ICS 13.020.01
CCS Z01

团 体 标 准

T/CSES 80—2023

淡水生物 DNA 条形码构建技术规程

Technical regulations for construction of DNA barcodes of freshwater organisms

发布稿

2023－01－04 发布　　2023－01－04 实施

中国环境科学学会　　发 布

目　　次

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：南京大学、江苏省环境监测中心、中国环境监测总站、浙江省生态环境监测中心、昆明学院、南京易基诺环保科技有限公司。

本文件主要起草人：张丽娟、张效伟、张咏、杨江华、金小伟、俞洁、徐杉、赵峥、杨雅楠、贾世琪、田颖、王志浩、孙晶莹。

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的特点。构建淡水生物DNA条形码有助于记录我国淡水生物物种分子遗传学特征，实现物种的快速准确鉴定。

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

淡水生物 DNA 条形码构建技术规程

1 范围

本文件规定了淡水生物DNA条形码构建方法和质量控制与质量保证。

本文件适用于我国淡水生态系统中浮游植物、着生藻类、水生维管束植物、浮游动物、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。

2 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

LY/T 3191 林木DNA条形码构建技术规程

SC/T 9402 淡水浮游生物调查技术规范

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 4279 国境口岸医学媒介昆虫DNA条形码鉴定实验室管理规范

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

DB 32/T 4178 河流水生态监测规范

3 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。

3.1

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。

[来源：GB/T 30989—2014，3.11]

3.2

聚合酶链式反应 polymerase chain reaction；PCR

一种体外酶促合成特异DNA片段的方法，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。

[来源：LY/T 3191—2020，3.3]

3.3

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。

[来源：GB/T 34265—2017，3.2]

3.4

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。

[来源：GB/T 30989—2014，3.19，有修改]

3.5

DNA条形码 DNA barcode

生物体细胞核或者细胞器中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，可用于生物的识别和鉴定。

[来源：SN/T 4278—2015，3.1，有修改]

3.6

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。

3.7

16S核糖体DNA 16S ribosomal DNA；16S rDNA

原核生物核基因组上编码核糖体小亚基16S rRNA的DNA序列。

3.8

18S核糖体DNA 18S ribosomal DNA；18S rDNA

真核生物核基因组上编码核糖体小亚基18S rRNA的DNA序列。

3.9

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。

3.10

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。

3.11

凭证标本 voucher specimen

获取DNA条形码等分子数据的来源标本。可以是一个标本的一部分组织，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。

[来源：SN/T 4279—2015，3.5，有修改]

3.12

FASTA格式 FASTA format

生物信息学术语，又称Pearson格式，是一种用文本表示核苷酸序列或者氨基酸序列的格式。核苷酸或者氨基酸碱基用单个字母表示，序列的第一行一般用“＞”起始，随后可以添加序列名或者注释，第二行为序列本身，只允许使用既定的核苷酸或者氨基酸编码符号，如核苷酸使用ATGC表示，序列中不能出现回车等字符。

[来源：SN/T 4278—2015，3.9]

注：实例参见附录B。

3.13

阴性对照 negative control

在实验过程中，与受试样品平行进行的，用确定不含DNA的样品代替受试样品进行的对照反应，用于观察整个反应体系是否正确，确认没有受到污染。

[来源：SN/T 4278—2015，3.7，有修改]

3.14

阳性对照 positive control

在实验过程中，与受试样品平行进行的，且预期产生已知阳性结果的样本，用于观察整个反应体系和反应过程是否正常。

[来源：SN/T 4278—2015，3.6，有修改]

4 试剂

4.1 无菌水。
4.2 无水乙醇：分析纯。
4.3 浮游植物培养基。
4.4 DNA 提取试剂。
4.5 通用 PCR 扩增试剂。
4.6 PCR 扩增引物。
4.7 PCR 产物纯化试剂。
4.8 通用 DNA 浓度测定试剂盒：主要成分包括缓冲液和染色剂。
4.9 凝胶电泳相关试剂。
4.10 测序试剂。

5 设备和材料

5.1 样品采集及前处理设备

5.1.1 竖式采水器：2 L 或 5 L。
5.1.2 浮游生物网：孔径 64 μm。
5.1.3 着生藻类采集器具：毛刷、刀片、托盘、洗瓶等。
5.1.4 水生维管束植物采集器具：枝剪、铁夹或铁耙、采集袋等。
5.1.5 大型底栖无脊椎动物采集装置：彼得生采泥器、D 型抄网、踢网、索伯网、40 目的筛网等。
5.1.6 鱼类采集装置：手抄网、捕捉网等捕捉工具、解剖刀、镊子等。

