对水稻第4号染色体长臂近端粒区一个 过氧化物酶基因簇的结构分析

陈泽华 周 波 韩 斌 钱跃民 洪国藩 *

(中国科学院国家基因研究中心,上海 200233)

摘要 通过对定位 BAC 克隆 q3037 (H0207F01)的序列测定和分析,在其中一个 22.5 kb 的区域发现一个由 5 个 第三类过氧化物酶基因(依次命名为 $\alpha p1$ 、 $\alpha p2$ 、 $\alpha p3$ 、 $\alpha p4$ 、 $\alpha p5$)组成的基因簇。分析表明, $\alpha p1$ 、 $\alpha p2$ 和 $\alpha p3$ 分别含 1 个内含子, $\alpha p4$ 和 $\alpha p5$ 分别含 2 个内含子。该 5 个基因分别编码 338、335、336、343 和 346 个氨基酸残基 的蛋白质,而且都具有 N 端信号肽序列,其中 OSP1、OSP4、OSP5 为阴离子过氧化物酶,OSP2、OSP3 为阳离子 过氧化物酶。对 5 个基因间的两两比较分析和进化分析结果表明:该基因簇是通过一系列的串联基因复制事件而 形成; $\alpha p5$ 与来自玉米的 ap1和来自大麦(Hordeum vulgare)的 prx7为潜在的直向同源基因,而且, $\alpha p1-5$ 与 ap1、prx7构成了分泌性植物过氧化物酶基因家族中一个新的分枝。

关键词 过氧化物酶;基因簇;串联基因复制;直向同源基因

含血红素的过氧化物酶(EC1.11.1.7)广泛分 布于动、植物和细菌、真菌中,它能氧化一系列的有 机和无机物,同时消耗 $H_2O_2^{[1]}$ 。根据序列相似性, 来源于细菌、真菌和植物的过氧化物酶可归为同一 个超家族,与动物的过氧化物酶家族起源不同,差 异显著。来源于细菌、真菌和植物的过氧化物酶基 因家族又可分为三类:胞内过氧化物酶(I类)、分泌 性真菌过氧化物酶(II类)和分泌性植物过氧化物酶 (III 类)^[2]。植物过氧化物酶被认为与植物的生长、 发育、逆境耐受、防御反应等有关,直接参与了细胞 壁的木质化和角质化、生长素代谢、受伤组织的栓化 愈合和对病原的防御反应^[2]。每一种植物拥有近 100 个第三类过氧化物酶^[3],由此形成一个庞大的 植物过氧化物酶基因家族。第三类过氧化物酶又可 根据其等电点分为中性过氧化物酶、阴离子过氧化 物酶和阳离子过氧化物酶。多数的第三类过氧化物 酶基因具有4个外显子、3个内含子,而且内含子插 入的位点严格保守^[3]。

基因家族的不同成员在基因组中有几种可能的 排布方式:或者弥散分布于整个基因组,或者在染 色体的特定区段成簇分布,或者两种方式兼而有

收稿日期: 2000-11-08 接受日期: 2000-12-21

之。家族基因的成簇分布现象在很多基因组中都被观察到过,并被广泛研究。例如,3个水稻的 -淀 粉酶基因被发现成簇分布于一个 28 kb 的区域^[4]; 对光敏感的、编码二磷酸核酮糖羧化酶小亚基的基 因家族在单子叶、双子叶植物中都存在成簇分布现 象^[5];在拟南芥菜基因组中,3个腈水解酶基因 (*nit2/ nit1/ nit3*)成簇分布于 13.8 kb 的区域^[6],5 个细胞壁偶联的受体激酶基因(*wak*1~5)集中分布 在一个 30 kb 的区域^[7]。

在基因家族进化过程中,普遍存在有基因复制现象^[8]。而复制后的基因位点就有可能进一步复制 而形成多基因簇。在基因簇中只要有一个基因还保 留着最初基因的功能,那么其他成员的功能有可能 发生较大的分化,而组织特异性表达或发育时间特 异性表达也随之发生变化。

植物过氧化物酶家族非常庞大,但是对于家族 成员在基因组中的分布情况却知道不多。在 Trametes versicolor 基因组中,两个木质素过氧化物酶基 因和一个锰离子过氧化物酶基因成簇分布于 10 kb 的区域^[9]。在 Populus kitakamiensis 基因组中,两 个阴离子过氧化物酶基因分布于一个 7 kb 的区 域^[10]。在水稻基因组的遗传图构建和大规模 EST 测序过程中,26 个过氧化物酶位点,114 个过氧化 物酶基因得到鉴定^[11],但基因在染色体上的详细 分布并不知道。其中有一个过氧化物酶位点

^{*} 联系人: Tel, 021-64516371; Fax, 021-64825775; e-mail, gfhong @newnetra.ncgr.ac.cn

(R2184S)位于第4号染色体长臂近端粒区。在以前 的工作中,利用 R2184S 作为探针将 BAC 克隆重叠 群 816 定位到该区域^[12]。从重叠群 816 中选择了 BAC q3037 (H0207F01)进行全序列测定并进行序 列分析,发现了一个由5个过氧化物酶基因组成的 基因簇。本文通过对该基因簇的详尽分析,揭示了 基因之间的演化关系。

1 材料和方法(Materials and Methods)

1.1 序列测定

采用双末端测序战略 (pairwise end sequer cing)^[13]对水稻 BAC q3037 进行了全序列测定。大 致的过程为:首先用超声波将纯化的 BAC DNA 打 碎,然后用绿豆核酸酶对打碎的 DNA 片段进行末 端修补,修补完毕回收 2~4 kb 的片段,回收的片 段再与经 *Sma*I 酶切并脱磷处理的 p KS 载体连接, 连接产物转化 DH5,然后随机挑选亚克隆进行双 末端 测 序,对 所 得 序 列 进 行 组 装 (Phred/ Phrap)^[14,15],最后再用引物定向步行的方法填补 余下的少数缺口。组装的正确性可从亚克隆框架和 *Not*I 酶切分析得到保证。

1.2 序列分析

利用 NCBI(http://www.ncbi.nlm.nih.gov/ BLAST)的 BLAST 程序进行数据库检索^[16]。GCG 软件包(10.1版本)被用来进行 DNA 和蛋白质序列 的分析。运用 Testcode^[17]和 Genscan^[18]进行潜在的 蛋白质编码区的预测和基因结构预测。多序列对准 由 PileUp 程序进行。信号肽预测由 SPScan 程序进 行。蛋白质的分子量和等电点计算由 Compute pI/ Mw(http://www.expasy.ch/tools/pi_tool.html)程 序进行。蛋白质氨基酸序列两两比较由 Gap 程序进 行。系统进化树的构建由 Grow Tree 程序完成。

2 结果(Results)

2.1 序列测定

完成以后的 q3037 全序列全长 124.3 kb(Gen-Bank 登录号 AJ245900),平均序列丰度为 10.5。经 亚克隆框架分析和 *Not* I 酶切分析,确保了组装的 正确性。同时,高丰度的序列覆盖也确保了最终完 成序列的质量。在对该序列的初步分析时发现,其 中 22.5 kb 的区域包含一个过氧化物酶基因簇,然 后就对该区域进行了详尽的分析。

2.2 基因鉴定

利用 GCG 软件包(10.1 版本)的 TestCode 程序

对该 22.5 kb 区域进行了分析,结果表明,5 个蛋白质编码区不均匀地分布于该区域[图1(A)]。将相应的蛋白质编码区分别进行 BLASTx 分析,显示都与数据库中的植物过氧化物酶有显著相似,表明该 22.5 kb 区域包含一个由 5 个过氧化物酶基因组成的基因簇。这 5 个基因依次命名为 $\alpha p1$ 、 $\alpha p2$ 、 $\alpha p3$ 、 $\alpha sp4$ 和 $\alpha p5$,其中 $\alpha p5$ 对应于遗传标记 R2184S。该 5 个基因具有相同的转录方向,基因间隔区分别为 2 、5.5 、1.7 和 3.3 kb。

2.3 基因结构预测

对该 22.5 kb 序列运用 Genscan 软件分析, 预 测了 5 个过氧化物酶基因的结构。结果表明, *asp*4 和 *asp*5 分别含 3 个外显子和两个内含子, 而 *asp*1, *asp*2 和 *asp*3 只有两个外显子, 被单一的内含子所 切断[图 1(B)]。



Fig. 1 Gene prediction in the 22.5 kb region

(A) TestCode analysis of the 22.5 kb sequence. The top region is supposed to predict coding regions to a 95 % level of confidence. The bottom region is supposed to predict non-coding regions to the same confidence level^[17]. (B) Genscan prediction of the 22.5 kb sequence. Gene structures are predicted using the matrix of *maize* (http://genes.mit. edu/ GENSCAN.html)^[18].

