

Sequencing of a 13.2 kb Segment Next to the Left Telomere of Yeast Chromosome XI Revealed Five Open Reading Frames and Recent Recombination Events with the Right Arms of Chromosomes III and V

DESPINA ALEXANDRAKI†‡* AND MARIA TZERMIA†

†Foundation for Research and Technology-HELLAS, Institute of Molecular Biology and Biotechnology and ‡University of Crete, Department of Biology, P.O. Box 1527, Heraklion 711 10 Crete, Greece

Received 15 August 1993; accepted 1 December 1993

We report the entire sequence of a 13.2 kb segment next to the left telomere of chromosome XI of *Saccharomyces cerevisiae*. A 1.2 kb fragment near one end is 91% homologous to the right arm of chromosome III and 0.7 kb of that are 77% homologous to the right arm of chromosome V. Five open reading frames are included in the sequenced segment. Two of them are almost identical to the known YCR104W and YCR103C hypothetical proteins of chromosome III. A third one contains a region homologous to the Zn (2)-Cys (6) binuclear cluster pattern of fungal transcriptional activators. The fourth one, part of which is similar to the mammalian putative transporter of mevalonate, has the structure of membrane transporters. The fifth one is similar to yeast ferric reductase. The sequence has been deposited in the EMBL data library under Accession Number X75950.

KEY WORDS — Genome sequencing; Saccharomyces cerevisiae; chromosome XI; chromosomal recombination; telomeres; gene redundancy.

INTRODUCTION

In the course of the European Community (BRIDGE) project to sequence Saccaromyces cerevisiae chromosome XI, we have determined the complete sequence of 13 213 base pairs on a DNA fragment mapped next to the left telomere (about 200 nucleotides away from a 1 kb telomeric sequence, unpublished results). This fragment contains five open reading frames (ORFs), the potential function of which will be discussed below.

MATERIALS AND METHODS

Strains and vectors

Cosmids pUKG040 and pEKG100 were provided in *Escherichia coli* strain TG1 ($\Delta(lac pro)$),

*Author to whom correspondence should be addressed.

CCC 0749-503X/94/SQ0081-11 © 1994 by John Wiley & Sons Ltd thil, supE44, hsdD5, F' (traD36, pro $A^{>+} B^{>+} lacI^Q$ lacZ Δ M15)) from Agnès Thierry and Bernard Dujon (Thierry and Dujon, in preparation). They are cosmids from libraries of chromosome XI, derivatives of cosmids pOU61 cos and pWE15 respectively, containing overlapping partial Sau3AI yeast DNA fragments. Escherichia coli strain DH5 α (supE44 Δ lacU169 (φ 80lacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1), and pUC18 and pUC19 vectors were used for all subsequent subcloning and sequencing steps.

Sequencing strategy

We have used directed sequencing of ordered restriction fragments. Cosmid DNAs were digested with *Eco*RI and electrophoresed in low melting point agarose. Four *Eco*RI fragments were purified and subcloned into pUC18 or pUC19 vectors.



Figure 1. (a) *Eco*RI restriction map of the 13 213 base pair segment. The arrows indicate the beginning (*Sau*3AI sites) of the sequences included in the two overlapping cosmids. The numbers below the bar indicate the size of each *Eco*RI fragment. (b) 6-phase ORF map of the 13 213 base pairs. Small bars indicate initiation codons and full bars indicate stop codons. The location and the direction of five ORFs are indicated by arrows. The number in the name of each ORF indicates its size in amino acids. The ORF names were assigned by MIPS.

The order of the EcoRI fragments is shown on the map of Figure 1A. The sequences of the two 5' EcoRI fragments and 2 kb of the third one have been determined from cosmid pUKG040. The rest of the reported sequence derives from cosmid pEKG100. The additional sequence of 24.6 kb included in the pEKG100 insert has been reported separately (Tzermia *et al.*, 1994).

Double stranded template DNAs were prepared by alkaline lysis followed by Qiagen-tip selection (Qiagen Inc.) or PEG precipitation (Ausubel et al., 1987). They were subsequently sequenced using ³⁵SldATP and the Sequenase kit (United States Biochemical Corp.) following the supplier's protocols. Sequencing of both strands of the EcoRI fragments, subcloned in both orientations, was performed by the 'universal' or the 'reverse' M13 primers on nested ExoIII-S1 nuclease (Ausubel et al., 1987) deletions. Synthetic oligonucleotides corresponding to internal sequences (prepared on an Applied Biosystems synthesizer by the Department of Microchemistry at I.M.B.B.-Crete) were used as primers to fill in the gaps. The junctions between the sequenced EcoRI fragments have been established by PCR sequencing of PCR products, which were synthesized from oligonucleotide primers near the ends of the fragments, using cosmid DNA as template, $[^{32}P]ATP$ -labelled primers and the *fmol* DNA sequencing kit (Promega Corp.). Samples of sequenced DNAs were analysed on 40 cm long 6% or 4% polyacrylamide gels with single or double loadings.

Sequence analysis software

Endonuclease restriction, 6-phase ORF mapping and hydrophobicity profiles of the sequences were accomplished by the DNA Strider software (Marck, 1988). Comparisons of nucleotide and amino acid sequences were made to the GenBank, EMBL, SWISS-Prot and NBRF libraries using the GCG package software by us on the I.M.B.B. MicroVAX and by the staff at MIPS.

RESULTS AND DISCUSSION

Sequence determination

The reported sequence was determined from overlapping Exo III produced deletions and

IENT OF CHROMOSOME XI

Sau3A I → pUKG040 1 GATCTATAAC GAAATGTCAA ATAATTTTAC GGTAATATAA CTTATCAGOG GCGTATACTG AAACGGACGT TACGATATTG TCTCACTTCA TCTTACCACC 101 CTCTATCTTA TTGTTGATAG AACACTAACC CTTCAGCTT ATTTCTGGTT ACGGTTACAC AAAAACTATC GCAACCCAGA AATTTTCATA TTTTCAGTGT 201 CAAAAAATGA GGGTCTCTAA ATGAGAGTTT GATACCATGA CTCCGTAACT CGCACTGCCC TGATCTGCAA TCTTGTTCTT AGAACTGACG CATATTCTAA 301 ACGCCCCGAC GCGACGCCCC AAAAAATGAA AAAAGAAGCA GCCACTCATT TTTATGGAAG GACAAAGTGC TGCGAAGTCA TACGCTTCCA ATTTCATGG 401 TGTTTATTGG ACATACTCTG TTAGCTTTAT TACCGTCCAC GCITTTTCTA CAATAGTGTA AAATTTCTTT CTATGTTTAT CATATTCATA AAATGCTTCA 501 TITGITCGTT AATTTTCAAT GTCTATGGAA ACCCGTTCGT AAAATTGGOG TITGITCTCTA ATTTCCGATA GTGTAGATAC OGTCCTTGGA TAGAGCACTG 601 23 GAGATAGETE GTETERATET GTAGAGTAE CATGGGACAE CAGTGATAAE TETEGEGAET GETEAGETE GAATAGEAGT CAACATGETE GTEAAGTCA 70 80 CATAGETGAA AACAGCTTCA GCAATTTCAA CEGGGTAGGT CEGACTEGCA TCACETGCTE GAAACAAGTA GTATTGGCCC AAATGACCTC TGATATCGG Ξ 901 GACGTAAACA CCCAATTCGA CCAAGTTAAC TCTTTCGTCA GATGGAGATA GACTGGTAGT GGCTCGGGCA GCGGCAACAC CAGCAGCGAT GCCGGCGACA CCAGCAGCGA TTGAAGTTAA TTTGACCAT ATATTTGTTT TGTTTCTTAG TGCTGATATA AGCTTAACAG GAAAGAAAGG AATAAAAACA TATTCTCAAA 1001 1101 GACATACAGT TGAAGCAGCT CTATTTATAC CONTECCTTC ATCAGTCATC ACTACTTAAA CGATTCGTTA ACAAGTCCTC ATTTAGCACC TCACATATCC 1201 TCTATATCTC ATCTTTCACA CAATCTCATT ATCACTATG AGATGCTCTT GTTTCTGAAC GAATCATACA TCTTTCATAG ACTTCGTATG TGGAGTACTG 2 1301 TTTTATGGCA CTCATGTGTA TTCGTATGCG TAGAATGTGA GAATGCCAAT TATAGGGTGC CGAGGTGCCT TATAAAACCC TTTTCTGTGC CTGTGACAT < 1401 TCCTTFTTCG GTCANANAGA MTATCCGANT TITAGATTIG GACCTCGIA CAGAAGCTIA TIGTCIAAGC CCATATICAG TCTGCTCTAN ACGCCTTCC EcoR I ECOR I EAAGAAAT ATTTCTATCT CIT<u>GAATTC</u> TACAACATTA AAGGTGTGTT GGGAGTCGTA TACTGTTAGC GTCTGGAAAT TCGT<u>GAATTC</u> TCCGCGAATG 1501 1601 CCTTGGTGCA ATTACGTAAT TGTAGCCGCT GAAGGCGGAT GGTATTGAGA GAAGTTGATG TTCAAAGAAT CGTGTATGAT AAAAATAAAA TTGTAACAGT 1701 TRATANATAC TANACITTAC ITGANANGIC ITTINGICCG ANAATGANIG IGGGAATACT ANTIACAAGC CICCCAIGGC ACTICITCAT ITCITITIGA 1801 CTACCAACAT GGCACTITTT CAATTAATTA TATTTAATAA ATATTTCTIT TTACTAAAAT TTTACAAAAC CCAAGGTCAT TATGGTCATT AAACCACAGT ATATATATAT ATATGTTGAA GAGGTTTCAA TICGGCCAAA ACGACAGTAC TATATCTGAG TCTGGACTAG ACAATATTGA ATTTACTTAC CAATTAGGAT 1901 2001 TAATATACCA GAAAAAATATC GAATTTCTTT TTCCGGAATA AAATACTTCT CTTTCAAAAT ATTCATTGC AAATTAGAAT ACATAGGATA TTATGATGTG 2101 CACCTARARAT ATAGAGARAC MAGTTAGAR CITACATATC ARACGTATC TRACITATE ARARAGARARA ATAGCTGGGC AGTGACACCT GARATACCCC 2201 ANANANAAGA GCAACTATCT AGAAAAGTAT GTAGAAGTCC GAAAGGAACA GAGGTGCCA CCAATAGCTT TCACAAAGAG TTCCTTTCGT GTTGGCAAAA 2301 ANAATGAGTC ACTTACTACA OGTTCAATTT TTATAATTCA ATTGTTGTAC CTTTCTGGAA AAGAAGTCCT CAAAATTACT CCACCCCGAA ATACCTGTAT 2401 CAAAGAAGGC ACTTGGAGAA CCTAATATAT CATCGATGAA ATTTTCCGAG AAATCTTCTA AAGGTECCTC TAAATCATTT AGTATAGTTG TTAAACTGTT 250 TTCAGTTGCA TCATTTTCGG ITTCAGGCCC ITCATCGATA ITTITCATIC CITIGITAAA AGCTCTTCIT CCTATITITT CCATTCATAT AATAATACAA 2601 ANGTANTAGE ACTITICCCA ANTITICACTA ACAGGITTICA ACTITICITIC GIGANTATGA TCANANTATET TITICCATTIC TITANATECA ATTITICACTO 2701 ANATACACTT ATTGTGTGGA CCCCAAAGT CTCCAGGACA TATAGAATAA GTTAGTGCGG ACGAATTTCT CTGAATGAAT AATTGGATAT CTTTATTTTC 2801 CATATACTGT AGAAAAATGG TATAAATTTG AATTAATGCT CTGAAAGCGT TCAGATAAAT AAGAAACAAC GCATTGTTTA AATGGGGGCCA TTTTTTGGA 290 АСССАТААТС АСТТАСААСА ТААТТТССТС ТІСТСТААТТ САЛААТАССТ СТСЛАССАСА ТТААЛССАСА ААТСТААТСА АТТТАСТССА СССАССАССА 3001 TANAATTGAA TATTCTCGCA TCAAAATCTT GATGTGTTAA GGTATAAAGA TTGGATAAAC TTCCTATCAT ATGGAGAGTT GCATGCCATA ATTGAAGTTC 3101 GETGAACTCA TTACCATTTA AATTAGATEC ATTTAAGTAA AAATCCAACE GETTAAACAT TTTAATGETE AACTTTTTTA AGTCCTCEAT GATGAGTGEA 3201 ATGTCAGGCG GTTTCTCACG AGCATGAATT TGTTTCATAA TATTTCTCAG TCTTAATGTT GTTTTATACA GAAGGATGTC CCCTGATGAA GAATAGTTT Ecor I стаслестае стерттаса матесност голтесское сс<u>ранте</u>ст старабаетс алагиттале стесствалт ассаелате ассаелатет 3301 F705 3401 CTCCAAGTAA GGTATTTCCT CTATTGGATG ATTTTTACTC AGATATAAAC GACGAAGATC TTCATTAAGG CCCATATGGA TAGCTGTGTC AATTGTCAG CCATCAATAA ATATGCAATG TGACTGATCG COCCEGTCAA GAOCGGCCAC ATTTATGTAC AGGTAACGTA AAAATAAAAA TTGCACACGC TCAGTATAAA 3501 360) AAACTITAGC CGAAACAAAA GACGITAAAA CIIIAIGGAA GACITCAATA GCCICIGGCA CIICIIIIGG AIGGGAIGCI AAGCACATAA IGGCIGIGAI 3701 AACACCCACT TTATAATAAT ITTTTTTACT GICCAGAITA AGGGAAATTA TTGIGTGCIG ACCAGTTTTA TGACTICITG GGCCTTTAAC AAAACAATC ТЕСАБАТССС ТТААТАСТТТ ТЕСТТТАТБА АСААТТТЕСТ ААСТАТЕСТА АЛААТСЕСТА БСАЛМАЛАТ САСТТАЛАТА СТЕЛЕДСАЛА АСТТЕЛТААТ 3801 3901 TTGGAAGCGA TTCACATAAG TATTCAATAA CACTTGGTGA GCCTGAGTGC GCTGGGGGAG TCTCAATGGA ACTAATCTCC GAGAGTGTTG AGTAATGATG 4001 CTCTCTTTTC CAGTTGTTAC GECTGAGTTT TAGGACTTEE CATATETECT CTCTGTATTT GGCAAAAGTA TETGTTTGAG TEGCTAATAT TECTEGATAC 4101 GACGTTEGTE CATAAMAAT ETECCTATES TECTTIGAAG ATAAATATET CATETIGGAA AGAGGETTIT ETECTTEETE AATETIAEGE TEGEATTETI 420 TTGTTAGETS AGATTCAAGE TETTCAATAE GAEGAGECAT ETCTAATTTE GAACATETEE TGTATTTTTE CETGAGITGG TEETECATGG COGAAGIGET TTETTTATAT TEACATTGET TEGATITEG AGAAGAGGAA CTICEACAGA ACGGEETTAE TEGATEGEAE TEGAGETTE TEACTETACA AAAGTGGGAC 430) 4401 GACTITICCAG CCTTTCTTCT ACTCTGCCTA TCTAATTTTT TTTTTTGACT AAGTTCCGTA TTCTTCATTT TCTATTGATT ATGGAAAGTA TATTGCTACA

Figure 2.

