

# 鳀科(Engraulidae)鱼类 DNA 条形码电子芯片研究

柳淑芳<sup>1,2</sup> 李献儒<sup>1,4</sup> 李 达<sup>1,5</sup> 庄志猛<sup>1,3①</sup>

(1. 农业部海洋渔业可持续发展重点实验室 中国水产科学研究院黄海水产研究所 青岛 266071; 2. 青岛海洋科学与技术国家实验室 海洋渔业科学与食物产出过程功能实验室 青岛 266071; 3. 青岛海洋科学与技术国家实验室 海洋生物学与生物技术功能实验室 青岛 266071; 4. 大连海洋大学水产与生命学院 大连 116023; 5. 上海海洋大学水产与生命学院 上海 201306)

**摘要** 本研究将基于线粒体 *COI* 部分序列的 DNA 条形码和 DNA 芯片技术相结合,以鳀科 11 属 30 种鱼类为研究对象,在比对分析其 DNA 条形码序列的基础上,利用软件 Oligo Array 2.1 筛选探针,经 OligoCalc 优化探针,去除易形成发夹(Hairpin)、茎环(Stem-loop)及自身二聚体结构(Homodimers)的探针,再利用 Oligo heat map 对探针与靶标序列进行虚拟杂交,共有 14 个物种的 24 条探针能与靶标序列特异性结合。利用 DNA 条形码芯片技术能将 14 个物种鉴定到种,虽然物种识别能力仅占总物种数的 46.7%,但鉴定准确率可达 100%。因此,基于 *COI* 基因的 DNA 芯片技术对鳀科鱼类物种鉴定有一定的实用价值,但该技术对物种的识别能力尚有较大提升空间,通过筛选和优化得到高质量的分子探针则是突破该项技术的关键。

**关键词** 鳀科; DNA 条形码; 探针; DNA 芯片; 电子杂交

**中图分类号** S917 **文献标识码** A **文章编号** 2095-9869(2016)06-0019-07

作为生物多样性保护的重要组成部分,物种的快速分类鉴定至关重要。形态学鉴定的局限性和不断缩减的形态分类学家队伍,使分类学的发展面临巨大的挑战。基于 DNA 条形码技术的分类系统是分类学中辅助物种鉴定的新技术(Austerlitz *et al.*, 2009),其原理为选择不同物种的一段同源 DNA 序列,通过序列比对和聚类分析,将物种鉴定为已知种或发现新物种(Xiao *et al.*, 2004; Moritz *et al.*, 2004)。Hebert 等(2003)通过 DNA 条形码技术成功区分了鳞翅目 200 个物种,认为以 *COI* 序列为基础的 DNA 条形码系统具有可靠、廉价、易推广等特点,并提出利用该技术进行生物分类鉴定的设想。随后,该团队又对 260 种北美鸟类进行了 DNA 条形码分析,发现种间差异约为种内差异的 18 倍,并成功将鸟类鉴定到种(Hebert *et al.*, 2004)。Ward 等(2005)利用 DNA 条形码技术对澳大利亚 207 种海洋鱼类进行分类,发现所有鱼类都能被成

功区分,认为该技术可应用于鱼类的分类鉴定。鉴于 DNA 条形码在物种鉴别上的广泛尝试和应用,2007 年,加拿大正式筹建了国际生命条形码数据库系统(Barcode of Life Data Systems, BOLD)。2008 年,我国成为 iBOL 四个中心节点之一。全球鱼类 DNA 条形码计划(Fish Barcode of Life Initiative, FISH-BOL)始于 2004 年,我国相关单位也积极参与其中,鱼类 DNA 条形码数据库不断得到扩充。本团队也陆续开发了一批近海鱼类、头足类等渔业生物 DNA 条形码(柳淑芳等, 2010; 王鹤等, 2011),并结合形态学分类结果探讨了 DNA 条形码在渔业生物分类鉴定上的可行性。

目前,国内外关于 DNA 条形码的研究仍然集中在积累数据,进而充实数据库。本研究尝试将 DNA 条形码与 DNA 芯片结合起来,以期利用 DNA 芯片的高通量、重复性好、操作性强等优点弥补 DNA 条

\* 山东省科技发展计划项目(2012GHY11531)、科技基础性工作专项(2013FY110700)和泰山学者建设工程专项经费共同资助。柳淑芳, E-mail: liusf@ysfri.ac.cn

