

· 研究报告 ·

水稻AT-hook基因家族生物信息学分析

张贵慰^{1†}, 曾珏^{1†}, 郭维¹, 罗琼^{2*}

¹云南农业大学农学与生物技术学院, 昆明 650201; ²云南农业大学植物保护学院, 昆明 650201

摘要 AT-hook是一种小型的DNA结合蛋白基序, 在哺乳动物非组蛋白染色体的高移动组蛋白(HMG-I/Y)中首次发现。对其它生物的研究表明, AT-hook蛋白在染色质结构组装、靶细胞特异性结合、转录调控和生长发育的调控中发挥重要作用。利用生物信息学方法, 从水稻(*Oryza sativa*)基因组中鉴定出了45个编码AT-hook蛋白的基因, 并对这些基因的系统进化、染色体定位及其编码蛋白的结构和功能等进行了系统的生物信息学分析。结果表明, 水稻AT-hook蛋白的结构和特性分化不显著; 45个水稻AT-hook基因可划分成5个亚族; 染色体复制是基因家族成员扩增的进化途径之一。基因数字表达分析结果显示, AT-hook基因主要在水稻幼穗中表达, 并通过qRT-PCR验证了部分基因的数字表达结果。

关键词 AT-hook基因, 生物信息学, 基因表达, 水稻

张贵慰, 曾珏, 郭维, 罗琼 (2014). 水稻AT-hook基因家族生物信息学分析. 植物学报 49, 49–62.

AT-hook基序是以精氨酸-甘氨酸-精氨酸-脯氨酸(RGRP)4个残基为中心的特征结构, 它是一种小型的DNA结合蛋白基序, 广泛存在于不同物种的DNA结合蛋白中(Churchill and Travers, 1991)。这个短的保守氨基酸序列对蛋白结合DNA和核定位是必需的(Aravind and Landsman, 1998; Do et al., 2006)。不同物种中, 这类蛋白所含AT-hook基序的数量存在差异。酿酒酵母(*Saccharomyces cerevisiae*)ORC2蛋白含有1个AT-hook基序(Duncker et al., 2009); 哺乳动物的HMGA蛋白含有3个AT-hook基序(Elías-Arnanz et al., 2010); 人(*Homo sapiens*)LEDGF/p75含有2个AT-hook基序(Shun et al., 2008); 玉米(*Zea mays*)BAF1含有1个AT-hook基序(Gallavotti et al., 2011); 棉花(*Gossypium hirsutum*)GhAT1含有2个AT-hook基序(Delaney et al., 2007)。

AT-hook蛋白与DNA的特异结合是通过AT-hook基序的氨基酸残基与双链DNA小沟中富含AT碱基的区域相互作用完成的(Aravind and Landsman, 1998)。如结核杆菌Lsr2蛋白与含5对AT碱基的序列相结合发挥作用(Gordon et al., 2010); 大豆(*Glycine max*)HMG-I蛋白通过与大豆根瘤素基因N23启动子和质体

蓝素基因启动子富含AT的序列结合而相互作用(Jacobsen et al., 1990; Webster et al., 2001); 小麦(*Triticum aestivum*)AHM1蛋白与染色体DNA富含AT的核基质附着区(matrix attachment region, MAR)相互作用(Morisawa et al., 2000); 果蝇(*Drosophila melanogaster*)TFA1蛋白结合hsp70启动子中富含AT碱基的DNA序列(Metcalf and Wassarman, 2006); 拟南芥(*Arabidopsis thaliana*)AHL1蛋白通过AT-hook基序与染色体的核基质附着区结合, 进而结合到拟南芥质体蓝素基因(PC-MAR1)的MAR序列上(Fujimoto et al., 2004)。

许多实验结果表明, AT-hook蛋白作为染色体结构和转录因子的辅助因子在基因转录活性的调控中起重要作用(Girard et al., 1998; Reddy et al., 2005; Burian et al., 2012a)。WhiB7蛋白可增强结核杆菌的耐药性(Burian et al., 2012b); 人HMGA1蛋白可负调节脂肪细胞的生长, HMGA2蛋白可正调控脂肪细胞的生长(Fedele et al., 2001); HMGI-C中AT-hook基序的破坏则会导致脂肪瘤(Ashar et al., 1995); 小鼠(*Mus musculus*)中HMGA2过量表达则会导致垂体腺瘤(Pierantoni et al., 2005); 棉花GhAT1蛋白抑制

收稿日期: 2013-03-11; 接受日期: 2013-06-08

基金项目: 国家自然科学基金(No.30760100)

† 共同第一作者。

* 通讯作者。E-mail: qiongbf@aliyun.com

*FSltp4*基因的表达,影响棉纤维的生长,因此利用该基因进行遗传改良可改善棉花的纤维质量(Delaney et al, 2007);拟南芥GIANT KILLER (GIK)直接调控AG下游多个与拟南芥繁殖器官分化相关的重要转录因子(Matsushita et al., 2007);玉米BAF1在叶芽分生组织诱导中发挥作用,是玉米雌穗形成所必需的(Morisawa et al., 2000)。

与动物相比,植物中关于AT-hook蛋白功能的研究报道较少。一些研究结果表明,AT-hook蛋白在植物生长发育中的功能涉及植物光形态发生(Lim et al., 2007)、茉莉酮酸酯信号转导(Street et al., 2008)、赤霉素生物合成(Ng et al., 2009)、细胞程序化死亡(Martínez-García and Quail, 1999)和下胚轴伸长(Vom et al., 2007)等方面。在双子叶模式植物拟南芥的基因组中鉴定了29个AT-hook蛋白(AT-hook nuclear localized protein, AHL),大多数AHL蛋白的功能未知,其相应的分子机制也不清楚(Ng et al., 2009; 肖朝文和傅永福, 2009)。水稻(*Oryza sativa*)是重要的粮食作物,也是单子叶模式植物,其基因组水平上AT-hook基因家族的生物信息学分析尚未见报道,其相关蛋白的功能研究也近乎空白。*DP1*是水稻中报道的第1个编码AT-hook DNA结合蛋白的基因(Jin et al., 2011),该基因不仅是调控水稻内稃发育的重要功能基因,而且在水稻营养器官的发育中也具重要作用,暗示其可能参与水稻多方面的发育调控,是水稻中一个重要的发育调控基因(唐雄杰等, 2011)。

水稻基因组测序的完成为利用生物信息学方法从基因组水平上研究和鉴定水稻重要功能基因提供了更加方便和快捷的手段。黄志刚等(2008)在Blast检索和生物信息学分析的基础上设计特异引物,以超级杂交水稻亲本1S为材料,通过RT-PCR扩增获得了编码水稻生长素受体TIR1的类似基因*OsTIR1*的cDNA。Macas等(2010)利用k-mer频率谱分析发现水稻smRNAs来源于着丝粒CentO序列。Shaik和Rama-krishna(2012)利用生物信息学分析了5 468个水稻干旱响应基因(DRGs),发现约有75%的DRGs参与染色质重塑时表达下调。本文从基因组水平上对水稻AT-hook基因家族进行了较系统的生物信息学分析,包括蛋白的类型、编码基因在染色体上的分布和系统进化等,以期水稻AT-hook蛋白功能的深入研究提

供参考。

1 材料与方法

1.1 水稻AT-hook基因家族成员的识别、保守结构域预测和基因结构分析

Pfam数据库中(Finn et al., 2006)标号为PF02178的结构域是AT-hook蛋白的特征结构域。首先从水稻TIGR数据库(<http://rice.plantbiology.msu.edu>, The TIGRRice Database release 5.0)(Yuan et al., 2003; Ouyang et al., 2007)中下载水稻全基因组序列,然后利用基于隐马尔科夫模型的HMMER2.3.2程序(Ed-dy, 1998)搜索水稻中含有AT-hook结构域的候选序列。利用SMART(<http://smart.embl-heidelberg.de/>)在线工具检测候选蛋白质序列,把显示不含AT-hook基序的蛋白序列剔除,最后得到编码AT-hook蛋白的基因序列,并同时获得这些基因的染色体位置信息。

