

DNA 条形码技术在河北保定、廊坊地区鳞翅目昆虫上的应用及小型区域数据库的构建*

宋韶彬^{1**} 石志勇¹ 金 倩¹ 韩辉林² 刘晓枫¹ 郝梦迪¹ 张爱兵^{1***}

(1. 首都师范大学生命科学学院 北京 100048; 2. 东北林业大学 哈尔滨 150040)

摘 要 【目的】为了探究 DNA 条形码技术和小型区域数据库在蛾类鉴定上的可行性, 本研究利用条形码通用引物扩增了来自河北保定、廊坊地区 10 种夜蛾 82 个样本的线粒体细胞色素 c 氧化酶亚基 I (Mitochondrial cytochrome c oxidase subunit I, CO I) 基因序列。【方法】基于进化树、距离、阈值和特征的方法。【结果】虽然整体分类效果较好, 但基于进化树、距离、阈值的方法都无法将二点委夜蛾 *Athetis lepigone* 进行较好的分类; 样本 LF110802.008 总是被分入标璫夜蛾 *Maliattha signifera* 类群, 与形态学分类结果发生分歧。基于特征的方法运用核基因 28S 进行分析, 结果与形态分类一致。同时还探讨了基于特征方法得到的诊断特征数目与样本数量之间的关系, 发现两者密切相关; 基于特征的方法对小样本量的鉴定也比较有效。本研究建立了小型区域的 DNA 条形码数据库, 使物种识别具有更强的针对性, 有利于提高地区性蛾类病虫害防治效果。【结论】在蛾类鉴定中, DNA 条形码有很好的分类效果, 小型区域数据库很有实际应用价值。

关键词 鳞翅目, 夜蛾, DNA 条形码, 系统发育树, ABGD, JMOTU, 诊断特征, 区域数据库

Species identification of Noctuidae moths (Insecta: Lepidoptera) from Baoding and Langfang, Hebei, China with DNA barcoding and establishment of a local DNA barcoding library

SONG Shao-Bin^{1**} SHI Zhi-Yong¹ JIN Qian¹ HAN Hui-Lin² LIU Xiao-Feng¹

HAO Meng-Di¹ ZHANG Ai-Bing^{1***}

(1. College of Life Sciences, Capital Normal University, Beijing 100048, China;

2. Forestry College, Northeast Forestry University, Harbin 150040, China)

Abstract [Objectives] To explore the feasibility of DNA barcoding and establishing a local DNA barcoding library for the identification of moths, the CO I genes of 82 samples from 10 Noctuidae species in Baoding and Langfang (Hebei Province) were amplified using universal barcoding primers. [Methods] Tree-based, distance-based, threshold-based and character-based methods. [Results] The results show that DNA-based classification was generally consistent with that of traditional methods. However, sample LF110802.008 was misclassified as *Maliattha signifera*, which differs from the morphological results. Using character-based methods to analyze the 28S gene produced results were consistent with the morphological taxonomy. The effect of sample sizes on the number of diagnostic characters was investigated. Character-based methods are relatively effective, even in the case of small sample sizes. We propose the establishment of a local DNA barcoding library in order to make species identification is more effective. [Conclusion] DNA barcoding produces good

* 资助项目: 国家自然科学基金项目 (31272340, 30700641); 北京市人才强教项目 (PHR201107120); 中华人民共和国科学技术部科技基础性工作专项重点项目 (2012FY110800); 高等学校博士学科点专项科研基金 (20101108120002)

**E-mail: ssbssg@sina.com

***通讯作者, E-mail: zhangab2008@mail.cnu.edu.cn

收稿日期: 2013-09-18, 接受日期: 2013-11-08

classification results and a local DNA barcoding library would be useful for moth identification.

Key words Lepidoptera, Noctuide, DNA barcoding, phylogenetic tree, ABGD, JMOTU, diagnostic characters, regional database

夜蛾科是鳞翅目中的一个本科,世界上已知 3.5 万多种 (Poole, 1989), 其中绝大多数种类是害虫。在各类农业害虫中,夜蛾已经逐渐被认为是一类比较突出的害虫,其原因不仅包括为害习性恶劣,而且种类较多,许多生物学特性尚未研究清楚,进行防治措施时不易掌握其规律,因此是昆虫学上值得重视的类群(朱弘复和陈一心, 1963)。物种的分类鉴定是对其科学研究的基础,但由于形态学鉴定不仅依赖于分类学家专业知识和经验,同时也要求标本具有比较齐全的外部形态特征,但在实际过程中很难同时具备上述两方面的条件(程希婷等, 2011)。并且,近些年来分类学家队伍不断缩减,使形态学发展面临巨大的挑战(肖金花等, 2004)。2003 年,加拿大学者 Hebert 提出 DNA 条形码技术,得到了国内外的广泛关注 (Hebert *et al.*, 2003a, 2003b; Remigio and Hebert, 2003)。DNA 条形码技术通过一段标准的基因序列(线粒体细胞色素 C 氧化酶亚基 I, mitochondrial cytochrome c oxidase subunit I, CO I)作为整个物种个体的标记来快速、准确、自动化地对物种进行鉴定 (Hajibabaei *et al.*, 2006)。DNA 条形码自问世以来,一直致力于建立标准、快捷、低成本的物种鉴定方法,大大提高了物种鉴定的效率和准确性 (Hebert *et al.*, 2003b; Smith *et al.*, 2005), 并且在鳞翅目昆虫的分类和鉴定中取得了良好的效果 (李青青等, 2010)。

本实验所选取的采样地点是河北省保定市清西陵(115°19'48.3"E, 39°21'36.3"N)和廊坊城郊(116°41'58.2"E, 39°28'13.4"N)。清西陵属全国重点文物保护单位,周围群峦叠嶂,树茂林密,有着良好的生态环境,离北京约 120 km,鳞翅目昆虫丰富(黄大庄, 2002)。廊坊市位于河北省中部偏东,地处北京、天津两大直辖市之间,离北京市中心约 40 km,两者的独特的地理位置对于北京周边地区害虫防治、生态多样性评价、

