# バイオインフォマティクスによる琵琶湖の固有種イサザの同定
## ～DNA バーコーディングと AI による画像認識～

滋賀県立膳所高等学校　佐藤瑠乃

## 1.研究の目的

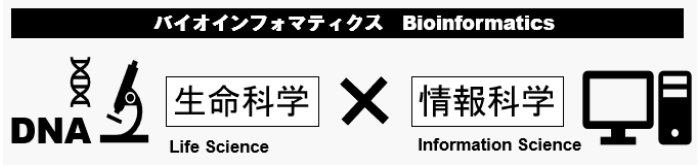
SUSTAINABLE DEVELOPMENT GOALS

SDGs

生物多様性を守る

生物多様性を守ることは、SDGs ゴール 14,15 のターゲットであり、将来に渡り持続することは、私たちが生きる上で非常に大きな意味がある。しかし、人間の経済活動や利便性の追求の結果、生態系の破壊が進行し、1975 年から毎年 4 万種が絶滅している。この状況の改善課題は、近年の自然科学分野における基礎研究の衰退から世界的に専門的な分類学者の数は減少したことだ。早急に種の同定を行うスキル向上や新しいツール開発の必要がある。今回我々の研究グループは、従来の形態分類学を補佐すべく新しい手法として、生命科学と情報科学の融合であるバイオインフォマティクスを活用する。現在の分類学における危機的な現状打破が本研究の目的である。

## 2. 研究仮説



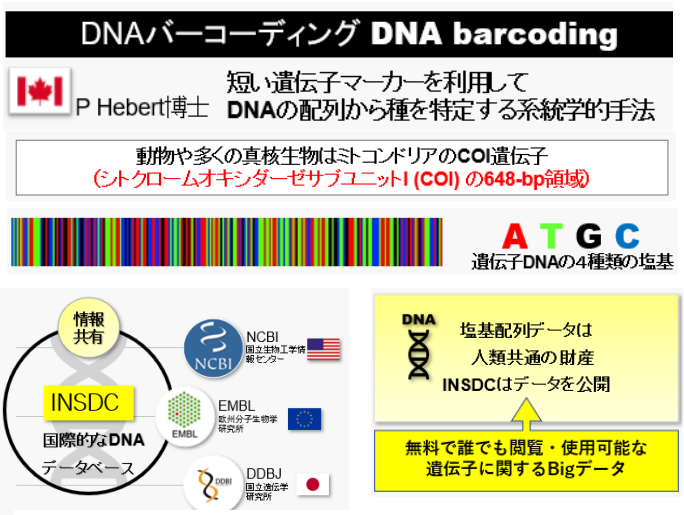

具体的には琵琶湖固有種であり絶滅危惧種であるイサザを対象として、DNA バーコードなどの遺伝子のビッグデータや AI を活用した画像認識などデータサイエンスを活用した手法から種の同定を試みる。DNA 情報と画像認識プログラムの現段階での同定の水準を実際に試みることで、より整合性を向上するには、これらを組み合わせることが必要であると仮定した。

## 3.イサザの生態と使用データ

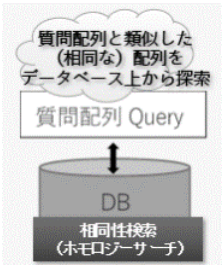
イサザとゴリの DNA（ミトコンドリアのシトクロームオキシダーゼサブユニット I の 648-bp 領域）を利用する。DNA アルゴリズム BLAST によるイサザの DNA の相同性が高い個体検索をデータとして使用した。　次に AI を用いた画像認識によるディープラーニングは、学習時に最適化するパラメータ数が多く数万枚、数十万枚の学習データが必要となる。我々はインターネット上から画像を集め、人工知能に判断させる教師なし学習を試した。まずイサザの画像を集めるために、インターネット上からスクレイピングをする Python プログラミングを作成した。次に、イサザの特徴である背びれを、教師あり学習で判断させた。琵琶湖水産試験場の方々の協力により、実際のイサザとゴリの多数の画像データを入手し、Python で頻繁に使用されるという画像処理ライブラリから openCV を使用し、データの数を増やすデータオーギュメンテーションを試みた。