401 ACCURTICAL ALTITUDE AUTOMACTI GALARCE CUNCTU CALCULAT CALCULAT CALCULAT CONCURS AND ALTACE CONCURS AND ALTACE AND ALT	4501	GATTCATTAT	TTCCAAAGCA	ACCGTTATTG	GTTACTAGTG	CTOGACAATT	GATATGTAAT	CAGTIGCTAT	ACATCINGAT	CAAGTTOCTC	TCTTCTTTTG		
4791 ORSCHATC ATGEMARCH REMATERT TRACEGOR TERMAGEN GARANTE CALUBRATE CALUBRATE ADMANDE CALUBRATE TEMPORE ADMANDE CALUBRATE ADMANDE AD	4601	ACGTTTCGAA	TATTTTTGA	CTTCAAACTT	GAAGTGCACA	CTCATCTTTT	CAAGGTTTTT	TATGATTCTC	CGTTTTATGG	CTATTCCCGC	тттаасата		
4000 STITUETAT TAITUCATE GEOGRAPHIC TERCACAT TRACEGATA MAGACAR TAIMANE CANCELLA TEALTICA TRANSAC 401 ACTABLAS TAIALCEA ATALGUAT TECCACA TAICCETT CONCELLA CONCELLA TEALGRAP ATALGUAT TAIALESA TELEVANA 402 ACTABLAS TAIALCEA ATALGUAT TECCACA TELEVANA CONCELLA CONCELLA ATALGUAT ATALGUAT TELEVANA 403 ACTABLAS TAIALESA ATALGUAT TELEVANA TELEVANA TAIACETT CONCELLA ATALGUAT ATALGUAT ATALGUAT 404 ACTABLAS TELEVANA GOLFATA AGOLFATA TELEVANA TELEVANA 405 ACTABLAS TELEVANA GOLFATA AGOLFATA TELEVANA TELEVANA ATALGUAT ATALGUAT ATALGUAT ATALGUAT 406 ACTABLAS TELEVANA GOLFATA AGOLFATA TELEVANA TELEVANA 406 ACTABLAS TELEVANA GOLFATA AGOLFATA TELEVANA TELEVANA 406 ACTABLAS TELEVANA GOLFATA AGOLFATA TELEVANA TELEVANA 406 ACTABLAS TELEVANA GOLFATA AGOLFATA TELEVANA ATALGUAT ATALGUAT ATALGUAT CANCELA ATALGUAT 406 ACTABLAS TELEVANA AGOLFATA TELEVANA ATALGUAT ACTAGLAS ATALGUATE ATALGUAT ATALGUAT ATALGUAT 406 ACTABLAS TELEVANA ATALGUAT AGOLFATA TELEVANA 406 ACTABLAS TELEVANA ATALGUAT AGOLFATA TELEVANA ATALGUAT ATALGUAT ATALGUAT ATALGUAT ATALGUAT ATALGUAT 406 ACTABLAS TELEVANA ATALGUAT AGOLFATA TELEVANA ATALGUAT ATALGUAT ATALGUAT ATALGUAT ATALGUAT ATALGUAT ATALGUAT 406 ATALGUAT AGOLFATICA TELEVANA AGOLFATA TELEVANA AGOLFATA TELEVANA ATALGUAT ATALGUAT ATALGUAT AGOLFATA 407 ATALGUAT AGOLFATICA TELEVANA AGOLFATI TELEVANA AGOLFATI AGOLFATI TELEVANA ATALGUAT ATALGUAT AGOLFATI AGOLFATI AGOLFATI 407 ATALGUAT AGOLFATICA TELEVANA AGOLFATI TELEVANA AGOLFATI TELEVANA ATALGUAT AGOLFATI AGOLFATI AGOLFATI 407 ATALGUAT AGOLFATICA TELEVANA AGOLFATI TELEVANA AGOLFATI TELEVANA ATALGUAT AGOLFATI AGOLFATI AGOLFATI 407 ATALAGUAT AGOLFATICA TELEVANA AGOLFATI TELEVANA AGOLFATI TELEVANA ATALGUAT AGOLFATI 407 ATALAGUAT AGOLFATICA TELEVANA AGOLFATI TELEVANA AGOLFATI 407 ATALAGUAT AGOLFATICA TELEVANA AGOLFATI TELEVANA AGOLFATI 407 ATALAGUAT AGOLFATICA TELEVANA AGOLFATI TELEVANA AGOLFATI 407 ATALAGUAT AGOLFATICA TELEVANA AGOLFATI 407 ATALAGUAT AGOLFATICA TELEVANA AGOLFATI 407 ATALAGUAT AGOLFATAL TELEVANA AGOLFATI 407 ATALAGUAT AGOLFATAL 4	4701	GAGGCGATGC	ATGGAAGACA	ATAATTATTT	TAAGGTGCTT	TTGAGATCAA	GATATTTTTA	AGAGTACTCA	CTATGOGTAT	GAGATAATAA	AATATATTCC		
991 ACTABLEME TENTACTION REALEMENT TOTOCOCCE TENTOCIC CONCERNESS ATMAGENESS ATMAGENESS ATTACHES ATTACHES ATTACHES A DECEMBER ATTACHES ATTACHES A DECEMBER A DECEMERA A DECEME	4801	GTTTTCTTAT	TATTTCCATT	CTTCAGATTT	TTACAGGCTA	AACACACATA	TATAAAAATG	CCATOGGATA	TTCAATTCAG	татсстоста	TIGATATAGA		
5901 MITCHAITT AMNORAM GAGARATA ASSGURANG CTOMUTCH ALCONFICE CEMERADE ANALOGIES TEMPTORE TEMPTORE TEMPTORE TEMPERATURE ANALOGIES TANTAGE TEMPERATURE CEMERATE ALCONFIGURATION ANALOGIES TANTAGE TEMPERATURE ALCONFIGURATION ANALOGIES TANTAGE TEMPERATURE ALCONFIGURATION ANALOGIES TANTAGE ALCONFIGURATION ANALOGIES ANALTAGE ALCONFIGURA ANALOGIES ANALTAGE ANALT	4901	ACTATATAAC	TGATACTAGA	ATATACTAAT	TOGTGCACTA	TTAACCGTTT	GGCGATGTCC	GAAGAACGGC	ATGAAGATCA	TCATAGOGAT	GTTGAAAATA		
1010 TETERACHAE THIETEETT GEOGGEGAM TETEGETARI GENITIAT TAGGGATTA TITARGAAT ANTACTITE CREATEGIN TAMATA ANTACTING 10200000000000000000000000000000000000	5001	AATTGAATTT	AAATGGTAAA	GACGATATTA	ACGGGAATAC	CTOAATCTOS	ATCGAGGTGC	CTGATGGAGG	ATATGOGTCC	TTTATTCTTC	TIGCTITIAT		
1200 TANGATCAL TAGGTOGGT ACARTONET TOTOGACTIT TITTOTOCOL AGTINTANCA TOGCTIANE ATALANTIC ANTENATE ANTENATE ATALANCE TATALOGAE TAGGTONET TOTOGACITY TAGGTONE ACARTONE ATTACANT ATALANCE ATALANC	5101	TCTGTACAAC	TTTTCTACTT	GGGGGGGCAAA	TICTGGTTAT	GCTATTTATT	TAGOSCATTA	TTTAGAGAAT	AATACTTTTG	CTGGTGGGAG	TAAATTAGAC		
1930. TAGGERING STITUMOUS GANGUELE TERTIONEC TITUTCHE ACATOLOG AMAITACT CLOCOMAGE GUITTANTE GANGUELE TERTIONAL GANGUELE TERTIONAL ACCOUNT AND ALLENDAR ANALYSIS CONCENTS ACCOUNTS TRACKAGE ACCOUNTS ACCOUNT AND ALLENDAR ANALYSIS CONCENTS ACCOUNTS ACCOUNT ACCOUNTS ACCOUNTS	5201	TATGCTTCTA	TAGGTGGGTT	AGCATTCAGT	TGTGGACTTT	TTTTTGCCCC	AGTTATAACA	TGGCTTTATC	ATATATTTC	AATTCAATTC	ATTATAGOCT		
140 AGENTTATT TUNINGUCA GUIDENCE GRUDENCE GRUDENCE ARGENGED TANGEGGGE TUNINGUE GUIDAGE GUIDAGE GUIDAGE GUIDAGE 150 SUBAT DE DESCRIPTION AGENTAL AGENTAL AND	5301	TAGGGATACT	GTTTCAAGGG	CACCGCTAC	TOCTTGCAGC	TTTTTCTGTC	ACACTCTGGG	AAATTTATCT	CACGCAAGGC	GITTTAATTG	GATTCGGTTT		
Solo Generative transformer transformer atterness and a second transformer atterness atteress at	5401	AGCATTTATT	TTCATACCCA	GIGTCACACT	CATCCCACTA	TGGTTCAGAA	ATAAAAGATC	TTTACCTCT	GGTATAGGAA	CTGCTGGAAG	CGGGTTAGGT		-
See 1 TOCOLITAL GITCADERA ADMENDIA MODICING TOMOLOGY ADMILITICA AND THEAT CHARTACIA CALOURS ADMILITICA AND THEAT CHARTALLY CALOURS ADMILITICA AND THEAT COUNTED CHARTAL CALOURS ADMILITICA AND THEAT COUNTED CHARTALLY CALOURS ADMILITICA THEAT CALOURS ADMILICAT TRANSCOLUCIUS ADMILICAT CALOURS ADMILICAT CALOURS ADMILICAT CALOURS ADMILICAT CALOURS ADMILICAT CALOURS ADMILICATI TACCOLUCUU ADMILICAT CALOURS ADMILICAT CALOURS ADMILICATI TACCOLUCUU ADMILICAT CALOURS ADMILICATI CALOURS ADMILICATI TACCOLUCUU ADMILICAT CALOURS ADMILICATI CALOURS ADMILICATI CALOURS ADMILICATICA ADMILICAT CALOURS ADMILICATICA CALOURS ADMILICATI TACCOLUCUU ADMILICATI CALOURS ADMILICATI CALOURS ADMILICATI CALOURS ADMILICATI CALOURS ADMILICATI TACCOLUCUU ADMILICATI ADMILICAT CALOURS ADMILICATI CALOURS ADMILICATI ADMILICATI CALOURS ADMILICATI TACCOLUCUU ADMILICATI TA	5501	GGTATTGTCT	TTAACTTGGG	AATGCAAAGT	АТТСТАСААА	AGAGGGGGCGT	TAAATGGGCG	CTCATTGCTC	AGTGCATAAT	ATGCACATCA	CTTAGCACCA		47.0
5101 GETCHEITA CITIETEGAN INGENENT TECHNENT TECHNENG GENERGEN CETTEREN INCENTIOL AND THE CENTRE AND	5601	TTGCGCTTAT	GTTGACCAGA	ACAACACATC	AAGGCCTACG	TCAACATAAG	Sau3A I→ I A <u>GATC</u> TTACA	DEKG100	GCTAGATTAT	GATGTGCITT	CAAATTTOGC		à
 MGEMAGGET CATACCTATE GEGENERTS ARTICLOSET CICLECTEGE ACCACCANT STUDENES A MATCORNER CHARACTER MGEMAGGET CATACCTATE GEGENERTS ARTICLOSET CICLECTEGE ACCACCANT STUDENES A MATCORNER CHARACTER CATALATT GENERTTE ACCACTTEC ARESECANCE TITETEGENE CATEGORAL CICLEGAM GEAMEENT ATTACCONE GEAMENTS ATTACCONE GARACTER CATACITAT GEAMENTER ACCACTATE CICLEMANT A CONTEGNA CANACTAR CICLEMAN ANTICOME CANACTAR CONTEGNA CANATTACE CONTENTS COTACTING CITLAGITTE TUCCANTOGE TRACAMAGE CITLATARIA CICLEMANT ACCONATE GEAMEENT ATTACOTING ACCANTER ACCANTER CONTENTS CATACITATE ALCOTANCE CICLEMANT GEAMANT CITLAGAGE TITLAGATI ACCONAGE CIATURACIA CANATTACE TATACAGA ANTEGNATA ALCOTANCE TUTLAGATA GETTATATA CONTACAT ACCONAGE CIATURACIA CANATTACE ACCANATEA ANTEGNATA ALCOTANCE TUTLAGATA CANATTACTITITAGE TATACAGA ATTACTACE ACANATARIA CANATTACA ANTERNA CANATTAC TUCAATARA CANATTA THITTEGUTI ALACTACAA ACATACAACA ANTERNA ANTERICA CANATTACAATA ANTERICA CANATACAATA ANTERNA ALCOTANTA TATACTARA ANGETANT TATACTAR ATTACTACAA ANTERNA CANATTACA ANTERNA ANALTACA TATACAGAT ANTERNA ANALANA ANALTATA ANGETANA CITLAGORIA TATACTAR ATTACTACA ANALTACA TATACAGAT ANTERNA ANALTARA ANGETATI TATACTARA ANGET TATACTAR ATTACTACAA ANALTATA TATACTACAA ANALTATA ANALTACAATA ANALTACAA TATACCAATA ANGETANT TATACTARA ANGETATACTACAATACTAA ATTACTACAA ANALTATA TATACTACAATA ANALTATA ANACCAATA TATACTAGA ANGETANT TATACTARA ANGETATACTACAATACTAA ATTACATA ANALTANA ANALANAA ANALANAA TATACTAGA ANGETANT ACACCATAT ATTACTACAA ANGETATATI TATACTACAATA ATTACACTA ANGETATA TATACTACAATA CANATAGAATA ANGETANT TATACTANA ANGETATA ATTACACTA ATTACACTA ATTACACTA ATTACACTA ANALANAA ANALANAA ANALANAA CANATTACAA CANATTACA ANGETANA ATTACAATA ANGENATATI TATACTACAATA ANALANAA ANALANAA CANATTACAA TAGGETANT ACCATATA TATACTACAA ANGETATA TATACTACAATA ANALACAATA ANALANAA ANACAATACAATACTACAATACAATA	5701	GGTCTGGTTA	CTTTTTGGAT	TIGTATCATT	TOCTATGTTA	GGATATGTTG	TCCTTTTGTA	TTCCTTGTCT	GATTTIACCG	TTAGTTTAGG	TIATACTAGT		
5001 5010 5020	5801	AAGCAAGGCT	CATACGTATC	GTGCATGGTG	AGTGTCGGCT	CTCTGCTGGG	ACGACCAATT	GTGGGTCACA	TTGCTGATAA	ATATGGATCA	CTAACAGTIG		
6001 TATTATEGEA ACANTIGGE CARANTIGE TRENATION ACKEGANE TEGETICA AMAGETECT GERACUTTE GEAGANE GANTTTREE 60000 600	5901	GCATGATATT	GCACCTTGTC	ATGGCCATCC	TTTGTTGGGC	CATGTGGATA	CCTIGTAAAA	ATTTOGCCAC	TGCGATACGT	TTTGGATTAT	TGGTTGGTTC		
6101 6000 CERTIFIC CETERATEC EXCENTIME GETERGANC TEGETEGANC TEGETEGANC TEGETEGAN GAUNCETT ATAMIGENC ACOMMENTS CECONTENT C CETACTIFIC TETTA TEGENTIA TEGENATEGE TATEGANGE GITTATATA GETEGANA ACTIGETE GEORGANGE ONTERATEG ACOMMENTA ATTEGATTA ALCETTANT TALEGANAGE CATTAGASGE GITTATATA GETEGANA ACTIGETE GEORGANGE ONTERATE ALCONTANTA ATTEGATTA ALCETTANT TALEGANAGE CATTAGASGE GITTATATA GETEGANA ACTIGETET TAGANGE CATTATATA TEGENTIA ATTEGATTA ALCETTANT TALEGANAGE CATTAGASGE GITTATATA GETEGANA ACTIGETET TAGANGE CATTATATA TEGENTIA ATTEGATTA ALCETTANTI TECANTEGA ACCTIANT GETATATA GETETTA ALMACANTA ACCHEGAN OF TAMAGET CATCATATA ATTEGATTA ALCETTANTI TECANTEGA ACCTIANT GETITITA ALMACANTA TACENCECT CIACCOGITA ALMOSTEGA CANATANAG ATAGANAGE GAUNAGANAGE TEGANAGITI TITAGENET ATAMATATI TACENTITA ALCENTERIE GAUNAGATI TALAGONAGET ATTEGATTA ALCETTATE TEGANTEGAT ATAGATA ATTEGATA ALCETTANT GETATEGAN ATAGANAGE TEGANTITAT TEGEATIGAN ALTITATA ALCENTAR CETALECTIA ALTITATA ALCENTAR TEGACINA ATTEGATI ALCEGATIANT GETATEGATA ALCENTAR ALCENTAR TEGACATA ALTITATA ALCENTAR TEGACATA ATTEGATAT ALCEGATIANT GETATEGAT ALCENTAR ALCENTAR TALENCERA ALTITATA ALCENTAR ACCONTENCE TEGECATI ALCENTARI ALCENTAR ALCENTAR TALENCERA ALTITATA ALCENTAR TALENCE ALCENTRA ALCENTARIA GENERATI ALCENTAR ALCENTAR ALCENTAR TALENCERA ALTITATA ACCONTENCE TEGECATI ALGENTITI GENATICIA ALTITATA ALCENTAR TEGACATA ALTITAGANG ALTITATAGE ACCONTENCE TEGECATAT ALCENTRA GENERATI ALGENTARIA ALCENTAR TALENCERA ALTITALAGUNA ALCANTEGA ALTICATA ALCENTARI GAUNAGANA ALTITAL GAUCANAGA ALTITATA ALAGENTA ALCENTAR ALCENTA ALAGENTIT ALCENTRA CALAGENTAR ALCENTRA ALAGENTAR ALCENTAR ALCENTAR TALENCERA ALTITAGAN ALCANTER CONCERNAR ALCENTRA ALCENTRA ALAGENTAR ALTITATE ALAGENTAR ALCENTRA ALAGENTAR ALCENTRA ALTIGATECH CALCEGARA ALCEGARAT ALGAGARA ALTITATA ALCENTAR ALCENTA ALTITAGANA ALCANTERCE ALCENTARIA ALCEGARATA ALGAGARAT ALTIGATARIA ALTITATA ALCONTENT ALCONTENCE ALTICATA ALCEGARAT ALGAGARAT ALTIGATARIA ALTITAGANA ALCANTERCE ALCONTENT ALCONTENCE ALTITICATI ALCONTENCE ALCONTENCE ALTI	6001	TATTATGGGA	ACAATTTGGC	CAACAATTGC	TTCAATTGTT	ACACGCATTG	TTGGTCTTCA	AAAGCTTCCT	GGTACCTTTG	GTAGTACCTG	GATTTTTATG		
