

• 研究报告 •

由春香 博士，山东农业大学教授、博士生导师。多年从事果树分子生物学与生物技术领域的相关研究，主要研究方向为果树逆境与果实品质分子生物学，以第一或通讯作者在《园艺学报》、*Plant Biotechnology Journal*、*Plant Journal*、*Plant Physiology*、*Horticulture Research*、*Plant Cell and Environment*、*Journal of Experimental Botany* 等国内外学术期刊上发表论文 70 多篇，参编著作 4 部。



多花海棠叶绿体基因组分析

王寻¹，冯资权¹，王大儒¹，韩月彭²，王小非¹，沈向¹，由春香¹

1 山东农业大学 园艺科学与工程学院 作物生物学国家重点实验室，山东 泰安 271018

2 中国科学院武汉植物园 中国科学院植物种质创新与特色农业重点实验室，湖北 武汉 430074

王寻，冯资权，王大儒，韩月彭，王小非，沈向，由春香. 多花海棠叶绿体基因组分析. 生物工程学报, 2022, 38(10): 3713-3727.

WANG X, FENG ZQ, WANG DR, HAN YP, WANG XF, SHEN X, YOU CX. Chloroplast genome in *Malus floribunda* Siebold. Chin J Biotech, 2022, 38(10): 3713-3727.

摘要：多花海棠 (*Malus floribunda* Siebold.) 是世界范围内广泛栽培的苹果属物种，具有较高的观赏价值和育种意义。对其进行叶绿体基因组比较分析，有利于完善苹果属系统进化以及种质利用的研究内容。基于全基因组测序数据，组装获得一个完整的具有四分体结构的多花海棠叶绿体基因组。该基因组包括大单拷贝区 (88 142 bp)、反向重复区 B (26 353 bp)、小单拷贝区 (19 189 bp) 与反向重复区 A (26 353 bp)，共计 160 037 bp。多花海棠叶绿体全基因组共注释到 111 个基因，包括 78 个蛋白编码基因、29 个 tRNA 基因和 4 个 rRNA 基因。此外，在其基因组中识别到大量的重复序列，与三叶海棠和变叶海棠略有差异。通过计算相对同义密码子使用度，发现其高频密码子共 30 种，并且密码子具有偏向 A/T 结尾的使用模式。种间序列比对、边界分析的结果

Received: February 23, 2022; **Accepted:** April 11, 2022

Supported by: National Natural Science Foundation of China (32172538, 32072520); National Modern Apple Industry Technology System Funded Program of China (CARS-27); Natural Science Foundation of Shandong Province, China (ZR2020MC132)

Corresponding authors: YOU Chunxiang. E-mail: youchunxiang@sdau.edu.cn

SHEN Xiang. E-mail: shenx@sdau.edu.cn

基金项目：国家自然科学基金 (32172538, 32072520); 国家苹果产业技术体系 (CARS-27); 山东省自然科学基金 (ZR2020MC132)

表明, 大单拷贝区序列变异较大, 8种苹果属植物 SC 区与 IR 区扩张收缩情况整体上较为相似。基于叶绿体基因组序列的系统进化分析, 将多花海棠、湖北海棠和变叶海棠聚为一类。多花海棠叶绿体基因组的研究可为今后遗传标记开发与种质资源利用等提供数据支持。

关键词: 多花海棠; 叶绿体基因组; 系统进化; 生物信息学; 比较基因组学分析

Chloroplast genome in *Malus floribunda* Siebold.

WANG Xun¹, FENG Ziquan¹, WANG Daru¹, HAN Yuepeng², WANG Xiaofei¹, SHEN Xiang¹, YOU Chunxiang¹

1 State Key Laboratory of Crop Biology, College of Horticultural Science and Engineering, Shandong Agricultural University, Tai'an 271018, Shandong, China

2 Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Chinese Academy of Sciences, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, Hubei, China

Abstract: *Malus floribunda* Siebold. (*Malus*) is widely cultivated all over the world, which is of high ornamental value and breeding significance. Comparative analysis of the chloroplast genome can help enrich the phylogenetic relationship and facilitate germplasm utilization of *Malus*. Based on the whole genome sequencing data, a complete chloroplast genome (*M. floribunda*) with tetrad structure was assembled. The chloroplast genome (160 037 bp) was composed of a large single-copy (LSC) region (88 142 bp), inverted repeat (IR) B (26 353 bp), a small single-copy (SSC) region (19 189 bp), and IRA (26 353 bp). A total of 111 genes were annotated: 78 protein-coding genes, 29 tRNA genes and 4 rRNA genes. In addition, a large number of repeat sequences were identified in the genome, which was slightly different from that of *M. sieboldii* and *M. toringoides*. As for the relative synonymous codon usage, 30 high-frequency codons were found, and the codons tended to end with A/T. The results of interspecific sequence alignment and boundary analysis suggested the sequence variation of the LSC region was large, and the expansion and contraction of the SC region and IR region of the eight *Malus* species were generally similar. According to the phylogenetic analysis of chloroplast genome sequences, *M. floribunda*, *M. hupehensis*, and *M. toringoides* were grouped into one clade. The findings in this study can provide data support for the development of genetic markers and utilization of germplasm resources in the future.

Keywords: *Malus floribunda*; chloroplast genome; phylogenetic; bioinformatics; comparative genomics analysis

多花海棠 (*Malus floribunda* Siebold.) 隶属蔷薇科 (Rosaceae)、苹果亚科 (Maloideae)、苹果属 (*Malus*)^[1-2], 多花海棠作为一类重要的苹果属栽培种, 在世界范围内广泛栽培, 因其花多且美, 具有极高的观赏价值。本种叶不分裂, 叶缘有锐锯齿; 花蕾深紫红色, 开花后颜色变

浅, 经由粉红色后至白色; 果为深洋红色^[2]。除了其优良的观赏特性^[3], 多花海棠在种质评价和抗性育种中也起到重要作用^[4-6]。研究发现, 多花海棠中存在苹果黑星病 (由苹果黑星病菌 *Venturia inaequalis* Wint.引起) 抗性基因^[7], 基于此材料的育种潜力, 多国科学家先后将多花海

棠抗黑星病特性通过杂交技术引入到新品种中,为苹果属抗病育种作出巨大贡献^[8]。然而,多花海棠系统进化及资源分类等有关内容却不甚明晰,因此,开展遗传学、基因组学等研究对其变异检测和种质利用尤为必要。

叶绿体基因组研究作为一种手段,为植物种质资源鉴定与分类、起源演化与系统发育、遗传转化与基因工程等提供新的方向^[9-10]。目前,部分苹果属植物叶绿体基因组已发布,涉及的材料包括山荆子 (*Malus baccata* (L.) Borkh.)、湖北海棠 (*Malus hupehensis* (Pamp.) Rehd.)、楸子 (*Malus prunifolia* (Willd.) Borkh.)、变叶海棠 (*Malus toringoides* (Rehd.) Hughes) 等。关于它们叶绿体基因组的比较分析也已有报道^[11-13],这极大地方便了苹果属植物的变异检测与种质利用。因此,完成叶绿体基因组图谱的构建和比较分析,对于揭示多花海棠的遗传特征和生物学地位等非常关键。

本研究基于全基因组测序数据,对多花海棠叶绿体基因组进行组装,并初步对其基因组组成、重复序列特征、密码子使用模式、序列相似性、系统进化关系等展开分析,为今后苹果属叶绿体比较基因组学研究提供参考。

