



# アマチュアがつくる菌類形質データベース：記載文からの自動形質情報抽出と Web アプリ上でのデータ可視化

Fungal Trait Database Created by an Amateur: Automatic Extraction of Trait Information from Taxonomic Descriptions and a Web App for Data Visualization

中島 淳志\*

Atsushi NAKAJIMA\*

概要：過去 250 年以上にわたって蓄積されてきた菌類分類学の知識は非構造化データとして広範な情報源に散在しており、アマチュアのみならず、専門家にとってもアクセスが容易ではない。菌類の同定には一般的に高度な技術と経験が必要とするが、誰もが有用な形質データを即座に利用可能になれば、同定精度の向上、ひいては菌類の生物多様性に対する理解の深化に繋がるであろう。本稿ではその目的で筆者が長年構築に取り組んできた形質データベース、「大菌輪 Trait Circus」を紹介する。これは、菌類（主に大型菌類）の英語・日本語の記載文から半自動的に抽出した形質情報を、独自のシソーラスに基づき構造化したデータベースである。形質情報の抽出には自然言語処理（構文解析）および機械学習（勾配ブースティング）の手法を利用した。さらに、本データベースを直感的に検索可能にし、菌類に関心のある非専門家に広く利用してもらうために、可視化およびインタラクティブな解析機能を搭載した Web アプリを作成した。Web アプリは以下の URL において、クリエイティブ・コモンズ・ライセンス (CC-BY) のもとで公開している。 <https://daikinrin-trait-circus.lolipop.io/>

Abstract: The knowledge of fungal taxonomy has been accumulated over the past 250 years. However, it is scattered across a wide range of sources as unstructured data and difficult to access not only for amateurs but also for experts. Although identification of fungi generally requires a high level of skill and experience, once trait data are instantly available to everyone, it will not only improve the accuracy of identification but also lead to a deeper understanding of fungal biodiversity. In this paper, I will introduce a fungal trait database, "Daikinrin Trait Circus", which I have been

working on for many years. The database contains semi-automatically extracted trait information from English- and Japanese-descriptions of fungi. The data were structured based on a specialized thesaurus and the methods of natural language processing (parsing) and machine learning (gradient boosting) were used. To make the database intuitively searchable and widely used by non-specialists interested in fungi, a web app with visualization and interactive analysis function was also created. The web app is available under a Creative Commons License (CC-BY).

## 背景と目的

一般市民を含む非専門家にとって、未知の生物との関わり第一歩が名前を知ること、すなわち「同定」だという場面は多いだろう。同定の手段は、図鑑や写真との絵合わせ、検索表の使用、あるいは詳しい人に聞くなど様々だが、近年はいわゆる人工知能の関連技術によって画像認識の精度が向上したことから、その生物群の予備知識が無くても一瞬で同定結果を得ることが可能になりつつある。「Google Lens」、「LeafSnap」、「pl@ntNet」、「Seek by iNaturalist」、国産のものでは「LINNÉ LENS」などに代表されるスマートフォンアプリが多数リリースされており、写真から自動で種を同定するツールは、今後はもはや当たり前の存在となるかもしれない。しかしながら、筆者の私見では、菌類に関しては科・属レベルですら、それらのツールにより満足いく結果が得られる機会は少ないのが現状である。そもそも菌類分類学が未だ  $\alpha$  分類学の段階を脱し得ないのに加え、同定の困難さは信頼性の高い教師データの欠如に繋がり、サイズの小ささや肉眼的形態の単純さ、種内変異の大

きさといった菌類特有の性質も阻害要因としてはたらいっている。また、深層学習（ディープラーニング）の技術に内在する説明可能性（explainability）の欠如という問題もある。

すなわち、尤もらしい候補種が提示されたとしても、種名のみで形質情報が得られなければ近縁種との差異が浮き彫りにならないため、確信度の高い同定には至らない。また、同定結果を検証する上でも、どのような形質を基に同定したのかが明らかでなければ客観性・再現性を担保することができないだろう。一方、経験と直感に基づく「職人芸的な同定」は、自動種同定ツールの発達により「コモディティ化する同定」とは一見して対極に位置しているように見える。しかしながら、そちらもまた往々にして伝統や権威に支えられ、「同定のエビデンスとしての形質データ」が軽視されている以上、客観性・再現性の欠如という批判は同様に当てはまるものである。同定という営為を「科学」の俎上に載せる上では、同定結果には常に形質データが紐づいていなければならない、というのが筆者の考えである。

