

DOI: 10. 11686/cyxb2024133

<http://cyxb.magtech.com.cn>

马超, 孙熙婧, 冯雅岚, 等. 紫花苜蓿 *GLK* 基因家族鉴定及渗透胁迫下的表达分析. 草业学报, 2025, 34(1): 174—190.

MA Chao, SUN Xi-jing, FENG Ya-lan, *et al.* Genome-wide identification of the *GLK* gene family in alfalfa and their transcript profiles under osmotic stress. Acta Prataculturae Sinica, 2025, 34(1): 174—190.

紫花苜蓿 *GLK* 基因家族鉴定及渗透胁迫下的表达分析

马超^{1**}, 孙熙婧^{1**}, 冯雅岚², 周爽¹, 琚吉浩¹, 吴毅¹, 王添宁¹, 郭彬彬¹, 张均^{1*}

(1. 河南科技大学农学院, 河南 洛阳 471000; 2. 武昌理工学院生命科学院, 湖北 武汉 430000)

摘要: *GLK* (Golden2-like 或 G2-like) 属于 GARP 超家族, 是植物生长发育过程中的重要转录因子, 在调节植物叶绿体发育、叶绿素生物合成和非生物胁迫响应等方面发挥着重要作用。目前, *GLK* 基因家族成员已在多个物种中被系统鉴定, 但在四倍体紫花苜蓿全基因组水平上仍然知之甚少。利用生物信息学的方法, 在“新疆大叶”紫花苜蓿基因组中鉴定到 100 个 *GLK* 基因 (*MsGLKs*), 并对其理化性质、染色体定位、系统进化关系、启动子顺式作用元件以及渗透胁迫和外源脱落酸 (ABA) 处理下的表达模式进行了分析。结果显示, 100 个 *MsGLK* 基因在 32 条染色体上不均匀分布, 蛋白序列长度为 201~860 个氨基酸。根据系统发育分析结果, 将 *MsGLK* 家族成员分为 13 个组。共线性分析表明, 在紫花苜蓿基因组中共发现 193 个 *MsGLK* 基因重复事件, 基因非同义替代数/同义替代数 (*Ka/Ks*) 分析显示, 大部分重复基因对经历了纯化选择。*MsGLK* 基因启动子的顺式作用元件广泛参与了植物生长发育、激素响应和胁迫反应。基因表达数据显示, 12 个基因的表达具有组织特异性, 25 个基因在所有组织中表达。RT-qPCR 检测发现, *MsGLK* 基因在干旱胁迫、盐胁迫和外源 ABA 处理下均有一定程度的响应。研究结果将为进一步探索 *MsGLK* 基因的功能和紫花苜蓿抗逆性遗传改良提供参考。

关键词: 紫花苜蓿; *GLK* 基因; 生物信息学; 渗透胁迫; 表达模式

Genome-wide identification of the *GLK* gene family in alfalfa and their transcript profiles under osmotic stress

MA Chao^{1**}, SUN Xi-jing^{1**}, FENG Ya-lan², ZHOU Shuang¹, JU Ji-hao¹, WU Yi¹, WANG Tian-ning¹, GUO Bin-bin¹, ZHANG Jun^{1*}

1. Agronomy College, Henan University of Science and Technology, Luoyang 471000, China; 2. College of Life Science, Wuchang University of Technology, Wuhan 430000, China

Abstract: *GLK* (Golden2-like or G2-like) transcription factors belong to the GARP superfamily, whose members play crucial roles in plant growth and development. Members of the *GLK* gene family play important roles in regulating plant chloroplast development, chlorophyll biosynthesis, and the abiotic stress response. The *GLK* gene family has been systematically identified in many plant species, but not yet in tetraploid alfalfa (*Medicago sativa*). In this study, we used bioinformatics methods to identify 100 *GLK* genes (*MsGLKs*) in the genome of the alfalfa cultivar “Xinjiang Daye”. Further analyses were conducted to explore the physicochemical properties of their putative encoded products, chromosome localization, phylogenetic relationships, *cis*-acting elements in their

收稿日期: 2024-04-18; 改回日期: 2024-06-05

基金项目: 国家自然科学基金 (32372227), 河南省高等学校青年骨干教师培养计划 (2021GGJS050), 河南省自然科学基金 (242300421319) 和河南省科技研发计划联合基金项目 (232103810021) 资助。

作者简介: 马超 (1984—), 男, 河南洛阳人, 副教授, 博士。E-mail: machao840508@163.com; 孙熙婧 (2003—), 女, 河南洛阳人, 在读本科。E-mail: lierizhaoming2003@163.com。 ** 共同第一作者 These authors contributed equally to this work.

* 通信作者 Corresponding author. E-mail: zhangjun0105@126.com

promoter regions, and transcript profiles under osmotic stress and exogenous abscisic acid (ABA) treatment. The results showed that the 100 *MsGLK* genes were unevenly distributed on the 32 chromosomes, and encoded polypeptides with sequence lengths ranging from 201 to 860 amino acids. In a phylogenetic analysis, the *MsGLK* family members were divided into 13 groups. A collinearity analysis detected 193 *MsGLK* gene duplicates in the alfalfa genome, and a analysis of the ratio of the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site (Ka/Ks) revealed that most of the duplicated gene pairs have undergone purification selection. A range of *cis*-acting elements were detected in *MsGLK* gene promoter regions, and were involved in plant growth and development, hormone responses, and stress responses. Analyses of gene transcript profiles revealed that 12 *MsGLKs* showed tissue-specific expression patterns, and 25 *MsGLKs* were expressed in all tissues. Further RT-qPCR analyses revealed that some *MsGLK* genes were activated to some degree in response to drought stress, salt stress, and exogenous ABA treatment. The results of this study provide reference information for further research on the function of *MsGLK* genes and for the genetic improvement of stress resistance in alfalfa.

Key words: alfalfa; *GLK* gene; bioinformatics; osmotic stress; expression patterns

转录因子(transcription factors, TFs)是指能够与特定 DNA 序列结合并对基因的转录有激活或抑制作用的蛋白质,在信号转导和基因表达调控中起着至关重要的作用^[1]。作为发育过程和应激反应的重要调节因子,TFs 通过调节下游胁迫响应基因的时空表达模式,在多种非生物胁迫反应中扮演着重要的角色^[2]。*GLK* 属于 GARP (Golden2、ARR-B、Psr1)超家族,在陆地植物中广泛存在,是植物生长发育中的重要转录因子^[3]。*GLK* 基因在 GARP 家族中是单系的,基因复制在单子叶和双子叶植物中独立发生^[4]。典型的 *GLK* 蛋白通常包括两个高保守性区域,分别为 Myb-DNA 结合结构域(Myb-DNA binding domain, DBD)和位于 C 端的 *GLK*/C-terminalbox (GCT-box)^[5]。

GLK 转录因子与植物叶绿体形成与发育、叶绿素生物合成密切相关^[6]。在拟南芥(*Arabidopsis thaliana*)和小立碗藓(*Physcomitrella patens*)中,*GLK1* 和 *GLK2* 基因以成对形式冗余地参与叶绿体的形成^[7]。拟南芥 *glk1/2* 双突变体表现为淡绿色,叶肉细胞叶绿体发育不良,类囊体膜稀少,基粒堆叠缺陷,叶绿素合成和光捕获相关的基因表达水平降低^[8]。而在玉米(*Zea mays*)和高粱(*Sorghum bicolor*)中,*GLK* 基因存在组织表达特异性,即 *GLK1* 和 *GLK2* 分别在维管束鞘(bundle sheath, BS)细胞和叶肉(mesophyll, M)细胞中表达,推测可能与部分 C_4 植物叶绿体发育特性相关^[9]。茶树(*Camellia sinensis*)*CsGLKs* 能参与紫外线-B(UV-B)介导的叶绿素合成和类黄酮积累的调节^[10]。大量研究发现,*GLKs* 可以与叶绿体光合电子传递链中的核编码基因(*PhANGs*)结合,特别是光系统 I 和 II 以及叶绿素生物合成途径中的基因^[11]。除叶片外,*GLKs* 也影响着果实和部分非绿色组织中的叶绿体发育。在番茄(*Solanum lycopersicum*)中,过量表达 *SlGLK1* 和 *SlGLK2* 能够使未成熟果实中叶绿体数量和叶绿素含量增加,并且与 *CsGLKs* 类似,*SlGLK2* 可以在 UV-B 的辐射下通过诱导 *SlUVB8* 从而促进果实叶绿体的发育^[12-13]。在水稻(*Oryza sativa*)中的研究发现,*OsGLK1* 的异位表达可以诱导非绿色组织中叶绿体的产生^[14]。

GLK 转录因子也可以通过优化不同生物和非生物环境胁迫条件下的光合作用,参与多种应激反应和抗病过程^[15]。*AtGLK1* 被证明能够增加拟南芥对禾谷镰刀菌的非宿主抗性,并参与对黄瓜(*Cucumis sativus*)花叶病毒感染的防御过程^[16-17]。*AhGLK1b* 的异位表达可以增强植物对真菌和细菌病原体以及其他环境压力的抗性^[18]。马铃薯(*Solanum tuberosum*)病毒 X(PVX)激活的细胞内免疫受体 Rx1 与 *NbGln1* 相互作用,介导烟草(*Nicotiana benthamiana*)的抗病毒活性^[19]。另外,在干旱和低温胁迫下,过量表达 *GhGLK1* 拟南芥植株对两种胁迫的抗性增加,而 *GhGLK1* 基因敲除的陆地棉(*Gossypium hirsutum*)植株抗旱性和抗寒性均降低^[20]。*AhGLK1* 还可通过调节叶绿素合成,间接影响花生(*Arachis hypogaea*)在干旱后的恢复过程^[21]。而在耐霜转基因甘蓝型油菜(*Brassica napus*)中,*GLK1* 的积累参与了植物的冷驯化过程^[22]。莲藕(*Nelumbo nucifera*)的研究显示,*GLK1* 和多酚氧化酶基因 *PPO1* 在非生物胁迫的响应中存在互作效应^[23]。此外,*GLK* 基因也参与了脱落酸(abscisic acid, ABA)的信号通路。拟南芥 *glk1/2* 双突变体在种子萌发和幼苗发育阶段对 ABA 敏感性增加,在幼苗时期表现出渗透胁迫

抗性;*GLK1/2*与*WRKY40*能够进行特异性结合,在ABA信号转导中起负调控作用,有助于提高植物在非生物胁迫下的适应性^[24]。

紫花苜蓿(*Medicago sativa*)是世界上最重要、经济价值最高的饲料作物之一,具有粗蛋白含量高、营养成分丰富、适口性好、适应性强等特点^[25]。干旱和盐分是威胁牧草生长发育的重要非生物胁迫因子,制约着牧草产业和草食畜牧业的快速发展^[26]。截止目前,*GLK*转录因子已在玉米^[22]、小麦(*Triticum aestivum*)^[27]、番茄^[28]、谷子(*Setaria italica*)^[29]、大豆(*Glycine max*)^[30]、烟草^[31]等植物中进行了相关鉴定,但在紫花苜蓿中未见报道。对于*GLK*转录因子的研究目前多集中在植物光合功能方面,对于其在非生物胁迫尤其是渗透胁迫相关的研究相对较少。本研究利用生物信息学的方法,对四倍体紫花苜蓿基因组中*GLK*基因家族进行鉴定和表征,在此基础上利用实时荧光定量PCR(quantitative real-time PCR, RT-qPCR)技术,分析该基因家族成员在紫花苜蓿特定生长阶段中响应干旱、盐、ABA等非生物胁迫下的表达模式,为进一步研究紫花苜蓿*GLK*基因家族生物功能提供理论依据,并期望在此基础上为紫花苜蓿种质的改良提供理论参考。