生态环境监测等方面有着极其重要的意义。本研究拟达到以下几个目标:1. 初步建立保定、廊坊小型区域的 DNA 条形码数据库,并以网页的形式呈现;2. 比较不同方法识别物种的有效性;3. 进一步探讨了基于特征方法得到的诊断特征数目与样本数量之间的关系。最终为提出快速高效的鳞翅目昆虫鉴定新方法,建立北京及周边地区生态环境监测与保护体系打下基础。

1 材料与方法

1.1 标本采集及形态鉴定

本实验所用的标本来自于 2011 年 7 月 30 日在河北保定清西陵(海拔 140 m, 115.33°E, 39.36°N), 2011 年 8 月 2 日在河北廊坊(海拔 10 m, 116.70°E, 39.47°N)利用高压汞灯诱集,然后低温冷冻致死。共包括鳞翅目夜蛾科 82 个标本(表 1)。标本采集后展翅保存,制作外生殖器玻片,形态学鉴定由东北林业大学韩辉林副研究员完成。标本保存于首都师范大学生命科学院遗传多样性与进化课题组。

1.2 DNA 提取、PCR 扩增及测序

采用样本腿部肌肉组织、利用 Biomed 组织/细胞基因组 DNA 快速提取试剂盒提取基因组 DNA,然后置于-20℃冰箱保存备用。

使用 DNA 条形码引物 LCO1490 (5'-GGTC AACAAATCATAAAGATATTGG-3')和 HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer *et al.*, 1994) 扩增线粒体 CO I 基因 5'端长度为 658 bp 的目的片段,用 25 μL 的 PCR 体系:2×Mastermix 12.5 μL,引物 LCO1490 和 HCO2198 各 0.5 μL (10 μmol/L),模板 DNA 2.5 μL, ddH₂O 9 μL。PCR 反应条件为:94℃预变性 5 min;30 个循环包括 94℃变性 30 s, 50℃复性 30 s, 72℃延伸 30 s;最后 72℃延伸 5 min。PCR 扩增产物经 1%琼脂糖凝胶电泳检测。北京

中科西林生物公司完成测序。

表 1 夜蛾科样本信息及其 CO I 序列的 GenBank 登录序列号

Table 1 Noctuidae samples used in this study and GenBank accession numbers of their CO I gene sequences

样本编号 Specimen no.	物种 Species	采集地点 Location	采集时间 Date	GenBank 登录号 GenBank accession no.
LF110802.045	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666520
LF110802.066	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666525
LF110802.070	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666527
LF110802.081	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666533
LF110802.091	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666536
LF110802.092	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666537
LF110802.094	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666539
LF110802.107	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666541
LF110802.109	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666542
LF110802.113	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666543
LF110802.114	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666544
LF110802.117	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666546
LF110802.125	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666549
LF110802.131	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666551
LF110802.153	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666559
LF110802.155	旋歧夜蛾 <i>Discestra trifolli</i>	廊坊	2011-08-02	KF666506
BD110730.054	旋歧夜蛾 <i>Discestra trifolli</i>	保定	2011-07-30	KF666478
BD110730.055	贯雅夜蛾 <i>Iambia transversa</i>	保定	2011-07-30	KF666479
LF110802.069	贯雅夜蛾 <i>Iambia transversa</i>	廊坊	2011-08-02	KF666526
BD110730.100	贯雅夜蛾 <i>Iambia transversa</i>	保定	2011-07-30	KF666488
BD110730.058	白斑赫夜蛾 <i>Acosmetia biguttula</i>	保定	2011-07-30	KF666480
BD110730.059	白斑赫夜蛾 <i>Acosmetia biguttula</i>	保定	2011-07-30	KF666481
BD110730.094	白斑赫夜蛾 <i>Acosmetia biguttula</i>	保定	2011-07-30	KF666485
BD110730.097	白斑赫夜蛾 <i>Acosmetia biguttula</i>	保定	2011-07-30	KF666486
LF110802.115	白斑赫夜蛾 <i>Acosmetia biguttula</i>	廊坊	2011-08-02	KF666545
BD110730.101	白斑赫夜蛾 <i>Acosmetia biguttula</i>	保定	2011-07-30	KF666489
LF110802.100	白斑赫夜蛾 <i>Acosmetia biguttula</i>	廊坊	2011-08-02	KF666540
LF110802.015	白斑赫夜蛾 <i>Acosmetia biguttula</i>	廊坊	2011-08-02	KF666511
BD110730.060	白线散纹夜蛾 <i>Callopistria albolineola</i>	保定	2011-07-30	KF666482
BD110730.091	白线散纹夜蛾 <i>Callopistria albolineola</i>	保定	2011-07-30	KF666484
BD110730.104	白线散纹夜蛾 <i>Callopistria albolineola</i>	保定	2011-07-30	KF666491
BD110730.106	白线散纹夜蛾 <i>Callopistria albolineola</i>	保定	2011-07-30	KF666492
BD110730.107	白线散纹夜蛾 <i>Callopistria albolineola</i>	保定	2011-07-30	KF666493
BD110730.109	白线散纹夜蛾 <i>Callopistria albolineola</i>	保定	2011-07-30	KF666494

BD110730.112	白线散纹夜蛾 <i>Callopietria albolineola</i>	保定	2011-07-30	KF666495
续表 1(Table 1 continued)				
样本编号 Specimen no.	物种 Species	采集地点 Location	采集时间 Date	GenBank 登录号 GenBank accession no.
BD110730.247	白线散纹夜蛾 <i>Callopietria albolineola</i>	保定	2011-07-30	KF666496
BD110730.081	姬夜蛾 <i>Phyllophila oblitterata</i>	保定	2011-07-30	KF666483
BD110730.098	姬夜蛾 <i>Phyllophila oblitterata</i>	保定	2011-07-30	KF666487
BD110730.102	姬夜蛾 <i>Phyllophila oblitterata</i>	保定	2011-07-30	KF666490
LF110802.008	姬夜蛾 <i>Phyllophila oblitterata</i>	廊坊	2011-08-02	KF666508
LF110802.005	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666507
LF110802.011	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666509
LF110802.013	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666510
LF110802.020	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666512
LF110802.032	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666515
LF110802.033	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666516
LF110802.034	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666517
LF110802.039	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666518
LF110802.041	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666519
LF110802.046	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666521
LF110802.054	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666497
LF110802.055	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666498
LF110802.062	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666523
LF110802.065	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666524
LF110802.072	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666529
LF110802.073	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666530
LF110802.088	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666502
LF110802.093	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666538
LF110802.105	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666504
LF110802.122	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666548
LF110802.126	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666550
LF110802.141	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666553
LF110802.142	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666554
LF110802.144	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666555
LF110802.146	标璫夜蛾 <i>Maliattha signifera</i>	廊坊	2011-08-02	KF666505
LF110802.024	二点委夜蛾 <i>Athetis lepigone</i>	廊坊	2011-08-02	KF666513
LF110802.047	二点委夜蛾 <i>Athetis lepigone</i>	廊坊	2011-08-02	KF666522
LF110802.058	二点委夜蛾 <i>Athetis lepigone</i>	廊坊	2011-08-02	KF666499
LF110802.079	二点委夜蛾 <i>Athetis lepigone</i>	廊坊	2011-08-02	KF666532
LF110802.084	二点委夜蛾 <i>Athetis lepigone</i>	廊坊	2011-08-02	KF666501

