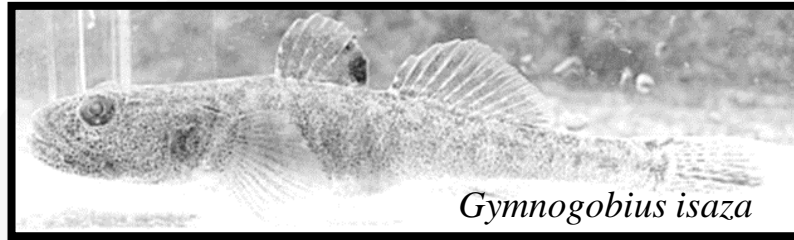


SCIENCE PROJECT



# 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

はじめに

# SUSTAINABLE DEVELOPMENT GOALS




# SDGs



## 生物多様性を守る



 **地球上の生物の種類**

認知済み  
約175万種  
未確認含む  
3,000万種  
~1億種以上

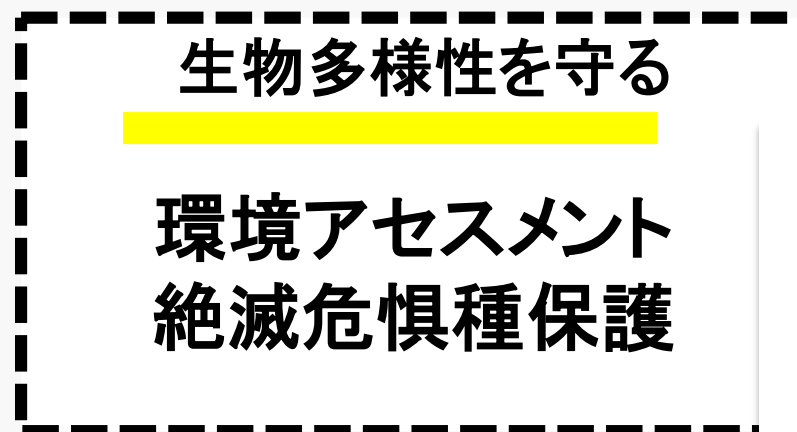
1975年以降

**毎年4万種が絶滅**

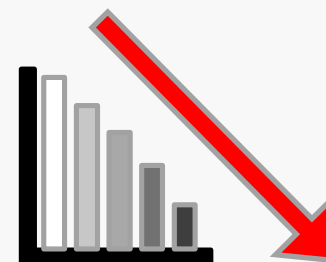


生物多様性を守るために科学的知見の充実が重要！

Science



## 自然科学分野における基礎研究の衰退



世界的に分類学者の数は年々減少

生物分類の早急なスキル向上 & ツールの開発が必要！

# バイオインフォマティクス Bioinformatics



生命科学

Life Science



情報科学

Information Science



```

Gymnogobius_uroteana  GTAGGTGGAACCTCTTGCCAATGAATTTGAGGAGGTTTCTCAGTAGATAATGCTACCCCT  80
Gymnogobius_isaza    GTAGGAGGAACTCTTGTTCAATGAATTTGAGGGGGCTTCTCAGTAGATAATGCCACCCCT  80
*****

Gymnogobius_uroteana  ACACGATTTTTGCATTTCAATTTCTCCCTTTGGGATTCTTGCCGCTACCCCTTCTG  120
Gymnogobius_isaza    ACACGATTTTTGCATTTCAATTTCTACTCCCTTTGTAGTTCTTGCTGCTACCCCTTCTG  120
*****

Gymnogobius_uroteana  CATCTTCTTTTCTTACATGAAACTGGCTCAAATAACCCAGCAGGATTAACCTCCGATGCC  180
Gymnogobius_isaza    CATCTTCTTTTCTTACATGAAACTGGCTCAAATAACCCGGCAGGCTTAACCTCCGATGCC  180
*****

