EXO70在拟南芥和水稻基因组中的倍增

杨昆^{①②}, 张毅^{①②*}, 吕俊^{①②}, 赵永斌^③, 张贺翠^①, 韩叙^①, 何光华^{①②*}

① 西南大学农学与生物科技学院, 重庆 400715;

② 西南大学水稻研究所, 重庆 400715;

③ 吉林师范大学生命科学学院, 四平 136000

* 联系人, E-mail: zhang569@swu.edu.cn; hegh@swu.edu.cn

2014-08-13 收稿, 2014-09-23 接受, 2014-11-20 网络版发表

中央高校基本科研业务费(XDJK2012B018, SWU112048, 2362014xk09)、重庆市基础与前沿研究计划(2013jcyjA80010, 2013jjB80003, 2012jjA80014)、国家自然科学基金(30970274, 31370349)和高等学校学科创新引智计划(B12006)资助

摘要 作为胞泌复合体的一个重要组件, EXO70在人类、小鼠及酵母中均由单基因编码. 为研究EXO70在拟南芥和水稻基因组中倍增的异同, 从拟南芥和水稻基因组中钓取所有的EXO70家族基因, 分析后发现, 拟南芥中包含23个EXO70基因, 其中15个基因位于染色体的负链; 选择性剪接的存在共产生27种成熟的mRNA. 水稻中共发现47个EXO70基因座, 它们分别分散于12条染色体的正负链, 其中的29个基因位于正链; 分别包含OsEXO70H3和OsEXO70H4在内的11号染色体和12染色体的5'端存在一段约 2×10^6 bp区段的序列高度同源; OsEXO70基因倍增的复杂性高于拟南芥EXO70. EXO70基因在拟南芥和水稻中均存在密码子使用偏好性, AtEXO70和OsEXO70的密码子偏好性存在较大的差异. Pfam03081是EXO70蛋白质共有的结构域, 且在三维结构上极为保守, 这可能决定不同的EXO70蛋白存在相同或相近的功能. 结果表明, EXO70在拟南芥和水稻2个物种进化中均存在多次加倍, OsEXO70倍增的次数、方式和复杂性都高于AtEXO70; 高度保守的Pfam03081可能是决定EXO70蛋白功能的关键因素; 像其他基因一样, EXO70同义密码子的偏性模式存在明显的物种特异性.

关键词

EXO70
pfam03081
基因倍增
密码子偏好性

胞泌作用在所有的真核细胞中普遍存在并且发挥着重要的作用, 如动物细胞的生长、细胞的极化、神经传递、植物和真菌细胞壁的形成^[1]、根尖和花粉管的极性生长等都需要经历胞泌过程将原料或成分由胞内运至细胞膜或者胞外^[2,3]. 胞泌作用涉及高尔基体衍生的胞泌囊泡和质膜的融合, 融合发生时囊泡膜插入细胞膜中, 同时朝向细胞外侧的囊泡膜张开, 而将囊泡中的内含物释放到胞外^[4]. 将高尔基体衍生的分泌囊泡拴系至质膜上需要胞泌复合体的参与^[5,6]. 这个保守的复合体最先发现于酵母细胞, 它在酵母细胞出芽生殖过程的分泌和极性生长中发挥着重要的作用^[7,8]. 随后它的同源复合体在哺乳动物

细胞中也被鉴别^[9]. 植物的胞泌复合体首先在拟南芥基因组中被鉴定^[10], 后来水稻和小立碗藓中也鉴别出胞泌复合体^[11]. 研究表明, 动物和酵母细胞的胞泌复合体均由Sec3, Sec5, Sec6, Sec8, Sec10, EXO84和EXO70等亚单位组成, 它们定位于活性囊泡与胞膜发生融合时的胞膜的特定位置^[12~15]. 作为胞泌复合体的重要组件, 在缺少胞泌复合体介导的囊泡拴系信号时, EXO70被定位于哺乳动物未分化的PC12细胞的核周质区^[16]. 在细胞膜上发生囊泡拴系时, EXO70在目标上的定位在一定程度上依赖F-肌动蛋白^[12,14]. 后来发现SEC3和EXO70在细胞膜的定位依赖胞膜内侧的4,5-二磷酸肌醇, 在这3种物质的协同

引用格式: 杨昆, 张毅, 吕俊. EXO70在拟南芥和水稻基因组中的倍增. 科学通报, 2015, 60: 38-51

Yang K, Zhang Y, Lü J. Comparison of EXO70 duplication between *Arabidopsis thaliana* and *Oryza sativa* (in Chinese). Chin Sci Bull, 2015, 60: 38-51, doi: 10.1360/N972014-00853

作用下建立一个朝向细胞膜靶点的极性定位,再与其他6个胞泌复合体组件互作而将胞泌囊泡栓系至细胞膜上^[17-19]。

基因重复是基因组中特定DNA序列的多个拷贝的重复排布,是在生物进化过程中由基因倍增或者基因组倍增引起的。EXO70和其他7个胞泌复合体组件在动物和酵母中均由单基因编码^[10]。在拟南芥基因组中除SEC6和SEC8,包括EXO70在内的其他6种胞泌复合体组件都由多基因编码^[10,11,18,20,21]。作为模式作物的拟南芥所具有的这种特性,自然引发人们对胞泌复合体基因在是否在其他植物基因组中出现基因倍增现象的思考。考虑到水稻是研究的热点作物,同时EXO70蛋白是胞泌作用的重要组件,我们通过分析已知的拟南芥和水稻全基因组序列,从而探讨水稻基因组中的EXO70及其倍增现象,并且进一步比较EXO70基因的重复在水稻以及拟南芥中的异同。

1 方法

(i) 预测和鉴别拟南芥和水稻EXO70蛋白及对应的基因。从NCBI的RefSeq蛋白数据库中下载4种已知的EXO70蛋白(表1)。应用Vector NTI Advance 11.5对下载的序列进行比对,找出高度保守的蛋白区段。以保守区段氨基酸序列为探针,运用blastP搜索拟南芥蛋白质数据库(<http://www.arabidopsis.org/Blast/index.jsp>),将搜索并鉴定的EXO70相关蛋白(表2)作为查询的探针序列对NCBI中水稻核苷酸数据库(NR/NT),HTG,WGS和GSS数据库进行TblastN比对,以钓取水稻基因组中所有的可能EXO70家族基因。同时以AtEXO70蛋白和对应编码序列用Blast先后搜索密歇根大学水稻基因组注解数据库(RGAP, http://rice.plantbiology.msu.edu/analyses_search_blast.shtml)和NCBI的水稻基因组数据库(http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome,数据库选择*O. sativa (japonica cultivar-group)* genome

(reference only)和*O. sativa (japonica cultivar-group)* RefSeq protein数据库),运用PSIBLAST搜索NCBI蛋白质数据库以检验搜索到的序列;并用Blast Like Alignment Tool (BLAT)^[22]和基因组定位去除重复和错误的基因模型。最后,利用搜索预测蛋白的mRNA或编码序列(CDS)再次利用Blast搜索水稻基因组数据库,对所有预测的拟南芥和水稻EXO70基因进行定位。

(ii) EXO70蛋白及编码基因的序列分析。应用Vector NTI Advance 11.5对目标蛋白质进行多序列比对,并运用MEGA5.2^[23]构建蛋白质序列的分子进化树,根据构建的进化树对EXO70基因进行命名。构建进化树的参数见表2。

