

DOI: 10.11913/PSJ.2095-0837.2020.60802

周皇经, 王翔伍, 吉逢逢, 张鹏辉, 郭清华, 郝云飞, 何世斌. 西瓜含有 JmjC 结构域的组蛋白去甲基化酶家族的生物信息学分析[J]. 植物科学学报, 2020, 38(6): 802-811

Zhou HJ, Wang XW, Ji FF, Zhang PH, Guo QH, Hao YF, He SB. Bioinformatics analysis of JmjC domain-containing histone demethylase family in watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) [J]. *Plant Science Journal*, 2020, 38(6): 802-811

西瓜含有 JmjC 结构域的组蛋白去甲基化酶家族的生物信息学分析

周皇经, 王翔伍, 吉逢逢, 张鹏辉, 郭清华, 郝云飞, 何世斌*

(河南大学生命科学学院, 棉花生物学国家重点实验室, 河南开封 475004)

摘要: 利用生物信息学方法, 对西瓜 (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) *JmjC* 基因家族的成员进行鉴定, 对该基因家族的染色体定位、基因结构、蛋白结构域、选择压力和酶活位点进行分析, 并对该基因家族与其它物种的系统进化及共线性关系进行研究。结果显示: 西瓜全基因组含有 17 个 *JmjC* 候选基因, 核苷酸序列长度为 1209 ~ 5541 bp; 这些基因均含有 *JmjC* 结构域, 分别位于 9 条染色体上, 归属 8 个亚族。系统进化、选择压力以及共线性分析结果表明, 西瓜与黄瓜 (*Cucumis sativus* L.) 亲缘关系较近, *JmjC* 家族基因数量相同, 其中 14 个成员呈现一对一的共线性关系; 而西瓜与拟南芥 (*Arabidopsis thaliana* (L.) Heynh) 亲缘关系较远, 但西瓜和拟南芥同一亚族中 *JmjC* 基因间 K_a/K_s 的比值均小于 1, 推测西瓜各个亚族成员的编码蛋白功能与同一亚族的拟南芥成员功能极为相似。酶活位点分析结果表明西瓜 *JmjC* 基因家族中有 10 个成员具有潜在的组蛋白去甲基化酶活性。

关键词: 西瓜; 组蛋白去甲基化酶; *JmjC* 结构域; 生物信息学分析

中图分类号: Q943.2

文献标识码: A

文章编号: 2095-0837(2020)06-0802-10

Bioinformatics analysis of *JmjC* domain-containing histone demethylase family in watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai)

Zhou Huang-Jing, Wang Xiang-Wu, Ji Feng-Feng, Zhang Peng-Hui,
Guo Qing-Hua, Hao Yun-Fei, He Shi-Bin*

(School of Life Sciences, State Key Laboratory of Cotton Biology, Henan University, Kaifeng, Henan 475004, China)

Abstract: Jumonji C (*JmjC*) domain-containing proteins are a class of histone demethylases responsible for removing histone methylation marks. They also play important roles in plant growth and development. Using bioinformatics, we identified the *JmjC* genes in *Citrullus lanatus* (Thunb.) Matsum. & Nakai, and analyzed their location, gene structure, protein domain, enzyme activity site, and selection pressure. In addition, the phylogenetic relationships and collinearity among *C. lanatus*, *Cucumis sativus* L., and *Arabidopsis thaliana* (L.) Heynh were also analyzed. A total of 17 *JmjC* genes were identified in *C. lanatus*, which ranged in size from 1209 to 5541 bp. These genes all contained the *JmjC* domain and were located on nine chromosomes and classified into eight subfamilies. Both *C. lanatus* and closely

收稿日期: 2020-04-10, 退修日期: 2020-05-18。

基金项目: 国家自然科学基金项目(31401044)。

This work was supported by a grant from the National Natural Science Foundation of China (31401044).

作者简介: 周皇经(1998~), 男, 本科生, 研究方向为植物表观遗传学(E-mail: 2373006425@qq.com)。

* 通讯作者(Author for correspondence. E-mail: sbhe@henu.edu.cn)

related species *C. sativus* shared the same number of *JmjC* genes, with 14 members showing a one-to-one collinear relationship, while *C. lanatus* and distant species *A. thaliana* showed lower collinearity. The K_a/K_s ratios of *JmjC* genes in the same subfamily were less than One between *C. lanatus* and *A. thaliana*, indicating that orthologous genes between the two species should have similar functions. In addition, according to enzyme activity site prediction, we speculated that 10 of them may confer histone demethylase activity.

Key words: *Citrullus lanatus*; Histone demethylase; *JmjC* domain; Bioinformatics analysis

在真核生物中,核小体是染色质的基本结构单元,每个核小体由145~147 bp的DNA缠绕组蛋白八聚体形成。组蛋白N端氨基酸残基易受到各种修饰,包括甲基化、乙酰化、磷酸化等,这些修饰影响染色质的构型,进而调节DNA复制、转录、修复等过程^[1]。其中,组蛋白甲基化主要发生在H3和H4的N端赖氨酸和精氨酸的残基上,赖氨酸可被一、二、三甲基化,精氨酸可被一、二甲基化,不同位置不同程度的甲基化,其功能也是不同的,如组蛋白H3K4的二甲基化、三甲基化与基因转录激活有关,而H3K27三甲基化与基因沉默有关^[2,3]。

组蛋白甲基化是一个可逆的过程,由组蛋白甲基化转移酶和去甲基化酶共同调控。根据组蛋白去甲基化酶的结构特征可分为两大类:LSD1(Lysine Specific Demethylase 1)和含有JmjC(Jumonji C)结构域的去甲基化酶家族。LSD1是以黄素腺嘌呤二核苷酸为辅因子的特异性赖氨酸去甲基化酶,只能去除H3K4的一、二甲基化,不能去除三甲基化^[4]。而含有JmjC结构域的去甲基化酶以二价亚铁离子和 α -酮戊二酸为辅因子,可以去除组蛋白赖氨酸多个位点的一、二、三甲基化^[5]。含有JmjC结构域的去甲基化酶在植物的生长发育、成花转变和逆境胁迫等方面具有重要的功能。如拟南芥(*Arabidopsis thaliana* (L.) Heynh)AtJMJ30负责调控开花时间、根的生长等^[6,7];而AtJMJ16、AtJMJ17等JmjC家族成员则与叶片衰老、响应干旱胁迫有关^[8,9]。

