Analysis of 5' region of glutelin genes from wild rice species

Hsin-Kan Wu¹, Tein-chin Chen, and Mei-Chu Chung

Institute of Botany, Academia Sinica, Nankang, Taipei, Taiwan, Republic of China

(Received June 7, 1995; Accepted September 13, 1995)

Abstract. The structure of the 5' flanking region of glutelin genes amplified from the various wild rice species was analyzed by cloning and sequencing. The results showed that beyond the essential boxes (legumin, CAAT, AACA and TATA), the 5' region of rice glutelin genes have numerous putative enhancers (long-direct and short-direct repeats) and putative regulatory segments (RY repeats, -300 bp elements, nuclear protein binding sites) though portions of a few elements have been deleted in some wild species. The possible roles of most of the putative elements in glutelin gene expression remain to be determined. The sequence length and structure of glutelin 5' regions vary among rice species. On the basis of the length, the degree of homolgy, and the corresponding base substitutions and deletions in the 5' regions of glutelin genes, the authors suggest that glutelin genes, in the subfamily *Glua* can be classified into three kinds of members, each with its 5' region of 0.5 kb, 0.9 kb, or 1.2 kb. A member gene in the subfamily may reside at one or more loci in a rice genome. The same member gene that appears among rice species with minor deletion, addition, or substitution may be designated as alleles of that gene.

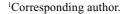
Keywords: Glutelin gene; 5' Region sturcture; Subfamily GluA; Wild rice.

Introduction

A rice glutelin cDNA was first isolated from cultivated rice by Takaiwa et al. (1986). More glutelin cDNAs were isolated and classified into two types, I and II, which can be distinguished respectively by stop coden TAA or TAG and by two and one polyadenylation signals (Takaiwa et al., 1987a). Since these early reports, three genomic clones (Gt1, Gt2 and Gt3) for rice glutelin have been isolated and studied in Okita's laboratory. Comparison of DNA sequences from relevant regions of these clones showed that two of them, Gt1 and Gt2, are closely related. Gt3 shows little or no homology to Gt1 and Gt2. All three clones had 5' flanking regions of less than 0.9 kb (Okita et al., 1989). Two new glutelin genes, Glua-3 and Glua-4 were later added to subfamily A (Takaiwa and Oono, 1991) that already contains Glua-1 (Type I) and Glua-2 (Type II). Takaiwa et al. (1991) proposed a new subfamily of glutelin gene, subfamily B, in which three member genes have been sequenced. Furthermore, direct repeat, enhancer core, and legumin box (Takaiwa et al., 1987b), -300 bp element, RY repeats and inverted repeats (Okita et al., 1989) and nuclear protein binding sites (Kim and Wu, 1990; Takaiwa and Oono, 1990) have been reported to be related to glutelin gene expression in cultivated rice. This paper reports the structure of 5' regions of glutelin genes with special emphasis on that of glutelin subfamily A genes from wild rice species.

Materials and Methods

The species, both cultivated and wild, used in this experiment are shown in Figure 1.



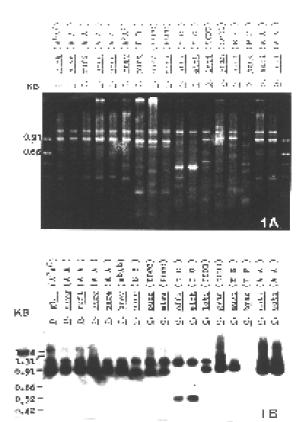


Figure 1. Southern blot analysis of the amplified rice glutelin 5' region. 1A, Gel electrophoresis of the PCR amplified 5' regions of glutelin genes from various rice species each indicated above the line. Each has two to three major bands; 1B, The blotted DNA was probed with the 5' region of a glutelin gene, isolated from rice cultivar Tainung 67 (genome AA). It shows that all wild rice species except *O. brachyantha* (genome FF) have two major bands positively reacted.

bot371-07.p65 41 2001/4/19, PM 01:48

Polymerase chain reaction (PCR) was used to amplify the 5' region of glutelin genes from various rice species. The two ends of the 5' region of a known glutelin gene (Takaiwa et al., 1986) were used to synthesize two primers, a2–1 (5' CAAGCTTTTGGAAAGGTGCCG3') and a2–2 (5'GCTCTAGAGTTGTTGTAGGACTAATGAA3') each with a *Xbal*I or a *Hind*III linker. The amplified 5' regions were cloned into the plasmid M13 *mp*19 to produce recombinant DNA molecules.

Deletion and sizing: Double-stranded recombinant DNA of M13 was extracted and digested with exonuclease III to produce successively shortened insert DNA using the Erase – a Base System (Promega). Transformation of *E. coli* JM101 was carried out with the deleted recombinant DNA using an *E. coli* Pluser Apparatus. Sizing was performed by electrophoresis of the single stranded shortened recombinant DNA from M13.

Sequencing and Analysis: The single-stranded recombinant DNAs were used as template and annealed to the fluorescent primers supplied in the Auto Sequencing Kit (Pharmacia). Sequencing reactions were carried out according to the procedures suggested by the supplier, and the products were loaded in an automated Laser Fluorescent DNA Sequencer (Pharmacia ALF). Data thus obtained were processed by a GCG (Genetics Computer Group) sequence analysis software package.

Results

For PCR amplification of the 5' regions of the rice glutelin genes, two oligomers were synthesized. The amplified DNAs were analyzed by gel electrophoresis. As shown in Figure 1A, each lane gave two major bands and several minor ones. Major bands were identified as true glutelin 5' regions of the species by Southern blot analysis using the 5' region from a glutelin gene isolated from Tainung 67 (cultivated rice of genome AA) as probe (Figure 1B). There were two major bands, 1.2 kb and 0.9 kb in length, amplified from each rice species except that from

O. eichingeri and O. officinalis in which the 0.9 kb band was substituted by a 0.5 kb band (Figure 1B). Only the 0.9 kb band has been shown to appear in glutelin genes of the cultivated rice (Takaiwa et al., 1987b, 1991). The other two bands are being described in thispaper for the first time.

Sequencing of the seven cloned 5' regions of glutelin genes from the wild species of five genomes revealed their lengths (Table 1) which correspond to their molecular weight estimated from the gel electrophoresis of the PCR products. These sequences can be grouped into three categories according to the length of the PCR amplified DNAs. These are 1.2 kb, 0.9 kb, and 0.5 kb respectively.

All seven 5' region sequences were arranged to give maximum alignment as shown in Figure 2. Each sequence listed in Figure 2 has at its upstream 5' end AGCTT or GCTT that is part of the cutting site of Hind III. This and the recovered primer sequence at both ends of the seven sequences assure that the amplified sequences are the true 5' regions of glutelin genes. The figure also includes the sequences of the Gt1, Gt2, and Gt3 reported by Okita et al. (1989). Comparing the regions with each other, it is clear that the 0.9 kb 5' region has a short deletion of 20 bp from -834 bp to -815 bp (from the translation initiation codon ATG) and another long deletion of 204 bp from -470 bp to -267 bp. As for the 0.5 kb 5' region, it has two long deletions; one is 463 bp in length from -1055 bp to -593 bp, and another is 211 bp, from -470 bp to -260 bp, that coincides with the long deletion in the 0.9 kb 5' region. Thus the 0.9 kb 5' region is about 200 bp shorter than the 1.2 kb 5' region, and the 0.5 kb 5' region is about 670 bp shorter than the 1.2 kb 5' region. Such long deletions combined with minor deletions of bases account for the actual length (in bp) of each 5' region sequence listed

There are 209 base-substitutions or -deletions that have occurred to corresponding positions in the sequence of the 1.2 kb and the 0.9 kb 5' regions. The same amount of such corresponding substitutions and deletions can also be

OC 11 4 6			1 (1 (. 1 , 1.	C	•		•
Table 1	' region	cedilence	length of	Gliitelin	genes tron	1 Warionic	species of r	100
Table 1. J	1 CEIOH	SCUUCIICC	ւնուբալ Ծւ	gruttiii	genes non	i various	SUCCICS OF I	ICC.

Species*		Genome	Length (kb) estimated from gel electrophoresis	Actual length (bp)
Oryza perrennis	(W0107)	AA	1.2	1,119
Oryza eichingeri	(W1519)	CC	1.2	1,116
Oryza punctata	(W1564)	BBCC	1.2	1,111
Oryza punctata	(W1564)	BBCC	0.9	911
Oryza grandiglumis	(W1194)	CCDD	0.9	913
Oryza australiensis	(W0008)	EE	0.9	912
Oryza sativa	(Gt1)	AA	_	779
Oryza sativa	(Gt2)	AA	_	878
Oryza eichingeri	(W1519)	CC	0.5	481
Oryza sativa	(Gt3)	AA	<u> </u>	842

^{*}Only some of the species indicated in Figure 1 was chosen to be cloned and sequenced. The length of the 5' region of *Gt1*, *Gt2*, and *Gt3* was calculated based on the sequences published (Okita, 1989). The sequence of the 5' region of glutelin gene of each wild rice species has been deposited to the DataBank of Japan (DDBJ) Tsukuba, Japan. The given accession number of each sequence is as follows: clone W0107-1.2: D26363; clone W1564-0.9: D26364; clone W1564-1.2: D26365; clone W1519-0.5: D26366; clone W1519-1.2: D26367; clone W1194-0.9: D26368; clone W0008-0.9: D26369.