```



## 同定 Identification

琵琶湖の固有種



絶滅危惧IA類 (CR)  
環境省レッドリスト

イサザ  
*Ghaenogobius isaza*

# 研究の仮説

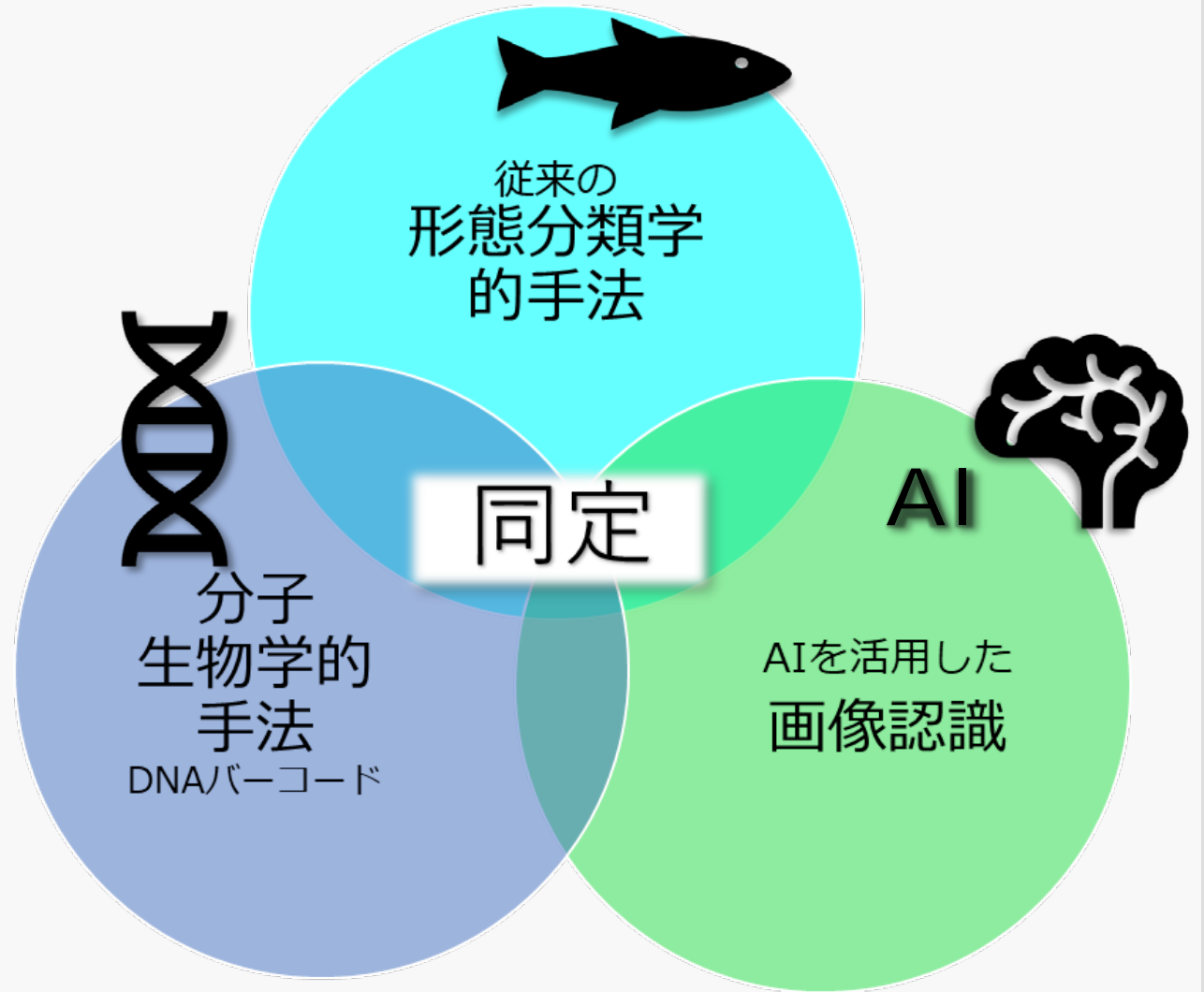
イサザ



ウキゴリ



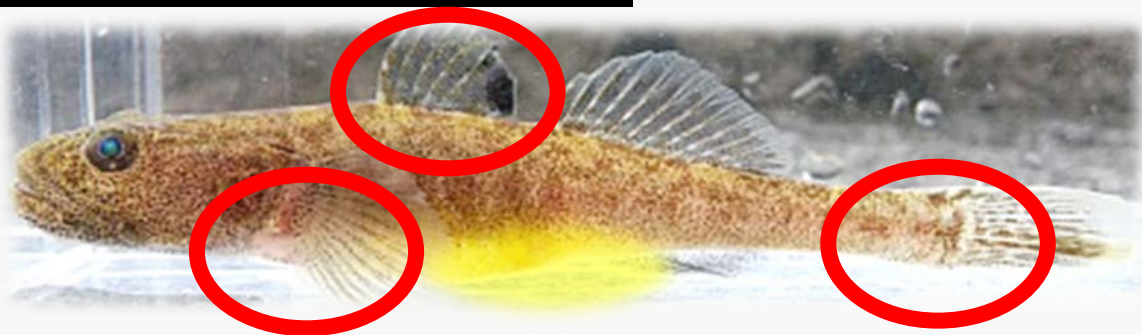
イサザかな？



組み合わせることにより、同定の水準が向上すると仮定

従来の  
形態分類学  
的手法

## イサザの生態



**イサザ**  
*Ghaenogobius isaza*  
・ハゼ (goby) の仲間  
・体長約5cm程度



イサザ

## イサザの特徴

3月～4月 産卵

琵琶湖  
の  
固有種

BIWAKO

琵琶湖にしか  
生息しない！

🌙 夜間

90m

☀️ 昼間

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る



イサザの同定

地球環境の指標

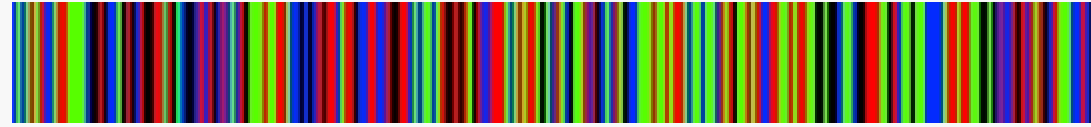
# DNAバーコーディング DNA barcoding



Paul Hebert博士 ⇨ 短い遺伝子マーカ―を利用してDNAの配列から種を特定する系統学的手法



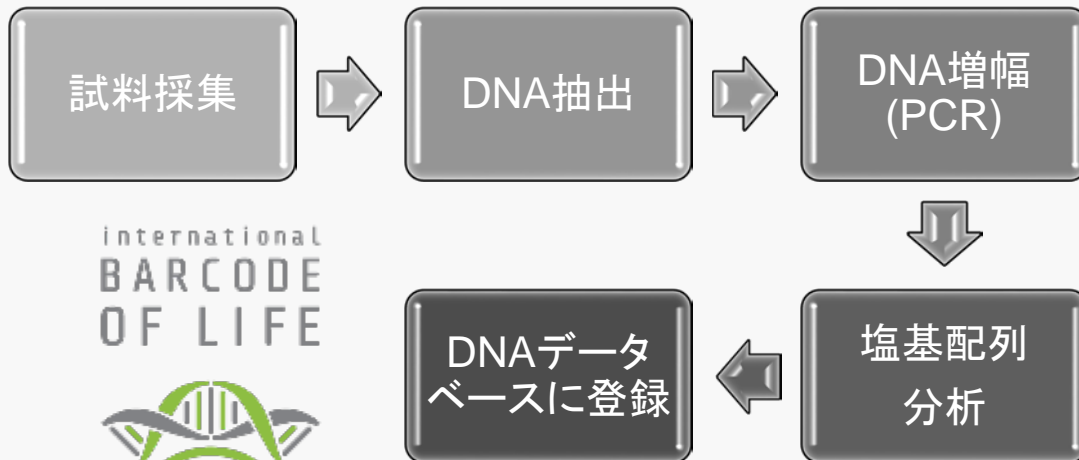
DNA



A T G C  
遺伝子DNAの4種類の塩基

動物や多くの真核生物はミトコンドリアのCOI遺伝子  
(シトクロームオキシダーゼサブユニットI (COI) の648-bp領域)

## DNAバーコードの取得方法



## DNAバーコードの貢献

生物種の同定, 環境アセスメント,  
絶滅危惧種保護,  
移入生物管理  
食品の安全性, 水質モニタリングなど

情報  
共有

国際的なDNA  
データベース

INSDC



NCBI 国立生物工学情報センター



EMBL 欧州分子生物学研究所



DDBJ 国立遺伝学研究所



塩基配列データは  
人類共通の財産  
INSDCはデータを公開

International  
Nucleotide  
Sequence  
Database  
Collaboration

**BIG  
DATA**

無料で誰でも閲覧・使用可能な  
遺伝子に関するBigデータ

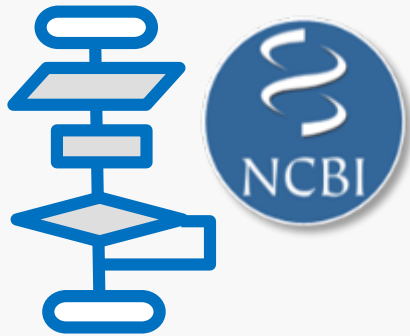


# BLAST

Basic Local Alignment Search Tool

相同性検索  
(ホモロジーサーチ)

