

DOI:10.11913/PSJ.2095-0837.2017.60864

邓娇, 陈庆富, 张启迪, 汪燕, 梁成刚, 黄娟. 苦荞全基因组 11S 种子储藏蛋白基因的鉴定及表达分析[J]. 植物科学学报, 2017, 35(6): 856-864

Deng J, Chen QF, Zhang QD, Wang Y, Liang CG, Huang J. Genome-wide identification and expression analysis of 11S seed storage protein genes in tartary buckwheat (*Fagopyrum tataricum*) [J]. *Plant Science Journal*, 2017, 35(6): 856-864

苦荞全基因组 11S 种子储藏蛋白基因的鉴定及表达分析

邓娇, 陈庆富, 张启迪, 汪燕, 梁成刚, 黄娟*

(贵州师范大学荞麦产业技术研究中心, 贵阳 550001)

摘要:以苦荞(*Fagopyrum tataricum* (L.) Gaertn)全基因组数据为平台,采用生物信息学方法,挖掘出 9 个 11S 种子储藏蛋白基因,并对其定位、蛋白结构、系统发育及表达模式进行了分析。结果表明,苦荞 9 个 11S 种子储藏蛋白基因编码的蛋白长度为 189 ~ 914 aa,等电点位于 5.18 ~ 9.82 之间,分子量为 21.27 ~ 103.33 kD;定位分析结果显示,这些成员位于苦荞基因组的 6 条连锁群上(Megascaffold2/5 以及 scaffold77/344/395/861);序列比对分析发现,除了 1 个 11S 种子储藏蛋白 sample1_00009513-RA 具有 1 个 cupin 保守结构域外,其余 8 个都含有 2 个 cupin 结构域,并且在 cupin 保守结构域中,苦荞和拟南芥(*Arabidopsis thaliana* (L.) Heynh)共有 14 个保守的氨基酸残基;蛋白结构预测表明,苦荞 11S 种子储藏蛋白的结构具有 2 种类型;苦荞与其它 6 个物种(拟南芥、花生(*Arachis hypogaea* Linn.)、大豆(*Glycine max* (Linn.) Merr.)、杏仁(*Armeniaca vulgaris* Lam.)、胡桃(*Juglans regia* L.)和芝麻(*Sesamum indicum* Linn.))11S 种子储藏蛋白以及苦荞过敏蛋白(TBb 和 TBt)系统发育分析结果表明,这些蛋白可以分为 3 类,共具有 4 对旁系同源蛋白和 3 对直系同源蛋白;与已报道的苦荞过敏性储藏蛋白以及其它 5 个物种(花生、大豆、杏仁、胡桃和芝麻)的 11S 过敏蛋白比较发现,5 个 11S 种子储藏蛋白(sample1_00013128-RA、sample1_00013130-RA、sample1_00021677-RA、sample1_00021668-RA 和 sample1_00021674-RA)与苦荞 2 个过敏蛋白的同源性较高,同时它们与胡桃 11S 过敏蛋白的同源性最高,但尚需进一步实验来确定这 5 个成员是否为食物过敏原;RNA-Seq 转录组数据显示,4 个基因(sample1_00018411-RA、sample1_00026786-RA、sample1_00021674-RA、sample1_00022718-RA)在 2 种荞麦属植物的灌浆期种子中表达水平较高,且在‘大苦 1 号’中的表达水平要高于‘大甜 1 号’。

关键词:苦荞; 11S 种子储藏蛋白; Cupin 保守结构域; 系统发育关系; 表达分析

中图分类号: Q943.2

文献标识码: A

文章编号: 2095-0837(2017)06-0856-09

Genome-wide identification and expression analysis of 11S seed storage protein genes in tartary buckwheat (*Fagopyrum tataricum*)

Deng Jiao, Chen Qing-Fu, Zhang Qi-Di, Wang Yan, Liang Cheng-Gang, Huang Juan*

(Research Center of Buckwheat Industry Technology, Guizhou Normal University, Guiyang 550001, China)

Abstract: Based on whole genome data of tartary buckwheat (*Fagopyrum tataricum* (L.) Gaertn), we used bioinformatics to exploit nine 11S seed storage protein genes, and

收稿日期: 2017-08-07, 退修日期: 2017-09-01。

基金项目: 贵州省科技计划项目(黔科合 LH 字[2016]7205 号/[2015]7766 号); 贵州师范大学资助博士科研项目(11904/0516026, 11904/0516031); 贵州省科学技术基金计划(黔科合基础[2016]1107 号); 贵州省荞麦工程技术研究中心(黔科合农 G 字[2015]4003 号); 国家自然科学基金项目(31760430)。

This work was supported by grants from the Foundation of Guizhou Natural Science (Qiankehe LH Zi [2016]1107/Qiankehe LH Zi [2015]7766), Scientific Research Foundation for Doctor of Guizhou Normal University (11904/0516026, 11904/0516031), Science and Technology Foundation of Guizhou Province (Qiankehe Jichu [2016]1107), Research Center of Buckwheat Engineering and Technology in Guizhou (QianKehe Nong G Zi [2015]4003), and National Natural Science Foundation of China (31760430).