さて、同定に欠かせないその形質情報はどこから入手可能かという点、250年以上にわたって分類学者を中心に蓄積されてきた膨大なデータが存在するものの、それらは多様な文献に散在しており、必要な情報の検索およびアクセスは容易ではない。加えて、近年は分類学的文献のデジタル化やオープンアクセス化が進み、「ZooKeys」や「MycKeys」のように高度なマークアップを施しているジャーナルも現れたものの、書誌情報やIMRADのタグ付け、生物の学名や遺伝子名などのアノテーション（例：Europe PMCの「SciLite」）が進んでいるのに対して、形質の情報は全く構造化されていないのが普通である。すなわち、記載文情報がセマンティック（機械可読）な形式に変換されておらず、人の目で個々の内容を理解する必要があり、大量のデータの集約や統計解析には膨大な労力を要する。分類群によってはLucidソフトウェアやDELTAシステムを基に構築されたインタラクティブ検索表が存在し、それらは大量の形質データのマトリックスから生成されているが、内部完結的であり、個々の形質データを横断的に検索・再利用可能な仕組みは整備されていない。この状況を打開するには、形質情報の大規模かつオープンなデータベースが必要であると考えられる。

筆者は、その構築に向けた取り組みは大きく2つに分けられると考える。一つは、今後収集する形質情報に体系的・標準的な枠組みを導入する「プロスペクティブ（前方視的）」な方向性である。Dawson et al. (2019) は「菌類の機能的

形質測定の一貫されたプロトコル作成に向けた最初の試み」として、木材腐朽性担子菌類を対象に、測定可能な形質の一覧および具体的な測定方法をハンドブックにまとめたが、このような事例が全菌群に広がれば、分類学・生態学的研究の成果をシームレスに形質データベースの蓄積に繋げていくことができるだろう。もう一つは、これまでに主に記載文の形で公表された形質情報を整理・加工してデータベースに集積する「レトロスペクティブ（後方視的）」な方向性である。しかしながら、日々膨大な情報が飛瀑のごとく押し寄せてくる中で、人の手で過去の文献の一つずつ目を通すのは大海を掬う愚行に等しいだろう。すなわち、記載文を自然言語処理の手法をもって機械可読な形式に変換し、情報抽出の過程を自動化することによりスループットの向上を図るのが現実的な方策といえよう。

既存の形質データベースとしては、Encyclopedia of Life (EoL) プロジェクトのフレームワークに構築された汎生物学的な形質データベース「TraitBank」のほか、植物の「TRY Plant Trait Database」や海洋生物の「EMODnet」等を挙げるができる。菌類学では「funfun」データベース (<https://github.com/traitecoevo/fungaltraits>) という試みがあるが、執筆時点でバージョンは「0.0.3」であり、未だ萌芽期にあると言える。近年は主に群集生態学分野で、従来の操作単位である生物種に代わって「機能的形質 (functional trait)」が注目されていることが形質データベース構築の推進力となっているが（現に funfun データベースも機能的形質の収集を掲げている）、菌類の形質情報に関してはほとんど手つかずと言ってよく、TraitBankにもGBIF等の外部データに由来する「地理的分布」や「宿主」を除いては、ほぼ全くデータが登録されていないのが現状である（中島、2019）。

記載文からの形質情報抽出は、少なくとも菌類学分野に関しては、筆者が渉猟した限りでは一例も見出されなかったが、生物全体で見ると新しいアイデアではなく、Cui (2010) によると30年以上前から取り組まれてきた。Cuiは植物の記載文に対して自然言語処理（機械学習）によるセマンティックアノテーションを検討し、80-90%以上の精度を記録した。しかしながら、生物のグループによって記載文の構成や専門用語は大きく異なるため、菌類に対してその手法をそのまま適用することはできず、菌類学のドメイン知識に基づく固有のオントロジー設計と、それを基にした分類器の構築が必須である。

本稿では筆者が過去8年間にわたって構築してきた形質

データベース、「大菌輪 Trait Circus」の概略を紹介する。その構築にあたっては、まず菌類学のドメイン知識に基づいて独自のシソーラスを作成し、さらに自然言語処理（構文解析）および機械学習（勾配ブースティング）の手法を利用することで、英語・日本語の記載文からの半自動的な形質情報抽出を可能にした。また、このデータベースを基に、分類階級を反映した階層的可視化およびインタラクティブな解析機能を搭載した Web アプリを作成した。なお、本データベースの第一の目的は、非専門家による同定支援に有用な形質情報の提供である。そのため、専門家の使用に堪えるクオリティを目標としているわけではなく、シソーラスの設計などにおいては学問上の厳密性よりも、非専門家目線での直感的な分かりやすさを重視していることを初めに断っておく。