(iii) 保守结构域分析。利用SMART工具(<http://smart.embl-heidelberg.de>)进行功能域预测及分析,用SOPMA (<http://pbil.ibcp.fr/>)对其进行二级结构分析,用SWISS-MODEL工具采用同源建模法(<http://swissmodel.expasy.org/>)对其三级结构进行分析,并用PyMOL 1.7处理所得的蛋白质三维结构图。

(iv) 密码子使用分析。利用CodonW 1.4.2 (<http://codonw.sourceforge.net/>)分别对水稻和拟南芥基因组中预测出的47种OsEXO70与23种AtEXO70蛋白编码序列和对应的全长基因序列进行密码子组成和使用偏好性分析^[24]。密码子使用偏好性的主要分析指标包括RSCU(同义密码子相对使用度)、ENC(有效密码子数)、CAI(密码子适应指数)等^[25-29]。

2 结果

2.1 拟南芥EXO70基因倍增及其编码蛋白的分析

拟南芥基因组共有23种EXO70基因(表3),它们分别分布于5条染色体,其中9种EXO70基因位于5号染色体,3号和1号染色体均包含5种不同的EXO70基因,2号染色体上包含3种EXO70基因,4号染色体上仅有1种EXO70基因。这表明EXO70基因在染色体位置的选择方面存在不均衡性。分析基因编码的方向后发现,8个EXO70基因位于DNA的正链,15个EXO70基因位于DNA的负链。在外显子构成方面,不同类别的AtEXO70基因有着较大的差异(表3)。

23种AtEXO70基因能够编码26种蛋白,其中的3种基因(EXO70A1, EXO70E2和EXO70H7)均存在转录后选择性剪接的现象(图1)。EXO70A1对应的cDNA序

表1 下载的4种EXO70蛋白

Table 1 Four EXO70 from NCBI protein database

物种	识别码	氨基酸数目
<i>S. cerevisiae</i>	NP_012450.1	623
<i>H. sapiens</i>	NP_001013861.1	684
<i>M. musculus</i>	NP_058553.2	697
<i>D. melanogaster</i>	NP_648222.3	693

表 2 蛋白质比对方法与进化树构建参数的设置
Table 2 Method of protein alignment and parameters of molecular phylogenetic analysis

比对方法	统计方法	系统发生检测		替代模型		Rates and Patterns		G/MDT
		TOP	NOBR	ST	M	RAS	PAL	
Clustal W	邻接法	Bootstrap方法	1000	氨基酸	泊松模型	Uniform rates	Same (Homogeneous)	部分剔除

a) TOP, Test of phylogeny; NOBR, No. of bootstrap replications; ST, Substitutions type; M, Model/method; RAS, Rates among Sites; PAL, Pattern among Lineages; G/MDT, Gaps/Missing data treatment

表 3 预测的拟南芥EXO70 蛋白及编码基因
Table 3 EXO70 genes of *Arabidopsis thaliana* and the corresponding proteins

基因名	基因座	基因模式	外显子数	RNA识别码	染色体	链	蛋白质识别码	氨基酸数	Pfam03081 (start to end)
<i>AtEXO70A1-1</i>	AT5G03540	AT5G03540.1	12	NM_120434.4	5	+	NP_195974.2	638	267~627
<i>AtEXO70A1-2</i>	AT5G03540	AT5G03540.2	11	NM_001125690.1	5	+	NP_001119162.1	523	152~512
<i>AtEXO70A1-3</i>	AT5G03540	AT5G03540.3	13	NM_001203287.1	5	+	NP_001190216.1	664	293~653
<i>AtEXO70A2</i>	AT5G52340	AT5G52340.1	11	NM_124613.3	5	+	NP_200047.3	631	267~620
<i>AtEXO70A3</i>	AT5G52350	AT5G52350.1	9	NM_124614.2	5	+	NP_200048.2	586	221~574
<i>AtEXO70B1</i>	AT5G58430	AT5G58430.1	1	NM_125229.3	5	-	NP_200651.1	624	264~615
<i>AtEXO70B2</i>	AT1G07000	AT1G07000.1	2	NM_100573.3	1	-	NP_172181.1	599	246~593
<i>AtEXO70C1</i>	AT5G13150	AT5G13150.1	1	NM_121318.2	5	-	NP_196819.1	653	272~641
<i>AtEXO70C2</i>	AT5G13990	AT5G13990.1	1	NM_121402.2	5	-	NP_196903.1	695	322~678
<i>AtEXO70D1</i>	AT1G72470	AT1G72470.1	1	NM_105906.3	1	+	NP_177391.1	633	246~618
<i>AtEXO70D2</i>	AT1G54090	AT1G54090.1	1	NM_104286.2	1	+	NP_175811.1	622	238~608
<i>AtEXO70D3</i>	AT3G14090	AT3G14090.1	1	NM_112265.4	3	-	NP_566477.2	623	241~610
<i>AtEXO70E1</i>	AT3G29400	AT3G29400.1	1	NM_113866.2	3	-	NP_189586.1	658	265~644
<i>AtEXO70E2-1</i>	AT5G61010	AT5G61010.1	1	NM_125494.2	5	+	NP_200909.1	639	272~623
<i>AtEXO70E2-2</i>	AT5G61010	AT5G61010.2	1	NM_001037039.1	5	+	NP_001032116.1	639	272~623
<i>AtEXO70F1</i>	AT5G50380	AT5G50380.1	1	NM_124420.2	5	-	NP_199849.2	683	303~668
<i>AtEXO70G1</i>	AT4G31540	AT4G31540.1	1	NM_119303.2	4	-	NP_194882.2	687	293~653
<i>AtEXO70G2</i>	AT1G51640	AT1G51640.1	1	NM_104042.2	1	-	NP_175575.1	660	253~624
<i>AtEXO70H1</i>	AT3G55150	AT3G55150.1	1	NM_115373.3	3	-	NP_191075.2	636	244~607
<i>AtEXO70H2</i>	AT2G39380	AT2G39380.1	1	NM_129495.2	2	-	NP_181470.1	637	243~607
<i>AtEXO70H3</i>	AT3G09530	AT3G09530.1	1	NM_111787.1	3	+	NP_187564.1	637	227~599
<i>AtEXO70H4</i>	AT3G09520	AT3G09520.1	1	NM_111786.1	3	+	NP_187563.1	628	232~603
<i>AtEXO70H5</i>	AT2G28640	AT2G28640.1	2	NM_128425.2	2	-	NP_180432.2	605	217~575
<i>AtEXO70H6</i>	AT1G07725	AT1G07725.1	2	NM_148445.3	1	-	NP_683286.2	615	237~588
<i>AtEXO70H7-1</i>	AT5G59730	AT5G59730.1	1	NM_125365.2	5	-	NP_200781.1	634	227~595
<i>AtEXO70H7-2</i>	AT5G59730	AT5G59730.2	1	NM_001037028.1	5	-	NP_001032105.1	632	227~595
<i>AtEXO70H8</i>	AT2G28650	AT2G28650.1	1	NM_128426.2	2	-	NP_180433.1	573	209~553

图 1 3 种 *AtEXO70* 的选择性剪接模式。黑色框部分代表编码序列，两端浅灰色框部分代表外显子部分的非编码区，折线为内含子部分，箭头代表转录方向。A1: *AtEXO70A1*; E2: *AtEXO70E2*; H7: *AtEXO70H7*

Figure 1 Three gene models of *AtEXO70* due to alternative splicing. Black boxes represented coding sequences; Grey boxes represented uncoding regions; fold lines represented introns; Arrowheads represented transcription directions.

