

(最終更新日 2019 年 11 月 26 日)

# Codon Usage Generator

## ver 2.3

# マニュアル

柳田駿介, 阿部貴志

新潟大学 工学部 情報工学科

問い合わせ先: [takaabe@ie.niigata-u.ac.jp](mailto:takaabe@ie.niigata-u.ac.jp)

## 目次

1. 概要 .....	3
2. 動作環境 .....	3
3. 基本操作手順 .....	3
3.1 起動と終了 .....	4
3.2 入力ファイルの選択 .....	5
3.3 計算結果の表示 .....	5
3.4 ファイルへ保存 .....	7
3.4.1 「All Codon」でのファイル保存 .....	7
3.4.2 「Codon Table」でのファイル保存 .....	9
4. 一括処理についての操作手順 .....	11
4.1 入力フォルダの選択 .....	11
4.2 出力フォルダの選択 .....	13
4.3 出力ファイルの種類選択 .....	14
5. 計算項目について .....	15
6. 参考文献 .....	16

## 1. 概要

本アプリケーションは、ユーザーが持っている遺伝子領域の塩基配列のマルチ FASTA 形式のデータセットについてローカルでより簡単にコドン組成を計算することができることを目的として作成されました。GUI で操作ができ、計算結果を表形式（CSV 形式）で簡単に取得することができます。

## 2. 動作環境

ダウンロードしたファイル「CUG.jar (Codon Usage Generator)」は、コドン組成を計算する GUI アプリケーションプログラムです。

本アプリケーションは、簡単に使用できることを目的としております。よって、お使いの PC の OS に関係なく動作しますが、Java で開発しておりますので、Java 実行環境である JRE(Java Runtime Environment)( $\geq$ Version 8)のインストールが必須となります。

