

黄小芳, 毕楚韵, 黄伟群, 等. 甘薯 β -淀粉酶家族基因的全基因组鉴定和表达分析 [J]. 华南农业大学学报, 2021, 42(5): 50-59.
HUANG Xiaofang, BI Chuyun, HUANG Weiqun, et al. Genome-wide identification and expression analysis of the β -amylase gene family in *Ipomoea batatas*[J]. Journal of South China Agricultural University, 2021, 42(5): 50-59.

甘薯 β -淀粉酶家族基因的全基因组 鉴定和表达分析

黄小芳^{1,2}, 毕楚韵^{1,2}, 黄伟群³, 刘江洪¹, 胡韵卓¹, 黄碧芳⁴, 林世强^{1,4,5}, 陈选阳^{1,2,6}
(1 福建农林大学 作物生物技术福建省高校重点实验室, 福建 福州 350002; 2 福建农林大学 农学院, 福建 福州 350002;
3 福建省种子总站, 福建 福州 350003; 4 福建农林大学 生命科学学院, 福建 福州 350002; 5 福建省特种作物育种与利用
工程技术研究中心, 福建 福州 350002; 6 福建农林大学 教育部作物遗传育种与综合利用重点实验室, 福建 福州 350002)

摘要:【目的】挖掘甘薯 *Ipomoea batatas* 基因组中 β -淀粉酶 (Beta-amylase) 基因家族序列信息, 分析结构与功能信息。【方法】基于甘薯栽培种 ‘泰中 6 号’ 全基因组测序数据, 利用生物信息学分析方法对鉴定到的 12 个甘薯 β -淀粉酶家族成员进行结构域保守性分析、染色体定位、潜在重复基因筛查、保守基序分析和系统进化树构建, 利用转录组数据进行低温胁迫下相关基因的表达分析。【结果】12 个 β -淀粉酶基因分布于甘薯第 2、4、5、6、11、12、13 和 14 号染色体上, 含有 8 个具有潜在重复关系的基因对。多重比对和功能结构域搜索结果显示, 甘薯 β -淀粉酶家族的氨基酸序列中含有 3 个保守性较高的区域和 10 个保守基序。甘薯与其他物种 β -淀粉酶蛋白的系统进化分析结果显示, 62 个 β -淀粉酶家族成员被分为 S1~S7 等 7 个亚组, 甘薯 β -淀粉酶家族成员主要分布在 S2、S4、S5、S6 以及 S7 亚组中, 且大多与拟南芥、马铃薯以及番茄的 β -淀粉酶为同一分支。转录组测序数据分析结果显示, 在低温贮藏的过程中有 6 个甘薯 β -淀粉酶基因表达量出现变化, 其中 ‘徐薯 15-1’ 有 2 个基因上调表达、4 个基因下调表达, ‘徐薯 15-4’ 仅有 2 个基因下调表达。【结论】 β -淀粉酶是一类关键的淀粉水解酶, 在甘薯生长发育和薯块贮藏阶段淀粉降解为还原糖的过程中发挥着重要作用, 本研究鉴定得到的 12 个甘薯 β -淀粉酶基因序列信息为进一步探讨甘薯 β -淀粉酶基因家族的生物学功能提供数据参考。

关键词: 甘薯; β -淀粉酶; 基因家族; 系统进化; 差异表达分析
中图分类号: S531 **文献标志码:** A **文章编号:** 1001-411X(2021)05-0050-10

Genome-wide identification and expression analysis of the β -amylase gene family in *Ipomoea batatas*

HUANG Xiaofang^{1,2}, BI Chuyun^{1,2}, HUANG Weiqun³, LIU Jianghong¹, HU Yunzhuo¹,
HUANG Bifang⁴, LIN Shiqiang^{1,4,5}, CHEN Xuanyang^{1,2,6}
(1 Key Laboratory of Crop Biotechnology in Fujian Province University, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2 College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 3 Seed Centre of Fujian, Fuzhou 350003, China; 4 College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 5 Fujian Engineering and Technology Research Center of Breeding and Utilization for Special Crop, Fuzhou 350002, China; 6 Key Laboratory of Genetics, Breeding and Multiple Application of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou 350002, China)

收稿日期: 2020-11-25 网络首发时间: 2021-06-17 16:14:51
网络首发地址: <https://kns.cnki.net/kcms/detail/44.1110.S.20210617.1606.004.html>
作者简介: 黄小芳 (1995—), 女, 硕士研究生, E-mail: 1102718600@qq.com; 通信作者: 林世强 (1979—), 男, 讲师, 博士, E-mail: linshiqiang@fafu.edu.cn; 陈选阳 (1972—), 男, 研究员, 博士, E-mail: cxy@fafu.edu.cn
基金项目: 福建省科技厅项目 (2019N0050)

Abstract: 【Objective】 To mine the sequence information of the β -amylase gene family of *Ipomoea batatas* genome, and analyze the structure and function of genes. 【Method】 Based on the whole genome sequence data of *I. batatas* cultivar ‘Taizhong 6’, the bioinformatic methods were applied to analyze the identified 12 members of the β -amylase gene family and conduct the domain conservation analysis, chromosomal localization, screening of potential duplication genes, conservative motif analysis, phylogenetic tree construction. The gene expression under low temperature stress was analyzed using the transcriptomics data. 【Result】 Twelve β -amylase genes were located on chromosomes No. 2, 4, 5, 6, 11, 12, 13 and 14 of *I. batatas*, and eight pairs showed potential duplication relationship. Multiple sequence alignment and functional domain search indicated that there were three highly conserved domains and 10 conservative motifs in the amino acid sequences of *I. batatas* β -amylase family. Phylogenetic trees of β -amylase proteins in *I. batatas* and other species showed that 62 β -amylase family members were divided into seven subgroups of S1–S7. The β -amylases of *I. batatas* were mainly distributed in the subgroups of S2, S4, S5, S6 and S7, most of which belonged to the same branches with *Arabidopsis thaliana*, *Solanum tuberosum* and *S. lycopersicum*. The results of transcriptomics data showed that six β -amylase genes expressed differentially during the low temperature storage period, of which two were up-regulated and four were down-regulated in ‘Xushu 15-1’, while only two genes were down-regulated in ‘Xushu 15-4’. 【Conclusion】 The β -amylases are a key class of starch hydrolyzing enzymes that play important roles in the degradation of starch into reducing sugars during the process of *I. batatas* growth, development and tuber storage stages. The sequences of the identified 12 sweet potato β -amylase genes provide data reference for further study on the biological functions of *I. batatas* β -amylase gene family.