1 材料与方法

1.1 实验材料

所用实验材料为多花海棠 (*Malus floribunda* Siebold.),样品取自国家苹果工程中心青岛苹果砧木研发中心。

1.2 实验方法

1.2.1 基因组建库及测序

利用CTAB法提取多花海棠叶片的全基因组DNA,待质量检测后于Illumina平台建库测序,其中DNA测序文库的插入片段大小为350 bp。

1.2.2 叶绿体基因组组装与注释

借助细胞器基因组组装工具 GetOrganelle (<https://github.com/Kinggerm/GetOrganelle>)^[14]获取多花海棠叶绿体基因组,选用从头组装工作流程,其中参数k值设定为21、45、65、85和125。Bandage (<http://rrwick.github.io/andage/>)^[15]用于组装结果的确认。叶绿体基因组序列的注释在PGA (<https://github.com/quxiaojian/PGA>)^[16]、PGAVAS2 (<http://www.herbalgenomics.org/pgavas/>)^[17]中完成,而后人工检查并修正注释信息。使用OGDRAW软件 (<https://chlorobox.mpimp-golm.pg.de/OGDraw.html>)进行叶绿体基因组的可视化^[18]。多花海棠叶绿体基因组已提交至NCBI (<https://www.ncbi.nlm.nih.gov/>),获得其登录号为OM115955。

1.2.3 叶绿体基因组特征及比较基因组学分析

对于多花海棠叶绿体基因组GC含量、基因分类与内含子分布分别利用Perl程序等分析,MISA-web (<https://webblast.ipk-gatersleben.e/misa/>)^[19]鉴定简单重复序列(参数保持默认),REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/eputer>)^[20]识别叶绿体基因组中的散在重复序列(最小重复单元大小为30 bp)。CodonW (<http://odonw.sourceforge.net/>)^[21]计算密码子使用情况。叶绿体基因组相似性比较利用Circoletto (<http://tools.bat.infospire.org/circoletto/>)完成,多序列比对在mVISTA (<https://genome.lbl.gov/ista/mvista/submit.shtml>)^[22]中进行,叶绿体基因组区域边界比较利用IRscope (<https://irscope.hinyapps.io/irapp/>)^[23],使用HomBlocks (<https://ithub.com/fenghen360/HomBlocks>)^[24]和MEGA_X (<https://www.megasoftware.net/>)^[25]对不同物种进行系统进化分析。本文比较基因组学分析所用的叶绿体基因组来自NCBI数据库,具体信息参见表1。

表 1 本文比较基因组学分析所用的叶绿体基因组

Table 1 Chloroplast genomes used for comparative genomics analysis in this study

Sample name	Family	Genus	Accession No.	Size (bp)
<i>Malus baccata</i>	Rosaceae	<i>Malus</i>	MK896774.1	160 024
<i>Malus hupehensis</i>	Rosaceae	<i>Malus</i>	MK020147.1	160 065
<i>Malus sieboldii</i>	Rosaceae	<i>Malus</i>	MT593044.1	160 040
<i>Malus toringoides</i>	Rosaceae	<i>Malus</i>	MT483999.1	160 093
<i>Malus yunnanensis</i>	Rosaceae	<i>Malus</i>	MH394387.1	160 067
<i>Malus prattii</i>	Rosaceae	<i>Malus</i>	MH929090.1	160 239
<i>Malus prunifolia</i>	Rosaceae	<i>Malus</i>	KU851961.1	160 041
<i>Malus floribunda</i>	Rosaceae	<i>Malus</i>	OM115955	160 037
<i>Crataegus hupehensis</i>	Rosaceae	<i>Crataegus</i>	MW201730.1	159 766
<i>Arabidopsis thaliana</i>	Brassicaceae	<i>Arabidopsis</i>	NC_000932.1	154 478
<i>Oryza sativa</i>	Poaceae	<i>Oryza</i>	NC_031333.1	134 502

2 结果与分析

2.1 多花海棠叶绿体基因组结构

经基因组测序并组装, 获得一个完整的苹果属多花海棠叶绿体基因组, 其总长为 160 037 bp。根据序列与结构特征, 多花海棠叶绿体基因组共分为 4 个区域, 分别为大单拷贝区 (large single-copy region, LSC, 1–88 142 bp), 反向重复区 B (inverted repeat B, IRB, 88 143–114 495 bp), 小单拷贝区 (small single-copy region, SSC, 114 496–133 684 bp) 与反向重复区 A (inverted repeat A, IRA, 133 685–160 037 bp), 结果见图 1。

2.2 多花海棠叶绿体基因组组成

2.2.1 叶绿体基因组基因特征

经统计, 多花海棠叶绿体基因组 GC 含量为 36.6%, 各区域 GC 含量组成分别为: LSC (34.2%)、IRB (42.7%)、SSC (30.4%)、IRA (42.7%)。对其序列进行注释发现, 多花海棠叶绿体全基因组共 129 个基因 (去重 111 个), 分别为 84 个蛋白编码基因 (去重 78 个)、37 个

tRNA 基因 (去重 29 个) 和 8 个 rRNA 基因 (去重 4 个)。大部分基因均为 1 个拷贝, 其中蛋白编码基因 *rpl2*、*rpl23*、*ycf2*、*ndhB*、*rps12*、*rps7*, tRNA 基因 *trnI-CAU*、*trnL-CAA*、*trnV-GAC*、*trnI-GAU*、*trnA-UGC*、*trnR-ACG*、*trnN-GUU*、*trnM-CAU*, rRNA 基因 *rrn16*、*rrn23*、*rrn4.5*、*rrn5* 各有 2 个拷贝。由注释结果可知, 蛋白编码基因依据功能可分为 4 类, 包括参与光合作用的 45 个基因、用于自我复制的 25 个基因、其他类的 5 个基因以及功能未知的 3 个基因 (表 2)。此外, 对多花海棠叶绿体基因的内含子进行分析, 发现大部分基因不含内含子, 其中 11 个蛋白编码基因和 8 个 tRNA 基因分别含有 1 个内含子, 另外 4 个基因 *rps12* (×2)、*ycf3* 和 *clpP* 则各含有 2 个内含子, 详细信息列于表 3。

与其他 4 种苹果属相比, 多花海棠叶绿体基因组大小与三叶海棠 (*Malus sieboldii*) 的 (160 040 bp) 为接近, 多花海棠、山荆子、湖北海棠和三叶海棠的 IR 区长度基本

一致,均小于变叶海棠的 26 362 bp (表 1 和表 4)。对于基因数目,多花海棠和山荆子叶绿体基因总数相同,而湖北海棠则多 1 个;多花海棠和湖北海棠具有相同的叶绿体编码基因数目;另外,5 种苹果属植物都包含

4 个 rRNA。

2.2.2 简单重复序列和散在重复序列

在多花海棠叶绿体全基因组中对简单重复序列 (simple sequence repeats, SSRs) 进行鉴定,结果显示共识别 73 个 SSR,包括 27 个 A、4 个

