

• 药材与资源 •

金铁锁叶绿体基因组序列及其系统发育分析

高亚芳¹, 刘莹莹², 杨从卫¹, 李国栋^{1*}, 钱子刚^{1*}

1. 云南中医学院中药学院, 云南 昆明 650500

2. 云南省食品药品监督检验研究院, 云南 昆明 650000

摘要: **目的** 解析金铁锁叶绿体基因组结构和确定其在石竹科的系统位置。**方法** 基于二代测序技术 Illumina Hiseq 平台, 对金铁锁叶绿体基因组进行测序, 使用生物信息学分析软件进行组装和注释, 并以苋科植物千穗谷为外类群, 通过 RAxML 8.2.11 软件构建 ML 系统发育树, 分析石竹科属间的亲缘关系。**结果** 金铁锁叶绿体基因组总长为 153 977 bp。LSC 和 SSC 区被 2 个反向重复区域 IRs 隔开, 其大小分别为 84 385、17 526、26 033 bp。除去重复后, 共编码 109 个基因, 包含 75 个 PCGs, 30 个 tRNA 和 4 个 rRNA 基因。总的 GC 含量为 36.5%, IR 区 (42.4%) GC 含量明显高于 LSC (34.2%) 和 SSC (30.1%) 区。在系统进化树中, 金铁锁与长萼瞿麦互为姊妹类群, 且各节点的支持率较高, 能够清晰反映属间的亲缘关系。**结论** 本研究对金铁锁叶绿体基因组结构进行了解析, 并探讨了石竹科属间系统发育关系, 为金铁锁种质资源综合评价、系统进化等研究提供有效的科学数据。

关键词: 金铁锁; 叶绿体基因组; 石竹科; 系统发育; 长萼瞿麦**中图分类号:** R282.12 **文献标志码:** A **文章编号:** 0253-2670(2019)22-5532-05**DOI:** 10.7501/j.issn.0253-2670.2019.22.022**Analysis on structure and phylogeny of complete chloroplast genomes in *Psammosilene tunicoides***GAO Ya-fang¹, LIU Ying-ying², YANG Cong-wei¹, LI Guo-dong¹, QIAN Zi-gang¹

1. Faculty of Traditional Chinese Pharmacy, Yunnan University of Traditional Chinese Medicine, Kunming 650500, China

2. Yunnan Institute for Food and Drug Supervision Inspection, Kunming 650000, China

Abstract: **Objective** To obtain the structural characteristics of chloroplast genome in *Psammosilene tunicoides* and analyze its phylogenetic position within the family of Caryophyllaceae. **Methods** High-through-put sequencing technology and bioinformatic analysis softwares were used to analyze the structures of complete chloroplast genome of *P. tunicoides*. RAxML 8.2.11 software was utilized to reconstruct the Caryophyllaceae phylogentic tree with *Amaranthus hypochondriacus* as outgroup. **Results** The complete chloroplast genome was 153 977 base pairs (bp) in size, including two inverted repeat (IR, 26 033 bp) regions separated by one large single copy region (LSC, 84 385 bp) and one small single copy region (SSC, 17 526 bp). The genome encoded 125 genes, of which 109 were unique, including 75 protein-coding genes (PCGs), 30 tRNA genes and 4 rRNA genes. The overall GC content of *P. tunicoides* was 36.5%, while those of IR regions (42.4%) were higher than LSC (34.2%) and SSC (30.1%) regions. The phylogenetic tree showed that *P. tunicoides* and *Dianthus longicalyx* were sister groups with 100% bootstrap value. All nodes of the phylogenetic tree of Caryophyllaceae were of high supports, and the ML tree had good resolution to reflect the phylogeny relationship among the Caryophyllaceae. **Conclusion** The complete chloroplast genome of *P. tunicoides* and the phylogenetic relationship within Caryophyllaceae were analyzed in this study. The results will provide effective molecular information for further studies on evaluation of germplasm and molecular phylogeny of Caryophyllaceae.

