

# 桑叶葡萄叶绿体基因组及其特征分析

杨亚蒙<sup>1</sup>, 焦 健<sup>2</sup>, 樊秀彩<sup>1</sup>, 张 颖<sup>1</sup>, 姜建福<sup>1</sup>, 李 民<sup>1</sup>, 刘崇怀<sup>1,\*</sup>

(<sup>1</sup>中国农业科学院郑州果树研究所, 郑州 450009; <sup>2</sup>河南农业大学园艺学院, 郑州 450002)

**摘 要:** 桑叶葡萄 (*Vitis ficifolia*) 是中国重要的葡萄属野生资源。以桑叶葡萄 ‘九里沟桑 4 号’ 为试验材料, 利用 Illumina HiSeq PE150 测序平台完成了其全基因组重测序, 并组装了其叶绿体基因组。注释结果表明, 桑叶葡萄叶绿体基因组全长 160 751 bp, 呈典型的四分体结构, 其大单拷贝区 (large single copy, LSC)、反向重复序列 (inverted repeat, IR) 和小单拷贝区 (small single copy, SSC) 的长度分别为 88 971、26 355 和 19 070 bp。注释共得到 133 个基因, 包含 88 个蛋白编码基因, 8 个 rRNA 基因, 37 个 tRNA 基因。在桑叶葡萄中共搜索到 84 个 SSR 位点, 其中单核苷酸、二核苷酸、三核苷酸、四核苷酸和五核苷酸重复基序个数分别为 55、8、9、9 和 3 个。用 MEGA 软件, 通过邻接法对包括 5 个葡萄种类的 32 个种的叶绿体基因组序列进行聚类分析, 结果表明, 桑叶葡萄与山葡萄 (*Vitis amurensis*)、欧亚种 ‘黑比诺’ (*V. vinifera* ‘Pinot Noir’)、夏葡萄 (*V. aestivalis*) 以及圆叶葡萄 (*V. rotundifolia*) 具有较近的亲缘关系。

**关键词:** 桑叶葡萄; 叶绿体基因组; 重测序; 聚类分析

**中图分类号:** S 633.1

**文献标志码:** A

**文章编号:** 0513-353X (2019) 04-0635-14

## Complete Chloroplast Genome Sequence and Characteristics Analysis of *Vitis ficifolia*

YANG Yameng<sup>1</sup>, JIAO Jian<sup>2</sup>, FAN Xiucui<sup>1</sup>, ZHANG Ying<sup>1</sup>, JIANG Jianfu<sup>1</sup>, LI Min<sup>1</sup>, and LIU Chonghuai<sup>1,\*</sup>

(<sup>1</sup>Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou 450009, China; <sup>2</sup>College of Horticulture, Henan Agricultural University, Zhengzhou 450002, China)

**Abstract:** *Vitis ficifolia* is an important wild resource in China. In this study, the whole genome resequencing of *V. ficifolia* was completed using Illumina HiSeq PE150 sequencing platform. The chloroplast genome was assembled. The results of the annotation indicated that the chloroplast genome of *V. ficifolia* has a total length of 160 751 bp, which is a typical tetrad structure. The length of LSC, IR and SSC regions are 88 971, 26 355 and 19 070 bp, respectively. The chloroplast genome contains a total of 133 genes, including 88 protein-coding genes, 8 rRNA genes, and 37 tRNA genes. We found 84 SSR loci in the *V. ficifolia* chloroplast genome, of which the number of mononucleotide, dinucleotide, trinucleotide, tetranucleotide and pentanucleotide repeat motifs was 55, 8, 9, 9 and 3, respectively. The NJ tree using chloroplast genome sequences of 32 species including 5 *Vitis* species was constructed by MEGA software. The results showed that *V. ficifolia* had a close genetic relationship with *V. amurensis*, Pinot Noir (*V.*

**收稿日期:** 2018-10-29; **修回日期:** 2019-04-02

**基金项目:** 国家现代农业产业技术体系建设专项资金项目 (CARS-29-yz-1); 中国农业科学院科技创新工程项目 (CAAS-ASTIP-2018-ZFRI); 农业部物种保护项目 (2130135-34); 中央级科研院所基本科研业务费专项 (1610192018105)

\* 通信作者 Author for correspondence (E-mail: liuchonghuai@caas.cn)

*vinifera*), *V. aestivalis* and *V. rotundifolia*.

**Keywords:** *Vitis ficifolia*; chloroplast genome; resequencing; cluster analysis

叶绿体是与植物光合作用直接相关的细胞器, 内共生起源学说认为叶绿体起源于细胞内共生的蓝藻 (Keeling, 2004)。与核基因组相比, 叶绿体基因组较小, 且属于单亲遗传, 其基因含量、基因顺序和碱基组成非常保守, 核苷酸替换频率较低, 因此, 叶绿体基因组在研究物种系统发育方面具有重要价值 (Wei et al., 2005; Leseberg & Duvall, 2009)。大多数被子植物的叶绿体基因组具有高度保守的四分体结构 (Jansen et al., 2005), 即由两个反向重复序列 (inverted repeat, IR) 将整个环状的叶绿体基因组分为一个大单拷贝区 (large single copy, LSC) 和一个小单拷贝区 (small single copy, SSC)。近年来, 随着大规模测序技术的快速发展, 使用 Roche 或 Illumina 平台的二代测序 (NGS) 技术为叶绿体和线粒体基因组组装提供了高通量、低成本和高效方法 (Moore et al., 2006; Cronn et al., 2008; Tangphatsornruang et al., 2009), 越来越多的植物叶绿体基因组被组装出来, 如美国蔓越莓 (Fajardo et al., 2013)、莱檬 (Su et al., 2014)、菠萝 (Nashima et al., 2015)、核桃 (Peng et al., 2015)、柿 (Fu et al., 2016)、板蓝根 (Chen et al., 2018) 等, 为研究物种进化 (Burke et al., 2016; Yang et al., 2016)、叶绿体转化等提供了技术支撑。

中国野生葡萄具有丰富的抗逆性和适应性, 是拓宽葡萄品种资源的重要基因库。然而由于种类繁多, 种间变异较为复杂, 中国野生葡萄的分类研究仍存在较大争议 (Wan et al., 2008)。桑叶葡萄 (*Vitis ficifolia*) 是重要的野生葡萄种类之一, 被认为是毛葡萄 (*Vitis heyneana*) 的一个亚种 (李朝奎, 1998; 张旭彤, 2012), 也有研究表明桑叶葡萄与毛葡萄亲缘关系较远, 应该划分为独立的种 (张永辉 等, 2011)。本研究中通过高通量测序获得桑叶葡萄的叶绿体基因组全序列, 通过桑叶葡萄与其近缘种的叶绿体基因组比较分析, 并结合系统发育分析探讨葡萄属叶绿体基因组结构特征和变异, 以期深入揭示葡萄属植物的系统演化关系提供数据基础。此外, 桑叶葡萄叶绿体基因组序列的鉴定还可为葡萄属的遗传多样性等研究提供重要依据。

