

## 高等植物线粒体基因组研究进展

张 晓, 张 锐, 侯思宇, 史 计, 郭三堆

(中国农业科学院生物技术研究所, 国家农作物基因资源与遗传改良重大科学工程, 北京 100081)

**摘 要:**高等植物线粒体基因组在目前已知生物中是最大、最复杂的, 阐述了高等植物线粒体基因组的最新研究进展。从全基因组的测序方法开始, 重点介绍了高等植物线粒体基因组的组成和结构特点, 特别是基因含量、内含子、基因间隔序列、RNA 编辑、重复序列、叶绿体和细胞核与线粒体之间的 DNA 交换以及线粒体基因的遗传起源和进化等热点问题, 以期为更好的了解核质互作、植物进化等科学问题奠定基础。

**关键词:**高等植物; 线粒体基因组; 重复序列; RNA 编辑; 异质性

doi:10.3969/j.issn.1008-0864.2011.04.04

中图分类号:Q75 文献标识码:A 文章编号:1008-0864(2011)04-0023-09

## Research Progress on Mitochondrial Genome of Higher Plant

ZHANG Xiao, ZHANG Rui, HOU Si-yu, SHI Ji, GUO San-dui

(Biotechnology Research Institute, National Key Facility of Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

**Abstract:** The mitochondrial genomes of higher plant are the largest and the most complex among reported species so far. This review summarized the recent research progresses on higher plant mitochondrial genome. Based on the newly sequenced whole mitochondrial genomes of 10 species of higher plants, this paper mainly focused on sequencing method, gene content and organization, intron content and splicing mode, variation of intergenic regions, RNA editing, repeat sequence, promiscuous DNA and horizontal transfer. Moreover, the genetic origin and evolution of mitochondrial DNA in higher plant were also discussed. Those were expected to provide basis for better understanding about the interaction of nuclear and cytoplasm and plant evolution, etc.

**Key words:** higher plant; mitochondrial genome; repeat sequence; RNA editing; hetemplasmy

真核生物细胞内有两种类型的遗传系统: 起主导作用的核基因组和半自主性的核外基因组(细胞器基因组)。在高等植物中, 细胞器基因组包括叶绿体基因组和线粒体基因组。目前研究发现高等植物线粒体基因组是所有生物中最大、最复杂的, 并且植物细胞质雄性不育(CMS)这种重要的性状与线粒体基因组的结构变化联系密切, 因此研究高等植物线粒体基因组可以使人类更好地了解核质互作、植物进化等有意义的科学问题, 而获得线粒体基因组的全序列对这些研究非常重要。本文就近年来高等植物线粒体基因组的测序方法、组织结构、基因组成、RNA 编辑等方面的研究进展作一评述。

### 1 全测序的高等植物

#### 1.1 基因组特征

对于植物线粒体中是否真的存在一个基因组“主环”目前仍有争议, 但通过基因组步移和鸟枪法测序(shotgun)策略所获得的一个大环状 DNA 分子的全部核酸序列仍然可以代表一种植物线粒体的全部遗传信息。1997 年, 拟南芥(*Arabidopsis*)线粒体基因组第一个完成全测序<sup>[1]</sup>, 截止到 2010 年 4 月 1 日, 共报道有 20 种高等植物的线粒体基因组完成全测序, 其中包括 8 种双子叶植物<sup>[1~7]</sup>、11 种单子叶植物<sup>[8~12]</sup> 和 1 种裸子植

收稿日期:2011-01-09; 接受日期:2011-03-30

基金项目: 国家转基因生物新品种培育重大专项(2008ZX08005-004)资助。

作者简介: 张 晓, 博士研究生, 研究方向为植物基因工程。Tel:010-82106128; E-mail:energy@126.com。通讯作者: 郭三堆, 研究员, 主要从事棉花基因工程研究。Tel:010-82106140; E-mail:gsdui@mail.caas.net.cn

物<sup>[13]</sup>。部分完成测序的植物线粒体基因组特征见表 1。

这些完成全测序的高等植物线粒体基因组从大小来看:最小的是油菜 (*Brassica napus* L.) (221 853 bp),最大的是西葫芦 (*Cucurbita pepo*) (982 833 bp),同样是高等植物,线粒体基因组大小却相差 4 倍;拟南芥和油菜这两种十字花科植物的线粒体基因组偏小。小麦、玉米和水稻这三种单子叶植物线粒体基因组的大小顺序是:玉米 > 水稻 > 小麦;籼稻比粳稻大 995 bp;野生玉米和野生水稻的线粒体基因组都比各自的栽培种要大。不同植物线粒体基因组虽然大小相差悬殊,但基因数目(52 ~ 60)和内含子数目(20 ~ 24)差别不大。在整个线粒体基因组中,编码序列所占比例为 3.9% (西葫芦) - 17.3% (油菜);内含子所占比例为 3.9% (葡萄) - 12.9% (油菜);叶绿体序列所占比例为 1.1% (拟南芥) - 11.5% (西

葫芦);核序列所占比例为 0.1% (玉米) - 13.4% (水稻);重复序列所占比例为 6.8% (拟南芥) - 58.3% (水稻)。

## 1.2 线粒体基因组测序方法

目前已完成线粒体基因组测序的拟南芥、甜菜、水稻、油菜、玉米、小麦、烟草、葡萄、西瓜和西葫芦,这 10 种高等植物的测序方法各不相同,但也有共性。

**1.2.1 植物材料的选取** 10 种植物中水稻、玉米、小麦、西瓜和西葫芦选用了黄化苗,甜菜选取了主根,其余 4 种选取了绿叶组织。这说明不必为了排除叶绿体基因组的干扰而硬性选择黄化材料,绿色组织也可以。

**1.2.2 mtDNA 纯化方法** 高纯度的 mtDNA 对文库的构建及全测序的完成至关重要。其中,CsCl 梯度法是最常用、最可靠的方法,表 1 所示

表 1 完成测序的部分高等植物线粒体基因组特征

Table 1 Key features of some sequenced higher plant mitochondrial genomes.

