



甘蓝型油菜ZF-HD基因家族的鉴定与系统进化分析

宋敏*, 张瑶, 王丽莹, 彭向永

曲阜师范大学生命科学学院, 曲阜 273165

摘要 ZF-HD是一类植物特有的转录因子, 在植物生长发育及胁迫响应过程中发挥重要作用。利用生物信息学方法, 在甘蓝型油菜(*Brassica napus*)基因组中鉴定到62个ZF-HD基因, 其中83.9%的基因缺乏内含子, 93.5%的BnZF-HD等电点大于7, 预测定位于细胞核的蛋白大多由100个以上氨基酸组成。根据进化关系可将其分为6个亚群, 在每个亚群中, 甘蓝(*B. oleracea*)和白菜(*B. rapa*)的ZF-HD基因数量相等或近似相等, 而甘蓝型油菜的ZF-HD基因数量接近或等同于甘蓝和白菜的ZF-HD基因数量之和。同一亚群的motif数量和类型高度相似。共线性分析结果显示, 全基因组三倍体化使ZF-HD基因在二倍体祖先得到扩张, 而异源多倍体化又进一步使甘蓝型油菜ZF-HD基因家族扩张。 K_a/K_s 值说明大多数ZF-HD基因在进化过程中受到了纯化选择。所有BnZF-HD基因都具有光响应元件, 2/3的基因具有MeJA、ABA和厌氧诱导顺式作用元件, 推测这些基因可能参与相关生物学过程。研究结果为进一步挖掘该家族基因的生物学功能奠定基础, 同时为揭示多基因家族在异源多倍体中的进化式样提供借鉴。

关键词 甘蓝型油菜, ZF-HD, 进化, 共线性分析, 顺式作用元件

宋敏, 张瑶, 王丽莹, 彭向永 (2019). 甘蓝型油菜ZF-HD基因家族的鉴定与系统进化分析. 植物学报 54, 699–710.

ZF-HD转录因子(zinc finger homeodomain, ZH-D)是植物特有的转录因子(Windhovel et al., 2001; Tan and Irish, 2006), 其在 C_4 植物黄顶菊(*Flaveria trinervia*)中首次被报道, 是磷酸烯醇式丙酮酸羧化酶(PEPCase)编码基因的潜在调控因子(Windhovel et al., 2001)。ZF-HD转录因子的命名是因其氨基酸序列中有保守的HD结构域(homeodomain)和C2H2型锌指结构(C2H2-type zinc finger motif)。HD结构域是含60个氨基酸的DNA结合域, 大多数具HD结构域的蛋白同时具有其它类型的结构域或基序, 可分为HD-ZIP、ZF-HD、WOX、Bell type HD、PHD和KNOX六个家族(Ariel et al., 2007)。含HD的蛋白质广泛存在于动植物体内, 在发育过程中发挥重要作用(Nam and Nei, 2005; Bhattacharjee et al., 2015)。锌指结构是由多个保守的氨基酸残基围绕核心的锌离子组成, 形成相对独立的指状四面体结构(Krishna et al., 2003)。根据其Cys和His残基的不同组合分成不同类

型, 如C3H、C2H2、C2C2和C4C6, 其中以C2H2型最为常见。锌指可以识别和结合DNA、RNA、DNA-RNA双链分子或蛋白质(Mackay and Crossley, 1998; Takatsuji, 1999; Krishna et al., 2003), 在转录和翻译水平调控基因表达, 在植物的应激反应和防御激活中发挥重要作用(Mackay, 1998)。在ZF-HD蛋白中, HD结构域能够结合DNA, ZF结构可以增强由HD结构域介导的蛋白质与DNA的相互作用(Windhovel et al., 2001; Hu et al., 2008)。

有趣的是, 有一类蛋白MIF (MINI ZINC FINGER)与ZF-HD蛋白的ZF结构域序列相似度高但缺乏HD结构域(Hu and Ma, 2006; Hu et al., 2008), 且是种子植物特有的蛋白。推测MIF蛋白与ZF-HD蛋白的进化关系有2种可能, 一是MIF蛋白由ZF-HD蛋白丢失HD结构域形成; 二是MIF获得HD结构域形成ZF-HD蛋白(Hu et al., 2008; Wang et al., 2016)。按照已发表的模式植物拟南芥(*Arabidopsis thaliana*)和水稻(*Oryza*

收稿日期: 2019-03-25; 接受日期: 2019-07-26

基金项目: 山东省高等学校科技计划(No.J15LE02)、中国博士后基金(No.2018M632646)和山东省高水平应用型立项建设专业(群)-生物工程专业群(No. [2016] 11-10)

* 通讯作者。E-mail: smin2000@qfnu.edu.cn

sativa)的ZF-HD蛋白的划分方法,本研究将MIF归入ZF-HD家族。

目前,国内外学者已经对水稻、大豆(*Glycine max*)、拟南芥、小麦(*Triticum aestivum*)、葡萄(*Vitis vinifera*)、白菜(*Brassica rapa*)、番茄(*Solanum lycopersicum*)和陆地棉(*Gossypium hirsutum*)等物种中的ZF-HD基因家族进行了鉴定(Jørgensen et al., 1999; Windhövel et al., 2001; Deng et al., 2002; Tan and Irish, 2006; Jain et al., 2008; Bhattacharjee and Jain, 2013; Wang et al., 2014, 2016; Khadiza et al., 2017; Abdullah et al., 2018)。已有研究表明,ZF-HD转录因子参与许多生长发育过程和非生物胁迫响应。拟南芥ZF-HD基因均在花组织中有表达,有些表现为花特异性表达(Tan and Irish, 2006)。拟南芥AtZHD1蛋白能特异性结合*ERD1* (*EARLY RESPONSE TO DEHYDRATION STRESS 1*)启动子,AtZHD1基因受ABA、盐和干旱诱导(Tran et al., 2007; Wang et al., 2014)。过表达AtZHD1和NAC基因的拟南芥植株耐旱性增强(Tran et al., 2007; Hu et al., 2008)。4个水稻ZHD基因与基因表达调控相关。大豆GmZHD1和GmZHD2能够结合钙调蛋白编码基因*GmCaM4*的启动子,在病原菌侵染后表达量升高(Park et al., 2007; Hu et al., 2008; Wang et al., 2014)。15个陆地棉ZHD基因参与纤维的早期发育和纤维色素合成(Abdullah et al., 2018)。大多数番茄ZF-HD基因在花芽部位表达,其中有多数响应非生物胁迫和激素处理(Khadiza et al., 2017)。多数白菜ZF-HD基因在花中表达,且明显受光周期、春化和非生物胁迫诱导(Wang et al., 2016)。

