BLSOM,
BLSOM Viewer
Ver 1.2 版

マニュアル

新潟大学 自然科学研究科
馬場一郎，阿部貴志

最終更新日 2018 年 8 月 28 日

BLSOM Viewer 更新履歴

2018.08.28 version 1.2 を公開しました.
- 分類マップのカラーコードをエクスポートする機能を追加しました.

分類マップのマップウィンドウのファイルメニューの,
[名前をつけて系統の配色のカラーコードを txt 形式で保存]を
選択していただくことでエクスポート機能を実行できます.
形式は以下の通りです.

0xFFFFFF,
0xFFFFFF,
…
0xFFFFFF

この形式は，BLSOM プログラムの DRAW のヘッダファイルの配列 color の要素に,
また，BLSOM Viewer のカラーコードのインポート機能にお使いいただけます.

2018.03.26 version 1.1 を公開しました.
- 出現頻度マップを表示する際のバグを改善しました.
- 分類マップのカラーコードをインポートする機能を追加しました.
- 出現頻度マップのパターンの一覧をインポートする機能を追加しました.

2018.03.01 version 1.0 を公開しました.

2017.03.01 version β を公開しました.

目次

1. 概要

我々は，ゲノムに潜む生物種固有の特徴を解明する目的で，大量かつ多次元データの 2 次元や 3 次元でのクラスタリングと可視化法として，コホネン博士らが開発した，教師なし学習アルゴリズム「自己組織化マップ（Self-Organizing Map, SOM）」に着目し，コホネン SOM の長所を生かしながら，再現性のある分類結果を得るアルゴリズムとして「一括学習型自己組織化マップ（Batch-Learning Self-Organizing Map, BLSOM）」を開発し，ゲノム配列解析に適用しています.

3 連や 4 連塩基といった連続塩基(オリゴヌクレオチド)の出現頻度に着目することで，生物種の情報を計算の途中で一切与えずに，連続塩基の出現頻度の類似性だけを基に，ゲノム配列断片を生物種ごとに高精度に分離(自己組織化)させる強力なクラスタリング能力を持ち，その結果を容易に可視化できます.

このマニュアルでは，ゲノム配列データから，ゲノム配列を断片化し，その配列断片の連続塩基組成の度数(もしくは，頻度)を計算し，BLSOM 解析を Linux 上で行うプログラム一式と，その BLSOM 解析結果を容易に閲覧可能な BLSOM Viewer（Windows or Linux で利用可能）について説明します.

BLSOM 解析プログラムは，Linux 版のみの提供となりますが，より多くのゲノムを対象とした解析が可能となります.

BLSOM Viewer は，BLSOM によって得られる解析結果を容易でかつ視覚的に分析することを目的としたソフトウェアです．お使いの PC の OS に関係なく動作しますが，Java の環境のインストールが必須です.

2. ダウンロードしたファイルの説明

「BLSOMtool.tar.gz」解凍後のディレクトリとファイルを説明します.

(ア) frq ディレクトリ

ゲノム配列データを断片化し，連続塩基組成計算行うプログラム一式です．詳しくは，「4.1 節 連続塩基組成計算」をご覧ください.

(イ) BLSOM ディレクトリ

BLSOM 解析用プログラム一式です．詳細については，「4.2 節 BLSOM プログラム群による解析」をご覧ください.

(ウ) BLSOMviewer_v1_0.jar

BLSOM Viewer プログラムです．詳細については，7 章以降をご覧ください.

3. 動作環境

連続塩基組成計算を行うプログラム一式，BLSOM 解析用プログラム一式は，64-bit Linux system(Kernel≧2.6) の環境下で，GNU C/C++ complier (≧4.4.7)，ruby (≧1.8.7) が必須になります.

また，BLSOM Viewer は Java で開発したソフトウェアのため，Java(≧1.8.0)環境のインストールが必須になります.

4. BLSOM による解析

本章では BLSOM による解析方法について説明します．塩基配列データを対象にした連続塩基組成に基づく BLSOM 解析の流れを図 1 のようになります．図 1 中の(1)連続塩基組成計算については 4.1 節，(2)の BLSOM プログラム群による解析については 4.2 節において詳しく説明します.

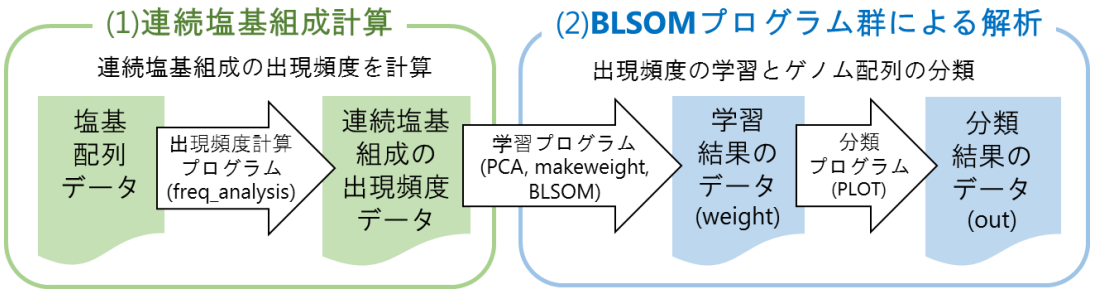


図 1　BLSOM による解析の流れ

4.1. 連続塩基組成計算

連続塩基組成計算では，塩基配列データ内の連続塩基組成について出現頻度の計算を行います．図 1 中の BLSOM による解析の流れの中では(1)の連続塩基組成計算の部分に当たります.

連続塩基組成計算では，frq ディレクトリ以下のプログラムソースを用います．言語は Ruby(≧1.8.7)です.

- freq_analysis_ver201803.rb
- DegeTetra_2ren.ptn
- DegeTetra_s4ren.ptn
- DegeTetra_4ren.ptn
- DegeTetra_s5ren.ptn

} いずれか 1 つを使用(./ptn 以下に格納)

① 連続塩基パターンの選択

解析したい連続塩基組成に合わせてパターンファイルを用意します(表 1).

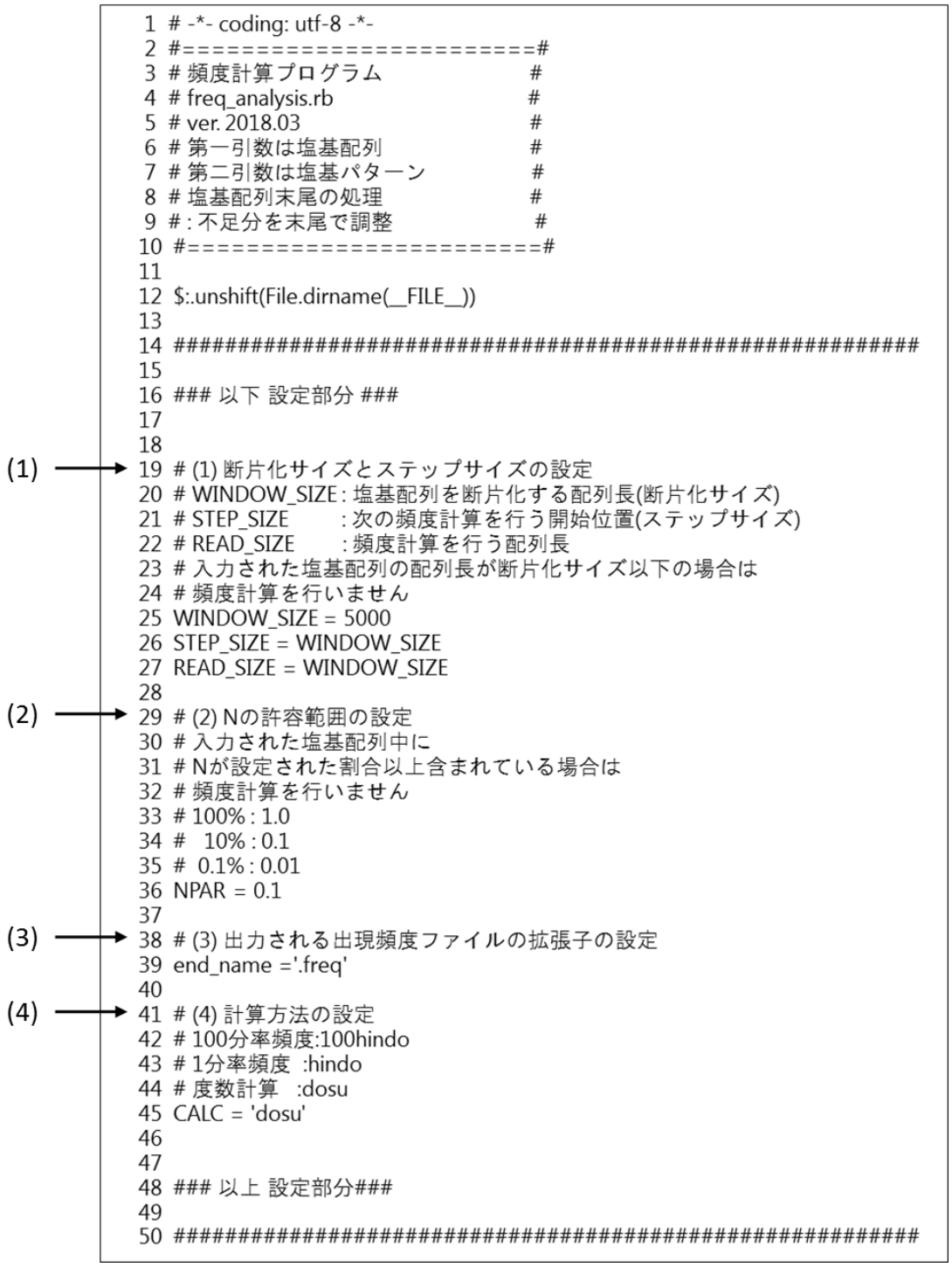
表 1 連続塩基組成とパターンファイルの対応

連続塩基組成	パターンファイル名
2 連続塩基組成	DegeTetra_2ren.ptn
縮退 4 連続塩基組成	DegeTetra_s4ren.ptn
4 連続塩基組成	DegeTetra_4ren.ptn
縮退 5 連続塩基組成	DegeTetra_s5ren.ptn

② 解析条件の設定

連続塩基組成計算を行うプログラム「freq_analysis_ver201803.rb」内の設定部分を変更します(図 2). 設定部分は図 2 の(1)～(4)部分になります. 詳細について以下で説明します.

なお, 現在, 連続塩基組成計算プログラム「freq_analysis_ver201803.rb」は fasta 形式にのみ対応しています(マルチ fasta 形式には対応していません).



```
1  # -*- coding: utf-8 -*-
2  #=========================#
3  # 頻度計算プログラム                #
4  # freq_analysis.rb                #
5  # ver. 2018.03                    #
6  # 第一引数は塩基配列                #
7  # 第二引数は塩基パターン            #
8  # 塩基配列末尾の処理                #
9  # : 不足分を末尾で調整              #
10 #=========================#
11
12 $:.unshift(File.dirname(__FILE__))
13
14 ##############################################################
15
16 ### 以下 設定部分 ###
17
18
19 # (1) 断片化サイズとステップサイズの設定
20 # WINDOW_SIZE : 塩基配列を断片化する配列長(断片化サイズ)
21 # STEP_SIZE       : 次の頻度計算を行う開始位置(ステップサイズ)
22 # READ_SIZE       : 頻度計算を行う配列長
23 # 入力された塩基配列の配列長が断片化サイズ以下の場合は
24 # 頻度計算を行いません
25 WINDOW_SIZE = 5000
26 STEP_SIZE = WINDOW_SIZE
27 READ_SIZE = WINDOW_SIZE
28
29 # (2) Nの許容範囲の設定
30 # 入力された塩基配列中に
31 # Nが設定された割合以上含まれている場合は
32 # 頻度計算を行いません
33 # 100% : 1.0
34 #  10% : 0.1
35 #  0.1% : 0.01
36 NPAR = 0.1
37
38 # (3) 出力される出現頻度ファイルの拡張子の設定
39 end_name ='.freq'
40
41 # (4) 計算方法の設定
42 # 100分率頻度:100hindo
43 # 1分率頻度  :hindo
44 # 度数計算   :dosu
45 CALC = 'dosu'
46
47
48 ### 以上 設定部分###
49
50 ##############################################################
```

(1) → 19 # (1) 断片化サイズとステップサイズの設定

(2) → 29 # (2) Nの許容範囲の設定

(3) → 38 # (3) 出力される出現頻度ファイルの拡張子の設定

(4) → 41 # (4) 計算方法の設定

図 2　連続塩基組成計算プログラム「freq_analysis_ver201803.rb」の設定部分

(1)　断片化サイズとステップサイズの設定

塩基配列の配列長に依存しない解析を行うために一定の配列長に分割することを断片化といいます．また，この一定の配列長のことを断片化サイズといいます．しかし，断片化を行うことにより重要な機能や意味を持つ連続した塩基の並びが分割され，塩基配列の情報が損なわれる可能性があります．そこで，断片化の際に次の断片化を行う開始位置を設定した配列長だけ移動させて行うことで情報の欠損を防ぐことができます．この断片化の開始位置を移動させることをステップといい，移動させる配列長をステップサイズといいます(図 3).

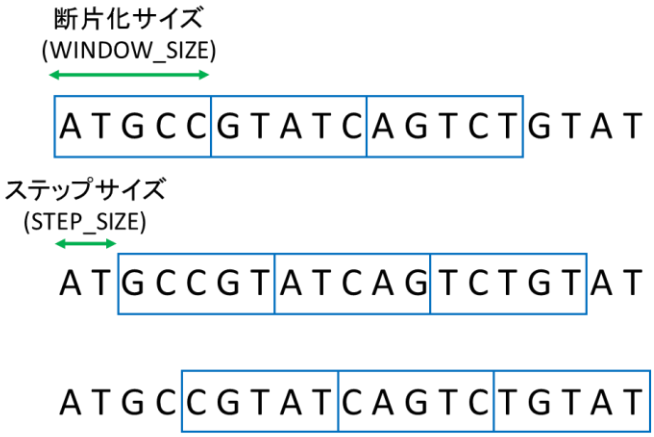


図 3　断片化サイズとステップサイズ
(配列長 19 の塩基配列を断片化サイズ 5，ステップサイズ 2 で断片化を行った際の例)

連続塩基組成計算プログラムでは，図 2(1)中の 25 行目の WINDOW_SIZE の値を変更することで断片化サイズを設定できます．また，図 2(1)中の 26 行目の STEP_SIZE を変更することでステップサイズを設定できます．初期設定では，ステップサイズが断片化サイズと等しくなるように設定されています．

なお，入力する総配列長より大きい値を設定すると計算が行われません．また，配列長の不足分は末尾部分で調節されます．

(例)総配列長:2,500，計算を区切る配列長:1,000 の場合：

[1-1000]と[1000-2000]と[1500-2500]の 3 パターンで出力

(2)　塩基配列データ中に現れる塩基 A，T，G，C 以外の文字列 N の許容範囲の設定

塩基配列の解析の際に塩基が A，T，G，C のいずれにも決定できない場合に用いられる N が，入力された塩基配列データに一定以上含まれていた場合，計算をしないように設定することができます．塩基配列中の N の許容範囲は図 2(2)中の 29 行目の NPAR の値を変更することで設定できます．初期設定では，10%(NPAR = 0.1)

で設定されています.