2.4 数据库检索

将 5 个过氧化物酶基因编码区分别用于数据库 检索,当 EST 全长的 90 %以上与待检索序列的相 同率大于 95 %以上时被认为是相互匹配的^[19]。在 对核酸数据库(nr)进行检索时,得到一个与 α_{p5} 匹配(相同 98.8 %)的全长 cDNA 序列 α_{scpx1} (Ger Bank 登录号: AF019743)。在对 EST 数据库 (dbEST)检索时,得到来源于不同组织的分别与 $\alpha_{p1}, \alpha_{p3}, \alpha_{p5}$ 相匹配的若干 EST 条目(表 1),这 些 EST 一部分与基因的 5 端相匹配,一部分与基 因的 3 端相匹配。而 α_{p2}, α_{p4} 则没有相匹配的 EST 或 cDNA 序列。

将 *asp*1、*asp*3、*asp*5 分别与相匹配的 cDNA 或 EST序列对准,可准确确定基因的内含子剪切位 点。而 *asp*2、*asp*4 与相似程度较高的 cDNA 或 EST 进行对准,也能确定出基因的内含子/外显子边界 Mar., 2001

(图 2)。所	有5・	个基因的内含子中,除 asp4的2个	个基因的外显于	子 GC 含量	量在 64.	8 % ~ (59.4%	,而内
内含子,as	p5 的	第二个内含子遵循 GT/AG的序列 💦	含子 GC 含量在	E 25 % ~	45.5%((表 2)。	。内含	子的高
模式以外,	其余	内含子都为 GC/AG的序列模式。 A	AT含量保证	了在基团	因转录后	「加工	时的1	与 效 剪
对 5 个基因	的外	显子和内含子计算 GC 含量发现,5 1	切 ^[20] 。					
osp1 AU075454 AU031972	436 1 12	gccacaagctagcacagcacagccggtaagagagtgag caagctagcacagcac	tgcggagagccaaa tgcggagagccaaa tgcggagagccaaa	ttgag atg ttgagatg ttgagatg	ggtgcagc ggtgcagc ggtgcagc	gagcag gagcag gagcag	gagact gagact gagact	515 76 91
osp1	516	ggccgtcctggagctcgtctccatcgtcgccgtcctcc	tcatctcctcacca	gcagcagc	agcagagt	tgtccg	tggact	595
AU075454	77	ggccgtcctggagcggntctccatcgtcgccgtcctcc	tcatctcctcacca	gcagcagc	agcagagt	tgtccg	tggact	156
AU031972	92	ggccgtnctggagnnnccctacatcgtcgccgtcctcc	tcatctcctcacca	gcagcagc	agcagagt	tgtccg	tggact	171
<i>os</i> p1	596	<pre>tccacgcggcgtcgtgcccgcagctggagtccatcgtg</pre>	egetecteegtgea	agcegete	tccagcag	gagatc	gecete	675
AU075454	157	tccacgcggcgtcgtgcccgcagctggagtccatcgtg	egeteeteegtgea	ageegete	tccagcag	gagatc	gecete	236
AU031972	182	tccacgcggcgtcgtgcccgcagctggagtccatcgtg	egeteeteegtgea	ageegete	tccagcag	gagatc	gecete	251
osp1 AU075454 AU031972	676 237 252	gccgccggcctcctccgcatcttcttccacgactgcttc gccgccggcctcctccgcatcttcttccacgactgctt gccgccggcctcctccgcatcttcttncacgactgctt	ecegeag <u>geaegea</u> eeegeag eeegeag	cgcaagaa	catttgca	cgtacc	gacgaa	755 281 296
<i>os</i> p1	756	caatgetgtaetgaaetgaaeataeatatatatatgea	ggggtgcgatgcgt	cggtgtat	ctgagagg	aggcag	caactc	835
AU075454	282		ggtngcgatgcgt	cggtgtat	ctgagagg	aggcag	caactc	322
AU031972	297		ggttgcgatgcgt	cggtgtat	ctgagagg	aggcag	caactc	337
<i>os</i> p1 AU075454 AU031972	836 323 338	ggagcagggcatgggcccaaacc-tcacgctgcagccg ggagcagggcatgggcccaaacc-tcacgctgcagccg ggagcagggcatgggcccaaaccttcacgctgcagccg	c-gggcgctgcagc cgggggcgctgcagc c-gggcgctgcagc	tggtggag tggtggag tggtggag	gacatceg gacat gacatceg	cgccaa agccaa	ggtgca ggtgca	913 386 416
osp1	914	cgccgcctgcggccccaccgtctcctgcgccgacatct	ccgccctcgccacc	cgcgacgc	tgtcgtcg	teteeg	geggee	993
osp1	994	cctcttacgccgtgccactcggccaaaaagacagcctc	gctccggcatcgtt	ggacctcg	tcggagat	eteeeg	ggeeeg	1073
osp1	1074	ggcacctccagagtccaggatctcatcgacctttttgc	aagcaggggactgc	gggacgcc	gccgacct	ggtgge	getete	1153
AU030963	1	agagtccag-anctcatcgacctttttgc	aaccaggggantgc	gggacgcc	gccgacct	ggtgge	gttete	70
<i>osp</i> 1	1154	eggegggcacaccgtegggaggacacgetgegeettet	tcgatgaccgcgcc	cgccgcca	ggacgaca	cettet	ccaaga	1233
AU030963	71	eggegggcacaccgtegggaggacacg-tgegeetttt	tcgatgaccgcgcc	cgccgcca	ggacgaca	cettet	ccaaga	149
<i>os</i> p1	1234	agetggegttgaactgeaceaaggaeeeeaaceggetge	cagaacctggacgt	gatcacco	cegaegee	ttcgac	aacgcc	1313
AU030963	150	agetggegttgaactgeaceaaggaeeeeaaceggetge	cagaacctggacgt	gatcacco	cegaegee	ttcgac	aacgcc	229
osp1	1314	tactacatcgcgctcacccacaaccagggcgtcttcacc	ctccgacatggcgc	tcatcaag	gaccggat	cacggc	gcccat	1393
AU030963	230	tactacatcgcgctcatccacaaccagggcgtcttcacc	ctccgacatggcgc	tcatcaag	gaccggat	cacggc	gcccat	309
osp1	1394	cgtcaggcagttcgccacggacaaggccgccttcttcac	cgcagttcgccaag	tccatggt	caaactaa	gcaatg	tgccga	1473
AU030963	310	cgtcaggcagttcgccacggacaaggccgccttcttca	cgcagttcgccaag	tccatggt	caaactaa	gcaatg	tgccga	389
osp1	1474	ggactgacaggaacgtcggcgagatccgccgcagctgct	ttcaggaccaacag	ccagagcci	togtogac	ttogoo	accagc	1553
AU030963	390	ggactgacaggaacgtcggcgagatccgccgcagctgct	ttcaggaccaacag	ccagagcci	togtogac	ttogoo	accagc	469
ospl	1554	gatgaggagggettegetgettetget tga teatteage	ctgtctgtcgttgc	togttaati	tagggagc	taggtg	cttgac	1633
AU030963	470	gatgagga-ggnttegetgettetgett <mark>ga</mark> teatteage	ctgtctgtcgttgc	togttaati	tagggagc	taggtg	cttgac	549
osp1	1634	gtegetagetgatectagtetetttagtetteaggeett	tggcttttgtacgca	atatgctt:	cgttatg	tccgtg	caataa	1713
AU030963	550	gtegetagetgatectagtetetttagtetteaggeett	tggcttttgtacgca	atatgctt!	cgttatg	tccgtg	caataa	629
The trans	slational	Fig. 2A The genomic sequence of osp1 an start (5-AGT) and stop codons (5-TGA or 5-TAA) are	nd its alignment w highlighted and under	vith matcher dined. The i	ed ESTs ntrons are u	nderlined	and in ital	ic.
<i>os</i> p2	277	agagtgagagcagagagccaaattaag atg gctgcagc	gaggagactgcccg	tcctggag	ctggtctc	cttcgt	cgccgt	356
AU075454	27	agagtgagtgcggagagccaaattgagatgggtgcagc	gaggagactggccg	tcctggag	cggntctc	catcgt	cgccgt	106
AU031972	42	agadtgagtgcggagagccaaattgagatgggtgcagc	gaggagactggccg	tnctggag	nnncccta	catcgt	cgccgt	121
osp2	337	cttccttatctcctcaccaacagcagcagcagccgagt	tgtctgtggacttc	cacgcggc	gtogtgoo	cgtcgc	tggagg	436
AU075454	107	cctcctcatctcctcaccagcagcagcagcagcagagt	tgtccgtggacttc	cacgcggc	gtogtgoo	cgcagc	tggagt	186
AU031972	122	cctcctcatctcctcaccagcagcagcagcagcagagt	tgtccgtggacttc	cacgcggc	gtogtgoo	cgcagc	tggagt	201
<i>osp</i> 2 AU075454 AU031972	397 147 162	ccatcgtgagctcctccgtgcaagccgcgctgcagcaa ccatcgtgcgctcctccgtgcaagccgctcccagcag ccatcgtgcgctcctccqtgcaagccgctctccagcag	gagatcgccctcgc gagatcgccctcgc gagatcgccctcgc	cgccggcc cgccggcc	teeteege teeteege teeteege	atette atette atette	ttccac ttccac ttncac	516 266 281
<i>osp</i> 2 AU075454 AU031972	517 267 282	gactgcctcccgcag <i>gcacgcacgcagaacatactcc</i> gactgcttcccgcag gactgcttcccgcag	tatattcgcacgta	cgtacgaa	caacaatg	ttgtac	tcaact	596 281 296
<i>osp2</i>	577	gaacatgetgeatatgeatgtgeagggggggggggggggg	cggtgtatcttagg	ggaggcag	caactcag	agcagg	gcatgg	676
AU075454	282		cggtgtatctgaga	ggaggcag	caactcgg	agcagg	gcatgg	336
AU031972	297		cggtgtatctgaga	ggaggcag	caactcgg	agcagg	gcatgg	351
<i>osp</i> 2 AU075454 AU031972	637 297 312	gcccaaacctgacgctgcagccgcggggcgctgcagctc gcccaaacctcacgctgcagccgcggggcgctgcagct gcccaaacctcacgctgcagccgcggggcgctgcagctg	gtcgacgacatccg gtggaggacatccg	cgccaggg agccaagg	tgcatgca tgca	gcatgc	ggeeee	756 374 416
osp2 osp2 osp2 AU030963	757 817 877 48	accgtctcctgcgccgacatctccgcgctcgccacccg tctcggccaaaaagacagcctcgccccggccccggtgc aggctctcctcgacaaattcggaagcaagggcctgaga	tgacgccgtcgtcg gcctcgtcaaccaa gaggccgccgacct gccgccgacct	teteeggg eteeeggg ggtggege ggtggegt	ggaccete ecegggea teteegge teteegge	gtacgc cctcca gcccac gggcac	agtgtc gcgtcc accgtc accgtc	836 916 996 86
osp2	997	gggagggcgcactgcgacttcttccgtgaccgcgccgc	ccgccaggacgaca	cettetee	aagaagct	ggccgt	gaactg	1076
AU030963	87	gggaggacacg-tgcgcctttttcgatgaccgcgcccg	ccgccaggacgaca	cettetee	aagaagct	ggcgtt	gaactg	165
osp2	$\begin{array}{c} 1057 \\ 146 \end{array}$	caccaaggaccccaaccggctgcagaatctggacgtcg	tcacgcccgacgcc	ttogadaa	cgcctact	acgtcg	cgctca	1156
AU030963		caccaaggaccccaaccggctgcagaacctggacgtga	tcacccccgacgcc	ttogadaa	cgcctact	acatcg	cgctca	245
osp2	1117	ccaggaagcagggggggtgttcacctccgacatggcgctc	atcaaggaccggat	cacggcgc	caatcgtc	aggcag	ttegee	1236
AU030963	206	tccacaaccagggcgtcttcacctccgacatggcgctc	atcaaggaccggat	cacggcgc	ccatcgtc	aggcag	ttegee	325
osp2	1237	gcggacaaggccgcattcttcacgcagttcgccaagtc	catggtcaagctca	gccaggtt	ccgaggac	tgacag	gaacgt	1336
AU030963	326	acggacaaggccgccttcttcacgcagttcgccaagtc	catggtcaaactaa	gcaatgtg	ccgaggac	tgacag	gaacgt	425
osp2	1297	cggcgagatecgccgcagctgcttcaggaccaacggac	cacgcctcgtcgac	ttggccac	cggcgatg	aggccg	cttctc	1396
AU030963	386	cggcgagatecgccgcagctgcttcaggaccaacagcc	agagcctcgtcgac	ttcgccac	cagcgatg	agg		476