列有2条，分别是AY072155和AY133751 (GenBank)，相关的EST序列有40条；3种不同编码的*AtEXO70A1*的外显子数目和基因编码序列都存在差异(表3)。在GenBank中*EXO70E2*包括4条cDNA序列和51条对应的EST序列，其中cDNA的登录号分别为BX841634.1, AY050411.1, AK317325.2和AY059656.1。2种不同*EXO70E2*的cDNA在5'-UTR端的-456 bp处存在104 bp的差异，因此这种选择性剪接的方式并不影响*EXO70E2*的编码序列，即选择性剪接的*EXO70E2*基因编码的蛋白质序列完全相同。GenBank中*EXO70H7*共有10条cDNA序列和51条EST序列，分析表明，*EXO70H7*存在2种不同的编码方式。对应的*EXO70H7*蛋白(NP_200781.1和NP_001032105.1)仅在最后15氨基酸存在差异。这缘于*EXO70H7*基因转录后位于起始密码子后的1858 bp处的选择性剪接，即相对于*EXO70H7-1*编码区，*EXO70H7-2*在起始密码子之后1858~1867处有10 bp被剪掉，由此引起两者的编码序列出现差异。

26种*AtEXO70*的序列长度介于523~695个氨基酸之间，每种蛋白都包含一个重要的结构域pfam03081。pfam03081由*AtEXO70*蛋白靠近C端的360个左右的氨基酸残基构成，结构模拟显示该结构域由十几个 α 螺旋形成近似棒状的结构。

2.2 水稻*EXO70*基因的加倍及其与*AtEXO70*的对比

从NCBI水稻基因组数据库和RGAP数据库中搜索并鉴定出47个*EXO70*基因座，分别位于水稻的12条染色体的正负2条链上，但不同的染色体上*EXO70*

基因座数目存在差异，其中1号染色体包含8个基因座，11号染色体包含6个基因座，而在3号和10染色体上仅包含1个基因座。其中38个基因座编码的蛋白长度超过412个氨基酸(表4)。*EXO70*基因座编码的蛋白质具有完整pfam03081结构域，依照以水稻*EXO70*基因的编码蛋白与拟南芥27种*EXO70*蛋白构建的进化树(图2)，将水稻*EXO70*基因分成12类(A~L)(表4)。12类*OsEXO70*基因的外显子数目呈现出较大的差异，以4种*OsEXO70A*基因外显子数目最多，大于或等于12。

和*AtEXO70*基因一样，*OsEXO70*基因重复既表现出它的基因座多样性，又表现出一定的染色体选择性。然而，*EXO70*基因在水稻基因组中的倍增方式比拟南芥基因组中更为复杂，这表现在4个方面。首先，*OsEXO70*基因座在数量上是*AtEXO70*基因座的2.04倍。其次，部分*OsEXO70*基因之间的序列相似度极高：同在11号染色体上2个不同的基因座Os11g42989和Os11g43049对应的编码序列完全相同；11号染色体上的Os11g0100800基因座和12号染色体上的Os12g01040基因座均包含3个外显子，分别编码590个氨基酸的蛋白，两者的编码区均由3900 bp构成，序列比对发现其碱基的一致性达到99.7%，蛋白质序列之间仅有3个氨基酸的差异。进一步对11号染色体和12染色体的全基因组比对分析发现，2个基因座分别所处的水稻的11号染色体和12号染色体短臂末端的DNA序列高度同源，高度核酸序列保守区段分别起始于11号染色体的2887bp和12号染色体的1277bp，长度约2 Mbp (图3)。第三，在进化中*OsEXO70*基因编码区的变异呈现多样化，使得蛋白质的结构与氨

表 4 搜索预测的水稻EXO70 蛋白及其编码基因信息
Table 4 Predicted EXO70 proteins and the corresponding gene from *Oryza sativa*

自命名	基因座	外显子数	RNA识别码	染色体	链	起点	终点	蛋白识别码	氨基酸数	Position of pfam03081
<i>OsEXO70A1</i>	Os04g0685600	12	NM_001060855.1	4	-	35421904	35427991	NP_001054320.1	634	263~626
<i>OsEXO70A2</i>	Os11g0157400	12	NM_001072333.1	11	-	2742977	2748161	NP_001065801	643	275~632
<i>OsEXO70A3</i>	Os12g0159700	18	NM_001072744.2	12	-	2981736	2992121	NP_001066212.2	976	282~636
<i>OsEXO70A4</i>	Os04g0685500	12		4	-	35416739	35420648	EEC78286	661	273~640
<i>OsEXO70B1</i>	Os01g0827500	1	NM_001051216.1	1	+	37165139	37167526	NP_001044681	652	289~646
<i>OsEXO70B2</i>	Os05g0473500	1	NM_001062357.1	5	-	23300759	23303351	NP_001055822.1	595	232~589
<i>OsEXO70B3</i>	Os01g0827600	4	NM_001051217.1	1	+	37168777	37173311	NP_001044682.1	553	188~532
<i>OsEXO70C1</i>	Os12g06840	1		12	-	3321591	3323693	ABA95888.1 EAZ19759.1	700	323~687
<i>OsEXO70C2</i>	Os11g0167600	1	NM_001072372.1	11	+	3249340	3251418	NP_001065840.1	692	306~679
<i>OsEXO70D1</i>	Os08g0455700	1	NM_001068497.1	8	+	22441274	22443172	NP_001061962.1	632	227~616
<i>OsEXO70D2</i>	Os09g0439600	1	NM_001069811.1	9	+	16948521	16950410	NP_001063276.1	638	236~622
<i>OsEXO70E1</i>	Os01g0763700	1	NM_001050871.2	1	+	33897350	33899158	NP_001044336.1	602	228~594
<i>OsEXO70F1</i>	Os02g0505400	1	NM_001053450.1	2	+	18754249	18756318	NP_001046915.1	689	300~674
<i>OsEXO70F2</i>	Os04g0382200	1	NM_001059141.1	4	+	18761710	18763776	NP_001052606.1	688	302~674
<i>OsEXO70F3</i>	Os01g0921400	3	NM_001051762.1	1	+	41991037	41994125	NP_001045227.1	556	189~548
<i>OsEXO70F4</i>	Os08g0530300	1	NM_001068860.2	8	+	26501957	26503777	NP_001062325.1	606	221~579
<i>OsEXO70F5</i>	Os10g33850	3		10	-	18423098	18425055	gb AAP54288.2	461	221~450
<i>OsEXO70G1</i>	Os02g0149700	1	NM_001052432.1	2	-	2735910	2738031	NP_001045897.1	494	99~459
<i>OsEXO70G2</i>	Os06g0698600	1	NM_001065002.2	6	+	30108006	30110861	NP_001058467.2	673	278~638
<i>OsEXO70G3</i>	Os08g0519900	2	NM_001068798.1	8	-	25933393	25935987	NP_001062263.1	687	295~652
<i>OsEXO70H1a</i>	Os11g0650100 /Os11g42989	1	NM_001189755.1	11	+	27754204	27755943	NP_001176684.1	579	209~472
<i>OsEXO70H1b</i>	Os11g43049	1	NM_001189755.1	11	+	27801505	27803244	NP_001176684.1	579	209~472
<i>OsEXO70H2</i>	Os03g33520	1		3	+	19782947	19784617	AAO66561.1	556	186~516
<i>OsEXO70H3</i>	Os12g01040	3		12	-	24538	28438	ABA95562.2	590	163~478
<i>OsEXO70H4</i>	Os11g0100800	3	NM_001189365.1	11	-	26139	30038	NP_001176294.1	590	163~478
<i>OsEXO70I1</i>	Os01g0905300	1	NM_001051662.1	1	+	41166899	41169036	NP_001045127.1	381	24~368
<i>OsEXO70I2</i>	Os01g0905200 /Os01g67810	2	NM_001051661.1	1	+	41164160	41166064	NP_001045126.1	557	215~547
<i>OsEXO70I3</i>	Os04g0111500	3	NM_001058568.1	4	-	652438	658115	NP_001052033.1	398	47~377
<i>OsEXO70I4</i>	Os07g10940	5		7	+	6020009	6029347	EAZ03199	691	326~687
<i>OsEXO70I5</i>	Os07g10960	5		7	+	6047375	6049551	BAC83686.1	588	311~583
<i>OsEXO70I6</i>	Os07g0210000	4	NM_001065705.1	7	+	6004291	6008770	NP_001059170.2	646	300~643
<i>OsEXO70J1</i>	Os08g0232700	1	NM_001067840.1	8	+	8060321	8062495	NP_001061305.1	526	155~523
<i>OsEXO70J2</i>	Os09g0347300	1	NM_001069478.1	9	+	11463763	11465938	NP_001062943.1	598	191~595
<i>OsEXO70J3</i>	Os05g0369500	1	NM_001061881.1	5	-	17763771	17765809	NP_001055346.1	528	121~525
<i>OsEXO70J4</i>	Os05g0369900	2	NM_001187461.1	5	+	17786692	17788517	NP_001174390.1	377	153~373
<i>OsEXO70J5</i>	Os05g0369300	1	NM_001061880.1	5	+	17758288	17760123	NP_001055345.1	520	155~517
<i>OsEXO70J6</i>	Os01g0383100	3	AK241094.1	1	-	16232624	16234985	gb BAH00950.1	681	331~678
<i>OsEXO70J7</i>	Os02g0575900	3	NM_001053750.1	2	+	22968001	22975517	NP_001047215.1	700	345~697
<i>OsEXO70K1</i>	Os06g0255900	1	NM_001063846.1	6	+	8115250	8118846	NP_001057311.1	412	60~409
<i>OsEXO70K2</i>	Os07g0211000 /Os07g10970	3	NM_001065707.2	7	-	6049678	6053148	NP_001059172.2	426	101~424
<i>OsEXO70L1</i>	Os06g08460	1	AK287613.1	6	+	4160338	4161798	13106.m00883	486	142~483
<i>OsEXO70L2</i>	Os11g0572200	2	NM_001074656.1	11	-	23304922	23307026	NP_001068124.1	433	236~402