ALGORITHM

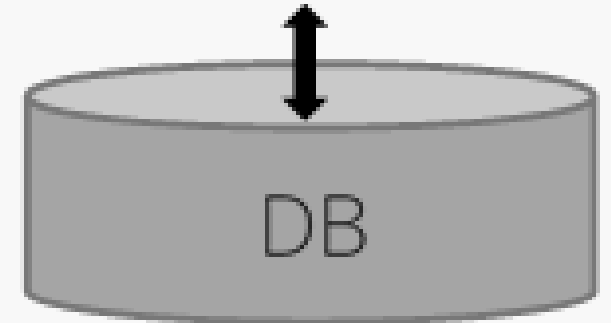


NCBIが開発した  
DNAの配列を  
検索するための  
特別なアルゴリズム

配列の  
類似性から  
類縁遺伝子を  
検索する  
配列解析の  
基本的方法

質問配列と類似した  
(相同な) 配列を  
データベース上から探索

質問配列 Query



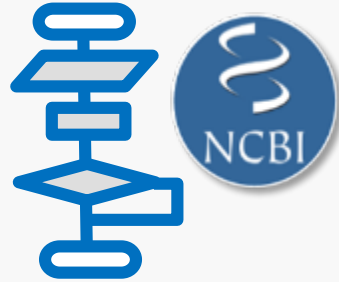
Gymnogobius\_uroteana  
Gymnogobius\_isaza

```
GTAGGTGGA ACTCTTGTCCAATGAATTTGAGGAGGTTTCTCAGTAGATAATGCTACCOCTT 60  
GTAGGAGGAACTCTTGTTCAATGAATTTGAGGGGGCTTCTCAGTAGATAATGCCACCOCTT 60  
***** ** *****
```

# ALGORITHM

# BLAST

Basic Local Alignment  
Search Tool



ウキゴリ



>Gymnogobius uroteana

```
1 gtaggtggaa ctctgtcca atgaattga ggaggttct cagtagataa tgctaccctt
61 acacgatttt tgcatttca tttctctc cccttggga ttctggccg cacccttctg
121 catctttt tctacatga aactggctca aataaccag caggattaa ctccgatgcc
181 gacaaaatcc ctttcacc ctactttcc tataaagatc ttctggctt tgccttata
241 ctctagccc tgcctcct tgcactttt tcccctaatt acctggaga tctgacaat
301 tttaccctg caaacccgt tgtactcct cccacatta agccagagt atatttctt
361 ttgcatatg ctattctcg ttcatcct aacaagctag gaggagtct agccctcct
421 gttccattt tggactact cttgtcct ctctacata cgtcaaaaca acgtagtgtg
481 acctccgcc cagtttca attctctc tgagccctg tagcagatg actattcta
541 actgaattg gaggcatacc tgtgaacac ccgtacatta ttattggaca aattgcatc
601 tcatctact tctcattt tcttg
```