4201 COTACTINGE INITIAGINA RECOMMENSE INITIAGANGE GITATAMA GEORGACIA GANTIGETE CONTINUE CONTINUE ACCAMMENT A RECOMMENSE ANALYSIS ACCOUNTER A RECOMMENSE ANALYSIS ANA	6101	GCGGCTTTTG	CCTTAGTTGC	CCCATAATC	GGTCTGGAAC	TTOGTTCAAC	TGADACGAAT	GGAAACGATT	ATTATOGTAC	AGCAATATTC	GTGGGTTTTG		
4301 ATECANTER ALCOTERACT TRUCKERER GETERARDE CHITTEGET ATEACHTE ACCESSION OF TRANSCE GETERARGE GETERARGET GENERARDET TRUCKERER ATEACTER ATTACTER A	6201	CGTACTTTGG	tgttagttta	TGCCAATGGC	TATTGAGAGG	GTTTATAATA	GCTOGAGATG	AGATIGCIGT	GCGTGAAGCC	TATTCAGCTG	ассалатса	¥	
6400 GATAAGAGE GGALAGETT GECAGAMAN TITTTECTT A KONCTITT TACKTTERE TITETGITT ATAMIDATE ATAATTAKE ACMATAADAG 6501 ATAATAMAA CATTETITT TEGAMATEA ANGETANCE GETTITTA ATAGMAATAT TAGEGACET CINCEGAA MATETINE ACMATAADAG 6501 ATAATAMAA CATTETITT TEGAMATEA ANGETANCE GETTITTA ATAGMAATAT TAGEGACET CINCEGAA MATETINE ACMATATADAG 6501 ATAATAMAA CATTETITTT TEGAMATEA ANGETANCE GETTITTA ATAGMAATAT TAGEGACET CINCEGAA TAGATATTE TAGEGAAA TITETITTA ATAGAAAA 6501 TAGEGEART TAGEGETTIG TEGATETIAT ANTAAACE TITALATETT ACTICACAAA ANTEMAACA TITETEAAC TITETEAAC TITETEAAC ACCACAAAAA 6501 TAGEGEART TAGEGETTIG TEGATETIAT ANTAAACE TITALATETT ACTICACAA CACAAATTIT CECTACETAA TETTEGETEGE GEGATITTE AGAAAAGGA 6501 CIAGETEAG ATAGEATE ATAGATEAT TALETICAA CECTAAC CITALACETT ACAACGAGE TATEGETEGE GEGATITTE AGAAAAGGA 6501 CIAGETEAG AETIGETATI TAGEGETTIT TEGAATETA ANTEATEA CACAAAGATATA TATEGACEA TATATEGAT TAGEGEAA GANETCACET 7100 GITETEAG AETIGETATI GEGAATETA TITETAGETA CACAAATTAT TEGAAGACE AAAAATATE TATEGAGETA AMAAAAAAA 6501 TEGTAMAGAC GACAACETEG GEGEGEGAA AATATATEA TITETAGETA CAMAAGAATATE TAGEGEAGE TATEGETTA TATAGGETA AGACCAATTA 7100 GITETEAG AETIGETATI GEGACETET TEGATATETA TEGAAGAAA TAGAAAAAAATTE TATEGAAGATA TATAGAGETA AAAAAAAAAA	6301	ATTGCATTTA	AACGTTAAGT	TATCACATAT	GAGTAAATGT	CTITTTCGTT	ATAAACAATT	ACCTAGGAGA	GTCTAAAGGT	CACTTTIATT	TCACACTGTA		
4501 AFAATAANA CATITITITI TGCAAATGA ARCTEARCE GOTTITITA AAGAAATAT TAGGACCT CIACGACTA MATCTICA CIAAAGCAC CIAAAGCA AAAGCAAA CAAGTACCA CIAAAGCAA AAGAACAA AAGAACAA AAAGAACA AAGAACAAA AAAAAAAGA CITTITITA AAGAAAACA TATATAAAAAAAAAAAAAAAAAAAAAA	6401	GATAAGAAGG	GGATAGAGTT	GCCAGAAAAT	TTTTTGCTTT	ATCAGCTTTT	TAGATTTGTC	TTCTOGTTTT	ATAAATAATG	ATAATTTAGT	ACAATAATAG		
6601 AGTAGAMAG TAAGATHAGG TIGAAACTIT TITAGGOGA TIKAHTITI AGTAGTATAG ATGTITGATG GGATAGTITI TAAGGAGAT TINTGAAG 7701 ATGTICCAGA TAGGTITAG TEGTATIGAT AIRTACTAT ANIBAAAGG TICHACGCAA AACATATIG GGATAGTATI TAAGGAAGA TITTITAAG AGTAGAAAAA 7601 TATGGGATAT TACGCGTITE GEATATIGAT AIRTACTAT ANITAAGAG CITAACGCA AACATATIG GGATAGTAT TITGGATG GGAGATITIG AGAAAAGGAA 7601 CAGGTITGAG ATATGANGGA TGGAATCTAT TATCTTCGAA CCGACATTIT AGAAGGAGG TIATGGAGG TATTGAGG MATTATAG AGAAAAGCA 7601 ACGTITGAGA ATATGANGGA TGGAATCTAT TATCTTGGA CCGACATTIT AGAAGGAAGGG TIATGGAGG TATTGAGG MATTATAGG MATTATAG 7601 ACGTITGAGA ATATGANGGA GGAGAATTI AGAGCATTIG AGAAACTATIG TIGGAGAACT AATAATGAAT AAACCAGGT AMAAAAAAA 7601 GATGTAGGA CAGAAAGTTG GGCTGGGAA AMAATATA GAACATATIG TIGGATAACT TAGGCGAGAT ATATGGAGA AMAAAAAAA 7601 GGTGAGGTAAT CACCTITTTA TICGGGAGAA AMAATATAG GAACATATIG TIGGATAAT TAGGGAGGT TATGGAGAT ATATGGAGGA CICGGTGGT 7610 GATATTACC CACAGTAGAA AMACAGGAAT ACGAGAAMA TIGAAAGAA ATATGGAGAG TATGGAGAT TATAGGAGAG CICGGTGGT 7610 GATATTACC CACAGTAGAA AMACAGGAAT ACGAGAAMA TIGAAAGAA ATTATGAAGGA ANGGAGA ATTAAGGTTI CATAAAATT 7610 GCCGAGGGGA CATTAGCTAT CACAGGAATA ACGAGAAMA TIGAAAAGACA TIGAGAAATG GAACAGACA ATTAACGATA ATTAAGGAAGT 7620 TICTTAAATA CACGGACATA KACAGGAAT ACGAGAAMA TIGAAAAGACA TIGAGAAATG CAAGAGAAGGA ATTAAAGGAT ATTAACGAAGA 7630 TICTTAAATA CACGGACATA KACAGGAATA ACGAGGAMA TIGAAAACAC TICGACAACA TAGGAAGCT TICTCAACGA ATTAACGTTIC ATAAAGATT 7640 TICTTAATA AAAAAGAC EGCTGAGCA CITCTTAAGA AMAAAACA TOCGCACAC TACGACATAC TAAAAATTCCA TIAAAAATATTAGA 7751 TICTTATAG GAAAAGACC EGCGTACTA CITCTTAAGA AMAAAAGA TAGCGACAC TACGACATTC AAAGTATTCCA TIGGCACAC ACACTAAAAT 7761 TICGGTTTAA TACAAAGGAC EGCTGAGCAT CITCTTAAGT CITCAAGGA ACCACAGATA ATGACGCAC TACGCACTA TACACTATA AGGCACCT TICTAAGGA GTGGCGCACAG CITCATAGA 7770 TICTTTATAG GAAAAGACC EGCGAGACA CITCTTAAGA CITCAAGGA TIGGGAGGCT TICCTAAGGA GTGGCACAC CAACTAGATT 7770 TICTTTATAG CAAAGTAG TITTAAGTA AAAAAGAAC ACAATAGAGGA TICCTAAATA TAGGCACACT TACGCACAC CAACTAGATT 7770 TICTTTAATT GAAAGAACAAC ACAATAGAT CITCTAAGA CAACACTACATA ATGACTACA TACTCTAA A	6501	атаатааааа	CATTTTTTTT	TGCAAATTGA	AAGCTTAACT	GGTTTTTTAA	AAGAAAATAT	TAGTGACCTT	CTACCGACTA	AAATCTTCCA	CTARATCOCC		
4701 AESTECACA EXAGGENETATE EXCENTEAE AESTACENE ANTACAME ANTACAME ANTACAME TETECOCAGA GANAGANA ESTECCE CECCATERE EXCENTE ACCACAMANA 4601 EXECCECTE EXECUTERE EXACCESSE EXCENTE ACCCESSE EXACTAR EXECUTERE EXECUTAR EXECUTERE ACCACAMANA 4601 EXECUTERACE ATALENESSE EXCANTERE EXACTEME EXECUTERE EXECUTAR EXECUTERE EXECUTAR EXECUTERE ACCACAMANA 4601 EXECUTERACE ATALENESSE EXCANTERE EXECUTERE EXECUTAR EXECUTERE EXECUTAR EXECUTERE ATALENESSE 4601 EXECUTERACE ATALENESSE EXCANTER A ACCESSE EXCENTER ACCACES TRAITEGACE ANALANAMA ACCALANCE 4601 EXECUTERACE ATALENESSE EXCANTER EXECUTERE EXECUTERE EXECUTAR EXECUTERE EXECUTAR EXECUTERE 4601 EXECUTERACE ACTIVITIES ACCUTERE EXECUTERE EXECUTERE EXECUTAR EXECUTERE ATALENCES A ANALANAMA ANONAMINE 4601 EXECUTERACE ACTIVITIE ACCUTERE A ARALENA A ATALENAS ANTENISATE ATALENASES ATALENASES A ANALANAMA ANONAMINE 4701 EXECUTERACE ACTIVITIE ACCUTERE ANALANES A ATALENASE ATALENASES ATALENASES ATALENASES ATALENASES 4701 EXECUTERACE ACTIVITIE EXECUTER EXECUTER ACCUTERE ATALENASES ATALENASES ATALENASES 4701 EXECUTERACE CACUAGES A ANTALESAN A TALENASE ATALENASES ATALENASES ATALENASES ATALESANCE 4701 EXECUTE ACCORDECAM ANALOGINAL TOWARDON TOTALESAN ATALENASES ATALENASES ATALESANCE ANALANATA ACAGGES ATALENASES 4701 EXECUTE ACCORDECAM ACCAGGANA ACCURENTES ETERACTORE TECRTORES ANALANATA ACAGGESET A TRANSCETTE TECRTORES 4701 EXECUTERACE ACCURERCE A ACCUARTE A TRANSCALA ATALENASES ATALENASES ATALESANCE ANTALESANCE ANTALESANCE ANTALESANCE 4701 EXECUTERACE ACCURERCE A ACCUARTE ACCURERT A ACCACES ATALENASES 4701 EXECUTERASES 4702 EXECUTERASES 4702 EXECUTERASES 4703 EXECUTERASES 4703 EXECUTERASES 4704 EXECUTERASES	6601	AGTAGAAACA	TAAGATAAGG	TIGAAACIIT	TTTAGGGCGA	TTATA TTTTT	AGTOCTATAC	ATGTTTGATG	GCATACTTTT	TAAGGAAGAT	TIATGAAGCT		
 1ATGOGGATT TAGOGGATT ALGOGGTAR AMACGTAR AMACCTARE CITARAGGA ALCONATTY CONTENTS A TOTOGTICE CITCATITY AGAAGGA ACCONTING A ALTONG A COMATCAR A ALTONG A ALTONG A COMATCAR A ALTONG A ALTONG A ALTONG A COMATCAR A ALTONG A ALTONG A ALTONG A ALTONG A ALTONG A COMATCAR A ALTONG A ALCONG A ALTONG A ALCONG A ALC	6701	ATGTTCCACA	TAGGTTTATG	TGCTATTGAT	АТАТАСТАТА	AATAAAACGC	TCTACCCAAG	AAATGAAACA	TTTTTTAAAC	TTTTGTCATC	ассасааааа		
6901 CTAGGTTCAG ALARCATGCA TEGAATCTAT TATCTTCOA CCOACTTT ACAMCGTT AGACCAGCG TTATTCACTG MITAAGGC AAGTTTTAC ACGTGTGACT CTGGGCATT AGACGTTT GACAATCTA TATCTTCOA CCOACTGCG TEADAGGAC AMAMATATG TATTTCGAGG TEADAGGAC AGATACTAT 7101 GTTCTTCAGC ACTTGCTAAT GACATATCTA TTTCTAAGTA CAACAGGATA GTTAACAACT AMAMATATG TATTTCGAGG TEADAGGATA GANTCTCATT 7101 GTTCTTCAGC ACTTGCTAAT GACATATCTA TTTCTAAGTA CAACAGGATA GTTAACAACT AMAMATATG TATTCGAGG TAAGGATA AMAAAAAAACC 7201 TGTAAMAGAC GACAACGTTG GEGTCGGGAA AMAATTAATA GAACCAATATG TTGGATAATC TAACGATTA TACCACGATA TATCGAGGA CTCGGTGTGT 7301 CGACGGTAAT CATCTTTTTA TECGGGACTC TGTAATCAAC TCTNTCTTA AGACCACTA ATGCATAGGAT TATGCAGGAG ATGCAGGA CTCGGTGTGT 7401 GATATTACTC CACGGTGGAA AMACAGGAAT AGAGAGAAA TTGAAAGAC TATGTAATCA TAGCATGGAG ATGAACGGAA ATAAGGTTT CAAGAAATTA 7501 CGACGGTAAT CATCTTTTTA TECAGGACC ATTGTTCACA CTTTTAATAT AGCGAAATTG CGAAGAAGG GTTTCATTGC AAGTATATCA GAAAATTAAC 7701 TCTTTBAATA CAGGGCCTAA TACGAGGCTA CTTCTTAAGA AMAAAAAT AGCGAAATG CCAAGAAAGG GTTTCATTGC AGGTTTATCC AGAATTAACT 7701 TCTTTBAATA CAGGGCGAA GTTCGTGACC ATTGTTCACA CTTTTAATAT AGCGAAATTG CCAAGAAAGGG GTTTCATTGC AGGTTTATCC AGAATTAAC 7701 TCTTTBAATA CAGGGGCTAT CAGAGGCTAA CTTCTTAAGA AMAAAAACC TGCGACACAC TACGACACTT TCCCAACGGA AATACGAAT TATGAGGAC TATTTGGAGGAC ATTTGGAGGAA ATTTAAGGAT 7701 TCTTTBAATA CAGGGGCTAT CAGAGGCAA CTTCTTAAGA AMAAAGAAC TGCGACACAC TAGGCACTT TCCCAACGGA AATACTAATA THTGGATAGG 7701 TTGTGGGTTTC ATTCOACACT GTGTTCAACA TATGTTCAA CAAGAGCAC TACCAACA CAGGGCGCTA CAACTAACTAT 7701 TTGCGGTTTC ATTCOACAT GGAAAGAAC ACCATCAAAT ACCACGCT TACGCACACCA TACGCACCAC TATCTACCC GACTAACAT TATGGAGGAC TCTTTAAGGAC TCCTTAAAAATA AGCAACACA ACCACAGATA AAGACACACT TATGTAGGAC ATTCCACCAC GACTACACTAC	6801	TATGCGGATT	TACGCGTTTG	TCATTCCTAT	AAACCTTAAC	CTTAACGCGA	AACCATATTT	CGCTACCTAA	TGTTGGTTCG	GTGCATTTTG	асалалссал		
7001 ACGTGGACT CEGGCACATT MAGCINETT CACANTCENA AGTETEGTE TEIMAGACE ANAATAATG TATTTGAGE TEATAGETAA GANTCTCATT 7101 GTTETTCAGC ACTTGGTAAT CACATATUEA TITCTAAGTA CACATAGETA CITAACAACT ANTAATGAAT ANACCAGETA ANAANAAAAAACC 7201 TGTAMAGAC GACAAGETEG GEGGGGGAA ANTATTAATA GAACCATAGE TEGGATAGET TAGCAGAT ATACCAGETA ANAANAAAAAACC 7301 CGACGETAAT CACCTTETTA TEGGGCACTE TGAATCAAAC TOTATCTTA ACTACCATA TEGCATATA TATCCAGET ATATCCAATA ATACCAGET 7401 GATATTACTC CACAGETAGA AMACAGGAAT ACAGAGAAAA TTGAAAACAT TATGCATATA TAGCAGET ATATCAGET ATATCCAGAA AACAGGAAT ACAGAGAAAA TTGAAAAGET TATATCAGE ATAGTGAGET TATGTCAGET ATATACAGET ATATCAGAGET ACACAGAGAAGA TCGAATTAGE TETTATCTC CACAGGAGET ATATCAGEGA AGTAACGTTA TACAGAGETA ACAGAGAAAA TTGAAAACAT CTGCACATCAT TAGCAGET TAGCAGAT AAAAGAACA AAAAAAACAA 7101 GGTAATTACTC CACAGGACTAT ACAGAGETAA CITETTCACA CTITAAATAT ACCCAAACTG CAAAAATTCCA TACAAAGETA ATATAACGA CCACAGGACTAT ACAGAGGATAT ACAGAGETAA CITETTAAGA TACCAAACTAC TACCAAAATTG CAAAAATTCCA TACAAAATTAC AAAAAGAAAA ATTAAAAAAAAA TTCAAAGAGATA ATTGAAGAACA TGCCAAATTA TAGGGGGGCTT TCTCAATCGA AAAAAGAAA AATATAAAAATTAAA CAAAAAAAAATTAACAAA AAAAAAGAACA ACAACAACAA TACCAAAAAAAATTAGCA TGCCAATTACAAAAAAAAAA	6901	CTAGGTTCAG	ATATCATGCA	TCGAATCTAT	TATCTTCCAA	CCGAACTTTT	ACAAACGTTT	AGACOGAGCG	TTATTGACTG	AATTAAGAGC	AAGTTTTTAC		
