

**AeDNA: 水生生物eDNA数据库**

陈凯 方成池 吴志刚 熊凡 俞丹 崔永德 张琪 王宝强 姜传奇 宋立荣 王洪铸 刘焕章 陈晓飞 凌海波 蔡俊雄 李涛 何舜平 缪炜 熊杰 曾宏辉

**AeDNA: AQUATIC ENVIRONMENTAL DNA DATABASE**

CHEN Kai, FANG Cheng-Chi, WU Zhi-Gang, XIONG Fan, YU Dan, CUI Yong-De, ZHANG Qi, WANG Bao-Qiang, JIANG Chuan-Qi, SONG Li-Rong, WANG Hong-Zhu, LIU Huan-Zhang, CHEN Xiao-Fei, LING Hai-Bo, CAI Jun-Xiong, LI Tao, HE Shun-Ping, MIAO Wei, XIONG Jie, ZENG Hong-Hui

在线阅读 View online: <https://doi.org/10.7541/2022.2022.0379>

**您可能感兴趣的其他文章**

**Articles you may be interested in**

[基于环境DNA宏条形码的洱海鱼类多样性研究](#)

INVESTIGATING THE FISH DIVERSITY IN ERHAI LAKE BASED ON ENVIRONMENTAL DNA METABARCODING

水生生物学报. 2020, 44(5): 1080–1086 <https://doi.org/10.7541/2020.125>

[微型生物群落eDNA–PFU监测](#)

MICROBIAL COMMUNITY BIOMONITORING– eDNA–PFU METHOD

水生生物学报. 2021, 45(6): 1400–1404 <https://doi.org/10.7541/2021.2021.0273>

[0#柴油对斑马鱼基因组DNA和环境DNA遗传毒性的RAPD比较](#)

COMPARISON OF DIESEL GENOTOXICITY ON ZEBRAFISH GENOMIC DNA AND ENVIRONMENTAL DNA FROM WATER ENVIRONMENT OF ZEBRAFISH BY RAPD

水生生物学报. 2020, 44(4): 764–773 <https://doi.org/10.7541/2020.092>

[舟山近海水样环境DNA获取方法的建立](#)

THE EDNA COLLECTION METHOD OF ZHOUSHAN COASTAL WATERS

水生生物学报. 2020, 44(1): 50–58 <https://doi.org/10.7541/2020.007>

[DNA条形码在淡水红藻中的应用评价——基于串珠藻科植物](#)

EVALUATING DNA BARCODE MARKERS FOR FRESHWATER RED ALGAE: A CASE STUDY USING FAMILY BATRACHOSPERMACEAE

水生生物学报. 2017, 41(3): 643–651 <https://doi.org/10.7541/2017.82>

[裸藻遗传转化技术的研究进展](#)

RESEARCH PROGRESS ON GENETIC TRANSFORMATION TECHNOLOGY OF *EUGLENA*

水生生物学报. 2018, 42(3): 655–662 <https://doi.org/10.7541/2018.081>



doi: 10.7541/2022.2022.0379

## AeDNA: 水生生物eDNA数据库

陈凯<sup>1\*</sup> 方城池<sup>1\*</sup> 吴志刚<sup>1\*</sup> 熊凡<sup>1</sup> 俞丹<sup>1</sup> 崔永德<sup>1</sup> 张琪<sup>1</sup> 王宝强<sup>1</sup>  
姜传奇<sup>1</sup> 宋立荣<sup>1</sup> 王洪铸<sup>1</sup> 刘焕章<sup>1</sup> 陈晓飞<sup>2</sup> 凌海波<sup>2</sup> 蔡俊雄<sup>2</sup>  
李涛<sup>1</sup> 何舜平<sup>1</sup> 缪炜<sup>1</sup> 熊杰<sup>1</sup> 曾宏辉<sup>1</sup>

(1. 中国科学院水生生物研究所, 武汉 430072; 2. 湖北省生态环境科学研究院, 武汉 430072)