目前,通过生物信息学方法已经完成了多种植物JmjC基因家族成员的鉴定,包括拟南芥^[10]、水稻(*Oryza sativa* L.)^[10]、玉米(*Zea mays* L.)^[11]、月季(*Rosa chinensis* Jacq.)^[12]等。2015年张亮生团队系统分析了真核生物中JmjC基因家族的起源和分子演化;将植物中该基因家族

成员划分为9个亚族,包括:JMJD6 (Jumonji domain containing protein 6)、KDM3 (Lysine-specific demethylase 3)、KDM5、PKDM7~PKDM9 (Putative-lysine-specific demethylase)以及PKDM11~PKDM13^[13]。而蛋白晶体结构和氨基酸突变酶活实验的研究发现,JmjC结构域中结合二价亚铁离子、 α -酮戊二酸以及组蛋白多肽的位点突变会显著影响其催化活性^[14]。曹晓凤团队通过拟南芥、水稻与人类同源蛋白的比较,推测分别有6、7个成员没有去甲基化酶活性^[10]。月季中RcJMJ40由于缺失2个二价亚铁离子和1个 α -酮戊二酸结合位点的片段,推测其不具有组蛋白去甲基化酶活性^[12]。系统性的鉴定JmjC基因家族成员并进行科学的分类,以及基于序列预测是否具有去甲基化酶活性,是开展其功能研究的基础。

西瓜(*Citrullus lanatus* (Thunb.) Matsum. & Nakai)属于葫芦科、西瓜属(*Citrullus*),是世界十大水果之一。其果肉营养丰富,富含瓜氨酸、番茄红素和其它有益人类健康的化合物^[15],但目前没有关于西瓜中JmjC基因的研究报道。本研究通过已公布的东亚西瓜栽培种‘97103’的三代测序基因组数据^[16],采用生物信息学方法对其JmjC基因家族进行系统分析,主要包括:JmjC家族成员的鉴定和分类,系统进化关系、选择压力和共线性分析,着重研究该家族蛋白序列中的关键酶活位点并初步确定其中可能具有去甲基化酶活的家族基因,以期为进一步开展西瓜JmjC基因家族的功能研究奠定基础。

1 材料和方法

1.1 西瓜JmjC基因家族成员的鉴定和分析

本研究从TAIR (<https://www.arabidopsis.org/>)和Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>)数据库中分别下载拟南芥和

水稻的 *JmjC* 家族的相关序列作为检索序列；在葫芦科基因组数据库 CuGenDB (<http://cucurbit- genomics.org/>) 下载西瓜栽培种‘97103’和黄瓜 (*Cucumis sativus* L.) 栽培种‘9930’的基因组数据进行 BlastP 检索(阈值为 1e-10)，获取相关候选基因。在 Pfam (<http://pfam.xfam.org/>) 网站中下载 *JmjC* 结构域(PE02373)的序列信息，进行本地建模，通过 Hmmscan 检索西瓜、黄瓜蛋白序列数据库。将两种方法得到的蛋白序列进行整合，利用 SMART (<http://smart.embl.de/>) 网站复筛含有 *JmjC* 结构域的序列。西瓜 *JmjC* 家族基因的蛋白等电点、分子量等基本参数通过在线工具 ExPasy (<https://www.expasy.org/>) 进行分析。采用 Tbtools 软件进行基因结构、结构域以及基因染色体定位图的绘制。

1.2 西瓜 *JmjC* 基因家族的系统进化分析

采用 MEGA7 软件的 MUSCLE 程序进行蛋白序列比对，p-distance 模型构建系统进化树，采用邻接法和最大似然法构树，Bootstrap 值设为 1000。采用 Itol (<https://itol.embl.de/>) 在线工具优化系统进化树。

1.3 西瓜 *JmjC* 基因家族的选择压力和共线性分析

将西瓜、拟南芥和黄瓜进行全基因组的蛋白序列双向比对，在 Java 环境下通过 MCScanX 软件进行共线性分析，设置共线性区块大小为 30，得

到的结果采用 Tbtools 进行可视化^[17]。通过 MEGA7 软件的 ClustalW 算法比对拟南芥和西瓜 *JmjC* 家族基因的 CDS 序列，比对结果采用 Tbtools 的 Simple K_a/K_s Calculator (NG) 模块进行选择压力计算。

1.4 西瓜 *JmjC* 基因家族的酶活关键位点分析及三级结构预测

在 PDB (<http://www.rcsb.org/>) 数据库中下载拟南芥的 AtPKDM7 (PDB ID: 6IP4) 和 AtPKDM8 (PDB ID: 5YKO) 的蛋白晶体信息。采用 ClustalW 软件分别比对西瓜的 PKDM7/PKDM8 亚族。并根据文献数据所鉴定的突变严重影响酶活的位点^[14, 18]，对 PKDM7 和 PKDM8 亚族内的成员比对后的家族基因关键位点进行标注。其它非拟南芥同源基因实验验证的关键位点，来自拟南芥、水稻与人类同源基因的序列比对^[10]。三级结构预测和模型构建通过在线网站 Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html>) 进行分析^[19]。

2 结果与分析

2.1 西瓜 *JmjC* 基因家族成员鉴定和分析

本研究经过序列比对，在西瓜基因组中鉴定出 17 个 *JmjC* 基因家族成员，分布在 8 个亚族中 (表 1)。结果显示，西瓜 *JmjC* 基因家族成员编码区长度差异较大，约在 1209~5541 bp，蛋白分子量为 44.86~212.89 kD，所有蛋白的等电点为 5.01~8.81。

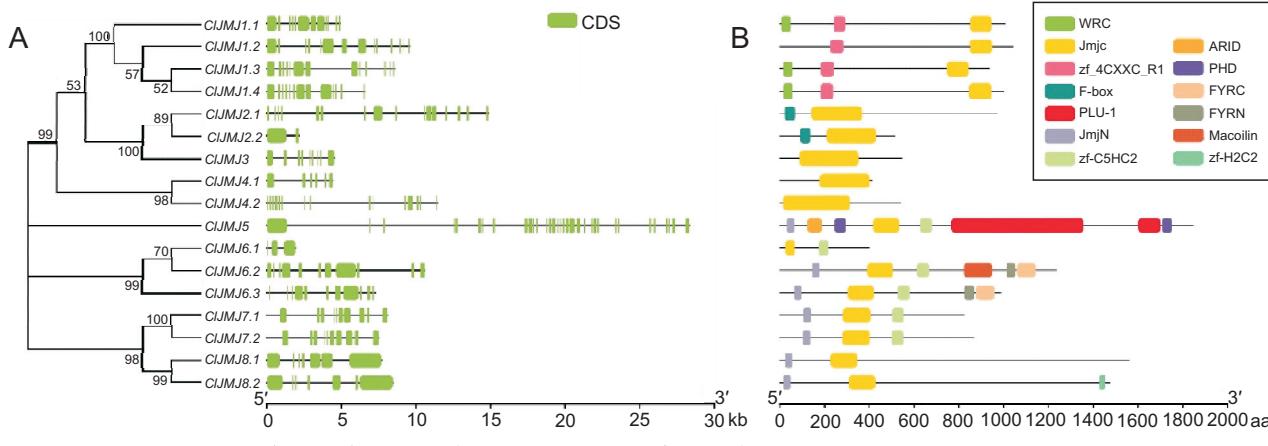
表 1 西瓜 *JmjC* 基因家族成员信息
Table 1 Information of *JmjC*-domain containing gene members in *Citrullus lanatus*