Key words: *Ipomoea batatas*; β -amylase; gene family; systematic evolution; differential expression analysis

淀粉是植物贮藏器官(块根、块茎)以及种子的主要营养成分之一,利用淀粉酶将储藏的淀粉分解为寡糖或单糖,可为植物的生长发育提供碳源和能源^[1]。根据淀粉酶水解淀粉方式的不同,可分为 α -淀粉酶和 β -淀粉酶。其中, β -淀粉酶是广泛存在于植物和少数微生物中的一种外切水解酶,只能从淀粉侧链的非还原性末端开始,以麦芽糖为单位切割 α -1,4-糖苷键,产生 β 型麦芽糖和限制性糊精^[2]。 β -淀粉酶作为自然界中将淀粉降解为麦芽糖的主要作用酶,广泛应用于啤酒、饴糖以及饮料等工业生产中,是食品加工和酿造行业的重要酶源^[3]。

甘薯 *Ipomoea batatas* 是我国重要的农作物,具有适应能力强、种植面积广以及营养价值高等特点。有研究表明, β -淀粉酶是甘薯块根中的一种蛋白质成分,含量仅次于甘薯贮藏蛋白,约占根可溶性蛋白的 5%^[4]。除此之外,甘薯在蒸煮过程中所产生的香气也部分来源于 β -淀粉酶水解淀粉所产生的麦芽糖,且 β -淀粉酶含量高的甘薯的甜度也明显高于含量低的甘薯^[5]。近年来甘薯 β -淀粉酶的研究陆续报道,例如梁新红等^[6]分析甘薯中 β -淀粉酶的酶解条件时发现,甘薯 β -淀粉酶酶解产物主要为麦芽糖,而麦芽糖可以诱导枯草芽孢杆菌发酵产生中温 α -淀粉酶;陈显让等^[7]对甘薯块根膨大后期

β -淀粉酶活性进行测定,发现在薯块膨大后期 β -淀粉酶活性多呈下降趋势,且不同品系间的酶活性差异较大。甘薯块根中的 β -淀粉酶虽然很早就被分离纯化,对其晶体结构也进行了较为深入的研究^[8],但是对甘薯 β -淀粉酶基因家族序列分析和 β -淀粉酶基因在甘薯生长发育过程中发挥的生物学功能作用的研究还较少。

2017 年中国科学院上海辰山植物科学研究中心和植物生理生态研究所联合德国马克斯普朗克分子遗传研究所和分子植物生理研究所,利用新型单倍型基因组装配方法对甘薯品种‘泰中 6 号’进行全基因组测序,获得了首个甘薯栽培种的全基因组测序结果^[9]。本研究以此测序数据为基础,从中挖掘出甘薯 β -淀粉酶基因家族序列,分析它们的序列结构域特征、系统进化关系以及低温贮藏期间基因表达量的变化情况,为进一步研究甘薯 β -淀粉酶基因的功能和进化奠定基础。

1 材料与方法

1.1 材料

以甘薯‘泰中 6 号’全基因组测序数据为研究对象,利用 snap 程序检索甘薯 15 条染色体序列

的 CDS 区,得到 178 458 个甘薯基因序列^[10]。基因组数据来源于 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/genome/?term=Ipomoea+batatas>),检索所采用的隐马尔可夫模型 (HMM) 为拟南芥 (At.hmm)、线虫 (Ce.hmm) 和水稻 (Os.hmm)。

1.2 甘薯 β -淀粉酶家族成员检索与鉴定

以从 TAIR 数据库 (<https://www.arabidopsis.org/index.jsp>) 和 RAP 数据库 (<http://rice.plantbiology.msu.edu/index.shtml>) 下载的拟南芥和水稻 β -淀粉酶蛋白序列为基础,通过 blastp^[11] 构建 β -淀粉酶序列数据库,将“1.1”获得的甘薯 DNA 序列翻译成蛋白序列,与数据库中序列进行比对,获得甘薯 β -淀粉酶基因家族候选序列。进一步从 Pfam 数据库 (<http://pfam.xfam.org/>) 下载 β -淀粉酶保守结构域 Glyco_hydro_14 的 HMM(PF01373),利用 hmmsearch 检索甘薯全基因组蛋白序列,获得甘薯 β -淀粉酶基因家族候选序列。合并 2 次的检索结果,利用 CDD^[12] 和 Interproscan^[13] 对获得的候选蛋白进行验证,删除其中不含有 β -淀粉酶保守结构域的序列。含有 β -淀粉酶保守结构域的序列利用 Clustal Omega^[14] 进行多重序列比对,将其中保守区域氨基酸缺失的序列与转录组测序数据中的 reads 进行匹配,修补缺失的区域,获得含有完整的 β -淀粉酶保守结构域的甘薯 β -淀粉酶基因家族序列。

1.3 甘薯 β -淀粉酶基因家族染色体定位和潜在重复关系分析

从甘薯全基因组 CDS 序列的 ZFF 文件中提取甘薯 β -淀粉酶基因序列的染色体位置信息,利用 circos^[15] 绘制甘薯 β -淀粉酶基因家族染色体分布图。为研究甘薯 β -淀粉酶基因序列之间的联系,利用 blastn 查找出其中具有潜在重复关系的基因对^[16],并在图中用线条进行连接。

1.4 甘薯 β -淀粉酶蛋白序列保守基序查找

以甘薯 β -淀粉酶家族成员蛋白序列为基础,利用 MEME^[17] 检索其中的保守基序,设置检索数目为 10,其他参数保持默认值。检索得到的保守基序氨基酸序列提交至 SMART 网站 (<http://smart.embl-heidelberg.de/>) 进行注释查询。

1.5 甘薯 β -淀粉酶家族蛋白序列系统进化分析

从 Uniprot 数据库 (<https://www.uniprot.org/>) 和 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>) 下载番茄、大豆、小麦和玉米 β -淀粉酶家族的蛋白序列;从 PGSC 数据库 (<http://solanaceae.plantbiology.msu.edu/index.shtml>) 下载马铃薯 β -淀粉酶家族的

蛋白序列。将下载得到的番茄、大豆、小麦、玉米和马铃薯的 β -淀粉酶序列与“1.2”获得的拟南芥、水稻以及甘薯的 β -淀粉酶序列合并,利用 ClustalO 进行序列多重比对,得到的比对结果导入 MEGA X 软件^[18] 中选择 Protein Sequences 进行分析;选择 Neighbor-Joining(NJ) 方法中的 p-distance 模型构建植物 β -淀粉酶基因家族系统进化树,Bootstrap 值取 1 000 次,对序列中的空位选择 Pairwise deletion 的处理方法。

1.6 甘薯 β -淀粉酶基因低温贮藏期基因表达量变化

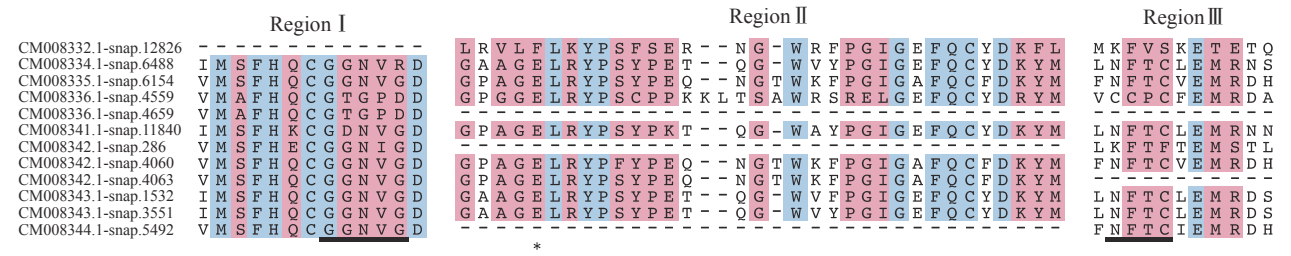
Ji 等^[19] 把‘徐薯 15-1’和‘徐薯 15-4’在 4℃ 条件下贮藏 0、2、6 周后,取其薯块进行转录组测序。本研究从 EBI 数据库 (<https://www.ebi.ac.uk/>) 下载该转录组测序数据,与 Hisat2^[20] 所构建的甘薯 β -淀粉酶基因索引文件比对,手动统计每个基因序列所匹配的 reads 数。获得的 reads 数值经过 DESeq2^[21] 标准化后,导入 R 软件中利用 pheatmap 绘制甘薯 β -淀粉酶基因在低温贮藏过程中的差异表达聚类图。

