

## Supporting information

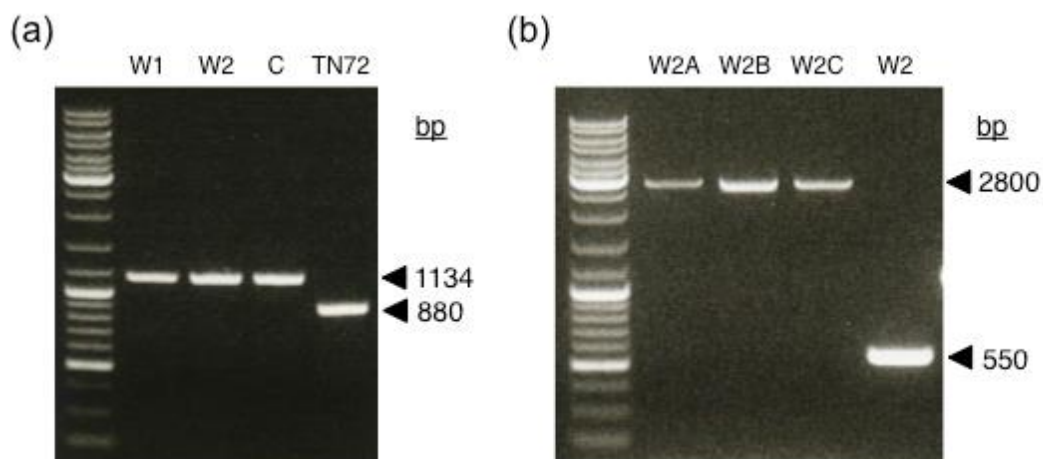
### **Codon reassignment to facilitate genetic engineering and biocontainment in the chloroplast of *Chlamydomonas reinhardtii***

Rosanna E. B. Young and Saul Purton

**Figure S1** PCR demonstrating homoplasmic integration of transgenes into the *C. reinhardtii* chloroplast genome. Primers are listed in Table S2.

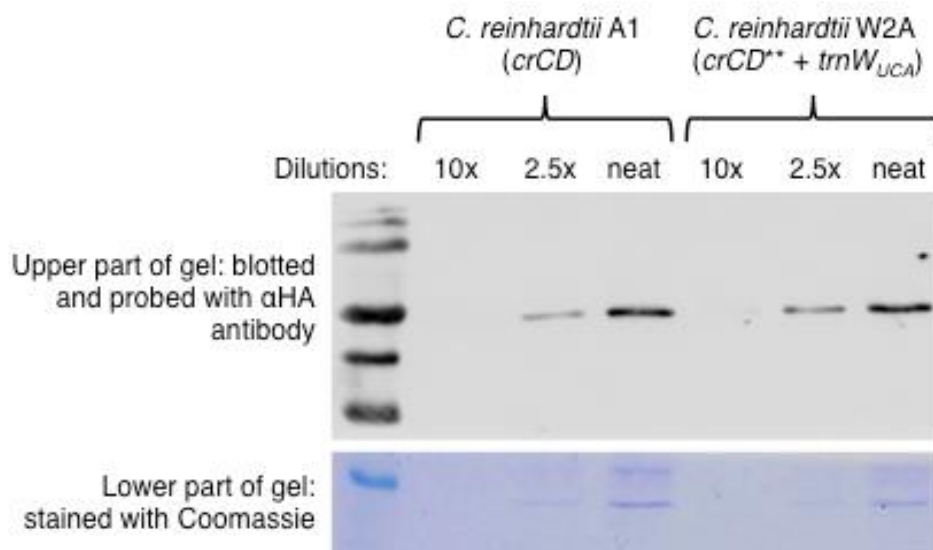
(a) Integration of pCD\*\* into strain TN72. W1 and W2 are two independent transformants. C is an empty-vector positive control strain.

(b) Integration of pWUCA1 into strain W2, introducing *trnW<sub>UCA</sub>* into a neutral region of the chloroplast genome next to *psaA3*. Three independent transformants are shown.



**Figure S2** SDS-PAGE of whole cell lysates for *C. reinhardtii* strains containing an intact *crCD* gene or a combination of *crCD\*\** with *trnW<sub>UCA</sub>*.

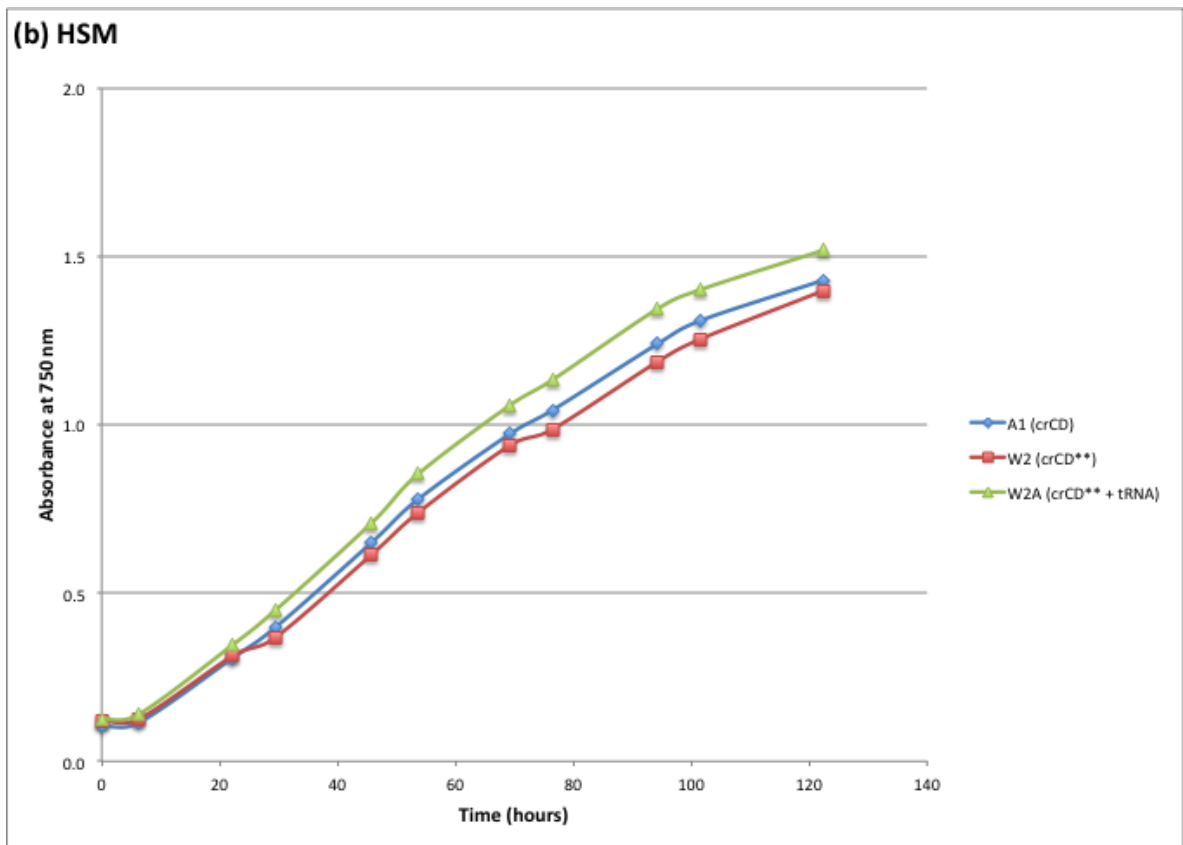
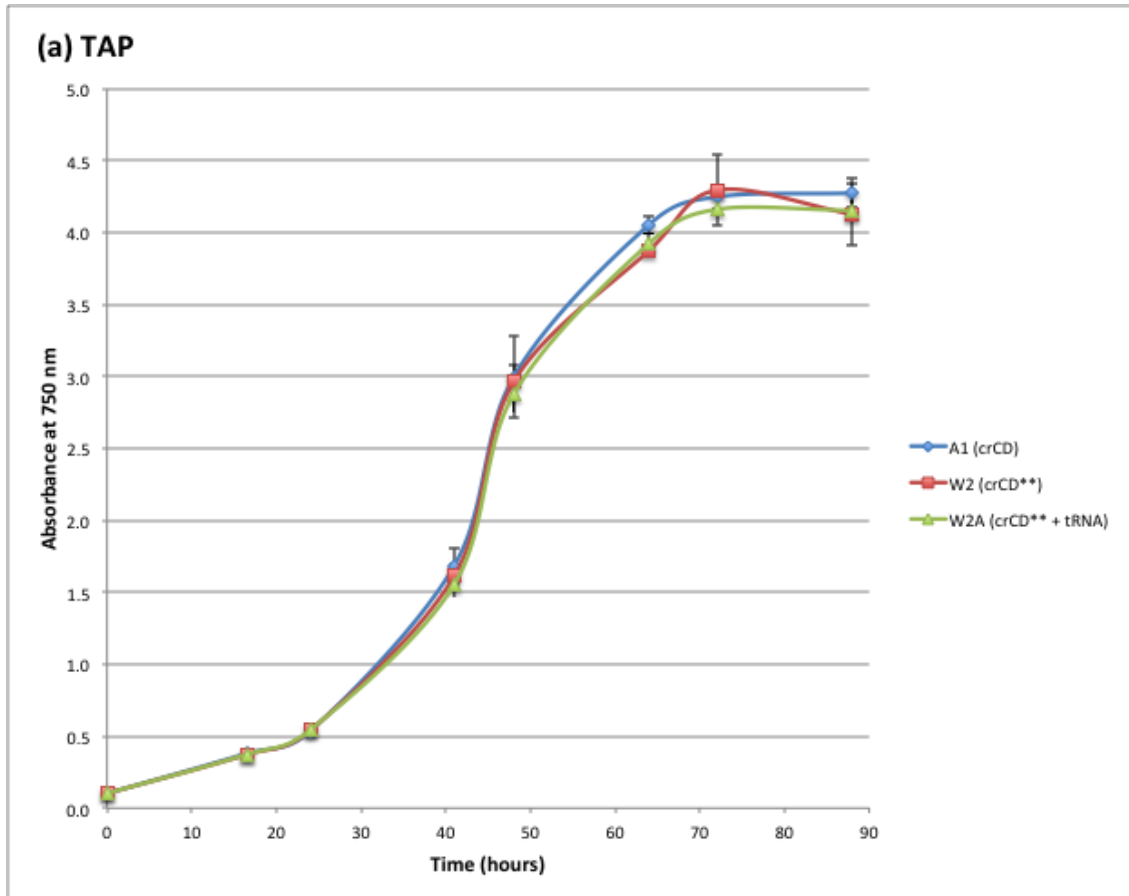
TAP cultures of the two strains were set up with a starting density of  $OD_{750} = 0.4$  and grown for 24 h, reaching an  $OD_{750}$  of 1.2-1.3. Cells were pelleted and resuspended in loading buffer to equal optical densities. The upper part of the gel was blotted onto membrane, which was then probed with an  $\alpha$ HA antibody to detect HA-tagged CrCD protein. The lower part of the gel was stained with Coomassie to confirm equal loading of the two strains. The left hand lane contains PageRuler Prestained Protein Ladder (Life Technologies): the bands visible on the blot are (from top) 170, 130, 100, 55, 40 and 35 kDa.