7101 GTICTICAGC ACTIGGTAAT CACATATCIA TITCIAAGTA CAAGATGATA GTIAACAACT AATAATGAAT AAACCAGGTA AAAAAAAAAA	7001	ACGTGTGACT	CTGGCACATT	AGAGCTGTTT	GAGAATCTAA	AGTCTTGCTG	TCTAAGGACC	AAAAATAATG	TATTTOGAGG	TCATAGCTAA	GAATCTCATT		
7201 TGTAMAGAC GACANGETEG GEGTEGGGAA ANTATTANEA GAACATATE TEGATANE TAGCETET TATACACTTA TECCGATEAC ATTETTEG 7301 CGACGETANT CATCITITEA TECGGCACTE TGAATCANAE TETRETTEA ASTACCTAT ATGCCATGAET TATACACTTA TECCGATEAE ATGCGTET 7401 GATATTACTE CACGGLEGAA MACAGGAAT ACAGAGAMA TEGAAAGAT ATATGAAAGAT ATATGAACGG ATGGENEGET AAATAMAATA ACAGCAATTA TAAAGCTET 7501 GETAMATTT CCACGGLEGAA AGTEAGCETA TGCAGTTTE TETRETTEC TTERCTECE TEGCAACHEG GATGAAAAAATA ACAGGAAT ACAGAGAAAA TEGAGTTTE CTEAAATAT AGCCAAATTGE GAAGAAAGG ATTAAACGETT CATAAAGETT 7601 GECTEAGGGEC CGTATAMAAA TECAGTGACC ATTGTTCACA CTETAAATAT AGCCAAATTGE CAAGAATGC A GAATACGAA AATTAAACAT 7701 TCTTTBAATA CAGGGACTAT ACAGGGETA CTECTTAAGA AAMAAACCA TEGCACACA TACGACTTTE ACAAATTCCA THAAGAAT ATTAAGAATT 7801 TAGATAAAT ATCAAAGGA TECTGTTACT TEGTEAAAT ACTCCTTAC ATAAAGATTA TAGGCAACTT TETCACAACTA ATTAAAGAT ATTAAAGAAT ATTAAAGAAT ATTAAAGAT 7801 TAGATAAAT ACAGGACTA CTECTTAAGA AAMAAGGAGE TCCCCCTTAC CAAAAAGTAA TGCCCCAAA TCGCCTTTCC AAGTACCTAA TGGCAGAAT ATTAAGAACA AAATAGGAGG TCCCCCAAA ACCAACAAAT ACCAACAAAA AAATAGGAGG TCCCCCAAA ACCAACAAAT ACCAACAAAA AAATAGGAGG TCCCCCAAA ACCAACAAT ATGCGAGAAT ATTACAACA AAATAGGAGAT CTCAAAGGAAT ATGCAAGGGAAT TTACAACAT TATGTAAATA TAGGCAGAAT ATGCAAGAAAA AACAACAACAACAACAACAACAACAACAACAACAA	7101	GTTCTTCAGC	ACTTGCTAAT	CACATATCIA	TTTCTAAGTA	CAAGATGATA	GTTAACAACT	AATAATGAAT	AAACCAGGTA	алалалала	ладалаласс		
7301 CGACGETANT CATCETETETA TECGGECACTE TGAATCAANE TECHTETETA AGTATECTAT ATGEGAGET TATEGEART ATATEGAAGA CTEGGTGTGT 7401 GATATTACTE CACAGETACA MACAGGAAT ACAGAGAMA TEGAAAGAT ATATGAAEGG ATGETATEGT AMATAMATA ACAGCENTTA THAAGCTET 7501 GGTAMITET CCAGGACGGA AGTTAGETTA TEGAGTTETE TETATETE TETATETE TEGATETET GATATATACT GAAACAGGA ATTAACGETA TAMAGTAT TAGAGETA ATATGAAAATA ACAGCEAATTA TAAAGCETTA CATAAAATTA 7601 GCCTGAGGGC CGTATAMAA TECAGTGACC ATTGETECAC CTTTAAATAT AGCCAAATTG CCAAGAAAGG GTTTCATTGE AGATTATEGE AMATATAAC 7701 TETTTAAATA CAGGGACTAT ACAGGGTAA CTECTTAAGA AAMAAACCA TGCCACACA TAGGACTTTE ACAAATTCGA TAAAAGAAT ATTAGGATGA 7801 TATATTAAT CAAAGGAC TGCTGTACCT TGCTTCAAAT ACCTECTTAC CATAMAGTTA TAGGGCGCT TCTCCACGGA CATACTAAAA TTTGGATAGGA 7801 TATATTAAT CAAAGGAC TGCTGTACTA GTTTAAGTA AAAATGACC AGGGGATTG ATACTCTAA AGGAGCATT CATACTAAAATTATTGGATGAT GTTTCAAACGAA AAAATAGAGAC CCAAAGGACT TGCGAGACAT CAATTCAAA TGGAGGACT TTTCCAAGCC GTTCATGTA CAAGTCCACT TTTCAAACGAA AAAATAGGAGG TCCAAACGAA ACCACAAGAAA ATATAGGAGG TCCAAACGAA ACCACAAGAATA ATGGAGGACT TTTCAAAGCAA CAGGTGCGTT TTCCAACGC GTTCAATGGA ACCACAAGAATA ACCAAGGGATT CCAAGGGAACTA CAAGTGCAATA ATGGAAGAC ACGGCGCAAA ACCGGCAAAAACAAC ACCAATGACTT TTTCAAAGCAA TATGGAAGTA ATGGAAGAC ATATGGAAG GTTCAAACAAC ACCAATGACTT CACAGGGAA TTTCAAACAAT ATGGAAGGGA TGCTGCAAAAACAACAACAACAACAACAACAACAACAACAACAA	7201	TGTAAAAGAC	GACAACGTTG	QCGTCGCGAA	AATATTAATA	GAACCATATG	TTGCATAATG	TAGCCTCTAT	татасастта	TCCCGATTAC	ATTGTTTTAG		
7401 GATATTACIC CACAGIACAA AMACAGGAAT ACAGAGAMA TIGAAAAGAT ATATGAACGG ATGGIATGGT AAATAAAATA ACAGCCATTA TAAAGCTIGT 7501 GGTAANTITI CCACGGACGGA ACTTAGCTIA TGCAGTTIG TITAICITCO TIGCATCIGI GATAMAACA CAAAACAGGA ATTAACGATI CATAAAATT 7601 GCCTGAGGGC CGTATAAAAA TICAGTGACC ATTGITCACA CITTAAATAT ACCCAAAATTG CCAAGAAAGG GITCATTGC AGGTTTAICG AAATAACAT 7701 TCITTAAATA CAGGGACTAT ACAGAGCTAA CITCITAAGA AAAAAAACCA TCCGACAACAC TACGACTITC ACAAATTGCA TAAAAGAAT ATTAGGATGG 7801 TAGATAAAAT AICAAAGGAC TGCTGTACTC TGCTTGCAATT ACCICCTTAC CATAAAGGTA TAGGGAGCTT TCTCCAICGA CATACTAAAA THTGGATAGG 7901 TAITATTAAT GAAATGACTA GITTTAAGTA AAAAAGACAC ACCICCTTAC CATAAAGGTA TAGGGAGCTT TCTCCAICGA CATACTAAAA THTGGATAGG 7901 TAITATTAAT GAAATGACTA GITTTAAGTA AAAATGACT ACCICCTTAC CATAAAGGTA TAGGGAGCTT CATCCAAAGGACTA THTGGATAAGAA AAAACAAC ACCAATACAAT ACCACAGCAA ACCACAAGAAT ATGGCAGCAC CAAATGACTA GATTCATATCTA AAAACTAATTA 8001 TTGCGGTTC AFTCCTCACT TGCTGAAATA COACACGGAT CAAATGAGGA TCCAATT GTGTAGAT GTGGAGACTG TACATGAGA THTCTAAACCA 8101 ACATCAACTC CGGCTTATAA GCTCGGAGA ACCGCGAAC CTCAATGGA TTGCGAAGAT CATATGGAACCA ATAGGGAGCT 8101 ACATCAACTC GGGCTTAAA GCTCGGAGA ACGGAAACC CAAATACTT TTTGAAAGT ATGTGAAATCA TTTGAAGAACA ACCGCCTAGGA TCTTAAGGATA ATGTAAGGAACA TATGGAAGCA TAATGGAAGC TATATGGAAGA CAAATACATT ATGTAATAT TTGACAAGT CTTTTAAGGATA ATACCAAGT 8101 TTCTCAACCC GGCCTTAAA GCTCGGAAA ACGGAACC CAAATACTT TTTGACAAGT CTTGTTACC CTTTTTTACGAACA CAAATACAGA TAAACAGA TAAACAGA TAAACAGA AACGGAAACCA AAATACGAAG TAAATACATT ATTGTAAAAT ATGTAACGAAGA ACATATTGCC CAACAGGAA	7301	CGACGGTAAT	сатетттта	TCCGGCACTC	TGAATCAAAC	TCTATCTTTA	AGTATCCTAT	ATGCATGAGT	TATGTOGATT	ATATCGAAGA	CTCGGTGTGT		
GETAMATTT CCAGGACGA AGTIAGCTIA TGEAGTTTG TTAICTAC TTGCATCGT GTATAATACT GAAACAGGA ATTAACGIT CATAAAATTT GCCTGAGGGC CGTATAMAMA TCCAGTGACC ATTGTTCACA CTTTAAATAT AGCGAAATTG CCAAGAAAGG GTTCATTGC AGGTTATCG AAATATACC TCTTTAAATA CAGGGACTAT ACAGAGGCTA CTTCTTAAGA AAAAAAACAG TGCGACACC TACGACTTC ACAAAATCGA TAAAAGAAA ATTTAAGAT TACTATAATA CAGGGACTAT ACAGAGGCTA CTTCTTAAGA AAAAAAACCA TGCGACACC TACGACTTC ACAAAATCGA TAAAAGAAA ATTTAGATT TAGATAAAT ACAAAGGAC TGCTGTACTC TGCTTCAACA CATAAAATAGA AGCACTTC ACAAAATCGA TGCGACACC TACGAAATTGCA CAAAATCGACTA ATTTAGATAG TATTAATTAAT GAAATGACTA GTTTTAACTA AAAAATGAC ACGAGGCATTG ATACTGTCA AAGTACTCAA TGGCTGTTTGC AGACTTATCT AAAACTAATT TTGCGGTTTC AFTCCTCACT TTGTCAACAA AAATAGGAGG TCCACAGGAA ACCACAGATA ATGAGCCACT CAATTCAGGC GCTTCATGTA GAAGTCCTT AAAATTGGT CGCCCGCAAT GAAAAACA ACCATCAAAA CCAACGGAT TGGCGACACT TTCTAAAGTA GTTGCGACAA CAGTGCCTTT TCCAAGCC TTCTATCGT GTTGAGAAAT ATCTAAACTA TCACTAGGGG TTAATGATGG GACTTCCATT GTGTTGTAGA GTTGCGACAA CAGTGCCTTT TCTAAACGA AACATCAACTC CGGCCTAATA GCTCCGAGGA ACCGGTACCT CCACGTGGGA TCGCAGGACT TTTCTAAGGA GTGCGACAA CATTCAGG TCTTTCCAAG AACATCAACTC CGGCCTAATA GCTCCGAGAA ACCGGTACCT CCACGGAGGA ACAATACATT ATTGTAATTA TTGCACGGGA ATGAAGAACC ATATGGAGCT TTTAATGCA TGTGCAATAG GTCCAAGCAA ACCGGTACCT CCACGTGGAA ACAATACATT ATTGTAATTA TGGACGGAA ACCAGCTAGGA ATAACCAGTT TCTATAGCTA GTCTCATAGA TGTCTAACGA TAATGGAGT TACCGAGAA ACAATACATT ATTGTAATACT ATTGTAATTCC CATTGTTGCC CATTTCTTCCAGGA ATAACAATA ATAACCAGTT TCTATAGCTA GTCCAAGAA TGCCAAGACA AACCGGTACCT CCACGTGGAC AACAATACATT ATTGTAATACT CCTTTTTTC TTTCAAGGAAA ACAATACTGA TGCCAGGAAA ACAATACATT ATTGTTATTCC CCTTTTTTTT TTTCAAGAAA ACAATATGGAT TACCAAGTG CACATTCTT GCTTACTGAA TGCAAGAG TAAATGGAT TACCGACA AACAATACTT TTTGACAAGT TTGGGGAAAC CAAAATAGAA ACAATATGCA CTCCAATGCC CTTGCCGGAT TTTTAACGAAT ATAACGAATG AGATCACCC CGATTAGCTG TAGGGAAGCC TTGGGGAAAC CAAAATAGAA ACCTTTGATA 88001 ATTCTAATATA TCCCGGTCAAC GTAAACGAAG GAATACAATG TGCCC CAATGCCA CTTAACTAA CAAACTGCT CCAAGGACT TGGGGAAAC CAAAAATAGAA ACCTTTGATA 88001 ATTCTAATA TCCCGGTCAAC GTACCAAATC GCAATACGAG TGTATACCC CCAATGCCA CTTAA	7401	GATATTACTC	CACAGTACAA	AAACAGGAAT	ACAGAGAAAA	TTGAAAAGAT	ATATGAACGG	ATGGTATGGT	алаталата	ACAGCCATTA	TAAAGCTIGT		
7601 GCCTCAGGGCC CGTATAAAAA TTCAGTGACC ATTGITCACA CTTTAAATAT AGCCAAATTG CGAGAAAGG GTTCATTGC AGGTTATCG AAATATTAAC 7701 TCTTTAAATA CAGGGACTAT ACAGAGGTAA CTTCTTAAGA AAAAAACA TGCGACACA TACGACTTTC ACAAATTCCA TAAAAGAAA ATTTAAGATT 7801 TAGATAAAAT ATCAAAGGAC TGCTGTACTC TGCTTCAAAT ACCTCCTTLC CATAAAGTTA TAGGGAGCTT TCTCCATCGA CATACTAAAA THTGGATAGGAT 7801 TAGATAAAAT ATCAAAGGAC TGCTGTACTC TGCTTCAAAT ACCTCCTTLC CATAAAGTTA TAGGGAGCTT TCTCCATCGA CATACTAAAA THTGGATAGGAT 7801 TATTATTAAT GAAATGACTA GTTTTAAGTA AMAATGATC GCGGGACTTC CATACTCAA AGGTACCAA TGGCTTTCC AGACTTATCT AAAACTATTT 8001 TTGCGGTTTC ATTCCTCACT TTGTCAACAA AMATAGGAGG TCGACAGCAA ACCACAGATA ATGAGCACT CAATTGAGC GTTCATGTA GAAGTCCTT 8101 AACATTGGCT GGCCCGCAAT GGAAAACAAC ACCATCAAAA CCAAGGGG TTAGGGGAGCTT TTCTAAAGTA GTTGCGACAA CAGTGCGTTA TTCATAACGC 8301 ACATCAAACTC GGGCTTATAA GCTCGGAGATA CGTCGGAATCC TCTAAGGGG TTATGGAGGT GTTGTTACTC CGTTTTTTC TTCCAGGA ACGGGAGCT TTTCAAGGATA AFAACCAGGTT 8401 TTTAATGCAA GTCCGAGCAA ACCGGTACCT CCATGGGAC AAATACATT ATTGTAATTA TTGCAGGGGA ACAATACGAGTT 8501 TCTATAGCAA GTCCAAGAA TAATGGAGT GTCACTGCCA AAATACATT ATTGTAATA CGAAAGAAAAAAAAAA	7501	GGTAAATTTT	CCAGGACGGA	AGTTAGCTTA	TCCAGTTTTG	TTTATCTTCC	TTGCATCTGT	GTATAATACT	GAAACACGCA	ATTAACGITT	CATAAAATTT		
TOTTTANATA CAGGGACTAT ACAGAGCTAA CITCITAMAA AAMAAACCA IGCCACACC TACGACTITC ACAAAITCCA TAAAAGAAT AITTAAGATT TAGATAAAAT AICAAAGGAC IGCTGTACIC IGCTICAAIT ACCICCITAC CATAAAGTA TAGGCAGCIT ICTCCAICGA CATACTAAAA ITTGGATAGG TAITAITAAT CAAAIGAAC IGCTGTACIC IGCTICAAIT ACCICCITAC CATAAAGTA TAGGCAGCIT ICTCCAICGA CATACTAAAA ITTGGATAGG TICICGGITIC AITCCICACI ITGICAACAA AAMAAGAGGC ICCACAGCAA ACCACAGATA AIGAGCACI CAAITCAGCC GCTICAIGIA GAAGTACTA AACAITIGGT CGCCCCGAAIT GAAAACAAC ACCAICAAAA CCAAGGGAT IGGCAGACIT ITCIAAAGTA CGACGGCTIT ITCAICAGCC ITCICAICCI GITGAGAAAT AICTAAACIA CACAICAAAA CCAAGGGGA ICGAGAGCAIT ITCIAAAGTA GAIGGCGCIT ITCCAICAGC ACAICGACCIC GGCCITAITA GCAICGAGAT AICTAAACIA TOXIIAGGGG ITAAIGAGG AITCCAIT ITCIAAAGTA GITGCGACAA CAGGGCGCIGAG ICTITICCAAGC ACAICGACCIC GGCCITAITAA GCICGGAGTA CGICGGAAICC ICCAAGGGG ITAAIGGA ACAITACAIT ITTCAAGATIG ITTICCAAGCA GCGCGCIGACG ICTITICCAAG ACAICGAACIC CGGCITAITAA GCICGGAGTA CGICGGAAICC ICCAAGGGA ACCAATACIT ITTCAAAATA TIGGAAGAGA ACGAGGAGCI ITTICAAAGAACA AACAGGAAAA AACGGIACCI CCACITGGCA ACAATACAIT AITGGAAGT CITTITACGAAG GICCAGGCAA ACGGGAACC AIAIGGAACA ACAICAACTIC CGGCCITAIAA IGICIAAACA IAAIGGAIG IGACIGACIT ITTICAACAATIA AITGGAAGT AITGGAAGAACA AACGAGTA ITTIAAITGCA ICTCCAAAACA TAAIGGAAG TAAIGGAIG IGACTACCIC CGATTAGCIG GAIGGAAAAAAAAAACAATAAGAACAGTA ACAICAACTIC GCCCATAGGA ICCAAAACAG TAAAIGGAIG IGACTGCACA AACGAGTG CITGITACCIC CCTTITITTC ITTCAGGAAA ACAATACGAGT ICTAIAGCTA GICICAAAACAG TAAAIGGAIG IGACTGCCA AACGAGAAC CAAAATAAAAA AAAACAAGT AAACCAGTT ITTIAACTAT ICCAAAACAG TAAAIGGAIG IGACTACCCC CGATAGGTG AAGAGGAAAC CAAAAAAAGA ACAITAIGCA CICCAAGGACT IGGGGAAAC CAAAATAAGA ACCITIGAIA ACTICIAATAA ICCCAGGCCI ITTITAACTGI TAAACGAAG GAACAATACCC CGATTAGGTG AGAGGAACC AIAAATAAGA AGCITIGAA AATCCAAGGC CITGCCAGGC TITTTAACTGA TAACGAAG GAACAAGCG IGAACGAAACCAAATAAGA ACCAAGTGCA ITAACGAAG IGAACAACCA CIAAAATAAGA AGCITIGAAA AATCCAAGCC CITGCCAGCC GATCCCAAACCG GAACGAAGCG IGAACGAAACAAACAAGAACA AAATCAAGAG IGAAGAACA ITAACCAAGTG IGAAACAAACAACCA CIAAAATAAGA ACCAAGTGCA ITAACGAAG IGAACAACAACAACAACAGAACAACAACAAGAACA IGAAACAACA CIA	7601	GCCTGAGGGC	CGTATAAAAA	TTCAGTGACC	ATTGTTCACA	CTTTAAATAT	AGCCAAATTG	CCAAGAAAGG	GTTTCATTGC	AGGTTTATCG	AAATATTAAC		
