

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

杨泽峰, 徐晔晖, 王一凡, 张恩盈, 徐辰武

扬州大学农学院; 教育部植物功能基因组学重点实验室; 江苏省作物遗传生理重点实验室, 扬州 225009

摘要 β -淀粉酶(β -amylase, BAM)是一类关键的淀粉水解酶,在禾谷类作物生长发育过程中起着重要作用,与植物多种非生物胁迫响应相关。本研究通过系统发育分析,将水稻、玉米、高粱、谷子、二穗短柄草5种禾本科植物中共54个BAM基因分为10个同源基因簇,每个同源基因簇都涵盖了这5个物种,因此推测在禾本科祖先物种中至少含有10个BAM基因,并且在禾本科植物分化后没有发生明显的基因丢失事件。基于对编码蛋白质序列的功能分化分析,表明同源基因簇间存在明显的进化速率的差异。对10个同源基因簇进行了适应性进化检测,发现有3个同源簇在禾本科植物的进化过程中经历了适应性进化。此外,对水稻 β -淀粉酶的表达分析发现,一些 β -淀粉酶具有组织特异性表达特征,并且至少有5个水稻的 β -淀粉酶基因具有受到非生物逆境的胁迫而表现出不同的表达模式。本研究结果为进一步探讨禾本科BAM基因的生物学功能提供了一定的理论基础。

关键词 β -淀粉酶;同源基因簇;功能分化;正选择作用

中图分类号 Q943.2

文献标志码 A

doi 10.3981/j.issn.1000-7857.2014.31.002

Molecular Evolution and Expression Patterns Under Abiotic Stresses of Beta-amylase Gene Family in Grasses

YANG Zefeng, XU Shuhui, WANG Yifan, ZHANG Enying, XU Chenwu

Jiangsu Key Laboratory of Crop Genetics and Physiology; Key Laboratory of the Ministry of Education for Plant Functional Genomics; Agricultural College, Yangzhou University, Yangzhou 225009, China

Abstract Beta-amylase (BAM) plays a central role in the complete degradation of starch to metabolisable or fermentable sugars during the germination or malting of cereal grains. It was found to be involved in the abiotic stress responses of crops. A genome-wide survey of BAM genes in 5 grass species was performed, including rice, maize, sorghum, *Setaria*, and *Brachypodium*, by describing their phylogenetic relationships, functional divergence, and adaptive evolution. The phylogeny classified the gramineous BAM genes into 10 clusters of orthologous genes (COGs), and the genes in the same COG shared the syntenic region. Functional divergence analysis provided statistical evidence that both the shift in the evolutionary rate pattern and cluster-specific alterations of amino acid physiochemical properties contributed to COG-specific functional evolution of BAM genes in grasses. In addition, 3 COGs were found to be influenced by positive selection through maximum likelihood analysis. The expression patterns of rice BAM genes were investigated, and the results revealed that they were differentially expressed under the treatments of abiotic stresses. These observations may provide useful references for further functional detection of BAM genes in grasses.

Keywords beta-amylase; cluster of orthologous genes; functional divergence; positive selection

收稿日期:2014-09-16;修回日期:2014-10-08

基金项目:高等学校博士学科点专项科研基金项目(20123250110001);国家重点基础研究发展计划(973计划)项目(2011CB100100);国家自然科学基金基金项目(31371632,31200943,31171187);江苏省自然科学基金项目(BK2012261);江苏省高校自然科学研究重大项目(14KJA210005);江苏省“青蓝工程”科技创新团队项目

作者简介:杨泽峰,副教授,研究方向为植物功能基因组学与生物信息学,电子信箱:zfyang@yzu.edu.cn;徐辰武(通信作者),教授,研究方向为数量遗传学与生物信息学,电子信箱:qtl@yzu.edu.cn

引用格式:杨泽峰,徐晔晖,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.

淀粉是一种葡萄糖的高聚体,是植物体内储存的主要养分。细胞中淀粉代谢过程参与了很多植物本身的重要的生长发育过程,包括禾谷类种子萌发、叶片的光合作用、块根中养分储存以及肉质果的生长发育过程等^[1]。植物种子萌发过程中需要消耗大量淀粉,淀粉降解需要4种有活性的水解酶(GH): α -淀粉酶、 β -淀粉酶、极限糊精酶以及 α -葡萄糖水解酶。在这些酶的协同作用下,淀粉可以水解成为葡萄糖,从而为种子萌发初期提供养分^[2]。 β -淀粉酶是一种外切酶,当作用于淀粉时,将从非还原性末端逐次以麦芽糖为单位切断 α -1,4-葡聚糖链,从而产生麦芽糖和有限糊精^[3]。很多研究表明, β -淀粉酶在植物萌发以及生长发育过程中起到重要作用。例如,在谷类种子萌发中, β -淀粉酶在淀粉水解成可发酵性糖的过程中发挥了关键性作用^[3]。

糖苷酶亦称糖苷水解酶,是作用于各种糖苷或寡糖使糖苷键水解的酶的总称。 β -淀粉酶属于糖苷水解酶家族14,含有一个保守结构域 Glyco_hydro_14 结构域(PF01373)。这个保守结构域含有2个高度保守的区域,分别位于N端以及另外一个相对中心的位置^[4]。研究证明, β -淀粉酶在许多植物中表现为多基因家族。例如,拟南芥基因组中可能含有9个 β -淀粉酶基因(*BAM1*~*BAM9*)^[5]。其中,*AtBAM1*~*AtBAM4*为叶绿体类型并且在淀粉降解中发挥不同的作用;*AtBAM3*是夜间叶片淀粉降解主要的催化基因;而*AtBAM4*对*AtBAM1*和*AtBAM3*基因编码的酶活性有调控作用。拟南芥中2个在根系生长发育过程中起关键作用的转录因子*AtBAM7*和*AtBAM8*分别结合了一个顺式作用元件,其中都包含一个G box 和一个类似BR-反应元件^[6]。通过分子克隆等生物学方法已分别从水稻、大麦、黑麦、小麦、玉米中克隆出多个 β -淀粉酶编码基因^[7]。根据形式不同可将禾本科 β -淀粉酶分为2类:一类主要在胚乳中,一类普遍存在于植物的各个器官中。