## 3. 基本操作手順

本アプリケーションの操作方法について説明します。まず、簡単な操作の流れを図 1 に示します。

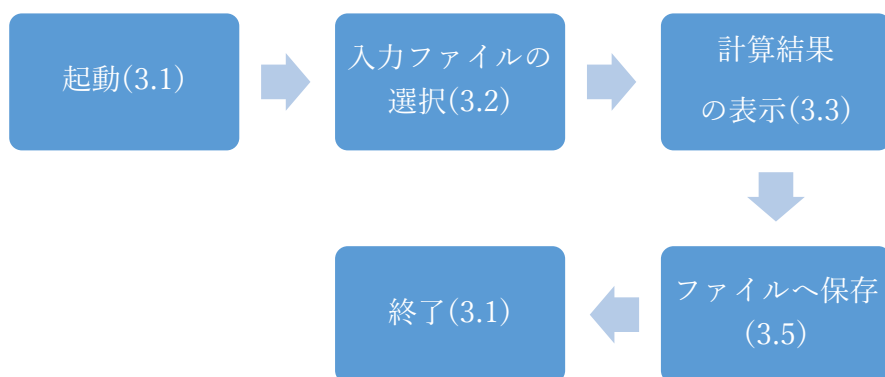


図 1 コドン組成計算システムの操作の流れ

この操作の流れに沿って操作の説明を行っていきます。

### 3.1 起動と終了

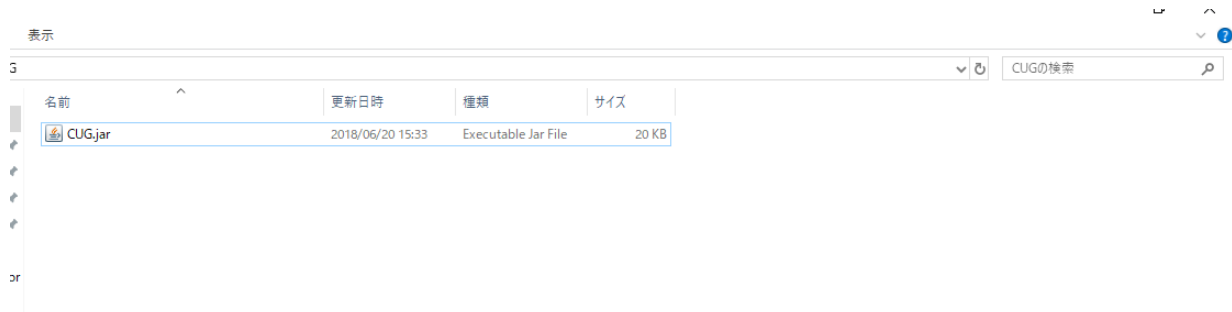


図 2 任意のフォルダにダウンロードした CUG.jar

図 2 は任意のフォルダに CUG.jar をダウンロードした画面です。この CUG.jar をダブルクリックするとコドン組成計算アプリケーションが起動します。



図 3 システム起動後の画面

図 3 はシステム起動直後の画面です。  
終了する場合は、ウインドウの閉じるボタンをクリックします。

### 3.2 入力ファイルの選択

入力ファイルを選択する時は「ファイル選択」ボタンをクリックします。

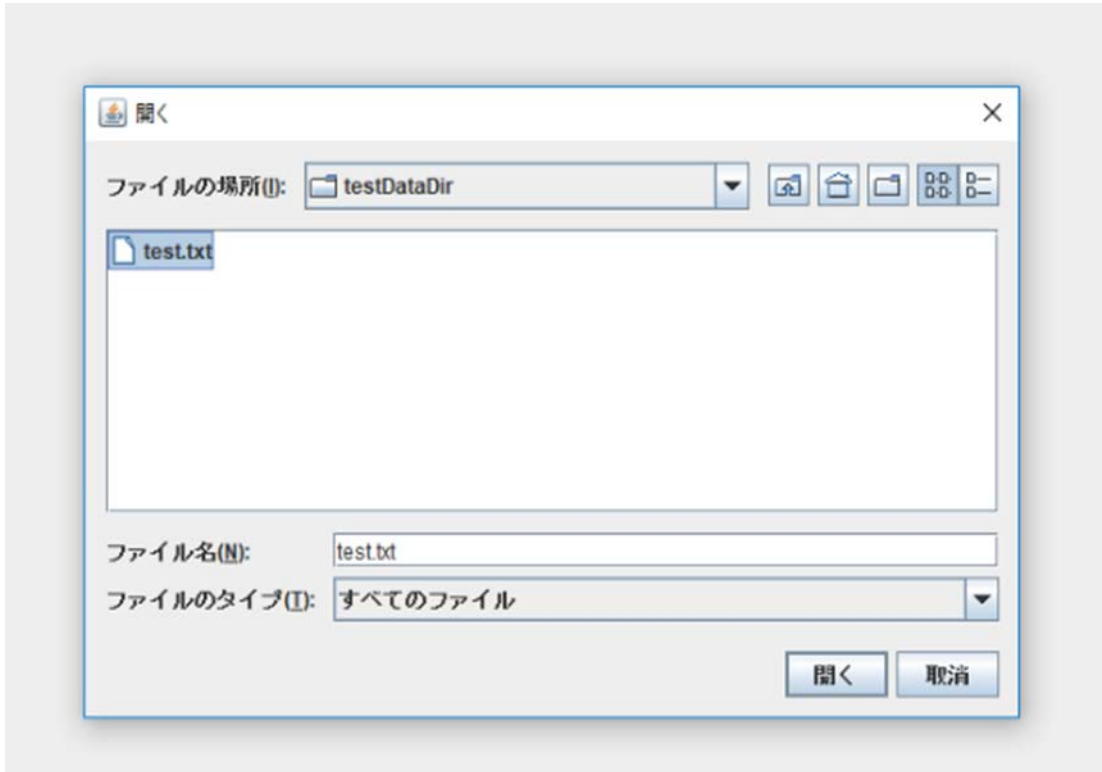


図 4 ファイル選択のダイアログ

図 4 は「ファイル選択ボタンをクリックすると図 3 の画面に表示されるファイル選択を行うダイアログです。このダイアログを用いてファイル選択を行います。

注意:本アプリケーションは遺伝子領域の塩基配列のマルチ fasta 形式のファイルのみを入力ファイルとして想定しています。どのような拡張子 (.txt, .ffn, .fasta など) も読み込み可能です。

### 3.3 計算結果の表示

図 4 のようにファイルを選択したら、ダイアログの右下にある「開く」ボタンをクリックします。そうすると計算が始まり、計算結果がタブ形式で表示されます。計算結果を表示するタブは「All Codon」と「Codon Table」の二つがあります。

タブ 1 「All Codon」  
タブ 2 「Codon Table」

Type	Count	percent	per thousand	RF	RSCU
Total CDS	12				
Total Codon	3494				
1st letter GC	2046	58.56			
2st letter GC	1421	40.67			
3st letter GC	2066	59.13			
TTT	84	2.40	24.04	0.55	1.09
TTC	70	2.00	20.03	0.45	0.91
TTA	38	1.09	10.88	0.09	0.57
TTG	57	1.63	16.31	0.14	0.85
TCT	27	0.77	7.73	0.13	0.81
TCC	39	1.12	11.16	0.19	1.16
TCA	22	0.63	6.30	0.11	0.66