bot371-07.p65 42 2001/4/19, PM 01:48

W0107120 W1619120		-1151		7 bp rept 5
W1564120	Oryza eichingeri (00) GC Oryza punctata (3900) T	110 101 120		
W1564085	Oryza punctata (EBCC)T			
W1194095	Orzya grandigiumis (CCCC) AGCT			
W0008085	Oryza australiensis (⊞)CT	W1194095		
Gt.1	Oryza sativa (AA)	W0009095		
Gt2	Oryza sativa (AA)	Gt1 . Gt2	GCATTCTCCA CTGACATAAT GCAAAATAAG ATATGATOSA TGACATAGC GCATTCTCCA C.GACATAAT GCAAAAGAAG ATATAATCTA TGACATAGC	
W1519050	Oryza eichingeri (CC) AGCT	W1519050		
St.3	Oryza sativa (AA)	Gt3	AATTCATCCT TOCTACCAAC TTGCATGATA TTATATTTGT GAATATCCT	
			RY rept 6	
W0107120	TITGGAAAGS TGCCGTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG	-1101 W0107120		F -701 RY repeat 5. 4
W1519120	TT7GGAAAGG TOCCGTGCAG TTC/VAGAGT TAGTTAGCAG TAGGATGAAG			
W1564120	TITOGAAAGG TOCCGTOCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG			
W1564085	TTTGGAAAGG TGCCGTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG			
W1194095	TTTGGAAAGG TGCCSTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG	, managed a		
#0006095 Gt1	TTTOGAMAGG TGCCGTGCAG TTCAAAGAGT TAGTTAGCAG TAGGATGAAG	Gt1		
Gt2		GLZ		
W1519050	TTTGGAAAG3 TGCCGTGCCG TTCAAACAAT TAGTTAGCAG TAGGATGTTG			
Gt3	***************************************	313	RY rept 4	
			Box V	
		W0107120	AGGTAAGTAT CTTAAGGTAA AGTGTTAGAA CTTCCCATAC ATAAGTCATA	
W0107120	ATTITIGCAC ATGGCAATGA GAAGTTAATT ATGGTGFAGG CANOOCAAAT	marray and	AGGTAAGTAT CITAAGCTAA AGTGTTAGAA CITCCCATAC ATAAGTCATA AGGTAAGTAT CITAAGCTAA AGTGTTAGAA CITCCATAC ATAAGTGATA	
W1519120	ATTITIGCAC ATGCCAATGA GAAGTTAATT ATGCTGTAGG CAACCCAAAT	WAEGAOOE	ACATAAGTAT CITGAGCTAA ASTGTTAGAA CATSAGACC ATAAGTCAC	
W1564120 W1564085	A.TTTTGCAC AT.GCAATGA GAAGTTAATT ATGGTGTA.G CAACCCAAAT A.TTTTGCAC ATGGCAATGA GAAGTTAATT ATGGTGTAGG CAACCCAAAT	W110400F	ACATAAGTAT CTTGAGCTAA A TGTTAGAA CAT AAAGCC ATAAGTCAG	
W11964085	A.TTTTGCAC ATGGCAATGA GAAGTTAATT ATGGTGTAGG CAACOCAAAT A.TTTTGCAC ATGGCAATGA GAAGTTAATT ATGGTGTAGG CAACOCAAAT	WXXXBX96	ACATAAGTAT CITCAGCTAA ALTGYTAGAA CATLAAACCC ATAAGTCAC	
W0008095	ATTITIGCAC ATGGCAATGA GAAGTTAATT ATGGTGTAGG CAACOCAAAT	GET	ACATAAGTAT CITEAACTAA AETGITAGAA CATEARAGC ATAAGTCAD	
Gt1	······································	GLZ	AGGTAAGTAT CETAAACTAA AGTGTTAGAA CTTOOCATAC ATAAGTCAT/	
Gt2	, ,,,,,,,,,		FATTTGAGAG CCTTTGTATA GCTGAAACCA ACGTATATOG AGCATGGAAA	
W1519050	GCTTTTGCTC ACAGCAATGA GAAGTTAATT ATGGTGTAOG CGTGA		RY rept 3	
Gt3			ACTOLTOGOL ATTOCOTOTA LOLOUTENA	004.004
		#0107120 #1519120	ACTGATGACA ATTGGGTGTA ACACATGACA AACCAGAGAG TCAAS ACTGATGACA ATTGGGTGTA ACACATGACA AACCAGAGAG TCAAG	
W0107120	CANACACONA AATATOCACA ACAGAATTTO TTOTATTOTO TACTOCO	9/456/4/20	ACTGATGACA ATTOGGTGTA ACACATGACA AACCAGAGAG TCAAG	
WU107120 W1519120	GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA	-1UU1 W19640R5	THIGATGAGT ATTAGOGGTG ACACATGACA AATCAGAGAG TCAAG	
W1564120	GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA	#11940e6	NAME OF A STATE OF A S	
W1564085	GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA	MAXXOUGU	THEATGAGA ATTAGGOSTO ACACATGACA AASCAGAC TCAAG	
W1194095	GAAACACCAA AATATGCACA AGACAATTTG TTGTATTCTG TAGTACAGAA	961	TITGATGAGI ATTAGGGGTG ACACATGACA AATCAGAGAC TCAAG ACTGATGACA ATTGGGTGT. ACACATGACA AACCAGAGAG TCAAG	
1400080095	GAAACACCAA AATATOCACA AGACAATTTG TTGTATTCTG TAGTACAGAA			
9t1	TTCTG TAGTACAGAC	Gts	AGAGAAGAAA ATOCAAGGAT FTTTTTATTC TGGTTCATGC CCTGGATGC	3
\$ t 2	TTCTG TAGTACAGAA		RY rept 2	
(11519050 Gt.3				
		up407420	CAA CATAAANNA AANNATTING TANATAAAN TANANING	6 554 7 ha namb 6
		W0407120		
#010/120 #1519120	TAAACTUAAA GTAATGAAAG AAGATOGT-STEAGAAAAT GAAACAATAT -861 Inw	werted rept W1519120	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT	'A
W010/120 W1519120 W1564120	TAAA <mark>CT.AA</mark> A GTAATGAAAG AAGA TIGGT: STE FAGAAAAT GAAACAATAT -661 I'm TAAACT.AAA GTAATGAAAG AAGATIGGT GTTAGAAAAT GAAACAATAT TAAACT.AAA GTAATGAAAAG AAGATIGGT GTTAGAAAAT GAAACAATAT	werted rept W1519120 hancer core W1564120 W1564085	CAA GATAAAGDAA AAGGATGT.G TACATAAAAC TACAGAGCTCAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCTCAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT	A A A
W010/120 W1519120 W1584120 W1584085	TAMACTUMA GTAATGAANG AAGA. TOST STEAGAAAAT GAACAATAT -851 Im TAMACTUMA GTAATGAANG AAGA. TOST GTUGAAAAT GAACAATAT EN TAMACTUMA GTAATGAANG AAGA. TOST GTUGAAAAT GAACAATAT TAMACTUMA GTAATGAANG AAGATGTGGT GTUGAAAAAG GAACAATAT	werted rept W1519120 hancer core W1564120 W1664085 W1194096		A A A
W010/120 W1519120 W1564120 W1564065 W1184065	TAMACT.AMA GTAATGAAMG AAGATOGT GTAGAAMAT GAMACAMTAT -661 Inn TAMACT.AMA GTAATGAAMG AAGATOGT GTAGAAMAT GAMACAMTAT ENP TAMACT.AMA GTAATGAAMG AAGATOGT GTAGAAMAT GAMACAMTAT TAMACT.AMA GTAATGAAMG AAGATGTGT GTTAGAAMA	werted rept W1519120 hancer core W1564120 W1664025 W11924026 W0008026		A A A A
W010/120 W1519120 W1584120 W1584085	TAMACTUMA GTAATGAANG AAGA. TOST STEAGAAAAT GAACAATAT -851 Im TAMACTUMA GTAATGAANG AAGA. TOST GTUGAAAAT GAACAATAT EN TAMACTUMA GTAATGAANG AAGA. TOST GTUGAAAAT GAACAATAT TAMACTUMA GTAATGAANG AAGATGTGGT GTUGAAAAAG GAACAATAT	werted rept W1519120 hancer core W1564120 W1664085 W1194096		A A A A A
W010/120 W15/9/120 W1564/120 W1564/065 W1184/065 W10005065 GE1 GE12	TAMACT.AMA GTAATGAAMG AAGATOGT GTAGAAAAT GAACCATAT -661 Inv TAMACT.AMA GTAATGAAMG AAGATOGT GTAGAAAAT GAACCATAT TAMACT.AMA GTAATGAAMG AAGATOGT GTAGAAAAT GAACCATAT TAMACT.AMA GTAATGAAMG AAGATGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAMG AAGATGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAAG AAGATGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAGA GAACTGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAGA GAACTGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAGA GAACTGTGT GTTAGAAAAG GAACCATAT	werted rept W1519120 hancer core W1564120 W1564025 W1192026 G11 G12 W1519050	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TECAGAGCT CAA GATAAACCAA AATGATGT.G TACATAAAAC TACAGAGCT CAA CATAAACCAA AATGATGT.G TACATAAAAC TACAGAGCT CAC CATAAACCAA AAGGATGT.G TACAAAAAAC TACAGAGCT	A A A A A A A
#310/120 #159120 #1564126 #1564085 #1192065 #510005065 \$12 W1519060	TAAACT.AAA GTAATGAAAG AAGA. TGGT GTAGAAAAT GAAACAATAT -861 IIM TAAACT.AAA GTAATGAAAG AAGA. TGGT GTAGAAAAT GAAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAACAATAT AAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAT GAACAATAT	werted rept W1519120 hancer core W1564120 W1564120 W1664025 W1134026 W000006 G11	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TECAGAGCT CAA GATAAACCAA AATGATGT.G TACATAAAAC TACAGAGCT CAA CATAAACCAA AATGATGT.G TACATAAAAC TACAGAGCT CAC CATAAACCAA AAGGATGT.G TACAAAAAAC TACAGAGCT	A A A A A A A
W010/120 W15/9/120 W1564/120 W1564/065 W1184/065 W10005065 GE1 GE12	TAMACT.AMA GTAATGAAMG AAGATOGT GTAGAAAAT GAACCATAT -661 Inv TAMACT.AMA GTAATGAAMG AAGATOGT GTAGAAAAT GAACCATAT TAMACT.AMA GTAATGAAMG AAGATOGT GTAGAAAAT GAACCATAT TAMACT.AMA GTAATGAAMG AAGATGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAMG AAGATGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAAG AAGATGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAGA GAACTGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAGA GAACTGTGT GTTAGAAAAG GAACCATAT TAMACT.AMA GTAATGAAGA GAACTGTGT GTTAGAAAAG GAACCATAT	werted rept W1519120 hancer core W1564120 W1564025 W1192026 G11 G12 W1519050	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TECAGAGCT CAA GATAAACCAA AATGATGT.G TACATAAAAC TACAGAGCT CAA CATAAACCAA AATGATGT.G TACATAAAAC TACAGAGCT CAC CATAAACCAA AAGGATGT.G TACAAAAAAC TACAGAGCT	A A A A A A A A A
W010/120 W1559/120 W1564/120 W1564/065 W1124/065 W10056065 GE1 W1519050 GE3	TAAACT.AAA GTAATGAAAG AAGA. TGGT GTAGAAAAT GAAACAATAT -861 IIM TAAACT.AAA GTAATGAAAG AAGA. TGGT GTAGAAAAT GAAACAATAT TAAACT.AAA GTAATGAAAG AAGAT. TGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAAACAATAT AAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAT GAAACAATAT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W159026 W100026 G11 G12 W1519050 G13	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TECAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAAC TACAGAGCT CAC GATAMAGCAA AATGATGT. G TACATAMAAC TACAGAGCT TAATATGGT GATCATCAAA AAGGATGT. CATAMAAT TAMAGTAAT TAGTCATGT TACAMACAA GAGAGCTTA TAMAGACACC CATGACTCA	A A A A A A A A A A A A A A A A A A A
W010/120 W1519120 W1562120 W1562085 W1124065 W0008065 Gt1 Gt2 W1519000 Gt3	TAMACT.AMA GTAATGAAMG AAGATGGT-GEFAGAAMAT GAMACAMTAT -661 Inn TAMACT.AMA GTAATGAAMG AAGATGGT GT-AGAAMAT GAMACAMTAT TAMACT.AMA GTAATGAAMG AAGATGGT GT-AGAAMAT GAMACAMTAT TAMACT.AMA GTAATGAAMG AAGAITGTGT GT-AGAAMAG GAMACAMTAT TAMACT.AMA GTAATGAAMG AAGAITGTGGT GT-AGAAMAG GAMACAMTAT TAMACT.AMA GTAATGAAMG AAGAITGTGT GT-AGAAMAG GAMACAMTAT TAMACT.AMA GTAATGAAMG AAGAITGTGGT GT-AGAAMAG GAMACAMTAT TAMACT.AMA GTAATGAMG AAGAITGTGGT GT-AGAAMAG GAMACAMTAT TAMACT.AMA GTAATGAMG AAGAITGTGGT GT-AGAAMAG GAMACAMTAT TAMACT.AMA GTAATGAMG AAGAITGTGGT GT-AGAAMAT GMAACAMTAT TAMACT.AMA GTAATGAMG AAGAITGTGGT GT-AGAAMAT GMAACAMTAT TAMACT.AMA GTAATGAMG AAGAITGTGGT GT-AGAAMAT GMAACAMTAT TAMACT.AMA GTAATGAMG AAGAITGTGTGT GT-AGAAMAT GMAACAMTAT TAMACT.AMA GTAATGAMG AAGAITGTGTGT GT-AGAAMAT GMAACAMTATAT TAMACT.AMA GTAATGAMG AAGAITGTGTGT GT-AGAAMAT GMAACAMTATAT TAMACT.AMA GTAATGAMG AAGAITGTGTGT GT-AGAAMAT GMAACAMTATAT TAMACT.AMA GTAATGAMG AAGAITGTGTGT GT-AGAAMAT GMAACAMTATATATATATATATATATATATATATATATATA	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W1519020 G13 G13 Crepeat 7 W1519120 W1519120	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAC CATAAACCAA AAGGATGT.G TACATAAAAC TCCAGAGCT TTAATATGGT GATCATCAAA AAAGATATGCATAAAAT TAAAGTAAT TATGTCATGT TGCGAAAAACA GGAGGCTTA TAAGACAACC CATGACTCA TATGTCATGT TGCGAAAAACA GGAGCTTA TAAGACAACC CATGACTCA TATGTCATGT TGCGAAAAACA GGAGCTTA TAAGACAACC CATGACTCA	A A A A A A A A A A A A A A A A A A A
W010/120 W1559/120 W1564/120 W1564/065 W1124/065 W10056065 GE1 W1519050 GE3	TAAACT.AAA GTAATGAAAG AAGA. TGGT GTAGAAAAT GAAACAATAT -861 IIM TAAACT.AAA GTAATGAAAG AAGA. TGGT GTAGAAAAT GAAACAATAT TAAACT.AAA GTAATGAAAG AAGAT. TGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAAACAATAT AAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAG GAAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTGGT GTAGAAAAT GAAACAATAT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W159006 G11 G12 W1519050 G13 repeat 7 W1519120 W1519120 W1519120 W1519120	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TCCAGAGCT TAATATOGT GATCATCAAA AAGGATGT.G TACATAAAAC TCCAGAGCT TAATATOGT GATCATCAAA AAAGATATGCATAAAAAT TAAAGTAAT TATGTCATGT TGCGAAAAGA GGAGAGCTTA TAAGAGAACC CATGACTCC TATGTCATGT TGCGAAAAGA G. GAGCTTA TAAGAGAACC CATGACTCC TATGTCATGT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCC TATGTCATGT TGCGAAAAGA GGAGAGCTTA TAAGACAACC CATGACTCC	A A A A A A A A A A A A A A A A A A Direct rept S A A
6010/120 6/15/91/20 6/15/91/20 6/15/91/20 6/16/16/91/20 6/13 6/13/91/20 6/13/91/20 6/15/91/20 6/15/91/20 6/15/91/20	TAMACT.AMA GTAATGAAMG AAGATGGT-SEEAGAAAAT GAAACAATAT TAAACT.AMA GTAATGAAMG AAGATGGT GTAGAAAAT GAAACAATAT TAAACT.AMA GTAATGAAMG AAGATGGT GTAGAAAAT GAAACAATAT TAAACT.AMA GTAATGAAMG AAGAISTGTGT GTAGAAAAG GAAACAATAT TAAACT.AMA GTAATGAAMG AAGAISTGTGT GTAGAAAAG GAAACAATAT TAAACT.AMA GTAATGAAMG AAGAISTGTGT GTAGAAAAG GAAACAATAT TAAACT.AMA GTAATGAAAG AAGAISTGTGT GTAGAAAAG GAAACAATAT TAAACT.AMA GTAATGAAMG AAGAISTGTGT GTAGAAAAG GAAACAATAT TAAACT.AMA GTAATGAAMG AAGAISTGTGT GTAGAAAAG GAAACAATAT TAAACT.AMA GTAATGAAMG AAGAITGTGTGTGAGAAAG GAAACAATAT TAAACT.AMA GTAATGAAMG AAGAITTATAGATAAT GTGTGAGCAT TATGGGGACCA CGAAATAAAA AAAGAACATT TATGACTAAT GTGTGAGCAT TATGGGGACCA CGAAATAAAA AAAGAACATT SATGGACTAAT GTGTGAGCAT TATGGGGACCA CGAAATAAAA AAAGAACATT SATGGACTAAT GTGTGAGCAT TATGGGGACCA CGAAATAAAA AAAGAACATT SATGGACTAAT GTGTGAGCAT TATGGGGACCA CGAAATAAAA AAAGAACATT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W159020 G11 G12 W1519020 G13 W0107120 W1519120 W1564120 W1564120 W1564120 W1564120 W1564120 W1564120	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TĞCAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TĞCAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TĞCAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCT CAC GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT TTAATATOGT GATCATCAA AAGGATGT. G TACAAMAC TACAGAGCT TATGTCATGT TACAGAMAGA GAGAGCTTA TAGAGAAGC CATGACTCA TATGTCATGT TACAGAMAGA GAGAGCTTA TAGAGAGAGC CATGACTCA TATGTCATGT TACAGAMAGA GAGAGCTTA TAGAGAGAGC CATGACTCA TATGTCATGT TACAGAMAGA GAGAGCTTA TAGAGAGAGC CATGACTCA TATGTCATGT TAGAGAAAGA GAGAGCTTA TAGAGAAGAG CATGACTCA TATGTCATATT TAGAGAAAGA GAGAGCTTA TAGAGAAGAG CATGACTCA TATGTCATATT TAGAGAAAGA GAGAGCTTA TAGAGAAGAG CATGACTCA TATGTCATGT TAGAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	A A A A A A A A A A A A A A A A A A A
