

长心卡帕藻(*Kappaphycus alvarezii*)表达 序列标记分析*

刘晨临¹ 黄晓航¹ 刘建国²

(1. 国家海洋局第一海洋研究所 海洋生态研究中心 青岛 266061;

2. 中国科学院海洋研究所 青岛 266071)

提要 分别采用 BlastX 和 Blast2go 软件对 453 条非冗余的长心卡帕藻 EST 序列进行匹配分析和 GO 注释。结果表明, 有 281 条序列与蛋白数据库中序列匹配, 有 132 条序列被归属为 3 个子本体: 分子功能、生物学过程和细胞组成。同时筛选到 24 个与逆境适应相关的基因, 包括应对胁迫反应和抗氧化反应的基因。采用 GCUA 软件, 进行长心卡帕藻基因密码子偏倚性的研究。结果表明其密码子偏倚性与其它红藻差别较大, 平均 G+C 含量和密码子的第三位的 G+C 含量均较其它藻类偏低。上述工作为进一步研究长心卡帕藻基因及其功能奠定了基础。

关键词 长心卡帕藻, 表达序列标记分析, 抗逆基因, 密码子使用频率

中图分类号 Q93

长心卡帕藻(*Kappaphycus alvarezii*)属红藻门、真红藻纲、杉藻目、红翎菜科、卡帕藻属, 是热带、亚热带多年生海藻。该藻经济价值较高, 是生产卡拉胶的主要原料。卡拉胶在食品和轻化工业领域有广泛的应用。另外, 卡拉胶还有一定的药学功效(Bixler, 1996)。目前世界上卡拉胶海藻的养殖水平还远远不能满足对卡拉胶的需求。我国卡拉胶养殖从 1987 年首次引进麒麟菜开始至今, 养殖规模仍然不大, 卡拉胶需求主要依赖进口。

卡帕藻的人工养殖主要是通过营养体繁殖。其系统比较脆弱, 易遭受鱼类、杂藻和微生物等侵害; 另外, 养殖环境的变化, 例如光照、盐度、温度和重金属等都会对卡帕藻的生物量及其所产卡拉胶的含量和强度产生影响(Ask *et al.*, 2002)。为提高我国产卡拉胶海藻的产量和卡拉胶质量, 除了要改进养殖技术, 发展有性生殖和实验室育苗技术外, 研究卡帕藻分子水平的抗逆机制, 筛选抗逆性强、生长速度快和卡拉胶产量高的优良品系并进行推广也是非常重要的工作。

表达序列标记(EST)代表生物体某种组织特定时间的某个表达基因。基因本体(GO)数据库提供各种物

种的基因及其编码蛋白的属性分析, 它包括三种不同的子本体, 反映生物过程、细胞组分和分子功能这三个结构层次。一个基因产物可能仅被一个子本体注释, 也可能被三个子本体同时注释(Ashburner *et al.*, 2000)。对 EST 序列进行 GO 注释, 有助于了解基因及其编码产物在大多数物种中的基本功能, 参与的生物学过程及其细胞定位。因此 GO 注释是筛选和开发物种抗逆功能相关基因的有效手段。

本文将建立卡帕藻 cDNA 文库, 获得 EST 数据进行 GO 分析, 筛选其可能与逆境胁迫相关的功能基因, 并对基因特征进行分析。为后续筛选卡帕藻抗逆品系, 研究功能基因在卡帕藻适应逆境胁迫环境的机制和作用奠定基础。

1 材料与方法

1.1 试验材料

长心卡帕藻(*Kappaphycus alvarezii*)于 2008 年 7 月采自海南省陵水县养殖场。低温保存带回实验室后, 于解剖镜下用灭菌海水和毛笔洗刷掉附着的杂藻等, 取嫩枝经 2h 低盐胁迫处理(盐度为 28)后, 冻存于-80℃

* 国家海洋公益项目, 2012418039 号、200705010 号; 国家海洋局基本科研业务经费项目, 2008G19 号。刘晨临, 博士, 研究员, E-mail: ch.liliu@163.com

收稿日期: 2010-05-13, 收修改稿日期: 2010-07-15

冰箱, 用于 RNA 提取和 cDNA 文库构建。

1.2 RNA 提取和 cDNA 文库构建

卡帕藻的 RNA 提取采用的是 CTAB-LiCl 法(Wu *et al.*, 2010)。提取的总 RNA, 用 mRNA 分离纯化试剂盒(Qiagen, 德国)纯化 mRNA, 用于 cDNA

合成和文库构建。使用 ZAP- cDNA Gigapack III Gold Cloning Kit 试剂盒(Stratagene, 美国)合成 cDNA (秦玉明等, 2010), 并构建文库, 具体方法参照试剂盒说明书。建成的噬菌体文库通过切割将目的片段转入 *E. coli* SOLR 菌株中, 随机挑选单菌落送交上海博尚生物技术有限公司测序。

1.3 数据处理和 GO 分析

用 Codon Code Aligner 软件裁去所测序列的载体部分, 并去除测序质量不佳的序列。用相同的软件对获得的 EST 序列中的冗余序列拼接成叠连群(contig)和其它非重复的 EST 一起作为假定的单一(非冗余)基因 TUGs(tentative unique genes)进行分析。

用 Blastcl3 软件对所有 TUGs 进行批量分析, 包括用 Blastx 程序与蛋白数据库(refseq_protein database)序列进行比对; 用 Blastn 与其它物种的 EST 数据库进行比对。用 Blast2go 软件对所有 TUGs 进行 GO 注释。密码子偏倚性和 G+C 含量分析等采用 GCUA: General Codon Usage Analysis v1.0 软件进行(<http://bioinf.may.ie/>)。

