草果叶绿体基因组特征及系统发育分析

马孟莉[1, 2]，张　薇[1, 2]，孟衡玲[1, 2]，卢丙越[1, 2*]

1. 云南省高校滇南特色生物资源研究与利用重点实验室，云南 蒙自　661199
2. 红河学院 生物科学与农学学院，云南 蒙自　661199

摘　要：目的　明确草果 *Amomum tsao-ko* 叶绿体基因组结构特征及其在姜科的进化地位。**方法**　利用 Illumina Hiseq 4000 测序平台对草果叶绿体基因组进行测序，通过生物信息学分析方法进行序列组装、注释和特征分析，并以绿苞闭鞘姜为外类群，通过 MEGA6 软件构建 ML 系统发育树，分析姜科物种间系统发育关系。**结果**　草果叶绿体基因组全长 163 648 bp，鸟嘌呤和胞嘧啶含量（guanine-cytosine content，GC）为 36.0%，包括一对 29 776 bp 的反向重复区（IRs）、一个大单拷贝区（LSC，88 741 bp）和一个小单拷贝区（SSC，15 355 bp）；共注释得到 113 个基因，包括 4 个 rRNA 基因、30 个 tRNA 基因和 79 个蛋白编码基因；在草果叶绿体基因组中共检测到 123 个 SSR 位点，大部分 SSR 均由 A 和 T 组成；豆蔻属物种叶绿体基因组大小和结构相似，IR 边界高度保守；聚类分析显示草果与豆蔻属的海南砂 *A. longiligulare*、阳春砂 *A. villosum*、绿壳砂 *A. villosum* var. *xanthioides*、爪哇白豆蔻 *A. compactum* 和白豆蔻 *A. kravanh* 亲缘关系最近；适应性进化分析发现 rpl20、rps11、ccsA、clpP、ycf1 和 ycf2 6 个基因在进化过程中被正向选择。**结论**　获得了完整的草果叶绿体基因组序列，明确了姜科物种间的亲缘关系，为研究豆蔻属植物的系统进化及物种鉴定提供科学依据。

关键词：姜科；草果；海南砂；阳春砂；绿壳砂；爪哇白豆蔻；白豆蔻；叶绿体基因组；系统发育

中图分类号：R282.12　　**文献标志码：**A　　**文章编号：**0253 - 2670(2021)19 - 6023 - 09

DOI：10.7501/j.issn.0253-2670.2021.19.025

Characterization and phylogenetic analysis of the complete chloroplast genome of *Amomum tsao-ko*

MA Meng-li[1, 2], ZHANG Wei[1, 2], MENG Heng-ling[1, 2], LU Bing-yue[1, 2]

1. Key Laboratory of Biological Resources Research and Utilization in Southern Yunnan, Mengzi 661199, China
2. College of Biological and Agricultural Sciences, Honghe University, Mengzi 661199, China

Abstract: Objective　To clarify the structural characteristics of chloroplast genome in *Amomum tsao-ko* and its evolutionary status within the Zingiberaceae. **Methods**　The chloroplast genome of *A. tsao-ko* was sequenced by Illumina Hiseq 4000 sequencing platform, and then assembled, annotated, and characterized by bioinformatic methods. The ML phylogenetic tree was constructed by MEGA6 software with *Costus viridis* as outgroup, and the phylogenetic relationship among Zingiberaceae species was analyzed. **Results**　The total length of chloroplast genome was 163648bp and the GC content was 36.0%, including a pair of 29776bp reverse repeat region (IRs), a large single copy region (LSC;88741bp) and a small single copy region (SSC;15355bp). A total of 113 genes were annotated, including 4 rRNA genes, 30 tRNA genes and 79 protein coding genes. 123 SSR loci were detected in the *A. tsao-ko* chloroplast genome, and most of the SSR was composed of A and T. The chloroplast genomes of genus *Amomum* are similar in size and structure, and the IR boundary is highly conservative. The cluster analysis showed that *A. tsao-ko* was closely related to *A. longiligulare*, *A. villosum*, *A. villosum* var. *xanthioides*, *A. compactum* and *A. kravanh*. rpl20, rps11, ccsA, clpP, ycf1 and ycf2 are undergoing positive selection in the process of evolution. **Conclusion** The complete chloroplast genome sequence of *A. tsao-ko* was obtained in this study, which clarified the genetic relationship among species of Zingiberaceae, and provided scientific basis for the study of phylogeny and species identification of genus *Amomum*.

Key words: Zingiberaceae; *Amomum tsao-ko* Crevost et Lemarie; *A. longiligulare* T. L. Wu; *A. villosum* Lour.; *A. villosum* var. *xanthioides* (Wall. ex Bak.) T. L. Wu et S. J. Chen; *A. compactum* Sol. Ex Maton; *A. kravanh* Pierre ex Gagnep; chloroplast genome; phylogenetic analysis

收稿日期：2021-02-13

基金项目：国家自然科学基金项目（31460380）；云南省地方本科高校基础研究联合专项资金面上项目（202001BA070001-181）；云南省万人计划青年拔尖人才专项；红河学院中青年学术骨干培养项目（2014GG0101）；红河学院重点学科开放基金项目（2018ZDXK02）

