

· 研究报告 ·

甘蓝型油菜MADS-box基因家族的鉴定与系统进化分析

高虎虎[†], 张云霄[†], 胡胜武, 郭媛^{*}

西北农林科技大学农学院, 旱区作物逆境生物学国家重点实验室, 杨凌 712100

摘要 MADS-box基因家族参与调控开花时间、花器官分化、根系生长、分生组织分化、子房和配子发育、果实膨大及衰老等植物生长发育的重要过程。基于甘蓝型油菜(*Brassica napus*)基因组测序数据, 利用生物信息学方法对甘蓝型油菜MADS-box基因家族进行鉴定和注释及基因结构与系统进化分析。结果显示, 在甘蓝型油菜中鉴定出307个MADS-box基因家族成员, 根据进化关系可将其分为两大类型, I型(M-type)包含 α 、 β 、 γ 三个亚家族, II型(MIKC-type)包括MIKC^C和MIKC^{*}两个亚家族, MIKC^C可进一步分为13个小类; 甘蓝型油菜A基因组染色体上分布的MADS-box基因多于C基因组。在基因结构上, MIKC-type亚家族基因序列普遍比M-type长且含有较多的外显子; M-type亚家族蛋白序列中的motif数量为2–5个, MIKC-type亚家族蛋白序列中平均含有7个motif。拟南芥(*Arabidopsis thaliana*)与甘蓝型油菜MADS-box基因共线性分析结果显示, 全基因组复制事件对MADS-box基因家族尤其是MIKC亚家族的扩张起重要作用; MIKC亚家族基因在进化过程中受到的选择压力约为M-type的2倍, 这表明MIKC-type亚家族在进化过程中被选择性保留。

关键词 甘蓝型油菜, MADS-box, 基因结构, 进化, 共线性

高虎虎, 张云霄, 胡胜武, 郭媛 (2017). 甘蓝型油菜MADS-box基因家族的鉴定与系统进化分析. 植物学报 52, 699–712.

MADS-box转录因子广泛存在于酵母、植物、昆虫、线虫、低等脊椎动物和哺乳动物等真核生物体内, 其命名来自酵母(*Saccharomyces cerevisiae*) *MCMI*基因、拟南芥(*Arabidopsis thaliana*) *AGAMOUS*基因、金鱼草(*Antirrhinum majus*) *DEFTCI*基因及人类(*Homo sapiens*) *SRF*基因的首字母缩写(Purugganan et al., 1995; Shore and Sharrocks, 1995)。MADS-box家族基因的典型特征是N端具有60个氨基酸的MADS-box结构域, 这一结构域是参与转录激活的重要组件。依据基因所含有结构域的类型可以将MADS-box家族基因分为两类, I型(M-type)包含 α 、 β 和 γ 三个亚家族, II型(MIKC-type)包含MIKC^C和MIKC^{*}亚家族。MIKC^C亚家族除了具有典型MADS结构域外, 还有K(Keratin-like)、I(Intervening)及C-terminal结构域(Gramzow et al., 2010)。这2个亚家族在基因结构上也有区别, M-type基因序列较短, 含有较少的外显子; MIKC-type基因一般有5–8个外显子, 序列较长(Alvarez-Buylla et al., 2000b)。

在植物中, MADS-box是一个庞大的转录调控因子家族, 参与许多生长发育的重要过程。例如, 花的形成、开花、果实膨大及衰老等(Alvarez-Buylla et al., 2000a; Theissen et al., 2000; Fang and Fernandez, 2002)。植物MADS-box转录因子的发现源于对拟南芥花器官分化的研究(Yanofsky et al., 1990)。拟南芥花器官分化ABC模型中的大部分基因都属于MADS-box家族的MIKC-type亚家族(Immink et al., 2010)。MIKC-type亚家族基因还参与许多生物和非生物胁迫的响应(Gan et al., 2005; Shao et al., 2010; Hemming and Trevaskis, 2011; Yao et al., 2015)。在拟南芥中, MIKC-type亚家族基因*AGL21*调控侧根的生长, 响应环境压力并受到多种植物激素的调控(Yu et al., 2014)。水稻(*Oryza sativa*) MIKC-type亚家族的OsMADS25和OsMADS27参与渗透胁迫响应, OsMADS25、OsMADS27和OsMADS57受硝酸盐诱导表达(Arora et al., 2007)。此外, 研究表明MIKC-type亚家族的MIKC^{*}类基因在水稻等单子叶植物花粉发

收稿日期: 2016-12-12; 接受日期: 2017-05-22

基金项目: 西北农林科技大学博士科研启动资金(No.Z109021614)、基本科研业务专项资金(No.Z109021703)和唐仲英作物育种基金(No.A212021713)

[†] 共同第一作者

* 通讯作者。E-mail: guoyuan2109@163.com

育的晚期起重要作用(Liu et al., 2013)。相比于MIKC-type亚家族,对于M-type基因的研究相对较少。与动物中相比,植物中M-type基因的产生和丢失远多于动物,因此早期的研究认为在植物中M-type基因的功能可能没有MIKC-type重要(De Bodt et al., 2003)。通过对拟南芥61个M-type基因的组织表达模式进行研究,发现只有5个基因表达(Kofuji et al., 2003)。但近年相继发现M-type基因与植物配子分化、胚和胚乳的发育相关(Day et al., 2008; Tiwari et al., 2010; Wuest et al., 2010; Masiero et al., 2011)。

随着二代测序技术的成熟,已完成测序的物种越来越多。国内外学者已经对拟南芥(Parenicová et al., 2003)、水稻(Arora et al., 2007)、大豆(*Glycine max*) (Shu et al., 2013)、芝麻(*Sesamum indicum*) (Wei et al., 2015)、二穗短柄草(*Brachypodium distachyrum*) (Wei et al., 2014)、葡萄(*Vitis vinifera*) (Díaz- Riquelme et al., 2009)和梅(*Armeniaca mume*) (Xu et al., 2014)等物种中的MADS-box基因家族进行了探究。利用生物信息学方法,在模式植物拟南芥中鉴定了107个MADS-box家族成员并将其分为5个亚家族: Ma (25 proteins)、M β (20 proteins)、M γ (16 proteins)、M δ (7 proteins)和MIKC (39 proteins)。目前已在水稻中鉴定了72个MADS-box基因,分为Ma (13 proteins)、M β (10 proteins)、M γ (9 proteins)、MIKC^C (38 proteins)和MIKC* (5 proteins)五个亚家族(Parenicová et al., 2003; Arora et al., 2007)。在基因组水平上用生物信息学手段鉴定MADS-box基因家族,可为深入研究MADS-box基因的生物学功能提供重要基础。

十字花科模式植物拟南芥和白菜(*Brassica rapa*)的MADS-box基因已被鉴定(Parenicová et al., 2003; Duan et al., 2015; Saha et al., 2015)。现有的研究表明,十字花科尤其是芸薹属植物的MADS-box基因与花器官分化(Fang and Fernandez, 2002)、开花时间(Chang et al., 2009; Greenup et al., 2009)及根尖分生组织分化(Tapia-López et al., 2008)有关。此外,在

低温和高温胁迫及赤霉素(GA)与水杨酸(SA)处理下,这些基因的表达量也发生显著变化(Duan et al., 2015; Saha et al., 2015)。甘蓝型油菜(*Brassica napus*)同属十字花科,其基因组测序结果已于2014年公布(Chalhoub et al., 2014)。本研究基于基因组测序结果,对甘蓝型油菜MADS-box基因家族进行鉴定,并利用一系列生物信息学软件对其进行注释和基本功能预测分析,初步探究其进化关系。研究结果将为深入揭示MADS-box基因家族的生物学功能奠定重要基础。