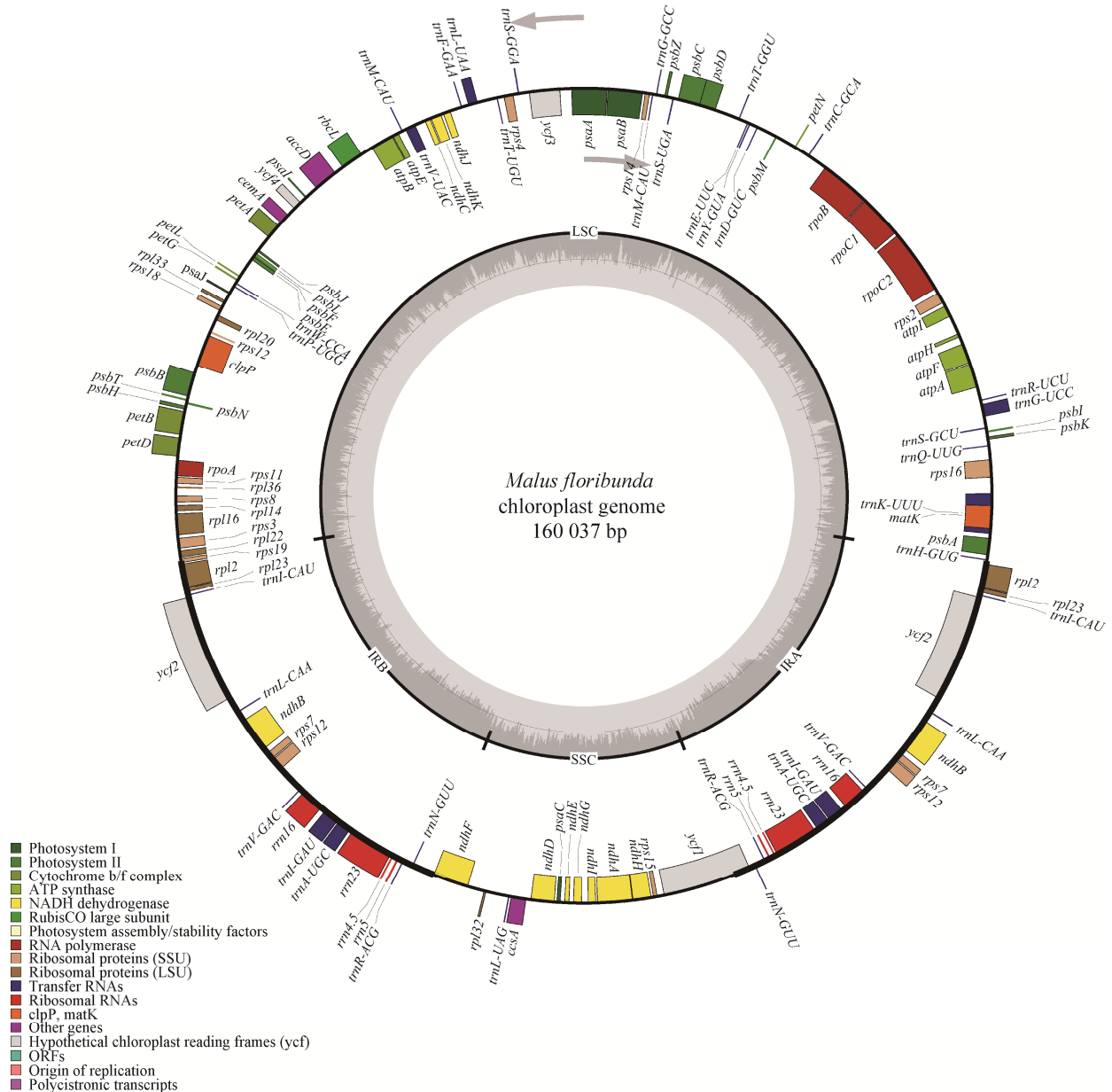


图 1 多花海棠叶绿体基因组图谱 箭头指示转录方向,阴影部分表示 GC/AT 含量

Figure 1 Genome map of *Malus floribunda* chloroplast. Arrows indicate the direction of transcription, and the shaded areas show GC/AT content.

表 2 多花海棠叶绿体基因组编码基因

Table 2 Coding genes of chloroplast genome in *Malus floribunda*

Gene category	Gene group	Gene name
Photosynthesis	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, ycf3</i>
	Subunits of NADH-dehydrogenase	<i>ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Subunits of cytochrome b/f complex	<i>petA, petB, petD, petG, petL, petN</i>
	Subunits of photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Subunit of rubisco	<i>rbcL</i>
Self replication	Large subunit of ribosome	<i>rpl14, rpl16, rpl2, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
	DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	Small subunit of ribosome	<i>rps11, rps12, rps14, rps15, rps16, rps18, rps19, rps2, rps3, rps4, rps7, rps8</i>
Others	Subunit of acetyl-CoA-carboxylase	<i>accD</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
	Envelop membrane protein	<i>cemA</i>
	Protease	<i>clpP</i>
	Maturase	<i>matK</i>
Unknown	Conserved open reading frames	<i>ycf1, ycf2, ycf4</i>

表 3 多花海棠叶绿体基因内含子分布

Table 3 Intron distribution of chloroplast genes in *Malus floribunda*

Gene	Strand	Start	End	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
<i>trnK-UUU</i>	Reverse	1 695	4 263	37	2 496	35		
<i>rps16</i>	Reverse	5 242	6 365	40	862	221		
<i>trnG-UCC</i>	Forward	9 073	9 841	23	697	48		
<i>atpF</i>	Reverse	12 509	13 796	144	732	411		
<i>rpoC1</i>	Reverse	21 933	24 719	435	740	1 611		
<i>ycf3</i>	Reverse	45 561	47 519	126	710	228	746	153
<i>trnL-UAA</i>	Forward	50 782	51 382	37	513	50		
<i>trnV-UAC</i>	Reverse	54 932	55 599	39	591	37		
<i>clpP</i>	Reverse	73 772	75 813	71	829	289	629	228
<i>petB</i>	Forward	78 751	80 195	6	796	642		
<i>petD</i>	Forward	80 386	81 592	8	723	475		
<i>rpl16</i>	Reverse	85 132	86 528	9	988	399		
<i>rpl2</i>	Reverse	88 333	89 843	390	685	435		
<i>ndhB</i>	Reverse	98 945	101 146	777	668	756		
<i>rps12</i>	Reverse	101 988	102 789	114		231	540	30
<i>trnI-GAU</i>	Forward	106 586	107 605	42	942	35		
<i>trnA-UGC</i>	Forward	107 670	108 549	39	806	34		
<i>ndhA</i>	Reverse	124 919	127 143	552	1 132	540		
<i>trnA-UGC</i>	Reverse	139 631	140 510	39	806	34		
<i>trnI-GAU</i>	Reverse	140 575	141 594	42	942	35		
<i>rps12</i>	Forward	145 391	146 192	114		231	540	30
<i>ndhB</i>	Forward	147 034	149 235	777	668	756		
<i>rpl2</i>	Forward	158 337	159 847	390	685	435		

C、39 个 T 和 3 个 AT 重复 (图 2A)。其中单核苷酸重复占到最大比重, 3 个二核苷酸重复分别是位于 LSC 区的 (AT) 单元重复 6 次的 2 个 SSR 和位于 SSC 区的 (AT) 重复 6 次的 1 个 SSR。比较发现, 三叶海棠和

变叶海棠中分布的 SSR 数目分别为 74 和 72, 与多花海棠相类似 (图 2A)。多花海棠叶绿体基因组中, SSR 集中分布于 LSC 区和 SSC 区, 尤以 LSC 区的 SSR 数目为最多 (图 2B)。

表 4 多花海棠和其他苹果属叶绿体基因组特征比较

Table 4 Comparison of chloroplast genome characteristics of *Malus floribunda* and other *Malus* species

<i>Malus</i> species	LSC length (bp)	SSC length (bp)	IR length (bp)	Gene number	CDS number	tRNA number	rRNA number
<i>Malus floribunda</i>	88 142	19 189	26 353	111	78	29	4
<i>Malus baccata</i>	88 134	19 182	26 354	111	76	31	4
<i>Malus hupehensis</i>	88 166	19 193	26 353	112	78	30	4
<i>Malus sieboldii</i>	88 149	19 183	26 354	110	76	30	4
<i>Malus toringoides</i>	88 183	19 186	26 362	110	76	30	4

