

念珠藻类(蓝藻) *psbA* 基因的进化分析

魏爱丽¹, 杨谢¹, 王捷¹, 巩超彦², 王清华¹, 李艳晖¹

(1.太原师范学院 生物系,山西 晋中 030619;2.山西中医药大学 实验管理中心,山西 晋中 030619)

摘要:蓝藻 *psbA* 基因家族编码不同形式的 D1 蛋白,该蛋白是光系统 II 反应中心的重要组成部分.以 39 条念珠藻属(*Nostoc*)及与其同源性较高的 *psbA* 基因序列为研究对象,构建最大似然树进行系统发育分析,然后运行 PAML4.9 软件,使用分支模型、位点模型和分支-位点模型估测氨基酸位点 ω 值,进一步探讨 *psbA* 基因所受到的选择压力.结果表明:(1)系统发育树呈现出内类群中念珠藻分为 2 个大分支.(2)在分支-位点模型和位点模型下检测出 13S,42V,75S,152R 和 255K 为统计学上显著的正选择位点,绝大多数为负选择位点.揭示了念珠藻 *psbA* 基因所经历的正选择可能在其适应极端环境中起着重要作用.

关键词:念珠藻;*psbA* 基因;正向选择;负向选择

中图分类号:Q941.2

文献标志码:A

念珠藻是由胶质外鞘包被的丝状固氮蓝藻,自身能合成多种特殊的化合物,有助于其在各种高度竞争的生态位中生存下来,尤其可在极度干旱、高温和寒冷等环境中生活^[1-2].该类蓝藻含有具免疫调节作用和抗氧化活性的三萜类氨基酸、肽类、多不饱和脂肪酸和色素等天然产物^[1,3].在医学中,还有很多用于不同治疗目的的新型药物化合物,而高度多样化的生态系统中生长的念珠藻成为了化学合成的替代品,在食品、医药和化妆品等行业具有极高的推广价值,经济效益可观^[3-4].

植物类囊体膜蛋白 D1 是光系统 II 反应中心的核心.在光照条件下,光合作用中水的分解涉及多步氧化和光化学反应,这些反应主要导致蓝藻、绿藻和高等植物光系统 II 反应中心 D1 蛋白的光致光损伤,促进新的 D1 蛋白来替换损失蛋白以维持其正常的生理功能.因此,D1 蛋白比其他光系统 II 蛋白有更高的周转率^[5].而蓝藻 *psbA* 基因家族编码不同形式的 D1 蛋白,每种亚型的不同调节确保其适应不断变化的环境条件^[6].另一方面,由于 *psbA* 基因保守性较强,植物分类学家把它作为一个重要的分子标记,尤其是属以上水平有较好的区分度^[7].但是对其功能研究还较少,尤其是藻类 *psbA* 基因是否存在正选择很少受到关注.

自然选择不断推动物种进化,维持着生态系统的物种多样性,而物种在进化过程中核心蛋白往往要承受巨大的选择压力.判断选择压力的分子进化参数用 ω 来表示,也就是核苷酸非同义替换率(dN)与同义替换率(dS)的比值.计算所得结果中,若 $\omega = 1$,即 $dN = dS$,表明选择不影响适合度,属于中性进化; $\omega < 1$,即 $dN < dS$,表明非同义突变降低了固定速率; $\omega > 1$,即 $dN > dS$,暗示蛋白质可能发生了适应性进化^[8-9].目前已有大量文献报道了蕨类植物多个科属 *rbcL*^[10-12], *psaA*^[13], *psbA*^[14], *psbD*^[15], *rps4*^[16], *rps12*^[17], *rpoC1*^[18], *matK*^[19] 等基因的适应性进化研究成果,结果表明 *matK*, *rpoC1*, *psaA*, *rbcL* 和 *rps12* 等 5 个基因中均存在正选择位点.森林等^[20]对裸子植物 4 亚纲 53 个代表种的 *psbA* 基因展开了适应性进化研究,检测到了 3 个正选择位点.目前对于低等植物基因的适应性进化研究仅集中到几种藻类上, MILLER 等^[21]分析了特殊生境下生长的聚球藻(蓝藻门)*rbcL* 基因的适应性进化情况,发现正选择位点 389 位氨基酸与多个

收稿日期:2021-10-29;修回日期:2021-12-04.

基金项目:国家自然科学基金(32000167;51709197);山西省“1331 工程”服务流域生态治理产业创新学科集群建设项目;山西省基础研究计划面上项目(自由探索类)(202103021224302);山西省高等学校教学改革项目(J2020273).

作者简介:魏爱丽(1969—),女,山西太谷人,太原师范学院教授,博士,研究方向为植物发育与生态毒理学, E-mail: 67514292@qq.com.

通信作者:王捷, E-mail: nostoc@126.com.