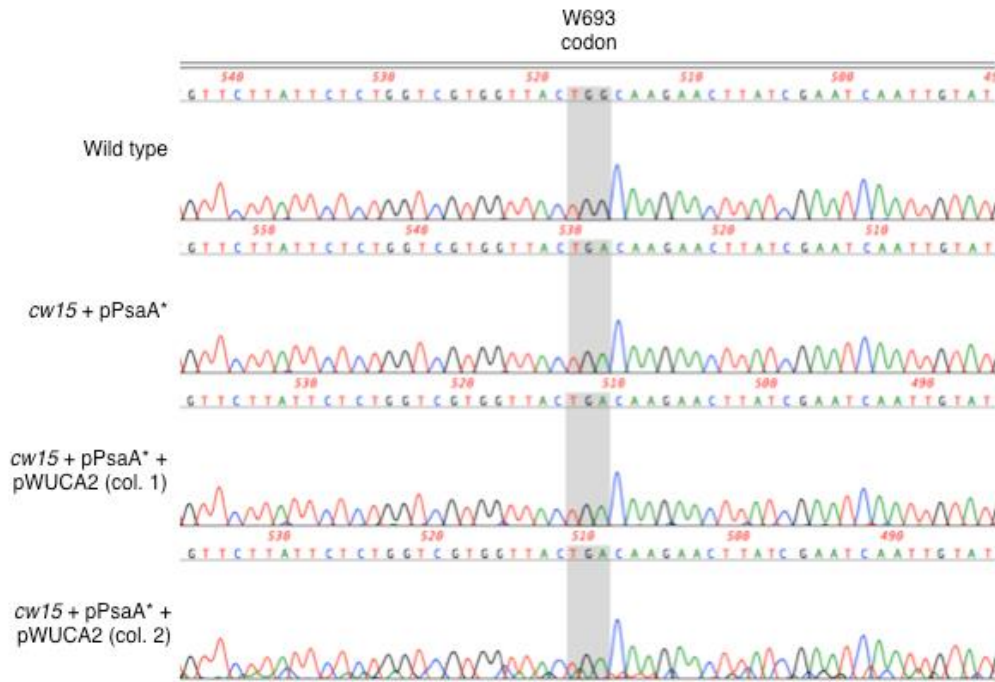
**Figure S3** Growth curves of *C. reinhardtii* cell lines with and without the synthetic tRNA gene, *trnW<sub>UCA</sub>*. All cultures were grown at 25 °C and shaken at 120 rpm under 90  $\mu$ E/m<sup>2</sup>/s light. Optical density was measured at 750 nm.

(a) Mixotrophic growth: cells were grown in 20 ml TAP (acetate-containing) medium. Error bars show  $\pm$  standard deviation of three cultures per cell line.

(b) Phototrophic growth: cells were grown in 20 ml HSM (minimal) medium.



**Figure S4** Confirmation of *psaA* exon 3 codon alteration in *C. reinhardtii* cell lines by DNA sequencing. The W693 codon was altered from TGG to TGA by genetic engineering to create the *cw15* + pPsaA\* cell line. When this mutant was transformed with the pWUCA2 plasmid, encoding *trnW<sub>UCA</sub>*, phototrophy was restored due to the readthrough of the TGA codon in *psaA*\*.



**Table S1** Primers used in the construction of plasmids.

Purpose	Reaction number	Primer sequences 5' to 3'
Construction of pCD**: introduce 2 TGG→TGA mutations into pCD (pRY127d) by 3-part Gibson assembly (split backbone in AmpR gene to facilitate PCR)	PCR 1 (0.4 kb)	AAGGTTTATGACAAATTCACCTTACAAGAC GTAAGTCAATTTCATGGAGCTACTTCTTG
	PCR 2 (2.9 kb)	AGTGAATTTGTTCATAAACCTTCTTCACC GCCTTCCTGTTTTTGGCTCA (in AmpR gene)
	PCR 3 (4.2 kb)	ACTTTCACCAGCGTTTCTG (in AmpR gene) TAGCTCCATGAAATTGACTTACAAATTGC
Construction of pPsaA*: introduce W693 TGG→TGA mutation into pBev1 by 1-part Gibson assembly	PCR 1 (9.3 kb)	GTGGTTACTGACAAAGAACTTATCGAATC ATAAGTTCTTGTTCAGTAACCACGACCAGA

**Table S2** Primers used to confirm homoplasmic integration of transgenes into *C. reinhardtii*.

Figure	Plasmid used for transformation	Primer sequences 5' to 3'	Product size for parental strain (bp)	Product size if homoplasmic integration (bp)
S1a	pCD**	GTCATTGCGAAAATACTGGTGC	880	1134
		CGGATGTA ACTCAATCGGTAG		
		ATAGGCTCTTCTCATGGATTCT CCTTATAATAAC		
S1b	pWUCA1	AGCACGTATCATTCTGTAGG	550	2800
		GGGAAAGAGGGACTTGAAC		
4a	pPsaA*	AGCACGTATCATTCTGTAGG	550	2511
		GGGAAAGAGGGACTTGAAC		
4b	pWUCA2	GTCATTGCGAAAATACTGGTGC	1256	2011
		GCAACAGGAACTTCTAAAGC		
5	pSty or pSde	GTCATTGCGAAAATACTGGTGC	880	1415
		CGGATGTA ACTCAATCGGTAG		
		ATAGGCTCTTCTCATGGATTCT CCTTATAATAAC		

Species	Strain	Accession No.	Size (bp)	GC Content	No. of CDS	Stop Codons			Identity of Genes with TGA Stop Codon
						TAA	TAG	TGA	
<i>Chlamydomonas reinhardtii</i>	137c background	BK000554	203828	34.5%	69	65	4	0	N/A
<i>Chlorella sorokiniana</i>	crs4	NC_023835	109811	34.0%	75	71	3	1	PROKKA_00016 (hypothetical protein) 5aa, may not be genuine
<i>Chlorella vulgaris</i>	C-27	NC_001865	150613	31.6%	174	124	31	19	All the short hypothetical proteins except 51aa, 100aa (LASTP hits) and poC1 (appears genuine from BLAST) from 997 have been overestimated?
<i>Coccomyxa sp.</i>	C-169	HQ693844	175731	50.7%	80	45	28	7	orf114, orf15, orf57, orf116, orf16, orf5A, orf33 and poC2.
<i>Dunaliella salina</i>	CCAP19/18	NC_016732	269044	32.1%	82	73	6	3	DUSAC_p041, three retron-coding ORFs and coding priming endonucleases. DUSAC_p043 DUSAC_p086
<i>Euglena gracilis</i>	Z	NC_001603	143171	26.1%	67	51	12	4	EugrCp014 (orf20); OS ribosomal protein 20) EugrCp028 (orf516)
<i>Lobosphaera incisa</i>	SAG:2007	KM462871	156031	27.8%	78	67	11	0	N/A
<i>Nannochloropsis gaditana</i>	CCMP526	KJ410682	114875	33.0%	132	108	21	3	Naga00012 (hypothetical protein) Naga00056 (psb28, orf79) Naga00145 (cpN; 1-domain protein haperone)
<i>Ostreococcus tauri</i>	OTTH0595	NC_008289	71666	39.9%	61	58	3	0	N/A
<i>Parachlorella kessleri</i>	SAG 11-11 1#	NC_012978	123994	30.0%	84	67	10	7	ftsH, orf5A, rpoA and hypothetical proteins
<i>Phaeodactylum tricorutum</i>	Not given	NC_008588	117369	32.6%	132	114	15	3	PhtrCp016 (orf90, hypothetical protein) PhtrCp067 (atpF, ATP synthase F0 subunit) PhtrCp092 (syfB; phenylalanyl-tRNA synthetase, beta chain)
<i>Porphyridium purpureum</i>	NIES 140	NC_023133	217694	30.3%	224	173	33	18	Y721_p115 (orfC; translation initiation factor) Y721_p031 (syfH; histidine RNA synthetase) Y721_p014 (sufC; ABC transporter) ...and many others.

