昆虫基因与进化研究专栏

基于 DNA 条形码的物种界定算法比较研究——以北京周边地区舟蛾科为例*

金 倩[1, 2**] 武春生[3] 陈 芬[1] 罗桂杰[1] 蔡卫佳[1]
刘 旭[1] 杨采青[2] 郝梦迪[2] 张爱兵[2***]

（1. 江苏省农业科学院宿迁农科所，宿迁 223800；2. 首都师范大学生命科学学院，北京 100048；
3. 中国科学院动物研究所动物进化与系统学重点实验室，北京 100101）

摘 要 **【目的】** 为了探究 3 种常用物种界定方法（jMOTU、ABGD、GMYC）的界定效果。**【方法】**本研究以中国北京周边地区 10 个采样点 483 个舟蛾科样品为例，利用线粒体细胞色素 C 氧化酶Ⅰ亚基基因（Cytochrome c oxidase subunitⅠ gene，COⅠ或 COX1，约 650 bp）的部分序列，进行 3 种物种界定算法（jMOTU、ABGD、GMYC）的实例比较研究。**【结果】**3 种物种界定方法的鉴定效力存在差异，与形态学结果相比较，ABGD 方法划分物种的准确率为 100%，基于 BEAST 的 GMYC 模型结果与形态学结果一致，产生的置信区间（64~68）覆盖了形态学的结果（67）。然而，基于 d8tree/MPLtree 的 GMYC 方法倾向于高估 MOTUs，jMOTU 方法倾向于低估物种数目。**【结论】** 结果显示，ABGD 方法和基于 BEAST 的 GMYC 模型方法对于本文研究对象舟蛾科能够较好地划分，可以对基于形态学的物种界定进行有效补充。

关键词 舟蛾科，DNA 条形码，物种界定，GMYC，ABGD，jMOTU

The reliability of DNA barcoding as a means of identifying Notodontid moth species (Insecta: Lepidoptera, Notodontidae) in Beijing, China

JIN Qian[1, 2**] WU Chun-Sheng[3] CHEN Fen[1] LUO Gui-Jie[1] CAI Wei-Jia[1]
LIU Xu[1] YANG Cai-Qing[2] HAO Meng-Di[2] ZHANG Ai-Bing[2***]

(1. Suqian Institute of Agricultural Sciences, Jiangsu Academy of Agricultural Sciences, Suqian 223800, China;
2. College of Life Sciences, Capital Normal University, Beijing 100048, China; 3. Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China)

Abstract **[Objectives]** To evaluate the performance of three species identification algorithms (jMOTU, ABGD, GMYC) with respect to the Notodontidae (Insecta, Lepidoptera). [**Methods**] COⅠ genes from 483 notodontid moth specimens collected from 10 collection sites in Northern China around Beijing were amplified using universal barcoding primers. [**Results**] Results obtained using each of the three algorithims differed. Using the results obtained by morphological species identification methods as a reference, the ABGD algorithm was 100% accurate. The results obtained using GMYC_BEAST were also generally consistent with those obtained by morphological methods, having a confidence interval of 64-68 c.f. 67 species identified by morphological methods. However, the GMYC_d8tree/MPLtree method tended to overestimate MOTUs and the jMOTU method tended to underestimate the number of species. [**Conclusion**] ABGD and GMYC_BEAST can

*资助项目 Supported projects：国家自然科学基金青年科学基金项目（31601877）；国家杰出青年基金项目（31425023）；国家自然科学基金项目（31272340，31471959）；宿迁市自主创新项目（SQCX2015-06）

**第一作者 First author，E-mail：jinhongyu2001@ 163.com

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

收稿日期 Received：2016-05-23，接受日期 Accepted：2016-07-04

reliably identify species within Notodontidae on the basis of CO Ⅰ gene variation.

Key words Notodontidae, DNA barcoding, species delineation, GMYC, ABGD, jMOTU

现有基于分子的物种界定算法可以理解为无导师（Unsupervised）的学习算法，即仅仅依赖分子本身信息，不需要借助参考数据库、自主地进行物种划分。基于单分子的物种界定算法主要包括基于距离的 jMOTU 方法（Jones *et al.*，2011）和 ABGD 方法（Puillandre *et al.*，2012），以及基于 GMYC 模型的方法（Pons *et al.*，2006）。Jones 等（2011）延伸了 Blaxter 等（1998）的工作，对 DNA 条形码信息进行 MOTUs 划分，并开发 jMOTU 和 Taxonerator 软件。该方法结合了 BLAST 比对的快速性与 NW（Needleman-Wunsch）比对的准确性，将序列归为不同的分子可操作分类单元（MOTUs），达到了高通量与高精确度，可以在普通台式电脑上，短时间内完成成千上万条序列的分析；Puillandre 等（2012）提出 ABGD 的算法（Automatic barcode gap discovery），主要基于 DNA 序列之间距离分布的特征找到 DNA 条形码空白区从而对 DNA 序列进行 MOTUs 划分。该算法首先用先验种内变异范围去推断种内遗传差异的单边置信区间界限，检测到的第一个 DNA 条形码空白区进行 MOTUs 划分，以此类推，递归地进行置信限的推断和 DNA 条形码空白区检测，直到不能划分 MOTUs 为止；Pons 等（2006） 提出 GMYC 模型对物种进行划分，该方法基于最大似然进行模型构建。该模型用于单分子标记的 MOTU 界定，也有学者尝试用于多分子的界定，但效果还不稳定（Boykin *et al.*，2012）。目前该方法需在一个系统发育框架内进行，主要是确定物种形成事件与溯祖事件之间发生变化的转折点，通过此点对物种进行划分。