1 材料与方法

1.1 紫花苜蓿*GLK*基因家族成员的鉴定

本研究使用的拟南芥*GLK*基因家族的蛋白质序列下载自拟南芥信息资源数据库(the arabidopsis information resource, TAIR, <https://www.arabidopsis.org/>)。紫花苜蓿(新疆大叶)的全基因组文件、蛋白序列和基因组注释文件由figshare数据库(https://figshare.com/articles/dataset/genome_fasta_sequence_and_annotation_files/12327602)取得。*GLK*蛋白的保守结构域(PF00249)文件从Pfam数据库(<http://pfam.xfam.org/>)下载得到^[32]。以拟南芥的*GLK*蛋白序列作为查询序列,通过TBtools v 2.010的Blastp功能在紫花苜蓿蛋白数据库中进行比对,并利用HMMER 3.0软件的hmmsearch命令鉴定紫花苜蓿*GLK*蛋白序列,设置E-value<1e⁻¹⁰。将两种方法获得的结果取并集,使用NCBI的CDD工具(<https://www.ncbi.nlm.nih.gov/cdd/>)进行结构域验证,保留含有完整*GLK*特征结构域的序列。

1.2 紫花苜蓿*GLK*家族成员蛋白理化性质、亚细胞定位及染色体定位

利用TBtools的Protein Paramter功能分析候选蛋白序列的长度、分子量、等电点等信息^[33]。通过WoLFPSORT网站(<https://wolfsort.hgc.jp/>)预测紫花苜蓿*GLK*蛋白的亚细胞定位。基于紫花苜蓿的基因组和注释文件,通过TBtools获得*GLK*基因的染色体分布及其位置图谱。根据染色体定位信息将筛选出的基因命名为*MsGLK1*~*MsGLK100*。

1.3 紫花苜蓿*GLK*家族成员系统发育分析、基因结构和基序分析

使用MEGA 11软件,利用MUSCLE将紫花苜蓿和拟南芥的全部*GLK*蛋白序列进行多序列比对,采用邻接法(neighbor-joining, NJ)构建系统发育树,设置重复值(bootstrap replications)为1000次^[34]。在MEME工具(<https://meme-suite.org/meme/index.html>)中分析保守基序,最大基序数设为10,最佳基序宽度为10~50。从紫花苜蓿gff3注释文件中获得基因结构信息。利用TBtools的Gene Structure View功能将上述信息进行综合可视化。

1.4 紫花苜蓿*GLK*基因启动子顺式作用元件分析

从紫花苜蓿gff3注释文件中提取目标基因编码区(coding sequence, CDS)上游2000 bp的推定启动子序列,提交PlantCare网站(<http://bioinformatics.psb.ugent.be/webtools/plantcare/>)进行启动子顺式作用元件预测。利用TBtools对顺式作用元件的数量及分布进行可视化显示。

1.5 紫花苜蓿*GLK*基因重复事件和共线性分析

通过TBtools的MCScanX插件分析紫花苜蓿*GLK*基因重复事件,对共线性分析结果进行可视化处理。利用TBtools的Ka/Ks Calculat功能计算非同义替代数(number of nonsynonymous substitutions per nonsynonymous site, Ka)和同义替代数(number of synonymous substitutions per synonymous site, Ks)及其比值,过滤无效数据,推断基因进化过程中的选择模式。

1.6 紫花苜蓿 *GLK* 基因在不同组织的表达模式分析

从 NCBI 数据库下载到紫花苜蓿花、叶、根、根瘤、短茎、伸长茎 6 个组织中的转录组数据 (SRP055547), 以及干旱胁迫、盐胁迫和外源 ABA 处理下 *MsGLK* 基因的转录组学数据 (SRR7091780-SRR7091794 和 SRR7160313-SRR7160357)^[35], 利用 Hisat2 建立基因组索引, 将 clean reads 对比到参考基因组上, 获得用于估计基因表达量的每千个碱基的转录每百万映射读取的转录本 TPM (transcripts per kilobase per million) 值, 使用 TBtools 对数据进行归一化处理, 并绘制热图。

1.7 紫花苜蓿 *GLK* 基因在渗透胁迫和外源 ABA 处理下的 RT-qPCR 分析

试验选用新疆大叶紫花苜蓿种子, 于 2023 年 10 月底在河南科技大学农学院分子生物学实验室开展。挑选大小均匀、饱满完整的种子, 选择营养土作为培养基质, 放置于 22 ℃, 光照 16 h 黑暗 8 h 的培养箱中进行盆栽培养。对培养 5 周后的幼苗分别用 30% 的 PEG-6000、300 mmol·L⁻¹ 的 NaCl 和 100 μmol·L⁻¹ 的 ABA 进行处理, 分别在处理 1、3、6、12、24、48 h 对幼苗的根组织进行取样。以正常条件下幼苗作为对照, 每个处理重复 3 次。取得的样品用液氮速冻后置于 -80 ℃ 冰箱保存。

利用 TRIzol (TaKaRa, 9108Q) 试剂提取总 RNA, 利用 PrimeScript RT Master Mix (TaKaRa, RR036) 反转录获得 cDNA。使用 Primer 5.0 设计引物, 选择 *MsActin* (AA660796) 作为内参基因^[36]。引物由上海生工生物工程有限公司合成, 引物序列见表 1。RT-qPCR 分析在 CFX96 Touch Real-time PCR Detection System (Bio-Rad, Hercules, CA, 美国) 进行, 反应体系参考 2×Ultra SYBR Green qPCR Mix (CISTRO, 广州) 试剂盒说明书, 基因相对表达水平由 2^{-ΔΔCt} 方法计算^[37]。

表 1 RT-qPCR 引物信息

Table 1 Primers used for RT-qPCR

基因 Gene	正向引物 Forward primer (5'-3')	反向引物 Reverse primer (5'-3')
<i>MsGLK10</i>	CAGATGAAGTCGGATGAGCCT	GGTGGTGCGGTAAGTTGGTC
<i>MsGLK17</i>	TACCACAGCAAGGGAAGCAG	GAATGGAACGCGCATGAAC
<i>MsGLK29</i>	GCAAGTGGAGGCAGAGAAGC	GGGTGGGATGGTTTGAGGA
<i>MsGLK33</i>	AAACCTGCCTCTTTACCGATTC	GTTGCTCTTTCTTGCCCTCC
<i>MsGLK36</i>	GCAAGTGGAGGCAGAGAAGC	GGGTGGGATGGTTTGAGGA
<i>MsGLK62</i>	CATCACTCATCCCTTTACCTC	CCTGCCAATCATTTTCGCTT
<i>MsGLK65</i>	AACACTCCGCTGTCCCAT	TCTCATTGCCGTCGTCATTT
<i>MsGLK75</i>	CTTCATCACTCATCCCTTTACCC	GCAGGCACCGCTCTCTACTT
<i>MsGLK83</i>	ATATCCAGGGTCATAGTATTTTCCA	CGTTGTTTTAGTTCTTCTCTCTTG
<i>MsActin</i>	ACTGGAATGGTGAAGGCTGG	TGACAATACCGTGCTCAATGG

1.8 数据处理

使用 Excel 2021 进行基因表达量数据的计算, 利用 SPSS 25.0 软件对数据进行统计分析, 绘图采用 Excel 2021 和 Adobe Illustrator 2024。

2 结果与分析

2.1 紫花苜蓿 *GLK* 基因家族成员的鉴定与染色体定位

通过 HMM 和 Blast 检索, 从“新疆大叶”紫花苜蓿基因组中共鉴定到 *GLK* 基因 100 个。“新疆大叶”紫花苜蓿为同源四倍体, 具有 32 条染色体。染色体定位分析结果表明, 这 100 个基因中有 3 个基因未能定位到染色体上, 剩余 97 个基因在 32 条染色体上不均匀分布 (图 1)。根据 *MsGLK* 基因在染色体上的分布顺序, 将 97 个基因命名为 *MsGLK1*~*MsGLK97*, 未能定位到染色体上的 3 个基因命名为 *MsGLK98*~*MsGLK100*。每条染色体上 *MsGLK* 家族成员的数量从 1~7 个不等, 其中 chr2.4 染色体上有 7 个家族成员, 分布数量最多, 其次为 chr2.1、

chr2. 2、chr2. 3、chr7. 1、chr7. 2 和 chr7. 3 染色体,各有 6 个家族成员,而 chr3. 1、chr4. 2、chr4. 3、chr4. 4、chr5. 2、chr6. 2、chr6. 3 和 chr6. 4 染色体仅有 1 个家族成员。此外,*MsGLK* 基因中存在一个串联重复基因对:*MsGLK94* 和 *MsGLK95*,位于 chr8. 4 染色体。

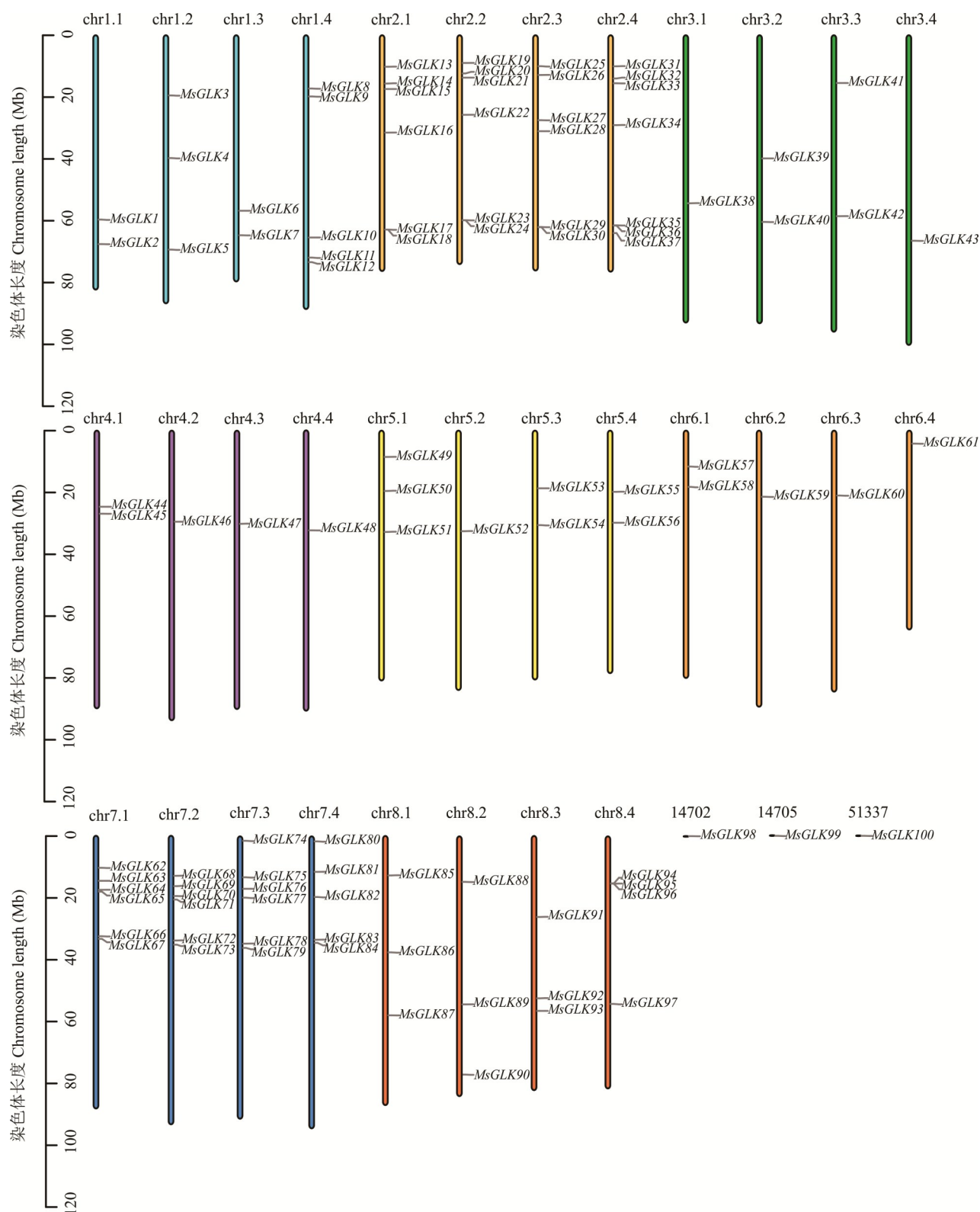


图1 紫花苜蓿 *GLK* 基因家族成员的染色体分布

Fig. 1 Chromosome distribution of *GLK* gene family members in alfalfa

Chr: 染色体 Chromosome.