Key words: *Psammosilene tunicoides* W. C. Wu et C. Y. Wu; complete chloroplast genome; Caryophyllaceae; phylogeny; *Dianthus longicalyx*

金铁锁 *Psammosilene tunicoides* W. C. Wu et C. Y. Wu 是石竹科金铁锁属的单种植物, 是我国西

南地区特有的多年生药用植物^[1]。到目前为止, 从金铁锁根中分离得到了 100 多种化学成分, 主要为

收稿日期: 2019-03-24**基金项目:** 国家自然科学基金项目 (81560613); 云南省科学技术厅-云南中医学院应用基础研究联合专项 [2015FB205 (-015)]**作者简介:** 高亚芳 (1992—), 女, 硕士研究生, 研究方向为分子生药学。Tel: 18213961630 E-mail: 306618006@qq.com***通信作者** 李国栋, 副教授, 硕士生导师, 主要从事分子生药学研究。E-mail: gammar116@163.com

钱子刚, 教授, 博士生导师, 主要从事中药资源学研究。E-mail: qianzig@aliyun.com

皂苷、环肽、呋伯啉生物碱等成分^[2]。药理活性研究结果显示,水煎浸膏和总皂苷在止痛、止血、抗炎和免疫调节等方面有较高价值^[3]。金铁锁作为“云南白药”和“百宝丹”系列中成药的主要组成之一,在临床运用中有显著疗效^[4]。由于金铁锁市场需求量较大,野生居群数量急剧下降,早在 1991 年作为稀有濒危物种,被列入《中国植物红皮书》中^[5]。随着对金铁锁野生资源的不断采挖,金铁锁分布范围越来越狭窄。在不断变化的环境中,自然居群的遗传多样性能够对物种长期生存和进化产生很大影响^[6]。前人对金铁锁遗传多样性的研究,主要集中于用叶绿体片段和核基因片段作为分子标记^[7]。由于所用片段较短,给出的变异位点相当有限,所得结果无法保证其准确性。然而,关于金铁锁基因组的研究仍未见报道。因此,本研究通过对金铁锁叶绿体基因组的结构进行解析,为进一步研究金铁锁的遗传背景、种质资源保护与开发利用、以及在植物系统发育进化等方面奠定基础。

1 材料

野生金铁锁 *Psammosilene tunicoides* W. C. Wu et C. Y. Wu 新鲜、幼嫩的叶片采自云南丽江(26°85'N; 100°25'E),经云南中医学院钱子刚教授鉴定,标本存放于云南中医学院种苗繁育中心。

2 方法

2.1 基因组 DNA 的提取和测序

使用改良的植物基因组快速提取试剂盒(百泰克有限公司)进行金铁锁总 DNA 的提取,操作步骤与郭蔚文^[8]的方法一致。利用琼脂糖凝胶电泳对提取到的 DNA 质量和完整性进行检测,然后用 NanoDrop 2000 (Thermo Scientific, 美国)进行浓度和含量检测,达到测序要求后,送到测序公司,采用 Illumina HiSeq 2500-PE150 平台进行建库测序。

2.2 叶绿体基因组的组装与注释

测序所获得的 raw reads 数据通过使用 NGSQCToolkit_v2.3.3^[9]软件滤过,除去低质量区,获得 clean reads。将筛选出的 reads 用在线软件

GetOrganelle pipeline (<https://github.com/Kinggerm/GetOrganelle>) 进行 *de novo* 组装,以 *Silene latifolia* (GnenBank 登录号: NC016730) 为参考序列, Kmer 值设置为 39,选择合适的 contig 后,基因组序列组装完成。接着,采用在线注释软件 Annotation tool GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>)^[10]对金铁锁叶绿体基因组进行注释,然后用 Geneious R11 软件^[11]进行手动校正。将注释好的序列转化为 sqn 格式文件,上传至 NCBI。最后,利用 OGDRAW 在线工具^[12]绘制金铁锁叶绿体基因组物理图谱。

2.3 系统进化分析

为了确定金铁锁在石竹科中的系统位置,本研究选择了 15 个已发表的石竹科叶绿体基因组为内类群和与石竹科亲缘关系较近的苋科中的植物千穗谷 *Amaranthus hypochondriacus* L. 为外类群,构建最大似然(maximum likelihood, ML)系统发育树。首先对 17 个叶绿体基因组用 MAFFT 7.308^[13]进行全基因组比对,之后使用 RAxML 8.2.11^[14]软件,选择核苷酸替代模型 GTR+I+G (Bootstrap 1 000 次重复)构建系统发育树。

