

小麦 FLA 基因家族鉴定及粒重相关基因单倍型分析

王建飞^{1,2}, 郭利建¹, 高维东¹, 胡城祯^{1,2}, 车卓¹, 陈涛^{1,2}, 张沛沛¹, 杨德龙^{1,2}

(1. 省部共建干旱生境作物学国家重点实验室, 甘肃兰州 730070; 2. 甘肃农业大学生命科学技术学院, 甘肃兰州 730070)

摘要: 类成束阿拉伯半乳糖蛋白 (fasciclin-like arabinogalactan-protein, FLA) 是一类高度糖基化的蛋白, 广泛参与植物生长发育及抗逆境胁迫等途径, 但其在小麦中的作用机制尚不清楚。为了解 FLA 在小麦中的分布及其生物学功能, 利用生物信息学手段在全基因组水平鉴定了小麦 FLA 家族成员并对其进行了系统分析, 利用 RNA-Seq 及 qRT-PCR 检测 *TaFLAs* 基因在不同组织和逆境胁迫中的表达模式, 通过单倍型分析候选基因与粒重相关性状的关系。结果表明, 在小麦中共鉴定到 71 个 *TaFLAs* 基因, 分布在 17 条染色体上, 依次命名为 *TaFLA1*~*TaFLA71*。71 个基因按照亲缘关系被聚类为 8 个亚族, 同一亚族的不同成员间具有相似的基因结构和保守基序, 推测其具有相似的功能。通过种内同线性分析, 71 个 *TaFLAs* 基因存在 85 个片段复制基因对, 其中 30 对为基因间重复, 55 对为基因内重复; 通过种间共线性分析, *TaFLAs* 基因与二穗短柄草的 *BdFLAs* 亲缘关系较近。经 RNA-Seq 及 qRT-PCR 分析, *TaFLAs* 在不同组织、非生物胁迫中普遍表达, 其中 *TaFLA15*、*TaFLA34* 和 *TaFLA62* 在穗和籽粒中表达量较高, *TaFLA5*、*TaFLA15*、*TaFLA34* 和 *TaFLA62* 在 PEG 6000 处理下上调表达, *TaFLA5* 和 *TaFLA34* 在 NaCl 处理下上调表达。针对 *TaFLA62* 开发了 SNP 分子标记, 通过结合地理时空分布进行关联分析, *TaFLA62-Hap II* 是增加粒重的优异单倍型, 且在小麦育种过程中受到正向选择。综合来看, 小麦 *TaFLAs* 的基因结构相似, 大多数能够响应逆境胁迫, 其 *TaFLA62-Hap II* 单倍体型可用于小麦产量性状的改良。

关键词: 小麦; FLA 基因家族; 表达模式; 单倍型分析; SNP 标记; 千粒重

中图分类号: S512.1; S330

文献标识码: A

文章编号: 1009-1041(2025)10-1293-13

Genome-Wide Identification of FLA Gene Family and Haplotype Analysis of Kernel Weight-Related Genes in Wheat (*Triticum aestivum* L.)

WANG Jianfei^{1,2}, GUO Lijian¹, GAO Weidong¹, HU Chengzhen^{1,2}, CHE Zhuo¹,
CHEN Tao^{1,2}, ZHANG Peipei¹, YANG Delong^{1,2}

(1. State Key Laboratory of Aridland Crop Science, Lanzhou, Gansu 730070, China;

2. College of Life Science and Technology, Gansu Agricultural University, Lanzhou, Gansu 730070, China)

Abstract: Fasciclin-like arabinogalactan protein (FLA), a class of highly glycosylated proteins, are widely involved in plant growth, development and stress responses, yet their functional mechanisms in wheat remain unclear. To investigate the members and biological roles of FLA in wheat, this study conducted genome-wide identification and systematic analysis of the wheat FLA family using bioinformatics approaches. The RNA-Seq and qRT-PCR were employed to examine the expression patterns of *TaFLAs* across tissues and under abiotic stresses, while haplotype analysis was performed to explore the association between candidate genes and kernel weight-related traits. A total of 71 *TaFLAs* were

收稿日期: 2025-02-21 修回日期: 2025-02-28

基金项目: 甘肃省农业农村厅种业攻关项目 (GYGG-2024-2); 甘肃省科技重大专项 (22ZD6NA009); 甘肃省科技厅基础研究创新群体项目 (24JRRA633); 国家自然科学基金项目 (32360518, 32260520, 32160487); 甘肃省高校科研创新平台重点培育项目 (2024CXPT-01); 中央引导地方科技发展资金项目 (23ZYQA0322); 甘肃省高等学校产业支撑计划项目 (2022CYZC-44); 甘肃省教育厅青年博士支持项目 (2024QB-062); 省部共建干旱生境作物学国家重点实验室开放基金项目 (GSCS-2023-07); 甘肃农业大学博士科研启动项目 (GAU-KYQD-2020-6); 甘肃省教育厅 2025 年研究生创新之星项目 (2025CXZX-853)

第一作者 E-mail: 17899318708@163.com (王建飞)

通讯作者 E-mail: yangdl@gsau.edu.cn (杨德龙)

identified in wheat, distributed across 17 chromosomes and designated *TaFLA1* to *TaFLA71*. Phylogenetic analysis classified these genes into eight subfamilies, with members within the same subfamily exhibiting conserved gene structures and motifs, suggesting functional similarity. Intraspecific synteny analysis revealed 85 segmental duplication pairs (30 interchromosomal and 55 intrachromosomal duplications), while interspecific synteny highlighted closer evolutionary relationships between *TaFLAs* and *BdFLAs* from *Brachypodium distachyon*. The RNA-Seq and qRT-PCR results demonstrated widespread expression of *TaFLAs* across tissues and under stress conditions. Notably, *TaFLA15*, *TaFLA34* and *TaFLA62* showed high expression in spikes and grains, while *TaFLA5*, *TaFLA15*, *TaFLA34* and *TaFLA62* were upregulated under PEG 6000 treatment; *TaFLA5* and *TaFLA34* were induced by NaCl stress. A SNP molecular marker developed for *TaFLA62* revealed, combined association analysis and geographical-temporal distribution studies, that the *TaFLA62-Hap II* haplotype significantly enhances kernel weight and has undergone positive selection during wheat breeding. Overall, the gene structure of wheat *TaFLAs* is similar, and most of them can respond to adversity stress, and its *TaFLA62-ap II* haplotype can be used to improve wheat yield traits.

Keywords: Wheat; FLA gene family; Gene expression; Haplotype analysis; SNP marker; Thousand-kernel weight

小麦 (*Triticum aestivum* L.) 为全球第二大粮食作物, 是全球 35%~40% 人口的主食, 为人类提供约 21% 的食物热量和 20% 蛋白质来源^[1]。随着全球极端气候变化、可耕地面积逐年减少、人口不断增加, 小麦可持续生产受到严重挑战, 供需矛盾日益突出, 亟需抗逆、高产育种新突破^[2]。因此, 深入挖掘小麦抗逆、高产优异基因资源, 进行分子遗传改良和新品种培育, 进一步提升小麦单产水平, 对保障全球粮食安全具有重要意义。

阿拉伯半乳聚糖蛋白 (arabinogalactan proteins, AGP) 广泛存在于植物体中, 是富含羟脯氨酸的糖蛋白^[3-4], 对植物的信号传导、抗非生物胁迫和生长发育具有重要作用^[5-7], 如花粉粒的生长发育^[8]、茎次生细胞壁的发育^[6]、根耐盐性和根毛的生长发育^[9]。根据蛋白质结构 AGP 被分为经典 AGPs (classical AGPs)、非经典 AGPs (non-classical AGPs)、富含赖氨酸的 AGPs (lys-rich AGPs)、阿拉伯半乳聚糖肽 (arabinogalactan peptides)、早期结节蛋白类 AGPs (early nodulin-like AGPs, eNod-like AGPs)、非特异性脂质转移蛋白类 AGPs (non-specific lipid transfer protein-like AGPs, nsLTP-like AGPs) 和类成束阿拉伯半乳糖蛋白 (fasciclin-like AGPs, FLAs) 七类^[8]。FLA 是 AGP 中糖基化程度最高的亚族, 含有特定的成束蛋白 (FAS) 结构域, 每个 FAS 结构域含有两个高度保守结构域 H1 和 H2^[10]。目前, 在拟南芥 (*Arabidopsis thaliana*)^[11]、烟草 (*Nicoti-*

ana tabacum L.)^[12]、水稻 (*Oryza sativa* L.)^[13]、棉花 (*Gossypium hirsutum* L.)^[14]、小麦^[10]、杨树 (*Populus* L.)^[15] 等物种中分别鉴定出 21、38、27、19、34 和 40 个 FLA。在拟南芥中, *AtFLA4* 基因可促进根细胞扩张^[16]; *AtFLA16* 基因可促进茎的次生细胞壁发育, 增加株高和茎的机械强度^[6]。*OsFLA1* 基因对水稻花粉正常发育具有重要作用^[8]。但有关小麦中 FLA 家族基因的研究相对较少, 其不同环境条件下的表达模式及其与小麦生长发育和抗逆性之间的关系尚未明确。因此, 本研究利用生物信息学方法对小麦 FLA 家族成员进行全基因组鉴定, 结合 *TaFLAs* 基因在小麦不同发育时期和非生物胁迫中的表达模式分析, 利用 WheatUnion 数据库中的小麦重测序数据和已有研究中小麦自然群体表型数据^[17] 进行重要 *TaFLA* 基因单倍型与籽粒性状间的关联分析, 以期对小麦的遗传改良和品种选育提供参考。

