

GEN 01771

## The primary structure of the mitochondrial genome of *Saccharomyces cerevisiae* — a review

(Yeast; genes; open reading frames; non-coding sequences; *ori* sequences; polymorphism)

Miklos de Zamaroczy and Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, Tour 43, 2 Place Jussieu, 75005 Paris (France)  
Tel. (1) 43 29 58 24; (1) 43 36 25 25, ext. 41.01

(Received July 24th, 1986)

(Accepted August 4th, 1986)

---

### SUMMARY

We have collated and compiled all the available primary structure data on the mitochondrial genome of *Saccharomyces cerevisiae*. Data concern 78 500 bp, namely 92% of the 'long' genomes; they are derived from several laboratory strains. Interstrain differences belong to three classes: (i) a small number of large deletions/additions, mainly concerning introns; (ii) a large number of small (10–150 bp) deletions/additions located in the intergenic sequences; (iii) 1–3 bp deletions/additions and point mutations; the interstrain sequence divergence due to the latter, is of the order of 2% for the strains compared; this low value is, however, an overestimate because of sequence mistakes.

---

### INTRODUCTION

The mitochondrial genome has become a favorite subject of investigations in the field of the molecular evolution of eukaryotes for a number of reasons.

(i) It is the only eukaryotic genome in which the primary structure is fully known in several species, including man (Anderson et al., 1981), mouse (Bibb et al., 1981), calf (Anderson et al., 1982), *Xenopus laevis* (Roe et al., 1985), and *Drosophila yakuba* (Clary and Wolstenholme, 1985) and more or less largely known in a number of other ones, such as *Saccharomyces cerevisiae* (see present work), *Schizosaccharomyces pombe* (Lang et al., 1983), *Torulopsis*

*glabrata* (Clark-Walker et al., 1985; Ainley et al., 1985), *Aspergillus nidulans* (see Brown et al., 1985), *Neurospora crassa* (see Breitenberger and Raj-Bhandary, 1985), *Trypanosoma brucei* (Benne et al., 1985), and *Chlamydomonas reinhardtii* (Vahrenholz et al., 1985).

(ii) It is a very ancient genome having, in all likelihood, a symbiotic, monophyletic origin (see Clark-Walker, 1985).

(iii) It contains a small number of genes, which are indispensable for cellular respiration and highly conserved in sequence.

(iv) It is characterized by a very high mutation rate in animals.

(v) It varies in size from extremely compact forms, almost exclusively containing coding sequences, to very large forms where the majority of the sequences are non-coding. This last feature is of special interest because it provides a unique oppor-

---

Abbreviations: bp, base pair(s); CRF, closed reading frame; kb, 1000 bp; nt, nucleotide(s); ORF, open reading frame; *ori*, origin of DNA replication; RFLP, restriction fragment length polymorphism.

TABLE I

## Sequence data

Gap <sup>a</sup>	Region <sup>b</sup>	Position <sup>c</sup>	Strain <sup>d</sup>	Ref. <sup>e</sup>
<b>a</b>	<i>Sal1 21S (R2)</i>	1	I	(1)
		908	I,D1	
<b>1</b>	<i>thr2-cys-his</i>	909	D1	(2)
		2284	D1	(3)
<b>1</b>	190	2725	D1	(3)
		3598	D1	(4)
<b>1</b>	<i>lys-arg1-gly</i>	5468	A	(5,22)
		6096	D1	(4)
<b>2</b>	<i>ser2-arg2</i>	6672	S	(6)
		8405	D1	(7)
<b>2</b>	480	9066	D1	(7)
		9067	D1	(7,8)
<b>3</b>	420	12871	D1	(7,9)
		12872	D1	(9)
<b>4</b>	500	13242	D1	(10)
		13243	D1	(10)
<b>4</b>	<i>val-oxi2</i>	13552	D2	(11)
		15316	D2	(12)
<b>4</b>	ORF2	17102	D2	(11)
		17261	M	(13)
<b>4</b>	<i>ori5<sup>h</sup></i>	18058	A	(14)
		19042	A	(15)
<b>7</b>	550	19443	A	(16)
		19834	A,M	(15,17)
<b>8</b>	<i>fmet-9S-pro<sup>i</sup></i>	19835	M	(17)
		22441	M,A,B	(5,18)
<b>8</b>	1800	22442	B	(19)
		23244	B,A	
<b>b</b>	<i>ori1<sup>j</sup></i>	23245	J	(38)
		24460	A,B	(5,14,18)
<b>9A</b>	.... <sup>k</sup>	24461	B	(14)
		24942	B,M	(5,14,18)
<b>9B</b>	800 <sup>k</sup>	24943	M	(21)
		24943	M	(21)
<b>10A</b>	ORF3- <i>trp<sup>m</sup></i>	27098	A	(5,22)
		600	A	(5,22)
<b>10A</b>	<i>ori8<sup>n,o</sup></i>	29294	A	(23)
		29976	K	(24)
<b>10B</b>	120	30791	K	(24)
		30792	D2	(25)
<b>10B</b>	.... <sup>p</sup>	39377	K	(26)
		42413	D2	(25)
<b>10B</b>	<i>aap1-oli2<sup>q</sup></i>	43798	J	(27)
		46386	D2	(28)
<b>10B</b>	ORF4	48086	D2	(29)
		48442	B	(14)
<b>10B</b>	48778	48778	D2	(29)
		48860	B	(15)
<b>10B</b>	ORF5	49265	B	(30)
		51053	B	(15)
<b>12</b>	<i>ori2</i>	51453	B	(16)
		51945	B	(14,16)
<b>12</b>	.... <sup>r</sup>	51946	B	(14)

(TABLE I, continued)

Gap <sup>a</sup>	Region <sup>b</sup>	Position <sup>c</sup>	Strain <sup>d</sup>	Ref. <sup>e</sup>
<b>13</b>	<i>glu<sup>s</sup></i>	53537	D1	(31)
		55059	D1	(32)
<b>13</b>	.... <sup>p</sup>	55474	F	(33)
		250	57268	F
<b>13</b>	250	57269	F	(33)
		57374	F	(34)
<b>13</b>	.... <sup>p</sup>	59070	D1	(32)
		63122	B	(15)
<b>13</b>	<i>ori6</i>	63468	D1	(32)
		64049	J	(35)
<b>c</b>	<i>oli1<sup>1</sup></i>	65585	J,D1	
		65586	D1	(36)
<b>c</b>	<i>ser1<sup>u</sup></i>	67352	E	(37)
		68567	D1	(36)
<b>16</b>	400	69128	D1,A	(16,36)
		69129	A	(16)
<b>17</b>	120 <sup>w</sup>	70877	A	(16)
		70878	A	(16)
<b>17</b>	<i>ori3</i>	71090	A	(15)
		71491	A	(16)
<b>17</b>	<i>ori4</i>	73092	A	(15)
		73554	A	(16)
<b>d</b>	73735	73735	A,I	
		73736	I	(1)
		21S <sup>x</sup>		
Total sequenced regions		78520 bp (92%)		
Total gaps (neglecting minigaps)		6630 bp (8%)		
Genome size ('long' genomes)		85150 bp (100%)		

<sup>a</sup> Sequence gaps are numbered according to [42] (numbers in square brackets refer to the references in footnote e). Recent sequence work has eliminated gaps **5**, **6**, **11**, **14**, **15** and split gaps **9** and **10**. Four minigaps (less than 30 bp long) are indicated by letters **a-d**; three of them (**a**, **b**, **d**) may correspond to local sequence rearrangements.

<sup>b</sup> Regions are indicated by the landmarks contained in them (see Fig. 1). Estimated gap sizes (bp) are indicated in the same column.

<sup>c</sup> Numbering neglects minigaps and gaps. The positions given indicate the first bp in the corresponding region. Transcription initiation sites ('nonnucleotides') and transcript processing sites have been listed in [42], except for those of *thr2* (*e*), *cys* (*i*) and *ser2* (*i.e.*) [39]; *ala* (*i*) [39,6]; *oxi1* (*i*) [6]; *aap1* (*e*) [44]. Clusters are numbered according to [46] (see below). The total number of GC clusters at the present time is 203, which are distributed in classes [46] as follows: 24, *ori*; 44, *a1*, 32, *a2*; 27, *a3*; 33, *a4*; 7, *v*; 14, *c*; and 22, *n*. Short sequences considered as uncertain in the original papers are the following: 1089-1140; 3805-3849; 8745-8761; 12235-12240; 70798-70886.

<sup>d</sup> Wild-type yeast strains are indicated with the single capital letter abbreviation system used in our laboratory: A, D243-2B-R1; B, C982-19d; C, NCYC-74S; D1, D273-10B/A21; D2, D273-10B/A48; E, A10; F, 777-3A; I, IL8-8C; J, J69-1B; K,

KL14-4A; M, MH41-7B; S, SM202. Strains A and J have a common origin (see [19] and [38]) and strains D1 and D2 are closely related.

\* References concern only the publications actually used in the sequence compilation or those providing supplementary information, like gap sizes. Additional references corresponding to regions or sub-regions sequenced in other strains or by other authors can be found elsewhere (see [42]) or in the following footnotes.

[1] Sor and Fukuhara (1983); [2] Berliani et al. (1980a); [3] Berliani et al. (1980b); [4] Bonitz and Tzagoloff (1980); [5] Bordonné (1982); [6] Martin, R.P. (pers. commun.); [7] Coruzzi et al. (1981); [8] Coruzzi et al. (1979); [9] Miller et al. (1979); [10] Li and Tzagoloff (1979); [11] Thalendorf and Tzagoloff (1980); [12] Michel (1984); [13] Miller, D.L. (pers. commun.); [14] Goursot et al. (1982); [15] de Zamaroczy et al. (1984); [16] de Zamaroczy, M. and Bernardi, G. (unpublished data); [17] Miller et al. (1983); [18] de Zamaroczy (1984); [19] Faugeron-Fonty et al. (1984); [20] Gaillard et al. (1980); [21] Sor and Fukuhara (1980); [22] Martin et al. (1983); [23] Faugeron-Fonty and Goyon (1985); [24] Osinga et al. (1984); [25] Bonitz et al. (1980); [26] Hensgens et al. (1983); [27] Novitski et al. (1984); [28] Séraphin et al. (1985); [29] Macino and Tzagoloff (1980); [30] Colin et al. (1985); [31] Bonitz et al. (1982); [32] Nobrega and Tzagoloff (1980); [33] Lazowska et al. (1980; 1983); [34] Holl et al. (1985); [35] Ooi and Nagley (1986); [36] Tzagoloff et al. (1980); [37] Hudspeth et al. (1982); [38] Maxwell et al. (1986); [39] Frontali et al. (1985); [40] Hudspeth et al. (1984); [41] Sor and Fukuhara (1982); [42] de Zamaroczy and Bernardi (1985); [43] Li et al. (1982); [44] Simon and Faye (1984); [45] Wakabayashi and Mabouchi (1984); [46] de Zamaroczy and Bernardi (1986); [47] Bolotin-Fukuhara et al. (1983); [48] Dujon and Jacquier (1983); [49] Dujon (1980).

<sup>f</sup> This region comprises an inserted sequence (GC cluster 20) in strain D1 [7] (brackets in Fig. 2), which is absent in strain S [6]; GC cluster 19 was a putative cluster [46] later shown not to exist.

<sup>g</sup> This region comprises a revised sequence (11 294–11 365) [12].

<sup>h</sup> Two regions comprise a revised sequence (GC cluster 41 and sequence 23 003–23 038, respectively) [16].