## データベースの構成

本データベースにおける形質データは、主に「要素 (element)」「属性 (attribute)」「値 (value)」の3つ組で構成されている。このうち、情報源に実際に含まれる情報は「要素」と「値」であり、基本的に原文のまま抽出する。「属性」は「要素」と「値」の組み合わせに対応するカテゴリである。例えば「Lamellae pale pinkish. (髷は淡桃色)」の要素は「lamellae」、値は「pale pinkish」である。属性は意味を踏まえて「color」を付与する。属性は要素と値の組み合わせを基に自動で付与し（詳細は後述）、適宜手動で修正している。属性の種類には「amount (数量)」「amyloidity (アミロイド性)」「color (色)」「development (発達)」「habitat (生息環境)」「position (位置)」「presence (有無)」「reaction (呈色反応)」「shape (形状)」「size (サイズ)」「structure (構造)」「surface (表面性状)」「taste/odor (味・臭い)」「texture (質感)」の14種類を設けている。属性未特定の情報是最終データには反映されないが、ノイズを減らすため、無理にいずれかに当てはめることはしていない。

「要素\_属性\_値」が揃ったデータに対し、「上位語辞書」と「同義語辞書」の2種類の辞書を用いて自動照合を行う。例えば「子実層托」は「髷」や「管孔」の上位語であり、「胞子」は「担子胞子」「子嚢胞子」「分生子」…などの上位語である。ある要素に上位語が存在する場合、上位語でも照合を行う（ただし、例えば「子実体」は「傘」の上位語だが、「傘の色=子実体の色」は問題ないとしても、「傘の形状=子実体の形状」とは言えない場合もあるので、個別のルールを適用する）。一方、同義語の例としては「pileus (傘)」に対す

る「cap」や複数形の「pilei」などを挙げることができる。「cap」や「pilei」が検出された場合、それらを「pileus」として扱う。「要素\_属性\_値」の「有効な組み合わせ」は事前に設定しておき、それに当てはまらないものは採用されない。例えば、「胞子\_形状\_円錐形」という組み合わせでヒットしたとしても、それは有効な組み合わせのリストにないので不採択となる（実際、この組み合わせは「胞子の形状」ではなく「胞子の装飾の形状」を指していると考えるのが尤もらしく、これを採用するとノイズになる可能性が高い）。このようにドメイン知識を踏まえて用語を適切に統制 (control) することで、単純な機械的照合よりも高精度の情報抽出が可能である。例えば「lemon」という単語は「傘」や「柄」などが要素であれば「レモン色」、「胞子」が要素であれば「レモン形」を指す場合が多い。「hazel」も生息環境という文脈では、「黄褐色」ではなく「ハシバミの木」の方が尤もらしいであろう。

## 記載文データの取得と構文解析

菌類では動植物ほどデジタル化された記載文の集積が進んでおらず、地衣類では「CNALH (Consortium of North American Lichen Herbaria)」が、若干の構造化もなされている有用なデータベースであるが、菌類全体をカバーする利用可能な情報源はほぼ「MycoBank」に限定される。MycoBank にない情報は現在のところ、個々の文献の PDF ファイルなどから収集する必要がある。同一の種であっても記載文の内容はほとんど重複していないという調査結果もあり (Lydon et al., 2003)、とにかく多様な情報源から豊富なデータを取得することが肝要である。しかしながら、日本語の情報源は一層乏しく、本データベースに含まれる情報の大半が紙媒体（スキャナー→OCR）由来である。残念ながら、現在のOCR精度では無謬のデジタル化には程遠く、MycoBankのデータも様々な言語が混じっていたり、記載文以外の情報が混入しているなど、完全な自動化を妨げる要因は多々存在する。クラウドソーシングによる手作業での修正なども提案されているが (Burleigh et al., 2013)、この段階は将来的に最大のボトルネックになりうるので、改善を検討すべきだろう。

収集した記載文データは機械可読な形式への変換が必要である。菌類の記載文は構造化データではないが、一方で完全な自然言語でもない。例外も多いものの、例えば「Pileus 5-10 cm broad, convex to campanulate, smooth, white...」のように、冠詞や動詞が基本的に用いられず、カンマ区切り

で主語を形容詞（句）が後置修飾する形式が一般的である。この構文は自然文のそれに比して単純かつ形式的であり、解析は筆者の経験から言えば比較的容易である。一方、標準的な英語の構文から逸脱しているため、既存の自然言語解析ツールの適用が困難であることも指摘されている (Cui, 2010)。