a) 加黑的基因座来自于密歇根大学的水稻注解数据库(Rice Genome Annotation Project) Release 7, 其他则来自于 GenBank. RNA 识别码和蛋白识别码均来自于 GenBank. 搜索 GenBank 的水稻基因组数据库(*O. sativa (japonica cultivar-group)* genome (reference only))确定基因的起点和终点. 在 38 条长度超过 412 个氨基酸的 EXO70 蛋白中, *Os10g33850*, *Os11g0572200* 和 *Os07g10960* 基因编码的 pfam03081 氨基酸序列较短, *Os06g0255900* 和 *Os02g0149700* 的 5'端对应存在一定数目的碱基缺失

图 2 EXO70 蛋白分子进化关系

Figure 2 Molecular evolutionary relationships of EXO70

基酸数量发生较大的差异. 如OsEXO70A3 (Os12g-06270)由976个氨基酸组成, 除了在282~636个氨基酸构成一个pfam03081结构域, 还包含一个pfam-00078结构域(824~962氨基酸), 这种情况在拟南芥 *EXO70*中表现的并不明显. 最后体现在倍增的机制上: 相对于其他的OsEXO70蛋白, 由412个氨基酸组

成的Os06g0255900的N端对应的编码区在进化过程中经历过一段DNA序列的缺失(表4)。另外,水稻EXO70蛋白质中肽链长度低于400个氨基酸的有9个,其中的6个短于300个氨基酸(表5),在其N端和pfam02081结构域出现程度更高的肽链缺失情况,表明其对应的*EXO70*基因在加倍后经历了不同程度的

图 3 水稻染色体 11 和染色体上基因组共有的倍增片段. 竖线代表区域为序列高度相似的倍增区段

Figure 3 A 2-Mb segmental duplication between chromosome 11 and 12. Regions marked with vertical bar represent highly similar sequences resulting from DNA duplication

DNA 片段丢失的过程; 这种缺失较为明显的影响所编码蛋白质的结构, 进而使蛋白质的功能异于其他 EXO70 蛋白. 相比较而言, 在 23 个拟南芥 EXO70 基因编码的蛋白质中最短的 523 个氨基酸组成, 局部 DNA 片段严重缺失的情况在 *AtEXO70* 基因中未曾出现.

尽管 EXO70 蛋白在拟南芥中存在较高的保守性, 但是不同类型的 EXO70 基因之间在内含子的数量上存在较大的差异, 如 3 种 *AtEXO70A* 包含 8 个以上的内含子, 17 种 *AtEXO70* 基因不存在内含子, 其余的 3 种

AtEXO70 基因仅含有 1 个内含子. 这种并行同源基因 (paralogous genes) 存在内含子数量较大差异的现象, 暗示着在进化史上拟南芥基因组可能经历过至少 2 次的大规模的倍增. 水稻 EXO70 基因在内含子方面表现出和 *AtEXO70* 基因相似的现象, 但是更为复杂 (表 4), 同时也意味着水稻基因组倍增的次数或者规模高于拟南芥.

在水稻和拟南芥基因组中, 同种类型的 EXO70 基因, 内含子与外显子的累积突变存在较大差异, 即表现为外显子的保守性远高于内含子.

2.3 Pfam03081 结构与进化分析

在所发现的 EXO70 蛋白质中均包含有 Pfam-03081 结构域, 二级结构则由绝大多数的 α 螺旋构成. 不同的 EXO70 蛋白的 Pfam03081 在对应位置上的氨基酸的组成存在一定的差异, 但是其二级结构和三级结构上存在极高的相似度, 如 *AtEXO70A1* 与 *OsEXO70A1* 在 Pfam03081 区仅有 72.3% 的序列一致性 (图 S1), 但是在二级结构和三维结构上极为相似 (图 S2 和 S3).

尽管在结构上较为保守, 但不同的 Pfam03081 序列长度依然存在一定的差异. 在拟南芥中 pfam03081 的氨基酸组成数量在最高为 379, 最低为 345. Pfam-03081 在一级结构上均位于靠近 EXO70 蛋白的 C 端, 不同种类的 *AtEXO70* 的 pfam03081 结构域的氨基酸保守性较高, 不同类型的 *AtEXO70* 蛋白的 pfam03081 结构域之间存在 2 个高变区. 3 种 *AtEXO70A1* 和 2 种 *AtEXO70H7* 的选择性剪接产物的差异区段, 全部位

表 5 搜索并鉴定的水稻 EXO70 蛋白 (短于 400 氨基酸) 及其编码基因信息

Table 5 Predicted OsEXO70 (shorter than 400 amino acids) and corresponding genes

基因座	外显子数	蛋白识别码	氨基酸数	染色体	链	起点	终点	备注
Os04g02070	3	NP_001052033.1	398	4	-	652438	658115	
Os01g67820	1	NP_001045127.1	381	1	+	41166899	41169036	
Os01g05580	1	Os01g05580.1	177	1	-	2664072	2663539	
Os01g69280								该基因不存在
Os11g43000	1							该基因不存在
Os05g30680	1	NP_001055347.1	229	5	+	17772524	17774287	
Os05g30620	1	EEE63468.1	164	5	+	17772524	17774287	
Os08g13270	1	EAZ41943.1	281	8	+	7892281	7894462	
Os07g10920	1	NP_001059171.1	320	7	-	6009034	6011220	
Os01g56210	2	BAC10769.1 BAD53330.1	243	1	-	32372869	32371982	
Os01g49460	1	BAD82552.1 EAZ13148.1	152	1	+	28444493	28444948	

