|  |  |
| --- | --- |
| ICS  Z01 | 13.020.01 |

团体标准

T/CSES XXXX—XXXX

水生生物DNA条形码构建规程

Regulations for construction of DNA barcodes of hydrobios

XXXX - XX - XX发布

XXXX - XX - XX实施

中国环境科学学会  发布

目次

[前言 II](#_Toc100608008)

[引言 III](#_Toc100608009)

[1 范围 1](#_Toc100608010)

[2 规范性引用文件 1](#_Toc100608011)

[3 术语和定义 1](#_Toc100608012)

[4 DNA条形码构建方法 2](#_Toc100608013)

[4.1 DNA条形码构建流程 2](#_Toc100608014)

[4.2 物种鉴定 3](#_Toc100608015)

[4.3 条形码扩增与测序 3](#_Toc100608016)

[4.4 条形码评估 4](#_Toc100608017)

[4.5 物种条形码 4](#_Toc100608018)

[5 质量控制和质量保证 4](#_Toc100608019)

[6 废弃物处理 5](#_Toc100608028)

[附录A （资料性） 常用DNA条形码引物及PCR扩增条件 6](#_Toc100608029)

[附录B （资料性） 遗传距离分析原理和方法 8](#_Toc100608030)

[附录C （资料性） DNA条形码数据表格 11](#_Toc100608035)

1. 前言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由南京大学提出。

本文件由中国环境科学学会归口。

本文件起草单位：

本文件主要起草人：

1. 引言

DNA条形码技术是利用生物共有的、但种间差异明显的遗传信息脱氧核糖核酸（DNA）序列来实现物种鉴定的技术，具有可标准化和可量化统计分析的特点。构建水生生物DNA条形码有助于记录我国水生生物物种分子遗传学特征，实现物种的快速准确鉴定。

为规范水生生物DNA条形码构建，促进DNA条形码技术在我国水生生物鉴定、生物监测技术体系中的应用推广，制定本文件。

水生生物DNA条形码构建规程

* 1. 范围

本文件规定了水生生物DNA条形码构建方法及质量控制和安全管理要求。

本文件适用于我国淡水生态系统中浮游植物、浮游动物、水生维管植物、大型底栖无脊椎动物和鱼类等生物类群的DNA条形码构建；其他生态系统生物DNA条形码构建可参照执行。

* 1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB 19489 实验室 生物安全通用要求

GB/T 30989 高通量基因测序技术规程

GB/T 34265 Sanger法测序技术指南

GB/T 35537 高通量基因测序结果评价要求

GB/T 37874 核酸提取纯化方法评价通则

HJ 710.7 生物多样性观测技术导则 内陆水域鱼类

HJ 710.8 生物多样性观测技术导则 淡水底栖大型无脊椎动物

HJ 710.12 生物多样性观测技术导则 水生维管植物

HJ 1216 水质 浮游植物的测定 0.1 ml计数框-显微镜计数法

SC/T 9402 淡水浮游生物调查技术规范

LY/T 3191 林木DNA条形码构建技术规程

SN/T 4278 国境口岸医学媒介昆虫DNA条形码鉴定操作规程

SN/T 4835 实验室生物废弃物管理要求

DB 21/T 2777 海洋浮游微藻分离和筛选技术规程

* 1. 术语和定义

GB/T 30989、GB/T 34265和SN/T 4278界定的以及下列术语和定义适用于本文件。

水生生物 hydrobios

全部或部分生活在各种水域中的生物。主要的淡水生物类群包括浮游植物、浮游动物、水生维管植物、大型底栖无脊椎动物和鱼类等。

引物 primer

在DNA复制过程中，结合于模板链上并作为复制延伸的起始位点和/或终止位点的，具有一定长度和顺序的寡核苷酸链。

聚合酶链式反应 polymerase chain reaction；PCR

一种体外酶促合成特异DNA片段的分子技术，由高温变性、低温退火及适温延伸等几步反应组成一个周期，循环进行，使目的DNA得以迅速扩增，具有特异性强、灵敏度高、操作简便省时等特点。

Sanger测序 sanger sequencing

用于基因测序的双脱氧链末端终止法，在链延伸过程中利用荧光标记双脱氧碱基随机阻断，产生以A、T、C、G结束的四组不同长度的核苷酸链，通过读取荧光信号实现对核酸碱基序列信息的读取。

