華中震業大学

# HUAZHONG AGRICULTURAL UNIVERSITY

# 博士学位论文 Ph D DISSERTATION

温州蜜柑线粒体基因组组装及胞质杂种'华柚2号'

# 细胞质雄性不育基因发掘

ASSEMBLY OF SATSUMA MANDARIN MITOCHONDRIAL GENOME AND IDENTIFICATION OF CANDIDATE CYTOPLASMIC MALE STERILITY ASSOCIATED GENES IN A SOMATIC CYBRID OF HB PUMMELO

研究生:	张 帅
CANDIDATE:	ZHANG SHUAI
学 号: STUDENT NO.:	2016305010036
专业:	果树学
MAJOR:	POMOLOGY
导 师: SUPERVISOR:	郭文武 教授 PROFESSOR GUO WENWU

中国 武汉 WUHAN, CHINA 二〇二〇年十二月

DECEMBER, 2020

分类号

密级

# 华中农业大学博士学位论文

# 温州蜜柑线粒体基因组组装及胞质杂种'华柚2号' 细胞质雄性不育基因发掘

### Assembly of Satsuma mandarin mitochondrial genomes and identification of candidate cytoplasmic male sterility associated genes in a somatic cybrid of HB pummelo

博士研究生:	张帅	
学 号:	2016305	010036
指导教师:	郭文武	教授
指导小组:	邓秀新	教授
	程运江	教授
	徐娟	教授
	徐强	教授
	伍小萌	副教授
	解凯东	讲 师

专业:果树学 研究方向:基因工程与分子生物学 获得学位名称:农学博士 获得学位时间:2020年12月 华中农业大学园艺林学学院

二〇二〇年十二月

# 目 录

摘要i
ABSTRACT iii
缩略词表vi
第一章 前言1
1 课题的提出1
2 前人研究进展
2.1 高等植物线粒体基因组2
2.1.1 线粒体细胞生物学特征及起源
2.1.2 植物线粒体基因组结构特点 2
2.1.3 植物线粒体基因组组成、转录及 RNA 编辑4
2.2 植物线粒体基因组高通量测序5
2.3 植物体细胞杂种线粒体基因组7
2.3.1 植物体细胞杂交技术的意义
2.3.2 体细胞杂种线粒体基因组遗传分析
2.4 植物雄性不育研究9
2.4.1 细胞核雄性不育9
2.4.2 细胞质雄性不育基因11
2.4.3 植物细胞质雄性不育分子机理 12
2.5 柑橘雄性不育与果实无核15
3 本研究的目的和内容16
第二章 '国庆1号'温州蜜柑和 HB 柚线粒体基因组测序与比较分析17
1 引言17
2 材料和方法
2.1 柑橘愈伤组织及黄化幼苗培养18
2.2 柑橘线粒体提取19
2.3 线粒体 DNA、总 DNA Illumina 及总 DNA PacBio 测序 DNA 文库构建与高
通量测序
2.4 LncRNA 文库构建与高通量测序 21

2.5	线粒体 de novo 组装策略	22
2.6	RNA 编辑位点鉴定	23
2.7	细胞内基因水平转移序列鉴定	23
2.8	基因注释	23
2.9	叶绿体基因组 de novo 组装策略	23
2.10	0 线粒体及叶绿体基因组共线性及变异(SNP/InDel)分析	24
3 结果.		24
3.1	'国庆1号'温州蜜柑和 HB 柚线粒体基因组 Illumina 测序及组装	24
	3.1.1 柑橘线粒体提取及鉴定	24
	3.1.2 线粒体完整性检测	25
	3.1.3 线粒体 DNA 检测	26
	3.1.4 线粒体基因组测序和拼接	26
	3.1.5 '国庆1号'温州蜜柑线粒体基因组结构	28
	3.1.6 HB 柚线粒体基因组结构	29
	3.1.7 '国庆1号'温州蜜柑和HB 柚线粒体基因组序列准确性分析	30
3.2	柑橘线粒体基因组三代测序组装策略	31
	3.2.1 柑橘线粒体基因组校正前后数据比较	33
3.3	'国庆1号'温州蜜柑和 HB 柚线粒体基因组 RNA 编辑	37
3.4	'国庆1号'温州蜜柑和 HB 柚线粒体基因组重复序列	. 39
3.5	'国庆1号'温州蜜柑和 HB 柚线粒体基因组变异和共线性分析	40
3.6	'国庆1号'温州蜜柑和 HB 柚叶绿体基因组比较	42
	3.6.1 '国庆1号'温州蜜柑和 HB 柚叶绿体基因组信息	42
	3.6.2 '国庆1号'温州蜜柑和 HB 柚叶绿体基因组变异分析	44
3.7	'国庆1号'温州蜜柑和 HB 柚线粒体和叶绿体序列转移	45
4 讨论.		47
4.1	柑橘线粒体 DNA 提取	47
4.2	柑橘线粒体基因组组装	48
4.3	柑橘线粒体基因组结构	49
第三章 HB	柚胞质杂种基因组来源分析及候选 CMS 基因发掘与功能解析	51
1 引言		51

2 材料和方法	. 52
2.1 材料	. 52
2.2 方法	. 53
2.2.1 HB 柚胞质杂种人工授粉	. 53
2.2.2 LncRNA 转录本拼接及预测	. 53
2.2.3 HB 柚胞质杂种及其亲本与甜橙参考基因组变异分析	. 54
2.2.4 柑橘不育材料特异 ORFs 鉴定	. 54
2.2.5 农杆菌介导的拟南芥遗传转化	. 55
2.2.6 拟南芥阳性苗筛选	. 55
2.2.7 目的蛋白原核表达	. 56
2.2.8 转录本 DN16978 诱饵载体构建	. 56
2.2.9 转录本 DN16978 互作蛋白筛选	. 56
2.2.10 酵母互作蛋白点对点验证	. 57
2.2.11 LUC 荧光检测实验	. 58
2.2.12 orf88 与 CsNADH 亚细胞定位	. 58
2.2.13 NAD <sup>+</sup> 与 NADH 含量测定	. 59
3 结果	. 60
3.1 HB 柚胞质杂种核质基因组与其融合双亲比较分析	. 60
3.1.1 HB 柚胞质杂种线粒体基因组结构	. 60
3.1.2 HB 柚胞质杂种与其融合双亲线粒体基因组比较	. 61
3.1.3 HB 柚胞质杂种线粒体基因组 16.8 kb 重复序列验证	. 63
3.1.4 HB 柚胞质杂种与其融合双亲叶绿体基因组及核基因组比较	. 65
3.1.5 冰糖橙胞质杂种线粒体基因组	. 67
3.2 HB 柚胞质杂种及'国庆1号'温州蜜柑不育线粒体基因筛选	. 68
3.2.1 HB 柚胞质杂种及其融合双亲线粒体转录本拼接及功能注释	. 68
3.2.2HB 柚胞质杂种、'国庆1号'温州蜜柑与 HB 柚 MSS 区鉴定	. 69
3.2.3 HB 柚胞质杂种和'国庆1号'温州蜜柑线粒体特异 ORF 预测	. 71
3.2.4 CMS 候选基因	. 72
3.2.5 CMS 候选基因在 HB 柚胞质杂种及其亲本中的表达	. 75
3.3 CMS 不育候选转录本功能验证	. 76

3.3.1 线粒体不育候选转录本异源表达拟南芥76
3.3.2 转录本 DN16978 核互作蛋白筛选77
3.3.3 CMS 候选 ORFs 亚细胞定位及跨膜结构域预测
3.3.4 orf88 毒性检验79
3.3.5 点对点验证 orf88 与 CsNADH 蛋白互作 80
3.3.6 LUC 荧光验证 orf88 与 CsNADH 蛋白互作 80
3.3.7 Orf88 与 CsNADH 亚细胞定位分析 81
3.3.8 HB 柚胞质杂种及其融合双亲花药 NAD <sup>+</sup> 、NADH 含量测定 82
3.3.9 Real-time PCR 检测 HB 柚胞质杂种及其融合双亲小孢子发育过程
CsNADH 表达
4 讨论
4.1 柑橘胞质杂种线粒体基因组发生融合84
4.2 HB 柚胞质杂种胞质雄性不育基因筛选
4.3 Orf88 可能的调控机理
后续研究设想
参考文献
附录 I 线粒体提取液配103
附录Ⅱ部分实验操作步骤104
附录 Ⅲ本研究中所使用到的引物序列106
附录 IV 本研究中使用的特异性 MSS 引物序列信息109
附录 V CMS 特异 ORF 信息111
附录 VI 扩增基因序列117
附录 VII 作者简介及博士学习期间发表论文119
致谢120

#### 摘要

果实无核是柑橘的优良性状, 雄性不育是柑橘无核的主要原因。温州蜜柑雄 性不育是线粒体基因与细胞核基因相互作用引起的雄性不育。以温州蜜柑为母本 进行有性杂交, 创制了许多无核柑橘新品种。但是, 柑橘线粒体基因组信息和细 胞质雄性不育(CMS)机制还不清楚。前人利用细胞融合技术, 将'国庆1号' 温州蜜柑雄性不育细胞质转移给 HB 柚, 获得 HB 柚胞质杂种, 是研究细胞质雄 性不育的理想试材。本研究以 HB 柚胞质杂种及其融合双亲为试材, 通过线粒体 基因组及全基因组测序, 组装和注释了高质量 HB 柚胞质杂种及其双亲线粒体基 因组; 从基因组水平确定了 HB 柚胞质杂种核、质基因组来源; 通过不育材料 HB 柚胞质杂种和'国庆1号'温州蜜柑与可育材料 HB 柚线粒体基因组和转录 组比较分析, 在不育材料中鉴定到 57 个与线粒体编码基因共转录的 CMS 特异 ORFs, 并挑选 orf88 作为候选基因, 对其潜在的调控细胞质雄性不育的机制进行 研究。主要研究结果如下:

1. '国庆1号'温州蜜柑和 HB 柚线粒体基因组基本信息

以'国庆1号'温州蜜柑愈伤组织和 HB 柚黄化苗为试材,提取线粒体 DNA; 用'国庆1号'温州蜜柑愈伤组织和 HB 柚叶片提取总 DNA。分别采用 Sanger、 Illumina 和 PacBio 三种测序技术,完整组装了'国庆1号'温州蜜柑和 HB 柚线 粒体基因组;大小分别为 521,559 bp 和 518,274 bp,GC 含量分别为 45.06%和 45.14%,线粒体编码蛋白数分别为 36 和 35 个。'国庆1号'温州蜜柑和 HB 柚 线粒体基因组有 28 个重排区域,相似片段总长 460,067 bp。同时,RNA 编辑位 点和编辑效率在这两个品种间具有物种特异性。

2. HB 柚胞质杂种细胞质和核基因组来源

HB 柚胞质杂种线粒体基因组大小为 538,434 bp, 与'国庆1号'温州蜜柑 线粒体基因组几乎一致,并且存在一个额外的 16.8 kb 同源重组片段; 与 HB 柚 线粒体基因组同源的长度为 472.3 kb (87.7%)。共线性分析结果显示, HB 柚胞 质杂种线粒体基因组 96.9%的序列与'国庆1号'温州蜜柑具有共线性,剩余 3.1% 是 HB 柚胞质杂种线粒体基因组的重复序列。HB 柚胞质杂种线粒体基因组与 HB 柚线粒体基因组有 29 个重排区域,其中 28 个重排区域与'国庆1号'温州蜜柑 和 HB 柚的一致。HB 柚胞质杂种、'国庆1号'温州蜜柑及 HB 柚叶绿体基因组

i

大小分别为160,423 bp、160,159 bp、160,590 bp,GC含量均为38.47%。HB 柚 胞质杂种及其亲本叶绿体基因组之间相似性达99%;HB 柚 胞质杂种与 HB 柚 叶 绿体基因组序列有11个 SNPs 和42个 InDels 变异位点,与'国庆1号'温州蜜 柑 叶绿体基因组序列有384个 SNPs 和120个 InDels 变异位点。此外,HB 柚 胞 质杂种及其亲本与甜橙参考基因组 SNP 和 InDel 统计结果显示,HB 柚 胞质杂种 及其亲本核基因组与甜橙基因组共有946,383个 SNPs 和 200,661个 InDels。其中,HB 柚 胞质杂种 99.8%的 SNPs 位点和96.9%的 InDels 位点与HB 柚一致。因此,HB 柚 胞质杂种叶绿体和细胞核基因组应来自HB 柚。

3. HB 柚胞质杂种胞质雄性不育候选基因筛选与功能验证

基于不育系 HB 柚胞质杂种与'国庆1号'温州蜜柑和可育系 HB 柚转录组 和线粒体基因组比较分析,在 HB 柚胞质杂种和'国庆1号'温州蜜柑线粒体基 因组中分别鉴定到 21 条和 20 条 CMS 线粒体特异序列,89 和 70 个 CMS 特异 ORFs。从这些特异 ORFs 中鉴定到和 rps1 3'端 138 bp 互补序列在同一个嵌合转 录本 DN16978 的 3 个特异 ORFs (orf55、orf59 和 orf88)。将转录本 DN16978 构 建超量表达载体转入拟南芥,成功获得拟南芥转基因系,并表现出一系列不育表 型。跨膜结构域和亚细胞定位预测这三个特异 ORFs 结果表明 orf88 编码蛋白含 有跨膜结构域,且自身含有线粒体定位信号肽,认为 orf88 可能是 HB 柚胞质杂 种 CMS 基因。利用酵母双杂交筛库技术获得 orf88 互作蛋白 CsNADH (NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12)。酵母双杂点对点实验 及 LUC 荧光检测实验表明 orf88 与 CsNADH 基因互作。亚细胞定位显示 orf88 和 CsNADH 均定位于线粒体。

综上所述,本研究组装了 HB 柚胞质杂种及其融合双亲的线粒体基因组;发掘了可能与 HB 柚胞质杂种细胞质雄性不育相关的基因 orf88,鉴定了与 orf88 互 作的基因 CsNADH,并对 orf88 调控细胞质雄性不育的内在机制进行了讨论。 关键词:柑橘;线粒体基因组;胞质杂种;细胞质雄性不育;线粒体型特异序列; orf88; CsNADH

ii

# ABSTRACT

Seedlessness is highly desirable traits for citrus. Male sterility is mainly responsible for seedlessness in widely-cultivated citrus varieties. Male sterility is caused by mitochondrial genes with coupled nuclear genes in Satsuma mandarin. Sexual crosses that utilized germplasm with Satsuma mandarin as the female parent yielded new seedless varieties of citrus. However, the mitochondrial genome infromation and mechanisms of cytoplasmic male sterility (CMS) remain to be elucidated in citrus. In our previous breeding program, male sterile and seedless cybrid pummelo (G1+HBP) was regenerated by fusing protoplasts from seedless 'Guoqing No. 1' Satsuma mandarin (G1, the callus parent) and seedy 'Hirado Buntan' pummelo (HBP, C. grandis (L.) Osbeck, the leaf parent), which is an ideal material for cytoplasmic male sterility analysis. The high-quality mitochondrial genomes of G1+HBP and its fusion parents were sequenced and assembled, and determined the origins of the nuclear and cytoplasmic genomes of the cybrid. Compared with the fertile HBP mitochondrial genome and transcriptome, 57 ORFs that co-transcribed with the same mitochondrial genes in the cybrid and G1 were identified as the CMS-specific ORFs. The orf 88 was then selected as a candidate gene, and its potential mechanism of regulating CMS was studied. The main results are as follows: 1. Characterization of the G1 and HBP mitochondrial genome

Purified mtDNA of G1 was extracted from a fresh subculture of embryonic callus while mtDNA of HBP was purified from 45-d-old etiolated seedlings. HBP total DNA was isolated from frozen leaves using a simplified CTAB protocol, and G1 total DNA was isolated from embryonic callus. The mitochondrial genomes of G1 and HBP were assembled into circular molecules of 521,599 bp and 518,274 bp with GC contents of 45.06% and 45.14% by using the total DNA PacBio Sequel sequencing reads and mtDNA sequencing reads, respectively. The mitochondrial genome of citrus encodes 35 to 36 proteins, three ribosomal RNAs and 21 tRNAs. A total of 460,067 bp (88.8%) of HBP is homologous to the mitochondrial genome from G1, and at least 28 apparent rearrangements in both the syntenic order and the

relative direction of the ORFs. In addition, species specificity of RNA sites and efficiency existed between G1 and HBP.

#### 2. Origins of the cytoplasmic and nuclear genomes of the cybrid pummelo G1+HBP

The mitochondrial genome of the cybrid (G1+HBP) was assembled into a circle of 538,434 bp. The cybrid inherited the entire mitochondrial genome of the G1 with an extra recombinant fragment of 16.8 kb, and 471.9 kb (87.6%) of the cybrid sequence is homologous to the HBP sequences. The synteny analysis showed that up to 96.9% the mitochondrial genome sequence from the cybrid is continuously co-linear with the mitochondrial genome sequence from G1; the remaining 3.1% of the mitochondrial genome sequence from the cybrid is an extra homologous recombinant fragment; and at least 29 rearrangements in the cybrid relative to HBP. A total of 28 of these rearrangements were identical to the rearrangements we observed in G1 relative to HBP. The chloroplast genomes of the cybrid, G1 and HBP were assembled into circular molecules of 160,423, 157,088 and 160,590 bp, respectively. Each of these genomes has a GC content of 38.47%. The chloroplast genome sequence similarity was >99% among the cybrid and its parents, and the cybrid had 11 SNPs and 42 InDels relative to the chloroplast genome sequence of HBP, which is much less than the 384 SNPs and 120 InDels relative to the chloroplast genome sequence of G1. Total DNA of the cybrid and its parents were sequenced using the Illumina platform to identify SNPs and InDels relative to the sweet orange genome. A total of 946,383 SNPs and 200,661 InDels relative to the sweet orange genome are shared by the parents and the cybrid, which are unidentical in parents. Among these polymorphisms, 99.8% of the SNPs and 96.9% of the InDels are shared by the genomic sequences of HBP and the cybrid. These data indicate that the chloroplast and nuclear genomes of the cybrid was inherited from HBP.

3. Screening and function verification of candidate CMS gene of the cybrid pummelo G1+HBP.

Compared with the fertile HBP mitochondrial genome and transcriptome, 21 and 20 CMS-specific mitotype-specific sequences (MSSs), and 89 and 70 CMS-specific ORFs were identified in the cybrid and G1 mitochondrial genomes, respectively. We

identified two reverse CMS-specific ORFs (*orf59* and *orf88*) and one forward CMS-specific ORF (*orf55*) co-transcribed with the 138 bp reverse-complementary sequence from the 3' end of *rps1* ( $\psi rps1$ ) on one 1705 bp transcript *DN16978* in the cybrid and G1. The transcript *DN16978* was ectopically expressed in Arabidopsis and transgenic lines shown typical male stertile. Transmembrane helices and subcellular location prediction shown the deduced amino acid sequence of *orf88* has mitochondrial localization signal peptide and a membrane-spanning helice, indicating that the *orf88* was a mitochondrial-anchored protein and more likely to the causal gene for CMS induction in citrus. The candidate protein CsNADH, subunit 12 of the NADH dehydrogenase 1 alpha sub complex, was identified to interact with DN16978 by yeast two hybrid screening and interact with orf88 validated by point to point and luciferase complementation imaging assays verification. The results of subcellular localization showed that orf88 and CsNADH were localized to the mitochondria.

In summary, the mitochondrial genomes of seedless cybrid pummelo (G1+HBP) and its parents were sequenced and assembled. A chimeric mitochondrial protein orf88 which interacts with the CsNADH in the G1+HBP was then verified; moreover, mechanism of orf88 for regulating CMS in the somatic cybrid pummelo was also discussed.

**Keywords:** Citrus; Mitochondrial genome; Cybrid; Cytoplasmic male sterility; Mitotype-specific sequence; *orf88*; *CsNADH* 

# 缩略词表

# Abbreviations

缩略符号	英文全称	中文含义	
Abbr.	English full name	Chinese full name	
cDNA	Complementary DNA	互补 DNA	
CMS	Cytoplasmic male sterility	细胞质雄性不育	
CpDNA	Chloroplast DNA	叶绿体 DNA	
G1	Citrus unshi. Marc. Cv. Guoqing	'国庆1号'温州蜜柑	
G1 + BTC	The cybrid between G1 and 'Bingtang Orange'	冰糖橙胞质杂种	
G1 + HBP	The cybrid between G1 and 'Hirado Buntan' pummelo	HB 柚胞质杂种	
HBP	'Hirado Buntan' pummelo	HB 柚	
HGT	Horizontal gene transfer	基因水平转移	
IGT	Intracellular gene transfer	细胞内基因转移	
InDel	Indel/Deletion	插入/缺失	
IR	Inverted repeat	反向重复区	
LB	Lauria-Bertani medium	LB 培养基	
lncRNA	Long non-coding RNA	长链非编码 RNA	
LSC	Large single copy region	长单拷贝区	
LUC	Luciferase	萤火虫荧光素酶	
mCherry	mCherry fluorescence protein	mCherry 红色荧光蛋白	
Mitogenome	Mitochondrial genome	线粒体基因组	
MSS	Mitotype-specific sequence	线粒体型特异序列	
MT	Murashige and Tucker medium	MT 培养基	
mtDNA	Mitochondrial DNA	线粒体 DNA	
ORF	Open reading frame	开放读码框	
PCR	Polymerase chain reaction	聚合酶链式反应	
qRT-PCR	Quantitive Real Time PCR	荧光实时定量 PCR	
rpm	Revolutions per minute	每分钟转数	
RT-PCR	Reverse transcription PCR	反转录 PCR	
SNP	Single nucleotide polymorphism	单核苷酸多态性	
SSC	Small single copy region	短单拷贝区	
YFP	Yellow fluorescence protein	黄色荧光蛋白	

# 第一章 前言

### 1课题的提出

柑橘在我国有着悠久的栽培历史和丰富的种质资源。2018 年,我国柑橘栽 培面积和产量超越苹果,成为我国第一大水果(国家统计局 2019)。无核是柑橘 果实的优良性状,也是柑橘育种的首要目标。柑橘具有童期长、雌/雄性败育、 珠心胚干扰、花期不育和自交不亲和等障碍,很难通过有性杂交方式进行柑橘雄 性不育育种。利用原生质体融合技术进行柑橘体细胞杂交可以很大程度上解决这 些问题。柑橘体细胞杂交技术能将杂交双亲遗传信息进行融合,能够转移胞质基 因组控制的优良性状,是创制柑橘新种质的一种快速、安全和有效的方法(Guo et al 2013)。

细胞质雄性不育是不能产生有功能性花粉的母性遗传现象(Sandhu et al 2007)。水稻、油菜等大田作物,通过三系配套,实现利用细胞质雄性不育系进行杂交育种。然而,柑橘遗传背景复杂,难以通过传统有性杂交方法,获得与细胞质雄性不育相对应的保持系和恢复系品种。因此,通过体细胞杂交技术转移雄性不育胞质基因组是创造柑橘无核新品种的可行方法,为柑橘雄性不育机理研究提供了理想试材。温州蜜柑是典型核质互作引起的雄性不育,即细胞质雄性不育(Yamamoto et al 1997)。本课题组据此通过细胞工程技术实现了温州蜜柑雄性不育胞质基因组向有核品种的转移,获得了'国庆1号'温州蜜柑(简称 G1)胚性愈伤组织原生质体与 HB 柚(简称 HBP)叶肉原生质体融合再生的二倍体 HB 柚胞质杂种,定名'华柚2号'(简称 G1+HBP),并获植物新品种权(品种权号: CNA20101034.2)(郭文武等 2019)。

HB 柚胞质杂种新品种'华柚2号'已稳定开花、结果多年。本实验室前期 通过对 HB 柚胞质杂种进行细胞学观察,确定了 HB 柚胞质杂种雄性不育关键时 期为雄蕊原基小孢子减数分裂时期, RNA-seq 及 iTRAQ 分析表明, HB 柚胞质 杂种大量核基因表达受到线粒体基因影响(郑蓓蓓 2015)。本研究以 HB 柚胞质 杂种及其融合双亲('国庆 1 号'温州蜜柑、HB 柚)为试材,利用二代和三代 测序技术组装与 Sanger 测序技术填补 gap 相结合的策略,拼接获得 HB 柚胞质 杂种及其融合双亲线粒体基因组,并构建温州蜜柑和柚线粒体参考基因组;比较

分析不育材料(HB 柚胞质杂种、'国庆1号'温州蜜柑)相比可育材料(HB 柚) 线粒体特异序列和特异 ORFs,结合不育材料线粒体转录组信息,筛选与 CMS 有关的候选 ORFs,并对其进行功能验证,以期深入理解 HB 柚胞质杂种胞质雄 性不育调控机理,为柑橘无核育种提供理论依据。

# 2前人研究进展

#### 2.1 高等植物线粒体基因组

#### 2.1.1 线粒体细胞生物学特征及起源

线粒体(mitochondria)是真核生物细胞的半自主细胞器,能够合成三磷酸 腺苷(ATP),为真核生物生命活动提供能量。1890年 Altanman 首次观察到细胞 内存在线粒体,并认为线粒体是一种细胞内共生菌。1898年,Benda 观察到细胞 中的线粒体有两种形式,即线状和颗粒状,将其命名为 mitochondrion,并且沿 用至今(Ernster and Schatz 1981)。随着显微镜技术的发展,研究人员观察到大 多数物种细胞中线粒体结构是直径为0.5~1 µm,长度在1~5 µm 的圆形或短杆状。

目前,人们普遍认为内共生学说是线粒体起源理论。原始真核细胞把线粒体 祖先原线粒体菌吞噬,形成寄主与宿主的共生关系(魏炽炬 2018)。这种共生关 系通过原始真核细胞提供营养,原线粒体产生能量的方式使其适应更多的生存环 境。在长期互利共生过程中,原始真核细胞与原线粒体遗传物质也在不断交换, 最终在真核生物细胞内形成线粒体,为宿主细胞提供生长所必须的能量物质 (Green and Reed 1998)。

#### 2.1.2 植物线粒体基因组结构特点

尽管所有真核生物各种生理活动所必须的能量物质是由细胞内线粒体提供, 但与哺乳动物相比,高等植物线粒体还参与氨基酸、脂类、维生素等与生命活动 有关物质的合成和代谢(Millar et al 2005)。1997年,被子植物拟南芥线粒体基 因组首次公布(Unseld et al 1997),随后一些大田作物和果树作物线粒体基因组 相继公布,例如:水稻(Notsu et al 2002)、油菜(Handa 2003)、玉米(Clifton et al 2004)、葡萄(Goremykin et al 2009)、苹果(Goremykin et al 2012)、猕猴桃(Liu et al 2017)等。目前,在 NCBI 细胞器基因组数据库(https://www.ncbi. nlm.nih.gov/genome/organelle/)可以找到 238 个植物线粒体基因组序列(Wu et al 2019);其中,Silenne conica 是目前已知线粒体最大的植物,大小为 11,300 kb (Millar et al 2005)。

一般而言,植物线粒体基因组主要为一个分布着全部基因的单一环状结构 (Gualberto and Newton 2017),也存在线状、分支链状、重叠或者多种并存结构 (Bendich 1993; Notsu et al 2002; Bendich 2007; Sloan 2013)。植物线粒体基因序 列中重复序列分为:大片段重复序列(>1 kb)和短片段重复序列(<1 kb)。线 粒体基因组中重复序列重组产生多种亚基因组结构是造成线粒体基因组结构多 样性的主要原因(Gualberto and Newton 2017)。在这些变异中,>1 kb 重复序列 同源重组对线粒体基因组多样性起着重要作用。相较于大片段重复序列,短片段 重复序列发生不可逆重排,产生新的重复序列或缺失原有序列,从而导致线粒体 基因组结构发生变异。因此,短重复序列对高等植物线粒体基因结构和进化起着 重要作用(Lilly and Havey 2001; Kubo and Mikami 2007; Arrieta-Montiel et al 2009; Davila et al 2011)。

植物线粒体基因组限制性酶切图谱分析表明,大多数植物线粒体基因组是一个含有整个基因组序列的环状结构,该环状结构中的一些同源序列之间重组可以 产生小的亚基因组环状分子,这些小分子与主环结构共存于植物线粒体(Jańska et al 1998)。此外,通过电子显微镜和凝胶电泳观察到线粒体基因组还存在大量 长线性 DNA 和分枝状 DNA 分子,表明植物线粒体基因组分子类型具有多样性 (Oldenburg and Bendich 1996; Backert et al 1997)(图 1.1)。



Cited from Gualberto and Newton 2017

#### 2.1.3 植物线粒体基因组组成、转录及 RNA 编辑

尽管植物不同物种线粒体基因组大小不同,但不同物种的线粒体基因组序列 同源性高,编码基因功能保守,因此不同物种植物线粒体基因数目也相近(约 51-60个),包括 30-37个保守蛋白编码基因。这些保守蛋白基因主要包括,参与 呼吸反应和代谢作用的电子传递链复合体 I-V、细胞色素 C 生物合成基因、核糖 体蛋白基因、核糖体 RNA、转运 RNA、matR、mttB 等(Kubo and Newton 2008)。 此外,植物线粒体基因组还存在未知功能 ORFs,其中一些 orf 是导致该物种细 胞质雄性不育的关键基因(Kazama et al 2016; Bhatnagar-Mathur et al 2018; Xie et al 2018)。除了功能未知 ORFs 不同外,复合体 II 亚基、核糖体蛋白基因、tRNA 及多拷贝基因数不同,是不同物种线粒体基因数目差异的主要原因(Alverson et al 2010)。比较分析不同被子植物线粒体基因组核糖体蛋白基因差异,发现植物 进化过程中,线粒体基因组通过细胞内基因转移(Intracellular gene transfer, IGT), 融合了一些源于叶绿体和核基因组的 DNA 序列,同时线粒体 DNA 序列也可以 转移到核基因组,且这一转移过程仍在进行(Kubo and Mikami 2007)。例如,在 拟南芥质体中发现编码定位于线粒体多肽基因 *rps13* 的一个拷贝,该基因来自核 基因组(Mollier et al 2002)。同时,线粒体基因组还可以通过水平基因转移 (Horizontal gene transfer, HGT) 融合外缘物种的 DNA 序列(Park et al 2015), 进而导致不同植物不同亚种间线粒体基因组结构差异。

虽然不同植物线粒体保守基因来源一致,但转运 RNA 组成相对复杂。植物 线粒体 tRNA 的起源由两部分组成:一部分是线粒体祖先基因;另一部分通过细 胞内基因转移获得叶绿体转运 RNA。即便如此,几乎所有高等被子植物需要从 细胞质摄入识别部分氨基酸的 tRNA 以进行蛋白合成,而低等苔藓类植物线粒体 基因组拥有一套完整的、足够识别所有密码子的 tRNA 基因,这可能与生存环境 有关 (Duchene and Marechal 2001; Adams et al 2002a, b)。

RNA 编辑是一种转录后水平修饰,使基因转录产生的 mRNA 核苷酸发生改变,导致其序列与模板序列不完全一致的现象,是动植物线粒体及植物叶绿体基因组的基本特征(Sun et al 2016)。早在 1989年,Covello 和 Gualberto 在植物线粒体基因中发现 RNA 编辑现象;随后,Hoch等(1991)发现叶绿体 RNA 编辑现象。植物线粒体和叶绿体 RNA 编辑有 C (胞嘧啶)到 U (尿嘧啶),U (尿嘧啶)到 C (胞嘧啶)和 A (腺嘌呤)到 I (次黄嘌呤)三种编辑类型。其中,A 到 I 编辑形式主要在叶绿体 tRNA 上发生 (Chateigner-Boutin and Small 2010)。线粒体 RNA 编辑主要为 C 到 U 的编辑,发生在基因编码区、tRNA 及非编码区;并且,RNA 编辑位点在不同物种中具有保守性(Edera et al 2018)。作为一种基因表达修正机制,RNA 编辑能将极性氨基酸转变为非极性氨基酸,改变代谢效率;也能使突变位点的密码子得到恢复或者产生新的起始或终止密码子,提高不同物种蛋白同源性(Stern et al 2010; Takenaka et al 2013)。起初,Giegé和Brennicke (1999)发现拟南芥 456 个编辑位点中有 441 个在蛋白编码区。随后,油菜和水稻线粒体中也分别发现编辑位点 427 和 491 个 (Notsu et al 2002; Handa 2003)。

#### 2.2 植物线粒体基因组高通量测序

随着测序技术发展,基因测序技术已成为现代生物学研究的必要条件(表 1.1)。科学家通过对不同材料基因组测序和序列分析,从中获得有效的基因信息。 此外,多组学联合分析可以对基因复杂多样性、物种性状变异及起源进化等问题 进行详细描述。

Sanger 于 1977 年提出了双脱氧核苷酸终止法用于 DNA 测序,即第一代测 序技术。起初,一代测序技术存在通量低、成本高等问题,难以用于基因组水平 测序。因此,科学家提出了 PCR 步移克隆和基因组鸟枪法。基于以上方法,早 期的线粒体基因组测序首先构建线粒体基因组 BAC/YAC 文库,再利用线粒体 DNA 探针从 BAC/YAC 文库挑选线粒体基因组克隆,最后采用鸟枪法进行测序 组装。由于该方法需要先分离得到线粒体 DNA,但植物细胞线粒体含量少、结 构复杂,造成线粒体提纯困难且难以获得纯度高的线粒体 DNA,并且构建 BAC/YAC 文库成本高,文库本身也存在插入片段小、不稳定及粘连等缺点,限 制了该方法的应用。

测序技术从 Sanger 测序逐步发展到双端测序,即第二代测序技术。相较于 Sanger 测序,第二代测序在降低测序成本的同时,提高了测序效率和准确性。当 前,已报道的植物线粒体基因组测序数据基本采用以第二代测序技术为主体, Sanger 测序技术辅助完成。根据二代测序 DNA 来源不同,分为以下两种策略进 行线粒体基因组测序:

1. 先对植物黄化材料线粒体 DNA 进行分离纯化,再利用二代测序技术进行 测序、组装。虽然提纯线粒体难度大且植物线粒体基因组有结构复杂、重复序列 多等因素对二代测序从头组装造成影响,但提纯后的线粒体 DNA 能有效减少质 体及核基因组干扰,利于线粒体基因组测序组装,并结合 Sanger 测序技术可以 有效进行植物线粒体基因组拼接。目前,采用这种策略完成了许多植物线粒体基 因序列测序、组装,例如:油菜(Heng et al 2014)。

2. 不分离植物材料线粒体 DNA,直接在全基因组测序基础上,分离得到与 线粒体基因组序列有关的原始 reads,随后序列比对拼接,得到线粒体基因组最 终序列。虽然这种方法避免了线粒体 DNA 提纯的困难,但由于物种间线粒体基 因组序列非编码区变异大,难以通过同源比对方法获得完整的线粒体序列。因此, 在物种本身已有核基因组序列及叶绿体基因组序列的基础上,才可以采取该策略 进行线粒体基因组研究,例如:苹果、葡萄(Goremykin et al 2009, 2012)。

随着各物种基因组深入研究,二代短片段测序难以满足测序需求,Pacific Biosciences 公司提出了单分子测序技术,即第三代测序技术。相较于前两代测序 技术,三代测序技术具有测序速度快、读长长等优势,但三代测序存在单碱基错

误率高等问题(许亚昆等 2019)。因此,可以采用三代测序技术与二代测序技术 相结合的策略对植物线粒体基因组进行研究。目前,已有一些植物线粒体基因组 研究采取该策略,例如:甘蔗(Shearman et al 2016),猕猴桃(Liu et al 2017)。

表 1.1 不同测序平台比较

引自 许亚昆等 2019

#### Table 1.1 Comparison of three generation sequencing technologies

Cited from Xu et al 2019

Technical	Principle of	Read length	Advantages	limitations
platform	sequencing			
Sanger	Chain-termi	600-1,000 bp	Long reads; high	Low throughput; high cost of
	nating		accuracy; good ability to	Sanger sample preparation;
	sequencing		deal with repetitive and	making massively parallel
			homopolymer regions	sequencing prohibitive
Roche/454	Pyrosequen	200-400 bp	Longest read lengths	Challenging sample preparation;
	cing		among the	hard to deal with
			second-generation; high	repetitive/homopolymer rgions
			throughput.	
Illumina	Sequencing	$2 \times 150 \text{ bp}$	Very high throughput.	Short reads
	by synthesis			
ABI/Solid	Sequencing	25-35 bp	High throughput; low	Long sequencing runs (days);
	by ligation		cost.	short reads, resulting in
				difficulties in subsequence data
				analysis and genome assembly.
PacBio	Sequencing	$\sim$ 1,000 bp	Long average read length;	Low accuracy; dependence on
SMRT	by		no amplification of	DNA polymerase activity.
	synthesis/D		sequencing fragments;	
	NA		longest individual reads	
	polymerase		approach 100 kb.	
Nanopore	Electronic	Maximum	Over-long read; electronic	High sequencing error
	signals	record 2.2 M	sequencing; portable.	
	sequencing/			
	exonuclease			

### 2.3 植物体细胞杂种线粒体基因组

#### 2.3.1 植物体细胞杂交技术的意义

体细胞杂交是指在一定外界条件诱导下将不同品种、种或属的原生质体融合并再生成新杂种植株的过程(Guo and Deng 2000)。这项技术可以克服传统有性杂交遇到的自交不亲和、多胚性或雌雄性不育等问题,创制的新种质为许多作物

育种提供了丰富材料(Wang et al 2013)。过去几十年来,体细胞杂交技术在柑橘 (Guo et al 2013)、小麦(Liu and Xia 2014)、茄属(Tiwari et al 2017)等作物改 良和种质创造方面得到了广泛应用。相较于传统杂交和转基因方法,体细胞杂交 技术还具有以下优点:(1)融合产生的体细胞杂种拥有双亲的优良性状(Luthra et al 2018);(2)不同材料的细胞核和细胞质基因组重组(Eeckhaut et al 2013);

(3)体细胞杂交是非转基因手段,无潜在的生物安全性争议。分子生物学技术 发展,为体细胞杂种核质遗传、核质互作及重要性状形成机理等研究提供了新的 契机。

#### 2.3.2 体细胞杂种线粒体基因组遗传分析

根据母系遗传特性,有性杂交的父本基本不会提供细胞质基因组给杂交后代; 而体细胞融合可以产生不同的同核体或异核体杂种,以及异源杂种(胞质杂种)

(Xia et al 2009)。因此,体细胞杂种可能遗传双亲的优良性状,也可能产生与 细胞质基因有关的细胞质雄性不育等新性状(Guo et al 2013)。一般而言,体细 胞杂种双亲叶绿体基因组不会同时遗传给体细胞杂种后代,且亲本叶绿体基因组 也很少发生重组(Trabelsi et al 2005);仅在小麦、茄属等作物体细胞杂种发现其 叶绿体基因组存在双亲共存或重组现象(Xiang et al 2004; Bidani et al 2007)。

不同于叶绿体基因组能相对稳定遗传自融合双亲之一,大多数细胞融合产生体细胞杂种的线粒体基因组是双亲线粒体基因组重组的结果(Landgren and Glimelius 1990; Leino et al 2003)。Kemble等(1986)对马铃薯与Solanum brevidens种间体细胞杂种后代的线粒体基因组进行分析,发现52%体细胞杂种线粒体DNA 是分子间重组形成,45%的体细胞杂种线粒体DNA是分子内重组产生。随后,Xu等(1993)对马铃薯不同种原生质体融合得到的体细胞杂种线粒体DNA进行分析,发现81%的线粒体DNA表现出不同于双亲的RFLP条带。同时,小麦与玉米体细胞杂种线粒体DNA也发现双亲线粒体基因组共同存在(Xu et al 2003)。随着越来越多物种的线粒体基因组公布,将体细胞杂种及其双亲线粒体基因组进行比较,能进一步解析体细胞杂种线粒体DNA来源。Wang等(2012)对芸薹属作物细胞融合后代及其亲本线粒体DNA进行比较分析,发现体细胞杂种后代线粒体DNA

基因组仅少部分序列发生融合;而茄科植物烟草与天仙子原生质体融合后代的线 粒体基因组比天仙子大40%,比烟草大64%,说明形成融合后代过程中,亲本线 粒体基因组大部分融合成一个新的线粒体基因组,少部分序列丢失

(Sanchez-Puerta et al 2015; Garcia et al 2019a).