4010/12/2 4/15/9/12/0 1/15/9/12/0 1/15/9/12/0 1/16/2/0 6/1 6/1 1/0/10/12/0 1/15/9/12/0 1/15/9/12/0 1/15/9/12/0 1/15/9/12/0 1/15/9/12/0	TAMACT.AMA GTAATGAAMG AAGA. TISET SEEAGAAMAT GAMACAATAT -661 IIM TAMACT.AMA GTAATGAAMG AAGA. TISET GTTAGAAMAT GAMACAATAT TAMACT.AMA GTAATGAAMG AAGA. TISET GTTAGAAMAT GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMG GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMG GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMG GAMACAATAT AMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMAG GAMACAATAT AMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMAG GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMAT GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMAT GAMACAATAT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W159006 G11 G12 W1519050 G13 repeat 7 W1519120 W1519120 W1519120 W1519120	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAC CATAAAGCAA AATGATGT.G TACATAAAAC TCCAGAGCT TTAATATCGT GATCATCAAA AAAGATATGCATAAAAT TAAAGTAAT TATGTCATGT TGCGAAAAGA GCAGAGCTTA TAAGACAACC CATGACTCA TATGTCATGT TGCGAAAAGA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAAGAA GGAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAAAGAA GGAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAAAGAA GGAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAAAGAA GGAGAGCTTA TAAGACAAGAG CATGACTCA TATGTCATAT TGCGAAAAAGAA GGAGAGCTTA TAAGACAAGAG CATGACTCA TATGTCATAT TGCGAAAAAGAA GAGAGCTTA TAAGACAAGAG CATGACTCA TATGTCATAT TGCGAAAAAGAA GAGAGCTTA TAAGACAAGAC	A A A A A A A A A A A A A A A A A A A
6010/120 6/15/91/20 6/15/91/20 6/15/91/20 6/16/16/91/20 6/13 6/13/91/20 6/13/91/20 6/15/91/20 6/15/91/20 6/15/91/20	TAMACTIMA GTAATGAAAG AAGA. TOST-SETAGAAAAT GAAACAATAT -661 Ink TAAACTI.AAA GTAATGAAAG AAGA. TOST GTAGAAAAT GAAACAATAT TAAACTI.AAA GTAATGAAAG AAGA. TOST GTAGAAAAT GAAACAATAT TAAACTI.AAA GTAATGAAAG AAGA. TOST GTAGAAAAG GAAACAATAT TAAACTI.AAA GTAATGAAAG AAGARITETIST GTAGAAAAG GAAACAATAT TAAGCTAAT GTGTGAGCAT TATGGGACCA GGAAATAAAA AAAGAACAT TATGACTAAT GTGTGAGCAT TATGGGACCA GAAATAAAA AAAGAACAT TATGACTAAT GTGTGAGCAT TATGGGACCA GAAATAAAA AAAGAACAT RATGACTAAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACAT RATGACTAAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACAT CATGAGTAAT GTGTGAGCAT TATGGACCA CAAATAAAA AAAGAACAT CATGAGTAAT GTGTGAGCAT TATGGACCA CAAATAAAA AAAGAACAT CATGAGTAAT GTGTGAGCAT TATGGACCA CAAATAAAA AAAGAACAT CATGAGTAAT TATGGACCA CAAATAAAA AAAGAACAT CATGAGTAAT TATGGACCA CAAATAAAA AAAGAACAT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564026 W100006 G11 G12 W1519050 G13 W0107120 W1519120 W1564120 W1564120 W1560006 W1134005 W1050006	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TECAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCT TAATATOGT GATCATCAAA AAGATATG. TACATAMAC TACAGAGCT TAATATGT GATCATCAAA AAGATATG. TACAGAAAAC CATGACTCA TATGTCATGT TACAGAAAGA GAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATGT TACAGAAAGA GAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TACAGAAAGAA GAGAGCTTA TAAGACAAGCACAGC CATGACTCA TATGTCATATATACAACAACAACAACAACAACAACAACAACAAC	A A A A A A A A A A A A A A A B A B A B
#010/12/0 #159-120 #159-120 #159-120 #159-120 \$11 \$11 \$11 \$11 \$11 \$11 \$11 \$11 \$11 \$1	TAMACT.AMA GTAATGAAMG AAGA. TISET SEEAGAAMAT GAMACAATAT -661 IIM TAMACT.AMA GTAATGAAMG AAGA. TISET GTTAGAAMAT GAMACAATAT TAMACT.AMA GTAATGAAMG AAGA. TISET GTTAGAAMAT GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMG GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMG GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMG GAMACAATAT AMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMAG GAMACAATAT AMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMAG GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMAT GAMACAATAT TAMACT.AMA GTAATGAAMG AAGATGETISET GTTAGAAMAT GAMACAATAT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAMA AMAGAACATT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564025 W1194025 G13 Crepeat 7 W1519120 W1564120 W1564120 W1564120 W1564120 W1564120 W1564120 G11 G11 G12 G11 G11 G12 G11 G11 G12 G11 G11	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TCCAGAGCT TAATATCGT GATCAACCAA AAGGATGT.G TACATAAAAC TCCAGAGCT TAATATCGT GATCATCAAA AAGGATGT.G TACATAAAAC TCCAGAGCT TAATGCATGT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATGT TGCGAAAAGA GGAGGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAGA GGAGGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAAGA GGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAAGA GGAGGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAGAA GGAGGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAAGA GGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAAGA CATGACTCA TATGTCATAT TAGACAACCACCATCATCATCATCATCATCATCATCATCATCA	A A A A A A A Direct S A Direct rept S A A A A A A A A A A A A A A A A A A
#010/120 #159120 #159120 #159208 #1192095 #1192096 #01075296 6t3 #0107120 #159120 #159208 #01022096 \$1192096 #1192096 #1192096 #1192096 #1192096 #1192096	TAMACTIMA GTAATGAAAG AAGA. TOGT GTAGAAAG GAACAATAT -661 Inv TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAGT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAGT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTTGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTTGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATTGTAGAAAAT GAACAATAT TATGACTAAT GTGTGACCAT TATGGGACCA CGAAATAAAA AAAGAACAT TATGACTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT TATGACTTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT QATGAGTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT CATGAGTAAT GTGTGACCAT TATGGACCAC CAAATAAAAAAAAAA	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W1194026 Gt1 Gt2 W1519050 Gt3 repeat 7 W1519120 W1564120 W1564120 W156006 W11340266 W11340266 W11340266 W11340266 W11340266 W11340266	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT TAATATGGT GATCATCAAA AAAGATATGCATAAAAAT TAAAGTAAT TATGTCATGT TGCGAAAAGA GGAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATGT TGCGAAAAGA GGAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAAGA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAAAGA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAAAGA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAAAGA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAAACAA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAAACAA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAAACAA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAACAAA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAACAAA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAACAAA GCAGAGCTTA TAAGACAAGC CATGACTCA CATGACTATAT TGCAAAACAAAACAAACAAACAAACAAACAAACAAACAAA	A A A A A A A A A A A A A A B A B B B B
#010/12/0 #15/9/20 #15/9/20 #15/206 #11/200 6t3 #010/5/96 6t3 #010/12/2 #15/9/20 #15	TAMACT.AAA GTAATGAAAG AAGA. TISET SEEAGAAAAT GAACCAATAT -661 IIM TAAACT.AAA GTAATGAAAG AAGA. TISET GTTAGAAAAT GAACCAATAT TAAACT.AAA GTAATGAAAG AAGA. TISET GTTAGAAAAT GAACCAATAT TAAACT.AAA GTAATGAAAG AAGA. TISET GTTAGAAAAG GAACCAATAT TAAACT.AAA GTAATGAAAG AAGAGETISET GTTAGAAAAG GAACCAATAT TAAACT.AAA GTAATGAAAG AAGAGETISET GTTAGAAAAG GAACCAATAT TAAACT.AAA GTAATGAAAG AAGAGETISET GTTAGAAAAG GAACCAATAT AAACT.AAA GTAATGAAAG AAGAGETISET GTTAGAAAAG GAACCAATAT TAAACT.AAA GTAATGAAAG AAGAGETISET GTTAGAAAAA GAACCAATAT TAAACT.AAA GTAATGAAAG AAGAGETISET GTTAGAAAAA GAACCAATAT TAAACT.AAA GTAATGAACA CAACACCACACCACACCACACCACA	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564025 W1194025 G13 Crepeat 7 W1519120 W1564120 W1564120 W1564120 W1564120 W1564120 W1564120 G11 G11 G12 G11 G11 G12 G11 G11 G12 G11 G11	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAA GATAAACCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAA GATAAACCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAA GATAAACCAA AATGATGT.G TACATAAAAC TACAGAGCT CAA CATAAACCAA AAGGATGT.G TACATAAAAC TACAGAGCT TAATATOGT GATCATCAAA AAAGATATGCATAAAAAT TAAAGTAAT TATGTCATGT TGCGAAAACA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATGT TGCGAAAAGA GGAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAAGA GGAGAGCTTA TAAGACAAGG CATGACTCA TATGTCATAT TGCGAAAAAGA GGAGAGCTTA TAAGACAAGG CATGACTCA TATGTCATAT TGCGAAAAGA GGAGAGCTTA TAAGACAAGG CATGACTCA TATGTCATAT TGCGAAAAGA GGAGAGCTTA TAAGACAAGG CATGACTCA TATGTCATGT TGCGAAAAGA GGAGACCTTA TAAGACAAGG CATGACTCA TATGTCATGT TGCGAAAAGA GGAGACCTTA TAAGACAAGG CATGACTCA TATGTCATGT TGCGAAAACGA CAGAGCCTTA TAAGACAAGC CATGACTCA TATGTCATGT TGCGAAAACGA CAGACCTTA TAAGACAAGC CATGACTCA TATGTCATGT TGCGAAAACGA CAGACCTTA TAAGACAAGC CATGACTCA TATGTCATGT TGCGAAACACA CAGACCCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAACAA	A A A A A A A A A A A A A A Direct rept S A A A A A A A A A A A A A A A A A A A
#010/120 #159120 #159120 #159208 #1192095 #1192096 #01075296 6t3 #0107120 #159120 #159208 #01022096 \$1192096 #1192096 #1192096 #1192096 #1192096 #1192096	TAMACTIMA GTAATGAAAG AAGA. TOGT GTAGAAAG GAACAATAT -661 Inv TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAGT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAGT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTTGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTTGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATTGTAGAAAAT GAACAATAT TATGACTAAT GTGTGACCAT TATGGGACCA CGAAATAAAA AAAGAACAT TATGACTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT TATGACTTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT QATGAGTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT CATGAGTAAT GTGTGACCAT TATGGACCAC CAAATAAAAAAAAAA	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W1519050 Gt3 V1519050 Gt3 V1519120 W1564120 W156412	CAA GATAMAGCAA AMGATET, G TACATAMAC TACAGAGCE CAA GATAMAGCAA AMGATET, G TACATAMAC TACAGAGCE CAA GATAMAGCAA AMGATET, G TACATAMAC TACAGAGCE CAA GATAMAGCAA AMTGATET, G TACATAMAC TECAGAGCE CAA GATAMAGCAA AMGATET, G TACATAMAC TACAGAGCE THATATOST GATCATCAAA AMGATET, G TACATAMAC TACAGAGCE TATATCAT TACAGAGAA AMGATET, G TACAMAMAC TACAGAGCE TATGTCATGT TACAMAGCA GAGAGCETTA TACAGAACC CATGACTC TATGTCATGT TACAMAGCA GAGAGCETTA TACAGAACC CATGACTC TATGTCATT TACAMAGCA GAGAGCETTA TACAGAGCAGC CATGACTC TATGTCATAT TACAMAGCA GAGAGCETTA TACAGCAGC CATGACTC TATGTCATAT TACAMAGCA GAGCAGCTTA TACAGCAGC CATGACTC TATGTCATAT TACAMACACA GAGCAGCTTA TACAGCAGC CATGACTC TATGTCATAT TACAGCACAC CAGCAGCTTA TACAGCAGC CATGACTC TATGTCATAT TACAGCACAC CAGCAGCTTA TACAGCAGC CATGACTC ATTTCCATAT TACAGCACAC CAGCAGCTTA TACAGCACAC CATGACTC ATTTCCATAT TACAGCACAC CAGCAGCTTA TACAGCACAC CATGACTC ATTTCCATAT TACAGCACAC CATGACTC TATGTCATAT TACAGCACAC TACACC TATGTCATAT TACAGCACAC TACACC TATGTCATAT TACAGCACAC CATGACTC TATGTCATAT TACACACAC CATGACTC TATGTCATAT TACACCACAC TATGTCATATAT TACACACAC CATGACTC TATGTCATATAT TACACACAC CATGACTC TATGTCATATATACC TATGTCATATATACC TATGTCATATATACC TATGTCATATATACC TATGTCATATACC TATGTCATATACC TATGTCATATACC TATGTCATATACC TATGTCATATACC TATGT	A A A A A A A A A A A A A A A A Direct sept s A A A A A A A A A A A A A A A A A A A
#010/12/0 #15/9/20 #15/9/20 #15/20/9 #15/20/9 6t1 #010055/9 6t3 #0107120 #15/9/20 #1	TAMACT.AMA GTAATGAAMS AAGA TOST GTTAGAAAAT GAMACAMTAT -661 INN TAMACT.AMA GTAATGAAMS AAGA TOST GTTAGAAAAT GAMACAMTAT TAMACT.AMA GTAATGAAMS AAGA TOST GTTAGAAAAT GAMACAMTAT TAMACT.AMA GTAATGAAMS AAGATST GTTAGAAAAS GAMACAMTAT TAMACT.AMA GTAATGAAMS AAGATSTOST GTTAGAAAAS GAMACAMTAT TAMACT.AMA GTAATGAAGA AAGATSTOST GTTAGAAAAT GAMACAMTAT TAMACT.AMA GTAATGAAGA AAGATSTOST GTTAGAAAAT GAMACAMTAT TAMACT.AMA GTAATGAAGA TATGGGACCA CGAMATAMAA AMAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAMATAMAA AAAGAACATT CATGACTAAT GTGTGAGCAT TATGGGACCA CGAMATTAMA AAGAACATT CATGACTAAT GTGTGAGCAT TATGGGACCA CGAMATAMAA AAAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAMATAMAA AAAGAACATT CATGACTAAT GTGTGACCAT TATGGGACCA CGAMATAMAA AAAGAACATT CATGACTAAT GTGTGACCAT TATGGGACCA CGAMATAMAA AAAGAACATT CATGACTAAT GTGTGACCA CGAMATAMAA AAAGAACATT CATGACTAAT GTGTGACCA CGAMATAMAA AAAGAACATT CATGACTAAT GTGTGACCA CGAMATAMAA AAAGAACATT CATGACTAATAT GTGTGACCA CGAMATAMAA AAAGAACATT CATGACTAATATAATATAATATATATATATATATATATAT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1505026 G11 G12 W1519050 G13 V0107120 W1519120 W1564120 W1564120 W1569050 G13 W159050 G13 W1564120 W1569050 G13 W1569050 W1569050 G13 W1569050 G13	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TECAGAGCT CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCT TACATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCT TAATATOGT GATCATCAAA AAGATATG. TACATAMAC TACAGAGCT TATGTCATGT TACAMAGCA GAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATGT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TACAMAACAA GAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TACAMAACAA GAGAGACTTA TAAGACAAGC CATGACTCA TATGTCATAT TACAMAACAA GAGAGCCTTA TAAGACAAGC CATGACTCA TATGTCATAT TACAMAACAA GAGAGCCTTA TAAGACAAGC CATGACTCA TATGTCATAT TACAMAACAA GAGAGCCTTA TAAGACAACC CATGACTCA TATGTCATAT TACAMAACAA GAGAGCCTTA TAAGACAACCA CATGACTCA TATGTCATAT TACAMAACAA GAGAGCCTTA TAAGACAACCA CATGACTCA TATGTCATAT TACAMAACAA TATGTCATA TATGTCATAT TACAMAACAA TATGTCATAAAAAA TATGTCATAAAAAAA TATGTCATAAAAAAAA TATGTCATAAAAAAAA TATGTCATAAAAAAAAAA	A A A A A A A A A A A A A A A A A A A