2 结果与分析

2.1 甘薯 β -淀粉酶基因鉴定和保守结构域分析

利用 HMM 检索甘薯全基因组蛋白序列,得到 16 个甘薯 β -淀粉酶家族候选成员;利用 blastp 检索得到 31 个甘薯 β -淀粉酶家族候选成员,其中包括 HMM 检索的全部结果,即 2 次检索一共得到 31 个 β -淀粉酶候选序列。甘薯 β -淀粉酶家族成员的确定主要根据 Interproscan 和 CDD 数据库鉴定 β -淀粉酶候选序列是否含有 Glyco_hydro_14 保守结构域,共鉴定到 12 个序列含有该保守结构域。Totsuka 等^[22] 分析了不同物种中的 β -淀粉酶氨基酸序列特征,发现 3 个高度保守的区域 (Region I、II 和 III),本研究将甘薯的 12 个 β -淀粉酶家族成员的氨基酸序列进行多重比对,同样定位到了类似的 3 个保守区域,结果如图 1 所示。图 1 中各保守区域内均有个别序列出现该区域缺失,这可能是由于基因组测序质量的限制^[23]。为了更全面地分析甘薯 β -淀粉酶基因家族,本研究将这些基因一并归入甘薯 β -淀粉酶基因家族中进行分析,甘薯 β -淀粉酶基因序列信息如表 1 所示。

3 个保守区域内均含有多个保守性较高的位点 (即氨基酸保守性在 70% 以上),Region I 中有 12 个,Region II 中有 25 个,Region III 中有 6 个。比较 3 个区域之间的氨基酸保守性,Region I 中的



蓝色阴影部分表示该位点氨基酸保守性为 100%，红色阴影部分表示该位点氨基酸保守性在 70% 以上；统计时各区域内由空位所代替的序列不纳入计算；Region I 画横线部分表示 flexible loop，Region II “*” 标注处表示催化活性位点，Region III 画横线部分表示 Inner loop

The blue shaded part indicates that the amino acid conservation of the site is 100%, and the red shaded part indicates that the amino acid conservation of the site is over 70%; The series replaced by vacancies in each region are excluded in the calculation; The horizontal line in Region I represents the Flexible loop, the “*” in Region II shows the catalytic active site, and the horizontal line in Region III denotes the Inner loop

图 1 甘薯 β -淀粉酶保守结构域分析

Fig. 1 Analysis of the conserved domains of β -amylase in *Ipomoea batatas*

表 1 甘薯 β -淀粉酶基因序列信息¹⁾

Table 1 The sequence information of β -amylase gene in *Ipomoea batatas*

基因名称 Gene name	外显子 Exon	链 Strand	长度/bp Length
CM008332.1-snap.12826	45177690-45177779; 45176575-45176835; 45176174-45176392; 45175888-45176091; 45175660-45175665; 45174982-45174999	—	798
CM008334.1-snap.6488	29784343-29784606; 29784692-29784856; 29785241-29785504; 29786090-2978630	+	909
CM008335.1-snap.6154	26986968-26987510; 26986519-26986728; 26986320-26986430; 26985282-26986064	—	1 647
CM008336.1-snap.4559	14557970-14558494; 14560135-14560344; 14560435-14560551; 14560653-14561441; 14561707-14561766	+	1 701
CM008336.1-snap.4659	14866852-14867396; 14867730-14867844; 14868220-14868330	+	771
CM008341.1-snap.11840	40434808-40434870; 40433953-40434363; 40433695-40433859; 40432752-40433015; 40432065-40432280; 40431587-40431988	—	1 521
CM008342.1-snap.286	1511146-1511589; 1512051-1512279; 1512390-1512481; 1512728-1512931; 1513310-1513504; 1514282-1514491; 1514642-1514861; 1515345-1515373	+	1 623
CM008342.1-snap.4060	14068276-14068818; 14069053-14069265; 14069351-14069461; 14069709-14070491	+	1 650
CM008342.1-snap.4063	14080849-14081390; 14081696-14081906; 14082004-14082114; 14082365-14082523	+	1 023
CM008343.1-snap.1532	7100890-7100952; 7101137-7101535; 7101690-7101854; 7101930-7102196; 7102266-7102487; 7102713-7102781	+	1 185
CM008343.1-snap.3551	12195237-12195299; 12195510-12195908; 12196093-12196257; 12196351-12196617; 12196701-12196940	+	1 134
CM008344.1-snap.5492	22226248-22226507; 22227548-22228337	+	1 050

1) 基因名称由以“-”分隔的染色体序列文件名和基因编号2部分组成,例如CM008332.1-snap.12826表示NCBI中‘泰中6号’基因组^[9]的染色体序列文件CM008332.1.fasta(第2号染色体)上根据snap预测^[10]的编号为snap.12826的基因;基因由外显子组成,表中外显子的位置区间均以正义链坐标为准,链为“+”的基因取正义链的序列,链为“-”的基因取正义链反向互补的序列,外显子排列顺序按照其在基因中出现的顺序,拼接外显子即可得到基因全长序列

1) The gene name consists of the name of chromosome sequence file and the gene number that are separated by the symbol “-”, for example, the CM008332.1-snap.12826 indicates the gene numbered snap.12826, which is predicted with snap^[10] in the chromosome sequence file CM008332.1.fasta (chromosome No. 2) of the ‘Taizhong 6’ genome^[9] in NCBI; The genes are made up of exons, in the table, the positional interval of the exon is set according to the sense strand, the sequence of gene with strand “+” is taken from the sense strand and the sequence of gene with the strand “-” is taken from the reverse compliment of the sense strand, the exons are arranged corresponding to the sequential position within the gene, the full-length gene can be obtained via merging the exons

氨基酸组成显示出更高的保守性,该区域总共有 13 个氨基酸位点,有 6 个位点的氨基酸组成高度保守,其中还含有 1 个 Flexible loop(画横线部分),共有序列的氨基酸组成为 GGNVG。Region II 和 Region III 的区域保守性虽然没有 Region I 高,但在 Region II 中含有与 β -淀粉酶催化活性高度相关的谷氨酸 (Glu, E) 催化位点^[24-25],即“*”标注处;Region III 中则含有一个 Inner loop(画横线部分),共有序列的氨基酸序列组成为 NFTC。

