

水稻新型潜在冷调节蛋白的生物信息学预测与分析

王鹏飞^{1,2}, 任相亮², 陈曦³, 王荣富^{1*}, 马飞^{2*}

(1. 安徽农业大学生命科学学院, 合肥 230036; 2. 南京师范大学生命科学学院, 南京 210046;
3. 辽宁师范大学生命科学学院, 大连 116029)

摘要: 冷调节蛋白是一类抗寒相关蛋白, 其种类十分广泛。几种水稻冷调节蛋白也曾在水稻中被发现, 但是目前深入了解的水稻冷调节蛋白种类较少。作者运用生物信息学方法, 用已知的冷调节蛋白作为电子探针, 预测出 16 种水稻新型潜在冷调节蛋白, 并且使用 NCBI-conserved domains、SWISS-MODEL 和 ExPASy ProtParam tool 等在线软件预测它们的功能结构域和二、三级结构以及氨基酸组成, 对其作为冷调节蛋白潜在家族成员的真实性、可靠性做出理论上的判断。分析表明, 这 16 种水稻新型潜在冷调节蛋白, 符合冷调节蛋白的一些基本的结构特征, 具有其他冷调节蛋白相同的功能结构域, 为这 16 种蛋白属于水稻冷调节蛋白家族提供了一些理论依据。

关键词: 冷害; 冷调节蛋白; 生物信息学; 水稻

中图分类号: Q811.4; S511

文献标识码: A

文章编号: 1672-352X (2011)01-0006-08

Prediction and analysis for new potential rice cold regulated proteins based on bioinformatic method

WANG Peng-fei^{1,2}, REN Xiang-liang², CHEN Xi³, WANG Rong-fu¹, MA Fei²

(1. Life Science College, Anhui Agricultural University, Hefei 230036;

2. Life Science College, Nanjing Normal University, Nanjing 210046;

3. Life Science College, Liaoning Normal University, Dalian 116029)

Abstract: Cold regulated protein is a kind of protein associated with cold resistant ability of plants. There are many types of cold regulated proteins in plant. However, very few cold regulated proteins from rice have been found recently. We have predicted 16 new potential cold regulated proteins in rice by bioinformatics methods as well as using the cold regulated proteins which have been found as probes. And then the structural domains, secondary structures, tertiary structures and amino acid compositions of these above proteins have also been predicted using these online programs, such as NCBI-conserved domains, SWISS-MODEL and ExPASy ProtParam tool. The bioinformatical analysis results show that the 16 new potential rice cold regulated proteins and other cold regulated proteins have the same basic structural characteristics and structural domains, implying that the 16 new proteins might belong to rice cold regulated protein family.

Key words: chilling injury; cold regulated protein; bioinformatics; rice

冷害的发生范围在世界上分布很广, 各种类型冷害的出现, 往往造成粮食大幅度减产^[1]。随着世界粮食危机的爆发和全球气候异常等问题的出现, 冷害问题正越来越受到各国政府的重视。Weiser 在 1970 年提出, 冷诱导能调节植物基因的表达, 从而产生具有抗寒性或冷适应性的蛋白^[2]。Guy 早在

1985 就在菠菜中首次发现了这种蛋白^[3], 但直到 1990 年, Hajela 等人才真正提出了冷调节蛋白(cold regulated proteins, CORPs)的概念^[4]。冷调节蛋白是指在冷驯化下诱导表达的一种特异性蛋白, 它具有调节植物细胞冷适应性与渗透压, 增强植物抗寒能力, 避免温度低于 0℃时植物细胞内部冰晶产生^[5],

收稿日期: 2010-06-21

基金项目: 国家转基因作物新品种培育重大专项(2009ZX08009-063B)和安徽省教育厅自然科学基金重大项目(KJ2010ZD04)共同资助。

作者简介: 王鹏飞, 男, 硕士研究生。

* 通讯作者: 王荣富, 男, 博士, 教授, 博士生导师。马飞, 男, 博士, 教授, 博士生导师。E-mail: feima1968@gmail.com

保证植物在低温环境下维持正常生理活动的功能。除了冷诱导之外, 盐胁迫处理、热诱导处理、ABA 处理等逆境胁迫, 均能诱导植物冷调节蛋白的产生^[6-8]。1999 年, Pearce 提出了冷信号转导与诱导生成冷调节蛋白过程的模型(图 1)^[9-10]。

冷调节蛋白的蛋白结构一般具有如下特征: 一级结构富含亲水性的氨基酸, 二级结构以 α -螺旋为主等。一种蛋白如果具有这些结构特点, 就决定了其具有亲水性, 流动性以及热稳定性。所以以上结构特征可以作为判断一种蛋白是否是潜在的冷调节蛋白家族成员的依据之一。但是, 冷调节蛋白是一类高级结构多元化, 种类繁多的多基因家族, 很难单凭几种简单的结构特征就能判断一种蛋白是否属于冷调节基因家族的成员, 所以新型冷调节蛋白的发掘一直以来都是一个难题。