## 1 材料与方法

### 1.1 试验材料及 DNA 的提取及测序

以桑叶葡萄 ‘九里沟桑 4 号’ 为试验材料, 其嫩梢幼叶于 2016 年 4 月采自中国农业科学院郑州果树研究所国家果树种质郑州葡萄圃。

使用 TIANGEN 植物基因组快速提取试剂盒提取其 DNA, 用琼脂糖凝胶电泳和分光光度计检测 DNA 的完整性和质量。

利用 Covaris 超声波破碎仪随机打断 DNA 序列, 然后在序列末端修复、加 A 尾、加测序接头, 再经过纯化、PCR 扩增等完成整个文库制备工作。对建好的文库进行质检, 合格的文库用 Illumina HiSeq PE150 (北京诺和致源生物信息科技有限公司) 双末端测序策略进行测序。

### 1.2 序列拼接及注释

从 GenBank 下载已发表的 5 个葡萄的叶绿体基因组序列 (GenBank 登录号: NC\_007957、NC\_023790、NC\_029454、AB856290 和 AB856291) 为参照, 利用 SPAdes v3.6.0 软件 (Bankevich et al., 2012) 对桑叶葡萄叶绿体基因组进行组装, 参数采用默认设置。

序列注释首先通过近源物种的蛋白编码序列直接进行 blastn 比对确认基因有无和边界, 随后利用在线工具 DOGMA (<http://dogma.cccb.utexas.edu/>) 注释叶绿体基因组中的基因, 默认参数, 并结合 NR (<http://www.ncbi.nlm.nih.gov/>) 数据库对基因进行功能注释。利用 tRNAscan-SE 在线网站对 tRNA 进行注释。利用 RNAmmer 1.2 Server (<http://www.cbs.dtu.dk/services/RNAmmer/>) 对 rRNA 进行注释。

注释完成之后, 输出 Sequin 文件, 提交至 GenBank, 获得序列登录号 MH638290。

利用 OGDRAW (<http://ogdraw.mpimp-golm.mpg.de/cgi-bin/ogdraw.pl>) 软件对桑叶葡萄叶绿体基因组进行作图。

### 1.3 叶绿体基因组数据分析

采用 MISA 软件 (<http://pgrc.ipk-gatersleben.de/misa/>) 对桑叶葡萄叶绿体基因组序列进行微卫星扫描分析, 参数设置为单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸的重复数分别为 10、5、4、3、3 和 3。利用 Tandem repeats finder v4.04 软件 (Benson, 1999) 检测串联重复, 默认参数。运用 CodonW1.4.2 软件 (<http://mobyle.pasteur.fr/cgi-bin/portal.py?form=codonw>) 对叶绿体基因的 RSCU (relative synonymous codon usage) 进行统计以及密码子偏好性分析。

### 1.4 系统发育分析

从 NCBI 数据库选取山葡萄 (KX499471)、欧亚种葡萄‘黑比诺’ (DQ424856)、夏葡萄 (KT997470)、圆叶葡萄 (KF976463)、三叶青 (KT033563)、蛇葡萄 (KT831767) 等 31 个种植物叶绿体基因组序列与桑叶葡萄进行聚类分析。使用 HomBlocks 软件 (<https://github.com/fenghen360/HomBlocks>) 进行共线性模块分析, 得到共有模块的基因。用 MEGA 软件的 NJ 法构建系统进化树, bootstrap 值设为 1 000。

## 2 结果与分析

### 2.1 桑叶葡萄叶绿体基因组的基本特征

桑叶葡萄的完整叶绿体基因组长度为 160 751 bp, 包括一对 26 355 bp 的反向重复序列 (IRA 和 IRB), 被一个长为 19 070 bp 的小单拷贝 (SSC) 区域和一个长为 88 971 bp 的大单拷贝 (LSC) 区域分开 (图 1)。

叶绿体基因组注释结果表明, 桑叶葡萄叶绿体基因组具有 133 个功能基因, 其中包括 88 个蛋白质编码基因 (包括 *ycf1* 的一份假基因拷贝), 8 个核糖体 RNA 基因和 37 个 tRNA 基因 (表 1)。17 个基因在 IR 区域复制, 包括 6 个蛋白编码基因 (*ndhB*、*rpl2*、*rpl23*、*rps7*、*ycf2* 和 *ycf15*), 7 个 tRNA 基因 (*trnA*-UGC、*trnI*-CAU、*trnI*-GAU、*trnL*-CAA、*trnN*-GUU、*trnR*-ACG 和 *trnV*-GAC) 和 4 个 rRNA 基因 (*rrn4.5*、*rrn5*、*rrn16* 和 *rrn23*)。IR 区域的 GC 含量 (43%) 高于 LSC (35.4%) 和 SSC (31.7%), 与先前的研究 (Qian et al., 2013) 一致, 4 个 rRNA 均分布在 IR 区域是造成其 GC 含量偏高的重要原因 (Asaf et al., 2017)。