(3)　出力ファイルの拡張子の設定

連続塩基組成の計算結果を出力する際の拡張子を設定します．拡張子は図 2(3)中の 38 行目の値を変更することで設定できます．初期設定では「.freq」になるように設定されています.

(4)　連続塩基組成の計算方法の設定

連続塩基組成の計算方法を設定します．図 2(4)中の 41 行目の値を変更することで計算方法を設定できます．初期設定は度数による計算に設定されています.

以上①と②(1)～(4)の設定の完了後，以下の(実行コマンド)のようにプログラムを実行します．freq_analysis_v201803.rb の第一引数には入力する塩基配列データを指定します．第二引数にはパターンファイル DegeTetra.ptn を指定します．実行例を図 4 に示します.

（実行コマンド） ruby freq_analysis_ver201803.rb 塩基配列データのファイル名 パターンファイル名

```
1  $ ls
2  fasta  freq_analysis_v2.rb  ptn
3  $ ls fasta
4  a.fasta  b.fasta  c.fasta
5  $ ls ptn
6  DegeTetra_2ren.ptn  DegeTetra_s4ren.ptn  manual.txt
7  DegeTetra_4ren.ptn  DegeTetra_s5ren.ptn
8  $ cat fasta/*.fasta > all_fasta.txt
9  $ ruby freq_analysis_ver201803.rb all_fasta.txt ptn/(パターンファイル名)
```

図 4　連続塩基組成の実行例

図 4 の実行例では塩基配列データをまとめ，連続塩基組成計算を実行しています.

- 実行ディレクトリ内のファイル
  - freq_analysis_ver201803.rb：連続塩基組成計算プログラムファイル
  - ptn：パターンファイルを格納したディレクトリ
  - fasta：入力する塩基配列データが格納されたディレクトリ
- 1，2 行目：実行ディレクトリ内の確認.
- 3～7 行目：ディレクトリ fasta および ptn 内の確認
- 8 行目：ディレクトリ fasta 内の複数の塩基配列データを任意のファイル名(実行例では「all_fasta.txt」)として 1 つにまとめる.

（実行コマンド） cat まとめたい塩基配列データファイル ＞ まとめた後のファイル名(任意)

- 9 行目：連続塩基組成計算の実行.
  （実行コマンド） ruby プログラムファイル 入力する塩基配列データファイル パターンファイル

4.2. BLSOM プログラム群による解析

BLSOM プログラム群による解析では，まず，連続塩基出現頻度について学習を行います．次に，学習結果と分類を行いたいゲノム配列の連続塩基出現頻度の類似度を比較し，最も類似度が高い格子点に分類します．これにより，2 次元格子状のマップに生物系統ごとに分類を行うことがきできます．図 1 中の BLSOM による解析の流れの中では(2)の BLSOM プログラム群による解析の部分に当たります.

① BLSOM プログラムファイルに含まれるファイルについて

- run.csh：BLSOM プログラム群の実行用シェル
  以下のプログラムの実行を行います.
  (ア) PCA：主成分分析
  (イ) makeweight：初期リファレンスベクトルの作成
  (ウ) BLSOM：BLSOM の実行
  (エ) PLOT：入力データの帰属
- param.dat：BLSOM 実行のためのパラメータファイル
- input.dat：縮退 4 連続塩基頻度データ(136 次元ベクトルデータ. 以下の Bacteria.frq, Virus.frq and Eukaryote.frq をマージしたファイル).
- Bacteria.frq：バクテリア由来データ
- Virus.frq：ウイルス由来データ
- Eukaryote.frq：真核生物由来データ
- weight.final ：BLSOM の実行結果（リファレンスベクトル情報）
- out.txt：input.dat の帰属結果.
- out2.txt：バクテリア，ウイルス，真核生物別の帰属結果

② パラーメータファイル（param.dat）の設定

BLSOM を実行する前に，入力データに応じて以下の書式で記載されたパラメータファイルにて条件設定を行います(図 5).

なお, ベクトルの次元数と件数, マップの横軸のサイズの最大サイズは実行環境のメモリによります. また, マップの横軸のサイズの設定については横軸×縦軸の値がデータ件数の 1/4～1/10 程度となるように設定して下さい. 推奨は, (横軸×縦軸≒データ件数×1/4)です. マップサイズの最大数は横軸，縦軸とも 400 となります.

```
//DIMENSION（入力データのベクトルの次元数）
        136
//NumberOfDataInLine（1 行中の列数：ベクトルの次元数と同じにして下さい）
        136
//NumberOfLines
        1
//MaxNumberOfData（最大データ件数）
        4000000
//NumberOfIteration（学習回数：近傍係数の+20 を目安）
        10
//MaxMapSize（マップの横軸のサイズ）
        40
//NeighborParam（近傍係数：マップサイズの横軸の 1/4 を目安）
        5
```

図 5　パラメータファイル param.dat

③　BLSOM プログラム群の実行

(ア) 実行環境

64-bit Linux system(Kernel≧2.6)

GNU C/C++ complier (≧4.4.7)

(イ) BLSOM プログラム群の実行

（実行コマンド） ./run.csh 頻度データファイル　使用するコア数

（実行例） ./run.csh input.dat 8

上記は，頻度データファイル input.dat に対し，8 コアで並列実行の場合の例です．入力ファイルは，1 つのみとして下さい．

実行結果は，学習結果のファイルとして「weight.final」，分類結果のファイルとして「out.txt」というファイルが出力されます．ファイル名は常に固定となりますので，必要に応じて変更してください．また，設定した学習回数分繰り返し学習が行われますが，その学習の奇数(ファイル名：weight.odd)と偶数時(ファイル名：weight.even)にバックアップファイルが作成されます．

入力データ件数によっては，BLSOM の実行に多大な時間がかかるので，バックグランドでの実行をお勧めします．（実行際に「&」を付けて下さい．）

(ウ) PLOT(入力データの帰属)の実行時の注意

BLSOM の学習結果として得られた「weight.final」に各入力データが帰属される格子位置を求めます．なお，各ベクトルデータのファイルごとにカテゴリを付与しています．

「run.csh」では，1 種類の input.dat に対して実行していますので，全て同じカテゴリとなります．なお，カテゴリを分けたい場合は以下のようにカテゴリごとにデータを分け

て「PLOT」を再度実行してください.

（実行例） cat weight.final | ./PLOT Bacteria.frq
Virus.frq Eukaryote.frq > out2.txt

または,

（実行例） cat weight.final | ./PLOT *.frq > out2.txt

5. BLSOM Viewer が対応する解析結果の形式

本章では，BLSOM Viewer で対応している連続塩基出現頻度の学習結果と生物系統別の分類結果の形式について説明します.

5.1. 連続塩基出現頻度の学習結果

5.1.1. 連続塩基出現頻度の学習データ

連続塩基出現頻度の学習結果は，公開版の BLSOM において初期設定では「weight.final」というファイル名で出力されるデータです．Viewer が対応する連続塩基出現頻度の学習結果のファイル形式を以下に示します(図 6).

```
120■60                                  //マップの格子数x_n■y_n (文頭に一度のみ)
0■0                                     //格子番号x_1■y_1
0.2068■0.4107■0.0743 ……                //連続塩基出現頻度からなるベクトルデータ
0■1                                     //格子番号x_2■y_2
0.3170■0.3455■0.0799 ……                //連続塩基出現頻度からなるベクトルデータ
……
```

図 6　連続塩基出現頻度の学習結果のファイル形式
(■は半角スペース，またはタブによるスペースを表す)

図 6 の 1 行目はマップのサイズx_n，y_nを表します．2 行目は格子点の座標(x_1，y_1)，3 行目は連続塩基組成ごとの出現頻度を表します．それ以降は格子点の座標の行，および連続塩基出現頻度の行がマップの格子点数分繰り返されます．公開版の BLSOM では初期設定でこの形式で出力されます.

5.1.2. 連続塩基出現頻度マップ

連続塩基出現頻度マップは，次元ごとの各格子点の出現頻度の値をグラデーションで表現することで作成する．出現頻度マップ上の各格子点において，出現頻度が高いほど赤が濃

くなり，低いほど青が濃くなる．中間的な出現頻度の場合は白を示す．なお，出現度数マップを作成する際は，出現度数の値ではなく，格子点全体の中で出現度数の高さが何番目かという指標を用いて評価を行う．これは，極端な値の出現度数に影響され，他の出現度数が同じような配色となり，特徴が検出しづらくなることを回避するためである．

5.2. コドン使用頻度の学習結果

5.2.1. コドン使用頻度の学習データ

コドン使用頻度の学習データは 5.1.1.節とほぼ同様の内容のため，5.1.1.節をご覧ください．

なお，同義コドンが存在する場合，各同義コドン内の頻度の最大値は同義コドンの総数となる(相対使用頻度)．そのため，コドン使用頻度の学習結果の値は 1.0 を超える場合があります．

5.2.2. コドン別の使用頻度マップ

コドン別の使用頻度マップは，次元ごとの各格子点の出現頻度の高さをグラデーションで表現することで作成する．出現頻度マップ上の各格子点において，出現頻度が高いほど赤が濃くなり，低いほど青が濃くなる．中間的な出現頻度の場合は白を示す

5.2.3. 同義コドン別の使用頻度マップ

同義コドン別の使用頻度マップは，まず，格子点ごとに同義コドンの使用頻度を比較し，任意の格子点において最も使用頻度の高いコドンに対応した配色で格子点の色づけを行う．そして，その出現頻度の高さを閾値によってグラデーションで表現することにより作成する．使用頻度マップ上の各格子点において，使用頻度が高いほど各配色が濃くなり，低いほど配色が薄くなる．

5.3. 生物系統別の分類結果

5.3.1. 生物系統別の分類データ

生物系統別の分類結果は，公開版の BLSOM において初期設定では「out.txt」というファイル名で出力されるデータです．別途作成した「out2.txt」でも利用可能です．生物系統別の分類結果のファイル形式を以下に示します(図 7).

```
120■60                                  //マップの格子数x_n■y_n (文頭に一度のみ)
species_1■10                            //系統名■系統の要素の総数
fragment_1■7■12■0.0743 …               //要素名■格子番号x■y■ユークリッド距離
fragment_2■12■18■0.1753 ……
……
species_2■22
fragment_3■8■5■0.3157 ……
fragment_4■23■35■0.0574 ……
……
```

図 7　生物系統別の分類結果のファイル形式
(■は半角スペース，またはタブによるスペースを表す)

図 7 の 1 行目はマップのサイズx_n，y_nを表します．2 行目は生物系統名とその系統のゲノム断片数，3 行目以降は 2 行目の生物系統のゲノム断片名と分類された格子点の座標(x，y)およびユークリッド距離の一覧が出力されます．それ以降は生物系統名とその系統の断片数の行，およびそのゲノム断片名とユークリッド距離の一覧の行が生物系統の数分繰り返されます．公開版の BLSOM では初期設定でこの形式で出力されます.

5.3.2.　生物系統別の分類マップ

生物系統別の分類マップでは，マップ上の各格子点において，同一の系統のみが分類されている場合にはその系統を示す色，複数の系統が混在する場合には黒色，そして配列が 1 つも分類されていない場合には白色を示す.

6.　BLSOM Viewer が対応する分割ファイルの形式

BLSOM Viewer では，解析結果の容量やソフトウェアの動作環境によっては各種機能が十分に動作しない場合があります．そこで，BLSOM による解析結果を分割し読み込むことによって機能性を確保することが可能です．本章では，BLSOM Viewer が対応している解析結果の分割ファイルの形式について説明します.

6.1.　連続塩基出現頻度の学習結果の分割ファイル

連続塩基出現頻度の学習結果を分割した際の形式は以下の通りです(図 8).

temp_(入力ファイル名)_dim(次元番号).txt

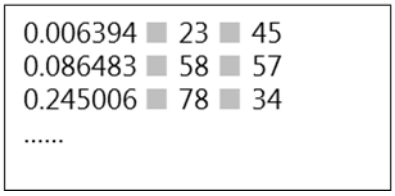


図 8　連続塩基出現頻度の学習結果の分割形式
(■は半角スペース，またはタブによるスペースを表す)

学習結果の分割ファイルは次元数分作成します．各ファイル名は「temp_(入力ファイル名)_dim(次元番号).txt」とし，次元の番号を書き込みます．図 8 で示す例のように．ファイル内では 1 項目目に連続塩基出現頻度，2 項目目は格子点のx座標，3 項目目は格子点のy座標を出力します．同様の形式が，マップの格子点数分繰り返されます．

6.2.　生物系統別の分類結果の分割

生物系統別の分類結果を分割した際の形式は以下の通りです(図 9).

temp_(入力ファイル名)_y(格子点のy座標).txt

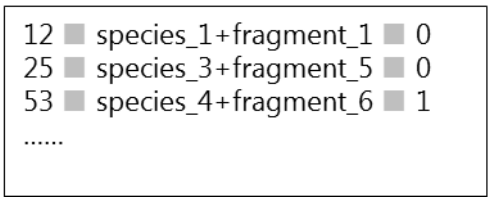


図 9　生物系統別の分類結果の分割形式
(■は半角スペース，またはタブによるスペースを表す)

分類結果の分割ファイルはマップサイズのy_n分作成します．各ファイル名は「temp_(入力ファイル名)_y(格子点のy座標).txt」とし，格子点のy 座標を書き込みます．図 9 で示す例のように，ファイル内では 1 項目目に格子点のx座標，2 項目目に生物系統名とゲノム断片名，3 項目目では系統番号を出力します．同様の形式が，格子点yに対応する格子点xに分類されたゲノム断片数分繰り返されます．

7.　BLSOM Viewer の機能

本ソフトウェアを用いることで，5.1.1.，5.1.2.節 BLSOM による解析にて得られる連続塩基出現頻度の学習結果と生物系統別の分類結果について分析を行うことができます．また，本ソフトウェアは 5.1.3.節コドンの使用頻度の学習結果にも対応しています．本章

ではそれぞれの分析機能について説明します.

