

藻类基因组研究进展与 EST 技术的应用

Progress of algal genome research and application of EST technology

方永军¹, 范晓蕾², 王广策², 胡松年¹

(1. 浙江大学 沃森基因组科学研究院, 浙江 杭州 310008; 2. 中国科学院 海洋研究所, 山东 青岛 266071)

中图分类号: Q943.2

文献标识码: A

文章编号: 1000-3096(2007)12-0076-07

藻类在自然界中广泛存在, 目前已知大约有 3 万余种。主要生长在水中、潮湿的岩石上、墙壁、树干上及土壤表面。它们对环境条件要求不高, 适应能力强, 是新生活区的先锋植物之一。藻体千差万别, 小至几微米, 大至几十米, 结构简单, 无根、茎、叶分化, 一般都具有进行光合作用的色素, 属于自养植物, 是地球上已知的最古老的一个生命类群之一。有化石记录大约在 35~33 亿年前, 地球的水体中就出现了原核蓝藻, 15 亿年前出现类似于现代藻类的有机体^[1]。

藻类与人类的生活生产有着密切的关系, 可以作为营养食品, 也可以用作工业原料, 具有重要的经济价值, 另一方面它们也可能危害水产养殖业, 例如赤潮^[2]。在进化关系上藻类植物跨越了原核和真核两界, 对于了解生物起源、演化及机制意义重大, 是生物学家共同关注的热点之一^[3]。同时对藻类进行分子水平的研究, 对于改善其经济性状、减少渔业危害以及了解地球上生物的进化都有着重要的意义。

1 藻类基因组研究现状

1.1 藻类基因组的研究

藻类染色体数变异较大, 自 1921 年 Ishikawa 首次报道了甘紫菜的核分裂以来, 已对 62 种紫菜做了染色体计数。分析这些报道过的染色体资料, 发现具 3 条染色体的种有 28 个, 占绝大多数。具 2 条的有 8 个, 4 条的有 12 个, 5 条的有 9 个, 具 6 条和 7 条染色体的种各有 1 个。尽管紫菜属的染色体组数很少, 但是由于染色体很小、难分辨, 所以对同 1 个种得出的染色体数目也不同。

由于紫菜种间或种内染色体数目存在变异。有人推测紫菜的原始种染色体数是 3 条(共 28 种)。2 条的物种是由于其中 1 条染色体丢失。4 条的物种则是由于其中 1 条染色体发生断裂所造成的, 5 条的物种可能是由含 3 条染色体的物种先加倍后丢失 1 条造成的。又发现在同 1 个种内虽然染色体数目不

同, 却具有基本相同的 2c DNA (2 倍染色体组) 含量, 证实染色体数目变异是由于融合或断裂而来, 而不是通过整条染色体丢失或获得产生的。

随着分子生物技术的发展, 越来越多的物种包括藻类被不同程度地测序。自 1990 年首次测序了斜生栅列藻 (*Scenedesmus obliquus*) 线粒体全基因组序列^[4], 到目前已经测序了 45 条藻类线粒体、叶绿体、染色体全序列和一个单细胞红藻 *Cyanidioschyzon merolae* 10D 的全基因组序列^[5]。

继 Ohta^[6]在 2003 年测序了单细胞红藻 *C. merolae* 10D 的质体基因组, Matsuzaki 等 2004 年又对其全基因组序列进行了测序, 也是目前唯一被完全测序的藻类。分析发现这种红藻的细胞核中有 20 条染色体, 这些染色体由超过 1 652 万个碱基对组成, 线粒体和叶绿体中碱基对的数量分别约为 3.2 万和 15 万个, 找出的 5 331 个基因中有 86.3% 是表达的, 是目前所知的最小的植物基因组。这种红藻只有生存所必需的最小限度的基因, 这为证明哪些是光合真核生物和植物必须的基因提供了重要信息。该物种的基因所含的内含子序列少, 核基因中有 3 个拷贝的核糖体 DNA, 缺少肌动蛋白基因。从遗传信息的角度来看, 这类红藻是最原始的。它为研究真核细胞起源、进化和功能机制提供了由简单基因组成的模式系统^[5]。

1.2 藻类线粒体基因组研究

内共生起源的学说认为单细胞的真核吞噬生物或原生生物吞噬原核生物形成了真核生物的细胞器, 该理论目前已经被普遍接受^[7,8]。通过 SSU (small subunit) rRNA 做的系统发生分析认为线粒体起源于 α -蛋白细菌^[9], 通过对线粒体 16sRNA 基因及热激

收稿日期: 2005-04-06; 修回日期: 2005-07-30

作者简介: 方永军(1977-), 男, 浙江人, 博士, 研究方向: 基因组学, E-mail: fangyj@genomics.com.cn; 胡松年, 通讯作者, E-mail: husn@genomics.org.cn

蛋白 HSP70 基因分析发现与真细菌 *Rickettsia* 高度相似^[10,11]。COX1 基因及其他编码蛋白基因的系统发生证明了高等植物及藻类的线粒体起源共同的祖先^[12]。线粒体基因组起源于真细菌的 α -蛋白细菌已被广泛认可。