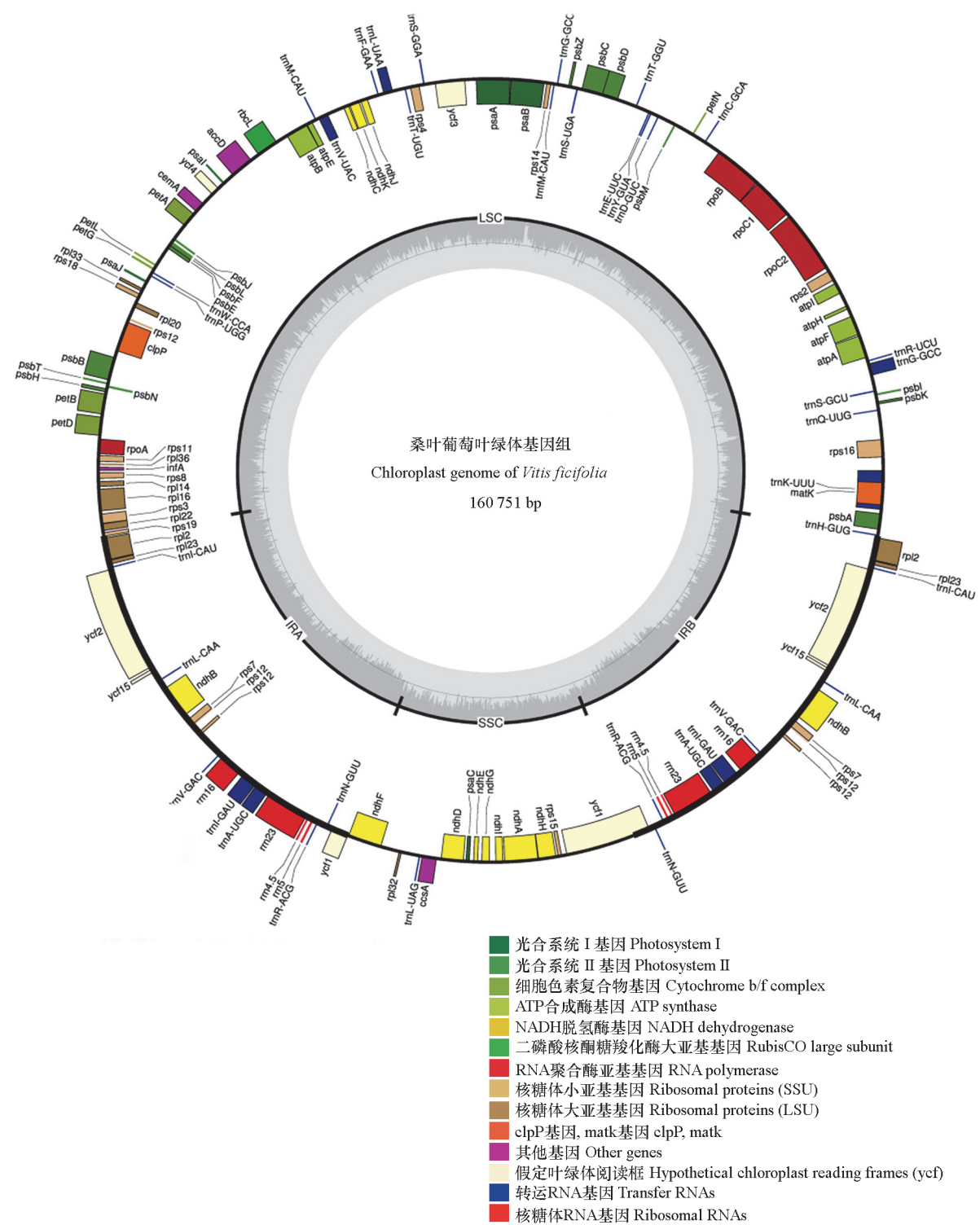


图 1 桑叶葡萄叶绿体基因组物理图谱

内侧基因顺时针；外侧基因逆时针转录。

Fig. 1 Gene map of the chloroplast genome of *Vitis ficifolia*

Genes drawn inside the circle are transcribed clockwise, and those outside are counter clockwise.

表 1 桑叶葡萄叶绿体基因组注释基因信息  
Table 1 Genes present in the chloroplast genome of *Vitis ficifolia*

基因分类 Category for gene	基因分组 Group of gene	基因名称 Name of gene
自我复制 Self replication	rRNA	<i>rrn4.5<sup>c</sup>, rrn5<sup>c</sup>, rrn16<sup>c</sup>, rrn23<sup>c</sup></i>
	tRNA	<i>trnH-GUG, trnK-UUU<sup>a</sup>, trnQ-UUG, trnS-GCU, trnR-UCU, trnG-GCC<sup>a</sup>, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnG-GCC, trnfM-CAU, trnS-GGA, trnT-UGU, trnL-UAA<sup>a</sup>, trnF-GAA, trnV-UAC<sup>b</sup>, trnM-CAU, trnW-CCA, trnP-UGG, trnI-CAU<sup>c</sup>, trnL-CAA<sup>c</sup>, trnI-GAU<sup>abc</sup>, trnV-GAC<sup>c</sup>, trnA-UGC<sup>abc</sup>, trnR-ACG<sup>c</sup>, trnN-GUU<sup>c</sup>, trnL-UAG</i>
	核糖体小亚基 Ribosomal protein (SSU)	<i>rps16<sup>a</sup>, rps2, rps14, rps4, rps18, rps11, rps8, rps3, rps19, rps7<sup>c</sup>, rps12<sup>bc</sup>, rps15</i>
	核糖体大亚基 Ribosomal protein (LSU)	<i>rpl33, rpl20, rpl36, rpl14, rpl16<sup>a</sup>, rpl22, rpl2<sup>ac</sup>, rpl23<sup>c</sup>, rpl32</i>
	RNA 聚合酶亚基 RNA polymerase	<i>rpoC2, rpoC1<sup>a</sup>, rpoB, rpoA</i>
光合作用相关 Genes for photosynthesis	光合系统 I Photosystem I	<i>psaA, psaB<sup>c</sup>, psaI, psaJ, psaC</i>
	光合系统 II Photosystem II	<i>psbA, psbB, psbK, psbI, psbM, psbD, psbC, psbJ, psbL, psbT, psbH, psbN, psbZ</i>
	细胞色素复合体 Cytochrome b/f complex	<i>petN, petA, psbF, psbE, petG, petB<sup>a</sup>, petD<sup>a</sup>, petL</i>
	ATP 合酶 ATP synthase	<i>atpA, atpF<sup>a</sup>, atpH, atpI, atpE, atpB</i>
	依赖 ATP 的蛋白酶单元 p	<i>clpP<sup>b</sup></i>
	ATP-dependent protease subunit p	
	二磷酸核酮糖羧化酶大亚基	<i>rbcL</i>
	RubiscoCO large subunit	
	NADH 脱氢酶	<i>ndhJ, ndhK, ndhC, ndhB<sup>ac</sup>, ndhF, ndhD, ndhE, ndhG, ndhI, ndhA<sup>a</sup>, ndhH</i>
其他 Other genes	NADH dehydrogenase	
	成熟酶 Maturase	<i>matK</i>
	包裹膜蛋白 Envelop membrane protein	<i>cemA</i>
	乙酰辅酶 A 羧化酶亚基	<i>accD</i>
	Subunit of acetyl-CoA-carboxylase	
	c 型细胞色素合成	<i>ccsA</i>
	c-type cytochrome synthesis ccsA gene	
	转录起始因子	<i>infA</i>
	Translation initiation factor IF-1	
未知功能 Genes of unknow function	假定叶绿体阅读框	<i>ycf3<sup>b</sup>, ycf2<sup>c</sup>, ycf4, ycf1<sup>c</sup>, ycf15<sup>c</sup></i>
	Hypothetical chloroplast reading frames	

注: <sup>a</sup>表示基因含有 1 个内含子; <sup>b</sup>表示基因含有 2 个内含子; <sup>c</sup>表示含有 2 个拷贝基因。  
Note: <sup>a</sup> and <sup>b</sup> present an intron and two introns in protein-coding genes, respectively. <sup>c</sup> present two copies of genes.

内含子可以调节基因的转录速率，在基因结构和功能方面有着重要作用。统计表明，桑叶葡萄叶绿体基因组中 18 个基因包含内含子。其中，10 个蛋白编码基因和 6 个 tRNA 基因含有一个内含子，2 个蛋白编码基因 (*ycf3* 和 *clpP*) 含有两个内含子，*rps12* 基因是一个反式剪接基因 (trans-spliced gene)，其 5'端位于 LSC 区域，3'端位于 IR 区域。这些内含子长度范围在 516 ~ 2 506 bp 之间，其中，*trnK-UUU* 基因的内含子最长，为 2 506 bp (表 2)。此外，桑叶葡萄含有内含子的数量和类型与欧亚种葡萄‘黑比诺’和山葡萄一致 (Jansen et al., 2006; Xie et al., 2016)，表明了葡萄属植物叶绿体基因组高度的保守性。

表 2 桑叶葡萄叶绿体基因组中含内含子的基因信息  
Table 2 Information of gene introns in the chloroplast genome of *Vitis ficifolia*

基因 Gene	位置 Location	外显子 I /bp Exon I	内含子 I /bp Intron I	外显子 II /bp Exon II	内含子 II /bp Intron II	外显子 III /bp Exon III
<i>trnK-UUU</i>	LSC	29	2 506	37		
<i>rps16</i>	LSC	40	900	236		
<i>trnG-GCC</i>	LSC	23	711	37		
<i>atpF</i>	LSC	144	747	414		
<i>rpoC1</i>	LSC	432	763	1 617		
<i>ycf3</i>	LSC	126	727	228	743	153
<i>trnL-UAA</i>	LSC	37	516	50		
<i>trnV-UAC</i>	LSC	39	573	37		
<i>clpP</i>	LSC	71	814	292	631	228
<i>petB</i>	LSC	6	688	642		
<i>petD</i>	LSC	8	734	475		
<i>rpl16</i>	LSC	9	1 063	399		
<i>rpl2</i>	LSC	390	664	435		
<i>ndhB</i>	IR	777	679	756		
<i>trnI-GAU</i>	IR	42	944	35		
<i>trnA-UGC</i>	IR	38	803	35		
<i>ndhA</i>	SSC	552	1 131	540		
<i>rps12*</i>	LSC-IR	114	—	232	540	26

注: \*反式剪接基因。

Note: \* trans-spliced gene.