目前, 越来越多的冷调节蛋白已经被人们发现。Breton 等根据 Danyluk 在小麦中发现的 Ta WCOR413 冷调节蛋白, 将 Ta WCOR413 进行分类并预测了水稻中的 COR413-TM1 和 COR413-PM1^[11]; Li 等在大豆中发现了 COR2 冷调节蛋白, 最近杨同文等又在大白菜中发现了一种新的冷调节蛋白 BpCOR^[12], 这些已被发现的冷调节蛋白已被实验验证能提高植物的抗寒能力。以上这些冷调节蛋白的发现都是将所研究的植物进行冷驯化后, 利用 RT-PCR 获得序列全长, 同时辅助生物信息学手段进行功能的预测, 继而预测并验证这个基因的功能。

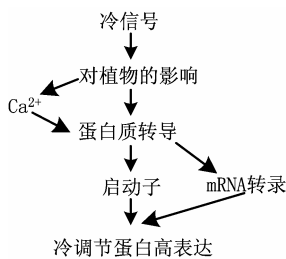


图 1 冷调节蛋白产生过程模型

Figure 1 Model of generating cold regulated proteins

水稻(*Oryza sativa*)是重要的粮食作物, 世界上近一半人口, 包括几乎整个东亚和东南亚的人口, 都以稻米为食。水稻喜高温, 对冷害敏感。因此, 预测和发掘水稻基因组中的抗寒相关基因及其表达的蛋白, 并使其过量表达或诱导其表达对水稻适应低温, 避免减产等方面具有重要的意义。作者运用生物信息学方法并结合多种分析软件, 预测了水稻中几种新型潜在冷调控蛋白, 并对这些蛋白进行了

初步的分析, 使之成为抗寒候选基因提供了一些理论依据。运用的这种预测方法(图 2)为预测、初步分析和验证水稻中新型潜在冷调节蛋白的功能提供了新的研究理念, 拓宽了研究思路。

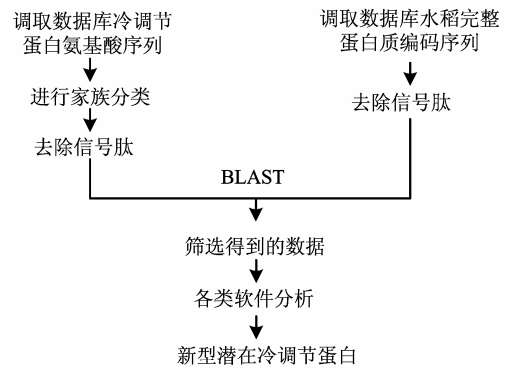


图 2 新型潜在冷调节蛋白预测流程

Figure 2 Model of predicting new potential cold regulated proteins

1 材料与方法

1.1 数据获得与分类

在 GenBank 数据库中获得所有物种中的冷调节蛋白的氨基酸序列, 将找到的这些氨基酸序列利用 MAGA3.1 进行同源建树。根据构建的系统发育树对这些冷调节蛋白进行分析并按其功能结构域的不同进行分类。

1.2 去除信号肽与探针的获得

用 SignalP 3.0 在线软件分析, 去掉冷调节蛋白的氨基酸序列的前导信号肽, 得到保守序列作为探针。

1.3 比对筛选

将得到的冷调节蛋白的探针与 GenBank 中的水稻各种 Complete CDS (完整蛋白质编码序列) 编码的氨基酸序列进行比对。剔除比对结果里的在水稻中已被发现并命名的序列, 筛选出氨基酸序列同源性达到 60% 以上, E-value 低于 $1e-20$ ^[13], 并且功能尚未被验证的蛋白质。然后将这些 Complete CDS 基因序列与水稻表达序列标签 (expressed sequence tags, ESTs) 进行比对, 寻找是否存在已被证实的可表达蛋白, 进一步验证这些基因编码的蛋白在生物体中能够表达的可能性。

1.4 软件分析判断

按不同类别分别对找到的水稻中新型冷调节蛋白进行功能域、二、三级结构(使用 NCBI- conserved domains、SWISS-MODEL 在线软件分析)、氨基酸

分子组成等分析(使用 ExPASy ProtParam tool 在线软件分析),对其功能性、亲缘关系加以初步预测和验证(使用 Clustalx1.81、MEGA3.1 进行亲缘性多序列比对分析、系统发育树分析等),并对其作为冷调节蛋白潜在家族成员的真实性、可能性进行初步预测和判断。