学名ゲノゴビウス・イサザ（Ghaenogobius isaza）。腹側の吸盤など形状は典型的なハゼ。特徴は淡い黒褐色、背・胸びれ・尾の形状、産卵期雌は腹が黄色。なぜか減少する時期があり、漢字では魚偏に少。シロウオの俗称も「イサザ」だが別の魚で琵琶湖固有種。昼間、琵琶湖湖底に生息、夜間餌のために、最深部のイサザは体長の二千倍近い 90m 湖面まで上昇する。春先、湖岸近い浅瀬に他の魚の産卵時期を避け産卵する。他の魚と生育環境が異なるイサザと近年のゴリとの交雑は、地球温暖化など湖底環境悪化を原因とした生態の変化が考えられ、イサザの同定には、地球環境の指標となる非常に大きな意味が存在するといえよう。

《2022 年 8 月 2 日琵琶湖水産試験場》

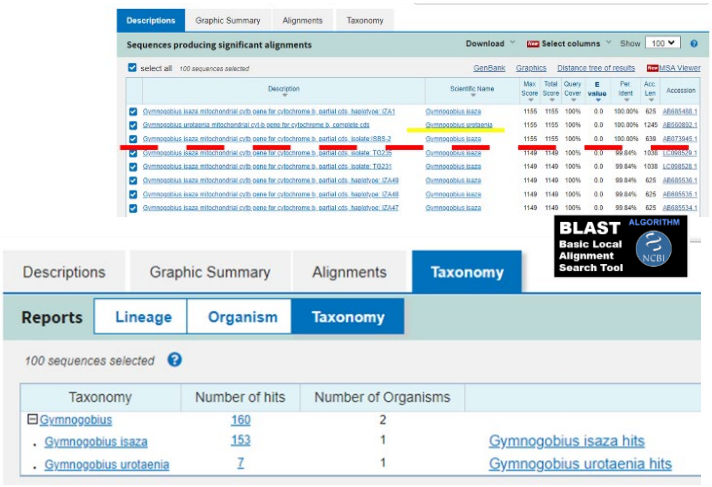




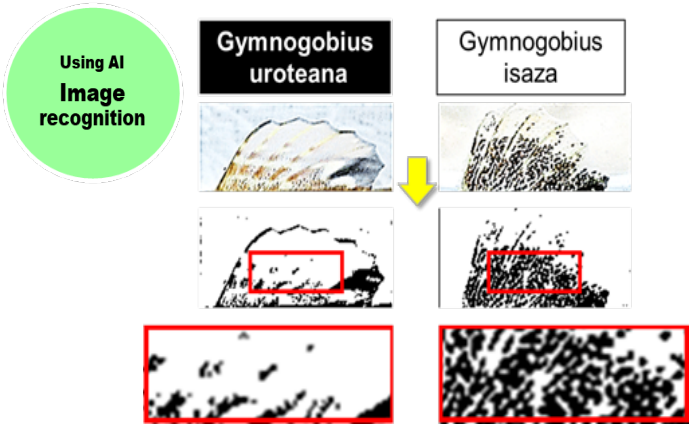



DNA の配列検索の特別なアルゴリズムＢＬＡＳＴは、検索対象と配列から局所的なアラインメントを行う塩基配列を決定し、相同性の高い塩基配列をおよびタンパク質のアミノ酸配列を持つ生物種を国際的な DNA レファレンス・データベースと照合し類似したものを選択するホモロジーサーチを行う。この機能を用いイサザとゴリの相同性検索を試みる。


