乌拉尔甘草*HSF*基因家族全基因组鉴定及表达模式分析

杨明明[1], 陈昶泞[1,2], 张乃曼[3], 王雪媛[3], 董林林[1*]

(1. 中国中医科学院中药研究所, 北京 100700; 2. 天津中医药大学中药学院, 天津 301617; 3. 盛实百草药业有限公司, 天津 300301)

摘要: 热激转录因子 (heat shock transcription factor, HSF) 家族在植物的生长发育以及应对生物和非生物胁迫中发挥着关键作用。乌拉尔甘草 (*Glycyrrhiza uralensis* Fisch.) 作为一种传统中草药, 富含多种生物活性成分, 包括黄酮类和萜类化合物。然而, 目前尚缺乏对乌拉尔甘草*HSF*基因家族的系统鉴定和表征。为明确该基因家族的生物学功能, 本研究基于乌拉尔甘草基因组, 系统鉴定了*HSF*家族成员, 并分析了其理化性质、染色体定位、系统发育关系、保守基序以及共线性。同时, 结合转录组与qRT-PCR分析, 解析了*GuHSF*基因的表达模式及其在干旱胁迫下的与黄酮合成相关的共表达基因。研究结果显示, 从乌拉尔甘草基因组中鉴定出26个*GuHSF*基因, 这些基因可分为3个亚家族, 并且C亚家族的HSF蛋白具有高度相似的保守基序。共线性分析识别出*GuHSF*家族内存在12对片段重复, 并且*GuHSF*基因与大豆*HSF*基因之间的共线性关系在所比较的物种中最为显著。qRT-PCR分析结果表明, *GuHSF2*、*GuHSF4*、*GuHSF14*、*GuHSF18*、*GuHSF21*、*GuHSF26*基因受到干旱胁迫的显著诱导。此外, 在干旱条件下, *GuHSF1*、*GuHSF4*、*GuHSF9*、*GuHSF11*、*GuHSF21*基因与黄酮合成相关基因显示出强相关性。本研究系统鉴定并分析了*GuHSF*家族基因, 为后续探究该家族核心成员的生物学功能提供了重要线索。

关键词: 乌拉尔甘草; *HSF*基因家族; 表达模式; 干旱; 黄酮

中图分类号: R931 **文献标识码**: A **文章编号**: 0513-4870(2026)04-1418-13

Genome-wide identification and expression pattern analysis of *HSF* gene family in *Glycyrrhiza uralensis*

YANG Ming-ming[1], CHEN Chang-ning[1,2], ZHANG Nai-jiao[3], WANG Xue-yuan[3], DONG Lin-lin[1*]

(*1. Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China; 2. College of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China; 3. China Medico Corporation, Tianjin 300301, China*)

Abstract: The heat shock transcription factor (HSF) family plays a crucial role in plant growth, development, and responses to biotic and abiotic stresses. As a traditional Chinese herbal medicine, *Glycyrrhiza uralensis* Fisch. is rich in various bioactive compounds, including flavonoids and terpenes. However, a systematic identification and characterization of the *HSF* gene family in *G. uralensis* is currently lacking. To clarify the biological functions of this gene family, we systematically identified *HSF* family members based on the genomic data of *G. uralensis*, and analyzed their physicochemical properties, chromosomal localization, phylogenetic relationships, conserved motifs, and collinearity. Meanwhile, combined with transcriptome sequencing and qRT-PCR analyses, we elucidated the expression patterns of *GuHSF* genes and their co-expressed genes associated with flavonoid biosynthesis under drought stress. The results showed that a total of 26 *GuHSF* genes were identified from the genome of *G. uralensis*,

收稿日期: 2025-11-08; 修回日期: 2026-01-10.

基金项目: 中央级公益性科研院所基本科研业务费专项资金资助 (ZZ13-YQ-049); 国家自然科学基金资助项目 (82504963); 中国博士后科学基金资助项目 (GZC20242025).

*通讯作者 E-mail: lldong@icmm.ac.cn

DOI: 10.16438/j.0513-4870.2025-1321

which can be classified into three subfamilies, and the HSF proteins within the C subfamily exhibited high similarity in conserved motifs. Collinearity analysis revealed 12 pairs of segmental duplications within the *GuHSF* family, and the syntenic relationships between *GuHSF* genes and soybean *HSF* genes were the strongest among the compared species. The results of qRT-PCR analysis indicated that *GuHSF2*, *GuHSF4*, *GuHSF14*, *GuHSF18*, *GuHSF21*, and *GuHSF26* genes were significantly induced by drought stress. In addition, under drought conditions, strong correlations were observed between *GuHSF1*, *GuHSF4*, *GuHSF9*, *GuHSF11*, and *GuHSF21* genes and flavonoid biosynthesis-related genes. This study systematically identified and analyzed the *GuHSF* family genes in *G. uralensis*, which provides important clues for subsequent investigations into the biological functions of core members of this family.

Key words: *Glycyrrhiza uralensis*; *HSF* gene family; expression pattern; drought; flavonoid

甘草是我国传统中药材, 属于豆科甘草属植物, 其药用及食用部位为干燥根和根茎。作为经典药用植物, 甘草根部因富含黄酮类化合物而具备显著的药用价值。已有研究证实, 甘草在抗菌、抗炎、镇痛及增强机体免疫力等方面具有明确的功效[1,2]。甘草广泛分布于我国的干旱及半干旱地区, 如内蒙古、新疆和甘肃等地, 具有较强的抗旱及防风固沙能力[3]。研究表明, 干旱胁迫是诱导甘草合成黄酮类物质的重要环境因子, 甘草根中黄酮类化合物的积累不仅提高了甘草对复杂环境的适应能力, 同时也提升了其药用成分的含量[4,5]。

植物作为典型的固着生物, 生长发育过程中常面临多种生物与非生物胁迫, 而干旱是影响植物生存与产量的主要非生物胁迫之一。为应对复杂多变的外界环境, 植物已进化出多层次的防御响应机制[6,7]。在遭受胁迫时, 植物迅速启动信号传导过程, 通过蛋白质级联磷酸化等方式实现胁迫信号的高效传递, 进而激活包括转录因子在内的下游功能蛋白[8]。被激活的转录因子可通过激活或抑制抗性及防御相关基因的表达, 直接增强植物对环境胁迫的适应能力[5,9]。其中, 热激转录因子 (HSFs) 是植物胁迫信号转导通路中的关键组分, 不仅参与热胁迫响应基因的激活, 还在其他非生物胁迫的应答过程中发挥重要的调控作用[10]。

植物基因组中含有大量热激转录因子 (*HSF*) 基因, 而酵母和动物基因组中*HSF*基因数量相对较少[11]。截至目前, 拟南芥 (*Arabidopsis thaliana*)、水稻 (*Oryza sativa*)、大豆 (*Glycine max*)、苜蓿 (*Medicago sativa*) 及黄芪 (*Astragalus membranaceus*) 等植物的*HSF*家族基因已得到系统的鉴定与分析[12-16]。研究显示, 所有HSF蛋白均含有高度保守的DNA结合结构域 (DBD) 以及两个疏水氨基酸重复片段 (HR-A区和HR-B区)。根据HR-A/B区的结构差异, *HSF*家族可进一步划分为A、B、C三个亚家族[11]。*HSF*家族成员在植物抵御逆境胁迫中发挥着关键的调控作用。例如, 在拟南芥中过量表达水稻*HSF2e*基因可以显著提升转基因植株的抗盐性[17]; 番茄中的HSFA1a通过上调自噬相关基因 (*ATG*) 的表达, 显著增强了抗旱能力[18]; 转基因玉米中*HSF21*的过表达株系在萌发期和苗期均表现出强大的耐冷性[19]; 而在苹果中过量表达*MdHSFA8a*不仅提高了植株的耐旱性, 还显著促进了类黄酮物质的积累[20]。这些研究表明, *HSF*家族基因在植物抗逆响应及次生代谢调控中具有保守且多样的功能, 为甘草*HSF*基因的功能研究提供了参考。

乌拉尔甘草作为甘草的主要栽培品种, 因其在治疗炎症、感冒、哮喘及肝肺疾病方面具有明确疗效而被广泛种植[1,2]。因此, 深入解析乌拉尔甘草*HSF*家族基因对于探讨其生境适应性机制及黄酮类物质的合成通路具有重要的理论和实践意义。本研究基于乌拉尔甘草基因组数据, 系统鉴定*HSF*家族基因, 并从系统发育关系、染色体定位、保守基序组成、基因结构特征、启动子顺式作用元件及基因表达模式等方面进行全面分析, 旨在为后续深入研究乌拉尔甘草HSF家族蛋白的功能提供基础参考。