十字花科芸薹属植物甘蓝型油菜(*B. napus*) ($A_nA_nC_nC_n$, $2n=38$)是一种重要的油料作物,由白菜(A_rA_r , $2n=20$)和甘蓝(*B. oleracea*) (C_oC_o , $2n=18$)杂交后进化而来(Chalhoub et al., 2014)。因此,甘蓝型油菜是研究异源多倍体化进程的重要材料。甘蓝型油菜基因组测序已在2014年完成(Chalhoub et al., 2014)。基于基因组测序结果,利用生物信息学方法,我们对甘蓝型油菜ZF-HD基因家族进行鉴定,对其蛋白结构、基本理化特征、顺式作用元件和基因共线性等进行分析,以期深入揭示ZF-HD基因家族的生物功能奠定基础,并为揭示多基因家族在异源多倍体中的进化式样提供理论依据。

1 材料与方法

1.1 数据来源

甘蓝型油菜(*Brassica napus* L.)、白菜(*B. rapa* L.)和甘蓝(*B. oleracea* L.)的基因组序列、CDS序列、蛋白质序列及其注释信息均下载自BRAD数据库(<http://brassicadb.org>) (Cheng et al., 2011)。拟南芥(*Arabidopsis thaliana* (L.) Heynh) ZF-HD蛋白序列下载自TAIR (<http://www.arabidopsis.org/>)。

1.2 ZF-HD基因的鉴定

通过3种途径搜索甘蓝型油菜和甘蓝的ZF-HD基因。

(1) 从植物转录因子网站PlantTFDB (<http://planttfdb.cbi.pku.edu.cn/>, V4.0) (Jin et al., 2017b)分别下载甘蓝型油菜和甘蓝全部ZF-HD家族成员蛋白序列;(2) 从Pfam数据库(<http://pfam.xfam.org/>) (Bateman et al., 2002; Finn et al., 2016)下载ZF-HD基因的PF-04770结构域种子数据stockholm文件,使用hmmsearch工具分别搜索甘蓝型油菜和甘蓝的蛋白质数据库,参数值为1.0;(3) 以拟南芥ZF-HD蛋白序列比对文件,进行本地Blast P检索,e值设为 $1e^{-3}$;整理获取的序列,去掉重复序列,然后利用在线工具CDD (<http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>)和Smart (Schultz et al., 1998) (<http://smart.emblheidelberg.de/>)检测蛋白序列的结构域,获得候选基因,用于后续分析。

1.3 ZF-HD蛋白的生物信息学分析

采用ProtParam (<http://web.expasy.org/protparam/>)获取蛋白质的相对分子质量、氨基酸长度以及等电点。利用WoLF PSORT (<http://www.genscript.com/wolf-psort.html>)在线工具预测蛋白质的亚细胞定位。利用Ensembl基因组数据库(<http://plants.ensembl.org/index.html>)获取cDNA序列、CDS序列以及启动子上游2 000 bp序列。通过PlantCare在线分析工具(<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)对ZF-HD基因的顺式作用元件进行搜索,在Excel软件中整理Tab结果文件,用TBtools软件的Simple BioSequence Viewer工具展示顺式作用元件的分布(Chen et al., 2018)。通过GSDS工具(<http://gsds.cbi.pku.edu.cn/>) (Hu et al., 2015)分析ZF-HD的基因结构。

1.4 *ZF-HD*基因家族的系统进化分析

利用Clustal X1.8软件对甘蓝型油菜、白菜、甘蓝和拟南芥的*ZF-HD*蛋白全长序列进行多重比对, 比对结果用MEGA 7.0进行系统进化分析(Kumar et al., 2016)。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)。利用最大似然法(maximum likelihood method, ML)构建进化树, 氨基酸替换模型为JTT+G, bootstrap设为1 000。利用在线工具Evolview (<https://www.evolgenius.info/evolview/>)进行树图编辑(He et al., 2016)。

1.5 *ZF-HD*基因的染色体分布

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 (<http://meme-suite.org/tools/meme>) (Bailey et al., 2015)对甘蓝型油菜*ZF-HD*蛋白的motif进行预测和分析。MEME参数设置: motif数量为15, motif长度为6–60。下载mast.xml结果文件, 用TBtools软件的Redraw motif pattern工具展示motif的分布(Chen et al., 2018)。

1.6 甘蓝型油菜与拟南芥共线性基因与选择压力分析

通过BRAD (<http://brassicadb.org/brad/searchSyn-teny.php>) (Cheng et al., 2012)在线检索获取拟南芥*ZHD*基因在甘蓝型油菜、白菜和甘蓝中的共线性基因及其在染色体上的位置。用Circos (Krzywinski et al., 2009)软件展示这些共线性基因的关系。用TBtools软件计算共线性基因之间的 K_a/K_s 值(Chen et al., 2018), 估算选择压力。

2 结果与讨论

2.1 *BnZF-HD*基因鉴定及基本理化特征

我们通过hmmsearch、Blast P搜索和在PlantTFDB数据库直接下载3种方法, 最终在甘蓝型油菜基因组中鉴定到62个*ZF-HD*基因, 根据其在染色体上的位置依次命名为*BnZF-HD1–62* (附录1)。我们利用ProtParam程序获取62个*BnZF-HD*基因家族成员编码蛋白质的分子量、氨基酸序列长度以及等电点预测信息。结果显示, 氨基酸数目最多的是*BnZF-HD11*, 有344个氨基酸, 预测分子量为37.30 kDa, 理论等电

点为8.23; 氨基酸数目最少的是*BnZF-HD13*, 有84个氨基酸, 预测分子量为9.22 kDa, 理论等电点为6.82; 共16个*BnZF-HD*的氨基酸数目低于100, 包括全部MIF蛋白; 等电点变化范围在5.77–9.78之间, 87.1%的*BnZF-HD*等电点大于7; 只有10个*BnZF-HD*基因含1个内含子, 其余基因无内含子。

我们采用WoLF PSORT程序对*BnZF-HD*基因家族成员进行蛋白亚细胞定位预测分析。结果表明, 41个*BnZF-HD*蛋白定位于细胞核, 其余定位于叶绿体、溶酶体或线粒体。16个氨基酸数目低于100的以MIF为主的蛋白中除*BnZF-HD48*外均定位于细胞质。

2.2 *BnZF-HD*基因系统进化分析

为了解*ZF-HD*基因家族的进化关系, 我们利用62个油菜、31个甘蓝(本研究鉴定, 附录2)、31个白菜(Wang et al., 2016)和17个拟南芥的*ZF-HD*蛋白序列(总计141个蛋白序列), 基于邻接法(图1)和最大似然法构建进化树(附录3), 根据分支情况并参照拟南芥*ZHD*基因的划分(Hu et al., 2008), 将其分为6个亚群(图1; 附录3), 且每个亚群所含蛋白数目相同。*ZHDIV*亚群包含的*BnZF-HD*最多, 为16个; *ZHDV*亚群只含有2个*BnZF-HD*; MIF蛋白单独聚成一个亚群, 有14个*BnZF-HD*属于这一亚群。每亚群都包含4个物种的*ZF-HD*蛋白, 其中每亚群的甘蓝和白菜的*ZF-HD*数量都相等, 且除了*ZHDVI*和*ZHDV*亚群以外, 甘蓝和白菜的*ZF-HD*数量之和等于同一亚群甘蓝型油菜的*ZF-HD*数目。系统发生树分支末端形成了44对姐妹基因对, 其中18对在甘蓝型油菜和白菜之间, 26对在甘蓝型油菜和甘蓝之间, 这些姊妹基因对表现为更近的亲缘关系。

