J. Lake Sci.(湖泊科学), 2026, 38(6):000-000
DOI 10.18307/2026.0635
©2026 by Journal of Lake Sciences

雅鲁藏布江流域鱼类 DNA 条形码数据集*

李雁羽[1,2]，王子旺[2,3]，张继峰[1,4,5]，钱建硕[2,3]，陈晋南[2,3]，何德奎[1,2,3**]
（1:西藏大学，拉萨 850000）
（2:中国科学院水生生物研究所，武汉 430072）
（3:中国科学院大学，北京 100049）
（4:西藏大学生态环境学院,青藏高原生物多样性与生态环境保护教育部重点实验室，拉萨 850000）
（5:拉萨城市湿地生态系统西藏自治区野外科学观测研究站，拉萨 850000）

摘要：分子物种鉴定与环境 DNA（environmental DNA, eDNA）技术近年来在生物多样性调查与环境监测中迅速发展，因其高效、非侵入式等优势而受到广泛关注，但其应用效果和结果准确性在很大程度上依赖于 DNA 条形码等参考数据库的完整性与可靠性。当前，我国多数流域尚缺乏系统、规范且覆盖全的 DNA 条形码数据资源，已成为制约分子监测技术在生态评估与生物多样性保护中应用的瓶颈。雅鲁藏布江作为世界上海拔最高的大型跨境河流之一，独特的气候条件和水文地貌孕育了丰富多样、极具特色的淡水鱼类区系，然而其流域尺度的鱼类 DNA 条形码数据库建设仍滞后，缺乏系统性数据支撑，限制了分子监测手段在该区域生物多样性调查与保护中的应用。本研究集成了研究团队 1998—2024 年采自雅鲁藏布江流域鱼类样品 DNA 序列，并整合公共数据库中的雅鲁藏布江流域鱼类序列，经统一数据处理、严格质量控制和基于遗传距离的可靠性评估，汇编形成了首个覆盖雅鲁藏布江全流域的鱼类 DNA 条形码参考数据集。数据集包含 3,174 条高质量 DNA 序列，其中自主测序序列 2,890 条（占总序列数的 91.1%），公共数据库序列 284 条（占 8.9%）。数据样点覆盖雅鲁藏布江干流、主要支流及附属湖泊、湿地，共计 82 个样点；海拔跨度为 155—4,600 m；物种涵盖 8 目 21 科 49 属 78 种，包括土著鱼类 62 种，外来鱼类 16 种；覆盖了 69%的雅鲁藏布江特有鱼类物种和 100%的泸公河汇口及其上游江段鱼类；序列平均长度约为 813 bp，核心分子标记为 COI 和 Cyt *b*，序列长度范围分别为 461-1779 bp 和 798-1390 bp；多数物种表现出清晰的 DNA 条形码间隙，可支持分子水平的物种鉴定。数据集采用“元数据—序列数据”分离管理模式，遵循 FAIR（可发现、可访问、可操作、可重用）和 CARE（集体利益、控制权、责任、伦理）国际数据治理原则，包含物种分类地位、标本凭证号、分布信息、采集时间、采集人及序列来源等标准化字段，并附有部分测序物种的彩色照片。数据集通过科学数据银行（ScienceDB；DOI：10.57760/sciencedb.36688）向用户开放访问。本数据集弥补了雅鲁藏布江流域鱼类分子鉴定参考数据的关键缺口，可为物种鉴定、生物多样性编目、外来物种监测及 eDNA 宏条形码研究提供参考依据，并为高原河流生态系统的科学保护与系统管理提供数据支撑。

关键词：青藏高原；跨境河流；鱼类 DNA 条形码；生物多样性监测；环境 DNA

A basin-scale fish DNA barcode dataset from the Yarlung Tsangpo River*

Li yanyu[1,2], Wang ziwang[2,3], Zhang Jifeng[1,4,5], Qian Jianshuo[2,3], Chen Jinnan[2,3], He Dekui[1,2,3**]
(1: *Xizang University, Lhasa 850000, P. R. China*)
(2: *Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, P. R. China*)
(3: *University of Chinese Academy of Sciences, Beijing 100049, P. R. China*)
(4: *Key Laboratory of Biodiversity and Environment on the Qinghai-Xizang Plateau, Ministry of Education, School of Ecology and Environment, Xizang University, Lhasa, 850000, P. R. China*)
(5: *Lhasa, Urban Wetland Ecosystem, Observation and Research Station of Xizang Autonomous Region, Lhasa, 850000, P. R. China*)

Abstract: Molecular species identification and environmental DNA (eDNA) technologies have advanced rapidly in recent

years. These tools are widely used in biodiversity surveys and environmental monitoring because they are highly efficient and non-invasive. Their effectiveness and accuracy depend strongly on the completeness and reliability of reference databases, especially DNA barcode libraries. In most river basins across China, systematic, standardized, and comprehensive fish barcode resources are still lacking. This limitation restricts the application of molecular techniques in ecological assessment and biodiversity conservation. The Yarlung Tsangpo River is one of the world's largest transboundary rivers at high altitude. Its unique climate, combined with complex hydrological and geomorphological conditions, supports rich and endemic freshwater fish diversity. However, a complete fish DNA barcode reference library for the entire basin has not yet been established. The absence of systematic genetic data further hampers the use of molecular monitoring tools in this alpine region. In this study, we compiled fish DNA sequences from specimens collected by our research group across the whole Yarlung Tsangpo River basin from 1998 to 2024, and supplemented them with validated sequences from public databases. All genetic data were processed following standard protocols and underwent strict quality control and reliability assessment based on genetic distance metrics. We constructed the first comprehensive fish DNA barcode reference dataset for the entire basin. This dataset contains 3174 high-quality DNA sequences, including 2890 newly generated sequences (91.1%) and 284 sequences (8.9%) from public databases. Specimens were collected from 82 sampling sites, covering river main stems, major tributaries, adjacent lakes, and wetlands, with elevations ranging from 155 to 4,600 m. The dataset covers 78 fish species belonging to 49 genera, 21 families, and 8 orders, including 62 native species and 16 non-native species. It includes 69% of the basin's endemic fish species and 100% of the fish species found upstream of the Lhagu River confluence. The average sequence length is approximately 813 bp. Cytochrome c oxidase subunit I (COI) and cytochrome *b* (Cyt *b*) are the core molecular markers, with sequence length ranges of 461–1779 bp and 798–1390 bp, respectively. Most species show clear DNA barcode gaps, ensuring reliable taxonomic identification at the species level. The dataset follows an independent metadata–sequence data structure and complies with the FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles. It integrates standardized metadata, including taxonomic information, voucher specimen records, species distribution, sampling date, collector details, and sequence origins, and provides in-situ color photographs for some fish species. The dataset is publicly available through the Science Data Bank (ScienceDB; DOI: 10.57760/sciencedb.36688). This dataset fills a critical gap in fish DNA barcode reference resources for the Yarlung Tsangpo River basin. It supports species discrimination, biodiversity inventory, non-native species monitoring, and eDNA metabarcoding studies. It also provides fundamental genetic data for the conservation of plateau river ecosystems and watershed management.

Keywords: Qinghai-Tibet Plateau; Transboundary river; Fish DNA Barcode; Biodiversity Monitoring; Environmental DNA (eDNA).

雅鲁藏布江（The Yarlung Tsangpo River, Yarlung Zangbo River）是世界上海拔最高的大型跨境河流之一，发源于喜马拉雅山中段北麓的杰玛央宗冰川，干流自西向东流经西藏自治区的日喀则市、拉萨市、山南市和林芝市，覆盖 22 个区、县，于墨脱县的巴昔卡出境后最终注入印度洋[1]。该流域具有巨大的海拔落差和高度复杂的地形—气候条件，形成了显著的水文与生态异质性，产生了以高原特有类群（裂腹鱼类、鳅鮀鱼类和高原鳅属鱼类）为主的鱼类区系[2-5]。该流域分布 10 目 25 科 70 属 155 种土著鱼类，其中 29 种为本流域的特有种[2]。当前，雅鲁藏布江正面临人类活动、气候变化引起的多重压力，保护流域生态系统健康与完整性是生物多样性保护面临的重要课题。准确掌握流域鱼类组成与变化趋势，是流域鱼类资源保护与管理的重要基础，也是支撑《昆明-蒙特利尔全球生物多样性框架》中“保护至少 30%的陆地、内陆水域、沿海及海洋区域”“恢复退化生态系统”等核心目标在高原流域落地的关键[6]。

传统的鱼类物种鉴定主要依赖形态学特征，但该方法易受标本保存状态、发育阶段和鉴定者经验等因素制约，在形态相似程度较高的类群中尤为显著，难以满足大范围、高频次生物多样性监测的精准鉴定需求。DNA 条形码 (DNA barcoding) 技术为此提供了突破性的解决方案[7]，生命条形码联盟 (the Consortium for the Barcode of Life, CBOL) 将 DNA 条形码定义为一段能够高效鉴别物种的标准 DNA 序

列[8]。该技术的有效性基于两项理论预期：其一，同一物种的个体通常具有高度相似的条形码序列；其二，物种间的遗传差异应大于种内变异。这一差异在理想状态下表现为“条形码间隙”，即在数据分布上种内遗传距离显著小于种间距离[9]。该技术仅需微量组织样本便可以进行精确鉴定，使物种鉴定不再受个体发育阶段或形态完整性的限制，通过与参考数据集结合，可以实现大批次、高通量的物种鉴定，还可以促进形态上难以区分的隐存种的发现或是追溯自然杂交物种的亲本来源[10]。

近年来 DNA 条形码技术已被广泛应用于长江流域、亚马孙流域和西北太平洋等多个区域，并构建了多个 DNA 条形码参考数据集或数据库，为物种鉴定与多样性评估提供了重要的数据支撑[11-15]，此外，DNA 条形码技术衍生出的 eDNA 宏条形码技术，仅通过采集少量水样，捕获和分析水体中游离的遗传物质，便可以实现对鱼类群落非侵入式、高效率的监测，特别适用于雅鲁藏布江这种空间异质性极高、传统鱼类调查难度较大的流域[16-17]。

