

日本動物分類学会・生物情報広域データベース化研究グループ
ジョイントワークショップ
生物分類データベースの現状と展望

2001 年 10 月 6 日 18:00 - 20:30 九州産業大学 G 会場
責任者：多田内 修（九州大学大学院農学研究院昆虫学分野）

趣 旨

OECD により提案された地球規模生物多様性情報機構(GBIF)の一環として、日本では 2002 年度より国内体制整備が本格的に始まります。今後 10 年間で、公的資金が投入されて生物データベースのてこ入れが行われることになる予定です。その中で、生物多様性を記載するうえにおいて欠くことのできない「分類」或いはその基盤ともなる「標本」は、基幹プロジェクトとしてデータベース化が進められます。それだけでなく、従来ボランティア活動に頼ってきたさまざまな生物データベースに対する公的なサポート、またデータベース構築のための人材育成も計画に組み込まれています。さて、10 年後に果たして日本は国際社会の中で、どれだけその役割を果たしたと評価されるのでしょうか？データベースの構築では、ゲノムプロジェクトとは性格が異なり、データを持つ各々の研究者の協力が必要になります。我々、動物分類学の関係者は、どのようにその計画にかかわっていくことができるのでしょうか？本会では、現在進行中の分類データベースのケーススタディと、生物データベース構築におけるノウハウの紹介を行います。

プログラム

座長	馬渡峻輔（北大・理学部）
司会	多田内 修（九大大学院・農学研究院）
18:05 - 18:25	GBIF の紹介と生物多様性データベースのあり方について 月井雄二（法政大・自然科学センター・生物）
18:25 - 18:45	日本ショウジョウバエデータベース（JDD）について 山本 雅敏（京都工繊大・ショウジョウバエ遺伝資源センター）
18:45 - 19:05	日本分類学データベース（DBTJ）の概要 酒井勝司（四国大・ETI-J）
19:05 - 19:25	昆虫学データベース（KONCHU）について 多田内 修（九大大学院・農学研究院）
19:25 - 19:45	分類以外の切り口で多様性をとらえるデータベースの可能性について 木原 章（法政大・自然科学センター・生物）
19:45 - 20:05	生物関連データベースのサポート 鵜川 義弘（宮城教育大学・環境教育実践研究センター）
司会	木原 章（法政大・自然科学センター・生物）
20:05 - 20:30	総合討論

発 表 要 旨

1. GBIF の紹介と生物多様性データベースのあり方について

月井雄二（法政大・自然科学センター・生物）

GBIF（Global Biodiversity Information Facility；地球規模生物多様性情報機構）は、OECD のメガサイエンスフォーラムで数年間検討された後、今年の3月に正式発足した国際組織です。その目的は、近年、国際的な関心を集めている生物多様性情報をデジタル化し、それらに誰もが自由にアクセスして利用できるシステムを地球規模で構築することにあります。ただし、GBIF そのものは、その目的を実現するための国際的な調整組織として位置付けられ、多様性データのデジタル化、ネットワークでの公開等は、各国が独自に目標を掲げて行うこととされています。

具体的には、生物標本のデジタル化、既知の生物名のカタログ作りを基礎とし、これらを世界全体で連携するためのシステム作り、またそれらを統合的に検索できるための検索エンジンの開発などが当面の目標となっています。さらに、GBIF は、すでに存在するゲノムデータベースや生態学関連の情報とも連携をはかり、生物多様性条約関連の活動とも連携することを目指しています（<http://www.gbif.org/>）。

日本国内では、GBIF の発足に伴い、その対応を検討するための組織として研究者および関係省庁の役人からなる GBIF 技術専門委員会が設置されました（設置場所：科学技術振興事業団）。この委員会では、すでに御存知かと思いますが、これまでに国内にある生物多様性関連のデータベースの実態調査を行っています。また、今後は、国内の生物多様性データベースを充実させるための様々なプロジェクトを実施していく予定です。

とはいえ、GBIF は発足したばかりですので、具体的な成果はまだ形には現れていません。このため、多くの研究者は、GBIF とは一体何なのか、自分たちの研究にどのような関連／影響をもつのかまったく想像もつかない状態なのではないでしょうか。