作者简介：马孟莉（1985—），女，讲师，主要从事药用植物资源研究。Tel: 18760753175　E-mail: mamlsky@126.com

***通信作者：**卢丙越，教授，主要从事药用植物种质资源挖掘与利用方面研究。Tel: 13769430165　E-mail: lby202@126.com

姜科为泛热带分布科，是多年生草本植物，由 52 属约 1377 种植物组成，在我国主要分布在云南、海南、广东和广西等省区。姜科中包括很多著名药材，如益智、草果、砂仁、姜、草豆蔻、郁金、姜黄、莪术、高良姜等，现代药学研究表明姜科除具有行气破瘀、温中散寒、舒筋活络、芳香健胃等功效外，还具有抗癌、抗炎、抗菌、抗氧化、镇痛、降血压等药理作用[1]。草果 *Amomum tsao-ko* Crevost et Lemarie 是姜科豆蔻属多年生常绿丛生草本植物，全株具有辛辣味，以干燥成熟果实入药，具有清湿化痰、温脾祛寒的功效，此外草果也是烹调佐料中的佳品，被誉为食品调料中的“五香之一”[2-4]。草果生长在海拔 1100～1800 m 的热带、亚热带荫蔽潮湿的林中，主要分布在中国的云南、广西和贵州，老挝、越南也有分布[5-6]。由于过度收获及草果原生境的破坏，草果野生资源近乎绝迹，2012 年已被世界自然保护联盟濒危物种红色名录列为“近危物种”。

叶绿体（chloroplast，cp）是植物细胞内的小细胞器，为绿色植物进行光合作用的场所，叶绿体有自己的遗传系统，由一个封闭的环状 DNA 分子组成。近年来，叶绿体基因组因其保守的基因序列和在植物中的重要作用而被广泛用于物种的鉴定和系统进化分析[7-8]。随着高通量 DNA 测序技术的发展，可用的叶绿体基因组序列数量越来越多，这为药用植物叶绿体基因组研究提供了重要参考。到目前为止，已有姜科姜属（姜、蜂巢姜）、山姜属（艳山姜、益智）、姜黄属（观音姜、黄花姜黄）和豆蔻属（爪哇白豆蔻、白豆蔻、阳春砂、绿壳砂和海南砂）叶绿体基因组测序的报道[9-14]。随着姜科植物叶绿体全基因组数量的增加，姜科其他物种的叶绿体基因组序列也越来越容易拼接。然而，草果作为姜科中重要的药食同源物种，国内外对草果遗传方面的研究甚少，尤其是分子遗传方面，严重限制了草果资源的保护、开发和利用。本研究首次详细报道了草果叶绿体基因组的完整结构，比较草果和其他豆蔻属植物叶绿体基因组差异及 IR 区的伸缩与扩张情况，此外通过叶绿体全基因组序列构建进化树来揭示草果与其他姜科植物的进化关系及其在系统发育中的地位。

1 材料

草果新鲜幼嫩的叶片采自云南省红河哈尼族彝族自治州金平苗族瑶族傣族自治县阿得博乡草果山村（22°54′30.34″N，103°13′16.39″E），经云南省高校滇南特色生物资源研究与利用重点实验室张薇教授鉴定为草果 *A. tsao-ko* Crevost et Lemarie，标本存放于红河学院生物科学与农学学院（标本号：LBY20180526）。

2 方法

2.1 基因组 DNA 提取及测序

取大约 5 g 的新鲜草果叶片，经 70%乙醇处理后用 TIANGEN 植物 DNA 提取试剂盒提取草果基因组总 DNA，通过琼脂糖凝胶电泳和 NanoDrop2000 微量分光光度计检测 DNA 的纯度和浓度，符合测序要求后送测序公司在 Illumina HiSeq4000 平台进行测序。

2.2 叶绿体基因组组装和注释

测序完成后得到的原始序列（raw reads）首先利用 NGS QC ToolKit 过滤去除接头及两端的低质量序列，得到高质量待分析序列（即 clean reads）。以白豆蔻（NC_036935.1）叶绿体基因组序列作为参考序列，利用 SOAPdenovo2.04 软件进行组装，再根据 reads 的 paired-end 和 overlap 关系对组装结果进行优化，经多次调整获得草果完整叶绿体基因组。使用 DOGMA 在线工具对叶绿体基因进行注释[15]，使用默认参数预测蛋白质编码基因、转移 RNA（tRNA）基因和核糖体 RNA（rRNA）基因，并结合京都基因和基因组百科全书（kyoto encyclopedia of genes and genomes，KEGG）、蛋白质直系同源簇（cluster of orthologous groups of proteins，COG）、非冗余蛋白序列（non-redundant protein sequences，NR）、Swiss-Prot 蛋白质序列（swiss-prot protein sequence，Swiss-Prot）、基因本体（gene ontology，GO）等数据库对基因进行功能注释，用 OGDRAW v1.2 在线软件绘制草果叶绿体基因组图谱。将注释好的序列用 BankIt 向 NCBI 在线提交，获得序列登录号 MK926774.1。

2.3 叶绿体基因组特征分析

利用 MISA 软件（http://pgrc.ipk-gatersleben.de/misa/misa.html）鉴定草果叶绿体基因组中的简单重复序列（simple sequence repeats，SSR），搜索参数设置为：含有完全重复的单核苷酸最小重复数为 10，二核苷酸最小重复数为 5，三核苷酸最小重复数为 4，四、五、六核苷酸最小重复数为 3；另外设置 2 个 SSR 之间的最小距离为 100 bp，如果距离小于 100 bp，则 2 个 SSR 被当做一个复合微卫星。利用 IRscope（https://irscope.shinyapps.io/irapp/）比较不同物种叶绿体基因组 IR 边界区特征[16]。

