

DOI: 10.3969/j.issn.2095-1787.2012.04.013

支持向量机和邻接法在夜蛾科昆虫条码研究中的应用

李俊¹, 韩辉林², 高 强¹, 金 倩¹, 迟美妍¹, 武春生³, 张爱兵^{1*}

¹首都师范大学生命科学学院, 北京 100048; ²东北林业大学林学院, 黑龙江 哈尔滨 150040;

³中国科学院动物研究所, 北京 100101

摘要:【背景】鳞翅目夜蛾科昆虫种类繁多, 目前已经超过 3.5 万种, 绝大多数是农林生产的主要害虫。由于多数近缘属种形态相似, 难以鉴定, 给农林害虫的防治工作带来了很大的困难。DNA 条形码技术是一种快速、准确鉴定物种的方法。支持向量机作为一种新的机器学习方法, 自 1995 年被提出以来已经在数据分类和高维模式识别等领域取得不错的效果。【方法】将北京妙峰山采集的 58 种夜蛾 101 个样品的 COI 序列分成 3 套数据集, 分别通过邻接法和支持向量机对其进行验证。【结果】通过对 DNA 条形码物种鉴定结果的验证表明, 邻接法优于支持向量机。但 DNA 条形码在鉴定夜蛾科的一些近缘种上, 效果不佳, 如棉铃虫和烟青虫。【结论与意义】DNA 条形码作为一种新兴的物种鉴定方法, 在分类学上具有很高的应用价值。通过邻接法和支持向量机的比较, 虽然支持向量机的成功率低于邻接法, 但是其在 DNA 条形码中的应用是对数据询问方式的一种探索。

关键词: DNA 条形码; 夜蛾科; 支持向量机; 邻接法; 近缘种

Species identification of Noctuidae (Insecta: Lepidopera) with DNA barcoding of support vector machine and neighbor-joining method

Jun LI¹, Hui-lin HAN², Qiang GAO¹, Qian JIN¹, Mei-yan CHI¹, Chun-sheng WU³, Ai-bing ZHANG^{1*}

¹College of Life Sciences, Capital Normal University, Beijing 100048, China; ²School of Forestry, Northeast Forestry University, Harbin, Heilongjiang 150040, China; ³Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

Abstract: 【Background】Several major pests in agriculture and forestry belong to the lepidopteran family of Noctuidae, which are often difficult to distinguish morphologically. As a fast and accurate approach to the identification of species, DNA barcoding is used in identification since 2003. This, accompanied with a machine learning method, the support vector machine (SVM) has achieved good results in data classification, high dimensional pattern recognition and other fields. The combination of these two methods holds promise to ease the correct identification of noctuids, which would improve pest control efforts. 【Method】DNA barcodes of the mitochondrial cytochrome oxidase c subunit I (COI) gene, divided into three sets of databases, were tested by the neighbor-joining (NJ) method and the SVM. We used 101 specimens of 58 species of Noctuidae which were identified by experts. These specimens were collected mainly from the Miaofeng Mountain, north of Miaofengshan Town, Beijing, China. 【Result】The NJ method proved to be superior to the SVM. However, some important species, such as *Helicoverpa armigera* and *H. assulta*, could not be distinguished in the NJ-tree. 【Conclusion and significance】As a new technology of species identification, DNA barcoding can provide assistance in problems in taxonomy; combination with the NJ method gave better support than the SVM.

Key words: DNA barcoding; Noctuidae; support vector machine; NJ method; closely related species

夜蛾科隶属鳞翅目, 为该目第一大科, 目前已知种类超过 3.5 万种 (Poole, 1989)。由于夜蛾科昆虫具有危害方式多样、发生时期不同、近缘种类繁多、个体变异丰富等特点, 且从事传统分类的学者日趋减少等, 导致除粘虫类、地老虎类、烟青虫 *Helicoverpa assulta* (Guenée)、棉铃虫 *Helicoverpa armig-*

收稿日期 (Received): 2012-08-31 接受日期 (Accepted): 2012-10-30
基金项目: 北京市自然基金重点项目 (KZ201010028028); 北京市人才强教项目 (PHR201107120); 国家自然科学基金项目 (31071963、31272340、31272355)
作者简介: 李俊, 男, 硕士研究生。研究方向: 分子生态学。E-mail: cnu_lijun@163.com
* 通讯作者 (Author for correspondence), E-mail: zhangab2008@mail.cnu.edu.cn

era (Hübner) 等重要的农林害虫外,其他大部分种类的相关研究极其薄弱(陈一心,1999)。近年来,随着形态分类学多元信息的整合以及分子生物学技术的快速发展,传统夜蛾科分类系统发生了很大变化(Zahiri *et al.*, 2010、2012)。但是,由于新系统还存在一定的不稳定性,本文以传统的分类系统为主予以研究。

