

# オリジナル DB の作成・更新方法

Mascot Server においてオリジナルのデータベースを作成し検索に利用できるようにする方法と、作成・データベース構築後にデータベースの内容を更新する方法について記します。作成方法は様々な方法がありますが、ここではテキストエディタを使ってデータベースファイルをご自身で作成する方法について説明します。そのほかの作成方法についてはデータベース管理マニュアル ([http://www.matrixscience.co.jp/supportpdf/MASCOTServer\\_ver26\\_sequencedbmanage.pdf](http://www.matrixscience.co.jp/supportpdf/MASCOTServer_ver26_sequencedbmanage.pdf)) の P.19~をご参照ください。

作成手順ですが、まず MASCOT サーバーのデータベース設定の新規作成を開始して途中まで進め、続いて指定フォルダ内でデータベースファイルを作成してから設定の続きを再開し、完了させます。

一度設定したオリジナルデータベースの内容を更新したい場合は、ファイルの中身を変更し上書き保存するだけで自動的に MASCOT 用のデータベースも更新されます。

## 1. MASCOT データベース設定の新規作成

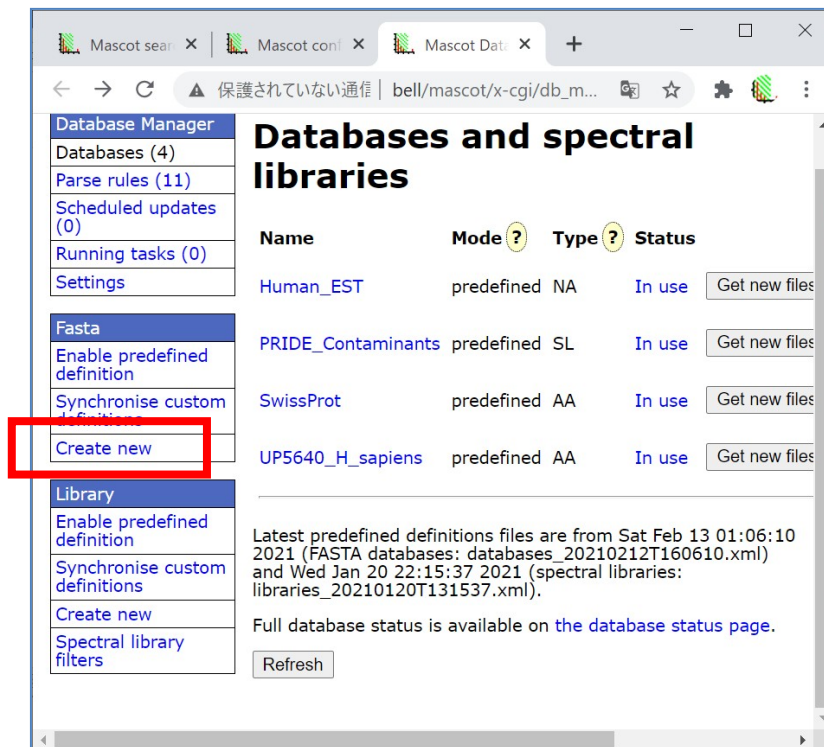
### 1-1. Database Manager ページで新規データベースの作成 (途中まで)

Home -> **Configuration Editor** -> **Database Manager** ページを開きます。

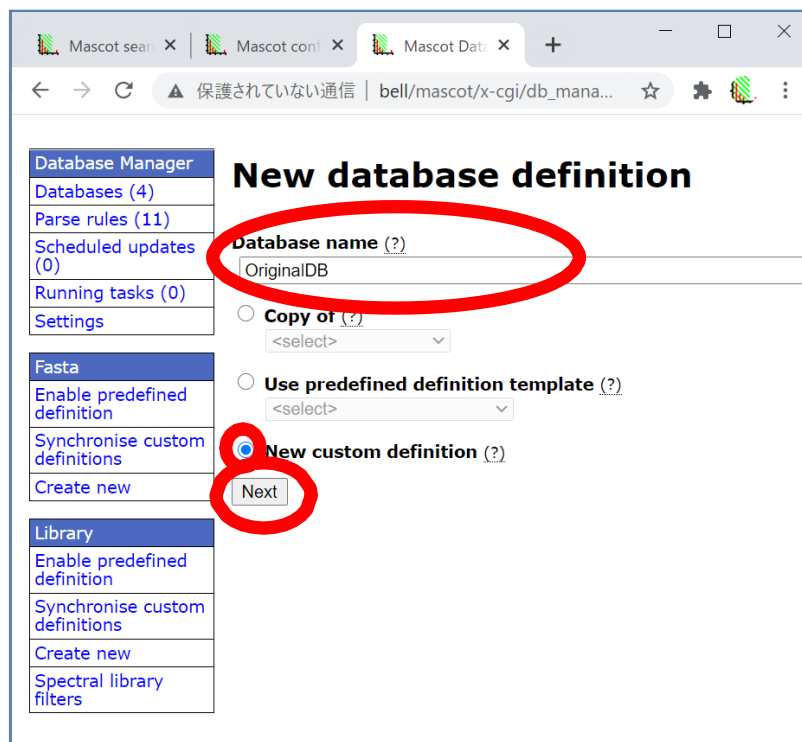
The image shows two overlapping screenshots from the Mascot web interface. The background screenshot is the 'Mascot Utilities' page, which has a sidebar with links: 'Database Status', 'Search Log', 'Configuration Editor' (highlighted with a red box), and 'Database Manager'. The foreground screenshot is the 'Mascot Configuration' page, which displays a table of configuration options. The 'Database Manager' link at the bottom of this table is also highlighted with a red box.

