

菌の輪を描く – GraPhlAn による種名リストの系統的可視化

中島淳志

2017.1.14 千葉菌類談話会スライド会要旨

菌類の分類を学ぶ上では、ある分類群の学名や和名とともに、それが属するグループを知ることも重要である。しかし、種名リストから上位分類の情報を得るためには、従来は一種ごとに文献などを参照するしかなく、特に対象種が増えると多大な労力を要していた。本発表では「MycoBank API」と円形系統樹描画ツール「GraPhlAn」（グラフィラン）を組み合わせることにより、任意の種名リストの分類階級を迅速に系統樹として可視化する手法を紹介する。

GraPhlAn 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 (Asnicar et al., 2015)。筆者はこのツールが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、菌類の分類階級の表現への応用を試みた。

①MycoBank からの分類階級データ取得

オランダの CBS 菌類多様性センターが運営するオンラインデータベース「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 API 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"](http://www.mycobank.org/Services/Generic/SearchService.svc/rest/xml?layout=14682616000000161&filter=_id=)」の URL にアクセスすると ID が「1」の分類群 (*Aboritporus 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。任意の方法で MycoBank データから種名リストの分類群のみを抽出する。

②GraPhlAn

GraPhlAn は Python 言語で書かれているが、そのスキルがなくてもブラウザ上で容易に使用可能である。まず、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「分類群」「アノテーション」の 2 ファイルをアップロードする (記述様式は GraPhlAn の公式サイトで確認していただきたい)。次に「GraPhlAn」→「Add rings to the tree」で 2 ファイルを指定し、新たに生成されたファイルを「Plot tree」で指定する。右の History 項目が全て緑色になったら「Plot tree on data～」のブロックの目のアイコンをクリックすれば系統樹が表示される。

③大菌輪 CCC データセットによる形質マッピング

「アノテーション」ファイルで系統樹の色や線幅などの変更ができるほか、外側に輪を重ねていく形で任意のデータを付与することが可能である。本発表では「大菌輪 CCC データセット」と組み合わせることにより、任意の形質を持つ種のみ「★」マークを付与したり、ある種が持つ形質の一覧 (例えば傘や孢子紋の色) を表示したり、数量データを棒グラフで表示したりすることができることを示す。

参考文献：

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