7801 TAGATAANAT AICAANGGAC TECTGTACIC IGCTICAATT ACCICITAC CATMANGITA TAGGCAGCIT ICTCCATCGA CATACIMAA TITGGATAGG 7801 TAITATTAAT GAAAGGAC TECTGTACIC IGCTICAATT ACCICITAC CATMANGITA TAGGCAGCIT ICTCCATCGA CATACIMAA TITGGATAGG 7901 TAITATTAAT GAAAGGAC TECTGTACIA GITTTAAGIA AMAATGAIC ACGAGGATTE ATACICITCA AACTACCA AGGCATTE CAATTCAGC CATTGATACA AGAGCITACIA AMAACTATIT 8001 TIGCGGTTIC AITCCTCACI TIGTCAACAA AMATGGAGG TCCACACGA AACCACAGATA ATGAGCCACI CAATTCAGC CACTGATATA AMAACTATIT 8001 TIGCGGTTIC AITCCTCACI TIGTCAACAA AMATGGAGG TCCACACGA AACCACAGATA ATGAGCCACI CAATTCAGC CACTGCATTITICAAGCC GTTCAATAA CAACATTGGT GGGAAACIC GACGCCGCACA CACTCCATTITICAAGCC TITCCAAGC 8201 TICTCATCCCT GTTGAGAAT AICTAAACIA TCACTAGGGG TTAATGATG GACTTCCAATT TITCCAGACA GCCGCGAGC TCTTTCCAAGC 8301 ACATCAACTC CGGCTAAGA TECCTAAGGA TAACGGATA CACATAGCT CAATAGGAT ATTGTAATA TCAGGAAAC AAAATAGAA AGCATTGGATA 8401 TITTAATGCA TGTCCAAGA TAACGGAAA ACGGTACCT CCAGTGGGA AACAATACATT TITGACAAGT CATTGTTATC CCATGGGATA ATAACCAGTT 8501 TCTATAGCTA GTCTCAAGA TECTTAACTG CAATGGATG TGTGTCCC CGATTTTTTC TITCAGGAAA ACAATATGCC CAGGTTTGCC 8601 CACCATTCTT GCTTACGGAT TAACGAATG AGATCATCCC CGATTAGCTG TAGGGAAACC AAAATAAGA AGCTTTGATA 8801 ATTCTAATA TCCGGGCTAAC GAATGCAATG GATGAATCCA CTAATGCCA CTAAACTAA CAAAAAAGAAGA CAAAAATAGA AGCTTGATA 8801 ATTCTAATA TCCGGTCAAC GACCAAATC GAAATAGCAG TGTATACCC CCAATGCCA CTAAACTAA CAAAATGGAAA CAAAATAGAA AGCTTGAATA	7701	TCTTTAAATA	CAGGGACTAT	ACAGAGCTAA	CTTCTTAAGA	алалаласса	TGCGACACAC	TACGACTTTC	ACAAATTCCA	TAAAAGAGAT	ATTTAAGATT		
7901 TATTATTAAT CAAATCACTA GTTTTAAGTA AMAATGATC ACCAGCATTG ATACTCTTCA AAGTACTCAA TGGCTTTTGC AGACTTATCT AMAACTATTT 8001 TTGCGGTTTC ATTCCTCACT TTGTCAACAA AMTAGGAGG TCCACAGCAA ACCACAGATA ATGAGCCACT CAATTCAGCC GCTTCATGTA GAAGTTCCTT 8101 AACATTTGGT CGCCCCCCAAT GGAAAACAAC ACCATCAAAA CCAAGGGAT TGGCGGAGCT TTCTAAGGAC GCTGCGCTT TTCATCAGCC 8201 TTCTCATCCT GTTGAGAAAT ATCTAAACTA TCACTAGGGG TTAATGATGG GACTTCCATT GTGTTGTAGA GTGGGAGCTG TACATTCAGA TTTTCCTAAC 8301 ACATCAACTC CGGCTTATAA GCCCGGGTA CGTCGAATCC TCTAACTGCA ATCACATTG TTTCCAGCA GCCGCTGACG TCTTTCCAGG 8401 TTTAATTGCA TGTGCAATAG GTCCTAGGGATA CGCGGGAGC TCTAGGGAGAC ATAAGGAGCA ACAGAGGACA ACCAGGTTG 8501 CCTATGCTA GTCCTCATGGA TGCCAGCCA ACCGGTACCT CCACTGGAGT ACAATACATT ATTGTAATTA TTGCAGGGAG ATGAAGAACC ATAAGGGATT 8601 CACCATTCTT GCTTACTGAA TGCAAACAG TAAATGGATG TGACTGCCAG AAGTACAGTG GATGTAAAA CGAAAGGAAA ACATATTGCC CAGGTTTGCC 8701 CCTCCATGCG CTTGCCGGGT TTTTTAACTGT TAACGAATG AGATCACTCC CGATTAGGTG TAGGGAAGC CAAAATAAGA AGCTTTGATA 8800 ATTCTAATA TCCCGGTCAAC GATCCCAAATC GCAATACCG TGTATCAC CTCAAGGAAC CAAAATAAGA AGCTTTGATA 8800 ATTCTAATATA TCCCGGTCAAC GATCCCAAATCG GATGGCAATGCA CTCAATGCCA CTAAACTAG CAAAATAGGA TGTTGATAA	7801	TAGATAAAAT	ATCAAAGGAC	TGCTGTACTC	TGCTTCAATT	ACCTCCTTAC	CATAAAGTTA	TAGGCAGCTT	TCTCCATCGA	CATACTAAAA	TTTGGATAGG		
8001 TTGCGGTTTC ATTCCTCACT TIGTCAACAA AMATAGGAGE TCACACAGAA ACCACAGATA ATGAGCCACT CAATTCAGCC GCTTCATGTA CAAGTCCTT 8101 AACATTGGT CGCCCGCAAT GGAAAACAA AMATAGGAGE TCACACAGAA ACCACAGATA ATGAGCCACT CAATTCAGCC GTTGCACAGA CAGTGCCTT TTCATCAGCC 8201 TTCTCATCCT GTTGAGAAAT ATCTAAACTA TCACTAGGGG TTAATGATGG GACTTCCATT GTGTTGTAGA GTTGCCACAA CAGTGCCTT TTCATAACC 8301 ACATCAACTC CGGCCTAATAA CCTCGAGTA CGTCGGATCC TCTAACTGCA ATCACTAATT TTACAGATG TTTCCAGCA GCCGCTGACG TCTTTCCAGG 8401 TTTAATTGCA TGTGCAATAG GTCCTAGCGA ACCGGTACCT CCACTGGGA ACCAATACATT ATTGTAATAT TTGCAGGGA ATGAAGAACC ATATGGAGT 8501 TCTATAGCTA GTCCTACGAA ACCGGTACCT CCACTGGCGA ACCAATACTT TTTGACAAGT CTTGTTACTC CCTTTTTTC TTTCAGGATA ATAACCAGTT 8601 CACCATTCTT GCTTACGAA TGCCTAGGAG TAATAGGATG TGCCCAC AACAATACATT ATTGTAATAA CGAAAGAACA ATAATGCAGT 8601 CACCATTCTT GCTTACTGAA TGCATAGCGA TAATGGATG TGCCCCAC AACAATACTT TTTGACAAGT CTTGTTACTC CCTTTTTTC TTTCAGGATA ATAACCAGTT 8601 CACCATTCTT GCTTACTGAA TGCAAAACG TAATGGATG TGCCACCACAATACTT TTTGACAAGT TTGGGGAAAC CAAAATAAGA ACAATATGCC CGATTAGCTG 8701 CCTCCATGCGC TTTTTAACTGA TAAACGAATGA TAAATCGAAGT GTCCCCC GAATGCAGTG GATGTAAAAAA CGAAACGAAA	7901	TATTATTAAT	GAAATGACTA	GTTTTAAGTA	AAAAATGATC	ACCAGCATTG	ATACTCTTCA	AAGTACTCAA	TGGCTTTTGC	AGACTTATCT	AAAACTATTT	¥	
8101 AACATTIGGT CGCCCCGAAT GAAAACAAC ACCAICAAAA CACAICAAAA CACAIGGAT TGGCAGACT TITCAAAGTA GTTGCCACAA CAGIGCCTT TICATCAGCC 8201 ITCTCATCCT GTTGAGAAT AICTAAACAA CACAICAGGG TTAATGATG GACTTCCATT GTGTTGTAGA TGTGCAGCAA CAGIGCCTT TICATCAGC 8301 ACATCAACTC CGGCTTATAA GCTCCGAGTA CGTCGAATCC TCTAACTGA ATCACTAATT TTACAGATG TTTTCCAGCA GCCGCTGCG TCTTTCCAGC 8401 TTTAATTGCA TGTGCAATAG GTCCAGCCAA ACCGTACCT CCACTGGAA AACAGTACATT ATTGTAATTA TTGCACGGAA GACGCTGCAA TAATGGAACT 8501 TCTATAGCTA GTCCTATGGA TGTCTTACCT CCATTGCGAC ACCATACTT TTTCACAAGT CTTGTTACTC CTTTTTTC TTTCAGGATA ATAACCAGTT 8601 CACCATTCTT GCTTACTGAA TCCAAAACAG TAATGGATG TGCTGCCAC AACATACTT TTTCACAAGT GAGACTT TGGGGAAAC CAAAATAAGA ACTTTGCTA 8701 CCTCCATGCG CTTGCCGGAT TTTTAACTGAT GAATGGATG GACTACCCC CGATTAGTGTG TAGGGAAGCT TTGGGGAAAC AAAATAAGA ACTTTGATA 8801 ATTCTAATA TCCCGGTCAAC GATCCAAATC GCAATAGCAG TGTATATCCA CTCAATGCCA CTAAACTAA CAAACAGTGCT CAAAATAAGA ACTTTGATA 8800 ATTCTAATA TCCCGTCAAC GATCCAAATC GCAATAGCAG TGTATACCA CTCAATGCCA CTAAACTAA CAAACAGTGCT CAAAATAAGA AACTTGAATG	8001	TTGCGGTTTC	ATTCCTCACT	TTGTCAACAA	AAATAGGAGG	L TCCACAGCAA	ACCACAGATA	ATGAGCCACT	CAATTCAGCC	GCTTCATGTA	GAAGTTCCTT		
8201 ITCTCATCCT GTTGAGAAAT AICTAAACTA TCACTAGGGG TTAAIGATGG GACTICCATT GTGTIGTAGA TGTGAAGCTG TACATTCAGA TTTTCTAAAC 8301 ACATCAACTC GGGCTBATAA GCCTGGAGTA CGTCGAATCC TCTBACTGGA ATCACTAATT TTACAGATTG TTTTCGAGA GCGCGCGCG TCTTTCCAGG 8401 TTTAATTGCA IGTCCAAGAG GTCCAGGCAA ACCGGTACC TCTBACTGCA ATCACTAATT TTACAGATTG TTTTCGAGGA GCGCGTGAC ATTATGGAGCT 8501 TTTAATTGCA IGTCCTAAGA IGTCTTACCT COATGCGAC ACACATACTT TTTGACAGCT CTTGTTACTC CCTTTTTTTC TTTGACGGAA ATAACCAGIT 8601 CACCATTCTT GCTTACTGAA ICCAAAACAG TAAATGGATG TGACTGCCC GAATGAGTG GATGTAAAA CGAAAGAAA ACATATECC CAGGTTTGCC 8701 CCTCCATGGC CTTGCGGGT TTTTTAACGATG AGATCATCCC CGATTAGTTG TAGGGAAGC TTTGGGGAAA CAAAATAAGA AGCTTTGATA 8800 ATTCTAATA TCCGGTCAAC GTCCAAATC GCAATGCGG TGTATACCA CTCAATGCCA CTAAACTAA CAACATGCT COAACATGCA TAAAACAAGACA	8101	AACATTTGGT	CGCCCGCAAT	GGAAAACAAC	ассатсаааа	CCAAGTGGAT	TGGCAGACTT	ттсталаста	GTTGCCACAA	CAGTGCCTTT	TTCATCACC		
8301 ACATCAACTC CGGCTATAA GCCTCGAGTA CGTCGAATCC TCTAACTGCA ATCACTAATT TTACAGATTG TTTTCCAGCA GCCGCTCACG TCTTTCCAGC 8401 TTTAATTGCA IGTCCAAGAG GTCCAGGCAA ACCGGTACCT CCAGTGAGA ACAATACATT ATTGIAATTA TTGACGGGAA CAAAAAACAAGTT 8501 TCTATAGCTA GTCTCATAGA IGTCTTACCT CCATTGCGAC ACACATACTT TTTGACAAGT CTTGTTACTC CCTTTTTTC TTTCAGGATA ATAACCAGTT 8601 CACCATTCTT GCTTACTGAA IGCCTTACTG TAAATGGATG TGACTGCCAG AAGTACAGTG GATGIAAAAA CGAAAGAAA ACATATIGCC CAGGTTTGGC 8701 CCTCCATGGC CTTGCCGGTT TTTTAACTGT TAAACGAATG AGATCATCCC CCAATGGCTA GAGCGAAGC TATGGGGAAC CAAAATAAGA ACCTTTGATA 88001 ATTCTAATAA TCCGGTCAAC GATCCCAAATCG GATAGCAG TGTATATCGA CTCAATGCCA CTAAACTAG CAACATGCCA TAAAAGAACA	8201	TTCTCATCCT	GTTGAGAAAT	атсталаста	TCACTAGGGG	TTAATGATGG	GACTTCCATT	GTGTTGTAGA	TGTGAAGCTG	TACATTCAGA	TTTTCTAAAC		
8401 TITRARTECA TETECARTAE ETECARECA ACCERTACE CONTEGER ACAATACATE ATTERATE TEGACEGAE ATGAAGACE ATATEGACE 8501 TETATAGETA ETECTATAGA TECCARACE CANTEGERE ACAATACET TITECACAAGE ETERTATE TEGACEGAE ATGAAGAAC ATATEGE 8601 CACCATECT ECTTACEGAA TECAAAACAE TAAAEGATE TEACTECCAE AAGTACAEE GATEBAAAA CGAAAGAAA ACATATECC CAGETTEGE 8701 CETECATEGE ETECCEGET TITTAAETE TAAAEGAATE GEATEATECE CEATAGETE TAGEGAAGE TIGEGGAAC CAAAATAAGA ACETTEGATA 88001 ATTERATA TECEGETAAE GATECAAACE GATAGEAGE TEATATECA CTCAATEGE CTAACTAE A CAACATEGE TAAAAGAACA	8301	ACATCAACTC	CGGCTTATAA	GCCTCGAGTA	CETCGAATCC	TCTAACTGCA	ATCACTAATT	TTACAGATTG	TTTTCCAGCA	GCCGCTGACG	TCTTTCCAAG		11
8501 TCTATAGCTA GTCTCATAGA IGTCTTACCT CONTIGCGAC ACACATACTT ITTCACAAGT CTTGTTACTC CCTTTITTC ITTCAGGATA AFAACCAGTT 8601 CACCATTCTT GCTTACTGAA TECAAAACAG TAAATGGATG TEACTGCCAG AAGTACAGTG GATGTAAAAA CGAAAGAAAA ACATATTECC CAGGTTTGCC 8701 CCTCCATGCG CTTGCCGGGTT ITTTAACTGT TAAACGAATG AGATCATCCC CGATTAGTTG TAGGGAAGCT TTGGGGAAC CAAAATAAGA AGCTTTGATA 8801 ATTCTAATAA TCCCGTCAAC GATCCCAAATG GATAGCAG TGTATATCCA CTCAATGCCA CTAAACTAA CAACATGCCA CAAAATGAA CAATGCAA TAAAAGAACA	8401	TTTAATTGCA	TGTGCAATAG	GTCCAGGCAA	ACCGGTACCT	CCAGTGAGTA	ACAATACATT	ATTGTAATTA	TTGACOGGAG	ATGAAGAACC	ATATGGACCT		E
8601 CACCATTETT GETTACTGAN TECANARCAG TANATGGATG TGATGCCAG ANGTACAGTG GATGTANAN CGANAGAAN ACATATIGEE CAGGTTTGEE 8701 CETECATGGE CITGECGGETT TITTAACTGET TANAEGANG AGATEATECE CGATTAGETG TAGGGAAGET TIGGGGAAC CAAAATAAGA AGETTTGATA 8801 ATTETAATA TEEGGTEAC GATECGAATE GOATAGEAG TGETATECCA CTEAMEGEA CAAACATGEE TAAAACAAGEAA	8501	TCTATAGCTA	GTCTCATAGA	IGTCTTACCT	CCATTGCGAC	ACACATACTT	TTTGACAAGT	CTTGTTACTC	CCTTTTTTTC	TTTCAGGATA	ATAACCAGIT		
8701 CCTCCATGGC CTTGCCGGTT ITTTAACTGT TAAACGAATG AGATCATCCC CGATTAGTTG TAGGGAAGCT TTGGGGAAAC CAAAATAAGA AGCTTTGATA 8801 ATTCTAATAA TCCGGTCAAC GATCCGAATG GGATAGCAG TGTATATCCA CTCAATGCCA CTTAAACTAA CAACATGCTC CCAAAGATGCA TAAAAGAACA	8601	CACCATTCTT	GCTTACTGAA	TCCAAAACAG	TAAATGGATG	TGACTGCCAG	AAGTACAGTG	GATGINANN	сдаласдала	ACATATIGCC	CAGGTTTGGC		
8801 ATTOTANTAN TECEGITEANE GATECEANTE GENINGEN TETRINECA CTEMPTECE CTANECTA CANENTEEL TANANGANEN	8701	CCTCCATGGC	CTTGCCGGTT	TTTTAACTGT	TAAACGAATG	AGATCATCCC	CGATTAGTTG	TAGGGAAGCT	TTGGGGAAAC	CAAAATAAGA	AGCTTTGATA		
	8801	ATTCTAATAA	TCCGGTCAAC	GATCCAAATC	GCAATAGCAG	TGTATATCCA	CTCAATGCCA	стталастаа	CAACAIGCTC	CCAACATGCA	T AA AAGA AC A	1	