jMOTU、ABGD、GMYC 等方法相继提出之后，相关学者对不同算法的适用范围等进行了实例和理论模拟的验证（Ratnasingham and Hebert，2013；Dellicour and Flot，2015；Thomsen and Willerslev，2015）。如，Ratnasingham 和 Hebert（2013）将这 3 种算法运用到蜜蜂、鸟类、鱼类、尺蛾、夜蛾、蝴蝶类群中，发现 GMYC 方法准确率最高（89%），其次为 ABGD 算法（85.2%）和 jMOTU 算法（74.3%）。杨聪慧等（2013）运用 ABGD 方法对草螟科类群进行物种界定，初始划分结果与形态鉴定结果进行比较，发现每个分组与形态种均呈现一一对应的关系，成功率为 100%（杨聪慧等，2013）。宋韶彬等（2014）运用 jMOTU 方法和 ABGD 方法对夜蛾科研究发现，虽然整体划分效果较好，但二者都无法将二点委夜蛾 *Athetis lepigone* 进行较好的分类，与形态学分类结果发生分歧（宋韶彬等，2014）。目前这 3 种方法对鳞翅目舟蛾科的实例验证还较少，故本文以舟蛾科为测试对象展开研究。

舟蛾科幼虫大多取食阔叶树树叶，常发生在森林、防护林、行道树和苗圃，如杨扇舟蛾、杨二尾舟蛾、榆掌舟蛾等。部分种类为害果树和竹林，如苹掌舟蛾 *Phalera flavescens*、竹镂舟蛾 *Loudonta dispar* 和竹篦舟蛾 *Besaia goddrica* 等。少数种类为害禾本科农作物，如高粱舟蛾 *Dinara combusta* 和豹舟蛾 *Togarishachia albistriga* 等。本研究以北京周边地区舟蛾科为例，利用线粒体细胞色素 C 氧化酶Ⅰ亚基基因（Cytochrome c oxidase subunit Ⅰ gene，CO Ⅰ或 COX1，约 650 bp）的部分序列，进行基于 DNA 条形码物种界定算法（jMOTU、ABGD、GMYC）的比较研究，不仅为了探索 3 种方法对于舟蛾科物种界定的可靠性，还可以对 DNA 序列信息在物种界定中的应用有一个更加清楚的认知。

1 材料与方法

1.1 标本采集及形态鉴定

本研究样品取自北京周边 10 个采样点（表 1），分别是：喇叭沟门、雾灵山、廊坊、保定、松山、鹫峰、涧沟、陇家庄、东灵山、百花山。为了尽可能多地覆盖舟蛾多样性，本研究选择了相对较长的一段时间采集蛾类，采集时间从 2010

表 1　采样点信息及采集数量

Table 1　Geographical coordinates of sample localities and specimen numbers in this study

采样点 Sampling sites	简称 Short title	纬度（°E） Latitude	经度（°N） Longitude	高程（m） Elevation	样品总数（个） Number of specimen
喇叭沟门 Labagoumen	LBG	40.95	116.50	475	52
雾灵山 Wulingshan	WLS	40.63	117.45	756	219
廊坊 Lanfang	LF	39.26	116.70	10	1
保定 Baoding	BD	39.36	115.33	140	2
百花山 Baihuashan	BHS	39.86	115.56	731	28
东灵山 Donglignshan	DLS	39.95	115.49	784	35
松山 Songshan	SS	40.51	115.79	769	100
鹫峰 Jiufeng	JF	40.07	116.09	147	9
涧沟 Jiangou	JG	40.06	116.03	797	31
陇家庄 Longjiazhuang	LJZ	39.98	116.06	146	6

年 5 月至 2013 年 8 月，共收集得到 483 个样品。样品利用高压汞灯诱集，采集后低温冷冻致死，展翅保存并制作外生殖器玻片，标本存放于首都师范大学遗传进化与多样性课题组。各样点标本收集数量（表 1）：喇叭沟门 52 头，雾灵山 219 头，廊坊 1 头，保定 2 头，百花山 28 头，东灵山 35 头，松山 100 头，鹫峰 9 头，涧沟 31 头，陇家庄 6 头。