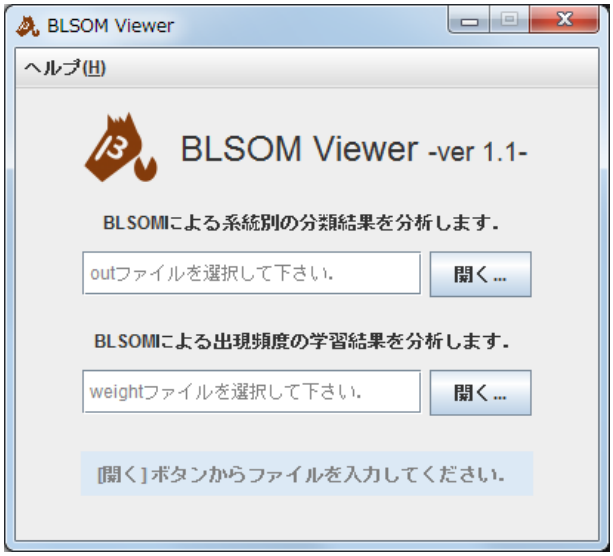


図 10　BLSOM Viewer トップ画面

7.1.　生物系統別の分類結果の分析

生物系統別の分類結果について分析を行うことができます．初期設定では[out.txt]というファイル名で出力されるデータです．別途作成した[out2.txt]でも分析可能です.

7.1.1.　ファイルの入力

生物系統別の分類結果の分析を行うには，図 10 のソフトウェアのトップ画面の上部の[開く]からファイルを入力します.

7.1.1.1.　マップの色の設定

図 10 のソフトウェアのトップ画面においてファイルを入力すると，ファイルの総行数を取得した後に，マップの色の設定ウィンドウが表示されます(図 11).

図 11　マップの色の設定ウィンドウ

図 11 の例に示すようなウィンドウを操作することにより，マップの作成の際に使用する色について設定を行うことができます．図 11 のウィンドウ中の各項目について以下で説明します．

① [初期設定を使用する]の項目を選択すると，図 11 の上部に表示される色の一覧を用いてマップが作成されます．

② [ファイルを選択]の項目を選択すると，HTML カラーコードの一覧が書き込まれたファイルを選択するダイアログが表示されます．HTML カラーコードについては 7.1.8.2.節をご覧下さい．対応する HTML カラーコードの一覧ファイルの形式は以下の通りです．

- 文頭：#，区切り：改行
- 文頭：0x，区切り：改行
- 文頭：0x，区切り：カンマと改行

7.1.1.2. ファイルの読み込み方法の設定

色に関する設定が完了すると，ファイルの読み込み方法に関するウィンドウが表示されます(図 12).

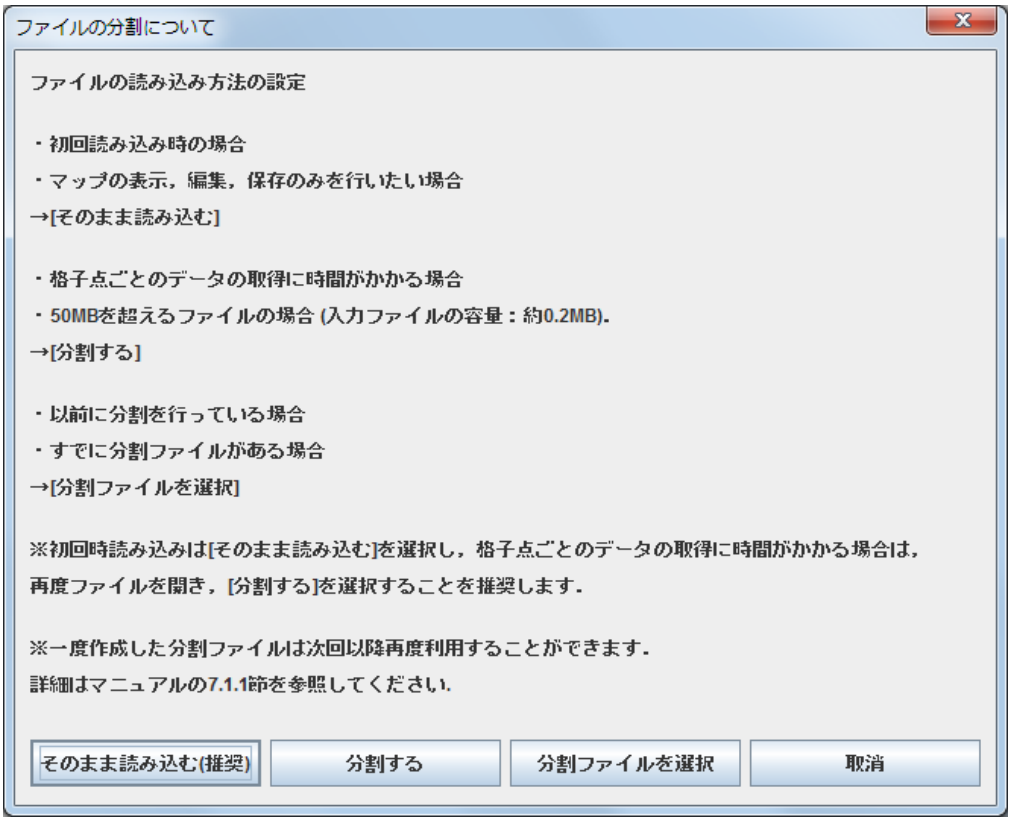


図 12　ファイルの読み込み方法に関するウィンドウ

初回読み込み時は，図 12 のウィンドウ中の[そのまま読み込む]を選択することを推奨します(7.1.1.節①項参照)．ただし，入力するファイルの容量やソフトウェアが動作する環境によっては，格子点ごとのゲノム断片データの表示(7.1.4.節参照)に時間がかかる場合があります．そのため，入力ファイルの分割ファイルを用意することで，読み込み時間を短縮することができます．ファイルの分割は図 12 のウィンドウ中の[分割する]を選択することで行うことができます(7.1.1.節②項参照)．また，一度作成した分割ファイルは次回以降に再利用することができます(7.1.1.節③項参照).

図 12 のウィンドウ中の各項目について以下で説明します.

① [そのまま読み込む]の項目を選択すると，マップが表示されるウィンドウに移行します．マップの表示，編集，保存(7.1.2.～7.1.3.，7.1.7.～7.1.8.節参照)は問題なく利用できますが，入力ファイルの容量やソフトウェアの動作環境によっては，ゲノム断片データの分析(7.1.4.～7.1.6.節参照)に時間がかかる場合があります.

② [分割する]の項目を選択すると，入力ファイルの分割ファイルを作成することができます．なお，入力ファイルの容量やソフトウェアの動作環境によっては分割に時間がかかる場合があります．初期設定では「temp_blsomViewer」というディレクトリ以下に，「temp_(入力ファイル名)_fragmentName」というディレクトリが作成され，さらにこのディレクトリ以下に「temp_(入力ファイル名)_y(マップの格子番号の y).txt」というファイル名で分割ファイルがマップの格子番号の y 分作成されます．作成された分割ファイルは，次回以降に[分割ファイルを選択する]を選択することによって再度利用することができます．分割処理中は進捗ウィンドウが表示されます(図 13).

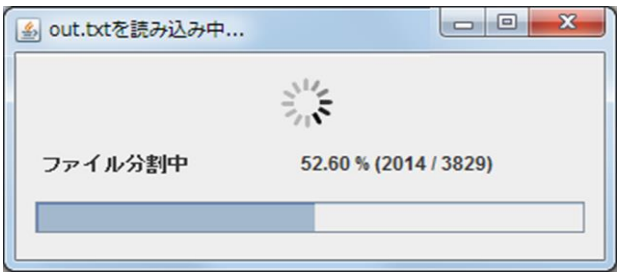


図 13　ファイル分割の進捗を表すウィンドウ

図 13 の進捗ウィンドウ中央の割合はファイル分割の進捗率を表します．また，カッコ内左の数値は現在読み込みが終わった行数，右の数値は分割を行っているファイルの総行数を表します．分割処理が完了し次第，マップが表示されるウィンドウに移行します.

③ [分割ファイルを選択]の項目を選択すると，ディレクトリの選択ダイアログが表示されます．そこで，分割ファイルが存在するディレクトリを選択することで，分割ファイルを読み込むことができます．ディレクトリに置かれている分割ファイル名は，「temp_(入力ファイル名)_y(マップの格子番号の y).txt」である必要があります．ソフトウェア上で作成された分割ファイルの場合，初期設定では，「temp_blsomViewer」というディレクトリ以下の，「temp_(入力ファイル名)_fragmentName」というディレクトリに格納されています．分割処理が完了し次第，マップが表示されるウィンドウに移行します.

7.1.2. 生物系統別の分類マップの表示

データの入力完了後，ウィンドウ上に生物系統別の分類マップと関連する各種データが表示されます(図 14).

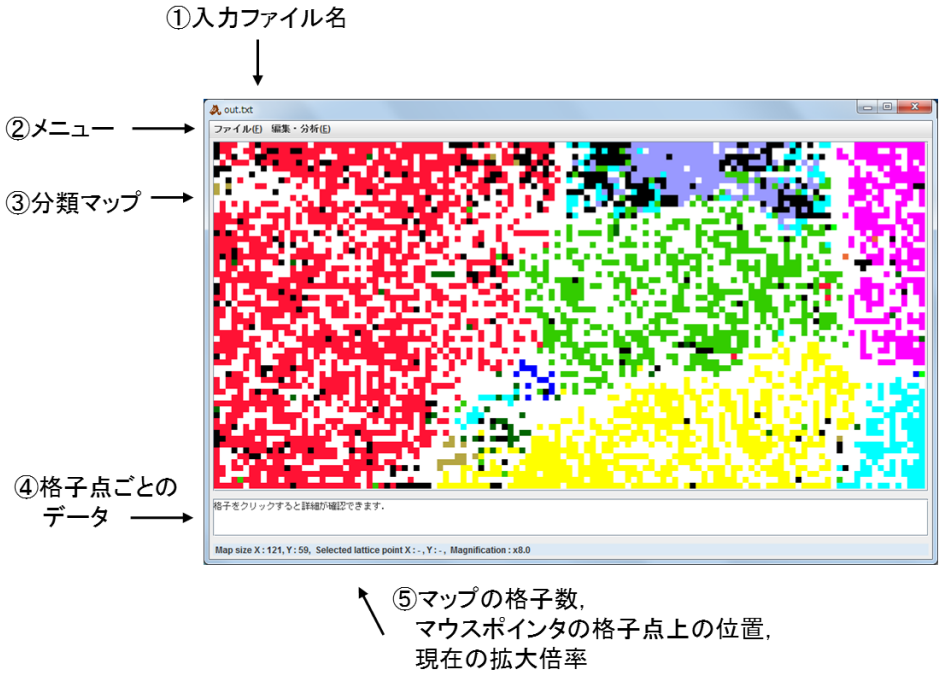


図 14　マップウィンドウ

図 14 のウィンドウ中の各項目について以下で説明します.

① ウィンドウ最上部のタイトルバーには入力ファイル名が表示されます.
② ウィンドウ上部のメニューを選択することで，データの保存やマップの編集が可能です(7.1.7.～7.1.8.節参照).
③ ウィンドウ中央では，生物系統別の分類マップが表示されます．マップは任意の倍率に拡大・縮小が可能です．また，マップがマップウィンドウに収まらなくなった場合，スクロールバーが表示され，マウス操作によってスクロールが可能です．マップ上の格子点を左クリックすると選択された格子点に分類されたゲノム断片データをウィンドウ下部のテキストエリアに表示することができます(7.1.4.節参照).
④ ウィンドウ下部のテキストエリアでは，格子点ごとのゲノム断片データが表示されます.
⑤ ウィンドウ最下部では，入力されたファイルのマップサイズ，現在マウスポインタが置かれている格子番号，および現在のマップの拡大倍率が表示されます(図 15).

Map size X : 121, Y : 59, Selected lattice point X : 1, Y : 1, Magnification : x8.0

図 15　図 14 のマップウィンドウ最下部に表示されるデータ

7.1.3. 生物系統に対する配色の凡例の表示

マップの表示と同時に生物系統ごとの配色がマップウィンドウとは別ウィンドウで表示されます(図 16). 図 16 の例で示す凡例ウィンドウでは，ソフトウェアに読み込まれた順に振られた系統番号と，系統ごとの配色，および生物系統名が表示されます．この凡例データは保存することが可能です(7.1.8.節参照).

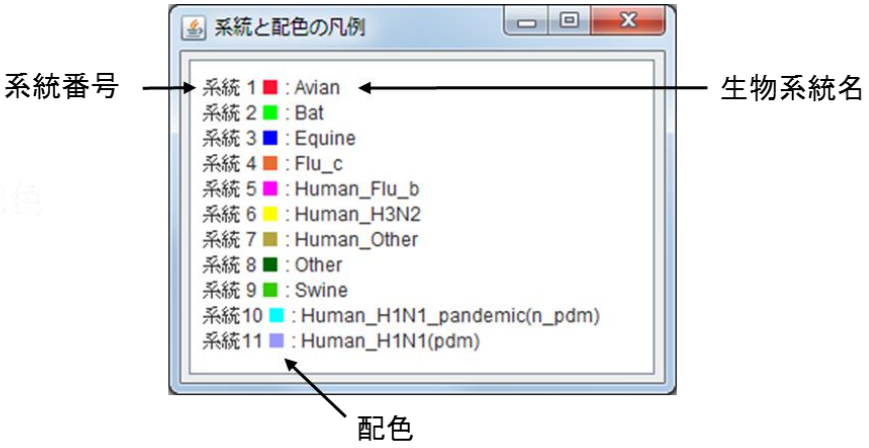


図 16　生物系統ごとの配色の凡例

7.1.4. 各格子点に分類されたゲノム断片データの表示

図 14 の例で示すウィンドウ上の分類マップ上において，格子点を左クリックするとマップウィンドウ下部のテキストエリアが更新され，各格子点に分類されたゲノム断片データを確認できます(図 17). 図 17 のテキストエリアでは，1 行目に選択されている格子番号と分類されたゲノム断片の総数，2 行目以降では，分類されたゲノム断片の一覧が表示されます.