<sup>i</sup> TSL is the tRNA synthesis locus [17].

<sup>j</sup> This region comprises an inserted sequence (brackets in Fig. 2) in *ori1* of strain B [19], which is absent in strain A [15]. GC cluster 59 is a revised sequence [38], initially presented in [20]; (this cluster shows perfect homology with clusters 71 and 79a over a 100-nt length). Sequence 24 304–24 459 corresponds to the first 134 bp of [20]; the last 20 bp remain uncertain [16].

<sup>k</sup> The total size of gaps **9A** + **9B** is about 800 bp. The actual orientation of the sequence bridging these two gaps in wt strains is unknown. GC clusters 65 and 66 were localized by restriction mapping in gap **9A** [46].

<sup>l</sup> This region comprises a mini-insert (GC cluster 70; brackets in Fig. 2) in strains M and D1 which is absent in strains A, K and FF1210-6C [5,41–43]. The primary structure of the *15S* RNA gene was also established in strain D1 [43], and, in part, in strain A [5,22]. Several conflicting data (corresponding to polymorphism and/or sequence mistakes) are not shown. In particu-

lar, an inserted sequence (TTAATATATATATAAAA) is present in strain D1 between positions 26 640–26 641. The paromomycin S → R mutation consists in a C → G transversion at position 26 925 [5,43].

<sup>m</sup> The 45-bp segment in strain M, immediately downstream from the *15S* RNA gene [21] is replaced by a non-homologous sequence in strain A [5,22].

<sup>n</sup> This region comprises an inserted sequence (brackets in Fig. 2) in *ori8* of strain A which is absent in strain B [23].

<sup>o</sup> Additional GC clusters 79c and 84, only present in strain B and A, respectively, [23] and indicated by vertical bars in Fig. 2 are not shown.

<sup>p</sup> These intron sequences account for the differences between long and short genomes (see Fig. 3 in [42]). Several point mutations and/or sequence mistakes exist in some exon sequences of *oxi3* (39 304–39 376 (A5 $\alpha$ ) and 40 742–40 876 (A5 $\beta$ )) from strains D2 [25] and K [26].

<sup>q</sup> This region comprises a short inserted sequence (brackets in Fig. 2) in strain J [27] which is absent in strain D2 [44]. Sequence divergence in intergenic region between the *oxi3* and *oli2* genes (1661 nt) from same strains is 2% (neglecting the short inserted sequence). The oligomycin 2-23r S → R mutation consists in a T → A transversion at position 46 128; the *oli4-r* S → R mutation is localized at position 46 301 [27].

<sup>r</sup> This region starts, in fact, at position 96 of [14]; therefore, GC cluster 120 does not exist (see [46]).

<sup>s</sup> This region comprises an inserted sequence of 100 bp (brackets in Fig. 2) in strain A [5,22] which is absent in strain D1 [31]. Accordingly, the size of GC cluster 132 is different in the two strains.

<sup>t</sup> This region comprises a revised sequence (GC cluster 155 from [46]) and an inserted sequence (brackets in Fig. 2) in strain D1 [36] which is absent in strain J [35]. Sequence divergence of intergenic regions around the *oli1* gene (1300 nt) from same strains is 3.2% (neglecting the inserted sequence). The oligomycin 1-625 S → R mutation consists in an A → T transversion at position 65 264 [35,36].

<sup>u</sup> This region comprises (1) a revised sequence of GC clusters 156, 158, 159 and of the left flanking sequence of cluster 156 [45], and (2) an inserted sequence (brackets in Fig. 2) in strain 992 [45], which is absent in strain D1 [36]. Accordingly, the size of GC cluster 158 is different in the two strains. The 11 'N' nt of cluster 157 probably correspond to sequence CCGGTTGTTCA [46].

<sup>v</sup> This region comprises two inserted sequences (brackets in Fig. 2) in strain D1 [36] which are absent in strain E [37]. A partial GC cluster 160 is present in strain E; an additional GC cluster, homologous to cluster 161, present in strain 1<sup>h</sup> at position 67 742 (not shown), is absent in strains D1 and E [40].

<sup>w</sup> GC cluster 171 is contained in gap **17** [46].

<sup>x</sup> This region comprises a mini-insert (GC cluster 192; brackets in Fig. 2) in strain I [1,49], which is absent in strain 55R5-3C [49]. Erythromycin S → R mutations 354 and 514 (A → G) are localized at positions 109 and 76 586 [47], respectively; chloramphenicol S → R mutations 323 (G → A) and 321 (A → C) at positions 77 320 and 78 519 [48], respectively.

tunity for investigating at the molecular level the open problem of the physiological function of non-coding sequences in eukaryotic genomes.

While a number of small mitochondrial genomes are fully known in their primary structure, only one large genome, that of *S. cerevisiae*, is almost completely known, the majority of the sequences of other large genomes being not yet available in the literature. The mitochondrial 'long' genome of laboratory wild-type yeast strains, is 85 kb in size, is formed for over two thirds by non-coding sequences (de Zamaroczy and Bernardi, 1985), and is known in its primary structure to the extent of 92%. The remaining 8% of as yet undetermined sequence is scattered over 14 gaps (de Zamaroczy and Bernardi, 1985; Table I). We thought it worthwhile, at the present time, to collate and compile this information, which is not easily available because it is scattered in over 60 reports, most of them only concerning small genome segments. We present here an annotated primary structure of the mitochondrial genome of *S. cerevisiae*; the data concern all the available sequences, namely 78 500 bp. About 50% of the data derive from two very closely related strains, D1 and D2 (see Table I for the nomenclature used) that have been essentially studied in Tzagoloff's laboratory. Another 20% of the sequence are derived from two other strains, A and J, sharing a common origin. The remaining 30% of the sequence comes from the mitochondrial genomes of other strains. The sequence presented comprises, in addition, about 7 kb of as yet unpublished results, 5 kb deriving from work of our laboratory on regions flanking some *ori* sequences.

A justification for pooling data from different strains is required. Indeed, the original view of no interstrain differences in the yeast mitochondrial genome was contradicted by the first investigations with restriction enzymes. These led to two unexpected discoveries (Bernardi et al., 1975), namely, interstrain differences in genome size and RFLP. Later work (Prunell et al., 1977; Sanders et al., 1977) showed that the first phenomenon was mainly due to a small number of large deletions/additions, which were localized on the genome map, and shown to correspond essentially to the presence or absence of intervening sequences in some genes; in contrast, RFLP was mainly due to a large number of small deletions/additions in intergenic sequences.

While the large deletions/additions were the results of neat processes not altering the flanking sequences, an accumulation of the small deletions/additions could have generated extensive sequence divergence among different yeast strains.

This possibility was investigated here by making interstrain comparisons of the primary structure of intergenic regions. This has shown that the small (10–150 bp) deletions/additions (several examples of which are presented) are flanked by conserved sequences. In other words, the situation was found to be the same as in the case of the large deletions/additions. Often these small deletions/additions concerned internally repeated sequences, which varied in numbers from one strain to another. Incidentally, similar differences are also known for the *var1* gene (Hudspeth et al., 1982).

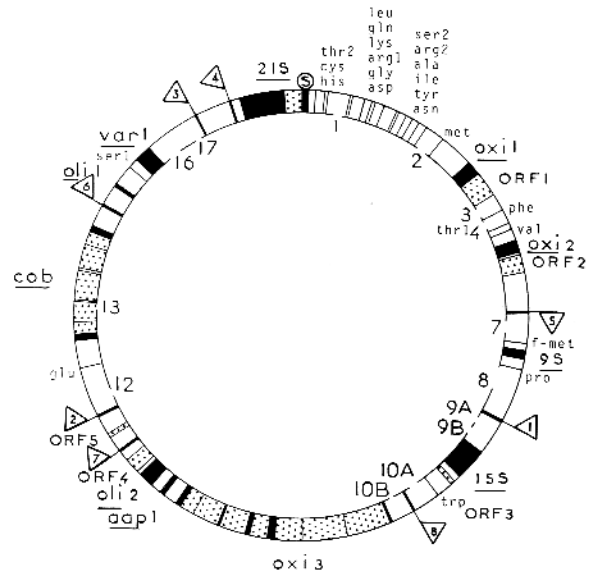


Fig. 1. Physical map of a long mitochondrial genome unit of wild-type *S. cerevisiae*, showing the remaining sequence gaps; these are numbered 1 through 17. Letter S corresponds to the *SalI* site used as the reference point (origin) of the map. In the sequenced regions, black areas correspond to mitochondrial genes or their exons (the last exon block of *oxi3* comprises exons A6-8 and introns a17-8); dotted areas to introns and intergenic ORFs; radial lines indicate tRNA genes (the *thr1* gene is the only one to have an anticlockwise orientation). Among mitochondrial genes, *oxi1*, 2 and 3 encode sub-units II, III and I, respectively, of cytochrome *c* oxidase; *cob*, cytochrome *b*; *aap1*, *oli2* and *oli1*, sub-units 8, 6 and 9 of ATPase; *var1*, a protein associated with the small mitochondrial ribosome sub-unit; *9S* corresponds to the central part of the tRNA synthesis locus; *15S* and *21S* are the genes for the small and large ribosomal RNAs, respectively. Triangles indicate the location of *ori* sequences 1–8; they point in the direction cluster C to cluster A.

Small deletions/additions may also concern GC clusters, but these are rare in the sequences which could be compared. They comprise (i) the presence or absence of a given GC cluster; such a situation has been seen occurring in the *15S*, *21S* and *var1* genes (see de Zamaroczy and Bernardi, 1986) as well as in *ori* sequences (see de Zamaroczy et al., 1984; Faugeron-Fonty et al., 1984), but can also occasionally be found in intergenic sequences (see, for example, the case of GC cluster 20); (ii) the presence, next to a GC cluster of an inserted G + C-rich sequence (which is sometimes flanked by an additional AT sequence; see GC cluster 132). In the case of cluster 158, the atypical G + C-rich sequence (which is 60 bp long) was also found to bridge clusters 168 and 169 (elsewhere in the genome); such a situation results from a recombination event leading to a GC cluster polymorphism.

A third phenomenon detected in interstrain comparisons concerns sequence divergence due to 1–3 bp deletions/additions and point mutations. A first assessment (de Zamaroczy and Bernardi, 1985) bearing on 8% of all intergenic sequence indicated a very limited extent of divergence (2.4%), which was confirmed by the present work bearing on 14% of them. Such divergence is, in fact, overestimated because it certainly comprises sequencing errors. These can be judged from comparisons of primary structures as determined on the same strain by different laboratories. For example, in the ORF4 region, a 0.9% difference was found in 1260 nt, another 0.9% difference being present in GC clusters (Macino and Tzagoloff, 1980; Séraphin et al., 1985). Mistakes often have characteristic features. For example, the sizes of identical oligo(A) or oligo(T) stretches located in runs of such sequences have been differently estimated by different authors. Likewise, GC clusters often show an apparent sequence divergence, which is essentially due to either the loss of very small fragments comprised between close *HpaII* or *HaeIII* sites (which are very frequent in them; Prunell et al., 1977), or to wrong estimates of the numbers of subsequent C's and G's.

The sequence presented here is that of a long genome, which comprises all 'facultative' introns, some long intergenic sequences (around ORF4 and *ori4*), as well as several short intergenic sequences, which are absent in some strains. The sequence presented is, therefore, not that of any given strain.

Fig. 1 presents the physical map of a long mitochondrial genome unit of *S. cerevisiae*, showing the approximate location of genes, introns, intergenic ORFs, *ori* sequences and of the gaps which still exist in the primary structure.

