

串珠藻目植物 *rbcL* 基因的适应性进化分析*

巩超彦 南芳茹 冯佳 吕俊平 刘琪 谢树莲^①

(山西大学 生命科学学院 太原 030006)

摘要 淡水红藻是藻类植物进化中的一个重要类群, 其中最主要的类群是串珠藻目。核酮糖 1,5 二磷酸羧化酶/加氧酶(Rubisco)是一种既负责碳同化又引发光呼吸, 在植物光合作用中起重要作用的酶, 其催化固定 CO₂ 的活性中心位于 Rubisco 大亚基, 由叶绿体 *rbcL* 基因编码。为了探究串珠藻目植物适应特殊生存环境的分子水平机制, 本文选取了串珠藻目及一些相近类群淡水红藻的 *rbcL* 基因 39 条, 利用 PAML4.8 软件, 运行位点模型、分支模型及分支-位点模型, 对选取类群的 *rbcL* 基因进行了适应性进化分析。结果表明: (1) 通过最大似然法构建的系统发育树显示, 所有内类群聚集为 6 个分支, 其中, 分支 A 为暗紫红毛茛, 分支 B 为胶串珠藻, 分支 C 为扁圆串珠藻, 后验概率同样为 100%, 分支 D 为熊野藻属, 分支 E 为连珠藻属, 分支 F 为弧形西斯藻; (2) 分支-位点模型中, 在 3 个分支中分别鉴定出 350S、277L 和 280L 为正选择位点; (3) 在构建出的 Rubisco 大亚基的参考三维模型中, 277L 和 280L 位于 Rubisco 大亚基羧基末端保守的 8 个 α 螺旋和 8 个 β 片层构成的 α/β 桶状结构域中第 7 个 α 螺旋和第 7 个 β 片层之间的 loop 结构上, 350S 位于羧基末端邻近 α/β 桶状结构域的一个 α 螺旋上。研究结果一方面显示了基于 ω 比值检验基因适应性进化的准确性和有效性, 另一方面也揭示了串珠藻目植物 *rbcL* 基因确实发生了适应性进化, 对串珠藻目适应特殊生存环境产生了有益的作用。

关键词 串珠藻目; *rbcL* 基因; 适应性进化

中图分类号 Q941 doi: 10.11693/hyhz20161200277

绝大多数红藻为海产, 只有少数生于淡水。这些淡水红藻是海陆演变过程中残留在淡水中的子遗生物, 多生长在清洁、温度偏低、环境比较稳定的水体中, 其分布区狭窄, 且具有一定的封闭性, 因此, 在对环境的长期适应过程中形成了一些珍稀特有种类(梅洪等, 2004)。淡水红藻是植物进化中的一个重要类群, 串珠藻目(Batrachospermales)是其中最主要的代表类群(施之新, 2006)。

核酮糖 1,5 二磷酸羧化酶/加氧酶(Rubisco)是植物中含量最丰富的蛋白质, 约占植物可溶性蛋白含量的 50%以上(Kapralov *et al*, 2007), 这也是对其相当低的催化效率的补偿(Spreitzer *et al*, 2002)。该酶在植物光合作用中具有双重功能, 一是催化核酮糖-1,5 二磷酸(RUBP)的羧化反应, CO₂ 与 RUBP 共价结合生成

不稳定的六碳化合物, 分裂出 2 分子 3-磷酸甘油酸, 二是在光呼吸中催化 RUBP 的加氧反应, O₂ 与 RUBP 反应生成 1 分子 3-磷酸甘油酸和 1 分子 2-磷酸乙醇酸(Portis, 1992; Gutteridge *et al*, 1995)。全酶由 8 个大亚基和 8 个小亚基组成, 其活性中心位于 Rubisco 大亚基, 特别是大亚基的 C 末端区域在功能上作用重要, 由叶绿体 *rbcL* 基因编码(Hartman *et al*, 1994)。以往的一些研究表明, 不同物种 Rubisco 的催化效率有差异(Jordan *et al*, 1981), 这种差异的原因主要是自然环境的选择压力, 而不是系统发生(Galmés *et al*, 2005)。

分析具有重要功能的蛋白质的适应性进化有助于深刻了解基因的变异和蛋白结构与功能的改变及物种的进化史(Nei *et al*, 2000)。在适应性进化的分析研究中, 可以通过度量核苷酸序列的非同义替换率

* 国家自然科学基金项目, 31670208 号, 31370239 号。巩超彦, 硕士研究生, E-mail: gongchaoyan1@163.com

通讯作者: 谢树莲, 博士生导师, 教授, E-mail: xiesl@sxu.edu.cn

收稿日期: 2016-12-14, 收修改稿日期: 2017-02-08

(dN)和同义替换率(dS)的比值(ω)来判断自然选择对氨基酸位点的选择压力(Yang, 2006)。同义替换指不导致氨基酸变异的核苷酸突变,一般发生在密码子的第三位,非同义替换指导致氨基酸变异的核苷酸突变,一般发生在密码子的第一、二位。当 $dN=dS$, 即 $\omega=1$, 表示非同义突变和同义突变的固定速率相同, 则认为选择对适合度没有影响; 当 $dN<dS$, 即 $\omega<1$, 认为非同义突变有害, 那么进化选择就会降低这些突变的固定速率; 当 $dN>dS$, 即 $\omega>1$, 认为发生的非同义突变有利于选择, 那么它们会以比同义突变更快的速率被固定。所以, $\omega>1$ 的显著性可被当做蛋白质发生适应性进化的证据。Nielsen 等(1998)、Yang 等(2000b)和 Wong 等(2004)针对编码蛋白质的 DNA 序列, 先后建立、改进了允许不同氨基酸位点取不同 ω 值的一系列似然进化模型。其中, Yang(2007)开发了 PAML(phylogenetic analysis by maximum likelihood)软件, 该软件是一个用最大似然法来对 DNA 和蛋白质序列进行系统发育分析的综合程序包, 其中的 Codeml 程序可以估算编码蛋白质的 DNA 序列的同义和非同义替代率并发现 DNA 序列中的正选择位点, 即对 DNA 序列进行适应性进

