

菌の輪を描く -GraPhIAnによる種名リストの系統的可視化-

中島 淳志

菌類を学ぶ上で、ある種がどのグループに含まれるかということは、その種の学名や和名を知ることとともに主要な関心事の一つである。しかし、種名の羅列からそれぞれの種の上位分類群の情報を得るには、基本的には一種ずつ文献などにあたって地道に調べるしかなく、特に対象種が多い場合には多大な労力を要していた。また、菌類の分類は日々改訂されており、最新の情報に更新するためにはその都度一から調べ直す必要があった。

本記事では「Mycobank」のWebAPI (Webサービス) と円形系統樹描画ツール「GraPhIAn」(グラフィアン) を組み合わせることにより、任意の種名リストを対象に、いわば「ボタン一つ」で最新の分類階級の情報を取得し、かつ迅速に系統樹の形で可視化することができることを示す。

円形系統樹描画ツール GraPhIAn

GraPhIAn は大規模遺伝子データを扱う「メタゲノミクス」の分野で開発された、分子系統樹データの高品質かつコンパクトな可視化を可能にするツールである [1]。本ツールは2015年にハーバード大学のグループにより発表された後、当該分野では多くの使用例が報告されているが、単純に生物の分類を表現する目的での利用は菌類のみならず、他の生物でも見出されなかった。筆者はこのツールが生物の分類のような階層構造を持つ概念に広く適用可能であることを見出し、今回の

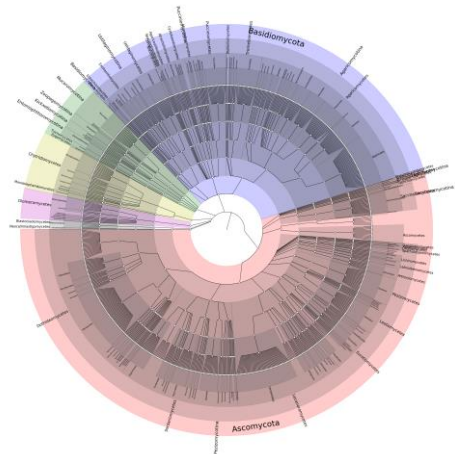


図1: GraPhIAnによるMycobank掲載の全ての「科」の系統樹(変形菌、卵菌、化石菌類、非合法名、シノニム等を除く)

応用を試みた。

GraPhIAnによる円形系統樹の描画例を図1に示す。この図は2017年1月7日にMycobank(詳細は後述)に収載されていた720の「科」とその上位分類群のデータを用いた系統樹である。円の内側から外側に向けて「門」→「亜門」→「綱」…の順に下位の分類階級を表す(末端は種ではなく科や属などでもよい)。上図から子囊菌門に属する科が過半数を占め、担子菌門がそれに続くことなどを一目で読み取ることが可能である。

Mycobankからの分類階級データ取得

菌類の最新の分類は国際菌学会(International Mycological Association,

IMA) が運営するオンラインデータベース「Mycobank」上に順次反映されている(他にも類似のデータベースは複数あるが、本稿では Mycobank のみを取り上げる)。

Mycobank は、収載データを分類群、文献、記載文、標本などいくつかのカテゴリに分けて WebAPI の形で提供している。例えば、

「http://www.mycobank.org/Services/Generic/SearchService.svc/rest/xml?layout=14682616000000161&filter=_id=1」の URL にアクセスすると ID が「1」の分類群 (*Abortiporus biennis*) の最新の分類階級データを取得できる (XML のパースと分類群 ID の照合を要する)。一度に取得できるデータは100件なので、時間がかかるが ID で filter するなどして繰り返しアクセスすることで全件のデータを得られる。データベースは日々更新されているが、クエリ文字列に「last_change_date」のパラメータを付与することにより任意の更新期間のデータを抽出できるので (2017年1月1日以降の場合「layout=14682616000000161」の後ろに「&filter=last_change_date>