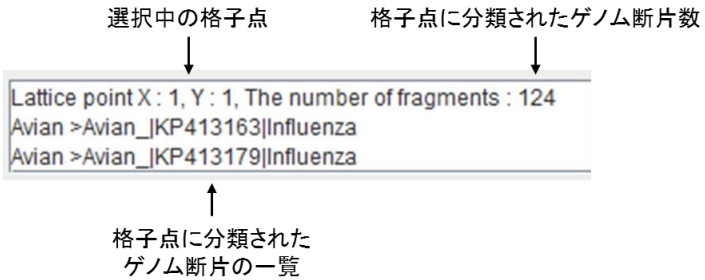


図 17　表示されるゲノム断片データ

7.1.5. 複数の系統が分類された黒色の格子点における系統の内訳の表示

図 14 の例で示すウィンドウ上の分類マップ上において，複数の系統が分類されたことを表す黒色の格子点を左クリックすると，黒色の格子点中に混在している生物系統やその内訳を確認することができます(図 18)．図 18 のウィンドウ上部の円グラフは，選択した黒色の格子点に分類されたゲノム断片の系統別の割合を表します．また，ウィンドウ下部のテキストエリアでは，系統ごとの割合やゲノム断片数の内訳を確認することができます．

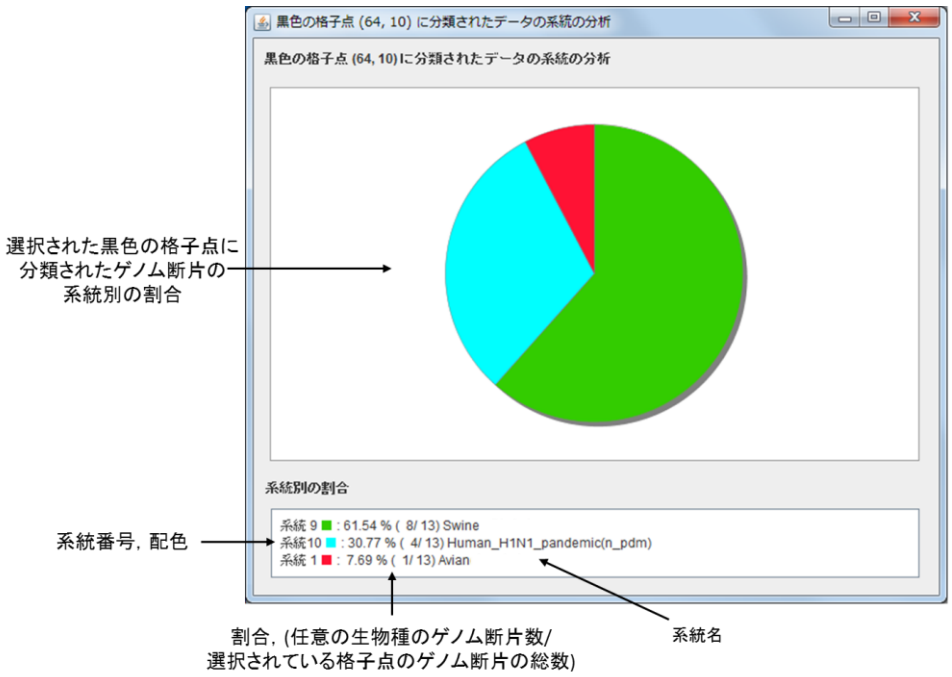


図 18　黒色の格子点に分類されたゲノム断片の系統別の割合

7.1.6. 特定の生物系統を含むすべての黒色の格子点の分析

特定の系統を含むすべての黒色の格子点に分類されたゲノム断片の系統を分析することができます．特定の系統と同じ格子点に分類されたゲノム断片は連続塩基出現頻度の類似度が高いため，関連性の高い可能性のある系統を抽出することができます．

分析を行うには図 14 のマップウィンドウ上の[編集・分析]メニューにおける[特定の系統を含む黒色の格子点の分析]を選択します(図 19)．そして，分析を行いたい系統を選択します(図 20)．または，マップ上の分析を行いたい系統上で右クリックをし，表示されるポップアップメニューの[系統 x を含む黒色の格子点の分析]を選択します(図 21)．なお，白

色の格子点，および黒色の格子点上では右クリックをしても黒色の格子点の分析メニューは表示されません．

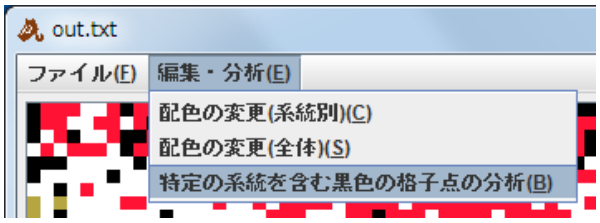


図 19　特定の系統を含む黒色の格子点の分析(編集・分析メニューから)

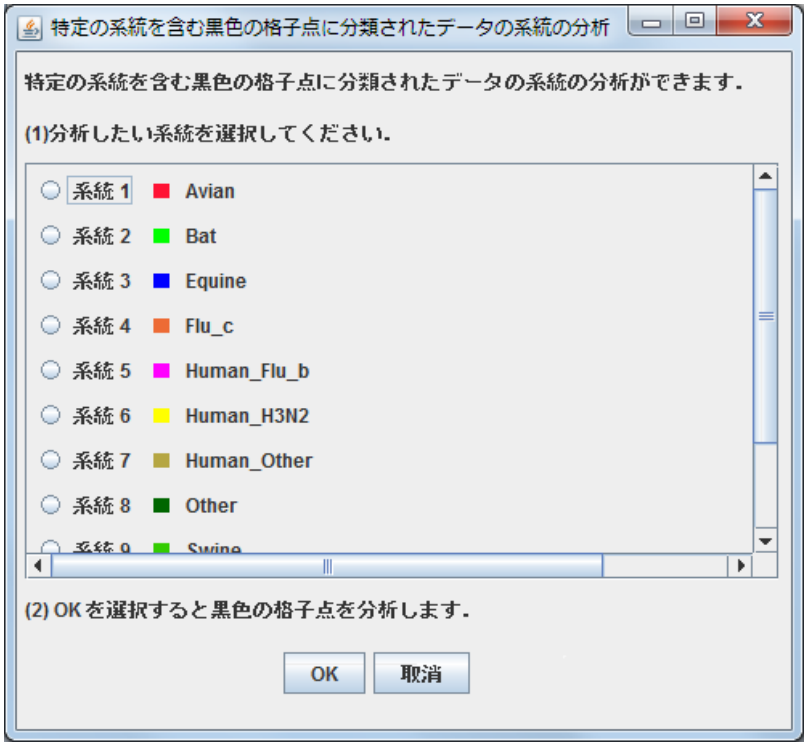


図 20　分析を行う系統の選択画面

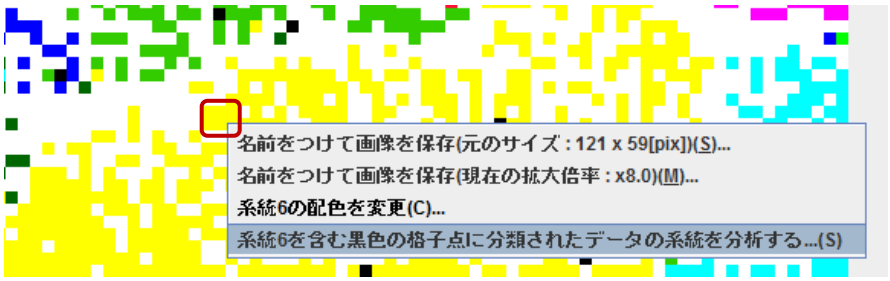


図 21　特定の系統を含む黒色の格子点の分析(マップ上のポップアップメニューから)
系統 6(黄色の領域)上で右クリック(赤枠部分)した際の例

系統を選択すると分析結果が表示されます(図 22)．図 22 のウィンドウ上部の円グラフは，選択した系統を含むすべての黒色の格子点に分類されたゲノム断片の系統別の割合を表します．また，図 22 のウィンドウ下部のテキストエリアでは，系統ごとの割合やゲノム断片数の内訳を確認することができます．

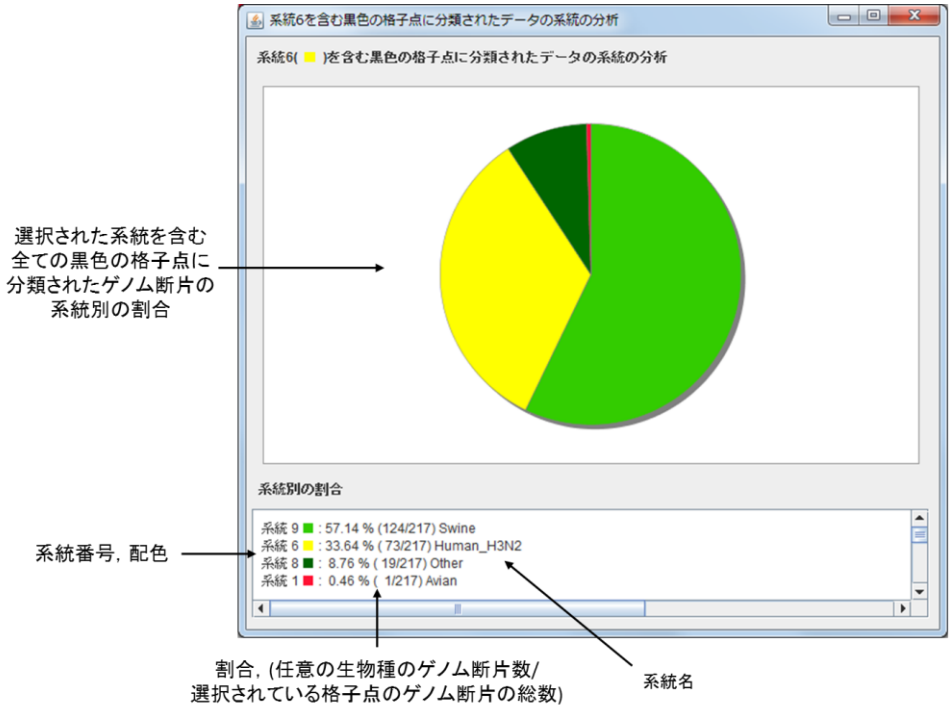


図 22　特定の系統を含むすべての黒色の格子点の分析結果

7.1.7.　生物系統に対する配色の変更

マップ上の系統別の配色を変更するには，[編集・分析]メニューの[配色の変更(系統別)]を選択します(図 23)．または，マップ上の配色の変更を行いたい系統上で右クリックをし，表示されるポップアップメニューの[系統 x の配色の変更]を選択します(図 24)．なお，白色の格子点，および黒色の格子点上では右クリックをしても配色の変更メニューは表示されません．

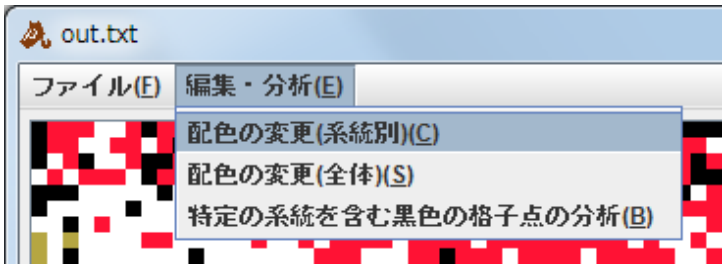


図 23　配色の変更(編集・分析メニューから)

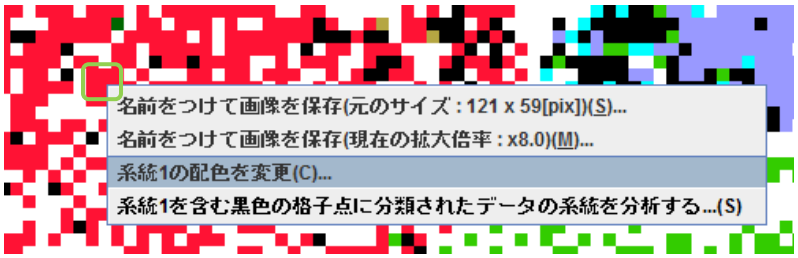


図 24　配色の変更(マップ上のポップアップメニューから)

系統別の配色の変更の項目を選択すると，配色の編集ウィンドウが表示されます(図 25).

図 25 のウィンドウを操作することで，系統ごとに配色を変更することができます．図 25 のウィンドウ上部の(1)項部分では，生物系統の一覧がチェックボックスとして表示されます．ここで配色の変更を行いたい系統を選択します．複数選択することも可能です．図 25 のウィンドウ中央部(2)項部分では，変更後の色を選択します．最後に(3)項部分で[OK]を選択することでマップ上の系統の配色が変更されます．なお，選択した系統を含む黒色の格子点も変更後の色で上書きされるため，上書きされた黒色の格子点を復元するには，[編集・分析]メニューを選択し，[配色の変更(全体)]内の [Only Black]ボタンを追加で選択してください.

マップ全体の配色を変更するには，[編集・分析]メニューの[配色の変更(全体)]を選択します(図 23)．全体の配色の変更の項目を選択すると，配色の編集ウィンドウが表示されます(図 26)．[Default Color]ボタンを選択すると初期設定の配色でマップ全体を描画します．[All White]ボタンを選択するとマップ全体を白色で描画します．　[Only Black]ボタンを選択するとマップの黒色の格子点のみを追加で描画します.