1.4 基因序列比对

基因序列比对采用 DNASTar 软件包中的 MegAlign 程序, 与卡帕藻热激蛋白 HSP70 序列比较的其他序列均来自 Boorstein 等(1994)。

2 结果

2.1 长心卡帕藻 cDNA 文库构建和 EST 序列的基本特征

从低盐胁迫处理后的长心卡帕藻提取了 RNA 并构建了 cDNA 文库。经检测, 所构建的卡帕藻的初始文库滴度为 1.24×10^6 , 重组效率为 98%。文库质量符合要求, 完全能够满足研究需要。文库转化至细菌质粒后, 挑取了 48 个克隆验证克隆到的序列片段的长度(图 1), 其中有 10 个克隆长度超过 2000bp, 有 5 个克隆不到 500bp, 表明构建文库的 cDNA 完整性是

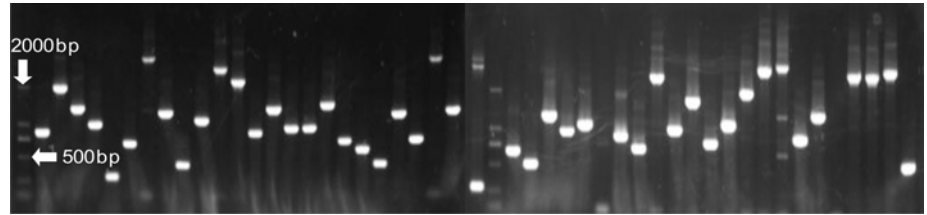


图 1 从 cDNA 文库中随机挑选克隆的插入片段长度分析

Fig.1 Analysis of the length of inserted sequences selected randomly from the cDNA library of *K. alvarezii*

符合要求的。

从卡帕藻的 cDNA 文库中随机挑选了 523 个克隆进行单向测序, 去掉 5 个测序结果不可用的克隆, 共获得 518 条 EST 序列。EST 序列的组成见表 1。经 CodonCode Aligner 分析, 518 条 ESTs 中的 109 条序列构成 44 个叠连群(contig), 另外 409 条序列为非重复序列。因此卡帕藻的 EST 分析共得到 453 个非冗余序列, 冗余率为 21%。通过 BlastX 比对分析, 453 条 ESTs 中与蛋白数据库中序列匹配的 TUGs 有 281 条($E = 10^{-3}$), 无匹配结果的序列有 172 条。通过 Blastn 比对 EST 数据库, 发现 453 条 TUGs 中有 112 条与其它红藻的 EST 序列具有较高的相似性($E = 10^{-3}$), 这些红藻包括麒麟菜、紫菜、江蓠和角叉菜等。

表 1 长心卡帕藻的表达序列标记特征
Tab.1 The characteristics of expressed sequence tags of *K. alvarezii*

项目	数目	百分比(%)
总 EST	518	—
叠连群	44	—
单一序列	409	—
非冗余序列(TUGs)	453	—
冗余率	—	21.0
已知的 TUGs	132	29.1
无 BLAST 结果	172	—
未注释	80	—
未成路径图	69	—
未知的 TUGs	321	70.9
与其它红藻 EST 匹配	112	24.5

注: 百分比计算方法为: 冗余率的百分比 = (总 EST 数 - 单一序列数) / 总 EST 数; 其它百分比均为占非冗余序列(TUGs)的百分比

2.2 长心卡帕藻 EST 序列的 GO 注释

用 Blast2go 软件对 453 条 TUGs 进行 GO 注释, 其中 132 条序列根据 GO 注释被归属为 3 类子本体: 分子功能、生物学过程和细胞组成(图 2)。在分子功能子本体内, 主要是参与结合活性和催化活性的基因, 在 GO 分类的条目中分别有 82(46%)个和 72(41%)个, 其次是参与细胞结构分子活性的 GO 条目有

12(7%)个;在生物学过程子本体内,以参与代谢过程和细胞过程的GO分类条目为主,分别为84(34%)和73(31%)个,其次是参与生物调节(15个)、刺激反应(13个)、过程调节(12个)和定位(11个)的GO分类条目,参与发育过程和多细胞有机体过程的GO条目分别有9和8个;在细胞组成子本体内,主要是位于细胞(75, 35%)、细胞区域(cell part)(72, 33%)和细胞器(53, 24%)这三类GO分类条目,属于大分子复合体的GO有12(5%)个。

根据BlastX和GO注释的结果,发现了卡帕藻一些潜在的与环境胁迫相关或参与活性氧反应的基因共24个(表2)。这些基因主要来自分子功能子本体的催化活性条目和生物学过程子本体的刺激反应条目。