蛋白活性位点相关。巩超彦等^[22]对淡水红藻中的 *psaA* 基因(主要编码光系统 I 反应中心蛋白)适应性进化研究中发现处于强烈的负选择作用下的位点占比较大。LIANG 等^[23]发现雨生红球藻(*Haematococcus pluvialis*)PDS 基因正选择位点所占比例为 7.9%。此外,王捷等^[24-25]也对念珠藻固氮作用相关基因 *hetR* 和 *nifH* 进行了适应性进化分析,在 *nifH* 基因中发现 14 个位点经历了正选择,而在 *hetR* 基因中仅发现 1 个位点经历了正选择。生物分子水平上正选择和负选择位点的准确剖析,不仅能深度挖掘基因结构和功能变异,还可进一步揭示其进化历程。

念珠藻可在多种逆境条件下生存,而逆境条件下的选择压力往往可以导致植物的光合作用相关基因发生进化^[10-15],因此本研究主要分析环境胁迫下念珠藻 *psbA* 基因在进化过程中的选择压力,为探索其精确的功能位点及进化历程提供一定的证据。

1 材料和方法

1.1 测序及系统树构建

本研究共选取了 39 条蓝藻 *psbA* 基因序列,其中测定了 31 株形态多样化的念珠藻基因序列(31 株念珠藻样品以及 PCR 扩增条件的详细信息参考文献^[24-26]),正、反向测序工作均由北京华大基因科技有限公司完成,其余 8 条同源性较高的基因序列是从 GenBank 数据库中下载获得。使用 Clustal X version 2.0^[27] 软件将以上 39 条序列进行对齐排列,并逐一仔细检查校正。运用 MEGA7.0^[28] 软件中的 K2P(Kimura2-parameter)模型来分析每一条序列,然后通过 Modeltest(Version3.7)^[29] 软件计算并获得最优进化模型,最后使用 PhyML3.0^[30] 软件构建最大似然树。

1.2 选择压力分析

基于念珠藻属 *psbA* 基因构建最大似然树的本源数据,打开 PAML4.9 软件包,分别运行位点模型、分支模型和分支-位点模型等 3 种模型中 Codeml 模块,以此结果来分析 *psbA* 基因在其演化历程中是否经受过正选择作用^[13-14]。分支模型中采用两种极端模型来检测各分支的选择压力,这两种模型为简单的单比率模型和最全面的自由比率模型。这两种极端模型之间还存在一些其他模型,本研究还使用了二比率模型进行了分析。位点模型主要通过 ω 值来估测 *psbA* 基因是否经历了正选择,如果 $\omega > 1$,说明可能存在正选择位点,如果 $\omega < 1$,则不存在正选择位点。主要分析 3 对模型(M1a 和 M2a, M0 和 M3, M7 和 M8),每对模型两两之间比较,然后依 χ^2 分布进行 LRT(likelihood ratio test)检验并确定位点是否发生过正选择。分支-位点模型可检验指定分支中位点的进化方式,经查阅文献发现,选择 test2 算法检验前景分支(所指定的分支)中的正选择位点更具有准确性^[15-16]。

1.3 三维结构建模

以普通念珠藻(*Nostoc commune* HK-02)PsbA 氨基酸序列(GenBank 登录号:AP018326)为参比序列,提交瑞士生物信息研究所网站^[24],基于同源建模原理来预测 PsbA 蛋白的三维结构。然后用 BioEdit^[31-32] 软件比对建模序列(AP018326)和鉴定出正选择位点的氨基酸序列,直观展现正选择位点。最后运用 Raswin^[33] 软件将正选择位点标记于晶体结构中。

2 实验结果

2.1 系统发育分析

选取无类囊体蓝藻 *Gloeobacter violaceus* PCC 7421(GenBank 登录号:BA000045)为外类群,利用最大似然法重建了念珠藻 *psbA* 基因的谱系关系。结果显示,内类群中的念珠藻明显聚集到 2 个大的分支 E 和 F 上,本研究所测序的大部分念珠藻菌株聚到 E 分支中,而 E 分支又包含 A, B, C 和 D 4 个较小的分支,其中 A 分支包含 3 株新疆古尔班通古特沙漠采集到的念珠藻(XA013, XA017 和 XA019),吉林省靖宇县采集到的未定种念珠藻(CHAB2813),我国山西忻州市神池县山坡、宁夏回族自治区土地和湖北省鹤峰市稻田等地分别采集到的普通念珠藻(*Nostoc commune*)、发状念珠藻(*Nostoc flagelliforme*)和拟球状念珠藻(*Nostoc sphaeroides*),以及中国淡水藻种库保存的来源于美国的 1 株普通念珠藻(FACHB261)和中国科学院植物

研究所的 1 株发状念珠藻(FACHB838).B 分支包含 C 和 D 两个小的分支,其中 C 分支包含 7 株新疆古尔班通古特沙漠采集到的念珠藻(XA009,XA010,XA011,XA012,XA014,XA015 和 XA016)、湖南省永州市采集到的普通念珠藻(CHAB2806)和中国淡水藻种库保存的来源于美国的 1 株灰色念珠藻(*Nostoc muscorum*,FACHB316).D 分支包含分别采集于浙江省宁波市东钱湖湖边、云南省洱源海西海湖边以及中国科学院水生生物研究所苏铁共生和土体上未定种的 4 株念珠藻、湖北省武昌区水稻田中采集到的 1 株沼泽念珠藻(*Nostoc paludosum*,FACHB88).F 分支包含分别来自于我国以及日本的 3 株普通念珠藻(CHAB2803,CHAB2807 和 CHAB2820),还有来自于我国内蒙古自治区和甘肃省的 2 株发状念珠藻.根据所计算出的树型结构,分析以上 6 个分支中位点的进化方式.

