

# 扁果草叶绿体基因组特征分析

热伊汉古丽·图尔迪, 慕丽红, 田新民

新疆大学 生命科学与技术学院 新疆生物资源基因工程重点实验室, 新疆 乌鲁木齐 830046

热伊汉古丽·图尔迪, 慕丽红, 田新民. 扁果草叶绿体基因组特征分析. 生物工程学报, 2022, 38(8): 2999-3013.

TURDI Rayhangul, MU LH, TIAN XM. Characteristics of the chloroplast genome of *Isopyrum anemonoides*. Chin J Biotech, 2022, 38(8): 2999-3013.

**摘要:** 为探讨扁果草 (*Isopyrum anemonoides*) 叶绿体基因组特征及该属物种的系统发育关系, 采用 Illumina Hiseq 高通量测序技术, 对扁果草进行全基因组测序, 组装、注释和完成扁果草叶绿体全基因组图谱绘制。结果显示, 扁果草叶绿体全基因组总长 161 034 bp, 为典型的四分体结构, 包含 85 个蛋白编码基因、37 个转运 RNA 和 8 个核糖体 RNA。共检测到 44 个散在重复序列和 47 个简单重复序列。此外, 扁果草叶绿体基因组中共包含 53 678 个密码子, 其中编码亮氨酸的密码子最多 (5 251 个), 编码色氨酸的密码子最少 (712 个)。共线性分析结果显示, 扁果草与近缘种叶绿体基因组中不存在倒位或重排现象。系统发育分析表明, 扁果草并未与该属的东北扁果草 (*I. manshuricum*) 聚在一个分支上, 而是与假蓎斗菜 (*Paraquilegia microphylla*) 有很近的亲缘关系。本研究为开展后续的扁果草属物种鉴定、系统发育研究提供基础资料。

**关键词:** 扁果草; 高通量测序; 简单重复序列位点; 叶绿体基因组; 系统发育; 密码子偏好性

**Received:** November 25, 2021; **Accepted:** March 30, 2022

**Supported by:** The Open Project of Xinjiang Key Laboratory of Biological Resources and Genetic Engineering (2020D04033); National Natural Science Foundation of China (31760102)

**Corresponding author:** TIAN Xinmin. E-mail: tianxm06@lzu.edu.cn

**基金项目:** 新疆维吾尔自治区重点实验室开放课题 (2020D04033); 国家自然科学基金 (31760102)

# Characteristics of the chloroplast genome of *Isopyrum anemonoides*

TURDI Rayhangul, MU Lihong, TIAN Xinmin

Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Sciences and Technology, Xinjiang University, Urumqi 830046, Xinjiang, China

**Abstract:** In order to characterize the chloroplast genome and phylogenetic relationships of *Isopyrum anemonoides*, we performed Illumina Hiseq high-throughput sequencing to sequence the complete chloroplast genome of this plant and constructed a whole-genome map based on contig assembly and annotation. The chloroplast genome of *I. anemonoides* is 161 034 bp in length and has a typical tetrad structure, comprising 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. The genome also contains a total of 44 dispersed repeat sequences and 47 simple sequence repeats. Among the genome's 53 678 codons, the largest proportion are leucine-encoding codons (5 251), whereas the smallest proportion encode tryptophan (712). Colinear analysis revealed an absence of inversions and rearrangements between *I. anemonoides* and related species at the chloroplast genome level. Whereas phylogenetic analysis indicated that *I. anemonoides* did not cluster in a clade with *I. manshuricum*, it did show a very close phylogenetic relationship with *Paraquilegia microphylla*. The findings of this study provide basic data that will contribute to further species identification and phylogenetic study of the genus *Isopyrum*.

**Keywords:** *Isopyrum anemonoides*; high-throughput sequencing; simple sequence repeats loci; chloroplast genome; phylogeny; codon preference

叶绿体是半自主性细胞器，具有独立的遗传物质，常见于陆生植物、藻类和少数原生动物。叶绿体基因组与线粒体基因组、核基因组相比较而言，在结构、基因数量和基因组组成上更加保守，进化速度相对适中<sup>[1]</sup>。叶绿体基因组有多种构型，最常见的结构是双链环状结构，包括短单拷贝区 (small single-copy region, SSC) 和长单拷贝区 (large single-copy region, LSC)，这两个区域被一对反向重复区域 (inverted repeat, IR) 分开，形成典型的四分体结构<sup>[2]</sup>。高等植物叶绿体基因组大约有 110–130 个编码基因，包含与光合作用、自我复制、开放阅读框 (open reading frame, ORF) 和一些其他蛋白编

码相关的基因<sup>[3]</sup>。有些植物叶绿体基因中包含两个内含子 (如 *clpP*、*ycf3*) 和外显子，内含子大部分为 II 型内含子，有时在基因中会存在一个反式剪接结构的内含子。近年来，基于叶绿体全基因组的物种鉴定及系统进化研究成为植物系统分类学的一个新趋势。

扁果草属隶属于毛茛科 (Ranunculaceae)，共包括 4 个种 (*Isopyrum*)，分布于亚洲和欧洲<sup>[4]</sup>。我国分布的扁果草属植物有两种，一种为扁果草 (*I. anemonoides* Kar. et Kir.)，分布于我国甘肃南部、青海、新疆、西藏和陕西等地；另一种为东北扁果草 (*I. manshuricum* Kom.)，分布于我国东北地区<sup>[5]</sup>。扁果草属植物是重要的

中药材资源,在民间常用于散结消肿、解毒等方面,其主要药用成分苜基异喹啉生物碱具有抗炎症和免疫调节等作用<sup>[6]</sup>。近年来,随着分子生物学的发展,越来越多的证据表明,基于分子系统学和形态学的物种鉴定在很多植物类群中有较大差异。目前,国内外对扁果草属植物的研究主要集中在形态学、化学成分和药理活性等方面<sup>[7]</sup>,且已建立了系统发育关系框架,但对于扁果草的叶绿体基因组研究仍处于一片空白。扁果草与东北扁果草在形态上较相似,其主要差异集中在根和花瓣,东北扁果草根状茎上有多数纺锤状小块根和密集的须根,花瓣下缘合生成浅杯状;而扁果草根状茎细长,横走,疏生少数须根,花瓣呈圆筒状。那么这两个物种是否为一个姐妹群?

基于此科学问题,本研究通过二代高通量测序技术,获得扁果草全基因组数据,对叶绿体基因组进行组装、注释并绘制其基因组图谱,分析其叶绿体基因组结构特征、散在重复序列、简单重复序列 (simple sequence repeat, SSR) 以及密码子偏好性。同时,比较分析扁果草与近缘种叶绿体基因组相似性和差异性,并建立系统发育树,揭示扁果草属物种间的亲缘和进化关系。该研究充实了扁果草属植物叶绿体基因组数据,为扁果草属植物物种鉴定与系统进化研究提供了基础资料。

## 1 材料与方法

### 1.1 实验材料采集与保存

本研究材料扁果草采集于新疆维吾尔自治区吉木萨尔县渭沪沟林场,标本保存于新疆大学植物标本馆,叶片材料经硅胶干燥后保存于实验室。

### 1.2 二代测序和叶绿体基因组组装、注释

使用Illumina HiSeq平台进行基因组测序,

最终获得原始测序数据。使用乳突拟耧斗菜 (*Paraquilegia anemonoides*) 叶绿体基因组序列作为参考,采用NOVOPlasty<sup>[8]</sup>程序进行组装,获得扁果草叶绿体全基因组,然后通过GeSeq对叶绿体基因组进行初步注释,用Sequin软件进行手动注释。使用在线软件OGDRAW绘制 (<https://chlorobox.mpimp-golm.mpg.de/GenBank2Sequin.html>) 扁果草叶绿体全基因组图谱<sup>[9]</sup>。