于pfam03081结构域之外,更暗示pfam03081结构域在发挥蛋白质的功能方面起到极为关键的作用. 在所有的拟南芥EXO70蛋白质的pfam03081结构域中有10处的氨基酸完全一致,它们可能是决定着蛋白质的结构或者功能的关键氨基酸残基(图S4). 在38个长度超过412个氨基酸的OsEXO70蛋白中,均包含1个pfam03081结构域,其结构域的氨基酸长度差异高于不同的AtEXO70蛋白,介于231~355个氨基酸之间(表4),如Os11g36400, OsEXO70F5 (Os10g33850)和OsEXO70I5 (Os07g10960)的pfam03081长度低于其他的OsEXO70蛋白. 序列比对后发现pfam03081存在3个序列高变区(图S5),不同类型的OsEXO70蛋白在这3个区域均表现出一定的序列长度和结构差异.

以pfam03081的序列构建的分子进化树(图S6和S7)与以完整EXO70蛋白构建出的进化树(图2)对比后发现,两者的树形仅有略微差异,表明pfam03081结构域区在与EXO70蛋白的N端序列呈现趋同的进化效应,而部分蛋白质则由于在进化过程中由于N端或者pfam03081内部的长片段的序列缺失而造成差异的进化效应(Os05g0369900和Os06g0255900). 氨基酸水平上的相对缺失可能在一定程度上影响OsEXO70的生物学功能.

2.4 EXO70在拟南芥和水稻中密码子使用偏性

RSCU能直观地反映密码子使用的偏好性程度,且其值与氨基酸的使用及密码子的丰度无关. 表6显示,除了甲硫氨酸和色氨酸之外,EXO70的18种氨基酸在59种密码子中的选择均存在偏好性,其偏性程度在水稻和拟南芥中呈现较为明显的差异. 在密码子选择方面,仅有5个使用频率最高的密码子为AtEXO70和OsEXO70所共有,其他的13个最高使用频率的密码子在两者中表现出明显的不同,这说明同种基因的密码子偏好性选择方面因物种不同而存在差异.

ENc值越小表明EXO70的偏好性越高,统计学显示ENc数值越低则表达水平越高^[30], ENc<30和ENc>55的基因可被预测为高表达和低表达基因,表7显示EXO70在拟南芥和水稻中的表达总体表现偏低. OsEXO70的ENc值略低于AtEXO70,可能是源于种类较多的OsEXO70存在部分表达量较高的基因,从而降低了水稻EXO70的ENc平均值. 基因表达水平的实际观测值与CAI值非常接近,而且显示表达量较高的基因具有较高的CAI值,而较低的CAI值显示目标

基因的表达量较低^[31,32]. 由此CAI值已广泛应用于基因表达水平的预测. FOP数值的高低同样也能反映基因表达水平的高低. 由于OsEXO70的Fop和和CAI值均高于AtEXO70,表明在总体水平上,EXO70基因在拟南芥中的表达水平要低于水稻.

3 讨论

3.1 pfam03081结构域

有趣的是,包括拟南芥和水稻EXO70在内,所有生物中已知的EXO70蛋白质均包含一个完整的pfam03081结构域, pfam03081由绝大多数的 α 螺旋组成,在拟南芥和水稻EXO70蛋白的三维结构上极为保守. 已发现的EXO70蛋白均位于胞质和核质内^[33-35],大多与其他蛋白质互作而发挥功能. 现在已知的EXO70蛋白多数能够参与细胞膜系统相关的功能,如细胞的极化、生长、分裂、分化、病原反应和依赖囊泡运输发挥作用的胞泌等功能^[19,36-39];还发现EXO70能够参加pre-mRNA的剪切^[35]. 迄今为止,所有已知EXO70参加的功能均是与其他蛋白质互作之后发挥功能,因此保守的pfam03081可能是与其他蛋白互作所必需的结构域,然而其具体的功能有待进一步的研究.

pfam03081结构域是cl03878一个成员,其序列和二级结构较为保守. 对EXO70蛋白结构分析表明,拟南芥23个并行同源EXO70基因和水稻38个EXO70编码的蛋白均包含完整的pfam03081结构域. 而所有AtEXO70的pfam03081的氨基酸长度介于345~379之间,而水稻中9个OsEXO70基因座的编码产物(NP_001174390.1)的pfam03081氨基酸残基构成短于230. 表明水稻EXO70基因在进化过程中数量加倍的同时发生了更多的变异(如碱基片段缺失或其他途径引起的无义突变).

3.2 EXO70基因倍增在动植物中的差异

前期研究表明,EXO70蛋白在酵母及动物中均由单基因编码,而本研究发现无论拟南芥还是水稻中EXO70都是多基因编码,这暗示EXO70基因在拟南芥和水稻为代表的植物基因组中倍增数目可能均高于动物,造成这种差异可能与环境、生物进化息息相关. 生物进化的驱动力表现为多元性,在进化过程中,多种因素伴随着自然选择起作用,某些非自然选

表 6 同义密码子在拟南芥和水稻EXO70 基因中的使用频率^{a)}
Table 6 The frequency of synonymous codons in EXO70 of Arabidopsis thaliana and Oryza sativa

氨基酸	密码子	拟南芥		水稻		氨基酸	密码子	拟南芥		水稻	
		数目	RSCU	数目	RSCU			数目	RSCU	数目	RSCU
Phe	UUU	276	0.86	369	0.71	Ala	GCU	376	1.71	476	0.67
	UUC	365	1.14	668	1.29		GCC	118	0.54	1028	1.45
	UUA	214	0.87	146	0.3		GCA	195	0.89	401	0.57
	UUG	311	1.26	466	0.94		GCG	189	0.86	926	1.31
Leu	CUU	330	1.34	449	0.91	Tyr	UAU	184	0.91	255	0.72
	CUC	271	1.1	971	1.97		UAC	220	1.09	451	1.28
	CUA	173	0.7	205	0.41	His	CAU	198	1.18	289	0.8
	CUG	177	0.72	727	1.47		CAC	139	0.82	431	1.2
Ile	AUU	318	1.1	413	0.93	Gln	CAA	232	0.96	336	0.66
	AUC	347	1.2	653	1.47		CAG	251	1.04	685	1.34
	AUA	206	0.71	268	0.6	Asn	AAU	271	0.96	355	0.83
Met	AUG	404	1	758	1		AAC	296	1.04	504	1.17
	GUU	385	1.69	315	0.69	Lys	AAA	465	1.01	366	0.61
Val	GUC	184	0.81	608	1.34		AAG	460	0.99	829	1.39
	GUA	112	0.49	143	0.31	Asp	GAU	613	1.42	640	0.8
	GUG	229	1.01	755	1.66		GAC	251	0.58	965	1.2
	UCU	414	1.62	311	0.82	Glu	GAA	581	0.94	589	0.61
Ser	UCC	220	0.86	528	1.39		GAG	659	1.06	1355	1.39
	UCA	298	1.17	320	0.84	Cys	UGU	113	1.2	132	0.66
	UCG	233	0.91	376	0.99		UGC	75	0.8	270	1.34
	AGU	198	0.77	209	0.55	Trp	UGG	162	1.00	374	1.00
	AGC	170	0.67	541	1.42		CGU	130	0.87	190	0.58
Pro	CCU	149	1.24	254	0.92	Arg	CGC	58	0.39	558	1.69
	CCC	44	0.37	233	0.84		CGA	91	0.61	184	0.56
	CCA	150	1.25	261	0.94		CGG	82	0.55	419	1.27
	CCG	136	1.14	360	1.3		AGA	350	2.35	210	0.64
	ACU	192	1.16	233	0.78		AGG	183	1.23	418	1.27
Thr	ACC	137	0.83	375	1.25	Gly	GGU	221	1.3	332	0.78
	ACA	191	1.16	272	0.91		GGC	81	0.48	729	1.72
	ACG	140	0.85	318	1.06		GGA	252	1.48	275	0.65
							GGG	126	0.74	363	0.85

