249

Journal of General Virology (1994), 75, 249–257. Printed in Great Britain

ps diffo

Nucleotide sequence and genome characterization of rice yellow mottle virus RNA

Martin Ngon A Yassi, Christophe Ritzenthaler,[†] Christophe Brugidou, Claude Fauquet and Roger N. Beachy^{*}

International Laboratory for Tropical Agricultural Biotechnology (ILTAB/ORSTOM-TSRI), Department of Cell Biology, Division of Plant Biology, MRC7, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037, U.S.A.

The genome of rice yellow mottle virus (RYMV) is a single-stranded positive-sense RNA that is not polyadenylated, and has an M_r of 1.4×10^6 . We present here the 4550 nucleotide (nt) sequence of RYMV RNA, and its predicted genomic organization. The RYMV genomic RNA contains four open reading frames (ORFs). The first (nt 80 to 553) encodes a protein containing 157 amino acids with a predicted M_r of 17.8K. No function has yet been attributed to this product. ORF2 (nt 608 to 3607) encodes a polyprotein of 999 amino acids, with a predicted M_r of 110.7K. The first 134 amino acids of ORF2 are predicted to be the genome-linked protein,

Introduction

Rice yellow mottle virus (RYMV) causes a severe disease of rice in Africa. It was first reported in Kenya (Bakker, 1974), subsequently in many western and eastern African countries (Hull, 1988), and more recently in Madagascar (J. L. Notteghem, CIRAD, personal communication). The virus infects a number of rice types (Oryzae spp.) causing symptoms that include yellowing or orange discoloration of the leaves, reduced tillering, stunting of the plants and sterility of the flowers (Bakker, 1974; Attere & Fatokun, 1983). The virus is naturally transmitted by several species of beetles, most commonly Sesselia pusilla and Chaetocnema pulla and is also mechanically transmissible. Purified preparations of RYMV consist of isometric particles of 25 to 28 nm in diameter (Bakker, 1974; Fauquet & Thouvenel, 1977) that are resolved as two components in isopycnic caesium sulphate sedimentation gradients (Hull, 1988).

† Present address: Institut de Biologie Moléculaire des Plantes, 12 rue Zimmer, 76084 Strasbourg, France.

Sequence data presented in this paper have been assigned the accession no. L20893 in the GenBank database.

0001-1850 © 1994 SGM



VPg, followed by the viral protease, the helicase and the RNA-dependent RNA polymerase. ORF3 is within the boundaries of ORF2 and is predicted to encode a polypeptide with 126 amino acids and an M_r of 13.7K. No function has yet been attributed to this protein. ORF4 (nt 3447 to 4166), which overlaps the 3' terminus of ORF2, encodes a 26K protein. This polypeptide has been identified as the RYMV coat protein. The data presented here confirm that RYMV belongs to the sobemovirus group and thus is a member of the picorna-like family of plant viruses.

It has been proposed that RYMV is a member of the sobemovirus group, of which southern bean mosaic virus (SBMV) is the type member (Matthews, 1982). Viruses in the sobemovirus group are characterized by isometric particles of 28 nm in diameter, a single coat protein of M_r approximately 29K, mechanical and beetle transmissibility, and narrow host range. The genome of sobemoviruses is a single-stranded messenger-sense RNA of M_r 1.4×10^6 . The 5' terminus of the RNA has a genome-linked protein (VPg) and the 3' end is not polyadenylated (Sehgal, 1981; Francki *et al.*, 1985; Hull, 1988).

In this report, we present the complete nucleotide (nt) sequence of RYMV RNA. Similarity of genome organization and sequence comparisons of the proteins predicted to be encoded by RYMV with those of SBMV and other plant RNA viruses confirm that RYMV is a member of the sobemovirus group and extend our knowledge of sobemovirus sequences.

Methods

Virus purification and viral RNA preparation. RYMV was obtained from infected rice fields in the Ivory Coast. The virus was routinely propagated in the rice variety IR8 and purified as previously described by Fauquet & Thouvenel (1977). Viral RNA was isolated by treating the virus with SDS and proteinase K (Dougherty & Hiebert, 1980),



followed by alkaline phenol-chloroform extraction and ethanol precipitation (Hari *et al.*, 1979).

Cloning of RYMV cDNA. The genome of RYMV consists of one single-stranded, positive-sense linear RNA molecule (Hull, 1988). To synthesize the first strand cDNA, 2 µg of RNA was primed with 0.5 µg of random primers (Promega) and the first strand cDNA was synthesized using avian myeloblastosis virus reverse transcriptase (AMV RT) (Promega) and $[\alpha^{-32}P]dCTP$ (NEN). The second strand cDNA was produced by using a mixture of RNase H and DNA polymerase I (Gubler & Hoffman, 1983). The cDNA was treated with T4 DNA polymerase followed by phenol-chloroform extraction and ethanol precipitation. The cDNA molecules were size-selected using a Push Column (Stratagene) and the largest fragments were ligated into the *SmaI* restriction site of the plasmid Bluescript II KS +/- [pBS(KS), Stratagene].

Screening the cDNA library. Approx. 45 bacterial colonies were selected following transformation and were screened by hybridization (Sambrook *et al.*, 1989). The probes used for this purpose were either $[\alpha^{-32}P]dCTP$ -labelled first strand cDNA from RYMV genomic RNA or an $\alpha^{-32}P$ -end-labelled 20-mer degenerate oligonucleotide deduced from sequencing the N-terminal amino acids of RYMV coat protein (CP). The oligonucleotide was labelled with $[\gamma^{-32}P]dATP$ using T4 polynucleotide kinase (Promega). Plasmid was purified from selected cDNA clones and sequenced as described below.

Poly(A) tailing and first strand cDNA synthesis of the 3' terminus of RYMV RNA. Viral RNA was polyadenylated with $[\alpha$ -³⁵S]ATP (Amersham) and poly(A) polymerase (Pharmacia LKB) as described by Smith *et al.* (1988). An oligonucleotide, 5' d[AATTCGCGGGCCG-C(T)15] 3', containing a NotI restriction site was used to prime the synthesis of the first strand cDNA of the poly(A)-tailed RNA. AMV RT and a Promega cDNA synthesis system kit were used according to the manufacturer's instructions.

First strand cDNA synthesis of the 5' terminus of RYMV RNA and dC tailing. To synthesize the first strand cDNA of the 5' terminus of RYMV RNA, a 30-mer oligonucleotide, 5' d(GCGCTCTGAGACT-ATCGCGGCCGCTATCAA) 3', corresponding to the sequence near position 700 of the antisense strand of the RNA was used as primer. The 5' RACE system (BRL) for rapid amplification of cDNA ends was used for the synthesis, purification and dC tailing of the first strand cDNA.

PCR amplification of the first strand cDNAs. For some of the cloning reactions, the first strand cDNA was used as a template for PCR amplification. Reaction mixtures (100 µl) contained 10 mM-Tris-HCl pH 8·3, 50 mм-KCl, 1·5 mм-MgCl₂, 0·001 % (w/v) gelatin, 0·2 mмdNTP (dATP, dCTP, dGTP and dTTP), 100 pmol of each primer, 10 μ l (half volume) of the first strand reaction mixture and 2.5 units of AmpliTaq DNA polymerase (Perkin-Elmer Cetus). After denaturation of the DNA at 95 °C for 3 min, the reaction mixtures were subjected to 13 cycles of 1 min at 94 °C, 2 min at 45 °C and 2 min at 72 °C. In some cases, the annealing temperature was lowered to 37 °C for the first three cycles of the programme. The second strand primer specific for and homologous to nt 2890 to 2905 of the RNA was used for the amplification of the cDNA at the 3' terminus of the RNA. The oligonucleotide used to synthesize the second strand cDNA of the 5' terminus of the RNA was the anchor primer provided by the manufacturer (BRL). After PCR amplification, the products were ethanol-precipitated (Sambrook et al., 1989), digested with suitable restriction enzymes and inserted into the plasmid pBS(KS). Cloned cDNAs were screened for insert size by digestion with the corresponding enzymes. Inserts of the predicted length were subjected to DNA sequence analysis.

Subcloning. The complete nucleotide sequence of the genome was first derived by sequencing the cDNA fragments obtained by the Gubler & Hoffman (1983) cloning procedure and PCR-based cloning. Oligonucleotides with suitable restriction sites were subsequently designed for sites along the genome and used to clone specific viral sequences. The oligonucleotides (as reported in Fig. 2) used included the sequences of nt 4435 to 4450, 3554 to 3579, 2281 to 2302, 1093 to 1117 and 695 to 724, complementary to the RNA, to prime the first strand cDNA synthesis. Oligonucleotides with the same polarity as the RNA and corresponding to nucleotide sequences at positions 1 to 15, 586 to 611, 1086 to 1111, 2281 to 2302 and 3443 to 3457 served as second strand primers. The first strand cDNAs were synthesized using either AMV RT (Promega) or Superscript (BRL). The second strand cDNAs were synthesized in all cases by PCR as described above. After PCR, the fragments were digested with suitable restriction enzymes and inserted into pBS(KS) prepared for this purpose. Clones were screened for insert size and those with the expected insert sizes were sequenced.

۴.

<u>)</u>

Plasmid preparation, cDNA sequencing and analysis. Plasmids were prepared by centrifugation in CsCl (Sambrook et al., 1989), or by a 'miniprep' method that included phenol-chloroform extraction (Serghini et al., 1989). Magic Maxiprep DNA preparation kits purchased from Promega were also used. Exonuclease III treatments of DNA sequences were used to subclone internal sequences of some cloned cDNAs and fragments resulting from the deletions were subjected to sequencing using the T7 and reverse pBS(KS) primers (Stratagene). The cDNA clones obtained by PCR amplification were sequenced using specific oligonucleotides designed from previously determined RYMV sequences. Single-stranded DNA produced by denaturation with NaOH was sequenced by the dideoxynucleotide method (Sanger et al., 1977) using Sequenase version 2.0 or T7 DNA polymerase (U.S. Biochemical Corp. and Pharmacia) and deoxyadenosine 5'-[a-35S]thiotriphosphate (Amersham) as the labelled nucleotide.

Sequencing of the N-terminal region of RYMV coat protein. RYMV CP was purified by first treating 10 μ g of virus at 80 °C in Laemmli sample buffer (62.5 mM-Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 5% 2-mercaptoethanol, 0.001% bromophenol blue). The CP was fractionated by SDS-PAGE (Laemmli, 1970) as previously described (Qu et al., 1991). After electrophoresis, the CP was transferred to Immobilon-P PVDF membrane (Millipore) as described by Matsudaira (1987) and the protein was subjected to microsequencing as previously described (Qu et al., 1991).

Results

Characterization of RYMV CP and RNA

RYMV CP migrated on SDS-PAGE as a double protein band of estimated M_r 28K and 29K (Fig. 1*a*). RYMV RNA isolated from purified virus was resolved as a single component of M_r approx. 1.4×10^6 when subjected to gel electrophoresis (Fig. 1*b*). This M_r is similar to those of the genomes of other sobemoviruses (Hull, 1988).

The purified RNA was tested for infectivity by inoculating rice plants (variety IR8) with different concentrations of RNA 18 days after planting. All plants inoculated with RNA at concentrations of 50 ng/ml or greater developed systemic symptoms typical of RYMV infections.

Since many viral RNAs are polyadenylated at the 3' terminus, we investigated whether RYMV RNA contains



Fig. 1. Analysis of RYMV CP and RNA. (a) SDS-PAGE analysis of RYMV CP. Lane 1, M_r markers; lane 2, RYMV CP. (b) Denaturing agarose gel electrophoresis of purified RYMV RNA. Lane 1, size markers; lane 2, RYMV RNA.

a poly(A) tail. A primer extension reaction using AMV RT (Sambrook *et al.*, 1989) with $oligo(dT)_{12-18}$ was performed. RNA isolated from cassava common mosaic potexvirus (CCMV; Costa *et al.*, 1972) was used as a positive control. No primer extension product was produced when RYMV RNA was used as the template in contrast to the result with CCMV RNA as template (data not shown), indicating that RYMV RNA is not polyadenylated. This result agrees with what is known about sobemovirus genomes (Hull, 1988).

Cloning and sequencing of RYMV RNA

The cloning method using hexadeoxynucleotides to prime first strand cDNA synthesis resulted in the isolation of three cDNA clones that hybridized with the first strand cDNA probe. However, no clone hybridized with end-labelled degenerate oligodeoxynucleotides derived from the 5' sequence of the CP gene. After DNA sequence analysis, it was found that sequences of the clones overlapped and covered the regions of the genome later designated nt 599 to 3105. Sequence alignments and comparison with the nucleotide sequence of SBMV RNA (Wu et al., 1987) and protein amino acid sequence (Hermodson et al., 1982; Wu et al., 1987) were used to determine the sense RNA strand of the cDNA in the overlapping clones. An oligonucleotide was designed to pair with a region where great sequence similarity was found between the genomes of SBMV and RYMV. This oligonucleotide (later identified as nt 2890 to 2905) was of the same polarity as RYMV RNA and was used to prime the synthesis of the second strand cDNA of the 3' terminus of RYMV RNA by PCR. The cloning of the PCR fragments derived from the 3' terminus of the RNA yielded cDNA of 1600 bp. The cloning of the 5' terminus of the genomic RNA using the 5' RACE system (as described in Methods) resulted in cloned cDNA fragments whose sizes were estimated, as predicted, to be 750 bp. The purified cDNA clones were sequenced, and the overlapping sequences were compiled. Every nucleotide of RYMV RNA was sequenced at least once on each strand of the cDNA from each of at least two clones from independent cDNA synthesis reactions.

The complete nucleotide sequence of RYMV RNA is shown in Fig. 2. The RYMV RNA sequence was searched for potential protein-coding sequences in both negative- and positive-strand orientations. No open reading frame (ORF) larger than 240 nt could be found in the negative orientation but four major ORFs were found in the positive-sense strand. The deduced amino acid sequences of the possible translation products of RYMV RNA that exceed 80 residues in length are shown in Fig. 2. The RYMV RNA sequence contains 4450 nt. and is slightly longer than that of SBMV strain C (4194 nt), the only other sobemovirus genomic RNA sequenced to date. The base composition of RYMV RNA shows a high guanine content (28.67%), followed by cytosine (26.29%), uracil (23.21%) and adenine (21.82%). The G+C content is therefore 55%. The calculated $M_{\rm r}$ of the RNA is 1.47×10^6 , in good agreement with the mass estimated by gel electrophoresis (Fig. 1b).

Coding capacity of RYMV RNA

The RYMV genome is compact and most of the predicted ORFs overlap each other (Fig. 3). The exceptions are ORFs 1 and 2, between which there is an intergenic region of six nucleotides. In total, only 330 of the 4450 nt are in non-coding regions.

The first AUG is located at base 80 and is a potential start codon for the first ORF (Fig. 2), which ends at the opal UGA codon at nucleotide 553. ORF1 encodes a protein containing 157 amino acids with a calculated M_r of 17.8K. Sixteen amino acids downstream of the UGA codon, and in the same reading frame, is a UAG amber stop codon at nt 599. This could extend the protein encoded by ORF1 to 172 amino acids with a calculated M_r of 19.5K. Our preliminary data on the *in vitro* translation of RYMV RNA show two proteins with M_r s estimated at 18K and 19K on SDS–PAGE analysis (M. Ngon a Yassi, unpublished). This result supports the hypothesis that the ORF1 could encode two polypeptides which possess common N-terminal sequences but two