>Gymnogobius isaza

```
1 gtagggaa ctctgttca atgaattga ggggcttct cagtagataa tgccaccctt
61 acacgatttt tgcatttca tttctactt ccctttag tcttgctgc tacccttctg
121 catctttt tctacatga aactggctca aataaccgg caggctaaa ctccgatgcc
181 gacaaaatcc ctttcacc ttactttcg tataaagatc ttctggctt tgccttata
241 ctctagccc tgcctcct tgcactttt tcccctaact acctggaga tctgacaac
301 tttaccctg caaacctact tgcactccc cccacatta aaccagagt atatttctt
361 ttgcatatg ctattctcg ttcatcct aacaagctag gaggagtct agctctcct
421 gttctatct tagtactact cttgtcct ctctgcata catcaaaaca acgaagtgtg
481 acctccgcc cagtttca attctctc tgaaccctg tagcagatg actcattcta
541 actgaattg gaggtatacc tgtgaacac ccgtacatta ttattggaca aattgcatc
601 tcatctact tctcattt tcttg
```

イサザ



The screenshot shows the BLAST web interface from the U.S. National Library of Medicine. It includes fields for entering a query sequence, selecting a database (currently set to 'Nucleotide collection (nr)'), and choosing a search program (currently set to 'megablast'). The interface is in Japanese and includes a 'BLAST' button at the bottom.

# 結果

## Sequences producing significant alignments

Download

New

Select columns

Show

100

select all 100 sequences selected

GenBank

Gr

イサザ



Description	Scientific Name	Max Score	Total Score
<input checked="" type="checkbox"/> <a href="#">Gymnogobius isaza mitochondrial cytb gene for cytochrome b, partial cds, haplotype: IZA1</a>	<a href="#">Gymnogobius isaza</a>	1155	1155
<input checked="" type="checkbox"/> <a href="#">Gymnogobius urotaenia mitochondrial cyt-b gene for cytochrome b, complete cds</a>	<a href="#">Gymnogobius urotaenia</a>	1155	1155 100% 0.0 100.00% 1245 <a href="#">AB560892.1</a>
<input checked="" type="checkbox"/> <a href="#">Gymnogobius isaza mitochondrial cytb gene for cytochrome b, partial cds, isolate: ISRS-2</a>	<a href="#">Gymnogobius isaza</a>	1155	1155 100% 0.0 100.00% 639 <a href="#">AB073945.1</a>
<input checked="" type="checkbox"/> <a href="#">Gymnogobius isaza mitochondrial cytb gene for cytochrome b, partial cds, isolate: TG235</a>	<a href="#">Gymnogobius isaza</a>	1149	1149 100% 0.0 99.84% 1038 <a href="#">LC098529.1</a>
<input checked="" type="checkbox"/> <a href="#">Gymnogobius isaza mitochondrial cytb gene for cytochrome b, partial cds, isolate: TG231</a>	<a href="#">Gymnogobius isaza</a>	1149	1149 100% 0.0 99.84% 1038 <a href="#">LC098528.1</a>
<input checked="" type="checkbox"/> <a href="#">Gymnogobius isaza mitochondrial cytb gene for cytochrome b, partial cds, haplotype: IZA49</a>	<a href="#">Gymnogobius isaza</a>	1149	1149 100% 0.0 99.84% 625 <a href="#">AB685536.1</a>
<input checked="" type="checkbox"/> <a href="#">Gymnogobius isaza mitochondrial cytb gene for cytochrome b, partial cds, haplotype: IZA48</a>	<a href="#">Gymnogobius isaza</a>	1149	1149 100% 0.0 99.84% 625 <a href="#">AB685536.1</a>
<input checked="" type="checkbox"/> <a href="#">Gymnogobius isaza mitochondrial cytb gene for cytochrome b, partial cds, haplotype: IZA47</a>	<a href="#">Gymnogobius isaza</a>	1149	1149 100% 0.0 99.84% 625 <a href="#">AB685536.1</a>