| 基因号 Gene ID | 基因名 Gene name | 亚族 Subfamily | 编码区长度 CDS length (bp) | 分子量 Molecular weight (kD) | 等电点 Isoelectric point |
|-------------------|------------------|-----------------|--------------------------|------------------------------|--------------------------|
| Cla97C01G022300.1 | CIJMJ1.1 | KDM3 | 3021 | 114.44 | 7.69 |
| Cla97C01G008510.1 | CIJMJ1.2 | KDM3 | 3126 | 118.51 | 6.13 |
| Cla97C02G041280.1 | CIJMJ1.3 | KDM3 | 2820 | 106.13 | 5.89 |
| Cla97C05G080920.1 | CIJMJ1.4 | KDM3 | 3009 | 114.07 | 8.11 |
| Cla97C08G154190.1 | CIJMJ2.1 | JMJD6 | 2910 | 111.23 | 5.26 |
| Cla97C10G200900.1 | CIJMJ2.2 | JMJD6 | 1539 | 58.15 | 8.45 |
| Cla97C06G126860.1 | CIJMJ3 | PKDM11 | 1638 | 62.85 | 5.01 |
| Cla97C03G063630.1 | CIJMJ4.1 | PKDM12 | 1239 | 46.53 | 5.07 |
| Cla97C09G170050.1 | CIJMJ4.2 | PKDM12 | 1626 | 61.37 | 5.13 |
| Cla97C01G002850.1 | CIJMJ5 | KDM5 | 5541 | 212.89 | 6.82 |
| Cla97C09G170690.1 | CIJMJ6.1 | PKDM7 | 1209 | 44.86 | 8.33 |
| Cla97C11G216880.1 | CIJMJ6.2 | PKDM7 | 3702 | 138.22 | 6.64 |
| Cla97C01G009140.1 | CIJMJ6.3 | PKDM7 | 2961 | 111.32 | 6.01 |
| Cla97C09G174380.1 | CIJMJ7.1 | PKDM8 | 2466 | 93.42 | 7.99 |
| Cla97C10G185960.1 | CIJMJ7.2 | PKDM8 | 2604 | 97.82 | 7.45 |
| Cla97C03G054490.1 | CIJMJ8.1 | PKDM9 | 4686 | 174.87 | 8.81 |
| Cla97C08G150370.1 | CIJMJ8.2 | PKDM9 | 4422 | 165.00 | 5.93 |

西瓜 *JmjC* 基因结构显示, 该家族所有成员均含有内含子, 其中 *CIJMJ6.1* 最短, *CIJMJ5* 最长(图 1: A)。西瓜 *JmjC* 基因家族蛋白共包含 13 种结构域, 其中 *CIJMJ5* 结构域组成最复杂。除 *CIJMJ6.1* 中 *JmjC* 结构域发生部分缺失, 其它成员都含有完整的 *JmjC* 结构域(图 1: B)。该基因家族的不同亚族或亚类之间的基因结构和结构域差异较大, 而同一亚族内的差异较小。推测同一亚族的成员功能相似, 不同亚族间出现了明显的功能分

化。其中 *PKDM7* ~ *PKDM9* 这 3 个亚族是植物中特有的, 且结构相似, 本研究也发现这 3 个亚族的基因结构和蛋白结构域比较接近, 与前人的研究结果一致^[13]。

本研究根据西瓜的基因组注释文件和 *JmjC* 基因家族成员信息, 将 17 个家族成员定位在 9 条染色体上, 但在 4 号和 7 号染色体上并未出现该基因家族成员(图 2)。说明该基因家族成员在染色体上的分布并不均匀, 大部分成员分布在染色体的端部。



A: 西瓜 *JmjC* 家族的进化树和基因结构。B: 西瓜 *JmjC* 家族蛋白结构域分布。

A: Phylogenetic tree and gene structure of *JmjC* family genes in *Citrullus lanatus*. B: Protein domain distribution of *JmjC* family genes in *C. lanatus*.

图 1 西瓜 *JmjC* 基因家族分析

Fig. 1 Analysis of *JmjC*-domain containing gene family in *Citrullus lanatus*

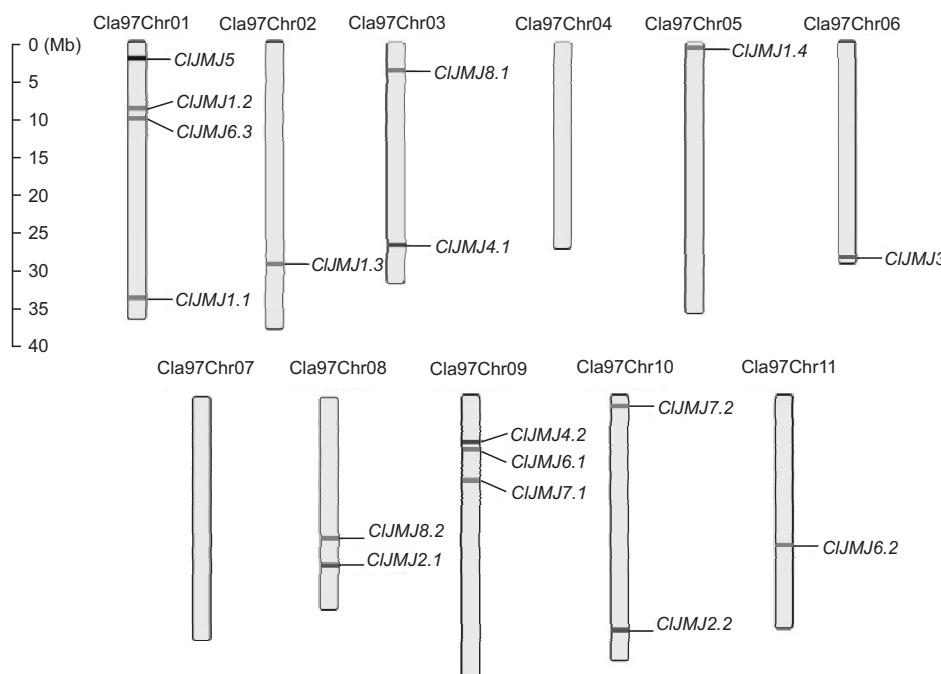


图 2 西瓜 *JmjC* 基因的染色体分布

Fig. 2 Chromosome location of *JmjC*-domain containing genes in *Citrullus lanatus*