材料与方法

植物材料 本研究使用的乌拉尔甘草转录组数据均来源于NCBI数据库, 项目编号为PRJNA1321174与PRJNA878610, 用于转录组测序的材料采用两种干旱处理方式, 分别为PEG6000诱导模拟干旱和自然干旱。其中模拟干旱处理为: 采用20% (*w*/*v*) PEG6000对生长50天的甘草幼苗进行胁迫处理, 分别在处理后0、2、6、12、24 h采集幼苗叶片, 以清水处理的幼苗作为对照组; 自然干旱处理为: 对生长4个月的乌拉尔甘草进行7天干旱胁迫 (DS), 胁迫后分别复水7天 (R1) 和14天 (R2), 并以正常浇水的植株作为对照 (CK)[4]。

HSF家族成员的全基因组鉴定及特征分析 乌拉尔甘草基因组来源于NCBI数据库 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030159175.1/)。

从 PFAM 数据库 (https://www.ebi.ac.uk/Tools/hmmer/search/phmmer) 获取 HSF 结构域 (PF00447) 对应的隐马尔可夫模型 (HMM) 文件。从 PlantTFDB 在线网站 (https://planttfdb.gao-lab.org/) 获取拟南芥 HSF 蛋白序列。通过 TBtools 软件的 HMM search 功能及 BLASTp 比对, 在乌拉尔甘草蛋白组中筛选潜在 HSF 蛋白。进一步利用 CD-Search 数据库 (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb /bwrpsb.cgi) 验证候选蛋白的 HSF 保守结构域, 最终确定乌拉尔甘草 HSF 家族成员。使用 ExPASy Prot-Param 网站 (https://web.expasy.org/protparam/) 和 Cell-PLoc 2.0 在线网站 (http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/) 预测 HSF 蛋白的理化性质及亚细胞定位。

染色体定位、系统发育分析和共线性 基于乌拉尔甘草基因组文件及注释文件, 使用 TBtools 软件完成 *HSF* 基因在染色体上的可视化分析。借助 TBtools 软件的 MuscleWrapper 模型, 对拟南芥、大豆与乌拉尔甘草的 HSF 氨基酸序列进行多重比对。随后通过 IQ-Tree Wrapper 程序构建系统发育树, 设置自展值为 1 000。利用 MCScanX 程序, 对拟南芥、大豆、鹰嘴豆、苜蓿及木豆基因组中的 *HSF* 家族成员, 开展基因组内与基因组间的同源性分析; 最终通过 TBtools 软件可视化呈现 *HSF* 成员间的同源性关系

基因结构、保守基序和顺式作用元件 采用 TBtools 软件预测乌拉尔甘草的 HSF 蛋白家族的保守基序, 设置保守基序搜索值为 10。结合保守基序的预测结果与乌拉尔甘草基因组注释文件, 通过 TBtools 软件绘制 HSF 蛋白的保守基序图以及 *HSF* 基因的结构图。通过 PlantCare 网站 (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) 预测 *HSF* 基因家族启动子序列 (ATG 上游 2 000 bp 区域) 中的顺式作用元件, 借助 TBtools 软件完成启动子顺式作用元件的可视化分析。

***K*-Means 聚类和相关性分析** 从 20% (*w*/*v*) PEG6000 处理的乌拉尔甘草转录组数据中, 依据基因注释信息筛选出 *GuHSF* 家族成员及黄酮合成相关基因。利用迈维生物云平台对筛选获得的目标基因进行 *K*-Means 聚类分析 (聚类数目设为 10, 其余参数默认)。通过联川生物云平台分析目标基因间的相关性, 最终利用 Cytoscape 软件 (v3.10.0) 可视化呈现基因相关性网络。

qRT-PCR 分析 采集经 20% (*w*/*v*) PEG6000 溶液 (Biosharp, BS902, 中国) 处理的乌拉尔甘草叶片 (取样时间点为处理后 0、2、6、12、24 h), 采用天根 RNA 提取试剂盒 (TIANGEN, DP441, 中国) 从甘草叶组织中提取总 RNA, 经天根反转录试剂盒 (TIANGEN, KR116, 中国) 反转录合成 cDNA。qRT-PCR 分析采用诺唯赞 ChamQ Universal SYBR qPCR Master Mix 试剂盒 (Vazyme, Q711, 中国), 在 ABI 7500 Fast Real-Time PCR 系统 (Thermo Fisher Scientific, 美国) 上完成。干旱处理后甘草叶中 *GuHSFs* 基因的表达量采用 $2^{-\Delta\Delta Ct}$ 法计算, 以甘草 *Guβ-tubulin* 作为内参基因。每个实验条件均设置 3 个生物学重复与 3 个技术重复, 实验数据取平均值后进行后续分析, 引物序列见表 1。采用单因素方差分析评估统计学差异, 显著性水平设定为 $P < 0.05$。

结果

1 乌拉尔甘草 HSF 家族成员鉴定

为鉴定乌拉尔甘草中 HSF 家族成员, 本研究结合 HMM 比对、BLASTp 比对及蛋白保守结构域预测方法, 从其基因组中鉴定出 30 个潜在的 HSF 蛋白。进一步借助 NCBI 数据库的 CD-search 工具对上述候选 HSF 蛋白序列进行保守结构域鉴定, 最终从乌拉尔甘草基因组中获得 26 个具有完整 HSF 保守结构域的蛋白成员 (表 2)。使用 TBtools 软件对这 26 个 HSF 家族成员进行染色体定位分析, 并根据基因在染色体上的排列顺序为其命名。结果如图 1A 所示, *HSF* 家族成员在甘草的 8 条染色体上呈不均匀分布, 其中第 3 号染色体上的 *HSF* 基因数量最多, 含有 5 个。为了揭示甘草 HSF 家族成员的进化关系, 本研究以拟南芥和大豆已报道的 HSF 蛋白序列为参考, 将其与甘草 26 个 HSF 蛋白序列构建系统发育树。参考拟南芥与大豆 HSF 家族的分类方法, 结合 HSF 各亚家族特有的结构域特征, 对甘草 26 个 HSF 蛋白序列进行系统分类。结果如图 1B 所示, 26 个 *GuHSF* 基因可划分为 A、B、C 三个亚家

Table 1 Sequences of primers for qRT-PCR

Gene name	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)
GuHSF2	GGTTGAGGGGTGCCTATGGTT	ATTTCCCCCCAGCAGTTTTTC
GuHSF4	GGGAGGAGCAGGTGGTTTCTT	CTTCGTCAACTGAATGTTCTCCCT
GuHSF14	GAGGGCACAGGCACCTTTTG	TCAACACAGTGACCTTGCCCTT
GuHSF18	CAGAGGCATTTGTTGAAGACCA	CTTCCATACCATACTCCCCCAA
GuHSF21	CGAGCGAGAACGAGAAACTGAA	TTCACACGGTCCCTCAGAAAAG
GuHSF26	GTTCGCCAACTCAACACCTATG	TTCAATAAGTGCTTCTGCCCTC
Guβ-tubulin	GCAATACAGGGCTCTCACCGT	TCAAATAACGACCGTGGCGA

族, 其中A亚家族成员占比最高 (53.85%), C亚家族成员次之 (3.85%), 而B亚家族成员占比最低 (42.31%)。

2 乌拉尔甘草HSF家族成员理化性质分析

为探究甘草HSF家族成员的理化性质, 本研究通过TBtools工具分析其蛋白序列理化参数。结果显示 (表2), 该家族成员的氨基酸长度介于190～503个氨基酸 (aa) 之间, 其相对分子质量为21 864.69～55 602.88 Da, 脂溶性指数为54.44～79.84, 平均亲水性为−0.895～−0.501。根据平均亲水性数值判断, 该家族所有成员均为水溶性蛋白。在其他理化参数方面, HSF家族蛋白的等电点为4.66～8.94, 其中包括7个碱性蛋白和19个酸性蛋白。蛋白的不稳定系数介于34.08～63.19之间, 仅GuHSF1和GuHSF23的不稳定系数小于40, 表明它们为稳定蛋白, 其余成员均为不稳定蛋白 (表2)。此外, 对HSF家族成员的亚细胞定位分析结果表明, 该家族所有蛋白均定位于细胞核。

