

天冬氨酸蛋白酶在拟南芥和水稻中的分子进化、表达模式以及在花药发育中的功能分析

汤龙军^{1,2}, 李鹏程², 朱璐², 王瑞丰², 陈旭², 李宗芸¹, 许杰^{2,*}

¹江苏师范大学生命科学学院, 江苏徐州221000; ²上海交通大学生命科学技术学院, 上海200240

摘要: 天冬氨酸蛋白酶(aspartic proteases, APs)是一类功能多样化的酶, 参与了植物大部分生理和发育过程, 但其基因家族在植物中的系统分类、分子进化以及功能方面研究仍很欠缺。本研究收集整理了拟南芥和水稻中APs基因序列、分类、基因结构和染色体定位等信息, 对比分析该基因家族在拟南芥和水稻中分子进化方面的差异; 利用转录组数据阐明该家族基因在不同组织部位和不同发育时期的表达差异和功能分化。此外, 本文还围绕花药发育的过程, 细致分析了各亚家族APs基因的表达变化, 推测APs在花药发育过程中的可能功能。

关键词: 天冬氨酸蛋白酶; 分子进化; 功能分化; 花药

Identification, Evolutionary and Expression Profile Analysis of the Aspartic Protease Gene Superfamily in *Arabidopsis thaliana* and Rice

TANG Long-Jun^{1,2}, LI Peng-Cheng², ZHU Lu², WANG Rui-Feng², CHEN Xu², LI Zong-Yun¹, XU Jie^{2,*}

¹School of Life Sciences, Jiangsu Normal University, Xuzhou, Jiangsu 221000, China; ²School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: Aspartic proteases (APs), a large family of proteolytic enzymes with diverse functions, play important roles in many aspects of physiological and developmental processes in plants. However, systematic studies on their classification, molecular evolution and function are still lacking. In this study, information about APs including gene sequence, subfamily classification, gene structure, sub-cellular localization and chromosomal location of all known APs in *Arabidopsis* and rice were collected, and analyzed by bioinformatics tools. In addition, the difference in spatio-temporal expression and function of these APs between *Arabidopsis* and rice was investigated using available transcriptomic data. Furthermore, the expression pattern of each of the APs genes along anther development was highlighted, which provides a framework for future studies defining the roles of APs in anther development.

Key words: aspartic protease; molecular evolution; functional differentiation; anther

天冬氨酸蛋白酶(aspartic proteinase, AP)是一类重要的蛋白水解酶, 其活性中心由2个催化性天冬氨酸残基组成(Davies 1990; Barrett 1992; Rawlings和Barrett 1999; 吕刚等2008), 主要参与机体的新陈代谢并起到生物调控作用。AP普遍存在于脊椎动物、植物、酵母、真菌以及病毒之中(Kamitori等2003; Faro和Gal 2005; Davies 1990; Chen等2009), 说明其在整个生物进化过程中具有极其重要的功能且非常保守。相比较而言, AP在动物和微生物中的功能研究和认识较深入, 而在植物中的研究相对滞后。尽管目前已有将近300个AP基因在植物中获得了克隆或序列信息, 也有研究证明植物AP在应激反应、衰老、细胞程序性死亡和

有性繁殖过程中都担任重要角色(Simoes和Faro 2004), 但其生化和遗传学功能方面的研究内容仍然较少。因此, 植物中AP基因的功能研究有待于进一步深化。

植物中的APs一般可以分为三大类: 典型的AP、类nucellin蛋白的AP以及非典型AP (Faro和Gal 2005)。典型的AP前体的特点是拥有一个大小为50~100个氨基酸分子的特殊蛋白域plant specific

收稿 2015-01-04 修定 2015-01-30

资助 国家自然科学基金面上项目(31000593、31370026和31271698)。

* 通讯作者(E-mail: xujie3000@sjtu.edu.cn; Tel: 021-34204869)。

insert (PSI), 这个片段插入到典型天冬氨酸蛋白前体的C端, 在成熟的过程中通常会被切割掉。虽然PSI区域的序列、结构以及功能均与动物以及微生物的AP没有同源性, 但它却和鞘脂激活蛋白样蛋白(saposin-like proteins)十分类似(Mutlu和Gal 1999)。类nucellin APs编码的蛋白与nucellin非常相似, 该类AP首次发现于大麦的珠心细胞中(Chen和Foolad 1997)。非典型的AP则具有介于前面两大类AP之间的序列特征(Faro和Gal 2005)。

AP蛋白作为一类重要的蛋白水解酶(EC3.4.23), 其生物学功能具有多样性, 其在不同植物发育过程和部位行使的功能也各有不同(陈荣和汲广东2011)。其中AP在细胞程序化死亡方面的功能是近期研究的主要关注点。众所周知, 动物细胞程序性死亡的开关是由一类Caspases (cysteine-aspartic acid proteases, 天冬氨酸特异性半胱氨酸蛋白酶)家族调控的, 而在植物体内, 大量的基因组测序和一级结构上的同源分析都没有发现动物Caspases的直系同源蛋白(Francischetti等2010), 然而, 植物的许多细胞程序性死亡过程中的确又能检测到Caspase酶(或类Caspase酶)活性(Hensel等1993; Chen等1993; Xu和Chye 1999; Abramovitch等2003; Solomon等1999; 丁宁等2008)。因此, 我们有理由相信植物体内存在类Caspase酶(比如与Caspase酶结构类似的Metacaspase), 或是由不同的半胱氨酸蛋白酶和天冬氨酸蛋白酶(比如APs)共同来行使某些Caspase蛋白酶活, 从而参与植物的程序性死亡过程。比如, 拟南芥中UNDEAD蛋白(属于A1类天冬氨酸蛋白), 在遗传学上, 通过RNAi方法降低UNDEAD基因的表达量, 就会引起绒毡层细胞的过早成熟从而提前发生细胞程序性死亡(Phan等2011)。此外, 以植物花药发育过程为例, 已有的研究报道和我们的研究数据都显示, 花药绒毡层的降解过程可能是一个典型的程序性细胞死亡过程, 在油菜(*Brassica napus*)和烟草(*Nicotiana tabacum*)花药发育中已经检测出Caspase-3酶活, 随着绒毡层的发育过程发生改变(Solis等2014)。我们实验室先前的研究分别报道了水稻花药转录因子TAPETUM DEGENERATION RETARDATION (TDR)和ETERNAL TAPETUM 1 (EATI)在花药绒毡层降解和发育调控过程中发挥重要的作用, TDR是

通过直接调控一个半胱氨酸蛋白酶(cysteine protease 1, OsCP1), 而EATI是通过直接调控2个天冬氨酸蛋白酶aspartic protease 25和37 (OsAP25和OsAP37)来调控绒毡层细胞死亡的功能(Li等2006; Niu等2013; 张虹等2008; 李晶和郑彩霞2012)。由此可见, 至少部分AP基因在花药绒毡层细胞死亡中可能起到了不可或缺的作用, 因此, 细致分析各个家族AP基因在花药绒毡层细胞死亡过程中的表达变化, 可能会增加人们对AP新功能的认识。

因此, 为全面了解拟南芥和水稻中APs的信息和功能, 本文采用生物信息学手段, 对比分析了拟南芥和水稻中AP家族蛋白的基因信息、蛋白定位预测、聚类分析、染色体定位, 并且对各类亚家族基因在不同组织器官的表达模式进行了分析。此外, 本文还重点分析了AP基因可能参与花药发育绒毡层细胞程序性死亡的功能研究, 旨在为拟南芥和水稻中AP家族的功能研究提供有意义的参考。

材料与方法

1 拟南芥和水稻中AP家族数据收集汇总

拟南芥数据的获得主要通过对信息数据资源库(TAIR, <http://www.arabidopsis.org/>)搜索; 水稻数据通过对水稻基因注释计划(Rice Gene Annotation Project, <http://rice.plantbiology.msu.edu>)和NCBI-protein blast两个数据库中搜索带有AP结构域的所有蛋白, 查找除去多余序列以及通过使用SMART、NCBI-CD和INTEPRO等基序搜索方法, 分别获得了拟南芥和水稻中AP家族的基本信息, 包含了基因信息、亚家族的分类、基因别名、蛋白的大小以及内含子数目信息。同时根据基因DNA序列和蛋白结构信息, 使用ATTED-II (<http://atted.jp/>)得到拟南芥AP蛋白的亚细胞定位预测的结果; 水稻中AP蛋白的亚细胞定位的预测是通过PSORT、TargetP 1.1、MITOPROT和ChloroP 1.1 (<http://cubic.bioc.columbia.edu/cgi/-var/nair/resonline.pl>)来完成的。