記載文からの情報抽出には条件付き確率場や LSTM のように系列ラベリング問題を解く手法も試みられているが、筆者が試した限りでは古典的な構文解析器（パーサー）で十分実用に堪える結果が得られた。具体的には、「Natural Language Toolkit (NLTK)」の正規表現パーサー (nlk.RegexpParser) を利用して「要素」と「値」が対になるようにフレーズチャンクを抽出している。例えば、要素の抽出には「{^<JJ>\*<NNINNSINNPINNPISCCITOI\\$\>+<ININNNINPNNSICDIJ>\*}」、値の抽出には「{<CCIDITPDTIRBRIPR\\$\IRBITOI\MDIWRBIJJJISIJRIINIRBIRBSIVBIVBPVIVBGIVBNIVBDI\(\I)ININNSINNPINNP\>+}」という正規表現パターンを用いている。記載文特有の工夫としては、事前に要素の一覧を作っておき、そのいずれかがヒットした場合、主語となる要素の直後に be 動詞を挿入し、文法を通常の英文に近づけることで精度の向上が認められた。日本語の記載文については、spaCy と GiNZA を使用して得られた品詞情報を基に、英語とほぼ同様の方法で要素と値の組み合わせを抽出している。

ここまでの過程で生の記載文を「要素」と「値」の対に

分けることができたが、その各々に対して以下の手順で「属性」を決定している。まず、既知の要素と値の組み合わせであれば、既にそれに付与されている属性を自動で採用する。未知の場合には、マルチクラス分類器に要素と値を入力し、所属確率の最も高い属性を出力する。ただし、この属性をそのまま採用するのではなく、あらかじめ個々の属性に対応するキーワード一覧を辞書として作成しておき、そのいずれかを含む場合のみ自動採用することで、分類精度の向上を図っている。例えば、「cylindrical sometimes curved」を分類器に入力すると「形状 (shape)」という属性が出力される。辞書には「形状」に対応するキーワードとして「cylindric」や「curv」が登録されているので、実際にそれが採用されることになる。誤分類により他の属性が出力されたとしても、それらに対応するキーワードが存在しなければ採用されず、最終データには反映されない。キーワード辞書は適宜見直しを行っているので、不採用のデータも将来的には属性が付与される可能性はある。

なお、分類器は Python の勾配ブースティングライブラリ「CatBoost」、とりわけ 2020 年 3 月 (v.0.22) から追加された「text\_features」オプションを主に活用して作成した。2020 年 6 月 21 日時点でデータ数（値の重複除去後）は「248,748 件」であり、クラス別の件数では「size」が 62,193 件と最多で、「shape」の 44,359 件、「color」の 42,988 件がそれに続いた。8 割を学習データ、2 割をテストデータとし、「要素 - 値」の組み合わせから属性を予測するマルチクラス分類モデル