1.2　DNA 提取、COⅠ扩增及测序

对于每一个样本，取其足部肌肉组织 3~4 条并储存在 100%的酒精中 4℃保存。用 Biomed 组织/细胞基因组 DNA 快速提取试剂盒提取基因组 DNA，置于 – 20℃冰箱保存备用。接下来进行 COⅠ基因的扩增和测序，具体的步骤为：使用 DNA 条形码引物 LCO1490（5′-GGTCAACA-AATCAT AAAGATATTGG-3′）和 HCO2198（5′-TAAACTTCAGGGTGACCAAAAAATCA-3′）（Vrijenhoek，1994）。扩增线粒体 COⅠ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 2 min；40 个循环包括 94℃变性 20 s，54℃复性 20 s，72℃延伸 45 s；最后 72℃延伸 10 min。PCR 扩增产物经 1%琼脂糖凝胶电泳检测，若结果为 700 bp 左右的亮带，将 PCR 产物送与北京中科西林公司测序，序列分析仪为 ABI 3130 全自动序列测序仪。所有的序列提交至 GenBank 系统中，序列号从 KX271893 至 KX272350。

1.3　数据校对

测序得到的 COⅠ条形码原始序列使用 Chromas 软件观察序列峰图质量，并对每个碱基逐一进行人工校对，确保 DNA 序列的可靠性。经校对的序列用首都师范大学遗传多样性与进化课题组自主开发的 Perl 脚本翻译测试（http://smkxxy.cnu.edu.cn/szll/zrjs/js/4557.htm），如果所得 COⅠ序列能够根据无脊椎动物线粒体密码子表翻译成氨基酸/蛋白质，证明所得序列为目标序列。每条序列在完成上述准备后，整理存放在一个.fasta 格式文件中，运用 BioEdit 7.0.1（http://www.mbio.ncsu.edu/BioEdit/BioEdit.html）软件进行序列比对分析。结果得到无碱基缺失并能够全部翻译通过的序列 483 条。

1.4　数据分析

1.4.1　序列分析　利用课题组自主开发的 Perl 脚本进行单倍型分析，将样本中相同的碱基序列去除并得到共享单倍型列表（http://smkxxy.cnu.edu.cn/szll/zrjs/js/4557.htm）。在 R 语言环境下计算碱基组成频率（A/T/C/G）、GC 含量、保守位

点、变异位点、简约信息位点、转换/颠换比率、核苷酸多样性等序列信息，具体使用函数及包：base.freq，GC.content，seg.sites，pis {phyloch}，nuc.div {pegas}，titv {spider}。

在 DNA 条形码研究中，种内与种间遗传距离的差异是一项非常重要的指标，当序列的种内变异足够小，而种间变异足够大时，会在中间形成一个明显的空白区，即 DNA 条形码空白区（DNA barcoding gap）。DNA 条形码空白区的存在很容易将物种分开，相反，若 DNA 条形码空白区重叠或者边界较为模糊时物种则很难被区分开。为了检验各个样本中种间遗传变异和种内遗传变异是否显著，使用课题组自主开发的 Perl 脚本进行了 DNA 条形码空白区分析（http://smkxxy.cnu.edu.cn/szll/zrjs/js/4557.htm）。

1.4.2 jMOTU 方法 利用 jMOTU 软件对样本进行可操作分类单元划分（Jones *et al.*，2011），序列差异小于指定阈值的样本被划分到一个 MOTU 内。由于不同阈值的选取会产生不同的 MOTU 划分结果，因此我们选取了多个序列差异阈值进行分析。分歧阈值设置为 1~80 bp，聚集参数为 97%，输入最短序列百分比为 95%。

1.4.3 ABGD 方法 利用 ABGD 软件对样本进行物种划分，使用 ABGD 软件（Puillandre *et al.*，2012）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。将样本在线提交至 ABGD 网站（http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html），基于 K2P 模型计算遗传距离，其余参数使用缺省值，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 ABGD 划分结果与形态鉴定结果比较对照。

1.4.4 GMYC 模型 首先将各个样本中重复的单倍型剔除。接下来通过两种方式构建超度量系统发育树（Ultrametric tree）：(1) 通过 PATHd8 软件构建。从 jModelTest2.2.2（Darriba *et al.*，2012）的 88 个核酸替代模型中依据 Akaike 信息准则（AIC）选择各个样本 COⅠ基因合适的进化模型。使用 RAxML7.0.3 构建 ML 树（Stamatakis，2006），利用 PATHd8 软件将枝长标准化获得超度量树（Britton *et al.*，2006，2007）。其中 PATHd8 软件产生两棵超度量树 d8tree/MPLtree。若最终得到的超度量树不是二叉分支，则用 R 语言环境下的函数 multi2di 进行转换。若 ML 树中一些枝长值为 0，在枝长标准化之前需将所有 0 替换为 0.000001；(2) 由 BEASTv1.8.0 直接计算获取（Drummond and Rambaut，2007）。主要参数设置：Yule 物种分化模型、ULRC 宽松分子钟（Drummond *et al.*，2006），2×10^8 次 MCMC 分析、后验分布样本每 10^4 保存一次。最佳碱基替代模型由 jModelTest2.2.2 评估获得（Darriba *et al.*，2012）。分析结果在 Tracer（Rambaut and Drummond，2007）中检验以确保每个参数的有效取样量（Effective sample size）均超过 200。在 BEAST 软件包中的 TreeAnnotator version 中生成最大谱系置信树（Maximum clade credibility tree）。