2 GO功能分类

Gene Ontology (GO)是基于数据库和网络工具的分析方法(Go Analysis Toolkit and Database for Agricultural community, <http://bioinfo.cau.edu.cn/agriGO/index.php>), 其中包括了分子功能(molecular

function)、生物过程(biological process)以及细胞组分(cellular component)。

3 AP家族在染色体分布和基因进化复制分析

采用TAIR网站的染色体画图工具(chromosome map tool, <http://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp>)绘制染色体图, 将预测的所有拟南芥天冬氨酸蛋白基因定位到染色体上; 在TIGR (<http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml>)上利用BLASTP将预测的水稻AP基因放置到水稻染色体上, 分析不同家族成员之间的相对位置关系, 并对基因家族连锁信息进行分析。

4 AP家族蛋白结构分析

所有拟南芥和水稻的AP基因的信号序列的长度都是通过SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>)和Signal Peptide Prediction (http://bioinformatics.leeds.ac.uk/prot_analysis/Signal.html)来预测的。活性位点的扫描是通过使用PROSITE得到的。AP结构域外的其他结构域是通过SMART、Pfam和Interproscan数据库来查找到的。内含子的数目是通过TIGR (http://www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice/)上的基因组浏览工具完成。

5 AP基因表达谱数据收集和聚类分析

对所有获得的拟南芥和水稻的AP基因, 使用公共资源信息数据库CREP Database (<http://crep.ncpgr.cn>)、水稻核苷酸芯片数据Rice Oligonucleotide Array Database (<http://www.ricearray.org/>)和在Bio-Array Resource for Plant Biology (BAR)中的Arabidopsis eFP Browser (<http://bar.utoronto.ca/welcome.htm>)网页工具, 查找和获得拟南芥和水稻的发育芯片表达数据, 包含拟南芥35个及水稻27个不同的生长发育阶段和组织器官, 对于多个探针号对应的基因取平均值作为表达值。使用微阵列数据分析Genesis (1.7.5)软件进行分析, 采用层次聚类的分析方法对这些基因表达数值进行处理。

6 花药中拟南芥和水稻AP基因表达信息分析和功能预测

在ROAD (www.ricearray.org)上找到所有AP家族基因在不同花药发育时期的表达值, 再利用程序Genesis制作出方便阅读的表达谱图, 再进行聚类分析。

实验结果

1 拟南芥和水稻中AP基因数据汇总和家族分类信息

我们分别在拟南芥和水稻中获得69个和89个AP基因, 以及它们的基本信息, 包括基因家族、蛋白编号、类型、别名、蛋白大小、内含子数量, 结合在线工具ATTED-II得到的蛋白亚细胞结构预测结果和蛋白定位信息(表1)。结果表明, 在拟南芥的基因组中, 69个AP基因家族成员可以分成四大类: 典型的AP类(A类, 3个基因), 类nucellin AP类(B类, 4个基因), 非典型的AP类(C类, 42个基因), 以及其他种类(Others, 20个基因)(表1)。在A类, 典型的AP类成员中, 根据他们是否含有PSI, 可以将其分为A1 (1个)和A2 (2个), 其中A1 (At1g11910)含有典型的PSI结构域, 而A2 (At4g22050和At1g69100)却不含有该结构域。在B类, 类nucellin AP类中, 共有4个基因(At1g77480、At1g44130、At4g33490和At1g49050)编码这类与nucellin十分类似的蛋白酶, 它们都不含有前导肽和PSI, 而且有趣的是nucellin本身仅仅是在同源性上和此类蛋白酶非常接近但却不具有天冬氨酸酶催化活性。C类, 非典型性的AP类是数量最大的一类, 在这里又分为6个小群组(C1、C2、C3、C4、C5、CX)。在C1中有10个基因, 它们编码的蛋白含有一个信号肽、一段25~30个氨基酸的前导肽, 而且预测这些蛋白都位于分泌途径中; C2有4个基因, 它们编码的蛋白含有丝氨酸富集的转运肽段, 但是缺少前导肽, 所以可能依靠pH或者内生抑制剂来控制其蛋白水解酶活性; C3有8个基因, 而且这一类的特点是跟CND41这个蛋白非常相似并且拥有一段较长的前导肽, 这个蛋白是在烟草的叶绿体中发现的(Nakano等1997); C4中共有10个基因, 它们的共同点是拥有一个C末端延长区段; C5类有3个基因, 含有的保守活性位点是DTS (D-天冬氨酸, T-苏氨酸, S-丝氨酸)而不是DTG (D-天冬氨酸, T-苏氨酸, G-甘氨酸), 其他都与C1类很相似, 所有无法归类的非典型AP基因都被划分到CX类(表1)。

水稻中的89个AP家族基因分类与拟南芥中的分类相似, 也分为四大类: 典型的AP类(A类, 6个基因), 类nucellin AP类(B类, 13个基因), 非典型的AP类(C类, 68个基因)以及其他种类(X类, 2个基因)

表1 拟南芥和水稻天冬氨酸家族基因基本信息

Table 1 The basic information of aspartyl protease family in *Arabidopsis* and rice

物种	亚家族	类型	蛋白大小/aa	内含子数	蛋白定位	蛋白数目/个	特点
拟南芥	A1	Typical	506	12	S, C	1	具有PSI结构域
	A2	Typical	354, 367	4	S, Y	2	无PSI结构域
	B	Nucellin-like	401~583	6, 7, 8	S, C, O, M, V	4	不含有前导肽和PSI结构域
	C1	Atypical	396~512	0, 1, 2, 9, 10	S, C, P, O, N	10	含一个信号肽, 前导肽
	C2	Atypical	392~491	0	C, C	4	含丰富丝氨酸的转运肽段
	C3	Atypical	464~535	0, 1, 2	S, C, V	8	结构与CND41蛋白相似
	C4	Atypical	481~632	9, 11	S, P, V, C	10	具有一个C末端延长区段
	C5	Atypical	425~455	1, 2	S, V, O, E, C	3	含有的保守活性位点是DTS不是DTG
	CX	未知	395~528	0, 1, 8, 9	S, P, C, V	7	未知
	其他	未知	386~630 (158, 756)	0, 1, 3, 5, 9, 11	O, N, S, C, V, P, Y	20	未知
水稻	A1	Typical	496~552 (52)	10, 12	Ex	5	具有PSI结构域
	A2	Typical	未知	未知	V	1	无PSI结构域
	B	Nucellin-like	410~632	7, 8, 9, 11	C, E, Ex, P	13	不含有前导肽和PSI
	C	Atypical	316~525	0, 1	C, E, Ex, M, N, P	42	未知
	C1	Atypical	463~519 (720)	1, 2, 3	C, Cy, E, P	9	CND41蛋白的同源
	C2	Atypical	372~472	0, 1	Cy, Ex, Pe	3	结构与S5相似
	C3	Atypical	441, 447	0	Ex	2	PCS1蛋白的同源
	C4	Atypical	454	0	Ex	1	CDR1蛋白的同源
	C5	Atypical	373~419	0	Cy, E, Ex, P	11	DTG基序中第一个天冬氨酸高度保守
	X	未知	未知	未知	C, P	2	未知

“类型”一列中, Typical代表典型的天冬氨酸蛋白酶类; Nucellin-like代表珠心细胞的类天冬氨酸蛋白酶类; Atypical代表非典型的天冬氨酸蛋白酶类。“蛋白定位”一列中, C代表叶绿体; Cy代表细胞质; E代表内质网; Ex代表胞外基质; M代表线粒体; N代表细胞核; P代表质膜蛋白; Pe代表过氧化物酶体; S代表分泌蛋白; V代表液泡; O代表未知。“特点”一列中, DTS代表一个氨基酸序列, D-天冬氨酸, T-苏氨酸, S-丝氨酸; DTG代表一个氨基酸序列, D-天冬氨酸, T-苏氨酸, G-甘氨酸; PSI代表植物特异性插入(plant specific insert); CND41代表41 kDa DNA结合蛋白; PCS1代表PROMOTION OF CELL SURVIVAL 1; CDR1代表CONSTITUTIVE DISEASE RESISTANCE 1。