作者简介: 邓娇(1986-), 女, 博士, 助理研究员, 研究方向为植物生理及分子生物学(E-mail: ddj613@163.com)。

* 通讯作者(Author for correspondence. E-mail: huang200669@163.com)。

analyzed their identification, location, protein structure, phylogenetic relationship, and expression. Results showed that the encoded protein lengths of these nine genes ranged from 189 to 914 aa, the isoelectric points ranged from 5.18 to 9.82, and the molecular weights ranged from 21.27 to 103.33 kD. Sequence alignment showed that one 11S seed storage protein (sample1_00009513-RA) contained only one cupin conservative domain, with the other eight members all containing two cupin conservative domains. In addition, there were 14 conservative amino acid residues in the cupin domain between tartary buckwheat and *Arabidopsis*. Location analysis demonstrated that these genes were mapped on six linkage groups of the tartary buckwheat genome (Megascaffold2/5 and scaffold77/344/395/861). Prediction analysis suggested that the tartary buckwheat 11S seed storage proteins exhibited two types of protein structure. The phylogenetic relationship of the 11S seed storage proteins from tartary buckwheat and six other species (*Arabidopsis*, peanut, soybean, almond, walnut, and sesame) indicated that these proteins could be classified into three groups, containing four pairs of paralogous and three pairs of orthologous. Compared with the reported allergic storage proteins in tartary buckwheat and 11S allergens from five other species (peanut, soybean, almond, walnut, and sesame), five 11S seed storage proteins demonstrated high similarity, with the highest similarity found with walnut; however, further experiments are needed to clarify whether these five members are food allergens or not. RNA-Seq analysis revealed that only four members (sample1_00018411-RA, sample1_00026786-RA, sample1_00021674-R, sample1_00022718-RA) had high expression levels in milk stage seeds of two kinds of buckwheat, and were more highly expressed in ‘Daku 1’ than in ‘Datian 1’.

Key words: *Fagopyrum tataricum*; 11S seed storage protein; Cupin conservative domain; Phylogenetic relationship; Expression analysis

荞麦属于蓼科荞麦属 (*Fagopyrum*), 是一种双子叶的小宗杂粮作物, 常见的栽培种有甜荞 (*F. esculentum* Moench) 和苦荞 (*F. tatarium* (Linnaeus) Gaertner)。我国荞麦主要种植于甘肃、宁夏、西藏和西南地区^[1]。苦荞又叫鞑靼荞麦 (Tartary buckwheat), 因其种子具有苦味而得名^[2]。研究发现, 苦荞种子富含黄酮类化合物, 具有降血脂、降血糖、抗氧化、抗炎、镇痛等多种功效^[3-6]。荞麦种子中的蛋白质含量较高, 平均约为 10%。根据溶解性不同, 荞麦种子蛋白可分为溶于水的清蛋白、溶于盐的球蛋白、溶于醇的醇溶蛋白和溶于稀酸或稀碱的谷蛋白^[2]。荞麦具有高含量的清蛋白和球蛋白以及低含量的醇溶蛋白和谷蛋白, 其蛋白质组分更接近豆类植物的蛋白质组分^[7]。此外, 荞麦种子中主要储存蛋白质的氨基酸组分均衡, 含有 8 种人体必须的氨基酸, 且配比合适^[8]。研究表明, 荞麦蛋白具有降低胆固醇、抑制胆结石、抗癌、抗白血病和抗癌等功效^[2]。

因此, 苦荞是一种集食用、药用和保健为一体的作物, 值得广泛开发利用。

11S 球蛋白和 7S 球蛋白同属于 cupin 超家族, 是种子中 2 种主要的储藏球蛋白, 也是人们饮食的主要蛋白成分。它们都具有一个由 2 个 jelly-roll β -barrels 组成的高度保守域和 2 个延伸螺旋的核心结构域, 被称为 cupin。有的蛋白含有 1 个这样的结构域, 有的含有 2 个^[9-11]。这 2 种球蛋白随种子的发育而合成直至种子成熟, 最终贮藏在蛋白储藏液泡中供种子萌发使用^[10]。双子叶植物的球蛋白都由 6 个亚基构成, 其中每一个亚基又由 2 条多肽组成, 一条为酸性多肽, 另一条为碱性多肽, 2 条多肽由二硫键连接起来^[12]。11S 球蛋白是由 6 个分子量为 50 ~ 60 kD 的蛋白亚基组成的 300 ~ 400 kD 的六聚异源寡聚蛋白, 其中酸性多肽链和碱性多肽链分别约为 40 kD 和 20 kD^[9,10]。据报道, 大多数 cupins 为食物过敏原, 其中 11S 球蛋白在花生

(*Arachis hypogaea* Linn.)、大豆(*Glycine max* (Linn.) Merr.)、杏仁(*Armeniaca vulgaris* Lam.)、胡桃(*Juglans regia* L.)和芝麻(*Sesamum indicum* Linn.)等物种中被鉴定为食物过敏原^[9,13-17]。目前,在荞麦中也鉴定到过敏性储藏球蛋白 TBb 和 TBt^[18-20]。

随着测序技术的快速发展,苦荞品种(‘晋荞麦 2 号’)基因组已经完成测序,这为在全基因组水平上分析 11S 种子储藏蛋白提供了条件。本研究基于苦荞全基因组数据对 11S 编码基因进行鉴定及表达分析,以期对 11S 种子储藏蛋白的结构和功能研究奠定基础,并为荞麦种子营养学和过敏性研究提供理论依据。

1 材料与方法

1.1 实验材料

荞麦 2 个栽培品种:甜荞(‘大甜 1 号’)和苦荞(‘大苦 1 号’),来源于贵州师范大学荞麦产业技术研究中心。2013 年 8 月种植于贵州师范大学百宜基地,采用常规大田管理。于同年 10 月分别收集两者的灌浆期种子,用液氮迅速冷冻,保存于-80℃超低温冰箱。