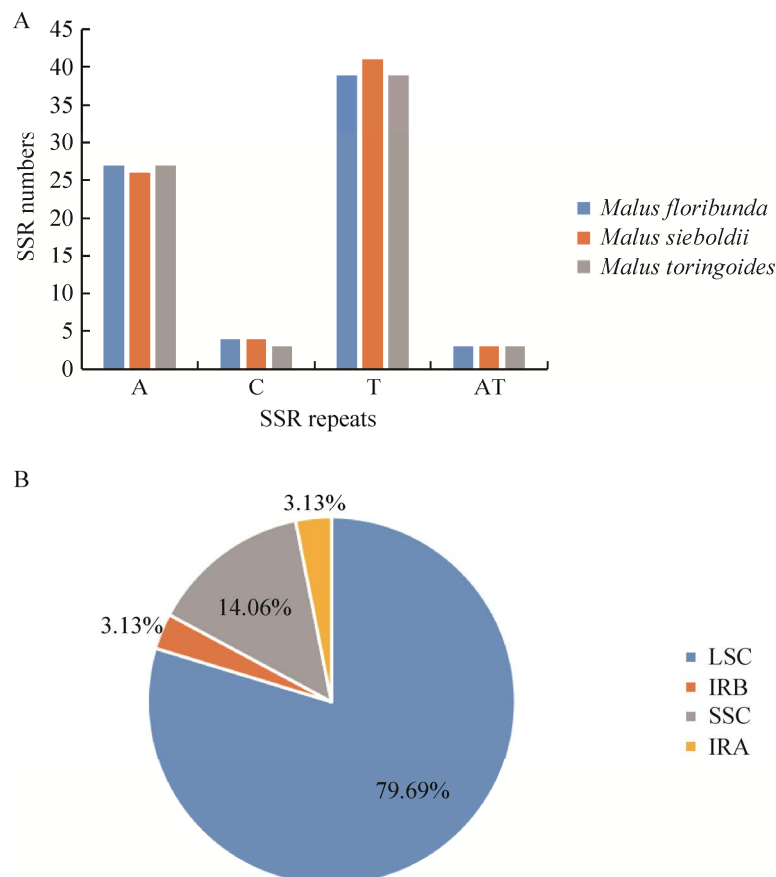


图 2 多花海棠叶绿体基因组简单重复序列 A: 多花海棠和 2 种苹果属植物简单重复序列重复单元比较; B: 多花海棠叶绿体中简单重复序列的基因组分布情况

Figure 2 SSRs of chloroplast genome in *Malus floribunda*. (A) The comparison of SSRs between *M. floribunda* and two *Malus* species. (B) The distribution of SSRs in chloroplast genome of *M. floribunda*.

通过在多花海棠叶绿体基因组中搜索散在重复序列 (interspersed nuclear elements, INEs) (图 3A), 识别正向重复 (forward, F)、反向重复 (reverse, R) 和回文序列 (palindromic, P), 各 28、5 和 24 个。相比而言, 变叶海棠叶绿体基因组中散在重复序列最多 (58 个); 而三叶海棠中则发现 1 个互补类型 (complement, C) 的 INE, 其在多花海棠和变叶海棠中均不存在。对 INE 的序列长度分析发现 (图 3B), 多花海棠叶绿体基因组中, 大部分散在重复集中在

30–35 bp, 尤以 30 bp 数目为最多 (17 个)。

2.2.3 密码子使用模式

对多花海棠叶绿体基因组中密码子使用情况进行分析, 通过计算相对同义密码子使用度 (relative synonymous codon usage, RSCU) 得知, 使用概率较高的密码子包括 UUA (亮氨酸, Leu)、GCU (丙氨酸, Ala)、AGA (精氨酸, Arg)、UCU (丝氨酸, Ser)、GAU (天冬氨酸, Asp)、GGA (甘氨酸, Gly)、UAU (酪氨酸, Tyr)、ACU (苏氨酸, Thr) 等, 它们的 RSCU 值均大于 1.6。

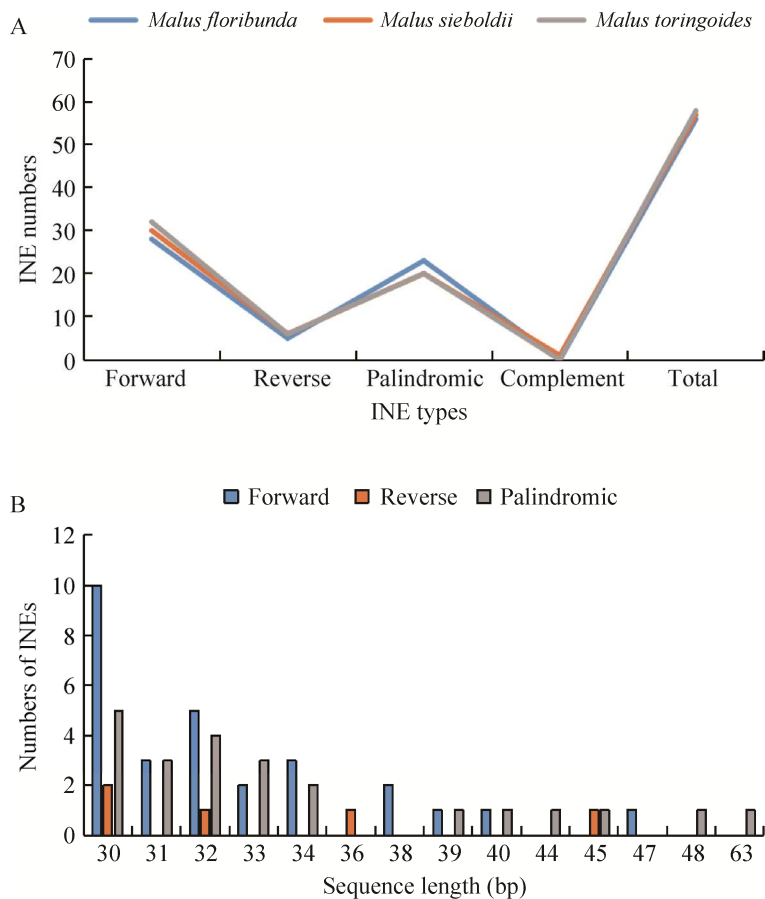


图 3 多花海棠叶绿体基因组散在重复序列 A: 多花海棠和 2 种苹果属植物散在重复序列重复类型比较; B: 多花海棠叶绿体基因组不同长度的散在重复数目

Figure 3 INEs of chloroplast genome in *Malus floribunda*. (A) The comparison of INE types between *M. floribunda* and two *Malus* species. (B) The numbers of different length of INEs in chloroplast genome of *M. floribunda*.

编码异亮氨酸 (Ile) 的 AUU、编码赖氨酸 (Lys) 的 AAA、编码谷氨酸 (Glu) 的 GAA 出现次数较多, 分别为 1 104、1 040 和 1 008 次; 而密码子 UGC (半胱氨酸, Cys) 在所有密码子中数目最少 (75 个)。由统计结果知 (图 4), 编码 Leu 的密码子 (共 6 种, 即 UUA、UUG、CUU、CUC、CUA 和 CUG) 在全部编码序列中数目最多, 总计 2 749 个, 编码 Cys 的密码子 (UGU 和 UGC) 数目最少 (共 298 个)。另外, 注意到高频密码子 (RSCU>1) 共 30 种, 而其中有 29 种以碱基 A/T 结尾, 这说明多花

海棠叶绿体基因组中密码子具有偏向 A/T 结尾的使用模式。

2.3 多花海棠叶绿体基因组序列比对

2.3.1 基因组相似性

基于多花海棠和同属叶绿体全基因组序列, 通过局部比对研究了它们的同源性和相似性。由图 5 可知, 多花海棠与同属的山荆子、湖北海棠、三叶海棠和变叶海棠具有良好的共线性关系, 比对区域基本覆盖各基因组。高相似的序列特征, 表明这些苹果属植物叶绿体结构较为稳定。

