Trends in Glycoscience and Glycotechnology
Vol.11 No.59 (May 1999) pp.119-127

Sequence Analyses Using DDBJ Home Page

DDBJのホームページを用いた配列解析

Kitano, Takashi ; and Saitou, Naruya
Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima 411-8540, Japan
FAX: 81-559-81-6789, E-mail: nsaitou@genes.nig.ac.jp

Key Words : *DDBJ, molecular evolution, sequence*

Abstract

The DDBJ (DNA Data Bank of Japan) home page features a submission system of nucleotide sequence data (SAKURA) and various tools for analyses of nucleotide and amino acid sequence data.

In this manuscript, we explain the methods of data retrieval by key words search, homology search, and multiple alignment using the DDBJ home page.

A. Accessing DDBJ Home Page

Start up your browser application and access the DDBJ home page (http://www.ddbj.nig.ac.jp/). Fig. 1 shows the DDBJ home page (The URL for the home page written in Japanese is http://www.ddbj.nig.ac.jp/Welcome-j.html).

B. SFgate-WAIS: Data Retrieval Using Key Words

Click "Database Searches and Data Analyses". The

要　約

DDBJのホームページでは、塩基配列の登録から、塩基配列やアミノ酸配列の解析に役に立つようなさまざまな内容が充実しています。

ここでは、DDBJのホームページへのアクセス方法から、データの取得、相同性検索、多重整列の方法について説明します。

A. DDBJのHome Pageへのアクセス方法

ブラウザーアプリケーションを用いて、アドレスhttp://www.ddbj.nig.ac.jp/を開くと、図1の画面に出ます。ここでは、日本語と英語の両方のページが用意されています（日本語のページはhttp://www.ddbj.nig.ac.jp/Welcome-j.html）。

B. SFgate-WAIS：キーワード検索による登録遺伝子の取得

"Database Searches and Data Analyses"をクリックして、"Da-

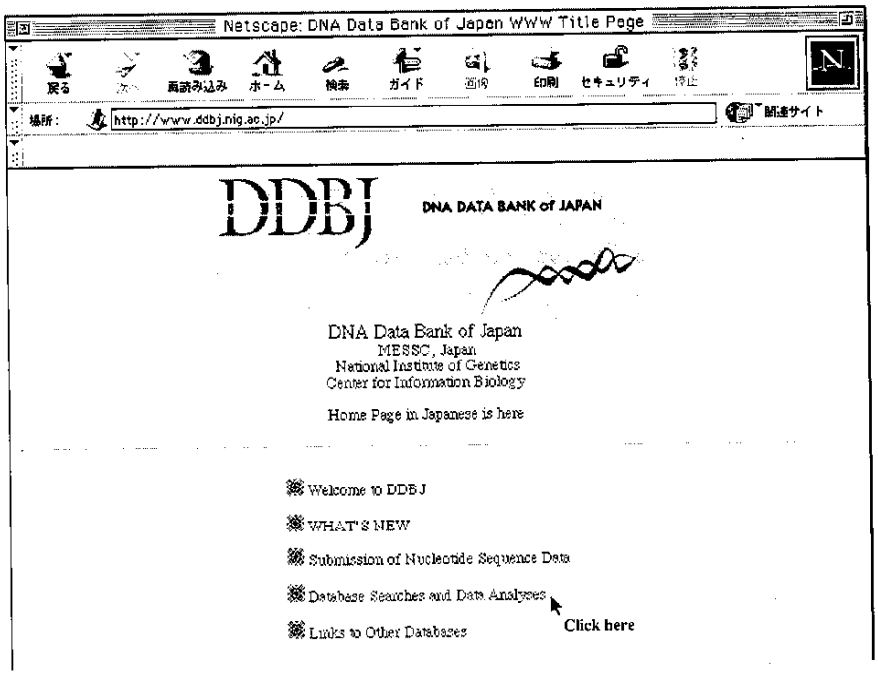


Fig. 1. DDBJ home page.

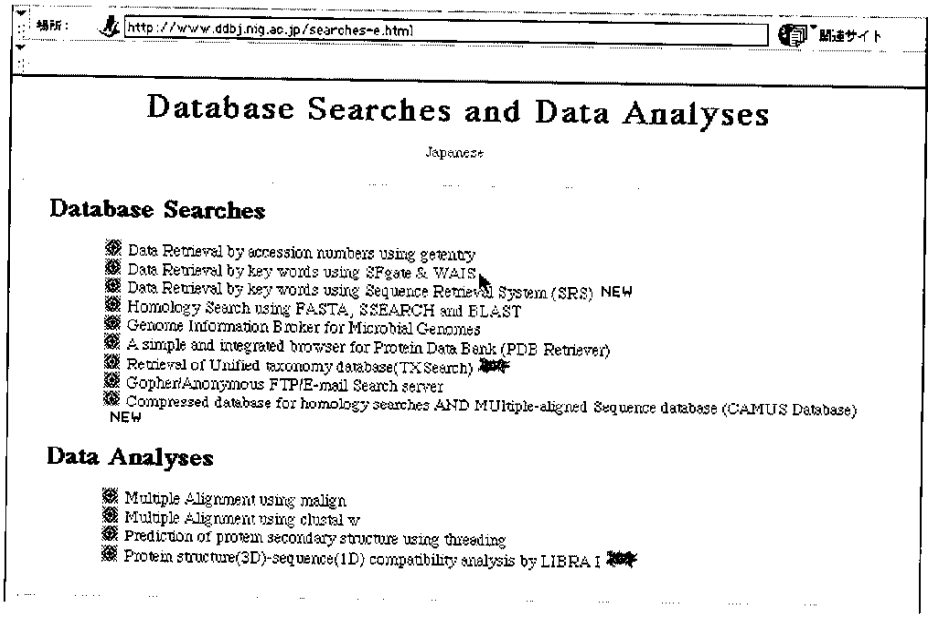


Fig. 2. Database searches and data analysis.