1.2 苦荞 11S 种子储藏蛋白基因的鉴定

从拟南芥(*Arabidopsis thaliana* (L.) Heynh.)数据库网站(<http://www.arabidopsis.org/>)中找到拟南芥 11S 种子储藏蛋白相关基因的序列信息,然后利用‘晋苦荞 2 号’全基因组数据(登录号: SRR5861725)进行 Blast 比对,筛选出候选基因,并确定它们所在的连锁群。利用 PRINTS 在线工具(<http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/index.php>)鉴定候选基因。通过 DNAMAN 软件分析苦荞 11S 种子储藏蛋白的等电点(Isoelectric point, pI)和分子量(Molecular weight, MW)。

1.3 苦荞 11S 种子储藏蛋白的结构预测、结构域和系统发育分析

根据候选基因编码的氨基酸序列,利用 SWISS-MODEL 在线软件(<http://swissmodel.expasy.org/interactive>)对苦荞 11S 种子储藏蛋白进行结构预测。利用 ClustalX 软件对苦荞和拟南芥 11S 种子储藏蛋白进行多重序列比对分析^[21],并找出 cupin 结构域。利用 MEGA 5.0 软件,采用邻

接法(Neighbor-Joining, NJ)构建苦荞和其它物种(花生、大豆、杏仁、胡桃和芝麻)的 11S 种子储藏蛋白系统发育树^[22], Bootstrap 值设定为 1000。

1.4 RNA 提取、cDNA 文库构建及转录组测序分析

采用 RNA 提取试剂盒提取‘大甜 1 号’和‘大苦 1 号’灌浆期种子的总 RNA。检测合格后进行测序分析。cDNA 文库采用 cDNA 合成试剂盒(Illumina, 美国)构建,检测合格的文库利用 HiSeq™ 2000 (Illumina, 美国)进行测序,RPKM 值(每百万 reads 中来自某一基因每千碱基长度的 reads 数目)> 1,且 P -value < 0.05 的基因被认为是可表达的基因。

2 结果与分析

2.1 苦荞 11S 种子储藏蛋白的鉴定

目前,苦荞品种‘晋荞麦 2 号’基因组测序已完成,基因组大小约为 450 M,拼接得到 13 条大的连锁群(Megasc scaffold)和一些小的连锁群(Scaffold),推测能编码蛋白的基因共 28 215 个。基于此数据库为平台,以拟南芥 3 个 11S 种子储藏蛋白基因(AT1G07750.1, AT1G03890.1 和 AT2G28680.1)为参照,共筛选出 9 个 11S 种子储藏蛋白基因(表 1)。通过 PRINS 数据库鉴定,这 9 个苦荞基因和 3 个拟南芥基因都属于 PR00439 家族,即 11S 种子储藏蛋白家族。这 9 个基因编码蛋白的氨基酸数目为 189 ~ 914 aa,其中序列最长的为 sample1_00018411-RA,最短的为 sample1_0009513-RA;蛋白的分子量范围为 21.27 ~ 103.33 kD,大多数在 51 ~ 58 kD;除 sample1_00018411-RA 和 sample1_0009513-RA 分别含有 8 个和 2 个外显子外,其余的基因均含有 4 个外显子;蛋白等电点的变化范围为 5.18 ~ 9.82,最高的是 sample1_0009513-RA,呈碱性,其余的都呈酸性。

上述 9 个 11S 种子储藏蛋白基因定位于 6 条连锁群上(Megasc scaffold 2/5 以及 scaffold 77/344/395/861)(表 1),且分布不均匀。其中大连锁群 Megasc scaffold 2 上含有 3 个基因(sample1_00021677-RA、sample1_00021668-RA 和 sample1_00021674-RA);小连锁群 scaffold 344 上含有 2 个基因(sample1_00013128-RA 和 sample1_00013130-RA);其余 4 条小的连锁群

上分别分布 1 个基因。*sample1_00021677-RA*、*sample1_00021668-RA* 和 *sample1_00021674-RA* 相似性很高，其中 *sample1_00021668-RA* 和 *sample1_00021674-RA* 具有 100% 的相似性，这两者和 *sample1_00021677-RA* 的相似度都高达 99.61%，且 3 个基因都在 Megasc scaffold 2 上，均编码 514 个氨基酸，因此推测它们可能是串联重复基因。另外，*sample1_00013128-RA* 和 *sample1_00013130-RA* 同时位于 scaffold 344，相似度也高达 96.2%，因此也可能是串联重复基因。

2.2 苦荞和拟南芥 11S 种子储藏蛋白的结构预测及 cupin 结构域分析

利用 SWISS MODEL 软件对 9 个苦荞 11S 种子储藏蛋白的结构进行预测，发现它们共有 2 种类型的结构(图 1)，具有第一种类型蛋白结构的包括 *sample1_00018411-RA*、*sample1_00013128-RA*、*sample1_00013130-RA*、*sample1_00026786-RA* 和 *sample1_00022718-RA*(图 1: A)；具有第二

种类型蛋白结构的有 *sample1_00009513-RA*、*sample1_00021677-RA*、*sample1_00021668-RA* 和 *sample1_00021674-RA*(图 1: B)。从图中预测的蛋白结构可以看出，苦荞 11S 球蛋白含有 6 个亚基，每个亚基由多个 β -链(β -strands)构成，每 2 个亚基之间由 α -螺旋(α -helices)连接。

对苦荞和拟南芥 11S 种子储藏蛋白的氨基酸序列进行多重序列比对分析发现，9 个苦荞 11S 球蛋白之间的相似度范围为 18% ~ 100%，其中 *sample1_00018411-RA* 和 *sample1_00026786-RA* 与其他成员的相似度最低。Cupin 保守结构域由 70 ~ 200 个氨基酸残基构成，在 9 个蛋白中，除 *sample1_00009513-RA* 只有 1 个 cupin 保守结构域，其他 8 个成员和拟南芥的 3 个 11S 种子储藏蛋白都有 2 个 cupin 保守结构域(图 2)。其中在第一个 cupin 结构域中，除了 *sample1_00009513-RA*(只有第二个 cupin 结构域)，其它 8 个与拟南芥中的 3 个 11S 种子储藏蛋白的氨基酸

