

中国储粮害虫 DNA 条形码鉴定系统研究

吴志刚¹ 李文欣¹ 赵紫华¹ 伍祎² 张涛² 曹阳² 李福君² 李志红^{1*}

(1. 中国农业大学 植物保护学院,北京 100193;

2. 国家粮食局科学研究院,北京 100037)

摘要 为进一步提升我国储粮害虫快速、准确鉴定能力,采用微软产品技术解决方案,应用 SQL SERVER 2008 数据库和 C# 编程语言,研制了中国储粮害虫 DNA 条形码鉴定系统 (Grain Pests DNA Barcode Identification System,GPDBIS)。GPDBIS 收集了我国将近 40 种储粮昆虫及仓储螨类的基础信息和 mtDNACO I 条形码信息,提供了基于 DNA 条形码的相似度比对鉴定功能和系统发育树分析鉴定功能、基础信息查询及管理功能。GPDBIS 初步应用效果显示,该系统界面友好、功能稳定,能够实现我国常见储粮害虫的快速、准确鉴定。

关键词 专家系统;储粮害虫;DNA 条形码;相似度比对;系统发育树

中图分类号 S433.89

文章编号 1007-4333(2017)05-0082-08

文献标志码 A

Development and primary application of the DNA barcode identification system of grain pests

WU Zhigang¹, LI Wenxin¹, ZHAO Zihua¹, WU Yi², ZHANG Tao²,
CAO Yang², LI Fujun², LI Zhihong^{1*}

(1. College of Plant Protection, China Agricultural University, Beijing 100193, China;

2. Academy of State Administration of Grain, Beijing 100037, China)

Abstract For rapid and accurate identification of stored-product pests, Grain Pests DNA Barcode Identification System (GPDBIS) was established by using SQL SERVER 2009 and C#. The basic information and mtDNACO I barcode information of about 40 species of stored-product pests and mites were collected in the GPDBIS system. This system provided similarity comparison, phylogenetic tree, information searching and management. The preliminary application showed that the GPDBIS system was interface friendly and function stably, which achieved rapid and correct identification of stored-product pests. GPDBIS developed in this study could further enhance the ability of rapid and correct identification of stored-product pests, improved the monitoring and management of stored-product pests. It could also facilitate the international trade of agricultural goods and grain safety of China.

Keywords expert system; stored-product pest; DNA barcoding; similarity comparison; phylogenetic tree

储粮害虫是重要的经济害虫,包括昆虫和螨类两大类。根据中国粮食部门的多次调查,共采集储粮昆虫 200 多种,其中储粮害虫有 100 多种^[1],粮食的储藏过程中,受虫蚀、霉变、鼠害等造成的损失较为严重,其中储粮害虫的危害是粮食储存中较为严重的问题之一^[2]。要保护储粮免受储粮害虫的危害或降低危害程度,首先要了解存在哪些储粮害虫,清

楚它们的名称、分类地位、生物学特性,以便采用恰当的防治方法来进行防治。目前,害虫的分类鉴定应用较多的有形态学鉴定和分子生物学鉴定。形态特征受到周围环境和发育时期等的影响,存在一定局限性^[3]。近年来,随着 DNA 检测和分析技术的快速发展,越来越多的分子生物学技术应用于昆虫分类鉴定中来,分子鉴定具有不受样品性别和发育

收稿日期:2016-07-25

基金项目:国家国际科技合作专项项目(2013DFG32350)

第一作者:吴志刚,博士,主要从事有害生物风险分析及专家系统研究,E-mail:wuzhigang@cau.edu.cn

通讯作者:李志红,教授,主要从事植物检疫与入侵生物学相关研究,E-mail:lizh@cau.edu.cn

阶段限制,对缺失形态鉴定特征的样本残体仍能适用的优点,可实现储粮害虫物种的有效鉴定,对于辅助非专业人员进行经济、快速地识别物种具有重要意义。中国农业大学和国家粮食局科学研究院近年来针对基于 DNA 条形码技术的分子鉴定技术进行了探索与尝试,并取得了一定的研究成果^[4-10]。