Table I lists sequence data and provides a number of details as well as the references used in the primary structure of the mitochondrial genome of *S. cerevisiae* (Fig. 2).

#### ACKNOWLEDGEMENTS

We are particularly grateful to Claude Mugnier for his constant help with computer programs. Sequence data treatments were performed using computer facilities at CITI2 in Paris, with the help of the French Ministère de la Recherche et de la Technologie.

We thank Dr. R.P. Martin for providing us with sequence data in *ala-ile-tyr* region prior to publication, Dr. D.L. Miller for unpublished data corresponding to gap regions 5 and 6. We thank Dr. G. Faugeron-Fonty for her help in some sequence experiments, Martine Brient for typing this manuscript and Philippe Breton for the artwork.

#### REFERENCES

- Ainley, W.M., Macreadie, I.G. and Butow, R.A.: *var1* gene on the mitochondrial genome of *Torulopsis glabrata*. *J. Mol. Biol.* 184 (1985) 565–576.
- Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roc, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G.: Sequence and organization of the human mitochondrial genome. *Nature* 290 (1981) 457–465.
- Anderson, S., De Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G.: Complete sequence of bovine mitochondrial DNA. *J. Mol. Biol.* 156 (1982) 683–717.
- Benne, R., Van den Burg, J., Brakenhoff, J., De Vries, B., Nederlof, P., Sloof, P. and Voogd, A.: Mitochondrial genes in trypanosomes: abnormal initiator triplets, a conserved frameshift in the gene for cytochrome oxidase subunit II and evidence for a novel mechanism of gene expression. In Quagliariello, E., Slater, E.C., Palmieri, F., Saccone, C. and Kroon, A.M. (Eds.), *Achievements and Perspectives of Mitochondrial Research, Vol. II: Biogenesis*. Elsevier, Amsterdam, 1985, pp. 325–336.

- Berlani, R.E., Pentella, C., Macino, G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system: isolation of mitochondrial transfer ribonucleic acid mutants and characterization of transfer ribonucleic acid genes of *Saccharomyces cerevisiae*. *J. Bacteriol.* 141 (1980a) 1086–1097.
- Berlani, R.E., Bonitz, S.G., Coruzzi, G., Nobrega, M. and Tzagoloff, A.: Transfer RNA genes in the *cap-oxi1* region of yeast mitochondrial DNA. *Nucl. Acids Res.* 21 (1980b) 5017–5030.
- Bernardi, G., Prunell, A. and Kopecka, H.: An analysis of the mitochondrial genome of yeast with restriction enzymes. In Puiseux-Dao, S. (Ed.), *Molecular Biology of Nucleocytoplasmic Relationships*. Elsevier, Amsterdam, 1975, pp. 85–90.
- Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A.: Sequence and gene organization of mouse mitochondrial DNA. *Cell* 26 (1981) 167–180.
- Bolotin-Fukuhara, M., Sor, F. and Fukuhara, H.: Mitochondrial ribosomal RNA mutations and their nuclear suppressors in yeast. In Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), *Mitochondria 1983: Nucleo-Mitochondrial Interactions*. De Gruyter, Berlin, 1983, pp. 455–467.
- Bonitz, S.G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. Sequences of yeast mitochondrial tRNA genes. *J. Biol. Chem.* 255 (1980) 9075–9081.
- Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for subunit I of yeast cytochrome oxidase. *J. Biol. Chem.* 255 (1980) 11927–11941.
- Bonitz, S.G., Homison, G., Thalenfeld, B.E. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. Processing of the apocytochrome *b* precursor RNAs in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 257 (1982) 6268–6274.
- Bordonné, R.: Structures primaires de gènes mitochondriaux de la levure *Saccharomyces cerevisiae*. Thèse de 3<sup>e</sup> cycle, Univ. L. Pasteur, Strasbourg, 1982.
- Breitenberger, C.A. and RajBhandary, U.L.: Some highlights of mitochondrial research based on analyses of *Neurospora crassa* mitochondrial DNA. *Trends Biol. Sci.* (1985) 478–482.
- Brown, T.A., Waring, R.B., Scazzocchio, C. and Davies, R.W.: The *Aspergillus nidulans* mitochondrial genome. *Curr. Genet.* 9 (1985) 113–117.
- Clark-Walker, G.D.: Basis of diversity in mitochondrial DNAs. In Cavalier-Smith, T. (Ed.), *The Evolution of Genome Size*, Chapter 10. Wiley, New York, 1985, pp. 277–297.
- Clark-Walker, G.D., McArthur, C.R. and Sriprakash, K.S.: Location of transcriptional control signals and transfer RNA sequences in *Torulopsis glabrata* mitochondrial DNA. *EMBO J.* 4 (1985) 465–473.
- Clary, D.O. and Wolstenholme, D.R.: The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* 22 (1985) 252–271.
- Colin, Y., Baldacci, G. and Bernardi, G.: A new putative gene in the mitochondrial genome of *Saccharomyces cerevisiae*. *Gene* 36 (1985) 1–13.
- Coruzzi, G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. DNA sequence of subunit 2 of yeast cytochrome oxidase. *J. Biol. Chem.* 254 (1979) 9324–9330.
- Coruzzi, G., Bonitz, S.G., Thalenfeld, B.E. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. Analysis of the nucleotide sequence and transcripts in the *oxi1* region of yeast mitochondrial DNA. *J. Biol. Chem.* 256 (1981) 12780–12787.
- de Zamaroczy, M.: L'organisation, et les origines de réplication du génome mitochondrial de la levure *Saccharomyces cerevisiae*. Thèse d'Etat, Université Paris VII, 1984.
- de Zamaroczy, M. and Bernardi, G.: Sequence organization of the mitochondrial genome of yeast — a review. *Gene* 37 (1985) 1–17.
- de Zamaroczy, M. and Bernardi, G.: The GC clusters of the mitochondrial genome of yeast and their evolutionary origin. *Gene* 41 (1986) 1–22.
- de Zamaroczy, M., Faugeron-Fonty, G., Baldacci, G., Goursot, R. and Bernardi, G.: The *ori* sequences of the mitochondrial genome of a wild-type yeast strain: number, location, orientation and structure. *Gene* 32 (1984) 439–457.
- Dujon, B.: Sequence of the intron and flanking exons of the mitochondrial *21S* rRNA gene of yeast strains having different alleles at the *ω* and *rib-1* loci. *Cell* 20 (1980) 185–197.
- Dujon, B. and Jacquier, A.: Organization of the mitochondrial *21S* rRNA gene in *Saccharomyces cerevisiae*: mutants of the peptidyl transferase centre and nature of the *omega* locus. In Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), *Mitochondria 1983: Nucleo-Mitochondrial Interactions*. De Gruyter, Berlin, 1983, pp. 389–403.
- Faugeron-Fonty, G. and Goyon, C.: Polymorphic variations in the *ori* sequences from the mitochondrial genomes of different wild-type yeast strains. *Curr. Genet.* 10 (1985) 269–282.
- Faugeron-Fonty, G., Le Van Kim, C., de Zamaroczy, M., Goursot, R. and Bernardi, G.: A comparative study of the *ori* sequences from the mitochondrial genomes of 20 wild-type yeast strains. *Gene* 32 (1984) 459–473.
- Frontali, L., Francisci, S., Palleschi, C., Stifani, S. and Zennaro, E.: Transcription initiation and processing sites in the tRNA region of the yeast mitochondrial genome. In Quagliariello, E., Slater, E.C., Palmieri, F., Saccone, C. and Kroon, A.M. (Eds.), *Achievements and Perspectives of Mitochondrial Research*, Vol. II: Biogenesis. Elsevier, Amsterdam, 1985, pp. 203–214.
- Gaillard, C., Strauss, F. and Bernardi, G.: Excision sequences in the mitochondrial genome of yeast. *Nature* 283 (1980) 218–220.
- Goursot, R., Mangin, M. and Bernardi, G.: Surrogate origins of replication in the mitochondrial genomes of *ori*<sup>-</sup> petite mutants of yeast. *EMBO J.* 1 (1982) 705–711.
- Hensgens, L.A.M., Bonnen, L., De Haan, M., Van der Horst, G. and Grivell, L.A.: Two intron sequences in yeast mitochondrial *cox1* gene: homology among URF-containing introns and strain-dependent variation in flanking exons. *Cell* 32 (1983) 379–389.
- Holl, J., Schmidt, C. and Schweyen, J.: *Cob* intron 3 in yeast mtDNA: nucleotide sequence and mutations in a novel RNA