Fig. 2B The genomic sequence of osp2 and its alignment with high similar ESTs

The translational start (5 - ATG) and stop codons (5 - TGA or 5 - TAA) are highlighted and underlined. The introns are underlined and in italic.

osp3	197	agtgaggagagtagtccacaactccggccggccactacgactctacgagtacgtac	276
AU062505	6		85
D48606	1		72
osp3	277	ccgcaacagcgggcgcattggccgtcctgcagctcgcctccatcgtcgccgtcgtcctcctcctcccccacctgcgcgcaacagcgggcgcattggccgtcctgcagctgcctccatcgtcgccgtcctcctcctcccccacctgcgcgcaacagcgggcgcattggccgtcctgcagctgcctccatcgtcgccgtcctcctcctcccccacctgcgcgcaacagcgggcgcattggccgtcctgcagctgcctccatcgtcgccgtcctcctcctccccccctgcgcgccgtcctcc	356
AU062505	86		177
D48606	73		164
osp3 AU062505 D48606	357 178 165	agaccgcgggc gctgctgctgctgctgtggacttcattgacgtggtggcgtgctcgcagtcgcagtggactccatcgtgcgctcagc gctgctgctgagccgtctgtggacttcattgacgtggtggcgtgctcgcagtcgcagtggactccatcgtgcgctcagc gctgctgctgagccgtctgtggacttcattgacgtggtggcgtgctcgcagtcgcaggtggactccatcgtgcgctcagc	436 257 244
osp3	437	cgtgcaggcgg-cgctccagcgtgagatcgccctcgccgccggcctcatccgtatcttcttccacgactgcttcccgcagcggcggacgctccagcgtgagatcgcctcgccgccggcctcatccgtatcttcttccacgactgcttcccgcagcggcgg-cgctccagcgtgagatcgcctcgncgncggcctcatccgtatcttcttccacgactgcttcccgcagcggcgg-cgctccagcgtgagatcgcctcgncgncggcctcatccgtatcttcttccacgactgcttcccgcagcggcggcgg-cgctccagcgtgagatcgcctcgncgncggcctcatccgtatcttcttccacgactgcttcccgcagcggcggcggcggcgcdgactgcttcccgcagcggcggcggcggcggcggcggcggcggcdgactgcttcccgcagcggcggcggcggcggcggcggcggcggcdgacggcggcggcggcdgacggcggcggcggcggcggcggcggcggcggcggcggcg	526
AU062505	258		337
D48606	245		323
osp3	527	$\begin{array}{c} gcacttgctactatatatatactctacctttttatgaactaaaccaacc$	595
osp3	596		675
AU062505	338		354
D48606	324		339
osp3	676	ttacttgagtggcgccaact-cggagcaggggatgccgcccaacgccaac-tcgctgcagccggggcgctgcagctggt ttacttgagtggcgccaacttcggagcaggggatgccgccaacgccaacntcgctgcagccgngggcgctgcagctggt ttacttgagtggcggcaact-cggagcaggggatgccggccaacgccaacgccaac-tcg	753
AU062505	355		434
D48606	340		391
osp3	754	$ggaggacatccgcgccaaggtgcacgccgcatgcgggcccaccgtctcctgcaccgacatctccgcgctggccacccgcg\\ggaggacatccgngccaaggtgcacgcc$	833
AU062505	435		462
osp3 osp3 osp3 osp3 osp3 osp3 AU091519	834 914 994 1074 1154 1	cggccgtcgtcctctccggcgggcccacctaccccgttcccctcggccagctcgacagcctcgccccggccccgctgggagaggggcatgggagagggcatggggagccgggcccggcgcccggcccacccgtggggaagtcaagggggcatgggaggggcatgggaggggcatgggaggggccttccccggaaggagggggggg	913 993 1073 1153 1233 68
озрЗ	1234	a actg cgg ccatcg t cgg cgg t t cg cg cagg a caagg ccg cct t ctt cacg cag t t cg t t a cg t catcg t caag ctg a a a ctg cgg ccatcg t t cg t catcg t cg cg cg g t cg cg cg g a caagg ccg cct t ctt cacg cag t t cg t t a cg t catcg t ca g ctg a a g c cg c c	1313
AU091519	69		148
osp3	1314	gcaaggtgccgaggcccggggggaacaaaggcgagatccgccgcaactgcttcaagaccaacagcggagcacgcetcgtg	1393
AU091519	149	gcaaggtgccgaggcccggcgggaacaaaggcgagatccgccgcaactgcttcaagaccaacagcggagcacgcctcgtg	228
osp3	1394	gaegtegtggagggettegetgeetetget <mark>taa</mark> tteatgteatgtegeteategagatgaatatetagggtegaeateat	1473
AU091519	229	gaegtegtgga-ggettegetgeetetget taa tteatgteatgtegeteategagatgaatatetagggtegaeateat	317
<i>os</i> p3 AU091519	$\begin{array}{c}1474\\318\end{array}$	$\begin{tabular}{lllllllllllllllllllllllllllllllllll$	1553 397
osp3	1554	tttcgttacggccgtgcgcgcgttaataagacgtactcccgcaccgtaataattatgttgttttaa	1619
AU091519	398	tttcgttacggccgtgcgcgcgttaataagatgtactcccgcaccgtaa	446

Fig. 2C The genomic sequence of osp3 and its alignment with matched ESTs

The translational start (5-AGT) and stop codons (5-TGA or 5-TAA) are highlighted and underlined. The introns are underlined and in italic.