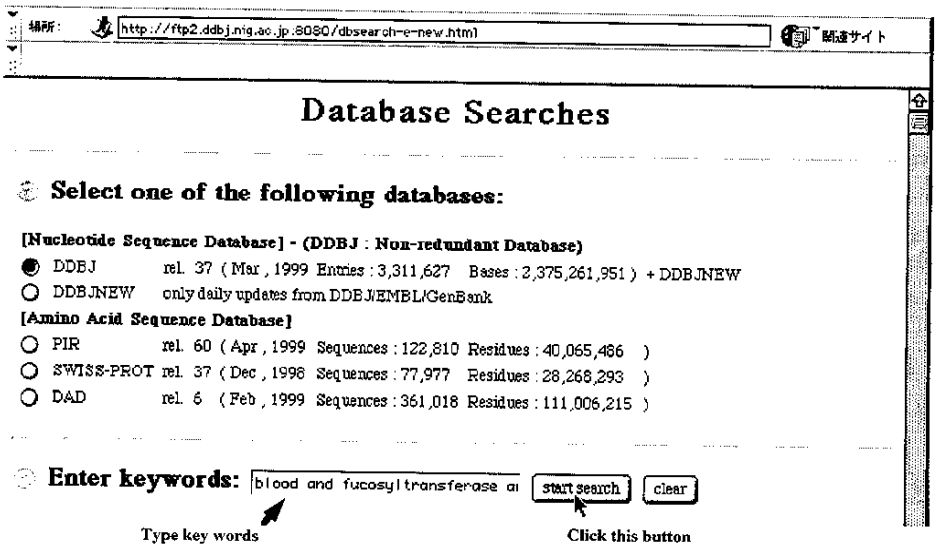
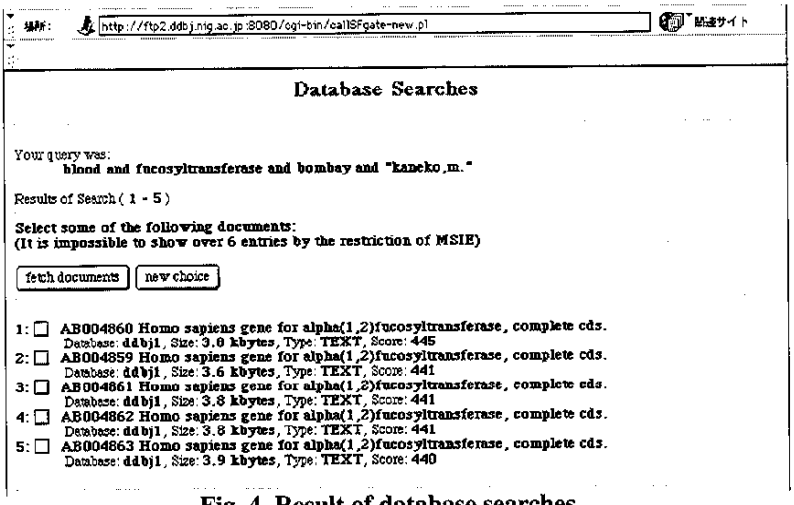


Fig. 3. Database searches.

"Database Searches and Data Analyses" page will appear (Fig. 2). Click "Data Retrieval by key words using SFgate & WAIS" in "Database Searches". The "Database Searches" page will appear (Fig. 3).

Select the database from "Select one of the following databases:". In this example, the database "DDBJ" is used. Input your key words in the "Enter keywords:" field . You can input plural words (Key words that appear very frequently in the database will be ignored.). In this example, four key words [blood and fucosyltransferase and Bombay and "Kaneko,M."] are given (Keywords including special characters such as -/,.* tabase Searches and Data Analyses"のページに入ります（図2）。"Database Searches"の項目の中の"Data Retrieval by key words using SFgate & WAIS"をクリックすると図3の画面に出ます。

"Select one of the following databases:"から、検索するデータベースを選択します。この例では、遺伝子を検索するのでデフォルトで"DDBJ"のままで用いています。まず、"Enter keywords:"のフィールド内にキーワードを入力します。これは、複数の語句を入力することができます（検索対象にならない語句があるので注意して下さい）。たとえば、blood and fucosyltransferase and bombay and "Kaneko,M."と入力して｢-(ハイフン) /(スラッシュ) ,(カンマ) .(ピリオド) *(アスタリスク)といった記号を含むキーワードを検索する場合には、そのキーワードを引用符で囲みます｣、"start search"をクリックします。



http://ftp2.ddbj.nig.ac.jp:8080/cgi-bin/callSFgate-new.pl

Database Searches

Your query was:
blood and fucosyltransferase and bombay and "kaneko,m."