- domain. In Quagliariello, E., Slater, E.C., Palmieri, F., Saccone, C. and Kroon, A.M. (Eds.), *Achievements and Perspectives of Mitochondrial Research, Vol. II: Biogenesis*. Elsevier, Amsterdam, 1985, pp. 227–236.
- Hudspeth, M.E.S., Vincent, R.D., Perlman, P.S., Shumard, D.S., Treisman, L.O. and Grossman, L.I.: Expandable *var1* gene of yeast mitochondrial DNA: in-frame insertions can explain the strain-specific protein size polymorphisms. *Proc. Natl. Acad. Sci. USA* 81 (1984) 3148–3152.
- Hudspeth, M.E.S., Ainley, W.M., Shumard, D.S., Butow, R.A. and Grossman, L.I.: Location and structure of the *var1* gene on yeast mitochondrial DNA: nucleotide sequence of the 40.0 allele. *Cell* 30 (1982) 617–626.
- Lang, B.F., Ahne, F., Distler, S., Trinkl, H., Kaudewitz, F. and Wolf, K.: Sequence of the mitochondrial DNA, arrangements of genes and processing of their transcripts in *Schizosaccharomyces pombe*. In Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), *Mitochondria 1983: Nucleo-Mitochondrial Interactions*. De Gruyter, Berlin, 1983, pp. 313–329.
- Lazowska, J., Jacq, C. and Slonimski, P.P.: Sequence of introns and flanking exons in wild-type and *box3* mutants of cytochrome *b* reveals an interlaced splicing protein coded by an intron. *Cell* 22 (1980) 333–348.
- Lazowska, J., Gargouri, A. and Slonimski, P.P.: The corrected sequence of the first intron of the *cob-box* gene in the yeast strain 777-3A. In Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), *Mitochondria 1983: Nucleo-Mitochondrial Interactions*. De Gruyter, Berlin, 1983, pp. 405–410.
- Li, M. and Tzagoloff, A.: Assembly of the mitochondrial membrane system: sequences of yeast mitochondrial valine and an unusual threonine tRNA gene. *Cell* 18 (1979) 47–53.
- Li, M., Tzagoloff, A., Underbrink-Lyon, K. and Martin, N.C.: Identification of the paromomycin-resistance mutation in the 15S rRNA gene of yeast mitochondria. *J. Biol. Chem.* 257 (1982) 5921–5928.
- Macino, G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system: sequence analysis of a yeast mitochondrial ATPase gene containing the *oli2* and *oli4* loci. *Cell* 20 (1980) 507–517.
- Martin, R.P., Sibley, A.-P., Bordonné, R. and Dirheimer, G.: Yeast mitochondrial tRNAs: structure, coding properties and gene organization. *Mol. Biol.* 17/6 (1983) 1126–1146.
- Maxwell, R.J., Devenish, R.J. and Nagley, P.: The nucleotide sequence of the mitochondrial DNA genome of an abundant petite mutant of *Saccharomyces cerevisiae* carrying the *ori1* replication origin. *Biochem. Int.* (1986) (in press).
- Michel, F.: A maturase-like coding sequence downstream of the *oxi2* gene of yeast mitochondrial DNA is interrupted by two GC clusters and a putative end-of-messenger signal. *Curr. Genet.* 8 (1984) 307–317.
- Miller, D.L., Martin, N.C., Pham, H.D. and Donelson, J.E.: Sequence analysis of two yeast mitochondrial DNA fragments containing the genes for tRNA-*ser* UCR and tRNA-*phe* UUY. *J. Biol. Chem.* 22 (1979) 11735–11740.
- Miller, D.L., Underbrink-Lyon, K., Najarian, D.R., Krupp, J. and Martin, N.C.: Transcription of yeast mitochondrial tRNA genes and processing of tRNA gene transcripts. In Schweyen, R.J., Wolf, K. and Kaudewitz, F. (Eds.), *Mitochondria 1983: Nucleo-Mitochondrial Interactions*. De Gruyter, Berlin, 1983, pp. 151–164.
- Nobrega, F.G. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. DNA sequence and organization of the cytochrome *b* gene in *Saccharomyces cerevisiae* D273-10B. *J. Biol. Chem.* 255 (1980) 9828–9837.
- Novitski, C.E., Macreadie, I.G., Maxwell, R.J., Lukins, H.B., Linnane, A.W. and Nagley, P.: Biogenesis of mitochondria: genetic and molecular analysis of the *oli2* region of mitochondrial DNA in *Saccharomyces cerevisiae*. *Curr. Genet.* 8 (1984) 135–146.
- Ooi, B.G. and Nagley, P.: The *oli1* gene and flanking sequences in mitochondrial DNA of *Saccharomyces cerevisiae*: the complete nucleotide sequence of a 1.35 kilobase petite mitochondrial DNA genome covering the *oli1* gene. *Curr. Genet.* 10 (1986) 713–723.
- Osinga, K.A., De Vries, E., Van der Horst, G.T.J. and Tabak, H.F.: Initiation of transcription in yeast mitochondria: analysis of origins of replication and of genes coding for a messenger RNA and a transfer RNA. *Nucl. Acids Res.* 12 (1984) 1889–1900.
- Prunell, A., Kopecka, H., Strauss, F. and Bernardi, G.: The mitochondrial genome of wild-type yeast cells, V. Genome evolution. *J. Mol. Biol.* 110 (1977) 17–52.
- Roe, B.A., Ma, D.-P., Wilson, R.K. and Wong, J.F.-H.: The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem.* 260 (1985) 9759–9774.
- Sanders, J.P.M., Heyting, C., Verbeet, M.Ph., Meijlink, F.C.P.W. and Borst, P.: The organization of genes in yeast mitochondrial DNA, III. Comparison of the physical maps of the mitochondrial DNAs from three wild-type *Saccharomyces* strains. *Mol. Gen. Genet.* 157 (1977) 239–261.
- Séraphin, B., Simon, M. and Faye, G.: A mitochondrial reading frame which may code for a maturase-like protein in *S. cerevisiae*. *Nucl. Acids Res.* 13 (1985) 3005–3014.
- Simon, M. and Faye, G.: Organization and processing of the mitochondrial *oxi3/oli2* multigenic transcript in yeast. *Mol. Gen. Genet.* 196 (1984) 266–274.
- Sor, F. and Fukuhara, H.: Séquence nucléotidique du gène de l'ARN ribosomique 15S mitochondrial de la levure. *C.R. Acad. Sci. Paris* 291 (1980) D-933–936.
- Sor, F. and Fukuhara, H.: Nature of an inserted sequence in the mitochondrial gene coding for the 15S ribosomal RNA of yeast. *Nucl. Acids Res.* 11 (1982) 1625–1633.
- Sor, F. and Fukuhara, H.: Complete DNA sequence coding for the large ribosomal RNA of yeast mitochondria. *Nucl. Acids Res.* 11 (1983) 339–348.
- Thalenfeld, B.E. and Tzagoloff, A.: Assembly of the mitochondrial membrane system. Sequence of the *oxi2* gene of yeast mitochondrial DNA. *J. Biol. Chem.* 255 (1980) 6173–6180.
- Tzagoloff, A., Nobrega, M., Akai, A. and Macino, G.: Assembly of the mitochondrial membrane system. Organization of yeast mitochondrial DNA in the *oli1* region. *Curr. Genet.* 2 (1980) 149–157.
- Vahrenholz, C., Prätze, E., Michaelis, G. and Dujon, B.: Mitochondrial DNA of *Chlamydomonas reinhardtii*: sequence and

arrangement of *URF5* and the gene for cytochrome oxidase subunit I. *Mol. Gen. Genet.* 201 (1985) 213–224.

Wakabayashi, K. and Mabuchi, T.: Nucleotide sequence involved in the replication of cloned yeast mitochondrial DNA. *J. Biochem.* 96 (1984) 171–177.

Communicated by M.R. Culbertson.

#### NOTE ADDED IN PROOF

Gap 10B located at position 30791 is about 120 bp long in strain K (Osinga et al., 1984). A 63-bp sequence filling the gap has been determined (B. Séraphin, personal comm.) in strain D2:

GTATTCATTAATAATTAATATTATAAATTCAACTATTGTTATATTTATAAATAGAATAATATA

The sequence between positions 30771–30790 (GC cluster 85) and an undetermined 3'-flanking sequence present in strain K are absent in strain D2.





GTAGGTATAAAATTTAATAAGAGTTTTATCCCAATGGAGTAATAATAATAATAATAAAATAAAGGATCTGTAGCTTAATAGTAAGGTACCATTTTG 5500 asp  
TCATAATGGAGGATGTCAGTGCAAACTGTATAGATCCGTATATTAATCTTAATATAAAAAATAAATAAATAATATATATTTAATAAATATTTTC 5600  
TTTATATATAAATATATAAAAAARAATAAATCTTTTTTTTATTATTATATTTATTAATAAATAATTAITTTGGTATATATTATAATTTATATAA 5700  
TTTTATATAAATTTATTTAATCTTTCATTATATATTAAATATTTAAATATTAATCTTTTATAAATAAATAAATAAATAAATTTAATA 5800 e  
TAAATCTCCCTCGGGGTCGGTCCCTCCCTCCATGATAGATAGGAGGGGGCCCTCCTCCCTCGGGGGCCGCGGGGGGGATTTCTTATTTT 5900 14  
TATATTTTAAATAAATAAATTTTATATAAAATTTATTTTCTACAATATATTTACTATTATTTTAAATAATCTTATATATAATATAAAAA 6000  
ATATATATATATATATATAAATAAATAATATATATATATTATTATAAATTTATAATCTTATTAATTAATTAATTAATTAATTAATTAATTAAT 6100  
ATATTATATATATATATATATATATATATATTATTAAATTTTATTTTTTATTTATATTAATAGTAAAAAATCATAAATTTATAAATTTAATTA 6200 (1)  
TATAAATTCATAAATATTTCTCTTTTTTATTATTATTTATTTATTTACTTATTAATAGTTCGGGGCCCGCCACGGGAAGCCGGACCCGAAAGGA 6300 15  
AAATAATAAAAAAATAAATAAATTTAATTAATTTAATTAATTTAATTAATTTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 6400  
TAATAAAGGAAAAATTAACATAAGGTAAGGGGATTTTCTCAAGTAAATGAATGTAAATCTTATGAGTTCGAATCTCATATTTCCGATATATCTT 6500 ser2  
TAATTTAATGGTAAAAATAGAAATACGAATCTAATTTATATAGGTCAAATCTCATARGATATTATTTATATATAAATTTATATATAAATAAAT 6600 arg2  
TATTAATAAATTTATTTATTTATTTAATAAATAAATTTAATAGNNCCGNNNNNNGCCNNNNNNCCGGAACCCCGAAGGAATAAATAAAA 6700 16  
ATATATAAATTAIITATATAAATTAATTTAATTTAATTTATTTATATAAAAAATATAAATTTATATTTTAAATAGGGTTAAATTAATTAAT 6800  
TAATTTTATAAATAAGATAAATAATATATAAAAACTTATTAATAATTTATAAATAAATAATTTATTTACTTTTGATATTTTAACTTTCATTAAT 6900  
ATATTTTATATAAGTAATAAATAGTTAATTTAAATTAATAAATACATAAAGATAAATATAAATAAATATTATATAAAGAAATAAATAATTT 7000 i  
ATATTTTATTTTTTTTATAAATAAATAAATAAATAAATAGGGGGTATAGTAAATTTGGTAGAAGGACCTGGGTTCGATGCATTTAATAGCTCA 7100 ala  
AGTCTCATTAACCCAATAATATATATATAAATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7200  
TAATAAATAATTCATCACTCAAGGAGATATTAATAATAAATAAATAATTTAATAAATTAATAAATAATTTATATAAATAAATAAATAAATAA 7300  
ATTATAAATAAATTTAATAATTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 7400 ile  
TCAAAAATAGGTTCAACTCTGTAGTTTCAATTAATATATCAATTAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7500  
ATAATTAATATATATATATAAATAAATAAATAAATTTAATTTATAAATAAATAATATATGTTCCCGGAAGGCGGAACCCATAAGGAGTTTAT 7600 17  
AATTAATTTAATAA 7700 18  
CCCGACTCCTTCGGGGTCCGGCCCTCCCTCCGGGAGGGGAGCGACTATTTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7800  
TAATAA 7900  
TTATATATATATATATAGATATAAATCTTAAATAAAGTAATATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8000  
ATATAAATA 8100 tyr  
AAAAATTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8200  
TATATAATAA 8300  
TAATA 8400  
TTTTTCTCTCGGGGTCGGGCTCCCGTGGCCGGGCTCCGGAACATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8500 (20)  
AGAGTTATATTTCTATATATTTATATTTTATTTATTTTCTCTCCGGAACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8600 21  
TATTTATCTTTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8700 asn  
ATATATATATATTTTAAATTTATATATATATATATATATATAAAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8800  
AAATTTTAT 8900  
TATAAATTTTATATATATTTTAAAGAGAAACTANNAT 9000  
AAAAATTTTATAA 9100 (2)  
TTTTTTTAAATAA 9200  
TATAITTTAAATAAAGGTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 9300 met  
TAAACATTTGCTCATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 9400  
ATTACTCTCTAGCAGGATTCACATCTCTCTCGGGCGGACTCTCTTTAAAGGGGTCGGTCCCTCCCATAGTATAGAGTATAGGGAGGGGCTC 9500 22-23  
CTCAGTCTCTCGGGTCCGGCCCGGGGGCCGGAATAATTTATATTTATTTTAAATAGATGTTCAATTAATAAATAAATAAATAAATAAATAAATA 9600  
AAATATATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 9700  
TTAAATTAATA 9800 24  
NNNAATAAATTTAT 9900  
TATAAATCACAATTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 10000  
ATAATATATATAAATA 10100 25  
GACCGAAGCCCGAAGGAGTTTATTTTTAGTATTTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 10200  
TTAATTTTTTTTAAATA 10300 26, 1  
TAGTATTAACATATATAAATAAGACAAAAGAGTCAAAGGTTAAGATTTAATAAATGTTAGATTTATAGGATACAAATTAACAACTTATTATGAA 10400 oxil  
GATGTACCAACACCTTATGATGTTATTTTACAGGATTCAGCAACACCAAAATCAAGAAGGATTTAGAAATACATGATAATATATGTTTTATTAT 10500  
TTATTTTAGGTTTAGATCTTGAATGTTATACAAATGTTTAAACATTTCAAAAAATCCCTATGCAATAAATAAATAAATAAATAAATAAATAAATA 10600  
AGTTATTTGAACAATTTTCCAGCTGAAATTTTATTAATTTATGCTTTCCCTCATTTATTTATTTATTTATGATGAGGTTATTTCCAGCACTATA 10700  
ACTATTAAGCTATTGGATTCATGATTTGAAAAATGAAATTCAGATTTTAAATGATAGTGGTGAACCTGCGTGAATTTGAAATCATATGTTTCT 10800





