Supplementary Information for Doyon et al., "Heritable Targeted Gene Disruption in

Zebrafish Using Designed Zinc Finger Nucleases"

Table of Contents

Methods	Describes design, assembly, and in vitro evaluation of ZFNs, describes in detail the yeast reporter system, and lists references
Supp. Fig. 1	Schematic of the PCR-based ZFN assembly procedure from individual 2- finger modules
Supp. Fig. 2	ELISA data for DNA binding by all ZFPs designed and assembled against the <i>golden</i> locus
Supp. Fig. 3	Schematic, restriction map, and sequence of the reporter plasmid used in the budding yeast ZFN activity screening system
Supp. Fig. 4	Vector map and full DNA/amino acid sequence of the ZFNs used for <i>golden</i> gene disruption
Supp. Fig. 5	ELISA data for DNA binding by all ZFPs designed and assembled against the <i>ntl</i> locus
Supp. Fig. 6	The full DNA/amino acid sequence of the ZFNs used for <i>ntl</i> gene disruption
Supp. Fig. 7	ZFN injection can cause hypomorphic phenotypes
Supp. Fig. 8	Injection of <i>ntl</i> ZFN-encoding mRNA into zebrafish embryos induces disruption of the <i>ntl</i> gene in somatic cells
Supp. Fig. 9	Injection of <i>ntl</i> ZFN-encoding mRNA into wildtype embryos induces tail phenotypes in juvenile and adult fish
Supp. Fig. 10	Injection of <i>ntl</i> ZFN-encoding mRNA into wildtype embryos induces biallelic disruption of the <i>ntl</i> gene in somatic cells
Supp. Fig. 11	Analysis of ZFN action at potential off-target sites
Supp. Table 1	ZFN pairs designed against the golden locus
Supp. Table 2	ZFN pairs designed against the <i>ntl</i> locus
Supp. Table 3	Primer sequences used to genotype potential off-target sites for ZFN action
Supp. Table 4	Plasmid designations for ZFN expression vectors

Supplementary Methods

Design, assembly, and in vitro evaluation of ZFN constructs

The published literature has accumulated a very extensive collection of methods for developing zinc finger proteins that target a sequence of one's choosing (Pabo et al, 2001). These include the rational elaboration of the protein-DNA interface (Desjarlais & Berg, 1992), use of affinity selection methodologies such as phage display (Rebar & Pabo, 1994; Greisman & Pabo, 1997; Segal et al, 1999; Beerli et al, 2000; Dreier et al, 2001; Isalan & Choo, 2001), biological selection methods such as 1 or 2-hybrid systems in budding yeast (Bartsevich & Juliano, 2000) or bacteria (Joung et al, 2000), and the use of naturally occurring zinc fingers (Bae et al, 2003). These strategies, combined with the modular reassortment of zinc fingers with pre-characterized specificites (Wright et al, 2006), have been successfully used by a large number of laboratories to develop DNA binding domains that engage investigator-specified targets in living cells in a broad range of organisms (Choo et al, 1994; Beerli et al, 2002; Ren et al, 2000; Zhang et al, 2000; Bibikova et al, 2001; Liu et al, 2001; Bibikova et al, 2002; Ren et al, 2002; Bartsevich et al, 2003; Porteus & Baltimore, 2003; Reynolds et al, 2003; Tan et al, 2003; Lloyd et al, 2005; Wright et al, 2005; Beumer et al, 2006; Morton et al, 2006).

The particular approach we have adopted for developing zinc finger proteins for genome editing applications relies on an archive of pre-characterized two-finger modules, each of which recognizes an experimentally validated 6 bp half-site (Isalan et al, 2001; Moore et al, 2001). To develop ZFNs directed against zebrafish *gol* and *ntl* cDNAs, these sequences were scanned for positions where modules exist in the archive that allow the fusion of two such modules to form a 4-finger protein (ZFN-R) with a composite 12 bp target site recognized on the Watson strand, and another such 4 finger protein (ZFN-L) that recognizes a 12 bp target site on the Crick strand, 5 or 6 bp away from the 5' most base pair recognized by ZFN-R.

While a detailed analysis of the relative merits of the various approaches to ZFP design is outside the scope of the present discussion, we note that the process described above reliably produces ZFPs that pass DNA binding ELISA assays (see Supplementary Figs. 2 and 5), pass functional assays for activity in yeast-based proxy systems (Fig. 2A and 3A), and, when fused to the FokI endonuclease domain, efficiently and specifically engage their intended, endogenous target loci in the context of complex genomes such as zebrafish (present work), hamster (Santiago et al, 2008), and human (Urnov et al, 2005; Lombardo et al, 2007; Miller et al, 2007; Perez et al, *Nature Biotechnology*, in press).

These ZFNs are assembled using a PCR-based procedure described in Supplementary Fig. 1. In brief, each two-finger module is amplified by PCR in a separate reaction. The two PCR products that correspond to the two-finger modules that compose each ZFN are then combined and joined by conventional restriction enzyme digestion-ligation into a ZFN expression vector to yield a gene encoding (NH₂ to COOH) a triple-FLAG tag, a nuclear localization signal, the ZFP module, and the endonuclease domain of the type IIS restriction enzyme FokI. The ZFN coding region is flanked with the cytomegalovirus (CMV) immediate-early promoter and bovine growth hormone (BGH) polyadenylation signal. The complete sequence of all ZFNs used in this work are shown in Supplementary Figs. 4 and 6. Each ZFN was evaluated for DNA binding using an ELISA assay performed as described (Bartsevich et al, 2003).

The *in vitro* consensus binding site for each ZFN was determined using a procedure described in detail elsewhere (Perez et al, *Nature Biotechnology* in press). In brief, HA-tagged ZFPs were synthesized by in vitro transcription-translation, and SELEX was performed by incubation with a pool of randomized DNA sequences and an anti-HA biotin-coupled antibody, capture of protein-bound DNA by streptavidin magnetic-coated beads, and PCR-based amplification of the bound DNA, and use of the resulting PCR pool in a second round of SELEX.

This procedure was repeated for a total of 4 rounds of selection, and DNA fragments amplified after the final round were cloned and sequenced.

Use of a Budding Yeast-Based Reporter System to Identify ZFN Pairs for Gene Disruption

Construction of the reporter plasmid

The reporter construct was targeted to the HO locus using the yeast integrating plasmid (YIp) HOpoly-KanMX-HO (Voth et al, 2001). To generate the SSA reporter (Supplementary Fig. 3), a fragment corresponding to nucleotides 1 to 750 of the *MEL1* gene (Liljestrom, 1985) (relative to the ATG) was cloned into the SalI and BamHI sites of HO-poly-KanMX-HO using the following primers: 5'-AATTGTCGACATGTTTGCTTTCTACTTTCTCACCGC-3' and 5'-AATTGGATCCCCCCATTGGAGCTGCC-3'. Then a fragment from nucleotides 299 to 2100 was cloned into the SacI and EcoRI sites using the following primers: 5'-AATTGAACTCAGACCACCTGCATAATAACAGC-3' and 5'-AATTGAAGCTCAGACCACCTGCATAATAACAGC-3' and 5'-AATTGAATTCGGGCAAAAATTGGTACCAATGC-3'. Finally, a 1489 bp fragment of the *PGK1* promoter was cloned into the BsiWI and SalI sites using the following primers: 5'-AATTCGTACGTCTAACTGATCTATCCAAAACTG-3' and 5'-

Construction of the reporter strain

Integration of the reporter construct into the 69-1B strain (S288C background; *MATa his3 200* $lys2-128\delta leu2\ 1$) was performed as described (Voth et al, 2001). Note that the designer deletion strain BY4741, available from Open Biosystems, provides the same characteristics as the 69-1B strain used in this study. Briefly, 2 g of the reporter construct containing the target sequence was

linearized with NotI and used to transform yeast using the lithium acetate method (Gietz & Schiestl, 2007). We confirmed correct integration by colony PCR using the following primers: HO-L: 5'-TATTAGGTGTGAAACCACGAAAAGT-3' and 5'-ACTGTCATTGGGAATGTCTTATGAT-3'; HO-R: 5'-ATTACGCTCGTCATCAAAATCA-3'; 5'-CATGTCTTCTCGTTAAGACTGCAT-3'.

ZFN expression vectors

The entire coding sequence of each ZFN pair was transferred to galactose inducible expression vectors using standard cloning procedures (Mumberg et al, 1994; Urnov et al, 2005; Moehle et al, 2007). These YCp vectors, p413prom and p415prom, are available through the ATCC (Mumberg et al, 1994). Transformation of the reporter strain with the ZFN expression vectors (plasmid names listed in Supplementary Table 4) was done in deep well blocks as described (Gietz & Schiestl, 2007).

Induction of ZFN expression

To derepress the *GAL1* promoter, the pools of transformants were diluted 1:10 into 1 ml of SC His-Leu- medium containing 2% raffinose as a source of carbon and incubated overnight at 30°C. ZFN expression was induced by diluting the raffinose cultures 1:10 into 1ml of SC His-Leu- medium containing 2% galactose. Cells were then incubated for 2 to 6 hours, before addition of 2% glucose to stop expression. Cells were then incubated overnight to allow for DSB repair and reporter expression.

Reporter assay (MEL1 assay)

The first step was to determine the cell density of the cultures in order to normalize the reporter signal to the amount of cells in the culture. This was done by a simple spectrophotometric reading

at 600 nm. The deep well block was then centrifuged at 3000g for 5 minutes to pellet yeast cells and 10 ul of the media was assayed for Mel1 activity as described (Ryan et al, 1998; Chen et al, 2004).

References for Supplementary Information

Bae KH, Kwon YD, Shin HC, Hwang MS, Ryu EH, Park KS, Yang HY, Lee DK, Lee Y, Park J, Kwon HS, Kim HW, Yeh BI, Lee HW, Sohn SH, Yoon J, Seol W, Kim JS (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* **21**: 275-280

Bartsevich VV, Juliano RL (2000) Regulation of the MDR1 gene by transcriptional repressors selected using peptide combinatorial libraries. *Mol Pharmacol* **58**: 1-10.

Bartsevich VV, Miller JC, Case CC, Pabo CO (2003) Engineered zinc finger proteins for controlling stem cell fate. *Stem Cells* **21**: 632-637

Beerli RR, Dreier B, Barbas CF, 3rd (2000) Positive and negative regulation of endogenous genes by designed transcription factors. *Proc Natl Acad Sci U S A* **97:** 1495-1500

Beerli RR, Segal DJ, Dreier B, Barbas CF, 3rd (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A* **95:** 14628-14633

Beumer K, Bhattacharyya G, Bibikova M, Trautman JK, Carroll D (2006) Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics* **172:** 2391-2403

Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, Chandrasegaran S (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21**: 289-297.

Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc- finger nucleases. *Genetics* **161**: 1169-1175.

Chen Q, Chen J, Sun T, Shen J, Shen X, Jiang H (2004) A yeast two-hybrid technology-based system for the discovery of PPARgamma agonist and antagonist. *Analytical biochemistry* **335**: 253-259

Choo Y, Sanchez-Garcia I, Klug A (1994) In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **372**: 642-645.

Desjarlais JR, Berg JM (1992) Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc Natl Acad Sci U S A* **89:** 7345-7349

Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF, 3rd (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **276**: 29466-29478.

Gietz RD, Schiestl RH (2007) Microtiter plate transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols* **2:** 5-8

Greisman HA, Pabo CO (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**: 657-661

Isalan M, Choo Y (2001) Rapid, high-throughput engineering of sequence-specific zinc finger DNA- binding proteins. *Methods Enzymol* **340:** 593-609

Isalan M, Klug A, Choo Y (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol* **19:** 656-660.

Joung JK, Ramm EI, Pabo CO (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A* **97:** 7382-7387

Liljestrom PL (1985) The nucleotide sequence of the yeast MEL1 gene. *Nucleic Acids Res* 13: 7257-7268

Liu PQ, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li PX, Chen B, Mendel MC, Zhong X, Lee YL, Eisenberg SP, Spratt SK, Case CC, Wolffe AP (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J Biol Chem* **276**: 11323-11334.

Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proc Natl Acad Sci U S A* **102**: 2232-2237

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* **25**: 1298-1306

Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA, Gregory PD, Pabo CO, Rebar EJ (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25:** 778-785

Moehle EA, Rock JM, Lee YL, Jouvenot Y, Dekelver RC, Gregory PD, Urnov FD, Holmes MC (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A* **104**: 3055-3060

Moore M, Choo Y, Klug A (2001) Design of polyzinc finger peptides with structured linkers. *Proc Natl Acad Sci U S A* **98**: 1432-1436.

Morton J, Davis MW, Jorgensen EM, Carroll D (2006) Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. *Proc Natl Acad Sci U S A* **103**: 16370-16375

Mumberg D, Muller R, Funk M (1994) Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res* **22:** 5767-5768

Pabo CO, Peisach E, Grant RA (2001) Design and Selection of Novel Cys2his2 Zinc Finger Proteins. *Annu Rev Biochem* **70**: 313-340.

Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science* **300:** 763.

Rebar EJ, Pabo CO (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**: 671-673

Ren D, Collingwood TN, Rebar EJ, Wolffe AP, Camp HS (2002) PPARγ knockdown by engineered transcription factors: exogenous PPARγ2 but not PPARγ1 reactivates adipogenesis. *Genes Dev* **16**: 27-32.

Reynolds L, Ullman C, Moore M, Isalan M, West MJ, Clapham P, Klug A, Choo Y (2003) Repression of the HIV-1 5' LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors. *Proc Natl Acad Sci U S A* **100**: 1615-1620.

Ryan MP, Jones R, Morse RH (1998) SWI-SNF complex participation in transcriptional activation at a step subsequent to activator binding. *Mol Cell Biol* **18**: 1774-1782.

Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN (2008) Targeted gene knockout in mammalian cells using engineered zinc finger nucleases. *Proc Natl Acad Sci U S A* **105**: 5809-5814

Segal DJ, Dreier B, Beerli RR, Barbas CF, 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* **96:** 2758-2763

Tan S, Guschin D, Davalos A, Lee YL, Snowden AW, Jouvenot Y, Zhang HS, Howes K, McNamara AR, Lai A, Ullman C, Reynolds L, Moore M, Isalan M, Berg LP, Campos B, Qi H, Spratt SK, Case CC, Pabo CO, Campisi J, Gregory PD (2003) Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. *Proc Natl Acad Sci U S A* **100**: 11997-12002

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**: 646-651

Voth WP, Richards JD, Shaw JM, Stillman DJ (2001) Yeast vectors for integration at the HO locus. *Nucleic Acids Res* **29:** E59-59

Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF, Joung JK (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nature protocols* **1**: 1637-1652

Wright DA, Townsend JA, Winfrey RJ, Jr., Irwin PA, Rajagopal J, Lonosky PM, Hall BD, Jondle MD, Voytas DF (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J* **44:** 693-705

Zhang L, Spratt SK, Liu Q, Johnstone B, Qi H, Raschke EE, Jamieson AC, Rebar EJ, Wolffe AP, Case CC (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. *J Biol Chem* **275**: 33850-33860.

Supplementary Figure 1: Schematic of the PCR-based ZFN assembly procedure from individual 2-finger modules. Each two-finger module is amplified by PCR in a separate reaction. The two PCR products that correspond to the two-finger modules that compose each ZFN are then combined and joined by conventional restriction enzyme digestion-ligation into a ZFN expression vector to yield a gene encoding, NH_2 to COOH, a triple-FLAG tag, a nuclear localization signal, the ZFP module, and the endonuclease domain of the type IIS restriction enzyme FokI.



NLS: nuclear localization signal F1-4: ZFP helices Fokl: Fokl cleavage domain **Supplementary Figure 2: ELISA data for DNA binding by all the ZFPs designed and** assembled against the *golden* locus. Each ZFN was evaluated for DNA binding using an ELISA assay performed as described (Bartsevich et al, 2003). The first sample is a negative control; the second and third sample, respectively, are positive controls and represent ZFN-R and ZFN-L from previous work (Urnov et al, 2005). Each ZFN is annotated by a specific base-pair position within the *golden* locus (the cDNA cannot be used as a reference in this context because some ZFNs span exon/intron boundaries). In this nomenclature, position #1 corresponds to the transcription start site of the *golden* gene as per the UCSC genome browser DanRer5 zebrafish genome annotation (July 2005); an "r" before the ZFN name indicates that its primary recognition is the Crick strand. The ZFNs highlighted in red form pairs 1 and 14/15, respectively, used in the primary text. Pair 1 ZFN-R is 5084a1 and ZFN-L is r5077a1; Pair 14/15 ZFN-R is 14368a1, Pair 14 ZFN-L is r14361b1, and Pair 15 ZFN-L is r14361a1..



Supplementary Figure 3: Schematic, restriction map, and sequence of the reporter plasmid used in the budding yeast ZFN activity screening system.



