植物学报 *Chinese Bulletin of Botany* 2025, **60** (4): 573–585, www.chinbullbotany.com
doi: 10.11983/CBB24146 cstr: 32102.14.CBB24146

·研究论文·

黑涩楠叶绿体全基因组的结构和比较分析及系统进化推断

王传永[1], 庄典[2], 宋正达[1], 翟恒华[1], 李乃伟[1*], 张凡[1*]

[1]江苏省中国科学院植物研究所(南京中山植物园), 江苏省植物资源研究与利用重点实验室, 南京 210014
[2]南京中医药大学, 南京 210023

摘要 黑涩楠(*Aronia melanocarpa*)因其观赏价值和经济价值而闻名, 但其与其它蔷薇科植物的系统进化关系仍不明确。该研究对黑涩楠叶绿体(cp)基因组进行测序, 并与13个蔷薇科物种的叶绿体基因组进行比较分析。结果表明, 黑涩楠的cp基因组大小为159 772 bp, 呈典型的四分结构; 其中大单拷贝区(LSC)长度为87 810 bp, 小单拷贝区(SSC)长度为19 200 bp, 中间含有2个26 381 bp的反向重复区(IRa和IRb)。共注释到132个基因, 包括87个蛋白质编码基因、37个tRNA和8个rRNA。还检测到76个简单重复序列(SSR)和50个长重复序列。系统进化分析表明, 黑涩楠与红涩楠(*A. arbutifolia*)的亲缘关系最近, 与榅桲(*Cydonia oblonga*)是姊妹支系。该研究提供的基因组信息将为后续的系统进化和种群遗传分析以及分子育种提供理论支持。

关键词 黑涩楠, 叶绿体基因组, 结构变异, 系统进化关系

王传永, 庄典, 宋正达, 翟恒华, 李乃伟, 张凡 (2025). 黑涩楠叶绿体全基因组的结构和比较分析及系统进化推断. 植物学报 **60**, 573–585.

黑涩楠(*Aronia melanocarpa*), 又名野樱莓、不老莓, 隶属蔷薇科涩楠属(*Aronia*) (Kulling and Rawel, 2008; Szopa et al., 2017)。黑涩楠原产于美国东北部和五大湖区(Hardin, 1973; Seidemann, 1993), 20世纪90年代被引入中国作为经济和观赏植物。近年来, 黑涩楠受到越来越多的关注, 已广泛用于制作果酱、果汁、葡萄酒和天然食品着色剂(汪娣等, 2023; 杨柳等, 2024)。此外, 黑涩楠作为一种草药, 还用于治疗高血压和动脉粥样硬化(Domarew et al., 2002; Lim et al., 2014; Shukla and Mehta, 2015)。

除黑涩楠外, 涩楠属还包括红涩楠(*A. arbutifolia*)和紫涩楠(*A. prunifolia*), 这两种植物都生长在北美的野外; 还有1种栽培类群涩石楠(*A. mitschurinii*), 原产于欧洲(Connolly, 2014; Shipunov et al., 2019)。然而, 涩楠属的属界和种间关系尚不明确(Brand, 2010; Shipunov et al., 2019)。此外, 涩楠属的分类历史也很复杂, 该属物种曾被归入许多属, 如欧楂属(*Mespilus*)、梨属(*Pyrus*)、花楸属(*Sorbus*)和石楠属(*Photinia*) (Brand, 2010)。Robertson等(1991)根据涩楠属与石楠属的花和果实相似的结论, 将其归入石楠属。然而, Campbell等(2007)研究表明, 涩楠属与石楠属在分子水平上几乎没有亲缘关系。此外, Guo等(2011)指出, 涩楠属既不是石楠属的姐妹群, 也不是落叶石楠属(*Pourthiaea*)的姐妹群, 他们建议将该属划分为独立的属。Li等(2012)根据核糖体DNA内部转录间隔区的分析, 发现涩楠属是白花楸属(*Aria*)和水榆属(*Micromeles*)的姐妹群。到目前为止, 涩楠属的系统进化位置仍不确定。

叶绿体(chloroplast, cp)是绿色植物的核心细胞器, 在光合作用和碳固定中发挥重要作用(Palmer, 1987)。在高等植物中, cp基因组的大小一般介于120–180 kb之间, 通常由2段反向重复序列(inverted repeat, IR)组成, 2个IR之间由1个大单拷贝区(large single copy region, LSC)和1个小单拷贝区(small single copy region, SSC)隔开(Ravi et al., 2008)。相对于核基因组, 叶绿体基因组进化速率较慢, 基因含量及结构高度保守(Daniell et al., 2016; Liu et al., 2016)。然而, 通过比较分析不同物种间cp基因组, 发

收稿日期: 2024-09-24; 接受日期: 2025-02-09
基金项目: 新优植物品种引进及推广示范(No.FY2024013)
* 通讯作者。E-mail: linaiwei@jib.ac.cn; mumizhongfeng@126.com

现在进化过程中种内和种间发生了重排，如IR收缩和扩展(褚振州等, 2023)。基于这些特点, cp基因组已广泛用于物种鉴定及系统进化分析(Liu et al., 2016; 赵月梅等, 2019)。在本研究中，我们对黑涩楠完整cp基因组进行测序和特征分析，并将其序列特征与蔷薇科13个物种的cp基因组进行比较，旨在揭示黑涩楠完整cp基因组的特征模式，并确定黑涩楠与其它蔷薇科植物cp基因组的系统进化关系。

1 材料与方法

1.1 植物材料和DNA提取

采集江苏省南京中山植物园温室栽培的黑涩楠(*Aronia melanocarpa* (Michx.) Ell.)新鲜叶片。采用CTAB法提取cp基因组DNA。用NanoDrop分光光度计(Thermo Scientific, Waltham, MA, 美国)对DNA质量进行评估。分别用1%凝胶电泳和Qubit荧光仪(Life Technologies, Darmstadt, 德国)检测其完整性和浓度。

1.2 叶绿体基因组测序、组装和注释

在Illumina HiSeq4000平台(广州佰数生物)基于边合成边测序进行DNA文库测序，测序读长为PE250，得到约50 Mb的高质量、干净的双端读数。用NOVOPlasty (Dierckxsens et al., 2017)对序列进行组装，以*Aronia arbutifolia*的叶绿体基因组全长为种子序列和参考基因组, K-mer分别设置为33、49、59和69，获得相同组装结果。为确保组装结果准确，进一步利用GetOrganelle v1.7.7.1 (默认参数)进行从头组装。结果2个软件的组装结果一致。使用CPGAVAS软件(Shi et al., 2019)注释基因组，并通过DOGMA (http://dogma.ccbb.utexas.edu)和BLAST (Wyman et al., 2004)检验注释结果。通过tRNA scan-SE验证tRNA注释结果。使用OGDRAWv1.2程序(http://ogdraw.mpimp-golm.mpg.de) (Lohse et al., 2007)绘制整个基因组的圈图。黑涩楠的cp基因组序列已上传至GenBank (MT527725)并在ScienceDB数据库中备份(数据集doi: 10.57760/sciencedb.26952)。

1.3 重复序列分析和密码子使用分析

使用MicroSAtellite (MISA, http://pgrc.ipk-gatersleben.de/misa/misa.html)软件(Mudunuri and Nagarajaram, 2007)鉴定简单重复序列(simple sequence repeat, SSR)，设置参数definition (unit_size, min_repeats): 1-10 2-6 3-5 4-5 5-5 6-5, interruptions (max_difference_for_2_SSRs): 100。重复序列(包括正向重复、反向重复和互补重复序列)通过REPuter (https://bibiserv.Cebitec.uni-bielefeld.de/reputer)在线软件(Kurtz et al., 2001)进行分析，设置参数minimum size=30, Hamming Distance=3。用CodonW1.4.2程序分析蛋白编码基因的相对同义密码子使用率(relative synonymous codon usage, RSCU)。