ウキゴリ



Descriptions

Graphic Summary

Alignments

**Taxonomy**

Reports

Lineage

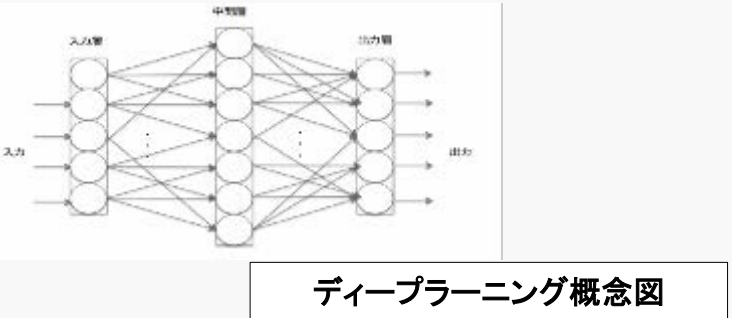
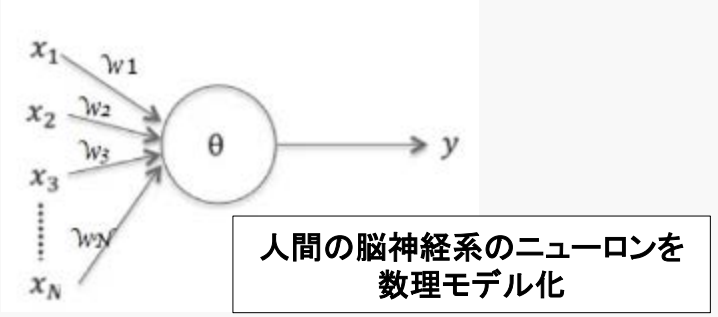
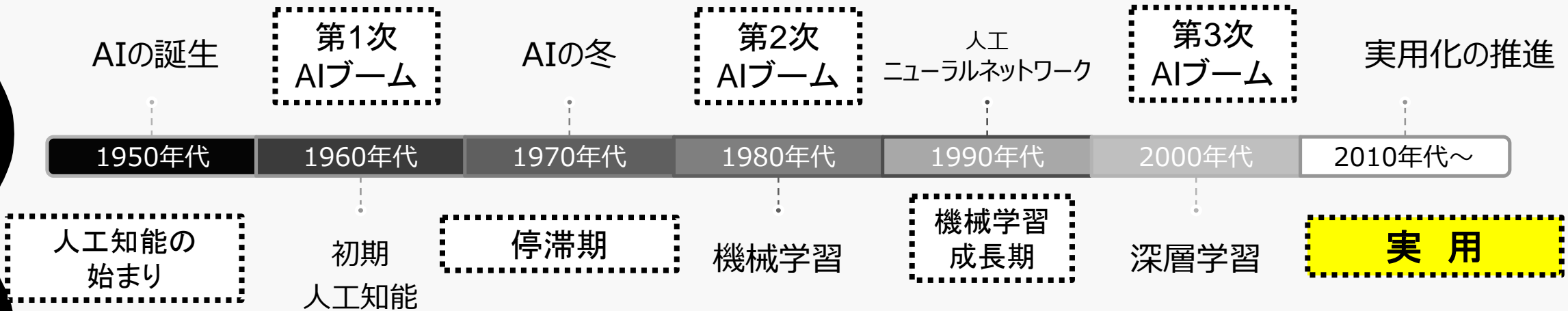
Organism

**Taxonomy**

100 sequences selected

Taxonomy	Number of hits	Number of organisms	Description
<input checked="" type="checkbox"/> <a href="#">Gymnogobius</a>	160	2	
• <a href="#">Gymnogobius isaza</a>	153	1	<a href="#">Gymnogobius isaza hits</a>
• <a href="#">Gymnogobius urotaenia</a>	7	1	<a href="#">Gymnogobius urotaenia hits</a>

