

坛紫菜(*Porphyra haitanensis*) 丝状孢子体 EST 的获取及其生物信息学分析*

庞国兴¹ 王广策² 胡松年¹⁾ 曾呈奎²

(中国科学院海洋研究所 青岛 266071; 中国科学院研究生院 北京 100039)

¹⁾(中国科学院基因组中心 北京 101300)

²⁾(中国科学院海洋研究所 青岛 266071)

摘要 采用改进的一步法提取坛紫菜丝状孢子体的总 RNA, 构建了坛紫菜丝状孢子体的 cDNA 文库, 对测序得到的 EST 进行分析并与条斑紫菜的 EST 相比较。结果表明, cDNA 文库共包含 2.5×10^5 个克隆, 测序得到 490 个 EST, 其中 275 个与已知功能基因同源, 占 56.1%, 260 个 EST 与已知的条斑紫菜 EST 同源, 占 53.0%。使用 GO (Gene Ontology) 分类方法对 275 个 EST 进行分类, 其中 226 个可归属于细胞组分 (Cellular component)、分子功能 (Molecular function) 和生物学过程 (Biological process) 三个大类, 而其余 49 个则为未知功能的 EST。挑选出与条斑紫菜同源性高的 28 个 EST, 这些 EST 所包括的密码子数为 3221 个, 发现密码子第三位的平均 GC 含量远高于第一位的平均 GC 含量和第二位的平均 GC 含量。统计坛紫菜丝状孢子体 3221 个密码子的使用频率, 发现同义密码子第三位碱基偏好使用 G 或 C。

关键词 坛紫菜, EST, GC 含量, 密码子

中图分类号 Q93

紫菜是红藻门红毛菜科紫菜属 (*Porphyra*) 海藻的总称, 全世界有一百三十余种, 广泛分布于寒带到亚热带海域潮间带 (Yoshida *et al.*, 1997)。紫菜生活史属于异型世代交替, 包括配子体世代和孢子体世代。它的配子体是大型叶片状组织, 是可食用的部分; 它的孢子体是较小的丝状体, 在自然条件下钻入贝壳生长。单倍叶状体产生雌雄配子, 受精形成二倍果孢子, 果孢子进一步萌发生长成为丝状体, 丝状体经过近半年的生长发育产生单倍壳孢子, 壳孢子萌发生长成紫菜叶状体 (费修纆等, 1999)。目前, 我国主要栽培品种有坛紫菜 (*Porphyra haitanensis*) 和条斑紫菜 (*Porphyra yezoensis*) 两种。

坛紫菜产于我国南方海域, 是我国特有的栽

培品种, 它的栽培面积及产量都大于条斑紫菜, 但创造的经济价值却低于条斑紫菜。因此, 需要加强对坛紫菜的基础研究, 以期将来可以改善坛紫菜的品质并提高经济价值。

表达序列标签 (Expressed Sequence Tags, ESTs) 来自 cDNA 克隆的末端序列, 代表生物体某种组织某一时期的表达基因。EST 数据在基因作图、克隆基因、新基因识别、蛋白质组研究等许多方面具有重要的用途 (陈红歌等, 2003; Adams *et al.*, 1991; 王广策等, 2002a; 高政权等, 2003)。目前, 条斑紫菜已进行了大规模的 EST 分析, 在 GenBank 的 dbEST 库中已存在 2 万多个条斑紫菜丝状孢子体和叶状配子体的 EST (Asamizu *et al.*, 2003; Nikaido *et al.*, 2000; Lee *et al.*, 2000; 杨官品

* 国家自然科学基金面上资助项目, 40476059、30170499、30250003、39890390 号; 中国科学院知识创新重要方向性项目, KZCX2-211 号; 中国科学院海洋研究所知识创新方向性项目 (No. 2002—2005) 和海洋 863 项目 (No. 2004AA603220)。庞国兴, 硕士, E-mail: guoxingpang@sina.com

