

GenRCA Rare Codon Analysis Report

Gene Name: YFP

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: Pichia pastoris (budding yeasts)

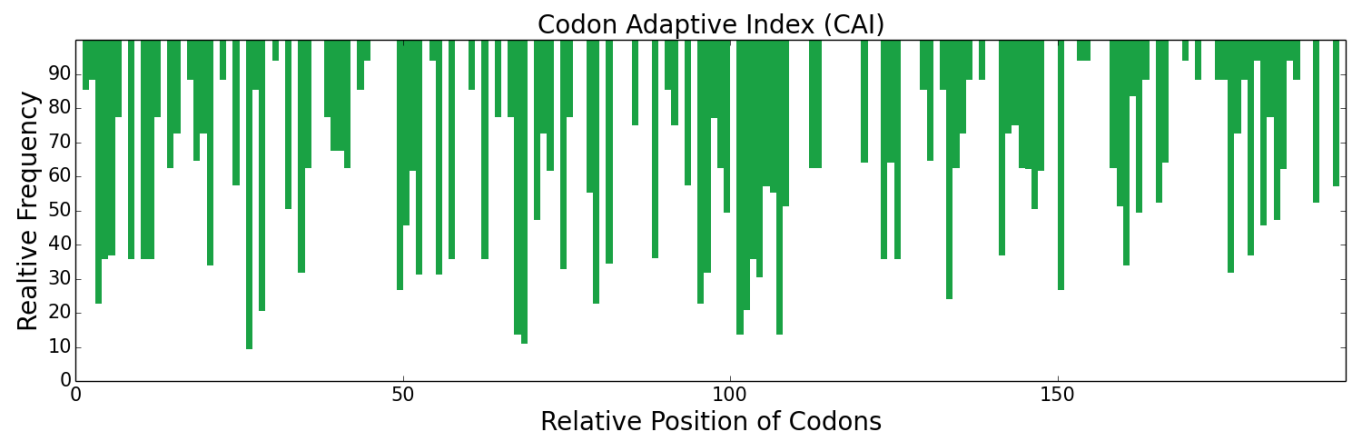
DNA Sequence Length: 582

DNA Sequence:

```
ATGTTCAAAGGGATAGTAGAGGGTATAGGTATAATAGAGAAGATCGACATTTATACCGAC
CTAGATAAATACGCCATTCGGTTCCCGGAAAAATATGCTTAACGGCATCAAGAAGGAGTCC
TCCATCATGTTCAATGGTTGTTTTTTGACGGTGACAAGCGTTAATAGCAACATAGTTTGG
TTTCGATATATTTGAGAAGGAGGCGCGCAAGCTGGACACATTTAGGGAGTACAAGGTCGGG
GATCGTGTTAACTTGGGAAGTTTTTCCCAAGTTCGGAGCTGCCTCTGGGGGCCACATCTTA
TCTGCGCGAATATCGTGCGTCGCGAGTATTATTGAAATCATCGAAAACGAAGATTACCAA
CAGATGTGGATACAGATACCAGAAAACCTTCACCGAATTCCTCATCGACAAAGATTATATT
GCTGTAGACGGAATCTCACTTACAATTGATACGATTAAGAATAATCAATTTTTTATCAGT
CTACCTTTAAAAATTGCACAGAACACTAATATGAAATGGAGAAAAAAGGCGACAAAGTA
AATGTGGAGCTGTCAAATAAAATTAACGCAAACCAATGCTGG
```

Protein Sequence:

```
MFKGIVEGIGIIEKIDIYTDLDKYAIRFPENMLNGIKKESIMFNGCFLTVTSVNSNIVW
FDIFEKEARKLDTFREYKVGDRVNLGTFPKFGAASGGHILSARISCVASIEIENEDYQ
QMWIQIPENFTEFLIDKDYIAVDGISLTIDTIKNNQFFISLPLKIAQNTNMKWRKKGDKV
NVELSNKINANQCW
```



Negative CIS Elements	Negative repeat Elements
3	0

Indices based on non-uniform usage of synonymous codon

Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.97	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥ 0	0.14	[41]
DCBS	Directional Codon Bias Score	≥ 1	1.43	[51]
CDC	Codon Deviation Coefficient	0~1	0.06	[61]
MILC	Measure Independent of Length and Composition	-1~1	-0.7	[71]
ICDI	Intrinsic Codon Deviation Index	0~1	0.07	[81]
SCUO	Synonymous Codon Usage Order	0~1	0.01	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.94	[111]
P	Codon Preference	≥ 1	1.02	[121]
MCB	Maximum-likelihood Codon Bias	≥ 0	0.01	[131]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.66	[11]
CFD	Codon Frequency Distribution	0~1	0.07	[21]
FOP	Frequency of Optimal Codons	0~1	0.36	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	-0.14	[51]
COUSIN18			-0.44	
CBI	Codon Bias Index	-1~1	-0.01	[61]
Dmean	Mean Dissimilarity-based Index	0~2	1.09	[71]
RCA	Relative Codon Adaptation	≥ 0	1.03	[81]
CUFS	Codon Usage Frequency Similarity	0~ $\sqrt{2}$	0.25	[91]
B	Codon Usage Bias	0~2	0.27	[101]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.48	[11]
gtAI	Genetic tRNA Adaptation Index	0~1	0.58	[21]
P2	P2 Index	0~1	0.52	[31]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
	GC Content at the Third Position			

GC3	of Synonymous Codons	0~1	0.5	[1]
GC	GC Content		0.4	
GC1	GC Content at the First Position of Synonymous Codons		0.4	
GC2	GC Content at the Second Position of Synonymous Codons		0.3	
ENcp	Effective Number of Codon Pairs	20~61	30.36	[2]
CPS	Codon Pair Score	-1~1	-0	[3, 4]
Codon Volatility	Codon Volatility	0.5~1	0.78	[5]

Gene Name: Beta-galactosidase

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: *Pichia pastoris* (budding yeasts)

DNA Sequence Length: 3072

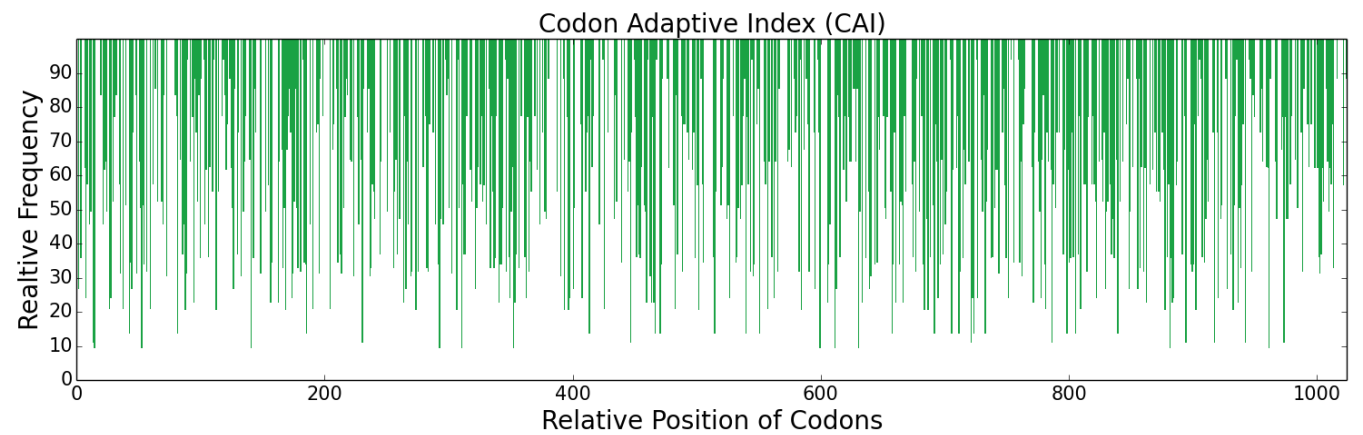
DNA Sequence:

```
ATGACGATGATAACTGATTCACTCGCCGTTGTGTTACAACGCCGGGATTGGGAAAACCCCT
GGTGTGACACAGTTAAACCGACTCGCTGCACACCCTCCTTTTGCCAGCTGGCGAAACAGT
GAAGAAGCGCGTACGGACAGACCTAGCCAACAGCTTCGGAGTCTAAACGGCGAATGGCGA
TTTGCCTGGTTCCAGCACCAGAAGCAGTGCCTGAATCGTGGTTGGAATGTGATTTGCCT
GAGGCGGATACCGTTGTAGTGCCGAGCAATTGGCAGATGCACGGGTATGACGCACCTATA
TATACTAATGTGACATACCCCATCACTGTCAATCCACCGTTTGTCCCACTGAGAATCCT
ACAGGATGTTATTCACTTACGTTCAACGTAGACGAATCGTGGTTACAGGAGGGTCAAACC
CGGATTATATTCGACGGTGTTAACAGCGCTTTTCATTTATGGTGCAACGGGCGTTGGGTT
GGTTACGGGCAGGATTCCAGGCTTCCGTCCGAGTTCGACCTCAGCGCATTCCCTTAGGGCT
GGCGAAAATCGTCTAGCGGTTATGGTGTGCGATGGTCTGACGGAAGCTATTTGGAGGAT
CAAGATATGTGGCGAATGTCCGGAATTTTCCGTGATGTAAGCTTGCTTCACAAACCTACT
ACCCAGATTTCCGATTTTCACGTGGCTACCCGCTTCAACGATGACTTCTCGAGGGCCGTC
CTGGAAGCTGAAGTACAGATGTGTGGTGAATTAAGAGACTACTTGAGGGTCACCGTATCT
CTGTGGCAAGGGGAGACGCAGGTGGCTTCGGGCACTGCTCCGTTTGGCGGTGAAATTATC
GATGAGAGGGGGCGGATACGCTGACAGAGTGACTCTACGGCTGAACGTGGAAAATCCTAAA
CTTTGGAGCGCTGAAATTCCGAATCTTTACCGGGCCGTAGTAGAGTTGCATACAGCAGAT
GGTACGCTTATCGAAGCCGAGGCATGCGATGTGGGTTTCAGGGAGGTCAGGATAGAGAAC
GGGCTACTACTTTTGAATGGCAAACCCCTCTTAATCCGGGGGGTAAACAGGCATGAGCAT
CATCCCTCCACGGCCAGGTCATGGATGAGCAAACAATGGTGCAAGACATTTTACTGATG
AAACAAAACAACTTTAACGCTGTGAGATGTTGCGATTATCCGAACCACCCGCTCTGGTAC
ACGCTTTGTGATAGATACGGAATCTACGTTGTGACGAGGCGAATATCGAGACTCATGGT
ATGGTGCCAATGAATCGATTGACTGACGATCCAAGATGGCTGCCTGCAATGTCTGAACGT
GTTACCAGAATGGTCCAGCGCGACAGAAATCACCCAGTGTAATAATCTGGAGTTTAGGG
AATGAATCGGGGCACGGGGCGAATCATGATGCGCTATATAGATGGATTAAATCAGTTGAT
CCAAGTCGACCTGTACAATACGAGGGCGGAGGAGCTGACACCACTGCTACAGACATAATT
TGCCCGATGTATGCCCGTGTGATGAAGATCAACCATTTCGCCGCGGTCCCAAAGTGGAGT
ATCAAAAAGTGGCTGAGTCTCCCAGGTGAGACTCGTCCCCTTATTCTGTGCGAGTATGCG
CACGCCATGGGCAATTCGCTAGGTGGATTTCGCGAAGTACTGGCAGGCCCTTCGACAGTAC
CCCAGACTCCAGGGTGGCTTCGTTTGGGATTGGGTTGATCAGTCCTTGATCAAGTATGAC
GAGAACGGCAATCCGTGGTCCGCTTATGGAGGCGATTTTGGTGACACGCCAAATGACCGG
CAATTTTGTATGAACGGGCTAGTGTTTGTGATCGGACGCTCACCCGCTTTGACAGAG
GCAAAGCACCAGCAGCAATTCCTCCAGTTCGCGTTATCTGGGCAAACTATAGAAGTCACG
TCGGAATACCTATTTTCGTCACTCCGATAACGAGCTTCTGCACTGGATGGTCGCGCTAGAC
GGGAAGCCACTAGCAAGTGGGGAGGTCCCATTGGATGTTGCCCCCAGGGCAAACAATTA
ATTGAGCTACCGGAATGCCGCAACCCGAAAGTGCGGGACAGCTCTGGCTAACCGTAAGA
GTGGTCCAACCAAATGCGACAGCTTGGTCAGAGGCGGGCCATATATCCGCCTGGCAACAG
TGGCGCCTCGCGGAAAATCTCTCTGTTACCTTACCTGCGGCATCCCATGCTATACCCCAT
CTTACAACAAGCGAAATGGACTTCTGCATAGAACTCGGTAATAAGCGTTGGCAATTTAAT
CGTCAGTCGGGATTCTTGTCTCAAAATGTGGATCGGGGACAAGAAGCAGTTACTCACCCCT
CTACGAGACCAGTTTACACGCGCCCCACTTGACAACGACATTGGTGTATCCGAAGCGACA
CGTATAGACCCCAACGCGTGGGTAGAGCGATGGAAGGCCCGGGCCACTACCAAGCCGAG
GCCGCACTCTTGCAGTGTACCGCAGACACGTTAGCAGATGCCGTACTGATAACTACCGCG
```