为探明黑涩楠的全基因组进化动态，用mVISTA (http://genome.lbl.gov/vista/mvista) (Mayor et al., 2000)程序在Shuffle-LAGAN模式下(Frazer et al., 2004)分析14个完整cp基因组之间的遗传差异，包括光叶石楠(*Photinia glabra*) (MK920277)、欧亚花楸(*Sorbus commixta* Hedl.) (MK920288)、河北梨(*Pyrus hopeiensis* T. T. Yu) (MF521826)、印缅红果树(*Stranvaesia nussia* (Buch.-Ham. ex D. Don) Decne.) (MK920284)、弗洛伦萨海棠(*Malus florentina* (Zuccagni) Stapf) (NC_035625)、中华落叶石楠(*Pourthiaea arguta* (Lindl.) Decne.) (NC_045413)、驱疝木(*Torminalis clusii*) (NC_045423)、红涩楠(*A. arbutifolia* (L.) Pers.) (NC_045391)、榅桲(*Cydonia oblonga* Mill.) (NC_045415)、木瓜海棠(*Chaenomeles cathayensis* (Hemsl.) C. K. Schneid.) (NC_045392)、华西小石积(*Osteomeles schwerinae* C. K. Schneid.) (NC_045420)、金绒梨(*Phippsiomeles mexicana* (Baill.) B. B. Liu & J. Wen) (NC_045422)和欧楂(*Mespilus germanica* L.) (MK920295)。利用IRscope软件对蔷薇科14个物种cp基因组中的单拷贝区(LSC和SSC)与IR区的边界进行可视化。

1.4 分子进化和系统进化分析

从美国国家生物技术信息中心(NCBI)数据库下载64个物种的cp基因组，含60个蔷薇科成员和4个外类群，包括暹罗桑(*Morus indica* L.)、裂叶蒙桑(*M. mongolica* (Bureau) C. K. Schneid.)、枣(*Ziziphus jujuba* Mill.)和滇刺枣(*Z. Mauritiana* Lam.)。根据64个物种的全部cp基因组构建系统进化树。使用MAFFT v.5软件对64个物种的cp基因组进行比对(Katoh et al., 2019)。使用TrimAL v1.4.1软件(默认参数)校正比对后的序列。分

别使用iQTree和FastTree软件构建最大似然(ML)树(Price et al., 2010)。其中, iQTree最优模型为TVM+F+R4, bootstrap设为1 000; FastTree最优模型选用GTR+G, 进行1 000次SH-like local support检测。

2 结果与分析

2.1 黑涩楠叶绿体基因组的特征

完整的黑涩楠cp基因组全长159 772 bp, 呈典型的四分结构, 这在多数植物中均有发现(Xu et al., 2017)。其包含1个长87 810 bp的LSC区和1个长19 200 bp的SSC区, 这2个区被1对长度为26 381 bp的反向重复区(IRa和IRb)隔开(图1)。整个cp基因组的总GC含量为36.6%, IR区的GC含量(42.7%)高于LSC区(34.3%)和SSC区(30.4%) (表1), 可能是由于该区域中重复的rRNA和tRNA的GC含量相对较高所致。这与前人的研究结果一致(He et al., 2016; Li et al., 2019)。

在黑涩楠cp基因组中, 共预测到132个基因, 包括87个蛋白编码基因、37个tRNA和8个rRNA。其中, 110个基因是单拷贝, 22个基因位于IR区(表1, 表2)。在22个重复基因中, 10个为蛋白编码基因, 8个编码tRNA, 4个编码rRNA (表2)。此外, 还发现45个参与光合作用的基因, 包括6个ATP合成酶亚基编码基因、12个NADH脱氢酶亚基编码基因、6个细胞色素*b/f*复合体亚基编码基因、5个光系统I亚基编码基因、15个光系统II亚基编码基因和1个编码Rubisco大亚基的基因(表2)。

此外, 黑涩楠cp基因组中有18个基因有内含子, 其中12个基因是蛋白编码基因, 6个基因编码tRNA。其中, 15个基因含有2个外显子, 3个基因(*ycf3*、*clpP*和*rps12*)含有3个外显子(表2)。值得注意的是, *rps12*是一个反式剪接基因, 其外显子1位于LSC区, 而外显子2和外显子3则位于IR区的2个位置, 这在许多物种中均很常见(Hildebrand et al., 1988)。此外, *trnK-UUU*含有最长的内含子(长度为2 490 bp), *matK*基因位于该内含子中。

2.2 重复序列分析

重复序列在基因组重排和序列分化中发挥重要作用。长重复序列经常出现在基因组序列中, 是重要的分析工具(Benson, 1999)。本研究发现, 在黑涩楠cp基因组中含有50个长重复序列, 其中35个为正向重复序列, 14个为反向重复序列, 还有1个复合长重复序列。大多数重复序列分布在LSC区(31个), 其次是IRs区(16个), 仅有6个重复序列分布在SSC区。此外, 大多数重复序列分布在基因间区(34个, 占68%), 少数分布在*ycf1*、*ycf2*、*ycf3*、*ndhA*和*rpl16*基因中(附表1)。

简单重复序列(SSR)通常由1–6个核苷酸重复单元组成(Liu et al., 2018)。SSR具有多种多态性, 因此可作为确定植物分类地位和系统进化关系的有效分子标记(Fu et al., 2016)。本研究在黑涩楠cp基因组中检测到76个SSRs。大部分已鉴定的SSR为单核苷酸SSR (57个, 占85.07%), 其中单核苷酸T的频率最高, 其次是单核苷酸A和C (表3)。然而, 除3个二核苷酸(4.47%)和7个复合SSR (0.44%)外, 在黑涩楠cp基因组中未发现其它多核苷酸SSR (附表2)。SSR主要分布在LSC区(60个, 78.94%), 其次是SSC区(12个, 15.79%)和IR区(4个, 5.97%)。基因间空间包含了大部分SSR (58个, 76.32%), 而编码序列中仅检测到18个SSRs, 分别是*matK*、*rps16*、*trnG-GCC*、*atpF*、*rpoC2*、*rpoB*、*atpB*、*clpP*、*rpl16*、*ndhA*和*ycf1* (附表2)。此外, *rps16*、*trnG-GCC*、*clpP*、*rpl16*、*ndhA*和*ycf1*比其它基因含有更多SSR位点。这些SSR可作为分子标记用于分析涩楠属种间的遗传变异。

2.3 黑涩楠叶绿体基因组中的密码子使用情况

密码子退化, 即同义密码子编码单一氨基酸的现象, 具有重要的生物学意义, 可使生物有效避免有害突变(Morton, 2003)。然而, 退化密码子家族中的同义密码子在植物进化过程中会出现使用偏差(Liu and Xue, 2005)。相对同义密码子使用度(RSCU)是检测密码子偏好程度的有效指标(Sharp and Li, 1987)。因此, 我们分析了黑涩楠cp基因组中的密码子使用频率和RSCU。编码蛋白质的基因共有79 848个碱基和26 616个密码子。根据编码21种氨基酸的64个可能密码子, 该cp基因组中出现频率最高和最低的氨基酸分别是亮氨酸(2 800, 10.52%)以及半胱氨酸(303, 1.14%)。这种现象在其它被子植物质体中也很常见(Chiapella et al., 2019)。通常情况下, 同义密码子偏好根据RSCU值可分为4类: 无偏好(RSCU≤1.0)、低偏好(1.0<RSCU<1.2)、中偏好(1.2≤RSCU≤1.3)和高

偏好(RSCU>1.3) (Zhao et al., 2010)。在本研究中，有30种密码子的RSCU值大于1.0，其中2种为低偏好(1.0<RSCU<1.2)，7种为中偏好(1.2≤RSCU≤1.3)，21种为高偏好(RSCU>1.3)；除密码子UUG外，其余密码子大多以A或U结尾。相比之下，34个无偏好(RSCU≤1.0)密码子大多以C或G结尾(表4)。

