

# 稻米贮藏蛋白家族的生物信息学分析

孙红正 郭凯 宋宁垣 赵全志\*

(河南农业大学农学院/河南粮食作物协同创新中心/河南省水稻生物学重点实验室, 郑州 450002;

第一作者: sunhongzheng@foxmail.com; \* 通讯作者: qzzhaoh@126.com)

**摘要:**贮藏蛋白是稻米中的第二大主要成分,是影响稻米食味品质和营养价值的重要因素。对水稻中谷蛋白、醇溶蛋白、清蛋白、球蛋白等 4 大类贮藏蛋白在基因组中的分布、基因结构、表达模式以及各个蛋白家族内不同拷贝的进化关系进行研究。结果表明,谷蛋白和醇溶蛋白在基因组上的拷贝数较多,分别有 15 个和 28 个基因拷贝检测到表达。谷蛋白家族可分为 3 个亚家族,醇溶蛋白分为 2 个亚家族,并根据不同拷贝的亲缘关系对未命名的基因进行命名。清蛋白和球蛋白拷贝数较少,清蛋白有 5 个拷贝,其中 SSA1、SSA5 表达量较高,其他拷贝表达相对较低,而球蛋白的 3 个拷贝表达量都极低。本文为水稻贮藏蛋白以及稻米品质的研究提供了理论参考。

**关键词:**水稻;贮藏蛋白;谷蛋白;醇溶蛋白;清蛋白;球蛋白

**中图分类号:**S511 **文献标识码:**A **文章编号:**1006-8082(2018)01-0052-06

水稻是我国重要的粮食作物,全国以稻米为主食的人口占总人口的 65%<sup>[1]</sup>。贮藏蛋白是大米的第二大主要成分,糙米中蛋白质含量占 8%~11%<sup>[2]</sup>,精米中蛋白质含量占 5.5%~12%<sup>[3]</sup>。稻米中的蛋白质根据其在不同溶剂中的溶解性不同可分为清蛋白、球蛋白、醇溶蛋白和谷蛋白。其中,清蛋白溶于水,球蛋白不溶于水但溶于稀盐溶液,醇溶蛋白不溶于水和稀盐而溶于乙醇,而谷蛋白则只能溶于酸或碱溶液,根据其溶解性不同可以将这 4 种蛋白依次提取出来。贮藏蛋白在稻米中的分布是不均匀的,球蛋白、清蛋白主要分布在糊粉层和次糊粉层,谷蛋白和醇溶蛋白主要分布于胚乳外层部分<sup>[4]</sup>。水稻精米中谷蛋白占 70%~80%,醇溶蛋白占 18%~20%<sup>[5]</sup>。贮藏蛋白于内质网合成,然后转运至不同蛋白体中,在水稻胚乳中,贮藏蛋白主要以蛋白体 PB-I、PB-II 的形式存在,PB-I 蛋白体是较为光滑的球形蛋白体,主要含难以被人体消化吸收的醇溶蛋白,PB-II 蛋白体则是性状不规则的非球形结构,主要含易消化的谷蛋白<sup>[6-7]</sup>。大量研究表明,稻米中蛋白质含量与稻米食味品质之间存在负相关关系,蛋白质含量过高的稻米食味品质往往较差<sup>[8]</sup>。但是,也有一些研究认为,稻米品质不仅与蛋白质含量有关,与谷蛋白和醇溶蛋白的比例也有关系<sup>[3,5,9]</sup>。目前,关于稻米中 4 类贮藏蛋白质的遗传及表达模式相关研究较少,关于水稻贮藏蛋白的命名及分类,不同的文献中对各个基因位点的命名不尽相同,甚至出现把谷蛋白基因与球蛋白基因、清蛋白基因与球蛋白基因混淆进行研究<sup>[10-14]</sup>。本文以水稻基

因组注释数据库和基因表达数据库为依据,分析水稻中 4 类贮藏蛋白在基因组的分布、不同拷贝之间的进化关系及其表达模式,对各个贮藏蛋白家族的分类及表达进行整理,为进一步研究稻米贮藏蛋白提供理论参考和基础。

## 1 材料与方法

### 1.1 水稻贮藏蛋白基因序列的获取

根据 RGAP(Rice Genome Annotation Project, <http://rice.plantbiology.msu.edu/>) 水稻基因组注释信息从 RGAP 数据库下载谷蛋白、醇溶蛋白、清蛋白和球蛋白基因序列<sup>[15]</sup>。基因在染色体上的位置及基因结构信息通过 RGAP 基因组浏览器获取。