**摘要:** 环境DNA (eDNA) 技术是一种生态和生物多样性监测和评价的新手段, 完整和准确的参考序列库是eDNA技术应用于水生生物多样性调查的基础。当前, 不同水生生物eDNA参考序列还存在诸多问题, 如不同类群使用的标记基因不同且资源较为分散, 部分参考序列分类不准确, 以及针对我国各类水体中水生生物eDNA参考序列不多等。针对上述问题, 研究构建了水生生物eDNA数据库(AeDNA, <http://aedna.ihb.ac.cn/>)。AeDNA整合了DNA条形码和基因组两种类型参考序列。其中18S、28S、ITS、COI、12S、rbcL等各类DNA条形码60余万条, 涉及2万余种鱼类、1万余种水生植物、1万余种底栖动物、1万余种浮游动物和1万余种浮游植物; 基因组包含线粒体、叶绿体等细胞器基因组6199个及万种鱼类基因组计划和万种原生生物基因组计划所产生的物种基因组。涉及的生境有江、河、湖、海、冰川和温泉等各类水环境, 尤其数据库构建团队贡献的6万余条参考序列, 具有我国丰富的各类水体生境信息。总体来说, AeDNA是一个数据量大、类群覆盖全、准确性高且具有我国水生生物特色的综合性eDNA参考序列库, 是水生态和水生生物多样性监测的重要基础资源。

**关键词:** eDNA数据库; DNA条形码; 基因组; eDNA技术

**中图分类号:** S932 **文献标识码:** A **文章编号:** 1000-3207(2022)11-1741-07



人类活动和气候变化造成了水生态系统受损、水生生物多样性衰退、物种组成改变等负面影响<sup>[1-5]</sup>。开展全面、准确的水生生物多样性调查是建立水生态评估方法并开展生物多样性保护的基础<sup>[6]</sup>。重要的水生生物包括鱼类、水生植物、底栖动物、浮游动物和浮游植物等<sup>[7-9]</sup>。当前, 水生生物调查主要是通过形态鉴定这一手段。然而, 这一方法存在诸多缺陷, 包括成本高、耗时、对从业人员专业水平要求高及方法难以标准化等<sup>[10]</sup>。环境DNA(eDNA)技术的产生则弥补了上述形态鉴定方法的不足, 是一种生态和生物多样性监测的新手段, 也是当前的前沿热点技术<sup>[6, 11-15]</sup>。eDNA技术通过采集水样, 富集并提取DNA, 进行建库和高通量测序, 最后通过将测序序列与数据库序列比对获得物种注释信息<sup>[15]</sup>。尽管eDNA技术在过去10来年

中被广泛应用, 但是用于物种鉴定的数据库各式各样, 并存在分类标准不统一、分类错误和覆盖度不高等问题<sup>[16]</sup>, 这些问题直接影响了应用eDNA技术调查水生生物多样性的准确性, 从而制约了eDNA技术的推广, 因此亟待建立准确的、综合性的水生生物eDNA数据库。

针对eDNA参考序列数据分散和整合度差, 且缺乏我国各类水环境水生生物eDNA参考序列库的问题, 我们通过梳理公共数据库中eDNA数据库的序列信息、Meta信息和分类信息, 同时整合自有的大量水生生物eDNA参考序列, 设计了eDNA数据的收录标准和存储格式, 矫正了部分数据的分类错误, 进而构建了水生生物eDNA数据库——AeDNA(<http://aedna.ihb.ac.cn/>)。AeDNA是中国首个综合性的水生生物eDNA数据库, 将大大促进eDNA技术在水生

**收稿日期:** 2022-09-07; **修订日期:** 2022-09-29

**作者简介:** 陈凯(1989—), 男, 博士后; 研究方向为比较基因组学。E-mail: chenikai@ihb.ac.cn 方城池(1984—), 男, 博士, 助理研究员, 研究方向为新基因的产生及其在适应性演化中的作用。E-mail: fangcc@ihb.ac.cn 吴志刚(1987—), 男, 博士, 助理研究员; 研究方向为湿地植物多样性及生态功能。E-mail: wuzg@ihb.ac.cn \*共同第一作者

**通信作者:** 曾宏辉, 高级工程师; E-mail: zhh@ihb.ac.cn

生物调查领域的应用。

## 1 材料与方法

### 1.1 数据收集

数据来源包括公共数据库和AeDNA团队贡献的序列。公共数据库包括NCBI(<https://www.ncbi.nlm.nih.gov/>)、BOLD(<http://www.boldsystems.org/>)、SILVA(<https://www.arb-silva.de/>)和MitoFish(<http://mitofish.aori.u-tokyo.ac.jp/download.html>)。AeDNA团队贡献的序列主要有各团队自行鉴定和测序的各类水生生物条形码序列数据。

### 1.2 数据处理

数据处理包括分类信息矫正和数据格式标准化。AeDNA数据库的分类信息以NCBI taxonomy为基础,并辅以人工矫正。首先通过NCBI taxid构建NCBI全部生物类群分类学数据库,分类系统以标准的界、门、纲、目、科、属、种共7个分类层级进行展示,然后将收录序列的物种名称与物种库进行匹配,继而得到序列的标准分类信息。最后将计算得到的标准分类信息进行人工校正。数据格式标准化包括制定标准Meta信息收集表,Meta信息(Meta-information)是指与序列相关的基础信息,包括序列名称、分类信息、序列类型、采样位置(经纬度)、样品采集地生境、数据来源信息和描述信息。标准化格式制定后采用自行编写的脚本进行统一汇编,并且统一自动化重命名。