### 1.3 叶绿体基因组基本特征与比较分析

通过Geneious软件对叶绿体基因组的总长度及各个区域 (LSC、SSC、IR) 的长度、基因组成 (蛋白编码基因、tRNA基因、rRNA基因、内含子、外显子)、碱基组成、GC (AT) 含量进行统计和比较分析,解析扁果草叶绿体基因组序列的特征。

利用 Reputer<sup>[10]</sup>软件预测扁果草叶绿体基因组散在重复序列。即正向重复 (forward repeats, F)、反向重复 (reverse repeats, R)、互补重复 (complement repeats, C)、回文重复 (palindromic repeats, P)。最小重复长度 (minimal repeat size) 设置为30, hamming 距离 (hamming distance) 为3。

简单重复序列 (SSR) 由长度1–6 bp重复单元组成,分布于基因组各个区域<sup>[11]</sup>。SSR 具有高度多态性,因此常作为群体遗传学和进化研究中的分子标记<sup>[12–14]</sup>。使用MISA (<https://webblast.ipk-gatersleben.de/misa/index.php>) 在线软件预测扁果草物种叶绿体基因组序列中的简单重复序列的类型和数量<sup>[15]</sup>。通过MEGA 7.0软件对扁果草叶绿体基因组密码子的使用偏好性进行统计。

利用Mauve软件对扁果草与近缘物种叶绿体基因组进行共线性分析,参数采用默认值。利用Circoletto程序,以E值1e-10,分别对扁果草与东北扁果草两个物种叶绿体基因组进行

比较,两个物种序列相似度按照相似区域与相似长度使用不同的颜色进行连接,图中连接线之间的空白区域表明两个物种没有相似性的区域。

#### 1.4 系统发育分析

从扁果草全基因组序列中提取出了nrDNA ITS序列和叶绿体基因片段 (*rbcL*、*matK*、*trnL-F*),并从NCBI数据库 ([https:// www.ncbi.nlm.nih.gov/](https://www.ncbi.nlm.nih.gov/)) 下载14个扁果草属近缘种序列(表1)。基于nrDNA ITS序列和3个叶绿体基因片段 (*rbcL*、*matK*、*trnL-F*),通过MAFFT对15种植物的叶绿体基因组序列完成序列比对,再使用MEGA 7.0进行校正,使用 MrBayes v.3.2.1软件<sup>[16]</sup>,分别和联合构建系统发育树,从而明

确该物种的系统位置和最近缘物种。

## 2 结果与分析

### 2.1 扁果草叶绿体基因组的基本特征

扁果草的叶绿体基因组与典型的陆地植物叶绿体基因组基本特征一样,均为典型的四分体结构,叶绿体基因组全长161 034 bp,由一对反向重复区 (IRA和IRB, 29 385 bp),一个长的单拷贝区 (LSC, 84 877 bp) 和一个短的单拷贝区 (SSC, 17 387 bp) 组成(图1)。扁果草叶绿体基因组的GC含量接近于其他植物叶绿体GC含量为38.8%;但其3个区域 (LSC、SSC、IR) 的GC含量具有明显的差异,其中IR区的GC含量最高为43.8%,其次是LSC区为36.6%,而SSC

表1 扁果草近缘种 NCBI 序列号

Table 1 NCBI sequence numbers of related species of *I. anemonoides*

Species	ITS	<i>matK</i>	<i>trnL-F</i>	<i>rbcL</i>	GenBank No.
<i>Aquilegia ecalcarata</i>	U75657.1	EF437127.1	EF437096.1	AY954495.1	NC_041529.1
<i>Aquilegia oxysepala</i>	EF437114.1	EF437128.1	EF437097.1	EF437140.1	
<i>Semiaquilegia adoxoides</i>	EF437123.1	EF437137.1	EF437106.1	EF437147.1	MK569498.1
<i>Isopyrum manshuricum</i>	EF437119.1	NC_041541.1	EF437102.1	EF437143.1	NC_041541.1
<i>Isopyrum thalictroides</i>	EF437120.1	EF437134.1	EF437103.1	EF437144.1	
<i>Enemion raddeanum</i>	EF437117.1	NC_041535.1	EF437100.1	AY954494.1	NC_041535.1
<i>Isopyrum anemonoides</i>	EF437118.1	EF437132.1	EF437101.1	EF437142.1	
<i>Paraquilegia microphylla</i>	EF437122.1	EF437136.1	EF437105.1	EF437146.1	
<i>Leptopyrum fumarioides</i>	EF437121.1	EF437135.1	EF437104.1	EF437145.1	NC_041542.1
<i>Thalictrum javanicum</i>	EF437124.1	DQ478615.1	EF437107.1	AY954496.1	
<i>Thalictrum robustum</i>	EF437125.1	EF437138.1	EF437108.1	EF437148.1	
<i>Dichocarpum dalzielii</i>	EF437115.1	KY235705.1	EF437098.1	EF437141.1	MK253459.1
<i>Dichocarpum sutchuenense</i>	EF437116.1	KY235714.1	KY235760.1	AY954493.1	
<i>Hydrastis canadensis</i>	MF349151.1	NC_034702.1	EF437112.1	AF093725.1	
<i>Thalictrum minus</i>					NC_041544.1
<i>Urophysa rockii</i>					NC_039742.1
<i>Paraquilegia anemonoides</i>					NC_041479.1
<i>Isopyrum anemonoides</i>	OM491213				OM457045