从TIGR数据库中下载水稻AT-hook基因家族的DNA及其cDNA序列,用在线工具GSDS (<http://gsds.cbi.pku.edu.cn/index.php>)(郭安源等, 2007)分析其基因结构,了解基因的内含子、外显子及基因不同剪接体的结构。

1.2 序列联配分析与系统进化树的构建

使用Clustal W软件(Larkin et al., 2007)对搜集到的水稻AT-hook蛋白序列进行多序列联配分析。以序列联配的结果为基础,用MEGA5.05软件生成水稻AT-hook蛋白的系统进化树。进化树生成采用邻接法(neighbor-joining method)。具体设置参数Test of Phylogeny: Bootstrap method; Replications: 1000; Method: Poisson model; Rates among Sites: Uniform rates; Gaps: Pairwise deletion。其它参数为默认。

1.3 AT-hook基因家族在染色体组上的定位

根据各AT-hook基序基因在染色体组中的位置信息,用MapInspect软件生成每个AT-hook基序基因在染色体组上的位置,从而得到各AT-hook基序基因在基因组中的分布状况。利用TIGR数据库中的水稻基因组片段复制信息,对水稻染色体片段复制情况进行分析,在染色体上标出属于片段复制的基因对。

1.4 水稻AT-hook基因家族数字表达分析

从TIGR数据库中收集与水稻AT-hook基序基因匹配的ESTs数据。使用现有的水稻发育芯片数据CREP database(<http://crep.ncpgr.cn/>)(包括29个发育阶段和39个组织器官类型), 分析AT-hook基序基因在水稻不同组织器官和不同发育阶段的表达情况。

1.5 qRT-PCR分析

将粳稻品种日本晴(*Oryza sativa* L. ssp. *japonica* cv. 'Nipponbare')种植于中国科学院遗传与发育生物学研究所温室。在营养生长阶段, 取根和叶并用水冲洗干净。用解剖针逐层剥除外层叶鞘和幼叶, 用刀片切取茎尖分生组织(shoot apical meristem, SAM)。在生殖生长阶段取不同生长期的穗(P1, 0–3 cm; P2, 3–5 cm; P3, 5–10 cm; P4, 10–15 cm; P5, 15–22 cm)。将以上材料用液氮速冻后放入–70°C冰箱中保存备用。

采用Trizol(Invitrogen)法提取总RNA。总RNA经DNaseI(Invitrogen)处理去除基因组DNA, 使用NanoDrop 2000c分光光度计(Thermo)检测RNA样品的浓度和纯度。吸取3 μg总RNA经M-MLV(Invitrogen)反转录合成第1链cDNA, 稀释成50 ng·μL⁻¹作为qRT-PCR的模板。使用Primer Premier 5.0软件设计引物。qRT-PCR采用Chromo4 Real Time PCR系统(PTC-2000, Bio-Rad), 使用SYBR Green1染料。qRT-PCR反应程序为: 95°C15秒, 58°C10秒, 72°C10秒, 40个循环; 76°C和82°C分别读板1次。采用2^{-ΔΔCt}方法计算基因表达相对量, 每个基因的表达反应重复3次, 以Actin作为内参基因。

2 结果与讨论

2.1 水稻AT-hook基因家族成员的识别

以Pfam数据库中的AT-hook特征基序为检索序列, 检索TIGR的水稻全基因组注释蛋白质数据库, 将得到的候选序列用SMART工具预测保守结构域, 把明显不含有AT-hook基序的蛋白剔除。结果共获得54条蛋白质序列, 它们分别由45条AT-hook基序基因编码, 基因重新以AT-hook基序基因在染色体上的位置顺序命名为OsAHL01–OsAHL45(表1)。由于可变剪接所编码的蛋白质包含的氨基酸序列几乎相同, 且在进化关系上保持一致, 所以在染色体上具有相同位置

的序列仅保留第1条。

2.2 水稻AT-hook蛋白的分类

根据序列相似性和AT-hook基序的旁侧序列及其与DNA结合的亲和力大小, 将这些基序分为3类(Aravind and Landsman, 1998)。水稻AT-hook蛋白根据其分子功能可分为3类, 第1类为结合蛋白, 大部分水稻AT-hook蛋白参与DNA的结合, 其它参与蛋白和染色质等的结合; 第2类为转录因子, 其中Q10J83和Q2R1K5具有转录因子活性; 第3类为未知功能蛋白。

大部分AT-hook基序位于AT-hook蛋白的中部或前端, 少数位于AT-hook蛋白的末端。水稻最长的AT-hook蛋白有1 953个氨基酸残基, 最短的有150个氨基酸残基。这些水稻AT-hook蛋白含有数量不等的AT-hook基序。大部分水稻AT-hook蛋白含有1–6个AT-hook基序, C7J2W2则含有13个AT-hook基序。每个水稻AT-hook蛋白除了含有AT-hook基序外, 还含有其它的结构域, 如WD40、Pfam:DUF296、AAA、RING、JmjC、H15、SRA和SET等。

2.3 水稻AT-hook基因家族的基因结构

水稻基因组中含有45个AT-hook基因, 它们的结构差异很大, 有的只有1个外显子, 有的则多达18个外显子。可变剪接体大多发生在5'UTR和3'UTR, 其它外显子和内含子的位置及长度则相对稳定。值得注意的是, 大多数具有多个剪接体的基因只编码1条相同的氨基酸多肽链, 各剪接体的差异在于5'UTR和3'UTR插入了不同数目和大小的内含子。例如, OsAHL38的2种剪接体编码的2条序列是完全相同的多肽链, 两剪接体的差异在于其中1个剪接体的5'UTR插入了内含子。个别基因的不同剪接体分别编码不同的多肽链。例如, OsAHL10的2种剪接体, 编码2条长度不同的多肽链, 原因在于其中1个剪接体的1个外显子缺失了一小段; OsAHL18有4种剪接体, 其中2种剪接体的5'UTR插入了长度不同的内含子, 编码2条相同的多肽链, 另外2个剪接体, 一个是倒数第3个外显子缺失了一小段, 另一个是3'UTR缺失了4个外显子, 分别编码2条不同长度的多肽链; OsAHL24有3种不同的剪接体, 一个剪接体的第2个外显子中插入了内含子, 另一个剪接体的5'UTR的第1个外显子缺失, 3个剪接体分别编码3条序列长度不同的多肽链; OsAHL28