## 1.3 数据库构建

本数据库基于CentOS 8系统,React与TypeScript用于网页前端页面展示,Django用于网页后端功能模块开发,Python用于文本操作和逻辑运行,MongoDB用于数据管理。

## 2 结果与讨论

### 2.1 数据库结构和功能呈现

AeDNA主体结构包括6部分:主页、DNA条形码、基因组、分析平台、数据提交和关于我们(图1)。数据资源包括收录的序列、分类学数据、序列相关的Meta信息及将序列构建成为用于同源分析的索引。数据存储包括MongoDB和BLAST database,其中MongoDB用于网页数据原始数据展现和Meta信息存储;BLAST database用于存储同源分析的索引序列。分析工具包括数据展示、数据提交和分析平台。主要功能包括搜索功能、数据统计功能、分析功能和数据更新功能。

**搜索功能** 用户可以通过分类信息在AeDNA数据库搜索数据,也可以结合通过分类信息、序列名称、样品采集地(国家/地区)和生境进行搜索。通过搜索用户可以获得目标序列及与目标序列有关的分类学信息和Meta统计信息。

**数据统计功能** 包括序列统计和Meta信息统计。序列统计呈现了各分类阶元的条形码或基因组数量;Meta信息统计呈现了各分类阶元条形码

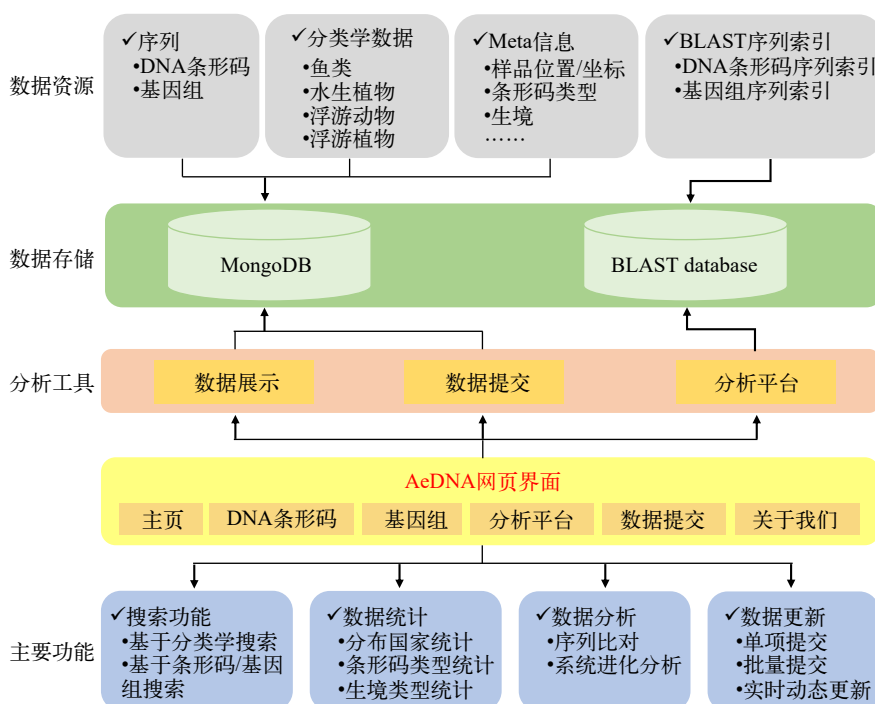


图1 AeDNA数据库网站构架示意图

Fig. 1 Web interface and framework of AeDNA database

序列类型和与其相关的样品分布数据, 如分布国家统计, 地理坐标呈现。

**分析功能** 数据库支持用户提交序列并通过BLAST程序搜索目标序列在AeDNA数据库中的同源序列, 进一步基于搜索到的同源序列构建系统发育分析(图 1)。BLAST默认程序为BLASTN, 默认比对序列为DNA条形码数据库的所有类群、所有序列, 默认比对E值为0.1。用户可以自行选择比对程序、比对序列类型、比对E值和其他参数。系统发育分析中多序列比对工具为Muscle v5, 系统树算法为TreeBeST, 可视化工具为MSAView。

**数据更新功能** 数据库提交分为单项提交和批量提交两种不同的模式(图 1)。单项提交允许用户将eDNA数据逐条提交到AeDNA, 批量提交允许用户通过填写标准信息收集表格提交序列。提交的序列需要经过AeDNA后台审核, 审核通过的序列汇入AeDNA数据库并反馈给用户, 审核不通过的序列提示用户再次按照要求提交。