目前我国基于 DNA 条形码技术的储粮害虫鉴定系统还未见报道,本研究拟通过建立中国储粮害虫 DNA 条形码鉴定系统 (Grain Pests DNA Barcode Identification System, GPDBIS) 能够在网络环境下实现对我国储粮害虫及其天敌的 DNA 条形码分子鉴定功能、基本信息浏览及查询功能,为我

国储粮害虫 DNA 条形码数据管理及快速鉴定提供服务平台,为储粮害虫防治人员提供技术支持。

1 GPDBIS 的设计与实现

1.1 系统结构设计

GPDBIS 系统是一个针对储粮害虫进行信息浏览、查询和采用 DNA 条形码进行鉴定与管理的网络专家系统。根据用户使用该系统的目的不同,划分为客户端用户和服务器端用户。客户端用户的目的是访问系统获取信息,而服务器端用户访问系统的目的是管理维护系统数据,因此两者之间具有数据层的联系,但功能层又相互分开,具体结构如图 1:

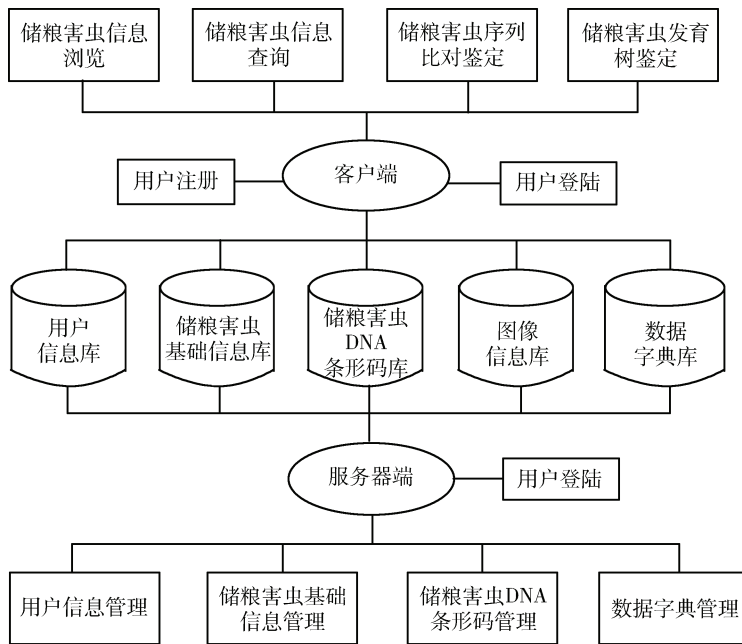


图 1 GPDBIS 系统总体结构设计

Fig. 1 The total design of GPDBIS system

1.2 系统功能设计

GPDBIS 系统把我国常见的储粮害虫条形码和对应的物种信息都录入到系统数据库中,作为本系统的储粮害虫 DNA 条形码数据库,当对某新的 DNA 条形码进行鉴定时,只需把新的条形码和本系统中已经录入数据库的 DNA 条形码进行序列比对或者统一构建发育树,可实时、快速的鉴定出该新的条形码是什么物种。还可采用构建系统发育树进行鉴定。点击鉴定结果物种名称的链接,可浏览物种的详细信息,包括中文名、英文名、异名、分类地位、地理分布、形态特征、危害特点、防治技术和图片等

物种基本信息。另外,GPDBIS 系统还提供了物种的信息查询功能,可快速浏览检索数据库中已录入的物种信息。具有管理权限的用户,可在后台管理页面对所有用户信息、物种信息、条形码信息等进行管理。

1.3 数据库设计

根据 GPDBIS 系统的功能设计,按照系统数据库结构的设计规范和原则,设计 GPDBIS 系统的数据库结构包括用户信息表、储粮害虫基础信息表、储粮害虫条形码表、数据字典表、公告表、系统日志表等。表 1 为储粮害虫基础信息表。