尽管技术优势明显，但 DNA 条形码技术在雅鲁藏布江流域的深入应用仍面临制约。目前在我国主要河流中，仅长江、黄河和红河流域已建成系统性的流域尺度鱼类 DNA 条形码数据集[11,18-19]，而包括雅鲁藏布江在内的其他大型河流，仍缺乏完整、规范的本地鱼类 DNA 条形码数据资源。这种数据缺口导致全面可靠的分子物种鉴定和环境 DNA 研究难以推广应用。 近期，已有研究开始尝试弥补这一短板。例如，2022 年有研究构建了覆盖雅鲁藏布江下游江段 22 种鱼类的 COI 条形码数据库，为局部区域的分子鉴定提供了基础支撑[20]。随后，Feng 等针对雅鲁藏布江中上游鱼类区系，系统比较了针对 12S、COI 和 Cyt *b* 基因设计的引物在物种分辨能力方面的表现，并且指出 Cyt *b* 在该流域具有更高的物种鉴别分辨率，并据此构建了用于 eDNA 宏条形码监测的 Cyt *b* 本地参考数据库，为珍稀及外来鱼类的高灵敏度检测提供了重要技术框架[21]。这些工作虽证实了本地化数据库或数据集的有效性，但现有成果仍呈碎片化状态，主要体现在空间覆盖不完整、物种覆盖不足、时空信息缺失、技术标准不统一、应用共享机制不完善等缺口。已有基于大规模线粒体 DNA 条形码序列的全球尺度研究表明，保护物种多样性并不等同于保护基因多样性，淡水鱼类遗传多样性的空间分布格局与流域地理结构密切相关，尤其大型河流、空间异质性等地形因子的显著影响[22]。序列数据空间信息的缺失将削弱对遗传分化格局及其生物地理学意义的解析能力。上述数据缺口阻碍了相关前沿技术从方法验证向规模化应用的推进。

为填补上述核心数据缺口，本研究整合团队 27 年全流域采样（覆盖雅鲁藏布江干流、主要支流及附属湖泊、湿地等生态系统）所获取的序列数据，整合公共数据库序列，旨在构建首个覆盖雅鲁藏布江全流域的、符合国际规范的开放共享鱼类 DNA 条形码参考数据集（表 1）。对于调查成本高、空间异质性强的高原大型河流系统，数据集的完整性与国际通用的 FAIR 原则（可发现 Findable、可访问 Accessible、可互操作 Interoperable、可重用 Reusable）及 CARE 原则（集体利益 Collective Benefit、控制权 Authority to Control、责任 Responsibility、伦理 Ethics）同等重要。为践行上述原则，本研究进一步系统阐述其获取方式、本地化部署方法及典型应用场景该数据集将直接服务于物种快速鉴定、生物多样性编目、外来种监测及非法贸易核查等工作，并为未来建立全流域 eDNA 监测网络、评估重大工程生态影响、制定生态保护策略，提供标准化的数据基础。

1. 数据来源与采集

本研究雅鲁藏布江鱼类 DNA 条形码数据集的构建（图 1），遵循 FAIR 与 CARE 国际数据治理原则，数据集的数据来源包括两部分：一是自主开展的全流域野外采样与测序数据，二是检索整合的公共数据库序列数据，通过双重数据源保障数据集的完整性与代表性。

1.1 研究区域概述

雅鲁藏布江的地理范围介于北纬 28°00′—31°16′、东经 82°00′—97°07′之间，全长约 2057 km，流域面积约 240,480 km²。在我国主要河流中，雅鲁藏布江长度居全国第六位，流域面积居第五位。雅鲁藏布江源头海拔约 5,590 m，流出中国国境处海拔约 155 m，在流域内形成约 5,435 m 的巨大海拔落差，干流单位河长水能蕴藏量与全流域平均单位面积水能蕴藏量，均居我国各大河流中的首位，蕴藏着极为丰富的水能资源。雅鲁藏布江沿途汇集多雄藏布、年楚河、拉萨河、尼洋河、帕隆藏布、锡约姆河等主要支

流，并在帕隆藏布汇口附近陡然南折，切穿喜马拉雅山东部，形成著名的雅鲁藏布大峡谷。河流自墨脱县巴昔卡出境后进入印度境内，改称布拉马普特拉河(Brahmaputra River)，进入孟加拉国后又称为贾木纳河(J amunā River)，与恒河相汇后，最终注入孟加拉湾[1]。

表 1 数据集基本信息简介

Tab.1 Basic Information Profile of the Database

类别	描述
数据集名称	雅鲁藏布江流域鱼类 DNA 条形码数据集
数据通信作者	何德奎(hedekui@ihb.ac.cn)
数据作者	李雁羽、王子旺、张继峰、钱建硕、陈晋南、何德奎
数据时间范围	1998—2024
地理区域	28°00′−31°16′ N、82°00′− 97°07′ E
数据量	5.91 MB
数据格式	*.xlx（元数据表） *.txt（DNA 序列） *.pdf（测序物种照片）
数据链接	https://doi.org/10.57760/sciencedb.36688
基金项目	第二次青藏高原综合考察研究(2024QZKK0200; 2019QZKK05010102) the Second Tibetan Plateau Scientific Expedition and Research Program (2024QZKK0200; 2019QZKK05010102); 西藏自治区科技计划基地与人才计划项目(XZ202501JD0019) Base and Talent Program Projects of Science and Technology Program of Tibet Autonomous Region (XZ202501JD0019)
数据组成	（1）雅鲁藏布江流域鱼类 DNA 条形码元数据表(*.xlsx) （2）雅鲁藏布江流域鱼类线粒体 COI 基因序列数据（*.txt） （3）雅鲁藏布江流域鱼类线粒体 Cyt *b* 基因序列数据（*.txt） （4）雅鲁藏布江流域鱼类其他分子标记数据（*.txt） （5）雅鲁藏布江流域鱼类测序物种照片（*.pdf）

图 1 编制雅鲁藏布江鱼类 DNA 条形码数据集的工作流程

Fig.1 Workflow for establishing a fish DNA barcode reference database for the Yarlung Tsangpo River

1.2 鱼类样本采集与鉴定

所有鱼类样本均采集自雅鲁藏布江流域，包括雅鲁藏布江干流、主要支流及该流域内的附属湖泊、湿地等生态系统。采样范围覆盖上游（仲巴县里孜村以上）、中游（仲巴县里孜至米林县派镇）和下游（米林县派镇至巴昔卡附近）各河段，涉及宽谷、窄谷和峡谷等雅鲁藏布江典型河流地貌。样品采集工作持续 27 年（1998—2024 年），形成了空间覆盖广泛且时间序列长的样本数据，共计采集 82 个样点（断面）（采样点信息详见图 1 和附件 1）。采集方式为地笼、流刺网、定置刺网以及低频电鱼机等方法。所有标本采集均在相关主管部门许可下进行。

采集到的鱼类标本首先使用丁香酚进行麻醉，随后采用断髓法实施安乐死[23-24]。标本经鉴定后，取其一侧的偶鳍或背部的肌肉组织作为组织样本，并用无水乙醇于固定保存，用于后续的 DNA 提取。标本的其余部分则保存于 8%的甲醛溶液中，长期保藏于中国科学院水生生物研究所水生生物博物馆

（Museum of Aquatic Organisms, IHB, CAS ），作为凭证标本以供后续的复核检视与研究。

物种的形态鉴定参照《青藏高原鱼类》、《西藏鱼类及其资源》、《中国动物志• 鲤形目(下卷)》、《中国动物志•鲇形目》[25-28]等专著和新种文献[29-33]。

1.3 鱼类样本 DNA 提取、扩增与测序

使用 Ezup 柱式动物基因组 DNA 抽提试剂盒(生工生物工程(上海)股份有限公司)，从鱼鳍或肌肉组织中提取总 DNA。提取的 DNA 首先使用 NanoPhotometer® N50 超微量分光光度计(IMPLEN，德国)进行浓度与纯度检测。随后，使用东胜龙 ETC811 PCR 仪(苏州东胜兴业科学仪器有限公司)进行目标片段的 PCR 扩增（表 2）。鉴于不同引物针对的目标类群及扩增片段存在差异，PCR 反应条件(包括退火温度和反应体系等)依据相应引物的原始文献进行设置[20,33-36]（附件 2）。PCR 产物经 1% TAE 琼脂糖凝胶电泳检测后，将条带清晰的 PCR 产物送往生工生物工程(上海)股份有限公司进行测序。

表 2 本研究 PCR 扩增所用引物信息汇总表

Tab. 2 Summary of primer information used for PCR amplification in this study

基因	引物名称	序列(5′—3′)	扩增片段长度
COI	FishF1	TCAACCAACCACAAAGACATTGGCAC	~655bp
	FishR1	TAGACTTCTGGGTGGCCAAAGAATCA	
	M13F	TGTAAAACGACGGCCAGT	注：M13F、M13R 为引物尾序列，与 FishF1、FishF2 组合使用
	M13R	CAGGAAACAGCTATGAC	
Cyt *b*	L14724	GACTTGAAAAACCACCGTTG	~1140bp
	H15915	CTCCGATCT CCGGATTACAAGAC	
	Cyt bL	CTAAAACACCGCAACCCACA	~1320bp
	Cyt bR	CAGTGGTGGGAGTTAAAATCTCCD	
	FZK	ATGACTTGAAAAACCACCGTT	~1135bp
	RZK	CGATCTCCGGATTACAAGACCG	

1.4 公共数据库序列信息获取与整理

为确保本研究构建的雅鲁藏布江鱼类 DNA 条形码数据库的完整性与准确性，在开展自主测序的同时，从公共数据库中检索所有本流域物种序列[37]，依据地理信息过滤与分类学复核策略，筛选分类信息、序列地理来源确定的序列[38]。

检索数据库包括 GenBank 数据库和生命条形码数据系统（Barcode of Life Data Systems）[39]。针对物种命名不统一、地理标签模糊等问题，我们采用了双重检索策略：(1)地理导向检索：以“Yarlung Tsangpo”、“Yarlung Zangbo”、“Yalu Zangbu”、“Brahmaputra”，主要支流名称(如“Lhasa River”、“Nyang River”)及相关地名（如“Lhasa”、“Pasighat”）等词汇作为关键词；(2)物种导向检索：依据流域已发表的鱼类物种名录，逐一使用拉丁学名在公开数据库进行精确查找，在 R 软件(V4.4.3)环境中使用 rentrez 包进行对这些数据进行批量下载[40-41]。随后进行分级递进式筛选(图 1)。首先进行 “地理信息过滤” ：仅保留采样地点明确标注位于雅鲁藏布江流域的序列。对于仅标注“Yarlung Tsangpo River”或“Xizang”等宽泛信息的记录，予以暂时保留并标记，以待后续与更详细信息交叉验证。其次执行 “分类学鉴定复核” ：对通过地理筛选的序列，我们核查其物种分类学标注的可靠性。优先采纳来源于有可供查验的标本凭证号并指明标本保存机构的记录，对于分类信息存疑(如疑似同物异名)的序列，进行单独备注，审慎评估后决定是否纳入。

2. 数据处理、质量控制与数据集构建

围绕“可复用、可扩展、可验证”的数据集构建目标，对自主测序数据与公共整合数据实施统一的处理流程、严格的质量控制及可靠性评估，最终构建结构规范的雅鲁藏布江鱼类 DNA 条形码数据集。

