

· 研究报告 ·

棉属光敏色素PHY基因家族的全基因组鉴定与驯化选择分析

顾家琦^{1,2†}, 朱福慧^{1,2†}, 谢沛豪^{1,2}, 孟庆营^{1,2}, 郑颖², 张献龙^{1,2}, 袁道军^{1,2,3*}

¹华中农业大学作物遗传改良全国重点实验室, 武汉 430070; ²华中农业大学植物科学技术学院, 武汉 430070

³新疆农业大学农学院, 乌鲁木齐 830052

摘要 光敏色素(PHY)作为植物体内重要的红光(R)和远红光(FR)受体, 在调节植物花期、提高作物产量以及调控植物抗逆性方面均具有重要作用。鉴定棉花PHY基因, 探究其驯化和改良的遗传调控网络, 发掘棉花光敏色素关键基因, 为早熟棉花品种的从头驯化和育种提供新见解。以5个拟南芥(*Arabidopsis thaliana*)光敏色素基因作为种子序列, 利用生物信息学方法, 对不同棉种中的PHY基因进行全基因组系统鉴定。系统发育分析表明, 锦葵科植物中的PHY基因包括PHYA、PHYB、PHYC与PHYE四个亚家族。PHY基因在陆地棉(*Gossypium hirsutum*)不同群体间的驯化选择分析结果表明, PHY基因的驯化过程可分为早期驯化和后期遗传改良2个阶段。对陆地棉野生种和栽培品种在长、短日照下的转录组分析显示, GhPHYA1Dt和GhPHYB1Dt在长、短日照下的表达差异显著; 在长日照处理14小时后, 栽培品种GhPHYC1At和GhPHYE1At的表达量显著低于野生种。研究结果为进一步揭示棉花PHY基因的驯化选择和功能机制奠定了基础, 为棉花早熟新品种选育和从头驯化提供了理论依据。

关键词 棉花, 驯化, 光敏色素, 基因家族

顾家琦, 朱福慧, 谢沛豪, 孟庆营, 郑颖, 张献龙, 袁道军 (2024). 棉属光敏色素PHY基因家族的全基因组鉴定与驯化选择分析. 植物学报 59, 34–53.

棉花属于锦葵科(Malvaceae)棉属(*Gossypium*), 包括45个二倍体($2n=2x=26$)和7个异源四倍体($2n=4x=52$)棉种, 异源四倍体棉种是由A和D基因组祖先棉种在约 1.5×10^6 年前经历天然杂交和染色体加倍后形成(Wendel and Cronn, 2003; Wendel et al., 2009)。不同棉种的植株形态、叶片形状以及纤维特征等差异较大, 四倍体棉种因具有高产、优质和适应性强等优势而逐渐代替二倍体棉种并成为主要的种植类型。目前, 世界上广泛栽培的棉种多为异源四倍体(陆地棉AD₁ (*G. hirsutum*)和海岛棉AD₂ (*G. barbadense*)), 其中陆地棉(AD₁)的产量和海岛棉(AD₂)的品质是其最主要的栽培优势(Wendel and Cronn, 2003; Wendel et al., 2009)。

作物驯化是复杂的进化过程。在这一过程中, 人

类对植物的性状选择导致植物在生理和形态上均发生变化, 使栽培品种与其野生祖先产生明显差异(Purugganan and Fuller, 2009)。在驯化过程中, 异源四倍体栽培棉种陆地棉(AD₁)和海岛棉(AD₂)的纤维品质均发生了显著变化。野生或半野生陆地棉(AD₁)和海岛棉(AD₂)的纤维表现出浅棕色、纤维长度较短(< 1 cm)以及紧密附着的性状, 经过长期驯化后其纤维转变为白色, 纤维长度增加(> 3 cm)且更易去除(Ma et al., 2018), 能够极大地满足纺织工业的需求(Fang et al., 2017b; Yang et al., 2020)。长期的驯化选择不仅使棉纤维性状发生了巨大改变, 同时也使栽培棉丧失了对光周期的敏感性, 使其从典型的短日照植物转变为日中性植物(Bao et al., 2019; Tian and Zhang, 2021), 从而可在全球广泛种植。

收稿日期: 2023-01-02; 接受日期: 2023-07-08

基金项目: 国家重点研发计划(No.2021YFF1000102)和华中农业大学中央高校基本科研项目(No.2021ZKPY013)

† 共同第一作者

* 通讯作者。E-mail: robert@mail.hzau.edu.cn

光是影响植物生长发育的重要环境因素之一,它不仅为光合作用提供能量,而且为其提供周围环境信息(Ballaré and Pierik, 2017)。光周期是一种约24小时的生物时钟节律,即植物感受昼夜变化,并随之调节自身生命活动,进而实现开花诱导、叶片运动和气孔开闭等生长发育变化(房迈莼等, 2005; 连欣, 2020)。植物通过光受体感知光的持续时间、强度、方向和光质(Briggs and Olney, 2001)。目前已知的光受体分为光敏色素、隐花色素和向光素等(魏俊梅, 2018)。光敏色素是植物生长发育过程中最重要的光受体之一,其通过介导光信号途径调节植物生长发育,在种子萌发、昼夜节律、开花时间和光形态建成等一系列发育过程中发挥重要作用(Chen and Chory, 2011; Li et al., 2011; Rao et al., 2011)。为了解光周期调控开花的分子基础以及光对棉花发育的影响,有必要对棉花光敏色素基因进行全基因组分析。

光敏色素在感知红光(R)和远红光(FR)方面具有重要作用,其也能感知蓝光(B)和紫外光(UV)(Chun et al., 2001)。PHYA介导远红光信号下幼苗的光形态建成,PHYB和PHYC介导红光信号下幼苗的光形态建成(李建平等, 2013; 李建平等, 2014; 岳晶等, 2015; 赵翔等, 2015; Peng et al., 2020)。研究表明,玉米(*Zea mays*)中PHYB基因与PIF结合,通过抑制miR156的表达激发避荫反应(Xie et al., 2017);而PHYC基因负调控玉米的避荫反应(丁武思等, 2021)。在调控开花方面,同一物种内不同PHY亚家族基因存在差异。在拟南芥(*Arabidopsis thaliana*)和菊花(*Dendranthema morifolium*)中,过表达PHYA导致花期提前(Johnson et al., 1994; 惠婕等, 2011)。研究表明,在拟南芥中PHYB抑制开花(Reed et al., 1993),而缺失该基因导致早花现象(詹克慧等, 2012)。不同物种内相同亚家族的基因其功能也有差异,在小麦(*Triticum aestivum*)中,PHYB及PHYC基因通过光周期途径诱导开花(Pearce et al., 2016)。

棉花是我国重要的经济作物之一,其主要产物棉纤维是纺织工业的重要原料(Abdellatif et al., 2012)。我国是世界上主要的棉花生产国和消费国,为使我国棉花产业在国际市场具有更强的竞争力,培育高产优质多抗的棉花新品种成为当前主要的育种目标(Zhang et al., 2011)。开花是从营养生长到生殖生长的关键转折点,提早开花可缩短生育期,有利于种植茬口的合

理安排,减少与粮争地。因此,本研究在全基因组水平系统鉴定了锦葵科物种的光敏色素PHY基因,对PHY基因的结构、保守基序、染色体定位、共线性、选择压力以及表达水平等进行生物信息学分析。在此基础上,结合陆地棉(AD₁)重测序数据对其驯化选择进行分析,利用陆地棉(AD₁)野生型和栽培品种在长、短日照下的转录组数据分析基因表达模式,以期获得驯化选择的关键PHY基因,为棉属PHY基因的功能鉴定及应用奠定理论基础,也为棉种遗传改良提供优质的基因资源。

1 材料与方法

1.1 数据来源

收集二倍体棉种草棉(*Gossypium herbaceum*, A₁)(Huang et al., 2020)、亚洲棉(*G. arboreum*, A₂)(Huang et al., 2020)、特纳氏棉(*G. turneri*, D₁)(Grover et al., 2019)、辣根棉(*G. armourianum*, D₂₋₁)(Grover et al., 2019)、戴维逊棉(*G. davidsonii*, D_{3d})(Yang et al., 2021)、克劳茨基棉(*G. klotzschianum*, D_{3-k})(Grover et al., 2019)、雷蒙德氏棉(*G. raimondii*, D₅)(Wang et al., 2021)、三裂棉(*G. trilobum*, D₈)(Grover et al., 2019)、特纳氏棉(*G. turneri*, D₁₀)(Udall et al., 2019a)、长萼棉(*G. longicalyx*, F₁)(Grover et al., 2020)和澳洲棉(*G. australe*, G₂)(Cai et al., 2020),四倍体棉种陆地棉(*G. hirsutum*, AD₁)(Saski et al., 2017)、海岛棉(*G. barbadense*, AD₂)(Wang et al., 2019)、毛棉(*G. tomentosum*, AD₃)(Chen et al., 2020b)和黄褐棉(*G. mustelinum*, AD₄)(Chen et al., 2020b),锦葵科的叉柱棉(*Gossypioides kirki*)(Udall et al., 2019b)、圆果种黄麻(*Corchorus capsularis*)、长果种黄麻(*C. olitorius*)(Zhang et al., 2021)、哥伦比亚锦葵(*Herrania umbratica*)(<http://magen.whu.edu.cn>)、可可(*Theobroma cacao*)(Argout et al., 2017)和木槿(*Hibiscus syriacus*)(Kim et al., 2017)等的全基因组蛋白质序列,其基因组蛋白质序列与基因组注释文件可从附表1中的网址下载。陆地棉不同组织的转录组数据来源于NCBI(登录号:PRJNA490626)(Zhang et al., 2015; Hu et al., 2019),陆地棉野生型和栽培品种在长、短日照处理下的转录组数据来自NCBI(登录号:PRJNA529417),

陆地棉与海岛棉重测序数据来自NCBI(登录号: PRJNA414461)(Yuan et al., 2021)。

1.2 PHY基因鉴定

利用拟南芥光敏色素基因家族成员PHYA (AT1G09570.1)、PHYB (AT2G18790.1)、PHYC (AT5G35840.1)、PHYD (AT4G16250.1)和PHYE (AT4G18130.1)的蛋白质序列作为查询序列,对上述物种的全基因组蛋白质序列进行blastp搜索。参数设置: e-value $\leq 1e-20$, qcovs>50, pident>50, bitscore>50。从PFAM (<http://pfam.xfam.org/>)网站下载PHY.hmm(登录号: PF00360)。利用HMMER 2.0软件对各物种全基因组蛋白质序列进行hmmsearch搜索,设置e-value $\leq 1e-5$ 。结合blastp与hmmsearch结果,进一步通过在线工具Search Pfam (<http://pfam.xfam.org/search>)进行验证,去除保守结构域不完整的序列,得到PHY家族成员。使用Perl脚本stat_protein_fa.pl(附录1)计算棉花PHY基因家族成员的分子量和等电点。

1.3 PHY基因进化与基因结构分析

利用MAFFT软件(Katoh and Standley, 2013)选择默认参数对棉花PHY氨基酸序列进行多序列比对。使用IQ-TREE (Nguyen et al., 2015)软件基于最大似然法构建进化树,使用-m参数自动计算适合的模型,Bootstrap值设为1 000。利用在线工具EvoView (<https://www.evolgenius.info/>)以及Adobe illustrator对进化树进行修饰和美化(Zhang et al., 2012)。使用Jalview软件(Procter et al., 2021)对多序列比对结果进行可视化。利用MEME工具(<http://meme.ebi.edu.au/meme/cgi-bin/meme.cgi>)预测蛋白质保守基序。参数设置:最大基序数量=15,最小基序数量=3,最大基序宽度=50,设置e-value $< 1e-30$ 。从基因组注释文件中检索基因的外显子-内含子结构信息,并使用TBtools v1.0 (Chen et al., 2020a)绘制基因结构和保守基序示意图。

1.4 染色体定位与顺式作用元件分析

利用TBtools v1.0软件(Chen et al., 2020a)绘制二倍体雷蒙德氏棉(D₅)、亚洲棉(A₂)和草棉(A₁)以及四倍体陆地棉(AD₁)和海岛棉(AD₂)的PHY基因家族成员染色体定位图。

从基因组序列中提取亚洲棉(A₂)、草棉(A₁)、雷蒙德氏棉(D₅)、陆地棉(AD₁)和海岛棉(AD₂)PHY基因上游2 000 bp序列作为启动子。利用在线软件PlantCARE (Lescot et al., 2002)分析启动子顺式作用元件。利用在线软件GSDS 2.0 (<http://gsds.gao-lab.org/>)绘制顺式作用元件分布图。