表1 GPDBIS 储粮害虫基础信息表

Table 1 Stored-product pests basic information table of GPDBIS

列名 Column	数据类型 Type	可以为空 Null Or Not Null	说明 Remarks
Pest ID	int	否	主键
Cn Name	varchar(100)	否	中文名
En Name	varchar(200)	是	英文名
Scientific Name	varchar(200)	是	学名
Peck Name	varchar(100)	是	定名人
Synonym Name	varchar(4000)	是	异名-作废
Class ID	int	是	分类地位 ID
Host Plant	varchar(1000)	是	寄主植物
Geography Distribute	varchar(8000)	是	地理分布
Form Characteristic	varchar(8000)	是	形态特征
Harm Feature	varchar(8000)	是	危害特点
Control Tech	varchar(8000)	是	防治技术
Info Source	varchar(8000)	是	信息来源
Demo	varchar(8000)	是	备注
Demo 1	varchar(8000)	是	备注 1
Demo 2	varchar(8000)	是	备注 2
Demo 3	varchar(8000)	是	备注 3
Demo 4	varchar(8000)	是	备注 4
Demo 5	varchar(8000)	是	备注 5
Input Time	datetime	是	入库时间
Adm ID	int	是	操作人 ID

注:GPDBIS,中国储粮害虫 DNA 条形码鉴定系统。

Note:GPDBIS,Grain Pests DNA Barcode Identification System.

1.4 系统开发

根据系统需求分析和系统设计,GPDBIS 系统是基于 B/S 结构,采用 ASP. Net 技术,基于 .NET Framework 2.0 类库,利用 C# 语言编程,使用 Microsoft Visual Studio 2010 工具进行开发。同时,GPDBIS 的发育树鉴定功能是基于 PHYLIP 的 3.69 版本开源代码,采用 C++ 编程技术,进行二次开发和接口封装,形成可被储粮害虫 DNA 条形

码鉴定系统前端调用的 DLL 动态链接库,使 GPDBIS 系统高效、可靠的实现发育树构建。

服务器操作系统为 Windows Server 2003,可支持 Windows Server 2003 或更高版本操作系统部署,如 Windows Server 2008 等。数据库系统采用了 Microsoft SQL Server 2008 作为数据库,存储储粮害虫的文本数据、DNA 条形码数据以及用户数据等(图 2)。



图 2 GPDBIS 首页界面

Fig. 2 The homepage interface of GPDBIS

2 推理机设计

推理机是专家系统的“思维”机构,是构成专家系统的核心部分,其任务是模拟专家的思维过程,控制并执行对问题的求解。它能根据当前已知的事实,利用知识库中的知识,按一定的推理方法和控制策略进行推理,求得问题的答案或证明某个假设的正确性^[11]。该系统的推理机主要涉及 DNA 序列比对鉴定的算法模型和系统发育树鉴定的发育树生成设计。

2.1 相似性比对

相似性是比较两者之间的相同程度。该系统中对于待鉴定 DNA 序列首先采用了相似性比对的方法,检测其与数据库内样本的相似程度,分别计算得分(或代价)函数,然后挑选一个得分最高(或代价最小)的比对作为最终结果,并使用%来定量标识,从而初步确定鉴定结果。2 条序列可能的比对数非常多,是序列长度的指数函数,随着序列长度的增长,计算量呈指数增长。为了进行高效的计算,找出最优的比对,系统调用的比对算法采用了动态规划算法(Dynamic programming),该算法把一个问题分解成计算量合理的子问题,并使用这些子问题的结果来计算最终答案。这个算法是生物信息学的基本算法之一^[12]。

2.2 系统发育树生成

为了弥补相似性比对的模糊结果的可靠性,GPDBIS 系统还利用系统发育树构建的功能,进一步提高未知样本 DNA 序列鉴定的准确性。系统发

育树鉴定的实现思路是将待鉴定样本 DNA 序列输入系统,生成比对文件,调用多序列比对 Web 服务,借用第三方的发育树构建服务和工具,绘制生成系统发育树图示(图 3)。

