

大黄药材基原物种叶绿体基因组分析与特异DNA条形码开发

李冉郡^{1,2}, 武立伟², 辛天怡², 廖海¹, 林余霖², 姚辉², 周嘉裕^{1*}, 宋经元^{2,3*}

(1. 西南交通大学生命科学与工程学院, 四川 成都 610031; 2. 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; 3. 中药资源教育部工程研究中心, 北京 100193)

摘要: 大黄是我国常用大宗药材之一, 其基原植物为唐古特大黄、药用大黄和掌叶大黄。不同基原的大黄药材活性成分和药效存在差异, 为快速准确鉴定大黄药材3个基原物种, 本文利用Illumina高通量测序技术对大黄药材基原物种叶绿体基因组进行测序, 完成其组装注释与结构特征解析, 并基于叶绿体基因组高变区开发精准鉴别大黄药材3个基原物种的特异DNA条形码, 最后进行验证。结果如下: 唐古特大黄、药用大黄和掌叶大黄叶绿体基因组全长分别为161 039 bp、161 093 bp和161 136 bp, 呈典型四分体结构, 均编码131个基因, 包括86个蛋白质编码基因、37个tRNA基因和8个rRNA基因。基于高变区设计的5对引物均可对42份样品有效扩增, 测序结果分析证明*rps16-trnQ*、*psaA-ycf3*、*psbE-petL*、*ndhF-rpl32*与*trnT-trnL*均可作为特异DNA条形码准确鉴定唐古特大黄、药用大黄和掌叶大黄。本文可为大黄药材3个基原物种分类鉴定、保证大黄药材临床用药安全及规范大黄药材市场提供依据。

关键词: 大黄; 叶绿体基因组; 基原鉴定; 特异DNA条形码

中图分类号: R 931 文献标识码: A 文章编号: 0513-4870(2022)05-1495-11

Analysis of chloroplast genomes and development of specific DNA barcodes for identifying the original species of *Rhei Radix et Rhizoma*

LI Ran-jun^{1,2}, WU Li-wei², XIN Tian-yi², LIAO Hai¹, LIN Yu-lin², YAO Hui²,
ZHOU Jia-yu^{1*}, SONG Jing-yuan^{2,3*}

(1. School of Life and Science, Southwest Jiaotong University, Chengdu 610031, China; 2. Key Laboratory 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; 3. Engineering Research Center of Chinese Medicine Resource of Ministry of Education, Beijing 100193, China)

Abstract: *Rhei Radix et Rhizoma* is one of the most used medicinal materials in China. Its original species are *Rheum palmatum*, *Rh. tanguticum*, and *Rh. officinale*. *Rhei Radix et Rhizoma* derived from different original species are significantly different in their active ingredients and pharmacological effects. To develop an accurate, rapid, and specific identification method, we obtained the chloroplast genomes of the three original species by Illumina Novaseq sequencing. We designed specific DNA barcodes from the hypervariable regions, which can accurately identify the three original species. The experimental results showed that the total length of the chloroplast genomes of *Rh. tanguticum*, *Rh. officinale* and *Rh. palmatum* were 161 039 bp, 161 093 bp, and 161 136 bp, respectively. All the three genomes were represented as typical quadripartite structures. A total of 131 genes, including 86 protein-coding genes, 37 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes were identified from each

收稿日期: 2021-10-22; 修回日期: 2022-02-16.

基金项目: 国家自然科学基金青年基金项目 (82003906); 国家重点研发计划 (2019YFC1711100).

*通讯作者 E-mail: spinezhou@home.swjtu.edu.cn; jysong@implad.ac.cn

DOI: 10.16438/j.0513-4870.2021-1525

chloroplast genome. Five pairs of primers based on the hypervariable regions were designed to efficiently amplify 42 samples. Results confirmed that five hypervariable regions, *rps16-trnQ*, *psaA-ycf3*, *psbE-petL*, *ndhF-rpl32*, and *trnT-trnL*, can be used as specific DNA barcodes for the identification of *Rh. tanguticum*, *Rh. officinale*, and *Rh. palmatum*. These results provided genetic information for further species identification of *Rhei Radix et Rhizoma*, and improve the safety of this clinical medication as well as standardize the market for *Rhei Radix et Rhizoma*.

Key words: *Rhei Radix et Rhizoma*; chloroplast genome; original species identification; specific DNA barcode

《中国药典》收载药材大黄 *Rhei Radix et Rhizoma* 为蓼科植物掌叶大黄 *Rheum palmatum* L.、唐古特大黄 *Rh. tanguticum* Maxim. ex Balf. 和药用大黄 *Rh. officinale* Baill. 的干燥根及根茎^[1]。大黄作为药材的记载始见于东汉的《神农本草经》，药用历史悠久^[2]。现代药理研究表明大黄有抗病毒、抗肿瘤、抗炎、抗氧化应激、抗细胞凋亡和调节机体免疫等功能^[3]。国家食品药品监督管理局公开中成药配方有 900 余种中成药含有大黄^[4]。大黄具有广泛药效作用主要与其含有大量活性成分有关，大黄主要化学成分为蒽醌类、蒽酮类、二苯乙烯类、鞣质类等化合物^[5]。不同基原的大黄药材在化学物质上存在差别^[6]，在不同功效上表现出专长，研究表明掌叶大黄专长于“清热泻火、利湿退黄”，而唐古特大黄专长于“泻下攻积、逐瘀通经”^[4]。因此大黄基原物种鉴别对于大黄临床精准用药具有重要意义。

大黄药材基原物种鉴定方法主要有性状鉴定、显微鉴定、理化鉴定及分子鉴定。大黄药材 3 个基原物种区别主要在于叶片分裂程度^[7]，药用大黄掌状浅裂，呈大齿形或宽三角形；掌叶大黄叶片浅裂至半裂，呈较窄三角形；唐古特大黄叶深裂，裂片狭长呈三角状披针形或狭长形^[8]。性状和显微鉴定对大黄药材基原物种或药材特征的完整性要求较高，鉴别过程需依靠长期经验积累，且经过深加工的大黄饮片及中成药，形态特征破坏会增加基原物种鉴定难度。理化鉴定 (HPLC、UPLC 等)^[9,10] 依据大黄药材主要成分进行鉴定，但操作较为繁琐。与传统鉴定方法相比，DNA 分子能为物种鉴定提供丰富信息，通过 ITS2、*psbA-trnH*、*matK* 等 DNA 条形码^[11-13] 可有效鉴别正品大黄与华北大黄 *Rh. franzenbachii* Munt.、河套大黄 *Rh. hotaense* C. Y. Cheng et Kao 等混伪品，但难以精确到基原物种^[14]。

叶绿体基因组可作为“超级条形码”进行物种鉴定^[15]。绝大多数被子植物叶绿体基因组表现为典型双链环状四分体结构^[16]，基因组长度通常介于 110~170 kb 之间^[17]，其编码基因数一般为 110~150 个，且在基因组成、基因含量及排列顺序上高度保守^[18]。与传统相对较短并易扩增的 DNA 条形码相比，叶绿体基因组具有更丰富变异位点信息和更强物种分辨率，可更准确反应各物种遗传特性。因此本研究拟基于叶绿体基因