2.4 系统发育分析

从 NCBI 下载姜科豆蔻属的海南砂 *Amomum longiligulare* T. L. Wu（MN067434.1）、阳春砂

Amomum villosum Lour.（MH161418.1）、绿壳砂 *Amomum villosum* var. *xanthioides* (Wall. ex Baker) T. L. Wu & S. J. Chen（MH161417.1）、爪哇白豆蔻 *Amomum compactum* Sol. ex Maton（MG000589.1）、白豆蔻 *Amomum kravanh* Pierre ex Gagnep.（NC_036935.1），山姜属的艳山姜 *Alpinia zerumbet* (Pers.) B. L. Burtt & R. M.Sm.（JX088668.1）和益智 *Alpinia oxyphylla* Miq.（KY985237.1），姜属的蜂巢姜 *Zingiber spectabile* Griff.（JX088661.1）和姜 *Zingiber officinale* Rosc.（MH161428.1），姜黄属的观音姜 *Curcuma roscoeana* Wall.（NC_022928.1）和黄花姜黄 *Curcuma flaviflora* S. Q. Tong（KR967361.1），山柰属的山柰 *Kaempferia galanga* Linn.（MK209001.1）和紫花山柰 *Kaempferia elegans*（Wall.）Baker（MK209002.1）共 13 个物种完整的叶绿体基因组序列，同时以姜目闭鞘姜科植物绿苞闭鞘姜 *Costus viridis* S.Q.Tong（MK262733.1）为外类群，利用 MAFFT7.037 软件进行序列多重比对，结果经手工检查与调整后用最大似然法（Maximum Likelihood，ML）法对系统进化关系进行分析。ML 系统发育树用 MEGA6 软件生成，选择最佳模型 GTR＋G＋I，自展值 Bootstrap value 设为 1000。

2.5 适应性进化分析

通过非同义替换位点替换次数（*Ka*）与同义替换位点替换次数（*Ks*）的比值（*Ka*/*Ks*）判断草果与同属的阳春砂、绿壳砂、海南砂、白豆蔻和爪哇白豆蔻之间叶绿体蛋白编码基因是否存在选择压力，当 *Ka*/*Ks*＞1，认为有正选择效应；*Ka*/*Ks*＝1，认为存在中性选择；*Ka*/*Ks*＜1，则认为有纯化选择作用。首先利用 PhyloSuite v1.2.1 提取草果、阳春砂、绿壳砂、海南砂、白豆蔻和爪哇白豆蔻的 79 个共有蛋白质编码基因[17]，提取的基因序列通过 MAFFT7.037 软件进行比对，然后用 DnaSP 6.12.03 软件计算 *Ka* 和 *Ks* 值，通过 Excel 软件统计各基因的 *Ka*/*Ks* 值，绘制不同功能基因的 *Ka*/*Ks* 图。

3 结果与分析

3.1 基因组基本特征

草果叶绿体基因组与绝大多数被子植物叶绿体基因组一样，为共价闭合的双链环状分子，全 163 648 bp，包括 1 对反向重复（IR）区（29 776 bp）、1 个大的单拷贝（LSC）区（88 741 bp）和 1 个小的单拷贝（SSC）区（15 355 bp）。全基因组的 GC 含量为 36.0%，其中 IR 区 GC 含量最高（41.1%），LSC 区（33.7%）和 SSC 区（29.8%）均较低。在草果叶绿体基因组中共注释到 113 个非重复基因，包括 79 个蛋白质编码基因、4 个 rRNA 基因和 30 个 tRNA 基因（表 1）。LSC 区包含的基因最多，包括 61 个蛋白编码基因和 21 个 tRNA 基因；SSC 区包括 1 个 tRNA 基因（trnL-UAG）和 11 个蛋白编码基因；所有的 rRNA 基因、8 个 tRNA 基因和 8 个蛋白编码基因均在 IR 区重复。其中 rps12 基因具有反式剪切结构，其 5’端位于 LSC 区，而 3’端位于 IR 区（图 1 和表 1）。

图 1 草果叶绿体基因组图谱

Fig. 1 Gene map of *Amomum tsao-ko* chloroplast genome

表 1　草果叶绿体基因组基因

Table 1　Genes encoded cp genome in *Amomum tsao-ko*

基因分类	基因分组	基因名称	基因数量
自我复制	rRNA	rrn4.5[c]、rrn5[c]、rrn16[c]、rrn23[c]	8
	tRNA	trnA-UGC[ac]、trnC-GCA、trnD-GUC、trnE-UUC、trnF-GAA、trnfM-CAU、trnG-GCC[a]、trnG-UCC、trnH-GUG[c]、trnI-CAU[c]、trnI-GAU[ac]、trnK-UUU[a]、trnL-CAA[c]、trnL-UAA[a]、trnL-UAG、trnM-CAU、trnN-GUU[c]、trnP-UGG、trnQ-UUG、trnR-ACG[c]、trnR-UCU、trnS-GCU、trnS-GGA、trnS-UGA、trnT-GGU、trnT-UGU、trnV-GAC[c]、trnV-UAC[a]、trnW-CCA、trnY-GUA	38
	核糖体小亚基	rps2、rps3、rps4、rps7[c]、rps8、rps11、rps12[bc]、rps14、rps15、rps16[a]、rps18、rps19[c]	15
	核糖体大亚基	rpl2[ac]、rpl14、rpl16[a]、rpl20、rpl22、rpl23[c]、rpl32、rpl33、rpl36	11
	RNA 聚合酶亚基	rpoA、rpoB、rpoC1[a]、rpoC2	4
光合作用相关	光合系统 1	psaA、psaB、psaC、psaI、psaJ	5
	光合系统 2	psbA、psbB、psbC、psbD、psbE、psbF、psbH、psbI、psbJ、psbK、psbL、psbM、psbN、psbT、psbZ	15
	细胞色素复合物	petA、petB[a]、petD[a]、petG、petL、petN	6
	ATP 合酶	atpA、atpB、atpE、atpF[a]、atpH、atpI	6
	二磷酸核酮糖羧化酶大亚基	rbcL	1
	NADH 脱氢酶	ndhA[a]、ndhB[ac]、ndhC、ndhD、ndhE、ndhF、ndhG、ndhH、ndhI、ndhJ、ndhK	12
其他基因	成熟酶	matK	1
	包裹膜蛋白	cemA	1
	乙酰辅酶 A 羧化酶亚基	accD	1
	c 型细胞色素合成	ccsA	1
	依赖 ATP 的蛋白酶单元 p	clpP[b]	1
	转录起始因子	infA	1
未知功能	假定叶绿体阅读框	ycf1[c]、ycf2[c]、ycf3[b]、ycf4	6

