リリース 56.0 以降の SwissProt 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

SwissProt 配列データベース(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 51.0 から 55.6 までは次のようになっています。

<リリース 51.0~55.6 までのタイトル行>

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

これに対し、2008 年 7 月 22 日に公開されたリリース 56.0 では、タイトル行に次の赤で示した文字列 (sp))が追加されています。

<リリース 56.0 のタイトル行>

>sp[Q4U9M9]104K_THEAN 104 kDa microneme/rhoptry antigen OS=Theileria annulata GN=TA08425 PE3 SV=1

タイトル行先頭の">"の次に"sp|"が追加されたため、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 ">[^|]*|¥([^]*¥)"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、リリース 56.0 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 56.0 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 - X00413:fasta] - Sun Jul 27 02:45:55 2008 - In ref file, expected '104K THEAN' to match with 'Q4U9M9|104K THEAN' in fasta file

Name SwissProt Family = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta = SwissProt_55.6.fasta Pathname = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_55.6.fasta Filename = In use Status <u>Statistics</u> <u>Unidentified taxonomy</u> 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。 Name = SwissProt Family = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta = <u>SwissProt</u> Family = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_56.0.fasta = SwissProt_56.0.fasta Pathname = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_56.0.fasta Filename check compression log or check error log fix fault and retry Status=Halted:Failed to compress となり更新されない。

2. リリース 56.0 に対応する ID 抽出ルール

リリース 56.0 のタイトル行から ID を抽出するための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」に従って SwissProt 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

- (3-1) Mascot Server のトップページの (http://ホスト名/mascot/) の [Configuration Editor]リンク をクリックし、新たに開いた「Mascot Configlation」ページの[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) [Add New Rule]ボタンを押すと新たにブランクの新規ルール定義欄(右図では Rule 33)が現れますので、リリース 56.0 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。

(3-3)「Edit Database Definitions」ページの [Select]リストから"SwissProt"を選択し てください。既存の設定内容が表示されま すので、[Rule to parse accession string from Fasta file:]リストから(3-2)で定義 した Rule 番号(右図では Rule 33)を選択し、 [APPLY]ボタンを押してください。設定変更 が有効になり(Mascot の設定ファイルであ る mascot.dat に変更内容が書き込まれ)、 [Mascot Database Maintenance] ページが 表示されます。

C	Matrix Scie	nce – Database Maintenance – Windows Internet Ex.	💶 🗖 🔀				
G	🔊 - 🧕	http://node21/masc 💙 🗲 🗙 Google	P -				
ファイル(E) 編集(E) 表示(V) お気に入り(<u>A</u>) ツール(T) ヘルプ(H)							
	* 😫 🕶	🌈 Matrix 🗙 🌇 Matrix Sc 👘 🔹 🔂 🛀 🖶 🔹	₽• @ • [»]				
	Rule 30:	">UniRef100_\([^]*\)"	^				
	Rule 31:	">[^]* \([^]*\)"					
	Rule 32:	">\([^\]*\)"					
	Rule 33:	">![^!]*\\([^]*\)"					
	Add No	ew Rule Return to database definitions					
./db	_gui.pl?1217222	03 🔰 🔤 🖳 🕄 🕄 🕄 🕄	🔍 100% 🔹 💡				

🖉 Matrix Science – Database Maintenance – Windows Internet Explorer 📃 🔲	x						
🚱 🗣 🔊 http://node21/mascot/x-cgi/v 😽 🗙 Google	•						
ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)							
😭 🍁 🔐 🗸 🍘 Matrix Scie 🗙 🗽 Matrix Science 🍡 🏠 🔹 🖶 🛛 🖓 👻 🥐							
Mascot Database Maintenance: Edit Database Definitions	^						
Help window							
Select SwissProt 💌							
Name SwissProt Active 💿 Inactive 🔿							
Path C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta							
AA 💿 NA 🔿 🛛 Mem map 🗹 Mem lock 🗖							
Threads 1 Local ref file 🗹							
Taxonomy Swiss-prot DAT							
Rule to parse accession string from Fasta file:							
Rule 33: "> [^]* \([^]*\)"							
Rule to parse description string from Fasta file:							
Rule 18: ">[^]* \(.*\)"							
Rule to parse accession string from local reference file:							
Rule 15: "^ID \([^]*\)"							
Source and parse rule for full text report (optional):							
Host localhost Port 80							
Path C:/inetpub/mascot/x-cgi/ms-getseq.exe SwissProt #ACCESSION#							
Rule 24: "*.*\(ID [-A-Z0-9_].*\)"							
Test this definition Delete this definition							
New definition Edit parse rules Edit options							
No changes are written to mascot.dat until you choose: APPLY							
😌 ローカル イントラネット 🔍 100% 🔻							
	-						

(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。

Mascot Database Maintenance	
Changes have been saved to	mascot.dat
The previous contents of mascot.dat are in	mascot.dat.7
To make further changes, <u>do not</u> press the browser back butt <mark>on, Follow</mark>	This link
Display	Database Status

 (3-5) [MASCOT search status page] ページの SwissProt 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 56.0 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
      = SwissProt
      Family
      = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta

      Filename
      = SwissProt_55.6.fasta
      Pathname = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_55.6.fasta

      Status
      = In use
      Statistics
      Unidentified taxonomy

      State Time = Thu Jul 10 003:14:15 # searches = 0
      Mem mapped = YES
      Request to mem map = YES
      Request unmap = NO

      Name
      = SwissProt
      Family
      = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta

      Filename
      = SwissProt_56.0.fasta
      Pathname = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_*.fasta

      State Time
      = SwissProt_58:13 # searches = 0
      Family
      = C:/inetpub/mascot/sequence/SwissProt/current/SwissProt_Sc.fasta

      State Time
      = Thu Jul 24 01:85:13 # searches = 0
      Inet compress
      Inet check error log
      fix fault

      Mem mapped = NO
      Request to mem map = YES
      Request unmap = NO
      Mem locked = NO
```

リリース 56.0 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 56.0 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。

🗿 Mascot search status page - Microsoft Internet Explorer							
- ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルブ(H)							
🚱 हुई • 🕥 - 💌 🙆 🏠 🔎 क्षेत्र 🌪 केन्न्राट्र्रेश 🤪 🔗 🌭 🔯 • 🔔 🚳							
:アドレス 🛛 🗃 http://smile/mascot/x-cgi/ms-status.exe?Autorefresh=true+Show=CLEARERROR+Database=D+ 💌 🔁 移動	ヵ (リンク >						
Restarting database Back to main status page							
	~						
ど ページが表示されました	<u>ان</u> ۱						

(3-6) [MASCOT search status page] ページにおいて、リリース 56.0 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. 動作確認

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

5. 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