

药用野生稻叶中花色苷合成基因的序列特点与表达差异

秦宗燕*, 刘石娟*, 包颖**

曲阜师范大学生命科学学院, 山东曲阜273165

摘要: 为全面把握花色苷合成基因在非模式植物绿色组织中的构成和表达, 本研究利用转录组重测序方法, 对药用野生稻 (*Oryza officinalis*) 绿叶中出现的花色苷合成基因进行了分析, 结果共鉴定出8个相关基因 (*CHS*、*CHI*、*F3H*、*F3'H*、*F3'5'H*、*DFR*、*ANS*和*UFGT*), 这些基因与栽培稻同源基因对之间在mRNA水平上的序列相似度达到了94%~96%。相对外类群的 K_a/K_s (非同义替换率与同义替换率的比值) 统计检验进一步表明这8个基因均处于严格的纯化选择之下。RPKM (每百万读段中每千个碱基长度的读段数) 值分析表明, 8个基因的表达水平不同, 其中*F3'5'H*的相对表达量最高, *CHI*和*CHS*次之, *F3'H*最低。

关键词: 药用野生稻; 叶转录组; 花色苷合成途径; 基因表达

花色苷是一类水溶性植物色素, 属于类黄酮 (flavonoids) 的一个亚类, 是具有类黄酮典型结构——2-苯基苯丙吡喃阳离子结构的衍生物, 具有 $C_6-C_3-C_6$ 碳骨架。花色苷广泛分布于植物的花、果实、茎、叶和根等器官中, 是植物体内一类重要的次生代谢产物, 其化学本质是花青素的糖苷类化合物。植物之所以呈现出各种颜色, 与花色苷有着重要的关系 (Winkel-Shirley 2001)。花色苷结构中有多个酚羟基, 属于羟基供体, 具有独特的抗氧化功能, 因此除显色外, 花色苷还可以作为植物抵抗应力条件的有效防护者, 在提高植物抗性、调节植物激素的运输等方面承担重要功能 (Shih等2008; 张红雨和陈德展2000; Smillie和Hetherington 1999; Nogués等1998; Tsuda等1996)。除作为天然染料外, 花色苷的抗氧化功能也被广泛用于食品、保健和医药等行业中, 被认为可以帮助清除人体内自由基, 并在预防疾病、维持体重、保护视力以及护肤养颜等方面发挥作用 (房欢和焦滨2012; 林晓霞和朱寿民2005; 彭姗姗等2005; 庞志申2000)。

随着在众多的植物中发现了花色苷合成基因和调控基因, 其合成途径和调控途径得到了详细的阐释。花色苷合成途径在被子植物中相对保守。在模式植物如拟南芥 (*Arabidopsis thaliana* (L.) Heynh.)、玉米 (*Zea mays* L.)、矮牵牛 (*Petunia hybrida* Vilm.) 中, 花色苷代谢途径的合成链条上主要包括8个关键结构酶, 由一套对应的基因编码。研究显示这些关键酶的编码基因通常被分成两类: 早期合成基因 (early biosynthesis genes, EBGs) 和晚期合成基因 (late biosynthesis genes, LBGs) (Zhang

等2016; Petroni和Tonelli 2011), 其中EBGs编码的关键酶包括查耳酮合成酶 (chalcone synthase, CHS)、查耳酮异构酶 (chalcone isomerase, CHI)、黄酮-3-羟化酶 (flavanone 3-hydroxylase, F3H)、类黄酮-3'-羟化酶 (flavonoid 3'-hydroxylase, F3'H) 和类黄酮-3',5'-羟化酶 (flavonoid 3',5'-hydroxylase, F3'5'H), 其共同催化合成类黄酮底物得以进行下游花色苷生成反应。而二氢黄酮醇-4-还原酶 (dihydroflavonol 4-reductase, DFR)、花青素合成酶 (anthocyanidin synthase, ANS) 和类黄酮-3-O-葡萄糖基酶 (UDP-glucose:flavonoid 3-O-glucosyltransferase, UFGT) 所对应的编码基因则属于LBGs, 经过这几个关键酶的催化从而产生多种组织中颜色各异的花色苷。虽然对花色苷合成途径中相关酶的研究在很多植物特别是栽培作物中已经进行得非常深入 (Hayashi等2005; Akada和Dube 1995; Franken等1991), 但是针对栽培作物的野生亲缘种类的研究却少之又少。同时, 由于以往的研究过分聚焦于植物的花与果实等显色部位, 而忽略了叶等器官中被绿色本身所掩盖的花色苷的存在, 这就使得在多数绿色组织中对合成花色苷所需要的各种关键酶基因的表达情况缺乏必要的研究, 从而限制了我们对该代谢途径在植物非显色部分如何正常运作和行使功能的全面和正确理解 (图1)。