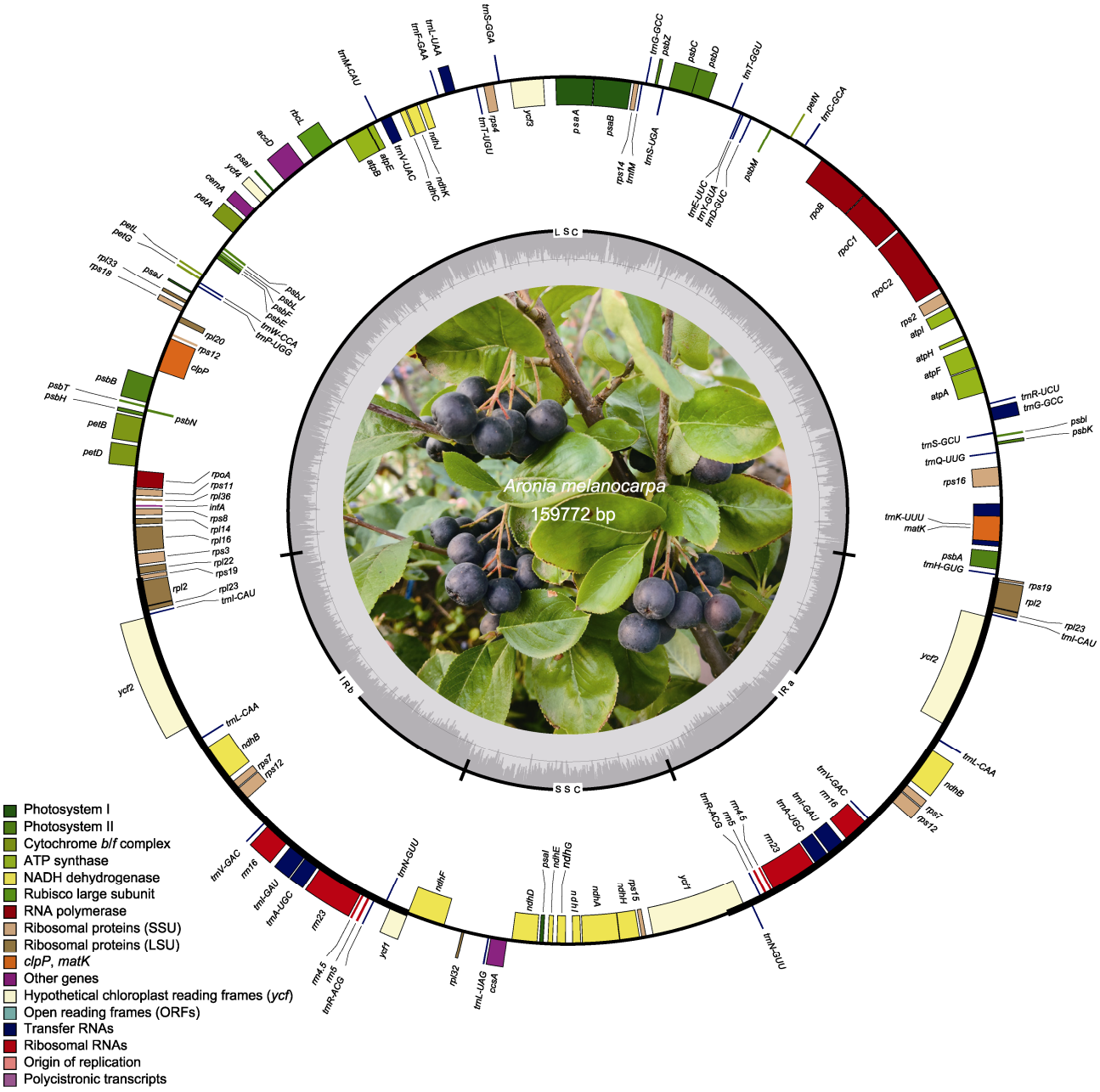


图1 黑涩楠叶绿体基因组图谱

圆圈内、外的基因分别按顺时针和逆时针方向转录。不同功能基因组用不同的颜色标记。内圈的GC和AT含量分别用深灰和浅灰色表示。

Figure 1 Map of the chloroplast genome of *Aronia melanocarpa*

Genes inside and outside of the circle are transcribed in the clockwise and counterclockwise directions respectively. Different functional gene groups are color-coded accordingly. GC and AT content are represented on the inner circle by darker and lighter gray, respectively.

表1 黑涩楠叶绿体基因组特征

Table 1 The *Aronia melanocarpa* chloroplast genome features

Genome features	*A. melanocarpa*	Genome features	*A. melanocarpa*
Genome size (bp)/GC content (%)	159772/36.6	Number of unique genes	110
LSC size (bp)/GC content (%)	87810/34.3	Protein-coding genes	87
SSC size (bp)/GC content (%)	19200/30.4	tRNAs	37
IR size (bp)/GC content (%)	52762/42.7	rRNAs	8
Total gene number	132	Genes duplicated in the IRs	22

LSC: 大单拷贝区; SSC: 小单拷贝区; IR: 反向重复区

LSC: Large single copy region; SSC: Small single copy region; IR: Inverted repeat region

表2 本研究测序的黑涩楠叶绿体基因组中的基因

Table 2 Genes in the chloroplast genome of *Aronia melanocarpa* sequenced in this study

Category	Gene group	Name of gene	Number
Self-replication	Proteins of the large ribosomal subunit	*rpl2*[ab], *rpl14*, *rpl16*[b], *rpl20*, *rpl22*, *rpl23*[a], *rpl32*, *rpl33*, *rpl36*	11
	Proteins of the small ribosomal subunit	*rps2*, *rps3*, *rps4*, *rps7*[a], *rps8*, *rps11*, *rps12*[ac], *rps14*, *rps15*, *rps16*[b], *rps18*, *rps19*[ab]	15
	Subunits of RNA polymerase	*rpoA*, *rpoB*, *rpoC1*[b], *rpoC2*	4
	rRNAs	*rrn23S*[a], *rrn16S*[a], *rrn5S*[a], *rrn4.5S*[a]	8
	tRNAs	*trnH-GUG*, *trnK-UUU*[b], *trnQ-UUG*, *trnS-GCU*, *trnG-GCC*[ab], *trnR-UCU*, *trnC-GCA*, *trnD-GUC*, *trnY-GUA*, *trnE-UUC*, *trnT-GGU*, *trnS-UGA*, *trnfM*, *trnS-GGA*, *trnS-GGA*, *trnT-UGU*, *trnL-UAA*[b], *trnF-GAA*, *trnV-UAC*[b], *trnM-CAU*, *trnW-CCA*, *trnP-UGG*, *trnI-CAU*[a], *trnL-CAA*[a], *trnV-GAC*[a], *trnI-GAU*[ab], *trnA-UGC*[ab], *trnR-ACG*[a], *trnN-GUU*[a], *trnL-UAG*	37
Photosynthesis	Subunits of photosystem I	*psaA*, *psaB*, *psaI*[a], *psaJ*	5
	Subunits of photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*[a], *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*	15
	Subunits of NADH dehydrogenase	*ndhA*[b], *ndhB*[ab], *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*	12
	Subunits of cytochrome *b*/*f* complex	*petA*, *petB*[b], *petD*[b], *petG*, *petL*, *petN*	6
	Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*[b], *atpH*, *atpI*	6
	Large subunit of Rubisco	*rbcL*	1
Biosynthesis	Maturase	*matK*	1
	Protease	*clpP*[c]	1
	Envelope membrane protein	*cemA*	1
	Acetyl-CoA carboxylase	*accD*	1
	C-type cytochrome synthesis gene	*ccsA*	1
	Translation initiation factor	*infA*	1
Unknown function	Conserved hypothetical chloroplast reading frames	*ycf1*[a], *ycf2*[a], *ycf3*[c], *ycf4*	6

a: 反向重复区中的2个基因拷贝; b: 包含2个外显子的基因; c: 包含3个外显子的基因

a: Two gene copies in inverted repeat regions; b: Genes containing two exons; c: Genes containing three exons

