

海洋蓝藻基因组学研究

Studies on marine cyanobacteria genomics

赵方庆,梁成伟,秦 松

(中国科学院 海洋研究所,山东 青岛 266071)

中图分类号:Q755 文献标识码:A 文章编号:1000-3096(2007)06-0079-08

1994年,日本 Kazusa 实验室首先启动了淡水蓝藻 *Synechocystis* sp. PCC6803 的基因组计划,并于1996年完成基因组测序的工作。*Synechocystis* sp. PCC6803 基因组大小为 3.57 Mb,共有 3 168 个基因,是目前研究最多也最为透彻的蓝藻基因组。除此之外,美国能源部下属的联合基因组研究所 (Joint Genome Institute),以及麻省理工学院 (Massachusetts Institute of Technology)、法国巴斯德研究所 (Institute Pasteur)、北京人类基因组中心 (Human Genome Centre) 也相继启动了淡水蓝藻基因组计划,如 *Anabaena variabilis* ATCC29413 (8.00 Mb), *Anacystis* (*Synechococcus*) *nidulans* PCC6301 (2.69 Mb), *A-*

throspira (*Spirulina*) *platensis* (5.40 Mb), *Gloeobacter violaceus* PCC7421 (4.66 Mb), *Microcystis aeruginosa* PCC7806 (4.80 Mb), *Nostoc* (*Anabaena*) sp. PCC7120 (6.41 Mb), *Nostoc punctiforme* ATCC29133 (9.20 Mb), (*Thermo*) *Synechococcus elongatus* BP-1 (2.59 Mb), *Synechocystis* sp. PCC6803 (3.57 Mb)。海洋蓝藻基因组计划启动相对较晚,目前完成的或正在进行中的海洋蓝藻基因组计划的种类见表1。蓝藻基因组计划的实施,使得已知序列数据空前丰富,目前在 Genbank 上登录的关于蓝藻的基因序列数据涵盖了诸如细胞分化、光合作用以及初级代谢和次级代谢的相关基因。

表1 海洋蓝藻基因组计划实施情况

种类	大小(Mb)	ORFs	GC 含量	进度	研究单位
聚球藻 PCC7002	3.33	n/ a	49.1	即将完成	PSU and PkU
聚球藻 WH8102	2.43	2 527	59.5	完成	DoE/ J GI
原绿球藻 NA TL2A	2.00	n/ a	32.0	进行中	DoE/ J G
原绿球藻 MIT9313	2.40	2 195	50.7	完成	DoE/ J GI
原绿球藻 MED4	1.67	1 694	30.8	完成	DoE/ J GI
原绿球藻 SS120	1.75	1 734	36.0	注释	Genoscope, France

注:PSU and PkU: 宾西法尼亚州立大学和北京大学;J GI/ DoE: 联合基因研究所 (Joint Genome Institute)/ 美国能源部 (Department of Energy); Genoscope: 法国 Genoscope 中心

1 蓝藻结构基因组学的主要研究方法

海洋原核微藻全基因组测序主要采用鸟枪法^[1]。首先,对目的藻株进行基因组文库的构建。文库所含的克隆数要达到一定数量,以保证经末端测序的克隆片段的碱基总数大于基因组碱基总数的5倍以上。其次,对基因组大片段进行测序。最后,将所得到的序列进行拼接,通过计算机软件将序列组合成数个连锁群(contigs),然后对连锁群进行排序和缺口填补^[2]。例如,对海洋聚球藻 (*Synechococcus* sp.

WH 8102) 进行鸟枪法测序^[3]。先提取基因组 DNA, 剪切得到大小 2 ~ 3 kb 的片断, 将之克隆到 pUC18 载体上进行测序。然后将测序产物进行拼接, 通过引物步移法补齐缺口。最终通过 PCR 验证拼接区。

收稿日期:2004-08-10;修回日期:2005-03-29

基金项目:中国科学院知识创新工程项目 (KZCX3-SW-215)

作者简介:赵方庆(1980-),男,博士研究生,研究方向为藻类基因组学和分子进化,E-mail:biofqzhao@yahoo.com.cn

完成基因组全部序列测定,得到全基因组序列图谱后,工作的重点就是对基因组全序列进行注释(annotation),分析全基因组所包含的遗传信息:首先是 GC(%)的分析;其次利用专门软件如 Glimmer^[4], Critica^[5]和 Generation 寻找开放读码框架(Open reading frame, ORF)。最后将测得的序列与已知数据库(如 BLASTP, KEGG GENES, Pfam, PROSITE, PRINTS, ProDom 和 COGs 等)中的序列进行比对。