Schematic and restriction map:

tures: total length, 10099 bp HO left homology arm: bases 67-984 PGK1 promoter: bases 985-2473 5' end of the *MEL1* gene: bases 2480-3229 MCS: bases 3231-3257 kanMX4 cassette: bases 3258-4705 3' end of the *MEL1* gene: bases 4707-6516 HO right homology arm: bases 6517-7029 pUK21 backbone: bases 7027-54 Fea

Vector sequence (also available as a Vector NTI file from <u>furnov@sangamo.com</u> upon request):

1	ATGACCATGA	TTACGCCACT	AGTCCGAGGC	CTCGAGATCC	GATATCGCCG	TGGCGGCCGC	CAGCTGAAGC	TTAATTATCC	TGGGCACGAG
	TACTGGTACT	AATGCGGTGA	TCAGGCTCCG	GAGCTCTAGG	CTATAGCGGC	ACCGCCGGCG	GTCGACTTCG	AATTAATAGG	ACCCGTGCTC
101	CTAAAACCTT CATGTGTACA	TATTTAGCAT	GGCCATTGAA	TGTAACAATT	ATATATATCG	CAAGCACAAA	AAATCAAGGA	GAGAGAACTA	CCACTTTGTT
	GATTTTGGAA GTACACATGT	ATAAATCGTA	CCGGTAACTT	ACATTGTTAA	TATATATAGC	GTTCGTGTTT	TTTAGTTCCT	CTCTCTTGAT	GGTGAAACAA
201	ATGTTCATTA TTGACATTGC	TCTCCATAAG	САААААААА	AAATAGAAAA	CATATGCTAT	AAGGTTGATA	TTCTCACGAG	TAAGCGGCAC	TTGCTACTTA
	TACAAGTAAT AACTGTAACG	AGAGGTATTC	GTTTTTTTTT	TTTATCTTTT	GTATACGATA	TTCCAACTAT	AAGAGTGCTC	ATTCGCCGTG	AACGATGAAT
301	AGATTTTTGG TCCTTAGTAG	CTACAGAAAT	AGTATATTAG	AGATTATAAT	TGCTAATCAA	ATCAAAATAT	AAAATTAGTA	AACCAAACCA	TTTATACCCT
	TCTAAAAACC AGGAATCATC	GATGTCTTTA	TCATATAATC	TCTAATATTA	ACGATTAGTT	TAGTTTTATA	TTTTAATCAT	TTGGTTTGGT	AAATATGGGA
401	AGATTGAACG	TTTTTTAATG	ATATTTCTGC	AAACCAAAGA	AAGATTGTTA	TCCAGATAGA	ATTTAGTTTT	GATATTCATT	TTTTTTGTTGA
5.01	TCTAACTTGC	CCCTCATAT	TCANAGACG	GTGCCATTAT	CCCTACCCT	TCCCATTCT	CTGGATTTCA	CIAIAAGIAA	AAAAACAACI
501	TCGAAAAGAA	CGGAGTATTA	AGTTTTCTCC	CACGGTAATA	GCCATCGCAA	AGCGTAACAT	GACCTAAAGT	CTTTAAAGTG	TCAACTACTT
601	AGCTTTTTCTT	GCAACACGTA	AGITTALGAT	GTCCCTTTTT	ACCATTATAG	GCALTAATG	DATCATAAAA	CGACCGTATA	CTGGTGALAT
001	AGTAGGGAGA	CGTTGTGCAT	TCCAATTCTA	CAGGGAAAAA	TGGTAATATC	CGTTATTTAC	TTAGTATTTT	GCTGGCATAT	GACCACTTTA
701	TCATCCCTCT ACGAGTACCT	GTAGTAAAAA	GTATAAATCA	TAGTTAATCG	GGCAATGTCC	CTCGATCAAG	GAGTATTGTG	TCATGTTCGA	GACAAACGCC
	AACATTTTTG TGCTCATGGA	CATCATTTTT	CATATTTAGT	ATCAATTAGC	CCGTTACAGG	GAGCTAGTTC	CTCATAACAC	AGTACAAGCT	CTGTTTGCGG
801	TTGTAAAAAC TTTCTTTTGG	ACAAATGTTG	TTTGCATTTA	TGATCCGTTA	TATTTTGATC	TAATGTAGAG	TTGCACGTAG	TTCTTACTGG	CAAAGAAATC
	GATGCATACC AAAGAAAACC	TGTTTACAAC	AAACGTAAAT	ACTAGGCAAT	ATAAAACTAG	ATTACATCTC	AACGTGCATC	AAGAATGACC	GTTTCTTTAG
901	CTACGTATGG AAAAAAGAAT	AAAGGTGATA	TTTGATCTTT	ACCGTTTAGT	TCCAACGTAA	AATTGTGCCT	TTGGACTTAA	AATGGCGTCG	TACGTCTAAC
	TGATCTATCC TTTTTTTCTTA	TTTCCACTAT	AAACTAGAAA	TGGCAAATCA	AGGTTGCATT	TTAACACGGA	AACCTGAATT	TTACCGCAGC	ATGCAGATTG
1001	ACTAGATAGG AAAACTGAAA	ATTACATTCT	TGATTAGGTT	TATCACAGGC	AAATGTAATT	TGTGGTATTT	TGCCGTTCAA	AATCTGTAGA	ATTTTCTCAT
	TTTTGACTTT	TAATGTAAGA	ACTAATCCAA	ATAGTGTCCG	TTTACATTAA	ACACCATAAA	ACGGCAAGTT	TTAGACATCT	TAAAAGAGTA
1101	ACCAGIGIAA	AATACTTTAT	CTACAATCAT	ACCATTCTTA	TAACATGTCC	CCTTAATACT	AGGATCAGGC	ATGAACGCAT	CACAGACAAA
	TGTTGGACTT	TTATGAAATA	GATGTTAGTA	TGGTAAGAAT	ATTGTACAGG	GGAATTATGA	TCCTAGTCCG	TACTTGCGTA	GTGTCTGTTT
1201	CAAACGTCAC CCGGTGATAG	AATTGATCCC	TCCCCATCCG	TTATCACAAT	GACAGGTGTC	ATTTTGTGCT	CTTATGGGAC	GATCCTTATT	ACCGCTTCAT
	GTTTGCAGTG GGCCACTATC	TTAACTAGGG	AGGGGTAGGC	AATAGTGTTA	CTGTCCACAG	TAAAACACGA	GAATACCCTG	CTAGGAATAA	TGGCGAAGTA
1301	ACCGCCACAG TTTGCATTCA	AGGGGCAGAG	AGCAATCATC	ACCTGCAAAC	CCTTCTATAC	ACTCACATCT	ACCAGTGTAC	GAATTGCATT	CAGAAAACTG
	TGGCGGTGTC AAACGTAAGT	TCCCCGTCTC	TCGTTAGTAG	TGGACGTTTG	GGAAGATATG	TGAGTGTAGA	TGGTCACATG	CTTAACGTAA	GTCTTTTGAC
1401	AAAATAGGTA TCAAAGAATG	GCATACAATT	AAAACATGGC	GGGCATGTAT	CATTGCCCTT	ATCTTGTGCA	GTTAGACGCG	AATTTTTCGA	AGAAGTACCT
	TTTTATCCAT AGTTTCTTAC	CGTATGTTAA	TTTTGTACCG	CCCGTACATA	GTAACGGGAA	TAGAACACGT	CAATCTGCGC	TTAAAAAGCT	TCTTCATGGA
1501	GGGTCTTATC TACTGTAATT	TTGTTTTGCA	AGTACCACTG	AGCAGGATAA	TAATAGAAAT	GATAATATAC	TATAGTAGAG	ATAACGTCGA	TGACTTCCCA
1.601	CCCAGAATAG ATGACATTAA	AACAAAACGT	TCATGGTGAC	TCGTCCTATT	ATTATCTTTA	CTATTATATG	ATATCATCTC	TATTGCAGCT	ACTGAAGGGT
1601	TAGATTCCTG	GIGTATITIT	AGTGTGCAAG	TTTCTGTAAA	TCGATTAATT	TTTTTTTTTTT	TCCTCTTTTT	пласстта	
1701	ATCTAAGGAC	ACACCACACA	GATATTATAA	CATCTGCATA	AGCIAAIIAA	GCALGAATTA	CTCGTGAGTA	AGGAAAGAGT	CACCAACTAT
1,01	CGCATACCTG	TTCTGCGTGT	CTATAATATT	GTAGACGTAT	TATCCGTAAA	CGTTCTTAAT	GAGCACTCAT	TCCTTTCTCA	CTCCTTGATA
1801	GCGTATGGAC CATTTAAAGA	TGCCGATTTG	GGCGCGAATC	CTTTATTTG	GCTTCACCCT	CATACTATTA	TCAGGGCCAG	AAAAAGGAAG	TGTTTCCCTC
	CTTCTTGAAT GTAAATTTCT	ACGGCTAAAC	CCGCGCTTAG	GAAATAAAAC	CGAAGTGGGA	GTATGATAAT	AGTCCCGGTC	TTTTTCCTTC	ACAAAGGGAG
1901	GAAGAACTTA TGATGTTACC	CTCATAAAGC	ACGTGGCCTC	TTATCGAGAA	AGAAATTACC	GTCGCTCGTG	ATTTGTTTGC	AAAAAGAACA	AAACTGAAAA
	AACCCAGACA ACTACAATGG	GAGTATTTCG	TGCACCGGAG	AATAGCTCTT	TCTTTAATGG	CAGCGAGCAC	TAAACAAACG	TTTTTCTTGT	TTTGACTTTT
2001	TTGGGTCTGT CGCTCGACTT	CCTGTCTTCC	TATTGATTGC	AGCTTCCAAT	TTCGTCACAC	AACAAGGTCC	TAGCGACGGC	TCACAGGTTT	TGTAACAAGC
	AATCGAAGGT GCGAGCTGAA	GGACAGAAGG	ATAACTAACG	TCGAAGGTTA	AAGCAGTGTG	TTGTTCCAGG	ATCGCTGCCG	AGTGTCCAAA	ACATTGTTCG
2101	TTAGCTTCCA TCTGGAATGG	CGGGAAAGGG	TTTAGTACCA	CATGCTATGA	TGCCCACTGT	GATCTCCAGA	GCAAAGTTCG	TTCGATCGTA	CTGTTACTCT
	AGACCTTACC GAGAGAAAGT	GCCCTTTCCC	AAATCATGGT	GTACGATACT	ACGGGTGACA	CTAGAGGTCT	CGTTTCAAGC	AAGCTAGCAT	GACAATGAGA

2201	AACAGAATTG	TCCGAATCGT	GTGACAACAA	CAGCCTGTTC	TCACACACTC	TTTTCTTCTA	ACCAAGGGGG	TGGTTTAGTT	TAGTAGAACC
	TTGTCTTAAC	AGGCTTAGCA	CACTGTTGTT	GTCGGACAAG	AGTGTGTGAG	AAAAGAAGAT	TGGTTCCCCC	ACCAAATCAA	ATCATCTTGG
2301	AGCACTTTGA TACATTTACA	ТАТАТАТААА	CTTGCATAAA	TTGGTCAATG	CAAGAAATAC	ATATTTGGTC	TTTTCTAATT	CGTAGTTTTT	CAAGTTCTTA
	GATGCTTTCT				ammammama			0,010,011,111	
	CTACGAAAGA	AIAIAIAIII	GAACGIAIII	AACCAGIIAC	GIICIIIAIG	TATAAACCAG	AAAAGAIIAA	GCAICAAAAA	GIICAAGAAI
2401	TTTTCTCTTT CTACTTTCTC	TTTACAGATC	ATCAAGGAAG	TAATTATCTA	CTTTTTACAA	CAAATATAAA	ACCAAAAGAT	CAAGTCGACA	TGTTTGCTTT
	AAAAGAGAAA	AAATGTCTAG	TAGTTCCTTC	ATTAATAGAT	GAAAAATGTT	GTTTATATTT	TGGTTTTCTA	GTTCAGCTGT	ACAAACGAAA
2501	ACCGCATGCA	TCAGTTTGAA	GGGCGTTTTT	GGGGTGTCTC	CGAGTTACAA	TGGCCTTGGT	CTCACTCCAC	AGATGGGTTG	GGACAACTGG
	AATACGTTTG TGGCGTACGT	AGTCAAACTT	CCCGCAAAAA	CCCCACAGAG	GCTCAATGTT	ACCGGAACCA	GAGTGAGGTG	TCTACCCAAC	CCTGTTGACC
2601	TTATGCAAAC	CAGTGAACAG	CTACTTCTAG	ACACCGCTGA	ТАСААТТТСТ	GACTTGGGGC	TAAAGGATAT	GGGTTACAAG	тататсаттс
2002	TGGATGACTG		a						
	ACCTACTGAC	GICACIIGIC	GAIGAAGAIC	IGIGGCGACI	AICIIAAAGA	CIGAACCCCCG	ATTICCIAIA	CCCAAIGIIC	AIAIAGIAAG
2701	CTGGTCTAGC TAATAACAGC	GGCAGAGATT	CCGACGGTTT	CCTCGTTGCA	GATGAACAAA	AATTTCCCAA	TGGTATGGGC	CATGTTGCAG	ACCACCTGCA
	GACCAGATCG ATTATTGTCG	CCGTCTCTAA	GGCTGCCAAA	GGAGCAACGT	CTACTTGTTT	TTAAAGGGTT	ACCATACCCG	GTACAACGTC	TGGTGGACGT
2801	TTTCTTTTCG	GTATGTATTC	GTCTGCTGGT	GAGTACACCT	GTGCTGGATA	TCCTGGGTCT	CTGGGTCGTG	AGGAAGAAGA	TGCACAGTTC
	AAAGAAAAGC	CATACATAAG	CAGACGACCA	CTCATGTGGA	CACGACCTAT	AGGACCCAGA	GACCCAGCAC	TCCTTCTTCT	ACGTGTCAAG
2901	ACCGCGTTGA	CTACTTGAAG	TACGATAATT	GTTACAATAA	GGGTCAGTTT	GGTACACCGG	AAATTTCTTA	CCACCGTTAC	AAGGCCATGT
	CAGATGCTTT TGGCGCAACT	GATGAACTTC	ATGCTATTAA	CAATGTTATT	CCCAGTCAAA	CCATGTGGCC	TTTAAAGAAT	GGTGGCAATG	TTCCGGTACA
3001	GTCTACGAAA GAATAAAACT	GGTAGGCCTA	TATTCTATTC	TCTATGTAAC	TGGGGTCAGG	ΑΤΤΤΑΑCΑΤΤ	TTACTGGGGC	TCTGGTATCG	CCAATTCTTG
	GAGAATGAGT				20000200000			20200200200	00000330330
	CTCTTACTCA	CCAICCGGAI	AIAAGAIAAG	AGAIACAIIG	ACCCCAGICC	TAAATIGTAA	AAIGACCCCG	AGACCATAGC	GGIIAAGAAC
3101	GGAGATGTTA TCTATTATGA	CTGCTGAGTT	CACTCGTCCA	GATAGCAGAT	GTCCCTGTGA	TGGCGATGAA	TACGATTGCA	AGTACGCCGG	TTTCCATTGT
	CCTCTACAAT	GACGACTCAA	GTGAGCAGGT	CTATCGTCTA	CAGGGACACT	ACCGCTACTT	ATGCTAACGT	TCATGCGGCC	AAAGGTAACA
3201	ATATTCTTAA	CAAGGCAGCT	CCAATGGGGG	GATCCCCGGG	TTAATTAAGG	CGCGCCAGAT	CTGTTTAGCT	TGCCTCGTCC	CCGCCGGGTC
	TATAAGAATT	GTTCCGTCGA	GGTTACCCCC	CTAGGGGCCC	AATTAATTCC	GCGCGGTCTA	GACAAATCGA	ACGGAGCAGG	GGCGGCCCAG
3301	TGGGCCGGTC CGACATGGAG	GCCCAGAATA	CCCTCCTTGA	CAGTCTTGAC	GTGCGCAGCT	CAGGGGCATG	ATGTGACTGT	CGCCCGTACA	TTTAGCCCAT
	ACATCCCCAT GCTGTACCTC	CGGGTCTTAT	GGGAGGAACT	GTCAGAACTG	CACGCGTCGA	GTCCCCGTAC	TACACTGACA	GCGGGCATGT	AAATCGGGTA
24.01	TGTAGGGGTA	TTCCATCCAT		GGCCGCACGG	CCCCAACCAA	A A ATTACCCC	TCCTCCCTCC	AGACCTCCCA	CCACCCAAAC
3401	GCTCCCCTCA	IIGCAICCAI		GGCCGCACGG	COCOARGCAA	AAAIIACGGC		AGACCIGCGA	GCAGGGGAAAC
	CATATTAGTA CGAGGGGAGT	AACGTAGGTA	TGTAAAACTA	CCGGCGTGCC	GCGCTTCGTT	TTTAATGCCG	AGGAGCGACG	TCTGGACGCT	CGTCCCTTTG
3501	CAGACGCGTT CTTTTAAAAT	GAATTGTCCC	CACGCCGCGC	CCCTGTAGAG	AAATATAAAA	GGTTAGGATT	TGCCACTGAG	GTTCTTCTTT	CATATACTTC
	GTCTGCGCAA	CTTAACAGGG	GTGCGGCGCG	GGGACATCTC	TTTATATTTT	CCAATCCTAA	ACGGTGACTC	CAAGAAGAAA	GTATATGAAG
3601	CTTGCTAGGA	TACAGTTCTC	ACATCACATC	CGAACATAAA	CAACCATGGG	TAAGGAAAAG	ACTCACGTTT	CGAGGCCGCG	ATTAAATTCC
	GAACGATCCT	ATGTCAAGAG	TGTAGTGTAG	GCTTGTATTT	GTTGGTACCC	ATTCCTTTTC	TGAGTGCAAA	GCTCCGGCGC	TAATTTAAGG
3701	TTGTACCTAC CTGATTTATA	TGGGTATAAA	TGGGCTCGCG	ATAATGTCGG	GCAATCAGGT	GCGACAATCT	ATCGATTGTA	TGGGAAGCCC	GATGCGCCAG
	AGTTGTTTCT	ACCCATATT	ACCCGAGCGC	TATTACAGCC	CGTTAGTCCA	CGCTGTTAGA	ТАССТААСАТ	ACCCTTCGGG	CTACGCGGTC
2001	TCAACAAAGA								
3801	CAAGCATGGC	AAAGGTAGCG	TIGCCAATGA	TGTTACAGAT	GAGATGGTCA	GACTAAACTG	GCTGACGGAA	TTTATGCCTC	TTCCGACCAT
	CTTTGTACCG GTTCGTAAAA	TTTCCATCGC	AACGGTTACT	ACAATGTCTA	CTCTACCAGT	CTGATTTGAC	CGACTGCCTT	AAATACGGAG	AAGGCTGGTA
3901	ATCCGTACTC GAAAATATTG	CTGATGATGC	ATGGTTACTC	ACCACTGCGA	TCCCCGGCAA	AACAGCATTC	CAGGTATTAG	AAGAATATCC	TGATTCAGGT
	TAGGCATGAG	GACTACTACG	TACCAATGAG	TGGTGACGCT	AGGGGCCGTT	TTGTCGTAAG	GTCCATAATC	TTCTTATAGG	ACTAAGTCCA
4001	TTGATGCGCT	GGCAGTGTTC	CTGCGCCGGT	TGCATTCGAT	TCCTGTTTGT	AATTGTCCTT	TTAACAGCGA	TCGCGTATTT	CGTCTCGCTC
	AGGCGCAATC AACTACGCGA	CCGTCACAAG	GACGCGGCCA	ACGTAAGCTA	AGGACAAACA	TTAACAGGAA	AATTGTCGCT	AGCGCATAAA	GCAGAGCGAG
4101	TCCGCGTTAG ACGAATGAAT	AACGGTTTGG	TTGATGCGAG	TGATTTTGAT	GACGAGCGTA	ATGGCTGGCC	TGTTGAACAA	GTCTGGAAAG	AAATGCATAA
	GCTTTTGCCA	TTCCCANACC			CTCCTCCC T	TACCALCCCC		CACACCTTTC	TTTTACCTATT
	CGAAAACGGT	IIGCCAAACC	AACIACGCIC	ACTAAAACTA	CIGCICGCAI	IACCGACCGG	ACAACIIGII	CAGACCITIC	IIIACGIAII
4201	TTCTCACCGG GGACGAGTCG	ATTCAGTCGT	CACTCATGGT	GATTTCTCAC	TTGATAACCT	TATTTTTGAC	GAGGGGAAAT	TAATAGGTTG	TATTGATGTT
	AAGAGTGGCC CCTGCTCAGC	TAAGTCAGCA	GTGAGTACCA	CTAAAGAGTG	AACTATTGGA	ATAAAAACTG	CTCCCCTTTA	ATTATCCAAC	ATAACTACAA
4301	GAATCGCAGA	CCGATACCAG	GATCTTGCCA	TCCTATGGAA	CTGCCTCGGT	GAGTTTTCTC	CTTCATTACA	GAAACGGCTT	TTTCAAAAAT
	CTTAGCGTCT	GGCTATGGTC	CTAGAACGGT	AGGATACCTT	GACGGAGCCA	CTCAAAAGAG	GAAGTAATGT	CTTTGCCGAA	AAAGTTTTTA
4401	TACCATAACT TAATCCTGAT	ATGAATAAAT	TGCAGTTTCA	TTTGATGCTC	GATGAGTTTT	TCTAATCAGT	ACTGACAATA	AAAAGATTCT	TGTTTTCAAG
	AACTTGTCAT ATTAGGACTA	TACTTATTTA	ACGTCAAAGT	AAACTACGAG	CTACTCAAAA	AGATTAGTCA	TGACTGTTAT	TTTTCTAAGA	ACAAAAGTTC
	TTGAACAGTA								