最后 GMYC 物种界定在 R 环境中使用 Splits{gmyc}函数进行单阈值（Single-threshold）的 GMYC 方法分析（Fujisawa and Barraclough，2013）。通过函数 gmyc 可获得零模型与 GMYC 模型之间的似然比值（LRT），种（Yule 模型）—种群（溯祖模型）的转折点，以及转折点对应的物种数量。

2 结果与分析

2.1 分类信息和序列特征

形态学鉴定该数据集共 43 个属，67 个物种，共获得 COⅠ序列 483 条，序列长度 594 bp，共生成 179 条单倍型，序列的碱基组成分别是：A 占 29.70%，T 占 38.35%，C 占 17.15%，G 占 14.80%，A+T 的含量（68.05%）高于 G+C 的含量（31.95%），在 594 个位点中，保守位点（Conservative sites）307 个，变异位点（Variable sites）287 个，简约信息位点（Parsimony information sites）235 个。核苷酸多样性为 11.99%。转换/颠换比率为 90.82%。种内遗传距离为 0~0.1594，平均为 0.0063；种间遗传距离为 0.05~0.2419，平均值为 0.1409，种间遗传距离分布与种内遗传距离分布之间存

在重叠。如图 1 所示，种内和种间没有严格定义的 DNA 条形码空白区。

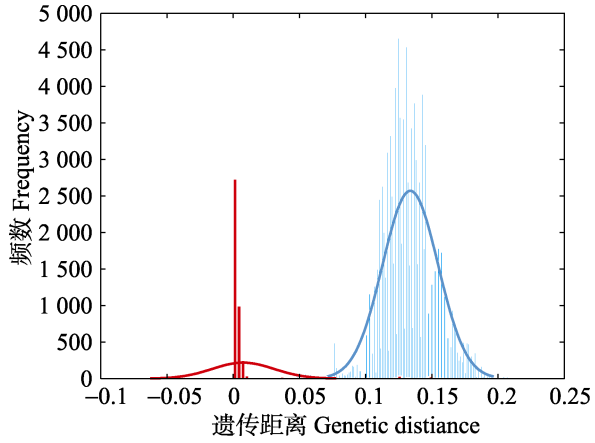

图 1　DNA 条形码空白区
Fig. 1　DNA barcoding gap

2.2　jMOTU 划分 MOTUs 结果

舟蛾科分子序列差异截取值（Cut-off value）与可操作分类单元数的分析结果如图 2 所示。MOTU 数目没有出现较为明显的富集平台，在 7~36 bp (1.18%~6.06%) 的序列差异内变化缓慢，划分 MOTUs 范围为 51~68。结果表明，与形态学分类结果相比，jMOTU 方法较为保守，低估了物种划分数目。

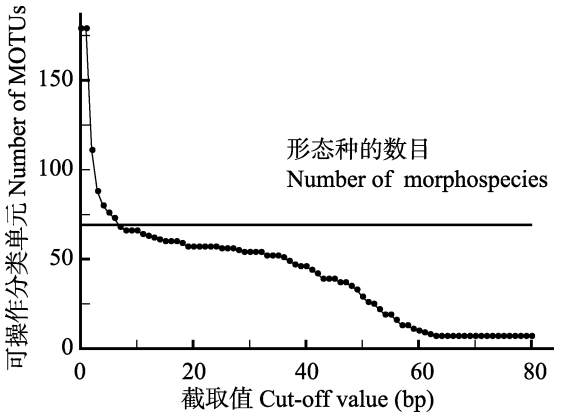

图 2　分子序列差异截取值与可操作分类单元数目的关系
Fig. 2　The relationship between cut-off value and the number of MOTUs

2.3　ABGD 划分 MOTUs 结果

利用 ABGD 软件以 0.001~0.1 的先验值 P 区间对 483 个舟蛾科样本进行划分，结果包含了初始划分（Initial partition）和递归划分（Recursive partition）两种情况，如图 3 所示。其中，初始

