

千斤拔属药用植物 DNA 条形码鉴定研究

张忠廉¹, 宋美芳¹, 李海涛¹, 管燕红¹, 牛迎风¹, 马小军^{2*}, 张丽霞^{1*}

1. 中国医学科学院北京协和医学院 药用植物研究所云南分所, 西双版纳州傣药南药重点实验室, 云南 景洪 666100

2. 中国医学科学院北京协和医学院 药用植物研究所, 北京 100193

摘要: 目的 筛选一种可准确、高效鉴别千斤拔属 *Flemingia* Roxb. ex Ait. et Ait. f. 药用植物的 DNA 条形码序列。方法 对 4 种 14 份千斤拔属药用植物的 ITS2、*rbcL*、*psbA-trnH* 序列进行扩增和测序, 同时查找 NCBI 数据库中千斤拔属植物的相应序列, 共计 7 种 20 份, 比较不同序列的扩增效率及测序成功率, 并对其进行种内、种间变异分析, barcoding gap 检验及 NJ 聚类图分析, 以评估不同序列对千斤拔属植物的物种鉴别能力。结果 测序成功率方面, ITS2 与 *rbcL* 序列对所分析样品的测序成功率为 100%, *psbA-trnH* 的测序成功率为 85.71%; 3 条序列中, 只有 ITS2 在 barcoding gap 检验中具有显著 gap; 从 NJ 聚类图来看, ITS2 可明显区分千斤拔属植物的不同物种 (除 *F. philippinensis* 与 *F. stricta* 外)。结论 ITS2 序列能准确鉴别千斤拔属药用植物, 可作为千斤拔药材基原植物鉴定的条形码序列。

关键词: 千斤拔属; DNA 条形码; ITS2; *rbcL*; *psbA-trnH*

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2015)01-0118-05

DOI: 10.7501/j.issn.0253-2670.2015.01.023

Identification of DNA barcoding on medicinal plants in *Flemingia* Roxb. ex Ait. et Ait. f.

ZHANG Zhong-lian¹, SONG Mei-fang¹, LI Hai-tao¹, GUAN Yan-hong¹, NIU Ying-feng¹, MA Xiao-jun², ZHANG Li-xia¹

1. Key Laboratory of Dai and Southern Medicine of Xishuangbanna Dai Autonomous Prefecture, Yunnan Branch of Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Jinghong 666100, China

2. Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China

Abstract: Objective To screen a DNA barcoding sequence that can identify the medicinal plants of *Flemingia* Roxb. ex Ait. et Ait. f. accurately and efficiently. **Methods** Four species and 14 individuals of *Flemingia* Roxb. ex Ait. et Ait. f. were collected, the (second internal transcribed spacer, ITS2) of ribosomal DNA, *rbcL*, *psbA-trnH* of chloroplast DNA were amplified and sequenced. Meanwhile, the NCBI data were retrieved and the according sequence was downloaded. The total numbers of species and individuals were six and twenty. The PCR amplification and sequencing efficiency, barcoding gap, and NJ trees were used to evaluate the efficiency of species identification. **Results** The sequencing success rates of ITS2, *rbcL*, and *psbA-trnH* were 100%, 100%, and 85.71%, respectively; Among the three DNA barcoding sequences, only ITS2 has remarkable barcoding gap; ITS2 could distinguish every species of *Flemingia* Roxb. ex Ait. et Ait. f. (except *F. philippinensis* and *F. stricta*). **Conclusion** ITS2 could identify the medicinal plants of *Flemingia* Roxb. ex Ait. et Ait. f. accurately and efficiently, and could be used as an ideal DNA barcoding of species identification for medicinal materials of *Flemingia* Roxb. ex Ait. et Ait. f.

Key words: *Flemingia* Roxb. ex Ait. et Ait. f.; DNA barcoding; ITS2; *rbcL*; *psbA-trnH*

收稿日期: 2014-08-11

基金项目: 云南省应用基础研究计划 (2011FZ301); 中国医学科学院药用植物研究所创新团队研究计划: 热带名贵药用植物种质创新与繁育 (121306)

作者简介: 张忠廉 (1982—), 男, 助理研究员, 研究方向为热带药用植物资源及分子生物学。Tel: 15924692393 E-mail: zzl0605@163.com