CATGCATGGCAGCATCAAGGAAAAACGTTATTTATCTCTAGAAAAACGTATAGGATCGAT
GGCTCAGGGCAAATGGCCATTACAGTTGACGTGCAAGTCGCATCTGACACACCGCATCCC
GCCCCGATAGGGCTCAACTGCCAGTTGGCTCAAGTAGCTGAGCGCGTTAACTGGCTAGGC
CTAGGACCGCAAGAGAATTATCCCGATCGTCTGACCGCAGCTTGTTTTGACCGCTGGGAC
CTCCCACTGAGTGATATGTATACGCCATACGTATTTCCGAGTGAGAATGGGCTTAGGTGC
GGAActCGCGAATTGAATTATGGCCCTCACCAATGGAGAGGTGACTTCCAGTTTAACATC
TCACGGTATTCTCAACAACAGCTGATGGAAACTTCCCACCGCCACCTGCTGCATGCCGAG
GAAGGTACCTGGCTTAACATTGACGGCTTCCATATGGGAATCGGTGGAGATGATTTCATGG
TCACCCAGCGTATCAGCAGAATTCCAGTTATCAGCCGGAAGGTACCATTATCAATTGGTT
TGGTGCCAAAAA

Protein Sequence:

MTMITDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWR
FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPIYTNVTYPITVNPFFVPTENP
TGCYSLTFNVDESWLQEGQTRIIIFDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLRA
GENRLAVMVLRWSDGSYLEDDQDMWRMSGIFRDVSLHHPPTTQISDFHVATRFDNDFSRAV
LEAEVQMCGEIRDYLRTVSLWQGETQVASGTAPFGGEIIDERGGYADRVTLRLNVENPK
LWSAEIPNLRYAVVELHTADGTLIEAEACDVGFREVRIENGLLLLNGKPLLRGVNRHEH
HPLHGQVMDEQTMVQDILLMKQNNFNAVRCSHYPNHPLWYTLCDRYGLYVVDEANIETHG
MVPNMRLTDDPRWLPAMSERVTRMVQRDRNHPSVIIWSLGNESGHGANHDALYRWIKSVD
PSRPVQYEGGGADTTATDIIICPMYARVDEDQFFPAVPKWSIKKWLSLPGETRPLILCEYA
HAMGNSLGGFAKYWQAFRQYPRLQGGFVWDWVDQSLIKYDENGPNWSAYGGDFGDTPNDR
QFCMNLVLFADRTPHPALTEAKHQQQFFQFRLSGQTIEVTSEYLFRRSDNELLHWMVALD
GKPLASGEVPLDVAPQGKQLIELPELPQPESAGQLWLTVRVVQPNATAWSEAGHISAWQQ
WRLAENLSVTLPAASHAIPHLTTSEMDFCIELGNKRWQFNRQSGFLSQMWIGDKKQLLTP
LRDQFTRAPLDNDIGVSEATRDPNAWVERWKAAGHYQAEAAALLQCTADTLADAVLITTA
HAWQHQQKTLFISRKTYRIDGSGQMAITVDVEVASDTPHPARIGLNCQLAQVAERNWL
LGPQENYPDRLTAACFDRWDLPLSDMYTPYVFPSENGLRCTRELNYGPHQWRGDFQFNI
SRYSQQQLMETSHRHLHLAAEEGTWLNIDGFHMGIGGDDSWSPSVSAEFQLSAGRYHYQLV
WCQK



Negative CIS Elements	Negative repeat Elements
0	0

Indices based on non-uniform usage of synonymous codon				
Index	Description	Range	Value	Reference

RSCU	Relative Synonymous Codon Usage	0~6	0.94	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥ 0	0.09	[4]
DCBS	Directional Codon Bias Score	≥ 1	1.31	[5]
CDC	Codon Deviation Coefficient	0~1	0.05	[6]
MILC	Measure Independent of Length and Composition	-1~1	-0.54	[7]
ICDI	Intrinsic Codon Deviation Index	0~1	0	[8]
SCUO	Synonymous Codon Usage Order	0~1	0	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.94	[11]
P	Codon Preference	≥ 1	1	[12]
MCB	Maximum-likelihood Codon Bias	≥ 0	0	[13]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.6	[11]
CFD	Codon Frequency Distribution	0~1	0.11	[2]
FOP	Frequency of Optimal Codons	0~1	0.34	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	0.05	[5]
COUSIN18			0.04	
CBI	Codon Bias Index	-1~1	0.01	[6]
Dmean	Mean Dissimilarity-based Index	0~2	0.42	[7]
RCA	Relative Codon Adaptation	≥ 0	1.01	[8]
CUFS	Codon Usage Frequency Similarity	0~ $\sqrt{2}$	0.18	[9]
B	Codon Usage Bias	0~2	0.27	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.45	[11]
gtAI	Genetic tRNA Adaptation Index	0~1	0.52	[2]
P2	P2 Index	0~1	0.5	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	0~1	0.52	[11]
GC	GC Content		0.51	

GC1	GC Content at the First Position of Synonymous Codons		0.58	
GC2	GC Content at the Second Position of Synonymous Codons		0.44	
ENcp	Effective Number of Codon Pairs	20~61	61	[21]
CPS	Codon Pair Score	-1~1	0	[3, 4]
Codon Volatility	Codon Volatility	0.5~1	0.75	[51]

Gene Name: Cre

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: Pichia pastoris (budding yeasts)

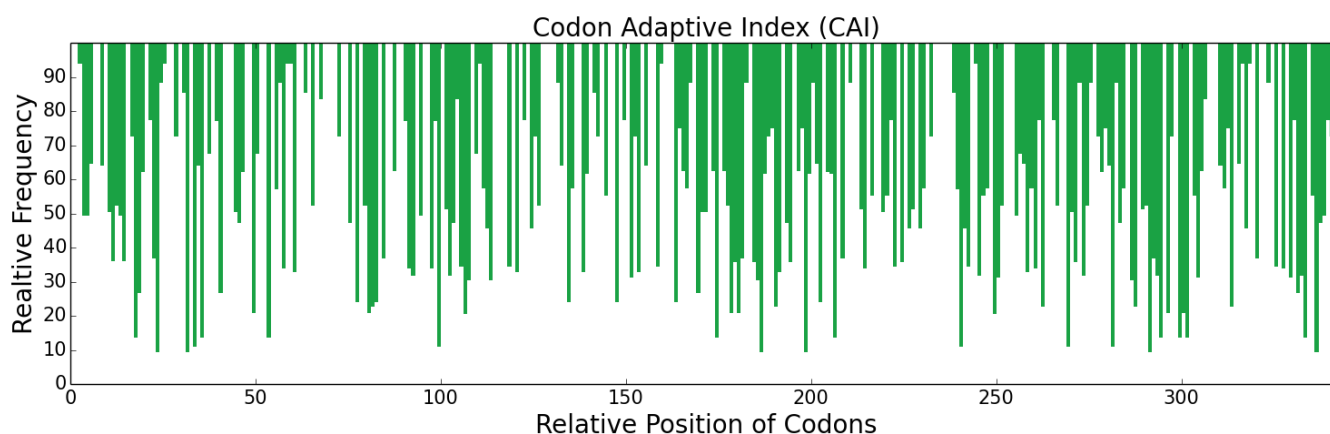
DNA Sequence Length: 1029

DNA Sequence:

```
ATGTCTAATTTATTAACCGTTCATCAGAACCTTCCCGCATTACCCGTTGACGCGACGTCA
GATGAGGTACGGAAAAATTTGATGGACATGTTCCGGGATCGCCAGGCGTTTTCCGAACAC
ACGTGGAAGATGCTTCTGTTCAGTTTGTTCGATCCTGGGCTGCGTGGTGCAAATAAATAAT
AGGAAGTGGTTCCCAGCAGAACCTGAAGATGTTAGAGACTACTTGCTGTACCTCCAAGCA
CGAGGGCTCGCTGTAAAGACTATCCAACAACACCTAGGCCAATTAAACATGCTACACCGC
AGAAGTGGCCTGCCTCGTCCGTCGGATTCCAATGCCGTGTCGTTGGTTATGAGACGTATT
AGGAAGGAGAACGTGGACGCAGGTGAAAGAGCTAAACAGGCTCTCGCCTTTGAAAGGACA
GATTTTCGACCAAGTCAGATCTCTCATGGAGAACAGCGACAGGTGTCAGGATATTCGTAAT
TTGGCTTTTCTCGGAATCGCCTATAACACGCTTCTTAGAATCGCGGAAATCGCACGAATA
CGAGTAAAGATATATCGCGGACAGACGGAGGGAGGATGCTGATACATATCGGACGGACA
AAAACCTCGTTTTCAACAGCGGGTGTAGAAAAAGCTTTGAGTCTAGGTGTCACCTAAGCTT
GTTCGAGCGTTGGATATCTGTGAGTGGTGTGGCCGATGACCCAAACAACTACTTGTTCGTC
CGCGTGCGTAAGAATGGCGTCGCCGCTCCGAGCGCAACTTCTCAATTATCCACCAGGGCC
CTAGAGGGGATTTTTGAGGCAACTCATCGCCTTATATATGGCGCAAAAGATGACTCAGGA
CAGCGCTATCTGGCCTGGTTCGGGGCATAGTGCACGGGTAGGCGCGGCTCGAGACATGGCG
CGAGCGGGTGTGAGCATCCCTGAAATTATGCAGGCCGGAGGGTGGACCAATGTGAATATT
GTAATGAACATATTCGTAACCTAGATAGCGAGACGGGCGCGATGGTCCGGCTGTTAGAG
GACGGAGAT
```

Protein Sequence:

```
MSNLLTVHQNLPALPVDATSDVRKNLMDMFRDRQAFSEHTWKMLLSVCRSWAAWCKLNN
RKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRI
RKENVDAGERAKQALAFERTDFDQVRSLMENS DRCQDIRNLAFLGIAYNTLLRIAEIARI
RVKDISRTDGGRM LIHIGRTKTLVSTAGVEKALSLGVTKLVERWISVSGVADDPNNYLC
RVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDMA
RAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGD
```



Negative CIS Elements	Negative repeat Elements
0	0

Indices based on non-uniform usage of synonymous codon

Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.95	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥0	0.08	[4]
DCBS	Directional Codon Bias Score	≥1	1.32	[5]
CDC	Codon Deviation Coefficient	0~1	0.04	[6]
MILC	Measure Independent of Length and Composition	-1~1	-0.62	[7]
ICDI	Intrinsic Codon Deviation Index	0~1	0	[8]
SCUO	Synonymous Codon Usage Order	0~1	0	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.94	[11]
P	Codon Preference	≥1	1.01	[12]
MCB	Maximum-likelihood Codon Bias	≥0	0	[13]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.56	[1]
CFD	Codon Frequency Distribution	0~1	0.13	[2]
FOP	Frequency of Optimal Codons	0~1	0.31	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	-0.07	[5]
COUSIN18			-0.12	
CBI	Codon Bias Index	-1~1	-0.01	[6]
Dmean	Mean Dissimilarity-based Index	0~2	0.95	[7]
RCA	Relative Codon Adaptation	≥0	0.98	[8]
CUFS	Codon Usage Frequency Similarity	0~√2	0.18	[9]
B	Codon Usage Bias	0~2	0.3	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.43	[1]
gtAI	Genetic tRNA Adaptation Index	0~1	0.51	[2]
P2	P2 Index	0~1	0.5	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	0~1	0.51	[1]
GC	GC Content		0.51	
GC1	GC Content at the First Position of Synonymous Codons		0.56	
GC2	GC Content at the Second Position of Synonymous Codons		0.44	
ENcp	Effective Number of Codon Pairs	20~61	61	[2]
CPS	Codon Pair Score	-1~1	-0.01	[3, 4]
Codon Volatility	Codon Volatility	0.5~1	0.74	[5]

