

# Taxonomy リストの変更方法

## 1. はじめに

SwissProt や NCBIprot のような総合配列データベースは様々な Taxonomy に由来するタンパク質エントリから構成されていますので、Mascot 検索を実行する際、Taxonomy(たとえば[Homo sapiens])を指定して検索範囲を絞ることができます。検索対象としたい Taxonomy は図 1 に示すように検索条件設定画面の[Taxonomy]リストから Taxonomy を選択することにより指定します。

検索条件設定画面の[Taxonomy]リストに表示される Taxonomy は「C:\inetpub\wwwroot\mascot\config\taxonomy」ファイルの内容が反映されています。「taxonomy」ファイルはテキストで書かれていますので、メモ帳のようなテキストエディタを使って開き(あるいは[スタート]→[すべてのプログラム]→[Mascot]→[admin]→[Mascot taxonomy file]を選択して開き)、Taxonomy の追加・削除・内容変更などの編集を行うことができます。

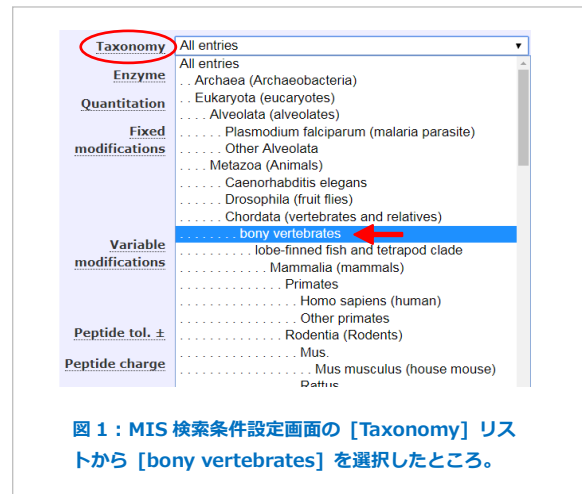


図 1 : MIS 検索条件設定画面の [Taxonomy] リストから [bony vertebrates] を選択したところ。

## 2. 「taxonomy」ファイルの構造

「taxonomy」ファイルは次のような構造になっています。

```
Title:All entries
Include: 1
Exclude: 0
*
Title:. . Archaea (Archaeobacteria)
Include: 2157
Exclude:
*
Title:. . Eukaryota (eucaryotes)
Include: 2759
Exclude:
*
(以下省略)
```

ひとつのエントリは「Title:」行で始まり、「\*」行で終わります。

「Title:」行の文字列は検索条件設定画面の [Taxonomy] リストに表示されます。通常は、生物種の名称や意味が把握できる短い文字列で構成します。また、Taxonomy のツリー構造を示すために、インデントとして「...」や「ひとつのスペース」を使うことができますが、タブや連続する複数のスペースを使用することはできません。

「Include:」行は、ひとつ以上の「Taxonomy ID」で構成され、複数の「Taxonomy ID」を記述するときは互いにカンマで区切ります。通常は、NCBI(米国 National Center for Biotechnology Information)が提供している

「Taxonomy ID」を適用します。

「Exclude:」行は、ひとつ以上の「Taxonomy ID」で構成され、複数の「Taxonomy ID」を記述するときは互いにカンマで区切ります。通常は、NCBI が提供している「Taxonomy ID」を適用します。

「Include」から「Exclude」を “差し引いた” Taxonomy が検索対象となります。

### 3. Taxonomy ID を調べる方法

[Taxonomy] リストに追加したい Taxonomy に対応する「Taxonomy ID」を調べるには次の 2 つの方法があります。ここでは「Homo sapiens」を例にとって説明します。

#### (1) 「names.dmp」ファイルから探す方法

「C:\inetpub\mascot\taxonomy\names.dmp」ファイルには「Taxonomy ID」に対応する「Taxonomy 名」が記載されています。テキストエディタを使って「names.dmp」ファイルを開き、「Homo sapiens」の文字列を検索してください。「Homo sapiens」の ID は 9606 ですが、次のように記載されています。

9606		Homo sapiens			scientific name	
9606		Homo sapiens Linnaeus, 1758			authority	
9606		human			genbank common name	
9606		man			common name	

#### (2) NCBI の「Taxonomy Browser」を利用する方法

Web ブラウザから次の URL にアクセスし、NCBI の「Taxonomy Browser」ページを開いてください。

<http://www.ncbi.nlm.nih.gov/Taxonomy/>

[Search] の入力欄から「Homo sapiens」で検索し、表示されたページの[Homo sapiens]のリンクをクリックしてください。次のような内容が表示され、「Homo sapiens」の「Taxonomy ID」は 9606 であることがわかります。

<b>Homo sapiens</b>	
<i>Taxonomy ID:</i> 9606	
<i>Genbank common name:</i> human	
<i>Inherited blast name:</i> primates	
<i>Rank:</i> species	
<i>Genetic code:</i> <a href="#">Translation table 1 (Standard)</a>	
<i>Mitochondrial genetic code:</i> <a href="#">Translation table 2 (Vertebrate Mitochondrial)</a>	
<i>Other names:</i>	
common name:	man
authority:	Homo sapiens Linnaeus, 1758