2.1 数据处理、质量控制和可靠性评估

首先将自 1.3 步骤获取的自主测序序列数据和使用 Bioedit（V 7.2）进行手动编辑和拼接[42]，并且参照序列的峰图在 Chromas 2.6.6 （http://technelysium.com.au/chromas.html）软件中核对原始峰图，人工修正测序误差，确保序列准确性。而后将自 1.4 步骤获取的公共数据集序列数据和自主测序序列数据合并，并对所有序列均进行长度筛选和翻译检查，以排除存在提前终止密码子或移码突变的异常序列，确保纳入数据集的序列均为合适的基因片段，按此流程构成初始数据集。而后在 R 软件（V4.4.3）环境中进行

序列特征分析，采用 Biostrings 包对 FASTA 格式序列进行读取[43]，并统计各序列的长度及 A、T、G、C 四种碱基的含量。以 A、T、G、C 四种碱基总数为分母，计算其相对含量并以百分比形式表示。在此基础上，分别统计各碱基含量及序列长度的最小值、最大值、平均值和标准误，并进一步比较 A + T 与 G + C 的含量差异，以分析序列的 AT 碱基偏倚特征。然后运用 MEGA-X 软件将本研究所用的序列进行对齐和裁剪[44]。

在确认序列长度与碱基组成符合条形码基因特征后，基于已对齐的序列进行 DNA 条形码数据集的可靠性评估。采用 ape 包中 dist.dna 函数，在 K2P（Kimura-2-parameter）模型下计算序列间两两遗传距离[45]，并对缺失位点采用成对删除策略[7,46]。以物种为单位计算平均和最大种内 K2P 距离，并将最近邻种间距离（nearest-neighbor distance, NN）定义为该物种与其他物种之间的最小种间距离；当 NN 小于或等于种内差异（平均或最大值）时，判定该物种缺乏条形码间隙。

2.2 数据集构建

在完成样品采集、序列获取及数据质量评估的基础上，构建雅鲁藏布江流域鱼类 DNA 条形码数据集。数据集采用元数据与序列数据分离的结构进行管理，其中元数据（metadata, data about data）以电子表格（.xlsx）形式存储，以样品和 DNA 条形码序列为基本记录单元，序列数据以文本文件（.txt）形式存储。通过唯一 RecordID（记录编号）实现元数据和序列数据之间的精准关联与可追溯。

数据集主要字段包括记录号、物种属性、数据来源、标本编号、物种分类信息（目、科、属、种）、登录号、基因名称、采集地以及采集年份。物种的分类信息及有效学名依据 Eschmeyer's Catalog of Fishes 进行校对，相关名称参考《拉汉世界鱼类系统名典》并结合钱建硕等（2026）基于大语言模型构建的世界鱼类拉汉对偶翻译系统进行核验与补充[47-49]。在数据集构建过程中，对所有字段进行统一命名和格式规范(表 3)，避免信息冗余与歧义。同时，对重复记录和缺失信息进行检查与整理，确保每条序列均具备完整或可追溯的基础信息。最终形成结构清晰、信息完整、便于查询和扩展的鱼类 DNA 条形码数据集，为后续的数据分析、序列比对及数据集应用提供基础支撑。

表 3 雅鲁藏布江流域鱼类 DNA 条形码数据集中所使用字段的描述

Tab. 3 Descriptions of the fields used in the fish DNA barcode dataset from the Yarlung Tsangpo River Basin

字段	描述
Record_ID	记录编号：作为唯一标识符，与序列文件（.txt）中对应序列的 Header 行编号一一对应，以确保元数据与序列数据之间的准确关联和可追溯性
NativeStatus	物种属性：指示物种是否为研究区域内的本地物种（Native species）或非本地物种（Non-native species）
DataSource	数据来源：指示序列数据的来源类型，包括本研究自行测序获得（Labgenerated）或来源于公共数据库（Publicdatabase）
Specimen_ID	标本编号：记录每份样品在采集单位或保存机构中的唯一编号，用于唯一标识对应的生物实体或凭证标本
Order	物种的目阶元分类信息
Family	物种的科阶元分类信息
Genus	物种的属阶元分类信息
Chinese_name	物种的中文正式名
Latin_scientific_name	物种的拉丁文学名
accession	条形码序列在公共序列数据库(如 GenBank、BOLD Systems)中对应的唯一登录号(Accession Number)
Gene	序列对应的基因的名称
Sequencing_primers	序列测序所用引物的信息
Collection_Site	获取样本的地理位置信息
Collection_Year	获取样本的时间信息

3. 数据描述与科学发现

本部分系统阐述数据集的基本特征、序列特征及遗传分化特征，明确数据集的概况、可靠性、科学价值与应用潜力。

3.1 数据集概况

本研究构建的雅鲁藏布江流域鱼类 DNA 条形码数据集共收录序列 3,174 条。其中，2,890 条序列为自主测序获得，约占总序列数的 91.1%；其余 284 条序列来源于 GenBank、BOLD 等公共数据库，约占总序列数的 8.9%。所有序列均通过长度筛选和翻译校验，未检测到提前终止密码子或移码突变，符合 DNA 条形码分析的质量要求，数据集不仅涵盖 COI 与 Cyt *b* 两种常用分子标记，还包括 12S、16S、ND6 等其他分子标记（图 3 C）。相较于既有的雅鲁藏布江流域鱼类 DNA 条形码公开数据，本研究在序列数量与物种覆盖度方面对雅鲁藏布江流域鱼类 DNA 条形码数据实现了显著提升；在序列规模、分类覆盖度与数据完整性等方面均达到较高水平。所有序列均配套完整的采集信息，包括采集地点与采集年份，序列涵盖的海拔范围为 155—4,600 m，时间范围为 1998—2024 年，并且配套部分测序物种彩色照片。

在分类组成上，本数据集系统覆盖了雅鲁藏布江流域已记录鱼类的主要分类类群（数据集鱼类名录详见附件 3，各阶元组成特征见图 3 A、B）。在目阶元水平上，雅鲁藏布江流域有记录的鱼类数据包括 13 个目，本数据集涵盖其中的 8 个目，分别为鲤形目（Cypriniformes）、鲇形目（Siluriformes）、颌针鱼目（Beloniformes）、鰕虎鱼目（Gobiiformes）、鲻形目（Mugiliformes）、攀鲈目（Anabantiformes）、鳉形目（Cyprinodontiformes）和合鳃鱼目（Synbranchiformes），目阶元覆盖率约 61.54%。不同目之间的序列数量分布存在明显不均衡。其中，鲤形目占据绝对优势，其序列数量约占数据集序列总数量的 85.01%；鲇形目次之，占比约为 13.57%；其余各目单独占比均不足 1%。这一数据特征反映出我国内陆水体鱼类以鲤形目和鲇形目为主体的典型特征。

在科阶元水平上，鲤科（Cyprinidae）为数据集中最具丰富的类群，其序列数量占数据集总量的 60% 以上；其次为条鳅科（Nemacheilidae）和鮡科（Sisoridae），占比分别为 15.02%和 13.35%。在属阶元水平上，裂腹鱼属（*Schizothorax*）、高原鳅属（*Triplophysa*）和墨头鱼属（*Garra*）的序列数量最为丰富。这些属不仅是雅鲁藏布江流域的优势类群，也是传统形态分类中争议较多的类群，本 DNA 条形码数据为后续开展物种界定、隐存种识别、系统发育关系解析及分类修订等研究提供了重要的数据支撑。

在种阶元水平上，数据集共包含 78 种鱼类，其中土著鱼类 62 种，外来鱼类 16 种。覆盖流域土著物种的 40%，已报道外来物种的 53.33%，对雅鲁藏布江特有鱼类的覆盖率为 69%，对模式产地位于雅鲁藏布江的鱼类覆盖率为 68.6%，对泸公河汇口及其上游江段鱼类覆盖率达到 100%。

3.2 序列特征

所有序列的碱基组成见表 4。DNA 序列长度的平均长度约为 813 bp，四种碱基的平均含量分别为：G 16.76%、C 27.33%、A 26.82%、T 29.08%，碱基组成分析结果表明，数据集的序列数据中 A＋T 含量（55.91%），明显高于 G ＋ C 含量（44.09%），整体呈现出显著的 AT 碱基偏倚性，符合脊椎动物线粒体 DNA 的典型特征[50]。COI 序列数据的序列长度范围为 461-1779 bp，Cyt *b* 序列数据的序列长度范围为 798-1390 bp。整体来看，数据集中各碱基含量的变化范围较小，标准误均处于较低水平，不同基因片段的序列长度分布存在一定差异，但整体离散程度较小，未发现明显异常值。表明所获序列质量较为均一稳定，为后续的遗传分析与物种鉴定提供了可靠的数据基础。

表 4 序列碱基组成

Tab.4 The base composition of the sample sequence

碱基类型	最小含量	平均含量	最大含量	标准误
G%	11.00	16.76	35.21	0.04
C%	9.06	27.33	36.23	0.04
A%	12.26	26.82	37.68	0.03
T%	16.67	29.08	41.38	0.03

