薛天源,何钰晴,夏忙,等.青蒿查尔酮合成酶(*AaCHS*)基因家族鉴定及光调控分析[J].山西农业大学学报(自然科学版),2024,44(1):1-13.
Xue T Y, He Y Q, Xia M, et al. Identification and photoregulation analysis of Artemisinin chalcone synthetase (*AaCHS*) gene family and analysis of light regulation in *Artemisia annua* L.[J].Journal of Shanxi Agricultural University(Natural Science Edition), 2024,44(1):1-13.
doi: 10.13842/j.cnki.issn1671-8151.202311021

青蒿查尔酮合成酶(*AaCHS*)基因家族鉴定及光调控分析

薛天源,何钰晴,夏忙,陈美珠,戴希刚,何思晓,朱洵,曾长立*

(江汉大学 湖北省汉江流域特色生物资源保护开发与利用工程技术研究中心,湖北 武汉 430056)

摘　要:[目的]为了解青蒿(*Artemisia annua* L.)查尔酮合成酶(chalcone synthase,CHS)基因家族的分子进化特性及其在不同组织部位的表达情况。[方法]从Pfam数据库下载CHS蛋白HMM模型,并使用HUMMER 3.0、Blastp和CD-search鉴定出青蒿基因组中的CHS基因家族成员。采用TBtools、EXPASy、CELLO v. 2.5、MEGA 7.0、MEME、SOPMA、SWISSMODEL和PlantCARE等软件对其氨基酸与碱基序列进行生物信息学分析,通过8个不同组织部位的转录组数据对*AaCHS*基因家族成员进行表达分析。[结果]青蒿基因组中共有16个*AaCHS*基因家族成员,蛋白结构分析显示AaCHS蛋白质结构并不稳定,将其分为4个组,基因结构分析显示所有AaCHS基因蛋白均具有外显子,数量为2~4,内含子数量为0~2,其中8个AaCHS基因不具有内含子。启动子顺式作用元件分析表明该基因家族成员启动子上含有大量光响应、植物激素响应元件;CD-search验证结果发现每个AaCHS蛋白均有Chal-sti-synt_N结构域。[结论]*AaCHS*家族成员表达分析结果表明,16个*AaCHS*基因在不同组织和光处理下表达不同。GO富集结果得到GO二级分类的生物过程和分子功能,生物过程共富集了80条序列;细胞组分共富集了19条序列,其中二级分类过程细胞质中富集了11条序列,数量最多,与亚细胞定位预测结果一致;推测AaCHS基因在细胞质中调控黄酮类化合物的累积。通过青蒿光调控差异表达分析,推测*AaCHS7*、AaCHS10在青蒿受到红光光处理中起正向调控作用。

关键词:青蒿;*CHS*基因家族;光调控;基因表达;差异基因表达
中图分类号:R932;S567.219　　**文献标识码**:A　　**文章编号**:1671-8151(2024)01-0001-13

青蒿(*Artemisia annua* L.)是菊科的成员,具有相当大的药用、观赏和经济价值[1]。青蒿素以青蒿为提取来源,是目前治疗疟疾的最有效的药物。疟疾是由原生动物寄生虫疟原虫引起的毁灭性疾病[2]。根据世界卫生组织2015年的数据,约有33亿人面临感染疟疾的风险,12亿人处于高风险[3]。有研究表明青蒿素和黄酮类化合物在治疗包括疟疾在内的各种疾病方面具有协同作用[4-5]。黄酮类化合物作为植物抗毒素或抗氧化剂,具有清除活性氧(ROS)的能力[6],并保护植物免受非生物和生物胁迫的损害,包括UV照射、冷胁迫、病原体感染和昆虫取食[7-9]。许多青蒿素和类黄酮生物合成途径的酶的增加,通常与产物丰度增加相关联[10]。但在青蒿中对于它们的生物合成途径及调控的研究报导还比较少,所以探究青蒿中黄酮化合物合成途径关键基因及其表达具有重要意义。

查尔酮合成酶(CHS)是植物聚酮合酶超家族的成员,其还包括芪合酶(STS)、吖啶酮合酶、吡喃酮合酶、联苄基合酶和对香豆酰三乙酸合酶[11]。黄酮类化合物在植物防御中起着重要的作用,而*CHS*作为黄酮类化合物生物合成的关键酶基因,在调节该途径中起着重要的作用。此外,查尔酮合成酶为在开花植物中具有重要作用的黄酮类化合物提供起始材料,如提供花色素、抗生素、UV保护剂和驱虫剂[12]。由于*CHS*在类黄酮生产中的重要性,在植物中已经得到了很好的研究,如矮牵牛

收稿日期:2023-11-15　　**修回日期**:2023-12-07
基金项目:湖北省重点研发计划项目(2021BBA097;2022BBA0064);江汉大学一流学科建设重大专项资助计划(2023XKZ026)
作者简介:薛天源(ORCID: 0000-0003-2127-6754),硕士研究生,研究方向:特色植物资源开发与利用,E-mail:xty99100911@163.com
*通信作者:曾长立(ORCID: 0000-0001-6947-3410),教授,博士,博士生导师,研究方向:植物逆境生理、珍稀濒危植物保护与利用,E-mail:zengchangli@jhun.edu.cn

(*Petunia hybrida*)*CHS*基因家族的每个成员由2个外显子组成,由1个大小和序列可变的内含子分隔,内含子位于保守位置[13];玉米中的2个*CHS*基因,具有高度同源的外显子序列,在编码区上游5′和下游3′序列以及内含子中有很大差异[14];菊花(*Chrysanthemum nankingense*)中鉴定出16个*CnCHS*基因家族成员,大多数含有2个外显子和1个内含子,具有Chal-sti-synt_N和Chal-sti-synt_C结构域,其中*CnCHS10*在花器官中的表达水平高于营养组织[15]。芒果(*Mangifera indica* L.)中*MiCHS4*、*MiCHS16*和*MiCHS17*响应UV-B和可见光处理[16]。

光作为一个重要的环境调控因子,不仅是植物光合作用的主要能量来源,同时也在调节植物生长发育过程中发挥关键作用[17]。有研究表明不同光质处理可使*AaCHS*表达上调,刺葡萄(*Vitis davidii*)中发现*VdCHS2*和*VdCHS3*在绿光、蓝光、暖黄光处理下表达量先升高后降低[18];非洲菊(*Gerbera hybrida*)中发现蓝光和红光促进*CHS*基因表达[19]。

目前,虽然对于CHS的研究已经取得了一些进展,但是*AaCHS*基因家族分析未见报道,其在青蒿中的功能尚不清楚。本研究基于青蒿已公布的参考基因组,鉴定了*AaCHS*基因家族成员。本研究验证了这些成员的序列特征、蛋白质保守基序、顺式作用元件和编码序列。同时,还分析了CHS蛋白结构域及其系统进化关系,并绘制了在青蒿不同组织器官和光调控条件下的表达量热图。最后,筛选出在不同光调控条件下差异表达的*CHS*基因。研究结果对挖掘具有特殊功能的青蒿中*CHS*基因家族成员具有重要意义。

1 材料与方法

1.1 *AaCHS*基因家族成员鉴定与序列特征分析

在NCBI数据库中下载青蒿数据库(PRJNA416223)的全基因组文件、GFF注释文件、CDS序列和蛋白序列。从Pfam数据库获取CHS结构域种子文件(PF00195),然后在Linux系统中使用HMMER 3.2.1软件创建Profile HMM,与青蒿的蛋白序列进行比对,去除冗余序列。接下来筛选出质量较高的基因家族候选基因,要求E-value$<1\times10^{-20}$,并使用clustalW2对这些高质多序列比对。随后利用hmmbuild构建隐马可夫模型,并使用该新建模型进一步筛选完整的CHS基因家族序列,筛选得到AaCHS蛋白序列。在NCBI下载拟南芥CHS的蛋白质序列文件,利用Blastp软件包进行同源比对,筛选目标物种中identity大于75%的序列,可筛选得到AaCHS蛋白序列。最后将Blastp同源比对的数据和HMMER筛选过后的数据取交集,得到AaCHS蛋白。将候选蛋白序列用NCBI中的CD-Search功能进行预测,对候选转录因子序列进行检验,最终鉴定出*AaCHS*基因家族成员。利用在线工具ExPAS软件对*AaCHS*基因家族成员的蛋白质长度、等电点、pH值等基本信息进行了分析[20];利用CELLO v.2.5软件进行亚细胞定位预测[21];利用SOPMA对*AaCHS*基因蛋白质α螺旋、β转角、延长线、随机线圈进行分析[22];利用SWISSMODEL对*AaCHS*基因蛋白质三级结构预测[23]。

