCAT V CCCAT V CCCAT V CCCAT V V CCAT V V V V V V V V V V V V V V V V V V V	TGG A CAA K GGT L ACC P GAA N CGT V CGA D CGA D CGA C G	CTGG G G G CGG C G CGC A CGC A CGC A CGC A CGC A CGC A CGC C A CGC C A CGC C A CGC C A CGC C A C C C C	GAAC G G A AATC I GATC M GAAC N CAGC G CGCC A GCAC	CATC I CGAC D ATTC F CATC R CATC A CATC A CATC A CATC A CATC A CATC C A CATC C A CATC C A C CATC C F	V I CGCZ A CGCC A CCGCC A S CCGCC A CCGCC A CCGCC A CCGCC A CCGCC A	FATT F G AGCCC A CGAC D SATC T CAGCC S CAAC K CCGCC R	CGA D S CTCC S CATT I CGAC C C CAAC K GGGCC G CAAC K	V GGGC G CGAC D CGAC D CCT? C G G G G G G G G G G G G G CAAC K I CCAAC F	R R R R R R R R R R R R R R R R R R R	GCAT I CATC I CCTA L CGTG V GGCCC P CCAG Q CCAG G CCAC I CCATC I CCAC C D	1620 1710 1800 1890 1980 2070 2160 2250 2340 2430

CGGCCGATCCGCATCCCTGGCGTAGGCGATCGGCTCACCGGGCTGAAGACCTCGGTCTTCCTGGAGGTCGAGGGTGCCGCCCACTACCTG 2520 R P I R I P G V G D R L T G L K T S V F L E V E G A A H Y L P A Y A G N L D I M T S A A L R T A E K M A Q R L L A T L T GCCTGAGGATGCCCCCATGACGACAAAGAAAATCTACATTTCTGATGTGACGCTGCGCGACGGCAGCCACGCCATCCGCCACCAGTACAG 2700 A * M T T K K I Y I S D V T L R D G S H A I R H Q Y S (bphF) CGTGGAACAGGTCAGGACCATCGCGAAAGAGTTGGACGAAGCCAAGGTCGATTCCATTGAGGTCGCCCATGGCAGCGCCTGCAGGGCTC 2790 V E Q V R T I A K E L D E A K V D S I E V A H G S G L Q G S CAGCTTCAACTACGGCTTCGGCGCCCACACCGACTTGGAGTGGATCGAGGCGGTGGCCAGCGTGGTCAAGCACGCGAAGATCGCCACCCT 2880 S F N Y G F G A H T D L E W I E A V A S V V K H A K I A T L L L P G I G T I H D L K A A Y D A G A R I V R V A T H C T E GGCGGACATCTCCAAACAGCACATCGAATACGCCCGCCCACCTGGGCATGGAGGCGGTCGGCTTTCTGATGATGAGCCACATGAGCACGCC 3060 A D I S K Q H I E Y A R H L G M E A V G F L M M S H M S T P GCAAGGCCTGGCCCAGCAGGCCAAGCTGATGGAAAGCTACGGAGCCACCTGCTGCTACGTGGTAGATTCCGGCGGCGCCCTTGAGCATGGA 3150 Q G L A Q Q A K L M E S Y G A T C C Y V V D S G G A L S M D TGACGTGCGGATGCGGTTTCGAGCCTTCAAGGACGTGCTCAAGCCCGAAACCGGAAACCGGCATCCACGCCAACAACCTCAGCCTTGG 3240 D V R M R F R A F K D V L K P E T E T G I H A H H N L S L G VANSIVAIEEGCDRVDASLAGMGAGAGNAP LEVFIAAAERMGWHHGTELYKLMDAADDIV GCGGCCGTTGCAGGATCGCCCCGTGCGCGTGGACCGCGAGACGCTGGCGCTGGGCTACGCCGGCGTGTACAGCAGCTTCCTGCGCCATTC 3510 R P L Q D R P V R V D R E T L A L G Y A G V Y S S F L GGAGGCGGCACAGAAGTACGGACTGAAGACCATGGACATCTTGGTCGAACTGGGGAGACGCCGCATGGTCGGCGGCCAGGAAGACAT 3600 E A A A Q K Y G L K T M D I L V E L G R R R M V G G Q E D M GATCGTCGATGTCGCTCTTGACCTGCTCAAGTCGCTGGAACATGAGCGCATCCATGCCCAGCCGGTTTCCAGCGAAGCGGGATGACATCC 3690 I V D V A L D L L K S L E H E R I H A Q P V S S E A G TCAAGGAGCACGCATGAAGCAAGGACGCATCTTTCTGTGGGCGCCTGGGTCTTTTTTACCTGGCGAACCTGCTGGGCACGCTGCCGTTTGC 3780 M K Q G R I F L W A L G L F Y L A N L L G T L P F A (ORF4) CTCGGCCTCGCTGTTCAGCCAGATGTACCCAGGCTTCGTGCCTGACGCGGCCACGCCCGGCCTTTCGGTTGCTCAGCAACGCCTGGGCCGT 3870 S A S L F S Q M Y P G F V P D A A T P G F R L L S N A W A V TGTGGGCCTGCAGCTGGGCGCCATCGGCATCGTGGCCTTGTGGGGTGCGCGCATCCGCTGCGCTGCCGCGGTGTTCGACATCGTCAT 3960 V G L Q L G A I G I V A L W G A R D P L R Y L A V F D I V I TGCCACCGAGGTGGTTGATGGCTTGTGGGACTTTTACAGCATCACCTGGAGCCACCGGCATTCGGGTTGACGACTTTGGTGAT 4050 A T E V V D G L W D F Y S I T W S H L A T A F G L T T L V I CCACCCTGTGTGGATCGTTTGGGCGCGCTGTACGCCAGGCGTGCAGTGTTGAAGTCGGCTCGGTGAGTACGCGCACTCGCCGGGCCTTTTAA 4140 H P V W I V W A L Y A R R A V L K S A R * TGTTGGCAGTTTGGGTACTAGGAGACATG