*通信作者 张丽霞, 副研究员。E-mail: 87050233@qq.com

马小军, 研究员。E-mail: mayixuan10@163.com

千斤拔 *Flemingiae Philippinensis Radix* 始载于《植物名实图考》，具有祛风除湿、舒筋活络、强筋壮骨、消炎止痛等功效，适用于风湿痹痛、腰肌劳损、气虚脚肿、肺虚久咳、白带过多、月经不调、跌打损伤等症^[1]，被广泛用于治疗妇科病、风湿痹痛等类型的中成药生产中；近年来又发现因其富含植物雌激素而在治疗妇女更年期综合征方面有很好的疗效。据统计，我国共有 100 多个中成药生产企业使用千斤拔原料生产中成药，年需求量超过 1 000 t，且有逐年上升趋势^[2]。由于目前千斤拔药材资源以野生资源为主，长期的肆意采挖使得千斤拔野生资源日益匮乏，药材市场供不应求现象日益严重。《中国药典》2010 年版^[3]所收载千斤拔药材基原植物有 3 种，分别为豆科 (*Flemingia*) 植物千斤拔 *Flemingia philippinensis* Merr. et Rolfe、大叶千斤拔 *F. macrophylla* (Willd.) Prain 或锈毛千斤拔 *F. ferruginea* Wall. ex Benth.。然而，千斤拔属 *Flemingia* Roxb. ex Ait. et Ait. f. 植物在我国共分布有 16 种及 1 变种，其中约 10 种可作药用^[1]，随着千斤拔野生资源的日益匮乏，药材市场以同属植物替代千斤拔正品药材的混用、代用现象日益严重，药材品质不一，其质量保障及用药安全问题令人堪忧。又因千斤拔药用部位为根，仅依靠传统的显微鉴别、理化鉴别等方法无法准确鉴别千斤拔药材的基原植物物种，化学指纹图谱等方法又因其操作繁琐，难以普及应用。因此，寻找一种快速、准确的千斤拔药材鉴定方法对于千斤拔药材的质量控制及安全使用意义重大。

DNA 条形码 (DNA barcoding) 是利用基因组中一段公认的标准短序列进行物种鉴定的分子诊断新技术^[4-6]，该技术自 2003 年由加拿大动物学家 Hebert^[7]首次提出后，现已发展成为生物分类学研究的热门方向 and 前沿技术，通过建立生物 DNA 条形码系统能够实现快速、准确和自动化的物种鉴定。10 多年间，研究者们曾提出过 10 余条植物候选 DNA 条形码序列^[8]，并在不同的科属种群中对各候选条形码序列的鉴定能力进行了考察，但到目前为止，研究者们仍未达成一致意见。2009 年，Consortium for the Barcode of Life (CBOL) 植物工作组提出使用“*rbcL*+*matK*”复合序列作为植物界的通用条形码序列^[9]，但由于扩增效率、鉴定成功率等问题，其在植物界的通用性受到质疑。2013 年，陈士林课题组建立“以 ITS2 为核心，*psbA-trnH* 为辅”的植物类中药材 DNA 条形码鉴定体系^[10]，并

建议将该方法纳入中国药典。

本研究对近几年研究较热门的 DNA 条形码候选序列 (如 ITS2、*psbA-trnH* 及 *rbcL*) 在千斤拔属植物间的鉴别能力进行了比较分析，考察不同 DNA 条形码候选序列在千斤拔属药用植物鉴定中的能力，以期建立一种千斤拔属药用植物的高效鉴定方法。

1 材料与方法

1.1 材料

共收集千斤拔属实验材料 4 种 14 份，均来自中国医学科学院药用植物研究所云南分所南药园内，经本单位张丽霞副研究员鉴定为千斤拔 *Flemingia philippinensis* Merr. et Rolfe、大叶千斤拔 *F. macrophylla* (Willd.) Prain、长叶千斤拔 *F. stricta* Roxb. ex Ait.、宽叶千斤拔 *F. latifolia* Benth.，凭证标本保存于中国医学科学院药用植物研究所云南分所标本馆内。实验材料见表 1。此外，为了在更为广泛的范围内考察千斤拔属药用植物的 DNA 条形码鉴别能力，从 GenBank 数据库中获得千斤拔属相关条形码序列用于物种的鉴定效率分析，所引用序列的序列号见表 2。

表 1 材料信息及其编号

Table 1 Specimen information and their code numbers

编号	植物名	编号	植物名
1	宽叶千斤拔	8	长叶千斤拔
2	宽叶千斤拔	9	大叶千斤拔
3	宽叶千斤拔	10	千斤拔
4	宽叶千斤拔	11	千斤拔
5	宽叶千斤拔	12	千斤拔
6	宽叶千斤拔	13	千斤拔
7	长叶千斤拔	14	大叶千斤拔