几乎所有的真核生物线粒体基因组都有编码呼吸作用复合体亚单位的基因,如细胞色素 b、细胞色素氧化酶、氧化磷酸化相关基因以及所有的 rRNAs 基因和大部分的 tRNAs^[13]。线粒体基因组以动物的最小,为 14~42 kb,呈环形,结构单一,基因排列紧密;真菌其次为 17~180 kb,环状排列^[14];植物的最大为 184~2 400 kb,结构也最复杂。

藻类的线粒体基因组研究相对较少,1990 年, Kuck 首次测序了斜生栅列藻(*S. obliquus*)线粒体全基因组序列(NC_002254),分析线粒体大亚基核糖体 RNA 基因(LSURRNA),发现了 2 个分别含有两类内含子的两个插入序列,并通过成熟 rRNA 确定了外显子和内含子的边界。实验证明含内含子的 LSURRNA 转录前体在体外高盐环境下有很强的自剪切活性,这是第一次报道了在 LSURRNA 基因中存在 II 型内含子并且是真核生物细胞器中最小的自剪切内含子^[4,15]。其后,Kuck^[16]对 *S. obliquus* 的线粒体全基因组进行的分析发现,总长 42 781 bp 的环形基因组除了含有一套高保守的其他植物都具有的 13 个线粒体基本基因外,还含 2 个不明确的 rRNA,27 个 tRNA 基因以及 4 个内含子序列,以及 5 个找不到同源性的 ORF,并证明终止子 UGA 曾经编码亮氨酸,而 UCA 曾是无义密码子。

1999 年,Burger 等^[17]测定了 *Porphyra purpurea* 的线粒体全序列,环形排列的 36 753 个碱基对中包含了 57 个紧密排列的基因,其中有与动物和真菌中特有基因相似的,也有与原生生物、植物线粒体特有基因相似的,还有与核糖体特异蛋白和琥珀酸酯亚单位相似的基因。大亚基 rRNA 含有 2 个与藻青菌(*Calothrix* sp.)中极为相似的 II 型内含子,推测细菌和线粒体基因组间在近期发生了遗传物质的横向转移。*P. purpurea* mtDNA 的另外一个典型特征是含有两段可能导致基因组重组的 291 bp 的反向重复序列,在编码区及基因间也发现大量的序列多态性。通过对 5 个进化位置不同的线粒体基因组进行比较分析,认为红藻的线粒体基因组有着特有的基因大小和排序^[18,19]。系统发生分析证明了红藻、绿藻和植物有着共同的祖先^[17]。

单细胞的莱茵衣藻(*Chlamydomonas reinhardtii*)线粒体基因组序列为 15.8 kb,为直线型,而同为衣藻的 *C. eugametos* 和 *C. moewusii* 分别为 24,22 kb,为环状分子^[20];绿枝藻纲(Prasinophyceae)的亚心形扁藻(*Tetraselmis subcordiformis*)和绿球

藻目(Chlorococcales)的威克海姆原藻(*Prototheca wickerhamii*)线粒体基因组是上述衣藻的 5 倍,分别为 42.8 kb 和 55 kb,但是在 *T. subcordiformis* 中未发现内含子,有一约 1.5 kb 的反向重复序列;而 *P. wickerhamii* 中则存在很多的内含子和基因间序列^[21,22]。卡氏棘阿米巴(*A. castellanii*)的 mtDNA 41 kb 却与 *P. wickerhamii* 含有几乎同样多的遗传信息,主要是因为它的基因分布更密集甚至发生重叠,基因间的序列减少^[23]。

Leblanc^[13]结合红藻线粒体数据分析了线粒体基因组,认为线粒体从十亿年前到现在结构和功能发生了很大的变化,植物线粒体庞大且有重排趋势,真菌、动物则有减少趋势。这种分化部分归因于平行同源进化,线粒体进化崇尚基因节俭的原则,因此越来越小,基因也越来越少,并且认为红藻和高等植物的线粒体关系更为亲近。

1.3 藻类质体基因组研究

质体是植物细胞中的光合细胞器,在高等植物及绿藻中称为叶绿体。基因组一般都是闭合环状的 DNA 分子,藻类质体基因组大小在 73~400 kb 之间,比高等植物的大。现在已建立了 20 多种藻类质体基因组的物理图谱与部分基因图谱,有 18 种已被完全测序,但总的来说仍落后于高等植物叶绿体的研究。

高等植物叶绿体基因组是由大小均一的 DNA 分子组成,但在藻类质体中发现 cpDNA 有异质性。1972 年 Nass 等^[24]首先在细小裸藻(*Euglena gracilis*)的无色突变体中发现有小的环状 DNA 分子。Heizmann 等^[25]1982 年进一步证实了叶绿体中有质粒状 DNA,且发现它与 cpDNA 主基因组有同源性。随后又在伞藻(*Parvocaulis*)、摩氏衣藻(*C. morganii*)、管枝藻(*Siphonocladales*)等绿藻中发现有质粒状 DNA。伞藻的研究清楚地表明其质粒状 DNA 就在叶绿体内,即 cpDNA 是由大小不均一的 DNA 分子组成。