收稿 2016-05-10 修定 2016-07-21

资助 国家自然科学基金(31570218)。

致谢 国际水稻研究所(International Rice Research Institute, IRRI)提供药用野生稻种子样品。

* 并列第一作者。

** 通讯作者(E-mail: baoyingus@126.com)。

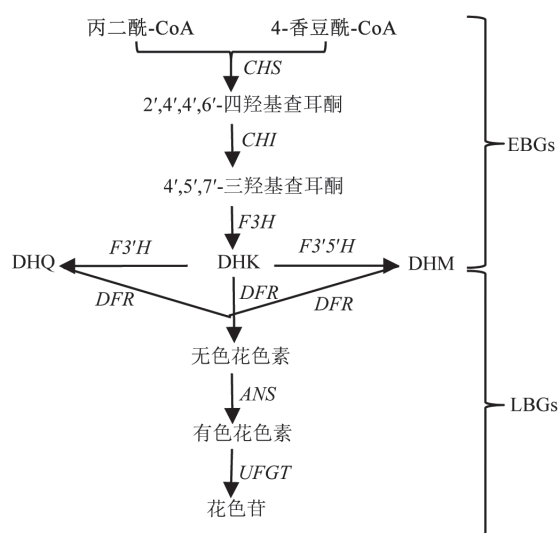


图1 花色苷生物合成途径

Fig.1 The anthocyanin biosynthesis pathway

DHK: 二氢山奈酚(dihydrokaempferol); DHM: 二氢杨梅素(dihydromyricetin); DHQ: 二氢槲皮素(dihydroquercetin)。

稻属(*Oryza* L.)隶属禾本科(Graminea), 包括重要的粮食作物亚洲栽培稻(*O. sativa* L.)和非洲栽培稻(*O. glaberrima* Steud.), 以及约22个野生种(中国科学院中国植物志编辑委员会2002)。作为栽培稻遗传改良最直接的基因库, 针对这些野生稻的研究已经日渐深入。例如稻属中具有最小基因组的物种短花稻(*O. brachyantha* A. Chev. & Roehr.)的全基因组序列(Chen等2013), 以及5个亲缘关系较近的具有AA基因组的物种全基因组重测序数据(Zhang等2014)等已经完成, 而纳入到国际稻属基因组计划(The International *Oryza* Map Alignment Project)中的另外11个物种的1、3和10号染色体的序列信息(Jacquemin等2013)也正在持续公布中。这些工作无疑为更好地了解稻属植物的遗传背景奠定了坚实的基础。在我国, 自然分布的野生稻仅有3种(中国科学院中国植物志编辑委员会2002), 它们是具有AA基因组的野生稻(*O. rufipogon* Griff.)、具有CC基因组的药用野生稻(*O. officinalis* Wall. ex Watt)和具有GG基因组的疣粒稻(*O. granulata* Nees et Arn. ex Hook. f.)。其中, 主要分布在亚洲和澳洲的药用野生稻和分布在非洲和斯里兰卡的另外2个物种拥有相同的基因组类型, 同时也被认为和5个多倍体的起源有关(Zou等2015)。换言之, 药用野生稻所具有的C基因组是稻属中除了A