表 1 苦荞 11S 种子储藏蛋白基因信息
Table 1 Information on the seed 11S storage protein genes from buckwheat

基因 Genes	蛋白长度 (aa) Protein length	外显子数 Exon number	等电点 pI	分子量 (kD) MW	位置 Site
<i>sample1_00018411-RA</i>	914	8	6.03	103.33	Megasc scaffold 5
<i>sample1_00009513-RA</i>	189	2	9.82	21.27	Scaffold 77
<i>sample1_00013128-RA</i>	515	4	5.83	58.32	Scaffold 344
<i>sample1_00013130-RA</i>	501	4	5.98	56.84	Scaffold 344
<i>sample1_00026786-RA</i>	355	4	6.92	37.88	Scaffold 395
<i>sample1_00021677-RA</i>	514	4	5.70	58.23	Megasc scaffold 2
<i>sample1_00021668-RA</i>	514	4	5.73	58.26	Megasc scaffold 2
<i>sample1_00021674-RA</i>	514	4	5.73	58.26	Megasc scaffold 2
<i>sample1_00022718-RA</i>	462	4	5.18	51.50	Scaffold 861

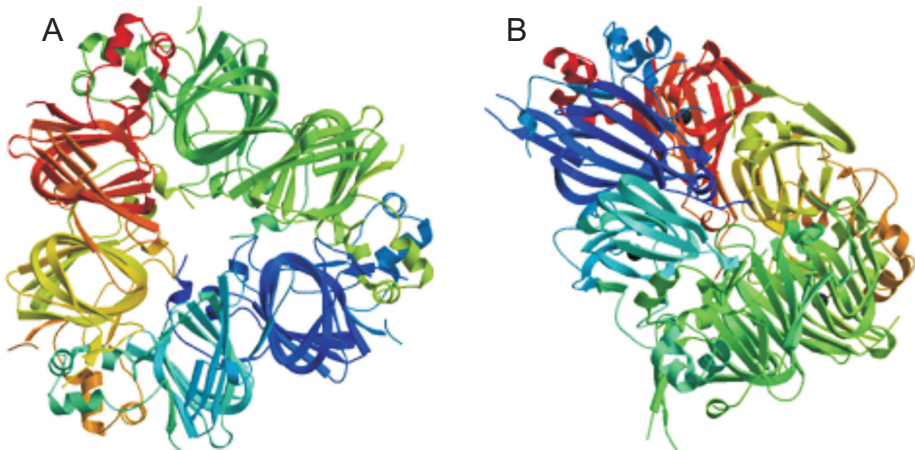
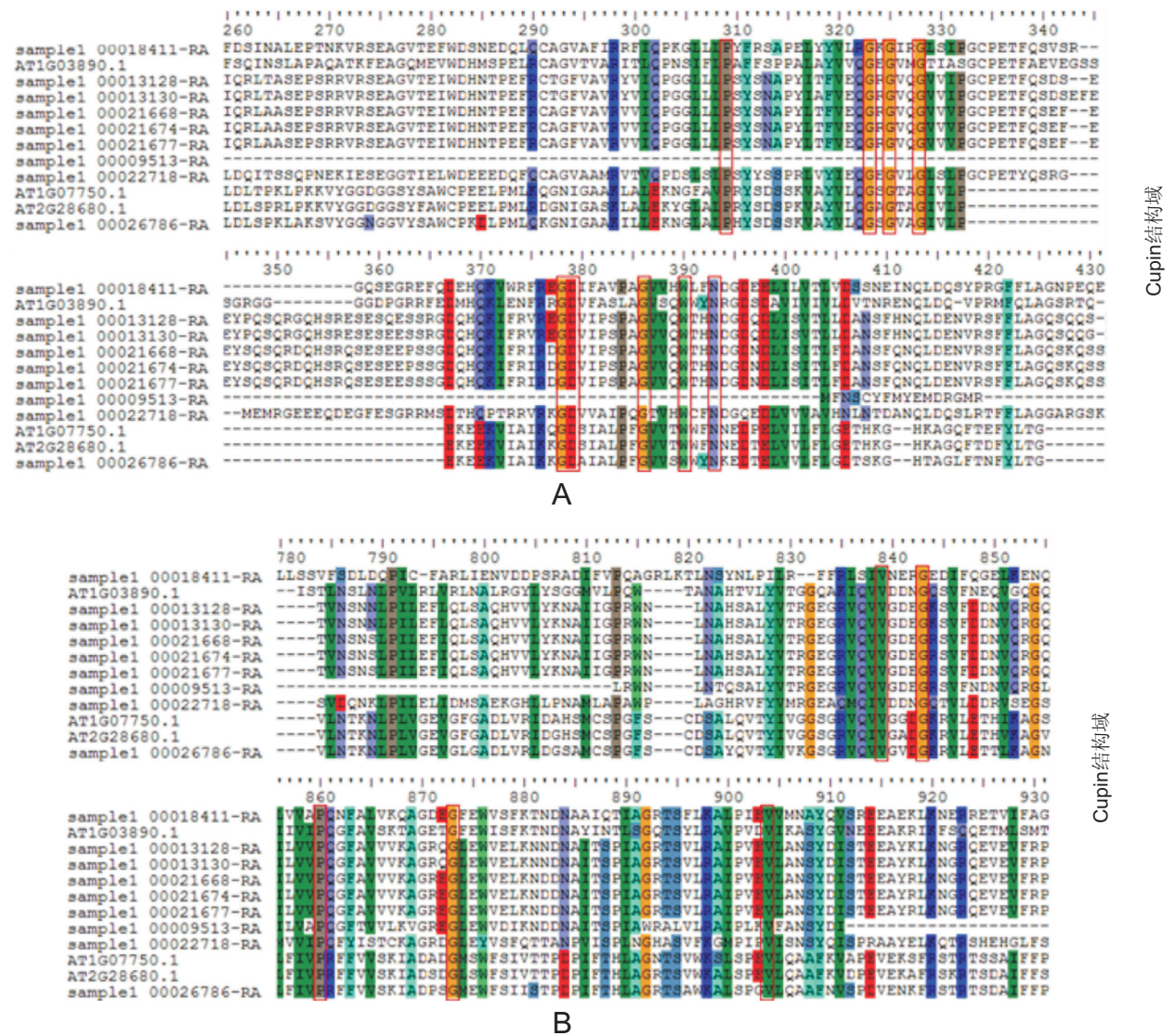