化分析。经由这些模型验证存在正向选择的位点后, 可进一步通过贝叶斯途径将发生正向选择的位点加以鉴定。

鉴于 Rubisco 大亚基的重要功能, 本文对串珠藻目的 *rbcL* 基因进行了适应性进化分析, 目的在于探讨自然选择对串珠藻目植物 Rubisco 大亚基的选择作用, 即在其进化过程中 Rubisco 大亚基是否经历过正选择及正选择发生的氨基酸。

1 材料与方法

1.1 序列数据采集及系统发育树构建

本研究所用数据包括作者采集样本提取和从 GenBank 收集的 *rbcL* 基因共 39 条(表 1)。用 Clustal X 软件(Thompson *et al.*, 1997)进行序列对位排列, 经人工检查校对, 得到每条序列均由 381 个密码子组成的序列数据供系统发育分析。

利用 MEGA5.0(Tamura, 2004)分析序列特征, 运行 Modeltest 软件(Posada *et al.*, 2004), 筛选核苷酸最优进化模型。采用最大似然法(ML)(Rannala *et al.*, 1996)运行 PhyML3.0(Guindon *et al.*, 2010)构建系统发育树。

表 1 用于本研究的种类及 *rbcL* 基因 GenBank 登录号
Tab.1 Species and the GenBank accession numbers of *rbcL* gene sequence used in the study

种类	GenBank 登录号
<i>Bangia atropurpurea</i>	DQ408155, DQ408162, KM363768, KM363769,
<i>Batrachospermum gelatinosum</i>	AF029141, KJ825967, KJ825969, KJ825970, KM077030, KM077034
<i>Batrachospermum helminthosum</i>	AB114646, AF244116, AF244118, AF244119, AF244120, KM593807, KJ825956
<i>Kumanoa capensis</i>	JX504698
<i>Kumanoa curvata</i>	JN590012
<i>Kumanoa faroensis</i>	JN590001
<i>Kumanoa gracillima</i>	JN590013
<i>Kumanoa gudjewga</i>	JN590002
<i>Kumanoa intorta</i>	JQ028695
<i>Sheathia arcuata</i>	JN086520, JN086521, JN086522, KM077039, KM077040, KM077041
<i>Sirodotia delicatula</i>	KC951855, KC951856, KC951861
<i>Sirodotia huillensis</i>	AF029157
<i>Sirodotia suecica</i>	AF029158
<i>Sirodotia tenuissima</i>	AF126420

1.2 统计分析

利用最大似然法得到 ML 树, 运行 PAML4.8 软件包(Yang, 2007)中的 Codeml 程序中的分支模型、位点模型、分支-位点模型进行分析。

在分支模型(Yang, 1998)中, 假定不同分支有不

同的 ω 值, 以检测各分支的选择压力。单比率(one ratio)模型是最简单的模型, 该模型假定进化树上所有分支的比率相同, 为 ω_0 。自由比率(free ratio)模型是最全面的模型, 该模型假定不同的分支有不同的比率。在这两种模型之间本研究还运用了二比率模型。

位点模型(Yang *et al.*, 2000a)是用于检验 *rbcL* 基因是否存在经受正选择($\omega > 1$)和负选择($\omega < 1$)的位点。在位点模型中, 允许不同位点有不同的选择压力, 而在系统树的不同分支之间则无差异。本研究采用了 3 对模型, 即 M1a(近中性)对 M2a(选择), M0(单一比值)对 M3(离散), M7(beta)对 M8(beta& ω)(Nielsen *et al.*, 1998; Yang *et al.*, 2000b), 后者均为备择假设, 前者是后者的零假设。对 3 对模型进行 LRT 检验(likelihood ratio testes), 是通过在相对的自由度(两模型参数之差)下, 二倍对数似然值之差的绝对值的 χ^2 检验来确定备择模型是否可靠(Yang *et al.*, 2000a)。

分支-位点模型(Hong *et al.*, 1997)用于检验指定分支中是否存在正选择位点。将系统发育树分为目标前景支和背景支(除了目标前景支之外的分支)。分支-位点模型的 LRT 检验, 根据文献(Yang, 2007)选择了 test2 模型, 将 MA 和无效模型(ω 设置为 1)进行比较。

1.3 三维结构建模

选取代表性物种胶串珠藻 *Batrachospermum gelatinosum* (串珠藻目, 串珠藻科, 串珠藻属的模式种, 登录号为 AF029141)的 *rbcL* 序列为参考序列, 翻译为氨基酸序列, 提交瑞士生物信息研究所(European Bioinformatics Institute: <http://www.isb-sib.cn/>) Swiss-model, 基于同源建模原理预测 Rubisco 大亚基的三维结构(周媛, 2011)。