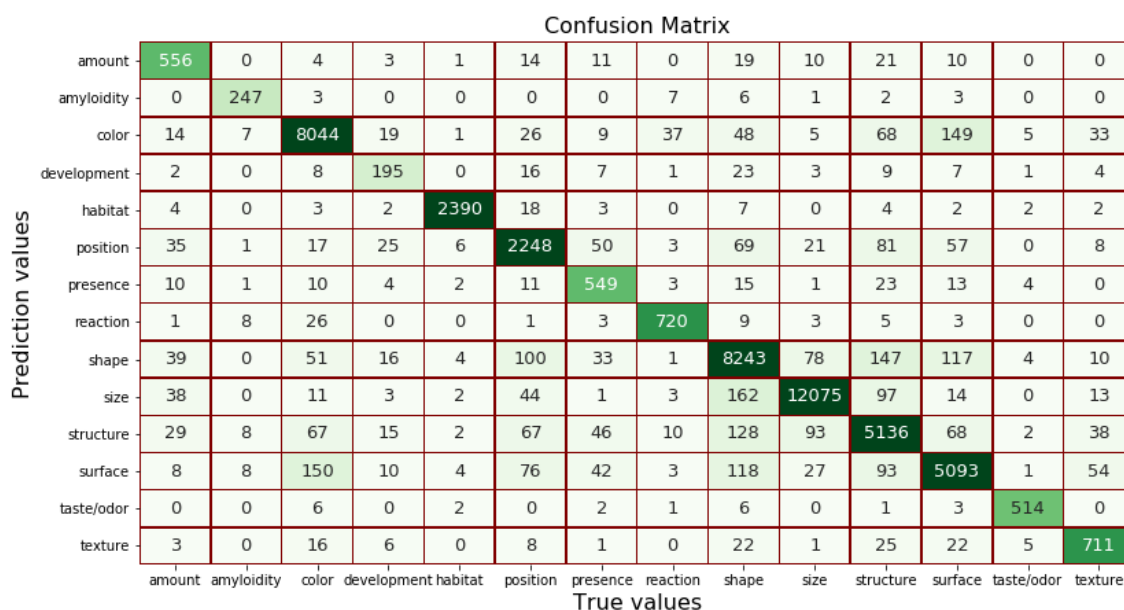


図 1：テストデータの混同行列。

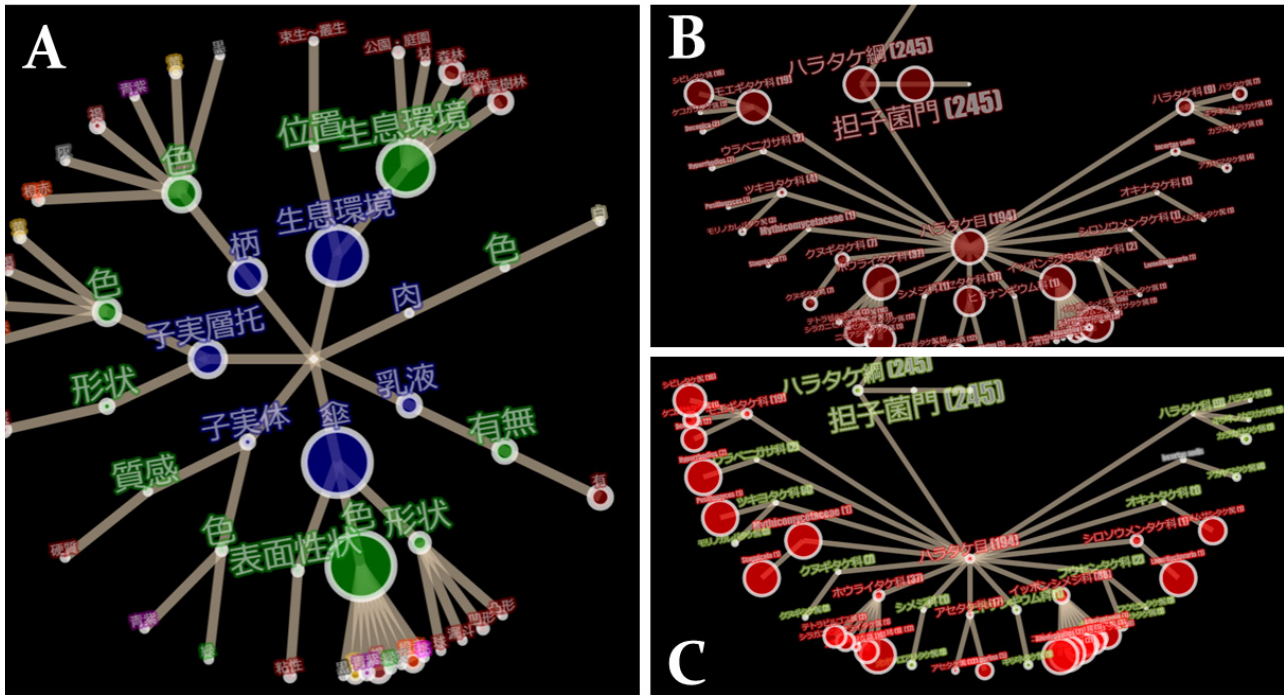


図 2：Web アプリでの樹形図表示。(A) taxon モードで「ハツタケ」を選択した状態。(B) trait モードで「傘\_形状\_乳頭」を選択した状態。(C) B の条件でノードサイズに陽性尤度比を反映させた状態。

を作成した。テストデータに対する正確度 (accuracy) は 0.93 であった (図 1)。

以上の手順により、記載文 1 件を平均 1-2 秒程度で処理することが可能になった。现阶段では、その後に人手によるチェックおよび修正を要するものの、1 時間に数千件のオーダーでデータベースに追加することができる。当初はほぼ目視と手入力で作業していたこともあり、データ数が 30 万件に達するまでには 7 年間の歳月を要したが、本手法を適用したところ、約 1 ヶ月でデータ数は 30 万件から 40 万件に増加し、飛躍的なスループットの向上が達成された。

## 可視化・Web アプリ作成

図鑑に代表される非専門家向け情報源においては、網羅性、正確性などの「情報の質」とともに、検索性や可読性、視覚的審美性といった「使いやすさ」も重要な要素である。また、本データベースに集積される大量の形質データには潜在的に有用な傾向や相関が含まれる可能性があるため、その価値を最大化するためにも、非専門家が容易にデータマイニングを実施できる機能が必要だと考えた。それらの要求を満たすために、本データベースを Web アプリ (ダッシュボード) の形で可視化することを試みた。

Web アプリは Plotly の「Dash」フレームワークを用いて

作成した。このアプリには大きく「taxon モード」と「trait モード」の 2 つのモードを用意し、前者では任意の分類群を選択すると、それが持つ形質の一覧を表示する一方、後者では任意の形質を選択すると、それを持つ分類群の一覧を表示するようにした。同定の際には、主に「trait モード」で複数の形質を順次選択し、候補分類群を絞り込んでいく使用法を想定している。形質を選択すると、各形質の選択肢にはそれを持つ残りの分類群の数が表示されるので、その数ができるだけ小さくなるような形質を選択すればよい。和名データのある分類群のみに限定した検索も可能である。