2 蓝藻功能基因组学的研究方法

为进行功能基因组研究,人们相继建立了一系列新的技术和方法,包括功能基因表达克隆(Functional cloning)、基因芯片(Gene-chip)或基因微阵列技术(Microarray)、基因表达系统分析(Serial analysis of gene expression, SAGE)或表达序列标签(Expressed sequence tag, EST)、蛋白质组技术(Proteomics)、质谱测序技术、生物信息学(Bioinformatics)、转基因技术(Transgenics)、基因敲除(Gene knock-out)等等,这些前沿技术为功能基因组学的研究提供了强有力的技术支持。在蓝藻功能基因组研究方面,目前应用较多的是基因微阵列技术和蛋白质组技术等。

2.1 基因微阵列技术

基因微阵列技术是一项新近发展起来的能够有效地研究基因转录和表达的技术手段^[6,7]。将寡核苷酸或 DNA 密集排列于硅片等固相支持物上,将要研究的样品标记后与其杂交,根据杂交位点及信号强弱来研究基因的转录和表达。到现在为止,蓝藻中该项研究主要集中在淡水的 *Synechocystis* sp. PCC6803,因为它是最早(1996)完成基因组测序的蓝藻。目前,第一张商业蓝藻基因芯片已经在日本宝生物公司(Takara Bio Inc., Japan)问世,叫做 IntelliGene™ CyanoCHIP,包括大约 2 800 个开放读码框,主要是用来进行光合作用的研究。通过对 *Synechocystis* sp. PCC6803 的 CyanoCHIP 等芯片的研究,揭示了胁迫条件下相关基因的表达情况,如高光强适应性^[8];氮、磷、硫的缺失;对 UV-B 的辐射效应^[9];加入电子传递抑制剂后的氧化还原反应^[10];盐胁迫^[11,12];碳源缺乏^[13];渗透压力^[12,14]以及其他环境胁迫因子对光合作用过程的影响。

2.2 蛋白质组技术

蛋白质组分析法首先利用双向电泳技术分离蛋白质组分,然后利用计算机软件对所得图像进行数据采集和分析,从胶上回收蛋白质并采用氨基酸

组成分析、微量蛋白质序列分析、质谱分析等技术进行精细鉴定,从而获得蛋白质组分的物理、化学及生物学参数,如分子量、等电点、表达量等。将获得的数据与已知蛋白质数据库中的数据进行比较,就可以知道某序列编码哪种蛋白质,从而建立起一个生物特异的全蛋白质数据库。在蓝藻当中,蛋白质组学研究也主要集中在淡水的集胞藻 PCC6803,地木耳 *Nostoc commune* DRH1,点形念珠藻 *N. punctiforme* PCC73102 和微囊藻 *Microcystis aeruginosa* PCC7806(表 2)。

2.3 代谢组学技术

代谢组学通过考察生物体受刺激或扰动(如某个特定的基因变异或环境发生变化)后代代谢物的变化,来研究生物体系的代谢途径。与基因芯片技术和蛋白质组学技术相比,代谢组学的研究对象主要集中在分子质量 1 000 以下的小分子。基因和蛋白表达的微小变化会在代谢产物上得到放大,从而使得检测更容易。

相对于链霉菌等微生物而言,代谢组学在蓝藻中刚刚起步。Christiansen 等^[26]利用简并 PCR 引物对 146 种蓝藻的非核糖体多肽合成酶基因(Non-ribosomal peptide synthetase)进行扩增,结果显示有 75.3% 的蓝藻具有相应的基因。在单细胞蓝藻中,只有管胞藻(*Chamaesiphon* sp. PCC6605)、两种粘球藻(*Gloeocapsa*),以及大多数的微囊藻具有该酶系,而隐杆藻(*Cyanotheca*)、粘杆藻(*Gloeotheca*)、聚球藻(*Synechococcus*)和集胞藻中均未检测到。相比之下,大多数的丝状蓝藻(*Prochloron*)如厚皮藻、颤藻目(*Oscillatoriales*)蓝藻以及具异型胞的丝状体种类具有该酶系。该研究结果为蓝藻代谢组学研究提供了一定线索。

3 海洋蓝藻基因组研究进展

3.1 原绿球藻基因组研究

海洋单细胞蓝藻——原绿球藻直径约为 0.5~0.7 μm,是目前已知最小的自养光合生物,它广泛分布于热带和亚热带海域。原绿球藻含有叶绿素 a 和 b,还有玉米黄素、α-胡萝卜素以及少量的叶绿素 c 类似物,它的捕光色素系统在功能上与绿藻和高等植物相似,但是在结构上却有着明显的不同^[27,28]。根据其分布及生理特征差异,原绿球藻可以分为高光适应型(high-light adapted)和低光适应型(low-light adapted)两种生态类型,前者以 *Prochlorococcus marinus* MED4 为代表,主要分布在大洋的表层海水中;后者以 *Prochlorococcus marinus* SS120, *Prochlorococcus marinus* MIT9313 为代表,主要分布在 100~

200 m 的水体中^[29,30]。原绿球藻和海洋聚球藻一起参与了 2/3 的海洋生态系统中的碳固定和地球上 1/3 的初级生产力^[31],此外它也是研究进化和海洋初级生产力的独一无二的材料,成为藻类研究的热点之一。