2 结果与分析

在 GenBank 数据库中,找到了 21 种冷调节蛋白,将这些冷调节蛋白的氨基酸序列利用邻接法构建系统发育树。该系统发育树的树枝末端分类(图 3)和不同功能结构域的补偿性突变推导,将 GenBank 中调取的冷调节蛋白分为 5 个主要小组(表 1)并拟命名为: WCOR413 superfamily 冷调节蛋白、Dehydrin superfamily 冷调节蛋白、C2 superfamily 冷调节蛋白、PEBP_RKIP superfamily 冷调节蛋白以及 Chloroa_b-bind superfamily。其中 WCOR413 superfamily 冷调节蛋白还分为类囊体膜 WCOR413 冷调节蛋白(TM 血统)和质膜 WCOR413 冷调节蛋白(PM 血统)2 个亚组。将无明确结构域的冷调节蛋白拟命名为 No putative domain 组,暂时归为一类。因 No putative domain 组的冷调节蛋白保守结构杂乱,故不对其进行分析研究。从图 3 的系统发育树可以看出这些冷调节蛋白并没有按照来源物种(包括单子叶植物和双子叶植物)聚类,说明冷调节蛋白在单、双子叶植物分化之前就存在独立的进化事件。WCOR413 superfamily 冷调节蛋白的家族成员与其他 3 组具有明确功能结构域的冷调节蛋白家族成员差别较大,可能是因其冷调节功能原理和其他 3 组冷调节蛋白的调节原理差异较大,作用的细胞区位没有重叠。

利用 SingnalP 3.0 在线分析软件,分别去掉找到的 21 种冷调节蛋白的信号肽序列,并以之作为探针,在 GenBank 的水稻数据库中初步预测出 16 种水稻的新型潜在冷调节蛋白: OsI_21013、OsJ_19572、OsJ_12721、OsI_10073、Os02g0669100、OsI_08410、OsI_07873、Os01g0934100、Os05g0370300、Os05g0149100、OsI_01954、OsJ_18292、OsJ_1712、OsJ_25519、Os05g0468800 以及 OsJ_18871,它们与探针的氨基酸序列同源性达到 60% 以上,且 E-value 低于 1e-20。

将这 16 种新型潜在冷调节蛋白的氨基酸序列放入水稻 ESTs 数据库(Gene Index-DFCI/DFCI Rice Gene Index)中进行在线 BLAST 比对,发现它们都能够比对上相应的同源性较高的水稻 ESTs,并且基

本上都能比对出功能类似蛋白(如 Cold acclimation protein、SK3-type dehydrin 1、C2 domain-containing protein-like 等)的 ESTs(表 2),进一步证明了其作为冷调节蛋白潜在家族成员的真实性、可能性。

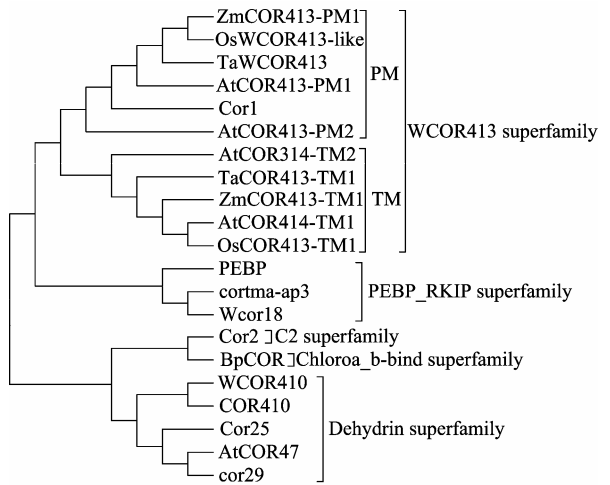


图 3 冷调节蛋白的系统发育树
Figure 3 Phylogenetic tree of cold regulated proteins

将这 16 种预测的水稻新型潜在冷调节蛋白与找到的 21 种冷调节蛋白,利用 MEGA3.1 进行系统发育树的构建,发现这 16 种预测的水稻新型潜在冷调节蛋白中与 WCOR413 superfamily 冷调节蛋白亲缘关系较为接近的为: OsI_21013、OsJ_19572、OsJ_12721 和 OsI_10073 蛋白(如图 4 所示)。其中 OsI_21013 来源于印度水稻种系(*Oryza sativa Indica Group*),其余为日本水稻种系(*Oryza sativa Japonica Group*)。与 WCOR413 superfamily 组成相同,这 4 个蛋白都具有 WCOR413 superfamily 功能结构域。从图 4 可以看出: OsJ_12721 和 OsI_10073 蛋白与质膜 WCOR413 冷调节蛋白(PM 血统)亲缘关系较接近,而 OsI_21013 和 OsJ_19572 与类囊体膜 WCOR413 冷调节蛋白(TM 血统)亲缘关系较接近。OsJ_12721 和 OsI_10073 蛋白,OsI_21013 和 OsJ_19572 蛋白两两之间的同源性高达 80% 以上,OsJ_12721 和 OsI_10073 蛋白的基因同样位于 3 号染色体,OsI_21013 和 OsJ_19572 蛋白的基因同样位于 5 号染色体,但是它们都位于基因组中不同的位置,因此排除了选择性剪接表达的可能性。但目前未知 OsJ_12721 和 OsI_10073 蛋白的基因、OsI_21013 和 OsJ_19572 蛋白的基因是否两两互为多拷贝基因,这就需要通过 northern 印迹实验进行进一步的验证。