Figure 2. (Continued).

8901	TTGCACCAAG	GACGATATGA	AGAAAAAGAA	AGCTTCATA	AAAATACTTC	CTGAATACTG	саааддаааа	GAAAACCATT	GTGCCAQCTA	AACAAAGTGC	¥
9001	TGCCACCCCG	AATTGCCAAT	ATAATCGGTT	TTTACTTGTT	GCCCAAGTTT	TATICGCTAC	CGTATAACTA	GTATATGCAG	AGCCATGAAT	CATAGCATCC	
9101	AAAAACATCA	TTCTTCCCAA	CATTIGIGA	AACATAATGA	Aggaagtata	TTTAACCCCA	GAAATATACT	CGAGAAAGTT	GTTTCTTCCC	GCGAAAAGAA	
9201	CTATTAATGG	AAAGTGTGCG	AATGCGAGTA	CACCACTTCT	GTCTGCCACA	TATOGAGCAA	CTTGAACTCT	ACGAGACTTG	AATATTATAT	TTTCAGGATC	
9301	GTATTCATAA	CCGTATGCCA	AAAAAACGGT	ATGAAGCACG	AGATATCCAA	GAATAATGAT	ACCCTCTAAT	CTTGTAGGTA	AATATCCTGT	GAATATTCTA	
9401	AAATAAGAGA	AATCTGACGC	AIGTTIGCTT	CCAATGGTAG	GTAGAGTGAG	GTATCCCCTT	ACATATCCCA	CAAGCTITTG	CTTCANDAGA	ACAGTTTIAA	
9501	AGGGAGTGTA	ATTCATGCAA	TGAAGAACGC	CTGCAAAAGC	CATGATAGOG	ACAAAATAAG	CACATATAAT	GCCACCATAT	ATGTTTCCGA	CATCTAAGIT	
9601	AGCATAGAAA	CCATGCAATG	CATGATAGTA	GGCTTTTCTT	AACTGTGTAT	CCATTICAAC	TGGADATGTT	AGGTTIGCAC	TGCCTTTGAC	ATAGGGCGEC	
9701	ATATGCCTCG	TTCCGTTGTT	TAGAACGTCA	TAAAACTCAG	AATTAGTCAT	GTTCTGAAGA	GCATCACTAT	AACTTOCACA	GTTTTTTTTG	ATGGTACTAA	
9801	ATGTTTTTTC	TAAGGTTTTA	TTCGAATAAC	CTITTTCATC	CAACGTTTCA	TAGATACAAT	ACAGCATTGA	TTGAAACGCT	GTTCGTAAG	AACATACTCG	
9901	TGATGAACGC	TTAGACTTGC	IGGTATACTC	CCAAGTAACT	TTTTGAAAAA	TCCTCGTGCA	GGCATTAGTG	ACAAGCAAGG	GCACCTTGTT	ACGAATAACT	
10001	GTCTTTGCAG	GTGATGCTCT	IGCTCCTGAC	AGGCAGAAAA	GCAAAATAGC	GCTCAAGATG	GACGICCAAT	GCATTATATT	CGTTGGGTTT	CATAAGCTIT	1
10101	ggataaaaaa	GACTTTATTA	CAACTGTAGC	AAAAAATCAT	GTTCATTTTT	TGCGACTTAT	ATATTTAGTG	GCAAATTAAT	GATACCTTGC	AATAACCGAA	
10201	TTTAGAGTAT	GTTGTCCAAG	AAAGGAGGGA	TTTTGTTCAT	CAGAAAA <u>GAA</u>	TTCAGAAAAG	CAAGGAAACA	GTACTATCGT	TTAGAAIGTA	GAATGATAGG	
10301	TTGCTTGCTA	ATTCTATTAT	GCACGAATG	ATACACCCAT	ATTITCAACA	AAATCAATAC	CCACTAGCAT	CATTGAGCCA	ACTATTIGTC	AATGCAACCA	
10401	TTACCGGTAC	TTCATOCTGA	TTTAACCAGT	CTACTTTTTT	ATCACGTCAA	AATTTACTTG	TTTTCCTGTA	AACCCGAAAT	аласссалал	AAGACCTOGG	
10501	TGCAATTACG	AATAAATGTA	CAATAATCAT	CCTGTTTGCA	TAGTAAACTT	CCAGTTAGAG	TCACACAACG	CAATGAATTT	TGACAGTTTT	CTGTGCGATA	
10601	TTCTTTGGTA	AACGTAAAGA	ACAGGCAACT	TTTGGTACAA	TGCATTCTAG	CCCATATGGT	TCATTTCTGG	TGCATTCGCA	AAGTCAGTAT	TIGTCTACCT	
10701	GTGTTTTCTG	GCTGAGAGAC	ATTATGATGT	TATTCATIGI	TAIGGATATC	TCTGTAGCTC	ATGCTGCTTA	TTTCTCCTA	AAAAAGTTTT	TTCTCTCGAA	
10801	TACATTCTTG	ACCATTTCAT	AGTGAAATTC	TIGTACTIAI	TTAAAACCAA	AAATGGAAGT	ATTCATACAT	CCCCCTATCA	AAAACACTCA	ATAAGTTTCG	
10901	AATTATTCGT	TCGTCTAAAC	AGTGTCCAAT	ACTCAAAGGG	GTATTCAAGA	CGGCACAAAA	TCAGCATCTT	CCCTTATCCG	TGTTCCAGAA	ATACCACOCT	
11001	AAGGTTTTTC	CTCCTACAAT	CATAAAATC	ATTAAGGAGG	CACCTTGAAA	AATCITGAAA	TTCAAAAGAG	TTATCTIGGG	CTAATCGAAA	TTAACGATAA	
11101	CCAGAGTAGA	ATATTCAAGA	TCACAGCTCC	ACCITAGITI	CGAGGTTGCT	CTCAATGGTC	TAGCTTAGTC	ATGCTITTTG	тгааасааат	CITGCCTITI	
11201	CTGATATTCI	ATAGTGTGGI	CAATCCAGA	TATTGGCGCA	TCCTARACCA	ATTATAGTTC	ACTTATGAAG	GGCAAATGAA	CACTCTTACC	AGTTTCACAA	L.
11301	ATAATTAGTA	CTAGCTAGGC	GTCAAACTG	ACTAGCTIAT	TATAATATTA	TCGATGTAAC	ATTTTTCCAT	AGTGAQGGAG	GAAAATAAAC	СТТАААААА	k.
11401	TATGCAAGGA	TTCTAGATTI	CAGGCTAGTC	TICAACCACI	TTOCACTTCC	GGATTAAAGG	CTTTIATAGA	AAACAICTTG	GTTTGGGACT	ATATTGCTAG	
11501	GTGAAAATT1	CAAAATTGTA	CAAAAGATCA	TTCTTGTTTT	TTIGCCTCAT	TAGCTGATAC	AATGOGGAAC	TACATOGCGO	TCTTTTATAA	GCATCTACTA	
11601	TGGTACATTO	TAAGTATGTC	GATCTTOGTI	AACACATAGA	CATTTTTTA	CTTCTTGATA	TCGGITCCGC	TAAATTIGCI	GATTATATT	GTTCACTAGG	;
11701	CCAGATCATA	ATGCCTGTAA	TATTGAGTAC	TTAATATOG	TACTTGATAT	CCGAACAGTA	AACGITTGTC	CGTAGTCATC	GTGACTOGTI	GACGTATINA	L.
11801	CTGCAAGAAA	TGATGGAAAA	AATTTAGGTA	TATTTGCAG	TGAAAACTCT	AGAACACTGO	CGTTGAAGAA	AGAATOGCTI	CAGCCGTCAT	GCTGTCTATC	:
11901	TGTTTGTAAF	CTTTTGGAAG	TTTTACATA1	TTTGCACTA	TTTCTTTAGC	ACGAGGTATA	TGTTTCTTT	TCAATAAAA	TCGTCCTTTI	GAAGAGCCAA	•
12001	GAACACTTTT	GATTTGAGTO	TATAGGATAA	TAATATTGC	TCTAGCGCTT	AATGACCATO	TTTGTACTTA	CCTCTCTCCC	ACCTTGTATA	GGCGGATTGG	;
12101	CTATAATAAC	GCGTTATATO	CATCAAAATA	TGTAGTTCTC	TTTCATTTDA	GCAAATTCGT	AATACGACTC	CAGATAATAA	TACCAAGTCA	TTTTTGGGTA	L.
12201	CCAGTGAGA	TTGGGTCTCF	TAACTGATTG	TGAATCTCA	TATTACAGTG	ATGTGAGGG	AAATACTTAA	TTAAAAAGG	AAAAAGAAAI	GTTACTTTTC	2
12301	TATTTTAAAI	GAAGATTCAG	G TTTGTTAAAC	ATGGATTITO	CATAGCATTA	GGTTTTTAT	TTTCCAAATC	CATTTUATT	ANACTITICA	ATCATGGGAT	5
12401	ATAAAGTTAC	C ATGCAATAGO	S AAAAGTTGTC	ATGAACTOCI	A TTIGAAACTA	AGGAAGAACI	GTAACAGAAG	G TTTTTIGTT!	TGAATTATT C	ATGAACTTAG	;
12501	AGCTGCTAAT	TTGGAGTGGG	G TCCAAGTCAT	CATTTTTT	G CGGGGCTATG	ATTACACAA	CTGCGCACA1	ANTATAATCO	GCTTTGTATC	TTTAACATC	4
12601	AGTAAGACCI	TGAAAATTT	A TTCTATCTT	AGACAATGA	CGGAGTTACA	GAACGATTT	CGTCTTCCCC	ATTTGATCA	AAACTAACTO	AAAATTGCAG	3
12701	TGTTGTAAT	TTTGAAACGO	COGAACGAAGT	TTTCACCAA	C TTAACAGAAG	CAGAAGACCO	TTCTTGAACT	R AGGCTGCACT	GTAGAACGC1	GAGCACTIT	4
12801	CCATTCACC	GCTATTTT	ACGGAAAA	TCAATAAATO	CTTGAATTCC	CTTCATTTA	ATGCCAATT	A CACCTOGTAT	ATCTAAAAT1	GCTGACATAT	r
12901	ATTATTTCT	T TTTGATACT	C CCTGGGAAI	CTCACAGAA	A GGTGGCTCTA	GAATACAGA	GAACAAAGT	A CATAGICCG	AATTGATGT	TGAAGTTAAT	r
13001	GGGAAAGTAG	GAATGGTACT	GICCACTCAP	GTAATACTT	GAGATTGANA	TTTGAAGCT	TTTCAAGTA	ATACCATAA	GATACCAAC	ATGATACCT	A
13101	CAATTTTAAA	R I	GTTAACTTG	TTTTCATA	A TCTCATTTAC	CGACTTGCA	TACTAAACGO	GTACCAGCAG	ACAATAGAAA	ACCTTCTCA	r
13201	TTGAAATGA	TTC									