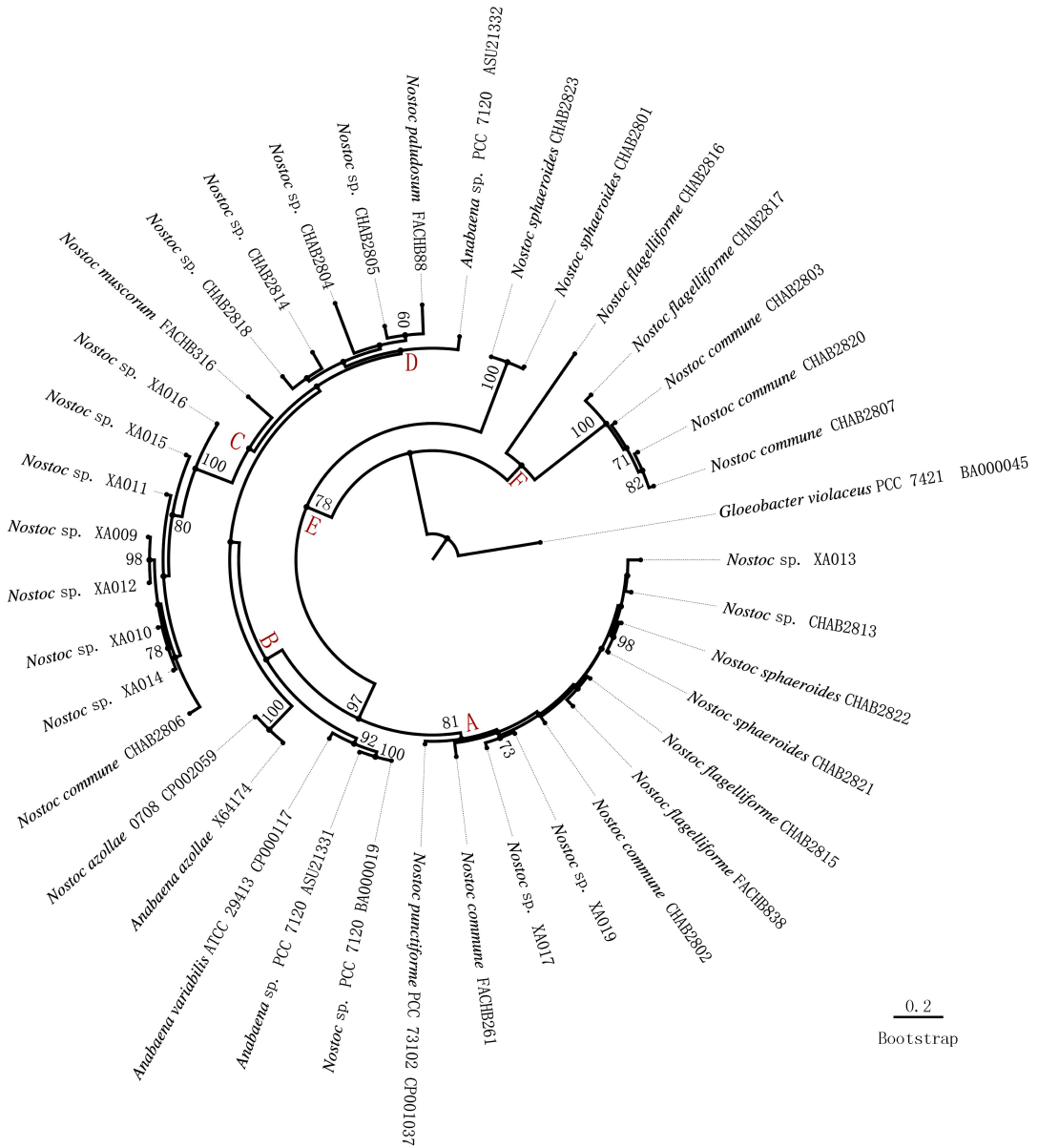


图1 基于psbA基因序列构建的ML分子系统树

Fig.1 Phylogenetic tree (ML) established based on the psbA gene sequence

2.2 选择压力分析

三种模型下所计算出的估计值、似然值和正选择位点等结果详见附表 1 和附表 2.分支模型中计算了单比率模型、二比率模型和自由比率模型,其中单比率模型结果显示似然值为-7 156.439 702, ω 估计值为 0.075 81.自由比率模型结果显示 A,B,C,D 和 E 分支 ω 值小于 1,仅有 F 分支的 ω 值大于 1(ω 值为 999.000 00),说明该分支位点可能经历了正选择,需对 F 分支进行进一步的分支-位点模型检测验证.二比率