1.5 PHY基因家族的共线性与选择压力分析

利用MCScanX软件(Wang et al., 2012)选择默认参数分析在陆地棉(AD₁)、海岛棉(AD₂)、亚洲棉(A₂)和雷蒙德氏棉(D₅)中PHY基因家族成员的共线性关系。使用TBtools v1.0软件(Chen et al., 2020a)(TBtools原生方法)选择默认参数批量计算PHY基因的非同义替换率/同义替换率(K_a/K_s)。

1.6 PHY基因家族的表达分析

使用Prefetch tool (<https://github.com/ncbi/sra-tools/tree/master/tools/prefetch>)下载陆地棉(AD₁)根、茎、叶、花托、萼片、雌蕊、雄蕊和花瓣的转录组数据,用fastq-dump (<https://github.com/ncbi/sra-tools/tree/master/tools/fastq-dump>)将sra数据转换为fastq格式,并选择默认参数进行质量控制。用Hisat2 v2.2.1软件(Kim et al., 2019)将质控后的转录组数据比对到参考基因组TM-1 (Saski et al., 2017)上,然后使用StringTie v2.1.4软件(Pertea et al., 2015)计算基因表达量(FPKM值),均选择默认参数。利用TBtools v1.0软件(Chen et al., 2020a)绘制基因表达热图。

1.7 陆地棉核苷酸多样性和遗传分化系数分析

在Yuan等(2021)分析的重测序数据基础上,提取731份陆地棉(23份野生型、193份半野生型和515份栽培品种)的变异数据,使用VCFtools (Danecek et al., 2011)和Pixy (Korunes and Samuk, 2021)软件计算陆地棉(AD₁)野生型、半野生型和栽培品种群体类型中PHY基因的核苷酸多样性(π)和群体间的遗传分化系数(F_{st})。

1.8 陆地棉单倍型分析

利用731份陆地棉重测序数据(Yuan et al., 2021),使用VCFtools软件(Danecek et al., 2011)提取PHY基因所在区段中的变异信息,利用R包pheatmap聚类,

得到4种单倍型。使用ggplot2、plyr和ggthemes等R包绘制热图和柱状图。

1.9 陆地棉转录组分析

使用fastq-dump (<https://github.com/ncbi/sra-tools/tree/master/tools/fastq-dump>)将陆地棉(AD₁)叶片在不同处理下的转录组数据从sra转换为fastq格式,并利用Trimmomatic (Bolger et al., 2014)去除接头和低质量的序列,获得干净序列。使用Hisat2 v2.2.1 (Kim et al., 2019)将干净序列比对到参考基因组TM-1 (Saski et al., 2017)上。使用featurecounts (Liao et al., 2014)统计比对到每个基因的干净序列数,获得基因表达量。利用SPSS软件中的成对样本T检验验证长、短日照下不同处理时间之间和At和Dt亚基因组之间,以及相同处理时间下长、短日照之间和栽培品种与野生型之间的差异表达基因。

2 结果与讨论

2.1 棉花*PHY*基因家族成员鉴定及其理化性质

综合BLASTP和hmmer程序分析结果,从二倍体棉种亚洲棉(A₂)、草棉(A₁)和雷蒙德氏棉(D₅)中分别鉴定到5个*PHY*基因,其中*PHYA*有2个,*PHYB*、*PHYC*和*PHYE*各1个;从四倍体棉种陆地棉(AD₁)和海岛棉(AD₂)中均鉴定到10个*PHY*基因,各亚家族的基因数均为二倍体棉种的2倍;在5个棉种中均未检测到*PHYD*基因。通过理化性质分析,发现棉花*PHY*蛋白包含884–1 199个氨基酸残基,大部分蛋白的分子量为99 430.1–133 314.7 Da,理论等电点为5.17–6.51。

由于*GhePHYA1At* (*Ghe11G33390*)、*GhePHYA2At* (*Ghe13G23940*)、*GaPHYB1At* (*Gar10G22470*)、*GaPHYA1At* (*Gar11G33210*)和*GaPHYA2At* (*Gar13G23780*)基因结构与*PHY*基因其它成员差异很大,经手工检查核实其基因结构预测有误,故对其进行基因注释矫正,并将注释后的基因用于后续分析(附图1)。

2.2 *PHY*蛋白系统发育分析

对4个四倍体棉种((AD)₁、(AD)₂、(AD)₃和(AD)₄)、11个二倍体棉种(A₁、A₂、D₁、D₂₋₁、D_{3d}、D_{3-k}、D₅、

D₈、D₁₀、F₁和G₂)和锦葵科6个物种(圆果种黄麻、长果种黄麻、哥伦比亚锦葵、可可、木槿和叉柱棉)进行系统发育分析,发现四倍体棉种和木槿均含有10个*PHY*基因,其中*PHYA*有4个,*PHYB*、*PHYC*和*PHYE*各2个;二倍体棉种及叉柱棉均含有5个*PHY*基因,其中*PHYA*有2个,*PHYB*、*PHYC*和*PHYE*各1个;锦葵科圆果种黄麻、哥伦比亚锦葵和可可均含有4个*PHY*基因,*PHYA*、*PHYB*、*PHYC*和*PHYE*各1个,而长果种黄麻仅有3个*PHY*基因,较前面3个物种缺失了*PHYB*亚家族基因(附表2)。

对1.2节中的5条拟南芥*PHY*蛋白序列以及126条锦葵科*PHY*蛋白序列进行系统发育分析,确定*PHY*蛋白家族的进化关系。结果显示,所有*PHY*蛋白分为4个家亚族,分别为*PHYA*、*PHYB*、*PHYC*与*PHYE*,每个亚家族分支内蛋白结构较为一致(图1)。

2.3 棉花*PHY*基因结构与保守基序分析

通过对草棉(A₁)、亚洲棉(A₂)、雷蒙德氏棉(D₅)、陆地棉(AD₁)及海岛棉(AD₂)的*PHY*氨基酸序列进行保守性分析,结果表明,35条*PHY*氨基酸序列包含15个高度保守的基序(图2A, B)。分析结果显示,*PHY*蛋白均含有4个外显子,说明该基因家族的结构较为保守(图2C)。

使用Jalview软件(Procter et al., 2021)对5条拟南芥*PHY*氨基酸序列以及上述35条棉属*PHY*氨基酸序列比对结果进行可视化,发现各亚家族内的多序列比对结果基本一致,说明同一亚家族内的序列高度保守。此外,陆地棉(AD₁)和海岛棉(AD₂)中相同*PHY*亚家族内At和Dt亚组间的基序基本一致,说明在同一棉种内不同亚组的*PHY*氨基酸序列高度保守(附图2)。

2.4 棉花*PHY*基因的染色体分布与共线性分析

*PHY*家族基因的染色体定位(附图3)结果表明,*PHY*家族基因均匀分布在3条或6条染色体上,二倍体棉种亚洲棉(A₂)、草棉(A₁)和雷蒙德氏棉(D₅)的*PHY*基因分别分布在第10、11和13号染色体上。10号(*PHYB1*和*PHYE1*)和11号(*PHYA1*和*PHYC1*)染色体分别含有2个*PHY*基因,这4个基因均分布在靠近染色体两端的位置;13号染色体上仅含有1个*PHY*基因(*PHYA2*),该基因分布在染色体的一端(附图3)。异源

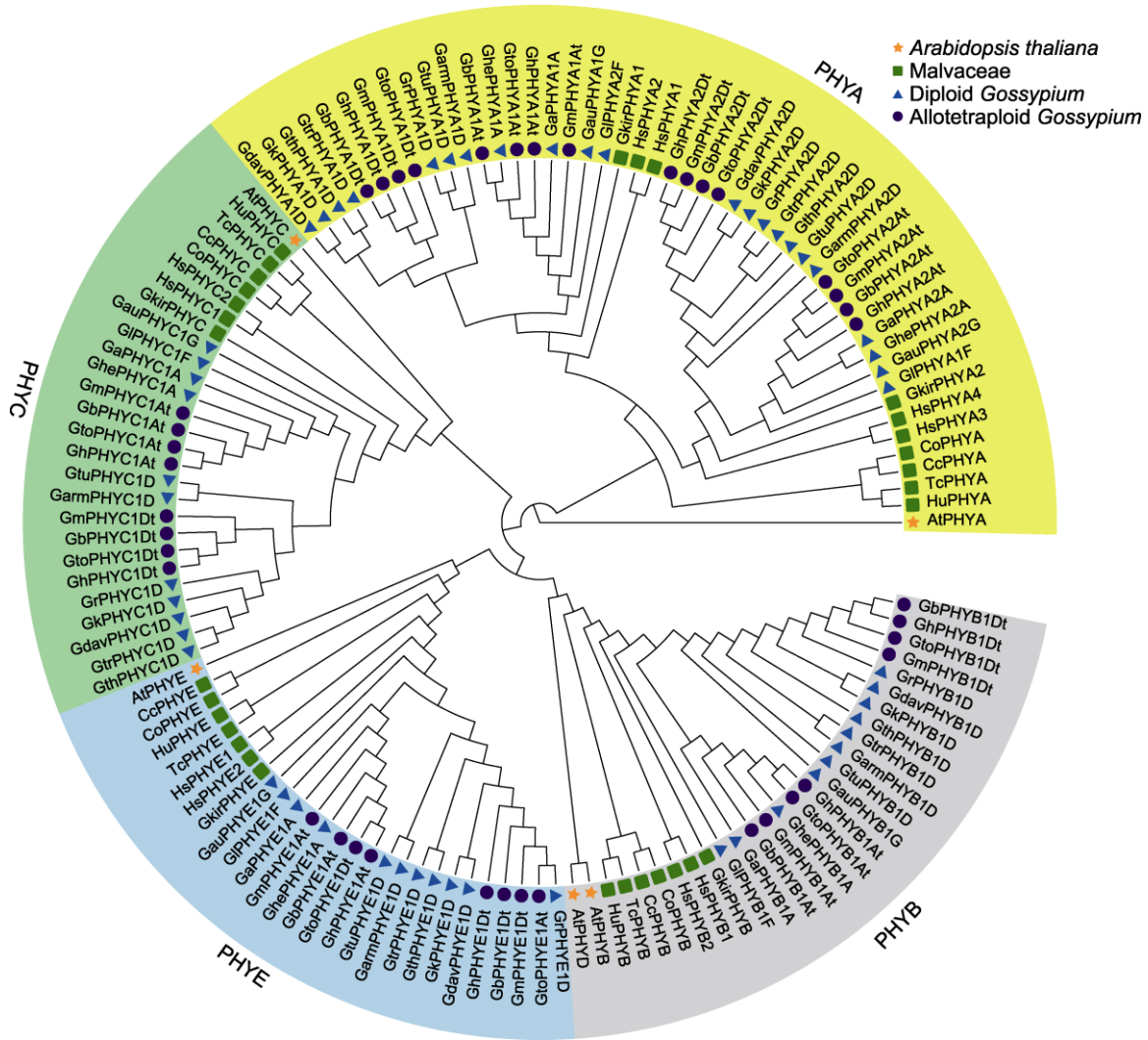


图1 棉属及其它植物中PHY家族成员的系统发育分析

At: 拟南芥; Cc: 圆果种黄麻; Co: 长果种黄麻; Ga: 亚洲棉; Garm: 辣根棉; Gau: 澳洲棉; Gb: 海岛棉; Gdav: 戴维逊棉; Gh: 陆地棉; Ghe: 草棉; Gk: 克劳茨基棉; Gl: 长萼棉; Gm: 黄褐棉; Gr: 雷蒙德氏棉; Gth: 瑟伯氏棉; Gto: 毛棉; Gtr: 三裂棉; Gtu: 特纳氏棉; Gkir: 叉柱棉; Hs: 木槿; Hu: 哥伦比亚锦葵; Tc: 可可

Figure 1 Phylogenetic analysis of PHY family members of *Gossypium* and other plants

At: *Arabidopsis thaliana*; Cc: *Corchorus capsularis*; Co: *C. olerius*; Ga: *G. arboreum*; Garm: *G. armourianum*; Gau: *G. australe*; Gb: *G. barbadense*; Gdav: *G. davidsonii*; Gh: *G. hirsutum*; Ghe: *G. herbaceum*; Gk: *G. klotzschianum*; Gl: *G. longicalyx*; Gm: *G. mustelinum*; Gr: *G. raimondii*; Gth: *G. thurberi*; Gto: *G. tomentosum*; Gtr: *G. trilobum*; Gtu: *G. turneri*; Gkir: *Gossypoides kirkii*; Hs: *Hibiscus syriacus*; Hu: *Herrania umbratica*; Tc: *Theobroma cacao*