有别于上述可以采用"叶肉亲本原生质体+叶肉亲本原生质体"进行体细胞 杂交的物种,柑橘体细胞杂种大多是在"二倍体悬浮细胞系亲本原生质体+二倍 体叶肉亲本原生质体"的融合模式下获得(肖诗鑫 2014)。研究表明,柑橘体 细胞杂交二倍体后代,细胞核遗传物质源自叶肉亲本,叶绿体基因组随机遗传自 融合亲本之一,线粒体基因组遗传自愈伤组织亲本(Guo et al 2002, 2013);柑 橘胞质杂种细胞核、叶绿体及线粒体基因组遗传规律与二倍体体细胞杂种相同

(Wang et al 2010; Bassene et al 2011; Guo et al 2013);此外,也有研究显示,柑橘体细胞杂种/胞质杂种线粒体基因组融合了双亲线粒体基因组序列(Dambier et al 2011; Aleza et al 2016)。随着三代测序技术广泛应用,解析柑橘体细胞杂种线粒体基因组遗传和变异机制将成为可能。

#### 2.4 植物雄性不育研究

植物雄性不育是一种在两性花植物中普遍存在的现象,是在有性生殖过程中, 雌蕊能正常发育而雄蕊生殖器官发育异常,导致花药形态、小孢子发育和花粉功 能异常的遗传现象。根据控制不育性状基因的来源,Sears(1947)将不育系分 为三种类型:细胞核雄性不育、细胞质雄性不育和核质互作雄性不育。随后,研 究人员发现玉米细胞质雄性不育属于核质互作雄性不育的一部分。因此,植物雄 性不育分为细胞核雄性不育(Genic male sterility, GMS)和细胞质雄性不育 (Cytoplasmic male sterility, CMS)(Vedel et al 1994)。

#### 2.4.1 细胞核雄性不育

细胞核雄性不育分为隐性核不育和显性核不育。目前,科学家在不同种间杂 交试验中,发现大部分核不育为隐性不育。虽然细胞核雄性不育植株不育育性相 对彻底,且不受细胞内胞质基因组的影响,但受不育系难以保持等问题制约,生 产中很少采用细胞核不育系进行三系配套育种(高春保 2010)。随着更多高表达 核基因启动子的发现及 CRISPR/Cas9 基因编辑技术发展,利用隐性核不育系统 生产杂交种子的体系得到快速发展(Du et al 2020; Wang et al 2020)。

雄蕊发育是一个精细复杂的过程,控制小孢子发育的任何一个基因功能异常 都可能导致雄性不育(Wilson et al 2009; Zhang et al 2011; Verma et al 2019)。例如, 水 稻 基 因 *MULTIPLE SPOROCYTE* (*MSP1*) 与 拟 南 芥 基 因 *EXCESS MICROSPOROCYTES1* (*EMS1*) 在小孢子发育早期起作用(Zhao et al 2002; Nonomura et al 2003)。同时,使 *EMS1* 磷酸化的基因 TAPETUM DETERMINANT 1(*TPD1*)和 SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES(*SERK1/2*) 也与小孢子母细胞发育有关,并调控绒毡层形成。研究表明,TPD1 作为配体与 EMS1 结合,诱导 EMS1 磷酸化,介导其信号转导,在绒毡层形成和发育中起作 用(Jia et al 2008; Huang et al 2016)。细胞学观察拟南芥一突(ems1-1)、二突 (serk1-1 serk2-1)和三突(ems1-1 serk1-1 serk2-1)发现,这三个突变体均为小 孢子母细胞增多,绒毡层缺失;体内及体外实验证实 EMS1 和 SERK1/2 相互作 用,进而发生磷酸化;*TPD1* 异位表达不会影响二突(serk1-1 serk2-1)表型,TPD1 信号是 EMS1 和 SERK 以受体-共受体形式所介导(Li et al 2017)。

绒毡层发育基因异常,同样会导致雄性不育。研究人员发现作用于 EMS1 下游编码 bHLH 转录因子的 *DYT1*,在 *DYT1* 下游和 *AMS* 上游发挥功能的转录因 子 *TDF1* 及 *EMS1* 新的下游基因 β*CAs*,这些基因突变会导致绒毡层细胞分化异 常,细胞液泡化提前,小孢子提前降解(Zhang et al 2006; Gu et al 2014; Huang et al 2017)。近年来,研究人员在水稻中发现 bHLH 转录因子家族基因 *ETERNAL TAPETUM1* (*EAT1*)通过促进天冬氨酸酶基因表达,导致绒毡层细胞程序化死 亡提前,植株表现雄性不育(Niu et al 2013; Ono et al 2018)。

从拟南芥不同突变植株还鉴定和分离了许多调控孢原细胞和花粉壁形成的 关键核基因,并构建花药发育基因组调控网络。例如:ABC 模型中调控雄蕊原 基起始的基因/转录因子:*AP1、AP3、AG、PI*等(Zahn et al 2006; Wellmer et al 2014); 花药发育早期,参与孢原细胞分化的基因/转录因子:*SPL/NZZ*和*BAM1/2*(Yang et al 1999; Hord et al 2006);花粉内/外壁形成及花药释放起作用的基因/转录因子: *MYB80、DEX1、NEF1、RPG1、NPU、MS2、CESA、NAC、RMS2*(Paxson-Sowders

et al 2001; Ariizumi et al 2004; Mitsuda et al 2005; Persson et al 2007; Guan et al 2008; Doan et al, 2009; Chang et al 2012; Zhao et al 2020).

#### 2.4.2 细胞质雄性不育基因

植物细胞质雄性不育主要是由细胞质中线粒体基因突变导致其功能丧失或 发生障碍而不能产生正常有功能花粉的母系遗传现象(Kim and Zhang 2018)。基 于线粒体基因组母系遗传的特点,育种工作者利用栽培品种和野生材料间杂交、 回交或通过种间、属间杂交及原生质体融合等方法,将外源细胞质代替正常细胞 质从而获得细胞质雄性不育系(Bohra et al 2016; Garcia et al 2019b)。

不育胞质在自然界广泛存在,由于植物核基因组存在恢复基因,大部分植株 是可育表型。植物细胞质雄性不育基因克隆最常用方法是从不育系、保持系及恢 复系线粒体转录组和蛋白组的差异中鉴定 CMS 基因(Chen and Liu 2014; Xie et al 2014)。迄今,研究人员在水稻、油菜、萝卜、辣椒等作物报道了细胞质雄性不 育基因。水稻 HL-CMS、BT-CMS 和 LD-CMS 系线粒体基因组存在与 atp6 共转 录的 orf79, 异源表达及基因敲除 orf79 实验证实该基因为不育基因(Wang et al 2006; Wang et al 2013; Kazama et al 2019)。 野败型水稻不育基因 WA352 直接抑制 核编码电子传递链蛋白 COXII 发挥功能,引起绒毡层内细胞色素 C 积累,活性 氧含量升高,使绒毡层PCD过程提前,花粉败育(Luo et al 2013)。从萝卜CMS-ogu 系发现与甘蓝型油菜不育基因 orf125 同源基因,研究认为 orf125 与 atp8 共转录 编码毒性蛋白,导致雄性不育(Iwabuchi et al 1999)。从芥菜型油菜不同 CMS 系,先后发现线粒体不育基因 orf220、orf108、orf288 与其它线粒体保守蛋白共 转录,造成不同油菜系花粉败育(Yang et al 2010; Kumar et al 2012; Jing et al 2012)。将辣椒 CMS-Peterson 系线粒体基因组中与 cox2 共转录的 orf456 和 orf507 转化可育材料, 能导致转化系雄性不育 (Kim et al 2007; Ji et al 2015)。一些 CMS 基因序列全部或者部分属于线粒体型特异序列(Mitotype-specific sequence, MSS), 直接比较不同育性材料的线粒体基因组序列,在不育材料中寻找与 MSS 序列有 关的特异 ORF, 是鉴定 CMS 基因最直接的方法。例如: 通过水稻 D1-CMS 系与 其它 10 个材料的线粒体基因组比较,在 D1-CMS 系线粒体的 MSS 序列中发现 不育基因 orf182 (Xie et al 2018); 洋葱 CMS-T 系与雄性可育系线粒体基因组序

列和基因含量基本一致,仅 orf725 存在差异,Kim 等(2019)认为 orf725 是洋葱 CMS-T 的不育基因。

#### 2.4.3 植物细胞质雄性不育分子机理

细胞质雄性不育是植物线粒体基因组重组致使线粒体基因功能异常或丧失, 并导致花粉败育的现象(Kaul 1998)。其中,线粒体重排产生的嵌合基因与线粒 体保守功能基因共转录产生异常转录本和蛋白质,是植物雄性不育的主要原因。 根据 CMS 不育基因作用机制不同(图 1.2),研究人员提出以下五种假说:

1)细胞毒害假说。呼吸传递链相关基因的表达直接或间接受到线粒体嵌合 基因编码细胞毒素蛋白抑制,导致花药发育过程中细胞线粒体功能紊乱,影响植 物花药正常发育,引起雄性不育;若细胞核存在育性恢复基因,对线粒体不育基 因转录及翻译起到抑制作用,降低不育蛋白质对细胞的毒害作用,恢复线粒体功 能,最终使植物育性得到恢复(Chase 2007; Horn et al 2014)。研究人员首先在玉 米 T-CMS 材料发现 URF13 毒蛋白,能抑制大肠杆菌和酵母正常生长(Dewey et al 1987);随后,相继从向日葵 CMS(Nakai et al 1995)、萝卜 ogu CMS(Duroc et al 2005)、油菜 hau CMS(Jing et al 2012)、BT 型水稻 CMS(Wang et al 2006) 和 WA 型水稻 CMS(Luo et al 2013)发现 *ORF522、ORF138、ORF288、ORF79、 WA352*等雄性不育基因编码毒蛋白对大肠杆菌生长有抑制作用。但 *ORF288*及 *WA352* 转基因实验表明,植株雄性不育与线粒体不育基因编码蛋白的毒性没有 直接关系(Luo et al 2013; Heng et al 2018),且其它作物 CMS 系统也缺少这方面 的直接证据。因此,CMS 系统细胞毒性蛋白与植物细胞质雄性不育之间的关联 有待进一步研究。

2)线粒体能量供应障碍假说。线粒体基因编码不育蛋白破坏线粒体结构和功能,影响花药发育过程中与细胞呼吸相关电子传递链异常,导致能量供应受阻,引起花粉败育。Lee 和 Warmke (1979)观察玉米不育系花药发育过程中线粒体分裂速率,发现线粒体数目快速增加,认为线粒体为花药发育提供更多能量。甜菜 CMS-G 系不育表型是 cox2 蛋白突变降低了电子传递链系统 cox2 功能导致(Ducos et al 2001)。由此说明,线粒体电子传递链障碍与雄性不育相关。线粒体不育基因结构解析发现这些基因是由电子呼吸传递链基因一部分和未知功能

orf 组成的嵌合基因,可能与电子呼吸链复合体亚基结合,影响其功能,造成电 子呼吸传递障碍,使细胞内 ATP 含量降低 (Chen and Liu 2014)。狗尾草和辣椒 不育系中均发现 F1Fo-ATP 酶和细胞色素 C 氧化酶功能受损现象(Kale and Munjal 2005; Ji et al 2015)。同时,向日葵 CMS-PET1 系线粒体含有嵌合基因 orf522,这个基因编码的蛋白包括一段与 ORFB(atp8 同源基因)蛋白 N 端同源 的序列(Horn et al 2003)。研究人员比较向日葵不育系与可育系发现,不育系 atp 酶活性明显低于可育系,这可能是 orf522 影响 F1Fo-ATP 酶导致(Sabar et al 2003; Delannoy et al 2007)。红莲型水稻线粒体编码的 orfH79 与核编码电子传递链亚基 P61 互作抑制复合物 III 活性, 使小孢子细胞内线粒体 ATP 合成受阻, 不能为小 孢子发育过程提供足够能量,导致小孢子败育。随后的研究中发现线粒体内膜蛋 白 P61 与水稻不育表型有关联(Wang et al 2013)。总之,这些不育 orf 分子机理 支持 CMS 基因与花药发育过程中能量不足有关的假说。还有一些包含 mtETC 同 源蛋白的CMS蛋白能与mtETC传递链组成元件或F1Fo-ATP酶复合物竞争结合, 损害呼吸电子传递链效率或使 F1Fo-ATP 酶复合物活性丧失,不能产生足量的 ATP 用于花粉发育。这些 CMS 不育蛋白可能影响线粒体膜完整性,导致质子传 递受阻、ATP 合成不足(Chen and Liu 2014)。

3)细胞程序化死亡(PCD)异常假说。线粒体释放细胞色素 C 能使细胞内活性氧(ROS)含量升高,导致器官/组织 PCD 异常(Yao et al 2002; Greenberg and Yao 2004)。植物雄配子体发育是一个复杂过程,是营养组织(孢子体)和小孢子细胞(配子体)协同作用的结果。药室壁、绒毡层、四分体的形成与降解都有可能影响小孢子发育。其中,小孢子败育与绒毡层 PCD 提前或者延迟密切关联(Ma 2005; Kawanable et al 2006; Ji et al 2013)。向日葵 CMS-PET1 和水稻 CMS-WA 系中均证实绒毡层细胞线粒体提前释放细胞色素 C,使绒毡层细胞 PCD 提前,导致小孢子败育(Balk and Leaver 2001; Luo et al 2013)。

4)反向调控假说。线粒体功能障碍产生线粒体反向信号(线粒体-细胞核) 与代谢通路或活性氧通路相关(Aken and Pogson 2017)。这些途径在植物中涉及 转录因子或蛋白质的激活,例如 *AOX1a* 基因能调控 *NAC* 转录因子和 RRL 蛋白 介导的线粒体反向信号表达(Ng et al 2013; Yao et al 2015)。CMS 蛋白可能影响 细胞内线粒体氧化还原状态,引起与花粉育性有关的核基因异常表达,导致孢子

体细胞或配子体细胞最终死亡(Fujii et al 2007, 2009)。萝卜同质异源 petaloid CMS 与 carpeloid CMS 系研究发现,线粒体基因通过反向调控 MADS box 家族 B 类和 C 类基因表达,使雄蕊出现心皮化和花瓣状差异表型(Linke et al 2003)。



图 1.2 细胞质雄性不育基因不育机理及恢复基因核质互作模型。 引自 Chen and Liu 2014

# Figure 1.2 A general model for different layers of mitochondrial-nuclear gene interactions in cytoplasmic male sterility (CMS)/restorer (Rf) systems.

Cited from Chen and Liu 2014

5) RNA 编辑假说。线粒体 RNA 编辑主要通过密码子前两个位点碱基的转换,使线粒体基因编码蛋白的氨基酸组成发生改变,产生的异常产物能影响线粒体功能,导致花粉败育(Ichinose and Sugita 2017; Xiong et al 2017)。Iwahashi 等(1993) 在 BT 型水稻线粒体基因组中发现与 *atp6* 共转录的 CMS 基因 *orf79*。通过不育系与保持系序列比较发现,不育系线粒体基因组中 *orf79* 与 *apt6* 共转录产生的新转录本在 *atp6* 3'末端有 9 个编辑位点。在含有恢复基因的恢复系中,这 9 个编辑位点完全编辑,并将 *atp6* 基因转录产物分成 1.5 kb 和 0.45 kb 两部分,而在不育系中只有一个 2.0 kb 的 mRNA 分子部分编辑。随后,Wei 等(2008) 报道了水稻线粒体基因 *atp9* 转录本编辑与细胞质雄性不育的关系。近年来,水稻中还发现 PPR-RNA 编辑因子参与 *nad3* 序列 5 个位点 RNA 编辑,导致 *nad3* 功能异常,损害线粒体复合物 I 活性,进而对线粒体功能产生影响,引起水稻花粉育性降低(Xiao et al 2018)。

#### 2.5 柑橘雄性不育主要研究

无核是柑橘果实的一个优良性状,也是育种工作者选育柑橘品种的重要指标。 雄性不育是大多数柑橘品种无核或少核的主要原因。迄今,通过常规实生选种、 芽变育种、利用遗传温州蜜柑不育细胞质材料做母本进行杂交育种以及原生质体 融合等方法已培育出大量无核柑橘新品种(郭文武等 2019)。其中,采用细胞工 程技术将温州蜜柑雄性不育胞质基因组转移到有核品种 HB 柚,创制了二倍体 HB 柚胞质杂种'华柚 2 号',表现出雄性不育,是研究柑橘雄性不育分子机理 的理想试材(郑蓓蓓 2015)。柑橘雄性不育研究主要在细胞和分子两个层面进行。

细胞层面研究主要集中在组织学和细胞学观察及花粉粒细胞化学和生理特 征方面。Yamamoto等(1997)通过正反交实验证实温州蜜柑是典型的CMS,随 后 Goto等(2016,2018)鉴定了一些与CMS有关的候选位点。本实验室胡志勇 (2007)对温州蜜柑进行深入研究发现,温州蜜柑花粉败育主要发生在小孢子从 四分体释放之后的花粉发育早期,与 ATP 合成相关的线粒体基因表达降低时期 一致,说明线粒体基因是导致温州蜜柑雄性不育的重要原因;郑蓓蓓等(2012) 对 HB 柚胞质杂种花器官形态进行细胞学观察发现,花粉发育过程中绒毡层和花 粉外壁发育异常、不能积累孢粉素、小孢子败育,并且还观察到 HB 柚胞质杂种 雄蕊原基发育受到抑制;'黔阳无核'椪柑花粉母细胞形成四分体时期出现异常, 且绒毡层及胚囊发育异常(邱文明 2013);进一步观察发现'黔阳无核'小孢子 与胚囊败育机制一致,均是胞母细胞减数分裂异常的结果(王瑶 2017); Zhang 等(2014)采用细胞学方法对'瓯柑'无核突变体花药发育和花粉形态进行观察, 认为花粉败育时期是从小孢子母细胞减数分裂至四分体时期。

分子层面研究主要集中在利用转录组、蛋白组、小RNA组、降解组、代谢 组等组学联合分析,以期挖掘调控柑橘雄性不育相关基因。邱文明(2013)利用 RNA-seq 技术,发现柑橘 MSLP 基因存在 4 个转录本,且在花药发育过程中,'黔 阳无核'MSLP 基因的表达水平比'鄂柑一号'高,推测该基因为柑橘雄性不育 相关基因。郑蓓蓓等(2012, 2014)通过高通量测序技术发现一系列与 HB 柚胞 质杂种雄性不育性状相关的基因和蛋白,如 CitAP3 基因和 CYP716 蛋白,这些 差异基因和蛋白质主要与能量代谢、核酸结合、蛋白质合成及降解过程相关;此 外,与花器官发育相关的转录因子也表现出差异表达;同时,方燕妮等(2014, 2016)

发现与'黔阳无核'椪柑及 HB 柚胞质杂种不育性状相关 miRNAs,并且认为 miRNA-PHAS-phasiRNA-target 调控网络可能参与了核质互作雄性不育过程,在 此基础上,王蓉等(2020)发现 miR399-*UBC24* 可能通过调控 SEP 家族基因影 响花器官发育,并通过 *ICE1* 调控气孔发育影响花药开裂; Zhang 等(2018)对

'无籽' 瓯柑进行转录组和蛋白质组联合分析,表明其雄性不育可能与苯丙氨酸 代谢有关。虽然柑橘雄性不育研究已取得一定进展,但与柑橘细胞质雄性不育有 关的 CMS 基因尚未报道,前期通过柑橘雄性不育的细胞和分子层面研究,为柑 橘线粒体雄性不育基因发掘提供了理论基础。

## 3本研究的目的和内容

温州蜜柑雄性不育是典型的细胞质雄性不育(Yamamoto et al 1997), HB 柚 胞质杂种是'国庆1号'温州蜜柑和 HB 柚(单胚)原生质体融合产生的胞质杂 种,表现雄性不育,隔离种植时果实无核,是研究柑橘细胞质雄性不育的理想试 材。本研究以 HB 柚胞质杂种及其融合双亲为试材,从线粒体基因组角度,对柑 橘线粒体基因组进行了测序、组装,发掘与柑橘细胞质雄性不育相关的线粒体基 因。主要研究内容如下:

1. '国庆1号'温州蜜柑和 HB 柚的线粒体 DNA Illumina 测序及总 DNA Illumina 和 PacBio 测序,拼接并注释其线粒体基因组信息,组装'国庆1号'温 州蜜柑和 HB 柚的线粒体基因组。

2. HB 柚胞质杂种及其融合双亲线粒体基因组比较分析,从基因组水平确定 HB 柚胞质杂种核基因组及胞质基因组亲本来源;进一步利用'国庆1号'温州 蜜柑与冰糖橙胞质杂种的线粒体基因组,明确柑橘胞质杂种后代线粒体基因组遗 传特点。

3. 基于 HB 柚胞质杂种及其双亲转录本和线粒体基因组信息,分析不育系 HB 柚胞质杂种和'国庆1号'温州蜜柑与 HB 柚 CMS 特异 ORFs,在这些特异 ORFs 中发掘与 CMS 有关的候选线粒体基因;通过异源表达、核互作蛋白筛选,分析候选 CMS 基因功能。

# 第二章 '国庆1号' 温州蜜柑和 HB 柚线粒体基因组测序与 比较分析

1 引言

柑橘作为多年生木本果树,具有童期长、遗传背景复杂、自交不亲和以及珠 心多胚等特点,采用杂交群体、正反交等正向遗传学方法发掘柑橘功能基因存在 诸多困难。高通量测序技术发展与应用为柑橘不同种基因组测序及组学研究提供 了契机。利用 Illumina 测序技术相继完成了克里曼丁橘和甜橙全基因组测序和拼 接(Xu et al 2013; Wu et al 2014)。近年来,随着读长更长的单分子实时 DNA 测 序技术发展,更多柑橘种类的全基因组陆续公布。本实验室以单倍体柚为试材, 完成柚全基因组测序和序列物理图谱拼接,并解析了柑橘单多胚关键序列差异 (Wang et al 2017);通过连续自交创建山金柑自交纯合系并利用 S3 代单株山金 柑为试材,获得的山金柑参考基因组极大的方便了柑橘功能基因组学研究(Zhu et al 2019)。基于以上柑橘不同种的基因组信息,柑橘色素功能基因发掘研究已 取得积极进展(Zhang et al 2020)。

除细胞核基因组外,植物细胞内的叶绿体和线粒体也包含各自的基因组。 2006年,Bausher等发布了甜橙叶绿体基因组信息;随后通过34个柑橘不同种 叶绿体基因组分析,阐明了柑橘属野生种和栽培种之间的关系 (Carbonell-Caballero et al 2015)。然而,柑橘线粒体基因组信息鲜有报道。目前, 己有多种果树线粒体基因组相继公布,如:葡萄(Goremykin et al 2009)、苹果 (Goremykin et al 2012)、猕猴桃(Liu et al 2017)。研究人员发现,被子植物线 粒体基因组具有分子内或分子间序列重排、基因组亚化学计量效应、线粒体基因 顺式剪切和 RNA 编辑等特点。其中,线粒体基因组重复序列之间重组产生的嵌 合基因可能与线粒体保守基因共转录,导致线粒体功能丧失而不能产生有功能的 花粉。

本实验室前期利用细胞融合技术将'国庆1号'温州蜜柑与HB 柚进行原生质体融合,获得HB 柚胞质杂种(G1+HBP),前期通过SSR 分子标记证实HB 柚胞质杂种细胞核和叶绿体基因组来自HB 柚,线粒体基因组来自'国庆1号'

温州蜜柑(Guo et al 2004; Wang et al 2010)。本研究以'国庆1号'温州蜜柑和 HB 柚两个融合亲本为材料,利用 Illumina 测序技术对这两个材料的线粒体 DNA、 总 DNA 测序,并采用 PacBio 三代测序技术对总 DNA 进行测序。生物信息学与 Sanger 测序技术结合,拼接获得完整的线粒体基因组,并以已发表物种的线粒体 蛋白注释信息为参考,对两个材料的线粒体基因组进行功能注释。通过两个材料 的线粒体基因组序列比较分析,为后续胞质杂种线粒体基因组来源及柑橘雄性不 育相关 CMS 基因发掘奠定基础。

# 2 材料和方法

#### 2.1 柑橘愈伤组织及黄化幼苗培养

本研究中,以'国庆1号'温州蜜柑('Guoqing No.1' Satsuma mandarin, *Citrus unshiu* Marc., 雄性不育)愈伤组织, HB 柚('Hirado Buntan' pummelo, *Citrus grandis* (L.) Osbeck, 雄性可育)黄化苗为材料提取线粒体 DNA(图 2.1)。实验所用材料'国庆1号'温州蜜柑愈伤组织培养在 MT 固体培养基上,每 20 天左 右继代一次,于培养室避光培养。HB 柚种植于华中农业大学柑橘研究所的柑橘 种质资源圃。从 HB 柚果实取出种子,用清水洗净种子表面果胶等杂质。用 1 mol NaOH 浸泡 10-15 min 后,清水洗净种子表面 NaOH,徒手剥去外种皮,再用 3%-5% NaClO 浸泡 15-20 min,超净台内用无菌水清洗 3-5 次。清洗消毒完后用镊子将 种子(胚朝下)放入播种培养基(MT+25g 蔗糖),为保证黄化幼苗质量,每个 试管仅放 1 颗种子,黑暗培养 40-50 天。



**图 2.1 柑橘线粒体 DNA 提取材料。**(A)'国庆1号'温州蜜柑愈伤;(B) HB 柚果实;(C) HB 柚黄化苗。

Figure 2.1 Extraction materials of citrus mitochondrial DNA. (A) Embryogenic callus of G1; (B) The fruit of HBP; (C) The etiolated seedlings of HBP.

### 2.2 柑橘线粒体提取

植物线粒体提取主要通过密度梯度离心和差速离心法,从植物不同组织提取 完整的线粒体,提取所用溶液配方见附录 I。选取柑橘愈伤组织/黄化幼苗作为材料,通过蔗糖梯度离心提取线粒体,具体步骤如下:

- 粗提取:
- 1. 柑橘种子播种在试管中, 暗光培养 40-50 天, 收集上胚轴, 约 60 g, 2-4 ℃ 预 冷;
- 客收集到的上胚轴放入到匀浆机中,并按照 2 mL/g 比例加入研磨缓冲液 (homogenization medium)进行研磨,医用滤膜加 Miracloth 过滤残渣,4 °C 保存滤液;
- 3.1,000 g 4 ℃ 离心 10 min 后将上清液转移至另一预冷的 50 mL 离心管;
- 4.18,000 g 4 °C 离心 20 min 后弃上清液得到棕黄色/淡绿色沉淀,加入 Wash buffer;
- 5.1,000 g 4 °C 离心 10 min 后将上清液转移到另一预冷的 50 mL 离心管;
- 6. 18,000 g 4 ℃ 离心 20 min 后弃上清液加入 1-2 mL 洗液, 4 ℃ 连续操作粗提取 完成;

精提取:

- 1.18,000 g 4 °C 离心 20 min 后收集粗提取的线粒体,并将线粒体溶于 buffer A;
- 2. 向其中加入 DNase 酶 50-100 μL (视沉淀量确定), 37 °C 2-3 h;
- 3. 18,000 g 4 ℃ 离心 20 min 后弃上清液,并加入 1-2 mL buffer B 重悬线粒体;
- 4. 配置蔗糖梯度;

10 mM Tris-HCL 20 mM EDTA

蔗糖	1.45 M	1.2 M
5 mL	2.48 g	2.052 g

 5. 将重悬在 buffer B 中的线粒体平铺在蔗糖梯度中,水平转头超速离心 72,000 g (25,00 rpm), 4 ℃,离心 1 h; 6. 用1mL一次性吸管小心将悬浮于1.45-1.2 M之间的线粒体吸出,18,000 g4 ℃ 离心 20 min,收集线粒体,用詹纳斯绿 B 染液及通过 western 技术对线粒体的 纯度和完整性进行检测;

7. 最后用 CTAB (附录 II) 法提取线粒体 DNA。

# 2.3 线粒体 DNA、总 DNA Illumina 及总 DNA PacBio 测序 DNA 文 库构建与高通量测序

用于 Illumina 和 PacBio 测序平台测序的 DNA 样品经 Qubit Fluorometer 和琼 脂糖凝胶电泳检验合格后,文库构建和测序工作委托北京诺禾致源有限公司完成。 Illumina 测序平台的 DNA 样品采用 NEBNext Ultra DNA library Prep Kit 提供的方 法进行建库。首先用 Covaris 超声波破碎仪将 DNA 随机打断成长度为 350 bp 的 片段,随后 DNA 片段末端修复、添加 ployA、测序接头、纯化、PCR 扩增等步 骤完成整个文库制备,构建好的文库通过 Illumina 平台测序(图 2.2A)。PacBio 测序平台的 DNA 样品采用 SMRTbell template prep Kit 提供的方法进行建库。首 先将 DNA 用 26 G Needle 片段化,随后用 BluePippin 选择 20 kb 以上的片段,经 末端修复和加 A 尾后,再在片段两端分别连接接头等步骤完成整改文库制备, 构建好的文库通过 PacBio 平台测序(图 2.2B)。





图 2.2 Illumina (A) 和 PacBio (B) 测序建库流程。 Figure 2.2 Flow for constructing Illumina (A) and Pacbio (B) library.

#### 2.4 LncRNA 文库构建与高通量测序

叶、花、果皮和汁胞 RNA 提取参考 Liu(2009)。提取的 RNA 使用 Nanodrop、 琼脂糖凝胶电泳和 Agilent 2100 进行浓度和完整性检测, 合格的 RNA 样品进行 LncRNA 建库。文库构建和测序工作委托北京诺禾致源有限公司完成。根据 Ribo-Zero™ kit 提供的方法从总 RNA 样品中去除 rRNA, 用 fragmentation buffer 将去除 rRNA 的样品打断成小片段。随后以片段化的 RNA 为模板, 加入随机引 物反转录成 cDNA 第一条链, 再以第一条链为模板, 通过缓冲液、dNTPs (dNTP 中的 dTTP 用 dUTP 取代)、DNA polymerase I 和 RNase H 合成 cDNA 第二条链。 合成的双链 cDNA 经过处理(纯化、末端修复、加 A、连接测序接头)后用 USER 酶降解含有 U 的 cDNA 链并对 PCR 富集, 最后用 AMPure XP beads 纯化 PCR 产物,得到链特异性文库。文库经质量检测 (Qubit 2.0 和 Q-PCR 对文库定量、 Agilent 2100 检测文库插入片段大小)合格后用 Illumina Hiseq4000 基因分析仪测 序(图 2.3)。



图 2.3 IncRNA 测序建库流程。 Figure 2.3 Flow for constructing IncRNA library.