图 1 苦荞 11S 种子储藏蛋白的 2 种结构类型
Fig. 1 Two structure types of 11S seed storage proteins in tartary buckwheat



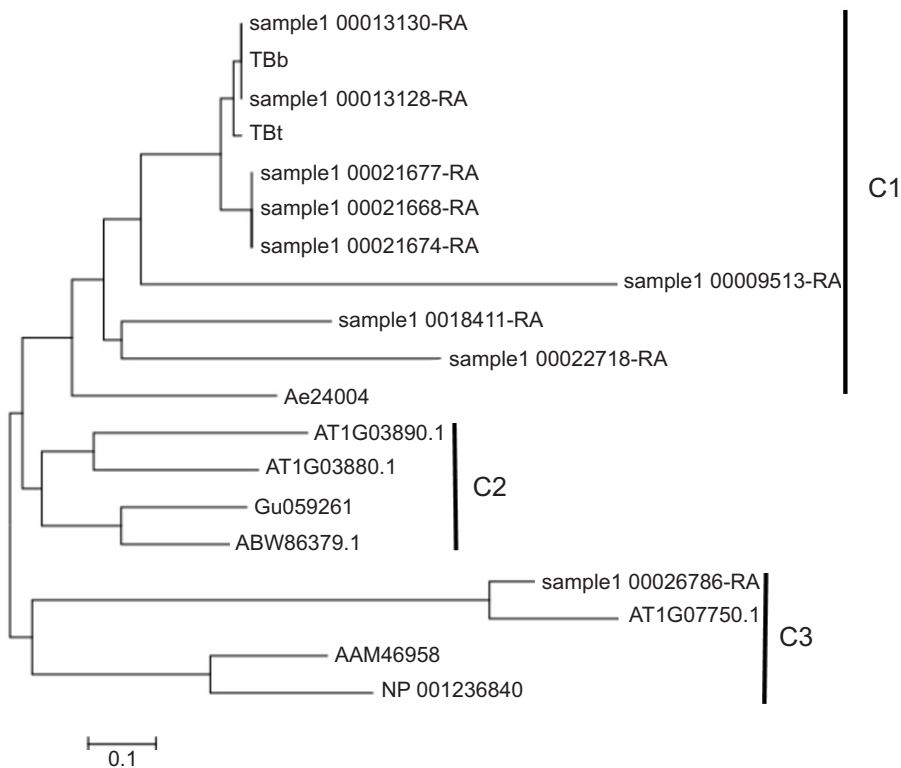
红框表示这些蛋白该位点氨基酸残基一致。
Red boxes mean amino acid residues in sites of these proteins are identical.

图 2 苦荞和拟南芥的 11S 种子储藏蛋白序列比较及 cupin 结构域分析
Fig. 2 Sequence comparison of 11S seed storage proteins from tartary buckwheat and *Arabidopsis* and analysis of the cupin domain

序列共有 9 个完全保守的位点(红框标记), 分别为: 脯氨酸(Proline, P)、天冬氨酸(Aspartic acid, D)、色氨酸(Tryptophan, W)和天冬酰胺(Asparagine, N)各 1 个; 甘氨酸(Glycine, G)5 个(图 2: A); 而在第二个 cupin 结构域中, 苦荞和拟南芥的氨基酸序列共有 5 个完全保守的位点(红框标记), 其中包括 2 个缬氨酸(Valine, V), 2 个甘氨酸(Glycine, G)和 1 个脯氨酸(Proline, P)(图 2: B)。研究表明, 甘氨酸涉及到球蛋白的正确折叠^[10], 这可能是苦荞与拟南芥的 11S 球蛋白序列中甘氨酸保守位点较多的原因。

2.3 苦荞和其它物种 11S 种子储藏蛋白系统发育分析

本研究构建的苦荞和其它物种 11S 种子储藏蛋白的系统发育树结果表明(图 3), 这些 11S 种子储藏蛋白可以分为 3 个进化枝(C1、C2 和 C3)。其中, 第一个进化枝 C1 包括 10 个苦荞种子蛋白(即 8 个 11S 种子储藏蛋白、2 个苦荞过敏蛋白 TBb 和 Tbt)和 1 个芝麻的 11S 过敏蛋白。C1 包含了 3 对旁系同源基因, 第一对包括 3 个蛋白: sample1_00013130-RA、TBb 和 sample1_00013128-RA; 第二对也包括 3 个蛋白:



Sample 1-00013130/00013128/00021677/00021668/00021674/00009513/00018411/00022718/00026786-RA 为苦荞 11S 种子储藏蛋白；TBb 和 TBt 为苦荞 2 个过敏蛋白；AAM46958.1、NP-001236840、GU059261、ABW86979.1 和 AF240004 分别为花生、大豆、杏仁、胡桃和芝麻的 11S 过敏蛋白。
Sample 1-00013130/00013128/00021677/00021668/00021674/00009513/00018411/00022718/00026786-RA are 11S seed storage proteins from tartary buckwheat; TBb and TBt are two allergic proteins of tartary buckwheat; AAM46958.1, NP-001236840, GU059261, ABW86979.1, and AF240004 are 11S allergens from peanut, soybean, almond, walnut, and sesame, respectively.