osp4	138	cagatcacatcaccacagtgtgggtgtaattaaggttgggaaacgagtaattagcttaggttaattaggattcgatccatcgatacgatggcgttgagaagaatgggcatcattcttgcgctcttggccattgctactctgctctcgccgcagtgtccgcacggttcattcttccacatcatcacatcgcgcgcccctcgccgcagggactcaccatcacatcgcgcgcccctcgccgcgcgcccccgccgcgcgcccccgccg	217
osp4	218		297
osp4	298		377
osp4	378		457
AU075454	168		241
BE405294	182		255
osp4	458	cggcctcctccgcatcttcttccacgactgcttcccccagg tacttgttactgatttgttgttgattaagtaatgaacacggcctcctccgcatcttcttccacgactgcttcccgcagcctcctccgtatcttcttccacgattgcttcccgcagcctcctccgtatcttcttccacgattgctttcctcagcagcctcctccgtatcttcttccacgattgctttcctcagcagcctcctccgtatcttcttccacgattgctttcctcagcagcctcctccgtatcttcttccacgattgctttcctcagcagcctcctccgtatcttcttccacgattgctttcctcagccctcctccgtatcttcttcctcacgattgctttcctcagccctcctccgtatcttcttcctcacgattgctttcctcacgccctcctccgtatcttcttcctcacgattgctttcctcagccctcctccgcagcctcctccgtatcttcttccacgattgctttcctcacgattgcttcctcacgattgcttcctcccccccc	537
AU075454	242		281
BE405294	256		295
osp4	538	$\frac{ctaggttagctagctctgtcaacctaacctaattgacctaattgaatttgcagggctgcgacgcgtcgctgctgctgctgacggtgtatctgaga}{ggttgcgacgcgtcgtcgtgtgtatctgaga}$	617
AU075454	282		308
BE405294	296		322
osp4	618	ggagccaacagcgagcagcagctgccgcccaacctgactctgcaaccacgggcgctgcagctcatcgaggacatccgcgc	697
BE404151	1	gcgctgcagctcatcgagtccatccgtgc	29
BE405294	323	ggacccaacagcgagcgggacctgccgccgaaccagacgctgcagccgcgtgcgatgcagctcatcgaggacgtccgcgt	402
osp4	698	ccaggtgcacgccgcatgcggccccaccgtctcctgcgccgacatcaccgccctcgccacccgcgacgccatcgtcgccgccgcgggcccgtcgtctcctgcgacgcccttgctacccgtgatgccgtggccttccaaggtgcacgccgccgtcgccgtcgtctcctgcgacatcatcgcccttgctacccgcgacgccgtcgtcttcgtg	777
BE404151	30		108
BE405294	403		481
osp4	738	tacgtacagatcgatacatgtttgcccattgtgcatatatat	857
osp4	858		937
BE404151	109		188
BE405294	507		572
osp4	938	cccgcagcccacctccgacgtctccacgctgctcagcgccttccagacccgcaacctcgacaacgtcgacctcgtcgcgccccccgcagcccacgccgccacgctcatcgacgtcttccagacccgcaacctcgacaacgtcgacctcgtcgcgccccgcagtccaccgccgacgccaacacgctcatcaacgccgttcaagagccgcaacctcgagccgatcgacctcgtcgcgccgcgccgcdccacgtcgtcgcgccgcgc	1017
BE404151	189		268
BE405294	573		652
osp4 BE404151 BE405294	1018 269 653	tctccggcggccactccatcggcagggcgcgatgcagctccttctccaaccgcttccgcgaggacgacgacttcgccaggtctccggcggccacaccatcggggacggccactgcagctccttctccaaccgcttctccgaggactccgacttcgtccgctcttggcg	$ \begin{array}{r} 1097 \\ 348 \\ 661 \end{array} $
osp4 BE404151	1098 349	aggetegecegecaactgetetaacgacggeagecegectgeaggagetegatgteaceaceceggacgtgttegacaacaa agectegecteceactgeteceagegactttaaceggttgeaggacetegatgteaceceeggacgtgttegacateaa	$\frac{1177}{428}$
osp4	1178	gtactacagcaacctggtggccggtcagggggtgttcacttccgaccagggggtctcaccggcgactggcgcaccagctggggtacttcacctccgaccagccgactggcgcaccgatggggtcttcacctccgaccagacgcgactggcgcaccgaatggg	1257
BE404151	429		508
osp4	1258	tggtcaatggcttcgccggcaaccactggtggttctacggccagttcggcagctccatggtgaagctgggacagctccagtggtcaacggcttcgccgggaaccactggtggttcttcggccagtttgccgcctccatgaccaagctcgggaatctccaggtggttcttcggccagtttgccgcctccatgaccaagctcgggaatctccaggtggttcttcggccagtttgccgcctccatgaccaagctcgggaatctccaggtggttcttcggccagtttgccgcctccatgaccaagctcgggaatctccaggtggttcttcggccagtttgccgcctccatgaccaagctcgggaatctccaggtggaatctccaggtggttcttcggccagtttgccgcctccatgaccaagctcgggaatctccaggtggttcttcggccagtttgccgcctccatgaccaagctcgggaatctccaggtggaatctccaggtggaatctccaggtggaccacggtggaatctccaggtggaccactggtggatctccaggtggaccactggtggatctccaggtggaccactggtggaccactggtggtggtggtggtggtggtggtgggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggtggtggtggtggtggtggtgggaccactggtggaccactggtggaccactggtggaccactggtggtggtggtggtggtggtggtggtgggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggtggtggtggtggtgggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggtggtggtgggccaggttgggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggaccactggtggtggtggtggtggtggtgggtg	1337
BE404151	509		588
osp4	1338	ggtccttcagggaacgtcggcgagatccgccgcaacagctgcttcgtgcccaacagccagaccatccttgcggccgccggggggccacaacagccagaccatccttgcggccgccgg	1417
BE404151	589		637
osp4	1418	$cgacgatggcttcacggcatctgct \\ \begin{tabular}{lllllllllllllllllllllllllllllllllll$	1497
osp4	1498		1577

Fig. 2D The genomic sequence of osp4 and its alignment with high similar ESTs

The translational start (5-ATG) and stop codons (5-TGA or 5-TAA) are highlighted and underlined. The introns are underlined and in italic.

<i>os</i> p5	47	gaggcagaggtcgagctgagagtgatcgacctttgttagctagagtgttgagcaatggcgtccaagctgggt atg gttgt	126
AF019743	1	gtcgagctgagagtgatcgatctttgttagctagagtgttgagcaatggcttccaagctgggtatggttg	71
BE040760	5	gaggcagaggtcgagctgagagtgatcgacctttgttagctagagtgttgagcaatggcgtccaagctgggtatggttgt	84
osp5	127	$gctactgatctcgggcctctttgctgcccgttgcgcggccgtggtgaccaccggcgaacccgtcgtcgccggcctctcct gctactgatctcgggcttctttgctgcccgtggccgtggtgaccaccggcgaacccgtcgtcgccggcctctcct gctactgatctcgggcctctttgctgcccgttgcgcggccgtggtgaccaccggcgaacccgtcgtcgccggcctctcct \\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ $	206
AF019743	72		151
BE040760	85		164
osp5	207	gggggttctatgacacgtcgtgcccgtcggtggagggcatcgtgaggtggcacgtcaccgaggccctccgccgcgacatc	286
AF019743	152	gggggttctatgacacgtcgtgcccgtcggtggagggcatcgtgaggtggcacgtcaccgaggccctccgccgcgacata	231
BE040760	165	gggggttctatgacacgtcgtgcccgtcggtggagggcatcgtgaggtggcacgtcaccgaggccctccgccgcgacatc	244
osp5	287	ggcatcgccgccgcctcgtccgcatcttcttccacgactgcttcccgcaggcacgtcctactaccattttactccagtaggcatcgccgccggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgccatcttcttccacgactgcttcccgcaggcctcgtccgccgccgcctgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgccatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctgctccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctgctccgcaggcctcgtccgcatcttcttccacgactgcttcccgcaggcctgctccgcaggcctcgtccgcaggcctcgtccgcaggcctgtcccgcaggcctgctccgcaggcctgctccgcaggcctgtcccgcaggcctgtcccgcaggcctgctccgcaggcctgctccgcaggcctgtcccgcaggcctgctccgccgcaggcctgtcccgccaggcctgctccgccgcaggcctgctccgccgccgccgccgccgccgccgccgccgcc	366
AF019743	232		282
BE040760	245		295
osp5	367	gttgccggttcggttgatgcatggtgcctccttcttacagatggcaatacatac	446
osp5	447		526
AF019743	283		318
BE040760	296		331
osp5	527	agcgagctgggtgagatacccaaccagacgctgcggccgtcggcgctgaagctcatcgaggacatccgccgccgtaca	606
AF019743	319	agcgagctgggtgagatacccaaccagacgctgcggccgtcggcgctgaagctcatcgaggacatccgcgccgccgtaca	398
BE040760	332	agcgagctgggtgagatacccaaccagacgctgcggccgtcggcgctgaagctcatcgaggacatccgcgccgtaca	411
osp5	607	$\label{eq:ctgcgcgccaaggtgtcctgcgccgacatcaccacgctcgccacgcgtgacgccatcgtcgccgtacgtctaa} at ccgcctgcggcgccaaggtttcttgcgccgacatcaccacgctcgccacgcgtgacgccatcgtcgct ctcgccgccacgcgtgacgccatcgtcgcc ctcgcgcgccaaggtgtnctgcgccgacatcaccacgctcgccacgcgtgacgccatcgtcggc catcgtcggc catcgtcggc catcgtcggc catcgtcgccacgcgtgacgccatcgtcggc catcgtcggc catcgtcgc catcgtcggc catcgtcggc catcgtcgc catcgtcggc catcgtcgc cat$	686
AF019743	399		468
BE040760	412		481
<i>os</i> p5	687	$\begin{tabular}{lllllllllllllllllllllllllllllllllll$	766
<i>os</i> p5	767		846
AF019743	469		534
BE040760	482		547
osp5	847	<pre>aaggtgggcctcctgccggcgcccttcttcgacgtgcccacgctcatccaggcgttcaaggaccgaaacctggacaagac</pre>	926
AF019743	535	aaggtggggctcctgccggcgcccttcttcgacgtgcccacgctcatccaggcgttcaaggaccgaaacctggacaagac	614
BE040760	548	aaggtgggcctcc	560
osp5	927	ggacctggtggcgctgtccggcgcgcacaccatcggactaggccactgcggcagcttcaacgaccgcttcgatggctcca	1006
AF019743	615	ggacctggtggcgctgtccggcgcgcacaccatcggactaggccactgcggcagcttcaacgaccgcttcgatggctcca	694
osp5	1007	agcccatcatggaccctgtgctggtgaagaagctgcaggccaagtgcgccaaggacgtgccggtgaactcggtcacgcag	1086
AF019743	695	agcccatcatggaccctgtgttggtgaagaagctgcaggccaagtgcgccaaggacgtgccggtgaactcggtcacgcag	774
osp5	1087	gagetggaegteegeacgeeceaacgeettegaeaacaagtaetaettegaeeteaageaggggatetteaagte	1166
AF019743	773	gagetggaegteegeaegeeceaacgeettegaeaacaagtaetaettegaeeteategeeaageaggggatetteaagte	852
BE039701	11	gagetggaegteegeaegeeceaaegeettegaeaacaagtaetaettegaeeteategeeaageaggggatetteaagte	90
osp5	1167	cgaccagggcctcatcgaggacgcgcagaccaaccgcaccgccgtccgcttcgccctcaaccaggccgccttcttcgacccgaccagggcctcattgaggacgcgcagaccaaccgcaccgccgtccgcttcgccctcaaccaggccgccttcttcgacccgaccagggcctcatcgaggacgcgcagaccaaccgcaccgccgtccgcttcgccctcaaccaggccgccttcttcgacccgaccgccgtcgccttcaaccaggccgccttcttcgacccgaccgccgtcgccttcaaccaggccgccttcttcgacccgaccgccgtcgccttcaaccaggccgccttcttcgacccgaccgccgtcgccttcttcgacccgaccgccgccttcttcgacccgccgtcgcctcaaccaggccgcctcttctcgacccgccgtcgccttcttcgacccgccgcctcaaccaggccgcctcttcttcgacccgaccgccgcctcaaccaggccgcctctcttcgacccgaccgccgccgcctcaaccaggccgcctcttctcgacccgaccgccgcctcaaccaggccgcctcttctcgacccgccgccgccgcctcaaccaggccgcctctcttcgacccgccgccgcctcaaccaggccgcctctctcgacccgccgccgccqccctcaaccaggccgcctctctcqacccgccgccgccqcccqccgccgccqccctcaaccaggccgcctctctcqaccaggccgcctccqccqccqccqccqccqccqccqccqccqccqccqc	1246
AF019743	853		932
BE039701	91		170
osp5	1247	agttcgcacgctccatggtcaagatgagccagatggacgtcctcaccggcaacgccggcgagatccgccaacaactgcgcc	1326
AF019743	933	agttcgcacgctccatggtcaagatgagccagatggacgtcctcaccggcaacgccggcgagatccgcaacaactgcgcc	1012
BE039701	171	agttcgcacgctccatggtcaagatgagccagatggacgtcctcaccggcaacgccggcgagatccgcaacaactgcgcc	250
osp5	1327	gctcccaaccgccgctcctccgacctcct-caacgctgccgacgacgaccaaggcttcgccgccgacgcc taa ttaactt	1405
AF019743	1013	gctcccaaccgccgctcctccgaacttcttcaacgctgccgacgacgaccaaggcttcgccgcggacgcctaattaactt	1092
BE039701	251	gctcccaaccgccgctcctccgacctcct-caacgctgccgacgacgacgacgacgacgcctaattaactt	329
osp5	1406	atggagtaattagtgatcgttttatgtttatgttttgtgctagtaataataattaagaggatgccatctgcgcgtggtgt atggagtaattagtgatcgttttatgtttatgttttgtgctagtaataataattaagaggatgccatctgcgcgtggtgt atggagtaattagtgatcgttttatgtttatgttttgtgctagtaataataattaagaggatgccatctgcgcgtggtgt atggagtaattagtgatcgttttatgtttatgttttgtgctagtaataataattaagaggatgccatctgcgcgtggtgt atggagtaattagtgatcgttttatgtttatgttttgtgctagtaataataattaagaggatgccatctgcgcgtggtgt atggagtaattagtgatcgtttatgtttatgttttgtgctagtaataataattaagaggatgccatctgcgcgtggtgt atggagtaattagtgatcgttttatgtttatgttttgtgctagtaataataattaagaggatgccatctgcgcgtggtgt atggagtaattagtgatcgtttatgtttatgttttgtgctagtaataataataattaagaggatgccatctgcgcgtggtgt atggagtaattagtgatcgttttatgtttatgttttgtgctagtaataataattaagaggatgccatctgcgcgtggtgt atggagtgt atggagtaattagtgtgtgt atggagtgt atggaggagtgt atggaggaggagtgt atggaggaggaggagtgt atggaggaggaggaggaggaggaggaggaggaggaggagg	1485
AF019743	1093		1172
BE039701	330		409
osp5 AF0197 4 3 BE039701	1486 1173 410	eq:ttgtttccatgcattctctgcttagttagaatggttttgcttcataaaaagtaaagttactactcgattcctcggtcgg	$1565 \\ 1252 \\ 489 \\$
osp5	1566	gacagagtaactgcatgtcaaatgtatcatcatcatcagttagtgctacatcatcacagctgtctttgtcacagcataaa	1645
AF019743	1253	gacagagtaactgcatgtcaaatgtatcatcatcatcagttagtgctacatcatcaca	1309
BE039701	490	gacagagtaactgcatgtcaaatgtatcatcatcatcagttagtgctacatcatcacagctgtctttgtcacagcataaa	569