组开发有效鉴别唐古特大黄、掌叶大黄和药用大黄的特异 DNA 条形码，以期用于大黄基原植物、药材、饮片及相关中成药的基原鉴定，为保证大黄药材临床用药安全、规范大黄药材市场及保护大黄种质资源提供依据。

材料与方法

植物材料 用于叶绿体基因组测序的样品，唐古特大黄新鲜叶片采自四川省阿坝州红原县邛溪镇，掌叶大黄新鲜叶片采自甘肃省陇南市宕昌县理川镇，药用大黄新鲜叶片采自四川省凉山州冕宁县冶勒乡。用于验证特异 DNA 条形码的样品除大黄药材 3 个基原物种外，还包括食用大黄 *Rheum rhabarbarum* Linnaeus、波叶大黄 *Rh. rhabarbarum* Linnaeus、苞叶大黄 *Rh. alexandrae* Batal.、塔黄 *Rh. nobile* Hook. f. et Thoms. 及皱叶酸模 *Rumex crispus* L. 共 42 份，详见表 1。样品由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。

基因组 DNA 提取与测序 利用 E.Z.N.A.[®] Plant DNA kit (D3485, OMEGA, USA) 提取总 DNA，1.0% 的琼脂糖凝胶电泳检测 DNA 质量，Qubit 3.0 (Thermo Fisher Scientific, USA) 检测 DNA 浓度。构建插入片段为 350 bp 的文库，利用 Illumina NovaSeq 测序平台进行双末端测序，双端各测 150 bp。

叶绿体基因组组装、拼接与注释 利用 FastQC 对测序所得原始序列进行质量检测，采用 Trimmomatic 软件^[19] 过滤去除接头序列和低质量区。使用 NOVO-Plasty^[20] 将原始读长组装成完整的叶绿体基因组。采用 BWA 将高质量读长比对回叶绿体基因组序列，并在 IGV 中进行人工检查以确保组装无误。利用 CPGAVAS 软件^[21] 和 GeSeq 软件^[22] 对完整叶绿体基因组序列进行注释，tRNA 基因用 tRNAscan-SE 软件^[23] 进行鉴定及人工校正。利用 Organellar Genome DRAW (<http://ogdraw.mpimp-golm.mpg.de/cgi-bin/ogdraw.pl>) 在线绘制叶绿体基因组图谱。

密码子偏好性分析 采用软件 CodonW (<http://codonw.sourceforge.net>) 分析密码子使用情况。通过

Table 1 Samples used to verify the specific DNA barcodes

Species	Sample part	Number	Sample No.	Location
<i>Rheum officinale</i>	Leaf	6	PS2902MT01-06	Mianning, Sichuan
	Root	4	PS2902MT07-10	Mianning, Sichuan
<i>Rh. palmatum</i>	Leaf	10	PS2903MT01-10	Dangchang, Gansu
	Seed	5	PS2903MT11-15	Dangchang, Gansu
<i>Rh. tanguticum</i>	Leaf	3	PS2904MT01-03	Hongyuan, Sichuan
	Seed	3	PS2904MT04-06	Hongyuan, Sichuan
	Leaf	5	PS2904MT07-11	Linchuan, Qinghai
	Root	1	PS2904MT12	Linchuan, Qinghai
<i>Rh. rhaponticum</i>	Leaf	1	PS2905MT01	Beijing
<i>Rh. rhabarbarum</i>	Leaf	1	PS2906MT01	Beijing
<i>Rh. alexandrae</i>	Leaf	1	PS2907MT01	Linzhi, Tibet
<i>Rh. nobile</i>	Leaf	1	PS2908MT01	Linzhi, Tibet
<i>Rumex crispus</i>	Leaf	1	PS2909MT01	Beijing

计算 Relative synonymous codon usage (RSCU) 获得密码子的偏好性值。密码子 RSCU 值等于 1.00, 表示该密码子无偏好性; RSCU 值大于 1.00 则表示该密码子使用频率较高, 反之亦然。

长重复序列 (long repeat) 和简单序列重复 (SSR) 分析 使用软件 REPuter (<http://bibiserv.techfak.uni-bielefeld.de/reputer/>) 进行长重复序列分析, 查找正向重复 F (forward)、反向重复 R (reverse)、互补重复 C (complement)、回文重复 P (palindromic) 四种重复类型。所有重复序列识别长度不小于 30 bp, 且序列相似度不小于 90%。利用 MISA 软件 (<http://pgrc.ipk-gatersleben.de/misa/>) 测定 SSR 位点类型与数目, 对于单、二、三、四、五和六核苷酸重复序列, 重复次数的最小值分别设置为 10、5、4、3、3 和 3, 两个 SSR 之间最小距离设置为 0 bp。

叶绿体基因组比较分析 应用在线比对工具 mVISTA 对测序获得的唐古特大黄、药用大黄、掌叶大黄叶绿体基因组进行全局比对分析, 利用 DnaSP 软件检测 3 条叶绿体基因组的核苷酸变异程度 (Pi)。

特异 DNA 条形码鉴定分析 为评估叶绿体基因组所筛选高变区的实际物种鉴定效率, 以筛选的高变区序列为模板设计引物 (表 2), 并利用收集的 42 份样本分别进行 PCR 扩增验证。PCR 扩增采用 25 μ L 反应混合体系, 包括 2 \times Taq MasterMix (北京艾德莱生物科技) 12.5 μ L, 上下游引物 (2.5 μ mol \cdot L⁻¹) 各 1 μ L, DNA 模板约 50 ng, 加灭菌双蒸水补足体积至 25 μ L。PCR 程序为 94 $^{\circ}$ C, 5 min; (94 $^{\circ}$ C, 30 s; 55 $^{\circ}$ C, 30 s; 72 $^{\circ}$ C, 45 s), 35 个循环; 72 $^{\circ}$ C, 10 min。扩增产物在 ABI3730XL 测序仪上进行 Sanger 测序。利用 CodonCode Aligner 软件对测序所得峰图文件进行序列拼接。为评估这些序列的鉴定能力, 利用 MEGA 7.0 软件进行序列比对、计算 K2P 遗传距离和构建邻接法 (neighbor-joining, NJ) 系统进化树, 设置 bootstrap 重复值为 1 000。

Table 2 Primer pairs used to distinguish original species of *Rheum Radix et Rhizoma*

Region	Forward/ Reverse	Primer sequence (5' to 3')
<i>rps16-trnQ</i>	1F	GGTTTAGCGGCTACGTACATAG
	1R	GTGATCTAATACAGGCAAGATAGG
<i>psaA-ycf3</i>	2F	CCGGACACGACATACAAAAGA
	2R	TTCCTACGGGAAGAACCAATCA
<i>psbE-petL</i>	3F	CGTGATACATGCATTCTCTGT
	3R	GAAGGAGCTAAATCAAATATGT
<i>ndhF-rpl32</i>	4F	TCGATTACCCGGCTCTTACC
	4R	TGAAATGAGACTGGTTATTTAG
<i>trnT-trnL</i>	5F	GATAAAGATTACTATACCGATC
	5R	ATCCTTTCGTTTGAACCGT

