

Supplementary Material for

Improving transgenesis efficiency and CRISPR-associated tools through codon optimization and native intron addition in *Pristionchus* nematodes

Ziduan Han, Wen-Sui Lo, James W. Lightfoot, Hanh Witte, Shuai Sun and

Ralf J. Sommer

Contents:

- Figure S1. Complete codon usage preferences in eight nematode species
- Figure S2. Complete codon usage preferences in three *Pristionchus* species
- Figure S3. Complete codon usage preferences in two *Caenorhabditis* species

Table S1. Genome annotation and transcriptome information used in this study

Table S2. A summary of codon usage bias in *Pristionchus Pacificus*

Table S3. The efficiency of GFP and TurboRFP constructs which codons were optimized based top 3% highly expressed genes

File S1. Full sequences of codon optimized and intron added constructs in *Pristionchus pacificus*

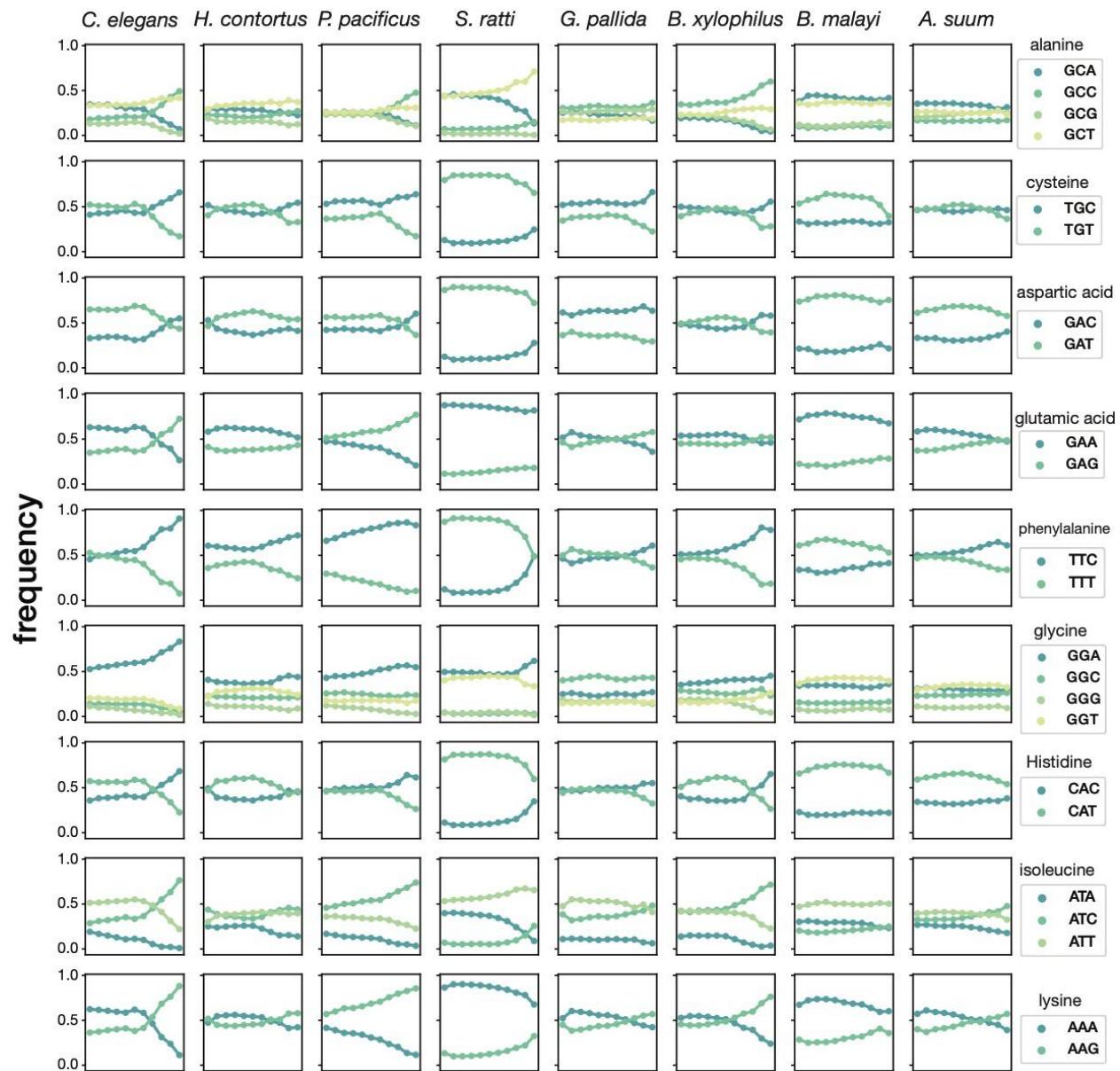


Figure S1. Complete codon usage preferences in eight nematode species as a function of gene expression levels. The codon usage bias in *Caenorhabditis elegans*, *Haemonchus contortus*, *Pristionchus pacificus*, *Strongyloides ratti*, *Globodera pallida*, *Bursaphelenchus xylophilus*, and *Brugia malayi*. The protein coding genes are binned based on the transcripts per kilobase million (TPM) values from a low to high expression level with a log2 scale into 11 bins (x axis). The dots represent the average codon usage frequency of a given bin.