2.3 *ZF-HD*家族基因在染色体上的位置

染色体定位显示, 52个*BnZF-HD*基因在染色体的位置明确, 在17条染色体上呈现不均匀分布(图2)。A、C亚基因组各有31个*BnZF-HD*, 在除A5以外的18条染色体上分布有1–7个数目不等的基因, A9染色体最多(7个), 其次为A7和C8染色体(各6个), A1、A10、C1和C4染色体仅有1个*BnZF-HD*基因。未在图中标识的10个基因中, *BnZF-HD4*和*BnZF-HD39*分别位于A02_random和C04_random, *BnZFHD26*和*BnZF-HD27*均位于A09_random, 其余6个基因位于scaffolds。

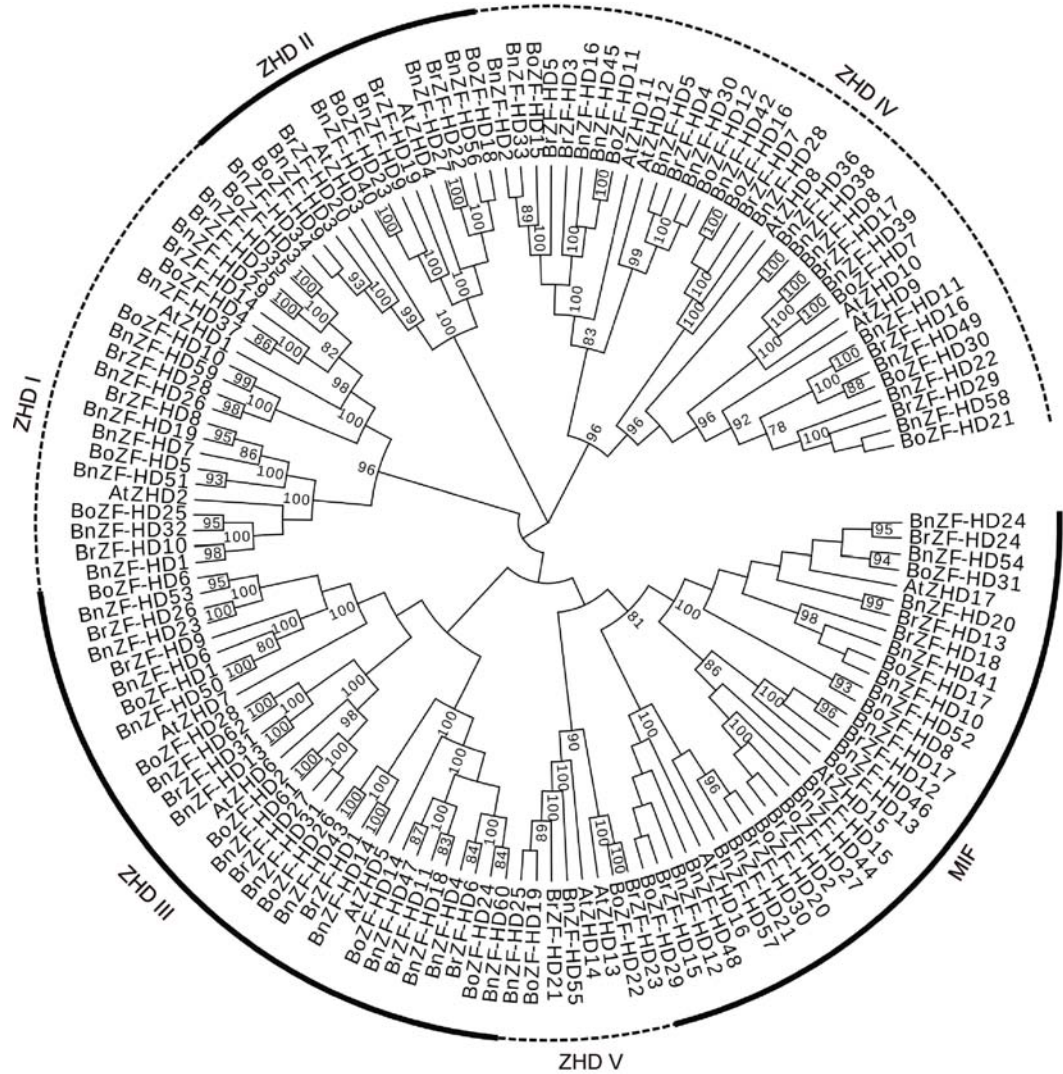


图1 甘蓝型油菜、拟南芥、白菜和甘蓝ZF-HD蛋白的系统发生树
利用MEGA 7.0软件构建邻接树, bootstrap值设为1 000。

Figure 1 Phylogenetic analysis of *Brassica napus*, *Arabidopsis thaliana*, *B. rapa* and *B. oleracea* ZF-HD proteins
This phylogenetic unrooted tree was constructed by the neighbor-joining (NJ) method with MEGA 7.0, with 1 000 bootstrap replicates.

同一亚族的ZF-HD基因在Ar与An亚基因组之间以及在Co与Cn亚基因组之间的定位基本相似(图2)。例如,白菜与甘蓝的MIF基因分别位于Ar6-Ar9及Co5-Co9染色体上,甘蓝型油菜的MIF基因分布于An6-An9及Cn5-Cn9染色体上,每条染色体上的MIF亚族基因数目及相对位置相似,说明这些基因排列方式高度保守。

2.4 甘蓝型油菜ZF-HD家族蛋白保守motif分布

通过MEME在线工具我们对ZF-HD的motif进行预测,

结果(图3; 附录4)显示, BnZF-HD的motif数目在1-5个之间, 25个BnZF-HD包含5个motif, 14个BnZF-HD包含4个motif, 18个BnZF-HD含3个motif, 其中MIF亚家族的蛋白都包含相同的3个motif, BnZF-HD38只有1个motif, BnZF-HD13、BnZF-HD25、BnZF-HD36和BnZF-HD55有2个motif。在预测到的10个motif中, 通过SMART在线工具检测发现motif 2和motif 3是ZF-HD_dimer结构域, 存在于57个ZF-HD蛋白中, motif 1是HOX (homeodomain)结构域, 除了BnZF-HD36、BnZF-HD38及MIF亚家族的14个成员以外,