3 结果与分析

3.1 金铁锁叶绿体基因组基本特征

金铁锁叶绿体基因组 (Genbank 登录号: MH521155) 是一个双链环状 DNA,呈典型四段式结构,总长度为 153 977 bp。1 个大的单拷贝区 (large single-copy region, LSC) 和 1 个小的单拷贝区 (small single-copy region, SSC) 被 2 个反向重复区域 (inverted repeat regions, IRa/IRb) 隔开,其大小分别为 84 385、17 526、26 033 bp。总的 GC (嘧啶) 含量为 36.5%, AT (嘌呤) 含量为 63.5%,有明显的 AT 偏向性。各区段的 GC 含量和长度见表 1。其中 IR 区 GC 含量最高,达到 42.4%,其次是 LSC 区 (34.2%) 和 SSC 区 (30.1%)。IR 区 GC 含量最高,是由于在 IR 区 4 个高 GC 含量 (55.2%) 的 rRNA 基因造成的。

表 1 金铁锁叶绿体基因组碱基组成

Table 1 Base composition of chloroplast genome in *P. tunicoides*

基因	T (U)/%	A/%	C/%	G/%	长度/bp	GC/%
LSC	33.5	32.3	17.6	16.6	84 385	34.2
SSC	34.6	35.4	15.6	14.4	17 526	30.1
IRa	28.3	29.3	20.2	22.2	26 033	42.4
IRb	28.3	29.3	22.2	20.2	26 033	42.4
合计	32.0	31.5	18.6	17.9	153 977	36.5

3.2 金铁锁叶绿体基因组各分区基因组成

金铁锁叶绿体基因组共包含 125 个基因，除去重复基因后有 109 个。其中有 75 个蛋白质编码基因，30 个编码 tRNA 基因，4 个编码 rRNA 基因。在这些基因中，有 6 个 tRNA 基因、4 个 rRNA 基因、2 个未知功能的基因（ycf3 和 ycf2）、1 个 NADH 脱氢酶

（ndhB）和 1 个核糖体蛋白小亚基（rps7）存在于 IR 区，这些基因在 IR 区各有一个拷贝。值得注意的是，在图中（图 1）有一部分 ycf1 基因在 SSC 和 IRb 的交界处重复。在 SSC 区共有 12 个 CDS 和 1 个 tRNA 基因（trnL-UAG）存在。然而绝大多数的基因存在于 LSC 区，有 59 个 CDS 和 22 个 tRNA 基因（表 2）。