区的GC含量最低, 仅为32.4%。整体而言, IR区的保守性高于LSC区和SSC区, 这可能是IR

区的GC含量高于LSC区和SSC区而造成的。各分区的碱基组成见表2。

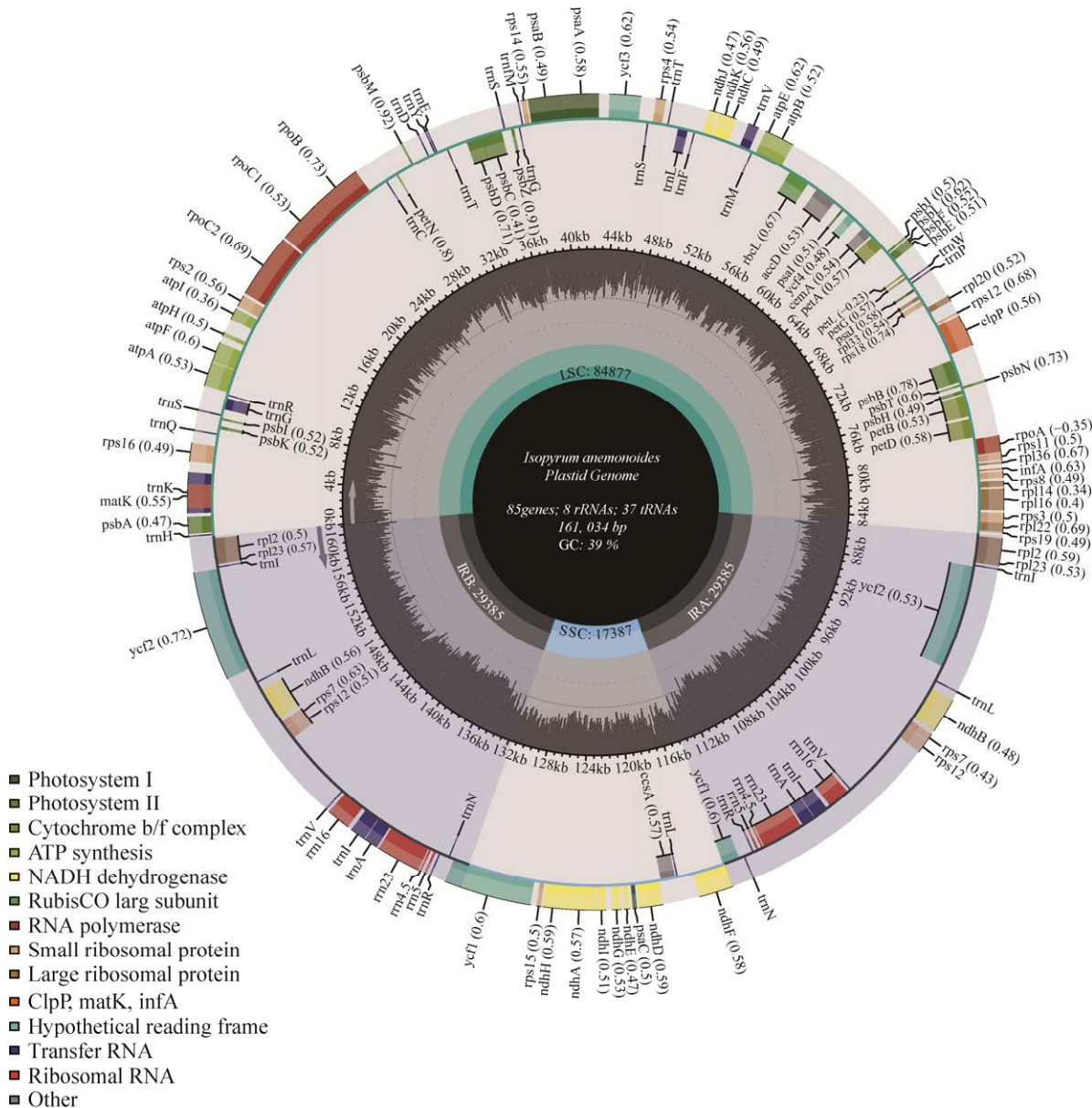


图 1 扁果草叶绿体基因组图谱 物种名称和有关基因组的特定信息(长度、GC 含量和基因数量)显示在图的中心. 不同功能的基因用不同的颜色表示. 内圈深色部分表示 GC 含量, 浅色部分表示 AT 含量

Figure 1 Chloroplast genome map of *I. anemonoides*. The species name and specific information regarding the genome (length, GC content, and the number of genes) are depicted in the center of the plot. Genes with different functions are represented by different colors. The dark part of the inner circle indicates GC content, and the light part indicates AT content.

表2 扁果草叶绿体基因组碱基组成

Table 2 The base composition of the chloroplast genome of *I. anemonoides*

Region	A%	T%	C%	G%	GC%	Length (bp)
Chloroplast genome	30.3	30.9	19.7	19.0	38.8	161 034
Small single-copy region, SSC	34.2	33.4	17.1	15.3	32.4	17 387
Large single-copy region, LSC	31.0	32.4	18.8	17.8	36.6	84 877
Inverted repeated region A, IRA	28.0	28.3	22.4	21.3	43.8	29 385
Inverted repeated region B, IRB	28.3	28.0	21.3	22.4	43.8	29 385

## 2.2 扁果草叶绿体基因组的基因组成

扁果草叶绿体基因组共包含130个基因, 包括8个rRNA基因、37个tRNA基因和85个蛋白编码基因。其中85个蛋白编码基因可分为以下4个大类。第1类: 45个与光合作用相关的基因, 包括6个ATP酶基因、12个NADH氧化还原酶基因、6个细胞色素b/f复合体基因、5个光合系统I基因、15个光合系统II基因、1个二磷酸核酮糖羧化酶大亚基基因。第2类: 除rRNA基因和tRNA基因外, 有28个与自我复制相关的基因, 包括4个RNA聚合酶亚基基因、10个核糖体蛋白大亚基基因和14个核糖体蛋白小亚基基因。第3类: 6个其他编码蛋白质的基因。第4类: 6个功能未知的基因(表3)。内含子在基因表达调控中起着重要作用, 在扁果草叶绿体基因组中, 共有12个蛋白编码基因和6个tRNA基因含有内含子, 除*clpP*、*ycf3*、*rps12*基因含2个内含子外, 其余15个基因均含1个内含子, 但大小存在一定差异。其中, *trnK-UUU* 拥有最大的内含子(2 534 bp), *matK* 蛋白编码基因位于其中, 这与其他绿色植物的一般特性相似(表4)。

## 2.3 密码子偏好性分析

使用MEGA 7.0软件分析扁果草叶绿体基因组蛋白编码序列密码子的使用偏好性(图2), 共预测了53 678个密码子, 其中编码亮氨酸(Leu)的密码子数量最多(5 251个, 9.78%), 编码色氨酸(Trp)的密码子数量最少(712个, 1.33%)。在扁果草叶绿体基因组中有20种氨基

酸, 除色氨酸(Trp)和甲硫氨酸(Met)使用一个密码子TGG和ATG外, 其余氨基酸均有2-6个同义密码子。亮氨酸(Leu)、精氨酸(Arg)、丝氨酸(Ser)使用6个同义密码子, 使用频率最高的密码子分别为TTG、AGA、TCT; 丙氨酸(Ala)、甘氨酸(Gly)、脯氨酸(Pro)、苏氨酸(Thr)、缬氨酸(Val)分别使用4个同义密码子, 使用频率最高的密码子分别为GCT、GGA、CCA、ACA、GTT; 异亮氨酸(Ile)使用3个同义密码子, 使用频率最高的密码子是ATT; 半胱氨酸(Cys)、天冬氨酸(Asp)、谷氨酸(Glu)、苯丙氨酸(Phe)、组氨酸(His)、赖氨酸(Lys)、天冬酰胺(Asn)、谷氨酰胺(Gln)、酪氨酸(Tyr)均使用2个同义密码子, 但使用频率并不相同, 这9个氨基酸使用频率较高的密码子分别为TGT、GAT、GAA、TTT、CAT、AAA、AAT、CAA、TAT。结果表明, 扁果草叶绿体基因偏好使用含有A、T碱基的密码子, 密码子第3位也偏好以A和T结尾。

## 2.4 叶绿体基因组重复序列分析

扁果草叶绿体基因组中预测出44个散在重复序列(表5)。其中, 正向重复17个(38.6%)、回文重复26个(59.1%)、反向重复1个(2.3%), 未发现互补重复(C)。在基因间隔区(*petA-psbJ*、*psbT-psbN*、*ndhG-ndhI*等)、基因编码区(*psaA*、*psaB*、*ycf2*等)、转运RNA(*trnQ-UUG*、*trnS-GGA*等)和*ycf3*基因内含子区域重复序列均有分布, 尤其在基因*ycf2*中出现了多个嵌套的

序列重复。

## 2.5 扁果草叶绿体基因组 SSR 分析

扁果草叶绿体基因组中共检测到47个SSR(表6),其中39个单核苷酸重复,3个二核苷酸重复,5个复杂核苷酸重复,但未发现三核苷酸重复、四核苷酸重复、五核苷酸重复和六核苷酸重复。大多数SSR位于LSC区(35个),仅有

少数分布在SSC区(6个)和IR区(6个)。此外,大部分的SSR位于基因间区31个(65.96%),其次是位于基因编码区8个(17.02%)和基因内含子区8个(17.02%),这证明了SSR主要分布在基因间隔区。这些SSR主要为A/T组成的单碱基重复,且该基因组的SSR具有极强的A、T偏好。