1.2 *AaCHS*基因家族成员顺式作用元件分析

用TBtools软件从青蒿的完整基因组序列中提取了16个*AaCHS*基因起始密码子上游2000 bp的启动子序列。使用PlantCARE在线数据库对这些启动子进行了顺式作用元件的分析[24],并最终在Tbtools软件中进行了可视化展示。

1.3 *AaCHS*家族成员保守结构域分析

使用MEGA7中ClustalW工具对AaCHS蛋白进行多序列比对,然后利用Weblogo软件获取其标识。最后,通过在线网站MEME和TBtools软件对AaCHS蛋白保守基序motif进行分析[25]。

1.4 *AaCHS*家族成员系统发育树的构建

在MEGA7中,以多序列比对的结果先计算遗传距离,之后在MEGA7软件的phylogeny(系统进化)中使用邻接法(NJ)法,设置其为Poisson model模型,此外设置bootstrap(n=1000)进行系统发育分析[26]。

1.5 *AaCHS*基因家族成员表达分析

从NCBI数据库中下载青蒿不同部位的转录组数据(SRP129502)以及青蒿光调控转录组数据(Pro-ject: PRJNA435470)。使用Trimmomatic软件去除测序数据中质量较低的序列,并利用Fast QC软件将测序数据转换为Fastq格式。使用hisat2重新构建索引文件,并将其与转录组数据在Fastq格式下进行比对。最后,通过定量分析计算

得到TPM值,进行$\log_{10}$(TPM+1)后生成基因表达量矩阵[27]。最后在TBtools中对该矩阵进行归一化处理并进行可视化分析。

1.6　AaCHS基因家族成员GO注释

根据eggNOG-mapper在线网站,对青蒿蛋白序列进行GO注释,所得注释文件以差异基因作为筛选范围,得到AaCHS基因功能注释文件,再通过Rstudio进行可视化。

1.7　*AaCHS*基因家族成员差异基因分析

差异表达基因在Linux系统下使用R语言,使用BiocManager中edgeR软件包进行差异表达分析,所得差异表达基因的P-Value去除NA值,并对筛选出的*AaCHS*基因家族成员进行标记,使用enhancedvolcano R包进行可视化。

2　结果与分析

2.1　*AaCHS*基因家族成员鉴定及理化性质分析

*AaCHS*基因家族成员鉴定以Pfam数据库下载的CHS结构域序列(PF00195)靶点进行筛选,与Blastp 2次筛选求交集,最后使用NCBI中CD-search功能筛选得到16个CHS蛋白(表1),根据命名规则将其基因依次命名为*AaCHS1*～*AaCHS16*[28]。对该基因家族成员蛋白长度、分子量、等电点、亲水性平均值、脂肪指数、亚细胞定位分析,结果表明AaCHS15蛋白长度最长为516个氨基酸,AaCHS3蛋白长度最短为266个氨基酸;分子量为29 392.96～56 149.6 Da;亲水性平均值从－0.149(*AaCHS16*)到0.432(*AaCHS15*);脂肪指数最高的为99.97(*AaCHS2*),最小的为89.60(*AaCHS12*)。亚细胞定位预测显示所有AaCHS蛋白都定位在细胞质中(表1)。根据AaCHS基因家族成员蛋白二级结构信息表(表2),可见α螺旋为39.47%～56.98%;β转角4.84%～8.64%;延长线13.18%～20.30%;随机线圈25.00%～36.50%。由二级结构可知,β转角均小于50%,可知*AaCHS*基因家族成员结构疏松。根据*AaCHS*基因家族成员三级结构预测图(图1),再次证明了AaCHS蛋白结构。

表1　青蒿CHS家族成员蛋白基本信息

Table1　Basic information of CHS family member proteins in *A. annua*

基因座标签 Locus tag	基因 Gene	蛋白ID Protein ID	蛋白质长度 Length of protein	分子量/Da Molecule weight	等电点 Isoelectric point	亲水性平均值 Grand average of hydropathicity	脂肪指数 Adipose index	亚细胞定位 Subcellular location
CTI12_AA076100	*AaCHS1*	PWA93008.1	401	43 893.75	5.96	－0.051	91.20	细胞质
CTI12_AA096970	*AaCHS2*	PWA90843.1	395	43 368.24	5.75	0.005	99.97	细胞质
CTI12_AA096980	*AaCHS3*	PWA90844.1	266	29 392.96	8.86	－0.107	92.41	细胞质
CTI12_AA096990	*AaCHS4*	PWA90845.1	401	43 832.69	5.56	0.045	99.93	细胞质
CTI12_AA158690	*AaCHS5*	PWA80332.1	389	42 749.28	5.71	－0.062	93.26	细胞质
CTI12_AA187870	*AaCHS6*	PWA80150.1	389	42 904.66	6.56	－0.066	94.29	细胞质
CTI12_AA187880	*AaCHS7*	PWA80151.1	401	43 863.66	5.96	－0.060	90.95	细胞质
CTI12_AA187900	*AaCHS8*	PWA80153.1	389	42 898.43	5.88	－0.068	96.56	细胞质
CTI12_AA210810	*AaCHS9*	PWA77763.1	387	42 412.68	5.21	－0.113	95.99	细胞质
CTI12_AA107510	*AaCHS10*	PWA65027.1	398	43 579.29	5.91	－0.063	92.41	细胞质
CTI12_AA107530	*AaCHS11*	PWA65029.1	301	33 354.92	9.41	0.005	96.21	细胞质
CTI12_AA381710	*AaCHS12*	PWA60393.1	378	41 539.81	5.74	－0.145	89.60	细胞质
CTI12_AA386950	*AaCHS13*	PWA59740.1	396	43 333.04	6.38	－0.013	99.49	细胞质
CTI12_AA460310	*AaCHS14*	PWA51752.1	403	44 312.32	7.63	－0.085	98.21	细胞质
CTI12_AA544590	*AaCHS15*	PWA42080.1	516	56 149.66	8.45	0.432	97.15	细胞质
CTI12_AA576510	*AaCHS16*	PWA38922.1	403	44 588.43	6.60	－0.149	93.15	细胞质

2.2　*AaCHS*基因家族成员系统进化分析

为了研究不同植物之间*CHS*基因家族成员的系统进化关系,我们使用邻接法(NJ)对包括青蒿(*Artemisia annua*)16条、雷蒙德氏棉(*Gossypium*

表2 青蒿CHS基因家族成员蛋白二级结构信息

Table2 Secondary structure of the proteins of the CHS gene family members in *A. annua*

基因座标签 Locus tag	基因 Gene	蛋白ID Protein ID	α螺旋/% Alpha helix	β转角/% Beta turn	延长线/% Extended strand	随机线圈/% Random coil
CTI12_AA076100	*AaCHS1*	PWA93008.1	43.64	5.99	15.21	35.16
CTI12_AA096970	*AaCHS2*	PWA90843.1	44.56	6.58	13.67	35.19
CTI12_AA096980	*AaCHS3*	PWA90844.1	39.47	8.27	20.30	31.95
CTI12_AA096990	*AaCHS4*	PWA90845.1	44.39	6.23	15.46	33.92
CTI12_AA158690	*AaCHS5*	PWA80332.1	43.19	7.97	15.94	32.90
CTI12_AA187870	*AaCHS6*	PWA80150.1	42.93	6.43	14.14	36.50
CTI12_AA187880	*AaCHS7*	PWA80151.1	44.39	6.73	15.71	33.17
CTI12_AA187900	*AaCHS8*	PWA80153.1	44.73	7.20	14.40	33.68
CTI12_AA210810	*AaCHS9*	PWA77763.1	42.38	6.98	16.80	33.85
CTI12_AA107510	*AaCHS10*	PWA65027.1	43.22	6.28	16.08	34.42
CTI12_AA107530	*AaCHS11*	PWA65029.1	46.84	8.64	16.28	28.24
CTI12_AA381710	*AaCHS12*	PWA60393.1	44.18	6.88	15.08	33.86
CTI12_AA386950	*AaCHS13*	PWA59740.1	45.45	7.32	15.66	31.57
CTI12_AA460310	*AaCHS14*	PWA51752.1	40.20	7.20	18.36	34.24
CTI12_AA544590	*AaCHS15*	PWA42080.1	56.98	4.84	13.18	25.00
CTI12_AA576510	*AaCHS16*	PWA38922.1	44.17	7.69	14.39	33.75