M translational start codon of bphA1

FIG. 3. Nucleotide sequence of the upstream region of *bphA1* and the deduced amino acid sequences of *bphE*, *bphG*, *bphF*, and ORF4. The deduced amino acid sequences of *bphE*, *bphG*, *bphF*, and ORF4 are shown in one-letter code, and asterisks indicates stop codons. The putative Shine-Dalgarno sequences complementary to the 3' ends of the 16S rRNAs of both *Pseudomonas aeruginosa* and *E. coli* (26) are double underlined. The putative promoter sequence, which is similar to the nitrogen fixation promoter (-24TGGC - 12TTGCT) (6, 31), is underlined.



FIG. 4. SDS-PAGE analyses of products of *bphE*, *bphG*, *bphF*, and ORF4 overexpressed in *E. coli* MV1190. The conditions for induction and electrophoresis are described in Materials and Methods. Total proteins of *E. coli* MV1190 harboring pKH402 (containing *bphE*, *bphG*, *bphF*, and ORF4) (lanes 2 and 3) and pKH406 (containing ORF4) (lanes 5 and 6) are shown. Lanes 2 and 5, noninduced cells; lanes 3 and 6, cells induced with IPTG. The low-range molecular weight standards (lanes 1 and 4) (Bio-Rad Laboratories) were used to estimate the sizes of products of *bphE*, *bphG*, *bphF*, and ORF4 (indicated by arrows).

teins corresponding to sizes of 39, 34, 29, and 16 kDa from *E. coli* harboring pKH402 and 16 kDa from *E. coli* harboring pKH406 were observed. These molecular sizes approximately match those determined on the basis of the deduced amino acid sequences of *bphE*, *bphG*, *bphF*, and ORF4 (28.2, 33.1, 38.3, and 15.0 kDa, respectively).

Assays of *meta*-cleavage enzymes activities. Activities of 2-hydroxypenta-2,4-dienoate hydratase, 4-hydroxy-2-oxovalerate aldolase, and acetaldehyde dehydrogenase (acylating) in crude extracts of biphenyl-induced *Pseudomonas* sp. strain KKS102 and *E. coli* strains harboring pKH plasmids were assayed. These enzyme activities of crude extract from *E. coli* harboring pAQN1 were also assayed as a control (Table 2).