四倍体棉种陆地棉(AD₁)和海岛棉(AD₂)的PHY基因分别分布在At和Dt亚组的第10、11和13号染色体上,其中A10、A11、D10和D11号染色体分别含有2个PHY基因,且分布在染色体靠近两端的位置,A13和D13号染色体均含有1个PHY基因,这2个基因分布在靠近一端的位置(附图3)。

为揭示棉花PHY基因在不同棉种间的进化关系,对陆地棉(AD₁)、雷蒙德氏棉(D₅)和亚洲棉(A₂)的PHY基因进行共线性分析,结果显示,5个基因在各棉种染色体上的位置一致且同源基因一一对应,表明这3个棉种间的PHY基因具有良好的共线性关系(附图4)。

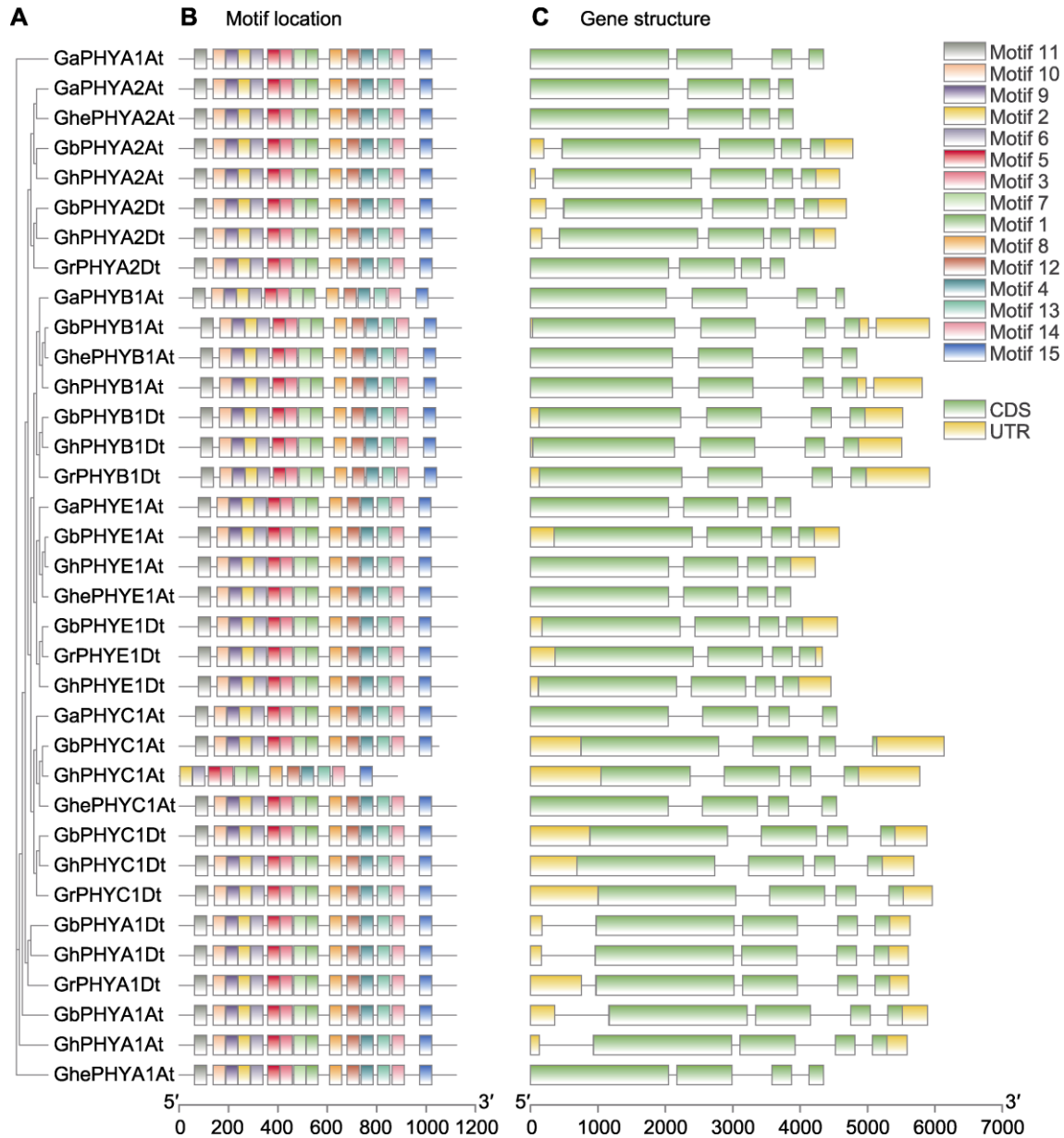


图2 棉花 *PHY* 基因结构和蛋白保守基序分析
(A) 系统进化树; (B) 蛋白保守基序; (C) 基因结构

Figure 2 Analysis of the *PHY* gene structure and protein conserved motifs of *Gossypium*
(A) Phylogenetic tree; (B) Protein conserved motif; (C) Gene structure

2.5 *PHY* 基因的选择压力分析

使用 TBtools (Chen et al., 2020a) 分别计算陆地棉 (AD₁) 与亚洲棉 (A₂) 和雷蒙德氏棉 (D₅) 重复基因的非同义替换率 (K_a)、同义替换率 (K_s) 以及二者的比值 (K_a/K_s)，结果显示，在本研究中发现的 10 个基因重复事件中，所有重复基因都经历了强烈的纯化选择 ($K_a/K_s < 1$) (附表 3)。

2.6 棉花 *PHY* 基因的顺式作用元件分析

从陆地棉 (AD₁)、海岛棉 (AD₂)、亚洲棉 (A₂)、草棉 (A₁) 和雷蒙德氏棉 (D₅) 中选取已鉴定的 35 个 *PHY* 基因，对其启动子区域的顺式作用元件进行分析。本研究鉴定到的顺式作用元件可分为 4 类：光响应 (light response)、胁迫响应 (stress response)、激素响应 (hormone response) 和生长调节 (growth regulation)

(附图5)。光响应包含不同种类与光反应相关的响应元件；胁迫响应包括应激响应元件、干旱响应元件和厌氧诱导响应元件；激素响应包括生长素响应元件、赤霉素响应元件、脱落酸响应元件和水杨酸响应元件；生长调节包括与栅栏组织发育相关、与分生组织发育相关、与胚乳发育相关和与昼夜节律控制相关的响应元件。除*GbPHYE1At*不含有与激素反应相关的顺式作用元件外，其余棉花*PHY*基因均包含4种顺式作用元件。

棉花*PHY*基因中与光反应相关的顺式作用元件数量较多且种类丰富，因此我们对其分布作了分析。异源四倍体棉种的*PHYA1*和*PHYB*中，*At*光响应元件数量多于*Dt*。*PHYA2*中*At*与*Dt*光响应元件数量基本一致，但分布位置略有差异。*At*分布靠近两端，*Dt*则均匀分布。*PHYC*中*Dt*光响应元件数量多于*At*，但*At*亚组内的作用元件分布位置与*Dt*一致。相比其它*PHY*成员，

*PHYE*含有的光响应元件较少，*At*光响应元件分布位置靠近两端，而*Dt*光响应元件分布在中间(图3)。

2.7 陆地棉*PHY*家族基因的表达

RNA-seq分析结果显示，*PHY*家族基因在不同器官及组织中呈特异性表达，且不同亚家族基因的表达规律不同。*PHYA*亚家族的*GhPHYA1At*和*GhPHYA1Dt*在根、茎、叶和花托中表达量均较高，但*GhPHYA2At*和*GhPHYA2Dt*仅在茎中高表达，在根、花萼和花托中表达量较低或几乎不表达。*PHYB*亚家族的*GhPHYB1At*和*GhPHYB1Dt*在根、叶、花托及雌蕊中表达量较高。*PHYC*亚家族的*GhPHYC1At*和*GhPHYC1Dt*在茎和花托中高表达。*PHYE*亚家族的*GhPHYE1At*和*GhPHYE1Dt*在根、花萼和雌蕊中表达量较高，在茎中表达量较低，而*GhPHYE1Dt*在花瓣和雄蕊中几乎不表达(图4)。

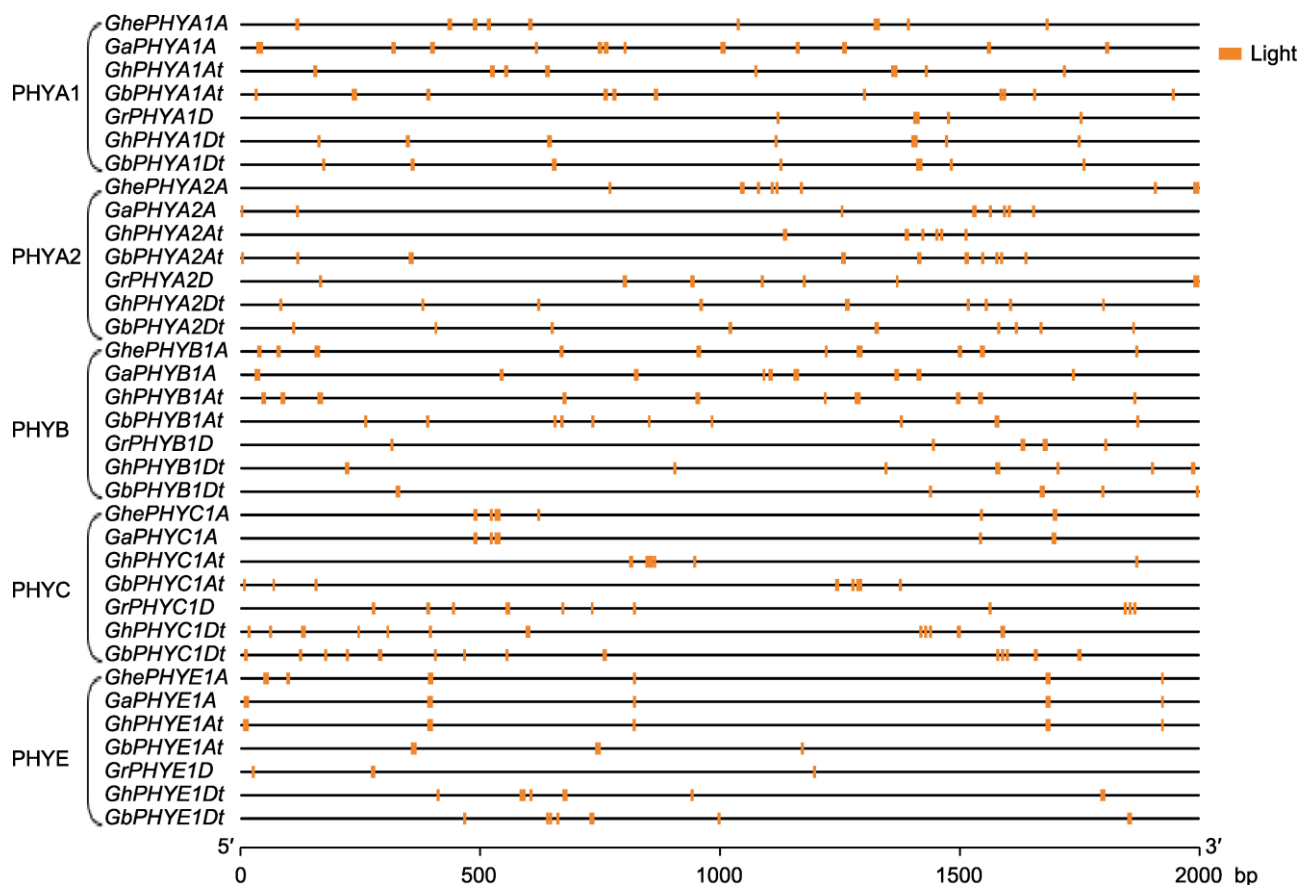


图3 棉花*PHY*基因光响应顺式作用元件分布

Figure 3 Distribution of light response *cis*-acting elements in the *Gossypium PHY* genes

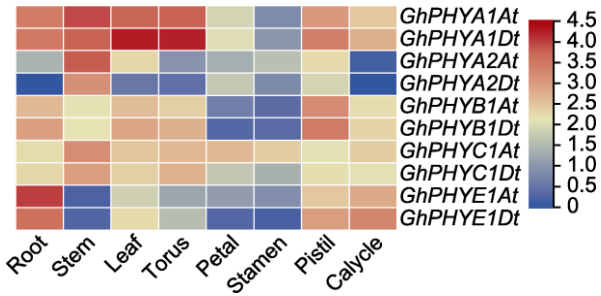


图4 陆地棉 *PHY* 基因在不同器官和组织中的表达聚类分析
不同颜色的矩形代表陆地棉不同 *PHY* 基因的表达水平, 即 $\log_2(1+\text{FPKM})$ 。

Figure 4 Clustering analysis of the expression of *Gossypium hirsutum* *PHY* genes in different organs and tissues
The rectangles with different colors represent the expression levels of different *PHY* genes in upland cotton, namely $\log_2(1+\text{FPKM})$.