#010/12/20 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/0 6t3 #010/7/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0 #15/9/12/0	TAMACTIMA GTAATGAAAG AAGA. TOGT GTAGAAAT GAACAATAT -661 Inv TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAAT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAAT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTTGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATTGTGT GTAGAAAAG AAACAATAT TAAGATTAAT GTGTGACAT TATGGGACCA CGAAATAAAA AAAGAACAT TATGAACTAAT GTGTGACCAT TATGGACCAC CGAAATAAAA AAAGAACAT TATGAACTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAAAA	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W1569025 G13 Frepeat 7 W1519120 W1564120 W	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TCAGAGGCT CAC CATAAACCAA AAGGATGT.G TACATAAAAC TCAGAGGCT TTAATATGGT GATCATCAAA AAGGATGT.G TACATAAAAC TCAGAGGCT TAATGCATGT TGCGAAAACA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATGT TGCGAAAACA GGAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAGA GGAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAACAC GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAACAC GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAAACA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAACACA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAACACA GAGAGCATTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAACACACA GAGGACCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAACACA GAGGACCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAACACACA GAGGACCTTA TAAGACAACC CATGACTCA AATTTGCTCA TAAGACCACA GAGACCTTA TAAGACAACC CATGCACAC AAAAAATTCA CATGCCTACT GTGCCCCATA TATCATCCAA CAATCCAACACACACACACAC	A A A A A A A A A A A A A A A A A A B A B A B A B A
#010/120 #159-120 #159-120 #159-120 #159-206 613 #0107-120 #059-109 #059-10	TAMACTAMA GTAATGAAAG AAGA TOGT GTAGAAAAT GAACAATAT TAAACTAAA GTAATGAAAG AAGA TOGT GTAGAAAAT GAACAATAT TAAACTAAA GTAATGAAAG AAGA TOGT GTAGAAAAT GAACAATAT TAAACTAAA GTAATGAAAG AAGA TOGT GTAGAAAAG GAACAATAT TAAACTAAA GTAATGAAAG AAGARTGTGTGT GTAGAAAAG GAACAATAT TAAACTAAA GTAATGAAAG AAGARTGTGTGT GTAGAAAAG GAACAATAT TAAACTAAA GTAATGAAAG AAGARTGTGT GAGAAAG GAACAATAT TAAACTAAA GTAATGAAAG AAGARTGTGT GTAGAAAAG AAAGAACATT TAAACTAAA GTAATGAACAT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAAATAAAA AAAGAACATT AATGACTAAT GTGTGAGCAT TATGGGACCA CGAAATAAAA AAAGAACATT QATGAGTAAT GTGTGAGCAT TATGGGACCA CGAAATAAAA AAAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATTAAAA AAAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATTAAAA AAAGAACATT CATGAGTAAT GTGTGAGCAT TATGGACCA CGAATTAAAA AAAGAACATT CATGAGTAAT GTGTGAGCAT TATGGACCA CGAATTAAAA AAAGAACATT CATGAGTAAT GTGTGAGCAT TATGGACCAT GAATGTTATC TCACCCAGGA -B51 TTTTATGAGCA GTGTGTCTC AATGAGCCTT GAATGTTATC TCACCCAGGA TTTTATGAGCA GTGTGTCTC AATGAGCCTT GAATGTTATC TCACCCAGGA TTTTATGAGCA GTGTGTCTC AATGAGCCTT GAATGTTATC TCACCCAGGA	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1505026 G11 G12 W1519050 G13 V0107120 W1519120 W1564120 W1564120 W1569050 G13 W159050 G13 W1564120 W1569050 G13 W1569050 W1569050 G13 W1569050 G13	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT TACATCATGT.GCCAAAACAA CAGGATGT.G TACATAAAAC TCCAGAGCT TAATATGGT GATCATCAAA AAAGATATGCATAAAAAT TAAAGTAAT TATGTCATGT TGCGAAAACAA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATGT TGCGAAAACAA GGAGCACTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAACAA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAACAA GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAACAAC GCAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAAAACAAC GCAGACCTTA TAAGACAAGC CATGACTCA AATTTTGCTCA TAAGAAACCAACCAACCAACCAACCAACCAACCAACCAAC	A A A A A A A A A A A A A A A A A A A
#010/120 #159120 #159120 #159206 #119206 #119206 #0105596 6t3 #0107120 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206 #159206	TAMACTIMA GTAATGAAAG AAGA. TOGT GTAGAAAT GAACAATAT -661 Inv TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAAT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAAT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATGTTGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGATTGTGT GTAGAAAAG AAACAATAT TAAGATTAAT GTGTGACAT TATGGGACCA CGAAATAAAA AAAGAACAT TATGAACTAAT GTGTGACCAT TATGGACCAC CGAAATAAAA AAAGAACAT TATGAACTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CAAATAAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAA AAAGAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAAACAT CATGAAGTAAT GTGTGACCAT TATGGACCAC CGAAATAAAAAAAAAA	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W156005 G13 Frepeat 7 W1519120 W1564120	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT TAATATGGT GATCATCAAA AAGGATGT.G TACATAAAAC TCAGAGAGCT TTAATATGGT GATCATCAAA AAGGATGT.G TACATAAAAC TCAGAGAGCT TAATGCATGT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATGT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAACAC GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAACAC GAGAGCCTTA TAAGACAACCAC CATGACTCA AAAAAATTCA CATGCCTACT GTGCCCCATA TATGATCCAA CAATCCAAA AAAAAATTCA ATGCCTACT GTGCCCCATA TATCATGCAA CAATCCAAA AAAAAATTCA TTTGCCTTACT GTGCCCCATA TATCATGCAA CAATCCAAA CAAAAATTCA TTTGCCTTTC GTGCCCCATA TATCATGCAAA CAATCCAAA CAAAAATTCA TTTGCCTTTC GTGCCCCATA TATCATCCAA CAATCCAAA CAAAAATTCA TTTGCCTTTC GTGCCCCATA TATCATCCAA CAATCCAAA CAAAAATTCA TTTGCCTTTC GTGCCCCATA TATCATCCAA CAATCCAAA CAAAAATTCA TTTGCCTTTC GTGCCCCATA TATCATCCAAA CAATCCAAA CAAAAATTCA TTTGCCT	A A A A A A A A A A A A A A A A A A A
#010/12/20 #159-120 #159-120 #159-120 #159-120 #1519:000 6t3 #0107-120 #1519:000	TAMACTAMA GTAATGAMG AAGA. TOST STEAGAMAT GAACCATAT -661 Int TAMACT. AAA GTAATGAMA AAGA. TOST GTTAGAMAT GAACCATAT TAMACT. AAA GTAATGAMA AAGA. TOST GTTAGAMAT GAACCATAT TAMACT. AAA GTAATGAMA AAGA. TOST GTTAGAMAG GAACCATAT TAMACT. AAA GTAATGAMAG AAGA. TOST GTTAGAMAG GAACCATAT TAMACT. AAA GTAATGAMAG AAGARTST GTGAGAMAG GAACCATAT TAMACT. AAA GTAATGAMAG AAGARTST GTGAGAMAG GAACCATAT TATGACTATA GTGTGAGCAT TATGGGACCA CGAATTAMA AAGAACATT TATGACTATAT GTGTGAGCAT TATGGGACCA CGAATTAMA AAGAACATT CATGAGTATAT GTGTGAGCAT TATGGGACCA CGAATTAMA AAGAACATT CATGAGCAT TATGGGACCA CGAATTAMA AAGAACATT CATGAGCAT TATGGGACCA CGAATTAMA AAGAACATT TATGAGCACA GTGTGTTCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCA GTGTGTCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCAG GTGTTCCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCAG GTGTTCCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCAG GTGTTCCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCTG GTGTATCCT GATGAGCCTC AAAGATTCCT CACCCAGGA TTTATGAGCAGC GTGTTCCC GATGAGCCTC AAAGATTCCT CACCCAGGA TTTATGAGCCTG GAAGTTTTCC CACCCAGGA TTTATGAGCCTG GAAGTTTTC CACCCCGGA TTTATGAGCCTG GAAGTTTTCC CACCCAGGA TTTATGAGCCTG GAAGTTTTC CACCCCGGA TTTATGAGCCTG GAAGTTTTCC CACCCAGGA TTTATGAGCCTG GAAGTTTTCC CACCCCGGA TTTATGAGCCTG GAAGTTTTCC CACCCCGGA TTTATGAGCCTG GAAGTTTTCC CACCCCGGA TTTATGAGCCTG GAAGTTTTC CACCCCGGA TTT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W1519050 G13 Frepeat 7 W1519120 W1564120	CAA GATAMAGCAA AMGATET, G TACATAMAC TACAGAGCI CAA GATAMAGCAA AMGATET, G TACATAMAC TACAGAGCI CAA GATAMAGCAA AMGATET, G TACATAMAC TECAGAGCI CAA GATAMAGCAA AMTGATET, G TACATAMAC TECAGAGCI CAA GATAMAGCAA AMGATET, G TACATAMAC TACAGAGCI TAATATOST GATCATAMACAA AMGATET, G TACATAMAC TACAGAGCI TTAATATOST GATCATAMA AMGATATG, CATAMAAT TAMAGTAAT TATGTCATGT TACAMAGCA GAGAGCTTA TAMAGCAAGC CATGACTCA TATGTCATGT TACAMAGCAA GAGAGCTTA TAMAGCAAGC CATGACTCA TATGTCATAT TACAMAGCAA GAGAGCTTA TAMAGCAAGC CATGACTCA TATGTCATAT TACAMAGCAAG GAGAGCTTA TAMAGCAAGC CATGACTCA TATGTCATAT TACAMAAGA GAGAGCTTA TAMAGCAAGC CATGACTCA TATGTCATAT TACAMAAGCA GAGAGCATTA TAMAGCAAGC CATGACTCA TATGTCATAT TACAMAAGCA GAGAGCTTA TAMAGCAAGC CATGACTCA TATGTCATAT TACAMAAGCA GAGAGCTTA TAMAGCAAGC CATGACTCA TATGTCATAT TACAMAAGCA GAGAGCATTA TAMAGCAAGC CATGACTCA TATGTCATAT TACAMAAGCA GAGAGCATTA TAMAGCAAGC CATGACTCA AATTTCATATA TACAMACCAA TATGATCAACAAAAATTCA AMAGAATTCA GAGCCCATA TATCATACCAA CATGCCAACAAAAATTCA AATGCCTACT GTGCCCCATA TATCATACCAA CAATCCAAA AAAAAATTCA CATGCCTTACT GTGCCCCATA TATCATACCAA CAATCCAAA AAAAAATTCA CATGCCTTACT GTGCCCCATA TATCATACCAA CAATCCAAA AAAAAATTCA CATGCCTTACT GTGCCCCATA TATCATACCAA CAATCCAAA CAAAAAATTCA TITGCCTTTC GTGTCCAAAAAAAAAAAAAAAACAAAAAAAAAA	A A A A A A A A A A A A A A A A A Inverted rept A A HS1 RY repeat 2, Inverted rept A A A A A A A A A A A A A A A A A A A
#010/12/0 #15/9/20 #15/9/20 #15/206 #119/206 #119/206 Gt3 #0107/20 #15/9/20	TAMACT. AMA GTAATGAMAG AAGA TOST GTT-GRAMAT GAMACANTAT -661 I INT TAMACT. AMA GTAATGAMAG AAGA TOST GTT-GRAMAT GAMACANTAT TAMACT. AMA GTAATGAMAG AAGA TOST GTT-GRAMAT GAMACANTAT TAMACT. AMA GTAATGAMAG AAGA TOST GTT-GRAMAG GAMACANTAT TAMACT. AMA GTAATGAMAG AAGARGTIST GTT-GRAMAG GAMACANTAT TAMACT. AMA GTAATGAMAG AAGARGTIST GTT-GRAMAG GAMACANTAT TAMACT. AMA GTAATGAMAG AAGARGTIST GTT-GRAMAG GAMACANTAT AMACT. AMA GTAATGAMAG AAGARGTIST GTT-GRAMAG GAMACANTAT TAMACT. AMA GTAATGAMAG AAGARGTIST GTT-GRAMAG GAMACANTAT TAMACT. AMA GTAATGAMAG AAGARGTIST GTT-GRAMAG GAMACANTAT TAMACT. AMA GTAATGAMAG AAGARGTIST GTT-GRAMAG GAMACANTAT TATGACTANT GTGTGMCCAT TATGGGGACCA CGAMATAMAA AAMGAACATT TATGACTANT GTGTGMCCAT TATGGGACCA CGAMATAMAA AAMGAACATT CATGAGTAAT GTGTGMCCAT TATGGGACCA CGAMATAMAA AAMGAACATT CATGAGTAAT GTGTGMCCAT TATGGGACCA CGAMATAMAA AAMGAACATT CATGAGTAAT GTGTGMCCAT TATGGGACCA CGAMATAMAA AAMGAACATT TATGAGTAAT GTGTGMCCAT TATGGGACCA CGAMATAMAA AAMGAACATT TATGAGTAAT GTGTGMCCAT TATGGGACCA CGAMATAMAA AAMGAACATT TATGAGTAAT GTGTGACCAT TATGGGACCA CGAMATAMAA AAMGAACATT TATGAGCA GTGGTTCC AATGAGCCTC AAMGATGTTATC TACCCCCGA TTTATGACCA GTGGTTCC AATGAGCCTC AAMGATGTTATC TACCCCCGA TTTATGACCA GTGGTTCC AATGAGCCTC AAMGATGTTATC TACCCCCGCA TTGATGACCT GATGACCCTC AAMGATGTTATC TACCCCCGCA TTGATGACCT GATGACCCTC AAMGATGTTC TACCCCCCGA TTGATGACCT GATGACCCTC AAMGATGTTC TACCCCCGCA TTGATGACCT GATGACCCTC AAMGATGTC TACCCCCGCA TTGATGACCT GATGACCCTC AAMGATGTC TACCCCCCGCA TTGATGACCT GATGACCCTC AAMGATGTC TACCCCCCCCCAATTCACCCTC AAMGATGTC TACCCCCCCCCAATTCACCCTC AAMGATGTC TACCCCCCCCCCAATTCACCCTC AAMGATGTC TACCCC	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564026 W1519050 G13 V1519050 G13 V1519050 G13 V1519050 W1564120 W1564120 W1564120 W15619050 G13 W1519120	CAA GATAMAGCAA ANGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA ANGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA ANTGATGT. G TACATAMAC TGCAGAGCI CAA GATAMAGCAA ANGGATGT. G TACATAMAC TACAGAGCI TAATATOGT GATCATCAAA AMGATATG. TACATAMAC TACAGAGCI TAATATOGT GATCATCAAA AMGATATG. TACATAMAC TACAGAGCI TATGTCATGT TGCAMAAGA GAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATGT TGCAMAAGA GAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCAMAAGA GAGAGCCTTA TAAGACAAGC CATGACTCA AATTTIGCTCATCT GTGCCCATA TATCATCAAC CATGACTCA CAAAAATTCA ATTGCCTTCT GTGCCCCATA TATCATGCAA CAATCCAMA CAAAAATTCA TTTGCCTTTC GTGCCCATA TATCATGCAA CAATCCAMA CAAAAATTCA CTTGCCTTTC GTGCCCATA TATCATGCAA CAATCCAMA CAAAAATTCA CTTGCCTTTC GTGCCCATA TATCATGCAA CAATCCAMA CAAAAATTCA CTTGCCTTTC GTGCTCAMAA CAAAAATTCA CTTGCCTTTC GTGCCCATA TATCATCAAAACAACAACAACAACAACAACAACAACAACA	A A A A A A A A A A A A A A B A B A B A
#010/12/0 #15/9/20 #15/9/20 #15/206 #119/206 #119/206 Gt3 #0107/20 #15/9/20	TAMACTAMA GTAATGAMG AAGA. TOST STEAGAMAT GAACCATAT -661 Int TAMACT. AAA GTAATGAMA AAGA. TOST GTTAGAMAT GAACCATAT TAMACT. AAA GTAATGAMA AAGA. TOST GTTAGAMAT GAACCATAT TAMACT. AAA GTAATGAMA AAGA. TOST GTTAGAMAG GAACCATAT TAMACT. AAA GTAATGAMAG AAGA. TOST GTTAGAMAG GAACCATAT TAMACT. AAA GTAATGAMAG AAGARTST GTGAGAMAG GAACCATAT TAMACT. AAA GTAATGAMAG AAGARTST GTGAGAMAG GAACCATAT TATGACTATA GTGTGAGCAT TATGGGACCA CGAATTAMA AAGAACATT TATGACTATAT GTGTGAGCAT TATGGGACCA CGAATTAMA AAGAACATT CATGAGTATAT GTGTGAGCAT TATGGGACCA CGAATTAMA AAGAACATT CATGAGCAT TATGGGACCA CGAATTAMA AAGAACATT CATGAGCAT TATGGGACCA CGAATTAMA AAGAACATT TATGAGCACA GTGTGTTCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCA GTGTGTCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCAG GTGTTCCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCAG GTGTTCCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCAG GTGTTCCC AATGAGCCTT GAATGTTATC CACCCAGGA TTTATGAGCTG GTGTATCCT GATGAGCCTC AAAGATTCCT CACCCAGGA TTTATGAGCAGC GTGTTCCC GATGAGCCTC AAAGATTCCT CACCCAGGA TTTATGAGCCTG GAAGTTTTCC CACCCAGGA TTTATGAGCCTG GAAGTTTTC CACCCCGGA TTTATGAGCCTG GAAGTTTTCC CACCCAGGA TTTATGAGCCTG GAAGTTTTC CACCCCGGA TTTATGAGCCTG GAAGTTTTCC CACCCAGGA TTTATGAGCCTG GAAGTTTTCC CACCCCGGA TTTATGAGCCTG GAAGTTTTCC CACCCCGGA TTTATGAGCCTG GAAGTTTTCC CACCCCGGA TTTATGAGCCTG GAAGTTTTC CACCCCGGA TTT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W1519050 G13 Frepeat 7 W1519120 W1564120	CAA GATAAAGCAA AAGGATGT.G TACATAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TACAGAGCT TAATATCGT GATAAACCAA AAGGATGT.G TACATAAAAC TACAGAGCT TAATATCGT GATAAACCAA AAGGATGT.G TACATAAAAC TACAGAGCT TAATGTCATGT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATGT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATGT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAAGAA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAAGAA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAAGAA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAAACAA GAGAGCATTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAAACAA GAGAGCATTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAAACAA GAGAGCATTA TAAGACAACC CATGACTCA TATGTCATAT TGCAAAACACAA GAGAGCATTA TAAGACAACC CATGACTCA AAAAAATTCA CATGCCTACT GTGCCCCATA TATCATGCAA CAATGCCAACACAAAATTCA ATTGCCTTACT GTGCCCCATA TATCATGCAA CAATCCAAA CAAAAATTCA CATGCCTACT GTGCCCCATA TATCATGCAA CAATCCAAA CAAAAATTCA CTTTCCTTTC GTGCCCCATA TATCATCCAA CAATCCAAA CAAAAATTCA TTTTCCTTTC GTGCCCCATA TATCATCCAA CAATCCAAAAAAAAAA	A A A A A A A A A A A A A A A A A A A