(表1)。C类中包括非典型的AP, 是最大的一组, 有68个成员。在这一组的一些成员根据APs已经报道过的或者预测的结构被更细分为6个组(C、C1、C2、C3、C4和C5), 9个在C1, 3个在C2, 2个在C3, 1个在C4, 11个在C5, 其他依然留在C大类中。C1组是CND41蛋白的同源, 是一个在叶绿体拟核中发现的烟草蛋白, 具有DNA结合以及蛋白水解酶活性(Nakano等1997)。S5被分入C2, 这个蛋白被报道是一个通过调节胚囊育性来调控粳稻-籼稻杂交不育性的蛋白(Chen等2008)。C3组是PCS1蛋白的同源, 一个拟南芥AP, 在胚胎发育和繁殖过程中对细胞命运的决定起重要的作用(Ge等2005)。C4组是CDR1蛋白明显的同源, 一个参与抗病性信号的拟南芥蛋白(Xia等2004)。C5中的APs在DTG基序中第一个天冬氨酸高度保守, 然而这个基序中的另一个残基保守性较弱。最后是X类包括2个无

法被分入任何一类的基因, AK107644和AK111185 (Chen等2009)。

2 拟南芥和水稻中AP基因基本结构、蛋白结构和亚细胞定位分析

从基因结构分析, 结果表明AP不同亚家族基因的内含子数目存在差异, 拟南芥中A类AP基因的内含子数目有4和12个两类; B类中内含子数目都在6~8个之间; C1、C2、C3类中, 除了At2g36670.1/2的内含子数目是9或10个, 其他基因内含子数目都小于等于2个, 而且大多数都是0个; C4的内含子数目较多且都是9或者11个; C5类基因内含子都是1或2个。对比水稻中: A类AP类基因的内含子数目有10和12个两类; B类中内含子数目都在7~11个之间; C类基因的内含子数目有0和1个两类; C1类多为1个、2个和3个, C2类中也仅有0和1个两类; C3、C4、C5类基因的内含子数目都为0个。

所以, 结果表明, 不管在拟南芥中还是水稻中, 非典型天冬氨酸家族基因的内含子数目比其他两类要少(除去拟南芥中的C4家族)(表1)。

我们通过对拟南芥和水稻APs的结构分析, 表明: 典型的AP都具有这样的基本结构, 即一个信号肽, 一个前肽, 一个含有两个活性位点的AP结构域(Barrett等2004)。我们通过对拟南芥AP的结构分

析, 发现有19个AP蛋白(27.54%)含有基本的结构, 包括1个信号肽和2个活性位点。47个蛋白(68.11%)缺少信号肽或者活性位点之一, 1个(1.45%)有缺少信号序列和活性位点的AP结构域。2个(2.90%)没有找到相关信息。只有3个AP鉴定出是典型的AP蛋白, 其中的1个含有两个Saq_B结构域, 位于植物特异性插入序列(PSI)(图1-A)。水

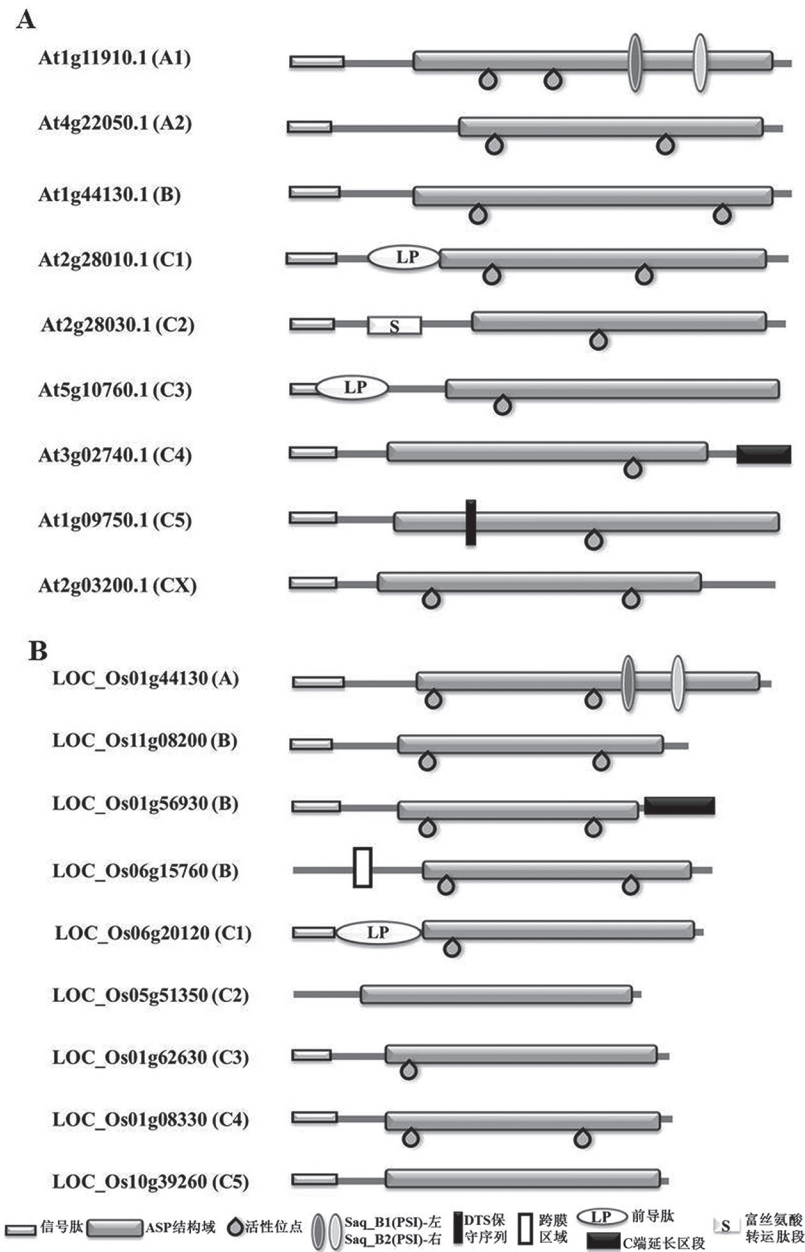


图1 拟南芥和水稻天冬氨酸蛋白酶家族基因的蛋白结构
Fig.1 Protein structure of AP proteins in *Arabidopsis* and rice

A: 拟南芥天冬氨酸蛋白酶家族基因的蛋白结构; B: 水稻天冬氨酸蛋白酶家族基因的蛋白结构。DTS代表一个氨基酸序列, D-天冬氨酸, T-苏氨酸, S-丝氨酸; S代表丝氨酸; LP代表前导肽; Saq_B1和Saq_B2表示植物特异性插入中的特殊结构域。

稻中有24个AP蛋白(25%)含有基本的结构, 包括一个信号肽和两个活性位点。51个蛋白(60.42%)缺少信号肽或者活性位点之一, 剩下的14个(14.58%)有缺少信号序列和活性位点的AP结构域。只有6个AP鉴定出是典型的AP蛋白, 其中的5个含有一个Saq_B结构域(图1-B)。拟南芥和水稻在蛋白的结构上存在显著的差异, 这可能是因为进化带来的功能多样性造成的。

从蛋白质中氨基酸数量来看, 拟南芥和水稻中的大多数AP氨基酸数量都在300~600 aa之间, 除了拟南芥C4/NP_190574、C4/NP_199124、其他类/NP_189700、其他类/NP_180389的蛋白质氨基酸数量分别是632、631、630和756 aa; 以及其他类/NP_680599氨基酸数目仅为158 aa。水稻A1/NP_001042785、B/NP_001060161、C1/NP_001056633的蛋白质氨基酸数量分别是52、642和720 aa; 同时, 通过蛋白亚细胞定位分析, 发现拟南芥的此家族中大多数成员都是分泌蛋白, 其中膜蛋白和叶绿体蛋白又占大多数, 有约一半(33/69)基因是集中在内