表1 水稻AT-hook蛋白及其编码基因

Table 1 The AT-hook proteins and their encoding genes in rice

Gene name	TIGR ID	Chromosome number	Protein	Length (aa)	Molecular weight (kDa)	PI	C ^b	Molecular function/Biological process
OsAHL01	LOC_Os01g11900	1	B8AAK8	1 953	215 862	7.195	1/1	DNA binding; 4 iron, 4 sulfur cluster binding; catalytic activity; endonuclease activity/DNA repair; base-excision repair
OsAHL02	LOC_Os01g13300	1	Q0JP99	496	55 530.4	7.192	1/1	DNA binding/Transcription, transcription regulation
OsAHL03	LOC_Os01g61890	1	Q5QMH2	284	31 552.5	9.496	1/4	DNA binding/Not known
OsAHL04	LOC_Os01g63910	1	UPI0001C7EC27	1 089	118 099	7.731	2/2	DNA binding/Transcription
OsAHL05	LOC_Os02g03270	2	Q6Z701	389	38 533.2	9.380	1/1	DNA binding/Transcription
OsAHL06	LOC_Os02g29230	2	UPI0000F17712	1 637	182 578	6.981	1/1	4 iron, 4 sulfur cluster binding; catalytic activity/DNA repair; base-excision repair
OsAHL07	LOC_Os02g39920	2	Q6K624	756	81 358.1	5.858	1/1	DNA binding/Not known
OsAHL08	LOC_Os02g57820	2	Q6K9W5	395	40 028.4	7.500	1/1	DNA binding/Transcription
OsAHL09	LOC_Os03g15580	3	Q10NR0	1 251	137 603	5.682	1/1	ATP binding; DNA binding/DNA repair
OsAHL10	LOC_Os03g31594	3	Q10J83	1 057	119 359	7.717	2/2	Transcription factor activity/Not known
OsAHL11	LOC_Os03g37915	3	UPI0001C7EEC3	150	16 085.4	11.965	1/1	Not known
OsAHL12	LOC_Os03g50355	3	UPI00001C7F115	234	23 858.2	9.915	1/1	Not known
OsAHL13	LOC_Os04g11510	4	UPI00001B0F26	492	54 668.5	8.528	1/1	Not known
OsAHL14	LOC_Os04g39520	4	UPI00001B150F	714	73 236.5	8.118	1/1	DNA binding; beta1-adrenergic receptor activity; hormone activity; zinc ion binding/G-protein coupled receptor signaling pathway
OsAHL15	LOC_Os04g42320	4	Q0JBZ3	847	92 507.3	6.365	1/1	DNA binding/Not known
OsAHL16	LOC_Os04g49990	4	Q7XLZ6	380	38 240.6	9.358	1/2	DNA binding/Transcription
OsAHL17	LOC_Os04g51700	4	UPI0001C7E40F	569	63 511.6	6.217	1/1	DNA ligase (ATP) activity; ATP binding/DNA replication; DNA repair; double-strand break repair; DNA recombination; response to X-ray; response to DNA damage stimulus
OsAHL18	LOC_Os04g53460	4	Q0J9Y0	415	44 794.2	9.193	3/4	DNA binding/Not known
OsAHL19	LOC_Os04g53470	4	Q0J9X9	208	22 517.7	10.370	1/1	DNA binding/Not known
OsAHL20	LOC_Os04g56720	4	Q0J9B4	550	58 408	8.624	1/1	Protein binding; chromatin binding/Response to stress
OsAHL21	LOC_Os04g58730	4	Q0J8W7	420	42 840	10.350	1/1	DNA binding/Transcription
OsAHL22	LOC_Os05g02590	5	UPI0001C7CC1F	344	37 028.1	6.789	1/1	Transferase activity/Metabolic process; biosynthetic process; cellular process; secondary metabolic process
OsAHL23	LOC_Os05g38810	5	UPI0001C7C1F9	1 413	152 389	4.026	1/1	Not known
OsAHL24	LOC_Os05g51850	5	C7J2W2	539	55 853.1	12.213	3/3	DNA binding/Metabolic process
OsAHL25	LOC_Os06g04540	6	Q5VQ11	329	33 416.3	7.089	1/1	Not known

表1 (续) Table 1 (continued)

Gene name	TIGR ID	Chromosome number	Protein	Length (aa)	Molecular weight (kDa)	PI	C ^b	Molecular function/Biological process
OsAHL26	LOC_Os06g14412	6	UPI0001C7C303	2 799	293 777	Not known	1/1	Nucleic acid binding; DNA binding; helicase activity; ATP binding/Metabolic process
OsAHL27	LOC_Os06g22030	6	Q69T83	379	37 450.1	11.057	1/1	DNA binding/Transcription
OsAHL28	LOC_Os07g08710	7	Q8H4Z0	279	28 530.6	11.297	2/3	DNA binding/Metabolic process
OsAHL29	LOC_Os07g27140	7	UPI0000E12AE0	1 010	109 704	8.652	1/1	DNA binding/Transcription
OsAHL30	LOC_Os07g37340	7	UPI0001C7CD64	100	10 421.3	12.442	1/1	DNA binding/Nucleosome assembly
OsAHL31	LOC_Os07g37800	7	Q7XIK4	495	53 351.1	10.541	3/3	DNA binding/Not known
OsAHL32	LOC_Os07g46690	7	UPI0001C7D336	1 674	187 061	7.205	1/1	Zinc ion binding; DNA binding; metal ion binding; ice binding; protein binding/Homoiothermy; metal ion transport
OsAHL33	LOC_Os07g49380	7	Q6Z3Y5	565	61 570.8	7.359	1/2	DNA binding/Not known
OsAHL34	LOC_Os08g02490	8	Q6ZJ55	373	37 813.5	9.867	1/1	DNA binding/Transcription
OsAHL35	LOC_Os08g33190	8	Q8H3M1	472	49 247	10.542	1/1	DNA binding/Metabolic process
OsAHL36	LOC_Os08g37270	8	Q6ZJD0	397	41 570.3	9.836	1/1	Protein binding/Protein metabolic process; cellular process
OsAHL37	LOC_Os08g37345	8	Q6ZJC3	325	32 282.4	8.502	1/1	DNA binding; ice binding; ligand-dependent nuclear receptor activity/Homoiothermy; response to freezing; regulation of transcription, DNA-dependent
OsAHL38	LOC_Os08g40150	8	Q6Z8N9	355	35 262.7	10.743	1/2	DNA binding/Transcription
OsAHL39	LOC_Os09g23730	9	Q69MW7	214	21 752.7	10.632	1/1	DNA binding/Metabolic process
OsAHL40	LOC_Os09g28930	9	A3BZP1	335	34 739.4	7.344	1/1	Not known
OsAHL41	LOC_Os09g31470	9	A1A6F0	360	35 491.1	10.831	1/1	DNA binding/Transcription
OsAHL42	LOC_Os10g36810	10	Q7XCX5	1 707	188 257	7.188	1/1	DNA binding; structural constituent of cell wall; zinc ion binding; protein binding/Histone modification; transcription initiation, DNA-dependent
OsAHL43	LOC_Os10g42230	10	Q336P9	406	41 398.8	10.025	1/3	DNA binding/Transcription
OsAHL44	LOC_Os11g36030	11	Q2R2F0	406	43 945.2	8.649	1/1	DNA binding/Not known
OsAHL45	LOC_Os11g38900	11	Q2R1K5	814	89 214.7	8.677	1/1	Binding; nucleic acid binding; transcription factor activity/Not known

C^b: 分母为基因可变剪接体数, 分子为编码蛋白质种类数; length (aa), molecular weight (kDa)和PI为预测值。

C^b: The common denominator for gene variable spliceosome number, molecular for coding protein species number; length (aa), molecular weight (kDa) and PI for predictive value.