DNA 条形码被 Hebert *et al.* (2003a、b) 和 Remigio & Hebert (2003) 提出之后,得到了国际上的广泛关注。许多研究者将 DNA 条形码应用于物种鉴定中,其利用一段基因序列作为分子标记来快速、准确、自动化地对物种进行鉴定(Hajibabaei *et al.*, 2006)。目前, DNA 条形码数据分析倡议协会(DNA Barcode Data Analysis Initiative, DBDAI) 把线粒体 COI 基因 5'端起始的长度为 658 bp 的一段序列作为动物物种鉴定的标准序列(Ratnasinghan & Hebert, 2007)。这段 COI 序列在后生多细胞动物群,包括脊椎动物(Francis *et al.*, 2010; Hubert *et al.*, 2008; Tavares & Baker, 2008) 和无脊椎动物(Jennings *et al.*, 2010; Johnson *et al.*, 2008; Zhou *et al.*, 2009) 的物种鉴定和隐存种鉴定方面都很有效(Efe *et al.*, 2009; Johnson *et al.*, 2008; Kerr *et al.*, 2009)。但是, DNA 条形码技术在物种鉴定方面仍然存在争议(Will & Rubinoff, 2004), 一些以形态鉴定为主的学者认为,物种鉴定需要依靠经验(Elias *et al.*, 2007; Hurst & Jiggins, 2005; Meyer & Paulay, 2005; Wiemers & Fiedler, 2007) 和理论知识(Hickerson *et al.*, 2006) 的积累。

支持向量机(support vector machine) 又称支持向量网络(Cortes & Vapnik, 2009), 于 1995 年被提出,是在统计学理论上发展起来的一种新的机器学习方法,在机器学习的许多方面表现出特有的优势,如解决小样本、非线性和高维模式识别等问题,并且在函数拟合等机器学习中得到推广和应用(郎宇宁和蔺娟如, 2010)。由于支持向量机具有理论完备、适应性强、全局优化、训练时间短、泛化性能好等优点,已经成为目前国内外研究的热点(边肇祺和张学工, 2000)。目前,支持向量机在数据分类(杜灿谊等, 2012; 李晟和何明一, 2012; 秦传东等, 2012)、回归估计、函数逼近等领域展开了应用并取得良好的效果(Brown *et al.*, 2000; Mukherjee, 2003; Walavalkar, 2003; Zhao & Principe, 2001)。

然而,支持向量机应用于 DNA 条形码方法的例子还较少。因此,本研究以北京妙峰山采集的 101 只夜蛾科昆虫的 COI 序列和网站下载的 14 条夜蛾 COI 序列作为研究材料,通过传统的形态鉴定方法和 DNA 条形码分类方法鉴定物种,利用支持向量机和邻接法(Mailund *et al.*, 2006) 对鉴定结果进行研究,以进一步验证 COI 基因对鳞翅目夜蛾科的鉴定效果。

1 材料与方法

1.1 试验材料

在北京妙峰山地区的鹫峰(E116°06.081'、N40°06.907')、陇驾庄(E116°03.347'、N39°58.321')、涧沟(E116°01'55.1"、N40°03'44.7") ,利用高压汞灯诱集夜蛾科昆虫,分别采样 21、20、60 头,然后冷冻致死。

1.2 样品的保存与鉴定

样本采集后展翅保存,制作外生殖器玻片,以进行形态学鉴定。经形态专家鉴定,101 个样本来自 58 种夜蛾(表 1)。

1.3 COI 序列的扩增及分析

DNA 模板选自样本腿部肌肉组织,通过 Biomed 组织/细胞基因组 DNA 快速提取试剂盒提取。COI 序列通过聚合酶链式反应(PCR) 扩增,反应体系 30 μ L: 模板 DNA 3 μ L, 引物 COI F 0.6 μ L, 引物 COI R 0.6 μ L, 耐热聚合酶链反应化合物 15 μ L, 双蒸水 10.8 μ L。采用的引物:上游引物 F1 ATTCAACCAATCATAAAGATAT, 下游引物 R1 TAAACTTCTGGATGTCCAAAAA。PCR 反应条件:94 $^{\circ}$ C 预变性 2 min; 40 个循环包括 94 $^{\circ}$ C 变性 20 s, 52 $^{\circ}$ C 退火 20 s, 72 $^{\circ}$ C 复性 45 s, 72 $^{\circ}$ C 延伸 10 min, 4 $^{\circ}$ C 保存。取 3 μ L 反应体系进行 1% 琼脂糖电泳检测,若有条带在 650 bp 左右,则送中科希林生物有限公司北京分公司测序。