陈泽华等:对水稻第4号染色体长臂近端粒区一个过氧化物酶基因簇的结构分析

167

Fig. 2E The genomic sequence of osp5 and its alignment with matched ESTs

The translational start (5-ATG) and stop codons (5-TGA or 5-TAA) are highlighted and underlined. The introns are underlined and in italic.

2.5 预测的过氧化物酶结构

Mar., 2001

5 个基因分别编码 338、335、336、343、346 个氨基酸残基的蛋白质。典型的第三类过氧化物酶大小通常在 330~360 个氨基酸。在 5 个过氧化物酶的 N 端预测了信号肽序列,该信号肽可负责将相应的 过氧化物酶导向细胞外。根据氨基酸序列计算 5 个 过氧化物酶的分子量分别为:33.3、22.7、32.1、34.7、34.8 kD。而且根据计算,OSP1、OSP4、OSP5 的等电点分别为 5.75、4.69、5.42,因此属于阴离子 过氧化物酶;而 OSP2、OSP3 的等电点分别为:8.17、8.18,因此属于阳离子过氧化物酶(表 3)。这 表明阳离子过氧化物酶基因和阴离子过氧化物酶基因可在染色体上紧密连锁。

通过多序列对准可发现在过氧化物酶家族中最 为保守的三个结构域[图 3(B),(D),(F)],其中 (B)、(F)为血红素结合区,而(D)为功能未知的保 守区^[21]。对于三类过氧化物酶都保守的9个氨基 酸残基中^[2],除了在 OSP1-3 中 123 位的 Arg(R)被 Gn(Q)所替代以外,其余8个残基在 OSP1~5 中 都被观察到(图 3)。绝大多数第三类过氧化物酶保 守的残基都在 OSP1~5、AP1、PRX7 中发现。第三 类过氧化物酶特异的二硫键形成残基也被观察到: Cys¹¹-Cys⁹¹、Cys⁴⁴-Cys⁴⁹、Cys⁹⁷-Cys³⁰¹、Cys¹⁷⁷-Cys²⁰⁹ (以成熟的辣根过氧化物酶 HRP C1 为参照)^[2],但 OSP3 的 Cys¹¹前移了两位。除此以外,OSP1~5、 AP1 和 PRX7 又表现出一些不同于其他亚家族的新