绿藻、高等植物叶绿体基因组可以编码 110 个左右基因,红藻的质体基因组的编码能力比前者多 70 个左右。在 *P. purpurea* 部分基因组(70%)上已定位了 125 个基因,这是迄今为止发现的基因容量最大的一个质体基因组。质体基因组编码的基因主要有两类:1) 质体自身遗传系统相关的基因,tRNA、rRNA 及核糖体蛋白基因等;2) 光合作用直接有关的基因,包括天线色素蛋白、光系统作用中心蛋白、RuBP 羧化酶及 atp 酶亚基的编码基因等。与陆地植物明显不同的是在红藻中有一类特殊的天线色素(藻胆蛋白)基因。另外,在藻类中还有陆地植物不具有的一类基因:rbcS、secY、dnaK、petF 等。研究显示从低等藻到进化程度较高的绿藻再到高等植物,存在着质体基因组容量由大到小,编码的基因由多到少,基

因发生重排和转移而逐渐稳定的规律。

质体起源于内共生的光合微生物,这是一个被普遍接受的假说。但争论在于红藻、绿藻、褐藻等不同类型藻的质体是起源于同一祖先还是有多个不同的祖先,即存在一个单一起源还是多起源的问题^[26]。比较多的研究认为不同类型的质体有不同的起源,但认为质体来自一个共同祖先的间接证据也不少。质体 DNA 序列都有着类似的组成及结构特征,如反向的 rRNA 重复序列、类似的质体基因操纵子进化^[27]。另外一个单一起源的有力证据是基因产物转移机制:由质体转移到核基因组上的基因的产物最终返回到来源细胞器上^[3]。Shivji^[28]比较了 2 种红藻与亲缘、地缘关系比较远的 2 种具色藻的质体基因组基本结构、基因容量及基因的组织结构后发现质体基因组之间有明显的衍生关系,即进化关系,也就是说有共同起源。

红毛茛亚纲即紫菜亚纲是其他藻类(如隐藻、定鞭藻)质体的起源,为了解它的系统发生及二次内共生, Viehmann^[29]测序分析了来自 9 种红毛茛亚纲和 2 种来自真红藻亚纲的质体 SSU rDNA,证明了红毛茛亚纲和真红藻亚纲均是单起源的。第一次内共生发生后质体基因组的遗传信息发生大量丢失,高达 90% 以上的信息被转移到核基因中去^[30,31],因此通过对所含质体物种的全基因组进行测序和分析对于了解质体的起源和进化有着非常重要的意义。Geifrey^[3]根据红藻 *C. merolae* 10D 基因组、疟原虫基因组、蓝藻基因组、硅藻假矮海链藻(*Thalassiosira pseudonana*)基因组及衣藻质体基因组分析也证明了所有的质体起源于同一个祖先。

为了解红藻质体的发生进化, Glockner^[32]于 2000 年测序分析了单细胞红藻 *Cyanidium caldarium* RK1 的叶绿体基因,发现了一些特殊的结构特征,例如非遗传信息区域的发夹结构,没有大的重复单元,并且认为这种叶绿体基因组的结构是最原始的。与其他已知的藻类质体基因组进行比较和系统发生分析显示二次内共生发生在 *C. caldarium* 和 *P. purpurea* 分离之后。2004 年 Sheveleva^[33]发现蓝细菌的自剪切自引导的 group II 内含子横向转移到了 *Euglena myxocylindracea* 叶绿体基因组中,这为了解叶绿体的遗传物质提供了新的视角。

Ohta 等^[6]2003 年测序了单细胞红藻 *Cyanidioschyzon* 10D 的质体基因组,找到编码包括 36 个 RNA 和 207 个蛋白的总共 243 个不存在内含子序列的基因。首次在红藻质体中发现低等绿藻质体中所有的 *cysY*。也是第一次在质体基因组中发现了编码硫酸盐转移酶的 *cysW*,推测它与核基因共同在生物体内的硫酸盐转移系统中起作用。*CysY* 和 *cysW* 的蛋白产物与蓝藻的高度相似。

莱茵衣藻(*Chlamydomonas reinhardtii*)是一种单细胞真核绿藻,只含有一个叶绿体,是研究光合作用过程的很好的模型。叶绿体全基因组 203395 bp,被一个 21.2 kb 的反向重复分成两个 80 kb 的单拷贝,共含有 99 个基因,基因组的典型特征是重复序列占 20% 以上,大量的短散在重复序列(SDRs)存在于基因间区域,它们对研究基因的结构和进化有着重要的意义。同样的特征只有在小球藻(*Chlorella vulgaris*)的叶绿体基因组中存在^[34]。比较分析莱茵衣藻叶绿体和其他物种的叶绿体基因组,认为 SDRs 可能来源于别的基因组片段。同时根据质体基因组遗传物质的变异作系统发生分析发现叶绿体基因组中基因丢失的频率呈增加的趋势^[35]。