模型指定的前景支 F 的 ω 估计值大于 1,数值同样为 999.000 00,表明这个分支可能存在正选择位点,前景支 A,B,C,D,E 的 ω 估计值均小于 1,表明这几个分支可能均不存在正选择位点.LRT 检验发现(见附表 2),B,C,E 和 F 分支的后验概率均大于 0.05,说明上述结果不可靠.位点模型下,鉴定出 255K 为具有统计学意义的正向选择位点,在两个模型 M2a 和 M8 中都有检出,后验概率均大于 99%.分支-位点模型中,在前景支 B 检测出位点 75S(后验概率为 100%)经历了正选择,前景支 E 检测出位点 94S(后验概率为 99.7%)经历了正向选择,分支 F 检测出 13S(后验概率为 96.8%),42V(后验概率为 95.5%),75S(后验概率为 99.7%)和 152R(后验概率为 99.7%)等 5 个正选择位点,其余分支均未检出正选择位点.经似然比检验发现(表 2),A,C,D 和 E 分支的后验概率均大于 0.05,拒绝检出正选择位点.

2.3 正选择位点定位

经同 PDB 数据库中蛋白序列比对发现,与一株嗜热蓝藻(*Thermosynechococcus vulcanus*, PDB ID: 6JLK)的 *PsbA* 蛋白结构相似度高达 88.95%.基于此来预测念珠藻 *PsbA* 蛋白的三维结构.经对位排列,图 2 直观展示了不同藻株的氨基酸序列所检测出的正选择位点.图 3 中标示出 13S,42V,75S,152R 和 255K 等 5 个正选择位点.

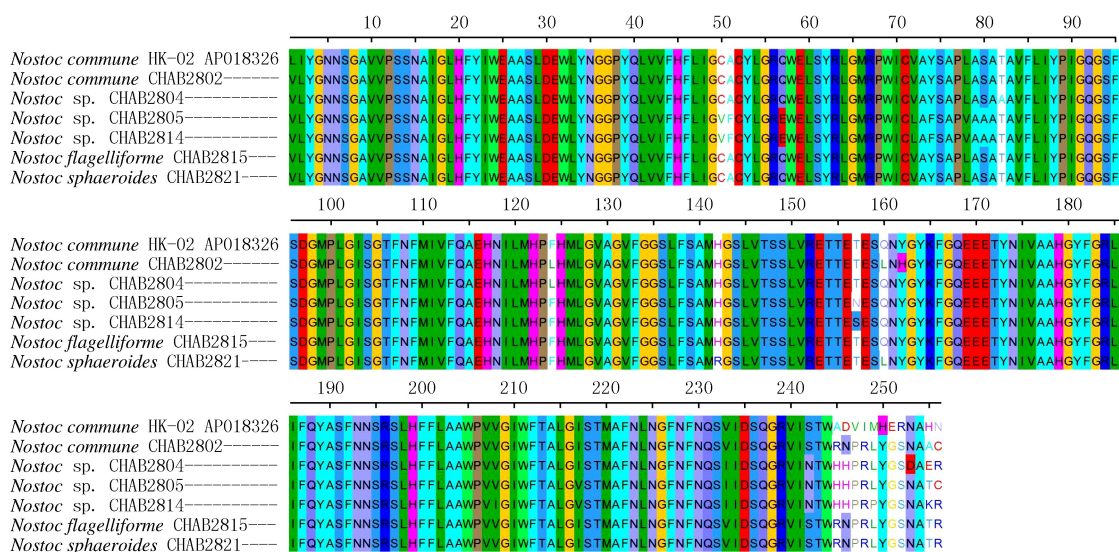


图2 *PsbA*氨基酸序列对位排列

Fig.2 Alignment of the amino acid sequence of *PsbA*

3 讨论

蓝藻 *psbA* 基因家族编码不同形式的 D1 蛋白,每种亚型不同调节确保其适应不断变化的环境条件^[6].当念珠藻的某些基因经受强烈的正选择压力时,则暗示该藻已适应所生存环境中的一些关键影响因子,并预示着在该进化时间点可能激发出部分蛋白的新功能,而强烈的负选择压力会使该藻的蛋白维系其固有的重要功能且变得高度保守^[12-16].正选择位点的筛选将有助于预测基因功能区或解决遗传育种等生物技术难题.

由于 *psbA* 基因保守性较强,植物分类学家把它作为一个重要的分子标记用于植物系统发育分析中,尤其是属以上水平有较好的区分度^[7].所构建的 ML 系统发育树显示,内类群中的念珠藻明显聚集到 2 个大的分支 E 和 F 上,本研究所测序的大部分念珠藻藻株聚到 E 分支中,而 E 分支又包含 A,B,C 和 D 4 个较小的分支,其中 A 分支包含 3 株新疆古尔班通古特沙漠采集到的念珠藻(XA013,XA017 和 XA019)、吉林省靖宇县采集到的未定种念珠藻(CHAB2813),我国山西忻州市神池县山坡、宁夏回族自治区土地和湖北省鹤峰市稻田等地分别采集到的普通念珠藻、发状念珠藻和拟球状念珠藻,以及中国淡水藻种库保存的来源于美国的 1 株普通念珠藻(FACHB261)和中国科学院植物研究所的 1 株发状念珠藻(FACHB838).普通念珠藻、发状念珠藻和拟球状念珠藻野外形态分别呈现片状、头发丝状和圆球状等,但是在实验室纯化培养条件下都长成圆球状,难以辨认.现系统发育树中的 3 种近似藻类聚为一支,且有较高的支持率.猜测这 3 种念珠藻是否