## 2.2 数据类型

AeDNA包含水生生物DNA条形码和基因组数据, 通过整合公共数据和AeDNA构建团队测定的条形码数据构建而成。数据库集成了60余万条序列, 其中数据库构建团队贡献了6万余条, 以我国各类水体水生生物条形码数据为特色。数据库涉及的生物类型包括2万余种鱼类、1万余种水生植物、1万余种底栖动物、1万余种浮游动物和1万余种浮游植物(图 2)。

**DNA条形码** DNA条形码是物种特有、能稳定遗传、可以作为身份标签的一段DNA序列。每个物种具有多种条形码序列, 本数据库收录多个物种多种条形码序列, 包括18S、ITS、COI、12S、rbcL等38种条形码数据。数据由AeDNA团队通过检索公用数据库、单一物种条形码测序、eDNA多

物种批量测序和挖掘基因组获得的条形码序列组成。DNA条形码数据将为水生生物鉴定、多样性调查、保护种/入侵种监测及研究提供参考。

数据库包含23418种鱼类、11750种水生植物、8817种底栖动物、10320种浮游动物和8953种浮游植物(图 3A)。总计592758条DNA条形码序列, 其中鱼类376201条、水生植物103227条、底栖动物11214条、浮游动物67689条、浮游植物34427条(图 3B)。不同生物类群收录的条形码序列有较大差异, 其中鱼类以线粒体基因为主, 例如COI、12S等, 还有少量核糖体基因序列和基因组上其他基因序列(图 3C); 水生植物包含叶绿体和ITS序列(核糖体基因); (图 3D); 浮游动物包含以COI为代表的线粒体序列和以18S为代表的核糖体序列(图 3E); 除蓝藻以外的浮游植物则包含rbcL为代表的叶绿体序列、COI为代表的线粒体序列和18S为代表的核糖体序列(图 3F)。我们发现高等动植物DNA条形码多采用进化速率更快的线粒体基因、叶绿体基因和ITS序列, 而浮游动植物除了上述基因外, 还采用更为保守的18S序列(真核藻类)或者16S序列(蓝藻), 这与已有研究结果一致<sup>[17-19]</sup>。另外, 最新的研究表明以线粒体12S序列作为鱼类鉴定的参考条形码序列相较于COI等条形码序列能鉴定更多的物种, 鉴定的结果更准确<sup>[19]</sup>。本数据库收录的12S序列5000余条, 后期还将持续进行增量更新。

AeDNA构建团队贡献了该数据库中的6万余条DNA条形码序列, 其中鱼类序列5.7万余条, 水生植物序列500余条, 底栖动物200余条, 浮游动物和浮游植物序列1200余条; AeDNA团队贡献的条形码数据具有完备的Meta信息, 除了条形码类型信息外还包括样品采集地点、生境类型、图片、视频等。从序列所在生境的分布图可以看出这些序列主要来源于我国各类水体中水生生物, 具有鲜明的