Figure 2. (Continued).

Figure 2. Complete sequence of the 13 213 bases of chromosome XI. The sequence reads 5' to 3' from the left telomere to the centromere. EcoRI sites are underlined. ORFs are boxed. The direction of each ORF is shown by arrow.



Figure 3. Diagram of the homologies between different chromosomes. Shaded bars indicate the chomosomes. Numbers at the bottom of each bar indicate the base coordinates as given in the BlastA analysis. Numbers on top of each bar indicate the percentage of base pair identities.

internal oligo-priming to fill in the gaps. An average length of 315 nucleotides was read from each sequencing reaction. Readings up to 400 bases were achieved on 4% polyacrylamide gels. Compressions seen at several specific positions were solved by repeating the sequencing reactions using dITP (5 different instances). Sequence assembly was performed manually according to restriction maps and the sequences obtained from PCR connecting fragments. Sequence alignments of both strands were done using the GCG program. The final sequence contained an additional 61-base EcoRI fragment following the first (5')EcoRI site, which had not been detected at the original gel electrophoretic analysis of the cosmid DNA (Figure 1a).

Sequence analysis

Six phase ORF map analysis of the 13.2 kb fragment revealed five ORFs>100 codons (Figure 1b). Their sizes range from 110 to 711 codons and they constitute 48.2% of the entire sequence (6366 bases). This percentage is a low compared to the average chromosome content in coding sequences and it is probably due to the location of this

fragment near the end region of the chromosome (Oliver et al., 1992).

The complete sequence of the 13 213 bases is given in Figure 2. FastA (Pearson and Lipman, 1988) and BlastA (Altschul et al., 1990) analyses of the sequenced segment revealed extensive homologies to known sequences on different yeast chromosomes (Figure 3). More specifically: a) a region of 1182 bases showed an overall identity of 90.7% (ranging from 75% to 100%) to the right arm of chromosome III; b) 159 bases of that showed 77% identity to a second region of chromosome III; c) 638 bases were found 77% identical to the right arm of chromosome V and d) homology has been detected to the right arm of chromosome II (Becker, personal communication). Obviously, these are due to recombination events between fragments near the ends of the mentioned chromosomes, as has been previously reported (Oliver et al., 1992).

Analysis of the putative ORF products

The putative translation products of the identified ORFs have been compared to protein databases using FastA (Table 1). For better evaluation

13.2 kb segment of chromosome XI

Table 1.	Best optimized FastA sco	ores obtained by the	comparison of	the putative t	ranslation pro	duct of each	ORF
with the	protein databases						

ORF	Homologous or Identical protein	Optimized score	Highest score	Reference
D123	S. cerevisiae Hypothetical protein YCR104W (124aa) 98·3% identity in 120aa	544	554	van der Linden <i>et al.</i> , 1992 EMBL: X59720
	S. cerevisiae SYGP-ORF12 (120aa) 86·2% identity in 123aa	485	554	Mulligan et al., 1993, unpublished EMBL: L10830
A110	S. cerevisiae Hypothetical protein YCR103C (111aa) 79·3% identity in 111aa	496	655	van er Linden <i>et al.</i> , 1992 EMBL: X59720
F705	S. cerevisiae CYP1 (HAP1) regulatory protein (1483aa) 28% identity in 130aa	155	3624	Verdiere, 1988 EMBL: X13793
B473	Chinese hamster (<i>Mev</i>) mevalonate transporter (494aa) 28·1% identity in 153aa	182	2515	Kim <i>et al.</i> , 1992 EMBL: S48888
F711	S. cerevisiae (FRE1) Ferric reductase (686aa) 24.5% identity in 693aa	542	3723	Dancis <i>et al.</i> , 1992 EMBL: M86908

of the significance of each obtained score, we have also included the highest FastA score, obtained by the comparison of each ORF to itself. Optimum scores higher than 200 have been considered as significant. Lower scores due to homologies in restricted areas of the protein sequences indicated conservation of specific domains. Protein patterns (motifs) have been identified by the ProSite program (Bairoch, 1991) of the GCG package. Our findings on each individual ORF are discussed below.