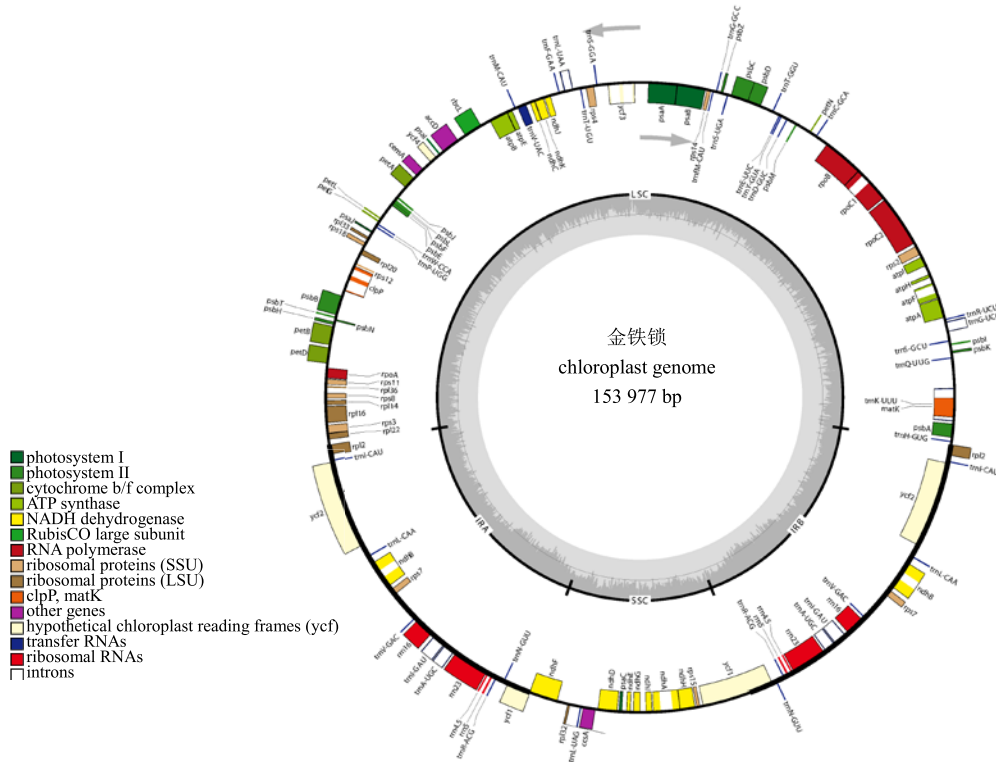


图 1 金铁锁叶绿体基因组环形基因图

Fig. 1 Gene map of chloroplast genome in *P. tunicoides*

表 2 金铁锁叶绿体基因组各分区的基因组成

Table 2 Gene composition of LSC, SSC, IR in *P. tunicoides* chloroplast genome

基因名称	Gene	CDS	tRNA	rRNA
LSC	81	59	22	0
SSC	13	12	1	0
IRa	16	5	7	4
IRb	16	5	7	4
合计	125	80	37	8

3.3 金铁锁叶绿体基因组编码的基因归类

金铁锁全部基因分类情况如表 3 所示，根据其功能可以把它们分为 4 大类：分别是与光合作用有关的基因、与自我复制有关的基因、未知功能的蛋白质基因，以及成熟酶基因（matK）、囊膜蛋白基因（cemA）等其他基因。在这些基因中含有 15 个内含子，有 3 个在 IR 区，剩余的除 ndhA 基因外，

均存在于 LSC 区。其中 ndhA、ndhB、rpl16、rpoc1、atpF、trnA-UGC、trnI-GAU、trnV-UAC、trnL-UAA、trnG-UCC 和 trnK-UUU 基因有一个内含子。而 clpP 和 ycf3 基因存在 2 个内含子，与其他高等植物类似。通过与其他内含子比较，存在于 LSC 区的 trnK-UUU 内含子最长，其片段大小为 2 496 bp，包含 matK 基因。

3.4 金铁锁叶绿体基因组系统发育分析

本研究为了确定金铁锁在石竹科的系统位置，基于 APG IV 分类系统，从 NCBI 上下载了 15 个石竹科叶绿体基因组作为内类群和 1 个苋科植物千穗谷为外类群进行系统发育分析，构建 ML 系统发育树。从图 2 中可以看出，分析结果强烈支持金铁锁与石竹属植物长萼瞿麦 *Dianthus longicalyx* Miq. 聚在一起，互为姊妹类群。ML 系统发育树的拓扑结构中同属于蝇子草族的麦仙翁属 *Agrostemma* L.、蝇

表 3 金铁锁叶绿体基因组编码的基因归类

Table 3 Genes category of chloroplast genome in *P. tunicoides*

功能	基因分组	基因名
photosynthesis	photosystem I	psaA、psaB、psaC、psaJ
	photosystem II	PsbA、PsbB、PsbC、PsbD、PsbE、PsbF、PsbH、PsbI、PsbJ PsbK、PsbL、PsbM、PsbN、PsbT
	cytochrome b/f complex	PetA、petB、petD、petG、petL、petN
	ATP synthase	AtpA、atpB、atpE、atpF*、atpH、atpI
	rubisCO large subunit	rbcL
self-replication	NADH dehydrogenase	ndhA*、ndhB*(x2)、ndhC、ndhD、ndhE、ndhF、ndhG、ndhH、ndhI、ndhJ、ndhK
	RNA polymerase	rpoA、rpoB、rpoC1*、rpoC2
	ribosomal proteins (SSU)	rps2、rps3、rps4、rps7(x2)、rps8、rps11、rps12 (x2)、rps14、rps15、rps18
	ribosomal proteins (LSU)	rpl2、rpl12、rpl14、rpl16、rpl20、rpl22、rpl32、rpl33、rpl36
	transfer RNAs	37 tRNAs (5 contain an intron, 7 in the IRs)
	ribosomal RNAs	rrn4.5 (x2)、rrn5 (x2)、rrn16 (x2)、rrn23 (x2)
other genes	other genes	clpP**、matK、accD、ccsA、cemA
unknown function	hypothetical chloroplast reading frames	Ycf1 (x2)、ycf2 (x2)、ycf3**、ycf4

