

## MacVector 基本操作(マルチプルアラインメント)

MacVector でマルチプルアラインメントをするために必要な操作の一部を紹介いたします。

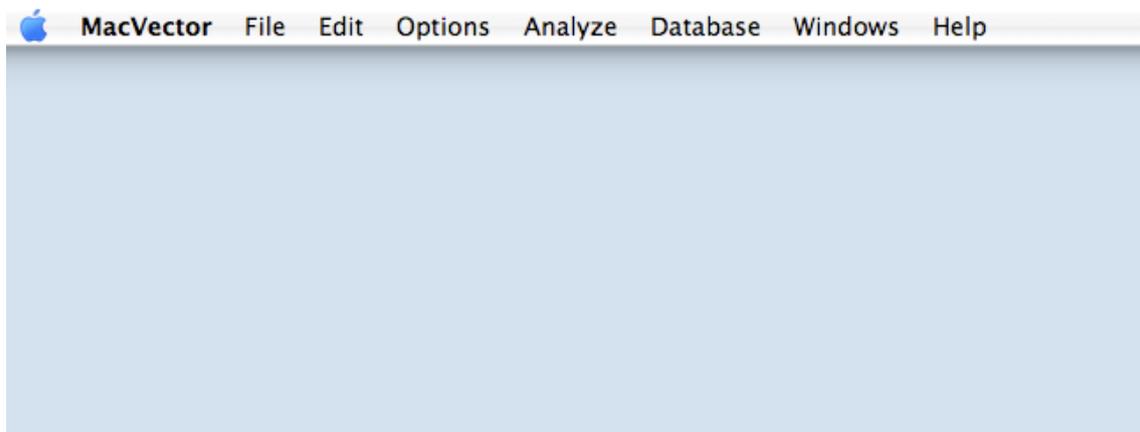
マルチプルアラインメントに関する主要な操作は下記のものであります。

- A. 配列情報のファイルの入手
- B. 配列情報ファイルの作成(新規)
- C. マルチプルアラインメント
- D. 系統樹の作成



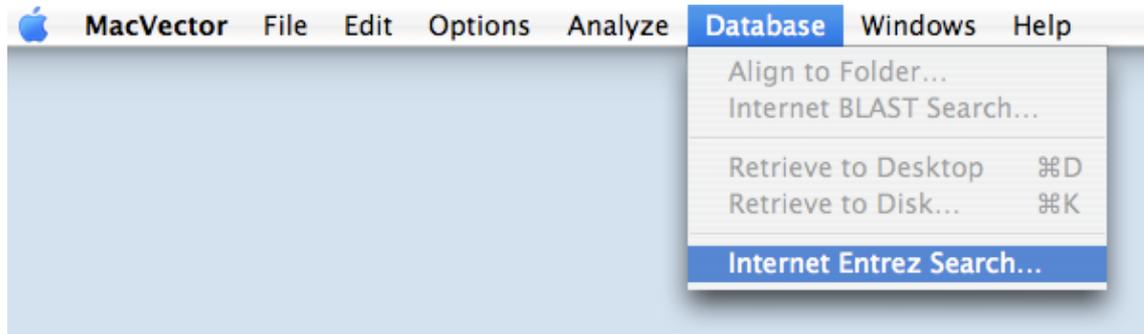
まず、MacVector を起動してください。

(注意！ MacVector は起動しただけでは何も新しいウィンドウは開きません。)



## A. 配列情報のファイルの入手

MacVector で利用する配列ファイルは、公知のデータの場合 NCBI の検索エンジンから入手することができます。



メニューバーから、Internet Entrez Search...を選択します。

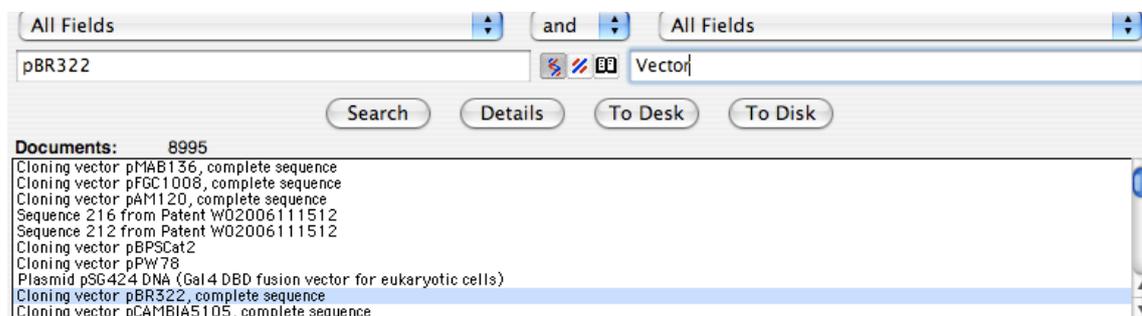


左右2つのカラムを利用して、検索するデータベースのフィールドを選択し、キーワードを選んでください。

2つのキーワードの組み合わせは、only、and、not、or の4つが可能です。

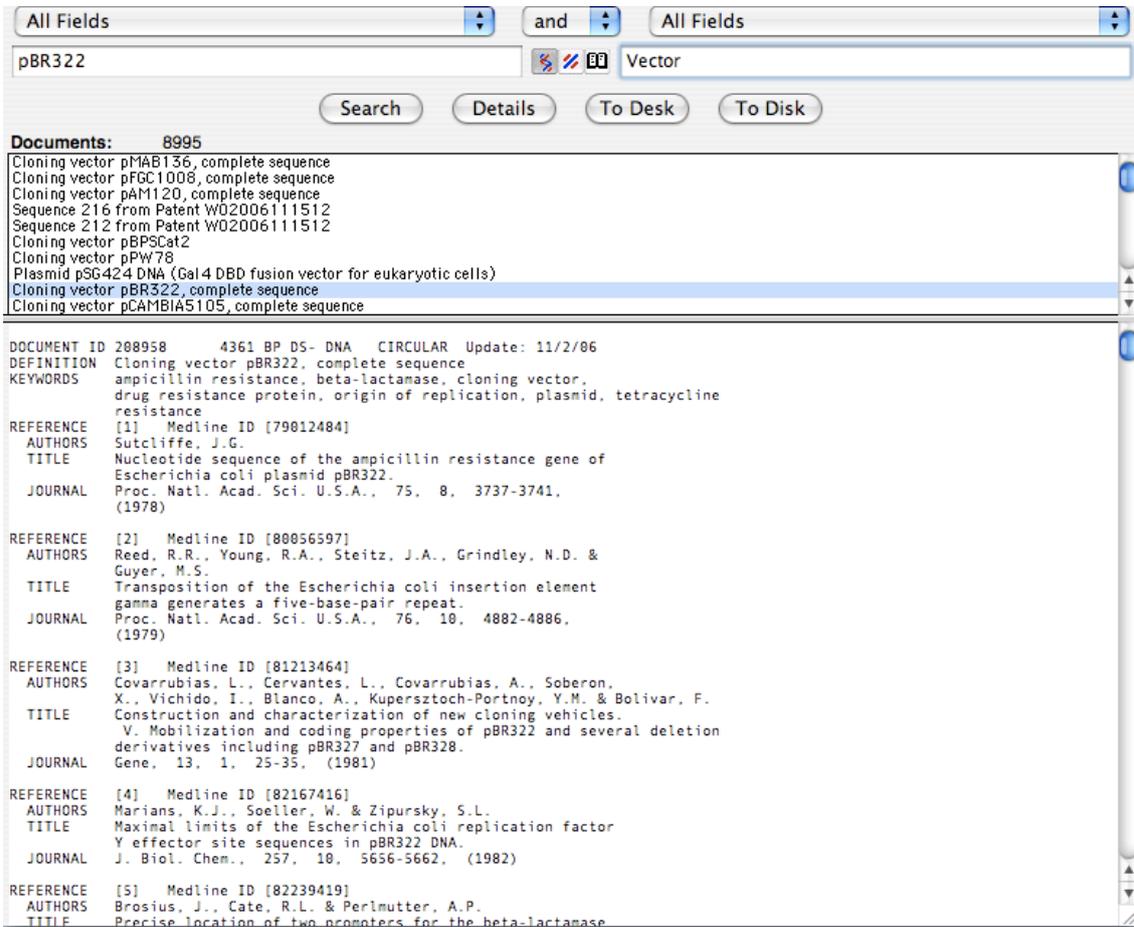
組み合わせ選択のカラムの下に、検索するデータの種別を指定するアイコンがあります。これを選んでください。(左から、核酸配列、アミノ酸配列、文献となります。)

以上の設定が終わったら、Search ボタンをクリックして検索開始です。



ヒットしたデータベースのエントリー(DE)が表示されます。

必要なファイルの入手方法は2つあります。

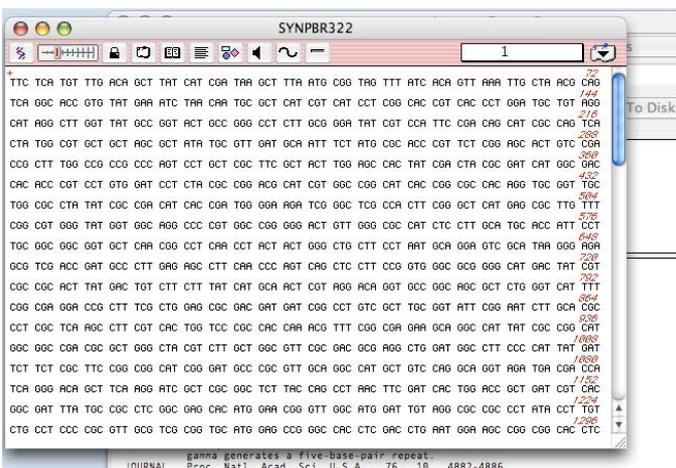


Documents: 8995

Cloning vector pMAB136, complete sequence  
 Cloning vector pFGC1008, complete sequence  
 Cloning vector pAM1120, complete sequence  
 Sequence 216 from Patent WO2006111512  
 Sequence 212 from Patent WO2006111512  
 Cloning vector pBPScaf2  
 Cloning vector pPW78  
 Plasmid pSG424 DNA (Gal4 DBD fusion vector for eukaryotic cells)  
 Cloning vector pBR322, complete sequence  
 Cloning vector pCAMBIA5105, complete sequence

