中国科学: 生命科学 2018 年 第 48 卷 第 5 期: 571 ~ 582
SCIENTIA SINICA Vitae lifecn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS

论 文

薯蓣和叉蕊薯蓣叶绿体基因组及特异性DNA条形码鉴定序列筛选研究

马双姣[1†], 周建国[1†], 李滢[1], 陈新连[1], 吴明丽[2], 孙伟[3], 李永华[4], 宋经元[1], 姚辉[1*]

1. 中国医学科学院北京协和医学院药用植物研究所, 国家中医药管理局中药资源保护重点研究室, 北京 100193;
2. 湖北中医药大学药学院, 武汉 430065;
3. 中国中医科学院中药研究所, 北京 100700;
4. 广西中医药大学药学院, 南宁 530001
† 同等贡献
* 联系人, E-mail: scauyaoh@sina.com

收稿日期: 2017-10-06; 接受日期: 2017-11-16; 网络版发表日期: 2018-02-05
国家科技支撑计划(批准号: 2011BAI07B08)和中国医学科学院医学与健康科技创新工程项目(批准号: 2016-I2M-3-016)资助

摘要 前期研究表明, 目前常用的DNA条形码序列对薯蓣属物种的鉴定效率低. 本研究利用高通量测序技术测定了薯蓣(*Dioscorea opposita*)和叉蕊薯蓣(*D. collettii*)叶绿体基因组, 完成了其物理图谱绘制, 基因组结构特征解析, 特异性DNA条形码序列的筛选. 薯蓣和叉蕊薯蓣叶绿体基因组总长度分别为152963和153870 bp, 均包含两个反向重复区(IRs)、一个大单拷贝区(LSC)和一个小单拷贝区(SSC) 4部分. 薯蓣和叉蕊薯蓣均含有125个基因, 其中包含87个蛋白编码基因、30个tRNA和8个rRNA. 薯蓣和叉蕊薯蓣的GC含量分别为37.04%和37.17%, 经多序列比对发现薯蓣属叶绿体基因组中非编码区变异高于保守的蛋白编码区域, LSC区、SSC区变异大于IR区, 筛选出了10个潜在的适合作为鉴定薯蓣属植物的特异性DNA条形码序列, 其中包括5个蛋白编码基因和5个基因间隔区. 系统进化树显示, 叶绿体全基因组序列也可作为薯蓣属物种鉴定的超级条形码. 本研究为薯蓣属系统进化以及物种鉴定等领域的研究提供依据.

关键词 薯蓣, 叉蕊薯蓣, 叶绿体基因组, DNA条形码, 鉴定

薯蓣属(*Dioscorea* L.)是薯蓣科(Dioscoreaceae)中种类最多的一属, 我国约有49种, 主产西南和东南部[1]. 薯蓣(*D. opposita* Thunb.)的干燥根茎为中药材山药, 具有补脾养胃, 生津益肺, 补肾涩精等功效, 用于脾虚食少, 久泻不止, 肺虚喘咳, 肾虚遗精, 带下, 尿频, 虚热消渴等症, 麸炒山药补脾健胃[2]. 2002年, 中华人民共和国卫生部公布的《关于进一步规范保健食品原料管理的通知》中, 山药被列入药食同源的物品名单. 现代药理研究表明, 山药含有许多活性成分, 如糖类、蛋白质、氨基酸、微量元素等, 具有活化神经细胞、降低血脂、调整体内激素分泌、抗衰老和抗氧化等多种作用, 具有很高的研发价值[3,4]. 叉蕊薯蓣(*D. collettii* J. D. Hooker)中含有薯蓣皂苷元和约莫皂苷元等甾体化合物[5,6]. 现代药理作用研究表明, 薯蓣皂苷元具有抗

引用格式: 马双姣, 周建国, 李滢, 等. 薯蓣和叉蕊薯蓣叶绿体基因组及特异性DNA条形码鉴定序列筛选研究. 中国科学: 生命科学, 2018, 48: 571–582
Ma S J, Zhou J G, Chen X L, et al. Complete chloroplast genomes of *Dioscorea opposite* and *D. collettii* and screening specific DNA barcodes (in Chinese). Sci Sin Vitae, 2018, 48: 571–582, doi: 10.1360/N052017-00160

© 2018 《中国科学》杂志社 www.scichina.com
https://engine.scichina.com/doi/10.1360/N052017-00160

过敏、抗肿瘤、抗心血管疾病等作用[7~9]. 薯蓣皂苷元不仅是人工合成甾体激素类药物不可替代的天然中间体, 同时也是治疗心血管病、糖尿病等药物的重要原料[10,11].

由于我国薯蓣属植物分布区域广泛, 种类丰富, 对不同地区的薯蓣种质进行系统的分析评价存在较大难度. 目前薯蓣栽培作物较少通过人工杂交进行育种, 生产上栽培的薯蓣主要是经过长期驯化和有针对性地选种得来. 薯蓣属植物生长发育易受到环境因子的影响, 同一物种在不同生长环境下, 形态特征及所含化学成分差异较大, 在长期驯化栽培过程中形成了一系列地方品种, 由于缺乏系统的鉴定和命名, 且大部分种质资源遗传背景复杂, 导致物种种质混乱[12,13]. 同时, 薯蓣属植物叶片形态极其相似, 这给传统方法在薯蓣属植物分类与鉴定中的研究带来一定的困难. 然而, 准确地鉴定物种是确保该属药用植物临床疗效、用药安全的关键. 因此, 需要建立准确、稳定的方法对薯蓣属植物进行鉴定.

DNA条形码分子鉴定法是一种新的生物鉴定方法, 具有通用性强、稳定性高、操作简单、鉴定结果可重复性良好的特点, 是传统鉴定方法的有效补充[14,15]. 目前对植物类药材来说, ITS2片段是最具优势的DNA条形码序列之一[16,17]. 已有多名学者利用DNA条形码序列对薯蓣属植物进行亲缘关系和物种鉴定等方面的研究[18~20], Sun等人[20]应用*matK*, *rbcL*和*psbA-trnH*对薯蓣属38个种、7个变种和1个亚种共138份样品进行研究, 结果显示叶绿体基因序列中*matK*鉴定效率较高, 为23.26%, 核基因ITS和ITS2序列的测序效率很低. 总体而言, 目前已有的条形码序列对薯蓣属的鉴定效率并不理想. 近年来, 随着高通量测序和生物信息技术的不断发展以及测序成本的降低, 大规模地利用二代及三代测序技术进行叶绿体基因组测序成为可能, 叶绿体基因组被越来越多地用在了条形码鉴定、分子标记和系统发育上, 在揭示物种进化及其种间亲缘关系等方面具有重要价值[21~24]. Li等人[25]提出叶绿体基因组可作为超级条形码用于物种鉴定特别是近缘种和种下等级的鉴定. 李滢等人[26]通过对贝母属4个物种叶绿体基因组进行全局分析, 筛选出了7个基因间区作为潜在的贝母属植物鉴定的特异性DNA条形码. Dong等人[27]筛选了12个属29个物种的叶绿体基因组序列的高变异区, 结果显示筛选出的高变异区对近缘物种的条形码鉴定提供了良好的数据支持. 本研究利用高通量测序技术对薯蓣属物种薯蓣(*D. opposita*)和叉蕊薯蓣(*D. collettii*)进行了叶绿体基因组测序和分析, 旨在为两物种的系统进化位置以及特异性DNA条形码序列的筛选提供依据.