3.3 种间与种内遗传距离特征

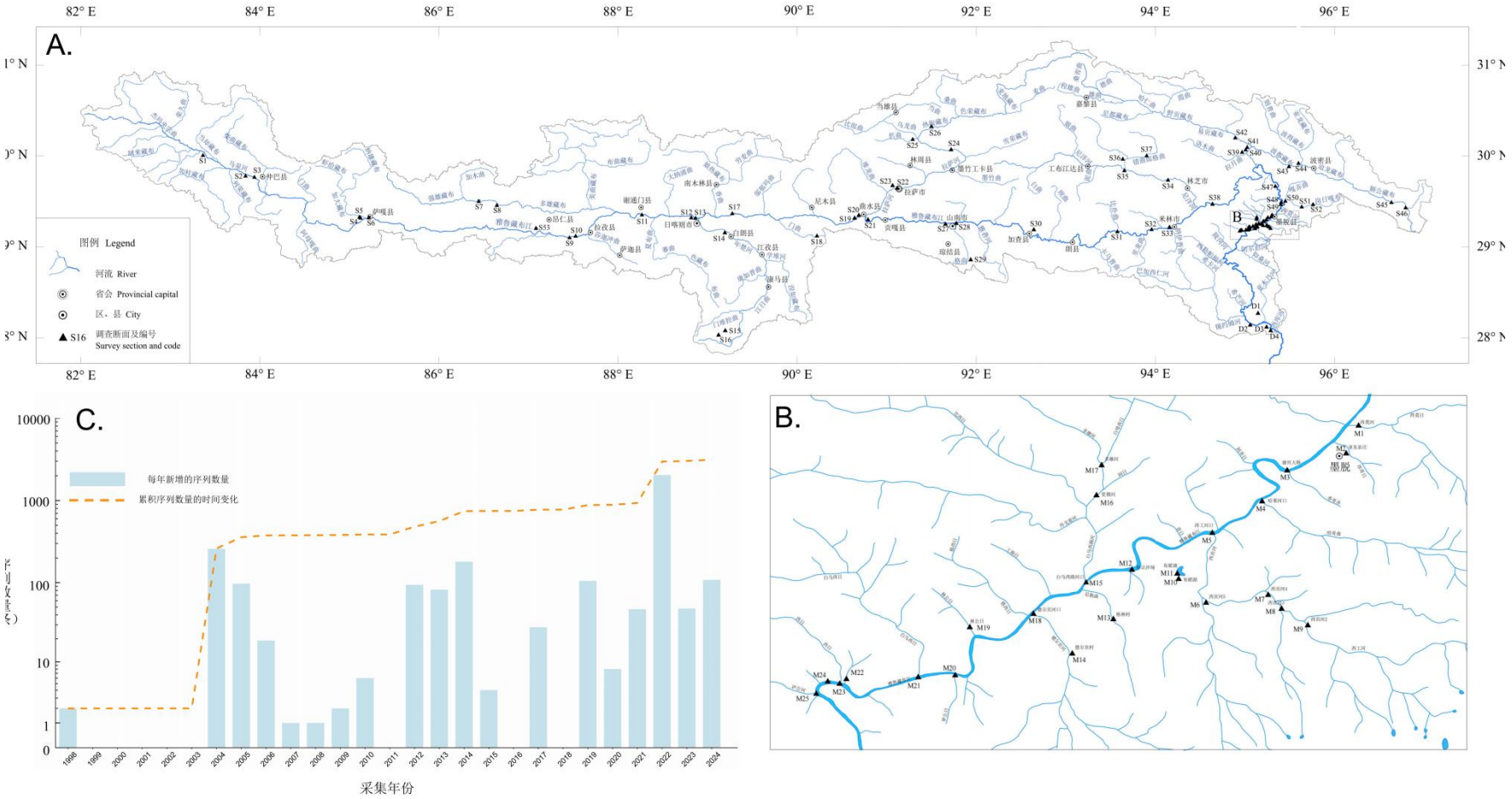

图 2 研究区域、采样点分布及 DNA 条形码序列数据的时间累积情况（A）雅鲁藏布江流域总体示意图，展示主要河流体系及采样断面分布；（B）雅鲁藏布江墨脱段局部放大图，显示研究区内采样点的空间分布；（C）DNA 条形码序列数据的时间累积情况，其中柱状图表示各年份新增序列数量，虚线表示累计序列数量随时间的变化

Fig. 2 Study area, sampling site distribution, and temporal accumulation of DNA barcode sequences (A) Overview of the Yarlung Tsangpo River Basin, showing the major river system and the distribution of sampling sections; (B) Enlarged view of the Motuo section of the Yarlung Tsangpo River, illustrating the spatial distribution of sampling sites; (C) Temporal accumulation of DNA barcode sequences, where bars represent the number of newly generated sequences per year and the dashed line indicates the cumulative number of sequences over time

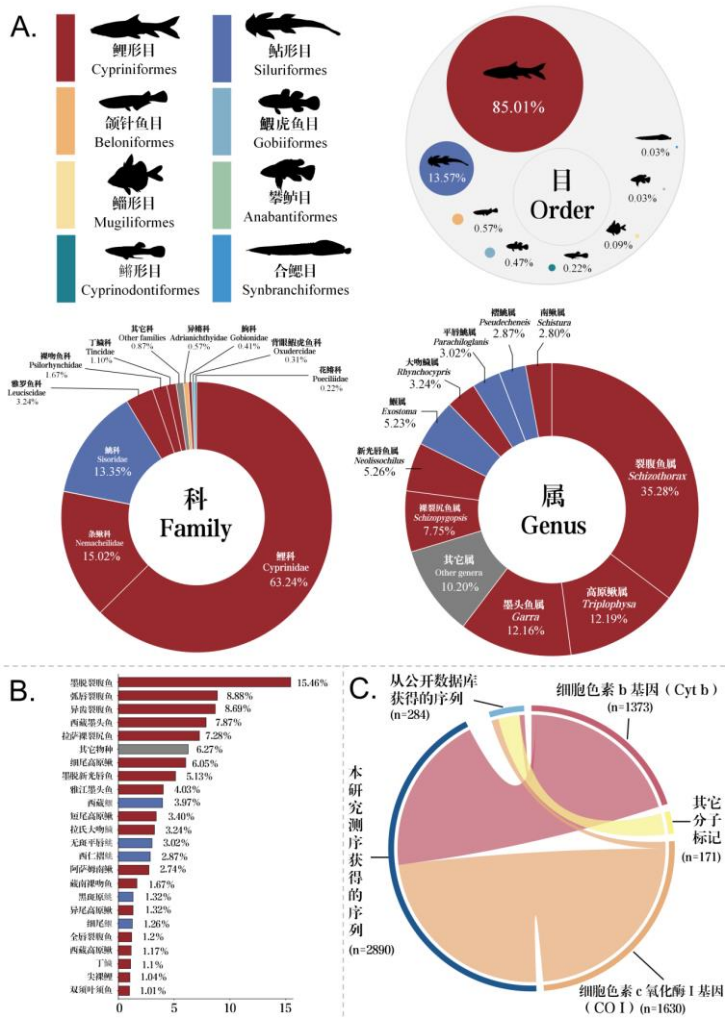

图 3 雅鲁藏布江流域鱼类 DNA 条形码数据集的分类组成特征

Fig.3 Taxonomic composition of the fish DNA barcode database for the Yarlung Tsangpo River Basin

种内与种间遗传距离的相对大小关系是评估 DNA 条形码参考数据集能否有效支撑物种分子鉴定的重要依据。一般认为，若物种间的遗传分化显著高于种内变异，即存在清晰的“条形码间隙（barcode gap）”，则该条形码体系具有较高的物种区分能力。本研究通过比较数据集中各物种的种内遗传变异与其最近邻种间遗传距离（nearest-neighbor distance, NN），对雅鲁藏布江流域鱼类 DNA 条形码数据集在物种区分层面的整体可靠性进行了检验。

针对 DNA 条形码数据集中占比最大的 COI 与 Cyt *b* 序列，采用 Kimura 2-parameter（K2P）模型分别计算了各物种的平均种内遗传距离（mean intraspecific distance, Mean Intra）、最大种内遗传距离（maximum intraspecific distance, Max Intra）以及最近邻种间遗传距离，并通过 NN-Mean Intra 和 NN-Max Intra 的散点图形式进行了可视化（图 4）。在散点图中，虚线表示种内与种间遗传距离相等的理论分界线，其上方区域代表存在条形码间隙。

结果显示，无论以平均种内遗传距离还是最大种内遗传距离作为比较指标，大多数物种的种内遗传距离均处于较低水平，而其最近邻种间遗传距离相对更高，散点主要分布于 $y = x$ 对角线以上，表明整体上存在较为清晰的 DNA 条形码间隙。这一结果说明，雅鲁藏布江流域鱼类 DNA 条形码数据集在总体上能够有效区分不同物种，具有较高的分子鉴定可靠性。尽管整体趋势符合这一理想状态，但在少数近缘类群中仍观察到种内与种间遗传距离的局部重叠现象，部分散点落于对角线下方。例如，裸裂尻鱼属（*Schizopygopsis*）和裂腹鱼属（*Schizothorax*）中有 5 个物种（拉萨裸裂尻鱼 *S. younghusbandi*、全唇裂腹

鱼 *S. integrilabiatus*、墨脱裂腹鱼 *S. molesworthi*、弧唇裂腹鱼 *S. curvilabiatus*、异齿裂腹鱼 *S. o'connori*）不存在条形码间隙，其最近邻种间遗传距离较低，与种内遗传变异范围存在一定程度的交叠。推测与近期分化、杂交或谱系未完全分离相关，也不排除公共数据库鉴定错误的可能。因此，对该类近缘物种进行分子鉴定时，建议结合形态学特征进行综合判定，以提升鉴定的准确性[51-55]。

总体而言，基于遗传距离比较的分析结果表明，本研究构建的雅鲁藏布江流域鱼类 DNA 条形码数据集在遗传分化层面具有良好的物种区分能力，为后续基于分子序列的鱼类物种鉴定与生物多样性研究提供了可靠的数据基础。

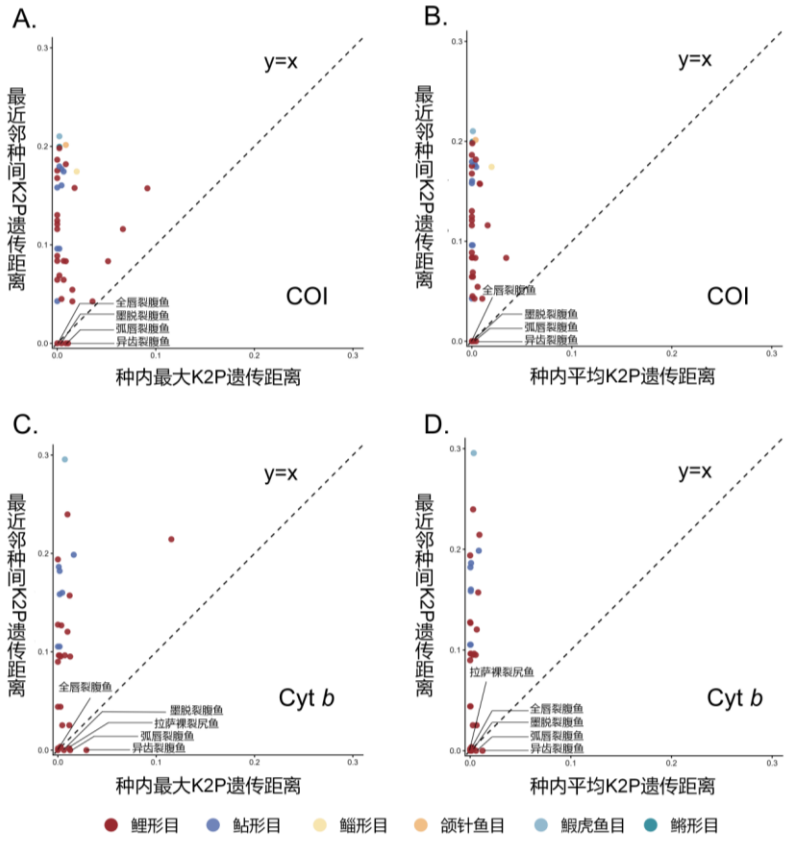

图 4 基于 Cyt *b* 与 COI 基因的 DNA 条形码间隙分析（K2P 距离）

Fig.4 DNA barcoding gap analyses based on Cyt *b* and COI genes using K2P distances