# AI 人工知能 Artificial Intelligence



**人工知能 AI**

**機械学習**  
(エキスパートシステム)

人工ニューラルネットワーク

**深層学習**  
(ディープラーニング)

人間の知的能力をコンピュータ上実現する  
様々な技術・ソフトウェア・コンピュータシステム

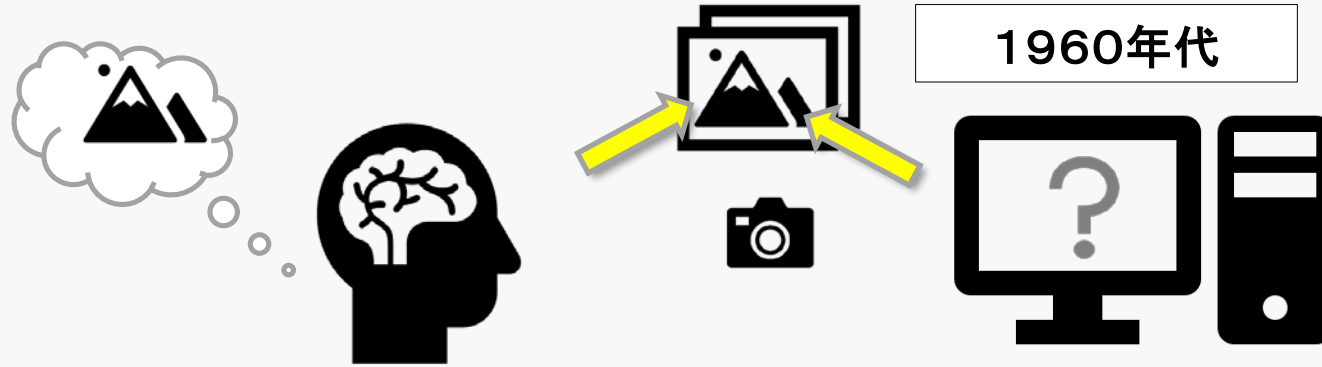


- 機械翻訳のための自然言語処理 (NPL: Natural language Processing)
- 音声認識

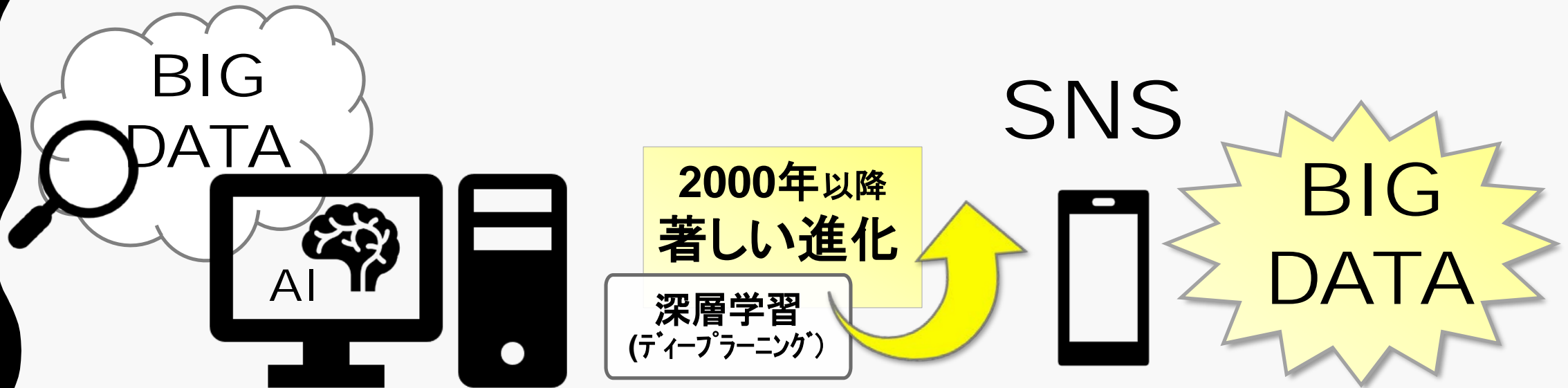


## 画像認識

# 画像認識 Image Recognition



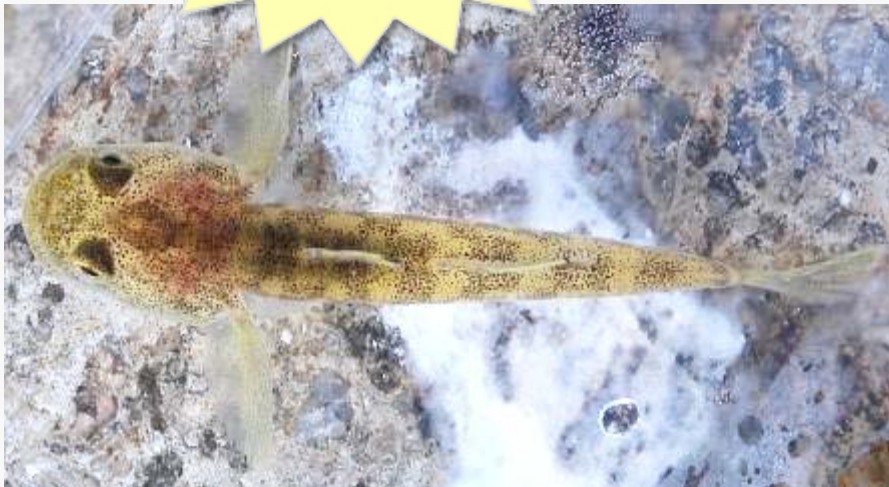
コンピュータに「画像に何が写っているのか？」を理解させる技術



## スクレイピング

スクレイピングで  
検出したイサザの画像

~~BIG  
DATA~~



```
# from icrawler.builttin import GoogleImageCrawler
# イサザの画像を100枚取得
# crawler = GoogleImageCrawler(storage={"root_dir": "Gymnogobius isaza"})
# crawler.crawl(keyword="Gymnogobius isaza", max_num=100)
from icrawler.builttin import BingImageCrawler

# イサザの画像を100枚取得
crawler = BingImageCrawler(storage={"root_dir": "Gymnogobius isaza"})
crawler.crawl(keyword="Gymnogobius isaza", max_num=100)
```

```
2022-01-26 07:21:13,518 - INFO - icrawler.crawler - start crawling...
2022-01-26 07:21:13,521 - INFO - icrawler.crawler - starting 1 feeder threads...
2022-01-26 07:21:13,525 - INFO - feeder - thread feeder-001 exit
2022-01-26 07:21:13,530 - INFO - icrawler.crawler - starting 1 parser threads...
2022-01-26 07:21:13,534 - INFO - icrawler.crawler - starting 1 downloader threads...
2022-01-26 07:21:13,872 - INFO - parser - parsing result page https://www.bing.com/images/async?q=Gymnogobius isaza&first=0
2022-01-26 07:21:14,421 - INFO - downloader - image #1 https://zukan.com/media/leaf/original/D13062.jpg
2022-01-26 07:21:14,881 - INFO - downloader - image #2 http://zakonomizube.web.fc2.com/images/isazal.jpg
2022-01-26 07:21:16,803 - INFO - downloader - image #3 https://tansuigyo.net/a/cb/c-595.jpg
2022-01-26 07:21:18,589 - INFO - downloader - image #4 http://www.biwahaku.jp/english/facts/images/isaza.jpg
2022-01-26 07:21:18,694 - INFO - downloader - image #5 https://i.ebayimg.com/images/s/i2gAAOSwk2lfvUjv/s-1300.jpg
2022-01-26 07:21:19,347 - INFO - downloader - image #6 https://tansuigyo.net/a/cb/c-834.jpg
2022-01-26 07:21:19,551 - INFO - downloader - image #7 https://cdn-ak.f.st-hatena.com/images/fotolife/n/nenjin/20190529/20190529014426.jpg
2022-01-26 07:21:20,064 - INFO - downloader - image #8 https://tansuigyo.net/a/cb/c-670.jpg
2022-01-26 07:21:21,638 - INFO - downloader - image #9 https://www.sumeshiya.com/blog/files/isaIMG\_7610.jpg
2022-01-26 07:21:23,057 - INFO - downloader - image #10 http://tansuigyo.net/a/cb/c-652.jpg
2022-01-26 07:21:23,397 - INFO - downloader - image #11 https://www.sumeshiya.com/blog/files/isaIMG\_7613.jpg
2022-01-26 07:21:23,941 - INFO - downloader - image #12 http://www.sumeshiya.com/blog/files/isaIMG\_7589.jpg
2022-01-26 07:21:24,986 - INFO - downloader - image #13 https://www.zukan-bouz.com/public\_image/Fish/641/Thumb630/isaza.jpg
2022-01-26 07:21:26,683 - INFO - downloader - image #14 https://tansuigyo.net/a/cb/c-508.jpg
2022-01-26 07:21:26,845 - INFO - downloader - image #15 https://www.sumeshiya.com/blog/files/isaIMG\_7617.jpg
2022-01-26 07:21:27,006 - INFO - downloader - image #16 https://www.sumeshiya.com/blog/files/isaIMG\_7593.jpg
```

