Zebrafish dystrophin and utrophin genes: Dissecting transcriptional expression during embryonic development

DANIEL LAI¹, CHUAN-CHING LAN¹, IVONE UN SAN LEONG¹ and DONALD R. LOVE^{1,2}

¹School of Biological Sciences, University of Auckland, Auckland 1142; ²Diagnostic Genetics, LabPlus, Auckland City Hospital, Auckland 1148, New Zealand

Received October 21, 2011; Accepted November 23, 2011

DOI: 10.3892/ijmm.2011.865

Abstract. Some genes can encode multiple overlapping transcripts, and this can result in challenges in identifying transcript-specific developmental expression profiles where tools such as RNA in situ hybrisations are inapplicable. Given this difficulty, we have undertaken a preliminary analysis of the developmental expression profile of selected transcripts of the dystrophin and utrophin genes of the zebrafish (Danio rerio) by targeting unique and common regions of each of these transcripts. The dystrophin and utrophin genes of zebrafish were identified by bioinformatic analysis and the dystrophin gene predictions were confirmed by transcript sequencing. These data enabled primer pairs to be designed in order to determine the expression profiles of unique, but overlapping transcripts, throughout embryonic development using quantitative real time reverse transcription PCR (qRT-PCR). The data indicated the early expression of the short carboxyl-terminal dystrophin transcript, with expression of the full length muscle transcript occurring during myogenesis. Importantly, a composite of these two profiles appeared to comprise the major transcriptional load of the zebrafish dystrophin gene. In contrast, utrophin gene expression was dominated by the full length transcript throughout embryogenesis. The approach described here provided a means by which a gene's transcriptional complexity can be deconvoluted to reveal transcriptional diversity during embryogenesis. This approach, however, required the identification of unique regions for transcript-specific targeting, and an appreciation of alternative splicing events that may compromise the design of primers for qRT-PCR.

Introduction

The human dystrophin gene spans 2.5 million base pairs and is the largest gene in the human genome (1). The dystrophin

E-mail: donaldl@adhb.govt.nz

gene contains at least 7 known promoters that drive the expression of multiple transcripts, together with multiple splice variants (Fig. 1). The largest and most studied isoform of dystrophin, designated Dp427m, is expressed as a 427 kDa protein in human muscle. This protein comprises 3685 amino acids and plays a pivotal role in maintaining the integrity of the dystrophin:glycoprotein complex that connects cytoskeletal actin to the extracellular matrix of muscle cells. The less abundant isoforms of dystrophin are localised to other tissues such as the brain (Dp427b/c and Dp427p), the retina (Dp260), brain and kidney (Dp140), and Schwann cell/peripheral nerve (Dp116); the smallest isoform, Dp71, is ubiquitously expressed (2,3). The dystrophin gene has an equally large paralogue, termed utrophin (4,5), which expresses multiple transcripts (2,3,6-18) (Fig. 1). The full length utrophin isoform is able to fulfil the role of dystrophin when overexpressed in the absence of dystrophin (19).

Human dystrophin orthologues have been identified in many animal and insect species to date, ranging from mammals such as dog (20), cat (21) and mouse (22) to fish such as pufferfish (23) and zebrafish (24-26), together with frog (27), sea urchins (28) and fruit flies (29). Utrophin has been confirmed in many of the species in which dystrophin has been identified (4,30,31).

Importantly, Duchenne and Becker muscular dystrophies (D/ BMD) are caused by mutations in the dystrophin gene [reviewed by Manzur *et al* (32)] and several models of D/BMD have been constructed using mice (33) and dogs (20), among others. In addition, there is increasing interest in using the zebrafish to model muscular dystrophies such as DMD (26,34-36).

Given the transcriptional complexity of the dystrophin gene and its role in diseases, mutations generally impact not only on Dp427m but also other dystrophin isoforms. In the absence of dystrophin expression, upregulation of full-length utrophin appears to occur (19). Against this background, little is known of the complexity of the developmental expression profile of dystrophin, as well as the utrophin, gene transcripts. In order to address this deficiency, we turned our attention to the zebrafish (Danio rerio). Zebrafish offers advantages over other model organisms in that they are optically transparent during embryonic development and ex utero development allows for easy manipulations (34,36-38). The main qualities of using zebrafish are their rapid generation cycles, low maintenance requirements and cost effectiveness at high densities.

Correspondence to: Dr Donald R. Love, Diagnostic Genetics, LabPlus, Auckland City Hospital, P.O. Box 110031, Auckland 1148, New Zealand

Key words: zebrafish, dystrophin, utrophin, quantitative real-time PCR



Figure 1. Human dystrophin gene exons and promoters. (A) The middle section shows the human dystrophin gene, the relative location of all exons, and the principal transcripts that are expressed by the gene. The exons involved in alternative splicing are shown in the expanded views. The unique first exons for the various dystrophin isoforms are shown, together with their relative positions in the gene. (B) The start sites of the various isoforms of the human utrophin gene are shown. In the case of Up116 and Up109, they have only been isolated from mouse to date (18).

The aim of our study was to provide for the first time a developmental expression profile of dystrophin and utrophin gene transcripts. Expression profiling posed challenges due to the transcriptional complexity of both genes: nucleotide sequences of each gene's transcripts are largely identical (albeit varying in length), and any differences are largely confined to short unique 5'-end sequences. These challenges highlighted the difficulty in applying a conventional RNA in situ hybridisation approach, which also suffers from sensitivity issues, therefore we targeted multiple exons in a quantitative real-time PCR (qRT-PCR) strategy.

Materials and methods

Primers. Primers were designed using the free online primer design software Primer3 (http://frodo.wi.mit.edu/cgi-bin/ primer3/primer3_www.cgi). Advanced settings were used where appropriate to minimise the formation of secondary structures while reducing non-specific products being amplified by the primers. All primers designed were subjected to manual quality analysis using NetPrimer (http://www.premierbiosoft.com/netprimer/). qRT-PCR primers (Table I) were designed to produce product sizes between 100-160 base pairs with an approximate T_m of 60°C. All primers were designed so that the product would encompass at least 2 exons, with preference given to primers that annealed to exon/exon boundaries, providing that primer quality was not compromised.

Primers were ordered from Invitrogen Corporation (http:// www.invitrogen.com) with all primers used for qRT-PCR subjected to quality checks. These checks involved the confirmation of only one PCR amplification product following electrophoresis on a 1.5% agarose 1X Tris/Borate/EDTA (TBE) gel stained with ethidium bromide. All PCR products and excised gel bands were subjected to spin-column purification using a GE Healthcare Illustra GFX PCR DNA and Gel Band Purification kit according to the manufacturer's recommended instructions (GE Healthcare Life Sciences).

Sequencing. Sequencing reactions were carried out by the Genomics Unit, School of Biological Sciences, University of Auckland. The ABI PRISM[™] BigDye Terminator sequencing kit version 3.1 was used to incorporate fluorescent dyes as recommended by the manufacturer. Thermal cycling was carried out using an Applied Biosystems GeneAmp[®] PCR System 9700. Unincorporated fluorescent dyes were removed using Agencourt[®] CleanSEQ[®] magnetic beads. Capillary electrophoresis of the sequencing products was carried using the ABI PRISM[™] 3130xl Genetic Analyzer.

Zebrafish husbandry. Wild-type zebrafish (Hollywood Fish Farms, Auckland, New Zealand) were grown and maintained at densities of 3-8 fish per 2.75 liters tank, in a water recirculation rack system (Aquatic Habitats) located inside a dedicated zebrafish facility. The zebrafish were exposed to 14 h of light per day with temperatures kept between 26-28°C. Water quality was monitored on a daily basis. Adult fish were fed a range of dry fish food supplemented with live Artemia, and juveniles were fed a mix of dry fish food and live Rotifer until at least two weeks of age. Adult male and female fish were

Primer designations	Target Gene	Target exons	Primer sequence
ZFdysX01F5 ZFdysX02R5	Dystrophin	1-2	GAAGCAGTGAACCCACAGGAC TGAGTTTATCCATTTGGTGAAGG
ZFdysX02F1 ZFdysX03R1	Dystrophin	2-3	CCTTCACCAAATGGATAAACTCAC GCCAACCAGAACCTCCAGAAG
ZFdysX75F5 ZFdysX76R5	Dystrophin	75-76	CGGCAACACAAGGGACGA GAGGCAGTGGAGGGTGAGGA
ZFdysDp71X01F2 ZFdysX63R2	Dystrophin	Dp71-63	GCTTTTTCCACTGTAGAGAGGG TGTTGTTTGGGTCTGGTGATTT
ZFutrX01F1 ZFutrX02R3	Utrophin	1-2	CGTGGGTGTGATGTCAGAG CTTGGAGAACTGAGCGTTGA
ZFutrX60F1 ZFutrX70R1	Utrophin	69-70	AGATTCTGCAGGCAGTGGA AGGCTGACCGTACTGAAACC
EF1aF EF1aR	Elongation factor 1α		CTGGAGGCCAGCTCAAACAT ATCAAGAAGAGTAGTACCGCTAGCATTAC
Rpl13aF Rpl13aR	Ribosomal protein L13a		TCTGGAGGACTGTAAGAGGTATGC AGACGCACAATCTTGAGAGCAG

Table I. Primers used for qRT-PCR.

kept mixed until one week prior to spawning at which point they were separated by gender. Embryos were collected after spawning and thoroughly rinsed. The embryos were visually inspected to ensure that they were at the same developmental stage and then kept immersed in 0.5X E3 media (1X E3: 5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl₂ and 0.33 mM MgSO₄) in 15 mm plastic Petri dishes at a density of 20/dish and kept at a constant 28°C. Embryo E3 media was changed every 24 h.

Developmental time course. The zebrafish embryonic developmental stages described in this paper are reported as hours post fertilisation (hpf) at 28°C: 3.5 hpf (high stage), 10 hpf (bud stage), 11 hpf (3-somite stage), 14 hpf (10-somite stage), 16 hpf (14-somite stage), 19.5 hpf (21-somite stage), 22 hpf (26-somite stage), 25 hpf (prim-6 stage), 31 hpf (prim-16 stage), 42 hpf (high pec stage), 60 hpf (pec fin stage) and 72 hpf (protruding mouth stage).

RNA extraction and purification. At each appropriate developmental stage, 20 embryos were inspected to ensure the correct developmental stage was reached, then they were pooled and snap frozen. RNA extraction and purification was carried out according to the protocol described earlier (39).

Reverse transcription. One microgram total-RNA from each extraction was reverse-transcribed using SuperScript III reverse transcriptase (Invitrogen Corporation) primed with random hexamers (Invitrogen Corporation) in a 20 μ l reaction volume according to the manufacturer's instructions. The final SuperScript III inactivation step (70°C for 15 min) was omitted and cDNAs were diluted 1:16 before use in qRT-PCR assays.

Quantitative real-time reverse transcription polymerase chain reaction. All qRT-PCR assays were carried out using

an Applied Biosystems 7900HT platform running SDS v2.3 software (Applied Biosystems). Reactions were set up using an Eppendorf epMotion 5075 (Eppendorf) automated liquid handling platform to minimise pipetting errors. Ten microliter reaction volumes comprised the following: 5 μ l Platinum SYBR-Green qPCR SuperMix-UDG with ROX (Invitrogen Corporation), 1 μ l of 2 μ M forward primer, 1 μ l of 2 μ M reverse primer and 2.6 μ l of template; the template was either 16X diluted cDNA, or water in the case of negative controls. The expression of two zebrafish reference genes were used as normalisers, *Rpl13a* and *EF1a* (40), with all reactions performed in triplicate. Cycling programs consisted of 40 cycles of 15 sec at 94°C and 60 sec at 60°C. Dissociation curve analysis was performed for each qRT-PCR run.

qRT-PCR data analysis. The data obtained from the SDS v2.3 software were inspected manually to omit any well that showed abnormal amplification curves, followed by manual baseline and threshold setting adjustments as recommended by the SDS manual. Amplification efficiencies were determined using LinRegPCR v11.1 (41,42). Baseline and threshold corrected Ct values, together with amplification efficiencies, were processed as described in the geNorm v3.5 manual (version: July 8, 2008) (43). This processing provided relative expression levels (arbitrary quantities) while allowing for normalisation against the two reference genes.

Results

Dystrophin and utrophin sequences. At the start of this study, the available sequences for zebrafish dystrophin were fragmented and incomplete. These limited data were used to design primers in order to amplify overlapping fragments of the full length zebrafish dystrophin muscle transcript, which

1	GGGGATTCAC	AAAAAATTCA	CTGCTGGATC	TGAGGTTCAG	TTTGTTCACA	50
51	GGGCGTTGGT	GAGGGCTCCT	TCCTTTTTTCT	GTTTATAAAA	TTGTAGGATT	100
101	AAAGGACCCA	ACGTTGGCTT	GGATCTTATG	TCATGGCCGA	AGCAGTGAAC	150
151	CCACAGGACC	AATGGGAGGA	AGGTTTTGAG	GACGAATTCG	GAGAAATCAT	200
201	CAAAACTCGA	TCAGATGAAA	GAGAAGACGT	TCAAAAGAAA	ACCTTCACCA	250
251	AATCCATAAA	CTCACACTTT	CCTAACACAA	CAACACCTCC	TATTCATCAC	300
201	CTCTTCACTC	ACCTOTOTO	TCCCCCACCT	GAAGACCICC	TATIGATGAC	300
301	CICITCACIG	ACCIGIGIGA	IGGCCGACGI	CITCIIGAAC	TICIGGAGGI	350
351	TCTGGTTGGC	CATGAAAT	TTAAAGAACG	TGGCTTCACT	CGAGTGCACT	400
401	CCCTTAACAA	TGTTAACAGG	GCCCTGCAGA	TCCTTCAGAA	GAACAATGTT	450
451	GATCTGGTGA	ACATTGGAGG	AGCTGACATT	GTAGATGGGA	ATCATAAACT	500
501	GACCCTGGGG	CTCATCTGGA	GCATCATTCT	CCACTGGCAC	TTAAGGATG	550
551	TGATGAAAGA	TGTCATGGCG	GACTTACAGC	AGACCAACAG	CGAAAAGATC	600
601	CTGTTAAGCT	GGGTCAGGCA	GTCCCTCAAA	AACTACCAAG	ACGTCAACGT	650
651	GGTCAACTTC	TCCAGCAGCT	GGGCCGACGG	TTTCGCTTTC	AACGCTCTCA	700
701	TCCACACCCA	CALOCCCCAC	CTGTTCAGCT	CCACTCTACT	CCACCAACAG	750
701	CATAATCCCA	TTCACACACT	CONTONTOCO	TTOCOTOTTO	GGAGCAACAG	000
751	GATAAIGUCA	TIGAGAGACI	GGATCATGCC	TICGGIGIIG	CGGAGAAGAG	000
801	TITAGGAATC	GACCGGCTGC	TGGACCCCGA	GENTGITGCT	ACAGICCATC	850
851	CAGATAAGAA	ATCCATCATC	ATGTACGTTA	CCTCCCTCTT	CAAAGTGCTG	900
901	CCCCATGGCG	TGAGTTTGGA	GGCCATTCAA	GAGGTGGAGA	CCCTCCCTCG	950
951	AGCCGCAGTG	ACCAAGGAGG	AGCACGTCCT	CTATCAAACC	CAACAGCGGT	1000
1001	ACTCTCAGCA	TCACAGTC	AGCGTGGCCC	AGAGTCGCGT	TCGCAGCCCG	1050
1051	TCGCCCTCCT	ATAAGCCGAG	ATTTAAAAGC	TATGCCTTCA	CTCAAGCCGC	1100
1101	CTACGTCAAG	ACACCTGAAC	AGCAGAGGAA	GTTCCTCATC	GCACACICTC	1150
1151	CAGACAAGGC	CGATGAACTG	CGACCGTCCC	CCAGTCCCCT	GCCGCAGGGG	1200
1201	CTABATCACC	TCCABACCTA	CCACACTCCC	CTCCACCACC	TOTTCACCTC	1250
1201	CIMAAIGAGC	IGGAAAGCIA	CCAGAGIGCC	CIGGAGGAGG	TCTTGACCTG	1200
1251	GCIGCICICI	GCGGAGGACG	GCCIGCAAGC	ACAGUCACCC	AICICIT	1300
1301	TTGTAGAAGA	AGTCAAAGAG	CAGTTTCACA	CTCATGA	CTACATGGTG	1350
1351	GAGCTGACGT	CTCACCAGGG	CAGCGTGGGC	CGGGTCCTGA	AAGCGGGAAG	1400
1401	TGTGTTGTTG	TCTGGAGGGC	AGCTGACTGA	TGATGAAGAG	AGGGAGGTGC	1450
1451	GGGAACAGAT	GAATCTGCTC	AACTCGCGCT	GGGAACATCT	GCGTGTGGCC	1500
1501	AGTATGGAGC	GACAGAGCAG	GCTTCATGAA	GTTTTGATGG	ACCTTCAGCA	1550
1551	CCAGCAGCTG	AAGCAGCTCT	CTGATTGGCT	GGACACGACG	GAGACTCGTA	1600
1601	TTABAACCAT	GGGAGCGCAG	GCTTTGGGGGC	CTGAACTTGA	TGACATAAAG	1650
1651	CCCCANATCC	AACACCACAA	NTCTTACAC	CARCATCTCC	ACCTCCAACA	1700
1701	CGGCAAAICG	AAGAGCAGAA	OTOTIACAG	GAAGAICIGG	AGCIGGAACA	1750
1701	AGIGCGAGIG	AATICACIGA	CICAIAIGGI	GGIGGIIGIA	GAIGAGAAIA	1/50
1751	GCGGGGGATGG	CGCCACCGCA	GCCTTGGAGG	AAAAGCTGCA	GAACCTTGGG	1800
1801	GAACGATGGG	CAGCGATCTG	TAAATGGACT	GAAGAGCGTT	GGATCCTCCT	1850
1851	TCAAAAGATC	CTGCTGTGCT	GGCAGCATTT	CTCTGAGGAG	CASCTGCTGT	1900
1901	TTGATTCTTG	GCTGACTCAA	AAGGAGGAGT	TGGTCCAGTC	CATCCAAAGC	1950
1951	AGTGGCACAA	ATGACCCAAA	TGAAGTAGCC	GCCAACCTGC	GGAAGCTGGC	2000
2001	TTTTTAAAA	GCAGACCTGG	AGCTGAAGAG	GCAGACCATG	GACAAGCTTT	2050
2051	GCTCGCTCGT	CCAGGATCTG	CTGACTAACA	TCAAGAGCAA	AGAGGCCGCT	2100
2101	GGGAAGCTGG	AAGCAAAGCT	GGAGAGGTTT	GCTCAGCGCT	GGGACAAATT	2150
2151	CCTCCACCCC	CTTCACCTCA	CCACCACCAA	COTTICANCT	ATTCTCACCA	2200
2201	CATCCCACTC	CONCATCACC	CACACAACCAA	TCCCAACTCT	CACCAACCTC	2200
2201	CATCCCAGIC	GGAGAICACG	CACACAACCA	IGGCAACIGI	CACCAAGGIG	2250
2251	ACCACGAACC	AGAAGAAGAT	GGTGAAGCAT	ACTAAGGAGG	GCATGTCCAC	2300
2301	CCCTCCACCT	CAGAAAAAGA	GACAGATTGT	CGTGGATTCA	GAACTGCGGA	2350
2351	AAA	TGTGGACTTC	ACTGAAATTC	ACAGTTTTAT	GACTCGGTCC	2400
2401	GAGGCTGTTT	TGCAAAATCC	CGAGTTCTCA	ATCTCCCGTA	AAGAAGGCAG	2450
2451	TGTGGCAGAT	CTCTATGAAA	AAGTGCTCC	AATTGACAGA	GAAAAACCCG	2500
2501	AAAAGTTCAG	GAAGTTGCAA	GAAGCCACTC	GCTCAGCCCA	GGCTTTGGTG	2550
2551	GATCAACTTA	CCAGC	TCAGAACTCA	GAGGACATCC	AGCAAGCGGC	2600
2601	TCAGCAGTTG	CGTGCTCGAT	GGGTGGACTT	CTGCGCTCTC	CTGGCCGAGC	2650
2651	CCCTCCCCTC	CTTCCCCTAC	CACACCAAAC	TCOTCCOTTT	CTACAATCTC	2700
2001	GGCIGGCGIG	TCCLCCLAC	TOCOCOARAG	IGCIGGCIII	CIACAAICIC	2700
2701	TITCAGCAGC	IGGAGCAGGC	IGCGGCCACG	GCGGAGAACI	GGCIGAAGGI	2/50
2751	GCAGTCGCCA	CCCGCATGTG	AACCCGAACC	TCTGAGGATC	CAGCIGGAGC	2800
2801	GCTGCAG	TGAAATCTCT	CGCCTCTCGG	CTCTTCAACC	TCAAGTGGCC	2850
2851	AAACTGCACG	AACAACTCGA	GGAGCTGCGG	CAGAAAGAAG	AGACGCCGGT	2900
2901	TCTGTTTGAC	GCGGACATTT	CTGCCTTCCA	AGAGCACTAC	CACCAAGTCC	2950
2951	TGGAGGACCT	GAGGGCCAGG	GAACGACAGC	TCGTTTTG	CCAGTCTAGT	3000
3001	CTTCCTCCAG	CACGTTATAA	GGATGTGATG	GCAGCGCTGT	TGGCCTGGCT	3050
3051	ACAGCAGTGT	GAGAATAAAC	TAGCCATCCC	TTCTACAGCA	GTGACTGAGT	3100
3101	ATCCCGTCAT	GGAGCAGAGA	CTCAAAGATA	TCAACCAAT	CCAAGCATCC	3150
3151	CAGAAGGAGC	ATCAGGGTGA	CGTGGATGAC	CTCAATAAGA	TGGCTGAGCA	3200
3201	GGTTTTTCCAG	AAAGCTCCAC	CTGAGATCTG	TCAGAAGTAC	CGGACAGAGC	3250
2251	TCCATAACCT	CATCOTOCOC	TCCACACCCA	TATCCCACCA	COTCOLOGIC	2200
2201	IGGATAACGI	GAIGGIGCGC	TGGAGACGCA	TATCGGAGCA	GCIGGAGGAG	3300
3301	MACATCCAGA	MACIGCAGGA	TCACATGACC	AAGCTACAGC	AGIICCAGAA	3350
3351	TGACACAAAG	ACTCTGCAGA	AGTGGATGGC	AGAGGTGGAT	GITITCCTGA	3400
3401	ATGAGGAATG	GCCAGCTCTG	GGAGACGCAG	AGGCTTTGGA	GAAACAGTTG	3450
3451	GAACAGTGCA	CCCCTCTGGT	GAATGACATC	CACACCGTCC	AACCCAGTCT	3500
3501	GAACGGCATC	AATGAGGTGG	GACTGGCCTT	GAAGAGGGAA	GCTGAGACGC	3550
3551	CGTTTGCCAT	CAAAATACAG	AAGATGCTTG	ATGAGCTCAA	TGCGCAGTGG	3600
3601	GAGCTCATTT	GCAAACA	TTATGCAAAG	AAATCAGCCC	TGAAAGGAGG	3650
3651	GCTGGACATG	ACGGTGAGTT	TGAGGAAGGA	GATGCAGGAG	ATGCAGGAGT	3700
3701	GGATCACTCA	GGCCGAGGAA	GAGTACOTOC	ACACACACTT	CCACTACAAA	3750
3701	BGATCACTCA	ACCURACIAA	GAGIACCIGG	AGAGAGACIT	CCAGIACAAA	3/50
3731	ACACCIGAGG	AGCITCATAA	AGCAGTGGAG	GAGCIGAAGO	GGGCTCAAGA	20000
3801	GGAAGTCCAT	CAAAAGGAAA	CAAAGGTGAA	GCTGCTGACG	GATAAAGTCA	3850
3851	CCAATTTCAT	TTCAAAGGCT	CCTCCCGCGG	CCCATGATGC	CCTGAAGGCA	3900
3901	GAACTGGATG	TTTTGACTTC	TAACTACCAG	CGCCTGTGCA	GCCGACTCGA	3950
3951	TGGAAAATGC	AAGACTTTAG	ACCAAGTGTG	GGCATGCTGG	TGTGAGCTGC	4000
4001	TGTCTTATTT	GGAGTTGGAG	AATGCCTGGA	TGGACCTACT	TGAAAAGAAA	4050
4051	CTTGATGAAA	CGGAAGGGCT	TCAAGGAGGC	ATAGAGGAAA	TTGAAGAAGC	4100
4101	TTTGACTECT	TTAGACACCA	TGATTAGAGA	GCATCCTGAA	TACAACCGTA	4150
4151	ACCAGATOCO	TGAATTACCC	CAGACCCTCA	TGGATGCCCC	AGTOCTOCAT	4200
1101	00010001000	10nniin000	MULJUUNDA	10001000000	110100100A1	1200