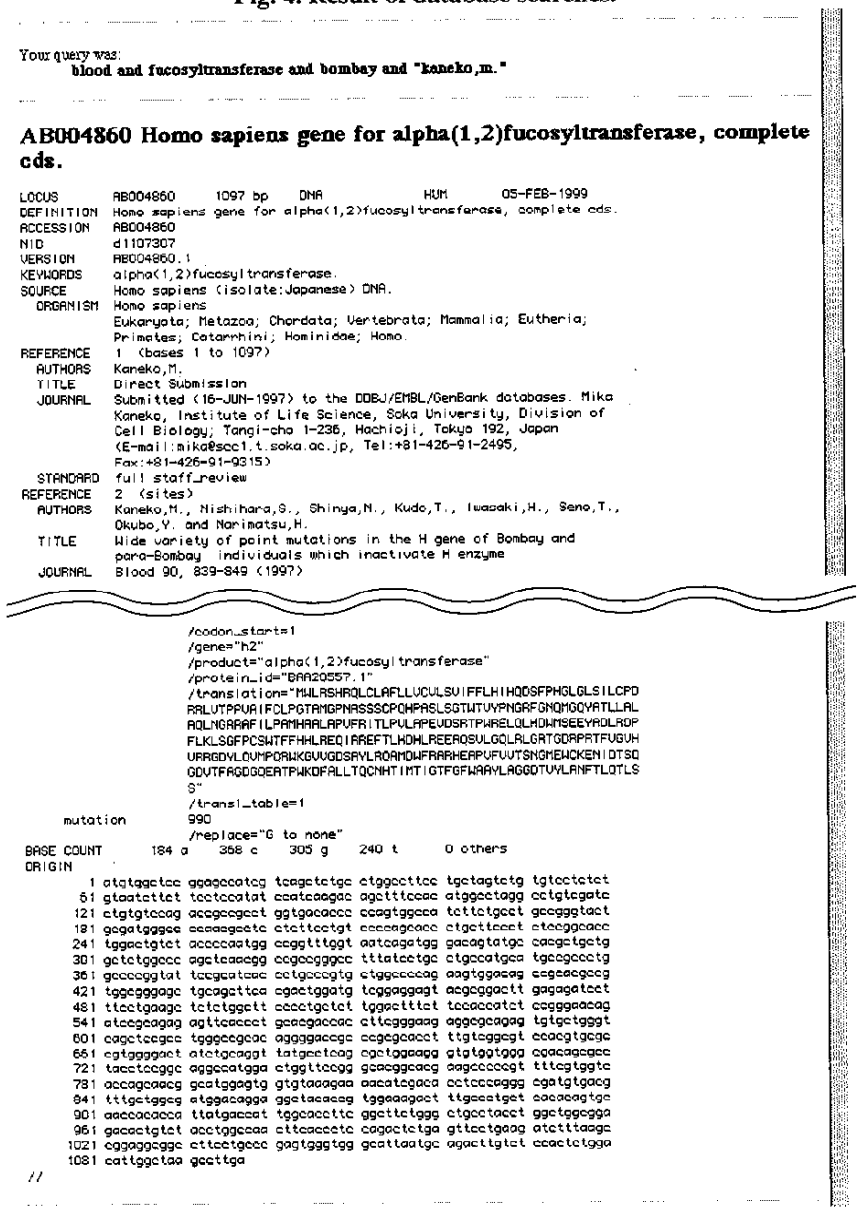
Results of Search (1 - 5)

Select some of the following documents:
(It is impossible to show over 6 entries by the restriction of MSIE)

fetch documents | new choice

1: ☐ **AB004860 Homo sapiens gene for alpha(1,2)fucosyltransferase, complete cds.**
Database: **ddbj1**, Size: **3.0 kbytes**, Type: **TEXT**, Score: **445**
2: ☐ **AB004859 Homo sapiens gene for alpha(1,2)fucosyltransferase, complete cds.**
Database: **ddbj1**, Size: **3.6 kbytes**, Type: **TEXT**, Score: **441**
3: ☐ **AB004861 Homo sapiens gene for alpha(1,2)fucosyltransferase, complete cds.**
Database: **ddbj1**, Size: **3.8 kbytes**, Type: **TEXT**, Score: **441**
4: ☐ **AB004862 Homo sapiens gene for alpha(1,2)fucosyltransferase, complete cds.**
Database: **ddbj1**, Size: **3.8 kbytes**, Type: **TEXT**, Score: **441**
5: ☐ **AB004863 Homo sapiens gene for alpha(1,2)fucosyltransferase, complete cds.**
Database: **ddbj1**, Size: **3.9 kbytes**, Type: **TEXT**, Score: **440**

Fig. 4. Result of database searches.



Your query was:
blood and fucosyltransferase and bombay and "kaneko,m."

AB004860 Homo sapiens gene for alpha(1,2)fucosyltransferase, complete cds.

```
LOCUS       AB004860     1097 bp    DNA             HUM       05-FEB-1999
DEFINITION  Homo sapiens gene for alpha(1,2)fucosyltransferase, complete cds.
ACCESSION   AB004860
NID         d1107307
VERSION     AB004860.1
KEYWORDS    alpha(1,2)fucosyltransferase.
SOURCE      Homo sapiens (isolate:Japanese) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1097)
  AUTHORS   Kaneko,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1997) to the DDBJ/EMBL/GenBank databases. Mika
            Kaneko, Institute of Life Science, Soka University, Division of
            Cell Biology; Tangi-cho 1-236, Hachioji, Tokyo 192, Japan
            (E-mail:mika@scc1.t.soka.ac.jp, Tel:+81-426-91-2495,
            Fax:+81-426-91-9315)
  STANDARD  full staff_review
REFERENCE   2  (sites)
  AUTHORS   Kaneko,M., Nishihara,S., Shinya,N., Kudo,T., Iwasaki,H., Seno,T.,
            Okubo,Y. and Narimatsu,H.
  TITLE     Wide variety of point mutations in the H gene of Bombay and
            para-Bombay  individuals which inactivate H enzyme
  JOURNAL   Blood 90, 839-849 (1997)
```

```
                     /codon_start=1
                     /gene="h2"
                     /product="alpha(1,2)fucosyltransferase"
                     /protein_id="BAA20557.1"
                     /translation="MWLRSHRQLCLAFLLVCVLSVIFFLHIHQDSFPHGLGLSILCPD
                     RRLVTPPVAIFCLPGTAMGPNASSSCPQHPASLSGTWTVYPNGRFGNQMGQYATLLAL
                     AQLNGRAAFILPAMHAALAPVFRITLPVLAPEVDSRTPWRELQLHDWMSEEYADLRDP
                     FLKLSGFPCSWTFFHHLREQIAREFTLHDHLREEAQSVLGQLALGATGDRPRTFVGVH
                     VRRGDVLQVMPQRWKGVVGDSAYLRQAMDWFRARHEAPVFVVTSNGMEWCKENIDTSQ
                     GDVTFAGDGQEATPWKDFALLTQCNHTIMTIGTFGFWAAYLAGGDTVVLANFTLQTLS
                     S"
                     /transl_table=1
     mutation        990
                     /replace="G to none"
BASE COUNT      184 a    368 c    305 g    240 t      0 others
ORIGIN
        1 atgtggctcc ggagccatcg tcagctctgc ctggccttcc tgctagtctg tgtcctctct
       61 gtaatcttct tcctccatat ccatcaagac agctttccac atggcctagg cctgtcgatc
      121 ctgtgtccag acagccgcct ggtgacaccc ccagtggcca tcttctgcct gccgggtact
      181 gcgatgggcc ccaaagccta ctcttcctgt ccccagcacc ctgcttccct ctccggcacc
      241 tggactgtct acccccaatgg ccggttttggt aatcagctgg gacagtatgc cacgctgctg
      301 gctctggccc aggtcaacgg ccgccgggcc tttatcctgc ctgccatgcc tgccgccctg
      361 gccccggtat tccgcatcac cctgccccgtg ctggccccag aagtggacag ccgaccgccg
      421 tggcggggag tgcagcttca cgactggatg tcggaggagt acgcggactt gagagatcct
      481 ttcctgaagc tctctggctt ccctgtctct tggcctttct tccaccatct cagggaccag
      541 atccgcagag agttcaccct gcacgaccac attcgggaag aggcgcagag tgtgctgggt
      601 cagctcgcgc tgggccgcac aggggaccgc cagcgcaccc ttgtcaggcg tccacgtgcg
      661 cgtggggact atctgcaggt tatgcctcag cgctggcagg gtgtggtggg cgacagcgcc
      721 tacctccggc aggccatgga ctggttccgg gcacggcacg aagccccagt tttcgtggtc
      781 accagcaacg gcatggagtg gtgtaaagaa aacatcgaca cctcccaggg cgatgtgacg
      841 tttgctggcg atggacagga ggctacacag tggcaagact ttgccatgct aacaacagtg
      901 aaccacacca ttatgaccat tggcacttc ggcttctggg ctgcctacct ggctggcgga
      961 gccactgtct acctggccaa cttcaccctc cagcctctga gttcctgaag atctttaagc
     1021 cggaggcggc cttcctgccc gagtgggtgg gcattaatgc agacttgtct ccactctgga
     1081 cattggctaa gccttga
//
```