禾本科植物是世界上分布最广泛、最丰富的开花植物,有着重要的经济价值。禾本科植物不仅仅是人类和牲畜粮食主要来源,其各部分器官还能加工成为各种副产品,例如,食糖、酒精等人类生活的必需品以及工业生产的原材料。 β -淀粉酶是禾本科植物淀粉分解过程中的一种关键酶。然而,关于 β -淀粉酶编码基因的适应性进化模式研究却较为鲜见。本研究利用生物信息学手段,对5种已完成测序工作的禾本科植物的 β -淀粉酶基因(*BAM*)的系统发育、适应性进化及不同非生物胁迫下的表达模式进行探讨。

1 材料与方法

1.1 禾本科植物 *BAM* 基因信息来源

从Phytozome V9.1数据库^[8]中下载获得5种禾本科植物(水稻、玉米、高粱、谷子、二穗短柄草) β -淀粉酶基因的相关序列信息,利用Pfam^[9]工具鉴定获得的候选*BAM*基因编码蛋白中是否含有保守结构域 Glyco_hydro_14 domain (PF01373),若存在,则确定为*BAM*基因并留做下一步研究。

1.2 序列分析软件

对蛋白质的多序列比对采用Clustal X软件^[10]进行,参数

为默认。系统发育树分别用MEGA 5.1^[11]软件中的Neighbor-joining (邻接法, NJ) 法、PhyML 3.0^[12]软件的Maximum likelihood (极大似然法, ML)法构建。ML树构建参数设置为:JTT模型,非可变位点估计比率,4种等级类型, gamma 分布参数估计以及最佳起始BIONJ树。JTT模型还可以用于构建NJ树。bootstrap重抽样进行了100次。进化树分枝长度和系统进化树构建采用PhyML软件完成。系统进化树的显示利用MEGA软件。

1.3 功能分化分析

为了估算基因家族内不同同源基因簇之间的功能分化,本研究采用DIVERGE^[13]软件。两个同源簇在某些位点的进化速率不同从而导致功能约束的改变,这样的功能分化称为I型功能分化;II型功能分化是指两个同源簇之间进化速率和早期不同,主要是理化性质的改变。同源簇之间功能分化系数 θ 是对这种进化速率的一种定量性描述。本研究利用DIVERGE软件通过计算不同同源簇之间I型和II型功能分化系数 θ_I 、 θ_{II} ,并通过似然比测验(LRT)对鉴定结果进行验证。

1.4 正选择作用检测

采用PAL2NAL软件^[14]将蛋白质序列的比对转换成相应的核苷酸序列,将结果输出到PAML软件^[15]作下一步处理。利用PAML软件中的codeml程序,采用极大似然检验法(M0/M3, M7/M8)检测不同位点的 ω 值(非同义替换率和同义替换率的比值, d_N/d_S)。M3/M0模型之间似然比检验了位点间是否存在选择压力的差异, M7/M8模型比较用于检测正选择作用,对这2对模型用统计量 $2\Delta\ln L$ 依 χ^2 分布进行似然比值检验(likelihood ratio test),自由度(DF)为两个模型参数数目的差值。本研究中M0/M3模型的自由度为3, M7/M8的自由度为2^[16]。

1.5 微阵列数据分析方法

微阵列数据从GEO开放数据库中下载得到,所用数据库的登录号分别为:GSE7951(水稻基因在9种组织中的表达数据)、GSE6901(水稻基因干旱、低温和盐胁迫处理下的表达数据)以及GSE14275(水稻基因在热激条件下的表达数据)。将所得微阵列数据用dChip v2010 (<http://www.hsph.harvard.edu/cli/complab/dchip/>)进行聚类分析并展示水稻*BAM*基因表达模式。采用集成在Microsoft Excel中的SAM^[17]软件比较所得基因表达量,标准采用2倍以上表达差异。

2 结果与分析

2.1 *BAM* 基因系统进化分析

大量研究表明,植物*BAM*是个多基因家族,本研究通过对5种禾本科植物(水稻、玉米、高粱、谷子以及二穗短柄草)全基因组研究证实了这一观点。首先,在Phytozome数据库中搜索了5种禾本科植物的*BAM*基因,并根据是否含有保守结构域(Glyco_hydro_14domain)对其进行鉴定,含有该保守结构域的基因被保留下来用做进一步分析。共鉴定出54个*BAM*基因(表1),其中水稻有10个,玉米13个,二穗短柄草11个,高粱和谷子各10个。这54个*BAM*基因分别根据它们在每个物种中的染色体位置命名。