Figure 2. Nucleotide sequence and predicted amino acid sequence of zebrafish Dp427m.

1.1111	CACCTCATCC	ACAAAACCT	CONCONCERC	AACACACCCT	CCCATCAACT	1250
4201	GAGCICATCC	ACAAAAAGGI	GGAGGACIAC	AACACACGCI	GGGATGAACT	4250
4251	GATGCAAAG	CTTCACAAA	GGCGCCAGCA	GCTGGAGAAG	AGCTTGCATT	4300
4301	GGGCCCAGGA	GAACGACAAA	ACCCTGCGTC	TCATTCAGGA	CTCTCTGAAC	4350
4351	ACCACTGACC	GACATCTGAC	CGCTTACATA	GCAGATGGTA	TAGACGCTGC	4400
4401	ACACATACCA	CACCAACCAC	ACTAATTCA	CACTCACTTC	AACCCCATC	1150
4401	ACAGATACCA	CAGGAAGCAC	AGAAAATICA	GACIGAGIIG	AACGGCCAIG	4450
4451	AGGTGACACT	GGATGACATG	AAGAAAAAGG	CTATGGAGGT	TGACGCCTCA	4500
4501	GAGAAAGTGA	TTGGAGAGAT	CGATGCAACA	TTAGATAAAC	TTTTGCAAGT	4550
4551	AAAGGGCAAG	TTCCGGCTTT	TCCAAAAGCC	AGCGAACTTT	GACCAGCGGC	4600
4601	TCACCCACTC	TCACCOACTC	CTCCACCACC	TCADACTCAA	COTOCOLOTO	4650
4001	IGAGGGAGIG	IGAGCGAGIG	CIGGAGGAGG	IGAAAGIGAA	GCIGGGAGIG	40.00
4651	CTGAGTATTC	GCAGTGTGGA	GCAGGAAGTG	GTGCAGTCAC	AGCTTGAGCA	4700
4701	GTGCATGAAA	TTTTTATAAGA	ACCTGAGTGA	GGTGAAGTCT	GAGGTGGAGA	4750
4751	CGGTGATAAA	GACGGGGAGG	CAGATTGTGC	AGAAGCAGCA	GACGGAGCAA	4800
4901	CCCAACCACT	TACACCACAC	COTCACCCCC	CTCAAACTCC	CTTACAATCA	1950
4001	CCCAAGGAGI	TAGACGACAG	GCIGACGGCC	CICAAACICG	CITACAAIGA	4000
4851	TITGGGTTCA	CAGGTAACAG	AAGGAAAGCA	GGAACIGGAG	AAGATTTTGA	4900
4901	AGTTACTGAG	GAAGTTTAGG	AAGGAGGTGA	ACAGTCTAAC	CGAATGGCTG	4950
4951	GCTACTACTG	ATGGGGAACT	GACGCGCCGC	TCATCGGTGG	AGGGTATGCC	5000
5001	AAGTGATTTG	GATGCGGAGT	TECCTTECCC	ANACCAGCG	CAGGAGACA	5050
5001	ANGIGATITO	TALOCOGRAT	1000110000	TANCOCACT	CAGGAGGAGA	5100
5051	CIGAGCAICA	TAAGCCACAG	TIGAAACIGG	IGAGGGAGII	GGCAGGGACC	5100
5101	CTGAAGGGTT	TACTTCATAG	CCAAGAGAAC	CTGATTGATG	ACAAGGTCAG	5150
5151	TTTGCTGAAC	TGCAACTGGA	TCGCAGTGAC	GTCACGGAGC	GAACAGTGGC	5200
5201	TCABACTCCT	CCT CATTAT	CAAAATCAAA	TGAAAACACT	TGAGCAGAAC	5250
5201	ATTACACA	TCAACACCTC	CATCONTOCO	CORCARCACI	1 GAUCADAACA	5200
5251	ATTICICAGA	ICAACACCIG	GAIGGAICGC	GCAGAAGAGA	AGCIGGAIGA	5300
5301	GATGGACAGC	CAGGGCTGCA	TTGAGCACGT	TATAAAGGT	TTGCGAATGG	5350
5351	AGTTGGAGGA	GATGAAGGGG	AAAGTTGAAG	CAGTGCAAGT	TCTAGCAGAG	5400
5401	GATCTTATAA	AGAACACAGG	GGAACATTGT	AAAGCTCAAG	TGAAACCAAA	5450
5 AE 1	COTTONACAO	CTTDACCARC	CTTTTCACAT	TOTTOCALCA	ACCATTTA	5500
5451	GCTTGAACAG	CTTAACCAAC	GITTIGACAT	IGIIGCAAGA	AGGATTTTAA	5500
5501	TGGGACAGGC	CTCCTCCCAA	GAGTTAGACG	AGTACCACAG	ACAGGCTAAT	5550
5551	ATATGGCTCC	AGGTGCTGGA	CGAGGAGATT	AAATTAGGTG	AAAGTCTAAA	5600
5601	AGAGGAAGAT	TTCCTGGAGG	ATGCCACTAT	AGACGAGGAG	GCTCTAAATG	5650
5651	AACTOTTOOT	AAAACCCCAC	AATCTCCTCA	ACACAACTCC	CCCTCCACAC	5700
2021	AACIGIICCI	MMMMGGGGAG	MAICICCICA	AGAGAACICC	CGGIGGAGAG	5700
5701	AAACGAGAGG	CCGTTCGAGA	GAAACACAAC	CTTCTCCATG	ACAGATATGA	5750
5751	CACTCTAAAG	ATCTGAGAG	CCCTGAGAAA	GAAGAAGGCT	CTGGCTCTCG	5800
5801	CATCGCAGTG	GTATCAGTTC	TGCAAGAAAA	GCGATGACAT	GATGCAGTGG	5850
5051	CTCCACACAA	TCCACAAAAC	CATACCACAC	CTTCCACATC	CCCCCCACCA	5000
5051	CIGGACAGAA	ICGAGAAAAAG	CATACGAGAG	CITCCAGAIC	CCCCGGAGGA	5900
5901	GCCCAGAGTG	AAAGCAGTTG	GATATGAGAT	TGACCAGCAG	CGGCCCAAGC	5950
5951	TGGAGGATTT	ACGTGGATTG	GGGCGTGTGC	TTTCAGAGGG	TGGAGCTAGC	6000
6001	AGAATGGTTG	AGCCCCGCCT	CCTGCCTATC	AATAAACGAT	GGACGGAGCT	6050
6051	ACACCTTAAC	TTCACTCCAC	TCCCCCACAA	AACTCACCTC	CACTTTCTCC	6100
6051	AGACGIIAAC	TICACICGAG	IGCGCCACAA	AAG	CAGITICIGC	0100
6101	TGCAGTGTAT	CTCTGAGAAT	GAGGCTCTGC	TGAACTCTCC	CGAGCGCTGG	6150
6151	TCTACTGCAT	TCATCAACCT	TCCCCAACAA	GAGAAGTGCC	TTAAGGAGGT	6200
6201	TAAAGTTAAC	CTGGATAAAC	TGGAGGGTCC	CGTGGCAGAG	GCTGTTGGAA	6250
6251	CACCTCCATC	TCAACCTCAC	CACCETTTCC	TCGTTCACCT	CCTCACCACC	6300
6201	GAGGIGCAIC	ICAACCIGAG	GAGGGIIIGC	IGGIICAGCI	GCICAGGACC	0300
6301	AACTGGGAAA	ACCTTAAAAT	GCTCTACCAG	GACAGACTCA	AACGCCTTGA	6350
6351	AAAAGCCAAG	AAGTTCAATG	AGGAGCTGAA	AATGCTGGAC	AACTGGCTCA	6400
	and the state in the size of a state and the					
6401	CAGATGCTGA	GAGGACGATT	ATGAAGTATG	AACAAGACCC	CATAAATAAT	6450
6401	CAGATGCTGA	GAGGACGATT	ATGAAGTATG	AACAAGACCC	CATAAATAAT	6450
6401 6451	AGAGACCACC	GAGGACGATT TCAA <mark>GG</mark> AGCT	ATGAAGTATG TCAGGCGGGT	AACAAGACCC TTGGAGAAAC	CATAAATAAT AGGAAGCAGC	6450 6500
6401 6451 6501	AGAGACCACC AGTGAAGGGA	GAGGACGATT TCAA <mark>GG</mark> AGCT CTCAATGCTC	ATGAAGTATG TCAGGCGGGT TTGGGACTGA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA	6450 6500 6550
6401 6451 6501 6551	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG	GAGGACGATT TCAA <mark>GO</mark> AGCT CTCAATGCTC GGACCACATT	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC	6450 6500 6550 6600
6401 6451 6501 6551 6601	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG	GAGGACGATT TCAAGGAGCT CTCAATGCTC GGACCACATT TGTCCAACCA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC	6450 6500 6550 6600 6650
6401 6451 6501 6551 6601 6651	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG	GAGGACGATT TCAACCAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC GAGTTTCAAT	6450 6500 6550 6600 6650 6700
6401 6451 6501 6551 6601 6651	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG	GAGGACGATT TCAA <mark>CC</mark> AGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC GAGTTCCAAT GAGTTTCAAT	6450 6500 6550 6600 6650 6700
6401 6451 6501 6551 6601 6651 6701	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA	GAGGACGATT TCAATGAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGGAG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC GAGTTTCAAT GGAAGCGGGT	6450 6500 6550 6600 6650 6700 6750
6401 6451 6501 6551 6601 6651 6701 6751	CAGATGCTGA AGGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC	GAGGACGATT TCAA <mark>SC</mark> AGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGGAG AGCTCAGTGC	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT	6450 6500 6550 6600 6650 6700 6750 6800
6401 6451 6501 6551 6601 6651 6701 6751 6801	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG	GAGGACGATT TCAAUAAGCTC CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGGAG AGCTCAGTGC AGCAGGAAGC	ATGAAGTATG TCAGGCAGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAAG TCAGATCAAC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT	6450 6500 6550 6600 6700 6750 6800 6850
6401 6451 6501 6551 6601 6651 6701 6751 6801 6851	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC	GAGGACGATT TCAA MAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGAGA AGCTCAGTGG AGCAGGAAGC CGCTGATAAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGA AAGCTCTTCA GTCAAACCAT	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAAG TCAGATCAAC TGGAGAAGCA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGGT AGCAAAGGCT ACTAAAGGTT	6450 6500 6600 6650 6700 6750 6800 6850 6850 6900
6401 6451 6501 6551 6601 6751 6701 6801 6851	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC	GAGGACGATT TCAA MAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGGAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTCGGCGAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAA TCGAGATCAAC TGGAGAAAGC CACTTCCTG	CATAAATAAT AGGAAGCAAGC CAGTGCAGTA CAACTCCCCGC GAGTTTCACAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT ACTAAAGGTT	6450 6500 6550 6600 6650 6700 6750 6800 6850 6900 6950
6401 6451 6501 6551 6601 6651 6701 6751 6801 6851 6901	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GCCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC	GAGGACGAT TCAA AGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGGAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAA TCAGATCAAC TGGAGAAGCA GATCTTCCTG	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTCAAT GGAAGCGGGT CCCGGGGGGGT AGCAAAGGCT AGCAAAGGTT AGAAGCAGAG	6450 6500 6550 6600 6700 6750 6800 6850 6900 6950
6401 6451 6501 6651 6601 6751 6801 6851 6901 6951	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CGCAAGGAAC GGAACTGCCG GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG	GAGGACGATT TCAA GACT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGG AGCTCAGTGG AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAACCAT GGGATTTGTCCACT	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TCGGACAACCA GATCTTCCTG TTTGTACCAA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGT AGAAGCAGAG GAACAGCTCT	6450 6500 6550 6600 6700 6750 6850 6850 6900 6950 7000
6401 6451 6501 6551 6601 6751 6801 6851 6801 6851 6901 6951 7001	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG CCAAACTCTC	GAGGACGATT TCAA MAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGGAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGCC	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GGGATTTGTC TCCAATACAA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG TTTGTACCAA AGACTCAGCT	CATAAATAAT AGGAAGCAGCC CAGTGCAGTA CAACTCCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAACGCGT TGAACAGCTCT TGAGCAGTCG	6450 6500 6550 6600 6700 6750 6850 6950 7000 7050
6401 6451 6551 6601 6751 6751 6851 6851 6851 6901 6951 7001 7051	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GCCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ACAAATCGAG CCAAACTGCC CCAACTGCCA	GAGGACGATT TCAA AGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGGAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA TGGACCCAAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GGCATTTGTC TCCAATACAA STTGAGGAT	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAA GATCTACCA GATCTTCCTG TTGGAGAAGCA GACGTAAAAG	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGGGGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT GAACAAGCTCT TGAGCAAGCCG AAAAGAAGCC	6450 6500 6550 6600 6700 6750 6800 6850 6900 6950 7000 7050 7100
6401 6451 6551 6651 6651 6751 6801 6851 6901 6951 7001 7051 7101	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGGACGACAG TGGGCCAAAG GCCAGCACAG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC ACCAAATCGAG CCAAACTCTC CCTACTGCAG AGCAGTGGAA	GAGGACGATT TCAA AAATGACT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTAG AGCTCAGTAG AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGCCAAA ACCCTTTTGG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GGGATTGACCAT TCCAATACCAT GTTGAGGAT CAAAGAGAG	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TAAAGTCAAC TGGAGAAGCA GATCTTCCTG TTGGACAAC AGACTCAGCA GACTCAACC GACGTAAAAG ACCCCCGTGC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT ACTAAAGGTT AGAAGCAGAG GAACAGCTCT TGAGCAGTCC AAAAGAAGCC CAGCCTGAGA	6450 6500 6550 6600 6750 6750 6800 6850 6900 6950 7000 7050 7100 7150
6401 6451 6551 6651 6651 6701 6751 6801 6851 6901 6951 7001 7001 7101 7101	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCCC ATCAACATGC ACAAATCGAG CCAAATCGAG AGCAGTGGAA	GAGGACGATT TCAA MAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGGAG AGCATGCGGAG AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGCC TGGACCCAAA ACCCTTTGG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GGGATTGTC TCCAATACAA AGTGCAGATT	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG TTTGTACCAA AGACTCAGCT GACGTAAAAG ACCGCCGTGC GGACATCCAT	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGAAGCAGAG GAACAGCTCT TGAGCAGTCC CAGCCTGAGA	6450 6500 6550 6600 6700 6750 6800 6850 6900 6950 7000 7050 7050 7150 7200
6401 6451 6551 6551 6601 6751 6801 6851 6901 7001 7051 7101 7151 7201	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GCCAAGAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCCC ATCAACATGC CCAAACTCTC CCTACTGCAG AGCAGTGGAA AGCAGTGGAA	GAGGACGATT TCAA GAGT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GATCTCCTCA AATTTGGGCC TGGACCCAAA ACCCTTTTGG GCGABACCCA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA AACTTTAGAA AAGCTCTTCA GTCAAACCAT GGGATTGTC TCCAATACCAA CAAAAGAGAG AGTGCAGATT	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAAC TGGAGAAGCA GATCTTCCTG TTTGTACCAA AGACTCAGCT GACGTAAAAG ACCGCCGTGC GGACATCCAT	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC GAGTTCAAT GGAAGCGGGT CCCGGGGGGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT GAACAAGCTCT TGAGCAGTCG AAAAGAAGCC CAGCCTGAGA	6450 6500 6550 6600 6700 6750 6800 6850 6950 7000 7050 7100 7150 7250
6401 6451 6551 6651 6651 6751 6851 6851 6851 6851 6951 7001 7051 7101 7151 7201	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGACGACAG TGGGCCAAAG GGCAAGAAAATG CGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG CCAAACTCTC CCTACTGCAG AGCAGTGGAA AGCGCCAGTA CTTAAAGACT	GAGGACGATT TCAA AAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGCCAAA ACCTTTTGG TGACCCAAA ACCTTTTGG GGAAAGACAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GGGATTTGAC GGGATTTGAC CAAAAGAGAG AGTGCAGATT ATGCCAGCTC	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG TTGGACAAC GACTCACCT GACGTAAAG ACCGCCGTGC GGACATCCAT GCAGTGTAAA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCGGT AGCAAAGGCT GAACAGCTCT TGAGCAGTCG AAAAGAAGCC CAGCCTGAGA TCAAGTTCTC CTTTGACTGG	6450 6500 6550 6600 6750 6750 6850 6850 6850 6950 7000 7050 7100 7150 7220 7250
6401 6451 6551 6551 6601 6751 6801 6851 6851 6851 7001 7051 7101 7151 7201 7251	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG CCAAACTCTC CCTAACGCAG AGCAGTGGAA ACGCCAGTA CTTAAAGACT TAGTGCAGCA	GAGGACGATT TCAA MAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG TGATGCGGAG AGCATGCGGAG AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGCC TGGACCAAA ACCCTTTAG GGAAAGACAA GGAAAGACAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GGGATTTGTC TCCAATACAA GTTGAGGAT CAAAAGAGAG AGTGCAGATT ATGCCAGCTC CTCTCGACAA	AACAAGACCC TTGGAGAAACC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG TTTGTACCAA AGACTCAGCT GACGTAAAAG ACCGCCGTGC GGACATCCAT GCACTGTAA ATTCAACAAA	CATAAATAAT AGGAAGCAGCC CAGTGCAGTA CAACTCCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGGT AGCAAAGGCT AGCAAAGGCTCA TGAGCAGACG GAACAGCTCT TGAGCAGTCCC CAGCCTGGACAG TCCTTGACTGG TCCTGGACAG	6450 6500 6550 6650 6700 6750 6800 6850 6900 6950 7000 7000 7150 7150 7200 7250 7300