	TTGTATAGTT	TTTTTATATT	GTAGTTGTTC	TATTTTAATC	AAATGTTAGC	GTGATTTATA	TTTTTTTCG	CCTCGACATC	ATCTGCCCAG
	ATGCGAAGTT AACATATCAA	ΑΑΑΑΑΤΑΤΑΑ	CATCAACAAG	ATAAAATTAG	TTTACAATCG	CACTAAATAT	AAAAAAAAGC	GGAGCTGTAG	TAGACGGGTC
4.6.01	TACGCTTCAA	<u>,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,</u>	CARCORCAA		A TO OTO OTO O		CTCC ATTCC A		CCCATCCACT
4601	GTCGAAAACG	AAAGIAAIAI	CAIGCGICAA	ICGIAIGIGA	AIGCIGGICG	CIAIACIGCI	GICGAIICGA	TACTAACGCC	GCCAICCAGI
	TTCACGCGTC CAGCTTTTGC	TTTCATTATA	GTACGCAGTT	AGCATACACT	TACGACCAGC	GATATGACGA	CAGCTAAGCT	ATGATTGCGG	CGGTAGGTCA
4701	AGCTCAGACC	ACCTGCATAA	TAACAGCTTT	CTTTTCGGTA	TGTATTCGTC	TGCTGGTGAG	TACACCTGTG	CTGGATATCC	TGGGTCTCTG
	TCGAGTCTGG	TGGACGTATT	ATTGTCGAAA	GAAAAGCCAT	ACATAAGCAG	ACGACCACTC	ATGTGGACAC	GACCTATAGG	ACCCAGAGAC
4801	CCAGCACTCC AAGAAGATGC	ACAGTTCTTT	GCAAATAACC	GCGTTGACTA	CTTGAAGTAC	GATAATTGTT	ACAATAAGGG	TCAGTTTGGT	ACACCGGAAA
	TTTCTTACCA	тстсаасааа	CGTTTATTCC	CGCAACTGAT	GAACTTCATC	статтаасаа	таттаттосс	AGTCAAACCA	TGTGGCCTTT
	AAAGAATGGT	IGICANGANA	COTTINITO	COCANCIONI	GAACIICAIG		IUIIMIICCC	NOICHAACCA	10100000111
4901	CCGTTACAAG CTGGGGGCTCT	GCCATGTCAG	ATGCTTTGAA	TAAAACTGGT	AGGCCTATAT	TCTATTCTCT	ATGTAACTGG	GGTCAGGATT	TAACATTTTA
	GGCAATGTTC	CGGTACAGTC	TACGAAACTT	ATTTTGACCA	TCCGGATATA	AGATAAGAGA	TACATTGACC	CCAGTCCTAA	ATTGTAAAAT
5001	GGTATCGCCA	ATTCTTGGAG	AATGAGTGGA	GATGTTACTG	CTGAGTTCAC	TCGTCCAGAT	AGCAGATGTC	CCTGTGATGG	CGATGAATAC
	CCATAGCGGT	TAAGAACCTC	TTACTCACCT	CTACAATGAC	GACTCAAGTG	AGCAGGTCTA	TCGTCTACAG	GGACACTACC	GCTACTTATG
5101	CTAACGTTCA ACGCCGGTTT	CCATTGTTCT	ATTATGAATA	TTCTTAACAA	GGCAGCTCCA	ATGGGGCAAA	ATGCAGGTGT	TGGTGGTTGG	AATGATCTGG
	ACAATCTAGA TGCGGCCAAA	GGTAACAAGA	TAATACTTAT	AAGAATTGTT	CCGTCGAGGT	TACCCCGTTT	TACGTCCACA	ACCACCAACC	TTACTAGACC
5201	TGTTAGATCT	GGGAATTTGA	СТСАССАТСА	GGAAAAGGCA	сатттстста	TGTGGGCAAT	GGTAAAGTCT	ССАСТТАТСА	ттастассаа
5202	TGTGAATAAC					202000000000000000000000000000000000000			
	ACACTTATTG	CCCTTAAACT	GACTGCTACT	CETTTTCCGT	GTAAAGAGAT	ACACCCGTTA	CCATTTCAGA	GGTGAATAGT	AACCACGGTT
5301	TTAAAGGCAT AGATATTATG	CTTCGTACTC	AATCTATAGT	CAAGCCTCTG	TCATCGCAAT	TAATCAAGAT	TCAAATGGTA	TTCCAGCAAC	AAGAGTCTGG
	AATTTCCGTA	GAAGCATGAG	TTAGATATCA	GTTCGGAGAC	AGTAGCGTTA	ATTAGTTCTA	AGTTTACCAT	AAGGTCGTTG	TTCTCAGACC
5401	TTTCAGACAC	AGATGAATAT	GGACAAGGTG	AAATTCAAAT	GTGGAGTGGT	CCTCTTGACA	ATGGTGATCA	AGTGGTTGCT	TTATTGAATG
	GAGGAAGCGT AAAGTCTGTG	TCTACTTATA	CCTGTTCCAC	TTTAAGTTTA	CACCTCACCA	GGAGAACTGT	TACCACTAGT	TCACCAACGA	AATAACTTAC
5501	CTCCTTCGCA ATCTAGACCA	ATGAACACGA	CCTTGGAAGA	GATTTTTTTT	GACAGCAATC	TGGGTTCAAA	GAAACTGACA	TCGACTTGGG	ATATCTACGA
	CCTATGGGCC	TACTTGTGCT	GGAACCTTCT	CTAAAAAAAA	CTGTCGTTAG	ACCCAACTTT	CTTTGACTGT	AGCTGAACCC	татасатост
	GGATACCCGG								
5601	AACAGAGTTG AAAGACGGTT	ACAACTCGAC	AGCGTCTGCT	ATCCTTGGAC	GGAATAAGAC	AGCCACCGGT	ATTCTCTACA	ATGCTACGGA	GCAATCCTAC
	TTGTCTCAAC TTTCTGCCAA	TGTTGAGCTG	TCGCAGACGA	TAGGAACCTG	CCTTATTCTG	TCGGTGGCCA	TAAGAGATGT	TACGATGCCT	CGTTAGGATG
5701	TGTCTAAGAA	TGATACAAGA	CTGTTTGGTC	AGAAAATTGG	TAGTCTTTCT	CCAAATGCTA	TACTTAACAC	GACTGTTCCA	GCTCACGGTA
	ACAGATTCTT	ACTATGTTCT	GACAAACCAG	TCTTTTAACC	ATCAGAAAGA	GGTTTACGAT	ATGAATTGTG	CTGACAAGGT	CGAGTGCCAT
	AGCGGAAGAT			TTCACCAAAC		AGTATTGATG	ATTGTTAAAA	AGTTCATGAA	АААААТАСТА
5801	TAGGTTGAGA	CCCTCTTCTT	GAGCTTATTG	I I GAGCAAAG	CAGGGCGAGA				
5801	TAGGTTGAGA CTCGAATATT ATCCAACTCT	CCCTCTTCTT GGGAGAAGAA	GAGCTTATTG CTCGAATAAC	AACTCGTTTC	GTCCCGCTCT	TCATAACTAC	TAACAATTTT	TCAAGTACTT	TTTTTATGAT
5801	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT	CCCTCTTCTT GGGAGAAGAA	GAGCTTATTG CTCGAATAAC	AACTCGTTTC	GTCCCGCTCT	TCATAACTAC	TAACAATTTT	TCAAGTACTT	TTTTTTATGAT
5901	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTTCT	CCCTCTTCTT GGGAGAAGAA AACTAAATAA	GAGCTTATTG CTCGAATAAC TAAACGACAG	AACTCGTTTC	GTCCCGCTCT CAGGTATTCC	TCATAACTAC AATAGTTTTC	TAACAATTTT	TCAAGTACTT TACATAATCT	TTTTTATGAT GAAGCCCTTG
5801	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTTCT ATAAGTCTCA TTGAAAAAGA	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC	AACTCGTTTC AATAGCCTAT TTATCGGATA	GTCCCGCTCT CAGGTATTCC GTCCATAAGG	ТСАТААСТАС ААТАGTTTTC ТТАТСААААG	TAACAATTTT GTTTTGTAGG CAAAACATCC	TCAAGTACTT TACATAATCT ATGTATTAGA	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC
5901 5901 6001	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTTCT ATAAGTCTCA TTGAAAAGA CGTTTACATA	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT	ТСАТААСТАС ААТАДТТТТС ТТАТСААААД GACTTTTGTA	ТААСААТТТТ GTTTTGTAGG CAAAACATCC AAATACTTTT	ТСААДТАСТТ ТАСАТААТСТ АТДТАТТАДА СТДССААТДТ	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA
5901 5901 6001	TAGGTTGAGA CTCGAATATT ATCCAACTCT GACCTTATAA TATTCAGAGT AACTTTTTCT ATAAGTCTCA TTGAAAAAGA CGTTTACATA TTGTATGACA GCAAATGTAT	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA	ТСАТААСТАС ААТАGTTTTС ТТАТСААААG GACTTTTGTA СТGAAAACAT	ТААСААТТТТ GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT
5901 5901 6001 6101	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAAGTCTCA TTGAAAAAGA CGTTATCATA TGTATGACA GCAAATGTAT TACAATCTGT TTTCACGAAG	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC	ТСАТААСТАС ААТАGTTTTС ТТАТСААААG GACTTTTGTA СТGAAAACAT ATCATATCAG	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG
5901 6001 6101	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT ATAGTCTCA TTGAAAAAGA CGTTTACAATA TTGTATGACA GCAAATGTAT AACATACTGT TTTCACGAAG CTAGTCTACA AAAGTGCTTC	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG	ТСАТААСТАС ААТАGTTTTС ТТАТСААААG GACTTTTGTA СТGААААСАТ АТСАТАТСАG ТАGTATAGTC	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC
5901 5901 6001 6101 6201	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAGTCTCA TTGAAAAAGA CGTTTACATA GCAAATGTAT AACATACTGT TTTCACGAAG CTTAGCTAC AAAGTGCTTC GATCACATCTG GGTATCTCTG	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC	ТСАТААСТАС ААТАGTTTTС ТТАТСААААG GACTTTTGTA СТGAAAACAT АТСАТАТСАG ТАGTATAGTC TGGGTCCCGA	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CCATTGGTCG
5901 5901 6001 6101 6201	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAGTCTCA TTGAAAAGG CGTTTACAAA GCAAATGTAT AACATACTGT GATCAGAGTGT GGTATCTCTG AACAATTGTT AACAATTGTT	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CCATTGGTCG
5901 5901 6001 6101 6201	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAGTCTCA TTGAAAAAGA CGTTACATA TGTATGACA GCAATGTGT TTCCACGAAG CTAGTCTACA AAAGTGCTTC GATCAGATGTT CCATCAGATGT TCCATAGACA	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC TTTGTTTTTG	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCCAAGC	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CCATTGGTCG GGTAACCAGC
5901 5901 6001 6101 6201 6301	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAAGTCTCA TTGAAAAAGA CGTTATCATA TTGTATGACA GCAAATGTAT TTCACGAAG CTAGTCTACA AAAGTGCTTC GATCAGATGTT GGTATCTCTG AACAATGTT CCATAGAGAC TTGTTAACAA	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG GCTCGCTCAT	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC TTTGTTTTTG AGCCGGTTTC	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG AAGTAAGCCC	ТСАТААСТАС ААТАGTTTTС ТТАТСААААG GACTTTTGTA СТGААААСАТ АТСАТАТСАG ТАGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CCATTGGTCG GGTAACCAGC AAAAGGGAGT
5901 5901 6001 6101 6201 6301	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAAGTCTCA TTGAAAAAGA CGTTATCATA TGTATGACA GCAAATGTAT TTCCACGAAG CTAGTCTACA GAACATCGTT GGTCACGATGT CCATAGAGAC TTGTTAACAA TTTGTCACCATCG AACAATCGTAC ATCACCATCG AACAACGATGGT	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG GCTCGCTCAT CGAGCGAGTA	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG AAGTAAGCCC TTCATTCGGG	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGA	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA TATCCCAGAT	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CCATTGGTCG GGTAACCAGC AAAAGGGAGT TTTTCCCTCA
5901 5901 6001 6101 6201 6301 6401	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAGTCTCA TTGAAAAGA CGTTTACAA GCAAATGTAT AACATACTGT TTTCACGAAG GCTAGTCTACA AAAGTGCTTC GATCAGAGAGTGT CCATAGAGAC TTGTCACAATGGT ACAATAGTG ATCACACATGG AAAACGATGG AAAACGATGG AAAACGATGG AAGAGGTGCCA	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG GCTCGCTCAT CGAGCGAGTA GATAGTCCTT	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG AGAGAACAGT	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC CATCAGATGT	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG AAGTAAGCCC TTCATTCGGG CCATATATTA	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGA CTACTACCCT	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT TCTAGACCTG	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA TATCCCAGAT ACCACGTACT	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CCATTGGTCG GGTAACCAGC AAAAGGGAGT TTTTCCCTCA GTTGGATGCA
5901 5901 6001 6101 6201 6301 6401	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAGTCTCA TTGAAAAGA CGTTTACAA GCAATGTAT AACATACTGT TTTCACGAAG GTAGTCTACA AAAGTGCTTC GATCAGAGTG GGTATCTCTG AACAATTGT CCATAGAGATG TTGTAACAA TTTGCTACCA ATGGGTAGC AACGCTTCCT TGGTACCACG ATGGGTAGC AACGCTTCCT	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG GCTCGCTCAT CGAGCGAGTA GATAGTCCTT CTATCAGGAA	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG AGAGAACAGT TCTCTTGTCA	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC CATCAGATGT GTAGTCTACA	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG AAGTAAGCCC TTCATTCGGG CCATATATTA GGTATATAAT	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGA CTACTACCCT GATGATGGGA	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT TCTAGACCGG	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA TATCCCAGAT ACCACGTACT TGGTGCATGA	TTTTTATGAT GAAGCCCTTG GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CGTTTGGTCG GGTAACCAGC AAAAGGGAGT TTTTCCCTCA GTTGGATGCA CAACCTACGT
5901 5901 6001 6101 6201 6301 6401 6501	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAGTCTCA TTGAAAAGA CGTTTACAATA TTGTATGACA GCAATGTGT TTCCACGAAG CTAGTCTACA AAAGTGCTTC GATCAGATGT GGTATCTCG AACAATTGT CCATAGAGAC TTGTTAACAA TTTGTAACAA TTTGTACAAC ATGTGTACCATCG TAGGGTAGC TGGTACCCATCG TGGTACCAA TTGGCAAGGT AACCATCGTT	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG GCTCGCTCAT CGAGCGAGTA GATAGTCCTT CTATCAGGAA AATTCCTGGG	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG AGAGAACAGT TCTCTTGTCA GGAACAACTT	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC CATCAGATGT GTAGTCTACA CACAGAATGT	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG AAGTAAGCCC TTCATTCGGG CCATATATTA GGTATATAAT	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGA CTACTACCCT GATGATGGGA TGTCGAAGTG	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT TCTAGACCTG AGATCTGGAC GTCACAAAAC	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA TATCCCAGAT ACCACGTACT TGGTGCATGA AAGAGAAGTT	TTTTTATGAT GAAGCCCTTG GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CGTCTTCCGC GGTAACCAGC AAAAGGGAGT TTTTCCCTCA GTTGGATGCA CAACCTACGT CCGCCAATTA
5901 5901 6001 6101 6201 6301 6401 6501	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATCAGAGT AACTTTTCT ATAAGTCTCA TTGAAAAAGA CGTTATCATA TTGTATGACA GCAATGTATT AACATACTGT TTTCACGAAG CTAGTCTACA AAAGTGCTTC GATCAGATGT CCATAGAGATG TAGTGTACCAC ATGTGTACCA ACACTTCCT TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TGGTACCAC TTTTGCCAC TAAGGGC TAAAAAGGGC	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGA AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG GCTCGCTCAT CGAGCGAGTA GATAGTCCTT CTATCAGGAA AATTCCTGGG TTAAGGACCC	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG AGAGAACAGT TCTCTTGTCA GGAACAACTT CCTTGTTGAA	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC CATCAGATGT GTAGTCTACA	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG AAGTAAGCCC TTCATTCGGG CCATATATAAT TTTGTCATAT AAACAGTATA	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGA CTACTACCCT GATGATGGGA TGTCGAAGTG ACAGCTTCAC	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT TCTAGACCTG AGATCTGGAC GTCACAAAAC CAGTGTTTTG	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA TATCCCAGAT ACCACGTACT TGGTGCATGA AAGAGAAGTT TTCTCTTCAA	TTTTTATGAT GAAGCCCTTG GATCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CGTTTCCGC GGTAACCAGC AAAAGGGAGT TTTTCCCTCA GTTGGATGCA CAACCTACGT CCGCCAATTA GGCGGTTAAT
5801 5901 6001 6101 6201 6301 6401 6501	TAGGTTGAGA CTCGAATATT TATCCAACTCT GAGCTTATAA TATTCAGAGT AACTTTTCT ATAAGTCTCA TTGAAAAAGA CGTTATCATA TTGTATGACA GCAAATGTAT TTCCACGAAG CTAGTCTACA AACATGCTTC GATCAGATGT GGTATCTCTG AACAATGTT CCATAGAGAC TTGTTAACAA TTTGGTACCA AACGCTTCCT TTGGTACCA TTGGTAGCAG AACATGGT TTTTGCCG TAAAAGGGT TTTTTTTTTCCCT	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG GGCTCGCTCAT CTATCAGGAA AATTCCTGGG TTAAGGACCC TTCAGGCTTCA	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG AGAGAACAGT TCTCTTGTCA GGAACAACTT CCTTGTTGAA	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC CATCAGATGT GTAGTCTACA CACAGAATGT GTGTCTTACA	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG AAGTAAGCCC TTCATTCGGG CCATATATTA GGTATATAAT TTTGTCATAT AAACAGTATA AGTACTGTAT	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGA CTACTACCCT GATGATGGGA TGTCGAAGTG ACAGCTTCAC	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT TCTAGACCTG AGATCTGGAC GTCACAAAAC CAGTGTTTTG	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA TATCCCAGAT ACCACGTACT TGGTGCATGA AAGAGAAGTT TTCTCTTCAA	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CGTCTTCCGC GGTAACCAGC AAAAGGGAGT TTTTCCCTCA GTTGGATGCA CGCCGATTAA GGCGGTTAAT
5901 5901 6001 6101 6201 6301 6401 6501 6601	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATCAAGATCT ATAGTCTCA TTGAAAAGA CGTTTACTA TGTATGACAA GCAATGTAT AACATACTGT TTTCACGATG GTATCTCTG GATCAGATGT GGTATCTCTG AACAATGTT CCATAGAGATGT TGTAACAA TTTGCTACCA ATGTGGTAGC ACCACGTTCT TTGGCAAGGA GACAGCTTCCT TTGGCAAGGA AACCATGGTA TTTGCCACCATGGT TTTTGCCACCATG TAAAAGGGC AACCATGGTA TTTTTGCCCG TAAAAAGGGA AAAACGGC ATTTTTCCCT CACCATAT	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG GATCGCTCAT CGAGCGAGTA AATTCCTGGG TTAAGGACCC TTCAGCTTCA	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG AGAGAACAGT TCTCTTGTCA GGAACAACTT CCTTGTTGAA	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC CATCAGATGT GTAGTCTACA CACAGGATGT GTGTCTTACA	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG CCC TTCATTCGGG CCATATATTA GGTATATAAT TTTGTCATAT AAACAGTATA AGTACTGTAT	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGA CTACTACCCT GATGATGGGA TGTCGAAGTG ACAGCTTCAC	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT TCTAGACCTG AGATCTGGAC GTCACAAAAC CAGTGTTTTG GATAGAAGGC	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT TATCCCAGAT ACCACGTACT TGGTGCATGA AAGAGAAGTT TTCTCTTCAA CATAAAAATT	TTTTTATGAT GAAGCCCTTG CTTCGGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CCATTGGTCG GGTAACCAGC AAAAGGGAGT TTTTCCCTCA GTTGGATGCA CCACCTACGT CCGCCAATTA GGCGGTTAAT
5901 5901 6001 6101 6201 6301 6401 6501 6601	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATCAGAGT AACTTTTCT ATAGTCTCA TTGAAAAAGA CGTTTACAATA TTGTATGACA GCAATGTTT AACATACTGT GGTATCTCCG AACAGATGTT CCATCAGAGAC TTGTTAACAA TTTGCCACG AACGCATCCT TTGGTAACCATG AACAGATGGT ACCATGGTACCA TTGGTACCAA GTAGTCCTCT TTGGTACCAA TTTTGCCCG TAAAAAGGG AACCATGGT TTTTTGCCCG TAAAAAGGG AACACCATGGT TTTTTGCCCG TAAAAAGGGC AACACCATGGT TTTTTCCCT ACCCGTATAT CAACCAATAT	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCAACTATG GCTCGCTCAT CGAGCGAGTA GATAGTCCTTG CTATCAGGAA AATTCCTGGG TTAAGGACCC TTCAGCTTCA AAGTCGAAGT	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG AGAGAACAGT TCTCTTGTCA GGAACAACTT CCTTGTTGAA CGGATGATTT GCCTACTAAA	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC CATCAGATGT GTAGTCTACA CACAGAATGT GTGTCTTACA CCAGGGTGAG GGTCCCACTC	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG CCATATATAAT TTTGTCATATA AAACAGTATA AGTACTGTAT TCATGACATA	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGA CTACTACCCT GATGATGGGA TGTCGAAGTG ACAGCTTCAC ATGGGCTTAC TACCCGAATG	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT TCTAGACCTG AGATCTGGAC GTCACAAAAC CAGTGTTTTG GATAGAAGGC CTACTTCCG	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA TATCCCAGAT ACCACGTACT TGGTGCATGA AAGAGAAGTT TTCTCTTCAA CATAAAAATT GTATTTTAA	TTTTTATGAT GAAGCCCTTG GATTCTAGAA CTTAGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG GGTAACCAGC AAAAGGGAGT TTTTCCCTCA GTTGGATGCA CAACCTACGT CCGCCAATTA GGCGGTTAAT TCTTGCTTGG AGAACGAACC
5901 5901 6001 6101 6201 6301 6401 6501 6601 6701	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATCAGAGT AACTTTTCT ATAGTCTCA TTGAAAAAGA CGTTACATA TTGTATGACA GCAATGTGT TTCCACGAAG CTAGTCTACA AAAGTGCTTC GATCAGATGT GGTATCTCG AACAATTGT CCATAGAGAC TTGTTAACAA TTTGGCACATCG TAGGGTAGC TAGGGAAGGA AAACCATCGT TTGGGAAGGA AACCATGGTT TTGGGAAGGA AAAACGGC TATTTTCCCT ACACCATCG TAAAAAGGCA	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGT AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCCACTATG GATAGTCCTT CTATCAGGAA AATTCCTGGG TTCAGGTCCA AAGTCGAAGT CATGTCGAGG	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG AGAGAACAGT TCTCTTGTCA GGAACAACTT CCTTGTTGAA CGGATGATTT GCCTACTAAA	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC CATCAGATGT GTAGTCTACA CACAGAATGT GTGTCTTACA CCAGGGTGAG GGTCCCACTC GGAGAACAGC	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG CCATATATAAG GGTATATAAT TTTGTCATAT AAACAGTATA AGTACTGTAT TCATGACATA TTAAAATATC	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGA CTACTACCCT GATGATGGGA TGTCGAAGTG ACAGCTTCAC TACCCGAATG ACAAAAAAAG	TAACAATTTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT TCTAGACCTG AGATCTGGAC GATAGTTTG GATAGAAGC CAGTGTTTTG GATAGAAGC CTATCTTCCG AATCTAAAAC	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA TATCCCAGAT ACCACGTACT TGGTGCATGA AAGAGAAGTT TTCTCTTCAA CATAAAAATT GTATTTTAA	TTTTTATGAT GAAGCCCTTG GATTCTAGAA CTTAGGAAC GATTCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CGTTTGCCCGC AAAAGGGAGC CAACCTACGT CCGCCAATTA GGCGGTTAAT TCTTGCTTGG AGAACGAACC TTGTCCCAGA
5801 5901 6001 6101 6201 6301 6401 6501 6601 6701	TAGGTTGAGA CTCGAATATT ATCCAACTCT GAGCTTATAA TATCAAGATCT ATAGTCTCA TTGAAAAGA CTTTTCT ATAAGTCTCA TTGAAAAAGA CGTTATCACA TTGTATGACA GCAAATGTAT TTCCACGAAG CTAGTCTACA AAAGTGCTTC GATCAGATGT CCATAGAGAC TTGTTAACAA TTTGCTACCA ACACTCCT TTGGTACCCA TTTTTGCCCCG TAAAAAGGGA AACAATGGTC CACCAATAG TTGGTATATA GTGGTATATA GTGGTATATA	CCCTCTTCTT GGGAGAAGAA AACTAAATAA TTGATTTATT CTTCATTGCA GAAGTAACGA AGGAACAACA TCCTTGTTGT TAGTTGATAC ATCCACTATG GATCACTCATG CTATCAGGAA AATTCCTGGG TTAAGGACCC TTCAGCTTCA AAGTCGAAGT CATGTCGAGG GTACAGCTCC	GAGCTTATTG CTCGAATAAC TAAACGACAG ATTTGCTGTC TTAGCGATAT AATCGCTATA GCTTCAGGAG CGAAGTCCTC AAACAAAAAAC TTTGTTTTTG AGCCGGTTTC TCGGCCAAAG GGAACAACTT CCTTGTTGAA CGGATGATTA GCCTACTAAA CTGCTGTGTG GACGACACAC	AACTCGTTTC AATAGCCTAT TTATCGGATA TTCACATGTG AAGTGTACAC TACATACAAA ATGTATGTTT AAAGAGTTCG TTTCTCAAGC AGTAAAATTG TCATTTTAAC CATCAGATGT GTAGTCTACA CACAGAATGT GTGTCTTACA CCAGGGTGAG GGTCCCACTC GGAGAACAGC CCTCTTGTCG	GTCCCGCTCT CAGGTATTCC GTCCATAAGG CTATACTAGT GATATGATCA CGACGAAATC GCTGCTTTAG AACACATTTC TTGTGTAAAG AAGTAAGCCC TTCATTCGGG CCATATATAA GGTATATAAAT TTTGTCATAT AAACAGTATA AGTACTGTAT TCATGACATA TTAAAATATC AATTTTATAG	TCATAACTAC AATAGTTTTC TTATCAAAAG GACTTTTGTA CTGAAAACAT ATCATATCAG TAGTATAGTC TGGGTCCCGA ACCCAGGGCT AGTTTCATCT TCAAAGTAGGA TGTCGAAGTG ACAGCTTCAC ATGGGCTTAC TACCCGAATG ACAAAAAAG TGTTTTTTC	TAACAATTT GTTTTGTAGG CAAAACATCC AAATACTTTT TTTATGAAAA CAAAACCGGA GTTTTGGCCT GTAGTGGCGT CATCACCGCA GAATTTGAGA CTTAAACTCT AGATCTGGAC GACACAAAAC CAGTGTTTTG GATAGAAGC CATCTTCCG AATCTAAAAC TTAGATTTG	TCAAGTACTT TACATAATCT ATGTATTAGA CTGCCAATGT GACGGTTACA TATCGAAGAA ATAGCTTCTT TGGTGTTCCA ACCACAAGGT ATAGGGTCTA TATCCCAGAT ACCACGTACT TGGTGCATGA CATAAAATT GTATTTTAA ACTGTGTTGC TGGACACAACG	TTTTTATGAT GAAGCCCTTG GATCTAGAA CTAAGATCTT GCAGAAGGCG CGTCTTCCGC CGTTTCCGC GGTAACCAGC AAAAGGGAGT CTTTCCCTCA GTTGGATGCA CGACCTACGT CCGCCAATTA GGCGGTTAAT TCTTGCTGG AGAACGAACC TTGTCCCAGA AACAGGGTCT