图 3 苦荞和其它物种的 11S 种子储藏蛋白以及 TBb 和 TBt 的系统发育关系
Fig. 3 Phylogenetic relationships of 11S seed storage proteins from tartary buckwheat and other species as well as TBb and TBt

sample1_00021677-RA、sample1_00021668-RA 和 sample1_00021674-RA；第三对为 sample 10018411-RA 和 sample1_00022718-RA。而 C2 进化枝包含 4 个蛋白，包括一对来自拟南芥的旁系同源蛋白(AT1G03890.1 和 AT1G07750.1)和一对直系同源蛋白，分别来自杏仁(GU059261)和胡桃(ABW86379.1)的 11S 过敏球蛋白。进化枝 C3 也包含 4 个蛋白，分为 2 对直系同源蛋白，第一对来自苦荞(sample1_00026786-RA)和拟南芥(AT1G03880.1)的 11S 种子储藏蛋白；另一对来自花生(AAM46958.1)和大豆(NP_001236840)的 11S 过敏蛋白。

2.4 苦荞 11S 种子储藏蛋白与苦荞过敏性储藏球蛋白同源性分析

目前，已报道的苦荞过敏球蛋白有 2 个，一个为 TBb，由 320 个氨基酸组成^[18]；另一个为 TBt，由 515 个氨基酸组成^[19,20]。将苦荞的 9 个 11S 种

子储藏蛋白与这 2 个已知的过敏性储藏球蛋白及其它 5 个物种(花生、大豆、杏仁、胡桃和芝麻)的 11S 过敏蛋白进行同源性分析(表 2)，结果发现，sample1_00018411-RA、sample1_00009513-RA、sample1_00026786-RA 和 sample1_00022718-RA 这 4 个 11S 种子储藏蛋白与 2 个过敏性储藏球蛋白的同源性较低，最低的一对为 sample1_00009513-RA 和 TBt(12%)。而 sample1_00013128-RA 与 TBb 和 TBt 的同源性最高，均达到了 98%；sample1_00013130-RA 与 TBb 和 TBt 的同源性也分别为 95% 和 91%；sample1_00021677-RA、sample1_00021668-RA 和 sample1_00021674-RA 这 3 个 11S 种子储藏蛋白与 TBb 的同源性都为 86%，与 TBt 的同源性都为 83%。与其它 5 个物种的 11S 过敏蛋白的同源性分析显示，sample1_00018411-RA、sample1_00009513-RA、sample1_00013128-RA、sample1_00013130-RA、sample1_

表 2 苦荞 11S 种子储藏蛋白、2 个苦荞过敏性储藏球蛋白及其它 5 个物种 11S 种子储藏蛋白的同源性分析
Table 2 Homology comparisons of 11S seed storage proteins of tartary buckwheat with two known allergic storage globulins of tartary buckwheat as well as 11S seed storage proteins from other species

11S 种子储藏蛋白 11 S seed storage proteins	TBb (%)	TBt (%)	花生(%) Peanut	大豆(%) Soybean	杏仁(%) Almond	胡桃(%) Walnut	芝麻(%) Sesame
sample1_00018411-RA	31	34	27	27	30	33	35
sample1_00009513-RA	44	12	23	21	22	28	25
sample1_00013128-RA	97	98	30	33	37	40	37
sample1_00013130-RA	95	91	31	33	37	40	37
sample1_00026786-RA	21	15	17	20	24	21	21
sample1_00021677-RA	86	83	31	32	36	39	37
sample1_00021668-RA	86	83	31	32	36	39	37
sample1_00021674-RA	86	83	31	32	36	39	37
sample1_00022718-RA	34	26	27	29	32	31	32

00021677-RA、sample1_00021668-RA、sample1_00021674-RA 都与胡桃的 11S 过敏蛋白的同源性最高；除了 sample1_00009513-RA，其它 8 个苦荞 11S 种子储藏蛋白与花生的 11S 过敏蛋白的同源性最低。

此外，经过序列比对分析发现，sample1_00013128-RA 与 TBb 和 TBt 的前 22 个氨基酸序列 (MSTKLILSFSLCLMVLSCSAQA) 一致，被称为信号肽序列，在蛋白成熟之前被切除^[19]。而与 TBb 和 TBt 相似度较高的 sample1_00013130-RA、sample1_00021677-RA、sample1_00021668-RA 和 sample1_00021674-RA 的前 21 个氨基酸序列与信号肽序列一致，第 22 个为亮氨酸(L)。以上 5 个与 TBb 和 TBt 相似度较高的 11S 种子储藏蛋白是否为过敏蛋白，还有待进一步验证。