Amino Acids	Amino Acid Data
Modifications	Modification definitions
Symbols	Symbols used in chemical formulae
Linkers	Linker definitions
Enzymes	Enzyme definitions
Instruments	Fragmentation Rules
Quantitation	Quantitation Methods
Crosslinking	Crosslinking Methods
Configuration Options	Global Options in mascot.dat
Database Manager	Sequence databases, Parse Rules

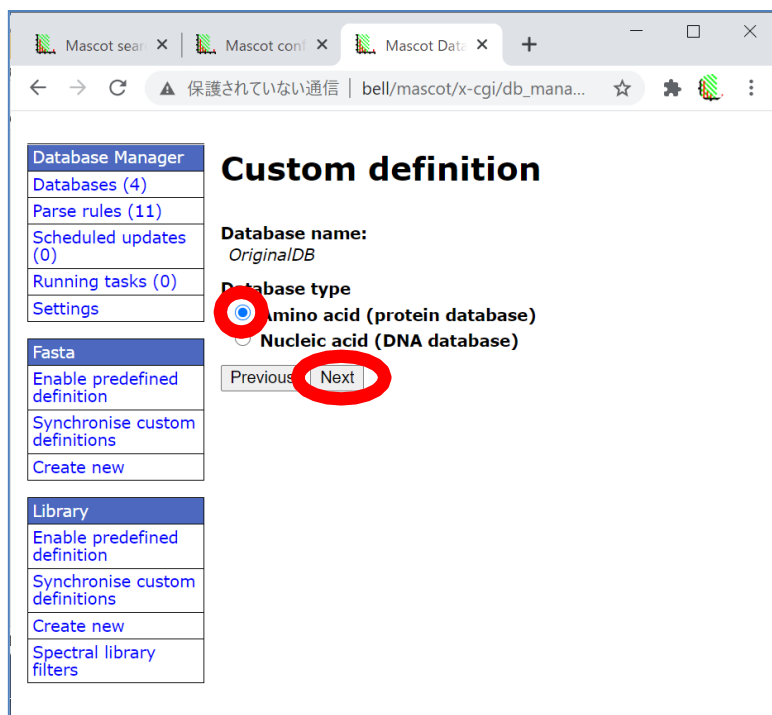
Database Manager 画面で、左フレームにある「Fasta」欄、「Create new」をクリックします。



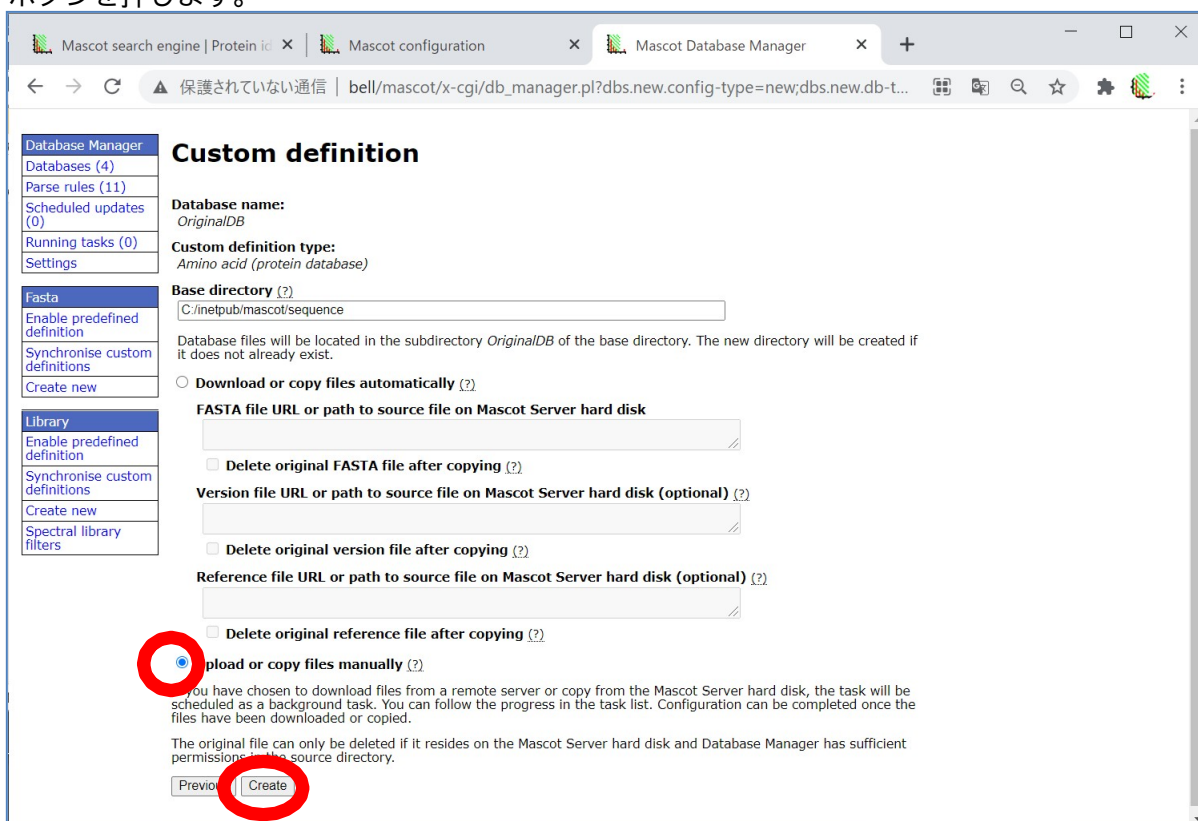
Database の名称(Database name)を入力、種類として「New custom definition」を選択し、「Next」ボタンをクリックします。