的3种剪接体也编码3条不同的多肽链,原因在于外显子分别由0位和2位内含子分隔,而且OsAHL28.2编码的多肽链不含AT-hook基序; OsAHL31的3种剪接体分别由外显子的缺失和2位内含子的插入所致,编码3种不同的多肽(图1)。从基因结构中我们发现一个有规律的现象,即AT-hook基序都是由其中较大的外显子编码。

2.4 水稻AT-hook蛋白系统进化分析

基于联配好的45条水稻AT-hook蛋白质序列和34条拟南芥AT-hook蛋白质序列,用MEGA 5.05软件,采用邻接法生成水稻AT-hook基因家族的系统进化树(图2)。根据进化树的结果,将AT-hook基因家族分为5个亚族(A-E)。水稻和拟南芥的AT-hook蛋白成员并不是单独聚类形成各自的进化支,而是每个亚族都同时含有水稻和拟南芥的AT-hook成员,每个亚族的成员具有相似的结构域和功能。例如, A亚族的A1进化支,自展值(bootstrap value)为99,水稻和拟南芥AT-hook蛋白序列都含有DUF296结构域,且都具有DNA结合的功能并参与转录;在D亚族的D3进化支中,水稻和拟南芥蛋白序列都含有H15结构域,参与代谢调控。据此,我们推测OsAHL12可能也具类似的

功能。

虽然,系统进化树分析显示,几乎每个亚族都同时含有水稻和拟南芥的AT-hook蛋白序列成员,但数量上有所不同。E亚族从数量上看E2进化支为一对一(one to one), E4进化支为二对二(two to two), E3进化支为多对多(many to many)。我们推测,同一组中数量上出现多对多(many to many)的情况可能与基因组复制有关(Cardon et al., 1999),故OsAHL01、OsAHL06和OsAHL45的进化可能与基因复制有关。

2.5 水稻AT-hook基因家族在染色体上的分布及复制情况

根据各个AT-hook基因家族成员的基因组位置信息,将水稻的45个AT-hook基因在染色体上标示出来(图3)。从图3可以看出,水稻AT-hook基因在不同染色体上呈不均匀分布,在第4号染色体上分布最多,有9个,且主要集中在下部,第12号染色体上分布为0。

利用TIGR上已有的染色体片段基因组区域复制信息,在水稻片段复制区域找到3对旁系的AT-hook同源基因。分别是位于1号染色体上的OsAHL01与2号染色体上的OsAHL06、位于1号染色体上的OsAHL04与7号染色体上的OsAHL29及位于4号染色体

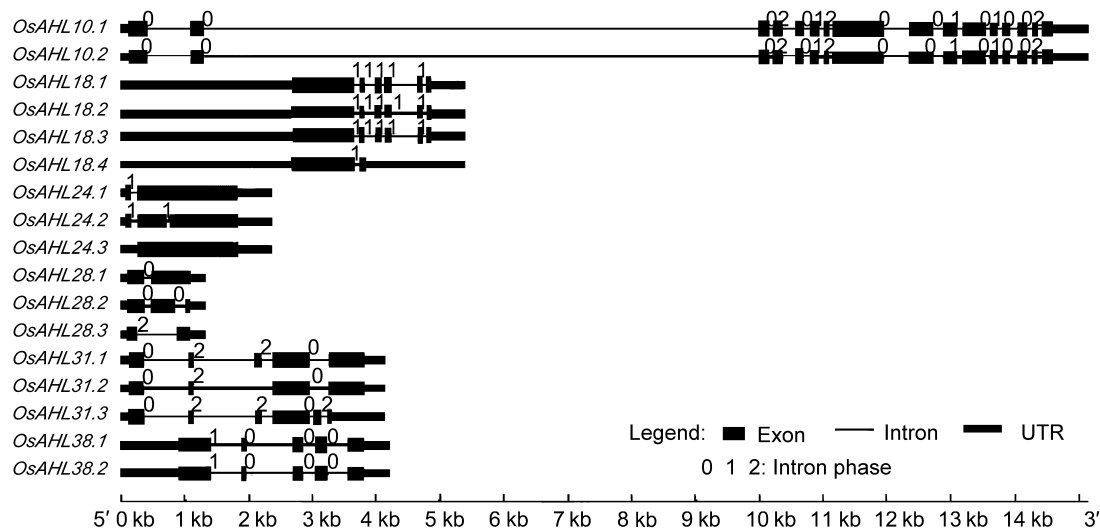


图1 多变剪接体

Figure 1 Polytopic spliceosome

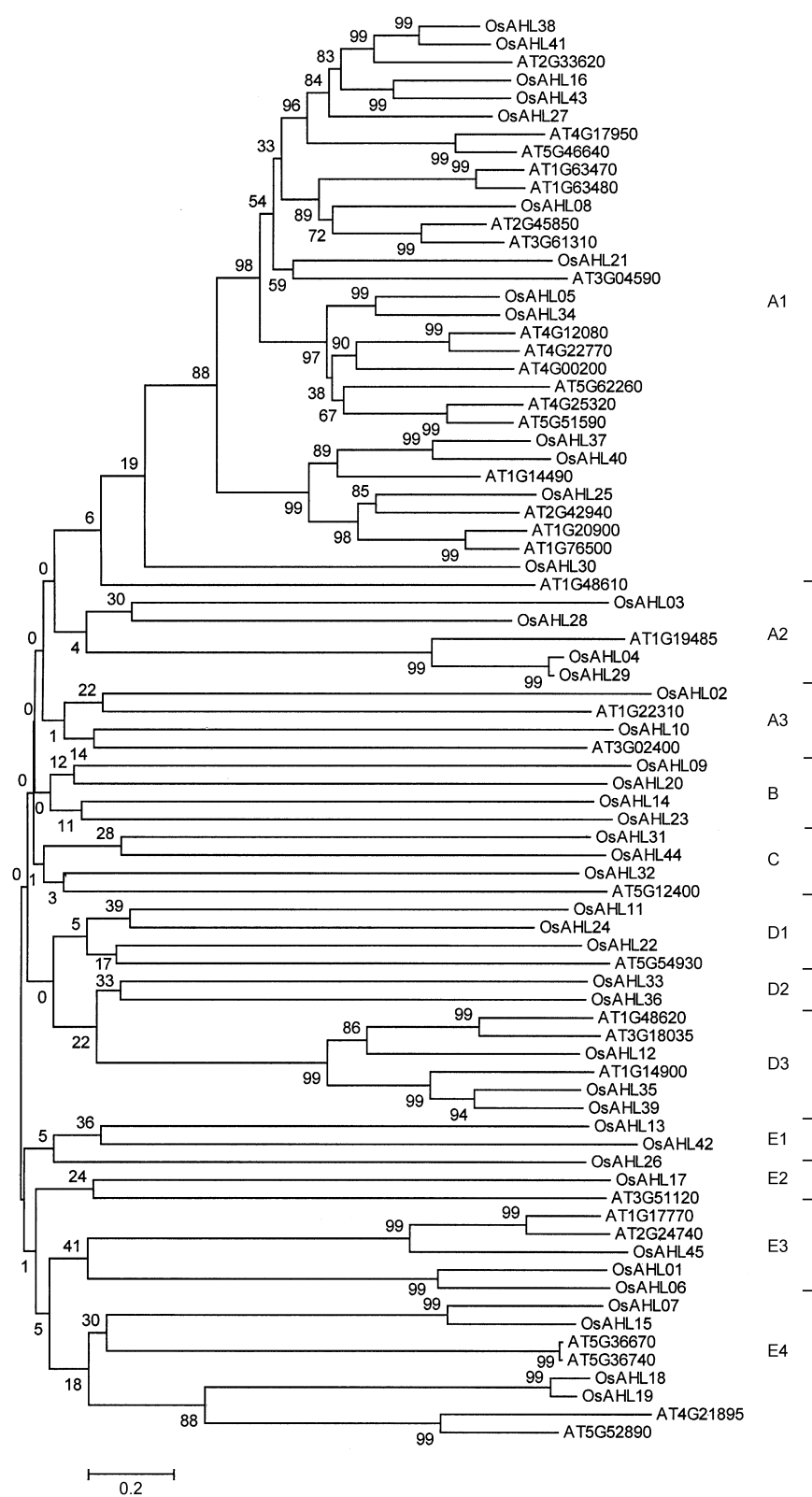


图2 水稻和拟南芥AT-hook蛋白的系统进化关系

Figure 2 Phylogenetic relationship among the *Oryza sativa* and *Arabidopsis thaliana* AT-hook proteins