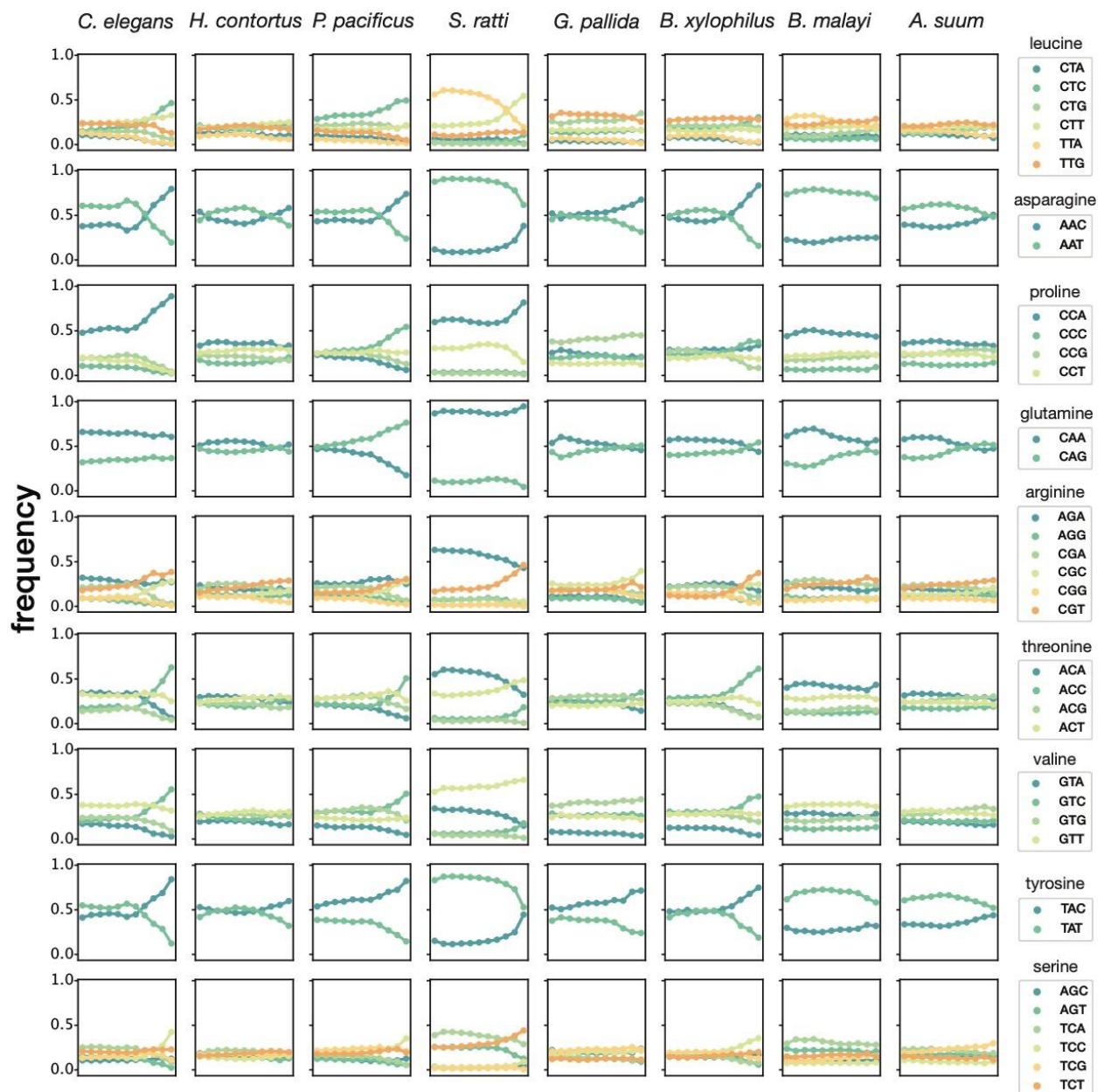


Figure S1. Continued.