1) 通讯作者: 胡松年, 教授, E-mail: husn@genomics.org.cn

收稿日期: 2004-05-19, 收修改稿日期: 2004-10-14

等, 2002)。然而, 坛紫菜丝状孢子体和叶状配子体的 EST 目前还未见报道。

作者以坛紫菜丝状孢子体为材料, 构建了 cDNA 文库, 并进行 EST 测序及分析, 这些工作对坛紫菜的育苗及完善紫菜 EST 数据库都具有重要的意义。

1 材料与方法

1.1 坛紫菜丝状孢子体的培养

坛紫菜 (*Porphyra haitanensis*) 叶状体取自福建省坛紫菜养殖场, 从叶状体上切取一块未分化的营养细胞组织, 经过镜检、刷洗、干燥处理、离体组织的隔离培养等步骤, 获得丝状体细胞 (费修缙等, 1999)。丝状体在合适条件下可以长时间保存于实验室中。坛紫菜丝状孢子体的培养基为海水加富培养基, 培养条件为 17℃, 充气培养, 光强约为 1800lx, 光暗周期为 12h: 12h, 每星期更换一次培养基。

1.2 总 RNA 及 mRNA 的分离

采用改进的一步法提取总 RNA (Chomczynski *et al.*, 1987)。取 1.0g 坛紫菜丝状孢子体, 液氮中充分研磨, 将粉末转入盛有 9ml 变性裂解液中的 50ml 离心管中, 加入 0.9ml ($V: V= 1: 10$) 3mol/L NaAc (pH= 5), 混匀后再加入 9ml ($V: V= 1: 1$) 水饱和酚, 充分振荡, 然后加入 3ml 氯仿, 振荡, 冰浴 10min 后离心 (4℃, 12000g, 10min), 转移上清到另一离心管中, 缓慢加入无水乙醇至终浓度为 10%—20% 以沉淀多糖 (Lewinsohn *et al.*, 1994)。用氯仿抽提后加入二倍体积无水乙醇, -70℃ 沉淀至少 2h, 同样条件下离心, 收集沉淀, 加入 2ml 4mol/L LiCl, 混匀后置于 4℃ 冰箱, 2h 后分装到 1.5ml 管中, 离心 (4℃, 12000g, 10min) 后各加入 400μl DEPC 水, 用酚仿抽提二次, 取上清到另一新 1.5ml 管, 加入 40μl 3mol/L NaAc (pH= 5) 和二倍体积无水乙醇, 混匀, -70℃ 沉淀 1h, 离心 (4℃, 12000g, 15min) 后弃上清, 70% 乙醇洗涤一次, 自然晾干后用 400μl 纯水溶解, 样品冻于 -70℃ 冰箱。mRNA 的分离采用 Oligotex mRNA Kits (QIAGEN), 操作按试剂盒的说明书进行。

1.3 cDNA 文库的构建

以 mRNA 为模板, 使用 Superscript II-RT 逆转录酶和 polyT 引物合成 cDNA 的第一条链, polyT 引物 5' 端引入 *Xho* I 酶切位点, 序列为: 5' > GAGAGAGAGAGAGAGACA ACTAG TCTCGACTF-

TTTTTTTTTTTTTT < 3'。使用 DNA 聚合酶 I 合成 cDNA 的第二条链, 通过 T4 DNA 聚合酶补平 cDNA 末端, T4 DNA 连接酶加 *EcoR* I 接头, T4 PNK 将双链 cDNA 末端磷酸化, *Xho* I 酶切双链 cDNA, 回收大于 500bp 的片段。使用 T4 DNA 连接酶连接 pBluescript II SK(+) 载体和插入片段, 连接产物纯化后, 电转化感受态细胞 DH10B, 电压为 1.8kV, 电击时间为 5ms, 加入等倍体积的 30% 甘油, -70℃ 冻贮。

1.4 EST 的获取

在含有氨苄青霉素、X-gal 和 IPTG 的平板上筛选重组子, 挑取白色菌落于 96 孔板中, 过夜培养至 OD_{600} 达到 0.6—0.8, 提取质粒后电泳检测。测序 PCR 反应使用 ET 试剂盒 (Pharmacia), 操作按试剂盒说明书进行。在 PCR 仪里完成测序反应, 程序为: 95℃ 15s 接 95℃ 15s, 50℃ 15s, 60℃ 1.5min, 40 个循环, 反应完成后, 用 70% 乙醇纯化反应产物, 晾干后每孔加毛细管电泳上样缓冲液 8μl, 在 MegaBace1000 上测序。