4. 数据集使用方法和建议

本节旨在详细介绍雅鲁藏布江鱼类 DNA 条形码数据集的获取与使用方法，并为不同研究场景提供应用建议，以期最大化其科学价值。

4.1 数据集存储位置

为严格遵循 FAIR 原则保障数据可发现性、可访问性与可重用性，本研究所构建的雅鲁藏布江鱼类 DNA 条形码数据集被存储于 ScienceDB(科学数据银行，DOI：10.57760/sciencedb.36688)，用户可自由下载用于本地化部署与研究。

4.2 本地化 BLAST 数据集的构建方法

对于涉及批量序列鉴定、敏感数据比对或需要高频次访问的研究需求，我们强烈建议用户在本地计算环境中构建并利用本数据集。相较于依赖网络的在线 BLAST 服务，本地化方案具有显著优势：（1）高速比对：不受网络延迟与服务器队列限制，特别适合海量 eDNA 测序数据的快速筛查；（2）高度可定制：研究者可根据具体课题，灵活筛选或增补数据集内容（如仅使用裂腹鱼属数据）；（3）数据安全与隐私：确保未公开的敏感序列数据（如新发现物种）在封闭环境中完成比对分析；（4）流程可重复与自动化：本地流程易于整合进生物信息学分析流程，保障研究结果的完全可复现。

本地 BLAST 数据集的构建与使用流程简明。首先，需在操作系统安装 NCBI BLAST+工具，并在 R 环境中安装 rBLAST 及相关程序包[43,54-56]。随后，以本数据集提供的标准化 FAST 文件(如 Yarlung_Fish_COI.fasta)作为参考序列，利用 makeblastdb()函数即可构建本地核苷酸数据集(dbtype＝‘nucl’)。数据集构建完成后，通过 blast()函数加载，并使用 predict()函数对待查询序列执行比对。该过程将输出包括匹配物种、序列相似性、比对长度及统计显著性(E 值)在内的关键信息，从而实现快速、准确的物种鉴定。完整示例代码见附件 4。

4.3 数据集应用场景与建议

为充分发挥本数据集在雅鲁藏布江流域这一特殊生态区域的科学价值，建议在以下应用中优先采用：（1）疑难类群鉴定：针对裂腹鱼属、高原鳅属等形态相似或存在杂交的类群，使用本地的、覆盖全面的序列集进行比对，可有效避免因全球数据集中高原鱼类代表性不足而导致的错误匹配或低分辨率问题；（2）支撑 eDNA 宏条形码研究：本数据集可作为开展雅鲁藏布江流域 eDNA 研究的重要参考数据集。其中包含的 COI 和 Cyt *b* 两种核心分子标记已被广泛应用于 eDNA 监测并发展出多种通用引物体系（如针对 COI 的 VCOI-U 以及针对 Cyt *b* 的 edm_Cytb-F/R[21,57]），该数据集还可以帮助对 eDNA 测序结果进行分类学注释，并为设计流域特异性引物与探针提供数据基础。需要指出的是，受样本获取条件及公开序列可用性限制，数据集目前未覆盖流域内全部已记录物种，这主要是公共数据库中泸公河口以下分布物种序列缺少可靠的地理分布。此外，部分近缘类群在单一线粒体标记下仍可能存在分辨能力有限的情况，相关问题有待在后续研究中通过扩展采样范围、引入更多分子标记予以进一步完善。

本数据集也可为其他流域物种间遗传分化、谱系演化等提供支撑，对评估全球 DNA 条形码数据库在高原河流鱼类研究中的适用性具有重要意义。未来可在本数据集基础上整合全球数据库中相关物种序列，开展跨流域比较分析，以评估不同流域种群间的遗传分化模式及外源序列在本流域鱼类 DNA 条形码研究中的适用性。

通过上述本地化、结构化的应用方式，本数据集旨在从一个静态的数据资源，转变为支持雅鲁藏布江流域鱼类多样性长期动态监测、保护生物学研究及生态健康评估的分子工具平台。本数据集也将持续更新与维护。通过版本控制，定期整合后续采集并经过严格质控的新序列数据和标本照片，以提供全面、准确的分子鉴定序列。

致谢：感谢中国科学院水生生物研究所水生生物博物馆黄俊豪、徐一扬参加野外样本收集和部分物种鉴定工作，研究生李茜茜、刘慧林、王勤参与野外考察和样本处理工作，研究生邴厚骅、姚佳琦和本科生李焱参与核酸提取和数据处理工作，在此一并表示感谢！

5 附件

附件见电子版（DOI 10.18307/2026.0635）。

6 参考文献

[1] 中国科学院青藏高原综合科学考察队. 西藏河流与湖泊. 北京: 科学出版社, 1984.

[2] He DK, Chen JN, Ding LY, et al. The status and distribution pattern of fish diversity in the Yarlung Tsangpo River. Biodiversity Science, 2024, 32(11): 24143. DOI: 10.17520/biods.2024143.[何德奎, 陈晋南, 丁刘勇等. 雅鲁藏布江鱼类多样性现状与分布格局. 生物多样性, 2024, 32(11): 24143.]

[3] He DK, Sui XY, Sun HY, et al. Diversity, pattern and ecological drivers of freshwater fish in China and adjacent areas. Reviews in Fish Biology and Fisheries, 2020, 30(3): 387-404. DOI: 10.1007/s11160-020-09600-4.

[4] Sui XY, Lin PC, Ding LY, et al. Genetic insights into drainage evolution: Late Miocene river capture in the eastern Himalaya. Journal of Systematics and Evolution, 2026, 64(1): 33-48. DOI: 10.1111/jse.70050.

[5] Li XX, Qian JS, Huang JH, et al. Genetic memory of fishes for Himalayan river evolution. Quaternary Sciences, 2023, 43(3): 819-837. DOI: 10.11928/j.issn.1001-7410.2023.03.11.[李茜茜, 钱建硕, 黄俊豪, 朱仁, 林鹏程, 何德奎. 鱼类遗传对喜马拉雅河流演化的记忆. 第四纪研究, 2023, 43(3): 819-837.]

[6] Convention on Biological Diversity (CBD). Kunming–Montreal Global Biodiversity Framework. Fifteenth meeting of the Conference of the Parties (COP15), 2022.

[7] Hebert PDN, Cywinska A, Ball SL, et al. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series B: Biological Sciences, 2003, 270(1512): 313-321. DOI: 10.1098/rspb.2002.2218.

[8] Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes, 2007, 7(3): 355-364. DOI: 10.1111/j.1471-8286.2007.01678.x.

[9] 陈双雅, 陈伟玲, 徐敦明, 等. 鱼类分子生物学物种鉴定技术. 厦门: 厦门大学出版社, 2012.

[10] Ma B, Li L, Wang JL, et al. Morphological and COI gene barcode analyses of *Schizothorax waltoni*, *S. o'connori* and their natural hybrids in the Yarlung Zangbo River. Journal of Fishery Sciences of China, 2018, 25(4): 753-761. DOI: 10.3724/SP.J.1118.2018.17359.[马波, 李雷, 王继隆, 等. 雅鲁藏布江拉萨裂腹鱼、异齿裂腹鱼及其自然杂交种的形态与 COI 基因条形码分析. 中国水产科学, 2018, 25(4): 753-761. DOI: 10.3724/SP.J.1118.2018.17359.]

[11] Shen Y J, Hubert N, Huang Y, et al. DNA barcoding the ichthyofauna of the Yangtze River: Insights from the molecular inventory of a mega-diverse temperate fauna. Molecular Ecology, 2016, 25(20): 5029-5044. DOI: 10.1111/mec.13853.

[12] Jiang PW, Li M, Zhang S, et al. Construction of a fish DNA metabarcoding reference database in the Pearl River Estuary based on mitochondrial COI and 12S rDNA genes. Journal of Shanghai Ocean University, 2023, 32(6): 1153-1165.[蒋佩文, 李敏, 张帅等. 基于线粒体 COI 和 12S rDNA 基因构建珠江河口鱼类 DNA 宏条形码数据库. 上海海洋大学学报, 2023, 32(6): 1153-1165.]

[13] Xu YQ, Liang XH, Chen XJ, et al. Construction of a DNA barcode reference database for common fishes in the Northwest Pacific based on the COI gene. Journal of Shanghai Ocean University, 2024, 33(4): 823-835.[徐颖琪, 梁绪虹, 陈新军等. 基于 COI 基因构建西北太平洋常见鱼类 DNA 条形码参考数据库. 上海海洋大学学报, 2024, 33(4): 823-835.]

[14] Xiao Y, Liu K, Wang YC, et al. Sequence and morphological diversity analysis of *Corbicula fluminea* in the Yangtze River Basin based on DNA barcoding. Journal of Lake Sciences, 2025, 37(1): 194-203.[肖杨, 刘凯, 王英才等. 基于 DNA 条形码的长江流域沼蛤序列和形态多样性分析. 湖泊科学, 2025, 37(1): 194-203.]

[15] Brosse S, Cuenot Y, Condachou C, et al. Near complete 12S DNA reference library for the freshwater fish of French Guiana, northern Amazonian region. Scientific Data, 2026, 13: 408. DOI: 10.1038/s41597-026-06811-5.

[16] Altermatt F, Couton M, Carraro L, et al. Utilizing aquatic environmental DNA to address global biodiversity targets. Nature Reviews Biodiversity, 2025, 1: 1-16. DOI: 10.1038/s44358-025-00044-x.

[17] Ji YQ, Diana A, Li XY, et al. High quality, granular, timely, trustworthy and efficient vertebrate species distribution data across a 30,000 km^2 protected area complex. Ecology Letters, 2025, 28(12): e70302. DOI: 10.1111/ele.70302.

[18] Zhou CJ, Ma K, Wang T, et al. Assessing current fish diversity in the Yellow River Basin by integrating large-scale barcoding and morphological data.Ecology and Evolution,2025, 15: e72617. DOI: 10.1002/ece3.72617.

[19] Shi X, Kou C, He C, et al. DNA barcode reference library and undetected diversity of fish species in the Yuanjiang River, China. Fishes, 2025, 10: 418. DOI: 10.3390/fishes10080418.

[20] 王子旺. 雅鲁藏布江下游鱼类 DNA 条形码数据库的构建及应用[学位论文]. 北京: 中国科学院大学, 2022.

[21] Feng X, Li B, Chen YF, et al. Species-level monitoring of rare and invasive fishes using eDNA metabarcoding in the middle and upper Yarlung Zangbo River, Tibet. Water Biology and Security, 2023, 2(1): 100089. DOI: 10.1016/j.watbs.2022.100089.