Gene Name: Luciferase

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: *Pichia pastoris* (budding yeasts)

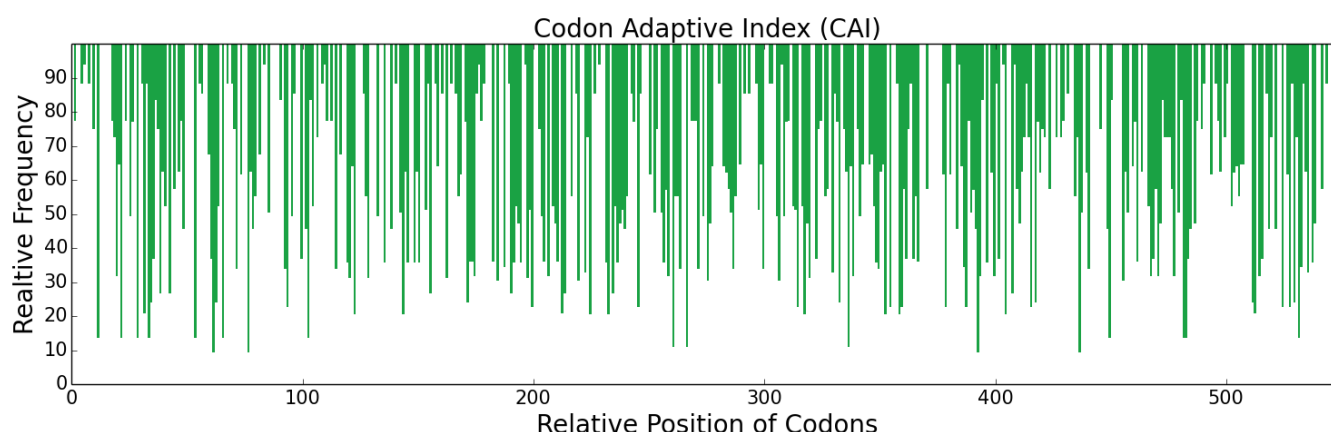
DNA Sequence Length: 1650

DNA Sequence:

```
ATGGAGGATGCTAAAAATATTAAAAAGGGACCAGCGCCATTTTACCCATTGGAGGACGGC
ACCGCGGGTGAGCAATTACACAAGGCGATGAAACGATATGCGCTCGTACCTGGAACGATC
GCATTTACGGATGCCCATATCGAGGTGAACATTACTTACGCGGAATATTTTCGAAATGTCC
GTACGGCTCGCAGAAGCGATGAAAAGATATGGACTAAACACAAACCATCGGATCGTGGTC
TGTTCCGAAAATTCTCTCAATTTTTTATGCCTGTTCTAGGGGCTTTATTCATTGGTGTA
GCTGTGGCGCCTGCAAACGACATTTATAATGAGAGAGAGTTGCTAAACTCCATGAACATA
AGCCAGCCGACTGTTGTTTTCGTCAGCAAGAAGGGTTTACAAAAGATATTGAACGTGCAA
AAAAAGCTTCCGATCATACAAAAGATAATCATAATGGATAGTAAAACGGATTATCAGGGT
TTCCAAAGCATGTATACTTTTCGTCACATCTCACCTCCCCCGGCTTCAATGAGTATGAT
TTTGTTCCCGAATCGTTTGATCGTGATAAAACGATAGCACTGATAATGAATAGCAGTGGG
TCTACTGGATTACCCAAGGGCGTTGCACTGCCCCATCGAACGGCTTGTGTCAGATTCTCG
CATGCTAGGGACCCGATTTTCGGTAATCAAATTATACCGGATACGGCAATACTGAGTGTG
GTCCCATTCCACCATGGGTTCGGTATGTTTACAACCTTTGGATACCTTATATGCGGCTTT
CGCGTCGTCCTAATGTACCGCTTTGAGGAGGAGCTATTTTAAAGATCGCTGCAGGATTAC
AAAATTCAAGTCAGCCCTTCTAGTCCCAACCTTGTTCTCTTTCTTTGCTAAAAGTACCCTA
ATTGATAAATATGATTTATCGAATTTACACGAGATTGCAAGTGGGGGTGCACCGCTGAGC
AAGGAAGTAGGAGAGGCTGTGCGCAAGAGGTTCCACCTCCCAGGAATCCGCCAGGGCTAC
GGATTAACCGAAACTACCTCCGCAATACTAATCACCCCGGAAGGGGATGATAAACCGGGG
GCCGTAGGAAAAGTAGTCCCTTTTTTTGAAGCCAAGGTTGTTGATTTGGATACAGGGAAA
ACATTGGGTGTGAATCAGCGTGGGGAGCTTTGCGTGCGGGGCCCTATGATAATGTCAGGC
TATGTAAACAATCCGGAAGCTACGAATGCCCTGATCGACAAAGACGGGTGGCTCCACTCA
GGAGACATTGCCTACTGGGACGAAGACGAGCATTCTTTATTGTCGACCGGCTTAAGTCA
CTAATTAAGTACAAGGGATACCAAGTGGCGCCTGCTGAATTGGAATCGATCCTTTTGCAG
CACCCCAACATCTTTGATGCAGGCGTAGCCGGCCTGCCTGACGACGACGCCGGCGAACTT
CCTGCGGCGGTAGTGGTTCTGGAGCATGGAAAACTATGACAGAAAAAGAGATCGTTGAC
TATGTTGCATCACAGGTCACCACCGCTAAGAAGCTCCGAGGTGGCGTAGTTTTTCGTGGAC
GAAGTGCCAAAGGGGTTGACAGGGAAACTCGACGCGCGTAAATCAGGGAAATACTGATT
AAGGCCAAGAAAGGTGGTAAATCTAAGCTC
```

Protein Sequence:

```
MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTDAHIEVNIITYAEYFEMS
VRLAEAMKRYGLNTNHRIVVCSSENSLQFFMPVLGALFIGVAVAPANDIYNERELLNSMNI
SQPTVVFVSKKGLQKILNVQKKLP I I Q K I I I M D S K T D Y Q G F Q S M Y T F V T S H L P P G F N E Y D
F V P E S F D R D K T I A L I M N S S G S T G L P K G V A L P H R T A C V R F S H A R D P I F G N Q I I P D T A I L S V
V P F H H G F G M F T T L G Y L I C G F R V V L M Y R F E E E L F L R S L Q D Y K I Q S A L L V P T L F S F F A K S T L
I D K Y D L S N L H E I A S G G A P L S K E V G E A V A K R F H L P G I R Q Y G L T E T T S A I L I T P E G D D K P G
A V G K V V P F F E A K V V D L D T G K T L G V N Q R G E L C V R G P M I M S G Y V N N P E A T N A L I D K D G W L H S
G D I A Y W D E D E H F F I V D R L K S L I K Y K Y Q V A P A E L E S I L L Q H P N I F D A G V A G L P D D D A G E L
P A A V V V L E H G K T M T E K E I V D Y V A S Q V T T A K K L R G G V V F V D E V P K G L T G K L D A R K I R E I L I
K A K K G G K S K L
```



Negative CIS Elements	Negative repeat Elements
3	0

Indices based on non-uniform usage of synonymous codon

Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.97	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥ 0	0.08	[41]
DCBS	Directional Codon Bias Score	≥ 1	1.29	[51]
CDC	Codon Deviation Coefficient	0~1	0.05	[61]
MILC	Measure Independent of Length and Composition	-1~1	-0.57	[71]
ICDI	Intrinsic Codon Deviation Index	0~1	0	[81]
SCUO	Synonymous Codon Usage Order	0~1	0	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.97	[111]
P	Codon Preference	≥ 1	1	[121]
MCB	Maximum-likelihood Codon Bias	≥ 0	0	[131]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.62	[11]
CFD	Codon Frequency Distribution	0~1	0.1	[21]
FOP	Frequency of Optimal Codons	0~1	0.34	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	-0.06	[51]
COUSIN18			-0.12	
CBI	Codon Bias Index	-1~1	0	[61]
Dmean	Mean Dissimilarity-based Index	0~2	0.7	[71]

RCA	Relative Codon Adaptation	≥ 0	1.01	[8]
CUFS	Codon Usage Frequency Similarity	$0 \sim \sqrt{2}$	0.18	[9]
B	Codon Usage Bias	$0 \sim 2$	0.28	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	$0 \sim 1$	0.46	[11]
gtAI	Genetic tRNA Adaptation Index	$0 \sim 1$	0.54	[2]
P2	P2 Index	$0 \sim 1$	0.49	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	$0 \sim 1$	0.5	[11]
GC	GC Content		0.47	
GC1	GC Content at the First Position of Synonymous Codons		0.55	
GC2	GC Content at the Second Position of Synonymous Codons		0.36	
ENcp	Effective Number of Codon Pairs	$20 \sim 61$	33.05	[2]
CPS	Codon Pair Score	$-1 \sim 1$	0.02	[3] , [4]
Codon Volatility	Codon Volatility	$0.5 \sim 1$	0.76	[5]

Gene Name: lacI transcription factor

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: Pichia pastoris (budding yeasts)

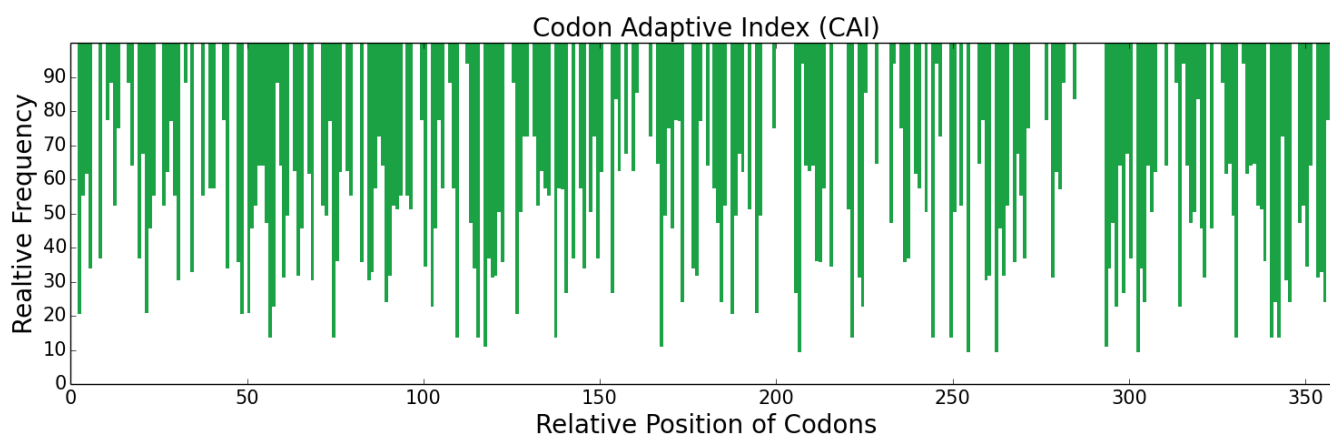
DNA Sequence Length: 1080

DNA Sequence:

```
ATGAAGCCGGTCACACTATACGATGTAGCTGAGTATGCAGGAGTTTCTTATCAGACTGTA
TCCCGAGTGGTCAACCAAGCATCACACGTCTCGGCTAAAACTAGGGAAAAGGTCGAAGCC
GCCATGGCTGAGCTAAACTACATAACCGAACCAGTGGCACAGCAGCTGGCGGGGAAACAG
AGCTTATTGATCGGCGTGGCTACATCGTCTTTGGCATTACACGCGCCCTCACAAATCGTC
GCTGCTATAAAGTCGAGGGCCGACCAGCTCGGCGCAAGTGTCGTTGTCAGTATGGTTGAG
CGTTCTGGGGTGGAGGCCTGTAAAGCCGCGGTTTATAATCTGCTAGCGCAACGCGTAAGC
GGCCTTATAATTAATATCCGCTTGACGACCAAGACGCAATCGCCGTCGAAGCGGCCTGC
ACGAACGTACCAGCCCTATTTCTTGACGTATCAGATCAAACGCCTATCAACTCCATTATC
TTCTCTCATGAAGACGGTACCCGCTTAGGAGTGGAGCACCTCGTTGCTCTAGGCCACCAA
CAGATTGCCCTGCTCGCAGGTCCGTTATCCTCAGTTAGTGTCTGATTAAGATTGGCTGGA
TGGCATAAGTACTTGACGCGGAATCAGATCCAGCCCATAGCCGAACGTGAAGGTGATTGG
AGTGCGATGAGCGGGTTCCAACAAACCATGCAAATGCTGAATGAAGGAATAGTACCAACA
GCCATGCTTGTTGCGAATGACCAAAATGGCGCTTGGTGCAATGCGGGCTATTACCGAGTCG
GGCTTGCGGGTGGGCGCAGATATATCCGTCGTAGGATACGATGATACTGAGGATAGCTCA
TGCTATATTCCACCTTTGACTACTATTAAGCAAGATTTTCGCCTACTGGGGCAGACGTCC
GTAGATCGGCTACTCCAGCTTTTACAAGGTCAGGCTGTTAAAGGGAATCAGCTGCTTCCT
GTGAGCTTGGTGAAGAGAAAAACAACCTTAGCGCCAAATACACAGACCGCAAGTCCCAGA
GCGCTCGCGGACTCGCTCATGCAACTGGCACGTCAGGTTAGCAGGCTCGAGTCTGGTCAA
```