表1 5种禾本科植物BAM基因基本信息

Table 1 Summary of 54 BAM genes identified in 5 gramineous genomes

物种	基因	编号	蛋白序列	内含子数目	染色体位置	位置
<i>Oryza sativa</i>	<i>OsBAM1</i>	LOC_Os01g13550	566	3	1	7578091~7582923
<i>Oryza sativa</i>	<i>OsBAM2</i>	LOC_Os02g03690	680	10	2	1539488~1548149
<i>Oryza sativa</i>	<i>OsBAM3</i>	LOC_Os03g04770	557	3	3	2272965~2276401
<i>Oryza sativa</i>	<i>OsBAM4</i>	LOC_Os03g22790	524	1	3	13161913~13164014
<i>Oryza sativa</i>	<i>OsBAM5</i>	LOC_Os07g35880	600	5	7	21471055~21473327
<i>Oryza sativa</i>	<i>OsBAM6</i>	LOC_Os07g35940	488	6	7	21503815~21506356
<i>Oryza sativa</i>	<i>OsBAM7</i>	LOC_Os07g47120	523	3	7	28178236~28181579
<i>Oryza sativa</i>	<i>OsBAM8</i>	LOC_Os09g39570	533	8	9	22713280~22717110
<i>Oryza sativa</i>	<i>OsBAM9</i>	LOC_Os10g32810	535	3	10	17180757~17183465
<i>Oryza sativa</i>	<i>OsBAM10</i>	LOC_Os10g41550	544	3	10	22324198~22327665
<i>Zea mays</i>	<i>ZmBAM1</i>	GRMZM2G450125	573	2	1	9057041~9060022
<i>Zea mays</i>	<i>ZmBAM2</i>	GRMZM2G347708	509	2	1	60022740~60024593
<i>Zea mays</i>	<i>ZmBAM3</i>	GRMZM2G082034	537	1	1	60087923~60089930
<i>Zea mays</i>	<i>ZmBAM4</i>	GRMZM2G007939	551	3	1	95188932~95192571
<i>Zea mays</i>	<i>ZmBAM5</i>	GRMZM5G803981	595	6	2	204170827~204173732
<i>Zea mays</i>	<i>ZmBAM6</i>	GRMZM2G175218	539	3	3	1536020~1540295
<i>Zea mays</i>	<i>ZmBAM7</i>	GRMZM2G069486	651	9	4	239081202~239093194
<i>Zea mays</i>	<i>ZmBAM8</i>	GRMZM2G025833	544	2	5	30775043~30777461
<i>Zea mays</i>	<i>ZmBAM9</i>	GRMZM2G058310	488	6	7	155357370~155360570
<i>Zea mays</i>	<i>ZmBAM10</i>	GRMZM2G181284	531	2	7	171860442~171862785
<i>Zea mays</i>	<i>ZmBAM11</i>	GRMZM2G446515	679	9	7	62300598~62305837
<i>Zea mays</i>	<i>ZmBAM12</i>	GRMZM2G077202	537	3	8	29211106~29216116
<i>Zea mays</i>	<i>ZmBAM13</i>	GRMZM2G035749	553	3	9	114176716~114179596
<i>Brachypodium distachyon</i>	<i>BdBAM1</i>	Bradi1g18390	537	2	1	14765285~14767168
<i>Brachypodium distachyon</i>	<i>BdBAM2</i>	Bradi1g25440	488	6	1	20669626~20672534
<i>Brachypodium distachyon</i>	<i>BdBAM3</i>	Bradi1g25447	580	4	1	20674474~20676627
<i>Brachypodium distachyon</i>	<i>BdBAM4</i>	Bradi1g62400	518	1	1	61645584~61647544
<i>Brachypodium distachyon</i>	<i>BdBAM5</i>	Bradi1g75610	573	1	1	72493249~72496562
<i>Brachypodium distachyon</i>	<i>BdBAM6</i>	Bradi2g00824	499	10	2	508715~513118
<i>Brachypodium distachyon</i>	<i>BdBAM7</i>	Bradi2g08187	556	3	2	6579347~6586085
<i>Brachypodium distachyon</i>	<i>BdBAM8</i>	Bradi3g02800	659	9	3	1756675~1762544
<i>Brachypodium distachyon</i>	<i>BdBAM9</i>	Bradi3g28210	534	3	3	29556156~29558022
<i>Brachypodium distachyon</i>	<i>BdBAM10</i>	Bradi3g33730	548	1	3	36109354~36111917
<i>Brachypodium distachyon</i>	<i>BdBAM11</i>	Bradi4g38630	677	8	4	43459903~43464924
<i>Sorghum bicolor</i>	<i>SbBAM1</i>	Sb01g019850	547	0	1	21751511~21753698
<i>Sorghum bicolor</i>	<i>SbBAM2</i>	Sb01g028700	557	3	1	50104773~50107537
<i>Sorghum bicolor</i>	<i>SbBAM3</i>	Sb01g035370	529	1	1	58904172~58906158
<i>Sorghum bicolor</i>	<i>SbBAM4</i>	Sb01g047500	564	2	1	70592880~70595781
<i>Sorghum bicolor</i>	<i>SbBAM5</i>	Sb02g012320	469	8	2	20582813~20592033
<i>Sorghum bicolor</i>	<i>SbBAM6</i>	Sb02g035590	604	6	2	69981839~69984625
<i>Sorghum bicolor</i>	<i>SbBAM7</i>	Sb02g035600	488	6	2	69991535~69994435
<i>Sorghum bicolor</i>	<i>SbBAM8</i>	Sb02g042220	531	1	2	75926839~75928531
<i>Sorghum bicolor</i>	<i>SbBAM9</i>	Sb03g000480	442	3	3	332927~336164
<i>Sorghum bicolor</i>	<i>SbBAM10</i>	Sb04g002450	566	7	4	2235466~2239975
<i>Setaria italica</i>	<i>SiBAM1</i>	Si016592m	655	9	1	8672419~8678661
<i>Setaria italica</i>	<i>SiBAM2</i>	Si028957m	810	9	2	15516601~15521488
<i>Setaria italica</i>	<i>SiBAM3</i>	Si029378m	563	3	2	42485434~42487777
<i>Setaria italica</i>	<i>SiBAM4</i>	Si033152m	531	7	2	42511222~42514352
<i>Setaria italica</i>	<i>SiBAM5</i>	Si032033m	524	1	2	47690524~47692184
<i>Setaria italica</i>	<i>SiBAM6</i>	Si000924m	544	3	4	302114~306028
<i>Setaria italica</i>	<i>SiBAM7</i>	Si035044m	543	2	9	17299082~17301322
<i>Setaria italica</i>	<i>SiBAM8</i>	Si034983m	557	3	9	36577745~36580943
<i>Setaria italica</i>	<i>SiBAM9</i>	Si034947m	563	1	9	46359875~46362007
<i>Setaria italica</i>	<i>SiBAM10</i>	Si034946m	563	2	9	56513552~56516833