植物种类 Plant species	线粒体基 因组大小(bp) Size of mitochondrial genomes(bp)	基因 数目 <sup>a</sup> Genes number <sup>a</sup>	编码序列 比例(%) Coding region ratio(%)	顺式内 含子数 No. of cis- splicing introns	反式内 含子数 No. of trans- splicing introns	内含 子序列 比例(%) Intron ratio(%)	叶绿体 DNA 序列比例(%) Chloroplast DNA sequence ratio(%)	核序列 比例(%) Nuclear DNA sequence ratio(%)	重复序列 比例(%) Repeat sequence ratio(%)
拟南芥 <sup>[1]</sup> <i>Arabidopsis</i>	366 924	55	10.6	18	5	8.0	1.1	4.0	6.8
甜菜 <sup>[2]</sup> Sugar beet	368 801	52	10.3	14	6	5.6	2.1	3.3	8.8
烟草 <sup>[3]</sup> Tobacco	430 597	60	9.9	17	6	6.5	2.5	ND	8.0
油菜 <sup>[4]</sup> Rapeseed	221 853	53	17.3	19	5	12.9	3.6	ND	ND
小麦 <sup>[8]</sup> Wheat	452 528	55	9.5	16	6	6.4	5.8	0.2	15.2
玉米 <sup>[9]</sup> Maize	569 630	52	7.4	15	7	4.3	4.4	0.1	23.0
水稻 <sup>[10]</sup> Rice	490 520	56	11.1	17	6	7.2	6.2	13.4	58.3
葡萄 <sup>[5]</sup> Grape	773 279	53	5.9	ND	ND	3.9	8.8	ND	6.8
西瓜 <sup>[6]</sup> Watermelon	379 236	58	10.3	19	5	8.6	6.0	6.4	10.0
西葫芦 <sup>[6]</sup> Zucchini	982 833	53	3.9	19	5	3.1	11.5	2.1	37.7

<sup>a</sup>不考虑拷贝数; ND:无具体数据。

<sup>a</sup>Not considering the copy number; ND: no data.

的 mtDNA 基本是此方法制备的。

**1.2.3 mtDNA 的片段化方法** 在以上 10 种植物中,其中 6 种选用酶切法(*Sau3A I*、*BamH I* 等限制性内切酶)进行 mtDNA 的片段化,另外 4 种选用物理剪切法(HydroShear DNA 剪切仪)。2008 年以后所完成的几项测序工作均使用物理剪切法。

**1.2.4 载体的选择** 早期测序多选用 Cosmid、Fosmid 等载体结合鸟枪法(shotgun)进行测序,但近年来测序工作中载体有简单化的趋势,如西瓜和西葫芦只选用 pUC18 这一简单的克隆载体。

**1.2.5 插入片段的大小** 早期测序主要是大插入片段结合小插入片段文库进行测序,近几年是直接建立 2~3 kb 插入片段文库,简化了测序步骤。

**1.2.6 测序覆盖倍数** 在 10 种植物中,测序覆盖倍数从 4 倍(甜菜)到 60 倍(葡萄)不等,多数在 10 倍左右。

## 2 植物线粒体基因组上的基因

高等植物线粒体基因组主要包括复合体 I 基因(*nad1*、*nad2*、*nad3*、*nad4*、*nad4L*、*nad5*、*nad6*、*nad7* 和 *nad9*)、复合体 II 基因(*sdh3* 和 *sdh4*)、复合体 III 基因(*cob*)、复合体 IV 基因(*cox1*、*cox2* 和 *cox3*)、复合体 V 基因(*atp1*、*atp4*、*atp6*、*atp8* 和 *atp9*)、Cytochrome c 生物合成基因(*ccmB*、*ccmC*、*ccmFC* 和 *ccmFN*)、核糖体蛋白基因(*rps1*、*rps2A*、*rps2B*、*rps3*、*rps4*、*rps7*、*rps10*、*rps12*、*rps13*、*rps14*、*rps19*、*rpl2*、*rpl5* 和 *rpl16*)、核糖体 RNA 基因(*rrn5*、*rrn18* 和 *rrn26*)、tRNA 基因(*trnN*、*trnD*、*trnC*、*trnE*、*trnQ*、*trnH*、*trnI*、*trnK*、*trnM*、*trnJ*、*trnF*、*trnP*、*trnS*、*trnW* 和 *trnY*)以及 *matR* 基因(编码类成熟酶)、*mttB* 基因(编码转运子)等。复合体 I、III、IV 和 V 中的基因、核糖体 RNA 基因、蛋白转位子和内含子成熟酶基因在线粒体基因组之间是保守的<sup>[1~13]</sup>,而其他种类的基因序列保守,但数目是变化的。

表 1 的 10 种植物线粒体基因组中,都缺乏一套完整的、足够阅读所有密码子的 tRNA 基因。它们普遍缺少 Ala、Arg、Leu、Thr、Val 所对应的

tRNA。有的植物缺少的更多,比如在拟南芥中,除了上述 5 种外,Phe、eMet、Trp 的 tRNA 也是丢失或失活的<sup>[14,15]</sup>。所以,相应的 tRNA 分子需要从细胞核或胞质中输入。

