

# リリース 51.0 以降の Swiss Prot データベースのタイトル行書式 変更に伴う Mascot Server 側の設定変更について

## 1. はじめに

Swiss Prot データベースの FASTA ファイルを構成する各エントリのタイトル行はリリース 50.9 以前では次のようになっています。

<リリース 50.9 までのタイトル行>

```
>104K_THEAN (Q4U9M9) 104 kDa microneme-rhoptry antigen precursor (p104)
```

これに対し、2006 年 10 月 31 日に公開されたリリース 51.0 では、次の赤い文字列部分（アクセッション番号の位置変更と生物種の追加）の変更が加えられています。

<リリース 51.0 のタイトル行>

```
>Q4U9M9|104K_THEAN 104 kDa microneme-rhoptry antigen precursor (p104) - Theileria annulata
```

エントリの固有番号 (ID) である“104K\_THEAN”の前にアクセッション番号 (AC) である赤い文字列“Q4U9M9”が付け加えられたため、現在 Mascot Server 側で設定されている ID を抽出するためのルール“>¥([<sup>^</sup>]\*¥)”を使用した場合“104K\_THEAN”を抽出せずに“Q4U9M9|104K\_THEAN”を抽出してしまい、50.9 以前のリリースから 51.0 に更新する際に、FASTA ファイルの“104K\_THEAN”エントリと対になるリファレンスファイル (dat ファイル) の“104K\_THEAN”エントリを特定できずに次のエラーが返され、リリース 51.0 のデータベースをセットアップすることができません。

なお、更新前のリリースは継続して使用できる状態になっています。

Error [M00078 - Job -16 - X00614:fasta] - Tue Nov 01 10:40:17 2006

- In ref file, expected '104K\_THEAN' to match with 'Q4U9M9|104K\_THEAN' in fasta file

MASCOT search status page

Version: 2.1.04 - Licensed to: Matrix Science in-house demonstration system (Tokyo), (4 processors).  
1 Intel processor (No hyper-threading in cpu, single core). [0 searches running]

[Search log](#) [monitor log](#) [error log](#) [Error message descriptions](#) [Do not auto refresh this page](#)

Name	= Sprot	Family	= D:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_*.fasta
Filename	= Sprot_51.0.fasta	Pathname	= D:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_51.0.fasta
Status	= Halted: Failed to compress <a href="#">check compression log</a> or <a href="#">check error log</a> fix fault and retry		
State Time	= Mon Nov 13 15:03:01 # searches = 0		
Mem mapped	= NO	Request to mem map	= YES
Request unmap	= NO	Mem locked	= NO
Number of threads	= 4	Current	= NO

Name	= Sprot	Family	= D:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_*.fasta
Filename	= Sprot_49.1.fasta	Pathname	= D:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_49.1.fasta
Status	= In use <a href="#">Statistics</a> <a href="#">Unidentified taxonomy</a>		
State Time	= Mon Nov 13 15:03:21 # searches = 0		
Mem mapped	= YES	Request to mem map	= YES
Request unmap	= NO	Mem locked	= NO
Number of threads	= 4	Current	= YES

Status=Halted:Failed to compress  
となり更新されない。

更新前のリリースは status=In use  
になっており使用することが可能。

## 2. リリース 51.0 に対応する ID 抽出ルール

リリース 51.0 のタイトル行に対応する ID 抽出ルールは次のようになります。

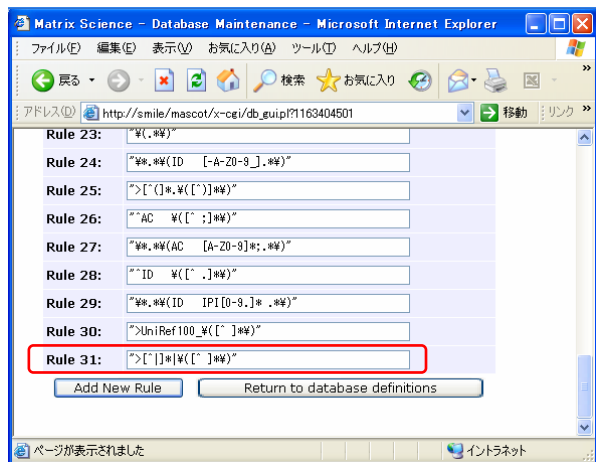
```
">[^]*|¥([^\ ]*¥)"
```

次の「3. ID 抽出ルールの変更方法：GUI を利用する方法」あるいは「4. ID 抽出ルールの変更方法：設定ファイルを編集する方法」で Swiss Prot の ID 抽出ルールを変更してください。