1 材料与方法

1.1 甘蓝型油菜MADS-box基因家族的鉴定和注释

甘蓝型油菜(*Brassica napus* L.)基因组序列、CDS序列、蛋白质序列及其注释信息均下载自BRAD (Cheng et al., 2011)数据库(<http://brassicadb.org>)。在Pfam (Finn et al., 2014)数据库(<http://pfam.xfam.org/>)中下载编号为PF00319的结构域数据。

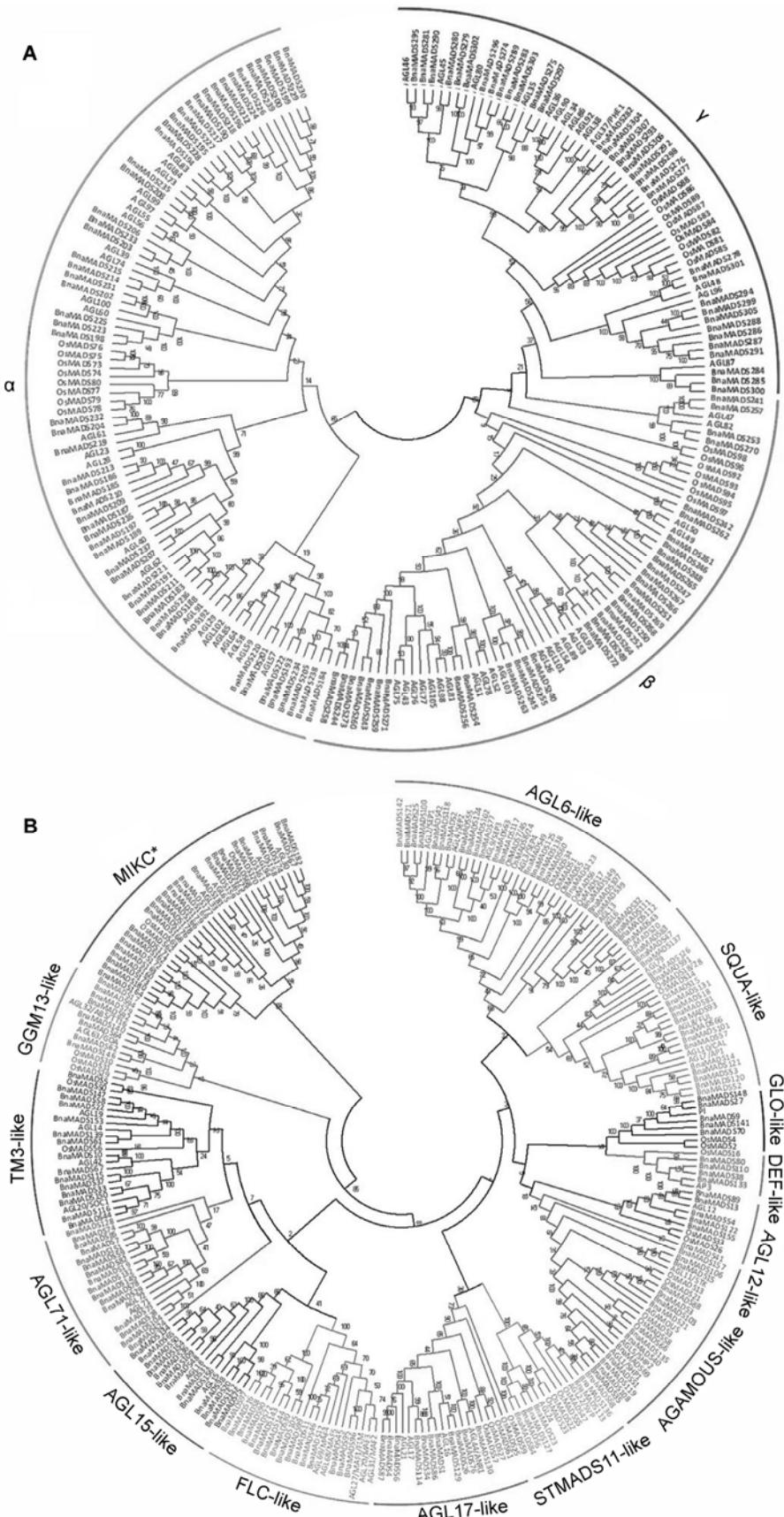
以Pfam下载的结构域数据为种子、甘蓝型油菜的蛋白组数据为库,进行本地BlastP检索,e值设为1e-3。在Excel 2013软件中对Blast结果进行初步整理,去除重复,获得候选基因。利用在线软件InterProScan (Mitchell et al., 2015) (<http://www.ebi.ac.uk/interpro/>)和Smart (Letunic et al., 2015) (<http://smart.embl-heidelberg.de/>)对候选基因的蛋白序列进行预测分析,确定其是否含有MADS-box结构域。通过在线工具ExPasy对含有MADS-box结构域的基因进行蛋白理化性质预测。

1.2 MADS-box基因家族的分类与进化树构建

从植物转录因子网站PlantTFDB (Jin et al., 2014) (<http://planttfdb.cbi.pku.edu.cn/>)下载拟南芥全部MADS-box家族成员蛋白序列。通过水稻基因组注释网站(<http://rice.plantbiology.msu.edu/>)下载得到水

图1 甘蓝型油菜、拟南芥和水稻的MADS-box基因系统发育分析
(A) M-type; (B) MIKC-type

Figure 1 Phylogenetic analysis of rapeseed, *Arabidopsis* and rice MADS-box genes
(A) M-type; (B) MIKC-type



稻MADS-box蛋白序列(Kawahara et al., 2013)。

用Perl脚本从BRAD网站下载的蛋白数据中提取鉴定属于MADS-box基因家族所有成员的蛋白序列。采用Clustal X1.8软件对蛋白序列进行多重序列比对, 比对结果用MEGA6 (Tamura et al., 2013)进行系统进化分析。用邻接法(neighbor joining, NJ)构建系统进化树; 遗传距离为泊松距离(poisson correction); 抽样次数为1 000 (bootstrap: 1 000 replications)。

1.3 MADS-box基因家族的染色体定位与基因结构分析

采用InterProScan软件鉴定具有典型MADS-box结构域的基因, 以基因编号在BRAD中检索其在染色体上的位置, 统计位置信息。在此基础上, 利用Mapchart软件绘制基因在染色体上的分布图。同时利用在线软件MEME对甘蓝型油菜MADS-box序列的蛋白保守基序(motif)进行预测和分析。MEME参数设置: motif数量为10, motif长度为6–60。

1.4 甘蓝型油菜与拟南芥共线性基因与选择压力分析

通过BRAD在线检索获取拟南芥MADS-box基因在甘蓝型油菜中的共线性基因及其在染色体上的位置。用Circos (Krzywinski et al., 2009)软件将这些共线性基因的关系进行展示。用Excel 2013软件对共线性基因在染色体上的分布进行统计并作图。

从BRAD网站上下载共线性基因的CDS序列。用MEGA6 (Tamura et al., 2013)内置的Muscle对两物种之间的共线性基因对做codon比对, 然后用KaKs_Calculator2.0 (Zhang et al., 2006)软件计算共线性基因之间的选择压力。