5.2 实验室分析设备

5.2.1 鉴定工具：解剖镜、光学显微镜等。
5.2.2 低温离心机：最大离心力 13000 rpm，温控范围低至 4 ℃。
5.2.3 超微量紫外分光光度计：测量体积最小 1 μL，波长范围包括 230 nm、260 nm 和 280 nm。
5.2.4 PCR 仪。
5.2.5 核酸电泳仪：水平式。
5.2.6 凝胶成像分析仪。
5.2.7 测序仪。
5.2.8 高压蒸汽灭菌仪：可达到 121 ℃、18 min 灭菌条件。
5.2.9 紫外线超净工作台。

5.3 分析软件

5.3.1 Sanger 测序数据分析软件：BioEdit、DNAStar 等。
5.3.2 高通量测序数据质量控制和比对软件：QIIME、R 语言、MEGA 等。

5.4 其他辅助设备和材料

5.4.1 低温冰箱：4 ℃，−20 ℃和−80 ℃。
5.4.2 采样瓶：100 mL 和 1000 mL 等。
5.4.3 照相器具：照相机或摄像机等。
5.4.4 培养皿、载玻片、盖玻片和镊子等。
5.4.5 微量移液器：0.5～10 μL、20～200 μL 和 100～1000 μL 等。
5.4.6 通用无菌手套。
5.4.7 离心管：1.5 mL、2 mL、5 mL 和 50 mL，无 DNA 和生物残留。
5.4.8 PCR 管、96 孔板等 PCR 反应容器：无 DNA 和生物残留。
5.4.9 记录工具：记录纸、防水笔等。

6 DNA 条形码构建方法

6.1 DNA 条形码构建流程

淡水生物DNA条形码构建流程见图1。

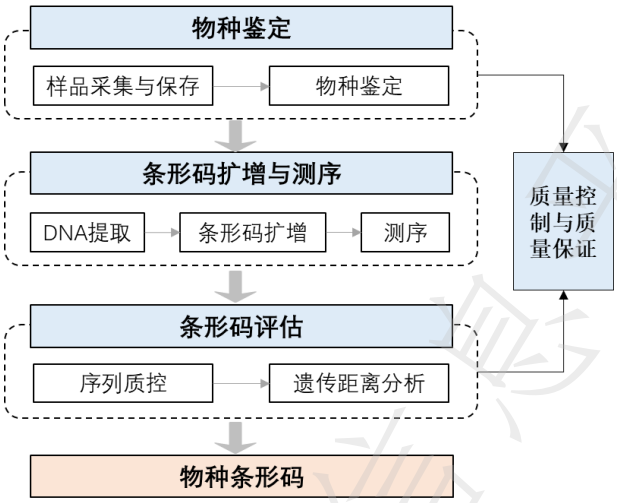

图1 淡水生物 DNA 条形码构建流程

6.2 物种采集与鉴定

6.2.1 浮游植物的采集、保存与鉴定按照 HJ 1216 的规定执行，浮游植物的分离培养参照 DB 21/T 2777 的规定执行。

6.2.2 着生藻类的采集与鉴定按照 DB 32/T 4178 的规定执行，野外冰袋保存，实验室 4 ℃避光条件下保存。

6.2.3 水生维管束植物的采集与鉴定按照 HJ 710.12 的规定执行，按照 LY/T 3191 的规定保存样品。

6.2.4 浮游动物的采集与鉴定按照 SC/T 9402 的规定执行，加乙醇溶液固定（乙醇浓度≥90%），常温保存。

6.2.5 大型底栖无脊椎动物的采集与鉴定按照 HJ 710.8 的规定执行，加乙醇溶液固定（乙醇浓度≥90%），常温保存。

6.2.6 鱼类的采集与鉴定按照 HJ 710.7 的规定执行，野外干冰或-20 ℃保存，实验室-20 ℃保存；或加乙醇溶液固定（乙醇浓度≥90%），常温保存。

6.2.7 浮游动物、大型底栖无脊椎动物、鱼类和水生维管束植物的每个物种个体数一般不少于 5 个（株）；浮游植物和着生藻类的纯培养藻种样品数一般不少于 5 个。

6.2.8 形态学鉴定应由三名具有丰富鉴定经验的专业人员独立完成后汇总，结果不一致的重新鉴定，并保留凭证标本。

6.3 条形码扩增与测序

6.3.1 DNA 提取和浓度及纯度测定

6.3.1.1 选取完整个体或合适部位的组织提取 DNA，避免外源污染。按照 LY/T 3191 和 SN/T 4278 分别进行植物和动物组织 DNA 提取。

6.3.1.2 按照 GB/T 37874 的规定评价提取 DNA 的浓度和纯度，一般要求浓度不低于 1 ng/μL，在 260 nm 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，OD260 nm/OD230 nm 应＞2.0。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。