Fig. 5. Example entry.

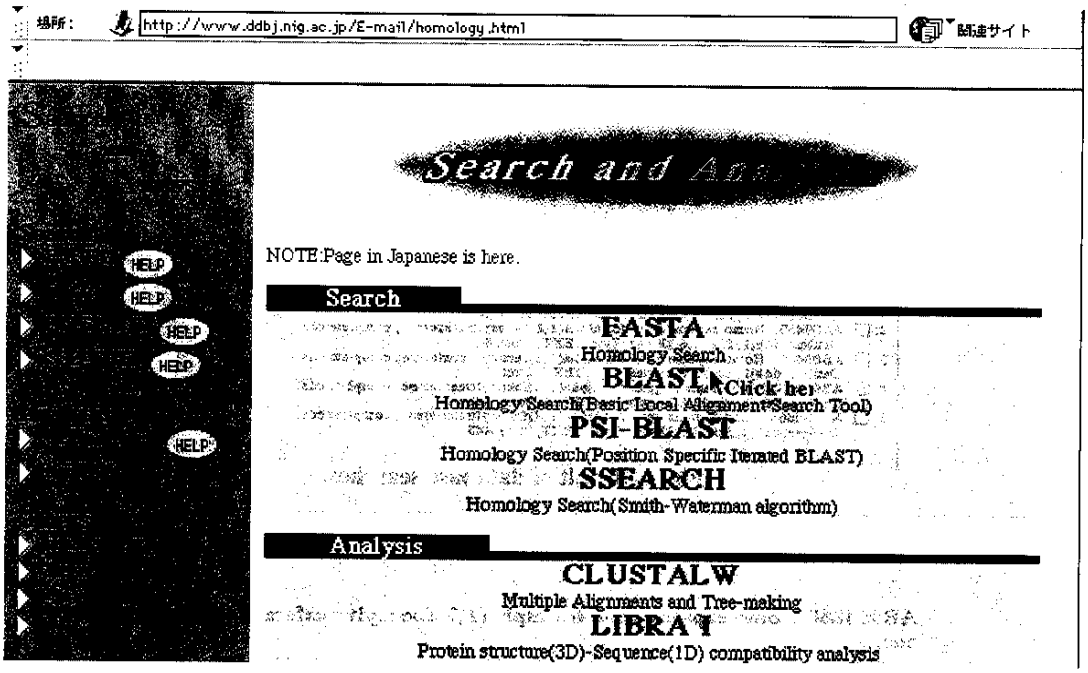


Fig. 6. Search and analysis.

