

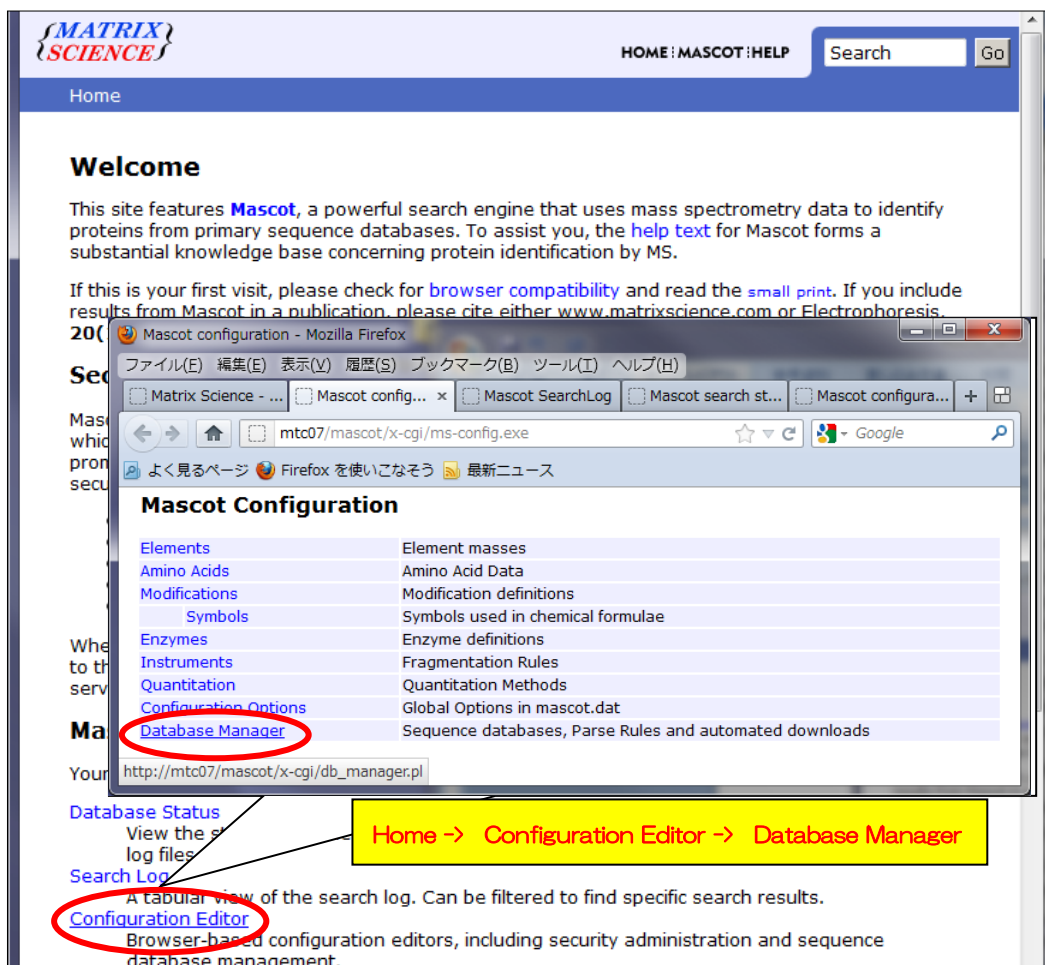
# 配列データベース管理マニュアル

## ● はじめに (Database manager 画面の位置)

本資料は、データベースの設定を行う管理画面「Dataabse Manager」に関する資料です。「Database Manager」ではデータベースファイルの追加・設定変更、データベース自動更新やそれに伴うネットワーク接続について設定できます。

Database manager のページを開くには、

Web ブラウザで Home → Configuration Editor → **Database Manager** とリンクをたどります。



## ● 本資料で取り扱う内容と掲載ページ

- **Database Manager 初期設定** **P.3**
  - ▼ 新規インストール後の初期設定画面 P.3
  
- **Section 別 設定内容 :: Settings Section** **P.7**
  - ▼ MASCOT Server のInternet 接続の可否 P.9
  - ▼ MASCOT Server の 外部公開プログラムへの接続の可否 P.9
  - ▼ プロキシ設定 P.10
  - ▼ 配列データベース設置場所の変更 P.11
  
- **Section 別 設定内容 :: Database Manager Section** **P.12**
  - ▼ 使用データベースのON/OFF, データベースファイルの取得（更新） P.12
  - ▼ **データベースファイルの定期的な自動取得・更新設定** P.15
  - ▼ **新規データベースの追加** P.17
    - predefined database definition **P.18**                      - template **P.22**
    - copy **P.29**    - custom **P.33**
  
- **Section 別 設定内容 :: Tasks Section** **P.43**
  - ▼ データベースダウンロード進捗の確認 P.43

## ● Database Manager 初期設定

### □ 新規インストール後の初期設定画面

新規インストール後最初に Database Manager を開くと、下図のように初期設定画面が現れます。使用可能な状態にあるデータベースについて、「Predefined database definition」（後述、→P. 18）と照らし合わせてどの項目にマッチするかを確認し項目を選択する画面です。通常は**デフォルト選択項目のまま「Import」ボタンを押します。**

**Mascot Database Manager Configuration Import**

**Note:**  
Please disable other programs and scripts that configure or update sequence databases before continuing.

Database Manager can keep your sequence database configuration up to date and schedule the download of new files.

If you choose to use Database Manager, you should not edit mascot.dat in future, because your changes will be lost. If you do not wish to use Database Manager now, please [return to the Configuration Editor](#).

Database Manager has two kinds of database definitions:

1. *Predefined database definition*: Automatic configuration by downloading up-to-date information from the Matrix Science website.
2. *Custom database definition*: Manually maintained configuration.

You can choose to keep the existing definition (Keep as Custom) or synchronise with a predefined definition (Synchronise). Synchronisation means the predefined database definition will be applied, and the database files (FASTA, reference, taxonomy) will be updated.

If you do not want to update the files now, choose Keep as Custom. You can synchronise at any time in the future by choosing Synchronise Custom Definitions from the Database menu.

Database definition	Action	Match quality
SwissProt	<input type="radio"/> Keep as Custom <input checked="" type="radio"/> Synchronise with SwissProt_ID <input type="radio"/> Synchronise with SwissProt_AC poor	perfect Differences

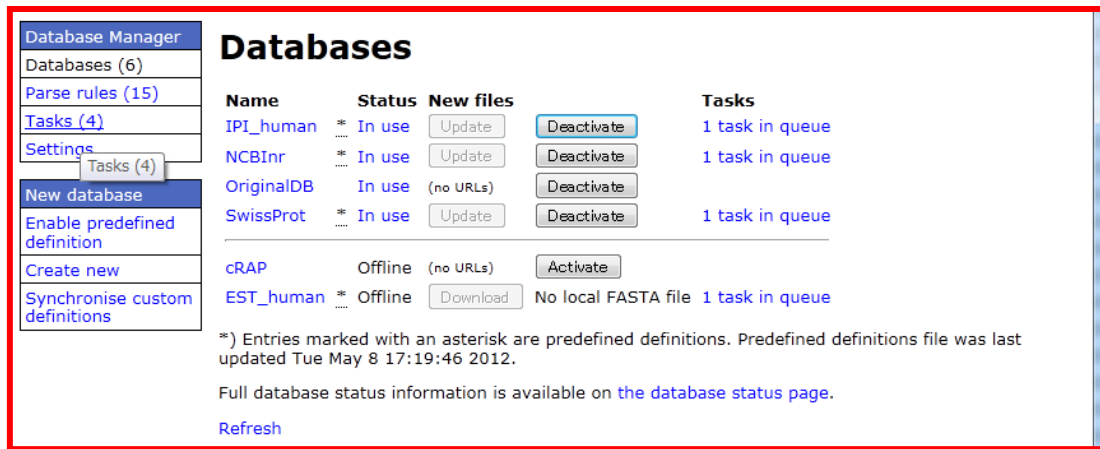
Import database configuration by choosing Import.

Import

通常はパラメータを変更することなく「Import」

Database Manager 初期設定画面

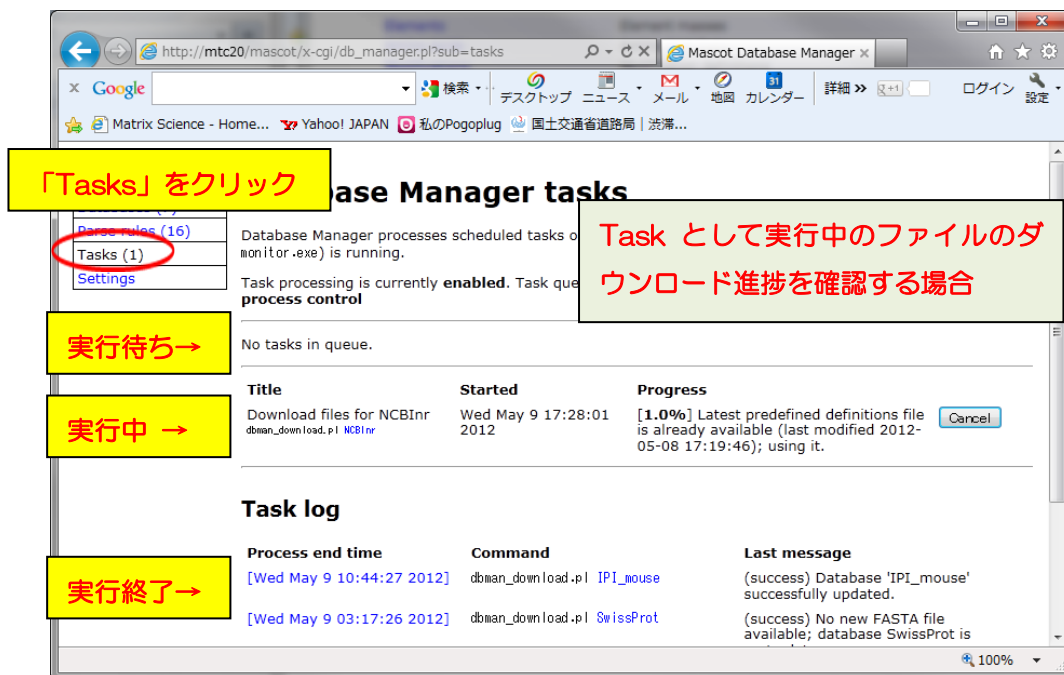
Importボタンを押すと画面が切り替わります。



Import ボタンクリック後自動的に表示される「Databases」画面（表示内容は上記と異なる場合があります）

### 【データベースファイルの自動取得とデータベース更新の進捗を確認する方法】

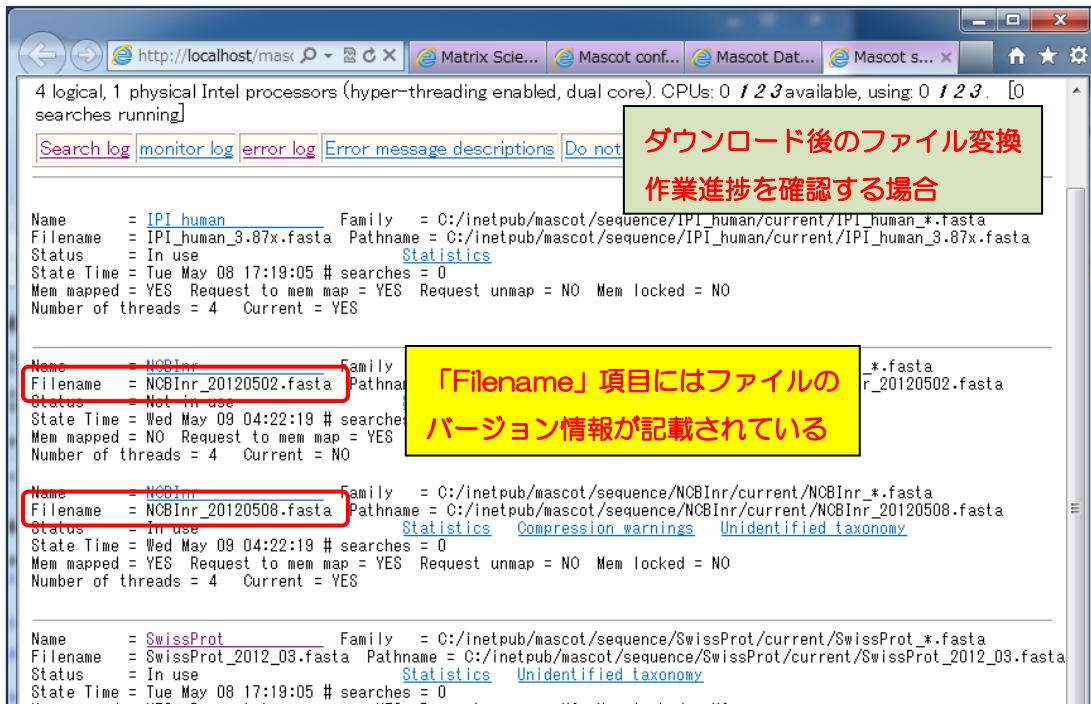
定義を確認したデータベースについて、データベースファイルのダウンロード・更新が自動的に試みられます。更新の進捗を確認するには下図のように「Tasks」セクションの情報を確認してください。左側のtable内にある「Database Manager」→「Tasks」をクリックすると、現在実行・実行待ちとなっているTaskが一覧で表示されます。Taskは「実行待ち」「実行中」「実行終了」の各セクションに分かれて表示されています。



Database Manager : Tasks 画面

ファイルダウンロード終了後、ダウンロードしたファイルをMASCOT で検索できるようバックグラウンド処理でファイル変換が行われます。変換状況は、**Database Status** (WEBブラウザを起動し、**Home** → **Database Status**)で各データベースの更新状況が確認できます。

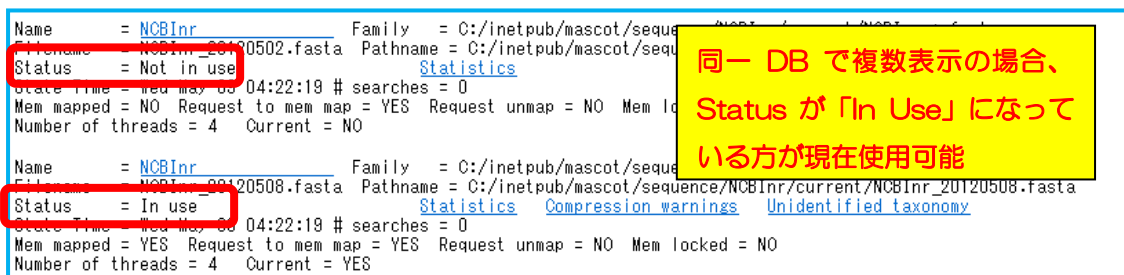
各データベースの「Filename」項目に、(Database名)\_(バージョンまたは日付).fasta と表示されています。後部の(バージョンまたは日付)でデータベースの新しさを確認してください。