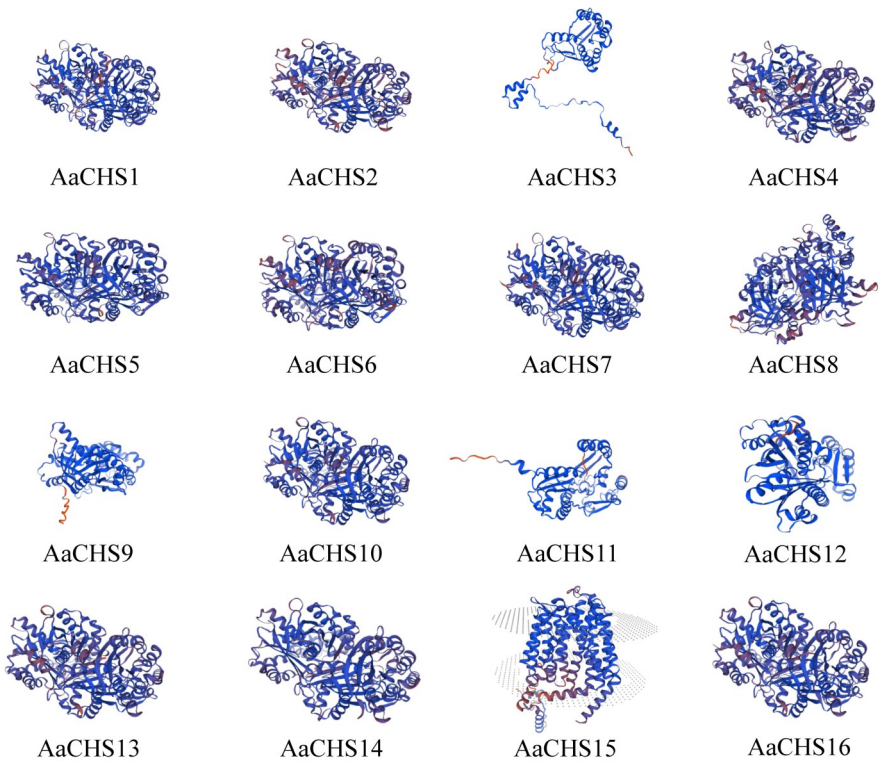
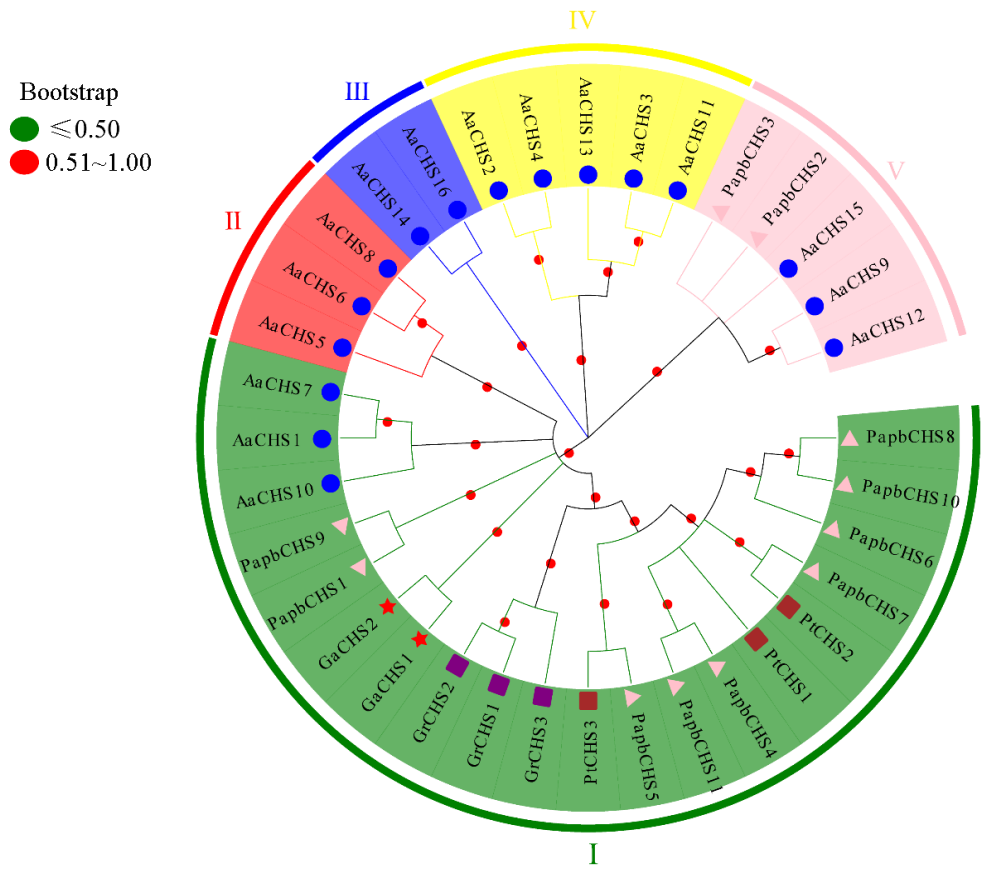


图1 青蒿CHS基因家族成员蛋白三级结构预测

Fig.1 Predicted tertiary structure of the proteins of the CHS gene family members in *A. annua*

raimondii)3条、毛果杨(*Populus trichocarpa*)3条、银中杨(*Populus alba*×*P. Berolinensis*)11条和大豆(*Glycine max*)2条在内的5个物种共35条CHS蛋白序列进行了系统进化分析(图2)。通过这项分析,我们将CHS蛋白划分为5个亚族:亚族Ⅰ(图2中绿色部分)中包含CHS蛋白数量最多,其

中包含雷蒙德氏棉、大豆和毛果杨全部CHS蛋白，还包含青蒿CHS蛋白(AaCHS1、AaCHS7、AaCHS10)；亚族Ⅱ(图2中红色部分)中只包含青蒿CHS蛋白(AaCHS5、AaCHS6、AaCHS8)；亚族Ⅲ(图2中蓝色部分)中只包含剩余的青蒿CHS蛋白(AaCHS14、AaCHS16)，亚族Ⅴ(图2中粉色部分)中银中杨蛋白和青蒿CHS蛋白(AaCHS9、AaCHS12、AaCHS15)聚为一支，雷蒙德氏棉、大豆和毛果杨都属于双子叶植物。因此这些物种的*CHS*基因家族成员之间存在较为密切的进化关系，并且归类在同一类别。



●:青蒿*CHS*基因;★:大豆*CHS*基因;▲:银中杨*CHS*基因;•:毛果杨*CHS*基因;■:雷蒙德氏棉*CHS*基因

●: *A. annua CHS* gene; ★:*Glycine max CHS* gene; ▲: *Populus alba*×*P. Berolinensis CHS* gene;•: *Populus trichocarpa CHS* gene; ■: *Gossypium raimondii CHS* gene

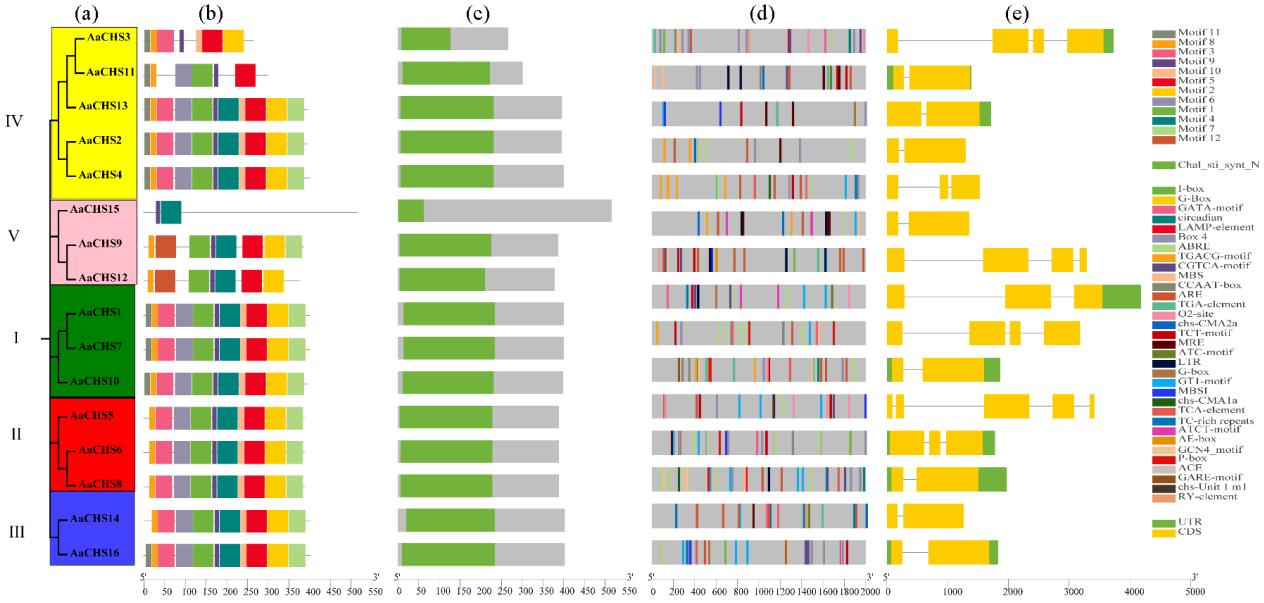
图2　青蒿、大豆、银中杨、毛果杨、大豆构建邻接法系统进化树

Fig.2　Phylogenetic tree constructed with *A. annua*, *Glycine max*, *Populus alba* × *P. Berolinensis*, *Populus trichocarpa* and *Gossypium raimondii*