表3 黑涩楠叶绿体基因组中简单重复序列(SSR)的鉴定频率

Table 3 The frequency of identified of simple sequence repeats (SSRs) in the *Aronia melanocarpa* chloroplast genome

Nucleotide(s)	Number of repeats													
	6	7	10	11	12	13	14	15	16	17	18	19	20	Total
A	–	–	5	5	4	2	2	4	2	–	–	–	1	25
C	–	–	2	1	–	1	–	–	–	–	–	–	–	4
G	–	–	1	1	–	–	–	–	–	–	–	–	–	2
T	–	–	13	9	6	2	1	4	1	2	2	1	–	41
AT	2		–	–	–	–	–	–	–	–	–	–	–	2
TA	1	1	–	–	–	–	–	–	–	–	–	–	–	2

– 不存在 – Absent

表4 黑涩楠叶绿体基因组中的密码子用法

Table 4 Codon usage in the *Aronia melanocarpa* chloroplast genome

Amino acids	Codon	No.	RSCU	Amino acids	Codon	No.	RSCU
Phe	UUU	975	1.3	Ala	GCU	645	1.84
Phe	UUC	526	0.7	Ala	GCC	217	0.62
Leu	UUA	912	1.95	Ala	GCA	390	1.11
Leu	UUG	565	1.21	Ala	GCG	148	0.42
Leu	CUU	593	1.27	TER	UAA	51	1.76
Leu	CUC	186	0.4	TER	UAG	21	0.72
Leu	CUA	363	0.78	TER	UGA	15	0.52
Leu	CUG	181	0.39	His	CAU	493	1.55
Ile	AUU	1123	1.47	His	CAC	145	0.45
Ile	AUC	440	0.58	Gln	CAA	727	1.54
Ile	AUA	730	0.96	Gln	CAG	217	0.46
Met	AUG	627	1	Asn	AAU	987	1.53
Val	GUU	524	1.44	Asn	AAC	305	0.47
Val	GUC	167	0.46	Lys	AAA	1068	1.49
Val	GUA	552	1.52	Lys	AAG	364	0.51
Val	GUG	208	0.57	Asp	GAU	889	1.62
Ser	UCU	573	1.69	Asp	GAC	208	0.38
Ser	UCC	330	0.97	Glu	GAA	1035	1.48
Ser	UCA	408	1.2	Glu	GAG	363	0.52
Ser	UCG	190	0.56	Cys	UGU	226	1.49
Ser	AGU	411	1.21	Cys	UGC	77	0.51
Ser	AGC	128	0.38	Trp	UGG	458	1
Pro	CCU	421	1.56	Arg	CGU	340	1.27
Pro	CCC	201	0.74	Arg	CGC	112	0.42
Pro	CCA	310	1.15	Arg	CGA	370	1.38
Pro	CCG	149	0.55	Arg	CGG	121	0.45
Thr	ACU	551	1.6	Arg	AGA	493	1.84
Thr	ACC	251	0.73	Arg	AGG	173	0.65
Thr	ACA	423	1.23	Gly	GGU	589	1.31
Thr	ACG	153	0.44	Gly	GGC	183	0.41
Tyr	UAU	801	1.61	Gly	GGA	725	1.62
Tyr	UAC	194	0.39	Gly	GGG	295	0.66

RSCU: 相对同义密码子使用率 RSCU: Relative synonymous codon usage

2.4 蔷薇科植物完整叶绿体基因组的比较分析

使用mVISTA在线软件，对14种蔷薇科植物cp基因组进行BLAST分析，并以黑涩楠为参考。比较基因组分析表明，入选物种的基因组整体上高度保守，尤其是在编码区(图2)。相反，在这些物种中，非编码区的差异更大(图2)。在非编码区，*trnK-rps16*、*rps16-trnQ*、*trnG-atpA*、*petN-psbM*、*trnT-psbD*、*psbZ-trnG*、*trnT-trnL*、*ndhC-trnV*和*accD-psaI*等区域的差异程度较高，这可能是蔷薇科物种鉴定的重要候选区域。当然，这些区域是否适合作为分子标记用于蔷薇科物种的系统进化研究仍需进一步验证。

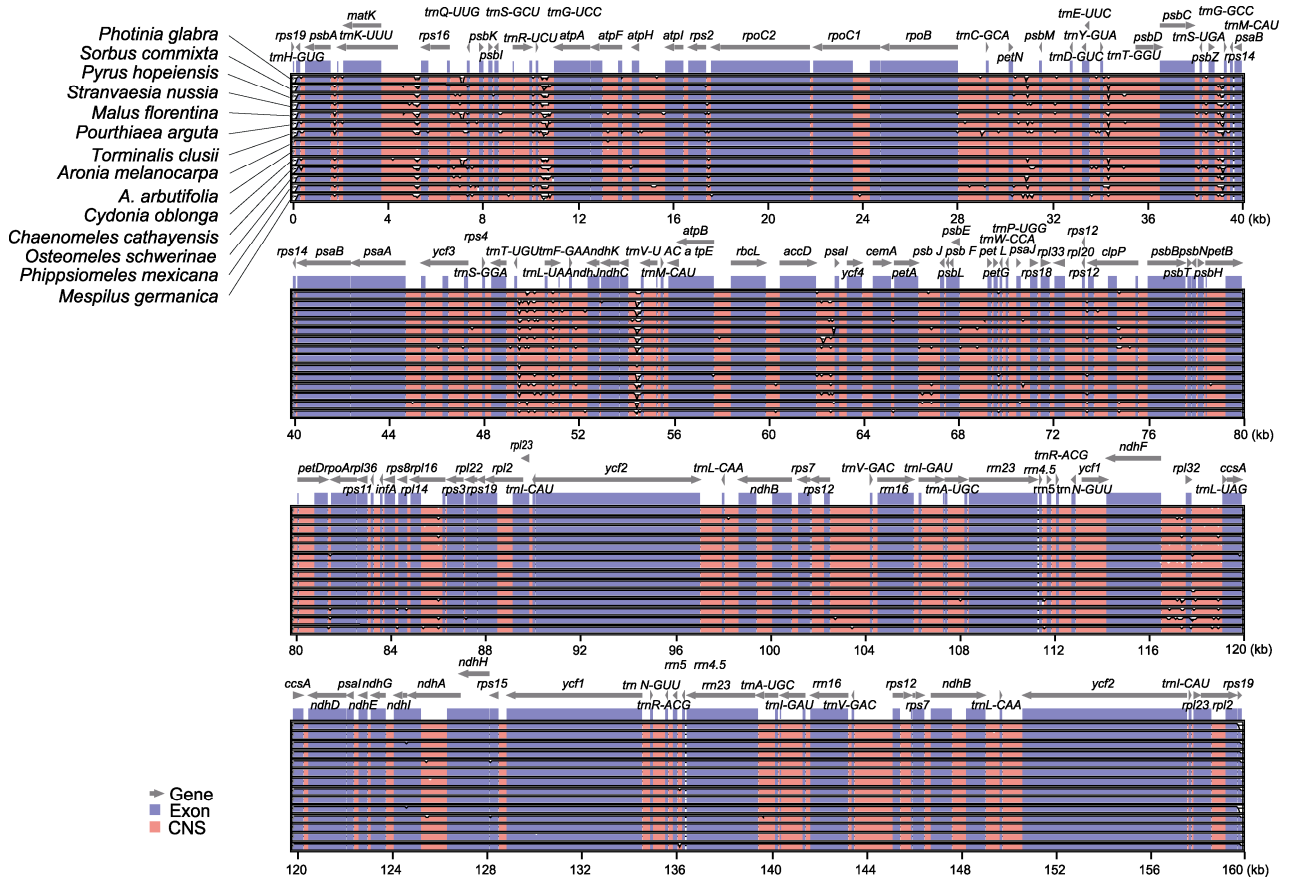


图2 14种蔷薇科植物叶绿体基因组比较

图中纵坐标和横坐标分别代表50%–100%的相同率和序列长度。注释基因显示在顶部。

Figure 2 A comparison of the chloroplast genomes among 14 Rosaceae species

The vertical scale and horizontal axes in the figure represent the percentage of identity ranging from 50% to 100% and the sequence length, respectively. Annotated genes are displayed on the top.