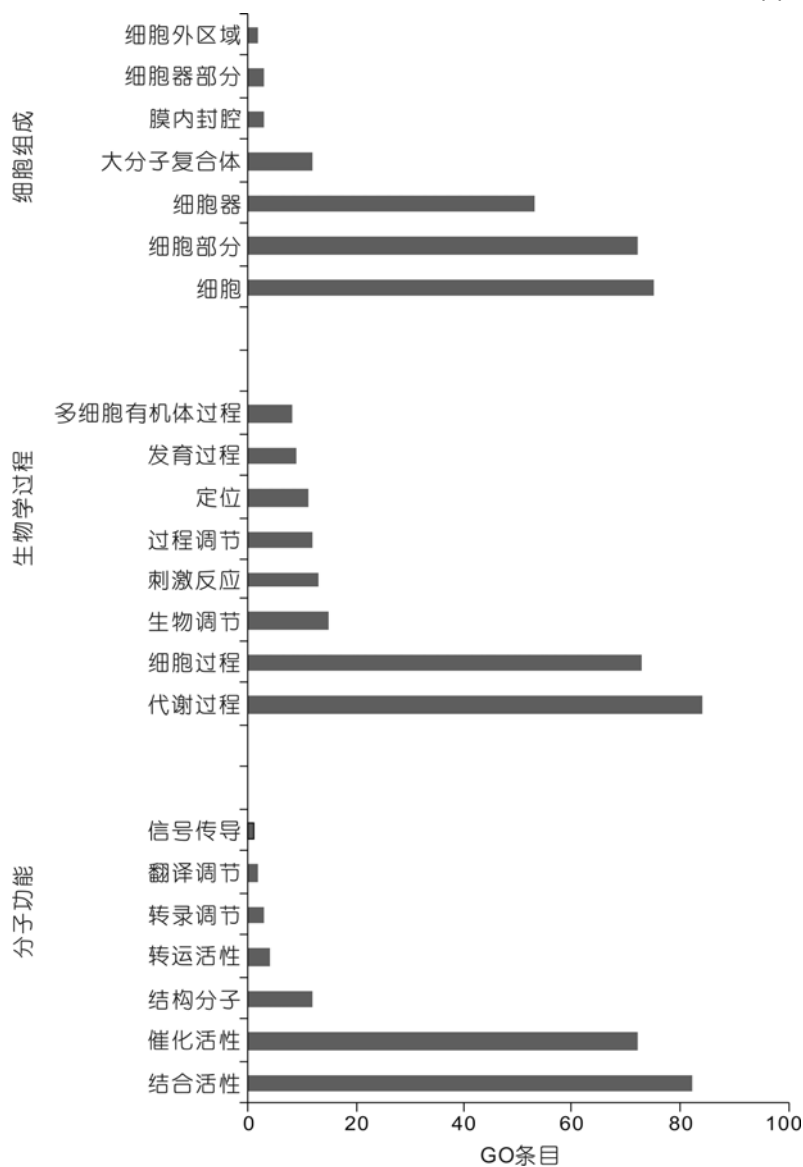


图2 长心卡帕藻132个已知序列的基因本体分类

Fig.2 Gene ontology annotation of 132 known sequences cloned from *K. alvarezii*

其中与镉离子胁迫相关的基因有3个,分别是ATP硫酸化酶、金属内肽酶和内质网ATP酶基因。热激蛋白基因2个,分别为HSP70和辅助伴侣GrpE基因。参与抗氧化系统的基因有13个,包括2个不同的硫氧还蛋白序列、抗氧化蛋白6、过氧化氢酶过氧化物酶HPI以及钒依赖的溴过氧化物酶基因等。

2.3 长心卡帕藻的密码子使用频率

选取132条已知EST序列编码区,用GCUA v1.0分析其密码子使用频率。选取的序列共编码了23251个氨基酸,其中每种氨基酸密码子出现的次数及同义密码子的使用频率(Relative Synonymous Codon Usage, RSCU)见表3。卡帕藻基因编码序列的平均G+C含量为0.473。密码子第三位为C、T、A和G的几率分别为0.259905、0.279969、0.199211和0.253287。第三位G+C含量为0.513。

3 讨论

基于红藻门物种重要的经济价值,对以紫菜为代表的许多红藻都开展了程度不同的基因组学研究。产琼胶和卡拉胶的红藻中,GenBank的EST数据库已经有包括江蓠、龙须菜、角叉菜和麒麟菜的EST序列约17000条。但是对卡帕藻EST分析显示,仍然有70%左右的基因为未知序列,与其它红藻的EST相比,也只有约24.5%具有同源性,其中包括与卡帕藻同属红翎菜科的异枝麒麟菜(*Euclima denticulatum*)的2000余条EST序列(Aspilla *et al.*, 2010)。说明对于红藻基因功能方面的研究,其中既包括对已知序列也包括对未知序列功能的探索还有很多工作要做。在构建卡帕藻cDNA文库,进行EST表达序列标记分析的基础上,作者还将通过构建基因芯片的方式,研究所得到的卡帕藻EST序列在不同生理条件和胁迫条件下的表达特征。

在所检测的518个与低渗胁迫相关的卡帕藻cDNA克隆中,有3个克隆属于热激蛋白HSP70基因(Kp141、Kp168和Kp513),其出现的频率较高。热激蛋白(HSP)作为分子伴侣参与蛋白质的正确折叠、聚合、运输和信号传递等重要生理过