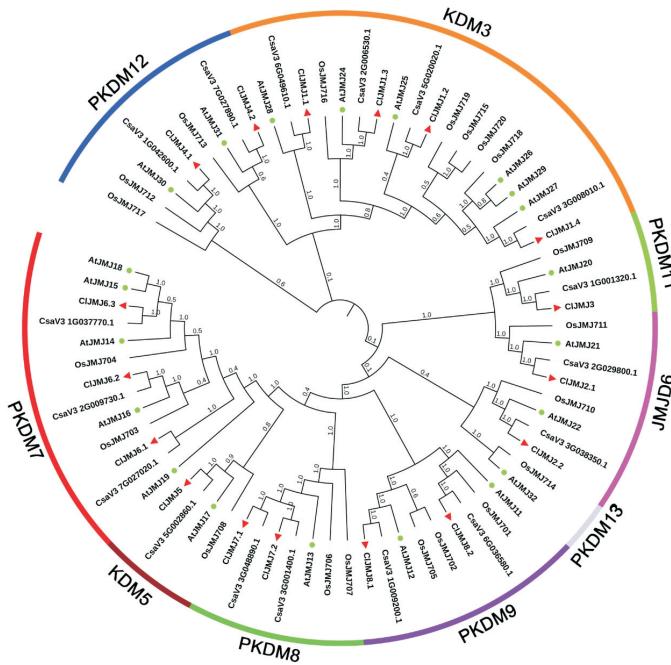
2.2 *JmjC* 基因家族成员的系统进化分析

研究结果表明, 拟南芥中有 21 个 *JmjC* 基因, 水稻中有 20 个^[10]。西瓜与黄瓜同属葫芦科, 亲缘关系较近, 本研究鉴定到黄瓜中有 17 个 *JmjC* 基因。西瓜、黄瓜和模式植物拟南芥、水稻 *JmjC* 基因构建的系统进化树分析结果显示(图 3), 西瓜与黄瓜亲缘关系较近, 拟南芥次之, 西瓜与水稻亲缘关系较远。西瓜 *JmjC* 基因成员相对于拟南芥、水稻的数量略偏少; 17 个成员分布在 8 个亚族中, 与其它被子植物相比, 缺少 PKDM13 亚族成员, KDM3 亚族成员数量也明显偏少(表 2)。

2.3 西瓜和黄瓜、拟南芥 *JmjC* 基因共线性分析

本研究将西瓜的基因组分别与黄瓜和拟南芥的基因组进行共线性分析(图 4), 以 30 个基因设为

一个共线性区块。结果显示, 西瓜和黄瓜亲缘关系近, 共线性区块较多, 而西瓜和拟南芥的共线性区块少, 与前人的研究结果相同^[20]。西瓜与拟南芥中仅有 2 个 *JmjC* 基因家族成员具有共线性关系, 分别为西瓜 1 号染色体上的 *CIJMJ5* 与拟南芥 1 号染色体上的 *AtKDM5*(黑色线条); 西瓜 8 号染色体上的 *CIJMJ8.2* 与拟南芥 5 号染色体上的 *AtPKDM9B*(蓝色线条)。表明西瓜与拟南芥的祖先分化时间较远, 染色体水平的结构变异已经破坏了祖先基因保守的共线性排布关系。西瓜和黄瓜之间的彩色线明显多于西瓜与拟南芥, 除了 *CIJMJ1.2*、*CIJMJ6.3* 和 *CIJMJ4.2* 之外, 其余 14 个 *JmjC* 基因家族成员均与黄瓜有一对一的共线性关系, 表明尽管染色体的对应位置发生改变, 但它们之间仍然



Cl: 西瓜; Csa: 黄瓜; At: 拟南芥; Os: 水稻。

Cl: *Citrullus lanatus*; Csa: *Citrullus sativus*; At: *Arabidopsis thaliana*; Os: *Oryza sativa*.

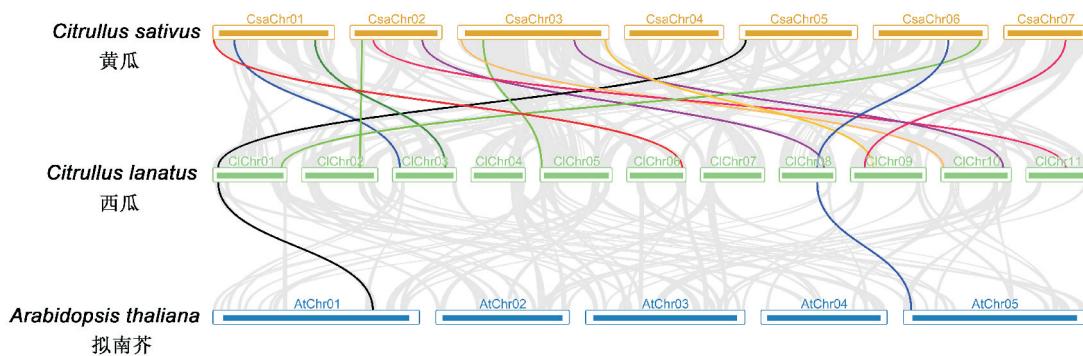
图 3 *JmjC* 蛋白的系统进化分析

Fig. 3 Phylogenetic analysis of *JmjC*-domain containing proteins

表 2 *JmjC* 基因家族各亚族成员分析

Table 2 Subfamily members of *JmjC*-domain containing proteins

| 物种 Species | 总数 All | 各亚组成员数目 Number of each subfamily | | | | | | | | |
|---------------------------------|-----------|----------------------------------|------|--------|-------|--------|-------|-------|------|-------|
| | | PKDM12 | KDM3 | PKDM11 | JMJD6 | PKDM13 | PKDM9 | PKDM8 | KDM5 | PKDM7 |
| 西瓜 <i>Citrullus lanatus</i> | 17 | 2 | 4 | 1 | 2 | 0 | 2 | 2 | 1 | 3 |
| 黄瓜 <i>Citrullus sativus</i> | 17 | 2 | 4 | 1 | 2 | 0 | 2 | 2 | 1 | 3 |
| 拟南芥 <i>Arabidopsis thaliana</i> | 21 | 2 | 6 | 1 | 2 | 1 | 2 | 1 | 1 | 5 |
| 水稻 <i>Oryza sativa</i> | 20 | 3 | 5 | 1 | 2 | 1 | 3 | 2 | 1 | 2 |