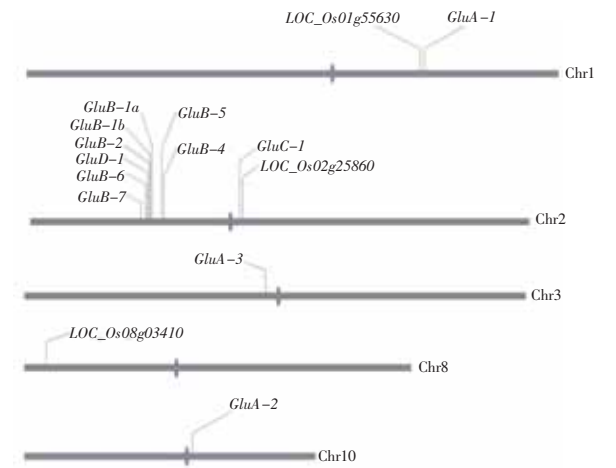
### 1.2 水稻谷蛋白与醇溶蛋白基因进化树的构建

谷蛋白与醇溶蛋白基因家族拷贝数较多,获取谷蛋白与醇溶蛋白基因编码序列后使用 MEGA6.06 软件对序列进行多重序列比对<sup>[16]</sup>,然后使用 Neighbor-Joining 法构建系统进化树,建树过程中碱基替换模型使用泊松分布模型,对多重比对中的空格采用 pair-wise deletion 处理,进化树的可靠性采用 500 次的 bootstrap 进行检验。

### 1.3 水稻贮藏蛋白基因的表达分析

谷蛋白、醇溶蛋白、清蛋白和球蛋白 4 类贮藏蛋白基因的表达数据从 RGAP Rice Gene Expression 数据库

收稿日期:2017-10-31



染色体示意图中线条长度代表染色体大小,灰色椭圆代表着丝粒,竖线位置指示基因所在染色体相应位置。

图1 水稻谷蛋白在基因组上的分布

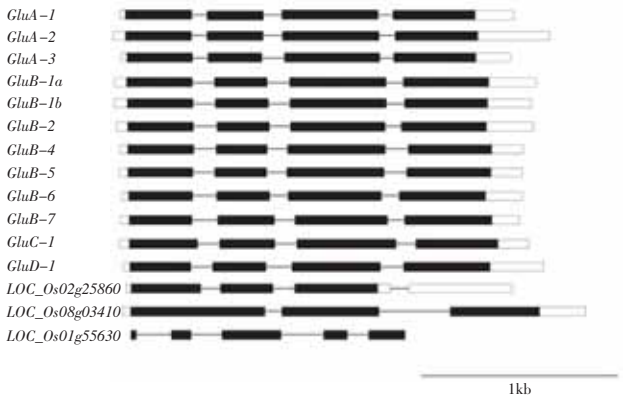
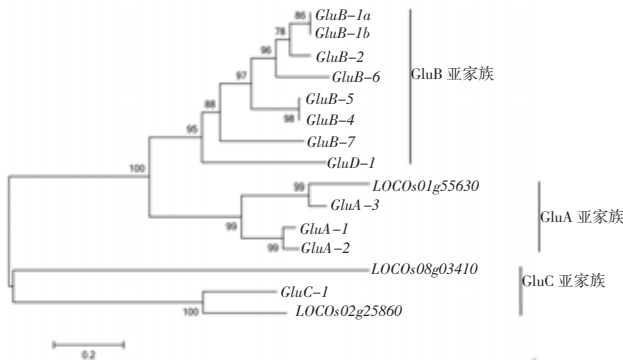


图2 水稻谷蛋白基因的基因结构示意图



进化树采用 NJ 法构建, 枝长上的数字代表 Bootstrap 值, 为 500 次 bootstrap 中树形结构一致的百分比例。

图3 水稻谷蛋白基因的系统进化树

(<http://rice.plantbiology.msu.edu/expression.shtml>)下载。

2 结果与分析

2.1 谷蛋白 (Glutelin)

在 RGAP 基因组注释数据库中, 有 15 个表达的基

因被注释为谷蛋白基因, 在这 15 个基因中, 有 12 个基因被命名分别属于 GluA、GluB、GluC 和 GluD 四种类型, 3 个基因未命名但属于谷蛋白家族。谷蛋白基因在水稻基因组中分别位于第 1、第 2、第 3、第 8 和第 10 染色体上, 第 1 染色体 2 个拷贝, 第 2 染色体 10 个拷贝, 第 3、第 8、第 10 染色体上各 1 个拷贝 (见图 1)。其中, 第 2 染色体上 10 个拷贝成簇分布, 8 个拷贝位于第 2 染色体短臂上, 分别是 GluB 类型 7 个拷贝和 GluD 类型 1 个拷贝。长臂靠近着丝粒位置 2 个拷贝分别是 GluC-1 和 LOC\_Os02g25860。