LF110802.147	二点委夜蛾 <i>Athetis lepigone</i>	廊坊	2011-08-02	KF666556
续表 1(Table 1 continued)				
样本编号 Specimen no.	物种 Species	采集地点 Location	采集时间 Date	GenBank 登录号 GenBank accession no.
LF110802.071	黑点贪夜蛾 <i>Simplicia rectalis</i>	廊坊	2011-08-02	KF666528
LF110802.101	黑点贪夜蛾 <i>Simplicia rectalis</i>	廊坊	2011-08-02	KF666503
LF110802.030	乏夜蛾 <i>Niphonyx segregata</i>	廊坊	2011-08-02	KF666514
LF110802.063	乏夜蛾 <i>Niphonyx segregata</i>	廊坊	2011-08-02	KF666500
LF110802.087	乏夜蛾 <i>Niphonyx segregata</i>	廊坊	2011-08-02	KF666534
LF110802.090	乏夜蛾 <i>Niphonyx segregata</i>	廊坊	2011-08-02	KF666535
LF110802.118	乏夜蛾 <i>Niphonyx segregata</i>	廊坊	2011-08-02	KF666547
LF110802.139	乏夜蛾 <i>Niphonyx segregata</i>	廊坊	2011-08-02	KF666552
LF110802.152	乏夜蛾 <i>Niphonyx segregata</i>	廊坊	2011-08-02	KF666558
LF110802.076	谐夜蛾 <i>Emmelia trabealis</i>	廊坊	2011-08-02	KF666531
LF110802.148	谐夜蛾 <i>Emmelia trabealis</i>	廊坊	2011-08-02	KF666557

1.3 数据分析

1.3.1 构建系统发育树 使用 Chromas 软件观察序列峰图质量,对每个碱基逐一进行人工校对。如果峰图质量差或遇到套峰而不能准确判断碱基,则重新进行扩增和测序。经校对的序列用 Clustal X (Thompson *et al.*, 1997) 软件进行多重序列比对分析。然后利用 Mega4 (Tamura *et al.*, 2007) 软件以 Kimura 双参数模型 (Kimura-2-parameter, K2P) (Kimura, 1980) 采用邻接法 (Neighbor-joining, NJ) 构建系统发育树,以毛翅目昆虫的 2 条相应序列作为外群。使用 1 000 次重复检测的 Bootstrap 值表示树上各节点的支持率。

1.3.2 利用距离法进行分类 使用 ABGD (Automatic barcode gap discovery) 软件 (Puillandre *et al.*, 2011) 基于遗传距离对样本进行划分,划分在同一组的样本被认定为是一个种。将 82 个夜蛾科样本的 K2P 遗传距离矩阵在线提交到 ABGD 网站 (<http://www.wabi.snv.jussieu.fr/public/abgd/abgdweb.html>), 其中,参数设置均使用缺省值,即种内差异先验值 P (Prior intraspecific divergence) 为 0.001 到 0.1,最小相

对 gap 宽度值 X (Minimum relative gap width) 为 1.5。然后将样本划分结果与其形态鉴定结果比较对照。

1.3.3 利用阈值进行分类 使用了基于序列差异的 DNA 分类的方法,即利用 CO I 序列之间的差异来鉴定可操作分类单元 (MOTU) (Floyd *et al.*, 2002; Blaxter *et al.*, 2005), 序列差异小于指定阈值的样本被划分到一个 MOTU 里,由于不同阈值的选取会产生不同的 MOTU 划分结果,因此选取了多个序列差异阈值进行分析,最后将样本划分结果与其形态鉴定结果比较对照。

1.3.4 利用特征法进行分类 使用 R 语言的 spider 数据包中的 nucDiag 函数,对夜蛾科 82 个样本进行分析,寻找可以作为种级诊断特征的特殊核苷酸位点,并计算成功率;为了进一步探寻诊断特征个数与样本个体数量是否有关,选择样本中数量最多的标璁夜蛾 (25 个个体) 和旋歧夜蛾 (17 个个体) 进行特征法分析。

1.3.5 区域数据库的构建 利用夜蛾科 82 个样本构建小型区域数据库,采用 mysql 软件完成数据库数据管理,然后利用 HMMer 软件,将抽取的数据建成 HMM 数据库,实现数据库查询功

能,最终以网页形式展现 (<http://192.168.102.63/BL/index.php>)

2 结果与分析

2.1 NJ 法聚类结果

以 82 个夜蛾科样本的 DNA 为模板,使用 DNA 条形码引物 LCO1490/HCO2198 进行 PCR 扩增,经序列测定,每个样本均得到一条长度约为 650 bp 的 CO I 序列,扩增成功率为 100%。所得序列用 Clustal X 软件比对后,修剪为长度

为 591 bp 的片段进行分析,序列提交至 GenBank 数据库,检索序列号见表 1。

根据外部形态特征,82 个夜蛾样本最终被鉴定为 10 个属,10 个种。使用邻接法以 82 条夜蛾序列构建的 NJ 系统发育树,聚类分析结果如图 1 所示。结果表明:(1)除二点委夜蛾 *Athetis lepigone* 被分为两支,其余 8 种夜蛾均各自聚成一个单系,与形态学分类结果保持一致,并且各支的支持率均比较高。(2) LF110802.008 样本被分入标瑙夜蛾 *Maliattha signifera* 类群,与形态学分类结果发生分歧。