Database Status 画面: 構築中のデータベース

図例のNCBIInrのように、2つの日付のデータベースが表示されているケースがあります。

「Filename」項目と「Status」項目を見比べてください。例図では、日付が20120502 のStatus項目は「Not in use」となっています。一方より新しい20120508の方は「In use」となっています。これは20120502の方は既に使用されておらず、新しい20120508の方が現在使用可能となっている事を表しています。



**【ネットワークご利用時Proxyサーバーをご利用のお客様のみ:Proxyサーバーの設定】**

自動更新プログラム利用のため Proxy サーバーの設定が必要です。詳細は Section 別設定内容:: Settings Section の「プロキシ設定」(P. 10) をご覧ください。

**【SwissProt以外のデータベースファイルを使用可能にする方法】**

インストール DVD 内に準備されている検索用データベースは SwissProt のみですが、MASCOT 用のデータベースとして SwissProt 以外にもインターネット上で公開されているデータベースやご自身で入手・作成されたデータベースを使用する事ができます。

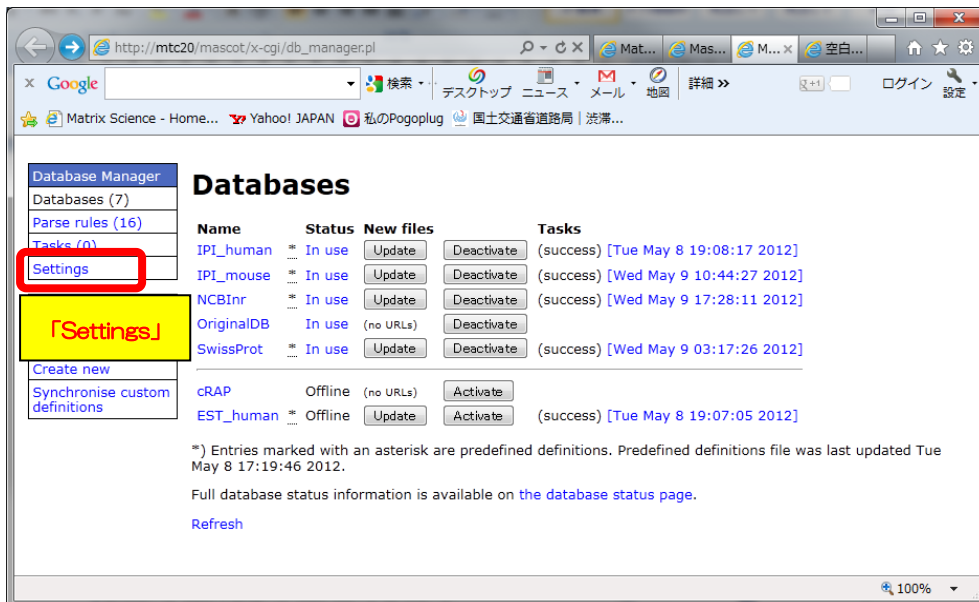
データベースを追加・使用する方法については、「**Section 別設定内容:: Database Manager Section(P. 17 )**」 をご覧ください。

## ● Section 別設定内容 :: Settings Section

Settings セクションでは大きく分けて以下の4項目の設定変更が可能です。

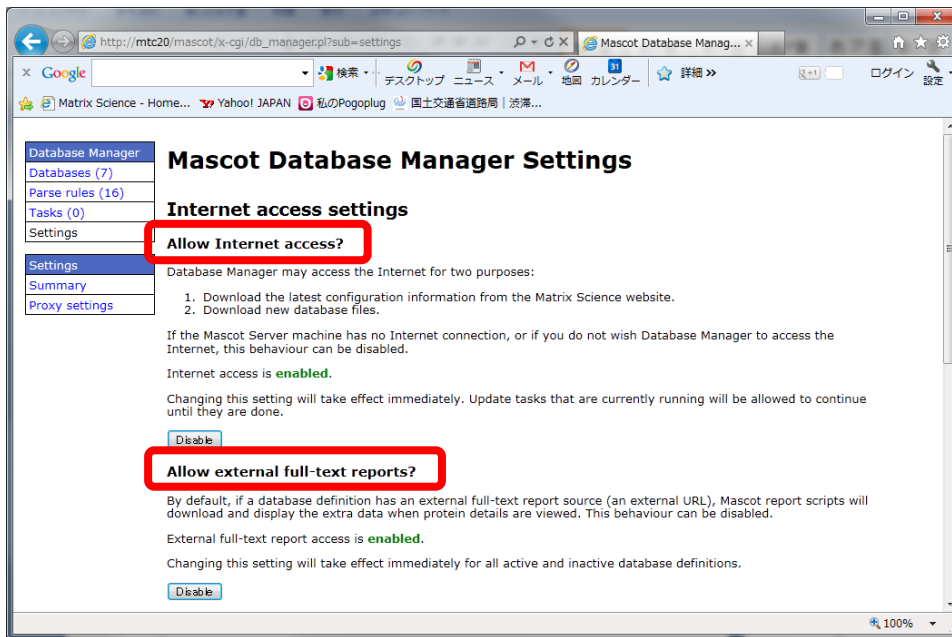
- MASCOT Server のInternet 接続の可否
- MASCOT Server の 外部プログラムへの接続の可否
- プロキシ設定
- 配列データベースの格納位置の設定

Settings セクション画面は、画面左側の「Database Manager」→「Settings」をクリックして開きます。

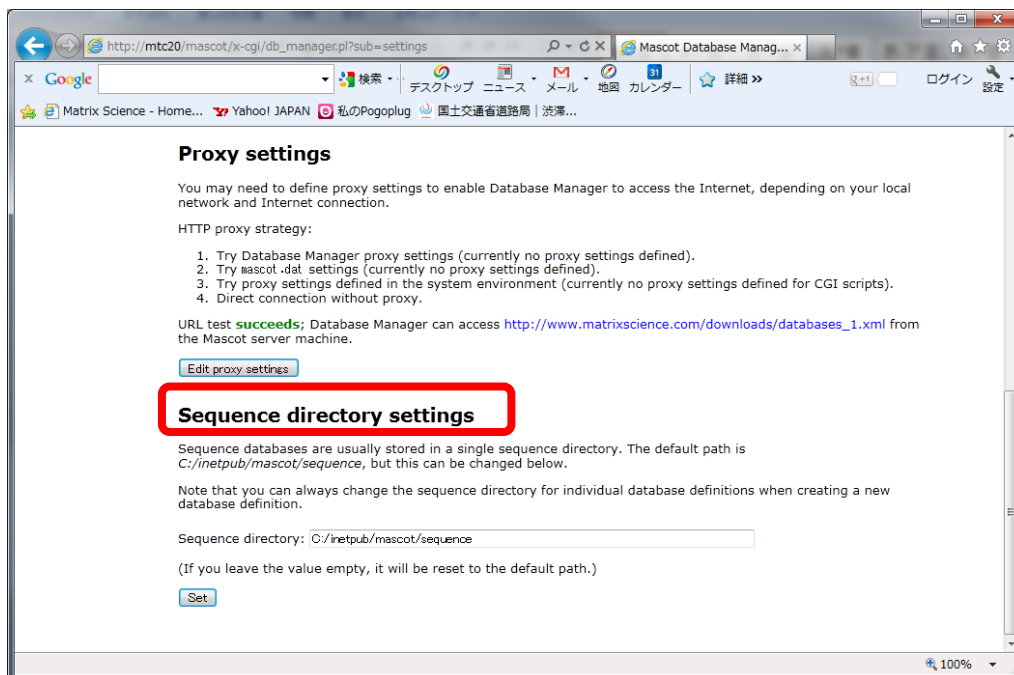


Settingsの画面（下図）は、上から順に、

「MASCOT Server のInternet 接続の可否」「MASCOT Server の 外部プログラムへの接続の可否」、「プロキシ設定」、「配列データベース格納位置の設定」に関する項目やリンクのボタンが並んでいます。



Settings 画面上部

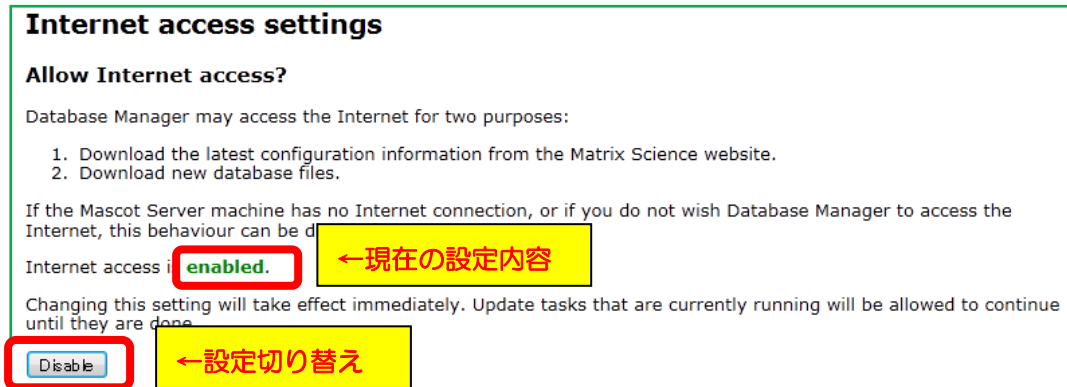


以降、各項目について詳しく説明しています。



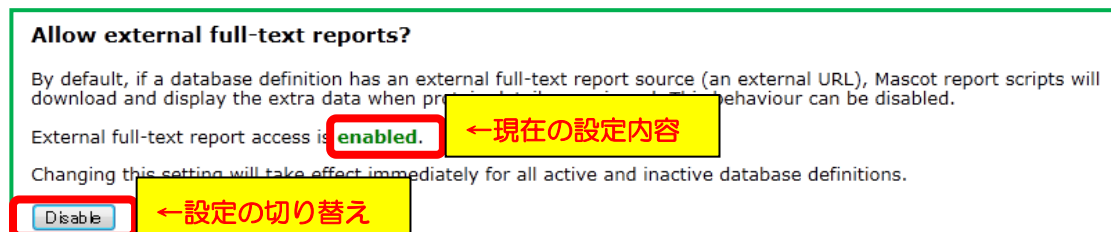
## □ MASCOTサーバーのインターネットへの接続許可 【Internet access settings】

MASCOT ではデータベースファイル取得やデータベースの定義ファイルチェックのためインターネットに接続します。これらの機能を使用せずインターネットへの接続を希望しない場合、【Internet access settings】項目で設定を変更し切り替える事ができます。



## □ MASCOT Server の 外部プログラムへの接続の可否 【Allow external full-text reports?】

MASCOT の結果画面の中でタンパク質の詳細情報を表示する箇所があります。詳細情報はローカルコンピュータにダウンロードされたファイルから取得する場合と、外部のサーバーから取得する場合があります。詳細情報を得るためにヒットタンパク質の情報を外部サーバーへ投げる事を好まない場合、【Allow external full-text reports?】項目で設定を変更する事ができます。



## □ Proxyサーバー設定【Proxy settings】

前述の通りMASCOTではデータベースファイル取得やdatabaseの定義ファイルのチェックのためインターネットに接続しますが、使用しているネットワーク環境によってはProxyサーバー経由でないとインターネットへ接続できないことがあります。【Proxy settings】項目ではProxyに関する設定を行います。

設定画面には、

【Proxy settings】

項目内の、「Edit

proxy settings」を

選択します。

**Proxy settings**

You may need to define proxy settings to enable Database Manager to access the Internet, depending on your local network and Internet connection.

HTTP proxy strategy:

1. Try Database Manager proxy settings (currently no proxy settings defined).
2. Try mascot.dat settings (currently no proxy settings defined).
3. Try proxy settings defined in the system environment (currently no proxy settings defined for CGI scripts).
4. Direct connection without proxy.

URL test **succeeds**; Database Manager can access [http://www.matrixscience.com/downloads/databases\\_1.xml](http://www.matrixscience.com/downloads/databases_1.xml) from the Mascot server ma

Edit proxy settings

プロキシ設定は「Edit proxy settings」

続いてプロキシサーバーの設定画面が現れます（下図）。画面が開く際に既に接続テストが行われており、リンク先 [http://www.matrixscience.com/downloads/databases\\_1.xml](http://www.matrixscience.com/downloads/databases_1.xml)

にアクセスできると「Succeeds」と表示します。設定画面では Proxy サーバーのアクセス先、ポートに加え、必要があればユーザー名やパスワードを入力する箇所があります。また、HTTPだけでなくFTP についても設定する箇所があります。設定後画面下部「Save」を押してください。

Database Manager

Databases (1)

Parse rules (6)

Scheduled updates

Running tasks

Settings

Proxy settings

### Mascot Database Manager Settings

#### Proxy settings

mascot.dat settings

The recommended proxy type proxy settings automatically d

**Proxy type**

Auto

**Host (HTTP proxy) (?)**

\_\_\_\_\_

**Port (HTTP proxy) (?)**

\_\_\_\_\_

**User (HTTP proxy) (?)**

\_\_\_\_\_

**Password (HTTP proxy) (?)**

\_\_\_\_\_

Use HTTP proxy for FTP (?)

Save

どこから proxy サーバーに関する情報を取得するか。基本的に「Auto」を推奨

Proxy サーバーアドレス。例: <http://proxy.mascot.jp/proxy>

Proxy ポート。例: 8080

認証ユーザー名とパスワード。使用していない場合、空欄で

## Sequence フォルダデフォルト設置箇所設定【Sequence directory settings】

MASCOT で使用するデータベースのファイルはファイル容量が大きく、しばしばファイルの置き場所に問題が生じる事があります。ファイルの置き場所を変更するには、【Sequence directory settings】の項目を書き換え、設定を適用するため「Set」ボタンを押してください(下図)。

### Sequence directory settings

Sequence databases are usually stored in a single sequence directory. The default path is `C:/inetpub/mascot/sequence`, but this can be changed below.

Note that you can always change the sequence directory for individual database definitions when creating a new database definition.

Sequence directory:  ←現在の設定内容。必要に応じて書き換える

(If you leave this field blank, the sequence directory will be reset to the default path.)