1.5 数据处理

使用 Cross_match 软件去掉 EST 中的低质量序列、载体序列、重复序列等, 使用 Phrap 序列拼接软件将测序得到的 EST 拼接成叠连群 (多拷贝 EST), 将拼接得到的叠连群和单拷贝 EST (未能拼接的 EST) 与 GenBank 非冗余蛋白库比对并使用 GO (Gene Ontology) 分类的方法做功能注释。

GO 分类是现在较常用的 EST 分类方法, 其大体步骤是先找到每个 EST 的 BlastX 最佳匹配的注释, 然后将注释词汇在网站上搜索, 结果一般会按照三种方式 (分子功能、生物学过程、细胞组分) 给出分类的梯形结构, 下一级的分类属于上一级的大类 (Ashburner *et al.*, 2000)。除此之外, 还进行 GC 含量、密码子使用频率等分析。

2 结果

2.1 坛紫菜总 RNA 和 mRNA 的提取

取 2μl 总 RNA 样品稀释 100 倍, 用分光光度计检测: $OD_{260} = 0.324$, $OD_{280} = 0.172$, $OD_{260}/OD_{280} = 1.874$, 样品浓度约为 1300μg/ml, 所得总 RNA 的量约为 500μg。mRNA 的分离采用 Oligotex mRNA Kits (QIAGEN), 操作按试剂盒的说明书进行, 取 3μl mRNA 样品, 稀释 13 倍, 用分光光度计检测: $OD_{260} = 0.313$, $OD_{280} = 0.269$, $OD_{260}/OD_{280} = 1.730$, 得到 mRNA 约为 8μg。

2.2 cDNA 文库的构建

以 mRNA 为模板, 逆转录合成 cDNA 的第一条链, 进而合成双链 cDNA。cDNA 两端连接 *EcoR*I 接头, 然后胶回收大于 500bp 的片段。连接载体后, 电转化感受态细胞 DH10B, 得到 2.5×10^5 个克隆。随机挑选 8 个克隆, 用菌落 PCR 方法检测其插入片段的大小在 600–800bp 之间。

2.3 EST 的获取和分析

测序得到 1280 个坛紫菜丝状孢子体 EST, 去除低质量的序列和载体序列后, 长度大于 100bp 的有 714 个 EST, 使用 phrap 默认参数拼接, 拼接出的叠连群有 91 个, 单拷贝 EST 有 399 个, 使用

BlastX 软件和 GenBank 的非冗余蛋白库比对 ($e\text{-value} < 1 \times 10^{-5}$), 490 个叠连群和单拷贝 EST 一共有 275 个比对上, 占 56.1%, 未能比对上 EST 占 43.9%。对 275 个叠连群和单拷贝 EST 做 GO 分类。275 个叠连群和单拷贝 EST 中有 226 个可归属于细胞组分 (Cellular component)、分子功能 (Molecular function) 和生物学过程 (Biological process) 三个大类, 而其余 49 个叠连群和单拷贝 EST 则为未知功能的 EST。226 个 EST 的 GO 分类结果见表 1。绝大部分 EST 与生理学过程、细胞结构、酶活性及细胞过程有关, 少数与转录调节活性、伴侣蛋白活性、信号传导活性及防御/免疫蛋白活性等功能有关。

表 1 坛紫菜丝状孢子体 490 个叠连群和单拷贝 EST 的 GO 分类结果

Tab. 1 GO classification of 490 contigs and singlets of filamentous sporophyte of *P. haitanensis*