1 材料与方法

1.1 生物信息学分析

通过 Ensembl Plants 数据库 (<https://plants.ensembl.org/index.html>) 下载小麦、大麦 (*Hordeum vulgare* L.)、二穗短柄草 (*Brachypodium distachyon* L.)、水稻、拟南芥、玉米 (*Zea mays* L.)、马铃薯 (*Solanum tuberosum* L.) 和番茄 (*Solanum lycopersicum* L.) 的全基因组文件、蛋

白文件和注释文件,其中全基因组文件和注释文件用于种间共线性分析,蛋白文件用于提取各物种 FLA 蛋白信息和构建系统进化树。利用 HMMER 文件和 Pfam 号 (PF02469) 获取各物种的 FLA 家族成员信息,筛选阈值为 $E < 1.0 \times 10^{-5}$ 。为了进一步提高筛选结果的准确性,对筛选得到的蛋白序列分别使用 NCBI CDD、InterPro 和 SMART 进行结构分析候选蛋白,筛选出 3 个数据库中均能鉴定出 FAS 结构域的蛋白,确定为最终的 FLA 家族成员。根据 *TaFLAs* 在染色体上的分布进行命名。

利用 MEGA 11^[18] 软件中的邻接法 (Neighbor-Joining, NJ), 构建小麦、大麦、二穗短柄草、水稻、拟南芥、玉米、马铃薯和番茄 FLA 家族成员蛋白序列的系统进化树, Bootstrap number 设置为 1 000。利用 WheatOmics 1.0^[19] (<http://wheatomics.sdau.edu.cn/>) 下载 *TaFLAs* 基因的启动子序列。从小麦多组学数据库 expVIP^[20] 下载中国春的转录组数据 (RNA-Seq)。利用 TBtools^[21] 软件对基因表达量进行可视化。使用 ExPaSy^[22] 在线工具对 *TaFLAs* 进行生物信息学分析 (表 1)。

表 1 生物信息学分析所用软件的网址

Table 1 Websites of software used in bioinformatics analysis

软件 Software	用途 Function	网址 Website
NCBI CDD	蛋白结构域分析 Protein structural domains analysis	https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi
Inter Pro	蛋白结构域分析 Protein structural domains analysis	https://www.ebi.ac.uk/interpro/
SMART	蛋白结构域分析 Protein structural domains analysis	http://smart.embl-heidelberg.de/smart/set_mode.cgi?NORMAL=1
ExPaSy	蛋白质理化性质分析 Protein physicochemical properties analysis	https://web.expasy.org/compute_pi/
GSDS ^[23]	基因结构分析 Gene structure analysis	http://gsds.gao-lab.org/
ChiPlot ^[24]	进化分析 Phylogenetic analysis	https://www.chiplot.online/chitree.html
Gpos-mPLoc ^[25]	亚细胞定位预测 Subcellular localization prediction	http://www.csbio.sjtu.edu.cn/bioinf/Gposmulti/

1.2 试验材料及处理

以小麦品种中国春为材料,选取 200 粒饱满、大小均匀种子,用 10% 的次氯酸钠 (NaClO) 消毒 30 min 后,用无菌水 (ddH₂O) 冲洗 5~7 遍,置于衬有滤纸的培养皿中;加入 10 mL ddH₂O,保证完全浸润滤纸,均匀摆放 20 粒小麦种子,放置于光照培养箱 (东南 GXZ-380, 宁波东南仪器有限公司),在光周期 16 h/8 h、温度 25 °C/23 °C、湿度 60% 条件下培养。待种子萌发后,挑选萌发整齐度一致的种子,定植于含有 Hoagland 营养液^[26] 的 96 孔水培盒中,每个水培盒定植 35 株,培养条件同上。待幼苗长至两叶一心 (12 d 左右) 时,分别用 200 mmol · L⁻¹ NaCl 和 20% PEG 6000 处理小麦幼苗^[27],在处理 0 h、1 h、3 h、6 h、12 h、24 h 和 48 h 时,分别剪取小麦幼苗第一片叶,液氮速冻后储存于 -80 °C 超低温冰箱中,用于总 RNA 提取。未处理的小麦幼苗转移

至含有有机育苗营养基质土 (壮苗 1 号,甘肃绿能农业科技股份有限公司) 的花盆中 (16 cm × 16 cm × 18 cm), 每盆移栽 4 株,共 20 盆,放置于培养间,在光周期 16 h/8 h、温度 22 °C、光照强度 120 μmol · m⁻² · s⁻²、湿度 60% 条件下培养。采集小麦植株孕穗期的根、茎、叶和开花期的幼穗及花后 5 d、10 d、15 d、20 d、25 d、30 d 的籽粒,液氮速冻后储存于超低温冰箱 (-80 °C) 中,用于总 RNA 提取。每个处理 3 次生物学重复。

1.3 总 RNA 的提取和 cDNA 合成

用 TIANGEN[®] RNAprep Pure Plant Kits 试剂盒 (天根生化科技有限公司,北京) 提取 1.2 中低温保存样品的总 RNA^[28]。用 NanoOne 超微量分光光度计 (杭州佑宁仪器有限公司,杭州) 测量 RNA 浓度,并通过琼脂糖凝胶电泳检测 RNA 质量。用 TIANGEN[®] FastKing gDNA Dispelling RT SuperMix 试剂盒 (天根生化科技有限公司,

北京)合成 cDNA [29]。

1.4 引物设计与合成

利用 *Ensembl Plants* 植物数据库[30] (<https://plants.ensembl.org/index.html>) 中的注

释信息获取随机选择的 4 个小麦 *TaFLAs* 的基因序列,通过 Primer 5 软件设计引物[31]。引物由生物工程(上海)股份有限公司合成,所用引物序列见表 2。

表 2 本研究所用的引物
Table 2 Primers used in this study

引物名称 Primer name	引物序列 Primer sequence(5'→3')	用途 Purpose
TaFLA5-qRT-PCR-F	CAAGTCCGGCTCCCTCAAT	荧光定量 PCR Fluorescent quantitative PCR
TaFLA5-qRT-PCR-R	GACGTTGAGCGGGTACTTG	
TaFLA15-qRT-PCR-F	TGTTCTGCCCGAAGACAAA	
TaFLA15-qRT-PCR-R	GATGGCGTGGCTCTTGTG	
TaFLA34-qRT-PCR-F	TCGACTGGGTGTCGGTGAA	
TaFLA34-qRT-PCR-R	CTCTTCGGACACCGTTTGGT	
TaFLA62-qRT-PCR-F	GTACTGCAAGAGCTTCGCG	
TaFLA62-qRT-PCR-R	TGGCCGTCAGGTTCTTGATC	
TaGAPDH-F	GACCCAGACAACCTCGAAC	qRT-PCR 内参(组织)[32]
TaGAPDH-R	GGAATCCATGACCACCTAC	qRT-PCR reference gene (for tissue)
TaACTIN-F	CCTTCCGTGTTCCCACTGTTG	qRT-PCR 内参(胁迫)[33]
TaACTIN-R	ATGCCCTTGAGGTTTCCTC	qRT-PCR reference gene(for stress treatment)

1.5 *TaFLAs* 基因的表达分析

以 1.3 中得到的中国春幼苗不同组织和非生物胁迫处理下的 cDNA 为模版,利用 TIANGEN® FastReal qPCR PreMix(SYBR Green)荧光定量试剂盒(天根生化科技有限公司,北京)进行 *TaFLAs* 基因的表达水平分析。反应体系和扩增程序参考高维东等[29]方法。*TaGAPDH* [32] 和 *TaACTIN* [33] 为内参基因,每个样品 3 次生物学重复,用 $2^{-\Delta\Delta C_T}$ 法[34] 计算基因相对表达量。

1.6 籽粒发育关键候选基因优异单倍型分析

基于 WheatUnion 1.0 数据库(<http://wheat.cau.edu.cn/WheatUnion/>)中获取 1 769 份小麦材料的测序数据[35-37],对 1.5 中筛选到的在穗和籽粒中高表达的 *TaFLA* 基因进行单核苷酸多态性(single nucleotide polymorphism, SNP)分析和基因单倍型分析。利用已公开发表的 122 份小麦种质资源的千粒重(thousand-kernel weight, TKW)、粒长(kernel length, KL)、粒宽(kernel width, KW)和粒厚(kernel thickness, KT)的表型数据对有不同单倍型的 *TaFLA* 基因进行关联分析[17]。在 WheatUnion 1.0 数据库中筛选到来自 14 个省份(北京、甘肃、河北、河南、江苏、宁夏、青海、山东、山西、陕西、四川、西藏、新疆、云南)的 328 份小麦种质资源,用其评估不同单倍型在中国的地理分