基因组之外在最多物种中出现的基因组。由于该基因组和栽培稻中的A基因组相对远缘, 因此避免了人工定向选择所导致的遗传改变, 因此以该物种作为材料开展研究不仅在稻属内具有代表性, 而且可以拓宽我们对植物关键次生代谢途径在自然状态下多基因进化模式的正确理解。为了进一步探究花色苷合成途径在野生稻非显色器官中执行功能的基本情况, 本研究通过对正常状态下药用野生稻叶转录组进行重测序的方法, 定性定量地调查了花色苷合成途径中8个关键基因的序列特点和表达变化。研究结果为更好地了解花色苷合成途径在稻属内的进化模式, 以及今后从花色苷合成途径入手对栽培稻进行品种改良等提供了必要的分子数据。

材料与方法

1 植物材料

用于本研究的药用野生稻(*Oryza officinalis* Wall. ex Watt) (登记号104973)种子来自国际水稻研究所(International Rice Research Institute, IRRI), 经50°C恒温处理5 d打破休眠, 然后将去壳种子置于种子萌发器内进行培养(25°C), 至胚根长至1 cm后移入温室花盆, 保持40%的湿度及28°C/22°C的昼夜温度下培养至成年个体。3株相同编号样品被设为一组生物学重复。

2 实验方法

2.1 总RNA提取与纯化

从3株长势良好的野生稻成年个体上各剪取33 mg新鲜叶子混合后置于液氮中迅速研磨成粉末, 再利用RNAprep Pure Plant Kit (cat No. DP441, TIANGEN)提取植物总RNA。取1 μL RNA样品在Agilent 2100 Bioanalyzer上分析确定所提RNA的浓度与质量。根据TruSeq RNA Sample Preparation Kit (RS-122-2001, Illumina)的操作步骤构建RNA文库, 并用Agilent 2100 Bioanalyzer检测文库的质量并通过Qubit和qPCR来扩增文库的数量。

2.2 cDNA链合成与转录组测序

用带有Oligo d(T)的beads对总RNA中mRNA进行纯化。将高质量的mRNA置于95°C高温10 min, 使其断裂成200 bp左右的片段, 然后利用Reverse Transcription Kit (cat No. M1705, Promega),

按照说明书将这些片段化的mRNA逆转录生成双链cDNA;末端补齐后,3'端加A,连接Illumina的双接头。AMPure XP beads进行两次纯化后,再进行12个循环的PCR扩增,然后利用2%琼脂糖凝胶电泳分离扩增产物,割胶回收长度在300~350 bp范围内的条带,并利用Min Elute Gel Extraction Kit (cat No. 28604, QIAGEN)进行纯化,得到双接头文库。利用2100 Bioanalyzer High Sensitivity Chip质控文库片段大小,并利用KAPA qPCR试剂盒(KAPA SYBR® FAST qPCR Kit Master Mix (2×) Universal, cat No. KK4602)质控文库准确浓度($\text{nmol}\cdot\text{L}^{-1}$),最后取10 μL 2 $\text{nmol}\cdot\text{L}^{-1}$ 的文库,在博奥生物有限公司的帮助下完成Illumina HiSeq 2000的转录组测序。

3 数据分析

根据GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>)中注释的栽培稻花色苷合成途径关键酶的基因*CHS* (登记号NM_001074541.1,下同)、*CHI* (NM_001072522.1)、*F3H* (NM_001068593.1)、*F3'H* (NM_001070873.1)、*F3'5'H* (AP004704.3)、*DFR* (NM_001052113.1)、*ANS* (AK067647.1)和*UFGT* (GQ280269.1)的mRNA作为参考序列,将过滤得到的高质量读段(reads)片段利用Bowtie2 2.2.9软件(Langmead和Salzberg 2012)进行对位排列(mapping)。由于药用野生稻与栽培稻的基因间会有一些的分歧,所以把参数mismatch从1向上依次设置到饱和,根据设置的要求进行精确的对位排列,输出SAM格式的文件,再利用SAMtools 1.3.1软件(Li等2009)将无法匹配的读段删除,并对每个基因的匹配读段进行计数得到可以匹配的读段数。筛出匹配到每个基因上的全部读段,将其拼接得到药用野生稻花色苷合成途径中基因编码序列的全长mRNA,通过SeaView软件(Gouy等2010)将其与栽培稻的同源序列进行比对。