DOCUMENT ID 208958 4361 BP DS- DNA CIRCULAR Update: 11/2/06  
 DEFINITION Cloning vector pBR322, complete sequence  
 KEYWORDS ampicillin resistance, beta-lactamase, cloning vector, drug resistance protein, origin of replication, plasmid, tetracycline resistance  
 REFERENCE [1] Medline ID [79012484]  
 AUTHORS Sutcliffe, J.G.  
 TITLE Nucleotide sequence of the ampicillin resistance gene of Escherichia coli plasmid pBR322.  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A., 75, 8, 3737-3741, (1978)  
 REFERENCE [2] Medline ID [80056597]  
 AUTHORS Reed, R.R., Young, R.A., Steitz, J.A., Grindley, N.D. & Guyer, M.S.  
 TITLE Transposition of the Escherichia coli insertion element gamma generates a five-base-pair repeat.  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A., 76, 10, 4882-4886, (1979)  
 REFERENCE [3] Medline ID [81213464]  
 AUTHORS Covarrubias, L., Cervantes, L., Covarrubias, A., Soberon, X., Vichido, I., Blanco, A., Kupersztoch-Portnoy, Y.M. & Bolivar, F.  
 TITLE Construction and characterization of new cloning vehicles. V. Mobilization and coding properties of pBR322 and several deletion derivatives including pBR327 and pBR328.  
 JOURNAL Gene, 13, 1, 25-35, (1981)  
 REFERENCE [4] Medline ID [82167416]  
 AUTHORS Mariani, K.J., Soeller, W. & Zipursky, S.L.  
 TITLE Maximal limits of the Escherichia coli replication factor Y effector site sequences in pBR322 DNA.  
 JOURNAL J. Biol. Chem., 257, 10, 5656-5662, (1982)  
 REFERENCE [5] Medline ID [82239419]  
 AUTHORS Brosius, J., Cate, R.L. & Perlmutter, A.P.  
 TITLE Precise location of two promoters for the beta-lactamase

まず、内容の確認を“Details”をクリックして行ってください。

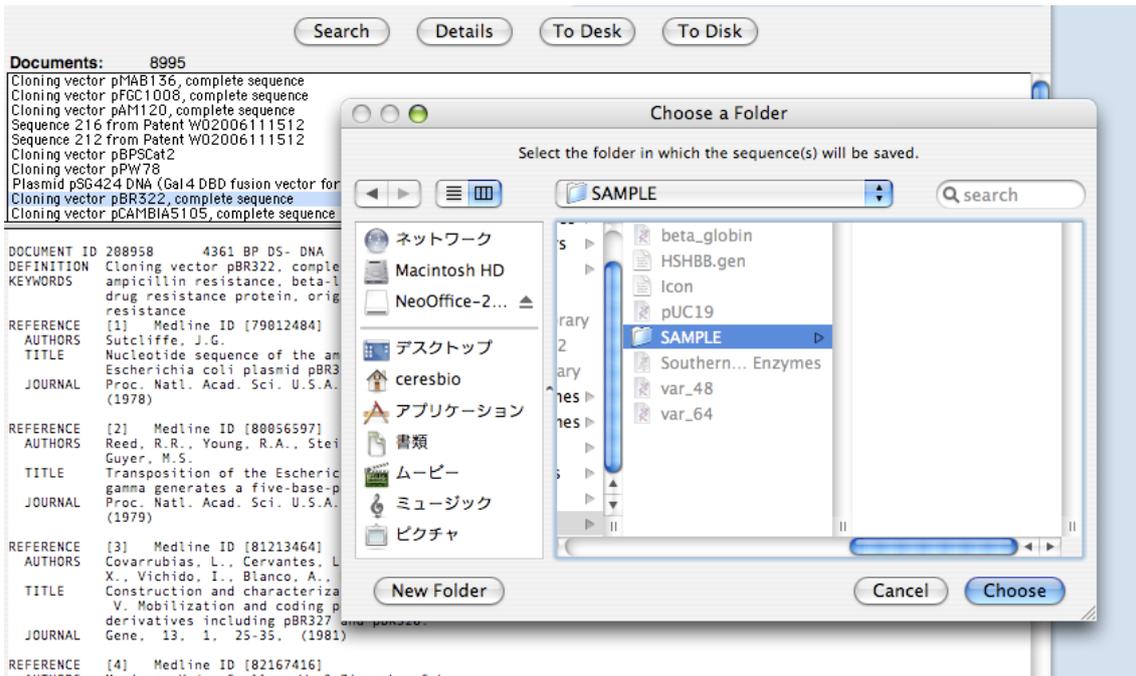


SYNPBR322

TTC TCA TGT TTG ACA GCT TAT CAT CGA TAA GCT TTA ATG CCG TAG TTT ATC ACA GTT AAA TTG CTA ACC 22  
 TCA GGC ACC GTG TAT GAA ATC TAA CAA TGC GCT CAT CGT CAT CCT CCG CAC CGT CAC CCT GAA TGC TGT AGG 44  
 CAT AAG CTT GGT TAT GCC GGT ACT GCC GGG CCT CTT GCG GAA TAT CGT CCA TTC CAA CAG CAT CGC CAG TCA 66  
 CTA TGG CGT GCT AGC GCT ATA TGC GTT GAT GCA ATT TCT ATG CCG ACC CGT TCT CCG AGC ACT GTC CAA 88  
 CCG CTT TGG CCG CCG CCC AGT CCT GCT CCG TTC GCT ACT TGG AGC CAC TAT CCA CTA CCG GAT CAT GGC GAC 110  
 CAC ACC CGT CCT GTG GAT CCT CTA CCG CCG AGC CAT CGT GGC CCG CAT CAC CCG CCG CAC AGG TGC GGT TCC 132  
 TGG CCG CTA TAT CCG CCA CAT CCA CCA TGG GAA AAG TCG GGC TCG CCA CTT CCG GCT CAT GAG CCG TTG TTT 154  
 CCG CGT GGG TAT GGT GGC AAG CCC CGT GGC CCG GGG ACT GTT GGG CCG CAT CTC CTT GCA TGC ACC ATT CCT 176  
 TGC GGC GGC GGT GCT CAA CCG CCT CAA CCT ACT ACT GGG CTG CTT CCT AAT GCA GAA GTC GCA TAA GGG AAG 198  
 GGG TCG ACC GAT GCC CTT GAG AGC CTT CAA CCC AGT CAG CTC CTT CCG GTG GGC GCG GGG CAT GAC TAT CTT 220  
 CCG CCG ACT TAT GAC TGT CTT CTT TAT CAT GCA ACT CGT AAG ACA GGT GCC GGC AGC GCT CTG GGT CAT TTT 242  
 CCG CAA GAA CCG CTT TCG CTG GAG CCG GAC GAT GAT CCG CCT GTC GCT TGC GGT ATT CCG AAT CTT GCA CCG 264  
 CCT CGC TCA AGC CTT CGT CAC TGG TCC CCG CAC CAA AGC TTT CCG CAA GAA GCA GGC CAT TAT CCG CCG CAT 286  
 GGC GGC CAA CCG GCT GGG CTA CGT CTT GCT GGC GTT CCG GAC GCG AAG CTG GAT GGC CTT CCC CAT TAT GAT 308  
 TCT TCT CCG TTC CCG CCG CAT CCG GAT GCC CCG GTT GCA GGC CAT GCT GTC CAG GCA GGT AAG TGA CAA CCA 330  
 TCA GGG ACA GCT TCA AGG ATC GCT CCG GGC TCT TAC CAG CCT AAC TTC GAT CAC TGG ACC COT GAT CGT CAC 352  
 GGC GAT TTA TGC CCG CTC CCG GAG CAG ATG GAA CCG GTT GGC ATG GAT TGT AAG CCG CCG CCA CTA CTT TGT 374  
 CTG CCT CCC CCG GTT CCG TCG CCG TGC ATG GAG CCG GGC CAC CTC CAG CTA GAT GAA AGC CCG CCG CAC CTC 396

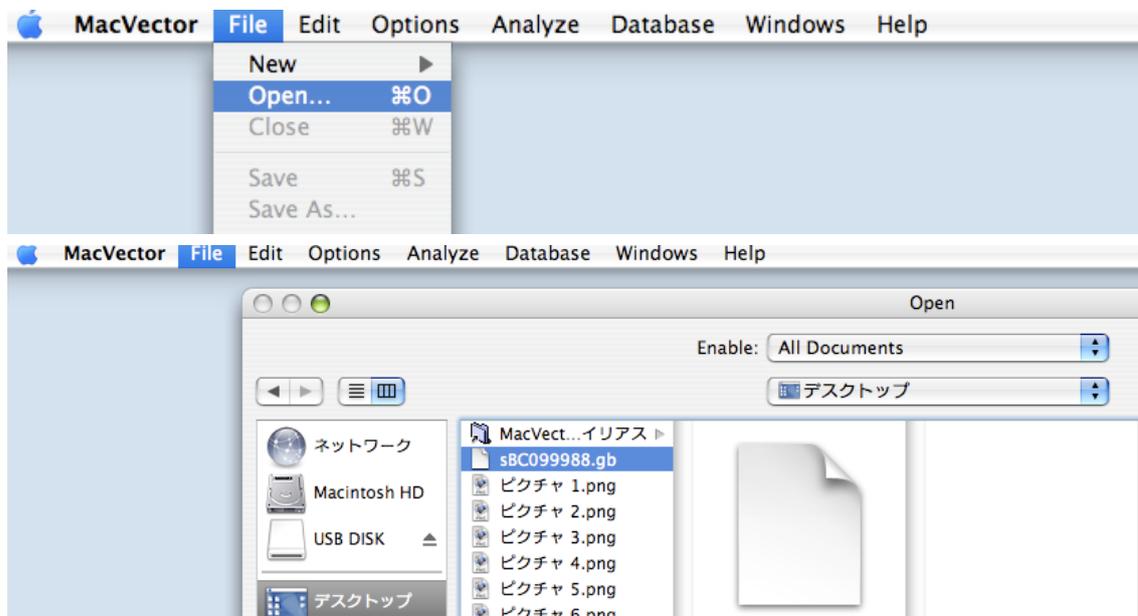
gamma generates a five-base-pair repeat.  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A., 76, 10, 4882-4886,