6401 6451 6501 6551 6601 6751 6701 6751 6801 6851 6901 7001 7051 7101 7151 7201 7251 7301	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ACAAATCGAG CCAAACTGCC CCAACTGCCA AGCAGTGGAA AGCAGTGGAA TTAAAGACT TAGTGCAGCA AGCTTGACGA	GAGGACGATT TCAA AGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA TGGACCCAAA ACCCTTTTGG GGAAAGACAA GGAGATGCGG TTGGTTGGCG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTGAAACCAT GGGATTGGC TCCAATACAA AGTGCAGAT CAAAAGAGAG AGTGCAGAT ATGCCAGCC CTCTCGACAA CTGCTGGATC	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAAC TGGAGAAAGCA GATCTTCCTG TTTGTACCAA AGACTCAGCT GACGTAAAAG ACCGCCGTGC GGACATCCAT GCACTGTAA ATTCAACAAA ACATGGTTCA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GATTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT GAACAGCTCT TGAGCAGTCG AAAAGAAGCC CAGCCTGAGA TCAAGTTCTC CTTTGACTGG GACCCAGAGA	6450 6500 6550 6650 6700 6750 6800 6850 6900 6950 7000 7050 7100 7150 7250 7250 7300 7350
6401 6451 6551 6551 6601 6751 6751 6801 6851 6801 6851 7001 7051 7101 7151 7201 7251 7301 7351	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGACGACAG TGGGCCAAAG GGCAAAAATG CGCAGGAGAC GGAACTGCCG CATCTCTCCC ATCAACATGC ACCAAATCGAG CCAAACTCTC CCTACTGCAG AGCAGTGGAA AGCGCCAGTA CTTAAAGACT TAGTGCACGA AGCTTGACGAG	GAGGACGATT TCAA AAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTAG AGCTCAGTAG GGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGCCAAA ACCCTTTTGG TGACCGATCTG GGAAAGACAA GGAGATGCGG TTGGTTGGCG GGGATCTGGA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAAACCAT GGGATTTGACGAT TCCAATACAA GTTGAGGAT CAAAGAGAG AGTGCAGATT ATGCCAGCTC CTCTCGACAA CTGCTGGATC	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG TTGGACAAC GACTCACCA GACTCACCA GACTCACCA GGACATCCAT GCACGTGTA ATCCAACAA ACATGGTTCA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCGGT ACTAAAGGT AGCAAAGGCTCT TGAGCAGTCG AAAAGAAGCC CAGCCTGAGAA TCAAGTTCTC CTTTGACTGG GACCCAGAGA TCCAAGACTCAA	6450 6500 6500 66700 6750 6800 6850 6900 7000 7050 7100 7150 7200 7250 7350 7350 7350 73400
6401 6451 6551 6551 6601 6751 6851 6851 6851 6851 6851 6901 6951 7001 7151 7201 7251 7301 7351 7351	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG AGCAGTGGAA AGCCCAGTA AGCAGTGGACA GTAATGGTGACGA GTAATGGTGACGA	GAGGACGATT TCAA MAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGCC TGGACCAAA ACCCTTTTGG GGAAAGACAA GGAGATGCGG TTGGTTGGCG GGGATCTGGG GGGATCTGCG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAAACCAT GGGATTGTC CCAATACAA GTTGAGGAT ATGCCAGCTC CTCTCGACAA CTGCTGGATCA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG TTTGTACCAA AGACTCAGCT GACGTCAA ACCGCCGTGC GGACATCCAT GCACGTGTAA ATTCAACAAA ACATGGTTCA GACATCGCCG	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGGT AGCAAAGGCT AGCAAAGGCTCT TGAGCAGTCCT TGAGCAGTCCT CCAGCCTGGACAG TCCAGGTCTCC CTTTGACTGG TCCTGGACAG GACCAGAGA	6450 6500 6650 6650 6700 6750 6850 6850 6850 7000 7050 7050 7050 7100 7250 7250 7350 7350 7450
6401 6451 6501 6551 6601 6751 6701 6751 6801 6851 6801 6851 7001 7051 7101 7151 7201 7251 7301 7351 7401	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GCCAAGAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ACAAATCGAG CCAAACTGCC CCTACTGCAGA AGCAGTGGAA AGCAGTGGAA TTAAAGACT TAGTGCAGCA GTAATGGTGG CAGCACTG CAGCACTG	GAGGACGATT TCAA AGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA TGGACCCAAA ACCCTTTTGG GGAAGGCCG GGAAAGACAAG GGGGATCTGGA CAAGACATGG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAAACCAT GGGATTGGC TCCAATACCA AGTGCAGATT ATGCCAGCTC CTCCGACAA CTGCTGGATC CGAGATCAAC AGGCAGATCAAC AGCAAAGATG	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TAAGTCAAC TGGAGAAAGCA GATCTTCCTG TTGTACCAA AGACTCAGCT GACGTAAAAG ACCGCCGTGC GGACATCCAT GCACTTCAACAA ACATGGTTCA GAAATGACGG CCCACAGCTG	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT GAACAGCTCT TGAGCAGTCG CAAGCTCGAC CCAGCCTGAGA TCCAAGTTCTC CTTTGACTGG GACCCAGAGA TCCAAACTCAA AACAAACAAC	6450 6500 6650 6650 6700 6750 6800 6850 6950 7000 7050 7100 7150 7250 7300 7350 7350 7400 7450
6401 6451 6551 6551 6601 6751 6801 6851 6851 6851 7001 7051 7101 7151 7201 7151 7201 7351 7301 7351 7401	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGC GGAACTGCCG GGAACTGCCG CATCTCTCCC ATCAACATGC ACCAAATCGAG AGCAGTGGAA AGCGCCAGTA CTTAAAGACT TAGTGCACGA GTAATGGTGG CCAACTGCGGG CCACACTGCGGC	GAGGACGATT TCAA AACA GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTGGCAACA AGCAGGAAGCC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGCCGAA ACCCTTTTGG TGGACGCTCTG GGAAAGACAA GGAGATGCGG TTGGTTGGGC GGGATCTGGA CAAGACATGG TCAGGACCTG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAAACCAT GGGATTTGAC GGATTTGAC CAAAAGAGAG AGTGCAGATT ATGCCAGCTC CTCTCGACAA CTGCTGGATC CGAGATCAAC AGCAAAGAG AAGAACAAAA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG TTGTACCAA GACTCATCA GACGTAAAG ACCGCCGTGC GGACATCCAT GCACTTGTAA ACTAGGTTCA ACTAGGTCA CCACCGTCC CCAACAGCG CCCACAGCTG	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCGGT ACTAAAGGT AGCAAAGCAGAG GAACAGCTCT TGAGCAGTCC CAGCCTGAGAA TCAAGTTCTC CTTTGACTGG GACCCAGAGA TCCAAGTCTA AACAAACAAA AACAAACAAA	6450 6500 6550 6600 6700 6750 6850 6900 6950 7000 7050 7150 7200 7150 7250 7350 7350 7350 7450 7450
6401 6451 6551 6651 6701 6751 6851 6851 6851 6901 6851 7001 7101 7101 7101 7151 7201 7301 7351 7451 7451	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG ACCAGTGCAG AGCAGTGGAA AGGCCAGTGACA AGCTTGACGA GTAAAGACT TAGTGCAGCA GTAATGGTGG CAACTATA	GAGGACGATT TCAASCAGCT CTCAATGCTC GGACCAACAT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGCC TGGACGAAA ACCTTTTGG GGAAAGACAA GGAGATGCGG GGGATCTGGC GGGATCTGGC CAGAACATGG CAGAACCTG CAGAACCTG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAAACCAT GGGATTGTC CCAATACAA GTTGAGGAT ATGCCAGCTC CTCTCGACAA CTGCTGGATCA AGGAACAAAA TGAGAAGTTA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG TTTGTACCAA AGACTCAGCT GACGTCAAC GACGTCAAC CGACGTGAA ATTCAACAAA ACATGGTTCA GAATGATCA CCAACAGCTG CCCAACAGCTG CCAATAACCC CAGGCACATT	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCTCT TGAGCAGTCCT TGAGCAGTCCT CAGCCTGAGA TCAAGTCTCC CTTTGACTGG TCCTGGACAG GACCAGAGA TCAAACTCAA AACAAACTAA	6450 6500 6500 6600 6700 6750 6850 6850 7000 7050 7000 7050 7150 7250 7350 7450 7550
6401 6451 6501 6551 6601 6751 6701 6751 6801 6851 6801 6851 7001 7051 7101 7151 7201 7251 7301 7351 7551	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ACAAATCGAG CCAAACTGCCG AGCAGTGGAA AGCAGTGGAA AGCAGTGGAGA TAATGGCAGCA GTAATGGTGG CCAACTGCGG GCAACTATTA TCATGCCAAG	GAGGACGATT TCAA GAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCTGAGGAAGC CGCTGATAAA GCTGGGCGAA AATTTGGGCC TGGACCCAAA ACCCTTTTGG GGAAAGACAA GGAGATGCGG GGGATCTGGA CAAGACATGG CAGACCCC CTGACGGCCA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAAACCAT GGGATTGCC TCCAATACAA AGTGCAGAGA AGTGCAGATC CTCCCGACAA CTGCTGGATC CGCAGATCAAC AGCAAAGATG AAGAACAAAA GAGTTCTTAC	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TAAGTCAAC TGGAGAAACCA GATCTTCCTG GACGTCAACA GACGTCAACA GACGTCAGCT GCACGTCAACA ACCGCCGTGC GGACATCAACA ACATGGTTCA GAAATGACGG CCCACAGCTG CAAGACACCA CAGGCACATT ACTGCAGAAC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GATTTCACT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT GAACAGCTCT TGAGCAGTCG CAGCCTGAGA TCAAGTTCTC CTTTGACTGG GACCCAGAGA TCCAAGCACAG GGCAAGACTCA AACAACAAA AGGACAAGG GGGAAGATTC	6450 6500 6550 6600 6700 6700 6850 6900 6950 7050 7100 7150 7200 7150 7200 7350 7350 7350 7350 7400 7450 7550 7600
6401 6451 6551 6601 6651 6701 6751 6801 6851 6801 6851 6901 7001 7051 7101 7151 7201 7351 7401 7351 7451 7551 7601	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGC GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG CCAAACTCTC CCTACTGCAGA AGCGCCAGTA AGCGCCAGTA AGCGCCAGTA AGCTTGACGGG GCAACTATGA CCACACTGCGGC GCAACTATTA	GAGGACGATT TCAA CACCA CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTGGCAGA AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGC TGGACCCAAA ACCCTTTTGG GGAAAGACAA GGAGATCCGG TTGGTCGGCG GGGATCTGGA CAAGACCATGG CCAGACCCG CTGACGCCA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTGTCCACT GGGATTTGAC GTCAAAGAGAG AGTGCAGATT ATGCCAGCTC CTCTCGACAA CTGCTGGATC AGCAAAGATG AAGAACAAAA TGACGAACTAA GAGTTCTTAC GCCAGAAAAA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TAAAGTCAAC GATCTTCCTG TTGGAGAAGCA GATCTTCCTG GACGTCACCT GACGTAAAG ACCCCCGTGC GGACATCCAT GCACTTCAAAA ACATGGTTCA CAA ACTGGTCA CCACCGTGC CCAACAGCT CAATGACCG CCAACAGCT ACTGCAGAAC GGGTGCACCC	CATAAATAAT AGGAAGCAGC CAGTGCAGGA CAACTCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCGGT ACTAAAGGT AGCAAAGCAGAG GAACAGCTCT TGAGCAGTCC CAGCCTGAGAA TCAAGTTCTC CTTTGACTGG GACCCAGAGA TCCAACTCAA AACAAACAAA TGAGACACGG GGGAAGATTC ATGTATAAGG	6450 6500 6550 6600 6700 6750 6850 6900 6950 7000 7050 7150 7200 7150 7250 7350 7350 7450 7550 7650
6401 6451 6501 6551 6601 6751 6701 6751 6801 6851 6901 6851 7001 7051 7001 7251 7301 7251 7351 7551 7551 7661	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG ACCAGTGCAGA AGCGCAGTGAA AGCCCAGTGACA AGCTTGACGA GTAAAGACT TAGTGCAGCA GTAATGGTGG CCAACTATTA TCATGCCAAG CCAACTATTA	GAGGACGATT TCAA GAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCTGATGCGGAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGCCGAA AATTTGGGCCA TGGACCCAAA ACCCTTTTGG GGAAAGACAAG GGAATGCGG TGGTTGGCG GGAACTCGG CAGAACCTG CAGAACCTG CAGACCGC CTGACGGCCA TTGGTTGGAG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTGACCACT GGGATTGTC CCAATACAA CAAAAGAGAG AGTGCAGATT ATGCCAGCTC CGAGATCAAC AGGAACAAA TGAGAAGTA AGGAACTAAC AGGAACTAAC	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC GGACATCACC GGACATCCAC GACTCTCCTG GACGTCAAC GACCCCGTGC GGACATCCAT GCACGTGAA ACTCGACACA ACATGGTTCA GAATGATCA CCAACTGGTCA GACACCCC CCAACACC CCAACACC CCACACCTG CCCACACCTG CCCACACCTG ACTGCAGAAC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGGGGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT GAACAGCAGTCG CAACAGCTCT CTGGACAGG GAACAGCAGCG GAACCAGGAG GACCCAGAGA TCATGACTCAA AACAAACTCAA AACAAACTCAA AACAAACACAA CAGCCCAGAGA CACCCAGAGA CAAACTCAA	6450 6500 6500 6650 6700 6750 6850 6850 6850 7000 7050 7050 7150 7250 7350 7450 7550 7550 7600 7550
6401 6451 6501 6551 6601 6751 6801 6851 6801 6851 6901 7051 7001 7051 7201 7301 7351 7301 7351 7401 7551 7551 7551	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ACAAATCGAG CCAAACTGCCG AGCAGTGGAA AGCAGTGGAA AGCTGACGAG GTAATGGTGG GCAACTATGA GCAACTGCGG GCAACTATTA TCATGCCAAG ACCTCCAGCGA ACTCCAGCGA AAAGCCAATG	GAGGACGATT TCAA GAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA TGGACCCAAA ACCCTTTTGG GGAAAGACAA GGAGATCGGG GGGATCTGGA CAAGACATGG CTGGACGCC CTGACGGCCA TTGGTTGGAG AAAAACTCGA AAAAACTCGA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAAACCAT GTCAAACCAT GTGTCACAC GGAGATTGTC CCAAAAGAGAG AGTGCAGATT ATGCCAGCTC CTGCTGGATC CGAGATCAAC AGCAAAGATG AAGAACAAAA TGAGAAGTTA GAGTTCTTAC GCCAGAAAAA GAGCTGCAGAG	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGCCTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG GACGTCAGCT GACGTCAGCT GCACGTCAGC GCACATCCAT GCACGTCAAAA ACTGGCTCA GACATCACA GAAATGACGG CCCACAGCTG CCACAGCTG CCAGCCACATC ACTGCAGAAC GGGTGGAGCC AAAGTTCCCC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC GAGTTCCAGG GAGATTCAAT GGAAGCGGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCTCT TGAGCAGTCG CAGCCTGAGA TCCAGGTCTC TCCTGGACAG GACCCAGAGA TCCAAACTCAA AACAACACA TGAAGCACCG GGGAAGATTC ATGTATAAG TCTCATTAAG ACAGCGTTGA	6450 6500 6550 6600 6700 6700 6850 6900 6950 7050 7150 7150 7150 7150 7200 7350 7350 7350 7350 7350 7400 7450 7550 7600 7550 7600
6401 6451 6551 6601 6651 6701 6751 6801 6851 6801 6851 7001 7051 7101 7151 7201 7151 7351 7401 7351 7401 7551 7501 7551 7651 7651	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGC GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG CCAAACTCTC CCTACTGCAGA AGCGCCAGTA AGCGCCAGTA AGCTGACGAG TCACTGCGGG GCAACTATGA CCACACTGCGGC GCAACTATTA ACTGCCAGGA ACTCCAGCGA ACTCCAGCGA ACTCCAGCGA ACTCCAGCGA ACACTTAAAA	GAGGACGATT TCAA CATC CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTGGCAGA AGCTCAGTAG AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGC TGGACCCAAA ACCTTTTGG GGAAAGACAA GGAGATCCGG TTGGTCGGCG GGGATCTGGA CAAGACATGG TCAGAACCTG CAGACCCC TTGGTCGGCA CAGACGCCA TTGGTTGGAG AAAAACTCGA GGCCAAAATG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTGTCCACT GGGATTTGAC GTCCAATACAA GTTCAGGAT ATGCCAGCTC CTCCTCGACAA CTGCTGGACA CTGCTGGACA AGCAAAGATG AAGAACTAAA GAGCTGGAAG CACATGCAA	AACAAGACCC TTGGAGAAAC TCGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC GATCTTCCTG TTGGAGAAGCA GATCTTCCTG GACGTCAAC GACTCACCA GACTCACCA GACTCACCA GGACATCCAT GCACGTGTAA ACATGGTTCA CCACCGTGC CCAACAGCT CCAACAGCT CCAACAGCT CCAACAGCT CCAACAGCT CCAACAGCT CCAACAGCT CCAACAGCC CAACATCCC CAGCACACT ACTGCAGAAC GGGTGGAGCC AAAGTTTCG CAACTTCT	CATAAATAAT AGGAAGCAGC CAGTGCAGGA CAACTCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGCGGGT ACTAAAGGT AGCAAAGCAGAG GAACAGCTCT CAGCTGGACAG GACCCAGAGA TCAAGTTCTC CTTTGACTGG GACCCAGAGA TCAAACTCAA AACAAACAAA TGAGACACGG GGGAAGATTC ATGTATAAGG TCCTATTAAG ACAGCGTTGA AAGGATCTAC	6450 6500 6550 6600 6700 6750 6850 6900 6950 7000 7000 7150 7200 7150 7250 7350 7350 7450 7550 7550 7650 7700 7750
6401 6451 6501 6551 6601 6751 6701 6751 6801 6851 6901 6851 7001 7051 7001 7051 7351 7301 7351 7551 7601 7751	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG CCAAACTCTC CCTACTGCAG AGCAGTGGAA AGCGCAGTA CTTAAAGACT TAGTGCAGCA GTAATGGTGG CCAACTATTA TCATGCCAGG CCAACTATTA TCATGCCAGA AAAGCCAATG ACACTCCAGCGA AAAGCCAATG ACACTCCAGCGA	GAGGACGATT TCAA GAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCTGATGAGAAGC CGCTGATAAA GCTCGCCAA GATCTCCCTA TGGACCCAAA ACCCTTTTGG GGAAAGACATGG GGGATCTGGA CAGACCGC TCGGTCGGC CAGACCGC CTGACGGCCA TTGGTTGGAG AAAAACTCGA AAAAACTGG GGCCAAAATG GACTCAGATG GACTCAGATG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGCG GCTCACTGAG AGTTAAATGA GCTGTGGCTG GCTACTGGCTG GTGACACAT GTCAAACCAT GGGATTTAGAG AGGGATTGTC TCCAATACAA AGGCCAGAT ATGCCAGCAC CTCTCGACAA CTGCTGGATC CGAGATCAACA AGGATCATAC GAGATCTAC GCCAGAAAAA GAGTCCTAC GCAGAAGATA GAGCTGCAAA AATGTCACTA	AACAAGACACC TTGGAGAAAC TCTGAGTCCA TCGACTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAAC TGGAGAAGCA GATCTTCCTG GACGTCAAC GACGTCAACT GCACGTCAGCT GCACGTCAT GCACGTCAACA GAAATGGTTCA GAAATGGTTCA GAAATGACGG CCCACAGCG CCAACACCC CAGGCACATT ACTGCAGAAC GGGTGGAGCC AAAGTTCCG	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCTC TGAGCAGTCG CAACAGCTCC CAGCCTGAGA CAAGTCTCC CTTTGACTGG GCACAGGAGAG TCCAAACTCAA AACAACAACAA CAGCGCTTGA CAGCGCTTGA CAGCGCTTAG ACAGCGTTAC AAGCATCTCA AAGCATCTCA	6450 6500 6650 6650 6700 6700 6850 6900 6950 7050 7100 7150 7150 7250 7300 7350 7350 7350 7550 7550 7550 75