① 通讯作者: 庄志猛, 研究员, E-mail: zhuangzm@ysfri.ac.cn

收稿日期: 2015-12-03, 收修改稿日期: 2015-12-30

形码需要测序、繁琐费时的不足。目前,开发了用于多种细菌病原体检测的特异性DNA芯片(Schena *et al.*, 1998; Kozal *et al.*, 1996; Dudda-Subramanya *et al.*, 2003; Santini *et al.*, 1999),但鲜见将该技术应用到渔业生物分类鉴定的研究报道。鳀科(Engraulidae)鱼类隶属于硬骨鱼纲(Osteichthyes)、辐鳍亚纲(Actinopterygii)、鲱形目(Clupeiformes),为世界海洋渔业最主要的类群之一,其总产量在各类鱼类中位居世界第3位。例如,日本鳀(*Engraulis japonicus*)既是日本、朝鲜、俄罗斯和中国在西太平洋远东海洋渔业中的重要捕捞对象,也是海洋掠食者的重要饵料生物。我国的日本鳀资源量曾高达300万t(张世义,2001),是我国海洋渔业资源的重要组成部分(刘瑞玉,2008)。全世界的鳀科鱼类约16属139种(Whitehead *et al.*, 1988),我国分布有7属24种(张世义,2001)。本研究综合运用多种分子生物学软件,在鳀科鱼类物种DNA条形码的基础上筛选特异性分子探针,并探讨其在DNA电子芯片杂交中的有效性。

## 1 材料与方法

### 1.1 实验材料

本研究样本来自中国水产科学研究院黄海水产研究所渔业生物样本库,主要采自我国黄海、东海。参考《中国鱼类系统检索》、《鱼类分类学》(成庆泰等,1987;孟庆闻等,1995)进行形态学分类鉴定。鳀科鱼类样品共计37份,初步鉴定为5种(表1)。参考FishBase (<http://fishbase.org/>)和综合分类学信息系统(Integrated Taxonomic Information System, ITIS)对物种的有效名以及分类地位进行确定。对每个样本拍照记录,取背部肌肉组织用于DNA提取。

### 1.2 DNA提取、PCR扩增及测序

采用酚-氯仿法提取DNA(Maniatis *et al.*, 1985), $-20^{\circ}\text{C}$ 保存备用。PCR扩增采用DNA条形码通用引物(Ward *et al.*, 2005),由华大基因公司合成。引物序列为:F1(TCA ACC AAC CAC AAA GAC ATT GGC AC);F2(TCG ACT AAT CAT AAA GAT ATC GGC AC);R1(TAG ACT TCT GGG TGG CCA AAG AAT CA);R2(ACT TCA GGG TGA CCG AAG AAT CAG AA)。PCR反应总体积为25  $\mu\text{l}$ , $10\times$ 缓冲液成分为100 mmol/L Tris-HCl(pH=8.0)、500 mmol/L KCl、15 mmol/L  $\text{MgCl}_2$ 、0.1%明胶,dNTPs终浓度为200  $\mu\text{mol/L}$ ,各引物终浓度为1.0  $\mu\text{mol/L}$ ,Taq聚合酶0.125 U,总DNA量为50 ng。采用升式PCR,反应程

序为: $94^{\circ}\text{C}$  3 min; $94^{\circ}\text{C}$  30 s, $53^{\circ}\text{C}$  30 s, $72^{\circ}\text{C}$  40 s,8个循环; $94^{\circ}\text{C}$  30 s, $54^{\circ}\text{C}$  30 s, $72^{\circ}\text{C}$  40 s,8个循环; $94^{\circ}\text{C}$  30 s, $55^{\circ}\text{C}$  30 s, $72^{\circ}\text{C}$  40 s,16个循环; $72^{\circ}\text{C}$  10 min。每次PCR反应设立不含DNA模板的空白对照。

阳性PCR产物送至华大基因公司采用ABI PRISM<sup>TM</sup>3730XL DNA Analyzer DNA测序仪进行序列测定,用PCR扩增引物作为测序引物进行双向测序。