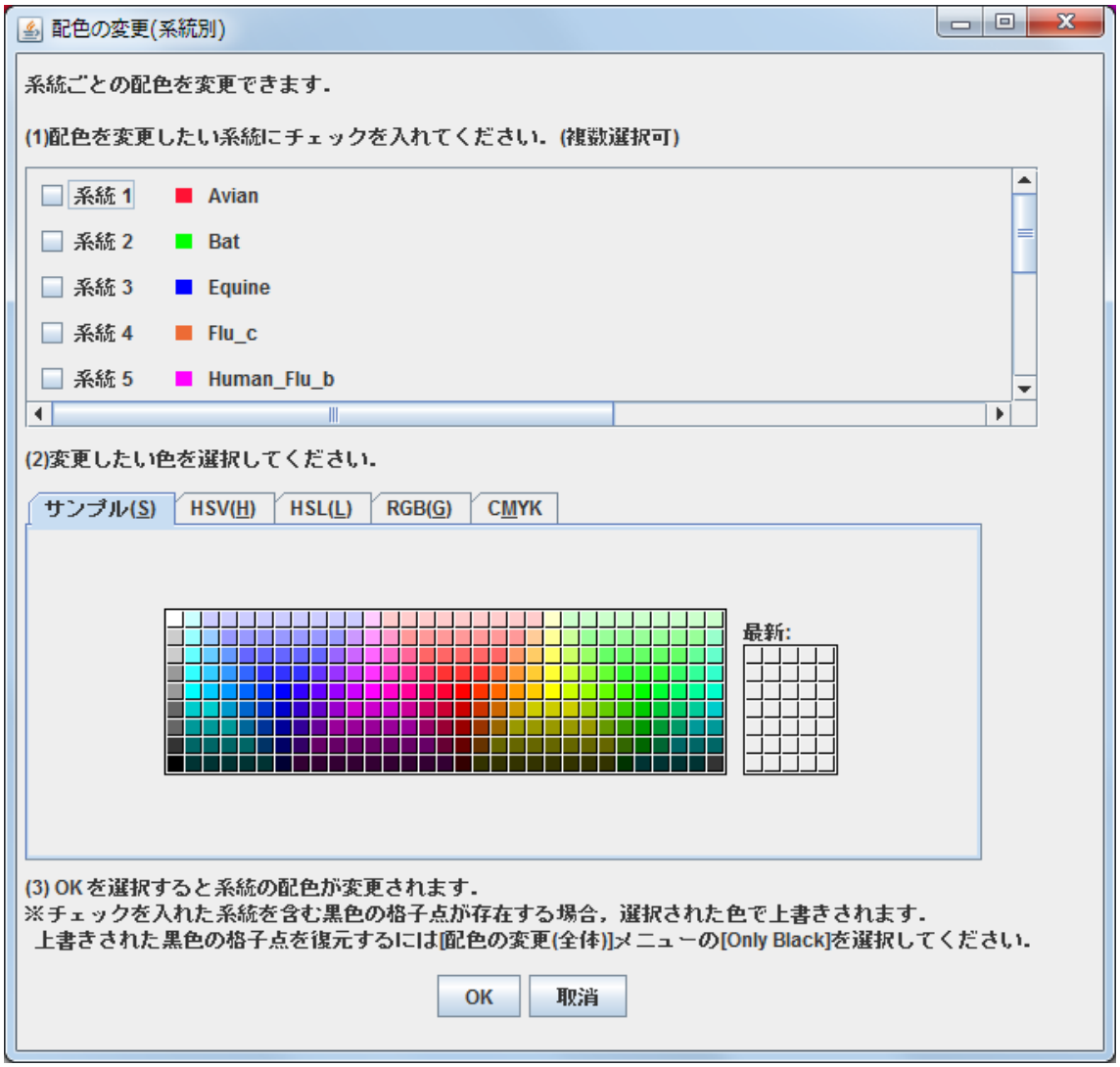


図 25　配色の編集ウィンドウ(系統別)

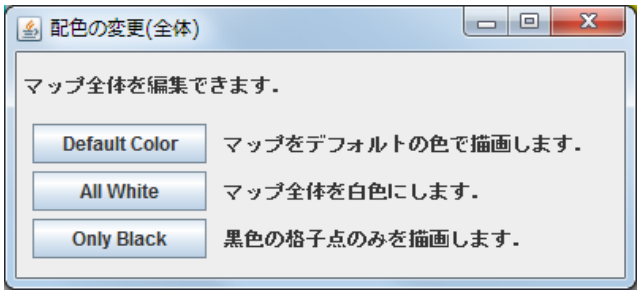


図 26　配色の編集ウィンドウ(全体)

7.1.8. データの保存

マップに関する各種データの保存は図 14 のマップウィンドウ上部のファイルメニュー(図 27)やマップ上で右クリックを行うことにより表示されるポップアップメニュー(図 28)等から実行することができます.

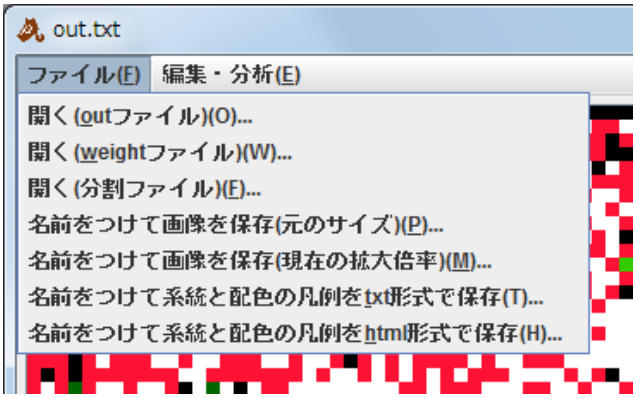


図 27　ファイルメニュー

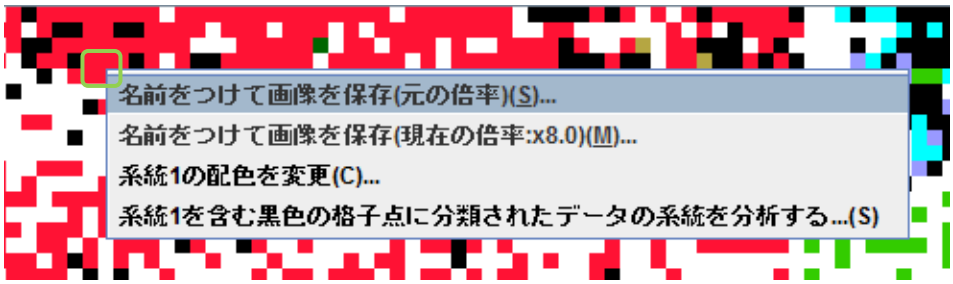


図 28　ポップアップメニュー

7.1.8.1. マップの保存

マップの保存を行うことができます．保存を行うには，図 14 のウィンドウ上の[ファイルメニュー]から[名前をつけて画像を保存]を選択します(図 27)．または，マップ上で右クリックをし，表示されるポップアップメニューの[名前をつけて画像を保存]を選択します(図 28)．保存されるファイル名は，初期設定では「(入力ファイル名)_blsomMap.png」になります.

7.1.8.2. 生物系統別の配色の凡例の保存

生物系統別の配色の凡例(7.1.3.節参照)を保存することができます．保存を行うには図 14 のウィンドウ上の[ファイルメニュー]から[名前をつけて系統と配色の凡例を txt 形式で保存]，または[名前をつけて系統と配色の凡例を html 形式で保存]を選択します(図 27)．保存されるファイル名は，初期設定では，「(入力ファイル名)_colorData.txt」，または「(入力ファイル名)_colorData.html」になります．

図 27 のファイルメニューにおいて，[名前をつけて系統と配色の凡例を txt 形式で保存]を選択すると，系統名と色情報(RGB 値，HTML カラーコード)をテキスト形式で保存します(図 29)．RGB 値とは，色の赤，緑，青の各要素を，10 進数の 0～255 の値で指定し，この 3 つの要素の組み合わせによって色を表現する値です．また，HTML カラーコードとは，色の赤，緑，青の各要素を「0x」以下から 2 桁ずつ 16 進数(0～9＋A~F)の値で指定し，この「0x」以下の 6 桁の組み合わせによって色を表現する文字列です．主に，HTML 形式において色を表現するために用いられます．

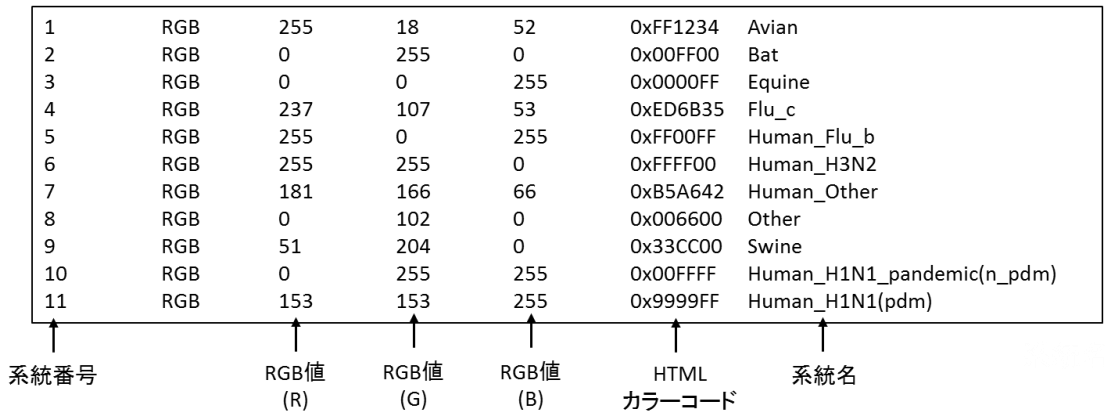


図 29　系統と配色の凡例の出力形式(テキスト形式)

図 27 のファイルメニューにおいて，[名前をつけて系統と配色の凡例を html 形式で保存]を選択すると，系統名と色情報(RGB 値，HTML カラーコード)を HTML 形式で保存します(図 30)．HTML 形式で保存することによって，HTML 形式に対応している環境中において色情報をテキストに保持したまま凡例のデータを扱うことができます．保存されるファイルは，一覧と羅列の 2 パターンで凡例のデータが出力されます．図 30 の例で示すファイル上部では系統番号と配色，系統名が一覧で出力されます．図 30 の例で示すファイル下部では，系統名と配色がカンマ区切りで羅列され出力されます．

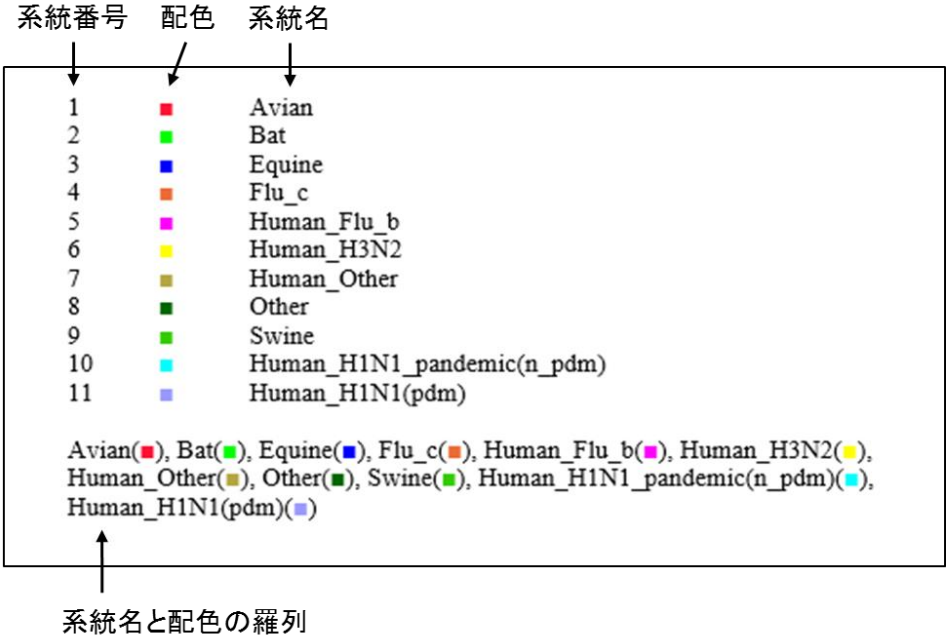


図 30　系統と配色の凡例の出力形式(HTML 形式)

7.2.　連続塩基出現頻度の学習結果の分析

連続塩基出現頻度の学習結果について分析を行うことができます．初期設定では「weight.final」というファイル名で出力されるデータです．

7.2.1.　ファイルの入力

連続塩基出現頻度の学習結果の分析を行うには，ソフトウェアのトップ画面の下部の[開く]からファイルを入力します(図 10).

7.2.1.1.　ファイルの種類とグラデーションの設定

図 10 のソフトウェアのトップ画面においてファイルを入力すると，ファイルの種類とグラデーションの設定ウィンドウが表示されます(図 31).

まず，図 31 上部においてファイルの種類の選択を行います．公開版の BLSOM を用いた連続塩基出現頻度の学習結果を入力している場合は，ウィンドウ内のメニューの[連続塩基出現度数の学習データ]を選択してください．

次に，図 31 下部においてグラデーションの段階数の選択を行います．初期設定では段階数は 11 に設定されています．

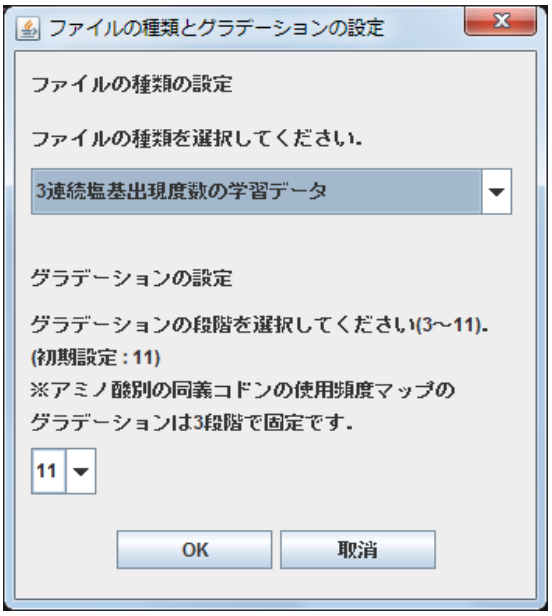


図 31 ファイルの種類とグラデーションの設定ウィンドウ

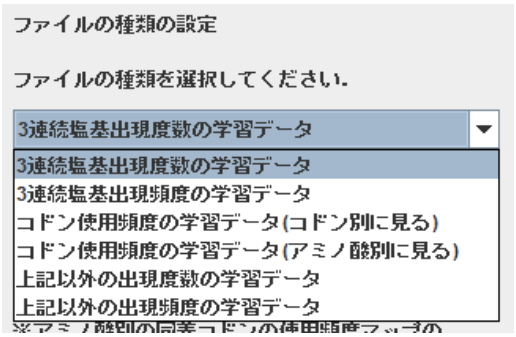


図 32　ファイルの種類の選択メニュー

※入力されたファイルが 64 次元データの場合は，図 32 のようにウィンドウ内のメニューの[コドンの使用頻度の学習データ]の項目が表示されます．コドン別の使用頻度マップの分析を行いたい場合は, 図 32 のメニューから[コドン使用頻度の学習データ(コドン別に見る)]を選択してください.