2.5 11S 种子储藏蛋白基因在甜荞和苦荞种子中的表达分析

甜荞和苦荞是荞麦的 2 种常见的栽培种，本研究以甜荞品种‘大甜 1 号’和苦荞品种‘大苦 1 号’的灌浆期种子为实验材料进行转录组测序，基于 RNA-Seq 数据，分析 9 个 11S 种子储藏蛋白基因在两者中的表达情况(表 3)。结果显示，sample1_00009513-RA、sample1_00013128-RA、sample1_00013130-RA、sample1_00021677-RA 和 sample1_00021668-RA 在‘大甜 1 号’和‘大苦 1 号’的灌浆期种子中未表达；其余 4 个基因在两者中都有较高的表达水平(RPKM 为 21.96 ~ 28 481.5)，其中只有 sample1_00026786-RA 在‘大甜 1 号’中的表达量高于‘大苦 1 号’，约为 4 倍；sample1_00022718-RA 在两者中的表达量都很高，且在‘大

苦 1 号’中表达水平稍高于‘大甜 1 号’；sample1_00018411-RA 和 sample1_00021674-RA 在‘大苦 1 号’中的表达量远高于‘大甜 1 号’，分别达到 6 倍和 5 倍(表 3)，这些结果表明 11S 种子储藏蛋白基因在不同的荞麦属植物中具有差异表达的特点。

表 3 11S 种子储藏蛋白基因在‘大甜 1 号’(A)和‘大苦 1 号’(B)种子中的表达分析
Table 3 Expression analysis of 11S seed storage proteins in the seeds of ‘Datian 1’ (A) and ‘Daku 1’ (B)

基因 Gene	表达量(A) RPKM (A)	表达量(B) RPKM (B)
sample1_00018411-RA	429.10	2683.08
sample1_00009513-RA	0	0
sample1_00013128-RA	0	0
sample1_00026786-RA	87.28	21.96
sample1_00013130-RA	0	0
sample1_00021677-RA	0	0
sample1_00021668-RA	0	0
sample1_00021674-RA	6052.92	28481.50
sample1_00022718-RA	3456.60	3720.80

注：RPKM 为每百万 reads 中来自某一基因每千碱基长度的 reads 数目。
Note: RPKM represents reads per kilo bases per million reads.

3 讨论

11S 种子储藏蛋白属于 cupin 家族，是日常食用蛋白的主要成分之一。而有些食物(如花生、大豆、杏仁和芝麻等)中的 11S 种子储藏蛋白被鉴定为能引起过敏反应的过敏原^[9]。

本研究在苦荞全基因组数据库中共鉴定了 9 个 11S 种子储藏蛋白基因，它们编码蛋白的氨基酸数目在 189 ~ 914 aa 间，等电点(pI)位于 5.18 ~ 9.82，分子量范围为 21.27 ~ 103.33 kD。这 9 个

基因大多数含有 4 个外显子, 且不均匀地分布于苦荞基因组的 Megascaffold 2/5 以及 scaffold 77/344/395/861 上。而同时分布于 scaffold 344 上的 *sample1_00013128-RA* 和 *sample1_00013130-RA* 的相似度高达 96.2%, 可能是基因的串联重复; 同样位于 Megascaffold 2 的 3 个基因(*sample1_00021677-RA*、*sample1_00021668-RA* 和 *sample1_00021674-RA*) 的同源性也很高, 也可能是基因的串联重复。

本研究通过对苦荞 11S 种子储藏蛋白的结构进行预测, 发现苦荞的 11S 球蛋白与其他植物球蛋白一样, 都由 6 个亚基组成。通过苦荞和拟南芥的多重序列比对发现, 除 *sample1_00009513-RA* 外, 都具有 2 个 cupin 结构域, 在 2 个结构域中, 苦荞和拟南芥共有 14 个极度保守的氨基酸残基, 其中甘氨酸(G)最多, 有 7 个。研究表明甘氨酸对球蛋白的正确折叠起关键作用^[10], 这可能是甘氨酸在物种间的保守性高的原因。

通过系统发育分析可以看出, 苦荞和拟南芥具有 1 对直系同源基因和 3 对旁系同源基因, 而苦荞中 11S 种子储藏蛋白基因比拟南芥的多 6 个, 这说明两者的 11S 种子储藏蛋白基因在进化中采用了不同的方式扩增。通过与已报道的苦荞 2 个过敏球蛋白比对发现, *sample1_00021677-RA*、*sample1_00021668-RA*、*sample1_00021674-RA*、*sample1_00013128-RA* 和 *sample1_00013130-RA* 这 5 个 11S 种子储藏蛋白与苦荞的过敏蛋白的相似性很高, 范围为 83% ~ 98%, 但需要进一步验证它们是否为过敏原。基于 RNA-Seq 数据分析发现, 有 4 个基因在‘大甜 1 号’和‘大苦 1 号’灌浆期种子中的表达水平较高, 且 3 个在‘大苦 1 号’中的表达水平要高于‘大甜 1 号’, 表明 11S 种子储藏蛋白在不同的荞麦属植物中表达水平存在差异。而其他 5 个成员(*sample1_00009513-RA*、*sample1_00013128-RA*、*sample1_00013130-RA*、*sample1_00021677-RA* 和 *sample1_00021668-RA*) 在‘大甜 1 号’和‘大苦 1 号’的灌浆期种子中均未表达, 推测可能为假基因。

参考文献:

- [1] 刘拥海, 俞乐, 肖迪. 荞麦种子蛋白质组分分析[J]. 种子, 2006, 25(12): 31-33.
Liu YH, Yu L, Xiao D. Analysis on content of seed protein

groups in buckwheat grain[J]. *Seed*, 2006, 25(12): 31-33.