膜系统中, 只有5个是分泌到细胞外的, 因此, 在拟南芥中AP家族很可能集中在液泡、叶绿体、或者细胞膜等内膜系统上发挥催化作用。而在水稻中却相反, 有多于一半(50/89)的基因集中在细胞外基质中, 其余大部分位于细胞质膜、内质网膜和叶绿体类囊体膜等内膜系统上。在水稻中AP家族既在内膜系统上又在细胞外基质中发挥作用, 主要是分泌到细胞外发挥催化作用(表1)。

3 AP基因家族GO功能分类

尽管我们知道拟南芥AP是一类重要的蛋白水解酶, 它的活性中心由2个催化性天冬氨酸残基组成, 但其可能参与的生物学过程我们并不清楚, 因此, 我们进行了这些基因的GO富集分析(图2), 结果表明: 拟南芥69个AP基因, 在细胞组成过程中, 这类家族蛋白比较集中于内膜系统(endomembrane system), 除此之外, 在植物细胞壁(plant cell wall)以及膜锚定(anchored to membrane)中也有富集; 在生物学过程中, 这类蛋白家族成员集中在蛋白水解功能(proteolysis); 在分子功能上, 它们则集中体现了

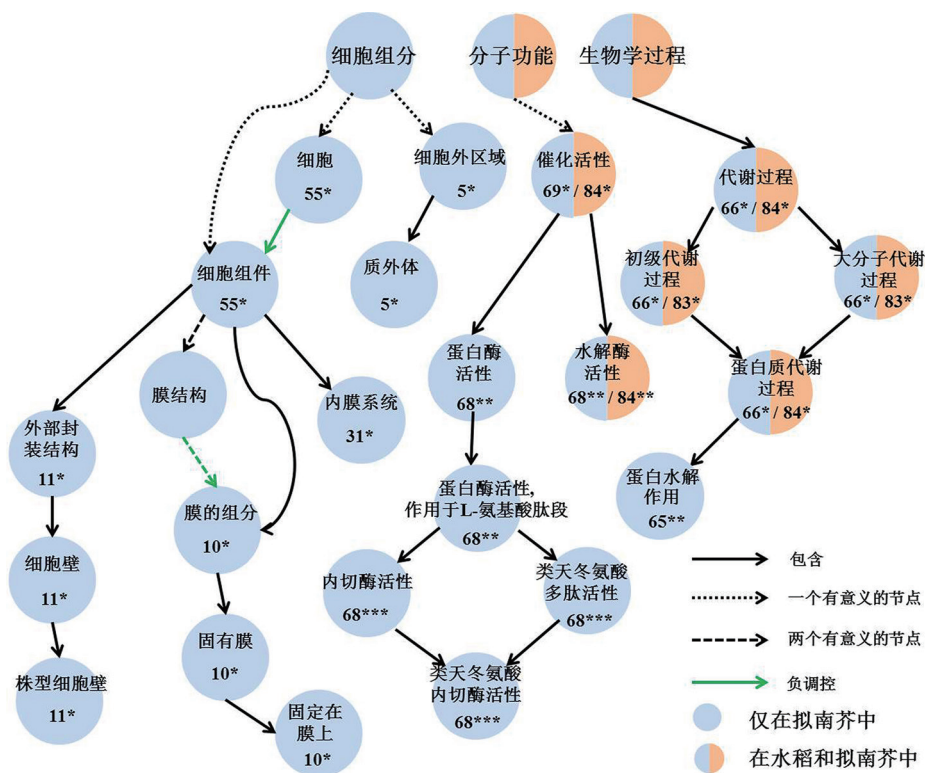


图2 拟南芥和水稻天冬氨酸蛋白GO生物学过程富集分析

Fig.2 GO biological process enrichment analysis of APs in *Arabidopsis* and rice

*表示P值在0.05~1.0e-60, **表示P值在1.0e-60~1.0e-120, ***表示P值在1.0e-120~1.0e-180。

天冬氨酸内切酶活性(图2)。在水稻中,我们也进行了89个AP基因的GO富集分析,结果表明,在生物学过程中,这类蛋白集中在初级代谢、大分子代谢过程和蛋白质代谢过程;在分子功能上,表现出催化活性和水解酶活性。通过拟南芥和水稻的对比分析,我们获知水稻和拟南芥在GO功能分类上存在一定差异,可能是基因进化结果的差异,也可能是水稻GO数据库功能注释不全所致。

4 拟南芥和水稻APs的染色体分布和基因簇

为了解AP基因家族各个亚家族分子进化来源,以及基因间可能存在的相互复制关系,我们分析了AP家族各个成员在染色体上的位置信息。结果表明,在拟南芥中69个基因分布在拟南芥第1至第5号染色体上,其中第1、3和5号染色体分布最多,分别是17、17和16个基因家族成员,而第2和4号染色体分布相对较少,分别是11和8个(图3-A),所有的染色体两臂上均有分布,但从图3-A可以看出,拟南芥AP基因在各个染色体上并不是均匀分布,某些区域分布密度很大,比如,拟南芥3号染色体有17个基因分布,其中下半部某个较小区域就聚集了9个,而物理长度长于3号染色体下部的5号染色体下部总共也只分布了6个基因。结果还表明某些基因是以基因簇的方式存在,尤其存在两个基因紧密连锁的关系,从图3-A中可以看出,共有6个基因簇存在,每个基因簇存在至少2个基因是紧密连锁的,尤其3号染色体下半部分就有3个C4类的基因(At3g50050、At3g51330.1、At3g51340.1)紧密连锁,说明他们在进化的过程中可能起源于同一片段和多拷贝复制。同样水稻中分析结果表明,水稻AP中的86个基因在所有12条染色体上各有分布,其中12个基因定位在1号染色体,8个基因定位在4号染色体,13个基因定位在6号染色体,11个基因定位在10号染色体,4个基因定位在12号染色体,2和7号染色体上分别包含了7个基因,3、5、9号染色体上分别包含了6个基因,8和11号染色体上分别包含了3个基因。水稻AP在染色体上也存在类似于拟南芥中的基因簇,共有9个基因簇存在,基因间存在紧密连锁的关系(图3-B)。

5 拟南芥和水稻AP家族不同发育时期和组织器官表达模式分析

同源基因或者结构序列相近基因在不同的组织器官中表达模式分析,可能为该类基因的生理

生化功能、基因功能分化和新功能的出现研究提供重要的信息。为了阐明拟南芥和水稻APs在植物生长发育过程中的可能的功能分化,我们对拟南芥和水稻基因芯片表达谱数据库的数据进行收集和聚类分析。我们在拟南芥中选择了35个组织器官和不同的发育时期,包括:子叶、莲座叶(1、2、4、6、8、10和12叶)、第7叶(叶柄,叶近端半、叶远侧半)、下胚轴、过渡期至开花后的整个莲座、衰老叶片、茎叶、干(第二节间)、茎的第一节点、茎尖(花序和营养叶)、根、不同花发育期(9、10/11、12和15期)、不同时期(花发育12期和15期)的不同组织(萼片、心皮、雄蕊、花瓣和15期的花梗)。除了9个拟南芥AP基因(At2g28040、At2g28220、At2g35615、At3g25700、At3g51340、At3g51350、At4g04985、At5g10080和At3g30770)由于缺少芯片探针设计的原因没有相关数据,我们获得了其他60个基因的全部35个组织以及不同发育时期的AP基因的表达数据(图4-A)。分析结果表明:拟南芥APs在各个组织中都有表达,但AP在不同组织部位以及发育时期,各个亚家族之间功能存在着差异的,聚类分析结果表明,4个亚家族中的60个AP基因可以分成3个亚组(I、II和III),每个亚组都由2~3类亚家族组成(图4-A)。分析的具体信息如下: I组共计17个基因,主要包含B类(3个成员)、C1类(3个成员)、C3类(2个成员)、C4类(1个成员)以及其他类(8个成员)这五类家族。主要表达特点是表达模式共性较为集中在根和衰老的叶片(图4-A)。此外,在花发育15期的萼片和花瓣以及莲座叶表达量较高(图4-A)。说明了这五类家族除了参与根和叶片的发育,还参与花瓣的发育过程。II组共有25个基因,主要包括A2类(2个成员)、B类(1个成员)、C1类(6个成员)、C2类(1个成员)、C3类(3个成员)、C4类(3个成员)、C5类(3个成员)、CX类(1个成员)、其他类(5个成员)这几类家族。从家族分布的角度来看十分广泛(除了不含唯一的A1类At1g11910基因)。主要的表达特点是表达模式共性比较集中在花发育9期和花发育15期的心皮(图4-A)。说明这类基因主要参与拟南芥花器官的发育,而对于拟南芥的营养生长参与较少。III组共有18个基因,主要包括A1类(1个成员)、C1类(1个成员)、C2类(3个成员)、C3类(3个成员)、C4类(5个成员)、CX(3个成员)以及其他类(2个成