被子植物中所有的线粒体基因使用通用遗传密码,并且第三密码子倾向于 A 或 T<sup>[16]</sup>。典型的翻译起始密码是 ATG,但是在 *rpl16*<sup>[4]</sup>、*mttB*<sup>[17]</sup>、*matR*<sup>[18]</sup> 基因中,可能有另外的起始密码子。

## 3 高等植物线粒体基因组特点

### 3.1 结构复杂

线粒体基因组中包括几种不同类型(环状、线性、大环、小环)的 DNA 分子,通过染色体步移方法获得的图谱非常复杂。小麦<sup>[8]</sup>和玉米<sup>[9]</sup>的线粒体基因组是由主环分子和多个亚环分子组成。亚环分子通常是由大同向重复(large direct repeat)序列介导的重组产生的。小麦 430 kb 的线粒体基因组主环上至少有 10 对重复序列;玉米主环分子含有一组特有的重复序列,重复序列间的同源重组导致出现不同大小的多个亚环分子。拟南芥线粒体基因组中含有的两个重复序列(6.5 kb 和 4.2 kb)参与了重组,导致产生两个同向的大小分别是 233 kb 和 134 kb 的亚环分子<sup>[1]</sup>。油菜线粒体基因组中存在一段 2 427 bp 的重复序列,该序列介导的重组产生了 125 kb 和 97 kb 两个亚环分子<sup>[4]</sup>。

除了同向重复序列,一些线粒体基因组也含有反向重复序列。这种情况下,不会产生小亚环,但产生了等尺寸的环(与母环不同,含有片段插入)<sup>[19]</sup>。

由于存在这些亚环和等尺寸的环,高等植物线粒体基因组被认为含有多种结构分子。需要指出的是,这些不同结构分子不是异质的,因为等尺寸环和亚环是主环的衍生形式。高等植物线粒体的异质性是由 sublimons 引起的,它们以低的化学计量学形式存在,与主环或任何其他亚环间没有共线性关系<sup>[19]</sup>。

对十字花科的油菜<sup>[4]</sup>和拟南芥<sup>[1]</sup>这两种植物的线粒体基因组进行比较分析发现,油菜线粒体 1/3 的基因组序列和拟南芥线粒体 2/3 的基因

组序列没有同源关系。这说明物种分化之后,两种植物的线粒体基因组发生了修饰、序列插入或丢失,这些序列达到了拟南芥整个线粒体基因组的2/3。在这个水平上,序列转移、插入或丢失被认为是进化中比基因组内部重排更重要的因子。这些非同源的序列中有的起源于叶绿体或细胞核基因组,但大多数序列与 GenBank 数据库中的序列都没有同源性。另外,油菜、拟南芥和甜菜这三种双子叶植物线粒体基因组序列之间的共有序列只存在于编码序列和内含子中<sup>[1,2,4]</sup>。这说明,植物线粒体基因组中基因间隔序列高度变化,流动性很强。

植物叶绿体基因组中基因排列顺序比较保守,例如水稻<sup>[20]</sup>、玉米<sup>[21]</sup>和小麦<sup>[22]</sup>叶绿体基因组中基因排列顺序完全相同,但植物线粒体基因组中基因排列顺序差别却很大,这也间接说明了植物线粒体基因组的复杂性。

### 3.2 基因密度低

高等植物线粒体基因组最明显的特征是:相对于全基因组的大小来说,它含有的基因太少。高等植物线粒体基因组是基因密度最小的线粒体基因组,其基因编码区只占整个线粒体基因组的10%左右。西葫芦的线粒体基因组密度最低,只有3.9%是基因编码序列<sup>[6]</sup>。这一比例相比植物叶绿体基因组来说也很低,在水稻和小麦叶绿体基因组中,编码序列分别占整个基因组序列的58.8%<sup>[20]</sup>和60.4%<sup>[22]</sup>,而在线粒体基因组中这个比例只有18.0%<sup>[10]</sup>和15.9%<sup>[8]</sup>,进一步说明线粒体基因组中基因密度很低。

从表1可以看出,甜菜的编码序列最短<sup>[2]</sup>,34.5 kb,小麦的最大<sup>[8]</sup>,43.1 kb,它们只相差8.6 kb,这个差别主要是由于某些基因的拷贝数不同,例如,小麦的线粒体基因组中 *rrn18* 有三个拷贝(每个拷贝1 955 bp), *rrn26* 有两个拷贝(每个拷贝3 467 bp), *atp6* 有两个拷贝(每个拷贝1 161 bp)。内含子最短的是甜菜<sup>[2]</sup>,18.7 kb,最长的是西瓜<sup>[6]</sup>,32.5 kb。基因数目相对保守,在52~60个之间,该数字的差别主要是由于 tRNA 基因的数目不同<sup>[1~10]</sup>。植物线粒体基因组 DNA 的 A + T 含量较高,这也与其基因密度低相符<sup>[23]</sup>。