6401 6451 6551 6601 6751 6701 6751 6801 6851 6801 6851 7051 7101 7151 7201 7351 7301 7351 7601 7651 7701 7751 7801	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GCCAGGCCAAAG GCCAGGCCAAAG GCCAGGCCGA CCTGCTGGC CGCAAGGAAC GGAACTGCCG ACAAATCGAG ACCAGTGGAA AGCAGTGGAA AAGCCAGTG GCAACTGCCG GCAACTGCGG GCAACTGCGG GCAACTGCGG GCAACTGCGG ACCCCAGCA ACCTCCAGCGA ACCCCAGCGA ACCCCAGCGA ACCCCAGCGA AAGCCCAATG AGCAGTGGCA CTGACTCTT	GAGGACGATT TCAA GAGT TCAATGCTC GGACCACATT TGTCCAACCA CTCTCTCGCAG AGCTCAGTGC AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGCCAA AATTTGGGCC TGGACCCAAA ACCCTTTTGG GGAAAGACACG GGGATCTGGA CAAGACCTGG CAGACCGC CTGACGGCCA TTGGTTGGAG AAAAACTCGA GGACTCAGAACGA GGACTCAGAGCGA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAAACCAT GTCAAACCAT GTGTCCACT TCCAATACCA GTTGAGGAT CAAAAGAGAG AGTGCAGATC CTGCTGGATC CTGCTGGATC CGCAGAACAAA TGAGAAGTTA GAGTTCTTAC GCCAGAAAAA GAGCTGGAAG CAGATGTCATA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGG TCAGATCAAC TGGAGAAGCA GATCTTCCTG TTTGTACCAA AGACTCAGCT GACGTAAAAG ACCGCCGTGC GACATCCAT GCACTCAAA ATTCAACAA ACATGGTTCA GAAATGACGG CCCACAGCTG CAAGTACACG CCACAGCTG CAAGTACACG CCACAGCTG CCAATAACCC CAGGCACATT ACTGCAGAAC GGGTGGAGCC AAAGTTTCCT TAGACTGGC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT GGAAGACAGCT TGAGCAGTGG CAGCCTGACAG TCCAGACTGA CAAGTCTC CTTTGACTGG GGCACACAGA TCCAAACTCAA AACAACACA ACAACCTGA AAGGATCTAC AAAGGATCTAC AAACAACTA AAACAACTA AAACAACTA	6450 6500 6550 6600 6700 6700 6850 6900 6950 7050 7150 7150 7200 7150 7250 7350 7350 7350 7350 7400 7550 7500 7550 7500 7550 7500 7550 7500 7550 7500 7550 7
6401 6451 6551 6601 6651 6701 6751 6801 6851 6801 6851 7001 7051 7001 7151 7201 7151 7201 7351 7351 7401 7351 7401 7551 7651 7651 7651 7651 7651	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGC GGAACTGCCG CATCTCTCCC ATCAACATGC CCAACATGC ACCAACTGCCA AGCAGTGGAA AGCGCCAGTA AGCGCCAGTA CTTAAAGACT TCACTGCGGC GCAACTATGA CGAACTACTA TCATGCCAGG ACACTGCCAGG AACTCCAGCGA AAGCCCATG AGATTTAAAA AGCAGTGGCA CGACTCTTC CACCAGCGA	GAGGACGATT TCAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTGTCCACT GGGATTTGAC GTCAAACCAT GTTGAGGAT ATGCCAGGAT CTACAGAGAG AGTGCAGATC AGCAAAGATG AAGAACAAAA GAGCTGGAAG CAGATGTCATA CGACACAACA	AACAAGACCC TTGGAGAAAC TCGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC GATCTTCCTG TTGGAGAAGCA GATCTTCCTG GACGTCAAC GACTCACCA GACTCACCA GACTCACCA GGACATCCAT GCACGTGTAA ACCACCGTGC GGACATCCAT CCACAGCTGCA CCACAGCTG CCAACAGCTCA CAGGCACCC AAAGTTCCC AAAGTCACC AAAGTGACC CAAGTGACC CAAGTGACC CAAGTGACC CAAGTGACC	CATAAATAAT AGGAAGCAGC CAGTGCAGGA CAACTCCCGC GAGTTTCAAT GGAAGCGGGGT CCCGGGTGGT AGCAAAGGCT ACTAAAGGT AGCAAAGGCT CAGACAGAG GAACAGCTCT CTTGACTGG GACCCAGAGA TCAAGTTCTC CTTTGACTGG GACCCAGAGA TCAAGTCTC ATGTATAAGG GGAACACCA AACAAACAAA AACAAACTAA AACGACTCAC AAGGATCTAC AAAGGATCTAC AAAGGATCTAC AAAGAACTAA	6450 6500 6550 6600 6700 6750 6850 6900 6950 7000 7150 7200 7150 7250 7350 7350 7450 7350 7450 7350 7450 7550 7550 7550 7650 7700 7750 7800 7800
6401 6451 6501 6551 6601 6751 6701 6751 6801 6851 6801 6851 7001 7051 7001 7051 7301 7351 7301 7351 7551 7601 7551 7601 7751 7801 7751	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA GGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG CCAAACTCTC CCTACTGCAG AGCATGCAGA AGCTTGACGA GTAATGGTGG CCAACATGC TCACTGCAGCA AGCTTGACGA AAGCTATA TCATGCCAGGA AAAGCCAATG AGATTTAAAA AGCATGGCA CTGACTCTT GAGCATGACA	GAGGACGATT TCAA GAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCTGAGGAGA AGCTCAGTGC AGCAGGAGAA GATCTCCTCA AATTTGGGCC TGGACCCAAA ACCCTTTTGG GGAAAGACATGG GGAATGCGG CAGACCGC TCGGTCGGA CAGACCCC TGGTCGGA CAGACCCC TGGTCGGA CAGACCCC TGGTCGGA AAAACTCGG GGCAAAATG GACCTAGATG GACTCAGATG ATGCAGACGA CTGGCTGGG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGCG GCTCACTGAG AGTTAAATGA GCTGTGGCTG GTGTGGCTG GTGTCACAT GTCAAACCAT GTCAAACCAT GGGATTTAGC CCAAACACAA CTGCCGGAT CCAAAGAGAG AGTGCAGATC AGGATCAACA CTGCTGGATC CGAGATCAACA GAGTTCTTAC GCCAGAAAAA GAGTCTCTAC GCCAGAAAAA CAGCTGGAAC AATGTCACTA CAGACTGCACA	AACAAGACACC TTGGAGAAAC TCTGAGTCCA TCGACTCGAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAAG GATCTACCA GATCTTCCTG GACGTCAAC GACGTCAGCT GACGTCAGCT GACATCACCA GACATCCACAA ACCGCCGTGC GGACATCCA CCAACAGCTG CCAACAGCTG CCAACAGCTG CCAACAGCC GGACGTCCACACT ACTGCAGAAC GGGTGGAGCC AAAGTTCCG AAAGTTCCG AAAGCCGCACATC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCTCA TGAGCAGTCG CAACAGCTCG CAGCCTGAGA CAAGTTCTC CTTTGACTGG CCCTGGACAGG GGCAAGACTCA AACAACAAC AACAACAAG TCCATTAAGG TCCCATTAAG CCTCATTAAG ACGCGTCGA AACGCGTCA AACGAACTA AACAACAACA	6450 6500 6550 6600 6700 6700 6850 6900 6950 7050 7100 7150 7100 7150 7200 7350 7350 7350 7350 7350 7550 7550 75
6401 6451 6551 6601 6751 6801 6851 6801 6851 6801 6851 7051 7051 7051 7301 7351 7401 7351 7601 7551 7601 7551 7601 7551 7851 7801	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GCCAGGCCAAAG GCCAGGCTGGA CCTGGCTGGA CCCAGCTGCC GCAACTGCCG ACAACTGCCG ACAAATCGAG ACCAGTGGAA AGCAGTGGAA AGCATGCAGA GTAATGGTGG GCAACTATG CAGCACTGG GCAACTATA AGCCAGCGGA ACTCCAGCGA AAGCCCAATG AGCATGCAAG AGCATGACAATG AGCATGGCA AGCATGACA	GAGGACGATT TCAA GAGCT TCAATGCTC GGACCACATT TGTCCAACCA CTCATGCGAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGCCAA AATTTGGGC TGGACCCAAA ACCCTTTTGG GGAAAGACATGG TGGATCGGG GGGATCTGGA CAAGACATGG CTGACACCG CTGACGCCA TTGGTTGGAG AAAAACTCGA GGCCAAAATG GACTCAGACGA GGCCAAAATG GACTCAGACGA CAGACCGAG CAGACCGAG CTGGCTTGGA GACTCAGACGA CTGGCTTGGA GGCAAGACGA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAG AACTTTAGAG GTCAAACCAT GTCAAACCAT GTCAAACCAT GTCAAACCAT GGCATTGTC CCAATACAA AGTGCAGAT CAAAAGAGAG AGTGCAGATC CTCCTGGCAAC AGGAAGTACAAC AGGAAGTACAAC AGGATGTCAC GAGATGTCATA CGACACAAGC CCAACATCAA CTTCGGCAGT	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGG TCAGATCAAC TGGAGAAGCA GATCTTCCTG GACGTAAAAG ACACCCGTGC GACGTAAAAG ACCGCCGTGC GGACATCCAT GCAATAACCC CAAGGTTCA GAAATGACGG CCAACAGCT ACTGCAGAAC GGGTGGAGCC AAAGTTCGC AAAGTTCGC AAAGTGCACAC AAAGCGCACATA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT CAGACAGCT TGAGCAGTGG CAAGCTGACAG TCCAGAGTCTC CTTGGACAGG GGCACAGAGA TCCAAACTCAA AACAACACAA AGGACACAG TCTCATTAAG ACAGCGTTGA AAAGAACTAA AAACAAACTA AAACAAACTA AAACAAACTA	6450 6500 6550 6600 6700 6700 6850 6900 6950 7050 7150 7200 7150 7250 7350 7350 7350 7350 7350 7350 7400 7350 7350 7400 7550 7500 7550 7600 7550 7600 7550 7600 7550 7600 7550 7600 7550 7600 7550 755
6401 6451 6551 6601 6751 6801 6851 6801 6851 6801 6851 7001 7051 7101 7151 7201 7151 7201 7351 7401 7351 7401 7551 7651 7651 7651 7651 7651 7651 765	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGAGACGACAG TGGGCCAAAG GGCAAGAAG CGCAGGCGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC CCAACATGC ACCAGTGGA AGCAGTGGA AGCAGTGGA AGCAGTGGA CTTAAAGACT TCACTGCAGG GCAACTATGA CGCAACTATTA ACCTGCCAGG ACACTGCCAGG ACACTGACA AGCTCAACG ACCCAGCGA ACTCCAGCGA ACTCCAGCGA ACTCCAGCGA ACTCCAGCGA CGAACTATTA AGCAGTGGCA CGACTCTTT GAGCATCTT CTGGAGAAT	GAGGACGATT TCAA CATC CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTGGCAAGA AGCTGGCAAGA GACTCAGTGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGC TGGACCCAAA ACCCTTTTGG GGAAAGACAA GGAGATCCGG CAGACCCG TTGGTTGGGC GGGATCTGGA CAAGACCAG CAGACCCG TTGGTTGGAG CAAGACCTGG CAGACCCCAA ATGGTTGGAG AAAAACTCGA AACCCAAATG GACTCAGATG ATGCAGACG CAGACTGGA CTGGCTGGGC CGGACTGGA CTGGCTGGGA CTGGCTGGGA CTGGCTGGAG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTGTCCACT GGGATTTGACGAT TCCAATACAAT ATGCCAGGAT CTCCTCGACAA CTGCTGGACA CTGCTGGACA AGGACTCACA GAGTCTTAC GCCAGAAAAA GAGCTGGAAG CAGATGTCATA CGACACAAGAC CCAACATCAA	AACAAGACCC TTGGAGAAAC TCGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TAAAGTCAAC GATCTTCCTG TTGGAGAAGCA GATCTTCCTG GACGTCACCA GACTCACCA GACTCACCA GACTCACCA GGACATCCAT GCACGTGAA ACTGGTTCA CCACCGTGC CCAACAGCTG CCAACAGCTG CCAACAGCTC CAGGCACACT ACTGCAGAAC CAGGTGGACC AAAGTCAAC AAAGCGGCA ACAGCATA ACCAGCATA ACCAGCATA	CATAAATAAT AGGAAGCAGC CAGTGCAGGA CAACTCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCGGT ACTAAAGGTT AGAAAGCAGAG GAACAGCTCT CAGCTGGACAG GACCCAGAGA TCAAGTTCTC CTTTGACTGG GACCCAGAGA TCAAACTCAA AACAAACAAA AGGACACGG GGGAAGATCT ATGTATAAGG CTCATTAAG ACAGCGTTGA AAACAAACTA AAACAAACTA AAACAAACTA AAACAAAC	6450 6500 6550 6600 6700 6750 6850 6900 6950 7000 7150 7200 7150 7250 7350 7450 7350 7450 7350 7450 7550 7450 7550 7550 7650 7650 7700 7550 7650 76
6401 6451 6501 6551 6601 6751 6801 6851 6801 6851 7001 7051 7001 7251 7301 7351 7551 7601 7551 7601 7551 7801 7851 7801 7951 8001	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA GGCAACTGCCG CATCTCTCCC ATCAACATGC CCAACATGCC ACAAATCGAG CCAAACTCTC CCTACTGCAG AGCATGCGAG AGCTTGACGA GTAATGGTGG TCACTGCGGC GCAACTATTA TCATGCCAAG AAGCTTGACGA AAGCCTATTA TCATGCCAGGA AAGCCAATG AGATTTAAAA AGCATGCGA AGATTTAAAA AGCATGCAAT CTGACACTGT GAGCATCTT GAGCATCTT CTGAGAAAT CTTACAGAT	GAGGACGATT TCAA AGCT TCAA GAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCTGGACGAG AGCTCAGTGC AGCAGGACGAA GATCTCCTCA AATTTGGGCC TGGACCCAAA ACCCTTTTGG GGAAAGACATGG GGGAATGCGG TCGGACGCCA TTGGTTGGAG CAGACCGC CTGACGGCCA TTGGTTGGAG GGCCAAATG GACTCAGATG ATGCAGACGA CAGAACTGG CAGACCACATG GGCCAAATG GACTCAGATG ATGCAGACGA CTGGCTTGGG GGAAGCTGG CGGAACTGG CAGACCTTAATG GCCACCTTA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG GTGTCACTGA AACTTTAGAA AACTTTAGAA AAGCTCTTCA GTGAAACCAT GTGTCACAT CCAAACAA AGTGCAGAT CTGCAGACA CTGCTGGATC CGAGATCAAC AGGACTGAAA GAGTTCTTAC GCCAGAAAAA GAGTTCTAC GCCAGAAAAA CGACACAAGC CCAACATCAA CTCCGGCAG AGGACGACA	AACAAGACACC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TAAGTCAAC TGGAGAAACCA GATCTTCCTG GACGTCAAC GACGTCAGCT GACGTCAGCT GACGTCAGCT GCACGTCAACA CCGCCGTGC GGACATCCA GAAATGACGG CCAACAGCTG CCAACAGCTG CCAGCACATT ACTGCAGAAC AGGTGGAGCC AAAGTCGCC AAAGTCGCC AAAGCGGCA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GATTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT CAGCCAGAGA CAAGCATCC CAGCCTGAGA GAACAGCAGTCG CCTGGACAGG GGCAAGAGATCC ATGTATAAGG TCCATTAAG GGGAAGATCT AAGCATCTA AAAGGATCTAC AAACAACTA AAAGGATCTAC AAACAACTA AAACAACTA AAACAACTA AAAGGATCAC AAACAACTA AAATGACCGA GGGACAAGG CTACTTGGA	6450 6500 6650 6650 6700 6700 6850 6900 6950 7050 7100 7150 7200 7150 7200 7350 7350 7350 7350 7350 7400 7550 7550 7550 7550 7550 7550 75
6401 6451 6551 6601 6551 6701 6751 6801 6851 6801 6851 7051 7101 7151 7201 7351 7301 7351 7401 7551 7601 7551 7601 7551 7851 7851 7851 7851 7851	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GCCAGGACAG CCTGGCTGGA CCCTGGCTGGA CCCAGCTGCC GCAACTGCCG CATCTCTCCC ACAAATCGAG CCAAACTGCC ACAAATCGAG AGCAGTGGAA AGCAGTGGAA AGCAGTGGCA GCAACTGCCAG GCAACTATTA AGCACACTG CGCACTATA AGCACACTG AGCATGACAATG AGCATGCAAG AGCATGCAAG AGCATGCAAT AGGCAGATCT CTGACAATC	GAGGACGATT TCAA GAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTCTCTCGCAG AGCTGGTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGCCGAA AATTTGGGCCAA AATTTGGGCCAA ACCCTTTTGG TGACGGCCTG GGAAAGACATGG CAGACCGCG CTGACGCCA CTGGTTGGAG AAAAACTCGA GGACACGCG CTGACGCCA ATGGTTGGAG GACTCAGAACG CTGGCTTGGG GGAAGCTGGA CTGGCTTGGG GGAAGCTGGA CCGCACTTTG GCCACCTTAATTG	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGTGGA GTCATAGAG AACTTTAGAG GTCAAACCAT GTCAAACCAT GTCAAACCAT GTCAAACCAT GGCATTTGTC CTCAATACAA AGTGCAGATT ATGCCAGCTC CGCAGAACAACA AGGAAGTAAC AGGAAGTAAC AGACTGGAAG CCAACAACAA AATGTCATCA CCCAACAACAAC CCCAACAACACA CTTCGGCAGT GCTAACGGAG AGGAGGGACT	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAA ACCACAGGGG CACTTCCTGA TCAGACAACGG CACTTCCTGA TCAGACAAC GACTCACTG GACGTAAAA AGACTCAGCT GACGTAAAAG ACCGCCGTGC GGACATCCAT GCACACAGCTG CCAACAGCTC CAGGCACACC CAGGCACACC AACTCACAA CCGCCGCG CCAACACCG CCACAGCTG CAATAACCC CAGGCGACACC AAGTTCGC AAAGTGGAACC AAACCGCGCACA TGCAGCATTA GCCAGCACACA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT ACTAAAGGTT AGCAAAGGCTA ACTAAAGGTT AGAAGCAGAGT TGAGCAGAGC CAGCCTGAGA TCAAGTCTC CTTTGACCAGAGA TCCAGACTCAA AACAAACAAA AGGACACAGG GGGAAGATTCA AAGAACAACAA AAGACACAGG TCTCATTAAG ACAGCGTTGA AAAGACACAG GGGACAAGG CTACTTGGAC CAGCCAAAGG CTACTTGGACCAA	6450 6500 6550 6600 6700 6700 6850 6900 6950 7000 7150 7200 7150 7250 7350 7400 7350 7400 7350 7450 7350 7450 7350 7350 7450 7550 7550 7650 7550 7650 7550 7650 7550 75
6401 6451 6551 6601 6751 6801 6851 6801 6851 6801 6851 7001 7051 7001 7051 7001 7151 7201 7151 7201 7351 7401 7551 7551 7651 7651 7651 7651 7651 765	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG GGCAAAAATG CCTGGCTGGA GGCAACTGCCG CATCTCTCCC ATCAACATGC ACAATCGAG ACAATCTC CCTACTGCAG AGCAGTGGAA AGGCCAGTGACA AGGCCAGTGACA AGGCTGACGA CTTAAAGACT TAGTGCAGCA GCAACTGCGGC GCAACTATTA TCATGCCAAG ACACTCCAGCGA AAAGCCAAG ACACTCCAGCGA AAAGCCAAGA CTGACACTATT GAGCATGAT CAGCAGGACCA CTGGAGAATC CTGGAGAATC CTGGAGAATC	GAGGACGATT TCAASAAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTGGCAGAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGC TGGACCCAAA ACCTTTTGG GGAAGATCGG CGGAACATGG TCAGAACCTG CAAGACCAG TTGGTTGGAG AAAACTCGA GACCCACATG GACCCAAATG GACTCAGATG ATGCAGACG CTGCCCACATC	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTGTCCACT GGGATTGTC TCCAATACAA GTTGAGGAT ATGCCAGGAT ATGCCAGGAT CTCCTCGACAA AGGACGACACA AGGAAGTCTTA GAGAAGTAT AGGCAAGATG AAGACCAAA CTGCTGGAAG CCAACATCAA CGCCAGAAAAA TGACACAACA CCCACACACACA CCAACATCAA CTCCGGCAGG CCAACATCAA	AACAAGACCC TTGGAGAAAC TCGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC GATCTTCCTG TTGGAGAAGCA GATCTTCCTG GACGTCAAC GACTCACCA GACTCACCA GACTCACCA GACTCACCA GGACATCCAT GCAGTGTAA ACCGCCGTGC CCACAGCTG CCACAGCTG CCACAGCTG CCACAGCCG AAGTTTCCC AAGTGAACC AAGTGACCA ACCGCCGTCC CAGCACATC ATGCAGCACA ACCGCCCACACTC ACGCGCACATCA ACCAGCACACA CCAGGAAACCA TCCGGACACACA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGC GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCGGT ACTAAAGGT AGCAAAGCAGAG GAACAGCTCT TGAGCAGTCC CAGCCTGGACAG GACCCAGAGA TCCAAGTTCTC CTTTGACTGG GACCCAGAGA TCCAAGTCTC ATGTATAAG GGGACACGG GGGACACGG CTCCTTGACAG GGGGACAACTA AAACAAACTA AAACAAACTA AAACAAACTA AAACAAAC	6450 6500 6550 6600 6750 6850 6850 6850 6950 7000 7050 7150 7200 7150 7250 7350 7450 7450 7450 7450 7550 7450 7550 75