高通量测序 high-throughput sequencing

区别于传统Sanger测序，能够一次并行对大量核酸分子进行平行序列测定的技术。

DNA条形码 DNA barcode

生物体细胞核或者细胞器中一段公认的能够代表该物种的标准的、有足够变异的、易扩增的短DNA序列，可用于实现生物的识别和鉴定。

遗传距离 genetic distance

通过遗传标记如DNA条形码对种群或分类单元间遗传相似性和进化关系的测度。



16S核糖体 DNA 16S ribosomal DNA；16S rDNA

原核生物编码核糖体小亚基16S rRNA的DNA序列。

18S核糖体 DNA 18S ribosomal DNA；18S rDNA

真核生物编码核糖体小亚基18S rRNA的DNA序列。

线粒体12S核糖体DNA mitochondrial 12S ribosomal DNA；Mt 12S rDNA

后生动物线粒体基因组上12S rRNA对应的DNA序列。

线粒体细胞色素c氧化酶I mitochondrial cytochrome c oxidase I；COI

后生动物线粒体基因组上的线粒体细胞色素c氧化酶I对应的DNA序列。

凭证标本 voucher specimen

获取DNA条形码等分子数据的来源标本。可以是标本的一部分组织，也可以是同一批同种标本中的一个个体，但必须与本标本的DNA条形码数据等分子凭证是一一对应的关系。凭证标本必须有唯一性的标本编号和存放信息，便于日后查证。

* 1. DNA条形码构建方法
     1. DNA条形码构建流程

水生生物DNA条形码构建流程包括：物种鉴定、条形码扩增与测序、条形码评估和物种条形码等过程，见图1。



图1 水生生物DNA条形码构建流程

* + 1. 物种鉴定
       1. 浮游植物的定性样品采集、保存与鉴定按照HJ1216的规定执行，浮游植物的分离培养按照DB21/T 2777的规定执行。
       2. 浮游动物的定性样品采集与鉴定按照SC/T 9402的规定执行，加乙醇溶液固定（乙醇浓度≥90%），常温保存。
       3. 水生维管植物的定性样品采集与鉴定按照HJ 710.12的规定执行，野外干冰或-20 ℃保存，实验室-20 ℃保存。
       4. 大型底栖无脊椎动物的定性样品采集与鉴定按照HJ 710.8的规定执行，加乙醇溶液固定（乙醇浓度≥90%），常温保存。
       5. 鱼类样品采集与鉴定按照HJ 710.7的规定执行，野外干冰或-20 ℃保存，实验室-20 ℃保存。
       6. 形态学鉴定建议由具有丰富鉴定经验的专业人员完成，并保留凭证标本。
    2. 条形码扩增与测序
       1. DNA提取和浓度及纯度测定
          1. 选取完整个体或合适部位的组织提取DNA，避免外源污染。植物和动物组织DNA的提取分别参考LY/T 3191和 SN/T 4278。
          2. 按GB/T 37874的规定评价提取的DNA浓度和纯度，一般要求浓度不低于1 ng/μL，在260nm和280 nm波长处的吸光度值比值（OD260 nm/OD280 nm）应在1.7~1.9范围内。有效的DNA样品分装为两份，一份在-80 ℃长期保存，另一份保存在-20 ℃用于后续实验，避免反复冻融。
       2. **条形码扩增与检测**
          1. 针对不同生物类群选择引物扩增目标DNA条形码序列，常用PCR扩增引物和扩增反应条件参照附录A。
          2. PCR扩增产物通过1%~2%琼脂糖凝胶电泳检测扩增的有效性，应在目标长度位置出现一条单一的条带。若样品出现多个条带，需纯化PCR产物。具体过程按照SN/T 4278的规定执行。
       3. **测序**

对目标条带进行Sanger测序，应满足GB/T 34265的规定。为减少Sanger测序出现双峰或者测序失败，可先将PCR产物克隆到质粒中再进行测序。针对大批量实验样品或Sanger测序失败的样品，可按照GB/T 35537的规定进行高通量测序，为每个样本提供不少于100条高质量序列。

* + 1. 条形码评估
       1. 序列质控

将Sanger测序的双端测序文件进行拼接，按照GB/T 34265的规定去除测序结果两端的低质量序列。高通量测序应保留数量最多的序列。测序序列应去除扩增引物，并保证序列方向与PCR扩增正向引物方向一致。