布,其中 86 份具有育成年份,用其评估不同单倍型在不同育种时期的出现频率。

1.7 数据处理

利用 Excel 2021 整理数据,用 SPSS Statistics 25.0 进行单因素方差分析和 *t* 检验,用 Origin 2016 绘图。

2 结果与分析

2.1 小麦 FLA 家族成员分布及其蛋白理化性质分析

利用 HMMER 程序以及拟南芥与水稻的 FLA 蛋白序列在小麦基因组中进行 Blastp 蛋白同源比对,从小麦全基因组数据库中共鉴定到 71 个 *TaFLAs*,根据其在染色体上的分布依次命名为 *TaFLA1*~*TaFLA71*。71 个 *TaFLAs* 均含有 FAS 结构域(包括 N 端的 H1 结构域和 C 端的 H2 保守功能结构域)。利用 ExPaSy 分析蛋白的理化性质,发现 *TaFLAs* 家族成员的氨基酸长度介于 214(*TaFLA27*)~483 aa(*TaFLA63*),分子量介于 19.370(*TaFLA21*)~51.262 kDa(*TaFLA64*),等电点介于 4.53(*TaFLA15*)~9.93(*TaFLA70*)。对 *TaFLAs* 蛋白的亚细胞定位预测显示,17 个定位于细胞外,16 个定位于液泡中,8 个定位于内质网中,11 个定位于叶绿体中,9 个

定位于质膜上,5 个定位于线粒体中,5 个定位于细胞核中。

2.2 多物种 FLA 家族成员系统进化分析

以 71 个小麦 TaFLAs、29 个大麦 HvFLAs、29 个二穗短柄草 BdFLAs、28 个水稻 OsFLAs 和 34 个玉米 ZmFLAs 作为单子叶禾本科代表,以 22 个拟南芥 AtFLAs、24 个番茄 SIFLAs 和 29 个马铃薯 StFLAs 作为双子叶代表,利用 MEGA11 构建系统进化树。结果(图 1)显示,所有 FLA 蛋白可分为 9 个亚族(Group I ~ IX)。其中,Group I 中没有 TaFLA 成员分布,Group II ~ IX 中分别含有 6、12、6、10、7、5、10 和 15 个 TaFLA 成员。与其他物种相比,TaFLAs 在一些

亚族中数量较多,如在 Group V 中,小麦有 10 个成员,大麦有 3 个,无水稻和拟南芥成员;在 Group VI 中,小麦有 7 个成员,大麦有 2 个,二穗短柄草只有 1 个。小麦是异源六倍体植物,其 3 个同源基因对应二倍体模式植物(水稻和拟南芥)中的 1 个基因,而本研究在 Group V 和 Group VI 亚族中没有发现水稻和拟南芥 FLA 蛋白。

2.3 小麦 TaFLA 家族成员的蛋白保守结构域和基因结构分析

基于系统进化树(图 1)可知,71 个 TaFLAs 归属于 8 个亚族(图 2A)。通过 MEME 在线网站对 71 个 TaFLAs 进行氨基酸序列分析,共鉴定出 20 个 motif,其中 motif 1、motif 2 和 motif 4 存在

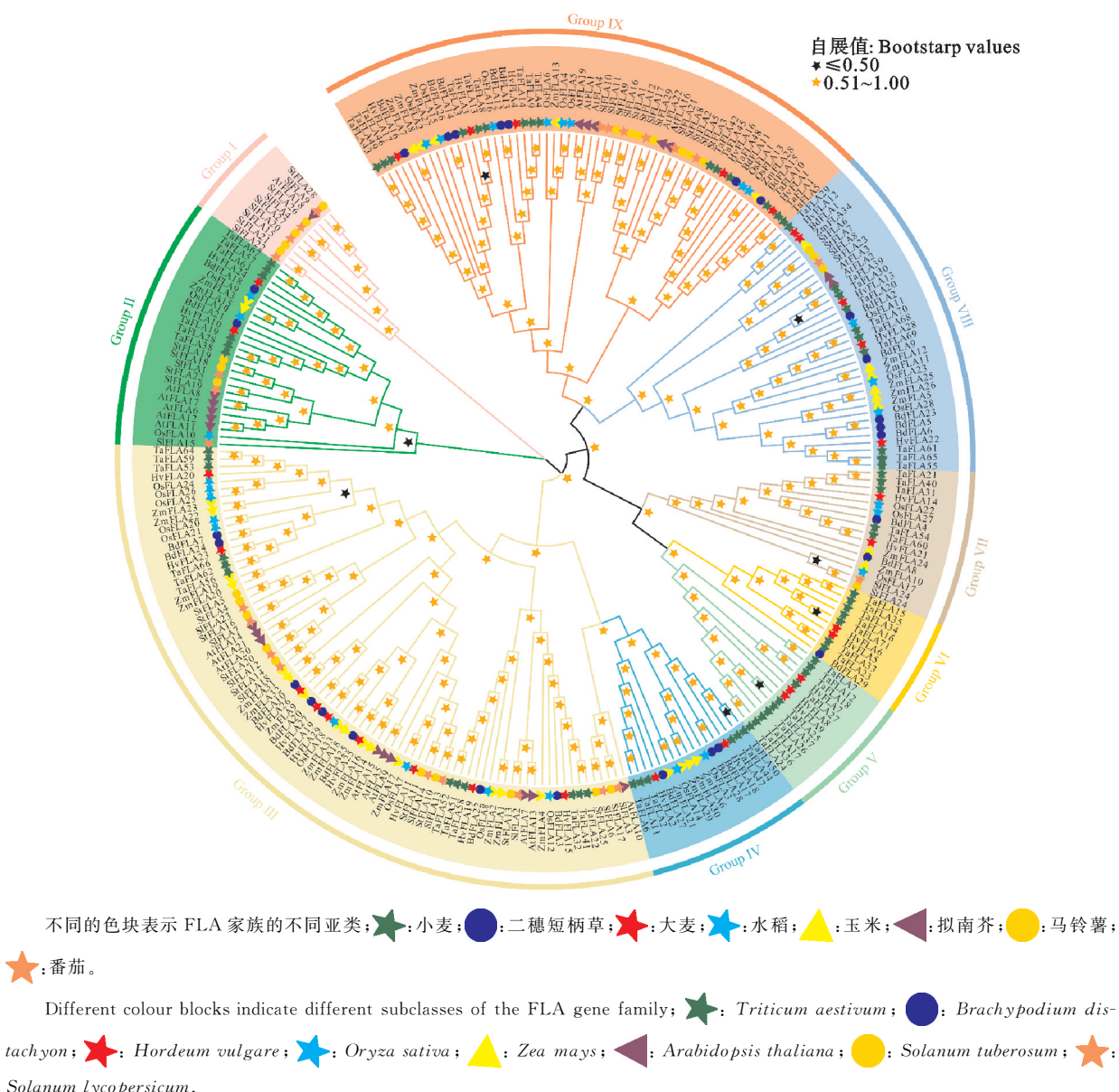


图 1 FLAs 的系统发育树
Fig. 1 Phylogenetic tree of FLAs

于所有 TaFLAs 蛋白中,它们是组成保守结构域 FAS 的核心元件,其中 motif 1 是组成 H1 的主要基序,motif 2 是组成 H2 的主要基序,motif 4 是组成[YF]H 的主要基序(图 2B)。此外,同一亚族中的 TaFLA 成员具有相似的 motif,例如 TaFLAs 蛋白的 Group II~IX 亚族中分别含有 12、10、10、4、6、6、5 和 5 个共有的 motif。

基于系统进化的分类结果,对 TaFLAs 不同亚族基因使用 GSDS 在线网站进行序列结构分析。结果显示,Group II 和 Group V 亚族中的所有基因都含有 2 个外显子和 1 个内含子;Group III 亚族中,*TaFLA66*、*TaFLA62*、*TaFLA56*、*TaF-*

LA53、*TaFLA59* 和 *TaFLA64* 含有 2 个外显子和 1 个内含子, *TaFLA51*、*TaFLA52*、*TaFLA58*、*TaFLA32*、*TaFLA22* 和 *TaFLA41* 仅含有 1 个外显子,但保守基序极为相似;Group IV 亚族中,除了 *TaFLA2* 以外,其他成员都只具有 1 个外显子;Group VI、Group VII 和 Group IX 亚族中,只有 1 个外显子;Group VIII 亚族中,除了 *TaFLA29* 和 *TaFLA68* 具有 1 个内含子和 2 个外显子以外,其他成员只具有 1 个外显子(图 2 C)。同一亚族中的大多数成员具有相似的结构,推测其具有相似功能,不同的亚族之间内含子数量和长度明显不同,这可能会促进其功能分化。

