

## 高等植物细胞器II类内含子剪接的研究进展

李秀兰, 姜曰水\*

曲阜师范大学生命科学学院, 山东曲阜273165

**摘要:** II类内含子是一类自我剪接内含子, 存在于真菌、细菌及古细菌中, 在植物细胞器中尤为普遍。典型II类内含子的剪接与剪接体内含子类似, 由内含子RNA及其编码的成熟酶介导完成。高等植物细胞器中, II类内含子成熟酶的结构发生变异, 导致自剪接功能丧失, 需要核编码蛋白因子辅助完成其剪接过程。自第一个参与叶绿体II类内含子剪接的核编码蛋白被发现以来, 越来越多的参与高等植物细胞器II类内含子剪接的核编码蛋白因子被报道。本文就高等植物细胞器II类内含子的分布、结构、剪接方式, 尤其参与它们剪接的核编码蛋白因子作一综述。

**关键词:** II类内含子; 叶绿体; 线粒体; 成熟酶; 剪接因子

叶绿体和线粒体是植物细胞的两个重要细胞器, 起源于宿主内共生的蓝藻细菌(Raven和Allen 2003)和 $\alpha$ -变形菌(Andersson等2003), 可产生ATP为细胞的生命活动提供能量。历经进化, 叶绿体和线粒体的大多数基因丢失或整合到宿主的核染色体上, 参与它们基因表达调控、生物发生及功能发挥的大多数蛋白由核基因组编码。高等植物叶绿体或线粒体基因的表达调控非常复杂(张琴等2015; 江芳等2016), 主要发生在转录后加工水平, 如RNA剪接、RNA编辑、RNA成熟等。

RNA剪接是指RNA前体中内含子的切除及编码多肽链的相邻外显子连接的过程。高等植物细胞器的多数基因含有内含子, 且通常位于保守结构基因内部, 因此, 这些内含子的精确剪接对细胞器基因的正常表达起关键作用。细胞器的内含子属于自我剪接内含子, 该类内含子RNA具有催化作用, 可催化自身的剪切及相邻外显子的连接。根据结构特征和剪接机制的不同, 自我剪接内含子可分为I类内含子(group I intron)和II类内含子(group II intron), I类内含子的二级结构主要由10个臂环状结构组成, 含有编码类似成熟酶的开放阅读框(open reading frame, ORF), 可借助游离的鸟嘌呤核苷酸, 通过两步转酯反应完成自我剪接; II类内含子的二级结构主要由呈辐射状分布的6个茎环结构域组成(Toor等2008), 含有编码成熟酶的ORF, 主要借助腺嘌呤核苷残基和两步转酯反应完成自我剪接。几乎所有高等植物细胞器内含子都属于II类内含子(Bonen 2008), 但它们编码成熟酶的ORF发生丢失或结构变异, 导致自剪接功能丧失(Zoschke等2010), 需要核编码蛋白因子辅助完成剪接过程。近年来, 多个参与高等植物细胞器II类

内含子剪接的核编码蛋白因子被鉴定。

### 1 II类内含子的分布与进化

II类内含子分布于原生生物、真菌、细菌及古细菌中, 在植物细胞器中尤为普遍。细菌的绝大多数II类内含子位于基因之间, 或插入到可移动的DNA元件内部, 极少数位于保守结构基因内部(Meng等2005)。植物细胞器中的内含子几乎都属于II类内含子, 拟南芥、玉米和水稻的叶绿体基因组约含有20个II类内含子和1个I类内含子, 开花植物的线粒体基因组约有25个II类内含子和1个I类内含子(Bonen 2008), 这些内含子一般位于高度保守基因内部。

关于细胞器II类内含子的起源, 一般认为它们来自细菌II类内含子, 后随内共生事件的发生进入真核生物的细胞器。基因组分析发现陆生植物细胞器II类内含子的进化非常复杂, 如地钱、苔藓和开花植物的线粒体基因组中虽然都有25个II类内含子, 但它们的同源性很低, 苔藓和开花植物中有8个II类内含子是相同的, 地钱和苔藓中只有3个, 地钱和开花植物间仅1个是相同的, 因此, 研究者认为这些内含子的进化独立于宿主, 且在进化过程中会分散性地缺失或获得某些序列(Groth-Malonek等2005)。即使在开花植物中, 同一II类内含子也存在很大差异(Kudla等2002), 如水菖蒲 $cox2$ 的内含子长462 bp, 而在银杏中则长达2 660 bp, 认为主要由于进化过程中内含子的某些结构