表 2 紫花苜蓿 GLK 家族的基本信息分析

Table 2 Basic information analysis of GLK gene family of alfalfa

基因名 Gene name	基因登录号 Gene ID	PL (aa)	Mw (kDa)	pI	Sl	基因名 Gene name	基因登录号 Gene ID	PL (aa)	Mw (kDa)	pI	Sl
<i>MsGLK1</i>	MS. gene005086. t1	388	43. 19	4. 73	Nuc	<i>MsGLK51</i>	MS. gene76602. t1	352	39. 86	5. 75	Nuc
<i>MsGLK2</i>	MS. gene005866. t1	378	42. 16	7. 72	Nuc	<i>MsGLK52</i>	MS. gene55246. t1	352	40. 14	5. 54	Nuc
<i>MsGLK3</i>	MS. gene041038. t1	281	32. 18	8. 66	Nuc	<i>MsGLK53</i>	MS. gene028039. t1	277	31. 24	6. 03	Nuc
<i>MsGLK4</i>	MS. gene37289. t1	231	26. 06	8. 63	Nuc	<i>MsGLK54</i>	MS. gene045645. t1	353	40. 08	5. 75	Nuc
<i>MsGLK5</i>	MS. gene60590. t1	379	42. 20	7. 18	Nuc	<i>MsGLK55</i>	MS. gene070117. t1	277	31. 32	5. 82	Nuc
<i>MsGLK6</i>	MS. gene43963. t1	388	43. 13	4. 76	Nuc	<i>MsGLK56</i>	MS. gene032930. t1	352	39. 83	5. 75	Nuc
<i>MsGLK7</i>	MS. gene001686. t1	378	42. 16	7. 73	Nuc	<i>MsGLK57</i>	MS. gene87987. t1	393	44. 55	6. 20	Nuc
<i>MsGLK8</i>	MS. gene070698. t1	437	48. 03	6. 98	Nuc	<i>MsGLK58</i>	MS. gene85671. t1	308	34. 34	9. 37	Nuc
<i>MsGLK9</i>	MS. gene026358. t1	201	23. 51	8. 81	Nuc	<i>MsGLK59</i>	MS. gene68923. t1	393	44. 48	6. 16	Nuc
<i>MsGLK10</i>	MS. gene036427. t1	388	43. 13	4. 76	Nuc	<i>MsGLK60</i>	MS. gene042587. t1	393	44. 45	6. 07	Nuc
<i>MsGLK11</i>	MS. gene21397. t1	297	33. 01	6. 43	Nuc	<i>MsGLK61</i>	MS. gene66964. t1	393	44. 48	6. 16	Nuc
<i>MsGLK12</i>	MS. gene60862. t1	378	42. 15	7. 72	Nuc	<i>MsGLK62</i>	MS. gene018653. t1	489	53. 80	5. 42	Nuc
<i>MsGLK13</i>	MS. gene055776. t1	423	46. 76	5. 86	Nuc	<i>MsGLK63</i>	MS. gene017974. t1	359	40. 43	6. 92	Nuc
<i>MsGLK14</i>	MS. gene35408. t1	373	41. 81	7. 74	Nuc	<i>MsGLK64</i>	MS. gene017688. t1	301	32. 94	6. 56	Nuc
<i>MsGLK15</i>	MS. gene35289. t1	257	28. 82	8. 91	Nuc	<i>MsGLK65</i>	MS. gene013905. t1	281	30. 65	5. 79	Nuc
<i>MsGLK16</i>	MS. gene62480. t1	365	40. 90	8. 66	Cyt	<i>MsGLK66</i>	MS. gene99484. t1	350	39. 92	5. 48	Nuc
<i>MsGLK17</i>	MS. gene003164. t1	860	94. 37	8. 47	Chl	<i>MsGLK67</i>	MS. gene84188. t1	238	26. 82	9. 64	Nuc
<i>MsGLK18</i>	MS. gene003165. t1	330	36. 16	6. 19	Nuc	<i>MsGLK68</i>	MS. gene035191. t1	489	53. 78	5. 42	Nuc
<i>MsGLK19</i>	MS. gene043457. t1	423	46. 73	5. 86	Nuc	<i>MsGLK69</i>	MS. gene52648. t1	351	39. 53	6. 70	Nuc
<i>MsGLK20</i>	MS. gene84326. t1	373	41. 81	7. 74	Nuc	<i>MsGLK70</i>	MS. gene013966. t1	301	32. 84	6. 56	Nuc
<i>MsGLK21</i>	MS. gene073152. t1	313	35. 17	6. 98	Nuc	<i>MsGLK71</i>	MS. gene068055. t1	283	30. 79	5. 94	Nuc
<i>MsGLK22</i>	MS. gene066206. t1	375	41. 71	8. 37	Nuc	<i>MsGLK72</i>	MS. gene069022. t1	350	39. 82	5. 48	Nuc
<i>MsGLK23</i>	MS. gene01976. t1	331	36. 21	6. 23	Nuc	<i>MsGLK73</i>	MS. gene77355. t1	256	28. 91	9. 60	Nuc
<i>MsGLK24</i>	MS. gene01975. t1	382	41. 62	8. 62	Chl	<i>MsGLK74</i>	MS. gene024551. t1	439	48. 79	5. 11	Nuc
<i>MsGLK25</i>	MS. gene028866. t1	423	46. 73	5. 86	Nuc	<i>MsGLK75</i>	MS. gene23439. t1	489	53. 78	5. 42	Nuc
<i>MsGLK26</i>	MS. gene034456. t1	373	41. 83	7. 25	Nuc	<i>MsGLK76</i>	MS. gene04612. t1	359	40. 43	6. 92	Nuc
<i>MsGLK27</i>	MS. gene040544. t1	581	63. 91	5. 80	Nuc	<i>MsGLK77</i>	MS. gene58831. t1	298	32. 76	6. 56	Nuc
<i>MsGLK28</i>	MS. gene045711. t1	375	41. 72	8. 37	Nuc	<i>MsGLK78</i>	MS. gene77413. t1	350	39. 94	5. 56	Nuc
<i>MsGLK29</i>	MS. gene048293. t1	314	34. 41	6. 60	Nuc	<i>MsGLK79</i>	MS. gene94669. t1	256	28. 93	9. 78	Nuc
<i>MsGLK30</i>	MS. gene048294. t1	330	36. 16	6. 19	Nuc	<i>MsGLK80</i>	MS. gene050550. t1	439	48. 79	5. 11	Nuc
<i>MsGLK31</i>	MS. gene31432. t1	423	46. 69	5. 86	Nuc	<i>MsGLK81</i>	MS. gene96951. t1	489	53. 73	5. 37	Nuc
<i>MsGLK32</i>	MS. gene43040. t1	373	41. 79	7. 25	Nuc	<i>MsGLK82</i>	MS. gene41865. t1	283	30. 79	5. 94	Nuc
<i>MsGLK33</i>	MS. gene94587. t1	315	35. 25	7. 30	Nuc	<i>MsGLK83</i>	MS. gene95656. t1	350	39. 85	5. 48	Nuc
<i>MsGLK34</i>	MS. gene038958. t1	375	41. 74	8. 37	Nuc	<i>MsGLK84</i>	MS. gene95601. t1	238	26. 83	9. 64	Nuc
<i>MsGLK35</i>	MS. gene02861. t1	330	36. 17	6. 24	Nuc	<i>MsGLK85</i>	MS. gene036866. t1	434	47. 87	6. 66	Nuc
<i>MsGLK36</i>	MS. gene02867. t1	314	34. 42	6. 60	Nuc	<i>MsGLK86</i>	MS. gene61025. t1	352	39. 45	6. 50	Nuc
<i>MsGLK37</i>	MS. gene002965. t1	238	26. 64	8. 87	Nuc	<i>MsGLK87</i>	MS. gene068940. t1	345	38. 77	6. 64	Nuc
<i>MsGLK38</i>	MS. gene008318. t1	461	50. 91	7. 22	Nuc	<i>MsGLK88</i>	MS. gene060554. t1	434	47. 88	6. 66	Nuc
<i>MsGLK39</i>	MS. gene019228. t1	383	42. 69	7. 20	Chl	<i>MsGLK89</i>	MS. gene42755. t1	345	38. 77	6. 64	Nuc
<i>MsGLK40</i>	MS. gene061558. t1	426	47. 13	7. 99	Nuc	<i>MsGLK90</i>	MS. gene36713. t1	221	25. 38	9. 67	Chl
<i>MsGLK41</i>	MS. gene057902. t1	394	43. 82	9. 10	Nuc	<i>MsGLK91</i>	MS. gene87427. t1	650	71. 83	5. 55	Nuc
<i>MsGLK42</i>	MS. gene049714. t1	424	46. 87	7. 99	Nuc	<i>MsGLK92</i>	MS. gene36099. t1	346	38. 76	6. 64	Nuc
<i>MsGLK43</i>	MS. gene05012. t1	425	47. 00	7. 99	Nuc	<i>MsGLK93</i>	MS. gene88392. t1	302	33. 85	8. 12	Cyt
<i>MsGLK44</i>	MS. gene26549. t1	420	47. 03	8. 17	Nuc	<i>MsGLK94</i>	MS. gene056957. t1	363	40. 53	9. 07	Nuc
<i>MsGLK45</i>	MS. gene006790. t1	340	37. 99	5. 97	Nuc	<i>MsGLK95</i>	MS. gene056956. t1	434	47. 87	6. 66	Nuc

续表 Continued Table

基因名	基因登录号	PL	Mw	pI	Sl	基因名	基因登录号	PL	Mw	pI	Sl
Gene name	Gene ID	(aa)	(kDa)			Gene name	Gene ID	(aa)	(kDa)		
<i>MsGLK46</i>	MS. gene031499. t1	340	38.06	5.97	Nuc	<i>MsGLK96</i>	MS. gene038713. t1	434	47.87	6.66	Nuc
<i>MsGLK47</i>	MS. gene96791. t1	340	37.99	5.97	Nuc	<i>MsGLK97</i>	MS. gene066880. t1	346	38.90	6.68	Nuc
<i>MsGLK48</i>	MS. gene56537. t1	339	37.89	5.97	Nuc	<i>MsGLK98</i>	MS. gene00817. t1	238	26.61	8.87	Nuc
<i>MsGLK49</i>	MS. gene54226. t1	387	43.44	7.27	Nuc	<i>MsGLK99</i>	MS. gene02381. t1	238	26.66	8.87	Nuc
<i>MsGLK50</i>	MS. gene72865. t1	277	31.26	5.95	Nuc	<i>MsGLK100</i>	MS. gene072393. t1	351	39.62	6.73	Nuc

PL: 蛋白质长度 Protein length; aa: 氨基酸 Amino acid; Mw: 分子量 Molecular weight; pI: 等电点 Isoelectric point; Sl: 亚细胞定位 Subcellular localization; Nuc: 细胞核 Nucleus; Chl: 叶绿体 Chloroplast; Cyt: 细胞质 Cytoplasm.