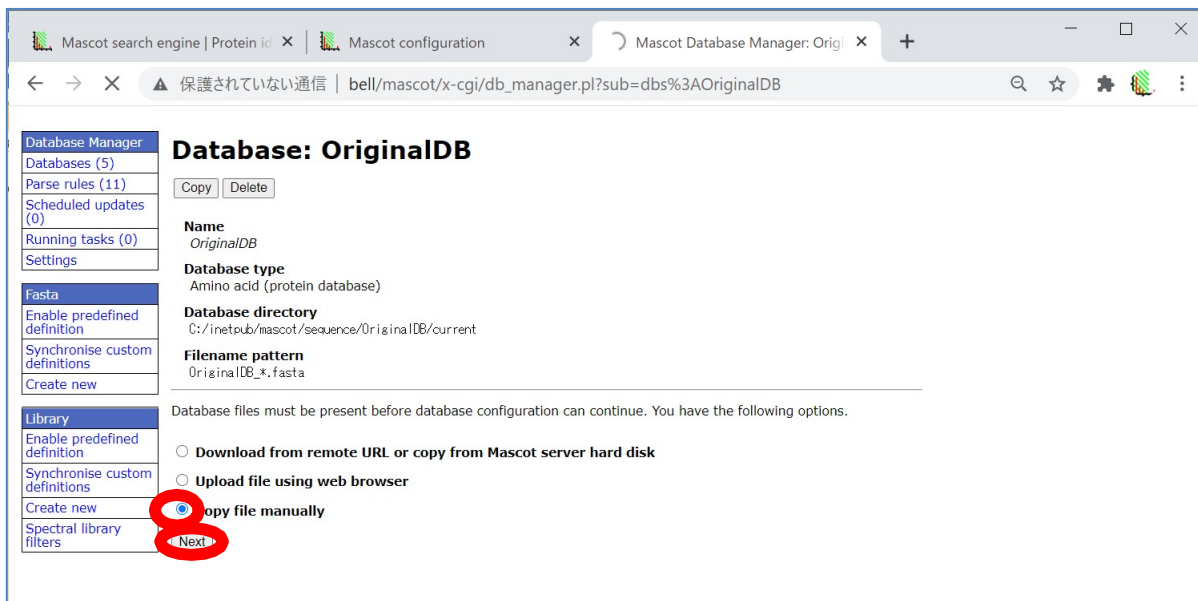
「Database type」でアミノ酸配列か塩基配列かを選びます（下図例ではアミノ酸を選択しています）。選択後、「Next」ボタンを押します。



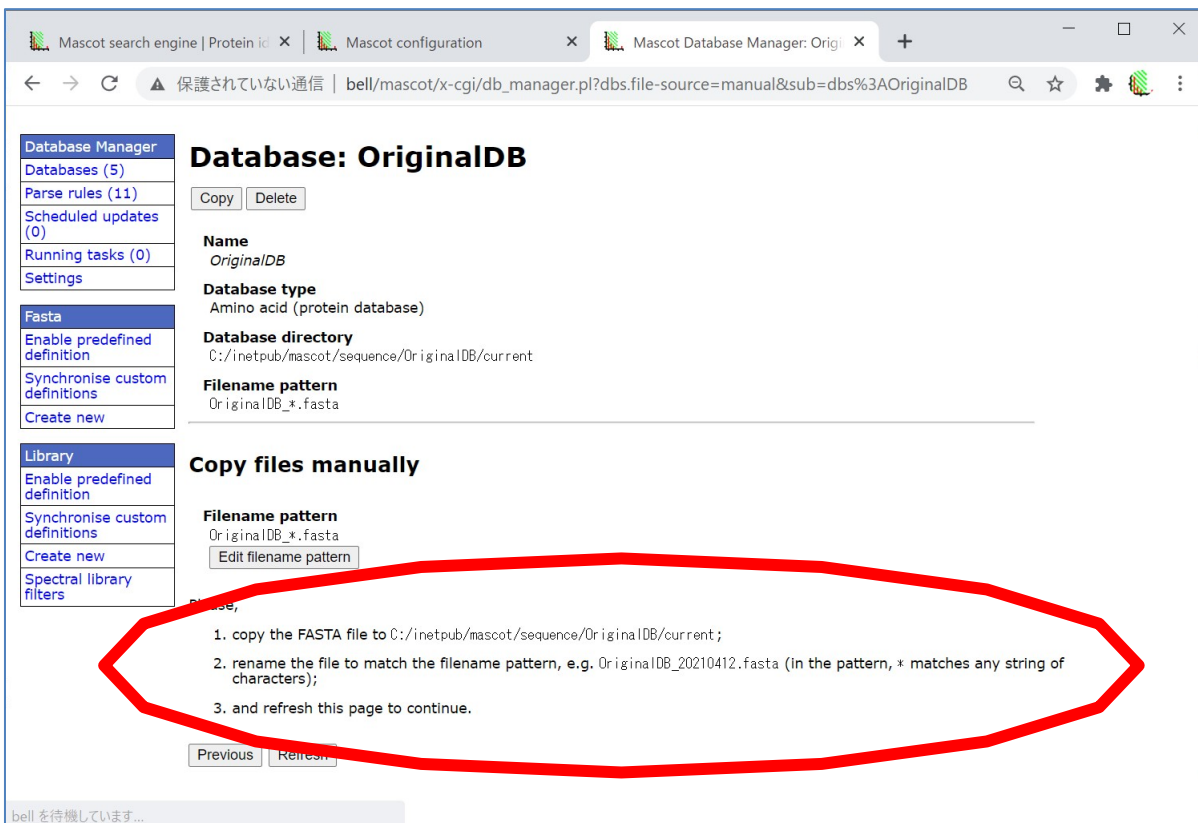
今回は自分でファイル作成する状況を想定した手順書なので、×「Download or copy files automatically」でなく、○「Upload or copy files manually」を選択します。選択後、「Create」ボタンを押します。