表 2 长心卡帕藻与胁迫相关的基因
Tab.2 Stress resistance related genes from *K. alvarezii*

ID	功能	E 值	相关胁迫
Contig22	丙酮酸脱氢酶(<i>Arabidopsis thaliana</i>)	7.00E-93	非生物胁迫
KP129	ATP 硫酸化酶(<i>Arabidopsis thaliana</i>)	1.72E-84	镉离子胁迫
KP136	金属内肽酶(<i>Vitis vinifera</i>)	3.65E-28	镉离子胁迫
KP45	ATP 酶 TER94 (<i>Hemiselmis andersenii</i>)	1.70E-35	镉离子胁迫
KP486	细胞凋亡相关蛋白激酶 ATM (<i>Branchiostoma floridae</i>)	1.28E-49	缺氧和离子放射性胁迫
Contig19	热激蛋白 70 (<i>Gracilaria tenuistipitata</i> var. <i>liui</i>)	6.00E-96	热和氧化胁迫
KP217	热激蛋白辅助伴侣 GrpE (<i>Synechococcus</i> sp. JA-3-3Ab)	8.09E-32	热和氧化胁迫
Contig36	甘油醛 3 磷酸脱氢酶(<i>Chondrus crispus</i>)	7.31E-86	渗透压胁迫
KP335	真核翻译起始因子(<i>Nematostella vectensis</i>)	6.76E-27	生长激素刺激
KP392	淀粉磷酸化酶(<i>Cyanophora paradoxa</i>)	3.80E-41	脱水作用
KP491	DNA 损伤识别蛋白 Rad 24 (<i>Arabidopsis thaliana</i>)	5.45E-49	外部刺激
KP012	硫氧还蛋白(<i>Cellvibrio japonicus</i> Ueda107)	3.00E-25	抗氧化
KP220	铁氧化还原蛋白依赖的谷氨酸合成酶(<i>Shewanella benthica</i> KT99)	2.78E-37	抗氧化
KP263	铁蛋白 1a (<i>Physcomitrella patens</i> subsp. <i>Patens</i>)	3.76E-26	抗氧化
KP299	抗氧化蛋白 6 (<i>Physcomitrella patens</i> subsp. <i>Patens</i>)	2.24E-33	抗氧化
KP302	硫氧还蛋白(<i>Porphyr a yezoensis</i>)	1.00E-42	抗氧化
KP304	磷酸葡萄糖酸盐脱氢酶(<i>Porphyr a yezoensis</i>)	3.35E-48	抗氧化
KP358	钒依赖的溴过氧化物酶(<i>Acaryochloris marina</i> MBIC11017)	8.30E-52	抗氧化
KP401	过氧化氢酶过氧化物酶 HPI (<i>Pseudoalteromonas tunicata</i> D2)	9.47E-26	抗氧化
KP409	谷氨酰-tRNA 还原酶(<i>Pseudoalteromonas haloplanktis</i> TAC125)	6.56E-109	抗氧化
KP454	4-羟苯丙酮酸二加氧酶(<i>Anoplopoma fimbria</i>)	1.65E-22	抗氧化
KP511	维生素 B ₁₂ 生物合成蛋白(<i>Picea sitchensis</i>)	5.48E-11	抗氧化
KP524	谷氨酸脱氢酶(<i>Pseudoalteromonas haloplanktis</i> TAC125)	1.42E-68	抗氧化
KP527	原卟啉原氧化酶(<i>Rhodothermus marinus</i> DSM 4252)	7.72E-17	抗氧化

表 3 基于 132 条 EST 序列计算的长心卡帕藻的密码子使用频率
Tab.3 Codon usage analysis of *K. alvarezii* based on 132 known ESTs

AA	Codon	N	RSCU	AA	Codon	N	RSCU	AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	UUU	545	1.1	Ser	UCU	283	1.1	Ile	AUU	648	1.47	Thr	ACU	341	1.09
	UUC	445	0.9		UCC	292	1.13		AUC	448	1.01		ACC	325	1.04
Leu	UUA	241	0.68	Cys	UCA	225	0.87	Met	AUA	230	0.52	Ser	ACA	302	0.97
	UUG	593	1.67		UCG	268	1.04		AUG	529	1		ACG	279	0.89
Tyr	UAU	385	1.02	ter	UGU	178	0.91	Asn	AAU	500	1.02	Arg	AGU	264	1.02
	UAC	371	0.98		UGC	212	1.09		AAC	483	0.98		AGC	218	0.84
ter	UAA	17	0	Trp	UGA	15	0	Lys	AAA	518	0.71	Arg	AGA	245	0.98
	UAG	11	0		UGG	284	1		AAG	939	1.29		AGG	264	1.05
Leu	CUU	381	1.07	Pro	CCU	209	0.85	Val	GUU	437	1.07	Ala	GCU	467	1.03
	CUC	351	0.99		CCC	208	0.85		GUC	444	1.08		GCC	464	1.03
	CUA	179	0.5		CCA	259	1.06		GUA	259	0.63		GCA	478	1.06
	CUG	382	1.08		CCG	303	1.24		GUG	499	1.22		GCG	397	0.88
His	CAU	342	1.14	Arg	CGU	338	1.35	Asp	GAU	730	1.13	Gly	GGU	458	1.14
	CAC	259	0.86		CGC	306	1.22		GAC	557	0.87		GGC	390	0.97
Gln	CAA	332	0.84	CGA	182	0.73	Glu	GAA	614	0.9	GGA	459	1.14		
	CAG	458	1.16		CGG	169		0.67	GAG	745		1.1	GGG	297	0.74

程(Vierling, 1991)。高温、低温、干旱和化学因子等环境胁迫条件都能促进 HSP70 的表达(Guy *et al*, 1998; Lin *et al*, 2001; Sung *et al*, 2001)。高 HSP70 表达量的烟草植株比对照更耐干旱和衣霉素胁迫(Cho *et al*, 2006)。热休克蛋白基因的表达量除了与耐热性呈正相关外, 同时还能够提高植物对盐、水和高温胁迫的耐受性(Guy *et al*, 1998)。因此, 有人提出将 Hsp70 作

为生物胁迫耐受的标志物(Ireland *et al*, 2004)。光合真核生物有 4 种类型的 HSP70 蛋白, 分别位于细胞质、线粒体、内质网和叶绿体(Renner *et al*, 2007)。根据 Boorstein 等(1994)的文献, 选取不同类型的 HSP70 氨基酸序列, 与作者克隆到的卡帕藻的 HSP70 序列进行比较, 发现卡帕藻热激蛋白 Kp513 为叶绿体分布类型, Kp168 和 Kp141 为细胞质分布类型(图 3)。