Crude extract of E. coli harboring pKH402 [containing bphEGF(ORF4)] or pKH403 (containing bphE) exhibited 2-hydroxypenta-2,4-dienoate hydratase activity much higher than that of E. coli harboring pKH404, pKH405, pKH406, or vector pAQN1. The level of 4-hydroxy-2-oxovalerate aldolase activity of crude extract from E. coli harboring pKH402 [containing bphEGF(ORF4)] or pKH405 (containing bphF) was remarkably higher than the background level. The background level of 4-hydroxy-2-oxovalerate aldolase was higher than those of 2-hydroxypenta-2,4-dienoate hydratase and acetaldehyde dehydrogenase (acylating). This has been observed in a Pseudomonas strain (23) and Pseudomonas sp. strain CF600 (27). Acetaldehyde dehydrogenase (acylating) activity was detected in crude extract of E. coli harboring pKH402 [containing bphEGF(ORF4)] or pKH404 (containing bphG). Acetaldehyde dehydrogenase (acylating) activity was detected only in crude extract of E. coli harboring bphG. These data show that bphE, bphG, and bphF encode 2-hydroxypenta-2,4dienoate hydratase, acetaldehyde dehydrogenase (acylating), and 4-hydroxy-2-oxovalerate aldolase, respectively. No metacleavage enzyme activity was detected in crude extract of E. coli harboring pKH406 (containing ORF4 alone).

Crude extract of biphenyl-induced *Pseudomonas* sp. strain KKS102 exhibited these three *meta*-cleavage enzyme activities. The activities were of the same level as those of crude extract from *E. coli* harboring *bphE*, *bphG*, and *bphF*.

## DISCUSSION

In this study, we sequenced an approximately 4-kb upstream region of the *bph* gene cluster and found three genes (*bphE*,

TABLE 2.	Enzyme acti	ivities of c	rude extr	racts from	m strains
expre	ssing various	genes of	the bph	gene clus	ster

	Enzyme activity ^a						
Extract	HPH (U/mg)	HOA (mU/mg)	ADA (mU/mg)				
Pseudomonas sp. strain KKS102 (biphenyl induced)	79	68	66 (223) ^b				
E. coli MV1190 carrying:							
pAQN1	0.5	8.2	$ND^{c}$				
pKH402	51	61	134				
pKH403	78	7.3	ND				
pKH404	1.3	8.0	127				
pKH405	0.7	45	ND				
pKH406	1.5	1.3	ND				

 a  HPH, 2-hydroxypenta-2,4-dienoate hydratase; HOA, 4-hydroxy-2-oxovalerate aldolase; ADA, acetaldehyde dehydrogenase (acylating).

^b The ADA activity with a 10-fold-higher concentration of NAD⁺ in the reaction mixture is shown in parentheses. Shingler et al. reported that the reaction rate decreased very rapidly (27).

^c ND, not detected.

bphG, and bphF) for the meta-cleavage pathway and an unknown ORF (ORF4). bphE was homologous to 2-hydroxypenta-2,4-dienoate hydratase of Pseudomonas sp. strain CF600 (27) and the TOL plasmid (14), and crude extract from E. coli harboring pKH402 or pKH403 showed 2-hydroxypenta-2,4dienoate hydratase activity. These results clearly show that bphE encodes 2-hydroxypenta-2,4-dienoate hydratase. Protein purification studies showed that 2-hydroxypenta-2,4-dienoate hydratase (XylJ) from the TOL plasmid was associated with 4-oxalocrotonate decarboxylase (XyII) (11). This probably ensures that unstable 2-hydroxypenta-2,4-dienoate is efficiently metabolized to 4-hydroxy-2-oxovalerate. The association of decarboxylase and hydratase was also observed in Pseudomonas sp. strain CF600, and this association seems to be required for both enzyme activities (27). In our study, however, crude extract of E. coli expressing bphE alone showed 2-hydroxypenta-2,4-dienoate hydratase activity, and its enzyme activity was of the same level as that in crude extract of biphenylinduced Pseudomonas sp. strain KKS102. BphE can probably express 2-hydroxypenta-2,4-dienoate hydratase activity stably independently of association with decarboxylase.