结果与讨论

1 叶绿体基因组基本特征、基因组成及分类

唐古特大黄、药用大黄、掌叶大黄分别得到 18 266 357 400、11 124 793 200、5 717 253 900 bp 的原始数据, 组装拼接后的叶绿体基因组大小为 161 039~161 136 bp (GenBank 登录号见表 3), 为常见四分体结构, 包含一对反向重复区域 IRa 和 IRb (30 929~30 963 bp), 一个大单拷贝区域 LSC (86 426~86 523 bp) 和一个小单拷贝区域 SSC (12 696~12 747 bp) (图 1)。叶绿体基因组总 GC 含量为 37.3%~37.4%, 且 GC 含量在不同区域分布不均衡, IR 区 GC 含量最高 (41.1%), LSC 区 GC 含量次之 (35.3%~35.4%), SSC 区 GC 含量最低 (32.6%)。3 个物种完整叶绿体基因组均编码 131 个基因, 包括 86 个蛋白质编码基因、37 个 tRNA 基因和 8 个 rRNA 基因。3 个物种叶绿体基因组的基因分布完全相同, 均发现 18 个具有内含子的基因, 其中 15 个基因包含一个内含子, 3 个基因 (*rps12*、*ycf3*、*clpP*) 包含两个内含子。在这 18 个含有内含子的基因中, 12 个基因分布在 LSC 区, 4 个基因 (*rpl2*、*ndhB*、*trnI-GAU*、*trnA-UGC*) 分布在 IR 区, 1 个基因 (*ndhA*) 分布在 SSC 区, 还有 *rps12* 基因是一个特殊的反式剪切基因, 5'端

位于LSC区, 3'端位于IR区(表4)。

2 叶绿体基因组密码子偏好性分析

根据RSCU值计算唐古特大黄、掌叶大黄和药用大黄叶绿体基因组蛋白编码基因的密码子使用模式。结果表明3个物种叶绿体基因组包含64种密码子, 其中61种密码子编码20种氨基酸, 其余3个为终止密码子。在所有氨基酸密码子中, 编码亮氨酸(Leu, L)的

密码子数最多, 半胱氨酸(Cys, C)的密码子数最少。RSCU为>1的密码子共30个, 其中以A/U结尾的密码子29个。RSCU≤1的密码子共34个密码子, 其中31个为G/C结尾的密码子。这些结果表明3个物种叶绿体基因组对A/U结尾密码子的偏好性高于G/C结尾密码子。编码蛋氨酸(Met, M)和色氨酸(Trp, W)的密码子均只有一种, RSCU值为1, 无使用偏好性。编

Table 3 The basic information of chloroplast genomes of the three *Rheum* species

Species	<i>Rheum tanguticum</i>	<i>Rh. officinale</i>	<i>Rh. palmatum</i>
Accession	MZ958816	MZ958817	MZ958818
Location	Hongyuan, Sichuan	Mianning, Sichuan	Dangchang, Gansu
Total length/bp	161 039	161 093	161 136
LSC length/bp	86 426	86 523	86 514
SSC length/bp	12 747	12 712	12 696
IR length/bp	30 933	30 929	30 963
CDS length/bp	82 557	82 557	82 557
GC content/%	37.4	37.3	37.4
GC content of LSC/%	35.4	35.3	35.4
GC content of SSC/%	32.6	32.6	32.6
GC content of IR/%	41.1	41.1	41.1
GC content of CDS/%	38.0	38.0	38.0

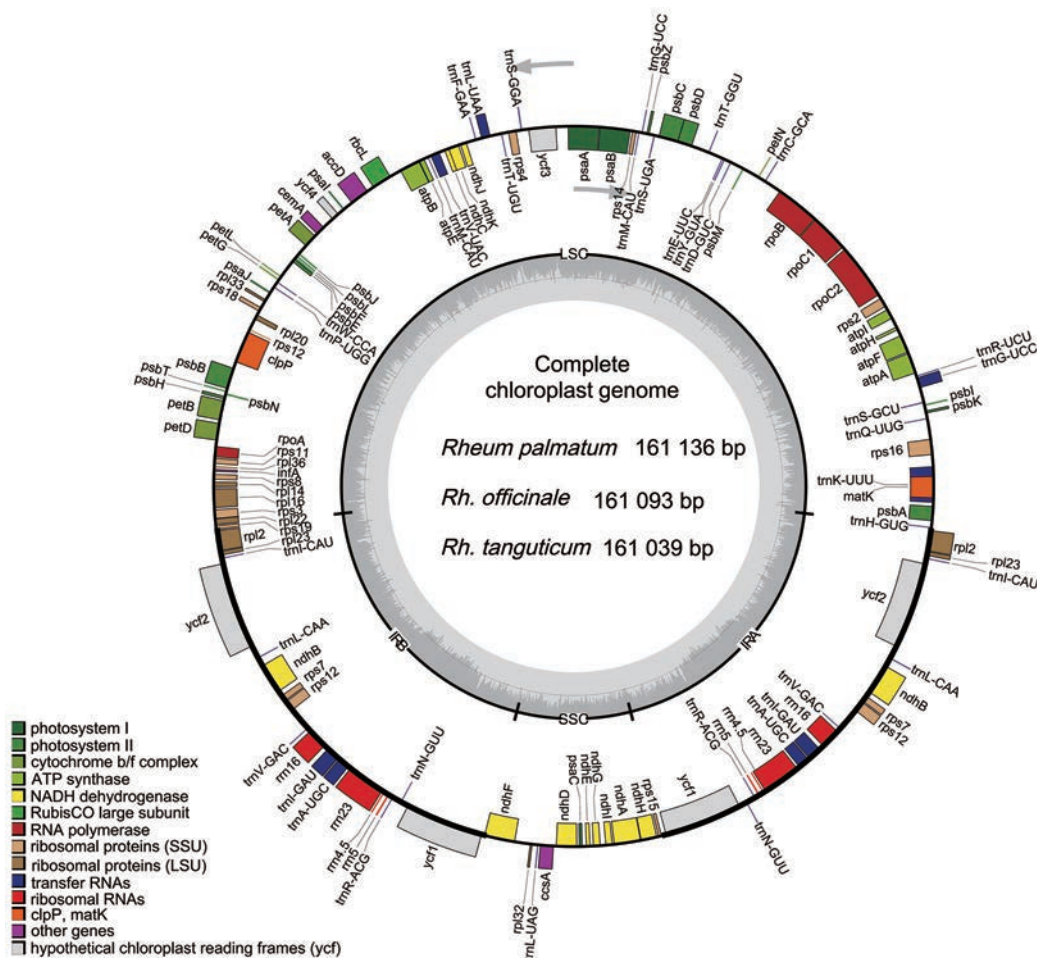


Figure 1 Chloroplast genome map of the three *Rheum* species. The length of the chloroplast genome was shown in the circle. The different colors on the circle represented different types of genes. Represented with arrows, the transcription directions for the inner and outer genes were listed clockwise and anticlockwise, respectively