ACAAU	10 WGAAGCU	20 AGGAAAGGAGC	30 AUAUUGCGA	40 AAGCAUCCCUC	50 CUUCCGACG/	60 ACAAUUGUAG	70 CCACACUGCAU	80 JCGUGUAUGA	90 CACGGUU
	100	110	120	130	140	150	160	M 170	T R L
GGAAG E	UUCUUAU	ACGACCGACUU	CGCAGACUG	UGGCAAAAGCC	AUCGCCGUGG	GCUAUACGCA	CGCACUCÁCCI	GGGUUUGGC	AUUCUCA
	190	200	210	220	230	240	250	260	H S Q 270
T	W D V	D A V	N D P 1	V L S A	D F N	CCGAGAAGGU PEKV	G W V	S V S	UUGCCUG F A C
UACUC	280 GGUGUAC	290 AGCCCACUACU	300 ACACGUGUG	310 AGCAGGUGAAA	320 UAUUUCACUA	330 AUAUUCCGCC	340 UGUUCAUUACO	350 ACGUGGUGU	360 GCGCUGA
T	R C T 370	A H Y 380	YTCI 390	EQVK 400	Y F T 410	N I P P 420	VHY 430	D V V	C A D
UUGCG	AGCGUAG	UGUUCAGCAGG	ACGACGAGA	UCGACCGCGAG	CACAACGAGO	GUAACGCAGA	GAUUUCUGCCI	GCAACGCUC	GAGCCCU
	460	470	480	490	500 S	510	1 S A 520	C N A 530	RAL 540
AAGUG S	AGGGAAG E G R	GCCAGCGAGUC	UGGUUUACC	UUUCGCGGGGAC L S R D	A C D	UACCCGAGCA I P E H	CUCCGGAACGI S G T	C R F	ACAAAUA D K Y
UCUCA	550 ACUUUUG	560 AGAGAAUCAGU	570 CCUGCCGAC	580 AUCUAAGGAGC	590 GGUCUUGGUI	600 ICAGACUAACG	610 CGGGAUGGGCI	620 ICOUCUGUUG	630
L	NF. 640	E N Q 650	S C R 1	H L R S	G L G	S D .	M G	S S V	V G R
cuuuc	UCACUCG	CACCAAGGAAA	CGACGAGGG	UGGACUCCCAC	CUNCUCANCO	UCUUGGCUGG	GUUGAUAGCGO	GCCGCGAUAG	UCUCAGA
F	ьт к 730	. T K E 740	T T R 7 750	VDSH 760	L I 770	V L A G 780	L I A 790	A A I 800	V S E 810
GCGCC R	UACCCGU L P V	GACGGCGUCAU	UGUGGGCAAI L W A :	UCCCCUCAGCU I P S A	AUCAUAGCA	ACUGGAUCGU N W I V	GCUGUCAGCUC L S A	ACGAGAGCU	UCAGCCG F S R
GUUCG	820 UUGAGGG	830 GGUUGAGAUUG	840 AGCCUAUGU	850 CCAUGCUUCGG	860 NAUGCAAGO	870	880 CCCUAGRINIC	890	900
F	V E G	VEI	EPM	SNLR	YGK	V Q S A	PRF	D P S	R G Y
UGUGG	UCGACGU	AUCUUACAACG	GCCACGUGA	940 UUCCUGUGAUA	950 UUGGACUUU	960 CCACGACAAC	GGCUUUGUCGO	980 SUUCCCCAAA	990 GAGUAAA
v	1000	S Y N 1010	GHV: 1020	I P V I 1030	L D F 1040	T T T T 1050	A L S 1060	V P Q 1070	R V N 1080
PCCCG	GGGUCUC G V S	GAUGGAGGCAA M <u>E A</u> S	GUCGUGGGG R G G	GGCUUCCACCA L P P	ACGUCCGUCA T S V 1	AACUCGAAGA	UGUGCCGCCG	AGCGUUGUGG	UUUUGUA
CCACG	1090	1100	1110	1120	1130	1140	1150	1160	1170
н	D S V	RLG	LGT	R V R T	P T G	R D L L	M T N	H H I	A A L
AGAAC	CCAAUGG	UAUCGCGUACA	1200 AGGGUCACC	1210 UUAAGAAAGUG	1220 GCUCUGGACO	1230 SCGCCAGUCAU	1240 CGCAUGUGACO	1250 AUCCACACA	1260 UAGACUG
Е	PNG 1270	1 A Y 1280	К G H : 1290	LKKV 1300	ALD 1310	A P V I 1320	A C D 1330	H P H 1340	I D C 1350
CGCGU	UUUACGA FYE	GGUGCCUCCCA	AGAUCUGGU K T W	CUCUCUUGGGG	GUCAAGUCUG	CCAGUCUGAA	GCCUCUGGUU	AGCAGACUG	CGGUGUC
ACUCU	1360	1370	1360	1390	1400	1410	1420	1430	1440
L	F G G	S S S	TDF:	S S C V	GGGAUCGCCC G I A	Q I G D	N P F	LUGAUAAGGC	AUCAAAG H Q S
CACGA	1450 CUUGCAG	1460 UGGGUGGUCGG	1470 GCUCCCCGC	1480 UCUACCACAAA	1490 GGUUGCGUGG	1500 SUUGGCUUGCA	1510 UAUAGGUGCUG	1520 SCGGAUGGUU	1530 ACAAUGU
т	T C S 1540	G W S 1550	G S P : 1560	LYHK 1570	G C V 1580	V G L H 1590	I G A	ADG 1610	Y N V
GGCAU	CUAAUGU	AGCCUGGUACU	UUCAUACUU	UCAAGAAGGAA	GUGAUCGUUG	AGUCUCCUUU	UGAGAUUUACO	GUAAAUUCC	GCGAAGC
	1630	1640	1650	1660	1670	1680	1690	<u>G K</u> F 1700	R E A 1710
N	S E E	AUAUGACGAGA Y D E	GUCUACGAC S L R	ACGGGGUGCAG H G V Q	UAUGCAGAGI YAE	JACGACUUCUC ¥ <u>D. F S</u>	UGGUGAUACCI	AUCCGGGGCUU I R A	CUUCGAA S S N
UACCU	1720 JGGGUGCG	1730 UGAGAGAGAGA	1740 GGUACCACG	1750 CUGAGGAACGU	1760 CGUAAGUCCO	1770	1780 ·	1790	1800
T	WVR		~ ** **				104004101101		ncenone
	1810	1820	1830	A E E R	RKS	GQPS	WAD	RFG	D D S
UGGCG	1810 AGGAUGU	1820 UGAUAUCGAGA	1830 CAUCGCAUC	A E E R 1840 CCGUAGCACCA	R K S 1850 UCAAUACCCI	G Q P S 1860 AGAACGCGGGCG	W A D 1870 GAAACGGUCAA	R F G 1880 AGCGCGUUG	D D S 1890 AGCAGUU
UGGCG G	1810 AGGAUGU E D V 1900	1820 IUGAUAUCGAGA DIE 1910	R Y H 1830 CAUCGCAUC T S H 1920	A E E R 1840 CCGUAGCACCA P V A P 1930	R K S 1850 UCAAUACCCF S I P 1940	G Q P S 1860 AGAACGCGGGCG R T R R 1950	W A D 1870 GAAACGGUCAJ K R S 1960	R F G 1880 AGCGCGUUG K R V 1970	D D S 1890 AGCAGUU E Q F 1980
ugged g cgueg V	1810 AGGAUGU E D V 1900 ACGCGGU D A V	1820 IUGAUAUCGAGA DIE 1910 UUCCAGAGUGCU SEC	R Y H 1830 CAUCGCAUC T S H 1920 CCUUCUCGU S F S	A E E R 1840 CCGUAGCACCA P V A P 1930 UCGAGUCUGCU F E S A	R K S 1850 UCAAUACCCI S I P 1940 ICACGAGGGGJ H E G	G Q P S 1860 AGAACGCGGGGG R T R R 1950 AUUGUGCCAGA I V P E	W A D 1870 GAAACGGUCAI K R S 1960 GACCUCAGCOU	R F G 1880 AGCGCGUUG K R V 1970 JAUGACCACG Y D H	D D S 1890 AGCAGUU E Q F 1980 OUCCUUU V P L
UGGCG G CGUCG V	1810 AGGAUGU E D V 1900 ACGCGGU D A V 1990	1820 UGAUAUCGAGA 'DIE 1910 UUCAGAGUGCU 'SEC 2000	R Y H 1830 CAUCGCAUC T S H 1920 CCUUCUCGU S F S 2010 CCUUCCGC	A E E R 1840 CCGUAGCACCA P V A P 1930 UCGAGUCUGCU F E S A 2020	R K S 1850 UCANUACCCI S I P 1940 ICACGAGGGGI H E G 2030	G Q P S 1860 AGAACGCGGGGG R T R R 1950 AUUGUGCCAGA I V P E 2040	W A D 1870 GAAACGGUCAA K R S 1960 GACCUCAGCUU T S 'A 2050	R F G 1880 AAGCGCGGUUG K R V 1970 JAUGACCACG Y D H 2060	D D S 1890 AGCAGUU E Q F 1980 SUUCCUUU V P L 2070
UGGCG G CGUCG V AAACU N	1810 BAGGAUGU E D V 1900 BACGCGGU D A V 1990 BGCCAGGG C Q G	IB20 UUGAUAUCGAGA DIE 1910 UUUCAGAGUGCU SEC 2000 GGCGGGCUCGA GC SGGCUCGA SAGS	I Y H 1830 CAUCGCAUC T S H 1920 CCUUCUCGU S F S 2010 GCCUUCGGG S L R	A E E R 1840 CCGUAGCACCA P V A P 1930 UCGAGUCUGCC F E S A 2020 CGAGUCCGCCC A S P P	R K S 1850 S I P 1940 CACGAGGGGJ H E G 2030 CUUGGACGGCC L D G	G Q P S 1860 RGAACGCGGCGG R T R R 1950 AUUGUGCCAGA I V P E 2040 JUAUCCAACUC L S N S	W A D 1870 GAAACGGUCAN K R S 1960 GACCUCAGCUU T S A 2050 GGAGAAUACCO E N T	R F G 1880 AAGCGCGUUG K R V 1970 JAUGACCACG Y D H 2060 GCUGGGACUC A G T	D D S 1890 AGCAGUU E Q F 1980 SUUCCUUU V P L 2070 SCCUCGGU P S V
UGGCG G CGUCG V AAACU N GAUUC	1810 AGGAUGU E D V 1900 ACGCGGGU D A V 1990 AGCCAGGG C Q G 2080 CCUCCCU	1820 IUGAUAUCGAGA DIE 1910 IUUCAGAGUGCU CSECGGGCUCGA GGCGGGCUCGA A G S 2090 IUCCGACGGAAU	R Y H 1830 CAUCGCAUC T S H 1920 CCUUCUCGU S F S 2010 GCCUUCGGG S L R 2100 GCCCUUCAG	A E E R 1840 CCGUAGCACCA P V A P 1930 UCGAGUCUGCU F E S A 2020 CGAGUCCGCCC A S P P 2110 CUACGUUGGAA	R K S 1950 UCAAUACCC/ S I P 1940 UCACGAGGGGG H E G 2030 UUUGGACGGCG L D G 2120 AAUCGGGUGG	G Q P S 1860 AGAACGCGCGCG R T R R 1950 AUUGUCCACAGA I V P E 2040 JUAUCCAACUC L S N S 2130 JCAUCUUUUGGA	W A D 1870 GAAACGGUCAJ K R S 1960 GACCUCAGCUI T S A 2050 GGAGAAUACCC E N T 2140 GAACAUGCUGG	R F G 1880 AAGCGCGUUG K R V 1970 JAUGACCACG Y D H 2060 SCUGGGACUC A G T 2150 SGAAAAGUGU	D D S 1890 AGCAGUU E Q F 1980 SUUCCUUU V P L 2070 CCUCGGU P S V 2160 SUUCCUCGCA
UGGCG G CGUCG V AAACU N GAUUC I	1810 AGGAUGU E D V 1900 ACGCGGU D A V 1990 AGCCAGGG C Q G 2080 CCUCCCU P S L	1820 1820 1920 1920 1910 1900	R Y H 1830 CAUCGCAUC T S H 1920 CCUUCUCGU S F S 2010 GCCCUCCGGG S L R 2100 GCCCUUCAGG P F S C P S	A E E R 1640 CCGUAGCACCA P V A P 1930 UCGAGUCGCUCGC F E S A 2020 CGAGUCCGCCC A S P P 2110 CUACGUUGGAA Y V G A T L E	R K S 1850 1850 1950 1940 1940 1040 1040 1940	G Q P S IGAACGCGGCG R T R R 1950 MUUGUGCCAGA I V P E 2040 JUAUCCAACUC L S N S 2130 JCAUCUUUGGA V I F G S S L E	W A D 1870 GAAACGGUCAJ K R S 1960 GACCUCAGCUG T S 'A 2050 GGAGAAUACCU E N T 2140 GAACAUGCUG E H A C N M L	R F G 1880 AAGCGCGUUG K R V 1970 JAUGACCACG Y D H 2060 GCUGGGACUC A G T 2150 GGAAAAGUGU 3 K S V G K V	D D S 1890 AGCAGUU E Q F 1980 SUUCCUUU V P L 2070 CCUCGGU P S V 2160 ICUGCGCA S A O
UGGCC G CGUCC V AAACU N GAUUC I GCUGU	1810 GAGGAUGU E D V 1900 GACGCGGU D A V 1990 GGCCAGGG C Q G 2080 CCUCCCU P S L 2170 IC2AAGAC	1220 1220 120 1000000000000 10000000000	R Y H 1830 CAUCGCAUC T S H 1920 CCUUCUCGU S F S 2010 GCCUUCGGG S L R 2100 GCCUUCGGG F F S C P S 2190	A E E R 1840 CCGUAGCACCA P V A P 1930 UCGAGUCGCCC F E S A 2020 CGAGUCGCCC A S P P 2110 CUACGUUGGAA Y V G A T L E 2200	R K S 1850 UCANUACCCI S I P 1940 ICACGAGGGGI H E G 2030 UUGGACGGCU L D G 2120 AAUCGGGUGU N R <u>V</u> 22210 UUGGCUGGCC	G Q P S 1860 IGAACGCGGCG R T R R 1950 IUUGUGCCAGA I V P E 2040 JUAUCCAACUC L S N S 2130 JCAUCUUUGGA V I F G S S L E 2220	W A D 1870 GAAACGGUCAJ K R S 1960 GACCUCAGCUU T S 'A 2050 GGAGAAUACCCC E N T 2140 GAACAUCCUGG E H A C N M L 2230	R F G 1880 1880 1880 18970 1970	D D S 1890 AGCAGUU E Q F 1980 OUCCUUU V P L 2070 CCUCCGU P S V 2160 CCUCCGU P S V 2160 CCUCCGCA C A S A Q 2250
UGGCC G CGUCC V AAACU N GAUUC I GCUGU A V	1810 AGGGAUGU E D V 1900 AGCCGGU D A V 1990 GCCAGGG C Q G 2080 CCUCCCU P S L 2170 CCAAAGAC K D	1220 1220 1220 120 1210 120 120 1	R Y H 1830 CAUCGCAUC T S H 1920 CCUUCUCGU S F S 2010 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S C P S 2190 AUUCCCAGA	A B E R 1840 CCGUAGCACCA P V A P 1930 UCGAGUCUGCC F E S A 2020 CGAGUCGCCCC A S P P 2110 CUACGUUGGAA Y V G A T L E 2200 UCUUGAAGGAU L E G	R K S 1850 UCANUACCCI S I P 1940 (CACGAGGGGI H E G 2030 UUGGACGGCU L D G 2120 AAUCGGGUGU K S G V N R <u>V</u> 2210 UUUGGCUGGCC F G W F	G Q P S 1860 IBACGCGGGGG R T R R 1950 IUUGUGCCAAR I V P E 2040 IUAUCCAACUC L S N S 2130 JCAUCUUUGGA V I F G S S L E 2220 UGAGAGGGGA P E R G	W A D 1870 GAAACGGUCAJ K R S 1960 GACCUCAGCUCI T S A 2050 GGAGAAUACCCC E N T 2140 GAAACGUCUGCUGC E H A C N M L 2230 AGUAAAGCACG S K A J	R F G 1880 1880 AAGCGCGUUG K R V 1970 1906 CUGGGACCCCG Y D H 2060 CCUGGGACUC GCUGGGACUC GCUGGGACUC G K V G K V C K V C K V C L D F	D D S 1890 1800 1980
UGGCC G CGUCC V AAACU N GAUUC I GCUGU A V L	1810 AGGAUGU E D V 1900 ACCCCGGU D A V 1990 MGCCAGGG C Q G 2080 CCUCCCU P S L 2170 MCAAAGAC K D S K T 2260	1220 1220 1220 120 120 120 120 12	R x H 1830 CAUCGCAUC T S H 1920 CCUUCCUCGU S F S 2010 GCCUUCGGG S L R 2100 GCCUUCGGG P F S C P F S 2190 AUUCCCGAN F P D Y S Q 2280	A B E R 1840 CCGUAGCACCA P V A P 1930 UCGAGUCUGCC F E S A 2020 CCGAGUCGGCCC A S P P 2110 CUACGUUGGAA Y V G A T L E 2200 UCUUGAAGGAU L E G I L K D 2220	R K S 1850 NICANUACCCI S I P 1940 (CACAAGAGGGE H E G 2000GACGGCC L D G 2120 AAUCGGGUCC K S G V N R V 2210 UUUGCUGGCC F G W I L A G 2300	G Q P S 1860 AGAACGCGGGGG R T R R 1950 AUUGUGCCAGA I V P E 2040 JUAUCCAACUC L S N S 2130 JCAUCUUUGGA S S L E 2220 JUGAGAGGGA P E R G L R G E 2310	W A D 1870 GAAACGGUCAJ K R S 1960 GACCUCAGGUCAJ T S'A 2050 GGAGAAUACCUG E H A Q140 GAACAUGCUGA N E H A Q10 M L Q230 AGUAACAGCAG S K A Q 2320 V K	R F G 1880 1880 1970 1970 1970 1970 1970 1906 2060 2000 20060 200	$ \begin{array}{ccccccc} & D & D & S \\ & & 1890 \\ & & 1AGCAGUU \\ & & AGCAGUU \\ & & 1980 \\ & & 1980 \\ & & 0UUCUUU \\ & & V & P & L \\ & & 2070 \\ & & 0UUCUCGCA \\ & & V & P & L \\ & & C & A \\ & & C & C \\ & & C & A \\ & & C & C \\ & & C & C \\ & & C & A \\ & & C & C \\ & & C & C \\ & & C & C \\ & & C & C$
UGGCC G CGUCC N AAACU N GAUUC I GCUGU A V L CCUCC P P	1810 AGGAUGU E D V 1900 ACGCGGU D A V 1990 GCCAGGG C Q G 2080 CCUCCU P S L 2170 V K D S K T 2260 CAACGCS CACCS C S R	1120,0 1120,0 100,0 1910 1	R Y H 1830 CAUCCCAUC T S H 1920 CCUUCUCGG S F S 2010 CCUUCGGG S L R 2100 CCUUCGGG S L R 2100 AUUCCCAGA F P S 2190 AUUCCCAGA F P D Y S Q 2280 GAACGGUAU N G H	A E E R 1840 CCGUAGCACCA P V A P 1930 UCGAGUCGCC CGAGUCGCCC CCACCCGUCGAC X P P 2110 CUACGUUGGAA Y V G A T L E 2200 UCUUGAAGGAU L E G 2200 UCUUGAAGGAU L E G 2200 S P R	R K S 1850 UCANUACCCJ S I P 1940 ICACGAGGGGG H E G 2030 H E G 2030 IUUGGACGGCG L D G 2120 N R Y 2210 UUUGGCGGGCG N R A S 2300 ICUCCGGCGAA A R A R	G Q P S 1860 AGAACGCCGCCG R T R R 1950 AUUGUGCCAGA 2040 UUAUCCAACUU L S N S 2130 I V P E 2130 I V P E 2130 CAUCUUUGGA CAUCUUUGA CAUCUUUGA CAUCUUCUUUGA CAUCUUCUUUGA CAUCUUUGA CAUCUUUUGA CAUCUUUU CAUCUUU CAUCUUUU CAUCUUUU CAUCUUUU CAUCUUUU CAUCUUUU CAUCUUUU CAUCUUUU CAUCUUUU CAUCUUUU CAUCUUUUU CAUCUUUUUUUU CAUCUUUUUUU CAUCUUUUUUUUUU	W A D 1870 GANACGGUCAJ K R S 1960 GACCUCAGCUU T S 'A 2050 GGAGAAUACCC E N T 2140 GAACAUCCUG GAACAUCCUG GAACAUCCUG GAACAUCCUG GAACAUCCUG S K A I V K Q 2320 CAUCAAGAA	R F G 1880 AGCCCCUUG K R V 1970 JAUGACCACC Y D H 2060 CUUGACACCA X D H 2050 CUUGACACCA X D H 2150 GT 2150 GT 2450 G K V C Z 230 GCUUGACCCAA S L D F S L T 2330 GGUACCCAA	D D S 1890 1890 1890 1990 1990 1980
UGGCC G CGUCC V AAACU N GAUUC I GCUGU A V L CCUCC P F L	1810 AGGAUGU E D V 1900 1900 1900 CCCAGGG C Q G 2080 P S L 2170 CCAAGGG C X D S K T 2260 SAAGCCG C S R Q A G 2350	2 1220 1230 2 1 2 1310 131	R Y H 1830 CAUCCCAUC T S H 1920 S F S 2010 GCCUUCCGG S L R 2100 GCCUUCAGA F F S 2190 AUUCCCAGA F F D 2190 AUUCCCAGA F F D 2290 GAACGGUAU GAACGGUAU N G M R T V 2370	A E E R 1840 CCGUAGCACCA P V A P 1930 UCGAGUCGCC CGAGUCGCCC CCACCCGUCGAC X P P 2110 CUACGUUGGAA Y V G A T L E 2200 UCUUGAAGAGAU L E G 2200 GUCCCCCCGGG S P R C P P G 2300	$ \begin{array}{c} {\sf R} \; {\sf K} \; {\sf S} \\ 1850 \\ 1850 \\ {\sf SI} \; {\sf IP} \\ 1940 \\ {\sf CACGAGGGGG} \\ {\sf H} \; {\sf E} \; {\sf G} \\ 2030 \\ {\sf H} \; {\sf E} \; {\sf G} \\ 2120 \\ {\sf AUUCGACGGCG} \\ {\sf L} \; {\sf D} \; {\sf G} \\ 2120 \\ {\sf AUUCGGGGGGG} \\ {\sf AUUCGGGGGG} \\ {\sf N} \; {\sf R} \; {\sf Y} \\ 2210 \\ {\sf UUUGCCGGCGC} \\ {\sf N} \; {\sf R} \; {\sf N} \; {\sf S} \; {\sf G} \\ {\sf Y} \\ {\sf N} \; {\sf R} \; {\sf Y} \\ 2210 \\ {\sf UUUGCCGGCGC} \\ {\sf G} \; {\sf W} \; {\sf I} \\ {\sf L} \; {\sf A} \; {\sf G} \\ 2300 \\ {\sf CUCCGCGCAAC \\ {\sf R} \; {\sf A} \; {\sf Q} \\ {\sf Z} \\ {\sf Z} \\ {\sf A} \; {\sf Q} \\ {\sf Z} \\ {\sf Z} \\ {\sf A} \; {\sf Q} \\ {\sf Z} \\ {\sf Z} \\ {\sf A} \; {\sf Q} \\ {\sf Z} \\ {\sf$	$ \begin{array}{ccccccc} G & Q & P & S \\ I & B60 \\ I & B60 \\ I & B60 \\ I & T & R \\ I & P50 \\ I & P50 \\ I & P & R \\ I & P \\ I \\ I & P \\ I \\ I & P \\ I \\$	W A D 1870 CAAACGGUCAA K R S 1960 GGACGUCAGGUU T S A 2050 GGAGAUACCG E N T 2140 GGAGAUACCG E N A 2050 GGAGAUACGG E N A 2140 AGUAAAGCAGI S K A V K Q 2320 AGUAAAGCAGI C P R I L Q E 2410	R F G 1880 JAGCGCGUUG K R V 1970 JAUGACCACG Y D H 2050 CUGGGACUC A G T 2150 GCUGGACUC A G Y 2150 GCUGGACUC S L D F S L J F 2330 GGUACCCAA K V P 2420	D D S B
UGGCC G CGUCG V AAACU N GAUUC I GCUGU A V L CCUCC P F L CCUCCA	1810 AGGANGU E D V 1900 1900 1990 C Q G 2080 P S L 2170 VCAAGGC V K D S K T 2260 AGCCAGGC C Q G 2080 P S L 2170 S K T 2260 C AGCCGG C AGCCGGG C AGCCGG C AGCCGGG C AGCCGG C AGCCGG	1020 1020 1020 1020 1020 1000	R X H 1830 CCUUCCGUC CCUUCCUCGU S F S 2010 GCCUUCCGG S L R 2100 GCCUUCCGG S L R 2100 GCCUUCCGG CCUUCCGG CCUCCGG S L R 2100 GCCUUCCGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGGGU S R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGG S L R 2100 GCCUUCGGGGU S L R 2100 GCCUUCGGGGU S L R 2100 GCACGGGU GCACGGGU GCACGGGU GCACGGGU GCACGGGU GCACGGGU GCACGGU GCACGGU GCACGGU GCUUCGGGGU S L R 200 GCACGGU GCACGGU GCUUCGGGGU GCACGGU GCACGGU GCACGGU GCACGGU GCACGGU GCACGGU CGUUCGGGU S S L S 200 CO CO CO CO CO CO CO CO CO CO	A E B R 1840 CCCGUACGACCA 1940 1930 UCCAGUCUGCA 2020 CCAGUCCGCCC A S P P 2110 CUCUGAUGAAGGAU V V G A T L E 2200 UCUUGAAGGAU L E G I L K D 2290 UCUUCCCCCCGCGG S P R C P P G 2300 UCGAGCACCUC	R K S 1850 S I P 1940 CACGAGGGG CACGAGGGG H E G 2030 CUUGGAGGGG K S G V N R V 2120 MAUCGGUGG K S G V N R V 2210 UUUGCAGGGC P G W I L A G 2300 CUUCGCGGCA A R A S L A Q 2390 CUUCGCGGAA	G Q P S 1860 IGANCGCGCC R T R R 1950 1005	W A D 1870 CANACGUCA K R S 1960 GACCUCAGCU T S A 2050 GGAGAUACCAGCU C L A C	R F G 1980 1980 1970	$ \begin{array}{c} D & D & S \\ 1890 \\ 1800 \\ 1800 \\ 1800 \\ 1980 \\ 000 \\$
UGGCC G CGUCG V AAACCU N GAUUC I GCUCGU P F L CCCCCA P F C P F	1810 SAGGAUGU SAGGAUGU SACGCGGU D A V 1990 SACGCGGGU C Q G 2080 SCCUCCCU P S L 2170 S K T 2250 S R Q A G 2350 GGCGAUG Q A G 2350 C A M R R C 2460	1020 1020 101 01 101 101 1010 1000000000000000000000000000000000000	R X H 1830 CAUCCGAUC T S H 1920 CCUUCUCGU S 2010 GCCUUCAGE S F S L R 2100 GCCUUCAGE F F Q P S Q V S Q S V S Q S V S Q S V S Q S N G N G X370 X370 X370	$ \begin{array}{cccc} A & B & B & R \\ 1840 \\ CCCGUACGACCACCCGUACGACCAP & P & A & P \\ 1930 \\ UCCAGUUCGACC \\ P & E & S & A \\ 2020 \\ CCAGUUCGGCCA \\ A & S & P & P \\ 2100 \\ CUCUGAAGAAGAU \\ V & V & G \\ A & T & L & E \\ 2200 \\ UCUGAAGAAGAU \\ L & E & G \\ I & L & K & D \\ 2200 \\ UCUGCACCCCCGCG \\ S & P & R \\ C & P & P & G \\ 2300 \\ UCGACGACAUC \\ R & R & H \\ F & D \\ D & D \\ D & D \\ \end{array} $	$ \begin{array}{cccc} R & K & S \\ 1850 \\ 1850 \\ S & I & P \\ 1940 \\ CACGAGGGGI \\ R & B & CACGAGGGGI \\ L & D & G \\ 2030 \\ AUGCGGUGG \\ K & S & G & V \\ 2120 \\ AUGCGGUGG \\ K & S & G & V \\ 2210 \\ UUGCAGGGC \\ P & G & W & I \\ L & A & G \\ 2300 \\ CUUCGCGGCAGAA \\ L & R & A \\ 2390 \\ CUUCGCAGGAGA \\ L & R & S \\ F & 0 \\ 210 \\ CUUCGCAGGA \\ R & A \\ S & S \\ L & R \\ R & S \\ F & 0 \\ R & S \\ R &$	G Q P S 1860 1860 1860 1860 1860 1950 2040 1950 2040 1950 2040 1950 2040 1950 2040 1950 2040 1950 2040 1950 2040 1950 2040 1950 2040 1950 2040 2050 20	W A D 1870 CAAAACGGUCAX K R S 1960 GACCUCAGGUU S A 2050 GGACAAUACCUC E N T 2140 GAACAUACCUG E N A 2230 AGUACAAGAA AGUACAAGAA CUCCCAAGAA T P R N 22410 CUCUCGCAAGAA CUCUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAA CUCUCGCAAGAAA CUCUCGCAAGAAA CUCUCGCAAGAAA CUCUCGCAAGAA CUCUCGCAAGAAA CUCUCGCAAGAAA CUCUCAAGAA CUCUCGCAAGAAA CUCUCAAGAA CUCUCAAGAA CUCUCAAGAAAA CUCUCAAGAAAA CUCUCAAGAAAA CUCUCAAGAAAAAAAA CUCUCAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	R F G 1980 1980 1970	$ \begin{array}{c} D & D & S \\ 1890 \\ 1800 \\ 1800 \\ 1800 \\ 1980 \\ 000 \\$
UGGCC G CGUCG V AAACCU N GAUUC I GGCUGU V L CCUCC P F C CCCCA P C CCCUC	1810 SAGGAUGU SAGGAUGU SACGCGGU D A V 1990 SACGCGGG C Q G 2080 SCCUCCCC P S L 2170 CAAAGACG P S L 2170 CAAAGACG P S R 2 250 S R Q A G Q A G C 2050	1020 1020 101 101 101 101 1010 1000000000000000000000000000000000000	R X H 1830 CAUCEGCAUC T S H 1920 CCUUCLCGC S 2010 GCCUUCAGE S 2 I R C 2 I R C 2 I R G S 7 S H T Y 8 G P S C 7 S Q 2190 GAACGGUAD 8 F P N G Q 3 GUGCAGGUAD N G 2370 AGUGGAGGU V V V V V 2 W C V V V V 0 W K V V V V V Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z	$ \begin{array}{cccc} A & E & B & R \\ 1840 \\ CCGUAGCACCA \\ DCGAGUCGCACA \\ DCGAGUCGCCC \\ F & E & S & A \\ 2020 \\ CGAGUCCGCCC \\ A & S & P & P \\ 2110 \\ CUACGUUGGAA \\ V & V & G \\ A & T & L & E \\ 2200 \\ CUCUGAAGUCCCCCCGG \\ R & A & H \\ F & D & C \\ R & H \\ 2470 \\ UUGCGAAUUCC \\ UGCGAAUCC \\ \end{array} $	$ \begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ 1960 \\ 1000 \\ $	G Q P S 1860 UGALCGCGCC UGALCGCGCC 1950 2040 1040 CALC 1040 CALC 1040 104	W A D 1870 CAAACGGUCA, K R S 1960 GACCUCAGCUI S A 2050 GGACAUACCC E N T 2140 CACACAUCCUG E N A L 2230 ACUCAAGAA X K Q 2320 ACUCCAAGAA T P R 1 240 ACUCCAAGAA P L R 1 L C E 2500 CCUCCGCAGAA	R F G 1880 1880 1970 19070	D D S 1890 AGCAGUU F Q P 1980 1980 CCUCGUUU Y P L 2070 CCUCGUU F S V 2160 ICUCCCAU F S L 2250 ICCUCAU S A Q 2250 ICCUCAU S A L I 2340 ICCUCAU S A L I 2340 ICCUCAU S C E L V N S 2520 IUCCUCAU ICCUCAU S A 255 ICCUCAU ICUCAU
UGGCC G CGUCC V AAACU N GAUUCC I GCUGU A V L CCUCCA P F L CCCCA P P C CCCU C R F A	1810 SAGGAUGU SAGGAUGU SAGGAUGU E D V 1900 SACGCGGU D A V 1990 SACGCGGU P S L 2170 CAAAGACG V K D S K T 2250 GGAGGUUGU Q A G Q A R C Q A R C Q A R C Q A M R R C 2440 CCUCCGCC S S P	1020 1220 1210 10000A3600000 2000 2000 2000 00000000000000 2000 000000000000000000000000000000000000	N X H 1830 CAUCEGCAUC 1920 CCUUCUCEGU 2010 GCCUUCAGE 5 L R 2100 GCCUUCAGE GC 0 F P S 7 P S C P 8 C P S C 9 S Q 2190 AUUCCAGEGUAD 040/CCAGEGUAD F P M G Y 300/GACGUAD S 2370 H C Y 2490 VUGCGGGGCG V V V V V Z 2400 UGCGGGGCG R T Z AGUGCAGGUAD Z <td>$\begin{array}{cccc} A & E & B & R \\ 1840 \\ CCGUAGCACCA \\ DCGUAGCUCCACACACACACACACACACACACACACACACA$</td> <td>$\begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ S & I & P \\ 1940 \\ 10$</td> <td>G Q P S 1860 UGALGCGGCG UGALGCGGCG 1950 2040 100 UGACAAUC L S N S 2130 UGALGCAUC UGAGAGGGCA 2210 CCAUCUUGGA CCAUC</td> <td>W A D 1870 CAAACGGUCAA CAAACGGUCAA K R S 1960 GACCUCAGCUU T S A 2050 GGACAUAACCUC E N T 2140 CACAUCAGUCUGG GACAUACUCCUG E H A C 2230 ACUCCAAGAA X R S CUCCAAGAA P L R 1 L C E 2500 UGCGAGAGAU A R D</td> <td>R F G 1880 1880 1970 19070 19070 19070 2060 2060 2000 200 2</td> <td>D D S 1890 AGCAGUU F Q P 1980 1980 CCUCGUUU V P L 2070 CCUCGUU CCUCGUU CCUCGUU CCUCGUU CCUCGUU CCUCCUU C A S A Q 2250 (CCUCCUU V P L L 2340 (CCUCCUU V P L L 2340 (CCUCCUU V P L L 2340 (CCUCCUU V P L L 2350 (CCUCCUU V P L L 2350 (CCUCUU V P L L 2350 (CCUCUUU V P L L 2350 (CCUCUUU V P L L 2350 (CCUCUUU V P L L 2350 (CCUCUUU V P L L 2350 (CCUCUUUU V P L 2350 (CCUCUUUU V P L 2350 (CCUCUUUU V P L 2350 (CCUCUUUU V P L 2550 (CCUCUUUU V P L 2550 (CCUCUUUU V P L 2550 (CUCUUUUUUU V V P 2550 (CCUCUUUUUU V V V 2550 (CCUCUUUUU V V V 2550 (CUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU</td>	$ \begin{array}{cccc} A & E & B & R \\ 1840 \\ CCGUAGCACCA \\ DCGUAGCUCCACACACACACACACACACACACACACACACA$	$ \begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ S & I & P \\ 1940 \\ 10$	G Q P S 1860 UGALGCGGCG UGALGCGGCG 1950 2040 100 UGACAAUC L S N S 2130 UGALGCAUC UGAGAGGGCA 2210 CCAUCUUGGA CCAUC	W A D 1870 CAAACGGUCAA CAAACGGUCAA K R S 1960 GACCUCAGCUU T S A 2050 GGACAUAACCUC E N T 2140 CACAUCAGUCUGG GACAUACUCCUG E H A C 2230 ACUCCAAGAA X R S CUCCAAGAA P L R 1 L C E 2500 UGCGAGAGAU A R D	R F G 1880 1880 1970 19070 19070 19070 2060 2060 2000 200 2	D D S 1890 AGCAGUU F Q P 1980 1980 CCUCGUUU V P L 2070 CCUCGUU CCUCGUU CCUCGUU CCUCGUU CCUCGUU CCUCCUU C A S A Q 2250 (CCUCCUU V P L L 2340 (CCUCCUU V P L L 2340 (CCUCCUU V P L L 2340 (CCUCCUU V P L L 2350 (CCUCCUU V P L L 2350 (CCUCUU V P L L 2350 (CCUCUUU V P L L 2350 (CCUCUUU V P L L 2350 (CCUCUUU V P L L 2350 (CCUCUUU V P L L 2350 (CCUCUUUU V P L 2350 (CCUCUUUU V P L 2350 (CCUCUUUU V P L 2350 (CCUCUUUU V P L 2550 (CCUCUUUU V P L 2550 (CCUCUUUU V P L 2550 (CUCUUUUUUU V V P 2550 (CCUCUUUUUU V V V 2550 (CCUCUUUUU V V V 2550 (CUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU
UGGCCC G CGUCCC V AAACUU I GGUUUC I GGUUUC I CCCCC P F L CCCCCC P CGCUUC P CGCCU C CCCCCC A C CCCCCC C C C CCUCCC V N C CUCCC V N C CUCCC V N AAACU C I C CUCCC V N AAACU C I C C C C C C C C C C C C C C C C C	1810 SAGGAUGU SAGGAUGU SAGGAUGU E D NacGCGGU V 1990 SAGGAGUCU D V 1990 SAGGAGU Q	UIGALUGAGGA UIGALUGAGGA J I E 1910 UUGAGAGGAGCUCGA 2000 UUGAGAGGCUCGAGCUCGA A C S C 2090 UUGCAGCGGAGCUCGAG A V S V 2 S Q 2 Q S Q CGAGUCGLAGA P V Q Q 2 R P M 2340 CGCGUCGAGCGAC P Q Q G R 2 K P Q Q G 2 K P Q Q C 2 K P Q	<pre>N X H 1 1030 CAUCGCAUC T S H 1 1920 CCUUCCGU S F S 2010 ACCUUCGGG S L R - 2100 ACCUUCGGG S L R - 2100 ACCUUCGGG F P S C P S - 2100 ACCUUCGGU F P D Y S Q - 2200 AAUGCAAGU N G M R T V - 2200 AAUGCAGAGU V B V E V F W R T 2200 AAUGCAGCU CAUCA V B V E V E W R C 200 AAUGCACU CAUCA V B V E V E W R C 200 AAUGCACU AUGCAGCU CAUCA V B V E V E W R C 200 AAUGCACU AUGCACU AUGCACU CAUCA V B V E V E W R C 200 AAUGCACU AUGCACU CAUCA V B V E V E W R C 200 AAUGCACU AUGCACU AUGCACU CAUCA V AUGCACU AUGCAU AUGCACU</pre>	$ \begin{array}{cccc} A & E & B & R \\ 1840 \\ CCGURACGACCA \\ P & V & A & P \\ 1930 \\ UCGAUCUGCA \\ CCAUCOGAUCUGCA \\ Z020 \\ UCGAUCUGCACCA \\ A & S & P & P \\ 2110 \\ COLGAUCUGCACA \\ L & B & G \\ 1 & L & K & G \\ 1 & L & K & G \\ 1 & L & K & G \\ 2290 \\ UCUGCAAGUC \\ R & R & H \\ F & D & 1 \\ 2470 \\ UUGCUAAUUCC \\ C & R & H \\ 1 \\ 2470 \\ UUGCUAAUUCC \\ C & C \\ $	$ \begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ 1900 \\ 1000 \\ $	G Q P S 1860 IGALCGCGGCG IGALCGCGGCG IGALCGCGGCG R T R R 1950 2000 2000 100	W A D 1870 CAAACGGUCAA CAAACGGUCAACUU S A 2050 GGACGUCAGCUU T S A 2050 GGACGAUACCUC E N T S 2140 CAACAUGCUG E N A L 2230 ACUUCAAGAAU X 2320 ACUUCAAGAAA T P R 1 2230 ACUUCAAGAAA Y L R 1 L Q E 2500 UGCGGAGAGAU A R D CUUCCUGGGAGAGAU A R D CUUCCUGGGAGAGAU	R F G 1880 1880 1405 1905 1405	D D S 1890 AGCAGUU E Q F 1980 SUUCCUUU C V P L 2070 CCUCCGUU P S V 2160 CCUCCGU CCUCCGU C A S A Q 2250 CCUCCUCU S L I 2340 AGAGUCCC S F K V P 2430 UGAACUCC S E L V N S 2520 UUAGUUU UAGCUUU