2.2 甘薯 β -淀粉酶基因家族在染色体上的分布

鉴定得到的 12 个甘薯 β -淀粉酶基因家族成员分布在第 2、4、5、6、11、12、13 和 14 号染色体上(图 2)。研究表明,2 个 DNA 序列比对后相似性 (Similarity) 大于 75% 且比对覆盖区域超过较长序列 75% 的基因之间存在潜在重复关系^[16]。本研究从甘薯 β -淀粉酶基因家族中鉴定得到 8 个具有潜在重复关系的基因对,并在图中用红色线条进行连接。已知 8 个具有潜在重复关系的基因对中,2 对位于同一染色体上,分别为第 12 号染色体上的 *CM008342.1-snap.4060* 和 *CM008342.1-snap.4063* 以及第 13 号染色体上的 *CM008343.1-snap.1532* 和 *CM008343.1-snap.3551*;

第 12 号染色体上的基因对在染色体上紧密排列,形成一个基因簇。处于染色体间的潜在重复基因有 6 对,其中第 4 号染色体上的 *CM008334.1-snap.6488* 分别与第 11 号染色体上的 *CM008341.1-snap.11840* 以及第 13 号染色体上的 *CM008343.1-snap.1532* 和 *CM008343.1-snap.3551* 构成 3 个基因对,而 *CM008341.1-snap.11840* 又与 *CM008343.1-snap.1532* 基因形成 1 个潜在重复基因对。除此之外,第 5 号染色体上的基因和第 12 号染色体上的 2 个基因之间也有 2 个具有潜在重复关系的基因对。

2.3 甘薯 β -淀粉酶保守基序分析

利用 MEME 从甘薯 β -淀粉酶氨基酸序列中检索到 10 个保守基序(图 3),各基序排列顺序基本遵循如下规律: Motif 8—Motif 7—Motif 1—Motif 3—Motif 6—Motif 4—Motif 5—Motif 2—Motif 10—Motif 9。从图 3 可以发现,甘薯 β -淀粉酶各成员的 motif 组成数量有差异,如 *CM008335.1-snap.6154*、*CM008336.1-snap.4559*、*CM008334.1-snap.11840* 和 *CM008342.1-snap.4060* 4 个成员含有

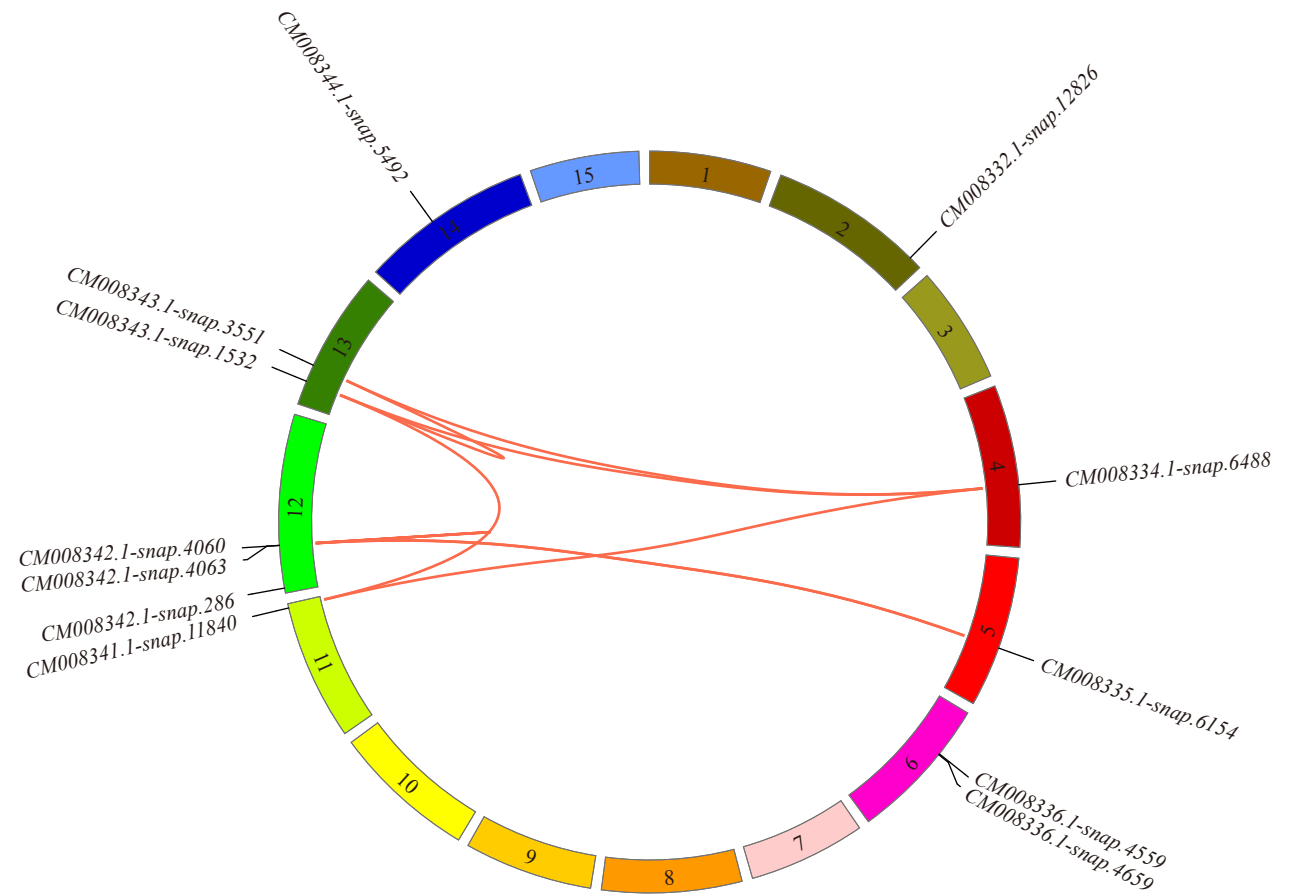


图 2 甘薯 β -淀粉酶基因家族染色体定位及潜在重复关系分析

Fig. 2 Chromosomal localization and potential duplication relationship of the β -amylase gene family in *Ipomoea batatas*

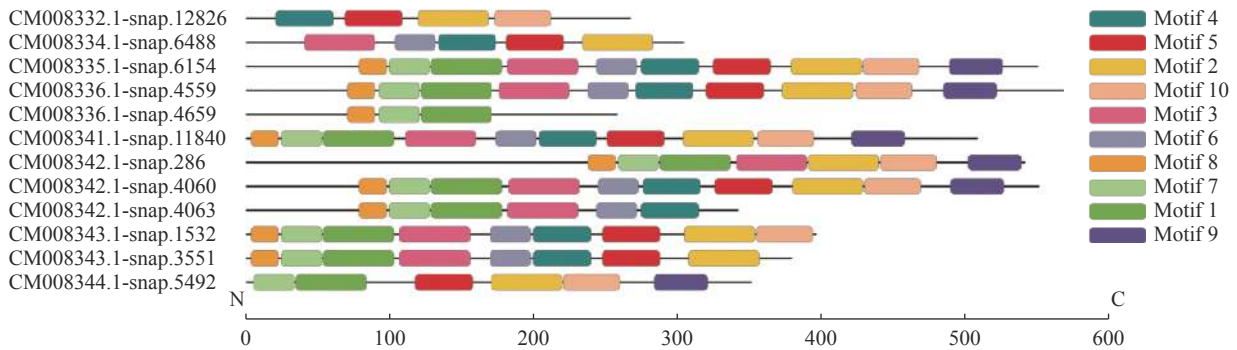


图 3 甘薯 β -淀粉酶蛋白序列保守基序分析

Fig. 3 Conservative motif analysis of protein sequences of *Ipomoea batatas* β -amylases