39 bp, 2个40~49 bp); 掌叶大黄叶绿体基因组共鉴定47个长重复序列, 包括23个正向重复F (18个30~39 bp, 3个40~49 bp, 2个大于70 bp)、22个回文重复P (19个30~39 bp, 3个40~49 bp)、2个反向重复R (30~39 bp)。在这3个物种中, 大多数重复序列长度都在30~39 bp之间, 只有掌叶大黄叶绿体基因组存在2个长度超过70 bp的长重复序列, 均未鉴定到互补重复序列(图3)。

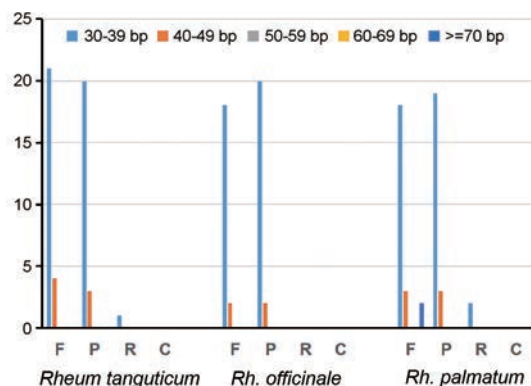


Figure 3 Repeat sequences in the chloroplast genomes of the three *Rheum* species. F: Forward repeats; P: Palindrome repeats; R: Reverse repeats; C: Complement repeats

简单重复序列 (SSR), 是一段短而串联的DNA序列, 大小约为1~6 bp。SSR广泛分布于叶绿体基因组中, 通常被用作物种鉴定的重要分子标记。唐古特大黄、药用大黄及掌叶大黄分别检测到71、84、80个SSR位点, 其中单核苷酸重复序列分别为35、46、44个, 二核苷酸重复序列分别为16、18、14个, 三核苷酸重复序列分别为11、10、11个, 四核苷酸重复序列均为8个, 五核苷酸重复序列分别为1、2、3个, 未检测到六核苷酸重复序列(表5)。重复最多的序列是单碱基重复, 且全部是A/T重复。在所有类型的SSRs中, A和T始终是最常用碱基, 叶绿体基因组SSRs通常由poyA或poyT重复序列组成, 这是叶绿体基因组A/T碱基偏好性的结果。

4 叶绿体基因组比较分析

为检测大黄药材3个基原物种叶绿体基因组的种间变异情况, 以注释过的掌叶大黄叶绿体基因组*Rh. palmatum*_MZ958818为参考, 利用mVISTA在线软件将获得的叶绿体基因组*Rh. tanguticum*_MZ958816、*Rh. officinale*_MZ958817、*Rh. palmatum*_MZ958818进行全局比对(图4)。结果显示, 3条叶绿体基因组序列中非编码区变异高于保守蛋白编码区域, LSC区、SSC区变异明显大于IR区, rRNA基因(*rrn4.5*、*rrn5*、*rrn16*、*rrn23*)高度保守几乎没有变异。基因的保守程度非常

Table 5 Types and amounts of SSRs in the chloroplast genomes of the three *Rheum* species

SSR type	Repeat type	<i>Rheum</i>	<i>Rh.</i>	<i>Rh.</i>
		<i>tanguticum</i>	<i>officinale</i>	<i>palmatum</i>
Mono	A/T	35	46	44
Di	AG/CT	1	1	1
	AT/AT	15	17	13
Tri	AAG/CTT	2	2	2
	AAT/ATT	9	8	9
Tetra	AAAT/ATTT	3	3	3
	AACC/GGTT	1	1	1
	AACT/AGTT	1	1	1
	AATC/ATTG	1	1	1
	AATT/AATT	1	1	1
	ACAG/CTGT	1	1	1
Penta	AAGAT/ATCTT	1	1	1
	AATAT/ATATT	0	1	1
	AAAGG/CCTTT	0	0	1
Total		71	84	80

高, 绝大多数的基因相似性都在90%以上。基因间区变异大于基因区, 如*rps16-trnQ*、*trnT-psbD*、*psaA-ycf3*、*trnT-trnL*、*psbE-petL*、*ndhF-rpl32*、*trnN-ycf1*、*ccsA-ndhD*等。这些高变区可能是唐古特大黄、药用大黄和掌叶大黄物种鉴定的潜在分子标记, 具有较高研究价值。

使用软件DnaSP计算3个物种叶绿体基因组所有基因和基因间区的核苷酸变异值(Pi), 进行高变区检测和序列差异水平分析。结果显示3个物种叶绿体基因组的基因间区Pi平均值(0.004)变异大于基因Pi平均值(0.0008), 突变热点通常位于LSC区和SSC区(图5), 与mVISTA分析结果一致。*psaB-psaA* (25 bp)、*rpoC1-rpoB* (26 bp)等基因间区Pi值较高与自身序列长度较短有关, 因此在筛选可作为潜在分子标记的高变热点区域时, 应保证序列长度大于100 bp, 结合Pi值与mVISTA结果分析, 本研究筛选*rps16-trnQ*、*psaA-ycf3*、*psbE-petL*、*ndhF-rpl32*、*trnT-trnL*为高变区。

5 基于叶绿体基因组高变区的特异DNA条形码鉴定分析

以样品总DNA为模板对筛选的叶绿体基因组高变区进行PCR扩增, 并对测序结果进行分析。实验结果表明基于高变区序列所设计的5对引物均可对42份样品进行有效扩增, 条带单一明亮(图6), 1F/1R扩增区域为*rps16-trnQ*, 扩增长度为1176 bp; 2F/2R扩增区域为*psaA-ycf3*, 扩增长度为751 bp; 3F/3R扩增区域为*psbE-petL*, 扩增长度为706 bp; 4F/4R扩增区域为*ndhF-rpl32*, 扩增长度为600 bp; 5F/5R扩增区域为*trnT-trnL*, 扩增长度为630 bp。根据MEGA 7.0计算测序所获高质量序列的种内及种间K2P距离, 结果显示种间最小遗传距离大于种内最大遗传距离, 不存在