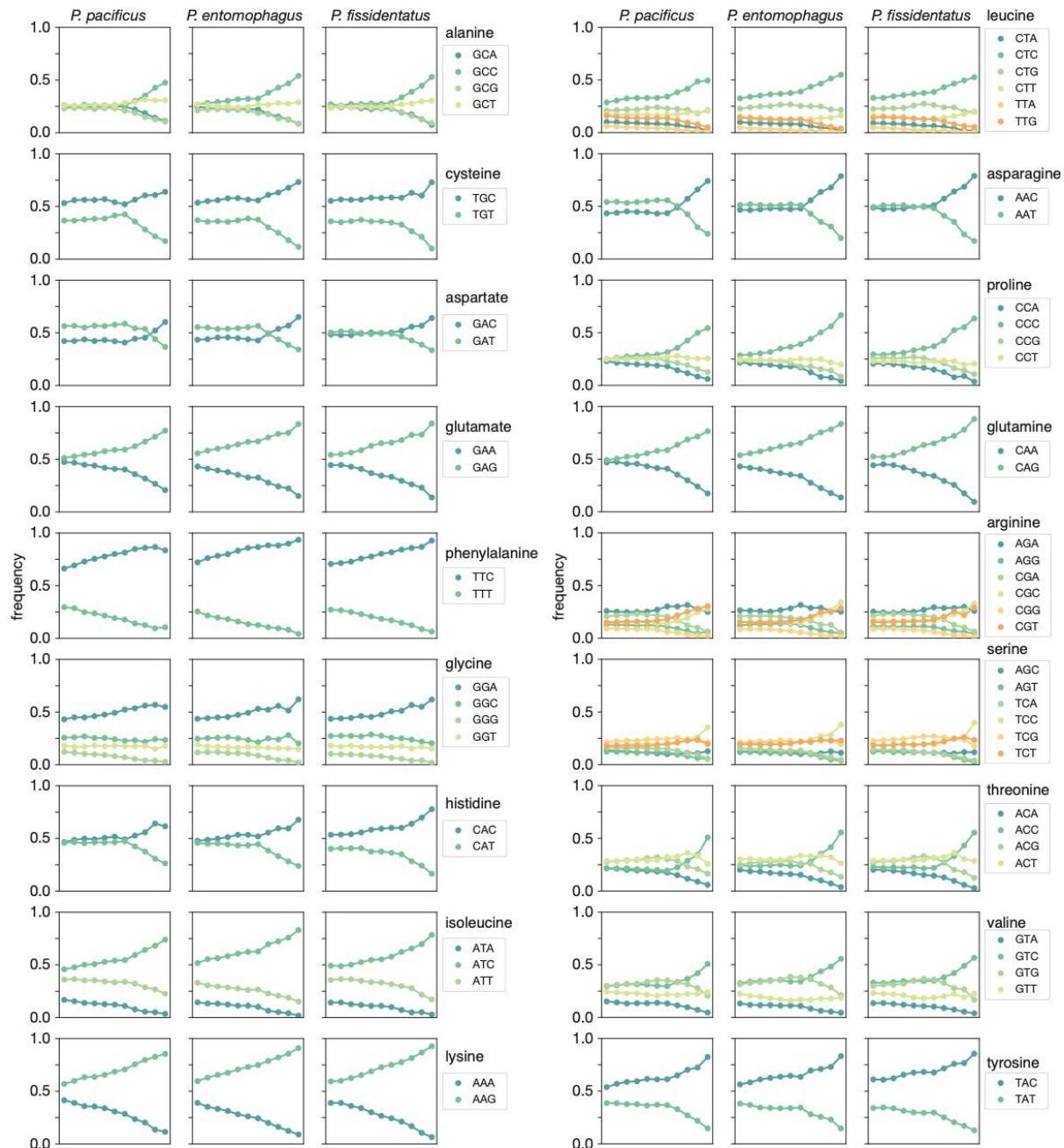


Figure S2. Complete codon usage preferences in three *Pristionchus* species as a function of gene expression levels. The protein coding genes are binned based on the transcripts per kilobase million (TPM) values from a low to high expression level with

a \log_2 scale into 11 bins (x axis). The dots represent the average codon usage frequency of a given bin.