灰色线代表西瓜基因组分别于黄瓜、拟南芥基因组之间的共线性区块；彩色线代表共线性区块中 *JmjC* 不同亚族基因的位置(各亚族基因所用的线条颜色与图 3 所用颜色对应)。

Gray lines represent collinear blocks between *Citrullus lanatus* genome and *Cucumis sativus* and *Arabidopsis thaliana* genomes, respectively; colored lines mark positions of different *JmjC* subfamilies in collinear blocks (color used for each subfamily corresponds to color used in Fig. 3).

图 4 *JmjC* 基因家族共线性分析
Fig. 4 Syntenic analysis of *JmjC* genes

具有较好的区块保守性。推测虽然在进化过程中染色体水平的融合、断裂，改变了西瓜和黄瓜基因的相对物理位置，甚至造成了两个物种染色体数目的变化，但家族基因所在位置的共线性排布的方式依然保守。同时西瓜和黄瓜 *JmjC* 基因家族之间的共线性关系中并没有出现一对多或多对一的现象，仅出现一对一的共线性关系，推测西瓜和黄瓜基因组分化之后，没有发生明显的家族基因复制事件。

2.4 西瓜 *JmjC* 基因家族各亚族成员的选择压力分析

编码蛋白质的核苷酸替换有两种：改变氨基酸的非同义替换和不改变氨基酸的同义替换。 K_a 和 K_s 分别用来表示非同义替换率和同义替换率。 K_a/K_s 的比值大小表示序列的进化方向，当比值小于 1 时，表明进化过程中受到纯化选择，等于 1 表明受到中性选择，大于 1 则表示受到正向选择^[21]。依据亚族的分类，对每一个亚族成员及其同一亚族中的拟南芥成员进行同源基因对之间的选择压力计算，将所有基因对之间的 K_a/K_s 绘制成散点图(图 5)。结果发现，各个亚族成员间的 K_a/K_s 比值均小于 1，表明西瓜 *JmjC* 基因家族都受到纯化选择，推测西瓜 *JmjC* 家族基因各个亚族成员的蛋白质功能与同一亚族的拟南芥成员具有极为相似的功能。

2.5 西瓜 *JmjC* 基因家族的关键酶活位点和三级结构预测

JmjC 类组蛋白去甲基化酶以二价亚铁离子(Fe(Ⅱ))和 α -酮戊二酸(α -KG)为辅助因子，其结构域中有 3 个氨基酸残基与 Fe(Ⅱ)结合，2 个

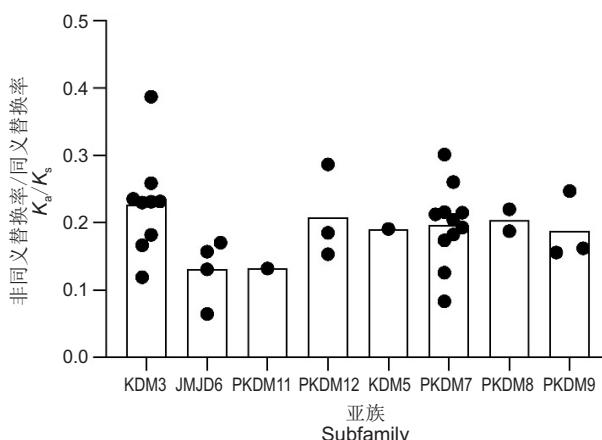


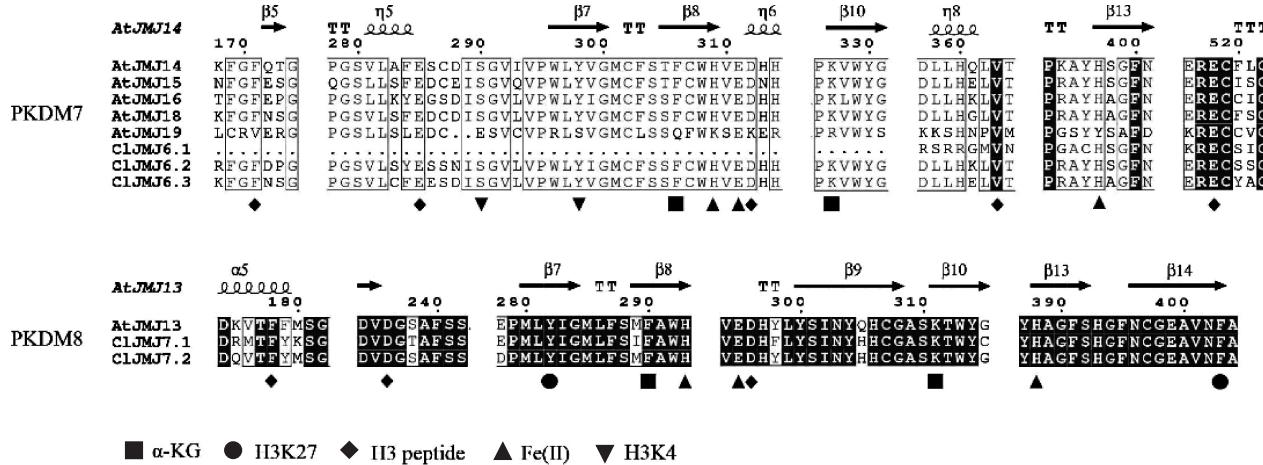
图 5 西瓜 *JmjC* 基因家族各亚族成员 K_a/K_s 分析
Fig. 5 K_a/K_s values of each subfamily member of *JmjC*-domain containing proteins in *Citrullus lanatus*

氨基酸残基与 α -KG 结合，这 5 个氨基酸残基较为保守，在发挥酶活功能中扮演重要角色^[22]。已有研究预测了拟南芥该家族中 15 个成员具有酶活性^[10]。虽然许多哺乳动物 *JmjC* 类家族基因蛋白的晶体已经解析，而植物中只有 3 个晶体信息，分别为 PKDM7 亚族的 OsJMJ703(水稻)、AtJMJ14(拟南芥)和 PKDM8 亚族的 AtJMJ13(拟南芥)^[14, 18]。