表 2 NCBI 数据库中的千斤拔属植物条形码序列

Table 2 Barcoding sequences of plants in *Flemingia* Roxb. ex Ait. et Ait. f. from NCBI database

序列	序列号	种名
ITS2	JX546292	大叶千斤拔
	FJ980288	大叶千斤拔
	JX546290	球穗千斤拔
	FJ980289	腺毛千斤拔
	GQ434347	腺毛千斤拔
<i>rbcL</i>	JX546291	千斤拔
	KF621123	大叶千斤拔
	GQ436332	大叶千斤拔
	GQ436333	腺毛千斤拔
<i>psbA-trnH</i>	GQ434959	腺毛千斤拔
	GQ434958	大叶千斤拔

1.2 仪器与试剂

DNA 提取试剂盒, Tiangen (DP305); 引物, 上海生工生物技术有限公司合成; rTaq DNA 聚合酶 (大连宝生物工程有限公司, DR001B), 包括: TaKaRa TaqDNA 聚合酶 (5 U/ μ L), 10 $\times$ 缓冲液 (Mg^{2+} free), dNTP mixture (各 2.5 mmol/L), $MgCl_2$ (25 mmol/L); Marker 购于北京鼎国生物技术有限公司 (B031); 其他试剂均为分析纯。

扩增仪 (Biometra Professional Standard), 电泳仪 (Biometra 公司), 凝胶成像系统 (Uvitec 公司)。

1.3 样品 DNA 的提取、PCR 扩增和测序

采集样品新鲜嫩叶并用水及 70%乙醇擦洗表面后, 取样约 0.1 g, 液氮充分研磨并用 DNA 提取试剂盒 (DP305, 天根生化) 提取基因组 DNA。继而用特定引物扩增相应的序列片段, PCR 反应体系为 20 μ L, 具体体系为 ddH₂O, 1 $\times$ PCR 缓冲液 (含 1.5 mmol/L $MgCl_2$), dNTP (0.2 mmol/L), 正向、反向引物 (各 0.2 μ mol/L), Taq DNA 聚合酶 (1.5 U), DNA (10~30 ng)。所用引物序列及扩增体系如表 3 所示。特定条形码序列扩增成功后经 PCR 纯化试剂

表 3 所用引物序列及相应 PCR 反应条件

Table 3 Sequences of used primers and corresponding PCR reaction conditions

引物名称	引物序列 (5'-3')	PCR 反应条件
ITS2	F: ATGCGATACTTGGTGTGAAT	94 $^{\circ}$ C, 5 min
	R: GACGCTTCTCCAGACTACAAT	94 $^{\circ}$ C、45 s, 55 $^{\circ}$ C、45 s, 72 $^{\circ}$ C、1 min, 35 个循环 72 $^{\circ}$ C、10 min
<i>rbcL</i>	F: ATGTCACCACAAACAGAAAC	94 $^{\circ}$ C、5 min
	R: TCGCATGTACCTGCAGTAGC	94 $^{\circ}$ C、45 s, 55 $^{\circ}$ C、45 s, 72 $^{\circ}$ C、1 min, 35 个循环 72 $^{\circ}$ C、10 min
<i>psbA-trnH</i>	F: GTTATGCATGAACGTAATGCTC	94 $^{\circ}$ C、5 min
	R: CGCGCATGGTGGATTCACAATCC	94 $^{\circ}$ C、45 s, 55 $^{\circ}$ C、45 s, 72 $^{\circ}$ C、1 min, 35 个循环 72 $^{\circ}$ C、10 min

盒 (AXYGEN, 美国) 纯化, 使用 ABI3730XL 测序仪 (Applied Biosystems Co., 美国) 进行双向测序。

1.4 数据处理

测序峰图采用序列拼接软件 CodonCode Aligner V 2.06 (CodonCode Co., 美国) 校对, 去除低质量序列及引物区, 后用 contig 软件拼接及 ClustalX V2.0 (Higgins D. G.) 进行多序列比对并查错。对于 ITS2 序列, 拼接得到一致序列后使用基于隐马尔可夫模型的 HMMer 注释方法, 去除两端 5.8 S 和 26 S 区段获得 ITS2 间隔区序列^[11]; 比对数据经 MEGA v 4.1^[12]软件处理计算 K2P 遗传距离, 并采用 NJ (neighboring joining) 法构建系统进化树; 为评价各候选序列的物种鉴别能力, 利用 Taxon DNA^[13]统计不同序列种间、种内遗传距离的分布频度, 评价不同 DNA 条形码候选序列的 barcoding gap。