收稿 2017-04-12 修定 2017-07-11

资助 山东省高等学校科技计划项目(J16LE09)和曲阜师范大学博士启动基金(BSQD20152493)。

\* 通讯作者(E-mail: jewelseeker@163.com)。

域发生片段插入所致。和细菌的II类内含子相比,有些植物细胞器的II类内含子被其他基因隔开,如线粒体基因*nad1*、*nad2*、*nad5*和叶绿体基因*psaA*、*rps12*中的内含子。

## 2 II类内含子的结构与剪接

### 2.1 II类内含子的结构

虽然II类内含子的长度差异很大,核苷酸一级序列的同源性很低,但它们都能形成保守的二级结构(Toor等2008),该结构主要由呈辐射状分布的6个茎环结构域(D1~D6)组成。D1为最大的结构域,包含外显子结合位点EBS1 (exon-binding site 1)、EBS2、EBS3和 $\delta$ ,它们可分别与5'外显子中的内含子结合位点IBS1 (intron-binding site 1)、IBS2和3'外显子中的IBS3、 $\delta'$ 配对形成双链区,帮助完成内含子RNA的折叠、外显子识别及催化区对接等(Waldsich和Pyle 2008)。D4为最不保守的结构域,又被称为可变区,许多II类内含子编码成熟酶的ORF位于该结构域内。D5是保守性最高的区域(Toor等2009),通常为34 bp的茎环结构,可与D1支架通过2个四元环-受体相互作用形成催化核心, $Mg^{2+}$ 的结合位点位于该结构域。D6为靠近3'-剪接位点的螺旋结构,包含参与套索结构形成的腺嘌呤核苷残基(分支点A),该残基一般位于内含子3'末端上游7~8 bp处。虽然D6的保守性不高,但有研究显示该结构域是二价金属离子和内含子核心结合的重要区域,在内含子的催化过程中起重要作用(Sigel等2004)。D2和D3结构域较小,在内含子折叠过程中,这2个区域首先完成折叠,随后作为支架参与其他结构域的组装(Fedorova和Zingler 2007)。

植物细胞器中,并不是所有的II类内含子都具备上述结构特征,如叶绿体*trnV*的内含子和线粒体*nad1-1*的D6缺少分支点A残基(de Longevialle等2007),线粒体*nad4-2*的分支点A残基并不出现在最常见的位置(Bonen 2008)。叶绿体*ycf3-2*的D5包含36个核苷酸(de Longevialle等2008),而线粒体*nad1-2*的D5的核苷酸数却很少(Li-Pook-Than和Bonon 2006)。

### 2.2 II类内含子的剪接

根据剪接反应及被释放形式的不同,II类内含子的剪接可分为三种途径。第一种途径与核剪接

体内含子的剪接类似(Toro等2007),通过两步转酯反应完成。首先,EBS1与IBS1及EBS2与IBS2的相互作用使5'-剪接位点靠近D6的分支点A,A的2'-OH作为亲核基团攻击5'-剪接位点磷酸二酯键中的磷酸基团,形成5'-剪接位点游离的3'-OH和内含子套索结构的中间产物。接着,EBS3与IBS3、 $\delta$ 与 $\delta'$ 等单碱基的配对使3'-剪接位点正确定位,5'-剪接位点游离的3'-OH作为亲核基团攻击3'-剪接位点磷酸二酯键中的磷酸基团,形成新的3'-5'磷酸二酯键,连接5'和3'外显子,内含子以套索状结构释放。这类途径中的两步转酯反应高度可逆,即被剪切的内含子可通过逆剪接插入含有与原始5'外显子相似序列的DNA或RNA分子中。第二种为水解途径(Vogel和Börner 2002)。首先,水分子作为亲核试剂攻击5'-剪接位点,并在5'-剪接位点形成游离的3'-OH;接着,以与第一种途径的第二步转酯反应相同的过程完成剪接,内含子以线性方式释放。第三种剪接途径(Molina-Sánchez等2006)由一个游离5'外显子(被认为是以上两种途径中连接外显子重新打开的产物)的3'-OH攻击3'-剪接位点,连接游离的5'外显子和3'外显子,并在内含子的3'端形成游离的-OH,游离的3'-OH进而攻击5'剪接位点,释放环状的内含子和新的游离的5'外显子。

细菌和酵母中,几乎所有的II类内含子都以第一种途径完成剪接,而在植物细胞器中,有的II类内含子以第二或第三种途径完成剪接,如叶绿体*trnV*的内含子和线粒体的*nad1-1*因D6缺乏分支点A残基,只能以水解途径完成剪接,内含子以线性方式释放(de Longevialle等2007);小麦线粒体*nad1-2*以第三种途径完成剪接,内含子以环状分子被释放(Vogel和Börner 2002)。