UGGCCC G CGUCCG V N GAUUC I GCUCCG P F L CCCCCC P CGCUCC P CGCUCC C CCCCCC C GGUGCG V	1810 1810 1800 1900 P 1900 0 A VAGCGGU 1990 0 CCUCCCU 2080 0 CCUCCCU 2170 0 CCALGCC 2170 0 S K 2250 2280 0 A V K D S K 2250 2250 0 A R C 2250 2350 2 A MAGGEGEGUC 2 A MAGGEGUC 2 A MAGGEGUC 3 S K 2350 2 A MAGGEGUC 3 S S 250 2 A MAGGEGUC 5 S F 2530 2 A MAGGEGUC 5 S S 250 2 A MAGGEGUC 5 S P 2530 0 A MAGGEGUC 5 S P 2530	UIGALUGAGGA UIGALUGAGGA J I E 1910 UUCAGAGGAGCUCAGA 2000 UUCAGAGGCUCAGCAGCA 2000 UUCAGAGGCAGCUCAGAG A V S V 2370 CGCAGUCAGCAGCA P V Q Q 2270 CGCAGUCAGCAGCA P V Q Q 2360 CGCGUCAGCAGCAC P Q G R 1 R P M 2360 CGCGUCAGCAGCAG P Q G R 2 R 2 M 2360 CGCGUCAGCAGCAG P Q G R 2 R 2 M 2360 CGCGUCAGCAGCAG P Q G R 2 R 2 M 2 8 0 2 8 0	N X H 1030 CAUCGCAUC T S H 1920 CCUUCLCGU S 2010 GCCUUCGGG S 2 D A 3 S L R 2 100 GCCUUCGGG GCCUUCGGG 3 C P S 2 190 AUUCCACAGA F 4 N GAACGGUAD A AGAACGGUAD N GAACGGUAGG GAACGGUAG 4 S Q Z450 0UCCCGGGGC C L A Q Y S Q 4 A G H C2 S50 UCCGCGGCGC C C2 S50 CCUCGUAG S 2 2 2 2 C	$ \begin{array}{cccc} A & E & B & R \\ 1840 \\ CCGURACGACCA \\ P & V & A \\ P & 1930 \\ UCGAUCUGCACCA \\ V & CA \\ 2020 \\ UCGAUCUCCGCC \\ A & S & P & P \\ 2110 \\ CUUGGAUGUCA \\ L & B & C \\ 2200 \\ UCUGGAUGUCA \\ L & B & C \\ 2210 \\ UCUGGAUGUCA \\ L & B & C \\ 2230 \\ UCUGGAUGAUC \\ C & P & P \\ 2210 \\ UCUGGAUGAUC \\ C & P & P \\ C & P & P \\ C & C & P \\ C & C & P & P \\ C & C & P & P \\ C & C & P & C \\ C & P & C & C \\ C & C & C & C \\ C & C & C & C$	R K S 1950 1950 1940 10CANAACCC; S I P 1940 10CACAGGGCS H E G 2030 10UGCACGGCC H E G 2030 10UGCACGGCC K S G V N R Y 210 10UGCACGCC F G W I 2300 10UGCACGACC F G W I 2300 10UGCACGACC F G D E F D E F D E F D E 100 100 100 100 100 100 100 10	G Q P S 1860 UGAACGCGCC UGACCACGCCC V T R T R R 1950 UUGUCCCCACC 2040 UUGUCCCACCCACUC 2040 V P F 2040 UGACCUCCACUC CACCCCACUC CACCACUC CACCCACUC CACCCACUC CACCCACUC CACCCACUC CACCACUC CACCACUC CACCACUC CACCACUC CACCACUC CACCACUC CACUC CACUC CACUC CACUC CACUC CACUC CACUC CACUC CACUC CACUC CACUC CACUC CACUC CACUC CACUC CACU	$ \begin{array}{c} \mbox{W} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$ \begin{array}{c} {\bf R} {\ \ F \ \ G} \\ 1880 \\ 1880 \\ 1880 \\ 1880 \\ 1880 \\ 1970 \\ 1970 \\ 100 \\$	D D S 1890 AGCAGUU E Q F 1980 IJ9800 IJ9800 IJ9800 IJ9800 IJ9800 IJ9800 IJ9800 IJ9800 I
UGGCCC G CGUCCG V N GAUUC I GCUCCC P F L CCCCCC C CCCCCC C CCCCCC C CCCCCC C C C	1810	UIGAUNCEAGA J I B2 1910 UIGACAUCAGA 1910 UIGACAUCAGA 2000 UIGACAUCAGA 2000 UIGACAUCAUCA 2000 0000000 0000000 0000000000 000000	N N	$ \begin{array}{cccc} A & E & B & R \\ 1840 \\ CCGURACCACP, \\ P & A & P \\ 1933 \\ UCGAUCUCCCCCC \\ P & E & S & A \\ 2020 \\ UCGAUCUCCCCCC \\ A & S & P & P \\ 2110 \\ CUGAUCUCCCCCCCCC \\ CUGAUCUCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$ \begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ S & I & P \\ 1940 \\ 1000 \\ 10$	G Q P S 1860 UGALCACCGGCC UGALCACCGGCC VIUGUCGCCACACUC UGUCGCCACACUC UJUAUCCACCACUC UJUAUCCACCACUC UJUAUCCACACUC CAUCUUGCAACUC CAUCUUGCAACUC CAUCUUGCAACUC CAUCUUGCAACUC CAUCUUGCAACUC CAUCUUGCAACUC CAUCUUGCAACUC CAUCUCAACUC CAUCUUGCAACUC CAUCUCAACUC CAUCUCAACUC CAUCUUGCAACUC CAUCUCAACUC CAUCUCAACUC CAUCUCAACUC CAUCUCAACUC CAUCUCAACUC CAUCUCAACUC CAUCUCAACUC CAUCUUCAACUC CAUCUCAACUC CAUCUUCAUCUUC CAUCUUCAUCUUC CAUCUUCAUCUUC CAUCUUCAUCUUC CAUCUUCAUCUUC CAUCUUCAUCUUC CAUCUUCAUCUUC CAUCUUCAUCUUC CAUCUUCUUCUUCUUCUUCUUCUUCUUCUUCUUCUUCUUC	W A D 1870 CAAACGGUCAI K R S 1960 GGACGHACGGUCA T S A 2050 GGACABAUACCC E N T 2140 M L 2230 M L 2320 CUCCUGCGGAA CUCCAGAGAU CUCCUGCGAA A R D 2590 CUCCUGCGAA A R D 2590 CUCCUGCAGAU A R D 2590 CUCCUGCAGAU	R F G 1880 1880 1970 1970 1970 1970 1006ACCAC 2050 CUGGGACUC 2150 CUGGGACUC 2150 CUGGGACUC 2150 CUGGGACUC 2240 CCANACUC 2310 CUGGUACUC 2310 COGUCACUC 2420 COGUCACUCAU 2510 16 E 2510 17 E 2510 10 C 2510 10 C 2510 10 C 2420 10 C 2510 10 C 2510 10 C 2510 10 C 2510 10 C 2510 10 C 2500 10 C 2500 10 C 2500 10 C 2500 10 C 2500 10 C 2500 10 C 2500 10 C 2500 10 C 2600 10 C 2600 10 C 2600 10 C 2500 10 C 2600 10 C 20 C 2	$ \begin{array}{c} {}^{\rm D} {}^{\rm D} {}^{\rm S} {}^{\rm S} {}^{\rm 1890} \\ {}^{\rm 1806} {}^{\rm culoccuvu} \\ {}^{\rm Culocuvu} \\ {}^{\rm Cul$
UGGCCC G CGUCCC V AAACUU I GAUUCC I GAUUCC I CGCUCC C CCCCA V C CCCCA C CCCCA C C CCCCA C C CCCCA C C CCCCA C	1810 1810 1900 1900 1900 1900 1900 1900 0 A 0 A 2250 A 10 A 11 A 12 A 12 A 12 A 12 A 12 A 12 A	UIGAUNACCAGA D I = D 19 I 2000 VICCAGAGGACUCAA 2000 VICCAGAGGACUCAA 2000 VICCAGAGGAAL P T E 2180 CCCAGUCACAACAA P V Q Q 270 CCCUCAGGACAA P V Q Q 270 CCCUCAGGACACAA P V Q Q 3 F N 2360 CCCUCAGGACACAA CCCUCAGGACACAA P Q G Q 2450 CCCUCAGGACCUCAA G G A A G V P 2450 CCUCAGGACCUCAA CCUCAGGACCUCAA CCUCAGGACCUCAA CCUCAGGACCUCAA CCUCAGGACCUCAA CCUCAGGACCUCAA CCUCAGGACCUCAA CCUCAGGACCUCAA CCUCAGACACCUCAA CCUCAGACACCUCAA CCUCAGACACCUCAA CCUCAGACACCUCAA CCUCAGGACCUCAA CCUCAGACACCUCAA CCUCAGACACUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGCACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGUCACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCAGGACCUCACCUCAA CCUCACGUCACCUCACCUCAA CCUCACCUCACCUCACCUCAA CCUCACCUCA	<pre>M 1 1930 1930 CAUCGCAUC T S H 1 1920 CCUUCUCGU S F S 2010 S L R - 2100 GCCUUCGGG S L R - 2100 GCCUUCGGG C P S - 2100 GCCUUCGGG P F S 2230 AUUCCCAAC S AUUCCCAAC S AUUCCCA</pre>	$ \begin{array}{cccc} A & B & B & R \\ 1840 \\ CCCURACCOCR \\ P & A & P \\ 1933 \\ UCGAUCUCCAUCCUCCCCC \\ P & B & A \\ 2020 \\ UCUGAUCUCCCCCCC \\ A & S & P & P \\ 2110 \\ CUCUGAUCUCCCCCCCC \\ A & S & P \\ 2200 \\ UCUUCAUCAUCCCCCCCCC \\ S & P & R \\ C & P & G \\ C & P & G \\ C & S & P \\ 2200 \\ UCUGAUCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$ \begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ 1940 \\ 1026 \\ $	G Q P S 1860 1860 1860 1860 1860 187 T R R 1950 1920 2040	$ \begin{array}{c} {\bf W} ~ {\bf A} ~ {\bf D} \\ 1870 \\ {\bf CAAACGGUCAI} \\ {\bf CAAACGGUCAI} \\ {\bf K} ~ {\bf R} ~ {\bf S} \\ 1960 \\ {\bf GACCUCACGUCAI} \\ {\bf C} \\ $	R F G 1880 1880 1970 1970 1970 1970 2050 2050 2050 2050 2050 2050 2050 2050 2050 2050 2240 36 V V 2330 36 V V 2420 36 V V 2420 36 C V P K R V P 2420 36 C V P K 2420 36 C V P K 2510 16 C V C 250 250 250 250 250 250 250 250	D D S 1890 AGCAGUU AGCAGUU CUCCUUU V P L CUCCUCUU V P L CUCCUCGU V P L COCCUCGU V P L COCUCGUC V C A COCUCGUC V P L COCUCGUC V C A COCUCGUC V P L COCUCGUC V C A COCUCGUCA V P L COCUCGUCA V P L COCUCGUCA COCUCGUCA COCUCGUCA COCUCGUCA COCUCGUCA COCUCGUCA COCUCGUCA COCUCGUCA COCUCGUCA V P L COCUCGUCA COCUCGUCA V P L COCUCGUCA V P L COCUCA V P L V P L
UGGCCC G CGUCCC V AAACU I GAUUCC I GAUUCC I CCCAGCC V V CCAGCC Q	1810 1810 Landariud MacGariud	U 102000 0 1 2 1 19 1 0 10 1 2 1 19 1 0 10000000000 20	$ \begin{array}{cccc} & x & y & y \\ 1030 \\ CAUCEGCAUC \\ CAUCEGCAUC \\ T & S & H \\ 1920 \\ CCUUCUCCU \\ S & T & S \\ 2010 \\ CCUUCCGG \\ S & L & R \\ 2100 \\ CCUUCCGG \\ S & L & R \\ C & P & S \\ 2100 \\ CCUCCGGG \\ C & P & S \\ 2100 \\ CCCCGGGG \\ C & S \\ C & P & S \\ C & S \\ C & P & S \\ C & S \\ C & P & S \\ C & S \\ C & P & S \\ C & S$	A E B R 1840 CCCURACCOCC P V A P 1933 UCCACUCCACC P E S A 2020 CCACUCCCCCC A S P P 2110 CCACGUCCACA Y V G A T L E 2200 UCUUGAGUGAGA Y V G A T L E 2200 UCUUGAGUGAGA S P R C P P G 2380 C S P R C P P G 2380 UCCACCACACA C A N S 2400 UCCACACACCCCG C A 2550 C A N S P R K 2550 CACCACCACACACC C A 2550 C C A 2550 C C A 2550 C C A 2550 C C A 2550 C C A 2550 C C C A C C A 2550 C C C A C C A	$ \begin{array}{l} {\rm R} \ {\rm K} \ {\rm S} \\ 1950 \\ 1950 \\ {\rm S} \ {\rm I} \ {\rm P} \ {\rm P} \ {\rm P} \ {\rm A} \\ {\rm S} \ {\rm I} \ {\rm P} \ {\rm P} \ {\rm A} \\ {\rm I} \ {\rm A} \ {\rm CACAGGEGE} \\ {\rm H} \ {\rm E} \ {\rm G} \\ {\rm CACAGGEGE} \\ {\rm H} \ {\rm E} \ {\rm G} \\ {\rm C} \ {\rm 2030} \\ {\rm C} \ {\rm CACGGEGEC} \\ {\rm H} \ {\rm R} \ {\rm G} \\ {\rm C} \ {\rm 2120} \\ {\rm A} \ {\rm CACGGEGECGEC} \\ {\rm A} \ {\rm R} \ {\rm A} \ {\rm G} \\ {\rm C} \ {\rm 2210} \\ {\rm C} \ {\rm C} \ {\rm C} \ {\rm C} \\ {\rm C} \ {\rm C} \ {\rm C} \ {\rm C} \ {\rm C} \\ {\rm C} \ {\rm C} \ {\rm C} \ {\rm C} \\ {\rm C} \ {\rm C} \ {\rm C} \\ {\rm C} \ {\rm C} \\ {\rm C} \ {\rm C$	G Q P S 1860 1860 1860 1860 1860 187 T R R 1950 1900 2040 2040 2040 2040 2040 2040 2040 2040 2040 18 S S 2130 2130 2210 230 2400 2580 2400 258	W A D 1870 CANACGUCAL CANACGUCAL CANACGUCAL CANACGUCAL T S A 2050 GAACAUGCUCA E H A T 2230 GAACAUGCUCA AU A L 2230 CUCUCAUGCAAGAAU V K Q 2320 CUCUCAUGCAAGAAU CUCUCAGAAGAAU A R D 2590 CUCGUAGUAGAGAU A R D 2590 CUCGUAGUAGAGAU A R D 2590 CUCGUAGUAGAGAU CUCUCAUGCAAGAAU A R D 2590 CUCGUAGUAGAGAU A R D 2590 CUCGUAGUAGAGAU CUCUAGUAGAGAU A R D 2590 CUCGUAGUAGAGAU CUCUAGUAGAGAU CUCUAGUAGAGAU CUCUAGUAGAGAU CUCUAGUAGUAGAGAU CUCUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGU	R F G 1880 1880 1970 1970 1970 1970 1900 10000CCCCC 2060 10000CCCCC 2060 10000CCCC 2060 10000CCC 2060 1000 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 2060 1000 10	D D S 1890 AGCAGUU AGCAGUU CUCCUU UV P L CUCUCUU V P L CUCCUCUU V P L CUCCUCUU V P L CUCCUCUU V P L CO CO CUCCUCUU V C A S A Q 2250 CCCUCAU S A Q 2250 CCCUCAU S A Q 2250 CCCUCAU S A Q 2250 CCCUCAU S A Q 2250 CCUCCAU S A Q 250 CCUCCAU S A Q 250 CCUCAU CUCCAU S A Q 250 CCUCCAU S A Q 250 CCUCAU S A Q 250 CCUCAU S A Q 2700 S CUCAU S A S 2700 S CUCAU S A S 2700 S CUCAU S A S 2700 S CUCAU S A S 2700 S CUCAU S A S S
UGGCCC G CGUCCC N GAUUCC I GAUUCC A U GCUCCC P F CCCCCC C CCCCCC C CCCCCC C Q UUCCCCCCC Q	1810 1810 Landariud 1900 1900 1900 Acccessu 1900 Jakaccessu 1900 Jakaccessu 1990 Jakaccessu 1990 Jakaccessu 1990 Jakaccessu 2080 C Q G C Q G C Q G C Q G C Q G C Q G C Q G C Q G C Q G Q G	UIGAUNAUCCAGA) J I = 1 1910 UUGAUNAUCCAGA) J I = 1 1910 UUGAUNAUCCAGA 2000 CCCCCAGCAUAU CCCCAGCAGAUA A P T E E 2180 CCCCAGCCAUCA CCCCAUCCAUCA A V S V C S 0 CCCCUCAGCAGAU CCCCUCAGCACAU CCCCUCAGCAUCAU CCCUCAGCAUCAU CCCUCAGCAUCAU CCCUCAGCAUCAU CCCUCAGCAUCAU CCCUCAGCAUCAU CCUCAUCAUCAU CCUCAU CCUCAUCAU CCUCAUCAU	$ \begin{array}{ccccc} & x & y & y \\ 1030 \\ cAUCGCAUC \\ cAUCGCAUC \\ T & S & H \\ 1920 \\ cCUUCUCCU \\ S & T & S \\ 2010 \\ CCUUCCGG \\ S & L & R \\ 2190 \\ AUUCCCAGA \\ T & S \\ 2190 \\ AUUCCCAGA \\ CCCUUCCAG \\ T & S \\ 2190 \\ AUUCCAGA \\ CCCUUCCAG \\ T & S \\ 2190 \\ AUUCCAGA \\ CCUUCCAGA \\ T & S \\ 2190 \\ AUUCCAGA \\ CCUUCCAGA \\ T & S \\ CUUCGAGA \\ CCUUCAGA \\ CUUCAGA \\ CCUUCAGA \\ CCUUCAGA \\ CUUCAGA \\ CUUU$	A E B R 1840 CCCURACCOCC P V A P 1930 UCCACUCCCCC P E S A 2020 CCACUCCCCCC A S P P 2110 CCACGUCCCCC A T L E 2200 UCUUGAAGAU V V G A T L E 2200 UCUUGAAGAU 2200 UCUUGAAGAU S P R A T L E 2200 UCUUGACGACAU S P R C P P G 2380 UCCACGACACAU C . L A N S 2470 UCCACGACACAU C . L A N S 2551 C . L A N S 2550 C . L C . L A S 2550 C . L C . L A S 2550 C . L A N S 2550 C . L C .	$ \begin{array}{l} {\rm R} \ {\rm K} \ {\rm S} \\ 1950 \\ 1950 \\ 1950 \\ 1940 \\ 1026 $	G Q P S 1860 1860 1860 1860 1860 1867 1950 1950 2040 20	W A D 1870 CAAACGGUCAI CAAACGGUCAI K R S 1960 GGACCHACGUCA 2050 GGACAAUACCC E H A T 2050 CACUCAGUCA E H A T 2230 CAUCACAGUCAG AU A R 2230 CUCCUCAGGAA AU R D 2590 UGCGAGAGAU A R D 2590 UGCGAGAGAU A R D 2590 UGCGAGAGAU CUCUUCCCAAGAAU A R D 2590 UGCGAGAGAU CUCUUCCCAAGAAU A R D 2590 UGCGAGAGAU CUCUUCCCAAGAAU A R D 2590 UGCGAGAGAU A R D 2590 CUCGUUACAGAU A R D 2770 CUCGUUACAGAU A R D 2770 CUCGUUACAGAU A R D 2770 CUCGUUACAU A R D 2770 CUCGUUACAU CUCGUUACAU A R D 2770 CUCGUUACUUACAU A R D 2770 CUCGUUACUUACAU CUCUUAU CUCUUAU CUCUUAU CUCUUAU CUCUUAU CUCUUAU CU	R F G 1880 1880 1970 1970 1970 1970 1900	D D S 1890 AGCAGUU AGCAGUU CUCCUU V P 1980 CUCUCUU V P 2160 CUCCGGU V P 2160 CUCCGGU CUCCGCCAU S A Q 2250 CCCCCAU S A Q 250 CCCCCCAU S A Q 250 CCCCCCAU CCCCCCU CCCCCU CCCCCCU CCCCCCU CCCCCCU CCCCCU CCCCCU CCCCCU CCCCCU CCCU CCU CCU CCU CCU CCU CCU CCU CCU CCU CCU CCU CCU CCU CCU CCU CCU C
UGGCC G V AAACUU N GAUUC I GCUGUU A V L CCUGU V CCCCA C Q UUUCAA L	1810 AscGALICUANCA E D V 1900 SAGGALICUANCA AscGALICUANCA AscGCCGUU D A D A V 1900 JGCCASGC C Q C CCCUCCCU P S L L 2170 CCALAGAC C C Q A C Q V K D Z250 CCALAGAC CCCUCCCU Z S S Q A C Q A Q A C C C C C C Q A C	$\begin{array}{c} & & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & &$	$ \begin{array}{ccccc} & & & & & & & \\ & & & & & \\ & & & & & $	A E B R 1840 CCCURACCOCC P V A P 1930 UCCACUCCCCC P E S A 2020 CCACUCCCCCC A S P P 2110 CCACGUCCACA Y V G A T L E 2200 UCUUGAAGUCCCCCGG S P R 210 UCUCCCCCGGCAAGUC S P R C P P G 2300 UCCACGUCCGACAUC C . L A N S 2470 UCCACGCCGCAA C 2550 CCCCGCGCCAA C 2550 CCCCGCCCACAUC C . L A N S 2550 C P R K 2550 C P R K 2300 D P R Q 2550 C C . L A N S 2550 C C . L C C . L A N S 2550 C C . L C C . C . L C C . L	$ \begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ 1950 \\ S & I & P \\ 1940 \\ 1626 \\ 1626 \\ 1626 \\ 1210 \\ 100 \\ 1210 \\ 100 \\ 1210 \\ 100 \\ 100 \\ 100 \\ 1210 \\ 10$	G Q P S 1860 1860 1860 1860 1860 187 T R R 1950 2040 1920 2040 1920 2040 1920 2040 1920 2040 1920 2040 1920 2040	W A D 1870 CANACGUCAX CANACGUCAX K R S 1960 GGACGUCACGUCA T S 'A 2050 GGACANACCUCA E H A C 2140 GAACANGCUG E H A C 2230 AGUAAAGCAG X K A S CUCCUCGGAA AGUAAAGCAG X K A S CUCCUCGGAA AGUAAAGCAG CUCCUCGGAA A R D 2590 CUCGUUCGGAAGAGAU A R D 2590 CUCGUUCGGAAGAGU A R D 2590 CUCGUUCGGAAGAGU A R D 2590 CUCGUUCGGAAGAGU A S C 2500 CUCGUUCGGAAGAGU A S C 2500 CUCGUUCGGAGAGUUCGGA A S C 270 CUCGUUCGGAGAGUUCGGA A S C 270 CUCGUUCGGAGAGUUCGGA A S C 270 CUCGUUCGGAGAGUUCGGA A S C 270 CUCGUUCGUUCGGAAGAGUUCGGAGUUCGGAGUUCGGA A S C 2800 CUCGUUCGUUCGUUCGGAGAGUUCGGAGUUCGGAGUUCGGAGUUCGGAGUUCGGUUCGUUC	$ \begin{array}{c} {\rm R} {\rm \ F} {\rm \ G} \\ {\rm \ 1880} \\ {\rm \ 1880} \\ {\rm \ 1880} \\ {\rm \ caseline} \\ {\rm \ 1880} \\ {\rm \ 188$	D D S 1890 AGCAGUU AGCAGUU D P 1980
UGGCCC G CCUCCC V AAACUU N GAUUCC I GCUCUU A V L CCUCCCA V CCAGCC V UUGCCC	1810 1810 AGGALIGUES D Y 1900 AGGACGEGUE D J 1900 J AGCGAGUE J 1900 J AGCGAGUE C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C S C S S S C A R L Z C C C C C	U GAUNAUCGAUGA) J I = 1 1910 UUGAUNAUCGAUGA) J I = 1 1910 UUGAUNAUCGAUGAUGAU) S E C 2000 UUGAUGAUGGAUGAUUAGU A G S C 2180 GCGAUUCUAGAUCGAU 2270 CCGGUUCAGAUGGAU CCUUAGAUCGGU) N A L 2540 CCUUAGAUCGGU) N A L 2540 CCUUAGAUCGGU) N A L 2540 CCUUAGAUCGGU) N A L 2540 CCUUAGAUCGGU) N A L 2700 CCUUAGAUCGGU	N N N N N 1030 CAUCEGCAUC T S N 11920 CCUUCUCCU S S N N 10100 S D S L R 1000 GCCUUCCGG S L R Y S Q 2100 GCCCUUCAGG GCCCUCACA F P D AUUCCCACA Y S Q 2370 AGUGCAGGAGGU N G W Y Z Y S Q 2370 AGUGCAGGGCC F M G 2550 CCUCEGGCCUC P H A G 2730 UCGGGGCCUC F G P 2420 UGGUGUGGGG 2420 UGGUGUGUGGG 2420 UGGUGUGUGGG 2210 UUGGUGUGGGGUGG 2210 X Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y <td>A E B R 1840 CCCURACCOCC P V A P 1933 UCCAQUUCCGCC A 210 CCAQUUCCGCC A 2 10 CCAQUUCCGCC A 7 L E 2200 UCUUGAAGGUUGAA A T L E 2200 UCUUGAAGGUUGAAGGAU L E G S P R C P P G 2380 UCUCCCCGACACUC C A 1 N S 2551 C A N S 2550 C A 1 N S 2550 C C C C C C C C C C C C C C C C C C C</td> <td>$\begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ 1950 \\ S & I & P \\ 1940 \\ 1620 \\ 16$</td> <td>G Q P S 1860 UGALCOCCGCCC UGALCOCCGCCC R T R R 1950 UUGUGCCCACUC 2040 UUGUGCCACUC 2040 UUGUGCCACUC 2130 2130 1 Y F E 2130 2130 1 Z F G E 2130 2130 1 Z F G E 2240 UCGALGCACUC 2 S S P I 2 CAUCUCCUCC 2 S S P I 2 400 UUCCAAGGGCCU V R G L 2 S C R R P R E 2 670 UUCCAAGGGCCU V R G L 2 580 UUCCAAGGGCCU V R G L 2 580 UUCCAAGGCCCACCC 2 60 UUCCAAGGCCCACCC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCU UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGU 2 80 UUCCAAGGCU 2 80 UUCCAAGGU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGU 2 80 UUCCAAGGU 2</td> <td>W A D 1870 1870 CAAACGGUCAX K R S 1960 GGACGUCAGCUG T S 'A 2050 GGACAUACCUG E N T 2140 CAACCUGUG E H A C 2230 CAUCCAAGAAUACCUG E H A C 2230 AGUAAAGCAG S K A I V K Q 2320 AGUCAAGAAUACCUG CUCUCGGAAGAUU V K Q 2320 CUCUCGGAAGAUU CUCUCGGAAGAUU A R D 2590 CUCUCGGAGAGAUU A R D 2590 CUCUCGGAGAGAUU CUCUCGGAGAGUU A F D 2590 CUCUCGCGUAGAGUU L I S 2770 CUCUCGGGAGGCUUCGCUAGAGUU V F S 2860 V F S 2850 VICUCAGGAUACCUGUGAGAUU</td> <td>R F G 1880 1880 1880 1880 1970 1970 1970 2050</td> <td>$\begin{array}{cccc} {\rm P} & {\rm D} & {\rm S} & {\rm S} \\ 1890 \\ {\rm AGCAGUU} \\ {\rm E} & {\rm Q} & {\rm F} \\ 1980 \\ {\rm SUUCUUU} \\ {\rm V} & {\rm P} & {\rm L} \\ 2070 \\ {\rm COUCGGUUU} \\ {\rm V} & {\rm P} & {\rm L} \\ 2070 \\ {\rm COUCGGUUU} \\ {\rm P} & {\rm S} & {\rm V} \\ 2160 \\ {\rm COUCGGUUU} \\ {\rm P} & {\rm S} & {\rm V} \\ 2160 \\ {\rm COUCGUUU} \\ {\rm P} & {\rm S} & {\rm V} \\ 2160 \\ {\rm COUCGUUU} \\ {\rm COUCGUUU} \\ {\rm S} & {\rm V} \\ 2250 \\ {\rm COUCCUU} \\ {\rm COUCGUUU} \\ {\rm AGUUCCC} \\ {\rm C} & {\rm P} \\ {\rm H} \\ {\rm S} & {\rm L} & {\rm I} \\ 2340 \\ {\rm COUCGUUU} \\ {\rm AGUUCCC} \\ {\rm C} & {\rm P} \\ {\rm S} & {\rm L} \\ {\rm I} \\ {\rm I} \\ {\rm C} \\ {\rm S} & {\rm V} \\ {\rm P} \\ {\rm S} \\ {\rm L} & {\rm I} \\ {\rm I} \\ {\rm C} \\ {\rm S} \\ {\rm COUCUUU} \\ {\rm I} \\ {\rm COUCUUU} \\ {\rm I} \\ {\rm C} \\ {\rm COUCUUU} \\ {\rm I} \\ {\rm C} \\ {\rm COUCUUU} \\ {\rm I} \\ {\rm I} \\ {\rm COUCUUU} \\ {\rm I} \\$</td>	A E B R 1840 CCCURACCOCC P V A P 1933 UCCAQUUCCGCC A 210 CCAQUUCCGCC A 2 10 CCAQUUCCGCC A 7 L E 2200 UCUUGAAGGUUGAA A T L E 2200 UCUUGAAGGUUGAAGGAU L E G S P R C P P G 2380 UCUCCCCGACACUC C A 1 N S 2551 C A N S 2550 C A 1 N S 2550 C C C C C C C C C C C C C C C C C C C	$ \begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ 1950 \\ S & I & P \\ 1940 \\ 1620 \\ 16$	G Q P S 1860 UGALCOCCGCCC UGALCOCCGCCC R T R R 1950 UUGUGCCCACUC 2040 UUGUGCCACUC 2040 UUGUGCCACUC 2130 2130 1 Y F E 2130 2130 1 Z F G E 2130 2130 1 Z F G E 2240 UCGALGCACUC 2 S S P I 2 CAUCUCCUCC 2 S S P I 2 400 UUCCAAGGGCCU V R G L 2 S C R R P R E 2 670 UUCCAAGGGCCU V R G L 2 580 UUCCAAGGGCCU V R G L 2 580 UUCCAAGGCCCACCC 2 60 UUCCAAGGCCCACCC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCCAUC 2 80 UUCCAAGGCCU UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCCU 2 80 UUCCAAGGCU 2 80 UUCCAAGU 2 80 UUCCAAGGCU 2 80 UUCCAAGGU 2 80 UUCCAAGGCU 2 80 UUCCAAGGCU 2 80 UUCCAAGGU 2	W A D 1870 1870 CAAACGGUCAX K R S 1960 GGACGUCAGCUG T S 'A 2050 GGACAUACCUG E N T 2140 CAACCUGUG E H A C 2230 CAUCCAAGAAUACCUG E H A C 2230 AGUAAAGCAG S K A I V K Q 2320 AGUCAAGAAUACCUG CUCUCGGAAGAUU V K Q 2320 CUCUCGGAAGAUU CUCUCGGAAGAUU A R D 2590 CUCUCGGAGAGAUU A R D 2590 CUCUCGGAGAGAUU CUCUCGGAGAGUU A F D 2590 CUCUCGCGUAGAGUU L I S 2770 CUCUCGGGAGGCUUCGCUAGAGUU V F S 2860 V F S 2850 VICUCAGGAUACCUGUGAGAUU	R F G 1880 1880 1880 1880 1970 1970 1970 2050	$ \begin{array}{cccc} {\rm P} & {\rm D} & {\rm S} & {\rm S} \\ 1890 \\ {\rm AGCAGUU} \\ {\rm E} & {\rm Q} & {\rm F} \\ 1980 \\ {\rm SUUCUUU} \\ {\rm V} & {\rm P} & {\rm L} \\ 2070 \\ {\rm COUCGGUUU} \\ {\rm V} & {\rm P} & {\rm L} \\ 2070 \\ {\rm COUCGGUUU} \\ {\rm P} & {\rm S} & {\rm V} \\ 2160 \\ {\rm COUCGGUUU} \\ {\rm P} & {\rm S} & {\rm V} \\ 2160 \\ {\rm COUCGUUU} \\ {\rm P} & {\rm S} & {\rm V} \\ 2160 \\ {\rm COUCGUUU} \\ {\rm COUCGUUU} \\ {\rm S} & {\rm V} \\ 2250 \\ {\rm COUCCUU} \\ {\rm COUCGUUU} \\ {\rm AGUUCCC} \\ {\rm C} & {\rm P} \\ {\rm H} \\ {\rm S} & {\rm L} & {\rm I} \\ 2340 \\ {\rm COUCGUUU} \\ {\rm AGUUCCC} \\ {\rm C} & {\rm P} \\ {\rm S} & {\rm L} \\ {\rm I} \\ {\rm I} \\ {\rm C} \\ {\rm S} & {\rm V} \\ {\rm P} \\ {\rm S} \\ {\rm L} & {\rm I} \\ {\rm I} \\ {\rm C} \\ {\rm S} \\ {\rm COUCUUU} \\ {\rm I} \\ {\rm COUCUUU} \\ {\rm I} \\ {\rm C} \\ {\rm COUCUUU} \\ {\rm I} \\ {\rm C} \\ {\rm COUCUUU} \\ {\rm I} \\ {\rm I} \\ {\rm COUCUUU} \\ {\rm I} \\$
UGGCCC G CCUCC V ANACU I GAUUC I GCUCC P F CCCCC V CCCCC V CCCCC Q UUCCC Q UUCCC Q UUCCC Q UUCCC	1810 1810 1800 Accounce 1900 Accounce 1900 Accounce 1910 Accounce 1910 Accounce 1910 Accounce 2010 P 1910 C C S S	□ A B 10 I I 10 I I I 10 I I I I 10 I I I I I 10 I </td <td>N N N N N N 1830 1830 CAUCEGAUC T N N 1920 CCUUCUCCU S S N N 2010 GCUUCUCEGE S S S N N 2120 GCCUUCCACA S Y S Q P S Z N G MUCCCACA S Y S Q 2370 ACUGCAGAGAGU Z S Y S Q Z Z S V E W R Q Z Z S V E W Z Z S V Z X G Q Z</td> <td>$\begin{array}{c} A & B & B & R \\ 1840 \\ CCCURACCUCCA \\ P & P & A & P \\ 1930 \\ UCCARUCUCCA \\ P & P & A & P \\ 2020 \\ CCARUCCACCUCACA \\ A & 2 & P \\ 2100 \\ CURACUUCACACUCACA \\ A & P & P \\ 2200 \\ CUCUCACACACACA \\ V & V & Q \\ 2200 \\ CUCUCACACACACACACACA \\ V & V & Q \\ 2200 \\ CUCUCACACACACACACACACACACACACACACACACAC$</td> <td>R K S 1850 1950 1950 1940 10000 10000 10000 10000 10000 1000000 1000000 1000000 100000 100000 1000000 100000 100000 100000 100000 100000 100000 100000 100000 100000 1000000 1000000 100000000</td> <td>G Q P S 1860 UGALCOCCGGCC UGALCOCCGGCC R T R R 1950 UUGUGCCACGUC 2040 UUGUGCCACUC L S N S 2130 UCALCUUUGCA CAUCUUUGCA CAUCUUGCA CAUCUUGCA CAUCUUGCA CAUCCAAGGCC N V C S R N R F R 2400 UUCAAAGGCCA CAUCCAAGGCC V R G L 2400 V R G L 2580 V R G L 2570 CAUCUUGCUGCCCACC 2570 CAUCCAAGGCCA V R G L 2570 CAUCUGCUGCCCACC 2850 CAUCCAAGGCCA V R G L 2840 ACACCCAAGGCA V N C 2590 CAUCUGCUGCCCACC 2850 CAUCUCGUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC CAUCUGCUGCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUCCACCACC 2850 CAUCUGCUCCACCCACC 2850 CAUCUCCACCCACCCACC 2850 CAUCUGCUGCUCCACCCACC 2850 CAUCUGCUGCUCCACCCACCC 2850 CAUCUCCACCCACCCACCCACC 2850 CAUCUCCACCCACCCACCCACCCACCCACCCACCCACCC</td> <td>W A D 1870 1870 GAAACGGUCAA K R S 1960 GACCUCAGCGUCAA 2050 GGACAAUACCC E N T 2050 C 2140 GGACAUACCUG E N A 2140 CAACAUACCUG E N A 2230 AGUAAAGACAG S K A 1 V K Q 2320 AGUCAAGAAA C C 2 2410 V K Q 2320 CCUCUGGGAAGAUU A R D CCUCUGGGAGAGAUU A R D 2590 CUGGGACGAGAUU A R D 2590 CUGGGCCUGGGAGAUU A R D 2590 CUGGGCCUGGGAGAUU A R D 2590 CUGGGCCUGGAGAGAU A R D 2590 CUGGGCCUGGAGAGAU A R C 2590 CUUGGAGAGAUU A R C 2590 CUUGGAGAGAUU A R C 2590 CUUGGGAGAGAUU A R C 2590 CUUGGGCCUGGAGAGAU A R C 2590 CUUGGGCCUGGAGAGAU A R C 2590 CUUGGGCCUGGAGAGAU A R C 2590 CUUGGGCCUGGAGAGAU A R C 2590 CUUGGAGAGAUU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAGAU CUUGGAGAGAGA CUUGGAGAGAGAU CUUGGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGA CUUGGAGAGAGA C</td> <td>R F G 1880 1880 1980 1970 1970 1970 1900 2060 2060 2060 2050</td> <td>$\begin{array}{c} {}^{\rm D} {\rm D} {\rm S} {\rm S} \\ 1890 \\ {}^{\rm A3GCAGUU} \\ {}^{\rm A3GCAGUU} \\ {}^{\rm D} {\rm E} {\rm Q} {\rm P} {\rm F} \\ {}^{\rm D} {\rm Q} {\rm Q} \\ {}^{\rm Q} Q}$</td>	N N N N N N 1830 1830 CAUCEGAUC T N N 1920 CCUUCUCCU S S N N 2010 GCUUCUCEGE S S S N N 2120 GCCUUCCACA S Y S Q P S Z N G MUCCCACA S Y S Q 2370 ACUGCAGAGAGU Z S Y S Q Z Z S V E W R Q Z Z S V E W Z Z S V Z X G Q Z	$ \begin{array}{c} A & B & B & R \\ 1840 \\ CCCURACCUCCA \\ P & P & A & P \\ 1930 \\ UCCARUCUCCA \\ P & P & A & P \\ 2020 \\ CCARUCCACCUCACA \\ A & 2 & P \\ 2100 \\ CURACUUCACACUCACA \\ A & P & P \\ 2200 \\ CUCUCACACACACA \\ V & V & Q \\ 2200 \\ CUCUCACACACACACACACA \\ V & V & Q \\ 2200 \\ CUCUCACACACACACACACACACACACACACACACACAC$	R K S 1850 1950 1950 1940 10000 10000 10000 10000 10000 1000000 1000000 1000000 100000 100000 1000000 100000 100000 100000 100000 100000 100000 100000 100000 100000 1000000 1000000 100000000	G Q P S 1860 UGALCOCCGGCC UGALCOCCGGCC R T R R 1950 UUGUGCCACGUC 2040 UUGUGCCACUC L S N S 2130 UCALCUUUGCA CAUCUUUGCA CAUCUUGCA CAUCUUGCA CAUCUUGCA CAUCCAAGGCC N V C S R N R F R 2400 UUCAAAGGCCA CAUCCAAGGCC V R G L 2400 V R G L 2580 V R G L 2570 CAUCUUGCUGCCCACC 2570 CAUCCAAGGCCA V R G L 2570 CAUCUGCUGCCCACC 2850 CAUCCAAGGCCA V R G L 2840 ACACCCAAGGCA V N C 2590 CAUCUGCUGCCCACC 2850 CAUCUCGUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC CAUCUGCUGCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUGCCCACC 2850 CAUCUGCUCCACCACC 2850 CAUCUGCUCCACCCACC 2850 CAUCUCCACCCACCCACC 2850 CAUCUGCUGCUCCACCCACC 2850 CAUCUGCUGCUCCACCCACCC 2850 CAUCUCCACCCACCCACCCACC 2850 CAUCUCCACCCACCCACCCACCCACCCACCCACCCACCC	W A D 1870 1870 GAAACGGUCAA K R S 1960 GACCUCAGCGUCAA 2050 GGACAAUACCC E N T 2050 C 2140 GGACAUACCUG E N A 2140 CAACAUACCUG E N A 2230 AGUAAAGACAG S K A 1 V K Q 2320 AGUCAAGAAA C C 2 2410 V K Q 2320 CCUCUGGGAAGAUU A R D CCUCUGGGAGAGAUU A R D 2590 CUGGGACGAGAUU A R D 2590 CUGGGCCUGGGAGAUU A R D 2590 CUGGGCCUGGGAGAUU A R D 2590 CUGGGCCUGGAGAGAU A R D 2590 CUGGGCCUGGAGAGAU A R C 2590 CUUGGAGAGAUU A R C 2590 CUUGGAGAGAUU A R C 2590 CUUGGGAGAGAUU A R C 2590 CUUGGGCCUGGAGAGAU A R C 2590 CUUGGGCCUGGAGAGAU A R C 2590 CUUGGGCCUGGAGAGAU A R C 2590 CUUGGGCCUGGAGAGAU A R C 2590 CUUGGAGAGAUU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAU CUUGGAGAGAGAU CUUGGAGAGAGA CUUGGAGAGAGAU CUUGGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGA CUUGGAGAGAGAGA CUUGGAGAGAGA C	R F G 1880 1880 1980 1970 1970 1970 1900 2060 2060 2060 2050	$ \begin{array}{c} {}^{\rm D} {\rm D} {\rm S} {\rm S} \\ 1890 \\ {}^{\rm A3GCAGUU} \\ {}^{\rm A3GCAGUU} \\ {}^{\rm D} {\rm E} {\rm Q} {\rm P} {\rm F} \\ {}^{\rm D} {\rm Q} {\rm Q} \\ {}^{\rm Q} Q}$