これは GBIF に直接関わっている者でさえ、具体的な成果があがるまでは、実際にどのようなものになるか予想がつかないのが現状ですので、当然といえば当然です。とはいえ、はっきりしていることは、「趣旨」にも書かれているように、生物多様性データベースは、研究者自身が積極的に関わっていかないと、他に誰もそれらを替わりに作ってくれる人はいない、ということです。

周知のとおり、ゲノムデータベースはいまや遺伝学ばかりではなく様々な研究分野で必要不可欠の存在となっています。しかし、データベース化が必要なのはゲノム情報だけに限りません。ゲノム研究よりもはるか以前から存在し、膨大な情報が蓄積されている生物多様性情報もデータベース化が待たれる分野といえます。

しかも、地球上にはこれまでに記載されている生物種を遙かに上回る膨大な数の生物種がいることがわかっています。また、その膨大な生物多様性情報を環境保護などの目的で皆が利用できるようにすることが社会的な要請にもなっています。このような膨大な情報を有効に活用するためには情報処理技術の導入は必要不可欠です。したがって、今後、それに併せて生物多様性研究のスタイルそのものも変えるべきであり、変わっていかざるを得ないでしょう。

しかし、データベースやネットワークといったことには関心がない、あるいは、必要性は理解できても、自分ではそのようなことは難しくて手が出せない、という方も多いでしょう。さらには、貴重な研究の時間を削ってまでそのような活動に協力する意味／意義があるのか、と訝る方もいるかも知れません。

たしかにデータベースというと一般には検索やソートなど様々な機能があって、それらの操作を覚えるだけでも大変なものが多いようです。そのような「データベース」を自分で作るなんてとんでもない、ことは間違いありません。しかし、ネットワーク上で公開されるデータベースは、基本的に「機能」と「データ」の両方ともに分散型のデータベースなのです。したがって、そこでは、自分でデータベースに必要なものすべてを準備する必要はありません。ある者はデータを提供し、ある者はそれを検索するインデックスを作る、という作業をネットワークを介して独自に行うことが可能です。これはいわゆる「検索エンジン」を考えてもらえば理解しやすいと思います。インターネットの世界では、検索エンジンの登場によっ

て、それまで世界中に散在していた様々な情報が、あたかもひとつのデータベースのように利用可能になったことは御存知でしょう。

これと同じことがGBIFにおける生物多様性データベースでも実現されると予想されます。したがって、各研究者がやるべきことは、各自がもつ多様性情報をデジタル化し、ネットワークで公開することだけで良いのです。コンピュータに自信のある人なら自分でそれらのデータを検索したりソートしたりする機能を追加してもよいでしょうが、あえてそこまでやる必要はありません。後は人まかせ、検索エンジンまかせにできるのが、ネットワークデータベースの強みといえます。

ただし、学術情報には一般の情報にはない重要な特徴があります。それは、情報が永続的に保存される必要があることと、情報の網羅的、統合的な検索ができなければならない、ということです。これらの特徴については図書館にある書誌情報を考えていただければよいと思います。情報の永続的な保存の必要性については、今回は省きますが、情報の網羅的かつ統合的な検索については、学術分野のみならず、現在、インターネットの世界全体で取り組むべき重要な課題となっています。

そのための対応策として、ネットワークで公開する情報に「メタデータ」を添付しよう、という呼び掛けが行われています (<http://dublincore.org/>; <http://www.diffuse.org/>)。この「メタデータ」の添付は学術研究関連の Website においては今後必要不可欠のものとなるはずであり、多様性情報をデータベース化する上でも避けては通れないテーマです。時間があれば、これについても紹介したいと思います。

20世紀の最後の四半世紀はゲノムデータベースの時代でしたが 21世紀は生物多様性データベースの時代になる、のかも知れません。

2. 日本ショウジョウバエデータベース (JDD) について

山本 雅敏 (京都工繊大・ショウジョウバエ遺伝資源センター)

ここで紹介します日本ショウジョウバエデータベース (Japan Drosophila Database, JDD) は下記の4つのデータで構成されている。画像データをできるだけ多く盛り込んだ、遺伝子、系統、種名などにより閲覧、検索可能なデータベースである。