	AGTATTTTA	TAAAGATTGG	AGTGGTAAAA	ATCGAGTATG	TGCTAGATGC	TATGGAAGAT	ACAAATTCAG	CGGTCATCAC	TGTATAAATT
	GCAAGTATGT	3 TTTCT 3 3 C C	TCACCATTT	TACCTCATAC	ACGATCTACC	3 T 3 C C T T C T 3	TOTTANCTO	CCACTACTC	አ ር አ ጥ አ ጥጥጥ አ አ
	CGTTCATACA	ATTICIANCE	ICACCATITI	INGCICATAC	ACGAICIACG	AIACCIICIA	IGIIIAAGIC	GCCAGIAGIG	ACATATITAA
6901	ACCAGAAGCA	CGTGAAGTGA	AAAAGGCAAA	AGACAAAGGC	GAAAAATTGG	GCATTACGCC	CGAAGGTTTG	CCAGTTAAAG	GACCAGAGTG
	TGGTCTTCGT	GCACTTCACT	TTTTCCGTTT	TCTGTTTCCG	CTTTTTAACC	CGTAATGCGG	GCTTCCAAAC	GGTCAATTTC	CTGGTCTCAC
	ATATTTTACA								
7001	GGCGGGAATCT	TACAGTGGCC	TATGCGGCCG	CTCTAGAACT	AGTGGATCGA	TCCCCAATTC	GCCCTATAGT	GAGTCGTATT	ACAATTCACT
	CCGCCTTAGA	ATGTCACCGG	ATACGCCGGC	GAGATCTTGA	TCACCTAGCT	AGGGGTTAAG	CGGGATATCA	CTCAGCATAA	TGTTAAGTGA
71.01	CCGGCAGCAA	GTGA CTGGGA	AAACCCTCCC	GTTACCCAAC	TTANTCCCCT	TGCAGCACAT	CCCCCTTTCC	CCACCTCCCC	TATACCAA
/101	GAGGCCCGCA	GIGACIGGGA	AAACCCIGGC	GIIACCCAAC	TIAATCGCCT	IGCAGCACAI	ceccerried	CCAGCIGGCG	IAAIAGCGAA
	AATGTTGCAG	CACTGACCCT	TTTGGGACCG	CAATGGGTTG	AATTAGCGGA	ACGTCGTGTA	GGGGGAAAGC	GGTCGACCGC	ATTATCGCTT
7201	CCGATCGCCC	TTCCCAACAG	TTGCGCAGCC	TGAATGGCGA	ATGGCGCCTG	ATGCGGTATT	TTCTCCTTAC	GCATCTGTGC	GGTATTTCAC
	ACCGCATACG	ACCOURCEC	A A COCOTOCO		TACCCCCA			COTACACACO	CCATAAACTC
	TGGCGTATGC	AAGGGIIGIC	AACGCGICGG	ACTIACCOCT	INCEGEGGAE	IACGCCAIAA	AAGAGGAAIG	CUINGACACG	CCATAAAGIG
7301	TCAAAGCAAC	CATAGTACGC	GCCCTGTAGC	GGCGCATTAA	GCGCGGCGGG	TGTGGTGGTT	ACGCGCAGCG	TGACCGCTAC	ACTTGCCAGC
	AGTTTCGTTG	GTATCATGCG	CGGGACATCG	CCGCGTAATT	CGCGCCGCCC	ACACCACCAA	TGCGCGTCGC	ACTGGCGATG	TGAACGGTCG
7401	CGGGATCGCG	CGCTTTCTTC	CCTTCCTTTC	TCGCCACGTT	CGCCGGCTTT	CCCCGTCAAG	CTCTABATCG	GGGTGGGCCA	TCGCCCTGAT
	AGACGGTTTT								
	GGCGAGGAAA	GCGAAAGAAG	GGAAGGAAAG	AGCGGTGCAA	GCGGCCGAAA	GGGGCAGTTC	GAGATTTAGC	CCCACCCGGT	AGCGGGACTA
7501	TCGCCCTTTG	ACGTTGGAGT	CCACGTTCTT	TAATAGTGGA	CTCTTGTTCC	AAACTGGAAC	AACACTCAAC	CCTATCTCGG	GCTATTCTTT
	TGATTTATAA AGCGGGAAAC	TGCAACCTCA	GGTGCAAGAA	ATTATCACCT	GAGAACAAGG	TTTGACCTTG	TTGTGAGTTG	GGATAGAGCC	CGATAAGAAA
	ACTAAATATT								
7601	ATTTTATGGT	CGATTTCGGC	CTATTGGTTA	AAAAATGAGC	TGATTTAACA	AAAATTTAAC	GCGAATTTTA	ACAAAATATT	AACGTTTACA
	CCCTAAAACG	GCTAAAGCCG	GATAACCAAT	TTTTTACTCG	ACTAAATTGT	TTTTAAATTG	CGCTTAAAAT	TGTTTTATAA	TTGCAAATGT
7701	GCACTCTCAG	TACAATCTGC	TCTGATGCCG	CATAGTTAAG	CCAGCCCCGA	CACCCGCCAA	CACCCGCTGA	CGCGCCCTGA	CGGGCTTGTC
	TGCTCCCGGC								
	ACGAGGGCCG	ATGTTAGACG	AGACTACGGC	GTATCAATTC	GGTCGGGGGCT	GTGGGCGGTT	GIGGGCGACI	GCGCGGGGACT	GCCCGAACAG
7801	ATCCGCTTAC	AGACAAGCTG	TGACCGTCTC	CGGGAGCTGC	ATGTGTCAGA	GGTTTTCACC	GTCATCACCG	AAACGCGCGA	GACGAAAGGG
	TAGGCGAATG	TCTGTTCGAC	ACTGGCAGAG	GCCCTCGACG	TACACAGTCT	CCAAAAGTGG	CAGTAGTGGC	TTTGCGCGCT	CTGCTTTCCC
7001	GGAGCACTAT							» mama a a a a a	3 3 C C C C T 3 T T
7901	TGTTTATTTT	TATAGGTTAA	IGICAIGAIA	AIAAIGGIII	CIIAGACGIC	AGGIGGCACI	TTTCGGGGGAA	AIGIGCGCGG	AACCCCIAII
	GCGGATAAAA	ATATCCAATT	ACAGTACTAT	TATTACCAAA	GAATCTGCAG	TCCACCGTGA	AAAGCCCCTT	TACACGCGCC	TTGGGGATAA
8001	ACAAATAAAA TCTAAATACA	TTCAAATATG	TATCCGCTCA	TGAGACAATA	ACCCTGATAA	ATGCTTCAAT	AATATTGAAA	AAGGAAGAGT	ATGAGTATTC
	AACATTTCCG							mmaammamaa	
	TTGTAAAGGC	AAGIIIAIAC	ATAGGCGAGI	ACICIGIIAI	IGGGACIAII	TACGAAGITA	TTATAACTIT	TICCITCICA	TACICATAAG
8101	TGTCGCCCTT	ATTCCCTTTT	TTGCGGCATT	TTGCCTTCCT	GTTTTTGCTC	ACCCAGAAAC	GCTGGTGAAA	GTAAAAGATG	CTGAAGATCA
	ACAGCGGGGAA	TAAGGGAAAA	AACGCCGTAA	AACGGAAGGA	CAAAAACGAG	TGGGTCTTTG	CGACCACTTT	CATTTTCTAC	GACTTCTAGT
9201	CAACCCACGT		GGATCTCAAC	2000002202	maammaaaaaa	TTTTCCCCCC	CAACAACCTT	TTCCAATCAT	
0201	CGWGIGGGII	N / · N · · / · / N / / · · ·		N/ · / · / · / · · · · N N / · N					
	AAAGTTCTGC	ACATCGAACT	GGAICICAAC	AGCGGTAAGA	ICCIIGAGAG	11110000000		TICCAAIGAI	GAGCACITI
	AAAGTTCTGC GCTCACCCAA	TGTAGCTTGA	CCTAGAGTTG	TCGCCATTCT	AGGAACTCTC	AAAAGCGGGGG	CTTCTTGCAA	AAGGTTACTA	CTCGTGAAAA
8301	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC	TGTAGCTTGA GGTATTATCC	CCTAGAGTTG CGTATTGACG	TCGCCATTCT CCGGGCAAGA	AGGAACTCTC GCAACTCGGT	AAAAGCGGGG CGCCGCATAC	CTTCTTGCAA ACTATTCTCA	AAGGTTACTA GAATGACTTG	CTCGTGAAAA GTTGAGTACT
8301	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG	ACATEGAACT TGTAGCTTGA GGTATTATCC CCATAATAGG	CCTAGAGTTG CGTATTGACG GCATAACTGC	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT	AGGAACTCTC GCAACTCGGT	AAAAGCGGGG CGCCGCATAC	CTTCTTGCAA ACTATTCTCA TGATAAGAGT	AAGGTTACTA GAATGACTTG	GAGCACITIT CTCGTGAAAAA GTTGAGTACT
8301	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG	ACATEGAACT TGTAGCTTGA GGTATTATCC CCATAATAGG	CCTAGAGTTG CGTATTGACG GCATAACTGC	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT	AGGAACTCTC GCAACTCGGT CGTTGAGCCA	AAAAGCGGGG CGCCGCATAC GCGGCGTATG	CTTCTTGCAA ACTATTCTCA TGATAAGAGT	AAGGTTACTA GAATGACTTG CTTACTGAAC	CTCGTGAAAA GTTGAGTACT CAACTCATGA
8301 8401	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG AGAAAAGCAT GACAACGATC	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGCCA	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT
8301 8401	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG AGAAAAGCAT GACAACGATC TCTTTTCGTA	ACATCGAACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGCCA TGACGCCGGT	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA
8301 8401 8501	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG GACAACGATC TCTTTTCGTA CTGTTGCTAG GGAGGACCGA	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGCCA TGACGCCGGT AACCGGAGCT	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGAC
8301 8401 8501	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC GTGGTCAGTG GGAGAACGATC TCTTTCGTA CTGTTGCTAG GGAGGACCGA ATACCAAACG	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGCCA TGACGCCGGT AACCGGAGCT	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC
8301 8401 8501	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC GTGGTCAGTG GGAGAACGATC TCTTTCGTA CTGTTGCTAG GGAGGACCGA ATACCAAACG CCTCCTGGCT TATGGTTGCC	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGGCCA TGACGCGGGCT AACCGGAGCT TTGGCCTCGA	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG
8301 8401 8501 8601	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC GTGGTCAGTG GGAGAACGATC TCTTTCGTA CTGTTGCTAG GGAGGACCGA ATACCAAACG CCTCCTGGCT ACGAGGCGTGA	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGGCCA TGACGCGGGCCA TTGGCCTCGA TACTCTAGCT	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC
8301 8401 8501 8601	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC GTGGTCAGTG GGAGAAAAGCAT GACAACGATC CTGTTGCTAG GGAGGACCGA ATACCAAACG CCTCCTGGCT TATGGTTTGC ACGAGCGTGA AATTAATAGA	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGGCCA TGACGCGGGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGGCCGTTG
8301 8401 8501 8601	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC GTGGTCAGTG AGAAAAGCAT GACAACGATC CTGTTGCTAG GGAGGACCGA ATACCAAACG CCTCCTGGCT TATGGTTTGC ACGAGCGTGA AATTAATAG TGCTCGCACT	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGGCCA TGACGCGGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGGCCGTTG
8301 8401 8501 8601 8701	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC GTGGTCAGTG GGAGAACGATC TCTTTCGTA CTGTTGCTAG GGAGGACCGA ATACCAAACG CCTCCTGGCT TATGGTTGCA ACGACGACG AATTAATAGA AGGACGACT TTAATTATCT CTGGATGGAG	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCCGCCCTTC	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC CGGCTGGCTG	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGGCCA TGACGCGGGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGGCCGTTG GAGCCGGTGA
8301 8401 8501 8601 8701	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG AGAAAAGCAT GACAACGATC CTGTTGCTAG GGAGGACCGA ATACCAAACG CCTCCTGGCT TATGGTTGC ACGAGCGTGC TTAATTATCT TGGATGGAG GCGTGGGGTCT GACCTACCTC	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG CGCCTATTC	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC AACGTCCTGG	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC TGAAGACGCG	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCGGCCCTTC AGCCGGGAAG	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC CGGCTGGCTG GCCGACCGAC	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT CAAATAACGA	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGGCCA TGACGCGGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG CTATTTAGAC	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGGCCGTTG GAGCCGGTGA CTCGGCCACT
8301 8401 8501 8601 8701	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG AGAAAAGCAT GACAACGATC CTGTTGCTAG GGAGGACCGA ATACCAAACG CCTCCTGGCT TATGGTTGC ACGAGCGTGA AATTAATAGA TGCTCGCACT TTAATTATCT CTGGATGGAG GCCTGGGTCT GACCTACCTC CGCACCAGA	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG CGCCTATTTC TTGCAGCACT	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC AACGTCCTGG GGGGCCCAGAT	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC TGAAGACGCG GGTAAGCCCT	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCGGCCCTTC AGCCGGGAAG CCCGTATCGT	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC CGGCTGGCTG GCCGACCGAC AGTTATCTAC	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT CAAATAACGA ACGACGGGGA	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGGCCA TGACGCCGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG CTATTTAGAC GTCAGGCAAC	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGGCCGTTG GAGCCGGTGA CTCGGCCACT TATGGATGAA
8301 8401 8501 8601 8701 8801	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG AGAAAAGCAT GACAACGATC CTGTTGCTAG GGAGGACCGA ATACCAAACG CCTCCTGGCT TATGGTTGCC ACGAGCGTGA AATTAATAGA TGCTCGCACT TTAATTATCT TAAGTAGC GCGTGGGTCT GACCTACCTC CGCACCCAGA CGCGCTACCA	ACATEGRACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG CGCCTATTTC TTGCAGCACT	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC AACGTCCTGG GGGGCCAGAT	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC TGAAGACGCG GGTAAGCCCT	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCGGCCCTTC AGCCGGGAAG CCCGTATCGT	AAAAGCGGGG CGCCGCATAC GCGCCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC CGGCTGGCTG GCCGACCGAC AGTTATCTAC	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT CAAATAACGA ACGACGGGGA	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGGCCA TGACGCCGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG CTATTTAGAC GTCAGGCAAC	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGGCCGTTG GAGCCGGTGA CTCGGCCACT TATGGATGAA
8301 8401 8501 8601 8701 8801	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG AGAAAAGCAT GCAAACGAT CTTTTCGTA CTGTTGCTAG GGAGGACCAA ATACCAAACG CCTCCTGGCT TAAGGTTTGC ACGAGCGTGA AATTAATAGA GGCTGGGCT TTAATTATCT GACCCACCA GGCGGTATCA CGCACCAG GCGTCATCT	ACATCGAACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG CGCCTATTTC TTGCAGCACT AACGTCGTGA	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC AACGTCCTGG GGGGCCAGAT CCCCGGTCTA	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC TGAAGACGCC GGTAAGCCCT CCATTCGGGA	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCGGCCCTTC AGCCGGGAAG CCCGTATCGT GGGCATAGCA	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC CGGCTGGCTG GCCGACCGAC AGTTATCTAC TCAATAGATG	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT CAAATAACGA ACGACGGGGA TGCTGCCCCT	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGGCCA TGACGCCGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG CTATTTAGAC GTCAGGCAAC CAGTCCGTTG	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGCCGGTGA CTCGGCCACT TATGGATGAA ATACCTACTT
8301 8401 8501 8601 8701 8801 8901	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG AGAAAAGCAT CTTTTCGTA CTGTTGCTAG GGAGGACCGA ATTACAAACG CCTCCTGGCT TATGGTTTGC ACGAGCGTGA AATTAATAGA GCGTGGGCT TTAATTATCT CTGGATGGAG GCGTGGGTATCA CGCCCCACAG GCGCCATAGT GCTTTATCTG AGAATAGAC	ACATCGAACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG CGCCTATTTC TTGCAGCACT AACGTCGTGA GATAGGTGCC	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC AACGTCCTGG GGGGCCAGAT CCCCGGTCTA TCACTGATTA	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC TGAAGACGCCG GGTAAGCCCT CCATTCGGGA AGCATTGGTA	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCGGCCCTTC AGCCGGGAAG CCCGTATCGT GGGCATAGCA ACTGTCAGAC	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC CGGCTGGCTG GCCGACCGAC AGTTATCTAC TCAATAGATG CAAGTTTACT	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT CAAATAACGA ACGACGGGGA TGCTGCCCCT CATATATACT	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGCCA TGACGCCGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG CTATTTAGAC GTCAGGCAAC CAGTCCGTTG TTAGATTGAT	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGCCGGTGA CTCGGCCACT TATGGATGAA ATACCTACTT TTAAAACTTC
8301 8401 8501 8601 8701 8801 8901	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG AGAAAAGCAT CTTTTCGTA CTGTTGCTAG GGAGGACCGA ATTACAACGG CCTCCTGGCT TATGGTTTGC ACGAGCGGGA AATTAATAGA TGCTCGCACT TTAATTATCT CACGAGCGCAGA CGCCGCATAGT GACCTCCTG CGCACCAAGA CGCCGCATAGT GCTTTATCTG AGATCGCTGA ATTTTAATT TCTAGCGACT	ACATCGAACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG CGCCTATTTC TTGCAGCACT AACGTCGTGA GATAGGTGCC CTATCCACGG	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC AACGTCCTGG GGGGCCAGAT CCCCGGTCTA TCACTGATTA AGTGACTAAT	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC GGTAAGCCCT CCATTCGGGA AGCATTGGTA	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCGGCCCTTC AGCCGGGAAG CCCGTATCGT GGGCATAGCA ACTGTCAGAC TGACAGTCTG	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC CGGCTGGCTG GCCGACCGAC AGTTATCTAC TCAATAGATG CTCCAATGA	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT CAAATAACGA ACGACGGGGA TGCTGCCCCT CATATATACT GTATATATGA	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGCCA TGACGCCGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG CTATTTAGAC GTCAGGCAAC CAGTCCGTTG TTAGATTGAT AATCTAACTA	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGCCGGTGA CTCGGCCACT TATGGATGAA ATACCTACTT TTAAAACTTC AATTTTGAAG
8301 8401 8501 8601 8701 8801 8901	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG AGAAAGCATC TCTTTTCGTA CTGTTGCTAG GGAGGACCGG ATACCAAACG CCTCCTGGCT TATGGTTTGC ACGACCGGCG AATTAATAGA TGCTCGCACT GACCTCCCCG GCGCGGGACT GACCTACCTC GCACCCAGA CGCGCATAGA GCGCCATAGT GCTTTATCTG AGATCGCCGACT TAAAAATTAA	ACATCGAACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG CGCCTATTTC TTGCAGCACT AACGTCGTGA GATAGGTGCC CTATCCACGG	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC AACGTCCTGG GGGGCCAGAT CCCCGGTCTA TCACTGATTA AGTGACTAAT	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC GGTAAGCCCT CCATTCGGGA AGCATTGGTA TCGTAACCAT	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCGGCCCTTC AGCCGGGAAG CCCGTATCGT GGGCATAGCA ACTGTCAGAC TGACAGTCTG	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC CGGCTGGCTG GCCGACCGAC AGTTATCTAC TCAATAGATG CAAGTTTACT GTTCAAATGA	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT CAAATAACGA ACGACGGGGA TGCTGCCCCT GTATATATGA GTTTTCCTTC	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGCCA TGACGCCGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG GTCAGGCAAC CAGTCCGTTG TTAGATTGAT AATCTAACTA	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGCCGGTGA CTCGGCCACT TATGGATGAA ATACCTACTT TTAAAACTTC AATTTTGAAG
8301 8401 8501 8601 8701 8801 8901	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG GACAACGATC TCTTTTCGTA CTGTTGCTAG GGAGGACCG CTCCTGGCT TATGGTTTGC ACGACCGTGA ATACCAAACG CCTCCTGGCT TTAATTATAGA TGCTCGCACT TTAATTATCT CTGGATGGAG GCGTGGGGTCT GACCTACCTC GCACCCAGA CGCCCAGA CGCCCAGA CGCCCAGA CGCCCACT CGCACCTAGT GACTTATCTG AGAATAGACT AAAAATTAA TAAAAGATC	ACATCGAACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG CGCCTATTTC TTGCAGCACT AACGTCGTGA GATAGGTGCC CTATCCACGG TAGGTGAAGA	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC AACGTCCTGG GGGGCCAGAT CCCCGGTCTA TCACTGATTA AGTGACTAAT TCCTTTTTGA	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC GGTAAGCCCT CCATTCGGGA AGCATTGGTA TCGTAACCAT TAATCTCATG	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCGGCCCTTC AGCCGGGAAG CCCGTATCGT GGGCATAGCA ACTGTCAGAC TGACAGTCTG ACCAAAATCC	AAAAGCGGGG CGCCGCATAC GCGGCGTATG CCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC CGGCTGGCTG GCCGACCGAC AGTTATCTAC TCAATAGATG CAAGTTTACT GTTCAAATGA CTTAACGTGA	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT CAAATAACGA ACGACGGGGA TGCTGCCCCT CATATATACT GTATATATGA GTTTTCGTTC	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGCCA TGACGCCGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG GTCAGGCAAC CAGTCCGTTG TTAGATTGAT AATCTAACTA CACTGAGCGT	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGCCGGTGA CTCGGCCACT TATGGATGAA ATACCTACTT TTAAAACTTC AATTTTGAAG CAGACCCCGT
8301 8401 8501 8601 8801 8901 9001	AAAGTTCTGC GCTCACCCAA TTTCAAGACG TATGTGGCGC CACCAGTCAC ATACACCGCG GTGGTCAGTG GACAACGATC TCTTTTCGTA CTGTTGCTAG GGAGGACCGA ATACCAAACG CCTCCTGGCT TATGGTTTGC ACGACCGACA AATTAATAGA TGCTCGCACT GACTACCTC GCACCACAGA CGCGCATAGT GCGTCACTG GCGCCATAGT GCTTTATCTG GCGCCATAGT GCTTTATCTG AAATCGCTG AATTATAATTAA TCTAGCGACT TAAAAATTAA TAAAAGATC AGAAAAGATC ATCTTCCTAC	ACATCGAACT TGTAGCTTGA GGTATTATCC CCATAATAGG CTTACGGATG GAATGCCTAC AGGAGCTAAC TCCTCGATTG CACCACGATG GTGGTGCTAC GCGGATAAAG CGCCTATTTC TTGCAGCACT AACGTCGTGA GATAGGTGCC CTATCCACGG TAGGTGAAGA ATCCACTTCT	CCTAGAGTTG CGTATTGACG GCATAACTGC GCATGACAGT CGTACTGTCA CGCTTTTTTG GCGAAAAAAAC CCTGTAGCAA GGACATCGTT TTGCAGGACC AACGTCCTGG GGGGCCAGAT CCCCGGTCTA TCACTGATTA AGTGACTAAT TCCTTTTTGA AGGAAAAACT	AGCGGTAAGA TCGCCATTCT CCGGGCAAGA GGCCCGTTCT AAGAGAATTA TTCTCTTAAT CACAACATGG GTGTTGTACC TGGCAACAAC ACCGTTGTTG ACTTCTGCGC GGTAAGCCCT CCATTCGGGA AGCATTGGTA TCGTAACCAT TAATCTCATG ATTAGAGTAC	AGGAACTCTC GCAACTCGGT CGTTGAGCCA TGCAGTGCTG ACGTCACGAC GGGATCATGT CCCTAGTACA GTTGCGCAAA CAACGCGTTT TCGGCCCTTC AGCCGGGAAG CCCGTATCGT GGGCATAGCA ACTGTCAGAC TGACAGTCTG ACCAAAATCC TGGTTTTAGG	AAAAGCGGGG GGCGGCGTATG GCGGCGTATG GCATAACCAT GGTATTGGTA AACTCGCCTT TTGAGCGGAA CTATTAACTG GATAATTGAC GGCCGACCGAC AGTTATCTAC TCAATAGATG CAAGTTTACT GTTCAAATGA CTTAACGTGA GAATTGCACT	CTTCTTGCAA ACTATTCTCA TGATAAGAGT GAGTGATAAC CTCACTATTG GATCGTTGGG CTAGCAACCC GCGAACTACT CGCTTGATGA GTTTATTGCT CAAATAACGA ACGACGGGGA TGCTGCCCCT GTATATATGA GTTTTCGTTC CAAAAGCAAG	AAGGTTACTA GAATGACTTG CTTACTGAAC ACTGCGGCCA TGACGCCGGT AACCGGAGCT TTGGCCTCGA TACTCTAGCT ATGAGATCGA GATAAATCTG GTCAGGCAAC CAGTCCGTTG TTAGATTGAT AATCTAACTA CACTGAGCGT GTGACTCGCA	CTCGTGAAAA GTTGAGTACT CAACTCATGA ACTTACTTCT TGAATGAAGA GAATGAAGCC CTTACTTCGG TCCCGGCAAC AGGCCGGTGA CTCGGCCACT TATGGATGAA ATACCTACTT TTAAAACTTC AATTTTGAAG CAGACCCCGT GTCTGGGGCA