表3 扁果草叶绿体基因组中的基因列表

Table 3 Genes in the chloroplast genome of *I. anemonoides*

Gene function classification	Group of genes	Gene name
Photosynthesis	Photosystem I gene	<i>psaA, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Photosystem II gene	<i>psaA, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Cytochrome b/f complex gene	<i>petA, petB*, petD*, petG, petL, petN</i>
	ATP synthase gene	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>
	Rubis CO large subunit gene	<i>rbcl</i>
	NADH dehydrogenase gene	<i>ndhA*, ndhB*(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
Self-replication	RNA polymerase gene	<i>rpoA, rpoB, rpoC1*, rpoC2</i>
	Ribosomal proteins (LSU) gene	<i>rpl2*(2), rpl14, rpl16*, rpl20, rpl22, rpl23(2), rpl33, rpl36</i>
	Ribosomal proteins (SSU)	<i>rps2, rps3, rps4, rps7(2), rps8, rps11, rps12*(2), rps14, rps15, rps16*, rps18, rps19</i>
	Ribosomal RNAs gene	<i>rrn4.5(2), rrn5(2), rrn16(2), rrn23(2)</i>
	Transfer RNAs gene	<i>trnA-UGC(2)*, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-CAU, trnG-GCC, trnG-UCC*, trnI-CAU(2), trnH-GUG, trnI-GAU*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(2), trnV-UAC*, trnW-CCA, trnY-GUA</i>
Other genes	Translational initiation factor gene	<i>infA</i>
	Maturase K gene	<i>matK</i>
	Subunit of acetyl-Co A gene	<i>accD</i>
	c-type cytochrome synthesis gene	<i>cemA</i>
	Envelop membrane protein gene	<i>ccsA</i>
	Protease gene	<i>clpP</i>
Unkown function	Hypothetical chloroplast reading frames (ycf)	<i>ycf1(2), ycf2(2), ycf3*, ycf4</i>

The number of genes in parentheses; \*: the intron-containing genes.

表 4 扁果草叶绿体基因组内含子的长度和位置

Table 4 Lengths and locations of introns in chloroplast genome of *I. anemonoides*

Genes	Location	ExonI (bp)	IntronI (bp)	ExonII (bp)	IntronII (bp)	ExonIII (bp)
<i>trnK-UUU</i>	LSC	37	2 534	35		
<i>rps16</i>	LSC	40	874	221		
<i>trnG-UCC</i>	LSC	23	696	48		
<i>atpF</i>	LSC	145	713	410		
<i>rpoC1</i>	LSC	432	755	1 619		
<i>ycf3</i>	LSC	124	712	230	765	153
<i>trnL-UAA</i>	LSC	35	560	50		
<i>trnV-UAC</i>	LSC	36	593	37		
<i>rps12<sup>b*</sup></i>	LSC	114	535	232		26
<i>clpP</i>	LSC	71	783	292	659	251
<i>petB</i>	LSC	6	809	642		
<i>petD</i>	LSC	8	724	475		
<i>rpl16</i>	LSC	9	940	399		
<i>rpl2<sup>a</sup></i>	IR	391	658	434		
<i>ndhB<sup>a</sup></i>	IR	777	703	756		
<i>trnI-GAU<sup>a</sup></i>	IR	37	923	35		
<i>trnA-UGC<sup>a</sup></i>	IR	38	800	35		
<i>ndhA</i>	SSC	553	1 060	539		

The *rps12* is a trans-spliced gene with the 5' end located in the LSC region and the duplicated 3' end in the IR regions. a: 2 copies, b: 3 copies.

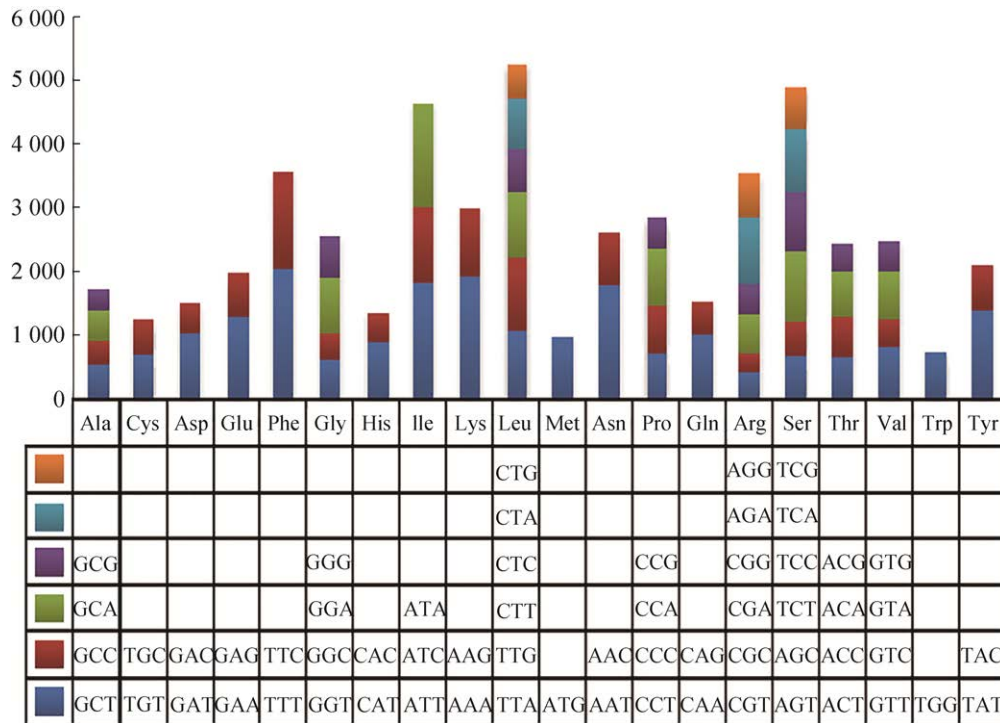


图 2 扁果草叶绿体基因组的密码子使用频率

*x* 轴代表密码子家族, 密码子使用频率绘制在 *y* 轴上

Figure 2 Codon usage frequency in the chloroplast genome of *I. anemonoides*. The *x*-axis represents codon families, frequency of codon usage is plotted on the *y*-axis.