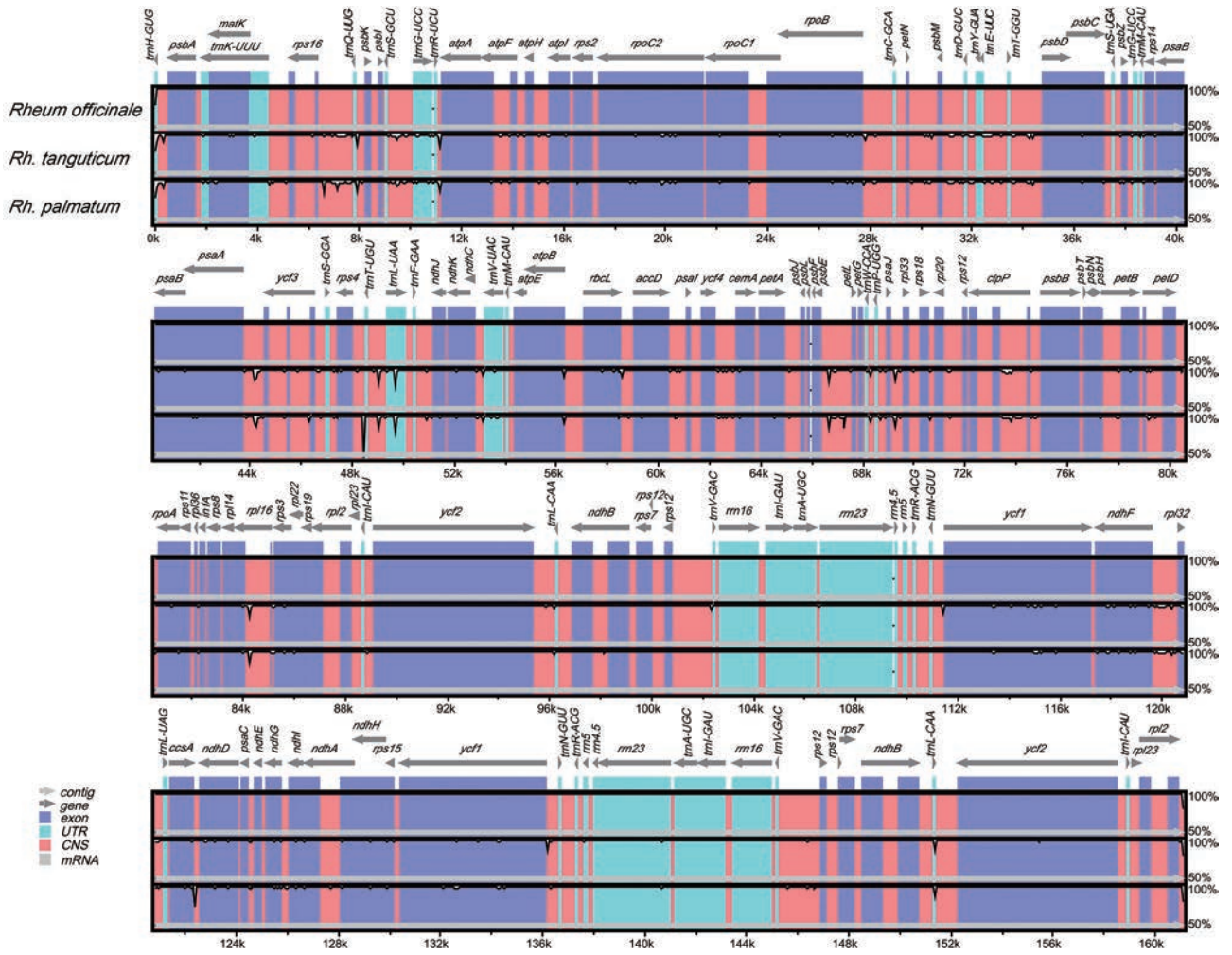


Figure 4 Global alignment in the chloroplast genomes of the three *Rheum* species

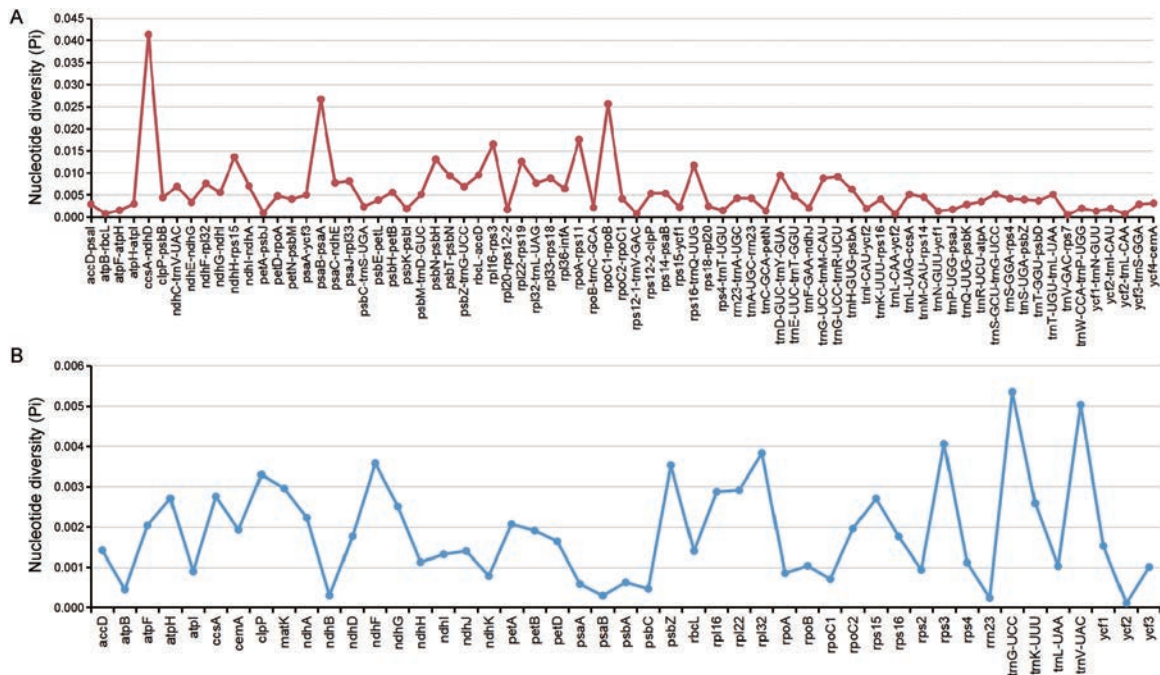


Figure 5 Nucleotide diversity of shared various regions with $P_i > 0$ in the chloroplast genomes of the three *Rheum* species. A: P_i values in the intergenic spacer regions; B: P_i values in the gene regions

重叠,表明所筛选片段符合DNA条形码的要求(表6)。NJ系统进化树分析证实唐古特大黄单独聚为一支,药用大黄单独聚为一支,掌叶大黄单独聚为一支(图7)。因此 *rps16-trnQ*、*psaA-ycf3*、*psbE-petL*、*ndhF-rpl32*、*trnT-trnL* 可以作为特异DNA条形码用于大黄药材3个基原物种的鉴定。*psaA-ycf3*、*ndhF-rpl32*、*trnT-trnL* 的NJ系统发育树显示唐古特大黄 PS2904MT01~PS2904MT06 单独聚为一支,采集地点为四川红原; PS2904MT07~PS2904MT12 单独聚为一支,采集地点为青海林川,表明 *psaA-ycf3*、*ndhF-rpl32* 与 *trnT-trnL* 可作为潜在特异DNA条形码用于同一物种不同产区的鉴定。