1.4 藻类其他方面的研究

藻类其他方面的研究主要集中在藻类营养成分及自我保护和自我修复机制。藻类在太阳辐射下会动态地产生保护性复合物。Helbling 等^[36]评估了 4 种红藻在辐射下产生保护性物质的能力,认为太阳辐射下会产生 UV 吸收物质和类胡萝卜素进行自我保护,从而影响水生植物的营养成分。MAA 是一种抗紫外线物质, Sinha^[37]将 3 种来源不同的蓝细菌分别置于不同的辐射下,测试吸收和御防中 MAA 的含量,发现 UV-B 比 PAR 和 UV-A 更有效刺激蓝细菌产生 MAA。

癌细胞中端粒酶活性偏高,抑制端粒酶活性是治疗癌症途径之一。Eitsuka 等^[38]研究了海藻和植物中的硫代异鼠李糖甘油二酯(SQDC:sulfoquinovosyl-diacylglycerol)对端粒酶的抑制作用,发现 SQDC 的硫酸盐聚合体和脂肪酸结构对抑制端粒酶活性很重要。因此认为可通过食疗法抑制端粒酶的活性。

Zhang^[39]从坛紫菜(*Porphyra haitanensis*)中分离出具有典型卟啉结构硫酸半乳糖片断 F1,能明显提高抗氧化和 SOD 的活性。Takenaka 等^[40]发现给 VB₁₂ 缺乏的小鼠喂紫菜能明显提高 VB₁₂ 的含量。

对于藻类的生殖也存在比较多的争论。Berthold 在 1882 首次发现紫菜发生了有性生殖,1978 年从亚显微结构和细胞学上也证明了紫菜的生活史中确实发生了有性生殖。在研究中发现了单倍的丝状体,认为紫菜也存在无性生殖方式。关于减数分裂发生的部位和时期也有 3 种不同的观点。曾呈奎认为受精果孢的第 1、2 次细胞分裂是减数分裂,果孢子是单倍体。Kito 提出了减数分裂在丝状体上孢子囊枝产生壳孢子时期。有些学者则认为紫菜减数分裂发生在壳孢子萌发时期。

2 EST 技术

所谓 EST,就是指对应于 mRNA 的 cDNA 上的一条序列片段,一般长度为 300 ~ 500 bp。通常是从

已有的 cDNA 文库中随机挑取进行克隆,对其 3' 或 5' 端进行一步法测序所获得的序列。目的是以 DNA 的部分序列代表特定时期特定组织中表达的基因。可代表生物体某种组织某一时间的一个表达基因,故被称之为“表达序列标签”^[41]。

2.1 EST 技术和分析方法

EST 技术是基于 cDNA 文库基础上的 cDNA 克隆的部分序列测定,包括建立生物体特定组织或细胞的高质量 cDNA 文库 → 从文库中随机挑取足够量的 cDNA 克隆进行 5' 端或 3' 端测序 → 生物信息学分析 → 进一步分析未知基因或归类分析全部 EST 以获得特异组织或细胞的基因表达概况。

cDNA 文库有普通文库和标准文库之分。普通 cDNA 文库指未经标准化的直接由 mRNA 反转录而来的。通过对它的分析可以了解特定组织在特定时期基因表达丰度情况。标准 cDNA 文库是去除了无信息(如 rRNA、线粒体 cDNA 和全部由多聚腺苷酸组成的插入序列)克隆,各种表达基因频率基本一致的 cDNA 文库,可以通过消减杂交,均等化来实现。理想的 cDNA 文库的另一特征是定向克隆,这样可选择性地得到编码序列或是 3' 端非编码序列。因此可以根据自己研究需要来构建和选择文库。

对随机挑取的克隆得到的 EST 进行处理和分析,包括:(1) 去除载体、宿主菌、线粒体或重复序列;(2) 由于 EST 序列代表随机取样的各种转录产物 mRNA,可能存在多个 EST 代表同一个转录产物,因此可以用 Phrap/ CAP3/ Stack PACK/ TIGR Assembler 程序对它们进行聚类 and 拼接。(3) 对拼接后的序列进行数据注释,主要方法为序列比对,根据序列的同源性推测功能上的相似性,分析的结果大致有 3 种:第一是已知基因,即为已鉴定和了解的基因;第二是以前未经鉴定的新基因;第三是未知基因。

2.2 EST 的应用

2.2.1 构建遗传学图谱

序列位置标签(Sequence Tagged Site, STS):一段短的 DNA 序列(200~500 个碱基对),这种序列在染色体上容易辨认,通常只出现一次,其位置和碱基顺序都是已知的。因此用于作基因组图谱的地标。3' 端 EST 是基因特异的,保守性较低,所以用 3' EST 完全可以充当 STS 界标,这样把大量的 EST 标记加入物理图中,构建了 EST 图谱。