[22] Manel S, Guerin PE, Mouillot D, et al. Global determinants of freshwater and marine fish genetic diversity. Nature Communications, 2020, 11: 692. DOI: 10.1038/s41467-020-14409-7.

[23] Grush J, Noakes DLG, Moccia RD. The efficacy of clove oil as an anesthetic for the zebrafish, Danio rerio (Hamilton). Zebrafish, 2004, 1(1): 46-53. DOI: 10.1089/154585404774101671.

[24] American Veterinary Medical Association. AVMA Guidelines for the Euthanasia of Animals: 2013 edition. Schaumburg, IL: American Veterinary Medical Association, 2013.

[25] 武云飞, 吴翠珍. 青藏高原鱼类. 成都: 四川科学技术出版社, 1992.

[26] 西藏自治区水产局. 西藏鱼类及其资源. 北京: 中国农业出版社, 1995.

[27] 乐佩琦. 中国动物志 • 硬骨鱼纲 • 鲤形目(下). 北京: 科学出版社, 2000.

[28] 褚新洛, 郑葆珊, 戴定远. 中国动物志 • 硬骨鱼纲 • 鲇形目. 北京: 科学出版社, 1999.

[29] Gong Z, Lin PC, Liu F, et al. *Exostoma tibetana*, a new glyptosternine catfish from the lower Yarlung Tsangpo River drainage in

southeastern Tibet, China (Siluriformes: Sisoridae). Zootaxa, 2016, 4527: 392-402.

[30] Gong Z, Deng SQ, Wang J, et al. A new species of genus *Garra* (Cypriniformes: Cyprinidae) from the lower Yarlung Tsangpo River drainage in southeastern Tibet, China. Chinese Journal of Zoology, 2018, 53(6): 857-867.

[龚正, 邓树庆, 王健等. 西藏雅鲁藏布江下游墨头鱼属一新种(鲤形目: 鲤科). 动物学杂志, 2018, 53(6): 857-867.]

[31] Gong Z, Freyhof J, Wang J, et al. Two new species of *Garra* (Cypriniformes: Cyprinidae) from the lower Yarlung Tsangpo River drainage in southeastern Tibet, China. Zootaxa, 2018, 4532: 367-384.

[32] Chen YX, Tan HM, Lin PC, et al. Taxonomic revision of the Sisoridae (Osteichthyes: Siluriformes) fishes of the lower Yarlung Tsangpo River, with descriptions of three new species and one new record in China. Acta Hydrobiologica Sinica, 2024, 48: 920-949.[陈咏霞, 谭慧敏, 林鹏程等. 雅鲁藏布江下游鮡科鱼类分类整理及三新种和中国一新纪录种. 水生生物学报, 2024, 48: 920-949.]

[33] Ward RD, Zemlak TS, Innes BH, et al. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2005, 360(1462): 1847-1857. DOI: 10.1098/rstb.2005.1716.

[34] Messing J. New M13 vectors for cloning. Methods in Enzymology, 1983, 101: 20-78. DOI: 10.1016/0076-6879(83)01005-8.

[35] Xiao W, Zhang Y, Liu H. Molecular systematics of Xenocyprinae (Teleostei: Cyprinidae): taxonomy, biogeography, and coevolution of a special group restricted in East Asia. Molecular Phylogenetics and Evolution, 2001, 18(2): 163-173.

[36] 郑新. 中国子陵吻虾虎鱼和小黄黝鱼的种群结构及遗传多样性分析[学位论文]. 上海: 上海海洋大学, 2016.

[37] Oliveira LM, Knebelberger T, Landi M, et al. Assembling and auditing a comprehensive DNA barcode reference library for European marine fishes. Journal of Fish Biology, 2016, 89(6): 2741-2754. DOI: 10.1111/jfb.13169.

[38] Du TQ, Ding CZ, Ke Y, et al. A global dataset on species occurrences and functional traits of Schizothoracinae fish. Scientific Data, 2024, 11: 727. DOI: 10.1038/s41597-024-03098-2.

[39] Porter TM, Hajibabaei M. Over 2.5 million COI sequences in GenBank and growing. PLoS One, 2018, 13(9): e0200177. DOI: 10.1371/journal.pone.0200177.

[40] R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2025. https://www.R-project.org/

[41] Winter DJ. rentrez: an R package for the NCBI eUtils API. The R Journal, 2017, 9(2): 520-526.

[42] Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series, 1999, 41: 95-98.

[43] Pagès H, Aboyoun P, Gentleman R, et al. Biostrings: efficient manipulation of biological strings. 2024. R package version 2.74.1. DOI: 10.18129/B9.bioc.Biostrings. https://bioconductor.org/packages/Biostrings.

[44] Kumar S, Stecher G, Li M, et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution, 2018, 35(6): 1547-1549. DOI: 10.1093/molbev/msy096.

[45] Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 2019, 35: 526-528. DOI: 10.1093/bioinformatics/bty633.

[46] Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution, 1980, 16(2): 111-120. DOI: 10.1007/BF01731581.

[47] Fricke R, Eschmeyer WN, Van der Laan R, eds. Eschmeyer’s catalog of fishes: Genera, species, references. San Francisco: California Academy of Sciences, 2026. http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp.

[48] 伍汉霖, 邵广昭, 赖春福等. 拉汉世界鱼类系统名典. 青岛: 中国海洋大学出版社, 2017.

[49] Qian JS, Ding LY, Li YY, et al. Latin-Chinese dual translation of global fish taxonomic names using large language models. J Lake Sci, 2026, 38(1): 425-434. DOI:10.18307/2026.0161.[钱建硕, 丁刘勇, 李雁羽等. 基于大语言模型的世界鱼类拉汉对偶翻译. 湖泊科学, 2026, 38(1): 425-434.]

[50] Saccone C, De Giorgi C, Gissi C, et al. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene, 1999, 238(1): 195-209. DOI: 10.1016/S0378-1119(99)00270-X.

[51] Yu D, Zhang Z, Zhang J, et al. Genetic diversity and population demography of *Schizothorax molesworthi* from the Motuo area of lower reaches of the Yarlung Zangbo River and Lohit River. Acta Hydrobiologica Sinica, 2019, 43(5): 923-930. DOI:

10.7541/2019.109.[俞丹, 张智, 张健等. 基于 Cyt *b* 基因的雅鲁藏布江下游墨脱江段及察隅河墨脱裂腹鱼的遗传多样性及种群历史动态分析. 水生生物学报, 2019, 43(5): 923-930.]

[52] Peng JL, Wang XZ, He SP. The progress and application of DNA barcoding. Acta Hydrobiologica Sinica, 2008, 32(6): 916-919.[彭居俐, 王绪桢, 何舜平. DNA 条形码技术的研究进展及其应用. 水生生物学报, 2008, 32(6): 916-919.]

[53] Zhou JS, Zhang C, Liu HP, et al. Studies on identification of Schizothoracinae population in Tibet water system by DNA barcode. Freshwater Fisheries, 2019, 49(1): 33-41. DOI: 10.13721/j.cnki.dsyy.2019.01.006.

[周建设, 张驰, 刘海平等. DNA 条形码在西藏水系裂腹鱼亚科鱼类鉴定中的研究. 淡水渔业, 2019, 49(1): 33-41.]

[54] Zhang Z, Schwartz S, Wagner L, et al. A greedy algorithm for aligning DNA sequences. Journal of Computational Biology, 2000, 7(1-2): 203–214.

[55] Hahsler M, Nagar A. rBLAST: R interface for the Basic Local Alignment Search Tool. 2019. R package version 0.99.2. https://github.com/mhahsler/rBLAST.

[56] Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics, 2009, 10: 421. DOI: 10.1186/1471-2105-10-421.

[57] Wang ZC, Liu XL, Liang D, et al. VertU: universal multilocus primer sets for eDNA metabarcoding of vertebrate diversity, evaluated by both artificial and natural cases. Frontiers in Ecology and Evolution, 2023, 11: 1164206. DOI: 10.3389/fevo.2023.1164206.

附件目录：

附件 1 本研究中样点编号和样点中文名称对应表

附件 2 本研究中所用基因及其 PCR 引物、反应体系与扩增条件

附件 3 雅鲁藏布江鱼类 DNA 条形码数据集物种名录

附件 4 建立本地 Blast 数据集的方法

参考文献

附件 1 样点编号和样点中文名称对应表

Appendix 1 Sampling site codes and Chinese names used in this study

样点编号	样点中文名称
S1	西藏自治区日喀则市仲巴县帕羊镇大桥马泉河
S2	西藏自治区日喀则市仲巴县拉让乡珠珠村马泉河
S3	西藏自治区日喀则市仲巴县拉让乡过娃拉布塘 219 过道边当却藏布
S4	西藏自治区日喀则市萨嘎县加加镇他细淌勒雅鲁藏布江
S5	西藏自治区日喀则市萨嘎县加加镇醒星格里加达电站下游加大藏布
S6	西藏自治区日喀则市桑桑镇奎琼村下嘎曲桥下游
S7	西藏自治区日喀则市昂仁县桑桑镇尼仁村强雄藏布
S8	西藏自治区日喀则市桑桑镇奎琼村下嘎曲桥下游多雄藏布
S9	西藏自治区日喀则市拉孜县查务乡生尖村雅鲁藏布江
S10	西藏自治区日喀则市拉孜县查务乡加嘎村水文站雅鲁藏布江
S11	西藏自治区日喀则市谢通门县荣玛乡亚波村雅鲁藏布江
S12	西藏自治区日喀则市桑珠孜区除松村 G349 边雅鲁藏布江
S13	西藏自治区日喀则市桑珠孜区年楚河入雅江口
S14	西藏自治区日喀则市白朗县洛江镇扎西岗村年楚河
S15	西藏自治区日喀则市亚东县堆纳乡曲美雄谷多庆错入湖河流
S16	西藏自治区日喀则市亚东县堆纳乡勒哇村多庆错入湖河流麻曲
S17	西藏自治区日喀则市桑珠孜区和平机场雅鲁藏布江
S18	西藏自治区山南市浪卡子县白地乡加塘门曲
S19	西藏自治区拉萨市曲水县达嘎镇雅江大桥上游雅鲁藏布江
S20	西藏自治区拉萨市曲水县曲水镇雅鲁藏布江
S21	西藏自治区拉萨市曲水县曲水镇桑布日雅鲁藏布江
S22	西藏自治区日拉萨市城关区藏热大桥上游拉萨河
S23	西藏自治区拉萨市城关区拉鲁湿地
S24	西藏自治区拉萨市墨竹工卡县扎雪乡格老窝村拉萨河
S25	西藏自治区拉萨市林周县旁多乡旁多村拉萨河
S26	西藏自治区拉萨市林周县唐古乡曲果多热振藏布
S27	西藏自治区山南市乃东区甲村下游雅鲁藏布江
S28	西藏自治区山南市乃东区泽当镇泽当大桥下游雅鲁藏布江
S29	西藏自治区山南地区乃东县亚堆乡热木纳村雅砻河
S30	西藏自治区山南市加查县拉姆拉错
S31	西藏自治区林芝市米林市卧龙镇下觉村雅鲁藏布江
S32	西藏自治区林芝市米林市里龙乡玉松村雅鲁藏布江
S33	西藏自治区林芝市米林市米林镇雅鲁藏布江
S34	西藏自治区林芝市巴宜区更章乡多布村尼洋河