讨论

长期以来大黄药材基原物种分类与鉴定主要依据外观性状,但地理和生态因素可能引起大黄药材基原

植物不同物种间形态特征的持续变化^[13],因而开发大黄药材3个物种的特异DNA条形码具有巨大价值。采用Illumina高通量测序和比较基因组学对大黄药材3个基原物种进行叶绿体基因组分析,结果显示叶绿体基因组可作为超级条形码准确鉴定唐古特大黄、药用大黄和掌叶大黄。叶绿体基因组的一些高变异区序列可作为物种鉴定及系统发育关系分析等相关研究的特异DNA条形码,本研究在唐古特大黄、药用大黄和掌叶大黄叶绿体高变基因间区筛选到5条序列,对42份样品进行PCR实验验证,扩增效率高,扩增片段长度为600~1200 bp, Sanger测序后所得高质量序列利用MEGA软件计算K2P距离和构建NJ系统进化树,结果证明所筛选的片段 *rps16-trnQ*、*psaA-ycf3*、*psbE-petL*、*ndhF-rpl32* 与 *trnT-trnL* 能作为特异DNA条形码鉴定大黄药材三个基原物种。

尽管已有近百篇研究基于叶绿体基因组筛选高变

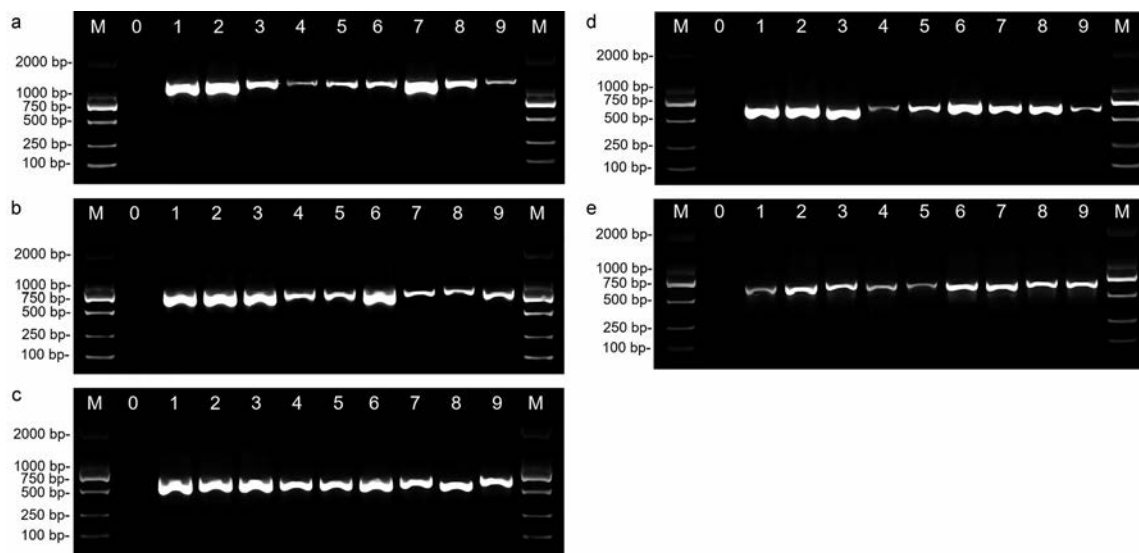


Figure 6 Agarose gel electrophoresis of five specific DNA barcodes of the original species of Rhei Radix et Rhizoma. a: *rps16-trnQ*; b: *psaA-ycf3*; c: *psbE-petL*; d: *ndhF-rpl32*; e: *trnT-trnL*. M: DL2000 DNA marker; 0: Negative control; 1: *Rheum officinale*_PS2902MT01; 2: *Rh. palmatum*_PS2903MT01; 3: *Rh. tanguticum*_PS2904MT01; 4: *Rh. tanguticum*_PS2904MT07; 5: *Rh. rhaponticum*_PS2905MT01; 6: *Rh. rhabarbarum*_PS2906MT01; 7: *Rh. alexandrae*_PS2907MT01; 8: *Rh. nobile*_PS2908MT01; 9: *Rumex crispus*_PS2909MT01

Table 6 Intra/Inter-specific K2P distance of five specific DNA barcodes of the original species of Rhei Radix et Rhizoma

Species	<i>rps16-trnQ</i>	<i>psaA-ycf3</i>	<i>psbE-petL</i>	<i>ndhF-rpl32</i>	<i>trnT-trnL</i>
<i>Rheum tanguticum</i>	0	0-0.001 7 (0.000 9)	0	0-0.001 7 (0.001 0)	0-0.001 7 (0.000 9)
<i>Rh. officinale</i>	0	0	0-0.001 9 (0.001 0)	0	0
<i>Rh. palmatum</i>	0-0.002 9 (0.000 6)	0	0-0.001 7 (0.000 9)	0	0
<i>Rh. officinale</i> and <i>Rh. palmatum</i>	0.016 7-0.018 7 (0.017 0)	0.015 5 (0.015 5)	0.008 3-0.010 4 (0.009 4)	0.015 0 (0.015 0)	0.005 2 (0.005 2)
<i>Rh. officinale</i> and <i>Rh. tanguticum</i>	0.007 8 (0.007 8)	0.015 6-0.017 3 (0.016 4)	0.003 8-0.005 7 (0.004 6)	0.011 2-0.013 1 (0.012 2)	0.005 2-0.006 9 (0.006 0)
<i>Rh. palmatum</i> and <i>Rh. tanguticum</i>	0.016 7-0.018 7 (0.017 0)	0.003 4-0.006 8 (0.005 1)	0.012 5-0.014 6 (0.013 6)	0.010 5-0.012 3 (0.011 4)	0.006 9-0.008 6 (0.007 7)