图 2 AeDNA数据总览

Fig. 2 Landscape of the data of AeDNA



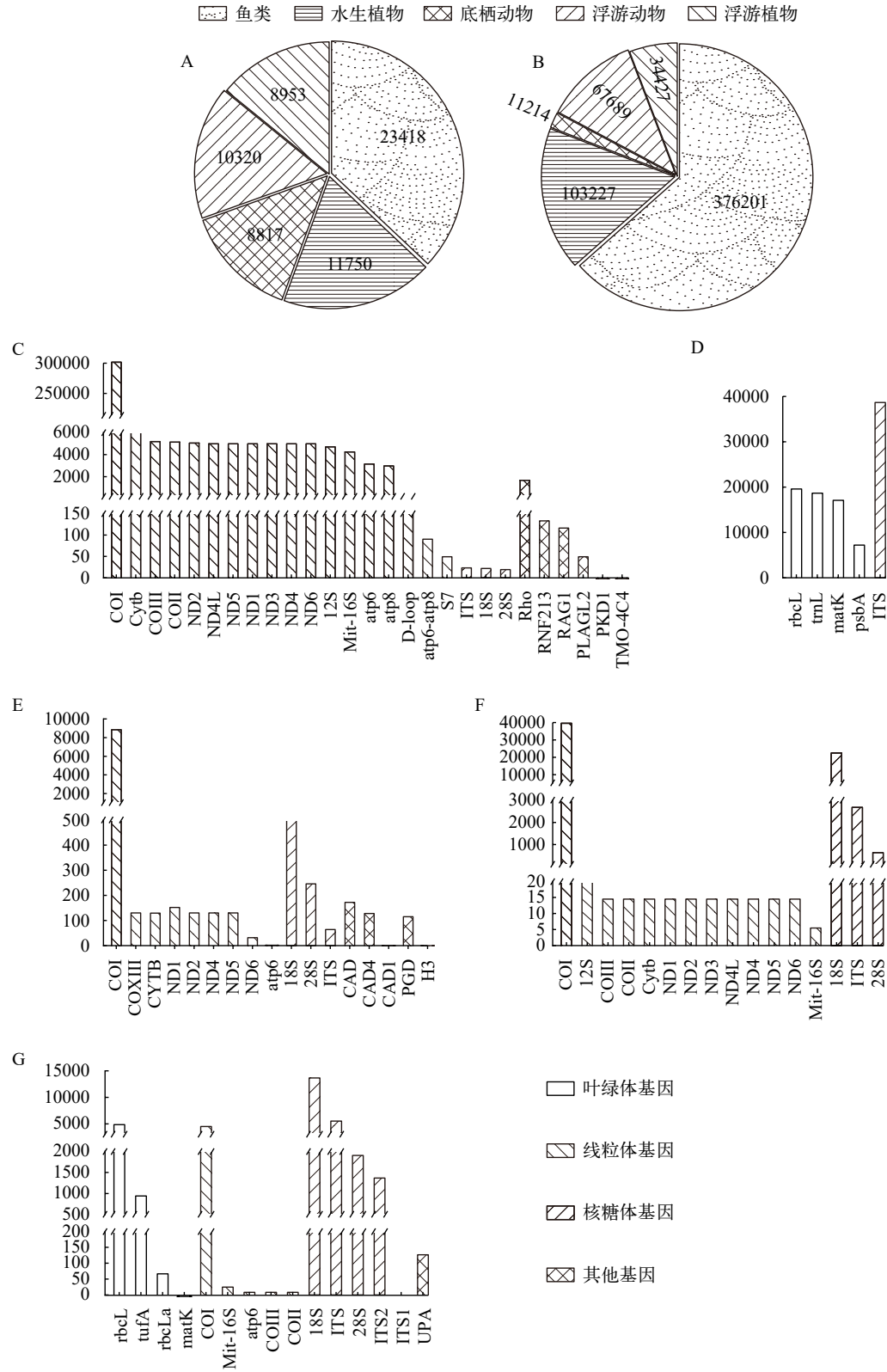


图 3 DNA 条形码数据统计

Fig. 3 Statistics of DNA barcodes

A. 数据库已收录DNA条形码的物种数量统计; B. DNA条形码数量统计; C—F. 不同水生生物类群DNA条形码类型统计[C、D、E、F和G分别代表鱼类、水生植物、底栖动物、浮游动物和真核浮游植物(仅含真核藻类)]

A. The number of species with DNA barcode in the database; B. Statistics of the number of DNA barcodes; C—F. Statistics of the types of DNA barcodes [C, D, E and F represent the DNA barcodes of the fish, plant, Zoobenthos, zooplankton and phytoplankton (only eukaryotic Algae, respectively)]

特色。样品采集地除中国台湾外的其他所有省份均有覆盖(图 4), 以长江流域和青藏高原为特色。内蒙古自治区、中国台湾和新疆维吾尔自治区等没有分布或分布较少, 未来需要加大采样力度, 使之覆盖更广、更全。

**基因组条形码** 如上文所述, 每个物种具有多种条形码序列, 不同类型的条形码序列长度不同, 序列保守性不同, 鉴定精度和解析度不一致, 因此, 研究人员针对各类群物种鉴定所用的条形码不一致。例如, 过去COI序列是鱼类鉴定的金标准, 最近发现12S序列比COI鉴定精度高<sup>[19]</sup>; 用于大部分真核藻类鉴定的条形码是18S、ITS、28S、tufA和rbcL等, 但甲藻和异鞭藻用COI, 轮藻则采用与很多高等植物类似的atpA/B、psbA/B等<sup>[18]</sup>; 蓝藻(原核浮游植物)通常采用16S、ITS、rbcL和rpoC1等。原生动物的鉴定通常用18S和COI<sup>[17]</sup>。因此, 采用DNA条形码进行物种鉴定缺乏统一标准。

基因组是指某一特定物种细胞内的一整套遗传物质, 是一个物种所有DNA序列信息的集合。基因组某种意义上说是一种“超级条形码”, 其鉴定精度和解析度将远远超过单独使用某一个标记基因。当前, 大量细胞器基因组和物种基因组的产生, 使得基因组的数目到达了一定的量级, 因此, 利用基因组进行物种鉴定也变为可行。