#### 2.5 线粒体 de novo 组装策略

使用 SPAdes 软件对总 DNA 的 PacBio 和线粒体 DNA 或者总 DNA 的 Illumina 测序结果进行 de novo 组装,其中设置的主要参数为: k=91、119、127(取最佳), 组装所用脚本参考 Wang (2019)所述。使用 Velvet 软件对线粒体 DNA 或者总 DNAIllumina 测序数据进行 de novo 组装其中设置的主要参数为: k=91、119、127 (取最佳), ExpCov=20、50、100、200、500、800(取最佳),组装所用脚本参 考 Zhu (2016)所述。将两种软件组装结果进行整合分析,获得 2 到 3 条线粒体 scaffolds。通过 PCR 和 Sanger 测序填补 scaffolds 之间的 gap,最终获得线粒体基 因组环状序列。使用 picard MarkDuplicates 去除所有 Illumina 中 reads 的重复序 列,利用 SAMtools 将剩余 reads 比对到线粒体基因组完整序列,以正向 reads 与 反向 reads 碱基出现的值为参考,利用 Pilon v1.22 对线粒体基因组序列进行校正, 得到线粒体基因组最终序列,其中三个软件设置的主要参数依次为: VALIDATION\_STRINGENCY=LENIENT REMOVE\_DUPLICATES=false MAX\_FILE\_HANDLES\_FOR\_READ\_ENDS\_MAP=8000; minimum mapping quality=30; fix=SNPs, InDels。

### 2.6 RNA 编辑位点鉴定

使用 HISAT2 软件将'国庆1号'温州蜜柑和 HB 柚 LncRNA-seq 数据比对 到各自组装的线粒体基因组(Kim et al 2015)。比对结果通过 SAMtools 转为二进 制 BAM 文件,并分析碱基比对结果,其中设置的主要参数为: adjust-MQ=50, mini-reads=2, gap-frac=0.02, max-depth=1000。使用 bcftools v1.3 鉴定二进制 BAM 文件的 RNA 变异,其中线粒体保守蛋白编码区变异位点表示线粒体 RNA 编辑位点,其中设置的主要参数为: multiallelic-caller=m。通过计算编辑位点的 变异频率推测该位点的 RNA 编辑水平(Picardi et al 2010),使用 SnpEff v4.3s (Cingolani et al 2012)对 RNA 编辑进行变异注释。

#### 2.7 细胞内基因水平转移序列鉴定

为了鉴定柑橘线粒体基因组中叶绿体序列来源的片段,使用 BLASTN 将柑橘线粒体基因组比对到甜橙叶绿体基因组(Xu et al 2013),参数使用: e-value 为 1e-5,长度大于 100 bp。

### 2.8 基因注释

利用 MITOFY 提供的线粒体基因数据库对柑橘线粒体基因组进行注释 (Alverson et al 2010)。使用 tRNAscan-SE v1.21 对 tRNA 和 rRNA 进行注释。使 用 BLASTN 鉴定线粒体基因组重复序列(相似度>95%)。Bioedit 软件计算线粒 体基因组 GC 含量,OGDRAW 绘制线粒体基因组物理图谱(Lohse et al 2013)。

#### 2.9 叶绿体基因组 de novo 组装策略

叶绿体基因组序列相比线粒体基因组序列更加保守,且 Illumina 和 PacBio 测序拼接得到的 scaffold 覆盖度也比线粒体基因组高。因此,首先对线粒体基因 组拼接剩余的 scaffold 进行覆盖度判断,初步确定属于叶绿体基因组序列的 scaffold;随后与从 NCBI 下载的柑橘叶绿体基因组数据比对,根据 scaffold 之间
序列与参考基因组重叠区信息,将上述 scaffold 手动连成叶绿体基因组初级环形序列;叶绿体基因组序列校正方法同 2.6 线粒体 *de novo* 组装策略所述。

# 2.10 线粒体及叶绿体基因组共线性及变异(SNP/InDel)分析

使用 lastz 分析 '国庆 1 号'温州蜜柑和 HB 柚线粒体和叶绿体基因组的共 线区域,使用 gmaj 对生成的 maf 文件进行分析,生成.Difference 文件用于同源 序列片段共线性的统计学分析,生成.rdotplot 文件通过 R.plot 生成点阵图(dotplot)。 Mummer 软件对线粒体和叶绿体基因组序列进行同源比对,生成.mums 文件提取 SNP 物理位置信息,用自编脚本将物理位置连续的 SNP 合并为差异 InDel。

# 3 结果

# 3.1 '国庆1号'温州蜜柑和 HB 柚线粒体基因组 Illumina 测序及组 装

## 3.1.1 柑橘线粒体提取及鉴定

取 HB 柚黄化幼苗鲜样 30 g 和 60 g,-80 °C 冰冻样品 60 g 进行线粒体提取。 通过蔗糖铺垫和密度梯度离心法,可将线粒体与质体、细胞器及细胞核分离,并 在蔗糖 1.2 M 与 1.45 M 界面富集纯净线粒体。当黄化幼苗用量为 60 g,在蔗糖 梯度界面能富集到大量高纯度线粒体(图 2.4A),当黄化幼苗用量为 30 g,在蔗 糖梯度界面有极少量的线粒体富集(图 2.4B),表明柑橘黄化幼苗提取线粒体所 需样品用量至少 60 g。然而,利用蔗糖铺垫方法在冰冻样品提取粗线粒体过程中, 上清液呈现大量白色絮状物(图 2.4C),在 1.2 M 与 1.45 M 界面没有明显环带出 现,表明柑橘黄化幼苗冰冻保存的样品无法用于线粒体提取。黑暗培养的'国庆 1 号'温州蜜柑愈伤组织线粒体提取采用同样的方法进行。



图 2.4 柑橘高纯度线粒体提取。(A)样品质量为 60g,密度梯度离心后线粒体富集在 1.2 M 与 1.45 M 浓度的蔗糖分层液之间;(B)样品质量为 30g,密度梯度离心后线粒体富集在 1.2 M 与 1.45 M 浓度的蔗糖分层液之间;(C)冰冻样品线粒体提取的杂质。

Figure 2.4 Extraction of high-quality mitochondria from HBP seedlings. 1.2 M and 1.45 M of sucrose density gradient were used to separate the purified mitochondria, and mtochondria enrichment between 1.2 M and 1.45 M sucrose liquids. (A) 60 g seedlings; (B) 30 g seedlings; (C) Impurities were presented at extraction of mitochondria from frozen seedlings.

## 3.1.2 线粒体完整性检测

细胞色素 C 是一种线粒体膜蛋白,可作为检测线粒体完整性和浓度的指标。 分别提取样品总蛋白及提纯线粒体蛋白,利用 Western blot 技术,用细胞色素 C 抗体检测不同样品蛋白中细胞色素 C 含量。提纯的线粒体蛋白细胞色素 C 含量 显著高于样品总蛋白,表明提纯的线粒体完整性好,纯度较高(图 2.5A)。

詹纳斯绿 B 是一种能将线粒体染成蓝绿色的染色剂。将詹纳斯绿 B 染液与 少量提纯的线粒体混合,常温放置 10-15 min,体视显微镜下观察(图 2.5B),视 野中蓝绿色絮状物为完整的线粒体。



图 2.5 柑橘线粒体纯度鉴定。(A) Western 检测不同材料 中细胞色素 C 含量。M: marker, 1: 20 ng 总蛋白, 2: 10 ng 线粒体蛋白, 3: 40 ng 总蛋白, 4: 20 ng 线粒体蛋 白; (B) 詹纳斯绿 B 对提取的柑橘线粒体进行染色。 Figure 2.5 Purity identification of the extracted mitochondria. (A) Western blot was used to detect cytochrome C content of different materials. M: marker; 1: 20 ng total protein; 2: 10 ng mitochondria protein; 3: 40 ng total protein; 4: 20 ng mitochondria protein. (B) The mitochondria extracted from citrus seedlings/callus that was dyed by Janus green B.

# 3.1.3 线粒体 DNA 检测

分别以线粒体 DNA 和总 DNA 为模板,用细胞内不同遗传物质标记引物进行 PCR 扩增。琼脂糖凝胶电泳检测显示,线粒体 DNA 主要检测到线粒体标记引物扩增条带,仅检测到 3 对核基因标记引物扩增条带,且没有叶绿体标记引物扩增条带出现;而总 DNA 几乎检测到所有引物扩增条带(图 2.6A)。由此表明,提取的线粒体 DNA 纯度较高,仅有少量其它 DNA 干扰。采用琼脂糖凝胶电泳检测提取的线粒体 DNA 完整性,从图 2.6B 可以看出,提取的线粒体 DNA 条带清晰,仅有轻微降解,片段大于 15 k,表明提取的线粒体 DNA 片段完整性高。同时,Qubit Fluorometer 检测线粒体 DNA 浓度及总量,均达到测序建库的要求。



**图 2.6 PCR 鉴定线粒体 DNA 纯度和线粒体 DNA 质量检测。**(A)细胞内不同遗传物质 标记鉴定线粒体 DNA 纯度。nuDNA:核 DNA; cpDNA:叶绿体 DNA;mtDNA:线粒体 DNA; (B)线粒体 DNA 琼脂糖凝胶检测图。CK: DNA 标样; M-1:5000 bp marker; M-2:23130 bp marker

Figure 2.6 Validation of the mtDNA through PCR analysis and quality test of the mtDNA from citrus. (A) PCR confirmed the mtDNA from citrus using difference intracellular genetic material markers. nuDNA: nuclear DNA; cpDNA: chloroplast DNA; mtDNA: mitochondrial DNA. (B) GeleleetroPhoresis detection of mtDNA. CK: DNA standard; M-1: 5,000 bp marker; M-2: 23,130 bp marker.

#### 3.1.4 线粒体基因组测序和拼接

'国庆1号'温州蜜柑线粒体 DNA 利用 Illumina MiSeq 技术测序,得到 120
Mb clean reads,平均读长 300 bp,相关测序数据统计见表 2.1。首先,利用
SOAPdenovo和 gapclose 将测序结果进行组装,得到 144 个 scaffolds。随后,将
测序深度 50X 以下的 scaffolds 与 NCBI上 nt/nr 库做比对,共有 122 条 scaffolds

没有注释到与线粒体基因组相关信息,因此将其剔除。最后,通过与甜橙叶绿体 基因组和克里曼丁橘参考基因组(NC\_008334.1, Celementina\_182\_v1)比对, 滤掉4条比对上甜橙叶绿体基因组和克里曼丁橘参考基因组序列的 scaffolds,最 终剩余18条 scaffolds为柑橘线粒体 DNA 序列,相关拼接结果统计如表2.2 所示。 PCR 步移法能有效填补>2 kb 的 gap,为减少<2 kb 的 scaffolds 可能是重复序列对 拼接产生干扰,将18条小于2 kb 的 scaffolds 剔除,并对剩余的13条 scaffolds 设计13 对线粒体基因组引物(引物序列见附录 III),采用引物两两随机组合, PCR 步移法对'国庆1号'温州蜜柑线粒体 scaffolds 进行拼接,最终将其连接 成一个完整、全长为489,644 bp 环状线粒体 DNA。下一步,对拼接完成的'国 庆1号'温州蜜柑线粒体基因组进行基因结构分析。

表 2.1 Illumina 平台 mtDNA-seq 和 RNA-seq 测序数据统计 Table 2.1 Summary of mtDNA-seq and RNA-seq data sets sequenced using the Illumina

platform					
Sample	Raw data	Clean data	Clean data	Clean data	Clean data
	(GB)	(GB)	(GC%)	Q20(%)	Q30(%)
G1 mtDNA	0.17	0.12	46.55	95.58	88.92
HBP mtDNA	0.68	0.60	41.19	93.39	87.82
G1 RNA	14.46	14.11	45.25	95.93	90.34
HBP RNA	13.20	12.95	45.50	95.50	89.53

表 2.2 柑橘线粒体基因组组装统计

Table 2.2 Statistical analysis of mitcohdnrial assembling data				
Scaffold (G1) Scaffold (HBP)				
18	22			
417,211	518,695			
63,952	120,621			
13,573	19,013			
88,419	153,420			
702	548			
45.26	45.40			
	ical analysis of mitcohdnria Scaffold (G1) 18 417,211 63,952 13,573 88,419 702 45.26			

## 3.1.5 '国庆1号'温州蜜柑线粒体基因组结构

'国庆1号'温州蜜柑线粒体基因组是一个全长为489,644 bp的环状 DNA, GC 含量 45.17%。根据 mitofy 总结的植物线粒体基因组序列,通过序列同源性 分析,利用 ORFFinder、BLASTX、BLASTN 和 tRNA-SE 等软件对'国庆1号' 温州蜜柑线粒体基因组进行功能注释,'国庆1号'温州蜜柑线粒体基因组共有 35 个编码蛋白基因、18 个 tRNA 基因和 3 个 rRNA 基因(图 2.7)。编码线粒体 蛋白基因包括:电子传递链复合体 I-V、细胞色素 C 生物合成基因、核糖体蛋白 基因及 matR、mttB。

'国庆1号'温州蜜柑编码线粒体蛋白部分基因之间序列被I型或者II型内 含子打断。其中, nad4、ccmFc、cox2、rps14 被1个I型内含子隔开,外显子之 间间隔<5 kb;而 nad7的外显子之间有5个内含子。基因 nad1、nad2和 nad5的 外显子被多个II型内含子隔开。nad1的5个外显子相距较远,分布在基因组间 隔159 kb、36 kb、65 kb的4个区; nad2的5个外显子分布在间隔100 kb的基 因组序列之间; nad5的4个外显子分布在基因组序列的两个区,相隔148 kb, 且这两个区的外显子转录方向相反,这种排列方式需要两次反式剪接和两次顺式 剪接将外显子相连。

'国庆1号'温州蜜柑含有 rrn5、rrn18 和 rrn26 三个 rRNA。与大多数植物 线粒体基因组编码 rrn5 和 rrn18 基因的方式相似,相距仅 152 bp 的 rrn5 和 rrn18, 先形成一个由同一条 DNA 链转录的双顺反子前体,随后产生 rrn18-rrn5 紧密排 列的成熟体。此外,相隔 176 kb 的 rrn26 也从相同方向转录。'国庆1号'温州 蜜柑共编码 18 个 tRNA。根据这些基因的起源,将其分为线粒体 tRNA,包括 trnM(cau)、trnK(uuu)、trnQ(uug)、trnG(gcc)、trnS(uga)、trnE(uuc)、trnY(gua)、trnP(ugg)、trnD(guc)、trnC(gca)、trnI(uau); 叶绿体 tRNA,包括 trnN(guu)、trnW(cca)、trnM(cau)、trnH(gug)、trnI(uau)、trnS(uga)。



**图 2.7 '国庆1号'温州蜜柑线粒体基因组结构图。**最外层红色圈表示线粒体基因组大小; 红色圈内侧表示线粒体基因在基因组的位置,颜色相同方框表示相同复合体的基因;灰色圈 外侧方框表示重复序列 (>100 bp);灰色圈内侧不同颜色圆圈表示 rRNA 和 tRNA。

Figure 2.7 A circular diagram of the G1 mitochondrial genome. The ourermost circle represents the physical map of the G1 mitochondrial genome. Coding sequences transcribed are drawn on the inside of the outermost circle. Different classes of conserved protein-coding genes were assigned with distinct colors. The locations of repeats larger than 100 bp were showed on the inside of the grey circle.

## 3.1.6 HB 柚线粒体基因组结构

从 HB 柚提取的线粒体 DNA 在 Hiseq 测序平台测序,得到 604 Mb、平均读 长 150 bp。以拼接完成的'国庆 1 号'温州蜜柑线粒体基因组为参考,对 HB 柚 进行线粒体基因组拼接组装。首先,将 HB 柚线粒体测序 clean reads 与'国庆 1 号'温州蜜柑线粒体基因组进行比对,滤掉不能比对上'国庆 1 号'温州蜜柑线 粒体基因组的序列。随后,采用与'国庆 1 号'温州蜜柑线粒体基因组组装相同 策略,对 clean reads 进行组装、优化和填补 gap,从而得到最初组装结果(表 2.2)。 将 22 条小于 2 kb 的 scaffolds 剔除,并对剩余 14 条 scaffolds 设计引物(引物序 列见附录 III),用 PCR 步移方法对 HB 柚线粒体基因组进行拼接。HB 柚线粒体 基因组大小 484,291 bp,GC 含量 45.19%,保守基因数为 34 个,tRNA 基因有 19 个,rRNA 基因 3 个(图 2.8)。与'国庆 1 号'温州蜜柑线粒体基因组编码基 因相比,HB 柚线粒体基因组含有 1 个 tRNA-Phe,缺少基因 *rps1*。此外,HB 柚 线粒体基因组 I 型和 II 型内含子分布与'国庆 1 号'温州蜜柑大致相同。



图 2.8 HB 抽线粒体基因组结构图。最外层红色圈表示线粒体基因组大小; 红色圈内侧表示 线粒体基因在基因组的位置, 颜色相同方框表示相同复合体的基因; 灰色圈外侧方框表示重 复序列 (>100 bp); 灰色圈内侧不同颜色圆圈表示 rRNA 和 tRNA。

Figure 2.8 A circular diagram of the HBP mitochondrial genome. The ourermost circle represents the physical map of the HBP mitochondrial genome. Coding sequences transcribed are drawn on the inside of the outermost circle. Different classes of conserved protein-coding genes were assigned with distinct colors. The locations of repeats larger than 100 bp were showed on the inside of the grey circle.

# 3.1.7 '国庆1号'温州蜜柑和 HB 柚线粒体基因组序列准确性分析

为保证基于柑橘线粒体基因组后续研究的准确,采用 SPAdes 组装、Mummer 共线性比较和 SAMtools 比对两种策略,将'国庆1号'温州蜜柑和 HB 柚线粒

体基因组原始 clean reads 与拼接完成的线粒体基因组进行比较,分析'国庆1号' 温州蜜柑和 HB 柚线粒体基因组准确性。从 SPAdes 组装结果中过滤得到'国庆 1号'温州蜜柑和 HB 柚的线粒体 Contigs 分别有 3 条和 7 条。将这些 Contigs 与 对应线粒体基因组进行共线性分析,表明 SPAdes 组装的 Contigs 序列与'国庆 1 号'温州蜜柑拼接完成的基因组整体共线性很好,但 Contig1、Contig2 存在倒位 序列,且有 20 kb 序列差异; HB 柚拼接完成的线粒体基因组同样能在 SPAdes 拼接的 Contigs 找到对应序列,但 Contig2、Contig3、Contig5、Contig6、Contig7 存在倒位序列,且 Contig4 序列丢失(图 2.9)。同时,IGV 分析 SAMtools 结果 表明,拼接完成的'国庆 1 号'温州蜜柑和 HB 柚线粒体基因组与原始 clean reads 之间有错配区域。以上结果表明,拼接完成的'国庆 1 号'温州蜜柑和 HB 柚线 粒体基因组序列是真实存在,受限于缺少可靠的柑橘线粒体参考基因组及线粒体 基因组自身重复序列干扰,拼接完成的'国庆 1 号'温州蜜柑和 HB 柚线粒体基 因组出现大量倒位的片段。因此,需要使用读长更长的三代测序技术对'国庆 1 号'温州蜜柑和 HB 柚线粒体基因组进行校正。



图 2.9 基于不同组装软件的'国庆1号'温州蜜柑和 HB 柚线粒体基因组共线性分析。 Figure 2.9 The synteny of G1 and HBP mitochondrial genome was assemblied from different softwares.

# 3.2 柑橘线粒体基因组三代测序组装策略

将继代 15 d 的'国庆 1 号'温州蜜柑愈伤组织、新萌发的 HB 柚嫩叶,送 至北京诺禾致源公司提取 DNA 并进行 PacBio 平台的三代建库、测序,相关测序 数据统计见表 2.3。结合线粒体 DNA 和总 DNA 二代测序结果,采取三代测序数 据(PacBio)与二代测序数据(Illumina)联合分析,一代测序技术(Sanger)补充 gap 的策略,对'国庆1号'温州蜜柑和 HB 柚的线粒体基因组进行校正。先通过 Illumina 测序数据对线粒体基因组数据进行富集拼接,同时利用 Illumina 测序数据与 PacBio 测序数据进行混合组装,随后将两种组装方式获得的结果整合,最后根据已获得的线粒体 scaffolds 片段利用 Sanger 方法拼接,从而得到完整的线粒体基因组。具体流程如下(图 2.10):

1) 用 Velvet 软件对 Illumina 测序数据进行组装,同时用 SPAdes 软件对 PacBio 和 Illumina 测序结果进行混合组装,得到约 3500 条 scaffolds 和 5000 条 scaffolds 组装结果;

2) 以序列最长 scaffold 的覆盖度为标准,对剩余 scaffolds 进行初步筛选。此步中,只要在标准覆盖度上下范围内 10 或 100 的 scaffolds 都认为是线粒体基因组序列。 植物线粒体序列编码的基因相对保守,进一步通过与已发表的线粒体序列比对, 对初步筛选得到的线粒体序列进行验证,将没有匹配线粒体序列的 scaffolds 剔除, 同时保留包含有叶绿体序列的 scaffolds,并做标记;

3) 将 2) 获得的 Velvet scaffolds 与 PacBio scaffolds 序列进行比对,根据 scaffolds 之间序列重叠区信息,将上述 scaffolds 组装获得初级线粒体基因组序列。在此过程中,碱基差异部分以二代组装方式 Velvet 得到的序列为主;

 4) 对线粒体基因组初级环首尾连接区域设计引物进行 Sanger 测序。测序结果填 补首尾相连序列,得到线粒体基因组环状序列;

5) 使用 picard MarkDuplicates 去除所有 Illumina 中 reads 的重复序列,利用 SAMtools 将剩余 reads 比对到线粒体基因组完整序列,以正向 reads 与反向 reads 碱基出现的值为参考,利用 Pilon v1.22 对线粒体基因组序列进行校正,得到线 粒体基因组最终序列。



Figure 2.10 Flow for mitogenome sequencing, assembly and annotation.

表 2.3 PacBio 平台 DNA-seq 测序数据统计

Table 2.3 Summary of DNA-seq data sets sequenced using the PacBio Sequel platform

Sample	Readbase	Read_number	Read_length	Read_length
	(GB)		(mean)	(N50)
G1	5.26	496,241	10,598	14,408
HBP	6.45	601,567	10,723	13,493

#### 3.2.1 柑橘线粒体基因组校正前后数据比较

经过三代测序数据校正,'国庆1号'温州蜜柑和 HB 柚线粒体基因组大小 分别为 521,559 bp(图 2.11)和 518,274 bp(图 2.12),比基于二代测序平台组装 的基因组大 31,915 bp和 33,983 bp,且这些序列大多与叶绿体基因组序列同源; GC 含量为 45.06%和 45.14%(表 2.4)。同时,利用 ORFFinder、BLASTX、BLASTN 和 tRNA-SE 等软件对'国庆1号'温州蜜柑和 HB 柚新组装的线粒体基因组进 行功能注释。'国庆1号'温州蜜柑和 HB 柚线粒体基因组分别编码基因 60/59 个,包括 36/35 个基因,21 个 tRNA 基因和 3 个 rRNA 基因(表 2.5),而 HB 柚 线粒体基因组缺少 *rps1* 基因。'国庆1号'温州蜜柑和 HB 柚新组装的线粒体基 因组与仅 Illumina 测序平台得到的线粒体编码基因相比,在'国庆1号'温州蜜 柑和 HB 柚均发现编码 *rps7* 基因序列。此外,'国庆1号'温州蜜柑线粒体 tRNA 编码序列还新发现 Val、Phe, HB 柚编码 tRNA 序列仅新发现 Val。为验证经三 代测序技术校正后的'国庆1号'温州蜜柑和 HB 柚线粒体基因组的准确性,我 们首先随机选取 5 个区域用 PCR 扩增这 5 个区域的连接关系;此外,利用 SAMtools 将 mtDNA 二代 Illumina 测序数据比对到'国庆1号'温州蜜柑和 HB 柚线粒体基因组,统计每条 reads 比对情况及每个碱基位点的测序深度,结果表 明比对到线粒体基因组上两条 reads 之间的距离符合设置阈值的数量占总 reads 数超过 96%,并且每个碱基位点具有相似的测序深度。以上结果表明,经三代测 序技术校正后的'国庆1号'温州蜜柑和 HB 柚线粒体基因组,准确性显著提高。

表 2.4 '国庆 1 号' 温州蜜柑和 HB 柚线粒体基因组比较 Table 2.4 Comparisons of the mitochondrial genomes between G1 and HBP

Feature	G1	HBP
Mitochondrial genome		
Genome size (bp)	521,559	518,274
GC content (%)	45.06	45.14
Length of coding region (%)	37,726 (7.23%)	37,023 (7.14%)
Length of tRNA genes (%)	1,591 (0.31%)	1,591 (0.31%)
Length of rRNA genes (%)	5,195 (1.00%)	5,195 (1.00%)
Length of cis-spliced introns (%)	21,637 (4.15%)	21,807 (4.21%)
Length of intergenic sequences (%)	455,410 (87.32%)	452,658 (87.34%)
Length of repeats (%)	3,686 (0.71%)	7,294 (1.41%)
Longest repeat (bp)	424	1,553
Length of plastid-derived sequences (%)	24,714 (4.73%)	16,285 (3.14%)
Number of rRNA genes	3	3
Number of tRNA genes	21	21
Number of protein genes	36	35
Total genes	60	59





#### Figure 2.11 Circular map of the mitochondrial genome of 'Guoqing No. 1' Satsuma

mandarin. Genes homologous to known protein-encoding genes are shown on the circle. Different classes of conserved protein-encoding genes are assigned with distinct colors. The genes shown inside the circle are transcribed in a counter-clockwise direction. The genes on the outside of the circle are transcribed in a clockwise direction.



图 2.12 HB 袖线粒体基因组结构图。圈两侧标有线粒体保守基因位置和万向,相同颜色万框 表示同一个复合体基因,方框大小代表基因大小,圈内侧基因表示反向转录,圈外侧基因表 示正向转录。

Figure 2.12 Circular map of the mitochondrial genome of 'Hirado Butan' pummelo. Genes homologous to known protein-encoding genes are shown on the circle. Different classes of conserved protein-encoding genes are assigned with distinct colors. The genes shown inside the circle are transcribed in a counter-clockwise direction. The genes on the outside of the circle are transcribed in a clockwise direction.

Intact open reading frames of mitochondrial origin			
Complex I	nad1 nad2 nad3 nad4 nad4L nad5 nad6 nad7 nad9		
Complex II	sdh3 sdh4		
Complex III	cob		
Complex IV	cox1 cox2 cox3		
Complex V	atp1 atp4 atp6 atp8 atp9		
Cytochrome c biogenesis	ccmB ccmC ccmFC ccmFN		
Ribosome large subunit	rpl5 rpl10 rpl16		
Ribosome small subunit	rps1 rps3 rps4 rps7 rps10 rps12 rps14		
Intron maturase	matR mttB		
rRNA genes	rrn5 rrn18 rrn26		
tDNA gamag	<pre>trnM(cau) trnK(uuu) trnQ(uug) trnG(gcc) trnS(uga)(2x) trnE(uuc)</pre>		
tkina genes	trnY(gua) trnF(gaa) trnP(ugg) trnD(guc) trnC(gca) trnI(uau)		
Potential genes of chloroplast of	rigin		
tRNAgnes	<pre>trnN(guu) trnW(cca) trnV(guc) trnM(cau) trnH(gug) trnI(uau)(2x)</pre>		
	trnS(uga)		

#### 表 2.5 柑橘线粒体基因组基因组成

#### Table 2.5 Gene content of cirus mitochondrial DNA

# 3.3 '国庆1号'温州蜜柑和 HB 柚线粒体基因组 RNA 编辑

RNA 编辑在线粒体基因组普遍存在。为研究柑橘线粒体基因组转录信息及RNA 编辑,分别提取'国庆1号'温州蜜柑和HB 柚的叶、花、果皮和汁胞 RNA,进行 RNA 混合测序,每个样品获得不小于11 GB 的 clean reads,测序得到的数据信息见表 2.1。将 RNA 测序数据比对到'国庆1号'温州蜜柑和 HB 柚线粒体基因组,使用 HISAT2、SAMtools、bcftools 和 SnpEff 分析 RNA 编辑位点。通过上述方法在'国庆1号'温州蜜柑和 HB 柚线粒体基因编码区检测到 C-U 的 RNA 编辑位点 303 个和 286 个。非同义突变编辑事件(Nonsynonymous,能使氨基酸序列发生改变)的 RNA 编辑有 265 个(87.4%,G1)和 259 个(90.6%,HBP)。其中,'国庆1号'温州蜜柑和 HB 柚各有 3 个和 2 个 RNA 编辑产生终止密码子,提前终止蛋白翻译。剩余 38 个(12.6%)和 25 个(9.4%)编辑位点属于同义突变(Synonymous)(表 2.6)。

01			0
		<b>G1</b>	HBP
Number of edits and position	Gene	303	286
	Intron	41	16
	Intergenic	38	60
	Downstream	29	32
	Upstream	43	64
Edit site location	First codon	110	114
	Second codon	159	153
	Third codon	34	19
Amino acid change	Synonymous	38	25
	Nonsynonymous	265	261
	Terminator	3	2

表 2.6 柑橘线粒体基因组 RNA 编辑统计 Table 2.6 RNA editing profile of citrus mitochondrial genomes

为进一步研究'国庆1号'温州蜜柑和 HB 柚 RNA 编辑对线粒体氨基酸的 影响,对非同义突变氨基酸变化进行分析表明,'国庆1号'温州蜜柑和 HB 柚 RNA 编辑引起氨基酸变化最多的均为 Ser 变为 Leu,有 62 个和 70 个位点;其次 是 Pro 变为 Leu,有 44 个和 45 个位点;剩余的氨基酸变化位点大于 15 个的有 Ser 变为 Phe, Pro 变为 Ser, Arg 变为 Cys, Arg 变为 Trp, His 变为 Tyr(图 2.13)。 此外,'国庆1号'温州蜜柑和 HB 柚 RNA 编辑位点氨基酸主要在 Ser、Pro 和 Arg 三个氨基酸,编辑后主要变成 Leu 和 Phe。



图 2.13'国庆 1 号'温州蜜柑和 HB 柚线粒体基因组氨基酸变化分布。

Figure 2.13 Distribution of amino acid changes in the mitochondrial genomes of G1 and

HBP.

RNA 编辑位点的密码子分布规律统计发现,RNA 编辑位点在第二位密码子 所占比率最高(>50%),其次是第一位密码子,只有少数 RNA 编辑位点发生在 第三位密码子(图 2.14A)。同时,统计分析发现,柑橘线粒体蛋白编码区 RNA 编辑效率超过 70%的位点占总编辑位点的 60%以上(图 2.14B)。上述结果表明, 与多数植物线粒体基因组类似,柑橘线粒体基因组大量存在 RNA 编辑位点,且 这些位点对蛋白变化可能产生影响。





Figure 2.14 RNA editing efficiency and the location of protein-coding regions in the mitochondrial genomes of G1 and HBP. (A) Distribution of RNA editing across codon regions. (B) Distribution of RNA editing efficiency.

# 3.4 '国庆1号'温州蜜柑和 HB 柚线粒体基因组重复序列

植物同一物种、不同种之间线粒体基因组结构差异可能是进化过程中线粒体 基因组发生重排所致。线粒体基因组重复序列是造成线粒体基因组序列或结构差 异的原因之一。将'国庆1号'温州蜜柑和 HB 柚线粒体基因组进行自身比对, 分析相似度>95%、序列长度>100 bp 的重复序列,分析结果见表 2.7。'国庆1号' 温州蜜柑线粒体基因组重复序列有 8 对,大小在 100 bp-500 bp 之间,最大的重 复序列424 bp。HB 柚线粒体基因组的重复序列有1对>1 kb 的重复序列(1,553 bp), 其余 11 对大小在 100 bp-500 bp 之间。'国庆1号'温州蜜柑和 HB 柚线粒体基 因组重复序列还包含大量 30 bp-100 bp 的短片段重复序列。'国庆1号'温州蜜 柑短重复序列有 50 对、HB 柚有 41 对。长短重复序列数量差异可能是造成'国 庆 1 号'温州蜜柑和 HB 柚线粒体基因组大小差异的主要原因。

No.	Type <sup>a</sup>	Size	<b>Coordinates of Copy-1</b>	Coordinates of Copy-2
		(bp)		
G1				
R1	IR	424	71180-70757	520211-520634
R2	IR	356	418254-417899	390230-390585
R3	IR	272	431945-431674	309642-309913
R4	IR	203	373496-373294	462390-462592
R5	DR	185	298604-298788	63889-64073
R6	DR	156	521404-521559	218883-219038
R7	IR	145	477358-477214	201290-201434
R8	IR	102	477416-477315	211510-211611
Total Leng	gth (bp)	1843		
HBP				
R1	IR	1553	133175-131623	471735-473287
R2	IR	347	122204-121858	3814-4160
R3	DR	279	139402-139680	204654-204932
R4	IR	272	245803-245532	182302-182573
R5	DR	245	205829-206073	438855-439099
R6	IR	189	276500-276312	104921-105109
R7	DR	157	171276-171432	431864-432020
R8	IR	145	396160-396016	369704-369848
R9	DR	128	205700-205827	438726-438853
R10	IR	127	297099-296973	1-127
R11	IR	103	297202-297100	518172-518274
R12	IR	102	369747-369646	146950-147051
Total Leng	gth (bp)	3647		

表 2.7 '国庆 1 号' 温州蜜柑和 HB 柚线粒体基因组中>100 bp 重复序列 Table 2.7 Repeats (>100 bp) identified in the G1 and HBP mitochondrial genomes

# 3.5 '国庆1号'温州蜜柑和 HB 柚线粒体基因组变异和共线性分析

利用 lastz 软件对'国庆1号'温州蜜柑和 HB 柚线粒体基因组进行比对, 根据共线性区段序列相似性>99%原则,'国庆1号'温州蜜柑和 HB 柚相似区域 范围为 48~10,327 bp,片段总长 460,067 bp(图 2.15A)。lastz 比对结果使用 Rstudio 做成点阵图,点阵图结果表明'国庆1号'温州蜜柑线粒体基因组与 HB 柚线粒 体基因组序列存在大量倒位区域(图 2.15B)。此外,'国庆 1 号'温州蜜柑和 HB 柚线粒体基因组还存在 747 个 SNPs 和 246 个 InDels 的变异。其中,线粒体 保守蛋白变异位点有 14 个,分布在 10 个线粒体基因(*atp1、atp8、ccmB、ccmC、ccmFN、nad5、nad7、rpl5、rps3、rps7*),非同义突变有 9 个,同义突变有 5 个 (表 2.8)。这些结果表明'国庆 1 号'温州蜜柑线粒体基因组与 HB 柚线粒体基因组存在重排现象。

A

Mitochondrial genomes of G1 and HBP



图 2.15 '国庆1号'温州蜜柑和 HB 柚线粒体基因组比较及共线性分析。(A)线粒体基因 组比较,蓝色部分代表'国庆1号'温州蜜柑线粒体基因组,紫色部分代表 HB 柚线粒体基 因组;(B)'国庆1号'温州蜜柑和 HB 柚共线性比较。

Figure 2.15 Comparison and collinearity analysis of G1 and HBP mitochondrial genomes. (A) Comparative composition of the mitochondrial genomes from G1 (blue) and HBP (purple); (B) Dot-plot comparisons of the G1 and HBP mitochondrial genomes.