表 1 冷调节蛋白分类
Table 1 Classification of cold regulated proteins

组别 Groups	生物体 Organism name	基因名 Gene name	蛋白质编码(GI) Protein identifier	数据库 Database
WCOR413 superfamily	<i>Arabidopsis thaliana</i>	At COR314-TM2	GI:18397074	GenBank
WCOR413 superfamily	<i>Arabidopsis thaliana</i>	At COR414-TM1	GI:18397077	GenBank
WCOR413 superfamily	<i>Oryza sativa Indica Group</i>	OsI_21013 *	GI:218197283	GenBank
WCOR413 superfamily	<i>Oryza sativa Japonica Group</i>	OsJ_19572 *	GI:222632584	GenBank
WCOR413 superfamily	<i>Oryza sativa Japonica Group</i>	OsJ_12721 *	GI:125588043	GenBank
WCOR413 superfamily	<i>Oryza sativa Indica Group</i>	OsI_10073 *	GI:125542458	GenBank
WCOR413 superfamily	<i>Oryza sativa Japonica Group</i>	Os COR413-TM1	GI:27902677	GenBank
WCOR413 superfamily	<i>Triticum aestivum</i>	Ta COR413-TM1	GI:27902669	GenBank
WCOR413 superfamily	<i>Zea mays</i>	Zm COR413-TM1	GI:162459387	GenBank
WCOR413 superfamily	<i>Triticum aestivum</i>	Ta WCOR413	GI:1657855	GenBank
WCOR413 superfamily	<i>Arabidopsis thaliana</i>	At COR413-PM1	GI:15226654	GenBank
WCOR413 superfamily	<i>Zea mays</i>	Zm COR413-PM1	GI:195636268	GenBank
WCOR413 superfamily	<i>Oryza sativa Japonica Group</i>	Os WCOR413-like protein	GI:10121845	GenBank
WCOR413 superfamily	<i>Arabidopsis thaliana</i>	At COR413-PM2	GI:15230325	GenBank
WCOR413 superfamily	<i>Glycine max</i>	Cor1	GI:213053826	GenBank
Dehydrin superfamily	<i>Arabidopsis thaliana</i>	At COR47	GI:15217937	GenBank
Dehydrin superfamily	<i>Capsella bursa-pastoris</i>	cor29	GI:47118064	GenBank
Dehydrin superfamily	<i>Brassica rapa subsp. Pekinensis</i>	Cor25	GI:33410861	GenBank
Dehydrin superfamily	<i>Triticum aestivum</i>	WCOR410	GI:1657845	GenBank
Dehydrin superfamily	<i>Oryza sativa Japonica Group</i>	Os02g0669100 *	GI:115447819	GenBank
Dehydrin superfamily	<i>Oryza sativa Indica Group</i>	OsI_08410 *	GI:125540619	GenBank
Dehydrin superfamily	<i>Oryza sativa Indica Group</i>	OsI_07873 *	GI:218191080	GenBank
Dehydrin superfamily	<i>Zea mays</i>	COR410	GI:195638184	GenBank
C2 superfamily	<i>Glycine max</i>	Cor2	GI:213053828	GenBank
C2 superfamily	<i>Oryza sativa Japonica Group</i>	Os01g0934100 *	GI:115442057	GenBank
C2 superfamily	<i>Oryza sativa Japonica Group</i>	Os05g0370300 *	GI:115463501	GenBank
C2 superfamily	<i>Oryza sativa Japonica Group</i>	Os05g0149100 *	GI:115462109	GenBank
C2 superfamily	<i>Oryza sativa Indica Group</i>	OsI_01954 *	GI:218188235	GenBank
C2 superfamily	<i>Oryza sativa Japonica Group</i>	OsJ_18292 *	GI:222631346	GenBank
C2 superfamily	<i>Oryza sativa Japonica Group</i>	OsJ_25519 *	GI:125601459	GenBank
C2 superfamily	<i>Oryza sativa Japonica Group</i>	OsJ_17126 *	GI:222630205	GenBank
PEBP_RKIP superfamily	<i>Hordeum vulgare subsp. Vulgare</i>	cortma-ap3	GI:10799810	GenBank
PEBP_RKIP superfamily	<i>Triticum aestivum</i>	Wcor18	GI:26017213	GenBank
PEBP_RKIP superfamily	<i>Oryza sativa Japonica Group</i>	OsJ_18871 *	GI:222631910	GenBank
PEBP_RKIP superfamily	<i>Oryza sativa Japonica Group</i>	Os05g0468800 *	GI:115464405	GenBank
PEBP_RKIP superfamily	<i>Arabidopsis thaliana</i>	PEBP	GI:15240946	GenBank
Chloroa_b-bind superfamily	<i>Brassica rapa subsp.</i>	BpCOR	GI:98991440	GenBank
No putative domains	<i>Arabidopsis thaliana</i>	COR78	GI:404668	GenBank
No putative domains	<i>Arabidopsis thaliana</i>	COR6.6	GI:18252153	GenBank
No putative domains	<i>Arabis pumila</i>	COR15a	GI:46452267	GenBank
No putative domains	<i>Arabidopsis thaliana</i>	cor15b	GI:456016	GenBank
No putative domains	<i>Arabidopsis thaliana</i>	COR15A	GI:15227963	GenBank
No putative domains	<i>Arabidopsis thaliana</i>	COR15B	GI:15227952	GenBank
No putative domains	<i>Brassica napus</i>	cor14	GI:42476965	GenBank

注: 基因名后面标有“*”号的为预测的水稻新型潜在冷调节蛋白。

Note: The gene name of protein with “*” is the predicted cold regulated proteins.