UGGCCC G CCUCCC N GAUUC I GCUCCC P I CCCCCC P P CCCCCC C CCCCC C CCCCCC C CCCCCC	1810 1810 1800 3AGGAUGUE 200 3AGGAUGUE 1900 3AGGAUGUE 1900 3AGGAUGUE 1900 3AGGAUGUE 1930 3AGGAUGUE 2080 2080 20170 2170 21717 2260 2350 2440 240 2620 2440 2872 2530 2620 2800 2800 L V E 2800 2800 CCCCCUGGUE 2800 M Y F 2800 2800 Sequedue 2800 Sequedue 2800 Sequedue 2800	UIGALUAUGAGAGA J I E J I E Z I E CCGAUGAGAGCUCA N S V Z I E S C Z 090 J UCCGAUGACACA N S V Z I E S C Z 0 S Q CCGAUGCACACA N V S V Z I E S C Z 2000 J UCCGAUGACACAC P V Q Q S R F N Z 180 J CCGAUGACACAC P V Q Q S R F V Q Z 360 J CCGAUGACACACA N A L Z 530 J CCGUAGAGACCCU CCGAUGACACACA N A L Z 540 J CCGAUGACACACA N A L Z 500 J CCGUAGAGACCCU R M L Z 200 J UCCAGGAUGCCU R M L Z 200 J UCCAGGAUGCCU R D M L Z 200 J CCGUAGACACCCU R D M L Z 200 J CCGUAGACCCU R D M L Z 200 J CCGUAGACCCU R M L Z 200 J CCGUAGACCU R M L Z 200 J CCGUAGACCCU R M L Z 200 J CCGUAGACCU R M L Z 200 J CCGUACCUAGACCU R M L Z 200 J CCGUACCU R M L	R X H 1830 CAUCECCAUC 1920 CCUUCUCCU CCUUCUCCU S 2010 GCCUUCCAGE S L 2100 GCCUUCCAGE 2100 GCCUUCCAGE 7 S 7 S 7 S 7 S 7 S 7 S 7 S 7 S 8 C 9 S 0 GCCUCCACA 8 R 7 S 0 GAACGAGUAU 2370 UCGCAGCCU C Y 2460 CCUCGAUUG A S 2730 UUGGUCUGGE 1000 UCGACCUC 8 Y 2910 AGAUUGUCGUCGE 1000 UGUCAUUGUCUCGUC 1000 UGUCAUUGUCUCUCH 100	A = B = R 1840 CCCURACCCC 1840 1940 P V A P 1930 UCCAGUCUCGCC CA 2210 CCAGUCGCCCCCC A 210 CUCUGAAGUCCCCCCCC A 7 L 2 2200 CCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	R K S 1850 1950 1950 1940 1000 1000GACGGCC L D G 2030 1000GACGGCC L D G 2030 1000GACGGCC L D G 2030 1000GACGGCGCA N R R V 2210 1000GACGGCGCA N R A G 2390 1000GACGGCGCA N R A G 2390 1000GACGGCGCA N G E 2390 1000GACGGCGCA N G E 2390 1000GACGGCGCA 1 N W 2570 1 S C S S S S S S S S S S S S S S S S S	G Q P S 1860 1860 1860 1860 1860 187 T R R 1950 2040 2040 1900 2040	W A D 1870 1870 GAAACGGUCAA K R S 1960 GGACGUCAGGUCAA 2050 GGACAAUACCUCA E N T 2140 GGACAAUACCUCA E N T 2140 E N A L 2230 AGUACAAGAAU V K Q 2320 AGUACAAGAAU P L R 1 2410 V K Q 2320 CUCUGGGAAGAAU P L R D 2590 UGCGACGUCGGAA A R D 2680 GGUGGCUUGCG W P S 2770 GGUGGCUUGCAAGAAU UUCAGGGAUUCCA 0 G Q 3040 CUCUCAGGAAUUCCA 0 G Z 2850 0 GGUGGCUUCCA 0 G Z 2950 UUUCAGGAAUUCCA	R F G 1880 1880 1980 1970 1970 2060 2060 2060 2150 30060 2150 30060 2150 30060 2240 3050 30	$ \begin{array}{cccc} {\rm P} & {\rm D} & {\rm S} & {\rm 1880} \\ {\rm 1800}{\rm Cuude} \\ {\rm V} & {\rm P} & {\rm L} \\ {\rm 2070} \\ {\rm Scouccuur} \\ {\rm V} & {\rm P} & {\rm L} \\ {\rm 2070} \\ {\rm Scouccuur} \\ {\rm V} & {\rm P} & {\rm L} \\ {\rm 2160} \\ {\rm Cuuccecca} \\ {\rm V} & {\rm P} & {\rm L} \\ {\rm 2250} \\ {\rm Cuuccua} \\ {\rm V} & {\rm R} & {\rm S} \\ {\rm A} & {\rm Q} \\ {\rm Cuuccua} \\ {\rm Cuucu} \\ {\rm Cuucu} \\ {\rm Cuucu} \\ {\rm Cuucu} \\ {\rm Cuucuu} \\ {\rm Cuucu} \\ {\rm Cuucuu} \\ {\rm Cuucu} \\ {\rm Cuucuu} \\ {\rm Cuucu$
UGGCC G CCUCC V AAACU I GAUUC P CCUCC P CCCCA V CCCCA V CCCCA V CCCCA V CCCA CCUCC V CCCA CCUCC V UUUCA L	1810 1810 1800 SacGaucu 2800 SacGaucu 1900 SacGaucu 1900 SacGaucu 1900 SacGaucu 1900 SacGaucu 1910 SacGaucu 1920 SacGaucu 1930 SacGaucu 2080 C Q G C Q G 2170 S K T 22170 S K T 22170 Q A G 2350 Q A G 2400 S S P 2530 S S P 2500 S S P 2800 CcccucaG 12 920 S S P 2800 S R T 2800 S R	UIGALUAUCGAGA J I B2 J I B2 J I B2 J I B2 J I B2 CCACGAGCUCAA N P T E B 2000 M P T E B 2000 M CCCCACGCACCA N A V S V 2 S O 2 S	<pre>N 1 8 1 1930 CAUCCCAUC T S H 1 1920 CCUUCUCCU S F S 2010 ACCUUCCGG S L R - 2100 ACCUUCCGG C P S - 2100 ACCUUCGGG C P S - 2190 AUUCCCACA F P D 2220 AUUCCCACA N G M N G M R T V 2300 AUUCCCACAC R G P 1 2460 V E V R 2400 V E V 2640 V E V 2640 V C C 2550 CCUCGGUCU R G P 2250 CCUCGGCUC R G P 2250 CCUCGGUCU R C P 2240 V C V 2640 V C V 200 V C V 20 V C V 2 V C V 2 V C V 2 V C V 2 V C V C V C V C V C V C V C V C V C V C</pre>	$ \begin{array}{cccc} A & E & B & R \\ 1840 \\ CCGURACGACCA \\ P & A & P \\ 1930 \\ UCGAUCUGCACCUCCA \\ P & S & A \\ 2020 \\ UCGAUCUCCACCUCA \\ A & S & P & P \\ 2110 \\ CUUCGAUCUCAC \\ CUUCGAUCACUCA \\ B & A & P \\ 2220 \\ UCUUCAACGUCA \\ C & P & P \\ 2220 \\ UCUCCACACUCA \\ R & H \\ P & D \\ 2230 \\ UCCCACCUCA \\ R & R \\ P \\ C & P \\ C & C \\ $	R K S 1850 1950 1950 1940 100000000 100000000000 10000000000	G Q P S 1860 UGALCOCCGCCC UGALCOCCGCCC UGALCOCCGCCC I V P I 2040 JULUCCALCU L S N S 2130 UCCALCUUGCA CCAUCUUGCA CCAUCUUGCA CCAUCUUGCA CCAUCUUGCA CCAUCUUGCA CCAUCUUGCA CCAUCUCA CCAUCA CCAUCUCA CCAUCA CC	W A D 1870 1870 GAAACGGUCAA K R S 2050 GGACGUCAGGUCAA 2050 GGACAUAGCUCA E N T 2140 GGACAUAGCUCA E N T 2140 E N A L 2230 GGACAUGUCAGA AUXAGCAG X K A S 2140 E N T 2140 E N T 2140 C L Y C 2230 CUCUCAGAAAAGCAG X K A S 2320 CUCUCGGAAAAGCAG 2320 CUCUCGGAAAAGCAG 2320 CUCUCGGAAAAGCAG 2250 CUCUCGGAAAAGCAG L V E 2680 CUCGGCAGAAUUUCCA L V E 2680 CUCGGCGCUGGAAAAGAGU V F S 2770 CUCUCGGAAAAAGAG UUUUCAGGAA V F S 2850 UUUUCAGGAAA	R F G 1880 1880 1970 1970 1970 1970 19060 2060 2060 2150 30060 2150 30060 2150 3007 3007 3007 3007 3007 3007 3007 3007 3007 3007 3007 3007 3007 3140 3007 3140 3007 3007 3007 3007 3007 3007 3007 3007 3007 3140 3007 3007 3007 3007 3007 3140 3007 3007 3007 3007 3140 3007 3140 31	D D S 1890 AGCAGUU AGCAGUU E Q F 1980 SUUCCUUGU V P L 2070 SCUCCGGU CCUCGGU CCUCGGU CCUCGGU CCUCGGU CCUCGU C A 2250 CCCUCGU C A 23400 AGCAGUCC C B L 23400 AGGACUCC C D F 24300 CCCCCCU C A 2430 CCCCCCU C A 2430 CCCCCCU C A 2430 CCCCCCU C A 2430 CCCCCCU C D F 2700 CCAC C D F 2700 CCCCCCU C D F 2700 CCCCCCU C D F 2700 CCCCCCC D F 2700 CCCCCCCC D F 2700 CCCCCCCCC D F 2700 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
UGGCC G CCUCC V AAACU I GAUUC I GCUGU I CCCCA V CCCCA V CCCCA V CCCCA V CCCCA V CCCCA V CCCCA V UUUCA I CCCCCA V V CCCA CCUCCA V I CCUCCA V I CCUCCA V I CCUCCA V I CCUCCA I CCUCCA V I CCUCCA V I CCUCCA I CCUCCA I CCUCCA V CCUCCA V CCUCCA V CCUCCA V CCUCCA V CCUCCA V CCUCCA V CCUCCA V CCUCCA V CCUCCA V CCUCCA V CCUCCA V C CCUCCA V C CCUCCA V I C CCUCCA V C CCUCCA V C CCUCCA V C CCUCCA V V C CCUCCA V V C CCUCCA V V C CCUCCA V V C CCUCCA V V C CCUCCA V V C CCUCCA V V C CCUCCA V V C CCUCCA V V C CCUCCA V V C CCUCCA V V C CCUCCA V C C CCUCA V V C CCUCCA V C C CCUCA V C C CCUCA V C C CCUCA V C C CCUCA V C C CCUCA V C C CCUCA C C C C C C C C C C C C C C C	1810 1810 200 Account 1900 Account 1910 Account 20170 C Q G C 20170 S K T 22350 Q A G 2350 Q A G 2400 Q A G 24200 Q A G 2520 Q A G 2520 Account	UUGAUAUCGAGA J I B2 J I B2 J I B2 J I B2 J I B2 CCCACCGAUCA P T E C 2000 M P T E C 2000 M CCCACCGAUCACACA P V Q Q 2300 CCCCACCGAUCACACAC P Q G R I R P M 2360 CCCUCACGAUCACACAC CCCACCCACACACACACACACACACACACAC	N N 1030 1030 CAUCCCAUC S 1920 CCUUCLCCU CCUUCCCGG S S L 2010 GCCUUCCGG S L 2100 GCCUUCCGG GCCUUCCGG S 2190 AUUCCCACAC AUUCCCACAC S Y Q AUUCCCACAC GAACGGUAD N G M T AUUCCCACAC R V S Q S Q ACGGAGGUA N CCCCC R Q S Q L A G L A G L S Q L A G L A G L A G L A G L A G L A G L A G L A G	A E B R 1840 CCGUAGCACCA 1940 CCGUAGCACCA P V A P 1930 UCGAGUCGCC F E S A 2020 CCACGUCGCC A S P P 2110 CCUGAGUGACA L B G 1 L K D 2220 UCUGAAGUGACA K P P 2210 UCUGAAGACACA C P P G 2310 UCUGACAGACA R R H F D 1 2220 UCUGACAGACA C P P G 2310 UCGACGACACA C P F C D 1 2256 CACGGAGAAAUUCC C P F G 2310 CACACACACCA C P R R K 2256 CACGGAGGAAA P R R K 2256 CACGGAGGAAA P R R K 2256 CACGGAGGAAA P R R K 2256 CACGGAGGAAA P R R K 2256 CACGGAGGAAAA P R R K 2256 CACGGAGGAGAAA P R R K 2256 CACGGAGGAAAACACCA C N N T 2250 CACGACCCUCCCC D V F 2250 CACGACCCUCCCC D V F 2310 UUUCCAACUCUCC C N R R K 2550 CACGGACGACAACACCAC C N R T 2310 UUUUCCAACUCUCCCUCCCC S D L A 3010 UUUUCCACCCUCCCCCCCCCCC V F Q L 3100	R K S 1950 1950 1950 1940 100000 100000 100000 10000 10000 10000 100	G Q P S 1860 UGALCACCGGCC UGALCACCGGCC V R T R R 1950 UUGUCCCCACUC 2040 VUGUCCCACUC 2040 V P E 2120 V P E 2220 CACUCACUC S S L L S CACUCACUCACU CACUCACUCACU CACUCACUCACU CACUCACUCACU CACUCACUCACU V R G L 2400 SUCGAAAGAL V R G L 2400 SUCGAAGAGACU V R G L 2400 SUCGAAGAGACU V R G L 2500 CACUCACGACU V R G L 2500 CACUCACUCACU CACUCACUCACU CACUCACUACUACU CACUCACUACUACU CACUCACUACUACU CACUCACUACUACU CACUCACUACUACU CACUCACUACUACU CACUCACUACUACU CACUCACUACUACU CACUCACUACUACU CACUCACUACUACU CACUCACUACUACUACU CACUCACUACUACUACUACUACUACUACUACUACUACUAC	$ \begin{array}{c} & \mbox{${\rm W}$ $ $ $ $ $ $ $ $ $ $ $ $ $ $ $ $ $ $$	R F G 1880 1880 1880 1880 1880 1970 1970 10000 2010	$ \begin{array}{ccccccc} {\rm P} & {\rm D} & {\rm S} & {\rm S} \\ 1890 \\ {\rm AGCAGUU} \\ {\rm E} & {\rm Q} & {\rm F} \\ {\rm S} & {\rm Q} & {\rm F} \\ {\rm S} & {\rm Q} & {\rm F} \\ {\rm S} & {\rm Q} & {\rm P} \\ {\rm S} & {\rm C2CUCGUU} \\ {\rm P} & {\rm S} & {\rm V} & {\rm P} \\ {\rm C100CGCC} \\ {\rm C100CGC} \\ {\rm$
UGGCCC G CCUCCC V GAUUCC I GCUCCC A V I CCCCCC P C P CCCCCC P C CCCCCC P C CCCCCC P C CCCCCC	1810 1810 1900 SacGaucu U 2000 SacGaucu U 2010 SacGaucu U 2017 SacGaucu U 2018 SacGaucu U 2019 SacGaucu U 2010 SacGaucu U 2018 SacGau	UUGAUAUCGAGA J I B2 J I B2 J I B2 J I B2 CCCACCGAUCAA P T E C 2000 MCCCACCGAUCAA P T E C 2000 MCCCACCGAUCAA P T S C 2000 MCCCACCGAUCAA P T S C 2000 MCCCACCGAUCAA P T S C 2000 MCCCACCCACCAC P C Q G R 2360 CCCCACCCACCAC P Q G R 2360 CCCCACCCACCAC CCCACCCACCAC CCCACCCAC	N N N 1030 CAUCGCAUC 1020 CCUUCUCGU S N 2010 GCCUUCGG S L 2100 GCCUUCGG GCCUUCGG F 2100 GCCUUCGG GCCUUCGG F 2100 GCCUUCGG 2230 GAACGGUAD ANUCCAACA GAACGGUAD N G M T AGAGGGACC G V S AGGGGACC G V S AGGGGACC F C V 2250 CCUCGGCC 2310 AGUGGACCUCC 2250 UCGGGCCCC 2250 UCGGGCCCC 2210 UCGGGCCCC 2210 UCGGGCCCC 2210 UCGGGCCCC 2210 UCGGGCCCC 2210 UCGGGCCCC 2210 UCGGGCCCC 3	$ \begin{array}{c} A & B & B & R \\ 1840 \\ CCGUAGCACCACCGUAGCACCACCGUAGCACCAF & B & A \\ 2020 \\ CCAGUCGCACCAF & S & A \\ 2110 \\ CCAGUCGACCAL & S P & P \\ 2110 \\ CCUGCAGUCGACAL & S P & P \\ 2110 \\ CUGCAGUCGACAL & C \\ 2290 \\ CCCCCCCGCS & P & R \\ CCAGUCGACAAC & P & R \\ C & P & R \\ C & C \\ C & N & S \\ C & C \\ C & S & S \\ C & S & C \\ C & S & S \\ C & S & C \\ C & S & S \\ C & S & C \\ C & S & S \\ C & S \\ C & S & S \\ C $	R K S 1950 1950 1950 1940 1000000000 10000000000 10000000000	G Q P S 1860 UGAACGCGGCC R T R R 1950 UJQUCCCCACUC 2040 1 V F E 2040 1 V F E 2040 1 S N S 2 L E 2 220 UGAACGUCAACUC 2 S L E 2 220 UGAACGUCAACUC 2 S L E 2 220 UGAACGUCAACUC 2 S C L E 2 2 2 C E 2 2 2 C E 2 2 2 C E 2 2 C E 2 400 UGAAAGAU V R G L 2 340 UGAAAGAU V R G L 2 340 UGAAAGAU V R G L 2 340 UGAAGCUCAGGCC V R G L 2 340 UGAACUCAGGCC V R G L 2 340 UGAACUCAGGCC V R G L 2 340 UGAACUCAGGCCAUCA C C C C C C C C C C C C C C C C C C C	$ \begin{array}{c} & \mbox{W A $ D} \\ 1870 \\ 6AAACGGUCAY \\ CAAACAGGUCAY \\ CACAGUCAGUCAGUCAY \\ CACAGUCAGUCAGUCAY \\ CACAGUCAGUCAY \\ CACAGUCAGUCAY \\ CACAGUCAY \\$	R F G 1880 1880 1880 1880 1880 1897 1897 1897 1897 1800 2010	$ \begin{array}{c} {}^{\rm D} {\rm D} {\rm D} {\rm S} \\ 1890 \\ 1800 \\ 1000 \\ $
UGGCCC G CCUCC V AAACU I GAUUC I CCCCC P P CCCCCC P C CCCCC P C CCCCCC P C CCCCCC	1810 1810 1800 AGGALICUE 200 AGGALICUE 1900 AGGALICUE 2010 AGGALICUE 2010 P 2170 CCAAGC 2170 CCAAGCE 2170 CCAAGC 2010 CCUACCCUC 1000 R 2010 CCUACCCUC 2010 CCUACCCCUC 2010 CCUACCCCCUC 2010 CCUACCCCCUC 2010 S 2010 S 2010 CCUACCCCCUC 2010 S 2010 S 2010 S 2010 S 2010 S 2010	UUGAUGAUGAU UUGAUGAUGAU J I E 1910 UUGAGAUGAUGAU S E C 2000 GGCCGGCCUCAA N P T E 2180 GGCCGGCUCAACA P V Q Q 2270 CGCAUCUCACGAU CAUCAU CGCAUCUCACGAU P V Q Q 23260 CGCUCAGGAUCA CGCUCAGGAUCA P V Q Q 23260 CGCUCAGGAUCAU P V Q Q 23260 CGCUCAGGAUCAU P V Q Q 23260 CGCUCAGGAUCAU 23260 CGCUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 23260 CGUCAGGAUCAU 2330 CGUCAGGAUCAU 2310 CGUCAGGAUCAU 2330 CGUCAGGAU 2330 CGUCAGGAU 2330 CGUCAGGAU 2330 CGUCAGGAU 2330 CGUCAGGAU 2300 CGUCAGGAU 2300 CGUCAGGAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGAU 2300 CGUCAGGUCAU 2300 CGUCAGGUCAU 2300 CGUCAGGAU 2300 CGUCAGGUCAU 2300 CGUCAGGAU 2300 CGUCAGGUCAU 2300 CGUCAGGAU 2300 CGUCAGGAU 2300 CGUCAGGAU 2300 CGUCAGGAU 2300 CGUCAGGAU 2300 CGUCAGGAU 2300 CGUCAGGAU 2300 CGUCAGGAU 2300 CGUCAGGUCAU 2300 CGUCAGGAU 2300 CGUCAGGUCAU CGUCAGGAU CGU	N N N 1830 CAUCECAUC 1920 CCUUCUCCU S N 2010 CCUUCUCCU S L 2100 GCCUUCCGG S L 2100 GCCUUCCGG CC P 2100 GCCUUCGG S L Y Q AUUCCAACA SAUCCAACA NAUCCACACU GAACGUAU V Y Q AAGGAAGGUAU Q V Y Q AAGGAAGGUAU Q Z460 AAGCCAACACU Z350 Z420 UUGGGGGGCCC P H Z200 UUGGGUGGCC Z Q S 400 AGCACUAU AGCAUGUGUGU Z 210 UUGGGUGGCUC UUGGUGUGGU L L S 200 UUGGUGUGGU L J 3090 JUGACUCCCC 3090 J100 J	A E B R 1840 CCGUAGCACCA P VA P 1930 UCGAUCUGCACCA CCAUCOCCCCCC F E S A 2020 UCGAUCUCCCCCC A S P P 2110 CCAUCUCCACUC L B C 2290 UCGACGAUCAA F V G A T L E C 2290 UCGACGAUCAA C P G C C C P G C C P G C C P G C C P G C C C C C C C C C C C P G C C C C C C C C C C C C C C C C C C C	R K S 1950 1950 1950 1940 10000 10000 10000 10000 100000 100000 100000 100000 1000000 100000 100000 100000 100000 100000 100000 100000 100000 100000 10000 10000 10000 10000 10000 10000	G Q P S 1860 UGALCOUCGUE UGALCOUCGUE VIADUCCICCAACUC 1 V F E 2040 VIADUCCAACUC 2040 V F E 2040 V F E 2220 UGALCOUCGAACUC V F E 2220 UGALCOUCGAACUC 22400 CAUCUCAACUC 2400 V C E 2400 V C E 2400 V C E 2400 UGCAAACAACUC 2400 V C E 2400 V C E 2550 CCCCAUCUCACG CCCCAUCUC 2550 CCCCAUCUCG A H F 2550 2500 25	W A D 1870 CAAACGGUCAA K R S 1960 GACAUCAGGUCA T S A 2050 GGACABAUACCC E N T 2140 M L 2230 GGACABAUACCC E N T 2140 M L 2230 CUCUCCGCAAGAAU CUCUCCGCAA A D 2320 CUCUCCGUCGC L V E 2680 CUCUCCCUCGC L S 2770 CUCUCCGUCGC L S 2770 CUCUCCCUCGCAA A E 2590 CUCGCCUCAGCA A E 2950 UUUUCAGGAA A E 2950 UUUCCAGGAAU CUCUCCCUCGCUCGC Q E L 3130 CUCCCCCUCGCUCGC S P N 3220 CCAGCGUCGCUCGCUCG CAGCGUCGCUCCG CAGCGUCGCUCCGUCG S P K 320 CUCCCCCCUCGCUCGCUCGCUCGCUCGCUCGCUCGCUCG	R F G 1880 1880 1970 1970 1970 1970 100GACCAG 2050 2050 100GACUCA 2150 100GACUCA 2150 100GACUCACUU 3 K S V G K V 2240 100GCUCAGACUU 3 K S V G K V 2250 100GCUCAGUU 1 K R V 22510 100GCUAGUU 1 K R V 22510 100GCUAUGUU L V C 2690 100GCUAUGUU L V C 2690 100GCUAUGUU 100GCUAUGUU 2780 100GCUAUGUU 2780 100GCUU 2780 100GCUU 100GUU 100GCUU 100GUU	D D S 1890 AGCAGUU AGCAGUU CUCCUUU V P L 2070 CCCCGGU V P L 2070 CCCCGGU V P L 2070 CCCCGGU V 2070 CCCCGGU V 2070 CCCCGGU V 2070 CCCCGGU V 2070 CCCCGGU V 2070 CCCCGGU V 2070 CCCCGGU V 2070 CCCCGGU V 2070 CCCCGGGU UGAGUGC CCCCCGU 2430 UGAGUGC CCCCCGGGU UGAGUGC CCCCCGGGU L A V 2790 UGAGUGC CCCCGGGU L 2790 CCCGGGGUU S G F 2970 CCCGGGGUU S G F 2970 CGGGGGGU S A A 3150 S A 3240 JAGACCCC K K T 3120 CCCCGGCA K T CCCCGGGU CCCCCGGCA S A C 2970 CCCGGGGU S S B S S A S S S A S S S S S S S S S S S S S S S S S S S
UGGCCC G G G G G G G G G G G G G G G G G	1810 1810 1800 SacGarucquest 2000 SacGarucquest 1900 SacGarucquest 1900 SacGarucquest 1900 SacGarucquest 1900 SacGarucquest 20170 C Q G C Q G C Q G C Q G C Q G C Q G C Q G C Q G C Q G C Q G C Q G C Q G C Q G C Q G Q G	UUGAUGAUGAU UUGAUGAUGAU J I E 1910 UUGAUGAUGAUGAU 2000 GGCCGGGCUCAA P V C Q 2070 Q S Q Q Q Q Q Z Z Z D Z Z D	N N 1830 CAUCECAUC 1920 CCUUCLCCU 1920 CCUUCLCCU S F 2010 CCUUCLCCU S L 2100 CCCUUCCGG CC P 2100 CCCUUCCGG CC P 2100 CCCUCCAUC NUUCCAACA F AUUCCAACA G N G N Y QAACGUAUA G N Y AUUCCAACA G AUUCCACAC Y V Y AGUGAAGUAU G A Y 2370 AGUACGUAUA AGUACCACCU Y V Y A G 2250 UCGCACCUCC Y S 220 UCGACUCCUCUCUC UCAACUCCUCUCUC S Y N 3000	A E B R 1840 CCCURACACCA P VA P 1933 UCGAUCUCCACC F E S A 2020 CCAUCUCCCCC A S P P 2110 CCAUCUCCCCC C C CCACUUCCACC CCACUCCCCCC S P R R A H F D I 2210 UCUCCAACCUCC S P P G 2310 UCCCCCCCCCCCC C P P G 2310 C AUCUCCACCUCCC C A 2550 C AUCUCCACCUCCC C A 2550 C AUCUCCACCUCCC C A 2550 C AUCUCCACCUCCC C A 2550 CCACCACCUCCCC C A 2550 CCACCACCUCCCCCCCCCCC C A 2550 CCACCACCUCCCCCCCCCCCC C A 2550 CCACCACCUCCCCCCCCCCCCC C A 2550 CCACCACCCUCCCCCCCCCCCCCCCCCCCCCCCCCCCC	$ \begin{array}{cccc} R & K & S \\ 1850 \\ 1950 \\ 1950 \\ S & I & P \\ 1940 \\ 1000 \\ 10$	G Q P S 1860	W A D W A D 1870 CANACGUCA, K R S 1960 GACCUCAGUCA T S A 2050 GGACABAUACC E N T 2140 CALCUCAUCOUS E N T CALCUCAUCOUS CALCUCAUCOUS CALCUCAUCAGUA CUCUCAAGUA CUCUCAAGUA CUCUCAAGUAAGUAC L V K Q 22410 CUCUCAAGUAAGUAC L V K Q 2500 CUCUCAGUAAGUAC L V E 2600 UUUCCAAGUAAGUAAC V K P 2590 UUUUCAAGUAAC S260 UUUUCAAGUAAC S260 UUUCAAGUAAGUAAC S200 3200 CACCUCUCUUGU S220 S130 CALCUCCUUCUGU S220 S120 CALCUUCUUGUAC S	R F G 1880 1880 1970 1970 1970 1970 1970 2050	D D S 1890 AGCAGGUU AGCAGUU CUCCUUU V P L CUCCUCU V P L CUCCUCU V P L CUCCUCU V P L CUCCUCU V P L CUCCUCU V P L CUCCUCU V P L 2250 CUCCUCU V C A 2250 CUCCUCUCA V C A 23240 CUCCUCA CC L P V P 2430 CUCCUCA V P 2790 2980 CUCCGCGCCC R A 3155 3156 3156 3156 3150
UGGCCC G G G G G G G G G G G G G G G G G	1810 1810 1900 346GAUGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAG	UUGAUAUCGAGA J I = 2 1910 UUGAGAGUCGAGA 2000 UGAGAGUCGAGCUCGA 2000 UGCAGGGCUCGAG 2000 M P T E 2180 GGCGGGCUCAGA P V Q Q 270 CGCGUCAGGGGAG P Q G N 2450 GGGUCGUCGAG CGCUCAGGGGAG P Q G R 2450 GGGAUGUCCGUCGAG CGUCAGGUCGUCGAG 2450 G A A G V P 2450 G G A A G V P 2450 CGCUCAGGCUCGAG 2450 CGCUCAGGCUCGAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 2450 CGCAGCUCCAG 250 CGCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG 250 CGCAGCUCCAG	N N N 1830 CAUCECAUC 1920 CCUUCUCCU CCUUCUCCU S S 2010 GCCUUCCGG S S L R 1 2100 GCCUUCCGG S L R 2100 GCCUUCCGG GCU S L R 1930 ACUGCAACAU F P D A S 2100 ACUGCAGAGG N N C S Q ACUGCAGAGG N Y S Q Z AS V E W R Z 2500 GCUCCGUGGAGG G AS V Z	A E B R 1840 CCCURACACCA P 1933 UCGAUCUCCACC P S A P 2110 CCAUCUCCCCC CAUCUCCCCC A S P P 2110 CCACCUCCCCC A S P P 2110 CCACCUCCCCCC S P R 2200 UCUCGAUCACC S P R 2210 UCUCGAUCACC S P R 2200 UCUCGAUCACC S P R 2210 UCUCGAUCACC S P R 2210 UCUCGACCUCCCC S P R 2310 D F R Q 2550 CCACCCUCCCC C N S 2550 CCACCACCCUCCC C N S 2550 CCACCACCUCCCC C N S 2550 CCACCACCUCCCC C N S 2550 CCACCACCUCCCC C N S 2550 CCACCACCUCCCC C N S 2550 CCACCACCUCCCC C N S 2550 CCACCACCCUCCCC C N S 2550 CCACCACCUCCCC C N S 2550 CCACCACCUCCCC C N S 2550 CCACCACCUCCCC C N S 2550 CCACCACCUCCCC C N S 2550 CCACCACCUCCCC C C S CCACCACCUCCCC C C S CCACCACCUCCCC C C S CCACCACCUCCCCC C C S CCACCACCUCCCCC C C S C C	R K S 1950 1950 1950 1950 1940 1000000000 10000000000 10000000000	G Q P S 1860 1860 1860 1860 1860 1867 187 T R R 1950 2040	W A D 1870 CAAACGGUCAI CAAACGGUCAI K R S 1960 GGACGAAGUCAI T S A 2050 GGACAAUACCC E H T 2107 CACUCCAGGA R L T 2107 CACUCCAGGA CACUCCAGGA CCCCCCCCCC L V E 2560 CCCCCCCCCCCCCCCCCC C S N A R D 2590 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	R F G 1880 1880 1860 1860 1860 1860 1860 1860 1860 1860 1800 2060 2060 2060 2060 2060 2060 2060 2060 2060 2060 2060 2060 2060 2060 2050	D D S 1890 AGCAGGUU B Q F 19980 SUDCCUUU V P L CCUCGGUU V P L CCUCGGU V P L CCUCGGU V P L 2250 CCUCGGUC V C A 2250 CCUCGGUC CCUCGGUCAU S A Q 2250 CCCUCAU S A Q 2250 CCCUCAU S A Q 2250 CCCUCAU S A Q 2340 CCCUCGGUCAU S A Q 2430 CCCUCAU CC A V P 2430 CCCUCAU CC A V P 2430 CC C C A V 2570 CCAU 2050 CC C C A V 2430 CCCUCAU CC C C A V 2430 CCCUCAU CC C C A V 2580 CCCUCAU CC C A V 2580 CCCUCAU CC C A V 2580 CCCUCAU CC C A V 2580 CCCUCAU CC C A V 2580 CCCCCCAU CC C A V 2580 CCCCCCAU CC C A V 2580 CCCCCCAU CC C A V 2580 CCCCCCAU CC C A V 2580 CCCCCCCAU CCCCCCCAU CCCCCCCAU CCCCCCCAU CCCCCCCCAU CCCCCCCCAU CCCCCCCCCC
UGGCC G G G G G G G G G G G G G G G G G	1810 1810 1900 1900 1900 1900 1900 1900 0 Accccscu 1910 1900 0 Accccscu 1900 Java 1900 Accccscu 1900 Accccscu 20170 C Q G C Q G 2170 CCUCCCU 2350 Q A G 2440 CCCCUCSCU 2 A G Q A G 2530 Acaccouu Z 2530 Acaccouu Z 2530 Acaccouu Z 2530 Acaccouu Z 2800 Deducceccusa Z 3160 Iccuacccus S V E 2 3160 Iccuaccus 3 2250 X V E 3 2250 X V E 3 2250 X V E	UIGAUNAUCGAGA J I S J	N N N 1830 1830 CAUCECAUC T S 1920 CCUUCLCCU S COUVELCE Z010 GCCUUCCGG S L R 1920 GCCUUCCGG S L R Y S Q R N Y S Q R T Y S Q R R R T V Z 2370 AGUGCAGGU Z S V K V E W R 2550 CCUCCEGGUGG A S 2240 UCGCAGCCUCE P H R G Z220 UUGGUGUGGU Z 2910 UUGGUGUGGU UUGACUCCUCUU N S 3000 3090 2000 3090 23270 AGUUCUCUGUN S 323270 323270 3240 32470 32470	A E B R 1840 CCCURACCACCA P 1933 UCCAQUEUCACCUCAC P E S A 2020 UCUUGAQUEUCAC A S P P 2110 CCACGUECACCAC A S P P 2110 UCUUGAGUEUCACCACCAC A T L E G 220 UCUUGAGUEAGUE S P R C P P G 2210 UCUCACCUCACCACAC C A 220 UCUCACCUCACCACACAC C P G 2310 UCCACACACACACACACACACACACACACACACACACAC	R K S 1950 1950 1950 1950 1940 100000 10000 10000 10000 10000 10000	G Q P S 1860 1860 1860 1860 1860 187 T R R 1950 2040 2050	W A D 1870 CANACGUCA, CANACGUCA, R R S 1960 GAACAUCAGUCA, T S A 2050 GAACAUCACUCAGUCA, E H A T 2230 GAACAUCACUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCA	R F G 1880 1880 1970 1970 1970 1970 1900 2060 2060 2060 2060 2060 2060 2060 2060 2060 2060 2060 2060 2060 2240 2330 2420 2510 2500 2500 2780	D D S 1890 AGCAGUU AGCAGUU CUCCUU V P 1980 CUCCUU V P 2160 CCUCGGU V P 2160 CCUCGGU V P 2160 CCUCGGU CCUCGGU CCUCGGU S V P S V P S V 22160 CCUCGGU CCUCGGUA S A Q 2250 CCCUCAU S A Q 2250 CCCUCAU CC A A Q 250 CCCUCAU CC A A Q 250 CCCUCAU CC A Q 2700 CCGCGCGC CC A A A 3060 CCCCCCAU CC A A A 3060 CCCCCCAU CC A A A 3060 CCCCCCCAU CC A A A 3060 CCCCCCCAU CC A A A 3060 CCCCCCCCAU CC A A A 3060 CCCCCCCCCCAU CC A A A 3030 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