9101	AAAGGATCTT GATCAAGAGC	CTTGAGATCC	TTTTTTTCTG	CGCGTAATCT	GCTGCTTGCA	ААСАААААА	CCACCGCTAC	CAGCGGTGGT	TTGTTTGCCG
	TTTCCTAGAA	GAACTCTAGG	AAAAAAAGAC	GCGCATTAGA	CGACGAACGT	TTGTTTTTTT	GGTGGCGATG	GTCGCCACCA	AACAAACGGC
9201	TACCAACTCT	TTTTCCGAAG	GTAACTGGCT	TCAGCAGAGC	GCAGATACCA	AATACTGTCC	TTCTAGTGTA	GCCGTAGTTA	GGCCACCACT
	ATGGTTGAGA	AAAAGGCTTC	CATTGACCGA	AGTCGTCTCG	CGTCTATGGT	TTATGACAGG	AAGATCACAT	CGGCATCAAT	CCGGTGGTGA
9301	TGTAGCACCG	CCTACATACC	TCGCTCTGCT	AATCCTGTTA	CCAGTGGCTG	CTGCCAGTGG	CGATAAGTCG	TGTCTTACCG	GGTTGGACTC
	ACATCGTGGC	GGATGTATGG	AGCGAGACGA	TTAGGACAAT	GGTCACCGAC	GACGGTCACC	GCTATTCAGC	ACAGAATGGC	CCAACCTGAG
9401	TTACCGGATA	AGGCGCAGCG	GTCGGGCTGA	ACGGGGGGTT	CGTGCACACA	GCCCAGCTTG	GAGCGAACGA	CCTACACCGA	ACTGAGATAC
	AATGGCCTAT	TCCGCGTCGC	CAGCCCGACT	TGCCCCCCAA	GCACGTGTGT	CGGGTCGAAC	CTCGCTTGCT	GGATGTGGCT	TGACTCTATG
9501	AGCATTGAGA	AAGCGCCACG	CTTCCCGAAG	GGAGAAAGGC	GGACAGGTAT	CCGGTAAGCG	GCAGGGTCGG	AACAGGAGAG	CGCACGAGGG
	TCGTAACTCT	TTCGCGGTGC	GAAGGGCTTC	CCTCTTTCCG	CCTGTCCATA	GGCCATTCGC	CGTCCCAGCC	TTGTCCTCTC	GCGTGCTCCC
9601	GGGAAACGCC	TGGTATCTTT	ATAGTCCTGT	CGGGTTTCGC	CACCTCTGAC	TTGAGCGTCG	ATTTTTGTGA	TGCTCGTCAG	GGGGGCGGAG
	CCCTTTGCGG	ACCATAGAAA	TATCAGGACA	GCCCAAAGCG	GTGGAGACTG	AACTCGCAGC	TAAAAACACT	ACGAGCAGTC	CCCCCGCCTC
9701	AACGCCAGCA	ACGCGGCCTT	TTTACGGTTC	CTGGCCTTTT	GCTGGCCTTT	TGCTCACATG	TTCTTTCCTG	CGTTATCCCC	TGATTCTGTG
	TTGCGGTCGT	TGCGCCGGAA	AAATGCCAAG	GACCGGAAAA	CGACCGGAAA	ACGAGTGTAC	AAGAAAGGAC	GCAATAGGGG	ACTAAGACAC
9801	TTACCGCCATT	TGAGTGAGCT	GATACCGCTC	GCCGCAGCCG	AACGACCGAG	CGCAGCGAGT	CAGTGAGCGA	GGAAGCGGAA	GAGCGCCCAA
	AATGGCGGAA	ACTCACTCGA	CTATGGCGAG	CGGCGTCGGC	TTGCTGGCTC	GCGTCGCTCA	GTCACTCGCT	CCTTCGCCTT	CTCGCGGGTT
9901	GCCTCTCCCC	GCGCGTTGGC	CGATTCATTA	ATGCAGCTGG	CACGACAGGT	TTCCCGACTG	GAAAGCGGGC	AGTGAGCGCA	ACGCAATTAA
	CGGAGAGGGG	CGCGCAACCG	GCTAAGTAAT	TACGTCGACC	GTGCTGTCCA	AAGGGCTGAC	CTTTCGCCCG	TCACTCGCGT	TGCGTTAATT
10001	CTCACTCATT	AGGCACCCCA	GGCTTTACAC	TTTATGCTTC	CGCGGCTCGT	ATGTTGTGTG	GAATTGTGAG	CGGATAACAA	TTTCACACAG
	GAGTGAGTAA CTTTGTCGA	TCCGTGGGGT	CCGAAATGTG	AAATACGAAG	GCGCCGAGCA	TACAACACAC	CTTAACACTC	GCCTATTGTT	AAAGTGTGTC

