Sequence Analyses Using DDBJ Home Page

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

Fig. 1. DDBJ home page.
Database Searches and Data Analyses

Database Searches

- Data Retrieval by accession number using Genbank
- Data Retrieval by key words using SFGate & WAIS
- Data Retrieval by key words using Sequence Retrieval System (SRS) NEW
- Homology Search using FASTA, SSEARCH and BLAST
- Generate information Bridge for Macintosh Computers
- A simple and improved browser for Protein Data Bank (PDB Retrieval)
- Retrieval of Related sequence database (TIGEFER) NEW
- Distributed databases for homology searches AND Multiple aligned sequence database (GAPUZ Database) NEW

Data Analyses

- Multiple Alignment using SALIGN
- Multiple Alignment using Clustal
- Prediction of protein secondary structure using neural
- Protein Structure(3D)-sequence(1D) comparability analysis by LIBSA

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 '-' and 'or' "Database Searches and Data Analyses")'s page will appear (Fig. 2). "Database Searches"'s project of "Data Retrieval by key words using SFGate & WAIS" will come up with and choose "Database Searches". Select one of the following databases: "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 '-' and 'or'). ©1999 Fcca (Forum: Carbohydrates Coming of Age)
Database Searches

Your query was:
DNA sequences of apricot (Prunus armeniaca) as determined by the restriction site analysis

Results of Search(1-5)

<table>
<thead>
<tr>
<th>No</th>
<th>Description</th>
<th>Source</th>
<th>Query</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AB004679 Apple 16S rRNA gene, complete cds.</td>
<td>DNA Data Bank of Japan (DDBJ)</td>
<td>Apple 16S rRNA</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>AB004680 Apple 23S rRNA gene, complete cds.</td>
<td>DNA Data Bank of Japan (DDBJ)</td>
<td>Apple 23S rRNA</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>AB004681 Apple 18S rRNA gene, complete cds.</td>
<td>DNA Data Bank of Japan (DDBJ)</td>
<td>Apple 18S rRNA</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>AB004682 Apple 25S rRNA gene, complete cds.</td>
<td>DNA Data Bank of Japan (DDBJ)</td>
<td>Apple 25S rRNA</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>AB004683 Apple 28S rRNA gene, complete cds.</td>
<td>DNA Data Bank of Japan (DDBJ)</td>
<td>Apple 28S rRNA</td>
<td>50</td>
</tr>
</tbody>
</table>

Fig. 4. Result of database searches.

Your query was:
blood and tissue sequences of apricot (Prunus armeniaca) and 'tanenho, m.'

Fig. 5. Example entry.

©1999 FCCA (Forum: Carbohydrates Coming of Age)
Fig. 6. Search and analysis.

Fig. 7. Blast search.
**Fig. 8. Result of blast search.**

A result of BLAST search is shown in the following table. The search was performed on the *human alpha(1,2)fucosyltransferase, complete cds.* database, and the search criteria were 

<table>
<thead>
<tr>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1700</td>
<td>0.0</td>
</tr>
<tr>
<td>135</td>
<td>9e-32</td>
</tr>
<tr>
<td>129</td>
<td>5e-20</td>
</tr>
<tr>
<td>121</td>
<td>1e-27</td>
</tr>
<tr>
<td>113</td>
<td>3e-25</td>
</tr>
</tbody>
</table>

The **Score** column represents the significance of the alignment, and the **E Value** column gives the probability of the alignment occurring by chance. A lower **E Value** indicates a more significant alignment.

**C. Homology Search Using FASTA, SSEARCH and BLAST**

You can analyze your nucleotide or amino acid sequences obtained by using homology searches (FASTA/SSSEARCH/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 "Search". 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, "pri-mates" has been selected, and it means data of non-human pri-

それぞれの単語はandでつながないので、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/SSSEARCH/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得られた遺伝子、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/SSSEARCH/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"の項目中の"Homology Search using FASTA, SSEARCH and BLAST"をクリックすると図6の画面に出現します。

"Search"の項目からのどのように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、検索オプションを設定します。解析の用途に合わせて、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「blastp」と「blastsx」の場合は「DDBJ release (DNA)」、「blastp」と「blastsx」の場合は「Protein ALL (Protein)」です。

DDBJ release (DNA)を選びとした場合、「DIVISION」からより詳細に検索条件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、ヒト以外の塩基配列のデータを意味し
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 “de-
tailed 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 mul-
tiple alignment by clicking this), and E value. Results of each
multiple alignment are shown on the button part of this page.
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 the "Align" page. Click this button.

Fig.10. Clastal W Analyzing system.

D. Clustal W を用いた多重列列

BLAST検索結果画面（図8）の"Select Sequences"をクリックすると、多重列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、多重列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:”.
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 (Phylib, clustal, Phylib or distance. Default is phylib)". 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).