両モードとも、「樹形図表示」ボタンを押すと、現在選択しているデータを円形樹形図の形で視覚化することができる (図 2)。「taxon モード」の樹形図は中心から周縁に向かって「要素→属性→値」、「trait モード」は分類階級が上位ほど中心に近くなっている。なお、グラフネットワークにおいては各形質または分類群を表している円のことを「ノード」、それらを繋ぐ線のことを「エッジ」という。ノードをクリックすると詳細情報が表示され、Google 画像検索や後述の「解析モード」に移行可能である。「trait モード」では選択したノードを中心にしたり、逆に一つ上の分類階級を中心にしたりして新たな樹形図を描画することができる。

ノードの大きさはデータ数を表すが、これを用いると単

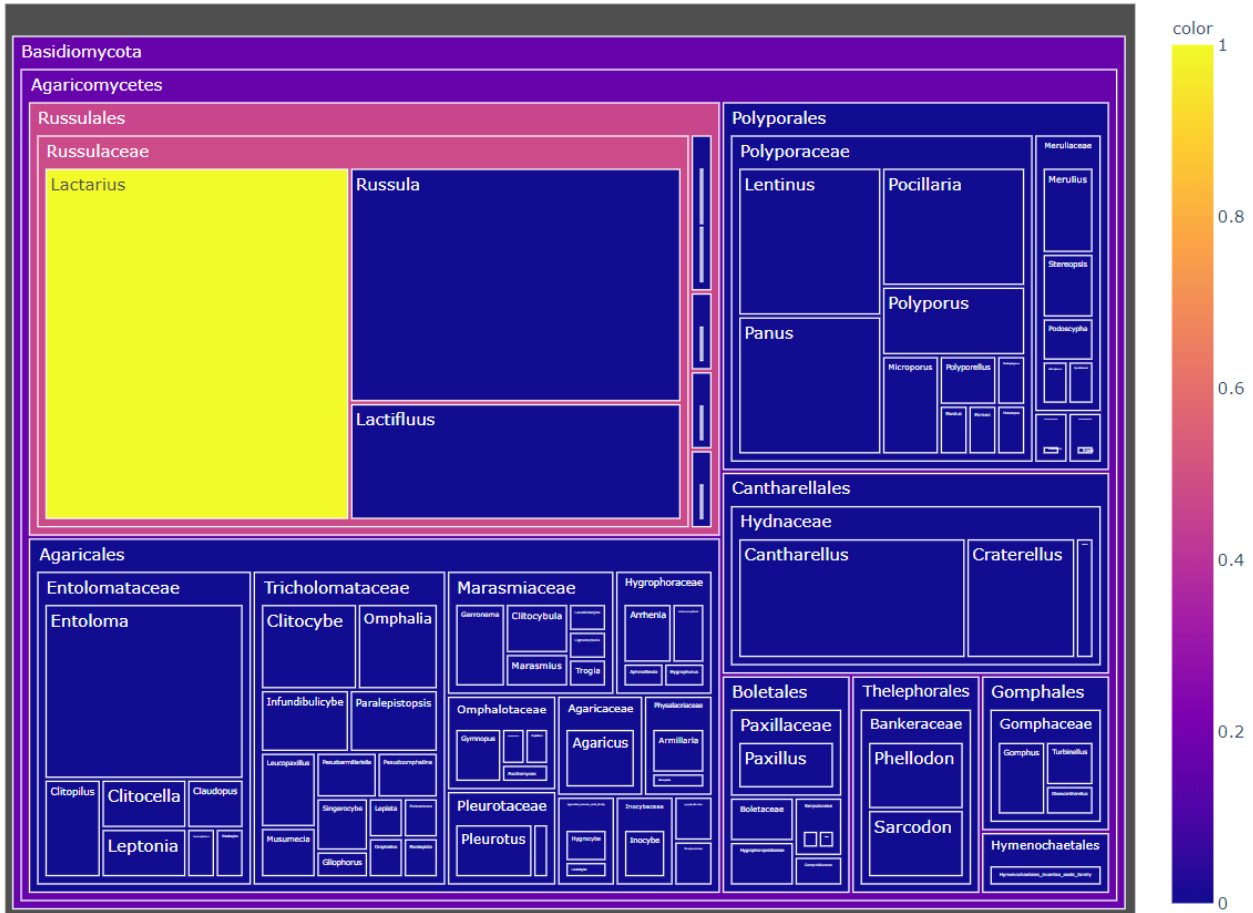


図3 解析モード (ヒートマップ)。「カラハツタケ属」で絞り込んだ状態で「傘\_形状\_漏斗」を選択。

純にデータ数が多い分類群が強調され、データ数の少ない分類群が埋もれてしまいがちになるという欠点がある。「traitモード」では、ノードサイズに陽性尤度比 (pLR) を反映させることで、その分類群に特徴的な形質をより浮き彫りにすることを可能にしている。pLRは、「その菌群であるならその形質を持つ」という確信度を指す「感度」、「その菌群ではないならその形質を持たない」という確信度を指す「特異度」の両方を加味した指標である。統計学的有意 (95% 信頼区間下限が1を超える) な pLR に対応するノードは赤色に変えて目立つように工夫した。