系统树的构建主要有 3 种方法:距离矩阵法(Distance matrix method)是根据每对物种之间的距离,其计算一般很直接,所生成的树的质量取决于距离尺度的质量。距离通常取决于遗传模型。最大简约(Maximum parsimony)法较少涉及遗传假设,它通过寻求物种间最小的变更数来完成的。对于模型的巨大依赖性是最小似然(Maximum likelihood)法的特征,该方法在计算上繁杂,但为统计推断提供了基础^[13-14]。该系统采用了距离矩阵法进行系统发育树的构建。

在测算遗传距离矩阵时,一般有平均连接聚类法(UPGMA 法)、Fitch-Margoliash 算法和邻接法(Neighbor-joining Method)。该系统调用的算法模型 Web 服务是采用了邻接法。邻接法是 1987 年 Nei 等提出的,该方法通过确定距离最近(或相邻)的成对分类单位来使系统树的总距离达到最小^[15]。其一般步骤如下:

1) 计算第 i 终端节点的净分歧度 r_i

$$r_i = \sum_{k=1}^N d_{ik} \quad (1)$$

式中: N 为终端节点数, d_{ik} 为节点 i 和节点 k 之间的距离,有 $d_{ik} = d_{ki}$

2) 计算并确定最小速率校正距离(rate-

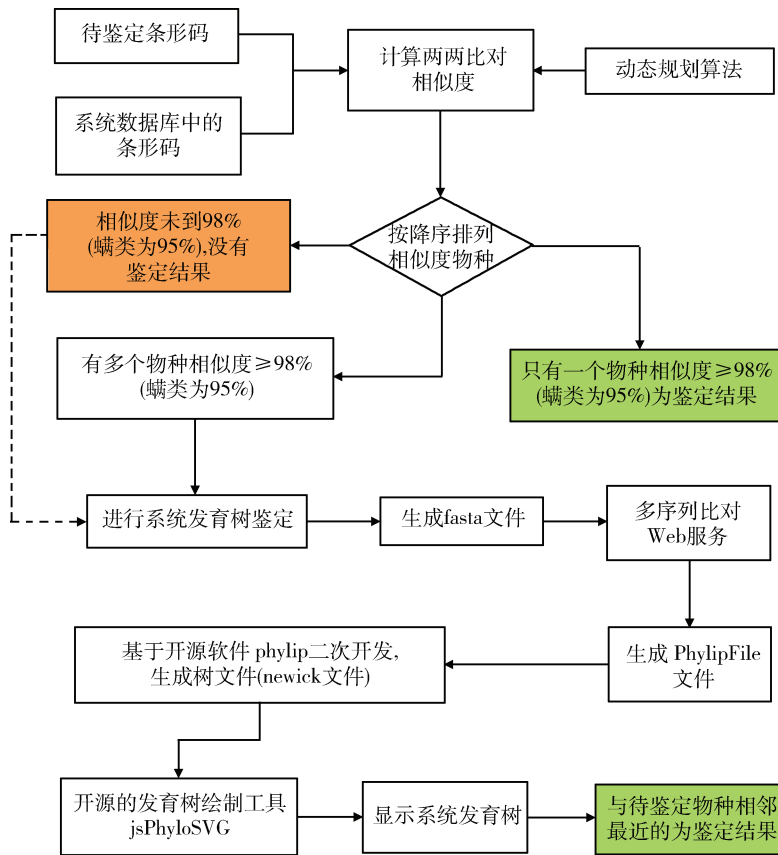


图3 GPDBIS 发育树鉴定的实现思路

Fig. 3 The identification of phylogenetic tree of GPDBIS

corrected distance) M_{ij} :

$$M_{ij} = d_{ij} - \frac{r_i + r_j}{N - 2} \quad (2)$$

3) 定义一个新节点 u , u 节点由节点 i 和 j 组合而成。节点 u 与节点 i 和 j 的距离为:

$$S_{iu} = \frac{d_{ij}}{2} + \frac{r_i + r_j}{2(N - 2)}$$

$$S_{ju} = d_{ij} - S_{ij} \quad (3)$$

式中: 节点 u 与系统树其他节点 k 的距离为:

$$d_{ku} = \frac{d_{ik} + d_{jk} - d_{ij}}{2} \quad (4)$$

4) 从距离矩阵中删除列节点 i 和 j 的距离, N 值(总节点数)减去 1

5) 如果尚余 2 个以上终端节点, 返回到步骤(1)继续计算, 直至系统树完全建成^[13]。

3 系统初步应用

经过测试, 系统在局域网部署后进行了初步应

用, 各个功能运行正常, 系统运行稳定。截至 2016 年 2 月, GPDBIS 收集的储粮昆虫和仓储螨类基础信息涉及 38 种, 来源于《储藏物害虫与防治》和《中国储粮昆虫图鉴》等专业书籍及文献的数字化处理, 内容包括中文名、英文名、学名、异名、分类地位、地理分布、形态特征、危害特点、防治技术; GPDBIS 收集的储粮害虫及仓储螨类的 mtDNACO I 条形码序列信息, 涉及 38 种、250 条序列, 全部序列由中国农业大学与国家粮食科学研究院检测、提供(表 2)。

为了验证系统的可靠性, 系统进行了实际案例测试应用。分别对采自于山东、广东等粮库的害虫样品进行了 DNA 条形码序列检测, 并采用 GPDBIS 的相似度比对和构建发育树功能进行鉴定, 鉴定结果准确、可靠(部分检测结果见图 4~6)。希望本系统在储粮害虫基本信息及 DNA 条形码序列信息进一步丰富后, 一定范围内开放对外访问服务, 发挥更大作用。

表2 GPDBIS 储粮害虫鉴定系统入库条形码种类

Table 2 Stored DNA codes of GPDBIS

序号 No.	种类 Class	样品来源 Sample source	DNA 条形码 数量/条 Number of DNA code (fragments)
1	麦蛾 <i>Sitotroga cerealella</i>	广西、江西、重庆	5
2	印度谷螟 <i>Plodia interpunctella</i>	新疆	5
3	欧洲谷蛾 <i>Nemapogon granella</i>	北京	5
4	杂拟谷盗 <i>Tribolium confusum</i>	广东	5
5	赤拟谷盗 <i>Tribolium castaneum</i>	广西、重庆、广东、福建、海南、云南	12
6	锯谷盗 <i>Oryzaephilus surinamensis</i>	广西、重庆、海南、新疆	12
7	米扁虫 <i>Ahasverus advena</i>	海南	5
8	锈赤扁谷盗 <i>Cryptolestes ferrugineus</i>	广西、江西、重庆、广东、海南、云南、辽宁、河南、福建、广东、山东、捷克、美国	26
9	长角扁谷盗 <i>Cryptolestes pusillus</i>	广东、海南、湖北、捷克、美国	9
10	土耳其扁谷盗 <i>Cryptolestes turcicus</i>	广东、河南、捷克	5
11	微扁谷盗 <i>Cryptolestes pusilloides</i>	云南	5
12	谷蠹 <i>Rhyzopertha dominica</i>	四川、江西、广西、重庆、海南	11
13	花斑皮蠹 <i>Trogoderma variabile</i>	北京	5
14	烟草甲 <i>Lasioderma serricorne</i>	北京、云南	5
15	米象 <i>Sitophilus oryzae</i>	澳大利亚、四川、江西、重庆	11
16	玉米象 <i>Sitophilus zeamais</i>	广东、广西	5
17	脊胸露尾甲 <i>Carpophilus dimidiatus</i>	海南	6
18	绿豆象 <i>Callosobruchus chinensis</i>	北京	5
19	米象金小蜂 <i>Lariophagus distinguendus</i>	广西	5
20	嗜虫书虱 <i>Liposcelis entomophila</i>	北京、重庆、广西、云南、山东、湖北、河北、克罗地亚、捷克、美国	11
21	嗜卷书虱 <i>Liposcelis bostrychophila</i>	重庆、北京、广西、广东、河南、捷克	7
22	无色书虱 <i>Liposcelis decolor</i>	重庆、云南、捷克、美国	5
23	小眼书虱 <i>Liposcelis paeta</i>	山东、贵州、河南、河北、湖北、浙江、捷克、克罗地亚、美国	10
24	啮书虱 <i>Liposcelis corrodens</i>	捷克、克罗地亚、美国、葡萄牙	5
25	暗褐书虱 <i>Liposcelis brunnea</i>	捷克、美国	5
26	虚伪书虱 <i>Liposcelis mendax</i>	江苏	5
27	皮氏书虱 <i>Liposcelis pearmani</i>	美国	5
28	三色书虱 <i>Liposcelis tricolor</i>	山东	5
29	红书虱 <i>Liposcelis rufa</i>	美国	5
30	线嗜酪螨 <i>Tyroborus lini</i>	捷克	5
31	椭圆食粉螨 <i>Aleuroglyphus ovatus</i>	捷克	5
32	拱殖嗜渣螨 <i>Chortoglyphus arcuatus</i>	捷克	5
33	家食甜螨 <i>Glycyphagus domesticus</i>	捷克	5
34	马六甲肉食螨 <i>Cheyletus malaccensis</i>	海南	10
35	普通肉食螨 <i>Cheyletus eruditus</i>	捷克	5
36	美洲室尘螨 <i>Dermatophagoides farinae</i>	捷克	5
37	褐蕈甲 <i>Cryptophilus integer</i>	山东	5
38	亚扁粉盗 <i>Palorus subdepressus</i>	中国、美国	5