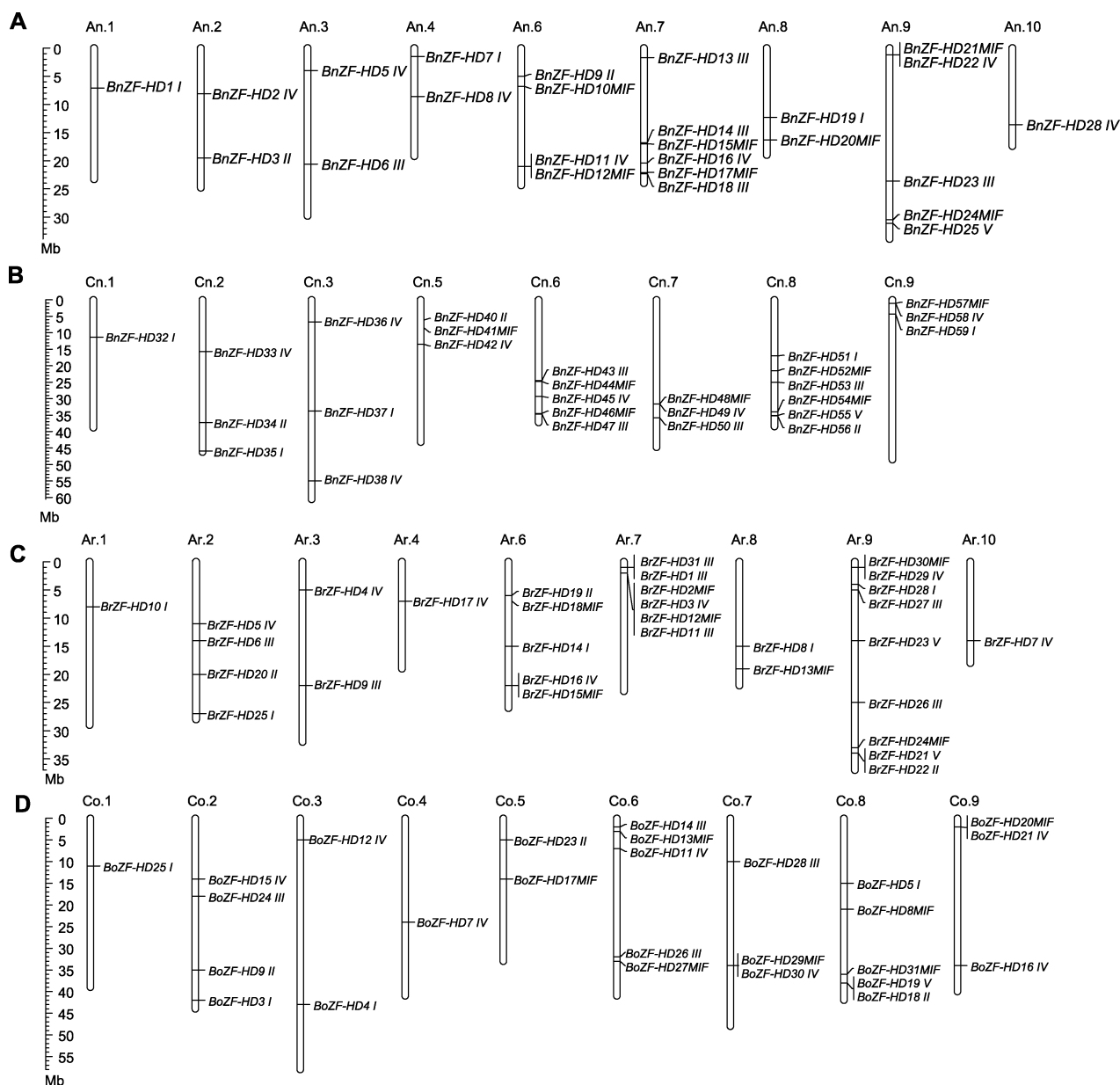


图2 甘蓝型油菜(A, B)、白菜(C)和甘蓝(D) *ZF-HD* 基因在染色体上的分布
基因名称后标注该基因所属亚群。

Figure 2 Chromosomal location of *ZF-HD* genes from *Brassica napus* (A, B), *B. rapa* (C) and *B. oleracea* (D)
The gene name was followed by its subgroup.

其它BnZF-HD都具有motif 1, 其余motif是一些非保守结构, 只存在于部分ZF-HD蛋白中。每个亚族的motif分布模式基本相似, 而亚群之间存在差异, 可能与亚群功能趋异有关。

2.5 甘蓝型油菜和拟南芥*ZF-HD*基因的共线性分析

我们利用BRAD网站中的Syntenic工具获取了拟南芥

所有*ZF-HD*基因在白菜、甘蓝和甘蓝型油菜中的共线性基因, 并分析其共线性关系(图4)。结果显示, 48个*BnZF-HD*基因在拟南芥中有同源基因, 其中64.5% (40个)为多拷贝基因, 即在甘蓝型油菜的亚基因组中具有2个及以上的拷贝, 其余8个*ZF-HD*基因在白菜、甘蓝及甘蓝型油菜的A亚基因组和/或C亚基因组中各有1个拷贝, 属于保守基因。14个*BnZF-HD*基因在拟

南芥、白菜和甘蓝中缺乏相应的同源基因。甘蓝型油菜缺乏*AtZHD13*的同源基因,而白菜和甘蓝中各有1个该基因的同源基因,说明甘蓝型油菜中存在基因丢失现象。

拟南芥与甘蓝型油菜的共线性基因对在各条染色体数量不一,其中拟南芥1号染色体的6个基因分别与甘蓝型油菜11个A组基因和12个C组基因具有共线性关系。甘蓝型油菜中A7、A9、C6和C8染色体上共线性基因的数量多于其它染色体。

我们在甘蓝型油菜和白菜A亚基因组之间鉴定出15个同源*ZF-HD*基因对,在油菜和甘蓝C亚基因组之间发现20个同源*ZF-HD*基因对。Chalhoub等(2014)报道了白菜和甘蓝祖先之间83.7%的同源基因对在油菜中仍保留为同源基因对。我们发现,在

白菜和甘蓝之间,64.5%的同源*ZF-HD*基因对在油菜中仍保留为同源基因对(附录5)。上述结果表明,在异源四倍体油菜形成和进化过程中,约2/3的*ZF-HD*基因保持完整。

2.6 选择压力分析

为确定是否有选择压力作用于甘蓝型油菜*ZF-HD*基因,我们以拟南芥、白菜和甘蓝*ZF-HD*基因为参照估算了在进化过程中甘蓝型油菜*ZF-HD*基因家族的选择压力(附录6–8)。如果异义替换(K_a)和同义替换(K_s)比值大于1,则认为有正选择效应;如果 $K_a/K_s=1$,认为存在中性选择;如果 $K_a/K_s<1$,则认为有纯化选择作用(Nekrutenko et al., 2002)。结果显示, *BnZF-HD5*、*BnZF-HD34*、*BnZF-HD53*、*BnZF-HD58*和