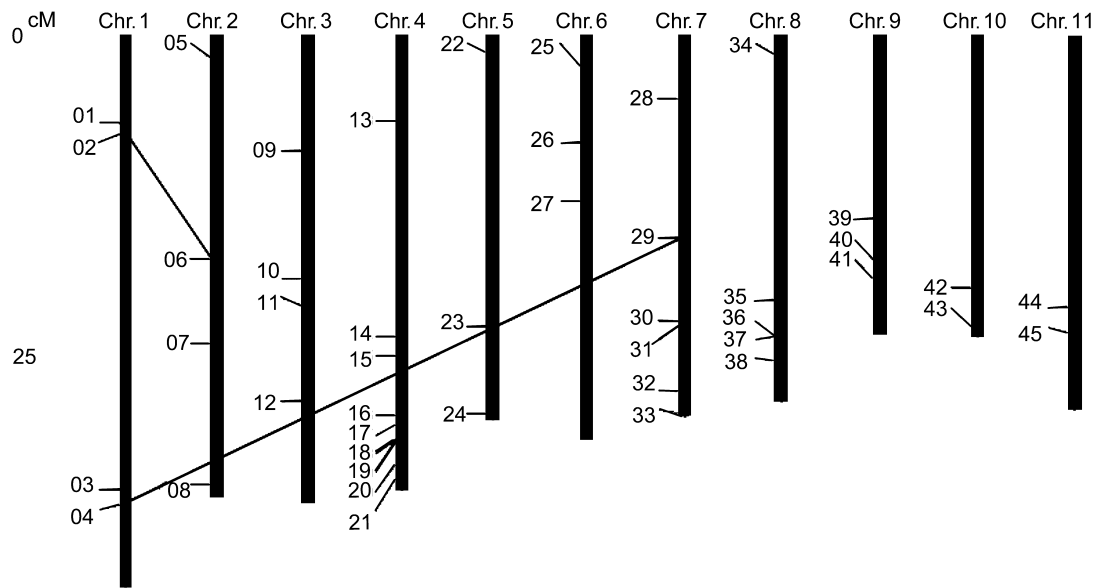


图3 水稻AT-hook基因在染色体上的分布
数字表示水稻AT-hook基因(对应表1: *OsAHL01–OsAHL45*), 直线连接标示存在于复制染色体片段上的AT-hook基因。

Figure 3 Genomic distribution of AT-hook protein-encoding genes on rice chromosomes
AT-hook genes are numbered 1 to 45 (Table 1: *OsAHL01–OsAHL45*), The AT-hook protein-encoding genes present on duplicated chromosomal segments are connected by lines.

上的*OsAHL18*与*OsAHL19*。AT-hook基因家族的基因组复制信息与系统进化树结果相吻合。

2.6 水稻AT-hook基因家族数字表达分析

从TIGR的水稻数据库中搜索AT-hook基因的ESTs序列, 未找到与*OsAHL03*、*OsAHL06*和*OsAHL11*等11个基因相匹配的EST序列。而约71%的基因有5条或5条以上的EST记录, 其中*OsAHL05*和*OsAHL36*匹配到5条ESTs, *OsAHL28*则匹配到240条ESTs, 只有*OsAHL04*和*OsAHL43*匹配到1条EST。这些ESTs序列来自不同水稻的EST数据库, 即水稻的AT-hook基因在各个组织中都有表达, 表明水稻的AT-hook基因可能参与了多种生长发育过程。从EST的组织分布来看, 在圆锥花序中水稻AT-hook基因的表达最丰富, 其次是愈伤组织。

在水稻发育芯片数据中, 除了没有与*OsAHL08*和*OsAHL18*匹配的探针号外, 其余基因都对应上了探针号。表达数据的分析结果表明, 绝大部分的基因都是在幼穗中高水平表达。而*OsAHL02*、*OsAHL13*、*OsAHL16*和*OsAHL18*的表达情况类似, 均在叶中高

水平表达; *OsAHL03*和*OsAHL06*的表达情况类似, 均在胚乳中高水平表达; *OsAHL04*则在雄蕊中高水平表达(图4A)。

根据水稻AT-hook基因的发育芯片表达谱数据, 用k均值聚类法(k=6), 对水稻AT-hook基因进行聚类分析, 将其分为6类(图4A)。总体上看, 大部分水稻AT-hook基因在幼穗发育阶段表达上调, 在幼穗期表达水平最高; 从苗期到分蘖期大部分基因维持在正常表达水平。而*OsAHL05*和*OsAHL25*表达水平偏低, 这2个基因被聚在一起, 它们共同含有另外一个结构域Pfam:DUF296。

我们通过qRT-PCR对*OsAHL05*、*OsAHL09*、*OsAHL17*和*OsAHL25*在不同组织中的表达情况进行了检测, 结果与发育芯片表达谱一致, 即上述基因在穗中的表达水平较高(图4B)。

