



中国科学院水生生物研究所

INSTITUTE OF HYDROBIOLOGY, CHINESE ACADEMY OF SCIENCES

高性能计算助力水生生物 物种鉴定

曾宏辉

水生生物数据分析管理平台

2023-12-4



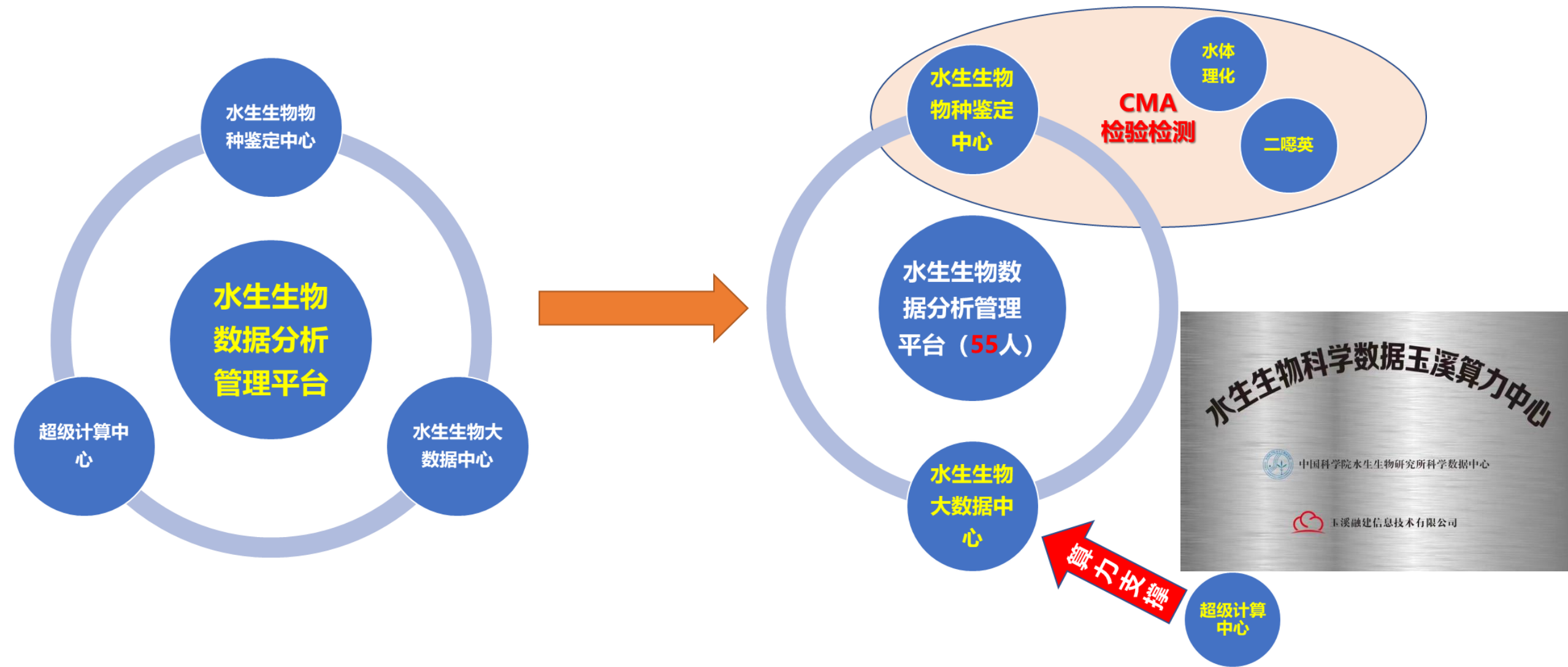
目录

- 1 中国科学院超级计算武汉分中心现状**
- 2 经典水生生物形态学鉴定
- 3 基于人工智能的水生生物形态学鉴定
- 4 基于eDNA技术的水生生物鉴定





中国科学院超级计算武汉分中心：现状



水生生物大数据中心是平台的核心，中国科学院超级计算武汉分中心为大数据中心提供算力支持。



中国科学院超级计算武汉分中心

中国科学院水生生物研究所

湖北省生态环境科学研究院

水生生物大数据中心共建合作协议

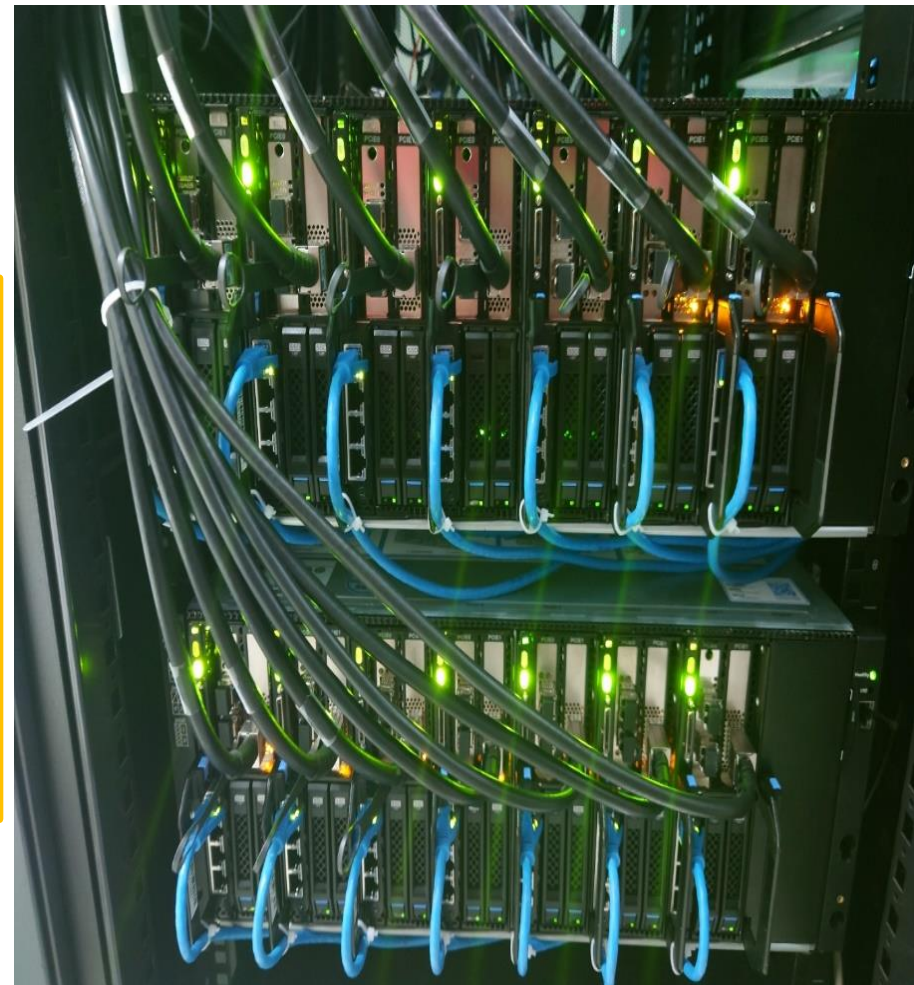
二零二二年三月

合作协议



水生生物大数据中心

中国科学院水生生物研究所
湖北省生态环境科学研究院



湖北省环科院提供
计算资源

与共建单位深度合作提高了超算分中心的算力



中国科学院超级计算武汉分中心



180m²

面积



2.2P

储存



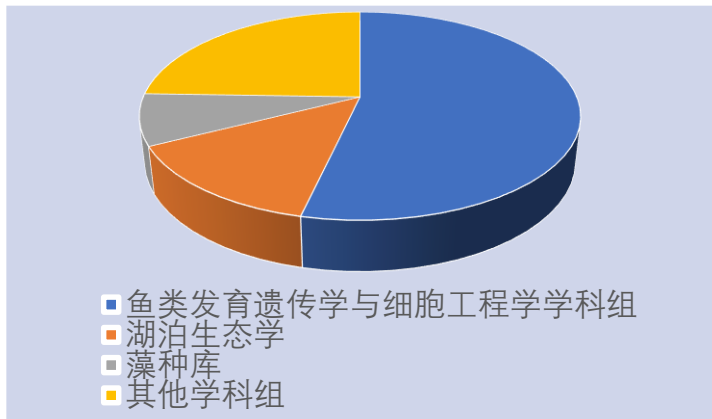
114TF

算力

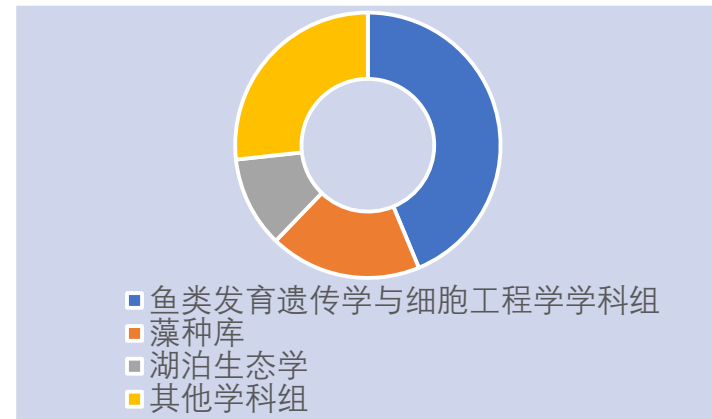
2023超算中心用户

| | |
|------|-----------|
| 用户 | 16 |
| 学科组 | 15 |
| 作业数 | 5002 |
| CPU时 | 553732.78 |

作业



CPU时



目录

- 1 中国科学院超级计算武汉分中心现状
- 2 经典水生生物形态学鉴定**
- 3 基于人工智能的水生生物形态学鉴定
- 4 基于eDNA技术的水生生物鉴定



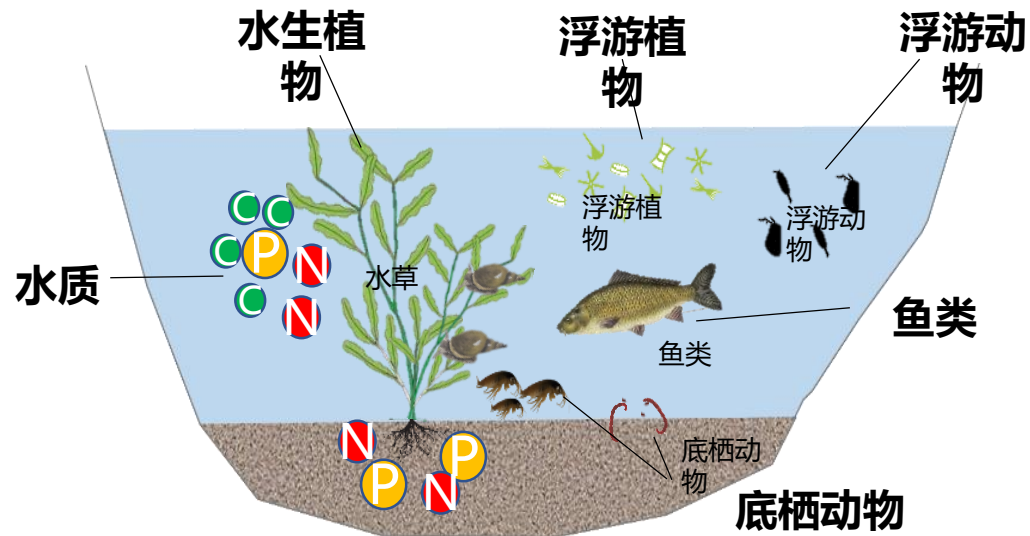


经典水生生物形态学鉴定：背景

开展水生态评估的必要性：

- 水生态综合评价是**落实习近平生态文明思想的必要举措**；
- 水生态综合评价是实现“十四五”期间“三水统筹”**生态环境综合治理的必然要求**；
- 水生态综合评价是保障《**中华人民共和国长江保护法**》**落地实施的必然需求**。

将水生态指标纳入考核指标体系，开展水生态综合评价，能夯实地方政府规划实施和水生态环境保护的主体责任，确保“十四五”规划各项保护措施落地生根，实现长江水生态明显向好，同时为“十五五”全面开展水生态考核探索**积累经验**，助力实现2035年美丽中国的远景规划。