a-基因含有 1 个内含子　b-基因含有 2 个内含子　c-有 2 个拷贝基因

a-genes with one intron　b-genes with two introns　c-genes with two copies

内含子在基因表达调控中发挥重要作用，草果叶绿体基因组中有 18 个含内含子的基因，包括 6 个 tRNA 基因（trnA-UGC、trnG-GCC、trnI-GAU、trnK-UUU、trnL-UAA、trnV-UAC）和 12 个蛋白质编码基因（rps12、rps16、rpl2、rpl16、rpoC1、petB、petD、atpF、clpP、ndhA、ndhB、ycf3），其中 rps12、clpP 和 ycf3 基因包含 2 个内含子（表 1）。

3.2　SSR 分析

SSR 广泛分布于叶绿体基因组中。草果叶绿体基因组中共检测到 123 个 SSR，其中单核苷酸重复 SSR 最多（62 个），其次是二核苷酸 SSR（37 个），四核苷酸 SSR 有 20 个，三核苷酸 SSR 最少，仅检测到 4 个。在所检测的 SSR 中以 A/T、AT/AT 和 AAAT/ATTT 为重复单元的占 83.74%，表明草果叶绿体 SSR 偏好使用 A 和 T 碱基（表 2）。

3.3　豆蔻属叶绿体基因组比较

将草果叶绿体基因组与已报道的豆蔻属的阳春砂、绿壳砂、海南砂、白豆蔻和爪哇白豆蔻进行比较（表 3），结果表明豆蔻属物种间基因组大小差异较小，其中阳春砂的叶绿体基因组最长，为 164 069 bp，白豆蔻的最短为 162 766 bp；除白豆蔻 LSC 区较短外，其余 5 个物种的 LSC 区长度差异较小，而 SSC 区长度在 6 个物种间相近；比较发现草果的 IR 区长度最短，为 29 776 bp，而阳春砂的 IR 区最长，为 29 959 bp；草果、阳春砂、绿壳砂和海南砂叶绿体基因组均包含 133 个基因，白豆蔻和爪哇白豆蔻则为 135 个基因；6 个物种的叶绿体基因组 GC 含量相似，均在 36.0%～36.1%。

将草果叶绿体基因组的 IR-LSC 和 IR-SSC 边界与已报道的 5 个豆蔻属物种进行比较（图 2）。豆蔻属叶绿体基因组的 6 个边界较为保守，其中 LSC/IRa 和 LSC/IRb 边界的侧翼基因相同，LSC/IRa 边界均在基因 rpl22 和 rps19 基因之间，而 LSC/IRb 边界位于 rps19 和 psbA 基因之间。SSC/IRa 边界 IRa 一侧都存在一个 ycf1 假基因（Ψycf1），除草果外，其余 5 个物种的 Ψycf1 基因均跨越 SSC/IRa 区，Ψycf1 假

表 2 草果叶绿体基因组 SSR 信息

Table 2 SSR in the *Amomum tsao-ko* cp genome

重复类型（数量）	重复序列	重复次数													总数
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	
单核苷酸（62）	A/T	–	–	–	–	–	–	–	33	16	8	2				59
	C/G	–	–	–	–	–	–	–	1		1				1	3
二核苷酸（37）	AG/CT	–	–	2												2
	AT/AT	–	–	20	7	3	3	2								35
三核苷酸（4）	AAG/CTT	–	3													3
	AGG/CCT	–	1													1
四核苷酸（20）	AAAC/GTTT	1														1
	AAAG/CTTT	4														4
	AAAT/ATTT	9														9
	AACT/AGTT	1														1
	AATG/ATTC	2														2
	AATT/AATT	2														2
	ACAT/ATGT		1													1

表 3 6 个豆蔻属叶绿体基因组一般特征比较

Table 3 Comparison of the general features of the six genus *Amomum* chloroplast genomes

基因组特征	草果	阳春砂	绿壳砂	海南砂	白豆蔻	爪哇白豆蔻
GenBank 登录号	MK926774.1	MH161418.1	MH161417.1	MN067434.1	NC_036935.1	MG000589.1
全长/bp	163 648	164 069	163 985	163 608	162 766	163 553
LSC 长度/bp	88 741	88 798	88 857	88 680	87 728	88 535
SSC 长度/bp	15 355	15 353	15 369	15 288	15 390	15 370
IR 长度/bp	29 776	29 959	29 948	29 820	29 824	29 824
总基因数	133	133	133	133	135	135
蛋白编码基因数	79	79	79	79	80	80
tRNA 基因数	30	30	30	30	30	29
rRNA 基因数	4	4	4	4	4	4
GC 含量/%	36.0	36.1	36.0	36.1	36.1	36.0