1 材料与方法

1.1 实验材料

本实验中用于DNA提取的薯蓣(*D. opposita*)和叉蕊薯蓣(*D. collettii*)的新鲜叶片分别采自河南焦作(35.22°N, 113.25°E)和四川峨眉山(29.57°N, 103.36°E)地区, 样品均经中国医学科学院药用植物研究所林余霖研究员鉴定, 凭证标本保存于中国医学科学院药用植物研究所.

1.2 DNA提取、叶绿体基因组序列的测序、拼接和验证

取新鲜幼嫩的叶片, 采用QIAGEN试剂盒法提取总DNA. 利用NanoDrop 2000微量分光光度计检测总DNA的浓度并用1%的琼脂糖凝胶电泳检测总DNA的质量. 将检验合格的总DNA样品通过Covaris超声破碎仪随机打断成长度约500 bp的片段, 经末端修复、加A尾、加测序接头、纯化、PCR扩增等步骤完成文库制备. 构建好的文库通过Illumina Hiseq 2000进行全基因组*de novo*测序. 对测序获得的数据进行过滤, 应用生物信息学方法挑出叶绿体基因组序列读长进行拼接和组装.

从拼接好的叶绿体基因组中选取6个相对保守的基因(*cemA*, *psbB*, *psbC*, *rpoA*, *rpoC1*, *rps4*)[28]以及4个连接区段(LSC-IRa, IRa-SSC, SSC-IRb, IRb-LSC), 分别设计引物(网络版附表1), 以基于Sanger测序平台对PCR产物进行直接测序, 将获得的序列与基因组组装结果比对, 以验证拼接结果的准确度.

1.3 叶绿体全基因组注释、分析和提交

将同属物种*D. elephantipes* (EF380353)[29], *D. rotundata* (KJ490011)[30]、穿龙薯蓣(*D. nipponica* (KP404629))[31]、盾叶薯蓣(*D. zingiberensis* (KP899622))[32]的叶绿体基因组作为参考基因组. 根据序列同源性, 按照Liu等人[33]介绍的方法, 采用在线注

https://engine.scichina.com/doi/10.1360/N052017-00160

释软件DOGMA (http://dogma.ccbb.utexas.edu/)对叶绿体全基因组进行基因预测和功能注释分类, 基因的起始和终止密码子需手工校正; 将注释的结果用Sequin软件分析序列数据和注释信息, 生成序列的Sqn文件, 并提交至GenBank上. 将薯蓣和叉蕊薯蓣序列的GenBank格式文件用GenomeVx在线绘图工具(http://wolfe.gen.tcd.ie/GenomeVx/)绘制叶绿体全基因组物理图谱. 使用MEGA 6.0 (molecular evolutionary genetics analysis 6.0)[34]分析密码子使用情况和GC含量.

简单重复序列(simple sequence repeats, SSR)是由长度为1~6 bp的重复单位组成的长达几十个核苷酸的序列, 其分布广泛, 存在于基因组的各个区域[35,36], 是基于微卫星DNA的高突变率、多等位性、共显性而开发的一种分子标记技术. 本研究利用MISA软件(http://pgrc.ipk-gatersleben.de/misa/)对薯蓣和叉蕊薯蓣进行叶绿体全基因组SSR位点搜索. 其参数设置为, 单核苷酸单元的重复数≥8, 二核苷酸单元的重复数≥4, 三到六核苷酸单元的重复数≥3. 本文中主要搜索完全重复型SSR位点, 循环排列或成反向互补的SSR被当作同一种类型.

1.4　序列变异、遗传距离及系统发育关系分析

本研究应用在线基因组比对工具mVista[37]对包括薯蓣(*D. opposita*)、叉蕊薯蓣(*D. collettii*)及*D. elephantipes*, *D. rotundata*, *D. villosa*、穿龙薯蓣(*D. nipponica*)、盾叶薯蓣(*D. zingiberensis*)在内的共7条薯蓣属叶绿体基因组进行全基因组比对. 应用ClustalW2[38]对7种薯蓣属叶绿体基因组特异性DNA条形码候选序列进行序列比对, 用软件MEGA 6.0对序列进行变异位点和遗传距离的分析. 利用MAFFT version 5[39]软件对包括薯蓣属在内的共24条叶绿体基因组序列进行全基因组比对, 用最大似然(maximum-likelihood, ML)法采用Tamura-Nei模型构建系统发育树(bootstrap 1000次重复).

2　结果及分析

2.1　叶绿体全基因组基本特征、基因组成及分类

薯蓣(*D. opposita*)和叉蕊薯蓣(*D. collettii*)的叶绿体全基因组均为常见的四分体结构: 两个反向重复区IRs、一个大单拷贝区LSC和一个小单拷贝区SSC (图1). 经拼接后薯蓣(*D. opposita*)叶绿体全基因组(GenBank 登录号KY996494)全长为152963 bp, GC含量为37.04%; 叉蕊薯蓣(*D. collettii*)叶绿体全基因组(GenBank 登录号KY996495)全长为153870 bp, GC含量为37.17%. 各区段长度及GC含量详见表1. 可以看出, 4个区段中GC含量最高的是IR区(42.96%), 其次是LSC区(34.98%)和SSC区(31.17 %).