Protein Sequence:

```
MKPVTLTYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNYIPNRVAQQLAGKQ
SLIIGVATSSLALHAPSQIVAAIKSRADQLGASVVVSMVERSGVEACKAAVHNLLAQRVS
GLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQ
QIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPT
AMLVANDQMALGAMRAITESGLRVGADISVVGYYDDTEDSSCYIPPLTTIKQDFRLLGQTS
VDRLLQLSQGQAVKGNQLLPVSLVKRKTTLAPNTQTASPRALADSLMQLARQVSRLESGQ
```



Negative CIS Elements	Negative repeat Elements
0	0

Indices based on non-uniform usage of synonymous codon

Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.97	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥ 0	0.08	[4]
DCBS	Directional Codon Bias Score	≥ 1	1.34	[5]
CDC	Codon Deviation Coefficient	0~1	0.06	[6]
MILC	Measure Independent of Length and Composition	-1~1	-0.61	[7]
ICDI	Intrinsic Codon Deviation Index	0~1	0.01	[8]
SCUO	Synonymous Codon Usage Order	0~1	0	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.96	[11]
P	Codon Preference	≥ 1	1.01	[12]
MCB	Maximum-likelihood Codon Bias	≥ 0	0.01	[13]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.57	[1]
CFD	Codon Frequency Distribution	0~1	0.11	[2]
FOP	Frequency of Optimal Codons	0~1	0.31	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	0.02	[5]
COUSIN18			-0.04	
CBI	Codon Bias Index	-1~1	0	[6]
Dmean	Mean Dissimilarity-based Index	0~2	0.89	[7]
RCA	Relative Codon Adaptation	≥ 0	0.98	[8]
CUFS	Codon Usage Frequency Similarity	0~ $\sqrt{2}$	0.2	[9]
B	Codon Usage Bias	0~2	0.31	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.43	[1]
gtAI	Genetic tRNA Adaptation Index	0~1	0.51	[2]
P2	P2 Index	0~1	0.5	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	0~1	0.51	[1]
GC	GC Content		0.52	
GC1	GC Content at the First Position of Synonymous Codons		0.62	
GC2	GC Content at the Second Position of Synonymous Codons		0.43	
ENcp	Effective Number of Codon Pairs	20~61	34.02	[2]
CPS	Codon Pair Score	-1~1	0.01	[3, 4]
Codon Volatility	Codon Volatility	0.5~1	0.74	[5]

Gene Name: TetR transcription factor

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: Pichia pastoris (budding yeasts)

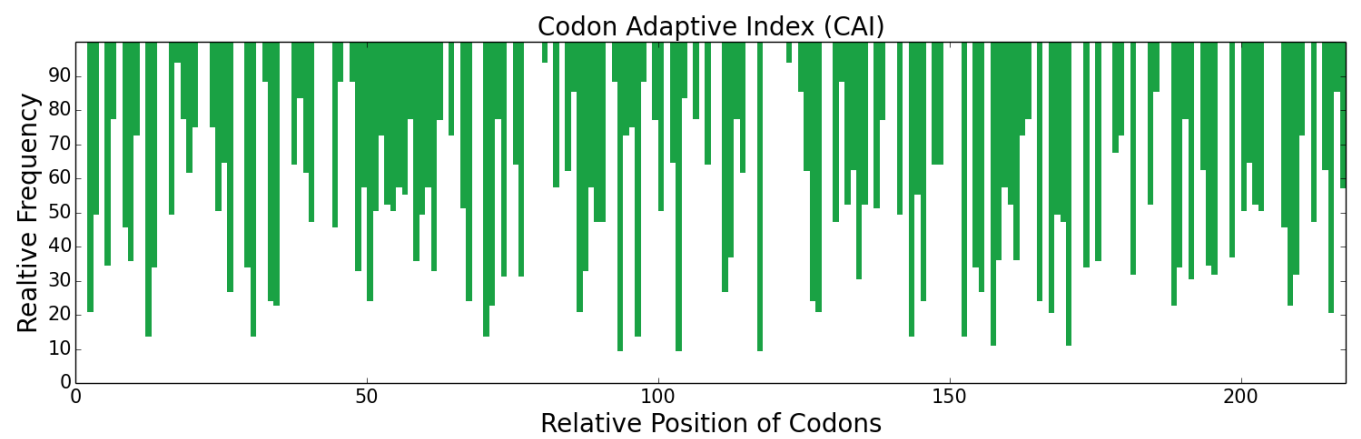
DNA Sequence Length: 654

DNA Sequence:

```
ATGGCTCGATTAAACCGTGAGTCTGTGATAGACGCTGCGCTAGAATTGTTAAATGAGACA
GGAATTGATGGACTTACCACGAGAAAAGCTAGCGCAAAAACTCGGGATTGAACAGCCTACA
CTGTACTGGCATGTGAAAAACAAAAGGGCCCTCCTTGACGCACTTGCCGTCGAGATATTA
GCCAGGCACCATGACTACAGTCTCCCAGCTGCGGGGGAGAGCTGGCAGAGCTTTTTGAGA
AATAACGCCATGTCATTCCGAAGGGCCCTGCTGAGATATCGGGACGGAGCGAAAGTTCAC
CTTGGTACCCGGCCTGATGAGAAGCAGTACGATACGGTAGAGACACAATTGCGGTTTATG
ACTGAAAATGGTTTCTCACTCCGAGATGGTCTGTATGCAATCTCGGCAGTTAGTCACTTT
ACTTTAGGTGCGGTCTCTGAACAGCAGGAACATACTGCGGCTCTAACGGATCGCCCCGCC
GCACCCGACGAGAACCTCCCACCGTTACTGCGCGAAGCTCTACAAATAATGGATTCCGAC
GATGGCGAACAAGCATTCTTGCATGGGCTAGAGTCGTTGATCCGTGGCTTTGAAGTACAA
CTTACCGCACTTTTGCAAATTGTGGGGGGCGACAAGCTGATTATCCCGTTCTGC
```

Protein Sequence:

```
MARLNRESVIDAALELLNETGIDGLTTRKLAQKLGIEQPTLYWHVKNKRALLDALAVEIL
ARHHDYSLPAAGESWQSFLRNNAMSFRRALLRYRDGAKVHLGTRPDEKQYDTVETQLRFM
TENGFSLRDGLYAISAVSHFTLGAVLEQQEHTAALTDSPAAPDENLPPLLREALQIMSDS
DGEQAFHLGLESIRGFVQLTALLQIVGGDKLIIPFC
```



Negative CIS Elements	Negative repeat Elements
0	0

Indices based on non-uniform usage of synonymous codon				
Index	Description	Range	Value	Reference

RSCU	Relative Synonymous Codon Usage	0~6	0.99	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥ 0	0.11	[4]
DCBS	Directional Codon Bias Score	≥ 1	1.36	[5]
CDC	Codon Deviation Coefficient	0~1	0.07	[6]
MILC	Measure Independent of Length and Composition	-1~1	-0.68	[7]
ICDI	Intrinsic Codon Deviation Index	0~1	0.06	[8]
SCUO	Synonymous Codon Usage Order	0~1	0.01	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.96	[11]
P	Codon Preference	≥ 1	1.01	[12]
MCB	Maximum-likelihood Codon Bias	≥ 0	0.02	[13]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.58	[11]
CFD	Codon Frequency Distribution	0~1	0.13	[2]
FOP	Frequency of Optimal Codons	0~1	0.33	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	0.17	[5]
COUSIN18			0.17	
CBI	Codon Bias Index	-1~1	0.01	[6]
Dmean	Mean Dissimilarity-based Index	0~2	1.11	[7]
RCA	Relative Codon Adaptation	≥ 0	1.01	[8]
CUFS	Codon Usage Frequency Similarity	0~ $\sqrt{2}$	0.26	[9]
B	Codon Usage Bias	0~2	0.32	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.43	[11]
gtAI	Genetic tRNA Adaptation Index	0~1	0.51	[2]
P2	P2 Index	0~1	0.51	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	0~1	0.51	[11]
GC	GC Content		0.51	

GC1	GC Content at the First Position of Synonymous Codons		0.62	
GC2	GC Content at the Second Position of Synonymous Codons		0.39	
ENcp	Effective Number of Codon Pairs	20~61	61	[2]
CPS	Codon Pair Score	-1~1	0	[3] , [4]
Codon Volatility	Codon Volatility	0.5~1	0.75	[5]

Gene Name: araC transcription factor

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: Pichia pastoris (budding yeasts)

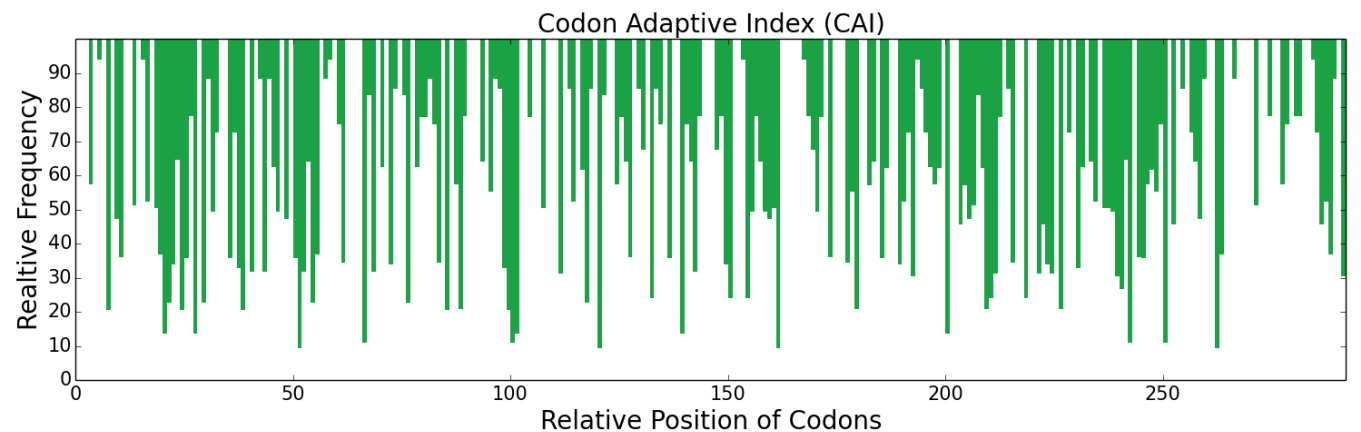
DNA Sequence Length: 876

DNA Sequence:

ATGGCTGAAGCCCCAAATGATCCGTTGCTGCCCGGTTACAGTTTTTAATGCACATCTTGTA
GCGGGGCTAACCCCGATAGAGGCGAACGGGTATTTAGACTTTTTTATAGACAGGCCGTTG
GGCATGAAAGGCTATATCTTAAACCTGACTATACGGGGCCAGGGGGTAGTTAAAAATCAA
GGACGTGAATTTGTTTGTCTGCCCTGGCGATATCTTGCTATTCCCACCTGGGGAAATCCAC
CACTATGGACGTCATCCGGAAGCCCGAGAGTGGTACCATCAGTGGGTCTATTTTCAGGCCG
CGCGCGTACTGGCACGAATGGCTTAACTGGCCAAGCATTTTCGCAAACACAGGGTTCTTT
CGGCCTGATGAAGCCCACCAGCCCCATTTCTCCGATCTCTTCGGACAAATAATTAACGCG
GGACAGGGCGAGGGTAGATACTCCGAGCTACTCGCTATTAATCTCTTAGAGCAGTTACTG
CTTCGGAGAATGGAAGCTATTAATGAGTCCTTACACCCACCCATGGATAACCGTGTCGA
GAAGCTTGCCAGTACATATCAGATCATCTAGCAGACTCGAATTTTCGACATCGCCTCAGTT
GCGCAACATGTGTGCCTGAGTCCTTCACGACTCAGCCACTTGTTCCGTCAACAACTCGGT
ATTAGCGTGCTAAGCTGGCGAGAAGACCAAAGGATCTCTCAGGCAAAGCTTCTTTTATCG
ACGACCCGCATGCCCATAGCCACAGTCGGACGCAACGTGGGTTTCGATGACCAGCTGTAT
TTTTCTCGGGTATTTAAGAAATGTACTGGTGCTAGTCCATCTGAGTTTAGAGCCGGATGT
GAGGAGAAGGTTAATGACGTGGCAGTAAAATTGTCTG