2 结果与分析

2.1 序列信息、PCR 扩增效率和测序成功率

所扩增 3 条 DNA 条形码序列长度及 GC 统计结果见表 4, 其中 ITS2 序列长度最短, 为 210 bp, *rbcL* 序列最长, 达 660 bp; 各序列 GC 量也不同,

表 4 不同条形码序列的序列信息及测序成功率

Table 4 Sequence information and sequencing success rates of different barcodings

条形码 序列	平均序列 长度/bp	GC / %	PCR 扩增 效率/%	测序成 功率/%	信息位点数/ 变异位点数
ITS2	210	58.9	100	100	36/43
<i>rbcL</i>	660	42.7	100	100	5/385
<i>psbA-trnH</i>	336	38.2	100	85.71	148/270

ITS2 的 GC 量最高, 为 58.9%, *psbA-trnH* 的 GC 量最低, 仅有 38.2%。PCR 扩增效率及测序成功率是评价 DNA 条形码序列的一个重要指标, 所分析 3 条 DNA 条形码序列在所分析样品中的扩增成功率均为 100%; 在测序成功率方面, ITS2 与 *rbcL* 序列的扩增成功率为 100%, 而 *psbA-trnH* 序列的扩增成功率为 85.71%。变异位点及信息位点数可反映所分析材料的碱基变异程度, 统计 3 条候选序列的变异位点及信息位点可知, *rbcL* 的变异位点最多 (385 个), 但其中信息位点仅为 5 个, 所提供的物种鉴别信息极为有限; 对于 *psbA-trnH* 序列, 其变异位点及信息位点数均比较多 (148/270), 说明其种间变

异及种内变异均比较丰富, 一个理想的 DNA 条形码序列, 其种间变异应明显大于种内变异, 根据 *psbA-trnH* 序列的变异位点数及信息位点数, 初步判断 *psbA-trnH* 序列对千斤拔属的物种鉴定能力有限; 而对于 ITS2, 其变异位点共 43 个, 说明 ITS2 在千斤拔属植物间碱基变异较少, 相对保守, 而其信息位点却可达到 36 个, 说明其在千斤拔属植物的物种鉴定方面可提供较多信息。综合比较序列长度、GC 量及序列扩增、测序成功率, 再结合序列变异位点及信息位点数, 初步认为, 在所分析 3 条序列间, ITS2 序列为最合适条形码序列。

2.2 不同 DNA 条形码候选序列 barcoding gap 检验

理想的 DNA 条形码序列, 其种间遗传变异应明显大于种内变异, 并在两者之间存在显著差异, 形成一个明显的间隔区, 即 barcoding gap^[14-15]。千斤拔属植物不同 DNA 条形码的 barcoding gap 见图 1。由图 1 可知, *rbcL*、*psbA-trnH* 序列的种间、种内变异重叠明显, 无明显 barcoding gap, 无法准确鉴

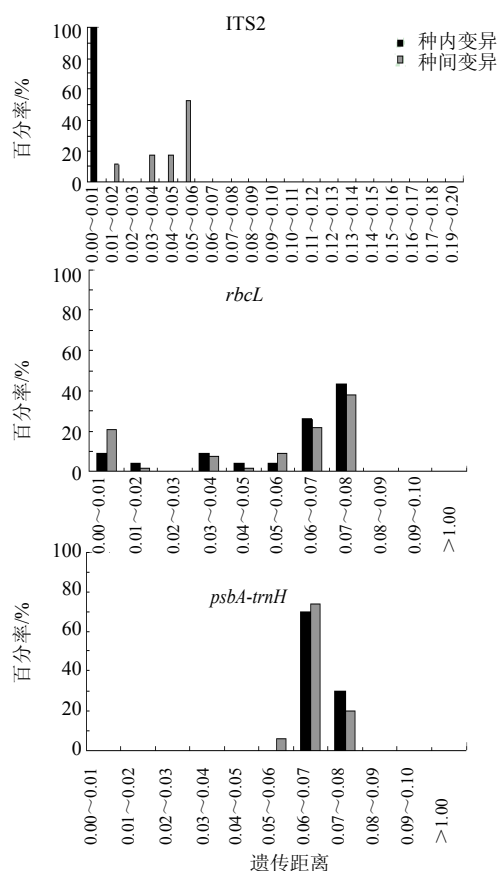


图 1 种内、种间变异相对分布图

Fig. 1 Relative distribution of inter-specific divergence and intra-specific variation