#010/120 #159/120 #159/120 #159/120 #119/120 #119/120 #15	TAMACTUMA GTAATGAAAG AAGA. TOST STEAGAAAAT GAAACAATAT -661 Ink TAAACT.AAA GTAATGAAAG AAGA. TOST GTTAGAAAAT GAAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOST GTTAGAAAAT GAAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TOST GTTAGAAAAG AAACAATAT TAAACT.AAA GTAATGAAAG AAGAITGTTGT GTTAGAAAAG GAAACAATAT TAAACT.AAA GTAATGAAAG AAGAITGTGTG GTTAGAAAAG AAAGAACATT TAAGACTAAT GTGTGAGCAT TATGGGACCA GAAATAAAA AAAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACATT CATGAGTTAAT GTGTGAGCAT TATGGGACCA CAAATTAAAA AAAGAACATT CATGAGTTAAT GTGTGAGCAT TATGGGACCA CAAATTAAA AAAGAACATT CATGAGTTAAT GTGTGAGCAT TATGGGACCA CAAATTAAAA AAAGAACATT CATGAGTTAAT GTGTGAGCAT TATGGGACCA CAAATTAAAA AAAGAACATT CATGAGTTAAT GTGTGAGCAT TATGGGACCA CAAATTAAAA AAAGAACATT CATGAGTTAAT GTGTGACCAT TATGGGACCA CAAATTAAAA AAAGAACATT CATGAGTTAAT GTGTGAGCAT TATGGGACCA CAAATTAAAA AAAGAACATT CATGAGTTAAT GTGTGACCAT TATGGGACCA CAAATTAAAA AAAGAACATT CATGAGTTAATGGACCA CAAATTAAAA AAAGAACATT CATGAGTTAAT GTGTGACCAT CAATGATTACT CACCCAGGA CTTTATGAGCA GTGTGTTCTC AATGAGCCTT GAATGTTTAC CACCCAGGA CTTTATGAGCA GTGTGTTCTC GATGACCCTC AAAGATTTC CACCCAGGA CTGTTTCTC GATGACCCTC AAAGATTTCT CACCCAGGA CTGTTCTCC GATGACCCTC AAAGATTTCT CA	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W156005 G13 Frepeat 7 W1519120 W1564120 W1	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TECAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI TATATATGT GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI TATATGTAT TACAGAGAA AAGGATGT. G TACAMAMAC TACAGAGCI TATATGTATT TACAGAAAAA AAGAATATGCATAMAAT TAMAGTAAT TATGTCATGT TACAMAGCA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATGT TACAGAAAAGA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGAA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTC AATTTCCTCA TAAGAAACCA GAGAGCTTA TAAGACAAGC CATGACTC AATTTCCTCA TAAGAAACCA GAGAGCCTTA TAAGACAACC CATGACTC AATTTCCTCA TAGGACAACC TAGGACCAACATCCAACAAAATTCA AATTCCTTCT GTGCCCATA TATCATCCAA CAATCCAAC AAAAAATTCA CATGCCTACT GTGCCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA TATCATCCAA CAATCCAACAACAACAACAACAACAACAACAACAACAACA	A A A A A A A A A A A A A A A A A I Inverted rept A A A I Inverted rept A A A A A A A A A A A A A A A A A A A
#010/12/20 #15/91/20 #15/91/20 #15/20/95 \$11 #15/10/05/96 \$13 #15/10/96 \$13 #010/12/96 #15/10/96 #15/10/96 \$11 \$12 #15/10/96 \$11 \$12 #15/10/96 \$13 #15/10/96	TAMACTIMA GTAATGAAAG AAGA. TOST GTAGAAAT GAACCATAT -851 INT TAAACTIMA GTAATGAAAG AAGA. TOST GTAGAAAAT GAACCATAT TAAACTIMA GTAATGAAAG AAGA. TOST GTAGAAAAT GAACCATAT TAAACTIMA GTAATGAAAG AAGA. TOST GTAGAAAAT GAACCATAT TAAACTIMA GTAATGAAAG AAGATISTET GTAGAAAAG GAACCATAT TAAACTIMA GTAATGAAAG AAGATISTEST GTAGAAAAG GAACCATAT TAAACTIMA GTAATGAAAG AAGATISTEST GTAGAAAAG GAACCATAT TAAACTIMA GTAATGAAAG AAGATISTEST GTAGAAAAG GAACCATAT TAAACTIMA GTAATGAAGA AAGATISTEST GTAGAAAAG GAACCATAT TAAACTIMA GTAATGAAGA AAGATISTEST GTAGAAAAG GAACCATAT TAAACTIMA GTAATGAACA AAGATISTEST GTAGAAAAT GAACCATAT TAAACTIMA GTAATGAACA AAGATICATA GAACCATAT TATGACTAT GTGTGACCAT TATGGGACCA CGAAATAAAA AAAGAACATT ATGACTATAT GTGTGACCAT TATGGGACCA CGAAATAAAA AAAGAACATT ATGACTATAT GTGTGACCAT TATGGGACCA CGAAATAAAA AAAGAACATT ATGACTATAT GTGTGACCAT TATGGGACCA CGAAATAAAA AAAGAACATT CATGAGTAAT GTGTGACCAT TATGGGACCA CGAAATAAAA AAAGAACATT CATGAGCA GTGTGTTCTC AATGAGCCTT GAATGTTATC TACCCACGGA TTATGAGCA GTGTGTTCTC AATGAGCCTT GAATGTTATC TACCCACGGA TTATGAGCA GTGTATTCCTC GATGACCCTC AAAGTTTCTC TACCCACGGA TGATGAGCAT GATGACCTTC AATGACCTTC AAAGTTTCTC TACCCACGGA TGATGAGCAT GAT	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W1564025 W1000026 G13 W1519050 G13 W1519050 W1564025 W1564025 W1564026	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TECAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI TATATATGT GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI TATATGTAT TACAGAGAA AAGGATGT. G TACAMAMAC TACAGAGCI TATATGTATT TACAGAAAAA AAGAATATGCATAMAAT TAMAGTAAT TATGTCATGT TACAMAGCA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATGT TACAGAAAAGA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGAA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTC AATTTCCTCA TAAGAAACCA GAGAGCTTA TAAGACAAGC CATGACTC AATTTCCTCA TAAGAAACCA GAGAGCCTTA TAAGACAACC CATGACTC AATTTCCTCA TAGGACAACC TAGGACCAACATCCAACAAAATTCA AATTCCTTCT GTGCCCATA TATCATCCAA CAATCCAAC AAAAAATTCA CATGCCTACT GTGCCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA CAATCCAAC CAAAAATTCA TITGCCTTTC GTGCCCATA TATCATCCAA TATCATCCAA CAATCCAACAACAACAACAACAACAACAACAACAACAACA	A A A A A A A A A A A A A A A A A I Inverted rept A A A I Inverted rept A A A A A A A A A A A A A A A A A A A
#010/12/20 #15/9/120 #15/9/120 #15/20/6 6t1 #15/10/05/96 6t3 #010/120 #15/10/6 #15/10/6 #15/10/6 #15/10/6 #15/10/6 #15/10/6 \$11 \$12 #15/10/6 \$13 #15/10/6 \$10 #15	TAMACTIMA GTANTGAMG AGGA. TOST GTAGAMAT GAMACMITAT -851 INT TAMACTIMA GTANTGAMG AGGA. TOST GTAGAMAT GAMACMITAT TAMACTIMA GTANTGAMG AGGA. TOST GTAGAMAT GAMACMITAT TAMACTIMA GTANTGAMG AGGA. TOST GTAGAMAG GAMACMITAT TAMACTIMA GTANTGAMG AGGARISTICST GTAGAMAG GAMACMITAT TAMACTIMA GTANTGAMG ANGLITICST GTAGAMAG GAMACMITAT TAMACTIMA GTANTGAMG ANGLITICST GTAGAMAG GAMACMITAT TAMACTIMA GTANTGAMG ANGLITICST GTAGAMAG GAMACMITAT TAMACTIMA GTOTOMICAT TATGGGACCA CGAMITAMA AMAGAACATT TATGACTIMI CTGTOMICAT TATGGGACCA CGAMITAMA AMAGAACATT QATGAGTAMI GTGTAGACAT TATGGGACCA CGAMITAMA AMAGAACATT QATGAGTAMI GTGTAGACAT TATGGGACCA CGAMITAMA AMAGAACATT QATGAGTAMI GTGTGAGCAT TATGGGACCA CGAMITAMA AMAGAACATT QATGAGTAMI GTGTGAGCAT TATGGGACCA CGAMITAMA AMAGAACATT QATGAGTAMI GTGTGAGCAT TATGGGACCA CGAMITAMA AMAGAACATT TATGAGTAMI GTGTGAGCAT TATGGGACCA CGAMITAMA AMAGAACATT TATGAGTAMI GTGTGAGCAT TATGGGACCA CGAMITAMA AMAGAACATT TATGAGTAMI GTGTGAGCAT TATGGGACCA CGAMITAMA AMAGAACATT TATGAGCA GTGTGTCTC AMTGGGCCTT GAMITATATC TAXCCAGGAA TTATGAGCA GTGTGTCTC AMTGGGCCTT GAMITATATC TAXCCAGGAA TTATGAGCA GTGTGTCTC AMTGGGCCTT GAMITATATC TAXCCAGGAA TTGATGAGCA GTGTGTTCTC GATGAGCCTT GAMITATATC TAXCCAGGAA TTGATGAGCA GTGTGTTCTC GATGAGCCTT GAMICTTGTC TAXCCAGGAA TTATGAGCA GTGTGTTCTC GATGAGCCTT GAMICTTGTC TAXCCAGGAA TTATGAGCAG GTGTGTTCTC GATGAGCCTT GAMICTTGTC TAXCCAGGAA TTATGAGCAG GTGTGTTCTC ATGAGCCTT GAMICTTGTC TAXCCAGGAA TTATGAGCAG GTGTGTTCTC ATGAGCCTT GAMICTTGTATC TAXCCAGGAA TAXCCAG	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W1564120 W1560026 G11 G12 W1519050 G13 W0107120 W1561025 W1564025 W1564026 W15	CAA GATAMAGCAA AMGATET, G TACATAMAC TACAGAGCI CAA GATAMAGCAA AMGATET, G TACATAMAC TACAGAGCI CAA GATAMAGCAA AMGATET, G TACATAMAC TECAGAGCI CAA GATAMAGCAA AMTGATET, G TACATAMAC TACAGAGCI CAA GATAMACAA AMGATET, G TACATAMAC TACAGAGCI TAATATOST GATAMACAA AMGATET, G TACATAMAC TACAGAGCI TATATATOST GATAMACAA AMGATET, TACAGAMAC CATGACTCA TATGTCATET TACAGAMACA GAGCABCTTA TAMGACAGC CATGACTCA TATGTCATET TACAGAMAGA GAGCABCTTA TAMGACAGC CATGACTCA TATGTCATAT TACAMAGAA GAGCAGCTTA TAMGACAGC CATGACTCA TATGTCATAT TACAMAMAGA GAGCABCTTA TAMGACAGC CATGACTCA TATGTCATAT TACAMAMAGA GAGCAGCTTA TAMGACAGC CATGACTCA TATGTCATAT TACAMAMAGA GAGCACCTTA TAMGACAGC CATGACTCA AMAMATICA TACAGCACA CAGCACCTTA TAMGACACAC CATGACTCA AMAMATICA CATGCCTACT GIGCCCCATA TATCATCACA CATGCCCAC AMAMATICA CATGCCTACT GIGCCCCATA TATCATCACA CATCCCAA AMAMATICA CATGCCTACT GIGCCCCATA TATCATCACA CAATCCAAA AAAAAATTCA CATGCCTTACT GIGCCCCATA TATCATCACA CAATCCAAA AAAAAATTCA TITGCCTTTC GIGCCCATA TATCATCACA CAATCCAAA AAAAAATTCA TITGCCTTTC GIGCCCATA TATCATCACAA CAATCCAAA AAAAAATCA TITGCCTTTC GIGCCCATA TATCATCACAA CAATCCAAA AAAAAATCA TITGCCTTTC GIGCCCATA TA	A A A A A A A A A A A A A A A Inverted rept A A Inverted rept A A A A A A A A A A A A A A A A A A A
#010/12/20 #15/91/20 #15/91/20 #15/20/95 \$11 #15/10/05/96 \$13 #15/10/96 \$13 #010/12/96 #15/10/96 #15/10/96 \$11 \$12 #15/10/96 \$11 \$12 #15/10/96 \$13 #15/10/96	TAMACTIMA GTANTGAMA AGA. TOST GTRIGAMAT GAMACANTAT -651 INT TAMACTIMA GTANTGAMA AGA. TOST GTRIGAMAT GAMACANTAT TAMACTIMA GTANTGAMA AGA. TOST GTRIGAMAT GAMACANTAT TAMACTIMA GTANTGAMA AGA. TOST GTRIGAMAT GAMACANTAT TAMACTIMA GTANTGAMA AGA. TOST GTRIGAMAG GAMACANTAT TAMACTIMA GTANTGAMA AGARISTICST GTRIGAMAG GAMACANTAT TAMACTIMA GTANTGAMAG AGARISTICST GTRIGAMAG GAMACANTAT TAMACTIMA GTANTGAMAG AGARISTICST GTRIGAMAG GAMACANTAT AMACTIMA GTANTGAMAG AGARISTICST GTRIGAMAG GAMACANTAT TAMACTIMA GTANTGAMAG AGARISTICST GTRIGAMAG GAMACANTAT TAMACTIMA GTANTGAMAG AGARISTICST GTRIGAMAG GAMACANTAT TAMACTIMA GTGTOMOCAT TATGGGGCOC GAMATAMAA AMAGAACATT TATGACTANT GTGTOMOCAT TATGGGGCOC GAMATAMAA AMAGAACATT CATGAGTANT GTGTGAGCAT TATGGGGCOC GAMATAMAA AMAGAACATT TATGAGTANT GTGTGAGCAT TATGGGGCOC GAMATAMAA AMAGAACATT TATGAGTAAT GTGTGAGCAT TATGGGCCOC GAMATAMAA AMAGAACATT TATGAGCAG GTGTGTTCC AATGAGCCTT GAMTGTTATC TAXCCOCAGA TTTATGAGCAG GTGTGTCC GATGAGCCTG AMAGTTTCC TAXCCOCAGA TTGATGAGCT GATGAGCCTG AMAGTTTCC TAXCCOCAGA TTGATGAGCA GTGTGTCC GATGACCCTG AMAGTTTCC TAXCCOCAGA TTGATGAGCA GTGTGTCC GATGACCCTG AMAGTTTCC TAXCCOCAGA TTGATGAGCA GTGTGTCC GATGACCCTG AMAGTTTCC TAXCCOCAGA TTATGAGGAG GTGTTCC GATGACCCTG AMAGTTTCC TAXCCOCAGA TTATGAGGAG GTGTTCC GATGACCTG AM	werted rept W1519120 hancer core W1564120 W1564120 W1564120 W1564120 W1564120 W1560026 G11 G12 W1519050 G13 W0107120 W1561025 W1564025 W1564026 W15	CAA GATAMAGCAA ANGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA ANGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA ANTGATGT. G TACATAMAC TECAGAGCI CAA GATAMAGCAA ANTGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA ANGATGT. G TACATAMAC TACAGAGCI TAATATOGT GATCATCAAA AMGATATG. TACATAMAC TACAGAGCI TTAATATOGT GATCATCAAA AMGATATG. TACATAMAC TACAGAGCI TATGTCATGT TACAMAGCA GAGCAGCTTA TAGAGACACC CATGACTC TATGTCATGT TACGAMAAGA GAGCAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGCAGCTTA TACAGCAAGC CATGACTC TATGTCATAT TACAMAAGA GAGCACCTTA TACAGCAAGC CATGACTC AATTTCATAT TACAMAACAA GAGCACCTTA TACAGCAAGC CATGACTC AATTTCATAT TACAGCAACACA GAGCACCTTA TACAGCAAGC CATGACTC AATTTCATAT TACAGCAACACA CAAGACACACACACACACACACACACAC	A A A A A A A A A A A A A A A A A A A
#0107-120 #159-120 #159-120 #159-120 #159-209 #1519-209 #1519-120 #1519-120 #1519-120 #1519-120 #159-1	TAMACTIMA GTANTGAMG AGGA. TOST GTINGAMAT GAMACANTAT -661 INT TAMACTIMA GTANTGAMG AGGA. TOST GTINGAMAT GAMACANTAT TAMACTIMA GTANTGAMG AGGA. TOST GTINGAMAT GAMACANTAT TAMACTIMA GTANTGAMG AGGATIST GTINGAMAG GAMACANTAT TAMACTIMA GTANTGAMG AGGATISTIST GTINGAMAG GAMACANTAT TAMACTIMA GTANTGAMG ANGLISTIST GTINGAMAG GAMACANTAT TAMACTIMA GTANTGAMG ANGLISTIST GTINGAMAG GAMACANTAT TAMACTIMA GTANTGAMG ANGLISTIST GTINGAMAG GAMACANTAT TAMACTIMA GTONTGAMG ANGLISTIST GTINGAMAG GAMACANTAT TAMACTIMA GTONTGAMG ANGLISTIST GTINGAMAG GAMACANTAT TAMACTIMA GTONTGAMG ANGLISTIST GTINGAMAG AMAGAACAT TATAGATAT GTINGACAT TATGGGACCO COMATIMA AMAGAACAT TATGAGTAT GTINGAMCAT TATGGGACCO COMATIMA AMAGAACAT PATGAGTAT GTINGAMCAT TATGGGACCO COMATIMA AMAGAACAT TATGAGTAT GTINGAMCAT TATGGGACCO COMATIMA AMAGAACAT TATGAGCA GTINGTOTIC ANTGGACCT GAMTITTAT CACCOCAGA TIGATGAGCA GTINGTOTIC ANTGACCTO AMAGTITIC CACCOCAGA TIGATGAGCA GTINGTOTIC GATGACCTO AMAGTITIC CACCOCAGA TIGATGAGCA GTINGTOTIC ATGACCOTO AMAGTITIC CACCOCAGA TIGATGAGCA GTINGTOTIC ATGACACTO AMAGTITIC CACCOCAGA TIGATGAGCA GTINGTOTIC AT	werted rept	CAA GATAAAGCAA AAGGATGT.G TACATAAAC TACAGAGCT CAA GATAAAGCAA AAGGATGT.G TACATAAAC TACAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCT CAA GATAAAGCAA AATGATGT.G TACATAAAAC TCCAGAGCT TACATCATGT TACATAAACAA AAGATATG. CATAAAAC TCCAGAGCT TAATATOST GATCATCAAA AAGATATG. CATAAAAC TCCAGAGCT TATGTCATGT TGCGAAAAGA GAGAGCTTA TAAGACAACC CATGACTCA TATGTCATGT TGCGAAAAGA GGAGAGCTTA TAAGACAACC CATGACTCA TATGTCATAT TGCGAAAACAC GGAGAGCTTA TAAGACAACC CATGACTCA AATTTGCTCA TAAGAACACAC GGAGAGCTTA TAAGACAACC CATGACTCA AATTTGCTCA TAAGAACACAC GGAGCACCTTA TAAGACACACC CATGACTCA AAAAAATTCA CATGCCTACT GTGGCCCATA TATCATGCAA CAATCCAAA AAAAAATTCA CATGCCTACT GTGGCCCATA TATCATGCAA CAATCCAAA CAAAAATTCA CTTTGCCTTGC GTGTCAAAAA CAAAAATTCA CTTTGCCTTGC GTGTCAAAAA CAAAAATTCA CTTTGCCTTGC GTGTCAAAAA CAAAAATTCA CTTTGCCTTGC GTGTCAAAAA AAAAAATTCA CTTTGCCTTGC GTGTCAAAAA AAAAAATTCA CTTTGCTTGC GTGTCAAAAA AAAAATTCA TTTGCTTTC GTGTCAAAAA AAAAATTCA TTTGCTTTC GTGTCAAAAA AAAAATTCA TTTGCTTTC GTGTCAAAAA AAAAATTCA TTTGCTTTC GTGTCAAAAA AAAAATTCA TTTTGTTTG TGTGTCAAAAA AAAAATTCA TTTGCTTTC GTGTCAAAAA ACTACACGCT CTCGGTGTTG ATCGTGTCAA CATGTGACCA CC	A A A A A A A A A A A A A A A A A A A
#010/120 #159120 #159120 #159206 #1192066 #1192060 6t3 #0107120 #159120 #1592060 6t3 #0107120 #1592060 6t3 #0107120 #1592060 6t5 #0100806 6t5 #1592060 6t5	TAMACTIMA GTAATGAAAG AAGA. TEGT GTAGAAAT GAACAATAT -661 Inv TAAACT.AAA GTAATGAAAG AAGA. TEGT GTAGAAAAT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TEGT GTAGAAAAT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TEGT GTAGAAAAG AAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TEGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGAITETEGT GTAGAAAAG AAACAATAT TATGATTAT GTGTGAGCAT TATGGGACCA GAAATAAAA AAAGAACATT TATGAGTAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTATA GTGTGACCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTATA GTGTGACCAT TATGGGACCA CAAATAAAA AAAGAACATT TATGAGTATA GTGTGACCAT TATGGGACCA CAAATAAAA AAAGAACATT TATGAGCA GTGTGTCC AATGACCCTC AAAGAACATTA CACCCCCACA TTTATGACCA GTGTGTCC AATGACCCTC AAAGAACATTC CACCCCACAA TTTATGACCA GTGTTTCCC AATGACCCTC AAAGATTCC CACCCACCAA TTTATGACC	werted rept	CAA GATAMAGCAA ANGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA ANGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA ANTGATGT. G TACATAMAC TGCAGAGCI CAA GATAMAGCAA ANTGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA ANGGATGT. G TACATAMAC TACAGAGCI TAATATOGT GATCATCAA AMGATGT. G TACATAMAC TACAGAGCI TATATATOGT GATCATCAA AMGATGT. TACATAMAC TACAGAGCI TATAGTCATGT TGCAMAGCA GGAGAGCTTA TAGAGACAGC CATGACTCA TATGTCATGT TGCAMAAGA GCAGAGCTTA TAGAGACAGC CATGACTCA TATGTCATAT TGCAMAAGA GCAGAGCTTA TAGAGACAGC CATGACTCA TATGTCATAT TGCAMAAGA GCAGAGCTTA TAGAGCAAGC CATGACTCA TATGTCATAT TGCAMAAGAC CAGAGCCTTA TAGAGCAAGC CATGACTCA CATGACTATA TGCAMAAGACA CAGAGCCTTA TAGAGCAACA CATGACTCA CATGACTATA TGCAGACCAC CAGAGCACACACACACACACACACACACAC	7 top rept 3 7 top