Protein Sequence:

MAEAQNDDLPLPGYSFNAHLVAGLTPIEANGYLDFFDIDRPLGMKGYILNLTIRGQGVVKNQ
GREFVCRPGDILLFPPGEIHYYGRHPEAREWYHQWVYFRPRAYWHEWLNWPSIFANTGFF
RPDEAHQPHFSDLFGQIINAGQGEGRYSELLAINLLEQLLLRRMEAINESLHPPMDNRVR
EACQYISDHLADSNFDIASVAQHVCCLSPSRLSHLFRQQLGISVLSWREDQRISQAKLLLS
TTRMPIATVGRNVGFDDQLYFSRVFKKCTGASPSEFRAGCEEKVNDVAVKLS



Negative CIS Elements	Negative repeat Elements
1	0

Indices based on non-uniform usage of synonymous codon

Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.97	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥ 0	0.09	[4]
DCBS	Directional Codon Bias Score	≥ 1	1.28	[5]
CDC	Codon Deviation Coefficient	0~1	0.04	[6]
MILC	Measure Independent of Length and Composition	-1~1	-0.64	[7]
ICDI	Intrinsic Codon Deviation Index	0~1	0.01	[8]
SCUO	Synonymous Codon Usage Order	0~1	0	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.96	[11]
P	Codon Preference	≥ 1	1.01	[12]
MCB	Maximum-likelihood Codon Bias	≥ 0	0.01	[13]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.59	[11]
CFD	Codon Frequency Distribution	0~1	0.11	[2]
FOP	Frequency of Optimal Codons	0~1	0.32	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	-0.12	[5]
COUSIN18			-0.15	
CBI	Codon Bias Index	-1~1	-0.01	[6]
Dmean	Mean Dissimilarity-based Index	0~2	0.89	[7]
RCA	Relative Codon Adaptation	≥ 0	1.01	[8]
CUFS	Codon Usage Frequency Similarity	0~ $\sqrt{2}$	0.19	[9]
B	Codon Usage Bias	0~2	0.31	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.44	[11]
gtAI	Genetic tRNA Adaptation Index	0~1	0.51	[2]
P2	P2 Index	0~1	0.5	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
-------	-------------	-------	-------	-----------

GC3	GC Content at the Third Position of Synonymous Codons	0~1	0.51	[1]
GC	GC Content		0.5	
GC1	GC Content at the First Position of Synonymous Codons		0.58	
GC2	GC Content at the Second Position of Synonymous Codons		0.4	
ENcp	Effective Number of Codon Pairs	20~61	61	[2]
CPS	Codon Pair Score	-1~1	0	[3, 4]
Codon Volatility	Codon Volatility	0.5~1	0.76	[5]

Gene Name: Regulatory protein GAL4

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: *Pichia pastoris* (budding yeasts)

DNA Sequence Length: 2643

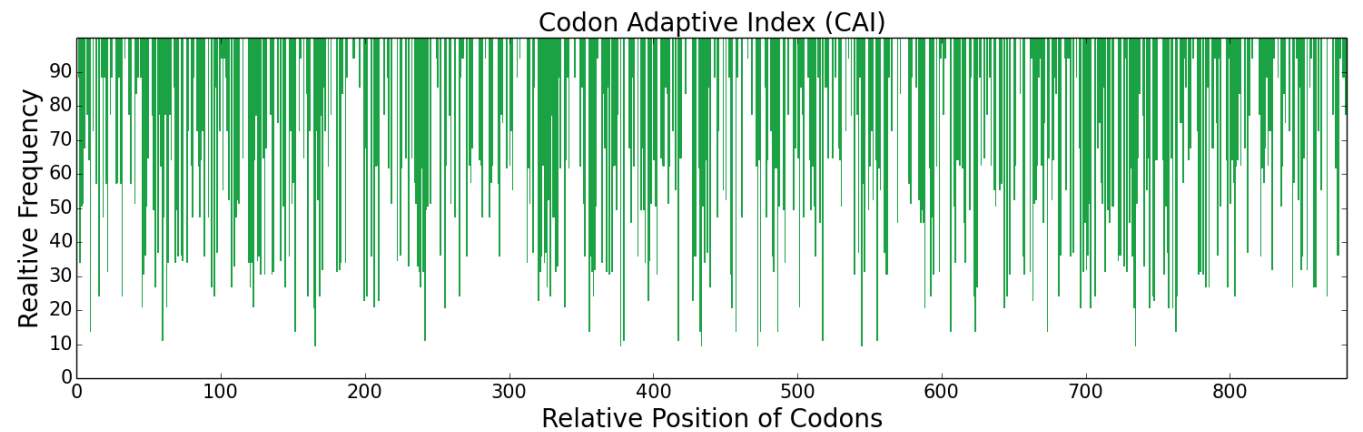
DNA Sequence:

```
ATGAAACTACTTAGTTCATTGAGCAGGCGTGTGACATTTGCAGACTCAAGAAACTGAAA
TGCAGCAAGGAGAAACCAAGTGCGCCAAATGCCTCAAGAATAACTGGGAGTGCAGATAC
AGTCCTAAAACTAAACGATCGCCCCCTTACCAGAGCTCACTTAACGGAAGTAGAGTCACGC
CTGGAGCGACTAGAGCAGTTGTTCTATTATAATTTCCCCGTGAAGATCTAGACATGATC
CTGAAAATGGACTCACTGCAGGACATAAAGGCTCTGTTGACGGGTCTCTTCGTACAAGAC
AATGTCAATAAAGACGCAGTTACGGATAGGCTGGCAAGTGTGAAACCGATATGCCACTA
ACGCTACGACAGCACCGTATATCGGCTACCTCGTCCTCTGAAGAGTCGAGCAACAAGGGA
CAACGTCAACTTACGGTTTCTATAGACAGTGCCGCGCATCATGACAATTCTACCATTCCCT
CTCGACTTTATGCCGCGGGACGCACTCCACGGCTTCGACTGGTCAGAAGAGGATGATATG
AGCGATGGCCTACCTTTTCTAAAAACTGATCCAAATAATAACGGTTTTTTTCGGTGATGGG
TCCCTCTTGTGTATATTGCGATCAATCGGGTTTAAGCCAGAGAACTATACAAACAGTAAC
GTTAACCGTTTGGCCACAATGATTACCGATAGGTACACCTTGGCCAGTAGGAGCACGACA
AGCCGCTTACTTCAAAGTTACTTGAACAATTTCCATCCCTACTGTCCGATCGTTCACAGT
CCAACTCTGATGATGCTCTATAATAATCAAAATAGAAAATCGCCTCCAAGGATCAATGGCAG
ATCCTGTTTTAATTGTATTCTGGCCATCGGTGCTTGGTGCATAGAGGGAGAATCTACAGAT
ATCGACGTCTTTTACTATCAAAATGCTAAGTCTCATCTAACAAGTAAGGTATTCTGAATCT
GGGAGCATAATCCTAGTAACGGCACTCCACCTGCTGAGCAGGTATACACAATGGCGACAG
AAAACAAACACTTCTTATAACTTTTCATTTCATTTCAGTATAAGAATGGCGATAAGCCTCGGC
CTTAATAGAGATCTACCTTCCAGCTTCTCGGATAGCTCAATCTTAGAGCAACGGAGACGC
ATTTGGTGGTCCGTGTATTTCATGGGAAATACAATTATCCTTACTATACGGGCGTAGTATT
CAGCTTTCGCAAAACACCATCTCATTCCCTAGTTCAGTTGACGATGTCCAACGCACCACC
ACTGGTCCTACTATTTACCATGGGATAATAGAAACAGCGCGGCTTCTACAGGTATTCACG
AAGATTTATGAACTGGACAAGACTGTCACTGCAGAAAAATCGCCGATTTGTGCGAAAAAG
TGTCTGATGATTTGTAATGAAATTGAGGAAGTTTCACGGCAGGCGCCAAAGTTCCTGCAG
ATGGACATAAGCACAACAGCGCTTACTAACTTACTAAAGGAGCATCCATGGTTATCCTTT
ACCCGATTTCGAGCTGAAGTGGAAACAGTTATCACTTATAATTTATGTGTTGCGCGATTTT
TTCACCAACTTTACCCAAAAAAAGTCCAGCTTGAACAAGACCAAAATGATCACCAATCG
TACGAGGTAAAACGGTGCAGCATTATGTTATCAGACGCAGCTCAGCGCACAGTCATGTCT
GTATCGTCGTATATGGACAACCATAACGTGACTCCTTACTTTGCTTGGAACCTGCAGTTAT
TACTTGTTCACGCAGTGTTAGTGCCGATTAAGACACTCCTGTGCAACTCAAAGAGCAAT
GCTGAGAATAACGAAACTGCGCAACTTCTACAGCAGATCAACACAGTGCTAATGTTGTTA
AAAAAGCTCGCGACGTTTAAATCCAAACCTGTGAGAAGTATATCCAAGTCCTTGAAGAA
GTCTGCGCTCCGTTTCTTGTGCAATGTGCAATCCCATTCGCTCATATATCGTACAAC
AACAGCAATGGGAGTGCAATCAAAAATATCGTGGGATCTGCGACCATTCGCACAGTATCCT
ACTCTCCCCGAAGAAAATGTCAATAATATATCTGTAAAGTATGTTTCCCCGGGTAGCGTG
GGCCCCAGTCCGGTTCCACTCAAGTCCGGAGCCAGTTTCAGCGATTTAGTGAAGCTTCTT
TCTAACCGTCCCCCTCTAGGAATAGCCCAGTGACAATACCGCGGAGTACCCCTTCTCAC
AGGTCTGTACCCCGTTTCTCGGGCAGCAGCAACAATTGCAGTCGCTTGTACCGTTGACC
CCGTCTGCGCTCTTTGGAGGAGCCAATTTCAACCAGTCCGGTAATATTGCTGACTCGAGC
TTGTGCTTTTACGTTTCACGAACTCCTCCAATGGACCCAATCTTATTACTACTCAGACGAAT
AGTCAGGCCCTCTCACAGCCTATCGCTTCTTCCAACGTACACGATAATTTTATGAACAAC
GAGATAACAGCCTCCAAATCGACGACGGCAATAACTCTAAACCACCTTTCACCTGGATGG
```

ACTGACCAAACGGCCTACAATGCATTTCGGCATAACCACTGGCATGTTCAATACTACGACG
ATGGACGATGTCTACAACTACCTCTTTGATGATGAGGATACACCCCCCAATCCAAAAAAA
GAG

Protein Sequence:

MKLLSSIEQACDICRLKKLKCSKEKPKCAKCLKNNWECRYSPKTKRSPLTRAHLTEVESR
LERLEQLFLLIFPREDLDMILKMDSLQDIKALLTGLFVQDNVNKDAVTDRLASVETDMPL
TLRQHRISATSSSESSNKGQRQLTVSIDSAAHHDNSTIPLDFMPRDALHGFDWSEEDDM
SDGLPFLKTDPNNGFFGDGSLLCILRSIGFKPENYTNSNVNRLPTMITDRYTLASRSTT
SRLQSYLNNFHPYCPIVHSPTLMMLYNNQIEIASKDQWQILFNCILAIGAWCIEGESTD
IDVFYYQNAKSHLTSKVFESGSIILVTALHLLSRYTQWRQKNTSYNFHSFSIRMAISLG
LNRDLPSSFSDSSILEQRRRIWWSVYSWEIQLSLLYGRSIQLSQNTISFPSSVDDVQRTT
TGPTIYHGIIETARLLQVFTKIYELDKTVTAEKSPICAKKCLMICNEIEEVSQAPKFLQ
MDISTTALTNLLKEHPWLSFTRFELKWKQLSLIIYVLRDFFTNFTQKKSQLEQDQNDHQS
YEVKRCSIMLSDAAQRTVMSVSSYMDNHNVTPTYFAWNCSYYLEFNAVLVPIKTLLSNSKSN
AENNETAQLLQQINTVLMMLKKLATFKIQTCCKYIQVLEEVCAPFLLSQCAIPLPHISYN
NSNGSAIKNIVGSATIAQYPTLPEENVNINISVKYVSPGSGVPSVPLKSGASFSDLVKLL
SNRPPSRNSPVTIPRSTPSHRSVTPFLGQQQQQLQSLVPLTPSALFGGANFNQSGNIADSS
LSFTFTNSSNGPNLITTQNTSQALSQPIASSNVHDNFMNNEITASKIDDGNNSKPLSPGW
TDQTAYNAFGITTMGFNTTTMDDVYNYLFDDEDTPPNPKKE