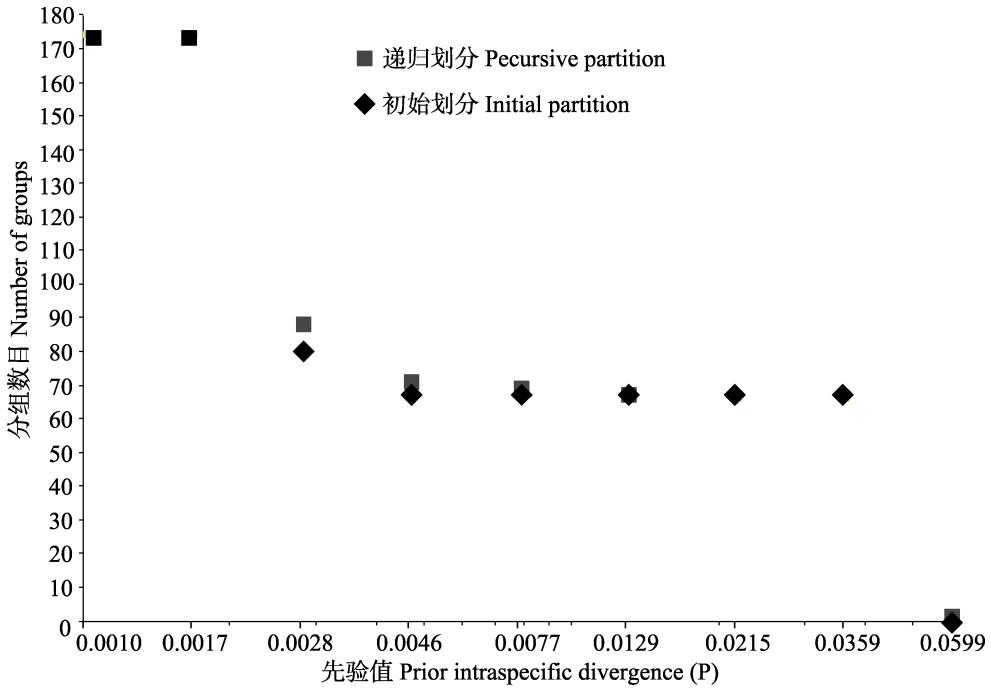

图 3　ABGD 方法划分结果
Fig. 3　The automatic partition results of dataset by ABGD

划分较为稳定，483 个样本均被分成 67 组；递归划分将样本分成了 1~25 组，但在一个很大的 P 值范围内维持在 14 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，发现每个分组与形态种均呈现一一对应的关系，基于样本遗传距离的 ABGD 方法划分物种的成功率为 100%。

2.4 GMYC 模型划分 MOTUs 结果

经 JModelTest2.2.2 检测发现该数据集适用于 HKY+I+G 碱基替代模型。基于 BEAST 的 GMYC 方法（图 4，表 2）零模型与 GMYC 模型之间的似然比值（LRT）为 182.9172***，种（Yule 模型）-种群（溯祖模型）的转折点发生在–0.0111 百万年前（Million years ago, Mya），以及转折点对应的物种数量为 66（64~68）。基于 d8tree（表 2）的 GMYC 方法零模型与 GMYC 模型之间的似然比值（LRT）为 167.931***，种（Yule 模型）-种群（溯祖模型）的转折点发生在 - 0.0159 Mya，以及转折点对应的物种数量为 73（69~77）。基于 MPLtree（表 2）的 GMYC 方法零模型与 GMYC 模型之间的似然比值（LRT）为 167.9826***，种（Yule 模型-种群（溯祖模型）的转折点发生在 - 0.0074 Mya，以及转折点对应的物种数量为 73（69~77）。结果表明，基于 BEAST 软件的 GMYC 方法较为可靠，与形态学划分结果一致。基于 PATHd8 软件的 GMYC 方法高估了物种划分数目。就计算速度而言，PATHd8 软件构建超度量的树速度比基于 BEAST 软件快；就稳定性而言，二者结果都较稳定，似然比检验结果极显著（表 2，***表示 P<0.001）。

3 讨论

DNA 条形码经历了从刚开始提出时的广泛关注与争议阶段到应用与争议并存阶段（金倩和张爱兵，2013）。在本研究中 DNA 条形码作为一个未知类群的勘测工具。当没有其余的独立数据（例如物种描述文献）跟条形码的结果进行比较时，条形码对于此类群的预测将会作为物种假设的第一步。我们以中国北京周边地区 10 个采样点 483 个舟蛾科样品为例，尝试基于 DNA 条形码进行物种界定算法（jMOTU、ABGD、GMYC）的比较研究，结果表明，基于 DNA 条形码的物种界定方法估计效力存在差异，ABGD 方法划分物种的准确率为 100%，基于 BEAST 的 GMYC 模型产生的置信区间覆盖了形态学的结果，说明 ABGD 方法和基于 BEAST 的 GMYC 模型能够较为准确地界定舟蛾科物种，然而，基于 d8tree/MPLtree 的 GMYC 方法倾向于高估 MOTUs，jMOTU 方法倾向于低估物种数目。