		3.	430		÷	844	n		345	0		34	160		3	470			348	10		7	490			3500	•		3510
JO	GG	NUA!	GC	CGAG	SUC	JCC	CAC	AAA	GAUG	GCC	AGG	AAC	GGG	CAAG	AAA	ACC	AAC	cuco	CAAC	CAC	GG	3CA	GCAF	GGZ		GAG	IAAC	AGO	76970
	G	I	G	R	v	s	н	к	D	G	0	E	G	0	Е	N	0	T.	0	P	G	A	A	R	ĸ	F	F	F	ם מ
	-	-	-			-			м	_A	°R	ี ห	G	ĸ	ĸ		ิท	ីន	- พ	٠ <u>،</u>	Ğ				Ĩĸ	~ "P	ĩĸ	~ e	·
		35	520		3	1530	'n		354	0		35	50		3	560			357	ດີ		3	580		٠,	1590	, [°]		3600
30	cci	CGI	JGGI	JCGC	auco	GCC	GA	seed	CAG	em	CAA	CGC	GCI	ICCA	GUG	GCC	CAC	IGCO	ucc	- CGC	AITZ	UC	IGGG	ACG	CUI	ICCI	IGGA	CCA	CUAL
	₽	T	W	s	v	G	G	A	P	A	s	т	G	s	s	G	P	G	L	P	D	I	W	D	G	s	พ	T	тт
ł	F	R	G	R	s	A	Е	P	o	L	0	R	Ā	p	v	A	0	A	s	R	r	s	G	т	v	P	G	p	ъ.
		36	510		3	620	ົ່	-	363	0	~	36	40	-	3	650	-		366	0	-	3	570	-		- 8680	, -	-	3690
20	UCL	AAC	ACO	CUGO	icco	cud	CAC	cuco	CGUU	GAG	uuc	CUA	GCG	GAC	UUC	AAG	CGG	AGU	ucc	ACA	uco	GC	GAU	GCG	ACO	ACA	IIAC	GAU	UGUG
	F																												
3	s	N	т	W	P	L	н	s	v	Е	F	L	А	Ð	F	к	R	S	s	т	s	А	D	А	т	т	Y	Ð	C
		31	700		3	3710	2		372	0		37	30		3	740			375	0	-	3	760		- 2	3770		-	3780
JG	ccd	ະບຸບເ	JAAU	UCUG	CCU	CGG	GUG	SUG	GAGU	cuu	GCG	CGU	UGL	UAC	UCC	AUG	UGG	AAG	ICCA	ACA	CGC	UG	GAU	GUC	GUI	IUAC	cuc	ccu	GAGG
,	P	F	N	L	P	R	v	W	s	ь	А	R	C	Y	s	м	w	ĸ	р	т	R	W	b	v	v	Y	τ.	P	R
		31	790		2	800	<u>َ</u>		381	0		38	20		3	830			1384	10		3	850		1	3860	, -	-	3870
JG	AG	GCC	CAC	GGU/	GCU	JGGG	SAG	JAU	CGAG	AUG	UGU	DUU	cuc	UAC	GAC	UAU	GCU	IGAC	ACC	AUC	ccr	AG	SUAC	ACG	GGC	TAAG	AUG	AGC	AGGA
7	s	A	т	v	A	G	s	I	Е	м	с	F	г	Y	D	Y	А	D	т	ï	Р	R	Y	т	G	K	м	s	R
		31	380		2	89	0		390	0		39	10		3	920		5	1393	30	-	3	940	-		395(,	-	3960
co	GCC	GGG	2000	CGUC	ACC	cuci	JAG	CGUI	UUGG	UAC	GGC	GCO	GAG	GGC	UGC	CAC	UU7	CUG	AGU	GGU	GGC	UC	AGCA	CGA	AAI	IGCC	dug	GUC	Igeeu
P	A	G	F	v	т	s	s	v	W	Y	G	A	Е	G	С	н	L	L,	s	G	G	s	А	R	N	А	v	v	A
		3	970		3	398	Ð		399	0		40	000	1	- 4	010			402	20		4	030			4040	5		4050
20	AU	GA	JUG	cuco	JCG2	GUG	266	CUG	GAAA	CGC	GUU	ACU	AGO	ucu	AUA	ccu	AGL	JAGC	GUG	GAU	ICCO	AA.	CGUC	GUA	AA	ACC	AUA	CUG	CCAG
5	М	D	С	S	R	v	G	W	к	R	v	T	s	s	Ι	Р	s	s	v	D	P	N	v	v	N	т	I	L	P
		4	060			107	0		408	80		40	90		4	100			411	0		4	120		÷.,	1130)		4140
Ċτ	AG	cu	AGCI	UGUC	GCG	/uco	JUC	GAU	CAAA	CCG	ACG	GUL	JAG	JGAU	ACG	CCG	GGC	SAAP	CUC	UAC	GUZ	AU	UGCU	AGU	AUG	GUC	CUG	CGC	GAUC
A	R	г	А	v	R	s	s	I	к	Р	т	v	S	D	т	P	G	к	L	Y	v	ĩ	A	s	М	v	L	R	D
		4	150		- 4	116	D		417	0		41	80		4	190			420	0		4	210			1220)		4230
20	GUI	JGA	JCC	AACI	CUC	AAI	JAC	GUG	AGCC	GGA	ccc	AGU	JAG	GAU	AGG	UAU	UGO	GCAC	CAG	UGA	UGI	UC	CACO	GAU	UGO	CUP	CGA	GAU	CGAC
P	v	D	P	т	L	n	т		A	G	Р	s	R	Ð	R	¥	W	H	Q										
		4	240			125	0 -		426	0		4:	270		4	280			429	0		4	300			1310)		4320
υι	JAC	CU	JAA	AAGO	JAU	SOU	GAG	CGU	UCUU	IGAA	CGC	GGG	ccz	AUC	CUA	CGA	GAU	JGAL	JACU	IGGU	GAG	UA	uccu	ACG	UG	GGA	GUU	IGCA	CUCA
		4:	330			134	0		435	0		43	360		4	370		-	438	0		4	390			4400	,		4410
co	CU	ЮĄ	ACC	CAN	\GC(GAC	CCA	UGC	ACUG	GAC	AGU	IGCU	JUAO		GAU	GGU	υυι	JCGG	SAGC	UAC	UG	GC	GGAC	cou	uca	ACCI	AAC	CGI	AUUC
		4	420			143	0		444	0		44	150																
~ ~																													