国内で維持されているキイロショウジョウバエ系統とその特性に関するデータ
系統情報として系統名、遺伝子型、形質等の遺伝的特性に加え系統の所在等の情報

キイロショウジョウバエの突然変異形質の画像データ

キイロショウジョウバエ以外のショウジョウバエ種の画像データ

国内ショウジョウバエ関係論文類の収録

キイロショウジョウバエ (以下ショウジョウバエとする) は遺伝学研究の長い歴史の中で他の生物と比較できないほど膨大な突然変異、染色体異常、トランスポゾン挿入、改変遺伝子導入系統が作成された。それらの系統は、開発を行った研究室や系統保存機関で維持され、研究者の要請に応じて自由に分譲されてきた。20世紀末のゲノム解析研究の結果、ショウジョウバエとヒトなど多くの生物の遺伝子は高い相同性を示すことが明らかにされ、ショウジョウバエの有用性がさらに高まった。その結果、ショウジョウバエ研究の専門家だけでなく、モデル生物として研究に応用しようとする研究者が増加し、研究に必要な系統の所在と遺伝的特徴を迅速に調査できるデータベースが求められるようになった。また専門的データベースへの入門版の必要性も高まっている。

ショウジョウバエ系統に関する遺伝学的情報と系統の管理がもっとも充実している研究機関はアメリカ合衆国インディアナ大学のショウジョウバエ系統センターである。インターネットを通じて研究者の要望に応え、約6,000系統を管理・分譲している。我国では京都工芸繊維大学ショウジョウバエ遺伝資源センターが、センター固有の系統とヨーロッパの系統センターとして長年機能してきたスウェーデン王国ウメオ大学から移管している約5,000系統を維持・管理することにしており、維持分譲の国際的2大センター体制を日米の両国に置く準備を整えている。特に日本で有用なデータベース作りをするため遺伝子型、突

然変異形質、遺伝的特性、系統の所在に関する情報を含んだ国内の系統データベースを完備することが必要である。ショウジョウバエ遺伝資源に関する情報データベースとしては、塩基配列や遺伝子領域などDNA分子データを扱ったものや、遺伝子機能情報として対立遺伝子の構造やそれらの形質を扱ったもの、また脳の構造とか、進化的情報に特化したデータベースが作成されている。特にアメリカと英国で共同運営で発信されているFlybaseはショウジョウバエの総合データベースであり、ショウジョウバエに関する全ての情報が網羅されていると言える。しかし、FlyBaseは研究者向けに作成されたデータベースであり、日本語で素早く検索が可能な、ショウジョウバエの研究者だけでなく、ショウジョウバエをモデル生物として利用する研究者にとっても有用な、系統情報データベースを完備し、維持されている系統と関連した情報を発信することで遺伝学研究的基盤を支援することを目標としたデータベースである。現在FlyBaseには突然変異の画像はほとんどない。

さらに、ショウジョウバエ研究の入門者の利便性を考え、日本語によるショウジョウバエの解説や論文をデータベース化している。

日本ショウジョウバエデータベース (<http://133.16.120.118/DroPicDB/DBindex.html>)

3. 日本分類学データベース (DBTJ) の概要

酒井勝司 (四国大・ETI-J)

十脚甲殻類、ゴカイ類 (今島実)、ホヤ類 (時岡隆、西川輝昭) などの種のリスト、シノニムリスト、記載、図版、分布図、文献リストなどをETI-Japanの開発した分類検索ソフトを使用して、そのデータベースをホームページ (<http://etij.c.shikoku-u.ac.jp>) 上で公開する一方、そのCD-ROMを実費で頒布している。