尽管高等植物线粒体基因组大小范围很广

(已测序的基因组大小范围是222~983 kb),但基因数目(52~60)没有明显变化。高等植物线粒体基因组含有的基因数目比人类<sup>[24]</sup>(16.6 kb,37个基因)多,但比地钱<sup>[25]</sup>(184 kb,66个基因)少。裸子植物苏铁<sup>[13]</sup>线粒体基因组415 kb,含64个基因,比被子植物的稍多。对陆地植物的分析说明,一些基因从高等植物中丢失,但基因组大小却没有相应变小。丢失的基因有些似乎是转移到了细胞核中,然后其翻译产物转运到线粒体中;或者线粒体的基因丢失通过细胞核中的同种异型基因进行补偿。例如,拟南芥线粒体中不存在 *rps13*,在细胞核中也没有迁移的拷贝,但是,细胞核中有一个来自于叶绿体的 *rps13* 拷贝,它编码一个线粒体 RPS13 多肽。tRNA<sup>Leu</sup>和 tRNA<sup>Arg</sup>在裸子植物苏铁<sup>[13]</sup>线粒体中存在,但在被子植物中,这两个 tRNA 分子丢失,需要从胞质中输入。需要指出的是高等植物线粒体基因组中有一些基因还没有被发现和鉴定。例如, *rpl10* 近来被发现存在于烟草和葡萄的线粒体基因组中,但在拟南芥、小麦、玉米和甜菜中并不存在<sup>[26]</sup>。

在已测序的高等植物线粒体基因组中,基因编码序列和内含子的核酸序列总长范围是:53~72 kb;这个数值与整个基因组大小相比是相当稳定的。这说明其余占整个线粒体基因组69%~93%的基因间隔区,是导致高等植物线粒体基因组大小差异的主要区域。

### 3.3 含有来自于叶绿体和细胞核的序列

高等植物线粒体基因组含有与叶绿体和细胞核 DNA 同源的序列。

在高等植物线粒体基因组中,普遍存在着来自于叶绿体的基因片段,并且占有较高的比例(表1)。但中共有的序列只有转运 Asn、eMet 和 Trp 的 tRNAs 序列。葡萄线粒体基因组中含有30个来自于叶绿体的片段,共68 237 bp,占整个线粒体基因组的8.8%,占其整个叶绿体基因组的42.4%,在所有已测序的植物线粒体基因组中,这个数值是最大的<sup>[5]</sup>。这些片段中只有9个与其他植物线粒体基因组有同源性,其余都是葡萄线粒体基因组所特有的。水稻线粒体基因组中有17段叶绿体来源的 DNA 片段,大小32~6 653 bp,相似性有61%~100%,共22 593 bp,占整个

线粒体基因组的 6.3%。这 22 593 bp 中,有 1 140 处碱基替换、45 处缺失和 23 处插入,由于序列的多处改变和 RNA 编辑的缺失,导致其中的蛋白质编码序列往往是无功能的<sup>[10]</sup>。但叶绿体来源的 7 个 tRNA 基因很可能有功能,因为有 6 种氨基酸的 tRNA 在线粒体中是不存在的,而来自叶绿体的 tRNA 基因很可能弥补了这一缺陷。在低等植物地钱的线粒体基因组中,有 27 种 tRNA,但其中没有叶绿体来源的,这说明叶绿体基因序列向线粒体转移这一现象很可能是开花植物所特有的,在开花植物进化过程中,线粒体与细胞核之间、线粒体与叶绿体之间很可能存在频繁的 tRNA 基因转移<sup>[25]</sup>。

植物线粒体基因组中,也存在着来自于细胞核的序列,这些序列主要是与反转座子序列同源,并且在线粒体基因组中占的比例较大(表 1)。在甜菜中,最短的序列是 25 bp,最长的序列是 2 827 bp,它们与 *Gypsy* 型反转座子、(R)-mandelonitrile lyase 基因、*At5g13390*(也称作 *nef1*)等基因有同源性<sup>[2]</sup>。在水稻中,来源于细胞核的序列与 *rps10*、*rps11* 和 *rps14* 等基因有同源性,并且包含 6 个 tRNA 基因序列<sup>[10]</sup>。在这些序列中,没有序列在任意两种植物间是共有的,这说明它们是独立整合的。叶绿体 DNA 向线粒体基因组中迁移是单向的,但细胞核 DNA 与线粒体 DNA 的迁移是双向的<sup>[10,27]</sup>。

高等植物核基因组、线粒体基因组和叶绿体基因组之间存在着广泛的交流。例如,在水稻中,有 6 段序列(I~VI)在细胞核、线粒体、叶绿体基因组中都存在,所包含的基因及序列长度分别是 *rpl2*(序列 I, 1 207 bp), *rpl2*(序列 II, 1 468 bp), *rpoB* 和 *rpoC1*(序列 III, 901 bp), *ndhK*(序列 IV, 229 bp), *atpH*(序列 V, 142 bp) 和 *rpl14*(序列 VI, 216 bp)。这 6 段序列很可能是来源于叶绿体,其中 I、II、III 片段是直接独立地从叶绿体转移到了其他基因组,而 IV、V、VI 片段是先从叶绿体转移到线粒体,再从线粒体转移到细胞核<sup>[10]</sup>。

### 3.4 存在 C-U RNA 编辑

1989 年,来自加拿大<sup>[28]</sup>、法国<sup>[29]</sup>、德国<sup>[30]</sup>的三个实验室几乎同时独立报道了植物线粒体 RNA 中 C-U 编辑这一现象。后来的研究发现,

RNA 编辑在高等植物线粒体中广泛存在,它是植物线粒体基因组中基因表达的必需步骤之一<sup>[31]</sup>。RNA 编辑属于转录后修饰范畴,其化学本质是脱氨基反应,在这个过程中,一个特异位点的胞嘧啶(C)被改变成尿嘧啶(U)。这种 C-U RNA 编辑事件倾向于出现在第二个密码子位置,并且大多数都是完全编辑,通过编辑,提高了不同物种之间线粒体蛋白质序列的同源性<sup>[32]</sup>。RNA 编辑位点中 92% 会改变氨基酸序列,通常是将亲水性氨基酸转变成疏水性氨基酸,使蛋白质能更好地折叠并发挥功能<sup>[33]</sup>。RNA 编辑还可以产生在基因组序列中所不存在的起始密码子和终止密码子,而通常产生新的起始密码子和终止密码子后,其编码的蛋白质更加保守,与其他物种相应蛋白质的同源性更高,从而使线粒体中的基因可以更好地进行表达。例如植物线粒体 *atp9* 基因的终止密码子,在 60% 的植物中都是通过 RNA 编辑产生的<sup>[34]</sup>。