ATTCATRAITAAITTTTTTATATATATATATATAAATTAATITATATATAAAAAAATAATITTTATTTAAATATTTTATCCCTTTAGGGTTC 21700

CGGCTCCCGTGGCCGGAACCCGGAACATTAATTAATAATAAATATATATAAATATTTATTTATTTTATTTAAATATATAAATAAATAAATA 21800

TTAAAAAGATAAAAAATAATGTTTATCTTTTATTAATAATATATATATATATATAAATTAATTAATTAATTAATTAATTAATTAATTAATAA 21900

TATAATATAAATAATAAATAATTTATCTTTTAAATAAATAATANNATATATATAAAGATCTTAAITTAATAAAAAATAAACTTTAATAATATGA 22000

ATTATATATATTTATTTATTTTATAAATTTATAATAAATCGTGAANNNNNNNNNNNNNNNNNNNAATTTTATAAAAAAGTAAGTAATTTTAC 22100

CTTATTAATAAATACCCCTTAATTAATAAATAAATAAAGGGACTTTATATTTATAAAGTAATAGATATATTTATTTATTTATTTATTTATTT 22200

TTTTATTTATTTTATATATATATATAATTAACAGATAGAAGCAAAAGCTCAGGCCCTTCTTTGGGAGAAAGACCTAGTGTAGTTCGAGCTATACC 22300

TATCTGATAAATAATTTAATTTAATTAACATTAAAAAAATAATATATATTTATCATAAATATTAATAATTTTATTAACATACATAAATAAATACTTT 22400

TATTTTATATTTATAAAAAATGAAACCTTCGGGTTCCGTATAAGTTCGGGGGCCGGCCAGGGAGCCGGAACCCGAAAGGAGTTATTTATATATAT 22500

ATAATTAATCTTATTAATTTATATATATATTTAATATTATTTTATATAATTTTATTAAGTATTATAATTTATTTAATATTATTTTAT 22600

ATAATTTTATATTTATTTATTTATTTATTTTATTTTAAACCTACTATAATCATATTTAATATATTAAATATATTTTATATTTATATCTTTTATG 22700

GTTATATATATATAAGGTTTATAAATATATATATATATATATAAATAATCAATATATATTTATTTATTTATTTATTTATTTACTATITTTAT 22800

TATATTAATAATATATATTTATTAATTAATGGGTACTTAATGGTATATTTATTTTAAATAAATTTATGATTTATGAATAAATAAAGTAGGA 22900

ATCGGTACGAATACGAAGGGAGTTATATTTAATTTTATAATTTTATATATTTATTAATTTAATAATTTTATATATTTATAATTTAT 23000

TATATACATAGGATAGATAGATAGATAGATAGATAGGGTCCCATTTATTTTACAATAAATAATTAATGAGGCCGATATCTTTTGTGTTAT 23100

AAATTTATATATATTCATATTTAATAATATATTTAATTAATAAATTTATTTATTTATTTATTTATTTAAAAAATAAATACTATTTATTTT 23200

TAATAATAATAAATTTTATATAATTTATCATTTTATTTATTTCTCTCTTGGGGTTCGGGCTCCCGTGGCCGGGCCCCGGAACCTCCCTTCCCT 23300

TCCGGGNNNTTACACCTTTATAATTAATAAAGGGTTCACATATATTTATTTTATATAAATTAATATATATATATATATATATAIACTT 23400

TTTTTTAATAATATATCTATATAAATAAATAATTTATATATATTTTATATAAATAATATTAATTTATTTTAAITTTCTATTTCTATTTGG 23500

GGGTCCCAATTTATTTTCAATAAATAATTTATTTGGGACCCGGATATCTTCTTGTTTATCATTATTTTATTAAATTTATTTATTTTAAATTT 23600

ATATTTATATATAAATTAATATATCGTTTATCTCCCTTCGGGGTTCGGTCCCTTCCATTAGTATAGAGTATAGGGAGGGGTCCTTCACTCCCTTC 23700

GGTCCCTCCCGGGGGGGGCGGACCGGAGTATTAATTTAAATAATTTTAAATGAATTTAATTAATAAATAAATAAATAATTTTAAAGATG 23800

AAATAAAAAATAAATAAATAAATAATTTAAGGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 23900

ATATTTAAGTTCGGGGGGGGTCTTTATTTATTTATATAATATATATTTATTTATTAATATATTTTATTTGATTATATATAAATTTATTTCTATG 24000

TGTTCTATATATTTTAAATTTCTGGTTATIGATCACCCACCCCTCCCTTATAAACTTAAITTTACTTATATTTTATAAATAAATCAACTT 24100

AATTAATAATTTAAATATATCTTAAATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 24200

ATTTTCTTTTTTTTCTTTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 24300

AGCGGGCCNNGAACTAAATAAATAATATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 24400

AATTAATAITTTATATGAAAAATATATCTTTTATTTATATCTCCCTTCGGGGTTCGGTAAATAAATAAATAAATAAATAAATAAATAAATAAATA 24500

ATATGGTCTAAACCTCTTTTATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 24600

TATTTTAAATA 24700

TATATAAATA 24800

ATTTTATAAATA 24900

CTTTCCGGGTTCCGGCTCCCGTGGCCGGGCCCCGGAACTTAAATACTCCCTTCGGGGTTCGGTCCCGCCGTAAGTATACGGGGGTCCTCACTCCCT 25000

GGCCGACTCTCTCTTAAAAAGGGTTCGGTCCCTCCCTCATTAGTATAGAGTATAGGGAGGGGTCCTCACTCTTCGGGGTCCGCCCCGGGGGGCC 25100

GTCGACTAATTTAACTTTAATATTAATTTAATATATTTATTTTAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 25200

CAATA 25300

ATTTTATTTATTTATAAATA 25400

AATTTATAGTTATATAAAGGAAAGTAAAAAATTTAAGAATATGATGTGTTCCAGATTAAGCGCTAAATAAGGACATGACGCATACGGATCAT 25500

ACGTTTATTTGATAGATAAATAAATGTTGGTGAACGTGAGTAAATTTATAGGAATTAATGAACATAGATAAGCTAAATACTTAATATATAT 25600

TATATAAATA 25700

AAPCCTAGCCACAGTCCATAATCGATAAGGATGAAGGTAGAACACTACCGTTGACTCGAATAATAGTCAATATCATAAATACAGCAGTGAGGAATTT 25800

GGACAATGATCGAAAGATTGATCCAGTTACTTTAGGATGATATATAAATAATTTTATTTTATTTAGTTCCGGGGCCCGCCAGCGGCGGAACCCGA 25900

AAGGAGAAATATAAATA 26000

TACCATATATTTTATATCGATATATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 26100

GTCGGGTAAACAAAAGAGGGCGAGCGTTAATCATAATGGTTAAAGGATCCGTAGAATGAATTTATATAATTTAGAGTTAATAAATAAATAAATAA 26200

GAATTAATAAGTAAAGATAAATA 26300

AGTGGAAACGGATTTCGATTCGTTGAGTTCTAGTAGTAACATGAAACAATTTTATAATATATATATATAAATAAATAAATAAATAAATAAATAAATA 26400

GATTTCCACCTGAAGAGTACGTTAGCAATAATGAACCTCAAAACATAGACGGTTACAGACTTAAGCCATGGAACATGTTAATTCGATATACCCAG 26500

ACTAACTTTACCATTTTTTGAAGTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 26600

AAATGTGATAAACGAGCAAAACCTCATATATAAATTTATATTTATTTAATAAATAAAGAAAGAAATTAAGACAATAAATAAATAAATAAATAAATA 26700

GGTAATAGACGTGCTATAAATA 26800

TTTTATATATATAAATA 26900

CTGTTTCGACATACTCATCACCGGTTGAACAATATATATCTTATTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 27000

55

21800

21900

22000

22100

22200

56, 57

pro

22300

22400

22500

22600

22700

22800

22900

23000

23100

23200

23300

(b), 59

23400

23500

23600

60

61a

(61b-61c)

23700

23800

23900

24000

(61d)

62, (i)

24100

24200

24300

63

24400

24500

64, (9A)

24600

24700

24800

24900

67

(9B), 68-69

25000

25100

25200

25300

25400

25500

25600

25700

25800

25900

(70)