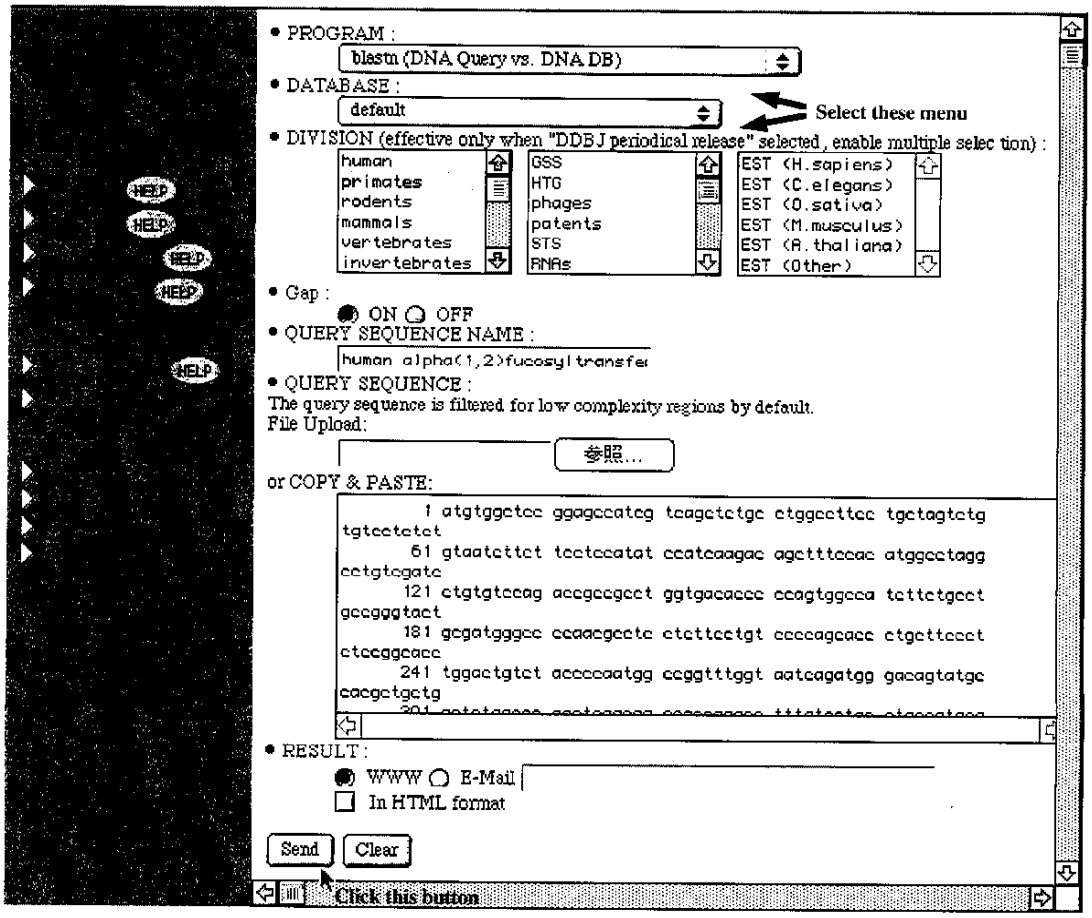


Fig. 7. Blast search.

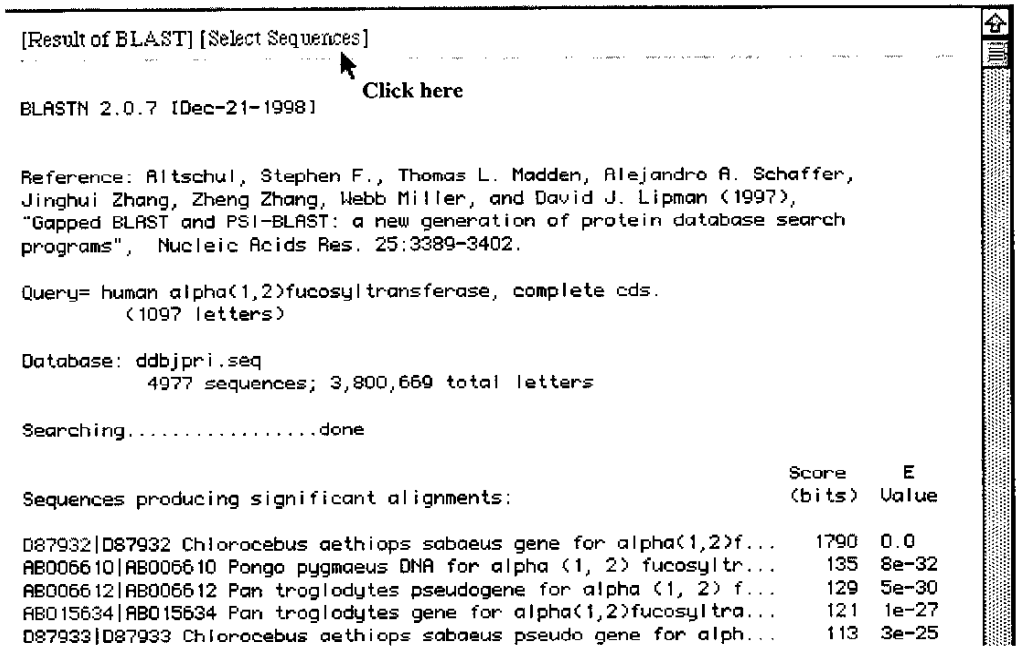


Fig. 8. Result of blast search.

can be specified by using double or single quotation marks), and click "start search".

Fig. 4 shows the result of database searches using the four key words above. In this example, five database entries were retrieved. Click an accession number and the corresponding entry will appear (Fig. 5).

C. Homology Search Using FASTA, SSEARCH and BLAST

You can analyze your nucleotide and/or amino acid sequences obtained by using homology searches (FASTA/SSEARCH/BLAST).

Click "Homology Search using FASTA, SSEARCH and BLAST" of "Database Searches" in the "Database Searches and Data Analyses" page (Fig. 2). The "Search and Analysis" page will appear (Fig. 6).

Select the homology search method in "Scarch". In this example, "BLAST" has been selected. Click "BLAST", and the "BLAST Search" page will appear (Fig. 7).

Specify the search program from "PROGRAM :". Choose one of four programs. The default is "blastn" (compares your DNA sequence with nucleotide sequence database). Specify the database in which homologous sequences are to be searched. Several databases are currently available. "DNA" is a non-redundant database which contains all available nucleotide sequences at present. Specify the division in which homologous sequences are to be searched. This option is effective only when "DDBJ periodical release" is selected. In this example, "primates" has been selected, and it means data of non-human pri-

それぞれの単語はandでつないだので、上記の例では4つの語句をすべて持つエントリーを検索します。なお、"Options"は通常、デフォルトのままで良いと思います。

図4は上記の4つの語句を用いたときの検索結果です。この例では、5つのエントリーが得られました。それぞれのエントリーの文字をクリックすると、図5のような個々のデータが得られます。

C. 相同性検索（FASTA/SSEARCH/BLAST）による相同遺伝子の検索

自分で配列決定した遺伝子やキーワード検索によって取得した遺伝子、またはアミノ酸配列などと相同性のある遺伝子を相同性検索（FASTA/SSEARCH/BLAST）によって分析することができます。

"Database Searches and Data Analyses"のページ（図2）の"Database Searches"の項目の中の"Homology Search using FASTA, SSEARCH and BLAST"をクリックすると図6の画面に出ます。

"Search"の項目からどの方法で相同性検索を行うかを決めます。この例では、"BLAST"を選択しています。"BLAST"の文字をクリックするとデータ入力画面に入ります（図7）。

まず"PROGRAM :"で、検索オプションを指定します。解析の用途に合わせ、4つのプログラムのうちのいずれかを指定します。デフォルトでは、"blastn"（これは、あなたの塩基配列を塩基配列データベースと比較します）が設定されます。次に、"DATABASE :"で、検索対象となるデータベースを指定します。デフォルトは、"blastn"と"tblastn"の場合は"DDBJ release (DNA)"、"blastp"と"blastx"の場合は"Protein ALL (Protein)"です。"DDBJ release (DNA)"を選択した場合、"DIVISION"からより詳細に検索対象を絞り込むこともできます。ここでは"primates"を選択しています。これは、ヒト以外の霊長類のデータを意味し