Gene	G1-HBP <sup>a</sup>	G1-HBP <sup>b</sup>	Mutation type <sup>c</sup>	SNP type
	(nucleic acid)	(amino acid)		JI
atp1	1324G-T	442Glu-Ser	Ν	transversion
atp1	1325A-C	442Glu-Ser	Ν	transversion
atp1	1341C-A	448Ile-Leu	Ν	transversion
atp1	1342A-C	448Ile-Leu	Ν	transversion
atp8	309A-C		S	transversion
ccmB	476T-C	159Leu-Pro	Ν	transition
ccmC	574T-C		S	transition
ccmFN	353A-T		S	transversion
ccmFN	1241T-G	444Leu-Arg	Ν	transversion
nad5	24A-C	8Phe-Leu	Ν	transversion
nad7	25C-A	9Gln-Lys	Ν	transversion
rpl5	119G-T	39Thr-Lys	Ν	transversion
rps3	873C-T		S	transition
rps7	360C-T		S	transition

表 2.8 '国庆 1 号'温州蜜柑和 HB 柚线粒体编码基因之间的 SNP Table 2.8 SNPs in protein-encoding mitochondrial genes from G1 and HBP

<sup>a</sup> Position of base mutation

<sup>b</sup> Position of amino acid mutation

<sup>c</sup>S, Synonymous; N, Non-synonymous

# 3.6 '国庆1号'温州蜜柑和 HB 柚叶绿体基因组比较

### 3.6.1 '国庆1号'温州蜜柑和 HB 柚叶绿体基因组信息

以甜橙叶绿体基因组为参考,采用与线粒体拼接组装相似的方法,对'国庆1号'温州蜜柑和 HB 柚叶绿体基因组进行拼接组装,最终组装成四段式结构的叶绿体基因组。'国庆1号'温州蜜柑和 HB 柚叶绿体基因组序列全长分别为160,159 bp/160,590 bp,GC 含量均为38.47%;四段式结构 LSC 区、SSC 区和两个 IR 区序列长度依次为 87,742 bp/87,774 bp、18,437 bp/18,396 bp、26,990 bp/27,210 bp。'国庆1号'温州蜜柑和 HB 柚叶绿体基因组编码基因 134 个,包括 89 个编码蛋白基因,37 个 tRNA 及 8 个 rRNA。其中,LSC 区编码蛋白基因 60 个,tRNA 22 个;两个 IR 区各有一个拷贝的编码蛋白基因 8 个,tRNA 7 个、rRNA 4 个;而 SSC 区有 13 个蛋白编码基因和 1 个 tRNA (图 2.16-2.17)。此外,

'国庆1号'温州蜜柑和HB柚的叶绿体基因均有17个基因含内含子,其中clpP、 rps12和 ycf3含有2个内含子。



图 2.16 '国庆1号'温州蜜柑叶绿体基因组结构图。圈两侧标有线粒体保守基因的位置和 方向,相同颜色的方框表示同一个复合体基因,方框大小代表基因大小,圈内侧基因表示反 向转录,圈外侧基因表示正向转录。

Figure 2.16 Circular map of the chloroplast genome of G1. Genes homologous to known protein-encoding genes are shown on the circle. Different classes of conserved protein-encoding genes are assigned with distinct colors. The genes shown inside the circle are transcribed in a counter-clockwise direction. The genes on the outside of the circle are transcribed in a clockwise direction.





Figure 2.17 Circular map of the chloroplast genome of HBP. Genes homologous to known protein-encoding genes are shown on the circle. Different classes of conserved protein-encoding genes are assigned with distinct colors. The genes shown inside the circle are transcribed in a counter-clockwise direction. The genes on the outside of the circle are transcribed in a clockwise direction.

# 3.6.2 '国庆1号'温州蜜柑和 HB 柚叶绿体基因组变异分析

'国庆1号'温州蜜柑和HB柚叶绿体基因组相互比较发现这两个材料叶绿体基因组之间共线性非常好(图 2.18A)。使用mVISTA软件以甜橙叶绿体基因组为参考序列进行'国庆1号'温州蜜柑和HB柚全局比对分析(图 2.18B),结

果显示'国庆1号'温州蜜柑与甜橙叶绿体基因组相比,变异主要集中在 LSC 区和 SSC 区; HB 柚和甜橙叶绿体基因组之间仅在 LSC 区和 IR 区有少量变异。 进一步对变异序列进行分析,发现这些变异序列集中在非编码区。上述结果表明 柑橘叶绿体基因组较为保守,不同种之间叶绿体基因组没有发生明显重排,非编 码区变异序列可开发分子标记,用于柑橘不同品种鉴定。



图 2.18 '国庆 1 号' 温州蜜柑和 HB 柚共线性分析及 mVISTA 比较。 Figure 2.18 Collinearity analysis of G1 and HBP chloroplast genomes (A), and structure of G1 and HBP chloroplast geneomes constructed using mVISTA (B), with the oranger chloroplast genome as the reference.

# 3.7 '国庆1号'温州蜜柑和 HB 柚线粒体和叶绿体序列转移

一般而言,植物细胞内水平基因转移主要是叶绿体基因组往线粒体基因组的 序列迁移。以甜橙叶绿体基因组为参考,统计'国庆1号'温州蜜柑和 HB 柚线 粒体基因组相似度>80%、序列长度>100 bp 来源叶绿体基因组的序列。'国庆 1 号'温州蜜柑和 HB 柚源自叶绿体基因组片段分别有 13 个和 21 个,序列总长 24,714 bp 和 16,285 bp,占总长 4.73%和 3.14%。进一步对这些来自叶绿体基因 组片段进行区域统计,'国庆 1 号'温州蜜柑和 HB 柚有 7/11 个,总长 13,533 bp/6,904 bp 来自 LSC 区;剩余 11,181 bp 和 9,381 bp 来自 IR 区,没有来自 SSC 区的序列(表 2.9)。上述结果表明相较于 HB 柚线粒体基因组,'国庆 1 号'温 州蜜柑线粒体基因组发生水平基因转移的次数较少。

Table 2.9 Chloroplast sequences identified in citrus mitochondrial genomes				
Fragment	Position in	Fragment	Position in G1	Position in HBP
	cpDNA	Size	mtDNA	mtDNA
1	17869-26857	8990	146132-155095	
2	27603-26844	760		350015-350773
3	40876-44144	3281	81157-84429	
4	40876-41783	908		449148-450055
5	41819-44144	2326		450049-452374
6	47439-47933	502	503965-504450	
7	47935-47439	504		221930-222417
8	48229-48398	177	505167-505343	221036-221212
9	58967-59203	238	407103-407340	303907-304144
10	59248-59456	210	407355-407561	303686-303892
11	61008-59774	1240		359553-360792
12	61391-61287	105		359446-359550
13	61703-61403	301		359153-359446
14	70813-70679	135	478719-478853	148345-148479
15	88394-88842	449	197014-197454	399992-400439
16	95547-96590	1044		358110-359153
17	100225-99933	299		169969-170267
18	100474-109781	9310	233155-242451	
19	103393-100474	2920		515266-518185
20	104846-104057	790		362158-362947
21	105886-106004	119	289602-289720	289921-290039
22	106301-106418	118	290387-290503	290706-290822
23	109776-110712	937	262337-263273	491931-492867
24	109781-107325	2457		512820-515276
25	111359-111606	248	321586-321833	111570-111817

表 2.9 柑橘线粒体基因组中鉴定到的叶绿体序列

4 讨论

# 4.1 柑橘线粒体 DNA 提取

植物黄化幼苗中类囊体膜含量高,叶绿体含量低,是提取线粒体 DNA 的理 想试材。线粒体遗传属于母系遗传,不论正交或反交,子代线粒体 DNA 序列和 母本一致。因此,通过对不育系授粉获得种子,培养黄化幼苗并提取线粒体 DNA, 是研究不同材料线粒体基因组信息切实可行的方法。

目前,线粒体 DNA 提取方法主要有差速离心法和密度梯度离心法(蔗糖和 percol)(陈健美 2011)。其中,蔗糖密度梯度离心与 percol 梯度离心相结合的方法能更大程度上消除质体 DNA 污染,但该方法步骤复杂、耗时长且需要大量试材。本实验室霍岩(2008)早期基于密度梯度离心和差速离心法构建了柑橘愈伤组织线粒体 DNA 提取技术,并尝试提取柑橘叶片线粒体 DNA。然而,以叶片为试材,提取的线粒体 DNA 浓度低、质量差。本研究基于霍岩(2008)提取柑橘愈伤组织线粒体 DNA 的方法,综合蔗糖梯度离心技术要点,改进了柑橘线粒体 DNA 提取方法,并通过细胞学、Western 及分子标记等技术对提取的线粒体及线粒体 DNA 进行检测,证明该方法提取的线粒体 DNA 可用于 Illumina 测序平台的建库测序。

虽然柑橘材料可以通过胚珠诱导获得胚性愈伤组织,为线粒体 DNA 提取提供材料,但不具有未发育胚珠的柑橘单胚材料无法诱导获得胚性愈伤组织。此外,柑橘属木本植物,黄化幼苗获取较其它大田作物更加困难,且无核品种黄化幼苗获得过程更为复杂,需要通过杂交授粉。因此,柑橘黄化幼苗材料用量及保存方式选取对线粒体 DNA 提取显得尤为重要。本研究对不同质量、不同保存方式的线粒体及线粒体 DNA 获得率进行比较,发现无论 30 g 还是 60 g 的黄化幼苗均能获得高质量线粒体,但 30 g 黄化幼苗得到的线粒体难以提取到足量线粒体DNA 用于 Illumina 测序。同时,冰冻材料提取过程中发现冻融的黄化幼苗细胞结构破坏,导致线粒体膜受损,无法提取到完整线粒体。因此,柑橘黄化幼苗需要至少 60 g 鲜样提取的线粒体 DNA 可用于 Illumina 测序。

# 4.2 柑橘线粒体基因组组装

目前,已报道的被子植物线粒体基因组大多是利用 Illumina 测序技术对植物 线粒体 DNA 或者总 DNA 进行测序,以细胞核和叶绿体基因组信息为参考,剔 除原始 reads 中细胞核和叶绿体 reads 干扰, 对从原始 reads 中分离得到的线粒体 基因 reads 进行拼接得到环状线粒体基因组序列(Van et al 2016; Lee et al 2018; Makarenko et al 2018)。这种测序数据只能将线粒体基因组组装出几个甚至十几 个候选片段, 需要 BAC 文库或基于链式聚合酶反应的 Sanger 测序技术提供序列 信息连接这些候选片段。然而,线粒体基因组中的重复序列和外源整合序列对线 粒体组装造成一定影响。本研究中,首先提取'国庆1号'温州蜜柑愈伤组织和 HB 柚黄化苗线粒体 DNA,利用二代测序技术对其线粒体基因组测序;随后,剔 除原始 reads 中核基因组和叶绿体基因组序列干扰, 对测序结果进行拼接并用 Sanger 测序技术填补 gap。尽管最终获得'国庆1号'温州蜜柑和 HB 柚完整线 粒体基因组序列,但是在这两个线粒体基因组序列准确性验证过程中发现,'国 庆1号'温州蜜柑和 HB 袖线粒体基因组与测序原始序列之间有大量倒位区域, 进一步分析发现这些错配区域序列是柑橘线粒体重复序列造成。这表明在缺少线 粒体参考基因组的情况下, 仅通过 Illumina + Sanger 测序技术相结合的方法是无 法获得高质量线粒体基因组序列。

PacBio 测序技术的 long-reads 能有效解决线粒体基因组大片段重复序列对线 粒体基因组组装带来的干扰。基于 PacBio 和 Illumina 测序数据混合分析, Sanger 测序技术填补 gap 的策略,能准确得到不同植物线粒体基因组完整序列(Wang et al 2019; Wu et al 2019)。本研究中,用'国庆1号'温州蜜柑、HB 柚 PacBio 和 Illumina 测序数据杂交得到的候选 scaffolds 长度均超过 100 kb,而这两个品种线 粒体基因组重复序列和叶绿体来源片段总长为 1,843 bp/3,647 bp,24,714 bp/16,285 bp,表明 long-reads 能消除线粒体基因组中的重复序列及叶绿体来源片 段干扰,从而降低了线粒体基因组拼接过程中连接错误的比例,减少剔除叶绿体 基因组 reads 造成的线粒体基因组中与叶绿体基因组同源序列的丢失,使校正后 的线粒体基因组序列更准确。

# 4.3 柑橘线粒体基因组结构

目前,Yu等(2018)基于细胞内不同基因组 raw reads 测序深度,利用已公 布甜橙 DNA 测序数据,获得大小为 640.9 kb 甜橙线粒体基因组。相比于该甜橙 线粒体基因组,本研究提取柑橘线粒体 DNA 和总 DNA,并融合一代、二代和三 代测序技术的优点,获得高质量柑橘线粒体基因组。其中,'国庆1号'温州蜜 柑(521.6 kb)和HB柚(518.3 kb)完整线粒体基因组可以作为温州蜜柑和柚线 粒体参考基因组。与其它已发表的线粒体基因组一样(Unseld et al 1997; Goremykin et al 2009, 2012), 柑橘线粒体基因组编码基因包括呼吸电子传递链组 成元件、两种内含子成熟酶、几种核糖体蛋白及 tRNAs。柑橘线粒体 DNA 不能 编码识别所有密码子所需的全部 tRNAs,可能依赖核编码的 tRNA 进入线粒体中 进行基因表达,且部分线粒体 DNA 编码的 tRNA 可能来源于叶绿体,如 trnN、 trnW。相关研究表明,植物细胞内遗传信息的传递分为不同物种间的基因水平转 移(HGT)和物种本身的细胞内水平转移(IGT)。被子植物基因水平转移现象 可能发生在不同物种的叶绿体和线粒体之间、也可能发生在线粒体之间、很少在 叶绿体之间发生 (Rice et al 2013; Chen et al 2017); 而细胞内水平转移是 DNA 在 细胞核和线粒体之间双向传播,也可以从叶绿体传播到细胞核和线粒体 (Richardson and Palmer 2007)。本研究中发现'国庆1号'温州蜜柑和 HB 柚线 粒体基因组序列与叶绿体同源的长度分别为 24.7 kb 和 16.3 kb,可能是'国庆 1 号'温州蜜柑和 HB 柚进化过程中叶绿体基因组转移至线粒体基因组序列长度大

小、数量不同所致。

线粒体基因组中存在大量重复序列,重复序列间重组能影响线粒体的功能及 生物性状。一般而言,不育系线粒体基因组比可育系线粒体基因组大2kb~100kb, 且重复序列的数量和总长都超过可育系(Tanaka et al 2012; Li et al 2018)。虽然 不育材料'国庆1号'温州蜜柑线粒体基因组比可育材料 HB 柚大3kb,但是线 粒体重复序列的总长和数目是 HB 柚比'国庆1号'温州蜜柑多,然而'国庆1 号'温州蜜柑和 HB 柚属于柑橘不同种(温州蜜柑和柚)。因此,不育系与可育 系线粒体基因组大小和重复序列的数量及长度关系不能在'国庆1号'温州蜜柑 和 HB 柚之间关联。'国庆1号'温州蜜柑线粒体基因组中与叶绿体基因组同源

的序列比 HB 柚长 8.4 kb,这可能是'国庆 1 号'温州蜜柑和 HB 柚线粒体基因 组大小差异的主要原因之一。

RNA 编辑在线粒体基因组中普遍存在,能对线粒体基因的分子功能和植物 生理过程产生影响(Yang et al 2017; He et al 2018)。RNA 编辑位点鉴定主要是通 过 cDNA 序列和 RNA-seq 短 reads 与模板序列比较的策略鉴定 RNA 编辑位点 (Takenaka and Brennicke 2007; Picardi et al 2010)。本研究中,采用 RNA-seq 数 据与'国庆1号'温州蜜柑和 HB 柚线粒体基因组序列比对的方法,鉴定这两个 材料 RNA 编辑位点。'国庆1号'温州蜜柑和 HB 柚鉴定到线粒体编码区和非编 码区 RNA 编辑的数量与西瓜相近 (Alverson et al 2010),产生的密码子及氨基酸 变化频率和变化规律与葡萄 (Picardi et al 2010)、猕猴桃 (Wang et al 2019)等作 物线粒体 RNA 编辑的规律相一致。此外,'国庆1号'温州蜜柑和 HB 柚线粒体 RNA 编辑还产生不同数量导致蛋白翻译提前终止的密码子。值得注意的是,水 稻中发现线粒体 RNA 编辑引起水稻花粉育性降低 (Xiao et al 2018)。因此,'国 庆1号'温州蜜柑和 HB 柚线粒体基因编码区 RNA 编辑的差异是否对这两个材 料花粉育性差异产生影响还需深入探索。

# 第三章 HB 柚胞质杂种基因组来源分析及候选 CMS 基因发 掘与功能解析

# 1 引言

细胞质雄性不育线粒体基因及恢复基因在大田作物已有相关报道(Kim and Zhang 2018)。相较于细胞核雄性不育基因在不同作物中有同源基因, CMS 基因 在不同 CMS 系统中几乎没有同源性。CMS 基因筛选常用的方法有三种:第一, 根据线粒体保守基因序列设计探针,利用 RNA 杂交技术在成套"三系"材料中 分析线粒体保守基因转录本的差异表达,筛选可能与 CMS 有关的线粒体基因 (Chen and Liu 2014); 第二,利用蛋白质组学技术,分析可育系与保持系线粒 体保守蛋白表达谱差异,确定与差异蛋白表达有关基因,也是筛选 CMS 基因的 重要手段(Kim et al 2019): 第三, 直接比较不育系和可育系线粒体基因组, 从 MSS 区筛选不育特异 ORF, 是发掘 CMS 基因最直接的方法(Xie et al 2018)。 筛选到的 CMS 基因,可以通过构建带有线粒体信号肽载体的转基因实验进行验 证。然而,对于大多数作物,这些方法存在一定的缺陷。首先,很多作物缺少天 然成套的"三系"材料:其次,虽然体细胞杂交技术可以克服不同亲缘(近缘和 远缘)杂交不亲和障碍并创制遗传 CMS 系线粒体基因组的不育新种质,但多数 植物利用体细胞杂交技术获得胞质杂种存在困难。即使获得花粉败育的胞质杂种, 胞质杂种能否完全遗传不育系的线粒体基因组且不发生同源重组,也充满不确定 性。

柑橘愈伤组织原生质体与叶肉原生质体融合获得的体细胞杂种和胞质杂种的叶绿体基因组表现为随机遗传,线粒体基因组多数情况来自愈伤组织亲本,也有部分融合后代线粒体基因组是双亲线粒体重组的结果(Dambier et al 2011; Aleza et al 2016);胞质杂种核基因组均来自叶肉亲本(Guo et al 2013; Xiao et al 2014)。这些发现是基于不同基因组 SSR 分子标记分析的结果。愈伤组织亲本线粒体基因组遗传给胞质杂种后代过程中,线粒体基因组是否发生重组以及是否融合叶肉亲本线粒体基因组序列还有待研究。以温州蜜柑作为愈伤组织亲本,柑橘不同品种为叶肉原生质体亲本,通过对称融合将温州蜜柑的不育细胞质转移至叶

肉亲本型胞质杂种后代,获得保持温州蜜柑不育性状或育性恢复的新种质,为发掘柑橘细胞质雄性不育及恢复基因和研究柑橘细胞质雄性不育机理提供了珍贵试材(郑蓓蓓 2015)。

本研究基于 HB 柚胞质杂种及其融合双亲核基因组和胞质基因组信息,明确胞质杂种核质基因组来源。通过比较分析不育材料(HB 柚胞质杂种、'国庆1号'温州蜜柑)相比可育材料(HB 柚)MSS 区和线粒体特异 ORFs,结合不育材料线粒体转录组信息,分离到包含 CMS 系特异 ORFs (orf55、orf59、orf88)和 ψrps1 的差异嵌合转录本 DN16978;构建含有线粒体定位信号肽的 DN16978 超表达载体,异源表达拟南芥获得的转基因植株表现出一些不育表型;亚细胞定位预测和跨膜结构域预测结果推测 orf88 可能是 HB 柚胞质杂种 CMS 不育基因。 首先,利用酵母双杂技术筛选与转录本 DN16978 互作的核编码蛋白;随后针对 候选基因 orf88 和筛选到的 CsNADH 基因,利用酵母双杂交点对点实验以及 LUC 荧光互补实验验证二者的互作关系。此外,利用烟草瞬时表达技术对 CsNADH 和 orf88 进行亚细胞定位。

# 2 材料和方法

# 2.1 材料

两个胞质杂种材料 HB 柚胞质杂种'华柚 2 号'(The cybrid between *C.unshiu* Marc. 'Guoqing No. 1' and 'Hirado Buntan' pummelo, G1+HBP, 雄性不育), 冰糖 橙胞质杂种 'G1+冰糖橙'(The cybrid between *C.unshiu* Marc. 'Guoqing No. 1' and *Citrus sinensis* Osbeck 'Bingtang Orange', G1+BTC, 雄性可育)种植于华中农业 大学柑橘研究所柑橘种质资源圃(图 3.1A)。遗传转化所用拟南芥为哥伦比亚野 生型拟南芥(*Arabidopsis thaliana*), 瞬时转化所用烟草为本式烟草(*Nicotiana bethamiana*)。



**图 3.1 柑橘果实。**(A) 冰糖橙胞质杂种果实,标尺 2 cm; (B) HB 柚胞质杂种×沙田柚授 粉果实; (C) HB 柚胞质杂种未授粉果实,标尺 3 cm。 **Figure 3.1 The fruits of citrus.** (A) The fruit of G1+BTC, Bar =2 cm; (B) The fruit of the G1+HBP from 'G1+HBP' × 'Shatian' pummelo; (C) The fruit of G1+HBP, Bar =3 cm.

# 2.2 方法

## 2.2.1 HB 柚胞质杂种人工授粉

花粉制备和人工授粉参考解凯东等(2013)的方法。2015 年 4 月下旬采取 HB 柚胞质杂种×沙田柚的组合人工授粉,授粉 142 朵,获得 30 个果实(图 3.1B)。 果实成熟后取 16 个果实,总共 720 余粒种子播种在试管中,培养黄化苗。

# 2.2.2 LncRNA 转录本拼接及预测

Trinity 软件无参拼接总 RNA(去除 rRNA)高通量测序数据,得到总 RNA (去除 rRNA)无参转录本序列文件 Trinity.fasta;同时使用 HISAT2 + StringTie 软件有参拼接总 RNA(去除 rRNA)高通量测序数据(结合线粒体基因组序列和 柚参考基因组注释文件),得到线粒体有参转录本序列文件 stringtie.fa 和线粒体 有参转录本注释文件 stringtie.gtf,其中 Trinity、HISAT2及 Stringtie 软件设置参 数为默认值。使用 PASA 软件将 Trinity.fasta 和 stringtie.fa 比对到线粒体基因组上, 参考 stringtie.gtf,整合分析获得线粒体综合转录组数据库序列文件 transcripts.fasta,设置的主要参数为:min\_per\_ID=95,min\_per\_aligned=30。采用 tblastn (e-value <1e-6)将线粒体综合转录组数据库序列文件 transcripts.fasta 比对 到植物线粒体基因组注释数据库 Mitofy,得到包含线粒体基因组编码蛋白的线粒体综合转录组数据。

## 2.2.3 HB 柚胞质杂种及其亲本与甜橙参考基因组变异分析

HB 柚胞质杂种、'国庆1号' 温州蜜柑和 HB 柚利用 Illumina HisSeq 测序获 得总 DNA clean reads 通过 BWA 软件比对到甜橙参考基因组,比对结果经 SAMtools 去除重复 reads,并按照 SNP 的 reads 支持数不低于 4 和 SNP 的质量值 (MQ)不低于 20 进行过滤,最终检测个体 SNP 和长度<50 bp 的小片段插入与 缺失(InDel),利用 ANNOVAR 对 SNP 和 InDel 检测结果进行注释。

## 2.2.4 柑橘不育材料特异 ORFs 鉴定

- 线粒体 ORFs 预测:使用软件 Unipro UGENE(Okonechnikov et al 2012)预测 HB 柚胞质杂种与其亲本'国庆1号'温州蜜柑和 HB 柚线粒体基因组可能编 码的 ORFs(氨基酸大小>50);
- 鉴定不育系相比可育系不完全相似的 ORFs:不育系 HB 柚胞质杂种与'国庆 1号'温州蜜柑预测的线粒体 ORFs 分别与可育系 HB 柚预测的线粒体 ORFs 比较(blastn e-value <1e-6),剔除长度一致,相似度 100%的 ORFs,得到不育 系相比可育系不完全相同的 ORFs;
- 3. 线粒体型特异序列(MSS)鉴定:不育系 HB 柚胞质杂种与'国庆1号'温州 蜜柑和可育系 HB 柚线粒体基因组序列比较(blastn e-value <1e-6),分别鉴定 相比 HB 柚线粒体基因组,HB 柚胞质杂种与'国庆1号'温州蜜柑的线粒体 型特异序列(Mitotype-specific sequence, MSS),要求 MSS 长度>100 bp(Xie et al 2014);
- 4. 鉴定包含不育系特异 ORFs 的嵌合转录本:为鉴定包含 ORFs 的嵌合转录本, 分别将不育系 HB 柚胞质杂种与'国庆1号'温州蜜柑相比可育系 HB 柚不完 全相同的 ORFs 比对到各自线粒体转录本,获得不育系特异 ORFs,且这些 ORFs 所在转录本含有线粒体基因编码序列,称"嵌合转录本"。比对不育 HB 柚胞 质杂种与'国庆1号'温州蜜柑相比可育系 HB 柚不完全相同的 ORFs、嵌合

转录本及 HB 柚胞质杂种与'国庆 1 号'温州蜜柑相比 HB 柚的 MSS,获得 HB 柚胞质杂种与'国庆 1 号'温州蜜柑相比 HB 柚特异 ORFs。

## 2.2.5 农杆菌介导的拟南芥遗传转化

农杆菌介导的拟南芥遗传转化方法参照郑蓓蓓(2015),略有修改。

- 1. 将 9 cm 滤纸浸湿, 放置于一次性塑料皿;
- 2. 拟南芥种子悬浮于 ddH<sub>2</sub>O, 用滴管点播于滤纸表面;
- 3.22 ℃ 培养 5-7 d, 待种子发芽且长出两片子叶后,将幼苗移入基质土,覆膜保湿 3 d;
- 4. 生长约 30 d 对植株进行剪薹,以保证植株花序数量及开花时间一致,约一周 后进行遗传转化;
- 5. 将-80 ℃ 保存的带有目的重组质粒的农杆菌进行活化,划线于加有固体 LB 的 培养皿(加抗生素),28 ℃ 培养 36-48 h;
- 6. 挑单克隆于加有 500 μL 液体 LB 培养基(加抗生素)的 1.5 mL 离心管,于 28 °C 震荡培养 24 h;
- 7. 吸 100 μL 菌液加入到 75 mL 液体 LB 培养基(加抗生素),于 28 ℃ 震荡培养 18-20 h;
- 8.4000 rpm 常温离心 15 min, 收集菌液;
- 9. 加入一定量转化渗透液(1/2 MS + 0.02% Silwet L-77) 悬浮菌液,测定菌液 OD 值至 0.6-0.8;
- 10. 拟南芥花序浸入农杆菌渗透液 90 sec, 侵染完成后平放置白色托盘黑暗培养 并覆膜保湿 1 d, 随后转入光下正常培养;
- 11. 根据植株新花序数量进行判断,确定后续转化时间与次数,通常一个基因转化 2-3 次;
- 12. 待种子完全成熟后收集, 常温保存。

# 2.2.6 拟南芥阳性苗筛选

1. 取一定量拟南芥转化种子倒入无菌 10 mL 离心管,加入 75%乙醇,上下晃动 10 mL 离心管,13 min,静置 1 min 后弃上清;

- 2. 加入100%无水乙醇后,重复上述步骤;
- 3. 用无菌水清洗种子 3-5 次,洗净种子表面的酒精;
- 4. 加入 1-2 mL 无菌水,将种子小心倒入加有 1/2 MS(400 μg/mL 头孢霉素和 30 μg/mL 潮霉素)培养基的培养皿,用 1 mL 移液器吸打均匀,并吸干多余水分, 在超净工作台内吹 1-2 h,直至表面水分吹干;
- 5. 光下培养 7-10 d, 待种子长出真叶且根系发达,移栽至基质中培养,覆膜保湿 3-5 d;
- 6. 待株系营养生长完成,开始生殖生长,取少量基叶提取 DNA,鉴定阳性,对 阳性株系进行表型鉴定并单株收种。

#### 2.2.7 目的蛋白原核表达

以HB 柚胞质杂种 cDNA 为模板,设计引物扩增 orf88 全长序列(引物序列 见附录III)。引物序列 5'端添加 EcoR I,3'端添加 Xho I 酶切位点。利用 ClonExpress II One Step Cloning 酶采用一步法将目的片段连接到原核表达载体 pET28a,构建 重组载体 pET28a-orf88。将原核表达重组载体转化至大肠杆菌表达菌株 BL21

(DE3),挑取单克隆 37 °C 培养过夜后,鉴定阳性。将阳性克隆按照 1:100 接种 到两瓶 50 mL 含有卡那霉素抗性的液体 LB 培养基, 37 °C、200 rpm/min 培养至 OD 值为 0.6-0.8,向其中一份加入 1.0 mmol/L 的 IPTG 诱导目的蛋白表达;每隔 30 min 用紫外分光光度计检测菌液 OD 值。

#### 2.2.8 转录本 DN16978 诱饵载体构建

根据 DN16978 转录本序列设计引物,5'端和 3'端添加接头序列(引物序列 见附录 III)。以 HB 柚胞质杂种叶片 cDNA 为模板扩增目的片段,限制性内切酶 *EcoR* I 和 *Not* I 将 pGBDT7 线性化,利用 ClonExpress II One Step Cloning 酶采用 一步法将目的片段连入载体 pGBDT7,构建成 BD 重组载体。

## 2.2.9 转录本 DN16978 互作蛋白筛选

通过酵母 mating 方法对转录本 DN16978 互作蛋白进行筛选, mating 筛选方 法参照肖雪硕士论文 (2017), 具体步骤如下:

- 1. 提前 1 天将钓饵菌液在 30 ℃、200 rpm 活化,获得 OD 值约为 0.8 的新鲜菌液, 3000 rpm/min 离心 2 min 富集菌液,随后加入 5 mL SD/-Trp 液体培养基备用;
- 2. 取-80 ℃保存的AH109文库菌1管置于冰上冻融,将上述钓饵菌液与文库菌 混匀并倒入2L无菌锥形瓶进行杂交;
- 用 1 mL 2x YPDA/Kan (50 μg/mL)洗涤文库离心管并把洗涤液倒入 2 L 锥形 瓶,重复 1 次,再加入 45 mL 2x YPDA 液体培养基于锥形瓶中混匀菌液;
- 4. 30 °C 40 rpm/min 培养 18-24 h;
- 5. 培养18h后,取100μL 菌液滴入载破片体式显微镜下观察杂交情况,若观察 到有大量结合体,则进行后面操作,若无则继续培养4h以上,直到有大量结 合体出现;
- 6. 将杂交后的菌液转移至 50 mL 灭菌离心管中, 3000 rpm/min, 离心 10 min, 弃上清;
- 7. 用 50 mL 0.5x YPDA/Kan (50 μg/mL)洗涤 mating 用过的锥形瓶,洗涤后的液体回收至上一步的 50 mL 离心管, 3000 rpm/min,离心 10 min,弃上清;
- 8. 用 10 mL 0.5x YPDA/Kan (50 μg/mL) 悬浮沉淀,吸打混匀后取 200 μL 涂于 四缺培养皿 (SD/-Trp-His-Leu-Ade);
- 9.30°C 培养 3-7 d,挑取单克隆阳性检测,并将阳性 PCR 产物送公司测序。

## 2.2.10 酵母互作蛋白点对点验证

根据 mating 筛选得到可能互作蛋白的基因序列设计引物(引物序列见附录 III),以 HB 柚胞质杂种叶片 cDNA 为模板 PCR 扩增目的基因,限制性内切酶 *EcoR* I和 *Xho* I将 pGADT7 线性化,采用一步法将目的片段连入载体 pGADT7, 构建成 AD 重组载体。将重组 AD 载体与 pGBKT7-16978 载体共同转化至酵母 AH109,涂在二缺(SD/-Trp-Leu)和四缺(SD/-Trp-Leu-His-Ade)固体培养基。 30 °C 培养 3-5 d,挑取单克隆摇菌并检测阳性。吸取阳性克隆菌液 5 μL 点在加 有 X-α-gal 的四缺培养基 30 °C 培养 3-5 d,观察是否有蓝色克隆出现,以此确定 两者互作关系。

#### 2.2.11 LUC 荧光检测实验

基于酵母点对点互作蛋白验证结果,利用 LUC 荧光检测实验进一步验证互 作蛋白之间的互作关系。根据基因序列设计引物(引物序列见附录 III), PCR 扩 增基因序列,将 orf88 连入载体 pCAMBIA-cLUC, CsNADH 连入载体 pCAMBIA-nLUC。其中,载体 pCAMBIA-nLUC 带有荧光酶蛋白(LUC) N 端 序列,与 orf88 形成融合蛋白;载体 pCAMBIA-cLUC 带有荧光酶蛋白(LUC) C端序列,与CsNADH形成融合蛋白。将载体pCAMBIA-nLUC、pCAMBIA-cLUC 及构建完成的重组载体转入农杆菌(GV3101), 28 ℃ 培养 12-16 h 后离心富集菌 体, 用 10 mM MES (含有 10 mM MgCl<sub>2</sub>和 0.5 μM 的乙酰丁香酮) 重悬菌液, 调整 OD600 为 0.4-0.6, 28 ℃ 黑暗静置 2 h 以上。分别将 pCAMBIA-nLUC 与 与 pCAMBIA-cLUC, pCAMBIA-orf88-cLUC pCAMBIA-nLUC, pCAMBIA-CsNADH-nLUC 与 pCAMBIA-cLUC 及 pCAMBIA-orf88-cLUC 与 pCAMBIA-CsNADH-nLUC 以 1:1 比例混匀,通过注射器将这些组合注入同一叶 片的四个不同区域,放在 28 ℃ 环境下培养 48-60 h 后,用 Roper 仪器检测荧光 酶蛋白活性。具体操作如下:

1. 取注射农杆菌的叶片置于 12 mm 方形培养皿;

- 2. 根据叶片大小加入 50-100 μL Beetle Luciferin 试剂,涂抹均匀后室温条件下黑 暗放置 5-10 min;
- 3. 使用 Roper 仪器检测荧光酶活性;
- 4. 通过调节曝光值,来表示 LUC 信号强弱。

#### 2.2.12 orf88 与 CsNADH 亚细胞定位

根据 orf88 与 CsNADH 序列设计引物,5'端和 3'端添加 Gateway 载体同源序 列(引物序列见附录 III)。以 HB 柚胞质杂种叶片 cDNA 为模板,PCR 扩增后用 Gateway 同源重组法连接到亚细胞定位载体 pCAMBIA-cYFP,得到重组载体 pCAMBIA-orf88-cYFP 和 pCAMBIA-CsNADH-cYFP。利用已报道的 mt-rk 作为 线粒体定位 marker (Nelson et al 2007)。农杆菌介导烟草瞬时转化方法同 LUC 荧光检测实验描述。取注入菌液叶片注射孔附近的部位放入载玻片,滴加少许蒸 馏水盖上盖玻片,激光共聚焦显微镜(Leica, Germany)下观察荧光并拍照。

## 2.2.13 NAD<sup>+</sup>与 NADH 含量测定

HB 柚胞质杂种、'国庆1号'温州蜜柑和 HB 柚花药 NAD<sup>+</sup>、NADH 含量测 定使用辅酶 I NAD(H)含量检测测定试剂盒(#BC0310, Solaribo,中国),具体 流程按照操作指南进行。过程如下:

- 1. 取 0.002 g 花药液氮研磨加 500 μL 提取液 (NAD<sup>+</sup>为酸性提取液, NADH 为碱 性提取液), 沸水浴 5 min, 冰浴冷却后 4 °C, 10,000 g 离心 10 min;
- 2. 取上清 200 μL 加等体积(500 μL)提取液(NAD<sup>+</sup>碱性提取液, NADH 为酸性 提取液)混匀, 4°C, 10,000 g 离心 10 min, 取上清冰上保存;

-	试剂名称	对照管	测定管	标准管	空白管
-	上清液	50	50		
	标准溶液			50	
	蒸馏水				50
	试剂六	500			
	试剂一	250	250	250	250
	试剂二	75	75	75	75
	试剂三	75	75	75	75
	试剂四	75	75	75	75
	试剂五	35	35	35	35
		充分混匀,	室温避光青	浄置 20 min	
	试剂六		500	500	500
	充分混匀,	静置 5 min,	15000 rpm	, 25 °C 离	心 15 min,
			弃上清		
	试剂七	1000	1000	1000	1000

3. 按以下配方依次往 1.5 mL 棕色离心管中加样:

4. 混匀样品在 OD570 nm 下比色测定。

注意事项:

1. 对照管和测定管为同一样品不同处理;

2.操作过程从试剂三开始避光,且不能将试剂一、二、三和四混合,需分开添加;
 3.吸光值大于1时,需稀释后测定。
样品测定公式: NAD<sup>+</sup>(nmol/g 鲜重)= $\Delta$ A 测定÷( $\Delta$ A 标准 1÷C 标)×V 提取 ÷W=1.25× $\Delta$ A 测定÷ $\Delta$ A 标准 1÷W。其中, $\Delta$ A 测定=测定管-对照管, $\Delta$ A 标准= 标准管-空白管

3 结果

### 3.1 HB 柚胞质杂种核质基因组与其融合双亲比较分析

#### 3.1.1 HB 柚胞质杂种线粒体基因组结构

HB 柚胞质杂种黄化幼苗培养、线粒体 DNA 及总 DNA 提取的材料方法、测序技术及拼接组装流程同 HB 柚。以'国庆1号'温州蜜柑线粒体基因组为参考 基因组,对 HB 柚胞质杂种线粒体基因组进行拼接组装,最终组装成环状式结构 的线粒体基因组(图 3.2)。HB 柚胞质杂种线粒体基因组序列全长为 538,434 bp, GC 含量为 45.02%。HB 柚胞质杂种线粒体基因组编码的基因种类与'国庆1号' 温州蜜柑完全一致(表 2.5)。HB 柚胞质杂种线粒体基因组相似度>95%、序列长 度>100 bp 的重复序列有 8 对,除 1 对 16,849 bp 可能是融合过程中亲本线粒体基 因组融合产生的新重复序列外,其余重复序列与'国庆1号'温州蜜柑完全相同。 HB 柚胞质杂种粒体编码区 C-U 的 RNA 编辑位点有 284 个。其中,230 个编辑 位点为非同义突变,包括2 个提前终止;剩余 43 个编辑位点为同义突变(表 3.1)。

		G1+HBP
Number of edits and position	Gene	284
	Intron	27
	Intergenic	49
	Downstream	37
	Upstream	48
Edit site location	First codon	94 (33.1%)
	Second codon	153 (53.9%)
	Third codon	37 (13.0%)
Amino acid change	Synonymous	41
	Nonsynonymous	230
	Terminator	2

表 3.1 HB 柚胞质杂种线粒体基因组 RNA 编辑统计 Table 3.1 RNA editing profile of the G1+HBP mitochondrial genomes



图 3.2 HB 柏胞质杂种线粒体基因组结构图。圈两侧标有线粒体保守基因的位置和方向,相同颜色方框表示同一个复合体基因,方框大小代表基因大小,圈内侧基因表示反向转录,圈 外侧基因表示正向转录。

**Figure 3.2 Circular map of the mitochondrial genome of G1+HBP.** Genes homologous to known protein-encoding genes are shown on the circle. Different classes of conserved protein-encoding genes are assigned with distinct colors. The genes shown inside the circle are transcribed in a counter-clockwise direction. The genes on the outside of the circle are transcribed in a clockwise direction.

