

生物調査における DNA 分析の活用

— DNA バーコーディングの未来 —

生態工学室 室長 大井 和之

はじめに

九環協が DNA 分析(自社分析)をはじめから 10 年になろうとしている。哺乳類の糞の同定やニッポンバラタナゴとタイリクバラタナゴの識別からはじまり、シアノバクテリアや魚類の胃内容物の同定、地域系統の判定、個体識別などのできることを増やしてきた。現在は、数年前から発展しはじめている環境 DNA 分析技術をいち早く取り入れ、最先端の技術動向に対応できる体制づくりを進めている。

環境 DNA 分析が広まるにつれ、種名の同定や特定の種の存在を分析する上で必要となる参照データのデータベース整備の問題など「環境 DNA 以前」からあった課題が新しいかたちで目にとまるようになってきた。本稿では、DNA 分析の中でも生物の種名の判定に用いられる、DNA バーコーディングと呼ばれる技術について概観し、その課題と将来像について考察していきたい。

1. DNA バーコーディングとは

バーコードと言えば、商店での商品管理に用いられる POS システムがまず思い浮かぶ。商品にあらかじめ印刷されたり、生鮮食品であればシールで貼り付けられたりしたバーコードをレジで読み取り、精算や在庫管理に用いられている。POS のバーコードは 13 桁の数字を示していて、その数字が商品を識別するコードになっている。POS システムでは計量商品など特別の場合を除いてバーコードには売価は書かれておらず、店舗ごとのデータベースで商品名や価格を呼び出して用いられる。

DNA バーコードは、生物細胞に含まれる遺伝情報を記録している分子「DNA」の構造(塩基配列)が生物種ごとに異なるため、特定の部分の塩基配列を分析し、既知の生物の塩基配列のデータベースと照合することで、その生物の種名を判定する助けとするシステムである。



図1 バーコード(POS)

DNA の塩基配列がちょうど POS システムの商品識別コードと同じような役割を果たすため、判定に用いる塩基配列のことを DNA バーコード、データベース検索で種名などを判定することを DNA バーコーディングと呼び習わしている。

生物の種名の判定は、この地球上の生物多様性を知る上で必須の技術であり、標本作製したり形態を観察したりすることは、生物多様性の調査で省略することができない大事なステップである。しかし、博物館の標本と比較しながら厳密に種を同定したり、新種を記載したりするには時間がかかる。地球規模の生物多様性情報の蓄積を加速するために、2001 年に設立された研究者が中心となった国際的な機関「地球規模生物多様性情報機構 GBIF」では、各生物の分布情報や標本のデータベースとともに、DNA バーコーディングの情報整備も行われてきた。

2. 分子系統解析

1990 年代に PCR 法と DNA シークエンサーが発達し、塩基配列解析が容易になったときに、様々な生物の相同な遺伝子の塩基配列を比較して生物の類縁関係を探る分子系統解析が広く行われるようになった。例えば植物では、葉緑体ゲノムにコードされたルビスコ大サ

ブユニット *rbcL* 遺伝子を用いて、種子植物全体の系統関係が明らかにされた(Chase et al. 1993)。一方、動物ではミトコンドリアゲノムのチトクロム b (*cytb*) 遺伝子やチトクロム c オキシダーゼサブユニット I (COI) 遺伝子の塩基配列を用いた分子系統解析が行われた。植物の葉緑体ゲノムと動物のミトコンドリアゲノムはどちらも細胞小器官(オルガネラ)が細胞核とは別に保持している遺伝情報で、真核細胞の起源とされる細胞内共生説の根拠となったものである。(植物にはミトコンドリアゲノムもある。)どちらもたいていの場合雌性配偶子(卵)を通じて子孫に伝わり(母系遺伝)、組換えが起こらないので 1 種類の DNA 分子が細胞内に多数存在し、核ゲノムに比べてコンパクト(植物の葉緑体ゲノムは 15 万塩基対程度、動物のミトコンドリアゲノムは 1 万 6 千塩基対程度の長さ)という特徴がある。このため、相同な遺伝子の DNA 分析が容易で、分子系統解析に広く用いられてきた。

しかし、植物の葉緑体ゲノムと動物のミトコンドリアゲノムでは大きな違いが 1 点あった。それは、植物の葉緑体ゲノムの進化速度が核ゲノムの進化速度の約 1/3 と遅い(塩基置換が起こりにくい)のに対し、動物のミトコンドリアゲノムの進化速度は核ゲノムの 5~10 倍速いという点である。このため、植物の *rbcL* 遺伝子では、同属の種間では塩基配列に 1 塩基の差異もない場合が多く、近縁な種間の系統解析は困難であった。一方、動物では、同一種でも異なる塩基配列になることがよくあり、種の同定にも近縁種を含めた各地の集団の塩基配列データが必要となった。また、哺乳類全体とか魚類全体といった大きな系統関係を解析する際には塩基置換が飽和してしまい、系統解析に必要な情報量が得られずに系統