2 结果与讨论

2.1 甘蓝型油菜MADS-box基因家族的鉴定与注释

为鉴定甘蓝型油菜中的MADS-box基因, 用BlastP程序($E\text{-value} < 10^{-3}$)在其蛋白数据中进行检索与识别, 共获得313个候选基因。将候选基因的蛋白序列提交至Interpro和Smart分别进行注释, 取二者注释结果的交集, 结果显示, 有307个基因具有MADS-box结构域。因此, 我们在甘蓝型油菜中鉴定出307个MADS-

box家族基因。根据这些基因在染色体上的位置和家族分类, 将其命名为BnaMADS1–BnaMADS307(附录1)。将这些基因的蛋白序列提交至在线软件ExPasy, 计算蛋白质等电点, 显示所有MADS-box蛋白等电点位于4–11之间, 平均值为7, MIKC-type基因的等电点略大于M-type。M-type氨基酸序列长度普遍小于MIKC-type, 但MIKC-type基因外显子个数多于M-type。MIKC-type亚家族基因数目约为M-type的1.5倍。

2.2 MADS-box基因家族的分类与系统发育

我们在PlantTFDB网站上对307个MADS-box基因进行初步预测后, 将其分为MIKC-type和M-type两个类型。参考水稻和其它物种MADS-box基因家族的分类结果(Arora et al., 2007), 我们将AGL33-like归为MIKC*。通过对分类结果进行整理, 最终将307个MADS-box基因分为两大类, 即包含125个MADS-box基因的M-type和包含182个基因的MIKC-type。

为探究甘蓝型油菜MADS-box基因与其它模式物种的进化关系, 我们分析了拟南芥M-type (59个)、水稻M-type (24个)和甘蓝型油菜M-type (125个)蛋白序列, 并构建系统进化树(图1A); 同时还建立拟南芥MIKC-type (42个)、水稻MIKC-type (38个)、甘蓝型油菜MIKC-type (182个)基因进化树(图1B)。依据3个物种基因进化树, 我们将甘蓝型油菜M-type分为 α (57)、 β (34)和 γ (34)三类; MIKC-type分为MIKC^C (157)和MIKC* (25)两类。MIKC^C可进一步分为13个小类: GGM13-like (8)、TM3-like (15)、AGL71-like (16)、AGL15-like (13)、FLC-like (16)、AGL17-like (11)、STMADS11-like (7)、AGAMOUS-like (15)、AGL12-like (5)、DEF-like (4)、GLO-like (5)、SQUA-like (17)和AGL6-like (25)。

2.3 MADS-box基因家族的染色体位置与基因结构

我们进一步依据从BRAD获取的染色体位置信息, 利用Mapchart软件绘制了241个MADS-box基因在甘蓝型油菜19条染色体上的位置示意图(图2), 并对其在染色体上的分布信息进行了统计(图3; 表1)。结果显示, MADS-box家族基因在19条染色体上都有分布, 且数量不一。甘蓝型油菜A组染色体上分布的MADS-box基因数目不论是总数还是2个小类下的数目都略多于C组。就单条染色体而言, A9染色体上分布22个MADS-box