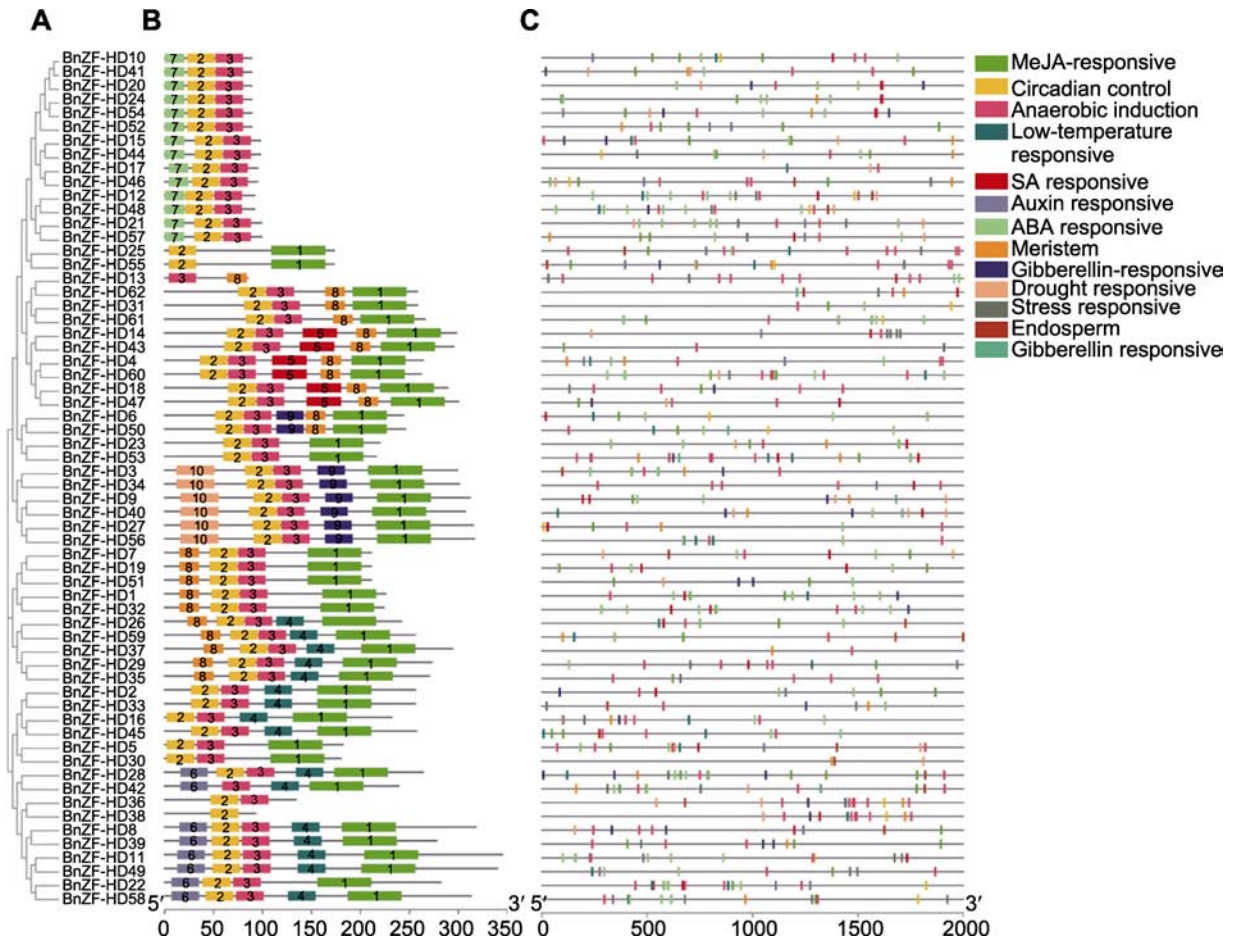


图3 甘蓝型油菜*ZF-HD*蛋白家族的聚类(A)、motif结构(B)和顺式作用元件(C)

Figure 3 Phylogenetic analysis (A), motif structure (B) and *cis*-acting element (C) of different *ZF-HD* protein families in *Brassica napus*

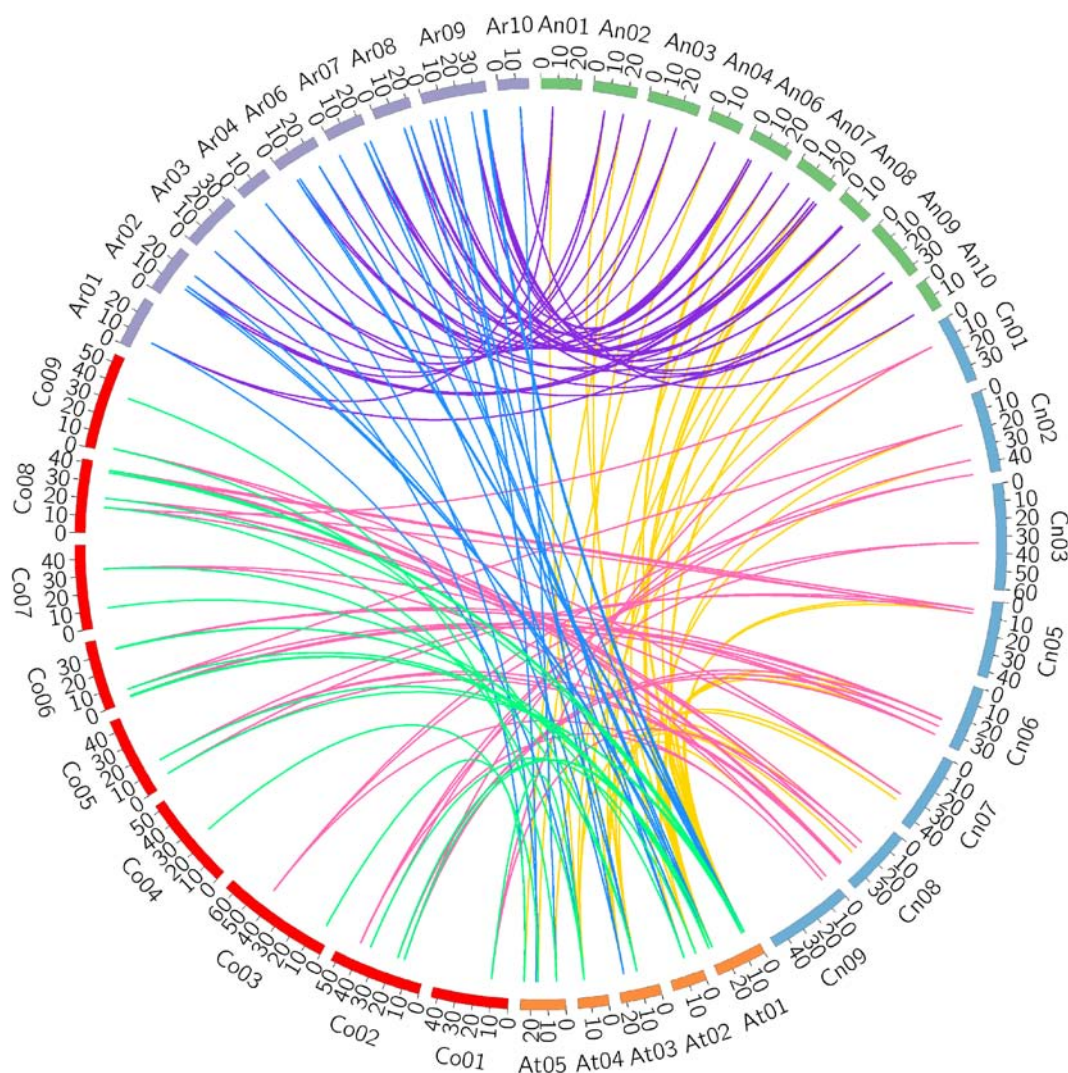


图4 甘蓝型油菜、白菜、甘蓝与拟南芥ZF-HD基因的染色体共线性关系

Figure 4 Collinear relationship of ZF-HD genes among *Brassica napus*, *B. rapa*, *B. oleracea*, and *Arabidopsis thaliana* chromosomes