母体であるアムステルダム大学ETI (The Expert Center for Taxonomic Identification) はオランダのユネスコ委員会の支援で設立され、Linnaeus II 入力ソフトを使用して種々のCD-ROMを作成している。平成5年には四国大学ETI-Japanの下で協定を締結したことから、独自にデータベースの作成にあたってきた。其の結果十脚甲殻類短尾類 (酒井恒、武田正倫、Tuerkay; Davie; Guinot)、タナシナ類 (酒井勝司)、異尾類 (McLaughlin)、長尾類 (林健一、Fransen) など3,500種、等脚類 (布村昇) 300種、端脚類 (石丸信一) 300種におよび、多毛類800種、ホヤ類 (時岡隆、西川輝昭) 300種などを入力している。種の記載には、英語入力で行っているが、文献にはドイツ語、フランス語等で記載されたものも多く英語への翻訳を必要とする。図版は出来る限り、原著からのコピーを使用した。生態写真、動画などを入力する事によって検索を容易に出来る。また分布図は原産地、分布の範囲を区別して入力した。文献リストは、必ず原記載の参照が必要となり、19世紀の文献には特に注意が必要となる。入力には、最初に十脚甲殻類短尾類のCD-ROMの完成を目指している。現在CD-ROM作成のため種の特徴の資料整理を行い、text key, picture key, identify keyのそれぞれの検索の作成を行っている。これら一連の作業には、分類学の専門家の最先端の知識を必要とするばかりか、使用するLinnaeus IIソフトの入力技術の習得を必要とする。アムステルダムのETIでは23人の分類学の知識を持った研究者が入力に携わっている。また、CD-ROMに使用する図版については、著作権、出版権が絡んでくる。

ETI-Japanの作成したCD-ROM, "Crabs of Japan", およびETIの作成した, "Macrobenthos of the North Sea II", M.J. de Kluijver etc. (2000)を使って、其の手法を紹介したい。この結果、分類学の資料整理に使用でき、且つ、生態学、環境生物、臨海実習などでの未知の動物の同定にこれら3種の検索を行う事によって容易にtaxaを見出す事ができよう。

ETI-Japanでは、平成5年以降の7年間、文部省の科学研究費 (平成11-12年学術振興会) の研究成果公開促進費の助成を得てCD-ROM及びの入力に携わってきたが、研究者の組織の活性化と、専門家による

Linnaeus IIソフトへの入力が求められているが、其の継続のためには時間と費用が問題となる。

日本分類学データベース (<http://etij.c.shikoku-u.ac.jp/>)

4 . 昆虫学データベース (KONCHU) について

多田内 修 (九大大学院・農学研究院)

昆虫類は既知の全動植物の約 2 / 3 を占め、これまで世界から 100 万種近くが記録されている。しかしスミソニアン研究所のアーウィンらの中南米の熱帯雨林での研究によると、昆虫類は 3,000 万種、後に 8,000 万種という推定値が出されており、熱帯地域を中心に莫大な数の昆虫類が棲息していると予想されている。このように膨大な種を含む昆虫類については、毎年多数の論文が発表され、情報の発見と整理、有効利用については、データベース構築の必要性の非常に高い分野と考えられる。

演者らは 1983 年より昆虫に関する種情報のデータベース化を開始し、九州大学大型計算機センター(現九州大学情報基盤センター)の公用データベース昆虫学データベース(KONCHU)として公開してきた。また 1989 年に出版された「日本産昆虫総目録」についても、MOKUROKU ファイルとして 1990 年にデータベース化を行い、1999 年には海外研究者にも対応できるよう第二版を構築した。さらに日本産全昆虫類の名称辞書データベース (DJI) を構築し、昆虫画像のデータベース化については、花粉媒介昆虫である日本産ハナバチ類の画像データベース (HANABACHI) を構築し、一部公開した。1999 年 6 月からは九州大学大型計算機センターだけでなく、九州大学農学部昆虫学教室のワークステーションにデータベース管理システム www 版昆虫学データベースシステムを開発し、データベースを移築してインターネットからアクセスを可能にした。その結果、従来の大学や研究機関ばかりでなく、行政機関、一般企業、一般愛好家、海外研究者からの上記データベースの利用者が急増し、アジア産昆虫種情報の重要な情報源として利用されている。本報告ではこれまで構築してきた KONCHU, MOKUROKU, DJI, HANABACHI の 4 つのデータベースの概略を紹介する。

昆虫学データベース (KONCHU) は日本の主要な昆虫学・動物学雑誌をもとに、昆虫 (クモ、ダニ類を含む) の分類単位 (種、属、科等) をレコード単位とした文献データベースで、書誌情報、分類情報、分布、キーワード等 13 項目からなり、日本だけでなく東アジア、大平洋地域産の昆虫を対象としている。