两物种叶绿体基因组结构和基因组成相似, 可以分为以下4类: 与光合作用有关的基因, 与自我复制有关的基因, 未知功能的蛋白基因和成熟酶基因(*matK*)、囊膜蛋白基因(*cemA*)等其他基因. 每个物种均注释到 87个蛋白编码基因, 30个tRNA基因和8个rRNA基因. 其中8个蛋白编码基因(*rpl2*, *rpl23*, *ycf2*, *ycf15*, *ndhB*, *rps7*, *rps12*和*ycf1*)、6个tRNA (*trnN-GTT*, *trnR-ACG*, *trnV-GAC*, *trnL-CAA*, *trnI-CAT*和*trnH-GTG*)和8个rRNA位于反向重复区, 基因组成详见网络版附表2. 所有的87个蛋白编码基因中有12个基因含有内含子, 其中3个基因(*ycf3*, *clpP*和*rps12*)含有两个内含子, 其余10个基因只有一个内含子, *rps12*基因为反式剪切基因, 该基因的5′端位于LSC区, 3′端位于IR区(表2).

2.2　叶绿体基因组密码子使用分析

以叉蕊薯蓣为例, 根据所有蛋白编码基因和转运RNA编码基因, 计算其叶绿体基因组密码子使用频率

表 1　薯蓣和叉蕊薯蓣叶绿体基因组各区长度及GC含量

	D. opposita		*D. collettii*	
	长度(bp)	GC含量(%)	长度(bp)	GC含量(%)
全长	152963	37.04	153870	37.17
LSC	83153	34.83	83826	34.98
SSC	18825	30.96	18863	31.17
IRs	25491	42.90	25589	42.96

https://engine.scichina.com/doi/10.1360/N052017-00160

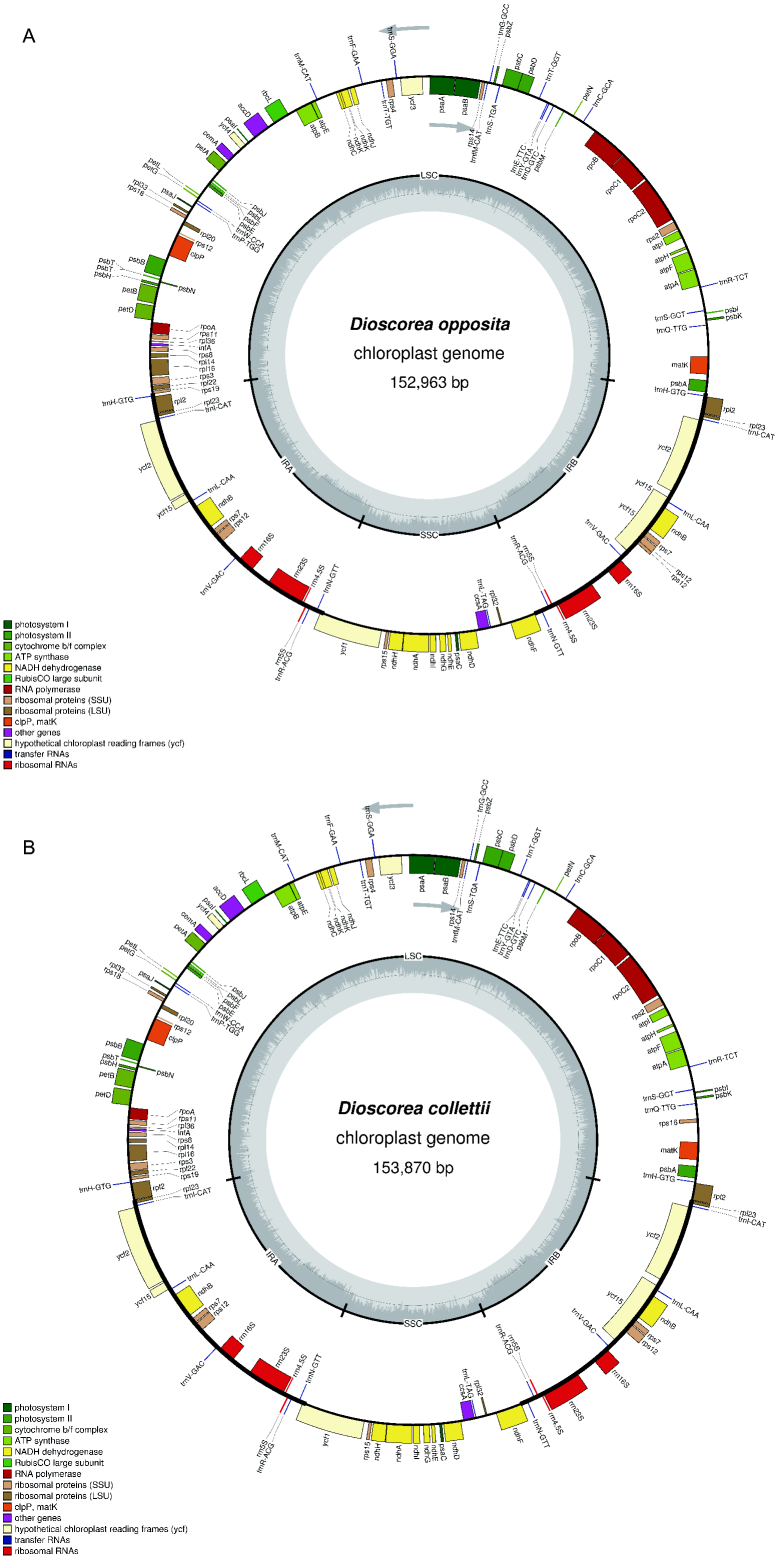


图 1　薯蓣(A)和叉蕊薯蓣(B)叶绿体基因组环形图

内侧基因顺时针转录, 外侧基因逆时针转录

https://engine.scichina.com/doi/10.1360/N052017-00160

表 2　以叉蕊薯蓣为例的12个包含内含子的叶绿体基因位置及长度

基因	位置	外显子 I (bp)	内含子 I (bp)	外显子 II (bp)	内含子 II (bp)	外显子III (bp)
rps16	LSC	87	210	39		
atpF	LSC	151	981	410		
rpoC1	LSC	436	736	1618		
ycf3	LSC	127	733	230	754	156
rps12	LSC	114	–	232	548	26
clpP	LSC	69	754	291	635	252
petD	LSC	6	785	521		
rpl16	LSC	399	1017	9		
rpl2	IR	428	666	394		
ndhB	IR	870	608	756		
ndhA	SSC	553	1083	539		
ndhB	IR	777	701	756		
rpl2	IR	428	669	397		

(relative synonymous codon usage, RSCU). 结果显示, 在所有的密码子中, 编码亮氨酸(Leu)的密码子最多, 为2750个. 密码子总数不超过1000的氨基酸有甲硫氨酸(Met)、酪氨酸(Tyr)、组氨酸(His)、谷氨酰胺(Gln)、半胱氨酸(Cys)、色氨酸(Trp)、丝氨酸(Ser)、终止子(TER). 半胱氨酸(Cys)密码子最少, 为320个(表3). 在烟草(*Nicotiana tabacum* L.)叶绿体基因组中, *rps19*, *psbC*和*ycf15*的起始密码子是GTG, *psbL*和*ndhD*的是ACG[40]. 在水稻(*Oryza sativa*)中, ACG和GTG是*rpl2*和*rps19*的起始密码子[41].