図 5 タブ「All Codon」における計算結果の出力画面

図 5 で画面に表示されているのは「All Codon」というタブで、このタブでは、入力中の全てのコドンに対する計算結果が表形式で表示されます。表示される項目は、「Count」、「percent」、「per thousand」、「RF」、「RSCU」の 5 項目です。各項目については、第 4 章で説明します。

次に、「Codon Table」に切り替えると図 6 の画面になります。

計算項目切替ボタン

Count	2nd letter	3rd letter	amino acid
Phe	TTT	84	Ser
Phe	TTC	70	Ser
Leu	TTA	38	Ser
Leu	TTG	57	Pro
Leu	CTT	39	Pro
Leu	CTC	45	Pro
Leu	CTA	9	Pro
Leu	CTG	213	Pro
Ile	ATT	113	Thr
Ile	ATC	94	Thr
Ile	ATA	8	Thr
Met	ATG	98	Thr
Val	GTT	65	Ala
Val	GTC	48	Ala
Val	GTA	32	Ala
Val	GTG	83	Ala

図 6 タブ「Codon Table」における計算結果の出力画面

図 6 は、タブを「Codon Table」に切り替えた時に表示される画面です。この画面では、計算結果をコドン表で見ることが出来ます。表示される値の項目は、コドン表上部に位置している図 6 の赤枠が示している「計算項目切替ボタン」(「Count」、「percent」、「per thousand」、「RF」、「RFCU」) を押して表示する項目を切り替えます。

### 3.4 ファイルへ保存

本システムでは、以下の三種類の出力をファイルに保存することが出来ます。

- ① 全コドンに対する計算結果
- ② 遺伝子毎の計算結果
- ③ コドン表に出力した計算結果

必要に応じて各結果をファイルに保存してください。

#### 3.4.1 「All Codon」でのファイル保存

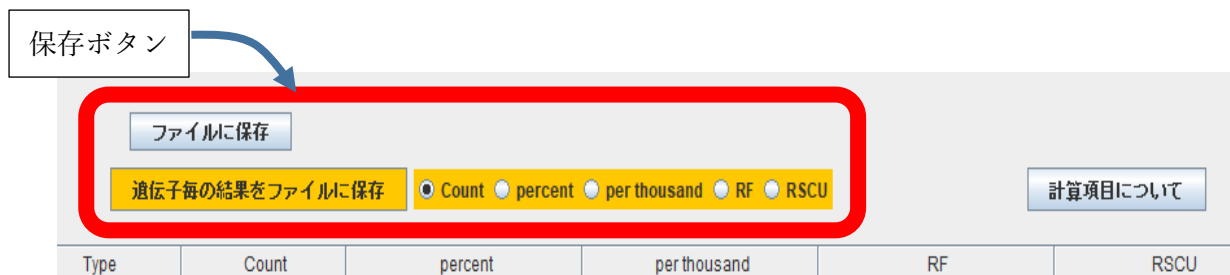


図 7 「All Codon」における保存ボタン

図 7 の赤枠部分は「All Codon」におけるファイルへの保存ボタンです。このタブでは、「①全コドンに対する計算結果」と「②遺伝子毎の計算結果」をファイルに保存することが出来ます。

「①全コドンに対する計算結果」をファイルに保存したい場合は、「ファイルに保存」ボタンをクリックします。

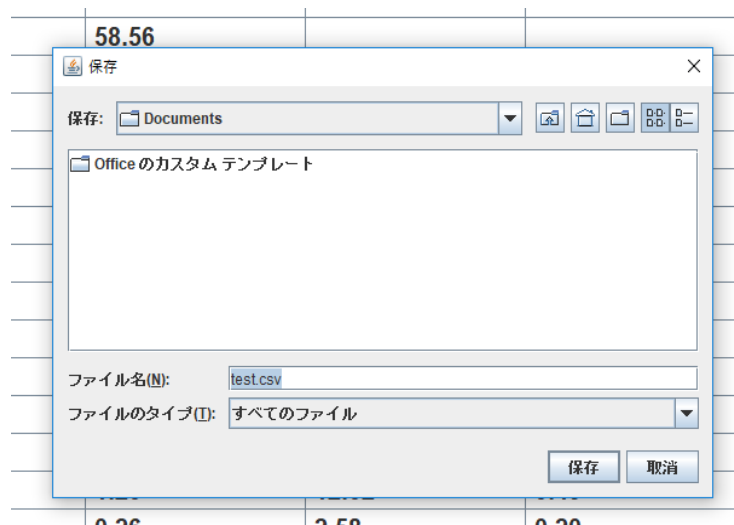


図 8 ファイル保存ダイアログ

図 8 は「ファイルに保存」ボタンをクリックすると表示されるダイアログです。このダイアログで保存する場所とファイル名を決めて保存ボタンをクリックすると保存が完了します。出力ファイルはカンマ区切りとなっています。