### 1.3 公共序列信息下载与筛选

从GenBank下载与本研究扩增区段一致的鳀科鱼类COI基因序列,通过BOLD数据库对下载序列进行准确性验证,同时在Fishbase中检验种名的有效性,最终从公共数据库中筛选出11属25种鳀科鱼类的64条序列,与本研究获取的5种鳀科鱼类37条序列合并为11属30种101条COI基因序列。相关信息见表1。

### 1.4 序列分析

**1.4.1 建立本地Blast数据库** 鳀科所有COI基因序列整合成序列文件seq.fasta,建立本地Blast数据库,将序列文件seq.fasta格式化,输出文件seq.fasta.nsq,用于搜索靶标序列和Blast数据库中所有序列的相似性,评估探针的特异性。

**1.4.2 筛选探针** 鳀科的11属30种101条序列为数据集1。利用软件Oligo Array 2.1为数据集分别筛选探针,输入序列为鳀科的序列文件及格式化文件。设置参数为:探针长度23-27 bp, $T_m$ 值 $70-82^{\circ}\text{C}$ ,GC含量40%-55%, $\text{Na}^+$ 和DNA浓度分别为1 mol/L和1  $\mu\text{mol/L}$ ,形成二级结构的最大 $T_m$ 值和交叉杂交最小 $T_m$ 值为 $65^{\circ}\text{C}$ ,2个探针间距离为25 bp,最大探针数为5,连续出现的单一碱基数不超过4个(如AAAA)。

**1.4.3 物种特异性探针的筛选** 本研究中使用NCBI-blast-2.2.24-win32.exe、Oligo Array 2.1和Oligo Array Aux 3.8软件筛选属级特异性及物种特异性探针。选取的标准是该探针为该属或物种所特有且与其他物种序列无同源性。所得探针在OligoCalc在线软件(<http://www.simgene.com/OligoCalc>)中优化,排除易形成发夹、茎环及自身二聚体结构的不合格探针。设定探针序列自身不存在大于5个碱基的互补配对,和大于4个碱基的反向互补配对。

**1.4.4 探针与靶标序列虚拟杂交** 各数据集序列利用ClustalX比对切齐。优化后的物种特异性探针与切齐序列在Oligo heat map (OHM)在线软件(<http://bioinfo.unice.fr/software/ohm/>)进行虚拟杂交,查看杂交结果。若探针能且只与其所代表物种的靶标序列

表 1 30 种鳀科鱼类 COI 基本信息  
Tab.1 Information of COI genes of 30 Engraulidae species

物种 Species	数量 Quantity	GenBank 序列号 GenBank Accession No.
欧洲鳀 <i>Engraulis encrasicolus</i>	3	AM911180; AM911181; AM911182
美洲鳀 <i>Engraulis mordax</i>	3	FJ164578; FJ164579; FJ164580
日本鳀 <i>Engraulis japonicus</i>	7	本研究(7)This study
阿根廷鳀 <i>Engraulis anchoita</i>	3	EU074422; EU074423; EU074424
新西兰鳀 <i>Engraulis australis</i>	2	EF609349; HQ167626
秘鲁鳀 <i>Engraulis ringens</i>	2	AM911170; HQ167627
银鳀 <i>Engraulis eurystole</i>	1	FJ918906
赤鼻棱鳀 <i>Thryssa kammalensis</i>	8	本研究(8) This study
长颌棱鳀 <i>Thryssa setirostris</i>	3	EF607597; EF607598; EF607599
汉氏棱鳀 <i>Thryssa hamiltonii</i>	3	EU148568; EU148569; EU148570
黄吻棱鳀 <i>Thryssa vitrirostris</i>	3	JF494691; JF494692; JF494693
马拉巴棱鳀 <i>Thryssa malabarica</i>	3	FJ347882; FJ347883; FJ347884
印度侧带小公鱼 <i>Stolephorus indicus</i>	3	EU595317; EU595318; EU595319
南非侧带小公鱼 <i>Stolephorus holodon</i>	3	JF494598; JF494599; JF494600
大眼小鳀 <i>Anchoa lamprotaenia</i>	3	JQ841067; JQ842365; JQ842366
墨西哥小鳀 <i>Anchoa ischana</i>	4	JQ398435; JQ398436; JQ398437; JQ398438
宽带小鳀 <i>Anchoa hepsetus</i>	3	JQ842002; JQ842003; JQ842004
绿小鳀 <i>Anchoa delicatissima</i>	2	GU440219; EU450645
高体小鳀 <i>Anchoa compressa</i>	2	EU450644; GU440218
浅湾小鳀 <i>Anchoa mitchilli</i>	3	JQ842367; JQ842368; JQ842369
刀鲚 <i>Coilia nasus</i>	7	本研究(7) This study
印度洋鲚 <i>Coilia ramacarti</i>	1	JX260855
太的黄鲫 <i>Setipinna taty</i>	7	本研究(7) This study
恒河黄鲫 <i>Setipinna phasa</i>	3	EU417790; EU417791; FJ459395
黄鲫 <i>Setipinna tenuifilis</i>	8	本研究(8) This study
宽带小公鱼 <i>Anchoviella lepidentostole</i>	3	JQ365220; JQ365221; JQ365222
尖吻半棱鳀 <i>Encrasicholina heteroloba</i>	3	EU392185; EU392186; EU392187
大西洋狼鳀 <i>Lycengraulis grossidens</i>	3	JX124803; EU074448; EU074449
太平洋鲸鳀 <i>Cetengraulis mysticetus</i>	1	JQ398441
大鳞多耙鳀 <i>Anchovia macrolepidota</i>	1	JQ398447