2.2.2 基因的分离和定位克隆

分离基因的经典方法为图位克隆和转座子标签克隆,是利用分子标记或表型变化来分析鉴定基因。但在 DNA 分子标记中 RAPD、SSR 等标记与目标基因间的连锁值的大小,所用的分子标记是否处于目标基因的编码区内,这些未知因素都将导致克隆基

因的失败。而以 EST 为重要来源的染色体物理图谱,可将基因确定在更狭小的染色体区段内,缩小了基因的筛选范围。同时,EST 标记是功能序列区的标记,处于目标基因的范围,而且是目标基因的编码区。因而 EST 标记是与目标基因共分离的,以这样的标记进行定位克隆将大大提高克隆基因的效率。将所获 EST 用生物信息学方法与各公共数据库中已知序列进行比较,可迅速而准确地确定基因功能。

2.2.3 EST 寻找新基因

将所得的 EST 数据与蛋白质数据库、dbEST 等数据库的数据比较,可以确定哪些代表已知基因、哪些代表已知 EST、哪些代表未知基因。从 Admas 在 1991 年首次应用表达序列标签的方法寻找人脑的新基因后,通过 EST 技术寻找新基因成为人类研究生物基因组的主要方法。

2.2.4 基因差异表达的研究

一般认为某一时期的基因表达数量通常占全部基因的 15%,细胞的分化由基因特异性的时空表达决定。近年来,对基因的差异表达研究发展了如差减杂交、代表性差异分析、mRNA 差别显示等技术,这些技术各有优势,而 EST 技术的优势在于其稳定性高和分析规模大。对一个未经标准化处理的 cDNA 文库随机挑选克隆进行大规模测序,可直接回答特定组织细胞在某一时期哪些基因表达了,丰度如何等问题,从而能在基因整体水平上研究相关的功能及代谢。

目前基因差异表达研究是家蚕 EST 研究的主流;比较大豆植株不同发育时期和每一个发育时期植株的不同组织或器官 cDNA 文库的 EST 信息,可以绘出大豆基因在时间和空间上基因表达的发育顺序谱;用来自曼氏血吸虫 4 个生活阶段的 7 个不同的 cDNA 文库的 1401 个 EST,对曼氏血吸虫不同发育阶段的基因表达概况作了初步研究;用于肿瘤细胞、正常人体组织和胚胎细胞的基因表达概况研究,系统分析肿瘤细胞基因表达的多样性和活性。

2.2.5 从 EST 中发掘单核苷酸多态

单核苷酸多态性(Single nucleotide polymorphisms, SNP)是指同一位点的不同等位基因之间只有一个核苷酸的差异或只有小的插入、缺失,SNP 标记又被称为第三代 DNA 分子标记。EST 是随机 cDNA 挑克隆单步法测序得到的,来自于不同 cDNA 文库和不同个体的大量冗余的 EST 是发掘 SNP 很好的资源。从 EST 中发掘 SNP 所用的软件方法各不同,但基本的思想都是先聚合同源序列,然后比较这些同源序列。

2.2.6 用于制备 DNA 芯片

芯片的制作必须知道大量的已知序列基因或基因片段作为探针,但对于没有测序,或者只有部分测序的基因组,或者对于那些有大量内含子的基因组,

知道每个 DNA 是不现实的。在这种情况下,可以将单个 EST 的 cDNA 克隆用作阵列 DNA 的来源。

3 藻类的 EST 研究

自从 EST 技术建立以来,dbEST 数据库从 1993 年成立之初的 22 537 条记录增加到了现在的 28022933 (2005-08-01) 条。搜索的结果表明藻类的 EST 数据达 236375 条,与高等植物和动物相比较,数据量仍较少。其中对模式生物莱茵衣藻(*Chlamydomonas reinhardtii*) 研究最为详细,共有 167707 条相关 EST 纪录,其次为条斑紫菜 20979 条 EST 纪录。

1999 年 Asamizu 等^[42]从光合自养生长的单细胞莱茵衣藻的均一化、大片段 cDNA 文库中筛选了 11 571 个 5' 端 EST,非冗余序列为 34 333 个。比对后 817 条为已知序列,其中 140 条为已知的衣藻基因。为了分离与碳胁迫相关的基因,2000 年又对在低 CO₂ 和高 CO₂ 条件下生长的莱茵衣藻进行了 EST 分析,分别构建了 12 969 和 13 450 个 5' 端 EST 标签^[43]。

2001 年 Becker^[44]从绿色鞭毛藻(*Scherffelia dubia*) cDNA 文库中分离到 361 个 EST,其中 138 个与已知序列相似,134 个 EST 为未知序列,大多数已知序列与分裂间期的营养积累和蛋白质生物合成相关。为了找出腰鞭毛虫叶绿素蛋白的编码基因和它的进化历史,Bachvaroff^[45]从两个 cDNA 文库中随机测序了 3 155 个克隆进行分析,找到 118 个与质体相关的一致序列,其中有 30 个是由红藻的质体基因组编码的。结果显示在鞭毛虫中有过大规模的基因从质体转移到核基因中,并且支持叶绿素型质体有一很小基因组的假说。