7.2.1.2. パターンの設定

ファイルの種類とグラデーションの設定が完了すると，入力ファイルの次元に対応するパターンの設定ウィンドウが表示されます(図 33).

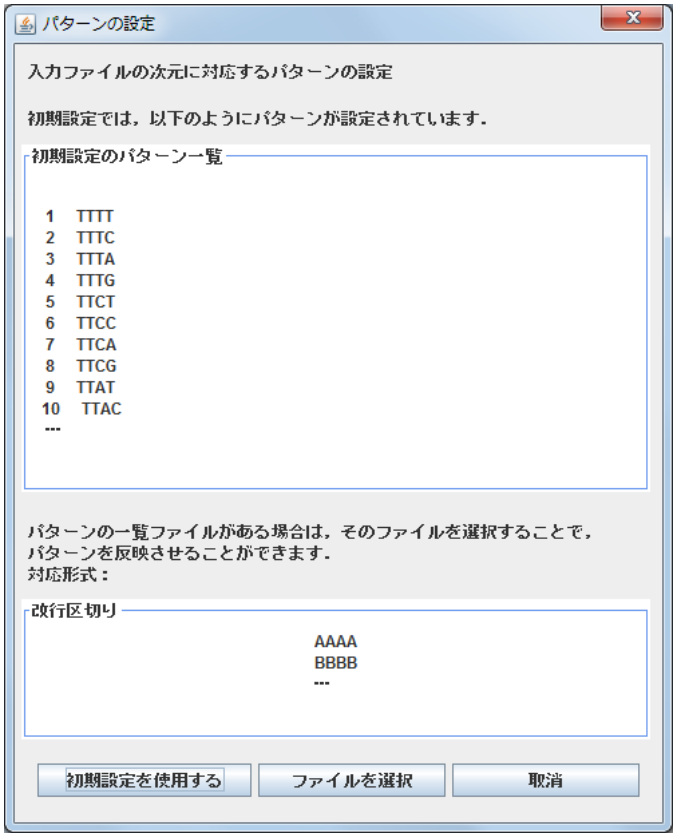


図 33　パターンの設定ウィンドウ

図 33 の例に示すようなウィンドウを操作することにより，入力ファイルの次元に対応するパターンの設定を行うことができます．図 33 のウィンドウ中の各項目について以下で説明します．

① [初期設定を使用する]の項目を選択すると，図 33 の上部に表示されるパターンの一覧を用いてマップが作成されます．
② [ファイルを選択]の項目を選択すると，パターンの一覧が書き込まれたファイルを選択するダイアログが表示されます．対応するパターンの一覧ファイルの形式は改行区切りになっているファイルです．なお，入力データの次元数に対し，パターンファイル内の項目数が多い場合はエラーになります．

7.2.1.3. ファイルの読み込み方法の設定

パターンに関する設定が完了すると，ファイルの読み込み方法に関するウィンドウが表示されます(図 34).

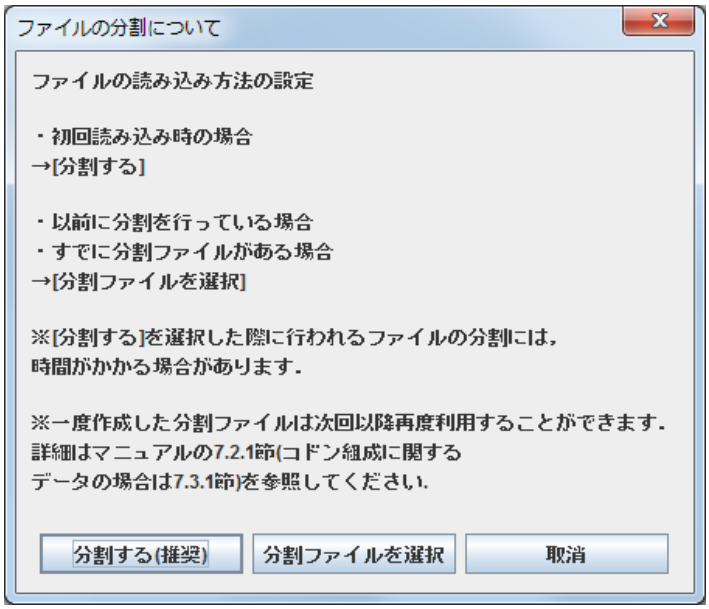


図 34　ファイルの読み込み方法に関するウィンドウ

初回読み込み時は，図 34 のウィンドウ中の[分割する]を選択します(7.2.1.節①項参照)．一度作成した分割ファイルは次回以降に再利用することができます(7.2.1.節②項参照)．

図 34 のウィンドウ中の各項目について以下で説明します．

① [分割する]の項目を選択すると，入力ファイルの分割を行います．なお，入力ファイルの容量やソフトウェアの動作環境によっては分割に時間がかかる場合があります．初期設定では「temp_blsomViewer」というディレクトリ以下に，「temp_(入力ファイル名)_freqencyData」というディレクトリが作成され，さらにこのディレクトリ以下に「temp_(入力ファイル名)_dim(入力ファイルの次元番号).txt」というファイル名で分割ファイルが入力ファイルの次元数分作成されます．作成された分割ファイルは，次回以降に[分割ファイルを選択する]を選択することによって再度利用することができます．分割処理中は進捗ウィンドウが表示されます(図 35)．

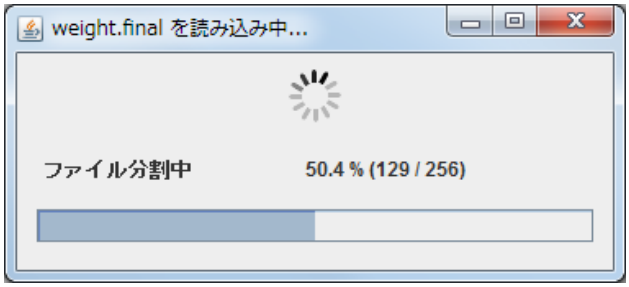


図 35　ファイル分割の進捗を表すウィンドウ

図 35 の進捗ウィンドウにおいて，中央の割合はファイル分割の進捗率を表します．また，カッコ内左の数値は現在読み込みが終わった次元数，右の数値は分割を行っているファイルの総次元数を表します．

② [分割ファイルを選択]の項目を選択すると，ディレクトリの選択ダイアログが表示されます．そこで，分割ファイルが存在するディレクトリを選択することで，分割ファイルを読み込むことができます．ディレクトリに置かれている分割ファイル名は，「temp_(入力ファイル名)_dim(入力ファイルの次元番号).txt」である必要があります．ソフトウェア上で作成された分割ファイルの場合，初期設定では，「temp_blsomViewer」というディレクトリ以下の，「temp_(入力ファイル名)_freqencyData」というディレクトリに格納されています．

7.2.2. 連続塩基出現頻度マップの表示

ファイルの入力完了後，ウィンドウ上に連続塩基組成のリストが表示されます(図 36)．

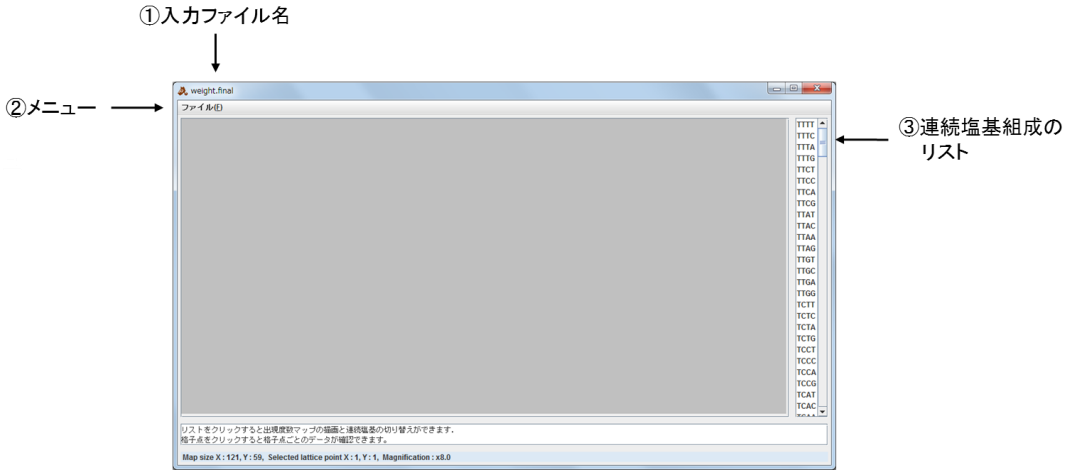


図 36　マップウィンドウ

図 36 のウィンドウ中の各項目について以下で説明します．

① ウィンドウ最上部のタイトルバーには入力ファイル名が表示されます．

② ウィンドウ上部のメニューを選択することで，データの保存が可能です(7.2.8.節参照)．

③ ウィンドウ右には，連続塩基組成のリストが表示されます．リストの項目は入力ファイルの次元数によって変化します(表 2)．このリストから着目したい項目を選択すると，ウィンドウ中央に選択された連続塩基組成の出現頻度マップが表示されます(図 37)．

表 2　入力ファイルの次元数と表示される連続塩基パターンの対応

入力ファイルの次元数	表示される項目
16	2 連続塩基組成
64	3 連続塩基組成
136	縮退 4 連続塩基組成
256	4 連続塩基組成
512	縮退 5 連連続塩基組成
1024	5 連続塩基組成
上記以外	次元番号

※図 32 において[コドン使用頻度の学習データ(コドン別に見る)]を選択した場合は，リスト内にはコドン組成の一覧が表示されます.

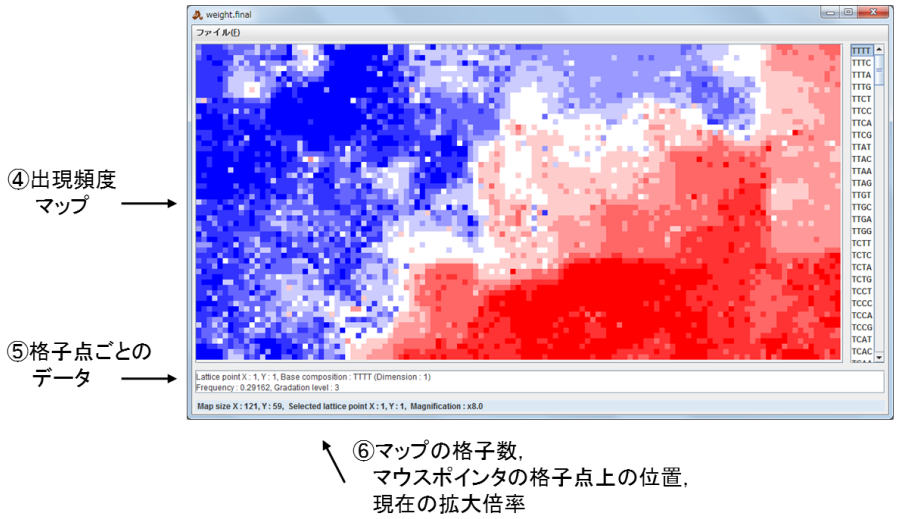


図 37　マップウィンドウ

図 37 のウィンドウ中の各項目について以下で説明します.

④ ウィンドウ中央では，連続塩基出現頻度マップが表示されます．マップは任意の倍率に拡大・縮小が可能です．また，マップがマップウィンドウに収まらなくなった場合，スクロールバーが表示され，マウス操作によってスクロールが可能です．マップ上の格子点を左クリックすると格子点ごとの出現頻度の値や，グラデーションの度合いをウィンドウ下部のテキストエリアに表示することができます(7.2.4.節参照).

⑤ ウィンドウ下部のテキストエリアでは，格子点ごとの出現頻度データが表示されます.

⑥ ウィンドウ最下部では，入力されたファイルのマップサイズ，現在マウスポインタが置かれている格子番号，および現在のマップの拡大倍率が表示されます(図 38).

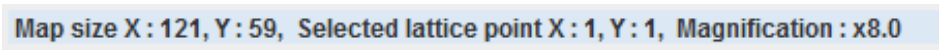


図 38　ウィンドウ最下部に表示されるデータ

7.2.3.　出現頻度に対する配色の凡例の表示

マップの表示と同時に出現頻度と配色の対応がマップウィンドウとは別ウィンドウで表示されます(図 39). 図 39 の例で示す凡例ウィンドウでは，グラデーションの度合い(値が大きいほど出現頻度が大きい)と出現頻度の高さに対する配色，およびその閾値が表示されます．この凡例データは出力することが可能です(7.2.5.節参照).

図 39　出現頻度に対する配色の凡例

7.2.4.　各格子点の出現頻度の表示

図 37 の例で示すウィンドウ上の出現頻度マップ上において，格子点を左クリックするとマップウィンドウ下部のテキストエリアが更新され，各格子点の出現頻度を確認できます(図 40). 図 40 のテキストエリアでは, 1 行目に選択中の格子番号と選択中の連続塩基組成, 2 行目では，出現頻度とグラデーションの度合いが表示されます.

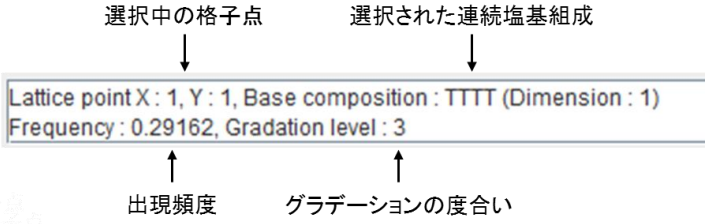


図 40　表示される出現頻度データ

7.2.5. データの保存

マップに関する各種データの保存は図 37 のマップウィンドウ上部のファイルメニュー(図 41)やマップやリスト上で右クリックを行うことにより表示されるポップアップメニュー(図 42，43)等から実行することができます.