## 2.2 桑叶葡萄密码子的偏好性

从桑叶葡萄叶绿体基因组中得到了 87 条 CDS (coding DNA sequence) 序列, 为保证结果的准确性, 剔除重复基因序列和长度小于 300 bp 序列, 剩余 53 条 CDS 序列用于密码子的偏好性分析。RSCU 值大于 1.00 的密码子为 28 个, 其中, 22 个以 A 或 U 结尾, 6 个以 G 或 C 结尾 (表 3)。说明桑叶葡萄叶绿体基因组密码子偏爱以 A 或 U (T) 结尾, 不偏好以 G 或 C 结尾的密码子, 编码亮氨酸 (Leu) 的密码子数量最多, 为 1 627 个。结果表明, 桑叶葡萄叶绿体基因组与大多数被子植物类似, 偏好使用 AT 结尾的密码子。

## 2.3 重复序列分析

简单序列重复 (SSR) 或微卫星 DNA 是指分布在整个基因组中 1 ~ 6 bp 的重复序列 (Asaf et al., 2017)。从桑叶葡萄叶绿体基因组检测出 84 个 SSR 位点, 其中, 单核苷酸、二核苷酸、三核苷酸、四核苷酸和五核苷酸重复分别有 55、8、9、9 和 3 个 (表 4), 且 54 个单核苷酸重复由 A 或 T 组成。欧亚种葡萄 ‘赤霞珠’ 叶绿体 SSR 同样偏向使用 A/T 碱基 (谢海坤 等, 2017), 这与先前的研究一致, 即植物叶绿体 SSR 标记富含 A/T 重复 (Kuang et al., 2011)。对 SSR 的位置进行分析发现, 大部分 SSR 都分布在基因组的非编码区 (表 5)。长度大于或等于 30 bp 的长重复序列, 在基因组的重排过程中可能扮演着重要的角色 (Cavalier-Smith, 2002; Nie et al., 2012)。桑叶葡萄叶绿体基因组中共检测出来 33 个串联重复, 且绝大多数长度范围在 30 ~ 89 bp 之间, 此外, 82% 的串联重复序列位于非编码区 (表 6)。

表 3 桑叶葡萄各氨基酸相对同义密码子使用度  
Table 3 RSCU analysis of protein coding region in *Vitis ficifolia*

RSCU				RSCU				
氨基酸 Amino acid	密码子 Codon	数量 Number	相对同义密码子使用度 Relative synonymous codon usage	氨基酸 Amino acid	密码子 Codon	数量 Number	相对同义密码子使用度 Relative synonymous codon usage	
Phe	UUU	845	1.03	Tyr	UAU	620	1.26	
	UUC	794	0.97			UAC	368	0.74
Leu	UUA	199	0.73	TER <sup>*</sup>	UAA	536	1.59	
	UUG	607	2.24			UAG	223	0.66
	CUU	330	1.22	His	CAU	472	1.49	
	CUC	225	0.83			CAC	162	0.51
Ile	CUA	98	0.36	Gln	CAA	468	1.50	
	CUG	168	0.62			CAG	156	0.50
	AUU	801	1.20	Asn	AAU	916	1.43	
	AUC	584	0.88			AAC	365	0.57
	AUA	614	0.92	Lys	AAA	733	1.28	
	Met	AUG	424		1.00		AAG	414
Val		GUU	270	1.34	Asp	GAU	497	1.52
	GUC	166	0.82			GAC	156	0.48
	GUA	229	1.13	Glu	GAA	479	1.44	
	GUG	143	0.71			GAG	184	0.56
Ser	UCU	453	1.23	Cys	UGU	309	1.08	
	UCC	544	1.48			UGC	263	0.92
	UCA	99	0.27	TER <sup>*</sup>	UGA	253	0.75	
	UCG	273	0.74			UGG	319	1.00
Pro	CCU	201	0.96	Arg	CGU	169	0.81	
	CCC	219	1.05			CGC	115	0.55
	CCA	319	1.52			CGA	157	0.75
	CCG	98	0.47			CGG	104	0.50
Thr	ACU	334	1.11	Ser	AGU	421	1.14	
	ACC	430	1.42			AGC	422	1.14
	Ala	ACA	208	0.69	Arg	AGA	392	1.88
ACG		236	0.78			AGG	313	1.50
GCU		222	1.55	Gly	GGU	298	1.21	
GCC		136	0.95			GGC	183	0.74
GCA		128	0.90			GGA	330	1.34
GCG		86	0.60			GGG	173	0.70

\*: 终止密码子。

\*: Termination codon.

表 4 桑叶葡萄叶绿体基因组简单重复序列 (SSR) 信息  
Table 4 Number of SSRs identified in the cp genome of *Vitis ficifolia*

[illegible]

表 5 桑叶葡萄叶绿体基因组 SSR 信息  
Table 5 Distribution of SSR in the cp genome of *Vitis ficifolia*