樹の信頼性が低くなるという問題も発生した。

植物も動物も、分子系統樹の解像度をよくするためには、解析に用いる情報を増やす必要があった。そのため、植物では葉緑体ゲノムでも *rbcL* 遺伝子よりも塩基置換が起こりやすい *matK* 遺伝子(Ooi et al. 1995)やスペーサー領域などが用いられ、動物ではミトコンドリアゲノム全体の塩基配列情報から相同な遺伝子のアミノ酸配列情報をとりだして系統解析が行われた。

3. ユニバーサルプライマー

ゲノムの中で遺伝子とよばれる領域は、通常タンパク質の構造を記録している部分であり、3 つの塩基の組み(コドン)が 1 種類のアミノ酸を示している。タンパク質を構成するアミノ酸は 20 種類しかないため、4 種類の塩基の 3 つ組みで 64 種類あるコドンの複数が同じアミノ酸を示すことになる。3 つ組みの塩基の 3 番目が A、C、G、T のいずれであっても同じアミノ酸を示す場合、この塩基にどんな塩基置換が起こっても遺伝子がコードしているタンパク質の構造は変わらないため、このような塩基置換を「同義置換」という。一方、コードしているアミノ酸が変化する塩基置換を「非同義置換」という。タンパク質をコードしている遺伝子では、同義置換は進化的に中立で、突然変異の発生率と同率で集団中に広がり固定していくが、アミノ酸が変わってしまう非同義置換は、それによりコードしているタンパク質の機能が損なわれる場合はその変異は子孫に伝わらない(機能的制約)ため、見かけの進化速度は遅くなる。

DNA 分析を行う際に、ゲノムの中の特定の領域だけ

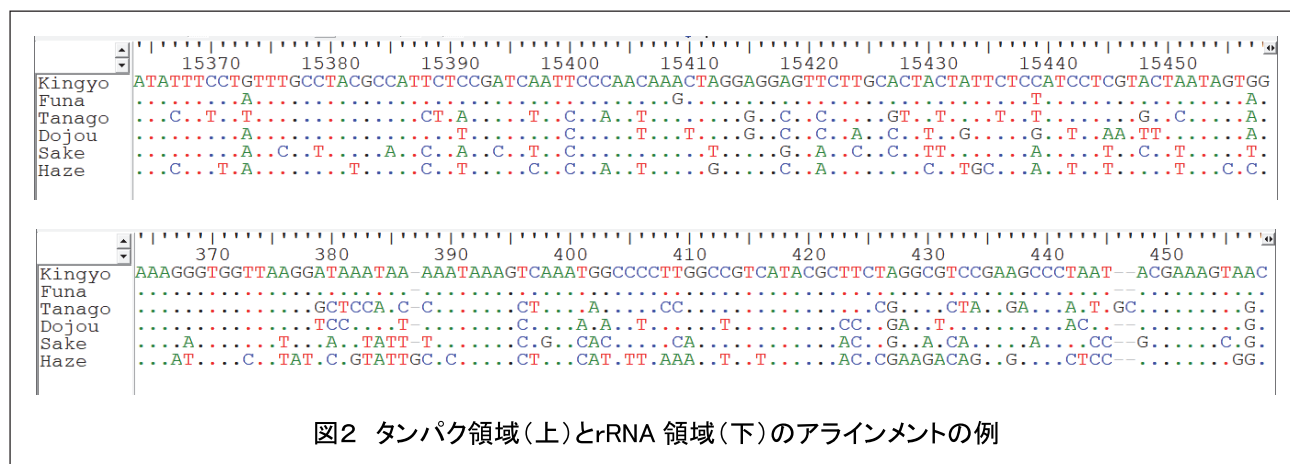


図2 タンパク領域(上)とrRNA領域(下)のアラインメントの例

をとりだすために、PCR (Polymerase chain reaction)を行う。PCR には増幅したい領域を挟むかたちでプライマーという約 20 個の塩基が連なった人工 DNA 分子が必要である。プライマーの塩基配列は、増幅したい生物の当該遺伝子領域の塩基配列と対応していなければならないが、系統解析や種の同定を行う際にはその種の塩基配列情報はまだわかっていない場合が多い。そのため、いろいろな種類の生物に対応できる「ユニバーサルプライマー」を設計する必要がある。ところが、タンパク質をコードしている遺伝子ではコドンの 3 番目の塩基が同義置換になりやすいため、広い分類群に対応するユニバーサルプライマーの設計が難しい。植物の葉緑体ゲノムは進化速度が遅いので、*rbcL* も *matK* も比較的広い分類群に対応するプライマーが知られているが、動物のミトコンドリアゲノムでは COI のユニバーサルプライマーだけが広く使われている。ただし、COI のユニバーサルプライマーは分類群によっては増えにくいので、対象の分類群に適したプライマーに替えた方がよい場合もある。