研究者曾试图建立种内种间差异的标准值（3%或 10 倍法则）对物种进行阈值划分（Hebert *et al.*，2004），但实际研究中种内种间经常存在重叠，定义一个阈值较为困难。本研究中种内遗传距离为 0~0.1594，平均为 0.0063；种间遗传距离为 0.05~0.2419，平均值为 0.1409，虽然平均种间遗传距离是平均种内遗传距离的 22.4 倍，但是最小种间遗传距离 0.05 小于最大种内遗传距离 0.1594 ，这使种内和种间遗传距离分布出现了重叠现象（Overlap），如图 1 所示，种内和种间没有严格定义的 DNA 条形码空白区。jMOTU 是通过设定阈值对序列进行划分的方法，序列差异小于指定阈值的样本被划分到一个 MOTU 里，根据不同阈值下的 MOTU 数目，可以直观的看到样本被划分成的类群数目。在本研究中，我们用一个 java 程序 jMOTU v1.0.6（Jones *et al.*，2011）来分析数据，通过多次预实验，选取 1~100 的分子序列差异截取值（Cut-off value）进行分析，然后根据选择不同的差异截取值来确定 MOTUs 个数。在理想的情况下，DNA 条形码空白区（DNA barcoding gap）是通过在一个阈值范围内的富集 MOTUs 的平台来显示，我们通过该平台来确定 MOTUs 的数目。本研究中所得 MOTUs 数量与用于定义物种的序列差异之间的关系如图 2 所示，并没有出现较为明显的 MOTUs 富集平台区，在 7 bp（1.18%）的序列差异时分成了 68 个 MOTUs，但是这种划分结果随着序列差异的增大，逐渐减小，当序列差异为 36 bp（6.06%）时划分结果为 51 个 MOTUs。这与 DNA

表 2　GMYC 模型界定物种

Table 2　GMYC delimitation in this study

形态种 Morphyospecies	方法 Method[a]	界定物种数量包含单个体数 Entities[b]	界定物种数量不包括单个体数 Clusters[c]	零假设的似然值 L_{null}[d]	GMYC 模型的似然值 L_{GMYC}[e]	似然比 Likelihood ratio[f]
67	BEAST	66（64-68）	39（39-40）	1 177.53	1 268.989	182.9172***
	d8tree	73（69-77）	38（37-38）	1 049.785	1 133.751	167.931***
	MPLtree	73（69-77）	38（37-38）	1 187.187	1 271.178	167.9826***

a: 基于 BEAST/d8tree/MPLtree 超度量树分析；b: GMYC 界定的物种数量，包括单个体数；c: 物种数量但不包括单个体数；d: 零假设的似然值；e: GMYC 模型的似然值；f: 似然比值检验(***表示 P<0.001，**表示 P<0.01，*表示 P<0.05)。

a: Different ultrametric tree obtention method; b: Number of entities including single individual; c: Number of entities without single individual; d: The likelihood value of null model; e: The maximum likelihood value of GMYC model; f: Likelihood ratio test (*** P<0.001, ** P<0.01, * P<0.05).