**Table S3** Stop codon distribution in microalgae. Compiled from annotated chloroplast genomes in Genbank. ‘No. of CDS’ is the number of protein-coding genes annotated in Genbank for each genome as of November 2014. The number of annotated TAA, TAG and TGA stop codons is listed. For comparison, the chloroplast genome of *Arabidopsis thaliana* Columbia is 154,478 bp long with 49 TAA, 19 TAG and 12 TGA stop codons according to Sato et al. (1999) *DNA Res* **6**, 283-290.

## Appendix S1 DNA and amino acid sequences.

### 1. Cytosine deaminase gene used in plasmid pCD\*\*

#### (a) DNA sequence of *crCD\*\** gene, with TGG→TGA codon alterations in green.

ATGTCTAACAAACGCTTTACAAACAATTATTAACGCTCGTTTACCAGGTGAAGAAGGTTTA  
TGA CAAATTCACCTTACAAGACGGTAAAATTTACAGCTATTGATGCTCAATCTGGTGTAAATG  
CCAATTACTGAAAACCTTTAGATGCTGAACAAGGTTTAGTTATTCCACCATTTCGTTGAA  
CCACACATTCACCTTAGATACTACACAAACAGCTGGTCAACCAAACCTGGAACCAATCAGGT  
ACTTTATTTGAAGGTATTGAGCGTTGGGCTGAACGTAAAGCTTTATTAACACACGACGAC  
GTTAAACAACGTGCTTGGCAAACATTAATAATGGCAAATTGCTAACGGTATTCAACACGTA  
CGTACTCACGTAGACGTTTCTGATGCTACTTTAACAGCTTTAAAAGCTATGTTAGAAGTT  
AAACAAGAAGTAGCTCCA TGA ATTGACTTACAAATTGCTGCTTTCCACAAGAAGGTATT  
TTATCATACCCAAACGGTGAAGCTTTATTAGAAGAAGCTTTACGTTTAGGTGCTGATGTT  
GTTGGTGTATTCCACACTTCGAATTTACACGTGAATATGGTGTGAATCTTTACACAAA  
ACATTTGCTTTAGCTCAAAAATATGATCGTTAATTGATGTTCACTGTGACGAAATTGAT  
GACGAACAACACGTTTCGTTGAAACAGTAGCTGCTTTAGCTCACCACGAAGGTATGGGT  
GCTCGTGTACTGCTTACACACTACAGCTATGCACCTTACAACGGTGCTTACACTTCT  
CGTTTATTCCGTTTATTAATAAATGTCTGGTATTAACCTTCGTTGCTAACCCATTAGTAAAC  
ATTCACCTACAAGGTCGTTTCGATACTTACCCAAAACGTCGTGGTATTACACGTGTTAAA  
GAAATGTTAGAATCAGGTATTAATGTTTGGTTCACGACGACGTTTGTGGTCCCTGG  
TACCCTTAGGTACTGTAACATGTTACAAGTTTACACATGGGTTTACACGTATGTCAA  
TTAATGGGTTACGGTCAAATTAACGACGGTTAAACTTAATTACTCACCCTCTGCTCGT  
ACTTTAACTTACAAGACTACGGTATTGCTGCTGTTAACTCAGCTAACTTAATTATTTTA  
CCAGCTGAAAACGGTTTCGATGCTTTACGTCGTCAAGTCCAGTACGTTACTCAGTTCGT  
GGTGGTAAAGTTATTGCTTCAACTCAACCAGCTCAAACAACCTGTTTATTTAGAACAACCA  
GAAGCTATTGACTACAAACGTTACCCATACGATGTTCCAGATTACGCTTAATAA

#### (b) Translated sequence of *crCD\*\**.

Tryptophans W21 and W147 that were mutated from TGG to TGA are shown in green. HA tag is shown in red. Residues in turquoise show where the synthetic CrCD enzyme differs from natural *E. coli* CodA for improved substrate binding; see Young and Purton (2014) *Plant J* **80**, 915-925.

MSNNALQTIINARLPGEGL W QIHLQDGKISAIDAQSGVMPITENSLDAEQGLVIPPFVEPHIHL  
DTTQTAGQPNWNQSGTLFEGIERWAERKALLTHDDVKQRAWQTLKWQIANGIQHVRTHVDV  
SDATLTALKAMLEVKQEVAP W IDLQIA AFPQEGILSYPNGEALLEEALRLGADVGAIPHFEFT  
REYGVESLHKTFALAKYDRLIDVHCDEIDDEQSRFVETVAALAHHEGMGARVTASHTTAM  
HSYNGAYTSRFLRLLKMSGINLVANPLVNIHLQGRFDYTPKRRGITRVKEMLESGINVCFGHD  
DV CGPWYPLGTANMLQVLHMGLHVCQLMGYQINDGLNLITHHSARTLNLQDYGIAAGNSA  
NLILPAENGFDALRRQVPVRYSVRGGKVIASQPAQTTVYLEQPEAIDYKR YPYDVPDYA

### 2. Sequence of the synthetic *trnW<sub>UCA</sub>* gene with 100 bp flanks, as used in plasmids pWUCA1 and pWUCA2.

Colours represent the tRNA sequence (green), anticodon (blue), and MluI sites used for cloning (red). The -10 and -35 consensus promoter elements are in bold.

tcgatgc **ACGCGT** taacctgattaacaactatatcaataaaatcaattgtagtgaaatactctgattgacattaaat  
aataccatgataaaaattataataacaaatctacgtccttagttcagtcggttagaacgcaggtttcaaaacctgatgctcgtg  
ggtcaattcctacagggcgtgttttctaatgtactttgtgtaaaagtgctggttaaccttttaggttcggattgaacaat  
aatggcagttaagagtcactaaagctgctgtatag **ACGCGT** tcgatgc

### 3. Endolysin SPN9CC\_0043 From phage from *Salmonella* Typhimurium

The natural gene is described in Lim *et al.* (2014) *J Microbiol Biotechnol* **24**, 803-811; accession number YP\_006383882. For the present paper, a codon optimised version of this endolysin gene was designed in which a single TGG codon was altered to TGA (green). This was cloned into pWUCA2 using SapI and SphI restriction sites (yellow) to make plasmid pSty. The HA tag sequence is shown in red.

#### (a) DNA sequence

GCTCTTCTATGGCTATGAGTCCAGCTTTACGTAACCTCAGTAATTGCTGCTATTTCTGGTGGT  
GCTATTGCTATTGCTTCAGTATTAATTACTGGTCCAGGTGGTAACGACGGTTTAGAAGGTG  
TTCGTTACAAACCATACAAAGACGTAGTAGGTGTTTTAACAGTTTGTACGGTCACGTAGG  
TAAAGATATTATGTTAGGTAAAACATACACAGAAGCTGAATGTGAAGCTTTATTAAACAA  
AGACTTAGCTACTGTAGCTCGTCAAATTAACCCATACATTAAAGTAGATATTCCAGAACT  
ACACGTGGTGCTTTATACAGTTTTGTTTACAACGTTGGTGCTGGTAACTTCCGTACTTCAAC  
TTTATTACGTAAAATTAACCAAGGTGATATTAAAGGTGCTTGTGACCAATTACGTCGT TGA  
ACTTACGCTGGTGGTAAACAATGGAAAGGTTTAAATGACACGTCGTGAAATTGAACGTGAA  
GTTTGTATGTTTATGGGGTCAACAA TACCCATACGATGTTCCAGATTACGCT TAATAA GCATGC

#### (b) Translated sequence

MAMSPALRNSVIAAISGGAIAIASVLITGPGGNDGLEGVRYKPYKDVVGVLTVCYGHVVKDI  
MLGKTYTEAECEALLNKDLATVARQINPYIKVDIPETTRGALYSFVYNVGAGNFRSTLLRKI  
NQGDIKGACDQLRR WTYAGGKQWKGLMTRREIEREVCLWGQQ YPYDVPDYA

### 4. *Shewanella denitrificans* Sden\_1266 (hypothetical protein)

A codon optimised version of this gene was designed in which two TGG codons were altered to TGA (green). This was cloned into pWUCA2 using SapI and SphI restriction sites (yellow) to make plasmid pSde. The HA tag sequence is shown in red.