表 2 蓝藻蛋白质组学研究概况

藻株	分析方法	研究对象或条件	蛋白质	参考文献
微囊藻 PCC7806 (野生株和突变株)	MALDI- TOF MS, 2DE	微囊藻毒素合成调控	MrpA	[15]
地木耳 DRH1	ESI-MS/ MS, 2DE	UV-B 刺激	有 493 种蛋白表达水平发生变化;诱导产生 27 种蛋白	[16,17]
点形念珠藻 PCC73102	MALDI-TOF MS, 2DE	细胞分化 固氮作用	Ser/ Thr 激酶, cdc2 调控基因 10 种蛋白上调(包含 1 种 PKS);约 1/3 的蛋白下调	[18] [19]
集胞藻 PCC6803	2D PAGE	正常条件	类囊体膜上的外周蛋白	[20]
	MALDI-TOF MS	低盐浓度	8 种蛋白与光合作用有关;4	[21,22]
	2D PAGE	高盐浓度 光诱导	种蛋白参与细胞过程;2 种蛋白上调(up-regulation), 但功能未知	
	MALDI-TOF MS, 2DE	高盐浓度	20 种蛋白诱导产生或积累	[22]
	MALDI-TOF MS, 1D SDS-PAGE 和 2DE RT-PCR,2DE	质膜蛋白 正常生长条件 不同 pH 不同培养条件下主要代谢途径的酶系调控	在 170 个斑点中确定了 67 种蛋白 对 17 种关键酶进行了定量分析	[23] [24]
集胞藻 PCC6803 HT-3	MALDI-TOF MS, N-端测序	纯化的光合系统	PS 相关的 31 种蛋白;发现 5 种新蛋白	[25]

2DE:双向电泳 (two-dimensional electrophoresis); MALDI-TOF MS:matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; PAGE: 聚丙烯酰胺凝胶电泳 (polyacrylamide gel electrophoresis); RT-PCR:逆转录 PCR (reverse transcriptase-polymerase chain reaction)

目前,完成序列测定工作的原绿球藻共有 3 种: *Prochlorococcus marinus* SS120^[32], *Prochlorococcus marinus* MIT 9313^[33], *Prochlorococcus marinus* MED4^[33]。和淡水蓝藻 (*Anabaena* sp. PCC7120, 6.4 Mb; *Synechocystis* sp. PCC6803, 3.57 Mb) 相比,原绿球藻基因组较小,约为 1.6~2.4 Mb,它的大小与古细菌 (*Methanococcus jannaschii* 1.66 Mb; *Methanopyrus kandleri* AV19, 1.69 Mb) 接近。其中,高光适应型的 *Prochlorococcus marinus* MED4 是目前已知基因组最小的光能自养生物。而测序完成的 3 种原绿球藻的 GC 含量差异很大 (30.8%~50.7%),但是和海洋聚球藻 (47.4%~69.5%) 相比,大多数原绿球藻的 GC 含量都是较低的^[29,34]。原绿球藻具体的基因组特征及比较见表 3。

表3 原绿球藻菌株基因组特征的比较

基因组特征	藻株		
	MED4	MIT9313	SS120
长度 (bp)	1 657 990	2 410 873	1 751 080
GC (%)	30.8	50.7	36.4
ORFs (%)	88	82	88.5
ORFs	1 716	2 275	1 884
tRNA	37	43	40
rRNA 操纵子	1	2	3

3.2 海洋聚球藻基因组研究

聚球藻 (*Synechococcus* sp.) 是一临时的分类单位, 包括球状和杆状蓝藻, 它主要限定于包含叶绿素 a1 和藻胆体结构的蓝藻。根据形态和生理特征、采集地差异、G+C 含量、藻红蛋白的有无, 聚球藻经常被分为以下几类: *Synechococcus*-cluster (PCC6301, PCC7942); *Cyanobium*-cluster (PCC6307); marine A-cluster (WH7805, WH8008, WH8102, WH8103, WH8020, WH8002); marine B-cluster (WH8101, WH5701)。但是, 在 marine A-cluster 中 *Synechococcus* 各品系之间在遗传和生理特征上都是多样的, PUB 和 PEB 的相对含量也有显著性的差异, 海洋聚球藻各品系之间的系统发育关系迄今都没有很好的解决。

目前正在进行海洋聚球藻基因组计划的有 *Synechococcus* PCC7002 和 *Synechococcus* WH8102。其中, *Synechococcus* WH8102 的基因组信息已在 NCBI 上公布出来。它的基因组大小为 2.43 Mb, GC 含量为 59.5%, 在其基因组上共找到 2 527 个 ORFs, 其中有 736 个 ORFs 是在原绿球藻基因组中没有发现的。这些基因大多与聚球藻的运动性 (motility)、糖转移酶 (Glycosyltransferase) 以及噬蓝藻体 (Cyanophage) 相关, 此外, 它还编码了藻胆体基因簇 (Phycobilisomes) 和硝酸盐还原酶 (Nitrate reductase) 相关基因, 这些基因都与它区别于原绿球藻的独特生态适应性有关^[31]。目前即将测序完成的海洋蓝藻 *Synechococcus* PCC7002 以前被称作 *Agmenium quadruplicatum* PR6 (ATCC27264), 它是一种单细胞不能固氮的海洋蓝藻, 基因组大小为 2.63 Mb^[35], GC 百分含量是 49%。它的基因组中包含着染色质外因子, 大小从 3 kb 到 150 kb 不等^[35, 36]。