←設定更新の適用

Proxy Settings 画面

## ● Section 別 設定内容 :: Database Manager Section

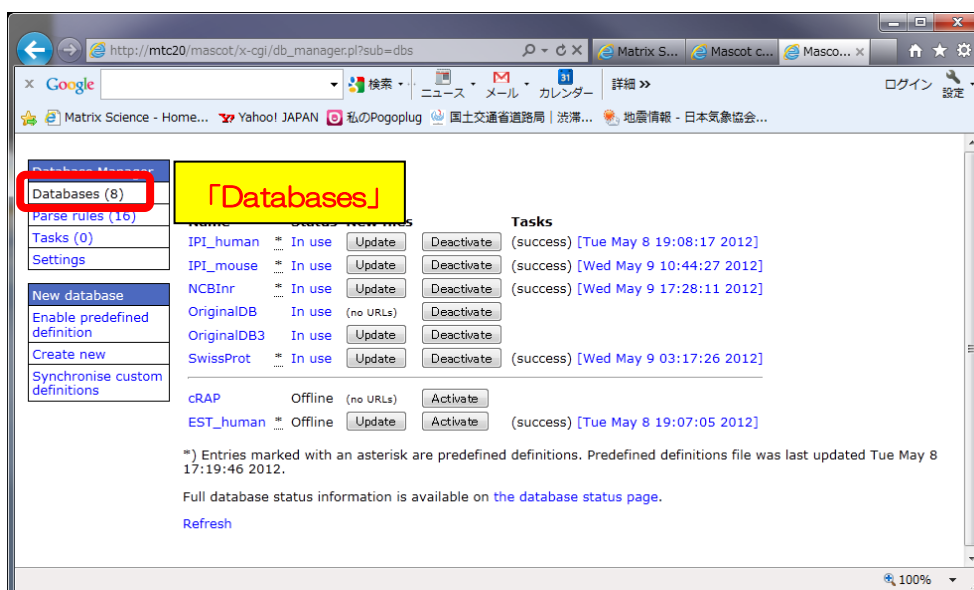
databases セクションでは、以下の内容の設定変更を行うことができます。

- ▼ 使用データベースのON/OFF
- ▼ データベースファイルの取得（ダウンロード開始）
- ▼ データベースファイルの定期的な自動取得・更新設定
- ▼ 新規データベースの追加

以下、Databases Manager Sectionで行う事のできる設定についてまとめています。

### □ 使用データベースのON/OFF、データベースファイルの取得（ダウンロード開始）【Databases section】

Database Manager の最初の画面、または左フレームの一番上、「Database Manager」→「**Databases**」をクリックすると現れる画面で、データベースについてその各種状態が一覧して確認できます。



Databases セクションでは、使用するデータベースの選択やデータベースファイルの更新（ダウンロードの開始）を行う事ができます（下図）。

**Databases**

Name	Status	New files	Tasks
IPI_human	* In use	Update	Deactivate (success) [Tue May 8 19:08:17 2012]
IPI_mouse	* In use	Update	Deactivate (success) [Wed May 9 10:44:27 2012]
NCBINr	* In use	Update	Deactivate (success) [Wed May 9 17:28:11 2012]
OriginalDB	In use	(no URLs)	Deactivate
OriginalDB3	In use	Update	Deactivate
SwissProt	* In use	Update	Deactivate (success) [Wed May 9 10:44:27 2012]
cRAP	Offline	(no URLs)	Activate
EST_human	* Offline	Update	Activate (success) [Tue May 8 19:07:05 2012]

\*) Entries marked with an asterisk are prepared by MASCOT. Full database status information is available in the database file.

Refresh

各データベースの名称部分（ハイパーリンク）をクリックすると、該当データベースの設定やバージョン、ファイル取得の状況についてより詳しく見る事ができます（次頁図）。

Mascot Database Manager: SwissProt - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(I) ヘルプ(H)

Matrix Science - Home x Mascot configuration x Mascot Database Ma... x

mtc20/mascot/x-cgi/db\_manager.pl?sub=d

よく見るページ Firefox を使ってみよう 最新ニュース

mtc20/mascot/x-cgi/db\_ma...r.pl?sub=dbs%3ASwissProt

## Database: SwissProt

**Database Manager**

- Databases (8)
- Parse rules (16)
- Tasks (0)
- Settings

**SwissProt**

- Copy
- Delete

**Name**  
SwissProt

設定ファイル configuration.xml 内での名称

**Using predefined definition**  
SwissProt\_ID

Configuration last updated Tue May 8 17:19:46 2012.

**Database type**  
Amino acid (protein database)

**Database directory**  
C:/inetpub/mascot/sequence/SwissProt/current

**Filename pattern**  
SwissProt\_\*.fasta

Database FASTA ファイルが置かれている場所

**Matching files in directory**

- SwissProt\_2012\_03.fasta (243 MB) (current)
- SwissProt\_2012\_03.dat (2.4 GB) (current)

**Database status**

In use

Deactivate

Activate/Deactivate 切り替え

**Scheduled updates**  
(no schedules defined)

Edit schedule

自動更新スケジュール編集

**Most recent finished task**  
[Wed May 9 03:17:26 2012] (success) No new FASTA file available; database SwissProt is up to date.

Update database now

データベースファイル更新

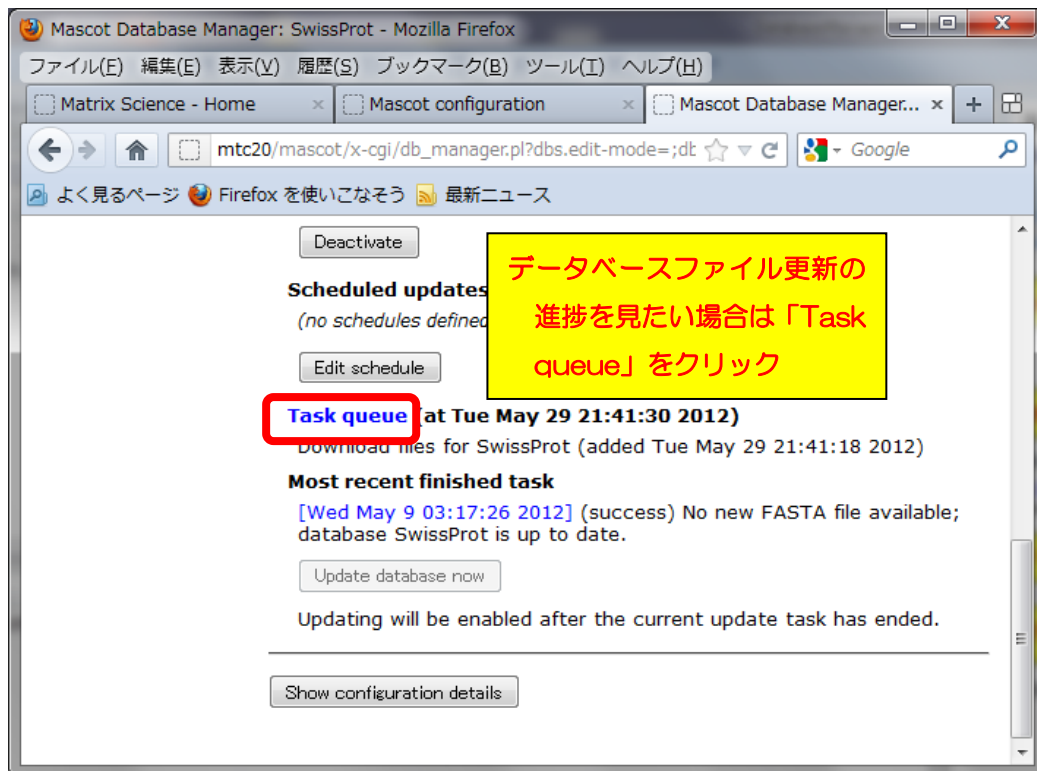
Show configuration details

さらに詳細な設定内容が見たい場合

Databases 個別データベース設定画面

各ページにある「**Update**」または「**Update database now**」 ボタンを押すと、データベースファイルの更新がその場でスタートします。スタートすると「Update」のボタンが押せなくなり、Taskとして進捗度合いが画面に表示されます。

- \* ネットワークにて Proxy サーバーをご利用されている場合、Database Manager ページで Proxy サーバーの設定が必要です。設定方法について詳しくは Section 別設定内容:: Settings Section の「プロキシ設定」(P.10) をご覧ください。



## □ データベースファイルの定期的な自動取得・更新設定【**databases section**】

前述の、各データベースの設定確認画面にある「**Edit Schedule**」を押すと、該当データベースの取得を定期的に行うための設定画面が現れます。

設定は、毎日<Daily>、毎週<Weekly>、毎月<Monthly>の3種類があります（次頁図）。

① 「databases」

Databases

Name	Status	New files	Tasks
IPI_human	In use	Update Deactivate	(success) [Tue May 8 19:08:17 2012]
IPI_mouse	In use	Update Deactivate	(success) [Wed May 9 10:44:27 2012]
NCBINr	In use	Update Deactivate	(success) [Wed May 9 17:28:11 2012]
OriginalDB	In use	Update Deactivate	(success) [Wed May 9 03:17:26 2012]
SwissProt	In use	Update Deactivate	(success) [Wed May 9 03:17:26 2012]

② database 選択

Deactivate

Scheduled updates  
(no schedules defined)

Edit schedule

③ 「Edit schedule」

Task queue (at Tue May 29 21:41:30 2012)  
Download files for SwissProt (added Tue May 29 21:41:18 2012)

Most recent finished task  
[Wed May 9 03:17:26 2012] (success) No new FASTA file available; database SwissProt is up to date.

Database configuration: SwissProt

Scheduled updates

None

Daily at HH : MM

Weekly on at HH : MM

Monthly on the <nth> at HH : MM

Cancel Save

④ 定期更新のスケジュール設定後「Save」

Databases 個別データベース Edit Schedule 画面



## □ 新規データベースの追加【databases section】

MASCOTの検索対象となるデータベースを追加するには、以下の4つの方法があります。

### 【新規データベースを追加する4つの方法と主な対象データベース】

1. すべての設定がひとまとめになった組み合わせ【predefined database definition】をそのまま使用（ファイル設置場所を除き各種設定の変更ができない。データベースの名称も準備されたものを使用する。）  
→ \* P. 18~ SwissProt, NCBIInr, IPI の他、NCBI のEST, EMBL のEST や IPI, trembl など
2. 上記項目1の「predefined definition」を使用するが、さらに名前を変更して利用【template 利用】  
→ \* P. 22~ ファイル公開サイト(NCBI,EBI など)で採用している共通フォーマットのファイルを利用する場合。または1でリストアップされたデータベースと種類は同じだが、日付を固定したり特定生物種のみ抜き出した別名のデータベースを作成したりする場合
3. 既にお手元のMASCOTで使用しているデータベース設定をベースに名前を変更して利用【copy 利用】  
→ \* P. 29~ 出発点が predefined のリストからでなく既に使用中のデータベースから選択。コピー元のデータベースで利用されている fasta ファイルをコピーして利用するかどうかも選べます。既存データベースの異なるバージョンを複数登録したり、単純な抽出ルールのオリジナルデータベースを作成したりする場合の利用等を想定
4. 完全に自分で各設定を記入・選択【Custom】  
→ \* P. 33~ 1~3以外、自分で作成した FASTA ファイルなどの利用を想定

---

## 【1.すべての設定がひとまとめになった組み合わせ【predefined database definition】を使用】

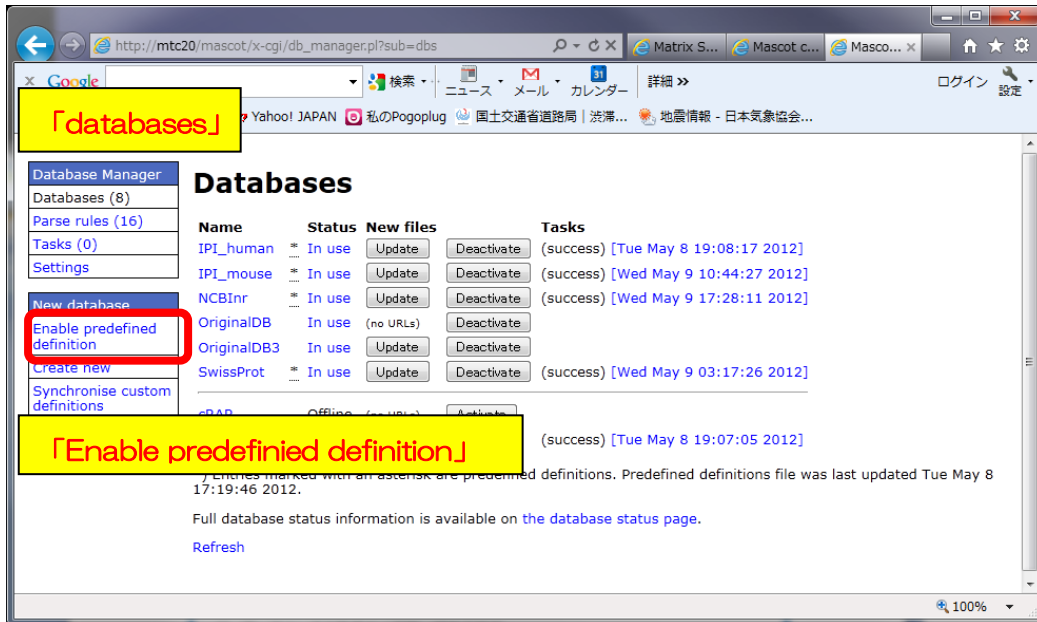
MASCOT 側で既に設定を準備しているいくつかのデータベースについて、名称も含めそのままその設定を利用する際には Predefined database definition を利用します。predefined definition が準備されているデータベースは以下の通りです。

---

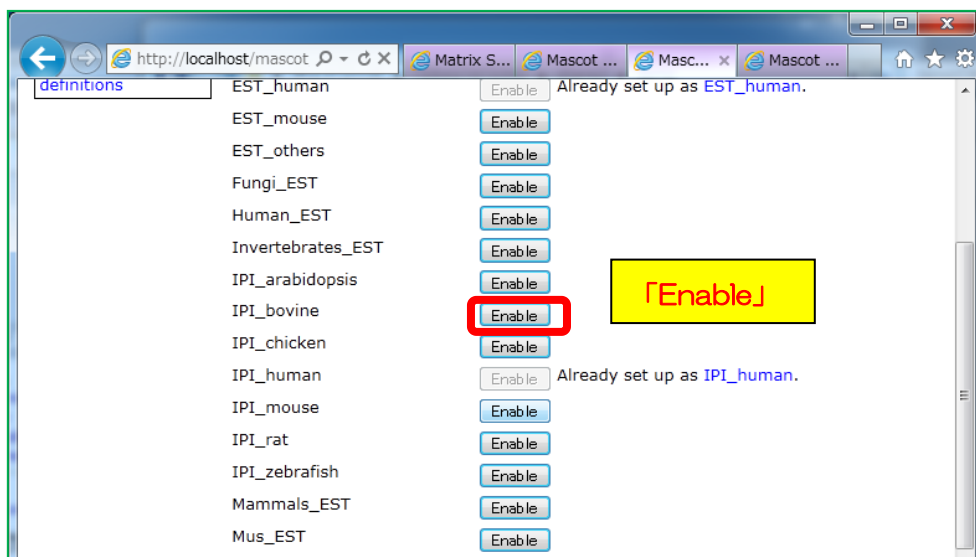
### 【Predefined database definition で定義されているデータベース一覧】