昆虫目録データベース (MOKUROKU) は「日本産昆虫総目録」(1989, 1990) をもとに構築した日本産の全昆虫を含む、科、属、種、和名、分布の 5 項目からなるデータベース。

日本産昆虫名称辞書データベース (DJI) は、日本産の全昆虫を対象に、学名から和名、その逆、そして、科名や属名を入力することにより、それに含まれる日本産の全種をリストし、MOKUROKU にリンクして分布を検索できるデータベース。

日本産ハナバチ類画像データベース (HANABACHI) は日本産のハナバチ類の画像を含む専門データベースで、形態の特徴、発生期、訪花植物、天敵、模式標本の雌雄、模式産地、模式標本の所蔵機関など 16 項目からなる、画像は雌雄各々 5 画像からなり、専門家以外にも分類同定の手助けとなる。

昆虫学データベース (KONCHU) (<http://konchudb.agr.agr.kyushu-u.ac.jp/>)

昆虫目録データベース (MOKUROKU) (<http://konchudb.agr.agr.kyushu-u.ac.jp/mokuroku/index-e.html>)

日本産昆虫名称辞書データベース (DJI) (<http://konchudb.agr.agr.kyushu-u.ac.jp/dic/index-e.html>)

5. 分類以外の切り口で多様性をとらえるデータベースの可能性について

木原 章（法政大・自然科学センター・生物）

多様性を記載する作業は、分類学の範疇と捕らえられがちですが、実際にはDNAデータベースに見られるように、客観的な基準を定めて系統的に多くのタクソンを比較検討できるシステムも重要です。例えば、私が大学院時代に行っていたカニの心臓神経系の研究では、単一のタクソンの性質を記載するというよりは、カニの結果を他の生物と比較する点に主たる意義が有りました。

無脊椎動物では「ヤリイカの巨大神経」とか「ザリガニの巨大軸索」「アメフラシの巨大神経節」といったように、タクソンに特徴的な神経細胞が同定・記載されています。それらの知見をまとめ、比較生物学的目的で利用するために、「同定された全ての神経細胞のデータベース」プロジェクトを立ち上げました。

例えば、ザリガニ1種の神経系でも

- 1) 解剖学的に同定されている神経細胞、多数、
 - 2) その生理学的機能まで同定されている個別の細胞、20～30例
 - 3) 神経細胞の集合として場所と機能が同定されている器官、約100例
 - 4) 免疫科学的手法から、その伝達物質と場所が同定されている細胞 多数、
- が、データの対象として挙げられます。

これらの神経細胞の名前、既知の知見、生物種名をまとめれば、特定種の神経だけを検索したり、また機能的に似ている細胞を横断的に検索することも可能になります。神経細胞の場合は、表現型として行動や生態学的な目的でも使うことができます。

今回は、この計画を具体化する上での手続きを、ご紹介したいと思います。

無脊椎動物神経細胞データベースプロトタイプ (http://BioDB.i.hosei.ac.jp/i_Neuron/default.htm)

6. 生物関連データベースのサポート（データベース作成課程と完成までのコツ）

鵜川 義弘（宮城教育大学・環境教育実践研究センター）

これまでに作成の一旦を担ったり維持管理スタッフとして関係した生物関連データベースのうち以下のものに関して、その規模（データ、担当者、SEの有無、使用計算機、予算）等についてまとめ、そのデータベース特有の作成過程や維持方法について解説する。

植物培養細胞文献データベース、日本産樹木検索用データベース、ハイブリドーマデータバンク、放線菌画像データベース、DDBJ/EMBL/GenBank DNA データベース、日本産アリ類カラー画像データベース、原生生物情報サーバ、ジスルフィド結合データベース、イネゲノム、スギゲノムcDNA プロジェクト、ライフサイエンス用語データベース（ライフサイエンス辞書）植物遺伝子シスエレメントデータベース PLACE、粘菌ゲノムcDNA プロジェクト、Bio-Mirror プロジェクト、仙台市カエルマップ、牧野標本館タイプ標本データベース、BioMetasearch生物資源系メタ検索サーバ、オーストラリア産アリ類データベース、マウス画像データベース、水中微小生物図鑑