## 3 植物细胞器II类内含子的剪接因子

II类内含子是可移动的遗传元件,完成自我剪接后可以归巢方式插入特异的靶位点,或以逆转座方式插入非特异位点。II类内含子的自我剪接和移动是由内含子RNA和内含子编码的成熟酶组成的复合物介导的。该成熟酶为多功能蛋白,在细菌中通常包括逆转录酶(reverse transcriptase, RT)、X、D (DNA-binding)和En (endonuclease) 4个功能区,RT结构域氨基酸序列保守,具有逆转录活性,可将内含子RNA逆转录成DNA;X结构域为

RNA结合区,具有成熟酶活性,在内含子的剪接中发挥功能;D结构域氨基酸序列保守性低,可与DNA结合;En结构域具有DNA核酸内切酶活性,与D结构域一起在内含子转移中起作用。与细菌相比,植物细胞器的II类内含子发生严重退化,开花植物线粒体和叶绿体基因组中仅存在1个内含子编码的成熟酶ORF, *mat-r*和*trnK*。*mat-r*编码线粒体的成熟酶MatR,它包含一个保守的X结构域,但RT结构域已退化,且En结构域缺失。叶绿体的*trnK*编码成熟酶MatK,多数MatK缺乏II类内含子移动所必需的结构域,可参与叶绿体少数内含子的剪接(Zoschke等2010)。因此,植物细胞器II类内含子的体内剪接需要核编码因子辅助完成,且近年来,越来越多参与植物细胞器II类内含子剪接的核编码蛋白因子被发现。

### 3.1 成熟酶

除细胞器基因组中的*mat-r*和*trnK*外,植物核基因组还有4个编码成熟酶的基因, *nMat1*、*nMat2*、*nMat3*和*nMat4*,它们在进化上独立于宿主,且所编码的蛋白与细胞器II类内含子编码的成熟酶高度同源(Keren等2009),推测这4个基因可能是由植物细胞器转移进入细胞核的。*nMat1~4*所编码的4个蛋白中, *nMAT1*、*nMAT2*和*nMAT3*定位于线粒体, *nMAT4*双定位于叶绿体和线粒体。根据拓扑学和进化起源可将这4个蛋白分为两类: *nMAT1*和*nMAT2*属于I型成熟酶,缺失D和En结构域, *nMAT3*和*nMAT4*属于II类成熟酶,包含成熟酶的4个特征结构域(Mohr和Lambowitz 2003),但D和En结构域发生变异。*nMAT*蛋白D/En结构域的变异或缺失致使其在内含子移动中的功能丧失,但仍可在植物细胞器部分II类内含子的剪接中发挥功能。*nMAT1*对线粒体*nad1-1*、*nad2-1*和*nad4-2*的剪接是必需的(Keren等2012), *nMAT2*参与*nad1-2*、*nad7-2*和*cox2*内含子的高效剪接(Keren等2009), *nMAT4*参与*nad1-1*、-2、-3的剪接(Cohen等2014),这些蛋白的功能缺失都会引起线粒体呼吸链复合物I缺陷的相关表型。

### 3.2 PPR蛋白

PPR (pentatricopeptide repeat)蛋白广泛存在于大多数真核生物中,在陆生植物中尤为普遍。PPR蛋白是由约35个氨基酸(PPR motif)为重复单位串

联而成的一类蛋白,根据PPR motif所包含氨基酸数目的不同,可将PPR蛋白分为P和PLS两类。PPR蛋白为RNA序列特异性结合蛋白,几乎都定位于线粒体或质体,可通过与特定核苷酸结合,参与RNA转录后加工,如编辑、剪接、成熟等(Lurin等2004)。

目前,报道了多个参与高等植物线粒体内含子剪接的PPR蛋白。拟南芥中的PPR蛋白OTP43、ABO5、TANG2、OTP439、BIR6、SLO3和MTL1对线粒体*nad1-1*、*nad2-3*、*nad5-2*和*nad7-1*等多个内含子的剪接是必需的(de Longevialle等2007; Liu等2010; des Francs-Small等2014; Koprivova等2010; Hsieh等2015; HaiLi等2016)。玉米中,PPR蛋白DEK2、EMP16和DEK35功能的丧失,导致*nad1-1*、*nad2-4*和*nad4-1*的剪接缺陷(Qi等2017; Xiu等2016; Chen等2017)。这些PPR蛋白的功能缺失,会导致线粒体复合物I缺陷,植株生长缓慢或种子发育畸形。除线粒体外,PPR蛋白也参与高等植物叶绿体内含子的剪接。拟南芥的PPR4、OTP51、THA8、PMD1参与内含子*rps12-1*、*ycf3-2*、*trnK*和*ndhA*的剪接(Schmitz-Linneweber等2006; de Longevialle等2008; Khrouchtchova等2012; Zhang等2015),HCF152结合在内含子和外显子的连接处,突变体中*petB*的剪接产物减少,被认为直接参与*petB*转录物的剪接(Meierhoff等2003),或通过影响成熟mRNA的降解而间接起作用(Pfalz等2009);水稻的WSL对*rpl2*前体mRNA的剪接是必需的(Tan等2014)。