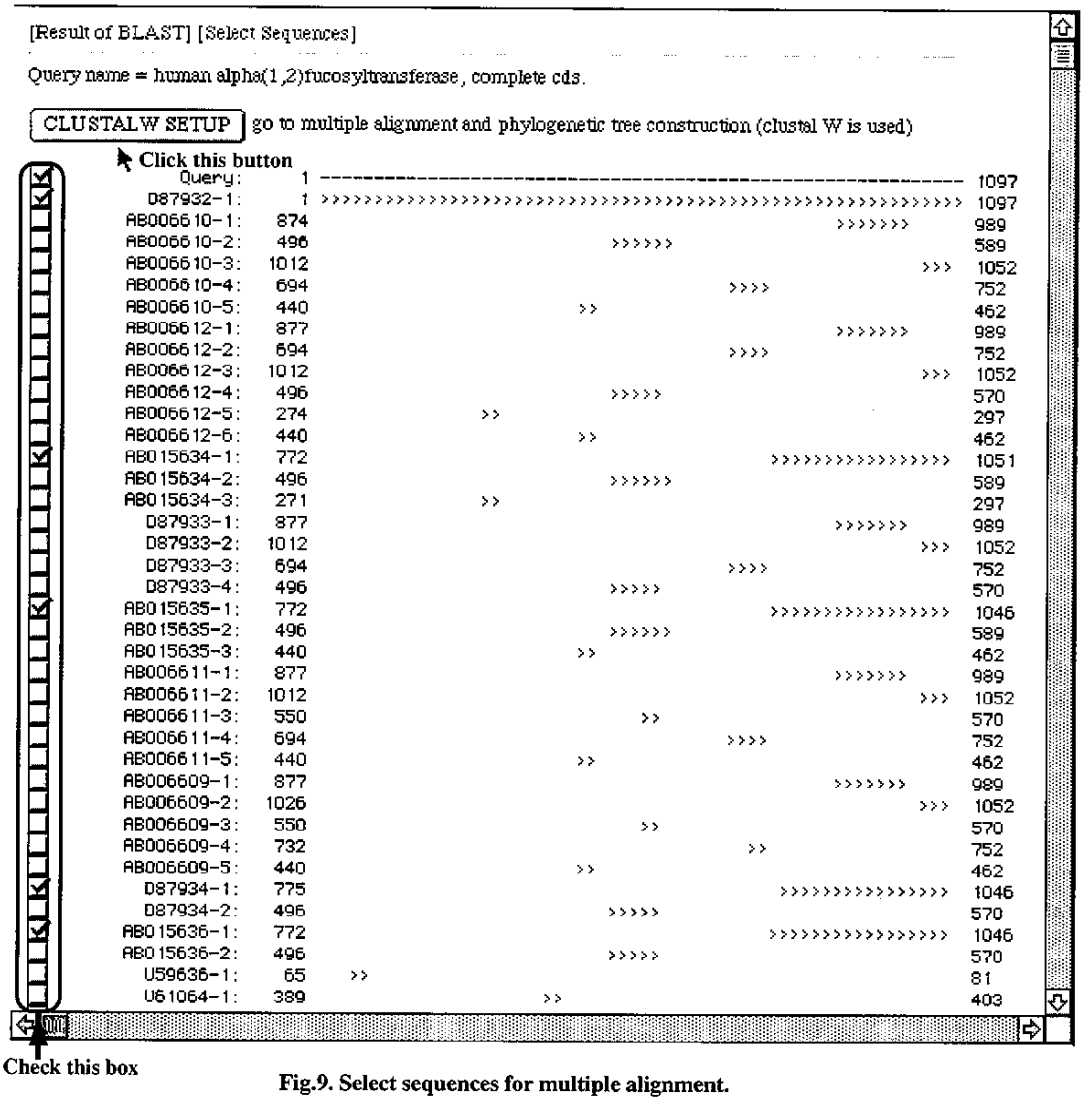


Fig.9. Select sequences for multiple alignment.

mates. Input a name to classify the query sequence in "QUERY SEQUENCE NAME :". Input a sequence only (The sequence is not FASTA FORMAT.) in "QUERY SEQUENCE :". You can copy&paste the nucleotide sequence data of Fig. 5. Select how to receive the results. When you select "WWW", the results are shown in WWW. When you select "E-Mail", you receive the results by E-mail. If "In HTML format" is checked, the results will be loaded into a web browser for viewing. Then click the "Send" button. You can select more detailed options in the "detailed option" section.

Figure 8 shows a sample result of a BLAST Search. From the left, accession number (you can see each data by clicking this), LOCUS and DEFINITION, Score (you can see each multiple alignment by clicking this), and E value. Results of each multiple alignment are shown on the button part of this page.

ています。"QUERY SEQUENCE NAME :"には、あなたの塩基配列の名前を指定します。"QUERY SEQUENCE :"に検索配列を指定します。ここでは、配列のみをテキストボックスに入力して下さい。(FASTAフォーマットではありません。) カットアンドペーストで図5の塩基配列の部分を入れることもできます。次に、検索結果の受け取り方法を指定します。"WWW"を指定した場合、検索結果をWWW画面に表示します。"E-Mail"を指定した場合、検索結果を指定されたE-Mailアドレス先に送信します。また、"E-mail"を指定したときに"In HTML format"をチェックするとHTML形式の添付ファイルが送られます。そして、"Send"ボタンをクリックします。なお、それ以下の項目 ("detailed option") で、より詳細な設定を行うことも可能です。

図8はBLASTの検索結果の例を示しています。左から、該当エントリーのACCESSION番号 (これをクリックするとそのデータを見ることができます) 、LOCUSとDEFINITION、Score (ここをクリックするとBLAST 結果の多重整列部分に移動します) 、およびE valueです。その下に、BLAST結果の多重整列部分が記入されています。