- **contaminants / cRAP** … コンタミ、または生体内に頻出するタンパク質。
- **XXXX\_EST** EMBL にあるEST データベース。詳細は [http://www.matrixscience.com/help/seq\\_db\\_setup\\_EMBLEST.html](http://www.matrixscience.com/help/seq_db_setup_EMBLEST.html) をご覧ください。  
<リスト> Environmental\_EST, Fungi\_EST, Invertebrates\_EST, Mammals\_EST, Mus\_EST, Plants\_EST, Prokaryotes\_EST, Rodents\_EST, Unclassified\_EST, Vertebrates\_EST
- **EST\_YYYY** NCBI にあるEST データベース。詳細は [http://www.matrixscience.com/help/seq\\_db\\_setup\\_EST.html](http://www.matrixscience.com/help/seq_db_setup_EST.html) をご覧ください。  
<リスト> EST\_human, EST\_mouse, EST\_others
- **IPI\_ZZZZ** IPIデータベース。(2011年で更新が止まっています) <リスト> IPI\_arabidopsis, IPI\_bovine, IPI\_chicken, IPI\_human, IPI\_mouse, IPI\_rat, IPI\_zebrafish
- **NCBInr**
- **Trembl**
- **SwissProt**
- **UniRef100**

「Predefined database definition」を使ってデータベースを追加する設定画面へアクセスするには、左フレームの「New database」 -> 「**Enable predefined definition**」をクリックします（下図）。



使用したいデータベースを選んで、「**Enable**」ボタンを押します。

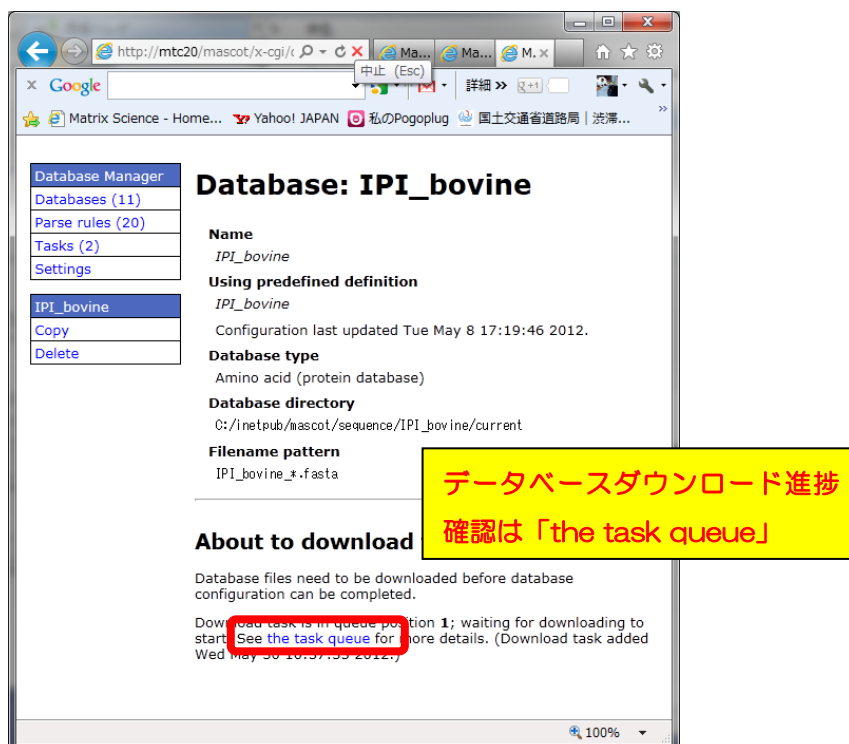


Enable predefined definition 画面

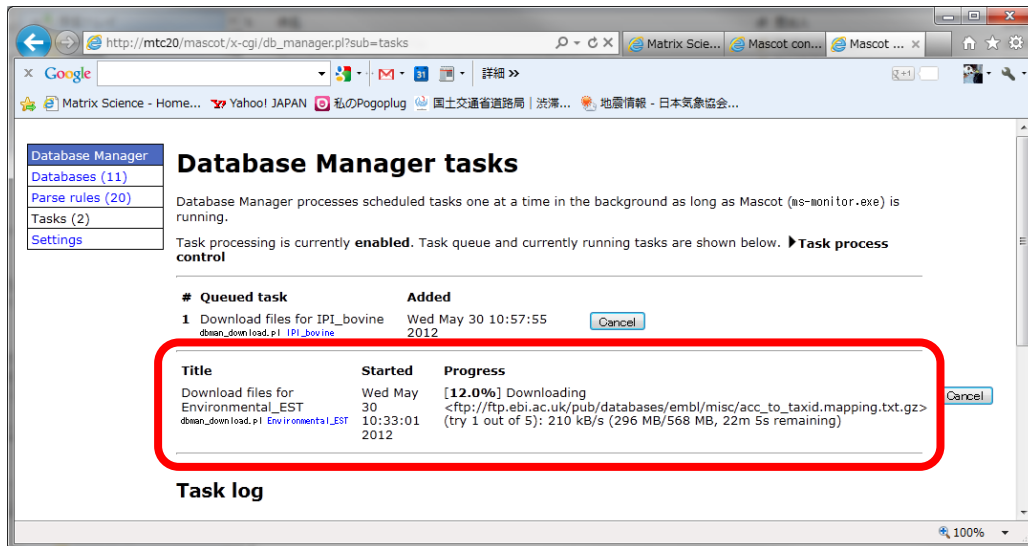
配列データベースファイルの置き場所を「Sequence directory」で確認し必要があれば変更します。確認後、「Create」ボタンを押します。



データベースの定義に関するSummaryが表示されます。この段階で既に該当データベースがMASCOTにて使用対象のデータベースとして認識されていますが、使用可能となるにはこれと並行して行われているデータベースファイルの取得、再構築の完了が必要です。データベースファイル取得状況を確認したい場合は、ハイパーリンクの「**the task queue**」をクリックします。



データベースファイルダウンロードの進捗状況が表示されます。



Database Status に表示され、Status が **“In Use”**となれば終了です(下図例は IPI\_mouse)。



Database Status 画面

## 【2. predefined definitionを利用するが、名前のみ変更して利用【template利用】】

template 利用は、同じ抜き出しルールなどのフォーマットとしてMASCOTが準備しているセット「predefined definition」を「template」として使うものの、名前のみ変更して利用する方法です。ファイル公開サイトで採用している共通フォーマット（ID（Accession）記述のルールが同一）のデータベースを利用する場合や、predefinedのバージョン違いをセットしたい場合に利用します。例えば2011年で更新が止まったIPIデータベースの代わりに、IPIに該当する情報も含まれるuniprot データベースから特定生物種のみを抜き出したデータベースを作成して代わりに使いたい方などがtemplateを利用します。

以下、template を利用したデータベースの登録方法を記しますがその前に関連事項として uniprot から生物種を抜き出して利用する方法を例に設定をご紹介します【[緑線箇所](#)】。

### ◇ uniprotからtaxonomy;humanを抜き出してFASTAファイルにする方法

データベース設定から少し話がそれますが、ユーザーからのニーズが高い、uniprotから特定生物種のエントリーだけを抜き出してFASTAファイルにする方法も併せてご紹介します。

1. 抜き出したい生物種のTax ID を確認します。Tax ID はMASCOTの生物種設定でも採用しているNCBIのtaxonomy ID です。調べたい場合は以下のサイトでキーワード検索をして番号をお確かめください。

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

Tax ID 例 ) Mus musculus : 10090, rattus norvegicus : 10116 など。

キーワード検索  
例1) rattus norvegicus  
例2) human

Entrez records		
Database name	Subtree links	Direct links
Nucleotide	270,518	270,517
Nucleotide EST	1,103,577	1,103,577
Nucleotide GSS	867,131	867,131
Protein	110,744	110,744
Structure	1,736	1,736
Genome	44	44

NCBI, Taonomy Browser <http://www.ncbi.nlm.nih.gov/taxonomy>

2. uniprot のサイト [www.uniprot.org](http://www.uniprot.org)へアクセス
3. キーワード **organism:TAX ID** と **keyword:"complete proteome"** で検索  
例) organism:9606 keyword:"complete proteome"

キーワード検索  
例) organism:9606 keyword:"complete proteome"

ダウンロードボタン

SwissProt データ数

Entry	Entry name	Protein names	Gene names	Organism	Length
B7Z1M9	C2D4D_HUMAN	C2 calcium-dependent domain-contain...	C2CD4D, FAM148D	Homo sapiens (Human)	353
		Uncharacterized protein C1orf53	C1orf53	Homo sapiens (Human)	145
		Putative uncharacterized protein ZN...	ZNF436-AS1, C1orf213	Homo sapiens (Human)	126
Q8N4J0	CI041_HUMAN	UPF0586 protein C9orf41	C9orf41	Homo sapiens (Human)	409
Q9H8K7	CJ088_HUMAN	Uncharacterized protein C10orf88	C10orf88	Homo sapiens (Human)	445
Q9Y471	CMAH_HUMAN	Putative cytidine monophosphate...	CMAHP, CMAH	Homo sapiens (Human)	72

検索後、FASTAファイルを手にするには、画面上部の「Download」をクリックし、formatとして「FASTA」を選択して「Go」ボタンを押します。

Download selected (0)  
 Download all (68511)

Format: FASTA (canonical)  
 FASTA (canonical & isoform)

Preview first 10

Go

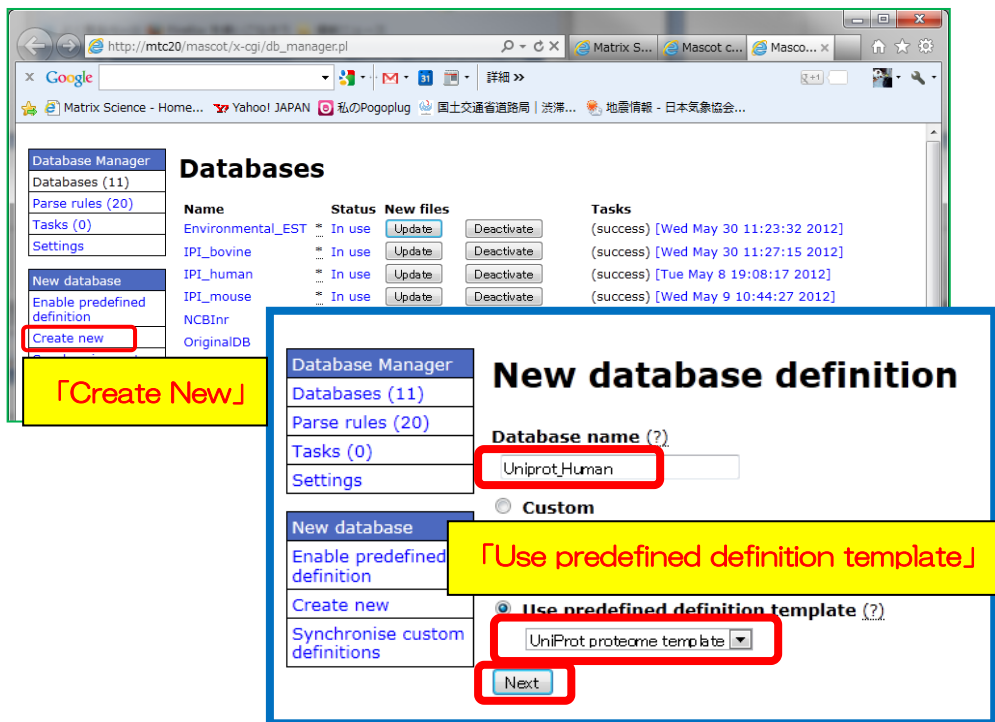
\* 2015年4月15日時点で68,511件のエントリーがヒットします。うちアノテーションがついているSwissProtのエントリーが 20,203件です。

取得したファイルはMASCOT の検索対象データベースファイルとして利用可能です。

<uniprotからtaxonomy:humanのデータを抜き出したFASTAファイル取得方法:おわり>

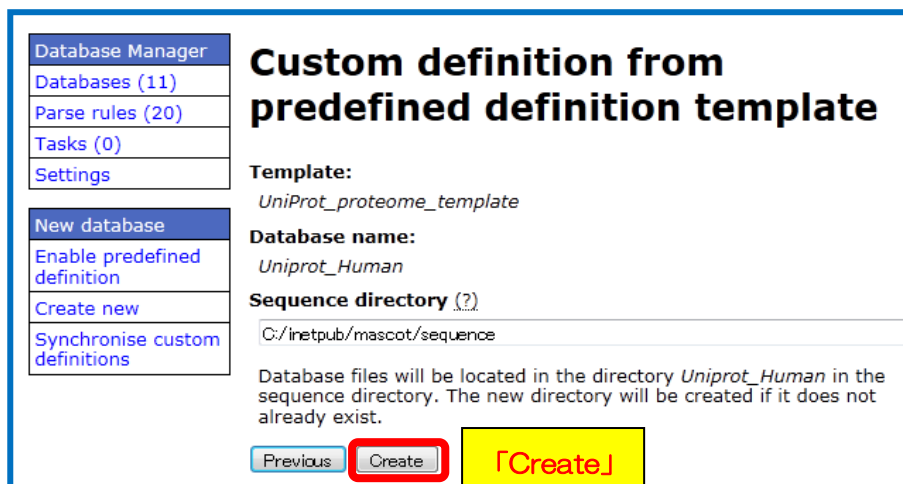
### ◇ Template 利用新規データベース作成例：uniprot\_human データベース作成

Database manager の、左フレーム「New database」→「**Create new**」を選びます。続いて現れる画面で、「Database name」に登録するデータベースの名称を入力し、「**Use predefined definition template**」で利用したいtemplateを選びます。例では「Uniprot proteome template」を選びます。選択後、「Next」を押します。