表 5 扁果草叶绿体基因组的重复序列

Table 5 Repetitive sequences of chloroplast genome of *I. anemonoides*

No.	Size	Position 1	Type	Position 2	Location 1	Location 2
1	84	64 434	P	64 434	<i>petA-psbJ</i>	<i>petA-psbJ</i>
2	76	92 176	F	92 194	<i>ycf2</i>	<i>ycf2</i>
3	76	92 176	P	153 617	<i>ycf2</i>	<i>ycf2</i>
4	76	92 194	P	153 635	<i>ycf2</i>	<i>ycf2</i>
5	76	153 617	F	153 635	<i>ycf2</i>	<i>ycf2</i>
6	67	92 176	F	92 212	<i>ycf2</i>	<i>ycf2</i>
7	67	92 176	P	153 608	<i>ycf2</i>	<i>ycf2</i>
8	67	92 212	P	153 644	<i>ycf2</i>	<i>ycf2</i>
9	67	153 608	F	153 644	<i>ycf2</i>	<i>ycf2</i>
10	63	14 068	P	116 830	<i>atpH-atpI</i>	<i>ndhF-trnL-UAG</i>
11	52	74 894	P	74 894	<i>psbT-psbN</i>	<i>psbT-psbN</i>
12	49	92 176	F	92 230	<i>ycf2</i>	<i>ycf2</i>
13	49	92 176	P	153 608	<i>ycf2</i>	<i>ycf2</i>
14	49	92 230	P	153 662	<i>ycf2</i>	<i>ycf2</i>
15	49	153 608	F	153 662	<i>ycf2</i>	<i>ycf2</i>
16	40	57 691	P	57 691	<i>rbcL-accD</i>	<i>rbcL-accD</i>
17	49	101 467	P	101 467	<i>rps12-trnV-GAC</i>	<i>rps12-trnV-GAC</i>
18	49	101 467	F	144 371	<i>rps12-trnV-GAC</i>	<i>trnV-GAC-rps12</i>
19	49	144 371	P	144 371	<i>trnV-GAC-rps12</i>	<i>trnV-GAC-rps12</i>
20	39	44 550	F	99 648	<i>ycf3*</i>	<i>rps12-trnV-GAC</i>
21	39	44 550	P	146 200	<i>ycf3*</i>	<i>trnV-GAC-rps12</i>
22	36	118 693	P	118 734	<i>ccsA-ndhD</i>	<i>ccsA-ndhD</i>
23	42	6 763	P	102 216	<i>trnQ-UUG</i>	<i>rps12-trnV-GAC</i>
24	42	6 763	F	143 629	<i>trnQ-UUG</i>	<i>trnV-GAC-rps12</i>
25	40	118 152	P	118 152	<i>ccsA</i>	<i>ccsA</i>
26	30	6 727	P	102 263	<i>rps16-trnQ-UUG</i>	<i>rps12-trnV-GAC</i>
27	30	6 727	F	143 594	<i>rps16-trnQ-UUG</i>	<i>trnV-GAC-rps12</i>
28	30	6 775	P	102 216	<i>trnQ-UUG</i>	<i>rps12-trnV-GAC</i>
29	30	6 775	F	143 641	<i>trnQ-UUG</i>	<i>trnV-GAC-rps12</i>
30	30	8 005	P	45 992	<i>psbI-trnS-GCU</i>	<i>trnS-GGA</i>
31	30	122 162	P	122 198	<i>ndhG-ndhI</i>	<i>ndhG-ndhI</i>
32	32	7 958	P	7 958	<i>psbI-trnS-GCU</i>	<i>psbI-trnS-GCU</i>
33	32	103 909	F	104 240	<i>rps12-trnV-GAC</i>	<i>rps12-trnV-GAC</i>
34	32	103 909	P	141 615	<i>rps12-trnV-GAC</i>	<i>trnV-GAC-rps12</i>
35	32	104 240	P	141 946	<i>rps12-trnV-GAC</i>	<i>trnV-GAC-rps12</i>
36	32	141 615	F	141 946	<i>trnV-GAC-rps12</i>	<i>trnV-GAC-rps12</i>
37	31	92 176	F	92 248	<i>ycf2</i>	<i>ycf2</i>
38	31	92 176	P	153 608	<i>ycf2</i>	<i>ycf2</i>
39	31	92 248	P	153 680	<i>ycf2</i>	<i>ycf2</i>
40	31	153 608	F	153 680	<i>ycf2</i>	<i>ycf2</i>
41	33	39 570	F	41 794	<i>psaB</i>	<i>psaA</i>
42	32	39 562	F	41 786	<i>psaB</i>	<i>psaA</i>
43	30	36 309	P	45 992	<i>psbC-trnS-UGA</i>	<i>trnS-GGA</i>
44	30	47 583	R	122 181	<i>trnT-UGU-trnL-UAA</i>	<i>ndhG-ndhI</i>

F: a forward repeat; R: an inverted repeat; C: a complementary repeat; P: a palindrome repeat; \*: the repeat sequence located in the intron region.

表 6 扁果草叶绿体基因组中的 SSR

Table 6 SSR in chloroplast genome of *I. anemonoides*

No.	Type	SSR	Size	Start	End	Location
1	p1	(T)10	10	4 318	4 327	IGS ( <i>trnK-UUU-rps16</i> )
2	p1	(A)11	11	4 478	4 488	IGS ( <i>trnK-UUU-rps16</i> )
3	p2	(AT)6	12	5 061	5 072	<i>rps16</i> *
4	p1	(T)12	12	6 048	6 059	IGS ( <i>rps16-trnQ-UUG</i> )
5	c	(T)10ataaaaggaaagtgaatggaatgctctaaatcat(C)10	53	6 440	6 492	IGS ( <i>rps16-trnQ-UUG</i> )
6	p1	(A)11	11	7 364	7 374	IGS ( <i>psbK-psbI</i> )
7	p1	(T)15	15	8 151	8 165	IGS ( <i>trnS-GCU-trnG-UCC</i> )
8	p1	(T)10	10	8 825	8 834	<i>trnG-UCC</i> *
9	p1	(T)10	10	11 931	11 940	<i>atpF</i> *
10	p1	(T)10	10	15 963	15 972	IGS ( <i>rps2-rpoC2</i> )
11	p1	(T)11	11	18 043	18 053	<i>rpoC2</i>
12	p1	(T)10	10	25 742	25 751	<i>rpoB</i>
13	p1	(A)10	10	30 804	30 813	IGS ( <i>PsbM-trnD-GUC</i> )
14	p1	(A)10	10	31 225	31 234	IGS ( <i>trnD-GUC-trnY-GUA</i> )
15	p1	(T)10	10	33 551	33 560	IGS ( <i>trnT-GGU-psbD</i> )
16	p1	(A)14	14	43 346	43 359	IGS ( <i>PsaA-ycf3</i> )
17	p1	(T)10	10	46 948	46 957	IGS ( <i>rps4-trnT-UGU</i> )
18	p1	(T)15	15	47 066	47 080	IGS ( <i>rps4-trnT-UGU</i> )
19	c	(AT)7tatataaatgt(A)13	38	47 248	47 285	IGS ( <i>rps4-trnT-UGU</i> )
20	p1	(A)14	14	48 051	48 064	IGS ( <i>trnT-UGU-trnL-UAA</i> )
21	p1	(A)11	11	48 215	48 225	<i>trnL-UAA</i> *
22	c	(T)11a(T)11	23	48 764	48 786	IGS ( <i>trnL-UAA-trnF-GAA</i> )
23	p1	(A)12	12	49 411	49 422	IGS ( <i>trnF-GAA-ndhJ</i> )
24	p1	(T)11	11	51 087	51 097	<i>ndhK</i>
25	p1	(T)10	10	52 048	52 057	IGS ( <i>ndhC-trnV-UAC</i> )
26	p2	(AT)6	12	57 684	57 695	IGS ( <i>rbcL-accD</i> )
27	p1	(T)11	11	60 543	60 553	IGS ( <i>PsaI-ycf4</i> )
28	p1	(T)12	12	69 012	69 023	IGS ( <i>rps18-rpl20</i> )
29	p1	(T)13	13	70 186	70 198	IGS ( <i>rpl20-rps12</i> )
30	p1	(T)11	11	71 185	71 195	<i>clpP</i> *
31	p1	(T)10	10	75 615	75 624	<i>petB</i> *
32	p1	(T)12	12	80 132	80 143	IGS ( <i>rps11-rpl36</i> )
33	c	(T)10g(A)13	24	80 727	80 750	IGS ( <i>infA-rps8</i> )
34	p1	(T)13	13	82 829	82 841	<i>rpl16</i> *
35	p1	(A)12	12	83 363	83 374	IGS ( <i>rpl16-rps3</i> )
36	p1	(T)14	14	84 984	84 997	IGS ( <i>rps19-rpl2</i> )
37	p1	(A)10	10	111 932	111 941	IGS ( <i>rrn5-trnR-ACG</i> )
38	p1	(A)10	10	114 117	114 126	<i>ycf1</i>
39	p1	(A)11	11	121 281	121 291	IGS ( <i>ndhE-ndhG</i> )
40	p1	(A)10	10	122 272	122 281	IGS ( <i>ndhG-ndhI</i> )
41	p2	(AT)6	12	123 657	123 668	<i>ndhA</i> *
42	c	(T)10atcaattatttcatatttcatataattattagta(T)11	55	128 370	128 424	<i>ycf1</i>
43	p1	(T)10	10	129 681	129 690	<i>ycf1</i>
44	p1	(T)11	11	130 358	130 368	<i>ycf1</i>
45	p1	(T)10	10	131 762	131 771	<i>ycf1</i>
46	p1	(T)10	10	133 947	133 956	IGS ( <i>trnR-ACG-rrn5</i> )
47	p1	(A)14	14	160 891	160 904	IGS ( <i>rpl2-trnH-GUG</i> )

Note: P: single SSR type; C: composite SSR type; IGS: intergenic spacer; \*: is located in the intron region.