### 3.3 CRM结构域蛋白

CRM (chloroplast RNA splicing and ribosome maturation)结构域起源于原核生物核糖体前体结合蛋白,与细菌中的YhbY同源,因参与叶绿体内含子剪接和核糖体大亚基催化核糖核蛋白的组装而被命名(Ostheimer等2003)。真核生物中,CRM结构域蛋白仅在植物中被发现,包含1~4个保守的CRM重复序列,信号肽预测多数定位于质体,少数定位于线粒体或细胞核。CRM结构域为RNA识别结合结构域,其识别、结合特征与RRM (RNA recognition motif)类似(Keren等2008)。近年来的研究发现,CRM结构域蛋白可与植物细胞器内含子RNA结合并参与其剪接。

叶绿体的CRS1 (chloroplast RNA splicing 1)包含3个CRM结构域,可通过体内高亲和、特异性地结合 $atpF$ 内含子的D1和D4,帮助内含子正确折叠并参与其剪接(Ostersetzer等2005); CRS2与细菌的肽基-tRNA水解酶同源,通过与包含2个CRM结构域的CAF1 (CRS2 associated factor)和CAF2蛋白互作,参与玉米叶绿体9个内含子的剪接(Ostheimer等2003); CFM2 (CRM family member 2)包含4个CRM结构域,参与叶绿体 $nadA$ 和 $ycf3$ 内含子的剪接(Asakura和Barkan 2007)。与叶绿体相比,参与线粒体内含子剪接的CRM结构域蛋白仅2个被鉴定,拟南芥的mCSF1 (mitochondria CAF-like splicing factor 1)和mCSF2包含2个CRM结构域,为CAF类剪接因子。mCSF1参与线粒体12个II类内含子的剪接,它的突变导致线粒体复合物I和复合物IV功能缺陷(Zmudjak等2013)。

### 3.4 DEAD-box RNA解旋酶

DEAD-box RNA解旋酶属于解旋酶II家族,含有保守的ATP结合域、水解结构域、RNA结合结构域和一个DEAD (Asp-Glu-Ala-Asp)序列。DEAD-box RNA解旋酶广泛存在于所有真核生物和许多原核生物中,主要参与ATP依赖的分子内和分子间RNA结构重排、核糖核蛋白复合物的重新组装,也有研究发现DEAD-box RNA解旋酶参与RNA合成、修饰、剪切、降解及核糖体的生物合成和翻译起始等(Linder和Jankowsky 2011)。预测高等植物中约有60个DEAD-box RNA解旋酶(Mingam等2004),但绝大多数功能未知。玉米和拟南芥的RH3 (RNA helicase 3)参与叶绿体 $rpl12$ 、 $rpl21$ 、 $atpF$ 等内含子的剪接(Asakura等2012; Gu等2014); ISE2参与拟南芥叶绿体 $rpl2$ 内含子的剪接(Carlotto等2015); 拟南芥的PMH2 (putative mitochondrial helicase 2)被认为是剪接体的成员之一,PMH2的纯合突变体中多个线粒体内含子的剪接效率明显下降,推测其作为RNA分子伴侣,通过促进形成或稳定RNA的二级结构而影响剪接(Köhler等2010)。

除上述核编码剪接因子外,还有些蛋白家族,如PORR (plant organellar RNA recognition)、RCC (regulator of chromosome condensation)等也参与高等植物细胞器II类内含子的剪接。玉米的PORR蛋白WTF1可启动叶绿体中约一半内含子的剪接

(Kroeger等2009),拟南芥的WTF9特异性地参与线粒体 $rpl2$ 和 $ccmFc$ 内含子的剪接(des Francs-Small等2012)。对于RCC蛋白家族,迄今仅发现一个成员RUG3参与拟南芥 $nad2-2$ 、 $-3$ 的剪接(Kühn等2011)。