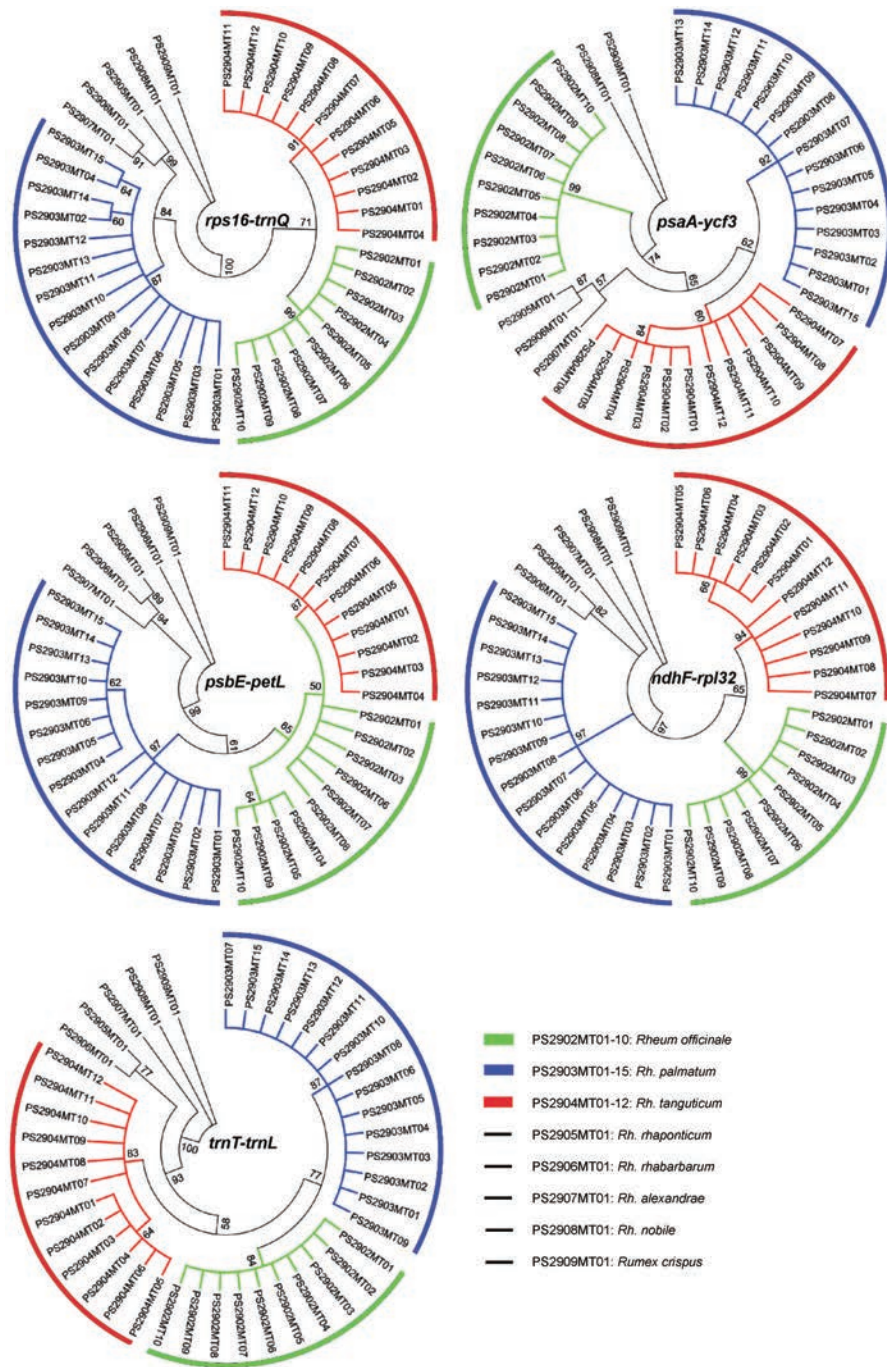


Figure 7 NJ phylogenetic tree based on five specific DNA barcodes of the original species of Rhei Radix et Rhizoma. The bootstrap scores (1 000 replicates) are shown (≥50%) for each branch

区尝试解决一些近缘物种的鉴定问题,但这些研究都均未进行实验验证。如刘昶等^[24]对川贝母、丹参等53种重要药用植物叶绿体基因组图谱及应用进行详细阐述,并利用ecoPrimers软件基于叶绿体基因组设计能够区分贝母属、葱属、芍药属及百合属等物种的DNA条形码序列,可作为潜在分子标记开发区域。武立伟等^[25]分析检测出细茎石斛复合种叶绿体基因组中的、*ndhF-rpl32*等高变异区。杨嘉鹏等^[26]推荐

石豆兰属植物叶绿体基因组高变区 *psbI-trnS*、*psbC-trnS*、*clpP-ex1-psbB*、*psaJ-rpl33*、*rpl33-rps18* 组合用于设计鉴定石豆兰属药用植物的专属DNA条形码。另外,在中国知网仅检索到10余篇文献对于叶绿体基因组所筛选高变区进行实验验证,如董博然等^[27]基于叶绿体基因组筛选 *trnT-trnL* 和 *ycf1* 两个高变区,91份样品实验验证后构建长梗秦艽、全萼秦艽的特异性引物PCR两步鉴定法,方法易于操作且准确性高。崔英

贤^[28]对筛选的高变区 *atpH-atpI* 设计两对特异性引物, 但扩增仍不成功, 表明该区段扩增效率低, 并不适合于豆蔻属分子标记开发; *accD-psaI* 和 *ycf4-cemA* 作为阳春砂、绿壳砂和海南砂的潜在分子标记, 但其实验所获序列构建的 NJ 树各分支支持率普遍较低, 鉴定效率不理想。由此可见, 对所筛选高变区进行实验验证的研究非常少, 结果不尽理想。综合文献研究, 建议在筛选出高变区域后, 应对实际扩增效率进行实验验证, 对所筛选高变区的物种鉴定能力进行确定。本研究对大黄药材 3 个基原物种叶绿体基因组进行测序分析, 开发特异 DNA 条形码并进行 PCR 验证, 从 K2P 遗传距离和 NJ 进化树两个方面评价所筛选高变区实际鉴定效率, 表明所筛选 5 个特异 DNA 条形码可以有效实施大黄基原物种鉴定, 为大黄基原物种分类鉴定奠定基础。

研究表明唐古特大黄、药用大黄及掌叶大黄在地理分布上出现重叠与替换^[29], ISSR 分析发现 3 个物种展示出丰富的遗传多样性^[30,31]。采用核基因组序列 ITS 的研究结果表明唐古特大黄存在种内遗传多样性^[32], 同一物种在不同地区间遗传物质有所差异。针对高变区设计的 5 对引物扩增片段长度为 600~1 200 bp, 旨在筛选包含丰富变异位点的特异 DNA 条形码, NJ 进化树结果证明四川与青海两处的唐古特大黄各自聚为一支, 表明所筛选特异 DNA 条形码可以用于大黄基原植物同一物种不同产区的分类鉴定。本研究实验采集对象为大黄药材基原植物, 基因组 DNA 较为完整, 对所筛选片段可进行有效扩增。DNA 条形码在中药全产业链各环节发挥重要作用^[33], 因此可采集不同产区的大黄药材基原植物样本, 基于本文筛选的 5 个特异 DNA 条形码, 进一步丰富大黄药材三个基原物种的遗传变异信息, 用于市售饮片与相关中成药的分子鉴定, 为保证大黄药材临床用药安全、规范大黄药材市场提供新依据。

致谢: 感谢中国中医科学院中药研究所孙伟研究员提供苞叶大黄和塔黄样品。

作者贡献: 李冉郡负责实验工作、数据分析、文章撰写与修改; 武立伟负责数据分析; 辛天怡负责样本收集及稿件修改; 廖海参与样本收集及稿件修改; 林余霖负责基原植物性状鉴定; 姚辉参与数据分析; 周嘉裕和宋经元负责实验设计、总体把关和稿件修改等工作。