种类组成、优势种、密度（或盖度）、生物量



经典水生生物形态学鉴定：方法



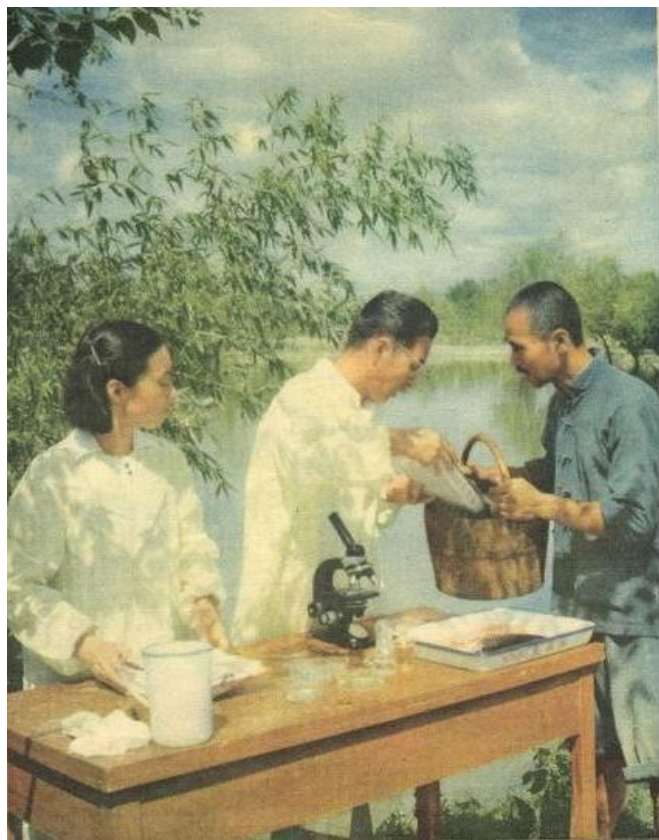
▲ 实验人员在显微镜下进行微型生物鉴定

▶ 鱼类照片以及部分关键特征

经典的分类学依据**关键形态特征**以及**比较解剖数据**，确定其的分类地位和系统关系。



经典水生生物形态学鉴定



1955年《人民画报》
4月号《防治鱼病》



沈蕴芬院士

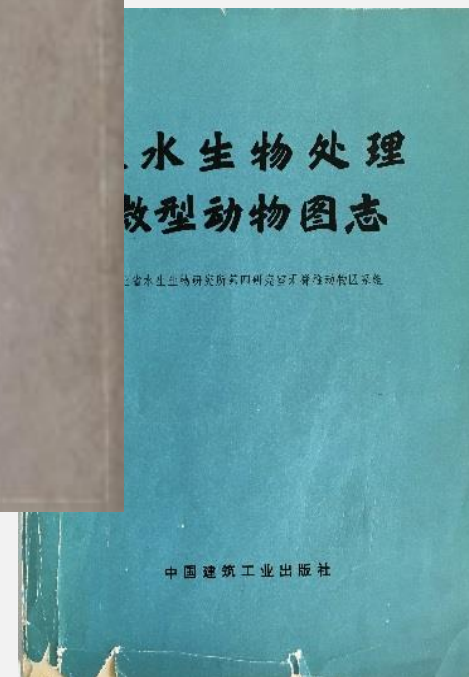
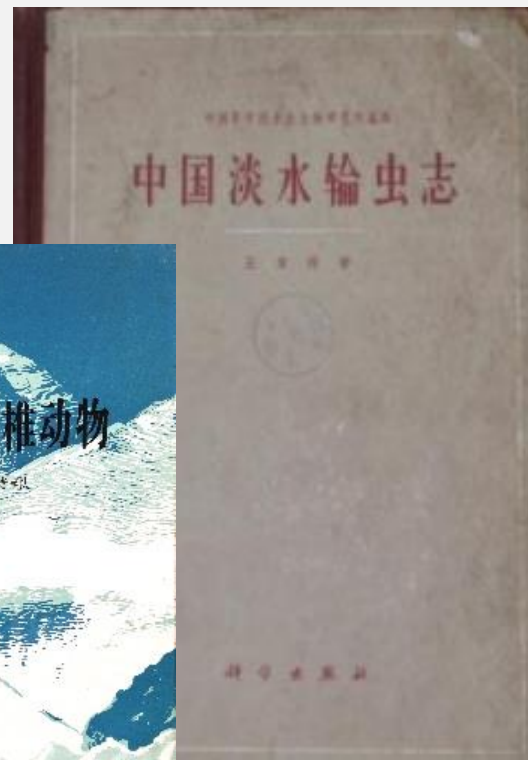
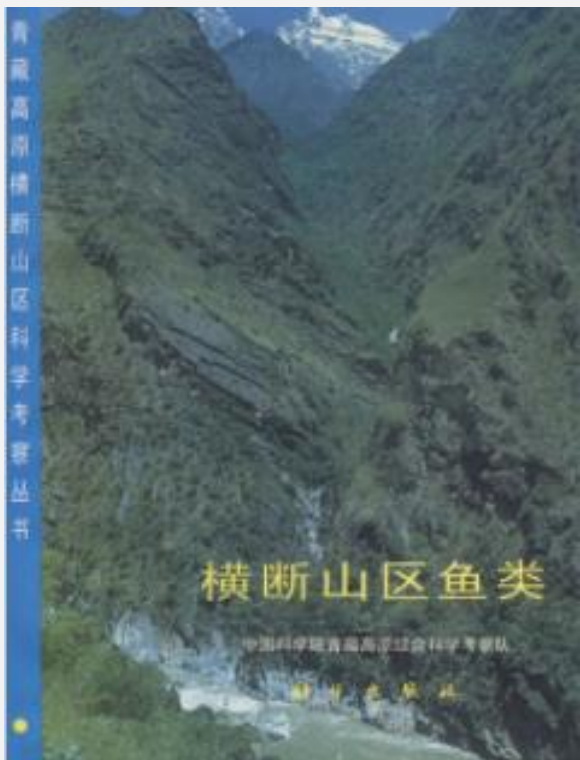


轮虫手绘图

水生所老一辈科学家为**水生生物分类鉴定**做出了卓远贡献。



经典水生生物形态学鉴定



发表的经典著作是水生生物分类鉴定的重要基础。



经典水生生物形态学鉴定

响应水生态综合评价需求

成立了专业的野外采样、物种鉴定
和数据分析团队

浮游动物 浮游植物 底栖动物
鱼类 水生维管束植物



样品采集 种类鉴定 数据分析 报告撰写



涉及水生生物门类齐全，技术水平专业
精湛，实验室管理科学规范





经典水生生物形态学鉴定

响应水生态综合评价需求

成立了专业的野外采样、物种鉴定
和数据分析团队

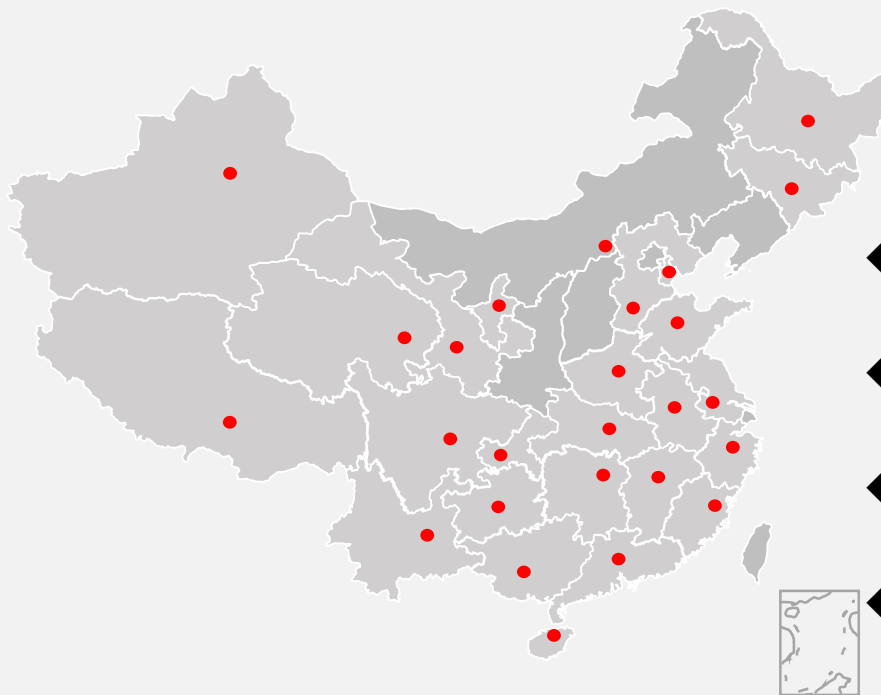
浮游动物 浮游植物 底栖动物
鱼类 水生维管束植物



样品采集 种类鉴定 数据分析 报告撰写



涉及水生生物门类齐全，技术水平专业
精湛，实验室管理科学规范



采集样点涉及省份

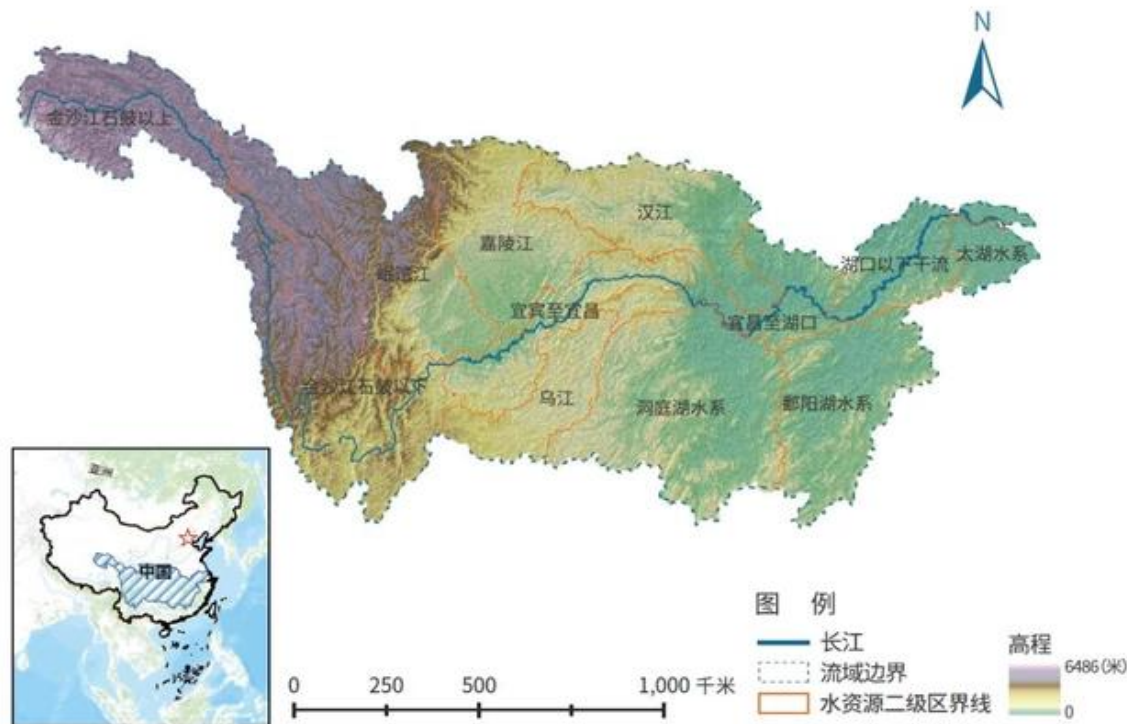
- ◆ 累计承接项目 **656** 个
- ◆ 采样点涉及 **26** 个省/直辖市
- ◆ 涉及水域 **400余** 个
- ◆ 样品 **40000** 余瓶

注：数据统计截至2023年11月25日



经典水生生物形态学鉴定

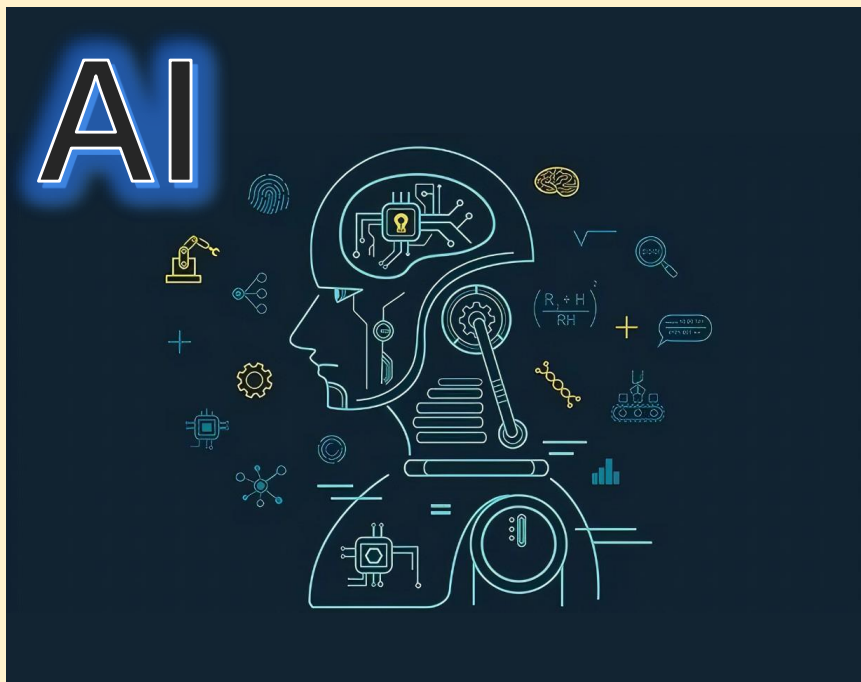
- 我国“十四五”规划纲要中，明确未来我国要加强生物监测，水环境管理要逐步从**水质目标向水生态转变**，水生态监测和保护成为当今发展的焦点。
- 传统生物调查和监测方法存在**成本高、效率低、费人力和破坏性**等问题。为应对水生态考核要求，**迫切需求研发新一代水生生物鉴定方法**。