综上所述,高等植物细胞器II类内含子的剪接是通过多个核编码蛋白因子共同参与完成的,核编码蛋白因子在细胞器基因表达调控中起重要作用。

### 4 问题与展望

进化过程中,高等植物细胞器II类内含子编码成熟酶的ORF发生缺陷或丢失,丧失了完成自我剪接的能力。越来越多参与高等植物细胞器II类内含子剪接的核编码蛋白的发现,证明这些内含子的剪接需要核编码蛋白辅助完成,然而它们的作用机制仍不清楚。核编码细胞器内含子剪接因子多为RNA识别结合蛋白,有的可参与植物细胞器多个II类内含子剪接事件的发生,如nMAT2参与线粒体 $nad2$ 、 $nad7$ 和 $cox2$ 中3个内含子的剪接(Keren等2009),WTF1对叶绿体 $atpF$ 、 $trnV$ 和 $ndhB$ 中等9个内含子的剪接是必需的(Kroeger等2009)。内含子RNA的空间结构是完成II类内含子剪接的关键因素,推测这些剪接因子可能通过识别多个内含子的共有或相似序列,或识别内含子的二级结构,参与内含子的剪接。有的核编码剪接因子只参与细胞器某个II类内含子的剪接,如EMP16和OPT43等(Xiu等2016; de Longevialle等2007),认为它们可能是通过识别某个内含子所特有的一级结构参与剪接的。虽然高等植物细胞器中的核编码剪接因子多数为RNA识别结合蛋白,但仅发现CRS1可与内含子的D1和D4结合(Ostersetzer等2005),而其他剪接因子的RNA识别位点尚未见报道,推测这类核编码剪接因子可能作为内含子RNA的分子伴侣,防止其错误折叠或被降解,辅助内含子RNA形成利于剪接的高级结构。由于外显子-内含子边界序列及剪接机制的相似性,研究者认为II类内含子为真核生物剪接体内含子的进化祖先(Martin和Koonin 2006)。Qu等(2016)最近通过冷冻电镜技术发现,细菌II类内含子剪接反应核心区域与剪接体在结构上存在高度的保守性,由此推测核编码的细胞器RNA剪接因子也可能通过招募其他蛋白因子发

挥作用, 形成类似于剪接体的复合物, 如CRS2通过与CAF1和CAF2互作参与剪接(Ostheimer等2003), CFM2和CFM3一起与CRS2/CAF形成复合物, 参与叶绿体内含子的剪接(Asakura等2008)。尽管II类内含子(Toor等2008)的晶体结构已被解析, 但是对核编码剪接因子的结构了解甚少, 目前仅解析了THA8 (Ban等2015)的晶体结构, 因此, 高等植物细胞器II类内含子与剪接体内含子间的关系, 以及核编码蛋白参与细胞器II类内含子剪接的作用机理尚需进一步研究, 相信随着更多核编码蛋白剪接因子的发现, 冷冻电镜、RNA免疫共沉淀等实验技术的发展, 高等植物细胞器II类内含子的剪接机制将会越来越明晰。