2.8 *PHY* 基因在陆地棉不同群体间的驯化选择

野生种经过早期驯化后被称为农家种或半野生种, 半野生种经过遗传改良成为栽培品种。通过计算 *PHY* 基因的核苷酸多样性以及遗传分化系数可分析 *PHY* 基因在驯化过程中受到的人工选择。

使用 Pixy 软件计算陆地棉 (AD_1) 23 份野生种 (wild)、193 份半野生种 (landrace) 和 515 份栽培品种 (cultivar) 的 *PHY* 基因在基因 (gene) 及其上、下游 2 000 bp、外显子 (exon) 和内含子 (intron) 区域的核苷酸多样性 (附表 4)。

通过对陆地棉 (AD_1) 全基因组核苷酸多样性进行分析, 发现 P_i 值在半野生种、野生种和栽培品种中逐渐降低 ($P_i \times 10^{-4}$ 值分别为 1.03 ± 0.55 、 1.02 ± 0.51 和 0.60 ± 0.17), 半野生种与栽培品种的核苷酸多样性比值 (1.71) 高于半野生种与野生种的比值 (0.99), 表明栽培品种比半野生种受到更强烈的人工选择。

对陆地棉 (AD_1) 不同基因区域的核苷酸多样性进行分析, 发现 *GhPHYA2At*、*GhPHYB1At* 和 *GhPHYA1Dt* 在所研究的 5 种区域内其 P_i 值在野生种、半野生种和栽培品种中呈逐渐递减, *GhPHYE1Dt* 在野生种、半野生种和栽培品种中呈逐渐递减 (除基因下游及内含子区域), *GhPHYC1At* 在半野生种、野生种和栽培品种中呈逐渐递减 (除基因下游及内含子区域), 而其余 5 个基因在 5 种区域内的核苷酸多样性变化规律不一致。*GhPHYA1At* 和 *GhPHYB1Dt* 在 4 个区域的 P_i 值均在半野生种、野生种和栽培品种中逐渐递减, *GhPHYA1At* 的外显子及 *GhPHYB1Dt* 的内含子的 P_i 值在野生种、半野生种和栽培品种中逐渐递减 (图 5)。

陆地棉 (AD_1) *PHY* 基因的核苷酸多样性分析表明, 基因区 P_i 平均值在 3 个群体中均显著低于陆地棉 (AD_1) 全基因组, 说明基因区的核苷酸多样性低于全基因组水平。7 个 *PHY* 基因的核苷酸多样性从野生种、半野生种到栽培品种逐渐递减, 3 个基因 (*GhPHYA1At*、*GhPHYB1Dt* 和 *GhPHYC1At*) 中半野生种的核苷酸多样性高于野生种。

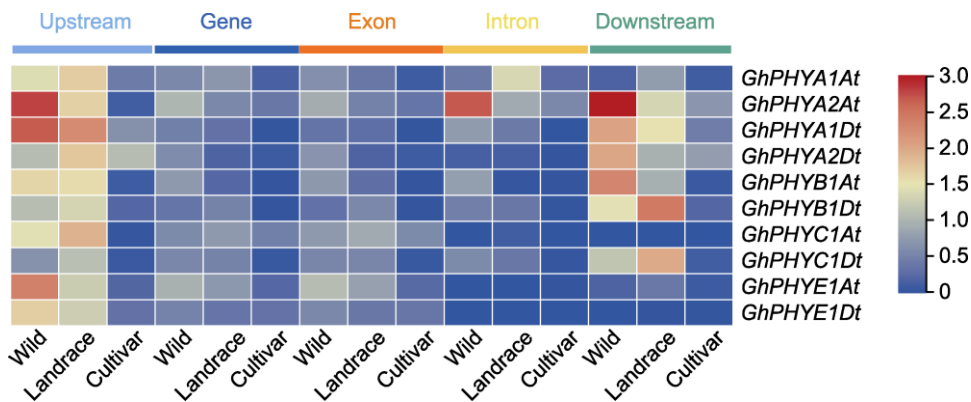


图5 陆地棉 *PHY* 基因在不同群体间的核苷酸多样性分析
不同颜色的矩形代表陆地棉不同材料中每个基因及其上、下游、外显子和内含子的核苷酸多样性 ($P_i \times 10^{-4}$)。

Figure 5 Analysis of nucleotide diversity of *Gossypium hirsutum* *PHY* genes among different populations
Different color rectangles represent the nucleotide diversity ($P_i \times 10^{-4}$ value) of each gene and its upstream, downstream, exon and intron in different materials of upland cotton.

*PHY*基因的核苷酸多样性在不同群体间的比值分析表明,野生种-半野生种组合中*GhPHYA2Dt*和*GhPHYB1At*的比值高于其它基因,说明这2个基因在驯化过程中受到了强烈的选择;半野生种-栽培品种组合中*GhPHYA1Dt*、*GhPHYB1Dt*、*GhPHYB1At*和*GhPHYC1Dt*的比值显著高于其它6条序列,*GhPHYA1At*和*GhPHYE1At*的比值略高于其余4条序列(*GhPHYA2At*、*GhPHYA2Dt*、*GhPHYC1At*和*GhPHYE1Dt*),表明在遗传改良过程中光周期相关基因受到了改造(表1)。

*PHY*基因在野生种与半野生种及半野生种与栽培品种之间的群体遗传分析表明,在野生种与半野生种之间,*GhPHYB1At*的群体间遗传分化系数最大(0.706 8),其次是*GhPHYA2Dt*(0.604 9)、*GhPHYA1At*(0.584 3)、*GhPHYC1Dt*(0.547 0)和*GhPHYB1Dt*(0.515 8),*GhPHYA1Dt*的遗传分化系数最小(0.056 9);在半野生种与栽培品种之间,*GhPHYC1Dt*的群体间遗传分化系数最大(0.709 2),其次是*GhPHYB1Dt*和*GhPHYB1At*,两者均大于0.5;*GhPHYB1At*、*GhPHYB1Dt*和*GhPHYC1Dt*三个基因在野生种与半野生种、半野生种与栽培品种之间的遗传分化系数均大于0.5,说明它们在驯化和遗传改良过程中均经历了强烈的人工选择,而*GhPHYA2Dt*和*GhPHYA1At*在驯化过程中受到较强的选择(表2)。

综合核苷酸多样性与遗传分化分析结果,我们发现*GhPHYB1At*在野生种与半野生种、半野生种与栽培品种之间的Pi比值和Fst比值均高于其它*PHY*基因,且在5个基因区域中的核苷酸多样性从野生种、半野生种到栽培品种呈逐渐降低的趋势,说明该基因在从野生种被早期驯化为半野生种以及从半野生种到栽培品种的过程中受到的驯化选择均最强烈。在野生种-半野生种组合中,*GhPHYA2Dt*的Pi比值和Fst比值均仅次于*GhPHYB1At*,但其在半野生种-栽培品种组合中数值较低,说明该序列在从野生种到半野生种的驯化过程中受到了较强烈的选择,但在后续遗传改良过程中受到的影响不大。在半野生种-栽培品种组合中,*GhPHYB1Dt*及*GhPHYC1Dt*的Pi比值和Fst比值均较高,且显著高于野生种-半野生种中对应的数值,说明相比从野生种到半野生种的驯化过程,这2个基因在从半野生种到栽培品种的遗传改良过程中受到较大影响。

表1 陆地棉不同群体间*PHY*基因的核苷酸多样性(Pi)比值

Table 1 Nucleotide diversity (Pi) ratio of *PHY* genes in *Gossypium hirsutum* among different populations

Gene name	Gene ID	Pi _{wild} / Pi _{Landrace}	Pi _{Landrace} / Pi _{Cultivar}
<i>GhPHYA1At</i>	Gohir.A11G261900	0.7891	5.1234
<i>GhPHYA2At</i>	Gohir.A13G172900	1.8274	1.4509
<i>GhPHYA1Dt</i>	Gohir.D11G271900	1.4826	99.2593
<i>GhPHYA2Dt</i>	Gohir.D13G178700	3.5917	2.3415
<i>GhPHYB1At</i>	Gohir.A10G172500	3.4313	16.8376
<i>GhPHYB1Dt</i>	Gohir.D10G179000	0.6808	38.3027
<i>GhPHYC1At</i>	Gohir.A11G149100	0.7955	1.5879
<i>GhPHYC1Dt</i>	Gohir.D11G156200	1.0780	15.8177
<i>GhPHYE1At</i>	Gohir.A10G037800	1.3580	3.4484
<i>GhPHYE1Dt</i>	Gohir.D10G038800	1.4530	1.0696

表2 陆地棉*PHY*基因在不同群体间的遗传分化系数(Fst)分析

Table 2 Analysis of genetic differentiation coefficients (Fst) of *PHY* genes of *Gossypium hirsutum* among different populations

Gene name	Gene ID	Fst _{wild} / Fst _{Landrace}	Fst _{Landrace} / Fst _{Cultivar}
<i>GhPHYA1At</i>	Gohir.A11G261900	0.5843	0.2320
<i>GhPHYA2At</i>	Gohir.A13G172900	0.4386	0.1680
<i>GhPHYA1Dt</i>	Gohir.D11G271900	0.0569	0.2137
<i>GhPHYA2Dt</i>	Gohir.D13G178700	0.6049	0.0374
<i>GhPHYB1At</i>	Gohir.A10G172500	0.7068	0.5372
<i>GhPHYB1Dt</i>	Gohir.D10G179000	0.5158	0.7092
<i>GhPHYC1At</i>	Gohir.A11G149100	0.2304	0.4461
<i>GhPHYC1Dt</i>	Gohir.D11G156200	0.5470	0.8466
<i>GhPHYE1At</i>	Gohir.A10G037800	0.4756	0.4709
<i>GhPHYE1Dt</i>	Gohir.D10G038800	0.4096	0.1872

这些数据支持陆地棉(AD₁)中*PHY*基因可能经历了2次驯化的观点。在早期驯化过程中,驯化选择的目标性状可能与满足人类需求、适合栽培及产量高等因素相关,因此对光周期相关基因的驯化选择不是主要需求,但在驯化的地理扩散阶段或后期目的性明确的选择育种过程中,随着驯化选择的目标性状从最初的地理起源区域到实现产量最大化,与光周期相关的表型则更多地受到了选择,因此*PHY*基因所受到的驯化选择在这个阶段较为强烈。

2.9 陆地棉单倍型分析

陆地棉经历了从野生种到半野生种及半野生种到栽培品种的驯化过程。为明确这2个驯化过程中*PHY*基

因的哪些变异导致其对光周期的敏感性发生变化, 我们对731份陆地棉材料进行聚类, 并对各基因的不同单倍型进行了分析(附图6)。

结果表明, *GhPHYA2At*、*GhPHYB1Dt*、*GhPHYC1Dt*及*GhPHYE1Dt*在从野生种到半野生种的驯化过程中, 各单倍型的种类及比例变化较小, 说明这些基因在早期驯化过程中受到的选择较弱; 而其它基因则受到不同程度的驯化选择。在从半野生种到栽培品种的驯化过程中, 所有*PHY*基因的各单倍型种类和(或)比例均发生了较大变化, 说明在遗传改良过程中, 这些基因受到了不同程度的人工选择。

结合*PHY*基因核苷酸多样性(π)和遗传分化系数(F_{st})分析, 我们总结了基因的单倍型分布。结果表明, *GhPHYA2Dt*和*GhPHYB1At*的单倍型1所占比例逐渐降低, 单倍型3所占比例逐渐增高, 表明这2个基因在早期驯化和后期遗传改良阶段均受到了较强的人工选择(图6A, B)。 *GhPHYB1Dt*和*GhPHYC1Dt*的单倍型1在野生种和半野生种中占比均很高, *GhPHYB1Dt*占比分别为100.00%和98.45%, *GhPHYC1Dt*占比分别为86.96%和81.87%; 而在栽培品种中2个基因的