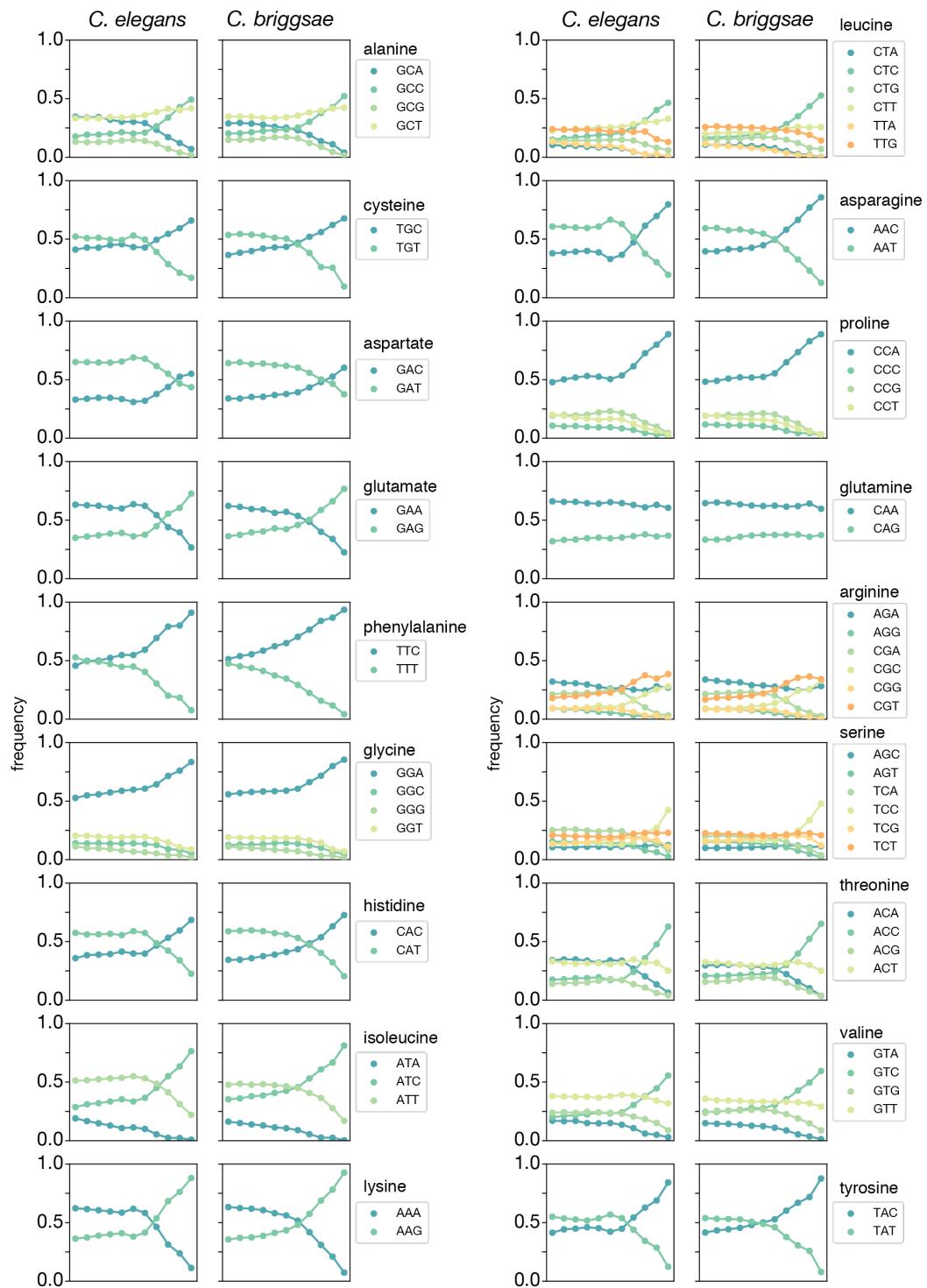


Figure S3. Complete codon usage preferences in two *Caenorhabditis* species as a function of gene expression levels. The protein coding genes are binned based on the transcripts per kilobase million (TPM) values from a low to high expression level with

a \log_2 scale into 11 bins (x axis). The dots represent the average codon usage frequency of a given bin

Species	Genome assembly & annotation version	Transcriptome Study Accession	Transcriptome stage
<i>Bursaphelenchus xylophilus</i>	WS271	PRJDB3458	mix stage*
<i>Haemonchus contortus</i>	WS271	PRJEB1360 PRJEB1376,	mix stage*
<i>Strongyloides ratti</i>	WS271	PRJEB3187	female adult
<i>Brugia malayi</i>	WS271	PRJEB2709 PRJNA1420	mix stage*
<i>Ascaris suum</i>	WS271	41	mix stage*
<i>Caenorhabditis elegans</i>	WS271	PRJNA5118 68 PRJNA2318	young adult
<i>C. briggsae</i>	WS271	38	mix stage*
<i>Globodera pallida</i>	WS271	PRJEB2896	mix stage*
<i>Pristionchus pacificus</i>	EI_Paco_annotation_v2_1	PRJEB20959	mix stage
<i>P. fissidentatus</i>	fissidentatus_annotations_prabh 2018	PRJEB20959	mix stage

* The original experiment contains samples from different stages. In our analysis, all the stages were pooled.

Table S1. Genome annotation and transcriptome information used in this study.

amino acid	amino acid abbrev.	10% highly expressed genes	3% highly expressed genes
alanine	A	GCT	GCC
cysteine	C	TGC	TGC
aspartic acid	D	GAT	GAC
glutamic acid	E	GAG	GAG
phenylalanine	F	TTC	TTC
glycine	G	GGA	GGA
histidine	H	CAT	CAC
isoleucine	I	ATC	ATC
lysine	K	AAG	AAG
leucine	L	CTC	CTC
methionine	M	ATG	ATG
asparagine	N	AAT	AAC
proline	P	CCT	CCC
glutamine	Q	CAG	CAG
arginine	R	AGA	CGT
serine	S	TCT	TCC
threonine	T	ACT	ACC
valine	V	GTG	GTC
tryptophan	W	TGG	TGG
tyrosine	Y	TAC	TAC
stop	*	TGA	TGA

Table S2. A summary of codon usage bias in *Pristionchus Pacificus* based on the top 10% highly expressed genes and top 3% highly expressed genes, respectively.

construct	No. of introns	injected P0s	No. of P0s with fluorescent + F1s	efficiency
<i>egl-20p::GFP</i>	3	55	5	9%
<i>egl-20p::TurboRFP</i>	3	63	18	29%

Table S3. The efficiency of GFP and TurboRFP constructs which codons were

optimized based top 3% highly expressed genes.