6.3.2 条形码扩增与检测

6.3.2.1 针对不同生物类群选择一对或多对引物扩增目标 DNA 条形码序列，常用 PCR 扩增引物和扩增反应条件见附录 A。

6.3.2.2　通过 1%～2%琼脂糖凝胶电泳检测 PCR 扩增产物的有效性，应在目标长度位置出现一条单一的条带。若出现多个条带，需重新扩增或纯化 PCR 产物。具体过程按照 SN/T 4278 的规定执行。

6.3.3　测序

对目标条带进行Sanger测序，应满足GB/T 34265的规定。为降低Sanger测序出现双峰或者测序失败的可能性，可按照SN/T 4278的规定先将PCR产物克隆到质粒中再测序。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1000条高质量序列。

6.4　条形码评估

6.4.1　序列质量控制

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。Sanger测序和高通量测序分别达到QV20和Q20。

6.4.2　遗传距离分析

6.4.2.1　序列准备

将有效的样本序列合并整理成FASTA格式的文件。

6.4.2.2　序列比对和修剪

采用ClustalW、Muscle等方法进行多序列核苷酸或氨基酸密码子比对，检查蛋白编码基因是否存在终止密码子，并修剪对齐双端序列。

6.4.2.3　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，具体分析过程见附录B。

6.4.3　条形码有效性

成功的物种条形码应同时满足以下条件：

a)　阴性对照 PCR 无条带；
b)　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；
c)　同一个样本 Sanger 测序获得的序列相似性≥99%；
d)　同一个样本高通量测序获得的第一优势序列占比≥90%；
e)　种间遗传距离显著大于种内距离。

6.5　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。

7　质量控制与质量保证

7.1　质量控制

7.1.1　设置一个空白对照，即没有加入任何生物组织的灭菌 1.5 mL 离心管，和样品同步操作，用于监控 DNA 提取及后续实验过程中的污染。

7.1.2　设置一个 PCR 扩增阴性对照，即用无菌水替代 DNA 样品进行同步 PCR 扩增，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。

7.1.3　设置一个 PCR 扩增的阳性对照，即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。

7.2　质量保证

7.2.1　严格按照本文件要求进行信息采集，填写各项数据记录。记录表格编页装订成册，内容齐全，填写详实，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）。
7.2.2　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记录表上，按照采样区域或点位对样品记录表进行统一编号。
7.2.3　实验场所应具备分子生物学实验室的基本条件。
7.2.4　为防止外源污染，实验前，应将实验耗材（如枪头、离心管和 PCR 管等）进行高压蒸汽灭菌，用漂白剂（次氯酸钠的浓度不低于 1.5%）擦拭桌面。
7.2.5　为防止外源污染，试剂配制宜在紫外线超净工作台中进行。紫外线超净工作台使用前，应紫外线消毒 30 min。
7.2.6　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。