2 结果与分析

2.1 系统发育分析

基于选取的 *rbcL* 基因序列, 通过 Modeltest 软件筛选得到核苷酸最优进化模型为 GTR+I+G(假设核苷酸位点替换具有可逆性, G 代表 gamma 分布的形状参数, I 代表不变位点的比例), 模型参数见(表 2)基于此模型构建了系统发育树, 使用的构树方法为最大似然法(ML)(图 1), 以 *Galdieria maxima* 和 *Cyanidium caldarium* 为外类群。从系统树可以看出, 除了外类群, 所有内类群聚集为 6 个分支。其中, 分支 A 为暗紫红毛菜(*Bangia atropurpurea*)的 4 个品系, 其他 5 个分支聚为 1 个大的分支, 包括所有串珠藻目的种类。分支 B 为 6 株胶串珠藻(*Batrachospermum gelatinosum*), 后验概率为 100%, 显示它们之间具有密切的亲缘关系。分支 C 为 7 株扁圆串珠藻(*Batrachospermum helminthosum*), 后验概率同样为 100%。分支 D 为熊野藻属(*Kumanoa*)的 6 个种, 后验概率为 99.2%, 与此前的有关报道一致(南芳茹等, 2015)。分支 E 为 4 种

共 6 株连珠藻属(*Sirodotia*)植物, 后验概率为 97.5%。分支 F 为 6 株弧形西斯藻(*Sheathia arcuata*), 后验概率为 100%。据此, 选定这 6 个分支供后面的适应性进化分析。

表 2 Modeltest 3.7 检验得到的 *rbcL* 基因最优进化模型参数

模型及参数	碱基频率	矩阵参数
GTR+I+G	freqA=0.3705	R[A-C]=4.0779
-lnL=7138.3921	freqC=0.1357	R[A-G]=4.6474
K=10	freqG=0.2091	R[A-T]=2.8539
(I)=0.4463	freqT=0.3247	R[C-G]=1.8602
(G)=0.9709		R[C-T]=24.2079
		R[G-T]=1.0000

注: GTR+I+G 为 *rbcL* 基因最优进化模型, GTR: general time reversible, 总体时间可逆模型; -lnL: 似然值; K: number of estimated parameters, 估算参数数目; I: invariable site, 保守位点的比率; R[A-C] 指碱基 A-C 替换的比率, 其余为对应碱基替换的比率。

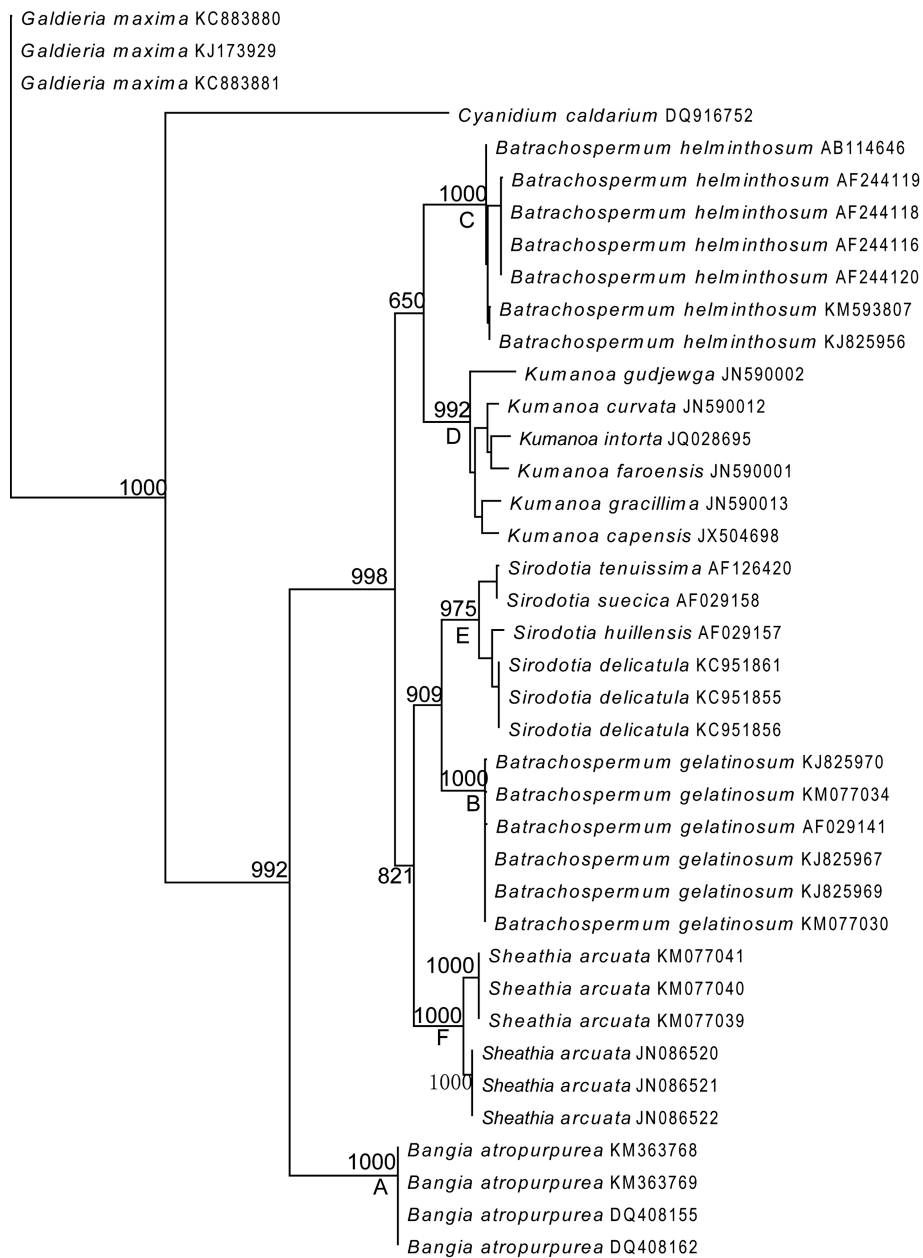
2.2 正选择位点鉴定

选择位点的鉴定结果见表 3 和表 4。

分支模型中自由比率模型显示, 分支 A、B、C、D、E、F 的 ω 值均小于 1, 意味着各分支均处于负选择下。但是, 经过与单比率模型进行 LRT 检验, 表明此模型并不可靠。二比率模型中, 指定分支 A、B、C、D、E、F 为前景枝, 其余为背景枝, 估算各前景枝的 ω 值。结果显示, 只有 F 分支的 $\omega > 1$, $\omega_F = 7.4382$, 显示这一分支中可能有正选择位点的存在, LRT 检验也支持这点。