2.7 讨论

AT-hook基序广泛存在于不同物种的DNA结合蛋白中。在小麦、玉米、水稻、拟南芥、豌豆(*Pisum sativum*)、大豆、棉花、果蝇和人等许多物种中均发

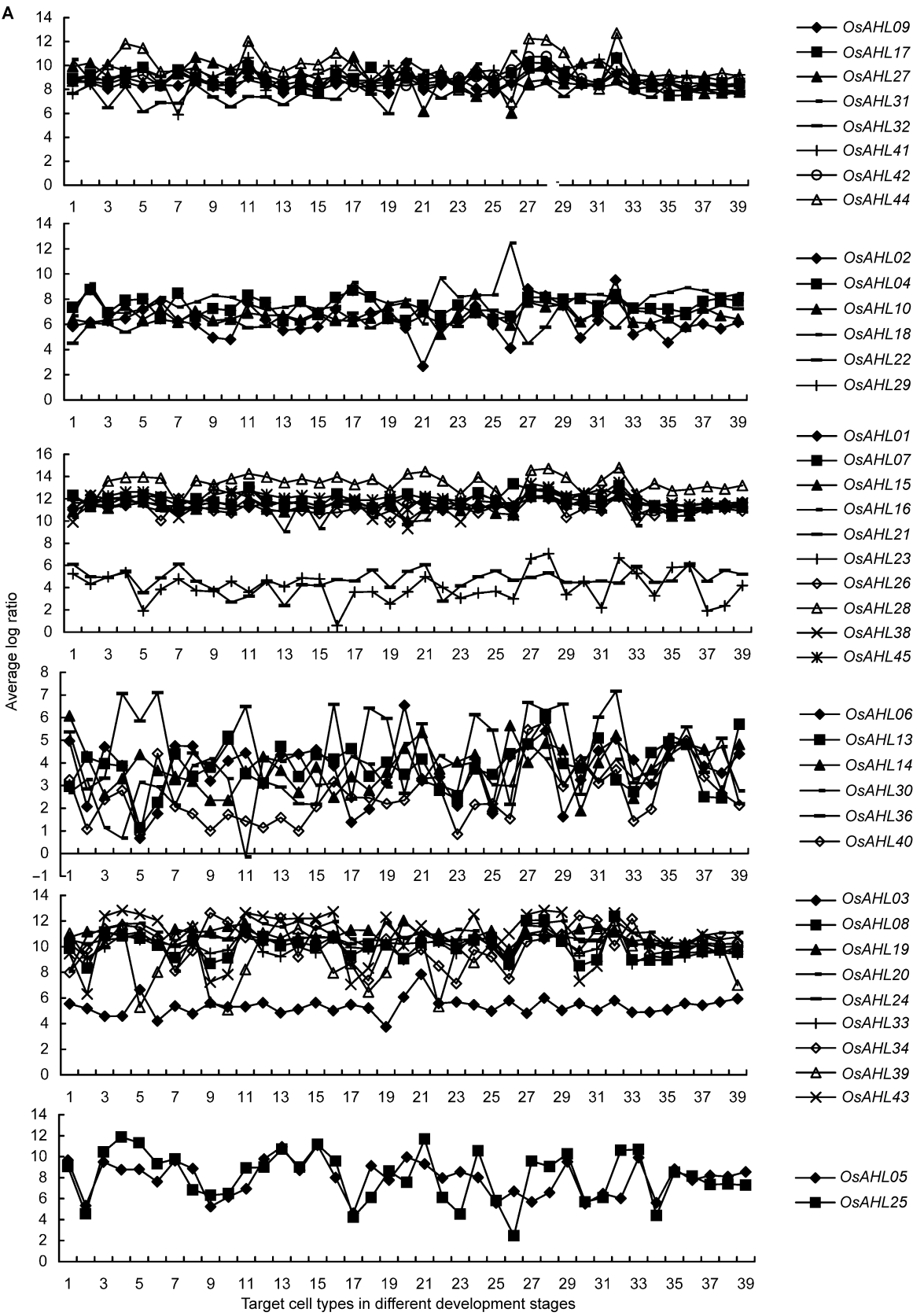


图4 Figure 4

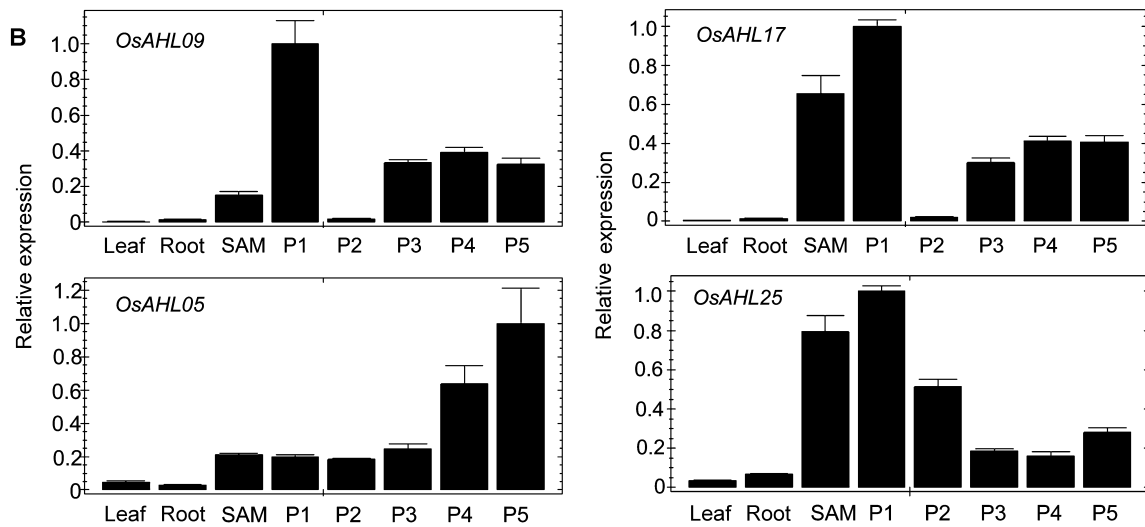


图4 水稻AT-hook基因的表达模式

(A) 基因组芯片表达k均值聚类(X轴表示水稻29个发育阶段及39个器官类型; Y轴表示以2为底数对芯片原始数据的对数值) 1: 胚乳(授粉后14天); 2: 旗叶(授粉后14天); 3: 愈伤组织(15天后, 处理2); 4: 愈伤组织(15天后, 处理3); 5: 愈伤组织(继代培养15天后); 6: 愈伤组织(感染1小时后); 7: 胚乳(授粉后21天); 8: 小花(授粉后3天); 9: 叶(幼穗生长至4–5 cm); 10: 叶鞘(幼穗生长至4–5 cm); 11: 幼穗(幼穗生长至4–5 cm); 12: 胚芽(暗处理种子萌发48小时后); 13: 胚根(暗处理种子萌发48小时后); 14: 胚芽(光照下种子萌发48小时后); 15: 胚根(光照下种子萌发48小时后); 16: 愈伤组织(再培养5天后); 17: 旗叶(抽穗期5天前); 18: 茎(抽穗期5天前); 19: 愈伤组织(感染6小时后); 20: 胚乳(授粉后7天); 21: 萌发种子(浸泡72 小时后发芽); 22: 穗(抽穗期); 23: 茎(抽穗期); 24: 愈伤组织(感染); 25: 颖壳(花期前1天); 26: 雄蕊(花期前1天); 27: 幼穗(雌蕊和雄蕊原基分化期, 阶段4); 28: 幼穗(花粉母细胞形成阶段, 阶段5); 29: 愈伤组织(筛选阶段); 30: 叶(幼穗二次枝梗原基阶段); 31: 叶鞘(幼穗二次枝梗原基阶段); 32: 幼穗(二次枝梗原基分化期, 阶段3); 33: 根(分蘖期); 34: 秧苗(分蘖期); 35: 胚和胚根(萌发3天后); 36: 根和叶(三叶期); 37: 秧苗(三叶期, GA3处理5、15、30和60分钟); 38: 秧苗(三叶期, KT处理5、15、30和60分钟); 39: 秧苗(三叶期, NAA处理5、15、30和60分钟); (B) 4个AT-hook基因的qRT-PCR表达分析(X轴表示叶、根、SAM和各发育阶段的穗; Y轴表示相对表达值) SAM: 茎尖分生组织; P1: 穗(0–3 cm); P2: 穗(3–5 cm); P3: 穗(5–10 cm); P4: 穗(10–15 cm); P5: 穗(15–22 cm)

Figure 4 Expression profiles of rice AT-hook protein-encoding genes

(A) The k-means clustering of genome chip (X-axis represents 29 developmental stages and 39 target cell types; Y-axis represents average \log_2 expression values). 1: Endosperm (14 days after pollination); 2: Flag leaf (14 days after pollination); 3: Calli (15 days after induction, treatment 2); 4: Calli (15 days after induction, treatment 3); 5: Calli (15 days after subculture); 6: Calli (1 h after infection); 7: Endosperm (21 days after pollination); 8: Spikelet (3 days after pollination); 9: Leaf (4–5 cm young panicle); 10: Sheath (4–5 cm young panicle); 11: Young panicle (4–5 cm young panicle); 12: Plumule (48 h after emergence, dark); 13: Radicle (48 h after emergence, dark); 14: Plumule (48 h after emergence, light); 15: Radicle (48 h after emergence, light); 16: Calli (5 days after regeneration); 17: Flag leaf (5 days before heading); 18: Stem (5 days before heading); 19: Calli (6 h after infection); 20: Endosperm (7 days after pollination); 21: Seed of germination (72 h after imbibition); 22: Panicle (heading stage); 23: Stem (heading stage); 24: Calli (just before infection); 25: Hull (one day before flowering); 26: Stamen (one day before flowering); 27: Young panicle (pistil and stamen primordium differentiation stage, stage 4); 28: Young panicle (pollen-mother cell formation stage, stage 5); 29: Calli (screening stage); 30: Leaf (young panicle at secondary branch primordium stage); 31: Sheath (young panicle at secondary branch primordium stage); 32: Young panicle (secondary-branch primordium differentiation stage, stage 3); 33: Root (seedling with two tillers); 34: Shoot (seedling with two tillers); 35: Embryo and radical (stage of 3 days after germination); 36: Root and leaf (three-leaf stage); 37: Seedling (trefoil stage, mixture of 5, 15, 30, 60 min after treating with GA3); 38: Seedling (trefoil stage, mixture of 5, 15, 30, 60 min after treating with KT); 39: Seedling (trefoil stage, mixture of 5, 15, 30, 60 min after treating with NAA); (B) qRT-PCR analysis of four AT-hook genes (X-axis represents leaf, root, SAM and the panicle of different development stages; Y-axis represents relative expression values) SAM: Shoot apical meristem; P1: 0–3 cm panicle; P2: 3–5 cm panicle; P3: 5–10 cm panicle; P4: 10–15 cm panicle; P5: 15–22 cm panicle