完整的 10 个 motif; CM008336.1-snap.4659 和 CM008332.1-snap.12826 中所含的 motif 数则较少, 分别仅有 3 个和 4 个。每个 motif 在同一个序列中最多仅出现 1 次, 不同 motif 在甘薯 β -淀粉酶家族序列中出现的次数并不相同。在 80% 以上的成员中均有检索到 Motif 7、Motif 1 和 Motif 2, 其中 Motif 7 和 Motif 1 总是串联出现, 在序列上紧密排列; Motif 3、Motif 4、Motif 5、Motif 6、Motif 8 和 Motif 10 出现的频率在 60% 以上; 个别 motif 出现的次数较少, 例如 Motif 9, 仅有 6 个 β -淀粉酶序列含有该保守基序。

表 2 中为 Motif 1~Motif 10 的氨基酸序列信息。通过 SMART 检索得到的功能注释结果。Motif 1、Motif 2、Motif 3、Motif 4、Motif 5 和 Motif 10 检索结果均为 Glyco_hydro_14 结构域; Motif 9 根据蛋白质的结构分类 (Structural classification of proteins, SCOP) 被归为 d1byb_结构域, 该结构域属于 β -淀粉酶蛋白^[26]。分析氨基酸组成特征, 发现在一些保守基序中可以找到与图 1 对应的 3 个保守区域 (画横线部分的氨基酸), 如 Motif 1 和 Motif 2 中分别含有 β -淀粉酶 3 个保守区域的 Region I 和 Region II; Motif 6 和 Motif 4 则包含了 Region III。

表 2 甘薯 β -淀粉酶保守基序氨基酸组成特征¹⁾

Table 2 Amino acid composition of the conservative motifs of the *Ipomoea batatas* β -amylase

基序 Motif	氨基酸组成 Amino acid composition	结构域 Structural domain
Motif 1	VDVWWGLVEKDSPREYNWAGYSELLQLAKKHGLKVQAVMSFHQCGGNVGD	Glyco_hydro_14
Motif 2	GIHWYWGTRSHAAELTAGYYNTRGRDGYLPIARMLARHGATLNFTCLEMR	Glyco_hydro_14
Motif 3	IPLPRWVLEEgDKNPdIFYTDRAGRRNYEYLSLgVDNQPLFKGRTPLQMY	Glyco_hydro_14
Motif 4	WYFPGIGEFQCYDKYMVASWKGAEEAAGHPewGMPGPTDAG	Glyco_hydro_14
Motif 5	TEFFRENGTYNTDYgKFflTWYSQMLIIHGDRILQEANKVF	Glyco_hydro_14
Motif 6	FLLGGTIVDIQVGMGPAGELRYPSPETQ	—
Motif 7	VKMDHTMNRKKAMEVSLQALKSAGVEGVM	—
Motif 8	IPKMMSRSRGVPVFVMLPLD	—
Motif 9	MCAFTYLRMNPELFearnWiqFVGfVKkMKEGEQRREC	SCOP domain d1byb_
Motif 10	DHEQPQHAQCAPEKLvWQvLLATWEARVPLAGENALPRYD	Glyco_hydro_14

1) 下划线表示与图 1 对应的保守区域部分; “—” 表示未匹配到相关信息

1) The underlines indicate the conservative regions corresponding to Fig. 1; “—” indicates no relevant information is matched

2.4 甘薯及其他物种 β -淀粉酶系统进化树分析

为了研究不同物种中 β -淀粉酶基因的系统进化关系, 下载了 7 个不同物种的 β -淀粉酶蛋白序列, 与甘薯 β -淀粉酶蛋白序列合并后进行多重序列比对, 利用 NJ 法构建系统发育树。这 7 个不同物种共 50 个蛋白序列分别为水稻 10 个、玉米 12 个、大

豆 3 个、拟南芥 9 个、马铃薯 7 个、番茄 7 个以及大麦 2 个。各个物种 β -淀粉酶蛋白序列的数据库查询编号以及所对应的简化后的蛋白名称如表 3 所示。

甘薯、拟南芥、马铃薯、番茄、大豆、水稻、玉米以及大麦 8 个物种共 62 个 β -淀粉酶家族成员被分

表 3 β -淀粉酶蛋白数据库查询编号
Table 3 The database query numbers of β -amylase proteins

物种 Species	蛋白名称 Protein name	数据库查询编号 Database query number	物种 Species	蛋白名称 Protein name	数据库查询编号 Database query number
水稻 <i>Oryza sativa</i>	OsBAM1	LOC_Os02g03690.1	拟南芥 <i>Arabidopsis thaliana</i>	AtBAM1	AT3G23920.1
	OsBAM2	LOC_Os07g35940.1		AtBAM2	AT4G00490.1
	OsBAM3	LOC_Os03g04770.1		AtBAM3	AT4G17090.1
	OsBAM4	LOC_Os07g47120.1		AtBAM4	AT5G55700.1
	OsBAM5	LOC_Os03g22790.1		AtBAM5	AT4G15210.1
	OsBAM6	LOC_Os09g39570.1		AtBAM6	AT2G32290.1
	OsBAM7	LOC_Os10g32810.1		AtBAM7	AT2G45880.1
	OsBAM8	LOC_Os01g13550.1		AtBAM8	AT5G45300.1
	OsBAM9	LOC_Os10g41550.1		AtBAM9	AT5G18670.1
	OsBAM10	LOC_Os07g35880.1	马铃薯 <i>Solanum tuberosum</i>	StBAM1	PGSC0003DMP400002800
玉米 <i>Zea mays</i>	ZmBAM1	NP_001148159.2		StBAM3	PGSC0003DMP400035625
	ZmBAM2	NP_001151271.2		StBAM4	PGSC0003DMP400021443
	ZmBAM3	XP_008658465.1		StBAM5	PGSC0003DMP400045472
	ZmBAM4	NP_001168436.1		StBAM7	PGSC0003DMP400000367
	ZmBAM5	NP_001130896.1		StBAM8	PGSC0003DMP400041754
	ZmBAM6	AQK60892.1		StBAM9	PGSC0003DMP400018848
	ZmBAM7	NP_001354441.1	番茄 <i>Solanum lycopersicum</i>	SIBAM1	A0A3Q7I9I2
	ZmBAM8	NP_001105496.2		SIBAM2	A0A3Q7EKY0
	ZmBAM9	NP_001170007.1		SIBAM3	A0A3Q7HEM6
	ZmBAM10	NP_001337631.1		SIBAM4	A0A3Q7IDQ8
	ZmBAM11	NP_001132696.1		SIBAM5	A0A3Q7IEI0
	ZmBAM12	XP_035818275.1		SIBAM6	A0A3Q7HUA2
				SIBAM7	A0A3Q7HV50
大豆 <i>Glycine max</i>	GmBAM1	P10538.3	大麦 <i>Hordeum vulgare</i>	HvBAM1	CAC16789.1
	GmBAM2	CAI39245.1		HvBAM2	AAX37357.1
	GmBAM3	CAI39244.1			