#### 3.1.2 HB 柚胞质杂种与其融合双亲线粒体基因组比较

将 HB 柚胞质杂种线粒体基因组与融合双亲'国庆1号'温州蜜柑和 HB 柚 线粒体基因组进行两两比较,并对 HB 柚胞质杂种线粒体基因组序列与'国庆1 号'温州蜜柑和 HB 柚相似序列进行区域划分,分为:完全来自'国庆1号'温 州蜜柑的片段 61.5 kb,'国庆1号'温州蜜柑与 HB 柚共有片段 460.1 kb,以及 同源重组新形成的 16.8 kb。同源重组新片段可拆分为仅来自'国庆1号'温州 蜜柑的片段 4.6 kb 和双亲共有的片段 12.2 kb,双亲共有片段在 HB 柚线粒体基 因组由 9.7 kb 和 2.5 kb 两部分组成(图 3.3A-B)。进一步对 HB 柚胞质杂种线粒 体基因组序列与'国庆1号'温州蜜柑和 HB 柚进行共线性分析。从共线性点阵 图可以看出,HB 柚胞质杂种线粒体基因组与'国庆1号'温州蜜柑线粒体基因 组高度一致,16.8 kb 同源重组新片段能在'国庆1号'温州蜜柑线粒体基因组 找到同源区域(图 3.3C);HB 柚胞质杂种线粒体基因组与 HB 柚线粒体基因组 相比,至少有29个重排区域,其中28个重排区域与'国庆1号'温州蜜柑相比 HB 柚线粒体基因组重排区域一致(图 3.3B)。此外,HB 柚胞质杂种线粒体基因 组与'国庆1号'温州蜜柑有5个 SNPs 和1个 InDel 变异位点,与HB 柚有743 个 SNPs 和 244 个 InDels 变异位点,表明 HB 柚胞质杂种线粒体基因组来自'国 庆1号'温州蜜柑(表 3.2)。



图 3.3 HB 柚胞质杂种线粒体基因组组成及共线性分析。(A) HB 柚胞质杂种线粒体基因组 组成,蓝色部分表示仅来自'国庆1号'温州蜜柑区域,绿色部分表示亲本共有区域;(B) HB 柚胞质杂种和 HB 柚线粒体基因组总共有 472.3 kb(87.7%)序列高度相似,但存在大量 重排区域。HB 柚胞质杂种额外 16.8 kb 中的 12.2 kb 在 HB 柚线粒体基因组由 9.7 kb 和 2.5 kb 两部分组成;(C) HB 柚胞质杂种和'国庆1号'温州蜜柑总共有 96.9% (521.6 kb)线粒体 基因组序列高度一致。HB 柚胞质杂种和'国庆1号'温州蜜柑存在 SNP 和 InDel 差异和一 段额外的 16.8 kb 序列。

Figure 3.3 Mitochondrial genome composition of the cybrid and dot-plot comparisons of the cytoplasmic genomes. (A) Mitochondrial genome composition of the cybrid (G1+HBP). G1
derived sequence (blue), parental-shared sequence (green). (B) Comparisons of the mitochondrial genome sequences of the leaf parent (HBP mtDNA) and the cybrid (G1+HBP mtDNA). The 472.3 kb (87.7%) of the cybrid sequence is highly similar to the HBP sequence. However, the syntenic order and direction are largely rearranged. The 16.8 kb fragment is homologous to 12.2 kb from the mitochondrial genome of HBP (two fragments circled with red in B, 9.7 kb and 2.5 kb). (C) Alignment of the mitochondrial genome sequences of the cybrid (G1+HBP mtDNA) and the callus parent (G1 mtDNA). Up to 96.9% (521.6 kb) of the

mitochondrial genome sequence from the cybrid is continuously co-linear with the entire mitochondrial genome sequence from G1. The only differences were SNPs and an extra 16.8 kb fragment (circled with red) in the cybrid.

	SNP	InDel		
Mitochondrial genome				
G1+HBP vs G1	5	1		
G1+HBP vs HBP	743	244		
Chloroplast genome				
G1+HBP vs G1	384	120		
G1+HBP vs HBP	11	42		

表 3.2 HB 柚胞质杂种胞质基因组 SNPs 和 InDels 来源分析 Table 3.2 Origin of SNPs and InDels from the cytoplasmic genomes from G1+HBP

#### 3.1.3 HB 柚胞质杂种线粒体基因组 16.8 kb 重复序列验证

根据 HB 柚胞质杂种线粒体基因组 16.8 kb 重复序列与其亲本线粒体基因组 序列的相似性,分为9.7 kb(S1)、4.6 kb(S2)和2.5 kb(S3)三部分。其中, S1 和 S3 在'国庆 1 号'温州蜜柑和 HB 柚中均有同源序列, S2 仅在'国庆 1 号'温州蜜柑中有同源序列(图 3.4A)。依据 HB 柚胞质杂种与'国庆1号'温 州蜜柑共有 16.8 kb、HB 柚胞质杂种特有 16.8 kb 及构成 16.8 kb 三部分(S1、S2 和 S3)的 5'和 3'端侧翼序列设计扩增引物(引物序列见附录 III)。P1 和 P9 是 HB 柚胞质杂种和'国庆1号'温州蜜柑线粒体基因组共有 16.8 kb 5'和 3'端侧翼 序列扩增引物; P5 和 P10 是 HB 柚胞质杂种特有 16.8 kb 5'和 3'端侧翼序列扩增 引物; P6/P2、P7/P3 和 P8/P4 分别是 S1、S2 和 S3 5'和 3'端侧翼序列扩增引物。 P1/P6 和 P4/P9 用于扩增 HB 柚胞质杂种和'国庆1号'温州蜜柑共有 16.8 kb 5' 和 3'端侧翼连接点序列: P4/P10 和 P5/P6 用于扩增 HB 柚胞质杂种特有 16.8 kb 5' 和 3'端侧翼连接点序列: P2/P7 用于扩增 S1 和 S2 连接点序列: P3/P8 用于扩增 S2 和 S3 连接点序列; P1/P9 用于扩增 HB 柚胞质杂种和'国庆1号'温州蜜柑 共有 16.8 kb 全长序列; P5/P10 用于扩增 HB 柚胞质杂种特异 16.8 kb 全长序列。 PCR 扩增结果显示, P1/P6、P2/P7、P3/P8 和 P4/P9 在 HB 柚胞质杂种和'国庆1 号'温州蜜柑有扩增产物,而 HB 柚没有扩增产物,表明 HB 柚胞质杂种和'国 庆1号'温州蜜柑存在 16.8 kb 的共有片段; P4/P10 仅 HB 柚胞质杂种有扩增产 物,表明 HB 柚胞质杂种特异的 16.8 kb 3'端序列是真实存在: P5/P6 在 HB 柚胞 质杂种和 HB 柚有扩增产物,表明 HB 柚胞质杂种特异 16.8 kb 5'端序列是真实 存在, 且与 HB 柚 S1 5'端序列同源(图 3.4B); P1/P9 在 HB 柚胞质杂种和'国 庆1号'温州蜜柑扩增出 16.8 kb 全长序列, 表明 HB 柚胞质杂种和'国庆1号' 温州蜜柑线粒体基因组中存在 16.8 kb 全长序列; 然而 P5/P10 没有在 HB 柚胞质 杂种中扩增出 16.8 kb 特异全长序列, 相反在 HB 柚胞质杂种和'国庆1号'温 州蜜柑中扩增出一个 379 bp 的短片段, 即'国庆1号'温州蜜柑 mtDNA 片段(图 3.4C)。以上实验结果表明 HB 柚胞质杂种线粒体基因组可能存在两种线粒体单 倍型基因组, 即'国庆1号'温州蜜柑线粒体单倍型基因组和含有 16.8 kb 重复 序列的线粒体单倍型基因组。



**图 3.4 HB 柚胞质杂种线粒体 DNA 特异 16.8 kb 鉴定。**(A) HB 柚胞质杂种和'国庆1号' 温州蜜柑共有 16.8 kb 片段的位置和结构,将其分为 9.7 kb(S1,蓝色)、4.6 kb(S2,绿色)

# 和 2.5 kb (S3, 紫色) 三个部分。P1-P10 表示根据 16.8 kb 不同位置序列设计的引物; (B-C) 表示不同引物组合 PCR 凝胶电泳图。

Figure 3.4 Characterization of the 16.8kb recombinant mtDNA fragment in the cybrid. (A) The location and structure of the original 16.8 kb fragment and its duplication in the cybrid, The 16.8 kb was divided into three segments, i.e. the 9.7 kb segment (S1, blue), the 4.6 kb segment (S2, green) and the 2.5 kb (S3, purple). Ten primers (P1 to P10) were designed for PCR. (B-C) Electrophoresis of PCR products. The templates are total DNA. The mitochondrial gene *atp6* was used as the control. M, 1 kb DNA ladder.

#### 3.1.4 HB 柚胞质杂种与其融合双亲叶绿体基因组及核基因组比较

以甜橙叶绿体基因组为参考基因组,对 HB 柚胞质杂种叶绿体基因组进行拼接组装,最终组装成四段式结构的叶绿体基因组(图 3.5)。HB 柚胞质杂种叶绿体基因组大小 160,423 bp,GC 含量为 38.47%。四段式结构 LSC 区、SSC 区和两个 IR 区序列长度依次为 87,917 bp、18,498 bp、27,004 bp。HB 柚胞质杂种叶绿体基因组编码基因的种类、分布区域与 HB 柚相近。从 HB 柚胞质杂种叶绿体基因组与'国庆 1 号'温州蜜柑和 HB 柚叶绿体共线性点阵图可以看出,HB 柚胞质杂种叶绿体基因组与'国庆 1 号'温州蜜柑叶绿体基因组共线性较好,但存在至少 10 个序列差异区域;而 HB 柚胞质杂种叶绿体基因组与 HB 柚叶绿体基因组与 HB 柚叶绿体基因组与 11 个 SNPs 和 42 个 InDels 变异位点(表 3.2)。表明 HB 柚胞质杂种叶绿体基因组来自 HB 柚。

65



图 3.5 HB 抽胞质杂种叶绿体基因组结构图。圈两侧标有线粒体保守基因的位置和方向,相同颜色方框表示同一个复合体基因,方框大小代表基因大小,圈内侧基因表示反向转录,圈 外侧基因表示正向转录。

Figure 3.5 Circular map of the chloroplast genome of G1+HBP. Genes homologous to known protein-encoding genes are shown on the circle. Different classes of conserved protein-encoding

genes are assigned with distinct colors. The genes shown inside the circle are transcribed in a counter-clockwise direction. The genes on the outside of the circle are transcribed in a clockwise direction.



图 3.6 HB 柚胞质杂种与'国庆1号'温州蜜柑(A)及 HB 柚胞质杂种与 HB 柚(B) 叶绿体基因组共线性分析。

# **Figure 3.6 Dot-plot comparisons of the chloroplast genome.** Comparisons of the chloroplast genomes from G1+HBP and G1 (A), and the G1+HBP and HBP (B).

HB 柚胞质杂种、'国庆1号' 温州蜜柑和 HB 柚与甜橙基因组 SNP 和 InDel 统计结果显示,HB 柚胞质杂种核基因组与 HB 柚相同且与'国庆1号'温州蜜 柑不同的 SNPs 有 946,238 个,仅 145 个核基因组 SNPs 与'国庆1号'温州蜜 柑相同且与 HB 柚不同;与之相类似,HB 柚胞质杂种核基因组与 HB 柚相同且 与'国庆1号'温州蜜柑不同的 InDels 有 195,099 个,仅 5,562 个核基因组 InDels 与'国庆1号'温州蜜柑相同且与 HB 柚不同。此外,HB 柚胞质杂种与双亲共同的 SNPs 有 29197 个,InDels 有 26421 个;与双亲均不同的 SNPs 有 1210 个,InDels 有 401 个;与双亲各有一个位点相同的 SNPs 有 201 个,InDels 有 24 个(表 3.3)。综上所述,HB 柚胞质杂种细胞核基因应来自 HB 柚,与早期分子标记鉴定结果一致。

	SNP	InDel
Same as G1	145 (0.01%)	5,562 (2.44%)
Same as HBP	946,238 (96.85%)	195,099 (85.76%)
Shared with parents	29197 (3.00%)	26421 (11.61%)
Differ for parents	1210 (0.12%)	401 (0.18%)
Half of each G1 and HBP	201 (0.02%)	24 (0.01%)

表 3.3 HB 柚胞质杂种细胞核 SNPs 和 InDels 来源分析 Table 3.3 Origin of SNPs and InDels from the nuclear from G1+HBP

#### 3.1.5 冰糖橙胞质杂种线粒体基因组

参照 HB 柚线粒体 DNA 提取方法,提取冰糖橙胞质杂种线粒体 DNA。利用 SPAdes 和 Velvet 软件对冰糖橙胞质杂种线粒体二代测序数据拼接组装,分别获 得 15683 个和 482 个 scaffolds;以'国庆 1 号'温州蜜柑线粒体基因组为参考,确定 13 个 scaffolds 为候选;再以甜橙叶绿体基因组为参考,剔除叶绿体 scaffolds 序列 5 个,剩余 scaffolds 设计引物(引物序列见附录 III),按照 2 (1) -12-3-14-2 (2) -8-7-5-1-2 (1)的顺序进行 Sanger 测序拼接。其中 NODE\_2 拆分为两部分,标为 2 (1) 和 2 (2)。最终冰糖橙胞质杂种线粒体拼接成大小为 523,775 bp,GC 含量为 45.06%的环状基因组,编码基因种类及数量与'国庆 1 号'温州蜜柑相

似(表 3.4)。冰糖橙胞质杂种线粒体基因组与'国庆1号'温州蜜柑线粒体基因 组同源性及共线性分析显示,冰糖橙胞质杂种线粒体基因组与'国庆1号'温州 蜜柑几乎一致, 仅有 2,662 bp 片段是'国庆 1 号'温州蜜柑线粒体基因组序列同 源重组的结果(图 3.7)。

表 3.4	'国庆1号'	温州蜜柑、	冰糖橙胞质杂种及 HB 柚胞质杂种线粒体基因组比较
Table 3.4	Comparison	ns of the mit	ochondrial genomes between G1, G1+BTC and G1+HBP

Feature	G1	G1+BTC	G1+HBP
Mitochondrial genome			
Genome size (bp)	521,559	523,775	538,434
GC content (%)	45.06	45.06	45.02
Number of rRNA genes	3	3	3
Number of tRNA genes	21	21	21
Number of protein genes	36	36	36
Total genes	60	60	60
	В		
2.7 kb Q	G1+BTC mtDNA	0 200 300 400 500	2.7 kb () 16.2



100

200

G1+HBP mtDNA

300

400

500

500

# 3.2 HB 柚胞质杂种及'国庆1号'温州蜜柑不育线粒体基因筛选

3.2.1 HB 柚胞质杂种及其融合双亲线粒体转录本拼接及功能注释

100

200

300

G1 mtDNA

400

选取 HB 柚胞质杂种、'国庆 1 号'温州蜜柑和 HB 柚三个材料的叶、花、 果皮和汁胞 4 个组织提取总 RNA,构建 4 个组织 RNA 样品混合池。根据前人报 道,线粒体转录组仅有 3 种 rRNA,因此选用 lncRNA 进行 HB 柚胞质杂种、'国 庆 1 号'温州蜜柑和 HB 柚的 RNA 混合池测序,采用 Ribo-Zero™ kit 从总 RNA 样品去除 rRNA,构建 lncRNA 文库,进行 lncRNA 高通量测序。每个样品得到 至少 36,139,798 条 reads,其中 98%以上为有效数据。将这些高质量 clean reads 采用 Trinity 无参拼接、HISAT2 + StringTie 有参拼接及 PASA 整合无参拼接和有

**Figure 3.7 Dot-plot comparisons of the chloroplast genome.** Comparisons of the mitochondrial genomes from G1+BTC and G1 (A), and G1+BTC and G1+HBP (B).

参拼接结果的方法,最终获得 HB 柚胞质杂种及其双亲线粒体综合转录本数据库 (图 3.8)。其中,HB 柚胞质杂种综合转录本有 185707 条,'国庆 1 号'温州蜜 柑综合转录本有 205515 条,HB 柚综合转录本有 189311 条。



图 3.8 线粒体转录本组装和综合转录本数据库注释。

Figure 3.8 Flow for assembly and annotation of integrative set of mitochondrial transcripts. 转录本组装完成后,为了解柑橘线粒体转录本信息,也为线粒体差异转录本 研究提供依据,对 HB 柚胞质杂种、'国庆 1 号'温州蜜柑和 HB 柚转录本进行 基因功能注释。选用 Mitofy 数据库提供的已发表植物线粒体蛋白序列为参考, 利用 tblastn 对转录本进行基因预测。结果表明,HB 柚胞质杂种、'国庆 1 号' 温州蜜柑和 HB 柚包含至少 1 个线粒体保守蛋白的转录本分别为 455、605 和 433 条,包含 2 个以上的线粒体保守蛋白完整或部分序列转录本分别有 20、24 和 20 条。剔除序列相似度>90%的重复转录本,保留最长转录本序列,HB 柚胞质杂种、 '国庆 1 号'温州蜜柑和 HB 柚分别剩余 14、14 和 11 条。

#### 3.2.2 HB 柚胞质杂种、'国庆1号'温州蜜柑与 HB 柚 MSS 区鉴定

目前,已报道的一些与细胞质雄性不育相关的嵌合基因序列全部或者部分属 于不育相比可育线粒体的 MSS,如玉米 urf13(Wise et al 1987)、水稻 orf182(Xie et al 2018)、油菜 orf288(Heng et al 2018)。为了能更有效挖掘与 HB 柚胞质杂 种有关的 CMS 基因,我们将 HB 柚胞质杂种、'国庆 1 号'温州蜜柑与 HB 柚线 粒体基因组进行序列对比,得到 HB 柚胞质杂种与 HB 柚线粒体特异的 MSS 区 M1-M21,长度范围 117-8,914 bp; '国庆 1 号'温州蜜柑与 HB 柚线粒体特异的 MSS 区 M1-M20,长度范围 117-8,914 bp (表 3.5)。HB 柚胞质杂种 MSS 序列 M21 是 M10 的一个拷贝,是 16.8 kb 重复片段的一段序列。将 M1-M20 序列与 甜橙叶绿体基因组序列进行比较,发现 M8、M11、M12 和 M13 在甜橙叶绿体基因组有同源序列。我们根据 M1-M20 5'和 3'序列设计引物(附录 IV),并在 HB 柏胞质杂种及其亲本中进行扩增,结果表明 M8、M11、M12 和 M13 引物在 HB 柏胞质杂种及其亲本中均有扩增产物,其余 16 对 MSS 引物仅在 HB 柏胞质杂种 1 (国庆 1 号) 温州蜜柑中有扩增产物(图 3.9)。

MSS	The location	The location	size(bp)	The location
NO.	in G1+HBP	in G1		in cpDNA
<b>M1</b>	1-288	1-288	288	
M2	50096-57667	50096-57667	7572	
M3	62058-63120	62058-63120	1063	
M4	103446-103706	103446-103706	261	
M5	114039-114271	114039-114271	233	
M6	124343-124595	124343-124595	253	
M7	139075-139334	139075-139334	260	
<b>M8</b>	146142-155055	146142-155055	8914	17879-26817
M9	175441-176400	175441-176400	960	
M10	228750-232999	228750-232999	4250	
M11	236075-236739	236066-236730	665	103413-104076
M12	237530-237953	237521-237944	424	104867-105290
M13	238808-240003	238799-239994	1196	106145-107340
M14	296660-298081	296651-298072	1422	
M15	316430-321780	316421-321771	5351	
M16	329039-330410	329030-330401	1372	
M17	345553-345669	345544-345660	117	
M18	395278-398234	395269-398225	2957	
M19	424079-428086	424070-428077	4008	
M20	471806-473100	471797-473091	1295	
M21	531280-535529		4250	

表 3.5 HB 柚胞质杂种和'国庆1号'温州蜜柑特异 MSS 信息 Table 3.5 Location of MSSs in the G1+HBP and G1 mitochondrial genome



图 3.9 琼脂糖凝胶电泳验证不育特异 MSS 区。



#### 3.2.3 HB 柚胞质杂种和'国庆1号'温州蜜柑线粒体特异 ORF 预测

HB 柚胞质杂种、'国庆1号'温州蜜柑和HB 柚分别预测到未知功能线粒体 ORFs 1111 个、1076 个和1083 个。HB 柚胞质杂种和'国庆1号'温州蜜柑分 别有730 个和699 个 ORFs 与 HB 柚完全一致;不完全一致的 ORFs HB 柚胞质 杂种有381 个,'国庆1号'温州蜜柑有377 个。这些不完全一致的 ORFs 中, HB 柚胞质杂种有251 个 ORFs,'国庆1号'温州蜜柑有280 个 ORFs 所在转录 本为嵌合转录本。其中,HB 柚胞质杂种相比 HB 柚特异 ORFs 共89 个,包括 66 个特异 ORFs 完全在 MSS 区,11 个特异 ORFs 有7-887 bp 在 MSS 区,剩余 12 个特异 ORFs 在 HB 柚中缺失;'国庆1号'温州蜜柑相比 HB 柚特异 ORFs 共70 个,44 个特异 ORFs 完全在 MSS 区,11 个特异 ORFs 有52-887 bp 在 MSS 区,15 个特异 ORFs 在 HB 柚中缺失。HB 柚胞质杂种与'国庆1号'温州蜜柑 共有的不育系特异 ORFs 63 个,其中 HB 柚胞质杂种和'国庆1号'温州蜜柑共 有的57 个 ORFs 对应转录本注释的线粒体基因相同(图 3.10)。因此,将这些 ORFs 作为 HB 柚胞质杂种和'国庆1号'温州蜜柑的候选 CMS ORFs(附录 V)。



Figure 3.10 Flow for identification of CMS-associated ORFs.

#### 3.2.4 CMS 候选基因

先前研究表明,HB 柚胞质杂种、'国庆1号'温州蜜柑与HB 柚线粒体基因 组编码线粒体基因序列差异主要是HB 柚缺少编码 rps1 的序列。而不育系HB 柚 胞质杂种和'国庆1号'温州蜜柑相比可育系HB 柚特异的 orf205 是 rps1 编码 序列。以ORFs 所在的转录本是否含有 rps1 基因为标准,在不育材料特异 ORFs 中发现 2 个反向 ORFs(编码 59 个氨基酸的 orf59 和编码 88 个氨基酸的 orf88) 和 1 个正向 ORF(编码 55 个氨基酸的 orf55)对应的转录本含有线粒体 rps1 3' 端 138 bp 反向互补序列(图 3.11A)。PCR 扩增和单克隆测序结果表明,这 3 个 ORFs 和 rps1 互补序列串联转录的转录本序列在HB 柚胞质杂种和'国庆1号' 温州蜜柑中一致,长度均为 1,705 bp,将这条转录本命名为 DN16978。此外, HB 柚胞质杂种和'国庆1号'温州蜜柑有长度 882 bp、能编码 rps1 完整信息的 转录本。该转录本编码 rps1 序列与 DN16978 有 138 bp 重叠区域,位于 DN16978 的 3'端(图 3.11B)。

HB 柚线粒体转录本存在与 DN16978 同源转录本, 起始位点位于 HB 柚线粒体序列 372,841 bp,转录本全长 1,100 bp。该转录本与 orf88 上游 741 bp 的碱基序列一致,与 orf88 有 86.7%序列相似性,与 orf88 下游 100 bp 碱基有 91.2%相

同。将 HB 柚线粒体转录本与 orf88 同源的序列转换为氨基酸,在 HB 柚 orf88 同源序列碱基 52-54 的位置出现终止密码子,而这个终止密码子在 orf88 是编码 脯氨酸 (P),结果如图 3.11C。此外,在碱基 1-51 的序列中,HB 柚与 orf88 序 列存在 2 个 SNPs,1 个 14 bp InDel 差别,PCR 和 Sanger 测序结果证实 SNPs 和 InDel 的差异是真实存在。这些差别可能是造成 HB 柚 orf88 同源序列提前终止的 主要原因。此外, orf55、orf59 和 ψrps1 位于 MSSs 区域,在 HB 柚中没有同源 序列。

A





图 3.11 CMS 特异转录本 DN16978 结构、序列及氨基酸信息。(A) HB 柚胞质杂种及其融合 双亲 DN16978 同源序列比对。DN16978 位于 HB 柚胞质杂种和'国庆1号'温州蜜柑 mtDNA 的反义链。颜色相同的片段表示 HB 柚胞质杂种/'国庆1号'温州蜜柑与 HB 柚的同源序 列;(B) 蓝色、绿色和紫色序列表示'国庆1号'温州蜜柑和 HB 柚胞质杂种与 HB 柚相似 序列。绿色序列表示 orf88 碱基序列及氨基酸序列;深蓝色序列表示 orf59 碱基序列及氨基 酸序列;茶色序列表示 orf55 碱基序列及氨基酸序列。灰色序列表示 ψrps1;(C) HB 柚胞 质杂种及其融合双亲 DN16978 同源转录本序列对比。

Figure 3.11 The structure, nucleotide sequence and deduced amono acid sequence of the CMS-specific transcript DN16978. (A) The structure of DN16978 in G1+HBP/G1 and the structure of the homologous sequence in HBP. The numbers indicate the postions on the mitochondiral genomes. orf55 (red and dark blue blocks) is on the forward strand, whereas orf59 (red and tawny blocks) is on the reverse strand. Orf55 and orf59 are overlaped in a 62-bp segment

(red block). *Orf55* and *orf59* are fully overlaped with MSS 20. The 1100-bp fragement that comprises *orf88* (green block) and the flanking regions (purple and blue blocks) are highly similar to a fragement of mtDNA from HBP. Nonetheless, in HBP, *orf88* was not predicted by ORFinder because of a pre-mature stop codon (TGA, underlined with red in C). In G1 and G1+HBP, the 882-nt *rps1* gene is located on the forward strand of mtDNA, which is upstream and partially overllaps (138 bp) *DN16978*. (B) Sequence of *DN16978*. The blue, green and purple sequences represent the three continuous segements in G1 and G1+HBP that are highly similar to HBP. The deduced amino acid residues of orf88 (green), orf59(dark blue) and orf55 (tawny) are indicated under the respective nucleotide sequences. The red 62-nt refer to the overlapping segement between *orf59* and *orf55*. The grey sequence refers to  $\psi rps1$ . (C) Multiple alignment of the transcript from *DN16978* from G1 and G1+HBP with the homologous transcript from HBP. *Orf88* (green box) is present in G1 and G1+HBP but not in HBP.

#### 3.2.5 CMS 候选基因在 HB 柚胞质杂种及其亲本中的表达

利用 RT-PCR 技术检测候选转录本 DN16978 在 HB 柚胞质杂种及其亲本中的表达。RT-PCR 结果表明转录本 DN16978 只在不育材料 HB 柚胞质杂种和'国 庆1号'温州蜜柑中有表达,在可育材料 HB 柚中不表达(图 3.12A)。在 HB 柚 胞质杂种及其亲本花发育的三个阶段(雄蕊原基时期、四分体时期和小孢子发育时期)进行 qPCR 检测 CMS 候选 ORFs (*orf55, orf59*和 *orf88*)的表达。这三个 ORFs 在可育材料 HB 柚中几乎不表达。在 HB 柚胞质杂种中,这三个 ORFs 在 雄蕊原基时期表达量最高,并随着花的发育逐渐降低;而在'国庆1号'温州蜜 柑中,这三个 ORFs 表达量随着花的发育略微增加,并且在小孢子发育时期表达量显著高于 HB 柚胞质杂种(图 3.12B-D)。



图 3.12 转录本 DN16978 在 HB 柚胞质杂种及其亲本花中的表达(A)和 orf55、orf59、 orf88 在 HB 柚胞质杂种及其亲本花三个发育时期的表达水平分析(B-D)。时期 1,雄 蕊原基时期;时期 2,四分体时期;时期 3,小孢子发育时期。

Figure 3.12 Expression of *DN16978* (1705 bp) in the cybrid ans its fusion parents (A), and relative expression level of *orf55*, *orf59* and *orf88* at three flower development stages (B-D). Stage 1, sepal primordial initiating. Stage 2, tetrad stage. Stage 3, microspore stage.

#### 3.3 CMS 不育候选转录本功能验证

#### 3.3.1 线粒体不育候选转录本异源表达拟南芥

为验证线粒体不育候选转录本是否引起雄性不育,本研究将线粒体不育候选转录本转化拟南芥,并进行与花粉育性相关表型观察。在 DN16978 转基因 T1 代植株中发现3个系表现出雄性不育表型,包括花丝变短、角果变短、花粉活力降低。每个转基因系选取3 株生长良好、长势相近的植株,每株分枝数超过3 个,每个分枝已结荚数为15,统计正常角果数。同时,选取同一状态的野生型拟南芥为对照。转基因植株每个分枝正常角果数约为7,野生型植株正常角果数为12,表明转基因系植株正常角果数降低可能是花粉育性降低的结果(图 3.13A-E)。

亚历山大染色液是一种能将植物有活力花粉染成紫红色的染料。取转基因系和野生型拟南芥植株刚开裂的花朵,用解剖针将花药从雄蕊剥离,滴加少量染液,常温避光放置 6-8 h 后置于体式显微镜下观察。结果发现野生型花药药室花粉被染成红色,转基因系花药药室花粉大多数被染成绿色且花粉粒形状不规则,但也有部分花粉被染成红色。分别统计了转基因系和野生型拟南芥 57 个花药被染成红色的花粉数,统计结果显示转基因系每个花药有 82 个花粉被染成红色,野生型每个花药有 251 个花粉被染成红色。以上结果表明,异源表达柑橘转录本*DN16978* 能导致拟南芥花粉育性降低(图 3.13F-G)。



**图 3.13** *DN16978* 的转基因拟南芥表型以及在转基因植株中的表达。(A) 拟南芥转基因系 T1 代阳性鉴定;(B) 野生型和转基因系植株;(C) 野生型和转基因系整花,标尺 5 μm;(D) 野生型和转基因系角果,标尺 5 mm;(E) 野生型和转基因系正常角果数统计;(F-G) 野生 型和转基因系成熟花粉亚历山大染色及可育花粉粒统计,标尺 5μm,共统计 57 个花药。 Figure 3.13 Male sterility phenotype of the transgenic lines overexpressed the chimeric transcript from *DN16978* and the transcript expression in the transgenic plants. A: Positive transgenic T1 lines confirmed by PCR amplification. OE, Overexpression line B: WT and OE lines during the reproductive stage of development. C: Flowers from WT and the OE lines. Bar =5 μm. D: Siliques of self-crossed WT (upper) and OE (lower) lines. Bar =5 mm. E: Number of normal siliques in the WT and OE lines. F, G: Visualization and quantification of viable pollen grains in WT and the OE lines. Viable pollen was identified by staining with Alexander's stain. Bar =5 μm, n =57 anthers per line.