本数据库的基因组数据包含线粒体基因组、叶绿体基因组以及由“万种鱼基因组计划”<sup>[20]</sup>和“万

种原生生物基因组计划”<sup>[21]</sup>产出的全基因组测序数据。数据库共搜集了5872条细胞器基因组序列, 其中线粒体基因组共3377条, 鱼类3118条, 水生植物18条, 浮游动物49条, 浮游植物192条; 叶绿体基因组共2495条, 水生植物1568条, 浮游植物927条(图 5)。后期数据库将大力整合基因组数据, 尤其是旗舰物种和入侵物种基因组数据, 同时结合基因组测序和通过大规模水环境样品采集、测序, 从宏基因组数据挖掘环境基因组来补充未测定基因组的空白。此外, AeDNA进一步将开发基于基因组水平进行物种鉴定和生物多样性研究的新方法、新标准。

### 3 展望

AeDNA为水生生物调查提供参考序列数据库, 整合了现有DNA条形码和基因组数据, 将促进eDNA技术在水生生物调查中的应用。当前, 数据库为第一版, 未来将进行持续的数据更新, 同时整合更多分析功能, 包括: (1)数据库将新增两栖动物、蓝藻等类群水生生物eDNA数据, 将继续收集、汇编国内同行测定的水生生物条形码数据和通过大规模采集样品、测序来完善中国境内条形码数据库, 构建更具综合性的水生生物eDNA数据库。样品来源将涉及中国大部分水域, 将实现对国内水生生物eDNA数据的全覆盖。数据库将收集水生生物图片、视频信息, 将展现序列、分类信息、

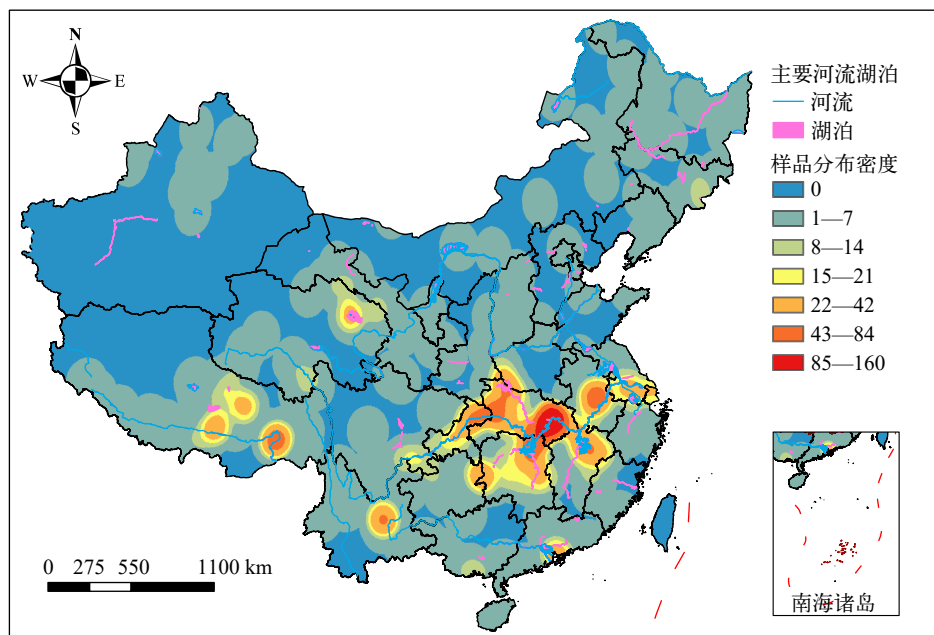


图 4 AeDNA团队贡献的DNA条形码序列地域分布图

Fig. 4 The distribution of DNA barcode sequences contributed by AeDNA team

根据中华人民共和国自然资源部监管的标准地图编辑[审图号为GS(2016) 1569]

Edited based on a standard map supervised by the Ministry of Natural Resources of the People's Republic of China [No. GS (2016) 1569]