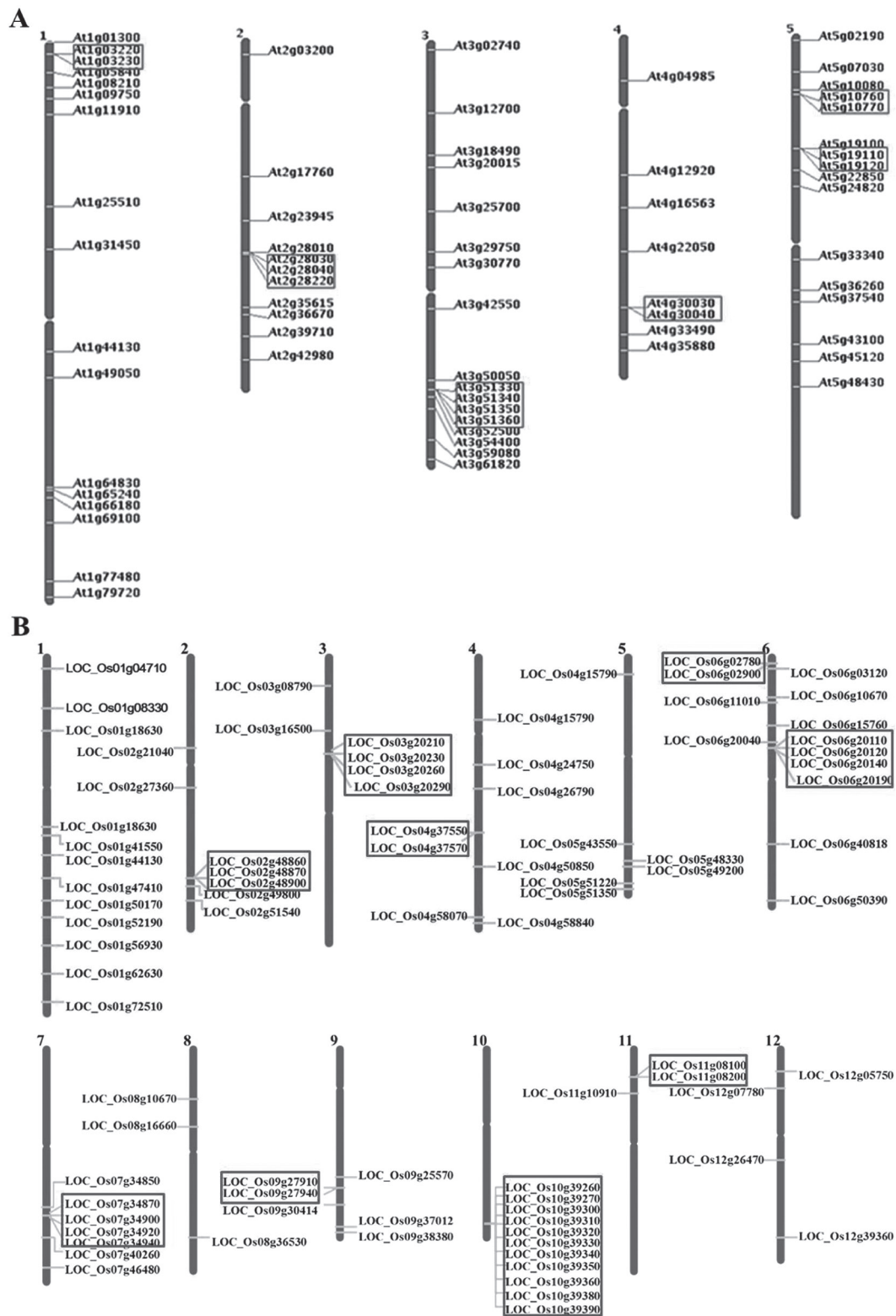


图3 拟南芥和水稻天冬氨酸蛋白酶基因的染色体分布示意图

Fig.3 Localization of AP genes on the *Arabidopsis* and rice chromosomes

A: 拟南芥中69个天冬氨酸蛋白酶基因在染色体上的分布; B: 水稻中86个天冬氨酸蛋白酶基因在染色体上的分布。图上1~12数字为染色体编号; 实线框中为含有紧密连锁基因的基因簇。

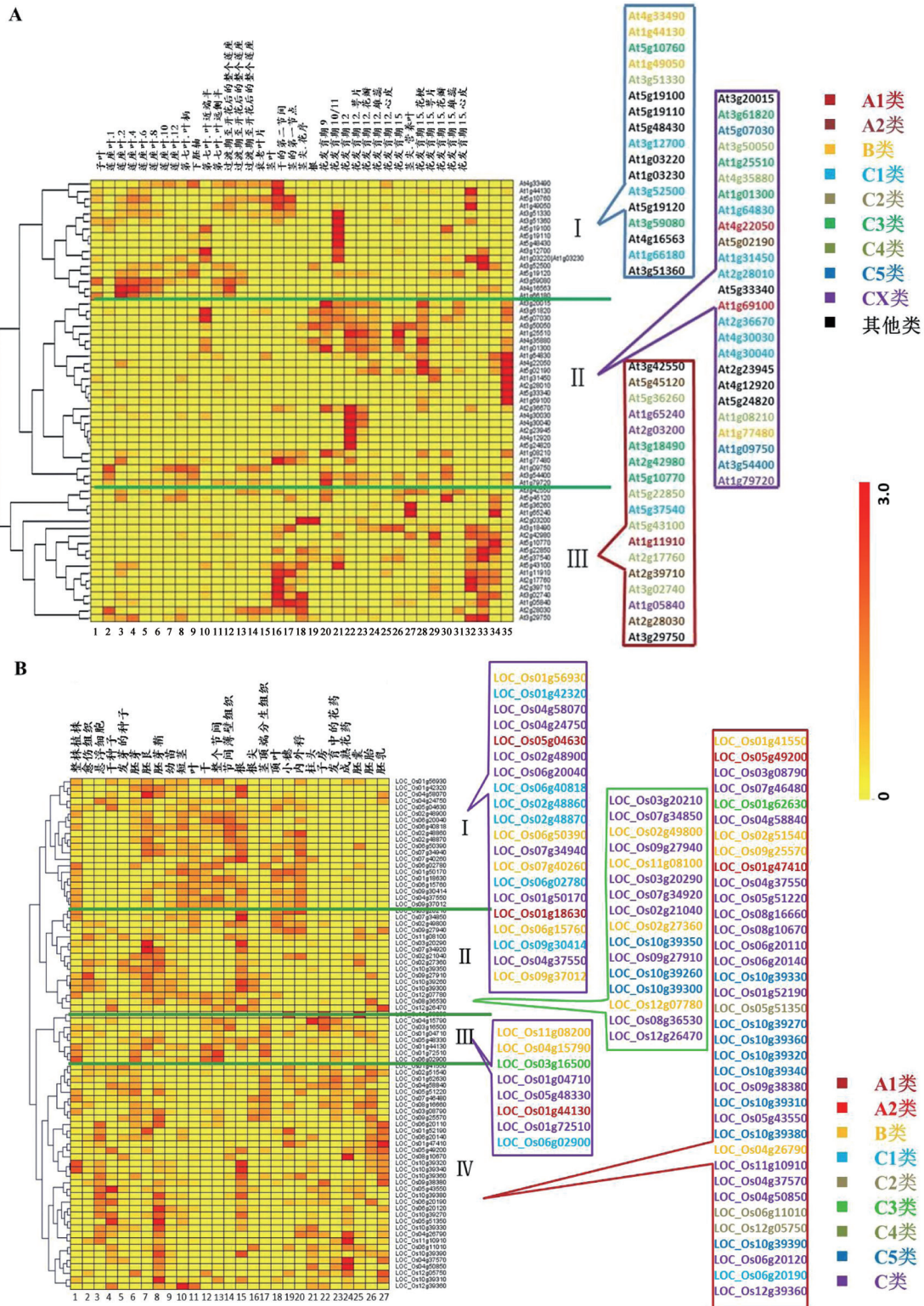


图4 拟南芥和水稻天冬氨酸蛋白酶基因在组织器官表达模式的系统聚类
Fig.4 Hierarchical cluster display of expression profiles for *Arabidopsis* and *rice* aspartyl protease genes in different tissues and/or organs