定物种; 而 ITS2 序列的种间、种内变异明显分离, 存在明显 barcoding gap, 有利于千斤拔属植物的物种鉴定。

2.3 系统进化分析

聚类树状图也是评估候选 DNA 条形码序列物种鉴别能力的方法之一。为增加条形码序列的物种鉴别分析可靠性, 将 GenBank 数据库中的千斤拔属相关条形码序列加入物种鉴定效率分析中。利用 MEGA v4.1 软件的 NJ 法构建各序列的聚类树状图, 由图 2 可知, ITS2 序列构建的聚类图能很好地将所分析物种的同种样品聚在一起, 而不同种间的分支又能很好地分开 (除 *F. philippinensis* 与 *F. stricta* 外, 其聚类分支无法明显分开), 充分说明 ITS2 对于千斤拔属植物具有很好物种鉴别能力; 而利用 *rbcL* 及 *psbA-trnH* 序列构建的系统进化树显示无法很好地反映不同种间的遗传变异及同种样品间的同源性。

由图 2 可知, 大叶千斤拔与宽叶千斤拔、长叶千斤拔与千斤拔亲缘关系较近, 2 分支聚合后又与球穗千斤拔聚在一起, 最后与腺毛千斤拔聚合。该分析结果与植物传统分类学中千斤拔属植物形态特征差异性结果基本一致, 说明分析结果可靠性较高。

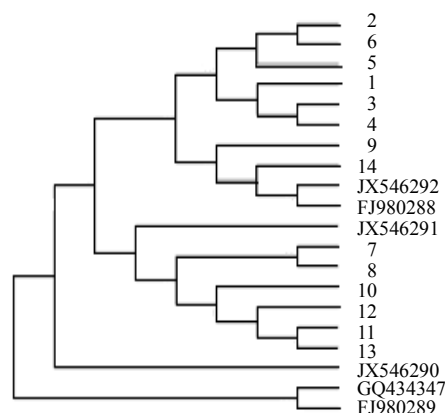


图 2 基于 ITS2 序列构建的 NJ 系统进化树

Fig. 2 NJ tree based on ITS2 sequences

3 讨论

3.1 候选 DNA 条形码序列对千斤拔属植物鉴别能力的分析比较

理想的 DNA 条形码应满足以下条件: (1) 具有显著的种间变异, 同时具有足够小的种内变异; (2) 有高度保守的引物设计区, 便于设计通用引物以扩增相应片段序列; (3) 片段应足够短, 以便于 DNA 提取和 PCR 扩增, 尤其是对部分降解 DNA 的

扩增。对千斤拔属 3 条候选序列进行比较分析, 综合 PCR 扩增和测序成功率、*barcoding gap* 和聚类图的分析结果可知, ITS2 较其他条形码序列, PCR 扩增和测序成功率均为 100%, 具有相对较明显的 *barcoding gap*, 且从聚类图可知, ITS2 对千斤拔属植物具有较好的鉴别能力。综上所述, 对于千斤拔属植物, ITS2 是比较理想的 DNA 条形码序列。

2009 年, CBOL 植物工作组通过大量实验发现, “*rbcL*+*matK*” 复合序列是较理想的植物 DNA 条形码序列组合, 其鉴定能力可达 72%。在实验过程中发现, 对于千斤拔属植物, 其 *matK* 序列的扩增成功率极低, 而 *rbcL* 序列的扩增成功率及测序成功率均为 100%, 但其序列较保守, 缺乏足够的信息位点 (序列长度为 660 bp, 但仅含 5 个信息位点), 无法鉴别千斤拔属植物; 因此, CBOL 所推荐的 “*rbcL*+*matK*” 复合序列无法在千斤拔属植物间成功应用。

2013 年, 陈士林课题组建议建立 “以 ITS2 为核心, *psbA-trnH* 为辅” 的植物类中药材 DNA 条形码鉴定体系^[10]。结果显示, 对于千斤拔属植物, ITS2 是较为理想的 DNA 条形码序列, 其 *psbA-trnH* 序列较其他序列, 测序成功率低, 且因种内变异较大, 无法成功鉴别所分析样品, 因此, 与 ITS2 单一片段相比较, “ITS2+*psbA-trnH*” 序列组合对于千斤拔属植物鉴定能力的改善无明显作用。