Negative CIS Elements	Negative repeat Elements
3	0

Indices based on non-uniform usage of synonymous codon				
Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.97	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥0	0.08	[41]
DCBS	Directional Codon Bias Score	≥1	1.28	[51]
CDC	Codon Deviation Coefficient	0~1	0.05	[61]
MILC	Measure Independent of Length and Composition	-1~1	-0.55	[71]

ICDI	Intrinsic Codon Deviation Index	0~1	0	[8]
SCUO	Synonymous Codon Usage Order	0~1	0	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.96	[11]
P	Codon Preference	≥ 1	1	[12]
MCB	Maximum-likelihood Codon Bias	≥ 0	0.01	[13]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.62	[1]
CFD	Codon Frequency Distribution	0~1	0.09	[2]
FOP	Frequency of Optimal Codons	0~1	0.33	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	-0.07	[5]
COUSIN18			-0.03	
CBI	Codon Bias Index	-1~1	-0	[6]
Dmean	Mean Dissimilarity-based Index	0~2	0.95	[7]
RCA	Relative Codon Adaptation	≥ 0	1.02	[8]
CUFS	Codon Usage Frequency Similarity	0~ $\sqrt{2}$	0.19	[9]
B	Codon Usage Bias	0~2	0.28	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.44	[1]
gtAI	Genetic tRNA Adaptation Index	0~1	0.52	[2]
P2	P2 Index	0~1	0.51	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	0~1	0.51	[1]
GC	GC Content		0.45	
GC1	GC Content at the First Position of Synonymous Codons		0.45	
GC2	GC Content at the Second Position of Synonymous Codons		0.4	
ENcp	Effective Number of Codon Pairs	20~61	42.82	[2]
CPS	Codon Pair Score	-1~1	0.02	[3, 4]
Codon Volatility	Codon Volatility	0.5~1	0.76	[5]

Gene Name: Ribonuclease Barnase

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: Pichia pastoris (budding yeasts)

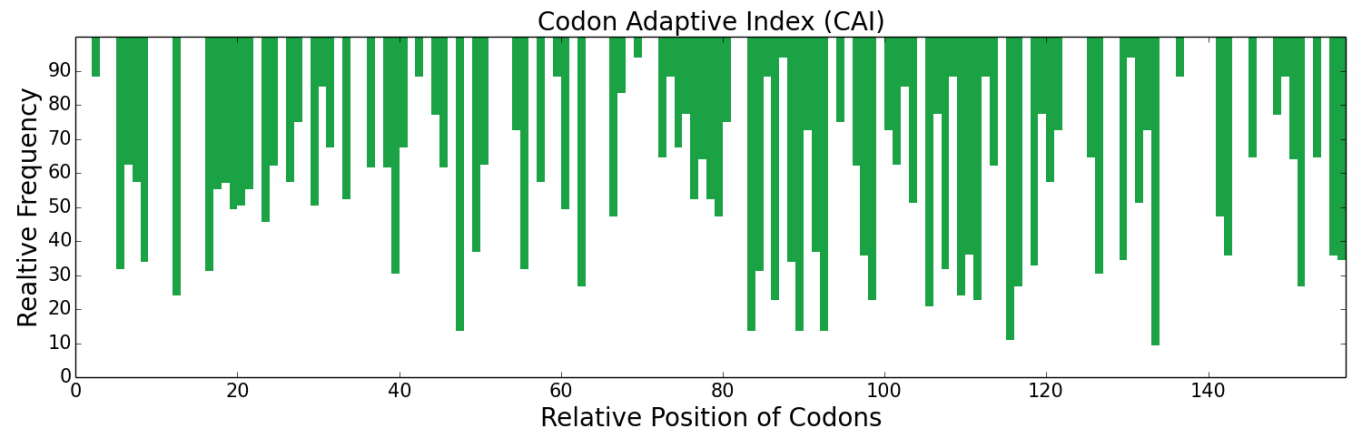
DNA Sequence Length: 471

DNA Sequence:

ATGATGAAAATGGAAGGCATCGCCCTAAAGAAGAGACTCTCTTGGATTAGCGTCTGCTTA
CTTGCTCTTGGTGTCTAGCTGCCGGAATGCTTTTCTCCACTGCAGCTAAGACAGAAACATCG
TCCCATAAAGCTCACACAGAAGCGCAAGTAATCAACACTTTTGACGGCGTTGCCGATTAT
TTACAAACGTACCATAAGCTGCCTGATAATTACATTACCAAATCCGAGGCACAGGCACTG
GGATGGGTTGCGAGCAAAGGGAATCTAGCGGACGTAGCGCCAGGAAAGTCAATAGGGGGT
GACATCTTCAGTAACCGAGAGGGGCAAACCTCCCCGGGAAATCAGGTCGCACGTGGAGGGAG
GCCGACATTAACCTACACCTCGGGTTTTCGTAATAGTGACCGGATTTTGTATTCTTCTGAT
TGGCTGATATACAAGACCACTGATCACTATCAGACGTTTACCAAGATACGT

Protein Sequence:

MMKMEGIALKKRLSWISVCLLVLSAAGMLFSTA AKTETSSHKAHTAQVINTFDGVADY
LQTYHKLPDNYITKSEAQALGWVASKGNLADVAPGKSIGGDIFSNREGKLPKGSGRTWRE
ADINYTSGFRNSDRILYSSDWLIYKTTDHYQTFTKIR



Negative CIS Elements	Negative repeat Elements
0	0

Indices based on non-uniform usage of synonymous codon				
Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.98	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥0	0.13	[41]

DCBS	Directional Codon Bias Score	≥ 1	1.37	[5]
CDC	Codon Deviation Coefficient	0~1	0.07	[6]
MILC	Measure Independent of Length and Composition	-1~1	-0.74	[7]
ICDI	Intrinsic Codon Deviation Index	0~1	0.07	[8]
SCUO	Synonymous Codon Usage Order	0~1	0.02	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.93	[11]
P	Codon Preference	≥ 1	1.03	[12]
MCB	Maximum-likelihood Codon Bias	≥ 0	0.03	[13]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.62	[1]
CFD	Codon Frequency Distribution	0~1	0.1	[2]
FOP	Frequency of Optimal Codons	0~1	0.34	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	0.29	[5]
COUSIN18			0.18	
CBI	Codon Bias Index	-1~1	0.01	[6]
Dmean	Mean Dissimilarity-based Index	0~2	0.7	[7]
RCA	Relative Codon Adaptation	≥ 0	1.02	[8]
CUFS	Codon Usage Frequency Similarity	0~ $\sqrt{2}$	0.25	[9]
B	Codon Usage Bias	0~2	0.28	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.46	[1]
gtAI	Genetic tRNA Adaptation Index	0~1	0.54	[2]
P2	P2 Index	0~1	0.49	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	0~1	0.52	[1]
GC	GC Content		0.48	
GC1	GC Content at the First Position of Synonymous Codons		0.46	
GC2	GC Content at the Second Position of Synonymous Codons		0.44	

ENcp	Effective Number of Codon Pairs	20~61	20	[2]
CPS	Codon Pair Score	-1~1	0.03	[3, 4]
Codon Volatility	Codon Volatility	0.5~1	0.75	[5]

Gene Name: Cas9

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: *Pichia pastoris* (budding yeasts)

DNA Sequence Length: 4104

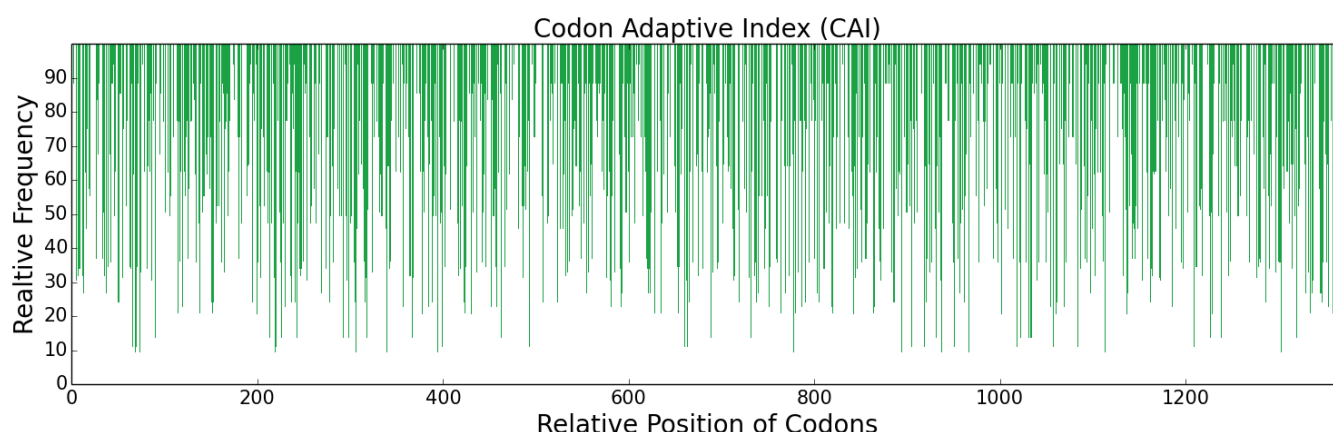
DNA Sequence:

```
ATGGATAAAAAGTACTCGATTGGCCTAGATATTGGCACGAACTCAGTGGGATGGGCCGTC
ATTACTGATGAATACAAGGTACCTTCCAAAAAGTTTAAGGTACTTGGCAACACGGATCGT
CATTCCATAAAAAAAATTTAATCGGTGCTCTCCTCTTCGATAGCGGAGAACTGCAGAG
GCTACTCGTCTAAAACGCACAGCTCGGCGCCGTTATACTCGGAGGAAGAATAGAATCTGT
TACCTACAGGAGATCTTCTCGAACGAGATGGCGAAAGTTGATGATTCCTTCTTTCACAGA
CTTGAAGAATCATTTTTAGTCGAAGAAGACAAGAAGCATGAGCGACACCCCATCTTTGGG
AACATCGTCGACGAGGTAGCCTACCACGAGAAGTATCCTACAATATACCATCTGCGAAAA
AAGCTTGTGCGATTCCACAGACAAGGCCGACCTCCGACTCATTTACCTGGCATTGGCACAC
ATGATAAAATTCAGGGGACACTTCTTAATCGAGGGTGATTTGAATCCTGATAACTCTGAC
GTAGACAAGCTGTTTATTCAATTGGTCCAGACCTATAATCAACTCTTCGAGGAAAATCCG
ATAAACGCAAGCGGAGTTGACGCAAAGGCTATCCTGTGACGCGAGACTGTCTAAAAGCCGG
CGCTTGGAGAACCTTATAGCGCAACTGCCAGGGGAGAAAAAGAATGGCCTCTTCGGAAT
CTCATCGCGCTGTCACTAGGCCTAACTCCCAATTTCAAAATCGAACTTTGACCTGGCAGAA
GATGCCAAACTGCAATTGTCCAAAGATACGTACGATGATGATCTAGACAACCTCTTAGCT
CAGATAGGAGATCAATACGCCGATTTATTTTTAGCTGCGAAGAATTTATCGGATGCGATT
CTGTTATCTGATATACTCCGGGTGAATACAGAGATTACCAAGGCCCTCAGCGCGAGT
ATGATTAAAAGGTATGACGAACATCACCAAGACCTTACATTATTAAGCTTTGGTGCGG
CAGCAGCTACCCGAGAAGTATAAGGAAATCTTTTTTGACCAATCAAAGAATGGGTATGCT
GGATATATTGACGGCGGTGCGAGCCAAGAAGAATTCTACAAGTTCATCAAGCCGATCCTT
GAGAAAATGGACGGGACCGAAGAGCTCTTAGTCAAATTAACCGGGAAGACCTGCTTCGC
AAGCAGCGTACATTTGATAACGGTAGTATTCCACACCAAATTCATCTGGGCGAGCTACAC
GCAATTCTCCGACGTGAGGAAGATTTCTACCCGTTCTTAAAGACAACAGGGAAAAAATT
GAGAAAATACTTACATTCCGTATTCTTACTACGTTGGGCCTTTAGCCCGTGGGAATTCA
AGATTCGCGTGATGACTCGTAAAAGTGAAGAAACAATTACTCCTTGGAATTTTGAAGAA
GTTGTGGACAAGGGTGCAAGCGCACAGAGTTTTATCGAGCGCATGACTAATTTTGACAAA
AACTTGCCAAACGAAAAGGTCTCCCAAAGCATTTCTCTGTTATACGAGTATTTACCCGTT
TATAACGAACTCACTAAGGTGAAATATGTGACAGAGGGCATGAGGAAACCCGCTTTTTTA
AGTGAGAGAGCAGAAGAAAGCAATCGTAGATTTGCTGTTCAAAACGAACAGGAAAGTCACC
GTAAAACAATTGAAAGAGGACTATTTCAAAAAGATCGAGTGCTTCGATAGCGTCGAAATC
TCGGGGGTCGAGGATAGGTTTAAACGCTTCCCTAGGGACGTATCATGACCTTCTTAAAATC
ATAAAGACAAGGACTTCTTGATAATGAAGAGAATGAAGACATTTTAGAAGACATCGTA
CTAACGCTAACGCTTTTTGAAGATCGAGAAATGATCGAAGAGCGACTGAAGACATACGCT
CATCTTTTCGATGACAAGGTTATGAAGCAGCTTAAGCGTCGACGTTATACAGGATGGGGT
CGCTTGTCGCGCAAGCTCATCAATGGTATAAGAGATAAGCAGTCCGGCAAAACTATCCTA
GATTTTTTGAAATCAGACGGATTTCGCGAACAGGAATTTTCATGCAGCTAATACACGATGAT
TCTCTGACCTTTAAAGAAGACATCCAGAAAGCCCAAGTGTGCGGCCAAGGTGATAGTCTC
CATGAACACATTGCCAATCTTGCTGGTAGCCAGCGATCAAAAAGGGCATACTCCAAACG
GTCAAGGTTGTGATGAGCTAGTGAAAGTCATGGGGCGTCAAAAGCCAGAAAACATTGTA
ATAGAAATGGCTCGAGAGAACCAGACGACTCAGAAGGGCCAGAAGAATTCCCGGGAGCGA
ATGAAGAGGATTGAGGAGGGGATCAAGGAGCTCGGATCTCAAAATATTGAAAGAGCATCCT
GTAGAGAACACGCAACTCCAAAACGAGAACTATACCTTTATTATCTGCAGAACGGGCGA
GACATGTATGTGGACCAAGAAGTGGATATTAATAGACTGTCTGATTACGACGTGGATCAT
```