ファイルの準備の仕方を選びます。今回は「Copy files manually」を選択後、「Next」ボタンを押します。

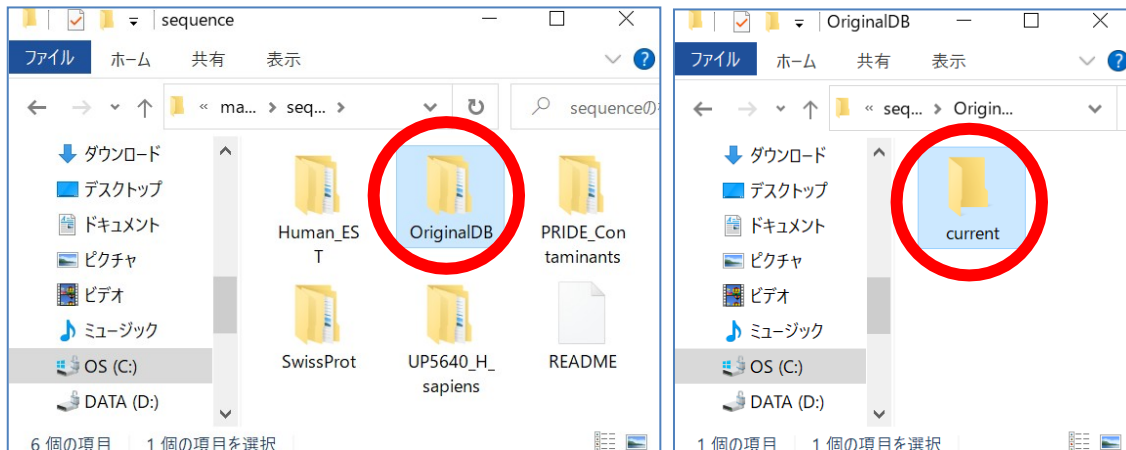


ここまででひとまず設定を中断します。図のように、所定の場所にデータベースのファイルがなく、所定のフォルダにデータベースファイルをコピーするよう促される画面が表示されます。



## 1-2. FASTAファイルの作成

MASCOT インストールフォルダ内の、sequence フォルダに、**新たに作成したデータベース名のフォルダ**（例では「OriginalDB」）がMASCOTによって自動的に作成されています。このフォルダを開き、フォルダ内にある **current** フォルダをさらに開いてください。



current フォルダ内に、データベースファイルを作成します。

操作) フォルダ内で右クリック、新規作成→テキストファイル

ファイル名は、「(DB名)\_(任意の文字列).fasta」という名称にします。

例) OriginalDB\_20210412.fasta

[次頁に続きます]

```

OriginalDB_20210412.fasta - メモ帳
ファイル(E) 編集(E) 書式(O) 表示(V) ヘルプ(H)
>ID0001 ProteinA
PLEASEENTERAASEQ
>ID0002 BSA
MKWVTFISLLLLFSSAYSARGVFRDRTHKSEIAHRFKDLGEEHFKGLVLI
FSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCK
VASLRETYGDMADCCCKQEPERNECFLSHKDDSPDLPKLKPDPNTLCDEF
KADEKFFWGKYL YE IARRHPYFYAPELLYYANKYNGVFQECQAEKDGAC
LLPKIETMREKVLASSARQLRCASIQKFGERALKAWSVARLSQKFPKAE
FVEVTKLVTDLTKVHKECCGGDLLECCADDRADLAKYICDNQDTISSKLE
CCDKPLLEKSHCIAEVEKDAIPENLPPLTADF AEDKDVCKNYQEAQDAFL
GSFLYEYSRRHPEYAVSVLLRLAKEYEATLECCAKDDPHACYSTVFDKL
KHLVDEPNLIKQNCDDQFEKLGEGYGFQNALIVRYTRKVPQVSTPTLVEVS
RSLGKVGTRCC TKPESERMPCTEDYLSLILNRLCYLHEKTPVSEKVTKCC
TESLVNRRPCFSALTPDETYVPKAFDEKLF TFHADICTLPDTEKQIKKQT
ALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
STQTALA
>ID0003 BSA2
MKWVTFISLLLLFSSAYSARGVFRDRTHKSEIAHRFKDLGEEHFKGLVLI
AFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCK
VASLRETYGDMADCCCKQEPERNECFLSHKDDSPDLPKLKPDPNTLCDEF
KADEKFFWGKYL YE IARRHPYFYAPELLYYANKYNGVFQECQAEKDGAC
LLPKIETMREKVLASSARQLRCASIQKFGERALKAWSVARLSQKFPKAE
FVEVTKLVTDLTKVHKECCGGDLLECCADDRADLAKYICDNQDTISSKLE
CCDKPLLEKSHCIAEVEKDAIPENLPPLTADF AEDKDVCKNYQEAQDAFL
GSFLYEYSRRHPEYAVSVLLRLAKEYEATLECCAKDDPHACYSTVFDKL
KHLVDEPNLIKQNCDDQFEKLGEGYGFQNALIVRYTRKVPQVSTPTLVEVS
RSLGKVGTRCC TKPESERMPCTEDYLSLILNRLCYLHEKTPVSEKVTKCC
TESLVNRRPCFSALTPDETYVPKAFDEKLF TFHADICTLPDTEKQIKKQT
ALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
STQTALA
26 行, 48 列 100% Windows (CRLF) UTF-8