### 4. 「taxonomy」ファイルの編集例

Mascot 検索条件設定画面の[Taxonomy] リストに、たとえば「Ferns or human」を追加したい場合は、「C:\inetpub\mascot\config\taxonomy」ファイルに次の行を追加し、保存してください。追加した Taxonomy は即座に反映され、図 2 のように、[Taxonomy] リストに表示されます。

Title: Ferns or human

Include: 9606, 3263

Exclude:

\*

同様に、「No human or mice」を表示させたい場合は次の行を追加してください。

Title: No human or mice

Include: 1

Exclude: 9606, 10088

\*

なお、「すべての Taxonomy」を表すルート「root」の Taxonomy ID は1です。

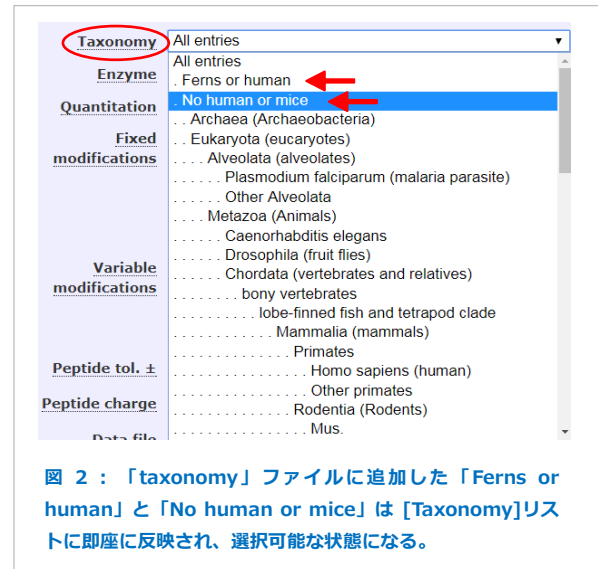


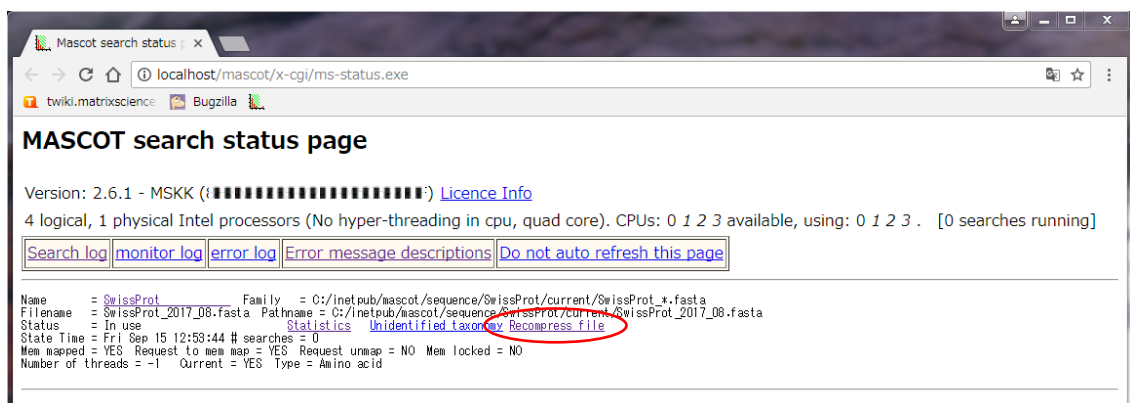
図 2 : 「taxonomy」ファイルに追加した「Ferns or human」と「No human or mice」は [Taxonomy] リストに即座に反映され、選択可能な状態になる。

## 5. 配列データベースの再セットアップ

Mascot Server は配列データベースをセットアップする際、「taxonomy」ファイルの内容に従ってタンパク質エントリーを Taxonomy 毎に仕分けしています。「taxonomy」ファイルを変更した場合は配列データベースを再度セットアップする必要がありますので、次の操作を行ってください。ここでは代表的な総合配列データベースである SwissProt を例にとって説明します。

### (1) Mascot Server version 2.6 の場合

「MASCOT search status page」ページ (Welcome トップページの [Database Status] リンクをクリック) の「SwissProt」ブロックにある [Recompress file] リンクをクリックしてください。即座に再セットアップが始まります。「Status = Creating compressed files 14% complete」のように表示されますので進捗状況を確認してください。「Status = In use」になればセットアップ完了です。



### (2) Mascot Server version 2.5 までの場合

Mascot Server を停止 ([スタート] → [すべてのプログラム] → [Mascot] → [admin] → [Stop Mascot service]) を選択させた後、「C:\inetpub\mascot\sequence\SwissProt\current\SwissProt\_日付.stats」ファイルを削除し、Mascot Server を起動 ([スタート] → [すべてのプログラム] → [Mascot] → [admin] → [Start Mascot service]) を選択) してください。即座に再セットアップが始まります。「Status = Creating compressed files 14% complete」のように表示されますので進捗状況を確認してください。「Status = In use」になればセットアップ完了です。