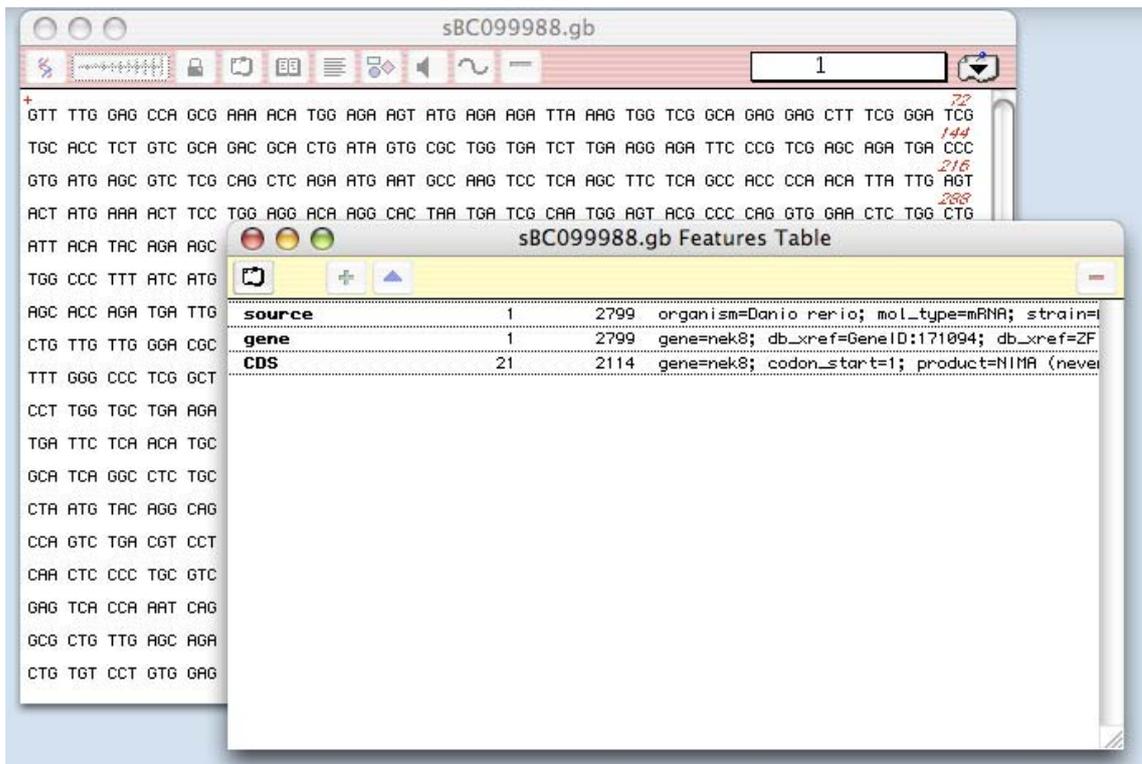
“To Desk”をクリックすると、MacVectorのエディターが表示されます。(方法1)



“To Disk”をクリックすると、エディターは表示されませんが、ファイルとしてセーブすることができます。(方法2)

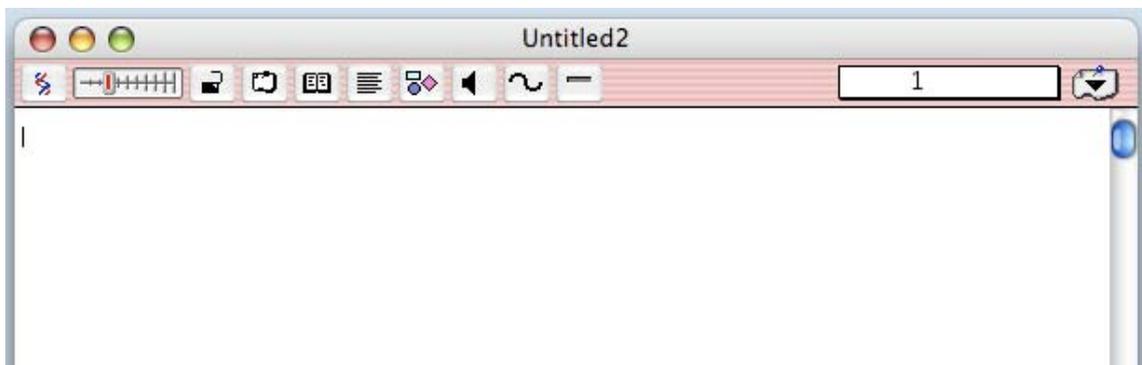
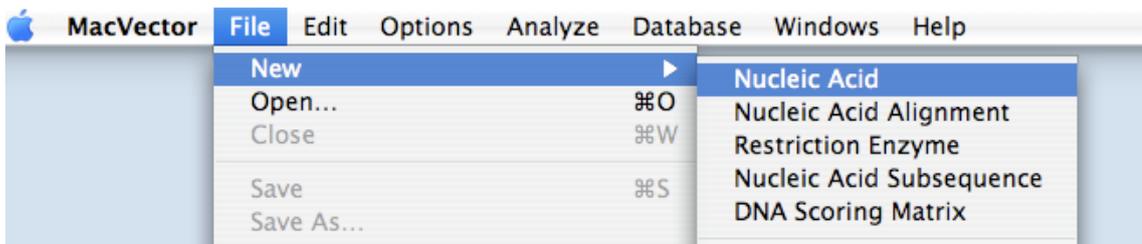
その他に、NCBI や EMBL からダウンロードした GeneBank や EMBL(SwissProt)フォーマットのテキストファイルも、アノテーションごと MacVector で利用することができます。





## B.配列情報ファイルの作成(新規)

新しい配列ファイルを作成するには、まず空のファイルを作成します。





配列ファイルウィンドウのアイコンについて簡単に説明します。



このアイコンをクリックすることで、DNA/RNA の変更を行います。



エディタに表示される塩基(残基)のブロック数を変更できます。(連続~10 塩基区切りが可能です。)



ファイル編集の“解除/ロックする”を選択します。



FT テーブルのウィンドウを表示します。



FT 以外のアノテーションのウィンドウを表示します。



このウィンドウの内容をプリンター出力形式にして表示します。



このウィンドウの内容をグラフィックで表示します。



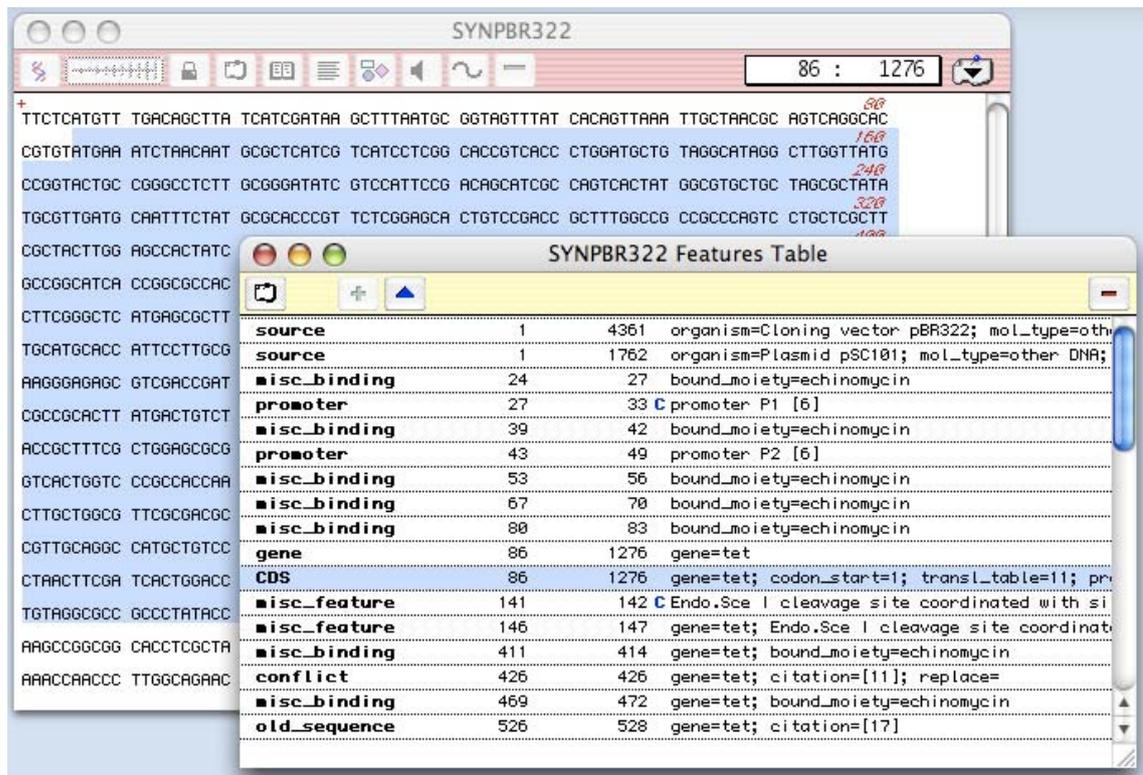
音声出力の切り替えを行います。



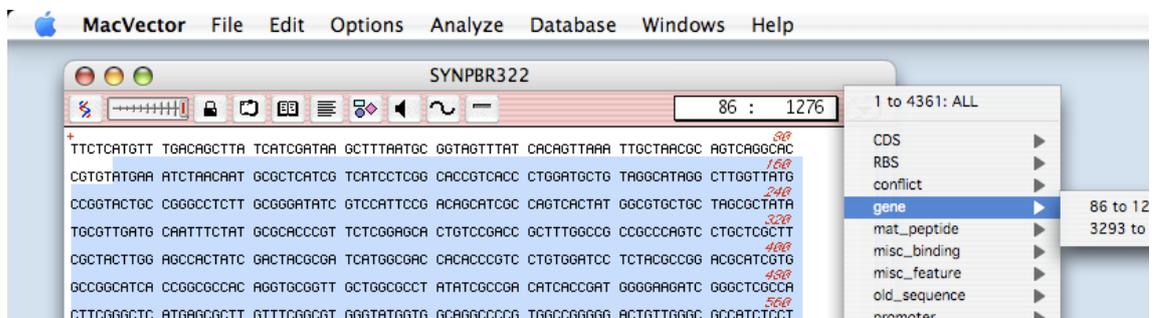
配列の形状がリニア(直鎖)状かサークル(環)状かを選択します。



相補鎖の表示を選択します。



FT のリストから任意の項目を指定すると、配列エディター上にその領域(位置)が表示されます。



配列エディター右上部のスクロールメニューにも、FT 情報が表示されます。

ここからも同様に、領域を選択することができます。

配列情報の入力にはいくつかの方法があります。

- a. 直接キーボードから入力
- b. 他のテキストエディターや Web ページからコピー&ペースト
- c. 他の MacVector のファイルから編集作成する