为 S1~S7 7 个亚组, 如图 4 所示, 树状图分枝上仅显示大于 70% 的 Bootstrap 值。图 4 中显示的蛋白名称与表 3 相对应。各个亚组中所包含的 β -淀粉酶家族成员数并不相同, S7 亚组包含的成员数最多, 有 15 个; S4 和 S3 亚组次之, 分别有 11 个和 8 个; S1、S2 和 S6 亚组成员数均为 7 个; S5 亚组最少, 仅有 6 个。甘薯 12 个 β -淀粉酶家族成员主要分布于 S2(2 个)、S4(4 个)、S5(1 个)、S6(1 个) 以及 S7(4 个) 亚组, S1 和 S3 亚组中不含有甘薯 β -淀粉酶家族成员。从图 4 可以发现, 甘薯的 β -淀粉酶与番茄、马铃薯和拟南芥的 β -淀粉酶亲缘关系较为接近, 在进化树上大多处于同一分支, 例如甘薯 CM008342.1-snap.286 与马铃薯 StBAM7、番茄 SIBAM2 以及拟南芥 AtBAM7 处于 S6 亚组的同一分支; S2 亚组中的 CM008336.1-snap.4559 和

CM008336.1-snap.4659 与马铃薯 StBAM4 和番茄 SIBAM7 为同一进化树分支, 且在图 2 的染色体定位结果中 CM008336.1-snap.4559 和 CM008336.1-snap.4659 所对应的基因在染色体上的位置相近。

2.5 甘薯 β -淀粉酶基因低温胁迫条件下的差异表达

利用转录组测序数据^[19] 分析得到了甘薯在低温贮藏期间 β -淀粉酶基因的表达量变化模式图(图 5)。图 5 共有 4 个分组, 分别为具有不同耐低温能力的甘薯品种 ‘徐薯 15-1(高)’ 和 ‘徐薯 15-4(低)’ 在 4 ℃ 条件下贮藏 2 和 6 周后, β -淀粉酶基因表达量相比于对照组(贮藏 0 周)的变化情况。甘薯 12 个 β -淀粉酶基因中共有 6 个基因在低温贮藏过程中表达量值出现上调(红色)或下调(绿色)变化。其中, ‘徐薯 15-1’ 有 4 个基因下调表

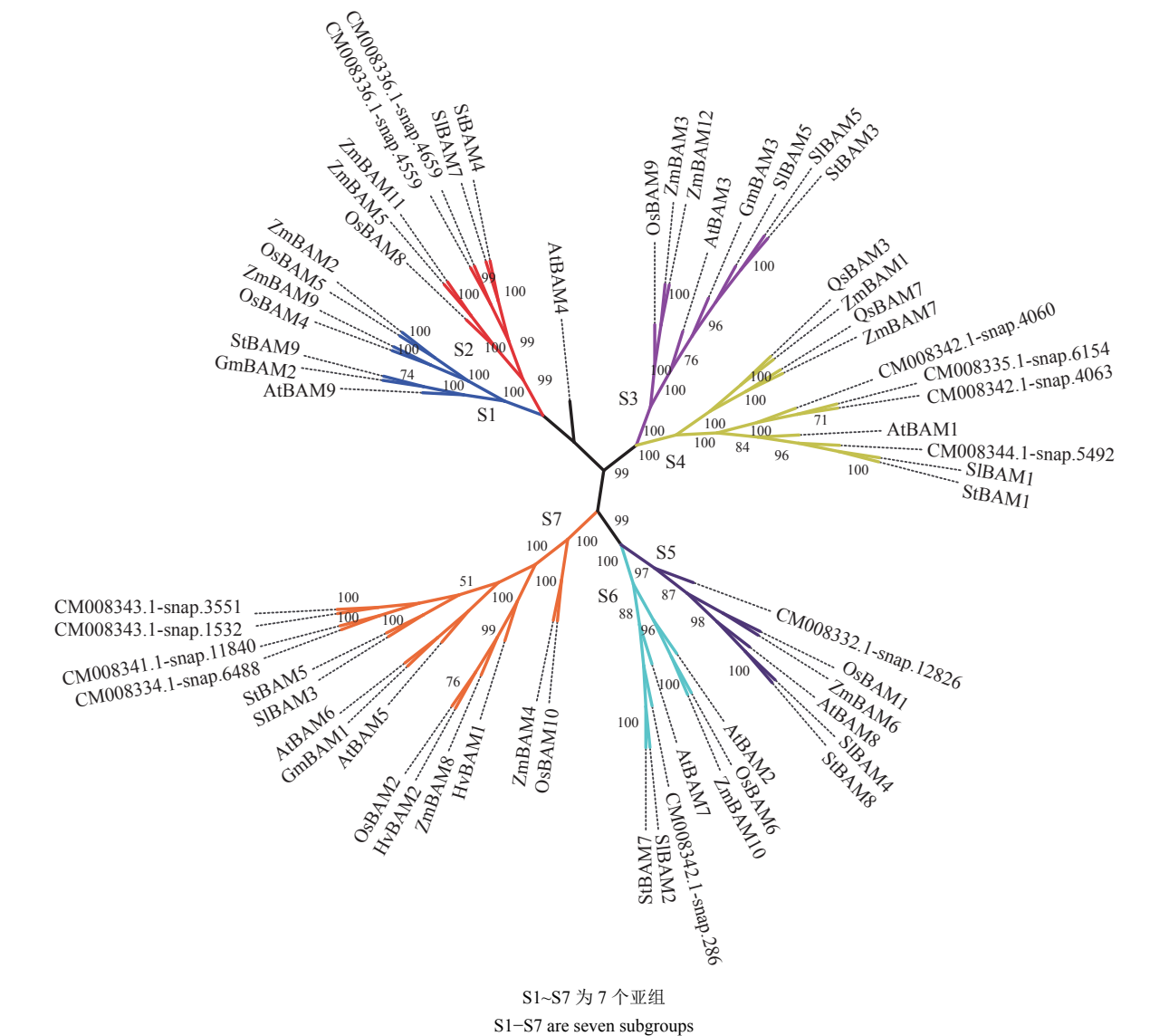


图 4 甘薯与拟南芥、马铃薯、番茄、大豆、水稻、玉米以及大麦的 β -淀粉酶系统进化分析

Fig. 4 Phylogenetic analysis of β -amylases in eight species, i.e. *Ipomoea batatas*, *Arabidopsis thaliana*, *Solanum tuberosum*, *Solanum lycopersicum*, *Glycine max*, *Oryza sativa*, *Zea mays* and *Hordeum vulgare*