结合, 则此探针可作为代表物种的鉴别探针; 若探针不能与其靶标序列结合, 或者与其靶标序列结合的同时还能与其他物种的非靶标序列结合, 这两种情况下的探针均不能作为物种的鉴别探针。

## 2 结果与分析

### 2.1 鳀科鱼类 DNA 条形码信息

通过对鳀科的 11 属 30 种 101 条 COI 基因序列进行比对, 保留共有序列 615 bp 用于 DNA 条形码分析。结果显示, 30 种鳀科鱼类的 DNA 条形码序列的平均碱基组成为 A(24.2%)、T(28.4%)、C(28.2%)和 G(19.2%), 可见碱基组成表现出明显的偏倚性, A+T

含量(52.6%)高于 G+C 含量(47.4%)。所分析序列核苷酸位点中保守位点 501 个, 变异位点 114 个, 其中, 转换位点 70 个, 颠换位点 44 个。87%的转换和 98%的颠换发生在第 3 密码子位点; 第 2 密码子位点最稳定, 没有转换和颠换。

依据 K2P 模型计算 30 种鳀科鱼类的种内和种间遗传距离。结果显示, 种内平均遗传距离为 0.004, 种间平均遗传距离为 0.146, 种间与种内平均遗传距离比为 37; 仅有日本鳀和新西兰鳀(*Engraulis australis*) 2 个物种的种间/种内遗传距离比为 2.3; 其他物种均符合 Hebert 等(2004)提出的“10×规则”。

采用邻接法对 30 种鳀科鱼类的 COI 条形码序列构建分子系统进化树, 可以看出, 鳀科 30 种鱼类均

聚类成独立的分支,且具有较高支持度,各物种均能得到有效区分,表明 *COI* 基因作为 DNA 条形码用于鳀科物种分类鉴定是可行的。

## 2.2 鳀科物种探针信息及杂交结果

鳀科 30 个物种通过 Oligo Array 2.1 进行第 1 轮筛选,得到 24 个物种的 46 条特异性探针;通过 OligoCalc 检验去除 20 条不合格探针,第 2 轮得到 16 个物种的 26 条探针(表 2)。第 3 轮筛选,利用 Oligo heat map(OHM)将探针与鳀科物种 *COI* 基因序列虚拟电子杂交,杂交结果如图 1 所示, TZ53 为银鳀的探针,但它并没有与银鳀的靶标序列杂交,探针灵敏性不强。新西兰鳀的探针 TZ52 不仅与新西兰鳀的序列杂交,还与新日本鳀、欧洲鳀、银鳀的序列杂交, TZ71 为印度洋鲚的探针,还能与黄吻棱鳀的序列结合。上

述 2 条探针在于其靶标序列结合的同时还能与其他非靶标序列结合,探针特异性不强,故不能作为对应物种的特异性探针。其余 14 个物种的 24 条探针均能且只能与靶标序列特异性结合,均可以作为对应物种的鉴别探针。因此,本研究对 30 种鳀科鱼类的 *COI* 基因探针进行 3 轮筛选,最终获得 24 条特异性探针,这些探针可以准确、特异地检测出 14 种鳀科鱼类。虽然这批探针对鳀科鱼类的检测效率仅为 46.7% (14/30),但对杂交成功的物种的鉴定准确率可达 100%。