将提取的 COI 序列组合分成 3 套数据集。所有夜蛾样本组成的数据集为数据集 I,剔除所有单序列的夜蛾种类组成的数据集为数据集 II,在数据集 II 的基础上增加网站上下下载的 14 条夜蛾科数据构成的数据集为数据集 III。将这些 COI 序列集分别通过 Chromas 进行编辑和修改,再用 Mega 4.0 (Tamura *et al.*, 2007) 对序列进行比对,转化为 Mega 4.0 (Tamura *et al.*, 2007) 可读文件。在此基

础上,使用 Mega 4.0 (Tamura *et al.* ,2007) 和 K2P 数据进行遗传距离分析并构建 NJ 树(Saitou & Nei, (Kimura,1980)模型中的计算和建树功能,对 3 套 1987)。

表 1 样本及其 COI 序列的 GenBank 登录序列号
Table 1 Noctuidea samples and GenBank accession numbers of their COI gene sequences

物种 Species	样本序列号 Sample ID	序列号 Accession no.	物种 Species	样本序列号 Sample ID	序列号 Accession no.
中赫夜蛾 <i>Acosmetia chinensis</i>	LJZ100726.623	JX509744	长须夜蛾 <i>Hypena tristalis</i>	LJZ100726.556	JX509823
	LJZ100813.016	JX509745	纹薹夜蛾 <i>Bryophita orthogramma</i>	LJZ100726.592	JX509742
	LJZ100813.154	JX509746	平嘴壶夜蛾 <i>Calyptra lata</i>	JG100812.041	JX509814
曲线奴夜蛾 <i>Paracolax tristalis</i>	JG100802.571	JX509755	朽木夜蛾 <i>Axylia putris</i>	JG100802.396	JX509769
线委夜蛾 <i>Athetis lineosa</i>	JF100809.052	JX509737		JG100812.206	JX509802
蒙阴夜蛾 <i>Hadula schawyra</i>	LJZ100726.480	JX509741	白钩粘夜蛾 <i>Mythimna proxima</i>	JG100812.448	JX509809
暗翅须夜蛾 <i>Polypogon gryphalis</i>	LJG100812.320	JX509806	独夜蛾 <i>Nikara castanea</i>	JG100802.149	JX509795
旋歧夜蛾 <i>Hadula trifolli</i>	JF100809.008	JX509736		LJZ100726.627	JX509796
	JF100726.049	JX509783		JF100809.214	JX509797
	JF100809.148	JX509784		JG100812.255	JX509804
	JG100802.609	JX509785	朽镰须夜蛾 <i>Zanclogna thalunalis</i>	JF110810.107	JX509831
	LJZ100813.093	JX509786	淡剑灰夜蛾 <i>Spodoptera depravata</i>	JF100809.166	JX509780
白线散纹夜蛾 <i>Calloplistria albolineola</i>	JF110810.061	JX509829		LJZ100726.594	JX509781
桑剑纹夜蛾 <i>Acronicta major</i>	JG100802.328	JX509774		LJZ100726.665	JX509782
	JG100812.518	JX509811	沪齐夜蛾 <i>Imosca coreana</i>	JG100802.267	JX509752
	JF100809.021	JX509819	榆剑纹夜蛾 <i>Acronicta Hercules</i>	JG100812.432	JX509813
焰夜蛾 <i>Pyrhia umbra</i>	JG100802.657	JX509770	麟角希夜蛾 <i>Eucarta virgo</i>	LJZ100813.009	JX509825
	JG100812.510	JX509810	纹希夜蛾 <i>Eucarta fasciata</i>	JG100802.048	JX509747
纶夜蛾 <i>Thalatha sinens</i>	JG100802.332	JX509753	大三角鲁夜蛾 <i>Xestia kollari</i>	JG100802.601	JX509759
大斑蕊夜蛾 <i>Cymatophoropsis unca</i>	JF100809.016	JX509818		JG100802.151	JX509760
银锭夜蛾 <i>Macdunnoughia crassisigna</i>	JF100726.072	JX509834		JG100812.378	JX509808