3.2 ITS2 序列对千斤拔药材准确鉴定的适用性及其意义

ITS2 序列作为 DNA 条形码候选序列具有以下特点: (1) ITS2 序列足够短, 降低了 DNA 提取、扩增和测序的难度, 尤其对那些已发生 DNA 降解的样品 (如药材、蜡叶标本等) 的分析鉴定工作具有重要意义; (2) ITS2 序列两端的 5.8 S 和 26 S 序列都很保守, 便于通用引物的设计; (3) Schultz 等^[16]建立的 ITS2 数据库能够预测 ITS2 二级结构并对间隔区进行注释, 极大地提高了 ITS2 序列的处理和分析的准确性、可靠性。本研究结果显示, ITS2 对千斤拔属植物具有很好的鉴别能力。因千斤拔药材以根入药, 传统的性状鉴别、显微鉴别、理化鉴别无法达到快速、准确鉴定的目的, 而化学指纹图谱等方法又因操作繁琐、专业背景要求高等特点, 无法满足中药材鉴别的实际应用。因此, 将 ITS2 序列作为千斤拔药材基原植物鉴定的条形码序列对于千斤拔药材的资源保护及安全用药具有重要意义。

随着分子生物学技术的快速发展, 高通量测序等技术的不断进步, DNA 条形码技术的应用不断拓展; 因中药研究的特殊性, 其应用安全性、种源确定、资源评价及濒危物种的保护问题日益成为中药发展不可避免的问题, 若 DNA 条形码技术能在药用植物鉴定中发挥作用, 必将对中药事业的发展起到至关重要的作用。

参考文献

- [1] 吴征镒. 新华本草纲要 [M]. 上海: 上海科学技术出版社, 1991.
- [2] 张丽霞, 彭建明, 马洁. 千斤拔研究进展 [J]. 中药材, 2007, 30(7): 887-890.
- [3] 中国药典 [S]. 一部. 2010.
- [4] Kress W J, Wurdack K J, Zimmer E A, *et al.* Use of DNA barcodes to identify flowering plants [J]. *Proc Natl Acad Sci USA*, 2005, 102(23): 8369-8374.
- [5] 俞超, 梁孝祺, 陈金金, 等. DNA 条形码技术鉴定贝母属植物 [J]. 中草药, 2014, 45(11): 1613-1619.
- [6] Han J P, Liu C, Li M H, *et al.* Relationship between DNA barcoding and chemical classification of salvia medicinal herbs [J]. *Chin Herb Med*, 2010, 2(1): 16-29.
- [7] Hebert P D N, Ratnasingham S, de Waard J R. Barcoding animal life: cytochrome oxidase subunit 1 divergences among closely related species [J]. *Proc Biol Sci*, 2003(270): 96-99.
- [8] Ning S P, Yan H F, Gang H, *et al.* Current advances of DNA barcoding study in plants [J]. *Biodiv Sci*, 2008, 16(5): 417-425.
- [9] CBOL Plant Working Group. A DNA barcode for land plants [J]. *Proc Natl Acad Sci USA*, 2009, 106(31): 12794-12797.
- [10] 陈士林, 姚辉, 韩建萍, 等. 中药材 DNA 条形码分子鉴定指导原则 [J]. 中国中药杂志, 2013, 38(2): 141-147.
- [11] Keller A, Schleicher T, Schultz J, *et al.* 5.8 S-28 S rRNA interaction and HMM-based ITS2 annotation [J]. *Gene*, 2009, 430(1/2): 50-57.
- [12] Kumar S, Nei M, Dudley J, *et al.* MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences [J]. *Brief Bioinform*, 2008, 9(4), 299-306.
- [13] Meier R, Kwong S, Vaidya G, *et al.* DNA Barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success [J]. *Syst Biol*, 55(5), 715-728.
- [14] Lahaye R, Bank M, Bogarin D, *et al.* DNA barcoding the floras of biodiversity hotspots [J]. *Proc Natl Acad Sci USA*, 2008, 105(8): 2923-2928.
- [15] Meyer C P, Paulay G. DNA barcoding: error rates based on comprehensive sampling [J]. *PLoS Biol*, 2005, 3(12): e422.
- [16] Schultz J, Müller T, Achtziger M, *et al.* The internal transcribed spacer 2 database—a web server for (not only) low level phylogenetic analyses [J]. *Nucl Acids Res*, 2006, 34: 704-707.