3 乌拉尔甘草HSF家族成员保守基序和基因结构分析

甘草HSF家族成员的保守基序分析结果显示 (图2), 除GuHSF6外, 所有HSF蛋白均包含motif 1和motif 3。进一步分析发现, 不同亚家族HSF蛋白的保守基序存在差异, 其中C亚家族成员的基序种类和数量最为保守, 均包含motif 1、motif 2、motif 3、motif 5。而A亚家族HSF成员的基序种类和数量差异较大, 仅系统发育树中聚类在同一分支的成员具有相似的保守基序 (图2)。对*HSF*家族基因的结构分析结果显示 (图2), 不同成员的基因结构存在显著差异, 外显子数量介于2到4个之间, 内含子数量为1到3个。其中, *GuHSF6*、*GuHSF15*、*GuHSF16*、*GuHSF18*、*GuHSF19*、*GuHSF22*和*GuHSF23*共7个成员不含UTR区域。

4 乌拉尔甘草*HSF*家族基因启动子顺式作用元件分析

利用TBtools软件从乌拉尔甘草基因组中提取*HSF*家族基因的启动子序列, 并通过PlantCare数据库分析其顺式作用元件。结果显示 (图3), *HSF*家族基因的启动子区域共检测到5大类顺式作用元件, 分别为激素响应元件、生物与非生物胁迫响应元件、昼夜节律控制元件、光响应元件以及类黄酮生物合成基因调控元件。启动子元件的预测结果表明, 不同*GuHSF*基因的启动子顺式作用元件组成存在显著差异, 其中*GuHSF14*的元件数量最多 (47个), 而*GuHSF15*的元件数量最少 (9个)。值得注意的是, 光响应元件是所有*GuHSF*基因共有的元件类型, 且占比最高 (60.29%, 图3)。此外, 脱落酸响应元件 (占比14.38%)、茉莉酸甲酯响应元件 (占比8.17%)、赤霉素响应元件 (占比5.39%)、水杨酸响应元件 (占比2.94%)、生长素响应元件 (占比2.45%)、干旱响应元件 (占比2.12%)、低温响应元件 (占比1.80%)、防御与应激响应元件 (占比

Table 2 Information and characteristics of GuHSFs. pI: Isoelectric point; GRAVY: Grand average of hydropathicity; SL: Subcellular localization

Gene name		Amino acids length	Molecular mass/Da	pI	Instability index	Aliphatic index	GRAVY	SL
Gur01G026360.1	*GuHSF1*	274	30 026.26	6.25	34.08	60.18	−0.826	Nucleus
Gur01G037730.1	*GuHSF2*	438	48 931.96	5.4	51.72	79.84	−0.585	Nucleus
Gur02G004950.1	*GuHSF3*	487	54 772.01	4.85	62.47	66	−0.641	Nucleus
Gur02G038620.1	*GuHSF4*	319	35 470.79	7.11	63.19	69.12	−0.562	Nucleus
Gur02G039400.1	*GuHSF5*	432	49 907.2	5.59	50.05	73.98	−0.695	Nucleus
Gur02G044350.1	*GuHSF6*	190	21 864.69	6.05	60.41	78.47	−0.726	Nucleus
Gur03G016760.1	*GuHSF7*	491	54 876.52	5.9	56.35	67.27	−0.603	Nucleus
Gur03G022420.1	*GuHSF8*	401	45 712.24	5.3	51.52	70.77	−0.697	Nucleus
Gur03G023500.1	*GuHSF9*	481	54 219.87	5.47	60.24	64.12	−0.77	Nucleus
Gur03G029060.1	*GuHSF10*	214	24 330.82	8.24	52.51	76.12	−0.525	Nucleus
Gur03G046130.1	*GuHSF11*	364	41 473.08	4.98	52.25	79.45	−0.573	Nucleus
Gur04G002410.1	*GuHSF12*	346	39 609.36	8.26	45.05	69.8	−0.693	Nucleus
Gur04G030570.1	*GuHSF13*	256	29 374.73	7.66	58.99	62.81	−0.895	Nucleus
Gur04G032650.1	*GuHSF14*	397	45 786.64	6.32	61.73	70.23	−0.802	Nucleus
Gur05G009740.1	*GuHSF15*	257	30 014.91	6.41	55.87	74.71	−0.654	Nucleus
Gur05G016450.1	*GuHSF16*	475	53 057.35	4.66	61.86	73.22	−0.648	Nucleus
Gur05G035540.1	*GuHSF17*	383	43 206.35	8.4	53.25	66.66	−0.702	Nucleus
Gur05G036720.1	*GuHSF18*	380	42 719.89	5.03	59.46	74.87	−0.569	Nucleus
Gur06G004150.1	*GuHSF19*	285	33 072.41	6.86	51.82	71.44	−0.663	Nucleus
Gur06G026620.1	*GuHSF20*	503	55 602.88	4.83	61.27	67.87	−0.682	Nucleus
Gur06G031620.1	*GuHSF21*	297	32 933.61	8.13	43.69	54.44	−0.863	Nucleus
Gur06G034790.1	*GuHSF22*	356	38 600.89	4.95	55.17	72.42	−0.501	Nucleus
Gur07G003410.1	*GuHSF23*	345	39 465.12	8.94	35.7	78.46	−0.598	Nucleus
Gur07G035110.1	*GuHSF24*	397	44 932	5.11	61.37	68.49	−0.735	Nucleus
Gur08G014500.1	*GuHSF25*	363	39 447.72	5.11	61.77	63.69	−0.661	Nucleus
Gur08G020270.1	*GuHSF26*	364	40 503.39	5.51	53.52	69.89	−0.542	Nucleus

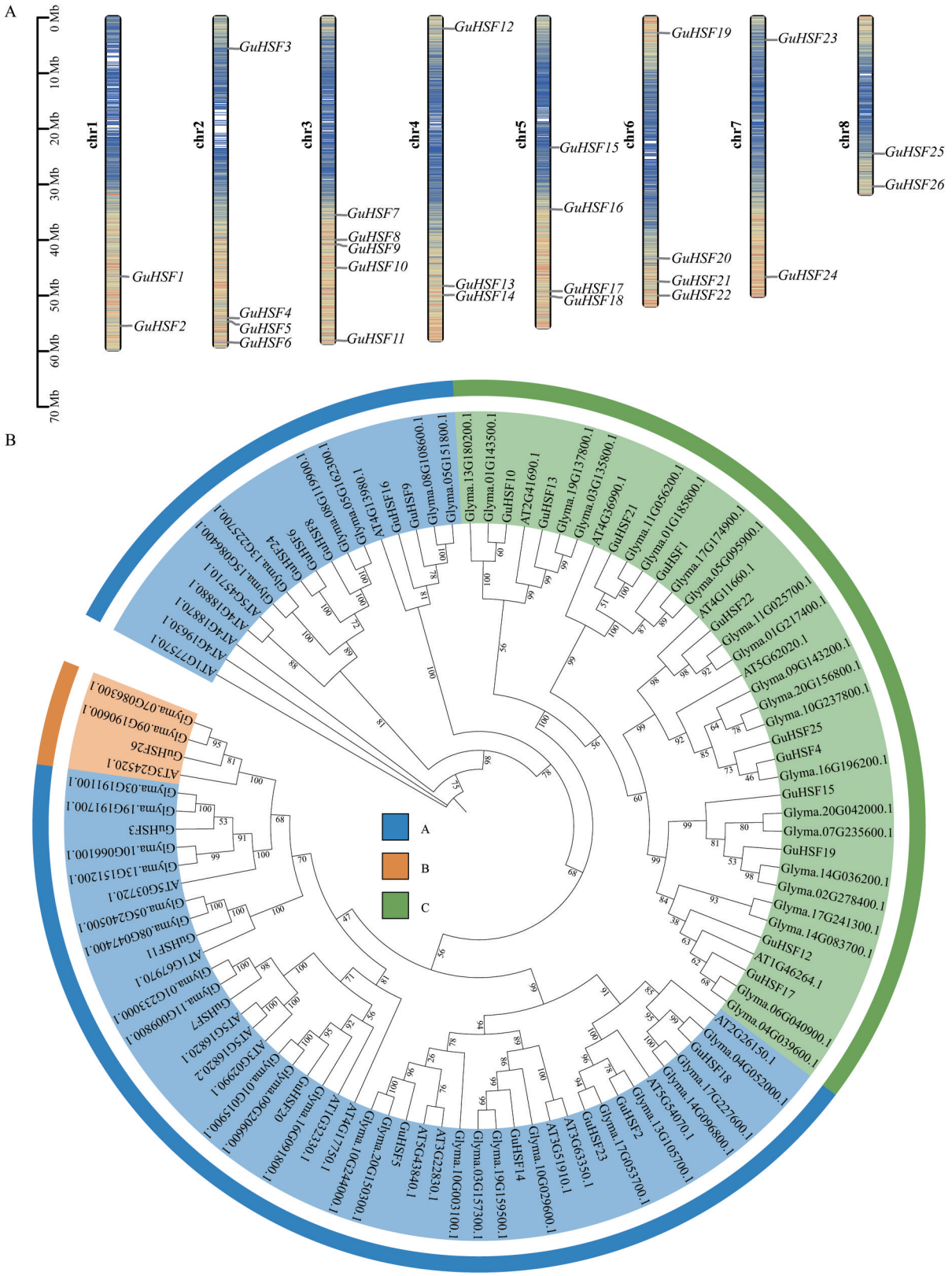


Figure 1 Chromosome localization (A) and phylogenetic analysis (B) of *HSF* genes in *Glycyrrhiza uralensis*