利益冲突: 本文的研究内容无任何利益冲突。

References

- [1] State Pharmacopoeia Committee. Pharmacopoeia of the People's Republic of China (中华人民共和国药典) [S]. 2020 Ed. Part I. Beijing: China Medical Science Press, 2020: 24.
- [2] Chen JQ, Chen YY, Tang YP, et al. Multifunctional regularity of Rhei Radix et Rhizoma in ancient and modern medicine [J]. Chin Tradit Herb Drugs (中草药), 2019, 50: 1485-1492.
- [3] Xu LH, Huang XJ, Li YR, et al. Effect and potential mechanism of rhubarb on COVID-19 based on the "wenbingzaixiaqiye" theory [J]. Pharmacol Clin Chin Mater Med (中药药理与临床), 2020, 36: 85-90.
- [4] Chen YY, Tang YP, Chen JQ, et al. Research progress and utilization strategy on resource chemistry of Rhei Radix et Rhizoma [J]. Chin Tradit Herb Drugs (中草药), 2018, 49: 5170-5178.
- [5] Wang Y, Yang X, Xia P, et al. Research progress on chemical composition and pharmacological effects of Rhei Radix et Rhizoma and predictive analysis on quality markers [J]. Chin Tradit Herb Drugs (中草药), 2019, 50: 4821-4837.
- [6] Komatsu K, Nagayama Y, Tanaka K, et al. Comparative study of chemical constituents of rhubarb from different origins [J]. Chem Pharm Bull (Tokyo), 2006, 54: 1491-1499.
- [7] Wang XM, Hou XQ, Zhang YQ, et al. Morphological variation in leaf dissection of *Rheum palmatum* complex (Polygonaceae) [J]. PLoS One, 2014, 9: e110760.
- [8] Lin YL, Chen SL. Illustrated Handbook of Chinese Medicinal Plants (中国药用植物原色图鉴) [M]. Fuzhou: Fujian Science and Technology Press, 2016.
- [9] Ge JH, Liu XH, Xu H, et al. Identification of different varieties of Rhei Radix et Rhizoma based on chemical analysis [J]. China J Chin Mater Med (中国中药杂志), 2015, 40: 2309-2313.
- [10] Du QT, Wen JL, Yan YS, et al. UPLC fingerprint of Rhei Radix et Rhizoma from different habitats [J]. J Chin Med Mater (中药材), 2013, 36: 725-731.
- [11] Li MN, Han RL, Han JP, et al. Identification of original plants of Radix et Rhizoma Rhei from its adulterants Rhizoma et Radix Polygoni Cuspidati and Radix Rumicis Obtusifolii by ITS2 sequences [J]. Global Tradit Chin Med (环球中医药), 2012, 5: 185-189.
- [12] Song JY, Yao H, Li Y, et al. Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique [J]. J Ethnopharmacol, 2009, 124: 434-439.
- [13] Zhang XQ, Liu CS, Yan XL, et al. Sequence analysis and identification of a chloroplast *matK* gene in Rhei Rhizoma from different botanical origins [J]. Acta Pharm Sin (药学报), 2013, 48: 1722-1728.
- [14] Li RJ, Xin TY, Song LK, et al. Research progress in original species identification in industry chain of Rhei Radix et Rhizoma [J]. China J Chin Mater Med (中国中药杂志), 2021, 46: 1060-1066.
- [15] Kane NC, Cronk Q. Botany without borders: barcoding in focus [J]. Mol Ecol, 2008, 17: 5175-5176.
- [16] Jansen RK, Raubeson LA, Boore JL, et al. Methods for obtaining and analyzing whole chloroplast genome sequences [J]. Methods Enzymol, 2005, 395: 348-384.

- [17] Zhu AD, Guo WH, Gupta S, et al. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates [J]. *New Phytol*, 2016, 209: 1747-1756.
- [18] Green BR. Chloroplast genomes of photosynthetic eukaryotes [J]. *Plant J*, 2011, 66: 34-44.
- [19] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data [J]. *Bioinformatics*, 2014, 30: 2114-2120.
- [20] Dierckxsens N, Mardulyn P, Smits G. Unraveling heteroplasmy patterns with NOVOPlasty [J]. *NAR Genom Bioinform*, 2019, 2: lqz011.
- [21] Liu C, Shi LC, Zhu YJ, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences [J]. *BMC Genomics*, 2012, 13: 715.
- [22] Tillich M, Lehwark P, Pellizzer T, et al. GeSeq-versatile and accurate annotation of organelle genomes [J]. *Nucleic Acids Res*, 2017, 45: W6-11.
- [23] Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs [J]. *Nucleic Acids Res*, 2005, 33: W686-689.
- [24] Liu C, Huang LF. Chloroplast Genome Illustrated Handbook of Chinese Medicinal Plants (中国药用植物叶绿体基因组图谱第一册) [M]. Beijing: Science Press, 2020.
- [25] Wu LW, Cui YX, Nie LP, et al. The characteristics of complete chloroplast genome sequence and phylogenetic analysis of *Dendrobium moniliforme* [J]. *Acta Pharm Sin (药学报)*, 2020, 55: 1056-1066.
- [26] Yang JP, Zhu ZL, Fan YJ, et al. Comparative plastomic analysis of three *Bulbophyllum* medicinal plants and its significance in species identification [J]. *Acta Pharm Sin (药学报)*, 2020, 55: 2736-2745.
- [27] Dong BR, Zhao ZL, Ni LH, et al. Molecular markers based upon whole chloroplast genomes and identifying alpine *Gentiana waltonii* and *G. lhasica* (Gentianaceae) [J]. *Acta Pharm Sin (药学报)*, 2021, 56: 2584-2591.
- [28] Cui YX. Structural Analysis of Complete Chloroplast Genome of Medicinal and Edible Original Plants of *Amomum*, Chinese Wolfberry, Hawthorn and Ginger (药食两用药材砂仁、枸杞、山楂和姜基原植物叶绿体基因组结构解析) [D]. Beijing: Peking Union Medical College, 2020.
- [29] Wang XM, Hou XQ, Zhang YQ, et al. Distribution pattern of genuine species of rhubarb as traditional Chinese medicine [J]. *J Med Plants Res*, 2010, 4: 1865-1876.
- [30] Wang XM, Yang R, Feng SF, et al. Genetic variation in *Rheum palmatum* and *Rheum tanguticum* (Polygonaceae), two medicinally and endemic species in China using ISSR markers [J]. *PLoS One*, 2012, 7: e51667.
- [31] Wang XM, Hou XQ, Zhang YQ, et al. Genetic diversity of the endemic and medicinally important plant *Rheum officinale* as revealed by Inter-Simple Sequence Repeat (ISSR) markers [J]. *Int J Mol Sci*, 2012, 13: 3900-3915.
- [32] Ma XC, Xie CX, Guan M, et al. High levels of genetic diversity within one population of *Rheum tanguticum* on the Qinghai-Tibet Plateau have implications for germplasm conservation [J]. *Pharm Crops*, 2014, 5: 1-8.
- [33] Lou Q, Xin TY, Song JY. Application of DNA barcoding technology in the whole industrial chain of traditional Chinese medicine [J]. *Acta Pharm Sin (药学报)*, 2020, 55: 1784-1791.