6401 6451 6501 6551 6601 6751 6801 6851 6801 6851 7001 7051 7001 7051 7301 7351 7301 7551 7551 7601 7551 7601 7551 7801 7951 8001 8051 8051	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ATCAACATGC CCAACATGCG CCAAACTCTC CCTACTGCAG AGCAGTGGAA AGCTTGACGA GTAATGGTGG TCACTGCGGA AAGCTTGACGA GCAATTAAA AGCATGCAGCA AAGCCAATG AGATTTAAAG AAGCCAATG AGATTCAAGAAT CTGACATCTT CAGCAGCACT CTGGAGAAAT CTTACAGAAT CTTACAGAGA	GAGGACGATT TCAA GAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCTCAGTGC AGCTGGCGAG AGCTCAGTGC AGCAGGACGCA ACTTGGGCCAA ACTCTCTCA GGACACCACA TGGACCCAAA CCCTTTTGG GGAACGCCG GGGATCTGGA CAGACCCG CTGACGGCCA TTGGTTGGAG CAGACCCC CTGACGGCCA ATGGTTGGAG GACTCAGATG AATGCAGACGA GGCCAAAATG GACCCACATGG GGAACCTGG ATGCAGACGA CTGGCTGGG GGAACCTGG CTGGCTGGG GGAACCTGA CCTCTAATG GCCACCTTA ACTCGAGCAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTCAAACCAT GTGTCACACA GTGTCACACA AGTGCAGAT CCAAAAGAGAG AGTGCAGATC ATGCCAGCAC AGGAGCAGAAGATG AGAGCTGCAAG CAGAAGTTCA GCCAGAAGATA GAGCTGCAAG CAGATGTCACA CTCCGGCAGA CCAACATCAA CTCCGGCAGC CCAACATCAA CTCCGGCAGC CCAACATCAA CTCCGGCAGC CCAACATCAA CTCCGGCAGC CCAACATCAA CTCCGGCAGC CCAACATCAA	AACAAGACACC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG GACGTCAGCT GACGTCAGCT GACGTCAGCT GCACTCAGCT GCACTCAGCT GCACTCAGCT GCACATCACA ACTGGCAGAC CCAACACCG CCAACACCG CCAACACCG CAAGCTCAGC CAAGCACACT ACTGCAGAAC GGGTGGAGCC AAAGTTCGC AAAGTTCGC AAAGTCGCC AAAGCGGCACATC CCAGCACATCA CCAGCACATCA CCAGCACATCA CCAGCACATCA CCAGCACACC TCCGGAAAACC TCCGAGCACACC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTCCATG GGAAGCGGGT CCCGGGGGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT AGCAAAGCACG GAACAGCTCT TGAGCAGTGG AAAAGAAGCC CAGCCTGAGA TCAAGTTCTC CTTTGACTGG GGCAAGACACG GGCAAGACTCA AACAAACAAA AAGGACTCAC AAGCATCAC AAAGCATCAC AAACAAACAA AAAGGATCTAC AAACAAACAA AAAGGATCTAC AAACAAACTA AAACGACTGA CAGCCCACGG CTACTGGAC CAGCCCACG GGGACAAGG CTACTGGAC	6450 6500 6550 6600 6700 6750 6850 6900 6950 7050 7100 7150 7200 7150 7200 7350 7350 7350 7350 7350 7350 7400 7550 7550 7500 7550 7500 7550 7500 7550 7500 7550 7500 7550 7500 7550 7500 7550 7500 7550 7500 7550 75000 75000 75000 75000 75000 75000 75000 75000 75000 7500000000
6401 6451 6551 6601 6751 6751 6801 6851 6801 6851 7001 7051 7101 7151 7201 7351 7301 7351 7401 7351 7401 7351 7601 7551 7601 7551 7601 7551 8001 8051 8051 8101 8151	CAGATGCTGA AGAGACCACC AGTGAAGGGA AAGACGACAG TGGGCCAAAG GCCAGGACC GGCAACTGCCG GGAACTGCCG CATCTCTCCC ACCAACATGC ACCAACTGCCG CCAACTGCCG ACCAACTGCCA AGCAGTGGAA AGCTGCAGA GTAATGGTGG CCAACTGCCAG GCAACTGCCG GCAACTATTG TCATGCCAGG CGCAACTATTA AGCCCAATG AGCTCCAGCGA AGCTTCCAGCGA CTGACTCTT CGACCATGACA CTGACACTGT CTGGAGAATT TTCGACAATT CTTGCAGAG GTCCTGGAGA	GAGGACGATT TCAASCAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTCTCTCGCAG AGCTGGTCAGTGC AGCAGGAAGCC CGCTGATAAA GCTGGGCGAA AATTTGGGCCAA AATTTGGGCC TGGACCCAAA ACCCTTTTGG GGAAAGACAAG GGGATCTGGA CAGACCGCG CTGACGGCCA CTGGTTGGAG AAAAACTCGA GGCCAAAATG GACCACATG ATGCAGACG ATGCAGACGA ATGCAGACGA ATGCAGACGA CTGGCTGGG GGAAGCTGGA TCGTTAATTG GCCACCTTTA ACTCGAGCAC	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGTGGTG AACTTTAGAG AAGTCCTTCA GTCAAACCAT GTCAAACCAT GTCAAACCAT GTCAAACCAT GTCAAACAAT CAAAAGAGAG AGTGCAGATT ATGCCAGCTC CGAGATCAAC AGGAACAAAA GAGCTGGAAG CAGATGTCAA CGCACAAAGAG CAGATGTCAA CGCACACACA CTTCGGCAGT GCCAACACCAA CTTCGGCAGT GCCAACACCAA CTTCGGCAGT GCCAACACCAA CTTCGGCAGT GCCAACACCAA CTTCGGCAGT GCCAACACCAA CTTCGGCAGT GCCAACACCAA CTTCGGCAGT GCCAACACCAA CTTCGGCAGT CACACCACAC	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAAT ATCAAGAGAC AGACATGGGG CACTTCCTGA TCAGATCAAC TGGAGAAACA GATCTTCCTG GACGTAAAAG ACACCAGCT GACGTAAAAG ACCGCCGTGC GGACATCCAT GCACATCAAA ACCAGCTCAA CCACCGCGGC CCACAGCTG CCAATAACCC CAGGCACATC ACTGCAGAAC ACTGCAGAC CAAGTTCGC AAAGTTCGC AAAGTTCGC AAAGTGGCA ACGGCGCATA CCAGGAAACC TCCAGGACATC GCAGAAACA TCCGGACATC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT ACTAAAGGTT AGCAAAGGCTA ACTAAAGGTT AGCAAAGCCGG GAACAGCTCT TGAGCAGAGA CCAGCCTGAGA TCCAGTCTG CCTGGACAGA GACCCAGAGA TCCAAACTCAA AACAAACAAA ACAAACAAACAA AGGAACACGG GGGAAGATTCA AAGGATCTAC AAAGGATCTAC AAAGGATCTAC AAAGGATCTAC AAAGGATCTAC AAAGGACCAGG CTACTTGGAC CAGCCAAAGG CTACTTGGAC CAGCCAAAGT	6450 6500 6550 6600 6700 6700 6850 6900 6950 7000 7150 7150 7200 7150 7250 7350 7400 7350 7450 7350 7450 7350 7450 7350 7350 7450 7350 7550 7650 7650 7650 7550 7650 7650 7550 7650 7550 75
6401 6451 6501 6551 6601 6751 6701 6751 6801 6851 6901 6851 7001 7051 7001 7251 7301 7251 7301 7251 7301 7351 7551 7601 7551 7601 7751 7801 7751 8001 8051 8001 8051 8101 8201	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG CCAAACTCTC CCTACTGCAG AGCAGTGGAA AGGCCAGTG TTAAAGACT TAGTGCAGCA GTAATGGTGG GCAACTATTA TCATGCCAAG ACACTGCGGC AGCATGACAG ACACTGCAGCA AAAGCCAATG AGCATGACA CTGACACTAT AGGCAGGACA CTGACTCTTT GAGCATGAT CTGCAGAAT CTTACAAGAT TCCGGCAGA GTCTCTGGGG CACTGCGGGC ACATGGGGCA	GAGGACGATT TCAA GGACCACATT TGCCAACCA CTCATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCTGATAAA GCTGGCGAA AACTTCCCTA AGACCCCTTTGG GGAACGCCCAA ACCCTTTTGG GGAACGCCG TGGTCGGC GGAACGCCG CAGACCGC CAGACCGC CAGACCGC CAGACCGC AAAAACTCGA AAAAACTGG ACACAACG GGCAAAATG GGCAACATG GGCACCATTA GCCACCTTA ACTCGAGCAA CGGAACGGCAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGGCTG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GTGTCCACT GGGATTGTC TCCAATACAA GTTGAGGAT ATGCCAGGAT ATGCCAGGAT CTCCTCGACAA CTGCTGGACA AGGACACAACA CTGCTGGACA CAGAAGATG AAGAACAAAA TGAGAAGTTA GACACAACA GCCAGAACATAA CGCCAGAACA CTCCGGCAGG CCAACATCACA CCAACATCAA CTTCGGCAGG CCAACATCAA CTTCGGCAGG CCAACATCAA AATGTCACTA AAGGAGGGACT TGGCAGGACA AGGAGGGACT ATTAGATGAA	AACAAGACCC TTGGAGAAAC TCGAGTCCA TCGACTCCA ATCAAGAGAAC AGACATGGGG CACTTCCTGT TCAGATCAAC GAACTCACT GGAGAAGCA GATCTTCCTG TTGTACCAA AGACTCACCT GACGTAAAG ACCGCCGTGC GGACATCCAT GCAGTGTAA ACATGGTTCA CCACAGCTG CCAACAGCTG CCAACAGCTG CCAACAGCTC AAGTGAACC AAGTGACCC AAGTGACCC AAGTTTCCC CAGCTGCACACT ACGCGCCATA GCCAGCACTA GCAGAAACCA TCTGGAAAAT TCCAGGCACAC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT AGCAAAGCACTCT TGACCAGCAGTCG CAACACCTCG CAGCCTGAGA CAAGTTCTCC CTTTGACTGG GCACACAGG GGGAAGATCCA AACAAACTAA AACAAACTAA AAGGACCAGG CTCCTTTAAAGG TCCTCATTAAGG TCTCATTAAGG CTCCTATAAGG CTCCTGCACAG GGGACAACTA AAAAACAAACTA AAAGAACTA AAAAGCAACTA CAACCCACG CTACTTGGAC CAGCCACGG CTACTGGACCAG GATTGATGATG CTGCCCCACG GATTGATGATG	6450 6500 6500 6650 6650 6700 6850 6850 6850 6950 7000 7050 7150 7200 7150 7250 7350 7450 7450 7450 7450 7450 7550 7450 7550 7450 7550 75
6401 6451 6501 6551 6601 6751 6801 6851 6801 6851 7001 7051 7001 7051 7301 7351 7301 7551 7551 7601 7551 7601 7551 7801 7851 7901 7951 8001 8051 8051 8101 8151 8051	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGAGACGACAG TGGGCCAAAG GGCAAAAATG GCCAGGCTGGC GGAACTGCCG CATCTCTCCC ATCAACATGC CCAACATGC CCAACTCTC CCTACTGCAG AGCAGTGGAA AGCAGTGGAA AGCTTGACGA GTAATGGTGG GCAACTGCGGG AGATTTAAA AGCATGCAGCA AGATTCAAGAT TCATGCCAGA AAGCCAATG AGATTAAAA AGCATGCGAAT CTGACACTTT CTGGAGAAT CTTACAGGAAT CTTACAGGAGA	GAGGACGATT TCAA GAGCT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCTGGCGAG AGCTCAGTGC AGCAGGAAGC CGCTGATAAA GCTGGCCAAA ACTCTCTCA TGGACCCAAA ACCCTTTTGG GGAACGCCG GGGATCTGGA GGAACACGG CTGACGGCCA TTGGTTGGAG CAGACCGCA ATGCTGGGG GACCCAAATG GACCACATGG GGACCCAAATG GGACCCAAATG GGACCCAAATG GGACCCAAATG GCACCCTTAATG GCACCCTTAATG GCACCCTTAA ACTCGAGACA CTGGCTGGG GAACGACAA CGTACCATCC GGAACGGCAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGCTG AACTTTAGAA AACCTCTTCA GTCAAACCAT GTCAAACCAT GTCAAACCAT GTGTCACACA AGTGCAGATG CCAAAACAAG AGTGCAGATC CTCCCGACAAC AGGACCAAAGATG AAGAACAAAA GAGTTCTTAC GCCAGAACATAA GAGCTGCAGAG CCAACATCAA CTCCGGCAGT CCAACATCAA CTCCGGCAGT CCAACATCAA CTCCGGCAGT CCAACATCAA CTCCGGCAGT CCAACATCAA CTCCGGCAGT CCAACATCAA CTCCGGCAGT CTAACGGGACT TGGCAGCACACA ATTAGATGAA ATGCTGTGGTG	AACAAGACACC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT TCAGATCAAC TGGAGAAGCA GATCTTCCTG GACGTCAAC TGGAGAAGCA GACGTCAGCT GACGTAAAAG ACCGCCGTGC GGACATCCAT CCAGCACACT CAGGTGAGAC CCAACACTTC ACTGCAGAAC GGGTGGAGCC AAAGTTCGC AAAGTTCGC AAAGCAGCA CCAGCACATA CCAGCACATC ACGCAGCATT ACGAGCAGCA ACAGGCACAT CCAGGCACAT CCAGGCACAT CCAGGCACAT CCAGGCACAT CCAGGCACAT CCCAGCACAT CCAGGCACAC CCACACCAC CCACACCTCC CCACACCAC CAACTTCCC CCACACCACAC CCACACCACAC CCACACCACAC CCACACACC CAACTTCCC CAACCACAC CCACACACA	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GATTTCACT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT AGCAAAGCTCT TGAGCAGTCG AAAAGAAGCCC CAGCCTGAGA TCAAGTTCTC CTTTGACTGG CCTGGACAGG GACCCAGAGA TCCAAGTTCTC ATGTATAAGG TCTCATTAAG ACAGCGTTGA AAAGGATCTAC AAACAACCAA AAAGGATCTAC AAACAACCAA CAGCCCACAG GGGACAAGG CTACTTGGAC CAGCCAAGG CTACTGGCC CAGCCCACAG GATTGATGCT GAGTATAAGC	6450 6500 6500 6650 6700 6750 6850 6900 6950 7050 7100 7150 7200 7150 7200 7350 7350 7350 7350 7350 7350 7350 73
6401 6451 6551 6601 6551 6601 6751 6801 6851 6801 6851 7001 7051 7001 7051 7101 7251 7001 7351 7301 7351 7401 7351 7601 7551 7601 7551 7601 7551 7601 7551 8001 8051 8001 8101 8151 8101 8151	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGAGACGACAG TGGGCCAAAG GCCAAGGACAG CCTGGCTGGA CGCAAGGAAC GGAACTGCCG CATCTCTCCC ACCAACTGCC ACCAACTGCC CCTACTGCAG AGCAGTGGAA AGCTGCAGA AGCTGCAGA GTAATGGTGG CCAACTATT TAGTGCAGCA GCAACTATTA AGCACACTGG CAACTATTA AGCACATGG ACTCCAGCGA ACTCCAGCGA ACTCCAGCGA ACTCCAGCAAT CTGACATTT CAGCCATGACAAT CTGACATTT CACGCGGACAAT CTGACATTT CACCGGGACA	GAGGACGATT TCAASCAGGAT CTCAATGCTC GGACCACATT TGTCCAACCA CTCTCTCGCAG AGCTGCAGTGC AGCAGGAAGCC CGCTGATAAA GCTGGGCGAA GATCTCCTCA AATTTGGGC TGGACGCCTG GGAAGACCAG CGGATCTGGA CAGGACCGGA CTGGTCGGGC CTGACGGCCA AAAACTCGA GGGATCTGGA GAACGGCCA ATGCTGGGGA AAAACTCGA GGCCAAAATG GACCGCCACTTA ACTCGAGCAG CGGACGGCAA ACCCTTAATTG GCCACCTTA ACTCGAGCAC GGAACGGCA CGGACGGCAT ACCCGGCCA CTGACGGCA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGC GCTCACTGAG AGTTAAATGA GCTGTGGTGG GACTTTAGAG AACTTTAGAG GTCAAACCAT GTCAAACCAT GTCAAACCAT GTCAAACCAT GTCAAACAT CCAAAAGAGAG AGTGCAGATT ATGCCAGCTC CGAGATCATAC AGGAACAAA GAGCTGGAAG CAGATGTCAA CGCACACACA CGCACACACA CGCACACACA CTTCGGCAGT GCCAGCACCA CGCACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTTCGGCAGG CCAACACACA CTCCGCCCC CCACCACCCC CCACCACCCC CCACCACCCC CCACCA	AACAAGACCC TTGGAGAAAC TCTGAGTCCA TCGACTCCAGT ATCAAGAGAC AGACATGGGG CACTTCCTAG TCAGATCAAC TGGAGAAACA GATCTTCCTG TTGGACAAC GACGTAAAAG ACCCCCGTGC GGACATCCAT GCACTCACT GCACTCACT GCACTCACT CCACAGCTCA CCACAGCTC CAATAACCC CAATAACCC CAATAACCC AAGTTTCG AAAGTTCGC AAAGTTCGC AAAGTTCGC AAAGTGACAC TCCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCAGGACATC CCACAGCAC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTTCAAT GGAAGCGGGT ACTAAAGGTT AGCAAAGGCGT ACTAAAGGTT AGCAAAGGCTC TGAGCAGAGG GAACAGCTCT CTGGACCAGAGA TCAAGTTCTC CTTGGACGG GGCACAGAGA TCCAAGTCTCA AACAAACAAA ACAAACAAACAAA TGAGCACAGG GGGAAGATTC ATGTATAAGG CTCCTGCACAGG CTACTTGGAC CAGCCAAGG CTACTTGGAC CAGCCAAGG CTACTTGGAC CAGCCAAGG GATTGATGTC CGGCTTGATG GAGTATAAGA	6450 6500 6500 6650 6700 6700 6850 6900 6950 7000 7150 7150 7200 7150 7250 7350 7400 7350 7450 7350 7450 7350 7450 7350 7450 7350 7450 7550 7450 7550 7450 7550 7450 7550 75
6401 6451 6501 6551 6601 6751 6801 6851 6801 6851 7001 7051 7001 7051 7001 7251 7301 7351 7501 7551 7601 7551 7601 7751 7801 7751 7801 7751 8001 8101 8101 8101 8251 8201 8201	CAGATGCTGA AGAGACCACC AGTGAAGGGA AGAGACGACAG TGGGCCAAAG GGCAAAAATG CCTGGCTGGA GGAACTGCCG CATCTCTCCC ATCAACATGC ACAAATCGAG CCAAACTCTC CCTACTGCAG AGCAGTGAGA AGGCCAGTG AGCAGTGACA GTAATGGCGG TCACTGCGGG CCAACTATTA TCATGCCAGA AAAGCCAATG ACAGCAGGGCA AGCATGACAT AGGCAGGAGAAT CTTACAAGAT TCGACACTTT GAGCATGAAT AGGCAGGACT CTGCAGGACA CTTACAAGAT TCGACACTTT CAGGCAGAAT CTTACAAGAT TCGACACTTT CAGCAGGGCA ACATGGGGCA ACATGGGGCA	GAGGACGATT TCAA GAGCT TCAA GACT CTCAATGCTC GGACCACATT TGTCCAACCA CTTCTCGCAG AGCTCAGTGC AGCTGATAAA GCTGGCGAA AGCTCCATAA GATCTCCTCA TGGACCCAAA ACCCTTTTGG GGAACGCCG GGAATGCGG GGAATGCGG CAGACCGC CTGACGGCCA TTGGTTGGAG GACACCGC CTGACGGCCA AAAAACTCGA GGCAAAATG GGCAACATGG GGAAGCTGGA AAAAACTCGA GGCAAAATG GGCAACATGG GGAAGCTGGA ATGCAGACGA CTGGCTTGGA GGCACCTTTA ACTCGAGCAA CGCACCTTTA ACTCGAGCAA	ATGAAGTATG TCAGGCGGGT TTGGGACTGA AAACAGCAGCG GCTCACTGAG AGTTAAATGA GCTGTGGCTG GCTACTGAG AACTTTAGAA AAGCTCTTCA GTCAAACCAT GGGATTGTC TCCAATACCAA GTGTGAGAT CAAAAGAGAG AGTGCAGAT ATGCCAGCAC CTCTCGACAA CTGCTGGATC CGAGATCAACA AGGATCATAC GAGATCATAC GAGACGAAA CTCCGGAGAC CAACATCAA ATGCCACACA CTCCGGAGC CCAACATCAA CTTCGGCAGC CCAACATCAA ATGCTGTGGA AGGAGGGACT TGGCACCACA CTTCGGCAGC CAACATCAA ATGCTGTGGA AGGACGGAC AGGACGGAC CAACATCAA CTTCGGCAGC CCACATCAA ATGCTGTGGT GAGTTGTGCA TGATCAGTGG TGACCAGTGAA	AACAAGACACC TTGGAGAAAC TCTGAGTCCA TCGACTCCAT ATCAAGAGAC AGACATGGGG CACTTCCTGT AAAGTCAAC TGGAGAAGCA GATCTTCCTG GACGTCAAC GACGTCAGCT GACGTCAGCT GACGTCAGCA GACATCCAC GGACATCCAT GCAGTGTCA GAAATGGTCA GAAATGGTCA GGAGACATCC CAAGCACACC CAAGCACACC AAAGTTCCG AAAGTTCCG AAAGTTCCG AAAGTTCCC ATGGACAACC CAGGACAACC ACGGACATC ATCGGACAAC ACCGGCACAT TCCGGAAACC CCAGGACACC CCAGGACACC CCAGGCACAC CAGGACACC CAGGCACACC CAGGCACAC CCAGGACACC CCAGGACACC CCAGGCACAC TCCGGAAACC CCCACACCC GCACACCC CAGCACACC CAGCCACACC CAGCCACACC CAGCACACC CAGCACACC CAGCACCC CAGCACACC CCAGCACACC CCCACACCC CAGCACCC AAAGTCCC CCACACCC CCCACACCC CCCCCCCCC CCCCCCC	CATAAATAAT AGGAAGCAGC CAGTGCAGTA CAACTCCCGG GAGTTCAAT GGAAGCGGGT CCCGGGTGGT AGCAAAGGCT AGCAAAGGCT AGCAAAGGCT AGCAAAGCACTCT TGACCAGCAGTCG AAAAGAAGCAC CAGCCTGAGA GAACAGCTCTC CTTTGACTGG GCAAAGTCTCA CAAACTCAA AACAAACAAA AACAAACAAA CAAGCGTTGA CAGCGTTGAC GGGAAAGCTCA CAACCCACGA CAACCCACGA CAACCCACGA CACCCACGAC CACCCACAGG CTCATTAAGG CTCATTAAGG CTCACTACA AAACAAACAA AAAAACAAACAA AAAGGACCCG GGGACAAGG CTACTTGGAC CAGCCCACGG GATTGATGCTC GATTGATGCT GAGTATAACCA AACTACCCACG GATTGATGCT CCGGCTCATG GAGTATAACCA	6450 6500 6500 6650 6700 6700 6850 6900 6950 7050 7100 7150 7150 7250 7300 7350 7350 7350 7350 7350 7350 73