20170101000000」を加える)、手元のデータを常に最新の状態に保つことができる。もちろん、データを一括で取得するのではなく、必要な種名のみをその都度データベースに問い合わせるという方法もある。

ちなみに、2017年1月1日から1月29日の期間に更新された分類群データを検索すると、29,844件のデータがヒットした。更新の全てが新規分類群の登録や分類階級の変更というわけではないが、少なくとも1日あたり平均約1,000件程度の更新が行われているということは、菌類の分類がまさに日々変化していることを端的に示す事実であろう。

データを取得した後は、Excel の VLOOKUP 関数など任意の方法で手元の種名リストと照合すればよい (筆者は Python の pandas ライブラリを用いて左外部結合を行っている)。

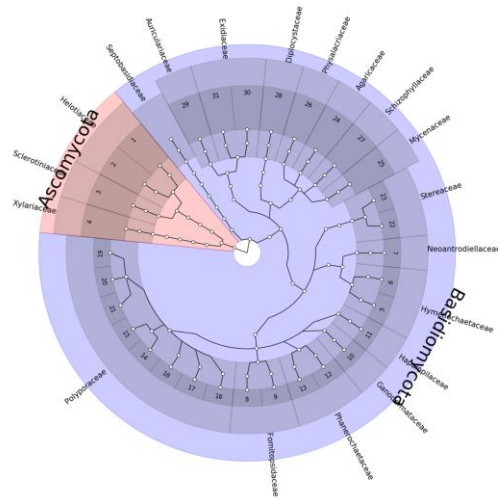


図2: 新治市民の森調査 (2017年1月8日) で採集された32種の系統樹

図2に新治市民の森 (神奈川県横浜市) で採集された32種の種名リストから生成した系統樹を示す。担子菌門と子囊菌門を色分けしているので両者の割合を容易に把握できるほか、サルノコシカケ科 (左下の辺り) が最も多くの種を含むことなども読み取れる (冬なので硬質菌が相対的に多い)。

GraPhlAn の使用手順

GraPhlAnはPython言語で書かれているが、プログラミングスキルを必要とせず、誰もがブラウザ上で容易に使用可能である。必要なのは2つのファイル (分類群リストとアノテーションリスト)、手順は3手順である。以下に最小限の説明を掲載するが、詳細なチュートリアルがオンラインで提供されているので併せて参照されたい (ただし英語のみ)。

<①分類群リストの準備>

新しいtxtファイルを作成し、Mycobankから取得した分類階級データを以下のようにピリオドで繋げる。分類群ごとに改行する。階層数を合わせるために、空白はダミーの文字で埋める必要がある (例では0を使用)。
Fungi. Basidiomycota. Agaricomycotina. Ag

aricomycetes. Agaricomycetidae. Agaricales. O. Amanitaceae

〈②アノテーションリストの準備〉

系統順にどの分類群を表示するか、フォントのサイズや線の太さはどうするか、背景色を加えるか、などの注釈・装飾的事項を2つ目の txt ファイルで指定する。以下の例は「ハラタケ目」のフォントサイズを「8」にする設定である（「―」はタブ区切りを表す）。

Agaricales―annotation_font_size―8

〈③ファイルのアップロード〉

ハーバード大学の Galaxy サーバー (<http://huttenhower.sph.harvard.edu/galaxy>) にアクセスし、左のサイドバーの「Get Data」→「Upload File」を選択して上記2 ファイルをアップロードする。

Get Data

Upload File from your computer

Name	Size	Type
annotatelist_MycoBank_allFamily.txt	26.7 KB	Auto-det...
taxonlist_MycoBank_allFamily.txt	63.4 KB	circl

ファイルをドラッグ&ドロップし、「分類群リスト」のみ「Type」のドロップダウンメニューから「circl」を指定する。右下の「Start」を押すと処理が開始し、完了すると画面右のタスクが灰色から緑色になる。