基因结构方面, 在 15 个谷蛋白基因中, 12 个已命名的谷蛋白基因均有 4 个编码外显子区, 编码氨基酸从 236~531 个氨基酸不等。LOC\_Os02g25860 有 4 个外显子但仅有 3 个编码外显子, 编码区长度与 GluA、GluB、GluC 和 GluD 基因相比明显变短。LOC\_Os08g03410 有 3 个外显子, 编码外显子区也有 3 个, 编码区长度与已命名谷蛋白基因相当。LOC\_Os01g55630 则有 5 个外显子, 编码长度与其他谷蛋白基因相比也较短 (见图 2)。

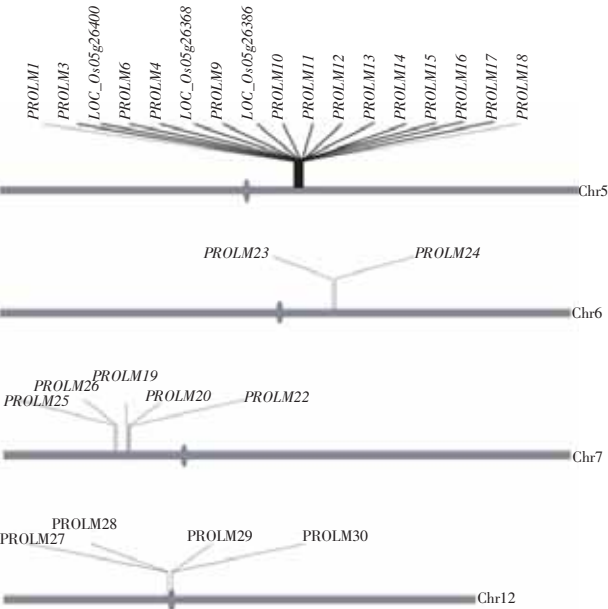
在谷蛋白基因的系统进化树中, 谷蛋白基因家族分为 3 大支: GluA 亚家族、GluB 亚家族、GluC 亚家族。GluA 亚家族除了已命名的 GluA-1、GluA-2、GluA-3 外还包括 LOC\_Os01g55630。GluB 亚家族包括 GluB-1a、GluB-1b、GluB-2、GluB-3、GluB-4、GluB-5、GluB-6、GluB-7 和 GluD-1 (见图 3)。因此, 谷蛋白基因 GluD-1 虽然被命名为 GluD 亚家族, 但在亲缘关系上与 GluB 亚家族较为接近。GluC 亚家族包括 GluC-1 以及 2 个未命名的拷贝 LOC\_Os08g03410 和 LOC\_Os02g25860。

水稻谷蛋白基因的表达量基本表现为灌浆早期表达量较低, 而后表达量急剧升高, 到后期表达量又有所降低。花后 25 d 的籽粒中谷蛋白基因表达除了 LOC\_Os08g03410 明显表现为胚乳中表达量远高于胚中的表达量, 也说明谷蛋白基因的表达并不是严格的胚乳特异性表达。LOC\_Os08g03410 的表达模式与其他谷蛋白基因明显不同, 在花后 25 d, 胚中的表达量达到 36 316.4 RPKM (Read Per Kilobase per Million), 而胚乳中的表达量只有 990.4 RPKM, 说明 LOC\_Os08g03410 倾向于在胚中特异表达。15 个谷蛋白基因中, GluA-1 和 GluA-2 的表达水平最高, 花后 10 d 籽粒中的表达量分别达到 55 337.8 RPKM 和 38 553.9 RPKM。GluA-3、GluB-1a、GluB-1b、GluB-2、GluB-4、GluB-5、GluC-1 的表达量处于同一水平, 花后 10 d 表达量都在 15

表 1 水稻谷蛋白基因在籽粒中的表达量(RPKM)

基因位点	基因名称	花后 5 d 籽粒	花后 10 d 籽粒	花后 25 d 胚	花后 25 d 胚乳
LOC_Os01g55690	<i>GluA-1</i>	10 598.8	55 337.8	1 697.1	51 968.9
LOC_Os10g26060	<i>GluA-2</i>	8 197.1	38 553.9	1 110.7	35 193.4
LOC_Os03g31360	<i>GluA-3</i>	7 940.6	18 950.8	564.4	11 573.2
LOC_Os01g55630	( <i>GluA-5</i> )	201.5	508.9	14.9	453.4
LOC_Os02g15178	<i>GluB-1a</i>	5 741.2	15 635.0	564.1	13 685.8
LOC_Os02g15169	<i>GluB-1b</i>	5 836.8	15 856.8	571.7	13 866.4
LOC_Os02g15150	<i>GluB-2</i>	5 774.5	15 444.8	581.0	14 467.1
LOC_Os02g16830	<i>GluB-4</i>	5 052.7	18 487.8	585.3	13 643.5
LOC_Os02g16820	<i>GluB-5</i>	5 214.2	19 423.9	634.2	14 483.5
LOC_Os02g15070	<i>GluB-6</i>	303.7	779.1	29.7	334.4
LOC_Os02g14600	<i>GluB-7</i>	1 664.5	1 410.0	67.3	855.5
LOC_Os02g15090	<i>GluD-1</i>	731.0	3 706.1	160.5	4 258.5
	( <i>GluB-8</i> )				
LOC_Os02g25640	<i>GluC-1</i>	4 775.8	16 286.4	393.2	15 253.6
LOC_Os02g25860	( <i>GluC-3</i> )	43.4	62.6	1.7	19.2
LOC_Os08g03410	( <i>GluC-4</i> )	332.3	687.2	36 316.4	990.4