2.3　*AaCHS*基因结构、顺式作用元件及鉴定分析

利用TBtools软件对AaCHS基因家族成员的编码序列基因结构进行了分析，结果显示16个*AaCHS*基因中有8个没有非编码区，而所有16个基因均包含CDS外显子(图3d)，经CD-search鉴定后发现16个*AaCHS*成员均有Chal-sti-synt_N结构(图3c)。通过使用PlantCARE在线网站对青蒿*CHS*基因家族成员的启动子区域进行分析，研究了转录起始位点上游约2000 bp的序列。根据分析结果，发现了77个顺式作用元件(图3d)，其中也包括常见元件TATA-box、CAAT-box以及A-box，这些元件中部分元件与生长发育、逆境胁迫响应、植物激素以及光响应相关(图4)，其中与生长发育相关的元件有6种，包含有参与玉米醇溶蛋白代谢调节(O_2-site)、调节各种细胞过程(CCAAT-box)等；与逆境胁迫有关的元件有4种，包括低温响应顺式作用元件(LTR)、抗病和胁迫响应元件(TC-rich repeats)、厌氧诱导作用元件

(ARE);与植物激素有关的元件有7种,其中包括参与ABA调节的基因表达(ABRE)、水杨酸顺式作用元件(TCA-element)、茉莉酸响应元件(CGTCA-motif)等;光响应元件种类最多有16种,其中包括调控靶基因的表达(MRE)、植物响应外界环境刺激的通用调控元件(G-box)。利用MEME在线网站在*AaCHS*基因家族成员中找到12个保守基序(Motif 1~Motif 12),通过对图5中的保守基序分析,结果表明:根据进化关系,在亚族Ⅱ中,3个*CHS*基因(*AaCHS5*、*AaCHS6*和*AaCHS8*)有相同的10个保守结构域;亚族Ⅰ中的3个*CHS*基因(*AaCHS1*、*AaCHS7*和*AaCHS10*)有相同的11个保守结构域;而在亚族Ⅳ中,*AaCHS2*、*AaCHS4*、*AaCHS11*和*AaCHS13*均具有相同的11个保守基序,亚族Ⅴ中*AaCHS15*仅有2个保守结构域,*AaCHS9*和*AaCHS12*中除了Motif7外的7个保守基序均相同;亚族Ⅲ中*AaCHS14*和*AaCHS12*除了Motif 11的10个保守结构域均相同(图3b)。



a:青蒿*CHS*基因家族成员邻接法系统进化树;b:Motif可视化;c:结构域分析;d:顺式作用原件;e:内含子外显子可视化

a: Phylogenetic tree of *A. annua CHS* gene family members by neighbor-joining method; b: Motif visualization; c: Domain analysis; d: Cis-element original; e: Visualization of intron and extron

图3 *AaCHS1-16*序列分析

Fig.3 Sequence analysis of *AaCHS1-16*

2.4 *AaCHS*家族成员表达分析

为了研究*AaCHS*基因家族成员的表达模式,我们对不同部位组织(嫩叶、老叶、根、表皮、花蕾、种子、花、毛状体)青蒿转录组测序结果进行分析(图6a)。研究结果显示,*AaCHS10*和*AaCHS16*属于亚族Ⅲ,其在嫩叶、老叶、表皮、花蕾、种子、花和毛状体均显著上调;在青蒿嫩叶中*AaCHS14*显著上调。此外对*AaCHS*基因家族成员进行光调控分析(图6b),结果表明,*AaCHS16*在白光和远红光条件下均显著上调,在蓝光条件下表达上调;*AaCHS10*在蓝光、红光和白光条件下均有表达,在黑暗下几乎不表达,这表明蓝光、红光和白光会促进*AaCHS10*的表达;*AaCHS6*在白光条件下表达均上调,这些结果能为进一步研究*AaCHS*基因家族成员的功能提供借鉴。

2.5 *AaCHS*基因家族成员GO注释

用eggNOG-mapper在线网站对青蒿蛋白序列进行GO注释,以*AaCHS*基因蛋白为筛选范围,*AaCHS*基因进行功能注释。选取P-value<0.05且富集Count最多的前12个GO二级分类过程,*AaCHS*基因在GO的二级分类的生物过程(biological process, BP)、细胞组分(cellular component, CC)2个大类并将其注释(图7)。从图5可见,生物过程共富集了80条序列;细胞组分共富集了19条序列,其中包括的细胞质(cytoplasm)富集了11条序列,数量最多;而在生物过程中富集了细