a) 粗体数字为使用频率最高的密码子

表 7 密码子组成和使用性参数^{a)}
Table 7 Composition and usage parameters of codons

名称	T3s	C3s	A3s	G3s	CAI	CBI	Fop	Nc	GC3s	GC	L_sym	L_aa	Gravy	Aromo
AtEXO70	0.3838	0.2615	0.3183	0.3127	0.206	-0.001	0.419	55	0.438	0.442	14030	14596	-0.35507	0.082694
OsEXO70	0.2316	0.4219	0.186	0.399	0.218	0.07	0.456	53.81	0.656	0.562	26708	27840	-0.21142	0.076042

a) 包含所有OsEXO70 基因的27839 个通用密码子和AtEXO70 基因的14596 个通用密码子, 不包含终止密码子

择的因素甚至可能起着决定性的作用。自然选择决定着物种进化的方向,包括生物自身的遗传特性、种群的多态性等内源动力和包含外界环境在内的外源驱动力在很大程度上影响生物的进化。物种形成和种群的性质与其所面临的外界环境密切相关,有时外源驱动力可能在很大程度上影响着生物在微观层次上的进化,即基因或整个基因组的改变。

环境变化与生物适应进化通常交互发展,这种互动表现在2个方面:一是生物缓慢且相对微弱的改变着其生存环境;二是生物的生存时刻都在受到环境的影响和制约,面对改变了的环境,生物从行为方式到表型等方面都需要做出适应性的调整,从而在适应中达到物种诞生与进化的目的,生物的进化在微观上则表现为基因或基因组结构的变化。动植物身体结构和行为能力等方面的巨大差异,导致动植物之间在应对气候、地质等环境的变化时,表现形式有很大的不同(如植物固定于某处,无法像动物那样能够迅速逃避周围环境变化带来的危害)。应对环境变化的行为能力的差异导致动植物拥有着不同的进化表现(例如,植物受环境的影响较动物更高,环境的剧烈变化可能会使其为了生存而发生更深层次的进化)。EXO70在植物基因组中的加倍可能也是适应性进化的结果。

3.3 EXO70重复与基因组倍增

基因倍增是指特定的基因片段在基因组中复制出一个或更多的拷贝,由此表明基因重复或重复基因序列在同一基因组的出现是基因倍增的一种表现,植物基因组序列的测序,为系统地研究植物功能基因的特征提供了机遇。2005年,水稻的籼稻和粳稻2个亚种的基因组的精细图宣告完成,发现水稻基因组中约有37544个非转座因子的编码基因^[40],比拟南芥基因组(其基因组大小约为157 Mb,共27416个基因)基因密度低^[41]。

水稻蛋白编码基因的总数为拟南芥蛋白编码基因总数目的1.37倍,低于水稻基因组与拟南芥基因组的长度之比(比值为3.7)。这可能缘于水稻基因组加倍的过程中,大量倍增的基因流失,或者因为某些新产生的包括EXO70基因座在内的基因由于插入和缺失等积累突变造成了基因座结构的较大改变,而导致倍增的基因获得新的功能或者退化成为没有功能的假基因。我们从非冗余的水稻基因组中搜索到47个

EXO70基因,约为AtEXO70基因座数目的2.04倍,其中只有38个OsEXO70基因编码的蛋白质序列具有完整的pfam03081,是拟南芥EXO70基因座数量的1.65倍,二者均高于水稻和拟南芥蛋白编码基因总数的比值。这表明EXO70基因的加倍过程比大规模的基因组加倍更为复杂,大规模基因组加倍之后基因加倍的过程可能仍在继续。拟南芥的23个EXO70基因8个位于正链,水稻中有29个EXO70基因座位于正链。另外,在拟南芥5号染色体上存在9个EXO70基因,其数目远高于其他几条染色体;水稻1号染色体上的座数目高于其他染色体。即EXO70基因在染色体和正负链的选择方面均表现出一定的“偏好”性。此外,2个物种不同EXO70基因间内含子的数量存在较为明显的差异(表3和4)。这些数据表明EXO70基因在2种基因组中发生了多次加倍,且基因加倍方式均存在包括基因组加倍在内的多种不同的机制。在水稻的进化史中,6种编码氨基酸长度低于300的OsEXO70基因在加倍后经历了不同程度的DNA缺失的过程,而这种DNA缺失的情况在拟南芥EXO70基因中表现的并不明显,由此而言,OsEXO70基因倍增的机制也较AtEXO70基因更为复杂;此类DNA缺失引起其编码的OsEXO70蛋白在N端或pfam03081的结构不完整,间接导致了蛋白功能的丧失或改变。

水稻基因组全序列为系统的研究基因倍增提供了便利。基于2002年公布的水稻基因组草图序列,Goff等人^[42]利用同义替代率分布方法提出水稻基因组可能发生过一次全基因组倍增。之后,Paterson等人^[43]以及Guyot和Keller^[44]指出,水稻在距今5500~7000万年前发生过一次全基因组倍增。2005年,基于籼稻精细图序列,水稻全基因组倍增事件再一次被证实^[45]。我们发现,在分别包含OsEXO70H4和OsEXO70H3两个基因座的水稻11号染色体和12号染色体的5'端存在的大片段高度重复序列,可能源自于进化史上最新的一次发生在约770万年前的基因组局部倍增过程^[46],Jacquemin等人^[47]则认为倍增区域的发生于500万年前。水稻基因组中这种大片段倍增是迄今为止发现的在动植物中最为清晰、完整的基因组倍增行为。相对于水稻基因组倍增而言,拟南芥的基因组片段的倍增行为则显得较为凌乱^[48,49]。

3.4 水稻与拟南芥EXO70密码子使用偏性

Lynn等人^[50]发现,细菌密码子的使用特性决定

于基因组的碱基组成,而且编码区的GC含量及密码子第3位的GC含量(GC3s)是影响密码子偏好性主要的决定因素.和多数基因类似,*EXO70*的密码子在物种间也存在使用偏性,水稻中不同基因的同义密码子的GC3s含量介于43%和92%之间^[51],*OsEXO70*基因编码区的GC3s值为65.6%,水稻和拟南芥*EXO70*基因在密码子使用上GC3s含量存在21.8%差异.作为双子叶植物和单子叶植物的2个代表,这种差异在一定程度上反映出2种基因组不同的碱基组成,不同的碱基组成又决定了水稻和拟南芥基因组之间在密码子偏好性选择的差异.