1.31%)、昼夜节律控制元件 (占比 0.65%) 以及类黄酮生物合成调控元件 (占比 0.49%) 也广泛分布于 *GuHSF* 基因启动子序列中。

5 *HSF* 家族成员的共线性分析

对 26 个 *GuHSF* 基因的片段重复分析结果显示，本研究共鉴定出 20 个同源基因座及 12 对片段重复事

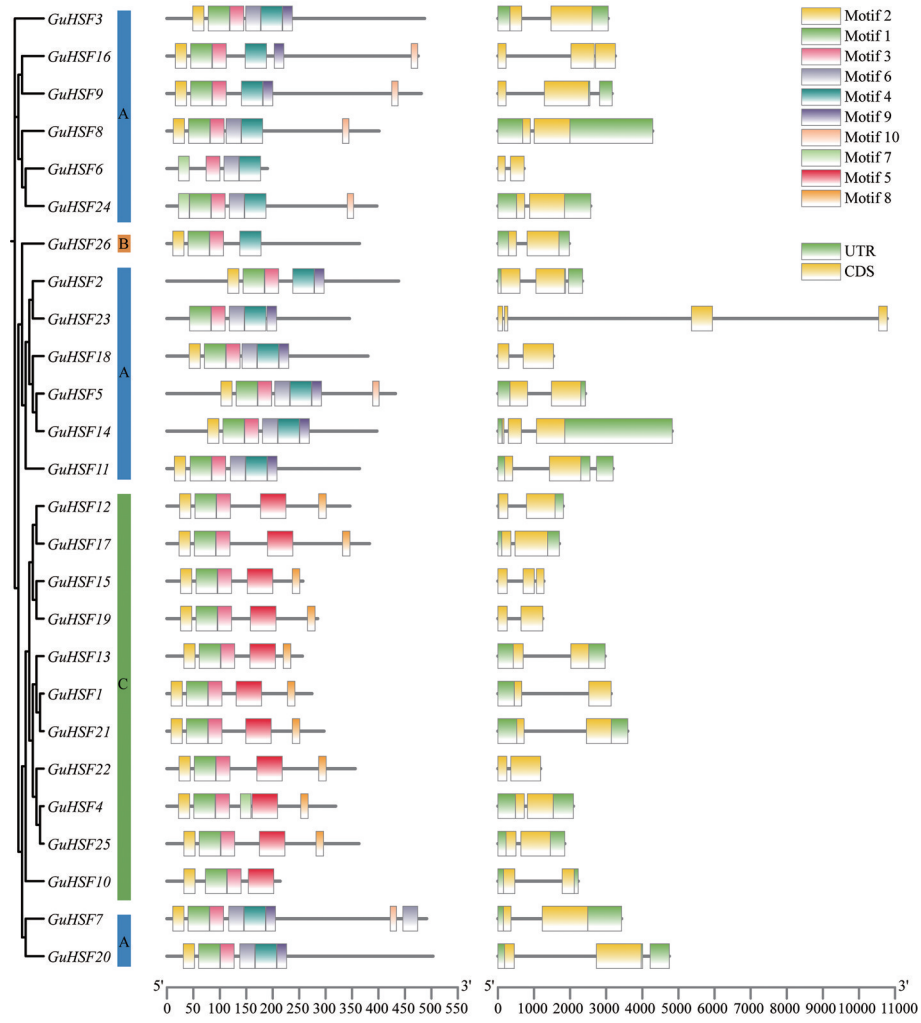


Figure 2 Conserved motifs and gene structures of *GuHSFs*

件 (图 4)。从亚家族分布特征来看, A、B、C 亚家族的片段重复数量存在显著差异: A 亚家族包含 7 对片段重复, C 亚家族为 5 对, B 亚家族未检测到片段重复事件 (图 1B)。这些同源基因对分布于 chr1～chr8 染色体 (如 *GuHSF2* 与 *GuHSF18* 分别定位于 chr1 与 chr5, *GuHSF4* 与 *GuHSF25* 分别定位于 chr2 与 chr8), 证实片段重复是驱动乌拉尔甘草 *HSF* 基因家族扩张的核心动力 (表 3)。对重复基因对的进化动力学分析表明, 除 *GuHSF1*/*GuHSF21*、*GuHSF4*/*GuHSF22*、*GuHSF7*/*GuHSF20*、*GuHSF18*/*GuHSF23* 因 Ks 值缺失未纳入选择压力分析外, 其余 8 组重复基因对的 Ka/Ks 值介于 0.119～0.389 之间, 均小于 1, 表明这些基因在进化过程中受到纯化选择作用, 功能趋于保守 (表 3)。基于重复基因对的 Ks 值估算进化时间, 结果显示片段重复事件发生于 20.448～81.933 Mya, 其中 *GuHSF5*/*GuHSF14* (21.073 Mya) 与 *GuHSF15*/*GuHSF19* (20.448 Mya) 分化时间相对较近, 而 *GuHSF4*/*GuHSF25* 分化时间最早 (81.933 Mya), 提示不同演化时期的片段重复事件共同推动了 *HSF* 基因家族的扩增与功能分化 (表 3)。这些结果表明片段重复不仅为乌拉尔甘草 *HSF* 基因家族的数量扩张提供了物质基础, 且纯化选择作用保障了基因功能的稳定性。

为探究 *HSF* 基因在物种间的进化关系, 本研究以乌拉尔甘草 *GuHSF* 基因为研究对象, 与 4 种豆科植物 (大豆、鹰嘴豆、苜蓿和木豆) 及模式植物拟南芥 *HSF* 同源基因进行了共线性分析 (图 5)。结果表明, *GuHSF* 基因与大豆 *HSF* 基因间共检测到 87 对共线性基因对, 数量显著高于其他物种, 提示二者亲缘关系最近; 其次为鹰嘴豆和苜蓿, 分别检测到 38 对和 33 对共线性基因对。与拟南芥和木豆的 *HSF* 基因的共线性基因对数量相对较少, 分别为 25 对和 22 对。这些结果表明, 在与其他物种的比较中, 甘草 *GuHSF* 基因与大豆 *HSF* 基因之间表现出最强的共线性关系。