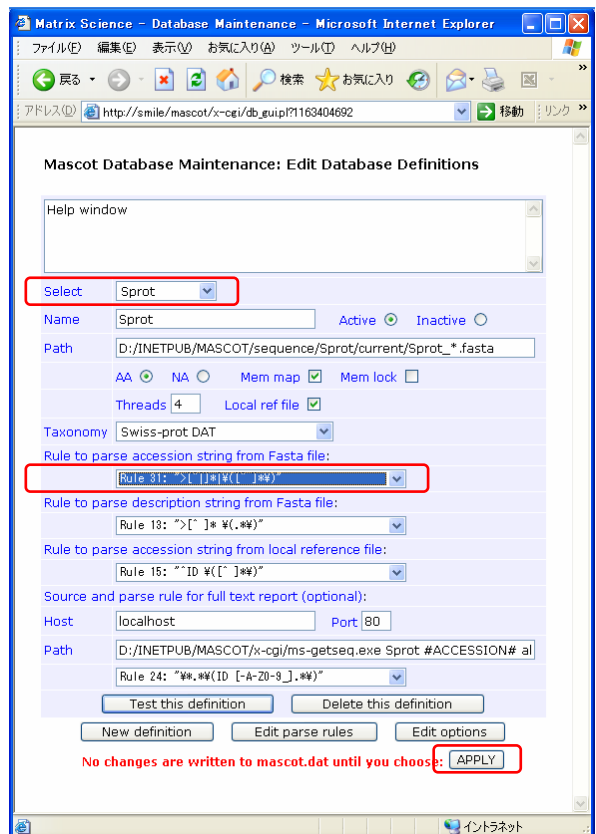
## 3. ID 抽出ルールの変更方法：GUI を利用する方法

(3-1) Mascot Server のトップページの (<http://ホスト名/mascot/>) の [Database Maintenance] リンクをクリックしてください。「Edit Database Definitions」ページが開きますので、ページ下方にある [Edit parse rules] ボタンを押してください。「Edit Parse Rules」ページが開きます。

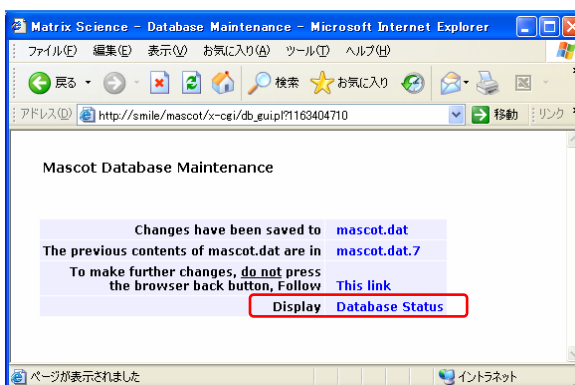
(3-2) 「Edit Parse Rules」ページ下方にある [Add New Rule] ボタンを押すと、新規ルール定義欄（右図では Rule 31）が現れますので、定義欄にリリース 51.0 に対応する ID 抽出ルール ">[^]\*|¥([^\ ]\*¥)" を入力し [Return to database definitions] ボタンを押してください。「Edit Database Definitions」ページが表示されます。



(3-3) 「Edit Database Definitions」ページの [Select] リストから "Sprot" を選択した後、[Rule to parse accession string from Fasta file:] リストから (3-2) で定義した Rule 番号を選択してください。右図では Rule 31 を選択しています。最後に [APPLY] ボタンを押してください。設定変更が有効になり、設定ファイルである mascot.dat ファイルにも変更が加えられ、[Mascot Database Maintenance] ページが表示されます。