a. 直接キーボードから作成する場合

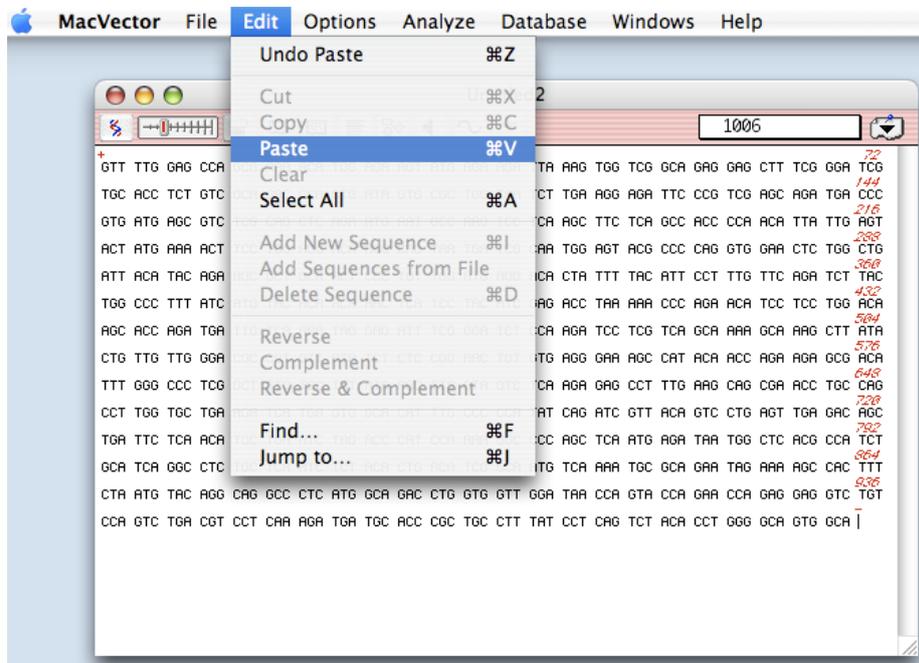
直接キーボードから入力します。この場合に  をスピーチモードにすることにより、入力した塩基が音声で確認できます。

b. 他のテキストエディターや Web ページからコピー&ペースト

MacVector はほとんどのテキストエディターからコピー&ペーストで配列を入力することができます。この場合、配列以外の数字や記号は自動的に除去されます。

```

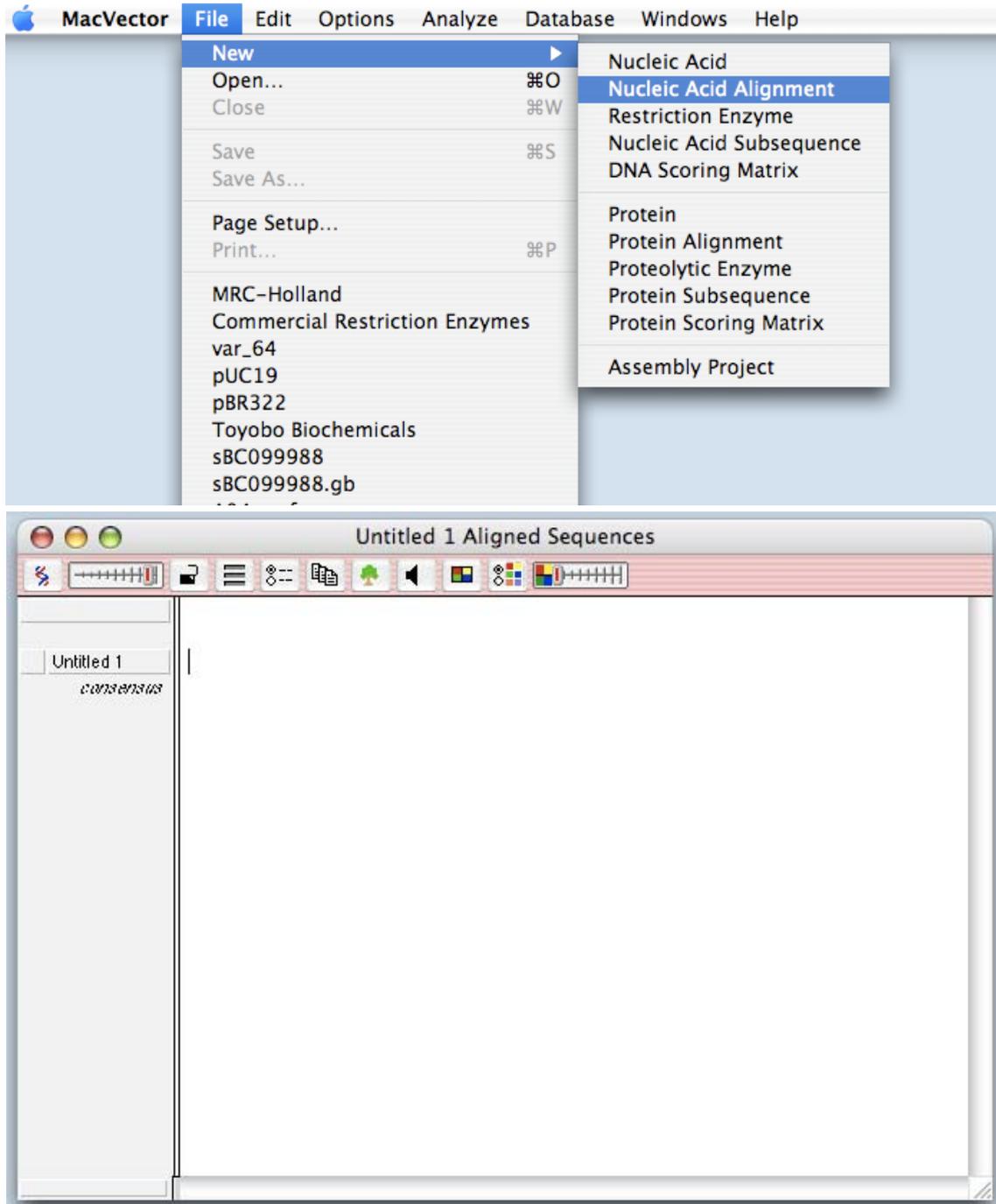
ORIGIN
1  gttttggagc cagcgaaaac atggagaagt atgagaagat taaagtggc ggcagaggag
61  ctttcgggat cgtgcacetc tgtcgcagac gcactgatg tggctgggt atcttgaagg
121 agatcccggt cagcagagat acccgtgat agcgtctgc agctcagaat gaatgccaag
181 tcctcaagct tctcagccac ccaaacatta ttggtacta tgaaaacttg ctggaggaca
241 aggcactaat gatcgaatg gatgacgcc caqgtggaac tctggctgat tacatacaga
301 agcgtcgcaa ctccctggtg gatgaggaca ctattttaca ttctttggt cagatcttac
361 tggcccttta tcatgtacac acaaaactca tcctacatcg agacctaaaa acccagaaca
421 tcctcctgga caagcaccag atgatgtca aaataggaga ttctggggtc tccaagatcc
481 tctcagccea aagcaagctt tatactgttg ttgggacgcc gtattatate tctccggaac
541 tctgtggagg aaagccatag aaccagaaga gccacatttg ggcctcgagg tctgtacttt
601 aagagcttac tactctaaag aagaccttgg aagcagcaaa cctaccagac ttggtctbga
661 agatcctcag tggcactatg gccccatgat caatcgtta caatctcag ttgagagacc
721 tcatctcaas catctgaat ctgagcccat ccaaaagccc ccaactcaat gqataatgg
781 ctccagccat ctgcatcagg cctctgctca atctctacac tgacatcgcc aatctcaaaa
841 tggcagaaat aqaaaagcaa ctttetaatg tacagggcag ccctcatggc agacctgggt
901 gttgataaac cagtlaccaga accagagagg gtctgtccag tctgagctcc tcaaaatgga
961 tgcaccctgt gcccttatcc tcaqcttaca cctggggcag tggcaattea actccctgpc
1021 gtctgcccac gttaaacact gagggtcatc aagtgctctc gggagccaca cagaagatgg
1081 gactcaccaa atcagggagg ctcactacat gggagggccc ctgggttggg tccggtgagc
1141 ccaactctgc tggcctgttt gaggagatgc agccccagtt catctccccc ttctctagagg
1201
  
```



また、Web のページから直接コピー&ペーストで入力することも可能です。

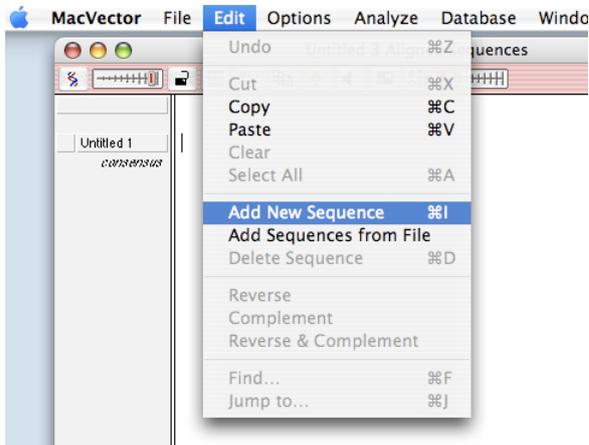
## C. マルチプルアラインメント

MacVector でマルチプルアラインメントを行うためには、まずアラインメントのエディターファイルを作成します。このファイルがプロジェクトファイルとして利用されます。