また、細菌類では 16S リボソーム RNA 遺伝子、真菌類では核ゲノムの 18S リボソーム RNA 遺伝子と 26S リボソーム RNA をコードしている遺伝子やその周辺のスパーサーなので、コドンとは関係なく RNA の高次構造に対応して塩基配列が変わりにくい保存領域と変化しやすい易変領域がある。保存領域にプライマーを設計することで、広い分類群に対応するユニバーサルプライマーを作ることができる(大井, 2015)。

4. データベース

分析対象の生物標本から DNA をとり、PCR によって目的の遺伝子領域の DNA 断片を増幅し、DNA シークエンサーで塩基配列を決定したら、次は既知の塩基配列情報と比較する必要がある。1990 年代の分子系統解析黎明期から、DNA 分析で得られた塩基配列情報は、論文を發表すると同時に DNA データバンク (Genbank) に登録・公開されてきた。(多くの論文誌で、Genbank の

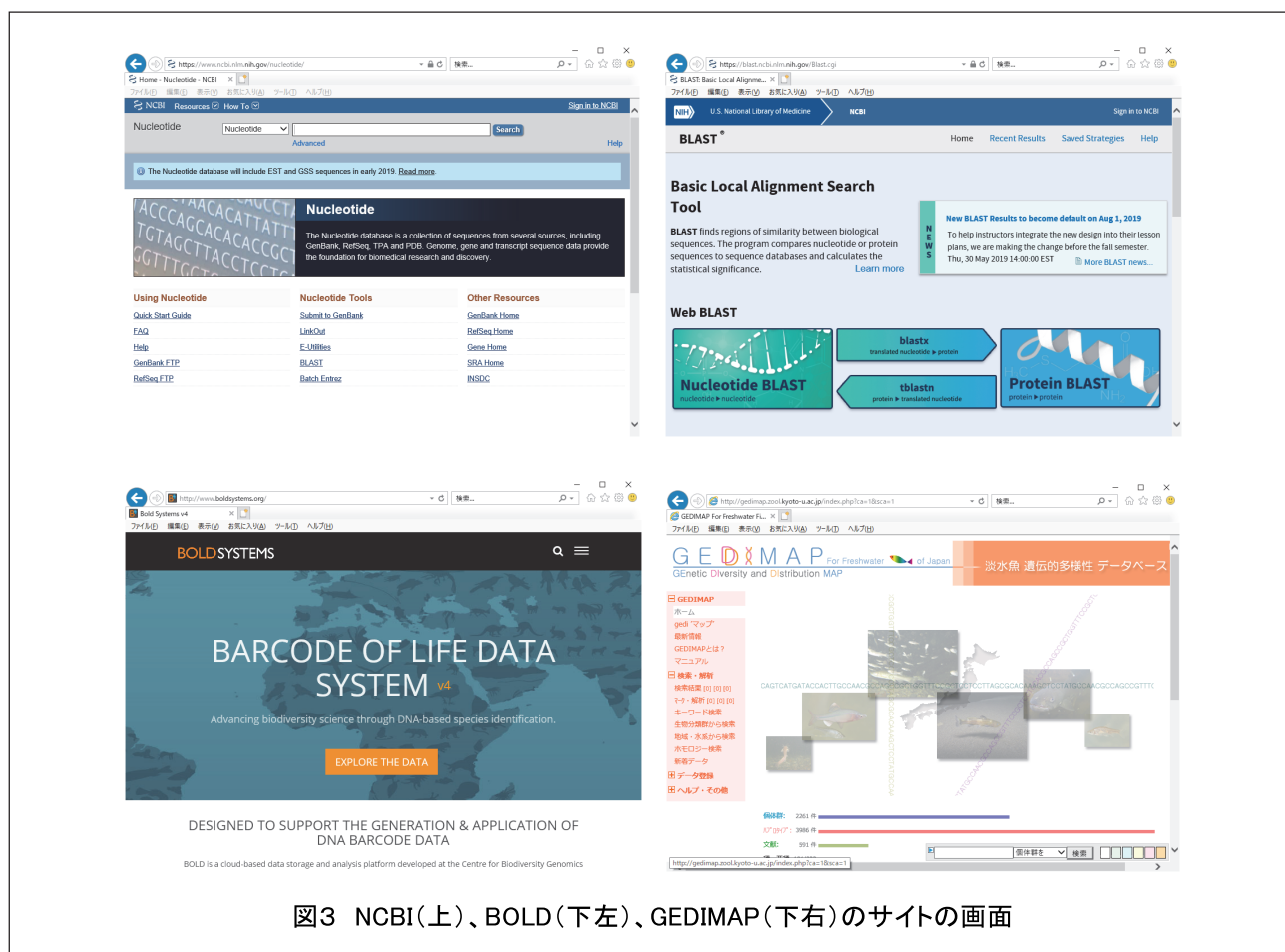


図3 NCBI(上)、BOLD(下左)、GEDIMAP(下右)のサイトの画面

アクセッション番号を記載しなければ論文が受理されない。) 日米欧 (NCBI/EMBL/DDBJ) の合同で運営されている Genbank は、世界最大の塩基配列情報データベースであり、塩基配列の類似性で検索する BLAST などのツールも充実している。早くから 16SrRNA 遺伝子の塩基配列を分類に用いてきた微生物 (細菌) では、BLAST 検索で 97% 以上の相同性がある配列が見つければ、その種名を当ててよいという判断がなされている。

しかし、Genbank に登録されている塩基配列は、もともと実験生物のデータなども多く証拠標本の登録が必要ないこと、試料の採取地点など登録時に付加する情報に対する登録者の裁量が大きいこと、付加された生物の学名などの情報を登録者以外の第 3 者が検証・修正することが難しいことから、誤った生物名に紐付けられて登録されている塩基配列も散見される。このため、DNA バーコーディングでよく用いられる領域について、キュレーターが学名を検証し、証拠標本と紐付けられた塩基配列のデータベースが構築されている。代表的なものが、GBIF のプロジェクトである Barcode of Life Database (BOLD) である。BOLD のウェブサイトからは、動物の COI、植物の *rbcL* と *matK*、菌類の ITS の塩基配列をキーとした検索が行える。また、日本産の淡水魚類の *cytb* などミトコンドリア DNA のハプロタイプについては、京都大学の渡辺先生らによる GEDIMAP という公開データベースがある。

5. メタバーコーディング