## 2.6 叶绿体基因组共线性分析

利用 Mauve 软件预测扁果草与近缘物种在叶绿体基因组水平的共线性关系 (图3): 扁果草和乳突拟耧斗菜有5个共性区域, 其他8个物种仅有4个共性区域, 叶绿体基因组均不存在倒位或重排现象。第1、2区域的长度十分相近, 第3、4区域的长度存在一定的差异。欧亚唐松草 (*Thalictrum minus*)、距瓣尾囊草 (*Urophysa rockii*)、天葵 (*Semiaquilegia adoxoides*)、蕨叶人字果 (*Dichocarpum dalzielii*)、乳突拟耧斗菜存在间隔区, 其大小可能是造成区域长度产生差异的主要原因。两个扁果草叶绿体基因组相似性为 95.51% (图4)。在扁果草与东北扁果草叶绿体基因组存在多处空白分割区域, 基因间存在内含子间

隔, 且内含子在位置、数量和长度上存在显著差异。

## 2.7 叶绿体基因组的系统发育分析

选取扁果草在内的15种植物进行系统发育树的构建, 以白毛茛 (*Hydrastis canadensis*) 作为外类群, 基于 nrDNA ITS 序列以贝叶斯法构建的发育树包含12个节点, 节点支持率均大于75%, 表明聚类结果可信度较高 (图 5)。然而, 通过系统发育分析发现扁果草属的几个物种并未如预期一样形成一个单系群, 而是聚为两支, 扁果草和假耧斗菜 (*P. microphylla*) 以98%的支持率聚在一个分支上, 而 *I. thalictroides*、东北扁果草聚在另一分支上, 其支持率也为100%。

基于 *matK*、*trnL-F* 和 *rbcL* 三个叶绿体基因组序列构建了系统发育树 (图6)。在系统发育树

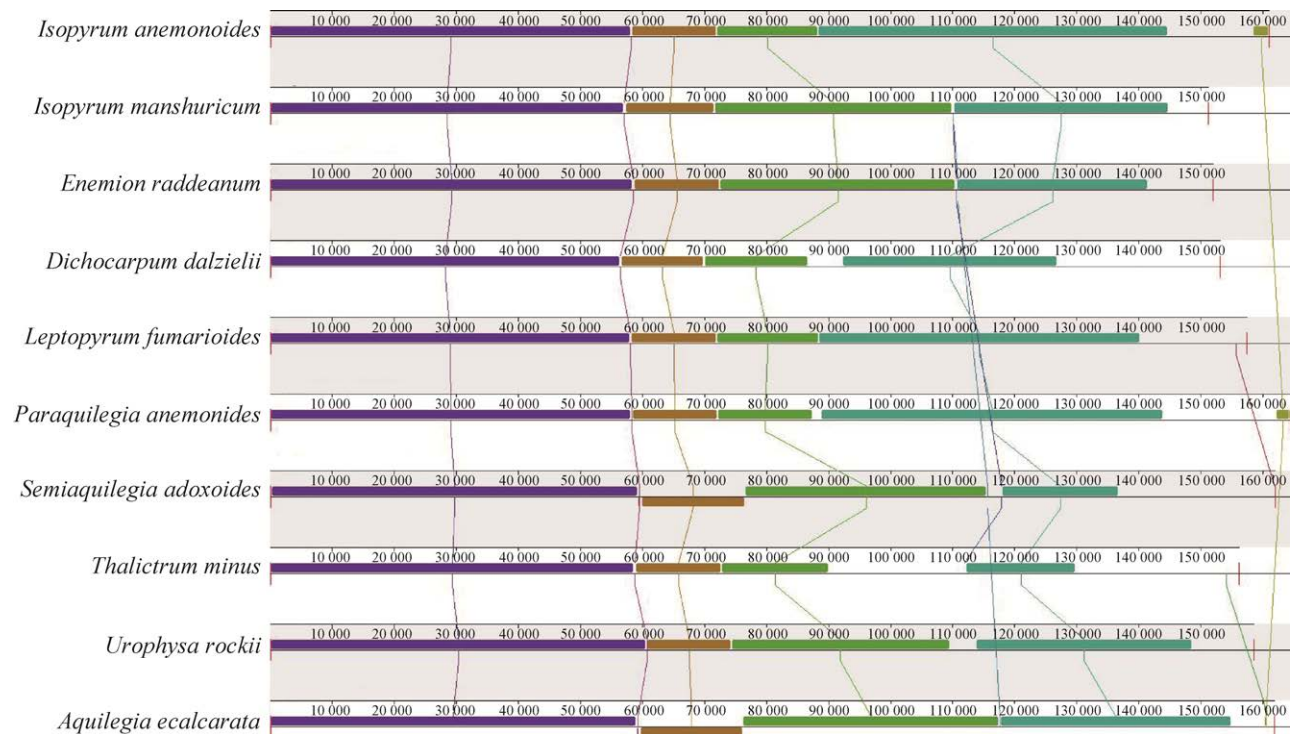


图3 利用 Mauve 对 10 个叶绿体基因组的共线性分析 不同的叶绿体基因组之间的同源区域由相同的色块表示, 并以线条连接

Figure 3 Colinearity analysis of 10 chloroplast genomes by Mauve. Homologous regions between different chloroplast genomes are represented by the same color blocks linked by lines.

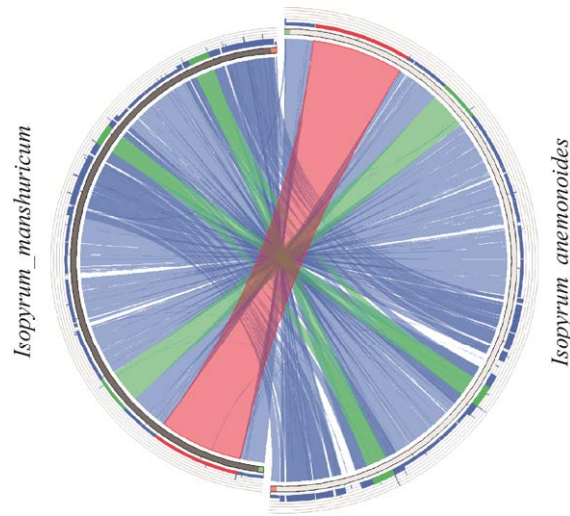


图4 两种扁果草相似性比较 序列根据相似的区域和相似的长度用不同的颜色连接起来，直方图中用得分的颜色表示。图中连接线之间的空白区域表示这两个物种没有任何相似性

Figure 4 Comparison of similarity between two species of *Isopyrum*. The sequences are connected by different colors according to similar regions and similar lengths, with scored colors shown in the histograms. The blank area between the connecting lines in the figure indicates that the two species do not have any similarity.