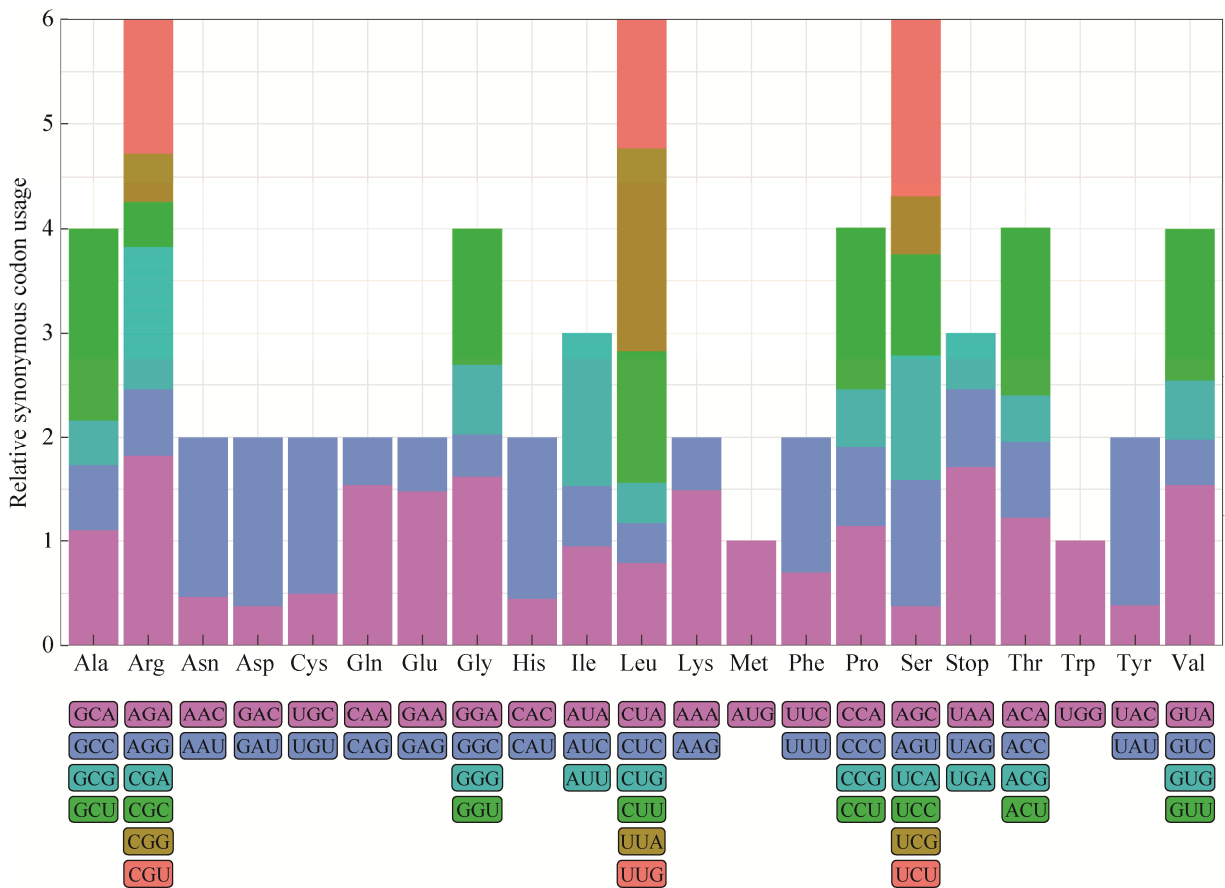


图 4 多花海棠叶绿体基因密码子使用模式 横轴表示氨基酸及其对应密码子, 纵轴代表相对同义密码子使用度

Figure 4 Codon usage mode of chloroplast genes in *Malus floribunda*. The horizontal axis represents amino acids and their corresponding codons, and the vertical axis represents relative synonymous codon usage.

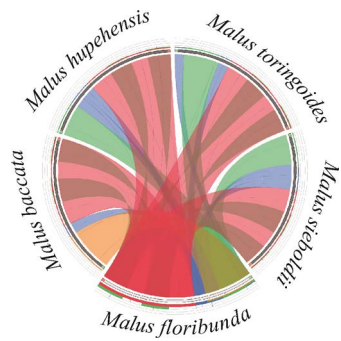


图5 多花海棠叶绿体基因组相似性比较 4种颜色的色带代表不同相对得分的比对区域,其中:红色>75%,蓝色≤25%

Figure 5 Similarity comparison of *Malus floribunda* chloroplast genome. The ribbons of four colors represent the comparison regions of different relative scores: red>75%, blue≤25%.

2.3.2 变异热点

为了比较多海棠和苹果属的其他物种叶绿体基因组的变异热点区域,本文在NCBI数据库中获取了湖北海棠、变叶海棠、西蜀海棠(*Malus prattii*)、楸子、山荆子、三叶海棠、滇池海棠(*Malus yunnanensis*)共7个叶绿体基因组。比对结果显示(图6),在叶绿体全基因组范围内,这些物种的序列较为一致,但是仍然存在部分区域差异较大,如 *rps16-trnK* (UUU)、*trnR* (UCU)-*atpA*、*atpH-atpF*、*trnT* (GGU)-*psbD*、*psbZ-trnG* (GCC)、*trnT* (UGU)-*trnL* (UAA)、*trnV* (UAC)-*ndhC*、*trnM* (CAU)-*atpE*、*accD-psaI*、*psaJ-rpl33*、*rps3-rpl16*、*rps19-rpl22*、*ndhF-trnN*

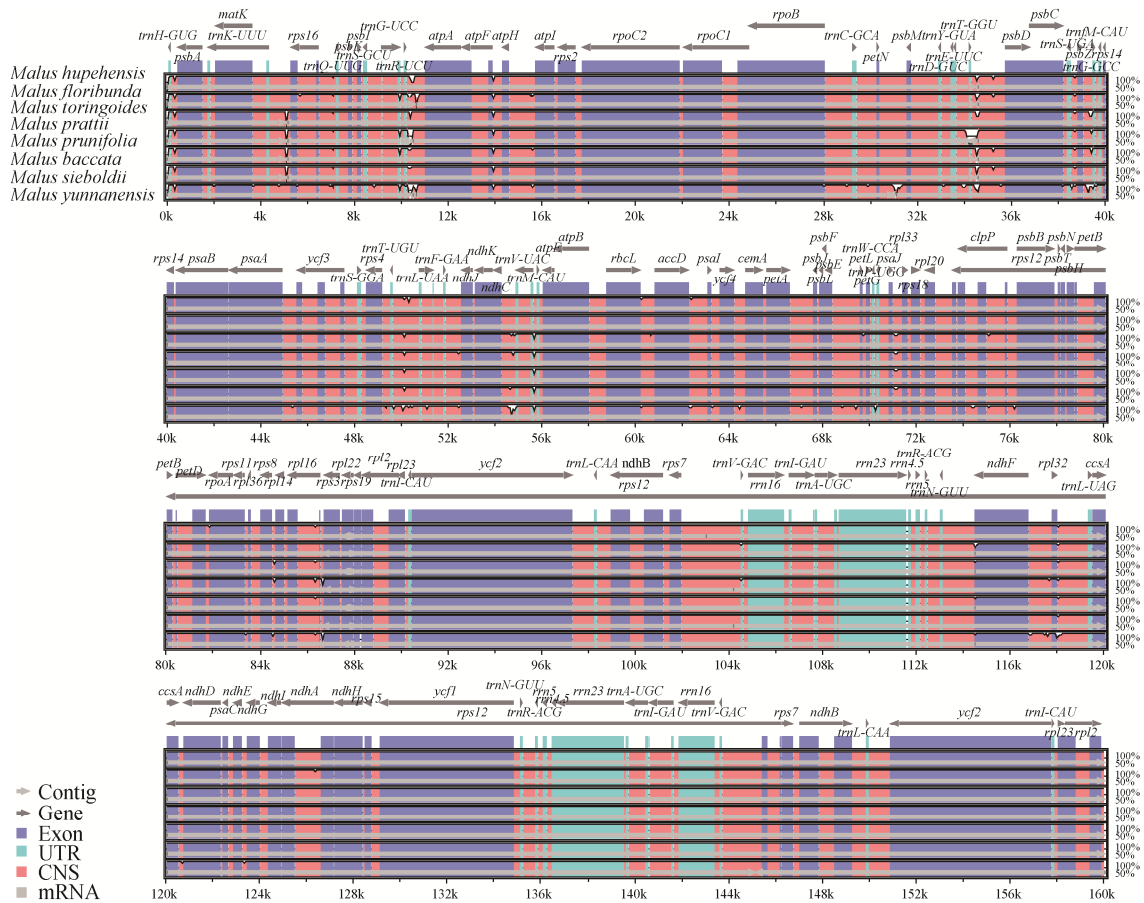


图6 多花海棠叶绿体基因组序列比对 横轴代表叶绿体基因组序列长度,纵轴表示相似性百分比
Figure 6 Sequence alignment of chloroplast genome in *Malus floribunda*. The horizontal axis represents the sequence length of chloroplast genome, and the vertical axis shows the percentage of similarity.