File S1. Full sequences of codon optimized and intron added constructs in
Pristionchus pacificus

>egl-20p::GFP (PZH008; codon optimization using top 10% highly expressed genes)
gacgaaaggccctcgatgcgcattttataggtaatgtcatgataataatggtttagacgtcagggtggcactttcg
ggaaatgtgcgcggAACCCtattgttattttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaa
tgcttcataataattgaaaaaggaagagtatgagtattcaacattccgtgtccccctattccctttgcggcatttgcctc
gttttgcacccagaacgcgtggaaagtaaaagatgtgaagatcagtgggtgcacgagtgggtacatgaaact
gatctcaacacagcggtaagatcctgagagttcgcggcaagaacgtttcaatgtgagcactttaaagttctgtatgt
ggcgcggattatccgtattgacgccggcaagagcaactcggcgcacactatttcagaatgacttgggtgag
actcaccagtcacagaaaagcatctacggatggcatgacagtaagagaattatgcagtgcgcataaccatgagtgat
aacactgcggccaacttacttgcacaacgatcggaggaccgaaggagctaaccgtttgcacaacatggggatca
tgtaactgcctgtatgtggaaaccggagctgaatgaagccataccaaacgacgagcgtacaccacgatgcctg
gcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagctccggcaacaattaatagactggatg
gaggcggataaaagttcaggaccactctgcgtcggccctccggctggctggttattgtgataaactggagccgg
agcgtgggtctcggtatcattgcagcactgggccagatggtaagccctccgtatcgtagttatcacacgacgg
gtcaggcaactatggatgaacgaaatagacagatcgtgagataggtgcctactgattaagcattgtactgtc
caagttactcatatatacttagattaaaacttcattttaaataaaaggatctaggtaagatcctttgataatctcatg
accaaaaatcccttaacgtgagtttgcgttccactgagcgtcagacccgtagaaaaagatcaaaggatcttgc
ttctgcgtaatctgcgttgcataacaaaaaccaccgcgtaccagcgggtggttttgcggatcaagagctac
ctctttccgaaggtaactggctcagcagagcgcagataccaaatactgttcttagtgcgttagccgttagt
tcaagaactctgttagcaccgcctacatacctgcgtctgctaattctgttaccagtggctgcgtcc
cttaccgggtggactcaagacgatagttaccggataaggcgcagcggcggctgaacgggggttcgt
ccagctggagcgaacgacactacccgaactgagatacctacagcgtgagctatgagaaagc
ggagaaaaggcggacaggatccgtaaagcggcagggtcggaaacaggagagcgcac
aacgcctggatctttagtgcgttgcggttcgcacctctgacttgcgtc
gcctatggaaaaacgcgcagcaacgcggccctttacggcccttgccttgc
cccctgattctgtggataaccgtattaccgccttgcgtgagctgatacc
gagtcgtgagcggaggaaagcggaaagagc
accccaggcttacactttatgcgtatgttgc
tgaccatgattacgccaagctgcgtc
GTGACGTGGTTCAGTGAGTCGCCATCC
ATAGTAATGAAAAGT GAGTAATTGTAATGTAATGTAATGGATCTACTGCTGTCGAATGT
TAACGTTGCTAAATATTGCGGTAAAGCGCTGGAGAAAGTTAGCGTTCGCGAG
TGGTACAATTGCTCTGCGAAGCTGATCGGTCTGTTCTGGATGACAAAACAATA
TTAACAGTGGAGGAGGTTATTCATTGTATTAAATTATGAATGCCTGTATTATTA
TTATTTATGGTTGGATATTCTGATGTAGGAGTAGGGTTACTGTTTCGAAGCTAG
TCAGCTTCGCGTTCGCGAACATTACAATTGGTGAGGGAATTAGCGGAATCCAATG
CAGTTGTCTTCGCGGTGCAAACCATGTTAATAGTGTGATGAGTTAGAGAATTAT