2.5 IR区的收缩和扩展

IR区的收缩和扩展导致cp基因组大小发生变化，在植物进化过程中起重要作用(Zhao et al., 2020)。在蔷薇科植物中检测到的14个cp基因组均含有4个典型的亚区，从而形成4个边界，分别为LSC-IRb交界处(JLB)、IRb-SSC交界处(JSB)、SSC-IRa交界处(JSA)和IRa-LSC交界处(JLA)。随着个体基因组的进化，边界区域的收缩和扩展非常常见，这也导致cp基因组大小有所不同(Guo et al., 2017)。我们比较了14种蔷薇科植物cp基因组的边界(图3)，发现了3个易变位点(*rps19*、*ycf1*和*ndhF*)，它们一般位于IR与SSC (JSA/JSB)和IR与LSC (JLA/JLB)的交界处。除光叶石楠和椤木外，所有物种的*rps19*基因都位于JLB，长度增加了120–124 bp (图3)。在IRa区还发现了1个额外的*rps19*基因，该基因与11个物种JLA处的1 bp间隙有关，除光叶石楠、弗洛伦萨海棠和椤木外，其它物种均无该基因。*ycf1*基因显示出显著的序列多样性。在其拷贝中，较短的1个拷贝位于JSB，长度从1 073 bp (金缕梨)到1 091 bp (驱疝木)不等，但在弗洛伦萨海棠中没有这个短的*ycf1*基因。JSA位于*ycf1*的第2个拷贝中，此较长拷贝的长度从5 633 bp到5 654 bp不等，只有椤木例外，其第2个拷贝的长度仅有2 297 bp。参与光合作用的基因*ndhF*也跨越了JSB，长度从2 243 bp到2 276 bp不等。而对木兰科植物的研究表明，*ndhF*主要位于SSC区(Amiryousefi et al., 2018)。此外，椤木cp基因组的边界显示出IRs的收缩，表明边界基因的变化可能由长度变化引起。

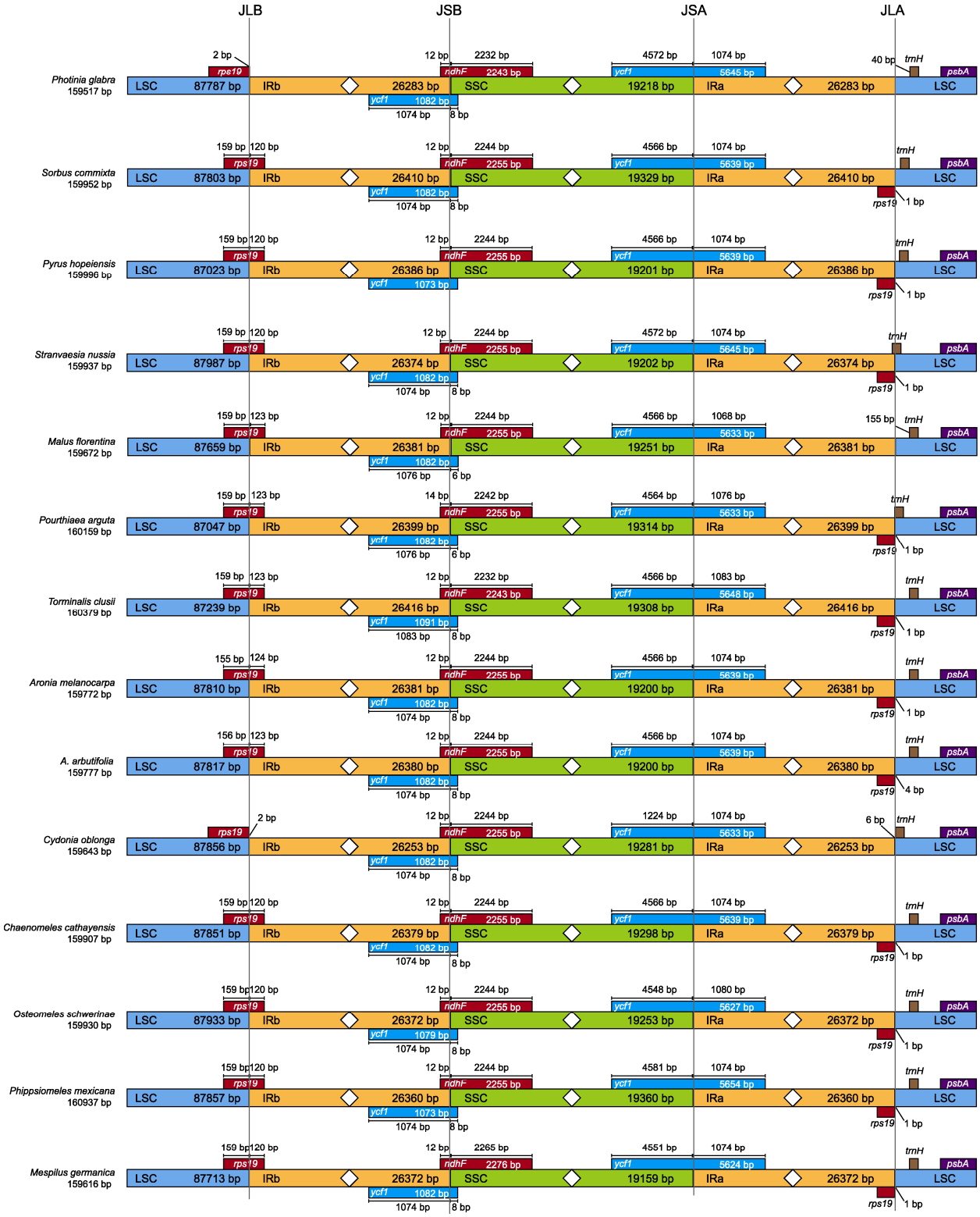


图3 蔷薇科14个物种的LSC、SSC和IR区域的边界距离比较
LSC、SSC和IR同表1。

Figure 3 Comparisons of the boundary distances for LSC, SSC, and IR regions among 14 species from the Rosaceae
LSC, SSC, and IR are the same as shown in Table 1.