Sequence analysis and enzyme activity assay show that bphGand bphF encode acetaldehyde dehydrogenase (acylating) and 4-hydroxy-2-oxovalerate aldolase, respectively. Studies of protein purification showed that DmpG (4-hydroxy-2-oxovalerate aldolase) and DmpF (acetaldehyde dehydrogenase [acylating]) were associated with each other, and this association could probably ensure efficient metabolism of short-chain acetaldehydes (22). High amino acid identities of BphG versus DmpF (78.3% identity) and BphF versus DmpG (80.8% identity) suggest that the properties of DmpF and DmpG proteins may be conserved in BphG and BphF.

The predicted amino acid sequence of ORF4 showed homology with ORF3 from *P. pseudoalcaligenes* KF707, whose function is unknown (29). ORF3 from KF707 was not necessary for biphenyl dioxygenase activity (29). ORF4 was not also necessary for biphenyl dioxygenase activity (16) by the *meta*cleavage pathway. Detailed experiments are needed to identify the function of ORF4.

The genes for enzymes of the branches of the *meta*-cleavage pathway, 2-hydroxymuconic semialdehyde dehydrogenase, 4-oxalocrotonate isomerase, and 4-oxalocrotonate decarboxylase, were identified in the TOL plasmid (10), plasmid NAH7 (35), and *Pseudomonas* sp. strain CF600 (27). These genes,



FIG. 5. Comparison of the *bph* gene structures in *Pseudomonas* sp. strain KKS102 and *P. pseudoalcaligenes* KF707. The *bphA1A2A3A4BCD* genes in KKS102 and those in KF707 encode the same respective enzymes (7, 16, 17, 29). The functions of ORF1 and ORF4 in KKS102 and ORF3 in KF707 are unknown (16, 29), and ORF4 in KKS102 is homologous to ORF3 in KF707. The *bphX* region in KF707 contains *meta*-cleavage genes (8).

however, have not been found in the *bph* gene cluster from KKS102. Computer analysis indicates that there is no ORF of appropriate size upstream of *bphE*, and functionally uncharacterized ORF1 and ORF4 are not homologous to these genes for the branched pathway. These genes may be downstream of *bphA4* or may not exist in KKS102.

The meta-cleavage genes (bphEGF) and functionally unknown ORF4 identified in this study are clustered, and there is no transcription terminator sequence (21) between ORF4 and bphA1. The average G+C content (63.5%) of these four genes is approximately the same as that (62.9%) of already analyzed genes in the bph cluster. These data suggest that the bphEG-F(ORF4) genes are also members of the bph gene cluster. Thus, the bph gene cluster of biphenyl and PCB degradation in KKS102 characterized so far comprises 12 genes, in the order bphEGF(ORF4)A1A2A3BCD(ORF1)A4.

The sequences related to transcriptional regulation were searched in the sequenced region of this *bph* gene cluster, and a putative promoter sequence, which is similar to the nitrogen fixation promoter (-24TGGC - 12TTGCT) (6, 31), was found upstream of *bphE* (Fig. 3). It was previously reported that another putative promoter sequence, which is also similar to the nitrogen fixation promoter (-24TGGC - 12TTGCT) (6, 31), was found upstream of *bphC* (18). No sequence similar to the terminator sequence (21) was found in the sequenced region. These data suggest that the *bph* gene cluster is transcribed into one transcript.

When the bph gene cluster of KKS102 is compared with that of P. pseudoalcaligenes KF707 (Fig. 5), the meta-cleavage genes (bphE, bphG, and bphF) are found upstream of bphA1 in KKS102, while the meta-cleavage genes in KF707, which are homologous to bphE, bphG, and bphF, are found in the bphXregion (7a). The gene order for 2-hydroxypenta-2,4-dienoate hydratase, acetaldehyde dehydrogenase (acylating), and 4-hydroxy-2-oxovalerate aldolase is preserved in KKS102, CF600 (27), and KF707 (7a). ORF4 in KKS102 is located upstream of bphA1, but ORF3 in KF707, which is homologous to ORF4, is located between bphA2 and bphA3 (29). Moreover, bphA4 (encoding ferredoxin reductase) in KF707 is clustered with bphA1A2 (encoding iron-sulfur proteins) and bphA3 (encoding ferredoxin) (29), but bphA4 in KKS102 is separated from bphA1A2A3 (16). These comparisons of bph gene organization in KKS102 and KF707 suggest that recombination and rearrangement events have occurred during formation of the bph gene clusters.