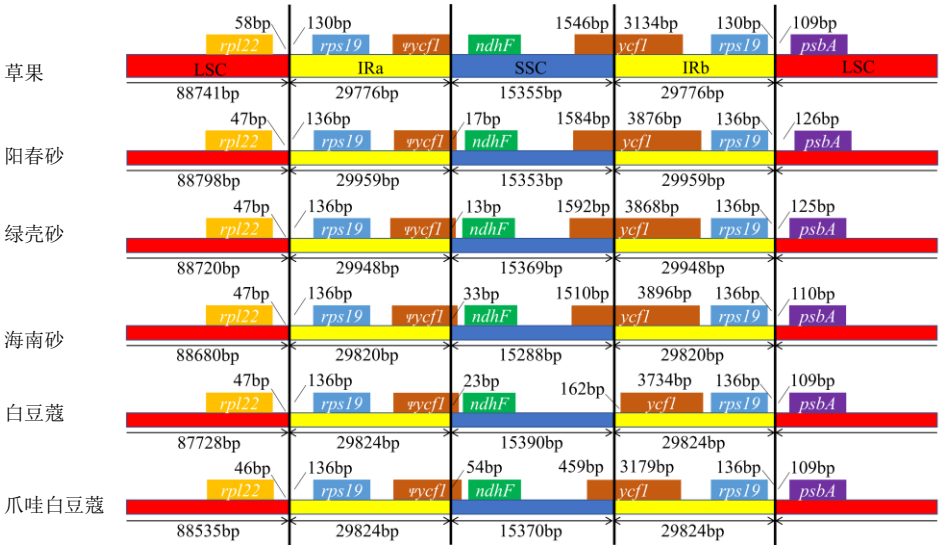

图 2 6 个豆蔻属叶绿体基因组 LSC、SSC 和 IR 边界比较

Fig. 2 Comparison of LSC, SSC and IR borders among six chloroplast genome of genus *Amomum*

基因进入 SSC 区的长度从 13 bp 到 54 bp，其中爪哇白豆蔻 Ψycf1 扩张最大；草果、阳春砂、绿壳砂、海南砂、爪哇白豆蔻的 SSC/IRb 均位于 ycf1 基因中，而白豆蔻 ycf1 基因已完全扩张到 IRa 区内部。

3.4 系统进化分析

为了明确草果在姜科植物的系统发育地位和进

化关系，选取已报道的 13 个姜科物种的叶绿体全基因组序列，以闭鞘姜科的绿苞闭鞘姜为外类群，构建 ML 系统发育树。所有姜科物种可分为 2 大类，豆蔻属和山姜属聚为一类，豆蔻属中同为砂仁来源的绿壳砂、阳春砂和海南砂亲缘关系最近，聚为一支，爪哇白豆蔻和白豆蔻聚为另一支，草果单独成一支，草果与其它豆蔻属的分支支持率为 99%；在另一类中包括姜属、姜黄属和山柰属，姜黄属的黄花姜黄先与山柰属的 2 个物种聚在一起，再与姜属的蜂巢姜和姜聚在一起，最后与姜黄属的观音姜聚类（图 3）。

3.5　适用性进化分析

为进一步研究草果与豆蔻属物种叶绿体基因在进化过程中受到的选择压力，利用 DnaSP 软件分析草果与阳春砂、绿壳砂、海南砂、白豆蔻和爪哇白豆蔻蛋白编码基因的 *Ka*/*Ks* 值（图 4）。79 个蛋白编码基因在草果 *vs* 阳春砂、草果 *vs* 绿壳砂、草果 *vs* 海南砂、草果 *vs* 白豆蔻和草果 *vs* 爪哇白豆蔻的 *Ka*/*Ks* 均值分别为 0.169、0.168、0.170、0.158 和