```

各タンパク質情報の入力方法は以下の通りです。

- エントリー（各タンパク質）はまず、“>”からはじめます。全角でなく半角です。以降入力はすべて半角文字でおこなってください。
- 以後スペースを入れずにデータベースの ID を記入します。次の半角スペースまでがすべて ID として認識されます。各エントリーでこの ID を重複させないでください。例では、最初のエントリーの ID として”ID0001”と記入していますが、以降 ID で ID0001 を使うとエラーになりますのでご注意ください。ID はどのような記述でも問題ありませんが、長すぎると検索速度が低下するので、重複しない限りで短くしてください。ID 入力後に半角スペースを入力します。
- 半角スペースの次に、タンパク質の名前などの情報を入れます。最初のエントリーの例では「ProteinA」と入力しています。最初のスペース以降は複数個のスペースが使用されても問題ありません。
- 続けて改行をし、ID0001 のアミノ酸配列を入力します。アミノ酸配列部分は任意に改行を入れても問題ありません。（ID0002 と ID0003 を比較してください。）次のエントリーの始まりを示す、“>”までがすべて前エントリーのアミノ酸配列と認識されます。

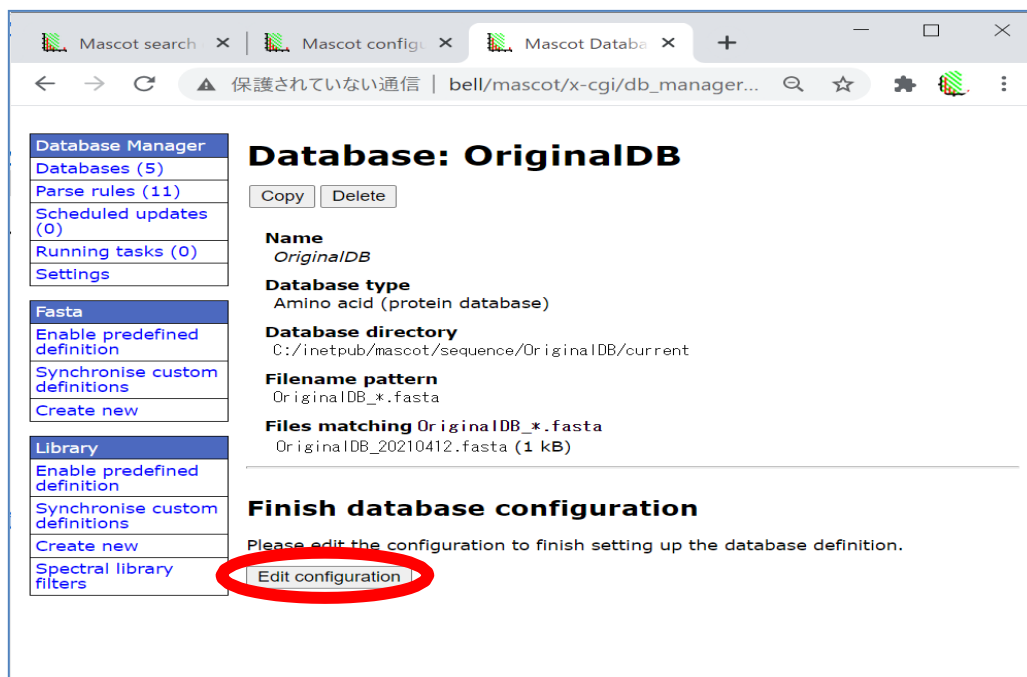
なお例ではアミノ酸配列を入力しましたが、塩基配列でも問題ありません。ただし塩基配列の場合は既に設定した項目である”database type”の選択が”Amino acid”でなく”Nucleic Acid”となりますので注意してください。

データベースの入力を終わったら、ファイルを保存してください。



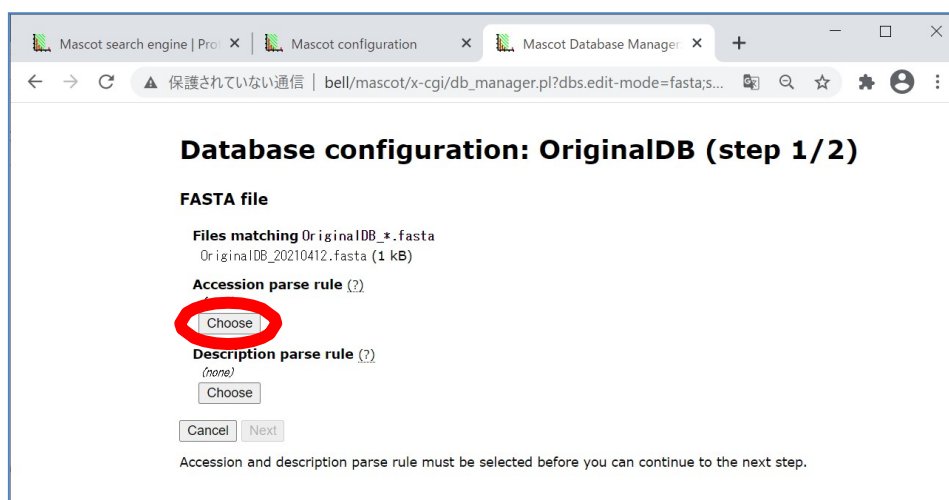
### 1-3. Database Manager 設定 (中断箇所からの続き)