为比较海带在不同生活史下基因表达的不同,Crepineau 等^[46]分别从掌状海带(*Laminaria digitata*) 配子体和孢子体的 cDNA 文库中选取了大约 500 条序列进行 EST 分析。在配子体的 412 个 EST 中有 39% 与已知序列相似,而孢子体的 493 个中有 48% 的相似。

红藻 EST 研究主要集中在紫菜(*Porphyra*),共 21 346 条数据,而又以条斑紫菜(*P. yezoensis*) 为主,有 EST 记录 20 979 条;其次为江蓠属(*Gracilaria*) 366 条。通过 EST 寻找与其生长发育、抗病和抗逆性相关的基因是一快速有效的方法。Lluisma^[47]在 1997 年首次报道了江蓠(*G. gracilis*) 的 200 条 EST,将这些 EST 与数据库中已知的多肽序列进行比对,结果 73% 的 EST 未找到显著的相似序列,鉴定的 EST 多为与碳代谢、氨基酸代谢、光合作用及蛋白质的合成及降解等有关。Sun^[48]在研究江蓠配子体得到的 180 条 EST(151 个非重复序列)时,与核苷酸数据库中的已知序列进行比较,找到了 30 条与已知基

因显著相似的 EST。用递交的 136 条 EST 与红藻 EST 数据进行比较,找到了 22 条相似的克隆。在蛋白质数据库搜索到 73 条(58.1%) 与已知功能蛋白相似的克隆,其中大部分与能量代谢、转录/翻译及光合作用相关。

Nikaido 等^[49]从条斑紫菜配子体的 cDNA 文库中建立了 10 154 条 EST,比较后发现其中 33.1% 的序列与已知序列有相似性,66.9% 的 EST 是首次报道的新序列。并对 101 个与已知基因显著相似的非重复 EST 的编码区密码子进行分析,结果显示密码子第一、第二和第三位 GC 含量分别为:62.2%、45% 和 79.4%,表明紫菜基因组在进化过程中遇到了较高的 GC 压力。为了找出与条斑紫菜配子体和孢子体时期不同形态和生理相关的基因,2003 年 Erika Asamizu 等^[42]进行大规模 EST 分析。从孢子体文库中测序得到 10 625 个 5' 端 EST,结合 Nikaido 从配子体测得的 10 154 个 EST,以 50 bp 有 95% 以上一致为标准,总共 20 779 个 EST 组成 4 496 个非冗余群。其中只有 1 013(22.5%) 是两个世代共有的,大部分的被各自 EST 分到了配子体世代(1 940, 43.1%) 或孢子体世代(1 543, 34.3%),统计分析表明配子体中有 89 个,孢子体中有 112 个高表达基因。对 12 个基因作了 RT-PCR,发现存在不同的表达。2002 年杨官品^[50]也报道了来自条斑紫菜丝状孢子体的 170 条 EST,其中新序列占 42.94%,在已知功能的 43 组 EST 中,大部分与能量代谢和蛋白质合成相关,有 3 组与生长发育相关,6 组与抗逆或防御相关。

坛紫菜作为红藻属的一个重要成员,是中国南方海域特有的栽培品种,然而国际上对其的研究报道很少,NCBI 中目前还没有相关 EST 收录。中国科学院海洋研究所和浙江大学沃森基因组科学研究院 2003 年联合对坛紫菜丝状孢子体进行了大规模测序,随机挑取克隆测序得到 7 064 个 EST 原始数据,经过处理最终获得 2 535 条唯一序列,包括 945 个 contigs 和 1 590 个 singlets,985(38.9%) 条序列被注释,找到了一些与蛋白的合成、加工及降解相关,与碳代谢、能量代谢相关的基因,并且首次揭示了紫菜中极有可能存在抗氰呼吸与 C4 固碳途径;另外还发现了与细胞信号传导密切相关的多种蛋白质及蛋白质家族基因,如 Hsp70 family、calceineurin、Rho family。此外,紫菜独特的光合色素——藻胆色素蛋白(如藻红蛋白)及其合成过程中的几个关键酶基因在作者的结果中也有所发现。

参考文献:

[1] 李孟军,赵建成. 蓝藻的起源和演化[J]. 生物学通

- 报, 2000, 35(1): 19-20.
- [2] Fleming L E, Backer L C, Baden D G. Overview of aerosolized Florida red tide toxins: exposures and effects[J]. **Environ Health Perspect**, 2005, 113(5): 618-620.
 - [3] McFadden G I, van Dooren G G. Evolution: red algal genome affirms a common origin of all plastids[J]. **Current Biology**, 2004, 14:514-516.
 - [4] Winkler M, Kuck U. The group IIB intron from the green alga *Scenedesmus obliquus* mitochondrion: molecular characterization of the in vitro splicing products[J]. **Curr Genet**, 1991, 20(6): 495-502.
 - [5] Matsuzaki M, Misumi O, Shin I T, *et al.* Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D[J]. **Nature**, 2004, 428:653-657.
 - [6] Ohta N, Matsuzaki M, Misumi O, *et al.* Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*[J]. **DNA Research**, 2003, 10: 67-77.
 - [7] Loiseaux-de Goër S. Plastid lineages[A]. Round E, Chapman DJ. Progress in phycological research 10 [C]. Bristol: Biopress Ltd, 1994. 137-177.
 - [8] Gray M W, Doolittle W F. Has the endosymbiont hypothesis been proven? [J]. **Microbiol Rev**, 1982, 46:1-42.
 - [9] Yang D, Oyaizu Y, Oyaizu H, *et al.* Mitochondrial origins[J]. **Proc Natl Acad Sci USA**, 1985, 82:4 443-4 447.
 - [10] Olsen G J, Woese C R, Overbeek R. The winds of (evolutionary) change: breathing new life into microbiology[J]. **J Bacteriol**, 1994, 176:1-6.
 - [11] Gupta R S. Evolution of the chaperonin families (Hsp60, Hsp10 and Tcp-1) of proteins and the origin of eukaryotic cells[J]. **Mol Microbiol**, 1995, 15: 1-11.
 - [12] Wolff G, Burger G, Lang B F, *et al.* Mitochondrial genes in the colourless alga *Prototheca wickerhamii* resemble plant genes in their exons but fungal genes in their introns[J]. **Nucleic Acids Res**, 1993, 21:719-726.
 - [13] Leblanc C, Richard O, Kloareg B, *et al.* Origin and evolution of mitochondria: what have we learnt from red algae? [J]. **Curr Genet**, 1997, 31: 193-207.
 - [14] Clark-Walker G D. Evolution of mitochondrial genomes in fungi[A]. Wolstenholme D R, Jeon K W. Mitochondrial Genomes[C]. San Diego: Academic press, 1992. 89-127.
 - [15] Kuck U, Godehardt I, Schmidt U. A self-splicing group II intron in the mitochondrial large subunit rRNA (LSUrRNA) gene of the eukaryotic alga *Scenedesmus obliquus*[J]. **Nucleic Acids Res**, 1990, 18(9): 2 691-2 697.
 - [16] Kuck U, Jekosch K, Holzamer P. DNA sequence analysis of the complete mitochondrial genome of the green alga *Scenedesmus obliquus*: evidence for UAG being a leucine and UCA being a non-sense codon [J]. **Gene**, 2000, 253(1):13-18.
 - [17] Burger G, Saint-Louis D, Gray M W, *et al.* Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*. Cyanobacterial introns and shared ancestry of red and green algae [J]. **Plant Cell**, 1999, 11(9): 1 675-1 694.
 - [18] Ohta N, Sato N, Kuroiwa T. Structure and organization of the mitochondrial genome of the unicellular red alga *Cyanidioschyzon merolae* deduced from the complete nucleotide sequence[J]. **Nucleic Acids Res**, 1998, 26: 5 190-5 298.
 - [19] Lang B F, Goff L J, Gray M W. 5S rRNA gene in the mitochondrial genome of the protist *Reclinomonas americana*: Absence of the same gene in red algal mitochondrial DNA [J]. **J Mol Biol**, 1996, 261: 607-613.
 - [20] Denovan-Wright E M, Lee R W. Comparative analysis of the mitochondrial genomes of *Chlamydomonas eugametos* and *Chlamydomonas moewusii* [J]. **Curr Genet**, 1992, 21:197-202.
 - [21] Kessler U, Zetsche K. Physical map and gene organization of the mitochondrial genome from the unicellular green alga *Platymonas (Tetraselmis) subcordiformis* (Prasinophyceae) [J]. **Plant Mol Biol**, 1995, 29:1 081-1 086.
 - [22] Wolff G, Plante I, Lang B F, *et al.* Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*[J]. **J Mol Biol**, 1994, 237:75-86.
 - [23] Burger G, Plante I, Lonergan K M, *et al.* The mitochondrial DNA of the amoeboid protozoan, *Acanthamoeba castellanii*: complete sequence, gene content, and genome organization[J]. **J Mol Biol**, 1995, 245:522-537.
 - [24] Nass M M, Ben-Shaul Y. A novel closed circular duplex DNA in bleached mutant and green strains of *Euglena gracilis*[J]. **Biochim Biophys Acta**, 1972, 272(1): 130-136.
 - [25] Nicolas P, Heizmann P, Nigon V. Different phenotypes of non-photosynthetic (photo-) phenotypes in *Euglena gracilis*: frequency of production by ultraviolet irradiation[J]. **C R Seances Acad Sci**, 1982,