经典水生生物形态学鉴定

浮游生物人工智能识别



形态学

环境DNA技术



分子生物学

在此背景下平台布局了**面向未来**的两个项目

目录

- 1 中国科学院超级计算武汉分中心现状
- 2 经典水生生物形态学鉴定
- 3 基于人工智能的水生生物形态学鉴定**
- 4 基于eDNA技术的水生生物鉴定





基于人工智能的水生生物形态学鉴定

人工智能识别关键步骤

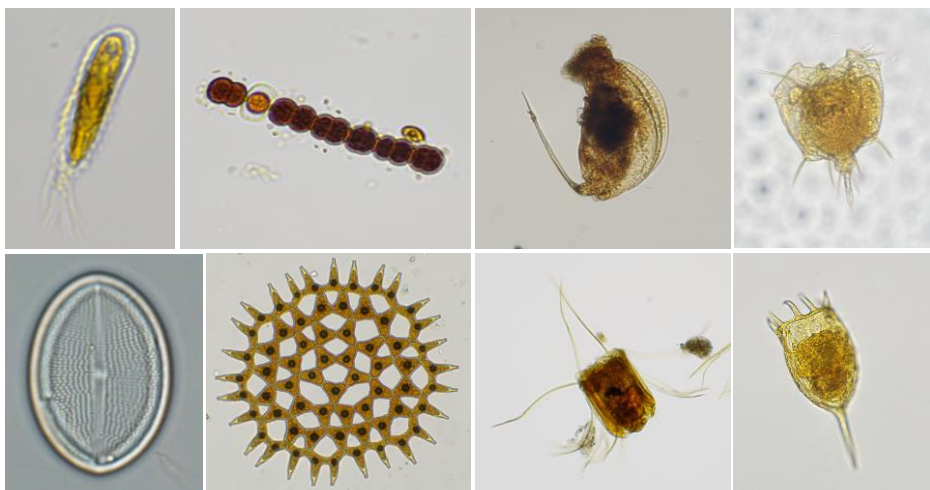




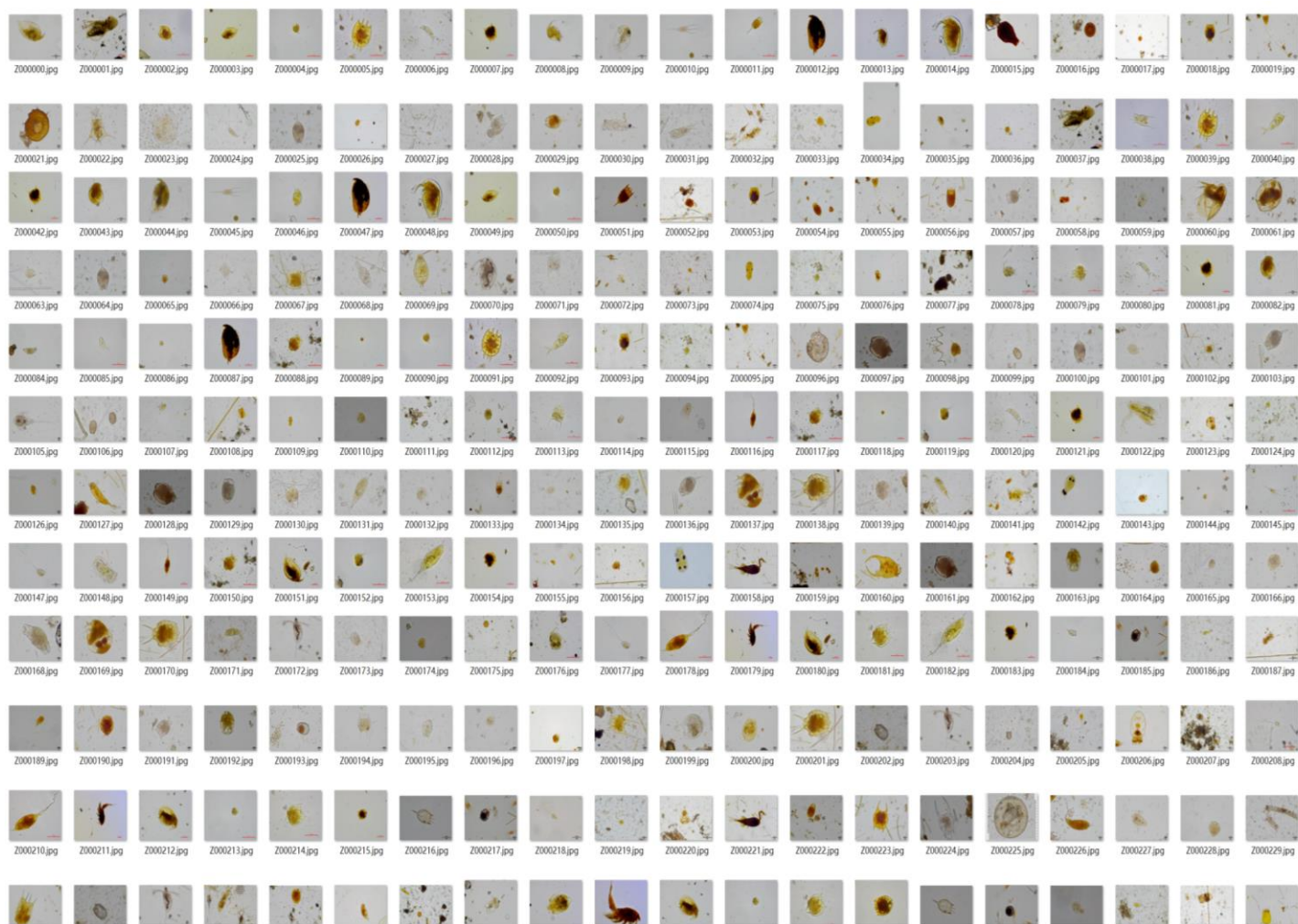
基于人工智能的水生生物形态学鉴定

数据库建设

- 收集照片：超过**12万**张
- 标记照片：超过**4万**张
- 浮游动物已标记类别：**337**种
- 浮游植物已标记类别：**244**种



典型照片示例



照片数据库（部分）



基于人工智能的水生生物形态学鉴定

算法设计

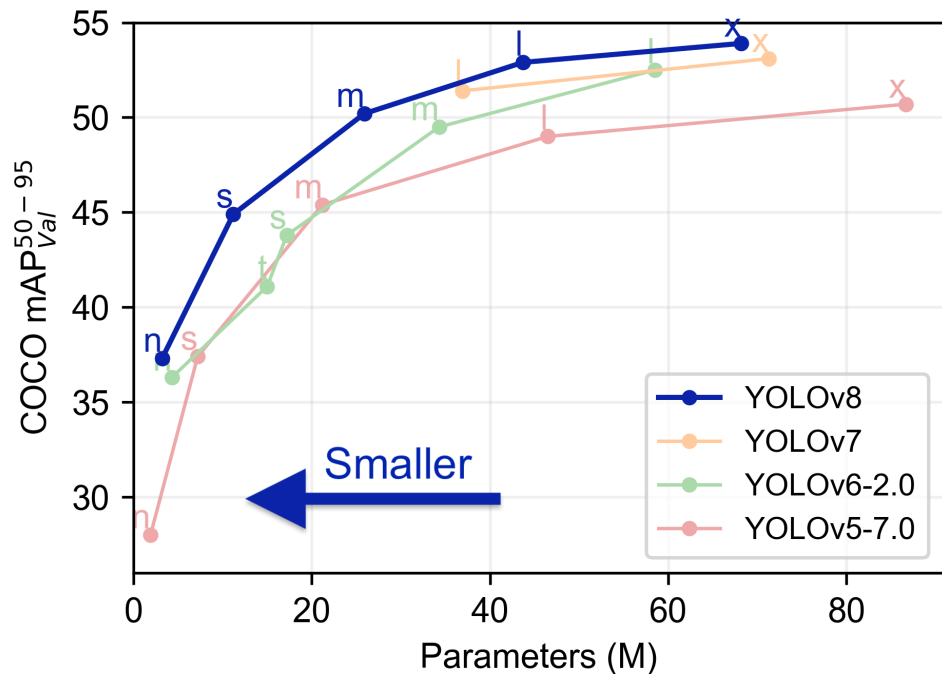
English | 简体中文

Ultralytics CI passing DOI [10.5281/zenodo.7347926](https://doi.org/10.5281/zenodo.7347926) docker pulls 10

Run on Gradient Open in Colab Open in Kaggle

Ultralytics YOLOv8, developed by Ultralytics, is a cutting-edge, state-of-the-art (SOTA) model that builds upon the success of previous YOLO versions and introduces new features and improvements to further boost performance and flexibility. YOLOv8 is designed to be fast, accurate, and easy to use, making it an excellent choice for a wide range of object detection, image segmentation and image classification tasks.

To request an Enterprise License please complete the form at [Ultralytics Licensing](#).

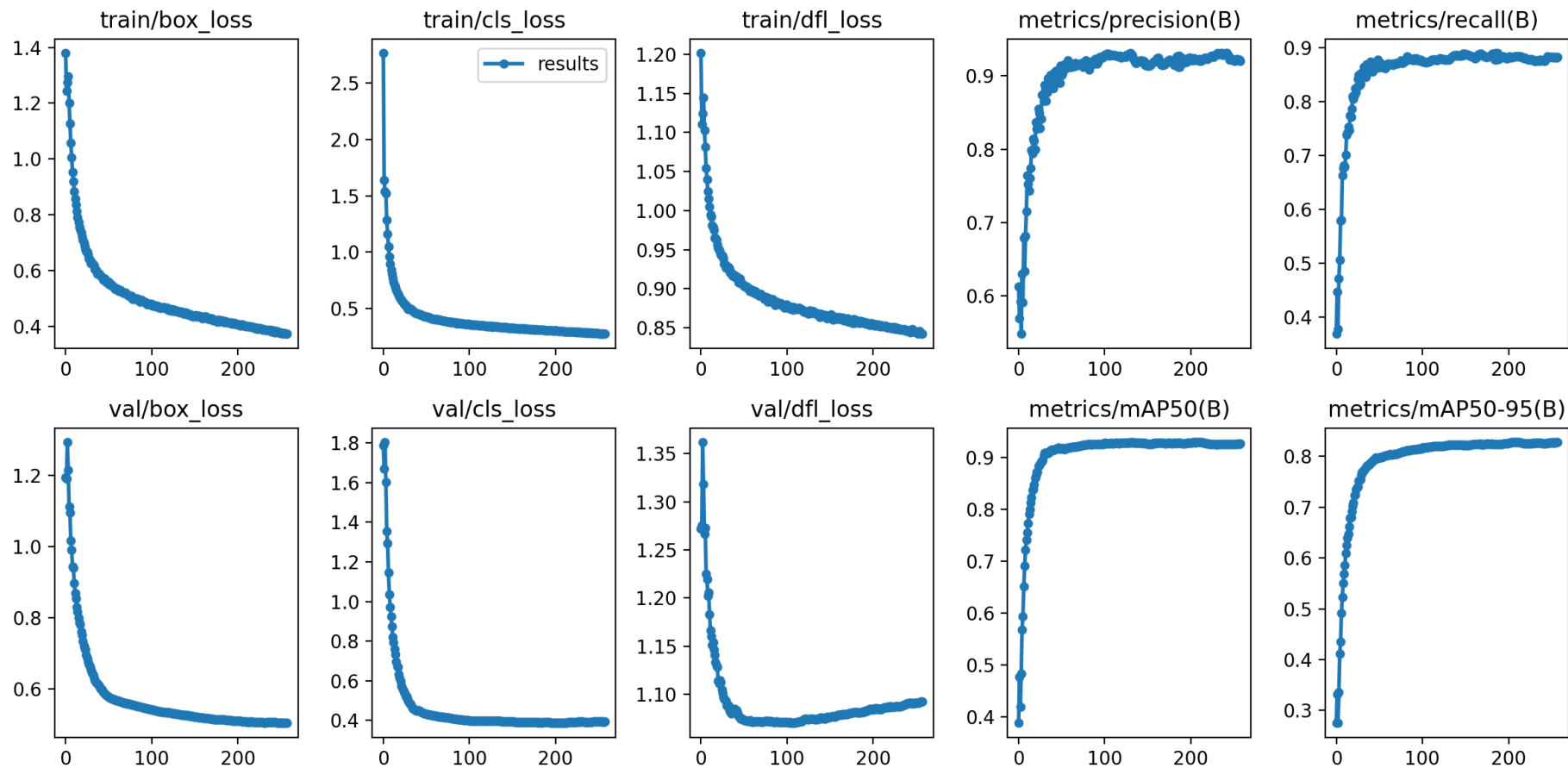


经过广泛测试后选择**YoloV8**作为浮游生物人工智能识别算法



基于人工智能的水生生物形态学鉴定

模型训练



Yolov8x训练结果



基于人工智能的水生生物形态学鉴定

模型训练

YOLOv8不同模型网络参数

| 模型 | 网络层数/Layers | 参数量/Parameters(百万) | 浮点运算数/GFLOPs |
|---------|-------------|--------------------|--------------|
| yolov8n | 225 | 3.82 | 12 |
| yolov8s | 225 | 11.41 | 29.9 |
| yolov8m | 295 | 25.94 | 79.5 |
| yolov8l | 268 | 43.72 | 165.4 |
| yolov8x | 268 | 68.26 | 258.1 |