*基因中有 1 或 2 个内含子

*gene contains one or two introns

子草属 *Silene* L. 和剪秋罗属 *Lychnis* L. 聚为一大支, 且支持率基本为 100%, 其分类结果与 APG IV 分类结果一致。在该研究中, 除了裸果木属的裸果木 *Gymnocarpus przewalskii* L. (属于指甲草亚科) 外, 其余均为石竹亚科, 分属于不同的族, 该系统树能够准确地将它们区分开, 并且支持率均为 100%, 可信度较高, 表明叶绿体基因组可以有效解决该科系统进化问题。

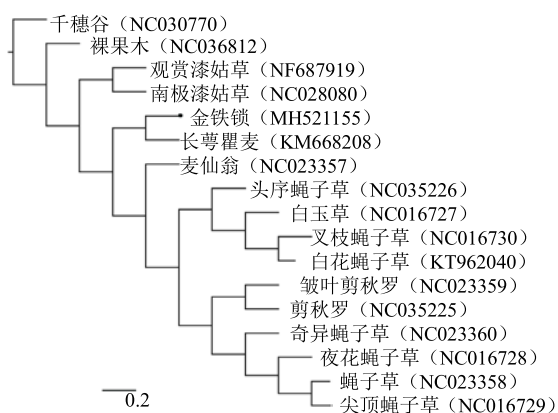


图 2 基于石竹科叶绿体全基因组序列构建系统进化树

Fig. 2 Phylogenetic tree based on chloroplast genome sequence of Caryophyllaceae

4 讨论

金铁锁是石竹科单种属植物, 其根具有较高的药用价值和经济价值。由于过度采挖, 其栖息地遭到严重破坏, 野生居群的数量急剧下降。为了保护金铁锁野生资源, 已有不少研究者对金铁锁组织培养、栽培技术、生物学特性等^[4]方面进行了大量研究。然而关于金铁锁叶绿体基因组的研究还处于空白状态。

本研究完成了金铁锁叶绿体基因组的测序、组装和注释, 并对其结构、GC 含量等进行了分析。研究结果显示, 金铁锁叶绿体基因组全长 153 977 bp, GC 含量为 36.5%, 共编码 125 基因等, 其结果与大多数高等植物类似。在该基因组中存在一些未知功能的基因, 如 *ycf1*、*ycf2*、*ycf3* 和 *ycf4*, 与大多数植物一样, IRb 和 SSC 边界处于 *ycf1* 基因上, 这可能与 IRb/SSC 边界的扩张有关。到目前为止, 其功能尚未清楚的 *ycf15* 基因在大多数被子植物中存在, 但在金铁锁及其大部分石竹科植物的叶绿体基因组中完全消失。IR 区与 LSC 和 SSC 区相比, GC 含量最高, 主要是因为该区域有 4 个 rRNA 基因存在。当基因进行选择性剪切时, 内含子发挥了重要作用, 金铁锁叶绿体基因组中共有 15 个内含