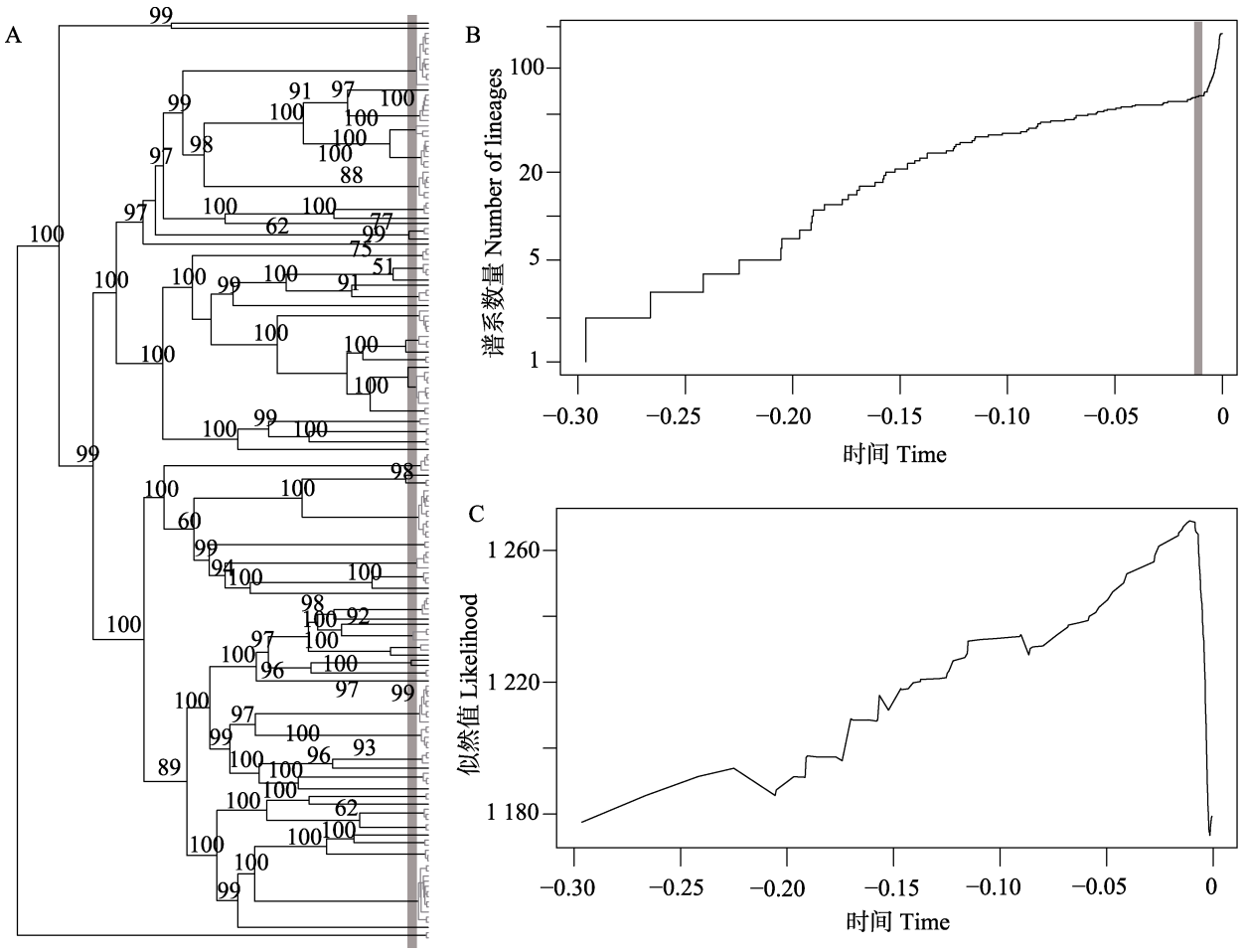

图 4　GMYC 物种界定

Fig. 4　GMYC model delimitation

A. 基于 BEAST 软件构建的超度量树，灰色竖线表示物种形成事件和种群溯祖事件的转折点，枝长为灰色表示种下种群溯祖事件，枝长为黑色表示种上物种形成事件；B. 时间和谱系间的关系，曲线表示枝长速率，枝长速率急剧上升的变化转折点对应物种界定的数量；C. 似然值和时间的变化曲线。

A. The ultrametric tree implemented with BEAST software. Gray line indicates transition from between-species to within-species. Gray branches represent coalescence processes. Black branches represent speciation and extinction processes; B. Relationship between time and lineages. The sharp increase in branching rate, corresponding to the transition from interspecies to intraspecies branching events; C. Relationship between time and likelihood value.

条形码空白区的结果相一致（图 1）。说明该方法不适用于本研究类群，易低估 MOTUs 数目。

基于阈值分类的 ABGD 方法在本研究类群中与形态学结果相一致，这归咎于 ABGD 算法本身的优势，它能够自动找到条形码空白区的位置，即使分布重叠，也可以划分数据集到候选物种。该方法从数据中统计推断并相应地划分数据集，并递归应用该算法直到新获得的类群序列，从而允许数据集多个阈值分类。对于本研究的舟蛾科类群，有效验证了该方法虽然没有较为明显的 DNA 条形码空白区，依然能够对数据集进行准确划分。ABGD 方法计算效率较高，操作也较为简单，只需将数据以 Fasta 格式在线提交到 ABGD 网站计算即可。如图 3 所示，本研究将初始划分和递归划分的结果都输出，递归划分能更好的处理数据集中差异较小的序列，但初始划分却在一个大范围 P 值内表现出稳定性，并且与形态种更加相近。我们需要注意的是，ABGD 方法在理论上也存在一定的局限性。例如：种内的序列条数最好大于 3~5 条时，该算法才奏效。通过模拟成种事件和种群遗传过程发现此算法对于最近成种事件较为敏感。此外，此算法必须进行先验推断，若没有先验知识则无法对物种进行正确划分。

基于 GMYC 模型的物种界定效力在很大程度上依赖于超度量树的构建质量，本研究通过两种方式构建超度量系统发育树：(1)通过 PATHd8 软件构建 d8tree 和 MPLtree；(2)由 BEASTv1.8.0 软件直接计算获取（Drummond and Rambaut，2007）。本研究结果发现，基于 BEAST 的 GMYC 方法较为稳定可靠，说明该途径构建的超度量树更适合本研究类群的物种划分。从计算效率上看，基于 BEAST 的 GMYC 方法虽然准确度更高，但整个计算过程耗时，尤其对于大数据而言往往不太适合。相反，基于 PATHd8 软件的 GMYC 方法因计算时间少往往更加受到使用着的青睐，但是该方法易高估 MOTUs 数目（Talavera *et al.*，2013），在本研究中也发现该现象。当数据集较小时，建议采用基于 BEAST 的 GMYC 对进行物种界定估计，该方法可以对基于形态学的物种界定进行有效补充。