导致密码子选择的偏好性以及偏好性差异的原因,不同的学者有着不同的认识. Sharp和Li^[52]通过对单细胞生物大肠杆菌、酿酒酵母等的基因的密码子数据的分析,认为密码子选择性使用取决于基因突变的偏好性和对最优密码子选择的平衡. Von Samson-Himmelstjerna等人^[30]基于大肠杆菌和线虫的基因表达研究结果,发现同义密码子的使用模式与基因的表达水平密切相关,认为基因的表达水平是决定密码子使用偏好性的重要因素.之后, Zhang和Li^[53]基于对哺乳动物基因的密码子使用的模式分析后,得出几乎相同的结论.而Plotkin等人^[54]对人类组织表达特异性基因的密码子使用和表达进行研究后,则得出不同的结论.他们认为:不同基因之间表现出的同义密码子的差异偏性,主要来源于翻译选择的“压力”.然而,之前Hastings和Emerson^[55]基于脊椎动物的组织特异性基因的cDNA序列的研究,却认为基因间密码子非均衡使用的模式并不存在较大的差别.法国学者Semon等人^[56]尽管认同Plotkin等人^[54]

所述的“组织特异性基因间密码子差异使用”,但对其差异来源的结论表示怀疑.此外,同种基因在不同物种中不同的碱基突变频率和模式会在很大程度上影响密码子非均衡使用以及偏性选择差异;物种或细胞内部tRNA的丰度亦可能会反向影响同义密码子的偏性模式.由此我们可以得知影响着密码子使用的原因并非唯一,可能有突变偏好、自然选择和遗传漂变等众多的因素参与其中,不同物种或同一物种的不同基因引起密码子使用偏性的原因也可能不相同,*EXO70*基因在水稻和拟南芥中的密码子选择偏好性的差异表现亦可能是多种原因促成.

4 结论

*EXO70*在拟南芥和水稻基因组中均发生了多次倍增,而且基因加倍方式均存在多种不同的机制.*EXO70*基因在水稻基因组中的倍增方式比拟南芥基因组中更为复杂;分别包含序列高度相似的*OsEXO70H4*和*OsEXO70H3*基因座的水稻11染色体和12染色体端部的序列高度相似,暗示水稻在进化史的近期发生过一次局部的基因组倍增.

不同的*EXO70*基因序列存在明显差异,编码的蛋白质在序列上呈现出较大差异,然而它们在结构上都存在一共同的结构域pfam03081,它在*EXO70*与其他蛋白质互作而发挥功能所必需,这种共有的结构域可能在一定程度上决定*EXO70*蛋白的功能.

*EXO70*基因在同义密码子使用上存在明显偏好性,使用频率最高的同义密码子在水稻*EXO70*基因和拟南芥*EXO70*基因中存在较大的差异,表现出明显的物种特异性.

参考文献

- 1 Lindeboom J, Mulder B M, Vos J W, et al. Cellulose microfibril deposition: Coordinated activity at the plant plasma membrane. *J Microsc-Oxford*, 2008, 231: 192–200
- 2 Geitmann A, Emons A M. The cytoskeleton in plant and fungal cell tip growth. *J Microsc-Oxford*, 2000, 198: 218–245
- 3 Miller D D, de Ruijter N C, Emons A M C. From signal to form: Aspects of the cytoskeleton-plasma membrane—Cell wall continuum in root hair tips. *J Exp Bot*, 1997, 48: 1881–1896
- 4 Zhang Y, Liu C M, Emons A M, et al. The plant exocyst. *J Integr Plant Biol*, 2010, 52: 138–146
- 5 Hsu S C, TerBush D, Abraham M, et al. The exocyst complex in polarized exocytosis. *Int Rev Cytol*, 2004, 233: 243–265
- 6 Munson M, Novick P. The exocyst defrocked, a framework of rods revealed. *Nat Struct Mol Biol*, 2006, 13: 577–581
- 7 TerBush D R, Maurice T, Roth D, et al. The exocyst is a multiprotein complex required for exocytosis in *Saccharomyces cerevisiae*. *EMBO J*, 1996, 15: 6483–6494
- 8 TerBush D R, Novick P. Sec6, Sec8, and Sec15 are components of a multisubunit complex which localizes to small bud tips in *Saccharomyces cerevisiae*. *J Cell Biol*, 1995, 130: 299–312

- 9 Kee Y, Yoo J S, Hazuka C D, et al. Subunit structure of the mammalian exocyst complex. *Proc Natl Acad Sci USA*, 1997, 94: 14438–14443
- 10 Elias M, Drdova E, Ziak D, et al. The exocyst complex in plants. *Cell Biol Int*, 2003, 27: 199–201
- 11 Chong Y T, Gidda S K, Sanford C, et al. Characterization of the *Arabidopsis thaliana* exocyst complex gene families by phylogenetic, expression profiling, and subcellular localization studies. *New Phytol*, 2010, 185: 401–419
- 12 Finger F P, Hughes T E, Novick P. Sec3p is a spatial landmark for polarized secretion in budding yeast. *Cell*, 1998, 92: 559–571
- 13 Guo W, Roth D, Walch-Solimena C, et al. The exocyst is an effector for Sec4p, targeting secretory vesicles to sites of exocytosis. *EMBO J*, 1999, 18: 1071–1080
- 14 Boyd C, Hughes T, Pypaert M, et al. Vesicles carry most exocyst subunits to exocytic sites marked by the remaining two subunits, Sec3p and Exo70p. *J Cell Biol*, 2004, 167: 889–901
- 15 Tsuboi T, Ravier M A, Xie H, et al. Mammalian exocyst complex is required for the docking step of insulin vesicle exocytosis. *J Biol Chem*, 2005, 280: 25565–15570
- 16 Vega I E, Hsu S C. The exocyst complex associates with microtubules to mediate vesicle targeting and neurite outgrowth. *J Neurosci*, 2001, 21: 3839–3848
- 17 He B, Xi F, Zhang X, et al. Exo70 interacts with phospholipids and mediates the targeting of the exocyst to the plasma membrane. *EMBO J*, 2007, 26: 4053–4065
- 18 Zhang X, Orlando K, He B, et al. Membrane association and functional regulation of Sec3 by phospholipids and Cdc42. *J Cell Biol*, 2008, 180: 145–158
- 19 Li S, van Os G M, Ren S, et al. Expression and functional analyses of EXO70 genes in *Arabidopsis* implicate their roles in regulating cell type-specific exocytosis. *Plant Physiol*, 2010, 154: 1819–1830
- 20 Synek L, Schlager N, Elias M, et al. AtEXO70A1, a member of a family of putative exocyst subunits specifically expanded in land plants, is important for polar growth and plant development. *Plant J*, 2006, 48: 54–72
- 21 Žárský V, Cvrčková F, Potocký M, et al. Exocytosis and cell polarity in plants-exocyst and recycling domains. *New Phytol*, 2009, 183: 255–272
- 22 Kent W J. BLAT—The BLAST-like alignment tool. *Genome Res*, 2002, 12: 656–664
- 23 Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, 2011, 28: 2731–2739
- 24 Zhang L, Jin L G, Luo L, et al. Analysis of nuclear gene codon bias on soybean genome and transcriptome (in Chinese). *Acta Agron Sin*, 2011, 37: 965–974 [张乐, 金龙国, 罗玲, 等. 大豆基因组和转录组的核基因密码子使用偏好性分析. *作物学报*, 2011, 37: 965–974]
- 25 Wright F. The effective number of codons' used in a gene. *Gene*, 1990, 87: 23–29
- 26 Gupta S K, Bhattacharyya T K, Ghosh T C. Synonymous codon usage in *Lactococcus lactis*: Mutational bias versus translational selection. *J Biomol Struct Dyn*, 2004, 21: 527–536
- 27 Sharp P M, Tuohy T M, Mosurski K R. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res*, 1986, 14: 5125–5143
- 28 Liu Q, Feng Y, Xue Q. Analysis of factors shaping codon usage in the mitochondrion genome of *Oryza sativa*. *Mitochondrion*, 2004, 4: 313–320
- 29 Sau K, Gupta S K, Sau S, et al. Factors influencing synonymous codon and amino acid usage biases in Mimivirus. *Biosystems*, 2006, 85: 107–113
- 30 von Samson-Himmelstjerna G, Harder A, Failing K, et al. Analysis of codon usage in beta-tubulin sequences of helminths. *Parasitol Res*, 2003, 90: 294–300
- 31 Peixoto L, Zavala A, Romero H, et al. The strength of translational selection for codon usage varies in the three replicons of *Sinorhizobium meliloti*. *Gene*, 2003, 320: 109–116
- 32 Romero H, Zavala A, Musto H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic Acids Res*, 2000, 28: 2084–2090
- 33 Hamburger Z A, Hamburger A E, West A P Jr, et al. Crystal structure of the *S. cerevisiae* exocyst component Exo70p. *J Mol Biol*, 2006, 356: 9–21
- 34 Cvrckova F, Grunt M, Bezvoda R, et al. Evolution of the land plant exocyst complexes. *Front Plant Sci*, 2012, 3: 159
- 35 Dellago H, Loscher M, Ajuh P, et al. Exo70, a subunit of the exocyst complex, interacts with SNEV (hPrp19/ hPso4) and is involved in pre-mRNA splicing. *Biochem J*, 2011, 438: 81–91
- 36 Cole RA, Synek L, Zarsky V, et al. SEC8, a subunit of the putative *Arabidopsis* exocyst complex, facilitates pollen germination and competitive pollen tube growth. *Plant Physiol*, 2005, 138: 2005–2018