2.3　SSR分析

由于SSR具有丰富的多态性, 现已成为一种高效的分子标记, 已在物种鉴定和系统进化分析等方面广泛应用[35,36]. 薯蓣叶绿体全基因组共有215个符合条件的SSR位点. 其中单碱基重复有140个, 二碱基重复有56个, 三碱基重复有10个, 四碱基重复有4个, 五碱基重复有5个. 重复最多的序列是单碱基重复, 主要是A/T的重复, 共有74处连续A及65处连续T的单碱基重复, 占所有单碱基重复(140处)的99.3%. 然后依次是AT/AT (39次), AG/CT (14次), AC/GT (3次), AAG/CTT (5次), AAT/ATT (5次), AAAT/ATTT (2次), AAAAT/ATTTT (2次), AATAT/ATATT (2次).

叉蕊薯蓣叶绿体全基因组共有236个符合条件的SSR位点. 其中单碱基重复有148个, 二碱基重复有63个, 三碱基重复有6个, 四碱基重复有12个, 五碱基重复有4个, 六碱基重复有3个. 重复最多的序列是单碱基重复主要是A/T的重复, 共有71处连续A及75处连续T的单碱基重复, 占所有单碱基重复(148处)的98.7%. 其次依次是AT/AT (44次), AG/CT (15次), AC/GT (4次), AAG/CTT (4次), AAAT/ATTT (4次), AGAT/ATCT (3次). 叉蕊薯蓣叶绿体全基因组中碱基重复序列比薯蓣中多, 薯蓣中未发现六碱基重复, 详情见网络版附表3.

2.4　薯蓣属叶绿体基因组比较及特异性DNA条形码序列筛选

本研究以叉蕊薯蓣为参照基因组利用在线基因组比对工具mVista[37]对7条薯蓣属叶绿体基因组序列进行全局比对(图2). 结果显示, 薯蓣属叶绿体基因组中非编码区(红色)变异高于保守的蛋白编码区域(蓝色), LSC区、SSC区变异明显大于IR区. 根据注释结果, 将7条叶绿体基因组中的蛋白编码基因和基因间隔区分别提取出, 比对分析后发现整体上看, 基因间隔区的种间变异大于蛋白编码基因区域. 综合比较序列长度以及变异程度, 筛选出了10个潜在的种属特异性DNA条形码序列, 其中包括5个蛋白编码基因(*rps19*, *ndhF*, *ycf1*, *rpl32*和*rpl20*)和5个基因间隔区(*matK-trnQ(TTG)*, *trnE(TTC)-trnT(GGT)*, *atpI-rsp2*, *trnT(TGT)-trnF(GAA)*和*ycf1-rps15*), 其序列变异和K2P距离统计分析详见表4. 根据筛选出的高变区序列及DNA条形码序列长度

https://engine.scichina.com/doi/10.1360/N052017-00160

表 3　叉蕊薯蓣叶绿体基因组中密码子

氨基酸	密码子	密码子个数	RSCU	tRNA	氨基酸	密码子	密码子个数	RSCU	tRNA
苯丙氨酸 (Phe/F)	UUU	913	1.22		酪氨酸 (Tyr/Y)	UAU	798	1.60	
	UUC	586	0.78	*trnF-GAA*		UAC	201	0.40	*trnY-GUA*
亮氨酸 (Leu/L)	UUA	839	1.83	*trnL-UAA*	终止(TER)	UAA	39	1.33	
	UUG	578	1.26	*trnL-CAA*		UAG	28	0.95	
	CUU	573	1.25		组氨酸 (His/H)	CAU	495	1.57	
	CUC	195	0.43			CAC	136	0.43	*trnH-GUG*
	CUA	386	0.84	*trnL-UAG*	谷氨酰胺 (Gln/Q)	CAA	713	1.50	*trnQ-UUG*
	CUG	179	0.39			CAG	240	0.50	
异亮氨酸 (Ile/I)	AUU	1129	1.45		天冬酰胺 (Asn/N)	AAU	997	1.57	
	AUC	459	0.59	*trnI-GAU*		AAC	271	0.43	*trnN-GUU*
	AUA	748	0.96	*trnI-CAU*	赖氨酸 (Lys/K)	AAA	1030	1.48	*trnK-UUU*
甲硫氨酸(起始) (Met/M)	AUG	641	1.00	*trn(f)M-CAU*		AGG	169	0.63	
缬氨酸 (Val/V)	GUU	512	1.43		天冬氨酸 (Asp/D)	GAU	876	1.62	
	GUC	156	0.44	*trnV-GAC*		GGC	151	0.34	*trnD-GUC*
	GUA	543	1.52	*trnV-UAC*	谷氨酸 (Glu/E)	GAA	1053	1.49	*trnE-UUC*
	GUG	220	0.61			GGG	290	0.65	
丝氨酸 (Ser/S)	UCU	586	1.67		半胱氨酸 (Cys/C)	UGU	250	1.56	
	UCC	341	0.97	*trnS-GGA*		UGC	70	0.44	*trnC-GCA*
	UCA	468	1.33	*trnS-UGA*	终止(TER)	UGA	21	0.72	
	UCG	188	0.54		色氨酸 (Trp/W)	UGG	456	1.00	*trnW-CCA*
脯氨酸 (Pro/P)	CCU	394	1.46		精氨酸 (Arg/R)	CGU	375	1.39	*trnR-ACG*
	CCC	241	0.90			CGC	87	0.32	
	CCA	316	1.17	*trnP-UGG*		CGA	352	1.31	
	CCG	125	0.46			CGG	111	0.41	
苏氨酸 (Thr/T)	ACU	549	1.60		丝氨酸 (Ser/S)	AGU	426	1.21	
	ACC	250	0.73	*trnT-GGU*		AGC	98	0.28	*trnS-GCU*
	ACA	412	1.20	*trnT-UGU*	精氨酸 (Arg/R)	AGA	521	1.94	*trnR-UCU*
	ACG	159	0.46			AAG	362	0.52	
丙氨酸 (Ala/A)	GCU	604	1.76		甘氨酸 (Gly/G)	GGU	604	1.36	
	GCC	218	0.63			GAC	203	0.38	*trnG-GCC*
	GCA	415	1.21	*trnA-UGC*		GGA	729	1.64	*trnG-UCC*
	GCG	137	0.40			GAG	357	0.51	