1-1の最後の過程の画面を改めて見ると、1-2 で作成されたファイルが画面内で認識され、設定の続きを行うためのボタン「**Edit Configuration**」が出ていますのでこのボタンを押して次の画面を開きます(下図)。もし「Edit Configuration」ボタンが画面に出ず、1-1 の最後の画面のままである場合、ファイル名の付け方やファイルの場所を間違えている可能性があります。もう一度ファイル名とファイルが置かれている場所をご確認ください。



データベースの、ID(Accession)部分、Description 部分を FASTA 先頭行から抜き出すための Parse Rule 定義を行います。

まず ID(Accession)部分の抽出ルールを設定するため、「Accession parse rule」下の「**Choose**」ボタンを押します。



ここでは FASTA の先頭行のうち、1 番目に注目します（下図赤線）。「ID0001 ProteinA」とありますが、このうち「ID0001」の部分を ID(Accession) としたいとします。

MASCOT が、候補となる Parse Rule を自動的に左側にリストアップし、そのルールを適用した場合の実際の ID (Accession)の例が右側に表示されています。ここでは目的としている抜き出し方である「ID0001」（下図青線）となっている、上から三番目か四番目の抽出ルールを選びます（上の赤丸）。どちらかをクリックしてラジオボタンを選択し、「Choose」ボタンを押します。

**Raw test data (3 samples)**

- >ID0001 ProteinA
- >ID0002 BSA
- >ID0003 BSA2

**Suitable parse rules (8)**

Parse rule	Match	Extracted data
<input type="radio"/> >[^ ]* ¥(.*¥)	3/3	1. ProteinA ▶ 2 more matches
<input type="radio"/> >¥(.*¥)	3/3	1. ID0001 ProteinA ▶ 2 more matches (Warning: contains spaces)
<input checked="" type="radio"/> ^(^,)*¥	3/3	1. ID0001 ▶ 2 more matches
<input type="radio"/> ^([ ])*¥	3/3	1. ID0001 ▶ 2 more matches
<input type="radio"/> >¥([ ])*¥	3/3	1. ID0001 ProteinA ▶ 2 more matches (Warning: contains spaces)
<input type="radio"/> [^ ]* ¥¥(.*¥)	3/3	1. ProteinA ▶ 2 more matches
<input type="radio"/> ¥(.*¥)	3/3	1. >ID0001 ProteinA ▶ 2 more matches (Warning: contains spaces)
<input type="radio"/> ¥([ ])*¥	3/3	1. >ID0001 ▶ 2 more matches