子。其中, trnK-UUU 内含子最长, 其片段大小为 2 496 bp, matK 基因包含在其中。

金铁锁作为古地中海子遗植物^[3], 早期由 Diels 鉴定为蝇子草属的 1 个新种。后由吴蕴珍和吴征镒先生发现其花和果实的结构与蝇子草属不同, 于是, 将其独立为一个单种属, 命名为金铁锁^[15]。后来, 这一观点从核型分析, 胚胎学和分子生物学特征等方面得到证实^[16]。屈燕等^[16]对金铁锁的大孢子进行切片研究, 并与其他近缘属进行比较发现, 金铁锁属和蝇子草属的亲缘关系最近, 与麦仙翁属 *Agrostemma* L. 和繁缕属 *Stellaria* L. 的关系较近, 与石竹属 *Dianthus* L. 的关系较远。本研究, 石竹科的叶绿体基因组 ML 系统发育树结果表明, 金铁锁与石竹属的长萼瞿麦互为姊妹类群, 而非蝇子草属植物。这一结果与前人的研究结果存在一定的分歧, 表明石竹科分子系统进化仍需进行深入的研究。叶绿体基因组 ML 树的拓扑结构显示金铁锁与蝇子草属的植物分布在不同的进化支上, 其结果证明了将金铁锁从蝇子草属中独立出来观点的准确性, 并且对石竹科的这 16 个植物的分类结果与 APG IV 的系统分类结果一致。在 14 个节点中, 有 13 个节点支持率为 100%, 另一个节点为 98%, 可信度非常高。表明了叶绿体基因组与传统的分类方法相比, 可以对该科植物进行简单、快捷、准确分类。叶绿体基因组研究除了用于物种分类外, 还可以用于物种鉴定、起源和进化, 遗传转化等研究。Zhang 等^[7]通过 2 个叶绿体片段 (rpL16, trnQ-5'rps16) 和核基因 (GPA1) 对金铁锁居群进行了谱系地理学研究, 结果显示, 所使用的片段变异位点有限。因此, 对居群间和居群内的遗传多样性进行研究, 可能会造成结果不准确的问题。本课题组在基于谱系地理学构建金铁锁核心种质的研究中, 用到 2 个叶绿体片段 (psbA-trnH、psbD-trnT) 对金铁锁野生和栽培居群进行研究时, 发现由于其变异位点的限制, 导致栽培居群仅有普通单倍型, 其他特有和私有单倍型无法呈现出来。本研究拟通过对金铁锁叶绿体基因组进行解析, 为金铁锁群体遗传方面的研究奠定基础。

参考文献

- [1] 黄春青, 林亚平. 金铁锁的研究进展 [J]. 贵阳中医学院学报, 2007, 29(6): 56-59.
- [2] 毛泽玲, 沈云亨, 周李刚. 金铁锁的化学成分与生物活

性的研究进展 [J]. 中华中医药学刊, 2016, 34(12): 2883-2887.

- [3] 张庆滢, 刘小烛, 毛常丽. 药用植物金铁锁的研究进展 [J]. 云南农业大学学报: 自然科学, 2009, 24(1): 140-141.
- [4] 李 斌, 李德龙, 尹丽莎, 等. 金铁锁研究进展综述 [J]. 安徽农学通报, 2016, 22(3/4): 29-32.
- [5] 傅立国. 中国植物红皮书 [M]. 北京: 科学出版社, 1991.
- [6] Jia X M, Liu X P. Characterization of the complete chloroplast genome of the Chinese yew *Taxus chinensis* (Taxaceae), an endangered and medicinally important tree species in China [J]. *Conserv Genet Res*, 2016, 9(2): 197-199.
- [7] Zhang Q Y, Zhao Y J, Gong X. Genetic variation and phylogeography of *Psammosilene tunicoides* (Caryophyllaceae), a narrowly distributed and endemic species in south-western China [J]. *Austral J Bot*, 2011, 59(5): 450-459.
- [8] 郭蔚文. 罗锅底及其混伪品的 DNA 条形码鉴定研究 [D]. 昆明: 云南中医学院, 2016.
- [9] Patel R K, Jain M. NGS Q C Toolkit: A toolkit for quality control of next generation sequencing data [J]. *PLoS One*, 2012, 7(2): e30619.
- [10] Tillich M, Lehwark P, Pellizzer T, et al. GeSeq-versatile and accurate annotation of organelle genomes [J]. *Nucl Acids Res*, 2017, 45(W7): 6-11.
- [11] Kearse M, Moir R, Wilson A, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. *Bioinformatics*, 2012, 28(12): 1647-1649.
- [12] Lohse M, Drechsel O, Kahlau S, et al. Organellar genome DRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. *Nucl Acids Res*, 2013, 41: 575-581.
- [13] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability [J]. *Mol Biol Evol*, 2013, 30(4): 772-780.
- [14] Stamatakis A, Hoover P, Rougemont J, et al. A rapid bootstrap algorithm for the RAxML web servers [J]. *System Biol*, 2008, 57(5): 758-771.
- [15] 潘跃芝, 杨志云, 龚 洵. 中国特有植物金铁锁的细胞学研究 [J]. 植物分类与资源学报, 2004, 26(2): 204-206.
- [16] 屈 燕, 虞 泓, 周湘玲. 珍稀濒危药用植物金铁锁研究进展 [J]. 中华中医药杂志, 2011, 26(8): 1795-1797.