属于同一种念珠藻,可进行二代测序或结合其他相关生理生化手段分析.B分支包含C和D两个小的分支,其中C分支包含7株新疆古尔班通古特沙漠采集到的念珠藻(XA009,XA010,XA011,XA012,XA014,XA015和XA016)、湖南省永州市采集到的普通念珠藻(CHAB2806)和中国淡水藻种库保存的来源于美国的1株灰色念珠藻(FACHB316).D分支包含分别采集于浙江省宁波市东钱湖湖边、云南省洱海海西海湖边以及中国科学院水生生物研究所苏铁共生和土体上未定种的四株念珠藻、湖北省武昌区水稻田中采集到的1株沼泽念珠藻(FACHB88).这几株未定种的念珠藻也可能是新种,可以进一步进行多基因分析.F分支包含分别来自于我国以及日本的3株普通念珠藻(CHAB2803,CHAB2807和CHAB2820),还有来自于我国内蒙古自治区和甘肃省的2株发状念珠藻.可以看出以上6分支中不同的系统发育位置分布着同一物种,不同国家、省份采集到的藻种聚到不同的分支上,表明利用 $psbA$ 单一基因的系统发育分析具有一定的局限性,未能呈现出念珠藻的进化关系.根据系统发育树树形结构,选定以上6个分支分析念珠藻 $psbA$ 基因是否经历了正选择.

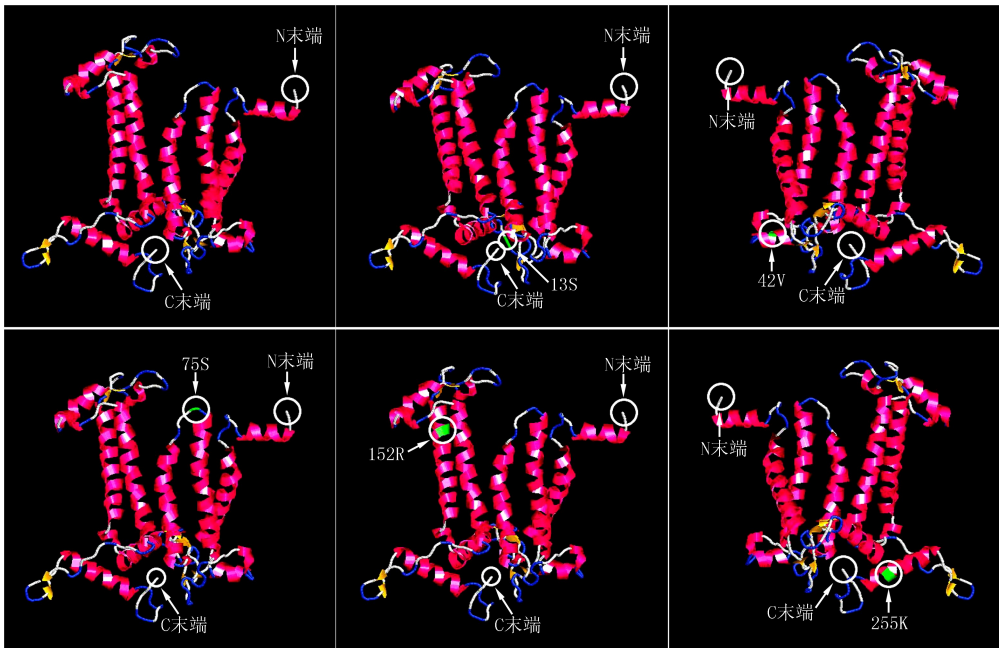


图3 普通念珠藻(AP018326)PsbA蛋白参考三维结构

Fig. 3 Reference three-dimensional structure of PsbA in *Nostoc commune* (AP018326)

本研究分支模型中,自由比率模型结果显示A,B,C,D和E分支 ω 值小于1,仅有F分支的 ω 值大于1(ω 值为999.000 00),说明该分支位点可能经历了正选择,需对F分支进行进一步的分支-位点模型检测验证.二比率模型指定的前景支F的 ω 估计值大于1,数值同样为999.000 00,表明这个分支可能存在正选择位点.前景支A,B,C,D,E的 ω 估计值均小于1,表明这几个分支可能均不存在正选择位点.LRT检验发现(见附表2),B,C,E和F分支的后验概率均大于0.05,说明上述结果不可靠.位点模型下,鉴定出255K为具有统计学意义的正向选择位点,在两个模型M2a和M8中都有检出,后验概率均大于99%.巩超彦等^[33]和韩雨昕等^[34]分别对不同红藻 $rbcL$ 基因适应性进化研究中,均并未能在位点模型下发现正选择位点.分支-位点模型中,在前景支B检测出位点75S(后验概率为100%)经历了正选择,前景支E检测出位点94S(后验概率为99.7%)经历了正向选择,分支F检测出13S(后验概率为96.8%),42V(后验概率为95.5%),75S(后验概率为99.7%)和152R(后验概率为99.7%)等5个正选择位点,其余分支均未检出正选择位点.经似然比检验发现(附表2),A,C,D和E分支的后验概率均大于0.05,拒绝检出正选择位点.