白痣眉夜蛾 <i>Pangrapta lunulata</i>	LJZ100726.689	JX509822	丘秘夜蛾 <i>Mythimna monticola</i>	JG100802.786	JX509761
谐夜蛾 <i>Emmelia trabealis</i>	JF100809.153	JX509833		JG100802.650	JX509762
肯髯须夜蛾 <i>Hypena kengkalis</i>	LJZ100813.106	JX509826		JG100802.003	JX509763
瓠首夜蛾 <i>Cranionycta oda</i>	JG100802.721	JX509832	锦夜蛾 <i>Euplexia lucipara</i>	JG100812.153	JX509833
长白镰须夜蛾 <i>Zanclognatha lui</i>	JG100802.208	JX509751	草禾夜蛾 <i>Oligia furuncula</i>	JG100812.373	JX509807
	JG100802.607	JX509798	棉铃实夜蛾 <i>Helicoverpa armigera</i>	JF100809.315	JX509739
				JG100812.126	JX509764
蒿冬夜蛾 <i>Cucullia fraudatrix</i>	LJZ100813.061	JX509827		JG100812.271	JX509765
烟青虫 <i>Helicoverpa assulta</i>	LJZ100726.679	JX509775		JG100812.394	JX509766
	JG100802.063	JX509776		JG100802.597	JX509756
	JG100802.734	JX509777	陌夜蛾 <i>Trachea atriplicis</i>	JG100802.534	JX509812
			焰实夜蛾 <i>Heliocheilus fervens</i>	JF100809.137	JX509738
东北巾夜蛾 <i>Dysgonia mandschuriana</i>	JG100812.505	JX509771	贯雅夜蛾 <i>Lambia transversa</i>	JF100809.296	JX509817
灰缘贫夜蛾 <i>Simplicia mistacalis</i>	JG100802.637	JX509757	宽胫夜蛾 <i>Schinia scutosa</i>	JG100802.307	JX509835
寒切夜蛾 <i>Euxoa sibirica</i>	JF110810.105	JX509830	直线希夜蛾 <i>Eucarta arcta</i>	JF100726.303	JX509816
线委夜蛾 <i>Athetis lepigone</i>	LJZ100726.354	JX509740	苇实夜蛾 <i>Heliothis maritime</i>	JG100802.205	JX509750
间盗夜蛾 <i>Hadena corrupta</i>	JG100812.141	JX509801	V 纹金翅夜蛾 <i>Autographa v-minus</i>	JG100802.651	JX509758
三斑蕊夜蛾 <i>Cymatophoropsis trimaculata</i>	JG100802.507	JX509828	兀鲁夜蛾 <i>Xestia ditrapezium</i>	JG100802.130	JX509772
	JG100812.070	JX509800		JG100802.549	JX509773
	JG100802.099	JX509748	烦寡夜蛾 <i>Sideridis incommoda</i>	JG100812.114	JX509836
女真首夜蛾 <i>Craniophora ligustri</i>	JG100812.227	JX509768	日雅夜蛾 <i>Lambia japonica</i>	JG100802.148	JX509749
	JF100809.161	JX509767	胞短栉夜蛾 <i>Brevipecten consanguis</i>	JF100809.010	JX509778
	JF110810.115	JX509830		JG100802.435	JX509779
乏夜蛾 <i>Nipponyx segregate</i>	JG100802.125	JX509788		LJZ100726.613	JX509743
	JG100802.274	JX509789		JG100812.304	JX509805
	JG100812.218	JX509791	暗杂夜蛾 <i>Amphipyra erebina</i>	JG100812.248	JX509803
	JG100802.662	JX509790		JF100809.353	JX509820
	JG100812.570	JX509792	苹梢鹰夜蛾 <i>Hypocala subsatura</i>	JG100802.932	JX509799
	LJZ100726.602	JX509793			
	LJZ100726.070	JX509794			
	JG100812.183	JX509815			
	LJZ100726.357	JX509824			