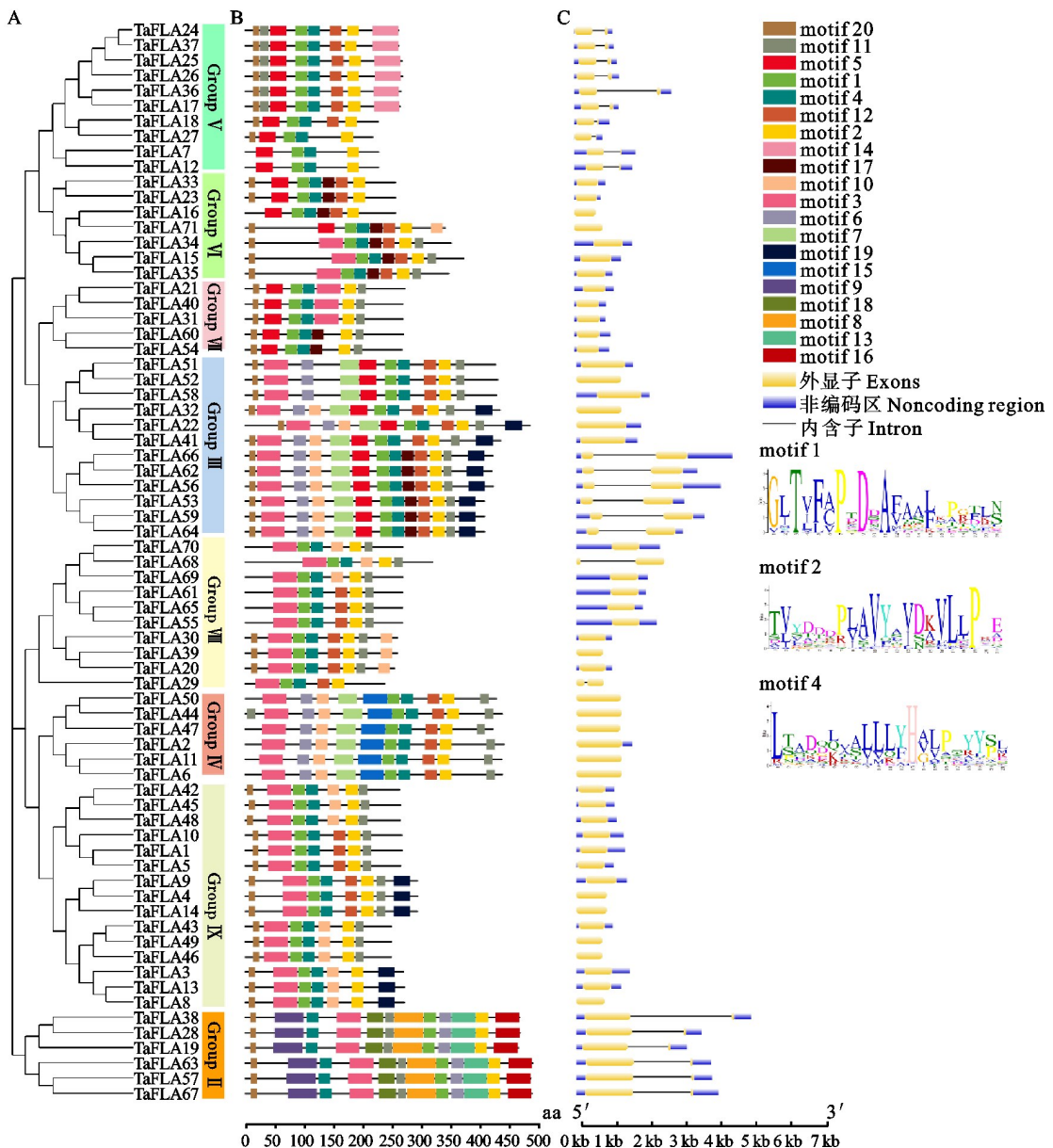


图 2 小麦 TaFLAs 的进化关系(A)、保守基序(B)和基因结构(C)。

Fig. 2 Phylogenetic relationships(A), conserved motifs(B) and gene structure(C) of TaFLAs in wheat

2.4 小麦 *TaFLAs* 的染色体分布和共线性分析

染色体分析表明,71 个小麦 *TaFLAs* 成员分布在 17 条染色体上,其中 4B、4D、7A、7B、7D 各定位到 1 个 *TaFLA* 基因,3A、3B、3D 各定位到 3 个 *TaFLAs* 基因,1A、5D 各定位到 4 个 *TaFLAs*

基因,1B、1D、5B 各定位到 5 个 *TaFLAs* 基因,5A 定位到 6 个 *TaFLAs* 基因,2A 定位到 8 个 *TaFLAs* 基因,2D 定位到 9 个 *TaFLAs* 基因,2B 定位到 10 个 *TaFLAs* 基因。此外, *TaFLA71* 基因未定位于已知染色体上(图 3)。

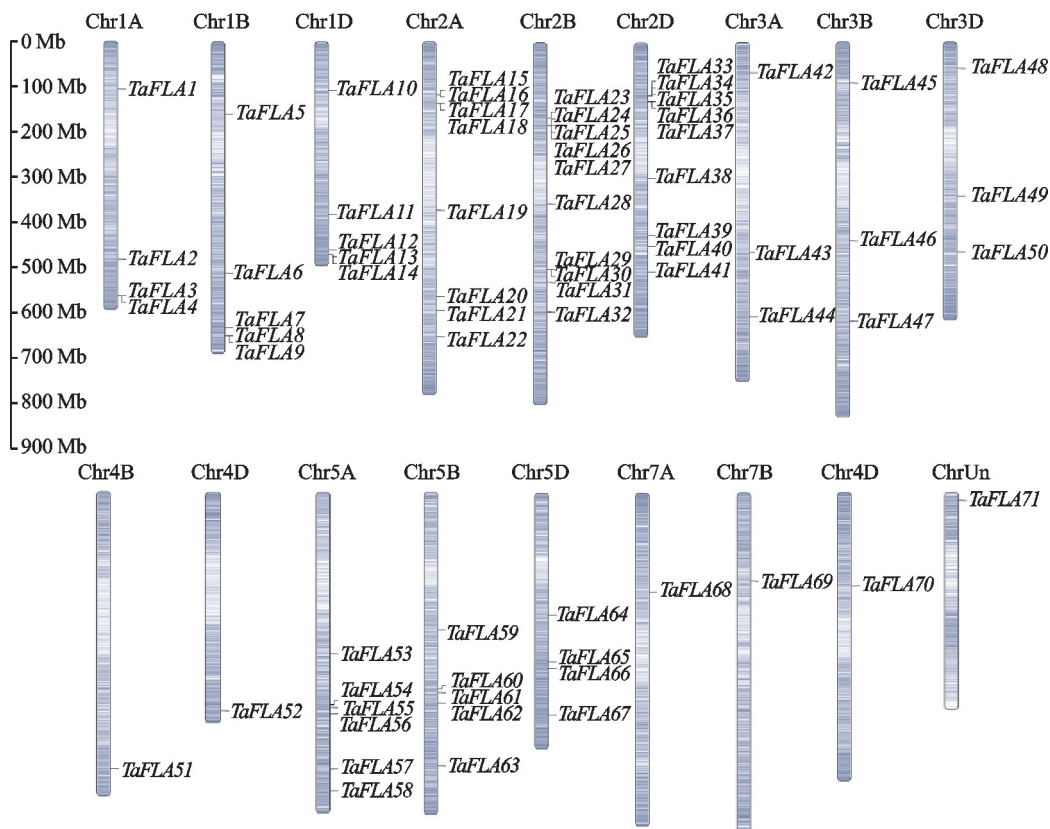


图 3 *TaFLAs* 基因在小麦染色体上的分布

Fig. 3 Distribution of *TaFLAs* genes on wheat chromosomes

种内同线性分析发现, *TaFLAs* 间存在 85 对基因重复事件,其中 30 对为基因间重复,55 对为基因内重复,均为片段复制,表明片段复制是 *TaFLAs* 基因扩张的主要原因(图 4 A)。片段复制的 K_a/K_s 值均小于 1,说明 *TaFLAs* 的进化主要发生在纯化选择下。对小麦 FLA 基因与二穗短柄草、水稻、大麦、玉米、马铃薯、番茄、拟南芥 FLA 基因进行共线性分析发现,小麦 *TaFLAs* 与二穗短柄草 *BdFLAs* 间有 84 对共线性基因,与水稻 *OsFLAs* 间有 73 对共线性基因,与大麦 *HvFLAs* 间有 23 对共线性基因,与玉米 *ZmFLAs* 间有 12 对共线性基因,与马铃薯 *StFLAs* 间有 12 对共线性基因,与番茄 *SIFLAs* 间有 6 对共线性基因,与拟南芥 *AtFLAs* 有 4 对共线性基因,说明小麦 FLAs 基因与二穗短柄草和拟南芥 FLAs 基因亲缘关系更近(图 4B)。

2.5 小麦 *TaFLAs* 在不同组织间的表达

利用小麦 RNA-Seq 数据对 *TaFLAs* 在中国春不同组织中的表达模式进行分析,发现 71 个 *TaFLAs* 基因在不同组织中表达存在显著差异 ($P < 0.05$),其中 *TaFLA20*、*TaFLA39*、*TaFLA30* 等 5 个基因在苗期、三叶期和挑旗期的根中高表达, *TaFLA67*、*TaFLA57*、*TaFLA63* 等 6 个基因在拔节期和开花期的茎中表达量较高, *TaFLA8*、*TaFLA13*、*TaFLA3* 等 6 个基因在苗期的叶中相对高表达, *TaFLA24*、*TaFLA37*、*TaFLA25* 等 17 个基因在挑旗期、开花期以及花后 2 d 的穗和籽粒中高表达(图 5A)。

为了进一步验证 *TaFLAs* 在不同组织中的表达情况,随机选择 4 个 *TaFLAs* (*TaFLA5*、*TaFLA15*、*TaFLA34*、*TaFLA62*),通过 qRT-PCR 检测其在不同组织的表达水平。结果显示, *TaFLA*

LA5 在茎中特异性高表达, 而 *TaFLA15*、*TaFLA34* 和 *TaFLA62* 主要在幼穗和花后 5 d 籽粒中高表达。以上结果与 RNA-Seq 分析结果基本一致(图 5B~5E)。