表2 预测蛋白的同源 ESTs (部分)

Table 2 Homologous ESTs of predicted cold regulated proteins (part)

基因名称 Gene name	EST 编号 EST id	数据库 Database
OsI_21013	rice TC328279 UP Q688W8_ORYSA (Q688W8) Cold acclimation protein	DFCI Rice Gene Index
	rice TC354362 homologue to UP Q688W8_ORYSA (Q688W8) Cold acclimation protein	DFCI Rice Gene Index
	rice BI795800 similar to GP 27902677 gb AAO24631.1 cold acclimation protein	DFCI Rice Gene Index
	rice TC295525 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	R rice TC300797 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	rice TC355443 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
OsJ_19572	rice TC328279 UP Q688W8_ORYSA (Q688W8) Cold acclimation protein	DFCI Rice Gene Index
	rice TC354362 homologue to UP Q688W8_ORYSA (Q688W8) Cold acclimation protein	DFCI Rice Gene Index
	rice TC295525 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	rice TC300797 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	rice TC355443 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	rice TC295525 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
OsJ_12721	rice TC300797 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	rice TC355443 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	rice TC328279 UP Q688W8_ORYSA (Q688W8) Cold acclimation protein	DFCI Rice Gene Index
	rice TC354362 homologue to UP Q688W8_ORYSA (Q688W8) Cold acclimation protein	DFCI Rice Gene Index
	rice BI811433 homologue to GP 10121845 gb AAG13395.1 cold acclimation protein	DFCI Rice Gene Index
	rice BI806314 homologue to GP 10121845 gb AAG13395.1 cold acclimation protein	DFCI Rice Gene Index
OsI_10073	rice TC295525 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	rice TC300797 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	rice TC355443 UP Q9FUW4_ORYSA (Q9FUW4) Cold acclimation protein	DFCI Rice Gene Index
	rice BI806314 homologue to GP 10121845 gb AAG13395.1 cold acclimation protein	DFCI Rice Gene Index
	rice TC328279 UP Q688W8_ORYSA (Q688W8) Cold acclimation protein	DFCI Rice Gene Index
	rice TC354362 homologue to UP Q688W8_ORYSA (Q688W8) Cold acclimation protein	DFCI Rice Gene Index
Os02g0669100	rice BI811433 homologue to GP 10121845 gb AAG13395.1 cold acclimation protein	DFCI Rice Gene Index
	rice TC292665 GB AAV49032.1 55274278 AY786415 SK3-type dehydrin 1	DFCI Rice Gene Index
	rice TC321959 GB AAV49032.1 55274278 AY786415 SK3-type dehydrin 1	DFCI Rice Gene Index
	rice TC326352 homologue to GB AAV49032.1 55274278 AY786415 SK3-type dehydrin 1	DFCI Rice Gene Index
	rice TC308412 homologue to GB AAV49032.1 55274278 AY786415 SK3-type dehydrin 1	DFCI Rice Gene Index
	rice TC313390 UP DH16D_ORYSA (P22913) Dehydrin Rab16D	DFCI Rice Gene Index
OsI_08410	rice TC292665 GB AAV49032.1 55274278 AY786415 SK3-type dehydrin 1	DFCI Rice Gene Index
	rice TC321959 GB AAV49032.1 55274278 AY786415 SK3-type dehydrin 1	DFCI Rice Gene Index
	rice TC308412 homologue to GB AAV49032.1 55274278 AY786415 SK3-type dehydrin 1	DFCI Rice Gene Index
	rice TC326352 homologue to GB AAV49032.1 55274278 AY786415 SK3-type dehydrin 1	DFCI Rice Gene Index
	rice TC313390 UP DH16D_ORYSA (P22913) Dehydrin Rab16D	DFCI Rice Gene Index
	rice TC323193 Oryza sativa (japonica cultivar-group) cDNA clone:001-123-C10	DFCI Rice Gene Index
Os01g0934100	rice TC297931 UP Q6K8N6_ORYSA (Q6K8N6) C2 domain-containing protein-like	DFCI Rice Gene Index
Os05g0370300	rice TC297931 UP Q6K8N6_ORYSA (Q6K8N6) C2 domain-containing protein-like	DFCI Rice Gene Index
Os05g0149100	rice TC297931 UP Q6K8N6_ORYSA (Q6K8N6) C2 domain-containing protein-like	DFCI Rice Gene Index
OsI_01954	rice TC297931 UP Q6K8N6_ORYSA (Q6K8N6) C2 domain-containing protein-like	DFCI Rice Gene Index
OsJ_18292	rice TC339518 UP Q8W0F9_ORYSA (Q8W0F9) C2 domain-containing protein-like	DFCI Rice Gene Index
	rice TC339518 UP Q8W0F9_ORYSA (Q8W0F9) C2 domain-containing protein-like	DFCI Rice Gene Index
OsJ_25519	rice TC339518 UP Q8W0F9_ORYSA (Q8W0F9) C2 domain-containing protein-like	DFCI Rice Gene Index
OsJ_17126	rice TC297931 UP Q6K8N6_ORYSA (Q6K8N6) C2 domain-containing protein-like	DFCI Rice Gene Index
	rice TC339518 UP Q8W0F9_ORYSA (Q8W0F9) C2 domain-containing protein-like	DFCI Rice Gene Index
OsJ_18871	rice TC310084 similar to UP Q8H0B8_WHEAT (Q8H0B8) Cold regulated protein	DFCI Rice Gene Index
	rice TC312955 similar to UP Q8H0B8_WHEAT (Q8H0B8) Cold regulated protein	DFCI Rice Gene Index
	rice TC356499 similar to UP Q8H0B8_WHEAT (Q8H0B8) Cold regulated protein	DFCI Rice Gene Index
Os05g0468800	rice TC310084 similar to UP Q8H0B8_WHEAT (Q8H0B8) Cold regulated protein	DFCI Rice Gene Index
	rice TC312955 similar to UP Q8H0B8_WHEAT (Q8H0B8) Cold regulated protein	DFCI Rice Gene Index