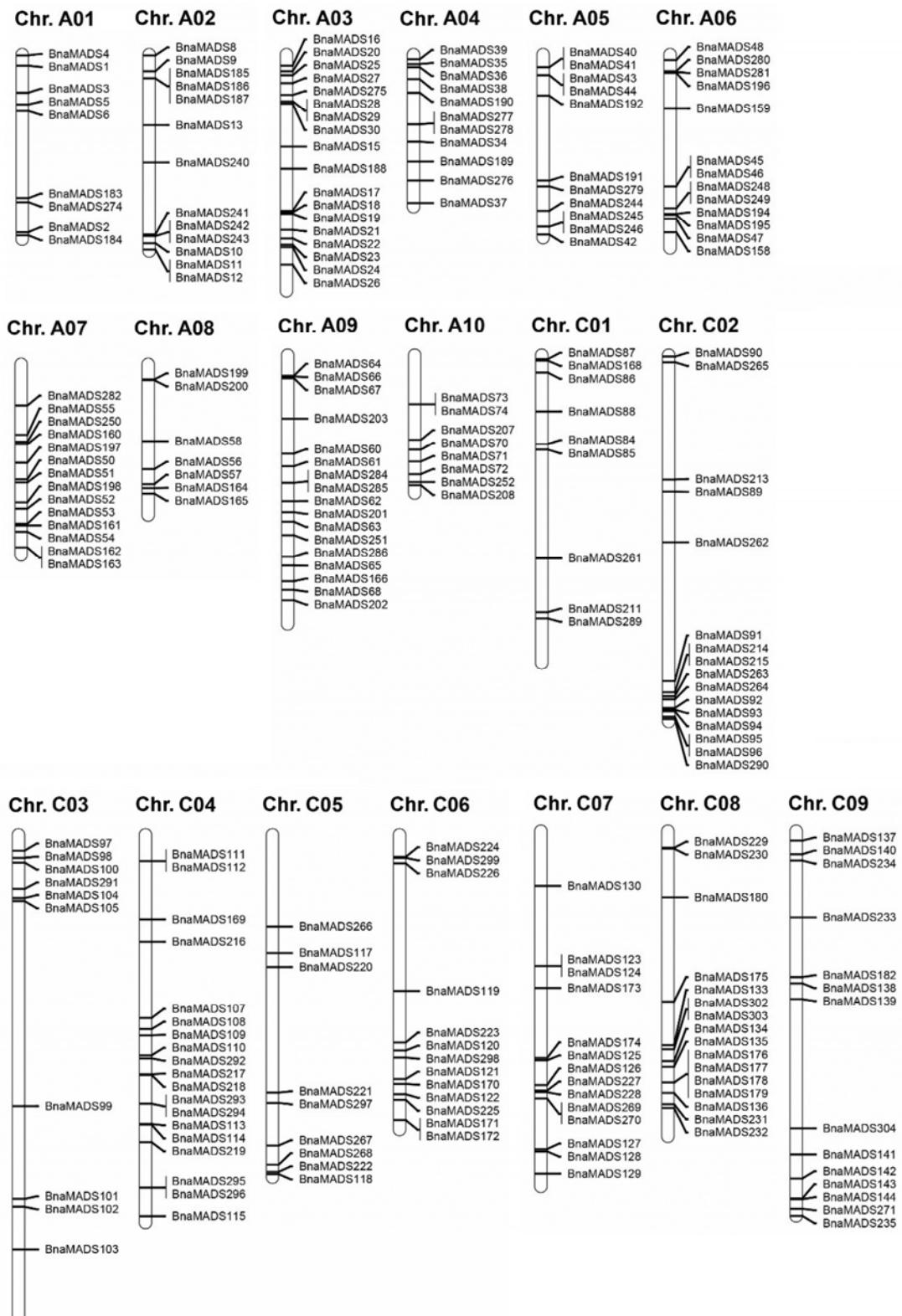


图2 甘蓝型油菜MADS-box基因在染色体上的分布

Figure 2 Chromosomal location of MADS-box genes in *Brassica napus*

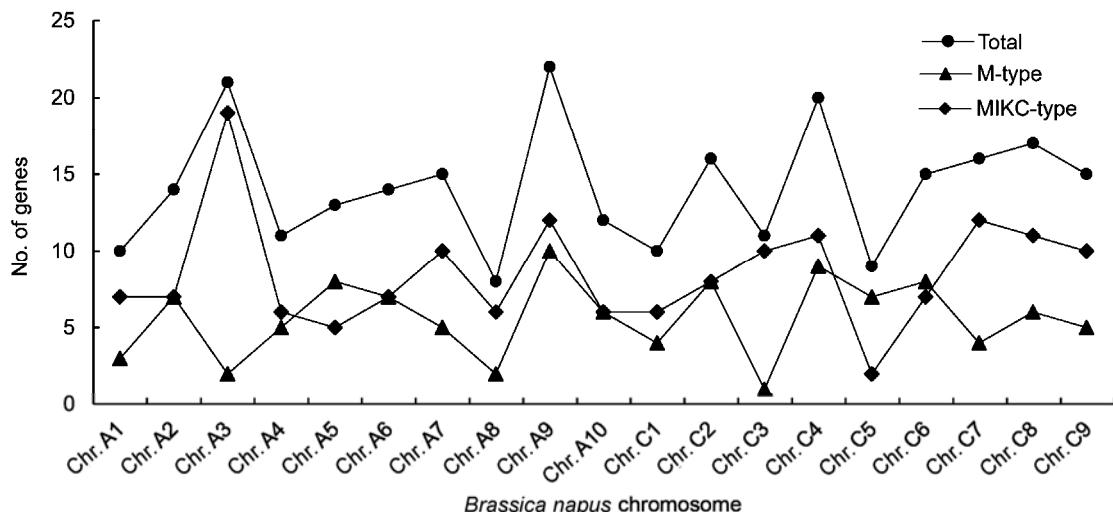


图3 甘蓝型油菜MADS-box基因在染色体上分布的统计

Figure 3 Statistics of MADS-box genes in chromosome of *Brassica napus*

表1 甘蓝型油菜MADS-box基因在亚基因组上的分布

Table 1 Distribution of MADS-box genes in subgenome of *Brassica napus*

Subgenome	M-type	MIKC-type	Total
A	55	85	140
C	52	77	129

基因，是所有染色体中最多的。此外，在A3、A7、C2、C4、C6、C7、C8和C9八条染色体上分布的MADS-box基因数都大于15。

为明确MADS-box基因编码蛋白的保守结构，我们将307个MADS-box基因分为3个类型：MIKC^C、MIKC*和M-type，并预测了保守基序(motif)结构(图4)。总体来说，同一亚家族所含有的motif数量和类型都极为相似。M-type型MADS-box基因具有2–5个保守motif，其中 α 和 β 亚家族具有3–5个motif， γ 亚家族多数具有2–3个motif。在预测到的10个motif中，以Smart软件进行注释发现motif 1和motif 3是MADS-box结构域，在所有的MADS-box蛋白中都出现，motif 10是C末端的卷曲结构，motif 2、4、5、6、7、8和9是一些非保守结构，只在部分MADS-box蛋白中出现。

MIKC类蛋白序列的MEME预测结果显示，MIKC-type亚家族明显含有较多的motif，其中MIKC*多于MIKC^C。在MIKC^C预测到的10个motif中，motif 1、3和5是MADS-box结构域，motif 2和4是K-box，motif 6

为卷曲结构，motif 7、8、9和10尚属未知。MIKC*的蛋白序列中明显含有更多的motif，平均为7个。利用Smart软件对10个motif进行预测，motif 1和4为MADS结构域，motif 2和5是卷曲结构。而Motif 3、6、7、8、9和10功能未知。

2.4 甘蓝型油菜与拟南芥MADS-box基因的共线性分析

我们利用BRAD网站Syntenic工具获取了拟南芥所有MADS-box基因在甘蓝型油菜中的共线性基因，并对这些共线性关系进行了统计和绘图(图5, 图6)。结果显示，拟南芥与甘蓝型油菜的共线性基因对在各条染色体间数量不一，其中拟南芥1号染色体与甘蓝型油菜C5和A7两条染色体间的共线性基因较多；拟南芥5号染色体与甘蓝型油菜多条染色体间共线基因较多，主要有C2、C9和A3。对单条染色体上共线性基因分布而言，拟南芥1、3和5这3条染色体的共线性基因数量大约是其它2条的2倍，甘蓝型油菜中A3、A9、C2和C9上共线性基因的数量多于其它染色体。

此外，我们还统计了拟南芥MADS-box基因在甘蓝型油菜中的拷贝数(图7)，发现80%的拟南芥MADS-box基因在甘蓝型油菜中有2个以上拷贝。1个MADS-box基因在甘蓝型油菜中最多有6个拷贝。因此我们认为MADS-box基因在芸薹属的多倍化事件中得