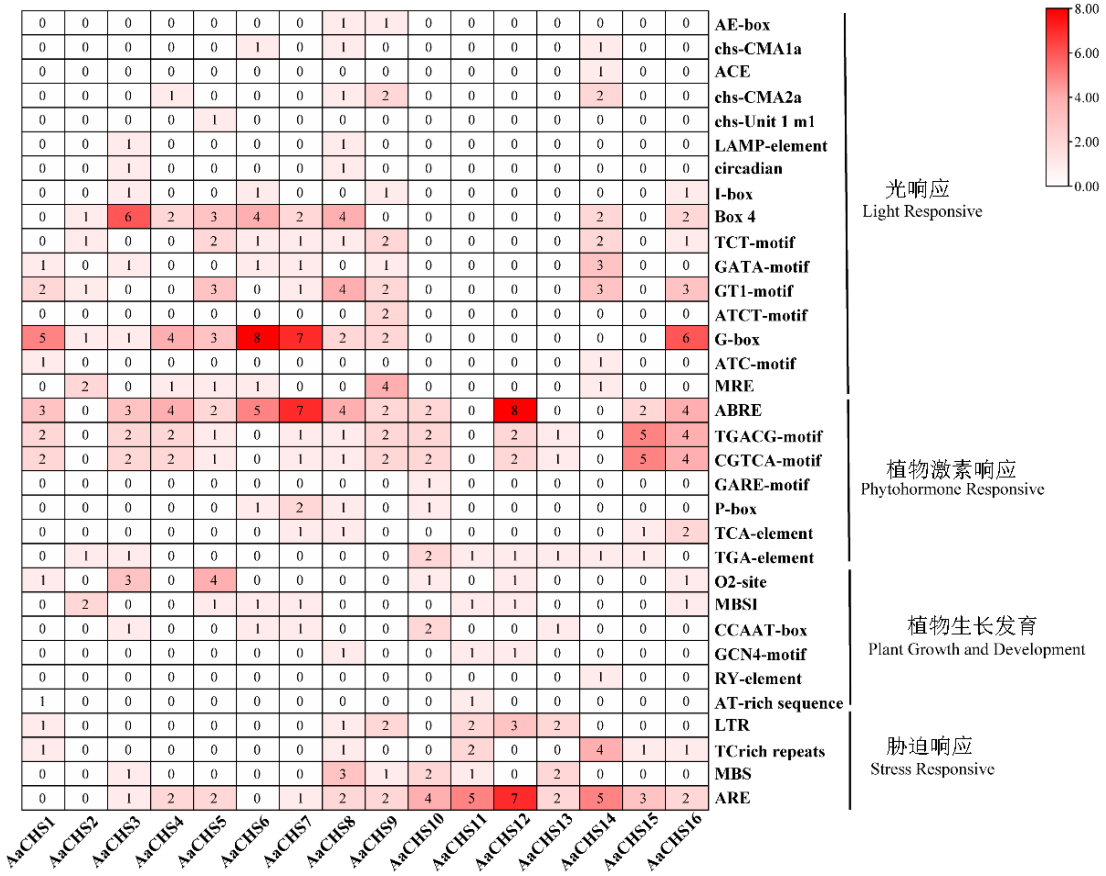


图5　*AaCHS*基因家族顺式作用元件分类

Fig.5　Classification of cis-acting elements of *AaCHS* gene family

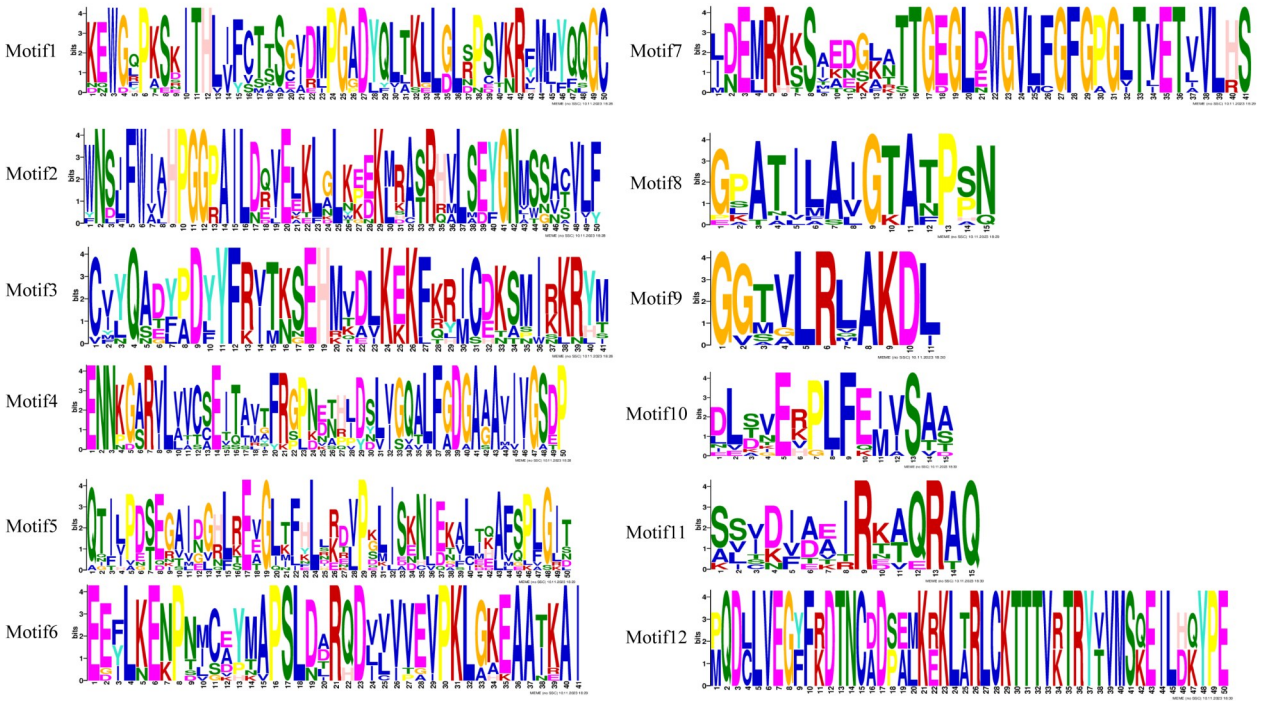
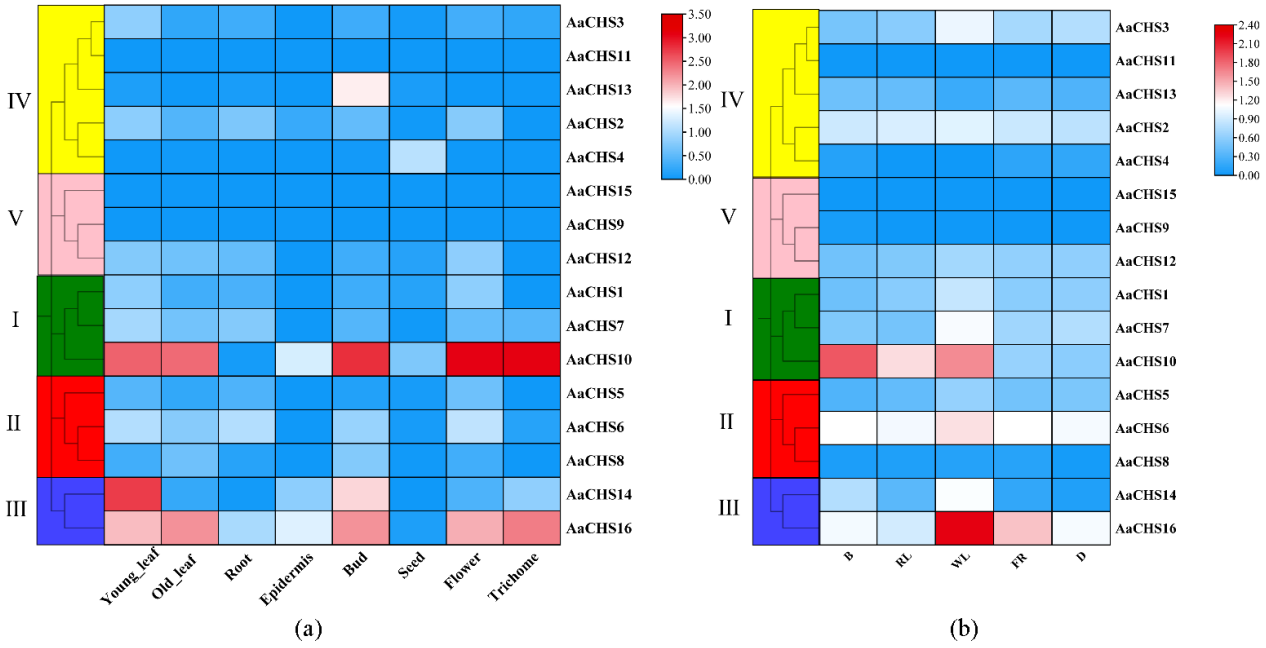


图4　*AaCHS*基因家族Motif序列标识

Fig.4　Motif sequence identification of *AaCHS* gene family



a:青蒿不同组织中CHS家族成员表达分析;b:青蒿CHS家族成员光调控分析(B:蓝光;RL:红光;WL:白光;FR:远红光;D:黑暗)

a: Expression analysis of CHS family members in different tissues of *A. annua*; b: Analysis of light regulation in CHS family members of *Artemisia annua* (B: blue light; RL: Red light; WL: white light; FR: far red light; D: Darkness)

图6 青蒿CHS家族成员表达及光调控分析

Fig. 6 Analysis of expression and photoregulation of CHS family members in *A. annua*

胞成分组装(cellular component assembly)、细胞发育过程(cellular developmental process)、解剖结构的形态发生(anatomical structure morphogenesis)、细胞成分的生物生成(cellular component biogenesis)、内膜系统(endomembrane system)、多细胞生物体的发育(multicellular organism development)、细胞成分组织(cellular component organization)、多细胞生物体的过程(multicellular organismal process)、细胞成分的组织或生物发生(cellular component organization or biogenesis)、解剖结构的发展(anatomical structure development)、发育过程(developmental)均富集了最多的8条序列。

2.6 *AaCHS* 基因家族成员差异基因分析

基于青蒿在不同光照下的调控,将青蒿在蓝光、红光、白光、远红光调控的转录组测序数据作为实验组,黑暗条件作为对照组,依次进行差异基因表达分析并将其分为4组,第1组为蓝光和黑暗条件下差异表达分析;第2组为红光和黑暗条件下差异表达分析;第3组为白光和黑暗条件下差异表达分析;第4组为远红光和黑暗条件下差异表达分析。将分组过后的2组青蒿器官进行差异表达分析;在Linux系统下使用R语言,利用BiocManager中的edgeR软件包对差异表达分析,将所得差异表达基因的P-Value去除NA值,将既是差异表达基因又是*AaCHS*基因家族成员筛选出来,使用enhancedvolcano R包将其可视化,以$Log_{10}FC>2.0$且P-Value>0.05为阈值筛选差异表达基因。分析结果发现,如图8b,*AaCHS7*、*AaCHS10*是红光和黑暗条件下差异表达分析下的正向差异表达基因,其$Log_{10}FC$值均大于2.0且P-Value值大于0.05;而在另外3组(图8a、图8c、图8d)中均未发现差异表达基因。

3 讨 论

黄酮类化合物是一类植物次生代谢产物。黄酮类生物合成途径包括类苯丙素途径和类黄酮途径。查尔酮合成酶(CHS)是苯丙素类化合物生物合成途径中的关键酶,在植物中普遍存在。它在类苯丙素途径的初始步骤中起作用,并使3个丙二酰-CoA分子与1个4-香豆酰-CoA分子缩合以产生柚皮素查尔酮,其是各种类黄酮的前体[28-31]。因此,研究查尔酮合成酶基因对于进一步分析影响青蒿中类黄酮化合物合成的环节十分必要。本研究对青蒿中*CHS*基因家族进行了识别,并进一步