基因名称中, 括号内基因名称为本研究建议命名名称。



染色体示意图中线条长度代表染色体大小, 灰色椭圆代表着丝粒, 竖线位置指示基因所在染色体相应位置。

图 4 水稻醇溶蛋白基因在基因组上的分布

000~20 000 RPKM 之间。*GluB-7*、*GluB-6*、*GluD-1*、*LOC\_Os01g55630*、*LOC\_Os02g25860* 的表达量相对较低(见表 1)。

2.2 醇溶蛋白(Prolamin)

水稻基因组注释数据库中有 28 个醇溶蛋白基因分别位于第 5、第 6、第 7 和第 12 染色体, 其中 17 个拷贝成簇分布于第 5 染色体, 其他染色体上, 第 7 染色体 5 个拷贝, 第 12 染色体 4 个拷贝, 第 6 染色体 2 个拷贝, 均呈簇状集中分布于染色体某一区段(图 4)。28 个醇溶蛋白基因中, 有 25 个以“PROLM+数字”命名, 有 3

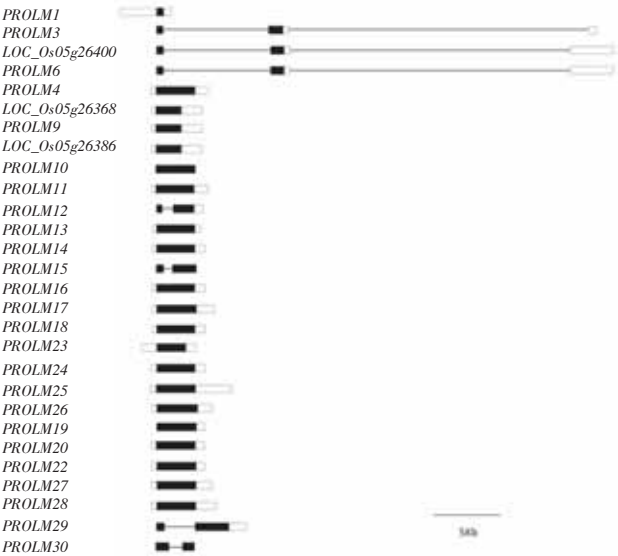
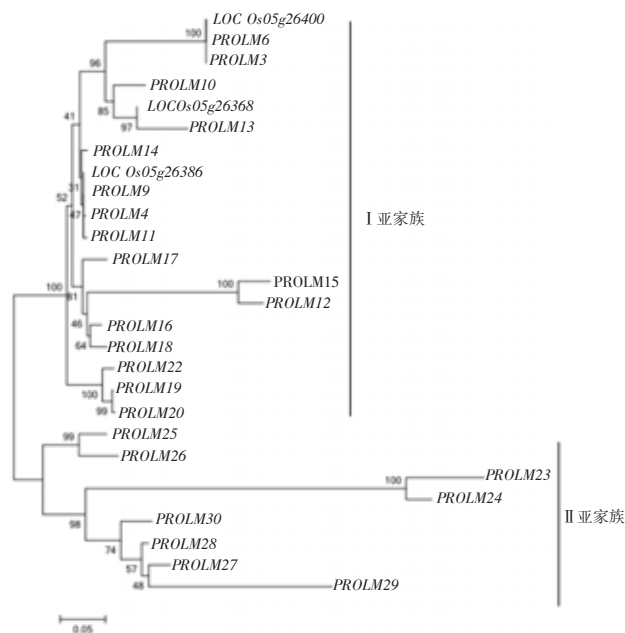


图 5 水稻醇溶蛋白基因的基因结构示意图

个拷贝 *LOC\_Os05g26400*、*LOC\_Os05g26368*、*LOC\_Os05g26386* 未被命名。其中 *PROLM3*、*LOC\_Os05g26400*、*PROLM6* 在基因组上的位置完全相同, *LOC\_Os05g26400* 与 *PROLM6* 完全相同, 而 *PROLM3* 与这两者编码区相同但转录本不同, 属于可变剪接版本。