605: taxonlist_MycoBank_allFamily.txt

604: annotatelist_MycoBank_allFamily.txt

〈④系統樹ファイルの生成〉

次に「GraPhlAn」→「Add rings to the tree」を選択し、ドロップダウンメニューから2 ファイルを指定する（「Input Tree」に分類群、「Ring input File」にアノテーション）。

GraPhlAn

Annotate tree

Add rings to the tree

Input Tree

605: taxonlist_MycoBank_allFamily.txt

Ring input File (loaded as tab-delimited file/text with the gene)

604: annotatelist_MycoBank_allFamily.txt

Execute

「Execute」を選択すると画面右に「Add rings to～」のタスクが追加され、正常に処理されると緑色に変化する。エラーが出たら赤色になるのでファイルの内容を修正する必要がある（重要：日本語など全角文字は使用できないのでエラーが出る）。ちなみに、この手順では phyloXML 形式のファイルを生成しており、「目のマーク」を選択すると中身を確認できる。

606: Add rings to the tree on data 604 and data 605

〈⑤系統樹の描画〉

「GraPhlAn」→「Plot tree」を選択し、「Source」に上記タスクが選択されていることを確認した後、「Execute」ボタンを押せば描画が行われる（画像ファイルの形式や解像度はここで選択できる）。

Source

606: Add rings to the tree on data 604 and data 605

Output data type

png

Set the dpi resolution of the output

150

Plot Size (Inches)

7.0

Execute

暫く待って「Plot tree on data～」のタスクが緑色になった後、「目のマーク」をクリックすれば系統樹が表示される。

607: Plot tree on data 606

データを表示

メタデータ付与

GlaPhlAn では系統樹の特定の部分を装飾したり、外側に輪を重ねたりすることで任意のメタデータを付与することが可能である。それらは全てアノテーションリストで指定することができる。例えば特定のノードに★マークを付与する、量的データを棒グラフで表す、などの表現が可能である。特定の形質を持つ分類群、例えば「傘が円錐形の種」や「胞子紋が桃色の属」などを任意に選んで系統樹上で強調することができる。

その目的では、筆者が作成し、オンラインで配布している、1万件以上の記載文を解析して形質情報を集約した「大菌輪CCCデータセット」を使用することができる。図3は「担子胞子が紡錘形」という形質の対数陽性尤度比 (logpLR) を属ごとに算出して棒グラフで示している (負の値は図示していない)。

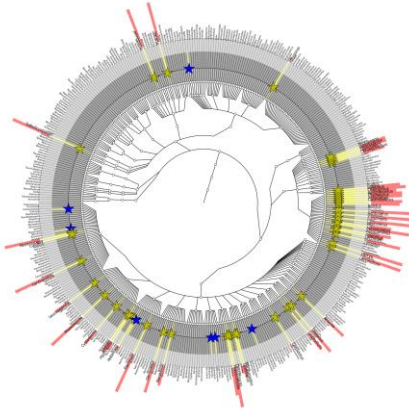


図3: ハラタケ目の属のうち「紡錘形の担子胞子」を有するものを示す系統樹

3時の方向付近に長い棒が偏っているが、これは「イグチ科」の系統である。この科の担子胞子に紡錘形が多い理由について、筆者は今のところ論理的な説明を持ち合わせていないが、生態や進化の観点から考察すれば興味深い知見に結びつくかもしれない。

ある形質が特定の系統に偏っていたとすれば、それはその系統を特徴づける重要な形

質かもしれない。一方、遠縁の系統に散在していたとすれば、それは収斂進化のような何らかの生物学上重要な現象を示している可能性もある。他にも、腹菌化が胞子の形質に与える影響、子実体のサイズと栄養摂取様式の関連など、菌類学における重要なトピックも、この手法を用いれば工夫次第で可視化や仮説生成に結びつけることができるだろう。

おわりに

本稿では種名リストから迅速かつ容易に複雑な系統樹を生成する手法を紹介した。上述の説明はやや学問的な内容に偏ってしまったが、より親しみやすい方面でも応用が考えられる。例えば、日本産の全種を系統樹に図示し、自分の目を見た種をチェックしていくという、鉄道ファンの「路線図塗りつぶし」に通じるような楽しみも考えられる。