基于人工智能的水生生物形态学鉴定

模型训练

浮游生物yolov8模型准确率

| 模型 | 精确率/P | 召回率/R | mAP50 |
|----------------|--------------|--------------|--------------|
| yolov8n | 0.831 | 0.776 | 0.865 |
| yolov8s | 0.820 | 0.767 | 0.852 |
| yolov8m | 0.908 | 0.804 | 0.851 |
| yolov8l | 0.771 | 0.714 | 0.799 |
| yolov8x | 0.968 | 0.959 | 0.981 |

nc=142

备注：精确率、召回率和mAP50等均基于测试数据集计算所得。
包括的类别(num classes)为142。
以上模型单张照片推理时间均低于3.6ms。



基于人工智能的水生生物形态学鉴定

模型训练

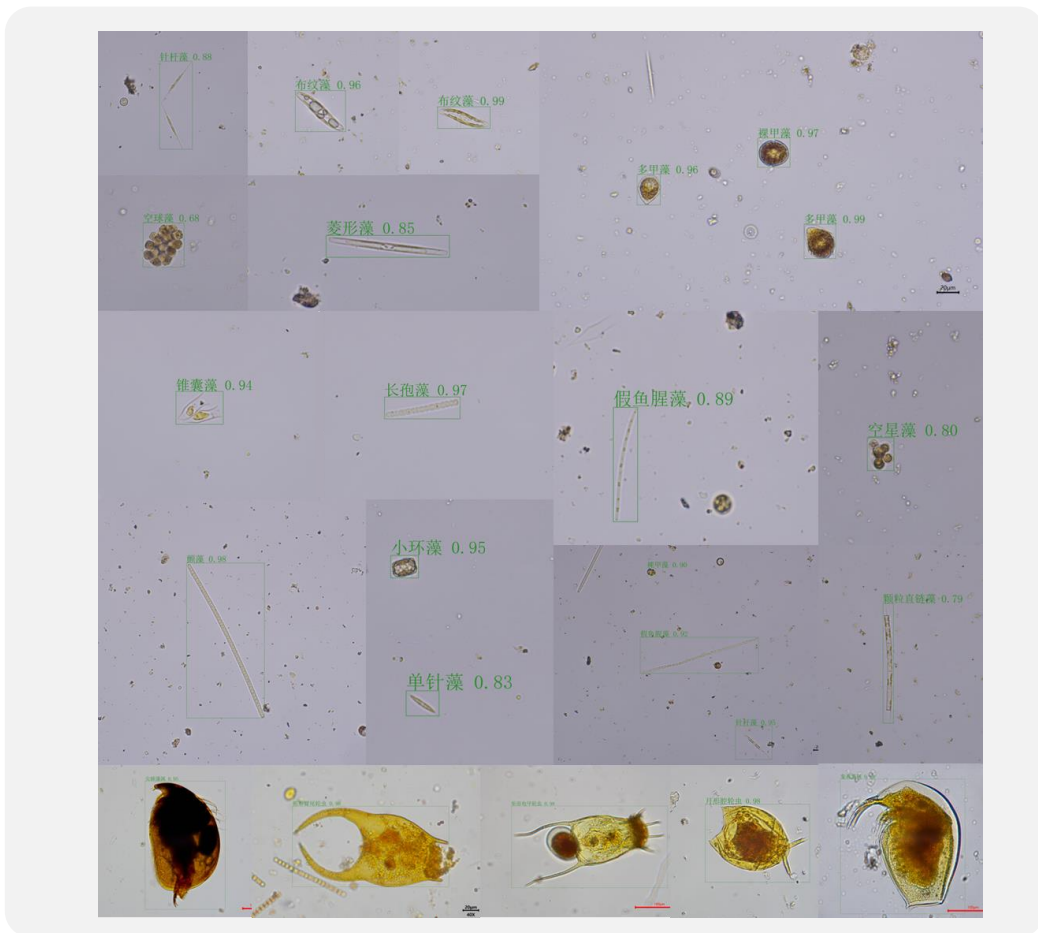
浮游生物yolov8模型准确率

| 模型 | 精确率/P | 召回率/R | mAP50 | |
|------------------|--------------|--------------|--------------|--------|
| yolov8n | 0.831 | 0.776 | 0.865 | nc=142 |
| yolov8s | 0.820 | 0.767 | 0.852 | |
| yolov8m | 0.908 | 0.804 | 0.851 | |
| yolov8l | 0.771 | 0.714 | 0.799 | |
| yolov8x | 0.968 | 0.959 | 0.981 | |
| ★ yolov8x | 0.958 | 0.938 | 0.968 | nc=242 |