图 3　基于 15 个物种叶绿体基因组序列构建的系统发育树

Fig. 3　Phylogenetic tree of 15 species based on chloroplast genomes

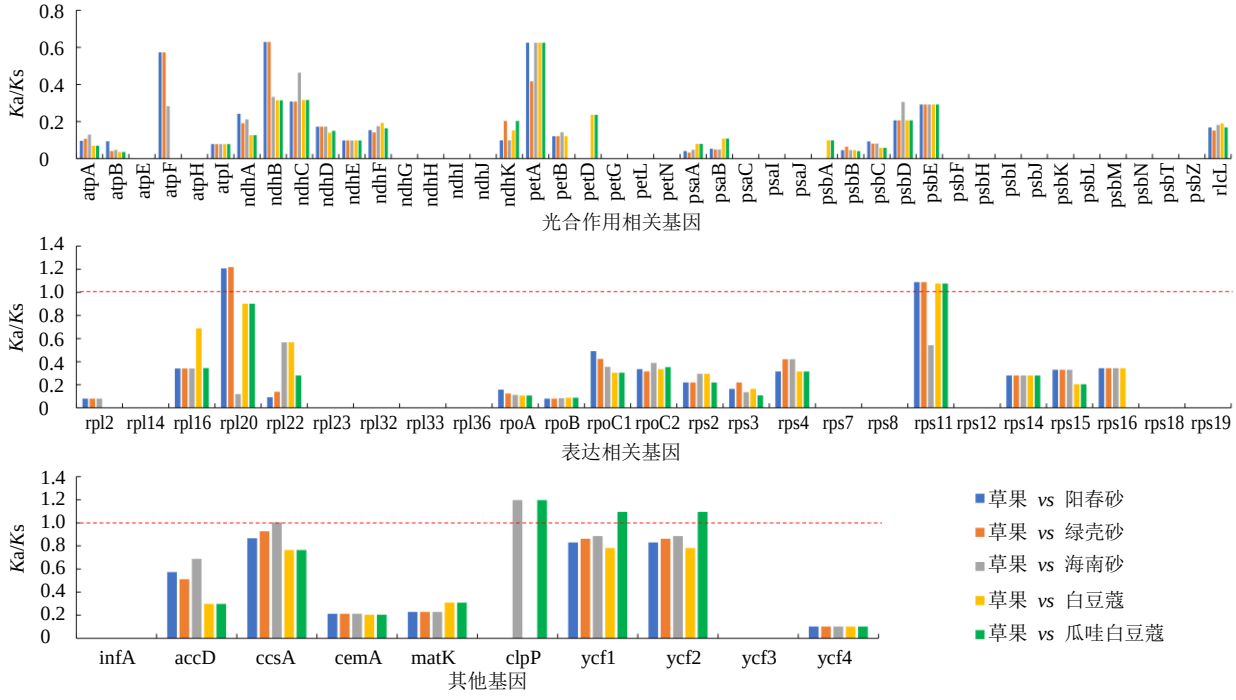

图 4　不同功能基因的 *Ka*/*Ks* 值

Fig. 4　*Ka*/*Ks* values of different functional genes

0.166，其中绝大多数基因 *Ka*/*Ks*＜1，表明豆蔻属物种叶绿体基因在长期的进化过程中受到了较强的纯化选择。进一步分析表明光合作用相关基因的 *Ka*/*Ks* 值均小于 1；表达相关基因 rpl20 在草果 *vs* 阳春砂和草果 *vs* 绿壳砂中 *Ka*/*Ks*＞1，除草果 *vs* 海南砂的 rps11 基因 *Ka*/*Ks*＜1 外，草果与其他豆蔻属物种的 rps11 基因 *Ka*/*Ks* 均大于 1；其他功能基因中 ccsA 基因在草果 *vs* 海南砂中 *Ka*/*Ks*＞1、clpP 基因在草果 *vs* 海南砂和草果 *vs* 爪哇白豆蔻中 *Ka*/*Ks*＞1、ycf1 和 ycf2 基因均在草果 *vs* 爪哇白豆蔻中 *Ka*/*Ks*＞1，表明这些基因在进化过程中受到正向选择作用。

4 讨论

通过比较草果和已报道姜科豆蔻属物种的叶绿体基因组，发现豆蔻属在叶绿体基因组上具有高度的保守性，已报道的豆蔻属 cp 基因组长度在 162 766～164 069 bp[9-11]，本研究组装的草果 cp 基因组全序列长度为 163 648 bp，表明草果叶绿体基因组大小符合豆蔻属物种的特点。已报道的豆蔻属叶绿体总基因数量在 133～135（非重复蛋白编码基因 79～80），在本研究中，草果叶绿体基因组注释到 133 个基因（113 个非重复基因，包括 79 个蛋白编码基因、30 个 tRNA 基因和 4 个 rRNA 基因）。豆蔻属物种间叶绿体基因组 GC 含量相似，均在 36.0%～36.1%，并且 LSC 区（33.7%～33.9%）和 SSC 区（29.8%～30.1%）的 GC 含量显著低于 IR 区（41.1%），其主要原因与 8 个 GC 含量较高的 rRNA 基因均分布在 IR 区有关。

一般来说，IR 区是叶绿体基因组中最保守的区域。IR 区、LSC 区和 SSC 区的扩张和收缩是进化过程中常见的现象，也是造成叶绿体基因组长度差异的主要原因。在本研究中，发现 6 个豆蔻属物种叶绿体基因组的 IR 与 LSC 或 SSC 之间的边界变化很小。6 个豆蔻属植物的 LSC/IRb 边界均在基因 rpl22 和 rps19 基因之间，IRa/LSC 边界均位于 rps19 基因和 psbA 基因之间，与姜科的观音姜、黄花姜黄、益智、姜的边界基因相似，说明姜科物种 IR 区与 LSC/IRa、IRa/SSC、IRb/LSC 区的边界保守性较高[13]，但也有姜科物种的 IR 边界基因存在较大的差异，如蜂巢姜的 IRb/LSC 边界只有 rps19 基因，且该基因位于 LSC 区，而 LSC/IRa 边界的 IRa 一侧没有 rps19 基因[10]，此外棕榈科和禾本科竹亚科也存在这一现象[18-19]。

叶绿体中 SSR 拷贝数的变异是一种重要的分子标记，比核和线粒体微卫星有更大的分类距离，在植物群体遗传学、多态性研究和进化研究中有着广泛的应用[20]。目前已有应用 cpSSR 研究药用植物丹参、三七、桃儿七、霍山石斛等的报道[21-24]。本研究利用 MISA 软件在一、二、三、四、五和六核苷酸最小重复数设置为 10、5、4、3、3、3 情况下共扫描到 123 个 SSR，其中有 105 个 SSR 由 A/T、AT/TA、AAAT/ATTT 和 AATT/AATT 组成，这表明草果 cp 基因组中的 SSR 包含频繁的 A 或 T 重复，这与已报到的其它植物叶绿体基因组相似[25-27]。叶绿体 DNA 在由亲代向子代传递过程中不涉及基因重组，且具有保守性、单亲遗传等特点，在进化过程中有独立的进化路线，因此，具有 cpDNA 特征的 cpSSR 技术在植物群体结构分析、种群分类、物种演化等研究领域有明显的优势[28]。草果 cpSSR 的获得对进一步研究姜科药用植物遗传多样性、群体结构、分子鉴定等方面具有重要意义。