## 3 讨论

本研究利用在线软件筛选鳀科鱼类 DNA 条形码探针,经过两轮筛选优化后将探针与鳀科鱼类 *COI*

表 2 鳀科物种 24 条特异性探针编号及序列  
Tab.2 IDs and sequences of 24 probes of Engraulidae species

探针编号 Probe IDs	探针序列 Probe sequences	物种 Species
TZ51	TTTACTTCTCTCGCTCCCGGTTTTA	阿根廷鳀 <i>E. anchoita</i>
TZ52	TGACTTATTCCTCTATACTAGGGG	新西兰鳀 <i>E. australis</i>
TZ53	ACATTCTTATTCTTCCCGGATTCGG	银鳀 <i>E. eurystole</i>
TZ54	TAGCTGGTATTTTCATCTATTCTCGG	赤鼻棱鳀 <i>T. kammalensis</i>
TZ55	CTGATATAGCCTTCCCGAATAAA	
TZ56	ACTGATTAGTACCGCTTATACTAGG	
TZ57	CCATCCTAGGAGCTATCAACTTCAT	汉氏棱鳀 <i>T. hamiltonii</i>
TZ58	GGAGCTATCAACTTCATCACTACAA	马拉巴棱鳀 <i>T. malabarica</i>
TZ59	TACCTCTCATGTTAGGAGCACCTGA	
TZ60	GGATCTCATCTATTCTAGGAGCAAT	长颌棱鳀 <i>T. setirostris</i>
TZ61	TGTTTCGTTTGATCAGTTCTCGTGAC	印度侧带小公鱼 <i>S. indicus</i>
TZ62	TTGGAGCAATCAACTTCATCACAAC	
TZ63	GTTGACCTGACAATCTTCTCCCTCC	
TZ64	GGTAGACCTAACAATCTTCTCTCTT	大眼小鳀 <i>A. lamprotaenia</i>
TZ65	TCCTAGGTGCCATCAATTTTCATCAC	宽带小鳀 <i>A. hepsetus</i>
TZ66	CCACTATCATCAATATGAAACCACC	墨西哥小鳀 <i>A. ischana</i>
TZ67	ACCACTATCATCAACATGAAACCAC	
TZ68	TTACTTCTTTCACTTCTGTCCTAG	浅湾小鳀 <i>A. mitchilli</i>
TZ69	CTTTTCACTTCACCTGGCCGGAATC	刀鲚 <i>C. nasus</i>
TZ70	ACTTAGTCAACCTGGGACACTCCTG	印度洋鲚 <i>C. ramacarti</i>
TZ71	TTCTCCTTTTATTAGCCTCACTGG	
TZ72	CCATGCTTTCGTCATAATCTTCTTC	黄鲫 <i>S. tenuifilis</i>
TZ73	CTTACGATCTTTTCTCTCCACCTAG	
TZ74	CTTATTGCTCGCCTCTTCTGGAGTT	
TZ75	CGGACCGTAACTTAAACACCACCTT	尖吻半棱鳀 <i>E. heteroloba</i>
TZ76	AGTTCTCATCACGGCAGTACTTTTA	
TZ77	CTGGGAGACGACCAAATTTATAACG	