现了含有AT-hook基序的DNA结合蛋白。AT-hook蛋白在染色质结构组装和对目标基因转录活性的调控中起着重要作用(肖朝文和傅永福, 2009)。我们的分析结果(表1)表明, 水稻AT-hook蛋白主要属于DNA结合蛋白, 功能分化不显著, 除去未知功能蛋白外, 只有OsAHL22和OsAHL36蛋白分化出不同功能。并且绝大部分蛋白除AT-hook基序外, 还含有其它具结合功能的结构域, 如Pfam:DUF296(其结构域中包含AT-hook基序)、H15、SRA、RING和SET等。

水稻基因组中45个编码AT-hook蛋白的基因可划分成5个亚族, 存在一定的分化。绝大多数在进化树中属于同一组的AT-hook基因, 其编码的蛋白质, 在结构或功能上具有非常高的相似性。水稻AT-hook基因在染色体上呈随机不均匀分布, 这与许多基因在染色体上呈不均匀分布相似。如日本晴中468个抗稻瘟病的NBS-LRR基因(Shang et al., 2009)、687个水稻F-box基因(Jain et al., 2007)及20个水稻SBP基因(Wang et al., 2011)等在12条染色体上也呈不均匀分布。其中SBP基因在水稻第10号和第12号染色体上的分布均为0。

基因复制主要包括片段复制、随机复制和反转录3种形式。植物多倍化使基因组中保留了大量复制的染色体片段, 由染色体片段复制造成的基因复制在基因组中最为常见(Adams and Wendel, 2005)。我们的分析发现, 水稻基因组中编码AT-hook蛋白的基因通过复制得到扩增。3个复制基因片段分别位于A和E亚族的同一分支内。无内含子的OsAHL19基因可能是在后来的进化过程中丢失了内含子。

AT-hook基因在水稻各发育时期的组织器官中都有表达, 幼穗中表达水平最高, 暗示水稻AT-hook基因很可能广泛参与水稻的生长发育过程, 是一类重要的功能基因。OsAHL09和OsAHL17聚在同一组, 表达趋势也相似(图4A, B), 暗示它们可能具有相似的功能。OsAHL05和OsAHL25聚在同一组, 但在穗发育阶段其表达趋势完全相反(图4B), 表明它们的功能存在一定的再分化, 在水稻幼穗发育的不同阶段, 各司其职, 协同调控幼穗的发育。

水稻是重要的粮食作物, 水稻穗的发育状况直接影响稻谷的产量。AT-hook基因在水稻各发育时期的组织器官中都有表达且在幼穗中表达水平最高, 说明AT-hook基因在调节水稻生长发育(尤其是穗部发育)

中发挥重要作用。研究这一基因家族中各成员在水稻生长发育中的功能及其作用的遗传网络, 对水稻生长发育的遗传调控和农业生产均具十分重要的意义。然而, 有关水稻中AT-hook基因的功能研究报道较少, OsAHL25(PAL1/DP1)是目前水稻中被鉴定和克隆的唯一AT-hook基因(Luo et al., 2005; Jin et al., 2011)。DP1通过调节AP1类基因(OsMADS1、OsMADS6和OsMADS17)和E类基因(OsMADS15)的表达, 影响花分生组织的发育(Jin et al., 2011)。唐雄杰等(2011)的研究表明, DP1在水稻所有器官中均有表达, 但在水稻的发育内稃中表达水平最高, 是水稻内稃正常发育所必需的。本文对水稻AT-hook基因家族进行了较系统的生物信息学分析, 并对其在水稻生长发育中的功能进行了预测和初步探讨, 为后续相关研究奠定了基础。

参考文献

- 郭安源, 朱其慧, 陈新, 罗静初 (2007). GSDS: 基因结构显示系统. 遗传 29, 1023–1026.
- 黄志刚, 胡眺, 赵玄之, 王若仲, 萧浪涛 (2008). 超级杂交水稻TIR1类似基因cDNA的克隆与生物信息学分析. 植物学通报 25, 695–700.
- 唐雄杰, 郭维, 王云月, 刘林, 李成云, 罗琼 (2011). 水稻内稃发育相关基因PAL1时空表达研究. 云南农业大学学报 26, 149–155.
- 肖朝文, 傅永福 (2009). AT-Hook蛋白的研究进展. 中国农业科技导报 11(5), 12–16.
- Adams KL, Wendel JF (2005). Polyploidy and genome evolution in plants. *Curr Opin Plant Biol* 8, 135–141.
- Aravind L, Landsman D (1998). AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res* 26, 4413–4421.
- Ashar HR, Fejzo MS, Tkachenko A, Zhou X, Fletcher JA, Weremowicz S, Morton CC, Chada K (1995). Disruption of the architectural factor HMGI-C: DNA-binding AT-hook motifs fused in lipemas to distinct transcriptional regulatory domains. *Cell* 82, 57–65.
- Burian J, Ramón-García S, Howes CG, Thompson CJ (2012a). WhiB7, a transcriptional activator that coordinates physiology with intrinsic drug resistance in *Mycobacterium tuberculosis*. *Expert Rev Anti Infect Ther* 10, 1037–1047.
- Burian J, Ramón-García S, Sweet G, Gómez-Velasco A,