从蛋白质长度来看,这54个基因编码的蛋白长度跨度从442~841aa不等,除了高粱*SbBAM1*基因,其余53个基因编码区中均含有内含子,且数目差异较大。

为了阐明禾本科植物*BAM*基因系统进化关系,本研究对供试的5种禾本科植物共54个*BAM*基因编码的蛋白质序列进行了多重序列比对。然后,分别利用ML法和NJ法构建系统发育树。这2个进化树拓扑结构相似,且NJ法具有较高的bootstrap支持,因此本文主要以NJ法构建的系统发育树进行分析(图1)。根据系统发育树,可以将54个禾本科植物*BAM*基因分成10个同源基因簇(COG1-10),每一同源基因簇均具有较高的bootstrap支持率。由于每个同源基因簇都涵盖了这5种禾本科植物,因此推测在禾本科祖先物种中至少含有10个*BAM*基因,并且在禾本科植物分化之后没有发生明显的基因丢失事件。但是有4个基因簇中分别都包含了6个*BAM*基因,说明在禾本科植物从祖先物种分化之后发生了基因重复事件。例如,COG8中包含2个短柄草*BAM*基因,说明在短柄草基因组中发生了一次基因重复事件,且发生在从其他禾本

科植物中分化后。玉米基因组中的*BAM*基因数目比其余4个物种的都多,证明玉米基因组进化过程中也发生了基因重复事件,并且不止一次。

在基因家族的系统进化研究中,外显子/内含子结构分析具有重要意义^[18]。本研究系统分析了5种禾本科植物基因组中共54个*BAM*基因的外显子/内含子结构,结果发现,除*SbBAM1*外,其他*BAM*基因编码区中至少含有1个内含子。并且每一同源簇中基因的外显子/内含子结构高度保守,表明这些*BAM*基因的外显子/内含子基本特征在禾本科植物分化之前就已经形成。例如,COG4中的6个*BAM*基因均含有3个内含子,并且这些内含子的位置高度保守(图1)。另外,在结构分析的过程中还发现存在内含子获得/丢失现象。例如,COG10中绝大多数基因只含有1个内含子,只有玉米*ZmBAM2*基因含有2个。通过比较发现,*ZmBAM2*基因中第一个内含子是通过复制产生的。此外,COG2中大部分基因均含有2个内含子,而*SbBAM1*没有内含子,应该是内含子丢失事件所致。

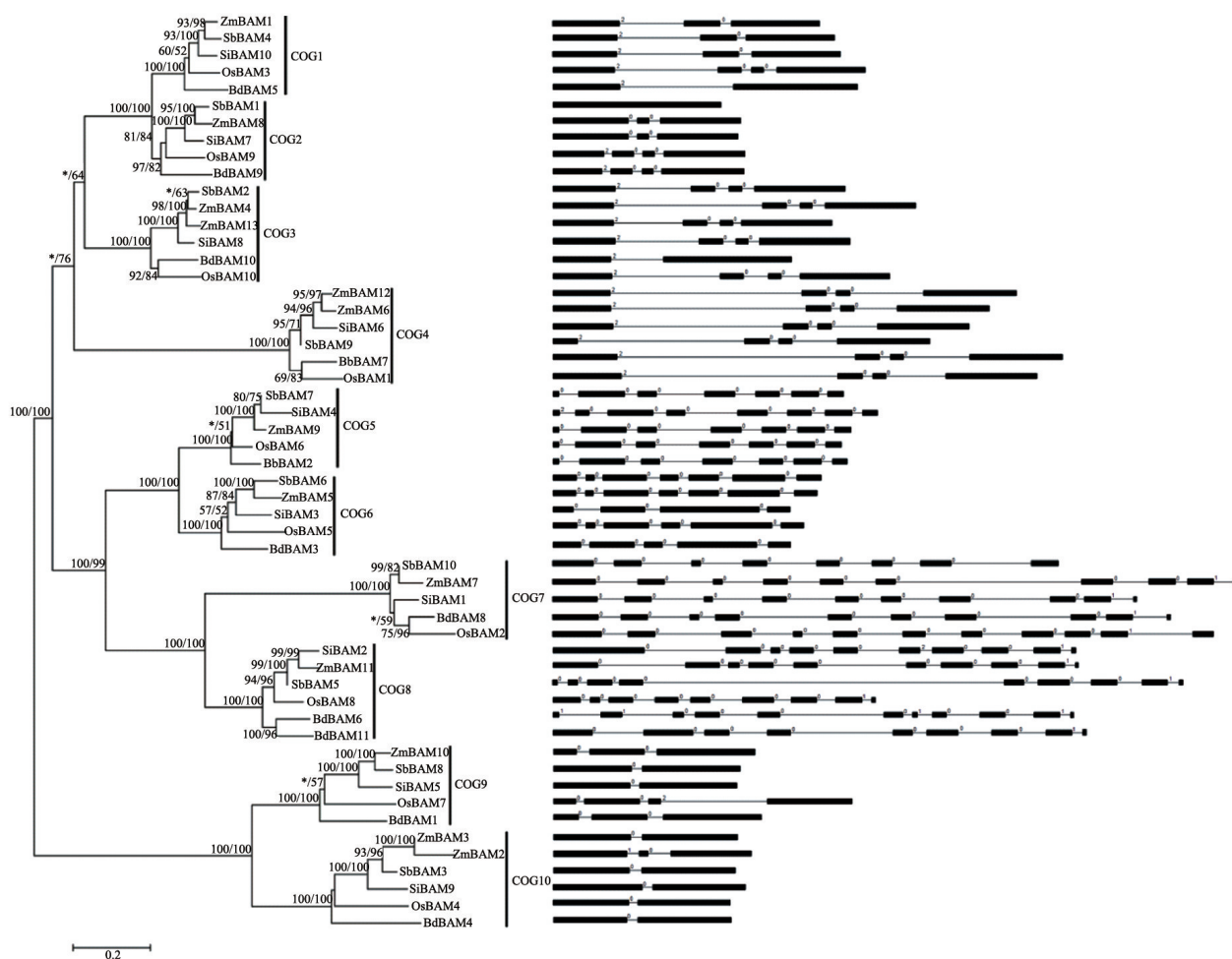


图1 5种禾本科植物*BAM*基因系统发育树及外显子/内含子结构

Fig. 1 Phylogenetic tree of gramineous *BAM* genes and their exon/intron structures