Databases -> Create New 画面

データベースファイルの置き場所を決め、「Create」を押します。





この時点ではFASTAファイルが所定の場所に置かれておらず、Inactive なデータベースとしてMASCOT側に登録されています。使用するためには別途取得したファイルを指定のファイル名に変えてから指定のフォルダに設置した上でActivateする必要があります。

**Database: Uniprot\_Human**

**Name**  
Uniprot\_Human

**Database type**  
Amino acid (protein database)

**Database directory**  
C:/inetpub/mascot/sequence/Uniprot\_Human/current

**Filename pattern**  
Uniprot\_Human\_\*.fasta

**No local FASTA file**

Database files must be present before the database can be configured and activated. You have the following options:

- Download from remote URL or...
- Upload file using web browser
- Copy file manually

**Next**

**FASTA ファイル名称**

**FASTA ファイル設置場所**

**FASTA ファイルが所定の場所がない事を表す**

**ファイルの設置方法**

- ダウンロードの URL 設定を作成し直しすぐにダウンロード
- web ブラウザからファイルをアップロード
- マニュアル操作で設置場所にファイルコピー

ファイルを設置する方法にはWEBブラウザ経由でファイルを置く「Upload file using web browser」と、フォルダにファイルを直接設置する「Copy file manually」があります。

### 【ファイルの設置方法で「copy file manually」を選択した場合】

「Database directory」で指定されているフォルダ、つまり

C:\inetpub\mascot\sequence\Uniprot\_Human\current

に、ダウンロードしたファイルを設置し、名称を「Filename pattern」で指定された形に書き換えます。但し名称中の「\*」部分は任意の文字列、例えばダウンロード日などに変更します。

Database Manager

Databases (18)

Parse rules (18)

Scheduled updates (0)

Running tasks (0)

Settings

New database

Enable predefined definition

Create new

Synchronise custom definitions

Uniprot\_Human

Copy

Delete

## Database: Uniprot\_Human

**Name**  
Uniprot\_Human

**Database type**  
Amino acid (protein database)

**Database directory**  
C:/inetpub/mascot/sequence/Uniprot\_Human/current

**Filename pattern**  
Uniprot\_Human\_\*.fasta

### Copy files manually

**Filename pattern**  
Uniprot\_Human\_\*.fasta

Edit filename pattern

Please,

- copy the FASTA file to C:/inetpub/mascot/sequence/Uniprot\_Human/current;
- rename the file to match the filename pattern, e.g. Uniprot\_Human\_20150512.fasta (in the pattern, \* matches any string of characters);
- and refresh this page to continue.

Previous Refresh

① 以下で指定されたフォルダに

② 指定された名称・ルールに基づきファイル名を変更した FASTA ファイルをコピーしたら

③ Refresh ボタンを押す

ファイル名変更例) Uniprot\_Human\_20120608.fasta (\*はファイル名に書かない事)

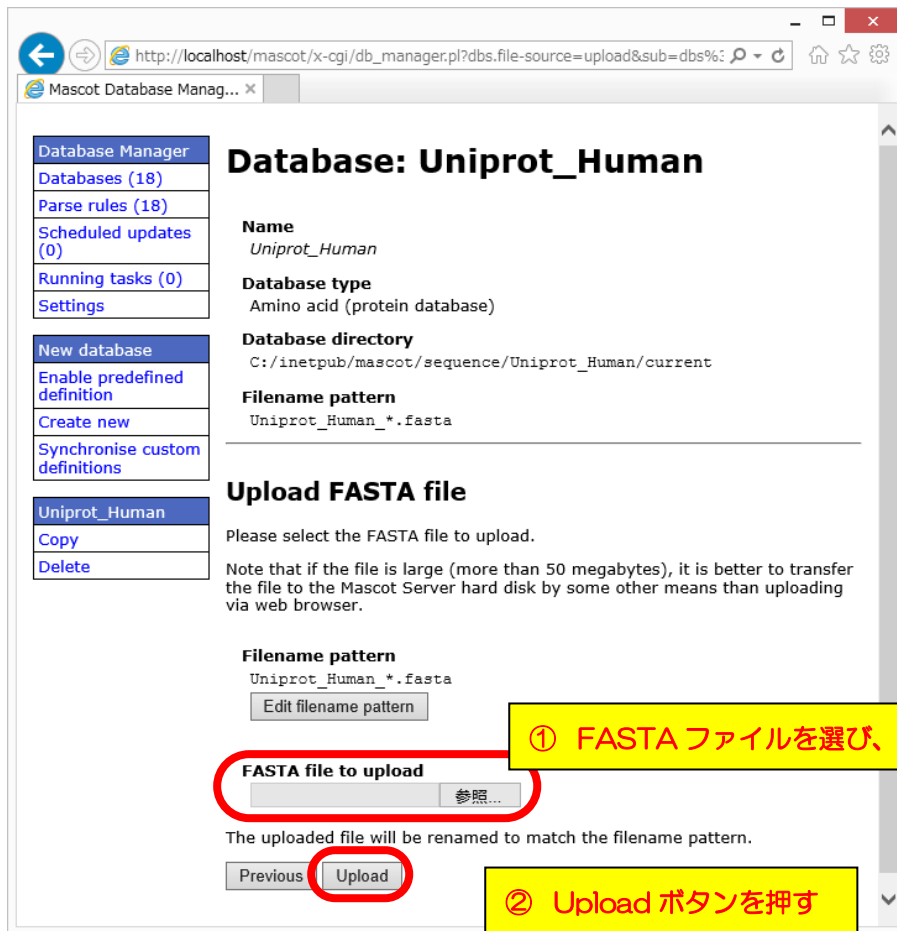
Uniprot\_human\_20120530.fasta  
FASTA ファイル  
43.7 MB

uniprot-organism%3A9606+complete+proteome\_iso.fasta  
FASTA ファイル

ファイル名を指示通り変更し、所定の場所へファイルをコピー。

### 【ファイルの設置方法で「Upload file using web browser」を選択した場合】

「Upload file using web browser」を選択した場合、次の画面（下図）でFASTAファイルを選択する画面が現れます。「FASTA file to upload」で使用するFASTAファイルを選び、「Upload」ボタンを押せばOKです。



### 【ファイル設置後】

ファイルを所定の方法でフォルダに設置すると、FASTAファイルが自動的に認識され、「No local FASTA file」の表記がなくなります。「**Activate**」ボタンを押せば、データベース構築が開始されます（次頁図）。データベースの構築状況は「Database Status」画面でご確認ください（MASCOT のHome画面→Database Status リンク）。

The image shows two screenshots of the Mascot database manager interface. The top screenshot displays the configuration page for the 'Uniprot\_Human' database. The 'Database status' section shows the database is 'Offline' and an 'Activate' button is highlighted with a red circle. A yellow callout box points to this button with the text: ファイル名が認識されたら「Activate」.

The bottom screenshot shows the 'Database Status' page. It lists the status of several databases. The 'Uniprot\_Human' database is highlighted with a red box, and a yellow callout box points to it with the text: Database Status 画面でデータベース構築確認.

**Database: Uniprot\_Human**

**Database Manager**

- Databases (12)
- Parse rules (21)
- Tasks (0)
- Settings

**Uniprot\_Human**

- Copy
- Delete

**Name**  
Uniprot\_Human

**Database type**  
Amino acid (protein database)

**Database directory**  
C:/inetpub/mascot/sequence/Uniprot\_Human/current

**Filename pattern**  
Uniprot\_Human\_\*.fasta

**Matching files in directory**  
Uniprot\_Human\_20120530.fasta (43.72 MB)

**Database status**  
Offline

**Activate**

**Scheduled updates**  
(no schedules defined)

**Database Status**

Filename = Environmental\_ES1\_111.fasta Pathname = C:/inetpub/mascot/sequence/Environmental\_ES1/Current  
Status = In use [Statistics](#) [Unidentified taxonomy](#)  
State Time = Wed May 30 11:31:57 # searches = 0  
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO  
Number of threads = -1 Current = YES

Name = IPI\_bovine Pathname = C:/inetpub/mascot/sequence/IPI\_bovine/Current  
Filename = IPI\_bovine\_3.73.fasta Pathname = C:/inetpub/mascot/sequence/IPI\_bovine/Current  
Status = In use [Statistics](#)  
State Time = Wed May 30 11:32:16 # searches = 0  
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO  
Number of threads = -1 Current = YES

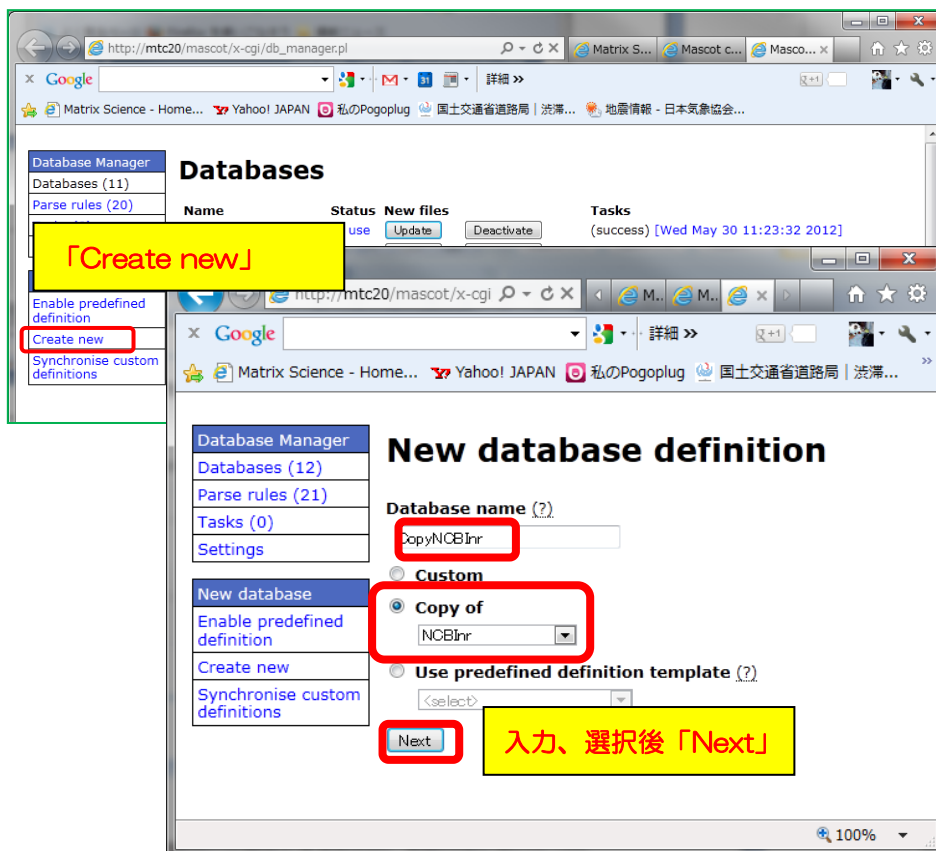
Name = Uniprot\_Human Pathname = C:/inetpub/mascot/sequence/Uniprot\_Human/current/Uniprot\_Human\_20120530.fasta  
Filename = Uniprot\_Human\_20120530.fasta Pathname = C:/inetpub/mascot/sequence/Uniprot\_Human/current/Uniprot\_Human\_20120530.fasta  
Status = Creating compressed files 39% complete  
State Time = Wed May 30 14:15:40 # searches = 0  
Mem mapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO  
Number of threads = -1 Current = NO

### 【3. 既にお手元の MASCOT で使用しているデータベース設定をベースに名前を変更して使用【copy 利用】】

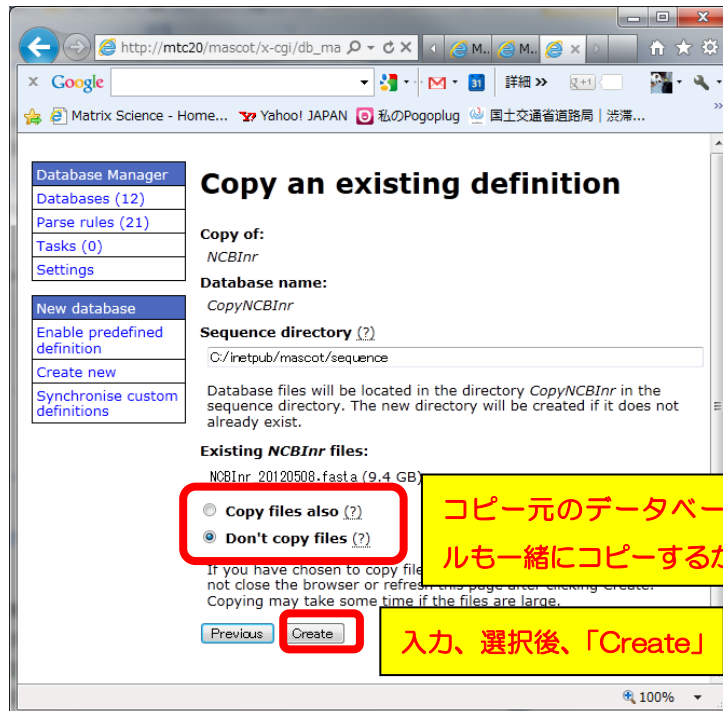
copy 利用は、設定ファイル中のテンプレートからでなく、既に利用しているデータベースの中から設定を選び、名前を変えてデータベース構築を行います。例えば既存データベースの異なるバージョンを複数登録したり、単純な抽出ルールのオリジナルデータベースを作成したりする場合の利用等を想定しています。

copyを利用してデータベースを新規作成するには、Database Manager においてDatabase manager ページの左フレーム「New database」→「**Create new**」を選びます（下図）。例では、NCBIInr と同じフォーマットで古いファイルを持ち続ける事を想定したデータベース「CopyNCBIInr」を作成します。現れた画面で「Database name」に名称を入力し、「**copy of**」で「**NCBIInr**」を選びます。記入・選択後、画面下の「Next」ボタンを押します。