テーブルに埋め込んだ巨大なタッチパネルに系統樹を表示して動的に変化させたり、特定の種を選んでその種の解説や画像などを表示させたりする装置のアイデアもあり、筆者はひそかに、その装置とともにある未来の野外観察会の光景を夢想している。

本手法は同定における実用性のみならず、菌類という生物群を個々の種ではなく「系統」のレベルで捉え、学ぶ上でも有効な手段であると考えられる。どの分類群同士が近縁か、という知識は図鑑を眺めていてもある程度身に付けることができるが、より本質的な理解には科ごと、属ごとの「概念 (concept)」の把握が不可欠である。GraPhlAnはその目的に対して親和性の高いツールであり、菌類の深遠な世界の包括的理解に向けて、一步進んだ視点をもたらす可能性を秘めている。

脚注

[1] Asnicar, F. et al., 2015. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ, 3, p.e1029.

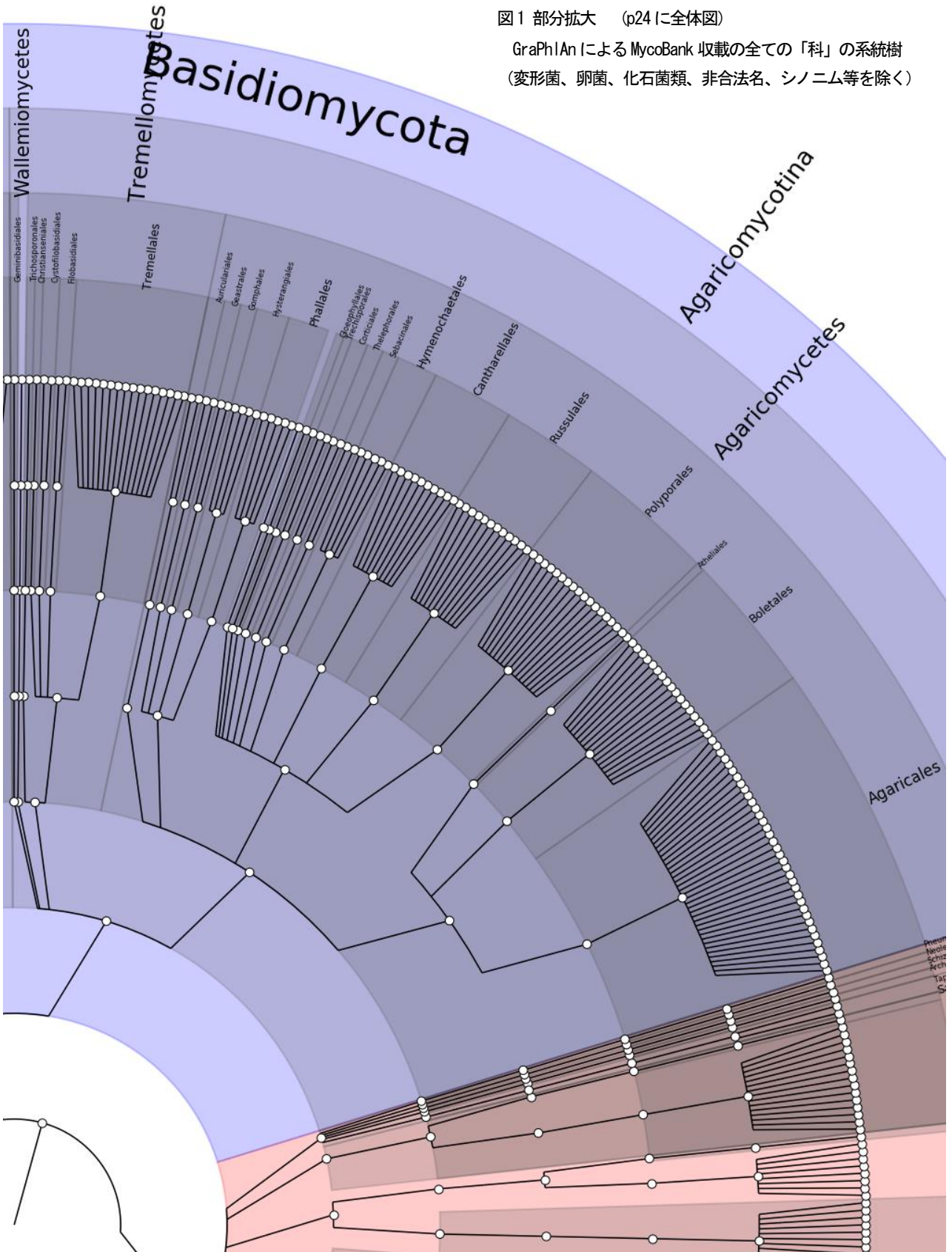


図1 部分拡大 (p24に全体図)

GraPhlAnによる MycoBank 収載の全ての「科」の系統樹
(変形菌、卵菌、化石菌類、非合法名、シノニム等を除く)