在位点模型中, 模型 M2a(选择)、模型 M3(离散)、模型 M8(beta& ω)均允许 ω 值大于 1。经 LRT 检验, M2a 和 M8 模型均没有对应的可靠零假设, M3 模型则显著优于其对应的零假设。但是, 比较模型并非用来检测正选择位点, 而是检验各位点是否取不同的 ω 值。本研究显示各位点的 ω 值是不同的, 在位点模型下均没有检测出正选择位点, 表现出 *rbcL* 基因受到强烈的负选择。

分支-位点模型是研究指定分支相较于其它分支是否存在适应性进化。本文 6 个分支中, 结果显示只有分支 A(*Bangia atropurpurea*)的 277L 和 280L, 分支 F(*Sheathia arcuata*)的 350S 被鉴定为正选择位点。LRT 检验也肯定了其存在。另外, 分支 C(*Batrachospermum helminthosum*)检测出了 277L 为正选择位点, 但经 LRT 检验, 认为备择假设不可靠, 拒绝其存在。分支 B、D、F 均不比单比率模型更合适。



0.1

图 1 基于 *rbcL* 基因序列构建的系统发育树

Fig.1 Phylogenetic tree established based on the *rbcL* gene sequence
 注: 节点处的数字代表最大似然法步靴值, A、B、C、D、E、F 代表选定的分支

2.3 正选择位点定位

以胶串珠藻 *Batrachospermum gelatinosum* (串珠藻目, 串珠藻科, 串珠藻属的模式种, 登录号为 AF029141) 的 *rbcL* 序列为参考序列, 基于同源建模原理预测出 Rubisco 大亚基的参考三维结构(图 2)。通过对 PDB 数据库进行 Blast 搜索模板, 得到一种喜温红藻 *Geldieria partita* 的 Rubisco 三维结构(PDB ID: 1BWV)(Sugawara *et al.*, 1999), 相似度 83.64%, 符合

同源建模的可靠性要求。

将用于建模的序列 AF029141、程序分析序列 PAML(379 个氨基酸位点)和模板 Rubisco 大亚基对应氨基酸序列用 Bioedit 软件进行比对, 确定了被检测出的正选择位点的相对位置(图 2)。运用 Raswin 软件(Sayle *et al.*, 1995)将正选择位点标定在构建出的 Rubisco 大亚基的参考三维模型中。如图 3 所示, 被选出的正选择位点 277L、280L 位于 Rubisco 大亚基

表 3 各模型参数估计值和 log-likelihood 值
Tab.3 Parameter estimates and log-likelihood values for different models