Figure 2. Continued.

8401	. CACAA <mark>GG</mark> CTT	TCAGAAGAGA	GCTGGGTGCC	AAAGAACCTG	TAATCAATGG	8450
8451	AACTCTTGAC	AATGCGAAAA	CCTTCCTCGC	TGAGATGCCT	CGTGAGGGCC	8500
9501	TCADACACAC	ACCTCCACAA	AAATOTCTAA	CTCCTCACCA	CCCCCTTCAA	9550
0501	. IGAMACAGAG	ACCIGGACAA	AAA	GICCIGAGGA	GCGGGTTCAA	0000
8221	. AACGTGGGGCC	GCATATTGCG	TAAAGAAGTG	GAGGATGTAA	CGGTGCGATG	8600
8601	GAAGAATCTG	GGTGCAGCTT	CTGTGGACTG	GCAGCAACAG	CTGGAACTGG	8650
8651	CCTTCCACAC	CCTCATCCAC	CTCCACCACC	CCCACCATCA	CCTCCACTAC	8700
0001		COLONIONO	CICCHOONCO	TOTTOONICA	oronoine	0750
8101	. AAGCTACGAC	AGACTGAGAG	CGTGAAGAAT	TCTTGGAAGC	CTGTTGGGGA	8750
8751	GCTGCTTGTA	GACGATTTGC	AAAACCACAT	AGACAGAGTG	AAGCATTTC	8800
9901	ACCAACACAT	ACCTCCCATT	CACCATAACC	TAAATCATCT	CAACCACCTC	9950
0001	. AGGAAGAGAI	AGCICCCATI	CAGGAIAACG	INANICAIGI	CHACCAGEIG	0000
8851	. GCCTCCACAT	TTAGACCGTC	TGACATTCAG	CTTTCTCCAG	ACAACCTGAG	8900
8901	CAGAATCGAT	GACCTCAACA	TGAGATGGAG	GCTTCTGCAG	ATCTCCATTG	8950
9951	ACCACCATCT	CACTCACCTC	ACCACACCCT	TTAACCACTT	CCCACCCTCT	9000
0551	AGGAGCAICI	GAGICAGCIG	ACCACAGCCI	TIANGGACII	GGGACCCICI	5000
9001	. CAGAATTTCC	TTCATGGTC	TGTTGAAAGC	CCTCTCGAAC	GCTCTATTTC	9050
9051	ACCCAACAAT	GTCCCCTACT	ATATCAATCA	CCAGACCCAA	ACAACATGTT	9100
0101	CCCACCACCC	AAACATCCCA	CAACTOTACC	ACTCATTACC	CCATCTCAAC	0150
9101	. GGGACCACCC	AAAGAIGGCA	GAACICIACC	AGICATIAGE	GGAICICAAC	9130
9151	. AACGTGCGGT	TCTCGGCATA	CAGGACGGCA	ATGAAGCTCA	GACGAATGCA	9200
9201	GAAAGCCCTC	TGTTTGGATC	TTCTGAGCAT	GCCTGCAGCC	TGTGAAGCCT	9250
0251	TTCACCACCA	CAATCTCAAA	CACAACCACC	ACTTCATCCA	CATCOTCOAC	0200
9251	. IIGAGCAGCA	CAAICICAAA	CAGAACGAGC	AGIICAIGGA	CATCGIGCAG	9300
9301	GTGATCAACT	GTCTGACCAG	CATCTACGAC	CGTCTGGAGC	AGCAGCACAG	9350
9351	CAGCCTGGTC	AACGTGCCTC	TCTGTGTGGGA	CATGTGTCTC	AACTGGCTGC	9400
0401	TONNOCTOTIO	CONTRACTOR OF	00101010000	ACA TROCTAC	oomamoommo	0450
9401	. ICAACGIIIA	CGATACAGGA	CGAGCIGGGA	AGAIICGIAC	CUTAICUTIC	9450
9451	AAAACAGGAA	TAATCTCTTT	GTGCAAAGCT	CACCTTGAAG	ATAAGTACAG	9500
9501	ATTTTTTTTTTT	CGAGAGGTGG	CCAGTGCCAC	AGGCTTCTGT	GACCAGCGGC	9550
0551	COOTOCOCOT	COTOCTOCAT	CATCOCATTO	ACATCOCCAC	COLCOTOCOT	0000
9551	. GCCICGGCCI	CUTUUIGCAT	GAIGCCAIIC	AGAICCCCAG	GCAGCIGGGI	9600
9601	GAAGTGGCGT	CTTTCGGAGG	GAGCAATATT	GAGCCCAGTG	TGCGCAGCTG	9650
9651	CTTTCAGTT	CCAATAACA	AACCGGAGTT	AGAGGCTTCA	GTCTTCCTGG	9700
0701	ACTOCATOOO	TTTACALOOT	CACTOCATCC	TTTCCCTTCC	TOTOOTTOLO	0750
9701	ACIGGAIGCG	TTTAGAACCT	CAGICGAIGG	TIIGGCIICC	IGICCLICAC	9750
9751	CGTGTAGCGG	CCGCTGAGAC	AGCAAAGCAC	CAGGCTAAAT	GCAACATTTG	9800
9801	TAAGGAATGT	CCTATTATTG	GCTTCA COTA	CCGAAGTTTA	AAGCACTTTA	9850
0051	. Innoonnioi	COMMINIO	Teamone	orocololom	ANOCACITIA	0000
9851	ACTATGATAT	CTGCCAAAGC	TGCTTCTTTT	CTGGCAGAGT	GGCCAAAGGT	9900
9901	CACAAGATGC	AGTACCCTAT	GGTTGAATAT	TGCACACCGA	CGACGTCAGG	9950
9951	AGAGGATGTG	AGAGACTTTC	CCAAGGTGTT	DAACDACAAC	TTCAGGACAA	10000
10001	AGAGGATIGIG	ROADACITIO	CCARGOIGII	AMAGAACAAG	COTOCH CAR	10000
10001	AGCGCTATTT	TGCCAAGCAC	CCTCGCATGG	GTTACCTTCC	CGTCCAGACC	10020
10051	ATCCTTGAAG	GAGACAACAT	GGAGACTCCT	GTTACACTGA	TCAACTTCTG	10100
10101	GCCTGTTGAC	CATCORCETC	CATCTTCCCC	TCAACTCTCC	CATGACGACA	10150
10101	COCTOTIONC	CATOCOLO	TRACOTRACOL	TOPINOTOTOC	CATORCORON	10100
10151	. CCCACICICG	CATCGAGCAT	TACGCTAGCA	GATGATGA	GCATCIGCIG	10200
10201	ATCCAGCACT	ACTGCCAGAG	TCTGAATCAA	GGCTCTCCTC	TCAGCCAGCC	10250
10251	GCAGAGCCCC	GCTCAGATCC	TCATCTCAAT	GGAGACTGAA	GAGAAGGGAG	10300
10201	LOOTOLOLOLO	1 or o ano 1 h	Charlos Color	LOODDADAG		10000
10301	AGCTGGAGAG	AGTGCTCAAT	GATCIGGAGC	AGGAAAACA	AAGCTGCAA	10350
10351	GCGGAGTACG	ATCGTCTGAA	AAAGGCACAC	GATCACAAGG	GTCTGTCACC	10400
10401	GTTGCCTTCA	CCTCCACAGA	TGCTTCCGGT	GTCGCCTCAA	AGTCCACGCG	10450
10451	1000101107	TAROOGLOSS	0001100001	TCCCCCICIT	all accil acc	10100
10451	ACGCAGAACT	AATCGCAGAA	GCCAAACTAC	TGCGGCAACA	CAAGGGACGA	10200
10501	TTGGAGGCAA	GAATGCAAAT	CCTGGAGGAT	CACAACAAGC	AGCTGGAATC	10550
10551	GCAGCTTACA	CCCCTAACAC	AGTTACTCGA	GCACACTGAG	TCCAAGGTGA	10600
10551	GUNGCITACA	COCCIANDAC	AGIIACICGA	GCACHCIGAG	ICCARGOIGA	10000
10601	ATGGCACTGC	CCTGTCCTCA	CCCTCCACTG	CCTCTCCGAG	ATCTGACACC	10650
10651	AGCCTGGCCT	CACTGCGTGT	GGCCGCAAGC	CAAACCACAG	AGACCATG	10700
10701	TCATCATCAC	OTOTOOLOTO	0020002002	TCCARCOACT	CCATTACAAC	10750
10701	IGAIGAIGAG	CIGICCAGIC	CCACCCAGGA	IGCANGCACI	GGATIAGAAG	10750
10751	ACGTCATCGA	GCAGCTCAAC	AACTCCTTCC	CTCACAGCCA	AGGCGGA	10800
10801	CGGCTAAATC	CATGAGAGGT	CCCAATGTGG	GAAGTCTTTT	CCACATGGTT	10850
10951	CATAACATTC	CACATCCCAT	CONCTOCTO	CTTCATCTAA	TRACACACCA	10000
10031	GATAACATIG	GACAIGUCAI	GGHGICGCIG	GITCAIGIAA	INACAGAGGA	10300
10901	GCAGGATTTA	GACTGAAGAC	GTCTTCTCCT	CGTTGCATGC	TTTTGTAGTG	10950
10951	TCAACAACTG	GACCGGATAT	GTTTACAATG	GGGAATATCA	ATAAAAATCT	11000
11001	ATTTTTOTCA	1003				11014
11001	AITITOIGA	AGGA				11014
1	MAEAVNPQDQ	WEEGFEDEFG	EIIKTRSDER	EDVQKKTFTK	WINSQFAKTR	50
51	RPPIDDLFTD	LCDGRRLLEL	LEVLVGHEIV	KERGFTRVHS	LNNVNRALOI	100
101	LOKNNUDI UN	TEGADTUDON	HKLTLGITMC	TTLHMOURDU	MKDUMADI OO	150
101	PSULLIN DIAM	LOONDIVDON	TTTTTTTTTM0	TTHINKARDA	In the standing of	100
151	INSEKILLSW	VRQSLKNYQD	VNVVNFSSSW	ADGFAFNALI	HSHRPELFSW	200
201	SVVEQODNAI	ERLDHAFGVA	EKSLGIDRLL	DPEDVATVHP	DKKSIIMYVT	250
251	SLEKVIDHOU	SLEATOEVET	LPRAAUTKEE	HULYOTOOPY	SOOTTUSUAO	300
2.51	SHERVERNEV	CHARTENEL	PE INTRA TUDE	mini Ki Kouli	PARTIADAHO	000
301	SKVRSPSPSY	RUPPECVART	A 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	ODVER	DKADELRPSP	350
351		KERERSTRET	QAAYVKTPEQ	QRKFLIAQSP		
401	SPLPQGLNEL	ESYOSALEEV	QAAYVKTPEQ	QRKFLIAQSP LOAOPPISSF	VEEVKEOFHT	400
401	SPLPQGLNEL	ESYQSALEEV	QAAYVKTPEQ LTWLLSAEDG	QRKFLIAQSP LQAQPPISSF	VEEVKEQFHT	400
1	SPLPQGLNEL HEGYMVELTS	ESYQSALEEV HQGSVGRVLK	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ	QRKFLIAQSP LQAQPPISSF LTDDEEREVR	VEEVKEQFHT EQMNLLNSRW	400 450
451	SPLPQGLNEL HEGYMVELTS EHLRVASMER	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP	400 450 500
451 501	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKROIF	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLOEDLE	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEOVRVNSLT	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALFF	400 450 500 550
451 501	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE	400 450 500 550
451 501 551	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA	HYDRERSIAET HYDRER	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL	400 450 500 550 600
451 501 551 601	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI	400 450 500 550 600 650
451 501 551 601 651	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLF	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAORW	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVT	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SOSEITHTTM	400 450 500 550 600 650 700
451 501 551 601 651	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVT	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTNI	400 450 500 550 600 650 700
451 501 551 601 651 701	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG	QAAVVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH	400 450 550 600 650 700 750
451 501 551 601 651 701 751	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK	QAAVVKTPEQ LTWLLSAEDG AGSVLLSAEDG LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR	400 450 500 550 600 650 700 750 800
451 501 551 601 701 751	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAOALVDOLT	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK	QAAVVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEA	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRKE VLAIDREKRE	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR	400 450 500 550 600 650 700 750 800 850
451 501 551 601 651 701 751 801	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV	400 450 500 550 600 650 700 750 800 850
451 501 551 601 651 701 751 801 851	SPLPQGLNEL HEGYMVELTS EHLRVASMES ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA	400 450 550 600 650 700 750 800 850 900
451 501 551 601 651 701 751 801 851 901	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL LAPVNLFQL	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW OLEELROKEF	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFDEHYHOVI	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTX CRDEISRLSA EDLFARFEOL	400 450 500 550 600 650 700 750 800 850 900 950
451 501 601 651 701 751 801 851 901	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL LQPQVAKLHE	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL	400 450 500 550 600 650 700 750 800 850 950
451 501 551 601 651 701 751 801 851 901 951	SPLPQGLNEL HEGYMVELTS EHLDVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL LQPQVAKLHE VLVQSSLPPA	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE RYKDVMAALL	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI	400 450 550 600 650 700 750 800 850 900 950 1000
451 501 551 601 651 701 751 801 851 901 951 1001	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL LQPQVAKLHE VLVQSSLPPA KAIQASQKEH	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE RYKDVMAALL QGDVDDLNKM	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL AEQVFQKAPP	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY EICQKYRTEL	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI DNVMVRWRRI	400 450 550 600 650 700 750 800 850 900 950 1000 1050
451 501 551 601 651 701 751 801 851 901 951 1001	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL LQPQVAKLHE VLVQSSLPPA KAIQASQKEN	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE RYKDVMAALL QGDVDDLNKM	QAAVVKTPEQ LITWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL AEQVFQKAPP FONDTKTLOF	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY EICQKYRTEL	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI DNVMVWRRI DNVMVWRRI	400 450 550 600 650 700 750 800 850 950 1000 1050
451 501 551 601 651 701 751 801 851 901 951 1001 1051	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVQQL LAFYNLFQQL LQPQVAKLHE VLVQSSLPPA KAIQASQKEH SEQLEENIQK	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE RYKDVMAALL QGDVDDLNKM LQDHMTKLQQ	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL AEQVFQKAPP FQNDTKTLQK	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY EICQKYRTEL WMAEVDVFLN	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI DNVMVRWRI EEWPALGDAE	400 450 550 600 650 700 750 800 850 900 950 1000 1050 1100
451 501 551 601 651 701 751 801 851 901 951 1001 1051 1101	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL LQPQVAKLHE VLVQSSLPPA KAIQASQKEH SEQLEENIQK ALEKQLEQCT	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE RYKDVMAALL QGDVDDLNKM LQDHMTKLQQ ALVNDIHTVQ	QAAYVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL AEQVFQKAPP FQNDTKTLQK PSLNGINEVG	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY EICQKYRTEL WMAEVDVFLN LALKREAETP	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI DNVMVRWRRI EEWPALGDAE FAIKIQKMLD	400 450 550 600 650 700 750 800 950 1000 1050 1100 1150
451 501 551 601 651 701 751 801 851 901 951 1001 1051 1101 1151	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL LQPQVAKLHE VLVQSSLPPA KAIQASQKEH SEQLEENIQK ALEKQLEQCT ELNAQWELIC	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE RYKDVMAALL QGDVDDLNKM LQDHMTKLQQ ALVNDIHTVQ KQAYAKKSAL	QAAYVKTPEQ LITWLLSAEDG AGSVLLSAGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL AEQVFQKAPP FQNDTKTLQK PSLNGINEVG KGGLDMTVSL	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY EICQKYRTEL WMAEVDVFLN LALKREAETP RKEMOEMOEW	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI DNVMVRWRI EWPALGDAE FAIKIQKMLD ITQAEEEYLE	400 450 550 600 650 700 750 800 850 900 950 1000 1000 1100 1150 1200
451 501 551 601 651 701 801 851 901 951 1001 1051 1101 1151	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL LQPQVAKLHE VLVQSSLPPA KAIQASQKEH SEQLEENIQK ALEKQLEQCT ELNAQWELDE	ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE RYKDVMAALL QGDVDDLNKM LQDHMTKLQQ ALVNDIHTVQ LHKAVEFI	QAAVVKTPEQ LTWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL AEQVFQKAPP FQNDTKTLQK PSLNGINEVG KGGLDMTVSL	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY EICQKYRTEL WMAEVDVFLN LALKREAETP RKEMQEMQEW	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI DNVMVRWRRI EEWPALGDAE FAIKIQKMLD ITQAEEEYLE	400 450 550 600 650 700 800 850 900 950 1000 1050 1100 1150 1200
451 501 551 601 701 751 801 851 901 951 1001 1051 1101 1151 1201	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQQL LQPQVAKLHE VLVQSSLPPA KAIQASQKEH SEQLEENIQK ALEKQLEQCT ELNAQWELIC RDFQYKTPEE	KENENSIAFI ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE RYKDVMAALL QGDVDDLNKM LQDHMTKLQQ ALVNDIHTVQ KQAYAKKSAL LHKAVEELKR	QAAVVKTPEQ LITWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL AEQVFQKAPP FQNDTKTLQK PSLNGINEVG KGGLDMTVSL AQEEVHQKET	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY EICQKYRTEL WMAEVDVFLN LALKREAETP RKEMQEMQEW KVKLLTDKVT	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI DNVMVRWRRI EEWPALGDAE FAIKIQKMLD ITQAEEEYLE NFISKAPPAA	400 450 500 650 650 750 800 850 900 950 1000 1050 1100 1150 1200
451 501 551 601 701 801 851 901 951 1001 1001 1101 1151 1201 1251	SPLPQGLNEL HEGYMVELTS EHLRVASMER KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVQQLT LAFYNLFQQL LQPQVAKLHE VLVQSSLPPA KAIQASQKEH SEQLEENIQK ALEKQLEQCT ELNAQWELIC RDFQYKTPEE HDALKAELDV	KENKISIAFI ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW DPNEVAANLR AKLERFAQRW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ QLEELRQKEE RYKDVMAALL QGDVDDLNKM LQDHMTKLQQ ALVNDIHTVQ LLKAVEELKR LTSNYQRLCS	QAAVVKTPEQ LITWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL AEQVFQKAPP FQNDTKTLQK PSLNGINEVG KGGLDMTVSL AQEEVHQKET RLDGKCKTLE	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY EICQKYRTEL WMAEVDVFLN LALKREAETP RKEMQEMQEW KVKLLTDKVT	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI DNVMVRWRRI EEWPALGDAE FAIKIQKMLD ITQAEEEYLE NFISKAPPAA SYLELENAWM	400 450 500 550 600 700 750 800 850 900 1000 1050 1100 1150 1200 1300
451 501 551 601 651 701 751 801 851 901 901 1001 1051 1101 1151 1201 1201	SPLPQGLNEL HEGYMVELTS EHLRVASMER ELDDIKRQIE KLQNLGERWA VQSIQSSGTN KSKEAAGKLE ATVTKVTTNQ SFMTRSEAVL SAQALVDQLT LAFYNLFQU LQPQVAKLHE VLVQSSLPPA KAIQASQKEH SEQLEENIQK ALEKQLEQCT ELNAQWELIC RDFQYKTPE HDALKAELDV DLLEKKLDET	HENRIGIATI ESYQSALEEV HQGSVGRVLK QSRLHEVLMD EQKLLQEDLE AICKWTEERW KKMVKHTKEG QNPEFSISRK SDGQNSEDIQ EQAAATAENW QLEELRQKEE RYKDVMAALL QGDVDDLNKM LQDHMTKLQQ ALVNDIHTVQ KQAYAKKSAL LHKAVEELKR LTSNYQRLCS EGLOGGIFET	QAAVVKTPEQ LITWLLSAEDG AGSVLLSGGQ LQHQQLKQLS LEQVRVNSLT ILLQKILLCW KLAILKADLE DKLVQALQLT MSTPPPQKKR EGSVADLYEK QAAQQLRARW LKVQSPPACE TPVLFDADIS AWLQQCENKL AEQVFQKAPP FQNDTKTLQK PSLNGINEVG KGGLDMTVSL AQEEVHQKET RLDGKCKTLE EEALTSJDTM	QRKFLIAQSP LQAQPPISSF LTDDEEREVR DWLDTTETRI HMVVVVDENS QHFSEEQLLF LKRQTMDKLC STKISTIVTT QIVVDSELRK VLAIDREKPE VDFCALLAER PEPLRIQLER AFQEHYHQVL AIPSTAVTEY EICQKYRTEL WMAEVDVFLN LALKREAETP RKEMQEMQEW KVKLLTDKVT EVWACWCELL IREHPEYNRM	VEEVKEQFHT EQMNLLNSRW KRMGAQALGP GDGATAALEE DSWLTQKEEL SLVQDLLTNI SQSEITHTTM RFDVDFTEIH KFRKLQEATR LAWLAYQTKV CRDEISRLSA EDLRARERQL PVMEQRLKDI DNVMVRWRRI EEWPALGDAE FAIKIQKMLD ITQAEEEYLE NFISKAPPAA SYLELEANTIM	400 450 550 600 700 750 800 850 950 1000 1050 1100 1150 1200 1350 1350