8　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，对含核酸染料的废液和废胶单独收集和处理。

附 录 A
（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'）	预期长度（bp）	退火温度（℃）
原核浮游植物、原核着生藻类	16S rDNA	16S_27F	AGAGTTTGATCCTGGCTCAG	1400～1600	55～62
		16S_1492R	RGMAACCTTGTACGACTT		
真核浮游植物、真核着生藻类	18S rDNA	18S_7F	ACCTGGTTGATCCTGCCAG	1500～2000	55～62
		18S_1534R	TGATCCTTCYGC AGGTTCAC		
水生维管束植物	RbcL	orbcLa_F	ATGTCACCACAAACAGAGACTAAAGCA	500～600	60～63
		orbcLa_R	TCATCYTTGGTAAAATCAAGTCCACCRC		
	ITS	oITS2_F	GCGAAATGCGATACTTGGTGTGAATTGC	300～500	60～63
		oITS2_R	CCTTGTAAGTTTCTTTTCCTCCGCTTATT		
浮游动物、大型底栖无脊椎动物	COI	LCO1490F	GGTCAACAAATCATAAAGATATTGG	～658	45～55
		HCO2198R	TAAACTTCAGGGTGACCAAAAAATCA		
鱼类	Mt 16S-12S rDNA	MetafishF1	TCGTGCCAGCCACCGCGGTTA	～2000	60～62
		MetafishR12	AACTNGGTNCGTTGATCGG		
	COI 2 对引物结合	FishF1	TCAACCAACCACAAAGACATTGGCAC	630～660	50～60
		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'）	预期长度（bp）	退火温度（℃）
原核浮游植物、原核着生藻类	16S_V3	16S_341F	ACCTACGGGRSGCWGCAG	100～200	55～62
		16S_518R	ATTACCGCGGCTGCTGG		
	16S_V4	16S_515F	GTGCCAGCMGCCGCGGTAA	～300	55～62
		16S_806R	GGACTACHVGGGTWTCTAAT		
	16S_V3_V4	16S_338F	ACTCCTACGGGAGGCAGCAG	400～500	55～62
		16S_806R	GGACTACHVGGGTWTCTAAT		
	CPC_IGS	PCβF	GGCTGCTTGTTTACGCGACA	350～400	50～55
		PCβR	GCTTCGGTRAKKGGRGTTTTCAT		
真核浮游植物	18S_V9	18S_1389F	TCCCTGCCHTTTGTACACAC	100～200	55～62
		18S_1510R	CCTTCYGCAGGTTCACCTAC		
	18S_V4-1	TAReuk454FWD1	CCAGCA(G/C)C(C/T)GCGGTAATTCC	300～400	55～62
		TAReukREV3	ACTTTCGTTCTTGAT(C/T)(A/G)A		
真核着生藻类	18S_V4-2	DIV4_F	GCGGTAATTCCAGCTCCAATAG	400～500	48～55
		DIV4_R	CTCTGACAATGGAATACGAATA		
水生维管束植物	18S_V7	Euka02_V7F	TTTGTCTGSTTAATTSCG	100～150	45～55
		Euka02_V7R	ACAGACCTGTTATTGC		
	RbcL	orbcL2_F	YGATGGACTTACNAGTCTTGATCGTTACAAAGG	200～300	60～62
		orbcL2_R	GNCCATAYTTRTTCAATTTATCTCTTTCAACTTGGATNCC		
浮游动物、大型底栖无脊椎动物	COI-1	mlCOIintF	GGWACWGGWTGAACWGTWTAYCCYCC	～313	45～55
		dgHCO2198	TAAACTTCAGGGTGACCAAAAAATCA		
大型底栖无脊椎动物	COI-2	mlCOIintF	GGWACWGGWTGAACWGTWTAYCCYCC	～313	45～55
		jgHCO2198	TAIACYTCIGGRTGICCRAARAAYCA		
鱼类	Mt 12S rDNA-1	MetafishF1	TCGTGCCAGCCACCGCGGTTA	150～200	60～63
		MetafishR1	ATAGTGGGGTATCTAATCCCAG		
	Mt 12S rDNA-2	Teleo_F	ACACCGCCCGTCACTCT	～100	55～60
		Teleo_R	CTTCCGGTACACTTACCATG		
	Mt 12S rDNA-3	Tele02_F	AAACTCGTGCCAGCCACC	150～200	55～60
		Tele02_R	GGGTATCTAATCCCAGTTTG		
	Mt 12S rDNA-4	Mifish_U_F	GTCGGTAAAACTCGTGCCAGC	150～200	55～60
		Mifish_U_R	CATAGTGGGGTATCTAATCCCAGTTTG		
	Mt 16S rDNA	Fish16S_F	GGTCGCCCCAACCRAAG	～100	55～60
		Fish16S_R	CGAGAAGACCCTWTGGAGCTTIAG		