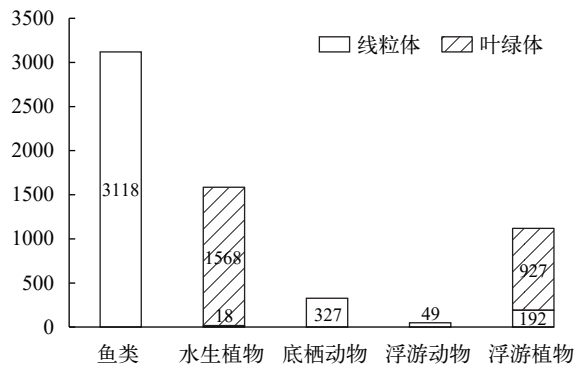


图5 AeDNA数据库中基因组数据的统计

Fig. 5 Statistics of genomes in AeDNA database

影像和Meta信息等数据。(2)除了实现对数据增量更新的同时,本数据库将着力于eDNA数据质量的提升。包括:基于NCBI分类学体系的分类系统矫正;基于聚类分析去除冗余序列和错误序列;基于HMM模型对序列进行修剪等。(3)数据库已建设有序列比对和系统发育树构建功能,后续将新增:多样性分析功能,如 $\alpha$ 和 $\beta$ 多样性等;入侵种和保护种监测功能;各类水生态指标分析功能和基于eDNA的水生生物评价功能。(4)数据库将建立用户管理系统,即用户可以在AeDNA数据库创建账号,实现对个人数据和分析任务的长期管理。总体来说,AeDNA将由专业团队进行长期维护,持续迭代,旨在实现水生生物的调查、监测、追溯和预警的综合能力。

## 致谢:

AeDNA网站首页视频由中国科学院水生生物研究所熊雄博士和王浩骅提供素材并进行剪辑。网页中各类水生生物照片由中国科学院水生生物研究所谷思雨、缪荣丽、魏朝军、王府臣、高欣欣、潘婷婷、王慧君、邴厚骅和王红霞提供。中国科学院水生生物研究所陈佳在网站网络管理和维护方面提供了帮助。网站构建得到了中国科学院超级计算环境武汉分中心的支持。

## 参考文献:

- [1] Barnosky A D, Matzke N, Tomiya S, *et al.* Has the Earth's sixth mass extinction already arrived [J]? *Nature*, 2011, **471**(7336): 51-57.
- [2] Butchart S H M, Walpole M, Collen B, *et al.* Global biodiversity: indicators of recent declines [J]. *Science*, 2010, **328**(5982): 1164-1168.
- [3] Dirzo R, Young H S, Galetti M, *et al.* Defaunation in the anthropocene [J]. *Science*, 2014, **345**(6195): 401-406.
- [4] Emer C, Galetti M, Pizo M A, *et al.* Defaunation precipitates the extinction of evolutionarily distinct interactions in the Anthropocene [J]. *Science Advances*, 2019, **5**(6): eaav6699.
- [5] Parkes M. Personal commentaries on "Ecosystems and human well-being: health synthesis - areport of the millennium ecosystem assessment" [J]. *EcoHealth*, 2006, **3**(3): 136-140.
- [6] Zhang H, Xian W W. Application of environmental DNA technology in ecological conservation and monitoring [J]. *Marine Sciences*, 2020, **44**(7): 96-102. [张辉, 线薇薇. 环境DNA技术在生态保护和监测中的应用 [J]. *海洋科学*, 2020, **44**(7): 96-102.]
- [7] Liu J K. Advanced Hydrobiology [M]. Beijing: Science Press, 1999: 151-278. [刘健康. 高级水生生物学 [M]. 北京: 科学出版社, 1999: 151-278.]
- [8] Shen Y F. Protozoan [M]. Beijing: Science Press, 1999: 1-10. [沈蕴芬. 原生动物学 [M]. 北京: 科学出版社, 1999: 1-10.]
- [9] Jiang C Q, Gu S Y, An R Z, *et al.* Morphology and phylogeny of two newly recorded ciliates (*Paramecium primaurelia* and *Tetrahymena mimbres*) from Tibetan hot springs in China [J]. *Acta Hydrobiologica Sinica*, 2020, **44**(1): 197-205. [姜传奇, 谷思雨, 安瑞志, 等. 西藏温泉两种中国新记录纤毛虫第一双小核草履虫和明布雷斯四膜虫的形态学和系统发育学研究 [J]. *水生生物学报*, 2020, **44**(1): 197-205.]
- [10] Beng K C, Corlett R T. Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects [J]. *Biodiversity and Conservation*, 2020, **29**(7): 2089-2121.
- [11] Deiner K, Bik H M, Mächler E, *et al.* Environmental DNA metabarcoding: Transforming how we survey animal and plant communities [J]. *Molecular Ecology*, 2017, **26**(21): 5872-5895.
- [12] Garrido-Sanz L, Senar M À, Piñol J. Relative species abundance estimation in artificial mixtures of insects using mito-metagenomics and a correction factor for the mitochondrial DNA copy number [J]. *Molecular Ecology Resources*, 2022, **22**(1): 153-167.
- [13] Pawlowski J, Apothéoz-Perret-Gentil L, Altermatt F. Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring [J]. *Molecular Ecology*, 2020, **29**(22): 4258-4264.
- [14] Ruppert K M, Kline R J, Rahman M S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA [J]. *Global Ecology and Conservation*, 2019(17): e00547.
- [15] Shan X J, Li M, Wang J W. Application of environmental DNA technology in aquatic ecosystem [J]. *Progress in Fishery Sciences*, 2018, **39**(3): 23-29. [单秀娟, 李苗, 王伟继. 环境DNA(eDNA)技术在水生生态系统中的应用