植物叶绿体基因组仅次于核基因组，含有大量的遗传信息，叶绿体全基因组测序技术为药用植物进化系统研究提供了新的平台及思路，近年来，姜科物种中的爪哇白豆蔻、白豆蔻、砂仁、姜、山柰、益智等叶绿体基因组序列的不断公布为确定姜科药用植物的进化和系统发育关系提供了重要依据。本研究从 GenBank 下载了 13 个姜科植物叶绿体基因组序列，以闭鞘姜科的绿苞闭鞘姜为外类群，构建了姜科物种的系统发育树，从系统发育分析看，姜科物种被分为 2 个大的类群，豆蔻属的 6 个物种与山姜属的 2 个物种被聚在一起，表明豆蔻属和山姜属在姜科中的亲缘关系较近，这一结果与 Wu 等[10]、Cui 等[11]、Li 等[12]和 Gao 等[13]的研究结果一致；聚类图也清晰的展现了豆蔻属 6 个物种之间的亲缘关系，《中国药典》2015 年版中作为砂仁来源的阳春砂、绿壳砂和海南砂被聚在一起，做豆蔻使用的白豆蔻和爪哇白豆蔻亲缘关系最近，草果则在豆蔻属中单独成一支，基于叶绿体基因组的聚类分析从分子角度证实了亲缘关系近的物种在药用成分具有相似性，可替代使用。此外研究也表明叶绿体基因组可有效区分姜科物种的系统发育关系，也为深入研究姜科植物进化、遗传多样性和群体结构提供了参考。

Ka/*Ks* 是评估蛋白质编码基因是否发生适用性进化的有效方法[29]。生物大多数基因的同义核苷酸

替换比非同义替换发生得更频繁，因此 *Ka*/*Ks* 值通常小于 1[30]。本研究共检测到 6 个正选择基因，正选择基因中没有检测到与光合作用相关的基因，这可能与豆蔻属物种生长于热带地区林下湿热环境，利用相似的光照条件进行光合作用有关[4]。在正选择基因中与基因表达相关的 rpl20 和 rps11 分别编码 L20 和 S11 亚基，特别是 rps11 基因在草果 *vs* 阳春砂、草果 *vs* 绿壳砂、草果 *vs* 白豆蔻、草果 *vs* 爪哇白豆蔻中 *Ka*/*Ks* 均大于 1，说明该基因近期正发生快速的进化，深入研究该基因对草果进化具有一定的意义。clpP 基因在草果 *vs* 海南砂和草果 *vs* 爪哇白豆蔻中也经历着正选择，该基因编码 Clp 蛋白水解酶亚基，其功能主要负责降解异常蛋白，与维持叶绿体正常代谢有关[31]；而在草果 *vs* 海南砂中检测出的 ccsA 基因负责编码细胞色素 c 合成蛋白，Xie 等[32]研究认为 ccsA 基因参与细胞色素 c 与血红蛋白的结合，这为理解该基因参与豆蔻属物种适用性进化提供了参考。ycf1 和 ycf2 作为叶绿体基因组中 2 个编码区序列最长的基因，其编码产物功能至今仍不清楚，ycf1 和 ycf2 基因在多个植物中均存在正向选择现象，推测这 2 个基因在陆生植物适应环境变化中发挥重要作用[10, 33-36]。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 张俊清, 李友宾. 中国药用姜科植物 [M]. 北京: 中国医药科技出版社, 2015: 123.

[2] Lim T K. *Amomum* tsao-ko [A] // *Edible Medicinal And Non-Medicinal Plants* [M]. Dordrecht: Springer Netherlands, 2012: 813-817.

[3] 胡一凡, 张雪梅, 石乃星 等. 云南草果种植资源 DNA 条形码序列分析 [J]. 中草药, 2019, 50(24): 6091-6097.

[4] 杨绍兵, 张金渝. 草果生产加工适宜技术 [M]. 北京: 中国医药科技出版社, 2018: 265.

[5] 吴德邻. 中国姜科植物资源 [M]. 武汉: 华中科技大学出版社, 2016: 263.

[6] Liu G Z, Jin M, Cai C T, *et al*. Soil microbial community structure and physicochemical properties in *Amomum tsaoko*-based agroforestry systems in the gaoligong mountains, southwest China [J]. *Sustainability*, 2019, 11(2): 546.

[7] Xue S, Shi T, Luo W J, *et al*. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina* [J]. *Hortic Res*, 2019, 6: 89.

[8] Cui Y X, Nie L P, Sun W, *et al*. Comparative and phylogenetic analyses of ginger (*Zingiber officinale*) in the family Zingiberaceae based on the complete chloroplast genome [J]. *Plants (Basel)*, 2019, 8(8): E283.

[9] Wu M L, Li Q, Xu J, *et al*. Complete chloroplast genome of the medicinal plant *Amomum compactum*: Gene organization, comparative analysis, and phylogenetic relationships within Zingiberales [J]. *Chin Med*, 2018, 13: 10.