高等植物线粒体基因组中几乎每一个编码蛋白质的基因中都存在 RNA 编辑位点。在 36 个拟南芥基因中检测到了 441 个 RNA 编辑位点<sup>[33]</sup>; 34 个油菜基因中检测到了 427 个<sup>[4]</sup>; 34 个水稻基因中检测到了 491 个<sup>[10]</sup>。将拟南芥和油菜这两种十字花科植物的编辑位点进行比较发现,油菜的 427 个编辑位点中 84% 与拟南芥相同,另外的 16% (69 个)是物种特异的<sup>[4]</sup>。

不同基因 RNA 编辑位点的数目相差很大。一般来说核糖体蛋白基因 RNA 编辑位点较少,而 *mttB*, *ccmB*, *ccmFN* 等基因较多,编辑位点最多的基因是 *ccmC*。水稻 *ccmC* 基因有 36 个位点,而 *rps12* 基因中没有编辑位点<sup>[10]</sup>。另外核糖体 RNA 基因(*rrn26*、*rrn18* 和 *rrn5*)中通常没有 RNA 编辑位点<sup>[10]</sup>。RNA 编辑位点不仅出现在基因编码区,也可能出现在基因间隔区或 UTR 区,比如拟南芥线粒体不仅在编码区存在 441 个 RNA 编辑位点,在非编码区也发现了 15 个,但非编码区的编辑位点通常是部分编辑<sup>[26]</sup>。

同一基因在不同植物中的编辑效率也相差很大。例如, *ccmFc*、*cob*、*matR* 和 *mttB* 在西葫芦<sup>[6]</sup>中完全编辑,在西瓜<sup>[6]</sup>中却是部分编辑,特别是 *mttB*, 在西葫芦中 20 个非同义编辑位点中有 18

个完全编辑,但在西瓜中,18个非同义编辑位点中只有3个完全编辑。反过来,*nad9*和*rps4*基因中的非同义编辑位点在西瓜中完全编辑,在西葫芦中部分编辑。

拟南芥和油菜这两种十字花科植物的编辑位点的一致性为84%,与二者基因编码序列的平均一致性高达99%相比,细胞核基因在不同物种RNA编辑差异中发挥了重要的调控作用<sup>[4]</sup>。拟南芥不同生态型之间线粒体*ccmB*基因RNA编辑存在差异也反映了这个问题<sup>[35]</sup>。研究表明,在细胞核中,存在一个细胞器靶向的RNA-结合蛋白家族,该家族是陆地植物所特有,具有35个氨基酸串联重复序列,称为PPR蛋白<sup>[36]</sup>。遗传与生化研究说明,PPR蛋白在植物细胞器RNA代谢加工的很广泛的范围中起作用,不仅包括编辑,还包括RNA的转录剪接、稳定和翻译<sup>[37]</sup>。一种假说认为,PPR蛋白直接与目标转录本上的一个特异位点互作,同时,PPR蛋白负责征募RNA编辑所需的酶类。第一个特异编辑因子是从陆地植物叶绿体和线粒体中鉴定出来的。研究证明这些特异因子是一类PPR蛋白;并且它们都是一个特异家族-PLS家族的成员,PLS具有特异的P(35氨基酸),L(长,通常是36氨基酸),S(短,通常是31氨基酸)模体的三联子重复,和一个扩展的C末端结构域<sup>[38]</sup>。

### 3.5 含有II型内含子和I型内含子

高等植物线粒体基因组中,*cox1*<sup>[9]</sup>、*cox2*、*nad1*、*nad2*、*nad4*、*nad5*、*nad7*、*rps3*、*ccmC*、*trnA*<sup>[5]</sup>、*ccmFC*、*rps10*、*rpl2*和*ccb438*<sup>[2]</sup>等基因中存在内含子。根据内含子结构特点、保守序列以及独特的二级结构特征,可以将其分成两类:I型内含子和II型内含子。I型内含子主要分布在原核生物、噬菌体、细胞器和细胞核基因组中;II型内含子存在于真菌和植物的线粒体基因组、以及植物叶绿体基因组中<sup>[39]</sup>。

在已测序的被子植物线粒体基因组中,内含子的数目是20~24个,除了西瓜*cox1*基因中存在I型内含子(通过水平转移的方式来自于真菌)外,其他都是II型内含子<sup>[6]</sup>。II型内含子已经在十几个线粒体基因中发现。在含有II型内含子的基因中,*nad1*、*nad2*和*nad5*显示反式剪接,

外显子单独转录,然后剪接成成熟转录本<sup>[23]</sup>。

内含子含量和剪接模式的多样性已在*nad1*、*nad4*、*nad7*、*cox2*和*rps3*中发现,这主要包括内含子的丢失、顺式剪接往反式剪接的转换<sup>[40]</sup>。