模型	参数个数	似然值	参数估计值	正选择位点
单比率模型 M0	77	-6261.664351	$\omega_0=0.0105$	无
二比率模型 A	78	-6261.398718	$\omega_0=0.0106, \omega_1=0.0065$	无
二比率模型 B	78	-6260.583442	$\omega_0=0.0111, \omega_1=0.0001$	无
二比率模型 C	78	-6261.557625	$\omega_0=0.0107, \omega_1=0.0085$	无
分支模型				
二比率模型 D	78	-6270.500853	$\omega_0=0.0081, \omega_1=0.0151$	无
二比率模型 E	78	-6261.347013	$\omega_0=0.0107, \omega_1=0.0064$	无
二比率模型 F	78	-6275.018856	$\omega_0=0.0094, \omega_1=7.4382$	无
自由比率模型 G	151	-6229.885950	$\omega_A=0.0061, \omega_B=0.0001$ $\omega_C=0.0086, \omega_D=0.0032$ $\omega_E=0.0070, \omega_F=0.0035$	无
位点模型				
M1a (近中性)	78	-6249.352513	$P_0=0.98305, P_1=0.01695$ $\omega_0=0.00754, \omega_2=1.00000$	无应答
M2a (选择)	80	-6249.352543	$p_0=0.98305, p_1=0.01695, p_2=0.00000$ $\omega_0=0.00754, \omega_1=1.00000, \omega_2=33.58472$	无
M3 (离散)	81	-6182.312217	$p_0=0.78505, p_1=0.17351, p_2=0.04145$ $\omega_0=0.00199, \omega_1=0.02932, \omega_2=0.11838$	无
M7 (beta)	78	-6184.435382	$P=0.19651, q=14.33722$	无应答
M8 (beta& ω)	80	-6184.439166	$P_0=0.99999, P=0.19651, q=14.33714$ $P_1=0.00001, \omega=2.42001$	无
备择假设 A	80	-6235.903461	$P_{2a}=0.02585, P_{2b}=0.00043$ $\omega_{b1}=0.00744, \omega_{b2}=1.00000$ $\omega_{f1}=999.00000, \omega_{f2}=999.00000$	277L* 280L*
零假设 A0	79	-6239.710868	$P_{2a}=0.02284, P_{2b}=0.00048$ $\omega_{b1}=0.00716, \omega_{b2}=1.00000$ $\omega_{f1}=1.00000, \omega_{f2}=1.00000$	无应答
备择假设 B	80	-6249.352525	$P_{2a}=0, P_{2b}=0$ $\omega_{b1}=0.00754, \omega_{b2}=1.00000$ $\omega_{f1}=1.00000, \omega_{f2}=1.00000$	无
零假设 B0	79	-6249.352513	$P_{2a}=0, P_{2b}=0$ $\omega_{b1}=0.00754, \omega_{b2}=1.00000$ $\omega_{f1}=1.00000, \omega_{f2}=1.00000$	无应答
备择假设 C	80	-6249.236553	$P_{2a}=0.003, P_{2b}=0.00005$ $\omega_{b1}=0.00744, \omega_{b2}=1.00000$ $\omega_{f1}=1.21234, \omega_{f2}=1.21234$	277L*
零假设 C0	79	-6249.239136	$P_{2a}=0.00331, P_{2b}=0.00006$ $\omega_{b1}=0.00743, \omega_{b2}=1.00000$ $\omega_{f1}=1.00000, \omega_{f2}=1.00000$	无应答
分支-位点模型				
备择假设 D	80	-6249.352521	$P_{2a}=0, P_{2b}=0$ $\omega_{b1}=0.00754, \omega_{b2}=1.00000$ $\omega_{f1}=1.00000, \omega_{f2}=1.00000$	无
零假设 D0	79	-6249.352519	$P_{2a}=0, P_{2b}=0$ $\omega_{b1}=0.00754, \omega_{b2}=1.00000$ $\omega_{f1}=1.00000, \omega_{f2}=1.00000$	无应答
备择假设 E	80	-6249.352513	$P_{2a}=0, P_{2b}=0$ $\omega_{b1}=0.00754, \omega_{b2}=1.00000$ $\omega_{f1}=1.00000, \omega_{f2}=1.00000$	无
零假设 E0	79	-6249.352513	$P_{2a}=0, P_{2b}=0$ $\omega_{b1}=0.00754, \omega_{b2}=1.00000$ $\omega_{f1}=1.00000, \omega_{f2}=1.00000$	无应答
备择假设 F	80	-6243.211594	$P_{2a}=0.00279, P_{2b}=0.00005$ $\omega_{b1}=0.00717, \omega_{b2}=1.00000$ $\omega_{f1}=321.61389, \omega_{f2}=321.61389$	350S**
零假设 F0	79	-6246.193907	$P_{2a}=0.00429, P_{2b}=0.00008$ $\omega_{b1}=0.00716, \omega_{b2}=1.00000$ $\omega_{f1}=1.00000, \omega_{f2}=1.00000$	无应答

注: 在 95%和 99%后验概率下检测出的正选择位点分别用*和**标出。表中分支模型下二比率模型中 ω_1 为估算出的指定支的 *dN*、*dS* 比值, ω_0 为估算出的其余分支的 *dN*、*dS* 比值; 分支-位点模型中 P_{2a} 表示估算出的背景分支中 ω 高度保守, 但前景分支中 ω 显著大于 1 的位点的比率; ω_{b1} 表示估算出的背景分支中此类位点的 ω 值; ω_{f1} 表示估算出的前景分支中此类位点的 ω 值; P_{2b} 表示估算出的背景分支中处于中性进化或者选择压力较小, 但前景分支中 ω 显著大于 1 的位点的比率; ω_{b2} 表示估算出的背景分支中此类位点的 ω 值; ω_{f2} 表示估算出的前景分支中此类位点的 ω 值。

表4 LRT 检验统计量
Tab.4 Likelihood ratio statistics

模型比较	$2\Delta L$	自由度 df	概率 P
M0 vs.A	0.531266	1	0.4661
M0 vs.B	2.1618	1	0.1415
M0 vs.C	0.2135	1	0.644
分支模型 M0 vs.D	17.673**	1	<0.01
M0 vs.E	0.6347	1	0.4256
M0 vs.F	26.709**	1	<0.01
M0 vs.G	3.5568	74	0.8015
M0 vs.M3	158.70**	4	<0.01
位点模型 M1a vs.M2a	0.00006	2	1
M7 vs.M8	0.00757	2	1
a vs.a0	7.61**	1	<0.01
b vs.b0	0	1	1
c vs.c0	0.01	1	0.92
分支-位点模型 d vs.d0	0	1	1
e vs.e0	0	1	1
f vs.f0	5.96*	1	0.01

注: * 显著, $P<5\%$; ** 极显著, $P<1\%$

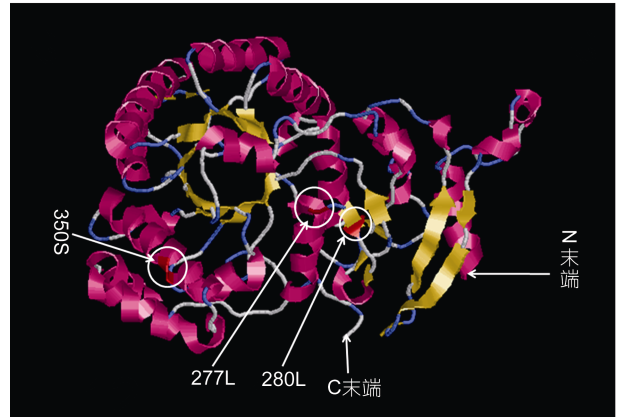


图2 胶串珠藻(AF029141)Rubisco 大亚基参考三维结构
Fig.2 Reference three - dimensional structure of Rubisco large subunits of *Batrachospermum gelatinosum* (AF029141)
注: 正选择位点在图中用白色圆圈标出