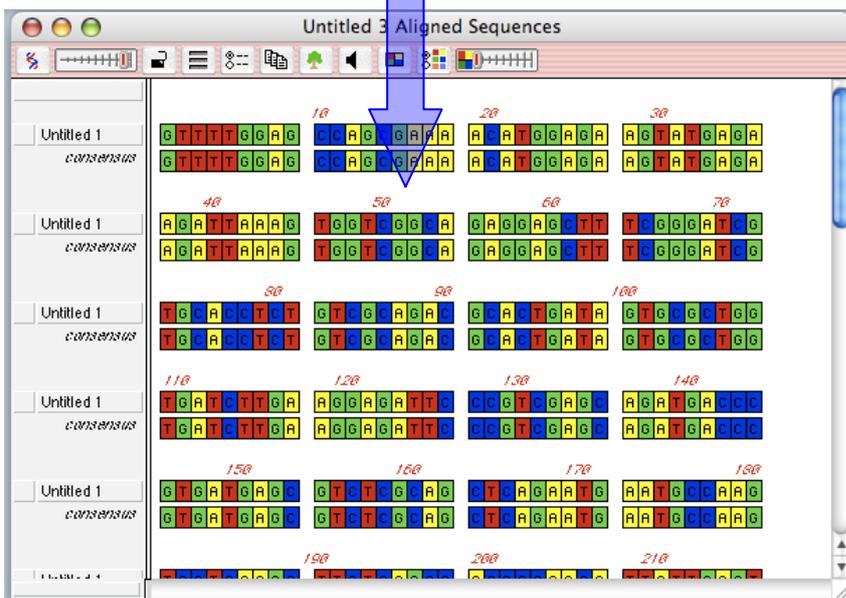
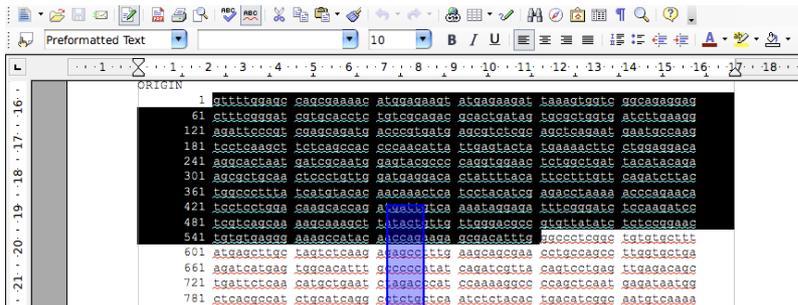


MacVector でマルチプルアラインメントを行うためには、あらかじめ MacVector のフォーマットで配列情報のファイルを保存しておくことが原則です。

しかし、直接配列を入力することも可能です。



この場合、前述の新規配列ファイル作成の要領で、キーボードから入力するか、他のファイルまたはエディターからコピー&ペーストで配列を入力してください。



ここで、マルチプルアラインメントのエディターの機能について簡単に説明します。

エディターの上部にはいくつかのアイコンが表示されています。



エディタに表示される塩基(残基)のブロック数を変更できます。(連続~10塩基区切りが可能です。)



このファイルの編集のロック/解除をおこないます。



配列をページに合わせて折り返して表示させるためにはこのボタンをクリックしてください。



エディターと結果出力ファイルの各種フォーマットの設定を行うためのボタンです。



出力する結果(ファイル、グラフィック)の種類を選択と、エディターで配列を編集した場合の再計算を行います。



系統樹を作成するための条件設定をおこないます。



エディターの音声出力の ON/OFF を選択します。



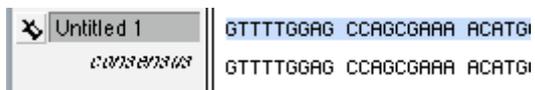
エディターの表示をブロックモード/テキストモードに切り替えます。



ブロックモードの各カラーの属性を設定します。



ブロックモードの幅を変更するために使用します。



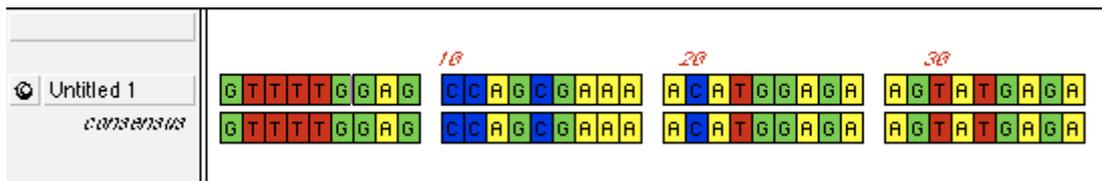
エディターにインポートされた配列のファイル名は、名称をマウスでダブルクリックすると編集できるようになります。