26000

26100

26200

26300

26400

26500

26600

26700

26800

26900

27000





TTCAGGCACATTGAGGACCTAGTGTAGATTABCAATTTTTGCATACATTTAAACATCAATTCATCATTTAGGTGCTATTAATTTGCTGTAACAAC 37900  
 ATTAATAATGAGGAACAATGGTATGACCAATGCATAAATACCATTATTTGATGATCAATTTTCATTACAGCGTCTTAATTTATTTATTCATTACCTGTA 28000  
 TTATCTGCTGGTATTACAATGTTATTATTAGATAGAAACTTCAACTCTTCATCTTTGAAGTAGCAGGAGGGTGTGACCCAACTTATACAGCAGCTTAT 38100  
 TTGATCTCTTGGTCAACAGTGGCCCTTATTTATATATAATAATATATATGATATGCATTTTTCTAAATGCTGGAAATATTAATAAAATGAAATAC 38200  
 AAATGATTAAGGCTATTAITTTAAAGGCTATTTGTAAGAAATTCATCATCTTATAAATCAGCAGGATAGATAAATAAATGATCTTATTTAAATAAAC 38300  
 GATAATATAAAAGACTCTCAGAGACTCACAAGAAAATATAAATAATTCATAAATAAATAAAATTTAATCAATGATTTAGCTGGATTAATTTGATGGTGAT 38400  
 GATATTTTGGTATTGTAAGTAAGAAATATGTATCATTAGAAATCTAGTAGCATTAGAAGATGAAATAGCTTTAAAAGAAATCAAAAATAAATTTGGTGG 38500  
 TTCTATTAATAATAGATCAGGTGTAAGGCTATTAGATATAGATTACTTAATAAAACTGGTATAATTAATTAATTAATGACGATTAATGGTAATATTAGA 38600  
 AATACTAAAAGATAGTACAATTTAATAAAGTTGTATTTTATTAGGATTTGATTTTATTATCCAATTAATTAACATAAGATAATAGTTGATTTGTTG 38700  
 GATTTTTGATGCTGATGGTACAATAAATTTATTCATTTAAAATAATCATCTCAATTAACAAAACCTGTAACATAAATTTTACAAGATGTACAAGA 38800  
 ATATAAAATATTTTAGTGGTAAATTTATTTTGATAAATCACAATAATGGTTATTATAAATGATCCATCAATCAAAAGATATAGTATTAATTTTATTT 38900  
 AATGATATATTAATAAATCCATCAAGAACACTAAAAATAAATAAATATATTTAAGTAAGAAATTTATAAATTTAAAAGAAATTAAGAGCTTATAATA 39000  
 AGACTCTGATTAACATACAAATTAAGGATGATTAATTTTTGAAAATAAAGAAAATAAATAAATTTAATAAAGATATAGTCCAAATTTATATATA 39100  
 TATAATATATATATATAATAAAGAGCCCTGAAGACTATATATTTAATTAATTCCTGGATTTGGTATTTTCCACTATAGTATCAACATATCTCAAAAAC 39200  
 CTGTATTTGGTGAATTTCAAGTGGTATTTGCTATGGCTCAATTTGGATTTATTAGGATCTTAGTATGATCAGCATATATGATATTTAGGATTTAGATGC 39300  
 AGATAGTACAGCATATTTCCCTATCCGCACTGATGATTTGCAATTTCCAACAGGAAATAAAATCTTTCTTGATTAATAAATCCCTTTAGCAGGATAAA 39400  
 AATAAAAAATAAAAAATAAAGTTGATCAGAAATTTATCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 39500  
 ATATAATAAAGGTAATTTATATGATATTTATCCAAGATCAATAGAAATATATTTCAACCATAAATAAATAAAGAAATAGTAGATATGGTTATAA 39600  
 TTTAGAACTCTGGTGGTATACCTCTATATACATAATATGTAACAATATAGTAGGATTTCCCTAATAATTTTTATATATATAAAGGATTTTATTA 39700  
 ACAGATGGTTCGATGATATCTATCAAAAAGATTTAGATAAAAAACAATTAAGAAATTAATTTGATGATTTAGATAAAAAACAATCAATAATCATATA 39800  
 GTGAATATTAATAATGATTTTATTTATTTATTCACATATTTGATAAAGTTATCCTCAAAAATAAAATTTGCTAAAGTTAAAGGTAAGGATTAATCAAT 39900  
 AGAATTTATACTAGATCAATCCATGTTTACTATTTAAGATATATATTTATAAATGGTAGAGTAAAAATTTGACCTAATAATTTATATGATTTATTA 40000  
 AATATGAATCTTAGCTCATATAAATTTATGATGGTTCATTTGTAAGGTTGGAGGTTTATATTTAATTTTACAATCTTTCTTAACTAAGAAATTA 40100  
 TTTTATTTATAAATTTTAAAATAAATAAATTTAATTTAAATTTGCTCTATTACATAAATCTAGAAATAAATATCTTATTAATAAGAGTAAATCTGTTAA 40200  
 AAGATTAATTTCCCTATAAATTTATAAATATATTTTACCCTCTATAAGATATAAATTTGATATATATTTGCAAAAAAATAAATAATGATTAATTTA 40300  
 ATTAATTAATTTAAITTTATTTATTTTACTTTTTTGATATATATAGAGGCAAACTCGAGGAAAACCATATAAATAGAAATAAGTAATTAATTTATAGCAAC 40400  
 CGTCGAATCAATCATATTCAGAATAAATAATGTAAGAGGCTAGAGATTTAGACCCCTCTGGTTATCTAAGTAATATATATATATATATATATATGATAAC 40500  
 ATAGGATATAATCCAATGAGATCAGTAAATGATTTAAAACAATAAATTTGTTTTAAGTATTAATAATAAATTAATATTTCCAGCTCTTAATTTGAGGATAT 40600  
 TATAAATCAATATTTTATATATAAATAAATAAATTTAACTAGCTAGATAAATTTATATAAATAAAAAAATAAATTTATATAAATAAATTTAAAATAATTT 40700  
 TTATTAATTTGAAACTGAATTTGTTTAAAGTTAAATAAAGAGCTCAATTTCCATGGTGGTTCAAATAGATTTAGCACATCCCTATGTTATATGCAATTTGCATTT 40800  
 CTATTTCTTATTTCACAATGGGTTGTTTAACTGGTGTGGCTTAGCTAACGCTCATTAGAGTAGGACTTCCAGATTTAAATTTAATAAGTGTGGTCTT 40900  
 AAAATTCACTAAAATAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 41000  
 TAAAAATTTAATAAATATTTGGTTATATATATATATATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 41100  
 TTTATATAAATAAAGACAATCCCGTGATAACTTTAATAAT 41200  
 AAATATTAATAAATAAGAGATAGTCCCACTTATATGTAATAATAAGTTAAATCCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 41300  
 AAATAA 41400  
 GGTCGGCCCGGAACTAAGAGATATAGAAATTTATAGATCAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 41500  
 ATAAATTAATAA 41600  
 ATGTTTTCAATTTACTTTTTCTTTAGAAAGTGTTTTTATAGCTAATTTATGCTTTTTATTTAAAATTTGGTCATATACAGCTAATAATAAATTTAAT 41700  
 AAAGAATTAACAGCTGTAATAAGAAATTTATAAAAAAATAAAGAACAAAGAGTATTTATAAATTTATTAATGGTAAATTAATAACATATAAAGAGATG 41800  
 ATCAATATTTAATAATAAATTTAATAAATCGTTTAAATTTAATTTAATTTAACCATAAAGAAATTTGATTTACTATTAATCCTTGATTAACAGGTTTAA 41900  
 TGATCTGATGGTATTTTTATCTAGGTTTTCAAAAACATAAATAAATGCTAATGATTAATAATTTCAATTTAGAAATACACAATAAAGATAGTTATATTTA 42000  
 GATATTTAATAAATAAATTTTAACTTGGTGGTATTTTAAAAGAGATATATAAATCTGGTCTACAGCTTATATTTATAAAGAGCTCAATCATCAAAAGCTA 42100  
 TAAAACCTTTTATGAAATTTTATAAATTTATCAACCATTAAGCTTTAGAAGATATAAACAATATTTATTTAATAATTTGCTTACTTATTAATAATTA 42200  
 TAAATTTACATATATTACTTAACTTTTATTAATAATTAAGAAATTAATTTATTTACAAGGTGTTAAAATAATATCTTTAGAAATAAATAAATAAATAA 42300  
 ATAGAGTTAAAATATTTAATAAATCTATATAAATAATGAAATAATGATAAATTTAAGAGATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 42400  
 TTTTTTTTATGACTACTACGTTGGGGGACTTTTCAGGCGCTGAAAGTTATCATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 42500  
 TCAATATAAGTCTAATTAACAAGTGTATTAATAATGGTAAACATAAATGCTTAAGCTGTAATGACAAAAGATCCCATATCTTGACAGTTATTTATATAT 42600  
 AAAAAAAGATGAAGAGACTTTGACTGATCTAATGCTCAACGAAAGTGAATCAAAATGTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 42700  
 AITCAATTTATTTATTTATTTATATAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 42800  
 TAATATAAATTTGAGAGCTAAGTTAGATATTTAGCTATTTATGATAAACAAGATAAACCCTATAAATAAATAAATAAATAAATAAATAAATAAATAA 42900  
 CCAATATATATATTTAATTTATTTATTTATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 43000  
 TTTCTTTATTTTGGAGGTAAGAGATGATAAAGGCTAATGCCATTTGTAATGATGGATAAGAAATTTATTTCTTAAAGATGAATAAATAAATAAATAA 43100  
 TTACTATAGGATGATGCCATCTTTATTTATATATATTTATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 43200

ORF a14

CRF a14  
A5α

ORF a15α

CRF a15α

93

A5β  
CRF<sub>1</sub> a15β

94

95

ORF a15β  
96

CRF<sub>2</sub> a15β

e  
A5γ,  
CRF a16



AAATATTATATATATTTTTTATAAATAAATAATAAAGATATCGGIGAGCCGATGCGATGAAAGTCGCACGTACGGTCTTACCGGGGGAAAC 4330D  
 1TGTAAAGGCTACCTACCGGGATACATGATGATTAACAATGGGTGCTATTTCTCTTTATTTGCAAGTACTACTATTGAAGCTCCTCAAAATTTAGGTTT 4340D  
AAACTATAAAGAAATTAGCTCAAATCAAACTCGATTAAATTTTCATCGGGCTAAATGTTATTTCTTCCCAATGCATTCTTAGGTATTAATGGTATG 4350D  
CCTAGAAGAATTCCTGATTATCCTGATGCTTTCCGAGGATGAAATATGCGCTTCTATGGTTCATTTCGCACTATATCATTTATCTTATTTAICT 4360D  
 ATATTTTATATGATCAATGTTAAATGGATTAATAATAAAGATTAATAATAACAGTATTTATGCTAAAGCACCGGATTTGTAGAATCTCAATCTTAT 4370D  
 CTTTAAATTAATAACAGTTAACTTCATCTACCGAATTTCTAACTTCTCCACAGCTGACACTCATTTAATACACACAGCTGACAACTTCTAAAGTT 4380D  
 ATAAATTTAAATATTTACTTAATAAATTAAGAAAAAAGTAAATATATATCTAAACCTTAATAATAAATAATAAATATCTTAAAAAAAATATATA 4390D  
 AAAAAAATATATAAAATTTATTAATAATCTTAATATATCTCTTTTGGGAAGCTATATAAAATAACTAATAATAAATCTATTATATATATATA 4400D  
 TATAAAATAATATATATATAAATAAATAAATAATATTTATAATATTTTTTAATAATATATAAATTAATATATTAATTAATATATATAATAT 4410D  
 TAAATATATATAAATATATATATTTTATAAATAAAATATTTTTAATACTAATATATTTATTTATTTATAAATATATAAATAGTAGTGTTAATATTA 4420D  
 TTAATACTAAAAAAGTAAATATAAATAGGATCTACAATACATTTATCTGATAAATATAAATATTAATATTTATAATATTTATATTAATAAAGGGATTA 4430D  
 ATTAATGTATCCAATTTAATTAATATAGATATATTTTATAAATTAATAATATATGTTTTAATAAAGGTAAAAATAGTTTTTATTTATATATAA 4440D  
 ATATAGGATATAAATAAT 4450D  
AGGATAAT 4460D  
 TGTAAATTTAATAAATTTTATTAATAATTTAATAAATAAATAAATTAACATTTAATAAATTTATATATATGTCAGTATTTTATATTAATGTTTAACT 4460D  
 ATTAATAATTTTTTTTTTATAAATAATATATTTATATATAAATAATATATATATTTTTTATAAATATATATATTTTTTATAAATATATATATTTTT 4470D  
 TATTAATTTATATAATGTTGTTTAACTTATTAATAAATATATATAAATAAATAGTCCATTTTATTTATGAATCAATTAATATATGTT 4480D  
TTCTTATTAATGATCTATTTAATTTTTATCTCCAAATTTCTTTTTACTATGATCTTAAGATATATGATCTAGATTTATTTATTTCAAAATATAAT 4490D  
 ATATATTTAATTTTTTTTTCATATAAATATATATATATAAATATTAATAATATTTACTTATTAAATAAATAAATAAATAAATAAATAAATA 4500D  
 ATTAATAATTTAATATATTTAATATATTTCTCTACAGCATATATATTTAT 4510D  
 TAAAGTATATATATAAATAAATAAATAATATATATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4520D  
 GAATATAATTTAATAATTTAATAAATAATTTATATAGTCCGGGCGGGCCCGGGCCCGGAGGAGGATTAATAAATAAATAAATAAATAAATAAATA 4530D  
 AATAATATATATAAATA 4540D  
 TATAATTTAT 4550D  
TATATTTAT 4560D  
 TTTATATGTTTAAATTTAATAACAT 4570D  
TGTTTAAATTTAACARCATTTCTCATATATCTATATATGAT 4580D  
TCAGAGTGAATTTTCCACAAGAGCTATTTATGATACTTATATAAATGCTTAAGGACAAATGGAGGTAAAAATGAGGTTTATATTTCCCTATGA 4590D  
TCCTTACATATTTATGTTTATTTTATATGCTAAATTAATATAGTATGATTTCCATATTCATTTGCAATACAGCTCAATATAGTATTTATATACCTTTAAG 4600D  
TATTTGTTATTTGATAGGTAATACTATTTTAGGTTTTATATAACATGGTTGAGTATCTCTCCATTTCTGTAACCTGCTGCTACACCATTACCATTAGTA 4610D  
CCCTTATAGTATATTTGAAACTTTATCTTATATGCTAGAGCTATTTTATAGGTTTAAAGATTAGGTTCTAATATCTTAGCTGGTCATTTATTAATGG 4620D  
1TATTTAGCTGGTTACTATTTAATTTATGCTAAATTAATTTATCTTTAGTATTCGGTTTTGACTTTAGCTTTAGCTTTAGCTTTAGCTTTAGCTTT 4630D  
HGAATTCGTTTGGTATTTCCCAATCTTATGCTTATCTTAAAGCATCAT 4640D  
TTATATAAATA 4650D  
ORF 4  
 ATTAATATATATTAATATTTTTAATAATAICATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4660D  
 ATATACTATGTTTTAATAAGAGATATAAATAAATTTATATATATATAGATATATATATAATGATATATAAATATATAAATATATAAATATATA 4670D  
 ATCTAATAAAGAGATTAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4680D  
 111TTATATTTAATA 4690D  
 AATAATTTAATTTATAAATATCTAATAATGAACTAATTTATCTAATAATTTTTAATTAATAGATAAATAAATAAATAAATAAATAAATAAATAAATA 4700D  
100  
TCCCGGCCACCGGGGACCCCGCAAGGAGAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4710D  
 AAACAGTATAGATTTAATA 4720D  
 TCAATGCTAATGATGATTAATAACTATCTTTTAAATAAAGAGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4730D  
 TTTGTTAAATGCTTTCTCTATTAATTAAGCTTAAATTTAGCTTAAGAAAGAAATTTTCTCTAATATTTATAAATAAATAAATAAATAAATAAATAAATA 4740D  
 AAATTAATAATCATATCCCTTAAATATTAATTTTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4750D  
102  
CCCCACAAGGAGATATAAATA 4760D  
 ATTTAAAGATATAAAGACTAGTTATGTTATTTCCACAGTGAACACGTAAGAATAAATCTAATTTAATCAAGAATCTTTGATTTATCTATTTCTAAT 4770D  
 GTTAAAAAGTTGGTAAATAGAAAAATAAAGATTTAAATTTATTTACCAGAACTCTTGTGATTAATAAATAAATAAATAAATAAATAAATAAATAAATA 4780D  
 CTTTACATGATAAATA 4790D  
 TGAATATATCAATA 4800D  
TAATTAATTTCTTACTATATAAATATCTATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4810D  
TTAATAATTTCTTCTTCAATAATATCTATTTTATTTAT 4820D  
 TTTATAAATA 4830D  
 TATATAAATA 4840D  
 ATATCTCATTAATATATAATAGTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4850D  
 ATAAGATTTATATAAATA 4860D  
103