* + - 1. 遗传距离分析

基于DNA条形码的遗传距离分析包括序列准备、序列比对和修剪及遗传距离计算。具体原理和分析过程见附录B。

* + - * 1. 序列准备

将有效的样本序列合并整理成fasta文件。

* + - * 1. 序列比对和修剪

采用ClustalW方法进行多序列比对，并修剪对齐双端序列。

* + - * 1. 遗传距离计算

遗传距离计算基于有差异的核苷酸位点在序列中所占的比例，同时需要考虑四种核苷酸的发生频率和核苷酸替换类型。

* + - 1. 条形码有效性

成功的物种条形码应同时满足以下条件：

1. 阴性对照PCR无条带；
2. 阳性对照的PCR产物应在预期的DNA条形码序列长度位置出现目的条带；
3. 测序获得高质量序列，Sanger测序和高通量测序分别达到QV20和Q20；
4. 种内遗传距离明显小于种间距离（一般5倍~10倍）。
   * 1. 物种条形码

成功的物种条形码信息由样本信息、DNA条形码信息和物种分类信息构成，在DNA条形码数据表中详实记录，见附录C。样本信息包括样本编号、凭证标本信息、采样信息、鉴定信息和3张高清样本图片等。DNA条形码信息包括分子实验记录、扩增引物、测序平台和序列信息。DNA条形码数据可按照SN/T 4279的规定上传至BOLD或NCBI等公共数据库。

* 1. 质量控制和质量保证
     1. 严格按标准要求进行信息采集，填写各项数据记录。记录表格编页装订成册，内容齐全，填写翔实，字迹工整、清晰。原始数据记录表、数码图片、样品和分类凭证标本应及时保存归档，并及时填写和归档电子数据（包括数据记录表和数码图片等）。
     2. 采样完成后，将所有样品运回实验室，与实验室人员交接，填写实验室样品记录表。将样品瓶上的所有信息抄写在实验室样品登记表上，按照采样区域或样点对样品登记表进行统一编号。
     3. 实验场所应具备分子生物学实验室的基本条件。
     4. 为防止外源污染，实验前应将实验用具进行高压蒸汽灭菌，并用75%乙醇擦洗实验桌面。
     5. DNA提取阴性对照（Negative isolation control，NIC），用于监控DNA提取过程中的污染：在DNA提取过程中设置一个空管，和样品同步操作。
     6. PCR扩增阴性对照（Negative amplification control，NAC），用于排除PCR反应混合物制备过程中由于污染造成的假阳性：用无菌水替代DNA样品进行同步PCR扩增。
     7. PCR扩增效率的阳性对照（Positive amplification control，PAC）：添加等体积已知浓度的靶向生物的基因组DNA，同步进行PCR扩增。
     8. 每个物种对应的凭证标本应按照SN/T 4279的规定妥善永久保存。
  2. 废弃物处理

废弃物的分类、收集、存放和集中处理应按照GB 19489和SN/T 4835的规定执行，其中生物废弃物在处理之前应采用高压灭菌、消毒或焚烧等方式灭活。此外，对含核酸染料的废液和废胶单独收集和处理。

2. （资料性）  
   常用DNA条形码引物及PCR扩增条件

表A.1给出了基于Sanger测序的常见条形码引物信息；表A.2给出了基于高通量测序的常见条形码引物信息。

表A.1 基于Sanger法测序的常见条形码引物信息

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 目标群落 | 基因名称 | 引物名称 | 引物序列（5’ - 3’） | 退火温度 |
| 蓝藻门 | 16S rDNA | 16S-27F | AGAGTTTGATCCTGGCTCAG | 55 °C ~62 °C |
| 16S-1492R | RGMAACCTTGTACGACTT |
| 真核藻类 | 18S rDNA | 18S-7F | ACCTGGTTGATCCTGCCAG | 55 °C ~62 °C |
| 18S-1534R | TGATCCTTCYGC AGGTTCAC |
| 水生维管植物 | RbcL | orbcLa-F | ATGTCACCACAAACAGAGACTAAAGCAAGTG | 60 °C ~61 °C |
| orbcLa-R | TCATCYTTGGTAAAATCAAGTCCACCRCG |
| ITS | oITS2-F | GCGAAATGCGATACTTGGTGTGAATTGC | 60 °C ~61 °C |
| oITS2-R | CCTTGTAAGTTTCTTTTCCTCCGCTTATTGATATGC |
| 浮游动物  底栖动物 | COI | LCO1490F | GGTCAACAAATCATAAAGATATTGG | 45 °C ~55 °C |
| HCO2198R | TAAACTTCAGGGTGACCAAAAAATCA |
| 鱼类 | Mt 12S rDNA | MetafishF1 | TCGTGCCAGCCACCGCGGTTA | 60 °C ~62 °C |
| MetafishR12 | AACTNGGTNCGTTGATCGG |
| COI  2对引物结合 | FishF1 | TCAACCAACCACAAAGACATTGGCAC | 50 °C ~60 °C |
| FishF2 | TCGACTAATCATAAAGATATCGGCAC |
| FishR1 | TAGACTTCTGGGTGGCCAAAGAATCA |
| FishR2 | ACTTCAGGGTGACCGAAGAATCAGAA |