- Av-Gay Y, Thompson CJ** (2012b). The mycobacterial transcriptional regulator whiB7 gene links redox homeostasis and intrinsic antibiotic resistance. *J Biol Chem* **287**, 299–310.
- Cardon G, Höhmann S, Klein J, Nettesheim K, Saedler H, Huijser P** (1999). Molecular characterisation of the Arabidopsis SBP-box genes. *Gene* **237**, 91–104.
- Churchill ME, Travers A** (1991). A protein motifs that recognize structural features of DNA. *Trends Biochem Sci* **16**, 92–97.
- Delaney SK, Orford SJ, Martin-Harris M, Timmis JN** (2007). The fiber specificity of the cotton *FSltp4* gene promoter is regulated by an AT-rich promoter region and the AT-hook transcription factor GhAT1. *Plant Cell Physiol* **48**, 1426–1437.
- Do HJ, Song H, Yang HM, Kim DK, Kim NH, Kim JH, Cha KY, Chung HM** (2006). Identification of multiple nuclear localization signals in murine Elf3, an ETS transcription factor. *FEBS Lett* **580**, 1865–1871.
- Duncker BP, Chesnokov IN, McConkey BJ** (2009). The origin recognition complex protein family. *Genome Biol* **10**, 214–214.
- Eddy Sean R** (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755–763.
- Elías-Arnanz M, Padmanabhan S, Murillo FJ** (2010). The regulatory action of the myxobacterial CarD/CarG complex: a bacterial enhanceosome? *FEMS Microbiol Rev* **34**, 764–778.
- Fedele M, Battista S, Manfioletti G, Croce CM, Giancotti V, Fusco A** (2001). Role of the high mobility group A proteins in human lipomas. *Carcinogenesis* **22**, 1583–1591.
- Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A** (2006). Pfam: clans, web tools and services. *Nucleic Acids Res* **34**, D247–D251.
- Fujimoto S, Matsunaga S, Yonemura M, Uchiyama S, Azuma T, Fukui K** (2004). Identification of a novel plant MAR DNA binding protein localized on chromosomal surfaces. *Plant Mol Biol* **56**, 225–239.
- Gallavotti A, Malcomber S, Gaines C, Stanfield S, Whipple C, Kellogg E, Schmidt RJ** (2011). BARREN STALK FASTIGIATE1 is an AT-Hook protein required for the formation of maize ears. *Plant Cell* **23**, 1756–1771.
- Girard F, Bello B, Laemmli UK, Gehring WJ** (1998). *In vivo* analysis of scaffold-associated regions in *Drosophila*: a synthetic high-affinity SAR binding protein suppresses position effect variegation. *EMBO J* **17**, 2079–2085.
- Gordon BR, Li Y, Wang L, Sintsova A, Bakel HV, Tian S, Navarre WW, Xia B, Liu J** (2010). Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* **107**, 5154–5159.
- Jacobsen K, Laumen NB, Jansen EO, Marcker A, Poulsen C, Marcker KA** (1990). HMG I-like proteins from leaf and nodule nuclei interact with different AT motifs in soybean nedulin promoters. *Plant Cell* **2**, 85–94.
- Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, Kapoor S, Tyagi AK, Khurana JP** (2007). F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol* **143**, 1467–1483.
- Jin Y, Lou Q, Tong HN, Wang AJ, Cheng ZJ, Tang JF, Li DY, Zhao XF, Li XB, Wan JB, Jiao YL, Chu CC, Zhu LH** (2011). An AT-hook gene is required for palea formation and floral organ number control in rice. *Dev Biol* **359**, 277–288.
- Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG** (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.
- Lim PO, Kim Y, Breeze E, Koo JC, Woo HR, Ryu JS, Park DH, Beynon J, Tabrett A, Wollaston VB, Nam HG** (2007). Overexpression of a chromatin architecture-controlling AT-hook protein extends leaf longevity and increases the post-harvest storage life of plants. *Plant J* **52**, 1140–1153.
- Luo Q, Zhou K, Zhao X, Zeng Q, Xia H, Zhai W, Xu J, Wu X, Yang H, Zhu L** (2005). Identification and fine mapping of a mutant gene for palealess spikelet in rice. *Planta* **221**, 222–230.
- Macas J, Neumann P, Novák P, Jiang J** (2010). Global sequence characterization of rice centromeric satellite based on oligomer frequency analysis in large-scale sequencing data. *Bioinformatics* **26**, 2101–2108.
- Martínez-García JF, Quail PH** (1999). The HMG-I/Y protein PF1 stimulates binding of the transcriptional activator GT-2 to the PHYA gene promoter. *Plant J* **18**, 173–183.
- Matsushita A, Furumoto T, Ishida S, Takahashi Y** (2007). AGF1, an AT-hook protein, is necessary for the negative feedback of AtGA3ox1 encoding GA 3-oxidase. *Plant*

- Physiol* **143**, 1152–1162.
- Metcalf CE, Wassarman DA** (2006). DNA binding properties of TAF1 isoforms with two AT-hooks. *J Biol Chem* **281**, 30015–30023.
- Morisawa G, Han-Yama A, Moda I, Tamai A, Iwabuchi M, Meshi T** (2000). AHM1, a novel type of nuclear matrix-localized, MAR binding protein with a single AT-hook and a J domain-homologous region. *Plant Cell* **12**, 1903–1916.
- Ng KH, Yu H, Ito T** (2009). AGAMOUS controls G1ANT KILLER, a multifunctional chromatin modifier in reproductive organ patterning and differentiation. *PLoS Biol* **7**, e1000251.
- Ouyang S, Zhu W, Hamilton J, Lin HN, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR** (2007). The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res* **35**, D883–D887.
- Pierantoni GM, Finelli P, Valtorta E, Giardino D, Rodeschini O, Esposito F, Losa M, Fusco A, Larizza L** (2005). High-mobility group A2 gene expression is frequently induced in non-functioning pituitary adenomas (NFPAs), even in the absence of chromosome 12 polyploidy. *Endocr Relat Cancer* **12**, 867–874.
- Reddy MC, Christensen J, Vasquez KM** (2005). Interplay between human high mobility group protein 1 and replication protein A on psoralen-cross-linked DNA. *Biochemistry* **44**, 4188–4195.
- Shaik R, Ramakrishna W** (2012). Bioinformatic analysis of epigenetic and MicroRNA mediated regulation of drought responsive genes in rice. *PLoS One* **7**, e49331.
- Shang J, Tao Y, Chen X, Zou Y, Lei C, Wang J, Li X, Zhao X, Zhang M, Lu Z, Xu J, Cheng Z, Wan J, Zhu L** (2009). Identification of a new rice blast resistance gene, Pid3, by genomewide comparison of paired nucleotide-binding site-leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. *Genetics* **182**, 1303–1311.
- Shun MC, Botbol Y, Li X, Nunzio FD, Daigle JE, Yan N, Lieberman J, Lavigne M, Engelman A** (2008). Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. *J Virol* **82**, 11555–11567.
- Street IH, Shah PK, Smith AM, Avery N, Neff MM** (2008). The AT-hook-containing proteins SOB3/AHL29 and ESC/AHL27 are negative modulators of hypocotyl growth in Arabidopsis. *Plant J* **54**, 1–14.
- Vom Endt D, e Silva MS, Kijne JW, Pasquali G, Memelink J** (2007). Identification of a bipartite jasmonate-responsive promoter element in the *Catharanthus roseus* ORCA3 transcription factor gene that interacts specifically with AT-Hook DNA-binding proteins. *Plant Physiol* **144**, 1680–1689.
- Wang Y, Hu ZL, Yang YX, Chen XQ, Chen GP** (2011). Bioinformatics analysis of SBP-box gene family in rice. *J Bioinformatics* **9**, 82–87.
- Webster CI, Packman LC, Gray JC** (2001). HMG-1 enhances HMG-1/Y binding to an A/T-rich enhancer element from the pea plastocyanin gene. *Eur J Biochem* **268**, 3154–3162.
- Yuan Q, Ouyang S, Liu J, Suh B, Cheung F, Sultana R, Lee D, Quackenbush J, Buell CR** (2003). The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. *Nucleic Acids Res* **31**, 229–233.

Bioinformatics Analysis of the AT-hook Gene Family in Rice

Guiwei Zhang^{1†}, Jue Zeng^{1†}, Wei Guo¹, Qiong Luo^{2*}

¹College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming 650201, China

²College of Plant Protection, Yunnan Agricultural University, Kunming 650201, China

Abstract The AT-hook is a small motif in DNA-binding protein that was first described in the high-mobility-group non-histone chromosomal protein HMG-I/Y. AT-hook family proteins play an important role in chromatin structure assembly, specific binding to target cells, transcription regulation and development regulation in other organisms. We identified 45 AT-hook genes from the rice genome by bioinformatics, analyzed the gene structure and chromosomal distribution, and examined the phylogenetic relationships among rice AT-hook proteins, finally constructing an unrooted tree from alignment of their full-length protein sequences in rice and Arabidopsis. AT-hook proteins grouped into 5 distinct clades (A–E), with no differences in structure or characteristics. The family members were recruited by chromosome replication. Digital gene expression analysis revealed that AT-hook genes mainly express in young panicles. We verified some of the gene expression by qRT-PCR.

Key words AT-hook gene, bioinformatics, gene expression, rice

Zhang GW, Zeng J, Guo W, Luo Q (2014). Bioinformatics analysis of the AT-hook gene family in rice. *Chin Bull Bot* **49**, 49–62.

† These authors contributed equally to this paper.

* Author for correspondence. E-mail: qiongbf@aliyun.com

(责任编辑: 孙冬花)