单倍型1占比均很低, 分别为4.08%和0.97%。 *GhPHYB1Dt*从半野生种到栽培品种的遗传改良过程中单倍型2 (25.63%)、单倍型3 (41.55%)和单倍型4 (28.74%)占比增高; *GhPHYC1Dt*在栽培品种中单倍型2 (66.99%)和单倍型3 (32.04%)占比增高, 说明这2个基因在后期遗传改良阶段受到了较强的人工选择(图6C, D)。

2.10 陆地棉转录组分析

本研究利用陆地棉(AD₁) 2个野生型(TX2094和TX2095)和2个栽培品种(CascadL7和Maxxa)在长日照(7:00 (LD7)、14:00 (LD14)和21:00 (LD21))和短日照(7:00 (SD7)、14:00 (SD14)和17:00 (SD17))条件下的叶片转录组数据(附表5), 分析*PHY*基因在野生种与栽培品种中对光照的响应模式。

2.10.1 陆地棉*PHY*基因亚基因组表达的偏好性

在长日照和短日照处理下, 栽培品种与野生种的*GhPHYA1*和*GhPHYB1*表现出明显的Dt亚基因组表达偏好性, *GhPHYA2*和*GhPHYC1*则表现出明显的At

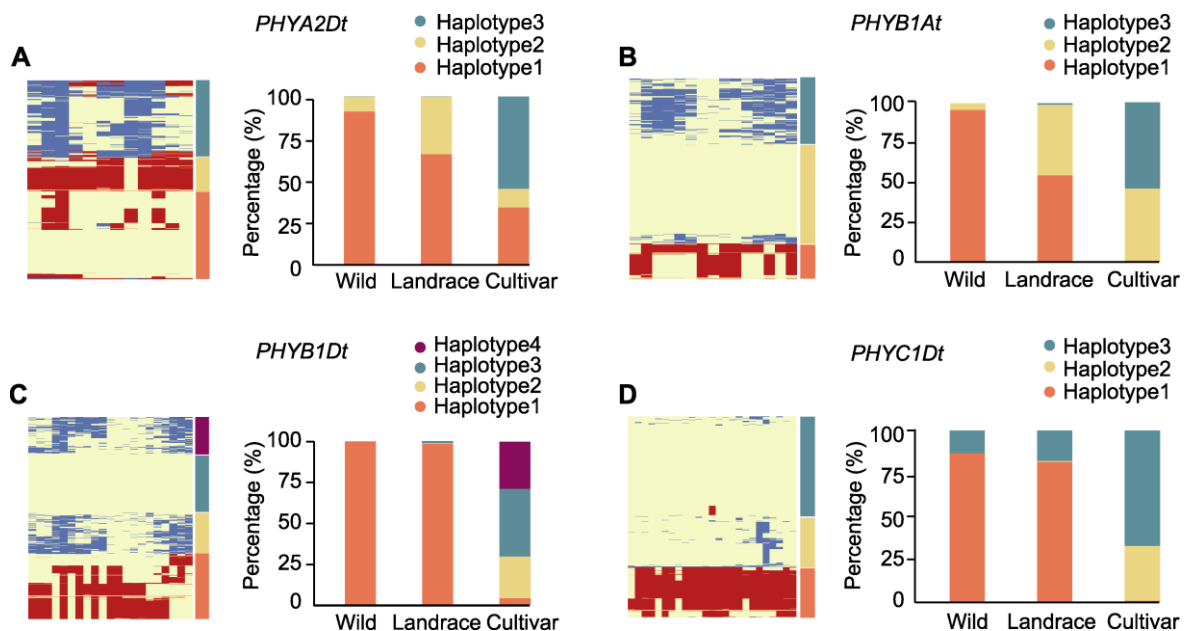


图6 *PHY*基因在野生种、半野生种及栽培品种中的单倍型分析

左图代表不同材料的聚类, 右图代表不同单倍型在野生种、半野生种及栽培品种中所占的比例。

Figure 6 Haplotype analysis of *PHY* genes in wild, landrace and cultivar

The left figure represents the clustering of different materials, and the right figure represents the proportion of different haplotypes in wild, landrace and cultivar.

亚基因组表达偏好性(附图7)。*PHYE*在长、短日照处理下的表达偏好性表现为: (1) 长、短日照处理开始(SD7和LD7)时,栽培品种和野生种*GhPHYE1At*的表达量均极显著高于*GhPHYE1Dt*; (2) 长、短日照处理7小时(SD14和LD14)和处理结束(SD17和LD21)时,栽培品种和野生种*GhPHYE1At*的表达量均低于*GhPHYE1Dt*(图7)。上述结果表明,*PHYA*、*PHYB*和*PHYC*亚家族在从野生种至栽培品种的驯化过程中表达偏好性并未发生较大变化;对于*PHYE*亚家族,在长日照处理7小时和14小时后栽培品种中*GhPHYE1Dt*的表达量显著高于*GhPHYE1At*,而在野生种中则差异不显著,说明在驯化过程中,随着光照处理时间的延长,*PHYE*亚家族基因的表达从At亚基因组向Dt亚基因组偏移,这可能是栽培品种能在长日照条件下正常开花的重要原因。

2.10.2 陆地棉*PHY*基因在长、短日照下的表达模式

栽培品种的*GhPHYA1At*、*GhPHYA1Dt*、*GhPHYB1At*和*GhPHYC1At*及野生种的*GhPHYA1At*、*GhPHYB1At*、*GhPHYB1Dt*、*GhPHYC1At*和*GhPHYC1Dt*在长、短日照处理下均表现相似的表达模式,即光照处理开始时(SD7和LD7)与光照处理7小时(SD14和LD14)后,*PHY*基因在长日照条件下的表达量低于短日照处理;但在光照处理结束(SD17和LD21)时,表达情况则相反,说明*PHY*基因的表达水平随着光照时间的延长而增高。在栽培品种与野生种中,*PHYE*亚家族基因在光照处理开始时,长日照条件下的表达量低于短日

照;但在光照处理7小时(SD14和LD14)和光照处理结束(SD17和LD21)时,长日照条件下的表达量均高于短日照(附图8)。

有趣的是,栽培品种的*GhPHYA1At*、*GhPHYA1Dt*、*GhPHYC1At*、*GhPHYE1At*和*GhPHYE1Dt*及野生种的*GhPHYC1At*和*GhPHYC1Dt*在长、短日照处理下,光照处理开始时的表达量均高于光照处理7小时,而光照处理结束时的表达量均高于光照处理7小时(附图8),说明*PHY*基因的表达在黑暗下被激活,转录产物积累,在恢复光照后,转录产物的富集抑制*PHY*基因的表达,随着光照处理时间的延长转录产物被消耗,从而促进*PHY*基因的表达。栽培品种的*GhPHYC1Dt*与野生种不同,其表达在长、短日照处理条件下均随光照处理时间的延长而降低,这说明长日照条件下栽培品种可能通过降低*GhPHYC1Dt*的表达来适应长日照环境,从而使植物正常开花。

在长、短日照条件下,*GhPHYA1At*、*GhPHYA1Dt*、*GhPHYE1At*和*GhPHYE1Dt*在野生种中的表达规律截然不同。在长日照条件下,光照处理开始时与光照处理结束时的表达量均高于光照处理7小时;在短日照条件下,则是处理开始时表达量最高,之后逐渐递减。这说明野生种在长日照条件下不开花的原因很可能是由于光照时间的延长,*GhPHYA1At*、*GhPHYA1Dt*、*GhPHYE1At*和*GhPHYE1Dt*的表达产物逐渐积累,从而抑制植物开花。值得注意的是,*PHYB*亚家族基因的表达则是随着光照时间的延长而增高,且栽培品种在长、短日照下的表达量无显著差异,而野

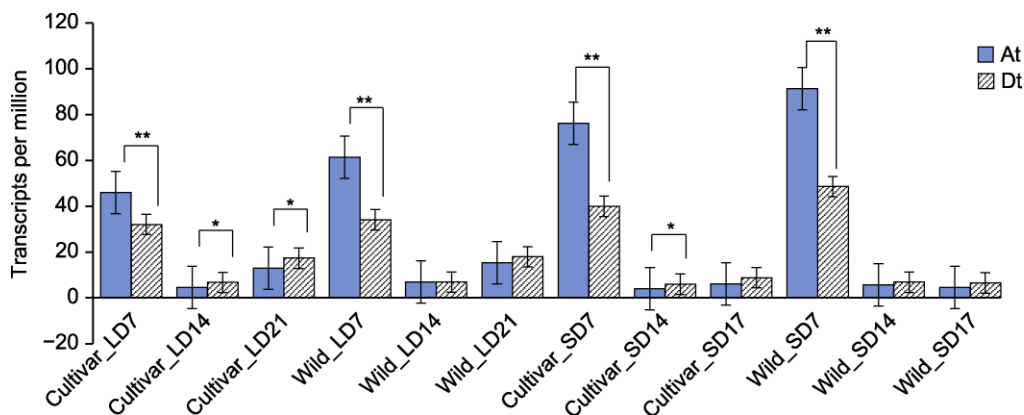


图7 陆地棉*GhPHYE1*基因在不同亚基因组中的表达

LD: 长日照; SD: 短日照; * $P < 0.05$; ** $P < 0.01$

Figure 7 Expression of *GhPHYE1* genes in different subgenomes of *Gossypium hirsutum*

LD: Long-day; SD: Short-day; * $P < 0.05$; ** $P < 0.01$

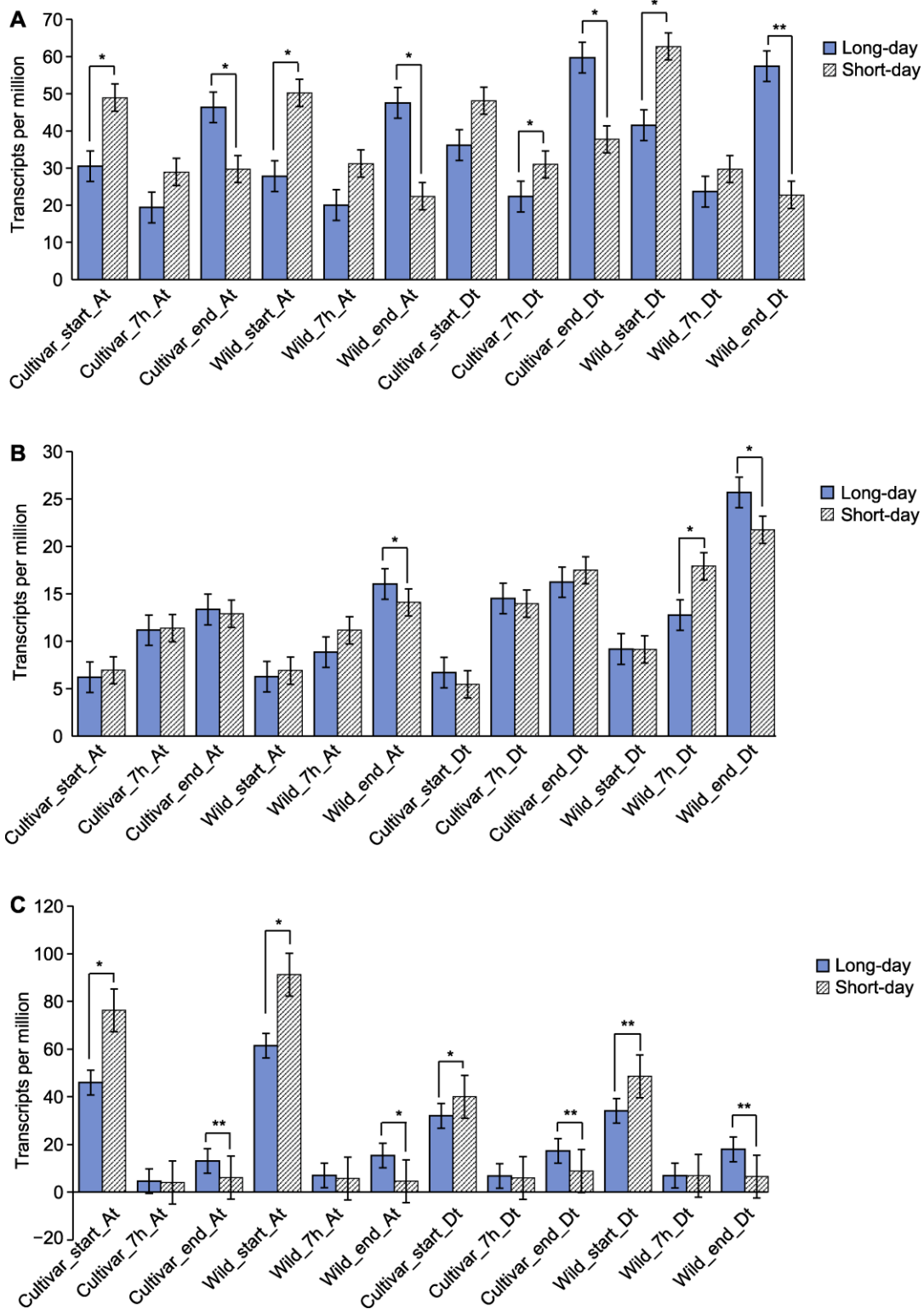


图8 陆地棉*PHY*基因在不同光照处理条件下的表达

(A) *PHYA1*基因的表达; (B) *PHYB1*基因的表达; (C) *PHYE1*基因的表达。* $P<0.05$; ** $P<0.01$

Figure 8 Expression of *PHY* genes in *Gossypium hirsutum* under different light treatment conditions