ATCGTCCCGCAGAGCTTCTCTAAAGGACGATAGTATAGACAACAAAGTACTGACCAGATCA
GATAAGAACCGAGGGAAATCAGACAACGTGCCCTCGGAAGAAGTGGTTAAAAAAATGAAA
AATTATTGGAGGCAACTCCTCAATGCCAAGTTAATCACACAACGGAAGTTTGACAACCTT
ACCAAGGCAGAGCGCGGTGGACTGTCTGAGCTTGATAAGGCTGGTTTCATTAAAGCGCCAG
CTCGTAGAAACCAGGCAAATAACTAAACATGTGGCGCAGATTCTTGACTCGCGGATGAAC
ACTAAATATGATGAGAACGATAAGCTGATACGCGAGGTTAAAGTAATTACGCTGAAAAGT
AAGCTCGTCTCCGATTTTCGGAAAGACTTTCAATTCTACAAGGTCAGAGAGATAAATAAT
TATCATCACGCACACGATGCCTATTTGAATGCTGTTGTAGGTACTGCCCTGATAAAGAAA
TACCCGAAATTGGAAAGTGAATTTGTGTACGGTGACTATAAAGTATACGACGTTTCGCAAA
ATGATAGCGAAATCTGAACAAGAGATAGGAAAGGCGACCGCGAAATACTTCTTTTATTTCG
AATATAATGAATTTTTTCAAAACAGAGATAACACTTGCTAACGGGGGAAATTCGCAAAAGA
CCGCTCATTGAAACCAACGGGAGAAACCGGGGAGATAGTTTGGGACAAAGGTAGAGACTTT
GCTACCGTTTCGAAGGTTCTAAGTATGCCACAGGTAAATATAGTCAAAAAGACGGAGGTG
CAAACAGGCGGATTTTCAAAAGAATCAATTTTACCCAAACGGAATTCTGATAAGCTTATT
GCTAGAAAAAAGGATTGGGACCCAAAGAAATATGGAGGCTTCGACAGTCCGACGGTTGCA
TATTCCGTACTGGTGGTAGCAAAAGTGAAAAAGGTAAGTCGAAAAAATTGAAATCAGTA
AAAGAACTCCTAGGCATCACAATCATGGAGAGAAGCTCGTTTGAGAAGAATCCTATCGAC
TTTTTAGAAGCCAAGGGGTATAAGGAGGTGAAAAAGGACTTGATTATAAAACTACCTAAG
TACTCTCTTTTCGAGTTGGAATAAGTTCGCAACGTATGTTGGCAAGTGCCGGCGAATTG
CAAAAGGGTAATGAATTAGCGCTTCCGTCCAAATACGTTAACCTTTTTATATTTGGCGTCT
CACTACGAGAACTTAAGGGCTCTCCCGAGGACAATGAGCAGAAGCAGTTATTTGTCTGAG
CAACATAAGCATTACTTAGACGAGATAATTGAGCAAAATTTCCGAATTTAGCAAGCGTGTA
ATATTGGCTGACGCCAATCTCGATAAGGTCTTATCCGCATATAACAAACATAGGGACAAG
CCCATACGGGAACAGGCCGAAACATCATCCACCTGTTTACCCTAACCAACTTAGGAGCG
CCAGCAGCCTTCAAATACTTTGATACCACGATTGACAGGAAGCGATACACGTCGACTAAA
GAAGTTTTGGATGCCACGCTCATAACCAATCCATAACCGGGTTGTACGAGACTCGAATC
GATCTTTCACAGCTAGGGGGTGAC

Protein Sequence:

MDKKYSIGLDIGTNSVGVAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFILQVQTYNQLFEEPNINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLKALVRQQLPKEYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAMWTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGYHDLKLI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAICKGILQTVKVVDLVKVMGRHKPENIVIMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVBREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSPQVNIKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVE
QHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDTLHQSIITGLYETRIDLSQLGGD



Negative CIS Elements	Negative repeat Elements
6	0

Indices based on non-uniform usage of synonymous codon

Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.98	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥ 0	0.1	[41]
DCBS	Directional Codon Bias Score	≥ 1	1.34	[51]
CDC	Codon Deviation Coefficient	0~1	0.06	[61]
MILC	Measure Independent of Length and Composition	-1~1	-0.53	[71]
ICDI	Intrinsic Codon Deviation Index	0~1	0	[81]
SCUO	Synonymous Codon Usage Order	0~1	0	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.98	[111]
P	Codon Preference	≥ 1	1	[121]
MCB	Maximum-likelihood Codon Bias	≥ 0	0	[131]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.64	[11]
CFD	Codon Frequency Distribution	0~1	0.09	[21]
FOP	Frequency of Optimal Codons	0~1	0.35	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	-0.05	[51]
COUSIN18			0	
CBI	Codon Bias Index	-1~1	-0	[61]
Dmean	Mean Dissimilarity-based Index	0~2	0.69	[71]

RCA	Relative Codon Adaptation	≥ 0	1.04	[8]
CUFS	Codon Usage Frequency Similarity	$0 \sim \sqrt{2}$	0.18	[9]
B	Codon Usage Bias	$0 \sim 2$	0.26	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	$0 \sim 1$	0.48	[1]
gtAI	Genetic tRNA Adaptation Index	$0 \sim 1$	0.56	[2]
P2	P2 Index	$0 \sim 1$	0.5	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	$0 \sim 1$	0.5	[1]
GC	GC Content		0.43	
GC1	GC Content at the First Position of Synonymous Codons		0.51	
GC2	GC Content at the Second Position of Synonymous Codons		0.29	
ENcp	Effective Number of Codon Pairs	$20 \sim 61$	50.2	[2]
CPS	Codon Pair Score	$-1 \sim 1$	0.01	[3] , [4]
Codon Volatility	Codon Volatility	$0.5 \sim 1$	0.77	[5]

Gene Name: Toxin CcdB

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: Pichia pastoris (budding yeasts)

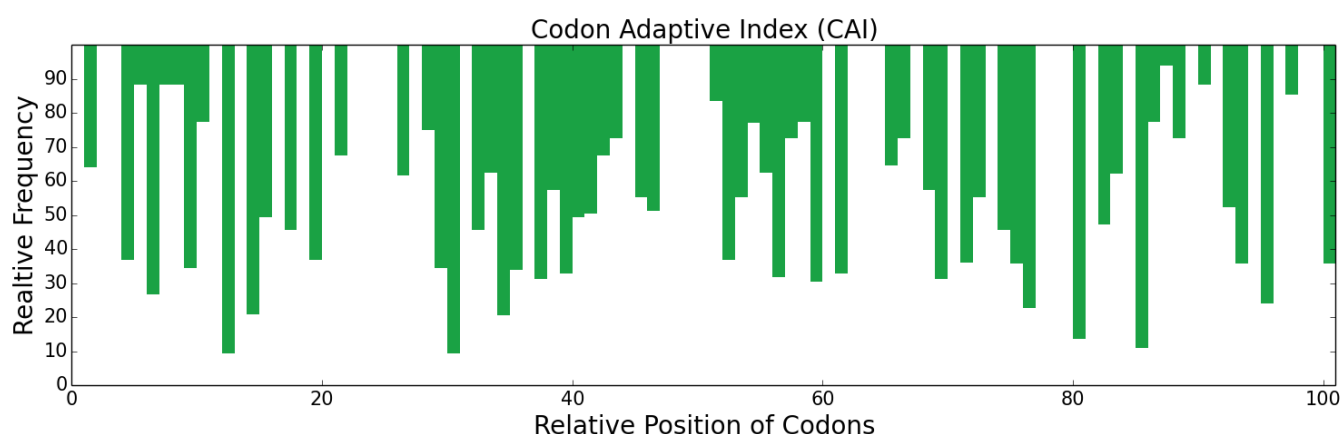
DNA Sequence Length: 303

DNA Sequence:

ATGCAGTTTAAAGGTATATACGTATAAACGTGAGTCTCGGTACCGATTATTTGTGGATGTA
 CAATCCGATATTATTGATACACCAGGACGTCGGATGGTGATCCCGCTAGCTAGCGCCAGG
 TTACTTTCCGACAAGGTCAGTAGAGAATTGTACCCTGTAGTCCACATCGGCGACGAGTCG
 TGGAGGATGATGACTACCGACATGGCCAGCGTTCCCGTCTCTGTGATAGGGGAAGAAGTT
 GCGGATCTGTACATCGCGAGAATGACATTAAAAACGCAATAAACCTCATGTTCTGGGGT
 ATA

Protein Sequence:

MQFKVYTYKRESRYRLFVDVQSDIIDTPGRRMVIPLASARLLSDKVSRELYPVVHIGDES
 WRMMTTDMASVPVSVIGEEVADLSHRENDIKNAINLMFWGI



Negative CIS Elements	Negative repeat Elements
0	0

Indices based on non-uniform usage of synonymous codon

Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.97	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥0	0.15	[41]
DCBS	Directional Codon Bias Score	≥1	1.5	[51]
CDC	Codon Deviation Coefficient	0~1	0.08	[61]

MILC	Measure Independent of Length and Composition	-1~1	-0.87	[7]
ICDI	Intrinsic Codon Deviation Index	0~1	0.02	[8]
SCUO	Synonymous Codon Usage Order	0~1	0.02	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.9	[11]
P	Codon Preference	≥ 1	1.03	[12]
MCB	Maximum-likelihood Codon Bias	≥ 0	0.04	[13]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.58	[1]
CFD	Codon Frequency Distribution	0~1	0.09	[2]
FOP	Frequency of Optimal Codons	0~1	0.32	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	0.43	[5]
COUSIN18			0.88	
CBI	Codon Bias Index	-1~1	0	[6]
Dmean	Mean Dissimilarity-based Index	0~2	0.91	[7]
RCA	Relative Codon Adaptation	≥ 0	0.98	[8]
CUFS	Codon Usage Frequency Similarity	0~ $\sqrt{2}$	0.21	[9]
B	Codon Usage Bias	0~2	0.34	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.45	[1]
gtAI	Genetic tRNA Adaptation Index	0~1	0.54	[2]
P2	P2 Index	0~1	0.48	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	0~1	0.53	[1]
GC	GC Content		0.47	
GC1	GC Content at the First Position of Synonymous Codons		0.51	
GC2	GC Content at the Second Position of Synonymous Codons		0.37	
ENcp	Effective Number of Codon Pairs	20~61	20	[2]
CPS	Codon Pair Score	-1~1	-0.01	[3, 4]