通过 PDB 中收录的拟南芥 AtJMJ13 和 AtJMJ14 的晶体信息，并结合酶活实验，本研究标注了西瓜 PKDM7、PKDM8 亚族中同源基因多肽的结合位点(图 6)。分析结果表明，PKDM7 亚族的 CIJM6.1 在关键的酶活位点附近出现了大段的序列缺失。虽然整体上氨基酸替换速率表明受纯化选择的同源基因对之间的蛋白功能应相对保守(图 5)，

但关键位点的缺失极有可能造成 CIJM6.1 的酶活功能丧失。CIJM6.2、CIJM6.3 和 PKDM8 亚族的家族成员 CIJM7.1、CIJM7.2 分别与 AtJM14 和 AtJM13 关键酶活位点完全相同, 因此这 4 个成员可能是具有酶活性的家族基因。JmjC 结构域是 2 个反向平行的 β 折叠片, 每个折叠片由 4 个 β

折叠组成。对 PKDM7 和 PKDM8 亚族成员进行三级结构预测, 发现除了 CIJM6.1 外, 整体上 PKDM7 和 PKDM8 亚族基因具有十分相似的三级结构(图 7)。提示西瓜 PKDM7 和 PKDM8 亚族有着共同的进化起源, 这与基因结构的分析结果一致(图 1)。



关键酶活位点依据序列比对(方形标记), 依据位点突变实验验证(其它图形标记)。下同。

Key enzyme sites based on sequence alignment are shown in square labels, and key enzyme sites based on site mutation experiments are given in other graphs. Same below.

图 6 西瓜 PKDM7 和 PKDM8 亚族成员的酶活关键位点分析

Fig. 6 Key enzyme sites of PKDM7 and PKDM8 subfamily members in *Citrullus lanatus*

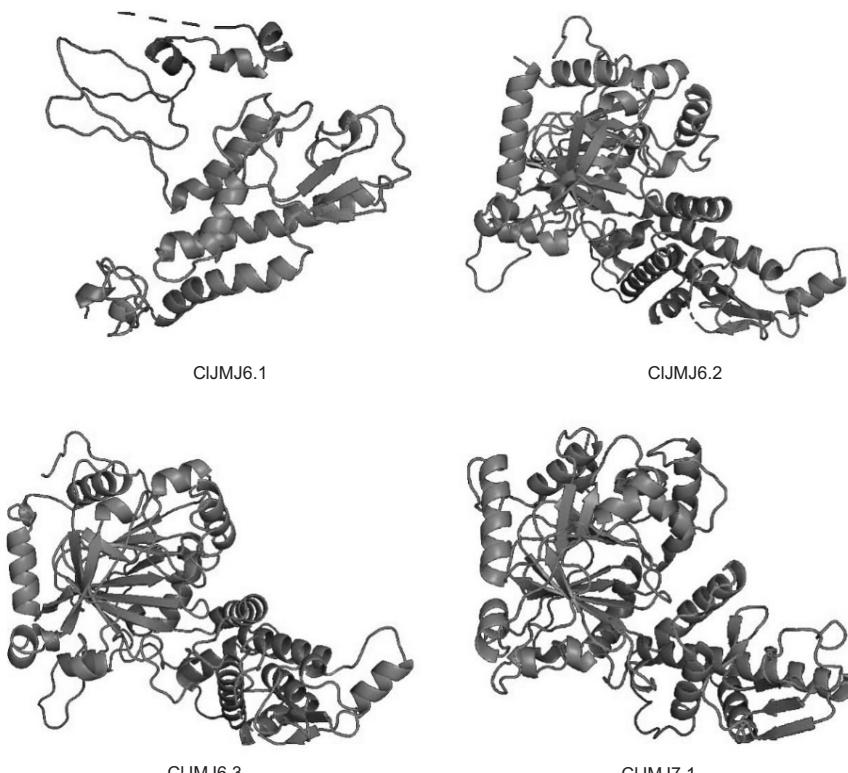
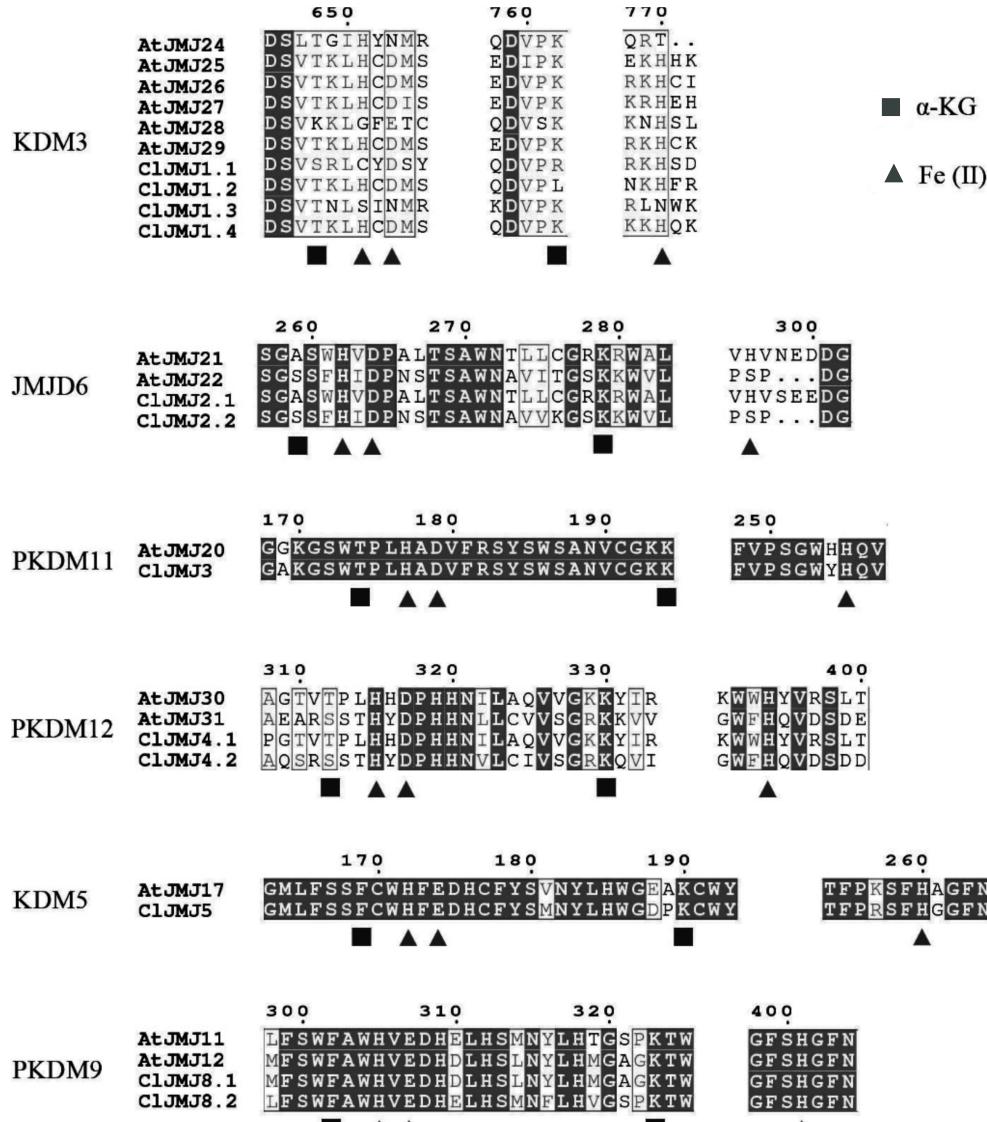