- 37 Hala M, Cole R, Synek L, et al. An exocyst complex functions in plant cell growth in *Arabidopsis* and tobacco. *Plant Cell*, 2008, 20: 1330–1345
- 38 Fendrych M, Synek L, Pecenkova T, et al. The *Arabidopsis* exocyst complex is involved in cytokinesis and cell plate maturation. *Plant Cell*, 2010, 22: 3053–3065
- 39 Wang J, Ding Y, Wang J, et al. EXPO, an exocyst-positive organelle distinct from multivesicular endosomes and autophagosomes, mediates cytosol to cell wall exocytosis in *Arabidopsis* and tobacco cells. *Plant Cell*, 2010, 22: 4009–4030
- 40 International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature*, 2005, 436: 793–800
- 41 The Arabidopsis Information Resource. The *Arabidopsis* genome annotation (TAIR10). Nov 17, 2010 https://arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp
- 42 Goff S A, Riche D, Lan T H, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Japonica*). *Science*, 2002, 296: 92–100
- 43 Paterson A H, Bowers J E, Chapman B A. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA*, 2004, 101: 9903–9908
- 44 Guyot R, Keller B. Ancestral genome duplication in rice. *Genome*, 2004, 47: 610–614
- 45 Yu J, Wang J, Lin W, et al. The genomes of *Oryza sativa*: A history of duplications. *PLoS Biol*, 2005, 3: e38
- 46 The Rice Chromosomes 11 and 12 Sequencing Consortia. The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. *BMC Biol*, 2005, 3: 20
- 47 Jacquemin J, Chaparro C, Laudie M, et al. Long-range and targeted ectopic recombination between the two homeologous chromosomes 11 and 12 in *Oryza species*. *Mol Biol Evol*, 2011, 28: 3139–3150
- 48 Simillion C, Vandepoele K, Van Montagu M C, et al. The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*, 2002, 99: 13627–13632
- 49 Bowers J E, Chapman B A, Rong J, et al. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, 2003, 422: 433–438
- 50 Lynn D J, Singer G A, Hickey D A. Synonymous codon usage is subject to selection in thermophilic bacteria. *Nucleic Acids Res*, 2002, 30: 4272–4277
- 51 Wang H C, Singer G A, Hickey D A. Mutational bias affects protein evolution in flowering plants. *Mol Biol Evol*, 2004, 21: 90–96
- 52 Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol*, 1986, 24: 28–38
- 53 Zhang L, Li W H. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol Biol Evol*, 2004, 21: 236–239
- 54 Plotkin J B, Robins H, Levine A J. Tissue-specific codon usage and the expression of human genes. *Proc Natl Acad Sci USA*, 2004, 101: 12588–12591
- 55 Hastings K E, Emerson C P Jr. Codon usage in muscle genes and liver genes. *J Mol Evol*, 1983, 19: 214–218
- 56 Semon M, Lobry J R, Duret L. No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Mol Biol Evol*, 2006, 23: 523–529

Comparison of *EXO70* duplication between *Arabidopsis thaliana* and *Oryza sativa*

YANG Kun^{1,2}, ZHANG Yi^{1,2}, LÜ Jun^{1,2}, ZHAO YongBin³, ZHANG HeCui¹,
HAN Xu¹ & HE GuangHua^{1,2}

¹ College of Agronomy and Biotechnology, Southwest University, Chongqing 400715, China;

² Rice Research Institute of Southwest University, Chongqing 400715, China;

³ College of Life Science, Jilin Normal University, Siping 136000, China

EXO70, an important component of the exocyst complex, is encoded by a single gene in humans, mice and yeast. To understand *EXO70* gene duplication in plant genomes, we identified and analyzed all *OsEXO70* and *AtEXO70* genes in *Arabidopsis thaliana* and *Oryza sativa*. A total of 23 different *EXO70* genes were identified in the *A. thaliana* genome, of which 15 were present on the ‘-’ strand. We detected 27 different mature mRNAs of *EXO70* arising from alternative splicing. In *O. sativa*, 38 *OsEXO70* genes coding for proteins comprising more than 412 amino acids were found to be distributed across the 12 rice chromosomes. Of these genes, 29 were located on the ‘+’ strand. Interestingly, a product of large-scale genome duplication—an approximately 2-Mbp-long fragment including both *OsEXO70H3* and *OsEXO70H4*, was highly conserved between chromosomes 11 and 12. The gene duplication mode of *OsEXO70* is more complex than that of *AtEXO70*. Although synonymous codon bias exists in both *AtEXO70* and *OsEXO70*, codon preferences differ between them. All EXO70 proteins contain a pfam03081 domain, indicating that these proteins may have the same or similar functions. We conclude that *EXO70* has been duplicated multiple times during the evolution of both *A. thaliana* and *O. sativa*, with more duplications occurring in *O. sativa* than in *A. thaliana*. Additionally, a greater number of gene duplication mechanisms are operative for *OsEXO70* than for *AtEXO70*. We also propose that the pfam03081 domain may play an essential role in EXO70 functional determination. Finally, many differences in synonymous codon preference exist between *AtEXO70* and *OsEXO70*.

EXO70, pfam03081, gene duplication, synonymous codon bias

doi: 10.1360/N972014-00853

补充材料

图 S1 AtEXO70A1 与 OsEXO70A1 的 pfam03081 的氨基酸序列比对

图 S2 AtEXO70A1 与 OsEXO70A1 的 pfam03081 区段的二级结构

图 S3 Pfam03081 的三维结构

图 S4 AtEXO70 蛋白的 pfam03081 结构域氨基酸序列比对

图 S5 OsEXO70 pfam03081 结构域比对情况

图 S6 AtEXO70 的 pfam03081 的分子进化分析

图 S7 OsEXO70 蛋白 pfam03081 结构域分子进化分析

本文以上补充材料见网络版 csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.