(A) Expression of *PHYA1* genes; (B) Expression of *PHYB1* genes; (C) Expression of *PHYE1* genes. * $P<0.05$; ** $P<0.01$

生种在光照处理结束时,长日照条件下 *GhPHYB1At* 和 *GhPHYB1Dt* 的表达量显著高于短日照,呈现较为明显的栽培品种与野生种之间的差异,说明 *PHYB* 在栽培品种与野生种光周期敏感性驯化过程中发挥重要作用(图8)。

2.10.3 陆地棉 *PHY* 基因在栽培品种与野生种之间的表达差异

对比不同光照处理时间下栽培品种与野生种 *PHY* 基因的表达情况,结果显示 *PHY* 基因表达的差异主要集中在暗期结束和光照处理结束。在栽培品种与野生种中, *GhPHYA1At* 在相同处理时间下的表达差异均不明显,而在短日照暗期结束时,栽培品种中 *GhPHYA1Dt* 的表达量显著低于野生种;同样, *GhPHYB1At* 在栽培品种与野生种中的表达量在相同处理时间下均无显著差异,栽培品种在长日照处理14小时和短日照暗期结束时 *GhPHYB1Dt* 的表达量极显著或显著低于野生种。综上,栽培品种与野生种中 *GhPHYA1At*、*GhPHYA1Dt*、*GhPHYB1At* 和 *GhPHYB1Dt* 均表现明显的Dt亚基因组表达偏好性,说明驯化选择可能更多发生在Dt亚基因组(图9)。

在长、短日照条件下, *PHYC* 和 *PHYE* 亚家族存在显著差异。在长日照条件下,处理7小时的栽培品种 *GhPHYC1Dt* 的表达量显著高于野生种;处理14小时后栽培品种 *GhPHYC1At* 和 *GhPHYE1At* 的表达量显著低于野生种。在短日照条件下,处理10小时的栽培品种 *GhPHYE1Dt* 的表达量显著高于野生种。值得注意的是, *GhPHYC1Dt* 与其它基因的表达规律不同,在各处理条件下,栽培品种 *GhPHYC1Dt* 的表达量均高于野生种,尤其在长日照处理7小时后(附图9)。

综上,在从野生种到栽培品种的驯化过程中, *GhPHYA1Dt* 和 *GhPHYB1Dt* 的表达量在长、短日照下差异显著,说明在光周期驯化过程中,这2个基因可能受到较强烈的驯化选择。 *GhPHYC1At* 和 *GhPHYE1At* 在长日照处理后期栽培品种中的表达量显著低于野生种,说明在经历长日照后,这2个基因低表达可能是光周期驯化的结果。

2.11 讨论与结论

约5 000年前,四倍体棉种陆地棉(AD₁)和海岛棉(AD₂) 在低纬度的墨西哥尤卡坦半岛和秘鲁西海岸地

区分别被人类驯化选择(Fang et al., 2017a; Yuan et al., 2021),之后从中、南美洲扩散至全球,成为世界主要的经济作物。在驯化过程中,棉花的植株形态、纤维产量及品质等重要农艺性状均发生改变,其对光周期的敏感性逐渐丧失,这使其能够在中高纬度的长日照环境下被广泛种植。研究调控棉花开花相关基因,尤其是光周期相关基因对于揭示棉花驯化的遗传机制具有重要意义。当前我国种植的棉花品种主要是早期从美国等国家引种后加以遗传改良(Westengen et al., 2005; Doebley et al., 2006)。目前,品种退化和遗传资源多样性狭窄等问题是棉花育种的瓶颈。因此,科学家从野生种、半野生种及农家种等材料中挖掘抗病、抗旱及调控开花的优良基因资源,以期进一步创造新种质、新材料,改良现有品系或品种的产量性状缺陷,进而培育出高产优质多抗的棉花新品种。然而,棉花野生或半野生种对光周期敏感,在长日照条件下较难开花。已有报道表明, *PHY* 基因通过光周期调控植物开花(Reed et al., 1993; Johnson et al., 1994; 惠婕等, 2011; 詹克慧等, 2012; Pearce et al., 2016)。棉花 *PHY* 基因的相关研究较缺乏,光敏色素是否调控棉花开花值得探讨。

拟南芥 *PHY* 基因亚家族分为A、B、C、D和E五类。本研究利用拟南芥及锦葵科多物种基因组数据,对 *PHY* 基因家族进行鉴定和分析,发现大部分锦葵科物种中仅含A、B、C和E四类 *PHY* 亚家族基因。研究表明,被子植物中具备 *PHYA*、*PHYB*、*PHYC*、*PHYD* 四个亚家族,如油菜(*Brassica napus*)、水稻、高粱(*Sorghum bicolor*)、玉米、谷子(*Setaria italica*) (申慧敏等, 2022)、小麦(李壮等, 2010; Pearce et al., 2016)。生物信息学分析表明,四倍体棉种和木槿的 *PHY* 基因数目是二倍体棉种及叉柱棉的2倍,说明随着锦葵科物种的进化, *PHY* 基因数目因物种染色体倍性而发生变化。此外,草棉(A₁)、亚洲棉(A₂)、雷蒙德氏棉(D₅)、陆地棉(AD₁)及海岛棉(AD₂) 中 *PHY* 基因均位于基因组或亚基因组的10、11和13号染色体上共线性位置,且基序与结构十分保守,说明 *PHY* 基因在不同棉种中可能具有相似的功能。

研究表明, *PHY* 是调控植物合成光敏色素、感受光信号和完成光周期途径的重要基因家族。在植物感受光信号的过程中,光敏色素在种子萌发、植株生长和开花等过程中扮演重要角色,是植物接受光信号的

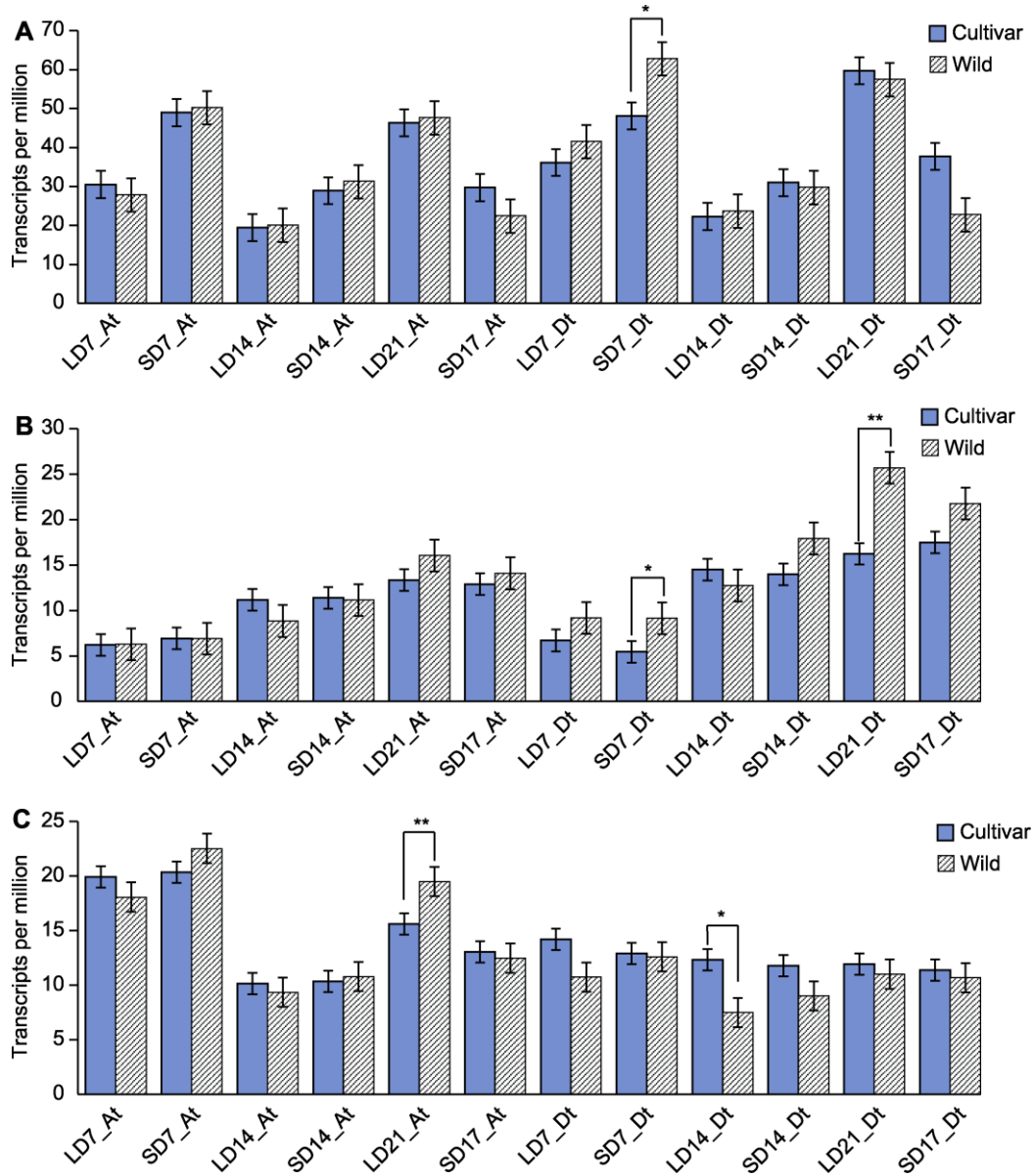


图9 陆地棉*PHY*基因在栽培品种与野生种中的表达

(A) *PHYA1*基因的表达; (B) *PHYB1*基因的表达; (C) *PHYC1*基因的表达。* $P < 0.05$; ** $P < 0.01$

Figure 9 Expression of *PHY* genes in cultivated and wild *Gossypium hirsutum*

(A) Expression of *PHYA1* gene; (B) Expression of *PHYB1* gene; (C) Expression of *PHYC1* gene. * $P < 0.05$; ** $P < 0.01$

重要受体(Hughes, 2013)。本研究通过对上述5个棉种的*PHY*基因顺式作用元件进行分析,发现*PHY*基因上游含有较多的光反应元件,推测有些*PHY*基因可能参与调控光周期敏感性。

对陆地棉(AD_1)中*PHY*基因的驯化选择及遗传分化研究表明,*GhPHYB1At*与*GhPHYA2Dt*在早期驯化过程中受到较强的驯化选择,而*GhPHYB1Dt*与*GhPH-*

*YC1Dt*在遗传改良过程中受到人工选择影响较大。本研究对*PHY*基因家族成员的表达分析表明,在不同日照条件下,*GhPHYA1Dt*和*GhPHYB1Dt*的表达量差异显著。研究表明,小麦中缺失*PHYB*的植株表现不同程度的延迟开花(Pearce et al., 2016)。棉花*GhPHYB1Dt*在长日照条件下的表达量高于短日照,我们推测在驯化选择作用下,由短日照作物演变为长