Kp618		0
Kp513	RQAVMNPENTFYYSVKRFITGR--KKEEVTNLSRQLSINVKTDSDNSNLIKLECPALN--KDFAPPEEISAQVLRKLVEDAGAYLGGAVTQAVITVPAYFNDSSQRQATK	100
Kp141		0
BOV. C	NQVAMNPTNTVFDAKRLIGRRFDAAVQSDMKIHPFMVYN-DAGRPKVQVEYKG-ETKSFYPEEVSMMVLTKMKEIAEAYLGKTVTNAVVTVPAYFNDSSQRQATK	159
SSA1. C	NQAAAMPNTVFDAKRLIGRRFNDEPVQADMKHFPPKFLID-VDGKPKIQVEFKG-ETKNFTPEQISSMVLGKMKETAESYLGAQVNDAVVTVPAYFNDSSQRQATK	157
HOMO. E	NQLTSPNPTVFDAKRLIGRTWNPSPVQDDIKFLPFKVVVE-KTKPYIQVDIGGGQTKTFAPEEISAMVLTMMKETAAYLGKVVTHAVVTVPAYFNDSSQRQATK	185
TOB. E	NLAAVNPRTVFDKRLIGRKFDDKEVQRDMKLVYKIVN-KDGKPIQVLIKDGGETKIFSPPEEISAMILTKMKETAAYLGKVKIKDAVVTVPAYFNDSSQRQATK	193
SSC1. M	RQAVVNPENTLFAKRLIGRRFDEAEVQRDIKQVPYKIVKHSNGDAWVEARGQ---TYSQAIGGFVNLKMKETAAYLGKPVKNNAVVTVPAYFNDSSQRQATK	183
PEA. M	RQAVNPTNTLFGTKRLIGRRFDAAQTQKEMKMPYKIVRAPNGDAWVEANGQ---QYSPSQIGAFVLTIKETAAYLGKTSKAVVTVPAYFNDSSQRQATK	205
POR. P	RQAVINSEBTFYYSVKRFITGR--PSKEVSDLELRITPKIETD-SEGKIRLKCPLN--KDFAPPEEISAQVLRKLVNDANKYLGEKVEKAVITVPAYFNDSSQRQATK	153
PAV. P	RQAVINSEBTFYYSVKRFITGR--PSKEVSDLELRITPKIETD-SEGKIRLKCPLN--KDFAPPEEISAQVLRKLVNDANKYLGEKVEKAVITVPAYFNDSSQRQATK	154
Kp618		0
Kp513	DAGKIAGLDVLRINEPTAASISYGLDK--KNNETILVFDLGGGTFDVSILEVGDGVFEVLSTSGDTILGGDDFDKIVSWLINEFRNDEGINLSQDRAQALQRL	202
Kp141		0
BOV. C	DAGTIAGLNLVLRINEPTAAAIAYGLDK-KVGAERNVLIFFDLGGGTFDVSILTIEDGIFEVKSTAGDTHLGGEDFDNRMVNHFIAEFKRKHKKIDISENKRVRRL	263
SSA1. C	DAGTIAGLNLVLRINEPTAAAIAYGLDK-K-GKEEHLVLIFFDLGGGTFDVSILTIEDGIFEVKATAGDTHLGGEDFDNRVNHFIQEFKRKKNKDLSTNQRALRRL	260
HOMO. E	DAGTIAGLNMVLRINEPTAAAIAYGLDK-R-EGKELVLIFFDLGGGTFDVSILTIIDNGVFEVATNGDTHLGGEDFDQRVMEHFIKLYKKTGPDVFKDNRVQKL	288
TOB. E	DAGVIAGLNVLRINEPTAAAIAYGLDK-K-GGKELVLIFFDLGGGTFDVSILTIIDNGVFEVLTNGDTHLGGEDFDQRIMEYFIKLIKKHGKIDSKDNRALGKL	296
SSC1. M	DAGQIVGLNLVLRVNEPTAAALAYGLEK---SDSKVAVFDLGGGTFDVSILDIIDNGVFEVKTNGDTHLGGEDFDIYLLREIVSRFKTETGIDLENDRMATQRI	285
PEA. M	DAGRIAGLDVQRIINEPTAAALSYGMN---KEG-LIAVFDLGGGTFDVSILEISNGVFEVKTNGDTHLGGEDFDNALDLVSEFKRTEIDLAKDKLALQRL	306
POR. P	DAGKIAGLDVLRINEPTAASISYGLDK--QNNETILVFDLGGGTFDVSILEVGDGVFEVLSTSGDTILGGDDFDQIVLWIKDFKQSEGDILGKDRQALQRL	255
PAV. P	DAGKIAGLDVLRINEPTAASISYGLDK--KDNETILVFDLGGGTFDVSILEVGDGVFEVLSTSGDTILGGDDFDEKIVKWLNEFEKEKESLKGDSQALQRL	256
Kp618		0
Kp513	TEASEKAKMELSSILQTDINLPIITSTDTGPKILKRTITRAKFEELCSDFIDRCKLPIVNAFEKNAQIKC	272
Kp141		56
BOV. C	RTACERAKRTLSSTQASIEIDSLYE---GIDFYTSITRARFEELNADLFRGTLDPVEKALRD-AKLDKQSHDITVLVGGSTRIPKIQKLLQDFENKELNKSIT	363
SSA1. C	RTACERAKRTLSSSAQTSVEIDSLFE---GIDFYTSITRARFEELCADLFRGTLDPVEKALRD-AKLDKQSDVEITVLVGGSTRIPKIQKLLQDFENKELNKSIT	360
HOMO. E	RREVEKAKRALSSQHQARIEIESFYE---GEDFSETLTRAKEELNMDLFRSTMKPVQKVLVD-SLKKSDIDEITVLVGGSTRIPKIQKLVKFEFNGKEPNSRGI	388
TOB. E	RREAERAKRALSSQHQVRYVEIESLFD---GVDSEPLTRARFEELNMDLFRSTMGVVKAMDD-AGLEKQIIDEITVLVGGSTRIPKIQKLLQDFDGGKEPNKGV	396
SSC1. M	REAAEKAKIELSSVSTIEINLPIITADASGPKHINMFSRAQFETLITAPLVKRTVDPVVKALKAD-AGLSTSDISEVLLVGGMSRMPKVVETVRSFLV-GKDPKAV	388
PEA. M	REAAEKAKIELSSVSTIEINLPIITADASGAKHILNITLRSKFEALVNLITERTKAPCKSLCKD-ANISIKDVDEVLLVGGMTRVPKVVQVSEIF-GKSPSKGV	409
POR. P	TEASEKAKIELSSVSTIEINLPIITATQDQPKILKRTITRAKFEELCSDFIDRCKLPIVNAFKALD-AKLEASSIDEVLLVGGSTRIPAIQVMKRLI-GKDPQSV	358