https://engine.scichina.com/doi/10.1360/N052017-00160

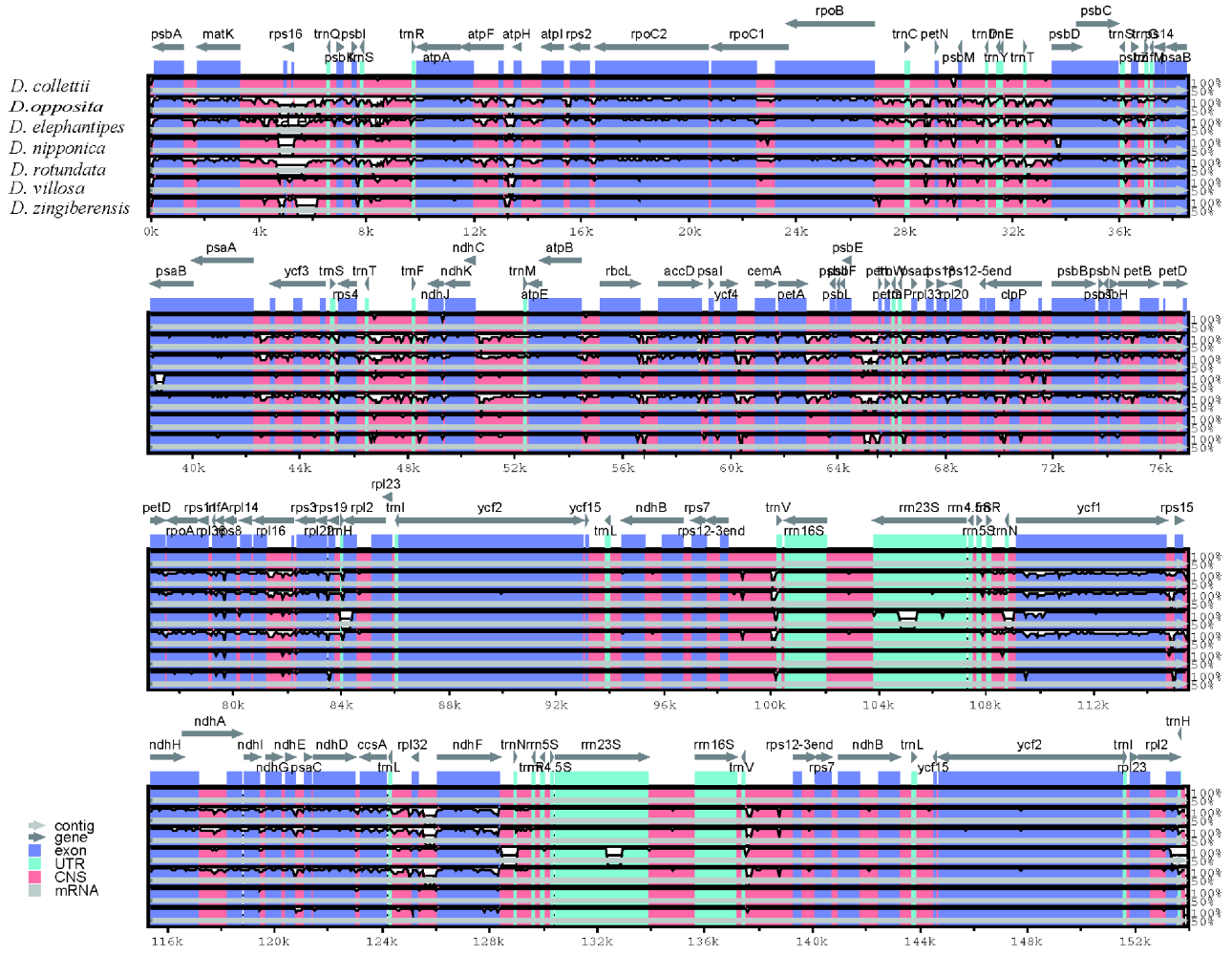


图 2　薯蓣属叶绿体基因组全局比对图

每个基因上的灰色箭头表示该基因的方向; *Y*轴表示50%~100%的相似性; *X*轴表示叶绿体基因组坐标; 不同颜色的波形图表示叶绿体基因组不同区域

的要求设计的特异性DNA条形码序列PCR扩增引物见网络版附表4.

2.5　系统进化分析

利用从GenBank下载的13条单子叶植物叶绿体基因组, 包括百合科青岛百合(*Lilium tsingtauense* (KU230438))、兰科铁皮石斛(*Dendrobium officinale* (KJ862886))、天门冬科知母(*Anemarrhena asphodeloides* (KX931449))和(*Yucca schidigera* (NC_032714))、水鳖科(*Najas flexiles* (NC_021936))、无叶莲科(*Petrosavia stellaris* (KF482381))、棕榈科油棕(*Elaeis guineensis* (NC_017602))、Haemodoraceae科(*Xiphidium caeruleum* (JX088669))、姜科(*Zingiber spectabile* (JX088661))、菖蒲科菖蒲(*Acorus calamus* (AJ879453))、禾本科绿竹(*Bambusa oldhamii* (FJ970915))和(*Oryza alta* (KF359913))、Cyclanthaceae科(*Carludovica palmate* (NC_026786))、2条核心真双子叶植物叶绿体基因组五加科人参(*Panax ginseng* (KF431956))和唇形科丹参(*Salvia miltiorrhiza* (JX312195))、1条基部真双子叶植物毛茛科牛扁(*Aconitum barbatum* var. *puberulum* (KT964698)), 以及1条早期分化的基部被子植物叶绿体基因组*Amborella trichopoda* (NC_005086)和薯蓣科7条叶绿体基因组序列构建ML树(图3). 以早期分化的被子植物*Amborella tri-*

https://engine.scichina.com/doi/10.1360/N052017-00160

表 4　候选序列的序列变异和K2P距离统计分析

候选序列	序列变异			K2P距离值		
	比对后长度(bp)	变异位点数	变异位点占比	最大值	最小值	$\bar{x}$
rps19	330	29	8.79%	0.090	0.004	0.055
ndhF	2333	129	5.53%	0.048	0.004	0.032
ycf1	6435	2201	34.2%	0.637	0.021	0.278
rpl32	174	13	7.47%	0.060	0.012	0.040
rpl20	372	26	6.99%	0.063	0.005	0.039
matK-trnQ(TTG)	3521	376	10.68%	0.054	0.007	0.038
trnE(TTC)-trnT(GGT)	934	65	6.96%	0.073	0.010	0.045
atpI-rsp2	269	18	6.69%	0.053	0.000	0.031
trnT(TGT)-trnF(GAA)	1818	124	6.83%	0.057	0.005	0.038
ycf1-rps15	413	140	33.90%	0.522	0.039	0.254