## 4.分析結果

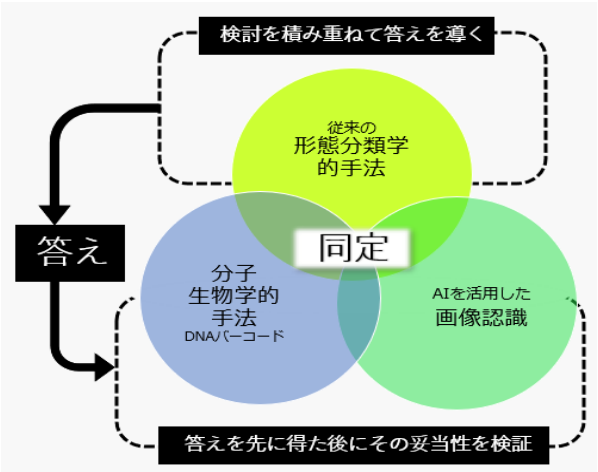



**イサザとゴリの DNA 検索の結果**　DNA が 100%一致する 3 つの個体の一つに、ウキゴリが見つけられる。また全体の結果を見ても、相同性が高いと判断された個体 160 の中に、ウキゴリが 7 個体混在する結果がでた。原因は現段階では、ＤＮＡ抽出時のコンタミネーションなどのミスや、ゴリとの交雑などの推測ができる。

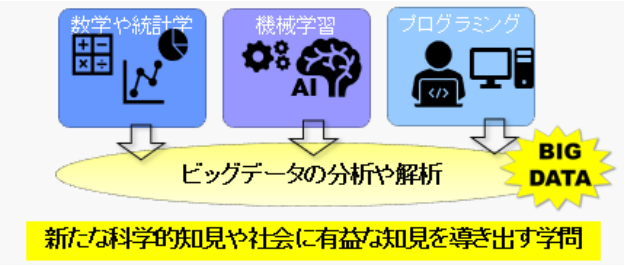



**AI による画像認識の結果**　イサザの画像は、インターネット上でかなり限定され、満足な結果が得られなかった。次に openCV を使用し、データの数を増やすデータオーギュメンテーションを試みた。特徴であるランダムな背びれ模様の認識は非常に高度であることを原因とし、現段階では、100%の水準を持つイサザの画像認識にはかなり課題があるという重大な結果を得ることができた。

## 5.今後の課題と結論




データサイエンスとは、データを用いて新たな科学的および社会に有益な知見を引き出そうとするアプローチのことであるが、多くの研究分野の集合体として成立する背景があり、その全貌は難解だと思われている。我々は松尾豊先生（東京大学大学院工学研究科教授）をはじめ、データサイエンス分野にとどまらない日本を代表する権威の方々のご教授により、次の 3 点の重要性に気づいた。①ミクロの視点ではなくマクロの視点②デザイン思考③オブジェクト指向。つまりデータサイエンスには、一つのことを掘り下げるミクロ的な視点の研究も重要だが、現在私たちの周りに溢れかえる情報を用いて、社会のニーズを考え、様々な研究を組み合わせて課題解決を図ることこそが今まさに求められているのである。この研究はその思考プロセスでありアプローチ手法をイサザの種の同定という環境問題の解決に繋がる問題を実践的に行った先行研究である。結論として、さらに同定手法の向上に努め、次年度以降の研究に繋げ、発展する重要性が存在するといえよう。




現在は、AI・IoT が高度な知的活動を担い、機械によって知的活動の自動化・個別生産化が行われる第四次産業革命（インダストリー4.0）中と言われている。しかし今世紀中には第五次産業革命が訪れる。生物学と情報科学の融合が進むことが予測され、さらにコロナ禍によりその歩みは早まっている。コンピュータ技術とバイオテクノロジーの融合”つまり生物学×情報学＝バイオインフォマティクスが未来を担っているのだ。一見対極にある生命と無機質な情報処理の世界は膨大な情報という点で非常に親密性があり、その解明にコンピュータは非常に適している。この探求により、バイオインフォマティクスこそ、データサイエンスの可能性が最も広がる領域であり、生物学の基礎となる分類学の発展に寄与し、生物多様性の保護に繋がると考えている。


## 6.謝辞

研究を進めるに際して、
京都大学/南直治郎教授、
滋賀県立大学/田辺祥子准教授
龍谷大学/山中裕樹准教授、
大阪電気通信大学/長瀧寛之特任准教授
株式会社バイオーム/藤木庄五郎様
琵琶湖水産試験場、琵琶湖博物館の皆様から
多くの助言をいただき研究が深まりました。深く感謝を申し上げます。