注：ファイル名は拡張子もつけて下さい。デフォルトでは「test.csv」となっています。

図 9 は全コドンに対する計算結果を保存したファイルの例です。

test.txt						
	A	B	C	D	E	F
1	test.txt					
2	Type	Count	percent	per thousand	RF	RSCU
3	Total CDS	12				
4	Total codon	3494				
5	1st letter C	2046	58.56			
6	2st letter C	1421	40.67			
7	3st letter C	2066	59.13			
8	TTT	84	2.4	24.04	0.55	1.09
9	TTC	70	2	20.03	0.45	0.91
10	TTA	38	1.09	10.88	0.09	0.57
11	TTG	57	1.63	16.31	0.14	0.85
12	TCT	27	0.77	7.73	0.13	0.81
13	TCC	39	1.12	11.16	0.19	1.16
14	TCA	22	0.63	6.3	0.11	0.66
15	TCG	31	0.89	8.87	0.15	0.93
16	TAT	44	1.26	12.59	0.51	1.02
17	TAC	42	1.2	12.02	0.49	0.98
18	TAA	9	0.26	2.58	0.2	0.39
19	TAG	0	0	0	0	0
20	TCT	10	0.54	5.44	0.1	0.10

図 9 全コドンに対する計算結果を保存した csv ファイルの例



続いて、「②遺伝子毎の計算結果」をファイルに保存したい場合は、「ファイルに保存」の右側にある「遺伝子毎の結果をファイルに保存」ボタンをクリックします。その際、右側にあるラジオボタンで保存するファイルに出力する計算項目の値を切り替えることが出来ます。(デフォルトでは「Count」が選択された状態となっています。)「遺伝子毎の結果をファイルに保存」ボタンをクリックして以降の操作は、「①全コドンに対する計算結果」をファイルに保存する時と同様です。出てきたダイアログで保存する場所と保存するファイル名を決めて、保存ボタンをクリックします。

図 10 は、出力する値に「Count」を選択して、遺伝子毎の計算結果を保存したファイルの例です。

1	test.txt	Count															
2	tag	Total Count	1st letter	2nd letter	3rd letter	TTT	TTC	TTA	TTG	TCT	TCC	TCA	TCG	TAT	TAC	TAA	TAG
3	>gil48994	22	14	15	5	0	0	0	0	0	0	0	0	0	0	0	0
4	>gil48994	821	473	331	503	11	19	10	13	11	10	6	9	12	8	0	
5	>gil48994	311	194	136	195	6	4	3	6	0	6	1	3	5	3	1	
6	>gil48994	429	253	164	263	13	11	5	12	0	2	3	5	4	5	1	
7	>gil48994	99	47	44	69	0	0	2	0	1	1	1	0	2	1	1	
8	>gil48994	259	156	90	140	8	8	2	6	0	4	2	1	5	5	1	