3.3 比较基因组学

3.3.1 基因组的一般特征比较

海洋蓝藻的基因组一般较淡水蓝藻的基因组 (3.5 ~ 15 Mb) 要小, 第一个完成基因组测序的 *Synechocystis* PCC6803 的基因组大小为 3.6 Mb^[37], 含有叶绿素 b 的 *Prochlorothrix hollandica* 的基因组则有 5.5 Mb^[38], 而土壤蓝藻 *Nostoc punctiforme* PCC73102 则超过了 9 Mb^[39]。以 *P. marinus* SS120 为例, 很多在淡水蓝藻中广泛存在的基因在原绿球藻中基因的拷贝数都减少甚至彻底缺失了^[32] (表 4)。尤其值得注意的是在 *P. marinus* SS120 中, 编码信号传导系统和应激反应系统的关键基因大都缺失了, 只剩下 5 个组氨酸激酶和 6 个反应调控因子。此外, 在海洋聚球藻 WH8102 中没有发现 *cpcC* 和 *cpcD* 基因, 这些基因在淡水蓝藻中编码藻蓝蛋白的 L_R 结合蛋白, 它是藻蓝蛋白组装所必需的。但是它在海洋聚球藻 WH8102 中的缺失说明海洋聚球藻 WH8102 中可能有一个藻蓝蛋白圆盘结构 (Disc)。海洋蓝藻基因组中藻蓝蛋白结构的简单化和藻红蛋白结构的复杂化, 与其适应海洋独特的生态环境有着密切的关系。在海洋环境中, 尤其是大洋海域, 由于营养盐和光照的限制, 海洋蓝藻有着适应其所处环境的独特的基因组结构。

除了基因组大小不同之外, 海洋和淡水蓝藻基因的排列顺序也有着显著的不同。例如, 在 *Synechococcus* PCC7002 中缺少 PS_{II}、藻胆体或者细胞色素 b₆-f 复合物的基因簇, 并且光合作用的相关基因大多是分散在基因组中, 相关的基因顺序为 *petCA psbB/DpsaB petBD*, 而在 *Anabaena* PCC7120, 该基因簇的顺序则为 *petCA recA petB*。此外, 由于噬蓝藻体的影响, 在海洋蓝藻 (*Synechococcus*) 中存在着较多的水平基因转移现象。

3.3.2 捕光系统 (Light-harvesting system)

在大多数的蓝藻中, 藻胆体 (phycobilisomes) 是主要的捕光色素系统, 而在原绿球藻 (*Prochlorococcus* sp.) 以及其他两种蓝藻如 *Prochloron* sp. 和 *Prochlorothrix hollandica* 中, 由于藻胆体结构的缺失, 叶绿素 a/b 复合物是主要的捕光系统。在海洋聚球藻中, 有大约 40 个基因位于几个操纵子和基因簇中, 编码藻胆体结构的合成、组装和降解的相关蛋白。目前报道的最大的藻胆体相关基因簇是在 *Synechococcus* WH8020 中^[40]。在长度约为 15 kb 的基因簇中, 有 18 个基因参与了藻蓝蛋白合成组装, 还有几个多肽与色素的合成及组装相关, 以及有两个不同的藻红蛋白编码基因 (*cpeAB*, *mpeAB*)。而在原绿球藻中