羧基末端保守的8个 α 螺旋和8个 β 片层构成的 α/β 桶状结构域中第7个 α 螺旋和第7个 β 片层之间的loop结构上,350S位于羧基末端邻近 α/β 桶状结构域的一个 α 螺旋上。

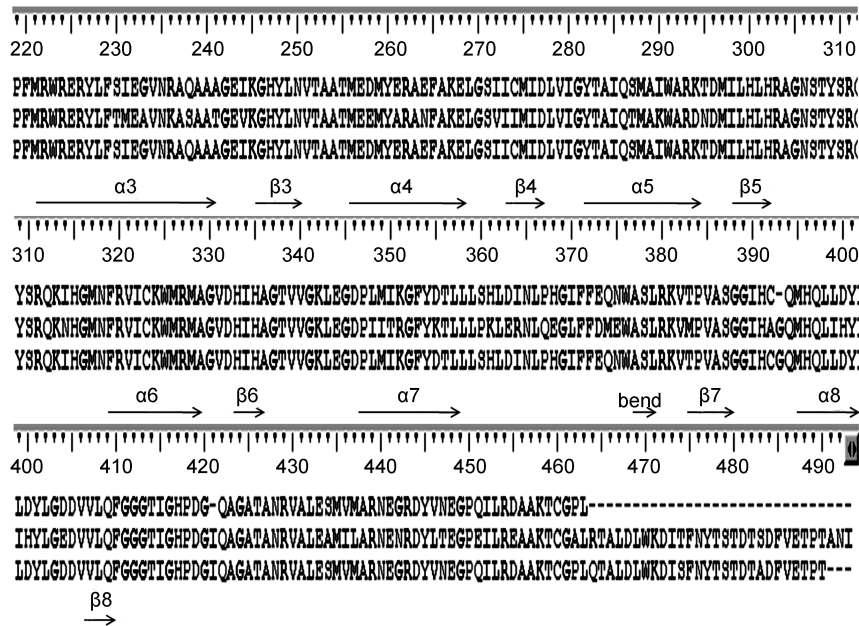


图3 Rubisco 大亚基氨基酸序列对位排列

Fig.3 Alignment of the amino acid sequence of Rubisco

注: 对应氨基酸下方箭头表示对应的 Rubisco 的二级结构。图中字母第一行代表 PAML 序列, 第二行代表模板序列, 第三行代表本文建模序列

3 讨论

被选出的正选择位点 277L、280L 位于 Rubisco 大亚基羧基末端活性中心区域, 350S 位于羧基末端邻近 α/β 桶状结构域的一个 α 螺旋上。已有研究表明

Rubisco 的一个大亚基和一个小亚基组成的二聚体是该酶最小的功能单位, 小亚基氨基端的功能结构域由 5 个 β 折叠和相邻的 2 个 α 螺旋组成(Farber *et al.*, 1990), 它们和大亚基羧基端的由 8 个 α 螺旋和 8 个 β 折叠组成功能结构域组成该酶的功能活性中心(Soper

et al, 1988; Knight *et al*, 1990)。这都提示这些位点可能对 Rubisco 功能有影响, 也为今后研究 Rubisco 的功能提供了优先选择位点。

本研究中分支-位点模型下, 分支 C 检测出了 277L 为正选择位点, 但经 LRT 检验, 备择假设不可靠, 拒绝了正选择位点的存在; 分支模型下检测出分支 D 可能有正选择位点, 但在分支-位点模型中没有检测出该支有正选择位点。Suzuki 等(2001, 2002)曾指出似然法分析可能得出假阳性结果, Wong 等(2004)基于模拟研究对似然法的可信度和统计效力也提出了质疑, Zhang(2004)曾怀疑 PAML 中分支-位点模型易产生假阳性(Yang *et al*, 2002)。之后, Zhang 等(2005)对该模型进行了改进, 很好地解决了假阳性问题。基于已有文献, 作者推测造成该结果的原因不是由于假阳性的问题, 而可能是由于选择压力放松造成的。同时, 这些结果也充分显示了基于 ω 比值检验 DNA 编码序列分子适应的可靠性和有效性。

位点模型下未能检测出正选择位点, 可能的原因是 *rbcL* 基因具有十分重要的功能, 需要保持相当的稳定性, 因此比较保守, 也有可能该基因适应性进化发生在较早期, 以至后来适应性进化信号被极其普遍的中性选择或者净化选择所淹没。相关研究也指出, 在藻类植物的 Rubisco 中几乎没有发现正选择位点的存在, 而在陆生植物中却广泛存在, 该酶结构上的差异不足以解释这个现象(Kapralov *et al*, 2007), 而可能的一个重要原因是它们生境的不同。藻类植物生长于稳定的水生环境, 具有特有的 CO_2 浓缩机制, 即碳酸氢盐泵, 该机制利用溶解在水中的碳酸氢盐并且抑制 Rubisco 的补氧活性, 使得 Rubisco 执行功能的气态环境相当稳定(Lambers *et al*, 2008)。

分支模型下检测出了分支 F 可能有正选择位点, 分支-位点模型也同时检测出了这一支确实有正选择位点。分支 A 在分支模型下显示没有正选择位点, 在分支-位点模型中却显示有两个正选择位点(277L 和 280L), 但它们的后验概率不高, 分别为 73.4%和 64%, 可能是分支模型未能检测出正选择的原因。另外, 分支 A 是暗紫红毛菜, 分支 F 是弧形西斯藻, 两种都是淡水红藻中分布范围较广泛的种, 说明其对环境有较好的适应, 这可能是在它们中检测出正选择位点的一个原因。