A: 拟南芥中60个天冬氨酸蛋白酶基因在35个组织以及不同发育时期的表达数据; B: 水稻中79个天冬氨酸蛋白酶基因在27个组织及不同发育时期的表达数据。

员)这七类家族。主要表达的特点是表达模式共性比较集中在花发育15期萼片、花瓣和雄蕊(图4-A)。此外,在衰老的叶片、茎叶和干(第二节间)也有比较集中的表达(图4-A)。说明这些基因的功能主要集中参与叶片和花器官发育这两个方面。

同样,我们对水稻中的89个基因进行了组织和器官的表达情况的分析,我们得到了其中79个基因的全部27个组织和器官的表达数据(图4-B)。聚类结果分析表明,这些基因可以分成4个亚组(I, II, III, IV),每个亚组包括2~4类亚家族。具体信息如下: I组共计20个基因,主要包括A类(2个成员)、B类(5个成员)、C类(7个成员)、C1类(5个成员)和C5类(1个成员)家族基因,主要表达特点,以LOC_Os06g02780为代表的10个基因在植物中具有广泛表达(图4-B),说明它们对于植物生长发育是必须的组成型基因。此外, I组其他基因还在幼根、胚芽鞘、幼苗、芽、叶子、茎、节间、根、顶叶这些部位集中表达且表达量较高(图4-B),说明这些基因可能在这些组织器官中有非常重要的作用。II组共计16个基因,主要包括B类(4个成员)、C类(9个成员)和C5类(3个成员)基因,这个组的主要表达特点是在柱头、子房、发育中的花药、成熟花药中几乎所有基因都不表达,部分基因在胚囊和胚中有少量表达(图4-B),说明这些基因都不参与生殖发育过程,但在其他营养器官中都有不同程度的表达。最突出的一点是,这个组中的几乎所有基因都在胚根和根中有表达,其中LOC_Os03g20290、LOC_Os07g34920在胚根中高表达,LOC_Os07g34850、LOC_Os03g20290、LOC_Os10g39260、LOC_Os10g39300在根中表达量非常高(图4-B),这些基因对于根的发育有重要的调控作用。III组共计8个基因,主要包括A类(1个成员)、B类(2个成员)、C类(3个成员)、C1类(1个成员)和C3类(1个成员)基因,主要表达特点是分散在胚、茎节间、小穗、柱头、子房、胚囊等部位表达,其中LOC_Os11g08200特异的在小穗、子房、胚囊中表达(图4-B),它参与了水稻花药的发育。LOC_Os04g15790主要在柱头、子房、花药中高表达(图4-B),说明这个基因对于水稻的生殖生长具有重要的影响。IV组共计36个基因,主要包括A类(2个成员)、B类(4个成员)、C类(17个成员)、

C1类(1个成员)、C2类(3个成员)、C3类(1个成员)和C5类(8个成员)家族基因。主要表达特点是大部分基因在小穗、子房、花药、胚囊、胚这些生殖组织器官中有集中表达,同时在悬浮细胞、干种子、胚芽、胚乳、胚芽鞘、根中也有较多基因表达。这些基因中LOC_Os08g10670、LOC_Os04g26790、LOC_Os11g10910、LOC_Os04g37570和LOC_Os04g50850在成熟花药中表达量非常高,是花药成熟相关的重要基因,LOC_Os10g39360、LOC_Os10g39320和LOC_Os10g39340在根中高表达,本组中近2/3的基因在胚芽鞘中都有不同程度的表达(图4-B),这些基因分别参与根和胚芽鞘的生长发育。

我们通过以上对AP基因在拟南芥和水稻中不同组织部位和发育时期的表达模式和聚类分析结果,可以推测,在进化的过程中拟南芥和水稻的AP家族基因的表达出现了显著差异分化,从而导致各个不同亚家族存在比较明显的功能差异,从而增加了AP功能多样性和丰富度,但AP基因表达并不是彼此完全独立的,而是存在一定的规律,这说明他们之间的功能具有一定的协作和相互作用(图4)。

6 拟南芥和水稻AP基因在花药绒毡层细胞死亡中的功能预测分析

为进一步说明AP基因在正常生理过程中存在功能多样性和协调性,我们选择了花药发育过程这一细胞发育特殊的过程来分析AP基因的功能。我们知道,花药的形成是一个十分复杂的过程,其中,花粉囊壁的最内一层细胞绒毡层细胞的降解是一个细胞程序化死亡的过程(Niu等2013; Li等2006),这层细胞的死亡过程可控制性对花粉的发育非常重要。已经有研究表明,在这个过程中,AP蛋白起到关键的作用(Niu等2013)。因此,为研究AP蛋白酶在细胞死亡中可能存在的功能,在拟南芥中,我们通过提取花药发育过程中的各个时期和组织部位的芯片,主要包括:花期9(早期)、花期10/11(中早期)和花期12(中晚期)、花期15(后期);还有特异性器官花期12-雄蕊(中期雄蕊)和花期15-雄蕊(后期雄蕊),以及衰老叶片(作为不同细胞死亡类型的比较)。研究结果表明,根据AP基因在不同组织和时期之间的特异性表达(图5-A),可以分成四大类:第一类(图5-A, I)中包含了14个基因,主要包括C1 (At2g28010)、C4 (At3g50050和

At5g36260)和CX (At2g03200、At1g65240和At3g42550)。它们的共同特点是在中期雄蕊集中表达, 并且表达量较高; 个别基因在前期有表达, 但是表达量不高, 说明这类基因参与花药中期的发育。第二类(图5-A, II)中包含了12个基因, 主要包括B、C1、C3、C4、C5、CX和其他类, 它们的特点是都在花发育的前中期表达, 其中, 约1/3的基因在花发育早期(花发育9期)有较高的表达量

(图5-A); 其他的基因则表达特异性不强且表达量不高, 说明这类基因参与拟南芥花药早期的发育。第三类(图5-A, III)中包括了15个基因, 主要包括A1、B、C2、C3、C4、CX和其他类, 它们特异性的在衰老的叶片中表达且表达量较高, 除了C4 (At5g22850)在花药发育后期的雄蕊中有特异表达(花发育15期-雄蕊) (图5-A), 其他的基因几乎没有在任何花特异器官和发育时期中表达, 这说明了这类