功能分类	分类	功能分类	分类
分子功能_转录调节活性	17	分子功能_不发育	3
分子功能_结合活性	113	生物学过程_生理学过程	214
分子功能_伴侣蛋白活性	5	生物学过程_细胞过程	144
分子功能_信号传导活性	22	生物学过程_发育	21
分子功能_运输活性	41	生物学过程_病毒生命循环	51
分子功能_翻译调节活性	15	生物学过程_未知生物学过程	3
分子功能_蛋白标记活性	2	生物学过程_不发育	6
分子功能_酶调节活性	2	细胞组成_毒粒	2
分子功能_抗氧化活性	7	细胞组成_免疫球蛋白复合体	3
分子功能_防御/免疫蛋白活性	4	细胞组成_细胞结构	216
分子功能_结构分子活性	59	细胞组成_细胞外结构	28
分子功能_酶活性	153	细胞组成_未定位	2
分子功能_细胞粘着分子活性	1	细胞组成_不发育	1
分子功能_传动活性	1		

将 490 个坛紫菜丝状孢子体的叠连群和单拷贝 EST 与 GenBank 中的 dbEST 中 2 万多条条斑紫菜 EST 进行比对, 同源序列有 260 个, 占 53.0%。

为了比较坛紫菜和条斑紫菜的密码子使用偏好性, 在 BlastX 的结果中挑选出与条斑紫菜同源性高的序列 ($\text{score} > 200$ or $P < 1.0 \times 10^{-15}$), 共有 28 个, 这些 EST 中所包括密码子数为 3221

个, 其中编码区的 GC 含量 (59.4%) 比 3' 非翻译区的 GC 含量 (57.1%) 略高。3221 个密码子第三位平均 GC 含量是 73.1%, 密码子第二位平均 GC 含量是 42.0%, 密码子第一位平均 GC 含量是 58.5%。

统计坛紫菜丝状孢子体 3221 个密码子的密码子使用频率表, 结果见表 2。从表中可以看出同义密码子第三位碱基偏好使用 G 或 C。

表 2 坛紫菜丝状孢子体 3221 个密码子的使用频率

Tab. 2 Usage frequency of 3221 codons of filamentous sporophyte of *P. haitanensis*

氨基酸	密码子	数量(个)	/1000	使用频率	氨基酸	密码子	数量(个)	/1000	使用频率
Gly	GGG	38	11.80	0.154	Trp	TGG	28	8.69	1.000
Gly	GGA	26	8.07	0.105	End	TGA	0	0	0
Gly	GGT	62	19.25	0.251	Cys	TGT	9	2.79	0.209
Gly	GGC	121	37.57	0.490	Cys	TGC	34	10.56	0.791
Glu	GAG	107	33.22	0.733	End	TAG	0	0	0
Glu	GAA	39	12.11	0.267	End	TAA	0	0	0
Asp	GAT	56	17.39	0.320	Tyr	TAT	18	5.59	0.222
Asp	GAC	119	36.95	0.680	Tyr	TAC	63	19.56	0.778
Val	GTG	122	37.88	0.473	Leu	TTG	15	4.66	0.052
Val	GTA	21	6.52	0.081	Leu	TTA	20	6.21	0.070
Val	GTT	47	14.59	0.182	Phe	TTT	42	13.04	0.393
Val	GTC	68	21.11	0.264	Phe	TTC	65	20.18	0.607
Ala	GCG	80	24.84	0.339	Ser	TCG	67	20.80	0.333
Ala	GCA	33	10.25	0.140	Ser	TCA	25	7.76	0.124
Ala	GCT	32	9.93	0.136	Ser	TCT	25	7.76	0.124
Ala	GCC	91	28.25	0.386	Ser	TCC	43	13.35	0.214
Arg	AGG	9	2.79	0.039	Arg	CGG	37	11.49	0.162
Arg	AGA	32	9.93	0.140	Arg	CGA	21	6.52	0.092
Ser	AGT	6	1.86	0.030	Arg	CGT	36	11.18	0.158
Ser	AGC	35	10.87	0.174	Arg	CGC	93	28.87	0.408
Lys	AAG	174	54.02	0.821	Gln	CAG	83	25.77	0.722
Lys	AAA	38	11.80	0.179	Gln	CAA	32	9.93	0.278
Asn	AAT	20	6.21	0.194	His	CAT	34	10.56	0.321
Asn	AAC	83	25.77	0.806	His	CAC	72	22.35	0.679
Met	ATG	98	30.43	1.000	Leu	CTG	122	37.88	0.425
Ile	ATA	15	4.66	0.082	Leu	CTA	11	3.42	0.038
Ile	ATT	50	15.52	0.273	Leu	CTT	50	15.52	0.174
Ile	ATC	118	36.63	0.645	Leu	CTC	69	21.42	0.240
Thr	ACG	87	27.01	0.414	Pro	CCG	61	18.94	0.389
Thr	ACA	13	4.04	0.062	Pro	CCA	30	9.31	0.191
Thr	ACT	27	8.38	0.129	Pro	CCT	20	6.21	0.127
Thr	ACC	83	25.77	0.395	Pro	CCC	46	14.28	0.293