また、配列の編集を行いたくないファイルは左端のクリップマークをクリックして、固定モードに変更してください。

<編集可能>

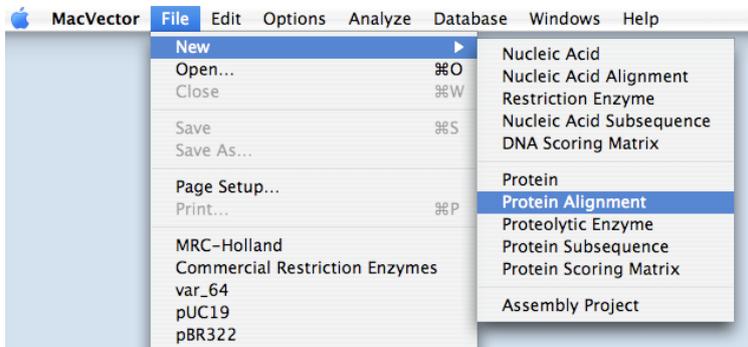


<編集不可能>

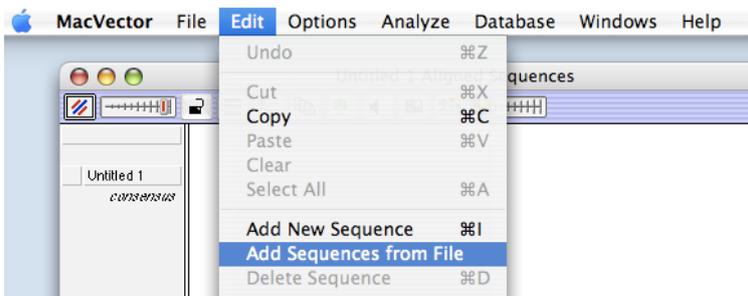


実際にアミノ酸配列を利用した一連の操作を紹介します。

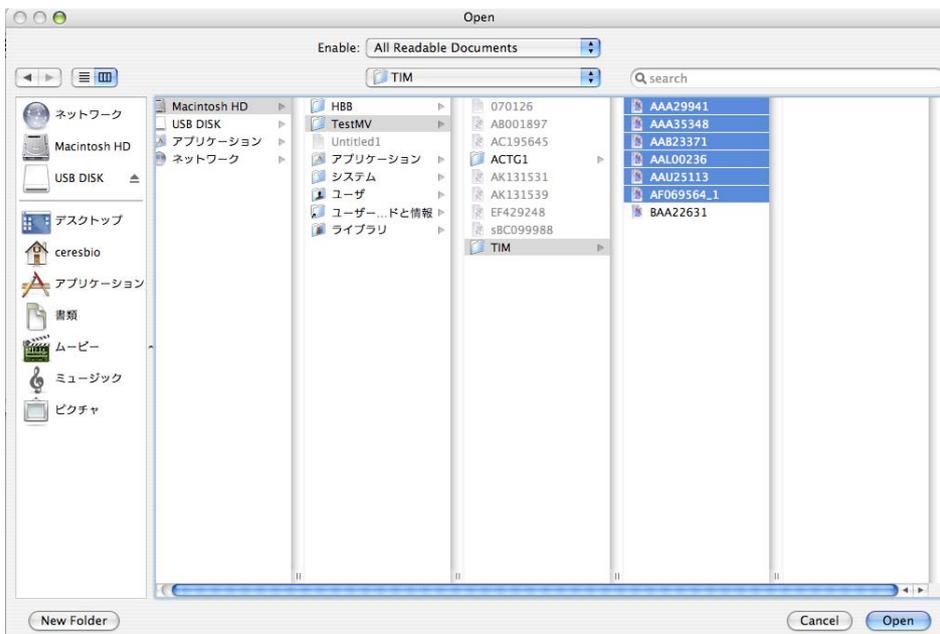
まず、新しいプロジェクト(タンパク質のアラインメントエディター)を作成します。



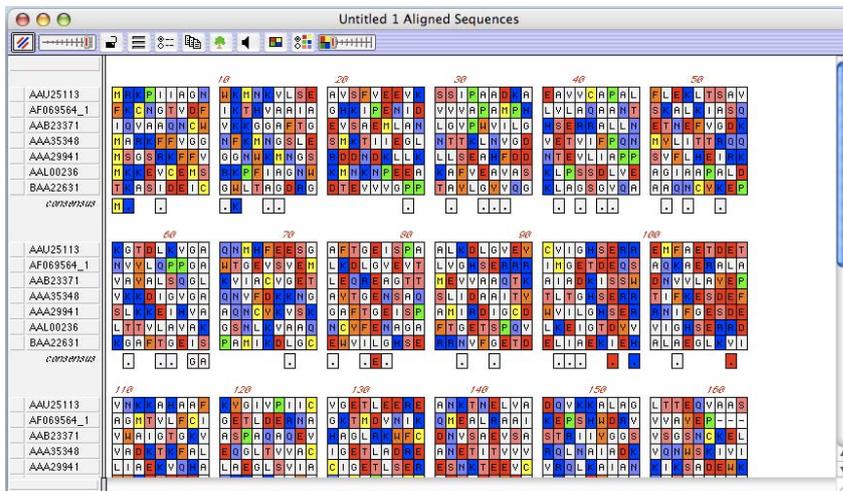
既存のアミノ酸配列をエディターに取り込みます。



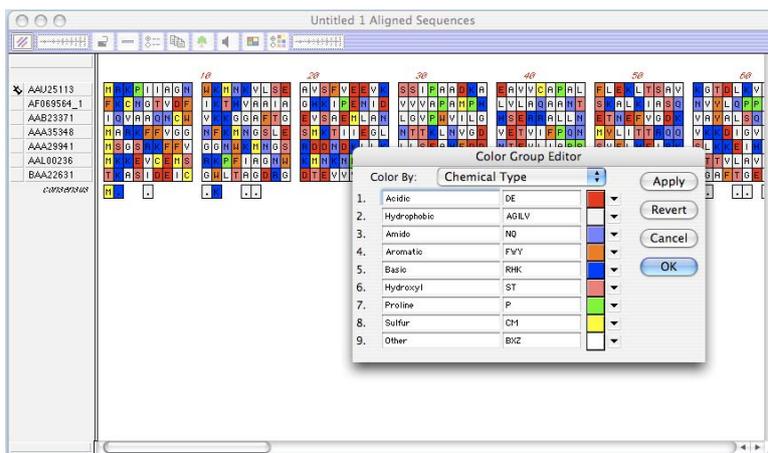
アップルマークを押しながら選択すると複数のファイルを選択することができます。



ファイルが取り込まれました。



 ボタンを利用して表示する各アミノ酸のカラータイプを選びます。

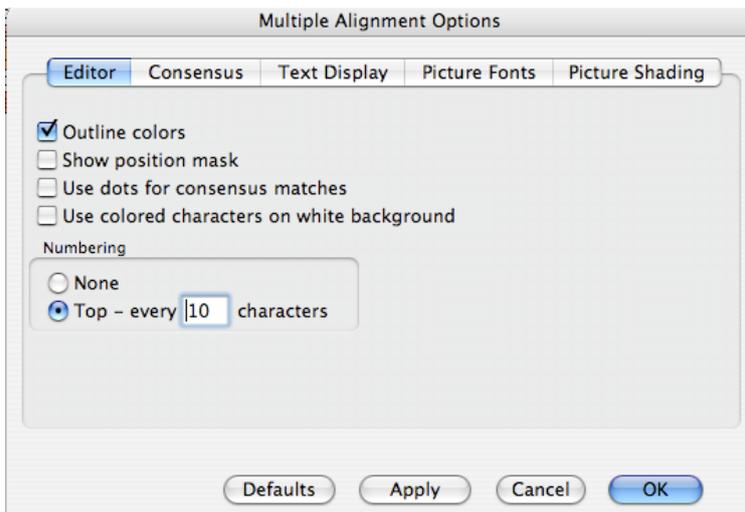


とりあえず、アラインメントを実行してみたい場合は出力等に関する設定の説明(P15-16)を飛ばしてご覧ください。

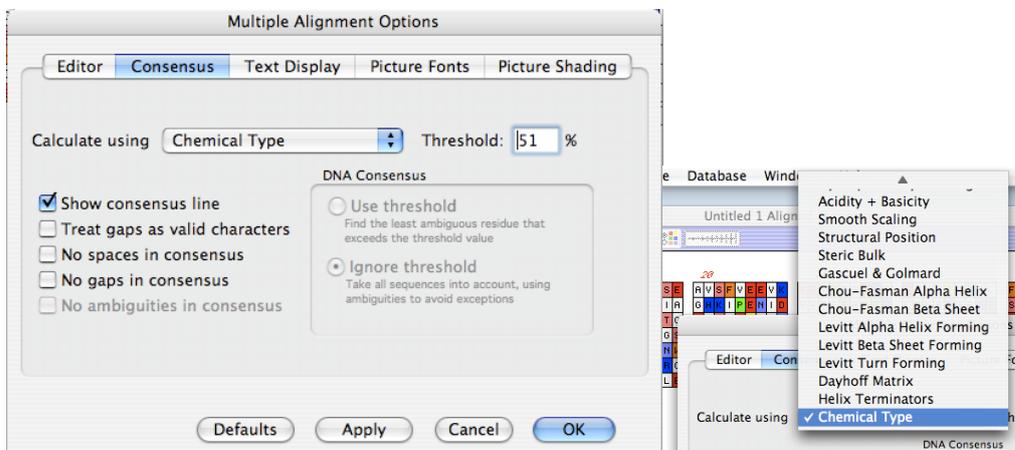


ボタンを押して、エディターや出力結果の書式を設定します。

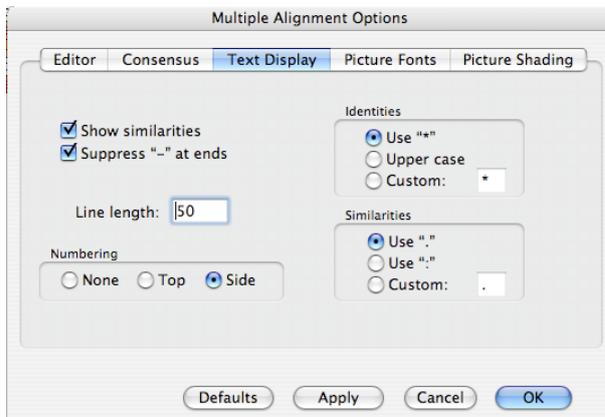
### <エディター画面の表示形式>



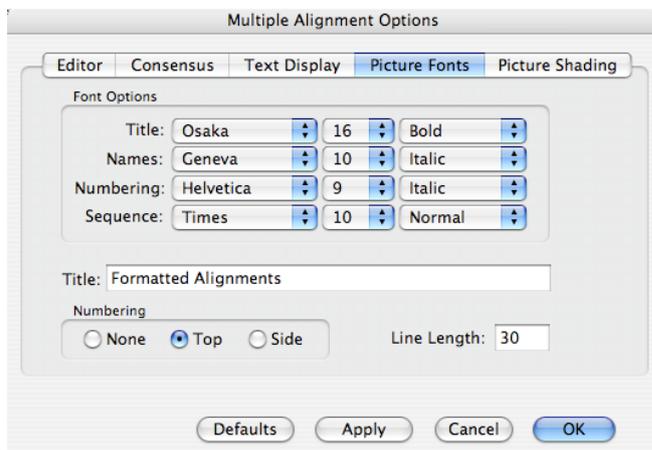
### <テキスト出力でコンセンサス配列を表示する形式>



<テキスト出力でアラインメントを表示する形式>

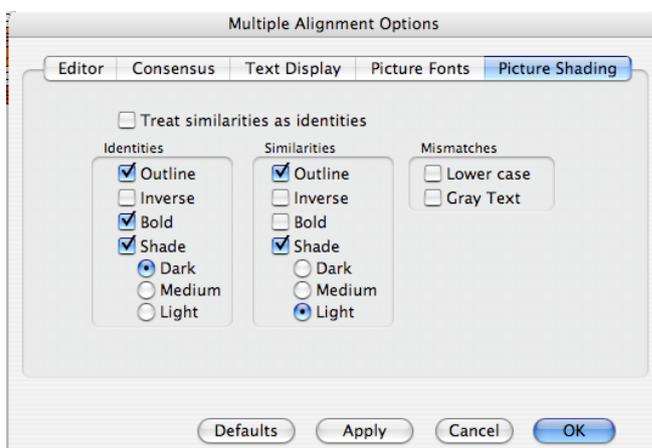


<ブロック形式-Formatted Alignment-出力でアラインメントを表示する形式(font)>



注！等幅フォントを使用すると、他のエディターでの編集が容易になります。

<ブロック形式-Formatted Alignment-出力でアラインメントを表示する形式(描画)>

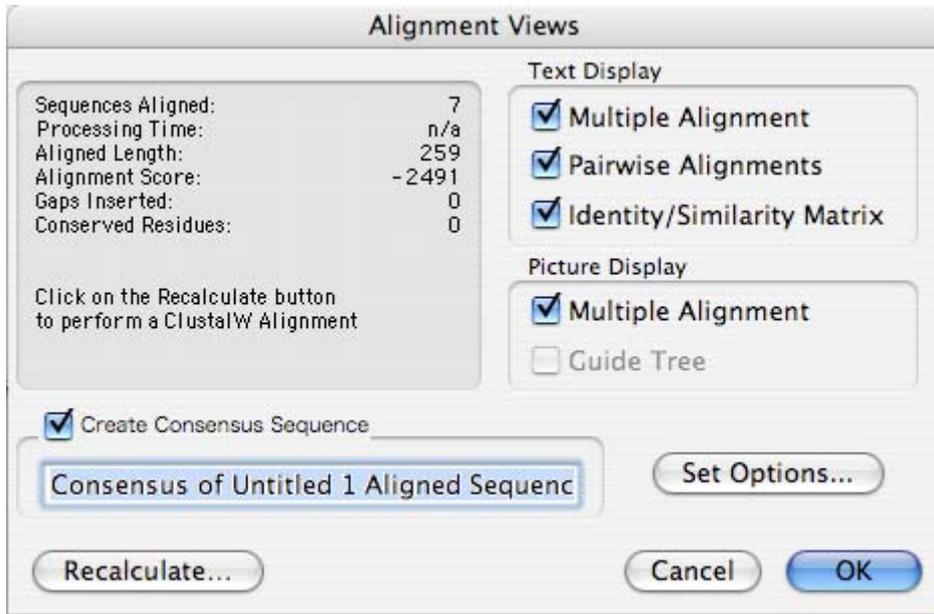


最後に OK をクリックしてください。



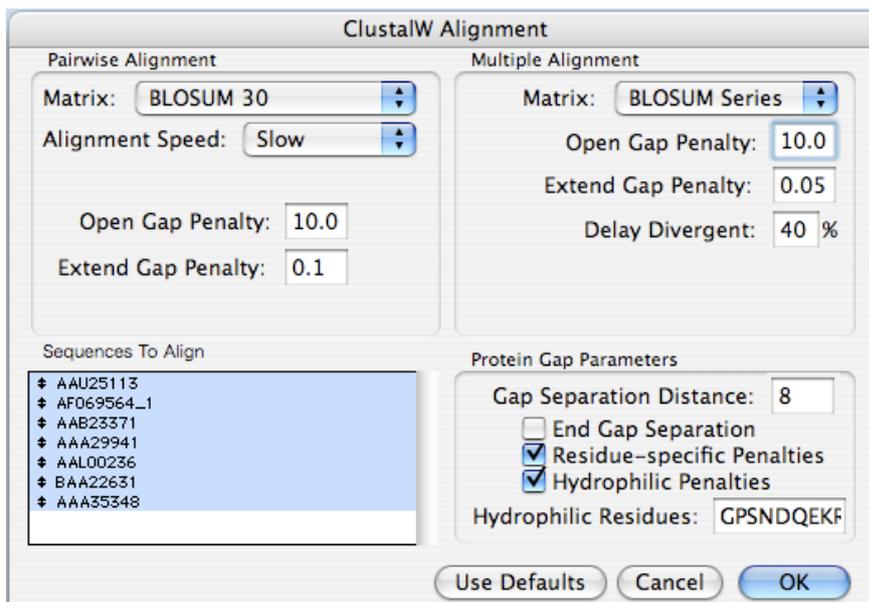
をクリックしてアラインメントを開始します。

出力する解析結果を選択してください。



各結果の表示は P15-16 で設定した条件に基づいて作成されます。

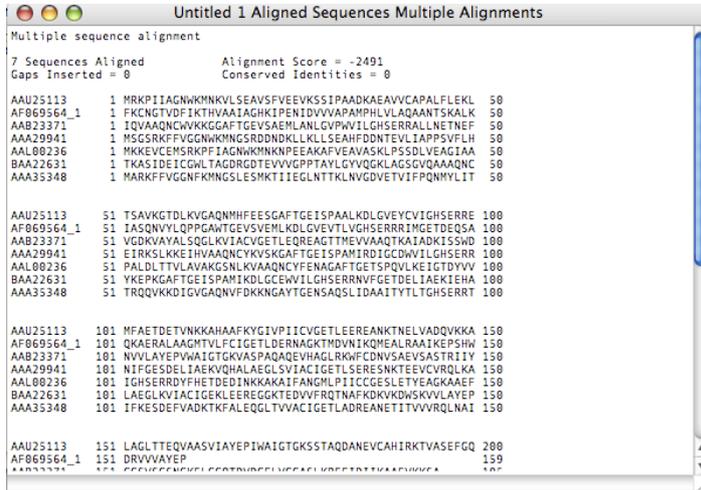
アラインメントのアルゴリズム「ClustalW」のパラメータを調整する場合、またエディターで何らかの変更を行った場合は必ず左下の“Recalculation...”を実行して結果の再構築を行ってください。