表A.2 基于高通量测序的常见条形码引物信息

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 目标群落 | 基因名称 | 引物名称 | 引物序列（5’ - 3’） | 退火温度 |
| 蓝藻门 | 16S\_V3 | 16S\_341F | CCTACGGGRSGCWGCAG | 55 °C ~62 °C |
| 16S\_518R | ATTACCGCGGCTGCTGG |
| 16S\_V4 | 16S\_515F | GTGCCAGCMGCCGCGGTAA | 55 °C ~62 °C |
| 16S\_806R | GGACTACHVGGGTWTCTAAT |
| 16S\_V3\_V4 | 16S-338F | ACTCCTACGGGAGGCAGCAG | 55 °C ~62 °C |
| 16S-806R | GGACTACHVGGGTWTCTAAT |
| 真核藻类 | 18S\_V9 | 18S\_1389F | TCCCTGCCHTTTGTACACAC | 55 °C ~62 °C |
| 18S\_1510R | CCTTCYGCAGGTTCACCTAC |
| 18S\_V4 | TAReuk454FWD1 | CCAGCA(G/C)C(C/T)GCGGTAATTCC | 55 °C ~62 °C |
| TAReukREV3 | ACTTTCGTTCTTGAT(C/T)(A/G)A |
| 水生维管  植物 | 18S\_ V7 | Euka02\_V7F | TTTGTCTGSTTAATTSCG | 45 °C ~55 °C |
| Euka02\_V7R | ACAGACCTGTTATTGC |
| RbcL | orbcL2\_F | YGATGGACTTACNAGTCTTGATCGTTACAAAGG | 60 °C ~62 °C |
| orbcL2\_R | GNCCATAYTTRTTCAATTTATCTCTTTCAACTTGGATNCC |
| 浮游动物  底栖动物 | COI-1 | mlCOIintF | GGWACWGGWTGAACWGTWTAYCCYCC | 45 °C ~55 °C |
| dgHCO2198 | TAAACTTCAGGGTGACCAAAAAATCA |
| 底栖动物 | COI-2 | mlCOIintF | GGWACWGGWTGAACWGTWTAYCCYCC | 45 °C ~55 °C |
| jgHCO2198 | TAIACYTCIGGRTGICCRAARAAYCA |
| 鱼类 | Mt 12S rDNA-1 | MetafishF1 | TCGTGCCAGCCACCGCGGTTA | 60 °C ~63 °C |
| MetafishR1 | ATAGTGGGGTATCTAATCCCAG |
| Mt 12S rDNA-2 | Tele02\_F | AAACTCGTGCCAGCCACC | 55 °C ~60 °C |
| Tele02\_R | GGGTATCTAATCCCAGTTTG |

1. （资料性）  
   遗传距离分析原理和方法
   1. 序列准备

合并整理后fasta文件格式见图B.1。



图B.1 fasta序列格式

* 1. 遗传距离计算模型

遗传距离分析模型包括p-distance模型、Jukes-Cantor模型和K2P模型，一般采用K2P模型。

* + 1. p-distance模型

利用两个同源基因DNA序列间的核苷酸差异数所占的比例来表示序列间的分歧度。

()

式中:

*p*——序列间的分歧度

*nd*——存在差异的核苷酸数目

*n*——总的核苷酸数目

* + 1. Jukes-Cantor模型

本模型假定4种核苷酸之间相互随机替代，即某一核苷酸替代变成其他3种核苷酸的概率是等同的。

()

式中:

*d*——遗传距离

*p*——序列间的分歧度

* + 1. Kimura 2-parameter（K2P）模型

本模型将核苷酸替代分成转换和颠换。其中转换指A、G间替代或G、C间替代。颠换指A和T/C间的替代或G和T/C间的替代。根据实际监测数据，核苷酸的转换替代率约为颠换替代率的2倍，该方法可以更加真实地反映核苷酸序列间的差异。

()

()

()