日照作物的过程中, 该基因的表达量增高, 从而促进棉花在长日照条件下开花, 并且这种对光周期敏感性的变化可能与该基因不同单倍型的比例及种类变化相关。对于野生种和栽培品种而言, *GhPHYC1At*和*GhPHYE1At*在长日照处理14小时后, 栽培品种中的表达量显著低于野生种。研究表明, 玉米*zmphyC1/zmphyC2*双突变体在长日照条件下促进开花, 而在短日照下无影响(李全权, 2020)。我们推测在驯化过程中, 棉花栽培品种中*GhPHYC1At*的低表达使其光周期敏感性降低, 促进棉花在长日照条件下开花, 这可能与栽培品种中单倍型1的比例升高及单倍型2的比例降低有关。综上, 我们推测*GhPHYB1Dt*在遗传改良过程中受到了较强的驯化选择, 导致该基因对光照时间的敏感性发生改变, 使其在不同日照长度下, 在野生种与栽培品种中的表达量有所不同, 因此我们认为其在陆地棉(AD₁)的光周期驯化选择及传播中具有正向作用。

本研究确定了锦葵科不同物种中*PHY*基因的数目, 结合系统发育分析将锦葵科*PHY*基因分为*PHYA*、*PHYB*、*PHYC*与*PHYE*四个亚家族, 并对陆地棉(AD₁)进行了驯化选择分析以及转录组分析, 明确了*GhPHYB1Dt*在光周期驯化过程中的重要作用, 发掘了棉花光周期敏感性候选基因。

作者贡献声明

顾家琦: 构思实验、分析数据及撰写论文; 朱福慧: 构思实验和分析数据; 谢沛豪: 分析数据; 孟庆营: 审稿、分析数据并提供技术支持; 郑颖: 分析数据; 张献龙: 论文修改润色; 袁道军: 构思实验、监督、审稿及论文修改。

参考文献

- 丁武思, 陈士瞻, 刘磊, 樊晓聪, 丁梦月, 孙广华, 王立建, 杨建平 (2021). 2个玉米光敏色素C基因的克隆及功能验证. 河南农业科学 50, 16–26.
- 房迈苑, 王小菁, 李洪清 (2005). 光对植物生物钟的调节. 植物学通报 22, 207–214.
- 惠婕, 黄丛林, 吴忠义, 张秀海 (2011). 拟南芥光敏色素基因*PHYA*转化菊花的研究. 江苏农业科学 39(2), 51–54.
- 李建平 (2013). 光敏色素D在拟南芥生长和发育中的功能与机理研究. 博士学位论文. 北京: 中国农业科学院. pp. 1–75.

- 李建平, 杨建平, 宋梅芳, 苏亮, 侯佩, 郝晓燕, 邵琳, 陈果, 足木热木, 黄全生 (2014). 拟南芥光敏色素D (*AtphyD*)在不同光质下的表达特性分析. 新疆农业科学 51, 305–310.
- 李全权 (2020). 玉米光敏色素C在光形态建成与开花中的功能解析. 博士学位论文. 泰安: 山东农业大学. pp. 1–131.
- 李壮, 马燕斌, 蔡应繁, 吴锁伟, 肖阳, 孟凡华, 付风铃, 黄玉碧, 杨建平 (2010). 小麦光敏色素基因*TaPhyB3*的克隆和表达分析. 作物学报 36, 779–787.
- 连欣 (2020). 棉花开花时间相关基因的挖掘与功能分析. 硕士学位论文. 泰安: 山东农业大学. pp. 1–78.
- 申慧敏, 吴年隆, 王亚敏, 王兴春 (2022). 谷子光敏色素家族基因的生物信息学及表达模式分析. 山西农业科学 50, 1–8.
- 魏俊梅 (2018). 棉花开花时间相关基因的挖掘鉴定与功能分析. 硕士学位论文. 泰安: 山东农业大学. pp. 1–57.
- 岳晶, 管利萍, 孟思远, 张静, 侯岁稳 (2015). 光敏色素信号通路中磷酸化修饰研究进展. 植物学报 50, 241–254.
- 詹克慧, 李志勇, 侯佩, 习雨琳, 肖阳, 孟凡华, 杨建平 (2012). 利用修饰光敏色素信号途径进行作物改良的可行性. 中国农业科学 45, 3249–3255.
- 赵翔, 赵青平, 杨煦, 慕世超, 张骁 (2015). 向光素调节植物向光性及其与光敏色素/隐花色素的相互关系. 植物学报 50, 122–132.
- Abdellatif KF, Khidr YA, El-Mansy YM, El-Lawendey MM, Soliman YA (2012). Molecular diversity of Egyptian cotton (*Gossypium barbadense* L.) and its relation to varietal development. *J Crop Sci Biotechnol* 15, 93–99.
- Argout X, Martin G, Droc G, Fouet O, Labadie K, Rivals E, Aury JM, Lanaud C (2017). The cacao Criollo genome v2.0: an improved version of the genome for genetic and functional genomic studies. *BMC Genomics* 18, 730.
- Ballaré CL, Pierik R (2017). The shade-avoidance syndrome: multiple signals and ecological consequences. *Plant Cell Environ* 40, 2530–2543.
- Bao Y, Hu GJ, Grover CE, Conover J, Yuan DJ, Wendel JF (2019). Unraveling *cis* and *trans* regulatory evolution during cotton domestication. *Nat Commun* 10, 5399.
- Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Briggs WR, Olney MA (2001). Photoreceptors in plant photomorphogenesis to date. Five phytochromes, two cryptochromes, one phototropin, and one superchrome. *Plant Physiol* 125, 85–88.
- Cai YF, Cai XY, Wang QL, Wang P, Zhang Y, Cai CW, Xu YC, Wang KB, Zhou ZL, Wang CX, Geng SP, Li B,

- Dong Q, Hou YQ, Wang H, Ai P, Liu Z, Yi FF, Sun MS, An GY, Cheng JR, Zhang YY, Shi Q, Xie YH, Shi XY, Chang Y, Huang FF, Chen Y, Hong SM, Mi LY, Sun Q, Zhang L, Zhou BL, Peng R, Zhang X, Liu F (2020). Genome sequencing of the Australian wild diploid species *Gossypium australe* highlights disease resistance and delayed gland morphogenesis. *Plant Biotechnol J* **18**, 814–828.
- Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, He YH, Xia R (2020a). TBtools: an integrative Toolkit developed for interactive analyses of big biological data. *Mol Plant* **13**, 1194–1202.
- Chen M, Chory J (2011). Phytochrome signaling mechanisms and the control of plant development. *Trends Cell Biol* **21**, 664–671.
- Chen ZJ, Sreedasyam A, Ando A, Song QX, De Santiago LM, Hulse-Kemp AM, Ding MQ, Ye WX, Kirkbride RC, Jenkins J, Plott C, Lovell J, Lin YM, Vaughn R, Liu B, Simpson S, Scheffler BE, Wen L, Saski CA, Grover CE, Hu GJ, Conover JL, Carlson JW, Shu SQ, Boston LB, Williams M, Peterson DG, McGee K, Jones DC, Wendel JF, Stelly DM, Grimwood J, Schmutz J (2020b). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat Genet* **52**, 525–533.
- Chun L, Kawakami A, Christopher DA (2001). Phytochrome A mediates blue light and UV-A-dependent chloroplast gene transcription in green leaves. *Plant Physiol* **125**, 1957–1966.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158.
- Doebly JF, Gaut BS, Smith BD (2006). The molecular genetics of crop domestication. *Cell* **127**, 1309–1321.
- Fang L, Gong H, Hu Y, Liu CX, Zhou BL, Huang T, Wang YK, Chen SQ, Fang DD, Du XM, Chen H, Chen JD, Wang S, Wang Q, Wan Q, Liu BL, Pan MQ, Chang LJ, Wu HT, Mei GF, Xiang D, Li XH, Cai CP, Zhu XF, Chen ZJ, Han B, Chen XY, Guo WZ, Zhang TZ, Huang XH (2017a). Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. *Genome Biol* **18**, 33.
- Fang L, Wang Q, Hu Y, Jia YH, Chen JD, Liu BL, Zhang ZY, Guan XY, Chen SQ, Zhou BL, Mei GF, Sun JL, Pan ZE, He SP, Xiao SH, Shi WJ, Gong WF, Liu JG, Ma J, Cai CP, Zhu XF, Guo WZ, Du XM, Zhang TZ (2017b). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet* **49**, 1089–1098.
- Grover CE, Arick II MA, Thrash A, Conover JL, Sanders WS, Peterson DG, Frelichowski JE, Scheffler JA, Scheffler BE, Wendel JF (2019). Insights into the evolution of the new world diploid cottons (*Gossypium*, Subgenus *Houzingenia*) based on genome sequencing. *Genome Biol Evol* **11**, 53–71.
- Grover CE, Pan MQ, Yuan DJ, Arick MA, Hu GJ, Brase L, Stelly DM, Lu ZF, Schmitz RJ, Peterson DG, Wendel JF, Udall JA (2020). The *Gossypium longicalyx* genome as a resource for cotton breeding and evolution. *G3* **10**, 1457–1467.
- Hu Y, Chen JD, Fang L, Zhang ZY, Ma W, Niu YC, Ju LZ, Deng JQ, Zhao T, Lian JM, Baruch K, Fang D, Liu X, Ruan YL, Rahman MU, Han JL, Wang K, Wang Q, Wu HT, Mei GF, Zang YH, Han ZG, Xu CY, Shen WJ, Yang DF, Si ZF, Dai F, Zou LF, Huang F, Bai YL, Zhang YG, Brodt A, Ben-Hamo H, Zhu XF, Zhou BL, Guan XY, Zhu SJ, Chen XY, Zhang TZ (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet* **51**, 739–748.
- Huang G, Wu ZG, Percy RG, Bai MZ, Li Y, Frelichowski JE, Hu J, Wang K, Yu JZ, Zhu YX (2020). Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat Genet* **52**, 516–524.
- Hughes J (2013). Phytochrome cytoplasmic signaling. *Annu Rev Plant Biol* **64**, 377–402.
- Johnson E, Bradley M, Harberd NP, Whitelam GC (1994). Photoresponses of light-grown phyA mutants of *Arabidopsis* (Phytochrome A is required for the perception of daylength extensions). *Plant Physiol* **105**, 141–149.
- Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915.
- Kim YM, Kim S, Koo N, Shin AY, Yeom SI, Seo E, Park SJ, Kang WH, Kim MS, Park J, Jang I, Kim PG, Byeon I, Kim MS, Choi J, Ko G, Hwang J, Yang TJ, Choi SB, Lee JM, Lim KB, Lee J, Choi IY, Park BS, Kwon SY,