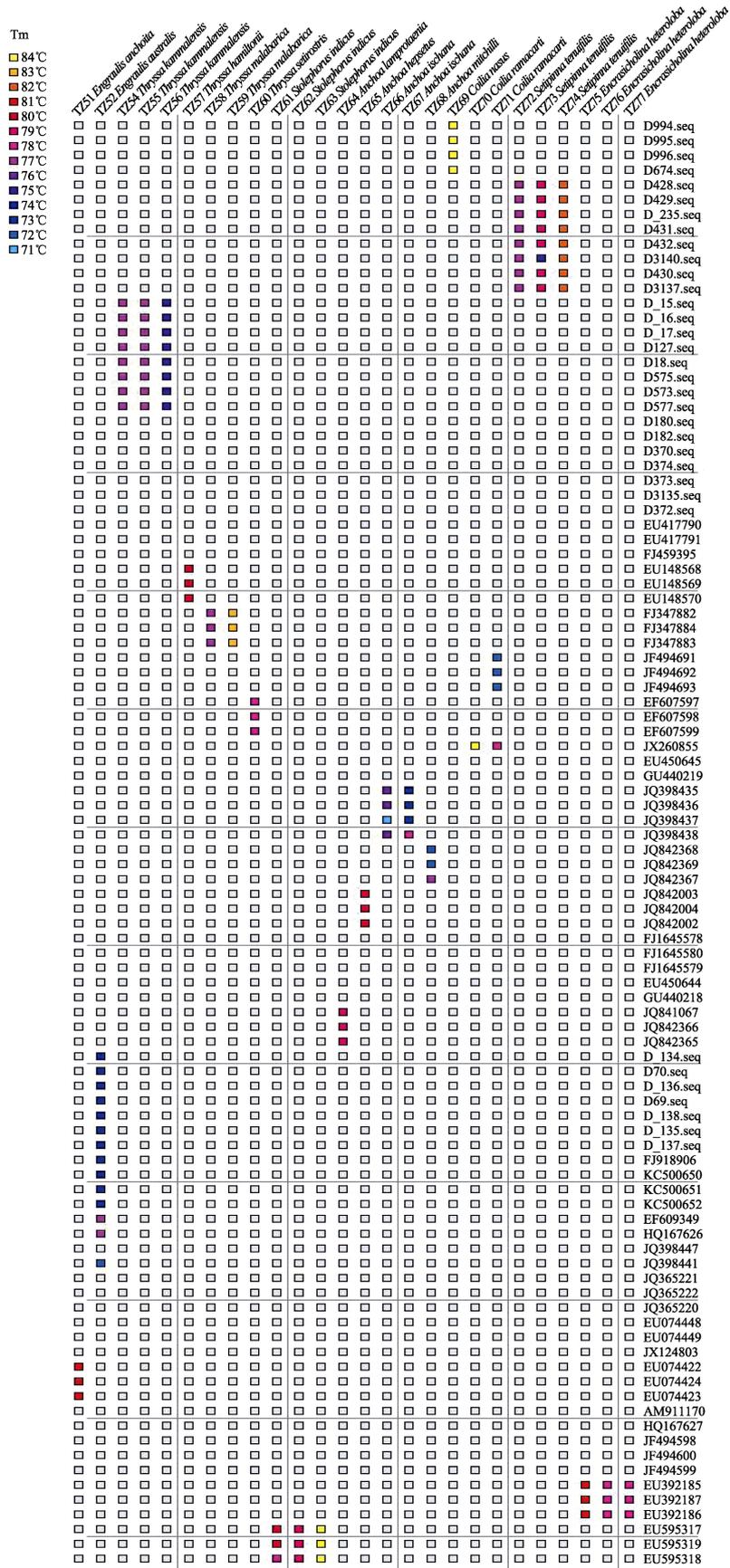


图 1 鯉科鱼类 DNA 条形码电子杂交

Fig.1 Electronic hybridization of DNA barcoding of Engraulidae species

基因序列进行虚拟电子杂交, 整个技术流程与实物 DNA 芯片大致相同, 包括探针的筛选、芯片的杂交和实验结果的统计分析 3 个步骤。作为流程的第 1 步, 探针筛选的质量高低直接影响最终实验结果, 准确的探针是芯片实验成功的必要条件(Hanharan *et al.*, 2003)。探针的筛选主要有三方面的要求: 同一性、灵敏性、特异性。由于探针均为在相同的条件下与靶标序列结合, 所以探针要有相似的特征, 即探针的同一性(包括探针长度、熔解温度、GC 含量等); 探针能与其靶标序列杂交即探针的灵敏性; 探针不能与其他物种的序列杂交即探针的特异性。寡核苷酸芯片探针的筛选需要复杂的生物信息学信息分析, 因此, 易于应用的探针设计软件非常重要(Li *et al.*, 2006)。Oligo Array 2.1 是在基因组水平上为寡核苷酸芯片的建立筛选探针的免费软件(Rouillard *et al.*, 2003)。该软件筛选探针基于三个标准: 一是寡核苷酸探针的熔解温度, 二是探针对靶标序列的特异性, 三是在杂交温度下探针不能形成稳定的二级结构。本研究利用 Oligo Array 2.1 对鳃科 11 属 30 种鱼类筛选特异性探针, 共筛选出 24 个物种的 46 条探针, 占所选物种数的 80.0%。未能筛选出有效探针的物种主要是因其 DNA 序列设计探针的区段与其他物种的相似性较高, 导致与其他序列共享探针, 这种情况下的探针不具有物种特异性。可见, 物种间的同源性高, 序列相似度高, 会影响到探针的筛选。Trewick(2008)通过对新西兰特有属 *Sigauss* 进行条形码分析, 发现两种在形态学上具有明显差异的物种却拥有相同的 mtDNA 单倍型, 这种情况也无法成功筛选出物种特异性探针。