2.2 紫花苜蓿 GLK 基因家族成员的理化性质分析

MsGLK 蛋白相对较小,序列长度为 201(MsGLK9)~860(MsGLK17)个氨基酸,对应的蛋白质分子量(molecular weight)为 23.51~94.37 kDa。蛋白质的理论等电点(theoretical isoelectric point, pI)为 4.73(MsGLK1)~9.78(MsGLK79),其中酸性蛋白质与碱性蛋白质数量大致相等,可见 MsGLK 蛋白质的酸碱分布相对均匀。此外,亚细胞定位的预测结果显示,2 个 MsGLK 蛋白位于细胞质中,4 个位于叶绿体中,剩余 94 个均定位于细胞核中(表 2)。

2.3 紫花苜蓿 GLK 基因家族成员的系统发育分析、基因结构和基序分析

为进一步探究 GLK 家族的进化关系,基于来自拟南芥^[38]与紫花苜蓿的 GLK 蛋白序列构建 NJ 系统发育树(图 2)。根据进化关系和基序分析,将 GLK 家族成员分为 13 个组,同一亚组的成员亲缘关系相近,并具有相似的结构。各组成员数量不等,最大组为第 V 组,含有 18 个 MsGLK 蛋白和 11 个 AtGLK 蛋白,最小组为第 VI 组和 XI 组,第 VI 组含有 4 个 MsGLK 蛋白,而第 XI 组含有 3 个 MsGLK 蛋白和 1 个 AtGLK 蛋白,表明其在进化过程中经历了动态变化。MsGLK 蛋白在所有亚组中皆有分布,AtGLK 蛋白仅在第 VI 组中未有分布,表明该组成员可能具有物种特异性。

根据 MsGLK 蛋白序列单独构建 NJ 系统发育树,这些蛋白也被分为 13 个亚组(图 3)。通过在线工具 MEME 预测了 MsGLK 蛋白的保守基序,共鉴定了 10 个潜在的基序(基序 1~基序 10),长度为 28~50 个氨基酸。基序 1 存在于所有蛋白中,而基序 2 仅在 MsGLK84 和 MsGLK67 中未被鉴定到,说明两个基序在 MsGLK 蛋白中高度保守。通过 NCBI 的 CD 检索工具对基序 1、2 进行注释,发现二者覆盖了 GLK 基因特异性基序 Myb-SHLQKYF。除上述 2 个保守结构域外,还观察到第 VI~VIII 组蛋白均存在基序 3,该基序被注释为 Myb-CC-LHEQLE,被认为参与磷酸盐饥饿信号等多种胁迫的信号传导^[39]。同一亚组内的基序组成相近,表明这些蛋白功能具有相似性,而不同组间基序的差异则揭示了 MsGLK 蛋白功能的多样性。

基因结构分析表明,MsGLK 基因的外显子数量从 1~17 个不等,内含子数量为 0~16 个,其中 MsGLK64、MsGLK70 和 MsGLK77 不含内含子。不同组间的基因结构存在差异,内含子数量差别较大,第 II、III 组 90% 的成员含有 4 个内含子,第 V、VII、VIII、X、XI、XII 和 XIII 组的大部分成员含有 5 个内含子,第 IX 组成员均具有 6 个内含子,第 VI 组的 4 个基因中 2 个含有 5 个内含子,2 个含有 6 个内含子,而分别含有 7~16 个和 0~5 个内含子。内含子的变异可能在 MsGLK 基因功能多样性的进化过程中起着关键作用^[40]。

2.4 紫花苜蓿 GLK 基因家族成员的启动子顺式作用元件分析

利用 PlantCare 在线工具分析了起始密码子上游 2000 bp 区域的顺式作用元件,预测了与植物生长发育、激素响应和胁迫反应相关的 13 类作用元件(图 4)。在所有作用元件中,与光响应相关的作用元件数量最多,种类最丰富,在所有基因中均有分布,表明光可能是调节 MsGLK 基因表达的重要因素。激素响应类元件数量仅次于光响应作用元件,包括生长素响应元件(TGA-element、AuxRE、AuxRR-core)、赤霉素响应元件(TATC-box、GARE-motif、P-box)、脱落酸响应元件(abscisic acid responsive element, ABRE)、茉莉酸响应元件(TGACG-motif、CGTCA-motif)和水杨酸响应元件(TCA-element),每个基因至少含有一个激素响应元件。胁迫响应元件有 4

类,分别为干旱响应元件(MBS)、低温响应元件(low temperature responsive element,LTR)、厌氧诱导调控元件(anaerobic induction responsive element,ARE)和应激防御反应调控元件(TC-rich repeats),其中LTR元件被认为同时参与盐胁迫反应的调控^[41]。69个*MsGLK*基因中至少含有3个胁迫响应元件,推测*MsGLKs*能够在胁迫的响应中起到积极作用。此外,本研究还鉴定到参与细胞周期调节(MSA-like)、昼夜节律(circadian)和类黄酮生物合成基因调控(MBSI)的顺式作用元件。

2.5 紫花苜蓿 *GLK* 基因家族成员的共线性分析

为了解*MsGLKs*基因的复制事件,利用MCScanX工具进行了共线性分析。在紫花苜蓿基因组中共发现193个*MsGLK*基因家族的片段重复事件,涉及88个*MsGLK*基因,有2个*MsGLK*基因(*MsGLK94*和*MsGLK95*)同时涉及片段重复事件和串联重复事件(图5)。这些结果表明,片段重复事件在*MsGLK*基因中广泛发生,可能在紫花苜蓿*GLK*基因家族的扩增中起着重要作用。此外,为研究*MsGLK*基因在进化中受到的选择压力,计算了涉及重复事件基因的*Ka*(每个非同义位点的碱基替换数)和*Ks*(每个同义位点的碱基替换数)值。共线性基因对中2对基因(*MsGLK45*和*MsGLK48*;*MsGLK66*和*MsGLK72*)的*Ka/Ks*大于1,表明其在进化过程中经历了正向选择,其余基因对的*Ka/Ks*均小于1,在进化过程中经历了纯化选择。