2

4	144	253		
RIAASILRLH RAGAKIIRLH RMGASILIRLH GIAAGLIRLF GIAAGLIRLF GUAAGLIRLF ALAAGLIRIF ALAAGLIRIF ALAAGLIRIF ALAAGLIRIF ALAAGLIRIF ALAAGLIRLF ALAAGLIRLF ALAAGLIRLF	ANLPAP.FFT SDIPSP.FET SDLPGF.TSS GTLPAP.FFD GLLPAP.FFD GLLPAP.FFD GLLPAP.FFD GDLPGPGTSS NQLPGPGTSS NQLPGPGTSS SDLPGPGTSS SDLPGPS.SN NQLPGPSS.SN	IOSDOELFST LOTDOELFST LHSDOVLFNN FKSDOGLIEH FKSDOGLIEH FYSDMALIKD FTSDMALIKD FTSDMALIKD FTSDMALIKD FTSDOGLTGD FVSDOLFTN MFSDQLFTN	*	
TITVNELRSDP VMDQRQRTDA GVMAAVSSDP HVTEALRRDI HVTEALRRDI HVTEALRRDI HVTEALRRDI SVQAALQGEI SVQAALQGEI AVQAALQGEI AVQAALQCEI TVQEAVRDI VVDSAIDAET	LQ.AFLDLAN LT.ANRSGAN TD.ANEAAAN LAPASS.DLV LAPASS.DLV LAPASS.DLV LAPASS.ALV LAPASL.DLV LAPASL.DLV LAPAPU.RLV LAPAPU.RLV FAPAPU.RLV FAPAPU.RLV FAPAPU.RLV FAPAPU.RLV FAPAPU.RLV FAPAPU.RLV FAPAPU.RLV	YVNLEEQKGL FTNLQSNQGL YTNLMSQKGL YFDLIAKQGL YFDLIAKQGL YVALTRKQGV YVALTRKQGV YVALTRKQGV YVALTRKQGV YSNLVAGQQV YSNLVAGQQV YIDLVNREGL	*	amily gran peptides central con- es(C) ^[2] in- d PRX7 and
CPNVSNIVRD CPNVTSIVRG CPRALAIIKS CPSVEGIVRW CPSVEGIVRW CPSVEGIVRW CPSVEGIVRW CPSLEAIVES CPGLESIVRS CPQLESIVRS CPQLESIVRS CPQLESIVRS CPQLESIVRS CPQLESIVRS	WRVPLGRRDS WQVLFGRRDS WQVLFGRRDS FEVPLGRRDG FDVPLGRRDG FDVPLGRRDG FDVALGRRDG FDVALGRRDG YAVPLGQLDS YPVPLGQLDS YRVPLGRLDS YRVPLGRLDS YSVALGRRDS	RTFTIFONKY STPNDFDNKY FTPNAFDNKY RTPNAFDNKY RTPNAFDNKY RTPNAFDNKY TTPDAFDNKY TTPDAFDNKY TTPDAFDNGY RTPNCFDNKY RTPNCFDNKY	* * ~ ~	t III peroxidase 1 tries. Predicted sig domain(F), and Conserved cystein (OSP1-5, AP1 an
QLTPTFIDNS QLSATFYDTS QLSATFYDTS QLSATFYDTS LSWSFYDTS LSWSFYDTS LSWSFYDTS QLSPNFHAAT .FHAAS	OSVTLAGGPS IGVVLAKGPS DSVVALGGPS DAVVASGGPS DAVVASGGPS DAVVVSGGPS DSLVKAGGPS DAVVVSGGPS DAVVUSGGPT DAVVUSGGPT DAVVUSGGPT DAVVVSGGPT DSVVVSGGPT DSVVVSGGPT	L.SALVDFDL NGNTFTNLDT GDGSLANLDT GDGSLANLDT GDGSLANLDT GNVTQVLDV GNVVTQVLDV GNVDV NNSVTQELDV NNSVTQELDV NNSVTQELDV RLONLDV RLONLDV TKQDLDV TLDV DRRTVLDV	M	dasses from class o maximize similar imal heme binding indicated by * . nserved residues in
ILHAS. LSDA VALFFG, ASNAS VALAT. AASA VALAT. AASA TUREP. LANG TTGEP. VVAG CAAVH. SSEG AAAEL. SVD. AAAEL. SVD. AAAEL. SVD. AAAEL. SVD. ITTPP. LADG SAAEP. PVAP NNNKKNSDK	CADLLTIAAO CADILLTIAAO CADILLTVAAR CADILTVAAR CADITTLATR CADITTLATR CADITTLATR CADISALATR CADISALATR CADITALATR CADITALATR CADILALAR CADILALAR CADILALAR CADILALAR CADILALAR CADILALAR	LRGLCPLNGN LOGICPOGGN LOGICPOGGN LTAKCASDPS LQAKCAKDVP LRNKCAGDNP LRNKCAGDNP LANCTKDPN LANCTKDPN LANCTKDPN LANCSNDGS LKRTCPVKGT LQCNCSATLT	MVEVUPFVSS 	nd other peroxi oduction of gaps to lomain (B), proxi tre highlighted and tre highlighted and dicated by \pm . Cor
ITLI. PLVCL LRFVGAILFL SASCLSLVVL SASCLSLVVL AAALSS. ATV FAARCA. AVV AAALSS. ATV AAALSS. ATV AAALSS. ATV AAALSS. ATV LLISSPT A FISLPTYHLP FISLPTYHLP FISLPTYHLP ESGVASILTL	ESACPR. TVB ENVCPG. VVB ENVCPG. VVB EAVC. GPTVB HAAC. GPTVB HAAC. GPTVB HAAC. GPTVB HAAC. GPTVB HAAC. GPTVS HAAC. GPTVS IAAC. GPTVS IAAC. GPTVS HAAC. GPTVS HAAC. GPTVS	PTLNTTYLQT LTVDATFLQT TNLDTAFRATS TMDPAFRQR GTMDPAFRQR PIMDPVLVKRR PIMDPVLVKR PIMDFVLVKK PITSKK DDTFSKK PTISPTFLSR PAAQ	VVNSNSLLHD RVNS RVNS RVNS RVNSSDLL VFNRRSSDLL VFNRRSSDLL RTN. SQSLVD RTN. SQSLVD RTN. SQSLVD RTN. SQSLVD RTN. SQSLVD RTN. SQSLVD RTN. SQSLVD RTN. VPNSQTIL.	idases OSP1-5 a en aligned by intr tal heme binding c ass III peroxidase a nighlighted and ind
FSSSSTLFTC 	PVIDRMKAAV DIVDDIKTAL GVIDSIKTQI KLIDDIRAAW KLIDDIRAAW KLIEDIRAAW DLIERIRVAW QLVEDIRAKW QLVEDIRAKW QLVEDIRARW QLVEDIRARW CLUEDIRACW KAVNDIRDRL EVIAQAKQSV	YNFSNTGLPD FNFNGSGNPD YNEPNAD.D LPPNAD.D FFDGS.K FRPVFDTN ARRQ .VRPV FR.E LFPRPD	QGQIR.LNCR NGQIR.LNCK QGQIR.LSCS AGEIRR.NCS AGEIRR.NCS KGEIRR.NCA VGEIRR.SCF VGEIRR.SCF VGEIRR.NCF VGEIRR.NCF VGEIRR.NCF QGEVRR.NCS QLEIRDV.CS	clustered peroxi code and have be d catalytic and dist ved residues in cla U peroxidses are h
	NANSAR. GF NV. GAG. GF NV. GAG. GF N. OTLRPEAL N. OTLRPEAL N. OTLRPEAL N. OTLRPEAL N. TLOPRAL NL. TLOPRAL NL. TLOPRAL NL. TLOPRAL NL. TLRPSAF	NOCRF IMDRL ARCGTFEORL ACCGTFEORL AOCGTFEORL GHCGSFNDR GHCGSFNDR AHCDFFEDR TRCAFFDR AHCDFFRDR ARCSSFSNR AHCSSFEDR AHCSSFEDR	MG# MG# LGNISPLTGT MGNIAPKTGT MGNIAPKTGT MGNIAPKTGT MSNMDILTGS MSNMDILTGR MSNMDILTGR LSXUPPRDRN LSXUPPRGGN LGQLQGPSGN MGQMRVRTSD MGDLPPSAGA	nparison of the given in one-letter epresent conserver r function. Conser e bridges in class
······································	SFRTEKDAFG GTQTEKDAPA GTQTEKDAPA MEONAIP N. MEONAIP O. SELGEIP G. SELNEIP G. SKLNEIP G. SKLNEIP N. SKOGMGP N. SKOGMGP N. SKOGMPP N. SKOGMPP TFTGKOOAPP	ALSGGHTFGK ALSGAHTFGR ALSGAHTTGCI ALSGAHTTGCI ALSGAHTTGCI ALSGAHTTGCI ALSGGHTTGCI ALSGGHTTVGR ALSGGHTTVGR ALSGGHTTVGR ALSGGHTTVGR ALSGGHTTGL	FINAFVEAMDR FDDFVSSMIK FDDFVSSMIK FDDFVSSMIK FDDFAKSMVK FDDFAKSMVK FTDFAKSMVK FTOFAKSMVK FTOFAKSMVK FTOFAKSMVK FTOFVSIVK FTOFVSIVK FTOFVSIVK FTOFVSIVK FTOFVSIVK FTOFVSIVK FGOFGSSMVK FGOFGSSMVK	ignment and con cid sequences are Horizontal lines r in (D) of unknow mation of disulfid
ALSFVALALA	ASTILLDN.TT GSTILLD.TD ASVILLSG. ASVILLSG. ASVILLS.GS ASVILLT.GS ASVILLT.GS ASVILLT.GS ASVILLT.GS ASVILLT.GA ASVILLT.GA ASVILLT.GA ASVILLT.GA ASVILLT.GA ASVILLT.GA	VCLNRSSDLV KCMD.LTDLV KCCL.LTDMV RCCL.LTDMV RCL.LTDMV RCL.LDKADLV RCL.DVADLV RCLRDADLV RCLRDADLV RCLREAADLV RCLREAADLV RCLREAADLV RCLREAADLV RCLREAADLV RCLREAADLV RUL.DNVDLV RUL.DNVDLV KNFT.LREAV	RSPANSTOTT NRYAGSOTOF NRYAGSOTOF TRFASNPAAF TRFALNDAAF TRFALNOAAF TRFSLNOAAF TRFSLNOAAF ROFATDKAAF ROFATDKAAF ROFADDKAAF RRFAQDKAAF NGFAQDKAAF NGFAQNHWWF TDYSNDVNVF TDYSNDVNVF	Fig. 3 All The armino 8 are in italic. served doma volved in for highlighted.
MGFRLISHLSL	FHDCFVNGCI FHDCFVNGCI FHDCFVQGCI FHDCFPQCI FHDCFPQCCI FHDCFPQCCI FHD	LPOLKDSFRY LAVMIPOFTN LAVMIPOFTN VPTLIESFKN VPTLIESFKN VPTLISSFAN VPTLISSFAN VPTLISSFAN VPTLISSFAN VOLLINKFGS VOALIDKFGS VOALIDKFGS VOALIDLFGS VSTLLSAFOT VOSLLALLGR	PNATDTIFLV SGSATIAIV DTTDNTV DTTDNTV PUTKRAA AQTKRAA PJTKRMA RITAPIV PQTAPIV PQTAPIV WRTAPIV MRTRPIV ATTRPIV	
HRPC1 TOPA WP1 WP1 WP1 OSP3 OSP1 OSP2 OSP2 OSP3 OSP3 OSP3 OSP3 OSP3 OSP3 OSP3 OSP3	HRPC1 TOPA WP1 WP1 WP1 WP1 OSP1 OSP1 OSP2 OSP2 OSP2 OSP2 OSP2 OSP2 OSP2 OSP2	HRPC1 TOPA WP1 WP1 WP1 WP1 OSP1 OSP1 OSP2 OSP2 OSP3 OSP3 OSP3 OSP3 OSP3	HRPC1 TOPA WP1 WP1 WP1 0SP1 0SP1 0SP1 0SP2 0SP2 0SP2 0SP2 0SP2 0SP2 0SP2 1AP1 TAP1	

osp5

Table 1	Matched	ESTs(> 95	% identity)	for	osp1,	osp3 and

-				
ACC. No.	Gene	Tissue	cDNA clone	Score
AU075805	osp1	Immature leaf	E60486	531
AU030963	osp1	Immature leaf	E60486	1168
D25030	osp1	Root	R2957	529
AU031972	osp1	Root	R2957	513
AU075454	osp1	Immature leaf	E60731	537
AU031073	osp1	Immature leaf	E60731	682
BE039997	osp1	Root		519
		~ "		
AU091519	osp3	Callus	C11468	785
AU062505	osp3	Callus	C11468	446
D48606	osp3	Green shoot	S14924	446
C20540	osp3	Green shoot	S14924	753
D41595	osp3	Etiolated shoot	S4191	329
C20517	osp3	Etiolated shoot	S4191	595
AU031487	osp3	Immature leaf	E61710	454
BE229129	osp3	Immature seed	98BS0186	729
BE229186	osp3	Immature seed	98BS0275	676
BE229213	osp3	Immature seed	98BS0313	410
1 5010742	-			1.600
AF019745	osp5	-		1602
D24571	osp5	Root	R2184	486
AU031855	osp5	Root	R2184	636
C20483	osp5	Root	R0894	759
D24028	osp5	Root	R0894	569
D20490	osp5	Root	R1777	559
C20491	osp5	Root	R1778	700
BE039701	osp5	Root		1134
BE040760	osp5	Entire plant		579
BE039233	osp5	Root		551
AW155066	osp5		mgie0001B21f	537
BE607327	osp5	Root		452