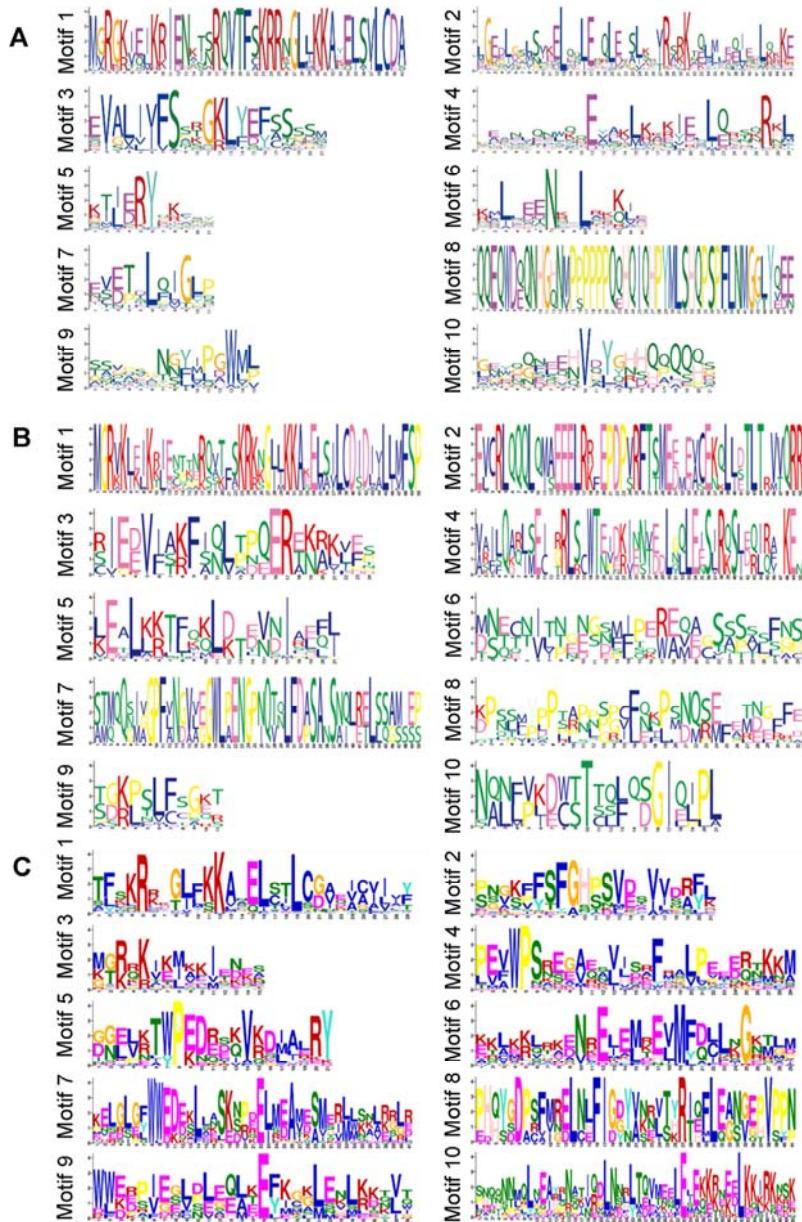


图4 甘蓝型油菜MADS-box基因家族的motif结构
(A) MIKC^c亚家族; (B) MIKCs*亚家族;(C) M-type亚家族

Figure 4 Motif patterns of different MADS-box gene families in *Brassica napus*

(A) MIKC^c subfamily; (B) MIKCs* subfamily; (C) M-type subfamily

到了扩张。

2.5 甘蓝型油菜MADS-box基因进化的选择压力与功能分化

在遗传学中, Ka/Ks 或 dN/dS 表示异义替换(Ka)和同义替换(Ks)之间的比例。根据 Ka/Ks 比值可判断是否

有选择压力作用于编码这个蛋白质的基因。同义突变不受自然选择作用, 而非同义突变则受到自然选择的作用。在进化分析中, 了解同义突变和非同义突变发生的速率具有重要意义。如果 $Ka/Ks > 1$, 认为有正选择效应; 如果 $Ka/Ks = 1$, 认为存在中性选择; 如果 $Ka/Ks < 1$, 则认为有纯化选择作用。我们以拟南芥

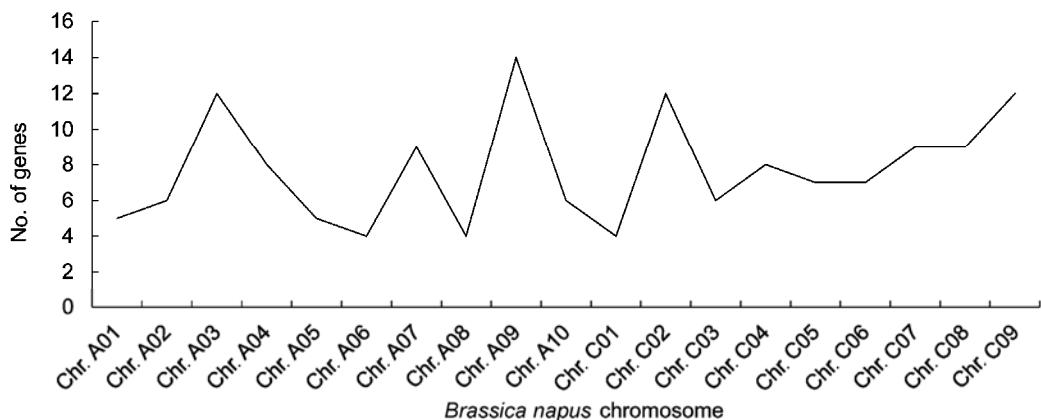


图5 与拟南芥共线性的MADS-box基因在甘蓝型油菜染色体上的分布

Figure 5 Distribution of MADS-box genes synteny with *Arabidopsis thaliana* in *Brassica napus*

MADS-box基因为参照估算了在进化过程中甘蓝型油菜MADS-box基因家族的选择压力(附录2, 表2)。结果显示, MADS-box基因在进化过程中的选择压力均小于1, 即受到纯化选择。有意思的是, 我们发现拟南芥MADS-box基因在油菜中的共线性基因并不注释为MADS-box。我们查阅其Interpro注释, 初步可分为2类: 丢失MADS-box结构域的(只具有K-box结构域)蛋白和Interpro没有注释、功能未知的蛋白。此外, 我们对甘蓝型油菜MADS-box基因家族内M-type和MIKC-type两个类型的拷贝数及选择压力分别进行统计, 结果显示, 2个类型MADS-box基因受到的平均选择压力都小于1, 但MIKC-type受到的选择压力约为M-type的2倍。同时MIKC-type的拷贝数明显多于M-type, 这表明MIKC-type在进化中更为保守。

2.6 讨论

甘蓝型油菜是由白菜和甘蓝(*B. oleracea*)天然杂交而来(Nagaharu, 1935), 是十字花科重要的油料作物。MADS-box家族基因是花期和花器官分化的主要调控因子(Fang and Fernandez, 2002; Chang et al., 2009; Greenup et al., 2009), 开花时间与花的分化是甘蓝型油菜育种和生产中重要的农艺性状, 直接决定其产量。因此, 深入研究MADS-box家族基因对于甘蓝型油菜遗传改良具有重要意义。前人利用生物信息学方法已在十字花科拟南芥中鉴定出107个MADS-box家族基因(Parenicová et al., 2003), 在白菜中鉴定出167个MADS-box家族基因(Saha et al., 2015)。

采用生物信息学方法, 本研究在甘蓝型油菜中鉴定出307个MADS-box基因, 是目前已报道物种中数目最多的。我们参考其它模式物种中的分类, 根据系统发育关系将这307个基因分为2类: M-type和MIKC-type。依据系统发育树(图1), M-type共包括125个基因, 可进一步分为 α 、 β 、 γ 三个类型; MIKC-type包含182个基因, 可分为MIKC*和MIKC^C两类, MIKC^C可进一步分为13个小类。白菜中 $AGL71$ -like类归于TM3-like, 我们发现甘蓝型油菜MIKC-type基因的进化树中 $AGL71$ -like并不能与TM-like归为一类。此外, 甘蓝型油菜M-type和MIKC-type MADS-box基因数目都是目前已报道物种中最多的, 且每一个小分类的基因数目也与拟南芥和水稻中的数目不同。这表明重复基因在不同物种进化过程中的保留情况不同, 因此这些不同物种中同一分类的MADS-box基因在进化过程中受到的约束不同(Nam et al., 2003; Airolidi and Davies, 2012)。

通常来说, 起源于同一拷贝的转录因子拥有类似的基因结构(Doebley and Lukens, 1998)。本研究表明, 甘蓝型油菜M-type基因通常具有较少的外显子, 序列长度也相对较短; MIKC-type基因序列普遍较长且含有较多的外显子。此外, 将307个MADS-box分为M、MIKC*和MIKC^C进行保守序列预测, 结果显示, 在进化树上位于同一个小类的MADS-box蛋白所含有的motif类型几乎完全一致。以上表明这些基因在甘蓝型油菜中很可能行使类似的功能。MIKC*与MIKC^C基因结构相似, 含有较多的motif和内含子, 但是缺少