最後に OK をクリックしてください。

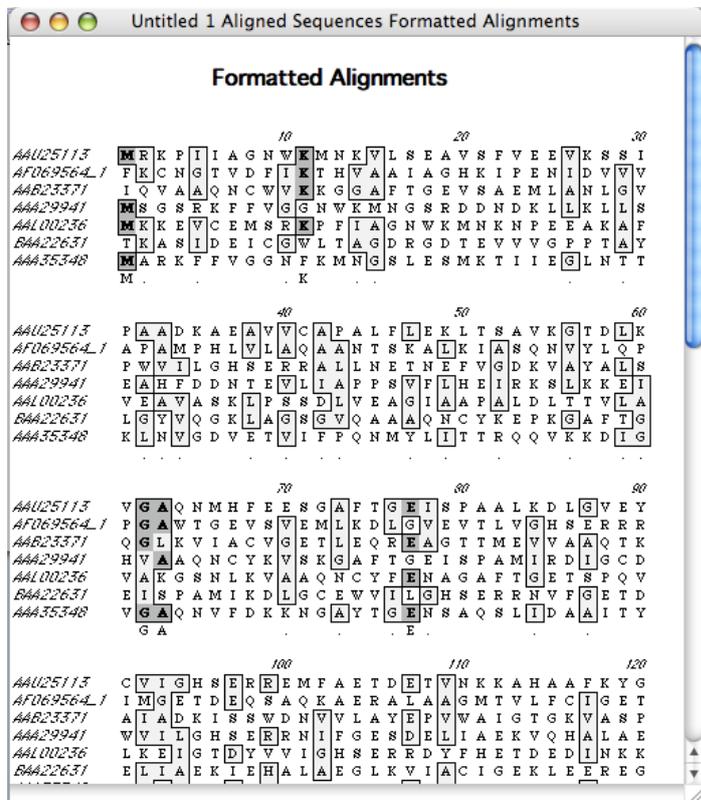
プリンターで出力できる形式の結果ファイルが表示されます。

(テキストアラインメント)



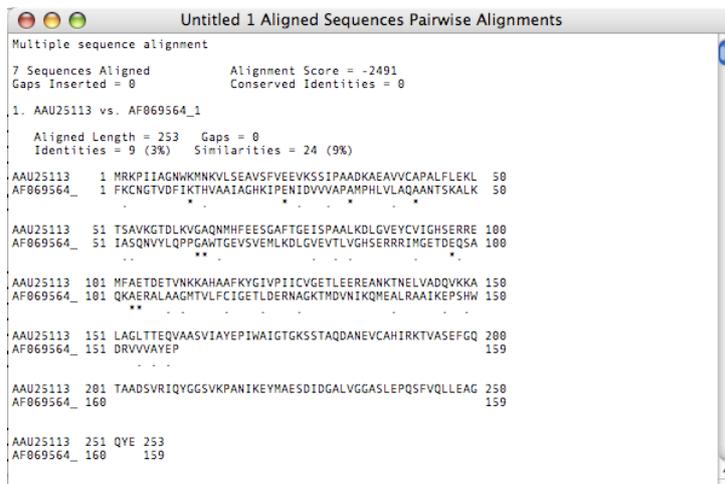
注！等幅フォントを使用すると、他のエディターでの編集が容易になります。

(ブロック形式のアラインメント)



注！等幅フォントを使用すると、他のエディターでの編集が容易になります。

(各配列間のアラインメントの詳細)



```

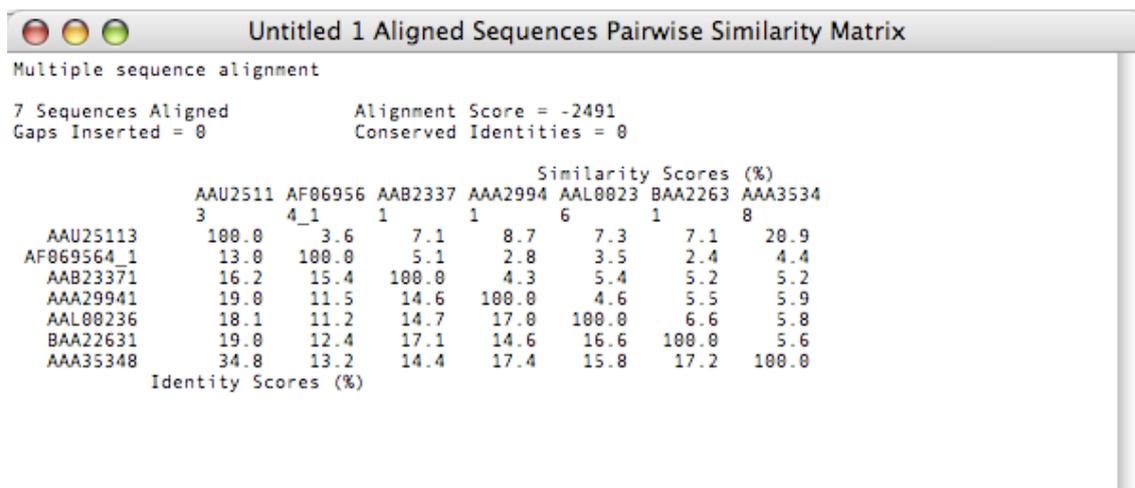
Multiple sequence alignment
7 Sequences Aligned      Alignment Score = -2491
Gaps Inserted = 0       Conserved Identities = 0

1. AAU25113 vs. AF069564_1
   Aligned Length = 253   Gaps = 0
   Identities = 9 (3%)   Similarities = 24 (9%)

AAU25113   1 MRKPIIAGNHKMKVLSAVSFVVEVKSSIPADKAEAVVCAPALFLEKL  50
AF069564_  1 FKCNGTVDFIKTHVAAIAGHKIPENIDVVVAPAMPHLVLAQAANTS KALK  50
           * * * * *
AAU25113   51 TSAVKGTDLKVGQNMHFEESGAFTEISPAALKDLGVEYCVIGHSERRE  100
AF069564_  51 IASQNVYLQPPGAWTGEVSVEMLKDLGVEVTLVGHSERRRIMGETDEQSA  100
           * * * * *
AAU25113  101 MFAETDET VNKKAHAAFKYGI VPIICVGETLEEREANKTNELVADQVKKA  150
AF069564_ 101 QKAERALAAGMTLVLCIGETLDERNAGKTM DVNLIKQMEALRAAIKEPSHW  150
           * * * * *
AAU25113  151 LAGLTTEQVVAASVIAYEPIWAIGTGKSS TAQDANEVCAHIRKTVASEFGQ  200
AF069564_ 151 DRVVVAYEP
           * * * * *
AAU25113  201 TAADSVRIQYGGSVKPAIKEYMAESDIDGALVGGASLEPQSFVQLLEAG  250
AF069564_ 160
           * * * * *
AAU25113  251 QYE 253
AF069564_ 160 159
    
```

注！等幅フォントを使用すると、他のエディターでの編集が容易になります。

(各配列間の相同性のマトリックス)



```

Multiple sequence alignment
7 Sequences Aligned      Alignment Score = -2491
Gaps Inserted = 0       Conserved Identities = 0

Similarity Scores (%)
      AAU2511  AF06956  AAB2337  AAA2994  AAL0023  BAA2263  AAA3534
AAU25113  100.0    3.6     7.1     8.7     7.3     7.1    20.9
AF069564_1 13.0   100.0    5.1     2.8     3.5     2.4     4.4
AAB23371   16.2   15.4   100.0    4.3     5.4     5.2     5.2
AAA29941   19.0   11.5   14.6   100.0    4.6     5.5     5.9
AAL00236   18.1   11.2   14.7   17.0   100.0    6.6     5.8
BAA22631   19.0   12.4   17.1   14.6   16.6   100.0    5.6
AAA35348   34.8   13.2   14.4   17.4   15.8   17.2   100.0
Identity Scores (%)
    
```