TAATATTATTCGTATGAATTAACTAATCTAATTACAGAACGCTGTGCAAGCAGAT
TGCAGGGCAAGCCGACGGATATGGCGAGCTGAACGAGCTGATCAATGCGAAGCA
GATCAGAAAGGTATGAGAGGGATTGACGACGCAAATGTCTCGATATTGATCTATT
TGCCGAAATTGAGCTTGCCAATGTGATTGAGTTCTCTACGATTACTAAGGATG
GTTAAATTCTCCTATATCTCGATAATTCTCAATGCGATCATTAACAAATGTTCGA
TTTCTCGTTATTCGCGTTGAGTTATTACAAATGCTGAGGGCTCATTGTTCT
GTGAAATCGAAGTTCATTGTTACATGAAGTGCTGAGCTGAAACTGATTGAT
AGTTTCTCAGGTGTTATTTAAAAATTATAAAATATTTATAATAAAAAATTTTTATT
TCAGATCTACCAGCTGACAGACGCTGTCTACAACGACGATTGCTGTCCGACAC
ATCACCTCCCCTCGTCTACAAGGACTTGCCTCCTGATCCCTCTGAGCTTG
ATCTTGATCTCAGATCTCTCCCTTTACAGTTAAACCCCCGTTCAGAGTAACCC
TAGTCCCAGACTAGCTCCCCCTTCTCCTAGAAGTTCCAATAATGCCCAATA
CCTCATAATGCACTCTCGCTGGTCAACCATTGTTACTACTCGGATGTTATC
CCTCCAGTCAGTTCTCGATAGTTCTCCTACCGTAATCCCCGAGTGTGTGTTGTC
ACTTGCTTATCAGCCCCGATACAGGTAGCTACCGTACTACTAGTGCCTTTGTG
TATTCACTCTCTCGATTCGCGTGCTATCTCGCCCATCATTGAAGGGAAACGA
AATAGTATTGTAGAGAGAGAAGGATGAATCATCCCTGACCCGCCCCCCCCCTT
CACCCATCTATTCGTATTTCAAGCTAATAAGTCATTCTGCCCTCTTATCCCTAG
TCCTCTGTTATTGATTGAGTTCTTAACCCCCCTGCTAGCTCCCTCCGCAAA
ATATTTACTGCTTTTGGCTGGAGAGTCAGGAGTTACTAATGAGATCCAGAT
TGGAGAGGAAAATAGGGACACAGGGGATCGATTCTACCCAAATTCCCTCTAT
TCCAATCCCTCATTCTACCCCGACTCTCTCCATTACACGCTATCTCACTCTGT
CTTACAGACGAGACAGCTGTGATTCACTGTTGCTCTACCGTACCCATTGTACCG
TATCTTGGCATGTATTCTGAATAAAAGGTTACTGTAATGATGTGGTGGTCAGTGC
AAGGGTCCCCTAGTGGACTAGTGAATATAGGGAGTGGTCGGAATGCGCACCATG
TCTGAATGTACACCACAGCAATTGAATCGTCTACGGCGCTCCTCGTTACA
CACTATCTCGTCGTCTGACCGTCTGCCGCACTCTCGATTCCCTGTGATCTCTAA
CCGCCCTCGAAGGGATAGGTTACCTCGATACACACACGGGACGAAATTATGG
CCCGTAAATGTAATGGAGGCCCTCGTCCCATCCCCCTATTGTCATGCCATTAC
AGAAGGGAGGGAGAAGAAGAAGGGTATTGATTGATCACTCGTGGCCGGTATTGATT
TTCATGAGAATACGAGATTGCGGTTCTCCcccgGGggAttttttttcagATGTCTAAGGG
AGAGGAGCTTCACTGGAGTGGTGCTATCCTCGTGGAGCTCGATGGAGATGTG
AATGGACATAAGTTCTCTGTGTTGGAGAGGGAGAGGGAGATGCTACTACGGAA
AGCTCACTCTCAAGTCATCTGCACTACTGTGAGCATTCTGGTTGTGAATGGGG
GTTGTAAAACCTCATGGGATTCTAACCTATTAAATTTCAGGGAAAGCTCCCTG
TGCCTTGGCCTACTCTCGTGAECTTTCTGCTACGGAGTGCAGTGCTTCTCTAGA
TACCCCTGATCATATGAAGCAGCATGATTCTCAAGTCTGCTATGCCTGAGGGATA
CGTGCAGGAGAGAACTATCTCTCGTAAGTCGTATACATTAGCGGGTGCTTTAC
GTGATATCCGGGGTTGGTTGAGAGAGGGAGATATTAAATAAAATATAATT
CAGAAGGATGATGGAAATTACAAGACTAGAGCTGAGGTGAAGTCGAGGGAGATA

CTCTCGTGAATAGAATCGAGCTCAAGGGAATCGATTCAAGGAGGATGGAAATATC
CTCGGACATAAGCTCGAGTACAATTACAATTCTCATATGTGGTGAGTGCTGTC
ATATTAAGTGACATGAAACTTTCTCAGTACATCATGGCTGATAAGCAGAAGAATG
GAATCAAGGTGAATTCAAGATCAGACATAATATCGAGGATGGATCTGAGCTC
GCTGATCATTACCAGCAGAATACTCCTATCGGAGATGGACCTGTGCTCCCTGA
TAATCATTACCTCTACTCAGTCTGCTCTAAGGATCCTAATGAGAAGAGAGA
TCATATGGTGCCTCGAGTTCGTACTGCTGCTGGAATCACTCATGGAATGGATG
AGCTCTACAAGTGAtcc gagctcgattgttattgtgattgtacgcgtatgcataaaaatgttatgcgtttaaac
cggattcactggccgtttacaacgtcgactggaaaaaccctggcgtaaccaactaatgcctgcagcacatc
ccccttcgccagctggcgtaatgcgaagaggcccgcaccgatcgccctccaaacagtgcgcagcctgaatggcg
atggcgccgtatgcggattttcccttacgcattgtcggtatttcacaccgcataatgtgcacttcgcataatgcctc
atgcgcatagttaagccagccccacacccgccaacacccgctgacgcgcctgacgggttgtctgcctccggcatc
cgcttacagacaagctgaccgtctccggagctgcattgtcagaggtttaccgtcatcaccgaaacgcgcga