- 294(3):145-148.
- [26] Palmer J. The symbiotic birth and spread of plastids: how many times and whodunit? [J]. **J Phycol**, 2003, 39, 4-12.
- [27] Stoebe B, Kowallik K. Gene-cluster analysis in chloroplast genomics[J]. **Trends Genet**, 1999, 15, 344-347.
- [28] Shivji M S, Li N, Cattolico R A. Structure and organization of rhodophyte and chromophyte plastid genomes: implications for the ancestry of plastids[J]. **Mol Gen Genet**, 1992, 232(1):65-73.
- [29] Oliveira M C, Bhattacharya D. Phylogeny of the bangiophycidae (rhodophyta) and the secondary endosymbiotic origin of algal plastids[J]. **American Journal of Botany**, 2000, 87(4):482-492.
- [30] Martin W, Stoebe B, Goremykin V, et al. Gene transfer to the nucleus and the evolution of chloroplasts[J]. **Nature**, 1998, 393, 162-165.
- [31] Richly E, Leister D. An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of Arabidopsis and rice [J]. **Gene**, 2004, 329: 11-16.
- [32] Glockner G, Rosenthal A, Valentin K. The structure and gene repertoire of an ancient red algal plastid genome[J]. **J Mol Evol**, 2000, 51(4):382-390.
- [33] Sheveleva E V, Hallick R B. Recent horizontal intron transfer to a chloroplast genome [J]. **Nucleic Acids Res**, 2004, 2(2):803-810.
- [34] Wakasugi T, Nagai T, Kapoor M, et al. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division [J]. **Proc Natl Acad Sci USA**, 1997, 94(11):5967-5972.
- [35] Jude E M, Jason W L, Cui Liying, et al. The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats[J]. **The Plant Cell**, 2002, 14: 2659-2679.
- [36] Helbling E W, Barbieri E S, Sinha R P, et al. Dynamics of potentially protective compounds in Rhodophyta species from Patagonia (Argentina) exposed to solar radiation [J]. **J Photochem Photobiol B**, 2004, 75(1-2):63-71.
- [37] Sinha R P, Ambasht N K, Sinha J P, et al. UV-B-induced synthesis of mycosporine-like amino acids in three strains of *Nodularia* (Cyanobacteria) [J]. **J Photochem Photobiol B**, 2003, 71(1-3):51-58.
- [38] Eitsuka T, Nakagawa K, Igarashi M, et al. Telomerase inhibition by sulfoquinovosyldiacylglycerol from edible purple laver (*Porphyra yezoensis*) [J]. **Cancer Lett**, 2004, 212(1):15-20.
- [39] Zhang Q, Li N, Liu X, et al. The structure of a sulfated galactan from *Porphyra haitanensis* and its in vivo antioxidant activity[J]. **Carbohydr Res**, 2004, 339(1):105-111.
- [40] Takenaka S, Takubo K, Watanabe F, et al. Occurrence of coenzyme forms of vitamin B12 in a cultured purple laver (*Porphyra yezoensis*) [J]. **Biosci Biotechnol Biochem**, 2003, 67(11):2480-2482.
- [41] 胡松年. 基因表达序列标签 (EST) 数据分析手册 [M]. 杭州:浙江大学出版社, 2005. 1-8.
- [42] Asamizu E, Nakajima M, Kitade Y, et al. Comparison of RNA expression profiles between the two generations of *Porphyra yezoensis* based on expressed sequence tag frequency analysis[J]. **Journal of Phycology**, 2003, 39(5):923-928.
- [43] Asamizu E, Nakamura Y, Sato S, et al. *Chlamydomonas Reinhardtii*. I. Generation of 3433 non-redundant expressed sequence tags[J]. **DNA Res**, 2000, 6: 369-373.
- [44] Becker B, Feja N, Melkonian M. Analysis of expressed sequence tags (ESTs) from the scaly green flagellate *Scherffelia dubia* Pascher emend. Melkonian et Preisig[J]. **Protist**, 2001, 152: 139-147.
- [45] Bachvaroff T R, Concepcion G T, Rogers C R, et al. *Dinoflagellate* expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome[J]. **Protist**, 2004, 155(1):65-78.
- [46] Crepineau F, Roscoe T, Kaas R, et al. Characterisation of complementary DNAs from the expressed sequence tag analysis of life cycle stages of *Laminaria digitata* (Phaeophyceae) [J]. **Plant Mol Biol**, 2000, 43: 503-513.
- [47] Lluisma A O. Expressed sequence tags (ESTs) from the marine red alga *Gracilaria gracilis* [J]. **Journal of Applied Phycology**, 1997, 9: 282-287.
- [48] Sun Xue, Yang Guanpin, Mao Yunxiang, et al. Analysis of expressed sequence tags of marine red alga, *Gracilaria lemaneiformis* [J]. **Nucleic Acids Res**, 2002, 12(7):518-523.
- [49] Nikaido I, Asamizu R, Nakajima M, et al. Generation of 10154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis* [J]. **DNA Research**, 2000, 7: 223-227.
- [50] 杨官品, 沈怀许, 许璞, 等. 条斑紫菜丝状孢子体表达序列标签分析[J]. 高技术通讯, 2002, 12(2):93-97.

(本文编辑:张培新)