また、対話的なデータマイニングを実現するための「解析モード」として、「ツリーマップ」と「パラレルダイアグラム」の2つの手段を用意した。「ツリーマップ」は、任意の形質で絞り込んだ際の分類群データ (taxonモード)、あるいは任意の分類群で絞り込んだ際の形質データ (traitモード) をそれぞれ階層的に可視化するものであり、選択した分類群/形質が全体のどれほどの割合を占めているかを一目で把握することができる。例えば、「傘\_形状\_漏斗」で絞り込んでツリーマップを作成すると、「ベニタケ目」「ハラ

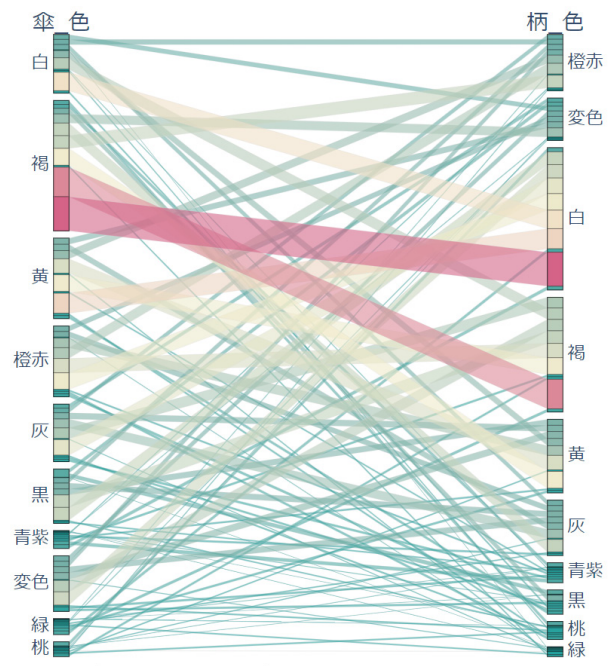


図4:解析モード(パラレルダイアグラム)。「傘\_形状\_中高」で絞り込んだ状態で「傘\_色」と「柄\_色」を選択。

タケ目」「タマチョレイタケ目」の順にこの形質を持つ分類群が多く、ベニタケ目の中では「カラハツタケ属」が「ベニタケ属」よりも若干多いことなどが見て取れる（図3）。「パラレルダイアグラム」では、ツリーマップと同様に形質または分類群で絞り込んだのち、任意の2つの「要素-属性」を選択することで、値の組み合わせを可視化することができる。例えば、「傘\_形状\_中高」で絞り込み、「傘\_色」と「柄\_色」を選択すると、「傘\_色\_褐」と「柄\_色\_白」の組み合わせが最多で、「傘\_色\_褐」と「柄\_色\_褐」がそれに続くことなどが見て取れる（図4）。本データベースには生息環境や基質の情報も収載しているため、例えば広葉樹では針葉樹に比べて、生息する菌類の胞子が大きく、形状も細長いものが多いことが知られているように (Purhonen et al., 2019)、特定の形質と環境要因の関係などを探索的に検討することで、菌類の生態や進化に関する未知の発見にも結びつくかもしれない。

## 今後の展望

本データベースは長年の作業を経てある程度形になったものの、未だ発展途上であり、当然要求される機能のいくつかを欠いた状態である。例えば、「傘の直径」や「襲の枚数」、「担子器・子嚢の胞子数」といった「数量 (numerical) データ」はほぼ全く抽出できていない。これは、特に計測値を表す文字列の OCR 精度が悪く、表記方法も多様であるためパースが困難であることに起因する。今後それらを利用できれば、散布図やヴァイオリンプロット、ヒストグラムなどの可視化や解析の幅も広がるだろう。菌類の子実体のサイズは一般的に変異が大きく、記載文における計測値の確度を疑問視する向きもあるが (Halbwachs and Karasch, 2019)、例えば微小菌類では胞子などのサイズの僅かな差異で種が分けられている例も多く、同定を目的とする以上、数量データの処理は避けて通れない道であろう。