6 乌拉尔甘草 *HSF* 家族基因在干旱条件下的表达模式分析

为探究 *GuHSF* 家族基因在干旱胁迫下的表达特

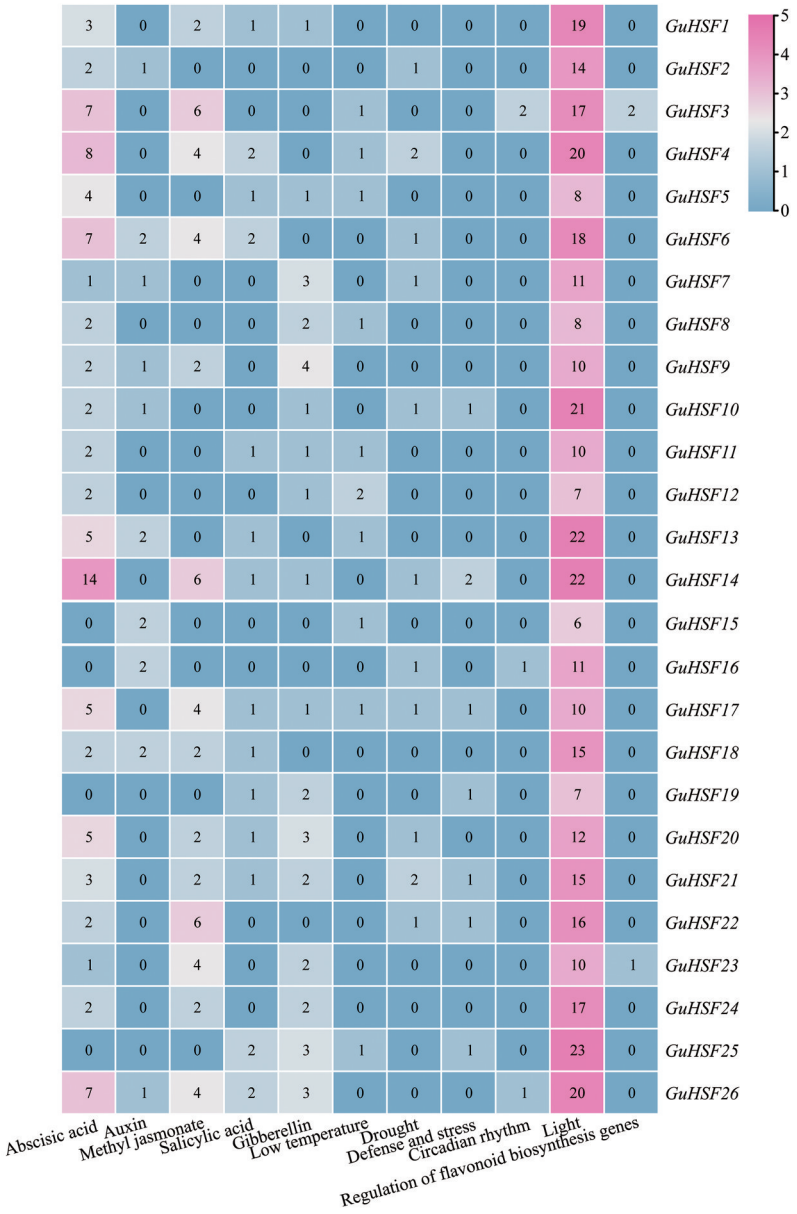


Figure 3 *Cis*-Acting elements in promoters of *GuHSFs*

征, 本研究系统分析了20% PEG6000模拟干旱与自然干旱胁迫条件下乌拉尔甘草的转录组数据。如图6A所示, 20% PEG6000胁迫处理下, 10个*GuHSF*基因 (*GuHSF1*、*GuHSF3*、*GuHSF4*、*GuHSF5*、*GuHSF10*、*GuHSF13*、*GuHSF14*、*GuHSF18*、*GuHSF21*、*GuHSF26*) 均呈现显著的干旱诱导表达模式, 且不同基因的响应时序存在差异。*GuHSF1*、*GuHSF14*、*GuHSF18*、*GuHSF21*为早期响应基因, 其表达水平在胁迫后2 h即显著上调; *GuHSF3*、*GuHSF5*、*GuHSF10*、*GuHSF13*为中期持续响应基因, 在胁迫后2～6 h表达量呈逐步升高的持续诱导趋势。*GuHSF26*表达水平于胁迫后6 h显著上调并达到峰值, 而*GuHSF4*表现出独特的持续响应特征, 在0～24 h胁迫期间内表达量均呈显著持续升高趋势。自然干旱胁迫下 (4月龄乌拉尔甘草植株) 的转录组分析结果显示 (图6B), *GuHSF2*、*GuHSF4*、*GuHSF18*、*GuHSF26*基因受干旱胁迫显著诱导, 其表达量均较对照上调2倍以上。复水处理后, 这些基因的表达水平可恢复至胁迫前的对照水平, 提示其表达调控具有干旱响应特异性及可逆性。

为进一步探究乌拉尔甘草*GuHSF*基因的表达模式, 本研究结合上述*GuHSF*基因的表达趋势及受20% PEG 6000诱导的转录组数据, 从中筛选FPKM值较高的6个*GuHSF*基因 (*GuHSF2*、*GuHSF4*、*GuHSF14*、*GuHSF18*、*GuHSF21*、*GuHSF26*) 进行qRT-PCR分析。结果如图6C显示, 上述基因均受20% PEG6000显著诱导, 且在处理后24 h表达量达到峰值。其中*GuHSF26*

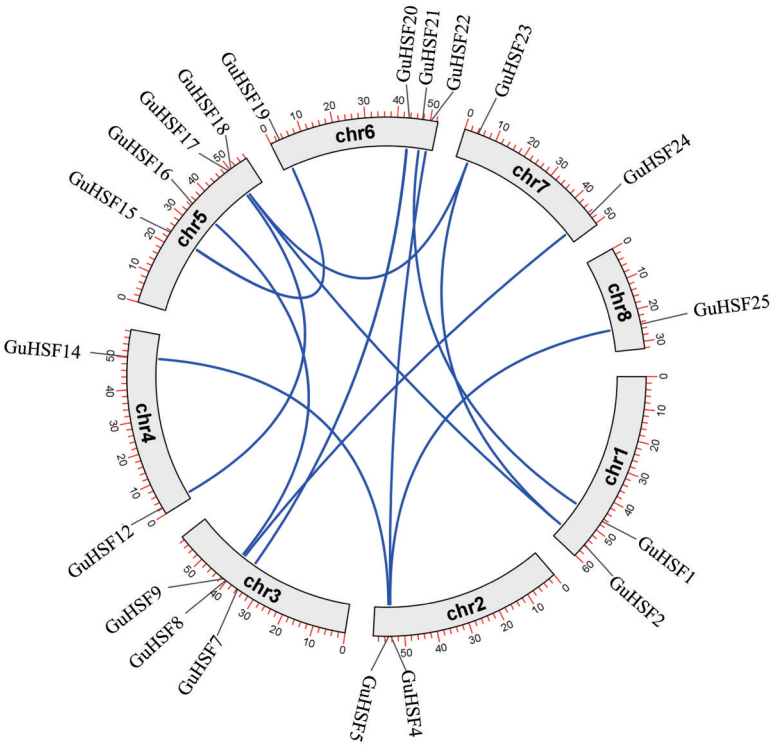


Figure 4 Segmental duplication of *GuHSF*s in *G. uralensis*

Table 3 Ka, Ks, and Ka/Ks values for duplicated gene pairs in *G. uralensis*. Ka: Nonsynonymous substitution rate; Ks: Synonymous substitution rate

Gene 1	Gene 2	Ka	Ks	Ka/Ks	Date/Mya	Chromosome
GuHSF2	*GuHSF18*	0.526	1.526	0.345	50.865	chr1/chr5
GuHSF2	*GuHSF23*	0.276	0.839	0.329	27.967	chr1/chr7
GuHSF4	*GuHSF25*	0.172	0.632	0.272	81.933	chr2/chr8
GuHSF5	*GuHSF14*	0.291	2.458	0.119	21.073	chr2/chr4
GuHSF8	*GuHSF24*	0.128	0.795	0.161	28.956	chr3/chr7
GuHSF9	*GuHSF16*	0.338	0.869	0.389	26.497	chr3/chr5
GuHSF12	*GuHSF17*	0.153	0.944	0.162	31.458	chr4/chr5
GuHSF15	*GuHSF19*	0.152	0.613	0.248	20.448	chr5/chr6

的表达水平在处理后6 h显著上调, 随后显著下降, 并于24 h再次升高, 呈现独特的表达特征。

7 干旱诱导下*GuHSF*基因与黄酮合成关键基因的相关性分析

为探究*HSF*家族成员是否介导干旱诱导的黄酮类物质生物合成, 本研究分析了干旱胁迫下乌拉尔甘草的转录组数据, 筛选出5个时间点的*HSF*基因及黄酮合成相关基因, 并开展*K*-Means聚类分析。如图7所示, 455个基因被划分为10个基因簇, 各基因簇呈现出特异性的干旱响应表达模式。其中, 基因簇4包含34个基因, 其表达水平随干旱诱导进程持续升高, 表现出对干旱胁迫的持续性响应特征。基于基因注释结果, 本研究进一步从基因簇4中筛选出黄酮合成相关基因与*HSF*基因, 并采用皮尔逊相关系数法分析这些基因表达水平的相关性, 以揭示*HSF*家族成员与黄酮合成途径的潜在关联。在基因簇4中, 共鉴定到5个*HSF*基因(*GuHSF1*、*GuHSF4*、*GuHSF9*、*GuHSF11*、*GuHSF21*), 以及7类黄酮合成关键基因, 具体包括*4CL2/10* (4-香豆酸辅酶A连接酶)、*I2'H* (异黄酮2′羟化酶)、*UGT13/88A19* (糖基转移酶)、*F5H* (阿魏酸-5-羟基化酶)、*UFGT5* (花青素3-*O*-葡萄糖基转移酶)、*F3'5'H* (黄酮3′,5′-羟化酶)、*F3'H* (黄酮3′-羟化酶)。相关性分析结果显示, *GuHSF*家族成员间存在显著强相关关系: *GuHSF1*与*GuHSF9*、*GuHSF1*与*GuHSF21*、*GuHSF4*与*GuHSF11*、*GuHSF9*与*GuHSF21*的Pearson相关系数均高于0.85 ($P < 0.05$)。更为重要的是, 部分*HSF*基因与黄酮合成关键基因呈现显著强相关性 ($r > 0.85$, $P < 0.05$), 具体关键关联基因对如下: *GuHSF1*分别与*F5H*、*UGT13*、*4CL10*、*I2'H*强相关; *GuHSF4*分别与*UFGT5*、*F3'5'H*、*F3'H*、*4CL2*、*I2'H*强相关; *GuHSF9*分别与*4CL10*、*I2'H*、*4CL2*、*UFGT5*、*F3'H*、*F3'5'H*强相关; *GuHSF11*分别与*UGT88A19*、*UFGT5*、*F3'5'H*、*4CL2*、*I2'H*、*F3'H*强相关; *GuHSF21*分别与*F5H*、*UGT13*、*4CL10*、*I2'H*强相关。上述结果初步表明, 部分*GuHSF*基因可能通过直接调控上述黄酮合成关键基因的表达, 参与乌拉尔甘草黄酮类物质的生物合成。