2.6 不同胁迫处理下小麦 *TaFLAs* 的表达

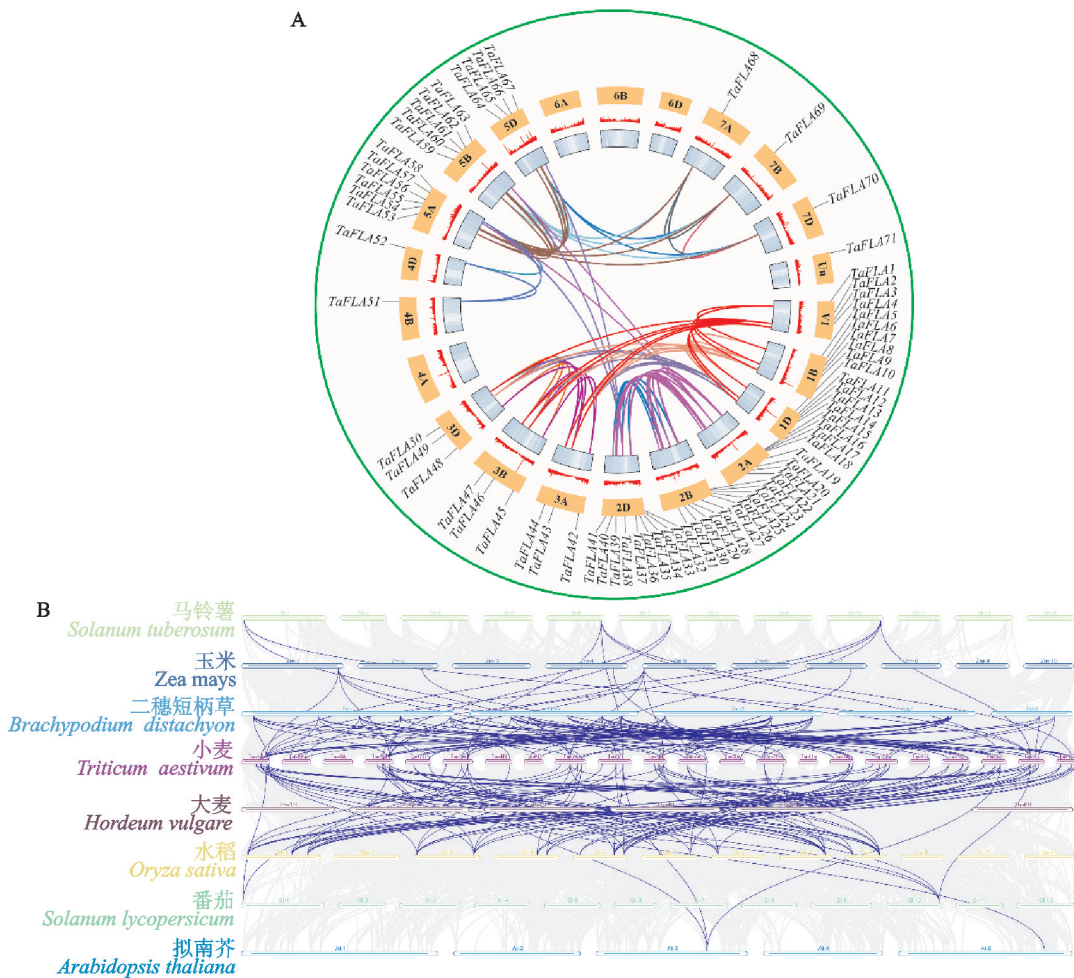
分析 *TaFLAs* 基因在干旱和盐胁迫下的表达模式, 发现 *TaFLA15* 和 *TaFLA34* 在 NaCl 处理 12 h、48 h 后表达量上调; *TaFLA5*、*TaFLA15*、*TaFLA34* 在 PEG 6000 处理 2 h 后表现出上调表达, *TaFLA4* 和 *TaFLA14* 在 PEG 6000 处理 12 h 后上调表达(图 6A)。

为了进一步验证 *TaFLAs* 基因是否响应非生物胁迫, 对 2.5 中随机选择的 4 个基因 (*TaFLA5*、*TaFLA15*、*TaFLA34*、*TaFLA62*) 通过 qRT-PCR 分别检测在 200 mM NaCl 和 20% PEG 6000 处理下的表达水平。结果显示, 在 200 mM

NaCl 处理下, *TaFLA5* 和 *TaFLA62* 的表达水平下调, 而 *TaFLA15* 和 *TaFLA34* 的表达显著上调 ($P < 0.05$); 其中, *TaFLA15* 在处理 6 h 后上调至峰值, 上调了 5.13 倍(图 6B~6E)。在 20% PEG 6000 处理下, *TaFLA5* 和 *TaFLA15* 的表达水平在 3 h 时显著升高, 而 *TaFLA62* 在 6 h 时表达水平上调(图 6F~6I)。以上结果与 RNA-Seq 分析结果基本一致。

2.7 小麦 *TaFLAs* 基因单倍型与籽粒性状关联分析

转录组和 PCR 定量分析结果均表明, *TaFLA15*、*TaFLA34*、*TaFLA62* 等基因在小麦幼穗及籽粒中显著高表达, 为了解这些基因是否参与小麦籽粒发育调控, 对 *TaFLA15*、*TaFLA34*、*TaFLA62* 基因及其启动子进行序列多态性分析, 发现 *TaFLA15* 和 *TaFLA62* 基因均存在两种单倍型。利用 122 份小麦籽粒的千粒重、粒长、粒宽和粒厚

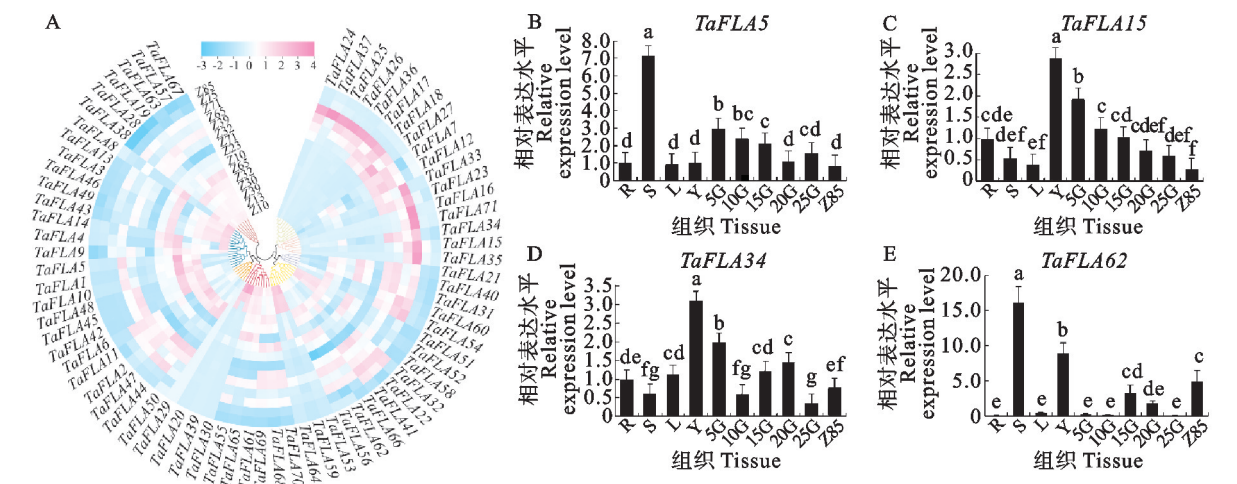


彩色连接表示片段复制; 图 A 最内圈为小麦基因组密度热图, 最外圈为小麦染色体的名称和大小。

Colored connections indicate segmental replication; Innermost circle in Fig. A shows a heat map of the wheat genome density, and the outermost circle shows the names and sizes of the wheat chromosomes.

图 4 *TaFLAs* 基因的种内共线性(A)和种间共线性(B)分析

Fig. 4 Distribution of *TaFLAs* genes analysis of intraspecific synteny(A) and interspecific collinearity(B)



R: 孕穗期的根; L: 孕穗期的叶; S: 孕穗期的茎; Y: 开花期的穗; 5G: 花后 5 d 的籽粒; 10G: 花后 10 d 的籽粒; 15G: 花后 15 d 的籽粒; 20G: 花后 20 d 的籽粒; 25G: 花后 25 d 的籽粒; Z10: 苗期的根; Z13: 三叶期的根; Z39: 挑旗期的根; Z30: 穗长 1 cm 的茎; Z32: 拔节期的茎; Z65: 开花期的茎; Z10: 苗期的叶; Z23: 分蘖期的叶; Z71: 花后 2 d 的叶; Z32: 拔节期的穗; Z39: 挑旗期的穗; Z65: 开花期的穗; Z71: 花后 2 d 的籽粒; Z75: 花后 14 d 的籽粒; Z85: 花后 30 d 的籽粒; 误差值用标准差 SD($n=3$) 表示。图柱上不同字母表示差异显著 ($P < 0.05$)。图 6 同。

R: Root at booting stage; S: Stem at booting stage; L: Leaf at booting stage; Y: Spike at anthesis; 5G: Grain at 5 d after anthesis; 10G: Grain at 10 d after anthesis; 15G: Grain at 15 d after anthesis; 20G: Grain at 20 d after anthesis; 25G: Grain at 25 d after anthesis; Z10: Roots at seedling stage; Z13: Roots at three-leaf stage; Z39: Roots at flag leaf stage; Z30: Stems at 1 cm spike length; Z32: Stems at nodule stage; Z65: Stems at anthesis stage; Z10: Leaves at seedling stage; Z23: Leaves at tillering stage; Z71: Leaves at 2 d after anthesis; Z32: Spikes at nodule stage; Z39: Spikes at flag picking stage; Z65: Spikes at anthesis stage; Z71: Grains at 2 d after anthesis; Z75: Grain at 14 d after anthesis; Z85: Grain at 30 d after anthesis; Error values are expressed as standard deviation SD($n=3$). Different letters on the columns indicate significant difference ($P < 0.05$). The same in Fig. 6.