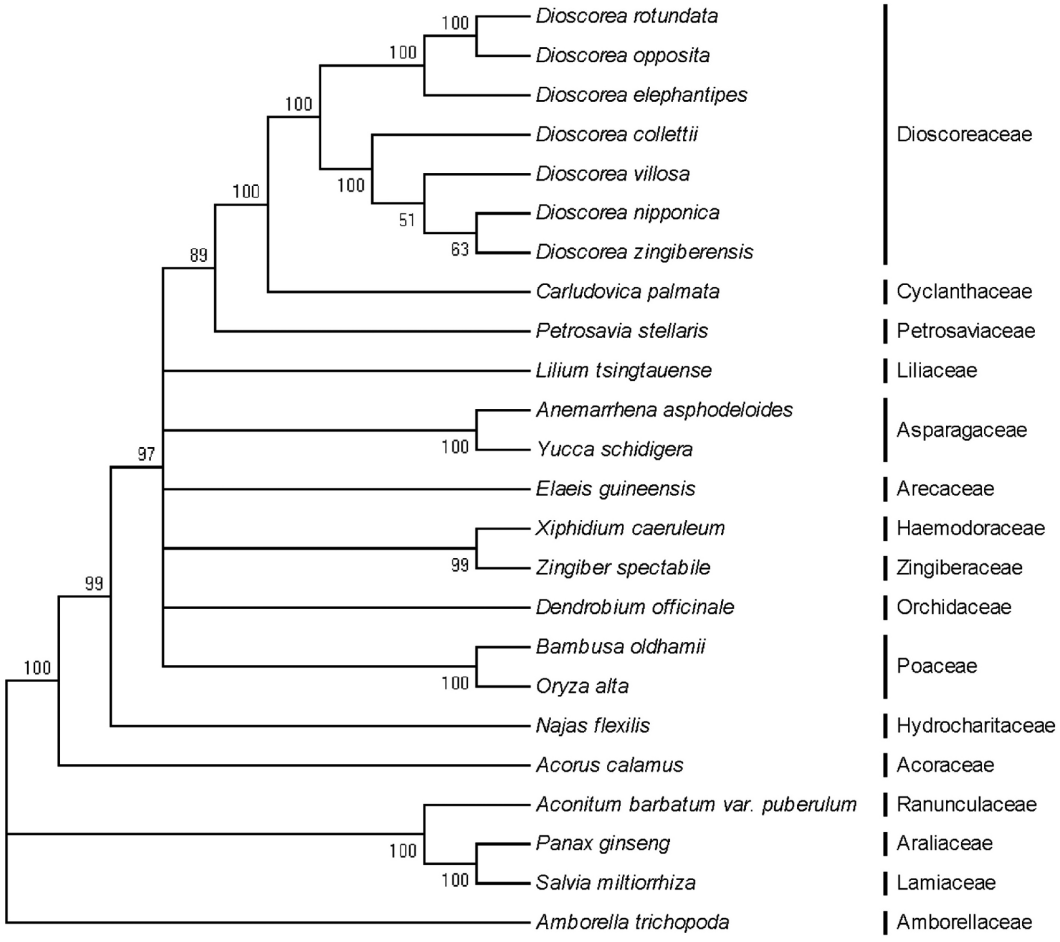


图 3　基于叶绿体基因组构建的薯蓣科和其他植物的ML树

https://engine.scichina.com/doi/10.1360/N052017-00160

*chopoda*作为外类群, 从进化树可以看出单子叶植物与真双子叶植物各自单独聚为一支, 支持率为100%. 在单子叶植物这一支中, 薯蓣科与Cyclanthaceae和无叶莲科亲缘关系更近. 系统发育上显示薯蓣科7个物种聚为一支, 且薯蓣与*D. rotundata*聚为一支, 表明薯蓣和*D. rotundata*亲缘关系较近, 薯蓣科7个种之间彼此能够区分开. 叶绿体基因组序列为薯蓣属植物的分子鉴定提供了一个新方法, 可以作为薯蓣属植物鉴定的超级DNA条形码序列.

3 讨论

根据“内共生”理论, 学术界普遍认为叶绿体是由被原始真核生物吞噬的古细菌蓝藻经过共生形成, 且越来越多的证据表明叶绿体的最终形成可能经历了多次“内共生”[42~44]. 叶绿体基因组不到核基因组大小的万分之一, 且具有多拷贝性, 可使目的基因高水平表达, 成为药用植物超级条形码鉴定的研究热点, 对揭示植物分子标记开发、物种亲缘关系以及推动植物在分子育种、遗传转化等叶绿体基因工程方面的研究进程具有重要意义[24,45].

本研究采用了先提取植物总DNA进行测序, 然后再抽取叶绿体基因组序列读长进行拼接的方法. 此方法与直接提取叶绿体DNA的方法相比, 减少了样品的使用, 避免了叶绿体DNA提取中容易出现的核DNA和其他细胞器DNA污染的现象, 而且省去了较为繁琐的纯化步骤, 很大程度上提高了DNA的提取效率. 其次, 对总DNA测序数据量达2G以上, 而叶绿体大小才150 kb左右, 占总DNA的5%, 所以采取对总DNA进行测序完全可以获得准确的叶绿体全基因组序列.