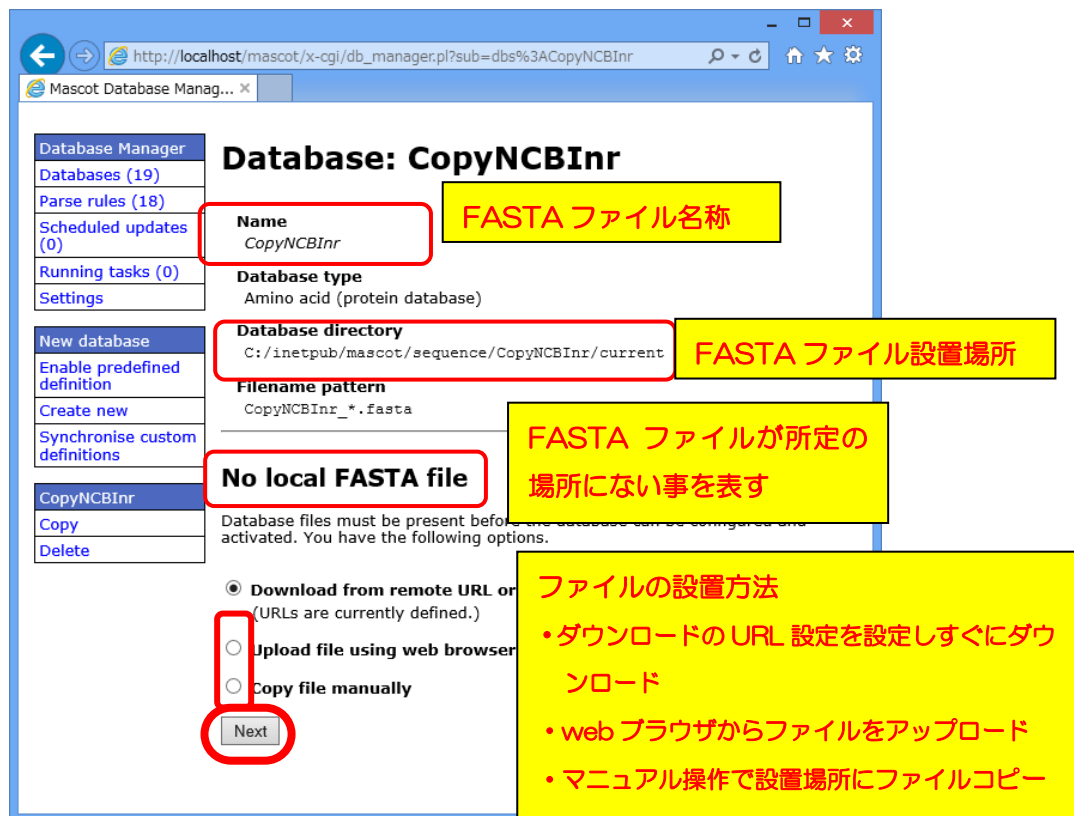
copy利用の場合、ファイルそのものもフォルダにコピーして名前を変えて利用するかユーザーに訪ねてきます。コピーの必要がない場合は、「don't copy files」を選択し、「Create」ボタンを押します。



Databases → Create New 画面



「Don't copy files」を選択した場合、所定の場所にFASTAファイルを自分で設置する必要があります。



**【ファイルの設置方法で「copy file manually」を選択した場合】**

「Database directory」で指定されているフォルダ、つまり

C:\inetpub\mascot\sequence\CopyNCBIInr\current

に、ダウンロードしたファイルを設置し、名称を「Filename pattern」で指定された形に書き換えます。但し名称中の「\*」部分は任意の文字列、例えばダウンロード日などに変更します。

The screenshot shows the Mascot Database Manager interface. The main content area displays the following information for the 'CopyNCBIInr' database:

- Name:** CopyNCBIInr
- Database type:** Amino acid (protein database)
- Database directory:** C:/inetpub/mascot/sequence/CopyNCBIInr/current
- Filename pattern:** CopyNCBIInr\_\*.fasta

The 'Copy files manually' section includes the following instructions:

- copy the FASTA file to C:/inetpub/mascot/sequence/CopyNCBIInr/current;
- rename the file to match the filename pattern, e.g. CopyNCBIInr\_20150512.fasta in the pattern. \* matches any string of characters);
- and refresh this page to contin

Three yellow callout boxes with red text provide additional instructions:

- ① 以下で指定されたフォルダに
- ② 指定された名称・ルールに基づきファイル名を変更したFASTAファイルをコピーしたら
- ③ Refresh ボタンを押す

### 【ファイルの設置方法で「Upload file using web browser」を選択した場合】

「Upload file using web browser」を選択した場合、次の画面（下図）でFASTAファイルを選択する画面が現れます。「FASTA file to upload」で使用するFASTAファイルを選び、「Upload」ボタンを押せばOKです。



### 【ファイル設置後】

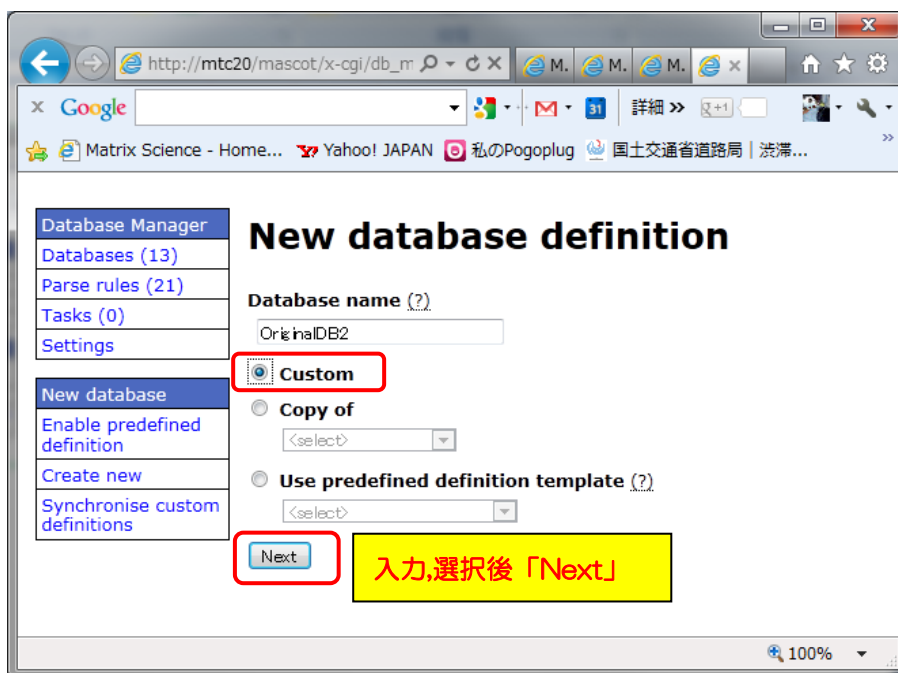
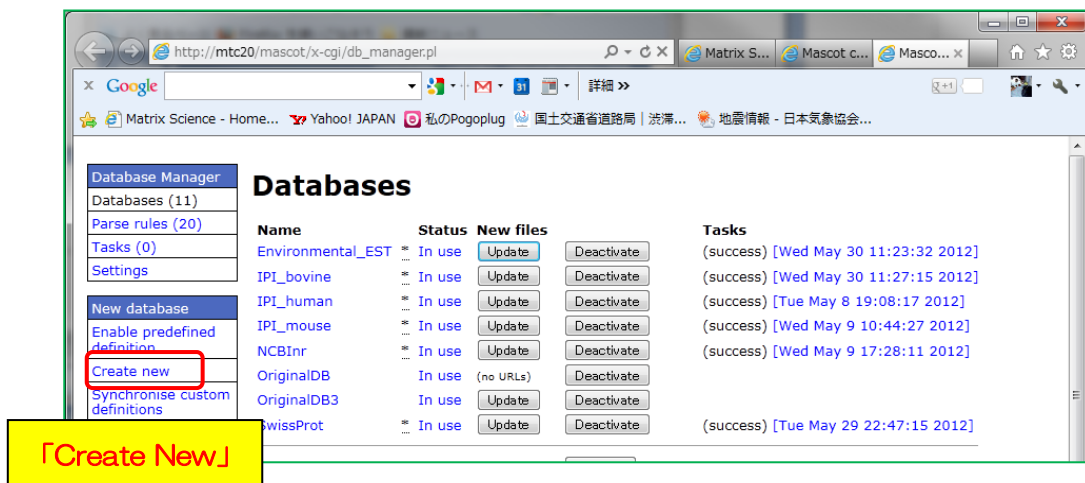
ファイルを所定の方法でフォルダに設置すると、FASTAファイルが自動的に認識され、「No local FASTA file」の表記がなくなります。「**Activate**」ボタンを押せば、データベース構築が開始されます。データベースの構築状況は「Database Status」画面でご確認ください(MASCOTのHome画面→Database Status リンク)。



#### 【4. 完全に自分で各設定を記入・選択 【Custom】】

1～3 で挙げた設定方法は、名前とファイルの設置場所以外、基本的に変更する事ができません。既存のデータベースのルールで当てはまらない抜き出しルールの組み合わせを適用する必要がある場合、本項目「Custom」を利用します。例では自らFASTAファイル「OriginalDB2」を作成しMASCOT用の検索データベースとして使用するケースを使って紹介します。

Database manager の、左フレーム「New database」→「**Create new**」を選びます（下図）。現れた画面で「Database name」にデータベースの名称を入力し、「**Custom**」を選びます。選択後、「Next」を押します。



Databases -> Create New 画面

データベースの設定画面に移行します。データベースの配列の種類のほか、インターネット上にあるfastaファイルを自動で取得する設定にしたい場合、「Download database files」を選び、FASTAファイル、versionファイル、reference ファイルの取得先を記入します。自動取得でなく、手動で更新または更新する意思がない場合、「Upload or Copy files manually」を選びます。各項目の選択後、「Create」ボタンを押します（下図）。

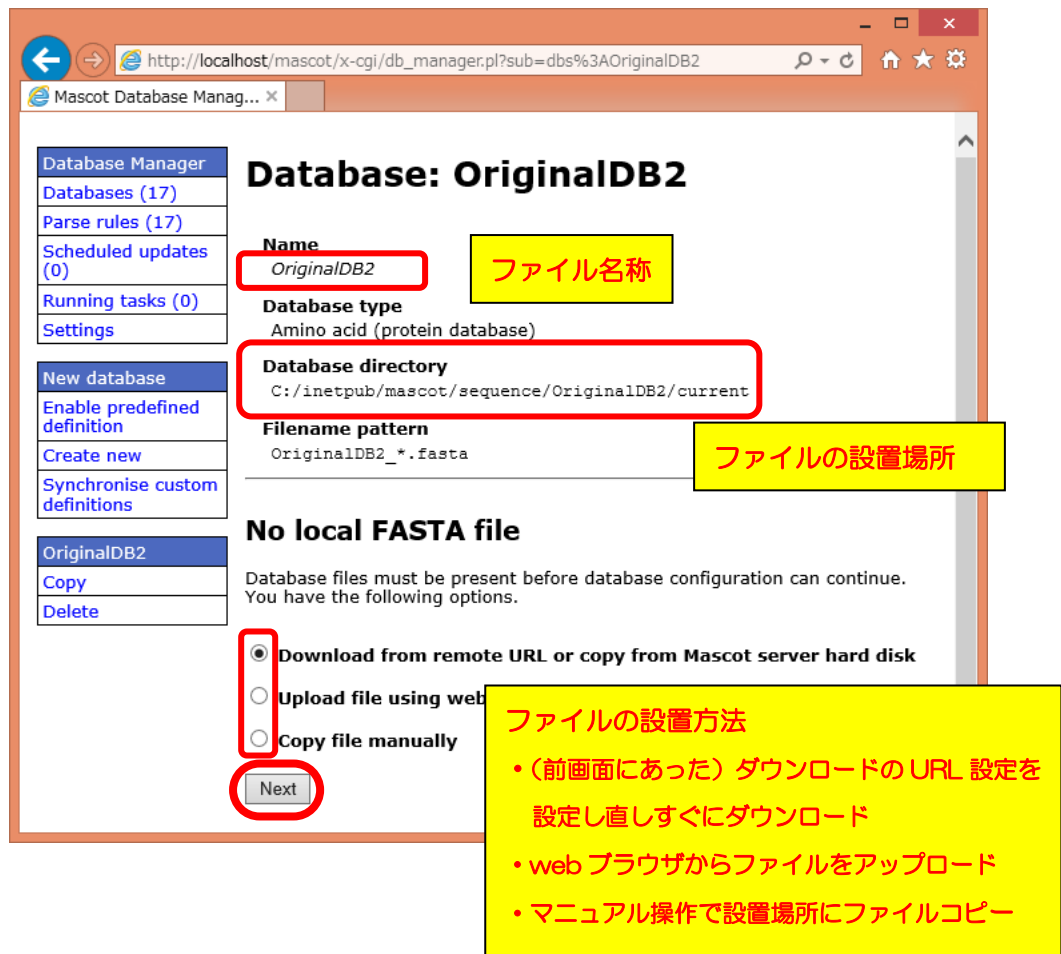
The screenshot shows the 'Custom definition' page in the Mascot Database Manager. The page is titled 'Custom definition' and has a sidebar on the left with navigation links: Database Manager, Databases (16), Parse rules (17), Scheduled updates (0), Running tasks (0), Settings, New database, Enable predefined definition, Create new, and Synchronise custom definitions. The main content area includes the following fields and options:

- Database name:** OriginalDB2
- Database type:** Radio buttons for 'Amino acid (protein database)' (selected) and 'Nucleic acid (DNA database)'. An annotation points to this section: 「Database type」 アミノ酸か核酸か
- Sequence directory (?):** Text input field containing 'C:/inetpub/mascot/sequence'. An annotation points to this field: 「Sequence directory」 配列データベースファイルの置き場所
- Download or copy files automatically (?):** Radio button (selected) for automatic download. Below it are three text input fields for 'FASTA file URL or path to source file on Mascot Server hard disk', 'Version file URL or path to source file on Mascot Server hard disk', and 'Reference file URL or path to source file on Mascot Server hard disk'. Each has a checkbox for 'Delete original file after copying'. An annotation points to the 'Download database files' option: 「Download database files」 自動更新プログラムにおけるファイルの取得先
- Upload or copy files manually (?):** Radio button (unselected). An annotation points to this option: 「Upload or Copy files manually」 上記設定にて自動ダウンロードしない場合に選択

At the bottom of the page, there are two buttons: 'Previous' and 'Create'. The 'Create' button is highlighted with a red box and an annotation: 入力、選択後「Create」

まだファイルを所定の場所に置いていないので、ページ内に「No local FASTA file」と表示されます。

検索対象のFASTAファイルを、「ファイル設置場所」に、ファイル名を「ファイル名称」に合わせた形に変更して設置する必要があります。設置の方法はフォルダに直接コピーする方法とwebブラウザからファイルをアップロードする方法があります（下図）。