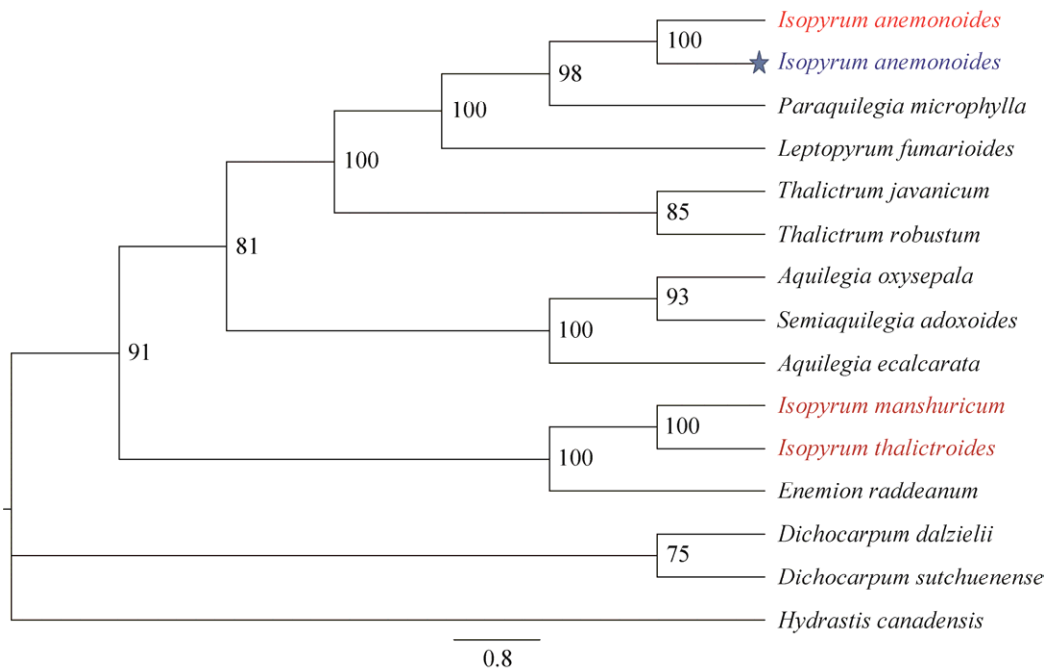


图5 基于 nrDNA ITS 序列构建的系统发育树 *H. canadensis* 为外类群，节点支持值是贝叶斯后验概率。本研究材料扁果草用深蓝色星号标记，用于系统发育分析的其他物种和 NCBI 序列号见表 1

Figure 5 Phylogenetic tree based on nrDNA ITS sequences. *H. canadensis* was used as an outgroup, node support values are Bayesian posterior probabilities. The material used in this study is marked with a dark blue asterisk. Species and NCBI accession number used for phylogenetic analysis are listed in Table 1.

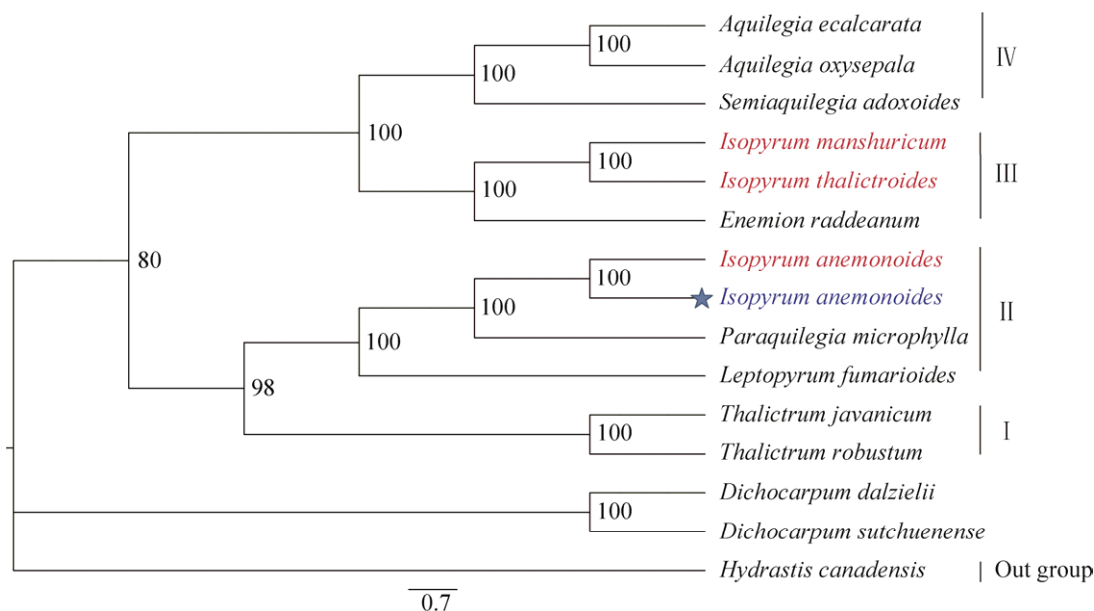


图 6 基于 3 个叶绿体片段 (*rbcL*、*matK*、*trnL-F*) 构建的系统发育树 *H. canadensis* 为外类群, 节点支持值是贝叶斯后验概率. 本研究材料扁果草用深蓝色星号标记, 用于系统发育分析的其他物种和 NCBI 序列号见表 1

Figure 6 Phylogenetic tree based on three chloroplast fragments (*rbcL*, *matK*, *trnL-F*). *H. canadensis* was used as an outgroup, node support values are Bayesian posterior probabilities. The material used in this study is marked with a dark blue asterisk. Species and NCBI accession number used for phylogenetic analysis in the present study are listed in Table 1.

上 15 种植物叶绿体基因组可分为两大类 4 个小类, 人字果属和白毛茛为一个大类, 其他的 12 个种为另一大类, 唐松草属物种爪哇唐松草 (*T. javanicum*) 和粗壮唐松草 (*T. robustum*) 聚于类 I; 蓝堇草 (*Leptopyrum fumarioides*)、假耧斗菜和扁果草聚于类 II; 拟扁果草 (*Enemion raddeanum*)、*I. thalictroides*、东北扁果草聚于类 III; 天葵、无距耧斗菜 (*Aquilegia ecalcarata*)、尖萼耧斗菜 (*A. oxysepala*) 聚于类 IV。基于 nrDNA ITS (图 5) 与 3 个叶绿体基因片段 (图 6) 的系统发育分析结果一致, 扁果草与假耧斗菜均聚在一个支上, 并与蓝堇草形成姐妹群, 未与扁果草属中 *I. thalictroides* 和东北扁果草形成一个单系群, 并且亲缘关系较远。

### 3 讨论

扁果草与扁果草属其他植物及毛茛科植物的叶绿体基因组相似, 均为四分体结构 (图 1), 包含 1 个 LSC 和 SSC 区以及 2 个 IR 区, 符合大多数被子植物叶绿体基因组的结构特征<sup>[17]</sup>。扁果草叶绿体基因组长度为 161 034 bp, 东北扁果草基因组长度为 151 243 bp, 表明扁果草具有较大的叶绿体基因组。通常情况下, 植物叶绿体基因组 LSC 区长度为 80–90 kb, SSC 区长度为 16–27 kb, IR 区长度为 20–28 kb<sup>[18]</sup>。扁果草叶绿体基因组 LSC 区长度约为 85 kb, SSC 区长度约为 17 kb, IR 区长度约为 29 kb, 可见扁果草叶绿体基因组较大的原因主要是由较大的 IR 区的长度引起的。此外, 扁果草叶绿体基

基因组蛋白编码基因第 3 位密码子具有明显的 A/T 偏好性, 这与其他高等植物相似。

重复序列对基因组重排和变异具有重要意义, 可能促进叶绿体基因组重排, 增加种群遗传多样性。散在重复序列可被应用到系统发育研究, 并作为种群研究的遗传标记<sup>[19]</sup>。重复序列普遍存在于基因间隔区 (intergenic spacer, IGS)<sup>[20-22]</sup>, 扁果草叶绿体基因组中共检测到 44 个散在重复序列 (表 5), 大多数重复序列分布在基因间隔区和 *ycf2* 基因中, 而东北扁果草中共检测到 49 个散在重复序列, 其中 35 个正向重复, 14 个回文重复, 未有反向重复。叶绿体基因组标记在进化和分类学研究方面比核 DNA 标记更有优势, 因为它们在大多数植物中是母系遗传, 突变率低<sup>[23-24]</sup>。由于叶绿体基因组 SSRs 具有高度的可重复性和多态性, 因而通常用于物种鉴定和遗传分析。在扁果草叶绿体基因组中, 共检测出 47 个 SSR, 其中包含单核苷酸重复、二核苷酸重复和复杂核苷酸重复。在东北扁果草叶绿体基因组中, 发现了与扁果草同样类型的 35 个 SSRs。