図 10 遺伝子毎の計算結果が保存された csv ファイル

### 3.4.2 「Codon Table」でのファイル保存

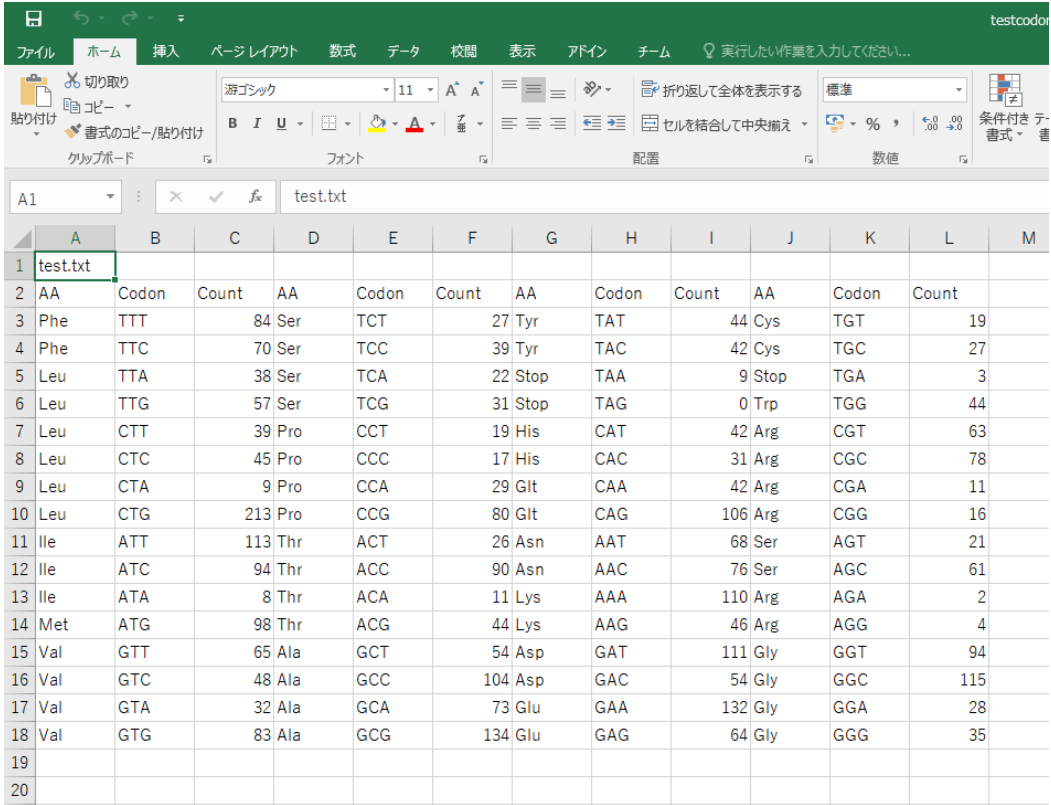


図 11 「Codon Table」における保存ボタン

図 11 の赤枠部分は、「Codon Table」における保存ボタンです。このタブでは、「③コドン表に出力した計算結果」をファイルに保存できます。「ファイルに保存」

ボタンをクリックすることで、ダイアログが出てきて、保存する場所と保存するファイル名を決めて、ファイルに保存します。

図 12 は、出力する値に「Count」を選択して、コドン表に出力した計算結果を保存したファイルの例です。



The screenshot shows a spreadsheet application with a green header bar. The file name 'test.txt' is visible in the top left. The spreadsheet contains a table of codon counts. The first row (row 1) has a green border around the cell 'test.txt' in column A. The table starts from row 2, column A, and continues to row 18, column M. The table is organized in groups of three columns each, representing different amino acids and their corresponding codons and counts.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	test.txt												
2	AA	Codon	Count	AA	Codon	Count	AA	Codon	Count	AA	Codon	Count	
3	Phe	TTT	84	Ser	TCT	27	Tyr	TAT	44	Cys	TGT	19	
4	Phe	TTC	70	Ser	TCC	39	Tyr	TAC	42	Cys	TGC	27	
5	Leu	TTA	38	Ser	TCA	22	Stop	TAA	9	Stop	TGA	3	
6	Leu	TTG	57	Ser	TCG	31	Stop	TAG	0	Trp	TGG	44	
7	Leu	CTT	39	Pro	CCT	19	His	CAT	42	Arg	CGT	63	
8	Leu	CTC	45	Pro	CCC	17	His	CAC	31	Arg	CGC	78	
9	Leu	CTA	9	Pro	CCA	29	Glt	CAA	42	Arg	CGA	11	
10	Leu	CTG	213	Pro	CCG	80	Glt	CAG	106	Arg	CGG	16	
11	Ile	ATT	113	Thr	ACT	26	Asn	AAT	68	Ser	AGT	21	
12	Ile	ATC	94	Thr	ACC	90	Asn	AAC	76	Ser	AGC	61	
13	Ile	ATA	8	Thr	ACA	11	Lys	AAA	110	Arg	AGA	2	
14	Met	ATG	98	Thr	ACG	44	Lys	AAG	46	Arg	AGG	4	
15	Val	GTT	65	Ala	GCT	54	Asp	GAT	111	Gly	GGT	94	
16	Val	GTC	48	Ala	GCC	104	Asp	GAC	54	Gly	GGC	115	
17	Val	GTA	32	Ala	GCA	73	Glu	GAA	132	Gly	GGA	28	
18	Val	GTG	83	Ala	GCG	134	Glu	GAG	64	Gly	GGG	35	
19													
20													

図 12 コドン表に出力した計算結果が保存された csv ファイル

#### 4. 一括処理についての操作手順

CUG ではコドン組成を計算したい遺伝子領域の塩基配列のマルチ fasta 形式ファイルがあるフォルダを指定すると、そのフォルダ内にある全てのマルチ fasta 形式ファイルについてコドン組成を計算することが可能です。まず、簡単な操作の流れを図 13 に示します。