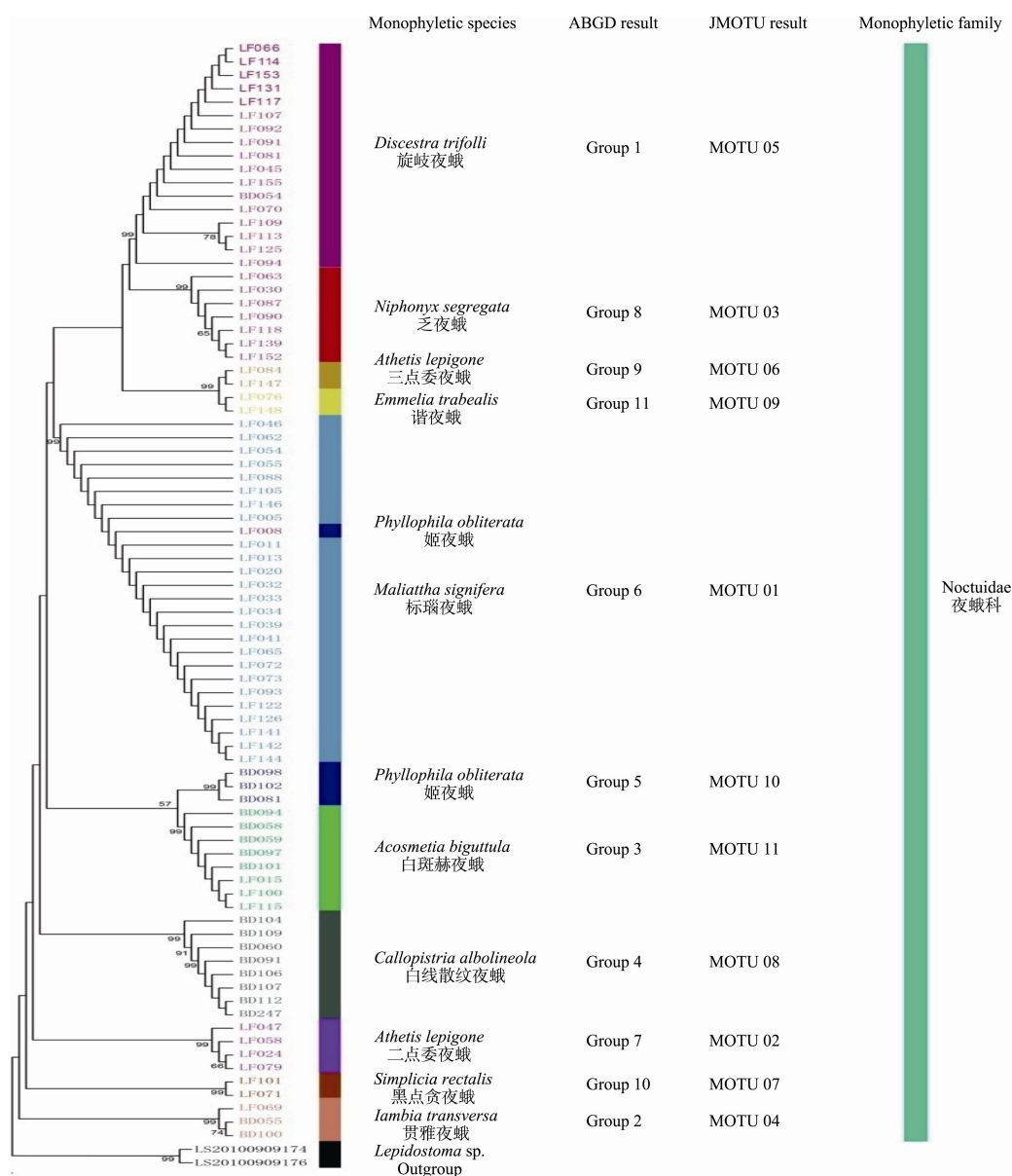


图 1 基于 CO I 序列构建的保定、廊坊地区夜蛾系统发育树

Fig. 1 Phylogenetic tree of the Noctuidae taxa in this study based on CO I gene sequences

2.2 基于距离方法的分类结果

利用 ABGD 软件以 0.001~0.1 的先验值 P 区间对 10 种夜蛾样本进行划分, 结果包含了初始划分 (Initial partition) 和递归划分 (Recursive partition) 两种情况, 如图 2 所示。其中, 初始划分较为稳定, 82 个样本被分成 11 组; 递归划分将样本分成了 1~22 组, 但在一个很大的 P 值范围内维持在 11 组。因此我们选择比较稳定的初始划分结果与形态鉴定结果进行比较发现: (1) 其中 8 种夜蛾分组结果与形态学分类结果保持一致; (2) 二点委夜蛾被分为两组 (Group 7 和 Group 9); (3) LF110802.008 样本被分入标瑙夜蛾类群 (Group 6), 与形态学分类结果发生分歧。

2.3 阈值分类结果

夜蛾科分子序列差异截取值 (Cut-off value) 与可操作分类单元数 (MOTU) 的分析结果如图 3 所示。MOTU 数目在 1%~6.5% 的序列差异内出现了一个平台, 82 条夜蛾序列被分为 11 个

MOTU, 每个 MOTU 分类结果已在系统发育树上标示 (图 2)。结果表明: (1) 其中 8 种夜蛾分组结果与形态学分类结果保持一致; (2) 二点委夜蛾被分为两个 MOTU (MOTU 2 和 MOTU 6); (3) LF110802.008 样本被分入标瑙夜蛾类群 (MOTU 1), 与形态学分类结果发生分歧。

2.4 基于特征方法的分类结果

本研究通过计算 82 个夜蛾样本长为 591 bp 的 CO I 基因序列, 共找到 33 个诊断核苷酸位点。共 10 种夜蛾, 仅有 6 个种找到了各自的诊断核苷酸位点 (表 2), 成功率为 60%。对于构建系统发育树、基于距离方法、阈值分类方法与形态学分类结果出现差异的 LF110802.008 个体, CO I 序列的特征法分析并未在标瑙夜蛾和姬夜蛾类群中找出特征, 对 LF110802.008 样本并未给出明确分类结果。