在 28 个醇溶蛋白基因中, 有 7 个拷贝存在内含子结构, 且均有 2 个编码区外显子, 其他 21 个拷贝均无内含子, 说明该基因家族可能是基因反转录插入重复的结果(见图 5)。*PROLM1* 编码区最短, 仅编码 24 个氨基酸, 其他拷贝编码氨基酸在 89~156 个氨基酸区间, 相对谷蛋白编码氨基酸较小。

在进化关系上, 28 个醇溶蛋白基因家族可分为两



进化树采用 NJ 法构建, 枝长上的数字代表 Bootstrap 值, 为 500 次 bootstrap 中树形结构一致的百分比例。

图 6 水稻醇溶蛋白基因的系统进化树

大亚家族, I 亚家族 19 个拷贝, 包括第 5 染色体成簇分布的除 *PROLM1* 外的其他 16 个拷贝和第 7 染色体 *PROLM19*、*PROLM20*、*PROLM22* 3 个拷贝。II 亚家族包括 *PROLM1*、*PROLM23*、*PROLM24*、*PROLM25*、*PROLM26*、*PROLM27*、*PROLM28*、*PROLM29*、*PROLM30* 共 9 个拷贝 (见图 6)。其中, *PROLM1* 序列较短, 整条序列与 II 亚家族的末端编码序列有同源性, 因此, 将 *PROLM1* 归入 II 亚家族。

在 28 个醇溶蛋白基因中, *PROLM22*、*PROLM24*、*PROLM25*、*PROLM26* 的表达量处于较高水平, 花后 25 d 胚乳中表达量均在 10 000 RPKM 以上, 最高的 *PROLM26* 表达量接近 59 998.8 RPKM, 因此, 这 3 个拷贝是分离胚乳高效表达启动子较好的候选基因。*PROLM3*、*LOC\_Os05g26400*、*PROLM6*、*PROLM10*、*PROLM13* 的表达量极低, *PROLM9*、*LOC\_Os05g26386* 无表达, 其他拷贝则处于中等表达水平。在表达的醇溶蛋白拷贝中, 均表现出胚乳中表达量远高于胚中的表达量, 表现出倾向但并不严格的胚乳特异性表达。

### 2.3 清蛋白 (Albumin) 和球蛋白 (Globulin)

水稻清蛋白和球蛋白基因在基因组的拷贝数相对较少, 清蛋白基因有 5 个拷贝被检测到表达: *SSA1*、*SSA2*、*SSA3*、*SSA4* 和 *SSA5*, 清蛋白基因 *SSA2*、*SSA3* 和 *SSA4* 在第 3 染色体成簇分布, 其他 2 个基因分别位于第 5 和第 11 染色体。在基因结构上 *SSA2* 有内含子, 其

他 4 个拷贝无内含子。清蛋白基因编码氨基酸从 69~361 个氨基酸不等。5 个清蛋白拷贝中, *SSA1* 和 *SSA5* 的表达量较高, *SSA2*、*SSA3* 表达量极低 (见表 3)。清蛋白占水稻籽粒中总蛋白的 5% 左右, 并且大部分清蛋白翻译后经过糖基化修饰, 在籽粒中最高表达量出现在花后 18~20 d<sup>[17]</sup>。

球蛋白基因有 3 个拷贝: *globulin 2*、*7S globulin* 和 *7S globulin 2*, 分别位于第 1、第 5 和第 11 染色体。3 个拷贝中, *7S globulin* 和 *7S globulin 2* 的基因结构中无内含子, 表达水平也较低, 而 *globulin 2* 在胚乳中没有检测到表达, 仅在花后 25 d 的胚中检测到极低水平表达 (见表 3)。

## 3 结论与讨论

前人研究将水稻谷蛋白分为 4 个亚家族 GluA、GluB、GluC 和 GluD。GluD-1 (*LOC\_Os02g15090*) 是 GluD 亚家族唯一的拷贝, 该基因由 Kawakatsu 等人于 2008 年命名<sup>[10]</sup>。但是从谷蛋白基因系统进化关系 (图 3) 上看, *LOC\_Os02g15090* 与 GluB 亚家族基因聚成一支, 并明显区别于 GluC 亚家族, 因此本研究将该基因以 *GluB-8* 命名。牛洪斌等于 2007 年克隆 1 个谷蛋白基因并将其命名为 *GluB-7*<sup>[8]</sup>, 但是经过序列检索, 该序列 (AY987390) 与 *GluB-2* 的相似程度最高, 因此判断该序列应为 *GluB-2*。另外 3 个没有命名的基因位点中 *LOC\_Os01g55630* (*GluA-5*) 由于与 *GluA-3* 相似程度最高, 鉴于之前 Takaiwa 报道 *GluA-4* 为假基因<sup>[12]</sup>, 因此将 *LOC\_Os01g55630* 命名顺延为 *GluA-5*。由于 Kawakatsu 等认为 *GluC-2* 是假基因<sup>[10]</sup>, 因此 *LOC\_Os02g25860*、*LOC\_Os08g03410* 分别根据其在系统进化树上的位置与其他谷蛋白基因的亲缘关系分别命名为 *GluC-3*、*GluC-4*。