>egl-20p::turborfp (PZH009; codon optimization using top 10% highly expressed genes)
gacgaaaggccctcgatgcctattttataggtaatgtcatgataataatggttcttagacgtcagggtggcactttcg
ggaaatgtgcgcgaaaccctattgttattttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaa
tgcttcaataatattgaaaaaggaagagtatgagtattcaacattccgtgtcgccctattcccttttgcggcatttgcctc
gttttgcacccagaaacgcgtgtgaaagtaaaagatgtgaagatcagttgggtgcacgagttgggtacatgcact
gatctcaacacgcgttaagatccgtggatggcattttccgttacgcgcggcaagagcaactcggtcgccgcataactatt
tcgcataatgcactttaaagttctgtatgtgggtgact
actcaccagtcacagaaaagcatttacggatggcatgacagtaagagaattatgcattgtgcgcataaccatgatgt
aacactgcggccaacttacttcgtacaacgcatttcgtggaggaccgaaggagctaaccgcattttgcacaacatggggatc
tgtaactcgccctgatcgttggaaaccggagctgaatgaagccataacgcgcgtgacaccacgatgcctgta
gcaatggcaacaacgtcgcaactattactggcactacttacttcgtccggcaacaattaatagactggatg
gaggcggataaagtgcaggaccacttcgtcgctcgccctccggctggctggatttgctgataaatctggagccgg
agcgtgggtctcgccgtatgcgcactggccagatggtaagccctccgtatcgtagttatcaca
gcacgcacgggg
gtcaggcaactatggatgaacgaatagacagatgcgtgagataggcctactgattaagcattggtaactgtc
caagttactcatatacttttagattgatattaaacttcattttaaaggatctaggtaagatccctttgataatctcatg
accaaaaatcccttaacgtgagtttcgttccactgagcgtcagacccgtagaaaagatcaaaggatcttgc
gatcccttt
tttgcgcgtatctgtctgcaaaacaaaaaccaccgcgtaccagcgggtggtttgcggatcaagagactacca
ctctttccgaaggtaactggctcagcagagcgcagataccaaactgttcttagtgcattggccacttggcc
tcaagaactgttagcaccgcctacatacctcgctcgtaatccgttaccagttggctgctgccc
acttgcgttggactcaagacgatagttaccggataaggcgcagcggcggctgaacgggggttcgtcacac
ccagctggagcgaacgcacccactacccgaactcgagataccctacagcgtgagctatgaga
aaacgcgcacgcgttccgaag
ggagaaaaggcggacaggtatccgttaagcggcagggtcggaacaggagagcgcacgaggagcttccagg
aacgcctggatcttatagtccgtcggttgcaccctcgacttgagcgtcgattttgcattgcgtc
gatgcgtcaggggggcgga
gcctatggaaaaacgcgcacgcacgcggccctttacgggtccgtggcccttgc
catacatgttcttcctgcgttat
ccctgttgcgttgcgatcaccgttaccgccttgcgttgc
gatgcgtcaggccaaacgcgcctcccccgcgttggccattatgc
gagtcagtgagcggagagcggaaagagcgcggcaatacgca
aaaccgcctcccccgcgttggccattatgc