同时,为比较不同基因的相对表达量,我们按照公式 $\text{RPKM}=10^9C/(NL)$ 计算每个基因的每百万读段中每千个碱基长度的读段数(reads per kilobase per million mapped reads, RPKM)来确定基因表达的丰度,式中*C*代表能够匹配到各基因上的读段数,*N*代表转录组总的匹配读段数,*L*代表基因长度;通常认为RPKM值在0.1~3.75之间的是低表达基因,在3.75~15之间的为中度表达基因,>15的为高表达基因(Mortazavi等2008)。

此外,为进一步检测各基因是否在药用野生稻内经受选择驱动,我们以全基因组数据已经释放的稻属*FF*基因组的短花稻作为外类群,利用*K_aK_s* Calculator软件(Zhang等2008)在GY模型(Goldman和Yang 1994)下统计和计算各基因碱基置换情况,确定它们的非同义替换率(non-synonymous substitutions rate, *K_a*)与同义替换率(synonymous substitutions rate, *K_s*)的比值并通过Fisher检验测试该值,进行中性检验,确定选择动力类型,依据为:*K_a/K_s*>1为正选择,*K_a/K_s*=1为中性选择,而*K_a/K_s*<1则为负选择或纯化选择。

实验结果

1 药用野生稻叶花色苷合成途径中关键酶基因序列特点

药用野生稻的叶转录组重测序后共得到大约 2.3×10^7 个原始读段,每个读段的长度为100 bp,其中能定位到已知参考基因组和基因转录本上的高质量读段约为 2.0×10^7 个,最终能够匹配到栽培稻花色苷合成代谢途径下不同结构基因转录本的读段共有9 503个,共涉及8个基因(*CHS*、*CHI*、*F3H*、*F3'H*、*F3'5'H*、*DFR*、*ANS*和*UFGT*)。

经过进一步拼接,我们得到了药用野生稻花色苷合成途径中这8个关键酶基因的完整mRNA序列(GenBank登记号KU572390~KU572397)。这8个基因mRNA的长度均在1~3 kb之间,最长的*DFR*基因长2 819 bp,最短的*CHI*基因长1 064 bp。当将这8个基因同它们在栽培稻中的同源基因进行比对后,发现两物种同源基因对在mRNA水平上的序列相似性为94%~96%,其翻译后得到氨基酸的序列同源率为93%~99% (图2)。我们也对药用野生稻中这8个基因的碱基组成进行了统计和分析,其中*CHI*序列的GC含量最低(46.6%),*UFGT*的GC含量最高(63.3%),其余基因的GC含量均在二者之间,但是这8个基因所对应编码区的GC含量都有所增加,最高的*DFR*基因编码区的GC含量高达70.2% (表1)。

2 基因的表达差异分析

为了比较各基因的表达差异,我们对每个花色苷基因读段的匹配情况进行了分析,结果发现,在8个基因中,匹配到*F3'5'H*基因的读段最多,达到3 162个,其次为*CHS* (2 220个),而*F3'H*最少,仅有262个。考虑到基因的长度有可能影响读段的匹