*BnZF-HD5*与相应共线同源基因的 K_a/K_s 值大于1,说明这5个*BnZF-HD*基因受到正选择压力的影响。正选择可以促进基因发生有利突变。其余基因的 $K_a/K_s < 1$,说明它们经历了纯化选择。*C*亚基因组*BnZF-HD*基因的 K_a/K_s 平均值高于在*A*亚基因组的基因,表明*C*亚基因组上的*BnZF-HD*基因在进化过程中承受了较高的选择压力。

2.7 顺式作用元件分析

基因启动子区域的顺式作用元件对于明确特定环境下基因的组织特异性或应激反应表达模式具有重要作用(Yamaguchi-Shinozaki and Shinozaki,

2005)。据报道,顺式元件密度与具有多重应激反应的基因之间显著正相关(Walther et al., 2007)。通过Plantcare在线分析工具对ZF-HD基因起始密码子上游2 000 bp的序列进行顺式作用元件搜索,去除一般性转录调控元件和功能未知元件,共检测到1 501个元件,其中光响应元件最多,总计752个,分布在所有*BnZF-HD*基因中。每个基因具有的光响应元件数量不一,*BnZF-HD3*的光响应元件最多(27个);*BnZF-HD52*最少(2个)。我们进一步分析了激素、逆境响应元件和组织特异性表达元件的分布(图3C),在43、47和57个*BnZF-HD*基因的启动子区分别检测到MeJA、ABA和厌氧(anaerobic)诱导元件,具有GA、SA、生

长素、低温、干旱和胁迫响应元件的基因数量在25–40之间不等,不足20个*BnZF-HD*基因有响应光周期元件、分生组织和胚乳表达元件。同一亚群的基因顺式元件分布模式并未呈现一致性和规律性。

2.8 讨论

*ZF-HD*基因在植物多种发育和生理过程中发挥作用。研究发现*ZF-HD*基因家族只存在于陆生植物中(Wang et al., 2016),这类基因可能在陆生植物祖先与单细胞藻类趋异后得到了进化。本研究在甘蓝型油菜中鉴定到62个*ZF-HD*基因,参考其它植物中的分类(Hu et al., 2008),发现MIF与*ZF-HD*蛋白相似度高,虽然缺乏HD结构域,也将14个MIF基因纳入*ZF-HD*基因家族的范畴。与拟南芥、白菜和甘蓝的*ZF-HD*蛋白聚类分析后,分成6个亚群,MIF单独形成1个亚群。同一亚群的motif分布模式基本一致,如MIF都具有相同的3个基序,不同亚群之间motif分布差异较大。甘蓝型油菜*ZF-HD*基因大多缺乏内含子,这与拟南芥、水稻、葡萄及白菜等植物中的研究结果一致(Hu et al., 2008; Wang et al., 2014, 2016)。

多倍化是物种进化的重要力量。比较基因组学研究显示,大约2千万年前与拟南芥分离后,芸薹属物种经历了基因组水平的三倍化(Chalhoub et al., 2014; Liu et al., 2014; Yang et al., 2016)。基于三倍化假说,每个拟南芥的*ZF-HD*基因对应3个白菜或甘蓝的同源基因,而实际上只有*AtZHD1/5/17*在白菜和甘蓝基因组各有3个同源基因,其它*AtZHD*基因在白菜和甘蓝基因组中有1或2个拷贝(附录5),表明白菜和甘蓝基因在芸薹属三倍化过程中发生了基因丢失。白菜与甘蓝的分化大约发生在400万年以前(Liu et al., 2014),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)。染色体加倍以后,理论上每个拟南芥*ZF-HD*基因在甘蓝型油菜中应该有6个同源基因,且甘蓝型油菜*ZF-HD*基因等于白菜和甘蓝的同源基因数目之和。实际上除了*AtZHD17*在甘蓝型油菜有6个同源基因外,其余*AtZHD*基因在甘蓝型油菜中的同源基因数目不超过4个,说明*ZF-HD*基因在芸薹属三倍化后发生了丢失。推测丢失的原因可能有以下3个:(1)芸薹属三倍化以后,在二倍体白菜和甘蓝祖先中发生了基因丢失,从而导致甘蓝型油菜中共线性同源基因丢失,如*AtZ-*

*HD3*和*AtZHD14*在白菜和甘蓝中各有1个共线性同源基因,在油菜中有2个共线性同源基因;(2)在异源四倍体化以后发生丢失,如*AtZHD1*和*AtZHD5*在白菜和甘蓝中各有3个共线性同源基因,在油菜中只有3或4个共线性同源基因;(3)两个过程中都有丢失现象,如*AtZHD13*在白菜和甘蓝中各有1个共线性同源基因,但油菜中没有相应的共线性同源基因,这可能与三倍化过程中发生大规模的基因丢失和染色体重排有关(Mun et al., 2009; 高虎虎等, 2017)。但总体来看,62个*BnZF-HD*基因中有48个与拟南芥、白菜和甘蓝存在共线性关系,从白菜和甘蓝祖先基因组各继承了24个,表明甘蓝型油菜*ZF-HD*基因家族扩张的主要力量是异源四倍体化。

比较基因组研究表明,在白菜和甘蓝基因组的不同区域,基因密度与表达水平存在差异,可划分为LF (least fractionated)、MF1 (medium fractionated)和MF2 (most fractionated) 3个亚基因组(subgenome) (Wang et al., 2011; Liu et al., 2014)。与拟南芥*ZF-HD*基因相比,白菜、甘蓝及甘蓝型油菜中大多数保留下来的*ZF-HD*基因位于LF亚基因组,这与整个基因组的保留趋势一致,其次是MF2亚基因组,MF1亚基因组保留的*ZF-HD*基因最少,而基因组整体水平上MF1亚基因组基因丢失的数目最多。

综上,本研究在甘蓝型油菜全基因组中鉴定到62个*ZF-HD*基因,这些基因在18条染色体上呈不均匀分布,A亚基因组和C亚基因组的基因数量相等。同一亚群的motif模式极为相似,异源多倍体化对*ZF-HD*基因的扩张起重要作用。甘蓝型油菜*ZF-HD*基因的顺式作用元件以光、MeJA、ABA和厌氧诱导元件为主。本研究结论为深入揭示*BnZF-HD*基因功能奠定了基础,同时为多基因家族在异源多倍体中的进化模式分析提供资料。

参考文献

- 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.
- Abdullah M, Cheng X, Cao YP, Su XQ, Manzoor MA, Gao JS, Cai YP, Lin Y (2018). Zinc finger-homeodomain transcriptional factors (ZHDs) in upland cotton (*Gossypium hirsutum*): genome-wide identification and expression