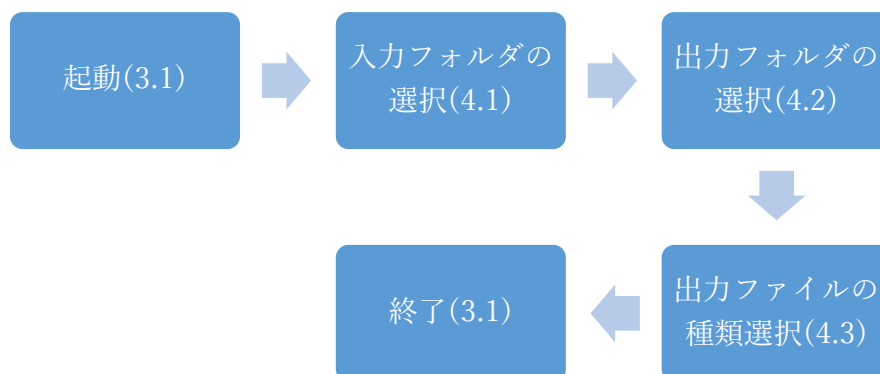


図 13 一括処理についての操作手順

この操作の流れに沿って操作の説明を行っていきます。なお、起動と終了に関しては「3.1 起動と終了」を参照してください。

##### 4.1 入力フォルダの選択

初めに、図 3 の画面で「フォルダ内一括処理」ボタンをクリックします。



図 14 「フォルダ内一括処理」ボタンクリック時の画面

図 14 は「フォルダ内一括処理」ボタンをクリックした時の画面です。まずは、1 番目の指示に従い、コドン組成を計算したマルチ FASTA 形式ファイルがあるフォルダを選択します。「フォルダ選択」ボタンをクリックしてください。

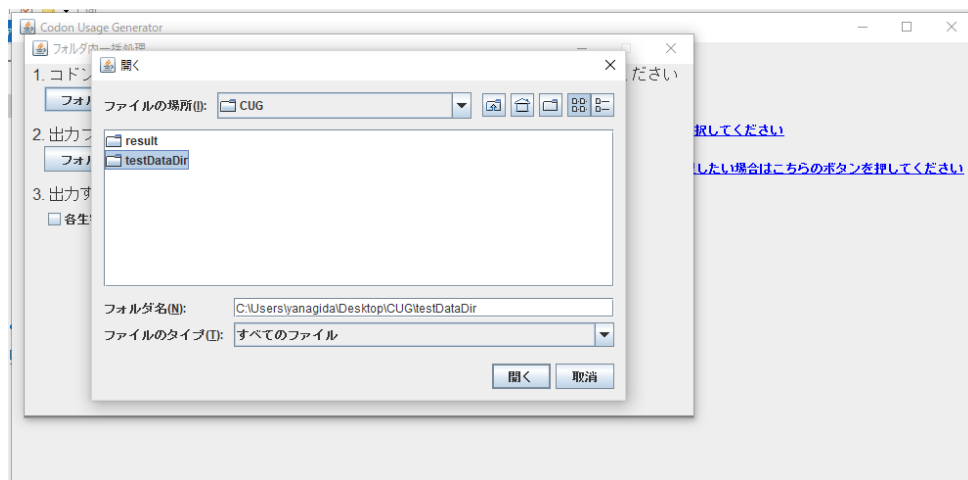


図 15 フォルダ選択ダイアログ

図 15 は「フォルダ選択」をクリックした際に出てくるフォルダ選択のダイアログボックスです。ここで任意のフォルダを選択すると図 16 の赤枠部分のように選択されたフォルダが入力されます。