PAV. P	TEAAEKAKIELSSVSTIEINLPIITANENKAKILKRTITRAKFEELCSDFIDRCKLPIVENALD-AKLPKQIDEVLLVGGSTRIPAVKLVKLDIL-GKEPNETV	359
Kp618		51
Kp513	NTTIPTKESSRYSPTLTSILVCFQVFEGERITRTRDNNILGKFEISGIPPA	272
Kp141		160
BOV. C	NPDEAVAYGAAVQAAILLAGDT-SEKTQDLLLDDVITPLSLGIEIETAGGVMTVLIPRNTIIPITKKSQIFSTYADNQPGVLIQVFEGERITRTRDNNILGKFEISGIPPA	467
SSA1. C	NPDEAVAYGAAVQAAILLAGD-SSKTQDLLLDDVITPLSLGIEIETAGGVMTKIPRNTIIPITKKSQIFSTYADNQPGVLIQVFEGERAKTKDNNILGKFEISGIPPA	464
HOMO. E	NPDEAVAYGAAVQAGVLSGD---QDTGDLVLLDVCPITLGIETVGGVMTKIPRNTIIPITKKSQIFSTYADNQPGVLIQVFEGERITRTRDNNILGKFEISGIPPA	490
TOB. E	NPDEAVAYGAAVQAGVLSGEG-GDETKDILLDDVAPLTLGIETVGGVMTKIPRNTIIPITKKSQIFSTYADNQPGVLIQVFEGERITRTRDNNILGKFEISGIPPA	500
SSC1. M	NPDEAVAIGAAVQAGVLSG---EVTDVLLDDVITPLSLGIEIETVGGVMTKIPRNTIIPITKKSQIFSTYADNQPGVLIQVFEGERITRTRDNNILGKFEISGIPPA	488
PEA. M	NPDEAVAMGAALQGGILRG---DVKELLLDDVITPLSLGIEIETVGGVMTKIPRNTIIPITKKSQIFSTYADNQPGVLIQVFEGERITRTRDNNILGKFEISGIPPA	509
POR. P	NPDEVVAIGAAVQAGVLAG---EVKDLLLDDVITPLSLGIEIETVGGVMTKIPRNTIIPITKKSQIFSTYADNQPGVLIQVFEGERITRTRDNNILGKFEISGIPPA	458
PAV. P	NPDEVVAIGAATQAGVLSG---EVKDLLLDDVITPLSLGIEIETVGGVMTKIPRNTIIPITKKSQIFSTYADNQPGVLIQVFEGERITRTRDNNILGKFEISGIPPA	459
Kp618	PRGVPQIEVTFDIDANGILNVSAEDKSGKKNKITTNDKGRITKEDIEMVQDAEKFAEDDAAKAKIERNLSENYCYTLRNTINE-DKVKDKISEGDKKSI	155
Kp513		272
Kp141	PRGVPQIEVTFDIDANGILNVSAEDKSGKKNKITTNDKGRITKEDIEMVQDAEKFAEDDAAKAKIERNLSENYCYTLRNTINE-DKVKDKISEGDKKSI	264
BOV. C	PRGVPQIEVTFDIDANGILNVSAEDKSGKKNKITTNDKGRITKEDIEMVQDAEKFAEDDAAKAKIERNLSENYCYTLRNTINE-DKVKDKISEGDKKSI	571
SSA1. C	PRGVPQIEVTFDIDANGILNVSAEDKSGKKNKITTNDKGRITKEDIEMVQDAEKFAEDDAAKAKIERNLSENYCYTLRNTINE-DKVKDKISEGDKKSI	566
HOMO. E	PRGVPQIEVTFDIDANGILNVSAEDKSGKKNKITTNDKGRITKEDIEMVQDAEKFAEDDAAKAKIERNLSENYCYTLRNTINE-DKVKDKISEGDKKSI	595
TOB. E	PRGVPQIEVTFDIDANGILNVSAEDKSGKKNKITTNDKGRITKEDIEMVQDAEKFAEDDAAKAKIERNLSENYCYTLRNTINE-DKVKDKISEGDKKSI	605
SSC1. M	PKGVQIEVTFDIDADGIINVSARDKATNKDSSITVAGSSG-LENEIEQMVNDAEKFKSQDEARQAETANKADQLANDTENSLEK---EFEGKVDKAEAKQVR	589
PEA. M	PRGLPQIEVTFDIDANGILNVSAEDKSGKKNKITTNDKGRITKEDIEMVQDAEKFAEDDAAKAKIERNLSENYCYTLRNTINE-DKVKDKISEGDKKSI	610
POR. P	PRGVPQIEVTFDIDANGILSVTAQDKGTSKQSSITISGAST-LPKDEEVMKAEAEQNAADKKEGENIRVKNEADLYCYQAEKQIS---ELPEALVNEQSLIK	559
PAV. P	PRGVPQIEVTFDIDANGILSVTAQDKGTSKQSSITISGAST-LPKDEEVMKAEAEQNAADKKEGENIRVKNEADLYCYQAEKQIS---ELPEALVNEQSLIK	560
Kp618	DKVQEVLTWMEENETAKEEYEAQKQSVFVATPIITKMYSGGGEGMPGMPGMPD--MGG-AAGATGGDDGVEISEVD	231
Kp513		272
Kp141	DKVQEVLTWMEENETAKEEYEAQKQSV	292
BOV. C	DKCNEIINWLDKNTAEKEEFHQKLEKVCNPIITKLYQSAGGMPGMPGMPGPPGGAPPAGPAPPAP-FAEGPTVEEVD	650
SSA1. C	KKAEETISWLDNNTASKEEFDKLLKELQDIANPMSKLYQ-AGGAPGGAAGGAPGPPGG-APPAP-FAEGPTVEEVD	642
HOMO. E	KAVEEKIEWLESHQDADIEDFKAKKKELEEIVQPIISKLYGSAGPPPTGEEDTAE-----KDEL	654
TOB. E	TATKEALEWLDNNSAEKEDYEELKEVEAVCNPIITAVYQKSGGAPGGESGASED-----D---DHDEL	667
SSC1. M	D---QITSLKELVARVQGGEEVNAEELKTKTEELQTSSMKLFEQ---LYKNSNNNN--NN-----NGNAESGETKQ	654
PEA. M	D---AVSDLRTAMAGENADDIKAKLDAANKAVSKIGQHMSGGSS---GGPSEGGSQ-GE-----QAPEAEYEEVKK	675
POR. P	N---LISELRSNLEKEELDSEANSEKLNALMEIGKNATS---AEKDTQNASN--DD-----TVIDTDFSEAK	620
PAV. P	ESKETVEMKENTIKKEDYDKIKENKLLQEKIMEIGQKAYAKKEPLKDEDSNKGSS---QD-----DFIDADFTESK	629