- [2] 陈庆富. 荞麦属植物科学[M]. 北京: 科学出版社; 2012.
- [3] 崔晓明, 耿非, 王建辉, 韩淑英. 荞麦黄酮对异丙肾上腺素致大鼠心肌缺血的预防作用[J]. 河北联合大学学报: 医学版, 2013, 15(4): 466-467.
Cui XM, Geng F, Wang JH, Han SY. Preventive effect of buckwheat flavones on isoprenaline-induced myocardial ischemia in mice[J]. *Journal of Hebei Association University: Medical Edition*, 2013, 15(4): 466-467.
- [4] 白静, 姜妍, 韩婷, 王建行, 余红, 等. 荞麦花叶黄酮对 2 型糖尿病 C57 小鼠糖脂代谢及胰岛素抵抗的影响[J]. 广东医学, 2012, 33(13): 1876-1879.
Bai J, Jiang Y, Han T, Wang JH, Yu H, *et al*. The effect of flavones of buckwheat flowers and leaves on the metabolic of glucose, lipis and insulin resistance in mice [J]. *Guangdong Medical*, 2012, 33(13): 1876-1879.
- [5] Watanabe M, Ohshita Y, Tsushida T. Antioxidant compounds from buckwheat (*Fagopyrum esculentum* Möench) hulls [J]. *J Agr Food Chem*, 1997, 45(4): 1039-1044.
- [6] 张博男, 储金秀, 韩淑英. 荞麦花叶总黄酮的舒张血管作用及其机制[J]. 中国药理学通报, 2010, 26(7): 952-956.
Zhang BN, Chu JX, Han SY. Relaxant effects of total flavonoids of buckwheat flowers and leaves on rat thoracic aorta and underlying mechanisms[J]. *Chinese pharmacology bulletin*, 2010, 26(7): 952-956.
- [7] Zheng GH, Sosulski FW, Tyler RT. Wet-milling, composition and functional properties of starch and protein isolated from buckwheat groats[J]. *Food Res Int*, 1997, 30(7): 493-502.
- [8] 杜双奎, 李志西, 于修焯. 荞麦蛋白研究进展[J]. 食品科学, 2004, 25(10): 409-414.
Du SK, Li ZX, Yu XZ. Research progress of buckwheat protein[J]. *Food Science*, 2004, 25(10): 409-414.
- [9] Mills ENC, Jenkins J, Marigheto N, Belton PS, Gunning AP, Morris VJ. Allergens of the cupin superfamily[J]. *Biochem Soc T*, 2002, 30(6): 925-929.
- [10] Tandang-Silvas MRG, Fukuda T, Fukuda C, Prak K, Cabanos C, *et al*. Conservation and divergence on plant seed 11S globulins based on crystal structures[J]. *BBA-Proteins Proteom*, 2010, 1804(7): 1432-1442.
- [11] Breiteneder H, Radauer C. A classification of plant food allergens[J]. *J Allergy Cli Immun*, 2004, 113(5): 821-830.
- [12] 白宝璋. 向日葵子实 11S-球蛋白的异质性与多型性[J]. 国外农学: 向日葵, 1991(4): 27-30.
Bai BZ. Heterogeneity and polymorphism of 11S-glubin of sunflower[J]. *Foreign Agriculture: Sunflower*, 1991(4): 27-30.
- [13] Zhao L, Zhao L, Zhang B, Robotham J, Roux K, Tang H.

- Identification of a common Ara h 3 epitope recognized by both the capture and the detection monoclonal antibodies in an ELISA detection kit[J]. *PLoS One*, 2017, 12(8): e0182935.
- [14] Wang T, Qin GX, Sun ZW, Zhao Y. Advances of research on glycinin and β -conglycinin: a review of two major soybean allergenic proteins[J]. *Crit Rev Sci*, 2014, 54(7): 850–862.
- [15] Willison LAN, Zhang Q, Su M, Teuber S, Sathe S, Roux K. Conformational epitope mapping of Pru du 6, a major allergen from almond nut[J]. *Mol Immunol*, 2013, 55(3): 253–263.
- [16] Sharma GM, Irsigler A, Dhanarajan P, Ayuso R, Bardina L, *et al.* Cloning and characterization of an 11S legumin, Car i 4, a major allergen in pecan[J]. *J Agr Food Chem*, 2011, 59(17): 9542–9552.
- [17] Wallowitz ML, Chen RJY, Tzen JTC, Teuber SS. Ses i 6, the sesame 11S globulin, can activate basophils and shows cross-reactivity with walnut *in vitro* [J]. *Clin Exp Allergy*, 2007, 37(6): 929–938.
- [18] 赵小珍, 张政, 景巍, 李玉英, 王转花. 苦荞麦主要过敏蛋白 N 端基因片段的克隆及序列分析[J]. 食品科学, 2006, 27(10): 41–44.
- Zhao XZ, Zhang Z, Jing W, Li YY, Wang ZH. Cloning and sequence analysis of N-terminal gene of major allergenic protein in tartary buckwheat[J]. *Food Science*, 2006, 27(10): 41–44.
- [19] 张昕, 崔晓东, 李玉英, 张德梅, 王转花. 苦荞过敏蛋白全长基因的克隆, 表达及免疫学活性研究[J]. 食品科学, 2009, 30(1): 203–206.
- Zhang Q, Cui XD, Li YY, Zhang DM, Wang ZH. Study on cloning and expression of a full-length gene of allergenic protein of tartary buckwheat and its immunological activity. *Food Science*, 2009, 30(1): 203–206.
- [20] Yang ZH, Li C, Li YY, Wang ZH. Effects of maillard reaction on allergenicity of buckwheat allergen Fag t 3 during thermal processing[J]. *J Sci Food Agr*, 2013, 93(6): 1510–1515.
- [21] Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, *et al.* Multiple sequence alignment with the clustal series of programs[J]. *Nucleic Acids Res*, 2003, 31(13): 3497–3500.
- [22] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. *Mol Biol Evol*, 2011, 28(10): 2731–2739.

(责任编辑: 周 媛)