对所构建的 NJ 树进行成功率的评价。在序列 Q 经形态鉴定为物种 A 的情况下,序列 Q 跟物种 A 聚在一起,而 A 没有其他的分支,认为 NJ 树聚类成功;Q 在物种 A 和 B 之外,认为 Q 属于其他物种, NJ 树聚类失败;物种 A 形成 2 个分支,而序列 Q 聚集在占物种 A 个体数大于或等于 50% 的分支上,认为 Q 属于 A 物种, NJ 树聚类成功;如果测试序列 Q 聚集在物种 A 个体数不到 50% 的分支上,则认为其不属于 A 物种, NJ 树聚类失败。这样计算 1000 次的成功率:

$$N_{\text{sample}} = N_{\text{seq}} \times 1000; \quad R_{\text{success}} = N_{\text{success}} / N_{\text{sample}}$$

支持向量机对 3 套数据进行分类,采用 libsvm 软件对所得到的妙峰山夜蛾科昆虫的 COI 序列数据集进行研究,将其转化为 libsvm 可读文件(A = 0.1, T = 0.2, C = 0.3, G = 0.4)。在支持向量机类型为 C-SVC 和核函数类型为径向基函数的基础上,得到夜蛾科 COI 序列的成功率。为了得到支持向量机的准确率(accuracy)、敏感性(sensitivity)、特异性(specificity)和精密度(precision)4 个参数,在原有的 3 套数据集上加入了 22 条舟蛾序列作为负样本,分别命名为数据集 I'、数据集 II' 和数据 III'。以数据集 II' 为例,数据集 II' 有 55 条序列,加入 22 条舟蛾科的 COI 序列,总共 77 条 COI 序列,随机抽取数据集 II' 80% 的序列(62 条序列)作为训练,数据集 II' 20% 的序列(15 条序列)作为测试,通过以下公式(Perruchet & Peereman, 2004; Velez *et al.*, 2007; Walker *et al.*, 1990)计算支持向量机的 4 个参数。重复随机抽取 1000 次,得到它们的平均值进而验证支持向量机的测试效果。

$$\text{precision} = \frac{\text{true_positive}}{\text{true_positive} + \text{false_positive}}$$
$$\text{sensitivity} = \frac{\text{true_positive}}{\text{true_positive} + \text{false_negative}}$$
$$\text{specificity} = \frac{\text{true_negative}}{\text{true_negative} + \text{false_positive}}$$
$$\text{accuracy} = \frac{\text{true_negative} + \text{true_positive}}{\text{true_negative} + \text{false_positive} + \text{true_positive} + \text{false_negative}}$$

2 结果与分析

2.1 基于 NJ 方法和支持向量机的夜蛾科鉴定

采用 NJ 方法的鉴定中,由于数据集 I 存在许多单序列样本(有些夜蛾种类仅有 1 个样本),在对 NJ 方法成功率的评估上效果不是很理想

(55.5%)。因此,在数据集 II 中剔除单序列样本后进行验证,成功率达到 89.1% (95% 的置信区间 CI: 82.4% ~ 95.8%) (Clopper & Pearson, 1934)。在对数据集 III 的评估中,成功率达到 91.3% (95% 的置信区间 CI: 86.3% ~ 96.3%) (Clopper & Pearson, 1934)。

通过支持向量机的方法对 3 套数据集(未加入舟蛾数据集)进行鉴定,得到数据集 I、数据集 II、数据集 III 的成功率分别为 41.3% (95% 的置信区间 CI: 40.7% ~ 42.0%)、83.8% (95% 的置信区间 CI: 83.1% ~ 84.5%)、83.5% (95% 的置信区间 CI: 82.9% ~ 84.1%) (Clopper & Pearson, 1934)。

舟蛾数据集对支持向量机的成功率影响较小(表 2)。从图 1 也可以看出,增加物种数量和同种个体数, NJ 方法的验证效果会提高,而对支持向量机的影响较小。总体来说, NJ 方法的效果优于支持向量机,主要由于它们的验证方法不同,支持向量机需要把数据集分为 2 份,一套进行训练,一套进行测试,很可能将 2 个同种数据作为预测的数据集,从而降低了成功率。如果样本容量足够大,同种个体数足够多,支持向量机的成功率可能会提高。

表 2 通过 3 套数据集计算出的支持向量机的 4 个参数
Table 2 Four parameters of SVM calculated by three databases

数据集 Database	精密度 Precision	特异性 Specificity	敏感性 Sensitivity	准确率 Accuracy
I'	0.40	0.26	1	0.51
II'	0.82	0.69	1	0.87
III'	0.83	0.65	1	0.87

3 套数据集是在原数据集的基础上增加 22 条舟蛾的 COI 序列作为负样本。各参数值的区间均为 0 ~ 1.00。
Three databases are negative sample added 22 Noctuoidea COI sequences based on the original databases. The range of every parameter is 0 ~ 1.00.