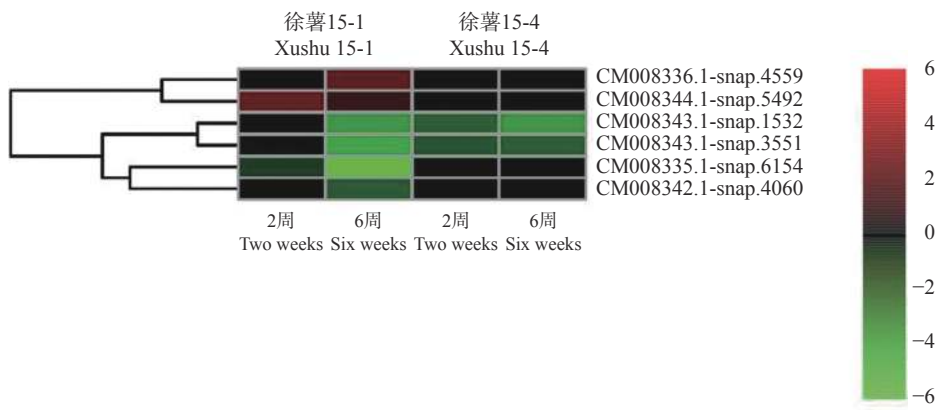


图 5 甘薯 β -淀粉酶基因低温胁迫下不同贮藏时间的表达分析

Fig. 5 Expression analysis of β -amylase gene in *Ipomoea batatas* at different storage time under low temperature stress

达, 2 个基因上调表达; 而在 ‘徐薯 15-4’ 中则仅有 2 个下调表达的基因。在 ‘徐薯 15-1’ 中, 储藏 2 周时仅有 2 个基因表达量发生变化, 储藏 6 周后则有 6 个基因表达量出现变化, 说明随着贮藏时间

的延长, ‘徐薯 15-1’ 中响应低温胁迫的基因数有所增加。

CM008343.1-snap.1532 和 *CM008343.1-snap.3551* 在 ‘徐薯 15-1’ 和 ‘徐薯 15-4’ 中均表现为下调, 在 ‘徐薯 15-1’ 中贮藏 2 周时并未出现表达量的变化, 6 周时才出现明显的下调趋势; 在 ‘徐薯 15-4’ 中贮藏 6 周时的表达量下调变化比 2 周时更明显, 复制基因分析发现这 2 个基因之间具有潜在重复关系 (图 2)。大部分甘薯 β -淀粉酶基因在低温贮藏过程中下调表达, 但 *CM008336.1-snap.4559* 和 *CM008344.1-snap.5492* 上调表达, 推测这 2 个基因可能在甘薯块根受到低温胁迫时发挥作用。

3 讨论与结论

本研究利用 HMM 法从甘薯全基因组中鉴定得到 12 个甘薯 β -淀粉酶家族成员, 这些成员均含有糖苷水解酶 14 家族的保守结构域 Glyco_hydro_14, 数量与拟南芥 9 个^[27]、水稻 10 个以及玉米 13 个等^[28] 差异不大。植物中 β -淀粉酶家族属于多基因家族, 序列中功能结构域的保守性和保守基序的组成是多基因家族的重要特征。研究表明, 甘薯 β -淀粉酶氨基酸序列中含有 3 个保守性较高的区域 (region I~III), 其中 region II 中的谷氨酸催化位点是影响甘薯 β -淀粉酶催化活性的有效部位^[2]。甘薯 β -淀粉酶家族序列中保守基序的组成并不完全相同, 若干存在于大多数 β -淀粉酶成员中的保守基序体现了甘薯 β -淀粉酶家族的共性, 例如 motif 1、motif 2 和 motif 7; 多基因家族在进化过程中往往会出现功能的分化, 一些仅存在于少部分 β -淀粉酶成员中的保守基序造成了甘薯 β -淀粉酶家族成员序列之间的差异, 这些差异可能与甘薯 β -淀粉酶家族在进化过程中功能的分化相关。

将甘薯与其他物种的 β -淀粉酶序列共同构建系统进化树, 以分析甘薯 β -淀粉酶的进化关系和可能的功能特征。根据树状图的分支特点和遗传距离的远近将得到的进化树结果分为 7 个亚组 (S1~S7), Thalmann 等^[24] 对陆地植物中的 β -淀粉酶家族进行分析, 系统进化分析结果中将 8 种不同植物的 β -淀粉酶分为 8 个亚组 (I~VIII), 其中拟南芥的 9 个 β -淀粉酶家族成员被归入其中的 7 个亚组, 分别为 BAM1(I 组)、BAM3(III 组)、BAM4(IV 组)、BAM9(V 组)、BAM5 和 BAM6(VI 组)、BAM2 和 BAM7(VII 组) 以及 BAM8(VIII 组)。本研究中, 拟南芥的 9 个成员同样分散于不同的亚组, 且各成员

的分类与 Thalmann 等^[24] 的分类结果相一致, 例如同样的 AtBAM2 和 AtBAM7 为一组 (S6), AtBAM5 和 AtBAM6 为一组 (S7), 其他成员则单独归入不同的亚组中。本研究中甘薯 β -淀粉酶成员所处的分支基本由拟南芥、马铃薯和番茄的 β -淀粉酶成员组成, 尤其与亲缘关系相近的马铃薯和番茄 β -淀粉酶家族成员在进化上拥有更紧密的联系。通过分析甘薯 β -淀粉酶家族成员遗传距离相近的马铃薯或番茄 β -淀粉酶家族成员的功能, 对甘薯 β -淀粉酶基因的功能进行预测, Hou 等^[29] 研究表明沉默马铃薯 *StBAM1* 基因会降低马铃薯中的 β -淀粉酶活性, 导致淀粉积累。甘薯 β -淀粉酶基因 *CM008344.1-snap.5492* 与 *StBAM1* 基因处于同一分支, 蛋白序列比对结果显示两者的相似性为 86%, 推测 *CM008344.1-snap.5492* 与 *StBAM1* 基因之间可能存在功能相似性, 转录组测序结果显示 *CM008344.1-snap.5492* 基因在低温贮藏过程中上调表达; 因此 *CM008344.1-snap.5492* 基因很可能对甘薯低温贮藏过程中淀粉的降解发挥作用。

本研究基于甘薯 ‘泰中 6 号’ 全基因组测序数据鉴定得到甘薯 β -淀粉酶家族成员, 由于基因组测序质量的限制^[23], 在个别序列组成上存在部分区域氨基酸的缺失, 因而本研究结合 SRA 中的转录组测序数据, 对鉴定得到的甘薯 β -淀粉酶序列进行修补。研究表明, β -淀粉酶是甘薯块根中的主要酶^[30]。甘薯收获后通常是在低温贮藏, 在贮藏的过程中甘薯薯块中的 β -淀粉酶可将淀粉降解为还原糖, 提高甘薯的甜度, 同时还可促进薯块发芽^[31]; 因此, 研究 β -淀粉酶在甘薯低温贮藏过程中的活性变化对指导淀粉专用型甘薯品种的育种以及控制鲜食甘薯品种的采后品质具有重要意义。Ji 等^[19] 通过高通量测序技术获得了 ‘徐薯 15-1’ 和 ‘徐薯 15-4’ 的块根在低温贮藏条件下的转录组数据, 并利用 qRT-PCR 技术进行了可靠性验证; 本研究基于该转录组数据分析了甘薯 β -淀粉酶基因在低温贮藏过程中基因表达量的变化情况, 获得了 6 个差异表达基因, 这些结果可为甘薯 β -淀粉酶基因功能的研究提供重要参考。

参考文献:

[1] 张勇为, 张义正, 谭文芳, 等. 甘薯贮藏期间淀粉酶种类变化及其部分性质分析[J]. 四川大学学报(自然科学版), 2018, 55(1): 197-200.
[2] TODA H, NITTA Y, ASANAMI S, et al. Sweet potato β -amylase: Primary structure and identification of the active-site glutamyl residue[J]. *European Journal of Bio-*

chemistry, 1993, 216(1): 25-38.