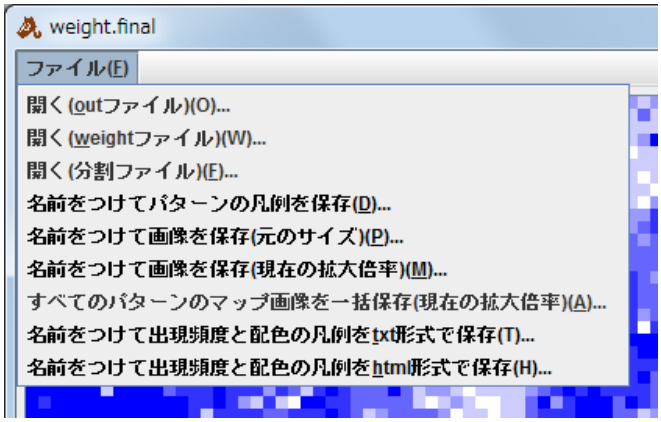


図 41　ファイルメニュー

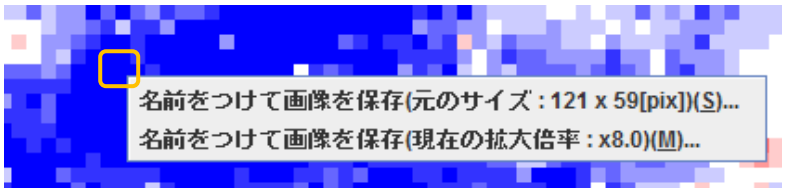


図 42　ポップアップメニュー

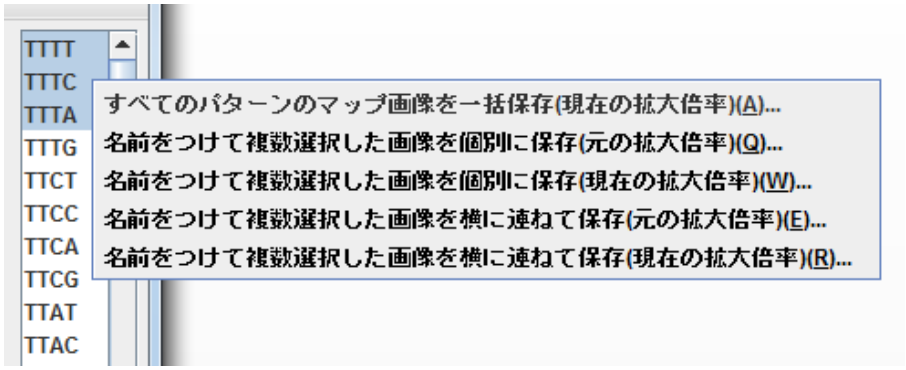


図 43　複数のマップを個別に保存

7.2.5.1. グラデーションの凡例画像の保存

図 37 のマップウィンドウが表示されると自動的に「temp_blsomVewer/temp_(入力ファイル名)_pic/」のディレクトリ下に[gradationPic.png]というファイル名でグラデーションの凡例画像が保存されます(図 44). グラデーションの段階数は 7.2.1.節で指定した段階数になります.

図 44　グラデーションの凡例画像(11 段階に設定した際の例)

7.2.5.2. 連続塩基組成の一覧の保存

連続塩基組成の一覧の保存を行うことができます. 保存を行うには図 37 のウィンドウ上の[ファイルメニュー]の[名前をつけて次元の凡例を保存]を選択します(図 41). 保存されるファイル名は, 初期設定では「(入力ファイル名)_ dimensionData.txt」になります.

7.2.5.3. マップの保存

単体のマップを個別に保存することができます. 保存を行うには, 図 41 の[ファイルメニュー]の[名前をつけて画像を保存]を選択します. または, マップ上で右クリックをし, 表示されるポップアップメニューの[名前をつけて画像を保存]を選択 (図 42)します. 保存される画像ファイル名は, 初期設定では「(入力ファイル名)_freqMap_(選択された連続塩基組成).png」になります.

また, すべてのパターンのマップを一括で保存することができます. 保存を行うには図 41 の[ファイルメニュー]の[すべてのパターンのマップ画像を一括保存]を選択します. または, 図 37 のマップウィンドウ右のリスト上で右クリックをし, 表示されるポップアップメニューの[すべてのパターンのマップ画像を一括保存]を選択します(図 43). 保存される画像ファイル名は, 初期設定ではそれぞれ, 「(入力ファイル名)_freqMap_(連続塩基組成).png」になります. 次元数やマップのサイズによっては時間がかかる場合があります.

さらに, 複数のマップを個別に保存することができます. 保存を行うにはまず, マップウィンドウ右のリストから保存を行いたい連続塩基組成を複数選択します. 次に複数選択がされた状態のリスト上で右クリックをし, 表示されるポップアップメニューの[名前をつけて複数選択した画像を個別に保存]を選択します(図 43). 保存される画像ファイル名は,

初期設定ではそれぞれ，「(入力ファイル名)_freqMap_(選択された連続塩基組成).png」になります．

加えて，複数のマップを横に連ねて保存することができます．保存を行うにはまず，マップウィンドウ右のリストから保存を行いたい連続塩基組成を複数選択します．次に複数選択がされた状態のリスト上で右クリックをし，表示されるポップアップメニューの[名前をつけて複数選択した画像を横に連ねて保存]を選択します(図 43)．

[名前をつけて複数選択した画像を横に連ねて保存]を選択すると入力ダイアログが表示されます(図 45)．図 45 中のダイアログの[画像間の余白の幅]の項目では，図 46 の①部分の余白の幅を指定することができます．[画像の枠線の幅]の項目では，図 46 の②の部分の画像の枠線の幅を指定することができます．幅に 0pix を指定すると枠線なしのマップになります．[画像の枠線の色]の項目では，画像の枠線の幅を 1pix 以上に指定した場合，枠線の色を選択することができます．

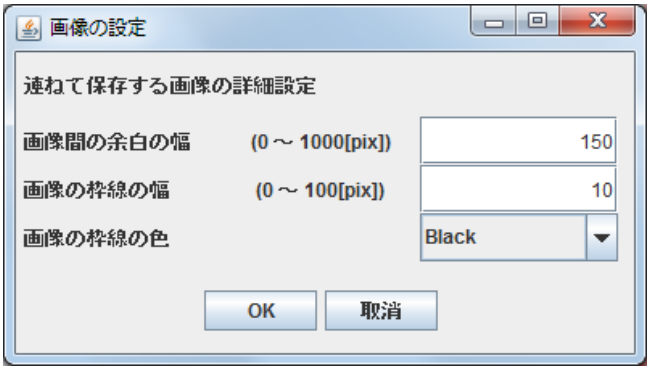


図 45　複数のマップを横に連ねて保存する際に表示される入力ダイアログ

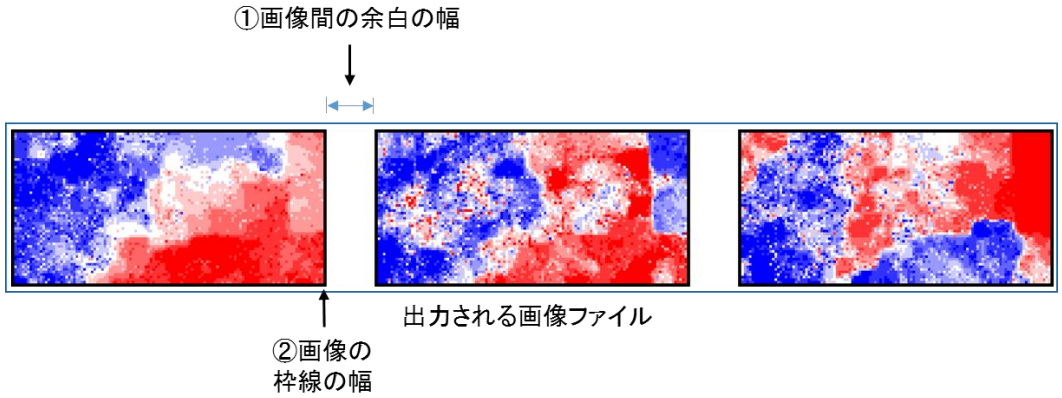


図 46　複数のマップを横に連ねて保存する場合に指定できる項目

保存される画像ファイル名は，初期設定では，「(入力ファイル名)_freqMap_(選択された複数の連続塩基組成).png」になります．保存に失敗する場合は並べるマップ数が多いか，マップの拡大倍率が高すぎることが原因として考えられます．マップ数を減らすか，マップの拡大倍率を低くした上で再度保存をお試しください．

7.2.5.4. 出現頻度に対する配色の凡例の保存

出現頻度に対する配色の凡例(7.2.3.節参照)を保存することができます．保存を行うには[ファイルメニュー]の[名前をつけて出現頻度と配色の凡例を txt 形式で保存]，または[名前をつけて出現頻度と配色の凡例を html 形式で保存]を選択します(図 41)．保存されるファイル名は，初期設定では，「(入力ファイル名)_freqData_(選択した連続塩基組成).txt」，または「(入力ファイル名)_freqData_(選択した連続塩基組成).html」になります．

図 41 のファイルメニューにおいて，[名前をつけて出現頻度と配色の凡例を txt 形式で保存]を選択すると，色情報(RGB 値，HTML カラーコード(7.1.8.2.節参照))と出現頻度の閾値をテキスト形式で保存します(図 47)．

11	RGB	255	0	0	0xFF0000	0.50085	0.65647
10	RGB	255	51	51	0xFF3333	0.48277	0.50085
9	RGB	255	102	102	0xFF6666	0.43611	0.48277
8	RGB	255	153	153	0xFF9999	0.40212	0.43611
7	RGB	255	204	204	0xFFCCCC	0.36871	0.40212
6	RGB	255	255	255	0xFFFFFF	0.33238	0.36871
5	RGB	204	204	255	0xCCCCFF	0.31000	0.33238
4	RGB	153	153	255	0x9999FF	0.29894	0.31000
3	RGB	102	102	255	0x6666FF	0.28369	0.29894
2	RGB	51	51	255	0x3333FF	0.26158	0.28369
1	RGB	0	0	255	0x0000FF	0.15417	0.26158
グラデーションの度合い		RGB値 (R)	RGB値 (G)	RGB値 (B)	HTML カラーコード	閾値の下限	閾値の上限

図 47　出現頻度に対する配色の凡例の出力形式(テキスト形式)

図 41 のファイルメニューにおいて，[名前をつけて系統と配色の凡例を html 形式で保存]を選択すると，色情報と出現頻度の閾値を HTML 形式で保存します(図 48)．HTML 形式で保存することによって，HTML 形式に対応している環境中において色情報をテキストに保持したまま凡例のデータを扱うことができます．

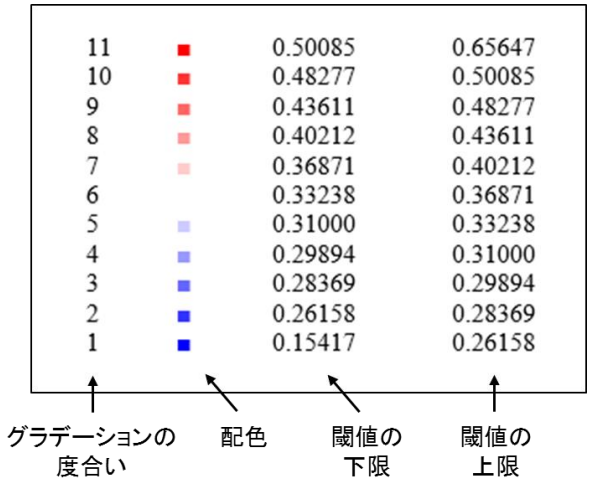


図 48　出現頻度に対する配色の凡例の出力形式(HTML 形式)

7.3.　コドン使用頻度の学習結果の分析

本ソフトウェアはコドンの使用頻度データにも対応しているため，BLSOM によるコドンの使用頻度の学習結果についても分析を行うことができます．コドン使用頻度マップにはコドン別の使用頻度マップとアミノ酸別の同義コドン使用頻度マップの 2 種類があります．前者のコドン別の使用頻度マップの分析機能については，7.3.2.節のコドン表の表示以外は 7.2.節の連続塩基出現頻度マップの分析機能の操作方法とほぼ同様の仕様のため，本節では主に後者のアミノ酸別の同義コドンの使用頻度マップの分析機能について説明します．

7.3.1.　ファイルの入力

コドン使用頻度の学習結果の分析を行うには，ソフトウェアのトップ画面の下部の[開く]からファイルを入力します．詳細については 7.2.1.節および 7.2.1.1.節をご覧下さい．

図 32 のウィンドウ中のメニューの[コドン使用頻度の学習データ(コドン別に見る)]を選択するとコドン別の使用頻度マップを分析することができます．なお，コドン別の使用頻度マップの分析機能については，7.3.2 節のコドン表の表示以外は 7.2.節の連続塩基出現頻度マップの分析機能の操作方法とほぼ同様の仕様のため，詳細については 7.2.節をご覧ください．

図 32 のウィンドウ中のメニューの[コドン使用頻度の学習データ(アミノ酸別に見る)]を選択するとアミノ酸別の同義コドンの使用頻度マップを分析することができます．

7.3.2.　コドン表の表示

7.2.1.節における一連の設定が完了すると，コドン表が表示されます(図 49)．図 49 のコドン表内のコドン(緑色の部分)を選択すると選択されたコドンの使用頻度マップを表示できます．また，図 49 のコドン表内のアミノ酸(青色の部分)をクリックすると選択されたアミノ酸の同義コドンの使用頻度マップを表示できます．

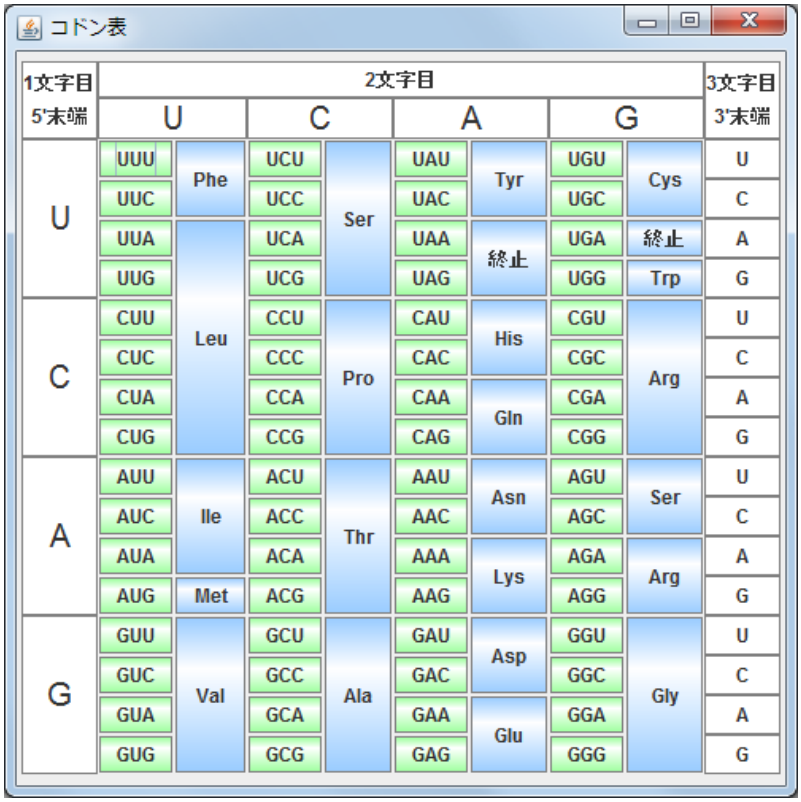


図 49　表示されるコドン表

7.3.3.　アミノ酸別の同義コドンの使用頻度マップの表示

7.2.1.節における一連の設定が完了すると，ウィンドウ上にアミノ酸のリストが表示されます(図 50).