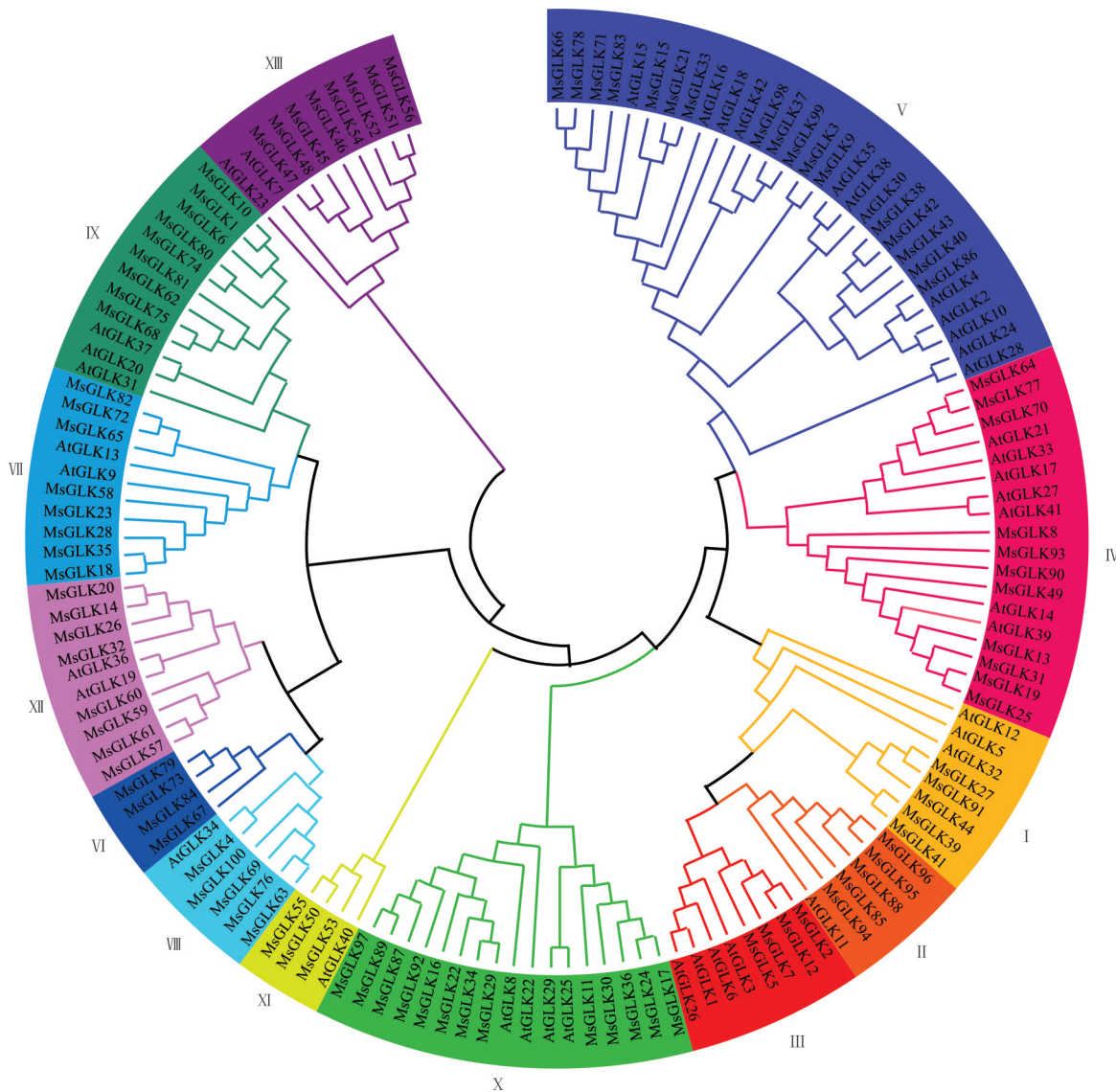


图2 紫花苜蓿与拟南芥 *GLK* 基因家族系统进化树

Fig. 2 Phylogenetic tree of *GLK* gene family in alfalfa and *A. thaliana*

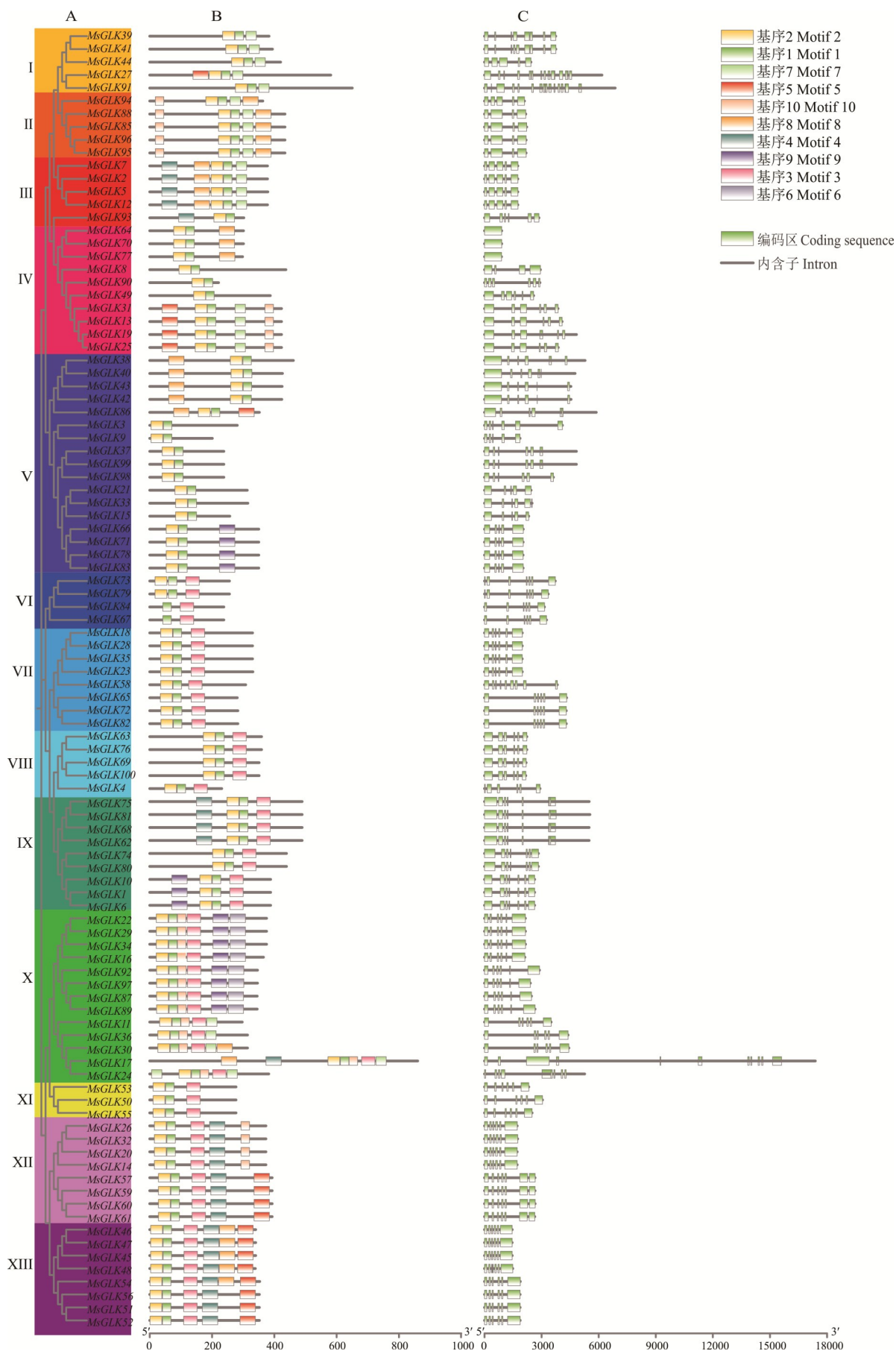
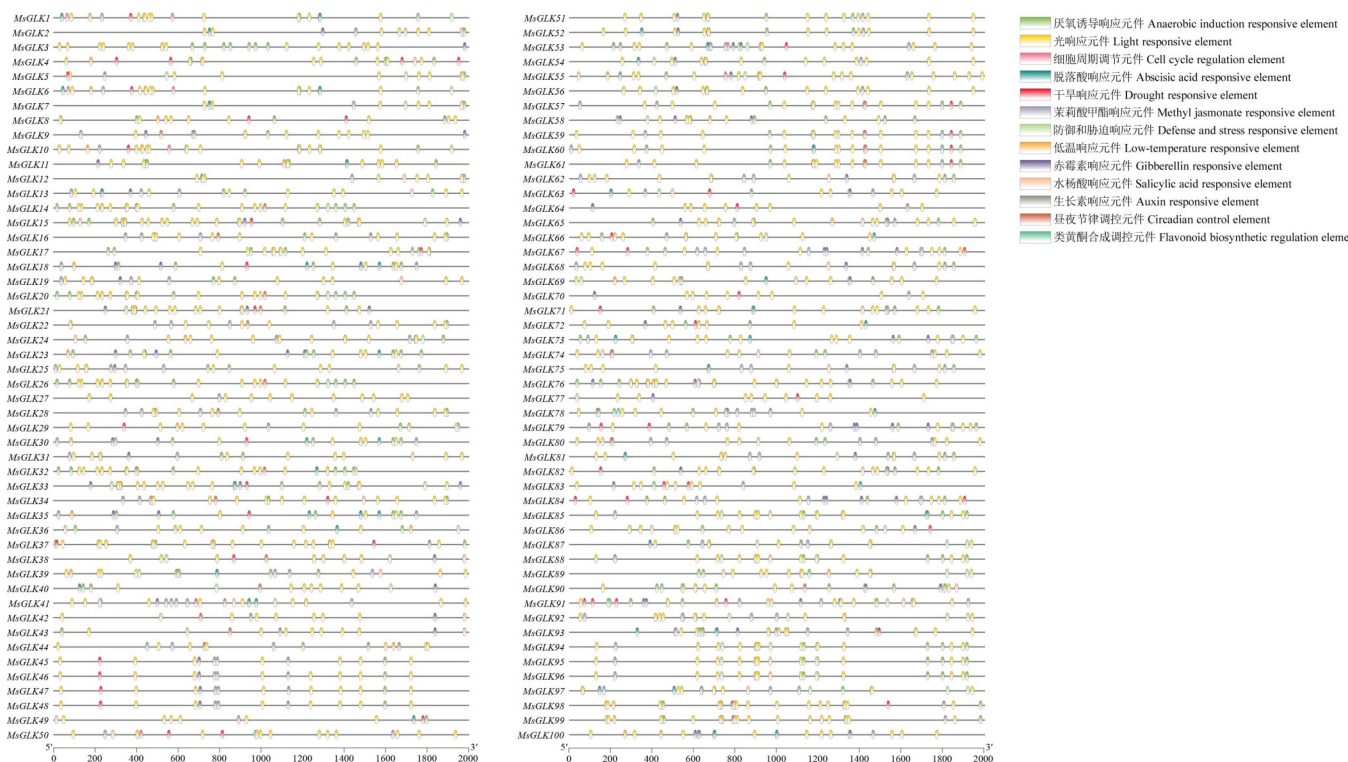


图3 紫花苜蓿 *GLK* 基因的保守基序和外显子-内含子结构

Fig. 3 The conserved motif and exon-intron structure of *GLK* genes in alfalfa

图4 紫花苜蓿 *GLK* 基因家族顺式作用元件分析Fig. 4 Analysis of *cis*-acting elements of alfalfa *GLK* gene family

2.6 紫花苜蓿 *GLK* 基因家族成员的组织表达模式分析

为了解 *MsGLKs* 基因在不同组织中的表达模式,通过公共数据库的转录组数据分析,获得了紫花苜蓿 6 个不同器官,包括:花、叶、根、根瘤、短茎、伸长茎的表达量数据。100 个 *MsGLKs* 基因中,有 12 个未在上述 6 种组织中表达,推测这些基因可能在其他组织中表达,或在某些胁迫条件下特异性表达(图 6)。其余 88 个基因在不同的组织中具有不同的表达丰度,部分基因的表达呈组织特异性。12 个 *MsGLKs* 基因仅在 1 个组织中表达,其中 8 个基因(如 *MsGLK15* 和 *MsGLK21*)仅在根中表达。5 个基因在 2 种组织中表达,如 *MsGLK51* 只在根和短茎中表达;18 个基因在 3 种组织中表达,如 *MsGLK31* 在叶、短茎和伸长茎中表达;14 个基因在 4 种组织中表达,如 *MsGLK7* 在根、伸长茎、花和根瘤中表达;8 个基因在 5 种组织中表达,如 *MsGLK13*、*MsGLK19* 和 *MsGLK25* 均在除根以外的组织中表达,且在叶中的表达量明显高于其他组织。在 6 种组织中均有表达的基因数量最多,为 25 个;其中部分基因的组织表达差异明显,如 *MsGLK74* 在叶中的表达量显著高于其他组织。这些结果显示,不同 *MsGLKs* 基因在不同部位和不同时期参与紫花苜蓿生长发育的调控,并发挥着不同的功能。

通过分析紫花苜蓿 RNA-Seq 数据,得到 *MsGLKs* 在外源 ABA、甘露醇、盐胁迫处理下的表达模式(图 7)。在 100 个基因中,ABA 处理下表达水平发生显著变化的有 31 个,响应甘露醇和盐胁迫的分别有 16 和 14 个。*MsGLK2*、*MsGLK5*、*MsGLK7*、*MsGLK29*、*MsGLK36*、*MsGLK62*、*MsGLK65*、*MsGLK68*、*MsGLK75* 和 *MsGLK77* 在 3 种处理下表达丰度均有变化。外源 ABA 处理下,观察到 *MsGLKs* 的表达水平形成两个不同的簇:一组中 6 个基因表达量先下调后上调,8 个基因表达量先上调后下调;另一组 10 个基因表达量先上调后下调,6 个基因的表达水平在 3~12 h 急剧升高。甘露醇处理下,7 个基因表达量波动后持续降低;4 个基因表达量持续升高,12 h 达到最高;5 个基因在 1~12 h 表达量剧增。在盐胁迫下,*MsGLK* 基因的表达水平也聚为 2 类,一个聚类的 8 个基因中 3 个于 3 h 表达量达到峰值,其余在 24 h 达到峰值;另一聚类 6 个基因在不同时间点表达量降低。

2.7 紫花苜蓿部分 *GLK* 基因在渗透胁迫和外源 ABA 处理下的表达模式分析

为探究 *MsGLK* 基因对干旱、盐胁迫和外源 ABA 的响应情况,根据组织表达量、启动子顺式元件分布和非生

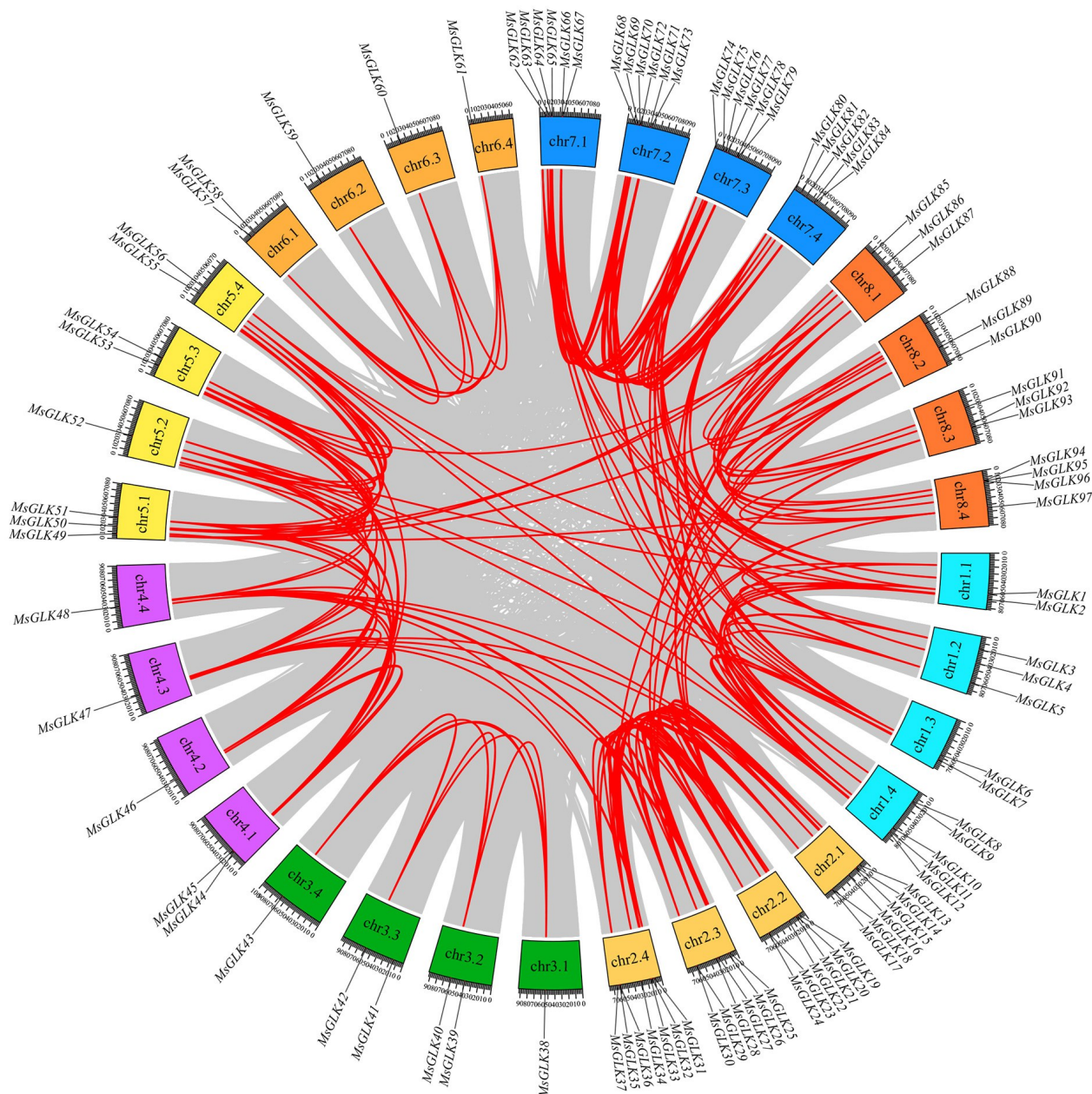


图5 紫花苜蓿 *GLK* 基因共线性分析

Fig. 5 Collinearity analysis of *GLK* gene in alfalfa

物胁迫下的转录组学分析结果,选取9个 *MsGLK* 基因进行了 RT-qPCR 表达分析(图8)。结果表明,外源 ABA 处理下, *MsGLK10*、*MsGLK17*、*MsGLK29*、*MsGLK33*、*MsGLK62* 和 *MsGLK75* 表达水平随处理时间增加呈先增高后降低的趋势, *MsGLK62* 和 *MsGLK75* 在 12 h 达到峰值;与 CK 相比,其余基因均在 3 h 表达显著上调,其中 *MsGLK17* 上调最明显,表达量是 CK 的 6.5 倍 ($P < 0.05$)。 *MsGLK36* 从 0~24 h 下调表达,在 48 h 显著升高,表达量是 CK 的 6.8 倍 ($P < 0.05$); *MsGLK83* 在 1 h 上调,3~48 h 下调表达。 *MsGLK65* 在 1 h 上调,随后表达水平降低,仅在 12 h 呈上调表达。