Fig. 2. Sequence of the 4450 nt of the RYMV RNA and deduced amino acid sequences of the four predicted ORFs. The numbers represent the nucleotides. Nucleotide differences between the sequences of cDNAs 4, 5 and 6 within the CP gene are given above the sequence using bold characters. Amino acids described in the text are underlined.

different C termini due to readthrough of the UGA stop codon. In this case, the ORF1 protein would be 13 residues shorter than the SBMV ORF1 protein which contains 185 amino acids and has an M_r of 21K (Wu *et al.*, 1987). RYMV and SBMV ORF1 proteins do not share extensive sequence homologies. The function of the ORF1 protein is not known for either SBMV or RYMV.

The second ORF initiates at nucleotide 608, and is the largest ORF encoded by RYMV RNA. Unlike SBMV, ORF2 of RYMV does not overlap with ORF1, and though its initiation codon is six nucleotides downstream of the ORF1 termination signal, it includes a strong initiation motif, according to the models proposed by Kozak (1989). ORF2 includes 3000 nt (bases 608 to 3607) and encodes a polypeptide with 999 amino acids and a calculated M_r of 100.7K. This polypeptide is presumed to contain the VPg, a protease and the polymerase by comparison with the sequences of proteases and polymerases of other RNA viruses (Nicklin *et al.*, 1986; Wellink & van Kammen, 1988; Koonin, 1991).