**► Unsuitable parse rules (42)**

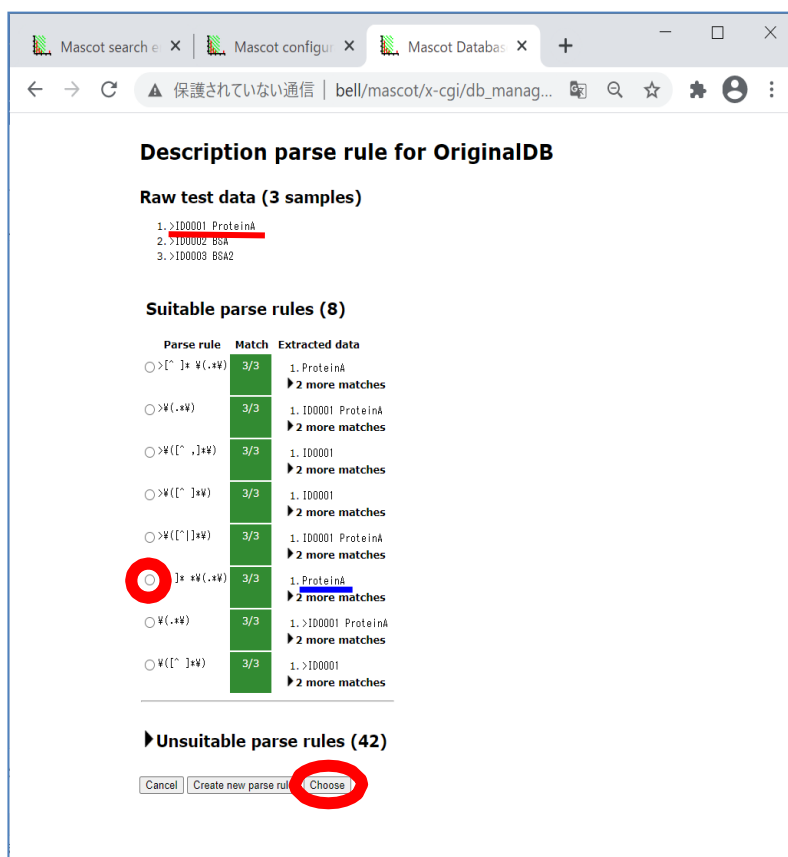
Cancel Create new parse rule Choose



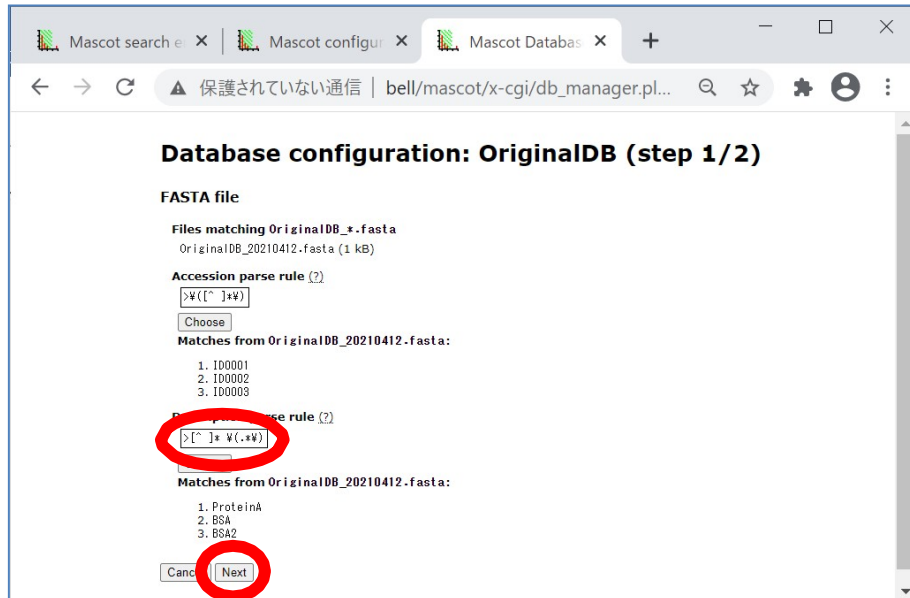
Choose ボタンを押すと元の画面に戻ります。「Accession parse rule」の下には、先ほど選択した Parse rule が表示されています（下図の上赤丸）。続いて、Description 部分の抽出ルールを設定するため、「Description parse rule」下の「Choose」ボタンを押します。



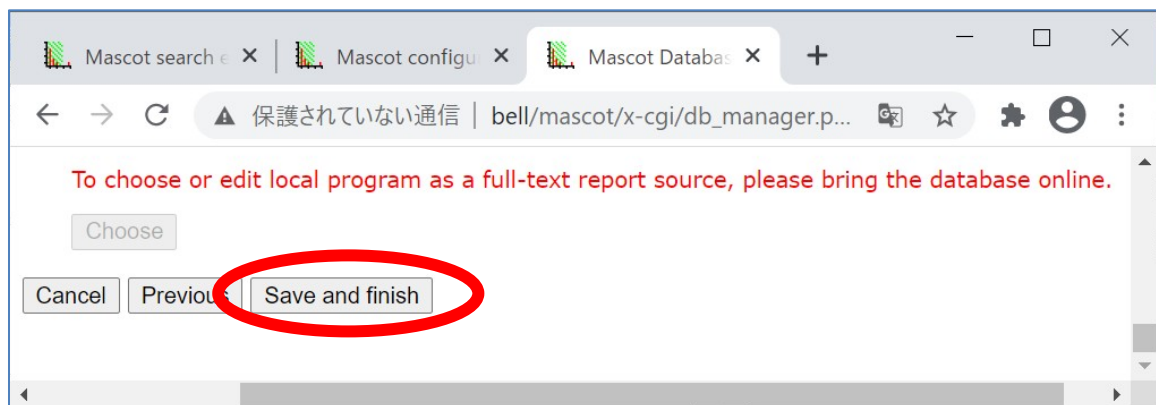
先程のAccessionの設定同様、まず FASTA の先頭行のうち 1 番目に注目します（下図赤線）。「I00001 ProteinA」の、「ProteinA」の部分を選択します。MASCOTが候補となる Parse Rule を左側に、そのルールを適用した場合の実際の Description の例が右側に表示されています。抜き出し方が「ProteinA」（下図青線）となっている、一番上の抽出ルールのラジオボタンをクリックしてラジオボタンを選択し「Choose」ボタンを押します。



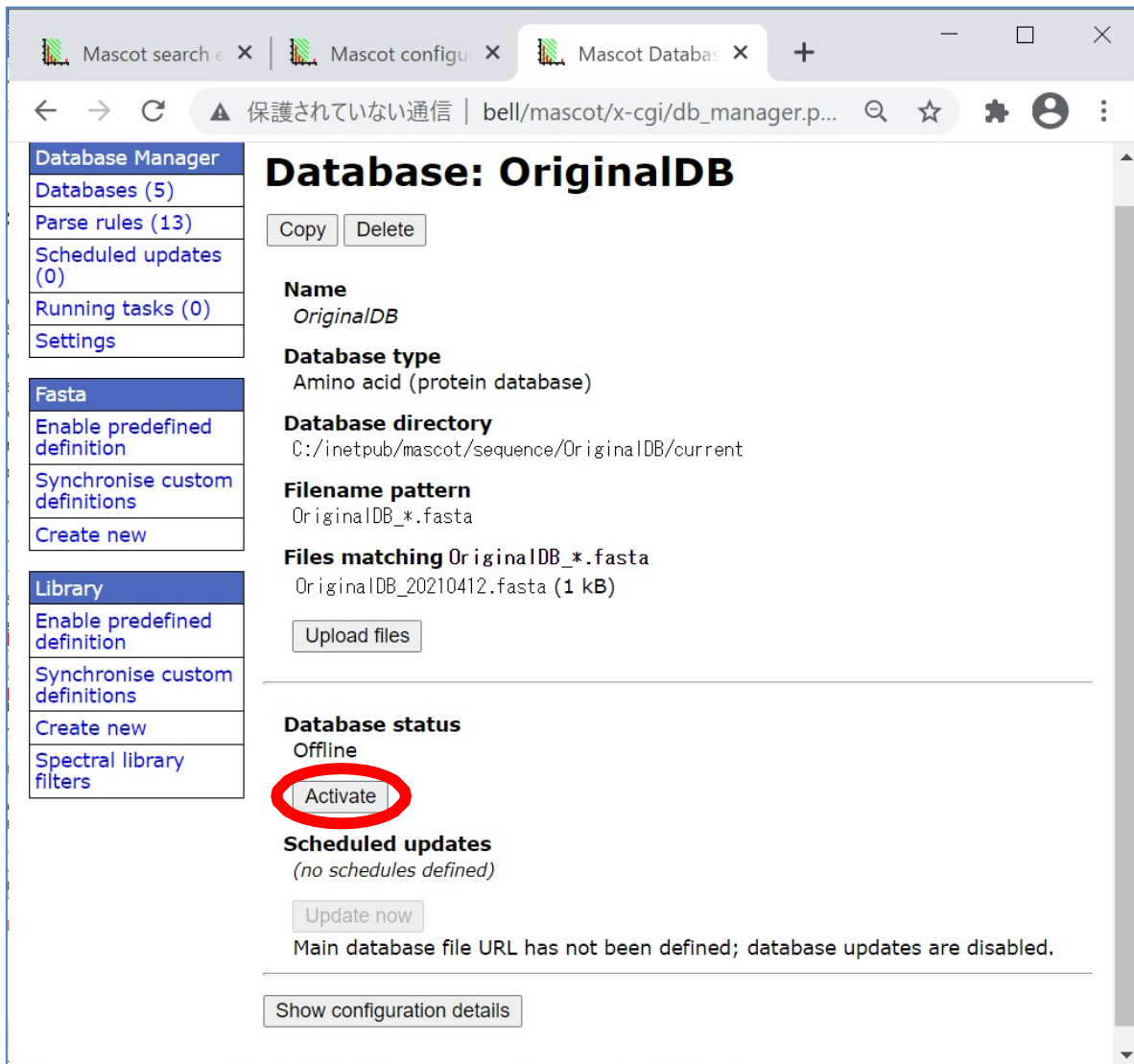
Description にも新たな Parse rule が入力されているのを確認したのち、画面最下部の「Next」ボタンを押してください。