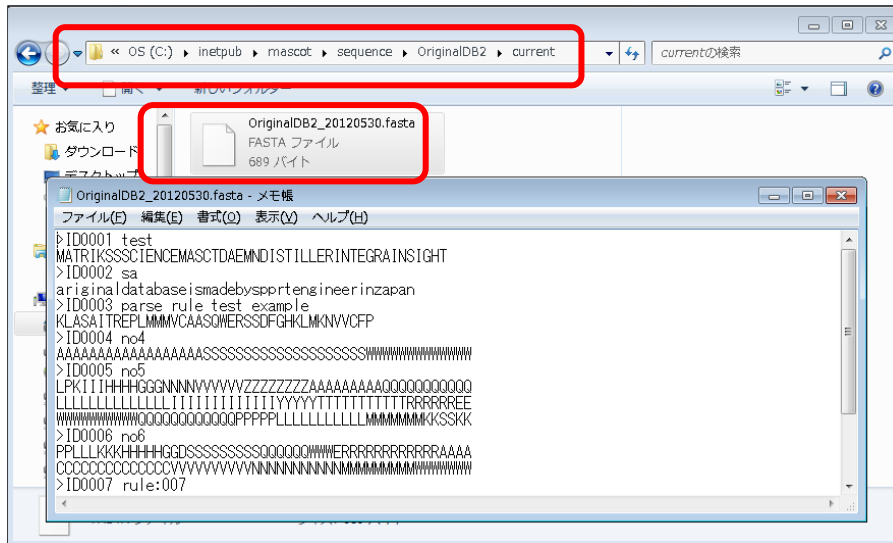


### 【ファイルの設置方法で「copy file manually」を選択した場合】

「Database directory」で指定されているフォルダ、つまり

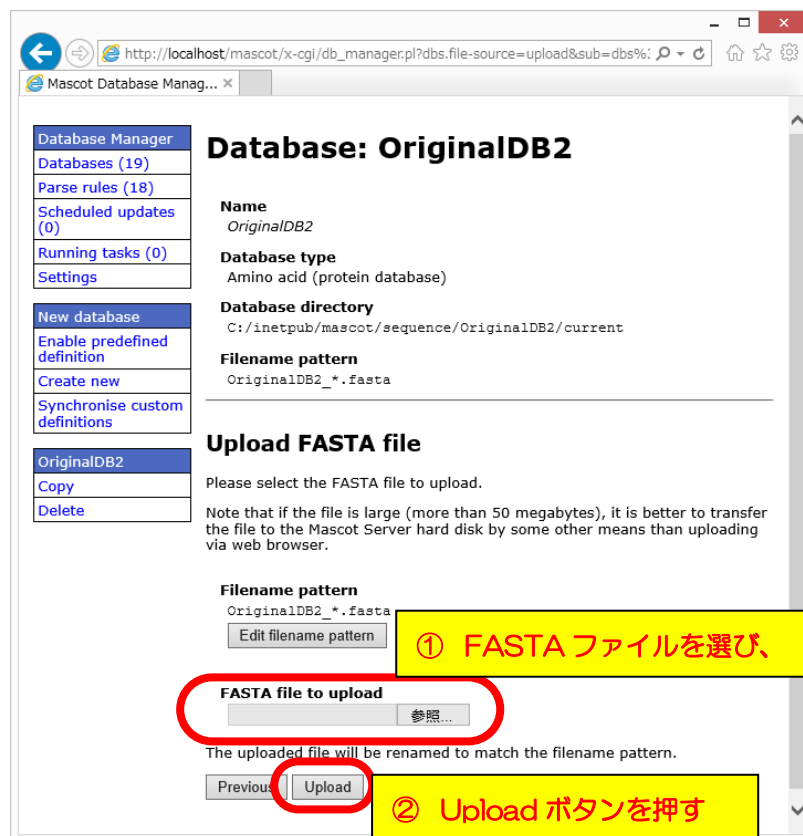
C:\inetpub\mascot\sequence\OriginalDB2\current

に、ダウンロードしたファイルを設置し、名称を「Filename pattern」で指定された形に書き換えます。但し名称中の「\*」部分は任意の文字列、例えばダウンロード日などに変更します。

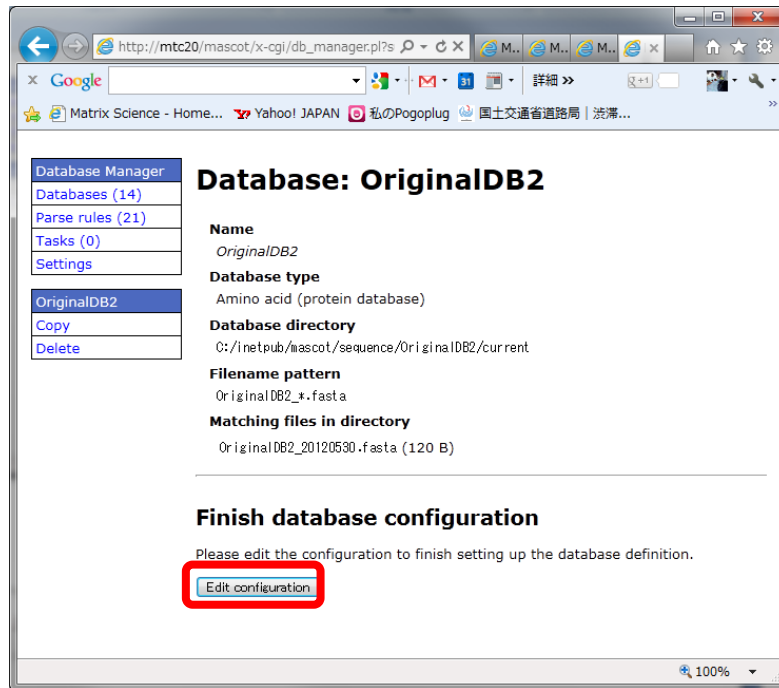


### 【ファイルの設置方法で「Upload file using web browser」を選択した場合】

「Upload file using web browser」を選択した場合、「Next」ボタンを押した次の画面（下図）でFASTAファイルを選択する画面が現れます。「FASTA file to upload」で使用するFASTAファイルを選び、「Upload」ボタンを押せばOKです。



FASTAファイルを正しい名称にして所定の場所に正しく置くと、Database Manager 画面で認識されます。設定をさらに進めるため、「**Edit Configuration**」ボタンを押します。



続いて、IDとDescription の抜き出しルール設定を行います。各々の抜き出しルール設定画面へ移行すると、MASCOT 側で候補となるルールを最大5つ選択しユーザーに提示します（次頁図）。ユーザーはFASTAファイル内の各行の記述と抜き出された内容を見比べた上で、最適な抜き出し方を行っているルールを選択し、「Choose」ボタンを押します。続いて現れる画面で最終的にルールを適用した場合のIDが5例表示されますので、問題がなければ「**Confirm**」ボタンを押します。同様の操作を「Description」でも実行します（次々頁図）

Accession の parse rule 設定

**Database configuration: OriginalDB2 (step 1/2)**

**FASTA file**

Matching files in directory  
OriginalDB2\_20120530.fasta (120 B)

**Accession parse rule (?)**  
(none)

**Choose** ① Accession の「Choose」

**Description parse rule (?)**  
(none)

**FASTA file accession parse rule for OriginalDB2**

**Raw test data**

1. >ID0001 test  
2. >ID0002 test  
3. >ID0003 parse rule test example  
4. >ID0004 no4  
5. >ID0005 no5

② FASTA の記述を確認

**Parse rules (5 suitable)**

Parse rule	Match	Extracted data
<input type="radio"/> >[^ ]* ¥(.*)	5/5	1. test ▶ 4 more matches
<input type="radio"/> >¥(.*)	5/5	1. ID0001 test ▶ 4 more matches
<input checked="" type="radio"/> >¥([^\ ]*)	5/5	1. ID0001 ▶ 4 more matches
<input type="radio"/> >¥([^\ ]*)	5/5	1. ID0001 test ▶ 4 more matches
<input type="radio"/> >¥([^\ ]*)	5/5	1. >ID0001 test ▶ 4 more matches

③ 左) ルール  
右) 抜き出し例  
最適なものを選ぶ

**FASTA file accession parse rule for OriginalDB2**

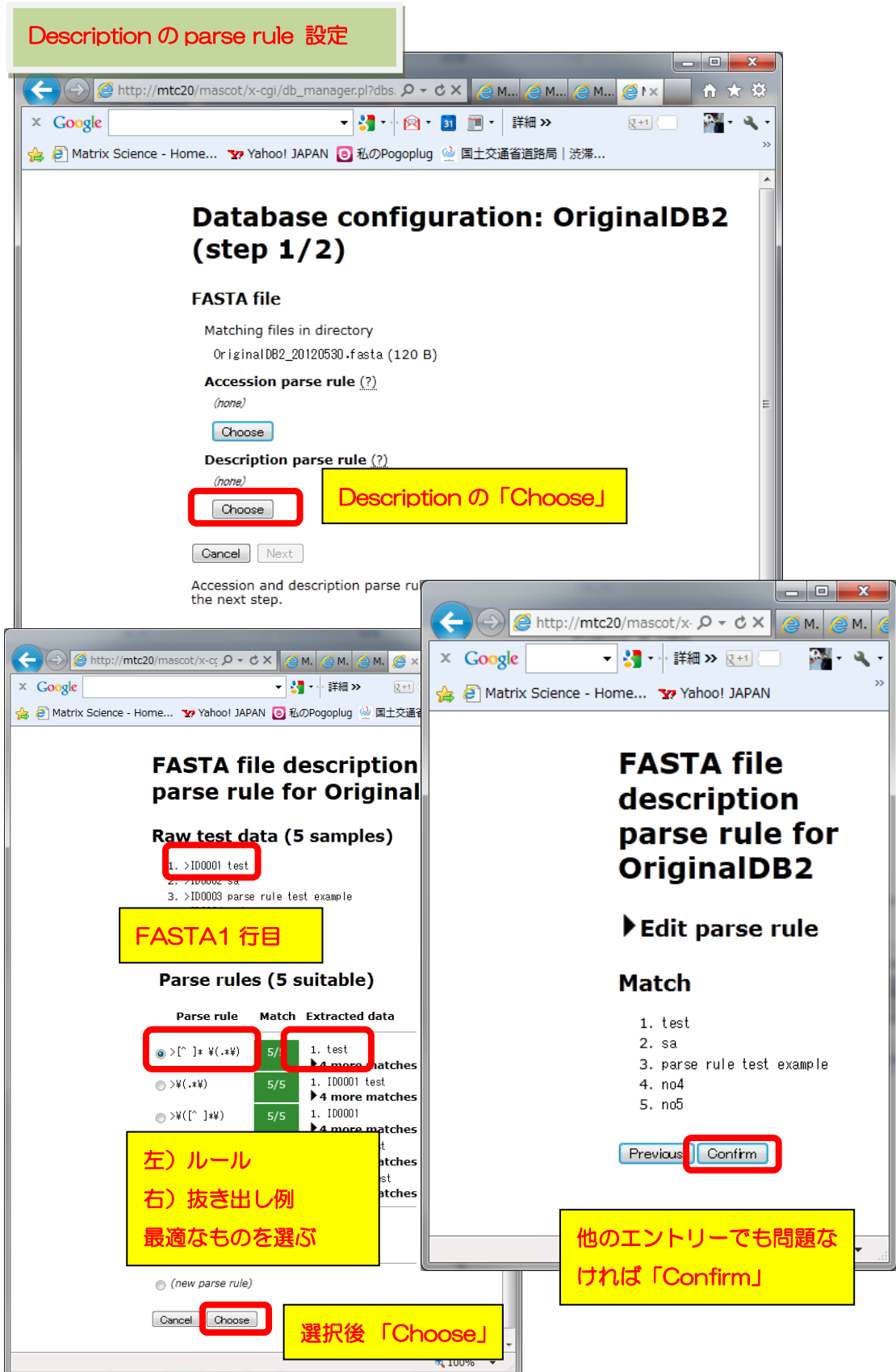
▶ Edit parse rule

**Match**

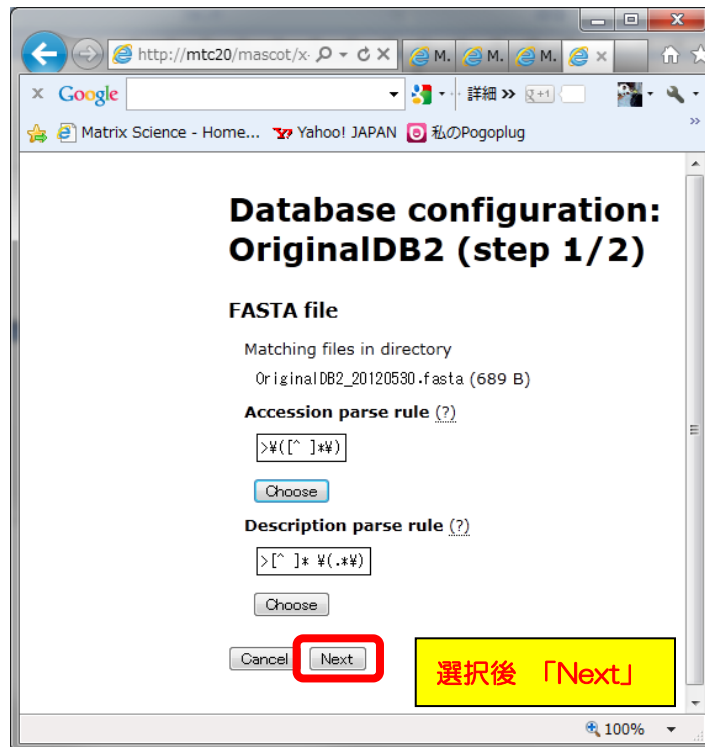
- ID0001
- ID0002
- ID0003
- ID0004
- ID0005