图 4 GPDBIS 序列比对鉴定结果界面

Fig. 4 The identification result interface based on DNA barcode comparison in GPDBIS



图 5 GPDBIS 序列比对鉴定界面

Fig. 5 The identification interface based on DNA barcode comparison in GPDBIS

参考文献 References

- [1] 沈兆鹏. 拉汉英对照中国 100 种重要储粮害虫名录[J]. 粮食科技与经济, 2006, 31(6): 35-38
Shen Z P. Pull the Chinese list of 100 important pests Chinese [J]. *Grain Science and Technology and Economy*, 2006, 31(6): 35-38 (in Chinese)
- [2] 胡丽华, 郭敏, 张景虎, 任娜. 储粮害虫检测新技术及应用现状 [J]. 农业工程学报, 2007, 23(11): 286-290
- [3] Sole C L, Scholtz C H. Did dung beetles arise in Africa? A phylogenetic hypothesis based on five gene regions [J]. *Molecular Phylogenetics & Evolution*, 2010, 56(2): 631-641
- [4] Li Z H, Kučerová Z, Zhao S, Stejskal V, Opit G, Qin M. Morphological and molecular identification of three geographical populations of the storage pest *Liposcelis bostrychophila*, (Psocoptera) [J]. *Journal of Stored Products Research*, 2011, 47(47): 168-172
- [5] Yang Q, Kučerová Z, Li Z, Kalinovič I, Stejskal V. Diagnosis of *Liposcelis entomophila*, (Insecta: Psocodea: Liposcelididae) based on morphological characteristics and DNA barcodes [C].