- Choi D, Kim RW** (2017). Genome analysis of *Hibiscus syriacus* provides insights of polyploidization and indeterminate flowering in woody plants. *DNA Res* **24**, 71–80.
- Korunes KL, Samuk K** (2021). Pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Res* **21**, 1359–1368.
- Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S** (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res* **30**, 325–327.
- Li JG, Li G, Wang HY, Deng X** (2011). Phytochrome signaling mechanisms. *Arabidopsis Book* **9**, e0148.
- Liao Y, Smyth GK, Shi W** (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.
- Ma ZY, He SP, Wang XF, Sun JL, Zhang Y, Zhang GY, Wu LQ, Li ZK, Liu ZH, Sun GF, Yan YY, Jia YH, Yang J, Pan ZE, Gu QS, Li XY, Sun ZW, Dai PH, Liu ZW, Gong WF, Wu JH, Wang M, Liu HW, Feng KY, Ke HF, Wang JD, Lan HY, Wang GN, Peng J, Wang N, Wang LR, Pang BY, Peng Z, Li RQ, Tian SL, Du XM** (2018). Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat Genet* **50**, 803–813.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ** (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268–274.
- Pearce S, Kippes N, Chen A, Debernardi JM, Dubcovsky J** (2016). RNA-seq studies using wheat *PHYTOCHROME B* and *PHYTOCHROME C* mutants reveal shared and specific functions in the regulation of flowering and shade-avoidance pathways. *BMC Plant Biol* **16**, 141.
- Peng H, Phung J, Zhai Y, Neff MM** (2020). Self-transcriptional repression of the *Arabidopsis* NAC transcription factor ATAF2 and its genetic interaction with phytochrome A in modulating seedling photomorphogenesis. *Planta* **252**, 48.
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL** (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295.
- Procter JB, Carstairs GM, Soares B, Mourão K, Ofoegbu TC, Barton D, Lui L, Menard A, Sherstnev N, Roldan-Martinez D, Duce S, Martin DMA, Barton GJ** (2021). Alignment of biological sequences with Jalview. In: Katoh K, ed. *Multiple Sequence Alignment: Methods and Protocols*. New York: Humana. pp. 203–224.
- Purugganan MD, Fuller DQ** (2009). The nature of selection during plant domestication. *Nature* **457**, 843–848.
- Rao AQ, Irfan M, Saleem Z, Nasir IA, Riazuddin S, Husnain T** (2011). Overexpression of the phytochrome B gene from *Arabidopsis thaliana* increases plant growth and yield of cotton (*Gossypium hirsutum*). *J Zhejiang Univ Sci B Biomed Biotechnol* **12**, 326–334.
- Reed JW, Nagpal P, Poole DS, Furuya M, Chory J** (1993). Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development. *Plant Cell* **5**, 147–157.
- Saski CA, Scheffler BE, Hulse-Kemp AM, Liu B, Song QX, Ando A, Stelly DM, Scheffler JA, Grimwood J, Jones DC, Peterson DG, Schmutz J, Chen ZJ** (2017). Sub genome anchored physical frameworks of the allotetraploid Upland cotton (*Gossypium hirsutum* L.) genome, and an approach toward reference-grade assemblies of polyploids. *Sci Rep* **7**, 15274.
- Tian Y, Zhang TZ** (2021). MIXTAs and phytohormones orchestrate cotton fiber development. *Curr Opin Plant Biol* **59**, 101975.
- Udall JA, Long E, Hanson C, Yuan DJ, Ramaraj T, Conover JL, Gong L, Arick MA, Grover CE, Peterson DG, Wendel JF** (2019a). *De novo* genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri*. *G3* **9**, 3079–3085.
- Udall JA, Long E, Ramaraj T, Conover JL, Yuan DJ, Grover CE, Gong L, Arick II MA, Masonbrink RE, Peterson DG, Wendel JF** (2019b). The genome sequence of *Gossypoides kirkii* illustrates a descending dysploidy in plants. *Front Plant Sci* **10**, 1541.
- Wang MJ, Li JY, Wang PC, Liu F, Liu ZP, Zhao GN, Xu ZP, Pei LL, Grover CE, Wendel JF, Wang KB, Zhang XL** (2021). Comparative genome analyses highlight transposon-mediated genome expansion and the evolutionary architecture of 3D genomic folding in cotton. *Mol Biol Evol* **38**, 3621–3636.
- Wang MJ, Tu LL, Yuan DJ, Zhu D, Shen C, Li JY, Liu FY, Pei LL, Wang PC, Zhao GN, Ye ZX, Huang H, Yan FL, Ma YZ, Zhang L, Liu M, You JQ, Yang YC, Liu ZP, Huang F, Li BQ, Qiu P, Zhang QH, Zhu LF, Jin SX, Yang XY, Min L, Li GL, Chen LL, Zheng HK, Lindsey K, Lin ZX, Udall JA, Zhang XL** (2019). Reference genome sequences of two cultivated allotetraploid cottons, *Goss-*

- ypium hirsutum* and *Gossypium barbadense*. *Nat Genet* **51**, 224–229.
- Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin HZ, Marler B, Guo H, Kissinger JC, Paterson AH** (2012). *MCSanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49.
- Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM** (2009). Evolution and natural history of the cotton genus. In: Paterson AH, ed. *Genetics and Genomics of Cotton*. New York: Springer. pp. 3–22.
- Wendel JF, Cronn RC** (2003). Polyploidy and the evolutionary history of cotton. *Adv Agron* **78**, 139–186.
- Westengen OT, Huamán Z, Heun M** (2005). Genetic diversity and geographic pattern in early South American cotton domestication. *Theor Appl Genet* **110**, 392–402.
- Xie YR, Liu Y, Wang H, Ma XJ, Wang BB, Wu GX, Wang HY** (2017). Phytochrome-interacting factors directly suppress *MIR156* expression to enhance shade-avoidance syndrome in *Arabidopsis*. *Nat Commun* **8**, 348.
- Yang ZE, Ge XY, Li WN, Jin YY, Liu LS, Hu W, Liu FY, Chen YL, Peng SL, Li FG** (2021). Cotton D genome assemblies built with long-read data unveil mechanisms of centromere evolution and stress tolerance divergence. *BMC Biol* **19**, 115.
- Yang ZR, Qanmber G, Wang Z, Yang ZE, Li FG** (2020). *Gossypium* genomics: trends, scope, and utilization for cotton improvement. *Trends Plant Sci* **25**, 488–500.
- Yuan DJ, Grover CE, Hu GJ, Pan MQ, Miller ER, Conover JL, Hunt SP, Udall JA, Wendel JF** (2021). Parallel and intertwining threads of domestication in allopolyploid cotton. *Adv Sci (Weinh)* **8**, 2003634.
- Zhang HK, Gao SH, Lercher MJ, Hu SN, Chen WH** (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res* **40**, W569–W572.
- Zhang LL, Ma XK, Zhang XT, Xu Y, Ibrahim AK, Yao JY, Huang HX, Chen S, Liao ZY, Zhang Q, Niyitanga S, Yu JX, Liu Y, Xu XM, Wang JJ, Tao AF, Xu JT, Chen SY, Yang X, He QY, Lin LH, Fang PP, Zhang LM, Ming R, Qi JM, Zhang LW** (2021). Reference genomes of the two cultivated jute species. *Plant Biotechnol J* **19**, 2235–2248.
- Zhang M, Zheng XL, Song SQ, Zeng QW, Hou L, Li DM, Zhao J, Wei Y, Li XB, Luo M, Xiao YH, Luo XY, Zhang JF, Xiang CB, Pei Y** (2011). Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. *Nat Biotechnol* **29**, 453–458.
- Zhang TZ, Hu Y, Jiang WK, Fang L, Guan XY, Chen JD, Zhang JB, Saski CA, Scheffler BE, Stelly DM, Hulse-Kemp AM, Wan Q, Liu BL, Liu CX, Wang S, Pan MQ, Wang YK, Wang DW, Ye WX, Chang LJ, Zhang WP, Song QX, Kirkbride RC, Chen XY, Dennis E, Llewellyn DJ, Peterson DG, Thaxton P, Jones DC, Wang Q, Xu XY, Zhang H, Wu HT, Zhou L, Mei GF, Chen SQ, Tian Y, Xiang D, Li XH, Ding J, Zuo QY, Tao LN, Liu YC, Li J, Lin Y, Hui YY, Cao ZS, Cai CP, Zhu XF, Jiang Z, Zhou BL, Guo WZ, Li RQ, Chen ZJ** (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol* **33**, 531–537.

Genome-wide Identification and Domestication Analysis of the Phytochrome *PHY* Gene Family in *Gossypium*

Jiaqi Gu^{1, 2†}, Fuhui Zhu^{1, 2†}, Peihao Xie^{1, 2}, Qingying Meng^{1, 2}, Ying Zheng²
Xianlong Zhang^{1, 2}, Daojun Yuan^{1, 2, 3*}

¹National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China; ²College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China; ³College of Agronomy, Xinjiang Agricultural University, Urumqi 830052, China

Abstract Phytochrome is an important receptor for red and far-red light sensing in plants, and it plays a vital role in regulating the plant flowering period, improving crop yield potential and regulating plant stress resistance. Identification of *PHY* family genes in *Gossypium*, exploration of the patterns of inheritance and regulatory network of domestication and improvement, and identification of the key phytochrome genes in *Gossypium*, provides insights into the *de novo* domestication and breeding of early maturing *Gossypium* species. To identify the phytochrome genes of *Gossypium*, we used bioinformatics methods to analyze 5 phytochrome genes in *Arabidopsis thaliana*. Phylogenetic analysis showed that the *PHY* genes in Malvaceae species consisted of 4 subfamilies (*PHYA*, *PHYB*, *PHYC* and *PHYE*). Moreover, the domestication selection analysis of *PHY* genes among different populations of *G. hirsutum* showed that the domestication process of *PHY* genes could be divided into two stages: domestication and improvement. Furthermore, the gene expression of the *PHY* gene family was analyzed using leaf RNA-sequencing data obtained from wild and cultivar genotypes of *G. hirsutum* under short-day (SD) and long-day (LD) conditions. The results showed that the expression of *GhPHYA1Dt* and *GhPHYB1Dt* were significantly different between SD and LD conditions. After 14 hours of long-day treatment, the expression of *GhPHYC1At* and *GHPHYE1At* in the cultivar was significantly lower than that in wild species. These results lay a foundation for further study on domestication selection and functional mechanisms of *Gossypium PHY* genes and provide a theoretical basis for breeding new early maturing *Gossypium* varieties and *de novo* domestication.

Key words cotton, domestication, phytochrome, gene family

Gu JQ, Zhu FH, Xie PH, Meng QY, Zheng Y, Zhang XL, Yuan DJ (2024). Genome-wide identification and domestication analysis of the phytochrome *PHY* gene family in *Gossypium*. *Chin Bull Bot* **59**, 34–53.

† These authors contributed equally to this paper

* Author for correspondence. E-mail: robert@mail.hzau.edu.cn

(责任编辑: 白羽红)

附表1 数据来源

Appendix table 1 Data sources

附表2 *PHY*基因家族成员及其理化性质

Appendix table 2 The members of *PHY* gene family and their physical and chemical properties

附表3 陆地棉*PHY*重复基因的选择压力分析

Appendix table 3 Selection pressure analysis of *PHY* duplicate genes in *Gossypium hirsutum*

附表4 陆地棉*PHY*基因在不同群体中的核苷酸多样性($Pi \times 10^4$)

Appendix table 4 Nucleotide diversity ($Pi \times 10^4$) of *Gossypium hirsutum PHY* genes among different populations

附表5 陆地棉叶片转录组测序数据统计

Appendix table 5 Statistics of *Gossypium hirsutum* leaf transcriptome sequencing data

附录1 Perl脚本

Appendix 1 Perl script

附图1 *GhePHYA1At* (*Ghe11G33390*)、*GhePHYA2At* (*Ghe13G23940*)、*GaPHYB1At* (*Gar10G22470*)、*GaPHYA1At* (*Gar11G33210*)和*GaPHYA2At* (*Gar13G23780*)基因矫正前后对比

Appendix figure 1 Comparison diagram of *GhePHYA1At* (*Ghe11G33390*), *GhePHYA2At* (*Ghe13G23940*), *GaPHYB1At* (*Gar10G22470*), *GaPHYA1At* (*Gar11G33210*), and *GaPHYA2At* (*Gar13G23780*) before and after correction

附图2 草棉(*A*₁)、亚洲棉(*A*₂)、雷蒙德氏棉(*D*₅)、陆地棉(*AD*₁)及海岛棉(*AD*₂)的*PHY*基因编码蛋白多序列比对

Appendix figure 2 Multiple sequence alignment of protein encoded by the *PHY* genes in *Gossypium herbaceum* (*A*₁), *G. arboreum* (*A*₂), *G. raimondii* (*D*₅), *G. hirsutum* (*AD*₁) and *G. barbadense* (*AD*₂)

附图3 二倍体及异源四倍体棉花*PHY*基因染色体定位

Appendix figure 3 Chromosome mapping of *PHY* genes in diploid and allotetraploid cotton

附图4 *PHY*家族基因在亚洲棉*A*基因组、陆地棉*AD*亚基因组以及雷蒙德氏棉*D*基因组之间的共线性关系

Appendix figure 4 Collinearity between *PHY* family genes in *Gossypium arboreum* *A* genome, *G. hirsutum* *AD* subgenome, and *G. raimondii* *D* genome

附图5 棉花*PHY*基因启动子顺式作用元件

Appendix figure 5 *Cis*-acting element of cotton *PHY* gene promoters

附图6 *PHY*基因在野生种、半野生种及栽培品种中的单倍型分析

Appendix figure 6 Haplotype analysis of *PHY* genes in wild, landrace and cultivar

附图7 陆地棉*PHY*基因在不同亚基因组中的表达

Appendix figure 7 Expression of *PHY* genes in different subgenomes of *Gossypium hirsutum*

附图8 陆地棉*PHY*基因在不同光照处理条件下的表达

Appendix figure 8 Expression of *PHY* genes in *Gossypium hirsutum* under different light treatment conditions

附图9 陆地棉*PHY*基因在栽培品种与野生种中的表达

Appendix figure 9 Expression of *PHY* genes in cultivated and wild *Gossypium hirsutum*

<https://www.chinbullbotany.com/fileup/1674-3466/PDF/23-004-1.pdf>