因此, 我们以 4 个姬夜蛾 (含 LF110802.008) 与 25 个标瑙夜蛾的 DNA 为模板扩增 28S 基因, 每个样本均得到一条长度约为 850 bp 的核基因序列。所得序列用 Clustal X 软件比对后, 修剪

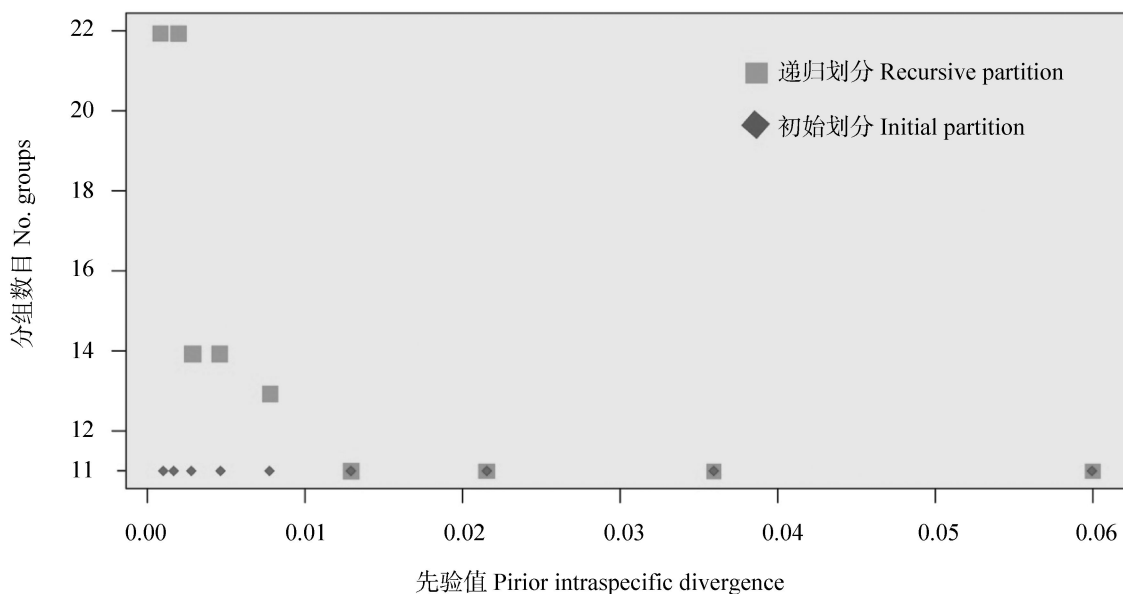


图 2 ABGD 软件对夜蛾样本的划分结果

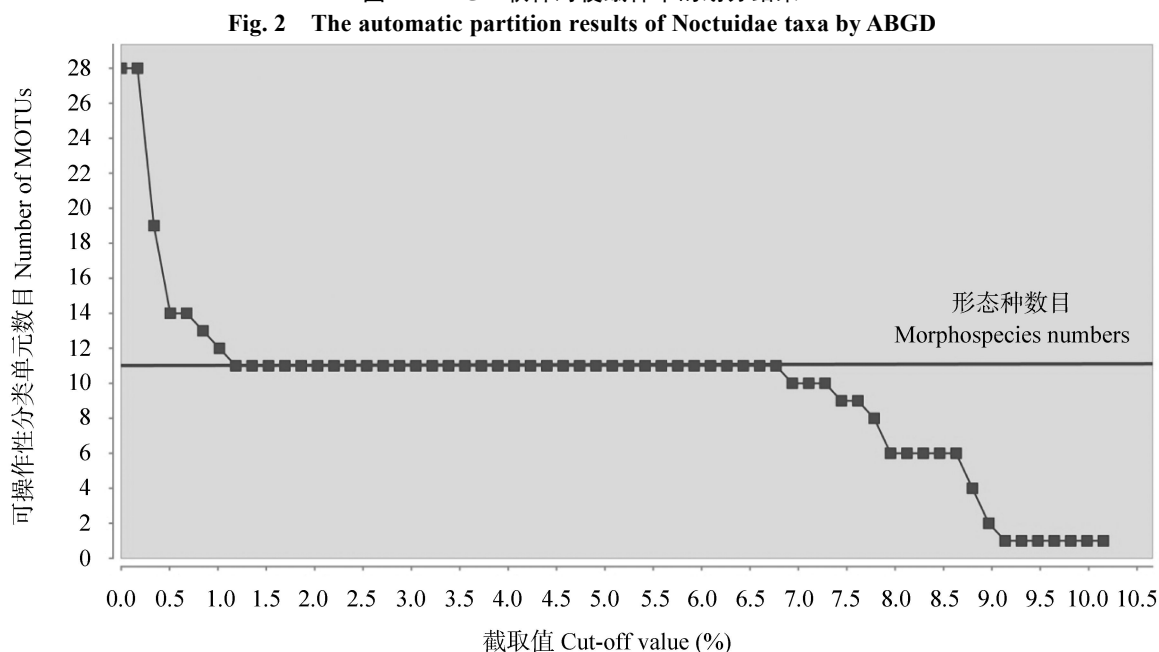


图 3 0~10%的分子序列差异截取值与可操作分类单元数目的关系

Fig. 3 Variation in the number of MOTUs defined at 0 to 10% cut-off values for DNA barcodes

表 2 10 种夜蛾 CO I 序列特征法分析

Table 2 Character-based DNA barcodes for 10 Noctuidae species of CO gene

物种 Species	诊断特征位点 Diagnostic characters	特征数目 Characters no.	样本数目 Species no.
白线散纹夜蛾 <i>Callopistria albolineola</i>	A_38* C_170 G_200 C_251 C_305 C_422 C_425 C_447 T_464 A_515 A_521 C_551	12	8
白斑赫夜蛾 <i>Acosmetia biguttula</i>	C_80 G_221 G_281 A_296 T_335 C_389 C_398 C_521	8	8
谐夜蛾 <i>Emmelia trabealis</i>	C_89 C_225 C_344 A_446	4	2
乏夜蛾 <i>Niphonyx segregata</i>	C_62 T_188 A_225 T_572	4	7
黑点贪夜蛾 <i>Simplicia rectalis</i>	T_164 G_176 C_368 T_512	4	2
贯雅夜蛾 <i>Iambia transversa</i>	A_302	1	3

* A₃₈: 表示第 38 个位点是 A, 下表同。

* A₃₈ means the 38th position is A, the same below.