agctggcacgacaggttcccgactggaaagcgggcagtgagcgcaacgcaattaatgttagtgcatttcactcattaggc
accccaggcttacacttatgctccggctcgatgtgtggaaattgtgagcgataacaattcacacagggaaacagcta
tgaccatgattacgccaagctgcatgcctgcagGTCGACGTCGGTGTTCAGTGAGTCGCCATCC
ATAGTAATGAAAAGTGAGTAATTGTAATGTAGAAATGGATCTACTGCTGTCGAATGT
TAACGTTGCTAAAATATTGCGGTAAAGCGCTGGAGAAAGTTAGCGTTCGCGAG
TGGTACAATTGCGTTCTGCGAAGCTGATCGGGTCTTGTGATGACAAAACAATA
TTAACAGTGGAGGAGGTTATTCAATTGTATTAATTGTAAATTGATGCCTGTATTATT
TTATTTATGGTTGGATATTCTGATGTAGGAGTAGGGTTACTGTTTCGAAGCTAG
TCAGCTTCGCGTTCGCGAACATTACAATTGGTGAGGGATTAGCGGAATCCAATG
CAGTTGTCTTCGCGGTGCAAACCATGTTAATAGTGTGATGAGTTAGAGAATTAT
TAATATTATTCTGATGAATTAATTAACTAATCTAATTACAGAACGCTGTGCAAGCAGAT
TGCAGGGCAAGCCGACGGATATGGCGAGCTGACGACGCAAATGTCTCGATATTGATCTATT
TGCGAAATTGAGCTTGCATGATTGAGTTGAGTTCTACGATTACTAAGGATG
GTTAAATTCTCCTATATCTGATAATTCCCTCAATGCGATCATTAACAAATGTTGCA
TTTCTCGTTATTCGCGTTGATTTTATTACAAATGCTGAGGGCTTCATTGTTCC
GTGAAATCGAAGTTTCATTGTTACATGAAGTGCTCAGTGGTGCAGAACTGATTGAT
AGTTTCTCAGGTGTTATTTAAAAATTATAAAATATTTATAATAAAAATTTTATTT
TCAGATCTACCAGCTGACAGACGCTGTCTACAAACGACGATTGCTGTCGAC
ATCACCTCCCCTCGTCTACAGGACTTGCCTCCTGATCCCTCTGAGCTTG
ATCTTGATCTCAGATCTCCCTTTACAGTTAAACCCCCGTTCAAGTAAACCC
TAGTCCCAGACTAGCTCCCCCTTCTCCTAGAAGTTCCAATAATGCC
CCTCATAATGCACTCTGCCTGGTCAACCATTGATTTCACTACTCGGATGTTATC
CCTCCAGTCAGTTCTCGATAGTTCTCCTACCGTAATCCCCGAGTGTGTTGTC
ACTTGCTTATCAGCCCCGATAACAGGTAGCTACCGTACTACTAGTGC
TATTCACTCTCGATTCGCGTGCTATCTCGCCCATCATTGAAGGGAAACGA
AATAGTATTGAGAGAGAGAAGGATGAATCATCCCTGACCCGCCCCCCCCCTT
CACCCATCTATTCGTATATTCAAGCTAATAAGTCATTCTGCCCTCTTTATCC
TCCTCTGTTATTGATTGAGTTCTTAACCCCCCTGCTAGCTCCCTCCGCAA
ATATTTACTGCTTTTCGGTGGCTGGAGAGTCAGGAGTTACTAATGAGATCC
TGGAGAGGAAAATAGGGACACAGGGGATCGATTGATCCAAATTCCCTCTAT
TCCAATCCCTCATTCTACCCCCGGACTCTCCATTCAACCGCTATCTCACT
CTTACAGACGAGACAGCTGTGATTCACTGTTGCTCTACCGTACCCATTG
TATCTGGCATGTATTCTGAATAAAAGTTACTGTAATGATGTGGTGGTCAGTGC
AAGGGTCCCCCTAGTGGACTAGTGAATATAGGGAGTGGTCGGAATGCG
TCTGAATGTACACCACAGCAATTGAATCGTCTATGGCGCTCCTCGTT
CACTATCTCGTCGCTCTGACCGTCTGCCGACTCTCGATTCCCTGTGATCT
CCGCCCTGCGAAGGGATAGGTTACCTCGATAACACACACGGGACGAA
CCCGTAAATGTAATGGAGGCCCTCGTCCCATCCCCATTGTCATGCC
AGAAGGGAGGGAGAAGAAGGGTATTGATTGATCACTCGTGGCGGGTATTGATT

TTCATGAGAATACGAGATTGC GGTTCTCCcccgGGggA ttttttttcagatgtctgagctcatcaa
ggagaatatgcataatgaagctctacatggagggactgtgaataatcatcattcaagtgcactctgagggagaggaa
agccttacgagggactcagactatgaagatcaagggtggagggaggacctGTGAGCATTCTGGTTG
TGAATGGGGTTGTGAAAACCTCATGGGATTCTAACCTATTAAATTTCAGGctccctt
tcgcttcgatatcctcgctacttcttcatgtacggatctaaggcttcatcaatcatactcagggaatccctgattctcaagca
gtcttccctgagggattcactggagagaatcactactacgaggatggaggaGTAAGTCGTATACATTAG
CGGGTGCCTTACGTGATATCCGGGTTGGTTTGAGAGAGGGAGATATTATTAA
AATAAAATATAATTTCAGgtgctactgctactcaggatacttccagaatggatgcatactacaatgtgaag
atacaatggagtgaattcccttaatggcacgtgtatgcagaagaactagaggatggaggctaatactgagatgct
taccctgctGTGAGTGCTGTCAAATATTAAGTGACATGAAACTTTCTCAGGatggaggact
cagaggacattctcagatggctctaagctcggtggaggagtagccattgtcttcaagactacttacagatctaag
aagcctgctaagaatctcaagatgcctggattccattcggtcatagactcgagagaatcaaggaggctataagga
gacttacgtggagcagcatgagatggctgtggctaagtactgcgatccctctaagctcggacatagataatccgagctc
gattgttattgtgattgtacgcgtatgcaataaaatgttatgcgtttaaaccggaattcactggccgttgttacaacgtc
gtgactggaaaaccctggcgtacccaacttaatgcctgcagcacatcccccttcgcagctggcgtaatagcgaag
aggcccgaccgatgcctcccaacagttgcgcagcctgaatggcgaatggcgcctgatgcggattttctcctacgc
atctgtgcgtatttcacacccgcatatggcacttcgtactacaatctgtatgccgtatgttaagccagccccgacac
ccgccaacacccgctgacgcgcctgacgggcttgtctcccggcatccgcttacagacaagctgtgaccgtccgg
gagctgcattgtcagaggtttaccgtatcaccgaaacgcgcga