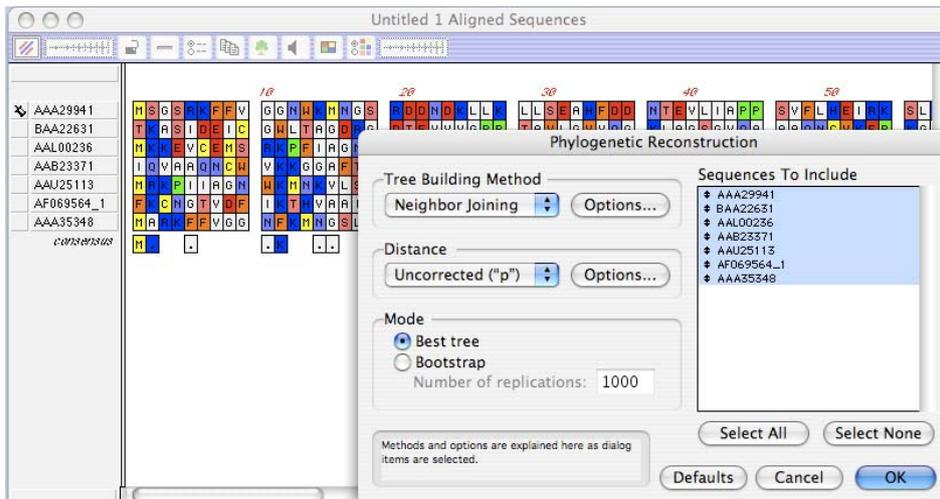
注！等幅フォントを使用すると、他のエディターでの編集が容易になります。

D. 系統樹の作成

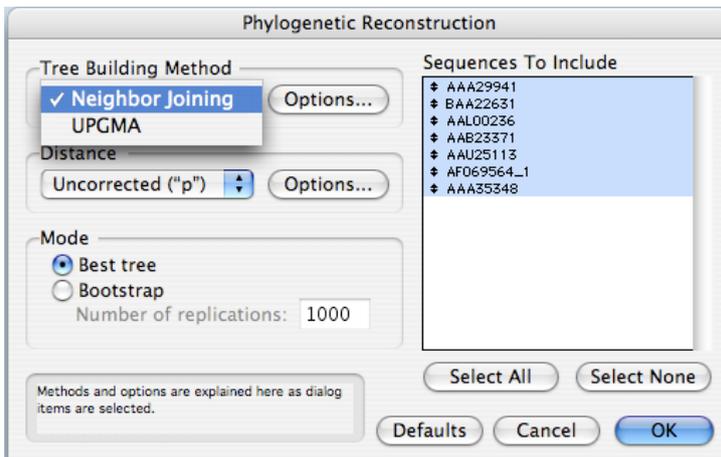
MacVector でマルチプルアラインメントの解析をおこなった結果から、系統樹を作成することができます。

マルチプルアラインメントのエディター上のアイコンボタン  をクリックしてください。

次に、系統樹に加える配列を右側のリストから選択します。



次にパラメータを設定します。

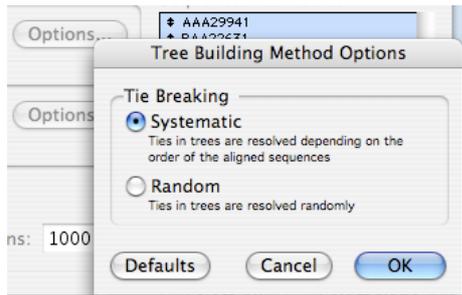


① 系統樹作成の手法としては以下のものが利用できます。

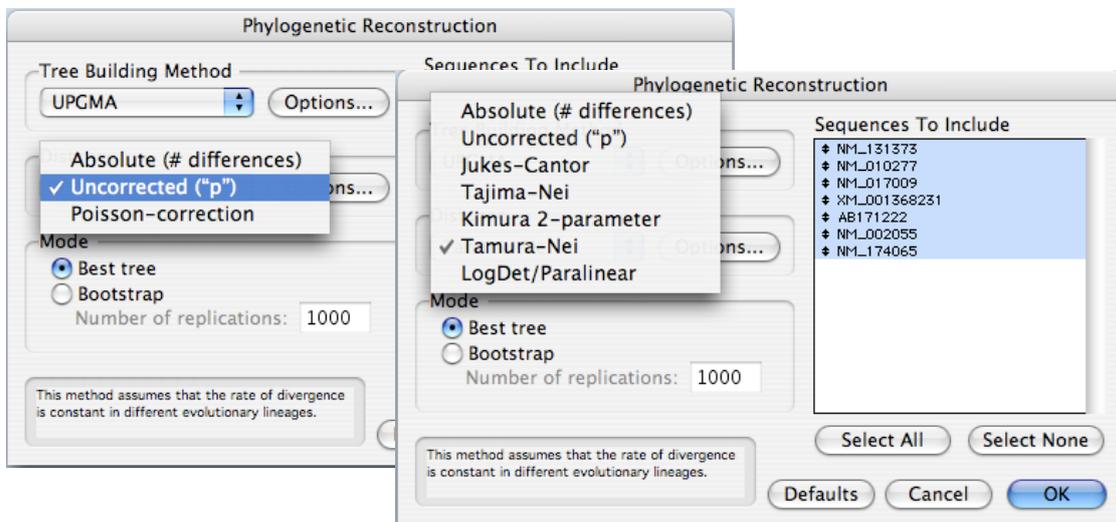
NJ (Neighbor Joining)法: 近隣結合法

UPGMA(Unweighted Pair Group Method with Arithmetic mean)法: 非加重結合法

(UPGMA 法は分子進化速度の異なる生物間ではその結果の誤差が大きくなるため、MacVector は NJ 法をデフォルトで利用しています。)



Option ボタンをクリックすると系統樹に表示する配列の順番を選択することができます。



②また、系統樹の距離を測定するための塩基(残基)置換推定法としては、

(核酸配列)

Absolute (# differences)

Uncorrected ("p")

Jukes-Cantoro

Tajima-Nei

Kimura 2-parameter

Tamura-Nei (default)

LogDet/Paralinear

(アミノ酸配列)

Absolute (# differences)

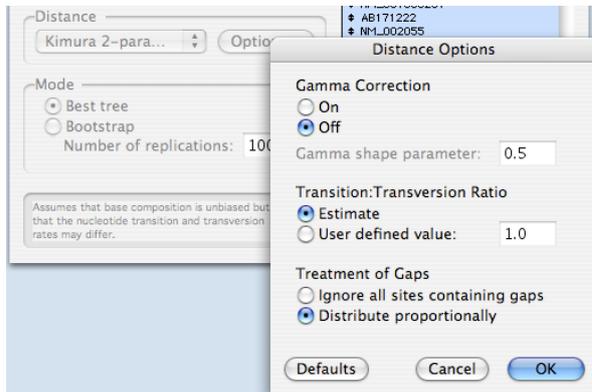
Uncorrected ("p") (default)

Poisson-correction

が選択できます。

Option ボタンで塩基(残基)置換推定法のさらに詳細な設定を行うことができます。

(利用できる設定値は各推定法によって異なりますので、詳細は UserGuide(英文)を参照してください。)



#### Gamma Correction:

配列範囲に均一な変異率を仮定できない場合に、領域の変異率に Gamma distribution を適用します。

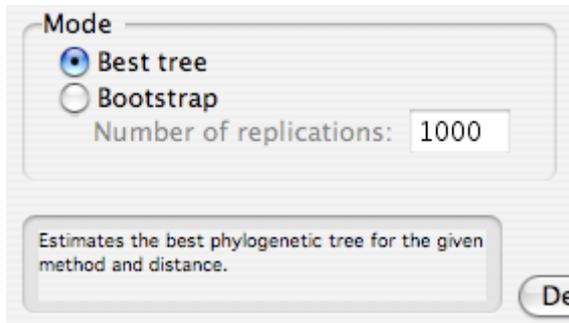
#### Transition: Transversion Ratio:

塩基転移速度の値を、アルゴリズムの標準値より任意に変更する場合に利用します。

#### Treatment of gaps:

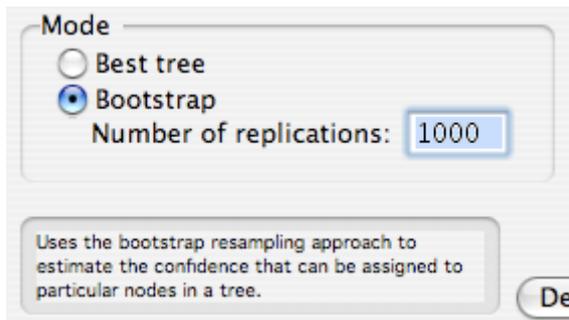
Gap を全て認めるか、配列の長さに応じてその比率を制限するかを選択します。

最後に、系統樹の最適化の方法を選択して OK をクリックします。



Best tree:

MacVector はプログラムが自動的に最適な系統樹を作成しますが、その評価をマニュアルで行うことができます。

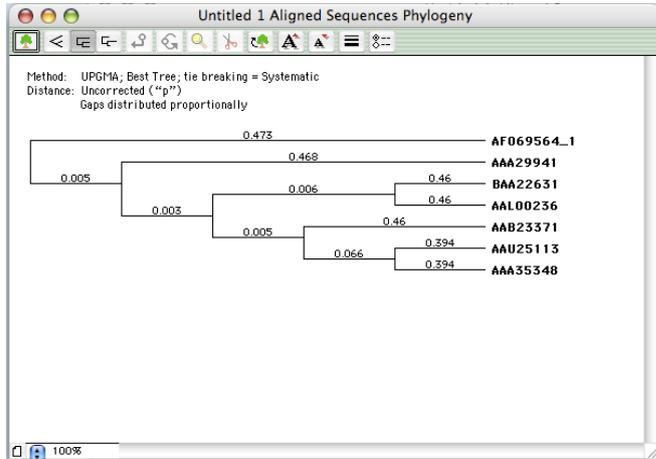


Bootstrap:

一般に Bootstrap 法と呼ばれている手法が一般的です。

Bootstrap 法のアルゴリズムは、系統樹を作成したアミノ酸配列の複製を大量に作成(replication)し、それぞれの replicant から推定される系統樹が元データの系統樹を支持する確率を求めるといふものです。(これを選択した場合は作成する複製の数を右のカラムに入力してください。)

以下に各パラメータで作成した系統樹の例と編集用のツールを紹介します。



系統樹の表示方法を選択します。左から順に、

slanted cladogram

rectangular cladogram

phylogram. が表示されます。



選択した配列を out-group として指定します。

(配列直近のラインをマークすると有効になります。)



1つの枝内にある配列の順番を入れ替えることができます。

(枝の元を示すラインをマークすると有効になります。)



選択した枝に含まれる部分だけの系統中を抜粋することができます。

(枝の元を示すラインをマークすると有効になります。)



特定の配列を排除することができます。

(配列直近のラインをマークすると有効になります。)



系統樹の条件を変更して再構築することができます。



系統樹の配列名表示のサイズを変更します。



系統樹のライン幅サイズを変更します。



系統樹表示画面の各種表示設定を変更します。

(配列名のフォント、枝間の距離表示、設定パラメータリスト(左上部)の表示、系統樹ルートを選択等)