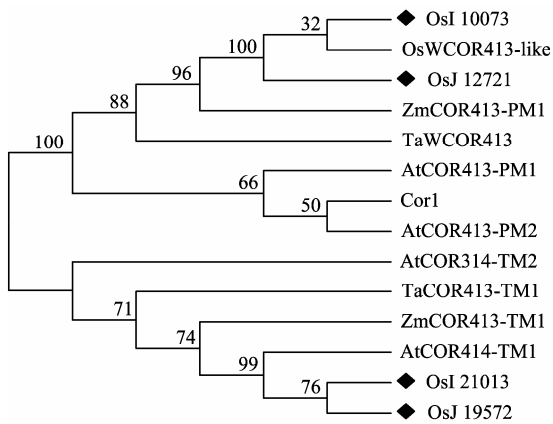


图 4 WCOR413 superfamily 冷调节蛋白氨基酸序列系统发育树

Figure 4 Phylogenetic tree of WCOR413 superfamily cold regulated proteins

与 Dehydrin superfamily 冷调节蛋白关系亲缘比较接近的为: Os02g0669100、OsI_08410 和 OsI_07873 蛋白(表 1)。其中 Os02g0669100 来源于日本水稻种,其余为印度水稻种系。这 3 个蛋白都具有 Dehydrin superfamily 功能结构域。Dehydrin superfamily 功能结构域又称脱水素功能结构域。含

有脱水素功能结构域的 Dehydrin superfamily 冷调节蛋白,其部分功能类似于一种被称为脱水素的胚胎发育后期丰富蛋白(LEA, late embryogenesis abundant proteins),具有高度水合能力,能与膜脂结合,防止细胞水分过度流失,维持膜结构水合体系的稳定性,避免了低温环境下膜融合以及膜脂相变导致的植物细胞死亡^[14]。除含有 Dehydrin superfamily 功能结构域外这一特征之外,有的 Dehydrin superfamily 冷调节蛋白还特有一段成串的丝氨酸区,又称 S 片段(S segment),共有序列为 SSSSSSSS(SS)。而这段特有的 S 片段可以通过磷酸化,使得 Dehydrin superfamily 冷调节蛋白在信号肽的引导下进入细胞核^[15]。而 Os02g0669100、OsI_08410 和 OsI_07873 这 3 种蛋白中,只有 Os02g0669100 和 OsI_08410 蛋白具有 S 片段(图 5),但是它们都被预测出有核定位信号序列(由 PredictNLS server 预测分析)。由于蛋白的入核机制十分复杂,不单单取决于蛋白的磷酸化水平^[16],并且核定位信号的预测概率不是 100%,所以关于它们是否都可以入核,在核内是否都具有功能,具体有哪些功能等疑问还需要进一步的研究。

```

Os02g0669100    MEDERNTEHQGGEEAEQVEVKDRGLFDNLLGRKKDDQPEEKKHEEELVTGMEKVSVEEP 60
OsI_08410      MEDERNTEHQGGEEAEQVEVKDRGLFDNLLGRKKDDQPEEKKHEEELVTGMEKVSVEEP 60
                *****

Os02g0669100    KKEEHHAEGEKESLLSKLHRSSSSSSSS--DEEEEVIDDNGEVVKKRKKKGLKEIKE 118
OsI_08410      KKEEHHAEGEKESLLSKLHRSSSSSSSSSDEEEEVIDDNGEVVKKRKKKGLKEIKE 120
                *****