本文采用分支模型、位点模型以及分支-位点模型探讨了环境胁迫下的念珠藻 $psbA$ 基因在进化过程中的选择压力,发现13S,42V,75S,152R和255K等5个位点经受了正向选择压力,绝大多数为负选择位点.大量文献表明高等植物在进化过程中大多数基因处于正向选择压力之下^[10,20].相比低等植物,蓝藻在地球上出现已有35亿年之久,分布广泛,种类繁多,有更广阔的研究前景.目前对蓝藻甚至更多藻类基因的选择压力研究较少,而藻类重要基因选择压力分析的新发现可推断其分子水平的进化.

附 录

附表见电子版(DOI:10.16366/j.cnki.1000-2367.2023.01.015).

参 考 文 献

- [1] NOWRUZI B, HAGHIGHAT S, FAHIMI H, et al. *Nostoc* cyanobacteria species; a new and rich source of novel bioactive compounds with pharmaceutical potential[J]. *Journal of Pharmaceutical Health Services Research*, 2018, 9(1): 5-12.
- [2] 胡鸿钧, 魏印心. 中国淡水藻类: 系统、分类及生态[M]. 北京: 科学出版社, 2006.
- [3] NOWRUZI B, AHMADI M A. A new report of N fixation by two species of cyanobacteria[J]. *Iranian Journal of Science & Technology, Transaction A*, 2008, 2: 1-4.
- [4] RIPPKA R, STANIER R Y, DERUELLES J, et al. Generic assignments, strain histories and properties of pure cultures of cyanobacteria [J]. *Microbiology*, 1979, 111(1): 1-61.
- [5] SRIVASTAVA A, SHUKLA P. Tightening the screws on *PsbA* in cyanobacteria[J]. *Trends in Genetics*, 2021, 37(3): 211-215.
- [6] SHERIDAN K J, DUNCAN E J, EATON-RYE J J, et al. The diversity and distribution of D1 proteins in cyanobacteria[J]. *Photosynthesis Research*, 2020, 145(2): 111-128.
- [7] 李春筱, 刘婷婷, 刘玉珊, 等. 云南高原异龙湖噬藻体 *psbA* 基因遗传多样性研究[J]. *生态与农村环境学报*, 2019, 35(10): 1322-1329.
- LI C X, LIU T T, LIU Y S, et al. Phylogenetic diversity of cyanophage *psbA* gene in the plateau eutrophic Yilong Lake in Yunnan Province[J]. *Journal of Ecology and Rural Environment*, 2019, 35(10): 1322-1329.
- [8] YANG Z H. *Computational molecular evolution*[M]. Oxford: Oxford University Press, 2006.
- [9] YANG Z H. PAML 4: phylogenetic analysis by maximum likelihood[J]. *Molecular Biology and Evolution*, 2007, 24(8): 1586-1591.
- [10] 陈洁, 张丽君, 王艇. 蕨类植物 *rbcL* 基因正选择和负选择位点的鉴定[J]. *西北植物学报*, 2009, 29(12): 2391-2400.
- CHEN J, ZHANG L J, WANG T. Diversifying and purifying selection sites in *rbcL* gene of fern species[J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2009, 29(12): 2391-2400.
- [11] 森林, 苏应娟, 张冰, 等. 凤尾蕨科植物 *rbcL* 基因的适应性进化分析[J]. *热带亚热带植物学报*, 2010, 18(1): 1-8.
- SEN L, SU Y J, ZHANG B, et al. Adaptive evolution of the *rbcL* gene in pteridaceous ferns[J]. *Journal of Tropical and Subtropical Botany*, 2010, 18(1): 1-8.
- [12] 王春波, 郭治友. 膜蕨科植物 *rbcL* 基因的适应性进化和共进化分析[J]. *广西植物*, 2017, 37(2): 145-152.
- WANG C B, GUO Z Y. Adaptive evolution and coevolution of the *rbcL* gene in Hymenophyllaceae[J]. *Guihaia*, 2017, 37(2): 145-152.
- [13] 吴筱娉, 森林, 陈楠, 等. 蕨类植物 *psaA* 基因的分子进化研究[J]. *植物科学学报*, 2017, 35(2): 177-185.
- WU X P, SEN L, CHEN N, et al. Study on the molecular evolution of the *psaA* gene from ferns[J]. *Plant Science Journal*, 2017, 35(2): 177-185.
- [14] 陈晓霞, 苏应娟, 王艇. 细鳞苔科 *psbA* 基因的适应性进化分析[J]. *西北植物学报*, 2010, 30(8): 1534-1544.
- CHEN X X, SU Y J, WANG T. Adaptive evolution analysis of the *psbA* gene in Lejeuneaceae[J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2010, 30(8): 1534-1544.
- [15] 许可, 王博, 苏应娟, 等. 蕨类植物 *psbD* 基因的适应性进化和共进化分析[J]. *植物科学学报*, 2013, 31(5): 429-438.
- XU K, WANG B, SU Y J, et al. Molecular evolution of *psbD* gene in ferns; selection pressure and co-evolutionary analysis[J]. *Plant Science Journal*, 2013, 31(5): 429-438.
- [16] 张丽君, 陈洁, 王艇. 蕨类植物叶绿体 *rps4* 基因的适应性进化分析[J]. *植物研究*, 2010, 30(1): 42-50.
- ZHANG L J, CHEN J, WANG T. Adaptive evolution in the chloroplast gene *rps4* in ferns[J]. *Bulletin of Botanical Research*, 2010, 30(1): 42-50.
- [17] 平晶耀, 祝铭, 苏应娟, 等. 蕨类植物叶绿体 *rps12* 基因的分子进化研究[J]. *植物科学学报*, 2020, 38(1): 1-9.
- PING J Y, ZHU M, SU Y J, et al. Molecular evolution of chloroplast gene *rps12* in ferns[J]. *Plant Science Journal*, 2020, 38(1): 1-9.
- [18] 彭阳, 苏应娟, 王艇. 蕨类植物 *rpoC1* 内含子缺失及其分子进化速率[J]. *植物学报*, 2020, 55(3): 287-298.
- PENG Y, SU Y J, WANG T. Intron loss and molecular evolution rate of *rpoC1* in ferns[J]. *Chinese Bulletin of Botany*, 2020, 55(3): 287-298.
- [19] 熊哲铭, 高一波, 任慧莹, 等. 40 种蕨类植物 *matK* 基因的系统分类及分子进化研究[J]. *植物科学学报*, 2020, 38(1): 10-22.
- XIONG Z M, GAO Y B, REN H Y, et al. Analysis on the phylogenetic classification and molecular evolution of the *matK* gene in 40 fern species[J]. *Plant Science Journal*, 2020, 38(1): 10-22.
- [20] 森林, 余坤, 胡志刚, 等. 裸子植物 *psbA* 基因分子进化式样的研究[J]. *热带亚热带植物学报*, 2016, 24(2): 151-159.
- SEN L, YU K, HU Z G, et al. Molecular evolutionary patterns of the *psbA* gene in gymnosperms[J]. *Journal of Tropical and Subtropical Botany*, 2016, 24(2): 151-159.
- [21] MILLER S R. Evidence for the adaptive evolution of the carbon fixation gene *rbcL* during diversification in temperature tolerance of a clade of hot spring cyanobacteria[J]. *Molecular Ecology*, 2003, 12(5): 1237-1246.