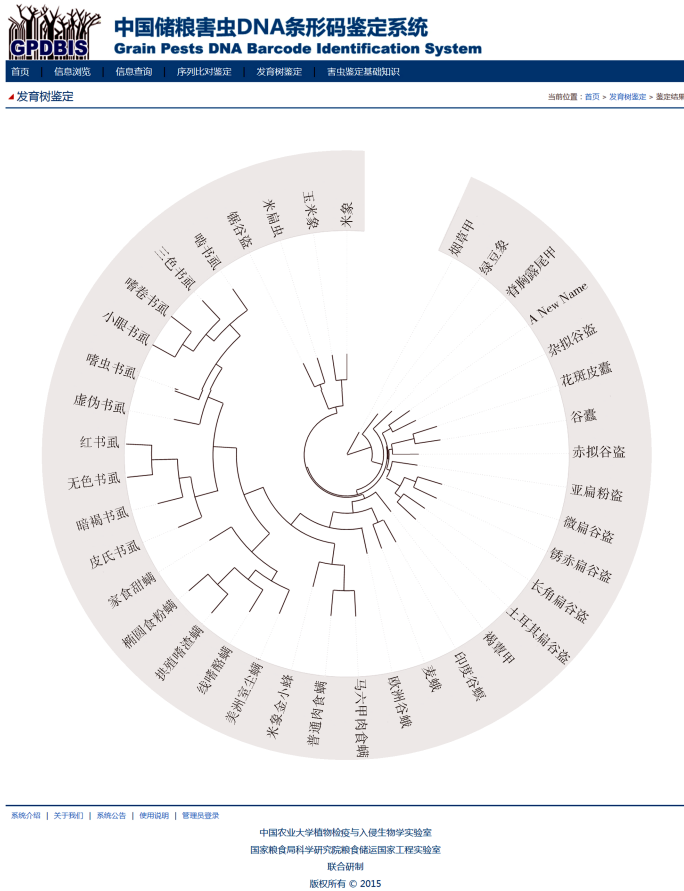


图 6 GPDBIS 发育树鉴定结果界面

Fig. 6 The identification result interface based on phylogenetic tree in GPDBIS

In: INTERSPEECH 2008, Conference of the International Speech Communication Association, Brisbane, Australia, 2012; 120-125

[6] 杨倩倩,李志红,伍祎,柳丽君. 线粒体 COI 基因在昆虫 DNA 条形码中的研究与应用[J]. 应用昆虫学报, 2012, 49(6): 1687-1695
Yang Q Q, Li Z H, Wu Y, Liu L J. Advance and application of mtDNA COI barcodes on insects [J]. *Chinese Journal of Applied Entomology*, 2012, 49(6): 1687-1695 (in Chinese)

[7] Yang Q, Zhao S, Kučerová Z, Stejskal V, Opit G, Qin M, Cao Y, Li F J, Li Z H. Validation of the 16S rDNA and COI DNA barcoding technique for rapid molecular identification of Stored product psocids (Insecta; Psocodea; Liposcelididae) [J]. *Journal of Economic Entomology*, 2013, 106(1): 419-425(7)

[8] Yang Q, Zhao S, Kučerová Z, Opit G, Cao Y, Stejskal V. Rapid molecular diagnosis of the stored-product psocid *Liposcelis corrodens*, (Psocodea; Liposcelididae): Species-specific PCR primers of 16S rDNA and COI[J]. *Journal of Stored Products Research*, 2013, 54(3): 1-7

[9] Wang Y J, Li Z H, Zhang S F, Varadinová Z, Jiang F, Kučerová Z, Stejskal V, Opit G, Cao Y, Li F J. DNA barcoding of five common stored-product pest species of genus *Cryptolestes* (Coleoptera; Laemophloeidae) [J]. *Bulletin of Entomological*

Research, 2014, 104(5): 1-8

[10] Zhang T, Wang Y J, Guo W, Luo D, Wu Y, Kučerová Z, Stejskal V, Opit G, Cao Y, Li F J, Li Z H. DNA barcoding, species-specific PCR and real-time PCR techniques for the identification of six *Tribolium* pests of stored products [J]. *Scientific Reports*, 2016, 6: 1-11

[11] Giarratano J C, Riley G D. *Expert Systems: Principles and Programming* [M]. New York: Pacific Grove, 1989

[12] 孙啸, 陆祖宏, 谢建明. 生物信息学基础 [M]. 北京: 清华大学出版社, 2005: 69-70
Sun X, Lu Z H, Xie J M. *Basics for Bioinformatics* [M]. Beijing: Tsinghua University Press, 2005: 69-70 (in Chinese)

[13] 樊龙江. 生物信息学札记 [BE/OL]. 3 版. Available From <http://ibi.zju.edu>, 2010
Fan L J. *Bioinformatics notes* [BE/OL]. 3rd. Available From <http://ibi.zju.edu>, 2010 (in Chinese)

[14] Weir B S. 遗传学数据分析 [M]. 北京: 中国农业出版社, 1996
Weir B S. *Genetic Data Analysis* [M]. Beijing: China Agriculture Press, 1996 (in Chinese)

[15] Nei M. *Molecular Evolutionary Genetics* [M]. New York: Columbia University Press, 1987: 150-71

责任编辑: 王燕华