2.6 系统进化分析

使用MAFFT软件对蔷薇科36属60种植物的完整cp基因组序列进行比对。基于最大似然法(ML)并分别使用iQTree和FastTree软件构建系统进化树。在系统进化分析中，将暹罗桑、裂叶蒙桑、枣和滇刺枣作为外类群。系统进化树显示，黑涩楠和红涩楠为姐妹群，支持率为100% (图4)。虽然系统进化分析表明榲桲与涩楠属聚类，但是支持率仅为36%。此外，苹果属(*Malus*)、落叶石楠属、驱疝木属(*Torminalis*)、涩石楠属、榲桲属(*Cydonia*)、木瓜属(*Chaenomeles*)、小石积属(*Osteomeles*)和金绒梨属(*Phippsiomeles*)被聚到同一分支中，支持率较高。

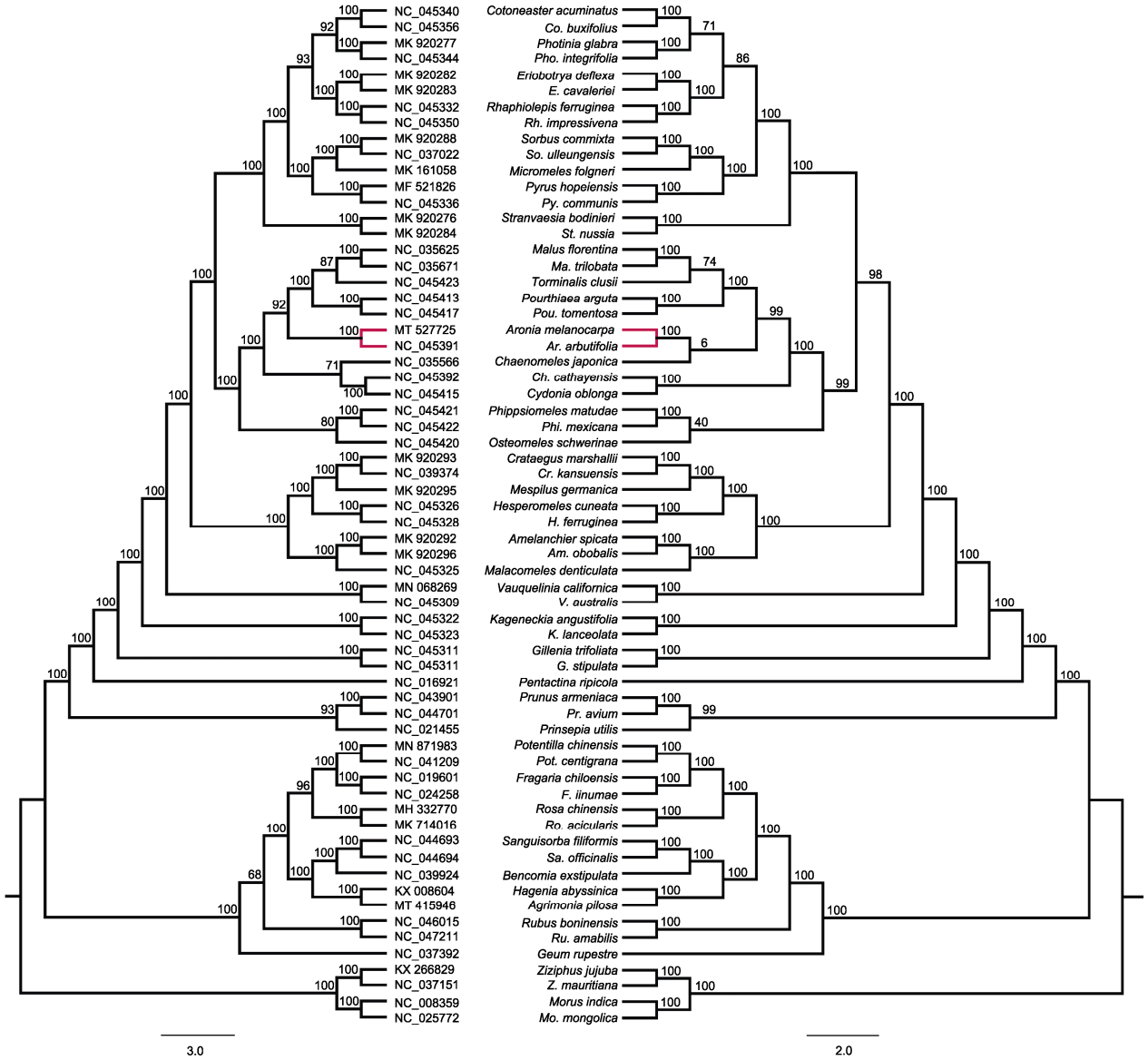


图4 利用完整的叶绿体基因组对60个蔷薇科物种进行系统进化分析

左侧为利用iQTree构建的系统发育树，节点上方的数字为自展支持率；右侧为利用FastTree构建的系统发育树，节点上方的数字为SH-like local支持率。

Figure 4 Phylogenetic analyses on 60 Rosaceae species using their complete chloroplast genomes

On the left, phylogenetic tree was constructed by iQTree, numbers above the nodes are bootstrap values; on the right, phylogenetic tree was constructed by FastTree, numbers above the nodes are SH-like local values.

3 讨论与结论

已有一些研究对涩楠属的系统进化位置进行了讨论(Campbell et al., 2007; Guo et al., 2011; Li et al., 2012)。本研究中，系统进化分析结果支持涩楠属作为一个独立的属，这与前人的研究结果一致(Guo et al., 2011; Liu et al., 2020)。例如，基于cp DNA树(*trnL-trnF*和*psbA-trnH*)，涩楠属与牛筋条属(*Dichotomanthes*)和落叶石楠属组成一个支系(Guo et al., 2011)。此外，基于完整cp基因组分析，红苦味果(*Aronia arbutifolia*)与枫棠属(*Eriolobus*)、苹果属、落叶石楠属、驱疝木属、木瓜属、榲桲属、牛筋条属、小石积属、枫棠属、花楸属、多依属(*Docynia*)和金绒梨属所在的一个大支系为姊妹关系(Liu et al., 2020)。由于分析的涩楠属物种数量有限(如缺乏紫涩楠的cp基因组数据)，因此需要增加更多的cp基因组来解决涩楠属内遗传分化和系统进化问题。

本研究利用Illumina高通量测序技术对黑涩楠完整cp基因组进行测序。黑涩楠cp基因组长度为159 772 bp；注释到132个基因，包括87个蛋白质编码基因、37个tRNA和8个rRNA。重复序列分析鉴定出76个SSRs和50个长重复序列，这些序列为进一步开发分子标记提供了支持。在基因间还检测到高变异区，如*trnK-rps16*、*rps16-trnQ*、*trnG-atpA*、*petN-psbM*、*trnT-psbD*、*psbZ-trnG*、*trnT-trnL*、*ndhC-trnV*以及*accD-psaI*，这些区域可广泛应用于遗传和系统进化研究。系统进化分析结果强烈支持黑涩楠与红涩楠亲缘关系最近的结论；其次是榲桲，但支持率较低。本研究获得的黑涩楠叶绿体基因组数据将为后续的种群遗传学和系统进化研究奠定基础，并为该属的分子育种和资源可持续利用提供理论依据。

作者贡献声明

王传永：撰写论文；庄典：完成实验；宋正达：植物材料培育；翟恒华：植物材料引种；李乃伟：构思并设计实验；张凡：分析数据并提供技术支持。

参考文献

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

Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580.

Brand M (2010). *Aronia*: native shrubs with untapped potential. *Arnoldia* **67**, 14–25.

Campbell CS, Evans RC, Morgan DR, Dickinson TA, Arsenault MP (2007). Phylogeny of subtribe Pyrinae (formerly the Maloideae, Rosaceae): limited resolution of a complex evolutionary history. *Plant Syst Evol* **266**, 119–145.

Chiapella JO, Barfuss MHJ, Xue ZQ, Greimler J (2019). The plastid genome of *Deschampsia cespitosa* (Poaceae). *Molecules* **24**, 216.

Chu ZZ, Gulbar Yisilam, Qu ZZ, Tian XM (2023). Comparative analyses on the chloroplast genome of three sympatric *Atraphaxis* species. *Chin Bull Bot* **58**, 417–432. (in Chinese)
褚振州，古丽巴哈尔·依斯拉木，屈泽众，田新民 (2023). 同域分布的3种木蓼属植物叶绿体基因组比较. 植物学报 **58**, 417–432.

Connolly BA (2014). Collection, Description, Taxonomic Relationships, Fruit Biochemistry, and Utilization of *Aronia melanocarpa*, *A. arbutifolia*, *A. prunifolia*, and *A. mitschurinii.* PhD dissertation. Connecticut: University of Connecticut. pp. 1–1.