编号 No.	SSR 类型 SSR type	SSR	大小 Size	起始 Start	终止 End	位置 Location
1	p3	(CAG)4	12	1 014	1 025	基因间区 psbA CDS
2	p1	(A)11	11	4 431	4 441	基因间区 IGS
3	p1	(C)11	11	5 425	5 435	内含子 intron(rps16)
4	c	(ATA)4gaatatctataat(TA)5	35	6 722	6 756	基因间区 IGS
5	p1	(A)11	11	10 307	10 317	基因间区 IGS
6	p1	(T)11	11	11 162	11 172	内含子 Intron(trnG-GCC)
7	p1	(A)12	12	13 330	13 341	基因间区 IGS
8	p1	(A)11	11	18 017	18 027	基因间区 IGS
9	p1	(T)10	10	19 603	19 612	内含子 Intron(rpoC2)
10	p1	(T)11	11	20 240	20 250	内含子 Intron(rpoC2)
11	p2	(AT)5	10	21 613	21 622	内含子 Intron(rpoC2)
12	p1	(T)10	10	27 969	27 978	内含子 Intron(rpoB)
13	c	(A)10gtgattccattataccattattatgattttacgtattccaatccgattggatatcag(A)11	79	28 899	28 977	基因间区 IGS
14	p1	(A)10	10	29 101	29 110	基因间区 IGS
15	p4	(TCTA)3	12	31 786	31 797	基因间区 IGS
16	p1	(T)12	12	32 054	32 065	基因间区 IGS
17	p1	(T)16	16	32 624	32 639	基因间区 IGS
18	p2	(AT)5	10	33 068	33 077	基因间区 IGS
19	p1	(A)10	10	34 894	34 903	基因间区 IGS
20	p1	(T)12	12	35 185	35 196	基因间区 IGS
21	p1	(T)11	11	38 636	38 646	基因间区 IGS
22	p4	(ATAG)3	12	38 981	38 992	基因间区 IGS
23	c	(A)10tggaa(T)14aatgttaataatgttaataataacaaatcaatttcaatatattatatatt ataatataataattattataaatatg(TAT)5ataattagaatgaataataatgcatttaataa atattaa(AT)8	185	39 542	39 726	基因间区 IGS
24	p2	(TA)5	10	39 879	39 888	基因间区 IGS
25	p3	(AAT)4	12	40 959	40 970	基因间区 IGS
26	p1	(A)10	10	47 995	48 004	intron(ycf3)
27	p1	(T)12	12	50 223	50 234	基因间区 IGS
28	c	(TATCC)3tctcttgcagaaattttttctcttcacaaagtctgtgg(TA)5tgatacgcgt acaaacgaacatctttgagcaaggaatctcatttgaatgattcacagtacatacattatcgtact gaaacttacaag(T)13	172	51 576	51 747	基因间区 IGS
29	c*	(T)10(TTTA)3*	19	52 527	52 545	基因间区 IGS
30	c	(T)10caagcgtggaaacccaggaccagaagtagtaggatttattctc(ATA)5	69	58 419	58 487	atpB CDS/IGS
31	c	(T)10ctcgagatc(T)14	33	62 956	62 988	基因间区 IGS
32	p1	(T)10	10	63 708	63 717	基因间区 IGS
33	p1	(A)13	13	65 495	65 507	基因间区 IGS
34	p1	(A)14	14	67 454	67 467	基因间区 IGS
35	p1	(A)10	10	67 962	67 971	基因间区 IGS
36	c	(AT)5tatactgtgtcatccagaaatcgagtgattcctcttcttctaactttgaaaggatcgata gatactatcgaggagcagatgttctg(AATC)3	111	68 111	68 221	基因间区 IGS
37	c	(T)13gaatc(T)11	29	69 698	69 726	基因间区 IGS
38	p1	(T)10	10	70 016	70 025	基因间区 IGS
39	p1	(T)11	11	71 475	71 485	基因间区 IGS
40	p1	(A)10	10	72 312	72 321	基因间区 IGS
41	c	(GAA)4atcc(T)10	26	73 153	73 178	基因间区 IGS
42	p1	(T)12	12	74 341	74 352	基因间区 IGS
43	p1	(A)10	10	75 441	75 450	内含子 Intron(clpP)
44	p1	(A)16	16	76 076	76 091	内含子 Intron(clpP)
45	p1	(A)14	14	82 553	82 566	基因间区 IGS
46	p1	(T)10	10	84 749	84 758	基因间区 IGS
47	p1	(T)15	15	89 033	89 047	基因间区 IGS
48	p4	(TTAT)3	12	104 015	104 026	基因间区 IGS
49	p1	(T)11	11	107 890	107 900	内含子 Intron(trnI-GAU)



续表 5

编号 No.	SSR 类型 SSR type	SSR	大小 Size	起始 Start	终止 End	位置 Location
50	p1	(T)11	11	113 092	113 102	基因间区 IGS
51	p5	(AGAAA)3	15	114 182	114 196	基因间区 IGS
52	p1	(T)10	10	116 140	116 149	ndhF CDS
53	p1	(A)12	12	117 699	117 710	基因间区 IGS
54	c	(TTA)4(T)10	22	118 689	118 710	基因间区 IGS
55	p1	(A)11	11	119 180	119 190	基因间区 IGS
56	c	(TA)6aaccttggtgaagccgggaaaaccagaaccatttgaatcactgcactacttgattcaa atgggttctcacaacgccaaacgtattcaattca(T)12acttaactcaagaggag(A)10	144	121 364	121 507	基因间区 IGS
57	p1	(A)15	15	123 139	123 153	基因间区 IGS
58	c	(TTAT)3cttgatatattgtaggatatagatagatcaaaatctaccggaaggtcccggaatta ggccaatggaattctgtctgtataatct(ATA)4	109	126 401	126 509	内含子 Intron(ndhA)
59	c	(AGA)3ataagaaaagattttttctattgtaattacattcttc(T)11	61	126 708	126 768	内含子 Intron(ndhA)
60	p4	(AATC)3	12	127 147	127 158	内含子 Intron(ndhA)
61	p3	(ATA)4	12	129 608	129 619	基因间区 IGS
62	p5	(TTTC)3	15	135 526	135 540	基因间区 IGS
63	p1	(A)11	11	136 621	136 631	基因间区 IGS
64	p1	(A)11	11	141 823	141 833	内含子 Intron(trnI-GAU)
65	p4	(AATA)3	12	145 696	145 707	基因间区 IGS
66	p1	(A)15	15	160 676	160 690	基因间区 IGS

表 6 桑叶葡萄叶绿体基因组串联重复信息

Table 6 Distribution of tandem repeats in the CP genome of *Vitis ficifolia*

编号 No.	总长度 Total Lenth	一致大小 Consensus size	拷贝数量 Copy number	起始 Start	终止 End	位置 Location
1	39	19	2.0	296	334	基因间区 IGS
2	30	15	2.0	1931	1960	内含子 Intron(trnK-UUU)
3	37	11	3.4	6645	6681	基因间区 IGS
4	86	35	2.5	6667	6752	基因间区 IGS
5	87	27	3.1	6745	6831	基因间区 IGS
6	33	17	1.9	6861	6893	基因间区 IGS
7	30	14	2.1	7595	7624	基因间区 IGS
8	25	12	2.1	7751	7775	基因间区 IGS
9	29	14	2.1	14685	14713	基因间区 IGS
10	34	17	2.0	14990	15023	基因间区 IGS
11	36	16	2.3	15800	15835	基因间区 IGS
12	40	20	2.0	28995	29034	基因间区 IGS
13	26	12	2.2	30788	30813	基因间区 IGS
14	25	11	2.3	31795	31819	基因间区 IGS
15	26	13	2.0	34063	34088	基因间区 IGS
16	50	23	2.2	34087	34136	基因间区 IGS
17	36	16	2.3	34664	34699	基因间区 IGS
18	63	20	3.3	39609	39671	基因间区 IGS
19	50	19	2.8	39724	39773	基因间区 IGS
20	38	12	3.3	39858	39895	基因间区 IGS
21	31	15	2.1	45866	45896	基因间区 IGS
22	44	14	3.0	54677	54720	基因间区 IGS
23	32	16	2.0	55134	55165	基因间区 IGS
24	48	24	2.0	61709	61756	accD CDS
25	31	14	2.2	63331	63361	基因间区 IGS
26	27	13	2.1	72577	72603	基因间区 IGS
27	41	21	2.0	72732	72772	rps18 CDS
28	46	23	2.0	79752	79797	内含子 Intron(petB)
29	89	21	4.1	93795	93883	ycf2 CDS
30	78	18	4.3	96255	96332	ycf2 CDS
31	37	17	2.2	129590	129626	基因间区 IGS
32	78	18	4.3	153391	153468	ycf2 CDS
33	89	21	4.1	155840	155928	ycf2 CDS