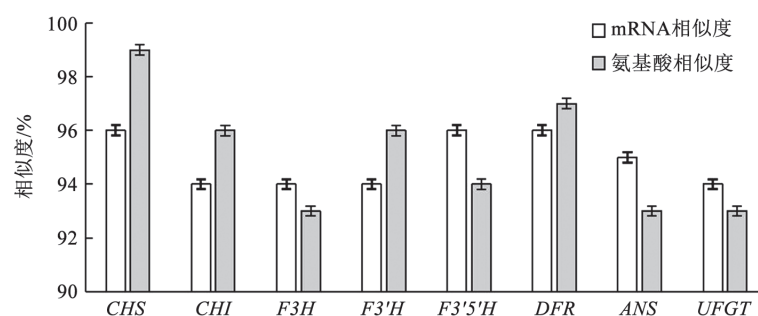


图2 药用野生稻和栽培稻中8个基因的同源性比较

Fig.2 Homologous comparisons of 8 genes from *O. officinalis* and *O. sativa*

表1 8个基因的序列特点

Table 1 Sequence characteristics of 8 genes

基因名	长度/bp	GC含量/%	开放阅读框			
			起点/bp	终点/bp	长度/bp	GC含量/%
CHS	1 761	60.1	1 366	53	1 314	67.6
CHI	1 064	46.6	43	933	891	48.0
F3H	1 346	58.5	104	1 150	1 047	61.6
F3'H	1 860	61.7	26	1 600	1 575	66.0
F3'5'H	2 029	61.9	1 901	351	1 551	68.9
DFR	2 819	55.8	1 127	126	1 002	70.2
ANS	1 564	50.8	1 475	360	1 116	55.2
UFGT	1 884	63.3	1 838	402	1 437	69.5

配数目, 因此我们采用RPKM法将不同基因的长度计算在内以校正各基因的相对表达量, 从而更好地评估药用野生稻中8个基因的相对表达水平。图3为各个花色苷合成途径中的关键结构基因的相对表达情况, 可以看出: *F3'5'H*的相对表达量最高, *CHS*和*CHI*次之, 但仍处于较高的表达水平, 随后为*F3H*、*ANS*和*UFGT*, 以上均属于高表达基因; *DFR*基因的RPKM值为14.7, 而*F3'H*的表达量最低, RPKM值为7.7, 都在3.75~15之间, 属于中度表达基因。

3 基因选择压分析

为了明确8个基因在药用野生稻中的选择驱动力, 我们检测了8个基因与外类群短花稻的同源基因(GenBank中对应的登记号分别为: *CHS*, XP_006662967.1; *CHI*, XP_006662673.1; *F3H*, XP_006660221.1; *F3'H*, XP_006665050.1; *F3'5'H*, XP_015695691.1; *DFR*, XP_006648840.1; *ANS*, XP_006651534.1; *UFGT*, XP_006656897.1)之间的 K_s 和 K_a , 发现基于GY这种统计模型计算出一一致的

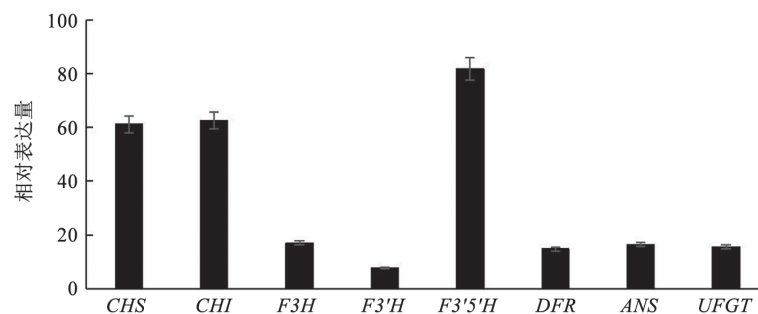


图3 药用野生稻花色苷合成途径中8个基因的相对表达情况

Fig.3 Relatively expression of 8 genes of anthocyanin biosynthetic pathway in *O. officinalis*

结果, 即这8个基因的 K_a/K_s 值全部显著小于1 (表2), 表明这8个关键酶基因在药用野生稻和栽培稻内均经历了严格的纯化选择。

表2 各花色苷合成基因在药用野生稻和外类群间的基因置换情况

Table 2 Gene substitution status in *O. officinalis* and outgroup

基因名	方法	K_a	K_s	K_a/K_s	P值 (Fisher)
<i>CHS</i>	GY-HKY	0.018	1.118	0.016	1.281×10^{-83}
<i>CHI</i>	GY-HKY	0.050	0.317	0.158	2.720×10^{-14}
<i>F3H</i>	GY-HKY	0.088	0.473	0.186	1.712×10^{-24}
<i>F3'H</i>	GY-HKY	0.060	0.696	0.086	1.462×10^{-76}
<i>F3'5'H</i>	GY-HKY	0.075	0.366	0.205	2.571×10^{-15}
<i>DFR</i>	GY-HKY	2.600×10^{-7}	5.900×10^{-7}	0.441	0
<i>ANS</i>	GY-HKY	0.144	0.519	0.277	3.847×10^{-22}
<i>UFGT</i>	GY-HKY	0.101	0.642	0.158	1.115×10^{-43}