Daniell H, Lin CS, Yu M, Chang WJ (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* **17**, 134.

Dierckxsens N, Mardulyn P, Smits G (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res* **45**, e18.

Domarew CA, Holt RR, Goodman-Snitkoff G (2002). A study of Russian phytomedicine and commonly used herbal remedies. *J Herb Pharmacother* **2**, 31–48.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32**, W273–W279.

Fu PC, Zhang YZ, Geng HM, Chen SL (2016). The complete chloroplast genome sequence of *Gentiana lawrencei* var. *farreri* (Gentianaceae) and comparative analysis with its congeneric species. *PeerJ* **4**, e2540.

Guo HJ, Liu JS, Luo L, Wei XP, Zhang J, Qi YD, Zhang BG, Liu HT, Xiao PG (2017). Complete chloroplast genome sequences of *Schisandra chinensis*: genome structure, comparative analysis, and phylogenetic relationship of basal angiosperms. *Sci China Life Sci* **60**, 1286–1290.

Guo W, Yu YJ, Shen RJ, Liao WB, Chin SW, Potter D (2011). A phylogeny of *Photinia* sensu lato (Rosaceae)

and related genera based on nrITS and cpDNA analysis. *Plant Syst Evol* **291**, 91–102.

Hardin JW (1973). The enigmatic chokeberries (*Aronia*, Rosaceae). *Bull Torrey Bot Club* **100**, 178–184.

He Y, Xiao HT, Deng C, Xiong L, Yang J, Peng C (2016). The complete chloroplast genome sequences of the medicinal plant *Pogostemon cablin*. *Int J Mol Sci* **17**, 820.

Hildebrand M, Hallick RB, Passavant CW, Bourque DP (1988). *Trans*-splicing in chloroplasts: the *rps 12* loci of *Nicotiana tabacum*. *Proc Natl Acad Sci USA* **85**, 372–376.

Katoh K, Rozewicki J, Yamada KD (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**, 1160–1166.

Kulling SE, Rawel HM (2008). Chokeberry (*Aronia melanocarpa*)—a review on the characteristic components and potential health effects. *Planta Med* **74**, 1625–1634.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* **29**, 4633–4642.

Li QY, Guo W, Liao WB, Macklin JA, Li JH (2012). Generic limits of Pyrinae: insights from nuclear ribosomal DNA sequences. *Bot Stud* **53**, 151–164.

Li YF, Sylvester SP, Li M, Zhang C, Li X, Duan YF, Wang XR (2019). The complete plastid genome of *Magnolia zenii* and genetic comparison to Magnoliaceae species. *Molecules* **24**, 261.

Lim JD, Cha HS, Choung MG, Choi RN, Choi DJ, Youn AR (2014). Antioxidant activities of acidic ethanol extract and the anthocyanin rich fraction from *Aronia melanocarpa*. *Korean J Food Cookery Sci* **30**, 573–578.

Liu BB, Liu GN, Hong DY, Wen J (2020). Eriobotrya belongs to *Rhaphiolepis* (Maleae, Rosaceae): evidence from chloroplast genome and nuclear ribosomal DNA data. *Front Plant Sci* **10**, 1731.

Liu HY, Yu Y, Deng YQ, Li J, Huang ZX, Zhou SD (2018). The chloroplast genome of *Lilium henrici*: genome structure and comparative analysis. *Molecules* **23**, 1276.

Liu QP, Xue QZ (2005). Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. *J Genet* **84**, 55–62.

Liu YC, Lin BY, Lin JY, Wu WL, Chang CC (2016). Evaluation of chloroplast DNA markers for intraspecific identification of *Phalaenopsis equestris* cultivars. *Sci Hortic* **203**, 86–94.

Lohse M, Drechsel O, Bock R (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* **52**, 267–274.

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I (2000). *VISTA*: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046–1047.

Morton BR (2003). The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. *J Mol Evol* **56**, 616–629.

Mudunuri SB, Nagarajaram HA (2007). IMEx: imperfect microsatellite extractor. *Bioinformatics* **23**, 1181–1187.

Palmer JD (1987). Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *Am Nat* **130**, S6–S29.

Price MN, Dehal PS, Arkin AP (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490.

Ravi V, Khurana JP, Tyagi AK, Khurana P (2008). An update on chloroplast genomes. *Plant Syst Evol* **271**, 101–122.

Robertson KR, Phipps JB, Rohrer JR, Smith PG (1991). A synopsis of genera in Maloideae (Rosaceae). *Syst Bot* **16**, 376–394.

Seidemann J (1993). Chockberries a fruit little known till now. *Deutsch Lebensm Rundsch* (*Ger*) **89**, 149–151.

Sharp PM, Li WH (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**, 1281–1295.

Shi LC, Chen HM, Jiang M, Wang LQ, Wu X, Huang LF, Liu C (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res* **47**, W65–W73.

Shipunov A, Gladkova S, Timoshina P, Lee HJ, Choi J, Despiegelaere S, Connolly B (2019). Mysterious chokeberries: new data on the diversity and phylogeny of *Aronia* Medik. (Rosaceae). *Eur J Taxon* **570**, 1–14.

Shukla S, Mehta A (2015). Anticancer potential of medicinal plants and their phytochemicals: a review. *Braz J Bot* **38**, 199–210.

Szopa A, Kokotkiewicz A, Kubica P, Banaszczak P, Wojtanowska-Krośniak A, Krośniak M, Marzec-Wróblewska U, Badura A, Zagrodzki P, Bucinski A, Luczkiewicz M, Ekiert H (2017). Comparative analysis of different groups of phenolic compounds in fruit and leaf extracts of *Aronia* sp.: *A*. *melanocarpa*, *A*. *arbutifolia*, and

A. × *prunifolia* and their antioxidant activities. *Eur Food Res Technol* **243**, 1645–1657.

Wang T, Zeng XY, Hu YQ, Li N (2023). Optimization of brewing technology and antioxidant activity of wild cherry berry wine. *Food Ind* **44**(6), 26–30. (in Chinese)
汪娣，曾雪莹，胡玉清，李楠 (2023). 野樱莓果酒酿造工艺优化及抗氧化活性分析. 食品工业 **44**(6), 26–30.

Wyman SK, Jansen RK, Boore JL (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**, 3252–3255.

Xu C, Dong WP, Li WQ, Lu YZ, Xie XM, Jin XB, Shi JP, He KH, Suo ZL (2017). Comparative analysis of six *Lagerstroemia* complete chloroplast genomes. *Front Plant Sci* **8**, 15.

Yang L, Ding X, Jiang ZP, Chen WW, Chen S, Ma N, Xu SY, Wang JY (2024). Optimization of fermentation technology of *Aronia melanocarpa* pomace enzyme and study on its antioxidant activity. *Chin Cond* **49**(9), 29–37. (in Chinese)
杨柳，丁雪，姜志鹏，陈文文，陈生，马宁，徐圣喻，王静怡 (2024). 野樱莓果渣酵素发酵工艺优化及抗氧化活性研究. 中国调味品 **49**(9), 29–37.

Zhao F, Drew BT, Chen YP, Hu GX, Li B, Xiang CL (2020). The chloroplast genome of *Salvia*: genomic characterization and phylogenetic analysis. *Int J Mol Sci* **181**, 812–830.

Zhao JL, Qi B, Ding LJ, Tang XQ (2010). Based on RSCU and QRSCU research codon bias of F/10 and G/11 Xylanase. *J Food Sci Biotechnol* **29**, 755–764.

Zhao YM, Yang ZY, Zhao YP, Li XL, Zhao ZX, Zhao GF (2019). Chloroplast genome structural characteristics and phylogenetic relationships of Oleaceae. *Chin Bull Bot* **54**, 441–454. (in Chinese)
赵月梅，杨振艳，赵永平，李筱玲，赵志新，赵桂仿 (2019). 木犀科植物叶绿体基因组结构特征和系统发育关系. 植物学报 **54**, 441–454.