### 参考文献

- Andersson SGE, Karlberg O, Canbäck B, Kurland CG (2003). On the origin of mitochondria: a genomics perspective. *Phil Trans R Soc Lond B*, 358: 165–179
- Asakura Y, Barkan A (2007). A CRM domain protein functions dually in group I and group II intron splicing in land plant chloroplasts. *Plant Cell*, 19: 3864–3875
- Asakura Y, Bayraktar OA, Barkan A (2008). Two CRM protein subfamilies cooperate in the splicing of group IIB introns in chloroplasts. *RNA*, 14: 2319–2332
- Asakura Y, Galarneau E, Watkins KP, Barkan A, van Wijk KJ (2012). Chloroplast RH3 DEAD box RNA helicases in maize and Arabidopsis function in splicing of specific group II introns and affect chloroplast ribosome biogenesis. *Plant Physiol*, 159: 961–974
- Ban T, Zhu JK, Melcher K, Xu HE (2015). Structural mechanism of RNA recognition: sequence-specific and non-specific RNA-binding proteins and the Cas9-RNA-DNA complex. *Cell Mol Life Sci*, 72: 1045–1058
- Bonen L (2008). *Cis-* and *trans-*splicing of group II introns in plant mitochondria. *Mitochondrion*, 8: 26–34
- Carlotto N, Wirth S, Furman N, Solari NF, Ariel F, Crespi M, Kobayashi K (2015). The chloroplastic DEVH-box RNA helicase INCREASED SIZE EXCLUSION LIMIT 2 involved in plasmodesmata regulation is required for group II intron splicing. *Plant Cell Env*, 39: 165–173
- Chen XZ, Feng F, Qi WW, Xu LM, Yao DS, Wang Q, Song RT (2017). *Dek35* encodes a PPR protein that affects *cis*-splicing of mitochondrial *nad4* intron 1 and seed development in maize. *Mol Plant*, 10: 427–441
- Cohen S, Zmudjak M, des Francs-Small CC, Malik S, Shaya F, Keren I, Belausov E, Many Y, Brown GG, Small I, et al (2014). nMAT4, a maturase factor required for *nad1* pre-mRNA processing and maturation, is essential for holocomplex I biogenesis in Arabidopsis mitochondria. *Plant J*, 78: 253–268
- de Longevialle AF, Hendrickson L, Taylor NL, Delannoy E, Lurin C, Badger M, Millar AH, Small I (2008). The pentatricopeptide repeat gene *OTP51* with two LAGLIDADG motifs is required for the *cis*-splicing of plastid *ycf3* intron 2 in *Arabidopsis thaliana*. *Plant J*, 56: 157–168
- de Longevialle AF, Meyer EH, Andrés C, Taylor NL, Lurin C, Millar AH, Small I (2007). The pentatricopeptide repeat gene *OTP43* is required for *trans*-splicing of the mitochondrial *nad1* intron 1 in *Arabidopsis thaliana*. *Plant Cell*, 19: 3256–3265
- des Francs-Small CC, de Longevialle FA, Li Y, Lowe E, Tanz SK, Smith C, Bevan MW, Small I (2014). The pentatricopeptide repeat proteins TANG2 and ORGANELLE TRANSCRIPT PROCESSING439 are involved in the splicing of the multipartite *nad5* transcript encoding a subunit of mitochondrial complex I. *Plant Physiol*, 165: 1409–1416
- des Francs-Small CC, Kroeger T, Zmudjak M, Ostersetzer-Biran O, Rahimi N, Small I, Barkan A (2012). A PORR domain protein required for *rpl2* and *ccmF<sub>c</sub>* intron splicing and for the biogenesis of *c*-type cytochromes in Arabidopsis mitochondria. *Plant J*, 69: 996–1005
- Fedorova O, Zingler N (2007). Group II introns: structure, folding and splicing mechanism. *Biol Chem*, 388: 665–678
- Groth-Maloney M, Pruchner D, Grewe F, Knoop V (2005). Ancestors of *trans*-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol Biol Evol*, 22 (1): 117–125
- Gu L, Xu T, Lee K, Lee KH, Kang H (2014). A chloroplast-localized DEAD-box RNA helicase AtRH3 is essential for intron splicing and plays an important role in the growth and stress response in *Arabidopsis thaliana*. *Plant Physiol Biochem*, 82: 309–318
- Haïli N, Planchard N, Arnal N, Quadrado M, Vrielynck N, Dahan J, des Francs-Small CC, Mireau H (2016). The MTL1 pentatricopeptide repeat protein is required for both translation and splicing of the mitochondrial *NADH DEHYDROGENASE SUBUNIT7* mRNA in Arabidopsis. *Plant Physiol*, 170: 354–366
- Hsieh WY, Liao JC, Chang CY, Harrison T, Boucher C, Hsieh MH (2015). The SLOW GROWTH3 pentatricopeptide repeat protein is required for the splicing of mitochondrial *NADH dehydrogenase subunit7* intron 2 in Arabidopsis. *Plant Physiol*, 168: 490–501
- Jiang F, Ding ZH, Dong L, Li PH (2016). Transcriptome analysis on the maize photosynthetic mutant *bsd2* (*bundle sheath defective II*). *Plant Physiol J*, 52 (8): 1214–1222 (in Chinese with English abstract) [江芳, 丁泽红, 董雷, 李平华(2016). 玉米光合突变体 *bsd2* (*bundle sheath defective II*)的转录组分析. *植物生理学报*, 52 (8): 1214–1222]
- Keren I, Bezawork-Geleta A, Kolton M, Maayan I, Belausov E, Levy M, Mett A, Gidoni D, Shaya F, Ostersetzer-Biran O (2009). AtnMat2, a nuclear-encoded maturase required for splicing of group II introns in Arabidopsis mitochondria. *RNA*, 15: 2299–2311
- Keren I, Klipcan L, Bezawork-Geleta A, Kolton M, Shaya F, Ostersetzer-Biran O (2008). Characterization of the molecular basis of group II intron RNA recognition by CRS1-CRM domains. *J Biol Chem*, 283: 23333–23342
- Keren I, Tal L, des Francs-Small CC, Araújo WL, Shevtsov S, Shaya F, Fernie AR, Small I, Ostersetzer-Biran O (2012). nMAT1, a