为长度为 788 bp 的片段进行特征法分析, 通过计算姬夜蛾 (含 LF110802.008) 与标璫夜蛾的 28S 基因序列, 每种均找到 25 核苷酸位点 (表 3)。结果 LF110802.008 个体 28S 基因片段与其余 3 个姬夜蛾样本 (BD110730.081、BD110730.098、BD110730.102) 相同, 与形态学分类结果保持一致。

为了进一步探讨基于特征方法所得到的诊断特征数目是否与样本数量有关, 分别选取了 25 个标璫夜蛾 (*Noctuidae Maliattha signifera*) 与 17 个旋歧夜蛾 (*Noctuidae Discestra trifolli*) 的 CO I 序列作为试验数据集进行分析 (表 4,

为避免 LF110802.008 存在分类差异影响, 整体数据集已将其剔除), 起始在数据集中随机抽取 2 个样本, 得到样本数目为 2 时的诊断特征数目, 此随机过程重复 500 次, 并计算平均诊断特征数目。依此类推, 随机抽取 3,4,5...直到全部样本抽取完毕。得到如下结果: (1) 由于 25 个标璫夜蛾完全共享单倍型, 因此诊断特征数目与样品数量无关, 随着样本量从 2 个增加至 25 个, 诊断特征均为 9 个, 保持不变 (图 4)。(2) 由于 17 个旋歧夜蛾存在种内差异, 因此诊断特征数目与样品数量有关, 随着样本量从 2 个增加至 17 个, 诊断特征数目逐渐下降, 最终降至为 0 (图 5)。

表 3 2 种夜蛾 28S 序列特征法分析
Table 3 Character-based DNA barcodes for 2 Noctuidae species of 28S gene

物种 Species	诊断特征位点 Diagnostic characters	特征数目 Characters no.	样本数目 Species no.
标璫夜蛾 <i>Maliattha signifera</i>	A ₃₃ * C ₅₈ A ₁₁₀ T ₁₅₃ T ₁₈₄ C ₄₂₉ T ₄₃₇ C ₄₄₀ C ₄₅₄ G ₄₆₂ G ₄₆₃ T ₄₆₄ C ₄₆₅ T ₄₈₉ A ₅₅₉ T ₅₇₀ G ₅₇₃ A ₅₇₄ T ₅₇₆ T ₆₁₃ C ₆₁₄ G ₆₃₆ A ₆₃₉ C ₆₈₇ T ₇₁₅	25	25
姬夜蛾 <i>Phyllophila obliterata</i>	G ₃₃ T ₅₈ G ₁₁₀ C ₁₅₃ C ₁₈₄ A ₄₂₉ C ₄₃₇ T ₄₄₀ T ₄₅₄ A ₄₆₂ C ₄₆₃ G ₄₆₄ T ₄₆₅ G ₄₈₉ C ₅₅₉ C ₅₇₀ A ₅₇₃ G ₅₇₄ G ₅₇₆ G ₆₁₃ T ₆₁₄ T ₆₃₆ G ₆₃₉ T ₆₈₇ C ₇₁₅	25	4

* A₃₃: 表示第 33 个位点是 A, 下表同。

* A₃₃ means the 59th position is A, the same below.

表 4 2 种夜蛾 CO I 序列特征法分析
Table 4 Character-based DNA barcodes for 2 Noctuidae species of CO gene

物种 Species	诊断特征位点 Diagnostic characters	特征数目 Characters no.	样本数目 Species no.
<i>Maliattha signifera</i> 标璫夜蛾	T ₄₇ A ₆₈ A ₁₄₇ C ₂₀₆ T ₂₆₃ T ₃₈₀ T ₄₁₆ G ₄₁₉ A ₄₃₀	9	25
<i>Discestra trifolli</i> 旋歧夜蛾		0	17

* T₄₇: 表示第 47 个位点是 T, 下表同。

* T₄₇ means the 47th position is T, the same below.

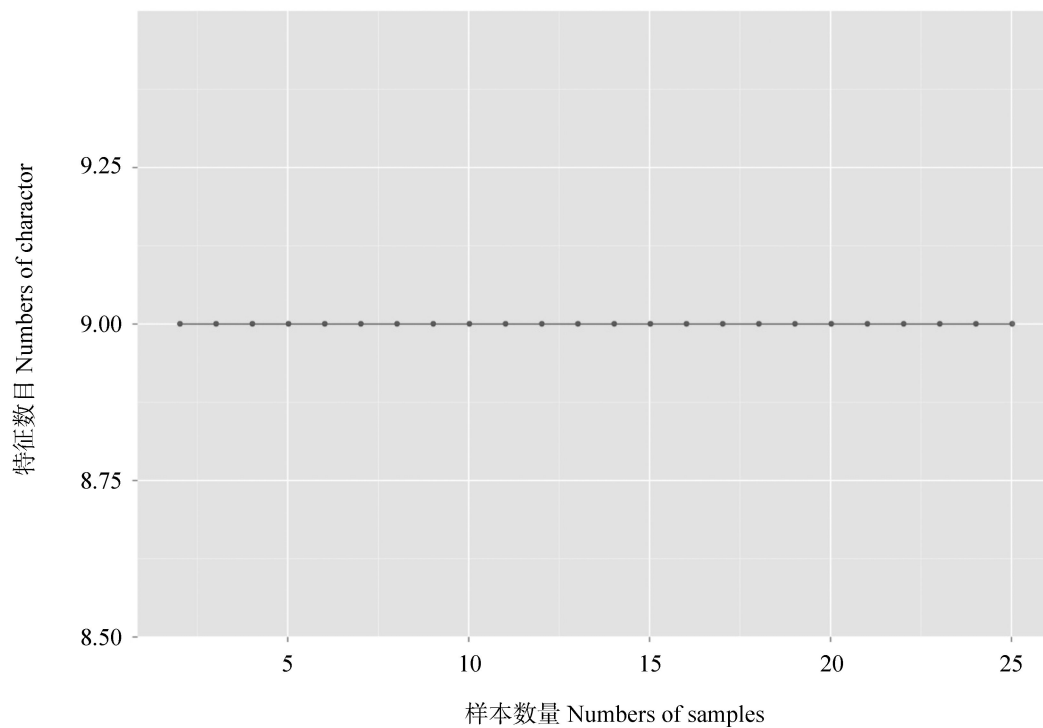


图 4 标璫夜蛾诊断特征数目与样本数量关系

Fig. 4 The relation of character numbers and samples numbers in *Malittha signifera*