- [22] 巩超彦,南芳茹,冯佳,等.串珠藻目植物 *psaA* 基因的适应性进化及共进分析[J].山西大学学报(自然科学版),2019,42(3):662-672.
GONG C Y,NAN F R,FENG J,et al.Adaptive evolutionary and co-evolutionary analysis on *psaA* gene of batrachospermales[J].Journal of Shanxi University(Natural Science Edition),2019,42(3):662-672.
- [23] LIANG C W,ZHAO F Q,MENG C X,et al.Molecular cloning,characterization and evolutionary analysis of phytoene desaturase(*PDS*) gene from *Haematococcus pluvialis*[J].World Journal of Microbiology and Biotechnology,2006,22(1):59-64.
- [24] 王捷,徐羽,石瑛,等.念珠藻属植物 *nifH* 基因的适应性进化分析[J].西北植物学报,2019,39(11):1961-1969.
WANG J,XU Y,SHI Y,et al.Adaptive evolutionary analysis on *nifH* gene of Nostoc[J].Acta Botanica Boreali-Occidentalia Sinica,2019,39(11):1961-1969.
- [25] 王捷,魏爱丽,石瑛,等.念珠藻属植物 *hetR* 基因的适应性进化分析[J].植物科学学报,2020,38(1):23-31.
WANG J,WEI A L,SHI Y,et al.Adaptive evolutionary analysis of *hetR* gene in Nostoc[J].Plant Science Journal,2020,38(1):23-31.
- [26] LARKIN M A,BLACKSHIELDS G,BROWN N P,et al.Clustal W and clustal X version 2.0[J].Bioinformatics,2007,23(21):2947-2948.
- [27] KUMAR S,STECHER G,TAMURA K.MEGA7:molecular evolutionary genetics analysis version 7.0 for bigger datasets[J].Molecular Biology and Evolution,2016,33(7):1870-1874.
- [28] POSADA D,BUCKLEY T R.Model selection and model averaging in phylogenetics:advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests[J].Systematic Biology,2004,53(5):793-808.
- [29] GUINDON S,DUFAYARD J F,LEFORT V,et al.New algorithms and methods to estimate maximum-likelihood phylogenies:assessing the performance of PhyML 3.0[J].Systematic Biology,2010,59(3):307-321.
- [30] CHENNA R,SUGAWARA H,KOIKE T,et al.Multiple sequence alignment with the Clustal series of programs[J].Nucleic Acids Research,2003,31(13):3497-3500.
- [31] HALL T A.BioEdit:a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT[J].Nucleic Acids Symposium Series,1999,41:95-98.
- [32] SAYLE R A,et al.RASMOL:biomolecular graphics for all[J].Trends in Biochemical Sciences,1995,20(9):374-376.
- [33] 巩超彦,南芳茹,冯佳,等.串珠藻目植物 *rbcL* 基因的适应性进化分析[J].海洋与湖沼,2017,48(3):527-535.
GONG C Y,NAN F R,FENG J,et al.Adaptive evolutionary analysis on *rbcL* gene of batrachospermales[J].Oceanologia et Limnologia Sinica,2017,48(3):527-535.
- [34] 韩雨昕,南芳茹,巩超彦,等.弯枝藻属 *rbcL* 基因的适应性进化分析[J].热带亚热带植物学报,2019,27(1):36-44.
HAN Y X,NAN F R,GONG C Y,et al.Adaptive evolutionary analysis of the *rbcL* gene from *Compsopogon*(rhodophyta)[J].Journal of Tropical and Subtropical Botany,2019,27(1):36-44.