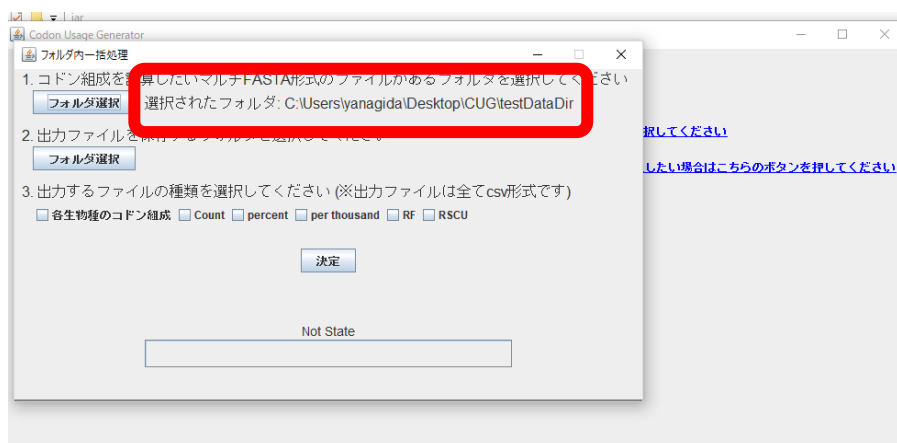


図 16 入力フォルダ選択後の画面

## 4.2 出力フォルダの選択

次に、2 番目の指示に従い、出力ファイルを保存するフォルダを選択します。入力フォルダと同様に「フォルダ選択」ボタンをクリックします。今回は、新しいフォルダ「test\_result」を作成し、その中に結果を出力します。新しいフォルダを作成するには、図 17 の画面にある赤枠のボタンをクリックします。

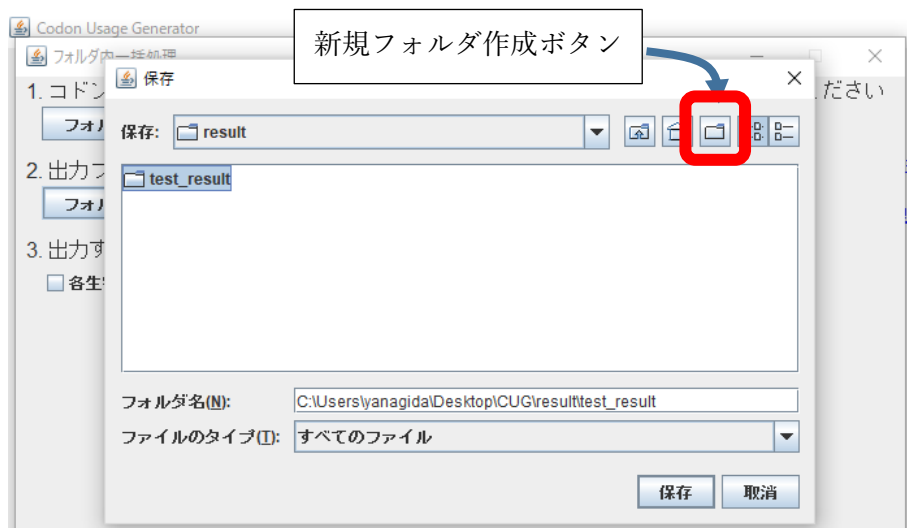


図 17 出力フォルダ選択ダイアログ

フォルダを選択すると図 18 のように出力フォルダが入力されます。



図 18 出力フォルダ選択後の画面

### 4.3 出力ファイルの種類選択

最後に、3 番目の指示に従い、出力するファイルの種類を選択します。図 19 の赤枠部分のコンボボックスから必要なファイルを選択し、「決定」ボタンをクリックすると一括処理が実行されます。