图3 不同亚细胞定位类型的HSP70蛋白编码序列比较

Fig.3 Amino acid sequence alignment of HSP70 from different organelle and species

细胞质型: BOV.C: 牛 *Bos taurus* P19120; SSA1.C: 酿酒酵母 *Saccharomyces cerevisiae* X12926。内质网型: HOMO.E: 人 *Homo sapiens* P11021; TOB.E: 烟草 *Nicotiana tabacum* X60057。线粒体型: SSC1.M: 酿酒酵母 *Saccharomyces cerevisiae* M27229; PEA.M: 豌豆 *Pisum sativum* X54739。质体型: POR.P: 脐形紫菜 *Porphyra umbilicalis* X62240; PAV.P: 定鞭金藻 *Pavlova lutherii* X59555

藻类在经受多种胁迫包括冷冻、干旱、重金属、渗透压和机械等胁迫的过程中, 都有活性氧反应(ROS)的参与(Dring, 2005)。藻类的钒依赖的溴过氧化物酶(vBPO)可以将溴化物催化氧化成次溴酸, 然后形成挥发性卤烃。这些挥发性的卤烃, 可能发挥激素或微生物防御的作用(Jordan *et al*, 1991)。对石莼的研究发现, vBPO 的主要作用是清除活性氧反应产生的过氧化氢, 避免其对生物体产生的氧化伤害, 在生物活性氧反应中发挥重要作用(Manley *et al*, 2001)。Kamenarska 等(2007)对卡帕藻部分纯化的 vBPO 活性进行了研究, 发现与其它大型海藻中分离得到的这种酶的作用底物谱完全不同。通过 EST 研究, 作者获得了卡帕藻 vBPO 3'端的部分序列(全长 639 个氨基酸中的 227 个), 拟通过 RACE 获得全长序列后, 与其它藻类的同类基因进行比较, 将可以探讨这种活性差异的分子机制。

密码子的使用偏倚性产生的生物学基础目前还不清楚, 近几年来研究者通过对模式生物的基因组

碱基组分和密码子用法特征所做的研究表明, 影响密码子使用偏倚性的可能因素包括 tRNA 丰度、基因表达特性、物种、基因组碱基组分及编码蛋白的结构和功能等, 而基因组环境的碱基组分、G+C 含量在其同义密码子使用偏倚性的产生方面占有较大的优势(Duret *et al*, 1999; Sajau *et al*, 2001)。作者以尽可能多的编码序列为对象, 计算了长心卡帕藻的密码子使用频率。与其它藻类相比, 卡帕藻基因编码序列的平均 G+C 含量和第三位密码子 G+C 含量均较低(表 4), 通常密码子偏倚性越强 G+C 含量越高。

与其它红藻的 8 个必须氨基酸相比, 编码卡帕藻氨基酸的第三位密码子偏好 U 或 A(表 5), 这也是卡帕藻 G+C 含量低的原因。对密码子偏倚性的研究将有助于基因克隆时兼并引物的设计。如果以基因全长序列为对象, 分别研究高表达量基因和低表达量基因的密码子偏倚性, 可能得到的结果更为准确。上述工作为进一步研究长心卡帕藻基因及其功能奠定了基础。

表 4 不同物种的 G+C 含量
Tab.4 G+C content of different algae

项目	拟南芥*	石莼*	衣藻*	条斑紫菜**	坛紫菜**	卡帕藻
第三位密码子 G+C 含量(%)	42.37	57.69	85.94	79.4	73.1	51.3
平均 G+C 含量(%)	44.60	66.05	56.30	65.2	58.9	47.3

* Stanley *et al*, 2005; ** 庞国兴等, 2005

表 5 几种红藻的 8 个氨基酸的偏倚密码子
Tab.5 Synonymous codon usage bias for codons of eight amino acids from several red algal species

红藻种类	Ala	Arg	Gly	Leu	Pro	Ser	Thr	Val
刁毛藻*	GCC	CGC	GGC	CUC	CCC	UCC	ACC	GUC
角叉菜*	GCC	CGC	GGC	CUC	CCC	UCC	ACC	GUC
紫菜*	GCG	CGC	GGC	CUG	CCG	UCG	ACG	GUG
卡帕藻	GCA	CGU	GGU, GGA	UUG	CCG	AGU	ACU	GUG