(GUU)、*rpl32-trnL* (UAG) 等。进一步暗示, 苹果属植物叶绿体中这些区域为物种变异提供了基础, 因此高度变异区域的比较对于研究物种进化以及鉴定具有重要的意义。

2.3.3 边界分析

叶绿体基因组中 4 个边界区对于基因组的结构稳定以及基因组大小起到关键作用。对此, 基于叶绿体基因组注释数据, 文中比较了 8 个

苹果属物种的 SC 区和 IR 区的边界扩张与收缩情况 (以单子叶模式植物水稻和双子叶模式植物拟南芥作参考)。如图 7 所示, 虽然苹果属叶绿体基因组长度变化从 160 024 bp (山荆子) 到 160 239 bp (西蜀海棠), 但是它们的 IR 区大小差异较小 (26 306–26 362 bp)。基因 *rps19* 在 8 个物种中均分布于 LSC 和 IRB, 并且在 LSC-IRB 边界都为 159 bp (LSC) 和 120 bp (IRB) 的

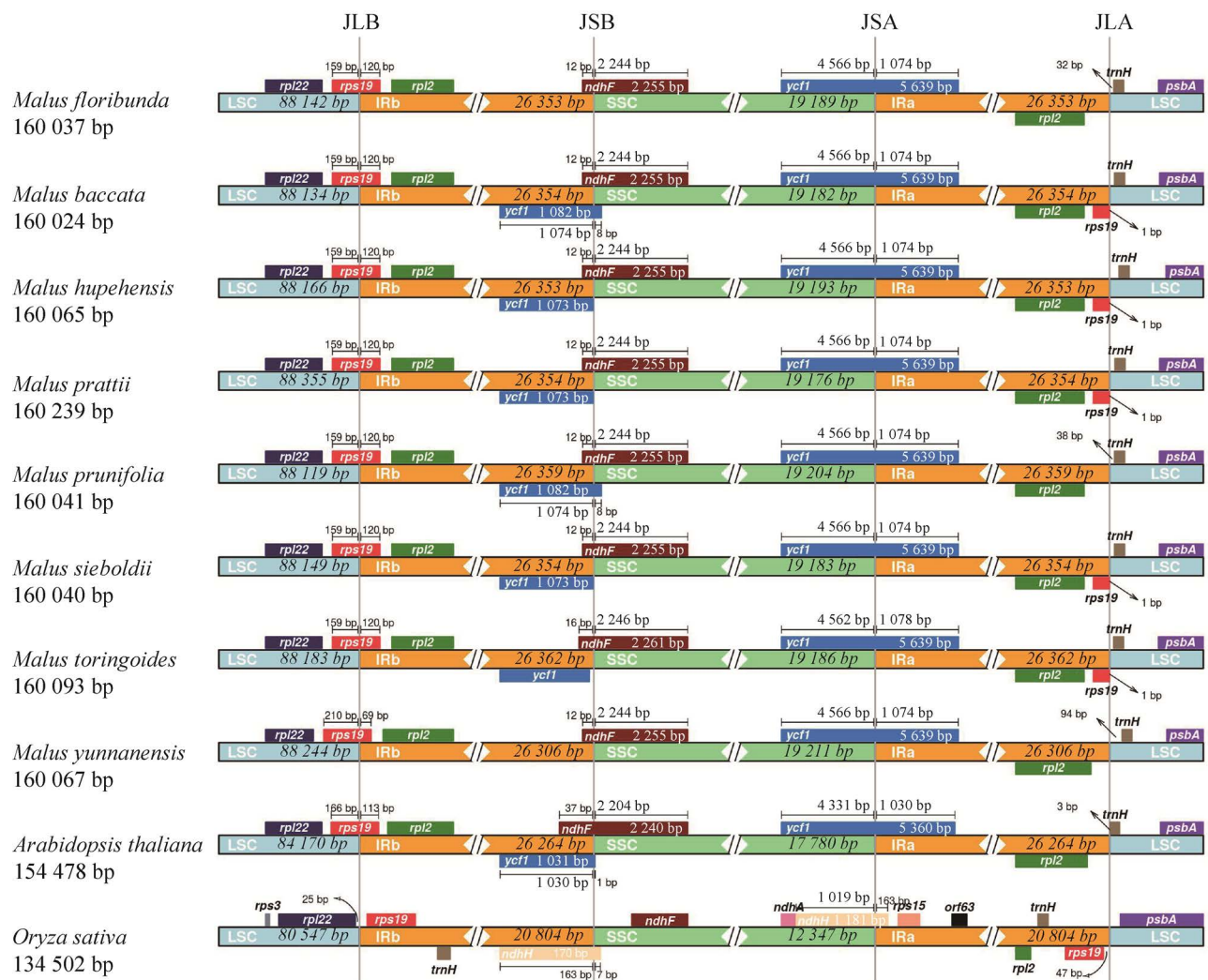


图 7 多花海棠叶绿体基因组边界分析 JLB、JSB、JSA、JLA 代表叶绿体基因组边界区域 4 个连接位点

Figure 7 Boundary analysis of chloroplast genome in *Malus floribunda*. JLB, JSB, JSA and JLA represent four junction sites in the boundary region of chloroplast genome.

组合方式 (滇池海棠除外)。在 IRB 区和 SSC 区分界处, *ycf1* 基因在不同植物中有所区别, 除多花海棠和滇池海棠未注释到 *ycf1*, 湖北海棠、西蜀海棠、三叶海棠和变叶海棠的 *ycf1* 都位于 IRB 一侧, 并未跨越边界, 而山荆子和楸子的均有 8 bp 落在 SSC 区域。另外一个值得注意的基因是 *trnH*, 虽然它在 8 个苹果属植物中都位于 LSC 区, 但是它们距离 IRA 和 LSC 边界的距离是不同的, 分别有 32、38、94 bp 等。此外, 比较分析发现, 单子叶植物水稻与双子叶的拟南芥和苹果属植物相比, 边界基因在长度和位置上均存在较大差异。与苹果属植物不同, 拟南芥 JSB 连接位点的 *ndhF* 基因向 IRB 区偏移量更大, 而水稻 *ndhF* 基因则全部位于 SSC 区; 在拟南芥中, LSC 与 IRB 边界的 *rps19* 更偏向 IRB 侧, 而水稻 *rps19* 则完全落入 IRB 区。

2.4 多花海棠叶绿体基因组系统进化分析

多花海棠叶绿体基因组的组装对其生物学进化地位的阐释有着重要价值。以山楂属

(*Crataegus*) 湖北山楂 (*Crataegus hupehensis*) 作外类群, 将多花海棠和其他 7 种苹果属植物基于叶绿体基因组构建系统进化树 (图 8)。由于邻接树分支上支持值均大于 60%, 说明结果是较为可靠的。图中显示所有物种共聚为两大类, 分别是外类的山楂属以及苹果属; 苹果属又分 3 支, 其中滇池海棠单独为一支, 多花海棠、湖北海棠和变叶海棠组成一支, 西蜀海棠、楸子、山荆子和三叶海棠划分为另一进化支。另外, 根据进化距离可以看出, 多花海棠在其所在聚类支中分化较早。本文的拓扑结构与先前高源等^[9]利用叶绿体片段序列进行的苹果属遗传多样性研究结果较为相似, 同时, 李亚楠^[11]对苹果属的划分也在本研究的进化关系中得到支持。

3 讨论与结论

多花海棠是世界上种植较为普遍的苹果属栽培种, 具有较高的观赏和育种价值。本研究

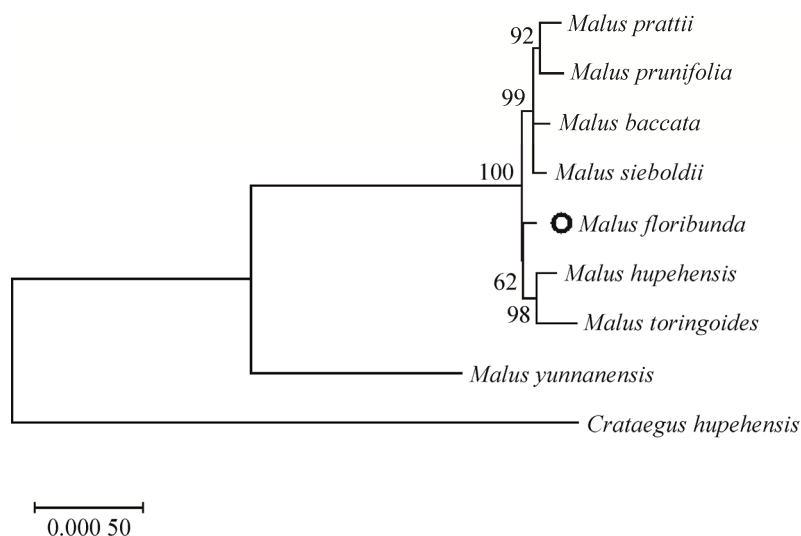


图 8 基于叶绿体基因组的多花海棠的系统进化和聚类分析 标尺代表进化距离, 多花海棠位置用黑色圆圈标出

Figure 8 Systematic evolution and cluster analysis of *Malus floribunda* based on chloroplast genome. The scale represents the evolutionary distance, and the position of *M. floribunda* is circled in black.