# 画像処理 Image processing



## 画像認識の前段階処理

画像のノイズ・歪みを除去

明るさ・色合いを調整

オブジェクトの輪郭を強調

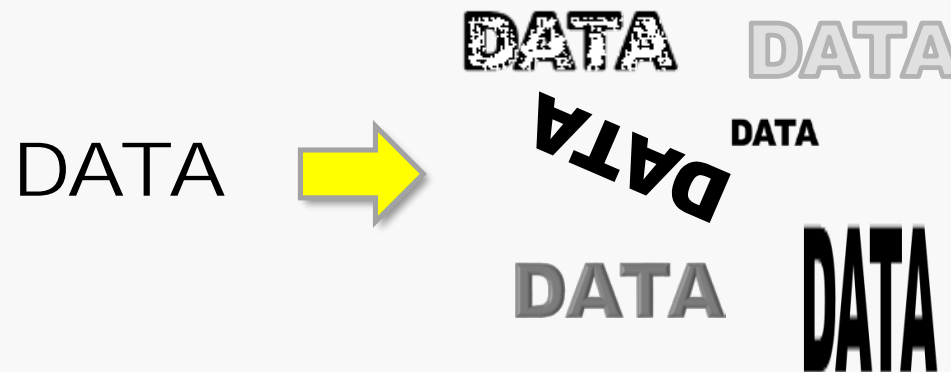
オブジェクトの領域抽出

# Data Augmentation データオーギュメンテーション(データ拡張)

ディープラーニングは学習時に最適化する  
パラメータ数が多いため  
数万枚、数十万枚の学習データが必要

データに様々な線形変換を加え  
データの数を増やす

データ数を水増しだけでなく  
画像のずれにたいしてもロバストになる



# 滋賀県水産試験場訪問と画像アプリ検証



滋賀県彦根市八坂町2138-3  
滋賀県水産試験場

滋賀県水産試験場



画像認識  
アプリ検証

14.4%

イサザ  
*Gymnogobius isaza*

イサザの正答率  
平均  
10~20%



# 結果



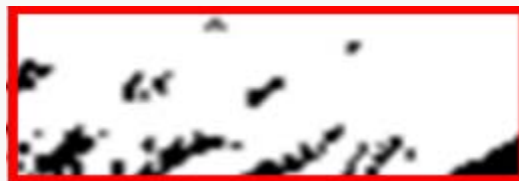
ウキゴリ



イサザ



ウキゴリ



イサザ



輪郭抽出

Cv2.findContours()

```

ファイル(F) 編集(E) 選択(S) 表示(V) 移動(G) 実行(R) ターミナル(T) ヘルプ(H)
import cv2 • Untitled-1 - Visual Studio Code
Get Started リリースノート: 1.61.1 import cv2 Untitled-1
1 import cv2
2 import matplotlib.pyplot as pyplot
3
4 img_bgr = cv2.imread('isaza01.jpg')
5 img_gray = cv2.cvtColor(img_bgr, cv2.COLOR_BGR2GRAY)
6 retval,thresh = cv2.threshold(img_gray,88 ,250,0)
7 img, contours, hierarchy = cv2.findContours(thresh,cv2.RETR_LIST,cv2.CHAIN_APPROX_SIMPLE)
8 result_img = cv2.drawContours(img, contours,-1,(0,0,255))
9
10 #
11 plt.imshow(result_img):
12 plt.show()
13