Table 2GC content of the introns/exons of the clusteredgenes, osp1-5

Gene		Length(bp)	GC content (%)
osp1	Exon1	228	66.7
	Intron	74	43.2
	Exon2	789	65.8
osp2	Exon1	228	65.8
	Intron	90	45.5
	Exon2	780	66.7
osp3	Exon1	243	67.9
	Intron	143	32.2
	Exon2	768	67.9
osp4	Exon1	273	64.8
	Intron1	93	34.4
	exon2	186	71.5
	Intron2	81	40.7
	Exon3	588	66.8
osp5	Exon1	237	67.5
	Intron1	153	35.9
	Exon2	186	69.4
	Intron2	104	25.0
	Evon3	618	66 0

的结构特征:具有一些不同于其他亚家族的保守残基,例如 $IIe(I)^{39}$ 、Phe(F)⁴⁰、Pro(P)⁴⁶、Gly(G)⁵⁸、

	OSP1	OSP2	OSP3	OSP4	OSP5
AA length	338	335	336	348	346
Potential signal pep- tide cleavage site	28	28	31	24	22
Estimated pI for ma- ture protein	5.76	8.17	8.18	4.69	5.28
Estimated molecular mass of mature pro- tein(kD)	33.3	32.7	32.1	34.7	34.8
Probable cellular location	Extracellular peroxidase				

Leu(L)⁷⁷、Leu(L)⁷⁹、Val(V)⁸⁷、His(H)⁸⁸等,以及一 个比较保守的特征性 C 末端序列: A(T)AS(D)A (M/P)。这表明, OSP1~5、AP1、PRX7 形成了第三 类植物过氧化物酶家族中一个新的分枝。

2.6 过氧化物酶的比较分析

对 5 个过氧化物酶及家族中部分有代表性的成员进行了两两比较分析。五个过氧化物酶间的相似程度按 OSP1~OSP5 呈梯度下降(表 4): $\alpha sp1$ 与其下游 4 个基因在氨基酸水平的相同率依次为: 87%、75%、61%、53%; $\alpha sp2$ 与其下游基因的相同率为: 72%、56%、51%; OSP3 与 OSP4、OSP5 的相同率分别为 56%、51%。该结果表明,一系列的串联基因复制导致了该基因簇的形成,即在进化过程中依次发生了 $\alpha p5-\alpha sp4-\alpha sp3-\alpha sp2-\alpha sp1$ 的基因复制事件。

将 5 个过氧化物酶氨基酸序列用于蛋白质数据 库检索发现, OSP5 与来自玉米的阴离子过氧化物 酶 AP1 (GenBank 登录号 Y13905)^[22]相同率为 72.54%, 与来自大麦(Hordeum vulgare)的 PRX7 (GenBank 登录号 AJ003141)^[23]相同率为 63.02%, 而 OSP1、OSP2、OSP3、OSP4 与数据库中所有过氧 化物酶的相同率在 55%以下。这表明 asp5、ap1、 prx7为潜在的直向同源基因。

对 OSP1~5, AP1, PRX7, 以及第三类过氧化 物酶中几个亚家族代表进行进化分析(图 4), 结果 显示, *asp*1~5 在进化上发生的先后顺序为从 *asp*5 依次到 *asp*1; 而且, *asp*1~5, *ap*1, *prx*7 与其他家 族代表有较大差距, 可归为一个新的分枝。

3 讨论(Discussion)

到目前为止,所发现的第三类过氧化物酶基因 都具有1~4个外显子,而且内含子插入位点严格 保守。在日本构建的水稻高密度遗传图谱中共鉴定

lanny												
	OSP1	OSP2	OSP3	OSP4	OSP5	PRX7	AP1	BP1	WP1	TOPA	HRPC1	TAP1
		87.13	75.38	60.84	53.75	49.39	55.09	47.61	43.79	38.11	38.65	33.12
OSP1		89.22	78.42	65.66	58.26	54.55	59.28	52.70	51.63	45.93	46.32	38.80
			71.73	56.97	51.66	47.85	54.46	50.15	43.61	37.05	35.15	34.18
OSP2			76.29	63.03	57.40	53.07	58.77	54.10	50.82	44.59	42.42	40.51
				56.46	51.52	46.11	55.79	46.34	40.26	36.63	38.34	33.97
OSP3				62.16	57.62	50.78	59.76	52.44	48.19	44.22	46.32	42.22
					55.82	48.77	54.73	48.34	43.36	35.29	36.96	36.45
OSP4					60.00	53.37	60.06	52.57	50.00	43.79	44.41	43.98
						63.02	72.54	48.83	40.51	38.66	38.02	32.41
OSP5						67.46	76.59	53.51	45.34	46.01	43.71	38.89
							64.41	45.73	43.41	38.99	38.21	32.69
PRX7							67.65	50.61	47.27	46.23	43.58	37.82
								55.09	46.05	38.80	41.42	33.44
AP1								59.28	51.97	45.74	46.45	38.70
									45.16	43.18	39.00	34.04
BP1									50.97	51.11	46.33	41.64
										46.45	50.64	42.05
WP1										53.87	55.77	49.34
											49.07	42.62
TOPA											57.45	48.53
												38.75
HRPC1												44.38

Table 4 Percentage similarities and identities between the clustered peroxidases and representatives from the class III peroxidase family

Percent amino acid identities are shown in row 1 and similarities are shown in row 2. Representatives from class III peroxidase family were included: BPI (M73234)^[24], WPI (X56011)^[25], TOPA (J02979)^[26], HRPC1 (M37156)^[27], TAPI (X15853)^[28].



Fig. 4 Phylogenetic analysis of amino acid sequences of OSP1-5 and other class III peroxidase family members

The tree was constructed using Grow Tree (GCG package 10.1).

出 26 个过氧化物酶位点^[11]。本文对其中的一个位 点标记(R2184S)所锚标的 BAC 克隆 q3037 (H0207F01)进行了全序列分析,发现除 R2184S 所 对应的过氧化物酶基因以外,在其上游还存在 4 个 紧密排列的过氧化物酶基因,5 个基因具有相同的 转录方向(图 1)。5 个基因分别命名为 $\alpha p1, \alpha p2, \alpha p3, \alpha p4, \alpha p5,$ 其中 $\alpha p5$ 对应于位点标记 R2184S。这是到目前为止所报道的最大的一个过氧 化物酶基因簇。

在大规模基因组测序过程中,利用数据库中大 量积累的 EST 数据进行基因组序列中基因结构的 注解已变得越来越重要^[19]。本文在软件预测的基 础上,再结合与相应 EST 的匹配,准确地鉴定了 5 个过氧化物酶基因的外显子/内含子组织结构: *asp*1,*asp*2,*asp*3分别含 1 个内含子,而 *asp*4,*asp*5 分别含 2 个内含子。所有内含子插入位点与其他第 三类过氧化物酶基因一致^[3,10],而且内含子都具有 较高的 AT 含量。高 AT 含量是植物基因内含子所 普遍具有的一个特征,被认为与内含子的有效剪切 有关^[20]。5 个基因中多数内含子所遵循的是 GC/ AG的剪切模式,这种模式在较少一些基因中被观 察到过^[4]。

从与 asp1, asp3, asp5 相匹配的 EST 来源,还 可揭示 asp1, asp3, asp5 的组织表达差异(表 1)。 asp1, asp3 可在根、叶等组织表达,而 asp5 主要在 根中表达。而数据库中却没有与 asp2 和 asp4 相匹 配的 EST,这表明 asp2 和 asp4 可能不表达或者极 低表达,也可能在极为特殊的条件下表达,这有待 于进一步的研究确证。

5 个基因分别编码 338、335、336、348 和 346 氨 基酸残基的蛋白质,都具有 N 端信号肽序列,属于 分泌性植物过氧化物酶。5 个蛋白质中 OSP2、OSP3 为阳离子过氧化物酶,而 OSP1、OSP4、OSP5 为阴 离子过氧化物酶。但在数据库检索时我们发现与 *asp5* 相匹配的 *ascpx*1(GenBank 登录号 AF019743) 被注解为阳离子过氧化物酶基因。经仔细比较后发 现,在相对于 *asp5* 终止密码上游 41 位处 *ascpx*1 多 引入了一个 t,由此导致了移框,使得表达产物成为阳离子过氧化物酶。根据几方面的证据判断,我 们倾向认为 $\alpha c p x 1$ 的序列存在错误: $\alpha s p 5$ 的序列 来源于高丰度、高质量的基因组测序,而 EST 或 cDNA 序列通常为低丰度的、单向的测序结果;数 据库中相应的 EST 在该位点上与 $\alpha s p 5$ 一致 [图 2 (E)]。