## 2.4 IR 边界的收缩与扩张

IR 区域的收缩和扩张是造成叶绿体基因组长度变化的重要原因 (Li et al., 2013)。比较 5 种葡萄属植物叶绿体基因组 IR 区与 LSC 和 SSC 区的边界, 并以蛇葡萄和拟南芥的叶绿体基因组作为参考。整体来说, 葡萄属植物具有相似的基因结构和基因顺序, 其基因组长度的差异主要发生在 LSC 区域 (图 2)。IR 边界比较表明, 5 种葡萄属植物的 LSC/IRA 边界均位于 *rps19* 基因内, 且 *rps19* 基因有一部分位于 IRA 区域, 长度范围在 39 ~ 49 bp。蛇葡萄的 *rps19* 基因有 46 bp 位于 IRA 区域, 与葡萄属植物类似。而拟南芥的 *rps19* 基因有 113 bp 位于 IRA 区域, 与其他植物明显不同。蛇葡萄和拟南芥的 IRA/SSC 边界位于 *ycf1* 假基因内, 而 5 种葡萄属植物的 IRA/SSC 边界均位于 *ycf1* 假基因和 *ndhF* 基因之间, 表明了 5 种葡萄属植物的 IR 区域在 IRA/SSC 边界发生了明显的收缩。桑叶葡萄和山葡萄的 *ycf1* 假基因距 IRA/SSC 边界 0 bp, 而其他葡萄属植物的 *ycf1* 假基因距 IRA/SSC 边界的长度范围在 86 ~ 106 bp 之间, 表明了中国野生葡萄与其他葡萄属植物明显的不同。此外, 所有植物的 SSC/IRB 边界均位于 *ycf1* 基因中, 从而导致这些植物在 IRA 区域均产生了一个 *ycf1* 假基因。

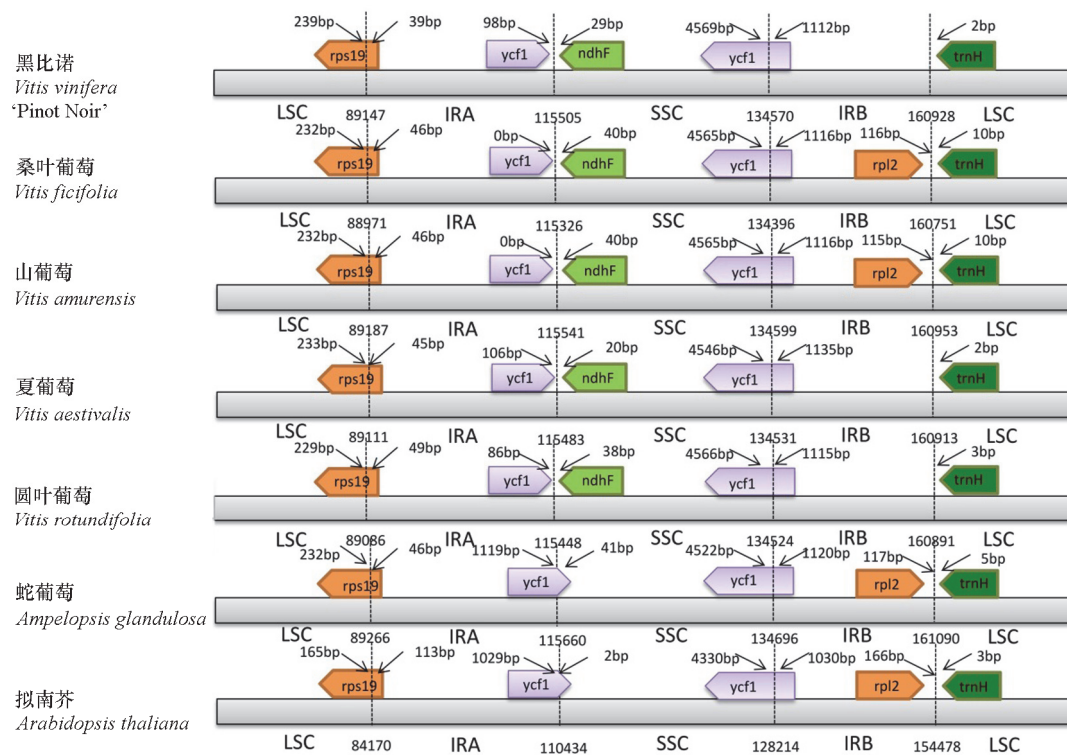


图 2 不同物种叶绿体基因组反向重复区 (IR)、大单拷贝区 (LSC) 和小单拷贝区 (SSC) 边界比对

Fig. 2 Comparison of the borders of the LSC, SSC, and IR regions of different species

## 2.5 系统进化分析

构建了桑叶葡萄等 32 种植物的系统进化树。结果 (图 3) 表明, 桑叶葡萄与山葡萄关系最近, 且与欧亚种 ‘黑比诺’ 和北美种夏葡萄组成了一个单系分支; 所有的真葡萄亚属植物聚在了同一支, 圆叶葡萄单独聚为一支, 即葡萄属可以分为真葡萄亚属和圆叶葡萄亚属两个分支 (Zecca et al.,

2012)。此外, 虎耳草目 (Saxifragales) 的所有物种与鼠李目的所有物种互为姐妹分支, 且达到 99% 的 BS 值, 这与 Raman 和 Park (2016) 的结论一致, 即虎耳草目与鼠李目的蛇葡萄属和葡萄属具有较近的亲缘关系。