本研究对基于 DNA 条形码物种界定算法进行了初步研究，发现不同的物种界定方法估计效力存在差异，其中 ABGD 方法和基于 BEAST 的 GMYC 模型适合本文研究对象舟蛾科，该研究结果将对生态学和生物防治的基础调查研究有所助益。当生态学家进行野外标本收集时，单纯依靠传统的分类需要耗费大量的人力、物力整理、厘定这些标本，而恰当地运用 DNA 条形码方法将会极大地加速这一进程。然而，本研究虽然历时 3 年涉及 10 个采样点大范围的收集样品，但野外采样过程中还是很难将所有类群收集全面，后续研究需持续进行样品的收集，对于野外收集不到的类群将从 GenBank/BOLD 等数据库中下载，将数据整合后，分为不同层次深入探讨，使得研究更加完善。另外，仅采用一段较短基因片段（COⅠ）尝试界定物种存在一定的局限性，短片段不能提供物种较完整的信息，后续研究应当尽可能的补充一些其他基因序列的信息，接受新证据的不断验证。

参考文献 (References)

Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature*, 392(6671): 71–75.

Boykin LM, Armstrong KF, Kubatko L, De Barro P, 2012. Species delimitation and global biosecurity. *Evolutionary Bioinformatics*, 8: 1.

Britton T, Anderson C, Jacquet D, Lundqvist S, Bremer K, 2006. PATHd8-a program for phylogenetic dating of large trees without a molecular clock. Avalilable at www math su se/PATHd8.

Britton T, Anderson CL, Jacquet D, Lundqvist S, Bremer K, 2007. Estimating divergence times in large phylogenetic trees. *Systematic Biology*, 56(5): 741–752.

Darriba D, Taboada GL, Doallo R, Posada D, 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8): 772–772.

Dellicour S, Flot JF, 2015. Delimiting species-poor datasets using single molecular markers: a study of barcode gaps, haplowebs and GMYC. *Systematic Biology*, 64(6): 900–908..

Drummond AJ, Ho SY, Phillips MJ, Rambaut A, 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5):

e88.

Drummond AJ, Rambaut A, 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1): 214.

Fujisawa T, Barraclough TG, 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent (GMYC) approach: a revised method and evaluation on simulated datasets. *Systematic Biology*, 62(5): 707–724.

Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W, 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(41): 14812–14817.

Jin Q, Zhang AB, 2013. Distance-based DNA barcoding methods for insects. *Chinese Journal of Applied Entomology*, 50(1): 283–287. [金倩, 张爱兵, 2013. 昆虫 DNA 条形码分析中的距离方法. 应用昆虫学报, 50(1): 283–287.]

Jones M, Ghoorah A, Blaxter M, 2011. jMOTU and Taxonerator: turning DNA Barcode sequences into annotated operational taxonomic units. *PLoS ONE*, 6(4): e19259.

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP, 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55(4): 595–609.

Puillandre N, Lambert A, Brouillet S, Achaz G, 2012. ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, 21(8): 1864–1877.

Rambaut A, Drummond A, 2007. Tracer v1. 4: MCMC trace analyses tool. See http://tree bio ed ac uk/software/tracer/[PMC free article][PubMed].

Ratnasingham S, Hebert PD, 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. *PLoS ONE*, 8(7): e66213.

Song SS, Shi ZY, Jin Q, Han HL, Lin XF, Hao MD, Zhang AB, 2014. Species identification of Noctuidae moths (Insecta: Lepidoptera) from Baoding and Langfang, Hebei, China with DNA barcoding and establishment of a local DNA barcoding library. *Chinese Journal of Applied Entomology*, 51(1): 156–168. [宋韶彬, 石志勇, 金倩, 韩辉林, 刘晓枫, 郝梦迪, 张爱兵, 2014. 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. 应用昆虫学报, 51(1): 156–168.]

Stamatakis A, 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21): 2688–2690.

Talavera G, Dincă V, Vila R, 2013. Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution*, 4(12): 1101–1110.

Thomsen PF, Willerslev E, 2015. Environmental DNA–an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.*, 183: 4–18.

Vrijenhoek R, 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5): 294–299.

Yang CH, Jin Q, Chen FQ, Wu CS, Zhang AB, 2013. Application of tree-based, distance-based and character-based DNA barcoding methods for the identification of Crambidae species (Insecta: Lepidoptera) in Baihuashan, China. *Chinese Journal of Applied Entomology*, 50(1): 61–70. [杨聪慧, 金倩, 陈付强, 武春生, 张爱兵, 2013. 基于进化树、距离和特征的 DNA 条形码方法研究－以百花山地区草螟科为例. 应用昆虫学报, 50(1): 61–70.]