图 5 小麦 TaFLAs 在不同发育时期 (A) 和组织中 (B~E) 的表达模式分析

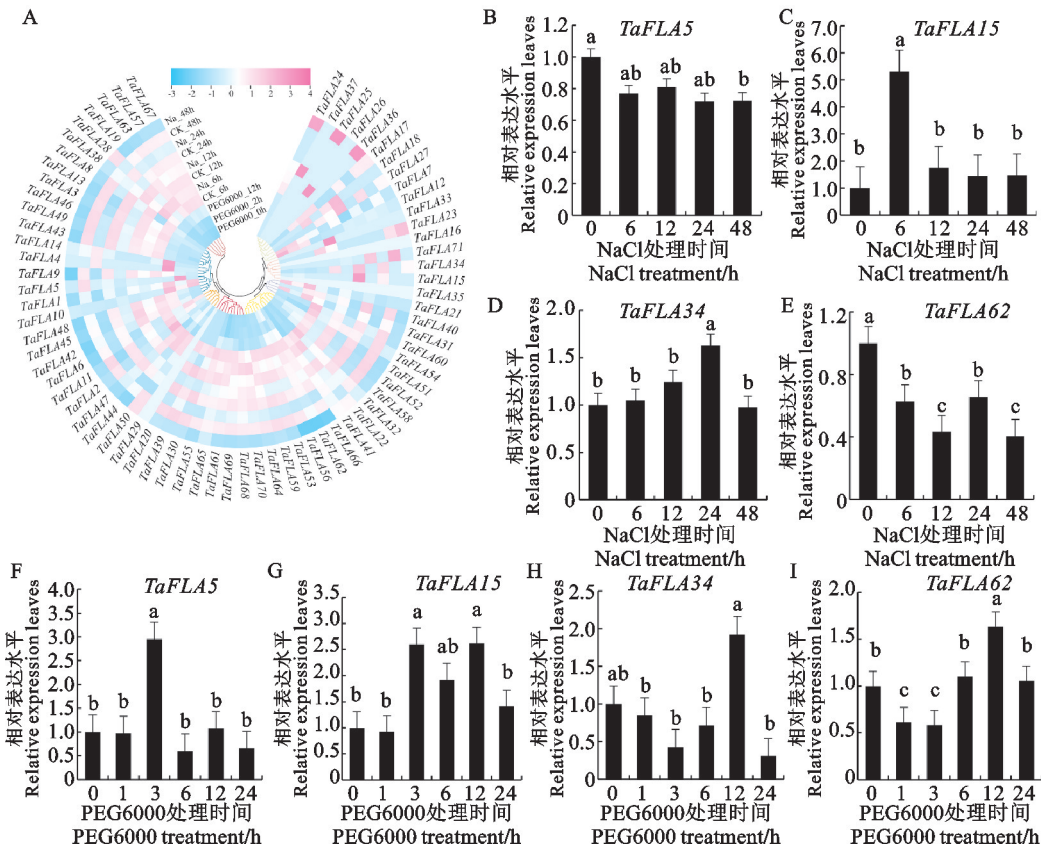
Fig. 5 Expression patterns of TaFLAs in wheat at different stages (A) and tissues (B~E).

表型数据进行关联分析发现,仅 *TaFLA62* 基因的两种单倍型之间存在显著差异 ($P < 0.05$)。 *TaFLA62* 基因在启动子区和基因编码区分别存在 4 个和 1 个 SNP 位点(图 7A)。因此,基于已公开发表的 2002 年与 2005 年河南洛阳,2010 年北京顺义)的粒重相关性状表型数据^[17],对 *TaFLA62* 基因的两种单倍型进行关联分析。结果表明,在上述 3 种环境条件下,携带 *TaFLA62-Hap II* 材料的千粒重显著高于携带 *TaFLA62-Hap I* 的材料 ($P < 0.01$);在 2 种环境条件下(2005 年、2010 年),携带 *TaFLA62-Hap II* 材料的粒长和粒宽均显著高于携带 *TaFLA62-Hap I* 的材料 ($P < 0.05$);在 2 种环境条件下(2002 年、2005 年),携带 *TaFLA62-Hap II* 材料的粒厚显著高于携带 *TaFLA62-Hap I* 的材料 ($P < 0.05$)(图 7B~图 7E),推测 *TaFLA62-Hap II* 为促进小麦籽粒发育的优异单倍型。为深入研究优异单倍型 *TaFLA62-Hap II* 在小麦育种过程中是否被积极选择,以小麦联合变异数据库获取的具有

明确选育地点的 328 份种质和具有育成年份的 88 份种质为材料,研究了 *TaFLA62* 基因两种单倍型的时空分布变化。结果显示, *TaFLA62-Hap II* 在中国小麦主产区(如河南、山东、山西和江苏等)得到广泛选择(图 7F)。将 86 份种质材料按照年份分为 5 组(1971 以前、1971—1980、1981—1990、1991—2000、和 2000 以后), *TaFLA62-Hap I* 频率依次为 28% (7/25)、25% (4/16)、24% (6/25)、25% (4/16)、0% (0/4), *TaFLA62-Hap II* 频率依次为 72%、75%、76%、75%、100%,表明在小麦育种过程中 *TaFLA62-Hap II* 被正向选择(图 7G)。

3 讨论

FLA 蛋白作为阿拉伯半乳糖亚族中高度糖基化的一类蛋白,广泛参与植物生长发育及逆境胁迫等多种生物学途径^[3,6]。目前,在小麦 EST 数据库 (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>) 中已鉴定到 34 个 TaFLAs 蛋白^[10]。但是小麦 EST 数据仅包含 580 155 个 EST,仍存在大



CK_6h,CK_12h,CK_24h,CK_48h;对照生长 6 h,12 h,24 h,48 h;Na_6h,Na_12h,Na_24h,Na_48h;用 200 mmol · L⁻¹ NaCl 处理并生长 6 h,12 h,24 h,48 h;PEG6000-0 h,PEG6000-2 h,PEG6000-12 h;PEG6000 处理后 0,2,12 h。

CK_6h, CK_12h, CK_24h, CK_48h: Chinese spring seedlings not treated with 200 mmol · L⁻¹ NaCl and grown for 6 h, 12 h, 24 h, and 48 h; Na_6h, Na_12h, Na_24h, Na_48h: Chinese spring seedlings treated with 200 mmol · L⁻¹ NaCl and grown for 6 h, 12 h, 24 h, and 48 h; PEG6000-0 h,PEG6000-2 h,PEG6000-12 h;0,2,12 h after PEG6000 treatments.

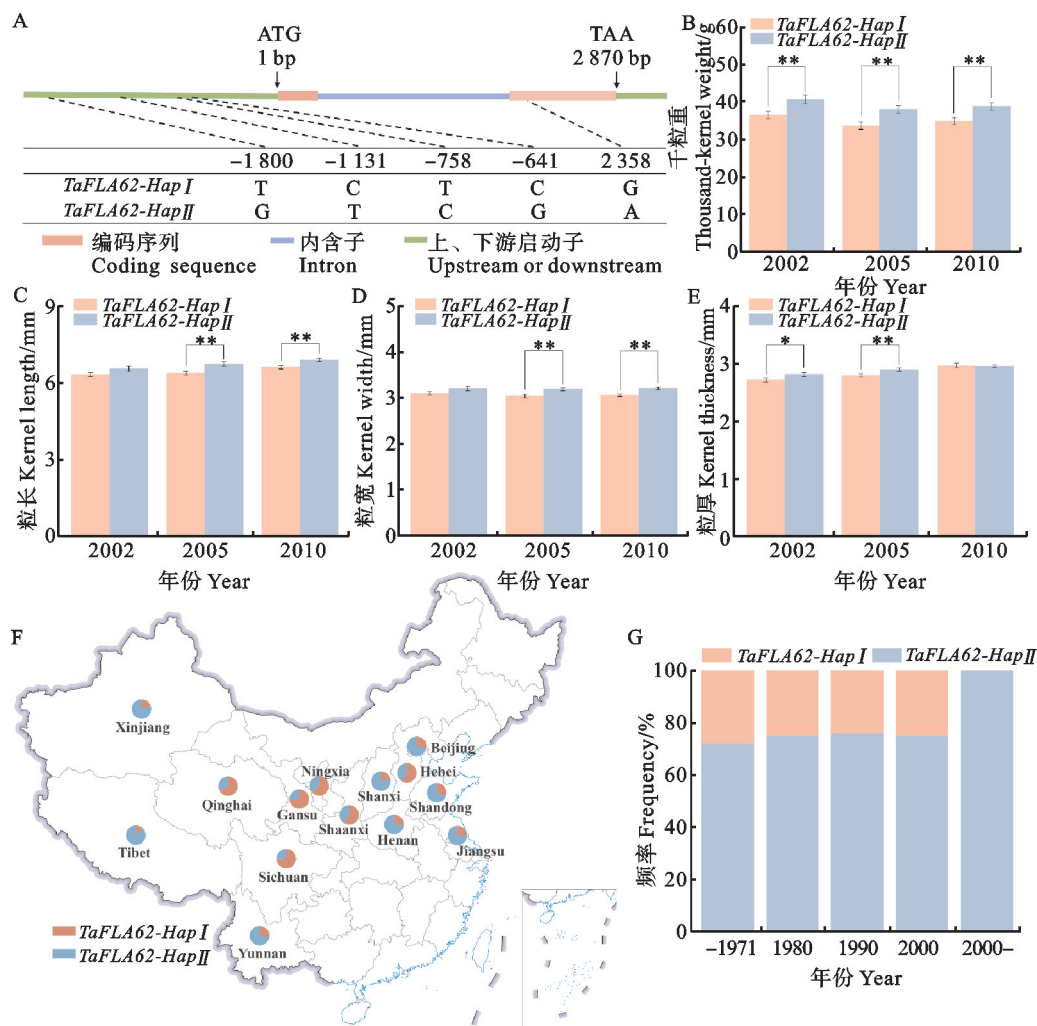
图 6 小麦TaFLAs在 200 mM NaCl 和 20% PEG 6000 胁迫下的模式分析(A)和表达谱分析(B~I)