**Choose** ④ 「Choose」

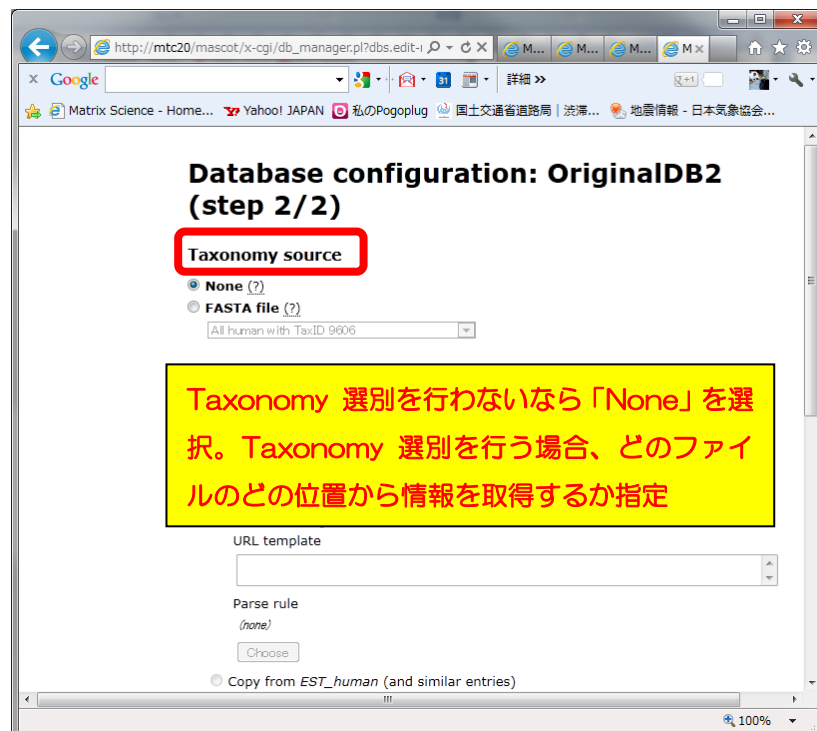
**Confirm** ⑤ 抜出内容を確認して「Confirm」



抜き出しルールを選んだら、「Next」ボタンを押します。

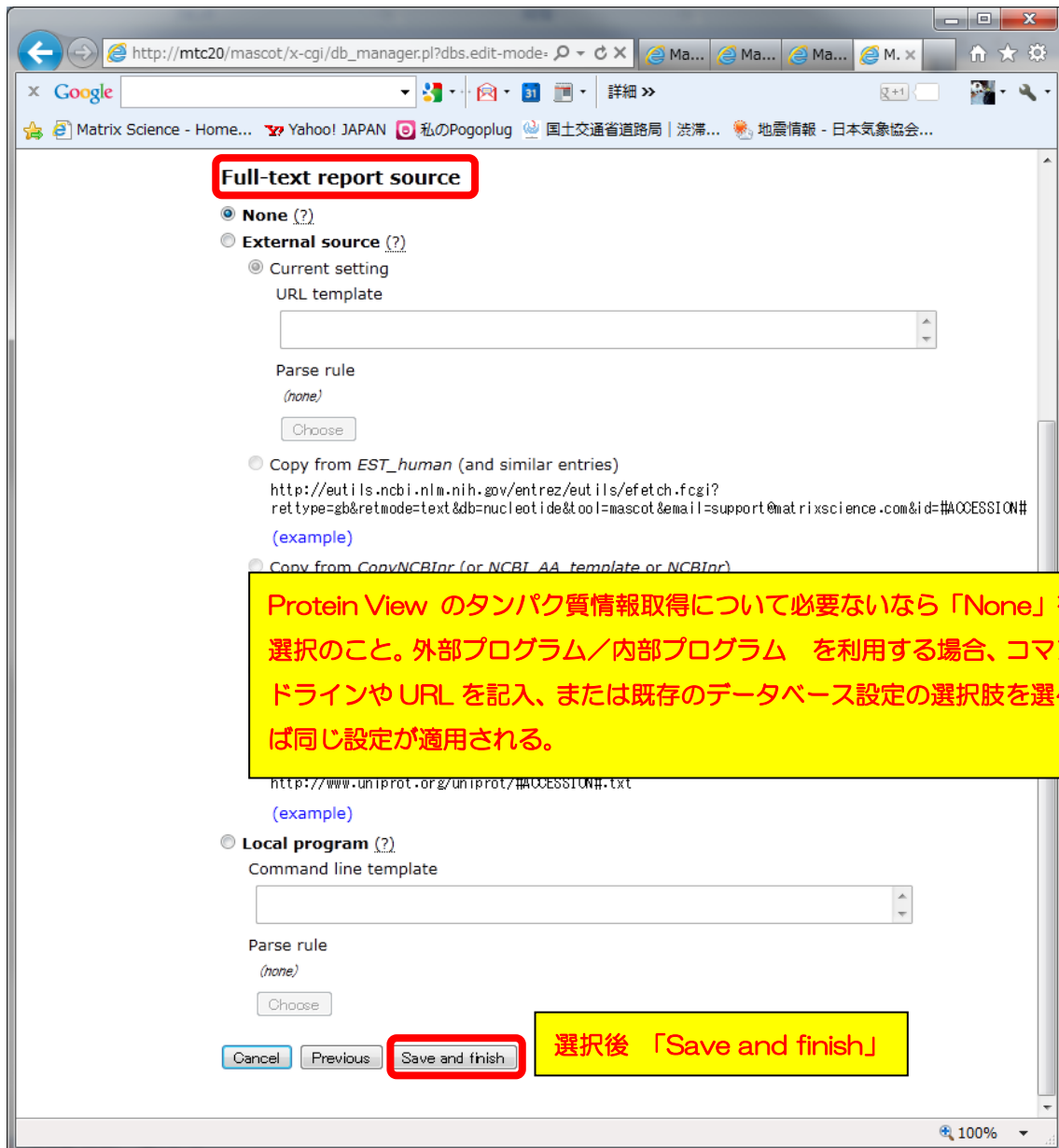


今度はTaxonomyとタンパク質の詳細情報取得先に関する設定画面となります。画面上部ではTaxonomy選別用のルールを選択します（下図）。





また、結果画面のProtein View 画面で表示されるタンパク質の詳細情報に関する設定も行うことができます。すべて選択を終えた後、画面最下部の「Save and finish」ボタンを押します。



個別のデータベース設定画面に戻ります。Save直後では Active状態になっていないので、「Activate」ボタンを押せば、データベースの構築が開始され、検索に利用する事ができるようになります。

The image shows two screenshots of the Mascot database manager interface. The top screenshot displays the configuration page for 'Database: OriginalDB2'. The 'Database status' section shows 'Offline' and an 'Activate' button, which is highlighted with a red box. A yellow callout box next to it says '「Activate」ボタンで使用開始'. Below this, the 'Scheduled updates' section shows '(no schedules defined)' and an 'Update database now' button. The bottom screenshot shows the 'ms-stat' status page. It lists the status of various databases. The 'OriginalDB2' entry is highlighted with a red box and has a yellow callout box that says 'Database 構築開始'. The status for 'OriginalDB2' is 'Creating compressed files 97% complete'.

**Database: OriginalDB2**

**Name**  
OriginalDB2

**Database type**  
Amino acid (protein database)

**Database directory**  
C:/inetpub/mascot/sequence/OriginalDB2/current

**Filename pattern**  
OriginalDB2\_\*.fasta

**Matching files in directory**  
OriginalDB2\_20120530.fasta (689 B)

**Database status**  
Offline

**Scheduled updates**  
(no schedules defined)

FASTA file URL has not been defined; database updates are disabled.

Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO  
Number of threads = -1 Current = YES

Name = Uniprot\_Human Family = C:/inetpub/mascot/sequence/Uniprot\_Human/current  
Filename = Uniprot\_Human\_20120530.fasta Pathname = C:/inetpub/mascot/sequence/Uniprot\_Humar  
Status = In use [Statistics](#)

State Time = Wed May 30 14:15:52 # searches = 0  
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO  
Number of threads = -1 Current = YES

**Database 構築開始**

Name = OriginalDB2 Family = C:/inetpub/mascot/sequence/OriginalDB2/current/  
Filename = OriginalDB2\_20120530.fasta Pathname = C:/inetpub/mascot/sequence/OriginalDB2/cur  
Status = Creating compressed files 97% complete  
State Time = Wed May 30 17:13:58 # searches = 0  
Mem mapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO  
Number of threads = -1 Current = NO

## ● Section 別 設定内容 :: Tasks Section

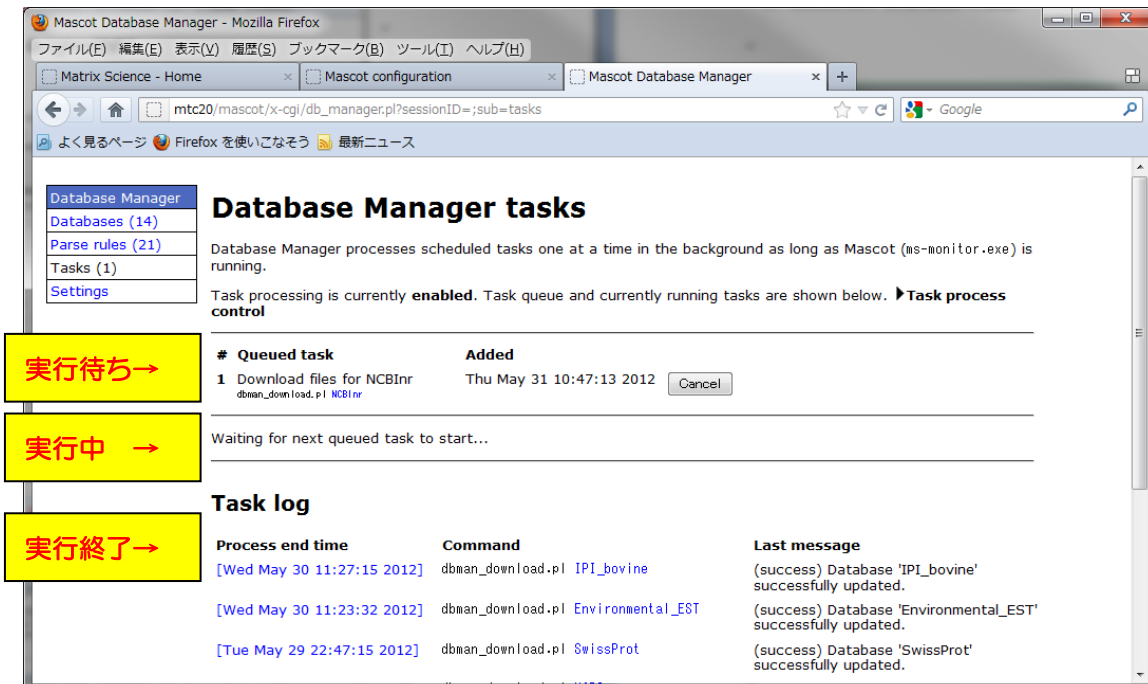
### □ データベースダウンロード進捗の確認

データベース更新プログラムのプロセスは「Task」として管理されます。更新の進捗を確認したい場合、「Tasks」セクションへアクセスしてください。

Database manager の、左フレーム「Database Manager」→「Tasks」を選びます。

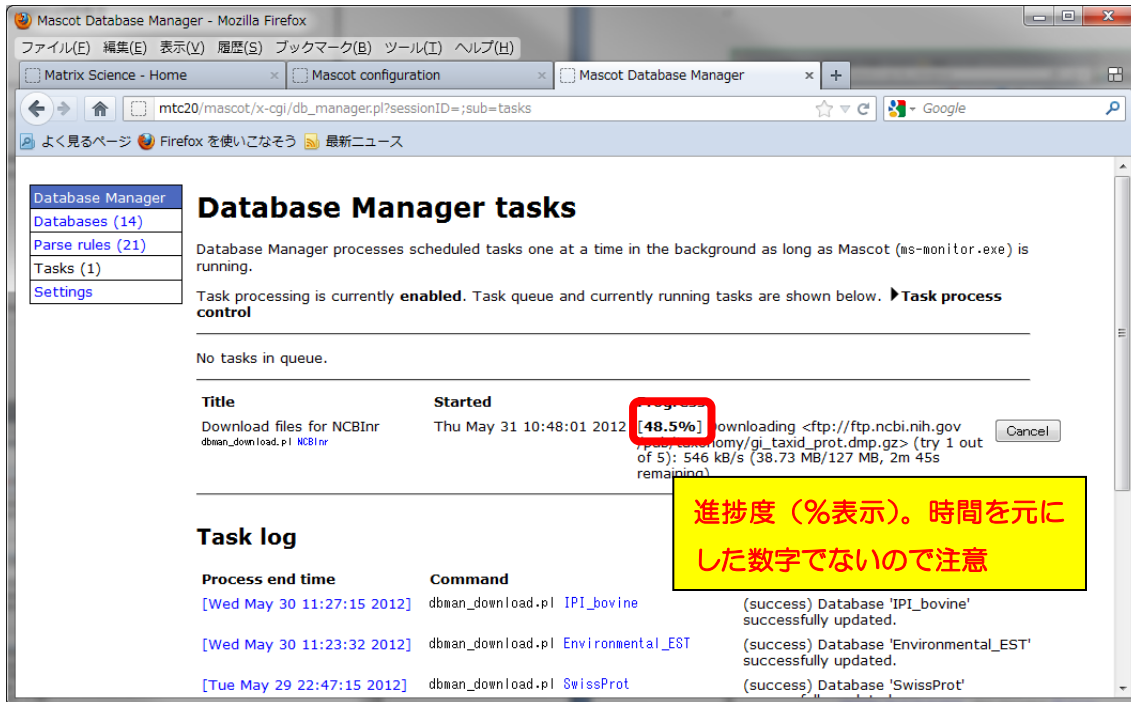


Task は 実行待ち/実行中/実行終了 の状態によって3つのセクションに分けて表示されます。



Tasks 画面

実行中のtaskについてはその進捗度が%表示で表され、ダウンロードするファイルのサイズと現段階でダウンロードしたファイルサイズ、並びに通信速度が併せて表示されます。



The screenshot shows the Mascot Database Manager interface in a Mozilla Firefox browser. The page title is "Database Manager tasks". The main content area displays a task in progress: "Download files for NCBItr" with a progress bar at 48.5%. The task details include the start time "Thu May 31 10:48:01 2012" and the command "dbman\_download.pl NCBItr". A yellow callout box highlights the progress percentage with the text: "進捗度 (%表示)。時間を元にした数字でないので注意". Below the task list is a "Task log" section showing previous successful updates for databases "IPI\_bovine", "Environmental\_EST", and "SwissProt".

Title	Started	Progress	Details
Download files for NCBItr <small>dbman_download.pl NCBItr</small>	Thu May 31 10:48:01 2012	48.5%	downloading <ftp://ftp.ncbi.nih.gov/pub/blast/blast_taxonomy/gi_taxid_prot.dmp.gz> (try 1 out of 5): 546 kB/s (38.73 MB/127 MB, 2m 45s remaining)

Process end time	Command	Details
[Wed May 30 11:27:15 2012]	dbman_download.pl IPI_bovine	(success) Database 'IPI_bovine' successfully updated.
[Wed May 30 11:23:32 2012]	dbman_download.pl Environmental_EST	(success) Database 'Environmental_EST' successfully updated.
[Tue May 29 22:47:15 2012]	dbman_download.pl SwissProt	(success) Database 'SwissProt' successfully updated.

## ● 技術サポート

アップグレードに関してご質問等ありましたら弊社技術サポートにご連絡ください。

電子メール : support-jp@matrixscience.com

電話 : 03-5807-7897                      ファックス : 03-5807-7896