讨 论

在本研究中, 我们从药用野生稻(登记号104973)的转录组中得到了8个花色苷代谢途径的关键结构基因, 包括*CHS*、*CHI*、*F3H*、*F3'H*、*F3'5'H*、*DFR*、*ANS*和*UFGT*, 将其与栽培稻中的同源基因进行比对后发现, 虽然各个同源基因对间具有较高的序列同源性, 但是其在物种间仍保持着不同程度的碱基变异, 这些变异体现在核苷酸水平是4%~6%的序列差异, 由于药用野生稻具有不同于栽培稻AA基因组的另外一个基础基因组, 即CC基因组, 因此, 这些序列上的差异有可能与基因组的特异性有关, 这为我们今后进一步利用这些差异对稻属不同基因组进行分子鉴别提供了可能。此外, 通过和外类群同源基因对之间的对比, 我们发现这8个基因全部处于严格的纯化选择之下($K_a/K_s < 1$) (表2), 说明这些碱基的变异并没有受到正向选择的改变, 充分显示了这些基因在功能上仍然具有高度的保守性。

同时, 我们也分析了这8个关键基因在药用野生稻中的相对表达情况(图3), 其中*F3'5'H*在叶中的相对表达量最高, *F3'H*的表达量最低。Huang等(2012)指出*F3'5'H*与*F3'H*不仅编码类黄酮代谢途径中相关的酶, 它们也是调节下游花色苷及类黄酮合成的“分水岭”。Ryu等(1998)分析了10个色稻品种, 认为在色稻籽粒中相对于*F3'5'H*来说, *F3'H*的表达占优势, 使得下游途径由二氢槲皮素(dihy-

droquercetin, DHQ)经过*DFR*、*ANS*以及*UFGT*基因的催化而合成花色苷。但在本研究中所用到的组织材料是新鲜的旗叶, 研究的组织器官不同, 而且花色苷及类黄酮的积累具有组织特异性(Robbins等2003), 因此可以预测在野生稻的绿叶中以二氢杨梅素(dihydromyricetin, DHM)为底物的途径也有可能发挥重要作用。另外值得一提的高表达基因为*CHS*和*CHI*, 这两个基因编码的酶除了参与花色苷合成外, 在高等植物类黄酮整个代谢网络中均作为关键酶催化最初的起始反应(Druka等2003; Reddy等1996), 因此, 作为早期的合成基因, 提供充分的代谢中间产物推动主体和不同代谢支路的顺畅运行可能是其高效表达的主要原因。而对于花色苷合成途径中的晚期合成基因如*DFR*、*ANS*和*UFGT*, 其明显降低的表达水平可能与下游基因的功能局限性相关, 由于这些基因更多在合成花色苷终产物上起关键作用, 而绿色幼叶中需要的花色苷含量往往较低, 因此基因表达水平相对上游基因总体偏低。总之, 通过对药用野生稻的绿叶转录组的全面分析, 我们获得了8个和花色苷合成相关基因的完整编码序列和各基因的表达特点, 这项工作为今后从花色苷物质防御角度提升栽培稻的遗传品质提供了最基本的分子数据。但遗憾的是, 由于材料的局限性, 我们没有能够获取该物种在花色苷代谢途径下有关转录因子的序列信息。由于花色苷代谢途径不但在转录水平上受多种转录因子的组合调控, 而且有些还要经历转录后水平的调节(Lepiniec等2006; Koes等2005), 因此, 后续对药用野生稻转录因子的深入了解和研究将有助于我们更深入地理解稻属花色苷合成途径的调控机理, 从而更好地认识花色苷及其他类黄酮类物质对于稻属植物正常生长的重要性。