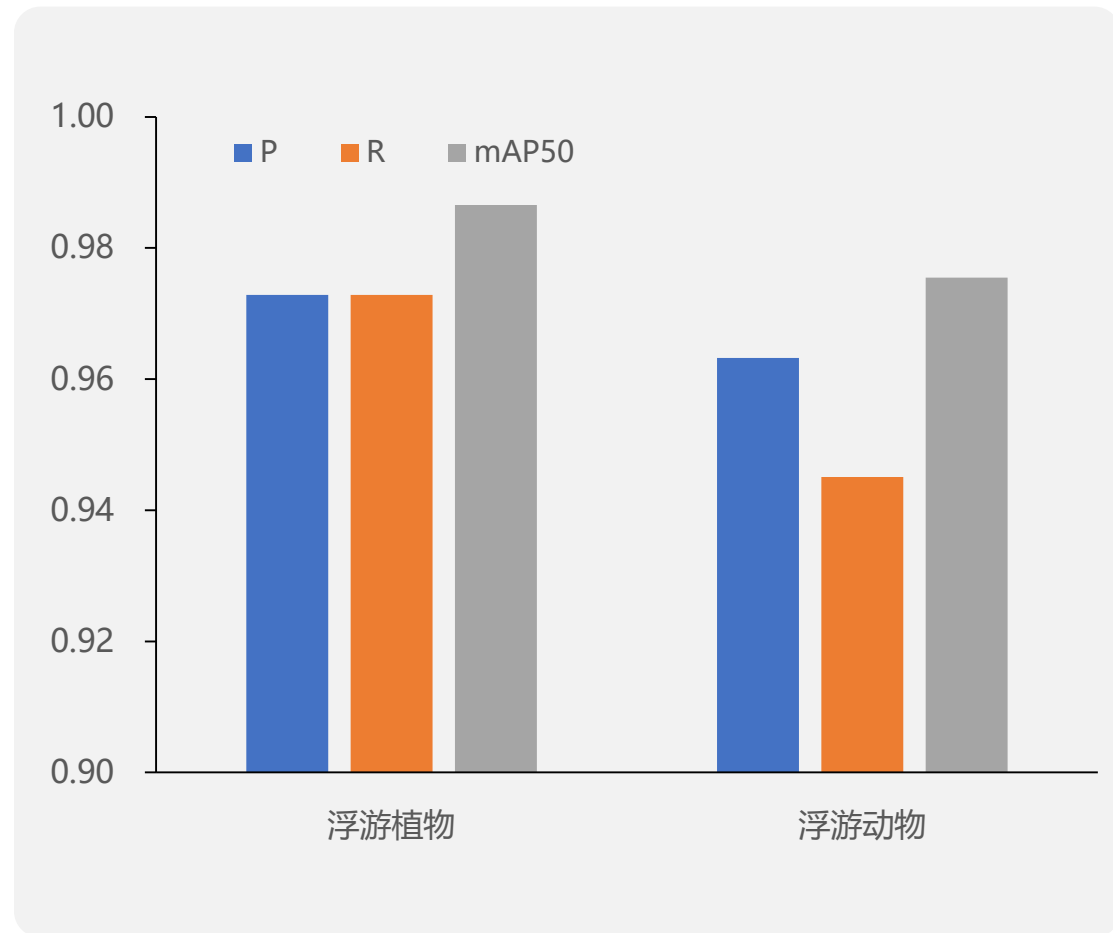


基于人工智能的水生生物形态学鉴定

模型识别案例



浮游动物植物识别案例

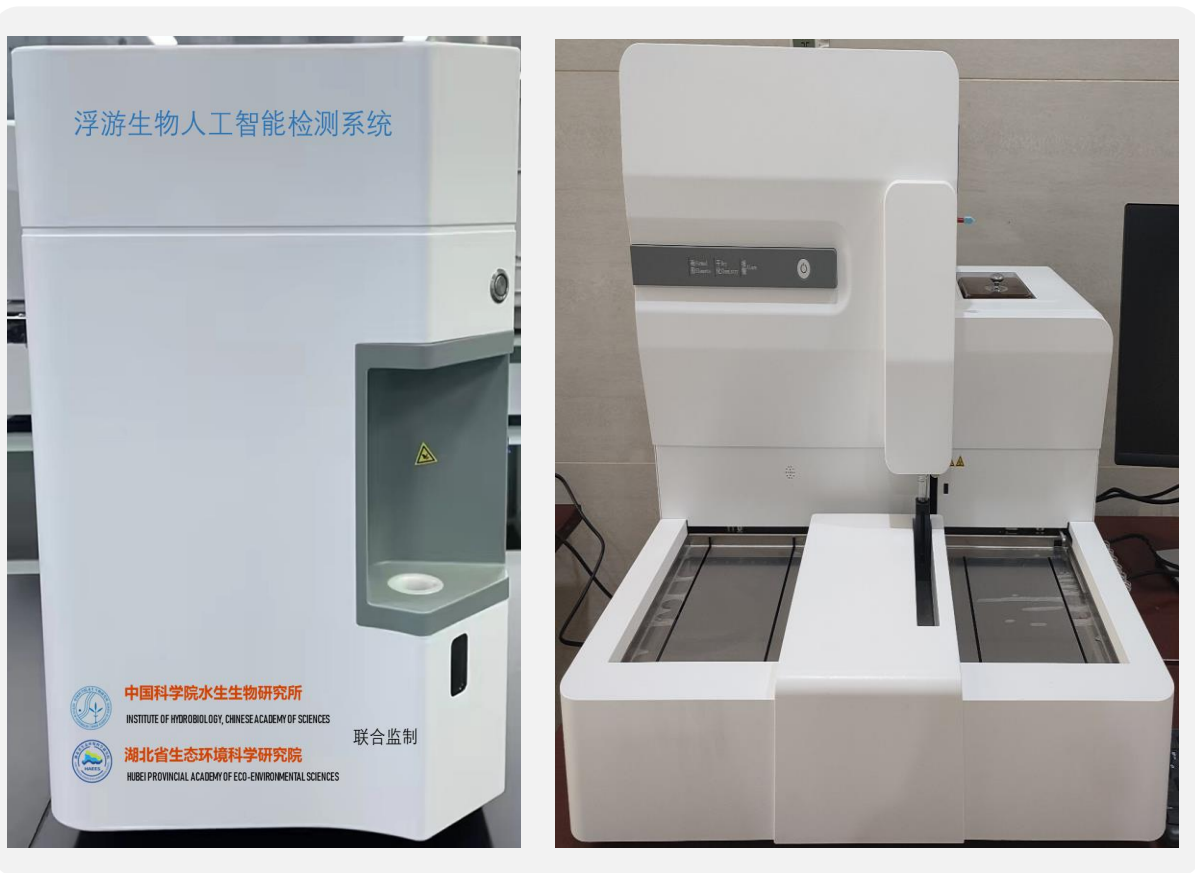


浮游动物植物准确率参数

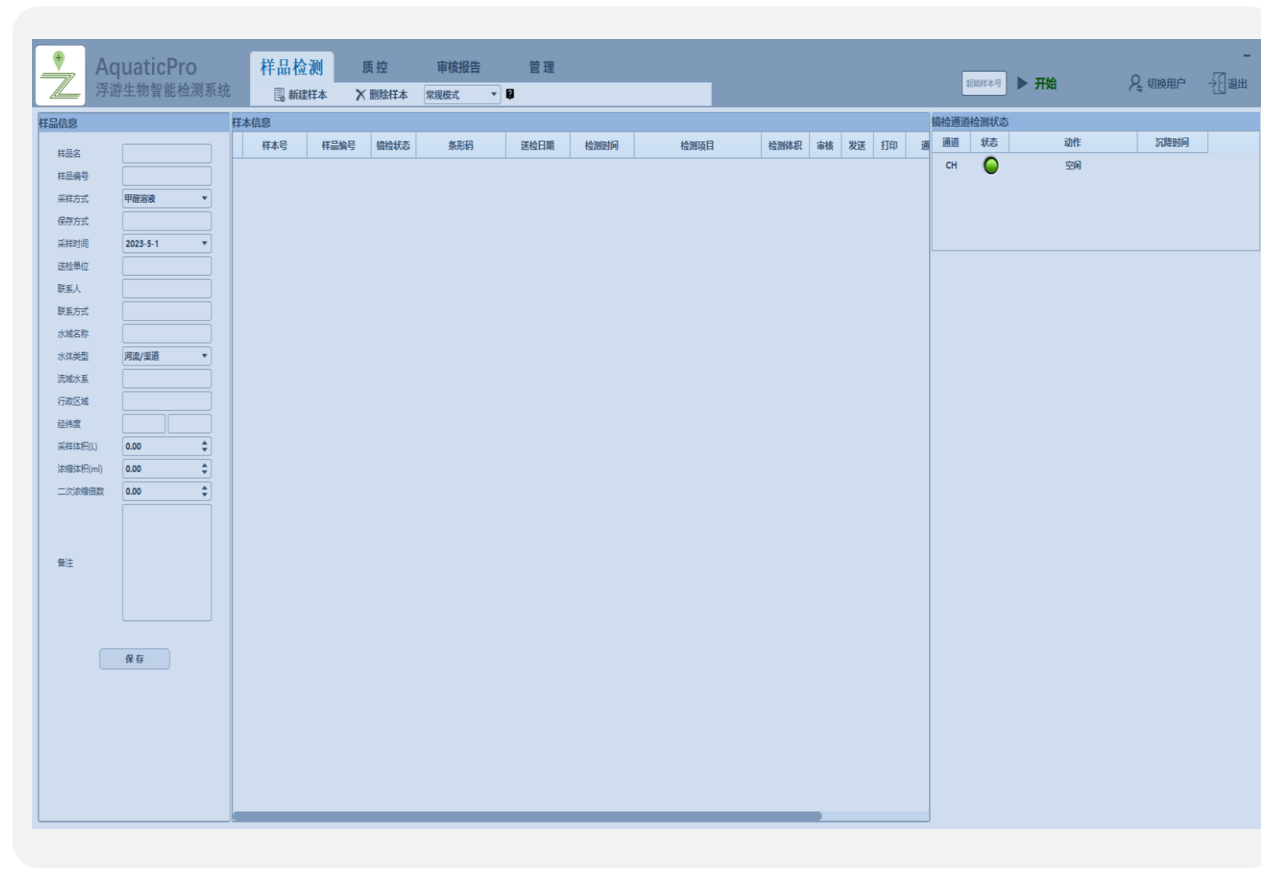


基于人工智能的水生生物形态学鉴定

设备研发



浮游生物智能检测设备



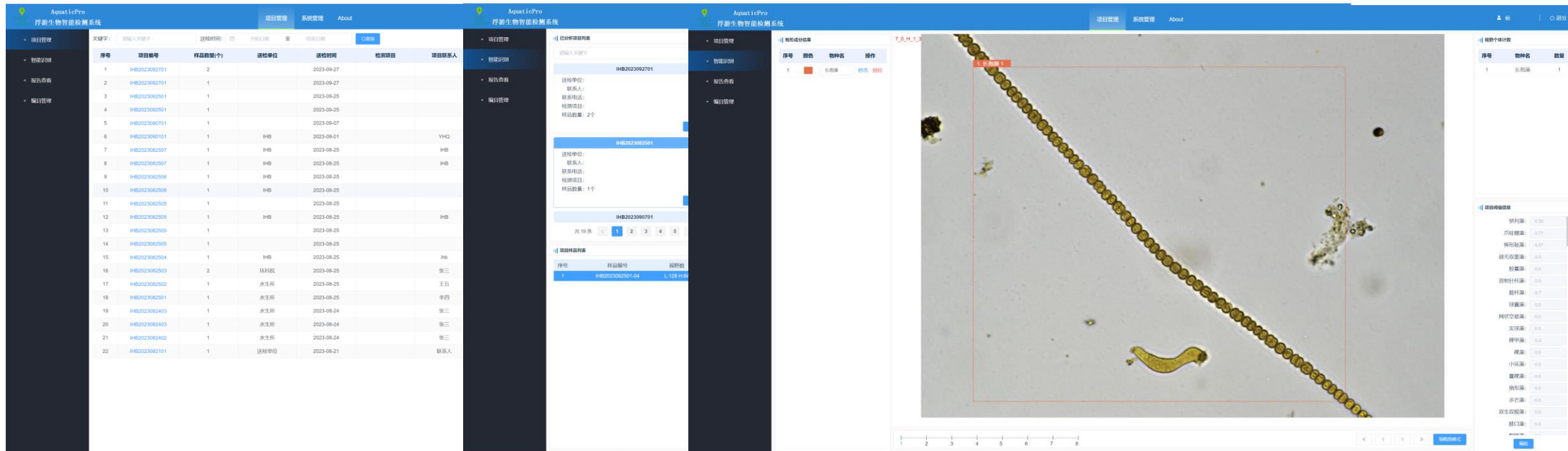
设备控制界面

整体实现从进样、智能识别到形成报告的全程自动化



基于人工智能的水生生物形态学鉴定

设备研发



项目管理、结果查看、审核界面

整体实现从进样、智能识别到形成报告的全程自动化

目录

- 1 中国科学院超级计算武汉分中心现状
- 2 经典水生生物形态学鉴定
- 3 基于人工智能的水生生物形态学鉴定
- 4 基于eDNA技术的水生生物鉴定**



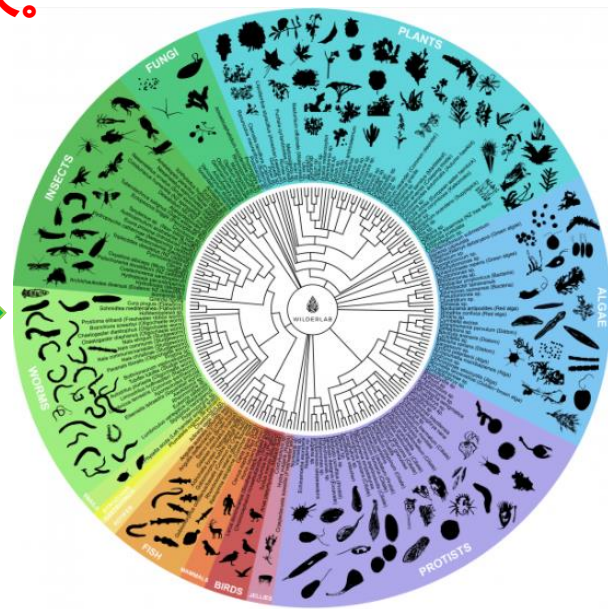
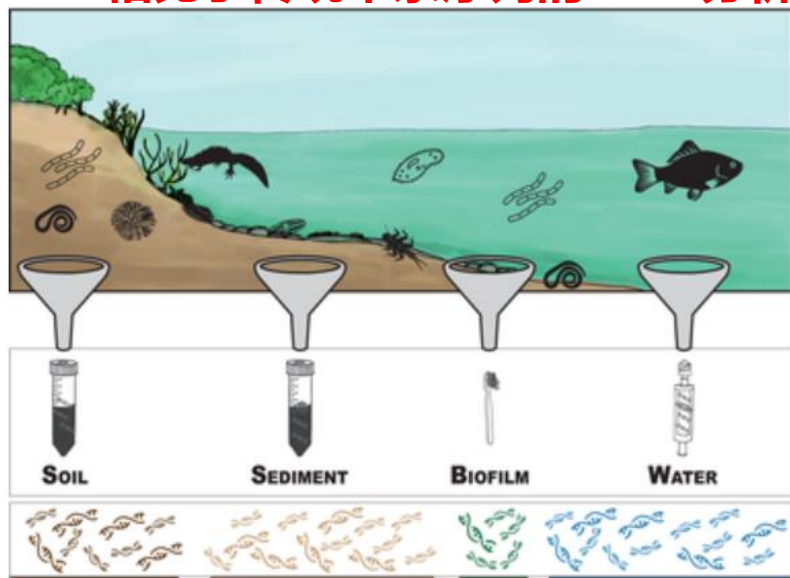


基于eDNA技术的水生生物鉴定

环境DNA (environmental DNA, eDNA) 是指在环境样品中所有被发现的不同生物的DNA片段的总和。

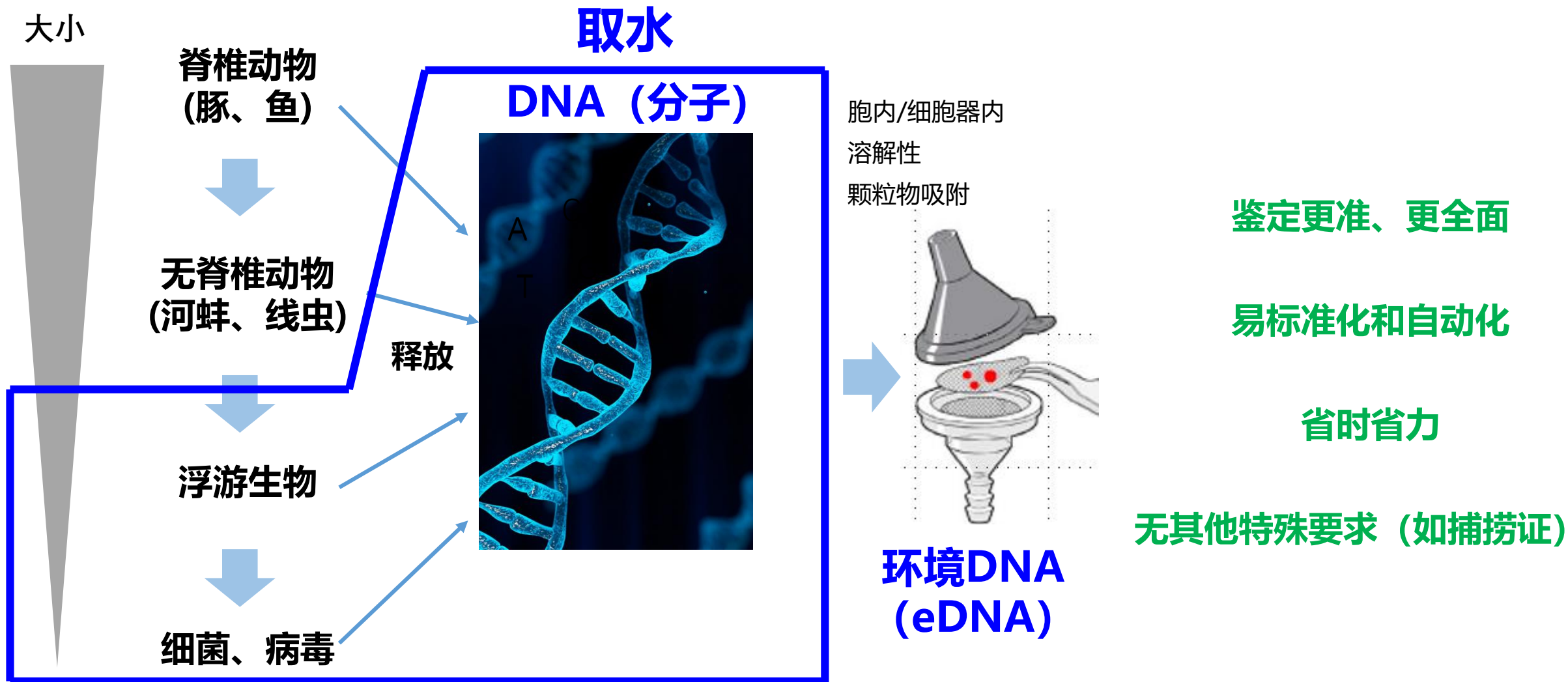
通过从环境介质 (水、土壤、沉积物等) 中**提取DNA**，对其中的特定DNA片段或是基因组进行**PCR扩增**和**高通量测序**，并根据已有的DNA数据库进行**比对和注释**，从而实现**对生物样品的监测**。

相比于传统单条序列的DNA分析方法，eDNA技术需要的算力需求成几何式增长。





基于eDNA技术的水生生物鉴定



环境DNA技术将是未来水生生物多样性监测和保护的最关键技术之一



基于eDNA技术的水生生物鉴定

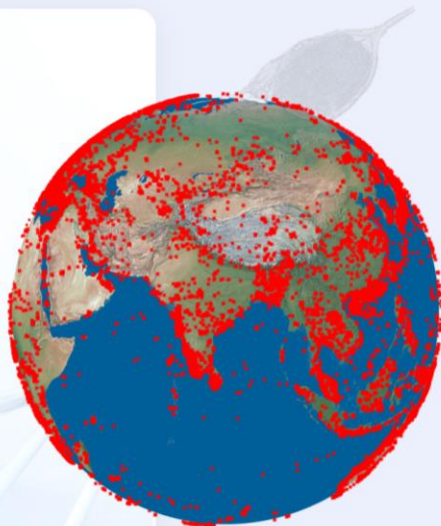


水生生物eDNA数据库 (Aquatic environmental DNA database, AeDNA)