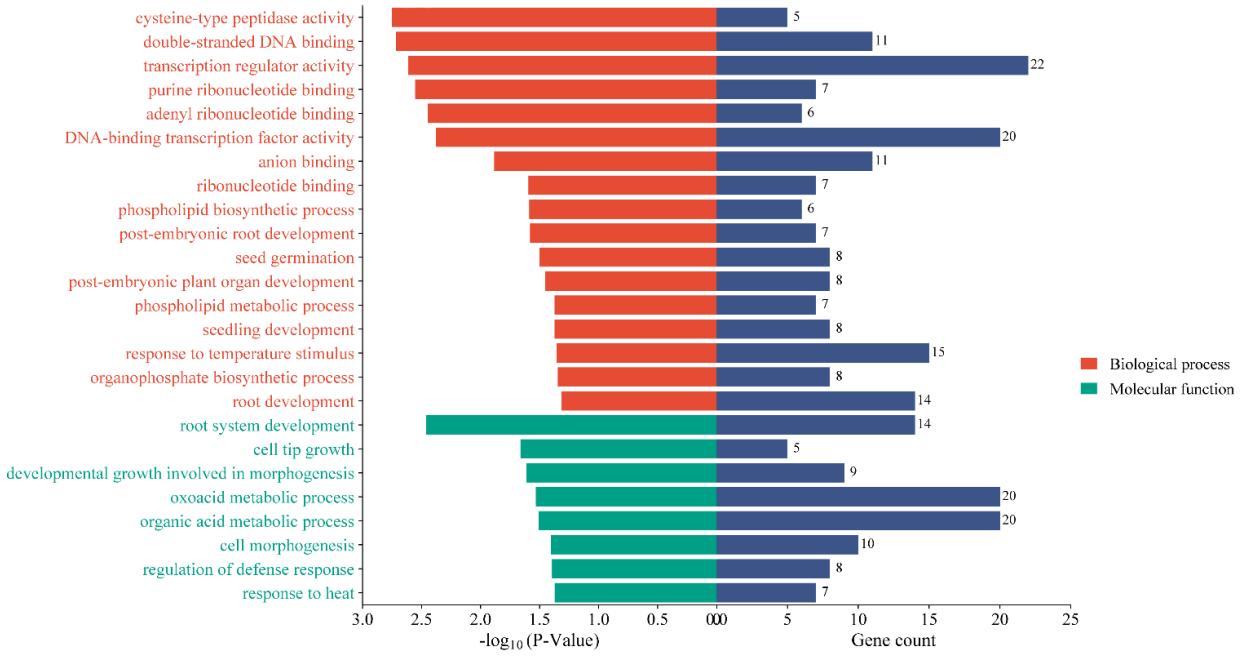
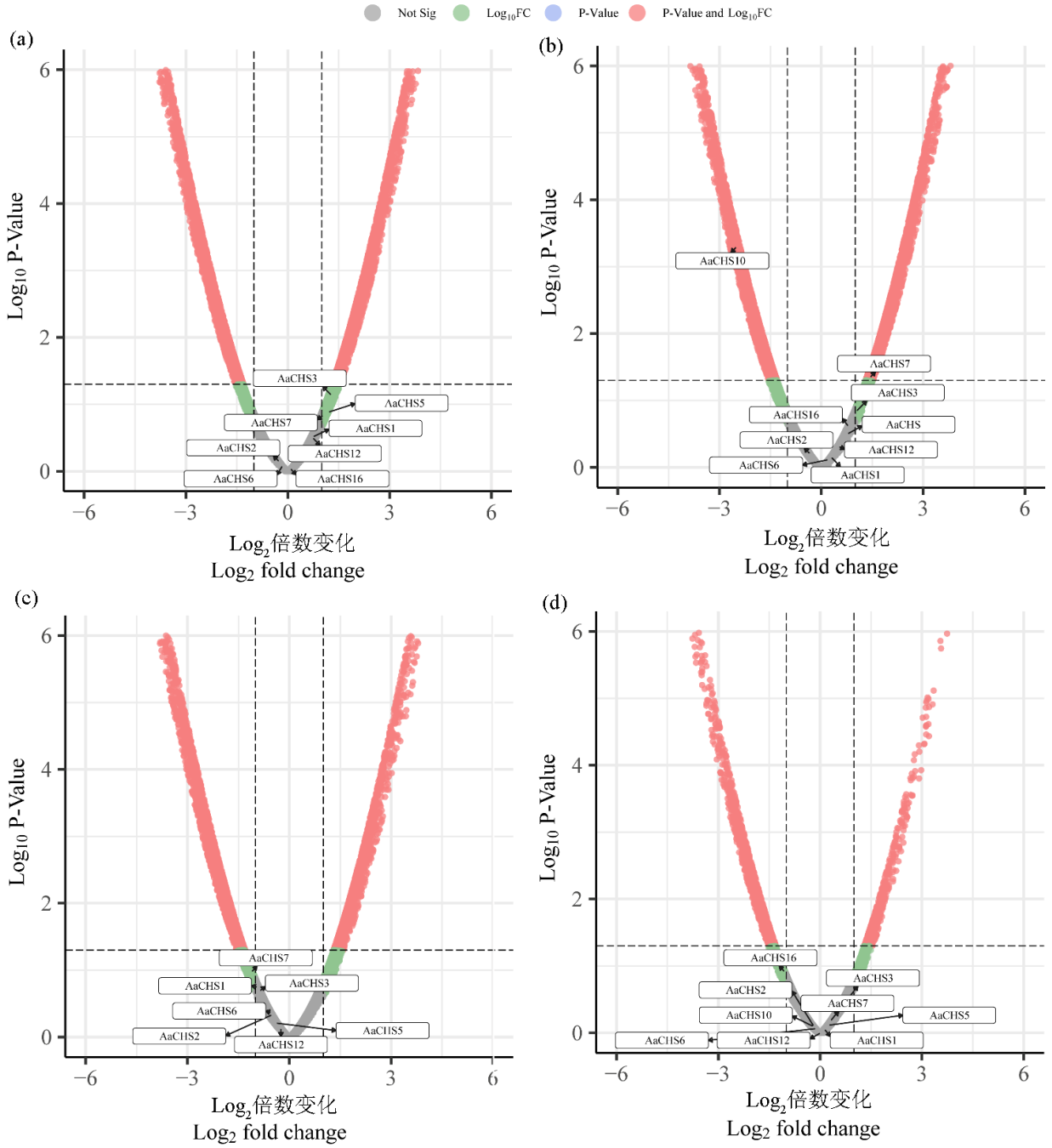


图7　AaCHS基因家族成员GO注释信息

Fig. 7　AaCHS gene family members GO annotation information

分析了*AaCHS*基因家族成员的理化性质结构、基因序列、进化关系及不同组织部位*AaCHS*基因表达模式。在青蒿中共发现16个*AaCHS*基因家族成员,本研究还对AaCHS蛋白一级结构进行了分析,包括蛋白质大小、分子量、等电点、平均亲水性系数及亚细胞定位分析,结果发现16个AaCHS蛋白均定位于细胞质;大多数为酸性不亲水蛋白。对AaCHS蛋白二级结构进行了分析,包括α螺旋、β转角、延长线、随即线圈分析,结果发现16个AaCHS蛋白的β转角均低于50%,说明16个AaCHS蛋白结构疏松;对AaCHS蛋白三级结构可视化分析,证明了这一观点。通过邻接法(NJ)对5个植物物种(包括青蒿、雷蒙德氏棉等)的35条CHS蛋白序列进行系统进化分析后发现,所有AaCHS被划分为5个亚族,并与银中杨CHS蛋白亲缘关系更近,推测他们具有相似的蛋白结构[32];对*AaCHS*基因序列分析发现:16条*AaCHS*基因有7条没有内含子,剩余基因的内含子数量为0~2个,16条基因均包含CDS外显子,含有2~5个外显子,何梦媛等[33]与本研究结果类似。另外,在所有青蒿*AaCHS*基因上游启动子区域内最常见的元件是光反应元件,其次是植物激素响应元件,最后是生长发育和逆境相关元件。本研究在*AaCHS*基因中找到12个保守基序(Motif 1~Motif 12),通过对其分析发现:根据进化关系,在亚族Ⅱ中,3个*CHS*基因(*AaCHS5*、*AaCHS6*、*AaCHS8*)有相同的10个保守结构域;亚族Ⅰ中的3个*CHS*基因(*AaCHS1*、*AaCHS7*、*AaCHS10*)有相同的11个保守结构域;而在亚族Ⅳ中,*AaCHS2*、*AaCHS4*、*AaCHS11*和*AaCHS13*均具有相同的11个保守基序;亚族Ⅲ中*AaCHS14*和*AaCHS12*除了Motif 11的10个保守结构域均相同;亚族Ⅴ中*AaCHS15*仅有2个保守结构域,*AaCHS9*和*AaCHS12*中除了Motif 7,其它7个保守基序均相同。在亚族Ⅰ、Ⅱ、Ⅲ和Ⅳ中,除了*AaCHS3*外的*AaCHS*家族成员均具有相同的10个保守结构域,根据进化树划分的亚族成员,其保守结构域存在高度保守性,推测它们有共同的起源,此外不同亚族成员间也发生了进化,如图3b中的Motif 11。