Fig. 6 Expression pattern analysis of wheat TaFLAs under 200 mM NaCl and 20% PEG 6000 stresses(A) and expression profile analysis(B-I)

量未解析的区域及基因,本研究基于最新版本的小麦完整基因组数据,根据 HMMER 程序和蛋白同源比对方法,在全基因组水平共鉴定到 71 个小麦 FLA 蛋白,比 EST 数据库中的 34 个 FLA 蛋白数量明显增多,可为小麦 TaFLAs 基因功能研究提供参考。FLA 家族成员已在多个植物物种中被鉴定,在番茄中鉴定出了 18 个 SIFLA 成员,被分为 4 个亚族^[38];在白菜(*Brassica rapa* var. *glabra* Regel)中鉴定了 33 个 BrFLA,被分为 4 个亚族^[39]。本研究将小麦中鉴定到的 71 个 FLA 分为 8 个亚族,与番茄和白菜中 FLA 的分组不同,可能因为小麦庞大的基因组和大量的同源基因导致,推测 TaFLA 在进化过程中经历了基因扩张,例如在 Group V 和 Group VI 亚族中分别含有 10 个和 7 个 TaFLA 基因,而没有发现拟南芥

AtFLA 和水稻 *OsFLA* 基因,说明 Group V 和 Group VI 亚族 TaFLA 基因在小麦进化过程中经历了基因扩张。

在同一亚族中保守基序的保守性,暗示其功能上的相似性^[40]。研究表明,FLAs 是 AGP 蛋白,具有一个或两个 FAS1 结构域和两个保守结构域 H1 和 H2^[3]。本研究发现,motif 1 和 motif 2 存在于所有 TaFLAs 中,它们分别是组成 H1 和 H2 的主要基序,说明 TaFLA 具有保守性。已有研究发现,在拟南芥中 *AtFLA4* 的 FAS 结构域发生变化,导致根尖细胞壁异常并停止生长^[16];在棉花中 *GhFLA1* 的 FAS 结构域点突变导致细胞壁棉纤维伸长受阻,纤维长度缩短^[41]。说明 FLA 蛋白在进化过程中结构相对保守,推测 TaFLAs 在小麦中也具有相似的功能。



* : $P < 0.05$; * * : $P < 0.01$.

图 7 *TaFLA62* 单倍型结构示意图 (A)、与籽粒相关性状关联分析 (B~E)、地理分布 (F) 和时空分布 (G)
 Fig. 7 *TaFLA62* haplotype structure diagram (A), correlation analysis with grain-related traits (B-E), geographical distribution (F), and spatio-temporal distribution (G)

共线性分析对揭示基因家族进化关系和预测新基因功能有重要作用,而基因复制是物种进化的主要驱动力^[42]。已有研究表明,在番茄和杨树中 FLA 分别发生了 7 对和 12 对基因重复事件,均为片段复制^[15,38]。本研究发现,在小麦 *TaFLA* 基因中发生的 85 对基因复制均为片段复制,揭示了基因片段复制在小麦 FLA 基因家族扩张中的重要贡献。 K_a/K_s 比率主要用于通过比较基因序列的突变模式,推断自然选择的作用形式和强度^[43]。已有研究表明, K_a/K_s 小于 1 时,*TaUBC* 和 *TaGGCT* 基因发生了纯化选择^[27,29]。本研究中,*TaFLA* 基因对的 K_a/K_s 比值均小于 1,说明小麦 *TaFLAs* 在进化过程中主要发生了纯化选择。种间共线性分析发现,小麦 *TaFLAs* 基

因与其他物种的 FLA 基因具有共线性,进一步说明不同物种 FLAs 基因可能由共同的祖先进化而来,并且具有相似的功能。

研究表明,FLA 蛋白在细胞分化、发育调控、逆境响应和信号转导中发挥重要作用^[6,13,41]。Lu 等^[3] 在油茶 (*Camellia oleifera* Abe) 中发现 *CoFLAs* 在各个组织中均表达。本研究基于小麦多组学数据库中的转录组数据分析发现,*TaFLA* 不同亚族成员在不同组织中均表达,说明 FLA 家族基因可能在植物整个生长发育时期内均发挥作用。此外,在盐胁迫和干旱胁迫下的差异表达模式反映了不同 *TaFLA* 基因在应对不同非生物胁迫时的复杂调控机制可能与启动子区域的顺式作用元件的差异有关。进一步通过 qRT-PCR 验证发现,*TaFLA5* 在茎中表达量较高,*TaFLA14*、

TaFLA34 在幼穗和花后 5 d 籽粒中表达量较高, *TaFLA62* 在幼穗和茎中表达量较高, 推测 *TaFLAs* 基因可能参与多个器官的发育, 其中 *TaFLA14*、*TaFLA34*、*TaFLA62* 可能主要参与籽粒发育。

基因启动子区作为转录因子与顺式作用元件结合的主要区域, 启动子区序列变异往往会导致转录因子与顺式作用元件结合发生改变, 从而使基因的表达发生偏差, 最终使表型发生差异^[44]。例如, *TaCHLI-7B* 启动子区的位点发生突变显著影响了千粒重和单株产量^[45]。本研究中, *TaFLA62* 编码区的变异位点未引起氨基酸的改变, 在启动子区发现了一个与粒重性状显著关联的 SNP(-1 800 bp, T/G), 但它并非位于预测的顺式作用元件中, 提示其可能通过复杂调控机制发挥作用, 或与功能性变异位点存在紧密连锁, 需进一步研究验证。分子标记辅助选择是提高育种精度和效率的有效途径^[46]。千粒重是影响小麦产量的关键因素之一, 较高的千粒重对小麦稳产和高产起决定性作用^[47]。目前, 许多粒重相关基因已被开发分子标记并用于分子标记辅助选择育种^[48-49]。在本研究中, 单倍型 *TaFLA62-Hap II* 相较于单倍型 *TaFLA62-Hap I*, 具有更高的千粒重、粒长、粒宽和粒厚; 优异单倍型 *TaFLA62-Hap II* 在西北麦区所占比例偏低, 但在全国大多数麦区中优异单倍型 *TaFLA62-Hap II* 所占比例仍较高, 说明 *TaFLA62-Hap II* 受到了正向选择。因此, *TaFLA62-Hap II* 作为影响小麦籽粒发育的优异单倍型, 可用于小麦产量改良。

4 结论

在小麦中共鉴定到 71 个 *TaFLAs*, 可分为 8 个亚族, 同亚族成员具有相似的基因结构; *TaFLAs* 基因普遍在小麦不同组织中表达, 大多数响应干旱和盐胁迫; *TaFLA62* 基因具有两种单倍型 *TaFLA62-Hap I* 和 *TaFLA62-Hap II*, 单倍型 *TaFLA62-Hap II* 具有更高的千粒重、粒长、粒宽和粒厚, 可用于分子辅助育种。

参考文献:

- [1] SONG L, LIU J, CAO B L, *et al.* Reducing brassinosteroid signalling enhances grain yield in semi-dwarf wheat [J]. *Nature*, 2023, 617(7959): 118.
- [2] LESK C, ROWHANI P, RAMANKUTTY N. Influence of extreme weather disasters on global crop production [J]. *Nature*, 2016, 529(7584): 84.
- [3] LU M Q, ZHOU J Q, JIANG S S, *et al.* The fasciclin-like Arabinogalactan proteins of *Camellia* oil tree are involved in pollen tube growth [J]. *Plant Science*, 2023, 326: 111518.
- [4] MA F F, XU Y F, WANG R F, *et al.* Identification of major QTLs for yield-related traits with improved genetic map in wheat [J]. *Frontiers in Plant Science*, 2023, 14: 1138696.
- [5] MA Y X, ZENG W, BACIC A, *et al.* AGPs through time and space [J]. *Annual Plant Reviews online*, 2018, 3(1): 767.
- [6] LIU E, MACMILLAN C P, SHAFEE T, *et al.* Fasciclin-like Arabinogalactan-protein 16 (FLA16) is required for stem development in *Arabidopsis* [J]. *Frontiers in Plant Science*, 2020, 11: 615392.
- [7] SEIFERT G J, ROBERTS K. The biology of Arabinogalactan proteins [J]. *Annual Review of Plant Biology*, 2007, 58: 137.
- [8] DENG Y, WAN Y C, LIU W C, *et al.* OsFLA1 encodes a fasciclin-like Arabinogalactan protein and affects pollen exine development in rice [J]. *Theoretical and Applied Genetics*, 2022, 135(4): 1247.
- [9] ELLIS M, EGELUND J, SCHULTZ C J, *et al.* Arabinogalactan-proteins; Key regulators at the cell surface? [J]. *Plant Physiology*, 2010, 153(2): 403.
- [10] FAIK A, ABOUZOUHAIR J, SARHAN F. Putative fasciclin-like Arabinogalactan-proteins (FLA) in wheat (*Triticum aestivum*) and rice (*Oryza sativa*): Identification and bioinformatic analyses [J]. *Molecular Genetics and Genomics*, 2006, 276(5): 478.
- [11] JOHNSON K L, JONES B J, BACIC A, *et al.* The fasciclin-like Arabinogalactan proteins of *Arabidopsis*: A multigene family of putative cell adhesion molecules [J]. *Plant Physiology*, 2003, 133(4): 1911.
- [12] WU X Y, LAI Y C, LV L Q, *et al.* Fasciclin-like Arabinogalactan gene family in *Nicotiana benthamiana*: Genome-wide identification, classification and expression in response to pathogens [J]. *BMC Plant Biology*, 2020, 20(1): 305.
- [13] MA H L, ZHAO J. Genome-wide identification, classification, and expression analysis of the Arabinogalactan protein gene family in rice (*Oryza sativa* L.) [J]. *Journal of Experimental Botany*, 2010, 61(10): 2647.
- [14] HOZUMI A, BERA S, FUJIWARA D, *et al.* Arabinogalactan proteins accumulate in the cell walls of searching hyphae of the stem parasitic plants, *Cuscuta campestris* and *Cuscuta japonica* [J]. *Plant & Cell Physiology*, 2017, 58(11): 1868.
- [15] ZHANG Y Y, ZHOU F W, WANG H, *et al.* Genome-wide comparative analysis of the fasciclin-like Arabinogalactan proteins (FLAs) in *Salicaceae* and identification of secondary tissue development-related genes [J]. *International Journal of Molecular Sciences*, 2023, 24(2): 1481.
- [16] SHI H Z, KIM Y, GUO Y, *et al.* The *Arabidopsis* SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion [J]. *The Plant Cell*, 2003, 15(1): 19.
- [17] MA L, LI T, HAO C Y, *et al.* TaGS5-3A, a grain size gene selected during wheat improvement for larger kernel and yield [J]. *Plant Biotechnology Journal*, 2016, 14(5): 1269.
- [18] SAITOU N, NEI M. The neighbor-joining method: A new method for reconstructing phylogenetic trees [J]. *Molecular*

- Biology and Evolution*, 1987, 4(4):406.
- [19] MA S W, WANG M, WU J H, *et al.* WheatOmics: A platform combining multiple omics data to accelerate functional genomics studies in wheat [J]. *Molecular Plant*, 2021, 14(12):1965.
- [20] BORRILL P, RAMIREZ-GONZALEZ R, UAUY C. ExpVIP: A customizable RNA-seq data analysis and visualization platform [J]. *Plant Physiology*, 2016, 170(4):2172.
- [21] CHEN C J, CHEN H, ZHANG Y, *et al.* TBtools: An integrative toolkit developed for interactive analyses of big biological data [J]. *Molecular Plant*, 2020, 13(8):1194.
- [22] WILKINS M R, GASTEIGER E, BAIROCH A, *et al.* Protein identification and analysis tools in the ExPASy server [J]. *Methods in Molecular Biology*, 1999, 112:531.
- [23] HU B, JIN J P, GUO A Y, *et al.* GSDB 2.0: An upgraded gene feature visualization server [J]. *Bioinformatics*, 2015, 31(8):1296.
- [24] XIE J M, CHEN Y R, CAI G J, *et al.* Tree visualization by one table(tvBOT): A web application for visualizing, modifying and annotating phylogenetic trees [J]. *Nucleic Acids Research*, 2023, 51(W1):587.
- [25] SHEN H B, CHOU K C. Gpos-mPLoc: A top-down approach to improve the quality of predicting subcellular localization of Gram-positive bacterial proteins [J]. *Protein and Peptide Letters*, 2009, 16(12):1478.
- [26] ZHANG L, SUN W T, GAO W D, *et al.* Genome-wide identification and analysis of the GGCT gene family in wheat [J]. *BMC Genomics*, 2024, 25(1):32.
- [27] MENG Y, LV Q, LI L Q, *et al.* E3 ubiquitin ligase TaSDIR1-4A activates membrane-bound transcription factor TaWRKY29 to positively regulate drought resistance [J]. *Plant Biotechnology Journal*, 2024, 22(4):987.
- [28] FAN X D, DONG Y F, ZHANG Z P, *et al.* First report of *Vitis* cryptic virus from grapevines in China [J]. *Plant Disease*, 2022, 106(11):3006.
- [29] GAO W D, ZHANG L, ZHANG Y Y, *et al.* Genome-wide identification and expression analysis of the UBC gene family in wheat (*Triticum aestivum* L.) [J]. *BMC Plant Biology*, 2024, 24(1):341.
- [30] FIONA C, ALLEN JAMES E, JAMIE A, *et al.* Ensembl 2022 [J]. *Nucleic Acids Research*, 2021, 50(D1):D988.
- [31] SINGH V K, MANGALAM A K, DWIVEDI S, *et al.* Primer premier: Program for design of degenerate primers from a protein sequence [J]. *BioTechniques*, 1998, 24(2):318.
- [32] GUO L J, MA M, WU L N, *et al.* Modified expression of TaCYP78A5 enhances grain weight with yield potential by accumulating auxin in wheat (*Triticum aestivum* L.) [J]. *Plant Biotechnology Journal*, 2022, 20(1):168.
- [33] HE J J, LI C Z, HU N, *et al.* ECERIFERUM1-6A is required for the synthesis of cuticular wax alkanes and promotes drought tolerance in wheat [J]. *Plant Physiology*, 2022, 190(3):1640.
- [34] LIVAK K J, SCHMITTGEN T D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method [J]. *Methods*, 2001, 25(4):402.
- [35] YANG Z Z, WANG Z H, WANG W X, *et al.* ggComp enables dissection of germplasm resources and construction of a multiscale germplasm network in wheat [J]. *Plant Physiology*, 2022, 188(4):1950.
- [36] HAO C Y, JIAO C Z, HOU J, *et al.* Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China [J]. *Molecular Plant*, 2020, 13(12):1733.
- [37] CHENG H, LIU J, WEN J, *et al.* Frequent intra- and interspecies introgression shapes the landscape of genetic variation in bread wheat [J]. *Genome Biology*, 2019, 20(1):136.
- [38] HU J H, WANG J, MUHAMMAD T, *et al.* Functional analysis of fasciclin-like Arabinogalactan in carotenoid synthesis during tomato fruit ripening [J]. *Plant Physiology and Biochemistry*, 2024, 210:108589.
- [39] LI J, WU X M. Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like Arabinogalactan proteins in Chinese cabbage (*Brassica rapa* L.) [J]. *Molecular Biology Reports*, 2012, 39(12):10541.
- [40] XIONG R Q, PENG Z H, ZHOU H, *et al.* Genome-wide identification, structural characterization and gene expression analysis of the WRKY transcription factor family in pea (*Pisum sativum* L.) [J]. *BMC Plant Biology*, 2024, 24(1):113.
- [41] HUANG G Q, GONG S Y, XU W L, *et al.* A fasciclin-like arabinogalactan protein, GhFLA1, is involved in fiber initiation and elongation of cotton [J]. *Plant Physiology*, 2013, 161(3):1278.
- [42] WANG Y P, TANG H B, DEBARRY J D, *et al.* MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity [J]. *Nucleic Acids Research*, 2012, 40(7):e49.
- [43] HURST L D. The K_a/K_s ratio: Diagnosing the form of sequence evolution [J]. *Trends in Genetics*, 2002, 18(9):486.
- [44] HAMMOND-KOSACK M C U, KING R, KANYUKA K, *et al.* Exploring the diversity of promoter and 5' UTR sequences in ancestral, historic and modern wheat [J]. *Plant Biotechnology Journal*, 2021, 19(12):2469.
- [45] WANG Z X, XU H Y, WANG F X, *et al.* EMS-induced mis-sense mutation in *TaCHLI-7D* affects leaf color and yield-related traits in wheat [J]. *Theoretical and Applied Genetics*, 2024, 137(10):223.
- [46] HASAN N, CHOUDHARY S, NAAZ N, *et al.* Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes [J]. *Journal of Genetic Engineering and Biotechnology*, 2021, 19(1):128.
- [47] QIN L, HAO C Y, HOU J, *et al.* Homologous haplotypes, expression, genetic effects and geographic distribution of the wheat yield gene *TaGW2* [J]. *BMC Plant Biology*, 2014, 14:107.
- [48] MIAO Y P, JING F L, MA J F, *et al.* Major genomic regions for wheat grain weight as revealed by QTL linkage mapping and meta-analysis [J]. *Frontiers in Plant Science*, 2022, 13:802310.
- [49] ZHANG R J, AN K X, GAO Y J, *et al.* The transcription factor CAMTA2 interacts with the histone acetyltransferase GCN5 and regulates grain weight in wheat [J]. *The Plant Cell*, 2024, 36(12):4895.