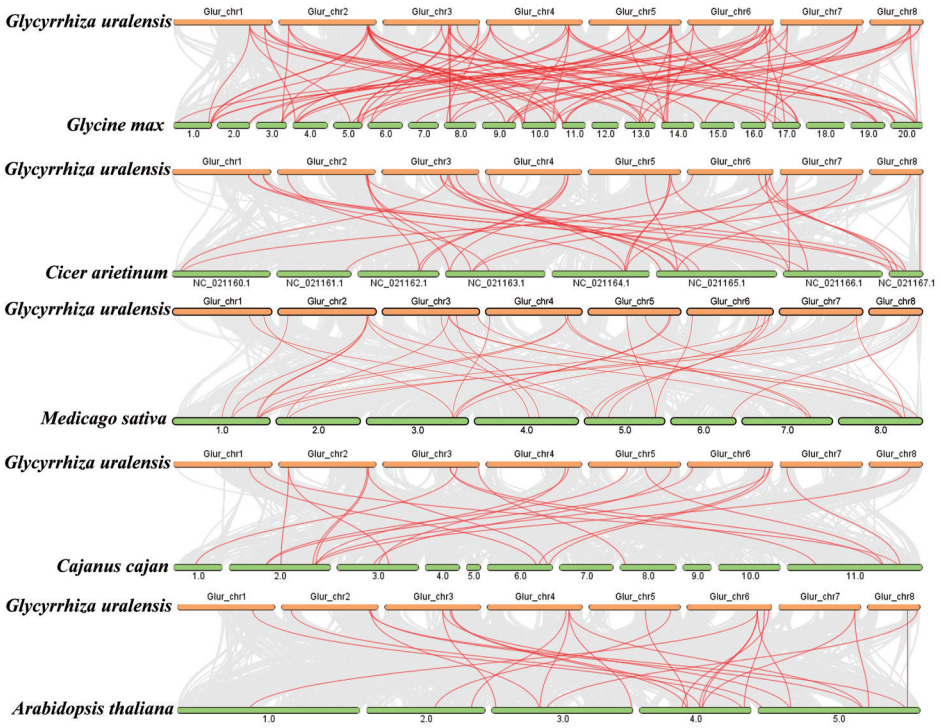


Figure 5 Collinear analysis of *GuHSFs* between *G. uralensis* and 5 other plant species (*Arabidopsis thaliana*, *Cajanus cajan*, *Glycine max*, *Medicago sativa*, and *Cicer arietinum*)

讨论

乌拉尔甘草是我国传统大宗中药材, 因其补脾益气、调和诸药等功效广泛应用于中医药领域[1]。黄酮类和萜类化合物是其核心活性成分, 已被证实具有多种药理活性, 是决定其品质的关键[1]。已有研究表明, 干旱、盐胁迫等非生物胁迫可诱导甘草黄酮合成关键基因表达, 促进黄酮积累[4,21]。干旱作为西北甘草主产区主要环境限制因子, 解析其调控甘草黄酮积累的分子机制, 对阐明甘草抗逆与品质形成的关联、指导优质抗逆甘草育种具有重要意义。HSF是植物逆境应答的核心转录调控因子, 通过调控下游基因表达参与胁迫响应及次生代谢[10,22,23]。推测HSF可能介导干旱信号, 调控甘草黄酮合成通路, 其相关机制是解析甘草干旱适应策略的关键。但目前乌拉尔甘草*HSF*基因家族的系统鉴定与功能研究仍较匮乏, 限制了对其介导黄酮合成调控通路的解析。本研究聚焦该基因家族, 从基因组鉴定、系统发育特征、保守基序、启动子元件、共线性及干旱响应表达特征等方面开展系统分析, 并结合黄酮合成相关基因的共表达关联挖掘其功能, 为后续解析其调控机制提供基础数据和线索。

本研究使用TBtools软件从乌拉尔甘草基因组中鉴定到26个*HSF*基因。对该家族蛋白的保守基序分析显示, 同一亚组成员的保守基序组成高度相, 尤其是C亚家族, 其保守基序的种类和数量最为一致, 表明这些保守基序可能是亚组成员功能相似性的关键。相反, 不同组之间的保守基序组成差异则反映了HSF蛋白在进化过程中经历扩张与分化的动态过程。启动子顺式作用元件在基因表达调控中具有关键作用[6]。对*HSF*家族成员启动子元件分析的结果显示, 启动子区域主要包括激素响应、生物和非生物胁迫响应元件、昼夜节律控制元件、光响应元件及类黄酮生物合成基因调控元件。这些结果进一步表明*GuHSF*基因在多种生物和非生物胁迫、昼夜节律及黄酮生物合成等途径中的广泛作用。

基因组进化和结构的多样性在很大程度上是由串联和片段重复事件驱动的, 这些事件构成了基因组演化的关键机制[24,25]。本研究鉴定的12对*HSF*家族片段重复事件中, A亚家族占比最高 (7对), 且未发现串联重复事件, 这与多数植物*HSF*基因家族依赖片段重复实现扩张的进化特征一致, 进一步证实片段重复是驱动乌拉尔甘草*HSF*家族扩增的核心机制。不同演化时期的片段重复事件及纯化选择作用, 既为家族扩张提供了物质基础, 也保障了基因功能的保守性, 这可能是甘草*HSF*家族在长期进化中维持逆境响应核心功能的重要原因。此外, 通过共线性分析, 发现*GuHSF*基因与大豆的*HSF*基因之间表现出最强的共线性关系, 提示这些基因可能在相似的生态适应性中发挥重要作用。