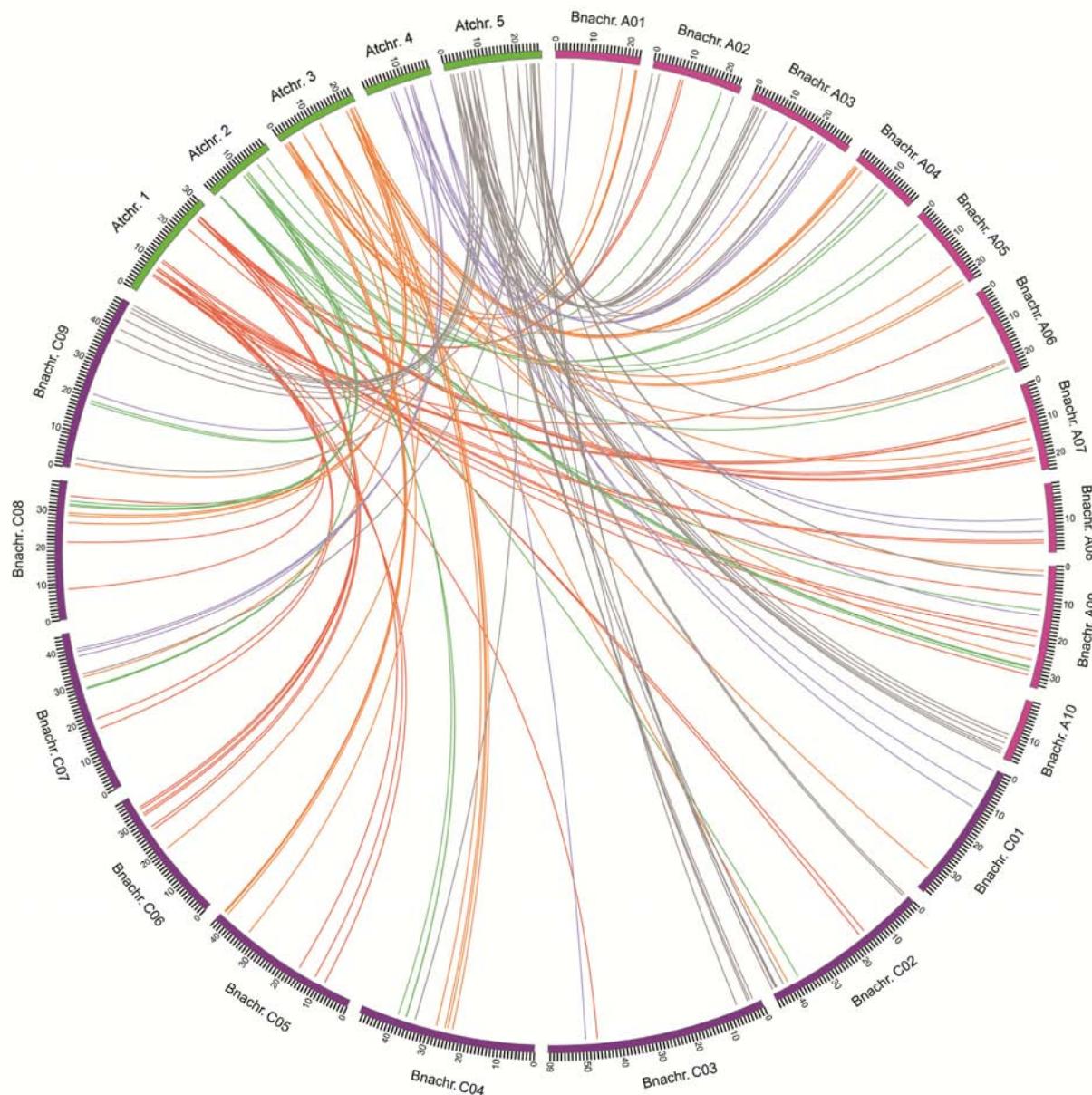


图6 甘蓝型油菜与拟南芥MADS-box基因的共线性关系

Figure 6 Syntenic relationship of MADS-box genes between *Arabidopsis thaliana* and *Brassica napus*

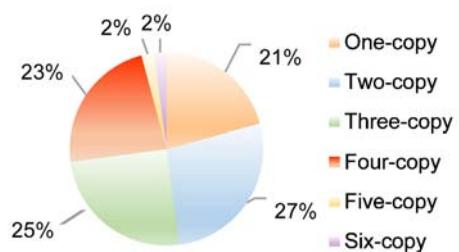


图7 拟南芥MADS-box基因在甘蓝型油菜中的拷贝数

Figure 7 Copy number of *Arabidopsis thaliana* MADS-box genes in *Brassica napus*

表2 甘蓝型油菜MADS-box基因亚家族的选择压力

Table 2 Select pressure of MADS-box gene subfamily in *Brassica napus*

Subfamily	<i>Arabidopsis thaliana</i>	<i>Brassica napus</i>	Average copy number	Average Ka/Ks
M-type	20	34	1.70	0.43
MIKC-type	29	95	3.27	0.22

K-box结构域。MIKC*同时具有M-type和MIKC^C的特性,因此在早期的研究中将其归于Mδ,后来在其它物种的MADS-box基因家族分类中将其归于MIKC-type (Parenicová et al., 2003; Arora et al., 2007; Díaz-Riquelme et al., 2009; Fan et al., 2013; Shu et al., 2013; Wei et al., 2014; Duan et al., 2015; Grimpel et al., 2016)。因此,有学者认为MIKC*是从MIKC^C进化到M-type的中间形态(Xu et al., 2014),即MIKC^C在长期进化选择过程中K-box结构域突变,但保留了较多内含子,形成了MIKC*, MIKC*在进一步的进化中丢失了内含子,形成短序列、少内含子的M-type。

多倍化是物种进化的重要动力,多数高等植物在进化过程中都经历了不同水平的多倍化(Tang et al., 2008)。十字花科物种都曾经历三倍化(Bowers et al., 2003)。比较基因组学研究显示,拟南芥与白菜分化自同一祖先物种。大约20~40 M,拟南芥和白菜的共同祖先发生了1次基因组水平的复制事件,称为α事件(Woodhouse et al., 2014)。芸薹属物种在此之后又经历了1次三倍化(Chalhoub et al., 2014; Liu et al., 2014; Yang et al., 2016)。在5~9 M,甘蓝和白菜的共同祖先又经历了1次基因组水平三倍化(The *Brassica rapa* Genome Sequencing Project Consortium et al., 2011)。基因组测序结果显示,白菜与甘蓝的分化约发生在4 M (Liu et al., 2014)。相对于拟南芥,甘蓝和白菜各发生了1次基因组复制。甘蓝型油菜大约在7 500~12 500年前由白菜和甘蓝天然杂交形成(Chalhoub et al., 2014)。物种进化的实质是基因的进化,多倍化事件是基因家族扩张的主要动力,在多倍化之后,基因数目呈倍数增长。依据甘蓝型油菜进化史可推断,甘蓝型油菜MADS-box家族基因数目应多于307个。在基因组复制事件发生后,植物为了代谢的平衡,往往会出现大面积的基因丢失和染色体重

排,基因功能有3种分化方向:保持原有功能、丢失原有功能及产生新功能(Nakano et al., 2006)。本研究显示,部分MADS-box基因在进化过程中丧失MADS-box结构域,只剩下K-box结构域。还有一部分序列,在Interpro中未注释,也没有匹配的domain,初步推测为假基因。因此这可能是在经历2次基因组水平的三倍化之后甘蓝型油菜中MADS-box基因家族数仅为307的原因。此外,M-type和MIKC-type在进化扩张上存在区别,MIKC-type主要来源于基因组水平的复制事件,M-type基因扩张主要来自基因组内小范围的复制事件(Edger and Pires, 2009)。我们对拟南芥和甘蓝型油菜MADS-box基因进行共线性统计,发现多数MADS-box基因在甘蓝型油菜中拥有多个拷贝,其中MIKC-type基因的平均拷贝数约为M-type的2倍。MIKC-type和M-type选择压力的平均值也显示,MIKC-type受到的选择压力更大。这表明基因组复制事件对MIKC-type的扩张起重要作用,在进化的过程中MIKC-type基因被有选择地保留(Maere et al., 2005)。

综上,本研究在甘蓝型油菜中鉴定出307个MADS-box基因,并利用生物信息学手段对这些基因的基本信息进行注释,探究其系统发育关系。染色体定位结果显示,MADS-box基因在染色体上都有分布,但数量不一。对307个MADS-box进行基因结构和保守序列预测表明,同一亚家族的基因结构极为相似。此外,我们对甘蓝型油菜和拟南芥MADS-box基因共线性及进化过程中的选择压力进行了探索,发现多倍化事件对MIKC-type扩张起重要作用。这些结论将为深入研究甘蓝型油菜MADS-box基因功能奠定基础,并为探索其进化过程提供借鉴。

