

○ ファイル「consensus.cons」について

consensus.cons はコンセンサス塩基配列の各塩基の組成等、ベースコールされた経緯に関する情報等が記述されます。ファイルの内容は以下のようになります。

```
49      ... [コンセンサス塩基配列長]
3      ... [CoverageDepth 最大値]
0  0  0  0  0  0  0  0  0  0  "n" "n" "n" 0  0  ... [塩基数情報]  1 塩基目
0  0  0  0  0  0  0  0  0  0  "n" "n" "n" 0  0  ... [塩基数情報]  2 塩基目
0  0  0  0  0  0  0  0  0  0  "n" "n" "n" 0  0  ... [塩基数情報]  3 塩基目
0  0  0  0  0  0  0  0  0  1  "n" "n" "n" 0  0  ... [塩基数情報]  4 塩基目
0  0  0  0  0  0  0  0  0  1  "n" "n" "n" 0  0  ... [塩基数情報]  5 塩基目
1  0  0  0  0  0  0  0  0  1  "n" "a" "n" 0  9  ... [塩基数情報]  6 塩基目
1  0  0  0  0  0  0  1  0  1  "n" "T" "n" 0  9  ... [塩基数情報]  7 塩基目
1  0  0  0  0  0  0  1  0  1  "n" "T" "n" 0 10  ... [塩基数情報]  8 塩基目
0  3  0  0  0  0  0  0  0  0  "n" "A" "n" 0 10  ... [塩基数情報]  9 塩基目
....
....
```

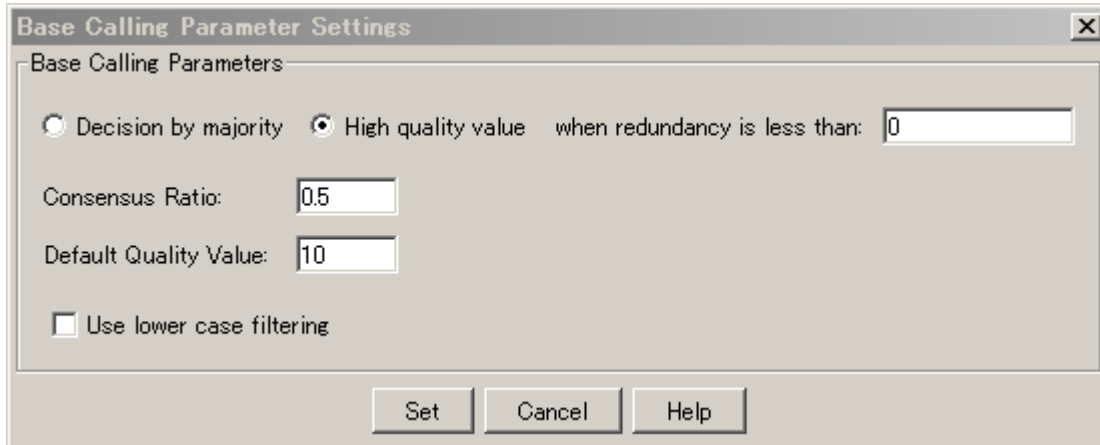
[コンセンサス塩基配列長]

[CoverageDepth 最大値]

[塩基数情報 (1 5 項目、タブ区切り)]

column01 ...塩基 a の数
column02 ...塩基 A の数
column03 ...塩基 c の数
column04 ...塩基 C の数
column05 ...塩基 g の数
column06 ...塩基 G の数
column07 ...塩基 t の数
column08 ...塩基 T の数
column09 ...ギャップの数
column10 ...その他 (塩基 N) の数
column11 ...規定塩基
column12 ...ベースコール結果
column13 ...クオリティ最大値を持つ塩基
column14 ...クオリティ最大値
column15 ...ベースコール結果に対するクオリティ値

○ ベールコール関連オプションについて



The image shows a Windows-style dialog box titled "Base Calling Parameter Settings". It contains several settings for base calling. At the top, there are two radio buttons: "Decision by majority" and "High quality value". The "High quality value" option is selected. To the right of this selection is a text field labeled "when redundancy is less than:" with the value "0". Below these are two more text fields: "Consensus Ratio:" with the value "0.5" and "Default Quality Value:" with the value "10". At the bottom left, there is a checkbox labeled "Use lower case filtering" which is currently unchecked. At the bottom right, there are three buttons: "Set", "Cancel", and "Help".

○ Decision by majority

多数決によるベースコール。 通常はこちらを選択します。

○ High quality value when redundancy is less than

指定された redundancy(coverage depth)値（規定値 3）よりも小さい場合には最もクオリティ値が高かった塩基をコールします。そうでない場合には多数決によるベースコールを行います。

■ Consensus Ratio（規定値 0.5）

コンセンサス割合指数値。

塩基数の多数決だけでベースコールしてしまうと、コールされた塩基の信頼性に問題が生じる事があります。

・ ケース 1 （塩基「A」がコールされます。信頼性は低い）

A: 40

G: 30

C: 30

T: 0

・ ケース 2 （塩基「A」コールされます。信頼性は高い）

A: 90

G: 5

C: 5

T: 0

2つのどちらのケースでもコールされる塩基は「A」となりますが、そのコールされた塩基に対する信頼性は異なります。そのため、GenomeTraveler ある一定以上の割合でベースコールされているか否かで、コンセンサス配列の品質を保つようにしています。

手順 1 多数決によるベースコール塩基の選出

はじめに大文字塩基のみでの多数決で塩基を選出してみます。大文字塩基のみの多数決で選出できなかった場合には大文字と小文字を合計した数での多数決で塩基を選出します。

手順 2 コンセンサス割合指数値の計算

手順 1 で選出されたベース塩基に対する塩基カウント数が占める割合を計算します。

Consensus Ratio = 選出された塩基のカウント数 / 塩基トータル数

条件をクリアした塩基がベースコールされます。
閾値よりも小さい場合は「N」となります。

- Default Quality Value （規定値 10）
クオリティデータがない場合に代わりに用いるクオリティ値。

- Use lower case filtering
チェックされている場合には、ベースコールを行う際に小文字の塩基を無視します。
たとえば、アセンブルやマッピングの際、実際に相同性があるアライメントもしくはマッピングされた領域の配列塩基は大文字で表記し、そうでない領域（クリッピング領域等）の塩基配列は小文字で表記していたりします。 クリッピング領域やクオリティが低く、ミスリードしている塩基をベースコールに関与させたくない場合等に利用できます。