[10] Wu M L, Li Q, Hu Z G, *et al*. The complete *Amomum kravanh* chloroplast genome sequence and phylogenetic analysis of the commelinids [J]. *Molecules*, 2017, 22(11): E1875.

[11] Cui Y X, Chen X L, Nie L P, *et al*. Comparison and phylogenetic analysis of chloroplast genomes of three medicinal and edible *Amomum* species [J]. *Int J Mol Sci*, 2019, 20(16): E4040.

[12] Li D M, Zhao C Y, Liu X F. Complete chloroplast genome sequences of *Kaempferia galanga* and *Kaempferia elegans*: Molecular structures and comparative analysis [J]. *Molecules*, 2019, 24(3): E474.

[13] Gao B M, Yuan L, Tang T L, *et al*. The complete chloroplast genome sequence of *Alpinia oxyphylla* Miq. and comparison analysis within the Zingiberaceae family [J]. *PLoS One*, 2019, 14(6): e0218817.

[14] Gui L J, Jiang S F, Xie D F, *et al*. Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution [J]. *Gene*, 2020, 732: 144355.

[15] Wyman S K, Jansen R K, Boore J L. Automatic annotation of organellar genomes with DOGMA [J]. *Bioinformatics*, 2004, 20(17): 3252-3255.

[16] Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics*, 2018, 34(17): 3030-3031.

[17] Zhang D, Gao F L, Jakovlić I, *et al*. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies [J]. *Mol Ecol Resour*, 2020, 20(1): 348-355.

[18] de Santana Lopes A, Gomes Pacheco T, Nimz T, *et al*. The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae [J]. *Planta*, 2018, 247(4): 1011-1030.

[19] Wang W C, Chen S Y, Zhang X Z. Whole-genome comparison reveals divergent IR Borders and mutation hotspots in chloroplast genomes of herbaceous bamboos (bambusoideae: Olyreae) [J]. *Molecules*, 2018, 23(7): E1537.

[20] 王化坤, 娄晓鸣, 章镇. 叶绿体微卫星在植物种质资源研究中的应用 [J]. 分子植物育种, 2006, 4(S1): 92-98.

[21] 沙秀芬, 彭芳, 陶珊, 等. 丹参线粒体和叶绿体微卫星标记开发及多样性分析 [J]. 西北植物学报, 2018, 38(12): 2215-2223.

[22] 韩岩. 基于 cpDNA 分子标记的三七居群遗传多样性研究 [D]. 昆明: 昆明理工大学, 2018.

[23] 张盾, 任梦云, 张银东, 等. 濒危藏药桃儿七 cpSSR-PCR 引物开发与反应体系优化 [J]. 分子植物育种, 2018, 16(9): 2886-2894.

[24] 刘枫, 赵群, 戴军, 等. 霍山石斛 cpDNA 全序列微卫星分布及分子鉴别研究 [J]. 中药材, 2016, 39(12): 2720-2724.

[25] 杨俏俏, 姜梅, 王立强, 等. 药食两用菡头叶绿体基因组解析、比较基因组学及系统发育研究 [J]. 药学学报, 2019, 54(1): 173-181.

[26] 张明英, 王西芳, 高静, 等. 美丽芍药叶绿体全基因组解析及系统发育分析 [J]. 药学学报, 2020, 55(1): 168-176.

[27] 左文明, 曾阳, 杨春芳, 等. 基于高通量技术的唐古特大黄叶绿体全基因组测序及应用研究 [J]. 中草药, 2019, 50(22): 5545-5553.

[28] Provan J, Powell W, Hollingsworth P M. Chloroplast microsatellites: New tools for studies in plant ecology and evolution [J]. *Trends Ecol Evol*, 2001, 16(3): 142-147.

[29] 王博, 高磊, 苏应娟, 等. 基于叶绿体基因组全序列分析真叶植物叶绿体基因的适应性进化 [J]. 中山大学学报: 自然科学版, 2012, 51(3): 108-113.

[30] Makalowski W, Boguski M S. Evolutionary parameters of the transcribed mammalian genome: An analysis of 2, 820 orthologous rodent and human sequences [J]. *PNAS*, 1998, 95(16): 9407-9412.

[31] Yu A Y, Houry W A. ClpP: a distinctive family of cylindrical energy-dependent serine proteases [J]. *FEBS Lett*, 2007, 581(19): 3749-3757.

[32] Xie Z, Merchant S. The plastid-encoded ccsA gene is required for heme attachment to chloroplast c-type cytochromes [J]. *J Biol Chem*, 1996, 271(9): 4632-4639.

[33] Huang Y, Wang J, Yang Y, *et al*. Phylogenomic analysis and dynamic evolution of chloroplast genomes in Salicaceae [J]. *Front Plant Sci*, 2017, 8: 1050.

[34] Dong W L, Wang R N, Zhang N Y, *et al*. Molecular evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution [J]. *Int J Mol Sci*, 2018, 19(3): 716.

[35] Yan C, Du J, Gao L, *et al*. The complete chloroplast genome sequence of watercress (*Nasturtium officinale* R. Br.): Genome organization, adaptive evolution and phylogenetic relationships in Cardamineae [J]. *Gene*, 2019, 699: 24-36.

[36] Lee H O, Joh H J, Kim K, *et al*. Dynamic chloroplast genome rearrangement and DNA barcoding for three apiaceae species known as the medicinal herb “Bang-Poong” [J]. *Int J Mol Sci*, 2019, 20(9): 2196.

[责任编辑 时圣明]