Codon Volatility	Codon Volatility	0.5~1	0.75	[5]
---------------------	------------------	-------	------	---------------------

Gene Name: T7 RNA Polymerase

Reference Source: Codon Usage Database - Kazusa

Expression Host Organism: *Pichia pastoris* (budding yeasts)

DNA Sequence Length: 2649

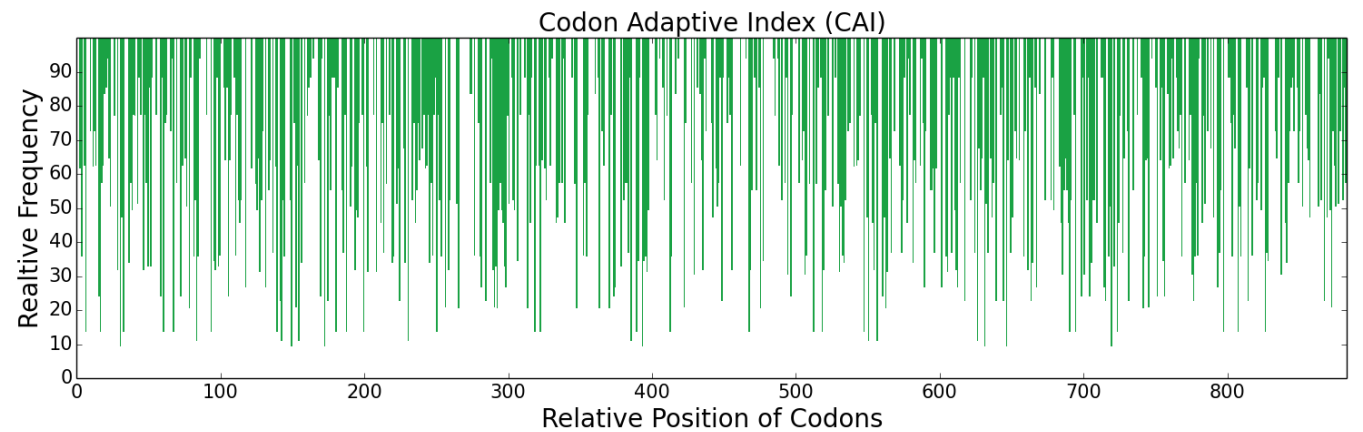
DNA Sequence:

```
ATGAACACAATAAACATCGCGAAGAACGACTTTTCAGACATCGAACTCGCGGCCATCCCT
TTCAATACCCTTGCTGATCACTACGGCGAACGGCTGGCGAGAGAACAACCTAGCCTTAGAG
CACGAAAGTTATGAGATGGGCGAGGCCAGGTTTCAGGAAAAATGTTTGAGAGACAACCTCAA
GCGGGAGAGGTTGCTGACAATGCGGCTGCCAAGCCACTCATCACTACCCCTTTTGCCGAAA
ATGATAGCACGCATAAATGATTGGTTTGAAGAGGTTAAGGCGAAGCGTGGCAAAAGGCC
ACTGCTTTCCAGTTCCTCCAGGAGATTAAACCCGAGGCAGTGGCATAATTACGATTAAAG
ACTACATTGGCTTGCTTAACAGCGCAGACAACACGACTGTCCAGGCTGTAGCTTCAGCG
ATTGGGCGCGCAATAGAAGATGAAGCACGGTTTGGACGAATCCGCGATCTAGAAGCCAAG
CACTTCAAAAAGAATGTTGAAGAACAGCTCAATAAGCGGGTTGGGCACGTCTATAAAAAA
GCGTTTCATGCAAGTCGTAGAGGCGGATATGCTTTCAAAGGGCTTACTGGGAGGTGAGGCG
TGGTCAAGCTGGCATAAGGAGGATAGCATTTCATGTGGGAGTAAGATGCATTGAGATGCTA
ATAGAAAGTACAGGGATGGTTTCCCTACATCGCCAAAAACGCAGGAGTGGTCGGACAGGAT
TCCGAGACAATCGAGCTAGCCCCCGAGTATGCGGAGGCAATAGCTACTCGAGCTGGCGCA
TTGGCTGGTATTAGTCCGATGTTTCAACCATGTGTTGTTCTCCTAAGCCCTGGACTGGA
ATAACGGGTGGTGGGTACTGGGCCAATGGCCGAAGGCCGTTAGCCTTAGTGAGGACGCAC
AGTAAGAAAGCATTAAATGCGTTACGAGGATGTTTATATGCCGGAGGTGTATAAGGCGATC
AACATCGCGCAGAACACAGCATGGAAAATCAATAAGAAGGTGCTGGCCGTTGCCAATGTG
ATTACTAAGTGGAACATTGCCCCGTTGAAGATATTCCCGCCATAGAGAGAGAAGAATTG
CCTATGAAACCCGAAGACATCGATATGAACCCGGAGGCTCTCACGGCTTGGAAGAGGGCT
GCAGCCGCCGTATATCGCAAGGATAAGGCGCGTAAATCTCGGCGTATAAGCTTAGAATTT
ATGTTGGAGCAGGCTAATAAGTTCGCAAAACCACAAGGCGATATGGTTTCCTTACAATATG
GATTGGCGAGGAAGAGTTTACGCCGTTTCGATGTTCAACCTCAGGGCAATGACATGACT
AAGGGACTGTTGACACTTGCCAAAGGGAAACCAATTGGAAGAGGGCTACTACTGGTTG
AAGATCCATGGTGCTAATTGTGCGGGCGTCGATAAAGTCCCATTCCCGGAACGTATTAAG
TTTATTGAAGAAAATCACGAAAATATCATGGCATGTGCCAAATCGCCACTCGAGAACACT
TGGTGGGCCGAACAGGACTCGCCCTTTTGCTTTTGGCGTTTGTCTCGAATATGCGGGC
GTCCAACACCATGGTCTTTCTTATAACTGCAGCCTTCCCCTAGCATTGACGGATCTTGT
TCAGGTATCCAGCACTTTTCTGCGATGCTGCGCGATGAGGTGGGCGGACGCGCTGTTAAC
CTCCTGCCGAGCGAAACCGTACAAGACATTTACGGTATCGTAGCAAAAAAGGTGAATGAG
ATTCTACAGGCCGATGCTATCAATGGAACGGACAACGAAGTCGTCACAGTAACAGATGAA
AACACGGGTGAAATAAGCGAGAAAGTAAATTAGGCACGAAAGCCTTGGCTGGGCAATGG
TTGGCATATGGTGTAACGCTCCGTCACCAACGGAGTGAATGACCCTGGCCTACGGG
TCTAAGGAGTTCGGGTTTCGGCAGCAAGTACTGGAAGACACCATTCAGCCAGCTATAGAC
TCTGGCAAAGGGCTAATGTTACGCAACCTAACCAAGCTGCAGGTTACATGGCAAAATTA
ATTTGGGAATCAGTGTGCGTACCCGTCGTGGCGGCAGTTGAGGCGATGAACTGGCTCAA
TCGGCTGCAAACTCCTAGCAGCAGAAGTGAAGGATAAAAAACGGGGGAAATACTTCGG
AAGAGGTGTGCGGTGCACTGGGTAACCCCGAGACGGGTTTCCAGTCTGGCAAGAGTACAAG
AAACCGATACAGACCCGATTGAATTTGATGTTTCTCGGTCAGTTCCGTCTCCAACCTACA
ATCAATACCAACAAAGACTCCGAGATAGATGCCCATAAAGCAGGAGTCGGGGATAGCCCC
AATTTTGTGCACAGTCAAGACGGTTCCCATCTGAGAAAGACGGTAGTCTGGGCGCATGAA
AAATATGGAATAGAGTCCTTTGCGCTTATACATGATTCTTTTCGGGACAAATCCCGCTGAT
GCAGCCAACCTATTTCAAAGCGGTACGTGAAACTATGGTTGACACCTATGAATCGTGTGAT
```

GTGCTAGCCGACTTCTATGACCAATTCGCCGACCAACTTCATGAGTCCCAGCTGGATAAG
ATGCCAGCTCTTCCTGCAAAGGGGAACCTGAATTTACGAGACATCCTTGAAAGTGACTTT
GCATTCGCC

Protein Sequence:

MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRKMFERQLK
AGEVADNAAAKPLITTLPLPKMIARINDWFEEVKAKRGKRPTAFQFLQEIKPEAVAYITIK
TTLACLTADNTTVQAVASAI GRAIEDEARFGRIRDLEAKHFKNVEEQLNKRVGHVYKK
AFMQVVEADMLSKGLLGGEAWSSWHKEDSIHVGVRCIEMLIESTGMVSLHRQNAGVVGQD
SETIELAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTH
SKKALMRYEDVYMPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPVEDIPAIEREEL
PMKPEDIDMNPEALTAWKRAAAAVYRKDKARKSRRISLEFMLEQANKFANHKAIWFPYNM
DWRGRVYAVSMFNPQGNDMTKGLLTAKGKPIGKEGYWLKIHGANCAGVDKVPFPERIK
FIEENHENIMACAKSPLENTWAEQDSPFCFLAFCFEYAGVQHHGLSYNCSLPLAFDGSC
SGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNEILQADAINGTDNEVTVTDE
NTGEISEKVKLGTKALAGQWLAYGVTRSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAID
SGKGLMFTQPNQAAGYMAKLIWESVSVTVVAAVEAMNWLKSAAKLLAAEVKDKKTGEILR
KRCVHVVTPDGFVPVQYKKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAP
NFVHSQDGSHLRKTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESCD
VLADFYDQFADQLHESQLDKMPALPAKGNLNLRDILESDFAF



Negative CIS Elements	Negative repeat Elements
1	0

Indices based on non-uniform usage of synonymous codon				
Index	Description	Range	Value	Reference
RSCU	Relative Synonymous Codon Usage	0~6	0.95	[11]
ENC	Effective Number of Codons	20~61	61	[2, 3]
RCBS	Relative Codon Bias Strength	≥0	0.08	[41]
DCBS	Directional Codon Bias Score	≥1	1.3	[51]
CDC	Codon Deviation Coefficient	0~1	0.06	[61]
MILC	Measure Independent of Length and Composition	-1~1	-0.55	[71]

ICDI	Intrinsic Codon Deviation Index	0~1	0	[8]
SCUO	Synonymous Codon Usage Order	0~1	0	[9, 10]
Ew	Weighted Sum of Relative Entropy	0~1	0.95	[11]
P	Codon Preference	≥ 1	1	[12]
MCB	Maximum-likelihood Codon Bias	≥ 0	0	[13]

Indices based on codon frequency in a reference set of genes

Index	Description	Range	Value	Reference
CAI	Codon Adaptation Index	0~1	0.61	[1]
CFD	Codon Frequency Distribution	0~1	0.11	[2]
FOP	Frequency of Optimal Codons	0~1	0.34	[3, 4]
COUSIN59	Codon Usage Similarity Index	∞	-0.01	[5]
COUSIN18			-0.02	
CBI	Codon Bias Index	-1~1	-0	[6]
Dmean	Mean Dissimilarity-based Index	0~2	0.83	[7]
RCA	Relative Codon Adaptation	≥ 0	1	[8]
CUFS	Codon Usage Frequency Similarity	0~ $\sqrt{2}$	0.19	[9]
B	Codon Usage Bias	0~2	0.28	[10]

Indices based on adaptation to the tRNA levels and their supply

Index	Description	Range	Value	Reference
tAI	tRNA Adaptation Index	0~1	0.46	[1]
gtAI	Genetic tRNA Adaptation Index	0~1	0.54	[2]
P2	P2 Index	0~1	0.49	[3]

Indices based on complex patterns of codon usage

Index	Description	Range	Value	Reference
GC3	GC Content at the Third Position of Synonymous Codons	0~1	0.52	[1]
GC	GC Content		0.49	
GC1	GC Content at the First Position of Synonymous Codons		0.55	
GC2	GC Content at the Second Position of Synonymous Codons		0.39	
ENcp	Effective Number of Codon Pairs	20~61	61	[2]
CPS	Codon Pair Score	-1~1	0.03	[3, 4]
Codon Volatility	Codon Volatility	0.5~1	0.76	[5]