- (3-4) [Mascot Database Maintenance] ページの [Database Status] リンクをクリックしてください。[MASCOT search status page] ページが表示されます。

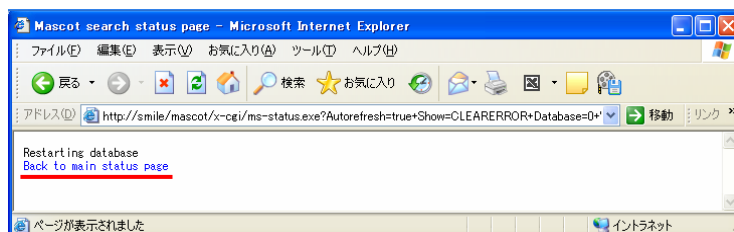


- (3-5) [MASCOT search status page] ページの Sprot ブロックの表示内容、特に status 行の内容を確認してください。リリース 51.0 の status は “Halted” になっていますが、更新前のリリースの status は In use になっています。

```
Name      = Sprot      Family = D:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_*.fasta
Filename  = Sprot_51.0.fasta Pathname = D:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_51.0.fasta
Status    = Halted: Failed to compress check compression log or check error log fix fault and retry
State Time = Mon Nov 13 15:03:01 # searches = 0
Mem mapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = 4 Current = NO

Name      = Sprot      Family = D:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_*.fasta
Filename  = Sprot_49.1.fasta Pathname = D:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_49.1.fasta
Status    = In use      Statistics Unidentified taxonomy
State Time = Mon Nov 13 15:03:21 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = 4 Current = YES
```

リリース 51.0 の status 行の最後にある [retry] リンクをクリックしてください。リリース 51.0 の更新プロセスが開始され、「Restarting database」ページに切り替わりますので、[Back to main status page] リンクをクリックしてください。[MASCOT search status page] ページが表示されます。



- (3-6) [MASCOT search status page] ページにおいて、リリース 51.0 の status 行は更新プロセスの進行とともに次のように変化し、“In use” が表示されれば更新作業は終了です。

```
Status = Creating compressed files    n% complete
Status = Running 1st test
Status = First test just run OK
Status = Trying to memory map files
Status = Just enabled memory mapping
Status = In use
```

#### 4. ID 抽出ルールの変更方法：設定ファイルを編集する方法

Mascot Server の設定ファイルである `mascot.dat` ファイルをテキストエディタで開き、次のように編集してください。

(4-1) [PARSE] ブロックにリリース 51.0 に対応する ID 抽出ルールを定義してください（ここでは `RULE_31` で定義しています。）。

```
# for parsing ID of r51.0 and later
RULE_31 ">[^\|]*|¥([^\|]*¥)"
```

(4-2) [Databases] ブロックの Sprot 行の “Index of parse rule for accession number from FASTA” に対応する部分の数字（ID 抽出ルール番号）を `12` から(4-1)で定義した番号（ここでは `31`）に変更してください。

<変更前>

```
Sprot C:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_*.fasta AA 1234 5 1 1 1 0 1 12 13 15 5
```

<変更後>

```
Sprot C:/INETPUB/MASCOT/sequence/Sprot/current/Sprot_*.fasta AA 1234 5 1 1 1 0 1 31 13 15 5
```

(4-3) [MASCOT search status page] ページを開き、(3-5) および (3-6) の操作を行い、再度リリース 51.0 の更新を行ってください。

#### 5. 動作確認

適当な質量データを使って、Sprot に対して Mascot 検索を実行してください。

#### 6. お問い合わせ

ご不明点などありましたら技術サポート担当者にご連絡ください。

マトリックスサイエンス株式会社  
 〒101-0021 東京都千代田区外神田 6-10-12 KNビル3階  
 電話：03-5807-7897  
 電子メール：support-jp@matrixscience.com