#0107-120 #159-120 #1	TAMACTAMA GTAATGAAAG AAGA. TOST GTAGAAAAT GAACAATAT TAAACT.AAA GTAATGAAAG AAGAISTIGGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGAISTIGGT GAACAAGAAAA AAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT ATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT QATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT QATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT QATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT TATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT CATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT TATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATAAAA AAGAACATT TATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATTAAAA AAGAACATT TATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATTAAAA AAGAACATT TATGAGTAAT GTGTGAGCAT TATGGGACCA CGAATTAAAA AAGAACATT TATGAGTAAT GTGTGAGCAT AATGAGCCTT GAATGTTATC TACCCCAGGA TTTATGAGCA GTGTGTTCTC AATGAGCCTT GAATGTTATC TACCCCAGGA TTTATGAGCA GTGTGTTCTC AATGAGCCTT GAATGTTATC TACCCCAGGA TTGATGAGCA GTGTGTTCTC AATGAGCCTT GAATGTTATC TACCCCAGGA TTGATGAGCA GTGTGTTCTC GATGAGCCTT GAATGTTATC TACCCCAGGA TTGATGAGCA GTGTGTTCTC GATGAGCCTT GAATGTTATC TACCCCAGGA TTGATGAGCA GTGTGTCTC AATGAGCCTT GAATGTTATC TACCCCAGGA TTGATGAGCA GTGTTTCTC GATGACCCTT GAATGTTATC TACCCCAGGA TTGATGAGCA GTGTTTCTC GATGACCCTT GAATGTTATC TACCCCAGGA TTGATGAGCAGT GATGACCCTT GAATGTTATC TACCCCAGGA TTGATGAGCAGT GATGACCCTT GAATGTTATC TACCCCAGGA TTGATGAGCAGT GATGAC	werted rept	CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCI CAA GATAAAGCAA AAGGATGT.G TACATAAAAC TACAGAGCI CAA GATAAAGCAA AATGATGT.G TACATAAAAC TGCAGAGCI CAA GATAAAGCAA AATGATGT.G TACATAAAAC TACAGAGCI CAA GATAAAGCAA AATGATGT.G TACATAAAAC TACAGAGCI TAATATOGT GATCATCAA AAAGATATGCATAAAACA TACAGAGCI TTAATATOGT GATCATCAA AAAGATATGCATAAAAC TACAGAGCI TAATGTCATGT TGCGAAAAGA GAGAGCTTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAAGA GAGAGCITTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAAGAA GAGAGCITTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAACAA GAGAGCITTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAAACAA GAGAGCITTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAACACA GAGAGCITTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAACACA GAGAGCITTA TAAGACAAGC CATGACTCA TATGTCATAT TGCGAAACACA GAGAGCITTA TAAGACAAGC CATGACTCA AATTTGCTCA TAAGAACACA GAGAGCITTA TAAGACAAGC CATGACTCA AATTTGCTCA TAAGAACACA GAGAGCITTA TAAGACAAGC CATGACTCA AAAAAATTCA AATGCCTACT GTGGCCCATA TATCATGCAA CAATCCCAAC AAAAAATTCA CATGCCTACT GTGGCCCATA TATCATGCAA CAATCCCAAC AAAAAATTCA CATGCCTTCT GTGGCCCATA TATCATGCAA CAATCCAAC CAAAAATTCA TTTGCCTTTC GTGTCCAAAA AAAAAATCA TTTGCCTTTC GTGTCAAAAA AAAAAACCA TTTGCTTTG GTGTCAAAAA AAAAAACCA TTTGCTTTG GTGTCAAAAA AAAAAACCA TTTGCTTTG GTGTCAAAAAAAAAAAAAAA	A A A A A A A A A A A A A A A A A A A
#010/120 #159120 #159120 #159206 #1192066 #1192060 6t3 #0107120 #159120 #1592060 6t3 #0107120 #1592060 6t3 #0107120 #1592060 6t5 #0100806 6t5 #1592060 6t5	TAMACTIMA GTAATGAAAG AAGA. TEGT GTAGAAAT GAACAATAT -661 Inv TAAACT.AAA GTAATGAAAG AAGA. TEGT GTAGAAAAT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TEGT GTAGAAAAT GAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TEGT GTAGAAAAG AAACAATAT TAAACT.AAA GTAATGAAAG AAGA. TEGT GTAGAAAAG GAACAATAT TAAACT.AAA GTAATGAAAG AAGAITETEGT GTAGAAAAG AAACAATAT TATGATTAT GTGTGAGCAT TATGGGACCA GAAATAAAA AAAGAACATT TATGAGTAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTAT GTGTGAGCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTATA GTGTGACCAT TATGGGACCA CAAATAAAA AAAGAACATT PATGAGTATA GTGTGACCAT TATGGGACCA CAAATAAAA AAAGAACATT TATGAGTATA GTGTGACCAT TATGGGACCA CAAATAAAA AAAGAACATT TATGAGCA GTGTGTCC AATGACCCTC AAAGAACATTA CACCCCCACA TTTATGACCA GTGTGTCC AATGACCCTC AAAGAACATTC CACCCCACAA TTTATGACCA GTGTTTCCC AATGACCCTC AAAGATTCC CACCCACCAA TTTATGACC	werted rept #1519120 hancer core #1519120 #1564120 #1564120 #154026 #1134026 #1519020 fit fit fit fit fit #1519020 #1564120	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TECAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI TATATATGT TACAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI TTAATATGT TACAMAGCAA AAGGATGT. G TACAMAAC TACAGAGCI TATATGTATT TACAMAGCAA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATGT TACAMAGCAA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAGCAA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGA GAGAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAMAAGAA GAGAGCTTA TAAGACAAGC CATGACTC AATTTCCTCA TAAGAAACCA CAGAGCCTTA TAAGACAAGC CATGACTC AATTTCCTCA TAAGAAACCA CAGAGCCTTA TAAGACAACC CATGACCTC AATTTCCTCA TACGACAAC CAGAGCCTA TAAGACACAC CATGACCTC AATTTCCTCA TACGACAAC CAGAGCCTA TAAGACACAA CATGCCTAC AAAAAATTCA CATGCCTACT GTGCCCCATA TATCATCCAA CAATCCAAA AAAAAATTCA CATGCCTTACT GTGCCCCATA TATCATGCAA CAATCCAAA AAAAAATTCA CATGCCTTACT GTGCCCCATA TATCATGCAA CAATCCAAA AAAAAATTCA CATGCCTTACT GTGCCCCATA TATCATGCAA CAATCCAAA AAAAAATCA TTTGCCTTTC GTGTCCAAAAA AAAAAATCA TTTGCCTTTC GTGTCCAAAA AAAAAATCA TTTGCCTTTC GTGTCCAAAAA AAAAAATCA TTTGCCTTTC GTGTCCAAAAA AAAAATCA T	A A A A A A A A A A A A A A A A A A A
#010/12/20 #159-120 #159-120 #159-120 #159-120 #159-120 #1519-000 6t3 #0107-120 #1519-	TAMACTAMA GTAATGAMA AGA. TOST GTAGAAAT GMACAMTAT -661 INT TAMACT.AAA GTAATGAMA AGA. TOST GTAGAAAT GMACAMTAT TAMACT.AAA GTAATGAMA AGA. TOST GTAGAAAT GMACAMTAT TAMACT.AAA GTAATGAMA AGA. TOST GTAGAAAAT GMACAMTAT TAMACT.AAA GTAATGAMA AGA. TOST GTAGAAAAT GMACAMTAT TAMACT.AAA GTAATGAMA AGARJETIGST GTAGAAAAG GMACAMTAT TAMACT.AAA GTAATGAMAG AGARJETIGST GTAGAAAAG GMACAMTAT TAMACT.AAA GTAATGAMAG AGARJETIGST GTAGAAAAA AAAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAATAAAAA AAAGAACATT ATTGACTAAT GTGTGAGCAT TATGGGACCA CGAATAAAAA AAAGAACATT ATTGACTAAT GTGTGACCAT TATGGGACCA CGAATAAAAA AAAGAACATT ATGAGTAAT GTGTGACCAT TATGGGACCA CGAATTAAAA AAAGAACATT ATGAGTAAT GTGTGACCAT CAATGACCAT CAACCACACAA TTTATGACCA GTGTGTTCC AATGACCCTT GAATGTTATC TACCACACGAA TTTATGACCA GTGTGTTCC AATGACCCTT GAATGTTATC TACCACACGAA TTGATGAGCA GTGTGTTCC AATGACCCTT GAATGTTATC TACCACACGAA TTGATGAGCA GTGTGTTCC AATGACCCTT GAATGTTATC TACCACACGAA TTGATGAGCA GTGTGTCC GATGACCCTT GAATGTTATC TACCACACGAA TTGATGAGCACCT TACCACCTT GAATGTTATC TACCACACGAA TTGATGAGACCC TTACCCCT GATGACCCTT GAATGTTATC TACCACAGTTG TACAAAACCC TTACCCCTT AAACCTTC TACCACACTTT TACCACACTTT TAAGAAACCC TTACCCC	werted rept	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TECAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI TATATATOGT GATCATCAA AAGGATGT. G TACATAMAC TACAGAGCI TTAATATOGT GATCATCAA AAGGATGT. G TACATAMAC TACAGAGCI TTATGTCATGT TACAGAAAGA GAGAGCTTA TAAGAAAAC CATGACTCA TATGTCATGT TACAGAAAGA GAGAGCTTA TAAGAAAAC CATGACTCA TATGTCATGT TACAGAAAGA GAGAGCTTA TAAGACAAG CATGACTCA TATGTCATAT TACAAAAGA GAGAGCTTA TAAGACAAG CATGACTCA TATGTCATAT TACAGAAAGA GAGAGCTTA TAAGACAAG CATGACTCA AATTTGCTCA TAAGAAACAC CAAGAGCTATA TAAGACAAGC CATGACTCA AATTTGCTCA TAAGACAACA CAAGAGCTTA TAAGACAAG CATGACTCA AAAAAATTCA TACACACAA CAAGACATA TATCATACAA CAATCCAAA AAAAAATTCA CATTCCTTACT GTGCCCCATA TATCATACAA CAATCCAAA AAAAAATTCA TITTCCCTTTC GTGCCCCATA TATCATACAA CAATCCAAA AAAAAATTCA TITTCCCTTTC GTGCCCCATA TATCATCAAA CAATCCAAA AAAAAATTCA TITTCCCTTTC GTGCCCCATA TATCATCAAA AAAAAATTCA TITTCCCTTTC GTGCCCCATA TATCATGCAA TAATCCAAA AAAAAATTCA TITTCCCTTTC GTGCCCCATA TATCATGCAA TAACCAAAA AAAAAATTCA TITTCCCTTTC GTGCCCCATA TATCATGCAA TAATCCAAAA AAAAAATTCA TITTCCCTTTC GTGCCCCATA TATCATGCAA TAATCCAAAA AAAAAATTCA TITTCCCTTTC GTGCCCCATA TATCATGCAA TAATCCAAAA AAAAAATTCA TITTCCCTTTC GTGTCCAAAA AAAAAATTCA TITTCCCTTTC GTGTCCAAAA AAAAAATTCA TITTCCCTTTC GTGTCCAAAA AAAAAATCA TATCCTTTC GTGTCAAAA AAAAAATCA	A A A A A A A A A A A A A A A A A A A
#0107120 #1599120 #1599120 #159206 613 #01005595 613 #0107120 #1519206 #1519206 #1519208	TAMACTIMA GTANTGAMA AGA. TOST STEAGAAAT GAACCATAT -661 INT TAMACTIMA GTANTGAMA AGA. TOST GTI-GGAAAT GAACCATAT TAMACTIMA GTANTGAMA AGA. TOST GTI-GGAAAT GAACCATAT TAMACTIMA GTANTGAMA AGA. TOST GTI-GGAAAT GAACCATAT TAMACTIMA GTANTGAMA AGA. TOST GTI-GGAAAAG GAACCATAT TAMACTIMA GTANTGAMA AGAGITETEST GTI-GGAAAAG GAACCATAT TAMACTIMA GTANTGAMA AGAGITETEST GTI-GGAAAAG GAACCATAT TAMACTIMA GTANTGAMA AGAGITETEST GTI-GGAAAAG GAACCATAT TAMACTIMA GTANTGAMA AGAGITEST GTI-GGAAAAG GAACCATAT TAMACTIMA GTANTGAMA AGAGITEST GTI-GGAAAAG GAACCATAT TAMACTIMA GTANTGAMA AGAGITEST GTI-GGAAAAG GAACCATAT TAMACTIMA GTICHGCOT TATGGGACCO CGAATTAAAA AAAGAACAT TATGACTAT GTICHGCAT TATGGGACCO CGAATTAAAA AAAGAACAT CATGGGTAAT GTICHGCAT TATGGGACCO CGAATTAAAA AAAGAACAT CATGGGTAAT GTICHGCAT TATGGGACCO CGAATTAAAA AAAGAACAT CATGGGTAAT GTICHGCAT TATGGGACCO CGAATTAAA AAAGAACAT CATGGGTAAT GTICHGCAT TATGGGACCO CGAATTAAAA AAAGAACAT CATGGGTAAT GTICHGCAT TATGGGACCO CGAATTAAAA AAAGAACAT CATGGGTAAT GTICHGCAT TATGGGACCO CGAATTAAAA AAAGAACAT CATGGGTAAT GTICHGACCAT TATGGGACCO CGAATTAAAA AAAGAACAT TATGACTAAT GTICHGACCAT TATGGGACCO CGAATTAAAA AAAGAACAT CATGGGTAAT GTICHGACCAT TATGGGACCO CGAATTAAAA AAAGAACAT TATGACTAAT GTICHGACCAT TATGGGACCO CGAATTAAAA AAAGAACAT TATGACCAG GTIGTTCTC AATGACCCT GAATGTTATC TACCCCAGGA TIGATGACCG GTITATCTC GATGACCCT GAATGTTATC TACCCCAGGA TIGATGACGT GTITATCTC GATGACCCT GAATGTTATC TACCCCAGGA TIGATGACGT GTITATCT GATGACCCT GAATGTTATC TACCCCAGGA TIGATGACGC GTITATCTC GATGACCCT GAATGTTATC TACCCCAGGA TIGATGACGC GTITATCTC GATGACCCT GAATGTTATC TACCCCAGGA TIGATGACGC GTITATCTC GATGACCCT GAATGTTATC TACCCCAGGA TIGATGACCC GTITATCT GATGACCTC GAATGTTATC TACCCCAGGA TIGATGACCC GTITATCT GATGACCTC GAATGTTATC TACCCCAGGA TIGATGACCC GTITATCT	werted rept	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TECAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI TATATATOGT GATCATCAA AAGGATGT. G TACATAMAC TACAGAGCI TATATATOGT GATCATCAA AAGGATGT. TACAGAAAAC CATGACTC TATGTCATGT TACAGAMAGCA GAGCAGCTTA TAAGAAAAC CATGACTC TATGTCATGT TACGAAMAGCA GAGCAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAAMAGCA GAGCAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAGAMAGCA GAGCAGCTTA TAAGACAAGC CATGACTC. AATTTGCTCA TACAGACAAC CAGCAGCTTA TAAGACAAGC CATGACTC. AATTTGCTCA TACAGACAC CAGCAGCTTA TAAGACAACC CATGACTC. AATTTGCTCA TACAGACAC CAGCAGCTTA TAAGACACAC CATGACTC. AATTTGCTCA TACAGACAC CAGCAGCTAT TATACTACAA CATGCCTACT AAAAAATTCA ATTGCCTTACT GTGCCCCATA TATACTACAAC CATGCCCACACACACACACACACACACACACACACACACA	A A A A A A A A A A A A A A B A A B A B
#010/12/20 #159-120 #159-120 #159-120 #159-120 #159-120 #1519-000 6t3 #0107-120 #1519-	TAMACTAMA GTAATGAMA AGA. TOST GTAGAAAT GMACAMTAT -661 INT TAMACT.AAA GTAATGAMA AGA. TOST GTAGAAAT GMACAMTAT TAMACT.AAA GTAATGAMA AGA. TOST GTAGAAAT GMACAMTAT TAMACT.AAA GTAATGAMA AGA. TOST GTAGAAAAT GMACAMTAT TAMACT.AAA GTAATGAMA AGA. TOST GTAGAAAAT GMACAMTAT TAMACT.AAA GTAATGAMA AGARJETIGST GTAGAAAAG GMACAMTAT TAMACT.AAA GTAATGAMAG AGARJETIGST GTAGAAAAG GMACAMTAT TAMACT.AAA GTAATGAMAG AGARJETIGST GTAGAAAAA AAAGAACATT TATGACTAAT GTGTGAGCAT TATGGGACCA CGAATAAAAA AAAGAACATT ATTGACTAAT GTGTGAGCAT TATGGGACCA CGAATAAAAA AAAGAACATT ATTGACTAAT GTGTGACCAT TATGGGACCA CGAATAAAAA AAAGAACATT ATGAGTAAT GTGTGACCAT TATGGGACCA CGAATTAAAA AAAGAACATT ATGAGTAAT GTGTGACCAT CAATGACCAT CAACCACACAA TTTATGACCA GTGTGTTCC AATGACCCTT GAATGTTATC TACCACACGAA TTTATGACCA GTGTGTTCC AATGACCCTT GAATGTTATC TACCACACGAA TTGATGAGCA GTGTGTTCC AATGACCCTT GAATGTTATC TACCACACGAA TTGATGAGCA GTGTGTTCC AATGACCCTT GAATGTTATC TACCACACGAA TTGATGAGCA GTGTGTCC GATGACCCTT GAATGTTATC TACCACACGAA TTGATGAGCACCT TACCACCTT GAATGTTATC TACCACACGAA TTGATGAGACCC TTACCCCT GATGACCCTT GAATGTTATC TACCACAGTTG TACAAAACCC TTACCCCTT AAACCTTC TACCACACTTT TACCACACTTT TAAGAAACCC TTACCCC	werted rept	CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TECAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AATGATGT. G TACATAMAC TACAGAGCI CAA GATAMAGCAA AAGGATGT. G TACATAMAC TACAGAGCI TATATATOGT GATCATCAA AAGGATGT. G TACATAMAC TACAGAGCI TATATATOGT GATCATCAA AAGGATGT. TACAGAAAAC CATGACTC TATGTCATGT TACAGAMAGCA GAGCAGCTTA TAAGAAAAC CATGACTC TATGTCATGT TACGAAMAGCA GAGCAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAAMAGCA GAGCAGCTTA TAAGACAAGC CATGACTC TATGTCATAT TACAGAMAGCA GAGCAGCTTA TAAGACAAGC CATGACTC. AATTTGCTCA TACAGACAAC CAGCAGCTTA TAAGACAAGC CATGACTC. AATTTGCTCA TACAGACAC CAGCAGCTTA TAAGACAACC CATGACTC. AATTTGCTCA TACAGACAC CAGCAGCTTA TAAGACACAC CATGACTC. AATTTGCTCA TACAGACAC CAGCAGCTAT TATACTACAA CATGCCTACT AAAAAATTCA ATTGCCTTACT GTGCCCCATA TATACTACAAC CATGCCCACACACACACACACACACACACACACACACACA	A A A A A A A A A A A A A A B A A B A B