#### 3.3.2 转录本 DN16978 核互作蛋白筛选

将 pGBKT7-16978 重组质粒转化的 AH109 酵母菌液与柑橘花药酵母文库菌 进行 mating。Mating 后的融合菌液涂布在四缺培养基(SD/-Trp/-Leu/-His/-Ade), 30 °C 培养 3-5 d 后挑单克隆并摇菌,菌液-80 °C 冰冻过夜,用 pGADT7 通用引 物进行菌液扩增,并送公司进行单克隆测序,测序结果在柑橘基因组数据库 (http://citrus.hzau.edu.cn/orange/)进行分析,获得 5 个与花发育有关的潜在互作 蛋白,分别是: *Cs5g24860* (NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12); *orange1.1t00826* (Pentatricopeptide repeat-containing protein); *Cs8g02500* (Transcription factor ABORTED MICROSPORES); *Cs5g03200* (Beta-carotene 3-hydroxylase 1); *Cs1g25350*(S locus F-box-S2C protein)。同时, 用 Mitoprot (<u>http://ihg.gsf.de/ihg/mitoprot.html</u>) 和 PredSL (Petsalaki et al 2006) 对这些可能互作蛋白进行线粒体靶向定位预测及亚细胞定位预测。这些蛋白的功 能、序列号、线粒体靶向性及亚细胞定位预测结果,见表 3.6。

Table 3.0 Candidate D1(10)70 binding proteins from a yeast two hybrid screen					
Gene ID Predicted enzyme or function		Mitochondrial targeting	Predicted subcellular		
		prediction score	localization		
Cs5g24860	NADH dehydrogenase [ubiquinone]	0.7595	mitochondrion		
	1 alpha subcomplex subunit 12				
orange1.1t00826	Pentatricopeptide repeat-containing	0.9748	Chloroplast or Nucleus		
	protein				
Cs8g02500	Transcription factor ABORTED	0.0139	Nucleus		
	MICROSPORES				
Cs5g03200	Beta-carotene 3-hydroxylase 1	0.9945	Cell membrane		
Cs1g25350	S locus F-box-S2C protein	0.0497	Nucleus		

表 3.6 DN16978 互作候选蛋白信息

	DNI CORO LI II			1 • 1
Table 3.6 Candidate	DN16978-binding	proteins from a	yeast-two-hy	brid screen

#### 3.3.3 CMS 候选 ORFs 亚细胞定位及跨膜结构域预测

通过网站 TMpred 对 orf55, orf59 和 orf88 编码蛋白的氨基酸进行跨膜结构 域预测发现, orf59 在 3-19 氨基酸之间有跨膜结构域; orf88 在 2-22 的氨基酸之 间有跨膜结构域; orf55 没有跨膜结构域(图 3.14)。利用软件 PredSL 对这三个 ORF 进行亚细胞定位预测,预测结果显示只有 orf88 的 mTP 得分为 0.83916,且 在第 39 个氨基酸有剪切位点,表明 orf88 定位在线粒体, orf55 和 orf59 的 mTP 得分分别为 0.008905 和 0,表明这两个 ORFs 编码的蛋白没有定位在线粒体(表 3.7)。综上所述,本研究将 orf88 作为 CMS 候选 ORF 进行后续研究。

表 3.7 orf55、orf59 和 orf88 亚细胞定位预测

sequence	cTP score	mTP score	SP score	prediction	cleavage site
orf55	0	0.008905	0.060917	other	-
orf59	0.000107	0	0.99921	secreted	26
orf88	0.087601	0.839159	0.000404	mitochondrion	39

cTP: chloroplast. mTP: mitochondrial. SP: secreted proteins



图 3.14 orf55、orf59 和 orf88 跨膜结构域预测及 orf88 毒性检测。得分>500 认为是疏水性区域;得分<500 认为是亲水性区域。i,代表线粒体内;o,代表线粒体外;CK,对照。 Figure 3.14 Transmembrane region of orf55, orf59 and orf88 predicted by TMpred and toxic effect of orf88 in E.coli. Scores >500 are considered hydrophobic. Scores <-500 are considered hydrophobic.

#### 3.3.4 orf88 毒性检验

为确定 orf88 编码蛋白是否具有抑制大肠杆菌生长的毒害作用,把 PCR 扩增的 orf88 编码序列连接到原核表达载体 PET28a,转化至大肠杆菌表达菌株 BL21 (DE3),挑单克隆于 500 μL 加有卡那霉素的 LB 培养。将等量的菌液加到两个 装有 20 mL 液体 LB (加卡那霉素)的三角瓶,37 °C 培养约 3 h,OD 值至 0.6 左右,向其中一个装有菌液的三角瓶中加入 0.5 mM 的 IPTG 诱导大肠杆菌表达 蛋白,每隔半小时测定一次 OD,共测 5 次。结果显示,加有 IPTG 和未加 IPTG 的大肠杆菌 OD 值变化趋势一致 (图 3.14D),表明 orf88 编码的蛋白没有毒性区 域且对大肠杆菌生长没有毒害作用。

#### 3.3.5 点对点验证 orf88 与 CsNADH 蛋白互作

先前的蛋白功能、线粒体靶向定位及亚细胞定位预测结果表明,CsNADH 可能是与 orf88 互作导致柑橘细胞质雄性不育的重要核蛋白。因此,进行酵母点 对点实验验证 orf88 与 CsNADH 的互作关系。

为验证 mating 文库筛选到的 CsNADH 与 orf88 互作关系,扩增 orf88 及 CsNADH 的编码区分别连接到 pGBKT7 和 pGADT7 载体,并共转入至酵母菌株 AH109,涂布在二缺培养基和四缺培养基,30 ℃ 培养 3-5 d 后挑取四缺培养基生 长的单克隆,PCR 鉴定单克隆阳性后,吸取 5 μL (分光光度计检测 OD 为 0.2) 菌液滴加于二缺培养基、普通四缺培养基,30 ℃ 培养 3-5 d 观察,结果如图 3.15 所示。共转 pGBKT7-orf88 与 pGADT7-CsNADH 和对照组合在二缺培养基长势 良好,仅共转 pGBKT7-orf88 和 pGADT7-CsNADH 在四缺培养基有明显斑点, 而对照组合在四缺培养基不能正常生长,表明 CsNADH 与 orf88 存在互作关系。



图 3.15 orf88 与 CsNADH 互作关系酵母点对点验证。 Figure 3.15 Interaction between orf88 and CsNADH in a yeast two-hybrid assay.

#### 3.3.6 LUC 荧光验证 orf88 与 CsNADH 蛋白互作

为验证酵母双杂交鉴定的核基因编码蛋白 CsNADH 与 orf88 互作是真实存 在,本研究采用农杆菌介导瞬时转化烟草表达系统进行 LUC 荧光验证。先前研 究已预测 orf88 的 N 端存在跨膜结构域,因此,构建 pCAMBIA-LUC 的 C 端与 orf88 融合载体 (pCAMBIA-orf88-cLUC), pCAMBIA-LUC 的 N 端与 CsNADH 融合载体 (pCAMBIA-NADH-nLUC)。把同一叶片划分四个区域,然后将构建好 的融合载体共同注射入其中的一个区域,剩余三个区域分别共同注射空载体 pCAMBIA-nLUC 和 pCAMBIA-cLUC, pCAMBIA-orf88-cLUC 和 空 载 体 pCAMBIA-nLUC, 以及 pCAMBIA-NADH-nLUC 和空载体 pCAMBIA-cLUC 作 为对照,根据瞬时转化 LUC 荧光强度判断互作关系。结果表明,同一烟草叶片 共同注射 orf88 和 CsNADH 区域有较高的 LUC 荧光强度,其余对照区域观察不 到 LUC 荧光强度(图 3.16)。以上结果表明,线粒体基因 orf88 与核编码基因 CsNADH 有相互作用。



图 3.16 LUC 荧光实验验证 orf88 与 CsNADH 互作关系。nLUC、cLUC 分别为 LUC 的 N 端和 C 端,共转化 nLUC 与 orf88-cLUC、CsNADH-nLUC 与 cLUC、nLUC 与 cLUC 作为阴性 对照。

**Figure 3.16 Interaction between orf88 and CsNADH in the luciferase complementation imaging assay.** Tobacco leaves were divided into four parts and infiltrated with *Agrobacterium* strains harboring orf88-cLUC and CsNADH-nLUC. The following three pairs of constructs were used as negative controls: nLUC+ orf88-cLUC, CsNADH-nLUC + cLUC and nLUC + cLUC.

#### 3.3.7 Orf88 与 CsNADH 亚细胞定位分析

为了验证 orf88 和 CsNADH 定位在线粒体,利用 Gateway 同源重组法构建 了 pCAMBIA-orf88-YFP 和 pCAMBIA-CsNADH-YFP 融合表达载体,以带有 mCherry 线粒体定位蛋白 mt-rk 作为阳性对照,通过农杆菌介导瞬时转化烟草表 达系统,观察 YFP 和 mCherry 在烟草细胞内发光信号的位置来确定 orf88 与 CsNADH 定位情况。分别将 pCAMBIA-orf88-YFP 和 pCAMBIA-CsNADH-YFP 与阳性对照同时转入烟草叶片,激光共聚焦显微镜观察发现,烟草叶片细胞线粒 体检测到 pCAMBIA-orf88-YFP 和 pCAMBIA-CsNADH-YFP 发出的荧光,与阳 性对照发出的荧光基本重合,表明 orf88 与 CsNADH 和阳性对照定位相同,均 定位在线粒体(图 3.17)。



**图 3.17 orf88 和 CsNADH 亚细胞定位。**分别将 orf88 和 CsNADH 与线粒体定位阳性对照 mt-rk 共注射烟草,激光共聚焦显微镜下观察 YFP 和 mCherry 发光。标尺 10 μm Figure 3.17 Subcellular localization of orf88 and CsNADH in tobacco. orf88-YFP and CsNADH-YFP were co-transformed with mt-rk, a mitochondrial marker with the red mCherry. Bar =10 μm.

#### 3.3.8 HB 柚胞质杂种及其融合双亲花药 NAD<sup>+</sup>、NADH 含量测定

为了确定 orf88 与 CsNADH 相互作用是否影响 HB 柚胞质杂种和'国庆 1 号'温州蜜柑小孢子内氧化还原特征物质的平衡,本研究将 HB 柚胞质杂种、'国 庆 1 号'温州蜜柑及 HB 柚小孢子发育早期、中期和晚期三个时期花药进行混合,测定混合样品 NAD<sup>+</sup>和 NADH 含量并计算 NAD<sup>+</sup>/NADH 比率。HB 柚胞质杂种和'国庆 1 号'温州蜜柑花药 NAD<sup>+</sup>含量分别为 2.69 mmol/g 和 2.05 mmol/g,明显 高于 HB 柚 (1.71 mmol/g);花药 NADH 含量,HB 柚 (0.61 mmol/g) 与 HB 柚 胞质杂种 (0.53 mmol/g) 接近,明显高于'国庆 1 号'温州蜜柑 (0.38 mmol/g) (图 3.18)。此外,HB 柚胞质杂种和'国庆 1 号'温州蜜柑花药线粒体 NAD<sup>+</sup>/NADH 比率分别高于 HB 柚的 1.8 和 1.9 倍。综合以上结果,小孢子发育早期、中期、晚期三个时期,HB 柚胞质杂种和'国庆 1 号'温州蜜柑花药 NAD<sup>+</sup>含量上升是 线粒体电子传递链异常的主要原因。



图 3.18 HB 柚胞质杂种及其融合双亲 NAD<sup>+</sup>和 NADH 含量的测定及花不同发育期 CsNADH 表达量分析。

Figure 3.18 (A-C) NAD<sup>+</sup> and NADH content and NAD<sup>+</sup>/NADH ratio in the cybrid and its fusion parents. (D) Relative expression level of *CsNADH* at three flower development stages. Stage 1, sepal primordial initiating. Stage 2, tetrad stage. Stage 3, microspore stage.

# 3.3.9 Real-time PCR 检测 HB 柚胞质杂种及其融合双亲小孢子发育过程 CsNADH表达

为进一步探究 CsNADH 是否在 HB 柚胞质杂种及其融合双亲花发育时期存 在差异表达,选取雄蕊原基、四分体时期及小孢子发育时期进行 Real-time PCR 检测。CsNADH 表达量分析结果显示,雄蕊原基和四分体时期 CsNADH 在 HB 柚胞质杂种及其融合双亲的表达没有明显差异,而在小孢子发育时期,CsNADH 的表达量在 HB 柚中显著高于 HB 柚胞质杂种及'国庆1号'温州蜜柑,由此表 明 HB 柚胞质杂种和'国庆1号'温州蜜柑小孢子发育异常可能与 CsNADH 表 达有关(图 3.18D)。 4 讨论

#### 4.1 柑橘胞质杂种线粒体基因组发生融合

HB 柚胞质杂种线粒体基因组与亲本线粒体基因组共线性分析表明,其线粒体基因组主要遗传自'国庆1号'温州蜜柑,且没有发生明显的序列重排。前人研究表明,原生质体培养条件(化学试剂或光照)能降低原生质体内细胞器分裂活性,可以使其中一个亲本线粒体活性丧失,忽略其对融合后代线粒体基因组的影响(肖诗鑫 2014)。本实验室前期检测柑橘愈伤组织和叶肉原生质体活性氧,发现愈伤组织原生质体细胞内没有过量积累活性氧,且氧化还原保持动态平衡,认为细胞内有清除活性氧的生理机制。而叶肉原生质体细胞内缺少清除活性氧的生理机制,导致其细胞内活性氧过量积累,对细胞生理代谢产生影响,阻碍细胞再生,表明叶肉原生质体线粒体受到培养环境胁迫,处于功能异常状态。同时,付婧(2018)利用激光共聚焦显微镜观察愈伤组织和叶肉原生质体线粒体形态,发现愈伤组织原生质体线粒体呈现点状或短棍状,分裂活性强;叶肉原生质体线 粒体表现为网络状,分裂活性较弱。因此,愈伤亲本比叶肉亲本线粒体分裂活性强可能是 HB 柚胞质杂种线粒体基因组来自愈伤组织亲本'国庆1号'温州蜜柑的原因之一。

原生质体融合再生植株的线粒体基因组往往是亲本线粒体基因组重组的结 果(Pelletier 1991; Arimura et al 2004)。烟草原生质体研究表明,新融合的原生 质体线粒体没有立即完成线粒体融合,而是培养一段时间,在细胞处于去分化状 态后不同来源线粒体大量融合;同时,原生质体线粒体 DAPI 染色研究表明,新 融合的原生质体大约四分之一线粒体没有线粒体遗传物质;而随着原生质体培养, 原生质体内的线粒体大量融合,分裂产生的几乎所有线粒体都包含线粒体 DNA, 说明线粒体融合可能通过分子内重组产生亚基因组动态平衡,确保所有线粒体都 能获得遗传物质 (Sheahan et al 2005)。线粒体融合可能促进线粒体基因组重组, 并在融合后代中得以体现。

HB 柚胞质杂种线粒体基因组,除完全遗传愈伤组织亲本线粒体基因组外,还有一个额外的 16.8 kb 重组片段。该重组片段与愈伤组织亲本'国庆 1 号'温州蜜柑线粒体基因组有完全一致的同源片段,而这个片段的 12.2 kb 在叶肉亲本

84

HB 柚线粒体基因组由 9.7 kb 和 2.5 kb 两个同源片段组成。虽然与处于脱分化状态的愈伤组织原生质体相比, 柑橘叶肉原生质体难以脱分化再生, 叶肉亲本线粒体很难与愈伤组织亲本线粒体发生融合, 但一些体细胞杂种/胞质杂种后代检测到叶肉亲本线粒体基因组(Dambier et al 2011; Aleza et al 2016)。HB 柚胞质杂种与'国庆1号'温州蜜柑线粒体基因组共有及 HB 柚胞质杂种线粒体基因组特有的 16.8 kb 序列 5'和 3'端序列扩增结果表明 HB 柚胞质杂种线粒体基因组存在包含和不包含 16.8 kb 重复片段的两种线粒体单倍型基因组序列。由此表明, HB 柚胞质杂种线粒体基因组重组片段可能是'国庆1号'温州蜜柑和 HB 柚线粒体基因组融合的结果。此外, 冰糖橙胞质杂种线粒体基因组相比愈伤亲本'国庆1号'温州蜜柑, 也存在一个 2.7 kb 的重组片段。因此, 柑橘原生质体融合产生胞质杂种的过程中, 胞质杂种线粒体基因组重组片段可能是愈伤亲本与愈伤亲本为子间或者愈伤亲本与叶肉亲本分子内线粒体基因组同源重组产生。

#### 4.2 HB 柚胞质杂种胞质雄性不育基因筛选

迄今,已经在13种作物中发现了至少28个CMS基因(Kim and Zhang 2018)。 一些CMS基因序列全部或者部分位于不育系相比保持系线粒体基因组的MSS区 (Hanson and Bentolila, 2004)。随着高通量测序技术的发展,直接比较不育系和 可育系线粒体基因组序列,从不育线粒体MSS区筛选CMS候选ORF已经成为一种 行之有效的策略,如:水稻CMS不育基因orf182、洋葱CMS不育基因orf725(Xie et al 2018; Kim et al 2019)。本研究统计了28个CMS基因碱基长度和编码氨基酸数 量,发现CMS基因碱基长度范围在237-1,566 bp,编码氨基酸数量范围在79-522, 由此表明目前报道CMS基因氨基酸数量最小的是水稻不育基因orf79,仅有79个 氨基酸。Igarashi等(2012)以水稻CMS基因orf79为参照,以编码氨基酸数>70 作为标准,对水稻RT-CMS系线粒体的ORF进行预测,从中鉴定到不育基因orf113。 因此,本研究扩大了HB柚胞质杂种及其亲本线粒体基因组特异ORF预测范围, 选取ORF编码蛋白氨基酸数量>50作为标准,对HB柚胞质杂种及其亲本线粒体基 因组进行ORF预测。Xie(2014)等对水稻24个材料线粒体MSS分析发现,水稻 CMS基因除orf113外,其余不育基因orfH79、orf79、CW-orf307部分或者全部序 列位于MSS区,这表明仅以ORF是否位于MSS区作为参考,可能会遗漏与CMS 有关的ORF。综上所述,本研究在特异ORF筛选过程中,首先保留不育材料中的特异ORF,其次对具有同源序列的ORF与MSS区进行比对,最终确定不育材料线粒体中可能与CMS有关的特异ORF 57个。进一步根据HB柚胞质杂种、'国庆1号' 温州蜜柑与HB柚线粒体基因组编码线粒体基因差异,筛选出转录本DN16978及 其含有的3个ORFs(orf55、orf59、orf88)作为CMS候选基因。

# 4.3 Orf88 可能的调控机理

CMS 基因编码毒素蛋白被认为是 CMS 材料不育的重要原因之一(Chen and Liu 2014)。CMS 基因编码蛋白是否具有毒性,可以通过原核表达观察大肠杆菌 生长状况确定。随着研究深入,发现一些 CMS 基因育性相关区域与毒性区域无 关。水稻不育基因 WA352 不同区域转化实验及原核表达毒性检测结果证实, WA352 毒性区域与 WA352 导致雄性不育的区段无关(Luo et al 2013);同时,油 菜不育基因 orf288 毒性区域与不育区域的相关研究也有类似结论(Heng et al 2018)。本研究通过异源表达拟南芥证实转录本 DN16978 能影响植株花粉育性;跨膜结构和亚细胞定位预测结果显示,转录本 DN16978 含有的三个 CMS 特异 ORFs,仅 orf88 编码蛋白同时满足含有线粒体定位信号肽和跨膜结构域,推测 orf88 可能对 HB 柚胞质杂种和'国庆 1 号'温州蜜柑雄性不育起作用;通过原 核表达实验确定 orf88 编码蛋白不能抑制大肠杆菌生长,表明 orf88 编码蛋白不

雄蕊发育进入小孢子减数分裂后期,绒毡层和小孢子内线粒体数量显著增加, 为小孢子发育提供足够的能量(Wang et al 2013)。这个时期线粒体电子传递链复 合体异常导致线粒体功能障碍,影响小孢子正常发育。研究人员在拟南芥呼吸电 子传递链 complex II 和 complex V 突变体观察到不育表型(Leno et al 2007; Li et al 2010)。由此表明,线粒体电子传递链复合体对配子体发育起着重要作用。相关 研究表明,CMS 基因与线粒体复合物结合能损害线粒体功能,导致线粒体不能 为小孢子发育提供足量的能量,进而不能发育成有功能的花粉。水稻 CMS 基因 orf79 能与 complex III 结合并损害其功能,小孢子内活性氧含量增加,ATP 浓度 降低,最终导致小孢子有丝分裂异常,植株败育(Wang et al 2013)。本研究通过 酵母双杂交、烟草瞬时表达 LUC 荧光强度及亚细胞定位等试验,证实核编码基 因 *CsNADH* 定位在线粒体,且与 orf88 互作。*CsNADH* 是 complex I NADH 脱氢 酶的一个亚基。HB 柚胞质杂种、'国庆 1 号'温州蜜柑和 HB 柚小孢子发育期花 药 NAD<sup>+</sup>和 NADH 测定显示不育材料小孢子发育时期 NAD<sup>+</sup>、NADH 含量及 *CsNADH* 表达水平与可育材料存在较大差异,表明小孢子发育过程中 NADH 脱 氢酶活性受到影响。我们推测 orf88 与 CsNADH 之间的相互作用可能会损害复 合物 I 的功能,扰乱线粒体能量代谢,最终不能形成有功能的花粉。



图 3.21 orf88 调控柑橘 CMS 可能机理。Orf88 与线粒体复合物 I 相互作用可能诱导小孢子母 细胞 PCD,导致雄性不育。

Figure 3.21 A working model of *orf88*-mediated CMS. The mitochondrial encoded *orf88* can interact with a nuclear encoded subunit of complex I (NADH) and that these interactions might induce PCD in microspores, which leads to male sterility.

# 5. 后续研究设想

本研究着力构建可靠的柑橘线粒体基因组并发掘细胞质雄性不育基因。利用 高通量测序组装了 HB 柚胞质杂种及其融合双亲的完整线粒体基因组,并从不育 材料 HB 柚胞质杂种、'国庆1号'温州蜜柑相比可育材料 HB 柚线粒体特异 ORFs 中筛选出1个线粒体基因 orf88,推测 orf88 与其靶基因互作可能与柑橘细胞质雄 性不育有关;遗传转化拟南芥实验表明,过量表达 orf88 所在的转录本 DN16978 能使拟南芥花粉育性降低。但缺少拟南芥表型授粉实验及 Northern blot 分析 orf88 在 HB 柚胞质杂种和'国庆1号'温州蜜柑的表达模式,难以完全证实 orf88 是 调控柑橘花粉育性的关键基因。关于 orf88 参与小孢子发育氧化还原特征物质及 能量水平调控,以及其它可能影响柑橘雄性不育的 orf 挖掘等诸多方面,还有待 深入研究。未来研究可以从以下5个方面入手:

1. 转录本 DN16978 拟南芥授粉实验

从遗传学角度概括,植物雄性不育分为孢子体不育和配子体不育两种类型。 三系作物研究中,用恢复系花粉给孢子体不育系授粉,杂种一代育性正常,自交 获得的杂种二代育性分离;而用恢复系花粉给配子体不育系授粉,杂种一代出现 1:1 育性分离比,杂种二代表现为育性正常(陆作楣等 2014)。因此,授粉实验 是判断植物雄性不育类型的重要手段。目前,受到技术手段限制还无法直接转化 CMS 基因进行转基因互补验证,需要借助线粒体靶向信号肽将不育基因编码的 蛋白在转化植株的线粒体中发挥作用;同时,受限于有些作物童期长、转基因体 系不成熟,只能借助模式植物拟南芥或烟草进行转基因验证(衡双平 2015)。关 于本研究,虽然拟南芥转化实验表明转录本 DN16978 能降低拟南芥花粉育性, 但缺少用野生型拟南芥花粉给 DN16978 转基因植株授粉实验,不能确定 DN16978 转基因植株是孢子体不育还是配子体不育,进而不能充分证实 DN16978 对拟南芥花粉育性产生影响。因此,需要对 DN16978 超表达植株进行授粉实验, 分析杂交后代育性分离规律,进一步验证 DN16978 是否影响 HB 柚胞质杂种育 性。

2. orf88 表达机制探究

已报道的 CMS 基因表达模式目前有三种: 特异型 mRNA 和蛋白质在雄性器 官(花药)表达和积累; 组成型 mRNA 和蛋白质在整个植物组织表达和积累; 组成型 mRNA 和蛋白质特异或优先在雄性器官(花药)表达和积累。多数 CMS 系统的基因表达模式属于第三种类型(Chen and Liu 2014)。对于大多数 CMS 系 统而言, CMS 蛋白时空表达模式决定了不同植物雄性不育的特异性。本研究观 察到 HB 柚胞质杂种和'国庆1号'温州蜜柑小孢子异常 PCD, 推测 orf88 编码 蛋白可能在小孢子细胞特异表达,从而造成 HB 柚胞质杂种和'国庆1号'温州 蜜柑雄性不育。未来研究拟利用 Northern blot 技术检测 orf88 在 HB 柚胞质杂种 及其亲本小孢子不同发育期、叶片及幼果的表达模式,以期通过深入了解 orf88 时空表达模式,为解析 orf88 的不育机制提供依据。

3. 解析柑橘线粒体基因组雄性不育机理

线粒体不育基因编码的蛋白能破坏线粒体结构和功能,影响小孢子发育过程 中氧化还原特征物质及能量水平的正常调节,导致能量供应受阻,PCD 异常, 引起花粉败育(Wang et al 2013)。本研究中,虽然筛选到与电子呼吸链相关基因 *CsNADH*,但小孢子发育过程中线粒体功能异常如何影响电子呼吸链功能还不清

88

楚。需要对 HB 柚胞质杂种及其亲本小孢子发育过程中线粒体氧化还原特征物质 及能量水平进行检测,分析不育材料与可育材料中这些物质变化规律,为进一步 解析 HB 柚胞质杂种与'国庆1号'温州蜜柑线粒体不育机理提供依据。 4. 柑橘线粒体其它不育相关 ORFs 发掘

细胞质雄性不育核恢复基因编码的蛋白定位于线粒体,通过抑制 CMS 基因 编码不育蛋白产生的负面影响,使植物育性恢复(Kim and Zhang 2018)。本研究 中,拼接组装的冰糖橙胞质杂种线粒体基因组与亲本'国庆1号'温州蜜柑线粒 体基因组基本一致,而冰糖橙胞质杂种雄蕊形态正常且花粉可育,我们推测冰糖 橙胞质杂种细胞核含有恢复基因,可以将冰糖橙胞质杂种视为 HB 柚胞质杂种 "恢复系"。未来的研究中,可以根据不育系特异 ORFs 序列设计探针,利用 northern blot 技术,分析特异 ORFs 在 HB 柚胞质杂种、HB 柚和冰糖橙胞质杂种 三个材料花蕾不同发育时期的表达模式,找到不育材料(HB 柚胞质杂种)与恢 复材料(冰糖橙胞质杂种)和可育材料(HB 柚)花蕾不同发育时期表达模式存 在差异的特异 ORFs,进一步对这些特异 ORFs 进行功能验证,确定柑橘 CMS 关键 ORF。

5. 研究线粒体基因与雄蕊原基发育核基因互作关系

本实验室以'国庆1号'温州蜜柑为愈伤组织亲本、沙田柚为叶肉亲本,原 生质体融合创制的沙田柚胞质杂种近年已开花结果,并且表现出雄性不育性状。 与 HB 柚胞质杂种相比,沙田柚胞质杂种雄蕊形态正常且花药与花丝明显分割开, 与'国庆1号'温州蜜柑雄蕊性状表型一致。沙田柚胞质杂种线粒体基因组测序 组装正在进行,且初步分析显示沙田柚胞质杂种线粒体基因组与'国庆1号'温 州蜜柑基本一致。前人研究认为核基因组与线粒体基因功能相互作用使 HB 柚胞 质杂种雄蕊原基发育受到抑制(郑蓓蓓 2015)。本研究中不育候选转录本 *DN16978* 酵母双杂实验没有找到调控雄蕊发育的互作蛋白。由此推测可能存在 其它的线粒体 orf 影响雄蕊发育。目前,利用激光显微切割技术解析 HB 柚胞质 杂种与 HB 柚花器官及雄蕊原基分化过程中关键基因调控网络已取得积极进展。 未来研究中,通过将调控 HB 柚胞质杂种雄蕊原基分化的关键基因与沙田柚胞质 杂种同源基因进行分析,进一步解析线粒体基因与雄蕊原基发育核基因互作调控 途径。

89

### 参考文献

- 1. 陈健美. 甘蓝型油菜 2 个细胞质雄性不育系的线粒体基因组研究. [博士学位论文]. 南京: 南京 农业大学图书馆, 2011
- 方燕妮. '黔阳无核'椪柑及胞质杂种'华柚2号'雄性不育相关 miRNA 挖掘与分析. [博士学位论 文]. 武汉: 华中农业大学图书馆, 2016
- 付婧. 柑橘体细胞杂种创制及原生质体再生过程中细胞器形态观察. [硕士学位论文]. 武汉: 华 中农业大学图书馆, 2017
- 高春保. 三种同核异质细胞质雄性不育小麦线粒体 DNA 差异性分析. [硕士学位论文]. 杨陵: 西北农林科技大学,2010
- 5. 国家统计局. 中国统计年鉴. 北京: 中国统计出版社, 2000~2019
- 6. 郭文武, 叶俊丽, 邓秀新. 新中国果树科学研究 70年——柑橘. 果树学报, 2019, 36: 1264-1272
- 7. 衡双平. 油菜 hau CMS 线粒体基因组和不育基因的研究. [博士学位论文]. 武汉: 华中农业大学图书馆, 2015
- 8. 霍岩. 柑橘线粒体 DNA 高效提取技术体系的建立及其 RAPD 分析. [硕士学位论文]. 武汉: 华 中农业大学图书馆, 2008
- 9. 陆作楣, 徐福海, 张莉. 种子世界, 2014, 09:19-22
- 10. 解凯东,王惠芹,王晓培,梁武军,谢宗周,伊华林,邓秀新,Grosser JW,郭文武. 单胚性 二倍体为母本与异源四倍体杂交大规模创制柑橘三倍体.中国农业科学,2013,46: 4550-4557
- 11. 刘丹. 柑橘优异资源胚性愈伤组织诱导及体细胞杂种创制. [硕士学位论文]. 武汉: 华中农业 大学图书馆, 2019
- 12. 邱文明. "黔阳无核"椪柑无核机理及雄性不育相关基因发掘. [博士学位论文]. 武汉: 华中农业 大学图书馆, 2013
- 13. 王瑶. 细胞学和比较蛋白组学研究'黔阳无核'椪柑不育机理. [硕士学位论文]. 武汉: 华中农业 大学图书馆, 2017
- 14. 魏炽炬. 质膜囊泡制备技术的改进及线粒体在肌肉细胞的递送. [硕士学位论文]. 汕头: 汕头 大学, 2018
- 15. 肖诗鑫. 柑橘胞质杂种创制及原生质体再生过程的细胞学研究. [博士学位论文]. 武汉: 华中农业大学图书馆, 2014
- 16. 肖雪. 柑橘 NAC036 互作蛋白的筛选与验证. [硕士学位论文]. 武汉: 华中农业大学图书馆, 2017
- 17. 许亚昆,马越,胡小茜,王军.基于三代测序技术的微生物组学研究进展.生物多样性,2019, 27: 534-542
- 18. 郑蓓蓓. 基于组学的 HB 柚胞质杂种雄性不育机理及功能基因发掘. [博士学位论文]. 武汉: 华中农业大学图书馆, 2015
- Adams KL, Daley DO, Whelan J, Palmer JD. Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell*, 2002a, 14: 931-943
- 20. Adams KL, Qui YL, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences*, 2002b, 99: 9905-9912

- 21. Aleza P, Garcia-Lor A, Jua'rez J, Navarro L. Recovery of citrus cybrid plants with diverse mitochondrial and chloroplastic genome combinations by protoplast fusion followed by in vitro shoot, root, or embryo micro grafting. *Plant Cell Tissue and Organ Culture*, 2016, 126: 205-217
- 22. Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution*, 2010, 27: 1436-1448
- 23. Ariizumi T, Hatakeyama K, Hinata K, Inatsugi R, Nishida I, Sato S, Kato T, Tabata S, Toriyama K. Disruption of the novel plant protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility in *Arabidopsis thaliana*. *Plant Journal*, 2004, 2: 170-181
- 24. Arimura S, Yamamoto J, Aida GP, Nakazono M, Tsutsumi N. Frequent fusion and fission of plant mitochondria with unequal nucleoid distribution. *Proceedings of the National Academy of Sciences*, 2004, 101: 7805-7808
- Arrieta-Montiel MP, Shedge V, Davila J, Christensen AC, Mackenzie SA. Diversity of the Arabidopsis mitochondrial genome occurs via nuclear-controlled recombination activity. *Genetics*, 2009, 183: 1261-1268
- 26. Backert S, Nielsen BL, Börner T. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trends in Plant Sciences*, 1997, 2: 477-483
- 27. Balk J, Leaver CJ. The PET1-CMS mitochondrial mutation in sunflower is associated with premature programmed cell death and cytochrome c release. *Plant Cell*, 2001, 13: 1803-1818
- 28. Bassene JB, Froelicher Y, Navarro L, Ollitrault P, Ancillo G. Influence of mitochondria on gene expression in a citrus cybrid. *Plant Cell Reports*, 2011, 30: 1077-1085
- 29. Bendich AJ. Reaching for the ring: the study of mitochondrial genome structure. *Current Genetics*, 1993, 24: 279-290
- 30. Bendich AJ. The size and form of chromosomes are constant in the nucleus, but highly variable in bacteria, mitochondria and chloroplasts. *Bioessays*, 2007, 29: 474-483
- 31. Bhatnagar-Mathur P, Gupta R, Reddy PS, Reddy BP, Reddy DS, Sameerkumar CV, Saxena RK, Sharma KK. A novel mitochondrial *orf147* causes cytoplasmic male sterility in pigeonpea by modulating aberrant anther dehiscence. *Plant Molecular Biology*, 2018, 97: 131-147
- 32. Bidani A, Nouri-Ellouz O, Lakhoua L, Sihachakr D, Cheniclet C, Mahjoub A, Drira N, Gargouribouzid R. Interspecific potato somatic hybrids between *Solanum berthaultii* and *Solanum tuberosum* L. showed recombinant plastome and improved tolerance to salinity. *Plant Cell Tissue and Organ Culture*, 2007, 91: 179-189
- Bohra A, Jha UC, Adhimoolam P, Bisht D, Singh NP. Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. *Plant Cell Reports*, 2016, 35: 967-993
- 34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014, 30: 2114-2120
- 35. Carbonell-Caballero J, Alonso R, Ibanez V, Terol J, Talon M, Dopazo J. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus. Molecular Biology and Evolution*, 2015, 32: 2015-2035
- 36. Chang HS, Zhang C, Chang YH, Zhu J, Xu XF, Shi ZH, Zhang XL, Xu L, Huang H, Zhang S, Yang ZN. No primexine and plasma membrane undulation is essential for primexine deposition and plasma membrane undulation during microsporogenesis in Arabidopsis. *Plant Physiology*, 2012, 25: 1541-1554