A6

a17, A7  
a18, A8

e

e ,97

cap1

98, e

99, e

oli2

ORF 4

100

101

102

e

AATATTCATTTTATTTTATATACAAATAAATATATTTATTTAGGATAAIAAAAAAAAAAATAAATAAATATATTTATTCATTTACTTTTATCA 48700  
 TTATTTATTAATAATTAATATATACAAITATAATATATAATATATTTATATAATATATAITTAATATTTAAATATAITATITATTTATTACTCTC 48800  
 ATGGAACTTATATTTTATGTCGGGCCCCCGGGGGCGGACCCCAAAGGAGATATTTTTATTTATTTATTAATAATTTATAATGTTATATTTTGGATTT 48900  
 AATAATATATACGGGTTCCTTACGAACCGGACCCTGGAGACGTAATATGGGGAGGGGGTGGTGATAATAAACCAGAAIATAAATAAATACAGAGAC 49000  
 ACAITAGATAAATTTTATAATATAACCAATATATAAATTTATAAATAATTAATTATATAATAAATAATGTATARAACCAATATAAATTAATGATATAA 49100  
 TATAAATAATAA 49200  
 TTGGGACCCCCACATAAATA 49300  
 TATAATATTTATAA 49400  
 ATAAATAATTAATA 49500  
 ATGATATAAATAAGATAA 49600  
 TATAAATAATAA 49700  
 TCCATAAAAAATTCCTTTGGAGGATTTTAAATAAATAAATACATCAATATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 49800  
 GTAATACGGGTAAACATTCACCGTGTTCACGGGTAATGTTTACCCTATTTATATAAATCTTAATAAATAAATAAATAAATAAATAAATAAATAA 49900  
 AATAAATTTATTAATTCCTCTTCGGGTTCCGGCTCCGGTGGCCGGAACTCCGGAACTATAAATAAATTTAATAAATTTATATATTTATGATTAAT 50000  
 AATAATATTTAATGTAACCTCTTCGGGATTCGGTCCCTCGTAAGTATATAGTATATAGTATATAGTATATAGTATATAGTATATAGTATATAGTAT 50100  
 TCGGGGTTCGGTCCCTCTACGGGTACGGATACGGATATAAAGAGTCCCTCACTCCCTATCACTNCGCTGAAGTGGAAATTTATTTATATATA 50200  
 TTTATAAATTTTATTTATTTAATTTATATATTTAAATATATATTTATATAAATAAACCCTAATTTATTAATGTTATATTTAATAAATAAATAAATA 50300  
 TCTTAAAAATTTTATAAATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 50400  
 GAAAGCGTTTAAAAATAA 50500  
 TACATATCAAGACCAATTCGAATGAACTATTATTTAATCTCCCTCCCTCCCACTATTTATAAGTACAATTAGGGCCCAACCCCGGAG 50600  
 TGTATTTACITGGCAATGTTTATCCCAATTAATATAAATACAGAGATTTAATAATTTATTTATAAATAAATAAATAAATAAATAAATAAATAAATA 50700  
 ATAGAACATTTATGTTGTTTACCAGTGTTAAGAATATTAATCCCAATTTTATTTAATAGTGAAGATATATTTTAAATTTAATGAAATCCATATAT 50800  
 ATTATTTAATAA 50900  
 AATA 51000  
 ATATATAA 51100  
 TTTGTAATATAAAGTATARAATTAAGITTTATAGGGGAGGGGTGGTGAATAGAACTAGAAGTAAATAAATAAATAAAGCATACATTAATCAAT 51200  
 ATTTAAATA 51300  
 ATCAATAAATAAGCTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 51400  
 TATAGGAATA 51500  
 TATTTAATA 51600  
 AATATAAATA 51700  
 TTTAATA 51800  
 ATATAAATA 51900  
 TAATA 52000  
 GGTTCGGGCTCCGTGGCCGGGCGCCGGAACATAAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 52100  
 CACTGGCAATGTTTATCCTATTTGATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 52200  
 TTAATTAATA 52300  
 TACTTTTATAAAGAGGAGTAGGAGACCCCCCTTACGGGGGGGAACCGAACCCCGCAGAGATTTATTTTAAATAAATAAATAAATAAATAAATA 52400  
 TAATA 52500  
 AATATATCAATA 52600  
 CACCTATTATATAAATA 52700  
 AAATA 52800  
 TCCGGGCCCCCGCCAGGGAGCCGGAACCCGAAAGGGGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 52900  
 ACAAGAATAAATAAGTCGGTGGATCGAACCCCTTTTATTTAATAATTAATAATTTAAGAAGGAATGTTTATATATATAAATCTTATTTGGGGAT 53000  
 TAATAAATAAATAAAGTTTGGATACCAAGCCCAAGACCGGAATCCCAAGGAGATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 53100  
 TTTTAAAAATA 53200  
 ATAATA 53300  
 TATTAATAATTAATA 53400  
 TATAAGGTAATATAAATA 53500  
 ATAAGATAAATA 53600  
 TTTCCGAAGGAGTAGATAGATCCCTTGGGGGGACCGAACCCCTTAAAGAGGAGTGGAAACCCCGTGGGACCGAACCCCTTTTATTTAAGAAGA 53700  
 AGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 53800  
 AATAACATTAATA 53900  
 CTAATGGTACGACATCACCTTCTCATGTGATAATATCGGTCGATCCGATTAAGGTTATTCATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 54000

104  
 105, 106  
 ori 7  
 107  
 108  
 i, 109  
 ORF 5  
 110  
 111  
 112  
 113  
 e  
 (i), 116  
 ori 2  
 117, 118  
 119, 120  
 121, 122  
 123  
 124, 125  
 126, 127  
 128, 129  
 130  
 131  
 132  
 glu



CAATGAAAATTTATAATAAAACTCTATATATTTTATTATTCAAATATTTTAAATACAAAACAATTAATAATTCGTATAAAATTTAATGAACAA 59500

AGCAATATAAIAAAAATAATATGVAAGTGATTTAATTAATCCCAATAGAGTAAATATATATATAAAGAGATAATCAGCAGGTAAACCAATAAATTC 59600

TCTAAATACCAACTACCGAGTAAATAAAGAGTTTATTAGTAGACTACTCAGAGACTACACCCACACTAGAAAAATAATTTATTTATTTAAATAT 59700

TTAAATATAAAAAATAAATAAATAATCTTAAAGAGCTATAGATTTTAAAGAGTAGTAATATAGATTTAAACCAATGATGGCCGGTTTAAATG 59800

ACGGAGATGGTATTTTGGTATCTAAAAATAAATAGCATCTTGTGAAATCTTGTAGAAATAAAAGATGAAAAATGTTAAGACAAATCCAAGATA 59900

ATTTGGTGGTCTGTAAAATTAAGATCAGGTGTTAAGACTATAGATATAGATACAAAATAAAGAGGTATAAATAAAATAATTAATGCCGTTAATGGT 60000

AATATCGTAAATAGAAAAGATTAGTACAATTTAATAAGATGATTTTATAAATATCGATTTTAAAGACCTATAAATAACTAAAGATAATGCTT 60100

GATTTAATAGGGTCTTTGATCGTGATGGTACTAATTAATTTATTTATTCGGTAAATAAAATAAGACCTCAATTAACATTTAGCGGTTACAATAAATA 60200

TTTACATGATGTTGAATCATATAGAGAAGTATTTGGTGGTAAATTTATTTTGTATAAGCTAAAATGGTATTTTAAATGATCTATTAAATAAAGAA 60300

TTACATAAATTTTATCTTTATAATAAAGTTGTCCTTCTAAATCTAATAAAGGTAAACGTTTATTTTAAATGATAAATTTTATTTATTTATGAT 60400

TATTAGCTTTTAAAGCACCTATAAATCTGCTTATATAAAGCTTGATTAATAATTAAGTAAATGAAATAAATAATTAATTTTCCGATTCATTA 60500

TATATATCTCAATTTATAAATAATTTAAAGATTCCTTATAATAATAAAGCATCTTGTAAATATTTGTTAAAGATAATAAATAATTTATTAATGAATCGG 60600

ATTATATTTTACAATCTTATAAATAAATCTGATCAATAAACATGATTAAGAAATAATAAATGTTTATGAAATAAGATAGTGTATATAAATTTT 60700

ATGAGATATAGTCCATTTTATTTATTTATAAAGCTCTGATAACTATATCTCGTAACTCTTTAGTAAACACCAGCATCTATGATTTAAAAATA 60800

TTAATAAATAATTAATTAATTAATTAATTTATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 60900