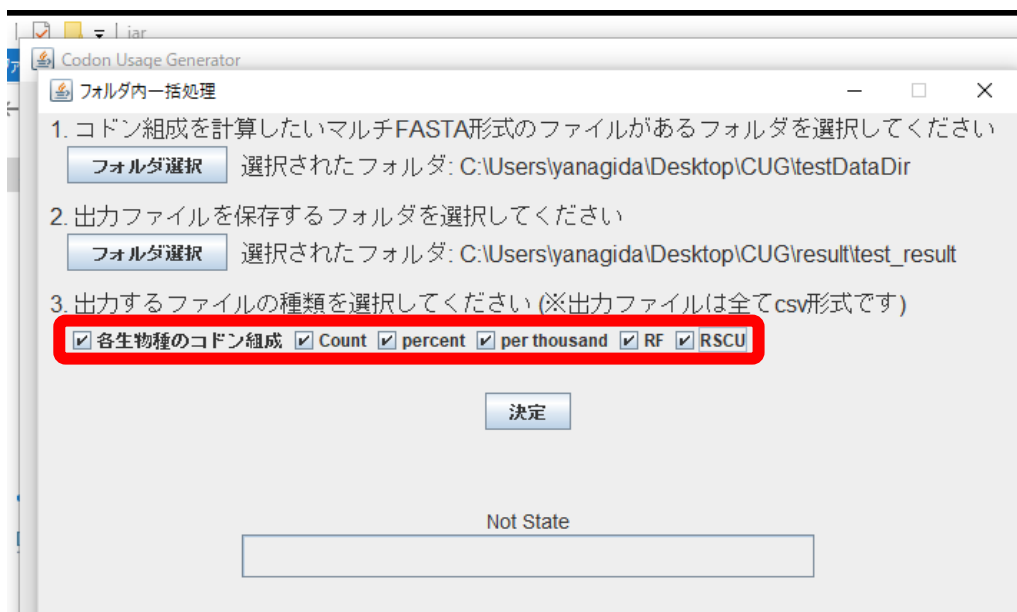


図 19 ファイルの種類選択

一括処理が実行されると図 20 のように指定したフォルダにファイルが出力されます。

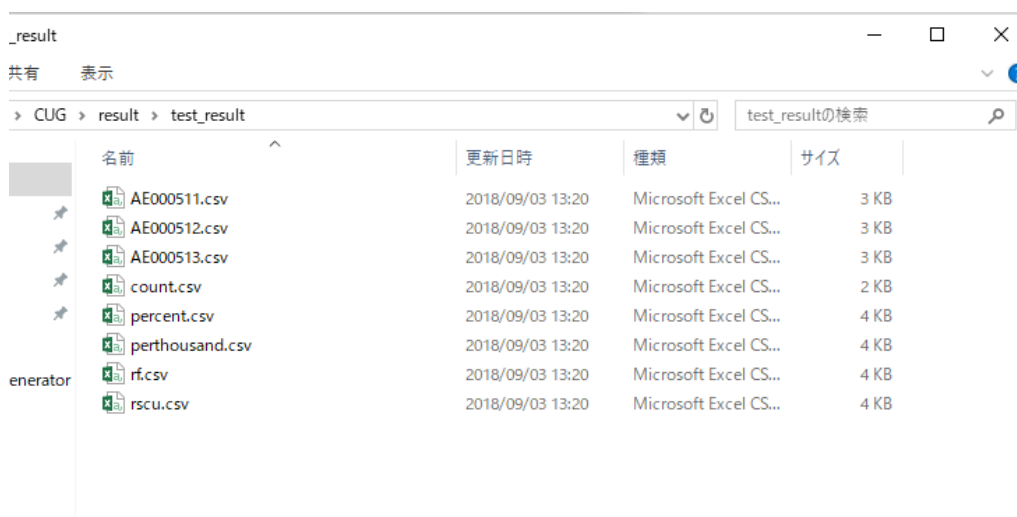


図 20 出力されたファイル群

## 5. 計算項目について

### ① 「Count」

「Count」は、対象のコドンの塩基配列中における個数を表しています。

### ② 「percent」、「per thousand」

「percent」、「per thousand」は、それぞれ対象のコドンの塩基配列中における割合を百分率と千分率で表したものです。

### ③ 「RF」<sup>[1][2]</sup>

「RF」は、「相対コドン頻度 (relative codon frequency)」の略称です。同義コドン中の、対象のコドンの割合を表します。

ある  $a$  番目のアミノ酸の縮重が  $d_a$  であるとします。このとき、同義コドン中の  $c$  番目のコドンの数を  $n_{ac}$  と表します。また、 $RF_{ac}$  を  $a$  番目のアミノ酸における  $c$  番目のコドンの  $RF$  の値とします。 $RF_{ac}$  は以下のように算出されます。

$$RF_{ac} = \frac{n_{ac}}{\sum_{c=1}^{d_a} n_{ac}}$$

全ての同義コドンが同頻度で、使用されている場合は以下ようになります。

$$RF_{ac} = \frac{1}{d_a}$$

また、同義コドンが1つしか使用されていない場合は、 $RF_{ac}$  は最大値1をとります。

### ④ 「RSCU」<sup>[1][2]</sup>

「RSCU」は、「相対同義コドン使用 (relative synonymous codon usage)」の略称です。同義コドン中の対象コドンの使用頻度の偏りを表します。

$RSCU_{ac}$  を  $a$  番目のアミノ酸における  $c$  番目のコドンの  $RSCU$  の値とします。 $RSCU_{ac}$  は以下のように算出されます。

$$RSCU_{ac} = \frac{n_{ac}}{(1/d_a) \sum_{c=1}^{d_a} n_{ac}} = d_a RF_{ac}$$

全ての同義コドンが同頻度で、使用されている場合は以下ようになります。

$$RSCU_{ac} = d_a \frac{1}{d_a} = 1$$

また、同義コドンが1つしか使用されていない場合は、 $RSCU_{ac}$  は最大値  $d_a$  をとります。

## 6. 参考文献

- [1] Haruo Suzuki, Celeste J. Brown, Larry j. Forney, and Eva M (2008) Comparison of Correspondence Analysis Methods for Synonymous Codon Usage in Bacteria. DNA research, 15, 357-365.
- [2] Kanaya S, Yamada Y, Kudo Y, Ikemura T (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene, 238, 143-155