严格的探针筛选会大大降低探针的获取数量, 甚至一些物种无法获得物种特异性探针, 但这样筛选可增加探针的准确性、灵敏性和实用性。筛选出的探针必须能与靶标序列结合, 即探针的灵敏性要强。探针本身如果形成发夹、茎环及自身二聚体等结构, 则会影响到其与靶标序列的结合, 灵敏性降低。OligoCalc 是用来获得单链和双链 DNA 或 RNA 序列性质的在线软件, 这些性质包括熔解温度、序列分子量、GC 含量以及光吸收系数等。本研究利用 OligoCalc 对筛选出的 46 条探针进行进一步筛选, 去除易形成发夹、茎环及自身二聚体的探针, 最后剩余 16 个种的 26 条探针, 占所选物种数的比例下降到 53.3%。

探针筛选是为了检验探针能否与靶标序列结合, 进而达到物种鉴定的目的。OHM 是用来评估和显示寡核苷酸探针与 DNA 序列杂交的在线工具(Croce *et al.*, 2008)。本研究将筛选优化后的 16 个种共 26 条探针与靶标序列在 OHM 中虚拟杂交, 最后有 14 个

物种的 24 条探针可与其靶标序列特异性杂交, 可以作为物种的特异性鉴别探针。经过三轮筛选得到可用 DNA 芯片技术鉴别的物种数占总物种的 46.7%, 虽然检测效率不高, 但探针的特异性强。Kochzius 等(2010)利用 DNA 条形码和芯片技术对 50 种欧洲海洋鱼类进行分类分析, 结果只能鉴定出 10% 的物种。可见, DNA 条形码和芯片技术结合对物种的识别能力尚有较大提升空间。如果在保证探针质量的前提下增加有效探针的数量, 或者根据 DNA 条形码序列在不同区段设计多个探针同时与靶基因杂交, 则有望大大提高 DNA 条形码芯片技术对物种的识别能力。因此, 突破 DNA 条形码电子芯片技术的关键是如何筛选和优化获得高质量的分子探针。

## 参 考 文 献

- 王鹤, 林琳, 柳淑芳, 等. 中国近海习见头足类 DNA 条形码及其分子系统进化. 中国水产科学, 2011, 18(2): 245-255
- 成庆泰, 郑葆珊. 中国鱼类系统检索. 北京: 科学出版社, 1987
- 刘瑞玉. 中国海洋生物名录. 北京: 科学出版社, 2008, 913-916
- 张世义. 中国动物志硬骨鱼纲(鲟形目、海鲢目、鲱形目、鼠鲭目). 北京: 科学出版社, 2001, 119-156
- 孟庆闻, 苏锦祥, 缪学祖. 鱼类分类学. 北京: 中国农业出版社, 1995, 153-172
- 柳淑芳, 陈亮亮, 戴芳群, 等. 基于线粒体 COI 基因的 DNA 条形码在石首鱼科(Sciaenidae)鱼类系统分类中的应用. 海洋与湖沼, 2010, 41(2): 223-232
- Austerlitz F, David O, Schaeffer B, *et al.* DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. BMC Bioinformatics, 2009, 10(S14): 296-300
- Croce O, Chevenet F, Christen R. OligoHeatMap (OHM): an online tool to estimate and display hybridizations of oligonucleotides onto DNA sequences. Nucleic Acids Res, 2008, 36: 154-156
- Dudda-Subramanya R, Lucchese G, Kanduc D, *et al.* Clinical applications of DNA microarray analysis. J Exp Ther Oncol, 2003, 3(6): 297-304
- Hanharan R. The analysis of microarray data. Pharmacogenomics, 2003, 4(4): 477-497
- Hebert PDN, Cywinska A, Ball SL. Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci, 2003, 270(1512): 313-321
- Hebert PDN, Stoeckle MY, Zemlak TS, *et al.* Identification of birds through DNA barcodes. PLoS Biol, 2004, 2(10): 1657-1663
- Kochzius M, Seidel C, Antoniou A, *et al.* Identifying fishes through DNA barcodes and microarrays. PLoS One, 2010, 5(9): 1-15
- Kozal MJ, Shah N, Shen N, *et al.* Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. Nat Med, 1996, 2(7): 753-759
- Li WJ, Ying XM. Mprobe 2.0 computer-aided probe design for oligonucleotide microarray. Appl Bioinformatics, 2006, 5(3):