参考文献

- Airoldi CA, Davies B (2012). Gene duplication and the evolution of plant MADS-box transcription factors. *J Genet Genomics* **39**, 157–165.
- Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF (2000a). MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *Plant J* **24**, 457–466.
- Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, de Pouplana LR, Martínez-Castilla L,

- Yanofsky MF** (2000b). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA* **97**, 5328–5333.
- Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S** (2007). MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* **8**, 242.
- Bowers JE, Chapman BA, Rong JK, Paterson AH** (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433–438.
- Chalhoub B, Denoeud F, Liu SY, Parkin IAP, Tang HB, Wang XY, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao MX, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier MC, Fan GY, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VHD, Chalabi S, Hu Q, Fan CC, Tollenaere R, Lu YH, Battail C, Shen JX, Sidebottom CHD, Wang XF, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu ZS, Sun FM, Lim YP, Lyons E, Town CD, Bancroft I, Wang XW, Meng JL, Ma JX, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowden RJ, Tost J, Edwards D, Zhou YM, Hua W, Sharpe AG, Paterson AH, Guan CY, Wincker P** (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953.
- Chang YY, Chiu YF, Wu JW, Yang CH** (2009). Four orchid (*Oncidium* Gower Ramsey) AP1/AGL9-like MADS box genes show novel expression patterns and cause different effects on floral transition and formation in *Arabidopsis thaliana*. *Plant Cell Physiol* **50**, 1425–1438.
- Cheng F, Liu SY, Wu J, Fang L, Sun SL, Liu B, Li PX, Hua W, Wang XW** (2011). BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biol* **11**, 136.
- Day RC, Herridge RP, Ambrose BA, Macknight RC** (2008). Transcriptome analysis of proliferating *Arabidopsis* endosperm reveals biological implications for the control of syncytial division, cytokinin signaling, and gene expression regulation. *Plant Physiol* **148**, 1964–1984.
- De Bodt S, Raes J, Van de Peer Y, Theißen G** (2003). And then there were many: MADS goes genomic. *Trends Plant Sci* **8**, 475–483.
- Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, Carmona MJ** (2009). Genome-wide analysis of MIKC^C-type MADS box genes in grapevine. *Plant Physiol* **149**, 354–369.
- Doebley J, Lukens L** (1998). Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**, 1075–1082.
- Duan WK, Song XM, Liu TK, Huang ZN, Ren J, Hou XL, Li Y** (2015). Genome-wide analysis of the MADS-box gene family in *Brassica rapa* (Chinese cabbage). *Mol Genet Genomics* **290**, 239–255.
- Edger PP, Pires JC** (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res* **17**, 699–717.
- Fan CM, Wang X, Wang YW, Hu RB, Zhang XM, Chen JX, Fu YF** (2013). Genome-wide expression analysis of soybean MADS genes showing potential function in the seed development. *PLoS One* **8**, e62288.
- Fang SC, Fernandez DE** (2002). Effect of regulated over-expression of the MADS domain factor AGL15 on flower senescence and fruit maturation. *Plant Physiol* **130**, 78–89.
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M** (2014). Pfam: the protein families database. *Nucleic Acids Res* **42**, D222–D230.
- Gan YB, Filleur S, Rahman A, Gotensparre S, Forde BG** (2005). Nutritional regulation of *ANR1* and other root-expressed MADS-box genes in *Arabidopsis thaliana*. *Planta* **222**, 730–742.
- Gramzow L, Ritz MS, Theißen G** (2010). On the origin of MADS-domain transcription factors. *Trends Genet* **26**, 149–153.
- Greenup A, Peacock WJ, Dennis ES, Trevaskis B** (2009). The molecular biology of seasonal flowering-responses in *Arabidopsis* and the cereals. *Ann Bot* **103**, 1165–1172.
- Grimplet J, Martínez-Zapater JM, Carmona MJ** (2016). Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC Genomics* **17**, 80.
- Hemming MN, Trevaskis B** (2011). Make hay when the sun shines: the role of MADS-box genes in temperature-dependant seasonal flowering responses. *Plant Sci* **180**, 447–453.
- Immink RGH, Kaufmann K, Angenent GC** (2010). The ‘ABC’ of MADS domain protein behaviour and interactions. *Semin Cell Dev Biol* **21**, 87–93.

- Jin JP, Zhang H, Kong L, Gao G, Luo JC** (2014). Plant TFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res* **42**, D1182–D1187.
- Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu JZ, Zhou SG, Childs KL, Davidson RM, Lin HN, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR, Matsumoto T** (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4.
- Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M** (2003). Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol* **20**, 1963–1977.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA** (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645.
- Letunic I, Doerks T, Bork P** (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res* **43**, D257–D260.
- Liu SY, Liu YM, Yang XH, Tong CB, Edwards D, Parkin IAP, Zhao MX, Ma JX, Yu JY, Huang SM, Wang XY, Wang JY, Lu K, Fang ZY, Bancroft I, Yang TJ, Hu Q, Wang XF, Yue Z, Li HJ, Yang LF, Wu J, Zhou Q, Wang WX, King GJ, Pires JC, Lu CX, Wu ZY, Sampath P, Wang Z, Guo H, Pan SK, Yang LM, Min JM, Zhang D, Jin DC, Li WS, Belcram H, Tu JX, Guan M, Qi CK, Du DZ, Li JN, Jiang LC, Batley J, Sharpe AG, Park BS, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong CH, Wang L, Li JP, Hu ZY, Zhuang M, Huang Y, Huang JY, Shi JQ, Mei DS, Liu J, Lee TH, Wang JP, Jin HZ, Li ZY, Li X, Zhang JF, Xiao L, Zhou YM, Liu ZS, Liu XQ, Qin R, Tang X, Liu WB, Wang YP, Zhang YY, Lee J, Kim HH, Denoeud F, Xu X, Liang XM, Hua W, Wang XW, Wang J, Chalhoub B, Paterson AH** (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun* **5**, 3930.
- Liu Y, Cui SJ, Wu F, Yan S, Lin XL, Du XQ, Chong K, Schilling S, Theissen G, Meng Z** (2013). Functional conservation of MIKC^{*}-type MADS box genes in Arabidopsis and rice pollen maturation. *Plant Cell* **25**, 1288–1303.
- Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y** (2005). Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA* **102**, 5454–5459.
- Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM** (2011). The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell* **23**, 865–872.
- Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesceat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong SY, Bateman A, Punta M, Attwood TK, Sigrist CJA, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang HZ, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi HY, Thomas PD, Finn RD** (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* **43**, D213–D221.
- Nagaharu U** (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn J Bot* **7**, 389–452.
- Nakano T, Suzuki K, Fujimura T, Shinshi H** (2006). Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol* **140**, 411–432.
- Nam J, dePamphilis CW, Ma H, Nei M** (2003). Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol Biol Evol* **20**, 1435–1447.
- Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L** (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* **15**, 1538–1551.
- Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF** (1995). Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* **140**, 345–356.
- Saha G, Park JI, Jung HJ, Ahmed NU, Kayum MA, Chung MY, Hur Y, Cho YG, Watanabe M, Nou IS** (2015). Genome-wide identification and characterization of MADS-box family genes related to organ development and stress resistance in *Brassica rapa*. *BMC Genomics* **16**, 178.
- Shao SQ, Li BY, Zhang ZT, Zhou Y, Jiang J, Li XB** (2010). Expression of a cotton MADS-box gene is regulated in anther development and in response to phytohormone signaling. *J Genet Genomics* **37**, 805–816.
- Shore P, Sharrocks AD** (1995). The MADS-box family of transcription factors. *Eur J Biochem* **229**, 1–13.
- Shu YJ, Yu DS, Wang D, Guo DL, Guo CH** (2013). Genome-wide survey and expression analysis of the MADS-box