- nuclear-encoded maturase involved in the *trans*-splicing of *nad1* intron 1, is essential for mitochondrial complex I assembly and function. *Plant J*, 71: 413–426
- Khrouchtchova A, Monde RA, Barkan A (2012). A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *RNA*, 18: 1197–1209
- Köhler D, Schmidt-Gattung S, Binder S (2010). The DEAD-box protein PMH2 is required for efficient group II intron splicing in mitochondria of *Arabidopsis thaliana*. *Plant Mol Biol*, 72: 459–467
- Koprivova A, des Francs-Small CC, Calder G, Mugford ST, Tanz S, Lee BR, Zechmann B, Small I, Kopriva S (2010). Identification of a pentatricopeptide repeat protein implicated in splicing of intron 1 of mitochondrial *nad7* transcripts. *J Biol Chem*, 285: 32192–32199
- Kroeger TS, Watkins KP, Friso G, van Wijk KJ, Barkan A (2009). A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing. *Proc Natl Acad Sci USA*, 106 (11): 4537–4542
- Kudla J, Albertazzi FJ, Blazević D, Hermann M, Bock R (2002). Loss of the mitochondrial *cox2* intron 1 in a family of monocotyledonous plants and utilization of mitochondrial intron sequences for the construction of a nuclear intron. *Mol Genet Genomics*, 267: 223–230
- Kühn K, Carrie C, Giraud E, Wang Y, Meyer EH, Narsai R, des Francs-Small CC, Zhang BT, Murcha MW, Whelan J (2011). The RCC1 family protein RUG3 is required for splicing of *nad2* and complex I biogenesis in mitochondria of *Arabidopsis thaliana*. *Plant J*, 67: 1067–1080
- Li-Pook-Than J, Bonen L (2006). Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic Acids Res*, 34 (9): 2782–2790
- Linder P, Jankowsky E (2011). From unwinding to clamping the DEAD box RNA helicase family. *Mol Cell Biol*, 12: 505–516
- Liu Y, He J, Chen ZZ, Ren XZ, Hong XH, Gong ZZ (2010). *ABA overly-sensitive 5 (ABO5)*, encoding a pentatricopeptide repeat protein required for *cis*-splicing of mitochondrial *nad2* intron 3, is involved in the abscisic acid response in *Arabidopsis*. *Plant J*, 63: 749–765
- Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B, et al (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, 16: 2089–2103
- Martin W, Koonin EV (2006). Introns and the origin of nucleus-cytoplasm compartmentalization. *Nature*, 440: 41–45
- Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G (2003). HCF152, an *Arabidopsis* RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. *Plant Cell*, 15: 1480–1495
- Meng Q, Wang YF, Liu XQ (2005). An intron-encoded protein assists RNA splicing of multiple similar introns of different bacterial genes. *J Biol Chem*, 280 (42): 35085–35088
- Mingam A, Toffano-Nioche C, Brunaud V, Boudet N, Kreis M, Lecharny A (2004). DEAD-box RNA helicases in *Arabidopsis thaliana*: establishing a link between quantitative expression, gene structure and evolution of a family of genes. *Plant Biotechnol J*, 2: 401–415
- Mohr G, Lambowitz AM (2003). Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucleic Acids Res*, 31 (2): 647–652
- Molina-Sánchez MD, Martínez-Abarca F, Toro N (2006). Excision of the *Sinorhizobium meliloti* group II intron RmInt1 as circles *in vivo*. *J Biol Chem*, 281: 28737–28744
- Ostersetzer O, Cooke AM, Watkins KP, Barkan A (2005). CRS1, a chloroplast group II intron splicing factor, promotes intron folding through specific interactions with two intron domains. *Plant Cell*, 17: 241–255
- Ostheimer GJ, Williams-Carrier R, Belcher S, Osborne E, Gierke J, Barkan A (2003). Group II intron splicing factors derived by diversification of an ancient RNA-binding domain. *EMBO J*, 22 (15): 3919–3929
- Pfalz J, Bayraktar OA, Prikryl J, Barkan A (2009). Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J*, 28 (14): 2042–2052
- Qi WW, Yang Y, Feng XZ, Zhang ML, Song RT (2017). Mitochondrial function and maize kernel development requires Dek2, a pentatricopeptide repeat protein involved in *nad1* mRNA splicing. *Genetics*, 205: 239–249
- Qu GS, Kaushal PS, Wang J, Shigematsu H, Piazza, CL, Agrawal RK, Belfort M, Wang HW (2016). Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol*, 23 (6): 549–557
- Raven JA, Allen JF (2003). Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol*, 4 (3): 209–233
- Schmitz-Linneweber C, Williams-Carrier RE, Williams-Voelker PM, Kroeger TS, Vichas A, Barkan A (2006). A pentatricopeptide repeat protein facilitates the *trans*-splicing of the maize chloroplast *rps12* pre-mRNA. *Plant Cell*, 18: 2650–2663
- Sigel RK, Sashital DG, Abramovitz DL, Palmer III AG, Butcher SE, Pyle AM (2004). Solution structure of domain 5 of a group II intron ribozyme reveals a new RNA motif. *Nat Struct Mol Biol*, 11 (2): 187–192
- Tan JJ, Tan ZH, Wu FQ, Sheng PK, Heng YQ, Wang XH, Ren YL, Wang JL, Guo XP, Zhang X, et al (2014). A novel chloroplast-localized pentatricopeptide repeat protein involved in splicing affects chloroplast development and abiotic stress response in rice. *Mol Plant*, 7: 1329–1349
- Toro N, Jiménez-Zurdo JI, García-Rodríguez FM (2007). Bacterial group II introns: not just splicing. *FEMS Microbiol Rev*, 31: 342–358
- Toor N, Keating KS, Pyle AM (2009). Structural insights into RNA splicing. *Curr Opin Struct Biol*, 19: 260–266
- Toor N, Keating KS, Taylor SD, Pyle AM (2008). Crystal structure of a self-spliced group II intron. *Science*, 320: 77–82
- Vogel J, Börner T (2002). Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *EMBO J*, 21 (14): 3794–3803
- Waldsich C, Pyle AM (2008). A kinetic intermediate that regulates