#### (a) DNA sequence

GCTCTTCTATGTCTATTAACAAAATTACTGATAACTTAACAACCTTCTGCTGAACACAGTGA  
AAACCACCACGACAACAACCAAGTAACTGCTGGTAAAAACATGTCTTCTCGTTCATTATTA  
TACGTTGCTGTAGCTTTAGGTTTATTCACTTGTGTAGCTGCTCAAGCTGAAACACAAGCTG  
CTTTAACACAAGAATTAAGACGAAGAAGGTTTATCTTTAGAACGTTTACAATTCTCTGT  
TTCTTTAGAATCATCTTCTACTACAAAAGGTACTGCTACTCCAAGTGGTGCAGAATCTACT  
GCTTTAACAACAGGTCTTCAATATAACGAATTAGCAAGTGAACCTAAACAATTACACCAA  
GGTGCTTCTTTTACGTTTA TGA TCACAAAAAAAACCTGCTTCTGTAGATTCTGTTAACTC  
TGTTGCTTCAATGAACAAAAGGTGATGCTGCAGGTTCTAATTTAGCAACTAAAGCTTCTGTA  
TTATCTGATCCAGCTGTAAGCATCTCCAGCTAACTTACGTACAGTAAACACTTTAGACC  
AAGGTTTATTAGGTATGACACGTGAACAAAAAATGGCTATTAAATCAGCTGAAAACCTCAG  
CTTCAGTACAAGCTTTACAATCTTCAGGTTTATTCCACTCTTTCAACATTTACGATGCTAAC  
ACTCACTTATTAGAAGATTTCGATGGTACTCATTCTACAGTACATTCTCAGTAACATTCTG  
ATGCTGATGTTGATGGTATTGGTTACAACGAATACGCTGATGTTTATGCTGAATTATACGT  
TTCTCAAGAAGGTGGTCCA TGA TTACTACTACTCAACTGAAGTATTCTCAATTGCTGGT  
AACTCATCTTACGATGATTACCGTGTTTTAAACAACCTTACAATCAGGTTACCAAACAGCTC  
ACTACGACGTTTAAATTGATTTATACGAAGTTGGTGTATCAAACCCAGTAGCTACTTTATC  
TTCAAACGATACTAACGCTTTATACGCTTTACCATTAGAATCTCGTGACCGTGACCCTATT  
TACGTTGAACCACACGTAGATACTTACATTGAAGTTGAAGCTGGTGGTCTTTATCTTGGT  
GGGAATTATTAGTATTAGTTTCTTTAGGTTTATTAGCTATTTCGTAAATCACCACAATGT TAC  
CCATACGATGTTCCAGATTATGCT TAATAA GCATGC

#### (b) Translated sequence

MSINKITDNLTTSAEHSNHHDDNNQVTAGKNMSSRSLLYVAVALGLFTCVAAQAETQAALTQ  
ELKDEEGLSLERSQFSVSLESSSTTKGTATPSGAESTALTTGLQYNELASELKQLHQGASLSRL



VSQKKPASVDSVNSVASMNKGDAAGSNLATKASVLSDPVKASPANLRTVNTLDQGLLGMT  
REQMAIKSAENSASVQALQSSGLFHSFNIDANTHLLLEDFDGSFYSTFSVTFDADVDGIGY  
NEYADVYAELYVSQEGGPV<sup>W</sup>LHYSTEVFSIAGNSSYDDYRVLTTLQSGYQTAHYDVLIDLIE  
VGVSNPVATLSSNDTNALYALPLESRDRDPIYVEPHVDTYIEVEAGGALSWWELLVLVSLGLL  
AIRKSPQC YPYDVPDYA

## 5. Plasmid pPsa\*

*PsaA* exon 3 (forwards orientation) is highlighted in turquoise. The TGG→TGA codon alteration is shown in green.

CCCATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTA  
TAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCCT  
GGCGTTACCCAACCTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCG  
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGAAT  
TGTAAGCGTTAATATTTTGTAAAATTCGCGTTAAATTTTTGTAAATCAGCTCATTTTTTA  
ACCAATAGGCCGAAATCGGCAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGG  
TTGAGTGTGTTCCAGTTTGGAAACAAGAGTCCACTATTAAGAAGCTGGACTCCAACGTCA  
AAGGGCGAAAAACCGTCTATCAGGGCGATGGCCACTACGTGAACCATCACCTAATCAA  
GTTTTTTGGGGTCGAGGTGCCGTAAGCACTAAATCGGAACCTAAAGGGAGCCCCCGAT  
TTAGAGCTTGACGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAGAAAGCGAA  
AGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACAC  
CCGCCGCGCTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGC  
GCGGAACCCCTATTTGTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACA  
ATAACCCTGATAAATGCTTCAATAATATTGAAAAGGAAGAGTATGAGTATTCAACATTT  
CCGTGTCGCCCTTATCCCTTTTTGCGGCATTTTGCCTTCTGTTTTTGTCTACCCAGAAA  
CGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAAC  
TGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGAT  
GAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGCAAGAG  
CAACTCGGTGCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTACCCAGTCACAG  
AAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGA  
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCG  
CTTTTTTGCACAACATGGGGGATCATGTAACCTGCCTTGATCGTTGGGAACCGGAGCTGAA  
TGAAGCCATACCAAACGACGAGCGTGACACCAGATGCCTGTAGCAATGGCAACAACGTT  
GCGCAAATTAACCTGGCGAACTACTTACTTAGCTTCCCGGCAACAATTAATAGACTGG  
ATGGAGCGGATAAAGTTGCAGGACCCTTCTGCGCTCGGCCCTTCCGGCTGGTGGTTA  
TTGCTGATAAATCTGGAGCCGGTGAGCGTGGCTCTCGCGGTATCATTGCAGCACTGGGG  
CAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGCGGGGAGTCAGGCAACTATGG  
ATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAGTGT  
CAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACCTCATTTTTAATTTAAAAG  
GATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCG  
TTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTC  
TGCGCGTAATCTGCTGCTTGCAAACAAAAAACCCGCTACCAGCGGTGGTTTTGTTTGGC  
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTACAGCAGAGCGCAGATACC  
AAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCG  
CCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGT  
GTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAAGGCGCAGCGGTCGGGCTGAA  
CGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACC  
TACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT  
CCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGC  
CTGGTATCTTTATAGTCCTGTGCGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT  
GCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC  
TGGCCTTTTGTGCTGGCCTTTTGTCTACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGAT  
AACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGGAGCGC  
AGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGC  
GCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAG  
TGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTT  
ATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAAC  
AGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTG  
GAGCTCCACCGCGGTGGCGGCCGCTCTAGGTATATACATTACCCCTTAAGGCTACCCGGC