为了深入探究*AaCHS*基因在不同组织部位中可能发挥的功能,通过挖掘NCBI数据库中青蒿8个不同组织部位(嫩叶、老叶、根、表皮、花蕾、种子、花、毛状体)转录组数据,我们筛选出了16个*AaCHS*基因家族成员,并发现它们在各个组织中均没有表达下调情况,与Shen等[34]研究结果类似,证明了我们筛选到的*AaCHS*基因家族成员是可靠的。生物信息学分析显示,AaCHS蛋白序列高



a:蓝光和黑暗条件下差异表达分析;b:红光和黑暗条件下差异表达分析;c:白光和黑暗条件下差异表达分析;d:远红光和黑暗条件下差异表达分析

a: Differential expression analysis under blue light and dark conditions; b: Differential expression analysis under red light and dark conditions; c: Differential expression analysis under white light and dark conditions; d: Differential expression analysis under far red light and dark conditions

图8 AaCHS基因家族成员差异基因分析

Fig.8 Differential gene analysis of AaCHS gene family members

度保守,它们的Motif分布也支持了系统进化树的结果。根据系统进化树,*AaCHS10*和*AaCHS16*在嫩叶、老叶、表皮、花蕾、种子、花和毛状体均显著上调;在青蒿嫩叶中*AaCHS14*基因显著上调;*AaCHS2*、*AaCHS3*、*AaCHS4*、*AaCHS11*和*AaCHS13*聚在一起,其表达水平类似,推测它们可能有共同的起源并且序列保守[35],Hassani等[36]研究发现查尔酮合成酶在青蒿的芽中显著表达,

与本研究所得结论一致。为进一步了解不同光质对AaCHS表达的调控,对蓝光、黑暗、红光、远红光、白光5个条件下*AaCHS*家族基因表达模式进行了分析。远红光处理下,*AaCHS1*~*AaCHS15*的表达量都很低,推测在青蒿生长发育和逆境适应过程中,它们可能并未发挥作用,但这一推测需要进一步验证。此外,我们还发现,与黑暗条件相比,光诱导可以显著提升植物体内*CHS*基因表达量。比如,与黑暗处理相比,蓝光诱导能够显著增加*AaCHS10*基因的表达量,Kitazaki等[37]通过分析莴苣叶的代谢途径发现,在蓝光下*CHS*表达量显著上调,这与本研究所得结果一致。白光诱导则能显著提高*AaCHS10*和*AaCHS16*基因的表达量。而在红光处理下,*AaCHS16*的表达受到抑制,赖恭梯等[18]研究发现刺葡萄经红光处理后*CHS*表达水平降低,与本研究所得结果一致。CHS与植物体内的类黄酮等次生代谢物的生成有关,其活性与类黄酮含量紧密相连[38]。因此,可以推测,作为环境因素之一的光质,可能是通过影响植物体内CHS基因表达量,进而影响青蒿中黄酮类物质的生成。这些结果也为深入研究*AaCHS*基因的功能,及光调控下黄酮化合物生物合成的调控机制提供了参考。

根据eggNOG-mapper在线网站,对*AaCHS*基因进行GO注释,以*AaCHS*基因蛋白作为筛选范围,对*AaCHS*基因进行功能注释,将其分为生物过程(biological process,BP)、细胞组分(cellular component,CC)2个大类,其中生物过程共富集了90条序列;细胞组分共富集了19条序列,其中包括的细胞质(cytoplasm)富集了11条序列,数量最多,与亚细胞定位预测结果一致;推测*AaCHS*基因在细胞质中调控黄酮类化合物的累积。为了研究*AaCHS*基因家族成员在不同光照下的差异表达,将青蒿在蓝光、红光、白光、远红光调控的转录组测序数据作为实验组,而黑暗条件则作为对照组。根据差异基因表达火山图结果显示,*AaCHS7*、*AaCHS10*是红光和黑暗条件下差异表达分析下的正向差异表达基因。推测它们在青蒿受到红光处理中起正向调控作用。此外,蔡伟等[39]对鱼腥草中的2种材料(鱼腥草6号和7号)进行了差异表达分析,并筛选出4个基因。本研究采用相同的筛选方法,即通过比较试验组与对照组之间的差异表达情况来确定目标基因。然而,在蓝光、白光、远红光以及黑暗条件下,未筛选出*AaCHS*差异表达基因。

4　结论

本研究在青蒿基因组中发现了16个*CHS*基因家族成员,并且这些家族成员有12个保守的基序。通过序列分析,发现*AaCHS*家族相对较为保守,其蛋白质结构也比较松散;其含有大量激素、逆境、生长和光响应元件;有8条没有非编码区,16条均包含CDS外显子;CD-search验证结果发现每个AaCHS蛋白均有Chal-sti-synt_N结构域;系统进化分析将青蒿(16条)、雷蒙德氏棉(3条)、毛果杨(3条)、银中杨(11条)、大豆(2条)在内的5个物种CHS蛋白划分为5个亚族。对*AaCHS*基因家族成员表达分析,*AaCHS10*和*AaCHS16*在嫩叶、老叶、表皮、花蕾、种子、花和毛状体均显著上调;在青蒿嫩叶中*AaCHS14*基因显著上调。对*AaCHS*进行光调控分析,发现与黑暗处理相比,蓝光诱导能显著提高*AaCHS10*基因的表达量,白光诱导则能显著提高*AaCHS10*和*AaCHS16*基因的表达量。*AaCHS*基因进行GO功能注释,富集结果分为2个大类BP和CC,其中包括的细胞质(cytoplasm)富集了11条序列,数量最多,与亚细胞定位预测结果一致,推测*AaCHS*基因在细胞质中调控黄酮类化合物的累积。通过差异表达分析,*AaCHS7*、*AaCHS10*是红光和黑暗条件下差异表达分析下的正向差异表达基因,推测其在青蒿受到红光处理中起正向调控作用。

参　考　文　献

[1]Vidic D, Čopra-Janićijević A, Miloš M, et al. Effects of different methods of isolation on volatile composition of *Artemisia annua* L [J]. International Journal of Analytical Chemistry, 2018,2018:9604183.

[2]Hao X L, Zhong Y J, Nützmann H W, et al. Light-induced artemisinin biosynthesis is regulated by the bZIP transcription factor *AaHY5* in *Artemisia annua* [J]. Plant and Cell Physiology, 2019, 60(8): 1747-1760.

[3]WHO Expert Committee on the Selection, Use of Essential Medicines, World Health Organization. The Selection and Use of Essential Medicines [M]. World Health Organization, 2014: 1-88.

[4]Bilia A R, Lazari D, Messori L, et al. Simple and rapid physico-chemical methods to examine action of antimalarial drugs with

hemin: its application to *Artemisia annua* constituents [J]. Life Sciences, 2002, 70(7): 769-778.

[5]Liu C S C, Yang S L, Roberts M F, et al. Antimalarial activity of *Artemisia annua* flavonoids from whole plants and cell cultures [J]. Plant Cell Reports, 1992, 11: 637-640.

[6]Cavaiuolo M, Cocetta G, Ferrante A. The antioxidants changes in ornamental flowers during development and senescence [J]. Antioxidants, 2013, 2(3): 132-155.

[7]Iwashina T. Flavonoid function and activity to plants and other organisms [J]. Biological Sciences in Space, 2003, 17(1): 24-44.

[8]Zhang P P, Du H Y, Wang J, et al. Multiplex CRISPR/Cas9-mediated metabolic engineering increases soya bean isoflavone content and resistance to soya bean mosaic virus [J]. Plant Biotechnology Journal, 2020, 18(6): 1384-1395.

[9]Pourcel L, Routaboul J M, Cheynier V, et al. Flavonoid oxidation in plants: from biochemical properties to physiological functions [J]. Trends in Plant Science, 2007, 12(1): 29-36.

[10]Colinas M, Goossens A. Combinatorial transcriptional control of plant specialized metabolism [J]. Trends in Plant Science, 2018, 23(4): 324-336.

[11]Flores-Sanchez I J. Polyketide synthases in *Cannabis sativa* L [M]. Amsterdam:PrintPartners Ipskamp B.V., 2008:1-169.

[12]Hahlbrock K, Scheel D. Physiology and molecular biology of phenylpropanoid metabolism [J]. Annual Review of Plant Physiology and Plant Molecular Biology, 1989, 40: 347-369.

[13]Koes R E, Spelt C E, van den Elzen P J M, et al. Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida* [J]. Gene, 1989, 81(2): 245-257.

[14]Franken P, Niesbach-Klösgen U, Weydemann U, et al. The duplicated chalcone synthase genes *C2* and *Whp* (*white pollen*) of *Zea mays* are independently regulated; evidence for translational control of *Whp* expression by the anthocyanin intensifying gene in [J]. The EMBO Journal, 1991, 10(9): 2605-2612.