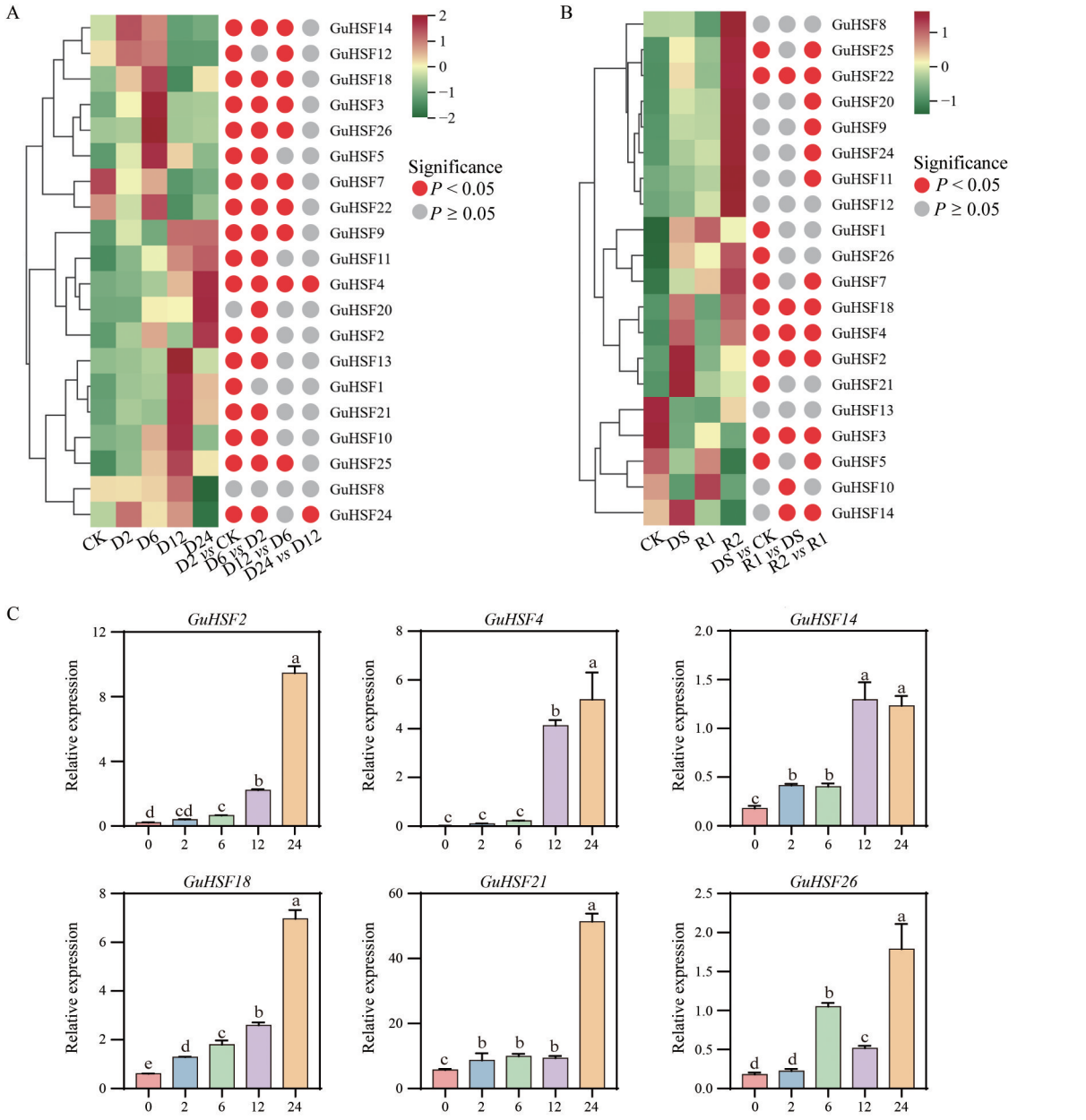


Figure 6 Expression patterns of the *GuHSF* genes under drought conditions. A: Expression pattern analysis of *GuHSF* genes under 20% PEG6000 treatment. CK, D2, D6, D12, and D24 represent 0, 2, 6, 12, and 24 h after treatment with 20% PEG 6000, respectively; B: Expression pattern analysis of *GuHSF* genes under water deficit and rehydration conditions. CK, DS, R1, and R2 represent the control group, 7-day drought stress, 7 days of rehydration after stress, and 14 days of rehydration after stress, respectively; C: The expression levels of the *GuHSF* genes under drought conditions. Significance analysis was performed using one-way ANOVA, and different lowercase letters indicate significant differences at $P < 0.05$

*HSF*家族基因通常在干旱胁迫下受到显著诱导[26,27]。在花生和芝麻等植物中, 部分*HSF*基因在干旱条件下的表达水平显著上升[26,27]。本研究基于先前的20% PEG6000溶液诱导的转录组数据分析表明, *GuHSF*家族基因普遍受到干旱胁迫的诱导, 如*GuHSF1*、*GuHSF3*、*GuHSF4*、*GuHSF5*、*GuHSF10*、*GuHSF13*、*GuHSF14*、*GuHSF18*、*GuHSF21*、*GuHSF26* (图6A)。此外, 对自然干旱处理的乌拉尔甘草转录组数据分析结果显示, *GuHSF2*、*GuHSF4*、*GuHSF18*和*GuHSF26*这4个基因在干旱条件下的表达显著上升, 并在复水处理过程中恢复至未干旱状态的水平, 进一步表明它们在干旱胁迫下的重要调节作用。进一步对上述干旱响应基因的启动子顺式作用元件分析发现, *GuHSF2*、*GuHSF4*、*GuHSF10*、*GuHSF14*和*GuHSF22*

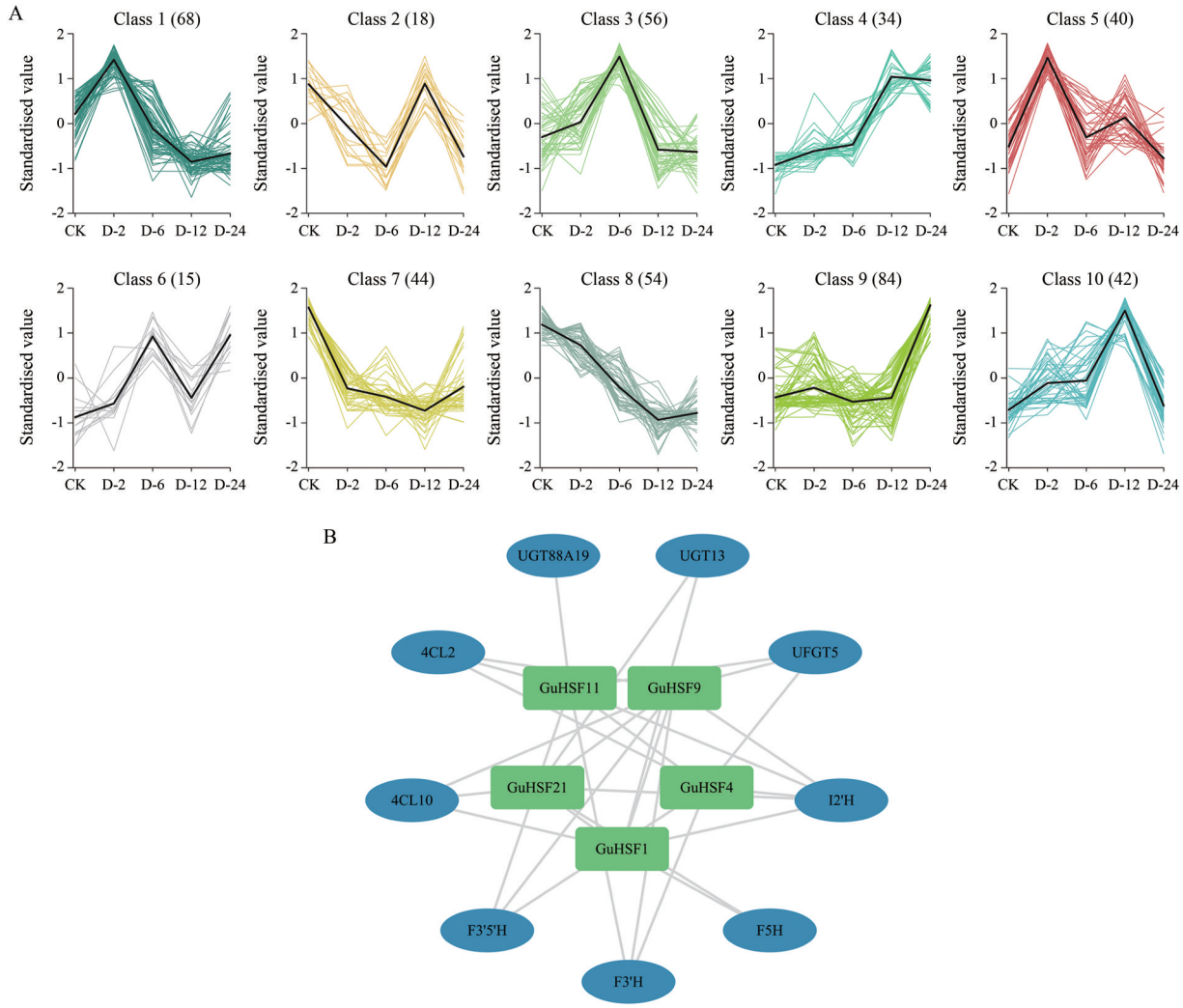


Figure 7 Identification of key flavonoid synthesis genes downstream of transcription factor HSF. A: *K*-Means analysis of *GuHSF* genes and flavonoid synthesis-related genes in *G. uralensis* under drought conditions; B: Correlation analysis of *GuHSF* genes and key flavonoid synthesis genes