参考文献

- Akada S, Dube SK (1995). Organization of soybean chalcone synthase gene clusters and characterization of a new member of the family. *Plant Mol Biol*, 29 (2): 189–199
- CAS Editorial Committee of the Flora of China (2002). *Flora of China*. 9 (2). Beijing: Science Press, 3–5 (in Chinese) [中国科学院中国植物志编辑委员会(2002). 中国植物志. 9 (2). 北京: 科学出版社, 3–5]
- Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, Bai Z, Goicoechea JL, Liang C, et al (2013). Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat Commun*, 4: 1595

- Druka A, Kudrna D, Rostoks N, Brueggeman R, von Wettstein D, Kleinhofs A (2003). Chalcone isomerase gene from rice (*Oryza sativa*) and barley (*Hordeum vulgare*): physical, genetic and mutation mapping. *Gene*, 302 (1–2): 171–178
- Fang H, Jiao Z (2012). Advances in anthocyanin biosynthesis and metabolic engineering. *Jiangsu Agric Sci*, 40 (7): 5–10 (in Chinese) [房欢, 焦焯(2012). 花色苷生物合成及代谢工程研究进展. *江苏农业科学*, 40 (7): 5–10]
- Franken P, Niesbach-Klösigen U, Weydemann U, Maréchal-Drouard L, Saedler H, Wienand U (1991). 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*. *EMBO J*, 10 (9): 2605–2612
- Goldman N, Yang Z (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*, 11 (5): 725–736
- Gouy M, Guindon S, Gascuel O (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*, 27 (2): 221–224
- Hayashi M, Takahashi H, Tamura K, Huang J, Yu LH, Kawai-Yamada M, Tezuka T, Uchimiya H (2005). Enhanced dihydroflavonol-4-reductase activity and NAD homeostasis leading to cell death tolerance in transgenic rice. *Proc Natl Acad Sci USA*, 102 (19): 7020–7025
- Huang W, Sun W, Wang Y (2012). Isolation and molecular characterization of flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase genes from a traditional Chinese medicinal plant, *Epimedium sagittatum*. *Gene*, 497 (1): 125–130
- Jacquemin J, Bhatia D, Singh K, Wing RA (2013). The International *Oryza* Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Curr Opin Plant Biol*, 16 (2): 147–156
- Koes R, Verweij W, Quattrocchio F (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci*, 10 (5): 236–242
- Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9 (4): 357–359
- Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M (2006). Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol*, 57: 405–430
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25 (16): 2078–2079
- Lin XX, Zhu SM (2005). Progress in research on anthocyanidin components in Chinese traditional medicine against neoplasm. *Chin J Chin Mater Med*, 30 (15): 1147–1150 (in Chinese with English abstract) [林晓霞, 朱寿民(2005). 中药花色苷类成分抗癌研究进展. *中国中药杂志*, 30 (15): 1147–1150]
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 5 (7): 621–628
- Nogués S, Allen DJ, Morison JIL, Baker NR (1998). Ultraviolet-B radiation effects on water relations, leaf development, and photosynthesis in droughted pea plants. *Plant Physiol*, 117 (1): 173–181
- Pang Z (2000). The overview of anthocyanin research. *Beijing Agric Sci*, 18 (5): 37–42 (in Chinese) [庞志申(2000). 花色苷研究概况. *北京农业科学*, 18 (5): 37–42]
- Peng SS, Yu HH, Zhao ZX (2005). Plant food with antioxidant effect. *Food Sci Technol*, (1): 44–48 (in Chinese) [彭珊珊, 于化泓, 赵哲霞 (2005). 具有抗氧化效果的植物性食品. *江西食品工业*, (1): 44–48]
- Petroni K, Tonelli C (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci*, 181 (3): 219–229
- Reddy AR, Scheffler B, Madhuri G, Srivastava MN, Kumar A, Sathyanarayanan PV, Nair S, Mohan M (1996). Chalcone synthase in rice (*Oryza sativa* L.): detection of the CHS protein in seedlings and molecular mapping of the *chs* locus. *Plant Mol Biol*, 32 (4): 735–743
- Robbins MP, Paolocci F, Hughes JW, Turchetti V, Allison G, Arcioni S, Morris P, Damiani F (2003). *Sn*, a maize bHLH gene, modulates anthocyanin and condensed tannin pathways in *Lotus corniculatus*. *J Exp Bot*, 54 (381): 239–248
- Ryu SN, Park SZ, Ho CT (1998). High performance liquid chromatographic determination of anthocyanin pigments in some varieties of black rice. *J food Drug Anal*, 6: 729–736
- Shih CH, Chu H, Tang LK, Sakamoto W, Maekawa M, Chu IK, Wang M, Lo C (2008). Functional characterization of key structural genes in rice flavonoid biosynthesis. *Planta*, 228: 1043–1054
- Smillie RM, Hetherington SE (1999). Photoabatement by anthocyanin shields photosynthetic systems from light stress. *Photosynthetica*, 36 (3): 451–463
- Tsuda T, Shiga K, Ohshima K, Kawakishi S, Osawa T (1996). Inhibition of lipid peroxidation and the active oxygen radical scavenging effect of anthocyanin pigments isolated from *Phaseolus vulgaris* L. *Biochem Pharmacol*, 52 (7): 1033–1039
- Winkel-Shirley B (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol*, 126 (2): 485–493
- Zhang HY, Chen DZ (2000). Theoretical characterization and its application of free radical scavenging activity for phenolic antioxidants. *Acta Biophys Sin*, 16 (1): 1–9 (in Chinese with English abstract) [张红雨, 陈德展 (2000). 酚类抗氧化剂清除自由基活性的理论表征与应用. *生物物理学报*, 16 (1): 1–9]
- Zhang QJ, Zhu T, Xia EH, Shi C, Liu YL, Zhang Y, Liu Y, Jiang WK, Zhao YJ, Mao SY, et al (2014). Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *Proc Natl Acad Sci USA*, 111 (46): 4954–4962
- Zhang X, Szydlowski N, Delvallé D, D'Hulst C, James MG, Myers AM (2008). Overlapping functions of the starch synthases SSII and SSIII in amylopectin biosynthesis in *Arabidopsis*. *BMC Plant Biol*, 8: 96
- Zhang YZ, Xu SZ, Cheng YW, Ya HY, Han JM (2016). Transcriptome analysis and anthocyanin-related genes in red leaf lettuce. *Genet Mol Res*, 15 (1): gmr.15017023
- Zou XH, Du YS, Tang L, Xu XW, Doyle JJ, Sang T, Ge S (2015). Multiple origins of BBCC allopolyploid species in the rice genus (*Oryza*). *Sci Rep*, 5: 14876