```

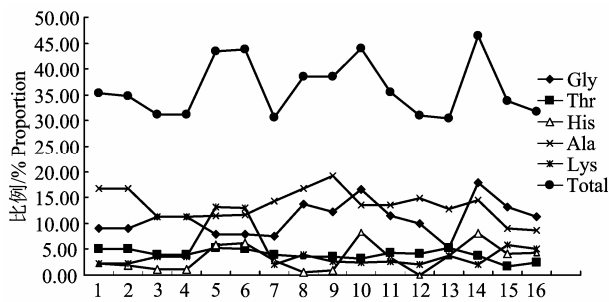
图 5 黑色标记部位为 Os02g0669100 和 OsI_08410 的 S 片段区域
Figure 5 The S segment of Os02g0669100 and OsI_08410 are blackened

与 C2 superfamily 冷调节蛋白亲缘关系比较接近的有: Os01g0934100、Os05g0370300、Os05g0149100、OsI_01954、OsJ_18292、OsJ_17126 和 OsJ_25519 等 7 种蛋白质(表 1)。这 7 种蛋白都具有 C2 superfamily 功能结构域。

与 PEBP_RKIP superfamily 冷调节蛋白亲缘关系比较接近的有: Os05g0468800 与 OsJ_18871 两种蛋白(表 1)。它们都具有 PEBP_RKIP superfamily 功能结构域。PEBP_RKIP superfamily 冷调节蛋白又称为磷脂酰乙醇胺结合蛋白(PEBP, phosphatidylethanolamine-binding family protein)。膜脂中磷脂含量,与植物的抗寒性密切相关(如磷脂酰乙醇

胺),而磷脂酰乙醇胺结合蛋白(PEBP_RKIP superfamily 冷调节蛋白)是调节膜冷稳定性的膜蛋白中的一种。

作者预测的 16 种水稻新型潜在冷调节蛋白的蛋白质结构符合已知冷调节蛋白的一般特征:一级结构富含亲水性的氨基酸(由 ExPASy ProtParam tool 在线软件分析),与小麦 COR39 冷调节蛋白、拟南芥 COR15 冷调节蛋白^[17]等氨基酸比例相似,都富含甘氨酸(Gly)、苏氨酸(Thr)、组氨酸(His)、丙氨酸(Ala)、赖氨酸(Lys)。在这些被预测的蛋白质的一级结构中,这 5 种氨基酸的总含量均大于 30%(图 6)。



1. OsI_21013; 2. OsJ_19572; 3. OsJ_12721; 4. OsI_10073; 5. Os02g669100; 6. OsI_08410; 7. OsI_07873; 8. Os01g0934100; 9. Os05g0370300; 10. Os05g0149100; 11. OsI_01954; 12. OsJ_18292; 13. OsJ_25519; 14. OsJ_17126; 15. OsJ_18871; 16. Os05g0468800

图 6 16 种水稻新型潜在冷调控蛋白的 5 种氨基酸含量以及总含量比例

Figure 6 Single and total proportion of 5 kinds of amino acids in the 16 new potential cold regulated proteins

预测的 16 种水稻新型潜在冷调节蛋白的 3 种主要的二级结构 (α -螺旋、 β -转角、 β -折叠) 中, α -螺旋占的比例最高, 如图 7 所示, 并且其共同的特征就是 β -折叠的含量都为 0.00% (由 SOPMA SECONDARY STRUCTURE PREDICTION METHOD 在线软件分析)。

利用 SWISS-MODEL 预测这 16 种水稻新型潜在冷调节蛋白的三级结构也与它们各自家族的其他成员相似, 例如: Os05g0468800 与 Wcor18 的三级结构就十分相似 (图 8)。

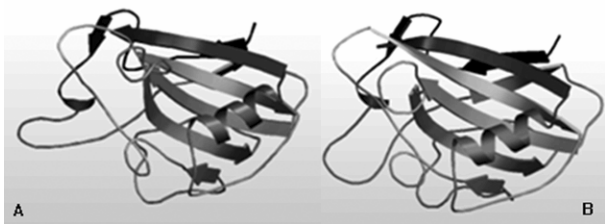
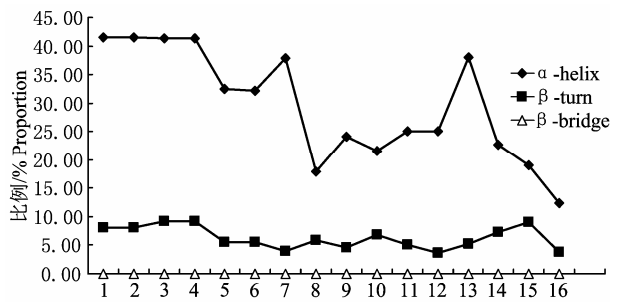


图 8 A、B 分别为 Os05g0468800 与 Wcor18 蛋白质三级结构
Figure 8 A shows the protein tertiary structure of Os05g0468800 and B shows the protein tertiary structure of Wcor18

3 讨论

作者利用生物信息学方法和一些生物学分析软件, 成功预测出 16 种水稻新型潜在冷调节蛋白, 分别与其探针蛋白同源性都高于 80%, 无论是他们的二、三级结构, 还是亲水氨基酸含量都符合已公布的冷调节蛋白的一般规律, 有着与各自亲缘关系较近、同源性较高的冷调节蛋白相同的结构域以及特征片段, 理论上属于冷调节蛋白家族成员的可能性