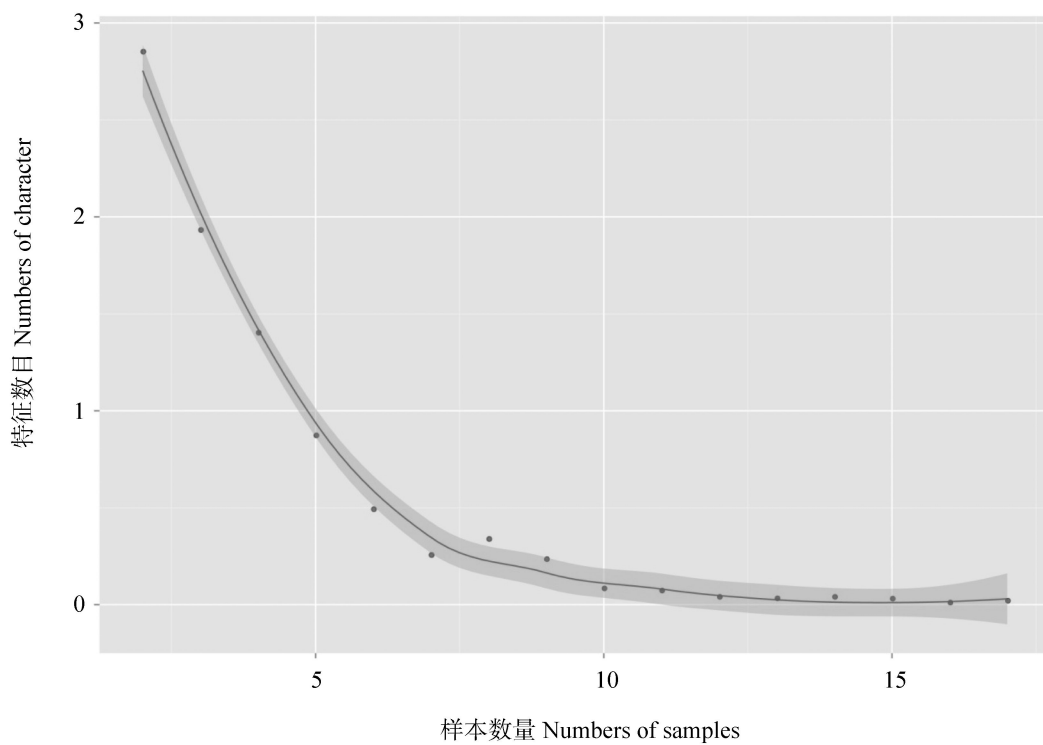


图 5 旋歧夜蛾诊断特征数目与样本数量关系

Fig. 5 The relation of character numbers and samples numbers in *Discestra trifolii*

2.5 区域数据库的构建

2.5.1 构建数据库

2.5.1.1 数据库管理系统采用 mysql 数据库。设计数据各个表的结构,包括字段名称、字段数据类型、字段大小、表主键设定及各表间的依赖关系。

2.5.1.2 采集整理样本数据,包括样本编号、样本 COI 基因分子序列、采样日期、采样地点信息(地点名称、经度、纬度、高度)、样本图片、样本采集人、样本鉴定人。

2.5.1.3 整理样本 COI 基因分子序列,存储为 fasta 格式,每条序列都包含每个分类层级信息。

2.5.1.4 将 fasta 格式数据和采集的样本数据转换为 sql 语句存储到 mysql 数据库中。

2.5.1.5 抽取样本编号和 COI 基因分子序列,利用 HMMer 软件,将抽取的数据建成 HMM 数据库。

2.5.2 数据库使用

2.5.2.1 功能 1: 问询序列为 fasta 格式,使用 HMMer 软件对构建的 HMM 数据库搜索,结果按分值由高到低排序,再根据结果中的样本编号检索 mysql 数据库得到物种信息。

2.5.2.2 功能 2: 设定检索字段(科、属、种、样本编号),输入关键词检索 mysql 数据库,得到物种信息和样本信息(包括样本编号、样本 COI 基因分子序列、采样日期、采样地点信息(地点名称、经度、纬度、高度)、样本图片、样本采集人、样本鉴定人)。

3 讨论

本研究利用系统发育树、基于距离、阈值、特征的方法分别对保定、廊坊地区的夜蛾科蛾类进行了比较分析:整体分类效果较好,但无论是系统发育树、基于距离的 ABGD 方法还是阈值分类的 JMOTU 方法都无法将二点委夜蛾类群进



图 6 区域数据库网站页面
Fig. 6 Regional database web page

行较好的分类,另外 LF110802.008 样本总是被分入标瑙夜蛾类群,与形态学分类结果发生分歧。

针对二点委夜蛾无法较好分类的问题,本研究认为该类群可能存在不完全谱系分选现象(Incomplete lineage sorting)(Floyd *et al.*, 2002; Wiemers and Fiedler, 2007; McFadden *et al.*, 2011)。不完全谱系分选是指在分子水平,对于一些分化时间较短的物种,它们基因序列的最近共祖事件早于物种形成事件,导致构建的基因树不能反映物种真实的进化关系(Van Velzen *et al.*, 2012)。不完全谱系分选会导致种内种间遗传距离出现重叠区域、近缘物种在基因树上不易区分等现象。除此之外,后生动物类群偶尔发生的高速率核苷酸置换也会导致 CO I 条码鉴定中的不可靠分类(Meyer and Paulay, 2005; Vences *et al.*, 2005)。

针对系统发育树、基于距离、阈值的方法对 LF110802.008 样本分类结果与形态学分类结果发生分歧的问题,本研究在对比了 LF110802.008 样本的 591 bp CO I 基因片段发现,与 25 个标瑙夜蛾完全共享单倍型,因此无论采用哪种分子分类的方法,CO I 基因都无法将 LF110802.008 样本进行准确的分类。