2.2 系统发育树分析

基于数据集 II COI 序列构建的 NJ 树如图 2 所示。NJ 树可以正确地按照亲缘关系将 55 个样本聚类,节点中的数值为各自的 Bootstrap 值。从图中可以发现,除棉铃虫和烟青虫、大三角鲁夜蛾 *Xestia kollari* (Laderer) 和兀鲁夜蛾 *X. ditrapezium* (Denis et Schiffermuller) 外,其他样本在 NJ 树上大体都有清晰的划分。从传统形态学上看,棉铃虫与烟青虫的亲缘关系比较近,生活的环境相似,它们之间出

现了杂交现象。DNA 条形码鉴定是依靠物种的种间基因遗传差异比种内基因遗传差异大进行区分的,而棉铃虫与烟青虫遗传交流比较多,因此它们在 NJ 树上聚为一类,线粒体 COI 基因很好地反映了这个事实。大三角鲁夜蛾和兀鲁夜蛾也有类似情况。乏夜蛾 *Nipponyx segregata* (Butler) 中编号为 LJZ100726.357 的样本在 NJ 树上为单独一支,可能由于基因杂交或样本问题,这有待进一步研究。

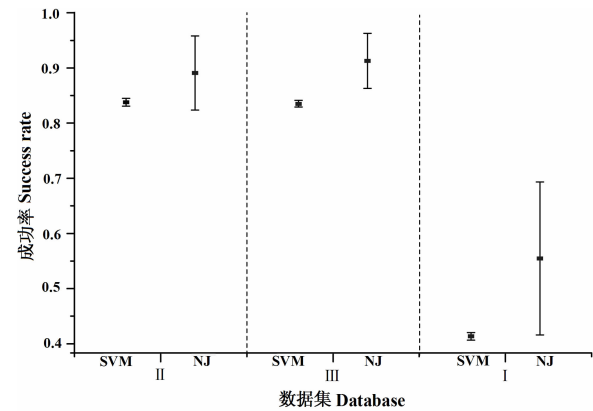


图 1 通过邻接法和支持向量机在夜蛾科 COI 序列数据集物种鉴定的成功率 (未加入舟蛾数据集)
Fig.1 Success rate in species identification, based on the COI gene, using either the SVM or NJ, performed on the non-joined Noctuoidea dataset

3 讨论

本研究中,邻接法 (Mailund *et al.*, 2006) 在数据集 II 和数据集 III 的成功率都为 90% 左右,而支持向量机的成功率为 83% 左右,两者都有很高的成功率。这验证了 DNA 条形码在分类学上的应用价值。已有证据表明,COI 序列在某些类群或者某个地区内有很好的鉴别能力,如 COI 序列鉴定热带鳞翅目昆虫的成功率达到 97.9% (Hajibabaei *et al.*, 2006)。

DNA 条形码作为一种新兴技术,在对物种分类识别上具有快捷、简单的优点。前人已有 DNA 条形码与机器学习方法相结合的研究,在 2008 年 Zhang *et al.* 将人工智能的思想引入 DNA 分子分类学领域,提出了基于 BP 人工神经网络的物种鉴定新方法。本研究将 DNA 条形码与支持向量机结合证明了 DNA 条形码在样品鉴定中具有较高的准确性。虽然现在普遍应用的邻接法 (Mailund *et al.*, 2006) 的成功率比支持向量机高,但是随着分子技术的不断更新,生物信息数据量不断扩增,将有越来越多的人依赖人工智能的方法对数据进行查询。因