* Lee *et al*, 2007

参 考 文 献

- 庞国兴, 王广策, 胡松年等, 2005. 坛紫菜(*Porphyra haitanensis*)丝状孢子体 EST 的获取及其生物信息学分析. 海洋与湖沼, 36(5): 452—458
- 秦玉明, 苏秀榕, 李 晔等, 2010. 缢蛏(*Sinonovacula constricta*)cDNA 文库的构建及肌动蛋白基因的研究. 海洋与湖沼, 41(1): 54—60
- Ashburner M, Ball C A, Blake J A *et al*, 2000. Gene Ontology: tool for the unification of biology. Nature Genet, 25(1): 25—29
- Ask E-I, Azanza R-V, 2002. Advances in cultivation technology of commercial eucheumatoid species: a review with suggestions for future research. Aquaculture, 206: 257—277
- Aspilla P S, Antonio A A C B, Zuccarello G C *et al*, 2010. A partial expressed sequence tag (EST) library of the economically important red alga *Eucheuma denticulatum* (N. L. Burham) F. C. Collins and Hervey. Philip Sci Lett, 3: 109—120
- Bixler H J, 1996. Recent developments in manufacturing and marketing carrageenan. Hydrobiologia, 326/327: 35—37
- Boorstein W R, Ziegelhoffer T, Craig E A, 1994. Molecular evolution of the HSP70 multigene family. J Mol Evol, 38: 1—17
- Cho E K, Hong C B, 2006. Over-expression of tobacco NtHSP70-1 contributes to drought-stress tolerance in plants. Plant Cell Rep, 25: 349—358

- Dring M J, 2005. Stress resistance and disease resistance in seaweeds: the role of reactive oxygen metabolism. *Advances in Botanical Research*, 43: 176—201
- Duret L, Mouchiroud D, 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila* and *Arabidopsis*. *Proc Natl Acad Sci USA*, 96(8): 4482—4487
- Guy C L, Li Q B, 1998. The organization and evolution of the spinach stress 70 molecular chaperone gene family. *Plant Cell*, 10: 539—556
- Ireland H E, Harding S J, Bonwick G A *et al*, 2004. Evaluation of heat shock protein 70 as a biomarker of environmental stress in *Fucus serratus* and *Lemnaminor*. *Biomarkers*, 9: 139—155
- Jordan P, Vilter H, 1991. Extraction of proteins from materials rich in anionic mucilages: partition and fractionation of vandate-dependent bromoperoxidases from the brown algae *Laminaria digitata* and *Laminaria saccharina* in aqueous polymer two-phase systems. *Biochim Biophys Acta*, 1073: 98—106
- Kamenarska Z, Taniguchi T, Ohsawa N *et al*, 2007. A vanadium-dependent bromoperoxidase in the marine red alga *Kappaphycus alvarezii* (Doty) Doty displays clear substrate specificity. *Phytochemistry*, 68: 1358—1366
- Lee H S, Lee H K, An G H *et al*, 2007. Analysis of Expressed Sequence Tags from the Red Alga *Griffithsia okiensis*. *The Journal of Microbiology*, 45(6): 541—546
- Lin B L, Wang J S, Liu H C *et al*, 2001. Genomic analysis of the Hsp70 superfamily in *Arabidopsis thaliana*. *Cell Stress Chaperones*, 6: 201—208
- Manley S L, Barbero P E, 2001. Physiological constraints on bromoform (CHBr₃) production by *Ulva lactuca* (Chlorophyta). *Limnology and Oceanography*, 46: 1392—1399
- Renner T, Waters E R, 2007. Comparative genomic analysis of the Hsp70s from five diverse photosynthetic eukaryotes. *Cell Stress Chaperone*, 12(2): 172—185
- Sajau H, Washio T, Saito R *et al*, 2001. Correlation between sequence conservation of the 5 untranslated region and codon usage bias in *Mus musculus* genes. *Gene*, 276(2): 101—105
- Stanley M S, Perry R M, Callow J A, 2005. Analysis of expressed sequence tags from the green alga *Ulva linza* (Chlorophyta). *J Phycol*, 41: 1219—1226
- Sung D Y, Vierling E, Guy C L, 2001. Comprehensive expression profile analysis of the *Arabidopsis* Hsp70 gene family. *Plant Physiol*, 126: 789—800
- Vierling E, 1991. Roles of heat shock proteins in plants. *Ann Rev Plant Physiol Plant Mol Biol*, 42: 579—620
- Wu G T, Liu C L, Liu S H, 2010. High-quality RNA preparation for cDNA library construction of the Antarctic sea-ice alga *Chlamydomonas* sp. ICE-L. *Journal of Applied Phycology*, online 22(6): 779—783

ANALYSIS OF *KAPPAPHYCUS ALVAREZII* EXPRESSED SEQUENCE TAGS

LIU Chen-Lin¹, HUANG Xiao-Hang¹, LIU Jian-Guo²

(1. *Research Center for Marine Ecology, the First Institute of Oceanography, State Oceanic Administration, Qingdao, 266061;*
2. *Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071*)

Abstract There were 281 sequences found significantly similar to the known proteins using the protein database BlastX aligning of 453 *Kappaphycus alvarezii* TUGs. GO annotation was conducted on 453 TUGs using Blast2go, 132 sequences were assigned to have molecular function, biological process and/or cellular component sub-ontology. According to the results of GO annotation, 24 genes were found associated with stress adaptation, including genes respond to stimulus and genes involve in antioxidant activity. Codon usage of *K. alvarezii* was also analyzed based on the 23251 codons from these 132 known genes. The codon usage bias was found quite different to the other red algae. The average GC content in codons and the GC usage in the third codon position were lower than the other algae. These data have laid a foundation for further study on gene functions of *K. alvarezii*.

Key words *Kappaphycus alvarezii*, Expressed sequence tag analysis, Stress resistance correlated gene, Codon usage