- analysis in fiber development. *Front Genet* **9**, 357.
- Ariel FD, Manavella PA, Dezar CA, Chan RL (2007). The true story of the HD-Zip family. *Trends Plant Sci* **12**, 419–426.
- Bailey TL, Johnson J, Grant CE, Noble WS (2015). The MEME suite. *Nucleic Acids Res* **43**, W39–W49.
- Bateman A, Birney E, Cerruti L, Durbin R, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer ELL (2002). The Pfam protein families database. *Nucleic Acids Res* **30**, 276–280.
- Bhattacharjee A, Ghangal R, Garg R, Jain M (2015). Genome-wide analysis of homeobox gene family in legumes: identification, gene duplication and expression profiling. *PLoS One* **10**, e0119198.
- Bhattacharjee A, Jain M (2013). Homeobox genes as potential candidates for crop improvement under abiotic stress. In: Tuteja N, Singh Gill S, eds. *Plant Acclimation to Environmental Stress*. New York: Springer. pp. 163–176.
- Chalhoub B, Denoeud F, Liu SY, Parkin IAP, Tang HB, Wang XY, Chiquet J, Belcram H, Tong CB, Samans B, Corr  a 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, Schnell 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, B  rard 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, Snowdon 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.
- Chen CJ, Xia R, Chen H, He YH (2018). TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *bioRxiv* [2018-03-27]. <https://www.biorxiv.org/content/10.1101/289660v1>.
- 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.
- Cheng F, Wu J, Fang L, Wang XW (2012). Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. *Front Plant Sci* **3**, 198.
- Deng X, Phillips J, Meijer AH, Salamini F, Bartels D (2002). Characterization of five novel dehydration-responsive homeodomain leucine zipper genes from the resurrection plant *Craterostigma plantagineum*. *Plant Mol Biol* **49**, 601–610.
- Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279–D285.
- He ZL, Zhang HK, Gao SH, Lercher MJ, Chen WH, Hu SN (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res* **44**, W236–W241.
- Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**, 1296–1297.
- Hu W, DePamphilis CW, Ma H (2008). Phylogenetic analysis of the plant-specific Zinc finger-Homeobox and Mini zinc finger gene families. *J Integr Plant Biol* **50**, 1031–1045.
- Hu W, Ma H (2006). Characterization of a novel putative zinc finger gene *MIF1*: involvement in multiple hormonal regulation of *Arabidopsis* development. *Plant J* **45**, 399–422.
- Jain M, Tyagi AK, Khurana JP (2008). Genome-wide identification, classification, evolutionary expansion and expression analyses of homeobox genes in rice. *FEBS J* **275**, 2845–2861.
- Jin JP, Tian F, Yang DC, Meng YQ, Kong L, Luo JC, Gao G (2017a). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res* **45**, D1040–D1045.
- Jin XL, Ren J, Nevo E, Yin XG, Sun DF, Peng JH (2017b). Divergent evolutionary patterns of NAC transcription factors are associated with diversification and gene duplications in angiosperm. *Front Plant Sci* **8**, 1156.
- J  rgensen JE, Gr  nlund M, Pallisgaard N, Larsen K, Marcker KA, Jensen E   (1999). A new class of plant homeobox genes is expressed in specific regions of determinate symbiotic root nodules. *Plant Mol Biol* **40**, 65–77.
- Khadiza K, Nath UK, Robin AHK, Park JI, Lee DJ, Kim MB, Kim CK, Lim LB, Nou IS, Chung MY (2017). Genome-wide analysis and expression profiling of zinc finger homeodomain (ZHD) family genes reveal likely roles in organ development and stress responses in tomato. *BMC Genomics* **18**, 695.
- Krishna SS, Majumdar I, Grishin NV (2003). SURVEY AND SUMMARY: structural classification of zinc fingers. *Nucleic Acids Res* **31**, 532–550.
- Krzywinski M, Schein J, Birol   , Connors J, Gascoyne R,

- Horsman D, Jones SJ, Marra MA (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645.
- Kumar S, Stecher G, Tamura K (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**, 1870–1874.
- 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.
- Mackay JP, Crossley M (1998). Zinc fingers are sticking together. *Trends Biochem Sci* **23**, 1–4.
- Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, Yu HJ, Kim DS, Kim N, Lim KB, Lee SI, Hahn JH, Lim YP, Bancroft I, Park BS (2009). Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol* **10**, 211.
- Nam J, Nei M (2005). Evolutionary change of the numbers of homeobox genes in bilateral animals. *Mol Biol Evol* **22**, 2386–2394.
- Nekrutenko A, Makova KD, Li WH (2002). The K_A/K_S ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. *Genome Res* **12**, 198–202.
- Park HC, Kim ML, Lee SM, Bahk JD, Yun DJ, Lim CO, Hong JC, Lee SY, Cho MJ, Chung WS (2007). Pathogen-induced binding of the soybean zinc finger homeodomain proteins GmZF-HD1 and GmZF-HD2 to two repeats of ATTA homeodomain binding site in the *calmodulin isoform 4* (*GmCaM4*) promoter. *Nucleic Acids Res* **35**, 3612–3623.
- Schultz J, Milpetz F, Bork P, Ponting CP (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* **95**, 5857–5864.
- Takatsuji H (1999). Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. *Plant Mol Biol* **39**, 1073–1078.
- Tan QKG, Irish VF (2006). The *Arabidopsis* zinc finger-homeodomain genes encode proteins with unique biochemical properties that are coordinately expressed during floral development. *Plant Physiol* **140**, 1095–1108.
- Tran LSP, Nakashima K, Sakuma Y, Osakabe Y, Qin F, Simpson SD, Maruyama K, Fujita Y, Shinozaki K, Yamaguchi-Shinozaki K (2007). Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the *ERD1* gene in *Arabidopsis*. *Plant J* **49**, 46–63.
- Walther D, Brunnemann R, Selbig J (2007). The regulatory code for transcriptional response diversity and its relation to genome structural properties in *A. thaliana*. *PLoS Genet* **3**, e11.
- Wang H, Yin XJ, Li XQ, Wang L, Zheng Y, Xu XZ, Zhang YC, Wang XP (2014). Genome-wide identification, evolution and expression analysis of the grape (*Vitis vinifera* L.) zinc finger-homeodomain gene family. *Int J Mol Sci* **15**, 5730–5748.
- Wang W, Wu P, Li Y, Hou X (2016). Genome-wide analysis and expression patterns of ZF-HD transcription factors under different developmental tissues and abiotic stresses in Chinese cabbage. *Mol Gen Genomics* **291**, 1451–1464.
- 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, Weissshaar B, Liu BH, Li B, Liu B, Tong CB, Song C, Duran C, Peng CF, Geng CY, Koh C, Lin CY, Edwards D, Mu D, 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, Poulain 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 S, 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 meso-

polyploid crop species *Brassica rapa*. *Nat Genet* **43**, 1035–1039.

- Windhövel A, Hein I, Dabrowa R, Stockhaus J (2001). Characterization of a novel class of plant homeodomain proteins that bind to the C_4 phosphoenolpyruvate carboxylase gene of *Flaveria trinervia*. *Plant Mol Biol* **45**, 201–214.
- Yamaguchi-Shinozaki K, Shinozaki K (2005). Organization of *cis*-acting regulatory elements in osmotic- and cold-

stress-responsive promoters. *Trends Plant Sci* **10**, 88–94.