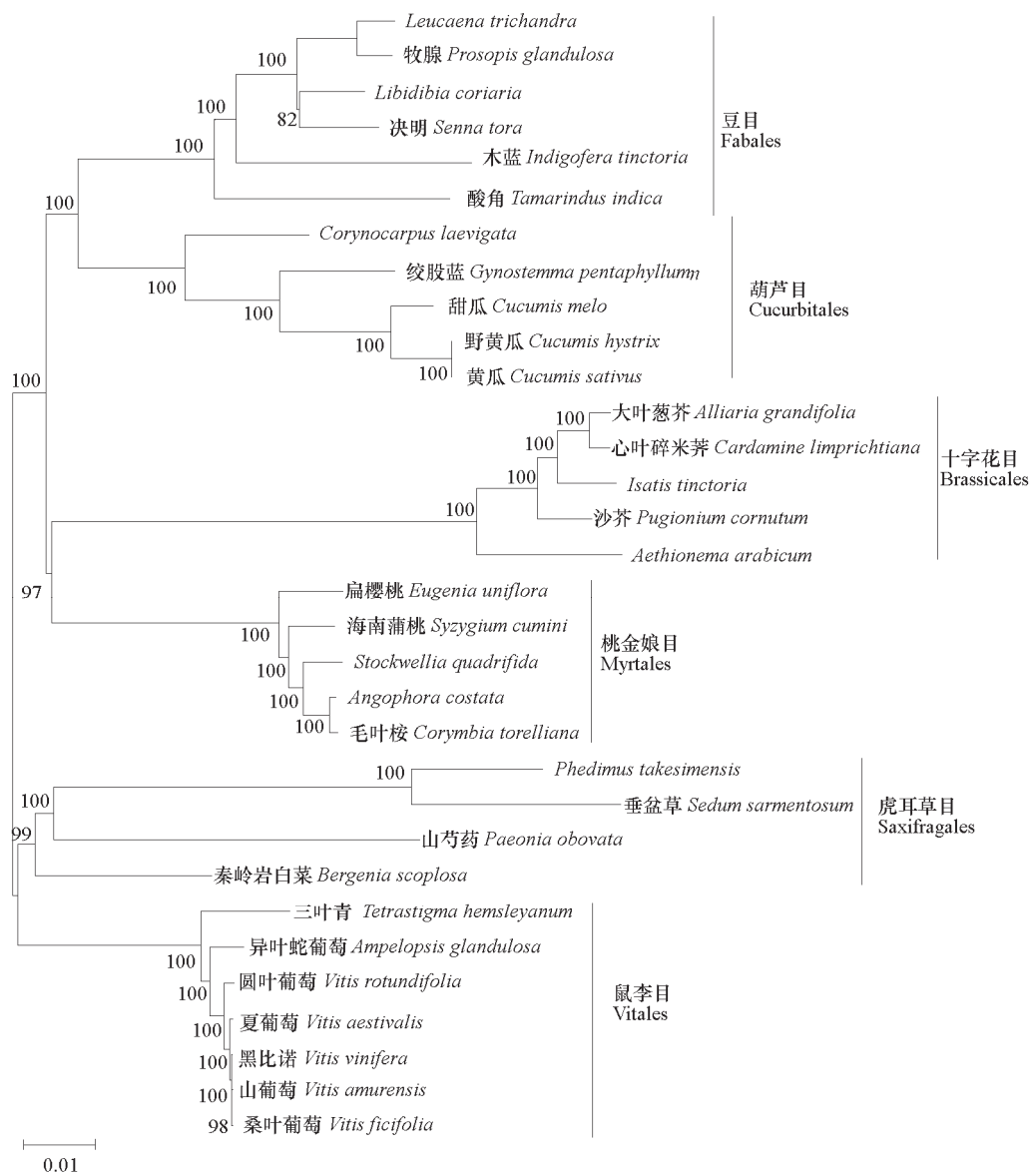


图 3 基于叶绿体基因组共线性模块用邻接法构建 32 个物种的系统进化树  
支上的数字为 1 000 次自展支持率 (%)。

Fig. 3 Phylogenetic tree of 32 species based on chloroplast genome collinear blocks using Neighbor-joining method (NJ)  
The bootstrap value was shown above the branch (%) .

### 3 讨论

传统获得植物叶绿体基因组一般有两种方法。一种是先从植物组织中分离叶绿体,然后再提取叶绿体 DNA,通过一代或者二代测序技术获得植物的叶绿体基因组,但中国野生葡萄叶片色素、单宁等物质较多,分离叶绿体较为困难,高质量的叶绿体 DNA 难以获得。另一种是提取植物的全基因组 DNA,然后利用叶绿体基因组的保守区域设计引物,进行一代测序,最终拼接得到植物的叶绿体基因组,此方法的弊端在于工作量大,且很难获得完整的叶绿体基因组序列。随着新一代测序技术(NGS)的发展,以及大量生物信息学软件的广泛使用,可以提取植物的全基因组 DNA,进行高通量测序,抽提出样本的叶绿体 reads 进行组装,得到植物叶绿体基因组,这种方法不用进行叶绿体 DNA 的分离,降低了劳动强度,且提高了试验的成功率。Illumina 测序平台通量高,具备近缘物种的叶绿体序列的前提下,获取叶绿体基因组具有独特的优势。因此,本研究中采用了 Illumina HiSeq 测序平台对桑叶葡萄进行了全基因组重测序,组装得到了桑叶葡萄叶绿体基因组。

为了解葡萄属植物叶绿体基因组间的差异,本研究中从 GenBank 下载了已发表的山葡萄、‘黑比诺’、夏葡萄、圆叶葡萄、蛇葡萄、拟南芥的叶绿体进行比较。结果表明,桑叶葡萄与其近缘种山葡萄、‘黑比诺’、夏葡萄、以及圆叶葡萄的叶绿体基因组长度差别不大,长度在 160 751 ~ 160 953 bp 之间,基因种类和数量大致相同,这证明了叶绿体基因组高度保守的特征。不同葡萄属植物叶绿体基因组长度差异主要发生在 LSC 区域,可能是由基因间隔区的插入和缺失造成,与大部分被子植物叶绿体一致。

本研究中得到了桑叶葡萄的叶绿体基因组,对其密码子偏好性、重复序列和 SSRs 特征分析,为中国野生葡萄叶绿体基因组的研究提供了数据基础。系统进化分析表明桑叶葡萄与山葡萄、‘黑比诺’、夏葡萄以及圆叶葡萄具有较近的亲缘关系。葡萄属植物种类繁多,属内种间杂交容易,中间型、过渡型相当普遍,分类工作难度较大,利用叶绿体基因组可为解决葡萄属内植物的分类问题提供参考,但是目前已发表的葡萄属植物叶绿体基因组数量还非常有限(<https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/vitis>),相关研究也仅停留在不同种叶绿体基因组的比较分析上,因此,还需要得到更多葡萄属植物的叶绿体基因组才能更好地解决葡萄属植物的系统发育问题。

### 4 结论

本研究中完成了桑叶葡萄全基因组重测序及其叶绿体基因组的组装、注释,并对桑叶葡萄叶绿体基因组的结构、基因组成、GC 含量、重复序列、密码子的偏好性进行了分析。结果表明,桑叶葡萄叶绿体基因组全长 160 951 bp,其基因结构和基因组成与大部分被子植物类似,体现了叶绿体基因组较为保守的特征;对桑叶葡萄及其近缘种的比较分析表明,葡萄属植物具有相似的基因结构和基因顺序,LSC 区是它们之间产生差异的主要区域。系统进化分析表明,葡萄属可分为真葡萄亚属和圆叶葡萄亚属两个分支,与传统分类一致。桑叶葡萄与同为中国野生葡萄的山葡萄具有较近的亲缘关系。本研究为利用叶绿体基因组解决葡萄属乃至葡萄科的系统发育问题提供了数据基础。

### References

- Asaf S, Waqas M, Khan A L, Khan M A, Kang S M, Imran Q M, Shahzad R, Bilal S, Yun B W, Lee I J. 2017. The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. *Frontiers in Plant Science*, 8: 304.