3 讨论

坛紫菜富含多糖等物质,而多糖的理化性质与 RNA 相似,在去除多糖的同时也去除了 RNA,造成总 RNA 产量的减少。使用普通一步法所提取的总 RNA 并不能去除多糖的污染,而使用试剂盒提取坛紫菜丝状孢子体总 RNA 时, RNA 的质量不高且产量较低。采用改进的一步法获得了质量较高的总 RNA,终浓度 10%—20% 的无水乙醇可以使多糖沉淀下来而 RNA 仍保留于溶液中,另外通过高浓度 LiCl (4mol/L) 沉淀 RNA 也可以去掉部分多糖(王广策等, 2002b)。

坛紫菜丝状孢子体 EST 在 MegaBace1000 上测序时成功率较低,仅为 50% 左右,可能原因是坛紫菜丝状孢子体 EST 的 GC 含量较高,高 GC 含量的模板在测序过程中形成较复杂的二级结构,使测序酶不能通过,这可能是导致测序成功率低的主要原因。坛紫菜丝状孢子体 490 个叠连群和单拷贝 EST 平均 GC 含量为 58.9%,而条斑紫菜叶状体 3267 条叠连群和单拷贝 EST 的平均 GC 含量为 65.2% (Nikaido *et al.*, 2000),这说明坛紫菜丝状孢子体中一些高 GC 含量的 EST 在测序过程中有可能没有测出来,导致坛紫菜丝状孢子体 EST 的 GC 含量低于条斑紫菜叶状体 EST 的 GC 含量。加入甜菜碱至终浓度 0.2—2.5mol/L,测序成功率显著提高。甜菜碱(Betain)化学名为甘氨酸三甲内盐,可以降低双螺旋 DNA 的稳定性,使二级结构易打开,从而促进高 GC 含量基因的扩增(白雪源等, 2000)。

使用 BlastX 软件和 GenBank 非冗余蛋白库比对($e\text{-value} < 1 \times 10^{-5}$),未能比对上的 EST 占 43.9%,其可能原因是在 GenBank 中红藻和紫菜的数据较少。490 个坛紫菜丝状孢子体的 EST 与 dbEST 中 2 万多条条斑紫菜叶状体和丝状体 EST 进行比对,未比对上的 EST 占 47.0%,这说明坛紫菜和条斑紫菜表达的基因有可能存在一些差异。

挑选出与条斑紫菜同源性高的 28 个 EST ($\text{score} > 200$ or $P < 1.0 \times 10^{-15}$),这些 EST 所包括密码子数为 3221 个。3221 个密码子第三位平均 GC 含量是 73.1%,密码子第一位平均 GC 含量是 58.5%,密码子第二位平均 GC 含量是 42.0%,与条斑紫菜叶状体的 EST (Nikaido *et al.*, 2000) 密码子 GC 含量的比较结果见表 3。从表中看出二种紫菜 EST 的密码子第三位的平均 GC 含量远高于第一位

的平均 GC 含量和第二位的平均 GC 含量。

表 3 坛紫菜丝状孢子体与条斑紫菜叶状体密码子 GC 含量(%) 的比较

Tab. 3 Comparison of GC content (%) of codons between filamentous sporophyte of *P. haitanensis* and Leafy gametophyte of *P. yezoensis*

项 目	坛紫菜丝状孢子体	条斑紫菜
EST 平均 GC 含量	58.9	65.2
密码子第三位平均 GC 含量	73.1	79.4
密码子第二位平均 GC 含量	42.0	45.0
密码子第一位平均 GC 含量	58.5	62.2