- 37. Chase CD. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends in Genetics*, 2007, 23: 81-90
- 38. Chateigner-Boutin AL, Small I. Plant RNA editing. RNA biology, 2010, 7: 213-219
- 39. Chen LT, Liu YG. Male sterility and fertility restoration in crops. *Annual Review of Plant Biology*, 2014, 65: 579-606
- 40. Chen ZW, Zhao N, Li SS, Grover CE, Nie HS, Wendel JF, Hua JP. Plant mitochondrial genome evolution and cytoplasmic male sterility. *Critical Reviews in Plant Sciences*, 2017, 36: 55-69
- 41. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 2012, 6: 80-92
- 42. Clifton SW, Minx P, Fauron CM, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti S, Tayloe C, Meyer L, Wilson RK, Newton KJ. Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiology*, 2004, 136: 3486-3503
- 43. Covello PS, Gray MW. RNA editing in plant mitochondrial. Nature, 1989, 341: 662-666
- 44. Dambier D, Benyahia H, Pensabene-Bellavia G, Aka Kacar Y, Froelicher Y, Belfalah Z, Beniken L, Handaji N, Printz B, Morillon R, Yesiloglu T, Navarro L, Ollitrault P. Somatic hybridization for citrus rootstock breeding: an effective tool to solve important issues of the Mediterranean citrus industry. *Plant Cell Reports*, 2011, 30: 883-900
- 45. Davila JI, Arrietamontiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, Xu YZ, Weigel D, Mackenzie SA. Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. *BMC Biology*, 2011, 9: 1-14
- 46. Delannoy E, Stanley WA, Bond CS, Small ID. Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochemical Society Transactions*, 2007, 35: 1643-1647
- De Paer CV, Hongwa C, Jeziorski C, Besnard G. Mitogenomics of *Hesperelaea*, an extinct genus of Oleaceae. *Gene*, 2016, 594: 197-202
- Dewey RE, Timothy DH, Levings CS III. A mitochondrial protein associated with cytoplasmic male sterility in the T cytoplasm of maize. *Proceedings of the National Academy of Sciences*, 1987, 84: 5374-5378
- Doan TT, Carlsson AS, Hamberg M, Bülow L, Stymne S, Olsson P. Functional expression of five *Arabidopsis* fatty acyl-CoA reductase genes in *Escherichia coli*. Journal of Plant Physiology, 2009, 166: 787-796
- 50. Du M, Zhou K, Liu Y, Deng L, Zhang X, Lin L, Zhou M, Zhao W, Wen C, Xing J, Li CB, Li C. A biotechnology-based male sterility system for hybrid seed production in tomato. *Plant Journal*, 2020, 102: 1090-1100
- 51. Duchene AM, Marechal DL. The chloroplast-derived trnW and trnM-e genes are not expressed in Arabidopsis mitochondria. Biochemical and Biophysical Research Communications, 2001, 285: 1213-1216
- 52. Ducos E, Touzet P, Boutry M. The male sterile G cytoplasm of wild beet displays modified mitochondrial respiratory complexes. *Plant Journal*, 2001, 26: 171-180
- 53. Duroc Y, Gaillard C, Hiard S, Defrance MC, Pelletier G, Budar F. Biochemical and functional characterization of ORF138, a mitochondrial protein responsible for Ogura cytoplasmic male sterility in Brassiceae. *Biochimie*, 2005, 87: 1089-1100

- 54. Edera AA, Gandini CL, Sanchez-Puerta MV. Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. *Plant Molecular Biology*, 2018, 97: 215-231
- 55. Eeckhaut T, Lakshmanan PS, Deryckere D, Von Bockstaele E, Van Huylenbroeck J. Progress in plant protoplast research. *Planta*, 2013, 238: 991-1003
- 56. Ernster L, Schatz G. Mitochondrial: a historical review. Journal of Cell Biology, 1981, 91: 227-255
- 57. Fujii S, Komatsu S, Toriyama K. Retrograde regulation of nuclear gene expression in CW-CMS of rice. *Plant Molecular Biology*, 2007, 63: 405-417
- Fujii S, Toriyama K. Suppressed expression of retrograde-regulated male sterility restores pollen fertility in cytoplasmic male sterile rice plants. *Proceedings of the National Academy of Sciences*, 2009, 106: 9513-9518
- Garcia LE, Zubko MK, Zubko EI, Sanchez-Puerta MV. Elucidating genomic patterns and recombination events in plant cybrid mitochondria. *Plant Molecular Biology*, 2019a, 100: 433-450
- 60. Garcia LE, Edera A, Marfil C, Sanchez-Puerta MV. Male sterility and somatic hybridization in plant breeding. *Preprints*, 2019b, doi: 10.20944/preprints201907.0330.v1
- 61. Giegé P, Brennicke A. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proceedings of the National Academy of Sciences*, 1999, 96: 15324-15329
- 62. Goremykin VV, Salamini F, Velasco R, Viola R. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution*, 2009, 26: 99-110
- 63. Goremykin VV, Lockhart PJ, Viola R, Velasco R. The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. *Plant Journal*, 2012, 71: 615-626
- 64. Goto S, Yoshioka T, Ohta S, Kita M, Hammada H, Shimizu T. Segregation and heritability of male sterility in populations derived from progeny of Satsuma mandarin. *PLoS One*, 2016, 11: e0162408
- 65. Goto S, Yoshioka T, Ohta S, Kita M, Hammada H, Shimizu T. QTL mapping of male sterility and transmission pattern in progeny of Satsuma mandarin. *PLoS One*, 2018, 13: e0200844
- 66. Green DR, Reed JC. Mitochondrial and apoptosis. Science, 1998, 28: 1309-1312
- 67. Greenberg JT, Yao N. The role and regulation of programmed cell death in plant-pathogen interactions. *Cellular Microbiology*, 2004, 6: 201-211
- 68. Gu JN, Zhu J, Yu Y, Teng XD, Lou Y, Xu XF, Liu JL, Yang ZN. DYT1 directly regulates the expression of *TDF1* for tapetum development and pollen wall formation in Arabidopsis. *Plant Journal*, 2014, 80: 1005-1013
- 69. Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annual Review of Plant Biology, 2017, 68: 225-252
- 70. Guan YF, Huang XY, Zhu J, Gao JF, Zhang HX, Yang ZN. *RUPTURED POLLEN GRAIN1*, a member of the MtN3/saliva gene family, is crucial for exine pattern formation and cell integrity of microspore in Arabidopsis. *Plant Physiology*, 2008, 147: 852-863
- Guo WW, Deng XX. Somatic hybrid plantlets regeneration between *Citrus* and its wild relative, *Murraya paniculata* via protoplast electrofusion. *Plant Cell Reports*, 1998, 18: 297-300
- 72. Guo WW, Cheng YJ, Deng XX. Regeneration and molecular characterization of intergeneric somatic hybrids between *Citrus reticulata* and *Poncirus trifoliata*. *Plant Cell Reports*, 2002, 20: 829-834

- 73. Guo WW, Xiao SX, Deng XX. Somatic cybrid production via protoplast fusion for citrus improvement. *Scientia Horticulturae*, 2013, 163: 20-26
- 74. Handa H. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucleic Acids Research, 2003, 31: 5907-5916
- 75. Hanson MR, Bentolila S. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell*, 2004, 16: S154–S169
- 76. Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, 2008, 319: 330-333
- 77. He P, Xiao GH, Liu H, Zhang LH, Zhao L, Tang MJ, Huang S, An YJ, Yu JN. Two pivotal RNA editing sites in the mitochondrial atp1 mRNA are required for ATP synthase to produce sufficient ATP for cotton fiber cell elongation. *New Phytologist*, 2018, 218: 167-182
- 78. Heng SP, Wei C, Jing B, Wan ZJ, Wen J, Yi B, Ma CZ, Tu JX, Fu TD, Shen JX. Comparative analysis of mitochondrial genomes between the *hau* cytoplasmic male sterility (CMS) line and its iso-nuclear maintainer line in *Brassica juncea* to reveal the origin of the CMS-associated gene *orf288. BMC Genomics*, 2014, 15: 322
- 79. Heng SP, Gao J, Wei C, Chen FY, Li XW, Wen J, Yi B, Ma CZ, Tu JX, Fu TD, Shen JX. Transcript levels of *orf288* are associated with the *hau* cytoplasmic male sterility system and altered nuclear gene expression in *Brassica juncea*. *Journal of Experimental Botany*, 2018, 69: 455-466
- Hoch B, Maier RM, Appel K, Igloi GL, Kössel H. Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*, 1991, 353: 178-180
- 81. Hord CL, Chen C, Deyoung BJ, Clark SE, Ma H. The BAM1/BAM2 receptor-like kinases are important regulators of *Arabidopsis* early anther development. *Plant Cell*, 2006, 18: 1667-1680
- 82. Horn R, Kusterer B, Lazarescu E, Prufe M, Friedt W. Molecular mapping of the *Rf1* gene restoring pollen fertility in PET1-based F-1 hybrids in sunflower (*Helianthus annuus* L.). *Theoretical and Applied Genetics*, 2003, 106: 599-606
- Horn R, Gupta KJ, Colombo N. Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion*, 2014, 19: 198-205
- 84. Huang J, Zhang TY, Linstroth L, Tillman Z, Otegui MS, Owen HA, Zhao DZ. Control of anther cell differentiation by the small protein ligand TPD1 and its receptor EMS1 in *Arabidopsis*. *PloS Genetics*, 2016, 12: e1006147
- 85. Huang J, Li ZY, Biener G, Xiong EH, Malik S, Eaton N, Zhao CZ, Raicu V, Kong HZ, Zhao DZ. Carbonic anhydrases function in anther cell differentiation downstream of the receptor-like kinase EMS1. *Plant Cell*, 2017, 19: 864-869
- 86. Ichinose M, Sugita M. RNA editing and its molecular mechanism in plant organelles. *Genes*, 2017, 8: 5
- 87. Iwabuchi M, Koizuka N, Fujimoto H, Sakai T, Imamura J. Identification and expression of the kosena radish (*Raphanus sativus* cv. Kosena) homologue of the ogura radish CMS-associated gene, *orf138. Plant Molecular Biology*, 1999, 39: 183-188
- Iwahashi M, Kyozuka J, Shimamoto K. Processing followed by complete editing of an altered mitochondrial *atp6* RNA restores fertility of cytoplasmic male sterile rice. *EMBO Journal*, 1993, 12: 1437-1446

- 89. Jańska H, Wołoszyńska M. The dynamic nature of plant mitochondrial genome organization. *Acta Biochimica Polonica*, 1998, 44: 239-250
- 90. Ji C, Li H, Chen L, Xie M, Wang F, Chen Y, Liu YG. A novel rice bHLH transcription factor, DTD, acts coordinately with TDR in controlling tapetum function and pollen development. *Molecular Plant*, 2013, 6: 1715-1718
- 91. Ji JJ, Huang W, Li Z, Chai WG, Yin YX, Li DW, Gong ZH. Tapetum-specific expression of a cytoplasmic orf507 gene causes semi-male sterility in transgenic peppers. Frontiers in Plant Science, 2015, 6: 272
- 92. Jia G, Liu X, Owen HA, Zhao D. Signaling of cell fate determination by the TPD1 small protein and EMS1 receptor kinase. *Proceedings of the National Academy of Sciences*, 2008, 105: 2220-2225
- 93. Jing B, Heng SP, Tong D, Wan ZJ, Fu TD, Tu JX, Ma CZ, Yi B, Wen J, Shen JX. A male sterility-associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. *Journal of Experimental Botany*, 2012, 63: 1285-1295
- 94. Kale AA, Munjal SV. Mitochondrial respiration associated with cytoplasmic male sterility in pearl millet. *Journal of Plant Biochemistry and Biotechnology*, 2005, 14: 161-165
- 95. Kaul, Mohan LH. Male sterility in higher plants. Springer Verlag Press New York, 1998, NY, USA
- 96. Kawanable T, Ariizumi T, Kawai-Yamada M, Uchimiya H, Toriyama K. Abolition of the tapetum suicide program ruins microsporogenesis. *Plant and Cell Physiology*, 2006, 47: 784-787
- 97. Kazama T, Itabashi E, Fujii S, Nakamura T, Toriyama K. Mitochondrial *ORF79* levels determine pollen abortion in cytoplasmic male sterile rice. *Plant Journal*, 2016, 85: 707-716
- 98. Kazama T, Okuno M, Watari Y, Yanase S, Koizuka C, Tsuruta Y, Sugaya H, Toyoda A, Itoh T, Tsutsumi N, Toriyama K, Koizuka N, Arimura SI. Curing cytoplasmic male sterility via TALEN-mediated mitochondrial genome editing. *Nature Plants*, 2019, 5: 722-730
- Kemble RJ, Barsby TL, Wong RSC, Shepard JF. Mitochondrial DNA rearrangements in somatic hybrids of Solanum tuberosum and Solanum brevidens. Theoretical and Applied Genetics, 1986, 72: 787-793
- 100. Kim B, Yang T, Kim S. Identification of a gene responsible for cytoplasmic male-sterility in onions (*Allium cepa* L.) using comparative analysis of mitochondrial genome sequences of two recently diverged cytoplasms. *Theoretical and Applied Genetics*, 2019, 132: 313-322
- 101. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 2015, 12: 357-360
- 102. Kim S, Lim H, Park S, Cho KH, Sung SK, Oh DG, Kim KT. Identification of a novel mitochondrial genome type and development of molecular markers for cytoplasm classification in radish (*Raphanus sativus* L.). *Theoretical and Applied Genetics*, 2007, 115: 1137-1145
- 103. Kim YJ, Zhang DB. Molecular control of male fertility for crop hybrid breeding. *Trends in Plant Science*, 2018, 23: 53-65
- 104. Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohnomurase J, Sakai T, Kawasaki S, Imamura J. Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. *Plant Journal*, 2003, 34: 407-415
- 105. Kubo T, Mikami T. Organization and variation of angiosperm mitochondrial genome. *Physiologia Plantarum*, 2007, 129: 6-13
- 106. Kubo T, Newton KJ. Angiosperm mitochondrial genomes and mutations. *Mitochondrion*, 2008, 8:5-14
- 107. Kumar P, Kumar VD, Sharma PC, Prakash S, Bhat SR. A novel *orf108* co-transcribed with the *atpA* gene is associated with cytoplasmic male sterility in *Brassica juncea* carrying *Moricandia arvensis* cytoplasm. *Plant and Cell Physiology*, 2008, 49: 284-289
- 108. Landgren M, Glimelius K. Analysis of chloroplast and mitochondrial segregation in three different combinations of somatic hybrids produced within *Brassicaceae*. *Theoretical and Applied Genetics*, 1990, 80: 776-784
- 109. Lee HO, Choi JW, Baek JH, Oh JH, Lee SC, Kim CK. Assembly of the mitochondrial genome in the campanulaceae family using Illumina low-coverage sequencing. *Genes*, 2018, 9: 383
- 110. Lee SJ, Warmke HE. Organelle size and number in fertile and T-cytoplasmic male-sterile corn. *American Journal of Botany*, 1979, 66: 141-148
- 111. Leino M, Teixeira R, Landgren M, Glimelius K. Brassica napus lines with rearranged Arabidopsis mitochondria display CMS and a range of developmental aberrations. Theoretical and Applied Genetics, 2003, 106: 1156-1163
- 112. Leon G, Holuigue L, Jordana X. Mitochondrial complex II is essential for gametophyte development in Arabidopsis. *Plant Physiology*, 2007, 143: 1534-1546
- 113. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*, 2009, 25: 2078-2079
- 114. Li SS, Chen ZW, Zhao N, Wang YM, Nie HS, Hua JP. The comparison of four mitochondrial genomes reveals cytoplasmic male sterility candidate genes in cotton. *BMC Genomics*, 2018, 19: 775
- 115. Li WQ, Zhang XQ, Xia C, Deng Y, Ye D. MALE GAMETOPHYTE DEFECTIVE 1, encoding the F<sub>A</sub>d subunit of mitochondrial F<sub>1</sub>F<sub>0</sub>-ATP synthase, is essential for pollen formation in Arabidopsis thaliana. Plant and Cell Physiology, 2010, 51: 923-935
- 116. Li ZY, Wang Y, Huang J, Ahsan N, Biener G, Paprocki J, Thelen JJ, Raicu V, Zhao DZ. Two SERK receptor-like kinases interact with EMS1 to control anther cell fate determination. *Plant Physiology*, 2017, 173: 326-337
- 117. Lilly JW, Havey MJ. Small, repetitive DNAs contribute significantly to the expanded mitochondrial genome of cucumber. *Genetics*, 2001, 159: 317-328
- 118. Linke B, Nothnagel T, Börner T. Flower development in carrot CMS plants: mitochondria affect the expression of MADS box genes homologous to *GLOBOSA* and *DEFICIENS*. *Plant Journal*, 2003, 34: 27-37
- 119. Liu Q, Zhu AD, Chai LJ, Zhou WJ, Yu KQ, Ding J, Xu J, Deng XX. Transcriptome analysis of a spontaneous mutant in sweet orange [*Citrus sinensis* (L.) Osbeck] during fruit development. *Journal of Experimental Botany*, 2009, 60: 801-813
- 120. Liu SW, Xia GM. The place of asymmetric somatic hybridization in wheat breeding. *Plant Cell Reports*, 2014, 33: 595-603
- 121. Liu YF, Li DW, Zhang Q, Song C, Zhong CH, Zhang XD, Wang Y, Yao XH, Wang ZP, Zeng SH, Wang Y, Guo YG, Wang SB, Li XW, Li L, Liu CY, McCann HC, He WM, Niu Y, Chen M, Du LW, Gong JJ, Datson PM, Hilario E, Huang HW. Rapid radiations of both kiwifruit hybrid lineages and their parents shed light on a two-layer mode of species diversification. *New Phytologist*, 2017, 215: 877-890
- 122. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research*, 2013, 41: 575-581

- 123. Luo DP, Xu H, Liu ZL, Guo JX, Li HY, Chen LT, Fang C, Zhang QY, Bai M, Yao N, Wu H, Wu H, Ji CH, Zheng HQ, Chen YL, Ye Shan, Li XY, Zhao XC, Li RQ, Liu YG. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nature Genetics*, 2013, 45: 573-577
- 124. Luthra SK, Tiwari JK, Kumar V, Lal M. Evaluation of interspecific somatic hybrids of potato (*Solanum tuberosum*) and wild *S. cardiophyllum* for adaptability, tuber dry matter, keeping quality and late blight resistance. *Agricultural Research*, 2019, 8: 158-164
- 125. Ma H. Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annual Review of Plant Biology*, 2005, 56: 393-434
- 126. Makarenko MS, Kornienko IV, Azarin KV, Usatov AV, Logacheva MD, Markin NV, Gavrilova VA. Mitochondrial genomes organization in alloplasmic lines of sunflower (*Helianthus annuus* L.) with various types of cytoplasmic male sterility. *Peer Journal*, 2018, 6: e5266
- 127. Millar AH, Heazlewood JL, Kristensen BK, Braun HP, Møller IM. The plant mitochondrial proteome. *Trends in Plant Science*, 2005, 10: 36-43
- 128. Mitsuda N, Seki M, Shinozaki K, Ohmetakagi M. The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell*, 2005, 17: 2993-3006
- 129. Mollier P, Hoffmann B, Debast C, Small I. The gene encoding *Arabidopsis thaliana* mitochondrial ribosomal protein S13 is a recent duplication of the gene encoding plastid S13. *Current Genetics*, 2002, 40: 405-409
- 130. Nakai S, Noda D, Kondo M, Terachi T. High-level expression of a mitochondrial *orf522* gene from the male-sterile sunflower (Helianthus annuus) is lethal to *E.coli. Japanese Journal of Breeding*, 1995, 45: 233-236
- 131. Nelson BK, Cai X, Nebenfuhr A. A multicolored set of *in vivo* organelle markers for co-localization studies in Arabidopsis and other plants. *Plant Journal*, 2007, 51: 1126-1136
- 132. Ng S, Ivanova A, Duncan O, Law SR, Van Aken O, De Clercq I, Wang Y, Carrie C, Xu L, Kmiec B, Walker H, Van Breusegem F, Whelan J, Giraud E. A membrane-bound NAC transcription factor, ANAC017, mediates mitochondrial retrograde signaling in *Arabidopsis. Plant Cell*, 2013, 25: 3450-3471
- 133. Niu N, Liang W, Yang X, Jin W, Wilson ZA, Hu J, Zhang D. EAT1 promotes tapetal cell death by regulating aspartic proteases during male reproductive development in rice. *Nature Communications*, 2013, 4: 1445
- 134. Nonomura KI, Miyoshi K, Eiguchi M, Suzuki T, Miyao A, Hirochika H, Kurata N. The *MSP1* gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. *Plant Cell*, 2003, 15: 1728-1739
- 135. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics*, 2002, 268: 434-445
- 136. Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, Miyashita N, Nasuda S, Nakamura C, Mori N, Takumi S, Murata M, Futo S, Tsunewaki K. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Research*, 2005, 33: 6235-6250

- 137. Oldenburg DJ, Bendich AJ. Size and structure of replicating mitochondrial DNA in cultured tobacco cells. *Plant Cell*, 1996, 8: 447-461
- 138. Ono S, Liu H, Tsuda K, Fukai E, Tanaka K, Sasaki T, Nonomura KI. EAT1 transcription factor, a non-cell-autonomous regulator of pollen production, activates meiotic small RNA biogenesis in rice anther tapetum. *PLoS Genetics*, 2018, 14: e1007238
- 139. Park S, Grewe F, Zhu A, Ruhlman TA, Sabir J, Mower JP, Jansen RK. Dynamic evolution of Geranium mitochondrial genomes through multiple horizontal and intracellular gene transfers. *New Phytologist*, 2015, 208: 570-583
- 140. Paxson-Sowders DM, Owen HA, Makaroff CA. A comparative ultrastructural analysis of exine pattern development in wild-type *Arabidopsis* and a mutant defective in pattern formation. *Protoplasma*, 1997, 198: 53-65
- 141. Pelletier G. Chloroplast and mitochondrial genomes: manipulation through somatic hybridisation. In: Advanced Methods in Plant Breeding and Biotechnology (Murray, D.R., ed.). Wallingford, UK: C.A.B. International, pp., 1991, 201-221
- 142. Persson S, Paredez A, Carroll A, Palsdottir H, Doblin M, Piondexter P, Khitrov N, Ausr M, Somerville CR. Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis. Proceedings of the National Academy of Sciences*, 2007, 104: 15566-15571
- 143. Petsalaki E, Bagos P, Litou Z, Hamodrakas S. PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. *Genomics, Proteomics and Bioinformatics*, 2006, 4: 48-55
- 144. Picardi E, Horner DS, Chiara M, Schiavon R, Valle G, Pesole G. Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Research*, 2010, 38: 4755-4767
- 145. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, Knox EB. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science*, 2013, 342: 1468-1473
- 146. Richardson AO, Palmer JD. Horizontal gene transfer in plants. *Journal of Experimental Botany*, 2006, 58: 1-9
- 147. Sabar M, Gagliardi D, Balk J, Leaver CJ. ORFB is a subunit of F1FO-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. *EMBO Reports*, 2003, 4: 381-386
- 148. Sanchez-Puerta MV, Zubko MK, Palmer JD. Homologous recombination and retention of a single form of most genes shape the highly chimeric mitochondrial genome of a cybrid plant. *New Phytologist*, 2015, 206: 381-396
- 149. Sandhu AP, Abdelnoor RV, Mackenzie SA. Transgenic induction of mitochondrial rearrangements for cytoplasmic male sterility in crop plants. *Proceedings of the National Academy of Sciences*, 2007, 104: 1766-1770
- 150. Sears ER. Genetics and farming. US Dept. Agr. Yearbool, 1947, 245-255
- 151. Sheahan MB, McCurdy DW, Rose RJ. Mitochondria as a connected population: ensuring continuity of the mitochondrial genome during plant cell dedifferentiation through massive mitochondrial fusion. *Plant Journal*, 2005, 44: 744-755
- 152. Shearman JR, Sonthirod C, Naktang C, Pootakham W, Yoocha T, Sangsrakru D, Jomchai N, Tragoonrung S, Tangphatsornruang S. The two chromosomes of the mitochondrial genome of a

sugarcane cultivar: assembly and recombination analysis using long PacBio reads. *Scientific Reports*, 2016, 6: 31533

- 153. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. Rapid evolution of enormous multi-chromosomal genomes in flowering plant mitochondrial with exceptionally high mutation rates. *PLoS Biology*, 2012, 10: e1001241
- 154. Sproule A, Donaldson P, Dijak M, Bevis E, Pandeya R, Keller WA, Gleddie S. Fertile somatic hybrids between transgenic *Nicotiana tabacum* and transgenic *N. debneyi* selected by dual-antibiotic resistance. *Theoretical and Applied Genetics*, 1991, 82: 450-456
- 155. Stern DB, Goldschmidt-Clermont M, Hanson MR. Chloroplast RNA metabolism. *Annual Review* of Plant Biology, 2010, 61: 125-155
- 156. Sun T, Bentolila S, Hanson MR. The unexpected diversity of plant organelle RNA editosomes. *Trends in Plant Science*, 2016, 21: 962-973
- 157. Takenaka M, Brennicke A. RNA editing in plant mitochondria: assays and biochemical approaches. *Methods in Enzymology*, 2007, 424: 439-458
- 158. Takenaka M, Zehrmann A, Verbitskiy D, Hartel B, Brennicke A. RNA editing in plants and its evolution. *Annual Review of Genetics*, 2013, 47: 335-352
- 159. Tanaka Y, Tsuda M, Yasumoto K, Yamagishi H, Terachi T. A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (*Raphanus sativus* L.). *BMC Genomics*, 2012, 313: 352
- 160. Tiwari JK, Devi S, Ali N, Luthra SK, Kumar V, Bhardwaj V, Singh RK, Rawat S, Chakrabarti SK. Progress in somatic hybridization research in potato during the past 40 years. *Plant Cell Tissue and Organ Culture*, 2018, 132: 225-238
- 161. Trabelsi S, Gargouri-Bouzid R, Vedel F, Nato A, Lakhoua L, Drira N. Somatic hybrids between potato *Solanum tuberosum* and wild species *Solanum vernei* exhibit a recombination in the plastome. *Plant Cell Tissue and Organ Culture*, 2005, 83: 1-11
- 162. Unseld M, Marienfeld JR, Brandt P, Brennicke A. The mitochondrial genome of Arabidopsis *thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics*, 1997, 15: 57-61
- 163. Uyttewaal M, Arnal N, Quadrado M, Martin-Canadell A, Vrielynck N, Hiard S, Gherbi H, Bendahmane A, Budar F, Mireau H. Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell*, 2008, 20: 3331-3345
- 164. Vedel F, Pla M, Vitart V, Gutierres S, Ch'etrit P, De Paepe R. Molecular basis of nuclear and cytoplasmic male sterility in higher plants. *Plant Physiology and Biochemistry*, 1994, 32: 601-608
- 165. Verma N. Transcriptional regulation of anther development in Arabidopsis. Gene, 2019, 202-209
- 166. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 2014, 9: e112963
- 167. Wang J, Jiang J, Li X, Li A, Zhang Y, Guan R, Wang Y. Complete sequence of heterogenous-composition mitochondrial genome (*Brassica napus*) and its exogenous source. *BMC Genomics*, 2012, 13: 675
- 168. Wang J, Jiang J, Wang Y. Protoplast fusion for crop improvement and breeding in China. *Plant Cell Tissue and Organ Culture*, 2013, 112: 131-142

- 169. Wang K, Gao F, Ji Y, Liu Y, Dan Z, Yang P, Zhu Y, Li S. ORFH79 impairs mitochondrial function via interaction with a subunit of electron transport chain complex III in Honglian cytoplasmic male sterile rice. *New Phytologist*, 2013, 198: 408-418
- 170. Wang L, Pan ZY, Guo WW. Proteomic analysis of leaves from a diploid cybrid produced by protoplast fusion between Satsuma mandarin and pummelo. *Plant Cell Tissue and Organ Culture*, 2010, 103: 165-174
- 171. Wang M, Yan W, Peng XQ, Chen ZF, Xu CJ, Wu J, Deng XW, Tang XY. Identification of late-stage pollen-specific promoters for construction of pollen-inactivation system in rice. *Journal of Integrative Plant Biology*, 2020, 62: 1246-1263
- 172. Wang R, Fang YN, Wu XM, Qing M, Li CC, Xie KD, Deng XX, Guo WW. (2020). The miR399-CsUBC24 module regulates reproductive development and male fertility in citrus. Plant Physiology, 2020, 183: 1681-1695
- 173. Wang SB, Li DW, Yao XH, Song QW, Wang ZP, Zhang Q, Zhong CH, Liu YF, Huang HW. Evolution and diversification of Kiwifruit mitogenomes through extensive whole-genome rearrangement and mosaic loss of intergenic sequences in a highly variable region. *Genome Biology and Evolution*, 2019, 11: 1192-1206
- 174. Wang X, Xu YT, Zhang SQ, Cao L, Huang Y, Cheng JF, Wu GZ, Tian SL, Chen CL, Liu Y, Yu HW, Yang XM, Lan H, Wang N, Wang L, Xu JD, Jiang XL, Xie ZZ, Tan ML, Larkin RM, Chen LL, Ma BG, Ruan YJ, Deng XX, Xu Q. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature Genetics*, 2017, 49: 765-772
- 175. Wang ZH, Zou YJ, Li XY, Zhang QY, Chen LT, Wu H, Su DH, Chen YL, Guo JX, Luo D, Long YM, Zhong Y, Liu YG. Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell*, 2006, 18: 676-687
- 176. Warmke HE, Lee SL. Pollen abortion in T cytoplasm male sterile corn (*Zea mays*): a suggested mechanism. *Science*, 1978, 200: 561-563
- 177. Wei L, Yan ZX, Ding Y. Mitochondrial RNA editing of F0-ATPase subunit 9 gene (ATP9) transcripts of Yunnan purple rice cytoplasmic male serile line and its maintainer line. *Acta Physiologiae Plantarum*, 2008, 30: 657-662
- 178. Wellmer F, Graciet E, Riechmann JL. Specification of floral organs in *Arabidopsis*. Journal of Experimental Botany, 2014, 65: 1-9
- 179. Wilson ZA, Zhang DB. From *Arabidopsis* to rice: pathways in pollen development. *Journal of Experimental Botany*, 2009, 60: 1479-1492
- 180. Wu Z, Hu K, Yan M, Song L, Wen J, Ma C, Shen J, Fu T, Yi B, Tu J. Mitochondrial genome and transcriptome analysis of five alloplasmic male-sterile lines in *Brassica juncea*. *BMC Genomics*, 2019, 20: 1-15
- 181. Xia GM. Progress of chromosome engineering mediated by asymmetric somatic hybridization. Journal of Genetics Genomics, 2009, 36: 547-556
- 182. Xiang F, Xia GM, Zhi D, Wang J, Nie H, Chen M. Regeneration of somatic hybrids in relation to the nuclear and cytoplasmic genomes of wheat and *Setaria italic. Genome*, 2004, 47: 680-688
- 183. Xiao HJ, Zhang QN, Qin XJ, Xu YH, Ni CZ, Huang JH, Zhu LL, Zhong FY, Liu W, Yao GX, Zhu YG, Hu J. Rice *PPS1* encodes a DYW motif-containing pentatricopeptide repeat protein required for five consecutive RNA-editing sites of *nad3* in mitochondria. *New Phytologist*, 2018, 220: 878-892

- 184. Xie HW, Peng XJ, Qian MJ, Cai YC, Ding X, Chen QS, Cai QY, Zhu YL, Yan LA, Cai YH. The chimeric mitochondrial gene *orf182* causes non-pollen-type abortion in Dongxiang cytoplasmic male-sterile rice. *Plant Journal*, 2018, 95: 715-726
- 185. Xiong J, Tao T, Luo Z, Yan SG, Liu Y, Yu XQ, Liu GL, Xia H, Luo LJ. RNA editing responses to oxidative stress between a wild abortive type male-serile line and its maintainer line. *Frontiers in Plant Science*, 2017, 8: 2023
- 186. Xu CH, Xia GM, Zhi DY, Xiang FN, Chen HM. Integration of maize nuclear and mitochondrial DNA into the wheat genome through somatic hybridization. *Plant Science*, 2003, 165: 1001-1008
- 187. Xu Q, Chen LL, Ruan X, Chen DJ, Zhu AD, Chen CL, Bertrand D, Jiao WB, Hao BH, Lyon MP, Chen JJ, Gao S, Xing F, Lan H, Chang JW, Ge XH, Lei Y, Hu Q, Miao Y, Wang L, Xiao SX, Biswas MK, Zeng WF, Guo F, Cao HB, Yang XM, Xu XW, Cheng YJ, Xu J, Liu JH, Luo OJ, Tang ZH, Guo WW, Kuang HH, Zhang HY, Roose ML, Nagarajan N, Deng XX, Ruan YJ. The draft genome of sweet orange (Citrus sinensis). *Nature Genetics*, 2013, 45: 59-66
- 188. Xu YS, Jones MGK, Karp A, Pehu E. Analysis of the mitochondrial DNA of the somatic hybrids of Solanum brevidens and S. tuberosum using non-radioactive digoxigenin-labelled DNA probes. Theoretical and Applied Genetics, 1993, 85: 1017-1022
- 189. Yamamoto M, Matsumoto R, Okudai N, Yamada Y. Aborted anthers of *Citrus* result from gene-cytoplasmic male sterility. *Scientia Horticulturae*, 1997, 70: 9-14
- 190. Yang J, Liu X, Yang X, Zhang M. Mitochondrially-targeted expression of a cytoplasmic male sterility-associated *orf220* gene causes male sterility in *Brassica juncea*. *BMC Plant Biology*, 2010, 10: 231
- 191. Yang WC, Ye D, Xu J, Sundaresan V. The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encondes a novel nuclear protein. Genes and Development, 1999, 13: 2108-2117
- 192. Yang YF, Zhu GN, Li R, Yan SJ, Fu DQ, Zhu BZ, Tian HQ, Luo YB, Zhu HL. The RNA editing factor SIORRM4 is required for normal fruit ripening in tomato. *Plant Physiology*, 2017, 175: 1690-1702
- 193. Yao N, Tada Y, Sakamoto M, Nakayashiki H, Park P, Tosa Y, Mayama S. Mitochondrial oxidative burst involved in apoptotic response in oats. *Plant Journal*, 2002, 30: 567-579
- 194. Yao X, Li J, Liu J, Liu K. An Arabidopsis mitochondria-localized RRL protein mediates abscisic acid signal transduction through mitochondrial retrograde regulation involving ABI4. Journal of Experimental Botany, 2015, 66: 6431-6445
- 195. Yu FG, Bi CW, Wang XL, Qian X, Ye N. The complete mitochondrial genome of *Citrus sinensis*. *Mitochondrial DNA Part B*, 2018, 3: 592-593
- 196. Zahn LM, Feng BM, Ma H. Beyond the ABC-model: Regulation of floral homeotic genes. *Advances in Botanical Research*, 2006, 163-207
- 197. Zhang C, Zhang M, Zhu Q, Liu Z, Yan F, Wu L, Xu X, Zhou X, Chen X. Cytological observation of pollen development in 'Ougan' (Citrus suavissima Hort. ex Tanaka) and its seedless mutant. *Journal of Fruit Science*, 2014, 31: 265-269
- 198. Zhang C, Yu D, Ke F, Zhu M, Xu J, Zhang M. Seedless mutant 'Wuzi Ougan' (*Citrus suavissima* Hort. ex Tanaka 'seedless') and the wild type were compared by iTRAQ based quantitative proteomics and integratedly analyzed with transcriptome to improve understanding of male sterility. *BMC Genomics*, 2018, 19: 106