基因启动子区域内至少包含一个响应干旱的顺式作用元件, 从调控元件层面佐证了这些基因参与干旱胁迫响应的潜在功能。

参与黄酮生物合成的转录因子通常通过直接调控黄酮合成关键基因的表达来发挥作用[5,28-30]。例如, 在香蕉中, HSF 转录因子能够调节 *FLS* 和 *F3'5'H1* 启动子的结合, 从而促进黄酮醇的积累[29]。在杨树中, HSFA5a 通过正向调控 *CHS1*、*F3'H2* 和 *FLS1/4* 的表达, 参与黄酮的生物合成[30]。据此推测, 乌拉尔甘草 HSF 可能通过直接调控黄酮合成相关基因的表达介导黄酮积累。为初步验证这一假设, 作者对 20% PEG6000 诱导的乌拉尔甘草转录组数据进行 *K*-Means 聚类分析, 筛选出干旱胁迫下持续上调的基因簇 (簇 4)。进一步分析发现, 簇 4 中 5 个 *GuHSF* 基因 (*GuHSF1*、*GuHSF4*、*GuHSF9*、*GuHSF11* 和 *GuHSF21*) 与 7 类黄酮合成关键基因 (*4CL2/10*、*I2'H*、*UGT13/88A19*、*F5H*、*UFGT5*、*F3'5'H* 和 *F3'H*) 具有较高的相关性, 显示出共表达特征。其中, *GuHSF4* 在 PEG6000 模拟干旱和自然干旱条件下均被显著诱导, 且与 *UFGT5*、*F3'5'H*、*F3'H*、*4CL2*、*I2'H* 的表达高度相关, 提示 *GuHSF4* 可能在干旱诱导的甘草黄酮生物合成中发挥核心调控作用。

综上所述, 本研究在乌拉尔甘草基因组中鉴定了 26 个 *HSF* 家族成员, 系统分析了其理化性质、系统发育关系、保守基序、启动子顺式作用元件及共线性特征, 并揭示了 *HSF* 家族成员的干旱响应表达模式及其与黄酮合成相关基因的共表达关联, 为后续深入探究干旱条件下 *HSF* 家族调控黄酮合成的分子机制提供了重要线索和研究方向。但本研究仅通过生物信息学分析和转录组数据挖掘推测基因功能及调控关联, 尚未开展体内外功能验证实验, 也未明确 HSF 与黄酮合

成基因的直接调控关系。鉴于*GuHSF4*在干旱响应和黄酮合成共表达分析中表现出核心特征, 未来将聚焦该基因开展功能验证, 结合转基因、ChIP-qPCR等技术明确其调控黄酮合成的分子机制, 为甘草抗逆育种及品质改良提供理论依据。

作者贡献: 杨明明负责实验设计、数据分析和手稿撰写; 陈昶泞、张乃曼和王雪媛负责实验和数据分析; 董林林负责研究项目整体规划和论文指导。

利益冲突: 所有作者声明不存在利益冲突。

References

[1] Wu LW, Ma TY, Zang CX, et al. *Glycyrrhiza*, a commonly used medicinal herb: Review of species classification, pharmacology, active ingredient biosynthesis, and synthetic biology [J]. J Adv Res, 2025, 75: 249-270.

[2] Kao TC, Wu CH, Yen GC. Bioactivity and potential health benefits of licorice [J]. J Agric Food Chem, 2014, 62: 542-553.

[3] Wu LW, Xu ZC, Wang Q, et al. Systematic screening and analysis of bZIP transcription factors in *Glycyrrhiza uralensis* and their response to ABA stress [J]. Acta Pharm Sin (药学学报), 2022, 57: 818-830.

[4] Li Y, Jiang D, Liu XY, et al. Multi-omics analysis provides crucial insights into the drought adaptation of *Glycyrrhiza uralensis* Fisch [J]. J Agric Food Chem, 2023, 71: 5391-5402.

[5] Yang MM, Liu L, Li MH, et al. Dehydration-responsive element-binding transcription factor, GuDREB35, positively regulates drought tolerance in licorice (*Glycyrrhiza uralensis*) [J]. Int J Biol Macromol, 2025, 306: 141829.

[6] Li GM, Shi XY, Lin QM, et al. Genome-wide identification and expression analysis of heat shock transcription factors in *Camellia sinensis* under abiotic stress [J]. Plants (Basel), 2025, 14: 697.

[7] Zhang H, Zhu J, Gong Z, et al. Abiotic stress responses in plants [J]. Nat Rev Genet, 2022, 23: 104-119.

[8] Zhang MM, Zhang SQ. Mitogen-activated protein kinase cascades in plant signaling [J]. J Integr Plant Biol, 2022, 64: 301-341.

[9] Yang L, Fang SY, Liu L, et al. WRKY transcription factors: Hubs for regulating plant growth and stress responses [J]. J Integr Plant Biol, 2025, 67: 488-509.

[10] Jacob P, Hirt H, Bendahmane A, The heat-shock protein/chaperone network and multiple stress resistance [J]. Plant Biotechnol J, 2017, 15: 405-414.

[11] Zhao MX, Hu XJ. The research progress of heat shock transcription factors in plants [J]. Botanical Res (植物学研究), 2018, (07): 158-163.

[12] Nover L, Bharti K, Döring P, et al. Arabidopsis and the heat stress transcription factor world: How many heat stress transcription factors do we need? [J]. Cell Stress Chaperones, 2001, 6: 177-189.

[13] Guo JK, Wu J, Ji Q, et al. Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis* [J]. J Genet Genomics, 2008, 35: 105-118.

[14] Li PS, Yu TF, He GH, et al. Genome-wide analysis of the HSF family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses [J]. BMC Genomics, 2014, 15: 1009.

[15] Ma J, Zhang GZ, Ye YC, et al. Genome-wide identification and expression analysis of HSF transcription factors in alfalfa (*Medicago sativa*) under abiotic stress [J]. Plants (Basel), 2022, 11: 2763.

[16] Wang Z, Wang PP, He JJ, et al. Genome-wide analysis of the HSF gene family reveals its role in *Astragalus mongholicus* under different light conditions [J]. Biology (Basel), 2024, 13: 280.

[17] Yokotani N, Ichikawa T, Kondou Y, et al. Expression of rice heat stress transcription factor OsHsfA2e enhances tolerance to environmental stresses in transgenic *Arabidopsis* [J]. Planta, 2008, 227: 957-967.

[18] He MM, Wu JT, Yang Y, et al. The HsfA1a-BAG5b module mediates thermotolerance through activating autophagy in tomato [J]. Plant Physiol, 2025, 199: kiaf526.

[19] Gao L, Pan LL, Shi YT, et al. Genetic variation in a heat shock transcription factor modulates cold tolerance in maize [J]. Mol Plant, 2024, 17: 1423-1438.

[20] Wang N, Liu WJ, Yu L, et al. HEAT SHOCK FACTOR A8a modulates flavonoid synthesis and drought tolerance [J]. Plant Physiol, 2020, 184: 1273-1290.

[21] Wang CC, Chen LH, Cai ZC, et al. Metabolite profiling and transcriptome analysis explains difference in accumulation of bioactive constituents in licorice (*Glycyrrhiza uralensis*) under salt stress [J]. Front Plant Sci, 2021, 12: 727882.

[22] Arrigo A. Human small heat shock proteins: Protein interactomes of homo- and hetero-oligomeric complexes: An update [J]. FEBS Lett, 2013, 587: 1959-1969.

[23] Scarpeci TE, Zanor MI, Valle EM. Investigating the role of plant heat shock proteins during oxidative stress [J]. Plant Signal Behav, 2008, 3: 856-857.

[24] Vision TJ, Brown DG, Tanksley SD. The origins of genomic duplications in Arabidopsis [J]. Science, 2000, 290: 2114-2117.

[25] Huang FY, Ye XW, Wang ZJ, et al. The prohibitins (PHB) gene family in tomato: Bioinformatic identification and expression analysis under abiotic and phytohormone stresses [J]. GM Crops Food, 2021, 12: 535-550.

[26] Wang Q, Zhang ZB, Guo C, et al. Hsf transcription factor gene family in peanut (*Arachis hypogaea* L.): Genome-wide characterization and expression analysis under drought and salt stresses [J]. Front Plant Sci, 2023, 14: 1214732.

[27] Dossa K, Diouf D, Cissé N. Genome-wide investigation of Hsf genes in sesame reveals their segmental duplication expansion

and their active role in drought stress response [J]. Front Plant Sci, 2016, 7: 1522.

[28] Chen C, Zhou XF, Cao B, et al. The transcription factor MYB8 positively regulates flavonoid biosynthesis of *Scutellaria baicalensis* in response to drought stress [J]. Plant Cell Environ, 2025, 48: 8898-8914.

[29] Song Q, He F, Kong LF, et al. The IAA17.1/HSFA5a module enhances salt tolerance in *Populus tomentosa* by regulating flavonol biosynthesis and ROS levels in lateral roots [J]. New Phytol, 2024, 241: 592-606.

[30] Naik J, Rajput R, Singh S, et al. Heat-responsive MaHSF11 transcriptional activator positively regulates flavonol biosynthesis and flavonoid B-ring hydroxylation in banana (*Musa acuminata*) [J]. Plant J, 2025, 121: e70084.