附　录　B
（规范性）
遗传距离分析原理和方法

B.1　序列准备

合并整理后FASTA格式的文件，见图B.1。

```
>Sample1
TGGGGAATTTTCCGCAATGGGCGCAAGCCTGACGGAGCAACGCCGCGTGAGGGATGAAGGCCTCTGGGCTGTAAAC
CTCTTTTCTCAAGGAAGAAGATCTGACGGTACTTGAGGAATAAGCCACGGCTAATTCCGTG
>Sample2
TAACGAATCTTCCGCAATGCGCGAAAGCGTGACGGGGCAATGCCGCGTGTGGGATGAAGCCCTTCGGGGTGTAAAC
CACTGTCAGGGTCCGCCAACAACATGAGGAGACCCAAAGGAAGAGCCGGCTAACTCTGTG
>Sample3
TGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAACGCCGCGTGAGGGATGAAGGCCTCTGGGCTGTAAAC
CTCTTTTATCAAGGAAGAAGATCTGACGGTACTTGATGAATAAGCCACGGCTAATTCCGTG
>Sample4
TGAGGAATATTGGTCAATGGGCGCAAGCCTGAACCAGCAATGCCGCGTGTGGACGACGGGGCTTCGCCTTGTAAA
CCACTGTCGGATGGGACGAGAGGGAGTAAGAACTCCTGGGACGGTACCATCAAAGGAAGGGTCGGCTAACTACGTG
>Sample5
TGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAATACCGCGTGAGGGAGGAAGGCTCTTGGGTTGTAAAC
CTCTTTTCTCAGGGAAGAACAGAATGACGGTACCTGAGGAATAAGCATCGGCTAACTCCGTG
>Sample6
TGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAATACCGCGTGAGGGAAGAAGGTCTGTGGATTGTAAAC
CTCTTTTGTTAGGGAAGATAATGACGGTACCTAACGAATAAGCATCGGCTAACTCCGTG
```

图B.1　FASTA 格式

B.2　遗传距离计算模型

包括p-distance模型、Jukes-Cantor模型和Kimura 2-parameter（K2P）模型，一般采用K2P模型。

B.2.1　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。

$$p = \frac{n_d}{n} \quad \cdots\cdots(B.1)$$

式中：

P ——序列间的分歧度；

n_d ——存在差异的核苷酸数目；

n ——总的核苷酸数目。

B.2.2　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。

$$d_j = \frac{-3\log_e\left(1-\frac{4p}{3}\right)}{4} \quad \cdots\cdots(B.2)$$

式中：

d_j ——遗传距离；

p ——序列间的分歧度。

B.2.3　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T、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。

$$d_k=-\frac{\log_e(1-2P-Q)}{2}-\frac{\log_e(1-2Q)}{4} \quad \text{(B.3)}$$

$$P=\frac{n_s}{n} \quad \text{(B.4)}$$

$$Q=\frac{n_v}{n} \quad \text{(B.5)}$$

式中：

d_k ——遗传距离；

P ——序列间的转换分歧度；

Q ——序列间的颠换分歧度；

n_s ——存在转换的核苷酸数目；

n_v ——存在颠换的核苷酸数目；

n ——总的核苷酸数目。

B.3 基于软件 MEGA 11 的遗传距离分析过程

B.3.1 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

B.3.2 序列比对和修剪

点击Alignment，选择Alignment by ClustalW或Alignment by MUSCLE（COI条形码序列选择Alignment by ClustalW（Codons）或Alignment by MUSCLE（Codons）），进行多序列比对。结果见图B.3。检查蛋白编码基因是否存在终止密码子，并手动修剪两端序列，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 MEGA 11 窗口截图

B.3.3 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。

B.4 基于 R 语言的遗传距离分析过程

B.4.1 FASTA序列导入

通过Biostrings包导入序列：

```
seq<- readDNAStringSet("sequences.fasta")
```

B.4.2 序列比对和修剪

a) 通过 msa 包比对序列：
```
seq.alin<- msa(seq,method ="ClustalW") # 选用 ClustalW 方法比对序列
seq.alin<- msa(seq,method = "Muscle") # 选用 Muscle 方法比对序列
```
b) 通过 ape 包将比对结果转化成 DNAbin 格式文件：
```
seq.dnabin<- as.DNAbin(seq.alin)
```
c) 通过 ips 包修剪双端序列：
```
seq.dnabin.trim<- trimEnds(seq.dnabin)
```

B.4.3 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")
```

B.4.4 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)
```

附 录 C
（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：		保存方式：	保存地点：
采样点位：		采样人/单位：	采样时间：
经纬度：		海拔（m）：	气温（℃）：
鉴定人/单位一：		鉴定时间：	分类特征：
鉴定人/单位二：		鉴定时间：	分类特征：
鉴定人/单位三：		鉴定时间：	分类特征：
门：		Phylum：	纲：
Class：		目：	Order：
科：		Family：	属：
Genus:		种：	Species：
分子实验室：		实验操作人：	DNA 保存地点：
DNA 条形码	引物信息	□16S □18S □RbcL □ITS □COI-1 □COI-2 □Mt 12S □其他：	
	测序平台		
	序列长度		
	GC 含量		
	序列		
	序列原始峰图		

注1：应记录和保留现场调查的照片、影像资料。
注2：分类特征应包括显微镜镜检图片和分类特征的语言描述。
注3：Phylum、Class、Order、Family、Genus、Species分别对应门纲目科属种的拉丁文名称。