- 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.

Genome-wide Identification and Phylogenetic Analysis of Zinc Finger Homeodomain Family Genes in *Brassica napus*

Min Song*, Yao Zhang, Liying Wang, Xiangyong Peng

School of Life Sciences, Qufu Normal University, Qufu 273165, China

Abstract Zinc finger-homeodomain (ZF-HD) genes encode plant-specific transcription factors that participate in the regulation of plant growth, development and abiotic stress response. In this study, a total of 62 *ZF-HD* genes were identified in *Brassica napus*. Among them, 83.9% lack introns, 93.5% of deduced proteins have the isoelectric point greater than 7. Most of the BnZF-HD proteins with nuclear localization signal were predicted to be more than 100 amino acid residues. Based on phylogenetic analysis, this gene family can be divided into six subgroups. In each subgroup, the number of *ZF-HD* genes in *B. oleracea* and *B. rapa* was the same or nearly so, while the number of *ZF-HD* genes in *B. napus* was close to or equal to the sum of *ZF-HD* genes in *B. oleracea* and *B. rapa*. The number and type of motifs in each subgroup were highly conserved. Collinear analysis indicated that the whole genome triplication allowed the *ZF-HD* gene family to expand in diploid ancestors, and allopolyploidization made the *ZF-HD* gene family to reexpand in *B. napus*. K_a/K_s analysis showed that the *ZF-HD* gene family has experienced purification selection in the evolution process. All BnZF-HD genes have light response elements, and two-thirds of them have MeJA, ABA and anaerobic inducible *cis*-acting elements. It is speculated that they may participate in related biological processes. This study laid a foundation for further exploring the molecular mechanism of this family in regulating plant growth and stress responses, and provided more reference for future research on the evolution of plant polyploidization.

Key words *Brassica napus*, ZF-HD, evolution, collinear analysis, *cis*-acting elements

Song M, Zhang Y, Wang LY, Peng XY (2019). Genome-wide identification and phylogenetic analysis of zinc finger homeodomain family genes in *Brassica napus*. *Chin Bull Bot* **54**, 699–710.

* Author for correspondence. E-mail: smin2000@qfnu.edu.cn

(责任编辑: 白羽红)

附录 1 甘蓝型油菜 *ZF-HD* 基因名称、ID 和基本信息

Appendix 1 Name, ID and various features of *ZF-HD* genes in *Brassica napus*

附录 2 白菜、甘蓝和拟南芥 *ZF-HD* 基因

Appendix 2 The members of *ZF-HD* gene family identified in *Brassica rapa*, *B. oleracea* and *Arabidopsis thaliana*

附录 3 基于最大似然(ML)法构建的甘蓝型油菜、拟南芥、白菜和甘蓝的 *ZF-HD* 基因系统进化树

Appendix 3 Phylogenetic tree of *Brassica napus*, *Arabidopsis thaliana*, *B. rapa* and *B. oleracea* *ZF-HD* genes constructed by the maximum likelihood (ML) method

附录 4 甘蓝型油菜 ZF-HD 蛋白 motif 保守氨基酸序列

Appendix 4 Motif sequences identified of ZF-HD proteins in *Brassica napus*

附录 5 ZF-HD 基因在白菜、甘蓝、甘蓝型油菜和拟南芥共线性区的同源基因

Appendix 5 Homologous ZF-HD genes in collinear regions of *Brassica rapa*, *B. oleracea*, *B. napus* and *A. thaliana* genomes

附录 6 甘蓝型油菜与拟南芥 ZF-HD 共线性基因的 K_a/K_s 比值

Appendix 6 The K_a/K_s ratios for orthologous ZF-HD genes between *Brassica napus* and *Arabidopsis thaliana*

附录 7 甘蓝型油菜与白菜 ZF-HD 共线性基因的 K_a/K_s 比值

Appendix 7 The K_a/K_s ratios for orthologous ZF-HD genes between *Brassica napus* and *B. rapa*

附录 8 甘蓝型油菜与甘蓝 ZF-HD 共线性基因的 K_a/K_s 比值

Appendix 8 The K_a/K_s ratios for orthologous ZF-HD genes between *Brassica napus* and *B. oleracea*

<http://www.chinbullbotany.com/fileup/PDF/t19055.pdf>

致谢审者

本刊编辑部衷心感谢以下审稿专家对我刊工作和支持和帮助！
(统计时间2018年11月1日至2019年11月1日。按姓名拼音排序)

白明义	白书农	白永飞	包 颖	曹扬荣	产祝龙	陈 凡	陈 惠	陈建群	陈凌云	陈明训
陈学伟	陈益芳	陈振夏	陈之端	程祝宽	迟 伟	种 康	储成才	崔洪霞	崔素霞	崔 勇
戴绍军	戴思兰	丁艳锋	丁杨林	丁 勇	丁兆军	范培格	付三雄	傅向东	傅 纓	高彩吉
高俊平	苟小平	顾红雅	郭 柯	郭 林	郭龙彪	郭文冰	郭亚龙	国春策	韩天富	何光存
何国振	何军贤	何晓明	侯岁稳	胡 江	黄慧润	黄建辉	黄小贞	黄小珍	贾玉鑫	姜博晨
姜闯道	姜里文	蒋才富	蒋德安	蒋细旺	焦雨铃	金京波	康菊清	康 明	孔 瑾	匡廷云
李传友	李春香	李 刚	李红菊	李洪清	李宏伟	李继刚	李来庚	李 玲	李美茹	李明军
李少芳	李思锋	梁成真	梁振昌	廖 红	廖景平	林荣呈	凌宏清	刘 宝	刘宝辉	刘长宁
刘凤霞	刘宏涛	刘静妍	刘秀群	刘玉乐	卢庆陶	卢善发	卢艳丽	鲁迎青	栾维江	马伯军
毛传澡	毛同林	孟 征	庞永珍	彭连伟	漆小泉	钱 前	秦 峰	秦国政	邱全胜	冉进华
饶玉春	任东涛	尚富德	沈锦波	施和平	施 慧	施怡婷	宋经元	孙丽莉	孙蒙祥	谭保才
唐定中	滕 胜	田 江	田世平	童红宁	万迎朗	王国栋	王国宏	王华芳	王慧中	王克剑
王 雷	王利军	王亮生	王勳骋	王宁宁	王鹏程	王若仲	王 台	王小菁	王永红	王 瑜
吴 鸿	夏 然	萧浪涛	谢 旗	谢庆军	辛明明	徐桂霞	徐明良	徐夙侠	徐云远	薛大伟
严顺平	阳成伟	杨淑华	杨永青	杨玉盛	杨元合	杨仲南	叶柯邑	印莉萍	袁晓辉	曾宋君
章成君	张光恒	张 鹏	张战营	赵桂仿	赵 忠	郑元润	钟春梅	钟 声	周华鹏	周庆源
周世良	周奕华	朱立煌	宗 媛	左建儒						