图 7 西瓜 PKDM7 和 PKDM8 亚族成员的三级结构预测

Fig. 7 Tertiary structure prediction of PKDM7 and PKDM8 subfamily members in *Citrullus lanatus*

本研究根据文献中预测的拟南芥该家族其它亚族基因的酶活位点^[10], 推测黄瓜中该家族其它亚族基因是否具有去甲基化酶活性(图8)。在KDM3亚族中, 通常认为该亚族中具有活性的JmjC蛋白第1个Fe(II)结合位点为H, 第2个α-KG结合位点为K, 通过比对, 推断KDM3亚族中CIJMJ1.4可能是有活性的家族基因, 而其它3个成员没有活性。在JMJD6亚族中, 黄瓜与拟南芥的家族成员具有相同的关键氨基酸位点, 但是目前动、植物JMJD6亚族蛋白的催化功能还有待进一步确定^[23, 24], 因此本研究无法预测西瓜中JMJD6

亚族成员的活性。在PKDM12亚族中, CIJMJ4.1、CIJMJ4.2分别和AtJM30、AtJM31的关键酶活位点完全相同。AtJM30可能具有H3K27或H3K36去甲基化的功能, 具体的催化特异性目前存在争议^[7, 25], 而AtJM31不具有酶活性。因此CIJMJ4.1是一个潜在的H3K27或H3K36类型的组蛋白去甲基化酶, 而CIJMJ4.2是一个功能失活的蛋白。在PKDM11、KDM5和PKDM9亚族中, 拟南芥和西瓜的家族成员关键位点均十分保守。PKDM11亚族中AtJM20是潜在的精氨酸的去甲基化酶^[23], KDM5亚族中AtJM17可以去除H3K4位



关键酶活位点依据序列比对。

Key enzyme sites based on sequence alignment.

图8 西瓜KDM3、KDM5、JMJD6、PKDM9、PKDM11和PKDM12亚族酶活关键位点分析

Fig. 8 Key enzyme sites of KDM3, KDM5, JMJD6, PKDM9, PKDM11, and PKDM12 subfamily members in *Citrullus lanatus*

点的甲基基团^[9]，而PKDM9亚族中AtJMJ11/12可以去除H3K27位点的甲基基团^[26, 27]。因此CIJMJ3、CIJMJ5、CIJMJ8.1和CIJMJ8.2都是潜在的具有酶活性的家族成员。

因此，黄瓜中除了CIJMJ1.1、CIJMJ1.2、CIJMJ1.3、CIJMJ4.2、CIJMJ6以及JMJD6亚族两个成员无法预测酶活性，其余10个JmjC家族成员可能具有去甲基化酶活性。

3 讨论

随着生命科学的发展，表观遗传调控在植物的生长发育以及压力应答过程中越来越引起人们的关注。含有JmjC结构域的蛋白是具有去除组蛋白甲基化的一类酶，在植物的生长发育及响应逆境胁迫中起重要作用。植物中关于JmjC基因的研究相对比较滞后，仅在拟南芥、水稻等模式植物中有报道，而西瓜中还没有关于JmjC基因的相关研究，鉴定与分析西瓜JmjC基因家族成员，对其功能研究具有重要意义。

在已经鉴定的JmjC成员的被子植物中，拟南芥有21个^[10]、水稻有20个^[10]、玉米有19个^[11]、月季有28个^[12]。本研究从西瓜品种‘97103’基因组中鉴定了17个JmjC基因，数量与拟南芥、水稻相比略偏少^[10]。拟南芥中已报道部分JmjC基因存在功能冗余现象，因此西瓜可能在进化过程中丢失了部分存在功能冗余的基因。根据系统进化分析结果，可将西瓜17个JmjC基因分为8个亚家族，各亚族相较于拟南芥有不同的扩张和收缩。说明在进化过程中，这些物种间可能发生了不同程度的基因重复或丢失。根据系统进化和选择压力分析结果，本研究推测拟南芥和西瓜直系同源基因间应该具有十分相似的功能，因此拟南芥家族基因的功能研究可以为西瓜家族同源基因的功能预测提供重要参考。

西瓜和黄瓜亲缘关系较近，同属于葫芦科，虽然染色体数目并不相同，但两个物种均有17个JmjC家族成员。根据共线性分析结果，有14个JmjC基因家族成员与黄瓜有一对一的共线性关系，表明西瓜JmjC基因的复制发生在与黄瓜分化之前，在西瓜中没有特异性JmjC基因的复制事件发生。西瓜和黄瓜中都缺失了PKDM13亚族成员，但PKDM13亚族在哺乳动物和植物中普遍存

在^[13]。有文献报道，拟南芥中PKDM13亚族中只有一个成员AtJMJ32，但是其在依赖FLC(Flowering locus C)的开花通路中与PKDM12亚族的AtJMJ30有部分功能冗余^[7]。AtJMJ30和AtJMJ32的功能在于高温下去除FLC位点的H3K27me3修饰，促进FLC表达，以防止拟南芥过早开花^[7]。AtJMJ30和AtJMJ32单个功能缺失不会造成明显的开花表型，但双突变在高温下早花，跟FLC突变体表型类似。因此西瓜中PKDM12亚族成员中的功能会在一定程度上弥补缺失的PKDM13亚族成员的功能。但该亚族成员的缺失是否是葫芦科的普遍现象，是否跟葫芦科植物某些独特性状的进化有关，是否发生在葫芦科的物种分化之前，需要未来更加深入的研究。