S35	西藏自治区林芝市工布江达县巴河镇丁当村尼洋河
S36	西藏自治区林芝市工布江达县巴河镇雪卡村错曲郎格曲
S37	西藏自治区林芝市工布江达县错高乡巴松措
S38	西藏自治区林芝市米林市丹娘乡朗嘎村雅鲁藏布江
S39	西藏自治区林芝市波密县通麦镇排龙村帕隆藏布
S40	西藏自治区林芝市波密县通麦镇 G318 小老虎嘴帕隆藏布
S41	西藏自治区林芝市波密县通麦镇帕隆藏布与易贡藏布汇合口
S42	西藏自治区林芝市波密县易贡乡柏村易贡湖大桥
S43	西藏自治区林芝市波密县古乡巴卡村帕隆藏布
S44	西藏自治区林芝市波密县扎木镇达卡村帕隆藏布
S45	西藏自治区昌都市八宿县然乌镇然乌错
S46	西藏自治区昌都市八宿县然乌镇仲巴村然乌错入湖河流真孔弄巴河/西藏自治区昌都市八宿县然乌镇来古冰川 3
S47	西藏自治区林芝市墨脱县加热萨乡加热萨村雅鲁藏布江
S48	西藏自治区林芝市墨脱县达木珞巴族自治乡达果桥金珠藏布
S49	西藏自治区林芝市墨脱县达木珞巴族自治乡卡布村雅鲁藏布江
S50	西藏自治区林芝市墨脱县达木珞巴族自治乡东久拉入金珠藏布瀑布
S51	西藏自治区林芝市墨脱县格当乡格当村金珠藏布
S52	西藏自治区林芝市墨脱县格当乡兴开村岗日嘎布
M1	西藏自治区林芝市墨脱县墨脱镇西漠河大桥西莫河
M2	西藏自治区林芝市墨脱县墨脱镇西漠河大桥西莫河亚东农庄
M3	西藏自治区林芝市墨脱县德兴乡德兴大桥上游雅鲁藏布江
M4	西藏自治区林芝市墨脱县墨脱镇哈果河口
M5	西藏自治区林芝市墨脱县背崩乡西工河口
M6	西藏自治区林芝市墨脱县背崩乡西贡河 5
M7	西藏自治区林芝市墨脱县背崩乡西贡河 4
M8	西藏自治区林芝市墨脱县背崩乡西贡河 3
M9	西藏自治区林芝市墨脱县背崩乡西贡河 1、西贡河 2
M10	西藏自治区林芝市墨脱县背崩乡亚让村布裙湖东南
M11	西藏自治区林芝市墨脱县背崩乡亚让村布裙湖南
M12	西藏自治区林芝市墨脱县背崩乡沙场雅鲁藏布江
M13	西藏自治区林芝市墨脱县背崩乡格林村尼勒河
M14	西藏自治区林芝市墨脱县背崩乡德尔贡村德尔贡河
M15	西藏自治区林芝市墨脱县背崩乡解放大桥白马西路河口
M16	西藏自治区林芝市墨脱县背崩乡易公白村更俄河
M17	西藏自治区林芝市墨脱县背崩乡易公白村多雄河
M18	西藏自治区林芝市墨脱县背崩乡江新村德尔贡河口
M19	西藏自治区林芝市墨脱县背崩乡地东村林公日
M20	西藏自治区林芝市墨脱县背崩乡地东村对面下游罗公日瀑布下
M21	西藏自治区林芝市墨脱县背崩乡地东村白马西日河口
M22	西藏自治区林芝市墨脱县背崩乡西让村下游无名溪流
M23	西藏自治区林芝市墨脱县背崩乡西让村西日河口
M24	西藏自治区林芝市墨脱县背崩乡西让村西日下游雅鲁藏布江
M25	西藏自治区林芝市墨脱县背崩乡西让村泸公河口

D1	西藏自治区林芝市墨脱县阿迪
D2	西藏自治区林芝市墨脱县罗登
D3	西藏自治区林芝市墨脱县仁更
D4	西藏自治区林芝市墨脱县巴昔卡

附件 2 本研究中所用基因及其 PCR 引物、反应体系与扩增条件

Appendix 2 Genes used in this study and their PCR primers, reaction systems, and amplification conditions

基因	引物名称	体系	反应条件	参考文献	适用类群
CO I	FishF1 FishF2 M13F M13R	2× Master Mix 12.5 μL，dNTPs 0.25 μL（10 mmol/L），Taq DNA 聚合酶 0.5 U，上、下游引物各 1 μL（10 μmol/L），基因组 DNA 模板 1 μL，无菌 ddH_2O 补足至 25 μL	94℃ 预变性 4 min，随后进行 37 个循环，每个循环包括 94℃ 变性 30 s、54℃ 退火 35 s、72℃ 延伸 50 s，循环结束后在 72℃ 终延伸 8 min	Ward et al., 2005 Messing, 1983	通用
Cytb	L14724 H15915 FZK RZK	2× Master Mix 12.5 μL，dNTPs 0.25 μL（10 mmol/L），Taq DNA 聚合酶 0.5 U，上、下游引物各 1 μL（10 μmol/L），基因组 DNA 模板 1 μL，无菌 ddH_2O 补足至 25 μL	94℃ 预变性 4 min，随后进行 37 个循环，每个循环包括 94℃ 变性 30 s、54℃ 退火 35 s、72℃ 延伸 50 s，循环结束后在 72℃ 终延伸 8 min	Xiao et al., 2001	通用 鳅属
	CytbL CytbR	PCR Master Mix：12.5 μL（含 $MgCl_2$、dNTP Mix、Taq 酶），上游引物：1.0 μL（55 ng/μL），下游引物：1.0 μL（55 ng/μL），DNA 模板：1.0 μL，ddH_2O：补足至 25 μL	94℃预变性 4 min，随后进行 35 个循环，每个循环包括 94℃变性 30 s、60℃退火 45 s、72℃延伸 1 min，循环结束后在 72℃终延伸 8 min，最后于 4℃保存	郑新, 2016	小黄黝鱼属

附件 3 雅鲁藏布江鱼类 DNA 条形码数据集物种名录

Appendix 3 Species list of fish included in the DNA barcode dataset from the Yarlung Tsangpo River

目	科	属	物种 中文名称	物种 拉丁文学名	物种属性
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Schizothorax* 裂腹鱼属	弧唇裂腹鱼	*Schizothorax curvilabiatus*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Schizothorax* 裂腹鱼属	墨脱裂腹鱼	*Schizothorax molesworthi*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Schizothorax* 裂腹鱼属	全唇裂腹鱼	*Schizothorax integrilabiatus*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Schizothorax* 裂腹鱼属	异齿裂腹鱼	*Schizothorax o'connori*	●

Cypriniformes 鲤形目	Cyprinidae 鲤科	*Garra* 墨头鱼属	墨脱墨头鱼	*Garra motuoensis*	●
Cypriniformes 鲤形目	Psilorhynchidae 裸吻鱼科	*Psilorhynchus* 裸吻鱼属	平鳍裸吻鱼	*Psilorhynchus homaloptera*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Bangana* 孟加拉鲮属	墨脱孟加拉鲮	*Bangana dero*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Garra* 墨头鱼属	西藏墨头鱼	*Garra tibetana*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Garra* 墨头鱼属	雅江墨头鱼	*Garra yajiangensis*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Neolissochilus* 新光唇鱼属	墨脱新光唇鱼	*Neolissochilus hexagonolepis*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Schistura* 南鳅属	阿萨姆南鳅	*Schistura scaturigina*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Schizopygopsis* 裸裂尻鱼属	拉萨裸裂尻鱼	*Schizopygopsis younghusbandi*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Triplophysa* 高原鳅属	短尾高原鳅	*Triplophysa brevicauda*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Triplophysa* 高原鳅属	细尾高原鳅	*Triplophysa stenura*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Triplophysa* 高原鳅属	异尾高原鳅	*Triplophysa stewarti*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Triplophysa* 高原鳅属	西藏高原鳅	*Triplophysa tibetana*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Triplophysa* 高原鳅属	达里湖高原鳅	*Triplophysa dalaica*	○
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Oxygymnocypris* 尖裸鲤属	尖裸鲤	*Oxygymnocypris stewartii*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Schizothorax* 裂腹鱼属	拉萨裂腹鱼	*Schizothorax waltoni*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Ptychobarbus* 叶须鱼属	双须叶须鱼	*Ptychobarbus dipogon*	●
Cypriniformes 鲤形目	Gobionidae 鮈科	*Pseudorasbora* 麦穗鱼属	麦穗鱼	*Pseudorasbora parva*	○
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Carassius* 鲫属	鲫	*Cyprinus auratus*	○
Cypriniformes 鲤形目	Gobionidae 鮈科	*Abbottina* 棒花鱼属	棒花鱼	*Pseudogobio rivularis*	○
Cypriniformes 鲤形目	Cobitidae 鳅科	*Misgurnus* 泥鳅属	泥鳅	*Misgurnus anguillicaudatus*	○
Cypriniformes 鲤形目	Cobitidae 鳅科	*Misgurnus* 泥鳅属	大鳞副泥鳅	*Paramisgurnus dabryanus*	○
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Cyprinus* 鲤属	鲤	*Cyprinus carpio*	○