また、本データベースは未だ菌界全体のカバーには至っておらず、特に記載文の独自性が高い地衣類や酵母は全く未着手である。特に後者の記載文については根本的な構造的差異があり、むしろ細菌や原生生物のそれに類似しているため、全く別のシステムを構築する必要があるだろう。加えて、本データベースで使用しているシソーラスはあくまで非専門家である著者の目線で設計されたものであり、既存のオントロジーとの互換性は考慮していない。今後拡張性を向上させ、GBIF など他の生物多様性データベースとのマッシュアップを実現させるためには、他の生物におけ

る形質データベースを参考に標準化を進めていかなければならないだろう。なお、記載文を唯一の情報源とすることによるバイアスにも留意する必要がある。記載文に従来記録されてきた形質は、生物が持つ無数の形質のうち「古典的」なものにほぼ限られているからである。今後は3D スキャナ、ハイパースペクトルカメラ、CT、MRI といった先端機器の適用拡大や廉価化により観測手段の多様化が進むと考えられるため、それらの「新しい」形質を柔軟に取り入れる基盤の構築を意識する必要がある。本データベースの構築には機械学習の手法を用いたが、自然言語のみならず、写真や動画からのマルチモーダルな形質情報抽出も見据えていかなければならない。

ところで、本稿は「アマチュアがつくる菌類形質データベース」と銘打ったが、率直に言えば、アマチュアがこのような取り組みを行うことには何ら優位性を見出せない。せいぜいプリミティブな「非専門家の目線」を反映させられるといった程度であろう。つまり、これは本来専門家が先導してなすべき仕事である。現に、米国国立医学図書館の「MEDLINE」のような学術文献のデータベース、あるいは遺伝子・タンパク質などのデータベースには専任のインデクサーやバイオキュレーターが従事しており、筆者が一銭も儲けていないのとは対照的に、相応の対価が支払われていることだろう。しかしながら、分類学を取り巻く厳しい人材不足・予算不足の現状において、形質データベースの構築のような地道な事業は評価されにくく、専門家の有限のリソースをそれに割くことは現実的でないのも確かである (Deans et al., 2012)。ただし、いかに機械学習による効率化を達成したとしても、学術文献由来のアノテーションには依然として多大な時間と労力を要し、かつ専門性の高さがクオリティを直接左右することには変わりはない。今後は生物多様性市民科学の入り口としての「同定」の重要性を強調するとともに、例えば日本語で書かれた分類学的情報を形質データベース作成の過程で「翻訳」することにより世界に発信できることなど、多面的な価値を具体的な成果とともに示していく必要があるだろう。

## 謝辞

本データベースの内容のうち、日本語文献由来のデータの一部は神奈川キノコの会の有志により入力されたものである。また、杉本泉氏には Web アプリの使用感やカビの形質について貴重なご意見を頂いた。ここに厚く感謝申し上げます。

## 引用文献

- Burleigh, J.G., Alphonse K., Alverson A.J., Bik H.M., Blank C., Cirranello A.L., Cui, H., Daly, M., Dietterich, T.G., Gasparich, G., Irvine, J., Julius, M., Kaufman, S., Law, E., Liu, J., Moore, L., O'Leary, M.A., Passarotti, M., Ranade, S., Simmons, N.B., Stevenson, D.W., Thacker, R.W., Theriot, E.C., Todorovic, S., Velazco, P.M., Walls, R.L., Wolfe, J.M., Yu, M. 2013. Next-generation phenomics for the Tree of Life. PLOS Currents Tree of Life. 2013 Jun 26 . Edition 1. doi: 10.1371/currents.tol.085c713acafc8711b2ff7010a4b03733.
- Cui, H. 2010. Semantic annotation of morphological descriptions: an overall strategy. BMC Bioinformatics 11: 278.
- Dawson, S.K. et al., 2019. Handbook for the measurement of macrofungal functional traits: A start with basidiomycete wood fungi. Functional Ecology 33: 372-387. <https://doi.org/10.1111/1365-2435.13239>
- Deans, A.R., Yoder, M.J. and Balhoff, J.P. 2012. Time to change how we describe biodiversity. Trends in Ecology & Evolution 27(2): 78-84.
- Halbwachs, H. and Karasch, P. 2019. Dimensionen: Ein Blick in die unsichere Welt der Pilzmorphologie. Zeitschrift für Mykologie 85(1): 93-108.
- Lydon, S.J. Wood, M.N.G., Huxley, R. and Suttonet, D. 2003. Data patterns in multiple botanical descriptions: Implications for automatic processing of legacy data. Systematics and Biodiversity 2: 151-157.
- 中島淳志 . 2019. 菌類の形質を掌握せよ ～形質データの構造化・統制・集積～ . 千葉菌類談話会通信 35: 42-49.
- Purhonen, J. et al. 2019. Morphological traits predict host-tree specialization in wood-inhabiting fungal communities. Fungal Ecology 46: 100863