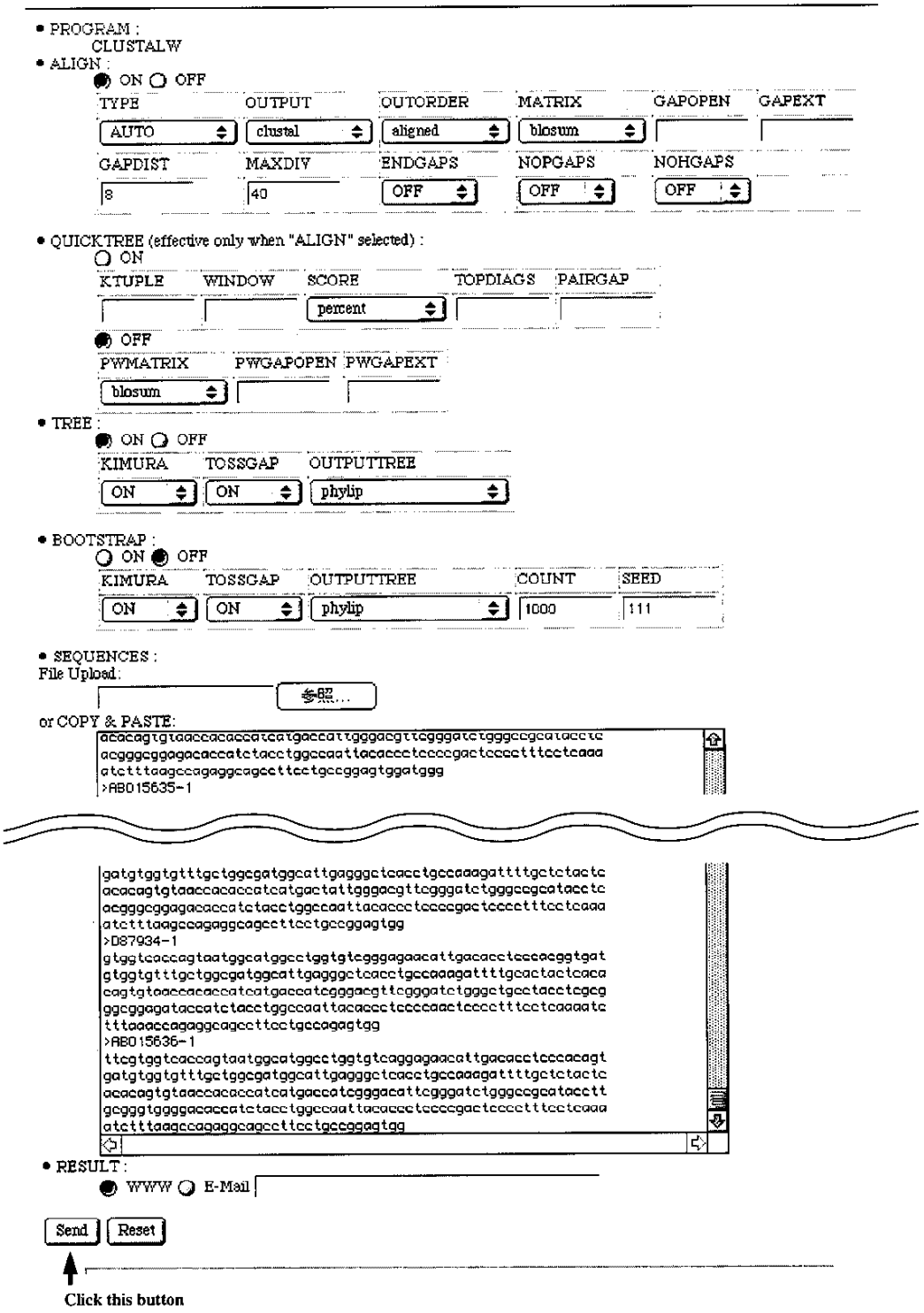


Fig.10. Clastal W Analyzing system.

D. Multiple Alignment Using Clustal W

Click the text "Select Sequences" of the results of BLAST (Fig. 8). The page for selecting sequences for the multiple alignment using Clustal W will appear (Fig. 9). Using this page, select sequences for multiple alignment, and check the boxes on

D. Clustal Wを用いた多重整列

BLAST検索結果画面（図8）の"Select Sequences"をクリックすると、多重整列を行う配列を選択する画面に入ります（図9）。この画面を用いて、多重整列を行う配列を選択し、左側のボックスをチェックします。ここでは、5つの配列をチェックし