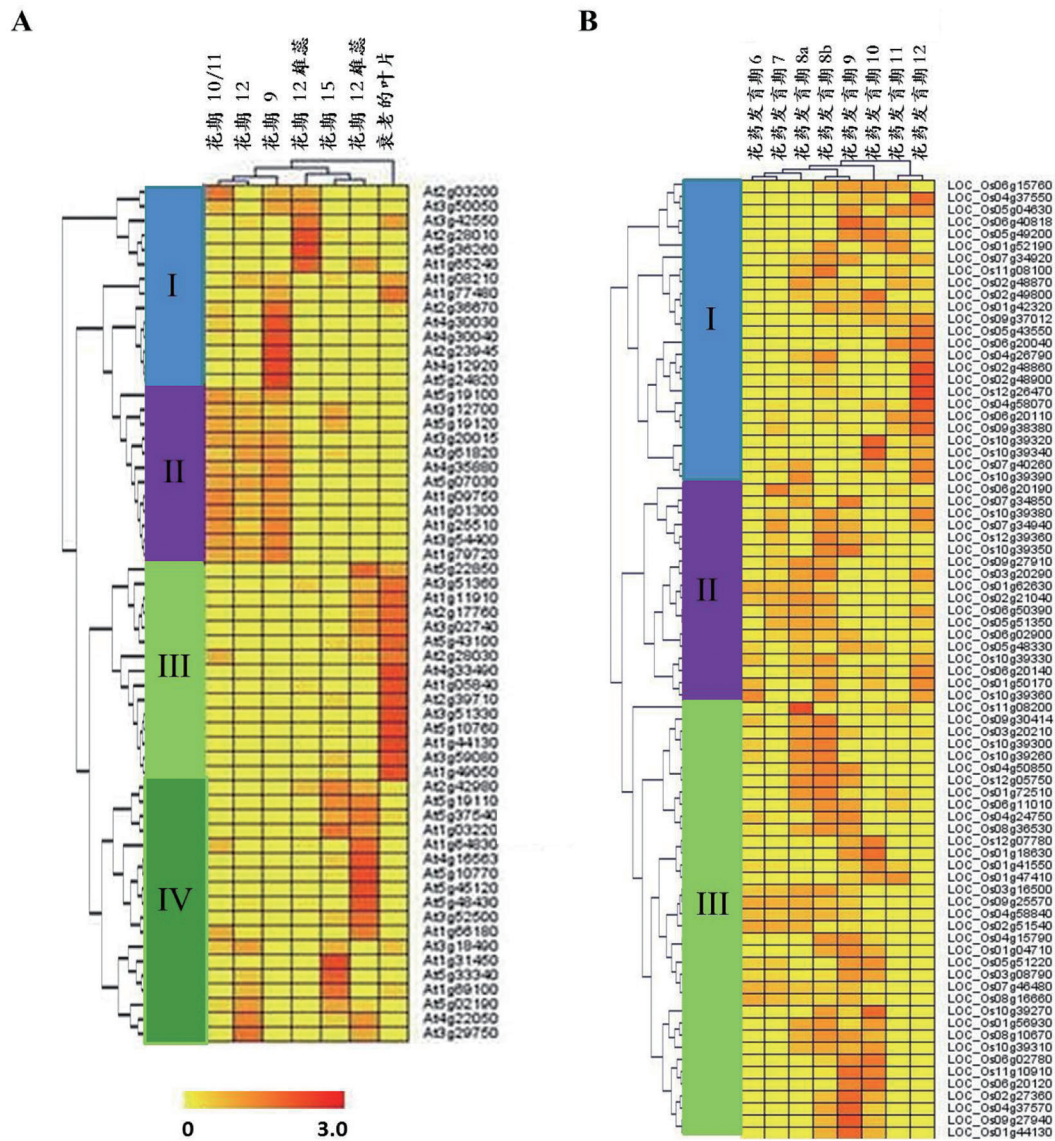


图5 拟南芥和水稻天冬氨酸蛋白酶家族基因在花药发育过程中的系统聚类

Fig.5 Hierarchical cluster display of expression profiles for *Arabidopsis* and rice aspartyl protease genes during the anther development

A, 拟南芥花的不同发育时期: 花期9 (早期)、花期10/11 (中早期)和花期12 (中后期)、花期15 (后期)以及特异性器官花期12-雄蕊(中期雄蕊)和花期15-雄蕊(后期雄蕊), 以及衰老叶片的芯片数据; B, 水稻花药不同发育时期: 花药发育期6-7 (早期)、花药发育期8-9 (中期)和花药发育期10-12 (晚期)的芯片数据; 图中I、II、III、IV表示根据AP基因在不同组织和时期之间的表达差异进行的分类。

基因不参与拟南芥花药发育过程。第四类(图5-A, IV)中包含了18个基因, 主要包括A2、C1、C2、C3和其他类, 这类基因中的大部分成员会在花药发育后期的雄蕊中特异性表达(花发育15期-雄蕊), 少部分会在拟南芥花药发育中期和后期特异性表达(花发育12期和15期)(图5-A), 不过表达量不是很高, 这说明了这类基因可能参与拟南芥花药的后期发育。

在水稻中, 同样选取了花药发育的8个不同时期(主要包括花药发育6、7、8a、8b、9、10、11和12期), 其中6~7期为对应的水稻花药发育早期, 8~9为水稻花药发育中期, 10~12为水稻花药发育晚期。研究结果表明(图5-B), *AP*基因在不同时期之间存在特异性表达, 可以分成三大类: 第一类(图5-B, I)中包含了25个基因, 主要包括A1、B、C、C1和C5类, 这些基因在花药发育的中期和后期表达量较高, 其中有18个基因在花药发育期12有不同程度的表达, 说明这些基因参与花药的中期和后期发育。第二类(图5-B, II)中包含了18个基因, 主要包括B、C、C1、C2、C3和C5类, 这一类别中的基因主要在花药发育的前期和中期表达较高, 可以推测这些基因参与了水稻花药前、中、后期的发育, 且集中在C类中。第三类(图5-B, III)中包含了36个基因, 主要包括A1、B、C、C1、C2、C3和C5类, 它们主要在花药的中期表达, 部分基因在后期的较早阶段(第10、11期)也有一定的表达, 花药发育期第6~11期是花药绒毡层发育的主要时期, 说明这些基因参与花药中期的发育, 对于绒毡层发育也具有重要的意义。

通过拟南芥和水稻花药芯片数据的分析, 我们可以获知大部分*AP*基因在花药的不同发育时期和组织部位都是有表达变化的, 说明*AP*蛋白在花药的发育过程中功能的广谱性, 但各个时期又彼此差异和互补表达, 也说明了功能分工和协调具有严格的控制。尤其考虑到花药绒毡层细胞的死亡特性, 以及根据目前的研究结果, 说明*AP*蛋白在绒毡层的降解过程中, 可能是起到一个关键的作用。

讨 论

本文通过对拟南芥和水稻中的*AP*家族基因的序列、分类、基因结构和染色体上分布等信息的分析, 发现尽管这两个物种中*AP*基因家族的分类

方法相似, 但每个类别中的成员之间从结构到功能都有着巨大的差异。从基因结构上来看, *AP*不同亚家族基因的内含子数目存在差异较大, 但其关键功能区域都是含有两个活性位点的*AP*结构域的, 这个区域的同源性极高, 进化上高度保守, 不同*AP*蛋白功能的改变来源于可能是一些特定部位氨基酸的差异, 从而确保*AP*蛋白功能的进化保守性和进化带来的功能适应多样性。而在蛋白亚细胞定位上, 拟南芥中大多数*AP*成员多是集中在内膜系统中分泌蛋白, 而在水稻中却相反, 在水稻中*AP*蛋白既在内膜系统上又在细胞外基质中发挥作用, 主要是分泌到细胞外发挥催化作用, 尽管我们还不清楚在进化上导致拟南芥和水稻*AP*蛋白亚细胞定位改变的动力和原因是什么, 推测可能与水稻和拟南芥生活的环境和*AP*蛋白在这两个物种中功能的差异相关。在进一步的GO功能分析来看, 拟南芥69个*AP*基因, 在生物学过程中, 这类蛋白家族成员集中在蛋白水解功能; 在分子功能上, 它们则集中体现了天冬氨酸内切酶活性。在水稻中, 89个*AP*基因在生物学过程中, 这类蛋白集中在初级代谢、大分子代谢过程和蛋白质代谢过程; 在分子功能上, 表现出催化活性和水解酶活性。也进一步说明了*AP*蛋白在水稻和拟南芥中功能的差异分化, 这种GO功能注释上的差异可能是基因进化的结果, 也可能是水稻GO数据库功能注释不全所致。

拟南芥中69个*AP*基因在染色体不均匀分布, 部分基因是以基因簇的方式存在, 尤其存在两个基因紧密连锁的关系, 说明他们在进化的过程中可能起源于同一片段和多拷贝复制。这种情况在水稻中也相似。综上说明这些基因在复制以后结构域发生变化, 导致基因产物的功能也发生相应的改变, 逐渐形成一些特有的功能。基因的复制是基因演化过程中的原始驱动力, 全基因组的基因重复性为多复制基因的序列和表达的进化提供了未加工的基因材料(Chen等2009; 朱璐等2012)。基因重复在*AP*基因家族的功能多样性形成中起到重要的作用。