在干旱胁迫下,与 CK 相比, *MsGLK10*、*MsGLK33*、*MsGLK65* 和 *MsGLK83* 在处理 1 h 时相对 CK 表达量显著上调且达到峰值,随后降低。 *MsGLK36* 在处理 1 h 后表达量持续下调,在 24 h 开始升高,直至 48 h 时表达量达到最高。该变化趋势与图 7 中结果存在一定差异,推测与试验样本来源和处理方式不同有关。 *MsGLK62* 和 *MsGLK75* 在处理 1 h 后表达量逐渐上升,并在 24 h 时达到最高水平。

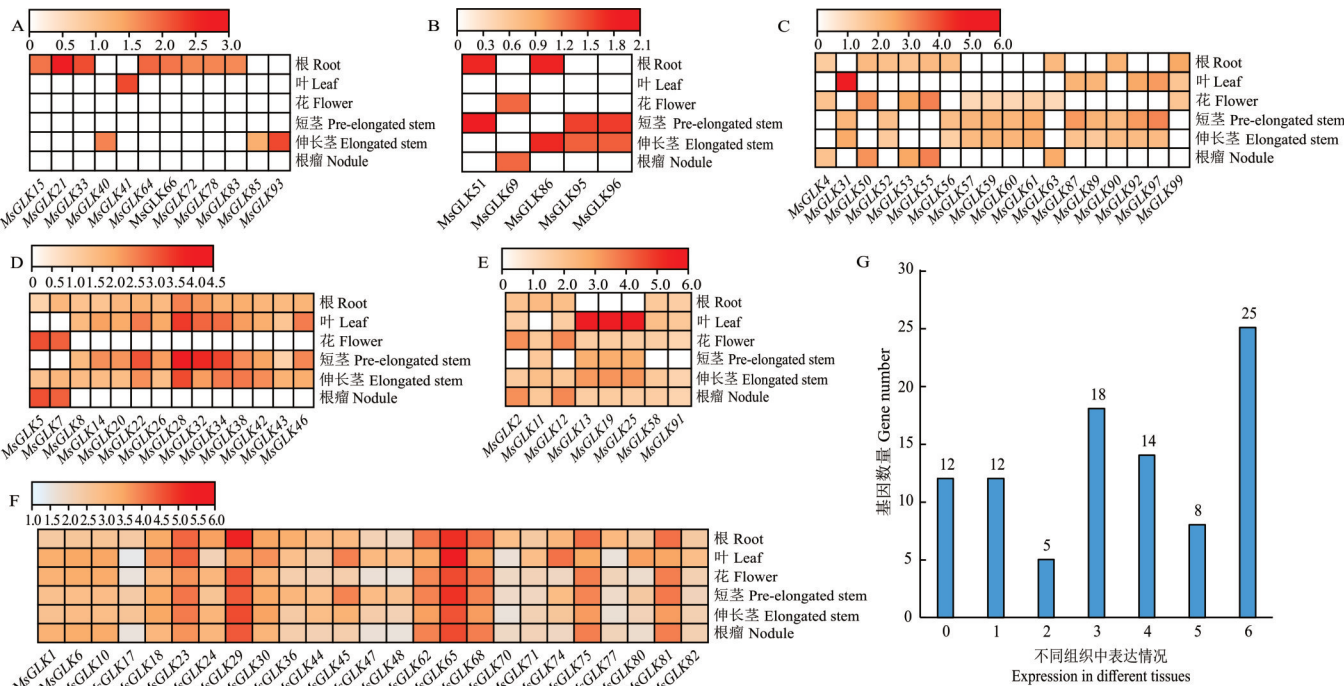


图 6 紫花苜蓿 *GLK* 基因在不同组织中的表达分析

Fig. 6 Analysis of *GLK* gene expression in different tissues of alfalfa

A~F: 在 1~6 种组织中 *MsGLK* 基因的表达 The expression of *MsGLK* gene in 1—6 tissues. 热图上方的色阶表示相对表达水平,从白色到红色的颜色渐变表示表达水平的增加 The color scale above the heatmap indicates the relative expression levels, and the color gradient from white to red indicates an increase in expression levels. G: 在 0~6 种组织中表达的 *MsGLK* 基因数量 Number of *MsGLK* genes expressed in 0—6 tissues. 下同 The same below.

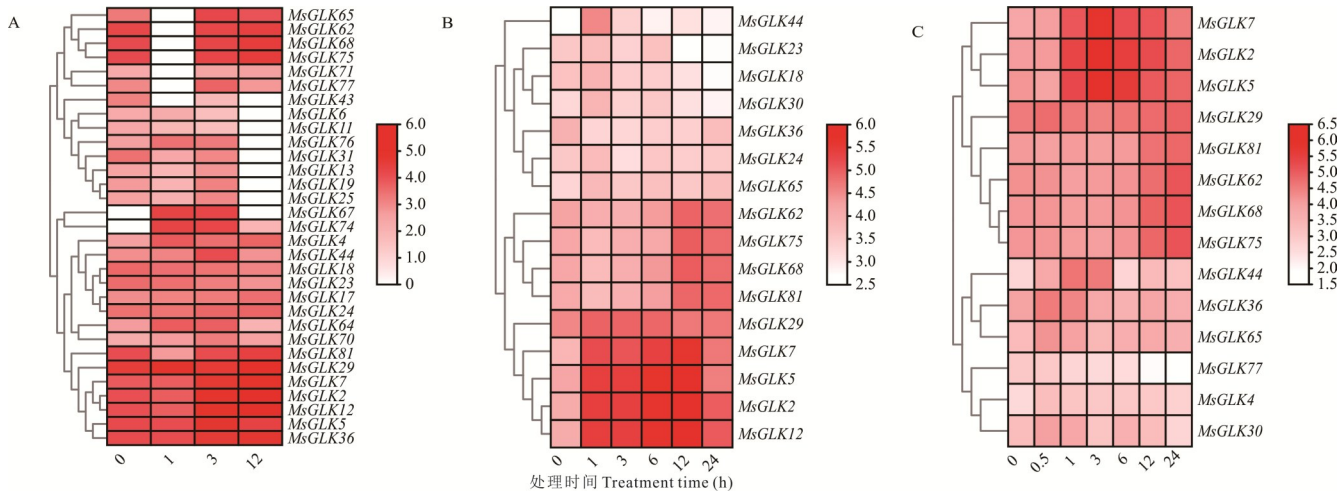


图 7 紫花苜蓿 *GLK* 基因在渗透胁迫下的表达分析

Fig. 7 Expression analysis of *GLK* genes in alfalfa under osmotic stress

A: $10 \mu\text{mol} \cdot \text{L}^{-1}$ 脱落酸 ABA; B: $400 \text{ mmol} \cdot \text{L}^{-1}$ 甘露醇 Mannitol; C: $250 \text{ mmol} \cdot \text{L}^{-1}$ 盐处理 NaCl treatment.

与 CK 相比,在盐胁迫下, *MsGLK10*、*MsGLK33*、*MsGLK65* 和 *MsGLK83* 的表达水平在 1 h 显著上调 ($P < 0.05$),除 *MsGLK33* 外随时间延长呈下降—升高,随后下降再升高趋势,且达到峰值的时间各不相同。其中 *MsGLK33* 对于盐胁迫的响应更为强烈,在所有时期均显著上调 ($P < 0.05$)。 *MsGLK29* 和 *MsGLK36* 在处理 1 h 后表达量下调,而后随时间增加逐渐升高,在 48 h 达到最高水平,呈显著上调表达趋势 ($P < 0.05$)。 *MsGLK62* 和 *MsGLK75* 随处理时间增长表达水平先升高再降低,在 24 h 时达到峰值,与转录组数据分析结果相似。

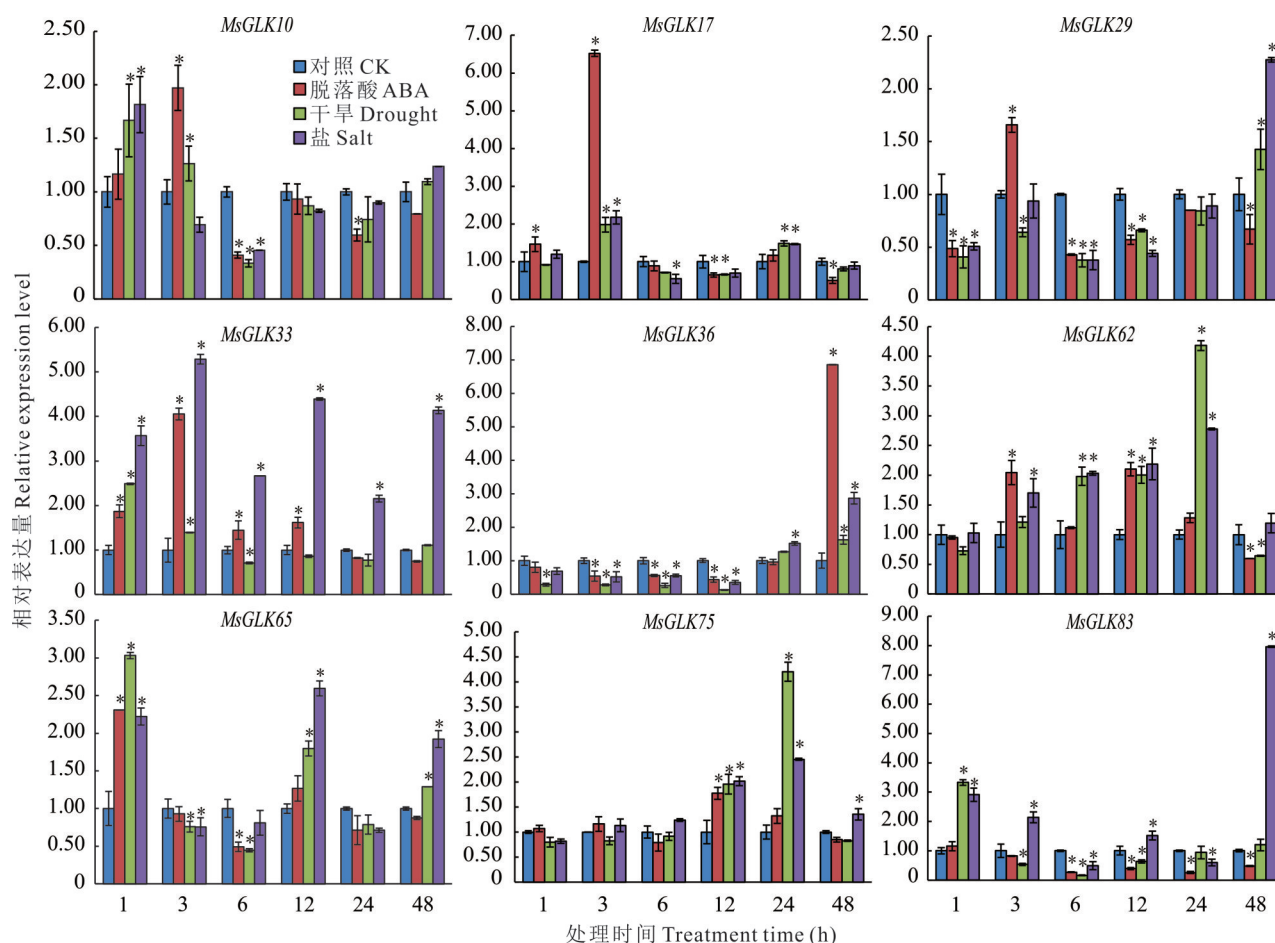


图8 *MsGLKs* 基因家族成员在外源ABA和渗透胁迫下的表达模式分析

Fig. 8 Expression patterns of *MsGLKs* gene family members under exogenous ABA and osmotic stress

*: 同一时间各胁迫处理与对照差异显著 ($P < 0.05$) At the same time, there were significant differences among stress treatment and CK ($P < 0.05$).