醇溶蛋白有 3 个拷贝 *LOC\_Os05g26400*、*LOC\_Os05g26368*、*LOC\_Os05g26386* 未被命名, 其中, *LOC\_Os05g26400* 与 *PROLM6*、*LOC\_Os05g26386* 与 *PROLM9* 这两对基因在基因组上的位置及转录本和编码区完全相同, 因此, *LOC\_Os05g26400*、*LOC\_Os05g26386* 这 2 个拷贝是错误注释造成的。另外一个醇溶蛋白基因 *LOC\_Os05g26368* 在系统进化关系上与 *PROLM13* 最近, 因此将 *LOC\_Os05g26368* 命名为 *PROLM13b*。醇溶蛋白基因拷贝数较多, 但大多数无内含子结构, 因此有可能是基因转录以后被反转录插入基因组的结果。醇溶蛋白翻译后形成的蛋白根据其在电泳图谱上的位置主要有 10 kDa、13 kDa 和 16 kDa 3



表 2 水稻醇溶蛋白基因在籽粒中的表达量(RPKM)

基因位点	基因名称	花后 5 d 籽粒	花后 10 d 籽粒	花后 25 d 胚	花后 25 d 胚乳
LOC_Os05g26240	<i>PROLM1</i>	8.4	107.5	8.9	234.1
LOC_Os05g26250	<i>PROLM3</i>	0.1	0	0	4.4
LOC_Os05g26350	<i>PROLM4</i>	125.6	1 575.3	92.3	3 315.3
LOC_Os05g26400	( <i>PROLM6</i> )	2.3	21.7	1.1	35.7
LOC_Os05g26359	<i>PROLM6</i>	1.8	1.9	1.0	2.7
LOC_Os05g26377	<i>PROLM9</i>	0	0	0	0
LOC_Os05g26386	( <i>PROLM9</i> )	0	0	0	0
LOC_Os05g26440	<i>PROLM10</i>	0.4	5.1	0.2	12.9
LOC_Os05g26460	<i>PROLM11</i>	63.4	439	17.3	822.3
LOC_Os05g26480	<i>PROLM12</i>	15.7	69.5	3.1	162.6
LOC_Os05g26490	<i>PROLM13</i>	1	4.5	0.5	9.5
LOC_Os05g26368	( <i>PROLM13b</i> )	20.5	114.8	9.5	200.3
LOC_Os05g26620	<i>PROLM14</i>	1.5	231.3	19.2	873.6
LOC_Os05g26690	<i>PROLM15</i>	40.3	324.0	12.7	752.9
LOC_Os05g26720	<i>PROLM16</i>	143.3	1 178.0	46.8	2 705.3
LOC_Os05g26750	<i>PROLM17</i>	184.5	602.5	43.3	1 192.3
LOC_Os05g26770	<i>PROLM18</i>	94.4	1 163.1	52.9	2 816.3
LOC_Os07g11900	<i>PROLM19</i>	1690	3 861.6	232.0	7 591.8
LOC_Os07g11910	<i>PROLM20</i>	1 902.8	5 051.8	283.8	8 878.9
LOC_Os07g11920	<i>PROLM22</i>	2 373.3	7744	388.7	13 021.7
LOC_Os06g31060	<i>PROLM23</i>	152.7	445.8	27.0	568.7
LOC_Os06g31070	<i>PROLM24</i>	7 054.4	9 912.1	668.4	14 520.0
LOC_Os07g10570	<i>PROLM25</i>	3 965.0	21 888.9	747.0	21 450.3
LOC_Os07g10580	<i>PROLM26</i>	9 594.1	52 046.3	1 689.0	59 998.8
LOC_Os12g16880	<i>PROLM27</i>	638.2	1 421.2	82.3	1 436.5
LOC_Os12g16890	<i>PROLM28</i>	3 155.0	7 236.5	395.9	6 671.8
LOC_Os12g17010	<i>PROLM29</i>	374.8	987.4	48.6	778.4
LOC_Os12g17030	<i>PROLM30</i>	168.9	368.6	15.1	260.0

基因名称中,括号内基因名称为本研究建议命名名称。

表 3 水稻清蛋白和球蛋白基因在籽粒中的表达量(RPKM)