### 3.6 基因编码序列变化慢,但基因间隔序列变化快

高等植物线粒体基因组中基因编码序列很保守,不同植物间基因序列同源性通常在95%以上<sup>[1~10]</sup>。Tian等<sup>[41]</sup>比较了籼稻和粳稻的线粒体基因组,发现这两个亚种之间共存在96个SNPs,25个indels,即SNPs出现的几率是0.02%,indels出现的几率是0.006%,这两个几率分别比两者叶绿体基因组之间的几率低2.5倍和3倍,比两者核基因组之间的几率低21倍和38倍,比动物线粒体基因组之间的几率低40~100倍。低突变率为线粒体基因组中非编码序列的积累提供了方便,进而方便了整个基因组序列的扩展。“突变压力学说”认为,基因组大小与突变率呈负相关。高等植物线粒体基因组之所以在所有生物中最大,可能要部分归功于它的低突变率<sup>[41]</sup>。

但是,高等植物线粒体基因组的结构变化很快,植物线粒体基因组的重排率比叶绿体基因组和动物线粒体基因组要高很多<sup>[5]</sup>。高等植物线粒体基因组间存在明显的分子内和分子间重组,导致不同植物线粒体基因组结构的差别非常大,甚至不同亚种之间的线粒体基因组结构差别也很大。由于结构变化大,因此不同植物线粒体基因组上的基因排列顺序差别很大,几乎找不到同源性,只能找到一些排列顺序相对保守的基因簇,比如*rrn18-rrn5*、*nad3-rps12*和*rps3-rpl16*<sup>[8]</sup>。

对整个线粒体基因组的序列分析结果显示,在任意两种植物之间,绝大多数基因间隔序列是不保守的,甚至在两种紧密相关的物种,比如拟南芥和油菜之间,或者任意两种禾本科植物之间也是如此。例如,尽管油菜线粒体基因组只有222 kb,但当它与拟南芥进行比较时,其特有序列达78.7 kb,占整个基因组的1/3。油菜特异序列的组成是:13.2%为线粒体起源(这些序列可以在非拟南芥的高等植物线粒体基因组中发现),5.9%为叶绿体来源,0.3%为细胞核起源,其余部分起源未知<sup>[4]</sup>。甜菜细胞质雄性不育系TK81-

MS 中总共有 68 kb(大约占总基因组的 13.6%) 序列是正常甜菜品种 TK81-O 所没有的<sup>[42]</sup>。在这 68 kb 序列中,线粒体起源、核起源、叶绿体起源和线粒体-episome 起源的序列分别占 7.6%、17.9%、0.1% 和 4.6%,而剩下 69.8% 的起源是未知的。这说明,物种特异线粒体序列的获得和丢失在进化层面上是非常快速的过程,并且序列迁移并不是这个过程的主导因素,未知起源序列的产生很可能存在其他的机制。进一步分析推测,大多数物种特异线粒体序列可能来自于已存在的序列,这些序列发生广泛的重排,导致线粒体基因组的“混乱”<sup>[42,43]</sup>。

拟南芥核基因组差不多是陆地植物中最小的,但是其线粒体基因组是中等大小,比同是十字花科的油菜<sup>[4]</sup>大了将近一倍,这说明线粒体基因组是独立进化的。

### 3.7 高等植物线粒体基因组中含有较多的重复序列

高等植物线粒体基因组中分布着大量的重复序列,这些重复序列从几 bp 到几十 kb 不等,占整个线粒体基因组的 6.84% ~ 58.34%<sup>[1~10]</sup>。不同植物线粒体基因组上的重复序列没有同源性,这说明它们在高等植物进化过程中是由每个物种独立获得的。

这些重复序列包括长重复序列和短重复序列。长重复序列又包括同向重复和反向重复,其中同向重复占多数。位于长重复序列上的基因便成为多拷贝基因。比如小麦<sup>[8]</sup>中的 *rrn26-trnQ-trnK*、*rrn5-rrn18-trnfM*、*atp6*、*atp8*、*trnD* 和 *trnP*; 玉米<sup>[9]</sup>中的 *atp1*、*nad1*、*rps3* 和 *nad2*; 拟南芥<sup>[1]</sup>中的 *atp6*; 甜菜<sup>[2]</sup>中的 *rrn26-trnfM*; 油菜<sup>[4]</sup>中的 *cox2*; 西瓜<sup>[6]</sup>中的 *sdh3*、*trnQ* 和 *trnG* 等。

除了长重复序列外,高等植物线粒体基因组上还分布着大量的短重复序列,这些序列的长度在几十 bp 至几百 bp。短重复序列对高等植物线粒体基因组的进化非常重要,包括组织结构和基因组总长度的变化。西葫芦是目前所测线粒体基因组中最大的,其原因就是它含有总长达 371 kb 的短重复序列<sup>[6]</sup>。虽然重复序列总长度很大,但单个最大的重复片段只有 621 bp,这在几种已测序的植物中是最小的<sup>[1~13]</sup>。这些短重复片段大

多位于基因间隔区域,只有 2% 的重复序列发生在基因编码区和内含子区,同时它们又可分成几十个不同的家族。需要指出的是,西葫芦<sup>[6]</sup>线粒体基因组重复序列中有相当一部分是与叶绿体起源的区域重叠的。

这些重复序列在高等植物线粒体基因组重组中非常活跃。表 1 所示的 10 种植物中,除了葡萄和西葫芦,线粒体基因组可能都通过长重复序列进行了分子内或分子间的重组,导致产生了亚基因组分子或异构体形式。而短重复序列在介导线粒体基因组重排方面非常活跃,它们频繁的加倍、“跳跃”,是导致高等植物线粒体基因组异质性的主要因素之一。