倒位是引起毛茛科叶绿体基因排列顺序变化的主要原因<sup>[25]</sup>, 之前研究显示, 天葵和无距耧斗菜等 8 个物种中共存在 5 种不同长度的大片段倒位现象<sup>[26]</sup>。通过对扁果草与近缘种叶绿体基因组共线性分析发现, 扁果草和乳突拟耧斗菜包含 5 个共性区域, 但其余物种仅有 4 个共性区域, 叶绿体基因组均不存在倒位或重排现象。通过 Circoletto 分别对扁果草与东北扁果草叶绿体基因组进行比较, 二者相似度为 95.51% (图 4), 而且扁果草蛋白编码基因数目比东北扁果草多 7 个, 但扁果草并没有与该属的东北扁果草聚为一支, 而是与假耧斗菜形成一个小分支。这说明目前划分的扁果草属不是一个单系群, 虽然扁果草与东北扁果草在形态上有一定的相似性, 但遗传差异较大。应该将该属物种进行

分类修订处理, 将扁果草独立为一个属, 东北扁果草和 *I. thalictroides* 独立为另一个属。

## 4 结论

本研究利用 NOVOplasty、Sequin 等软件进行扁果草叶绿体基因组组装和注释, 获得完整的叶绿体基因组序列。扁果草叶绿体基因组全长为 161 034 bp, 有 130 个基因, 包括 85 个蛋白编码基因、37 个 tRNA 基因和 8 个 rRNA 基因。叶绿体基因组中, 共有 44 个散在重复序列, 以回文重复和正向重复为主, 检测到 47 个 SSR 位点主要分布于基因间隔区 (IGS), 以单碱基重复为主, 其主要由 A/T 组成。密码子偏好性分析表明, 亮氨酸 (Leu) 为使用频率最高的氨基酸 (9.78%), 色氨酸 (Trp) 为使用频率最低的氨基酸 (1.33%)。基于叶绿体全基因组的共线性分析表明, 扁果草与其他近缘种共线性良好, 均未发现倒位或重排现象。此外, 系统发育分析显示, 扁果草属不是一个单系群, 扁果草并没有与该属的东北扁果草和 *I. thalictroides* 聚在一支, 而是与假耧斗菜有很近的亲缘关系。该研究为扁果草属分子标记开发、种间鉴定以及该物种的演化过程及分类学地位提供科学依据。

## REFERENCES

- [1] Dong WP, Xu C, Cheng T, et al. Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biol Evol*, 2013, 5(5): 989-997.
- [2] Cheng H, Li JF, Zhang H, et al. The complete chloroplast genome sequence of strawberry (*Fragaria × ananassa* Duch.) and comparison with related species of Rosaceae. *PeerJ*, 2017, 5(10): e3919.
- [3] Wicke S, Schneeweiss GM, DePamphilis CW, et al. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*, 2011, 76(3-5): 273-297.
- [4] 中国科学院中国植物志委员会. 中国植物志, 第 27 卷. 北京: 科学出版社, 1974.

- Chinese Flora Commission of the Chinese Academy of Sciences. Flora of China, volume 27. Beijing: Science Press, 1974 (in Chinese).
- [5] Ivanovska N, Philipov S, Istatkova R, et al. Antimicrobial and immunological activity of ethanol extracts and fractions from *Isopyrum thalictroides*. J Ethnopharmacol, 1996, 54(2/3): 143-151.
- [6] Ivanovska N, Nikolova P, Hristova M, et al. Complement modulatory activity of bisbenzylisoquinoline alkaloids isolated from *Isopyrum thalictroides*--I. Influence on classical pathway in human serum. Int J Immunopharmacol, 1999, 21(5): 325-336.
- [7] 张少峰, 武彦义. 扁果草属植物的化学成分和药理活性研究进展. 阴山学刊(自然科学版), 2005, 19(1): 49-52.  
Zhang SF, Wu YY. Phytochemical and pharmacological research progress in medicinal plant *Isopyrum*. Yinshan Acad J (Nat Sci Ed), 2005, 19(1): 49-52 (in Chinese).
- [8] Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Res, 2017, 45(4): e18.
- [9] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res, 2019, 47(W1): W59-W64.
- [10] Lohse M, Drechsel O, Kahlau S, et al. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res, 2013, 41: (W1)W575-W581.
- [11] Jiao Y, Jia HM, Li XW, et al. Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). BMC Genomics, 2012, 13: 201.
- [12] Yang AH, Zhang JJ, Yao XH, et al. Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*. American J Botany, 2011, 98(5): e123-126.
- [13] Tang DQ, Lu JJ, Fang W, et al. Development, characterization and utilization of GenBank microsatellite markers in *Phyllostachys pubescens* and related species. Mol Breeding, 2010, 25: 299-311.
- [14] Powell W, Morgante M, McDevitt R, et al. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. PNAS, 1995, 92(17): 7759-7763.
- [15] 曾淑云. 地黄属植物叶绿体基因组研究[D]. 西安: 西北大学, 2017.  
Zeng SY. Chloroplast genome of *Rehmannia* (Orobanchaceae)[D]. Xi'an Northwest University, 2017 (in Chinese).
- [16] Ronquist F, Huelsenbeck JP. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics, 2003, 19(12): 1572-1574.
- [17] Jansen RK, Raubeson LA, Boore JL, et al. Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol, 2005, 395: 348-384.
- [18] Raubeson LA, Peery R, Chumley TW, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics, 2007, 8: 174.
- [19] Park I, Yang S, Choi G, et al. The complete chloroplast genome sequences of *Aconitum pseudolaeve* and *Aconitum longecassidatum*, and development of molecular markers for distinguishing species in the *Aconitum* Subgenus *Lycocotnum*. Molecules, 2017, 22(11): 2012.
- [20] Yan C, Du JC, Gao L, et al. The complete chloroplast genome sequence of watercress (*Nasturtium officinale* R. Br.): genome organization, adaptive evolution and phylogenetic relationships in Cardamineae. Gene, 2019, 699: 24-36.
- [21] Hong SY, Cheon KS, Yoo KO, et al. Complete chloroplast genome sequences and comparative analysis of *Chenopodium quinoa* and *C. album*. Front Plant Sci, 2017, 8: 1696.
- [22] Huang YY, Matzke AJM, Matzke M. Complete sequence and comparative analysis of the chloroplast genome of Coconut Palm (*Cocos nucifera*). PLoS One, 2013, 8(8): e74736.
- [23] Holwerda BC, Jana S, Crosby WL. Chloroplast and mitochondrial DNA variation in *HORDEUM VULGARE* and *HORDEUM SPONTANEUM*. Genetics, 1986, 114(4): 1271-1291.
- [24] Vanichanon A, Biake NK, Sherman JD, et al. Multiple origins of allopolyploid *Aegilops triuncialis*. Theor Appl Genet, 2003, 106(5): 804-810.
- [25] Johansson JT, Jansen RK. Chloroplast DNA variation and phylogeny of the Ranunculaceae. Plant Syst Evol, 1993, 187(1-4): 29-49.
- [26] 王媛媛, 百万富, 杨美青. 毛茛科植物叶绿体基因组结构变异及系统发育分析. 分子植物育种, 2020, 18(21): 7030-7043.  
Wang YY, Bai WF, Yang MQ. Structural variation and phylogenetic analysis of chloroplast genome in Ranunculaceae. Mol Plant Breed, 2020, 18(21): 7030-7043 (in Chinese).

(本文责编 郝丽芳)