基因位点	基因名称	花后 5 d 籽粒	花后 10 d 籽粒	花后 25 d 胚	花后 25 d 胚乳
LOC_Os03g55730	<i>SSA2</i>	38.3	15.7	25.8	13.5
LOC_Os03g55734	<i>SSA3</i>	315.8	95.8	0	15.2
LOC_Os03g55740	<i>SSA4</i>	203.7	33.2	0	1.5
LOC_Os05g41970	<i>SSA1</i>	9 887.5	19 339.0	910.1	25 524.9
LOC_Os11g33000	<i>SSA5</i>	10 258.7	9 233.9	302.8	4 696.3
LOC_Os01g71094	<i>7S globulin 2</i>	3.1	2.7	1.7	1.0
LOC_Os05g33400	<i>7S globulin</i>	5.8	1.4	2.7	4.1
LOC_Os11g34780	<i>globulin 2</i>	0	0	3.7	0

类蛋白,不同种类醇溶蛋白的表达对蛋白体 PB-I 的形成起重要作用<sup>[19]</sup>。

水稻籽粒中,处于高水平表达量(>10 000 RPKM)的谷蛋白基因有 10 个,而醇溶蛋白基因处于高水平表达的基因也有 4 个拷贝,因此,水稻贮藏蛋白基因是分离高效籽粒表达启动子的理想候选基因。Qu 等<sup>[11]</sup>曾研究比较了 6 个水稻谷蛋白启动子在水稻胚乳中的表达特异性及启动活性。在应用方面,前人曾用水稻谷蛋白启动子 Gt13a 在水稻胚乳中特异表达人血清白蛋白和人粒细胞巨噬细胞集落刺激因子<sup>[20-21]</sup>,而 Gt13a 胚乳特异启动子即为 *GluA-1*(*LOC\_Os01g55690*)的启动子。从

谷蛋白基因的表达量来看,*GluA-1*、*GluA-2*、*GluA-3*、*GluB-1a*、*GluB-1b*、*GluB-2*、*GluB-4*、*GluB-5*、*GluC-1* 这几个谷蛋白基因在胚乳中的表达量都相对较高,是克隆高效胚乳特异启动子的理想候选基因。而本研究中命名的 *GluC-4*(*LOC\_Os08g03410*)谷蛋白基因在花后 25 d 胚中的表达量远高于胚乳中的表达量(见表 1),因此该谷蛋白基因启动子可以作为候选胚特异启动子来研究胚的发育进程及其对籽粒灌浆的影响。除了谷蛋白基因,醇溶蛋白基因中,*PROLM22*、*PROLM24*、*PROLM25*、*PROLM26* 的表达量也比较高,特别是 *PROLM26* 在花后 25 d 的表达量达到将近 60 000 RP-

KM, 比 *GluA-1* 的启动活性还要高, 因此也是比较理想的候选高效胚乳特异启动子。

稻米中蛋白质含量是影响稻米品质的重要因素之一, 而且蛋白质组分也会影响稻米品质。精米中蛋白质主要以谷蛋白和醇溶蛋白为主<sup>[2]</sup>, 研究者大都认为醇溶蛋白对稻米食味品质有负面影响<sup>[4]</sup>。稻米中蛋白质的合成处于动态平衡状态, 在某一贮藏蛋白被抑制的情况下, 其他贮藏蛋白会在转录和翻译水平上进行补偿, 抑制醇溶蛋白表达能够提高稻米的营养价值<sup>[7,23]</sup>, 因此, 可以通过筛选低醇溶蛋白的品种或采用基因工程的手段抑制醇溶蛋白以提高稻米的营养和食味品质。