bot371-07.p65 43 2001/4/19, PM 01:48

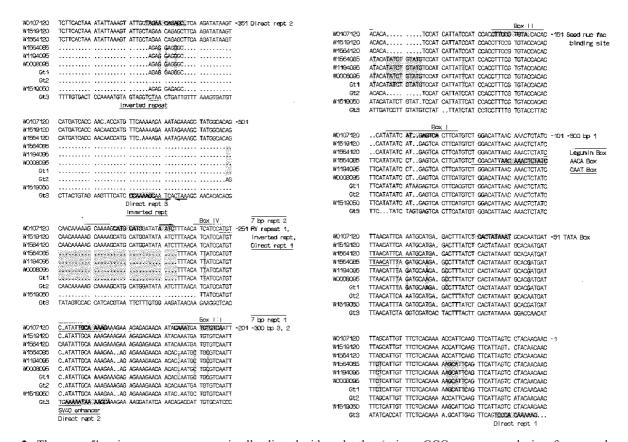


Figure 2. The seven 5' region sequences are maximally aligned with each other (using a GCG sequence analysis software package) including the three 5' regions of Gt genes (Okita et al., 1989) and arranged in descending order starting from the most 5' end. The dot within the sequence denotes base deletion. The shaded, upper, or underlined bases stand for the repeat, element, box, etc., the name of which is listed to the right side. A pair of arrows indicates inverted repeats. The vertical bar represents base substitution or deletion in the 0.9 kb and *Gt1* 5' region with respect to that of the 1.2 kb and *Gt2*. The *Gt3* 5' region has its own repeats, etc. with its name indicated below the sequence.

found between the Gt1 and Gt2 sequences. The substituted or deleted bases in the 5' region sequence of Gt1 are like those in the sequence of the 0.9 kb 5' region (marked with a vertical bar in Figure 2). The same can be found between the sequences of Gt2 and the 1.2 kb region. For example, at the -976 bp position the 0.9 kb and Gt1 sequences have the same T and the 1.2 kb and Gt2 sequence have a base deleted in common. Furthermore, many boxes, repeats, and elements of the 0.9 kb and Gt1 sequence and of the 1.2 kb and Gt2 sequence have similar correspondence (Figure 2, Table 2). As a result, the 5' region of Gt1 and Gt2 has a 24.39% and 21.64% higher homology respectively to that of the 0.9 kb and the 1.2 kb than the homology between the 0.9 kb and the 1.2 kb 5' region. Due to the two long deletions in the 0.5 kb 5' region, it has 66 instead of 209 base-substitutions or -deletions. Among them, 36 bases are the same as the correspondent bases in the 0.9 kb 5' region; 9 bases follow that in the 1.2 kb 5' region, and 21 bases follow that in neither of the two 5' regions (Figure 2).

In view of the long direct repeats, four have been identified in *O. sativa* cultivar Mangetsumochi (Takaiwa et al., 1987b) and three in *O. sativa* cultivar M201 (Okita et al., 1989). Figure 2 shows the sequences and positions of the four direct repeats. All four are conserved in the sequences

examined except that the 0.5 kb sequence lacks direct repeat 4 and *Gt2* lacks direct repeat 2. *Gt3* sequence has its own four direct repeats, and their positions do not coincide with those identified in the other sequences. Its first direct repeat is located from -15 to -4 bp relative to ATG.

The SV40 enhancer TGAAAAA, identified in the Gt3 sequence (Okita et al., 1989), is shown to be superimposed with its own direct repeat 2, located from -248 bp to -236 bp, but can not be found in the other 5' region sequences. However, one kind of 7 bp direct repeat (T/AGCA/ GAAA/G) with high homology to the SV40 enhancer can be found in the 5' region sequences examined. There are six such short direct repeats in the 1.2 kb sequence (Figure 2), two of them (repeats 2 and 4) are independent. The TGCA motifs of the remaining four are superimposed upon those of their respective -300 bp elements and/or long direct repeats. The 0.5 kb sequence lacks the short direct repeats 6, 5, and 2; the 0.9 kb sequence and Gt1 lack short direct repeat 2. Gt3 also has six short direct repeats but with low or very low homology to that of the 7 bp direct repeat located at the corresponding positions in the 5' region of the various wild rice species.

A -300 bp element was reported to be present in the 5' region of barley prolamin gene (Forde et al., 1985). There are six -300 bp elements that can be traced in the 5' re-

bot371-07.p65 44 2001/4/19, PM 01:48

Table 2. Base substitution, deletion and homology in the boxes, repeats and elements of glutelin gene flanking 5' region.

		AA CC BBCC B (1.2 kb)	BCC CCDD EE (0.9 kb)	Gt1	Gt2	CC (0.5 kb)	Gt3
Direct repeat	4 3	•GACATAATGCAAAAG ATGTCATGTTGCGAAA	T, T A	H 0.9 kb ^b H 0.9 kb	H 1.2 kb H 1.2 kb	Del H 0.9 kb	VLH ^a VLH
	2	AGAGGAGAG TAGAACAGAGC ATGTCATATTGCAAAAGAA AG AAAG	G, G	H 0.9 kb H 1.2 kb	Del H 1.2 kb	H 0.9 kb H 0.9 kb	VLH VLH
Box	VI V	GTCA TAAGTCA TAAC TGATGA	C CGTT	H 0.9 kb H 0.9 kb	H 1.2 kb H 1.2 kb	Del Del	VLH VLH
	IV	ATCATCCATGTCATATTG ACAAATGATGTGTCAATTA CTTCCGTGTACCACA ATATCAT••GAGTCACTTCA	T •, C, C Conserved Conserved	H 0.9 kb H 0.9 kb Conserved	H 1.2 kb H 1.2 kb Conserved Conserved	H 1.2 kb H 1.2 kb Conserved Conserved	VLH VLH HH HH
-300 bp element	6 5 4 3 2	TGCAAAGTT TGCAAAAG TGCGAAAG TGCAAAAG CAAA TGTGTCA AT••GAGTCA	Conserved T A Conserved •, C C	Conserved H 0.9 kb H 0.9 kb Conserved H 0.9 kb	Conserved H 1.2 kb H 1.2 kb Conserved H 1.2 kb H 1.2 kb	Del Del H 0.9 kb conserved H 0.9 kb H 1.2 kb	LH NH LH LH LH LH
RY repeat	7 6 5 4 3 2	TATG CATGCAAG CATGCATC CATG CATG CATG CA	C Del Conserved Conserved Conserved Del Del	H 0.9 kb Del Conserved Conserved Conserved Del Del	H 1.2 kb H 1.2 kb Conserved Conserved H 1.2 kb H 1.2 kb	Del Del Del Del Del Del	NF H H H H H
Enhancer core 13 bp AACA box Legumin box	(•TGGTGTT AACAAACTCTATC CTTAACATTCAAATGATG	G Conserved T, G, A	H 0.9 kb Conserved H 0.9 kb	H 1.2 kb Conserved H 1.2 kb	Del Conserved H 0.9 kb	Del H 0.9 kb
Immuture seed nuc fac BS		CTTCCGTGTA	Conserved	Conserved	Conserved	Conserved	НН
CAAT Box		TGGACATTAACAAACTCTATCTTAACA	Conserved	Conserved	Conserved	Conserved	
Inverted repeat		AATC (-271) AATC (-459) AATC (-766) CTAA(-408) CTAA(-684) CTAA(-996)	Del Del C Del Conserved	Del Del H 0.9 kb Del H 1.2 kb H 1.2 kb	H 1.2 kb H 1.2 kb H 1.2 kb H 1.2 kb H 1.2 kb	Del Del Del Del Del Del	AATC (-324) NF NF CTAA (-317) CTAA (-374) CTAA (-504)
SV40 enhancer		Not found	Not found	Not found	Not found		TGAAAAA
7 bp direct repeat	5 4 3	TGCAAAG TGCAAAA AGCAAAA TGCGAAA	Conserved Conserved Conserved A	Conserved Conserved Conserved H 0.9 kb	Conserved Conserved Conserved H 1.2 kb	Del Del Conserved H 0.9 kb	LH VLH LH LH
	2	AGCAAAA TGCAAAA	Del Conserved	Del Conserved	H 1.2 kb Conserved	Del Conserved	LH LH

*NH stands for nonhomology; LH, low homology; VLH, very low homology; HH, high homology of *Gt3* 5' region sequence to the corresponding sequence of 1.2 kb 5' region of AA, CC and BBCC genomes. H, homology; NF, not found; Del, deleted. The *Gt3* 5' region has its own direct repeat 4, CCAAAACCAAAAGCA; direct repeat 2, AAAAATAAAAGCA and direct repeat 1, TCCCACAAAAAC and six RY repeats, CATG and four inverted repeats. Positions of these repeats do not coincide with those found in other 5' region sequences. Boldfaced bases in the lane of 1.2 kb 5' region are shown to have been respectively substituted by the base or deleted shown in the lane of 0.9 kb, for example, --T, G-T., AG-.. etc.

bot371-07.p65 45 2001/4/19, PM 01:48

 $^{^{\}mathrm{b}}$ Sequence homology of Gt1 5' region to that of 0.9 kb 5' region.

gion of glutelin genes in the wild rice species. The sixth (located from -810 bp to -802 bp) is independent; five of them, i.e., the elements 5, 4, 3, 2, and 1 are superimposed upon part of the Takaiwa's long direct repeat 4, 3 and upon the sequence of Box IV III and I respectively, which are the bind site of nuclear proteins as reported (Kim and Wu, 1990). In addition to these, superimposition can be found between the -300 bp elements 6, 5, 4, 3, and the four 7 bp repeats. Each of the six -300 bp elements have been conserved with minor base substitutions in the sequences examined but not in the 0.5 kb 5' region, in which it lacks -300 bp elements 6 and 5. Only low homology remains in most of the elements in *Gt3* (Figure 2, Table 2).