表 1 中 GO 分类后的数据多于 226 个,这是因为基因编码的蛋白可能具有多种不同的功能,所以来源于基因的 EST 在 GO 分类中可以同时属于多个类群。在已知功能的 226 个 EST 中,有 4 个和防御/免疫蛋白活性相关,21 个与发育相关,153 个与酶相关,通过这些 EST 可以得到一些重要功能的基因,这对于将来指导坛紫菜育苗和提高坛紫菜的品质具有重要意义。

由表 2 可以看到,编码同一个氨基酸的同义密码子的第三位碱基偏好使用 G 或 C。这一性质与条斑紫菜叶状体的 EST 的密码子使用频率十分相似(Nikaido *et al.*, 2000),类似的现象在其它的红藻中也有发现,这种特性对寻找坛紫菜蛋白质在基因组中的编码区域具有重要的指导意义。研究结果表明:虽然坛紫菜丝状孢子体与条斑紫菜叶状体生长环境不同,而且属于不同的种和不同的世代,但二者 EST 的 GC 含量、密码子使用偏好性非常相似。

参 考 文 献

- 王广策,孙海宝,曾呈奎, 2002a. 三角褐指藻磷酸甘油酸变位酶基因可能侧翼序列的筛选、克隆以及序列测定. 海洋与湖沼, 33(3): 259—264 [Wang G C, Sun H B, Zeng C K, 2002. Screening, cloning and sequencing for the putative flanking region of gene encoding for phosphoglycerate mutase of *phaeodactylum tricornutum*. *Oceanologia et Limnologia Sinica*, 33(3): 259—264]
- 王广策,曾呈奎, 2002b. 两性绵霉线粒体中存在丙酮酸激酶;另一种丙酮酸激酶基因(pyk)的克隆及序列分析. 中国生物化学与分子生物学报, 18(6): 720—727

- 白雪源, 孟令英, 崔红等, 2000. 甜菜碱促进高 GC 含量 p16 基因的 PCR 扩增. 中华医学遗传学杂志, 17(3): 224—225 [Bai X Y, Meng L Y, Cui H *et al*, 2000. PCR product of p16 gene which has high GC content can be improved by betain. Chinese Journal of Medical Genetics, 17(3): 224—225]
- 陈红歌, 贾新成, 2003. 表达序列标签及其应用. 生物技术通讯, 14(1): 82—85 [Chen H G, Jia X C, 2003. Expressed Sequence Tags and their applications. Letters in Biotechnology, 14(1): 82—85]
- 杨官品, 沈怀舜, 许璞等, 2002. 条斑紫菜丝状孢子体表达序列标签分析. 高技术通讯, 12(2): 93—97 [Yang G P, Shen H S, Xu P *et al*, 2002. Analysis of expressed sequence Tags of filamentous sporophyte of *Porphyra yezoensis*. High Technology Letters, 12(2): 93—97]
- 费修缙, 曾呈奎, 1999. 经济海藻种质种苗生物学. 济南: 山东科学技术出版社, 50—53, 62—66
- 高政权, 王广策, 曾呈奎, 2003. 蛋白质组学的研究概况及其在海洋生物学中的应用, 海洋与湖沼, 34(3): 334—344 [Gao Z Q, Wang G C, Zeng C K, 2003. Outlines of studies on proteomics and its application on marine biology. Oceanologia et Limnologia Sinica, 34(3): 334—344]
- Adams M D, Kelley J M, Gocayne J D *et al*, 1991. Complementary DNA sequencing: expressed sequence Tags and Human Genome Project. Science, 252(5013): 1651—1656
- Asamizu E, Nakamura Y, Sato S *et al*, 2003. Comparison of RNA expression profiles between the two generations of *Porphyra yezoensis* (Rhodophyta), based on expressed sequence tag frequency analysis. Journal of Phycology, 39(5): 923—930
- Ashburner M, Ball C A, Blake J A *et al*, 2000. Gene ontology: tool for the unification of biology. Nat Genet, 25(1): 25—29
- Chomczynski P, Sacchi N, 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem, 162(1): 156—159
- Lee E K, Seo S B, Kim T H, 2000. Analysis of expressed sequence tags of *Porphyra yezoensis*. Mol Cells, 10(3): 338—342
- Lewinsohn E, Steele C L, Groteau R, 1994. Simple isolation of functional RNA from woody stems of gymnosperms. Plant Mol Biol Rep, 12(1): 20—25
- Nikaido I, Asamizu R, Nakajima M *et al*, 2000. Generation of 10154 expressed sequence tags from a leafy gamatophyte of a marine red alga, *Porphyra yezoensis*. DNA Research, 7(3): 223—227
- Yoshida T, Natoya M, Kikuchi N *et al*, 1997. Catalogue of species of *Porphyra* in the world, with special reference to the type locality and bibliography. Nat Hist Res, 3(special issue): 5—18