虽然目前已报道了薯蓣属5个物种的叶绿体基因组, 但本研究是首次对薯蓣属植物叶绿体基因组序列的高变区进行分析筛选. 叶绿基因组具有高度保守性, 但物种间仍存在一定的差异. 本文测定并分析了薯蓣属植物薯蓣和叉蕊薯蓣的叶绿体全基因组, 并利用在线基因组比对工具mVista对7条薯蓣属叶绿体基因组进行高变区分析和筛选. 由序列比对结果可知, 薯蓣属叶绿体基因组中蛋白编码区序列变异程度低于非编码区. tRNA及rRNA序列高度保守, 在属内变异程度小. 另外*psbA*, *rpoB*, *psbC*, *psaB*, *psaA*, *ycf4*, *ycf2*, *ndhB*, *rps7*基因的保守性高, 变异程度较小; *rps16*, *ndhF*, *accD*及*ycf1*基因变异程度较大; *matK-trnR*, *atpF-atpI*, *trnL-ndhF*基因间隔区变异较大. 综合比较序列长度以及变异程度, 筛选出了10个适合作为分子标记的候选序列, 其中包括5个蛋白编码基因和5个基因间隔区. 对于筛选的DNA条形码序列, 不仅需要考虑序列的变异程度, 还要综合序列长度和PCR扩增和测序的实验要求等[26]. 前期研究表明, 目前常用的DNA条形码序列对薯蓣属物种的鉴定效率很低[20], 本研究通过叶绿体基因组筛选的DNA条形码在薯蓣属物种中的鉴定效率有待于收集更多的该属物种通过常规DNA条形码方法进一步验证, 最终确定该属物种鉴定的特异性DNA条形码. 由此可见, 对叶绿体全基因组进行比对能够更加全面地反映薯蓣属物种的序列变异情况, 为从叶绿体基因组序列中筛选特异性DNA条形码鉴定序列提供了参考. 叶绿体基因组作为植物特有的相对独立于核基因组之外的基因结构, 其本身富含大量的遗传信息, 对于研究植物系统进化分析具有重要的意义[46,47], 极大地提高了对被子植物系统进化关系的认知. 近年来, 已有大量研究通过比较多种蛋白编码基因及叶绿体基因组完整序列的方法来解决被子植物的系统进化问题, 为研究被子植物之间神秘复杂的进化关系提供了新方法[48~50]. 本研究利用薯蓣属5条已发表的叶绿体基因组序列及本研究测得的2条序列在内的共24条叶绿体基因组序列构建ML系统进化树, 以探讨薯蓣属植物与其他被子植物之间的系统进化位置关系. 总的来说, 在ML树中单子叶植物与双子叶植物均单独聚为一支; *Amborella*作为被子植物的早期分支, 在ML树中单独聚为一支, 是其他被子植物的姐妹群, 表现出单系性, 此结果与Hansen等人[29]的研究结果相同; 在单子叶植物中, Cyclanthaceae与薯蓣科聚为一大支, 亲缘关系较近; Haemodoraceae与姜科聚为一大支. 同时, 薯蓣科物种单独聚为一支, 能够很好地与其他物种区分开, 且属内7个物种各自单独聚为一小支, 能够彼此分开. 分析结果为了解薯蓣科物种的进化位置提供了参考.

薯蓣属药用植物叶绿体基因组测序不仅可以丰富叶绿体基因组序列的数量, 也能为该属药用植物抗性蛋白的表达、代谢途径的改造等方面提供参考, 更重要的是可以推动薯蓣属药用植物在分子育种、遗传转化以及系统进化等方面的研究进程. 另外, 为研究薯蓣属药用植物的生长发育、培育抗逆等优良品种、实现该属药用资源的可持续利用奠定基础.

https://engine.scichina.com/doi/10.1360/N052017-00160

参考文献

1 中国科学院中国植物志编辑委员会. 中国植物志. 北京: 科学出版社, 1985, 16: 54
2 国家药典委员会. 中华人民共和国药典(一部). 北京: 中国医药科技出版社, 2015. 28–29
3 冯学锋, 黄璐琦, 格小光, 等. 山药道地药材形成源流考. 中国中药杂志, 2008, 33: 859–862
4 蔡金辉, 严淅子, 黄晓辉, 等. 山药品种资源的分类研究. 江西农业大学学报, 1999, 21: 55–57, 59
5 郑玉红, 夏冰, 杭悦宇, 等. 山药原植物薯蓣及其近缘种的分子鉴别和亲缘关系研究. 南京农业大学学报, 2007, 30: 55–59
6 孙华钦, 罗科, 邹文俊, 等. 穿龙薯蓣、黄山药和盾叶薯蓣*psbA-trnH*片段序列分析. 应用与环境生物学报, 2006, 12: 792–797
7 王晓荣, 李湧健, 程彬彬. 薯蓣皂苷元抗肿瘤作用及其机制研究. 西部中医药, 2014, 27: 140–143
8 Huang C H, Ku C Y, Jan T R. Diosgenin attenuates allergen-induced intestinal inflammation and IgE production in a murine model of food allergy. Planta Med, 2009, 75: 1300–1305
9 Li F, Fernandez P P, Rajendran P, et al. Diosgenin, a steroidal saponin, inhibits STAT3 signaling pathway leading to suppression of proliferation and chemosensitization of human hepatocellular carcinoma cells. Cancer Lett, 2010, 292: 197–207
10 张楠, 张杰, 刘志礼, 等. 盾叶冠心宁片治疗高脂血症的临床试验研究. 中西医结合心脑血管病杂志, 2009, 7: 225
11 许丽娜, 卫永丽, 彭金咏. 天然产物薯蓣皂苷的研究进展. 中国中药杂志, 2015, 40: 36–41
12 Gao X, Zhu Y P, Wu B C, et al. Phylogeny of *Dioscorea* sect. *Stenophora* based on chloroplast *matK*, *rbcL* and *trnL-F* sequences. J Syst Evol, 2008, 46: 315–321
13 吴志刚, 范传颖, 包晓青, 等. 山药种质资源核糖体rDNA-ITS区序列分析. 中草药, 2014, 45: 1136–1142
14 陈士林, 姚辉, 韩建萍, 等. 中药材DNA条形码分子鉴定指导原则. 中国中药杂志, 2013, 38: 141–148
15 Chen S, Pang X, Song J, et al. A renaissance in herbal medicine identification: from morphology to DNA. Biotech Adv, 2014, 32: 1237–1244
16 Yao H, Song J, Liu C, et al. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE, 2010, 5: e13102
17 Chen S, Yao H, Han J, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE, 2010, 5: e8613
18 郑玉红, 夏冰, 杭悦宇, 等. 山药原植物薯蓣及其近缘种的分子鉴别和亲缘关系研究. 南京农业大学学报, 2007, 30: 55–59
19 梁欣健. 南板蓝根与山药的DNA条形码鉴定研究. 广州中医药大学, 2012
20 Sun X Q, Zhu Y J, Guo J L, et al. DNA barcoding the *Dioscorea* in China, a vital group in the evolution of monocotyledon: use of *matK* gene for species discrimination. PLoS ONE, 2012, 7: e32057
21 Wu F H, Chan M T, Liao D C, et al. Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. BMC Plant Biol, 2010, 10: 68
22 Kuang D Y, Wu H, Wang Y L, et al. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. Genome, 2011, 54: 663–673
23 Nock C J, Waters D L E, Edwards M A, et al. Chloroplast genome sequences from total DNA for plant identification. Plant Biotech J, 2011, 9: 328–333
24 Takano A, Okada H. Phylogenetic relationships among subgenera, species, and varieties of Japanese *Salvia* L. (Lamiaceae). J Plant Res, 2011, 124: 245–252
25 Li X, Yang Y, Henry R J, et al. Plant DNA barcoding: from gene to genome. Biol Rev, 2015, 90: 157–166
26 李滢, 姚辉, 宋经元, 等. 基于叶绿体全基因组的贝母属特异性DNA条形码的筛选. 世界科学技术-中医药现代化, 2016, 18: 24–28
27 Dong W, Liu J, Yu J, et al. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS ONE, 2012, 7: e35071
28 Pauwels M, Vekemans X, Godé C, et al. Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). New Phytol, 2012, 193: 916–928
29 Hansen D R, Dastidar S G, Cai Z, et al. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). Mol Phylogenet Evol, 2007, 45: 547–563
30 Mariac C, Scarcelli N, Pouzadou J, et al. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. Mol Ecol Resour, 2014, 14: 1103–1113