2.2 BAM同源簇之间的功能分化分析

本研究利用 DIVERGE 2.0 软件^[13],对经过比对的氨基酸序列进行功能分化分析。根据构建的系统发育树,将禾本科植物 54 个 BAM 基因划分为 10 个同源基因簇。利用 DIVERGE 2.0 软件,对这 10 个同源簇的功能分化参数(I 型和 II 型)进行极大似然法估计。ML 法可以分别估算不同同源簇之间进化速率以及氨基酸位点的显著差异^[19],这种方法的一个优点是可以利用氨基酸序列估算进化速率,从而降低对同义替换位点的饱和性的敏感性^[18]。本研究分别对 10 个同源簇两两之间进行功能分化比较,因此共进行了 45 次估算比较。对 I 型功能分化系数(θ_i)估算发现,除了 COG1/COG2,COG2/COG3 和 COG3/COG4,其余 42 对同源簇之间 θ_i 都达到显著水平($P < 0.05$)。同时,有 43 对同源簇之间的 θ_{ii} 也达到了显著水平。其中有 41 对的 θ_i 和 θ_{ii} 都达到显著水平,说明同源簇间进化速率和氨基酸特性的特异性差异都对它们的功能性分化起到重要作用。值得一提的是,COG1/COG2 之间 θ_i 和 θ_{ii} 估算

结果都不显著,很可能这 2 个同源簇中的基因在某些方面发挥了相似的功能。

利用功能分化的后验概率(Q_k)来鉴定不同同源簇之间的功能分化位点。 Q_k 值越大,表明两个同源簇之间相同位点存在进化速率和氨基酸特性差异的可能性越大^[19,20]。本研究利用 $Q_k > 0.7$ 和 $Q_k > 0.9$ 分别作为检测同源簇之间 I 型、II 型功能分化的后验概率阈值^[21]。在本研究中检测到的 42 对存在 I 型功能分化的同源簇中,有 39 对能鉴定到 $Q_k > 0.7$ 的位点;在 43 对存在 II 型功能分化的同源簇中,有 39 对能鉴定到 $Q_k > 0.9$ 的位点。例如,COG1/COG7 这 2 个同源簇间有明显的 I 型功能分化,并且检测到 42 个有较高概率的氨基酸位点(图 2)。其中有 38 个位点在 COG1 中保守而在 COG7 中变化较大,其余 4 个位点反之。但是在 77 个与 II 型功能分化显著相关的位点则有不同的表现,在成对比较的两个同源簇中,这些位点在其中一个同源簇中高度保守,在另一个同源簇中则高度分化。

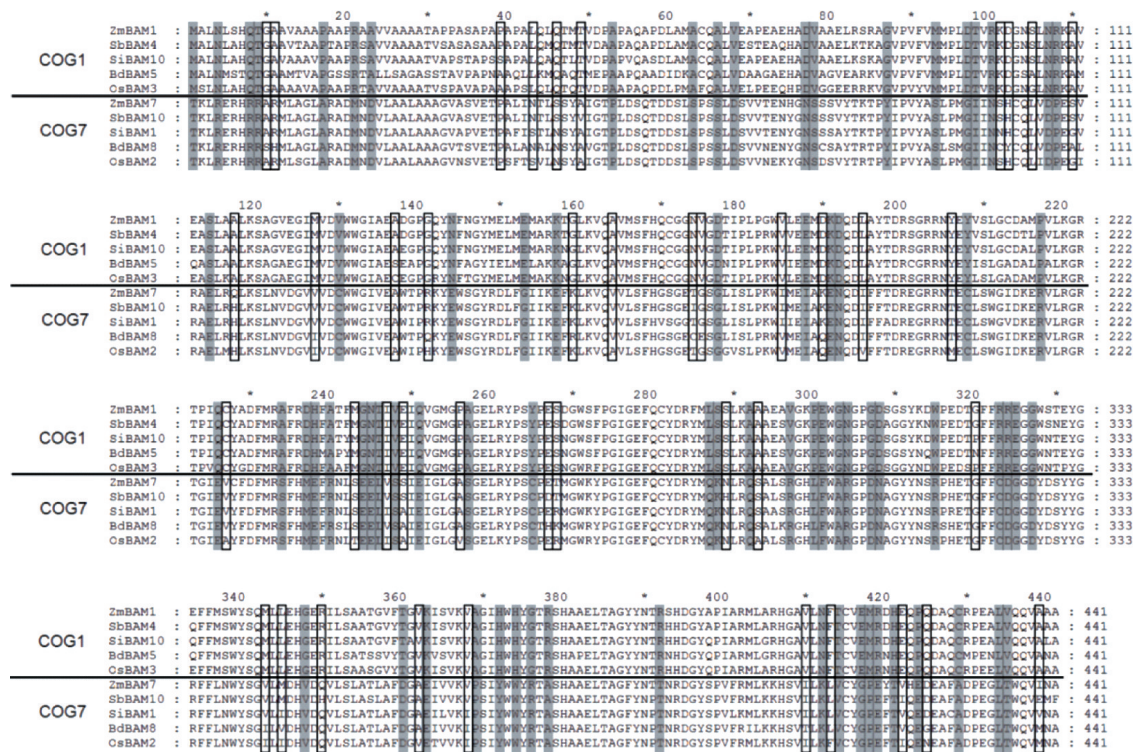


图2 COG1和COG7氨基酸位点对比

Fig. 2 Alignment of BAM proteins in COG1 and COG7

2.3 BAM同源簇之间的适应性进化分析

由于位点特异性模型能有效地估算同源基因的正选择作用,所以本研究选择该模型鉴定经历正选择作用的 BAM 基因以及相应的正选择位点。在位点特异性模型下,每个位点上的 d_n/d_s (ω) 值不同,当 $d_n/d_s > 1$,即 $\omega > 1$ 时,则认为经历了正选择作用^[22]。为了深入研究禾本科植物 BAM 基因经历的正选择作用,分别对每个 COG 都进行了适应性进化分析。