AeDNA是由水生生物DNA条形码和基因组组成的eDNA数据库，涉及生物类群包括2万余种鱼类、1万余种水生植物、1万余种浮游动物和1万余种浮游植物；生境类型覆盖江、河、湖、海、冰川和温泉等各类水环境；数据类型包含18S、28S、ITS、CO1、12S、rbcL等条形码序列以及线粒体、叶绿体等细胞器基因组和物种全基因组序列。

数据库收集了近60万余条eDNA序列，其中AeDNA团队贡献了6万余条。AeDNA依托中国科学院水生生物研究所水生生物鉴定中心、水生生物数据分析管理平台、超级计算平台和湖北省生态环境科学研究院，旨在实现水生生物的调查、监测、追溯和预警的综合能力。

| | DNA条形码 | 基因组* |
|--|---------|-------|
|  浮游植物 | 34,427 | 1,119 |
|  水生植物 | 103,227 | 1,586 |
|  浮游动物 | 67,690 | 49 |
|  底栖动物 | 11,214 | 327 |
|  鱼类 | 376,201 | 3,118 |
|  两栖动物 | 建设中 | 建设中 |
|  其他 | 建设中 | 建设中 |



AeDNA共收录60余万条eDNA序列，其中水生所团队提供6万条，总收录物种5.6万种，包含几乎所有的鱼类和水生植物。

**包含浮游植物、水生植物、浮游动物、底栖动物、鱼类等各类水生生物；
覆盖江、河、湖、海等各类水体环境**



基于eDNA技术的水生生物鉴定

1. 项目团队通过物种形态鉴定并测序获得的数据:

- 照片、meta信息、序列、标本、种质资源



2. 公共数据库

- **权威公共数据库**: NCBI (全球最大的核酸数据库)、PR2 (原生生物)、BOLD (后生动物)
- **取高质量序列 (去冗余)**: 完整度高、聚类关系、meta信息全、来自权威团队
- **分类学家校对、更正错误**: 结合NCBI taxonomy和P10K计划对物种分类阶元进行修正, 对模糊的分类阶元进行标准化和重定义
- **标准化**: 序列、物种分类、meta信息



藻类和原生动物种质资源库
National Culture Collection of Algae and Protozoa



基于eDNA技术的水生生物鉴定

核心成员



何舜平 研究员

中国科学院水生生物研究所
鱼类系统发育和生物地理学、比较基因组学



缪伟 研究员

中国科学院水生生物研究所
原生动功能基因组学



李涛 研究员

中国科学院水生生物研究所
藻类组学与环境宏基因组学



曾宏辉 高级工程师

中国科学院水生生物研究所
水生生物数据分析管理



陈晓飞 高级工程师

湖北省生态环境科学研究院
水生态环境保护规划与水污染治理

数据管理团队



熊杰 研究员

浮游生物
xiongjie@ihb.ac.cn



陈凯 博士

浮游生物
chenkai@ihb.ac.cn



熊凡 博士

鱼类
xiongf@ihb.ac.cn



吴志刚 博士

水生植物
wuzg@ihb.ac.cn



方成池 博士

鱼类
fangcc@ihb.ac.cn

贡献团队



期待你们的加入



刘焕章 研究员

中国科学院水生生物研究所
鱼类生态学与保护生物学



王洪铸 研究员

中国科学院水生生物研究所
无脊椎动物学、河流-泛滥平原生态学

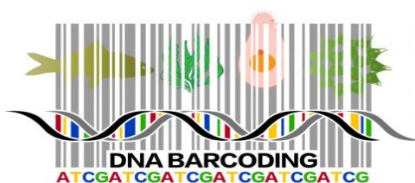
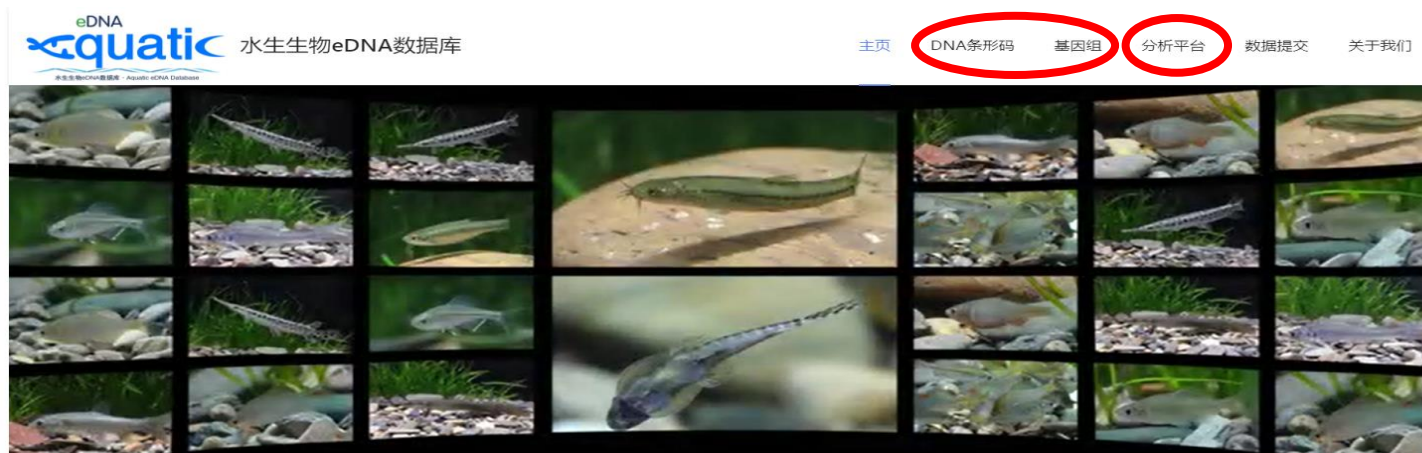


宋立荣 研究员

中国科学院水生生物研究所
蓝藻水华生态学和生物学



基于eDNA技术的水生生物鉴定



水生生物eDNA数据库 (Aquatic environmental DNA database, AeDNA)

AeDNA是由水生生物DNA条形码和基因组组成的eDNA数据库，涉及生物类群包括2万余种鱼类、1万余种水生植物、1万余种浮游动物和1万余种浮游植物；生境类型覆盖江、河、湖、海、冰川和温泉等各类水环境；数据类型包含18S、28S、ITS、CO1、12S、rbcL等条形码序列以及线粒体、叶绿体等细胞器基因组和物种全基因组序列。

数据库收集了近60万余条eDNA序列，其中AeDNA团队贡献了6万余条。AeDNA依托中国科学院水生生物研究所水生生物鉴定中心、水生生物数据分析管理平台、超级计算平台和湖北省生态环境科学研究院，旨在实现水生生物的调查、监测、追溯和预警的综合能力。

样品采集

核酸提取

文库制备

高通量测序

数据处理

结果输出

华大智造



基于eDNA技术的水生生物鉴定



便携式环境eDNA采集仪

八通道水样过滤仪

eDNA自动采样设备与关键组件

主要功能：

- 实现水样**定量采集**
- 过滤后将样品膜及时密封，**避免交叉污染**
- 可结合平台eDNA测序、eDNA数据库方便快捷获得检测数据

主要优势：

- 保证eDNA**采样标准化**
- 提高eDNA**采样效率**
- 高度集成体积小，**易于携带**，减轻野外作业负担



基于eDNA技术的水生生物鉴定



便携式环境eDNA采集仪

八通道水样过滤仪

eDNA自动采样设备与关键组件

| 序号 | 项目名称 | 申请日 | 申请号 | 类型 |
|----|-----------------------------|------------|----------------|------|
| 1 | 一种环境DNA远程连续智能采集系统 | 2022-01-12 | 202210029592.4 | 发明 |
| 2 | 一种环境DNA的智能采集方法 | 2022-01-12 | 202210029618.5 | 发明 |
| 3 | 一种可用于环境DNA采样的弹匣式滤膜更换系统 | 2022-02-16 | 202210142610.X | 发明 |
| 4 | 一种可用于环境DNA采样的夹取式滤膜更换系统 | 2022-02-18 | 202210149471.3 | 发明 |
| 5 | 一种可用于环境DNA采样的轮盘式滤膜更换系统 | 2022-02-18 | 202210149460.5 | 发明 |
| 6 | 一种可用于环境DNA采样的提拉式滤膜更换装置 | 2022-02-21 | 202210156593.5 | 发明 |
| 7 | 基于自动回位的环境DNA采样的弹匣式滤膜更换装置及系统 | 2022-02-21 | 202210156614.3 | 发明 |
| 8 | 轮盘式环境DNA滤饼更换装置的采样方法、控制方法及系统 | 2022-02-18 | 202210149474.7 | 发明 |
| 9 | 用于环境DNA采样的夹取式滤膜更换方法及系统 | 2022-02-18 | 202210149466.2 | 发明 |
| 10 | 一种可用于环境DNA采样的弹匣式滤膜更换系统 | 2022-02-16 | 202210142650.4 | 发明 |
| 11 | 一种用于环境DNA采集的滤膜装载结构 | 2022-02-21 | 202220345867.0 | 实用新型 |

稳步推进了专利申请工作



基于eDNA技术的水生生物鉴定



研发团队在各种水体测试设备性能，并积极与潜在客户团队交流，确保设备满足专业人员实际作业需求



基于eDNA技术的水生生物鉴定



应用生态学

- 生物多样性评估
- 生态群落功能评估
- 环境污染物诊断
- 生态健康评价
- 污染物阈值推导



保护生物学

- 濒危、珍稀、隐匿物种的监测与评价



入侵生物学

- 外来入侵物种监测
- 入侵路径分析
- 潜在生态影响研究



生物监测

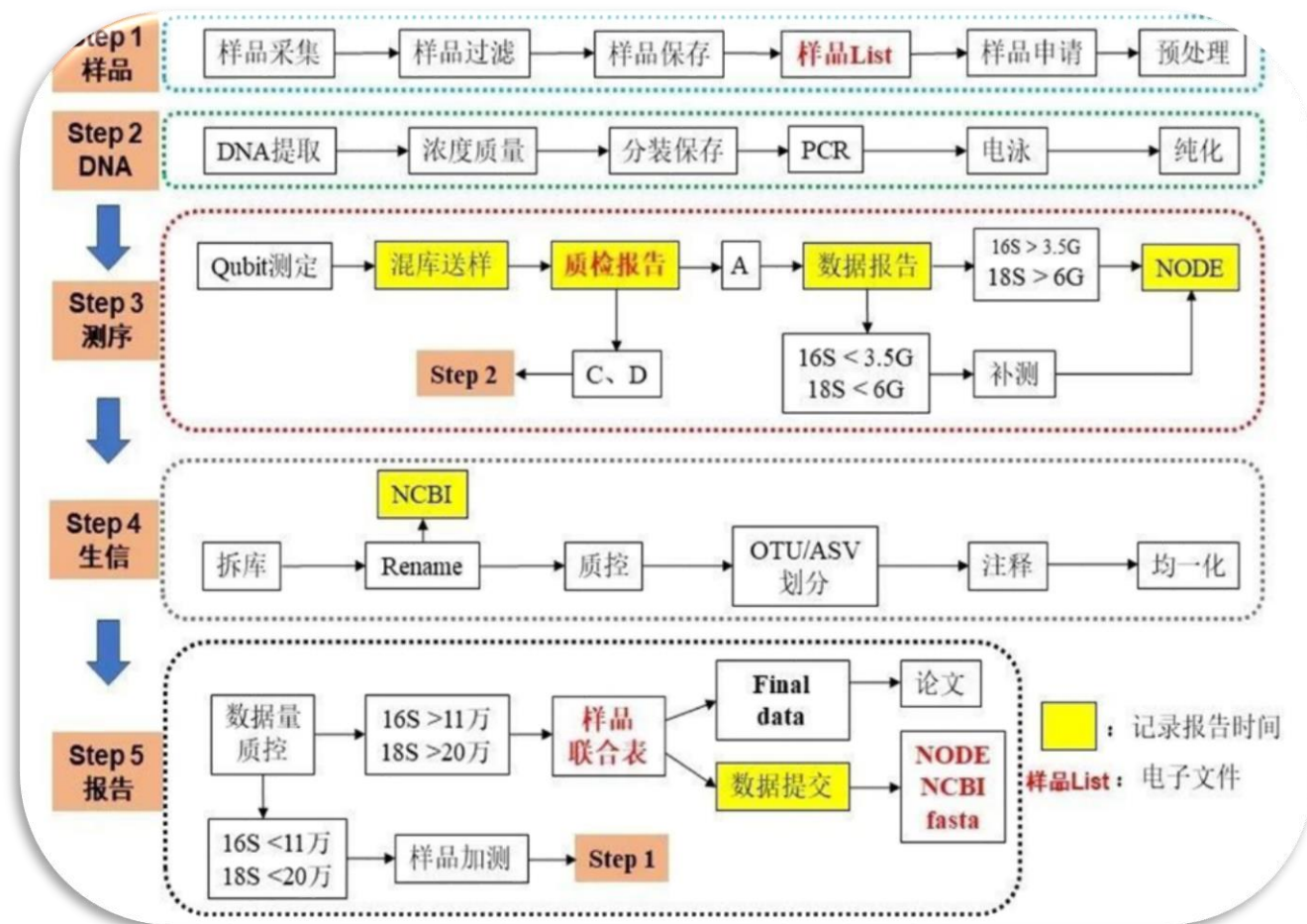
- 鱼类
- 藻类
- 浮游动物
- 底栖动物
- 微生物

环境DNA技术弥补了传统形态学生物监测的缺陷，显著提升了水生生物群落的监测能力，能大尺度、高频率、高灵敏度、自动化地获取生态监测信息，而且能准确地识别水生态系统的变化趋势，进而改变对水生态系统的认识与管理方式。