This study shows the catabolic pathway of biphenyl and PCBs in KKS102 (Fig. 1). Chlorobenzoic acid is produced by the function of 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hy-

drolase (the product of *bphD*). KKS102 cannot grow on chlorobenzoic acid as a sole source of carbon, but *P. fluorescens* KKL101, which was isolated with KKS102, can (15). KKS102 appears not to have the genes for chlorobenzoic acid degradation. These data suggest that both KKS102 and the symbiont KKL101 synthesize a biphenyl and PCB degradation system.

## ACKNOWLEDGMENTS

We thank V. Shingler for the kind gift of *Pseudomonas* sp. strain CF600 and for advice on synthesizing L(S)-4-hydroxy-2-oxovalerate. We also thank K. Furukawa for providing unpublished data and for helpful discussion.

This work was partly supported by a fund from the Railway Technical Research Institute and by a grant-in-aid from the Ministry of Education, Science, and Culture of Japan.

#### REFERENCES

- Assinder, S. J., and P. A. Williams. 1988. Comparison of the *meta* pathway operons on NAH plasmid pWW60-22 and TOL plasmid pWW53-4 and its evolutionary significance. J. Gen. Microbiol. 134:2769–2778.
- Assinder, S. J., and P. A. Williams. 1990. The TOL plasmids: determinants of the catabolism of toluene and the xylenes. Adv. Microb. Physiol. 31:1–69.
- Burlage, R. S., S. W. Hooper, and G. S. Sayler. 1989. The TOL (pWW0) catabolic plasmid. Appl. Environ. Microbiol. 55:1323– 1328.
- Collinsworth, W. L., P. J. Chapman, and S. Dagley. 1973. Stereospecific enzymes in the degradation of aromatic compounds by *Pseudomonas putida*. J. Bacteriol. 113:922–931.
- Dagley, S., and D. T. Gibson. 1965. The bacterial degradation of catechol. Biochem. J. 95:466–474.
- Dixon, R. 1986. The xylABC promoter from the Pseudomonas putida TOL plasmid is activated by nitrogen regulatory genes in Escherichia coli. Mol. Gen. Genet. 203:129–136.
- 7. Fukuda, M., Y. Yasukochi, K. Kimbara, M. Takagi, and K. Yano. Unpublished data.
- 7a.Furukawa, K. Personal communication.
- Furukawa, K., J. Hirose, A. Suyama, T. Zaiki, and S. Hayashida. 1993. Gene components responsible for discrete substrate specificity in the metabolism of biphenyl (*bph* operon) and toluene (*tod* operon). J. Bacteriol. 175:5224–5232.
- Harayama, S., and M. Rekik. 1989. Bacterial aromatic ringcleavage enzymes are classified into two different gene families. J. Biol. Chem. 264:15328–15333.
- Harayama, S., and M. Rekik. 1990. The meta cleavage operon of TOL degradative plasmid pWW0 comprises 13 genes. Mol. Gen. Genet. 221:113–120.
- Harayama, S., M. Rekik, K.-L. Ngai, and L. N. Ornston. 1989. Physically associated enzymes produce and metabolize 2-hydroxy-

2,4-dieonate, a chemically unstable intermediate formed in catechol metabolism via *meta* cleavage in *Pseudomonas putida*. J. Bacteriol. **171**:6251–6258.