- 181–186
- Moritz C, Cicero C. DNA barcoding: promise and pitfalls. *PLoS Biol*, 2004, 2(10): 1529–1531
- Maniatis T, Fritsch EF, Sambrook J. Molecular cloning: A laboratory manual. *Acta Biotechnol*, 1985, 5(1): 104
- Rouillard JM, Zuker M, Gulari E. Oligo Array 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res*, 2003, 31(12): 3057–3062
- Santini JT, Cima MJ, Langer R. A controlled-release microchip. *Nature*, 1999, 397(6717): 335–338
- Schena M, Heller RA, Theriault TP, *et al.* Microarray: biotechnology's discovery platform for functional genomics. *Trends Biotechnol*, 1998, 16(7): 301–306
- Trewick SA. DNA barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera: Acrididae). *Cladistics*, 2008, 24(2): 240–254
- Ward RD, Zemlak TS, Innes BH, *et al.* DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci*, 2005, 360(1462): 1847–1857
- Whitehead PJP, Nelsonand GJ, Wongratana T. FAO species catalogue. Vol 7. Clupeoid fishes of the world (Suborder Clupeoidei). An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf-herrings. Part 2-Engraulidae. FAO Fish Synop, 1988, 125(7/2): 305–579
- Xiao JH, Xiao H, Huang DW. DNA barcoding: new approach of biological taxonomy. *Acta Zool Sin*, 2004, 50(5): 852–855

(编辑 冯小花)

## Development of DNA Barcode-Microarray for Identification of Engraulidae Fishes

LIU Shufang<sup>1,2</sup>, LI Xianru<sup>1,4</sup>, LI Da<sup>1,5</sup>, ZHUANG Zhimeng<sup>1,3①</sup>

(1. Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071; 2. Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071; 3. Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266071; 4. College of Fisheries and Life Science, Dalian Ocean University, Dalian 116023; 5. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306)

**Abstract** Species identification plays a key role in the protection of biodiversity. However, there have been great challenges in traditional morphological identification because of its innate limitations and the lack of experienced techniques. Recently thanks to the development of molecular biology, DNA-based identification is becoming a more accurate alternative method. The cytochrome oxidase I (*COI*) barcode is a short piece of DNA from a standard region of the mitochondrial genome, which can be used to identify and classify different species and has been proved to be highly effective in many species including fish. The advantage of *COI* is that it is long enough to identify variations among species yet short enough to be sequenced quickly at low cost. In this study, we analyzed DNA barcoding of Engraulidae species in order to explore the feasibility of DNA microarray technology in species identification. Oligonucleotide probes were designed based on the *COI* sequences of mitochondrial obtained from 101 individuals of 30 fish species in 11 genera. Forty-six specific probes obtained from 24 species were selected with online software Oligo Array 2.1, among which the probes potentially forming hairpin, stem-loop, or homo-dimers were identified with Oligonucleotide Properties Calculator and excluded. The remaining probes were virtually hybridized with the target sequences in Oligo Heat Map (OHM), and subsequently a total of 26 probes from 16 species bound to the target sequences specifically. The DNA microarray technology identified 14 species, accounting for 46.7% of the total with 100% accuracy. Therefore, DNA microarray based on *COI* sequences should have practical values in the identification of Engraulidae fish species. Nevertheless, this technique needs further improvement and the key step would be screening and optimization of high quality molecular probes.

**Key words** Engraulidae; DNA barcode; Probe; DNA Microarray; Electric hybridization

① Corresponding author: ZHUANG Zhimeng, E-mail: zhuangzm@ysfri.ac.cn