基于eDNA技术的水生生物鉴定

建立一套成熟的数据分析方法



助力实现一站式eDNA生物监测服务

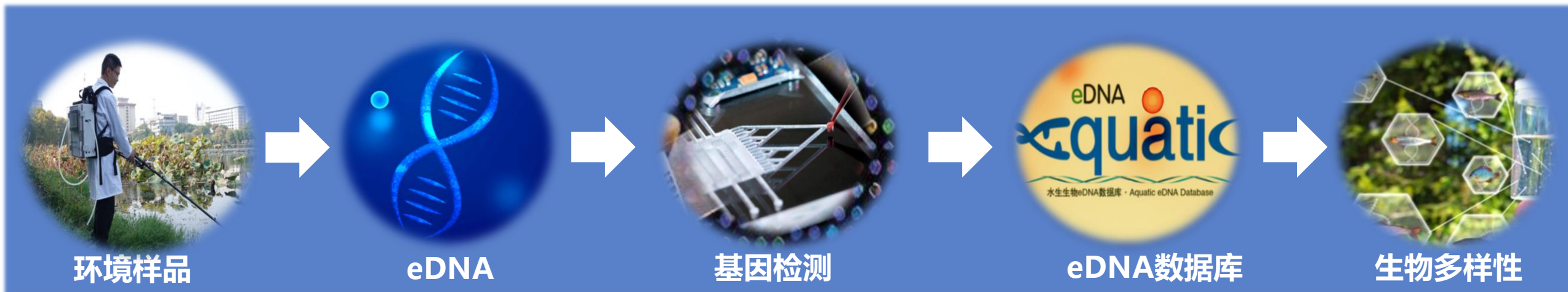
- 生态环境部长江流域生态环境监督管理局
- 中国检验检疫科学研究院
- 国家市场监督管理总局信息中心
- 湖北省生态环境科学研究院等相关部门

目前，利用该分析方法，已完成了整个长江流域第一阶段的资源调查和生态评估工作



基于eDNA技术的水生生物鉴定

掌握从设备到数据库再到算法的全部技术



开发快速连续
eDNA采集设备

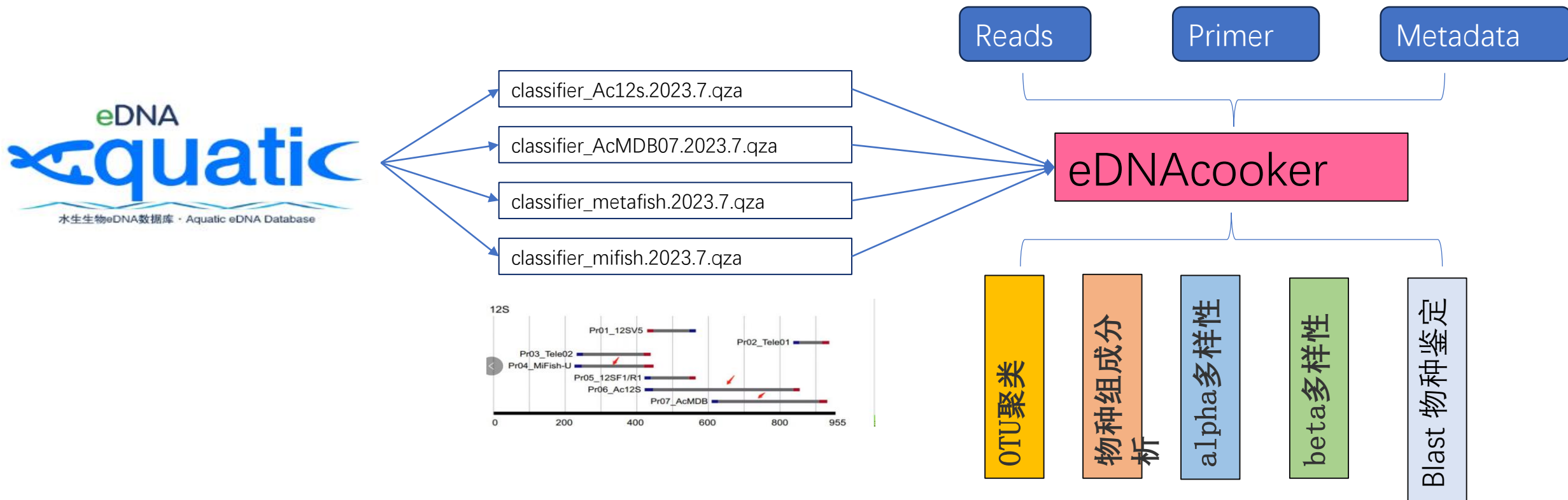
推进建设权威
eDNA检测标准

自建大型全面生
物DNA数据库

自研先进快速
物种鉴定算法



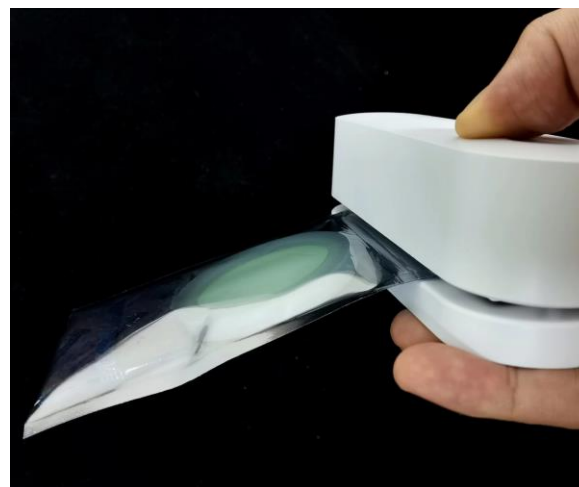
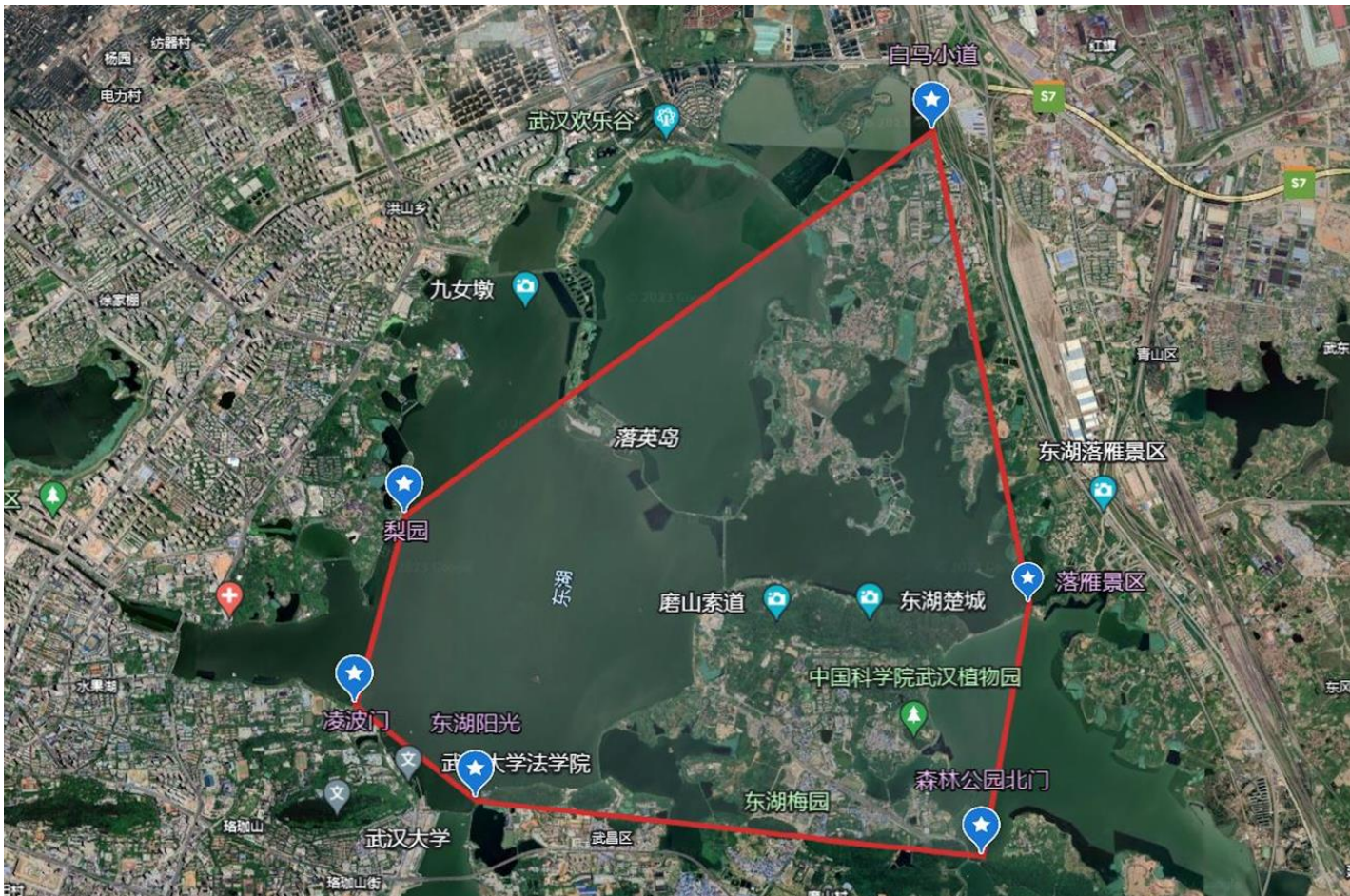
基于eDNA技术的水生生物鉴定



基于水生生物eDNA数据库，开发了一键式数据分析软件eDNAcooker。使用者只需要提供原始的测序序列、引物以及采样信息表格，就能快速的鉴定出物种的种类、组成成分以及多样性等，并提供可视化的图表分析结果。近期，该软件将会搭载到AeDNA网站上，方便eDNA的广泛使用。



基于eDNA技术的水生生物鉴定



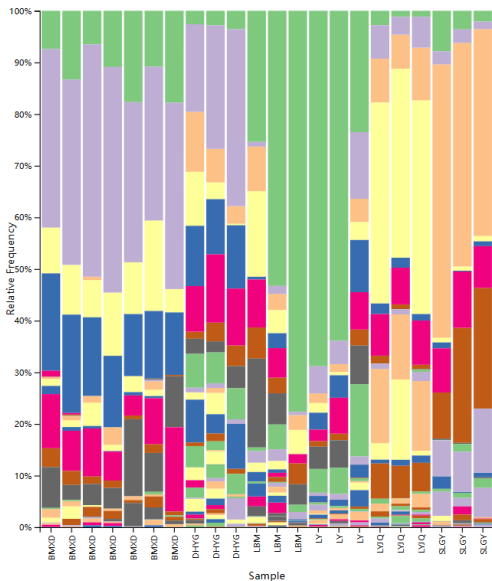
- 6个采样点
- 3-7重复/点

- 2套引物 (metafish, mifish)
- 44组eNDA数据

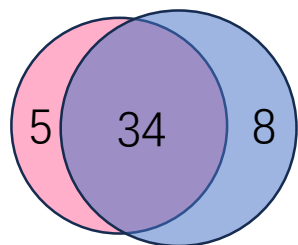
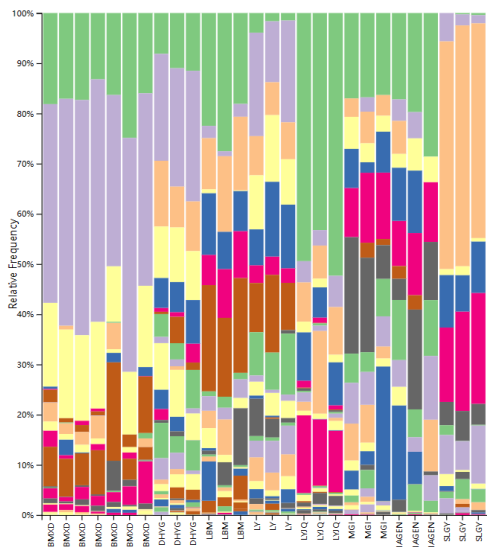