Supplementary Figure 4: Vector map and full DNA/amino acid sequence of the ZFNs used for *golden* gene disruption.

Vector map (sequence shown on next page):



ZFN expression vector sequence

- 1) The ZFP DNA binding helices are shown in lower case and underlined.
- 2) The 2A peptide sequence is double-underlined.
- 3) The Eag I restriction site and the T7 promoter used to linearize the plasmid and drive mRNA transcription, respectively, are indicated.

golden pair 1

CTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAG T7 promoter <u>G</u>GAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTGATCCACTAGTCCAGTGTGGTGGAA М Κ D D G D Y Κ D D D Κ D Y Η Η Ι Y TTCGCCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG КМАРК Κ D D D D Κ Κ R VG І Н G V Ρ GATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAGGTGGGCATTCACGG<u>GGTACC</u>C Kpn I ZFP-L A A R PFO С R Ι С М R Ν Μ Α Ε F S t S GCCGCTATGGCTGAGAGGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTACCTCC r R Т Η Т G Ε Κ Ρ F Α С Ι q s 1 S Н Ι D GGCTCCCTGTCCCGCCACATCCGCACCCCACACAGGCGAGAAGCCTTTTGCCTGTGACATT Κ A<u>rsdnlre</u>HT Κ С G R F Ι Η Т G TGTGGGAGGAAGTTTGCCCGCTCCGACAACCTGCGCGAGCATACCAAGATACACGGGA Ρ F Q С R I С М R Ν F S r s Ο Κ d а TCTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGTAGTGACGCCCTG Η Ι R Т Η Т G Ε Κ Ρ F А С D Ι С G R s e AGCGAACACATCCGCACCCACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGG <u>q n a t r t k</u> Κ Κ F Α Η Т Ι Η T. R AAATTTGCCCAGAACGCCACCCGCACAAAGCATACCAAGATACACCTGCGG Fok-L G S 0 L V Κ S Ε L Ε Ε Κ Κ S Ε L R Η Κ L GGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTG BamH I V Ρ Ε Κ Y Η Y Ι Ε \mathbf{L} Ι Ε Ι Α R Ν S т Q D AAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGAC R Ι L Ε М Κ V М Ε F F М Κ V Y G Υ R G Κ CGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAG Η L G G S R Κ Ρ D G Α Ι Υ Т V G S Ρ Ι D CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGAT

Y	G	V	I	V	D	Т	Κ	А	Y	S	G	G	Y	Ν	L	Ρ	I	G	Q
TAC	GGC	GTG	ATC	GTG	GAC	ACA	AAG	GCC	TAC	AGC	GGC	GGC	TAC	AAT	CTG	ССТ	ATC	GGC	CAG
A	D	Е	М	Е	R	Y	V	Е	Е	Ν	0	Т	R	Ν	K	Н	L	Ν	Ρ
GCC	GAC	GAG	ATG	GAG	AGA	TAC	GTG	GAG	GAG	AAC	CÃG	ACC	CGG	AAT	'AAG	CAC	СТС	AAC	CCC
NT	Ţ	T 47	TAT	V	77	v	П	C	C	τ7	m	Ţ	Ţ	V	Ţ	т	Ţ	77	C
NAAC	e GAG	w TGG	w TGG	r. AAG	v GTG	ı TAC	Р ССТ	'AGC	S AGC	v GTG	ACC	e GAG	F TTC	r. AAG	г TTC	ь CTG	г TTC	v GTG	S AGC
G	H	F	K	G	N	Y	K	A	Q	L	Т	R	L	N	H	I	Т	N	С
GGC	CAC	.1.1.C	AAG	GGC	:AAC	TAC	AAG	GCC	CAG	CTG	ACC	AGG	CTG	AAC	CAC	ATC	ACC	AAC	TGC
Ν	G	А	V	L	S	V	Е	Е	L	L	I	G	G	Ε	М	I	Κ	А	G
AAT	GGC	GCC	GTG	CTG	AGC	GTG	GAG	GAG	CTG	CTG	ATC	GGC	GGC	GAG	ATG	ATC	AAA	GCC	GGC
т	т.	т	т.	E	E	V	R	R	ĸ	ਸ	N	N	G	E	т	N	ਸ	R	S
ACC	CTG	ACA	CTG	GAG	GAG	GTG	CGG	CGC	AAG	TTC	AAC	AAC	GGC	GAG	;ATC	AAC	TTC	AGA	TCT
															B	gl II			
G	G	G	Е	G	R	G	S	L	L	Т	С	G	D	V	Е	Е	Ν	Ρ	G
GGC	GGC	GGA	GAG	GGC	CAGA	GGA	AGT	CTI	CTA	ACA	TGC	GGT	'GAC	GTG	GAG	GAG	AAT	CCC	GGC
П	п	ъл	П	37	17	П	тт	П	C	П	37	v	П	тт	П	т	П	37	V
CCT	AGG	™ ATG	GAC	ı TAC	AAA	.GAC	н САТ	'GAC	GGT	D 'GAT	ı TAT	ہ AAA'	d. GAT	н 'CAT	ש GAC	L ATC	d GAT	ı TAC	r. AAG
			-				-									-		-	
D	D	D	D	K	М	A	P	K	K	K	R	K	V	G	I	H	G	V	P
GAT	GAC	GAT	GAC	AAG	FATG	GCC	CCC	CAAC	iaag	AAG	AGG	AAG	GTG	GGC	Kpn: Kpn	I I	66 <u>6</u>	GTA	<u>CC</u> C
7FP.	R																		
A	A	М	A	Е	R	Ρ	F	Q	С	R	I	С	М	R	Ν	F	S	d	r
GCC	GCT	ATG	GCT	'GAG	AGG	CCC	TTC	CAG	TGT	'CGA	ATC	TGC	ATG	CGI	'AAC	TTC	AGT	GAC	CGC
g	Ь	1	g	r	н	т	R	т	н	т	G	E	ĸ	P	ਸ	Δ	C	Л	т
TCC	GAC	CTG	TCC	CGC	CAC	ATC	CGC	ACC	CAC	ACA	.GGC	GAG	AAG	CCI	TTT	GCC	TGT	GAC	ATT
-		_		_	_			_	_	_				_		_		_	
С	G	R	K NAC	F	A ICCC	r r	S TCC	d CDC	d CDC	<u>1</u> ירידים			Н Слт	T 'NCC	K NAC	I גידע	H	Т	G
101	000	AUU	AAU				ICC	JAU.	JAC.		ACC		CAI	ACC			CAC	ACU	000
G	G	G	S	Q	Κ	Ρ	F	Q	С	R	I	С	М	R	Ν	F	S	r	S
GGA	GGC	GGA	TCT	'CAG	AAG	CCC	TTC	CAG	TGT	'CGA	ATC	TGC	ATG	CGI	'AAC	TTC	AGT	CGC	TCC
d	d	1	t	r	Н	I	R	Т	Н	Т	G	Е	K	Ρ	F	A	С	D	I
GAC	GAC	CTG	ACC	CGC	CAC	ATC	CGC	CACC	CAC	ACA	GGC	GAG	AAG	CCI	TTT	GCC	TGT	GAC	ATT
a	a	P	77	-	7				-1	-	-			-	T.7	-		-	Ð
C TGT	GGG	r AGG	k AAG	F TTT	A 'GCC	<u>q</u> CAG	s TCC	g :GGC	a 'GAC	L CTG	t ACC	<u>r</u> CGC	н CAT	'ACC	к ААС	l ATA	н CAC	ь СТG	к CGG
						0110				010			0111				0110	010	
Fok-	R	\circ	т	77	V	C	ъ	т	Ū	ч	v	v	C	ч	т	P	U	v	т.
GGA	TCC	Q CAG	L CTG	v GTG	م AAG	AGC	ь GAG	ь SCTG	₽ GAG	ь GAG	л AAG	л AAG	TCC	L GAG	ь СТG	к CGG	л САС	л AAG	CTG
Baml	41														•				-
K	Y	V	Ρ	н	Е	Y	I	Е	L	I	Е	I	A	R	N	S	т	0	D

AAG	TAC	GTG	CCC	CAC	GAG	TAC	ATC	GAG	CTG	ATC	GAG	ATC	GCC	AGG	AAC	AGC	ACC	CAG	GAC
R	I	L	Е	М	К	V	М	Е	F	F	М	K	V	Y	G	Y	R	G	K
CGC	ATC	CTG	GAG	ATG	AAG	GTG	ATG	GAG	TTC	TTC	ATG	AAG	GTG	TAC	GGC	TAC	AGG	GGA.	AAG
Н	L	G	G	S	R	K	Ρ	D	G	А	I	Y	т	V	G	S	Ρ	I	D
CAC	CTG	GGC	GGA	AGC	AGA	AAG	ССТ	GAC	GGC	GCC	ATC	TAT	ACA	GTG	GGC	AGC	CCC	ATC	GAT
Y	G	V	I	v	D	т	K	A	Y	S	G	G	Y	Ν	L	Р	I	G	0
TAC	GGC	GTG	ATC	GTG	GAC	ACA	AAG	GCC	TAC	AGC	GGC	GGC	TAC	AAT	'CTG	ССТ	ATC	GGC	CAG
А	D	Е	М	0	R	Y	v	K	Е	Ν	0	т	R	Ν	K	Н	I	Ν	Ρ
GCC	GAC	GAG	ATG	CÃG	AGA	TAC	GTG	AAG	GAG	AAC	CÃG	ACC	CGG	AAT	'AAG	CAC	ATC	AAC	CCC
N	E	W	W	ĸ	V	Y	P	S	S	V	т	E	F	ĸ	F	т.	ਸ	V	S
AAC	GAG	TGG	TGG	AAG	GTG	TAC	ССТ	AGC	AGC	GTG	ACC	GAG	TTC	AAG	TTC	CTG	TTC	GTG.	AGC
G	н	ਸ	ĸ	G	N	v	ĸ	Δ	0	т.	Ŧ	R	т.	N	н	ĸ	т	N	C
GGC	CAC	TTC	AAG	GGC	AAC	TAC	AAG	GCC	CÃG	CTG	ACC	AGG	CTG	AAC	CAC	AAA	ACC	AAC	TGC
N	G	Δ	V	т.	g	V	F	E	т.	т.	т	G	G	F	м	т	ĸ	Δ	G
AAT	GGC	GCC	GTG	CTG	AGC	GTG	GAG	GAG	CTG	CTG	ATC	GGC	GGC	GAG	ATG	ATC	AAA	GCC	GGC
т	т.	т	т.	F	E	V	R	R	ĸ	F	N	N	G	F	т	N	ਸ	*	
ACC	CTG	ACA	CTG	GAG	GAG	GTG	CGG	CGC	AAG	TTC	AAC	AAC	GGC	GAG	ATC	AAC	TTC	TGA	TAA
<u>CTC</u>	GAG	TCT	AGA	GGG	CCC	GTT	ТАА	ACC	CGC	TGA	TCA	GCC	TCG	А <u>СТ</u>	'GTG	ССТ	ТСТ	AGT	TGC
Xho	I												BGH	poly(A) signa	al			
<u>CAG</u>	CCA	TCT	GTT	GTT	TGC	CCC	тсс	CCC	GTG	ССТ	TCC	TTG	ACC	CTG	GAA	GGT	GCC	АСТ	<u>CCC</u>
аст	GTC	CTT	тсс	ͲΔΔ	ТАА	ልልጥ	GAG	GAA	ልጥጥ	GCA	тсg	САТ	тGт	СТС	AGT	AGG	тдт	ידבי	тст
<u>110 1</u>	010		100	1111		<u> </u>	0110	01111	<u></u>	0011	100	0111	101	010	1101	1100	101	0111	<u>1 U I</u>
<u>ATT</u>	CTG	GGG	GGT	GGG	GTG	GGG	CAG	GAC	AGC	AAG	GGG	GAG	GAT	TGG	GAA	GAC	AAT	AGC.	<u>AGG</u>
<u>CAT</u>	GCT	GGG	GAT	GCG	GTG	GGC	тст	ATG	<u>G</u> CT	тст	ACT	GGG	CGG	TTT	'TAT	GGA	CAG	CAA	GCG
AAC	CGG	AAT	TGC	CAG	CTG	GGG	CGC	ССТ	CTG	GTA	AGG	TTG	GGA	AGC	CCT	GCA	AAG	TAA.	ACT
GGA	TGG	CTT	тст	CGC	CGC	CAA	GGA	тст	GAT	GGC	GCA	GGG	GAT	CAA	GCT	CTG	ATC	AAG.	AGA
<u> </u>		<u> </u>		COT	יייייי	007	т с л	ጥጥረ	770	770	አምሮ			7 CC		Cumm	CTTC		
CAG	GAI	UAU	GAI	CGI	ттС	GCA	тGА	тīС	AAC	AAG	AIG	GAI	IGC	ACG	Ea	agl			

golden pair 14

CTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAAT<u>TAATACGACTCACTATAG</u> *T7 promoter*

<u>G</u>GAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTGATCCACTAGTCCAGTGTGGTGGAA

M D Y K D H D G D Y K D H D I D Y K TTCGCCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG

D D D K M A P K K K R K V G I H G V P GATGACGATGACAAGATGGCCCCCAAGAAGAAGAAGAGGAAGGTGGGCATTCACGG<u>GGTACC</u>C *Kpn I*

ZFP-L

<u>g n l a r H</u> I R T H T G E K P F A C D I GGCAACCTGGCCCGCCACATCCGCACCCCACACAGGCGAGAAGCCTTTTGCCTGTGACATT

C G R K F A <u>t s a n l s r</u> H T K I H T G TGTGGGAGGAAGTTTGCCACCTCCGCCAACCTGTCCCGCCATACCAAGATACACACGGGA

S Q K P F Q C R I C M R N F S $\underline{r \ s \ d \ t \ l}$ TCTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGTAGTGACACCCTG

<u>se</u> H I R T H T G E K P F A C D I C G R AGCGAACACATCCGCACCACACGGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGG

K F A <u>r s q t r k t</u> H T K I H L R AAATTTGCCCGCAGCCAGACCCGCAAAACCCATACCAAGATACACCTGCGG

Fok-L

G S Q L V K S E L E E K K S E L R H K L <u>GGATCC</u>CAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTG *BamH I*

K Y V P H E Y I E L I E I A R N S T Q D AAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGAC R I L E M K V M E F F M K V Y G Y R G K CGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGGAAAG

H L G G S R K P D G A I Y T V G S P I D CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGAT

Y G V I V D T K A Y S G G Y N L P I G Q TACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAG

A D E M E R Y V E E N Q T R N K H L N P GCCGACGAGATGGAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACCTCAACCCC

N E W W K V Y P S S V T E F K F L F V S AACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGC

G H F K G N Y K A Q L T R L N H I T N C GGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGC N G A V L S V E E L L I G G E M I K A G AATGGCGCCGTGCTGAGCGTGGAGGAGGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGC T L T L E E V R R K F N N G E I N F R S ACCCTGACACTGGAGGAGGTGCGGCGCGAGGTTCAACAACGGCGAGATCAACTTCAGATCT Bal II G G G E G R G S L L T C G D V E E N P G GGCGGCGGAGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGC <u>PR</u>MDYKDHDGDYKDHDIDYK CCTAGGATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG D D D K M A P K K K R K V G I H G V P GATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAGGTGGGCATTCACGG<u>GGTACC</u>C Kpn I ZFP-R A A M A E R P F O C R I C M R N F S d r GCCGCTATGGCTGAGAGGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTGACCGC ShlsrHIRTHTGEKPFACDI TCCCACCTGTCCCGCCACATCCGCACCACACAGGCGAGAAGCCTTTTGCCTGTGACATT CGRKFA<u>rsdalar</u>HTKIHTG TGTGGGAGGAAGTTTGCCCGCTCCGACGCCCTGGCCCGCCATACCAAGATACACGGGGA S Q K P F Q C R I C M R N F S d r s n l TCTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTGACCGCTCCAACCTG S T H I R T H T G E K P F A C D I C G R TCCCGCCACATCCGCACCACACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGG K F A t s q s l t r H T K I H L R AAGTTTGCCACCTCCGGCTCCCTGACCCGCCATACCAAGATACACCTGCGG Fok-R G S Q L V K S E L E E K K S E L R H K L GGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTG BamH I K Y V P H E Y I E L I E I A R N S T O D AAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGAC R I L E M K V M E F F M K V Y G Y R G K CGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAG H L G G S R K P D G A I Y T V G S P Т D CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGAT Y G V I V D T K A Y S G G Y N L P I G O

24

TAC	GGC	GTG.	ATC	GTG	GAC	ACA	AAG	GCC.	ГАC/	AGC	GGC	GGCI	rac <i>i</i>	AAT(CTG	CCTA	ATC(GGC	CAG
A GCC	D GAC	E GAG.	M ATG	Q CAG	R AGA'	Y FAC(V GTG2	K AAG(E GAG2	N AAC(Q CAGA	T ACCO	R CGG <i>I</i>	N AAT <i>I</i>	K AAG(H CACA	I ATCI	N AAC(P CCC
N AAC	E GAG	W TGG	W TGG	K AAG	V GTG	Y FAC(P CCT2	S AGC <i>I</i>	S AGC(V GTG2	T ACC(E GAGI	F [TC]	K AAG:	F FTC(L CTG:	F FTC(V GTG <i>i</i>	S AGC
G GGC	H CAC	F TTC.	K AAG	G GGC	N AAC'	Y FACA	K AAG(A GCC(Q CAG(L CTG2	T ACC2	R AGG(L CTG <i>i</i>	N AAC(H CAC2	K AAAA	T ACC <i>I</i>	N AACI	C [GC
N AAT	G GGC	A GCC	V GTG	L CTG	S AGC(V GTG(E GAG(E GAG(L CTG(L CTG2	I ATC(G GGC(G GGC(E GAG <i>i</i>	M ATG2	I ATCZ	K AAA(A GCCC	G GGC
T ACC	L CTG.	T ACA	L CTG	E GAG	E GAG(V GTG(R CGG	R CGC <i>I</i>	K AAG:	F FTC2	N AACI	N AACO	G GGC(E GAG <i>I</i>	I ATCZ	N AAC:	F FTC:	* FGA1	ГАА
<u>CTC</u> Xho	<u>GAG</u> I	TCT.	AGA	GGG	CCC	GTTI	ΓAA)	ACC	CGC	[GA]	TCA	GCCI	ГСG <i>I</i> ВGН	A <u>CT(</u> poly(A	<u>GTG(</u>) sign	<u>CCT:</u> al	<u>гст</u> 2	AGTI	<u>rgc</u>
<u>CAG</u>	CCA	ТСТ	GTT	GTT	TGC		<u>rcc</u>	<u> 2220</u>	GTG	CCT	<u>rcc</u> :	rtg <i>i</i>	ACCO	CTGO	GAA	GGT(<u>GCC</u>	ACTO	<u> 202</u>
ACT	GTC	CTT	TCC	TAA	TAA	AAT	GAG	GAA	ATT	GCA'	TCG	CATI	[GT(CTGA	AGTZ	AGG	<u>rgt(</u>	CATI	<u> </u>
<u>ATT</u>	CTG	GGG	GGT	GGG	GTG	GGG	CAG	GAC <i>I</i>	AGCZ	AAG	GGG	GAG	GAT	rggo	GAA	GAC	AAT <i>I</i>	AGCI	<u>AGG</u>
CAT	GCT	GGG	GAT	GCG	GTG	GGC	ГСТЛ	ATG	<u>G</u> CT	ГСТІ	ACT	GGG	CGG	CTT:	[AT(GGA	CAG	CAAC	GCG
AAC	CGG.	AAT	TGC	CAG	CTG	GGG	CGC	ССТО	CTG	GTA	AGG	ГТGC	GGA	AGC	CCT	GCA	AAG	ΓΑΑ	ΑСТ
GGA	TGG	СТТ	ТСТ	CGC	CGC	CAA	GGA'	ГСТС	GAT	GGC	GCA	GGG	GAT	CAAC	GCT	CTG	ATCI	AAGA	AGA
CAG	GAT	GAG	GAT	CGT	TTC	GCA	ΓGΑ'	TTG <i>I</i>	AAC	AAG	ATG	GATI	ſGĊł	4CG(CAG Eag	GTT(gl	CTC	CGGC	<u>CCG</u>

golden pair 15

CTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAAT<u>TAATACGACTCACTATAG</u> *T7 promoter*

<u>G</u>GAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTGATCCACTAGTCCAGTGTGGTGGAA

M D Y K D H D G D Y K D H D I D Y K TTCGCCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG

ZFP-L

<u>g n l a r</u> H I R T H T G E K P F A C D I GGCAACCTGGCCCGCCACATCCGCACCCACACAGGCGAGAAGCCTTTTGCCTGTGACATT

C G R K F A <u>t s g n l t r</u> H T K I H T G TGTGGGAGGAAGTTTGCCACCTCCGGCAACCTGACCCGCCATACCAAGATACACGGGA

S Q K P F Q C R I C M R N F S $\underline{r \ s \ d \ t \ l}$ TCTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGTAGTGACACCCTG

<u>se</u> H I R T H T G E K P F A C D I C G R AGCGAACACATCCGCACCACACGGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGG

K F A <u>r s q t r k t</u> H T K I H L R AAATTTGCCCGCAGCCAGACCCGCAAAACCCATACCAAGATACACCTGCGG

Fok-L

G S Q L V K S E L E E K K S E L R H K L <u>GGATCC</u>CAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTG BamH I

K Y V P H E Y I E L I E I A R N S T Q D AAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGAC

R I L E M K V M E F F M K V Y G Y R G K CGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAG

H L G G S R K P D G A I Y T V G S P I D CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGAT

Y G V I V D T K A Y S G G Y N L P I G Q TACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAG

A D E M E R Y V E E N Q T R N K H L N P GCCGACGAGATGGAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACCTCAACCCC

N E W W K V Y P S S V T E F K F L F V S AACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGC

G H F K G N Y K A Q L T R L N H I T N C GGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGC N G A V L S V E E L L I G G E M I K A G AATGGCGCCGTGCTGAGCGTGGAGGAGGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGC T L T L E E V R R K F N N G E I N F R S ACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCT Bal II G G G E G R G S L L T C G D V E E N P G GGCGGCGGAGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGC <u>PR</u>MDYKDHDGDYKDHDIDYK CCTAGGATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG D D D K M A P K K K R K V G I H G V P GATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAGGTGGGCATTCACGG<u>GGTACC</u>C Knn I ZFP-R A A M A E R P F O C R I C M R N F S d r GCCGCTATGGCTGAGAGGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTGACCGC shlsrHIRTHTGEKPFACDI TCCCACCTGTCCCGCCACATCCGCACCACACAGGCGAGAAGCCTTTTGCCTGTGACATT CGRKFA<u>rsdalar</u>HTKIHTG TGTGGGAGGAAGTTTGCCCGCTCCGACGCCCTGGCCCGCCATACCAAGATACACGGGGA S Q K P F Q C R I C M R N F S d r s n l TCTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTGACCGCTCCAACCTG S T H I R T H T G E K P F A C D I C G R TCCCGCCACATCCGCACCACACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGG K F A t s q s l t r H T K I H L R AAGTTTGCCACCTCCGGCTCCCTGACCCGCCATACCAAGATACACCTGCGG Fok-R G S Q L V K S E L E E K K S E L R H K L GGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTG BamH I K Y V P H E Y I E L I E I A R N S T O D AAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGAC R I L E M K V M E F F M K V Y G Y R G K CGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAG H L G G S R K P D G A I Y T V G S P Т D CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGAT Y G V I V D T K A Y S G G Y N L P I G O

27

TAC	GGC	GTG	ATC	GTG	GAC	ACA	AAG	GCC	TAC	AGC	GGC	'GGC	TAC	'AAT	'CTG	CCT	'ATC	GGC	CAG
A GCC	D GAC	E GAG	M ATG	Q CAG	R AGA	Y .TAC	V GTG	K AAG	E GAG	N AAC	Q CAG	T ACC	R CGG	N AAT	K 'AAG	H CAC	I ATC	N AAC	P CCC
N AAC	E GAG	W TGG	W TGG	K AAG	V GTG	Y TAC	P CCT	S AGC	S AGC	V GTG	T SACC	E 'GAG	F STTC	K 'AAG	F TTC	L CTG	F TTC	V GTG	S AGC
G GGC	H CAC	F TTC	K AAG	G GGC	N SAAC	Y TAC	K AAG	A GCC	Q CAG	L CTC	T SACC	R AGG	L CTG	N AAC	H CAC	K AAA	T ACC	N AAC	C TGC
N AAT	G GGC	A GCC	V GTG	L CTG	S AGC	V GTG	E GAG	E GAG	L CTG	L CTG	I SATC	G GGC	G GGC	E 'GAG	M ATG	I ATC	K AAA	A GCC	G GGC
T ACC	L CTG	T ACA	L .CTG	E GAG	E GAG	V GTG	R CGG	R CGC	K 'AAG	F TTC	N LAAC	N 'AAC	G GGC	E 'GAG	I ATC	N 'AAC	F TTC	* TGA	ТАА
<u>СТС</u> Хho	<u>GAG</u> I	TCT	AGA	GGG	CCC	GTT	ТАА	ACC	CGC	TGA	TCA	GCC	TCG BGF	A <u>CT</u> I poly()	' <u>GTG</u> A) sigi	<u>CCT</u> nal	TCT	AGT	<u>TGC</u>
<u>CAG</u>	CCA	TCT	GTT	GTT	'TGC	CCC	TCC	CCC	GTG	CCI	TCC	TTG	ACC	CTG	GAA	GGT	GCC	ACT	<u>CCC</u>
<u>ACT</u>	GTC	CTT	TCC	TAA	TAA	AAT	GAG	GAA	ATT	GCA	TCG	CAT	TGT	CTG	AGT	'AGG	TGT	CAT	TCT
<u>ATT</u>	CTG	GGG	GGT	GGG	GTG	GGG	CAG	GAC	AGC	AAG	GGG	GAG	GAT	TGG	GAA	GAC	AAT	AGC	<u>AGG</u>
<u>CAT</u>	GCT	GGG	GAT	GCG	GTG	GGC	TCT	ATG	<u>IG</u> CT	TCI	ACT	'GGG	CGG	TTT	'TAT	'GGA	.CAG	CAA	GCG
AAC	CGG	AAT	TGC	CAG	CTG	GGG	CGC	CCT	CTG	GTA	AGG	TTG	GGA	AGC	CCT	'GCA	AAG	TAA	ACT
GGA	TGG	СТТ	TCT	CGC	CGC	CAA	GGA	TCT	'GAT	GGC	GCA	GGG	GAT	'CAA	GCT	CTG	ATC	AAG	AGA
CAG	GAT	GAG	GAT	CGT	TTC	GCA	TGA	TTG	AAC	AAG	ATG	GAT	TGC	ACG	CAG E	GTT Eag I	CTC	<u>CGG</u>	<u>CCG</u>

Supplementary Figure 5: ELISA data for DNA binding by all the ZFPs designed and assembled against the *ntl* locus. Each ZFN was evaluated for DNA binding using an ELISA assay performed as described (Bartsevich et al, 2003). The first sample is a negative control; the second and third sample, respectively, are positive controls and represent ZFN-R and ZFN-L from previous work (Urnov et al., 2005). Each ZFN is annotated by a specific base-pair position within the *ntl* locus (the cDNA cannot be used as a reference in this context because some ZFNs span exon/intron boundaries). In this nomenclature, position #1 corresponds to the transcription start site of the *ntl* gene as per the UCSC genome browser DanRer5 zebrafish genome annotation (July 2005); an "r" before the ZFN name indicates that its primary recognition is the Crick strand. The ZFNs highlighted in red form pairs 2 and 3, respectively, used in the primary text. Pair 2 ZFN-R is 1607a1, Pair 3 ZFN-L is 1607b1, and Pair 2/3 ZFN-L is r1600a1.



Supplementary Figure 6: The full DNA/amino acid sequence of the ZFNs used for *ntl* gene disruption (the vector backbone and overall arrangement is identical to that shown in Supplementary Fig. 4).

ZFN expression vector sequence

1) The ZFP DNA binding helices are shown in lower case and underlined.

2) The 2A peptide sequence is double-underlined.

3) The Eag I restriction site and the T7 promoter used to linearize the plasmid and drive mRNA transcription, respectively, are indicated.

ntl pair 2

CTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAG T7 promoter GGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTGATCCACTAGTCCAGTGTGGTGGAA Y DHDG D Y K D Η D М D Κ Ι D Υ Κ TTCGCCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG Κ A P Κ Κ Κ R Κ VG I H Ρ D D D D М G V GATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAGGTGGGCATTCACGGGGTACCC Kpn I ZFP-L Α А Α Ε R Ρ F Q С R Ι С М R Ν F М S r S GCCGCTATGGCTGAGAGGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGTAGT s r H I R т Η Т G Ε Κ Ρ F Α С Ι d n 1 D GACAACCTGAGCCGGCACATCCGCACCACACAGGCGAGAAGCCTTTTGCCTGTGACATT Α <u>d s s t r k k</u> Κ G С G R Κ F н Т Ι Η Т TGTGGGAGGAAATTTGCCGACAGCAGCACCCGCAAAAAGCATACCAAGATACACGGGA S 0 Κ Ρ F Q С R I С М R Ν F S r s d h l TCTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGCTCCGACCACCTG GΕ F R Т Н Т Κ Ρ A C D Ι С R Η Ι G S а TCCGCCCACATCCGCACCCACACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGG <u>s n a r k</u> t Т Κ F Α h Η Κ Ι Η L R AAGTTTGCCCACTCCAACGCCCGCAAGACCCATACCAAGATACACCTGCGG Fok-L G K S ΕL Ε Ε Κ S L V Κ S Ε L R Η Κ L 0 GGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTG BamH I Ε R Κ Υ V Ρ Η Ε Y Ι L Ι Ε Ι А Ν S Т 0 D AAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGAC R Ι L Ε М Κ V Μ Ε F F М Κ V Υ G Υ R G K CGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGGAAAG

H L G G S R K P D G A I Y T V G S P I D CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGAT Y G V I V D T K A Y S G G Y N L P I G Q TACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAG A D E M O R Y V E E N O T R N K H I N P GCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCC N E W W K V Y P S S V T E F K F L ਸ V S AACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGC G H F K G N Y K A Q L T R L N H I T N C GGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGC N G A V L S V E E L L I G G E M I K A G AATGGCGCCGTGCTGAGCGTGGAGGAGGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGC T L T L E E V R R K F N N G E I N F R S ACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCAGATCT Bgl II <u>G G E G R G S L L T C G D V E E N P G</u> GGCGGCGGAGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCGGC <u>PR</u>MDYKDHDGDYKDHDIDYK CCTAGGATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG D D D K M A P K K K R K V G I H G V P GATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAGGTGGGCATTCACGG<u>GGTACC</u>C Kpn I ZFP-R A A M A E R P F O C R I C M R N F S r s GCCGCTATGGCTGAGAGGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGTAGT d t l s q H I R T H T G E K P F A C D I GACACCCTGAGCCAGCACATCCGCACCACACAGGCGAGAAGCCTTTTGCCTGTGACATT C G R K F A <u>d r s a r t r</u> H T K I H T G TGTGGAAGGAAATTTGCCGACAGGAGCGCCCGCACACGGCATACCAAGATACACGGGA S O K P F Q C R I C M R N F S <u>r s d d l</u> TCTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGTAGTGACGACCTG S K H I R T H T G E K P F A C D I C G R AGCAAGCACATCCGCACCACACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGG K F A d n s n r i k H T K I H L R AAATTTGCCGACAACAGCAACCGCATAAAGCATACCAAGATACACCTGCGG Fok-R

G S Q L V K S E L E E K K S E L R H K L <u>GGATCC</u>CAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTG BamH I

K Y V P H E Y I E L I E I A R N S T Q D AAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGAC K V M E F мкvуg R I L Ε М F Y R G K CGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGGAAAG H L G G S RKPDGA I Y T V G S Ρ I D CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGAT Y G V Ι V D T K A Y S GGYNL Ρ Ι G Ο TACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAG ORYVE Ε Ν QТ R N K Ρ Δ D Ε М Η Ι Ν GCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCC K V Y P S SVTEFKF F Ν E W W L V S AACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGC Т K G N Y K A O L R L N H Т С G H F I Ν GGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGC N G A V L S V E E L L I G G E M Т KAG AATGGCGCCGTGCTGAGCGTGGAGGAGGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGC Т L Т L E Ε V R R Κ F Ν Ν G Ε Ι Ν F * ACCCTGACACTGGAGGAGGTGCGGCGCGAGTTCAACAACGGCGAGATCAACTTCTGATAA CTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGC Xho I BGH poly(A) signal CAGCCATCTGTTGTTTGCCCCTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCC ACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCT CATGCTGGGGATGCGGTGGGCTCTATGGCTTCTACTGGGCGGTTTTATGGACAGCAAGCG AACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACT GGATGGCTTTCTCGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGCTCTGATCAAGAGA CAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCCGGCCG Eagl

CTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAAT<u>TAATACGACTCACTATAG</u> *T7 promoter*

<u>G</u>GAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTGATCCACTAGTCCAGTGTGGTGGAA

M D Y K D H D G D Y K D H D I D Y K TTCGCCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG

D D D K M A P K K K R K V G I H G V P GATGACGATGACAAGATGGCCCCCAAGAAGAAGAAGAGGAAGGTGGGCATTCACGG<u>GGTACC</u>C *Kpn I*

ZFP-L

A A M A E R P F Q C R I C M R N F S <u>r s</u> GCCGCTATGGCTGAGAGGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGTAGT

<u>d</u> n <u>l</u> s <u>r</u> H I R T H T G E K P F A C D I GACAACCTGAGCCGGCACATCCGCACCCACACAGGCGAGAAGCCTTTTGCCTGTGACATT

C G R K F A <u>d s s t r k k</u> H T K I H T G TGTGGGAGGAAATTTGCCGACAGCAGCACCCGCAAAAAGCATACCAAGATACACACGGGA

S Q K P F Q C R I C M R N F S $\underline{r \ s \ d \ h \ l}$ TCTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGCTCCGACCACCTG

<u>s</u> a H I R T H T G E K P F A C D I C G R TCCGCCCACATCCGCACCACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGG

K F A <u>h s n a r k t</u> H T K I H L R AAGTTTGCCCACTCCAACGCCCGCAAGACCCATACCAAGATACACCTGCGG

Fok-L

G S Q L V K S E L E E K K S E L R H K L <u>GGATCC</u>CAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTG *BamH I*

K Y V P H E Y I E L I E I A R N S T Q D AAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGAC R I L E M K V M E F F M K V Y G Y R G K

CGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAG

H L G G S R K P D G A I Y T V G S P I D CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGAT

Y G V I V D T K A Y S G G Y N L P I G Q TACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAG

A D E M Q R Y V E E N Q T R N K H I N P GCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCC

N E W W K V Y P S S V T E F K F L F V S AACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGC

G H F K G N Y K A Q L T R L N H I T N C GGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGC N G A V L S V E E L L I G G E M I K A G AATGGCGCCGTGCTGAGCGTGGAGGAGGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGC T L T L E E V R R K F N N G E I N F R S ACCCTGACACTGGAGGAGGTGCGGCGCGAGGTTCAACAACGGCGAGATCAACTTCAGATCT Bal II G G G E G R G S L L T C G D V E E N P G GGCGGCGGAGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCCCCGGC <u>PR</u>MDYKDHDGDYKDHDIDYK CCTAGGATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG D D D K M A P K K K R K V G I H G V P GATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAGGTGGGCATTCACGG<u>GGTACC</u>C Kpn I ZFP-R A A M A E R P F O C R I C M R N F S r s GCCGCTATGGCTGAGAGGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGTAGT d t l s q H I R T H T G E K P F A C D I GACACCCTGAGCCAGCACACCGCACCACACGGCGAGAAGCCTTTTGCCTGTGACATT C G R K F A <u>d r s a r t r</u> H T K I H T G TGTGGAAGGAAATTTGCCGACAGGAGCGCCCGCACACGGCATACCAAGATACACGGGGA S Q K P F Q C R I C M R N F S r s d s 1 TCTCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTAACTTCAGTCGCTCCGACTCCCTG S K H I R T H T G E K P F A C D I C G R TCCAAGCACATCCGCACCACACAGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGG KFA<u>dnsnrik</u>HTKIHLR AAGTTTGCCGACAACTCCAACCGCATCAAGCATACCAAGATACACCTGCGG Fok-R G S Q L V K S E L E E K K S E L R H K L GGATCCCAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTG BamH I K Y V P H E Y I E L I E I A R N S T Q D AAGTACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGAC R I L E M K V M E F F M K V Y G Y R G K CGCATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAG

H L G G S R K P D G A I Y T V G S P I D CACCTGGGCGGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGCAGCCCCATCGAT

Y G V I V D T K A Y S G G Y N L P I G Q TACGGCGTGATCGTGGACACAAAGGCCTACAGCGGCGGCTACAATCTGCCTATCGGCCAG A D E M Q R Y V E E N Q T R N K H Ι N P GCCGACGAGATGCAGAGATACGTGGAGGAGAACCAGACCCGGAATAAGCACATCAACCCC E W W K V Y P S s v т EFKF L F V S Ν AACGAGTGGTGGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCGTGAGC G Н F K G N Y K A Q L T R L N H Т т N С GGCCACTTCAAGGGCAACTACAAGGCCCAGCTGACCAGGCTGAACCACATCACCAACTGC SVEE T. тддем N G AVT. T. Т Κ Α G AATGGCGCCGTGCTGAGCGTGGAGGAGGCTGCTGATCGGCGGCGAGATGATCAAAGCCGGC T L T L E E V R R K F N N G E I N F * ACCCTGACACTGGAGGAGGTGCGGCGCAAGTTCAACAACGGCGAGATCAACTTCTGATAA CTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGC Xho I BGH poly(A) signal CAGCCATCTGTTGTTTGCCCCTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCC **ACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCT** CATGCTGGGGATGCGGTGGGCTCTATGGCTTCTACTGGGCGGTTTTATGGACAGCAAGCG AACCGGAATTGCCAGCTGGGGCGCCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACT GGATGGCTTTCTCGCCGCCAAGGATCTGATGGCGCGCGGGGGATCAAGCTCTGATCAAGAGA CAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCG Eagl

Supplementary Figure 7: ZFN injection can cause hypomorphic phenotypes.

(A) mRNA encoding high-fidelity, obligate-heterodimer (Miller et al., 2007) ZFNs was injected into 1-cell embryos from a cross between wildtype and ntl^{b195} heterozygous fish. ZFN-injected embryos have diminished or no notochord, misshapen somites, and reduced tail mesoderm, with phenotypes resembling both the null ntl^{b195} homozygous mutant (see Fig. 3C) and the hypomorphic ntl^{b487} homozygous mutant (top right panel). (B) When processed by in situ hybridization to detect notochordal ntl expression, ntl ZFN-injected embryos display a range of phenotypes, from normal expression (i) to disrupted and patchy expression (ii, iii), to completely absent expression (not shown). The patchy notochord phenotype is typical of that in hypomorphic ntl^{b487} homozygous mutant embryos (v, vi); normal wildtype siblings are shown in panel (iv).



Supplementary Figure 8: Injection of *ntl* **ZFN-encoding mRNA into zebrafish embryos induces disruption of the** *ntl* **gene in somatic cells.** Nine individual *ntl*-like embryos (#1-#9) were assayed to assess the percentage of *ntl* loci that had undergone NHEJ using an assay based on the mismatch-sensitive endonuclease ("Surveyor"; Miller et al., 2007), and most showed clear evidence of ZFN-induced NHEJ, ranging from 3-17% of chromatids.



Supplementary Figure 9: Injection of *ntl* **ZFN-encoding mRNA into wildtype embryos induces tail phenotypes in juvenile and adult fish.** DNA was prepared from small posterior tissue samples obtained from tailless fish like those shown in Fig. 4A. The *ntl* locus surrounding the ZFN-cleavage site was amplified and sequenced from each of three tailless embryos. In every case, *ntl* mutant-bearing amplicons represent a substantial fraction of the total (Sample 1, 5/25 (20%) *ntl*-bearing chromatids, 2 different alleles; Sample 2, 3/30 (10%) *ntl*-bearing chromatids, 1 allele; Sample 3, 8/29 (28%) *ntl*-bearing chromatids, 4 different alleles). The frequency of each allele is indicated after the allele description.

Tail sample #1	ZF	N-L	_			ZF.	N-R								
	Leu Asp	Pro Asn	Ala	Met	Tyr	Ser	Val	Leu							
	CTCGAC	CCTAAI	GCA	ATG	TAC	TCG	GTC	CTG							
TCAGAGCCAGTGTCACCGGI	CTCGAC	CCTAAI	G::		TAC	TCG	GTC	CTG	CTGG)	ATTT	TGT	GGCG	GC (Δ	15)2	2x
TCAGAGCCAGTGTCACCGGI	CTCGAC	CCTAAI	GCA	ATC	aat	GTA	СТС	GGT	CCTG	CTGG	ATT	TTGT	GG (+	4)3	}x

Tail sample #3	ZFN-L	ZFN-R	
	Leu Asp Pro Asn Ala Met	. Tyr Ser Val Leu	
	CTCGACCCTAATGCAAT	GTACTCGGTCCTG	
TCAGAGCCAGTGTCACCGG	CTCGACCCTAATG::::	: TACTCGGTCCTGC	TGGATTTTGTGGCGGC (Δ 5)2x
TCAGAGCCAGTGTCACCGG	CTCGACCCTAATGC:::	::::TCGGTCCTGC	TGGATTTTGTGGCGGC (Δ 7) 1x
TCAGAGCCAGTGTCACCGG	FCTCGACCCTAATGCAAT	caatGTACTCGGTC	CTGCTGGATTTTGTGG (+4) $4x$
TCAGAGCCAGTGTCACCGGI	CTCGACCCTAATGtaat	gtaagaacatTGTA	CTCGGTCCTGCTGGAT (Δ 3+14) 1x

Supplementary Figure 10: Injection of *ntl* ZFN-encoding mRNA into wild-type embryos induces biallelic disruption of the *ntl* gene in somatic cells. (A) Wild-type embryos injected with mRNA encoding *ntl*-targeting ZFN Pair 2 (carrying wild-type, rather than obligate-heterodimer FokI domains) show *ntl*-like phenotypes (middle panel), and some show additional mild necrosis (right panel); uninjected embryos are shown in the left panel. In situ hybridization of representative embryos to detect notochordal *ntl* expression are inset. (B) Allelic diversity for one representative *ntl*-targeting ZFN mRNA-injected embryo from (A). In all 3 embryos, a large number of unique *ntl* alleles were observed (>20) and 64-81% of the sequenced chromatids carried an induced mutation (Embryo 1, 38/59 *ntl*-bearing chromatids, Embryo 2, 44/63 *ntl*-bearing chromatids).



Supplementary Figure 11: Analysis of ZFN action at potential off-target sites.

(A) Experimental scheme to evaluate ZFN action at potential off-target sites. (B) The ZFN recognition site in the *ntl* gene, and a qualitative representation of *in vitro* site selection results showing the ZFN recognition sites determined experimentally. (C) A listing of the potential off-target sites for the ZFNs; the mismatches relative to the ZFN consensus sites are shown in lowercase.



C

ZFN-treated animals

#	Pair	Loc	Pos	Sequence	Gene	Coding?	# of <i>ntl</i> progeny analvzed	Wild-type chromatids	Mutant chromatids
		chr19	20,038,174	TCTCGACCCTAATGCAATGTACTCGGTCCTGC	nti	Exon 2	15	0/15	15/0
1	2/3	chr8	15,071,183	c <mark>CTCGgCCCgAAc</mark> aCc_gG <mark>cACTCcGTCgcc</mark> C	prkcz	Intron 12	7	26/26	0/26
2	2/3	Unmap.	125,224,504	c <mark>CTCGcCaCacAc</mark> aCAcac <mark>cACTCGGaCCgG</mark> g		Desert	7	25/25	0/25
3	2/3	chr11	1,589,395	T <mark>CTCtgCCCgccc</mark> GCcgcG <mark>ctCTCGGTCCcG</mark> g	zgc:136857	Exon 6	9	34/34	0/34
4	2	chr20	2,693,549	a <mark>CTCGgCCaactc</mark> aCtggG <mark>cACTCGGTtCaG</mark> t		Intergenic	7	28/28	0/28
5	2	chr20	285,141	c <mark>tgCGgCCCgAAc</mark> tCctcG <mark>TACTCGGcggcG</mark> C	si:ch211-241j12.3	Exon 7 (stop)	2	7/7	0/7
6	3	chr3	26,364,497	a <mark>CTCtGCtCTAAT</mark> attgaG <mark>TACTtGGTCaTG</mark> g		Intron 1	4	15/15	0/15
7	3	chr19	25,689,552	g <mark>CTCacCCCaAAc</mark> aatcTt <mark>gACTCGGTtCaG</mark> g		Intergenic	3	12/12	0/12

Supplementary Figure 11: Analysis of ZFN action at potential off-target sites, continued.

Sequencing of the *ntl* progeny fails to identify mutant chromatids at potential off-target sites. Each of the following panels presents the sequencing and additional genotyping data for all the potential off-target sites, listed by number as in Fig. S11C. The chromatids were genotyped by amplifying the respective stretches by PCR (primers listed in Supplementary Table 3), followed by TOPO cloning and sequencing. (a) Representative chromatograms selected from *ntl* progeny of founders A-D (when applicable) are shown. In each case, the potential off-target sites are indicated with black boxes. In some cases, in addition to direct sequencing, we were able to use a mismatch sensitive endonuclease assay (unless the animal is heterozygous for a naturally occurring SNP in the region), or, if the ZFN target site overlaps with a restriction enzyme site, to do a loss-of-RFLP assay. (b) Gel-based genotyping assays using restriction enzymes or mismatch sensitive nucleases. In both cases, we used DNA extracted from wild-type progeny as a control.

Supplementary Figure 11, off-target 1

1a



b



Supplementary Figure 11, off-targets 2 and 3



Sequencher	
le Edit Select Contig Sequence View Window Help	
Contig[0023]	
Overview Summary Cut Map Find Show Chromatograms ReAlign	
3 SH 128124-00_26_17_F04 GCATTTATCCTTTTTAGATCCGGTTTCACCCTC 3 SH 128125-18 24 17 F05 GCATTTATCCTTTTTAGATCCGGTTTCAC	GCCACACACACCACCCGGCACCGGGGATACAGCTCTCAGTGTGCAGCGTCCCGATCAAC GCCACACACACCACCGGGCACCGGGGATACAGCTCTCAGTGTGCAGCGTCCCGATCAAC
SH_125134-27_33_17_604 GCATTTATCCTTTTTAGATCCGGTTTCACCCTC	GCCACACACACCACTCGGACCGGGATACAGCTCTCAGTGTGCAGCGTCCCGATCAAC
g sh_120120-29_20_17_HH4 GCATTTATCCTTTTTAGATCCGGTTTCACCGET] widtype	GECACACACACCACTEGGACCEGGATACAGCTETEAGTGTGEAGCGTECEGATEAAC GECACACACACCACTEGGACCEGG
000 frag bases \$30 1240 1250 1260	
# consensus position 1,203	
	,
Chromatograms from Contig[0023]	-30 35 17 J04 Prognent basis #118 207
TTEACCCTCGCCACACA	
1 0 0 0 0 0 0 0 0 0	
	$\Lambda \Lambda $
	CIG 7 00 Fragment base #1.241200 CACACACCACCACICGGACCGGGACACAC
* A A A T A A A A A A A A A A A A A A A	a T A T C C C G A G T A G A G T A G A G T A G A G A G
	A A A A A A A A A A A A A A A A A A A
T T C A C C C T C G C C A C A C A	7 30 17 C04 Fragment bases #1,2561,255
	An Anna Marah, AAn An AAAAAAAA
T T C A C C C T C G C C A C A C A C A	CACACCACICGGACCGGGGAAC
	ΛοΛοοΛοοΛοΑΛοΛοΛοΛοΛ

3a

Sequencher	
File Edit Select Contig	Sequence View Window Help
Contig[0036]	
Overview Summer	y Cut Map Find Show Divonatograms Rentligner
2 54,12578-38,1120,17,148 2 54,2578-38,142,157,643 2 54,25786-39,1421,17,884 2 54,25778-27,1118,17,004 2 645494	COTICESCATESCATELAATEACECESA DETETESEGEESEGEESEGEESEGEESEGATATECETTSSECASEATTASTECETTASTEAS SECTESSATESECETTAATEACECESATETTSECESECESEGEESEGEESEGTATECETTSSECASEATEATECETTASTEAS SECTESSATESECETTAATEACECESSATETETSECESECESSECES
to stop bases 8.92 consensus bases selected at consensus position 241	1320 1330 1240 1250 1260 1270 1288 1290 1300 CCTCCSGCATCGCCCTCAATCACCCCGCA
🔄 Chromatograms fro	m Cantig[0036]
	h e h e h e e e e e e e e e e e e e e e
1: EEEA	A A B B B B C C (Statuti singet 10) (Statut
	hmmmmmmmmm
<u>****</u> [₩^^^	

Supplementary Figure 11, off-target 4

4a

b





Supplementary Figure 11, off-target 5

5a



BseRI - + + + + + + + +

Supplementary Figure 11, off-targets 6 and 7





7a



Supplementary Table 1

ZFN pairs designed against the *golden* locus (data on their testing is shown in Fig. 2A)

ZFN pair # The codon in the <i>slc24a5</i> locus at which the DSB is induced is indic	
_	by the cognate amino acid number in the ORF
1	Ile 166
2	Ile 166
3	Ser 355
4	Ser 355
5	Asp 381
6	Asp 381
7	Asp 381
8	Asp 381
9	Asp 397
10	Asp 397
11	Val 399
12	Ala 400
13	Ala 400
14	Val 437
15	Val 437
16	His 471
17	His 471
18	Glu 500
19	Glu 500
20	Glu 500
21	Glu 500

Supplementary Table 2

ZFN pairs designed against the *ntl* locus (data on their testing is shown in Fig. 3A)

ZFN pair #	The codon in the <i>ntl</i> locus at which the DSB is induced is indicated by the
	cognate amino acid number in the ORF
1	Leu 14
2	Ala 79
3	Ala 79
4	Trp 95
5	Asn 124

Supplementary Table 3 Primer sequences used to genotype potential off-target sites for ZFN action

#	Primers
1	CAGAAGTAAAGCCTGTTCAGTTTCATC
	TGCTTCTCTGTCTGTTCTGTGCTG
2	ACTCTCACACACACAAGATTAACGATG
	GCATGAAATCTCTATTCTCATCTATTCTGC
3	GTGTGTGTGGGGGATGTGAAGATG
	ATTAGGCCGGATCTTACACCAGC
4	ACCTTGTATCCGCACACTTGATTT
	TTGCAACTCTAATCCAGCACCTC
5	GTGTTGTACAGGTATGACCGCTCTT
	GCTAAAAGGGTCCGTCAGAAGGC
6	ACATAAGACTTTTTGTGGTCC
	TCAGATATGCAAACCAACATGTGC
7	GGCATACTGAACTAGCGGCTACC
	CTTCACTCTGCAGTCCAGCTCAC

Supplementary Table 4 Plasmid designations for ZFN expression vectors

Target gene	ZFN pair #	Yeast expression vectors	Zebrafish expression vectors
gol	1	pSGYDLeu-12776, pSGYDHis-12775	pVAX-12776-2A-12775
gol	14	pSGYDLeu-12805, pSGYDHis-12804	pVAX-12805-2A-12804
gol	15	pSGYDLeu-12806, pSGYDHis-12804	pVAX-12806-2A-12804
ntl	2	pSGYDLeu-13370, pSGYDHis-13368	pVAX-13370-2A-13368
ntl	3	pSGYDLeu-13370, pSGYDHis-13369	pVAX-13370-2A-13369