以往在研究酶的重要功能位点时往往采用生物化学方法对目标位点进行突变筛选以及突变体回复试验(Hong *et al*, 1997; Du *et al*, 2003)。例如, 有研究

鉴定出凤尾蕨科旱生植物 *rbcL* 基因的 365F 为正选择位点(周媛等, 2011), 该位点位于 βH 折叠区, 苯丙氨酸可能与 βH 折叠中的其他氨基酸形成氢键, 该区域的空间变化可能与植物对生境的适应有关(刘念等, 2010)。而 365F 位点与本研究鉴定出的位点相近。此外有实验表明, Rubisco α/β 桶状结构域中的 loop6 对该酶的催化作用、决定羧化作用和氧化作用的比值至关重要(Chen *et al*, 1989, 1991)。本研究鉴定出的 277L 和 280L, 虽然不位于 loop6 但位于 loop7, 由此可见, 本研究鉴定出的位点可能对酶的功能的研究产生积极意义。这些正选择位点仍有待更深入的研究, 今后的定向诱变研究和遗传相关试验也可以优先关注这些位点。

适应性突变的发生虽然是偶然的, 但突变位点的正选择作用则是自然环境的选择, 环境条件必然对酶的功能及活性产生影响。淡水红藻是海陆演变过程中残留在淡水中的子遗生物, 在世界各地的淡水中分布稀少且分布区狭窄。海洋生境同淡水生境差异巨大, 淡水红藻对生境的要求较高, 一般为弱光照下洁净、清冷、高溶氧的流动水体(施之新, 2006)。因此, 淡水红藻 *rbcL* 基因的进化压力来源可能就是这些环境因素, *rbcL* 的适应性进化使其更加适应淡水环境。

4 结论

本文利用 PAML4.8 软件, 运行位点模型、分支模型及分支-位点模型, 对选取淡水红藻类群的 *rbcL* 基因进行适应性进化分析, 得到如下结论:

(1) 通过最大似然法构建的系统发育树显示, 所有内类群聚集为 6 个分支, 其中, 分支 A 为暗紫红毛菜, 分支 B 为胶串珠藻, 分支 C 为扁圆串珠藻, 后验概率同样为 100%, 分支 D 为熊野藻属, 分支 E 为连珠藻属, 分支 F 为弧形西斯藻;

(2) 分支-位点模型中, 在 3 个分支中分别鉴定出 350S、277L 和 280L 为正选择位点;

(3) 在构建出的 Rubisco 大亚基的参考三维模型中, 277L 和 280L 位于 Rubisco 大亚基羧基末端保守的 8 个 α 螺旋和 8 个 β 片层构成的 α/β 桶状结构域中第 7 个 α 螺旋和第 7 个 β 片层之间的 loop 结构上, 350S 位于羧基末端邻近 α/β 桶状结构域的一个 α 螺旋上。

(4) 研究结果显示了基于 ω 比值检验基因适应性进化的准确性和有效性, 同时也揭示了串珠藻目植物 *rbcL* 基因确实发生了适应性进化, 对串珠藻目适应特殊生存环境产生了有益的作用。