Five protein binding sequences (boxes; Kim and Wu, 1990), have been well conserved because they can be found in all the sequences examined except in the 0.5 kb 5' region that lacks Box V. In addition to the five boxes, we found a new one, the sixth box with its core motif GTCA from -749 bp to -746 bp in the 1.2 kb 5' region, but the first nucleotide G of the motif has mutated to C in the case of 0.9 kb 5' region. In the 0.5 kb 5' region, the fifth and sixth boxes have been deleted. In *Gt*3, only boxes Iand II are conserved (Figure 2, Table 2).

In the 5' region sequences of 1.2 kb, there are seven RY repeats. The seventh is between -950 bp and -947 bp and has a nucleotide C mutated to T. From -274 bp to -267 bp, the first RY repeat CATGCATG can be traced. Between these two positions, it accommodates the other five RY repeats (repeats 6, 5, 4, 3, and 2). The 0.9 kb 5 region lacks RY repeats 1, 2, and 6. None of the RY repeats can be found in the 5' region sequence of 0.5 kb. Okita et al. (1989) reported that there were respectively one, two and one RY repeat in the 5' region of Gt1, Gt2, and Gt3. In Figure 2, however, it is shown that Gt1 and Gt2 have four and seven RY repeats respectively. Gt3 has six RY (all CATG) repeats dispersed in a segment of 310 bp, from -780 bp to -476 bp. None is located at the same positions shown in the other 5' region sequences (Figure 2 and Table 2). It is interesting to note that the number of RY repeats in the various 5' regions vary substantially among species and that some of the repeats are superimposed with the -300 bp element and the long direct repeat.

Okita et al. (1989) identified two pairs of inverted repeats (AATC and CTAA) in the 5' region of *Gt2* and suggested that the DNA segments between the two components of the two repeat pairs could have been transposed from somewhere else. Figure 2 shows that the position of the three inverted repeat pairs we found in the sequence of 1.2 kb 5' region start at -271 bp, -408 bp; -459 bp, -684 bp; and -766 bp, -996 bp, respectively. It is interesting to note that the sequence between -271 bp and -408 bp including the inverted repeat pair has been deleted in the 0.9 kb and 0.5 kb 5' region (Figure 2, Table 2). In the *Gt3* sequence, we identified one AATC starting at -322 bp but three CTAA at -317 bp, -374 bp, and -504 bp. In the case of 0.5 kb sequences, all the inverted repeats have been deleted.

Proximal to the translation initiation codon, ATG, within a range of -1 to -261 bp, the 13 bp AACA Box (AACAAACTCTATC, Takaiwa and Oono, 1991), the legumim Box (CTTAACATTTAGATGCAAG, Takaiwa et al., 1987b), the CAAT Box (TGGACATTAACAA ACTCTATCTTAACA, Okita et al., 1989) the immature seed nuclear factor binding site (CTTTCGTGTA, Takaiwa and Oono, 1990), and Boxes I, II, III and IV of Kim and Wu (1990) all reside in and are well conserved in each of the 5' region sequences examined, in addition to the TATA Box (TCACTATAAAT). The first three of the above mentioned boxes are superimposed.

Discussion

Rice glutelin genes have numerous putative enhancers dispersed though the 5' flanking region within a span of about 950 bp. Beyond the four long direct repeats, the 7 bp short direct repeat may have a putative function as enhancer because its sequences are similar to that of the SV40 enhancer that was superimposed upon the direct repeat 2 in the 5' region of Gt3 (Okita et al., 1989). In this study, the 7 bp direct repeat 3 and the long direct repeat 3 are superimposed upon each other in all the sequences examined. SV40 enhancer core homology was observed in gliadin (wheat storage protein) genes (Reeves and Okita, 1987). Chen et al. (1986) reported that 4 short direct repeats of AA/GGCCA in the 5' region of β -conglyccinin α subunit could increase expression 20 fold. It would be worthwhile to see how each of the many direct repeats enhances the expression of rice glutelin genes.

Two -300 bp elements were found in the 5' region of prolamin (B1 hordein) genes of barley. The elements are composed of the conserved core motif GTCATG and were proposed to be endosperm specific (Forde et al., 1985). In the 5' region of rice glutelin genes, two (elements 1 and 2, Table 2) of the six -300 bp elements have GTCA motif; the other four (elements 3 to 6, Table 2) are similar to the two -300 bp elements (TGTAAAGT and TGTAAAAG) that are endosperm specific in wheat prolamin (LMW glutenin) genes (Colot et al., 1987). The differential functions of the two groups of the sequences that are similar to that of barley and wheat respectively in glutelin gene expression are not known. Superimposition of the four -300 bp elements (elements 3, 4, 5, and 6) respectively upon the four 7 bp short direct repeats (repeats 1, 3, 5, and 6) and the two long direct repeats combined with the fact that the elements 4, 5, and 6 are dispersed at relatively far 5' upstream localities (extending from -530 bp to -810 bp) may suggest that the four -300 bp elements play a role as enhancer. A sequence similar to that of the SV 40 enhancer however, did not significantly increase the level of expression of soybean storage protein genes (Chen et al., 1986). The superimposition of -300 bp element 1 and 2 upon that of nuclear protein binding boxes I and IV (Kim and Wu, 1990) may imply that some nuclear protein molecules are needed to bind GTCA core motif and enable the two -300 bp elements to carry out endosperm specific expression.

bot371-07.p65 46 2001/4/19, PM 01:48

In the 5' region of 1.2 kb, we identified one more pair of inverted repeats that was not found in *Gt2* by Okita et al. (1989). The pair, located at -408 bp and -271 bp and found in the 1.2 kb 5' region (Figure 2), might be responsible for a ~200-base deletion occuring in the 0.9 kb and 0.5 kb 5' regions. The fact that the 0.5 kb and 0.9 kb 5' regions lost three and one and a half pairs of inverted repeats respectively (Table 2) and that the large blocks of deletions border on the inverted repeat may reflect the involvement of the lost inverted repeat pairs in deletion. In CC genomic species (*officinalis* and *eichingeri*) the 0.9 kb 5' region was replaced by that of the 0.5 kb (Figure 1B). It might be inferred that the 0.5 kb 5' region was derived from that of the 0.9 kb due to the occurrence of large blocks of deletion.

The RY repeat (CATGCATG) plays a role in the regulation of legume seed protein gene expression (Dickinson et al., 1988). The seven RY repeats (four of them have the CATG motif only) in the 5' region of glutelin genes are not the only case of multiple copies of the RY repeat present in the 5' region of several legume seed protein genes (Dickinson et al., 1988). Whether the 0.5 kb 5' region of glutelin genes that totally lacks all seven RY repeats—beyond one long direct repeat, one nuclear protein binding site, one enhancer core, two -300 bp elements, and three 7 bp direct repeats (Table 2)—could still perform its normal regulation of glutelin gene transcription is not known.

Legumin, CAAT, AACA, and TATA boxes in all the ten 5' region sequences of glutelin genes observed (including Gt1, Gt2, and Gt3) are located between nucleotide -1 and -150. They are well conserved and aligned. Furthermore, superimposition of the first three boxes upon each other is shown in Figure 2. Legumin box is an element that may have a function in the regulation of legumin gene expression in pea (Baunnlein et al., 1986). The function of this box in rice glutelin genes has not been determined. The CAAT box (TGTTGACAATTT) was designated as the site where the interaction between specific factor and RNA polymerase occurs (Benoist et al., 1980). In rice glutelin genes, the CCAAT sequence was shown to have no significant homology to the eukaryotic model sequence TGTTGACAATTT (Okita et al., 1989). However, the CCAAT-like sequence is important for maximal gene expression in Kalanchoe and tobacco plant (Shaw et al., 1984; Odell et al., 1985). CACA box consists mostly of C and A (GTGCCACCAAACACACATACCAAAA) and was observed in the 5' region of wheat gliadin genes (Reeves and Okita, 1987) though its function was not known. The 13 bp AACA box (Takaiwa and Oono, 1991) is also CA rich and well conserved. Superimposition of the AACA box upon the legumin box and the CAAT box, located proximal to the TATA box in the 5' region of glutelin genes, may mean that the AACA box has an essential function in the expression of glutelin genes.

Study on expression of glutelin genes has been done by comparing the amount of mRNA at developmental stages of rice endosperm (Okita et al., 1989), and by detecting CAT enzyme (Leisy et al., 1989) or GUS activities (Zhao et al., 1994) regulated by a glutelin gene 5' region in transgenic tobacco. It has also been carried out by transient expression assay using immature rice seed protein (Kim and Wu, 1989). Our sequence analysis showed that the 5' region glutelin genes are different in nature from each other. It would be worthwhile first to compare the expression capacity of the three naturally existing 5' regions. Then an artificial 5' region containing well-designed sets of regulating motifs in the glutelin gene 5' region should be constructed and tested to define the expression capacity of the motifs. A thorough understanding of this may be essential to the genetic engineering of rice glutelin genes sooner or later.

The results of our analysis also help to clarify classification of rice glutelin genes. Okita et al. (1989) postulated three gene subfamilies: Gt1, Gt2, and Gt3 for glutelin. Takaiwa et al. (1991) classified the glutelin genes so far that they were isolated and sequenced into two subfamilies, A and B. From our PCR experiments (Figure 1A, B), it can be inferred that there are at least three distinct sequence lengths for the 5' regions of glutelin genes. Southern blot analysis, using specific segments as probes, showed that these 5' region sequences can each be accommodated at specific locations in a genome (data not shown) and may represent three glutelin gene loci. Beyond the length differences, corresponding substitution and deletion of the 209 bases in the 0.9 kb and Gt1 with respect to the 1.2 kb enabled us to find a closer relationship between the 1.2 kb and Gt2 and between the 0.9 kb and Gt1 sequences. Total homology at all these 209 bases between the 0.9 kb and Gt1, and between the 1.2 kb and Gt2 support our suggestion that the 5' region of the 1.2 kb and Gt2 can be assigned as one locus and that of the 0.9 kb and Gt1 as another locus in various rice genomes. The same is true of the 5' region of the 0.5 kb in wild rice species. When the coding region sequences of glutelin genes with the known $0.5~\mathrm{kb},\,0.9~\mathrm{kb},\,\mathrm{or}\,1.2~\mathrm{kb}\,5'$ region were compared , it revealed that those coding sequences were highly homologous (90-95%) to each other. This suggests that all these glutelin genes can be grouped into one subfamily, i.e., the Glua subfamily.

The length of the three 5' regions of 1.2 kb examined in our analysis varies from 1,111 to 1,119 bp among three species, i.e., *O. perrennis, O. eichingeri*, and *O. punctata* (Table 1). This is due to minor deletion or addition of bases varying in number and position occuring in a unique 5' region. For example, in the 1.2 kb 5' region one-base deletion can be examined at position -364 bp in *O. perrennis*, two-base deletion at -529 bp in *O. eichingeri* but four-base deletion at position -791 in *O. punctata*, etc. (Figure 2). In the case of the 0.9 kb 5' region, a similar situation can be observed. We suggest designating these sequences distributed in different species with minor base deletion, addition, or substitution as alleles of a glutelin locus.

Literature Cited

Baumlein, H., U. Wobus, J. Pustell, and F.C. Kafatos. 1986. The legumin gene family: structure of a B type gene of *Vicia*

bot371-07.p65 47 2001/4/19, PM 01:48

- *faba* and a possible legumin gene specific regulatory element. Nucl. Acid Res. **14:** 2707–2719.
- Benoist, C., K. O'Hare, R. Breatnach, and P. Chambon. 1980. The ovalbumin gene sequence of putative control regions. Nucl. Acids Res. 8: 127–142.
- Chen, Z.L., M.A. Schuler, and R.N. Beachy. 1986. Functional analysis of regulatory elements in a plant embryo-specific gene. Proc. Natl. Acad. Sci. USA 83: 8560–8564.
- Colot, V., L.S. Robert, T.A. Kavanagu, M.W. Bevan, and R.D. Thompson. 1987. Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO 6: 3559–3564.
- Dickinson, C.D., R.P. Evens, and N.C. Nielsen. 1988. RY repeats are conserved in the 5'-flanking regions of legume seed-protein genes. Nuc. Acid. Res. 16: 371.
- Forde, B.G., A. Heyworth, J. Pywell, and M. Kreis. 1985. Nucleotide sequence of a Bi Hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. Nucl. Acids Res. 13: 7327–7337.
- Kim, S.Y. and R. Wu. 1990. Multiple protein factors bind to a rice glutelin promoter region. Nucl. Acids Res. 18: 6845-6852.
- Leisy, D.J., J. Hnilo, Y. Zhao, and T.W. Okita. 1989. Expression of a rice glutelin promoter in transgeneic tobacco. Plant Mol. Biol. 14: 41–50.
- Odell, J.T., F. Nagy, and N.H. Chua. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature **313**: 810–812.
- Okita, T.W., Y.S. Hwang, J. Hnilo, W.T. Kim, A.P. Aryan, R. Larson, and H.B. Krishnan. 1989. Structure and expression

- of the rice glutelin multigene family. J. Biol. Chem. **264**: 12573–12581.
- Reeves, C.D. and T.W. Okita. 1987. Analyses of α/β type gliadin genes from diploid and hexaploid wheats. Gene **52**: 257–266.
- Shaw, C.H., C.H. Cater, M.D. Watson, and C.H. Shaw. 1984. A functional map of the nopaline synthase promoter. Nucl. Acid. Res. 12: 7831–7846.
- Takaiwa, F., S. Kikuchi, and K. Oono. 1986. The structure of rice storage protein glutelin precursor deduced from cDNA. FEBS Lett. 206: 33–35.
- Takaiwa, F., S. Kikuchi, and K. Oono. 1987a. A rice glutelin gene family- a major type of glutelin mRNAs can be divided into two classes. Mol. Gen. Genet. **208**: 15–22.
- Takaiwa, F., H. Ebinuma, S. Kikuchi, and K. Oono. 1987b. Nucleotide sequence of a rice glutelin gene. FEBS Lett. 221: 43–47.
- Takaiwa, F. and K. Oono. 1990. Interaction of an immuture seeds specific trans-acting factor with the 5' upstream region of a rice glutelin gene. Mol. Gen. Genet. **224**: 289–293.
- Takaiwa, F. and K. Oono. 1991. Genomic DNA sequence of two new genes for new storage protein glutelin in rice. Jap. J. Genet. **66:** 161–171.
- Takaiwa, F., K. Oono, D. Wing, and A. Kato. 1991. Sequence of three members and expression of a new major subfamily of glutelin genes from rice. Plant Mol. Biol. 17: 875–885.
- Zhao, Y., D.J. Leisy, and T.W. Okita. 1994. Tissue-specific expression and temporal regulation of the rice glutelin *Gt3* gene are conferred by at least two spatially separated *cis*-regulatory elements. Plant Mol. Biol. **25**: 429–436.

bot371-07.p65 48 2001/4/19, PM 01:48

野生稻穀蛋白基因 5' 區的分析

吳信淦 陳添進 鍾美珠

中央研究院植物研究所

野生稻穀蛋白基因 5' 區構造曾用選值及定序的方法予以定出。該基因的 5' 區除了與基因表現有關的主要序列片段(如豆素、CAAT、 AACA、 TATA 等)之外,尚有很多推論為可增進表現的片段(如長及短的單向重複片段)及調控片段(如 RY 重複片段、300 bp 片段及能與核蛋白結合的片段等)。各野生稻物種中該基因 5' 區的構造不盡相同,主要是若干大片段的缺失以及序列片段中 209 個對應氮基的取代。從穀蛋白基因 5' 區序列的長度、同源程度以及各序列對應氮基的取代及缺失數據,作者等建議水稻穀蛋白基因族 Glua 可區別為三種成員基因,其 5' 區長度各為 0.5 kb、 0.9 kb 及 1.2 kb。每一成員基因在各染色體組各占一或多個基因座。野生稻物種中同一成員基因則有因其 5' 區序列中少數氮基發生缺失、外加或取代而形成等位基因 (allele)。

關鍵詞:穀蛋白基因; 5'區構造;野生稻; GluA 次族。

bot371-07.p65 49 2001/4/19, PM 01:48