微生物 (細菌類) では、平板培地で 1 細胞由来のコロニーを作らせるなどして分離しない限り、様々な種類の菌が混合した状態で存在している。このような混合状態の試料から直接 DNA を抽出し、16SrRNA 遺伝子のようなバーコード領域を増幅するユニバーサルプライマーを使って PCR を行うと、種々の菌由来の DNA 断片が混合した状態のもの (PCR 産物) が得られる。これを直接 DNA シークエンサー (従来型の「サンガー法」で塩基配列を決める機器) にかけても塩基配列を読むことはできないが、PCR 産物をプラスミドという 2000~3000 塩基の特別な DNA 断片に結合し、大腸菌に取り込ませてか

ら平板培地で 1 細胞由来のコロニーを作らせることで、PCR 産物を 1 分子ずつに分けて増やすことができる (クローニング)。1 コロニーのプラスミド DNA から 1 種類の塩基配列を、DNA シークエンサーを使って読むことができる。1 つの試料から得た PCR 産物をクローニングし、100 個程度のコロニーを拾ってそれぞれ塩基配列を決めるという方法で、混合状態の試料に含まれている種の組成を調べること (メタバーコーディング) ができ、1990 年代からつい最近まで広く使われていた。例えば、この方法で土壌中の細菌叢を分析すると、培養が困難なためにこれまで調べられていなかった菌が多数あることがわかってきている。しかし、この方法は手間とコストがかかるため、多数の環境試料を一斉に分析するようなことは難しかった。

従来型の DNA シークエンサーと全く異なる仕組みで、DNA を 1 分子ごとに光感受素子の上の基盤に結合させ、そこで 1 塩基ごとに伸長反応をしながら塩基の種類により異なる蛍光を発生させ、その蛍光を観測して塩基配列を決める「次世代シークエンサー (NGS)」が、2010 年代になって普及してきた。NGS を用いると、クローニングを行わなくても混合状態の PCR 産物の塩基配列を解析すること (アンプリコンシーケンス) が可能で、1 つの試料から簡単に数万個、数十万個の塩基配列を得ることができる。NGS では従来型の DNA シークエンサーに比べて短い塩基配列しか読めないという問題があるため、これまで用いられてきた DNA バーコーディング領域の一部分だけを増幅するプライマーを用いたり、12SrRNA 遺伝子など新しい領域を解析対象としたりする。このため、バーコードの情報量が減り近縁種が同じ配列になってしまったりして区別できなくなったり、リファレンスとなる標本と紐付いた塩基配列データがデータベースに登録されていなかったりする課題があるが、NGS の急速な普及によりデータの蓄積が加速しており、現時点の課題は早晩解決されるものと思われる。