基于eDNA技术的水生生物鉴定

Metafish



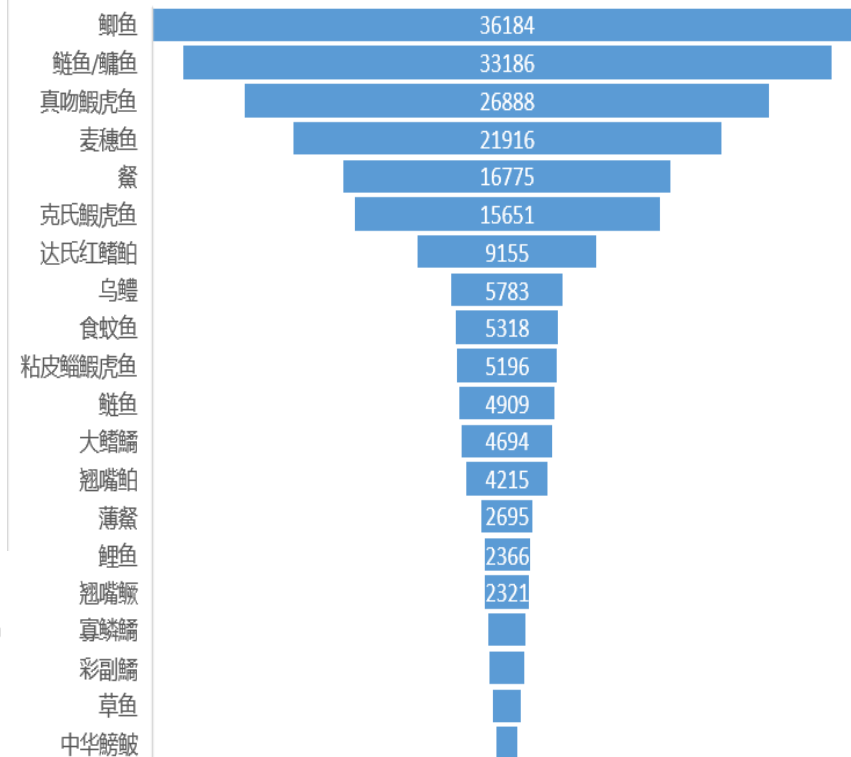
Mifish



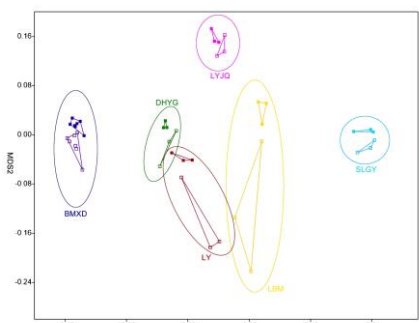
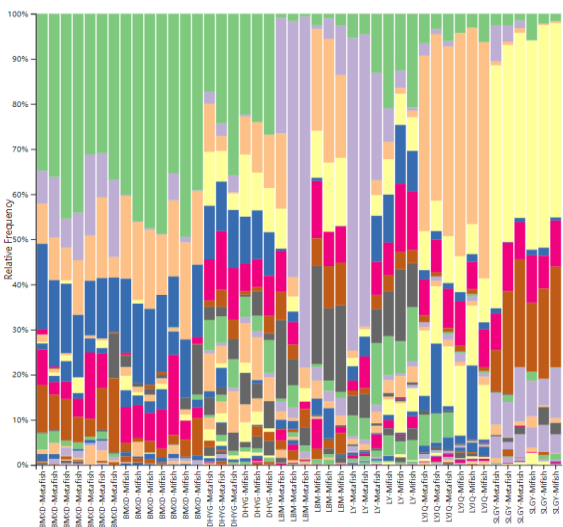
传统捕捞统计结果39种，非常接近

| Class | Order | Family | Genus | Species |
|-------|-------|--------|-------|---------|
| 1 | 8 | 11 | 31 | 47 |

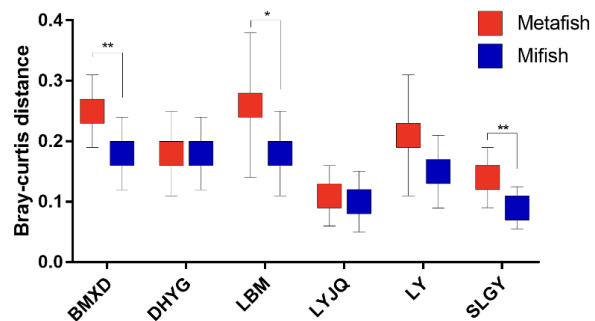
Top 20 of OUT



入侵物种(食蚊鱼, 齐氏罗非鱼)

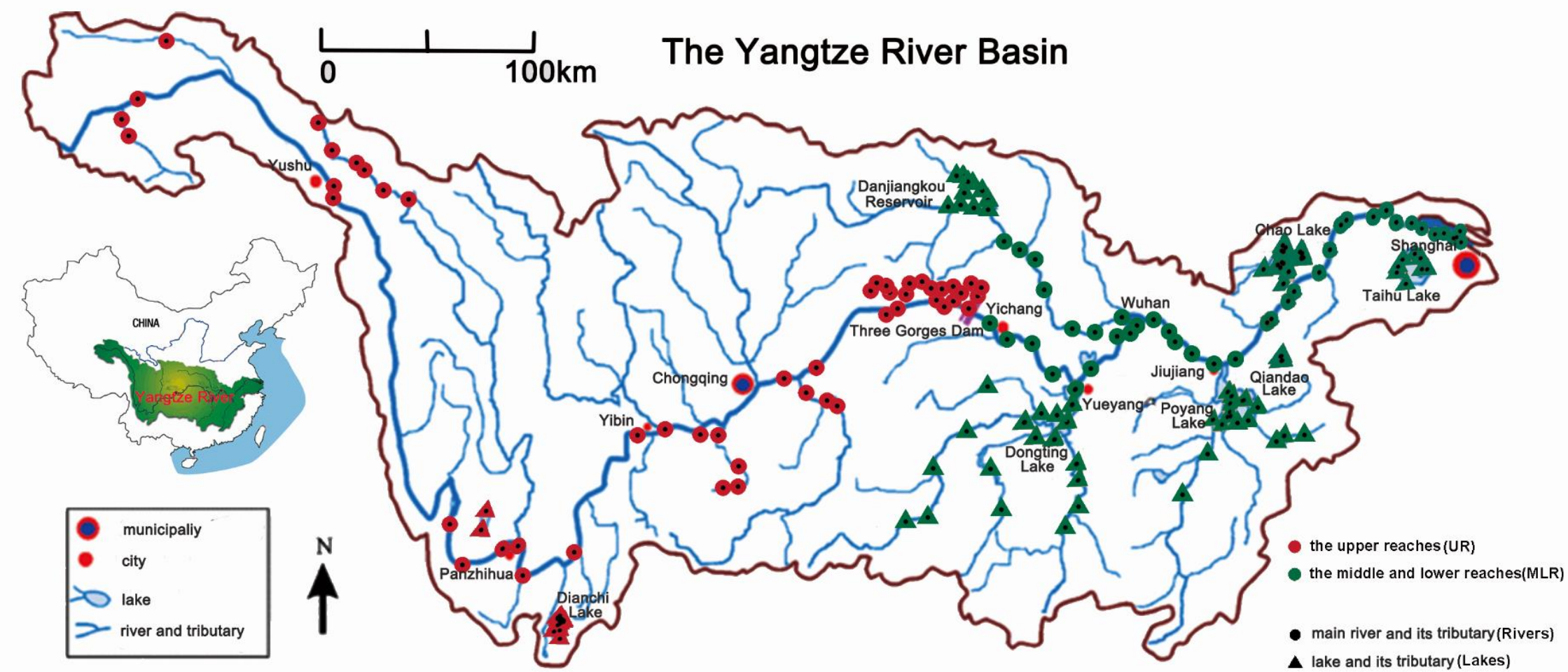


The Bray-curtis distance of inner groups





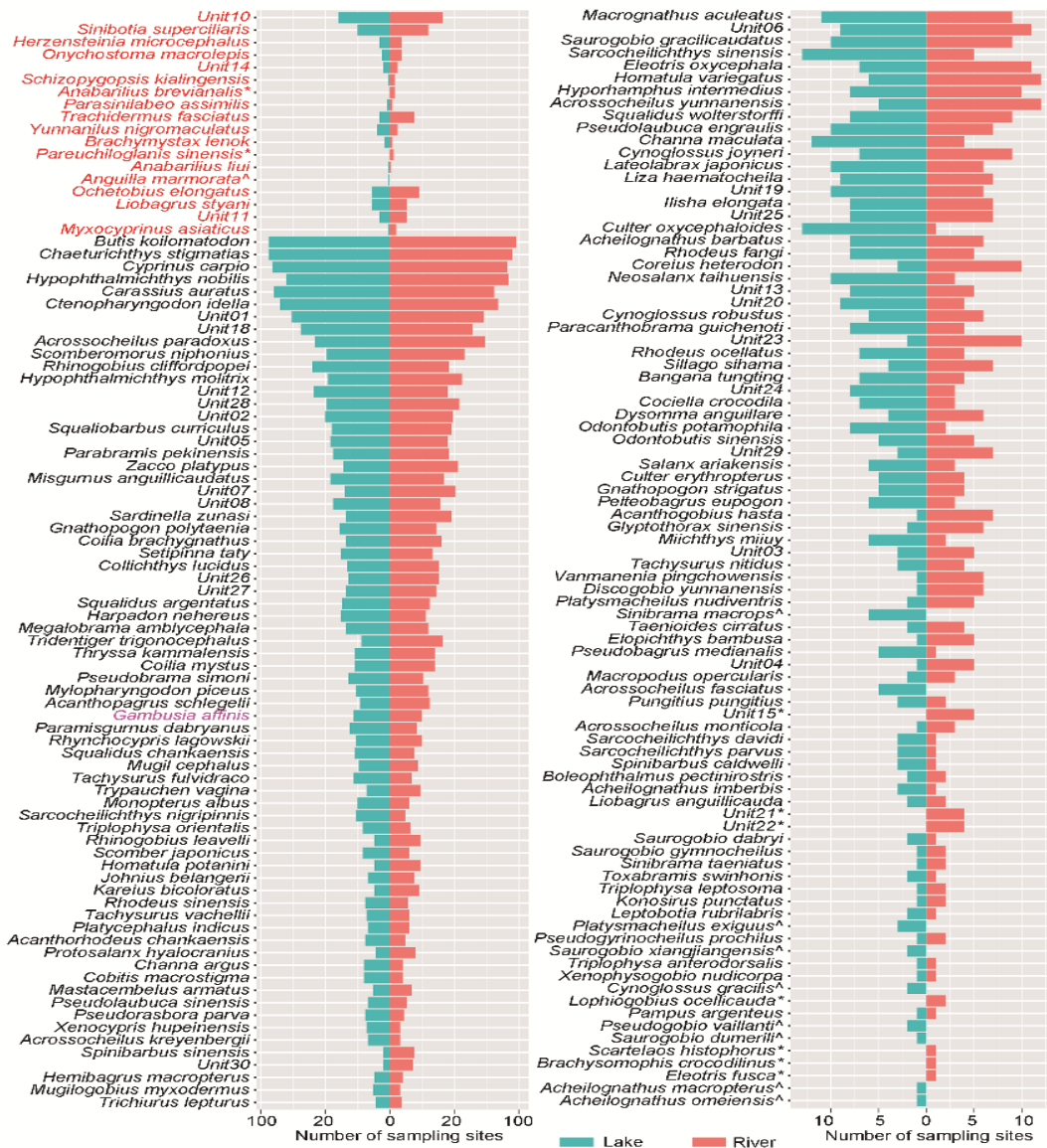
基于eDNA技术的水生生物鉴定



- 共194个样点，其中98个河流及其支流样点和96个湖泊及其支流样点
- 76个位于长江上游，118个位于长江中下游；
- 2L水样，1L用于环境DNA检测（引物Ac12S），1L用于水质检测



基于eDNA技术的水生生物鉴定



- 检测到**176**个分类单元 (其中**27**个由多物种组成, 共代表近**234**种鱼类)
- **158**个分类单元为河流和湖泊共有, **9**个仅在河流组存在, 多为鲈形目鱼类; 另**9**个仅在湖泊组存在, 多为鲤科鱼类
- 常见鱼类占**157**个分类单元, 主要为鲤科、鳅科、鲮科、钝头鮠科、平鳍鳅科和鲃科鱼类
- 同时检测到**18**个濒危鱼类, 包括鳢和中华胭脂鱼
- 检测到一些入侵物种, 如**食蚊鱼**、**长尾鳊**和**大眼华鳊**

人工智能识别 VS 环境DNA技术



主要优势

所见即所得、直观

主要劣势

不同类群样本采集方法差异巨大

定量分析

可准确定量分析

难易程度

操作简单

检测耗时

较短, <30min

应用方向

水生态检测、过鱼设施、关键保护物种
监测、入侵物种监测等

采样简单, 不同类群采样方法相近,

DNA直指核心, 破坏性小

当前数据库质量参差不齐层次不齐, 准
确度有待提高

当前定量分析存在较难

DNA提取、生信分析相对复杂

较长, >1d?

应用前景广阔

人工智能识别 VS 环境DNA技术



未来**人工智能识别**和**环境DNA技术**将会作为最具潜力的下一代监测方法
相辅相承共同为水生态评价提供技术支撑

请批评指正, 谢谢!