図 50　マップウィンドウ

図 50 のウィンドウ中の各項目について以下で説明します.

① ウィンドウ最上部のタイトルバーには入力ファイル名が表示されます.
② ウィンドウ上部のメニューを選択することで，データの保存が可能です(7.3.6.節参照).
③ ウィンドウ右には，アミノ酸 20 種＋終止コドンのリストが表示されます．このリストから着目したい項目を選択すると，ウィンドウ中央に選択されたアミノ酸の同義コドン使用頻度マップが表示されます(図 51).

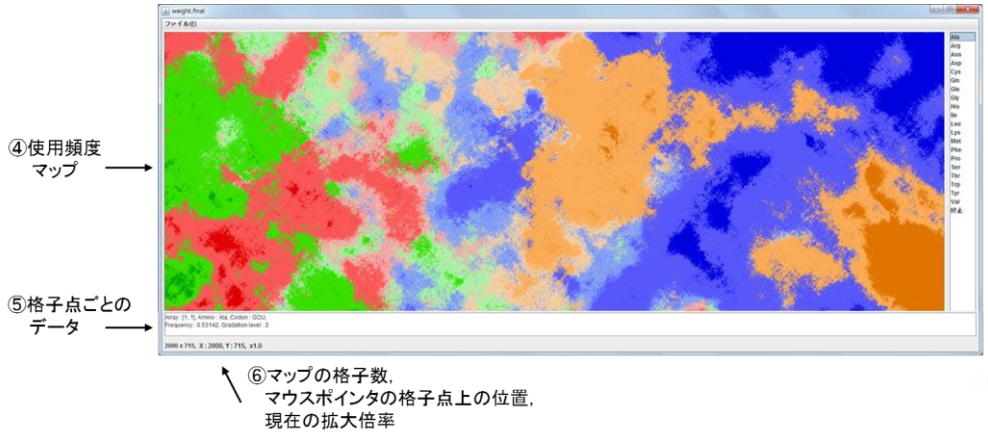


図 51　マップウィンドウ

図 51 のウィンドウ中の各項目について以下で説明します.

④ ウィンドウ中央では，アミノ酸別の同義コドンの使用頻度マップが表示されます．マップは任意の倍率に拡大・縮小が可能です．また，マップがマップウィンドウに収まらなくなった場合，スクロールバーが表示され，マウス操作によってスクロールが可能です.

マップ上の格子点を左クリックすると格子点ごとの使用頻度の値や，グラデーションの度合いをウィンドウ下部のテキストエリアに表示することができます(7.3.5.節参照).

⑤ ウィンドウ下部のテキストエリアでは，格子点ごとの使用頻度データが表示されます.

⑥ ウィンドウ最下部では，入力されたファイルのマップサイズ，現在マウスポインタが置かれている格子番号，および現在のマップの拡大倍率が表示されます(図 52).

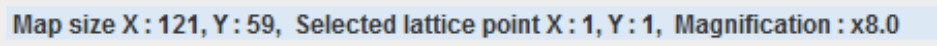


図 52　ウィンドウ最下部に表示されるデータの例

7.3.4.　使用頻度に対する配色の凡例の表示

マップの表示と同時に使用頻度と配色の対応がマップウィンドウとは別ウィンドウで表示されます(図 53). 図 53 の例で示す凡例ウィンドウでは，同義コドンとグラデーションの度合い(値が大きいほど使用頻度が大きい)，同義コドンに対応する配色とグラデーション，およびその閾値が表示されます. この凡例データは出力することが可能です(7.3.6.節参照).

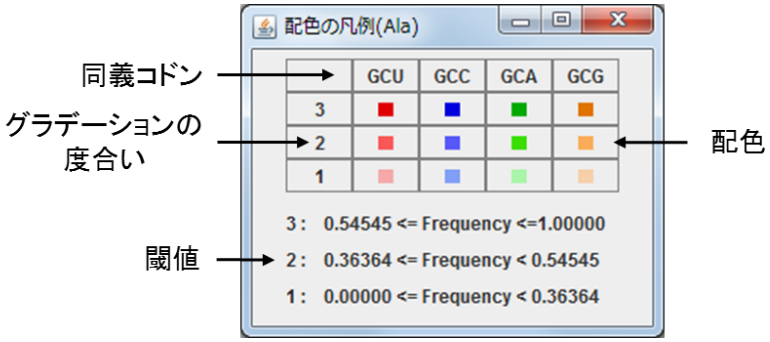


図 53　使用頻度に対する配色の凡例

7.3.5.　各格子点の使用頻度の表示

図 51 の例で示すウィンドウ上の使用頻度マップ上において, 格子点を左クリックすると，マップウィンドウ下部のテキストエリアが更新され，各格子点の使用頻度を確認できます(図 54). 図 54 のテキストエリアでは，1 行目に選択中の格子番号と選択中のアミノ酸，および選択中の格子点において最も使用頻度が高かったコドンが表示されます. 2 行目では，使用頻度とグラデーションの度合いが表示されます.

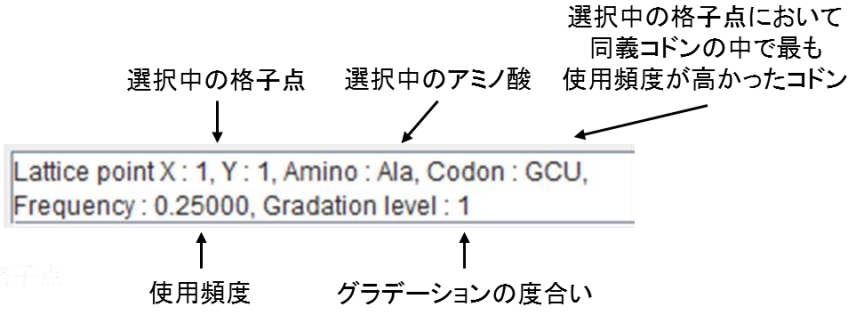


図 54　表示される使用頻度データ

7.3.6.　データの保存

マップに関する各種データの保存は図 51 のマップウィンドウ上部のファイルメニュー(図 55)等から実行することができます.

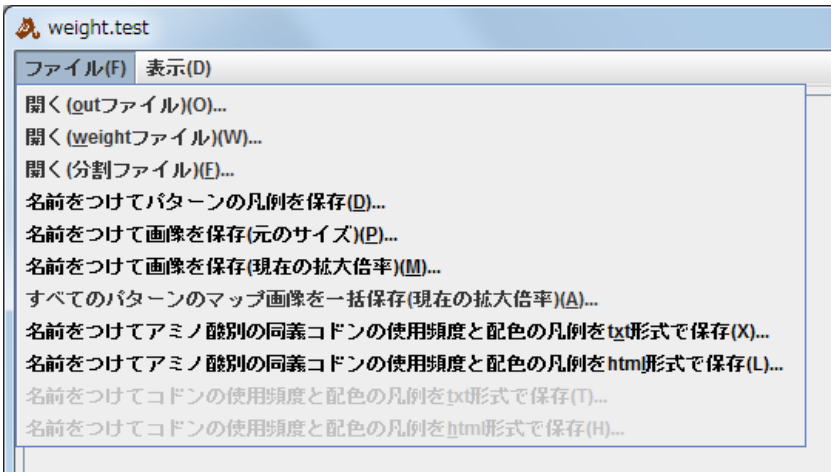


図 55　アミノ酸の一覧の保存

7.3.6.1.　アミノ酸の一覧の保存

アミノ酸の一覧の保存を行うことができます．保存を行うには[ファイルメニュー]の[名前をつけて次元の凡例を保存]を選択します(図 55)．保存されるファイル名は，初期設定では「(入力ファイル名)_ dimensionData.txt」になります.

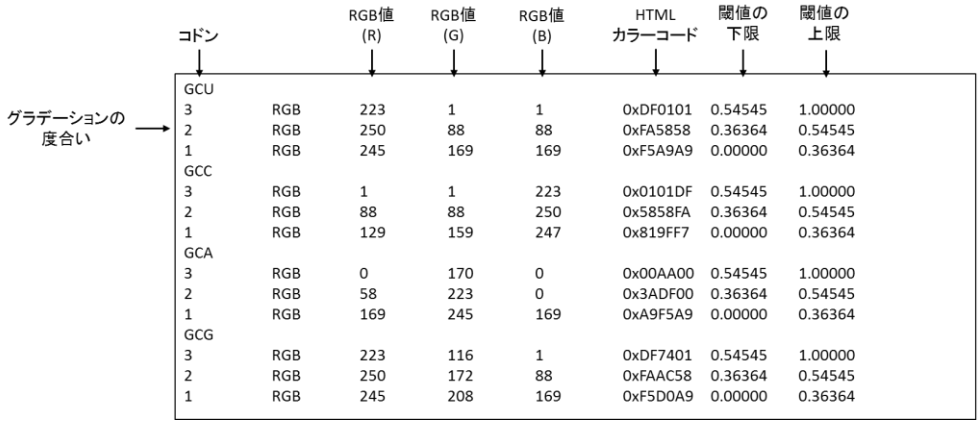
7.3.6.2.　マップの保存

マップの保存については 7.2.5.3.節を参照してください.

7.3.6.3. 使用頻度に対する配色の凡例の保存

同義コドン別の使用頻度に対する配色の凡例(7.3.4.節参照)を保存することができます. 保存を行うには[ファイルメニュー]の[名前をつけてアミノ酸別の同義コドンの使用頻度と配色の凡例を txt 形式で保存]，または[名前をつけてアミノ酸別の同義コドンの使用頻度と配色の凡例を html 形式で保存]を選択します(図 55)．保存されるファイル名は，初期設定では，「(入力ファイル名)_freqData_(選択したアミノ酸).txt」，または「(入力ファイル名)_freqData_(選択したアミノ酸).html」になります.

[名前をつけてアミノ酸別の同義コドンの使用頻度と配色の凡例を txt 形式で保存]を選択すると，同義コドンごとの色情報(RGB 値，HTML カラーコード(7.1.8.2.節参照))と使用頻度の閾値をテキスト形式で保存します(図 56).



GCU							
3	RGB	223	1	1	0xDF0101	0.54545	1.00000
2	RGB	250	88	88	0xFA5858	0.36364	0.54545
1	RGB	245	169	169	0xF5A9A9	0.00000	0.36364
GCC							
3	RGB	1	1	223	0x0101DF	0.54545	1.00000
2	RGB	88	88	250	0x5858FA	0.36364	0.54545
1	RGB	129	159	247	0x819FF7	0.00000	0.36364
GCA							
3	RGB	0	170	0	0x00AA00	0.54545	1.00000
2	RGB	58	223	0	0x3ADF00	0.36364	0.54545
1	RGB	169	245	169	0xA9F5A9	0.00000	0.36364
GCG							
3	RGB	223	116	1	0xDF7401	0.54545	1.00000
2	RGB	250	172	88	0xFAAC58	0.36364	0.54545
1	RGB	245	208	169	0xF5D0A9	0.00000	0.36364

図 56　使用頻度に対する配色の凡例の出力形式(テキスト形式)

[名前をつけてアミノ酸別の同義コドンの使用頻度と配色の凡例を html 形式で保存]を選択すると，同義コドンごとの色情報と使用頻度の閾値を HTML 形式で保存します(図 57)．HTML 形式で保存することによって，HTML 形式に対応している環境中において色情報をテキストに保持したまま凡例のデータを扱うことができます.

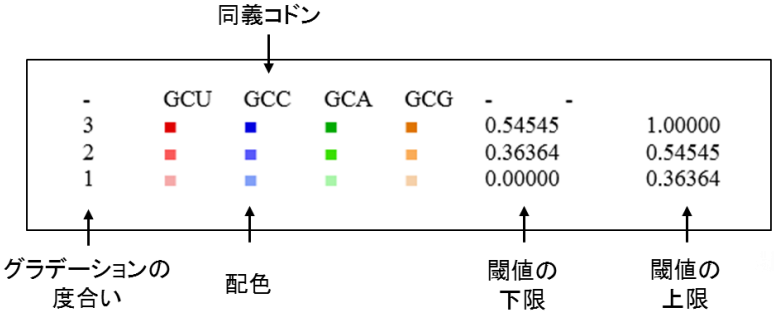


図 57　使用頻度に対する配色の凡例の出力形式(HTML 形式)

8.　参考文献

[1] Kanaya S, Kinouchi M, Abe T, Kudo Y, Yamada Y, Nishi T, Mori H, Ikemura T (2001) Analysis of codon usage diversity of bacterial genes with a self-organizing map: characterization of horizontally transferred genes with emphasis on E. coli O157 genome. Gene, 276:89-99.

[2] Abe T, Kanaya S, Kinouchi M, Ichiba Y, Kozuki T, Ikemura T (2003) Informatics for unveiling hidden genome signatures. Genome Research, 13:693-702.

[3] Abe T, Sugawara H, Kinouchi M, Kanaya S, Ikemura T (2005) Novel Phylogenetic Studies of Genomic Sequence Fragments Derived from Uncultured Microbe Mixtures in Environmental and Clinical Samples. DNA research, 12:281-290.

[4] Abe T, Sugawara H, Kanaya S, Ikemura T (2006) Sequences from almost all prokaryotic, eukaryotic, and viral genomes available could be classified according to genomes on a large-scale Self-Organizing Map constructed with the Earth Simulator. Journal of the earth simulator, 6:17-23.

[5] Abe T, Kanaya S, Uehara H, Ikemura T (2009) A novel bioinformatics strategy for function prediction of poorly-characterized protein genes obtained from metagenome analyses. DNA Research, 16:287-298.

[6] (Review) Iwasaki Y, Abe K, Wada K, Wada Y, and Ikemura T. (2013) A Novel Bioinformatics Strategy to Analyze Microbial Big Sequence Data for Efficient Knowledge Discovery: Batch-Learning Self-Organizing Map (BLSOM). Microorganisms, 1:137-157.