https://engine.scichina.com/doi/10.1360/N052017-00160

31 Wu L, Wang B, Yang J, et al. The chloroplast genome sequence of an important medicinal plant *Dioscorea nipponica*. Mitochondr DNA A DNA Mapp Seq Anal, 2016, 27: 2559–2560

32 Zhou W, Chen C, Hua W P, et al. The complete chloroplast genome sequence of *Dioscorea zingiberensis* (Dioscoreceae). Mitochondr DNA A DNA Mapp Seq Anal, 2016, 27: 2730

33 Liu C, Shi L, Zhu Y, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics, 2012, 13: 715

34 Tamura K, Stecher G, Peterson D, et al. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol, 2013, 30: 2725–2729

35 杨丽, 赵韩生, 彭镇华, 等. 竹子叶绿体基因组SSR分子标记的开发及其应用. 热带亚热带植物学报, 2014, 22: 263–269

36 Tang D Q, Lu J J, Fang W, et al. Development, characterization and utilization of GenBank microsatellite markers in *Phyllostachys pubescens* and related species. Mol Breeding, 2010, 25: 299–311

37 Frazer K A, Pachter L, Poliakov A, et al. Vista: computational tools for comparative genomics. Nucleic Acids Res, 2004, 32: W273

38 Panchal H, Anjaria K, Patel S. Phylogenetic analysis of some leguminous trees using CLUSTALW2 bioinformatics tool. IEEE Computer Society, 2012. 917–921

39 Katoh K, Kuma K, Toh H, et al. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res, 2005, 33: 511–518

40 Sugiura M, Hirose T, Sugita M. Evolution and mechanism of translation in chloroplasts. Annu Rev Genet, 1998, 32: 437–459

41 Liu Q B, Xue Q Z. Codon usage in the chloroplast genome of rice (*Oryza sativa* L. ssp. *japonica*). Acta Agron Sin, 2004, 30: 1220–1224

42 于惠敏, 解相林. 叶绿体的“内共生”与“基因转移”现象. 生物学通报, 2007, 42: 17–19

43 燕安, 俞利凤. 叶绿体基因组: 起源, 结构与表达调控. 生物学通报, 2004, 39: 6–8

44 Margulis L. The colonization hypothesis. (Book reviews: origin of eukaryotic cells. Evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth. Bryologist, 1972, 75: 675–676

45 Tangphatsornruang S, Sangsrakru D, Chanprasert J, et al. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. DNA Res, 2009, 17: 11–22

46 Wu M, Lan S, Cai B, et al. The complete chloroplast genome of *Guadua angustifolia* and comparative analyses of neotropical-paleotropical bamboos. PLoS ONE, 2015, 10: e0143792

47 Qian J, Song J, Gao H, et al. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. PLoS ONE, 2013, 8: e57607

48 郭豪杰, 刘久石, 罗丽, 等. 五味子叶绿体基因组结构解析与比较分析. 中国科学: 生命科学, 2017, 47: 728–739

49 Zhou J, Chen X, Cui Y, et al. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two aristolochia medicinal species. Int J Mol Sci, 2017, 18: 1839

50 Shen X, Wu M, Liao B, et al. Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. Molecules, 2017, 22: 1330

https://engine.scichina.com/doi/10.1360/N052017-00160

Complete chloroplast genomes of *Dioscorea opposite* and *D. collettii* and screening specific DNA barcodes

MA ShuangJiao[1], ZHOU JianGuo[1], LI Ying[1], CHEN XinLian[1], WU MingLi[2], SUN Wei[3], LI YongHua[4], SONG JingYuan[1] & YAO Hui[1]

1 Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People's Republic of China, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China;
2 Pharmacy Faculty, Hubei University of Chinese Medicine, Wuhan 430065, China;
3 Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China;
4 Department of Pharmacy, Guangxi University of Chinese Medicine, Nanning 530001, China

Previous studies of DNA barcoding *Dioscorea* species have shown that the identification efficiency was low. In this study, the complete chloroplast genomes of *D. opposite* and *D. collettii* were sequenced using Illumina HiSeq 2000. Genome structure and content and high variation regions were analyzed. The complete chloroplast genome lengths of *D. opposite* and *D. collettii* were 152963 and 153870 bp, which include four parts: two inverted repeats (IRs), one large single copy (LSC) and one small single copy (SSC). Both of *D. opposite* and *D. collettii* encoded 125 genes, including 87 protein coding genes, 30 tRNA genes and eight rRNA genes. The GC contents of *D. opposite* and *D. collettii* were 37.04% and 37.17%, respectively. The result showed that the sequence variation of non-coding regions is higher than that of the protein coding region and the variations of the LSC and SSC are higher than that of the IRs. Ten molecular markers of *Dioscorea*, which include five protein coding regions and five non-coding regions were screened to be used as the potential DNA barcodes for authenticating *Dioscorea* species. Phylogenetic analyses showed that the chloroplast genome can be used as an ultra-barcode to distinguish *Dioscorea* species from each other. This study laid the foundation for super barcode utilization in the *Dioscorea* and provided a molecular basis for further investigation on this important medicinal species.

***Dioscorea opposite*, *Dioscorea collettii*, complete chloroplast genome, ultra-barcode, identification**

doi: 10.1360/N052017-00160

https://engine.scichina.com/doi/10.1360/N052017-00160