表 4 某些功能基因在海洋原核微藻和淡水蓝藻基因组中拷贝数的差异比较

功能	基因	COG 序号	拷贝数			
			Pma	Tel	Syn	Ana
信号转导机制						
组氨酸激酶	-	0642	5	16	42	124
反应调节蛋白	-	2197	6	26	41	84
PAS 域	-	2202	1	13	23	57
GAF 域	-	2203	0	15	28	62
HPt 域	-	2198	0	4	7	9
GGDEF 域	-	2199	0	10	23	14
EAL 域	-	2200	0	5	13	7
HD- GYP 域	-	2206	0	1	2	2
Ser/ Thr 蛋白激酶	-	0515	0	10	9	35
Ser/ Thr 磷酸酶 1	-	0631	0	3	3	4
Ser/ Thr 磷酸酶 1	-	0639	0	1	1	2
cAMP-结合蛋白	<i>crp</i>	2114	0	1	2	7
腺苷酸环化酶	<i>cyaA</i>	0664	2	3	12	14
趋光性						
Hybrid 组氨酸激酶	<i>taxA Y</i>	0643/ 2198/ 0784	0	3	3	3
趋光受体蛋白	<i>taxD</i>	2203/ 0840	0	3	3	3
趋光蛋白	<i>tax P</i>	0784	0	3	3	3
CheW- 类似信号转导蛋白	<i>tax W</i>	0835	0	3	3	3
Che Y- 类似信号转导蛋白	<i>tax Y</i>	0784	0	3	3	3
RNA 聚合酶调控						
σ28	<i>fliA</i>	1191	0	1	1	2
σ24	<i>rpoE</i>	1592	0	3	3	2
反义 Sigma 因子	<i>rsbW</i>	2172	1	2	2	5
丝氨酸磷酸酶调节蛋白	<i>rsbU</i>	2208	1	2	5	5
其他						
藻胆体降解蛋白 A	<i>nblA</i>	-	0	1	2	2
藻胆体降解蛋白 B	<i>nblB</i>	-	0	2	1	1
普遍胁迫蛋白	<i>uspA</i>	0589	0	4	4	7
盐胁迫蛋白	CDS	P34	-	0	1	2
锚蛋白重复序列	-	0666	0	1	1	2
Forkhead 结构域	-	1716	0	2	2	2

注: Pma: 原绿球藻 SS120; Tel: 细长聚球藻; Syn: 集胞藻 PCC6803; Ana: 鱼腥藻 PCC7120

大部分藻胆体相关基因都已经缺失了,只剩下编码藻红蛋白的相关基因 *peBA*。令人疑惑的是在原绿球藻中还有 *ho1* 的类似基因^[41]。藻胆蛋白的发色团是由亚铁血红素经过一系列的生物合成途径进行合成的。而其中第一步则是由开环的四吡咯环(正铁血红素, protoheme)经由 *ho1*(亚铁血红素加氧酶)催化裂解成胆绿素(biliverdin IX α)的^[42, 43]。在 *Synechococcus* 中, *ho1* 与 *pebAB* 都是藻胆蛋白基因簇的一部分,而在 *Prochlorococcus* sp. MED4 和 SS120 中,它们则位于 200 kb 外的区域。

通过对原绿球藻和海洋聚球藻藻红蛋白基因编码区域的比较,发现在 *Prochlorococcus* SS120 中, *mpeX*(S-腺苷甲硫氨酸转移酶)和 *uvrD*(DNA 螺旋酶)分别位于 PE 基因簇的两端,在高光适应型的 MED4 中也有同样的现象,只不过后者的 PE 基因簇只有 4 493 bp,而 SS120 则有 11 500 bp。在 9 个编码藻红蛋白合成的相关基因中, MED4 只保留了 2 个: *orf181* 和 *cpeB*,在 *orf181* 前的 175nt 的区域可能是它的启动子编码区。通过对不同原绿球藻及其他蓝藻藻红蛋白的序列比较发现 MED4 的 *pe* 基因在多个位点都发生了突变,丢失了两个高度保守的色基结合位点(Cys)。MED4 中藻红蛋白基因的这些显著变化以及 *peA* 的缺失说明藻红蛋白在原绿球藻中处于丢失的过程当中^[44]。有荧光光谱分析的证据显示原绿球藻中藻红蛋白可以将激发能传递至叶绿素^[45],但是这种效率是非常低的,大概只占原绿球藻光合效率的 1.8 %^[46]。目前,藻红蛋白在原绿球藻中的功能目前并不清楚,还有待于进一步的实验验证。

3.3.3 二元信号转导系统(Two-component signal transduction systems)

对光照和营养盐限制的响应是光合生物适应环境变化所必需的,尤其是在缺乏营养盐的大洋环境中。当外界环境因素发生变化时,有很多调控机制参与了基因的表达及相关产物的调控作用,二元信号转导系统就是主要的调控机制之一。它包括组氨酸蛋白激酶(Histidine kinase, Hik)和反应调节蛋白(Response regulator, Rer),前者能够感受外界环境变化并将该变化传递至后者,而后者一旦被激活就能够调控特定基因的表达从而引发对外界刺激的响应。二元信号转导系统在绝大多数细菌中都存在,它们一般占编码蛋白质的 1 %。大肠杆菌基因组中含有 62 个编码二元系统的基因,参与调节不同的生理过程,如趋化性、渗透调节、代谢和运输。二元系统在致病的革兰氏阳性菌和革兰氏阴性菌中都存在,除了调节一些基本的功能外,它们还控制毒素和一些对致病性很重要的蛋白质的表达。通过对不同种类细菌的二元信号转导系统的比较,发现它们不仅在功能上相似,而且还具有高度保守的氨基酸位点^[47]。