TTTATATATATCTCCTTTATTAATTTATTTATATATATTAATTAATAATGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 61000

TAAAAGTTCAGTAGCATTAATATCTTAAAGTTTCCGGTCTTGTGGGACTCCCATAGGAGTTFATGATAAAATGGTAAATTTGCTCARGAAAATCTA 61100

AGGTATAAATA 61200

AATATAAATA 61300

AAAGATAGATTTAATTTTTTTTTATATAGTACCGGNNNNNNNCCGGGAGCCGGAACCCCGAAGGAGTAATATAATATATAATAAATAAATAAATA 61400

ATATATAATATAATA 61500

TAATCTTATGTTAACATAATAAATTTGTTGACTGAAATGATACTATACCATCTATGCTATTTAAGACTCTATCCCTGATAAATAATAGGAGTTAT 61600

CTAATGTTTGGACCTATTTTAGTATTATAGTTTACCATTACTGATAGAGGTGATGTAAGAGGTAATACTTTTAAAGATAAATAAATAAATAAATAAATA 61700

TTATCTTTGATTTCAATTTGATTTATAGGACAAATGGAGCATGCCATGTAGAAGTACCTTATGCTCTAATGGGACAAATCGCTACATTTATCTACT 61800

CGCTATTTCTTAATTTATGTAACCTGTTACTCTACTATTTGAAATGTTTATCTTATATCGGTAGAGTAATAAATAAATAAATAAATAAATAAATA 61900

AGTATAAATA 62000

TCTCTAATTTAATA 62100

ATAAATGAATATAAATA 62200

TATATATATAGTTCGGGACCCGGTACGGAAACCGGACCCCGGACGAGATTTAATTAATTTATCATTATTTTTTATTTAATCTTATTTA 62300

TTATAAATA 62400

TTAATTAATAAAGAGGATTTAAAAAAGGAGTGGGAGCCCGCTCCCGTGGGTGGGGGACCGAACCCTCGAAGGAGTACTATTAAATAAATAAATA 62500

AAATA 62600

AAATTTTAAATCGGGTTTCCCCACCTACATAAATCTACGTATACCTACATATACTTATGATACTTACATATACTTACCTTACTTATATATACTTAT 62700

GATACTTACGTATACTTACATATAAAGGGATCCCTCTCTCATCCGGCTCCTACTACCCCTATTTTAAATCATTAAAGAAATTTAATAAATAA 62800

ATATAAATA 62900

TAATTTATCTTAGGAAGGATTTATTTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 63000

ATATCTAATAATATATATATATATATATTTTTTTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 63100

ATTTTTGATTTCCAAATATATAGTCCGGTTCTTACGAAACCCGGGACCTCGGGAAGTAATAAGGGAGGGGTGGGTGATAGAACCAACTATTCAR 63200

TAATAATACAGCACACATAGTAAATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 63300

CGGGGACCCCGAAGGAGTATAACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 63400

AGAGATATCCGGTCCCAATA 63500

CCCGGGGCCCCGAACTAATA 63600

AGATATAATATATAAATA 63700

TATCTCTCTCGACGGGACTCTCTTTAAAGGGGTTGGTCCCTCCATTAGTATAGAGTATAGGGAAGGCTCCCTCACTCTCGGGGTCCCGCC 63800

CCCGGGGCCCCCGGACTAATA 63900

TTTTTATTTAATTTTATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 64000

TATATACTCTTATAAAGATTCGGTTCCTCCATCTCTATGGATCCCTCACTCTCTGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 64100

TAAATATATATATTTTATTTACAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 64200

TCAAATATGTTAATATGATACCTACAGAGATTTAATATTTTTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 64300

TTTTTATTTATTTATTTATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 64400

TGAAGGAGAAATA 64500

TAATA 64600

TAATATATTTATTTATTTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 64700

GTCCGGGCTCCCGCCACGGGAGCCGGAACCCCGAAAGGAGTTATTTATATATATATATATATGAATTAATTTAATAAATAAATAAATAAATAAATA 64800

CRF b14

B5, CRF b15

138

139

140

B6

e

141

142

143

144

145

146, 147

148

ori6

149, 150

151

152 - 153

154

155



TTATTCCTTTTCCCGATGGGACTTATATATTATATTATATATATTCGCTCGATCTTCATATTTATTTTATAAAAAAATTTATATATTAT 70300 167

TATTTACAAATTAATATTAATAGTCCGGCCCGCCCGGGGGGGAACCGAAGGAGTCCGGACCCCGGGGACCCGATCCCTTTTATTTTAAAT 70400 168

AGGAAGGAGTGAUGGACCCCGGGACCGAACCCCGAAGGAGTCTTTTTCATTATTAATAATAACTATAAAITATAITTAATAATAATTTACTTGG 70500 169

TATAAATCTTAATGTTCCGGGGCCCCCGGAGCCGGGAACCCCGAAAGGAGAAATATATAAATATTACTTGTGTATAATTATATATATTATAACC 70600 170

TCCTTTTAAAAATATCTTACTTTATAAATAAAATTAATAATAANCTGATAGTAATCGAATTTTATTATTTAAITTAATTAATAATAGACAAT 70700

TATTAATTAATTTACTTATTAATAATTAATTTAGAAITTTATATATAATAATATATTTATTTATTTAAITTTTATTATTTATTTATTTATTT 70800

ATATATTTTATATACTTATTAACAATAATAAATTAAGACGTAAATATATATGTGAAGGAATCGGAAGTGGCTATTATATAAATATATATATAT 70900 (17)

AAATATTAATTAATAATTAATATTAT 71000

ACTTTATAATCTTAT 71100

ATTTAATAATAAATATTTATATTTTATATTTTATATTAAGGGGGTCCCAATTTATTTTCAATAAATAATTTATCATGGGACCCGGATCTCTCTG 71200 172, 173

TTTATCATTTATTTATTTCTTATTTTGGTTTTTATTTAATATTATAAATTTATTTTATACAATTTTATTATATGTTTATACCTTATTTATTTATATAA 71300 ori 3

ATATATATTTATATAATAATTTAATTAATTTATATATAAAAAATTAACCTAATGCTGCTCAIATAIATTCATTCCTAGTTTCCCAATCCACCACCC 71400 174

CTCCCCCTAAACCTATATATCCGAAATTAATAACCTTACCTATATTTAAATATATATAATAATATATAATAAATTTATTTATATAAATATATA 71500 (i)

ATATAIATAIATAIATAIATAIATAAATAATAATAAATAATTAATCTCCATTAGGGGTTTTGGTCCATATCGGGAACCGCAATAATAATATAATA 71600 175

TATAAATAAGATAITCTTATTTATAATAIATTAATAAATAATAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 71700

TTTATTTATTAJAAAAATAAATTTTATTTATTTAGNTTTTTTTTATAAACATTTTTATAAAAAATAAATGTAATAATAATATGATTAATTA 71800

GTAATAATAAATTTATTTATTTTATAAATTTACTTTTTCGAGGTATAGATCAGTATCAGTATCGTAAAAAACGGGTGACTAAAAATATATA 71900 176

TATATAAAAAIATAAAAAATAAATATAATAAATTTAATAAATAAATAATCAATATATTTATTTATTTATTTATAATAAATAATTAATTAATAGT 72000 177

CCGGCCCGCCCGCCGGGGCCGACCCCGAAGGAGTCCGAAACCCCTTTTATTTAATTTTTATTTAAAGAAGGAGTGGGGACCCCTCCCGTTAGGGACCC 72100 178

ARCCCCGAAGGAGTAATTAATAGATTATAATTTATTTTATTTTATATAAATAIATAAATAATAATAATAATAATAATAATAATAATAATAATA 72200

ATAATAATGTAATAATAATCATATCTCCCTTATAAATGGAATCTTATAAATAATTTAATAAATAATAATAATAATAATAATAATAATAATAATA 72300

TTAATTAATA 72400

TTTATTAATTAATAATTCGTATTAGATCTTATTTAATTTAATAACCTTTAAAAAACTTTAATAAATAATAATAATAATAATAATAATAATAATA 72500

TTATATATTTTATTTATTTTAAAAATATATATATTCCTTATAAATAATTTTAAAAAGTTTATAAATAATACTGATATATATTTATAAATAAAT 72600

TTATTTATTTTATTT 72700

AAATTAATTTGGCGCCCAAGCCGGTGTCCACCGACTGGTCCCAATATAAATGATATAATAATAATACTATATGATATAATAATAATAATAATA 72800 179

ATATGATATAACATAAATAATAATACTCTCCGGGGTCCCGCCCGGGGGGAACCGGACTATATGAAATATATTTATTTATTTATAAATAATAAATA 72900 180

TATTAATAATAATAAATTTCTTAAATAATTAATA 73000 181

CCCGTGGCCGGGCGGGAATTAATA 73100

TTTATTAATAIATCAATA 73200 182, 183

TCGTTAATCAATATATATATTTTATTTATTTTCTTATTTAATTTTATTTATATAAATTTTATTTATTTATTTATTTATTTATTTATTTATTT 73300 184

CGCGGGGGACTTTATTTATTTTAAITATAIATATATATTTATAAIAIATTAATTTGATATATATAAAATAAATACTAATGATGCTTTGATTTAT 73400 ori 4

TTAATAGTTGGTCTTATACCCACCCCTCCCGCTATTCGCTCCGAGGTCCCGGTTTCGTAAGAAACCGGACTATATATTTGTAATAAAAAATA 73500 185, 186

TACCTATATAATAATTTAATA 73600

ATAAATA 73700 187

TAATTAATAIATAAATA 73800 (d)

AAATTTCTCCCTTCTTAAAAATAAATA 73900 188

GGGGCCCTACTCTTTTGGAGACTTAAITATAAATA 74000

TTTTTAAATA 74100 189

TCACATTTGGAGCCGAGTAAAAAGGAGATAAGAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 74200 190

TAGTCCGACCGAAGGAGATGAGATTATATATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 74300

TTATGATA 74400

TAAITAAIAAIIATAIATAATAATCCGGCGAAGGAGTGAAGGACCCCTCCGACTACTATCTAATGGGAGGGGGACCGAACCCCTATTTAAGAAGGAG 74500 191

TTATATATAIATAAATAAGATTATAAATAAATA 74600

IATATATTTTATTTATTTATTTATTTATATAAGATGTAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 74700 i, 215 RI

ATAATAATAAATTAATAAATA 74800

ATTAATAATGATTTTATTTATTTATTTAATAAAGAAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 74900

CAACAGAGATATATAGGATTTGGTGAAGAAATAATAAAGGCTTAATAAGTATTTGTGAAAATAATAATAATAATAATAATAATAATAATAATA 75000

AAATACTATTAATAAGTATAGTGAAGTACCCTAAGGGAAGATAGAAAGATTTTATAAGCAATCAATGAATATAIATAIATAIATAATGATGATCC 75100

TTTTGTATAAGGGICAGCAAGTAATTAATAIATAGTAAAAATAAAGTATAAATAATAAAGAAATAATAATAATAATAATAATAATAATAATAATA 75200

ATTAATATTAATGACCCGAAAGCAACGATCTAATATGATARGATGGATAACGATCGAACCGGTTGATGTTGCAATATCATCTGATTAATTTGGGTT 75300

AGTAGTGAAGACAAATCGGTTTTGCAGATAGCGGTTTTCTAGAAATATATGTAAGTATAGCCCTTATAAATAAATAAATATATAATAATAATAATA 75400

ATATATAIATAAAGATGGTACAGCAATTAATATATATAGCGAACATATAAAGTTTTATTAATAAATAATAAATCTCGAAATAATTTAATTAATAATA 75500

AGAGTCAGATATGTCGCAATAAGTAAATACTCAAAGGGAACACGCCAGATTAAGATATAAAGTCTTAATAAATAATAAGTGAATAAATAATAATA 75600