6. DNA バーコーディングの利活用と発展に向けて

九環協で 2010 年から行ってきた DNA 分析の多くは、本稿で紹介した DNA バーコーディングに相当する。そ

の対象生物は植物プランクトン、鉄バクテリアといった微生物から、大型藻類(ノリ、コンブ)、真菌類、植物(種子)、貝類、昆虫類、魚類、両生類、哺乳類などの動植物まで多岐にわたっている。毒素を出すような有害な種類があったり、外来種と在来種が見た目で区別できなかったりするときに、標本の同定を確実にするために DNA 分析を行っている。種子や幼生、仔魚のほか、糞や胃内容物の同定にも DNA バーコーディングが用いられる。

また、九環協には高い生物分類技能を持った職員が複数在籍していて、貴重な生物標本が得られる機会も少なくない。特に DNA バーコーディングのリファレンス配列の登録がまだあまり進んでいない分類群については、業務等の制約がない希少種の標本が得られた場合、DNA バーコーディング領域の塩基配列を決めて DNA データバンクに登録するとともに、標本を博物館へ登録するようにしていく計画である。

今年 2 月には、イルミナ社の「ジャパンアグリ iSeq グラント 2018」で受賞した NGS (iSeq100) が九環協の実験室に設置され、メタバーコーディングが簡単に行えるようになった。特に農業に関連する植物病原体の検出と同定に活用できるよう、試料の採取、DNA 抽出方法の確立、プライマーの設計など、試験手法の研究開発を進めていくことにしている。

環境 DNA 分析、すなわち土壌や水中の微生物叢や魚類など大型生物の網羅的解析において、NGS を用いたメタバーコーディングの需要が高まってきている。九環協では、分析機関として DNA バーコーディング、メタバーコーディングに対応するだけでなく、試験手法の開発、データベースの充実などを通じて、DNA バーコーディング技術の発展に寄与していきたいと考えている。

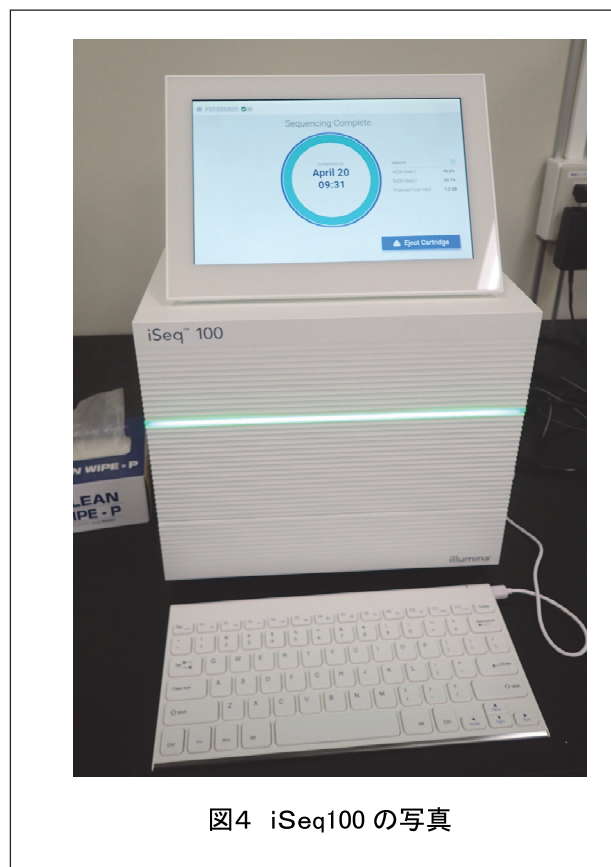


図4 iSeq100 の写真

文献

- Chase, M.W. et al. (1993) Phylogenetics of Seed Plants: An Analysis of Nucleotide Sequences from the Plastid Gene *rbcL*. *Ann. Missouri Bot. Gard.* 80: 528-580.
- Ooi, K. et al. (1995) Useful Primer Designs to Amplify DNA Fragments of the Plastid Gene *matK* from Angiosperm Plants. *J. Jpn. Bot.* 70: 328-331.
- 大井和之 (2015) 環境 DNA 分析におけるプライマー設計. *環境管理* 44: 60-66.