The ORF2 polyprotein is the most conserved polypeptide between RYMV and SBMV. The N-terminal third of the RYMV and SBMV ORF2 proteins (amino acids 1 to 303 of both polypeptides) show 50% sequence similarity. In this region are found sequences that are highly conserved in the proteases of the picorna-, potyand comoviruses and with cellular proteases (Gorbalenya *et al.*, 1988; Fig. 4). This region is therefore thought to contain a serine protease, because the catalytic site involves a serine, instead of a cysteine, as is the case for 1



Fig. 3. Genome organization of RYMV with the predicted ORFs indicated. The numbers indicate the nucleotide positions in the genome.

cysteine proteases of picorna-, poty- and comoviruses (Gorbalenya et al., 1988).

The internal region of the predicted polymerases of RYMV contains motifs similar to the well conserved GXXXXGK and DXXG domains (nt 1587 to 1610 and nt 1676 to 1687, respectively) thought to be associated with a purine NTP binding side or with helicase activity (Hodgman, 1988; Gorbalenya & Koonin, 1989; Stanway, 1990; Fig. 2). Between nucleotides 2120 and 3137 is located a third conserved motif also identified as an NTP-binding site in the picornavirus and SBMV polymerases (Wu *et al.*, 1987; Fig. 2).

The C-terminal regions of both polymerases (RYMV amino acid residues 510 to 940 and SBMV residues 458 to 887) share 50% sequence identify and 78% sequence similarity. This domain contains the four most conserved blocks, DXXXXD, GXXXTXXXN, GDD and K, identified in the polymerases of positive-sense RNA viruses (Poch *et al.*, 1989; Fig. 2). It has been suggested that these blocks are characteristic of the proteins associated with the replication of viral RNA and could function as an active binding and/or recognition site of an RNA-dependent RNA polymerase (Kamer & Argos, 1984).

ORF3 (nt 2092 to 2470) is initiated within the Cterminal region of ORF2. It encodes a 126 amino acid protein with a calculated M_r of 13.7K. The ORF3 protein is smaller than the SBMV ORF3 product which is 18.3K. The SBMV and RYMV ORF3 polypeptides share two conserved regions. The first region includes the N-terminal 49 amino acids of both proteins and shows 51% sequence identity. The second block, RYMV amino acids 79 to 103 and SBMV residues 59 to 83, contains only 21% sequence identity; however, the area is surrounded by many other similar amino acids. The function of this ORF is not known.

The AUG at position 3447 is likely to be the initiation codon for the fourth ORF because it is in the context ACAAAGAUGGC, which is similar to the consensus sequence for translation initiation in plants (AACAAU-GGC) described by Lütcke et al. (1987). ORF4 ends at a stop codon at nt 4166. At 36 nt downstream is a UGAUGA (double stop codon) in frame with the ORF4 coding sequence. Considering that the RYMV CP is released as a doublet when subjected to SDS-PAGE (Fig. 1a), it is possible that ORF4 encodes two proteins with the same N terminus but two different C termini, due to leaky termination. ORF4 is predicted to encode a protein of 239 amino acids when the first stop codon is used and a protein of 251 amino acids when the second stop codon is used, encoding proteins of 26K and 27K, respectively. The ORF4 protein has been identified as the CP by comparison with the N-terminal amino acid sequence of the RYMV CP.

The N-terminal first 22 amino acid sequence of the RYMV CP contains the sequence RKGKKTNSNQG-QQGKRKSRR (amino acids 3 to 22), which is identical to the bipartite nuclear targeting motif (Dingwall & Laskey, 1991).

Computer alignment of the CP sequence of SBMV strain C with that of RYMV shows 50% similarity between the two sequences and 26% identical amino acids. Comparison of the sequence with the known structure of SBMV CP (Hermodson *et al.*, 1982) suggests that these homologies play a significant role in virus structure (Fig. 5). The homologous sequences are arranged such that the α -helix and β -sheets of the SBMV CP appear to be conserved in RYMV CP. Thus, the tertiary structure of RYMV is likely to be similar to that of SBMV. When the RYMV CP sequence was compared with CP sequences of several other small spherical

		* *		*	
SBMV	LGFGARVYHE-GMDVLMVI	PHHV-31-RIDFV	LVKVPTA-52-PTA	KGW S GTPLYTRE	G-IVGMHTGYVD
		11 11			
RYMV	LGLGTRVRTPTGRDLLMT	NHHI-29-HIDCA	FYEVPPK-52-TTC	SGW S GSPLYHK-	GCVVGLHIGAAD
Cons	- <u>G</u> - <u>G</u>	H xD	к <u>А</u> - у	-G- <u>C</u> G(GG
	SS		P	S	

Fig. 4. Alignment of the amino acid sequence surrounding the catalytic sites (*) of the putative protease of RYMV ORF2 with the homologous sequences of SBMV and the consensus sequence (Cons) obtained from several picornaviruses, cowpea mosaic virus proteases and selected cellular proteases as described by Gorbalenya *et al.* (1988). The vertical bars indicate amino acids identical between SBMV and RYMV. Hyphens (-) represent gaps to allow maximal alignment; x and y represent the variable numbers of amino acids. The leucine (L) residue at the left is encoded by RYMV nt 1097 to 1099 (see Fig. 2). The serine (S) change in the catalytic site of RYMV and SBMV proteases in comparison to cellular and cysteine proteases of the other virus is shown in bold.