首先通过比较 M0 和 M3 模型估算每个氨基酸位点的 ω 值,从而判断每个同源簇在进化过程中是否存在不同的选择压力(表 2)。结果表明,M0 模型下获得的 ω 平均值为 0.1339,表明纯化选择是禾本科植物 BAM 基因在进化过程中受到的主要选择压力。然而所有同源簇的 M3/M0 似然比检验结果都显著,说明这些同源簇在进化过程中经历的选择压力可能差异性较大。

表2 “位点”特异性模型下5种禾本科植物BAM同源基因簇的正选择检测

Table 2 Detection of positive selection under site-specific models for each BAM COG in grasses

同源簇	M0模型 ω 值	$2\Delta l$ (M3/M0)	$2\Delta l$ (M8/M7)	$2\Delta l$ (M8/M8a)	M8模型下参数估计值	正选择位点
COG1	0.0747	146.0849**	1.8371	0.1605	$\beta(0.2377, 2.5970)$, $P_1 = 0.0143, \omega = 1.6720$	NAN
COG2	0.0930	171.2139**	0.7677	0.0016	$\beta(0.2494, 2.3429)$, $P_1 = 0.0371, \omega = 1.0000$	NAN
COG3	0.0966	154.7483**	4.4181	4.1010*	$\beta(0.2544, 1.9785)$, $P_1 = 0.0047, \omega = 13.5839$	21K, 46A, 49S, 53A, 54T, 60E, 69Q, 70Q, 174P, 327H, 388D
COG4	0.1480	25.1146**	1.1673	0.0019	$\beta(1.0438, 7.2717)$, $P_1 = 0.0388, \omega = 1.0000$	NAN
COG5	0.1905	110.1256**	3.5924	2.6080	$\beta(0.1528, 0.6507)$, $P_1 = 0.0182, \omega = 2.9248$	NAN
COG6	0.1257	101.8675**	7.9661*	7.3297**	$\beta(0.3731, 2.1066)$, $P_1 = 0.0031, \omega = 394.8958$	16A, 17T, 31P, 34F**, 275M, 358G, 546V
COG7	0.2014	129.7076**	6.1525*	3.6122*	$\beta(0.2631, 1.2559)$, $P_1 = 0.0393, \omega = 2.2291$	2K, 6Q, 7R, 8D*, 10E, 11P, 12S, 18Q, 24F, 27T, 31S, 105N, 114L, 115H, 117V, 191M, 193S, 293I, 297L, 560E
COG8	0.1271	134.3305**	5.0598	4.1233*	$\beta(0.1220, 0.8892)$, $P_1 = 0.0198, \omega = 2.9347$	3P, 4V*, 5A, 6E, 8A*, 9A, 13P, 114V, 221N, 224R, 246M, 265L, 394S, 395K
COG9	0.0871	217.9859**	23.6633**	16.8038**	$\beta(0.3497, 3.0064)$, $P_1 = 0.0159, \omega = 9.5580$	8S, 14Q, 18A, 9P*, 21G, 22R, 30P, 505N**, 506T, 507R
COG10	0.1951	172.5371**	0.7926	5.0585*	$\beta(0.4022, 1.7130)$, $P_1 = 0.0731, \omega = 1.1122$	24R, 37V, 44S, 68T, 223Q, 232Q, 448A, 463P

注: *表示在0.05水平上的差异显著性, **表示在0.01水平上的差异显著性; ☆表示后验概率大于95%, ☆☆表示后验概率大于99%。

通过M8/M7、M8/M8a模型的似然比检验来确定禾本科植物每个BAM同源簇中是否有正选择位点。检测结果发现,在10个同源簇中有6个在M8/M8a的似然比检验下检测到受到正选择作用影响;M8/M7下只检测到3个(COG6, COG7和COG9)。另外,将所有在M8/M7下的检测结果均在M8/M8a模型下检出。这些结果说明正选择作用对禾本科植物BAM基因中至少3个同源簇的进化起到重要作用。为了鉴定出这些同源簇中经历正选择作用的位点,在M8模型下做了贝叶斯^[23]后验概率估计。通过估算发现,这3个同源簇中分别有7,20和10个后验概率大于50%的氨基酸位点,这一结果表明正选择作用对这3个同源簇(COG6, COG7和COG9)的功能性分化具有重要作用。

2.4 非生物胁迫下水稻BAM基因的表达谱分析

非生物胁迫对于植物正常生长、发育以及结实过程具有极大影响作用^[24]。为阐明BAM基因在植物逆境过程中的作用,进一步分析了水稻BAM基因在不同非生物胁迫下的表达特征。结果表明,水稻BAM基因的表达模式受多种非

生物胁迫的调控,如干旱、高温、低温以及盐胁迫等(图3)。微阵列数据显示,在不同胁迫处理下,水稻BAM基因表达模式差异较大。例如,在4种胁迫处理下,OsBAM4和OsBAM10基因表达都明显上调;OsBAM3基因在低温处理下无表达上调趋势;而OsBAM9基因仅在干旱和盐处理下表达量明显增加。同时,本研究设定2倍差异作为基因表达的标准,发现没有基因的在逆境条件下表达下调。本研究都表明禾本科植物BAM基因家族可能参与了多种非生物胁迫的响应过程。