- 199. Zhang D, Luo X, Zhu L. Cytological analysis and genetic control of rice anther development. *Journal of Genetics and Genomics*, 2011, 38: 379-390
- 200. Zhang W, Sun Y, Timofejeva L, Chen C, Grossniklaus U, Ma H. Regulation of *Arabidopsis* tapetum development and function by *DYSFUNCTIONAL TAPETUM1 (DYT1)* encoding a putative bHLH transcription factor. *Development*, 2006, 133: 3085-3095
- 201. Zhang Y, Ye JL, Liu CY, Xu Q, Long LC, Deng XX. Citrus PH4-Noemi regulatory complex is involved in proanthocyanidin biosynthesis via a positive feedback loop. *Journal of Experimental Botany*, 2020, 71: 1306-1321
- 202. Zhao DZ, Wang GF, Speal B, Ma H. The *Excess Microsporocytes1* gene encodes a putativeleucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the Arabidopsis anther. *Genes and Development*, 2002, 16: 2021-2031
- 203. Zhao J, Long T, Wang Y, Tong X, Tang J, Li J, Wang H, Tang L, Li Zhi, Shu Y, Liu X, Li S, Liu H, Li J, Wu Y, Zhang J. *RMS2* encoding a GDSL lipase mediates lipid homeostasis in anthers to determine rice male fertility. *Plant Physiology*, 2020, 182: 2047-2064
- 204. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist*, 2016, 209: 1747–1756
- 205. Zhu CQ, Zheng XJ, Huang Y, Ye JL, Chen P, Zhang CL, Zhao F, Xie ZZ, Zhang SQ, Wang N, Li H, Wang L, Tang XM, Chai LJ, Xu Q, Deng XX. Genome sequencing and CRISPR/Cas9 gene editing of an early flowering mini citrus (*Fortunella hindsii*). *Plant Biotechnology Journal*, 2019, 17: 2199-2210

# 附录 Ⅰ 线粒体提取液配

	Homogenization medium		V	Vash buffer	精提取 buffer	
药物	终浓度	配 500mL 用量	终浓度	配 500mL 用量	终浓度 (A)	终浓度(B)
甘露醇	0.4 M	36.44 g	0.4 M	36.44 g	-	-
EDTA	5 mM	0.93 g	5 mM	0.93 g	-	20 mM
Cysteine	8 mM	0.703	-	-	-	-
Tricine 麦 黄酮	10 mM	0.896	-	-	-	-
BSA 牛血 清蛋白	0.1 %	0.5 g	0.1 %	0.5 g	-	-
PVP-40 聚						
乙烯吡咯	1 %	5 g	-	-	-	-
烷酮						
MOPS	25 mM	2.62 g	25 mM	2.62 g	-	-
蔗糖	-	-	-	-	0.5 M	0.5 M
Tris-HCL	-	-	-	-	50 mM	10 mM
MgCl <sub>2</sub>	-	-	-	-	10 mM	
NADH	调 pH 至 7.8		调	J pH 至 7.2	调 ph	至 7.5

Appendix I The formula of mitochondrial extract liquid

#### 附录 II 部分实验操作步骤

#### 方法 I 线粒体 DNA 提取

- 提取的线粒体约 200 µL 装入 1.5 mL 离心管,加入 500 µL 预热的(65 ℃ 水浴) CTAB 提取液(1.5%巯基乙醇),65 ℃ 水浴 60-90 min,每隔 15 min 摇 匀一次。
- 加入 700 μL 氯仿/异戊醇(24:1)溶液,上下混匀 10 min, 12000 g 离心 20 min, 吸上清于另一离心管,且重复该步骤。
- 加入 60 μL 5 M NaCl 和 1 mL 冰冻无水乙醇(-20 °C),轻轻混匀后于-20 °C
  冷冻 30 min 以沉淀 DNA。
- 4. 12000g离心10min,弃上清,加入1mL70%乙醇浸泡过夜。
- 弃乙醇,适当风干 DNA 沉淀,加入 50 μL TE 和 5 μL 20 μg/mL RNaseA, 37 ℃ 孵育过夜。
- 6. 检测 DNA 浓度及质量。

#### 方法 Ⅱ 载体构建及 TA 克隆

载体构建使用一步法定向无缝克隆试剂盒(CloneExpress, Vazyme), 具体操作流程为:

- 1. 制备线性化载体。使用限制性内切酶切消化,获得线性化载体。
- 插入片段扩增引物设计及 PCR 扩增目的片段。引物设计的原则是:在引物的
  端引入线性化克隆载体末端 20 bp 左右同源序列,使 PCR 扩增产物的目的片段 5'和 3'末端分别带有和线性化克隆载体两末端相对应的完全一致的序列。
- 3. 回收目的片段。
- 冰上配制重组反应体系: 4 μL 5×CE II Buffer, 2 μL Exnase<sup>TM</sup> II,线性化克隆 载体与 PCR 目的片段回收产物浓度按照 1/3 加入,加入 ddH<sub>2</sub>O 补齐至 20 μL。
- 5. 轻轻吸打混匀, 37 ℃ 温育 30 min。反应完成后冰上冷却 5 min, 同时从-80 ℃ 冰箱中取出大肠杆菌感受态解冻。
- 将 50 μL 大肠杆菌感受态细胞和 10 μL 重组反应液加入冷冻好的离心管中, 轻轻吸打混匀,冰上放置 30min。

- 7. 将冰浴的离心管置于42°C水浴锅热激90s,然后迅速置于冰中冷却2-3 min。
- 8. 加入 400 μL 液体 LB 培养基, 37 °C 摇床中摇动 45-60 min。
- 加入 400 μL 液体 LB 培养基混匀后倒入相应抗生素的 LB 固体平板上, 37 ℃ 培养 12-16 h。
- 10. 阳性克隆鉴定。

## 附录 III 本研究中所使用到的引物序列

Appendix III The sequences of the primers used in this study

Primer Name	Primer Sequences	Primer Name	Primer Sequences
For mitochondrial genome ass	embling		
G1-1F	TCCTGTAGACCGTGAG	HB-1F	TACCAAGCCACTCTTCAAACCC
G1-1R	TGCCCATCTTGGTCTC	HB-1R	TTATCGAACGAACCGCACTCCT
G1-2F	TGACTTATCGGCTGCCTTGCTT	HB-2F	ACAGCTACTTTGCCATTCCAGC
G1-2R	GTAAGGGAAATGGAAG	HB-2R	TCAGTTTGGGCACTAGGCAGTA
G1-3F	GGATTTAGAGGACCCA	HB-3F	GGGAACCGAAAGAGCAGAAATG
G1-3R	TATGCAAGTCCGCAGCATTGTG	HB-3R	GGCTCTGAACAGGAGTAACGCAAAC
G1-4F	TGCTACGGAACTACCT	HB-4F	TACCATGAACGGACTCTTTGCC
G1-4R	ACGGAGCAGATTTGAC	HB-4R	TGTACCACAACACCTGGACAATCAG
G1-5F	AAAGGTCCATTTACAG	HB-5F	CTCGGTTCTTGAGACGGAGATTAGG
G1-5R	ATGCCACAGAAGGAAC	HB-5R	TCATTAGTGGATTCCCATTGGTTAG
G1-6F	AACCGATAGCGGAGAC	HB-6F	AGCACGAACCAAGATCCATACC
G1-6R	ATGTCGGCTTGAGTAA	HB-6R	AGGAAGCGGCGTGAACTATTCT
G1-7F	CTTGAGCATTTCCCTCTAATACCTT	HB-7F	AACCCACAGGAATCCATAATAGAAC
G1-7R	TCCCTCTCCGGATTCTCGTATA	HB-7R	GTGCCCGACCTGAGTAAAGTTGTAG
G1-8F	TCTCCCGAACCGAAGT	HB-8F	CTAATCCGCAAGGAAAGGACAAGAA
G1-8R	CATCGTCCCTCGCAACAACATT	HB-8R	CCGTAATCACCCGATTGAAGAG
G1-9F	TTCATTGCCCTTACCC	HB-9F	CGAAGAAGGGCAGAAACTTACATAC
G1-9R	GCCCGTGTTTCATCCATTTC	HB-9R	ACATTTCTATGGATTGTAGCGTCAG
G1-10F	ATTCCGACAGGTCTCACTCATC	HB-10F	AGTGTAACCGACTTGTTTCATCAGC
G1-10R	ACCTGCTACGGAACTA	HB-10R	AACCAGGAGACGGAATTGGAAC
G1-11F	TCCGTGGCACGAAATCCATCAT	HB-11F	GCAGTCTACGGCACCTAAACAAACA
G1-11R	ATATGACAGGCCCAGAAGATGG	HB-11R	AAGGACGAAGCGGCTCTAATCT

G1-12F	AGAGTCAGTAGCTGGGAATCTC	HB-12F	GAAGCAATCTGGACGGAACCTA					
G1-12R	CCATTCCTCGTGAGCCACTTAT	HB-12R	GAGTTCTTCCGTTGGAGTTTCG					
G1-13F	TTCTGTGCCTCAAGTT	HB-13F	TGGCTGCGTTAGGAAATACTGG					
G1-13R	AGTGTTCTTGCCTCTA	HB-13R	GAAGAACGGGAACGAAACGAAATAA					
G1+BTC-1F	AGTGATGGAAGCAATTTTGACC	HB-14F	TATGACACGGTCAGGTTCAGCA					
G1+BTC-1R	CGATTTAGCATGTGGTGGATCA	HB-14R	GTCATTATTGTTACGCCGACCC					
G1+BTC-2-1F	AGCGTTCTGTGCCTCAAGTT	G1+BTC-3F	CTAGCTACTGACTATGGTACAC					
G1+BTC-2-1R	CTCCTGGTTGAGGAGACGAAAT	G1+BTC-3R	AAGCTGCTTGCTGACTTATAC					
G1+BTC-2-2F	GAATTCTCAACCCGAGATGTTA	G1+BTC-5F	ATGGAAAGAGATGGGAGGCTAG					
G1+BTC-2-2R	CACGTCCGAAGGAATCAATGTT	G1+BTC-5R	AAAAGTCTTCTTCGTGTCTCCT					
G1+BTC-7F	AGAAGGTTCTCCTGAGAAGGAA	G1+BTC-12F	GTCTGTACGGTGAAGTACATTG					
G1+BTC-7R	CGCATACCCTCATTCAGCTTAA	G1+BTC-12R	AGTCGGTCAGTAGTGAGTCCTA					
G1+BTC-8F	CAGCGAGTCCTTTCTTCCTGTT	G1+BTC-14F	ACTGCTTGGCGTTGCATATTAG					
G1+BTC-8R	ATCTGATCTATCTACTCGGCA	G1+BTC-14R	CGAGAGCGTCTCTGAATACTCA					
G1-3r-1F	CTAGAGTACATGACTAGACTGC	HBP-3r-1F	AACTTTTCCGCCCACTGCCACTA					
G1-3r-1R	GTCGTAGAAGCTGTTCCTAGGA	HBP-3r-1R	CCAACTCATTGAGTGTAACCGA					
G1-3r-2F	ATTGGCGCTTAATAGCCTGTAG	HBP-3r-2F	CATACCCTCATTCAGCTTAAGG					
G1-3r-2R	TTCACCCAAGCTTCATCCTGGT	HBP-3r-2R	CAGGAATGAAACCTCTCGTAGT					
G1-3r-3F	GTGTATAAGGACCAACGACGAG	G1+HBP-3r-1F	CAATGTACTCCACGGATAGAGG					
G1-3r-3R	TCATAACACTGATGTCAAGCCC	G1+HBP-3r-1R	AGAGTTCTTATGTCTTTCCGCG					
For the homologous fragement								
P1	CGGAATCGCGTTAAGAAGATCA	P6	TGACTTCACGGTCGCCAAAGAA					
P2	CACCTACTGTTCGTTAACAGTT	P7	CAAGGAGTTCATCGATTGAAGT					
Р3	ACTTCAATCGATGAACTCCTTG	P8	ATGGCTATAACAGAGTTTCTGT					
P4	CTAATCGATTACGATAAATTGGATCGGAT	Р9	CAATTTGGGATCCAATTCGGGA					
P5	TGCGGCATTTCCACGATATCGT	P10	CCGCTACTATTACTATGTGAAAACGTCCG					
For protein subcellualr localiza	For protein subcellualr localization							
CsNADH-207F GGGGACAAGTTTGTACAAAAAGCAGGCTTCATGGCTATGAAGAGCGCGT								

CsNADH-207R	GGGGACCACTTTGTACAAGAAAGCTGGGTGCTCTTTCTTGGAGG
orf88-207F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTTAGTTA
orf88-207R	GGGGACCACTTTGTACAAGAAAGCTGGGTGGCTCTCAAGCTGTTGACAATCTA
For Y2H assays	
16978-pGB-F	ATATGGCCATGGAGGCCGAATTCATGCCTCTCATCAATGTCATCC
16978-pGB-R	GGGGTTATGCTAGTTATGCGGCCGCAGGTTTCATTACTTTTCTTCCA
orf88-pGB-F	ATATGGCCATGGAGGCCGAATTCATGTTAGTTATGACTAGTCT
orf88-pGB-R	GGGGTTATGCTAGTTATGCGGCCGCTTAGCTCTCAAGCTGTTGACAATCTA
CsNADH-pAD-F	GCCATGGAGGCCAGTGAATTCATGAGATTTCCATCAAGAAGT
CsNADH-pAD-R	ACGATTCATCTGCAGCTCGAGTTACTCTTTGGAGG
For LUC	
orf88-cLUCF	TACGCGTCCCGGGGCGGTACCATGTTAGTTATGACTAGTCT
orf88-cLUCR	ACGAAAGCTCTGCAGGTCGACGCTCTCAAGCTGTTGACAATCTA
CsNADH-nLUCF	ACGGGGGACGAGCTCGGTACCATGGCTATGAAGAGCGCGT
CsNADH-nLUCR	CGCGTACGAGATCTGGTCGACCTCTTTCTTGGAGG
For the binary vector	
OE-DN16978-F	GGTTTCACTGCAGGATGCCTCTCATCAATGTCATCC
OE-DN16978-R	GGGGAAATTCGAGCTGGTTACCAGGTTTCATTACTTTTCTTCCA

### 附录 IV 本研究中使用的特异性 MSS 引物序列信息

Primer names	Primer sequences	Fragment Size (bp)	<b>Corresponding MSS</b>	
MG1	F: GTTGTCTGATCACACTCGAAAT	200	MCC1	
WIS1	R: TCACGTCCTCAGAGCCCGAGAA	288	141991	
MG2	F: ATTTGATCATCTTCTTAACTAA	7570	MCCO	
W152	R: TCCGTATCGATAGTGGAAGGTT	1312	M552	
MG2	F: GAATTTGTAGAGCAAGAACAAG	1062	MCC2	
W155	R: TTCTTCAAACCCCCAATCCGCT	1005	M222	
MGA	F: AATCAAGGGCTAAACCTGAAGC	GAAGC 2(1	MCCA	
W154	R: CCACTCGCCCTTGGTTAGAGAC	201	M554	
M85	F: CCGCTGACTGTTCTCCGAACGG	222	MSS5	
W155	R: TGGTTCTTGAAAGAACAATTCT	233	101000	
MS6	F: TCGCTGGCTCCGTGACAACAGA	253	MSS6	
14150	R: TATGGTGGATCAATAATAGGCC	255	111550	
MS7	F: TAATTTGAAGACAAAATGGTAT	260	MGG7	
141.57	R: ACAGAAGAGAAGAAGTCATTTT	200	111557	
MS8	F: AAAAGAATGTTGTTATGCTGCC	8014	MSS8	
14150	R: AGCAATTCATGCGAGAATTGGT	0714	W550	
MSO	F: CAGTGTAACTGAGAGTAAAAGT	960	MSSO	
141.57	R: TTTACGAATGGTTCATTCCCTA	900	111557	
MS10	F: GTTGACCTCACCAACTACGTGC	4250	MCC10	
MIGIN	R: CTAGGGCTGGTCTCTCTGGCCT	4230	MISSIU	

Appendix IV Sequences and information regarding specific MSS primers used in this study

MC11	F: ATTTTATTTTATTTATTCCCT	665	MSS11	
WISTI	R: AATCCTCTTCCCGCCGAGCCCC	005	MSSII	
MC13	F: TGACTTGCCCCCCCCGCCGTG	424	MSS12	
WIS12	R: CAGACGCGAGCCCCTCCTCGGG	424	WISS12	
MC12	F: GCTTAGTTGCCACCGTCGAGTT	1106	MCC12	
W1515	R: GGCTCGAATGGTACGATCCCTC	1190	M3313	
MC14	F: ACAGAGATCTCATTCGTCCCAC	1422		
IVIS14	R: GGGCAGGAAATCGTACAGGCCC	1422	MS814	
M61 <i>5</i>	F: GGTCTTTTAATAACTCATTGAT	5251	MSS15	
W1515	R: CTGTAAGGGGGGGATGGAATAGT	5551		
MG17	F: AGAAGACAGCTGGACTTCAGAG	1270	MSS16	
W1510	R: TTGAGTCGCCTTTCTCTTTCCT	13/2	M3310	
M617	F: AATCGATTCGATTCTCGTTCTC	117	10017	
<b>WIST</b> /	R: ACTATGGGAAGGGACGAGACAA	11/	M3317	
MC10	F: AGCTTCAGAGGTGGGGAAACCC	2057	MCC10	
W1518	R: CAACGAGGTGGAGGGCCATCGC	2937	M3318	
MC10	F: GATCTGTTCCAACGTCGGCTAG	4008	MCC10	
W1519	R: TAAGAAAGAGTCGTGCTGTGAG	4008	M5519	
MCOO	F: GAGTGGGTATGCGGGCTTCTTT	1205	MCCOO	
111520	R: AGCGTTAGCACTTTACTAATAA	1295	INIS520	

# 附录 V CMS 特异 ORF 信息

Appendix V Information of CMS-specific ORF

G1 ORFs	G1+HBP ORFs	Locations in G1 mitochondrial genome	Locations in G1+HBP mitochondrial genome	Orientation	Annotation in G1 ORFs of corresponding transcripts	Annotation in G1+HBP ORFs of corresponding transcripts	MSS
orf309	orf309	24-953	24-953	Direct	atp1,nad5	atp1,nad5	M1
orf70	orf70	224-436	224-436	Reverse	atp1,nad5	atp1,nad5	M1
orf90	orf90	49993-50265	49993-50265	Direct	ccmFn,ccmFC, ccmC,nad5	nad5	M2
orf260	orf260	50110-50892	50110-50892	Reverse	ccmFn,ccmFC, ccmC,nad5	nad5	M2
orf74	orf74	50515-50739	50515-50739	Direct	ccmFn,ccmFC, ccmC,nad5	nad5	M2
orf55a	orf55a	50627-50794	50627-50794	Direct	ccmFn,ccmFC, ccmC,nad5	nad5	M2
orf141	orf141	50961-51386	50961-51386	Reverse	ccmFn,ccmFC, ccmC,nad5	nad5	M2
orf69	orf69	51195-51404	51195-51404	Direct	ccmFn,ccmFC, ccmC,nad5	nad5	M2
orf161	orf161	51675-52160	51675-52160	Reverse	ccmFn,ccmFC, ccmC,nad5	nad5	M2

orf52	orf52	51896-52054	51896-52054	Reverse	ccmFn,ccmFC,	nad5	M2	
					ccmC,nad5			
out101	ouf104	orf104 52141 52455	52141 52455	Davanaa	ccmFn,ccmFC,	m a d 5	MO	
0rj104	01/104	52141-52455	52141-52455	Keveise	ccmC,nad5	naas	IVIZ	
ortssh	orf55h	52645-52812	52645-52812	Direct	ccmFn,ccmFC,	nad5	M2	
013550	013550	52015 52012	52045 52012	Direct	ccmC,nad5	nuus	1112	
orf143	orf143	52747-53178	52747-53178	Reverse	ccmFn,ccmFC,	nad5	M2	
01/145	01/145	52747-55176	52747-55176		ccmC,nad5	nuus	1112	
orf210	orf210	52863-53495	52863-53495	Direct	ccmFn,ccmFC,	nad5	M2	
01/210	01/210	52005 55475	52005 55475	Direct	ccmC,nad5	nuus	1012	
orf78	orf78	53033-53269	53033-53269	Reverse	ccmFn,ccmFC,	nad5	M2	
01770	01770	55055 55207	55055 55207	iceverse	ccmC,nad5	nuus	1012	
orf69a	orf69a	orf69a 53653-53862	53653-53862	Direct	ccmFn,ccmFC,	nad5	M2	
orjosu					ccmC,nad5	naus	1112	
orf51	orf51	53698-53853	53698-53853	Reverse	ccmFn,ccmFC,	nad5	M2	
01351	01301	55078-55855	55090-55055	Reverse	ccmC,nad5	nuus	1,12	
orf76	orf76	54685-54915	54685-54915	Reverse	ccmFn,ccmFC,	nad5	M2	
01970	01,770	01000 01910	51005 51915	iceverse	ccmC,nad5	nuus	1112	
orf54	orf54	55771-55935	55771-55935	Direct	ccmFn,ccmFC,	nad5	M2	
01951	01301	00771 00900	55771 55555	Direct	ccmC,nad5	nuus	1112	
orf50	orf50	56279-56431	56279-56431	Direct	ccmFn,ccmFC,	nad5	M2	
01950	01,50	56275 56151	50277 50 151	Direct	ccmC,nad5	nuus	1112	
orf50a	orf50a	57655-57807	57655-57807	Reverse	ccmFn,ccmFC,	nad5	M2	
013000	01300	5/055-5/00/	5/055-5/00/	NEVEISE	ccmC,nad5	nuus	IVIZ	
orf81	orf81	orf&1 58070 58221	58079-58324	Davarse	ccmFn,ccmFC,	nad5		
01301	01301	5007 <i>5-</i> 5052 <del>4</del>	50077-5052 <b>4</b>	iceverse	ccmC,nad5	nuus		

orf70a	orf70a	62042-62254	62042-62254	Reverse	ccmFn,ccmFC, ccmC,nad5	nad5	M3
orf320	orf320	62124-63086	62124-63086	Direct	nad5	nad5	M3
orf114	orf114	62299-62643	62299-62643	Reverse	nad5	nad5	M3
orf67	orf67	62596-62799	62596-62799	Direct	nad5	nad5	M3
orf131	orf131	62698-63093	62698-63093	Reverse	nad5	nad5	M3
orf68	orf68	88160-88366	88160-88366	Direct	ccmB	ccmB	
	orf62		114262-114450	Reverse		atp4,nad4L	M5
	orf61		115406-115591	Reverse		atp4,nad4L	
orf86	orf86	118177-118437	118177-118437	Direct	atp4,nad4L	atp4,nad4L	
orf56		125870-126040		Reverse	atp4,nad4L		
orf52a	orf52a	138971-139129	138971-139129	Direct	atp4,nad4L	atp4,nad4L	M7
orf102	orf102	139037-139345	139037-139345	Reverse	atp4,nad4L	atp4,nad4L	M7
orf74a	orf74a	139095-139319	139095-139319	Direct	atp4,nad4L	atp4,nad4L	M7
orf1244	orf1244	146102-149836	146102-149836	Reverse	atp6	atp4,nad4L	M8
orf50b	orf50b	146148-146300	146148-146300	Direct	atp6	atp4,nad4L	M8
orf60	orf60	147963-148145	147963-148145	Direct	atp6	atp4,nad4L	M8
orf53	orf53	148179-148340	148179-148340	Direct	atp6	atp4,nad4L	M8
orf119	orf119	149829-150188	149829-150188	Reverse	atp6	atp4,nad4L	M8
orf71		190082-190297		Reverse	rpl2		
orf56a		191235-191405		Reverse	rpl2		

	orf549		150358-152007	Reverse		orf25	M8
	orf75		150995-151222	Direct		atp4,nad4L	M8
	orf138		151562-151978	Direct		orf25	M8
	orf65		152520-152717	Reverse		orf25	M8
	orf149		152710-153159	Reverse		orf25	M8
	orf344		153165-154199	Reverse		orf25	M8
	orf53a		153554-153715	Reverse		orf25	M8
	orf62a		153635-153823	Direct		orf25	M8
	orf55c		154245-154412	Direct		orf25	M8
	orf217		154329-154982	Reverse		orf25	M8
	orf66		154770-154970	Direct		orf25	M8
	orf82		229742-229990	Direct		nad5	M10
	orf70b		230121-230333	Reverse		nad5	M10
	orf54a		230284-230448	Reverse		nad5	M10
	orf50c		230361-230513	Direct		nad5	M10
	orf72		231096-231314	Direct		nad5	M10
orf99	orf99	232918-233217	232918-233217	Direct	nad2	nad5	M10
orf60a		236630-236812		Reverse	rps7		M10
orf53a		358004-358165		Direct	nad4		
orf66	orf66a	385854-386054	385863-386063	Reverse	rps4,nad6,rpl5, rps14	nad2,nad6,rps4, nad2	

orf59a	orf59a	393067-393246	393076-393255	Reverse	rps4,nad6,rpl5, rps14	rpl5,rps14	
orf52b	orf52b	395175-395333	395184-395342	Reverse	atp1	rps3	M18
	orf79		396975-397214	Reverse		rps3	M18
	orf71		397717-397932	Reverse		rps3	M18
orf89	orf89	423934-424203	423943-424212	Reverse	nad1	nad1	M19
orf88a	orf88a	424130-424396	424139-424405	Reverse	nad1	nad1	M19
orf54a	orf54b	424375-424539	424384-424548	Reverse	nad1	nad1	M19
orf95	orf95	424773-425060	424782-425069	Reverse	nad1	nad1	M19
orf75	orf75a	424837-425064	424846-425073	Direct	nad1	nad1	M19
orf71a	orf71a	424874-425089	424883-425098	Direct	nad1	nad1	M19
orf70b	orf70c	425346-425558	425355-425567	Direct	nad1	nad1	M19
orf53b	orf53b	426106-426267	426115-426276	Direct	nad1	nad1	M19
orf255	orf255	426373-427140	426382-427149	Reverse	nad1	nad1	M19
orf105	orf105	426380-426697	426389-426706	Direct	nad1	nad1	M19
orf91	orf91	426780-427055	426789-427064	Direct	nad1	nad1	M19
orf51a	orf51a	427625-427780	427634-427789	Reverse	nad1	nad1	M19
orf86a	orf86a	428029-428289	428038-428298	Direct	nad1	nad1	M19
orf128		471465-471851		Direct	rps1		M20
orf205	orf205	472020-472637	472029-472646	Direct	rps1	rps1	M20
orf86b		472031-472291		Reverse	rps1		M20

 orf59	orf59	472740-472919	472749-472928	Reverse	cox2,rps1	rps1	M20
orf55	orf55	472858-473025	472867-473034	Direct	rps1	rps1	M20
orf88	orf88	473213-473479	473222-473488	Reverse	cox2,rps1	rps1	
orf51b	orf51b	482430-482585	482439-482594	Reverse	nad9	cox2	
orf68a	orf68a	498542-498748	498551-498757	Reverse	nad9	cox2	
orf78a	orf78a	504331-504567	504340-504576	Reverse	nad9	cox2	
orf112	orf112	506967-507305	506976-507314	Reverse	cox2	cox2	
orf99a	orf99a	510109-510408	510118-510417	Reverse	cox2	cox2	
	orf82a		532272-532520	Direct		nad5	M21
	019020		552272 552526	Direct		naus	(G1+HBP)
	orf70d		532651-532863	Reverse		nad5	M21 (G1+HBP)
	<i>(</i> <b>-</b> <i>i</i>		522014 522050	D		15	M21
	orf54c		532814-532978	Reverse		nad5	(G1+HBP)
	orf50d		532891-533043	Direct		nad5	M21
	-						(GI+HBP) M21
	orf72a		533626-533844	Direct		nad5	(G1+HBP)
	~~~00h		525449 525747	Diment			M21
	<i>or</i> J99b		333448-333/4/	Direct		ndas	(G1+HBP)

### 附录 VI 扩增基因序列

#### >orf55

>orf59

>orf88

ATGTTAGTTATGACTAGTCTCAAAATAGACTTTTTTCTTTGGTTACCTTGTAT AGGGGGGGCAACCACTATGACTATTAAAGTCAGGAGTTCAAGCCTGCGCTC TAGCAGTCGTCAAAGCACAAGCCCAGCTGGATTGCTAACTACTGTCCTTTT TGGAGAGAGGAAAGGGGCAAAGGCATTGGTCGGAAAAGTCTTCTTTCAAT CAGTCCTAGGCGGGTCTCTTTTGAGAGTTCCATTGGTAGATTGTCAACAGC TTGAGAGCTAA

>DN16978

TTGAATTTAAGGATGCCTCTCATCAATGTCATCCATTATCTTTTCTGTACCAT CTTGGATGACATTGCCACAATGAAGGTAGAATGTATTTGGCTAAGGATGCCC AGAGACAAATCAGGATGGGGGCTTCGATGGGATTCTTCCATCTAAGAATCA GAAAAAATGGACTGATTAACCTCTTACTGCTTGAGAACCTATCAGGTGAG ATGGATTGACCAACATTAACAGGCCTACTATATAAGTATGAAGCGATACATC GGTAAACACTGCTATAGTAGCCCAGCCTAGAGAAAGTTATGGGACAAAACA TGACAAGACTCTTTCTATTTATTCGTCCTTATCAGAAATCTGCTACAGAAAG ATTCCGCTTAAATTTAAAGCAGTTAATCATTATTATGAATTGGAAGATGAAA TCTCTCTGGTGATAGATGGAGAAAGGTCTCCTTTTGTTTCTATTCAGGAAGA AGATAATTGCTATTCCCAAGGAAGAACGTTAACTCTCGCTATTCTATTTGAAT GTGGTGTAGGTGAAGTGAGGATTTAGATAAATCTAAAAAGACCTCTCCCGG GTTTGAGGGCTTAAATCGTTTTTAGTCTTCAACCCAGTCGAATATAAATGTT GAGACCAACCTTCTTATGTTAGTTATGACTAGTCTCAAAATAGACTTTTTTC TTTGGTTACCTTGTATAGGGGGGGGGCAACCACTATGACTATTAAAGTCAGGAG TTCAAGCCTGCGCTCTAGCAGTCGTCAAAGCACAAGCCCAGCTGGATTGCT AACTACTGTCCTTTTTGGAGAGAGGGAAAGGGGCAAAGGCATTGGTCGGAA AAGTCTTCTTTCAATCAGTCCTAGGCGGGTCTCTTTTGAGAGTTCCATTGGT AGATTGTCAACAGCTTGAGAGCTAAAAAAAGTAGTTCTTGCTCTTCCGAT AGGGAAGACGTAATCTGGCTTGACCATTACGGAATTTCCTTTGATGACTGA CTACAGATGCCTGCCAACATCTCGCGATAGATTTAGCAAACTAAAGCGTTA GCACTTTACTAATAAAATATATAGGGCTTTTCCCTATCTTACTAATAATAAGG GGGCTGCTAGTTGTAGCGCGCTAACGCCCCTATAGAGTAAGGCTCTTCTTCT

#### 致谢

盛夏时节,懵懂无知,满怀憧憬,进入柑橘课题组,开始研究生生涯,探求 未知领域;凛冬将至,受益匪浅,念念不舍,离开柑橘课题组,奔赴新的工作岗 位,开启新的篇章。回首过去七年多的研究生时光,无论是成功的喜悦,还是失 败后的迷茫,身边总会有老师、同学、朋友和亲人们的鼓励、支持和帮助。值此 论文完成之际,向他们致以最诚挚的感谢。

首先, 衷心感谢我的研究生导师郭文武教授! 自研究生入学以来, 郭老师无 论在生活还是研究上都给予我严格的要求和悉心的指导。在我实验遇到瓶颈、思 想上开始松懈的时候, 郭老师为我指点迷津的同时还鞭策我努力向前; 在我文章 写作语言逻辑、学术逻辑与规范问题频出时, 郭老师以其严谨细致的学术思维, 教我如何写出科学规范的学术论文, 让我的学术习惯和学术素养得到质的提升。 郭老师出差开会还心系我的课题, 为我的研究寻求合作交流, 他这份对科研执着 和热爱的精神深深感染了我, 激励我不断向前。郭老师除了在学术上给予我指导, 还对我的为人处事等方面进行谆谆教导, 使我受益匪浅。

特别感谢伍小萌老师!伍老师在实验设计和论文写作修改上给予指导和帮助。 在我课题研究遇到困难时,在我写出的学术论文问题不断时,伍老师倾注了大量 时间和心血,一点一滴对我实验进行指导,帮我梳理文章逻辑和修改语言,给予 我极大的帮助。伍老师严谨的治学态度、兢兢业业的工作精神是我学习的榜样! 感谢科研助理尹朝平师兄,在生物信息学方面给予的帮助。线粒体基因组组装拼 接的那段时间,跟他一起学习、讨论组装拼接软件的选择和使用,收获颇大,为 博士阶段研究奠定了基础。

真诚地感谢实验室邓秀新院士、程运江、徐娟、徐强、王鹏蔚、叶俊丽、柴 利军、曾云流、解凯东等老师对我课题研究的热心指导。感谢 Larkin 教授在论 文写作上给予的指导和建议。感谢蔬菜系邓颖天副教授、余长春博士在拟南芥转 基因表型观察及雄性不育课题研究的困惑给予的解答及建议。感谢教育部重点实 验室实验平台屈晓璐老师和石春梅师姐在激光共聚焦显微镜使用提供的帮助。感 谢国家柑橘育种中心谢宗周老师在材料采集和管理上提供的帮助。

感谢郑蓓蓓和方燕妮师姐在我初入实验室时,对我的指导和包容,教会我基因克隆、组织培养、酵母双杂及遗传转化等实验技术,为我独立开展课题研究奠

定基础。感谢雄性不育小组成员,姜楠、程来超、王蓉、李超超、樊燕杰在课题 研究内容提供的无私帮助;感谢黄跃和杨宏宾师弟在生物信息分析方面的帮助; 感谢宫金礼师妹和李柽钖师弟在烟草瞬时表达实验上提供了很多的建议和帮助; 感谢生科院的吴增祥在线粒体提取中提供的帮助;感谢中国科学院昆明植物研究 所朱安丹研究员、华南植物园王帅斌博士在线粒体基因组组装方面的帮助;感谢 马金香师傅在 MT 培养基配制及愈伤组织材料保存过程中提供的帮助。感谢实验 室许让伟、王萍、王淑明、杨晓明、何义仲、龙健梅、谭丰全、郑雄杰等师兄师 姐在课题研究方面给予的指导和帮助;感谢夏强明同学在我家庭遭遇挫折后给予 的开导、关心和照顾;感谢贾慧慧、王蓉、朱凯杰、梁梅、柯丽丽、徐远涛、朱 晨桥、张海鹏等同级好友在科研和生活中相互关心和照顾;感谢汤小美、宋鑫、 冯梦琦、刘丹、杨雯惠、张苗、高虎、马雨婷、任杰、谢善鹏、卿美、和晶晶、 朱虹娴、王鹏博等师弟师妹在实验中给予的热心帮助。

感谢亲爱的母亲多年来的理解与关怀。特别是在家庭发生挫折后,我的母亲 一直默默付出,从未给我压力,并给予支持与鼓励,让我得以顺利完成学业。

感谢华中农业大学园艺植物生物学教育部重点实验室、柑橘细胞工程与改良 遗传课题组为本课题研究提供优秀的实验平台、齐全的科研设备和干净、舒适的 科研环境!本研究得到国家重点研发计划项目(2018YFD1000106)、国家自然 科学基金(31530065,31820103011)、中央高校自主科技创新基金(2662018PY007) 等项目资助。

> 张 帅 2020 年 12 月 于武昌狮子山

121