### 参考文献

- [1] 胡忠孝. 中国水稻生产形势分析[J]. 杂交水稻, 2009, 24(6): 1-7.
- [2] 陈能, 罗玉坤, 谢黎虹, 等. 我国水稻品种的蛋白质含量及与米质的相关性研究[J]. 作物学报, 2006, 32(8): 1 193-1 196.
- [3] 张欣, 施利利, 丁得亮, 等. 稻米蛋白质相关性状与 RVA 特征谱及食味品质的关系[J]. 食品科技, 2014, 39(10): 188-191.
- [4] 张启莉, 谢黎虹, 李仕贵, 等. 稻米蛋白质与蒸煮食味品质的关系研究进展[J]. 中国稻米, 2012, 18(4): 1-6.
- [5] 吴洪恺, 刘世家, 江玲, 等. 稻米蛋白质组分及总蛋白质含量与淀粉 RVA 谱特征值的关系 [J]. 中国水稻科学, 2009, 23 (4): 421-426.
- [6] 王继馨, 张云江, 程爱华, 等. 水稻蛋白亚基含量对米饭食味的影响[J]. 中国农学通报, 2008, 24(1): 89-92.
- [7] Kim H J, Lee J Y, Yoon U H, et al. Effects of reduced prolamin on seed storage protein composition and the nutritional quality of rice[J]. *Int J Mol Sci*, 2013, 14(8): 17 073-17 084.
- [8] 王忠, 顾蕴洁, 陈刚, 等. 稻米的品质和影响因素[J]. 分子植物育种, 2003, 1(2): 231-241.
- [9] 孙平. 蛋白质含量多会降低稻米食味吗?——试析日本产销界关于稻米食味和是否追肥问题的争议 [J]. 中国稻米, 1998, 4(5): 31-33.
- [10] Kawakatsu T, Yamamoto M P, Hirose S et al. Characterization of a new rice glutelin gene *GluD-1* expressed in the starchy endosperm [J]. *J Exp Bot*, 2008, 59(15): 4 233-4 245.
- [11] Qu L Q, Xing Y P, Liu W X et al. Expression pattern and activity of six glutelin gene promoters in transgenic rice[J]. *J Exp Bot*, 2008, 59 (9): 2 417-2 424.
- [12] Takaiwa F, K Oono. Genomic DNA sequences of two new genes for new storage protein glutelin in rice [J]. *Jpn J Genet*, 1991, 66(2): 161-171.
- [13] Duan M, Sun S S. Profiling the expression of genes controlling rice grain quality[J]. *Plant Mol Biol*, 2005, 59(1): 165-178.
- [14] Lee H J, Jo Y M, Lee J Y, et al. Lack of globulin synthesis during seed development alters accumulation of seed storage proteins in rice [J]. *Int J Mol Sci*, 2015, 16(7): 1 4717-1 436.
- [15] Yuan Q, Ouyang S, Liu J et al., The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists[J]. *Nucleic Acids Res*, 2003, 31(1): 229-233.
- [16] Tamura K, Stecher G, Peterson D, et al. MEGA6: molecular evolutionary genetics analysis version 6.0 [J]. *Mol Biol Evol*, 2013, 30 (12): 2 725-2 729.
- [17] Mawal Y R, Mawal M R, Ranjekar P K. Biochemical and immunological characterization of rice albumin [J]. *Biosci Rep*, 1987, 7(1): 1-9.
- [18] 牛洪斌, 覃怀德, 王益华, 等. 水稻谷蛋白的一个新基因克隆及表达分析[J]. 作物学报, 2007, 33(3): 349-355.
- [19] Saito Y, Shigemitsu T, Yamasaki R, et al. Formation mechanism of the internal structure of type I protein bodies in rice endosperm: relationship between the localization of prolamin species and the expression of individual genes[J]. *Plant J*, 2012, 70(6): 1 043-1 055.
- [20] He Y, Ning T, Xie T, et al. Large-scale production of functional human serum albumin from transgenic rice seeds[J]. *Proc Natl Acad Sci U S A*, 2011, 108(47): 19 078-19 083.
- [21] Ning T, Xie T, Qiu Q, et al. Oral administration of recombinant human granulocyte-macrophage colony stimulating factor expressed in rice endosperm can increase leukocytes in mice [J]. *Biotechnol Lett*, 2008, 30(9): 1 679-1 686.
- [22] 刘奕, 程方民. 稻米中蛋白质和脂类与稻米品质的关系综述[J]. 中国粮油学报, 2006, 21(4): 6-10.
- [23] Kawakatsu T, Hirose S, Yasuda H, et al. Reducing rice seed storage protein accumulation leads to changes in nutrient quality and storage organelle formation[J]. *Plant Physiol*, 2010, 154(4): 1 842-1 854.

## Bioinformatics Analysis of Rice Storage Proteins Families

SUN Hongzheng, GUO Kai, SONG Ningyuan, ZHAO Quanzhi

(College of Agronomy, Henan Agricultural University/Collaborative Innovation Center of Henan Grain Crop/Key Laboratory of Rice Biology in Henan Province; 1st author: sunhongzheng@foxmail.com; \*Corresponding author: qzzhaoh@126.com)

**Abstract:** Storage protein is the second most important component of rice grain, which is one of the important evaluation factor affecting the tasting and nutrition quality. By investigation of the genome location, gene structure, expression mode and phylogenetic relationship of rice glutelin, prolamin, albumin and globulin, the results showed that glutelin and prolamin genes had multiple copies in the rice genome, which had 15 and 28 copies, respectively. The glutelin family could be classified into three subfamilies, and the prolamin family could be divided into two subfamilies. According to the relationship of gene copies, unnamed copies were assigned proper nomenclature. The albumin and globulin families had fewer copies. The albumin family had five copies. Of the five copies, the *SSA1* and *SSA5* copies had relatively higher expression level, and the other three copies were expressed at low level. The expression of three copies of the globulin family were extremely low. This study provide a theoretical guide for rice seed storage protein research.

**Key words:** rice; storage protein; glutelin; prolamin; albumin; globulin