[15]Zhu L L, Ding Y Q, Wang S X, et al. Genome-wide identification, characterization, and expression analysis of *CHS* gene family members in *Chrysanthemum nankingense* [J]. Genes, 2022, 13(11): 2145.

[16]Hu H F, Shi B, Zhu W C, et al. Genome-wide identification, characterization and expression analysis of mango (*Mangifera indica* L.) chalcone synthase (*CHS*) genes in response to light [J]. Horticulturae, 2022, 8(10): 968.

[17]贾昆鹏. 植物激素独脚金内酯和茉莉酸信号与光信号互作的分子机制研究[D].上海:上海交通大学,2014.
Jia K P. The molecular mechanism of cross-talking between light- and phytohormones strigolactone- and Jasmonate-signaling [D]. Shanghai:Shanghai Jiao Tong University, 2014.

[18]赖恭梯,阙秋霞,潘若,等. 刺葡萄查尔酮合成酶基因*CHS*对不同光质的响应及转录因子调控分析[J].生物技术通报,2022,38(11):129-139.
Lai G T, Que Q X, Pan R, et al. Response of *Chalcone* synthase gene(*CHS*)to different light quality and transcription factor regulation in *Vitis davidii* [J]. Biotechnology Bulletin, 2022, 38(11): 129-139.

[19]Meng X C, Xing T, Wang X J. The role of light in the regulation of anthocyanin accumulation in *Gerbera hybrida* [J]. Plant Growth Regulation, 2004, 44(3): 243-250.

[20]Mariethoz J, Alocci D, Gastaldello A, et al. Glycomics@ExPASy: bridging the gap [J]. Molecular & Cellular Proteomics, 2018, 17(11): 2164-2176.

[21]Yu C S, Lin C J, Hwang J K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions [J]. Protein Science, 2004, 13(5): 1402-1406.

[22]Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments [J]. Bioinformatics, 1995, 11(6): 681-684.

[23]Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes [J]. Nucleic Acids Research, 2018, 46(W1): W296-W303.

[24]Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences [J]. Nucleic acids research, 2002, 30(1): 325-327.

[25]Chen C J, Chen H, Zhang Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data [J]. Molecular Plant, 2020, 13(8): 1194-1202.

[26]Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11 [J]. Molecular Biology and Evolution, 2021, 38(7): 3022-3027.

[27]Bray N L, Pimentel H, Melsted P, et al. Near-optimal probabilistic RNA-seq quantification [J]. Nature Biotechnology, 2016, 34(5): 525-527.

[28]Jin J F, Wang Z Q, He Q Y, et al. Genome-wide identification and expression analysis of the *NAC* transcription factor family in tomato (*Solanum lycopersicum*) during aluminum stress [J]. BMC Genomics, 2020, 21: 288.

[29]Jez J M, Bowman M E, Noel J P. Structure-guided programming of polyketide chain-length determination in chalcone synthase [J]. Biochemistry, 2001, 40(49): 14829-14838.

[30]Kreuzaler F, Ragg H, Heller W, et al. Flavanone synthase from Petroselinum hortense: Molecular weight, subunit composition, size of messenger RNA, and absence of pantetheinyl residue [J]. European Journal of Biochemistry, 1979, 99(1): 89-96.

[31]Jez J M, Ferrer J L, Bowman M E, et al. Structure and mechanism of chalcone synthase-like polyketide synthases [J]. Journal of Industrial Microbiology and Biotechnology, 2001, 27(6): 393-398.

[32]何锐杰,方庭,余伟军,等. 西番莲查尔酮合成酶(*CHS*)基因家族全基因组鉴定及表达模式[J].应用与环境生物学报,2022,28(4):1066-1075.

He R J, Fang T, Yu W J, et al. Genome-wide identification and expression analysis of the CHS gene family in passion fruit [J]. Chinese Journal of Applied and Environmental Biology, 2022, 28(4): 1066-1075.

[33]何梦媛,姚华,李国治,等. 甘草*CHS*基因家族鉴定、表达特性分析及其与甘草查尔酮A积累的关系研究[J].植物生理学报,2022,58(1):141-154.
He M Y, Yao H, Li G Z, et al. Identification of *CHS* gene family and analysis of its expression characteristics in relation to the accumulation of licochalcone A in Chinese licorice (*Glycyrrhiza uralensis*) [J]. Plant Physiology Journal, 2022, 58(1):141-154.

[34]Shen Q, Zhang L D, Liao Z H, et al. The genome of *Artemisia annua* provides insight into the evolution of Asteraceae family and artemisinin biosynthesis [J]. Molecular Plant, 2018, 11(6): 776-788.

[35]戢强强,陈伟达,陈敬东,等. 青蒿苯丙氨酸解氨酶(*AaPAL*)基因家族鉴定及光调控分析[J].分子植物育种,2023,21(11):3544-3551.
Ji Q Q, Chen W D, Chen J D, et al. Family identification and photoregulation analysis of phenylalanine ammonia lyase (*AaPAL*) gene of *Artemisia annua* [J]. Molecular Plant Breeding, 2023, 21(11): 3544-3551.

[36]Hassani D, Fu X, Shen Q, et al. Parallel transcriptional regulation of artemisinin and flavonoid biosynthesis[J]. Trends in Plant Science, 2020, 25(5): 466-476.

[37]Kitazaki K, Fukushima A, Nakabayashi R, et al. Metabolic reprogramming in leaf lettuce grown under different light quality and intensity conditions using narrow-band LEDs [J]. Scientific Reports, 2018, 8(1): 7914.

[38]孟祥春.非洲菊花生长、花色素苷积累及CHS、DFR基因表达的光调控研究[D].广州:华南师范大学,2004.
Meng X C. Flower growth, anthocyanin accumulation and *CHS*、*DFR* gene expression regulated by light in *Gerbera hybrida* [D]. Guangzhou: South China Normal University, 2004.

[39]蔡伟,肖敬忠,余小丽,等. 鱼腥草bZIP基因的鉴定与分析[J].中草药,2023,54(13):4295-4305.
Cai W, Xiao J Z, Yu X L, et al. Identification and analysis of *bZIP* gene in *Herba Houttuynia* [J]. Chinese Traditional and Herbal Drugs, 2019,54(13):4295-4305.

Identification and photoregulation analysis of Artemisinin chalcone synthetase (*AaCHS*) gene family and analysis of light regulation in *Artemisia annua* L.

Xue Tianyuan, He Yuqing, Xia Mang, Chen Meizhu, Dai Xigang, He Sixiao, Zhu Xun, Zeng Changli

(*Development and Utilization of Characteristic Biological Resources in Hanjiang River Basin, Hubei Engineering Research Center for Conservation, Jianghan University, Wuhan 430056, China*)

Abstract: [Objective] The aim of this study was to understand the molecular evolutionary characteristics of the chalcone synthase (*CHS*) gene family of *Artemisia annua* L. and its expression patterns in different tissues. [Methods] The CHS protein HMM model was downloaded from Pfam database, and *CHS* gene family members in *A. annua* L. genome were identified using HUMMER 3.0, Blastp, and CD-search. Bioinformatic analyses of amino acid and nucleotide sequences were conducted using TBtools, EXPASy, CELLO v.2.5, MEGA 7.0, MEME, SOPMA, SWISSMODEL, and PlantCARE. The expression analysis of *AaCHS* gene family members was performed using transcriptome data from eight different tissues. [Results] The results showed that there were 16 members in the *AaCHS* gene family in *A. annua* genome. Protein structure analysis revealed that the protein structure was unstable, categorizing them into four groups. Gene structure analysis indicated that all *AaCHS* gene had exons, with the number ranging from 2 to 4, and eight *AaCHS* genes lacked introns, with the number ranging from 0 to 2. Promoter cis-element analysis revealed a high abundance of photoresponsive and phytohormone responsive elements. CD-search verified the presence of the Chal-sti-synt_N domain in each AaCHS protein. [Conclusion] Expression analysis of *AaCHS* family members showed different expression patterns in various tissues and under light treatment. GO enrichment analysis of biological processes and molecular functions revealed 80 enriched sequences in the biological process category and 19 sequences in the cellular component category, with the cytoplasm being the most enriched subcellular component (11 sequences). It is speculated that *AaCHS* genes regulate the accumulation of flavonoids in the cytoplasm. Differential expression analysis under light conditions regulation suggested that *AaCHS7* and *AaCHS10* played a positive regulatory role in *A. annua* L. under the red light treatment.

Keywords: *Artemisia annua* L., *CHS* gene family, Photoregulation, Gene expression, Differential gene expression

(编辑:吕俊俐)