ACQUIREMENT AND ANALYSIS OF EXPRESSED SEQUENCE TAGS OF FILAMENTOUS SPOROPHYTE OF *PORPHYRA HAITANENSIS*

PANG Guo-Xing, WANG Guang-Ce¹, HU Song-Nian², ZENG Cheng-Kui (C. K. Tseng)¹

¹(Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071; Graduate School, Chinese Academy of Sciences, Beijing, 100039)

²(Genome Center, Chinese Academy of Sciences, Beijing, 101300)

¹(Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071)

Abstract *Porphyra haitanensis* is only cultured in south China. The yield of *P. haitanensis* is much larger than that of *P. yezoensis*, but the economic value of *P. haitanensis* is much less than that of *P. yezoensis*. Therefore, it is important and helpful to study on *P. haitanensis* in order to improve the quality and economic value of *P. haitanensis*.

Polysaccharide is abundantly contained in *P. haitanensis*. Because physical and chemical characters of polysaccharide are similar to those of RNA, it is difficult to extract RNA from *P. haitanensis*. In this paper, RNA was extracted by improved one-step method. Ethanol was slowly added to the aqueous phase until the concentration of ethanol reached 10%—20%. After centrifugation, polysaccharide was precipitated and RNA remained in the solution. In addition, high concentration of LiCl solution could also remove some polysaccharide. mRNA was separated by Oligotex mRNA Kits (QIAGEN).

cDNA library of filamentous sporophyte of *P. haitanensis* was constructed. Recons from cDNA library were taken out on Petri dish which contained penicillin, X-gal and IPTG. White colonies were taken to 96-well plates and cultured over night until OD_{600} reached 0.6—0.8. Plasmids were extracted and examined by electrophoresis. ET kits (Pharmacia) were used in the reactions for sequencing and the products of reactions were precipitated and purified. Dissolved with 8 μ l loading buffer (Pharmacia), the products were sequenced on MegaBace1000.

Low quality sequences, vector sequences and repeated sequences of ESTs were removed by Cross match and ESTs were clustered into contigs by Phrap. 91 contigs and 399 singlets were obtained.

Each sequence from the 490 contigs and singlets was subjected to similarity search against the NCBI-provided non-redundant protein database, nr, using the BlastX program. 275 (56.1%) of the 490 clustered ESTs showed sequence similarity to genes registered in the public databases (e -value $< 1 \times 10^{-5}$) and new sequences occupied 43.9%. 275 contigs and singlets were classified via GO (Gene Ontology) method, among which 226 contigs and singlets could be classified into three sorts: cellular component, molecular function and biological process. The function of the other 49 contigs and singlets were unknown.

Similarity search against over 20 000 ESTs of *P. yezoensis* showed that 260 (53.0%) of the 490 clustered ESTs showed sequence similarity to ESTs registered in the GenBank. 28 ESTs showed sequence similarity with the ESTs of *P. yezoensis* were selected and 3221 codons were contained in these ESTs. The average GC content of the third letter in the codons (73.1%) is much higher than that of the first letter (42.0%) and the second letter (58.5%). Frequency of usage of 3221 codons was analyzed. Research results shows that the GC content and usage frequency of ESTs from *P. haitanensis* and *P. yezoensis* are quite homological though they live in different environment and belong to different genus.

Key words *Porphyra haitanensis*, EST, GC content, Codons