## 4 展望

模式植物和重要农作物线粒体基因组全测序的完成,丰富了高等植物线粒体基因组数据库,拓展了对植物线粒体基因组的认识。如果将已测序的基因组中大小相近、基因结构相似的占多数的几种植物的线粒体基因组作为“典型”的线粒体基因组,会发现在其他植物中存在“非典型”基因组。比如,西瓜(*Citrullus melo* L.)<sup>[43]</sup>具有超大的线粒体基因组;牻牛儿苗科植物(*Geraniaceae*)<sup>[44]</sup>的线粒体基因组与较近物种相比核酸替换率明显高;互叶梅(*Amborella trichopoda*)<sup>[45]</sup>(最原始被子植物)的线粒体基因组序列中存在频繁的水平转移。这些“非典型”的线粒体基因组为研究者提供了研究高等植物线粒体基因组进化机制的线索。这些尚未探明的进化机制包括:线粒体基因组中未知起源的序列是如何出现的;与叶绿体基因组相比为何线粒体的核酸替换率特别低;核酸序列是如何通过平行转移的方式获得的。利用与线粒体基因组结构完整性或基因组分布有关的突变体,比如拟南芥 *chm* 突变体<sup>[46]</sup>和玉米 P2 系统<sup>[47]</sup>可以研究以上问题。研究植物线粒体基因组的目标之一,是对线粒体基因组进行人工修饰或线粒体转化。特别是线粒体转化,目前尚未成功。

随着 DNA 测序技术的发展,将会有越来越多的植物线粒体基因组被测定,这将使研究者更好

地了解植物线粒体基因组及其功能,并有助于研究高等植物核-质互作,而 CMS 等与线粒体基因组突变有关性状的机理也将被揭示得越来越清晰。

### 参 考 文 献

- [1] Unseld M, Marrieffeld J R, Brandt P, *et al.*. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366, 924 nucleotides[J]. *Nat. Genet.*, 1997, 15:57-61.
- [2] Kubo T, Nishizawa S, Sugawara A, *et al.*. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA<sup>Cys</sup> (GCA) [J]. *Nucleic Acids Res.*, 2000, 28:2571-2576.
- [3] Sugiyama Y, Watase Y, Nagase M, *et al.*. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome; comparative analysis of mitochondrial genomes in higher plants[J]. *Mol. Genet. Genomics*, 2005, 272:603-615.
- [4] Handa H. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rape seed (*Brassica napus* L.); comparative analysis of the mitochondrial genomes of rape seed and *Arabidopsis thaliana* [J]. *Nucleic Acids Res.*, 2003, 31:5907-5916.
- [5] Goremykin V V, Salamini F, Velasco R, *et al.*. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer[J]. *Mol. Biol. Evol.*, 2009, 26:99-110.
- [6] Andrew J A, Wei X X, Rice D W, *et al.*. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae) [J]. *Mol. Biol. Evol.*, 2010, 27(6):1436-1448.
- [7] Satoh M, Kubo T, Nishizawa S, *et al.*. The cytoplasmic male-sterile type and normal type mitochondrial genomes of sugar beet share the same complement of genes of known function but differ in the content of expressed ORFs [J]. *Mol. Genet. Genomics*, 2004, 272:247-256.
- [8] Ogihara Y, Yamazaki Y, Murai K, *et al.*. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome [J]. *Nucleic. Acids Res.*, 2005, 33:6235-6250.
- [9] Clifton S W, Minx P, Fauron C M, *et al.*. Sequence and comparative analysis of the maize NB mitochondrial genome [J]. *Plant Physiol.*, 2004, 136:3486-3503.
- [10] Notsu Y, Masood S, Nishikawa T, *et al.*. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome; frequent DNA sequence acquisition and loss during the evolution of flowering plants [J]. *Mol. Genet. Genomics*, 2002, 268:434-445.
- [11] Allen J O, Minx P, Fauron C M, *et al.*. The complete mitochondrial genomes of five close relatives of maize (Unpublished <http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi>); taxid = 33090& opt = organelle).
- [12] Allen O J, Fauron C M, Minx P, *et al.*. Comparison among two fertile and three male sterile mitochondrial genomes of maize [J]. *Genetics*, 2007, 177:1173-1192.
- [13] Chaw S, Shih A C, Wang D, *et al.*. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites [J]. *Mol. Biol. Evol.*, 2008, 25:603-615.
- [14] Duchene A M, Marechal D L. The chloroplast-derived *trnW* and *trnM-e* genes are not expressed in *Arabidopsis mitochondria* [J]. *Biochem. Biophys. Res. Commun.*, 2001, 285:1213-1216.
- [15] Marienfeld J, Unseld M, Brennicke A. The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information [J]. *Trends Plant Sci.*, 1999, 4:495-502.
- [16] Sugiyama Y, Watase Y, Nagase M, *et al.*. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome; comparative analysis of mitochondrial genomes in higher plants [J]. *Mol. Genet. Genomics*, 2005, 272:603-615.
- [17] Sunkel S, Brennicke A, Knoop V. RNA editing of a conserved reading frame in plant mitochondria increases its similarity to two overlapping reading frames in *Escherichia coli* [J]. *Mol. Gen. Genet.*, 1994, 242:65-72.
- [18] Thomson M C, Macfarlane J L, Beagley C T, *et al.*. RNA editing of *mat-r* transcripts in maize and soybean increases similarity of the encoded protein to fungal and bryophyte group II intron maturases; evidence that *mat-r* encodes a functional protein [J]. *Nucleic Acids Res.*, 1994, 22:5745-5752.
- [19] Kitazaki K, Kubo T. Cost of having the largest mitochondrial genome; evolutionary mechanism of plant mitochondrial genome [J]. *J. Bot.*, 2010, doi:10.1155/2010/620137.
- [20] Hiratsuka J, Shimada H, Whittier R, *et al.*. The complete sequence of the rice (*Oryza sativa*) chloroplast genome; Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals [J]. *Mol. Gen. Genet.*, 1989, 217(2-3):185-194.
- [21] Maier R M, Neckermann K, Igloi G L, *et al.*. Complete sequence of the maize chloroplast genome; gene content, hotspots of divergence and fine tuning of genetic information by transcript editing [J]. *J. Mol. Biol.*, 1995, 251:614-628.
- [22] Ogihara Y, Isono K, Kojima T, *et al.*. Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA [J]. *Mol. Gen. Genomics*, 2002, 266:740-746.
- [23] Kubo T, Mikami T. Organization and variation of angiosperm mitochondrial genome [J]. *Physiologia Plantarum*, 2007, 129:6-13.
- [24] Anderson S, Bankier A T, Barrell B G, *et al.*. Sequence and organization of the human mitochondrial genome [J]. *Nature*,