海洋蓝藻是最早进行放氧光合作用的生物,它在漫长的进化过程中有着极强的适应性,例如低氧、

紫外辐射、高温、对低光强的适应和高效的营养吸收等。近年来,随着越来越多的海洋蓝藻基因组序列的测定,使得研究它们的信号转导模式,揭示海洋蓝藻独特生态适应性成为可能。原绿球藻是目前已知最小的、在海洋中丰度最高的光合生物,分布于南北纬度 40 ° 之间的水域,并且在深度 100 ~ 200 m 的水体中都有分布。在高光适应型的 MED4 中,有 11 个基因参与了二元信号转导系统,而在淡水的 *Synechocystis* PCC6803 中则有 80 个相关基因。*Prochlorococcus* MED4 具有高度紧缩的基因组(1.66 Mb),但这并不是造成二元信号转导系统基因数目减少的唯一的原因。事实上,通过对淡水蓝藻的比较发现,这些基因的数目与基因组的大小并无直接关系。此外,目前测序完成的海洋蓝藻大都像 MED4 一样,含有少量数目的 *rer* 和 *hik* 基因。相对于临时性的池塘和土壤环境来说,海洋蓝藻所处的环境相对来说还是稳定的,它们可能不需要太多的基因来参与信号转导的调控。海洋蓝藻中, Hik 和 Rer 的种类多样性的降低也支持了这一观点^[48]。

在 *Escherichia coli* 中, *hik/ rer* 基因一般是位于同一个转录单位或操纵子中,并且在同一个信号应答中共同发挥作用。但是在淡水的 *Synechocystis* PCC6803 和海洋蓝藻中,仅有一小部分 *hik/ rer* 基因是连锁在一起的,如 *Prochlorococcus* MED4 中的 *hik03/ rer06*。通过对其序列进行分析,发现它与 *Bacillus subtilis* 的 *PhoR/B* 区域具有相当大的同源性。*PhoR/B* 调控着一些重要功能基因的表达,这些基因都参与了磷酸盐限制条件下细胞的继续生存和生长过程^[49]。在 *Prochlorococcus* MED4 中 *hik03* 基因邻近还有一个与 *B. subtilis pstS* 基因高度相似的读码框,它仅仅在磷酸盐不足的情况下表达,有可能是原绿球藻对寡营养条件的一种适应^[48]。

3.3.4 氮源利用相关基因

通过对氮源利用相关基因簇的结构分析^[32],发现 *Prochlorococcus* 在进化过程中逐渐丧失了利用硝酸盐和亚硝酸盐的能力,而 *Synechococcus* WH8102 的氮源利用相关基因都在同一基因簇中。MIT9313 丢失了编码硝酸盐还原酶和硝酸盐、亚硝酸盐转运的相关蛋白等 15 个基因,而保留了亚硝酸盐还原酶基因,但是在其侧翼有一个类似于原细菌的亚硝酸盐转运子,而不是典型蓝藻类型的亚硝酸盐透性酶基因,这说明 MIT9313 可能是通过水平基因转移获得该功能的。在 MED4 中,氮源利用的基因簇进一步缩减,甚至丢失了亚硝酸盐还原酶基因,大小只有约 10 kb。因此, MIT9313 不能利用硝酸盐作为氮源, MED4 既不能利用硝酸盐也不能利用亚硝酸盐作为氮源。而 *Synechococcus* 在硝酸盐大量增加的情况下可以形成赤潮,多发生于春季北大西洋和红海海域。

Prochlorococcus 利用有机氮的能力也要比 *Synechococcus* 弱, MED4 具有利用尿素、氰酸盐和小分子

多肽的相关基因,但是不能利用单个氨基酸,而MIT9313虽可以利用氨基酸但是缺乏氨基酸的转运和裂解的相关基因。此外,二者都具有高亲和性的铵盐转运子 *amt1*,但是缺乏固氮酶基因。

参考文献:

- [1] Fleischmann R D, Adams M D, Smith H O, *et al.* Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd [J]. **Science**, 1995, 269: 496-512.
- [2] Venter J C, Smith H O, Hood L. A new strategy for genome sequencing [J]. **Nature**, 1996, 381:364-366.
- [3] Palenik B, Brahamsha B, Larimer F W, *et al.* The genome of a motile marine *Synechococcus* [J]. **Nature**, 2003, 424(28):1 037-1 042.
- [4] Delcher A L, Harmon D, Kasif S, *et al.* Improved microbial gene identification with GLIMMER [J]. **Nucleic Acids Res**, 1999, 27:4 636-4 641.
- [5] Badger J H, Olsen G J. CRITICA: Coding region identification tool invoking comparative analysis [J]. **Mol Biol Evol**, 1999, 16:512-524.
- [6] Fodor S P A, Read J. Light-directed, spatially addressable parallel chemical synthesis [J]. **Science**, 1991, 251:767-773.
- [7] Schena M, Heller R A, Thieriault T P, *et al.* Microarrays: biotechnology's discovery platform for functional genomics [J]. **Trends Biotechnol**, 1998, 16: 301-306.
- [8] Hihara Y, Kamei A, Kanehisa M, *et al.* DNA microarray analysis of cyanobacterial gene expression during acclimation to high light [J]. **Plant Cell**, 2001, 13:793-806.
- [9] Huang L, McCluskey M P, Ni H, *et al.* Global gene expression profiles of the cyanobacterium *Synechococcus* sp. strain PCC6803 in response to irradiation with UV-B and white light [J]. **J Bacteriol**, 2002, 184: 6 845-6 858.
- [10] Hihara Y, Sonoike K, Kanehisa M, *et al.* DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechococcus* sp. strain PCC6803 [J]. **J Bacteriol**, 2003,185:1 719-1 725.
- [11] Allakhverdiev S, Hishiyama Y, Miyairi S, *et al.* Salt stress inhibits the repair of photodamaged photosystem by suppressing the transcription and translation of *psbA* genes in *Synechocystis* [J]. **Plant Physiol**, 2002, 130:1 443-1 453.
- [12] Kanesaki Y, Suzuki I, Allakhverdiev S I, *et al.* Salt stress and hyperosmotic stress regulate the expression of different sets of genes in *Synechocystis* sp. PCC6803 [J]. **Biochem Biophys Res Commun**, 2002, 290:339-348.
- [13] Heffelfinger G S, Martino A, Gorin A, *et al.* Carbon sequestration in *Synechococcus* sp.: from molecular machines to hierarchical modeling [J]. **OMICS**, 2002, 6:305-330.
- [14] Mikami K, Kanesaki Y, Suzuki I, *et al.* The histidine kinase Hik33 perceives osmotic stress and cold stress in *Synechocystis* sp. PCC6803 [J]. **Mol Microbiol**, 2002, 46:905-915.
- [15] Dittmann E, Erhard M, Kaebernick M, *et al.* Altered expression of two light-dependent genes in a microcystin-lacking mutant of *Microcystis aeruginosa* PCC7806 [J]. **Microbiol**, 2001, 147:3 113-3 119.
- [16] Krehenbrink M, Oppermann-Sanio F B, Steinbuechel A. Evaluation of noncyanobacterial genome sequences for occurrence of genes encoding proteins homologous to cyanophycin synthetase and cloning of an active cyanophycin synthetase from *Acinetobacter* sp. strain DSM 587 [J]. **Arch Microbiol**, 2002, 177: 371-380.
- [17] Ehling-Schulz M, Schulz S, Wait R, *et al.* The UV-B stimulon of the terrestrial cyanobacterium *Nostoc commune* comprises early shock proteins and late acclimation proteins [J]. **Mol Microbiol**, 2002, 46: 827-843.
- [18] Shibata M, Katoh H, Sonoda M, *et al.* Genes essential to sodium-dependent bicarbonate transport in cyanobacteria: function and phylogenetic analysis [J]. **J Biol Chem**, 2002, 277:18 658-18 664.
- [19] Liaimer A, Matveyev A, Bergman B. Isolation of host plant induced cDNAs from *Nostoc* sp. strain PCC9229 forming symbiosis with the angiosperm *Gunnera* spp. [J]. **Symbiosis**, 2001, 31:293-307.
- [20] Wang Y, Sun J, Chitnis P R. Proteomic study of the peripheral proteins from thylakoid membranes of the cyanobacterium *Synechocystis* sp. PCC6803 [J]. **Electrophoresis**, 2000, 21:1 746-1 754.
- [21] Sazuka T, Yamaguchi M, Ohara O. Cyano2Dbase updated: linkage of 234 protein spots to corresponding genes through N-terminal microsequencing [J]. **Electrophoresis**, 1999, 20:2 160-2 171.
- [22] Fulda S, Huang F, Nilsson F, *et al.* Proteomics of *Synechocystis* sp. strain PCC6803. Identification of periplasmic proteins in cells grown at low and high salt concentrations [J]. **Eur J Biochem**, 2000, 267: 5 900-5 907.
- [23] Huang F, Parmryd I, Nilsson F, *et al.* Proteomics of *Synechocystis* 6 803: identification of plasma membrane proteins [J]. **Mol Cell Proteomics**, 2002, 1(12):956-966.
- [24] Yang C. Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis* [J]. **Appl Microbiol Biotechnol**, 2002, 58:813-822.
- [25] Kashino Y, Lauber W M, Carroll J A, *et al.* Proteomic analysis of a highly active photosystem II preparation from the cyanobacterium *Synechocystis* sp. PCC6803 reveals the presence of novel *Polypeptides* [J]. **Biochem**, 2002, 41:8 004-8 012.
- [26] Christiansen G, Dittmann E, Ordorika L V, *et al.*