5 个过氧化物酶都具有植物过氧化物酶的一些 基本结构特征,如两个保守的血红素结合区(B和F 区,图3),以及功能未知的D区;同时,又具有一 些新的结构特征,如一些其他过氧化物酶所不具有 的保守残基,以及一个保守的特征性的C末端序列 (图3)。再结合对过氧化物酶氨基酸序列的两两比 较分析和进化分析结果,我们认为:本文所报道的 过氧化物酶基因簇起源于 $\alpha p5$,经一系列的串联基 因复制事件所形成,同时在复制过程中丢失了一个 内含子($\alpha p4 \sim \alpha p3$),在过氧化氢酶基因家族 (cat)进化过程中也曾发生过类似的现象^[29]; $\alpha p5$ 与 ap1、prx7为潜在的直向同源基因,分别在各自 的基因组中行使相同的功能; $\alpha p1 \sim 5$ 、ap1和 prx7构成了第三类过氧化物酶基因家族中一个新 的分枝。

另外,我们观察到: OSP5 与 AP1 的相似性 (76.59%)高于 OSP5 与 OSP4 的相似性 (60.00%)、OSP4 与 OSP3(62.16%)的相似性;而 与 OSP3 和 OSP2 的相似性水平相近。由此我们推 测, *asp5-asp4-asp3* 的基因复制事件发生在水稻和 玉米基因组分化之前,即在现在的玉米基因组中应 当存在一个与本文报道的基因簇相对应的基因簇, 在该基因簇中除 *ap1* 以外,至少还存在分别与 *asp4、asp3* 互为直向同源基因的两个过氧化物酶基 因。同理可推测,在大麦基因组中也存在一个相应 的基因簇,其中除 *prx7* 以外,至少还存在与 *asp4* 互为直向同源基因的过氧化物酶基因。以上推论建 立在以下的前提上:不同的基因在不同的基因组中 进化速率大致恒定。至于真实情况如何则有待于进 一步研究确定。

植物过氧化物酶家族非常庞大,阐明整个家族 的结构及进化有赖于更多家族成员的发现,而对结 构进化的了解将会对认识基因功能及演化提供线 索。尽管随着大规模基因组测序的展开,数据库中 积累了大量的过氧化物酶基因数据,但距离全局性 的认识过氧化物酶基因家族还有一定差距。本文在 现有的基础上,通过对水稻第4号染色体长臂近端 粒区一个过氧化物酶基因簇的结构及进化分析,为 认识过氧化物酶基因家族的结构及进化提供了一种 范例。

References

- Dunford H B. Horseradish peroxidase: Structure and kinetic properties. Everse J , Everse K E , Grisham M B eds. *Peroxidases in Chemistry and Biology*, Boca Raton: CRC press, 1991, 2:1–23
- 2 Welinder K G. Superfamily of plant, fungal and bacterial peroxidase. Curr Opin Struct Biol, 1992, 2: 388-393
- 3 Justesen A F, Jespersen H M, Welinder K G. Analysis of two incompletely spliced A rabidopsis cDNAs encoding novel types of peroxidase. Biochi et Biophy Acta, 1998, 1443: 149-154
- Sutliff T D, Huang N, Litts J C, Rodriguez R L. Characterization of an -amylase multigene cluster in rice. *Plant Molec Biol*, 1991, 16: 579-591
- 5 Dean C, Pichersky E, Dunsmuir P. Structure, evolution, and regulation of *rbcs* genes in higher plants. Annu Rev Plant Physiol Plant Molec Biol, 1989, 40: 415-439
- 6 Hillebrand H, Bartling D, Weiler E W. Structural analysis of the nit2/ nit1/ nit3 gene cluster encoding nitrilases, enzymes catalyzing the terminal activation step in indole-acetic acid biosynthesis in A rabidopsis thaliana. Plant Molec Biol, 1998, 36: 89-99
- 7 He Z H, Cheeseman I, He D, Kohorn B D. A cluster of five cell wall-associated receptor kinase genes, wak1-5, are expressed in specific organs of Arabidopsis. Plant Molec Biol, 1999, 39: 1189-1196
- 8 Maeda N, Smithies O. The evolution of multigene families: Human haptoglobin genes. Ann Rev Genet, 1986, 20: 81–108
- 9 Johansson T, Nyman P O. A cluster of genes encoding major isozymes of ligin peroxidase and manganese peroxidase from the white rot fungus *Trametes versicolor*. Gene, 1996, **170**: 31-38
- 10 Osakabe K, Koyama H, Kawai S, Katayama Y, Morohoshi N. Molecular cloning of two tandemly arranged peroxidase genes from *Populus kitakamiensis* and their differential regulation in the stem. *Plant Molec Biol*, 1995, 28: 677–689
- Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T et al. A high-density rice genetic map with 2275 markers using a single F₂ population. *Genetics*, 1998, 148: 479 494
- 12 Hong G, Qian Y, Yu S, Hu X, Zhu J, Tao W, Li W et al. A 120 kilobase resolution contig map of the rice genome. DNA Seq, 1997, 7(6): 319-335
- 13 Roach J C, Boysen C, Wang K, Hood L. Pairwise end sequencing: A unified approach to genomic mapping and sequencing. *Genomics*, 1995, 26(2): 345-353
- 14 Ewing B, Hillier L, Wendl M C, Green P. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res, 1998, 8(3): 175-185
- 15 Ewing B, Green P. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.*, 1998, 8(3): 186-194
- 16 Altschul S F, Madden T L, Schaffer A A, Zhang J, Zhang Z, Miller W, Lipman D J. Gapped BLAST and PSFBLAST: A new

generation of protein database search programs. *Nucleic Acids Res*, 1997, **25**(17): 3389-3402

- 17 Fickett J W. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res, 1982, 10(17): 5303-5318
- 18 Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol, 1997, 268: 78-94
- 19 Bailey L C, Searls D B, Overton G C. Analysis of EST-driven annotation in human genomic sequence. *Genome Res*, 1998, 8: 362-376
- 20 Luehrsen K R, Walbot V. Addition of A- and U-rich sequence increases the splicing efficiency of a deleted form of a maize intron. *Plant Molec Biol*, 1994, 24(3): 449-463
- 21 Welinder K G. Plant peroxidases. Their primary, secondary and tertiary structures, and relation to cytochrome peroxidase. *Eur J Biochem*, 1985, **151**: 497 - 504
- 22 Teichmann T, Guan C, Kristoffersen P, Muster G, Tietz O, Palme K. Cloning and biochemical characterization of an anionic peroxidase from *Zea mays. Eur J Biochem*, 1997, 247: 826-832
- 23 Kristensen B K, Bloch H, Rasmussen S K. Barley coleoptile peroxidases. Purification, molecular cloning, and induction by

pathogens. Plant Physiol, 1999, 120(2): 501-512

- 24 Johansson A, Rasmussen S K, Harthill J E, Welinder K G. cDNA, amino acid and carbohydrate sequence of barley seed-spcific peroxidase BP1. *Plant Molec Biol*, 1992, 18: 1151–1161
- 25 Rebmann G, Hertig C, Bull J, Mauch F, Dudler R. Cloning and sequencing of cDNAs encoding a pathogen-induced putative peroxidase of wheat (*Triticum aestivum* L). *Plant Molec Biol*, 1991, 16(2): 329-331
- 26 Lagrimini L M, Burkhart W, Moyer M, Rothstein S. Molecular cloning of complementary DNA encoding the lignin-forming peroxidase from tobacco: Molecular analysis and tissue-specific expression. *Proc Natl Acad Sci USA*, 1987, 84: 7542–7546
- 27 Fujiyama K, Takemura H, Shibayama S, Kobayashi K, Choi J K, Shinmyo A, Takano M *et al.* Structure of the horseradish peroxidase isozyme *c* genes. *Eur J Biochem*, 1988, **173**: 681–687
- 28 Roberts E, Kolattukudy P E. Molecular cloning, nucleotide sequence, and abscisic acid induction of a suberization-associated highly anionic peroxidase. *Mol Gen Genet*, 1989, 217 (2-3): 223-232
- 29 Frugoli J A, McPeek M A, Thomas T L, McClung C R. Intron loss and gain during evolution of the catalase family in angiosperms. *Genetics*, 1998, 149: 355-365

Structural Analysis of a Gene Cluster Encoding Two Cationic and Three Anionic Peroxidases from Rice Chromosome 4

CHEN Ze-Hua, ZHOU Bo, HAN Bin, QIAN Yue-Min, HONG Guo-Fan *

(National Center for Gene Research, Chinese Academy of Sciences, Shanghai 200233, China)

Abstract Sequence analysis of a rice BAC q3037 (H0207F01) identified a cluster of five tandemly arranged peroxidase genes, $\alpha sp1$, $\alpha sp2$, $\alpha sp3$, $\alpha sp4$ and $\alpha sp5$, within a 22.5 kb region. $\alpha sp4$, $\alpha sp5$ each have three exons interrupted by two introns, while $\alpha sp1$, $\alpha sp2$ and $\alpha sp3$ each have two exons interrupted by a single intron. The five genes were predicted products of 338, 335, 336, 343 and 346 amino acid residues, respectively, including putative signal peptide sequence at the amino-termini. And OSP1, OSP4 and OSP5 were predicted to be anionic peroxidase, OSP2 and OSP3 are cationic. Comparative analysis and evolutionary analysis of the clustered genes and other peroxidase family members revealed that the gene cluster occurred by tandemly gene duplications (from $\alpha sp5$ to $\alpha sp1$); and that $\alpha sp5$, $\alpha p1$ and prx7 were potential orthologies, and $\alpha sp1-5$, $\alpha p1$ and prx7 constituted a novel evolutionary branch of class III peroxidases.

Key words peroxidase; gene cluster; tandemly gene duplication; ortholog

Received: November 8, 2000 Accepted: December 21, 2000

^{*} Corresponding author: Tel, 86-21-64516371; Fax, 86-21-64825775; e-mail, gfhong @newnetra.ncgr.ac.cn