参 考 文 献

- 刘念, 王庆彪, 陈婕等, 2010. 麻黄属 *rbcL* 基因的适应性进化检测与结构建模. 科学通报, 2010, 55(14): 1341—1346
- 周媛, 王博, 高磊等, 2011. 凤尾蕨科旱生蕨类 *rbcL* 基因的适应性进化和共进化分析. 植物科学学报, 29(4): 409—416
- 南芳茹, 冯佳, 谢树莲, 2015. 基于叶绿体 *psaA* 和 *psbA* 基因的中国熊野属植物系统发育分析. 水生生物学报, 39(1): 155—163
- 施之新, 2006. 中国淡水藻志(第十三卷)——红藻门和褐藻门. 北京: 科学出版社
- 梅洪, 赵先富, 郭斌等, 2004. 中国淡水藻类生物多样性研究进展. 生态科学, 22(4): 356—359
- Chen Z X, Spreitzer R J, 1989. Chloroplast intragenic suppression enhances the low CO₂/O₂ specificity of mutant ribulose-bisphosphate carboxylase/oxygenase. The Journal of Biological Chemistry, 264(6): 3051—3053
- Chen Z X, Yu W Z, Lee J H *et al*, 1991. Complementing amino acid substitutions within loop 6 of the α/β -barrel active site influence the carbon CO₂/O₂ specificity of chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase. Biochemistry, 30(36): 8846—8850
- Du Y, Peddi S R, Spreitzer R J, 2003. Assessment of structural and functional divergence far from the large subunit active site of ribulose-1, 5-bisphosphate carboxylase/oxygenase. Journal of Biological Chemistry, 278(49): 49401—49405
- Farber G K, Petsko G A, 1990. The evolution of α/β barrel enzymes. Trends in Biochemical Sciences, 15(6): 228—234
- Galmés J, Flexas J, Keys A J *et al*, 2005. Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. Plant, Cell & Environment, 28(5): 571—579
- Guindon S, Dufayard J F, Lefort V *et al*, 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology, 59(3): 307—321
- Gutteridge S, Gatenby A A, 1995. Rubisco synthesis, assembly, mechanism, and regulation. The Plant Cell, 7(7): 809—819
- Hartman F C, Harpel M R, 1994. Structure, function, regulation, and assembly of D-ribulose-1, 5-bisphosphate carboxylase/oxygenase. Annual Review of Biochemistry, 63(1): 197—232
- Hong S, Spreitzer R J, 1997. Complementing substitutions at the bottom of the barrel influence catalysis and stability of ribulose-bisphosphate carboxylase/oxygenase. Journal of Biological Chemistry, 272(17): 11114—11117
- Jordan D B, Ogren W L, 1981. Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. Nature, 291(5815): 513—515
- Kapralov M V, Filatov D A, 2007. Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evolutionary Biology, 7: 73
- Knight S, Andersson I, Brändén C I, 1990. Crystallographic analysis of ribulose 1, 5-bisphosphate carboxylase from spinach at 2.4 Å resolution: subunit interactions and active site. Journal of Molecular Biology, 215(1): 113—160
- Lambers H, Chapin III F S, Pons T L, 2008. Plant Physiological Ecology. 2nd ed. New York: Springer
- Nei M, Kumar S, 2000. Molecular Evolution and Phylogenetics. Oxford: Oxford University Press
- Nielsen R, Yang Z, 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, 148(3): 929—936
- Portis A R, 1992. Regulation of ribulose 1, 5-bisphosphate carboxylase/oxygenase activity. Annual Review of Plant Physiology and Plant Molecular Biology, 43: 415—437
- Posada D, Buckley T R, 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology, 53(5): 793—808
- Rannala B, Yang Z H, 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. Journal of Molecular Evolution, 43(3): 304—311
- Sayle R A, Milner-White E J, 1995. RASMOL: biomolecular graphics for all. Trends in Biochemical Sciences, 20(9): 374—376
- Soper T S, Mural R J, Larimer F W *et al*, 1988. Essentiality of Lys-329 of ribulose-1, 5-bisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum* as demonstrated by site-directed mutagenesis. Protein Engineering, Design and Selection, 2(1): 39—44
- Spreitzer R J, Salvucci M E, 2002. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. Annual Review of Plant Biology, 53(1): 449—475
- Sugawara H, Yamamoto H, Shibata N *et al*, 1999. Crystal structure of carboxylase reaction-oriented ribulose 1, 5-bisphosphate carboxylase/oxygenase from a thermophilic red alga, *Galdieria partita*. Journal of Biological Chemistry, 274(22): 15655—15661
- Suzuki Y, Nei M, 2001. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. Molecular Biology and Evolution, 18(12): 2179—2185
- Suzuki Y, Nei M, 2002. Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. Molecular Biology and Evolution, 19(11): 1865—1869
- Tamura K, Peterson D, Peterson N *et al*, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28(10): 2731—2739
- Thompson J D, Gibson T J, Plewniak F *et al*, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 25(24): 4876—4882
- Wong W S W, Yang Z, Goldman N, *et al*, 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics, 2004, 168(2): 1041—1051

- Yang Z, 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15(5): 568—573
- Yang Z H, Swanson W J, Vacquier V D, 2000a. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Molecular Biology and Evolution*, 17(10): 1446—1455
- Yang Z H, Bielawski J P, 2000b. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, 15(12): 496—503
- Yang Z H, Nielsen R, 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, 19(6): 908—917
- Yang Z H, 2006. *Computational Molecular Evolution*. Oxford: Oxford University Press
- Yang Z H, 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8): 1586—1591
- Zhang J Z, 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. *Molecular Biology and Evolution*, 21(7): 1332—1339
- Zhang J Z, Nielsen R, Yang Z H, 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, 22(12): 2472—2479

ADAPTIVE EVOLUTIONARY ANALYSIS ON *rbcL* GENE OF BATRACHOSPERMALES

GONG Chao-Yan, NAN Fang-Ru, FENG Jia, LÜ Jun-Ping, LIU Qi, XIE Shu-Lian
(School of Life Science, Shanxi University, Taiyuan 030006, China)

Abstract Freshwater red algae are an important lineage in algal evolution. Batrachospermales is an order with the most species. Ribulose-1,5 bisphosphate carboxylase/oxygenase (Rubisco) is an enzyme responsible for both carbon assimilation and photorespiration, and plays an important role in plant photosynthesis. The Rubisco large subunit containing active sites for catalyzing CO₂ fixation, is encoded by chloroplast *rbcL* gene. To explore the molecular mechanism of algal adaption to special living environment, 39 *rbcL* genes of Batrachospermales and other relative taxa were selected to analyze the adaptive evolution. Using software PAML4.8, the site model, branch model, and branch-site model were tested. The results show that (1) in the phylogenetic tree constructed by the maximum likelihood method, all the internal groups cluster into six branches, i.e., Cluster A is *Bangia atropurpurea*, B is *Batrachospermum gelatinosum*, C is *B. helminthosum*, D is the genus *Kumanoa*, E is the genus *Sirodotia*, and F is *Sheathia arcuata*; (2) 350S, 277L, and 280L are identified as positive selection sites on three branches in the branch-site model; (3) in the reference three-dimensional model of Rubisco large subunit, 277L and 280L are located in a loop between the 7th α helix and the 7th β sheet of the conserved α/β -barrel domain that consists of 8 α -helices and 8 β -sheets at the C-terminus, and 350S is located on an α helix in the C-terminus adjacent to the α/β -barrel domain. Therefore, the results demonstrated the accuracy and effectiveness of examining the adaptive evolution of gene based on ω ratio, and revealed that the *rbcL* gene has undergone adaptive evolution, which has been beneficial to Batrachospermales for the adaption to the special living environment.

Key words Batrachospermales; *rbcL* gene; adaptive evolution