- Harayama, S., M. Rekik, A. Wasserfallen, and A. Bairoch. 1987. Evolutionary relationships between catabolic pathways for aromatics: conservation of gene order and nucleotide sequences of catechol oxidation genes of pWW0 and NAH7 plasmids. Mol. Gen. Genet. 210:241-247.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351–359.
- 14. Horn, J. M., S. Harayama, and K. N. Timmis. 1991. DNA sequence determination of the TOL plasmid (pWW0) *xylGFJ* genes of *Pseudomonas putida*: implications for the evolution of aromatic catabolism. Mol. Microbiol. 5:2459–2474.
- 15. Kikuchi, Y. Unpublished data.
- Kikuchi, Y., Y. Nagata, M. Hinata, M. Kimbara, M. Fukuda, K. Yano, and M. Takagi. 1994. Identification of the *bphA4* gene encoding ferredoxin reductase involved in biphenyl and polychlorinated biphenyl degradation in *Pseudomonas* sp. strain KKS102. J. Bacteriol. 176:1689–1694.
- Kimbara, K., T. Hashimoto, M. Fukuda, T. Koana, M. Takagi, M. Oishi, and K. Yano. 1988. Isolation and characterization of a mixed culture that degrades polychlorinated biphenyls. Agric. Biol. Chem. 52:2885-2891.
- Kimbara, K., T. Hashimoto, M. Fukuda, T. Koana, M. Takagi, M. Oishi, and K. Yano. 1989. Cloning and sequencing of two tandem genes involved in degradation of 2,3-dihydroxybiphenyl to benzoic acid in the polychlorinated biphenyl-degrading soil bacterium *Pseudomonas* sp. strain KKS102. J. Bacteriol. 171:2740–2747.
- 19. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680-685.
- Norrander, J., T. Kempe, and J. Messing. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101-106.
- 21. Platt, T. 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. 55:339–372.
- Powłowski, J., L. Sahlman, and V. Shingler. 1993. Purification and properties of the physically associated *meta*-cleavage pathway enzymes 4-hydroxy-2-ketovalerate aldolase and aldehyde dehydrogenase (acylating) from *Pseudomonas* sp. strain CF600. J. Bacteriol. 175:377-385.

- Sala-Trepat, J. M., K. Murray, and P. A. Williams. 1972. The metabolic divergence in the *meta* cleavage of catechols by *Pseudomonas putida* NCIB 10015. Physiological significance and evolutionary implications. Eur. J. Biochem. 28:347–356.
- 24. 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.
- 25. Schell, M. A. 1990. Regulation of the naphthalene degradation genes of plasmid NAH7: example of a generalized positive control system in *Pseudomonas* and related bacteria, p. 165–177. *In S. Silver, A. M. Chakrabarty, B. Iglewski, and S. Kaplan (ed.), Pseudomonas*: biotransformations, pathogenesis, and evolving biotechnology. American Society for Microbiology, Washington, D.C.
- Shine, J., and L. Dalgarno. 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) 254:34–38.
- 27. Shingler, V., J. Powlowski, and U. Marklund. 1992. Nucleotide sequence and functional analysis of the complete phenol/3,4-dimethylphenol catabolic pathway of *Pseudomonas* sp. strain CF600. J. Bacteriol. 174:711-724.
- Smith, M. R. 1990. The biodegradation of aromatic hydrocarbons by bacteria. Biodegradation. 1:191-206.
- Taira, K., J. Hirose, S. Hayashida, and K. Furukawa. 1992. Analysis of *bph* operon from the polychlorinated biphenyl-degrading strain of *Pseudomonas pseudoalcaligenes* KF707. J. Biol. Chem. 267:4844–4853.
- Terada, I., S.-T. Kwon, Y. Miyata, H. Matsuzawa, and T. Ohta. 1990. Unique precursor structure of an extracellular protease, aqualysin I, with NH₂- and COOH-terminal pro-sequences and its processing in *Escherichia coli*. J. Biol. Chem. 265:6576–6581.
- Thöny, B., and H. Hennecke. 1989. The -24/-12 promoter comes of age. FEMS Microbiol. Rev. 63:341-358.
- 32. van der Meer, J. R., W. M. de Vos, S. Harayama, and A. J. B. Zehnder. 1992. Molecular mechanisms of genetic adaptation to xenobiotic compounds. Microbiol. Rev. 56:677–694.
- Vieira, J., and J. Messing. 1987. Production of single-stranded plasmid DNA. Methods Enzymol. 153:3–11.
- 34. Vieira, J., and J. Messing. Methods Enzymol., in press.
- Yen, K.-M., and I. C. Gunsalus. 1982. Plasmid gene organization: naphthalene/salicylate oxidation. Proc. Natl. Acad. Sci. USA 79: 874–878.