为了进一步确定 LF110802.008 样本的分类地位,本研究采用了基于特征的方法,特征法的原理是在条码序列中寻找可以区分不同物种的特殊核苷酸位点,同时忽略同一物种的种内变异,其在果蝇属(Yassin *et al.*, 2010)和蜻蜓目(Rach *et al.*, 2008; Damm *et al.*, 2010)的应用中已取得一定成效。由于基于特征的方法在 CO I 基因水平上并未在标瑙夜蛾和姬夜蛾类群中找出特征,因此无法对 LF110802.008 样本给出明确分类结果。另外由于怀疑 LF110802.008 样本线粒体 CO I 基因发生基因渗透现象,因此改用更为保守的核基因 28S 进行基于特征方法的研究。研究结果表明 LF110802.008 样本在 28S 基因水平上更趋近于姬夜蛾类群,与形态学分类结果一致。另外

本研究还进一步探讨基于特征方法所得到的诊断特征数目与样本数量之间的关系,由于样本数据量较小,因此可能会具有一定的局限性,但基于特征的方法与采样(如采样数量、采样是否均匀等)密切相关,因此诊断特征不是绝对的,而是相对的;但相较于构建系统发育树、基于距离的 ABGD 方法、基于阈值的 JMOTU 方法,基于特征的方法更适于小样本量的快速检测差异,为物种分类提供了快速鉴定的可能(如 LF110802.008 样本在 28S 基因水平上的分类)。

本研究建立鳞翅目蛾类小型区域的 DNA 条形码数据库,着力于在病虫害高发区域建立一套鉴定与防治体系,由于蛾类大多数是农林害虫,区域数据库更具有地理代表性,减少了在鉴定过程中由于地理因素导致的种内变异的影响,因此有利于提高地区性蛾类病虫害防治效果,特别是在病虫害快速鉴定、病虫害常识普及、共享防治方法等方面,具有广阔的前景。

参考文献 (References)

- Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E, 2005. Defining operational taxonomic units using DNA barcode data. *Phil. Trans. R. Soc. B*, 360(1462): 1935–1943.
- Damm S, Schierwater B, Hadrys H, 2010. An integrative approach to species discovery in odonates: from character-based DNA barcoding to ecology. *Mol. Ecol.*, 19(18): 3881–3893.
- Floyd R, Abebe E, Papert A, Blaxter M, 2002. Molecular barcodes for soil nematode identification. *Mol. Ecol.*, 11(4): 839–850.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R, 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.*, 3(5): 294–299.
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs M, Hebert PDN, 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. U S A*, 103: 968–971.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR, 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. B*, 270: 313–322.
- Hebert PDN, Ratnasingham S, DeWaard JR, 2003b. Barcoding animal life: cytochrome c oxidase subunit I divergences among

- closely related species. *Proc. R. Soc. B*, 270(Suppl. 1): S96–S99.
- Kimura M, 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, 16(2): 111–120.
- McFadden CS, Benayahu Y, Pante E, Thoma JN, Nevarez PA, France SC, 2011. Limitations of mitochondrial gene barcoding in Octocorallia. *Mol. Ecol. Resour.*, 11(1): 19–31.
- Meyer CP, Paulay G, 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.*, 3(12): 2229–2238.
- Poole RW, 1989. Noctuidae. Lepidopterorum Catalogues (New Series). Fascicle 118. E.J. Brill, Leiden. Part 1: v–xii, 1–500; part 2: 501–1013; part 3: 1014–1314.
- Puillandre N, Lambert A, Brouillet S, Achaz G, 2011. ABGD, automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.*, 21(8): 1864–1877.
- Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H, 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proc. R. Soc. B*, 275(1632): 237–247.
- Remigio EA, Hebert PDN, 2003. Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships. *Mol. Phylogenet. Evol.*, 29: 641–647.
- Smith, MA, Fisher BL, Hebert PDN, 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Phil. Trans. R. Soc. B*, 360(1462): 1825–1834.
- Tamura K, Dudley J, Nei M, Kumar S, 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, 24(8): 1596–1599.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25(24): 4876–4882.
- Van Velzen R, Weitschek E, Felici G, Bakker FT, 2012. DNA barcoding of recently diverged species: relative performance of matching methods. *PLoS ONE*, 7(1): e30490.
- Vences M, Thomas M, Bonett RM, Vieites DR, 2005. Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Phil. Trans. R. Soc. B*, 360(1462): 1859–1868.
- Wiemers M, Fiedler K, 2007. Does the DNA barcoding gap exist?—a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front. Zool.*, 4(8): 16.
- Yassin A, Markow TA, Narechania A, O’Grady PM, DeSalle R, 2010. The genus *Drosophila* as a model for testing tree- and character-based methods of species identification using DNA barcoding. *Mol. Phylogenet. Evol.*, 57(2): 509–517.
- 程希婷, 王爱民, 顾志峰, 王嫣, 战欣, 石耀华, 2011. DNA 条形码研究进展. *基因组学与应用生物学*, 30(6): 748–758. [Cheng XT, Wang AM, Gu ZF, Wang Y, Zhan X, Shi YH, 2011. Current progress of DNA barcoding. *Genomics and Applied Biology*, 30(6): 748–758.]
- 黄大庄, 2002. 清西陵地区自然资源研究. 东北林业大学出版社 35–36. [Huang DZ, 2002. Studies on the natural resource in Qingxiling. *Northeast Forestry University Press*, 35–36]
- 李青青, 李地艳, 段焰青, 李加敏, 刘晓飞, 曹能, 叶辉, 2010. DNA 条形码在鳞翅目昆虫中的应用. *生命科学*, 22(4): 307–312. [Li QQ, Li DY, Duan YQ, Li JM, Liu XF, Cao N, Ye H, 2010. Application of DNA barcoding in lepidopteran insects. *Chinese Bulletin of Life Sciences*, 22(4): 307–312.]
- 肖金花, 肖晖, 黄大卫, 2004. 生物分类学的新动向——DNA 条形码编码 3. *动物学报*, 50(5): 852–855. [Xiao JH, Xiao H, Huang DW, 2004. DNA barcoding: new approach of biological taxonomy. *Acta Zoologica Sinica*, 50(5): 852–855.]
- 朱弘复, 陈一心, 1963. 中国经济昆虫志, 第三册, 鳞翅目: 夜蛾科. 北京: 科学出版社. 15–17. [Zhu HF, Chen YX, 1963. Economic Insect Fauna China, the third: *Lepidoptera Noctuidae*