式中：

*d*——遗传距离

*P*——序列间的转换分歧度

*Q*——序列间的颠换分歧度

*ns*——存在转换的核苷酸数目

*nv*——存在颠换的核苷酸数目

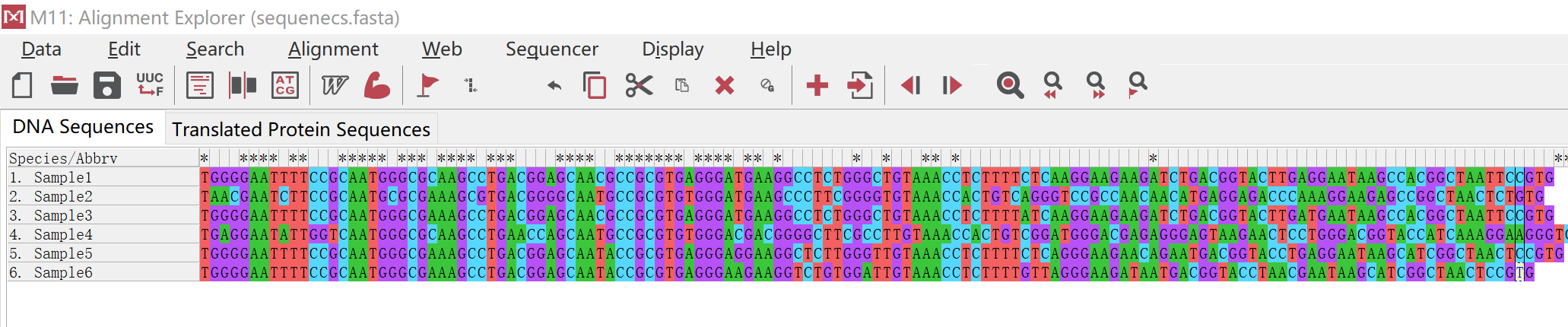
*n*——总的核苷酸数目

* 1. 基于软件MEGA 11的遗传距离分析过程

遗传距离分析通过软件MEGA 11实现。

* + 1. fasta 序列导入

打开MEGA 11，点击File，选择Open A File/Session，上传fasta文件，见图B.2



图B.2 序列导入MEGA 11窗口截图

* + 1. 序列比对和修剪

点击Alignment，选择Alignment by ClustalW（COI条形码序列选择Alignment by ClustalW（Codons）），进行多序列比对。结果见图B.3。一般手动修剪两端序列，保持序列对齐。点击Data，选择Export alignment，将比对结果保存成MEGA 格式文件。



图B.3序列比对后窗口截图

* + 1. 遗传距离计算

导入MEGA文件，点击DISTANCE，选择Compute Pairwise Distances，其中Model/Method选择“Kimura 2-parameter model”计算基于K2P模型的遗传距离。

* 1. 基于R语言的的遗传距离分析过程

遗传距离分析也可通过R语言实现。

* + 1. fasta序列导入

通过“Biostrings”包导入序列：

seq<- readDNAStringSet("sequences.fasta")

* + 1. 序列比对和修剪

通过“msa”包比对序列：

seq.alin<- msa(seq,method ="ClustalW")

通过“ape”包将比对结果（seq.alin）转化成DNAbin格式文件：

seq.dnabin<- as.DNAbin(seq.alin)

修剪双端序列：

seq.dnabin.trim<- trimEnds(seq.dnabin)

* + 1. 遗传距离计算

通过“ape”包计算基于K2P模型的遗传距离：

dist.seq<- dist.dna(seq.dnabin.trim, model = "K80")

* + 1. 结果导出

遗传距离数据导出为 “genetic\_distance.csv”文件：

dist.output<- as.matrix(dist.seq)

write.csv(dist.output,"genetic\_distance.csv",quote = F,row.names = T)

1. （资料性）  
   DNA条形码数据表格

表C.1给出了DNA条形码数据表。

表C.1 DNA条形码数据表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 样本编号 |  | | | 样本照片 |  |
| 凭证标本编号 |  | | | 凭证标本保存地点及保存方式 |  |
| 采样位点 |  | 采样人/单位 |  | 采样时间 |  |
| 经纬度 |  | 海拔 |  | 温度 |  |
| 鉴定人/单位 |  | 鉴定时间 |  | 分类特征 |  |
| 中文名 |  | 拉丁名 |  | 分类信息（门纲目科属） |  |
| 分子实验室 |  | 分子实验操作人 |  | DNA保存地点 |  |
|  | 引物信息 | □16S □18S □RbcL □ITS □COI □Mt 12S □其他： | | | |
| 测序平台 |  | | | |
| 序列长度 |  | | | |
| GC含量 |  | | | |
| 序列 |  | | | |
| 序列原始峰图 |  | | | |