已有研究证明,很多BAM基因之间的表达模式差异较大。如大麦Bmy1基因只在胚乳中表达,而其同源基因Bmy2的表达并不表现组织特异性^[25]。为了更好地了解禾本科植物BAM基因的功能,本研究组调查了水稻9种组织中BAM基因的表达模式,发现部分水稻BAM基因呈现明显的组织特异性。OsBAM3、OsBAM6、OsBAM7、OsBAM8和OsBAM9基因主要在花药中表达(图3),这些基因可能在水稻生殖生长过程中发挥了重要作用。

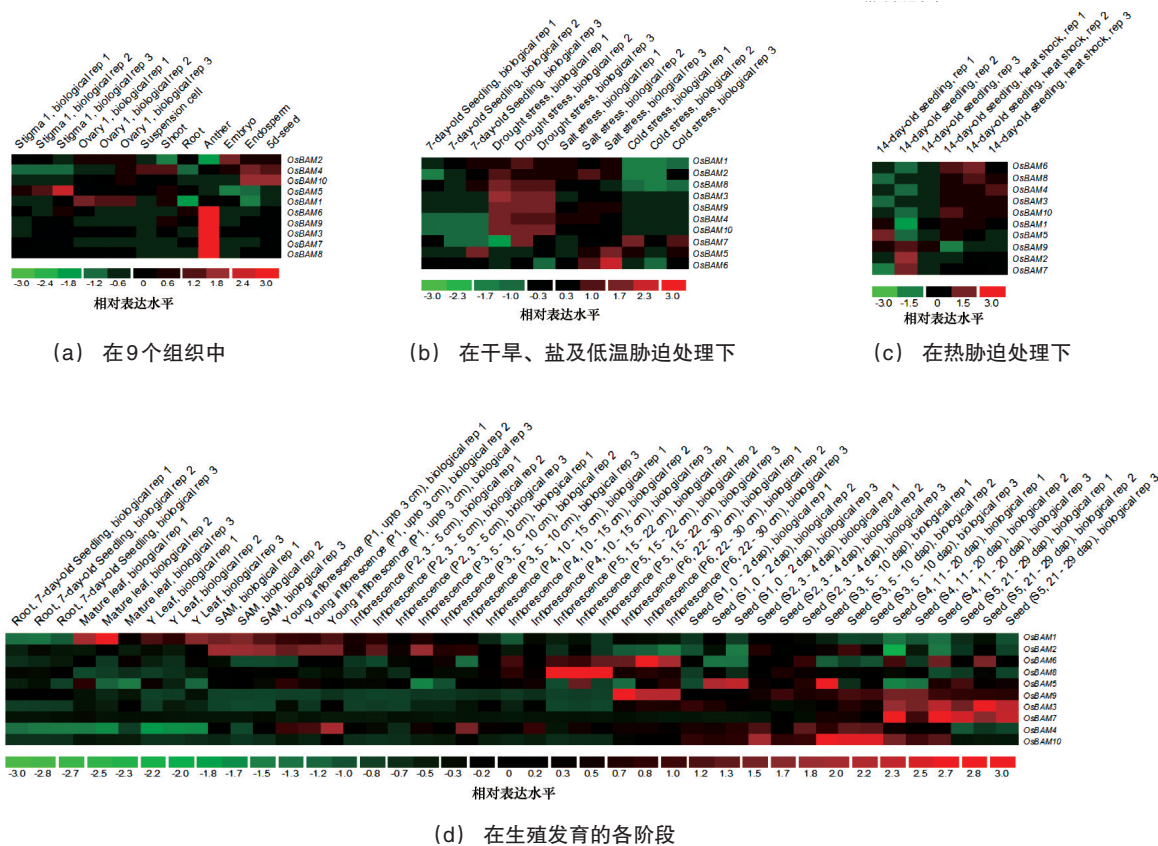


图3 水稻 BAM 基因表达谱分析

Fig. 3 Gene-expression patterns of rice BAM genes

3 讨论

基因家族(gene family)是由一个祖先基因经过重复和变异产生的,其基因序列尤其是编码的蛋白质序列具有明显的相似性。基因家族的成员分布主要有两种形式,一种是形成同源簇,如玉米中22-kD醇溶蛋白基因家族的大部分成员成簇分布在同一条染色体的168kb区域内^[20];另一类是基因家族成员成簇地分布在不同的染色体上,如珠蛋白基因家族。在不同物种中,BAM基因数目的不同主要是由BAM基因家族的物种特异性扩张所致,而物种特异性扩张的主要形成方式是基因重复。为了进一步分析鉴定出的禾本科植物BAM基因家族成员的基因重复模式,本研究组分别查找了这些旁系同源基因的物理位置,若位于染色体的同一区段或者相邻区段则起源于串联重复;接着通过查找一对旁系同源基因的上、下游是否存在其他旁系同源基因来判断是否起源于片段复制。据此标准发现,COG10中ZmBAM2/ZmBAM3这对玉米旁系同源基因起源于一次串联重复事件,因为它们位于同一染色体区域。而另外两对,玉米ZmBAM4/ZmBAM13基因与短柄草BdBAM6/BdBAM11基因则起源于片段复制,因为在这两对同源基因的两侧都发现了其他高度保守的同源基因。

基因重复为新功能的产生提供了丰富的遗传物质基础^[27]。通过遗传和变异,可以产生具有“新功能化”或者“亚

功能化”的基因,使物种获得新的功能从而更好地适应生存环境。例如,本研究通过序列比对以及构建系统发育树检测到禾本科植物基因组中存在多次基因重复事件,功能分化分析发现不同同源基因簇间存在明显的进化速率的差异,同时表达分析也证实了水稻BAM基因的表达模式具有多样性。尤其当对水稻BAM基因在生殖发育各阶段的表达模式进行研究发现,大部分基因主要在开花到种子阶段表达,因此水稻BAM基因的表达很可能与生长发育过程有密切关系。