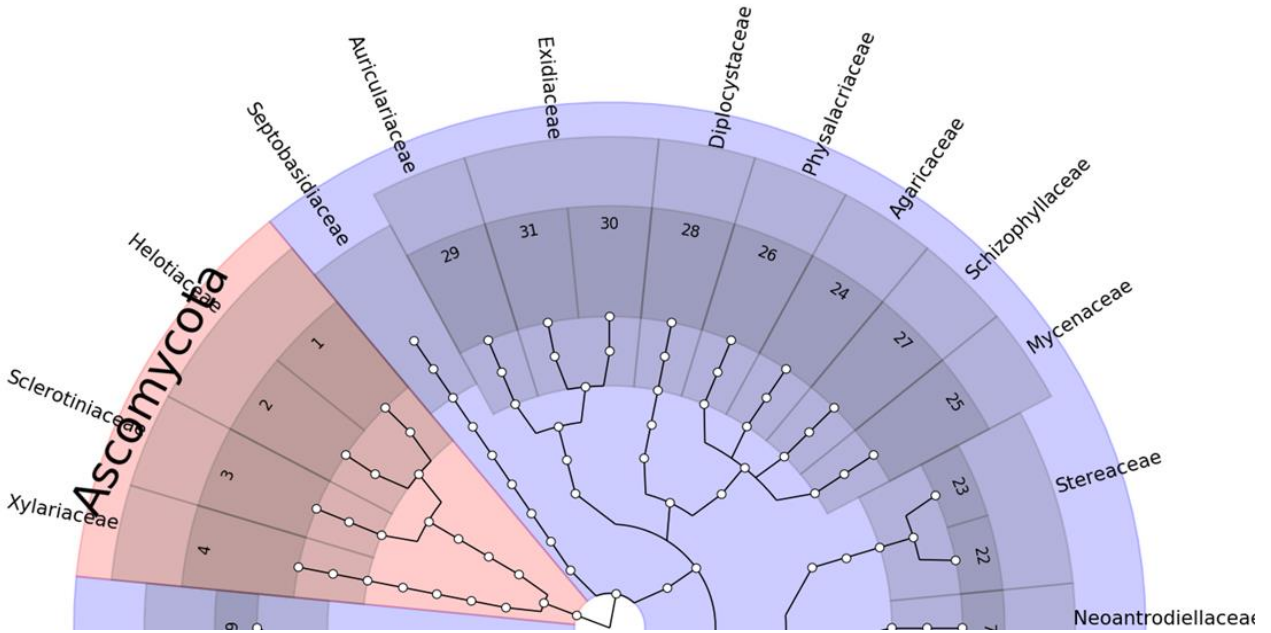


図2 部分拡大：新治市民の森調査（2017年1月8日）で採集された32種の系統樹（p25に全体図）

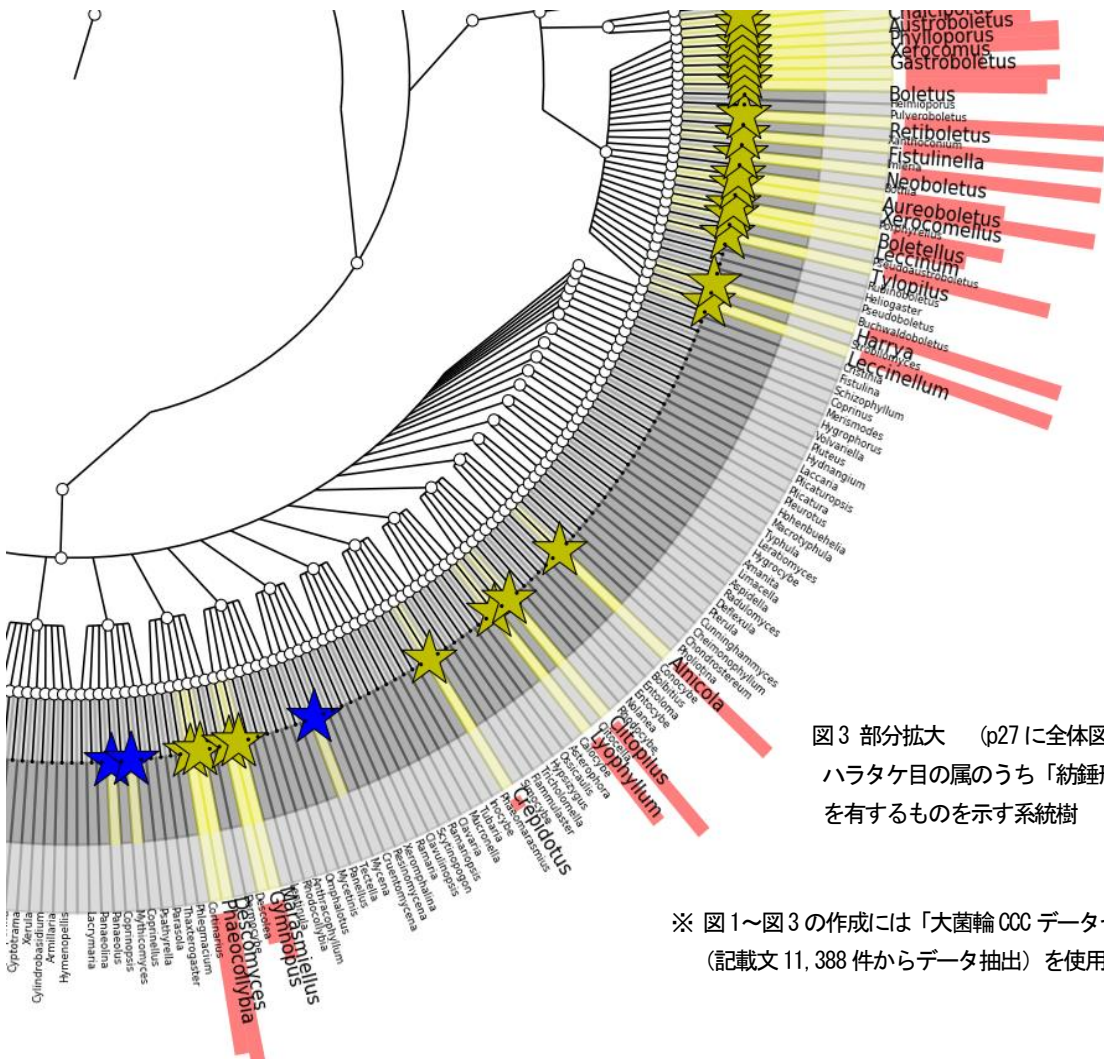


図3 部分拡大（p27に全体図）
ハラタケ目の属のうち「紡錘形の担子孢子」を有するものを示す系統樹

※ 図1～図3の作成には「大菌輪CCGデータセット」ver. 1.1（記載文11,388件からデータ抽出）を使用した。