- 研究进展 [J]. *渔业科学进展*, 2018, **39**(3): 23-29.]
- [16] Thomsen P F, Willerslev E. Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity [J]. *Biological Conservation*, 2015, **183**(1): 4-18.
- [17] Abraham J S, Sripoorna S, Maurya S, *et al.* Techniques and tools for species identification in ciliates: a review [J]. *International Journal of Systematic and Evolutionary Microbiology*, 2019, **69**(4): 877-894.
- [18] Leliaert F, Verbruggen H, Vanormelingen P, *et al.* DNA-based species delimitation in algae [J]. *European Journal of Phycology*, 2014, **49**(2): 179-196.
- [19] Xiong F, Shu L, Zeng H, *et al.* Methodology for fish biodiversity monitoring with environmental DNA metabarcoding: The primers, databases and bioinformatic pipelines [J]. *Water Biology and Security*, 2022, **1**(1): 100007.
- [20] Fan G, Song Y, Yang L, *et al.* Initial data release and announcement of the 10, 000 Fish Genomes Project (Fish10K) [J]. *Gigascience*, 2020, **9**(8): giaa080.
- [21] Miao W, Song L, Ba S, *et al.* Protist 10000 genomes project [J]. *Innovation*, 2020, **1**(3): 100058.

## AeDNA: AQUATIC ENVIRONMENTAL DNA DATABASE

CHEN Kai<sup>1</sup>, FANG Cheng-Chi<sup>1</sup>, WU Zhi-Gang<sup>1</sup>, XIONG Fan<sup>1</sup>, YU Dan<sup>1</sup>, CUI Yong-De<sup>1</sup>, ZHANG Qi<sup>1</sup>,  
WANG Bao-Qiang<sup>1</sup>, JIANG Chuan-Qi<sup>1</sup>, SONG Li-Rong<sup>1</sup>, WANG Hong-Zhu<sup>1</sup>, LIU Huan-Zhang<sup>1</sup>,  
CHEN Xiao-Fei<sup>2</sup>, LING Hai-Bo<sup>2</sup>, CAI Jun-Xiong<sup>2</sup>, LI Tao<sup>1</sup>, HE Shun-Ping<sup>1</sup>, MIAO Wei<sup>1</sup>,  
XIONG Jie<sup>1</sup> and ZENG Hong-Hui<sup>1</sup>

(1. Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China; 2. Hubei Provincial Academy of Eco-Environmental Sciences, Wuhan 430072, China)

**Abstract:** Environmental DNA (eDNA) technology is an emerging method for ecological and biodiversity monitoring and evaluation, and is one of the frontier technologies. A complete and accurate reference sequence database is the basis of the application of eDNA technology to the investigation of aquatic biodiversity. At present, there are still some problems in eDNA reference sequences of different aquatic organisms, such as the use of different marker genes for different groups of aquatic organisms, the scattered distribution and inaccurate classification of reference sequences, and the lacking of reference sequences for aquatic organisms in China. To address these issues, the Aquatic eDNA database (AeDNA, <http://aedna.ihb.ac.cn/>) was constructed in this study. AeDNA integrates two types of reference sequences: DNA barcode and genome database. The DNA barcode database contains more than 600000 records of different marker genes, including 18S, 28S, ITS, CO1, 12S and rbcL, etc., which covers about 20000, 10000, 10000, 10000 and 10000 species of fishes, plants, Zoobenthos, zooplankton and phytoplankton, respectively. The genome database contains 6199 organelle genomes of mitochondria and chloroplast, as well as the genomes of fish and protist which generated by the 10000 Fish Genomes Project and the Protist 10000 Genomes Project. The habitats cover various water environments such as rivers, lakes, seas, glaciers and hot springs. In particular, AeDNA contains more than 60000 reference sequences of aquatic organisms in various water environments of China. In summary, AeDNA is a comprehensive eDNA reference sequence database with large amount of data, high coverage of taxa, high accuracy and more data for aquatic organisms in China. It will be a fundamental resource for aquatic ecology and biodiversity monitoring.

**Key words:** eDNA database; DNA barcode; Genome; eDNA technology