- gene family in soybean. *Mol Biol Rep* **40**, 3901–3911.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S** (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.
- Tang HB, Bowers JE, Wang XY, Ming R, Alam M, Paterson AH** (2008). Synteny and collinearity in plant genomes. *Science* **320**, 486–488.
- Tapia-López R, García-Ponce B, Dubrovsky JG, Garay-Arroyo A, Pérez-Ruiz RV, Kim SH, Acevedo F, Pelaz S, Alvarez-Buylla ER** (2008). An AGAMOUS-related MADS-box gene, *XAL1* (*AGL12*), regulates root meristem cell proliferation and flowering transition in *Arabidopsis*. *Plant Physiol* **146**, 1182–1192.
- The *Brassica rapa* Genome Sequencing Project Consortium, Wang XW, Wang HZ, Wang J, Sun RF, Wu J, Liu SY, Bai YQ, Mun JH, Bancroft I, Cheng F, Huang SW, Li XX, Hua W, Wang JY, Wang XY, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu BH, Li B, Liu B, Tong CB, Song C, Duran C, Peng CF, Geng CY, Koh C, Lin CY, Edwards D, Mu DS, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang HB, Wang HP, Belcram H, Zhou HL, Hirakawa H, Abe H, Guo H, Wang H, Jin HZ, Parkin IAP, Batley J, Kim JS, Just J, Li JW, Xu JH, Deng J, Kim JA, Li JP, Yu JY, Meng JL, Wang JP, Min JM, Poulaire J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao MX, Jin MN, Ramchiary N, Drou N, Berkman PJ, Cai QL, Huang QF, Li RQ, Tabata S, Cheng SF, Zhang S, Zhang SJ, Huang SM, Sato SS, Sun SL, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li YR, Du YC, Liao YC, Lim Y, Narusaka Y, Wang YP, Wang ZY, Li ZY, Wang ZW, Xiong ZY, Zhang ZH** (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* **43**, 1035–1039.
- Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, Saedler H** (2000). A short history of MADS-box genes in plants. *Plant Mol Biol* **42**, 115–149.
- Tiwari S, Spielman M, Schulz R, Oakey RJ, Kelsey G, Salazar A, Zhang K, Pennell R, Scott RJ** (2010). Transcriptional profiles underlying parent-of-origin effects in seeds of *Arabidopsis thaliana*. *BMC Plant Biol* **10**, 72.
- Wei B, Zhang RZ, Guo JJ, Liu DM, Li AL, Fan RC, Mao L, Zhang XQ** (2014). Genome-wide analysis of the MADS-box gene family in *Brachypodium distachyon*. *PLoS One* **9**, e84781.
- Wei X, Wang LH, Yu JY, Zhang YX, Li DH, Zhang XR** (2015). Genome-wide identification and analysis of the MADS-box gene family in *sesame*. *Gene* **569**, 66–76.
- Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M, Wang XW** (2014). Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. *Proc Natl Acad Sci USA* **111**, 5283–5288.
- Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenführer J, von Mering C, Grossniklaus U** (2010). *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr Biol* **20**, 506–512.
- Xu ZD, Zhang QX, Sun LD, Du DL, Cheng TR, Pan HT, Yang WR, Wang J** (2014). Genome-wide identification, characterisation and expression analysis of the MADS-box gene family in *Prunus mume*. *Mol Genet Genomics* **289**, 903–920.
- Yang JH, Liu DY, Wang XW, Ji CM, Cheng F, Liu BN, Hu ZY, Chen S, Pental D, Ju YH, Yao P, Li XM, Xie K, Zhang JH, Wang JL, Liu F, Ma WW, Shopan J, Zheng HK, Mackenzie SA, Zhang MF** (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat Genet* **48**, 1225–1232.
- Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM** (1990). The protein encoded by the *Arabidopsis* homeotic gene AGAMOUS resembles transcription factors. *Nature* **346**, 35–39.
- Yao QY, Xia EH, Liu FH, Gao LZ** (2015). Genome-wide identification and comparative expression analysis reveal a rapid expansion and functional divergence of duplicated genes in the WRKY gene family of cabbage, *Brassica oleracea* var. *capitata*. *Gene* **557**, 35–42.
- Yu LH, Miao ZQ, Qi GF, Wu J, Cai XT, Mao JL, Xiang CB** (2014). MADS-box transcription factor AGL21 regulates lateral root development and responds to multiple external and physiological signals. *Mol Plant* **7**, 1653–1669.
- Zhang Z, Li J, Zhao XQ, Wang J, Wong GKS, Yu J** (2006). KaKs_calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* **4**, 259–263.

Genome-wide Survey and Phylogenetic Analysis of MADS-box Gene Family in *Brassica napus*

Huhu Gao[†], Yunxiao Zhang[†], Shengwu Hu, Yuan Guo^{*}

State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Northwest A&F University,
Yangling 712100, China

Abstract The MADS-box gene family is involved in many processes during plant growth and development, such as flowering time, floral organ differentiation, root growth, meristem differentiation, ovary and gamete development, fruit enlargement and senescence. In this study, we used rape (*Brassica napus*) genome sequencing data with bioinformatics methods to identify and annotate the MADS-box genes. Rape contains 307 members of MADS-box gene family. According to the evolutionary relationships, these genes can be divided into two subfamilies: I-type, also known as M-type, containing three subclades, α , β , and γ ; II type, also known as MIKC-type containing two subclades, MIKC^C and MIKC*. MIKC^C can be further divided into 13 groups. The number of MADS-box genes is greater in the A than C subgenome chromosome of *B. napus*. For the gene structure, the sequence is longer for MIKC-type than M-type genes and contains more exons. The number of motifs in M-type genes is about 2–5, and MIKC-type genes contain an average of 7 motifs. Synteny analysis revealed that whole-genome duplication played a major role in the expansion of the *BnaMADS* gene family, especially the MIKC-type subfamily. The selection pressure of the MIKC-type subfamily was about 2 times that for the M-type, which resulted in the selective preservation of MIKC-type subfamily genes during evolution.

Key words *Brassica napus*, MADS-box, gene structure, evolution, synteny

Gao HH, Zhang YX, Hu SW, Guo Y (2017). Genome-wide survey and phylogenetic analysis of MADS-box gene family in *Brassica napus*. Chin Bull Bot **52**, 699–712.

[†] These authors contributed equally to this paper

* Author for correspondence. E-mail: guoyuan2109@163.com

(责任编辑: 白羽红)

附录1 甘蓝型油菜MADS-box基因家族成员

Appendix 1 The members of MADS-box gene family identified in *Brassica napus*

<http://www.chinbullbotany.com/fileup/PDF/t16244-1.pdf>

附录2 甘蓝型油菜与拟南芥共线性基因的选择压力

Appendix 2 Selection pressure of syntenic MADS-box genes between *Arabidopsis thaliana* and *Brassica napus*

<http://www.chinbullbotany.com/fileup/PDF/t16244-2.pdf>