- Bankevich A, Nurk S, Antipov D, Gurevich A A, Dvorkin M, Kulikov A S, Lesin V M, Nikolenko S I, Pham S, Pribelski A D, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler G, Alekseyev M A, Pevzner P A. 2012. SPAdes: a new genome assembly algorithm and its application to single-cell sequencing. *Journal of Computational Biology*, 19 (5): 455 – 477.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27 (2): 573 – 580.
- Burke S V, Wysocki W P, Zuloaga F O, Craine J M, Pires J C, Edger P P, Mayfield- Jones D, Clark L G, Kelchner S A, Duvall M R. 2016. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). *BMC Plant Biology*, 16 (1): 140.
- Cavalier-Smith T. 2002. Chloroplast evolution: secondary symbiogenesis and multiple losses. *Current Biology*, 12 (2): 62 – 64.
- Chen H, Shao J, Zhang H, Jiang M, Huang L, Zhang Z, Yang D, Molly H, Mostafa R, Luo X, Sun B, Wu W, Liu C. 2018. Sequencing and analysis of *Strobilanthes cusia* (Nees) kuntze chloroplast genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in angiosperm. *Frontiers in Plant Science*, DOI: 10.3389/fpls.2018.00324.
- Cronn R, Liston A, Parks M, Gernandt D S, Shen R, Mockler T. 2008. Multiplex sequencing of plant chloroplast genomes using solexa sequencing-by-synthesis technology. *Nucleic Acids Research*, 36 (19): e122.
- Fajardo D, Senalik D, Ames M, Zhu H, Steffan S A, Harbut R, Polashock J, Vorsa N, Gillespie E, Kron K, Zalapa J E. 2013. Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. *Tree Genetics & Genomes*, 9 (2): 489 – 498.
- Fu J, Liu H, Hu J, Liang Y, Liang J, Wuyun T, Tan X. 2016. Five complete chloroplast genome sequences from *Diospyros*: genome organization and comparative analysis. *PLoS ONE*, 11 (7): e0159566.
- Jansen R K, Raubeson L A, Boore J L, Depamphilis C W, Chumley T W, Haberle R C, Wyman S K, Alverson A J, Peery R, Herman S J, Fourcade H M, Kuehl J V, McNeal J R, Leebens-Mack J, Cui L. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology*, 395 (6): 348.
- Jansen R K, Kaitanis C, Saski C, Lee S B, Tomkins J, Alverson A J, Daniell H. 2006. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. *BMC Evolutionary Biology*, 6 (1): 32.
- Keeling P J. 2004. Diversity and evolutionary history of plastids and their hosts. *American Journal of Botany*, 91 (10): 1481.
- Kuang D Y, Wu H, Wang Y L, Gao L M, Zhang S Z, Lu L. 2011. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for dna barcoding and population genetics. *Genome*, 54 (8): 663 – 673.
- Leseberg C H, Duvall M R. 2009. The complete chloroplast genome of *Coix lacrymajobi* and a comparative molecular evolutionary analysis of plastomes in cereals. *Journal of Molecular Evolution*, 69 (4): 311 – 318.
- Li Chao-luan. 1998. *Flora of China · Vitaceae*. Beijing: Science Press: 140 – 198. (in Chinese)
- 李朝銮. 1998. *中国植物志 · 葡萄科*. 北京: 科学出版社: 140 – 198.
- Li R, Ma P, Wen J, Yi T. 2013. Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. *PLoS ONE*, 8 (10): e78568.
- Moore M J, Dhingra A, Soltis P S, Shaw R, Farmerie W G, Foltz K M, Soltis D E. 2006. Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology*, 6 (1): 17.
- Nashima K, Terakami S, Nishitani C, Kunihisa M, Shoda M, Takeuchi M, Urasaki N, Tarora K, Yamamoto T, Katayama H. 2015. Complete chloroplast genome sequence of pineapple (*Ananas comosus*). *Tree Genetics & Genomes*, 11 (3): 60.
- Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar S S, Tan X, Wan F, Weining S. 2012. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS ONE*, 7 (5): e36869.
- Peng S, Yang G, Liu C, Zhai M. 2015. The complete chloroplast genome of the *Juglans regia* (Juglandales: Julandaceae). *Mitochondrial DNA*, 28 (3): 407 – 408.
- Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, Yao H, Sun C, Li X, Li C, Liu J, Xu H, Chen S. 2013. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE*, 8 (2): e57607.
- Raman G, Park S J. 2016. The complete chloroplast genome sequence of ampelopsis: gene organization, comparative analysis, and phylogenetic

- relationships to other angiosperms. *Frontiers in Plant Science*, 7: 341.
- Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthapaisanwong P, Yoocha T, Jomchai N, Tragoonrun S. 2009. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. *DNA Research an International Journal for Rapid Publication of Reports on Genes & Genomes*, 17 (1): 11 – 22.
- Su H J, Hogenhout S A, Alsadi A M, Kuo C H. 2014. Complete chloroplast genome sequence of Omani Lime (*Citrus aurantiifolia*) and comparative analysis within the rosids. *PLoS ONE*, 9 (11): e113049.
- Wan Y, Schwaninger H, Li D, Simon C J, Wang Y. 2008. A review of taxonomic research on Chinese wild grapes. *Vitis Journal of Grapevine Research*, 47 (2): 81 – 88.
- Wei W, Zheng Y, Chen L, Wei, Yan Z, Yang R. 2005. PCR-RFLP analysis of cpDNA and mtDNA in the genus *Houttuynia*, in some areas of China. *Hereditas*, 142: 24 – 32.
- Xie H, Jiao J, Fan X, Zhang Y, Jiang J, Liu C. 2016. The complete chloroplast genome sequence of Chinese wild grape *Vitis amurensis* (Vitaceae: *Vitis* L.). *Conservation Genetics Resources*, 9 (1): 1 – 4.
- Xie Hai-kun, Jiao Jian, Fan Xiu-cai, Zhang Ying, Jiang Jian-fu, Sun Hai-sheng, Liu Chong-huai. 2017. Assembling and characteristic analysis of the complete chloroplast genome of *Vitis vinifera* cv. Cabernet Sauvignon from high-throughput sequencing data. *Scientia Agricultura Sinica*, 50 (9): 1655 – 1665. (in Chinese)
- 谢海坤, 焦健, 樊秀彩, 张颖, 姜建福, 孙海生, 刘崇怀. 2017. 基于高通量测序组装‘赤霞珠’叶绿体基因组及其特征分析. *中国农业科学*, 50 (9): 1655 – 1665.
- Yang Y, Tao Z, Dong D, Jia Y, Li F, Zhao G. 2016. Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Frontiers in Plant Science*, 7: 959.
- Zecca G, Abbott J R, Sun W B, Spada A, Sala F, Grassi F. 2012. The timing and the mode of evolution of wild grapes (*Vitis*). *Molecular Phylogenetics & Evolution*, 62 (2): 736 – 747.
- Zhang Xu-tong. 2012. Phylogenetic analysis of Chinese wild grapes germplasm resources [M. D. Dissertation]. Yangling: Northwest A & F University. (in Chinese)
- 张旭彤. 2012. 中国野生葡萄种质资源的亲缘关系研究[硕士论文]. 杨凌: 西北农林科技大学.
- Zhang Yong -hui, Liu Chong-huai, Fan Xiu-cai, Zhang Ying, Sun Hai-sheng, Peng Bin, Jiang Jian-fu. 2011. Application of ISSR markers in taxonomy of Chinese wild grape. *Journal of Fruit Science*, 28 (3): 406 – 412. (in Chinese)
- 张永辉, 刘崇怀, 樊秀彩, 张颖, 孙海生, 彭斌, 姜建福. 2011. ISSR 标记在中国野生葡萄分类中的应用. *果树学报*, 28 (3): 406 – 412.