в

1351 DGRVLDELIH KKVEDYNTRW DELMQRASQR RQQLEKSLHW AQENDKTLRL 1400 1401 IQDSLNTTDR HLTAYIADGI DAAQIPQEAQ KIQTELNGHE VTLDDMKKKA 1450 343

1451	MEVDASEKVI	GEIDATLDKL	LQVKGKFRLF	QKPANFDQRL	RECERVLEEV	1500
1501	KVKLGVLSIR	SVEQEVVQSQ	LEQCMKFYKN	LSEVKSEVET	VIKTGRQIVQ	1550
1551	KQQTEQPKEL	DDRLTALKLA	YNDLGSQVTE	GKQELEKILK	LLRKFRKEVN	1600
1601	SLTEWLATTD	GELTRRSSVE	GMPSDLDAEL	AWAKAAQEDT	EHHKPQLKLV	1650
1651	RELAGTLKGL	LHSQENLIDD	KVSLLNCNWI	AVTSRSEQWL	KLLLDYQNQM	1700
1701	KTLEQNISQI	NTWMDRAEEK	LDEMDSQGCI	EHVIKGLRME	LEEMKGKVEA	1750
1751	VQVLAEDLIK	NTGEHCKAQV	KPKLEQLNQR	FDIVARRILM	GQASSQELDE	1800
1801	YHRQANIWLQ	VLDEEIKLGE	SLKEEDFLED	ATIDEEALNE	LFLKGENLLK	1850
1851	RTPGGEKREA	VREKHNLLHD	RYDTLKNLRA	LRKKKALALA	SQWYQFCKKS	1900
1901	DDMMQWLDRI	EKSIRELPDP	PEEPRVKAVG	YEIDQQRPKL	EDLRGLGRVL	1950
1951	SEGGASRMVE	PRLLPINKRW	TELDVNFTRV	RHKSELQFLL	QCISENEALL	2000
2001	NSPERWSTAF	INLPQQEKCL	KEVKVNLDKL	EGPVAEAVGR	GASQPEEGLL	2050
2051	VQLLRTNWEN	LKMLYQDRLK	RLEKAKKFNE	ELKMLDNWLT	DAERTIMKYE	2100
2101	QDPINNRDHL	KELQAGLEKQ	EAAVKGLNAL	GTDLSPQCSK	DDRDHIKQQL	2150
2151	ASINSRWAKV	SNQLTEIKRR	SAGAKMLLAE	LNEDMGEFQS	WLDDAEAVAA	2200
2201	LPVEAGRKEQ	LSATLEKVKA	RVVELPSRKQ	ALHQINSKGS	SLPADKVKPL	2250
2251	EKQLKVINMR	WAKVSTDLPE	KQRQIEDLLR	DLSLYQEQLS	KLSIWASNTK	2300
2301	TQLEQSPTAV	DPKIEDDVKE	KKPAVETLLA	KERPPCQPEK	GQYDGLSADW	2350
2351	TSIQVLLKDW	KDKCQLAAVT	LTGSAAGDAA	LDKFNKSWTE	LDDWLALLDH	2400
2401	MVQTQRVMVG	DLDEINEMTV	KLKSALQDME	QRCPQLNKQI	TAAQNLKNKT	2450
2451	NNPETRATIT	DRIEKLQAHW	EDSHAKLTAR	VLTLQNMYKD	SSDWLEARKR	2500
2501	VEPLIKKANE	KLESWKKVSH	SVEDLKGQNA	DVKQLSKDLQ	QWQTQMNVTN	2550
2551	ELANKLLTLY	ADDDTSKVKQ	MTESMNLAWA	NIKKRAGDKE	ADLEAGLRQL	2600
2601	QHYYLDLEKF	LNWLTEAETT	ANVLQDATFK	EGLLENPATV	RHLLEQWQDL	2650
2651	QAEIDAHRET	YHSLDENGHR	IVSSLEGTDN	AVVLQKRLDD	MGQRWHELCN	2700
2701	KVMSIRPYLD	AGVDQWKHLH	MSLQELLNWL	QLKREELEKQ	KPVGGDVPTV	2750
2751	HQQLLTHKAF	RRELGAKEPV	INGTLDNAKT	FLAEMPREGL	KQRPGQKDVS	2800
2801	PEERVQNVGR	ILRKEVEDVT	VRWKNLGAAS	VDWQQQLELA	LERLMELQDA	2850
2851	QDQLDYKLRQ	TESVKNSWKP	VGELLVDDLQ	NHIDRVKAFQ	EEIAPIQDNV	2900
2901	NHVNQLASTF	RPSDIQLSPD	NLSRIDDLNM	RWRLLQISIE	EHLSQLTTAF	2950
2951	KDLGPSQNFL	HASVESPLER	SISPNNVPYY	INHQTQTTCW	DHPKMAELYQ	3000
3001	SLADLNNVRF	SAYRTAMKLR	RMQKALCLDL	LSMPAACEAF	EQHNLKQNEQ	3050
3051	FMDIVQVINC	LTSIYDRLEQ	QHSSLVNVPL	CVDMCLNWLL	NVYDTGRAGK	3100
3101	IRTLSFKTGI	ISLCKAHLED	KYRFLFREVA	SATGFCDQRR	LGLLLHDAIQ	3150
3151	IPRQLGEVAS	FGGSNIEPSV	RSCFQFANNK	PELEASVFLD	WMRLEPQSMV	3200
3201	WLPVLHRVAA	AETAKHQAKC	NICKECPIIG	FRYRSLKHFN	YDICQSCFFS	3250
3251	GRVAKGHKMQ	YPMVEYCTPT	TSGEDVRDFA	KVLKNKFRTK	RYFAKHPRMG	3300
3301	YLPVQTILEG	DNMETPVTLI	NFWPVDHPPA	SSPQLSHDDT	HSRIEHYASR	3350
3351	DDEHLLIQHY	CQSLNQGSPL	SQPQSPAQIL	ISMETEEKGE	LERVLNDLEQ	3400
3401	ENRKLQAEYD	RLKKAHDHKG	LSPLPSPPQM	LPVSPQSPRD	AELIAEAKLL	3450
3451	RQHKGRLEAR	MQILEDHNKQ	LESQLTRLRQ	LLEQTESKVN	GTALSSPSTA	3500
3501	SPRSDTSLAS	LRVAASQTTE	TMGDDELSSP	TQDASTGLED	VIEQLNNSFP	3550
3551	HSQGGGRLNP	*				3561

Figure 2. Nucleotide sequence and predicted amino acid sequence of zebrafish Dp427m. (A) Dp427m nucleotide sequence based on double-stranded sequencing of overlapping amplicons derived from reverse-transcribed RNA isolated from zebrafish embryos. Untranslated regions are shown in red, translated regions in blue, bases at exon-exon boundaries are underlined in turquoise and shaded in yellow, and translation start and stop codons are shaded in blue. (B) The translated protein sequence.



Figure 3. Differences between the actual and predicted sequences of zebrafish Dp427m. Five regions of the dystrophin amino acid sequence are shown that differ between the Vega predicted zebrafish dystrophin sequence (Vega Transcript ID: OTTDART00000044458; Vega Zdys) and that predicted from our cDNA sequence (Zdys).

were subsequently sequenced. The entire experimentally determined nucleotide sequence together with the predicted amino acid sequence are shown in Fig. 2. During the course of this study, updates to the zebrafish genome sequence assembly allowed for a more accurate prediction of the full length dystrophin gene; however, while this predicted sequence is largely correct, there are a few regions in the distal half of the full length transcript that do not agree with our sequence data (Fig. 3). This discrepancy may be due to alternative splicing or minor sequencing artefacts. Our sequence also differs from that reported by Guyon et al (44), which is a composite of direct cDNA sequence data and EST sequences. This earlier sequence carries an additional exon (number 72, compared to annotations of the human Dp427m transcript, and so probably represents an alternatively spliced transcript), but also 45 amino acids at positions 495-539 (inclusive), which appears to be a repeat of amino acids at positions 585-629 (inclusive).

The utrophin gene proved difficult to predict using bioinformatic approaches, and the sequences appeared to be fragmentary using the currently available zebrafish genome builds. The utrophin sequence presented here is based on an incomplete utrophin gene prediction, which primarily covers the amino and carboxyl-terminal regions (Fig. 4). A comparison of the predicted exons for zebrafish utrophin and those from human utrophin via ClustalX DNA alignment showed 64.3 and 55.9% of the base pairs to be identical for the amino and carboxyl-terminal regions, respectively.

Developmental expression profiling. Expression profiling of the zebrafish dystrophin gene involved targeting four distinct regions of this gene in order to allow for the simultaneous analysis of its transcriptional diversity; our own sequence data was used for designing most of the primers. The muscle-specific first exon was targeted to follow the expression of the muscle isoform, the second exon of Dp427m allowed the profiling of all full-length transcripts, the expression of Dp71 was followed by targeting this transcript's unique first exon (determined bioinformatically), and the full transcriptional load of the dystrophin gene was determined by targeting exons 75-76 (Fig. 5).

The expression profile of Dp427m approximates that of total full length dystrophin transcripts (Fig. 5A and B) with expression starting at 16 hpf, coinciding with myogenesis, and a reduced rate of expression after 30 hpf. These data suggest that for the most part, the Dp427m isoform is dominant among full length isoforms expressed during zebrafish embryo development. The expression profile of the ubiquitous Dp71 isoform (Fig. 5C) suggests maternal inheritance as post-zygotic transcription begins at 3-3.5 hpf (45). Dp71 transcripts decrease to their lowest level at approximately 16 hpf, and are then largely maintained at a low steady state level. Interestingly, the expression profile of total dystrophin gene transcripts (Fig. 5D) appears to be a composite of Dp71 and Dp427 expression profiles.

Total utrophin gene transcript expression and full length Utr427 expression exhibit similar profiles (Fig. 5E and F), suggesting that the full length Utr427 is the dominant isoform that is expressed by the utrophin gene during embryonic development. The data suggest maternal inheritance of Utr427 transcripts with low level de novo expression starting at approximately 16 hpf.