植物体内组蛋白甲基化水平由组蛋白甲基化转移酶和去甲基化酶共同调节，而含有JmjC结构域的去甲基酶在维持体内组蛋白甲基化的稳态中起关键作用，因此鉴定该家族成员是否具有去甲基化酶活性具有重要意义。JmjC类组蛋白去甲基化酶的活性依赖JmjC结构域的完整性，除CIJMJ6.1有部分JmjC结构域缺失，其余基本都含有完整的JmjC结构域。而JmjC类组蛋白去甲基化酶以Fe(Ⅱ)和α-KG为辅助因子，其结构域中的重要氨基酸的结合位点对发挥去甲基化酶活性至关重要。本研究依据拟南芥、水稻和人类序列比对的结果，初步确定了10个可能具有组蛋白赖氨酸去甲基化酶活性的JmjC蛋白，具体的酶活性以及功能有待进一步研究。

参考文献：

- [1] Pikaard CS, Mittelsten Scheid O. Epigenetic regulation in plants[J]. *Cold Spring Harb Perspect Biol*, 2014, 6(12): a019315.
- [2] Berr A, Shafiq S, Shen WH. Histone modifications in transcriptional activation during plant development[J]. *Biochim Biophys Acta*, 2011, 1809(10): 567–576.
- [3] Zheng B, Chen X. Dynamics of histone H3 lysine 27 trimethylation in plant development[J]. *Curr Opin Plant Biol*, 2011, 14(2): 123–129.
- [4] Shi Y, Lan F, Matson C, Mulligan P, Whetstone JR, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1[J]. *Cell*, 2004, 119(7): 941–953.
- [5] Accari SL, Fisher PR. Emerging roles of JmjC domain-containing proteins[J]. *Int Rev Cell Mol Biol*, 2015, 319:

- 165–220.
- [6] Wu J, Yamaguchi N, Ito T. Histone demethylases control root elongation in response to stress-signaling hormone abscisic acid [J]. *Plant Signal Behav*, 2019, 14 (7) : 1604019.
- [7] Gan ES, Xu Y, Wong JY, Goh JG, Sun B, et al. Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of FLC in *Arabidopsis* [J]. *Nat Commun*, 2014, 5: 5098.
- [8] Liu P, Zhang S, Zhou B, Luo X, Zhou XF, et al. The Histone H3K4 demethylase JMJ16 represses leaf senescence in *Arabidopsis* [J]. *Plant Cell*, 2019, 31 (2) : 430–443.
- [9] Huang S, Zhang A, Jin JB, Zhao B, Wang TJ, et al. *Arabidopsis* histone H3K4 demethylase JMJ17 functions in dehydration stress response [J]. *New Phytol*, 2019, 223 (3) : 1372–1387.
- [10] Lu F, Li G, Cui X, Liu C, Wang XJ, et al. Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in *Arabidopsis* and rice [J]. *J Integr Plant Biol*, 2008, 50 (7) : 886–896.
- [11] Qian Y, Chen C, Jiang L, Zhang J, Ren Q. Genome-wide identification, classification and expression analysis of the JmjC domain-containing histone demethylase gene family in maize [J]. *BMC Genomics*, 2019, 20 (1) : 256–256.
- [12] Dong Y, Lu J, Liu J, Jalal A, Wang C. Genome-wide identification and functional analysis of JmjC domain-containing genes in flower development of *Rosa chinensis* [J]. *Plant Mol Biol*, 2020, 102 (4–5) : 417–430.
- [13] Qian S, Wang Y, Ma H, Zhang L. Expansion and functional divergence of Jumonji C-containing histone demethylases: Significance of duplications in ancestral angiosperms and vertebrates [J]. *Plant Physiol*, 2015, 168 (4) : 1321–1337.
- [14] Yang Z, Qiu Q, Chen W, Jia B, Chen X, et al. Structure of the *Arabidopsis* JMJ14-H3K4me3 complex provides insight into the substrate specificity of KDM5 subfamily histone demethylases [J]. *Plant Cell*, 2018, 30 (1) : 167–177.
- [15] Subburaj S, Lee K, Jeon Y, Tu L, Son G, et al. Whole genome resequencing of watermelons to identify single nucleotide polymorphisms related to flesh color and lycopene content [J]. *PLoS One*, 2019, 14 (10) : e0223441.
- [16] Guo S, Zhao S, Sun H, Wang X, Wu S, et al. Rese-
- quencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits [J]. *Nat Genet*, 2019, 51 (11) : 1616–1623.
- [17] Wang Y, Tang H, Debarry JD, Tan X, Li J, et al. MCS-canX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity [J]. *Nucleic Acids Res*, 2012, 40 (7) : e49.
- [18] Zheng S, Hu H, Ren H, Yang Z, Qiu Q, et al. The *Arabidopsis* H3K27me3 demethylase JUMONJI 13 is a temperature and photoperiod dependent flowering repressor [J]. *Nat Commun*, 2019, 10 (1) : 1303–1303.
- [19] Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis [J]. *Nat Protoc*, 2015, 10 (6) : 845–858.
- [20] Guo S, Zhang J, Sun H, Salse J, Lucas WJ, et al. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions [J]. *Nat Genet*, 2013, 45 (1) : 51–58.
- [21] Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, et al. KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging [J]. *Genom Proteom Bioinf*, 2006, 4 (4) : 259–263.
- [22] Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation [J]. *Nat Rev Genet*, 2006, 7 (9) : 715–727.
- [23] Cho JN, Ryu JY, Jeong YM, Park J, Song JJ, et al. Control of seed germination by light-induced histone arginine demethylation activity [J]. *Dev Cell*, 2012, 22 (4) : 736–748.
- [24] Blanc RS, Richard S. Arginine methylation: the coming of age [J]. *Mol Cell*, 2017, 65 (1) : 8–24.
- [25] Yan Y, Shen L, Chen Y, Bao S, Thong Z, et al. A MYB-domain protein EFM mediates flowering responses to environmental cues in *Arabidopsis* [J]. *Dev Cell*, 2014, 30 (4) : 437–748.
- [26] Lu F, Cui X, Zhang S, Jenuwein T, Cao X. *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase [J]. *Nat Genet*, 2011, 43 (7) : 715–719.
- [27] Crevillén P, Yang H, Cui X, Greeff C, Trick M, et al. Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state [J]. *Nature*, 2014, 515 (7528) : 587–590.

(责任编辑: 周媛)