[3] 孙俊良, 梁新红, 贾彦杰, 等. 植物 β -淀粉酶研究进展[J]. 河南科技学院学报(自然科学版), 2011, 39(6): 1-4.

[4] LI H S, ÔBA K. Major soluble proteins of sweet potato roots and changes in proteins after cutting, infection, or storage[J]. Agricultural and Biological Chemistry, 2014, 49(3): 737-744.

[5] NAKAMURA K, OHTO M A, YOSHIDA N, et al. Sucrose-induced accumulation of beta-amylase occurs concomitant with the accumulation of starch and sporamin in leaf-petiole cuttings of sweet potato[J]. Plant Physiology, 1991, 96(3): 902-909.

[6] 梁新红, 李英, 孙俊良, 等. β -淀粉酶酶解甘薯淀粉条件分析[J]. 食品工业科技, 2014, 35(7): 178-181.

[7] 陈显让, 李红兵, 康乐, 等. 甘薯块根膨大后期 β -淀粉酶和淀粉含量相关性分析[J]. 食品工业科技, 2013, 34(19): 93-96.

[8] CHEONG C G, EOM S H, CHANG C, et al. Crystallization, molecular replacement solution, and refinement of tetrameric beta-amylase from sweet potato[J]. Proteins, 1995, 21(2): 105-117.

[9] YANG J, MOEINZADEH M H, KUHL H, et al. Haplotype-resolved sweet potato genome traces back its hexaploidization history[J]. Nature Plants, 2017, 3(9): 696-703.

[10] 黄小芳, 毕楚韵, 石媛媛, 等. 甘薯基因组 NBS-LRR 类抗病家族基因挖掘与分析[J]. 作物学报, 2020, 46(8): 1195-1207.

[11] ALTSCHUL S F, GISH W, MILLER W, et al. Basic local alignment search tool[J]. Journal of Molecular Biology, 1990, 215(3): 403-410.

[12] LU S, WANG J, CHITSAZ F, et al. CDD/SPARCLE: The conserved domain database in 2020[J]. Nucleic Acids Research, 2020, 48(D1): D265-D268.

[13] QUEVILLON E, SILVENTOINEN V, PILLAI S, et al. InterProScan: Protein domains identifier[J]. Nucleic Acids Research, 2005, 33: W116-W120.

[14] SIEVERS F, HIGGINS D G. Clustal Omega for making accurate alignments of many protein sequences[J]. Protein Science, 2018, 27(1): 135-145.

[15] KRZYWINSKI M, SCHEIN J, BIROL I, et al. Circos: An information aesthetic for comparative genomics[J]. Genome Research, 2009, 19(9): 1639-1645.

[16] GU Z, CAVALCANTI A, CHEN F C, et al. Extent of gene duplication in the genomes of drosophila, nematode, and yeast[J]. Molecular Biology and Evolution, 2002, 19(3): 256-262.

[17] BAILEY T L, BODEN M, BUSKE F A, et al. MEME SUITE: Tools for motif discovery and searching[J]. Nucleic Acids Research, 2009, 37: W202-W208.

[18] KUMAR S, STECHER G, LI M, et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms[J]. Molecular Biology and Evolution, 2018, 35(6): 1547-1549.

[19] JI C Y, KIM H S, LEE C J, et al. Comparative transcriptome profiling of tuberous roots of two sweetpotato lines with contrasting low temperature tolerance during storage[J]. Gene, 2020: 727. doi: 10.1016/j.gene.2019.144244.

[20] KIM D, LANGMEAD B, SALZBERG S L. HISAT: A fast spliced aligner with low memory requirements[J]. Nature Methods, 2015, 12(4): 357-360.

[21] LOVE M I, HUBER W, ANDERS S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2[J]. Genome Biology, 2014, 15(12). doi: 10.1186/s13059-014-0550-8.

[22] TOTSUKA A, NONG V H, KADOKAWA H, et al. Residues essential for catalytic activity of soybean beta-amylase[J]. European Journal of Biochemistry, 1994, 221(2): 649-654.

[23] WU S, LAU K H, CAO Q, et al. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement[J]. Nature Communications, 2018: 9. doi: 10.1038/s41467-018-06983-8.

[24] THALMANN M, COIRO M, MEIER T, et al. The evolution of functional complexity within the β -amylase gene family in land plants[J]. BMC Evolutionary Biology, 2019: 19. doi: 10.1186/s12862-019-1395-2.

[25] KANG Y N, ADACHI M, UTSUMI S, et al. The roles of Glu186 and Glu380 in the catalytic reaction of soybean beta-amylase[J]. Journal of Molecular Biology, 2004, 339(5): 1129-1140.

[26] CHEN Y, CRIPPEN G M. An iterative refinement algorithm for consistency based multiple structural alignment methods[J]. Bioinformatics, 2006, 22(17): 2087-2093.

[27] VALERIO C, COSTA A, MARRI L, et al. Thioredoxin-regulated beta-amylase (BAM1) triggers diurnal starch degradation in guard cells, and in mesophyll cells under osmotic stress[J]. Journal of Experimental Botany, 2011, 62(2): 545-555.

[28] 杨泽峰, 徐暑晖, 王一凡, 等. 禾本科植物 β -淀粉酶基因家族分子进化及响应非生物胁迫的表达模式分析[J]. 科技导报, 2014, 32(31): 29-36.

[29] HOU J, ZHANG H, LIU J, et al. Amylases StAmy23, StBAM1 and StBAM9 regulate cold-induced sweetening of potato tubers in distinct ways[J]. Journal of Experimental Botany, 2017, 68(9): 2317-2331.

[30] HATTORI T, FUKUMOTO H, NAKAGAWA S, et al. Sucrose-induced expression of genes coding for the tuberous root storage protein, sporamin, of sweet potato in leaves and petioles[J]. Plant and Cell Physiology, 1991, 32(1): 79-86.

[31] 唐君, 周志林, 林冬兰, 等. 甘薯贮藏过程淀粉酶活性变化及对薯块芽萌发的影响[J]. 福建农业学报, 2010, 25(6): 699-702.