我们通过对拟南芥和水稻*AP*基因表达信息, 拟南芥*AP*基因在各个组织中都有表达, 但在不同组织部位以及发育时期、各个亚家族之间功能存在着差异的; 水稻*AP*基因在幼根、胚芽

鞘、幼苗、芽、叶子、茎、节间、胚根、根和顶叶等营养器官中表达量较高,可能对这些组织器官功能有非常重要的作用。我们可以推测,在进化的过程中拟南芥和水稻的AP家族基因的表达出现了显著差异分化,导致各个不同亚家族存在比较明显的功能差异,从而增加了AP蛋白功能多样性和丰富度,但AP基因表达并不是彼此完全独立的,而是存在一定的规律,这说明他们之间的功能具有一定的协作和相互作用。

花药的形成是一个十分复杂的过程,尤其花粉囊壁的最内一层细胞绒毡层细胞程序化死亡的过程对花粉的发育非常重要,根据花药绒毡层细胞的死亡特性,以及目前相关的研究结果,已经初步证明AP蛋白在绒毡层的降解过程中可能起到一个关键的执行者功能(Niu等2013; Li等2006)。因此,我们进一步利用公共芯片数据,全面分析花药发育过程中AP基因的表达情况,以期进一步了解和挖掘AP蛋白在绒毡层细胞死亡中的功能,为相关的方向遗传学提供依据。我们的分析结果表明,在拟南芥中,至少有12个AP基因参与花药早期的发育、14个AP基因参与花药中期的发育、18个AP基因参与拟南芥花药的后期发育。在水稻中,25个AP基因参与花药的中期和后期发育、18个AP基因参与了水稻花药前、中、后期的发育、36个AP基因主要在花药的中期表达。以上结果说明,不管在水稻和拟南芥花药发育中,大部分AP蛋白在花药的不同发育时期和组织部位都是有表达变化的,这可能与花药发育过程中,涉及多层细胞的形成、快速降解等变化是密切相关的。尽管AP蛋白在花药的发育过程表达具有广谱性,但具体到各个发育时期,不同AP亚家族之间的表达还是存在一定的差异性,说明AP蛋白之间彼此功能差异分化,也说明了AP基因的协调表达具有严格的控制。

参考文献

陈荣,汲广东(2011). 天冬氨酸蛋白酶家族的分类、功能和演化. 鲁东大学学报(自然科学版), 27 (4): 328~334
 丁宁,沈元月,冯永庆,秦岭(2008). 植物细胞程序性死亡分子机制和信号转导. 生物技术通报, (1): 25~29
 李晶,郑彩霞(2012). 高等植物珠被绒毡层的研究. 植物生理学报, 48 (12): 1135~1140
 吕刚,韩笑,张秀娟,杜培增(2008). 天冬氨酸蛋白酶的研究进展. 吉林化工学院学报, 25 (1): 13~18

张虹,梁婉琪,张大兵(2008). 花药绒毡层细胞程序性死亡研究进展. 上海交通大学学报(农业科学版), 26 (1): 86~90
 朱璐,许杰,张大兵(2012). 拟南芥ABC转运类蛋白家族的分子进化、表达模式和蛋白功能网络预测分析. 植物生理学报, 48 (12): 1151~1166
 Abramovitch RB, Kim YJ, Chen S, Dickman MB, Martin GB (2003). *Pseudomonas* type III effector AvrPtoB induces plant disease susceptibility by inhibition of host programmed cell death. EMBO J, 22: 60~69
 Barrett AJ (1992). Cellular proteolysis: an overview. Ann N Y Acad Sci, 674: 1~15
 Barrett AJ, Woessner JF, Rawlings ND (2004). Handbook of Proteolytic Enzymes. London: Elsevier
 Chen F, Foolad MR (1997). Molecular organization of a gene in barley which encodes a protein similar to aspartic protease and its specific expression in nucellar cells during degeneration. Plant Mol Biol, 35 (6): 821~831
 Chen JJ, Ding JH, Ouyang YD, Du HY, Yang JY, Cheng K, Zhao J, Qiu SQ, Zhang SL, Yao JL et al (2008). A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of indica-japonica hybrids in rice. Proc Natl Acad Sci USA, 105: 11436~11441
 Chen JJ, Ouyang YD, Wang L, Xie WB, Zhang QF (2009). Aspartic proteases gene family in rice: Gene structure and expression, predicted protein features and phylogenetic relation. Gene, 442 (1-2): 108~118
 Chen Z, Sliva H, Klessing DF (1993). Active oxygen species in the induction of plant systemic acquired resistance by salicylic acid. Science, 262: 1883~1886
 Davies DR (1990). The structure and function of the aspartic proteinases. Annu Rev Biophys Biophys Chem, 19: 189~215
 Faro C, Gal S (2005). Aspartic proteinase content of the *Arabidopsis* genome. Curr Protein Pept Sci, 6 (6): 493~500
 Francischetti I, Moreno JB, Scholz M, Yoshida WB (2010). Os leucócitos e a resposta inflamatória na lesão de isquemia-reperusão. Rev Bras Cir Cardiovasc, 25 (4): 575~584
 Ge XC, Dietrich C, Matsuno M, Li G, Berg H, Xia Y (2005). An *Arabidopsis* aspartic protease functions as an anti-cell-death component in reproduction and embryogenesis. EMBO Rep, 6: 282~288
 Hensel LL, Grbic V, Baumgarten DA, Bleecker AB (1993). Acclimation of photosynthetic proteins to rising atmospheric CO₂. Plant Cell, 5: 553~564
 Kamitori S, Ohtaki A, Ino H, Takeuchi M (2003). Crystal structures of *Aspergillus oryzae* aspartic proteinase and its complex with an inhibitor pepstatin at 1.9 Å resolution. J Mol Biol, 326 (5): 1503~1511
 Li N, Zhang DS, Liu HS, Yin CS, Li XX, Liang WQ, Yuan Z, Xu B, Chu HW, Wang J et al (2006). The rice *tapetum degeneration retardation* gene is required for tapetum degradation and anther development. Plant Cell, 18 (11): 2999~3014
 Mutlu A, Gal S (1999). Plant aspartic proteinases: enzymes on the way to a function. Physiol Plant, 105 (3): 569~576
 Nakano T, Murakami S, Shoji T, Yoshida S, Yamada Y, Sato F (1997). A novel protein with DNA binding activity from tobacco chloro-

- plast nucleoids. *Plant Cell*, 9: 1673~1682
- Niu NN, Liang WQ, Yang XJ, Jin WL, Wilson ZA, Hu JP, Zhang DB (2013). EAT1 promotes tapetal cell death by regulating aspartic proteases during male reproductive development in rice. *Nat Commun*, 4: 1445
- Phan G, Remaut H, Wang T, Allen WJ, Pirker KF, Lebedev A, Henderson NS, Geibel S, Volkan E, Yan J et al (2011). Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. *Nature*, 474 (7349): 49~53
- Rawlings ND, Barrett AJ (1999). *MEROPS*: the peptidase database. *Nucleic Acids Res*, 27 (1): 325~331
- Simoes I, Faro C (2004). Structure and function of plant aspartic proteinases. *Eur J Biochem*, 271 (11): 2067~2075
- Solís MT, Chakrabarti N, Corredor E, Cortés-Eslava J, Rodríguez-Serano M, Biggiogera M, Risueño MC, Testillano PS (2014). Epigenetic changes accompany developmental programmed cell death in tapetum cells. *Plant Cell Physiol*, 55 (1): 16~29
- Solomon M, Belenghi B, Delledonne M, Menachem E, Levine A (1999). The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. *Plant Cell*, 11: 431~443
- Xia Y, Suzuki H, Borevitz J, Blount J, Guo Z, Patel K, Dixon RA, Lamb C (2004). An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. *EMBO J*, 23: 980~988
- Xu FX, Chye ML (1999). Expression of cysteine proteinase during developmental events associated with programmed cell death in brinjal. *Plant J*, 17 (3): 321~328