- Nonribosomal peptide synthetase genes occur in most cyanobacterial genera as evidenced by their distribution in axenic strains of the PCC [J]. **Arch Microbiol**, 2001, 176:452-458.
- [27] Georick R, Repeta D J. The pigments of *Prochlorococcus marinus*: the presence of divinyl chlorophyll a and b in a marine prochlorophyte [J]. **Limnol Oceanogr**, 1992, 37:425-433.
- [28] La Roche J, van der Staay G W M, Partensky F, et al. Independent evolution of the prochlorophyte and green plant chlorophyll a/b light-harvesting proteins [J]. **Proc Natl Acad Sci USA**, 1996, 93:15 244-15 248.
- [29] Partensky F, Hess W R, Vaulot D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance [J]. **Micro Mol Biol Rev**, 1999, 63(1): 106-127.
- [30] Ting C S, Rocap G, King J, et al. Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies [J]. **Trends in Microbiol**, 2002, 10(3):134-142.
- [31] Bryant D A. The beauty in small things revealed [J]. **Proc Natl Acad Sci USA**, 2003, 100(17): 9 647-9 649.
- [32] Dufresne A, Salanoubat M, Partensky F, et al. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome [J]. **Proc Natl Acad Sci USA**, 2003, 100(17): 10 020-10 025.
- [33] Rocap G, Larimer F W, Lamerdin J, et al. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation [J]. **Nature**, 2003, 424(28):1 042-1 047.
- [34] Watson G, Tabita F R. Regulation, unique gene organization, and unusual primary structure of carbon fixation genes from a marine phycoerythrin-containing cyanobacterium [J]. **Plant Mol Biol**, 1996, 32: 1 103-1 115.
- [35] Chen X, Widger W R. Physical genome map of the unicellular cyanobacterium *Synechococcus* sp. strain PCC7002 [J]. **J Bacteriol**, 1993, 175:5 106-5 116.
- [36] Akiyama H, Kanai S, Hirano M, et al. A novel plasmid recombination mechanism of the marine cyanobacterium *Synechococcus* sp. PCC7002 [J]. **DNA Res**, 1998, 5:327-334.
- [37] Kaneko T, Sato S, Kotani H, et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement) [J]. **DNA Res**, 1996, 3:185-209.
- [38] Schyns G, Rippka R, Namane A, et al. *Prochlorothrix hollandica* PCC9006: Genomic properties of an axenic representative of the chlorophyll a/b-containing oxyphotobacteria [J]. **Res Microbiol**, 1997, 148:345-354.
- [39] Meeks J C, Campbell E L, Summers M L, et al. Cellular differentiation in the cyanobacterium *Nostoc punctiforme* [J]. **Arch Microbiol**, 2002, 178(6): 395-403.
- [40] Wilbanks S M, Glazer A N. Rod structure of a phycoerythrin IF-containing phycobilisome. I. Organization and sequence of the gene cluster encoding the major phycobiliprotein rod components in the genome of marine *Synechococcus* sp. WH8020 [J]. **J Biol Chem**, 1993, 268:1 226-1 235.
- [41] Willows R D, Mayer S M, Foulk M S, et al. Phytybilin biosynthesis: The *Synechocystis* sp. PCC6803 heme oxygenase-encoding ho1 gene complements a phytochrome-deficient *Arabidopsis thaliana* hy1 mutant [J]. **Plant Mol Biol**, 2000, 13:113-120.
- [42] Richaud C, Zabulon G. The heme oxygenase gene (*pbsA*) in the red alga *Rhodella violacea* is discontinuous and transcriptionally activated during iron limitation [J]. **Proc Natl Acad Sci USA**, 1997, 94:11 736-11 741.
- [43] Cornejo J, Willows R D, Beale S I. Phytybilin biosynthesis: cloning and expression of a gene encoding soluble ferredoxindependent heme oxygenase from *Synechocystis* sp. PCC6803 [J]. **Plant J**, 1998, 15: 99-107.
- [44] Ting C S, Rocap G, King J, et al. Phycobiliprotein genes of the marine photosynthetic prokaryote *Prochlorococcus*: evidence for rapid evolution of genetic heterogeneity [J]. **Microbiol**, 2001, 147: 3 171-3 182.
- [45] Lokstein H, Steglich C, Hess W R. Light-harvesting antenna function of phycoerythrin in *Prochlorococcus marinus* [J]. **Biochim Biophys Acta**, 1999, 1 410:97-98.
- [46] Steglich C, Mullineaux C W, Teuchner K, et al. Photophysical properties of *Prochlorococcus marinus* SS120 divinyl chlorophylls and phycoerythrin in vitro and in vivo [J]. **FEBS Lett**, 2003, 553:79-84.
- [47] Stock J B, Robinson V L, Goudreau P N. Two-component signal transduction [J]. **Annu Rev Biochem**, 2000, 69:183-215.
- [48] Mary I, Vaulot D. Two-component systems in *Prochlorococcus* MED4: Genomic analysis and differential expression under stress [J]. **FEMS Microbiol Lett**, 2003, 226:135-144.
- [49] Hulett F M. The signal-transduction network for *Pho* regulation in *Bacillus subtilis* [J]. **Mol Microbiol**, 1996, 19:933-939.

(本文编辑:张培新)