ORF D123 is included in the region which is closely related to the right arms of yeast chromosomes III and V. It is almost identical to YCR104W a hypothetical protein in the HMR 3' region on chromosome III, and very similar to the STGP-ORF12 encoded by a gene contained in a region of 36 772 base pairs of chromosome V between the known genes MAK10, AFG18 on its 5' site and CYC7 on its 3' site (Mulligan *et al.*, unpublished, Mortimer *et al.*, 1989) (Figure 4a). The function of both of these hypothetical proteins remains unknown. As has been already reported for YCR104W (Bork et al., 1992), D123 showed also similarities to the yeast temperature-shock inducible protein TIP1 (Kondo and Inouye, 1991) (27.3% identity in 99 overlapping amino acids, FastA score: 144) and to the yeast serine rich, glucose induced protein SRP1 (Marguet et al., 1988) (26.3% identity in 99 overlapping amino acids, FastA score: 121), the function of which is similarly unknown. All of these proteins, including D123, start with a putative hydrophobic signal sequence, which is followed by a conserved domain of about 90 residues including the stressinduced protein motif (P-W-Y-[ST](2)-R-L). This domain is followed, in SRP1 (total length of 254 amino acids) and TIP1 (total length of 210 amino acids) proteins only, by a repetitive serine and alanine rich region (Figure 4b). According to Kondo and Inouye (1991) and Marguet et al. (1988), there is a family of several genes in different chromosomes which cross-hybridize to both TIP1 and SRP1 sequences but some of these may not be highly expressed genes, since only three distinct transcripts have been detected. We have not

Α

D123 YCR104W SYGP-ORF12	MVKLTSIAAGVAAIAAGVAAAPATTTLSPSDERVNLVELGVYVSDIRAHLAQYYLFQAAH MVKLTSIAAGVAAIAAGIAAAPATTTLSPSDERVNLVELGVYVSDIRAHLAQYYLFQAAH MVKLTSIAAGVAAIAATASATTTLAQSDERVNLVELGVYVSDIRAHLAQYYSFQAAH
D123 YCR104W SYGP-ORF12	PSETYPVEIAEAVFNYGDFTTMLTGIPAEQVTRVITGVPWYSTRLRPAISSALSKDGIYT PTETYPVEIAEAVFNYGDFTTMLTGIPAEQVTRVITGVPWYSTRLRPAISSALSKDGIYT PTETYPIEVAEAVFNYGDFTTMLTGIAPDQVTRMITGVPWYSSRLKPAISSALSKDGIYT *.****.*
D123 YCR104W SYGP-ORF12	- IAN A IPK - IAN *
B	
D123 YCR104W SYGP-ORF12 TIP1 SRP1	MVKLTSIAAGVAAIAAGVAAAPATTTLSPSDERVNLVELGVYVSDIRAHLAQYYLFQAAH MVKLTSIAAGVAAIAAGIAAAPATTTLSPSDERVNLVELGVYVSDIRAHLAQYYLFQAAH MVKLTSIAAGVAAIAATASATTTLAQSDERVNLVELGVYVSDIRAHLAQYYSFQAAH MS-VSKIAFVLSAIASLAVADTSAAETAELQAIIGDINSHLSDYLGLETGN MA-YTKIAL-FAAIAALASAQT-QDQINELNVILNDVKSHLQEYISLASDS
D123 YCR104W SYGP-ORF12 TIP1 SRP1	PSETYPVEIAEAVFNYGDFTTMLTGIPAEQVTRVITGVPWYSTR; RPAI PTETYPVEIAEAVFNYGDFTTMLTGIPAEQVTRVITGVPWYSTR; RPAI PTETYPIEVAEAVFNYGDFTTMLTGIAPDQVTRMITGVPWYSSR; KPAI S-GFQIPSDVISVYQQVMTYTDDAYTTLFSELDFDAITKTIVKI, WYTTR; SSEI SSGFSLSSMPAGVIDIGMALASATDDSYTTLYSEVDFAGVSKMLTMVPWYSSR; EPALKS *
D123 YCR104W SYGP-ORF12 TIP1 SRP1	
D123 YCR104W SYGP-ORF12 TIP1 SRP1	SSALSKDGIYT-IAN SSALSKDGIYTAIPK SSALSKDGIYT-IAN SSALSKDGIYT-IAN SSALSKDGIYT-IAN TSSAAPSSSAAPSSSAAPSSSAESSSKAVSSSVAPTTSSVSTSTVETASNAG SSEAKSSSAAPSSSEAKSSSAAPSSTEAKITSAAPSSTGAKTSAISQITDQQIQATKAVS **
D123 YCR104W SYGP-ORF12	

 TIP1
 QRVNAGAA----SFGAVVAGAAALLL

 SRP1
 EQTENGAAKAFVGMGAGVVAAAAMLL

Figure 4. (a) Multiple alignment of the sequences of the D123 ORF, the YCR104W hypothetical protein and SYGP-ORF12 using the CLUSTAL program. (b) Multiple alignment of the sequences of the D123 ORF, YCR104W, SYGP-ORF12, TIP1 and SRP1 proteins. The stress-induced protein motif is shadowed. Asterisks indicate residue identities and dots indicate conservative substitutions.

A110	MEMLLFLNESYIFHRLRMWSTVLWHSCVFVCVECENANYRVPR-CLIKPF-SVPVTFPFS
YCR103C	MEMLLFLNESYIFHRFRMWSIVLWHSCVFVCAECGNANYRGAG-VPCKTLLRAPVKFPLS
DYC101	MEMLLFLNESYIFHRFRMWSIVLWHSCVFVCAECGNAYYRGAGGCLEKPF-CAPVKFPFS

A110	VKKNIRILDLDPRTEAYCLSPYSVCSKRLPCKKYFYLLNSYNIKRVLGVVYC
YCR103C	VKKNIRILDLDPRSEAYCLSINSVCFKRLPRKKYFHLLNSYNIKRVLGVVYC
DYC101	VKKNIRILDLDPRSEAYCLSHHLVCPKRFPCKATSLLLIPEG

Figure 5. Multiple alignment of the sequences of the A110 ORF, YCR103C and DYC101 hypothetical proteins using the CLUSTAL program.

examined the expression of the gene encoding D123 protein. However, its proximity to the telomere could indicate that it is expressed at flow levels or under specific conditions (Sandell and Zakian, 1992). In fact, we have noticed, by FastA analyses, that the TIP1 and SRP1 proteins are more similar to the SYGP-ORF12 (30.6% identity in 108 overlapping amino acids and 29.3% identity in 99 overlapping amino acids, respectively) than they are to the D123 and YCR104W ORFs.

ORF A110 is also contained in the region of homology between chromosomes XI and III. It is similar to the hypothetical protein YCR103C. It contains several non conservative amino acid substitutions which may imply that the two products serve different functions. A third ORF of 101 codons (DYC101 sequence communicated by H. Domdey), that was identified on chromosome II, is also homologous to A110, exhibiting 76.3% identity in 97 overlapping amino acids (FastA score: 442). That ORF resembles YCR103C as much as the A110 does (79.4% identity in 97 overlapping amino acids). A multiple alignment of the three sequences revealed more residue substitutions between A110 and the other two proteins, except for the last 14 amino acids at the carboxy terminus which are identical in A110 and YCR103C ORFs and absent from the chromosome II ORF (Figure 5).

The FastA alignment of the F705 ORF product showed homology to the yeast protein CYP1 (HAP1) only in 130 overlapping residues of its amino terminus (residues 7 to 136 in F705 and 49 to 170 in CYP1). In fact, the alignment varied depending on the program used, except for a stretch of 36 residues, 30 of which make up the fungal Zn (2)-Cys (6) binuclear cluster domain. F705 was also regionally similar to several proteins that contain the same motif (FastA scores: 100–155). All of these proteins are transcriptional activators (GAL4, MAL63, LEU3, etc.) that bind DNA with their cysteine-rich amino terminus in a zinc dependent fashion (Coleman, 1992). The identified motif in F705 starts at residue (SCHFCRVRKLKCDRVRPFCGSCSSRN-23 RKOC) and follows the consensus sequence: [GAS]-C-x(2)-C-[RKH]-x(2)-[RK]-x-[RK]-C-x(5,9)-C-x(2)-C-x(6,8)-C F705 contains also, near its carboxy terminus, a second rare motif characterizing membrane proteins involved in sugar transport, starting at residue 629 (MEKIGRRAFNKG) and following the consensus sequence: [LIVMST]-[DE]-x-[LIVMFA]-G-R-[RK]-x(4,6)-G. However, its significance is questionable since the hydrophobicity profile of F705 protein does not reveal the typical transmembrane domains (data not shown), unless it is a novel type of protein that can mediate both interaction with a sugar and transcriptional regulation.

The FastA search for ORF B473 product revealed low similarities to a number of membrane proteins. The most significant similarity was between its amino terminus and a mammalian membrane protein, the putative transporter of mevalonate (Figure 6a). This similarity does not necessarily indicate a directly homologous molecule but some sort of membrane transporter. In fact, the hydrophobicity profile of ORF B473 showed 12 membrane spanning stretches typical of such proteins (Figure 6b). The profile also included one central and one carboxy terminal hydrophilic region which followed the transmembrane segments 6 and 12 respectively, similarly to the profile of MEV protein and other membrane transporters (Culham et al., 1993). The main difference was at the amino terminus (~ 30 residues) or ORF B473 which appeared hydrophilic. A transcript corresponding to the ORF B473 was not detected in extracts of cells grown in standard YPD (Guthrie and Fink, 1991) growth medium (data not shown).

B473 Mev	MSEERHEDHHRDVENKLNLNGKDDING MPPAIGG * *.*	NTSISIEVPDGGYGWFILL-AFILYNFSTWGAN PVGYTPPDGGWGWAVVVGAFISIGFS-YAFP ***** *** .**
B473 Mev	SGYAIYLAHYLENNTFAGGSKLDYASIC KSITVFFKEIEGIFNATTSEVSWIS	GGLAFSCGLFFAPVITWLYHIFSIQFIIGLGIL SSIMLAVMYAGGPISSVLVNKYGSRPVMIAGGC
B473 Mev	FQGAALLLAAFSVTLWEIYLTQGVLIG LSGCGLIAASFCNTVQELYLCIGVIGG . * .*. *. *. *. *. *. *. *	FGLAFIFIPSVTLIPLWFRNKRSLASGIGTAGS LGLAFNLNPALTMIGKYFYKKRPLANGLAMAGS .**** . **.* .* .**.**.* ***



Figure 6. (a) Alignment of the 180 aminoterminal residues of ORF B473 with the 156 aminoterminal residues of the mammalian membrane transporter MEV. (b) Hydrophobicity profiles (Kyte and Doolittle, 1982) of ORF B473 and MEV protein.

This may indicate gene expression under specific conditions or gene repression related to its position near the telomere.

The product of ORF F711 showed a significant similarity to the known yeast FRE1 ferric reductase. We have proven by biochemical, genetic and structural analyses that it is also a membrane protein that can reduce the environmental ferric iron to its intracellular ferrous form. The expression of this non-essential gene is down-regulated by the presence of iron in the growth medium and its RNA is not detectable under standard YPD medium growth conditions (Georgatsou and Alexandraki, submitted for publication).

ACKNOWLEDGEMENTS

We thank Bernard Dujon for the excellent coordination of the Chromosome XI sequencing project. We thank Irmi Becker and all MIPS staff for help with the sequence analysis. We thank Horst Domdey for allowing us to use the unpublished

Δ

13-2 kb SEGMENT OF CHROMOSOME XI

DYC101 hypothetical protein from chromosome II. We also thank Yannis Papanikolaou for help with the computer analyses at IMBB, Georgia Houlaki for help with the preparation of the figures and the co-contractor George Thireos for helpful discussions. This work was supported by the Commission of the European Communities under the BRIDGE program of the Division of Biotechnology and by the Greek Ministry of Industry, Energy and Technology.

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403-410.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (Eds) (1987). Current Protocols in Molecular Biology. Green Publishing Associates and Wiley-Interscience.
- Bairoch, A. (1991). A dictionary of sites and patterns in proteins. Nucl. Acid Res. 16, 2241–2245.
- Bork, P., Ouzounis, C., Sander, C., Scharf, M., Schneider, R. and Sonnhammer, E. (1992). Comprehensive sequence analysis of the 182 predicted open reading frames of yeast chromosome III. *Protein Sci.* 1, 1677–1690.
- Coleman, J. E. (1990). Zinc proteins: enzymes storage proteins, transcription factors and replication proteins. Annu. Rev. Biochem. 61, 897–946.
- Culham, D. E., Lasby, B., Marangoni, A. G., Milner, J. L., Steer, B. A., van Nues, R. W. and Wood, J. M. (1993). Isolation and sequencing of *Escherichia coli* gene *proP* reveals unusual structural features of the osmoregulatory proline/betaine transporter, ProP. J. Mol. Biol. 229, 268–276.
- Dancis, A., Roman, D. J., Anderson, G. J., Hinnenbush, A. G. and Klausner, R. D. (1992). Ferric reductase of Saccharomyces cerevisiae: Molecular characterization, role in iron uptake and transcriptional control by iron. *Proc. Natl. Acad. Sci. U.S.A.* 89, 3869–3873.
- Georgatsou, E. and Alexandraki, D. Two distinctly regulated genes are required for ferric reduction, the first step of iron uptake in *Saccharomyces cerevisiae*. (Submitted for publication)

- Guthrie, C. and Fink, G. R. (Eds) (1991). Guide to yeast genetics and molecular biology. *Methods in Enzymol*ogy, vol. **194.** Academic Press, New York.
- Higgins, D. G. and Sharp, P. M. (1988). Clustal: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237–244.
- Kim, C. M., Goldstein, J. L. and Brown, M. S. (1992). cDNA cloning of MEV, a mutant protein that facilitates cellular uptake of mevalonate and identification of the point mutation responsible for its gain of function. J. Biol. Chem. 267, 23113-23121.
- Kondo, K. and Inouye, M. (1991). TIP1, a cold shock inducible gene of Saccharomyces cerevisiae. J. Biol. Chem. 266, 17537-17544.
- Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. 157, 105-132.
- Marck, C. (1988). 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computer. *Nucl. Acid Res.* 16, 829–1836.
- Marguet, D., Guo, X. J. and Lauquin, G. J.-M. (1988). Yeast gene SRP1 (serine rich protein)-Intragenic repeat structure and identification of a family of SRP1 related DNA sequences. J. Mol. Biol. 202, 455–470.
- Mortimer, R. K., Schild, D., Contopoulou, C. R. and Kans, J. A. (1989). Genetic map of Saccharomyces cerevisiae, Edition 10. Yeast 5, 321-403.
- Oliver, S. G. et al. (1992). The complete DNA sequence of yeast chromosome III. Nature 357, 38-46.
- Pearson, V. R. and Lipman, D. J. (1988). Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. U.S.A. 85, 2444–2448.
- Sandell, L. L. and Zakian, V. A. (1992). Telomeric position effect in yeast. *Trends Cell Biol.* 2, 10–14.
- Tzermia, M., Horaitis, O. and Alexandraki, D. (1994). The complete sequencing of a 24.6 kb segment of yeast chromosome XI identified the known Loci URAI, SACI and TRP3, and revealed 6 new open reading frames including homologues to the threonine dehydratases, membrane transporters, hydantoinases and the phospholipase A₂-activating protein. Yeast 10.