1. OsI_21013; 2. OsJ_19572; 3. OsJ_12721; 4. OsI_10073; 5. Os02g669100; 6. OsI_08410; 7. OsI_07873; 8. Os01g0934100; 9. Os05g0370300; 10. Os05g0149100; 11. OsI_01954; 12. OsJ_18292; 13. OsJ_25519; 14. OsJ_17126; 15. OsJ_18871; 16. Os05g0468800

图 7 16 种水稻新型潜在冷调控蛋白的 3 种主要二级结构所占的比例

Figure 7 Proportion of 3 kinds of secondary structures in the 16 new potential cold regulated proteins

较大, 极有可能是新类型水稻冷调节蛋白的候选基因, 但不排除假阳性的存在, 有待进一步研究讨论。

作者运用的这种预测方法, 基于传统方法的成果, 不需要 EST 片段拼接、兼并引物克隆、制备基因芯片、RACE 法全长扩增等生物信息及实验方法, 即可找到蛋白质编码序列, 较之传统方法节省了一定的时间和成本。这种预测方法能将前人的大量冗长的实验结果更为有效的回收和利用, 并且更加便于后期对这类冷调节蛋白基因进行进一步分析与验证。但是, 由于缺乏实验事实作为依据, 仅仅依靠生物信息学手段预测新型的蛋白质还是具有一定局限性的。因此, 为了验证预测的这 16 种水稻新型潜在冷调节蛋白是否是水稻中的冷调节蛋白, 进而具有提高水稻抗寒能力的功能, 作者下一步的工作将利用转基因等实验手段检验这 16 种蛋白的功能, 也为这 16 种水稻新型潜在冷调节蛋白是否属于水稻冷调节蛋白家族成员提供更为可靠的论据。

参考文献:

[1] Pearce R S. Molecular analysis of acclimation to cold[J]. Plant Growth Regulation, 1999, 29: 47-76.
 [2] Weiser C J. Cold resistance and injury in woody plants[J]. Science, 1970, 169: 1269-1273.
 [3] Guy C L, Niemi K J, Brambl R. Altered gene expression during cold acclimation of spinach[J]. Proc Natl Acad Sci USA, 1985, 82: 3673-3677.
 [4] Hajela R K, Horvath D P, Gilmour S J, et al. Molecular cloning and expression of *cor* (Cold-regulated) genes in *Arabidopsis thaliana*[J]. Plant Physiol, 1990, 93: 1246-1252.
 [5] Franks F. Biophysics and Biochemistry at Low Temperatures[M]. Cambridge University Press, 1985.

- [6] Hinch D K. Rapid induction of frost hardiness in spinach seedlings under salt stress[J]. *Planta*, 1994, 194: 274-278.
- [7] Didierjean L, Frendo P, Nasser W, et al. Heavy-metal-responsive genes in maize: identification and comparison of their expression upon various forms of abiotic stress[J]. *Planta*, 1996, 199: 1-8.
- [8] Mantyla E, Lang V, Palva E T. Role of abscisic acid in drought-induced freezing tolerance, cold acclimation, and accumulation of LT178 and RAB18 protein in *Arabidopsis thaliana*[J]. *Plant Physiol*, 1995, 107: 141-148.
- [9] 康国章, 王正询, 孙谷畴. 植物的冷调节蛋白[J]. *植物学通报*, 2002, 19(2): 239-246.
- [10] Pearce R S. Molecular analysis of acclimation to cold[J]. *Plant Growth Regulation*, 1999, 29: 47-76.
- [11] Breton G, Danyluk J, Charron J B, et al. Expression profiling and bioinformatic analyses of a novel stress-regulated multispinning transmembrane protein family from cereals and *Arabidopsis*[J]. *Plant Physiol*, 2003, 132(1): 64-74.
- [12] 杨同文, 王红星, 刘天学. 大白菜一个冷相关基因的分离与逆境诱导表达[J]. *西北植物学报*, 2010, 30(2): 215-222.
- [13] Niimura Y, Nei M. Extensive gains and losses of olfactory receptor genes in mammalian evolution[J]. *PLoS ONE*, 2007(8): e708.
- [14] Ampbell S A, Close T J. Dehydrins: gene, proteins, and associations with phenotypic traits[J]. *New phytol*, 1997, 137(1): 61-74.
- [15] Jaen A B, Goday A, Figueras M, et al. Phosphorylation mediates the nuclear targeting of the maize RAB17 protein[J]. *plant J*, 1998, 13(5): 691-697.
- [16] Lee S H, Hannink M. Molecular mechanisms that regulate transcription factor localization suggest new targets for drug development[J]. *Adv Drug Deliv Rev*, 2003, 55: 717-731.
- [17] 王艇, 唐振亚. 植物冷驯化和热激反应的分子基础[M]// 刘良式. *植物分子遗传学*. 北京: 科学出版社, 1997: 499-549.