- proper folding of a group II intron RNA. *J Mol Biol*, 375: 572–580
- Xiu ZH, Sun F, Shen Y, Zhang XY, Jiang RC, Bonnard G, Zhang JH, Tan BC (2016). EMPTY PERICARP16 is required for mitochondrial *nad2* intron 4 *cis*-splicing, complex I assembly and seed development in maize. *Plant J*, 85: 507–519
- Zhang HD, Cui YL, Huang C, Yin QQ, Qin XM, Xu T, He XF, Zhang Y, Li ZR, Yang ZN (2015). PPR protein PDM1/SEL1 is involved in RNA editing and splicing of plastid genes in *Arabidopsis thaliana*. *Photosynth Res*, 126: 311–321
- Zhang Q, Jia N, Li H, Gao HB (2015). Identification and analysis of *ARC6* gene in a chloroplast division mutant *cpd44* in *Arabidopsis thaliana*. *Plant Physiol J*, 51 (4): 501–508 (in Chinese with English abstract) [张琴, 贾宁, 李慧, 高宏波(2015). 拟南芥叶绿体分裂突变体*cpd44*和*ARC6*基因的鉴定与分析. 植物生理学报, 51 (4): 501–508]
- Zmudjak M, des Francs-Small CC, Keren I, Shaya F, Belausov E, Small I, Ostersetzer-Biran O (2013). mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in *Arabidopsis*. *New Phytol*, 199: 379–394
- Zoschke R, Nakamura M, Liere K, Sugiura M, Börner T, Schmitz-Linneweber C (2010). An organellar maturase associates with multiple group II introns. *Proc Natl Acad Sci USA*, 107 (7): 3245–3250

## Research progress of group II intron splicing in higher plant organelles

LI Xiu-Lan, JIANG Yue-Shui\*

*College of Life Science, Qufu Normal University, Qufu, Shandong 273165, China*

**Abstract:** Group II introns are large self-spliced introns that are found in fungi, bacteria and archaeobacteria, but are particularly prevalent within the organelle genomes in plants. The splicing of typical group II introns is closely resembles to that of the nuclear spliceosomal introns, which is carried out by the complex of intron RNA and intron-encoded maturase. However, group II introns in higher plant organelles have lost the ability to self-splice *in vivo* and require nucleus-encoded proteins as cofactors. Since the first splicing factor was identified in chloroplasts, more and more nuclear-encoded proteins have been shown to be involved in the splicing of the introns in chloroplasts or mitochondria. In this review, we will summarize the data on the distribution, structure and splicing of group II introns in higher plant organelles, with a focus on their nuclear-encoded splicing cofactors.

**Key words:** group II intron; chloroplast; mitochondrion; maturase; splicing factor

Received 2017-04-12 Accepted 2017-07-11

This work was supported by Shandong Province Higher Educational Science and Technology Program (Grant No. J16LE09) and Qufu Normal University Doctoral Fund Project (Grant No. BSQD20152493).

\*Corresponding author (E-mail: jewelseeker@163.com).