- 1981, 290:457–465.
- [25] Oda K, Yamato K, Ohta E, *et al.*. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome[J]. *J. Mol. Biol.*, 1992, 223:1–7.
- [26] Kubo N, Arimura S. Discovery of the *rpl10* gene in diverse plant mitochondrial genomes and its probable replacement by the nuclear gene for chloroplast RPL10 in two lineages of angiosperms[J]. *DNA Res.*, 2010, 17(1):1–9.
- [27] Stupar R M, Lilly J W, Town C D, *et al.* Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats [J]. *Proc. Natl. Acad. Sci. USA*, 2001, 98:5099–5103.
- [28] Covello P S, Gray M W. RNA editing in plant mitochondria [J]. *Nature*, 1989, 341:662–666
- [29] Gualberto J M, Lamattina L, Bonnard G, *et al.*. RNA editing in wheat mitochondria results in conservation of protein sequences[J]. *Nature*, 1989, 341:660–662.
- [30] Hiesel R, Wissinger B, Schuster W, *et al.*. RNA editing in plant mitochondria[J]. *Science*, 1989, 246:1632–1634.
- [31] Gray M W. Diversity and evolution of mitochondrial RNA editing systems[J]. *IUBMB Life*, 2003, 55:227–233.
- [32] Gray M W. RNA Editing in Plant Mitochondria; 20 Years Later[J]. *IUBMB Life*, 2009, 61(12):1101–1104.
- [33] Giege P, Brennicke A. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs [J]. *Proc. Natl. Acad. Sci. USA*, 1999, 96:15324–15329.
- [34] Handa H. RNA editing of rapeseed mitochondrial *atp9* transcripts: RNA editing changes four amino acids, but termination codon is already encoded by genomic sequence [J]. *Jpn. J. Genet.*, 1993, 68:47–54.
- [35] Bentolila S, Chateigner-Boutin A L, Hanson M R. Ecotype allelic variation in C-to-U editing extent of a mitochondrial transcript identifies RNA-editing quantitative trait loci in *Arabidopsis*[J]. *Plant Physiol.*, 2005, 139:2006–2016.
- [36] Small I D, Peeters N. The PPR motif – a TPR-related motif prevalent in plant organellar proteins [J]. *Trends Biochem. Sci.*, 2000, 25, 46–47.
- [37] Saha D, Prasad A M, Srinivasan R I. Pentatricopeptide repeat proteins and their emerging roles in plants [J]. *Plant Physiol. Biochem.*, 2007, 45, 521–534.
- [38] Rudinger M, Polsakiewicz M, Knoop V. Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts [J]. *Mol. Biol. Evol.* 2008, 25, 1405–1414.
- [39] Ohshima K, Takemura M. Molecular evolution of mitochondrial introns in the liverwort *Marchantia polymorpha* [J]. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.*, 2008, 84:17–23.
- [40] Qiu Y L, Palmer J D. Many independent origins of trans splicing of a plant mitochondrial group II intron [J]. *J. Mol. Evol.*, 2004, 59:80–89.
- [41] Tian X, Zhang J, Hu S, *et al.*. The rice mitochondrial genomes and their variations [J]. *Plant Physiol.*, 2006, 140:401–410.
- [42] Satoh M, Kubo T, Mikami T. The Owen mitochondrial genome in sugar beet (*Beta vulgaris* L.); possible mechanisms of extensive rearrangements and the origin of the mitotype-unique regions [J]. *Theor. Appl. Genet.*, 2006, 113:477–484.
- [43] Lilly J W, Harvey M J. Small, repetitive DNAs contribute significantly to the expanded mitochondrial genome of cucumber [J]. *Genetics*, 2001, 159:317–328.
- [44] Parkinson C L, Mower J P, Qiu Y L, *et al.*. Multiple major increases and decreases in mitochondrial substitution rates in the plant family *Geraniaceae* [J]. *BMC Evol. Biol.*, 2005, 5:73.
- [45] Bergthorsson U, Richardson A O, Young G J, *et al.*. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella* [J]. *Proc. Natl. Acad. Sci. USA*, 2004, 101:17747–17752.
- [46] Abdelnoor R V, Yule R, Elo A, *et al.*. Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to *MutS* [J]. *Proc. Natl. Acad. Sci. USA*, 2003, 100:5968–5973.
- [47] Kuzmin E V, Duvick D N, Newton K J. A mitochondrial mutator system in maize [J]. *Plant Physiol.*, 2005, 137:779–789.