4 结论

在植物中, β -淀粉酶能将直链淀粉分解成麦芽糖,在植物萌发以及生长发育过程中起到重要作用。本研究基于生物信息学方法,对 β -淀粉酶在禾本科植物中的进化模式及水稻BAM基因的非生物胁迫表达模式进行分析。研究结果表明,在供试5个禾本科植物基因组中分别含有10~13个BAM基因,并形成了10个同源基因簇,表明禾本科植物祖先物种含有10个BAM基因。在供试基因组中,该基因家族没有明显的基因丢失现象,但在玉米和二穗短柄草基因组中分别有3次和1次基因重复事件。同源簇之间存在明显的进化速率的差异,可能有助于产生同源簇特有的功能。适应性进化分

析表明,至少有3个BAM基因粗受到正选择作用的影响。基于微阵列数据的表达分析表明,水稻部分BAM基因表现出组织特异性表达模式,并受到干旱和盐害等非生物胁迫的诱导。

参考文献(References)

- [1] Zhang D, Wang Y. Beta-amylase in developing apple fruits: Activities, amounts and subcellular localization[J]. Science in China Series C: Life Sciences, 2002, 45(4): 429-440.
- [2] Rejcek M, Stevenson C E, Southard A M, et al. Chemical genetics and cereal starch metabolism: Structural basis of the non-covalent and covalent inhibition of barley beta-amylase[J]. Molecular BioSystems, 2011, 7(3): 718-730.
- [3] Vinje M A, Willis D K, Duke S H, et al. Differential expression of two beta-amylase genes (*Bmy1* and *Bmy2*) in developing and mature barley grain[J]. Planta, 2011, 233(5): 1001-1010.
- [4] 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.
- [5] 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.
- [6] Reinhold H, Soyk S, Simkova K, et al. Beta-amylase-like proteins function as transcription factors in Arabidopsis, controlling shoot growth and development[J]. Plant Cell, 2011, 23(4): 1391-1403.
- [7] Mason-Gamer R J. The β -amylase genes of grasses and a phylogenetic analysis of the Triticeae (Poaceae)[J]. American Journal of Botany, 2005, 92(6): 1045-1058.
- [8] Goodstein D M, Shu S, Howson R, et al. Phytozome: A comparative platform for green plant genomics[J]. Nucleic Acids Research, 2012, 40:D1178-D1186.
- [9] Robert D, Finn, John Tate, Jaina Mistry, et al. The Pfam protein families database[J]. Nucleic Acids Research, 2008, 36(1): D281-D288.
- [10] Larkin M A, Blackshields G, Brown N P, et al. Clustal W and Clustal X version 2.0[J]. Bioinformatics, 2007, 23(21): 2947-2948.
- [11] Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. Molecular Biology and Evolution, 2011, 28(10): 2731-2739.
- [12] Guindon S, Delsuc F, Dufayard JF, et al. Estimating maximum likelihood phylogenies with PhyML[J]. Methods in Molecular Biology, 2009, 537(1): 113-137.
- [13] Gu X, Velden K V. DIVERGE: Phylogeny-based analysis for functional-structural divergence of a protein family[J]. Bioinformatics, 2002, 18(3): 500-501.
- [14] Suyama M, Torrents D, Bork P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments [J]. Nucleic Acids Research, 2006, 34: W609-W612.
- [15] Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood[J]. Molecular Biology and Evolution, 2007, 24(8): 1586-1591.
- [16] Yang Z, Bielawski J P. Statistical methods for detecting molecular adaptation[J]. Trends in Ecology & Evolution, 2000, 15(12): 496-503.
- [17] Tusher V G, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response[J]. Proceedings of the National Academy of Sciences, 2001, 98(9): 5116-5121.
- [18] Yang Z, Wang Y, Xu S, et al. Molecular evolution and functional divergence of soluble starch synthase genes in cassava (*Manihot esculenta crantz*)[J]. Evolutionary Bioinformatics Online, 2013, 9: 239-249.
- [19] Gu X. Functional divergence in protein (family) sequence evolution[J]. Genetica, 2003, 118(2/3): 133-141.
- [20] Liu Q, Wang H, Zhang Z, et al. Divergence in function and expression of the NOD26-like intrinsic proteins in plants[J]. BMC Genomics, 2009, 10: 313.
- [21] Suárez-Castillo E C, García-Arrarás J E. Molecular evolution of the ependymin protein family: A necessary update[J]. BMC Evolutionary Biology, 2007, 7: 23.
- [22] Yang Z, Nielsen R, Goldman N, et al. Codon-substitution models for heterogeneous selection pressure at amino acid sites[J]. Genetics, 2000, 155(1): 431-449.
- [23] Yang Z, Wong W, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection[J]. Molecular Biology and Evolution, 2005, 22(4): 1107-1118.
- [24] Singh A, Pandey A, Baranwal V, et al. Comprehensive expression analysis of rice phospholipase D gene family during abiotic stresses and development[J]. Plant Signal & Behavior, 2012, 7(7): 847-855.
- [25] Vinje M A, Willis D K, Duke SH, et al. Differential expression of two β -amylase genes (*Bmy1* and *Bmy2*) in developing and mature barley grain[J]. Planta, 2011, 233(5): 1001-1010.
- [26] Song R, Llaca V, Linton E, et al. Sequence, regulation, and evolution of the maize 22-kD alpha zein gene family[J]. Genome Research, 2001, 11(11): 1817-1825.
- [27] Lynch M, O'Hely M, Walsh B, et al. The probability of preservation of a newly arisen gene duplicate[J]. Genetics, 2001, 159(4): 1789-1804.

(责任编辑 吴晓丽)



SCIENCE & TECHNOLOGY REVIEW

《科技导报》“卷首语”栏目征稿

“卷首语”栏目每期邀请一位中国科学院院士和中国工程院院士就重大科技现象、事件,以及学科发展趋势、科学研究热点和前沿问题等,撰文发表个人的见解、意见和评论。本栏目欢迎院士投稿,每篇文章约2000字,同时请提供作者学术简历、工作照和签名电子文档。投稿邮箱:kjdbbjb@cast.org.cn。