254 M. Ngon A Yassi and others

RYMV SBMV	$\begin{array}{cccc} \mbox{MARKGKKTNSNQGQQGKRKSRRPRGRSAEPQLQRAPVAQASRISGTVPGPLSS-NTWPLHSVEFLADFKRSSTSADATTYDCVPFNLPRVWSLARCYS} \\ A: :::T::N.::::R ::R : ::::::. ::AP:AQ::: :: .P:::S :::L:: E. ::. ::T::::::. PF : W L:::: : LAQAIQNTLPNPPRRKRAKRRAAQVPKPTQAGVSMAPIAQGTMVKLRPPMLRSSMDVTILSHCELSTELAVTDTIVVTSELV-MPFTVG-TW-LRGVAQ \\ \mbox{$
RYMV	MWKPTRWDVVYLPEVSATVAGSIEMCFLYDYADTIPSDTGKMSRTAGFVTSSVWYGAEG-CHLLSGGSARNAVVASMDCSRVGWKRVTS-
	:W:::.W : YLP. ::T :G:I:M:F:YD ADT P: ::: S:::G.VT::VW.G:.G C: .:: ::: A: :: D:::V: .K::T:
SBMV	NWSKYAWVAIRYTYLPSCPTTTSGAIHMGFQYDMADTLPVSVNQLSNLKGYVTGPVWEGQSGLCFVNNTKCPDTSRAITIALDTNEVSEKRYPFKTATDY
	$\Box \sqsubseteq_{\beta D} \{\beta D} \{\beta E} _ \Box c_a c_a [\{\alpha B} _] \{\beta F} _ \Box \{\alpha C} _ \Box [\{\beta G} _] c_a [\]$
RYMV	SIPSSVDPNVVNTILPARLAV-RSSIKPTVSDTPGKLYVIASMVLRDPVDPTLN
	: :::V :N::N. PARL:: :::::G:LY::::.L::P::::LN
SBMV	ATAVGVNANIGNILVPARLVIAMEGGSSKTAVNTGRLYASYTIRLIEPIAAALN

Fig. 5. Alignment of the deduced amino acid sequences of the CPs of RYMV (ORF4 amino acids 1 to 239) and SBMV-C (ORF4 amino acids 9 to 259). Sequence identities are shown in the middle line. Colons (:) indicate residues that have the same polarity; hyphens (-) represent gaps to allow maximal alignment. The known structural domains of SBMV C (i.e. R, arm, S), and Ca²⁺ binding sites (Ca) are also represented below the sequences. The secondary structure of SBMV (α -helices and β -sheets) is also indicated. The R domain, arm and S domain are represented according to the SBMV coat protein structure as defined by Hermodson *et al.* (1982).



Fig. 6. Predicted folding of the 5' UTR of the RYMV genomic RNA sequence using the STAR 100 computer program. The box between nt 30 and 35 represents sequences complementary to sequences found in the maize mitochondrial 5S ribosomal RNA (rRNA) (Chao *et al.*, 1983), the wheat mitochondrial 18S rRNA (Spencer *et al.*, 1984) and the soybean chloroplast 18S rRNA (de Lanversin & Pillay, 1988). Sequences in the grey background have been identified by alignment of the 5' UTR nucleotide sequences with those complementary to sequences near the 3' terminus of the maize mitochondrial 5S rRNA.

viruses from different groups, sequence homologies were primarily found in the S domain of the CPs. The highest similarity score was between RYMV and tobacco necrosis virus (TNV). This result agrees with the proposals of Meulewaeter *et al.* (1990) and Dolja & Koonin (1991) that CPs of sobemoviruses and the CP of the necrovirus TNV are phylogenetically related (Fig. 8a).



Ð

v

Y.

Fig. 7. Computer-predicted secondary structure of the 3'-terminal 147 nt of RYMV RNA. The analysis was made with the software STAR 100. Long, extended lines indicate continuity of the sequence regions; numbers correspond to the nucleotide position relative to the 3' nucleotide.

The non-coding region at the 5' terminus of RYMV RNA contains 79 nt. Within this region, at residues 30 to 37, the sequence 3' CGAAAGCA 5' is partially complementary to sequences at the 3' end of the 18S or the 5S ribosomal RNA of plant organelles (Fig. 6). Furthermore, this region is predicted to possess two stem and loop structures with predicted Δ G of $-132\cdot3$, 56.7 kJ for the stem–loop at the 5' region, and $-152\cdot9$, 74.8 kJ for the internal one, based on the algorithms of Abrahams *et al.* (1990), using the STAR 100 program.



Fig. 8. Relationship of RYMV with other small icosahedral viruses on the basis of amino acid sequences of coat proteins (a) and their putative polymerases (b). The computer program DNA Star was used to make a pairwise multiple alignment of all viruses based upon their amino acid sequences. (a) Amino acid sequences of BWYV strain F, PLRV, barley yellow dwarf virus (BYDV-MAV; Ueng et al., 1992), TNV, SBMV C, RYMV, tomato bushy stunt virus (TBSV; Hopper et al., 1984), cucumber necrosis virus (CNV; Rochon & Tremaine, 1989), turnip crinkle virus (TCV; Carrington et al., 1989) and carnation mottle virus (CarMV; Guilley et al., 1985).

The 3' untranslated region of RYMV RNA contains 245 nt. Using the same computer program and algorithm, the last 147 nt of the RNA show a structure with a ΔG of 532·1, 105 kJ with a pseudoknot domain (nt -92 to -147, relative to the 3' nucleotides) and a stem-loop domain (nt -1 to -91, relative to the 3' nucleotides) (Fig. 7).

Discussion

.

The genome of RYMV, a virus that infects rice in Africa, has been characterized by cDNA cloning and sequence analysis. RYMV is the second sobemovirus whose genome has been sequenced, the first being SBMV strain C (Wu *et al.*, 1987). The genomic organization of RYMV RNA as shown in Fig. 3 is very similar to that of SBMV strain C (Wu *et al.*, 1987). With the exception of ORF1, RYMV proteins are highly homologous to their counterparts in SBMV.

It is probable that the nucleotide sequences partially complementary to the 3'-terminal sequences of 18S or 5S rRNA in the 5' untranslated regions (UTR) of both RYMV (Fig. 6) and SBMV RNA (Wu et al., 1987) are involved in ribosomal binding and translation of ORF1. This hypothesis is supported by the observation that, first, the context sequences surrounding the AUG initiation codons of ORF1 in both RYMV and SBMV strain C are weak in comparison with the consensus sequences for initiation of translation of eukarvotic mRNAs as proposed by Kozak (1989) and Lütcke et al. (1987). These authors predicted conserved A or G residues at position -3 relative to the first nucleotide of the initiation codon, and G at position +4. Both RYMV and SBMV C have nucleotides U at -3 and A at +4. Second, within this region, the sequence ACAAUUG is repeated twice. This sequence is also found in the 5' UTR of tobamoviruses (Avila-Rincon et al., 1989) and in those

of the leader sequences of chloroplast mRNAs (Danon & Mayfield, 1991). Third, this region contains an AUU codon in the loop at nucleotide positions 26 to 28, which could act as a ribosome entry sequence, as is the case in the 5' UTR of tobacco mosaic virus RNA (Tyc *et al.*, 1984). Computer-assisted folding of the first 121 nt of the RYMV RNA using the STAR 100 computer program shows a stem-loop structure similar to that of the 5' UTR of the mRNA of the chloroplast *psbA* gene (Danon & Mayfield, 1991). In this case, the sequence complementary to the 18S rRNA in RYMV and SBMV C RNA might act as a ribosome binding site to allow efficient initiation of translation of ORF1. How the viral RNA might gain access to the ribosomes of either the chloroplast or mitochondria, if at all, is unknown.

The last 90 nt of the RYMV genome are predicted to form a stem-loop domain. It should be noted that the folding of the last 120 nt of SBMV C RNA (Wu *et al.*, 1987) also shows a conserved structure, similar to that of RYMV (data not shown). The sequence immediately upstream of this region (nt 91 to 147 from the 3' terminus; Fig. 7) is similar to the upstream pseudoknot domain in tobacco mosaic virus RNA (Avila-Rincon *et al.*, 1989), whereas in SBMV C RNA, there is only one pseudoknot predicted.

By analogy to SBMV C (Gorbalenya *et al.*, 1988), the viral VPg could be included in the N-terminal 134 amino acids of the putative polyprotein. In this region, there are two conserved blocks in both RYMV and SBMV, of 14 residues (positions 34 to 47 in RYMV and 38 to 52 in SBMV) for the first segment, and of nine amino acids (positions 81 to 90 in RYMV and 61 to 70 in SBMV) for the second block. In the second domain, a conserved tyrosine at RYMV position 83 and SBMV position 63 could form a phosphodiester bond with the 5' nucleotide of the genome, as proposed by Tobin *et al.* (1989) in the case of the poliovirus VPg.

Based upon computer-assisted pairwise comparisons (DNA Star), the predicted RNA-dependent RNA polymerase of RYMV is closely related to its counterpart of SBMV, followed by those of the luteoviruses potato leafroll virus (PLRV) and beet western yellows virus (BWYV). These findings support the suggestion that these viruses are phylogenetically related to each other at the polymerase level (Miller *et al.*, 1988; Mayo *et al.*, 1989; Koonin, 1991; Fig. 8*b*).

The direct significance of the nuclear targeting motif located at the N-terminal first 26 amino acids of the RYMV CP sequence is not known. However, as in RYMV, SBMV C also contains a bipartite nuclear targeting motif at the N-terminal first 28 amino acids of its mature CP (residues 6 to 28; Hermodson *et al.*, 1982). This finding may explain the observation that, in sobemovirus infections, virus particles have been found in nuclei of infected cells (Francki *et al.*, 1985). It would be interesting to investigate whether the virus CP enters the nucleus as a CP subunit, in which case it is probable that the particles found in nuclei might be empty, or whether the particles assemble in the cytoplasm before entering the nuclei.

The necrovirus TNV CP sequence (Meulewater *et al.*, 1990) is more closely related to that of SBMV C (34% identity) than is the RYMV CP (26% identity). Sequence alignment and comparison predicts that the tertiary structure of RYMV and TNV CPs (Meulewater *et al.*, 1990) could be identical to that of SBMV C as described by Hermodson *et al.* (1982). RYMV, SBMV C and TNV all lack a P domain and have identical Ca²⁺ binding sites (Meulewater *et al.*, 1990; Fig. 5).

Most of the results discussed here are based upon sequence analysis of RYMV RNA and comparisons with viruses that have been more thoroughly studied. A fulllength cDNA clone for RYMV has recently been obtained in our laboratory (C. Brugidou *et al.*, unpublished), and it will be interesting to investigate the possibility of translation and/or replication enhancement by both the 5' and the 3' UTRs of RYMV RNA. Other experiments involving the *in vitro* transcription and translation of sequences that include the ORF2 polyprotein should be performed to confirm the activity of a protease in this sobemovirus.

We thank Christine Smith, Michael Jennings and the Monsanto Company for the sequencing of the RYMV CP and for providing the degenerate oligonucleotides used in this work. We also thank Dr John Fitchen for critical review of the manuscript. This research was supported by a grant from the Rockefeller Foundation and by the French Scientific Institute for Development through Cooperation (ORSTOM).

References

ABRAHAMS, J. P., VAN DEN BERG, M., VAN BATENBURG, E. & PLEIJ, C. (1990). Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. Nucleic Acids Research 18, 3035-3044.

٤.

1

Ų.

1

- ATTERE, A. F. & FATOKUN, C. A. (1983). Reaction of Oryza glaberrima accessions to rice yellow mottle virus. Plant Disease 67, 420–421.
- AVILA-RINCON, M. J., FERRERO, M. L., ALONSO, E., GARCÍA-LUQUE, I. & DÍAZ-RUÍZ, J. R. (1989). Nucleotide sequences of 5' and 3' noncoding regions of pepper mild mottle virus strain S RNA. *Journal of General Virology* 70, 3025–3031.
- BAKKER, W. (1974). Characterization and ecological aspects of rice yellow mottle virus in Kenya. Agricultural Research Reports, Wageningen 829.
- CARRINGTON, J. C., HEATON, L. A., ZUIDEMA, D., HILLMAN, B. I. & MORRIS, T. J. (1989). The genome structure of turnip crinkle virus. *Virology* **170**, 219–226.
- CHAO, S., SEDEROFF, R. R. & LEVINGS, C. S. (1983). Partial sequence analysis of the 5S to 18S rRNA gene region of the maize mitochondrial genome. *Plant Physiology* **71**, 190–193.
- COSTA, A. S. & KITAJIMA, E. W. (1972). Cassava common mosaic virus. CMI/AAB Descriptions of Plant Viruses, no. 90.
- DANON, A. & MAYFIELD, S. (1991). Light regulated translational activators: identification of chloroplast gene specific mRNA binding proteins. *EMBO Journal* 10, 3993–4001.
- DE LANVERSIN, G. & PILLAY, D. T. (1988). Primary structure and sequence organization of the 16-23S spacer in the ribosomal operon of soybean (*Glycine max* L.) chloroplast DNA. *Theoretical and Applied Genetics* **76**, 443-448.
- DINGWALL, C. & LASKEY, R. A. (1991). Nuclear targeting sequence a consensus. Trends in Biochemical Sciences 16, 478–481.
- DOLJA, V. V. & KOONIN, E. V. (1991). Phylogeny of capsid proteins of small icosahedral RNA plant viruses. *Journal of General Virology* 72, 1481–1486.
- DOUGHERTY, W. G. & HIEBERT, E. (1980). Translation of potyvirus RNA in a rabbit reticulocyte lysate: cell-free translation strategy and a genetic map of the potyviral genome. *Virology* **104**, 183–194.
- FAUQUET, C. & THOUVENEL, J. C. (1977). Isolation of the rice yellow mottle virus in Ivory Coast. *Plant Disease Reporter* 61, 443–446.
- FRANCKI, R. I. B., MILNE, R. G. & HATTA, T. (1985). Atlas of Plant Viruses, pp. 153–169. Boca Raton: CRC Press.
- GORBALENYA, A. E. & KOONIN, E. V. (1989). Viral proteins containing the purine NTP-binding pattern. *Nucleic Acids Research* 17, 8413-8440.
- GORBALENYA, A. E., KOONIN, E. V., BLINOV, V. M. & DONCHENKO, A. P. (1988). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Letters* 236, 287–290.
- GUBLER, U. & HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* 25, 263–269.
- GUILLEY, H., CARRINGTON, J. C., BALAZS, E., JONARD, G., RICHARDS, K. & MORRIS, T. J. (1985). Nucleotide sequence and genome organization of carnation mottle virus RNA. *Nucleic Acids Research* 13, 6663–6677.
- HARI, V., SIEGAL, A., ROZEK, C. & TIMBERLAKE, W. E. (1979). The RNA of tobacco etch virus contains poly(A). *Virology* **92**, 207– 216.
- HERMODSON, M. A., ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., PUNDAK, S., ROSSMANN, M. G. & TREMAINE, J. H. (1982). Amino acid sequence of southern bean mosaic virus coat protein and its relation to the three-dimensional structure of the virus. *Virology* 119, 133–149.
- HODGMAN, T. C. (1988). A new superfamily of replicative proteins. Nature, London 333, 22-23.
- HOPPER, P., HARRISON, S. C. & SAUER, R. T. (1984). Structure of tomato bushy stunt virus. V. Coat protein sequence determination and its structural implications. *Journal of Molecular Biology* 177, 701-713.
- HULL, R. (1988). The sobemovirus group. The Plant Viruses, vol. 3, Polyhedral Virions with Monopartite RNA Genomes, pp. 113–146. Edited by R. Koenig. New York: Plenum Press.
- KAMER, G. & ARGOS, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Research* 12, 7269–7282.

- KOONIN, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology* 72, 2197–2206.
- KOZAK, M. (1989). A scanning model for translation: an update. Journal of Cell Biology 108, 229-241.
- LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, London 227, 680-685.
- LÜTCKE, H. A., CHOW, K., MICKEL, F. S., MOSS, K. A., KERN, H. F. & SCHEELE, G. A. (1987). Selection of AUG codons differs in plants and animals. *EMBO Journal* 6, 43–48.
- MATSUDAIRA, P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *Journal of Biological Chemistry* 262, 10035–10038.
- MATTHEWS, R. E. F. (1982). Classification and nomenclature of viruses. Intervirology 17, 1–199.
- MAYO, M. A., ROBINSON, D. J., JOLLY, C. A. & HYMAN, L. (1989). Nucleotide sequence of potato leafroll luteovirus RNA. *Journal of General Virology* 70, 1037–1051.
- MEULEWATER, F., SEURINCK, J. & VAN EMMELO, J. (1990). Genome structure of tobacco necrosis virus strain A. Virology 177, 699–709.
- MILLER, W. A., WATERHOUSE, P. M. & GERLACH, W. L. (1988). Sequence and organization of barley yellow dwarf virus genomic RNA. Nucleic Acids Research 16, 6097-6111.
- NICKLIN, M. J. H., TOYODA, H., MURRAY, M. G. & WIMMER, E. (1986). Proteolytic processing in the replication of polio and related viruses. *Bio/Technology* 4, 33–42.
- POCH, O., SAUVAGET, I., DELARUE, M. & TORDO, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO Journal* 8, 3867–3874.
- QU, R., BHATTACHRYYA, M., LACO, G. S., DE KOCHKO, A., SUBBA RAO, B. L., KANIEWSKA, M. B., SCOTT ELMER, J., ROCHESTER, D. E., SMITH, C. E. & BEACHY, R. N. (1991). Characterization of the genome of rice tungro bacilliform virus: comparison with commelina yellow mottle virus and caulimoviruses. *Virology* 185, 354–364.
- ROCHON, D. M. & TREMAINE, J. H. (1989). Complete nucleotide sequence of cucumber necrosis virus genome. *Virology* 169, 251–259. SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular*

1

Cloning: A Laboratory Manual, 2nd edn. New York: Cold Spring Harbor Laboratory.

- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences, U.S.A. 74, 5463–5467.
- SEHGAL, O. P. (1981). Southern bean mosaic virus group. In *Handbook* of *Plant Virus Infection and Comparative Diagnosis*, pp. 91–121. Edited by E. Kurstak. Amsterdam: Elsevier/North-Holland.
- SERGHINI, M. A., RITZENTHALER, C. & PINCK, L. (1989). A rapid and efficient 'miniprep' for isolation of plasmid DNA. Nucleic Acids Research 17, 3604.
- SMITH, O. P., HARRIS, K. F., TOLER, R. W. & SUMMERS, M. D. (1988). Molecular cloning of potato leafroll virus complementary DNA. *Phytopathology* 78, 1060–1066.
- SPENCER, D. F., SCHNARE, M. N. & GRAY, M. W. (1984). Pronounced structural similarities between the small subunit ribosomal RNA genes of wheat mitochondria and *Escherichia coli*. Proceedings of the National Academy of Sciences, U.S.A. 81, 493–497.
- STANWAY, G. (1990). Structure, function and evolution of picornaviruses. Journal of General Virology 71, 2483–2501.
- TOBIN, G. J., YOUNG, D. C. & FLANEGAN, J. B. (1989). Self-catalyzed linkage of poliovirus terminal protein VPg to poliovirus RNA. *Cell* 59, 511–519.
- TYC, K., KONARSKA, M., GROSS, H. J. & FILIPOWICZ, W. (1984). Multiple ribosome binding to the 5' terminal sequence of tobacco mosaic virus RNA. Assembly of an 80S ribosome mRNA complex at the AUU codon. *European Journal of Biochemistry* 140, 503-511.
- UENG, P. P., VINCENT, J. R., KAWATA, E. E., LEI, C.-H., LISTER, R. M. & LARKINS, B. A. (1992). Nucleotide sequence analysis of the genomes of the MAV-PS1 and P-PAV isolates of barley yellow dwarf virus. *Journal of General Virology* 73, 487–492.
- WELLINK, K. J. & VAN KAMMEN, A. (1988). Proteases involved in the processing of viral polypeptides. Archives of Virology 98, 1–26.
- WU, S., RINEHART, C. & KAESBERG, P. (1987). Sequence and organization of southern bean mosaic virus genomic RNA. *Virology* 161, 73–80.

(Received 24 May 1993; Accepted 21 September 1993)