```



考察

検討を積み重ねて答えを導く

従来の  
形態分類学  
的手法

同定

分子  
生物学的  
手法  
DNAバーコード

AIを活用した  
画像認識

答え

答えを先に得た後にその妥当性を検証

具体的な同定に対するアプローチ

従来の  
形態分類学  
的手法

イサザの特性について調査



分子  
生物学的  
手法  
DNAバーコード

国際的な  
DNAレファレンス・データベース照合  
BLASTによるホモロジー検索

AIを活用した  
画像認識

AIによる  
画像認識プログラムを作成

AI



多角的視点での検証が正しい同定につながる

# 今後の課題

Bioinformatics



**DNA**

Life Science



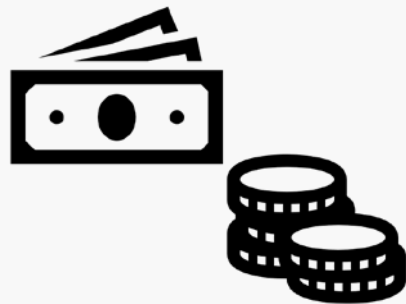
Information  
Science



**AI**

BIG  
DATA

研究における  
メリット



コスト削減



時間短縮

**重要**

技術向上

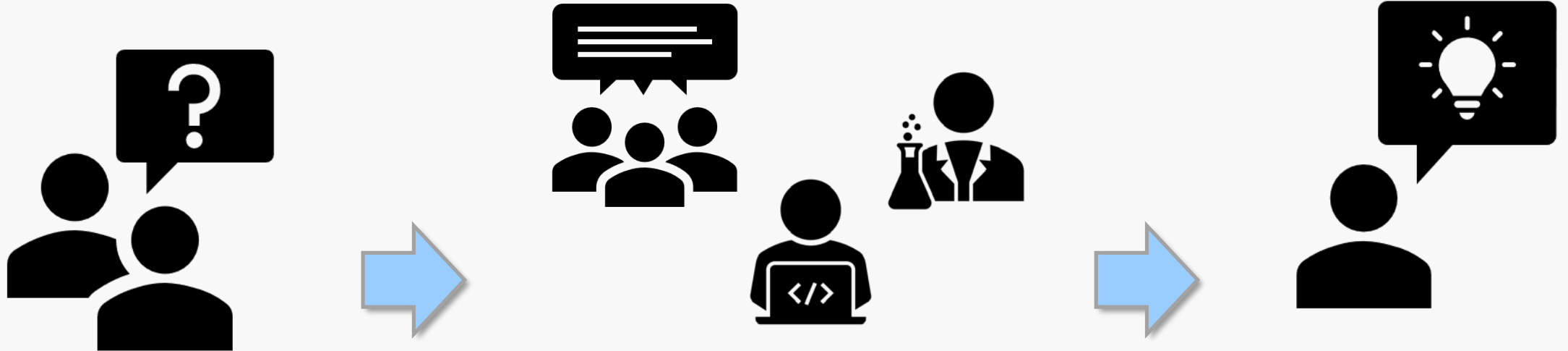


多角的視点での  
顕彰

自ら課題を発見

多角的視点で検証

課題解決

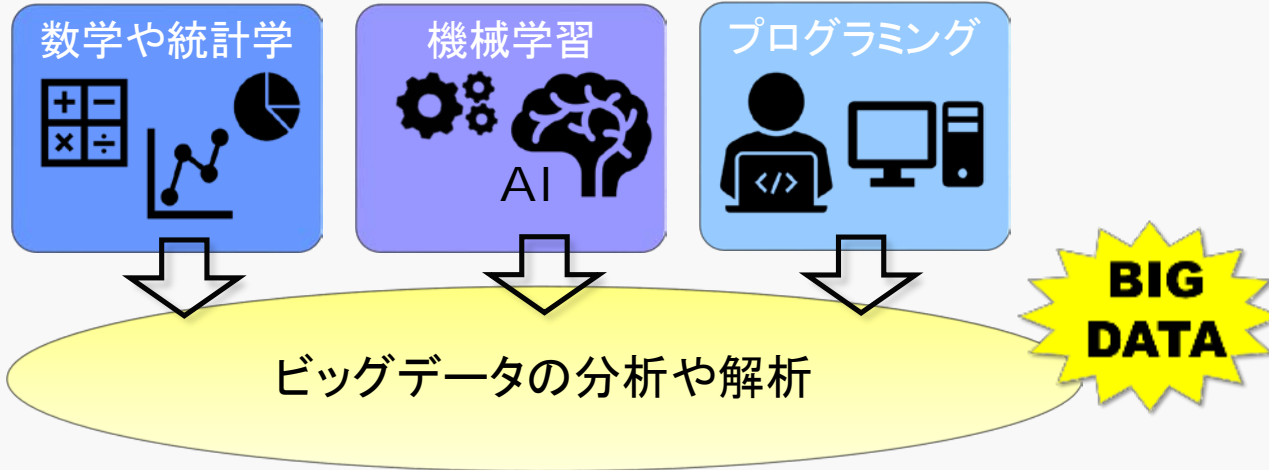


データサイエンス Data Science

実際に自ら考え、  
多角的視点で検証し、  
課題解決につなげる

# データサイエンス Data Science

生物学、医学、工学、経済学、社会学、人文科学...



新たな科学的知見や社会に有益な知見を導き出す学問



世界が抱える問題は  
答えが一つではない

データサイエンスは 多角的視点が必要!

# Bioinformatics



Life Science



Information  
Science



生物多様性を守る



# 謝 辞

研究を進めるに際して

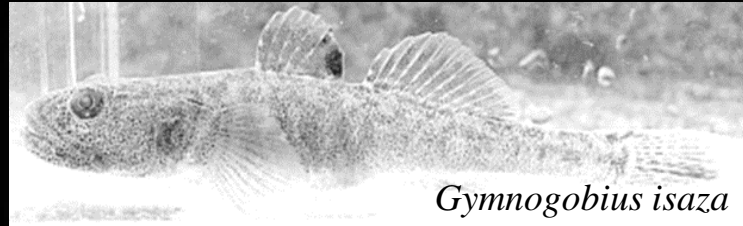
京都大学	南 直治郎	教授
滋賀県立大学	田辺 祥子	准教授
龍谷大学	山中 裕樹	准教授
大阪電気通信大学	長瀧 寛之	特任准教授

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

ご清聴ありがとうございました。



Gymnogobius isaza

SCIENCE PROJECT