AGTTAGTTACGGCTTACGTTCCATAAAAATATTGGCATATTTTATAAATTATTTTATAGATCA  
TATATTTTGTAAATATATAATATTACGCATAACACATTTATTTAAAAACATATAATATTAC  
GCATAACACATTTATTTAAAAACAGCAAAAACCTTGGCTCAAATCCCTATAGGATATTTTTT  
ATTTATGGCAGCAGCAGGTTTACACGCTCCGTCAGGACGCCGGCACGTAGTTGGAAAGTA  
TGCCCCCTTGCCCGGAAGGGGAAAGGAGGAGACAAATTTATTTATTGTATATAAATAGC  
ATGACTTTCCAAGCGAGTTAACATAAAACAACTGCGCGAGTTAACATAAAACAACTTCCT  
CTCTGGGGAGGCATGGCAGCAAATGGCACCCTAAATACATGCGCTTTTAAACACAGATAAT  
TATAAACAAGCATAGCGTTAAATCTGCTACCTTGGATTAAATCATTAGAAAAGAAATTTGAG  
CCGTGTGCAGTGAAAATTGCATGCACGGCTCTTAAGGTTTAAATAAATTTTTAAAGAAGA  
AAATTTAACTCCTAACTATGTACTTCCATGGTGCACGTTTTTCAAACCTATGAAGCTTGGTT  
AAGTGACCCTACTCACATTAACCAAGTGCTCAAGTAGTATGGCCTATTGTAGGTCAAGA  
AATTTTAAACGGTGATGTAGGTGGTGGTTCCAAGGTATTCAAATTACTTCTGGTTCTTCC  
AATTATGGCGTGCTAGTGGTATTACTAGTGAATTACAACCTTATACTACAGCAATTTGGTGG  
TTAGTAATGGCTGCTGCAATGTTCTTTGCTGGTTGGTTCCACTACCACAAGCTGCTCCA  
AAACTAGAATGGTTCCAAAACGTTGAATCAATGTTAAACCACCCTTAGTGGTCTTCTTG  
GTTTAGGTAGTTTAGCTTGGGCTGGTCACCAAATTCACGTTTCTTACCAGTAAACAAATT  
ATTAGATGCTGGTGTAGATCCAAAAGAAATTCCACTTCCTCATGATTTATTATTAATCGT  
GCTATTATGGCTGACTTATACCCAAGTTTTGCTAAAGGTATTGCTCCTTTCTTACTTTAAA  
CTGGAGTGAATACAGTGATTTCTTAACATTTAAAGGTGGTTTAAACCCTGTTACTGGTGGT  
CTTTGGTTAAGTGATACTGCTCACCACCACGTAGCTATTGCTGTATTATTCTTAGTAGCTGG  
TCACATGTATCGTACTAACTGGGGTATTGGTCACAGTATGAAAGAAATTTAGAAGCTCAC  
CGTGGTCCATTTACAGGTGAAGGTCACGTTGGTTTATATGAAATTTTAAACACTTCTTGGC  
ATGCACAATTAGCTATTAACCTTAGCTTTATTTGGTTCGTTATCAATTATTGTAGCTCACCAC  
ATGTACGCAATGCCTCCATACCCTTATTTAGCTACTGATTACGGTACACAATTATCATTATT  
TACACACCACACATGGATTGGTGGTTTCTGTATTGTTGGTGTGGTGTGCTCAGCAGCTATT  
TTCATGGTTCGTGACTACGATCCTACTAATAACTACAACAATTATTAGACCGGTGAATTC  
GTCACCGTGATGCTATTATTTCTCACTTAAACTGGGTTTGTATTTTCTTAGGTTTCCACAGC  
TTTGGTTTATACATCCACAACGATACAATGAGTGCTTTAGGTGCTCCTCAAGACATGTTCT  
CAGATACTGCTATCCAACCTCAACCAGTATTTGCTCAATGGATTCAAAAATACACACTTCTT  
AGCTCCACAATTAACAGCACCAATGCTTTAGCTGCTACAAGTTTAACTTGGGGTGGTGA  
GCTAGGTGCTCATGGCGGTAAAGTAGCTATGATGCCTATTTCTTTAGGTACTTCTGACTTT  
ATGGTTCACCACATTCACGCTTTCACAATTCACGTAACCTGTGTTAATTCTTCTGAAAGGTG  
TTTTATTTGCTCGTAGCTCTCGTCTTATCCAGATAAAGCTAACTTAGGTTTCCGTTCCCT  
TGTGACGGTCTGGTCTGGCGGTACTTGTACGTTTCTGCTGGGACCACGTATCTTAG  
GCTTTTTCTGGATGTACAACAGCTTATCAATTGTAATTTTCCACTTCAGCTGGAAGATGCA  
ATCTGATGTTTGGGTACGGTTACAGCTTCTGGTGTCTCACATTACTGGTGGTCAACTTTG  
CACAAGCGTAAACACAATCAACGGTTGGTTACGTGACTTCTTATGGGCACAATCATCAC  
AAGTAATCCAATCATACGGTTCAGCTCTATCTGCTTATGGTTAATTTTCTTAGGTGCTCAC  
TTCGTATGGGCATTCTCGTTAATGTTCTTATTCTCTGGTCTGGTTACTGACAAGAACTTAT  
CGAATCAATTGTATGGGCTCACAACAACTTAAAGTTGCACCTGCAATTCACCACGTGCT  
TTAAGTATTACTCAAGGTCGTGCTGTTGGTGTAGCTCACTACCTTTTAGGTGGTATTGCTAC  
TACATGGTCTGTTCTTCTTAGCACGTATCATTCTGTAGGTTAAACATTTAATACTTTTTAATA  
CATATATGCCTAAGTTTATCTTTAAAGATAAACTTAGCCATATGTGTTAAGTTATCTAACA  
AGGTTACCTTTTTATTTCTCTTTAGATATATAAACATTTAAAACTACCGTGATCGTTACT  
TTAGATAACTGGAAGGGGGAAAAATCATGTATTCGCTGGAAGGCGCACCTCCTACTGCCT  
ACTGCGCAGCATTAAAAATGCTGTAGATATTGGTATCTTACAAAGGACAGTAGTACACAAT  
TAAACGCGTTGGAATATTTATATTTTCATTAGAGAATTGGGTACCGAGCTCCACCGCGGTG  
GCGGCCGCTCTAGCTAGAAGTGGATCGCACTCTACCGATTGAGTTACATCCGCTTTAG  
TATGTTACTATTTCTTTTATTATAACTTATAAAAATATAATACATAAAGATAAATTCTATAAT  
AAAAAGCTAAGATTTTATTTTCTGGCACATCGTAATTTATAAAGACAGGCAAATTTAAAC  
AAAAGATAACTTTAGAACTTAATTTTAAAAATGTAAAATGATGTTTAGGTATTTAACCTAA  
ACACCATAAAAATAAAAACGATGTTTATGCTATTACATAAAACATCATGAAAAATAAAAA  
TTAAAGTTTGTCAATAGTATCAAATTCGAATTTAATTTCTTTCCAACTTCACATGCAGCA  
GCAAGTTCTGGAGACCATTTACAAGCTGAACGAATTACGTGCGCACCTTCACGAGCAAGG  
TCACGACCTTCGTTACGAGCTTGAGTACAAGCTTGCATGCCTGCAGTTATTTGCCAACTAC  
CTTAGTGATCTCGCCTTTCACGTAGTGGACAAATCTTCCAACTGATCTGCGCGGAGGCC  
AAGCGATCTTCTTCTGTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACT  
GGGCCGGCAGGCGCTCCATTGCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGG  
TTACTGCGCTGTACCAAATGCCGGACAACGTAAGCACTACATTTGCTCATCGCCAGCCC  
AGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGATCCTGTT

AGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAGGCAACGCTATGTTCTCT  
TGCTTTTGTGTCAGCAAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCA  
AGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACG  
GAATGATGTCGTCGTCGACACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCC  
AGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCAAG  
CCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGGCTTCAGGCCGCCATCCACT  
GCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTTCGAGATGGCGCTCGATGACGCCA  
ACTACCTCTGATAGTTGAGTTGATACTTCGGCGATAACCGCTTCACGAGCCATGGACATTT  
TCACTTCTGGAGTGTATTGTTCAATTAATCTTTAATAAGATTACTAAGTTCTTCTGGAGTA  
CGCATTGCCATAAAAAAGAAAAATAAATAAAAAGATTAAAAAGTTTATTTTTAAATCT  
TTCTCGAGAATTTTAAATAAGTTTAAAATTCAACAAAAATAGTGAGTGGTAAGATCACTTG  
TTAACAAAAGTAATGGTTCACCCTTGTCAATTTAAATACTAAAATTCATTTGCCCGAAGA  
GGACAAATTTATTTATTGCATTAATAATCCCTAAGTTTACTTGCCCGTAAGGGGAAGGGGG  
GGACGTCCACAGGCGTCGTAAGCAACTAAAGTTTATGACGCCGATTGCTTTGTTAGGAAA  
ATATAAATATCCCATAAGAAAAGGTCCTTTAAAGTTTTATGGACTAAATAAAAAAGATA  
GCATAAGCATTAAAATCATGCAAATTAAAAAAAGGTAAATGTATTTATAAAAAGGTAAA  
TGTATTTATATAGTATTTATATTATAGCATAATAATAAATATATTTATAAATGATTGTTCT  
TAGAGCTAAAAGAGAAGAACAATGGGTTTATAGGTATTTTGGAGATCCAGTTATAAAAATG  
ACTTTTGACGTTTATGGTATATAAACACTGCCTCTAATAAAGTCATCGATAAGCTTGATCC  
CATATAGCCAATGGCTTAAGGAGTGTATAGGAATAACTAGTCATGCACATTTTCCTAAA  
AATCTAAAATGTTATTAGAAGCTATACAAAATTAAAAATTATGTTATTATATATTTATTA  
AAATTTAGACTTATCACGGGGATATGGCGGAATGGTAGACGCTACGGACTTAAAATCCGT  
TCTTGTGCGAACAAGGTGAGGGTTCAAGTCCCTCTTTCCCATTTATAAATAAAAAATCAAG  
TCAATATTTTAAACATTGTAACTAAAAATAAGTGTAAAATAACTAGGGTAAATAGTTAA  
GAATTGGTAGTTTTTGAACATAACAGTGACATAAAACTGGTTAAAAGACATCCCTGTAAG  
AGAAATGCATATGGTGAATTACACAATAAATTTAAAAATAAAGCTCTGACAAGCCTCTTC  
CCCTTTCGATATACATGCTCTGTTAGATATAATTCTGCCTAAGTTCCTAACAAAGGGTTT  
ACTTTTTCTCGGGGAATGGTTTTGCTACTTTGTTTTAAATCTTTTTCAAGAATGTCAATCT  
CCTATAAAGGGAGGAACATTACTCATGGTCTCACAATTTAATAAACTAGTTTCTTATAAT  
CTATTTTATAAAAATAAATCTAAATATATTTCTTATGAATTTTGTTAATTTAGAACAAATG  
AAAATCTTTACGTAATGCTACTTTTTGCATGCTTTTTTTAACAACATTTTTATATTGGTTTT  
ATACTGCATTTTACAGTACAAATCCTCAACAAATTATAAATCCATTGTCATTAAACAACAT  
TAAAATAATTTTACTTATCCTGACGGGTTTCAAGCAAACAATATTAATGTTTCTACAAC  
TATTTACCGATTAACCCTGTATTAATACCGAAAGAGAAGAGCAACCCGAAGCTAACGGA  
ACGAATGGGCTGTTGGGTGTTTCTTATTAGTAGTAAATTTAAAATCAATAGCAATCCCAC  
GCATTATGATGGGTGTTTCTAATTTATTATTAGTATTATTACTTGTTCGTTGGGAAAAA  
TCAGGTCATTTCCCATTAAGTAATTTATATGAATCTTTAATGTTTTTAGCGTGGTGTGTAC  
ATTTTATATTTATTATTGTACAAGTTTTACTTTATTGGTTGAAAAAATGTTAGGTTTAT  
TAATTCACCTTGAGTTTATTAATGAATGCATTTGCTACTTTTATGTTTACAAAAGAAAT  
GCAACAAGCATCACCATAGTACCAGCTTTACAATCAAATTGGTTAATGATGCATGTTACT  
GTAATGATTATTAGTTATGCCACATTAATTATTGGATCGTTATTGTTCGATTTTTATTTTTGAT  
TTTTATTTAAACACAAAAAAGGTACACCCAAAAAGTATGATAACTTTATTAACAATTTAGAT  
GCATTAAGTTATCGCATTATTGGATTAGGTTTTCTTTTTTAACTATTGGGATTTTATCTGG  
GGCTGTGTGGGCTAATGAAGCATGGGGATCATATTGGAGTTGGGATCC

## 6. Plasmid pWUCA1

The *trnW<sub>UCA</sub>* region is colour coded as in section 2 above. *PsaA* exon 3 (forwards orientation) is highlighted in turquoise. The *aadA* spectinomycin resistance gene (reverse orientation) is highlighted in pink and its promoter/UTRs in green.

CCCATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATTCGCCCTA  
TAGTGAGTCGTATTACAATTCAGTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCT  
GGCGTTACCCAACCTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCG  
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGAAT  
TGTAAGCGTTAATTTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTA  
ACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAAGAATAGACCGAGATAGGG  
TTGAGTGTGTTCCAGTTTGGAAACAAGAGTCCACTATTAAGAACGTGGACTCCAACGTCA  
AAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAA

GTTTTTTGGGGTCGAGGTGCCGTAAGCACTAAATCGGAACCCTAAAGGGAGCCCCGAT  
TTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAA  
AGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCACCACAC  
CCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGC  
GCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACA  
ATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTT  
CCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTITGCCTTCTGTTTTTGCTCACCCAGAAA  
CGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAAC  
TGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAAGTTTTCCAATGAT  
GAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAG  
CAACTCGGTGCGCCGATACACTATTCTCAGAATGACTTGGTTGAGTACTACCAGTACAG  
AAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGTGCCATAACCATGA  
GTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGTAAACCG  
CTTTTTTGCACAACTATGGGGGATCATGTAACCTCGCTTGATCGTTGGGAACCGGAGCTGAA  
TGAAGCCATACCAACGACGAGCGTACACCACGATGCCCTGTAGCAATGGCAACAACGTT  
GCGCAAACTATTAAGTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGG  
ATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGTTTA  
TTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGC  
CAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTACAGGCAACTATGG  
ATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGT  
CAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAG  
GATCTAGGTGAAGATCCTTTTTGATAATCTCATGACAAAATCCCTAACGTGAGTTTTCG  
TTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTC  
TGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGTACCAGCGGTGGTTTTGTTGCC  
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCCAGCAGAGCGCAGATAACC  
AAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCG  
CCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCTGT  
GTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAAGGCGCAGCGGTCCGGCTGAA  
CGGGGGGTTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATAACC  
TACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTAT  
CCGGTAAGCGGCAGGGTTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGC  
CTGGTATCTTTATAGTCCTGTCGGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT  
GCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCC  
TGGCCTTTTTGCTGGCCTTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGAT  
AACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGACGCCGAACGACCGAGCGC  
AGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGC  
CGTGTGGCCGATTCAATTAATGCAGCTGGCAGCAGAGGTTTTCCCGACTGGAAAGCGGCGAG  
TGAGCGCAACGCAATTAATGTGAGTTAGTCTACTCATTAGGCACCCAGGCTTTACTCTT  
ATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAAC  
AGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTG  
GAGCTCCACCGCGGTGGCGGCCGCTCTAGGTATATACATTACCCCTTAAAGGCTACCCGGC  
AGTTAGTTACGGCTTACGTTCCATAAAATATTGGCATATTTTTATAAATTATTTTTATAGATCA  
TATATTTTGTAATATATAATATTACGCATAACACATTTATTTAAAAACATATAATATTAC  
GCATAACACATTTATTTAAAAACAGCAAAAACCTTGGCTCAAATCCCTATAGGATATTTTTT  
ATTTATGGCAGGAGCAGGTTTACACGCTCCGTCAGGACGCCGGCACGTAGTTGGAAAGTA  
TGTCCCCCTTGGCCGGAAGGGGAAAGGAGGAGACAAATTTATTTATTGTATATAAATAGC  
ATGACTTTCCAAGCGAGTTAACATAAAACAACTGCGCGAGTTAACATAAAACAACTTCTCT  
CTCTGGGAGGCATGGCAGCAAATGGCACCCCTAAATACATGCGCTTTTAAACACAGATAAT  
TATAACAAGCATAGCGTTAAATCTGCTACCTTGGATTAAATCATTAGAAAGAATTTGAG  
CCGTGTGAGTAAAATTGCATGCACGGCTCTTAAGGTTTAAATAAATTTTTAAAGAAGA  
AAATTTAACTCCTAACTATGTACTTCCATGGTGCACGTTTTTCAAACCTATGAAGCTTGGTT  
AAGTGACCCTACTCACATTAACCAAGTGCTCAAGTAGTATGGCCTATTGTAGGTCAAGA  
AATTTTAAACGGTGATGTAGGTGGTGGTTTTCCAAGGTATTCAAATTAATCTGGTTTCTTCC  
AATTATGGCGTGCTAGTGGTATTACTAGTGAATTACAACCTTATACTACAGCAATTGGTGG  
TTAGTAATGGCTGCTGCAATGTTCTTTGCTGGTTGGTTCCACTACCACAAAGCTGCTCCA  
AACTAGAATGGTTCCAAAACGTTGAATCAATGTTAAACCACCACTTAGGTGGTCTTCTTG  
GTTAGGTAGTTAGCTTGGGCTGGTCACCAAAATTCACGTTTCTTTACCAGTAAACAAAT  
ATTAGATGCTGGTGTAGATCCAAAAGAAATTCACCTTCTCATGATTTATTATTAATTCGT  
GCTATTATGGCTGACTTATACCAAGTTTTGCTAAAGGTATTGCTCCTTTCTTACTTTAAA  
CTGGAGTGAATACAGTGATTTCTTAACATTTAAAGGTGGTTTAAACCCTGTTACTGGTGGT

CTTTGGTTAAGTGATACTGCTCACCACCACGTAGCTATTGCTGTATTATTCTTAGTAGCTGG  
TCACATGTATCGTACTAACTGGGGTATTGGTCACAGTATGAAAGAAATTTTAGAAGCTCAC  
CGTGGTCCATTTACAGGTGAAGGTCACGTTGGTTTATATGAAATTTTAACTTCTTGGC  
ATGCACAATTAGCTATTAAGCTTTATTTGGTTTCGTTATCAATTATTGTAGCTCACCAC  
ATGTACGCAATGCCTCCATACCCTTATTTAGCTACTGATTACGGTACACAATTATCATTATT  
TACACACCACACATGGATTGGTGGTTTCTGTATTGTTGGTGTCTGGTGTCTCACGCAGCTATT  
TTCATGGTTTCGTGACTACGATCCTACTAATAACTACAACAATTATTAGACCGTGTAAATTC  
GTCACCGTGTAGCTATTATTTCTCACTTAACTGGGTTTGTATTTTCTTAGGTTTCCACAGC  
TTTGGTTTATACATCCACAACGATACAATGAGTGTCTTAGGTCGTCTCAAGACATGTTCT  
CAGATACTGCTATCCAACCTCAACCAGTATTTGCTCAATGGATTCAAAATACACACTTCTT  
AGCTCCACAATTAACAGCACCAATGCTTTAGCTGCTACAAGTTAACTTGGGGTGGTGA  
GCTAGGTGCTCATGGCGGTAAAGTAGCTATGATGCCTATTTCTTTAGGTACTTCTGACTTT  
ATGGTTCACCACATTCACGCTTTCACAATTCACGTAAGTGTGTTAATTCTTCTGAAAGGTG  
TTTTATTGCTCGTAGCTCTCGTCTTATCCAGATAAAGCTAACTTAGGTTTCCGTTTCCCT  
TGTGACGGTCTGGTGTGGCGGTACTTGTCAAGTTTCTGCTGGGACCACGTATTCTTAG  
GTCTTTTCTGGATGTACAACAGCTTATCAATTGTAATTTTCCACTTCAGCTGGAAGATGCA  
ATCTGATGTTTGGGGTACGGTTACAGCTTCTGGTGGTTTCTCACATTACTGGTGGTAACTTGC  
CACAAAGCGCTAACACAATCAACGGTTGGTTACGTGACTTCTTATGGGCACAATCATCAC  
AAGTAATCCAATCATAACGGTTCAGCTCTATCTGCTTATGGTTTAAATTTTCTTAGGTGCTCAC  
TTCGATGGGCATTCTCGTTAATGTTCTTATTCTCTGGTGTGGTACTGGCAAGAACTTAT  
CGAATCAATTGTATGGGCTCACAACTTAAAGTTGCACCTGCAATTCAACCACGTGCT  
TTAAGTATTACTCAAGGTCGTGCTGTTGGTGTAGCTCACTACCTTTTAGGTGGTATTGCTAC  
TACATGGTCTGTTCTTCTTAGCACGTATCATTCTGTAGGTTAACTTTAATACTTTTTAATA  
CATATATGCCTAAGTTTATCTTTAAAGATAAACTTAGCCATATGTGTTAAGTTATCTAACA  
AGGTTACCTTTTTATTTCTCTTTAGATATATAAACATTAATAAACTACCGTGATCGTTACACT  
TTAGATAACTGGAAGGGGGAAAAATCATGTATTGCTGGAAGGCGCACCTCCTACTGCCT  
ACTGCGCAGCATTAAAATGCTGTAGATATTGGTATCTTACAAAGGACAGTAGTACACAAT  
TAA**ACGCGT**taaccatgattaacaactatatcaataaaatcaattttagtgaaatactctgattgacattaaataatcatgataaaatta  
taataacaaatttacgtccttagtctagtcggtagaacgcaggtttcaaaacctgatgctggttcaattcctacagggcgtgttttctcaatgtact  
ttgtgaaaagtggctggttaaccttttaggttcggtgaacaataatggcagttaagagtcactaaagctgctgtag**ACGCGT**TGGA  
ATATTTATATTTTTCATTAGAGAATTGGGTACCGAGCTCCACCGCGGTGGCGGCCGCTCTAG  
CTAGAACTAGTGGATCGCACTCTACCGATTGAGTTACATCCGCTTTAGTATGTTACTATTT  
CTTTTATTATAACTTATAAAATATAATACATAAAGATAAAATCTATAATAAAAAGCTAAGA  
TTTTATTTTCTGGCACATCGTAATTTATAAAGACAGGCAAATTTAAACAAAAGATAACTT  
TAGAACTTAATTTTAAAAATGTAAAATGATGTTTAGGTATTTAACCTAAACACCATAAAAA  
TAAAACGATGTTTATGCTATTCACATAAACACTCATGAAAAATAAAAATTAAGTTTGTG  
AATAGTATCAAATTCGAATTTAATTTCTTTCCAACTTCACATGCAGCAGCAAGTTCTGGA  
GACATTTACAAGCTGAACGAATTACGTCGCCACCTTCACGAGCAAGGTCACGACCTTCG  
TTACGAGCTTGAGTACAAGCTTGCATGCCTGCAGTTATTTGCCAACTACCTTAGTGATCTC  
GCCTTTCACGTAGTGGACAAATTTCCAAGTATCTGCGCGCAGGCCAAGCGATCTTCT  
TCTTGTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGACTGGGCCGGCAGG  
CGCTCCATTGCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTACTGCGCTGT  
ACCAAATGCGGGACAACGTAAGCACTACATTTGCTCATCGCCAGCCAGTCGGGCGGGC  
AGTTCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGATCCTGTTACAGGAACCGGATC  
AAAGAGTTCCTCCGCCGCTGGACCTACCAAGGCAACGCTATGTTCTTCTGCTTTTGTGAGC  
AAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTG  
CGCTGCCATTCTCAAATTTGAGTTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGT  
CGTGACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAAGCCG  
AAGTTTCAAAGGTCGTTGATCAAAGCTCGCCGCTTGTTCATCAAGCCTTACGGTCAC  
CGTAACCAGCAAATCAATATCACTGTGTGGCTTACAGGCCCATCCACTGCGGAGCCGTA  
CAAATGTACGGCCAGCAACGTCGGTTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGA  
TAGTTGAGTTGATACTTCGGCGATAACCGCTTACAGGCCA**TGGACATTTTCACTTCTGGA**  
GTGTATTGTTCAATTAATCTTTAATAAGATTACTAAGTTCTTCTGGAGTACGCATTGCCAT  
AAAAAGAAAAATAAAATAAAGATTAAAAAAGTTATTTTTAAATCTTTCTCGAGAAT  
TTAAATAAGTTTAAATTCACAAAAATAGTGAGTGGTAAGATCACTTGTTAACAAAAG  
TAATGGTTCACCCTGTCAATTTAAATACTAAAAATCATTGCCCCGAAGAGGACAAATTT  
ATTTATTGCATTAATCCCTAAGTTTACTTGGCCGTAAGGGGAAGGGGGGGACGTCCAC  
AGGCGTCGTAAGCAACTAAAGTTTATGACGCCGATTGCTTTGTTAGGAAAAATAAAATAT  
CCCATAAGAAAAGGTCCTTTAAAGGTTTATGGACTAAATAAAAAAGATAGCATAAGCAT  
TAAAATCATGCAAATTAAAAAAAGGTAATGTATTATAAAAAAGGTAATGTATTATA