続いて現れる設定画面について、今回の例ではデフォルト値から変更する必要はありません。画面最下部の「Save and finish」ボタンを押します。



最後に設定内容を確認する画面が現れます。このままこのデータベースを使用可能な状態とするには「**Activate**」ボタンを押します。



### 1-4. データベース構築の確認

新規に作成したデータベースが正しく認識されているか、確認をします。

Database Status 画面にて、先ほど作成した DB が新たに追加され、**Status が “In Use”**となっているかを確認してください（下図）また、サンプルデータなどで検索テストを行い、実際に検索がうまくいくかどうかをご確認ください。

```

Name      = PRIDE_Contaminants  Family   = C:/inetpub/mascot/sequence/PRIDE_Contaminants/current/PRIDE_Contaminants
Filename  = PRIDE_Contaminants_20160908.msp Pathname = C:/inetpub/mascot/sequence/PRIDE_Contaminants/current/PRIDE_Contaminants_20160908.msp
Status    = In use
State Time = Tue Apr 13 09:33:38 # searches = 0
Mem mapped = NO Request to mem map = NO Request unmap = NO Mem locked = NO
Number of threads = 1 Current = YES Type = Spectral library

Name      = SwissProt          Family   = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta
Filename  = SwissProt_2020_05.fasta Pathname = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_2020_05.fasta
Status    = In use
State Time = Tue Apr 13 09:33:38 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = YES Type = Amino acid

Name      = UP5640_H_sapiens   Family   = C:/inetpub/mascot/sequence/UP5640_H_sapiens/current/UP5640_H_sapiens
Filename  = UP5640_H_sapiens_20201007.fasta Pathname = C:/inetpub/mascot/sequence/UP5640_H_sapiens/current/UP5640_H_sapiens_20201007.fasta
Status    = In use
State Time = Tue Apr 13 09:33:38 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = YES Type = Amino acid

Name      = OriginalDB        Family   = C:/inetpub/mascot/sequence/OriginalDB/current/OriginalDB_*.fasta
Filename  = OriginalDB_20210412.fasta Pathname = C:/inetpub/mascot/sequence/OriginalDB/current/OriginalDB_20210412.fasta
Status    = In use
State Time = Tue Apr 13 11:20:04 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = -1 Current = YES Type = Amino acid

```

なお、下図のように Status が “In use” 以外の表示となっている場合は、恐れ入りますがそのメッセージ内容と、“check error log” をクリックし表示されているデータベース構築時当日のエラー内容を弊社までご連絡ください。

```

Name      = OriginalDB        Family   = C:/INETPUB/MASCOT/sequence/OriginalDB/current/OriginalDB_*.fasta
Filename  = OriginalDB_20210412.fasta Pathname = C:/INETPUB/MASCOT/sequence/OriginalDB/current/OriginalDB_20210412.fasta
Status    = Halted: First test failed. check error log fix fault and retry
State Time = Tue Apr 13 11:20:04 # searches = 0
Mem mapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = 4 Current = NO

```

## 2. オリジナルデータベースの更新

オリジナルデータベースの内容を変更したい場合は、  
 C:¥inetpub¥mascot¥sequence¥(DB名)¥current

にあるデータベースファイルのうち、拡張子が.fasta のファイル（最初に作成したファイル）をテキストエディタで開き、その内容を書き換えて保存してください。

保存後、MASCOT が自動的に認識し、検索用 DB を作成し直します。作成のやり直しには1分ほど時間がかかります。

以上です。ご不明な点がございましたら下記問い合わせ先までご連絡ください。

<問い合わせ先>

### 技術サポート

電子メール : support-jp@matrixscience.com  
 電 話 : 03-5807-7897  
 ファックス : 03-5807-7896