3 讨论

GLK 是一类植物特异性转录因子,主要参与植物叶绿体形成发育、叶绿素生物合成的正向调控,同时也被证实能够提高作物产量和果实品质,提高植物对于生物和非生物胁迫的抗性^[15]。已有研究表明,不同植物的 *GLK* 基因家族成员数量存在较大差异,例如,在甜橙(*Citrus sinensis*)中鉴定到 27 个 *GLK* 成员^[40],玉米中鉴定到 59 个成员^[22],小麦中鉴定到 87 个^[27],烟草中 89 个^[31],大豆 130 个^[30],陆地棉 146 个^[32]。本研究在同源四倍体“新疆大叶”紫花苜蓿基因组中鉴定到 100 个 *MsGLK* 基因,少于同源四倍体大豆的 130 个和异源四倍体陆地棉的 146 个,但多于异源四倍体烟草的 89 个。据观察,本研究使用的紫花苜蓿参考基因组大小为 3.15 Gb,大于陆地棉(2.3 Gb)和大豆(978.4 Mb),小于烟草基因组(4.5 Gb)。据此可以认为 *GLK* 家族成员的数量与基因组大小基本没有相关性。不同物种中 *GLK* 数量的多样性可能与基因在进化过程中特异性的重复与丢失有关。有研究指出,谱系特异性扩展(lineage-specific expansion, LSE)是导致物种间基因家族数量差异较大的因素之一^[42]。

多基因家族通常由基因复制产生^[43]。早期研究指出,祖先陆地植物很可能只具有单个 *GLK* 基因,大多数含有多个 *GLK* 基因的物种基因组经历了全基因组复制^[9]。基因复制对于新功能的出现和分化至关重要,*GLK* 基因的复制扩增能够引起其亚功能化和在 C_4 植物二态叶绿体中功能特异性的出现^[9,44]。片段复制和串联复制则是基因家族的进化扩增中的重要动力^[43]。既往大多数物种的 *GLK* 基因家族鉴定表明,片段复制在该家族的扩增中起到了重要作用。小麦 87 个 *GLK* 基因中 74 个基因涉及片段复制事件,6 个基因涉及串联复制事件^[27]。关于大豆的研究中,73.07% 的 *GmGLK* 基因参与片段复制事件,仅有一对参与串联复制事件^[30]。而烟草中观察到 12 对

NtGLK 片段重复基因对,未观察到串联重复基因对^[31]。本研究中,在紫花苜蓿基因组中共发现 88 个 *MsGLK* 基因参与的 192 个片段复制事件,仅有 2 个 *MsGLK* 基因参与一个串联复制事件,与前人研究结果一致。有研究认为,片段复制事件经常发生在进化缓慢的基因组中,推测 *MsGLK* 基因在进化过程中保守性较强^[22]。绝大多数 *MsGLK* 重复基因对的 *Ka/Ks* 小于 1,证明这些基因拷贝经历了纯化选择后,维持其基本功能的结构进一步保留并稳定下来。染色体定位结果显示,*MsGLK* 基因在紫花苜蓿基因组中分布广泛且不均匀,或与进化过程中染色体的缺失、插入、重复、倒置有关^[28]。

系统发育树根据拟南芥和紫花苜蓿 *GLK* 蛋白序列构建,进一步分为 13 个组 (I ~ XIII),多于陆地棉和大豆的分组^[30,32],较多的分组意味着更多的功能分化和更精确的分类。一般而言,同一亚组的蛋白承担着更相似的功能。在 V 组中发现 4 个拟南芥 KANADI 样 (KAN) 基因 (*AtGLK4*、*AtGLK30*、*AtGLK35*、*AtGLK38*),该类基因被发现在拟南芥芽顶端分生组织 (shoot apical meristem, SAM) 表达,能够抑制生长素的生物合成、转运和信号传导,从而抑制器官形态发生^[45]。属于 *NIGT1/HRS1/HHO* 家族的 *NIGT1.1*、*NIGT1.2*、*NIGT1.3* 和 *NIGT1.4* 被观察到聚集在第 III 组中,这些基因被认为是能够整合根尖氮、磷缺乏信号,并抑制初级根的生长以做出响应;部分研究表明,*NIGT1/HRS1/HHO* 转录因子可能通过响应光、温度和激素信号调控植物生长发育,并与植物耐盐性和低温胁迫抗性相关^[46-47]。第 IV 组则包括 *AtLUX* (*AtGLK27*)、*AtBOA* (*AtGLK17*)、*AtMYBC1* (*AtGLK41*),这些基因被认为参与了昼夜节律振荡^[48]。

保守基序和基因结构的差异在一定程度上体现了蛋白质功能的差异性和基因家族的进化趋势。分析 *MsGLKs* 的保守基序发现,基序 1 和基序 2 在除 *MsGLK67* 和 *MsGLK84* (仅含基序 1) 的基因中共享,表明两个基序在维持 *GLK* 基因的基本功能中至关重要。二者对应的 *GLK* 保守基序 Myb-SHLQKYF 能够与位于 C 末端的 I-box 结合,作为转录激活因子发挥作用^[49]。此外,对应 Myb-CC-LHEQLE 基序的基序 3 在大多数亚组中被观察到,该结构可能参与多种胁迫信号的传导^[39]。除此以外,不同亚组具有不同的基序,或许意味着其在非生物胁迫中能起到更多样的作用。真核生物基因结构的进化过程涉及许多内含子的增加和缺失,其中缺失发生的频率更高。许多证据表明,内含子的变异是基因进化的主要因素^[50]。在本研究中,*MsGLK* 基因的内含子数量变化大,在第 I 和 IV 组中表现尤为明显,推测 *GLK* 基因的演化与内含子分布的改变密切相关。此外,研究表明,真核生物基因组中广泛存在的移动遗传元件—长末端重复逆转录转座子 (LTR 逆转录转座子) 能够插入内含子、启动子、下游等基因组区域,进而影响豆科植物基因的结构与功能,或许也对紫花苜蓿 *GLK* 基因的结构有一定影响^[51]。

通过分析紫花苜蓿 *GLK* 基因家族的启动子顺式作用元件,共发现 13 类顺式作用元件。先前的研究指出,光能够诱导拟南芥、玉米、谷子中的 *GLK* 基因的表达^[29,52]。在 *MsGLKs* 中,光反应元件分布最为广泛,数量也最多,其中 G-box 是最重要的光响应元件之一。研究发现,*GLK* 转录因子能够与 G-box 结合因子 (G-box binding factor, GBF) 相互作用,协同发挥作用^[53]。*MsGLK* 基因的启动子还预测到大量生长素 (auxin, IAA)、脱落酸、赤霉素 (gibberellin, GA)、茉莉酸甲酯 (methyl jasmonate, MeJA)、水杨酸 (salicylic acid, SA) 等激素响应元件和干旱、低温、厌氧等胁迫响应元件。其中,ABA 响应元件 (ABRE) 同时也受到干旱胁迫的诱导,而低温响应元件 (LTR) 也与盐胁迫响应相关^[41,54]。此外,还存在少部分与细胞周期、昼夜节律以及类黄酮合成相关的调控元件。基因启动子区域中顺式作用元件的类型、数量和分布的多样性反映了 *MsGLK* 基因反应调控机制具有复杂性。*GLK* 基因对叶绿素的合成和叶绿体发育具有正调控作用。组织表达量分析结果显示,大部分 *MsGLK* 基因在叶片中表现出较高的表达水平。根据 *GLK* 基因控制叶绿素生物合成的功能推测,在叶片中较高的表达水平能够满足 *GLK* 对于相关基因活性及代谢功能的调控作用。此外,观察到 8 个 *MsGLK* 基因 (*MsGLK15*、*MsGLK21*、*MsGLK33*、*MsGLK64*、*MsGLK66*、*MsGLK72*、*MsGLK78*、*MsGLK83*) 在紫花苜蓿根组织中特异性表达,可能与根系的生长发育相关,可以作为相应功能研究的优秀靶基因。

许多研究表明,*GLK* 家族在非生物胁迫和植物激素处理反应中发挥了重要作用^[15, 20-21]。本研究分析了两种渗透胁迫 (干旱、盐) 和一种外源激素 (ABA) 处理下的 9 个 *MsGLK* 基因的表达情况。其中,5 个 *MsGLK* 基因 (*MsGLK10*、*MsGLK17*、*MsGLK33*、*MsGLK65*、*MsGLK83*) 在干旱和盐胁迫的 1~3 h 内显著上调,表明这些基因

能够在短时间的渗透胁迫中起到正向调控作用;4个基因(*MsGLK29*、*MsGLK36*、*MsGLK62*、*MsGLK75*)在胁迫处理24~48 h显著上调,推测它们能够响应更长时间的胁迫。*MsGLK*基因在胁迫的不同阶段响应情况不尽相同,可能更有利于植物适应长期逆境环境。脱落酸(ABA)在植物针对不利环境因子的适应性反应中起着重要作用,被认为是介导植物生长和逆境反应的重要调节剂^[55]。研究发现,ABA信号通路在渗透胁迫引起的植物生理反应和应激基因的诱导中十分关键^[56]。已有研究显示,GLK1/2-WRKY40转录模块能够在ABA反应中起负调控作用,影响拟南芥幼苗的渗透胁迫敏感性^[24]。被检测的基因中,大多数在外源ABA诱导下,呈现出与两种渗透胁迫相似的变化趋势,表明这些*MsGLK*基因在调控植株渗透胁迫抗性过程中可能依赖ABA的信号传导途径,该观点可以进一步通过试验进行验证。所分析的9个基因中,*MsGLK33*和*MsGLK75*分别在盐和干旱处理下变化幅度最高,响应最强烈,可能在紫花苜蓿耐旱抗盐方面具有较好潜力。上述结果表明,*MsGLK*基因在紫花苜蓿应对渗透胁迫方面起着重要作用,具体的生物学功能及调控机制有待进一步研究和验证。

4 结论

本研究从四倍体“新疆大叶”紫花苜蓿参考基因组中鉴定到100条*MsGLKs*基因,97个不均匀分布在32条染色体上,其余3个未定位在染色体上。*MsGLKs*蛋白相对较小,序列长度为201~860个氨基酸,蛋白质分子量为23.51~94.37 kDa。根据系统发育关系、保守基序及基因结构,可将*MsGLKs*蛋白分为13组,所有成员均含有GLK特异性基序Myb-SHLQKYF。*MsGLKs*成员含有参与植物生长发育、激素响应和胁迫反应相关的顺式作用元件。共线性分析表明,片段重复在该家族的扩增中起到了重要作用。转录组数据揭示了不同*MsGLKs*基因可能在紫花苜蓿生长发育的过程中发挥着不同的功能。RT-qPCR结果表明9个基因均能参与响应干旱和盐胁迫,并且该响应过程可能与ABA信号传导相关。本研究结果可为进一步探究*MsGLKs*基因功能提供理论依据,以期对紫花苜蓿抗旱耐盐种质的改良和创新提供新的候选基因。

参考文献 References:

- [1] Zaikina E A, Rumyantsev S D, Sarvarova E R, *et al.* Transcription factor genes involved in plant response to abiotic stress factors. *Ecological Genetics*, 2019, 17(3): 47–58.
- [2] Joshi R, Wani S H, Singh B, *et al.* Transcription factors and plants response to drought stress: Current understanding and future directions. *Frontiers in Plant Science*, 2016, 7: 1029.
- [3] Riechmann J L, Heard J, Martin G, *et al.* *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science*, 2000, 290(5499): 2105–2110.
- [4] Yasumura Y, Moylan E C, Langdale J A. A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *The Plant Cell*, 2005, 17(7): 1894–1907.
- [5] Rossini L, Cribb L, Martin D J, *et al.* The maize *Golden2* gene defines a novel class of transcriptional regulators in plants. *The Plant Cell*, 2001, 13(5): 1231–1244.
- [6] Tachibana R, Abe S, Marugami M, *et al.* BPG4 regulates chloroplast development and homeostasis by suppressing GLK transcription factors and involving light and brassinosteroid signaling. *Nature Communications*, 2024, 15(1): 370.
- [7] Waters M T, Wang P, Korkaric M, *et al.* GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. *The Plant Cell*, 2009, 21(4): 1109–1128.
- [8] Cackett L, Luginbuehl L H, Schreier T B, *et al.* Chloroplast development in green plant tissues: The interplay between light, hormone, and transcriptional regulation. *New Phytologist*, 2022, 233(5): 2000–2016.
- [9] Wang P, Fouracre J, Kelly S, *et al.* Evolution of *GOLDEN2-LIKE* gene function in C₃ and C₄ plants. *Planta*, 2013, 237(2): 481–495.
- [10] Liu X Y, Chen X, Cao J J, *et al.* GOLDEN 2-LIKE transcription factors regulate chlorophyll biosynthesis and flavonoid accumulation in response to UV-B in tea plants. *Horticultural Plant Journal*, 2023, 9(5): 1055–1066.
- [11] Tu X Y, Ren S B, Shen W, *et al.* Limited conservation in cross-species comparison of GLK transcription factor binding suggested wide-spread cisome divergence. *Nature Communications*, 2022, 13(1): 7632.
- [12] Powell A L, Nguyen C V, Hill T, *et al.* *Uniform ripening* encodes a *Golden 2-like* transcription factor regulating tomato fruit

- chloroplast development. *Science*, 2012, 336(6089): 1711–1715.
- [13] Li H, Li Y X, Deng H, *et al.* Tomato UV-B receptor SIUVR8 mediates plant acclimation to UV-B radiation and enhances fruit chloroplast development via regulating SIGLK2. *Scientific Reports*, 2018, 8(1): 6097.
- [14] Zubo Y O, Blakley I C, Franco-Zorrilla J M, *et al.* Coordination of chloroplast development through the action of the GNC and GLK transcription factor families. *Plant Physiology*, 2018, 178(1): 130–147.
- [15] Kim S, Choi H, Yi T, *et al.* Pleiotropic properties of GOLDEN2-LIKE transcription factors for crop improvement. *Applied Biological Chemistry*, 2023, 66(1): 1–14.
- [16] Singh J, Savitch L, Subramaniam G, *et al.* The GLK1 ‘regulon’ encodes disease defense related proteins and confers resistance to *Fusarium graminearum* in *Arabidopsis*. *Biochemical and Biophysical Research Communications*, 2007, 359(2): 234–238.
- [17] Han X Y, Li P X, Zou L J, *et al.* GOLDEN 2-LIKE transcription factors coordinate the tolerance to cucumber mosaic virus in *Arabidopsis*. *Biochemical and Biophysical Research Communications*, 2016, 477(4): 626–632.
- [18] Ali N, Chen H, Zhang C, *et al.* Ectopic expression of *AhGLK1b* (GOLDEN2-like transcription factor) in *Arabidopsis* confers dual resistance to fungal and bacterial pathogens. *Genes*, 2020, 11(3): 343.
- [19] Townsend P D, Dixon C H, Sloatweg E J, *et al.* The intracellular immune receptor Rx1 regulates the DNA-binding activity of a Golden2-like transcription factor. *Journal of Biological Chemistry*, 2018, 293(9): 3218–3233.
- [20] Liu J N, Mehari T G, Xu Y C, *et al.* *GhGLK1* a key candidate gene from GARP family enhances cold and drought stress tolerance in cotton. *Frontiers in Plant Science*, 2021, 12: 759312.
- [21] Liu X, Li L M, Li M J, *et al.* *AhGLK1* affects chlorophyll biosynthesis and photosynthesis in peanut leaves during recovery from drought. *Scientific Reports*, 2018, 8(1): 2250.
- [22] Liu F, Xu Y X, Han G M, *et al.* Molecular evolution and genetic variation of *G2-like* transcription factor genes in maize. *PLoS One*, 2016, 11(8): e0161763.
- [23] Li T T. Expression analysis of *PPO*, *GLK1* and *NADH* in lotus and identification of its interaction domain. Yangling: Northwest A & F University, 2021.
李婷婷. 莲藕 *PPO* 与 *GLK1*、*NADH* 的表达分析及互作结构域确定. 杨凌: 西北农林科技大学, 2021.
- [24] Ahmad R, Liu Y T, Wang T J, *et al.* GOLDEN2-LIKE transcription factors regulate *WRKY40* expression in response to abscisic acid. *Plant Physiology*, 2019, 179(4): 1844–1860.
- [25] Chen L, He F, Long R C, *et al.* A global alfalfa diversity panel reveals genomic selection signatures in Chinese varieties and genomic associations with root development. *Journal of Integrative Plant Biology*, 2021, 63(11): 1937–1951.
- [26] Huang X F, Shi P L, Yu C Q, *et al.* A review of abiotic stress resistance of forages. *Acta Agrestia Sinica*, 2023, 31(5): 1293–1301.
黄小芳, 石培礼, 余成群, 等. 非生物胁迫下牧草抗逆性研究进展. *草地学报*, 2023, 31(5): 1293–1301.
- [27] Jia Y K, Gao H H, Feng J C, *et al.* Genome-wide identification and expression analysis of *G2-like* transcription factors family gene in wheat. *Acta Agronomica Sinica*, 2023, 49(5): 1410–1425.
贾玉库, 高宏欢, 冯健超, 等. 小麦 *G2-like* 转录因子家族基因鉴定与表达模式分析. *作物学报*, 2023, 49(5): 1410–1425.
- [28] Wang Z Y, Zhao S, Liu J F, *et al.* Genome-wide identification of tomato Golden 2-like transcription factors and abiotic stress related members screening. *BMC Plant Biology*, 2022, 22(1): 82.
- [29] Chen H F, Qin L, Tian J G, *et al.* Identification and evolutionary analysis of the *GOLDEN 2-LIKE* gene family in foxtail millet. *Tropical Plant Biology*, 2022, 15: 301–318.
- [30] Alam I, Manghwar H, Zhang H Y, *et al.* Identification of GOLDEN2-like transcription factor genes in soybeans and their role in regulating plant development and metal ion stresses. *Frontiers in Plant Science*, 2022, 13: 1052659.
- [31] Qin M Y, Zhang B H, Gu G, *et al.* Genome-wide analysis of the *G2-like* transcription factor genes and their expression in different senescence stages of tobacco (*Nicotiana tabacum* L.). *Frontiers in Genetics*, 2021, 31(12): 626352.
- [32] Zhao Z L, Shuang J R, Li Z G, *et al.* Identification of the Golden-2-like transcription factors gene family in *Gossypium hirsutum*. *PeerJ*, 2021, 9: e12484.
- [33] Chen C J, Chen H, Zhang Y, *et al.* TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, 2020, 13(8): 1194–1202.
- [34] Tamura K, Stecher G, Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, 2021, 38(7): 3022–3027.
- [35] O’Rourke J A, Fu F L, Bucciarelli B, *et al.* The *Medicago sativa* gene index 1.2: A web-accessible gene expression atlas for

- investigating expression differences between *Medicago sativa* subspecies. BMC Genomics, 2015, 16(1): 502.
- [36] Liu H, Li X Y, Zi Y F, *et al.* Characterization of the heat shock transcription factor family in *Medicago sativa* L. and its potential roles in response to abiotic stresses. International Journal of Molecular Sciences, 2023, 24(16): 12683.
- [37] Kang C H, Jung W Y, Kang Y H, *et al.* AtBAG6, a novel calmodulin-binding protein, induces programmed cell death in yeast and plants. Cell Death and Differentiation, 2006, 13(1): 84–95.
- [38] Wu R H, Guo L, Guo Y Y, *et al.* The *G2-Like* gene family in *Populus trichocarpa*: Identification, evolution and expression profiles. BMC Genomic Data, 2023, 24(1): 37.
- [39] Meng X X, Liang Z K, Dai X R, *et al.* Predicting transcriptional responses to cold stress across plant species. PNAS, 2021, 118(10): e2026330118.
- [40] Xiong B, Gong Y, Li Q, *et al.* Genome-wide analysis of the *GLK* gene family and the expression under different growth stages and dark stress in sweet orange (*Citrus sinensis*). Horticulturae, 2022, 8(11): 1076.
- [41] Zhang X Y, Zhao L J, Li Y J, *et al.* Expression of *AtPUB18* after salt stress treatment and analysis of its promoter from *Arabidopsis thaliana*. Acta Botanica Boreali-Occidentalia Sinica, 2014, 34(1): 54–59.
张新宇, 赵兰杰, 李艳军, 等. 盐胁迫对拟南芥 *AtPUB18* 基因的诱导表达及其启动子分析. 西北植物学报, 2014, 34(1): 54–59.
- [42] Panchy N, Lehti-Shiu M, Shiu S H. Evolution of gene duplication in plants. Plant Physiology, 2016, 171(4): 2294–2316.
- [43] Cannon S B, Mitra A, Baumgarten A, *et al.* The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biology, 2004, 4: 10.
- [44] Birchler J A, Yang H. The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. The Plant Cell, 2022, 34(7): 2466–2474.
- [45] Ram H, Sahadevan S, Gale N, *et al.* An integrated analysis of cell-type specific gene expression reveals genes regulated by REVOLUTA and KANADI1 in the Arabidopsis shoot apical meristem. PLoS Genetics, 2020, 16(4): e1008661.
- [46] Kiba T, Inaba J, Kudo T, *et al.* Repression of nitrogen starvation responses by members of the Arabidopsis GARP-type transcription factor NIGT1/HRS1 subfamily. The Plant Cell, 2018, 30(4): 925–945.
- [47] Li Q, Zhou L Y, Li Y H, *et al.* Plant NIGT1/HRS1/HHO transcription factors: Key regulators with multiple roles in plant growth, development, and stress responses. International Journal of Molecular Sciences, 2021, 22(16): 8685.
- [48] Zhao X Y, Yang J J, Li X Z, *et al.* Identification and expression analysis of GARP superfamily genes in response to nitrogen and phosphorus stress in *Spirodela polyrhiza*. BMC Plant Biology, 2022, 22(1): 308.
- [49] Rose A, Meier I, Wienand U. The tomato I-box binding factor LeMYBI is a member of a novel class of myb-like proteins. The Plant Journal, 1999, 20(6): 641–652.
- [50] Rogozin I B, Carmel L, Csuros M, *et al.* Origin and evolution of spliceosomal introns. Biology Direct, 2012, 7: 11.
- [51] Yang L L, Zhang X Y, Wang L Y, *et al.* Lineage-specific amplification and epigenetic regulation of LTR-retrotransposons contribute to the structure, evolution, and function of Fabaceae species. BMC Genomics, 2023, 24(1): 423.
- [52] Fitter D W, Martin D J, Copley M J, *et al.* GLK gene pairs regulate chloroplast development in diverse plant species. The Plant Journal, 2010, 31(6): 713–727.
- [53] Tamai H, Iwabuchi M, Meshi T. Arabidopsis GARP transcriptional activators interact with the pro-rich activation domain shared by G-Box-Binding bZIP factors. Plant and Cell Physiology, 2002, 43(1): 99–107.
- [54] Lee S C, Kim S H, Kim S R. Drought inducible *OsDhn1* promoter is activated by OsDREB1A and OsDREB1D. Journal of Integrative Plant Biology, 2013, 56: 115–121.
- [55] Wang J L, Du C X, Zhou H K, *et al.* Exogenic ABA and plant abiotic stress resistance mechanism. Journal of Anhui Agricultural Sciences, 2019, 47(13): 12–15.
王金丽, 杜晨曦, 周华坤, 等. 外源 ABA 与植物非生物胁迫抗逆机制. 安徽农业科学, 2019, 47(13): 12–15.
- [56] Yoshida T, Fernie A R, Shinozaki K, *et al.* Long-distance stress and developmental signals associated with abscisic acid signaling in environmental responses. The Plant Journal: For Cell and Molecular Biology, 2021, 105(2): 477–488.