>Zebrafish_Utrophin_Exon_01

ATGCCCACCGTGGGTGTGATGTCAGAGGCGTCTGAGGGCGGGGGGGAGACAACGAATTCACT GACATCATCAAGTGGAGATCAG

>Zebrafish_Utrophin_Exon_02

ATGAACATGACGCGGTGCAGAAGAAGAACCTTCACCAAGTGGATCAACGCTCAGTTCTCCA AG

>Zebrafish_Utrophin_Exon_03

ACGGGGAAAGCGGCGATCAAAGACATGTTCACTGACCTGAGAGACGGCAGAAAACTGCTG GACCTGCTGGAGGGACTGACCGCAAACACACTG

>Zebrafish_Utrophin_Exon_05

GTGGAGCTGGTGAATATTGGAGGGACAGATATAGTTGATGGCAATCATAAGCTGACTCTG GGCCTCATCTGGAGCATCATCCTGCACTGGCAG

>Zebrafish_Utrophin_Exon_07

GCCTCATGGCTTCAGCTGGGAGAAGGTGGTCAGTCTGACGCCTGTGGAGAGACTGGAGCA CGCTTTCACCTTCGCCAAAGACCAGCTCAACATCGAGAAGCTGCTGGATCCAGAAG

>Zebrafish Utrophin Exon 08

ATGTGGCGGTGCAGCTGCCGGATAAGAAGTCCATCATCATGTACGTGACCTCACTGTTCG CCGTCTTACCCAATGACATCACCATGGATGACATCAGAGAGGTGGAGACGCTGCCCCGAC GCTACAAGGTGGAGACCGACGAGGGACACCCGGTGAGT

>Zebrafish_Utrophin_Exon_09

GTGGAGGTGGACAGCAGCAGCAGCAGCACGCGGGCCGAGACCCTGGCGGAGCTGCAG A T G G A G G T G A A T C T G G A C T C T G G A G G A G G T G A C T G G C T G C TCGGCCGAAGACACGCTGCACATGCAGGAGGACGTTTCTGAAGACGTGGAGGAGGTCAAG GACCAATTCCACACGCATGAG

>Zebrafish_Utrophin_Exon_11

GCTTCATGAGGTCCTGATGGACCTGCAGCAGCAGCAGCTTCAGCAGCTCTCTGATTGGCT GACACAGACAGAGGGGGGGGGATCCGGAAGATGGAGATTGAGCCAATAGCAGGAGACATACA GGGATACCTGACCCAGATAGAGCAGCACAAG

>Zebrafish_Utrophin_Exon_12

GTGGTTGTGGATGAGAACAGCGGTGAGAGCGCCACTGCTGCCCTGGAGGAGCAACTGCAG

>Zebrafish_Utrophin_Exon_13

TCACTGGGTGAGCGCTGGGCGGCAGTGTGTCGCTGGACGGAGGAGCGCTGGAACAAACTG GATGAAACTCAGCTGGTGTGGCAGAGGCTGCTTGATGACCAG

>Zebrafish_Utrophin_Exon_14

AGTTTATTTGGGTCATGGCTAACCGAAAAGGAGAGGCCTTGGGTGAAGTGCAGACCTGT GATTTTAAAGACCCCAGTGACATTAATGCCAGCGTCCGCAGATTGGCT

>Zebrafish_Utrophin_Exon_15

AGTCTTAAGGAGGATATGGCCCAGAAGCGCAGGACTTTGAGCGCTCTGTCAGACGCTGGT CAAGATGTGATGGTGCTGTTGAACAGTCCAGCCGCAGCTCAAAGGATCGAATCCGACACA

>Zebrafish_Utrophin_Exon_63

CAGTCTATGGTTTGGCTGCCGGTTCTACACCGAGTTGCTGCGGCAGAGACGGCGAAACAT CAGGCCAAGTGCAACATCTGCAAAGAGTTCCCCATCGTCGGCTTCAG

>Zebrafish_Utrophin_Exon_64

ATATCGCAGTTTAAAGCACTTCAACTATGACGTCTGCCAGAGCTGTTTCTTCTCTGGTCG CACAGCGAAGGGTCATAAACTGAACTACCCGATGGTGGAGTACTGCACACCG

>Zebrafish_Utrophin_Exon_66

TCCGGTCACTCTCATCAGCATGTGTCCGGAGCACTATGA

>Zebrafish_Utrophin_Exon_67

CAACAG

>Zebrafish_Utrophin_Exon_68

GCTGGCACAGATGGAGCGATCCAATGGCTCTCTGCCCACCGACAGCAGCTCGGCCACTGG AAGCAT

>Zebrafish_Utrophin_Exon_69

TTTCAGCCAGCCACAAAGCCCCGCCCATATCAACAGGACCCCCGCCCACAACACCCCACAG AGGAGAGCTGGAGCGAATCATAGCACGGCTGGAGGACGATCAGCG

>Zebrafish_Utrophin_Exon_70

GGTTCTGCAGAGGGAGTATGAAGAACTGCGGTTTCAGTACGGTCAGCCTGGCGCCGCAGC TGCTAGCGGAGCCCCAGATGAAGCTGATCTGCTAGCTGAAGCTAAACTCCTGCGACAGCA CAAGGGACGACTGGAGGCCCGCATGCAGATTCTGGAGGACCACAACAACAGCTGGAGTC TCAGCTTTACCGACTGCGCCAGCTGCTGCATCAG

>Zebrafish_Utrophin_Exon_72

ACGAGTTTCTCCATCAGTGTGCTAACAGTAACTCTGCTCTGGCGGAGGTTATGGATCAGA TTAACCACAGTTTTTCCTGCATGCTCTC

>Zebrafish_Utrophin_Exon_73 CGTCGAGTGTGTCCTCAAGACCGAAG

Figure 4. Predicted zebrafish utrophin gene exon sequences.



Figure 5. Developmental expression profiles of dystrophin and utrophin gene transcripts. The developmental stage in hpf is shown on the X-axis, and the relative expression levels (arbitrary units) are shown on the Y-axis. The latter have been normalised against the two reference genes Rpl13a and EFla.

Discussion

The data presented here used an approach that allowed for the simultaneous expression profiling of the Dp427m transcript, all full-length dystrophin transcripts, the Dp71 transcript, and all transcripts expressed by the dystrophin gene. This approach involved targeting common exons that are expressed among isoforms, and targeting unique exons to allow for the profiling of specific isoforms. In addition, the profiles of all full length utrophin gene transcripts, as well as all transcripts expressed by the utrophin gene, were examined throughout zebrafish embryonic development. In terms of the former, it was considered critical to avoid exons involved in alternative splicing, which comprised exons 4 (46), 39, 41-42, 45-50 (47), 66-68 and 71-74 (24,25,48) and 78 (49). As such, the applicability of the approach described here requires knowledge of alternative splicing of target genes. The results reflect expression across the whole embryo as development progresses, and represent an average across all tissues and cell types.

The total expression profile of the dystrophin gene comprises that of Dp71 and the full length expression profiles, with Dp71 being the most abundant isoform present during the early stages of embryonic development, up until 16 hpf. From 16 hpf onwards, full length dystrophin expression appears to be present in higher abundance relative to other dystrophin isoforms. It should be noted that while other potential dystrophin transcripts were not investigated, they do not appear to be major players based on the expression profiles reported here. Experimental validation of this conclusion would be difficult in the context of qRT-PCR analysis as the sequences of the unique first exons of other dystrophin isoforms thought to exist are currently unknown or are too short to design transcriptspecific primers.

In contrast to dystrophin, in which the predominantly expressed isoforms during embryonic development change from Dp71 to the full length dystrophin, the dominant isoform of utrophin that is expressed throughout development appears to be the full length Utr427. This conclusion does not rule out the presence of other as yet uncharacterised isoforms, but their expression levels would not be expected to be significant given our analysis herein.

The caveat to looking at gene expression of a whole embryo, as opposed to individual tissues, is that an absence of noticeable expression changes between two developmental stages does not imply the absence of differences between individual tissues. Conversely a small change in expression across the whole embryo may represent significant changes in a single tissue type. For example, a moderate increase in the levels of total full length dystrophins could reflect a significant increase in the expression of non-muscle isoforms. Localised expression, not undertaken here, might be evident by RNA in situ analysis (24,25), but as suggested earlier, the design of appropriate probes suffers from sensitivity and specificity issues that are not easily resolved in the case of the two genes studied. With these effects in mind, the total expression profile should be interpreted with caution as it does not reflect expression in different tissues, which is driven by tissue-specific promoters. Ideally for genes with tissue-specific expression such as dystrophin, tissue-specific investigations would be preferred over whole embryo studies. However, attempting to isolate individual tissue types from a developing embryo would be technically challenging. In conclusion, the expression data are the first reported attempt to follow multiple dystrophin and utrophin gene transcripts simultaneously throughout vertebrate development. The approach offers opportunities to determine the impact of targeted dystrophin gene mutagenesis in the zebrafish on the expression profile of the dystrophin gene, and the possible upregulation of utrophin gene transcripts.

Acknowledgements

We thank the Genomics Unit, DNA Sequencing Facility at the School of Biological Sciences, University of Auckland for their sequencing services. We acknowledge the assistance of Dr Andrew Dodd and Dr Rongying Tang who provided the foundation for the research presented herein. We acknowledge the financial assistance of the Maurice and Phyllis Paykel Trust (http://www.paykeltrust.co.nz/) and the School of Biological Sciences, The University of Auckland (http://www.sbs.auckland.ac.nz/), for partial funding of the study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Tinsley JM, Blake DJ, Zuellig RA and Davies KE: Increasing complexity of the dystrophin-associated protein complex. Proc Natl Acad Sci USA 91: 8307-8313, 1994.
- Lederfein D, Levy Z, Augier N, Mornet D, Morris G, Fuchs O, Yaffe D and Nudel U: A 71 kd protein is a major product of the Duchenne muscular dystrophy gene in brain and other nonmuscle tissues. Proc Natl Acad Sci USA 89: 5346-5350, 1992.
- Austin RC, Howard PL, D'Souza VN, Klamut HJ and Ray PN: Cloning and characterization of alternatively spliced isoforms of Dp71. Hum Mol Genet 4: 1475-1483, 1995.
- Love DR, Hill DF, Dickson G, Spurr NK, Byth BC, Marsden RF, Walsh FS, Edwards YH and Davies KE: An autosomal transcript in skeletal muscle with homology to dystrophin. Nature 339: 55-58, 1989.
- Tinsley JM, Blake DJ, Roche A, Fairbrother U, Riss J, Byth BC, Knight AE, Kendrick-Jones J, Suthers GK, Love DR, et al: Primary structure of dystrophin-related protein. Nature 360: 591-593, 1992.
- Nudel U, Zuk D, Einat P, Zeelon E, Levy Z, Neuman S and Yaffe D: Duchenne muscular dystrophy gene product is not identical in muscle and brain. Nature 337: 76-78, 1989.
- Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kurnit DM and Kunkel LM: Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature 323: 646-650, 1986.
- Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener CA and Kunkel LM: Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50: 509-517, 1987.
- Gorecki DC, Monaco AP, Derry JM, Walker AP, Barnard EA and Barnard PJ: Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. Hum Mol Genet 1: 505-510, 1992.

- Holder E, Maeda M and Bies RD: Expression and regulation of the dystrophin Purkinje promoter in human skeletal muscle, heart, and brain. Hum Genet 97: 232-239, 1996.
- D'Souza VN, Nguyen TM, Morris GE, Karges W, Pillers DA and Ray PN: A novel dystrophin isoform is required for normal retinal electrophysiology. Hum Mol Genet 4: 837-842, 1995.
- Lidov HGW, Selig S and Kunkel LM: Dp140: a novel 140 kDA CNS transcript from the dystrophin locus. Hum Mol Genet 4: 329-335, 1995
- Byers TJ, Lidov HGW and Kunkel LM: An alternative dystrophin transcript specific to peripheral nerve. Nat Genet 4: 77-81, 1993.
- Burton EA, Tinsley JM, Holzfeind PJ, Rodrigues NR and Davies KE: A second promoter provides an alternative target for therapeutic upregulation of utrophin in Duchenne muscular dystrophy. Proc Natl Acad Sci USA 96: 14025-14030, 1999.
- Zuellig ŘA, Bornhauser BC, Knuesel I, Heller F, Fritschy JM and Schaub MC: Identification and characterisation of transcript and protein of a new short N-terminal utrophin isoform. J Cell Biochem 77: 418-431, 2000.
- Blake DJ, Schofield JN, Zuellig RA, Górecki DC, Phelps SR, Barnard EA, Edwards YH and Davies KE: G-utrophin, the autosomal homologue of dystrophin Dp116, is expressed in sensory ganglia and brain. Proc Natl Acad Sci USA 9: 3697-3701, 1995.
- Wilson J, Putt W, Jimenez C and Edwards YH: Up71 and up140, two novel transcripts of utrophin that are homologues of short forms of dystrophin. Hum Mol Genet 8: 1271-1278, 1999.
- Jimenez-Mallebrera C, Davies K, Putt W and Edwards YH: A study of short utrophin isoforms in mice deficient for full-length utrophin. Mamm Genome 14: 47-60, 2003.
- Matsumura K, Ervasti JM, Ohlendieck K, Kahl SD and Campbell KP: Association of dystrophin-related protein with dystrophin-associated proteins in mdx mouse muscle. Nature 360: 588-591, 1992.
- Cooper BJ, Winand NJ, Stedman H, Valentine BA, Hoffman EP, Kunkel LM, Scott MO, Fischbeck KH, Kornegay JN, Avery RJ, *et al*: The homologue of the Duchenne locus is defective in X-linked muscular dystrophy of dogs. Nature 334: 154-156, 1988.
- 21. Carpenter JL, Hoffman EP, Romanul FC, Kunkel LM, Rosales RK, Ma NS, Dasbach JJ, Rae JF, Moore FM, McAfee MB, *et al*: Feline muscular dystrophy with dystrophin deficiency. Am J Pathol 135: 909-919, 1989.
- 22. Hoffman EP, Brown RH Jr and Kunkel LM: Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell 51: 919-928, 1987.
- Pozzoli U, Elgar G, Cagliani R, Riva L, Comi GP, Bresolin N, Bardoni A and Sironi M: Comparative analysis of vertebrate dystrophin loci Indicate intron gigantism as a common feature. Genome Res 13: 764-772, 2003.
- 24. Bolaños-Jiménez F, Bordais A, Behra M, Strähle U, Mornet D, Sahel J and Rendón A: Molecular cloning and characterization of dystrophin and Dp71, two products of the Duchenne Muscular Dystrophy gene, in zebrafish. Gene 274: 217-226, 2001.
- Bolaños-Jiménez F, Bordais A, Behra M, Strähle U, Sahel J and Rendón A: Dystrophin and Dp71, two products of the DMD gene, show a different pattern of expression during embryonic development in zebrafish. Mech Dev 102: 239-241, 2001.
 Chambers SP, Dodd A, Overall R, Sirey T, Lam LT, Morris GE
- Chambers SP, Dodd A, Overall R, Sirey T, Lam LT, Morris GE and Love DR: Dystrophin in adult zebrafish muscle. Biochem Biophys Res Commun 286: 478-483, 2001.
- Kramarcy NR and Sealock R: Dystrophin as a focal adhesion protein. Collocalization with talin and the Mr 48,000 sarcolemmal protein in cultured Xenopus muscle. FEBS Lett 274: 171-174, 1990.
- Wang J, Pansky A, Venuti JM, Yaffe D and Nudel U: A sea urchin gene encoding dystrophin-related proteins. Hum Mol Genet 7: 581-588, 1998.
- Neuman S, Kaban A, Volk T, Yaffe D and Nudel U: The dystrophin/ utrophin homologues in Drosophila and sea urchin. Gene 263: 17-29, 2001.
- Vater R, Young C, Anderson LV, Lindsay S, Blake DJ, Davies KE, Zuellig R and Slater CR: Utrophin mRNA expression in muscle is not restricted to the neuromuscular junction. Mol Cell Neurosci 10: 229-242, 1998.
- 31. Urasawa N, Wada MR, Machida N, Yuasa K, Shimatsu Y, Wakao Y, Yuasa S, Sano T, Nonaka I, Nakamura A, *et al*: Selective vacuolar degeneration in dystrophin-deficient canine Purkinje fibers despite preservation of dystrophin-associated proteins with overexpression of Dp71. Circulation 117: 2437-2448, 2008.
- Manzur AY, Kinali M and Muntoni F: Update on the management of Duchenne muscular dystrophy. Arch Dis Child 93: 986-990, 2008.

- 33. Sicinski P, Geng Y, Ryder-Cook AS, Barnard EA, Darlison MG and Barnard PJ: The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. Science 244: 1578-1580, 1989.
- Bassett DI and Currie PD: The zebrafish as a model for muscular dystrophy and congenital myopathy. Hum Mol Genet 12: 265-270, 2003.
- 35. Bassett DI, Bryson-Richardson RJ, Daggett DF, Gautier P, Keenan DG and Currie PD: Dystrophin is required for the formation of stable muscle attachments in the zebrafish embryo. Development 130: 5851-5860, 2003.
- 36. Dodd A, Chambers SP and Love DR: Short interfering RNA-mediated gene targeting in the zebrafish. FEBS Lett 561: 89-93, 2004.
- Dodd A, Curtis PM, Williams LC and Love DR: Zebrafish: bridging the gap between development and disease. Hum Mol Genet 9: 2443-2449, 2000.
- Guyon JR, Mosley AN, Zhou Y, O'Brien KF, Sheng X, Chiang K, Davidson AJ, Volinski JM, Zon LI and Kunkel LM: The dystrophin associated protein complex in zebrafish. Hum Mol Genet 12: 601-615, 2003.
- 39. Lan CC, Tang R, Un San Leong I and Love DR: Quantitative real-time RT-PCR (qRT-PCR) of zebrafish transcripts: optimization of RNA extraction, quality control considerations, and data analysis. Cold Spring Harb Protoc 2009: pdb.prot5314, 2009.
- 40. Tang R, Dodd A, Lai D, McNabb WC and Love DR: Validation of zebrafish (*Danio rerio*) reference genes for quantitative real-time RT-PCR normalization. Acta Biochim Biophys Sin (Shanghai) 39: 384-390, 2007.

- 41. Ramakers C, Ruijter JM, Deprez RH and Moorman AF: Assumption-free analysis of quantitative real-time PCR data. Neurosci Lett 339: 62-66, 2003.
- 42. Ruijter JM, Ramakers C, Hoogaars WM, Karlen Y, Bakker O, Van den Hoff MJ and Moorman AF: Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res 37: e45, 2009.
- 43. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A and Speleman F: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: Research0034, 2002.
- 44. Guyon JR, Goswami J, Jun SJ, Thorne M, Howell M, Pusack T, Kawahara G, Steffen LS, Galdzicki M and Kunkel LM: Genetic isolation and characterization of a splicing mutant of zebrafish dystrophin. Hum Mol Genet 18: 202-211, 2009.
- Kane DA and Kimmel CB: The zebrafish midblastula transition. Development 119: 447-456, 1993.
- 46. Torelli S and Muntoni F: Alternative splicing of dystrophin exon 4 in normal human muscle. Hum Genet 97: 521-523, 1995.
- 47. Sironi M, Cagliani R, Pozzoli U, Bardoni A, Comi GP, Giorda R and Bresolin N: The dystrophin gene is alternatively spliced throughout its coding sequence. FEBS Lett 517: 163-166, 2002.
- 48. Lidov HG and Kunkel LM: Dp140: alternatively spliced isoforms in brain and kidney. Genomics 45: 132-139, 1997.
- 49. Ceccarini M, Rizzo G, Rosa G, Chelucci C, Macioce P and Petrucci TC: A splice variant of Dp71 lacking the syntrophin binding site is expressed in early stages of human neural development. Brain Res Dev Brain Res 103: 77-82, 1997.