此,本研究也是对数据查询方法的一种探索。DNA 条形码快速识别数据网络的建立,将为以后的研究提供便利,也给生物多样性研究创造了有利条件。

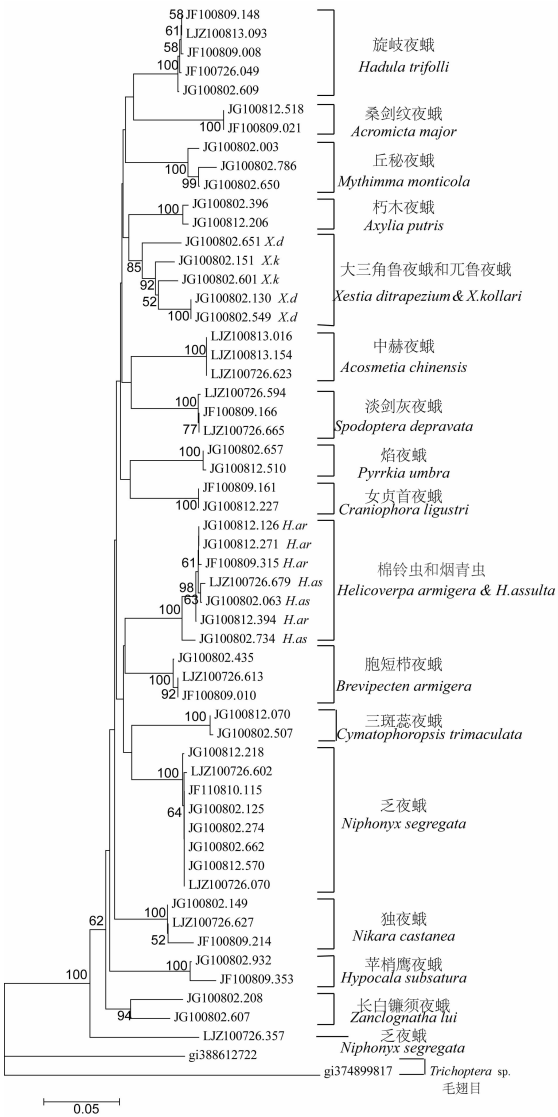


图 2 基于数据集 II 的 COI 序列构建的部分夜蛾种类的 NJ 树
Fig.2 NJ-tree of 55 slected Noctuidae species, based on the COI gene of database II

参考文献

边肇祺, 张学工. 2000. 模式识别. 北京: 清华大学出版社.
陈一心. 1999. 中国动物志. 北京: 科学出版社.
杜灿谊, 杨翠丽, 潘威. 2012. 支持向量机在汽车自动变速器故障识别中的应用. 汽车工程, 34(3): 241–244.
郎宇宁, 蔺娟如. 2010. 基于支持向量机的多分类方法研究. 中国西部科技, 9(17): 28–29.
李晟, 何明一. 2012. 基于向量机的数字图像鉴别技术研究. 科学技术与工程, 12(10): 2323–2327.