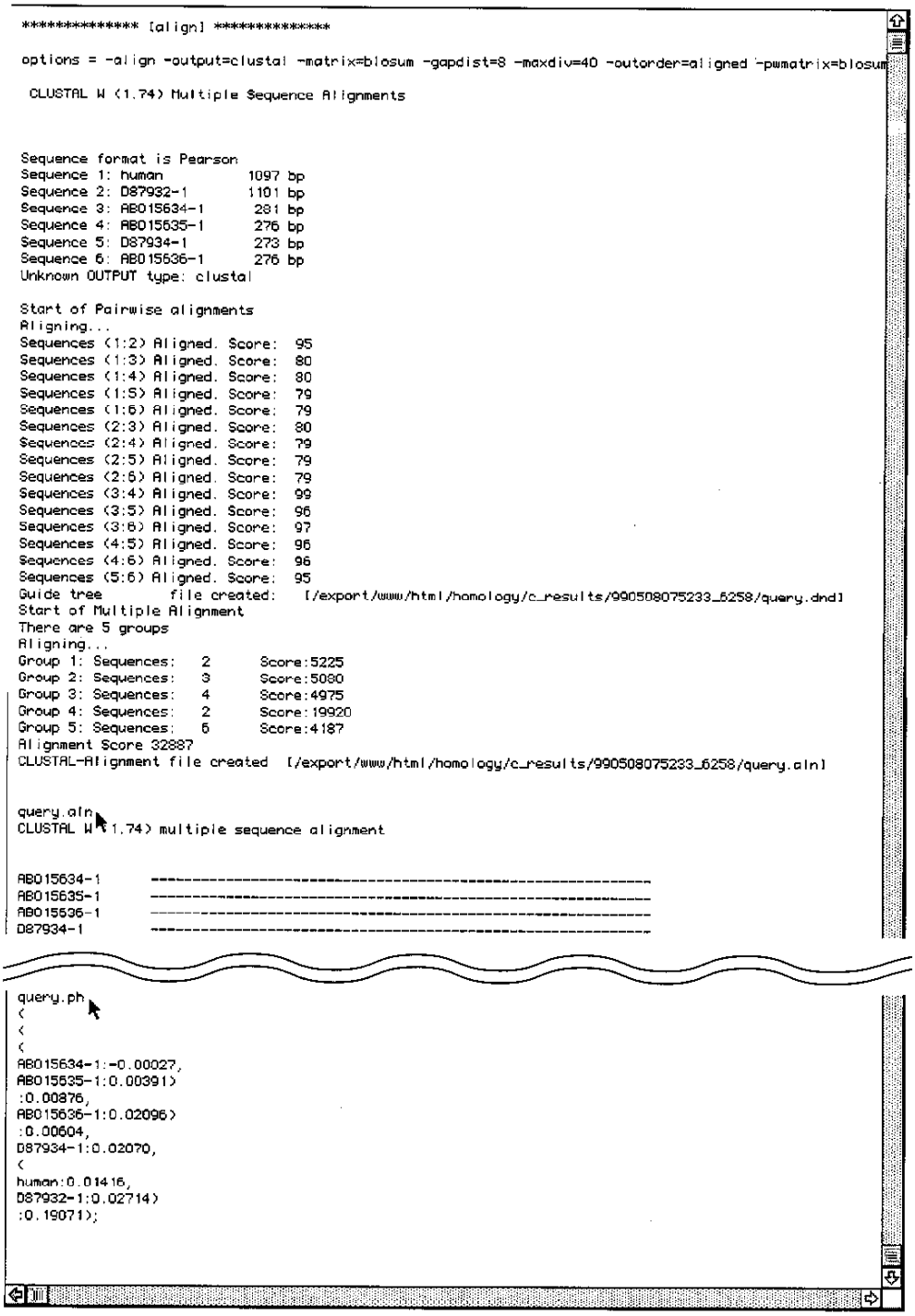


Fig. 11. Result of multiple alignment.

the left. In this example, five sequences has been checked. Then, click "CLUSTALW SETUP", and the "Clustal W Analyzing System" screen will appear (Fig. 10).

Select "type of sequences", "output format", "order of sequences", "model matrix (in the case of amino acid sequences)", and some gap parameters of "ALIGN :".

ています。それから、"CLUSTALW SETUP"ボタンをクリックすると、Clustal Wを用いた多重整列を行う画面に入ります（図10）。

まず、"ALIGN :"の項目で、解析する配列のタイプ（塩基かアミノ酸か）、出力形式、配列の出力順序、アミノ酸を用いたときの遷移置換行列表、およびいくつかのギャップ変数を設定します（Optionsは通常、そのままで良いと思います）。次の

"QUICKTREE (Effective only when "ALIGN" is selected) : " use FAST algorithm for the alignment guide tree. In "TREE :", select "KIMURA (use Kimura's correction. Default: ON)", "TOSSGAPS (ignore positions with gaps. Default: ON)", and "OUTPUTTREE (Phylip, clustal, Phylip or distance. Default is phylip)". If you want to obtain bootstrap probabilities, you should check "BOOTSTRAP :". Selected sequences from Fig. 9 are inputted in the "SEQUENCES :" window. Then click the "Send" button.

Figure 11 shows the result of Clustal W analyzing system. The multiple alignment is shown below "query.aln". You can download this data by clicking "query.aln". The tree data is shown below "query.ph". You can download this data by clicking "query.ph". The corresponding tree can be shown by some application such as TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) and DendroMaker (http://www.cib.nig.ac.jp/dda/timanish/dendromaker/home. html).

You also enter the "Clustal W Analyzing System" from the screen of "Search and Analysis" (Fig. 6).

"QUICKTREE (ALIGN 指定時のみ有効) :"は、FAST algorithmで実行するときに用います。速く計算したいときに用いると良いかもしれません。次の"TREE :"の項目では、"KIMURA"（系統樹計算時に、Kimura's correction を使用するかどうか）、"TOSSGAPS"（系統樹計算時に、gapを無視するかどうか）、"OUTPUTTREE"（系統樹計算結果の出力フォーマット）を指定します。ブートストラップの計算を行ないたいときは"BOOTSTRAP :"の項目で設定します。"SEQUENCES :"のウインドウにはさきほどの図9で選択した配列が入っています。そして、"Send"のボタンをクリックします。

図11はClustal Wによる解析結果を示しています。"query.aln"の下に多重整列の結果が表示されています。また、"query.aln"をクリックすると、このデータがダウンロードされます。計算された系統樹のデータは一番下の"query.ph"に示されています。"query.ph"をクリックすると、このデータがダウンロードされます。これはTreeView（ダウンロードできるアドレスは、http://taxonomy.zoology.gla.ac.uk/rod/treeview.html）やDendroMaker（アドレスは、http://www.cib.nig.ac.jp/dda/timanish/dendromaker/home.html）などのアプリケーションで系統樹を見ることができます。

なお、Clustal Wによる多重整列は、図6のAnalysisの項目の中の"CLUSTAL W"からも入ることができます。

Profile of the Authors

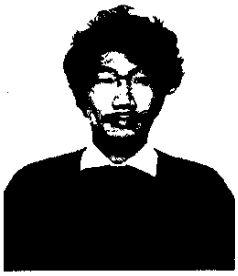
KitanoTakashi

Graduated from Hirosaki University in 1994, and obtained a Ph.D. in Genetics at the Graduate University for Advanced Studies in 1999. He is currently COE postdoctoral fellow at National Institute of Genetics.

(Main themes)

He studied the evolution of Rh blood group genes using both experimental and statistical techniques.

HHHHHHHHHH



Saitou Naruya

Graduated from the University of Tokyo in 1979, and obtained a Ph.D. in Genetics at University of Texas at Houston in 1986. He is currently associate professor at the National Institute of Genetics and at the Graduate University for Advanced Studies (joint appointment).

(Main themes)

He is studying the evolution of genes in various organisms, in particular humans.