Cypriniformes 鲤形目	Cyprinidae 鲤科	*Schizothorax* 裂腹鱼属	齐口裂腹鱼	*Schizothorax prenanti*	○
Cypriniformes 鲤形目	Tincidae 丁鲹科	*Tinca* 丁鲹属	丁鲹	*Tinca tinca*	○
Cypriniformes 鲤形目	Leuciscidae 雅罗鱼科	*Rhynchocypris* 大吻鲴属	拉氏大吻鲴	*Rhynchocypris lagowskii*	○
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Tariqilabeo* 塔氏鲮属	彩花塔氏鲮	*Tariqilabeo latius*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Schistura* 南鳅属	裂条南鳅	*Schistura rupecula*	●
Cypriniformes 鲤形目	Danionidae 鲃科	*Salmostoma* 鲑口波鱼属	孟加拉鲑口波鱼	*Salmostoma bacaila*	●
Cypriniformes 鲤形目	Danionidae 鲃科	*Salmostoma* 鲑口波鱼属	费洛鲑口波鱼	*Salmostoma phulo*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Puntius* 小鲃属	斑尾小鲃	*Puntius sophore*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Puntius* 小鲃属	沼泽小鲃	*Puntius chola*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Osteobrama* 骨鳊属	孟加拉骨鳊	*Osteobrama cotio*	●
Cypriniformes 鲤形目	Cobitidae 鳅科	*Lepidocephalichthys* 似鳞头鳅属	斑尾似鳞头鳅	*Lepidocephalichthys goalparensis*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Garra* 墨头鱼属	阿氏墨头鱼	*Garra arupi*	●
Cypriniformes 鲤形目	Danionidae 鲃科	*Devario* 神鲃属	等鳍神鲃	*Devario aequipinnatus*	●
Cypriniformes 鲤形目	Danionidae 鲃科	*Danio* 鲃属	孟加拉鲃	*Danio dangila*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Chagunius* 沙昆鲤属	沙昆鲤	*Chagunius chagunio*	●
Cypriniformes 鲤形目	Danionidae 鲃科	*Barilius* 低线鱲属	低线鱲	*Barilius barila*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Aborichthys* 阿波鳅属	墨脱阿波鳅	*Aborichthys kempi*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Triplophysa* 高原鳅属	山南高原鳅	*Triplophysa shannanensis*	●
Cypriniformes 鲤形目	Cyprinidae 鲤科	*Schizothorax* 裂腹鱼属	巨须裂腹鱼	*Schizothorax macropogon*	●
Cypriniformes 鲤形目	Nemacheilidae 条鳅科	*Triplophysa* 高原鳅属	硬刺高原鳅	*Triplophysa scleroptera*	○
Siluriformes 鲇形目	Sisoridae 鮡科	*Pseudecheneis* 褶鮡属	西仁褶鮡	*Pseudecheneis sirenica*	●
Siluriformes 鲇形目	Sisoridae 鮡科	*Parachiloglanis* 平唇鮡属	无斑平唇鮡	*Parachiloglanis immaculata*	●

Siluriformes 鲇形目	Sisoridae 鮡科	*Glyptothorax* 纹胸鮡属	墨脱纹胸鮡	*Glyptothorax medogensis*	●
Siluriformes 鲇形目	Sisoridae 鮡科	*Exostoma* 鳅属	西藏鳅	*Exostoma tibetanum*	●
Siluriformes 鲇形目	Sisoridae 鮡科	*Exostoma* 鳅属	细尾鳅	*Exostoma tenuicaudatum*	●
Siluriformes 鲇形目	Sisoridae 鮡科	*Creteuchiloglanis* 异鮡属	扁头异鮡	*Creteuchiloglanis kamengensis*	●
Siluriformes 鲇形目	Sisoridae 鮡科	*Glyptosternon* 原鮡属	黑斑原鮡	*Glyptosternon maculatum*	●
Siluriformes 鲇形目	Siluridae 鲇科	*Silurus* 鲇属	鲇	*Silurus asotus*	○
Siluriformes 鲇形目	Bagridae 鲿科	*Sperata* 诺鲿属	刀诺鲿	*Sperata aor*	●
Siluriformes 鲇形目	Bagridae 鲿科	*Mystus* 鲿属	条纹鲿	*Mystus vittatus*	●
Siluriformes 鲇形目	Sisoridae 鮡科	*Glyptothorax* 纹胸鮡属	条纹纹胸鮡	*Glyptothorax striatus*	●
Siluriformes 鲇形目	Sisoridae 鮡科	*Gagata* 黑鮡属	尾斑黑鮡	*Gagata cenia*	●
Siluriformes 鲇形目	Amblycipitidae 钝头鮠科	*Amblyceps* 钝头鮠属	藏南钝头鮠	*Amblyceps arunachalense*	●
Siluriformes 鲇形目	Ailiidae 艾利鲇科	*Ailia* 艾利鲇属	光背鲇	*Ailia coila*	●
Gobiiformes 虾虎鱼目	Oxudercidae 背眼虾虎鱼科	*Rhinogobius* 吻虾虎鱼属	波氏吻虾虎	*Rhinogobius cliffordpopei*	○
Gobiiformes 虾虎鱼目	Gobiidae 虾虎鱼科	*Glossogobius* 舌虾虎鱼属	舌虾虎鱼	*Glossogobius giuris*	●
Gobiiformes 虾虎鱼目	Odontobutidae 沙塘鳢科	*Micropercops* 小黄黝鱼属	小黄黝鱼	*Micropercops swinhonis*	○
Mugiliformes 鲻形目	Ambassidae 双边鱼科	*Pseudambassis* 拟双边鱼属	印度拟双边鱼	*Parambassis lala*	●
Mugiliformes 鲻形目	Ambassidae 双边鱼科	*Chanda* 玻璃鱼属	溪流玻璃鱼	*Chanda nama*	●
Beloniformes 颌针鱼目	Adrianichthyidae 异鳉科	*Oryzias* 青鳉属	中华青鳉	*Oryzias sinensis*	○
Cyprinodontiforme s 鳉形目	Poeciliidae 花鳉科	*Gambusia* 食蚊鱼属	食蚊鱼	*Gambusia affinis*	○
Synbranchiformes 合鳃目	Mastacembelidae 棘鳅科	*Macrognathus pancalus* 吻棘鳅属	潘卡尔吻棘鳅	*Macrognathus pancalus*	●
Anabantiformes 攀鲈目	Channidae 鳢科	*Channa* 鳢属	翠鳢	*Channa punctata*	●

注：附件 3 中，●表示本土物种，○表示外来物种

附件 4 DNA 条形码本地 BLAST 分析的完整流程及示例代码
Appendix 4 Complete workflow and example code for local BLAST analysis of DNA barcodes

Blast（Basic Local Alignment Search Tool）局部序列比对基本检索工具，是 NCBI 开发的一款基于序列相似性的数据库搜索程序（Zhang et al., 2000）。Blast 是一款集成软件，包括 BLASTp、BLASTn、BLASTx、tBLASTn 和 tBLASTx 模块，通过调用不同的比对模块，满足不同的研究需求。使用 Blast 检索工具需具备三个基本要素：比对数据库（database）、待比对序列（query）和目标序列（subject）。Blast 工具使用有两种方式：在线版本（https://blast.ncbi.nlm.nih.gov/Blast.cgi）和本地版本。与在线 Blast 相比，本地 Blast 具有以下优势：1）比对速度：由于在线 Blast 序列比对受限于网络和比对序列的数量，本地 Blast 的序列比对速度远超在线版本；2）自定义：可根据特定的研究需要，建立本地 Blast 数据库进行序列比对和检索；3）数据安全性：建立本地 Blast 数据库可为敏感或专用数据提供更多的隐私和安全保护。下文介绍如何在 R 语言环境中构建本地 Blast 数据库，其步骤主要包括：1）下载安装 Blast 工具、R 语言和 Rstudio 软件；2）安装 R 语言程序包；3）构建本地 Blast 数据库并检验。为了保证 Blast 本地数据库构建的可重复性和可操作性，该项目提供了 R 语言代码和示例文件（附件 S1）。

1. 下载安装 Blast 工具、R 语言和 Rstudio 软件

1.1 下载 Blast 工具

针对不同操作系统，选择不同的安装方式。Windows 系统，从网址 https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST 下载后安装，注意环境配置；而 Linux (Debian/Ubuntu)系统，通过以下脚本安装：

```
## 下载 blast
conda install -c bioconda blast
conda install -c bioconda/label/cf201901 blast
## 查看版本
blastn -version
```

1.2 下载安装 R 语言和 Rstudio 软件

R 语言是用于统计分析，统计绘图的编程语言和操作环境。为了在 R 语言环境中使用 Blast 工具，需安装 R 语言和 Rstudio 软件，其下载网址分别为：https://cran.r-project.org、https://www.rstudio.com/products/rstudio。

2. 安装 R 语言程序包

安装 R 语言程序包 rBLAST（Hahsler and Nagar, 2019）和所需依赖包 Biostrings（Pagès et al., 2022）。安装方法见 00_pkg_install.R。

```
# 安装所需 R 语言程序包
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("Biostrings")
devtools::install_github("helixcn/seqRFLP")
devtools::install_github("mhahsler/rBLAST")
# 检查 blast 工具是否被安装
Sys.which("blastn")
# 检 blast 版本
system("blastn -version")
```

3. 构建本地 Blast 数据库并检验

3.1 构建本地 Blast 数据库

首先获取建库的目标基因数据，可从 NCBI 下载或者用户收集。该项目收集了大量中国淡水鱼类的 Co1 和

Cytb 基因（图 1 和图 2），通过 makeblastdb 函数分别构建 Co1 和 Cytb 数据库，支持数据格式为*.FASTA。

```
# 建库的目标基因数据的相对路径
reference_co1 = "db/co1.fasta"
# 参数 dbtype 选项’nucl’表示核苷酸基因，’prot’表示蛋白质
makeblastdb(reference_co1, dbtype='nucl')
```

3.2 检验本地 Blast 数据库

为了检验已构建的 Blast 数据库可靠性，通过待比对序列（query）在本地比对。示例见 02_coi_makeblastdb.R 和 03_cytb_makeblastdb.R。

```
# 打开已建立 Blast 数据库
bl <- blast(db = reference_co1, type='blastn')
# 读取待比对序列
query <- readDNAStringSet('example/query_co1.fasta')
# 读取待比对序列在本地 Blast 数据库比对
predict(bl, query)
```

参考文献：

Messing J. New M13 vectors for cloning. Methods in Enzymology, 1983, 101: 20-78.

Ward RD, Zemlak TS, Innes BH, et al. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2005, 360(1462): 1847-1857.

Xiao W, Zhang Y, Liu H. Molecular systematics of Xenocyprinae (Teleostei: Cyprinidae): taxonomy, biogeography, and coevolution of a special group restricted in East Asia. Molecular Phylogenetics and Evolution, 2001, 18(2): 163-173.

郑新. 中国子陵吻虾虎鱼和小黄黝鱼的种群结构及遗传多样性分析[学位论文]. 上海: 上海海洋大学, 2016.

Zhang Z, Schwartz S, Wagner L, et al. A greedy algorithm for aligning DNA sequences. Journal of Computational Biology, 2000, 7(1-2): 203–214.

Hahsler M, Nagar A. rBLAST: R interface for the Basic Local Alignment Search Tool. 2019. R package version 0.99.2. https://github.com/mhahsler/rBLAST.

Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: efficient manipulation of biological strings. 2024. R package version 2.74.1.