通过 DNA 测序获得了完整的多花海棠叶绿体基因组,与此同时,借助生物信息学技术,对其进行基因注释、序列比对、边界分析等。利用二代短读测序数据,从头组装得到一个长度为 160 037 bp 具有环状结构的多花海棠叶绿体 DNA,序列总长与同属的山荆子 (160 024 bp)^[26]、湖北海棠 (160 065 bp)^[27]、楸子 (160 041 bp)^[28]、变叶海棠 (160 093 bp)^[13]等相比是较为接近的。而与不同属的漆树科盐肤木属盐肤木 (*Rhus chinensis* Mill., 158 809 bp)^[29]、蔷薇科桃属长柄扁桃 (*Amygdalus pedunculata* Pall., 157 851 bp)^[30]、蔷薇科移 [木衣] 属长爪 [木衣] (*Docynia longiunguis* Q. Luo & J. L. Liu, 158 914 bp)^[31]等叶绿体基因组长度差别较大。

分析获知多花海棠叶绿体基因组包含 4 个保守的经典区域,分别为大单拷贝区 (88 142 bp),反向重复区 B (26 353 bp),小单拷贝区 (19 189 bp)与反向重复区 A (26 353 bp)。边界分析发现,其与苹果属的其他植物在 IR 区域长度差别不大,进一步说明反向重复区在维持叶绿体基因组的结构稳定与功能保守等方面发挥积极作用。通过对多花海棠叶绿体基因组注释,共获得 78 个蛋白编码基因,与多脉铁木 (79 个)^[32]、矮牡丹 (78 个)^[33]中的数目是接近的。经对比,苹果属的湖北海棠叶绿体基因组也具有 78 个编码基因。进一步对多花海棠叶绿体编码基因的密码子使用模式进行计算,发现密码子具有偏向 A/T 结尾的特征,并且使用 UAA 作为终止密码子的概率更高 (RSCU=1.71)。类似的结果在球花石斛^[34]、大花君子兰^[35]、杜梨^[36]中也已报道。

基于叶绿体 DNA 比较植物种间多样性和遗传变异,一直以来都是研究热点问题。目前,在苹果属中进行了大量报道。高源等借助 *trnH-psbA*、*trnS-trnG* spacer+intron、*trnT-5'trnL* 和

5'trnL-trnF 这 4 个叶绿体基因间区对新疆野苹果^[37]、山荆子^[38]和楸子^[39]等多态性变异位点进行了检测并对不同居群种质的系统演化展开探讨;李亚楠通过对 91 个叶绿体基因组构建系统发育树并结合前人分类,建议将苹果属划分为 2 亚属,8 组,25 种^[11]。在本文的研究中,通过叶绿体全基因组比对,发现 8 个苹果属植物中存在 *trnV* (UAC)-*ndhC*、*trnT* (GGU)-*psbD* 和 *trnR* (UCU)-*atpA* 等高度变异区域,此外,系统聚类关系表明多花海棠、湖北海棠和变叶海棠在进化上较为紧密且独立。

由比较基因组学分析得知,相较于单子叶模式植物水稻和双子叶模式植物拟南芥,多花海棠与苹果属内其他种序列相似性更高。同时注意到,不同物种叶绿体基因在排列顺序和位置上不尽相同,例如禾本科水稻中 SSC 与 IR 区边界处 *ycf1* 基因发生退化缺失^[40],使得 SSC 区缩小、基因偏移。这一情况与拟南芥和苹果属植物等明显不同,推测可能与物种分化有关。因此,在今后的研究中,应多加关注物种间的序列差异以及变异特性。

本研究基于高通量测序获得了多花海棠叶绿体全基因组序列,对其序列特征、基因功能、全局比对等展开分析,并借助生物信息学和比较基因组学手段,在叶绿体基因组层面比较了不同的苹果属种质的遗传多样性,可为今后深入挖掘分子标记、揭示物种进化提供参考。

REFERENCES

- [1] 周婷,沈星诚,周道建,等.海棠品种分类研究进展.园艺学报,2018,45(2):380-396.
Zhou T, Shen XC, Zhou DJ, et al. Advances in the classification of crabapple cultivars. Acta Horticult Sin, 2018, 45(2): 380-396 (in Chinese).
- [2] 李育农.苹果属植物种质资源研究.北京:中国农业出版社,2001.
Li YN. Researches of Germplasm Resources of *Malus*

- Mill. Beijing: Chinese Agriculture Press, 2001 (in Chinese).
- [3] 姜文龙, 范俊俊, 张丹丹, 等. 观赏海棠不同叶位色彩特征及特异种质挖掘. 园艺学报, 2017, 44(6): 1135-1144.
Jiang WL, Fan JJ, Zhang DD, et al. Research on the characteristics of different leaf positions of crabapple's leaf color and elite germplasm excavation. Acta Horticult Sin, 2017, 44(6): 1135-1144 (in Chinese).
- [4] 郑杨, 曲晓玲, 郭翎, 等. 观赏海棠资源谱系分析及育种研究进展. 山东农业大学学报(自然科学版), 2008, 39(1): 152-160.
Zheng Y, Qu XL, Guo L, et al. Advances on ornamental crabapple resources. J Shandong Agric Univ (Nat Sci Ed), 2008, 39(1): 152-160 (in Chinese).
- [5] 聂继云, 吕德国, 李静, 等. 22种苹果种质资源果实类黄酮分析. 中国农业科学, 2010, 43(21): 4455-4462.
Nie JY, Lü DG, Li J, et al. A preliminary study on the flavonoids in fruits of 22 apple germplasm resources. Sci Agric Sin, 2010, 43(21): 4455-4462 (in Chinese).
- [6] 孙希生, 丛佩华, 薛光荣. 德国苹果抗性育种概况. 中国果树, 2002(4): 57-58.
Sun XS, Cong PH, Xue GR. General situation of resistance breeding of apple in Germany. China Fruits, 2002(4): 57-58 (in Chinese).
- [7] Papp D, Singh J, Gadoury D, et al. New North American isolates of *Venturia inaequalis* can overcome apple scab resistance of *Malus floribunda* 821. Plant Dis, 2020, 104(3): 649-655.
- [8] 闫忠业, 姜汉平, 田颖辉. 苹果抗病虫育种研究概况. 中国果树, 2005(6): 48-49, 58.
Yan ZY, Lou HP, Tian YH. The survey overview of apple breeding for resistance to disease and insect. China Fruits, 2005(6): 48-49, 58 (in Chinese).
- [9] 高源, 王大江, 王昆, 等. 苹果属 15 个种的叶绿体 DNA 变异与遗传分化. 果树学报, 2021, 38(1): 1-12.
Gao Y, Wang DJ, Wang K, et al. Genetic divergence and population structure of chloroplast DNA in fifteen species of *Malus* Mill. J Fruit Sci, 2021, 38(1): 1-12 (in Chinese).
- [10] 李慧峰. 泰沂山区苹果属植物系统学研究[D]. 沈阳: 沈阳农业大学, 2012.
Li HF. Studies on the taxonomy of the genus *Malus* Mill.(Rosaceae)of Taiyi-mountains[D]. Shenyang: Shenyang Agricultural University, 2012 (in Chinese).
- [11] 李亚楠. 苹果属(*Malus* Mill.)种间关系的分子系统学研究[D]. 哈尔滨: 东北林业大学, 2021.
Li YN. Molecular systematics study on interspecific relationship of *Malus* Mill[D]. Harbin: Northeast Forestry University, 2021 (in Chinese).
- [12] Wang S, Cai XR, Zhang B, et al. Characterization of the complete chloroplast genome sequence of *Malus toringoides* (Rosaceae). Mitochondrial DNA B Resour, 2020, 5(3): 2603-2604.
- [13] Li YN, Liu YL, Wu P, et al. The complete chloroplast genome sequence of *Malus toringoides* (Rosaceae). Mitochondrial DNA B Resour, 2020, 5(3): 2787-2789.
- [14] Jin JJ, Yu WB, Yang JB, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol, 2020, 21(1): 241.
- [15] Wick RR, Schultz MB, Zobel J, et al. Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics, 2015, 31(20): 3350-3352.
- [16] Qu XJ, Moore MJ, Li DZ, et al. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods, 2019, 15: 50.
- [17] 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. BMC Genomics, 2012, 13: 715.
- [18] 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.
- [19] Beier S, Thiel T, Münch T, et al. MISA-web: a web server for microsatellite prediction. Bioinformatics, 2017, 33(16): 2583-2585.
- [20] Kurtz S, Choudhuri JV, Ohlebusch E, et al. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res, 2001, 29(22): 4633-4642.
- [21] Sharp PM, Li WH. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res, 1987, 15(3): 1281-1295.
- [22] Frazer KA, Pachter L, Poliakov A, et al. VISTA: computational tools for comparative genomics. Nucleic Acids Res, 2004, 32(Web Server issue): W273-W279.
- [23] Amirouf A, Hyvönen J, Pocza P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics, 2018, 34(17): 3030-3031.
- [24] Bi GQ, Mao YX, Xing QK, et al. HomBlocks: a multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block