TAGTATTTATATTATAGCATAATAATAAATATATTTATAAATTGATTGTTCTTAGAGCTAA  
AAGAGAAGAACAATGGGTTTATAGGTATTTGAGATCCAGTTATAAAAATGACTTTTGAC  
GTTTATGGTATATAAACACTGCCTCTAATAAAGTCATCGATAAAGCTTGATCCCATATAGCC  
AATGGCTTAAGGAGTGTACATAGGAATAACTAGTCATGCACATTTTCCTAAAAATCTAAAA  
TGTTATTAGAAGCTATACAAAAATTAATAAATTATGTTATTATATATTTATTAATAAATTTAGA  
CTTATCACGGGATATGGCGGAATGGTAGACGCTACGGACTTAAAAATCCGTTCTTGTGCG  
AACAAAGGTGAGGGTTCAAGTCCCTCTTTCCCATTTATAAATAAAAAATCAAGTCAATATTT  
TAACATTGTAAACTAAAAATAAGTGTAAAATAACTAGGGTAAATAGTTAAGAATTGGTA  
GTTTTTGAACATAACAGTGACATAAACTGGTTAAAAGACATCCCTGTAAGAGAAATGCA  
TATGGTGAATTACACAATAAATTTAAAAATAAAGCTCTGACAAGCCTCTTCCCCTTTGCGA  
TATACATGCTCTGTTAGATATAATTCTGCCTAAGTTCCTAACAAAGGGTTTACTTTTTCTCG  
GGGAATGGTTTTGCTACTTTGTTTTAAATCTTTTTCAAAGAATGTCAATCTCCTATAAAGG  
GAGGAACATTACTCATGGTCTCACAAATTTAATAAATAACTAGTTTCTTATAATCTATTTTATAA  
AAATAAATCTAAATATATTTCTTATGAATTTTGTAAATTTAGAACAATTTGAAAATCTTT  
ACGTAATGCTACTTTTTGTCATGCTTTTTTAAACAACATTTTTTATATTGTTTTTACTGCATT  
TTACAGTACAAATCCTCAACAAATTATAAATCCATTGTCATTAACAAACATTAAAATAAT  
TTACTTATCCTGACGGGTTTCAAGCAAACAATATTAATGTTTCTACAACCTATTTACCGAT  
TAACCCTGTATTAATAACCGAAAGAGAAAGCAACCCGAAGCTAACGGAACGAATGGGC  
TGTTGGGTGTTTTCTTCATTAGTAGTAAATTTAAAATCAATAGCAATCCCACGCATTATGAT  
GGGTGTTTCTAATTTATTATTAGTATTATTACTTGTTCGTTGGGAAAAATCAGGTCATT  
TCCATTAAGTAATTTATATGAATCTTTAATGTTTTTACGCTGGTGTGTACATTTTTATAT  
TTATTATATTGTACAAGTTTTACTTTATTGGTTGAAAAAATGTTAGGTTTCAATTAATTGCACC  
TTGTAGTTTATTAATGAATGCATTTGCTACTTTTAGTTTACCAAAGAAATGCAACAAGCA  
TCACCATTAGTACCAGCTTACAATCAAATTTGGTTAATGATGCATGTTACTGTAATGATTA  
TTAGTTATGCCACATTAATTATTGGATCGTTATTGTGCTGTTTTATTTTTGATTTTTATTTAAAC  
ACAAAAAAGGTACACCCAAAAAGTATGATAACTTTATTAACAATTTAGATGCATTAAGTT  
ATCGCATTATTGGATTAGGTTTTCTTTTTTAACTATTGGGATTTTATCTGGGGCTGTGTGG  
GCTAATGAAGCATGGGGATCATATTGGAGTTGGGATCC

## 7. Plasmid pWUCA2

The *trnW<sub>UCA</sub>* region is colour coded as in section 2 above. The *psaA* exon 1 promoter and 5' UTR are highlighted in pink and the SapI and SphI sites for inserting a gene of interest are highlighted in yellow. The *psbH* gene (reverse orientation) is highlighted in turquoise.

AGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACCTTA  
ATCGCCTTGCAGCACATCCCCCTTTGCGCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTAC  
GCATCTGTGCGGTATTTTACACCCGCATATGGTGCACACTCTCAGTACAATCTGCTGTGATGCCGCA  
TAGTTAAGCCAGCCCCGACACCCGCAACACCCGCTACGCGCCCTGACGGGCTGTCTGTCTCC  
CGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGAGCTGCATGTGTCAGAGGTTTTTACC  
GTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTC  
ATGATAATAATGGTTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTA  
TTTTGTTTTTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG  
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCT  
TTTTTGCGGCATTTTGCCTTCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAAGATGCT  
GAAGATCAGTTGGGTGCACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAGATCCTT  
GAGAGTTTTCGCCCCGAAGAACGTTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCG  
CGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAA  
TGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGA  
ATTATGCAGTGCTGCCATAACCATGAGTGATAAACAATGCGGCCAACTTACTTCTGACAACGATC  
GGAGGACCGAAGGAGCTAACCCTTTTTTGCACAACATGGGGGATCATGTAACCTGCCTTGAT  
CGTTGGGAACCGGAGCTGAATGAAGCCATAACAAACGACGAGCGTGACACCACGATGCCTGTA  
GCAATGGCAACAACGTTGCGCAAATTAATACTGGCGAACTACTTACTCTAGCTTCCCGGCAAC  
AATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGG  
CTGGCTGGTTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGC  
ACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAAC  
TATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACT  
GTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACCTTATTTTTAATTTAAAAGGA

TCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCAC  
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAA  
TCTGCTGCTTGCAAACAAAAAACACCGCTACCAGCGGTGGTTTGTGGCCGATCAAGAGCT  
ACCAACTCTTTTTCCGAAGGTAAGTGGCTTACAGCAGAGCGCAGATACCAAATACTGTCCTTCTA  
GTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGC  
TAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAG  
ACGATAGTTACCGGATAAGGCGCAGCGGTCCGGCTGAACGGGGGGTTCGTGCACACAGCCCAG  
CTTGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAAGCGCCAC  
GCTTCCCGAAGGGAGAAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC  
GCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTATAGTCCTGTCCGGTTTCGCCACCT  
CTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAG  
CAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGTGGCCTTTTGTCTCACATGTTCTTTCCTGCGTT  
ATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGC  
CGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCAATTCGAATCCGCGTTTTCTCCGTG  
AAAGGGAGGTGTCTTAGGCCTTAGACGATGGGGGCTTTTTGTTATATTTTACTAAATATATAT  
TATAATTAATAAATAATTGAATTGTCAATTTTTAATGTACACTTAGTTGAAAGTGCCCCTGTCCC  
CTTGCCATATTTAACAGAAGTTATTTATAACGCAGCTGTTTTTTGGAGTCTATAAATTTATAAC  
ATCAGTTACTATGGATTTCCCTTTAGTTTTATGGCCTAGGACGTCCCCTTCCCCTTCGATGCTGG  
AGGCATCCTTTTACGGGACAATAAATAAATTTGTTGCCCTCGCCTATCGGCTAACAAAGTTCCTTC  
GGAGTATATAAATATAGGATGTTAATACTGCTATAAACTTTAGTTGCCAATATTTATATTAGG  
ACGCCAGTGGCAGTGGTACCGCCACTGCCTGCTTCGCAGTATATAAATATAGGCAGTTGGCAG  
GCAACTGCCACTGACGTCCTATTTAATACTCCAAGTTTACTTGCCTAGGCAGTTGGCAGGCA  
ACAAATTTATTTATTGTCCACTAAAATTTATTTGCCGAAGGGGACGTCCACTAAAATTTATTT  
ACCCGAAGGGGACGTCCTAATAAATATGGGGATGTCAATGCTCCGTTAGGAAGTAACTAAC  
GTTTTTCAAATAAATTTATCCCGGAGGGAAGTAGGCAGTAGCCCGCCACTGTCATCCTTTAAG  
TGGATCTCTCGTCAGGCAATTTGCTTACACCTTTAAATTAATAAATTAAGAAAAGTGA  
GCTATTAACCGGTaaccatgattaacaactatatcaataaaatcaattttagtagaaataactctgattgacattaaataatcatgataaaatfa  
taataacaaatttacgtccttagttcagtcggttagaacgcaggtttcaaaaactgatgctggttcaattcctacagggcgtgttttctcaatgactttgtg  
aaaagtggctgtttaaccttttagtttcgattgaacaataatggcagttaaagatcactaaagctgctgtagACCGGTAAGCTTTCTTAA  
TTCAACATTTTTAAGTAAATACTGTTAATGTTATACTTTTACGAATACACATATGGTAAAAAA  
TAAAACAATATCTTTAAAATAAGTAAAAATAATTTGTAAACCAATAAAAAATATATTTATGGT  
ATAATATAACATATGATGTAAAAAAACTATTTGTCTAATTTAATAACCATGCATTTTTTATGA  
ACACATAAATAAAGCGTTGCTAATGGTGTAAATAATGTATTTATTAATTAATAATTTGT  
TATTATAAGGAGAAATCCATGGGAAGAGCTACTCCATGGATCCTCTAGAGTCGACCTGCAGGC  
ATGCAAGCTTGTACTCAAGCTCGTAACGAAGGTCTGACCTTGCTCGTGAAGGTGGCGACGTA  
ATTCGTTACGCTTGTAAATGGTCTCCAGAACTTGTGCTGCATGTGAAGTTTGGAAAGAAATTA  
AATTCGAATTTGATACTATTGACAAACTTAAATTTTTATTTTTCATGATGTTTTATGTGAATAGCA  
TAAACATCGTTTTTTATTTTTATGGTGTAAAGTTAAATACCTAAACATCATTTTACATTTTTTAAA  
ATTAAGTTCTAAAGTTATCTTTTGTAAATTTGCCTGTCTTTATAAATTACGATGTGCCAGAAA  
AATAAAATCTTAGCTTTTTATTATAGAATTTATCTTTATGTATTATATTTTATAAGTTATAATAA  
AAGAAATAGTAACATACTAAAGCGGATGTAGCGCGTTTATCTTAACGGAAGGCCAGTGGCAGT  
GGCGGTGCCACTGCCGAATATAAATATGGTTGAGTTGCTTAGTTTACCTTAGCGAAAAGAAGA  
CTTAGCAGCTAGCCTTAACAAACAGTTTTTATATTTTATGTTTGTGTTAAATAAAAATTAAGAAAC  
TTAGCTAAAGTTTCCCAACTCATAGAAACGTCATCTAAAATTAAGAAGTGTGTAATTTCT  
AAAATGATTAATAAGAATGCTGCAAATAAAAGGATAAATAACAGCCATTAACAGTTGTACCC  
CAGCCTGGTAATACTTTACCTGCTTCTGAGTTAAGTGGACGTAATAAAGTACCTAATGGTGTAA  
CTAAACCAGGTTCTTGGAAAGTCTGAATTTACTTTTGTGTTAGCTTTAGAAGTTCCTGTTGCC  
ATAATTGATTAATGAATTAAGCGTTATTAGCGCTATTTTATTTACTTTCTGTAAAAAATAAGG  
AAAATATTCTTCAGTGCATTCCTCTCAGGATTATAAATACTCTGAGGATAACGTTCTCTCGTC  
AAGGGGTGCTTCTTGTGAGTATAGAAACCTACTAGCACAAGAAATAAATTGCATAAAAATGT  
ATTTACCTAGGACCGCAGTAGGCAGTCCCTTTTCCCCTCAGAAGTGCCTGCTTTAAAAGAATG  
AAAAAACTGCCTTGTCTGGTAAGTAAAACTTTAATTACTCACTAAAGACGATCTTAGAAGTT  
CTTTGTTCATTTTTTATTTAATATAAATATTTGTTATATAAAAAATTAATAATTTTTAATTAATGTT  
TAACCTTTGTAAGGACAGTTTCAAAGTGACATGAATGGCTACTGCAAAAACGAAGTAAGTTATT  
CTTTCTCAGGGCAAAATTTGAGTAGATTAATTTTGTTTAAAAATGTGGGACACAGTCGTCAAG  
TCTTTTGAACATCTAAGAGATATGTTGAAAAGAGAATAATTTTATTATTAATGAGCTATGGA  
AAGTCCAGCTTTTTTCTTTACCTTTTTTTTATGGTTTCTTCTGTTAAGTGTAACTGGCTATTCAGT  
TTATGTTAGTTTTGGTCCACCTTCAAAAAAATTACGTGATCCTTTTGAAGAACACGAAGATTAA  
ACAAGTTAAAAAGTACTATTTTTACAAGTGACTTCGGTGCCTCTGAGAACCCTAGTTATAGTGA  
TATAAATAACTAGCTAACTACTTTATTTTTATGAAAGTCATTTTGTGAGCATATAAACA

AAACAAAATTGCTATACTAGGCAGTCACAGTGCAACTGTCTCCGTCTCCTTAACCGAGAAAGG  
GTAAACGTCTTCGGTAAAGTAACAAACTTTAGTTATGTAACTGCTTGCGAGTTAACCATTTTTT  
TTCTCCGAAGGACAACAGTTGGCAGTTGCCAAACTTTAGTGGTCTAATATTTATATTAGGCAG  
TTGGCAGGCAACTGCACTGACGTCCCGAAGGGGAAGGGGTTTACTTACCTCCTAACGGAGTAT  
ATAAATAGAATAAAATTTATTTCTGCGCTAGCAGATTTACATACTAGGATTTAATACTCCGA  
AGGAGGCAGTGGCGGTACCACTGCCACTGGCGTCCTCCTTCCCCTTCGGGCAAATGCATTTTAG  
TGCCACTTAAGTTTACTTGCCTAGGCAGTTGGCAGGACGTCAGTGGCAGTGGTACCGCGACTGC  
CTATATTTATATACTCCTAAGTTTACTTGCCTAGGCAGTTGGCAGGCAACTGCCACTGACGTCC  
TCCCCTTCCCCTTCGGGACGTCCCCTTACGGGAATATAAATATTAGTGGATATTTATATACTGC  
GATGTTTACATACTCCGAAGGAGGAGAGCTAGCAGTTGCCTGCCAACTGCCTAATATAAATATT  
GGGCAAGTAAACTTAGAATGTTTACATACTCCGAAGGAGGACGTCCCTTACGGGAATATAAAT  
ATTAGTGGCAGTGGTACCGCCACTGCCTCCTTCGGAGTATTAATAATCCTAGTATATAATATACC  
GTAAGGGACGTCCCTCCGACGGTGGCAGTGGCGGTACCACTGCCACCGGCGTCCTAATATACAT  
ATTGAAGTATTTAAACCTGTTAGCGCACGCTCTAACGAGTCAGTAAACTTCCCTTTTGGGGCTT  
CTAGGCAGCGCATAAATTTTCTAGGACGAACGTCCACTGGCGTCTCGTAAGGAGCAGTGACAG  
GCCACTAATGTCCCCTTAATGGGTAAATAAATGGCTATCGTCTATCCATGAAGAGACCATATAT  
TCCAGTAGCACCGTTATGATCCTCAAAGGGTAACACCATTTGTATAGTATTATGGTGAAATGCA  
TCCCCTTCAGGGTAGATTTATATCTTACAG