- 秦传东, 刘三阳, 张市芳. 2012. 一种肿瘤基因的支持向量机提取方法. 西安电子科技大学学报: 自然科学版, 39 (1): 191 – 196.
- Brown M, Lewis H G and Gram S R. 2000. Linear spectral mixture models and SVM for remot sensing. *IEEE Trans Geo-science and Remote Sensing*, 38: 2346 – 2360.
- Clopper C J and Pearson E S. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26: 404 – 413.
- Cortes C and Vapnik V. 1995. Support vector networks. *Machine Learning*, 20: 273 – 297.
- Efe M A, Tavares E S, Baker A J and Bonatto S L. 2009. Multigene phylogeny and DNA barcoding indicate that the Sandwich tern complex (*Thalasseus sandwicensis*, Laridae, Sternini) comprises two species. *Molecular Phylogenetics and Evolution*, 52: 263 – 267.
- Elias M, Hill R I, Willmott K R, Dasmahapatra K K, Brower A V Z, Mallet J and Jiggins C D. 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 274: 2881 – 2889.
- Francis C M, Borisenko A V, Ivanova N V, Eger J L, Lim B K, Guillén-Servent A, Kruskop S V, Mackie L and Hebert P D N. 2010. The role of DNA barcodes in understanding and conservation of mammal diversity in southeast Asia. *PLoS ONE*, 5: e12575.
- Hajibabaei M, Janzen D H, Burns J M, Hallwachs M and Hebert P D N. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences USA*, 103: 968 – 971.
- Hebert P D N, Cywinska A, Ball S L and deWaard J R. 2003a. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 270: 313 – 322.
- Hebert P D N, Ratnasingham S and Waard J R D. 2003b. Barcoding animal life: cytochrome c oxidase subunit I divergences among closely related species. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 270: 596 – 599.
- Hickerson M, Meyer C P and Moritz C. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, 55: 729 – 739.
- Hubert N, Hanner R, Holm E, Mandrak N E, Taylor E, BurrIDGE M, Watkinson D, Dumont P, Curry A, Bentzen P, Zhang J, April J and Bernatchez L. 2008. Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE*, 3: e2490.
- Hurst G D D and Jiggins F M. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic, and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 272: 1525 – 1534.
- Jennings R M, Bucklin A and Pierrot B A. 2010. Barcoding of arrow worms (phylum Chaetognatha) from three oceans: genetic diversity and evolution within an enigmatic phylum. *PLoS ONE*, 5: e9949.
- Johnson S B, Waren A and Vrijenhoek R C. 2008. DNA barcoding of Lepetodrilus limpets reveals cryptic species. *Journal of Shellfish Research*, 27: 43 – 51.
- Kerr K C R, Birks S M, Kalyakin M V and Red'kin Y A. 2009. Filling the gap-COI barcode resolution in eastern Palearctic birds. *Frontiers in Zoology*, 6: 29.
- Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *Journal of Molecular Evolution*, 16: 111 – 120.
- Mailund T, Brodal G S, Fagerberg R, Pedersen C N S and Phillips D. 2006. Recrafting the neighbor-joining method. *BMC Bioinformatics*, 7: 29 – 36.
- Meyer C P and Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLOS Biology*, 3: 2229 – 2238.
- Mukherjee S. 2003. *Classifying Microarray Data Using Support Vector Machines. A Practical Approach to Microarray Data Analysis*. Boston: Kluwer Academic.
- Perruchet P and Peereman R. 2004. The exploitation of distributional information in syllable processing. *Journal of Neuro-linguistics*, 17: 97 – 119.
- Poole R W. 1989. *Noctuidae. Lepidopterorum Catalogues (New Series)*. New York: Leiden.
- Ratnasingham S and Hebert P D N. 2007. BOLD: The barcode of life data system (www.barcodinglife.org). *Molecular Ecology Notes*, 7: 355 – 364.
- Remigio E A and Hebert P D N. 2003. Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships. *Molecular Phylogenetics and Evolution*, 29: 641 – 647.

- Saitou N and Nei M. 1987. The neighbor-joining method; a new method for reconstructing phylogenetic tree. *Molecular Biology and Evolution*, 4: 406 – 425.
- Tamura K, Dudley J, Nei M and Kumar S. 2007. MEGA 4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24: 1596 – 1599.
- Tavares E S and Baker A J. 2008. Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. *BMC Evolutionary Biology*, 8: 81.
- Velez D R, White B C, Motsinger A A, Bush W S, Ritchie M D, Williams S M and Moore J H. 2007. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genetic Epidemiology*, 31: 306 – 315.
- Walavalkar L. 2003. Support vector learning for gender classification using audio and visual cues. *International Journal of Pattern Recognition and Artificial Intelligence*, 17: 417 – 439.
- Walker H K, Hall W D and Harst J W. 1990. *Clinical Method*. 3rd edition. Boston: Butterworths.
- Wiemers M and Fiedler K. 2007. Does the DNA barcoding gap exist? —A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, 4(8): 16.
- Will K W and Rubinoff D. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics*, 20: 47 – 55.
- Zahiri R, Holloway J D, Kitching I J, Kaila L, Lafontaine J D, Mutanen M and Wahlberg N. 2012. Molecular phylogenetics of Erebidæ (Lepidoptera, Noctuoidea). *Systematic Entomology*, 37: 102 – 124.
- Zahiri R, Kitching I J, Lafontaine J D, Mutanen M, Kaila L, Holloway J D and Wahlberg N. 2010. A new molecular phylogeny offers hope for a stable family level classification of the Noctuoidea (Lepidoptera). *Zoologica Scripta*, 40: 158 – 173.
- Zhao Q and Principe J C. 2001. SVM for SAR ATR. *IEEE Transactions on AES*, 37: 643 – 653.
- Zhang A B, Sikes D S, Muster C and Li S Q. 2008. Inferring species membership using DNA sequences with back-propagation neural networks. *Systematic Biology*, 57: 202 – 215.
- Zhou X, Adamowicz S J, Jacobus L M, DeWalt R E and Herbert P D N. 2009. Towards a comprehensive barcode library for arctic life —Ephemeroptera, Plecoptera, and Trichoptera of Church Manitoba, Canada. *Frontiers in Zoology*, 6: 30.

(责任编辑:杨郁霞)