Structural and Comparative Analysis of the Complete Chloroplast Genome of the *Aronia melanocarpa* and Its Phylogenetic Inference

Chuanyong Wang[1], Dian Zhuang[2], Zhengda Song[1], Henghua Zhai[1], Naiwei Li[1*], Fan Zhang[1*]

[1]*Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden Mem. Sun Yat-Sen, Nanjing 210014, China*
[2]*Nanjing University of Chinese Medicine, Nanjing 210023, China*

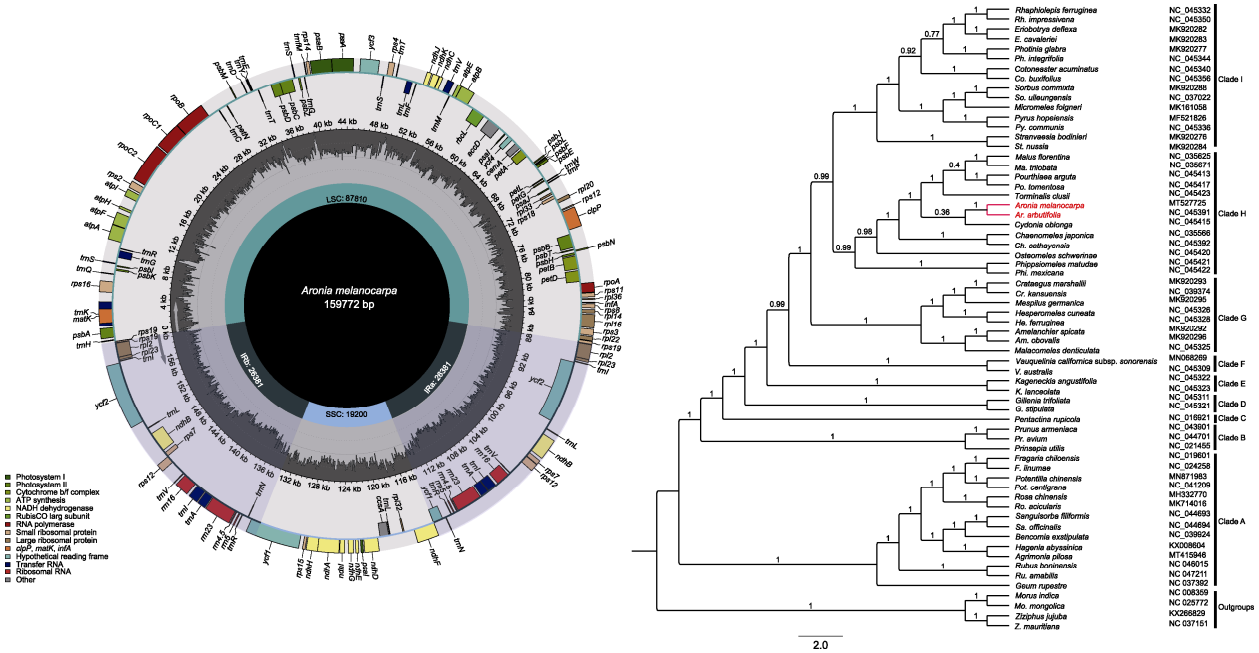
INTRODUCTION: *Aronia melanocarpa* also known as black chokeberry, belongs to the genus *Aronia* (Rosaceae). In addition to *A. melanocarpa*, *Aronia* includes *A. arbutifolia* or red chokeberry and *A. prunifolia* or purple chokeberry, both distributed naturally in North American, and an additional cultivated taxon, *A. mitschurinii* or Mitschurin's chokeberry, originating from Europe. However, the species boundaries and relationships among the species of *Aronia* are not clear. Moreover, the taxonomic history of *Aronia* is complex, as species of this genus have formerly been placed in many different genera, such as *Mespilus*, *Pyrus*, *Adenorachis*, *Sorbus*, and *Photinia*. In the present study, we first sequenced and characterized the complete chloroplast (cp) genome of *A. melanocarpa* and compared its sequence with those of the cp genomes from 13 species of the family Rosaceae. The aims of this study were: (1) to increase our understanding of the structural patterns of complete cp genome of *A. melanocarpa*; (2) to investigate the phylogenetic relationships of *A. melanocarpa* with other Rosaceae species based on their cp genomes.

RATIONALE: The chloroplast is a unique and essential organelle in green plants with vital roles in photosynthesis and carbon fixation. Comparative analyses of cp genomes between different plant species reveal intra- and inter-species rearrangements that have occurred during evolution, such as inverted repeat (IR) contraction and expansion. Based on these characteristics, the cp genome has been wildly used for species identification, phylogenetic analysis, and exploring the genetic basis of environmental adaptation.

RESULTS: The complete *A. melanocarpa* cp genome was sequenced, analyzed, and compared with that from 13 other species in the Rosaceae. The cp genome is 159 772 bp and has a total guanine-cytosine (GC) content of 36.6%. It exhibits a typical quadripartite structure with four separate regions, including a large single copy (LSC) region of 87 810 bp and a small single copy (SSC) region of 19 200 bp separated by two inverted repeats (IRa and IRb) regions of 26 381 bp each. A total of 132 genes were annotated, including 87 protein-coding genes, 37 tRNAs, and eight rRNAs, with 22 duplicates in the IR re-

gions. In total, 76 simple sequence repeats (SSRs) and 50 long repeats were detected. Phylogenetic analysis indicated that *A. melanocarpa* is most closely related to *A. arbutifolia* and forms a sister clade to *Cydonia oblonga* with weak support.

CONCLUSION: We analyzed the complete cp genome of *A. melanocarpa* by using Illumina high-throughput sequencing technology. The sequence of *A. melanocarpa* cp genome could be further used for the development of molecular markers. Highly variable regions were detected in intergenic regions, such as *trnK-rps16*, *rps16-trnQ*, *trnG-atpA*, *petN-psbM*, *trnT-psbD*, *psbZ-trnG*, *trnT-trnL*, *ndhC-trnV* and *accD-psaI*, which might be useful for broad applications in genetic research studies as well as phylogenetic studies. Phylogenetic construction results strongly supported that *A. melanocarpa* was closest related to *A. arbutifolia*, followed by *C. oblonga* with weak support. This newly available genomic data for *A. melanocarpa* will provide a basis for future research on the population genetics and phylogenomics and will benefit the breeding studies and utilization of the genus *Aronia*.



Map of the chloroplast genome of *Aronia melanocarpa* and phylogenetic analyses among the 60 Rosaceae species using their complete chloroplast genomes. *Aronia* formed a clade with Dichotomanthes and Pourthiaea based on cpDNA tree. Moreover, *A. melanocarpa* is most closely related to *A. arbutifolia* and forms a sister clade to *Cydonia oblonga* with weak support.

Key words *Aronia melanocarpa*, chloroplast genome, structural variation, phylogenetic relationships

Wang CY, Zhuang D, Song ZD, Zhai HH, Li NW, Zhang F (2025). Structural and comparative analysis of the complete chloroplast genome of the *Aronia melanocarpa* and its phylogenetic inference. *Chin Bull Bot* **60**, 573–585.

---

* Authors for correspondence. E-mail: linaiwei@jib.ac.cn; mumizhongfeng@126.com

(责任编辑: 白羽红)

附表 1 黑涩楠叶绿体基因组的重复序列分析

Appendix table 1 Analysis of repeat sequence in the *Aronia melanocarpa* chloroplast genome

附表 2 黑涩楠叶绿体基因组中的简单重复序列

Appendix table 2 Simple sequence repeats (SSRs) in the *Aronia melanocarpa* chloroplast genome

https://www.chinbullbotany.com/fileup/1674-3466/PDF/24-146-1.pdf

扫一扫看附表