Sequence characteristic and expression analysis of genes related to anthocyanin biosynthesis in *Oryza officinalis* leaf transcriptome

QIN Zong-Yan*, LIU Shi-Juan*, BAO Ying**

School of Life Sciences, Qufu Normal University, Qufu, Shandong 273165, China

Abstract: In order to completely understand member composition and expression of the genes about anthocyanin synthesis in green tissues of non-model plant, the leaf transcriptome of *Oryza officinalis* was re-sequenced, and the genes in the anthocyanin synthetic pathway were analyzed. A total of 8 genes (*CHS*, *CHI*, *F3H*, *F3'H*, *F3'5'H*, *DFR*, *UFGT* and *ANS*) were identified. These genes shared high-level sequence similarities (94%~96%) regarding to the mRNA with their homologous genes in cultivated rice, respectively. Moreover, the statistical tests of the ratio of non-synonymous substitutions rate to synonymous substitutions rate (K_a/K_s) relative to the outgroup show that the 8 genes were all influenced by the purified selection. In addition, reads per kilobase per million mapped reads (RPKM) comparison revealed a divergent expression among the 8 genes, among which *F3'5'H* had the highest expression level, following by *CHI* and *CHS*, and *F3'H* had the lowest expression level.

Key words: *Oryza officinalis*; leaf transcriptome; anthocyanin biosynthetic pathway; gene expression

Received 2016-05-10 Accepted 2016-07-21

This work was supported by the National Natural Science Foundation of China (Grant No. 31570218).

*Co-first authors.

**Corresponding author (E-mail:baoyingus@126.com).