- searching. *Genomics*, 2018, 110(1): 18-22.
- [25] Kumar S, Stecher G, Li M, et al. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*, 2018, 35(6): 1547-1549.
- [26] Liu BB, Hong DY, Zhou SL, et al. Phylogenomic analyses of the *Photinia* complex support the recognition of a new genus *Phippsiomeles* and the resurrection of a redefined *Stranvaesia* in Maleae (Rosaceae). *J Syst Evol*, 2019, 57(6): 678-694.
- [27] Zhang X, Rong CX, Qin L, et al. Complete chloroplast genome sequence of *Malus hupehensis*: genome structure, comparative analysis, and phylogenetic relationships. *Molecules*, 2018, 23(11): 2917.
- [28] Bao L, Li K, Liu Z, et al. Characterization of the complete chloroplast genome of the Chinese crabapple *Malus prunifolia* (Rosales: Rosaceae: Maloideae). *Conserv Genet Resour*, 2016, 8(3): 227-229.
- [29] 左瑞华, 蒋平, 孙传伯, 等. 基于全基因组从头测序技术盐肤木叶绿体基因组的测序分析. *生物工程学报*, 2020, 36(4): 772-781.
Zuo RH, Jiang P, Sun CB, et al. Analysis of the chloroplast genome characteristics of *Rhus chinensis* by *de novo* sequencing. *Chin J Biotech*, 2020, 36(4): 772-781 (in Chinese).
- [30] 段春燕, 张凯, 段义忠. 长柄扁桃叶绿体基因组比较. *生物工程学报*, 2020, 36(12): 2850-2859.
Duan CY, Zhang K, Duan YZ. Comparison of complete chloroplast genome sequences of *Amygdalus pedunculata* Pall. *Chin J Biotech*, 2020, 36(12): 2850-2859 (in Chinese).
- [31] 李连星, 彭劲谕, 王大玮, 等. 长爪柊[木衣]叶绿体基因组特征系统发育及密码子偏好性分析. *生物工程学报*, 2022, 38(1): 328-342.
Li LX, Peng JY, Wang DW, et al. Chloroplast genome phylogeny and codon preference of *Docynia longiunguis*. *Chin J Biotech*, 2022, 38(1): 328-342 (in Chinese).
- [32] 蒋明, 柯世省, 王军峰. 多脉铁木叶绿体基因组的序列特征和系统发育. *林业科学*, 2020, 56(5): 60-68.
Jiang M, Ke SS, Wang JF. Characterization and phylogenetic analysis of *Ostrya multinervis* chloroplast genome. *Sci Silvae Sin*, 2020, 56(5): 60-68 (in Chinese).
- [33] 周晓君, 张凯, 彭正锋, 等. 矮牡丹与芍药属其他5个种叶绿体基因组特征比较. *林业科学*, 2020, 56(4): 82-88.
Zhou XJ, Zhang K, Peng ZF, et al. Comparative analysis of chloroplast genome characteristics between *Paeonia jishanensis* and other five species of *Paeonia*. *Sci Silvae Sin*, 2020, 56(4): 82-88 (in Chinese).
- [34] 朱斌, 甘晨晨, 王洪程. 球花石斛(*Dendrobium thyrsiflorum*)叶绿体基因组特征及亲缘关系解析. *生物技术通报*, 2021, 37(5): 38-47.
Zhu B, Gan CC, Wang HC. Characteristics of the complete chloroplast genome of *Dendrobium thyrsiflorum* and its phylogenetic relationship analysis. *Biotechnol Bull*, 2021, 37(5): 38-47 (in Chinese).
- [35] 郑祎, 张卉, 王钦美, 等. 大花君子兰叶绿体基因组及其特征. *园艺学报*, 2020, 47(12): 2439-2450.
Zheng Y, Zhang H, Wang QM, et al. Complete chloroplast genome sequence of *Clivia miniata* and its characteristics. *Acta Horti Sin*, 2020, 47(12): 2439-2450 (in Chinese).
- [36] 李泳潭, 张军, 黄亚丽, 等. 杜梨叶绿体基因组分析. *园艺学报*, 2020, 47(6): 1021-1032.
Li YT, Zhang J, Huang YL, et al. Analysis of chloroplast genome of *Pyrus betulaefolia*. *Acta Horti Sin*, 2020, 47(6): 1021-1032 (in Chinese).
- [37] 高源, 王大江, 王昆, 等. 新疆野苹果叶绿体 DNA 变异与遗传进化分析. *植物遗传资源学报*, 2020, 21(3): 579-587.
Gao Y, Wang DJ, Wang K, et al. Chloroplast DNA variation and genetic evolution of *Malus sieversii* (Ledeb.) M. Roem. *J Plant Genet Resour*, 2020, 21(3): 579-587 (in Chinese).
- [38] 高源, 王大江, 王昆, 等. 基于叶绿体 DNA 变异的山荆子种质遗传多样性和系统演化. *中国农业科学*, 2020, 53(3): 600-611.
Gao Y, Wang DJ, Wang K, et al. Genetic diversity and phylogenetics of *Malus baccata* (L.) borkh revealed by chloroplast DNA variation. *Sci Agric Sin*, 2020, 53(3): 600-611 (in Chinese).
- [39] 高源, 王大江, 王昆, 等. 基于叶绿体 DNA 分析的楸子种质遗传多样性研究. *园艺学报*, 2020, 47(5): 853-863.
Gao Y, Wang DJ, Wang K, et al. Genetic diversity of *Malus prunifolia* germplasms based on chloroplast DNA analysis. *Acta Horti Sin*, 2020, 47(5): 853-863 (in Chinese).
- [40] 李裕华, 任永康, 赵兴华, 等. 禾本科主要农作物叶绿体基因组研究进展. *生物技术通报*, 2020, 36(11): 112-121.
Li YH, Ren YK, Zhao XH, et al. Research progress on chloroplast genome of major gramineous crops. *Biotechnol Bull*, 2020, 36(11): 112-121 (in Chinese).

(本文责编 陈宏宇)