Evolutionary analysis on *psbA* gene of *Nostoc*-like(Cyanobacteria)

Wei Aili¹, Yang Xie¹, Wang Jie¹, Gong Chaoyan², Wang Qinghua¹, Li Yanhui¹

(1. Department of Biology, Taiyuan Normal University, Jinzhong 030619, China; 2. Laboratory Management Center, Shanxi University of Chinese Medicine, Jinzhong 030619, China)

Abstract: The *psbA* gene family in cyanobacteria encodes different forms of the D1 protein which is part of the Photosystem II reaction center. In this study, 39 *psbA* gene sequences with high homology of *Nostoc* were used to construct a maximum likelihood tree for phylogenetic analysis. Then using PAML4.9 software, the branch model, site model and branch-site models were estimated ω values, to explore the selective pressure of the *psbA* gene. Our results indicate that: (1) the phylogenetic tree showing the inner group could be divided into two large branches of *Nostoc*. (2) five significant positively selected sites such as 13S, 42V, 75S, 152R and 255K were detected in branch-site models and site model, indicating that most of the sites were under negative selection pressure. Therefore, the results showed that positive selection of *psbA* might have played an important role in the adaption of *Nostoc* to the extreme environments.

Keywords: *Nostoc*; *psbA* gene; positive selection; purifying selection

[责任编辑 刘洋 杨浦]

附表 I 各模型参数估计值和对数似然值

Attached tab. I Parameter estimates and log-likelihood values for different models

| 模型 | 数量 | 似然值 | 估计值 | 正选择位点 | |
|---------|--------------------------|-----|----------------|---|---|
| 分支模型 | 单比率模型 M0 | 77 | -7 156.439 702 | $\omega = 0.075\ 81$ | 无 |
| | 二比率模型 A | 78 | -7 156.437 454 | $\omega_0 = 0.075\ 76, \omega_1 = 0.079\ 25$ | 无 |
| | 二比率模型 B | 78 | -7 156.035 993 | $\omega_0 = 0.076\ 49, \omega_1 = 0.042\ 20$ | 无 |
| | 二比率模型 C | 78 | -7 156.301 649 | $\omega_0 = 0.075\ 43, \omega_1 = 0.110\ 24$ | 无 |
| | 二比率模型 D | 78 | -7 156.437 011 | $\omega_0 = 0.075\ 85, \omega_1 = 0.070\ 86$ | 无 |
| | 二比率模型 E | 78 | -7 156.107 373 | $\omega_0 = 0.075\ 12, \omega_1 = 0.136\ 04$ | 无 |
| | 二比率模型 F | 78 | -7 154.590 676 | $\omega_0 = 0.074\ 17, \omega_1 = 999.000\ 00$ | 无 |
| | 自由比率模型 H | 151 | -7 094.214 857 | $\omega_A = 0.104\ 062, \omega_B = 0.039\ 588\ 7, \omega_C = 0.110\ 542, \omega_D = 0.077\ 621\ 3, \omega_E = 0.120\ 435, \omega_F = 999.000\ 00$ | 无 |
| 位点模型 | M1a: 近中性 | 78 | -6 982.621 498 | $P_0 = 0.917\ 11, P_1 = 0.082\ 89, \omega_0 = 0.042\ 35, \omega_1 = 1.000\ 00$ | 无应答 |
| | M2a: 选择 | 80 | -6 975.978 165 | $P_0 = 0.916\ 83, P_1 = 0.079\ 26, P_2 = 0.003\ 91, \omega_0 = 0.042\ 76, \omega_1 = 1.000\ 00, \omega_2 = 4.628\ 29$ | 255K 1.000 * * |
| | M3: 离散 | 81 | -6 941.903 208 | $P_0 = 0.462\ 90, P_1 = 0.452\ 39, P_2 = 0.084\ 71, \omega_0 = 0.000\ 00, \omega_1 = 0.079\ 57, \omega_2 = 0.618\ 34$ | 无 |
| | M7(β) | 78 | -6 952.802 586 | $P = 0.216\ 43, q = 1.787\ 36$ | 无应答 |
| | M8(β & ω) | 80 | -6 937.617 455 | $P_0 = 0.996\ 08, P_1 = 0.239\ 24, q = 2.306\ 22, P_1 = 0.003\ 92, \omega = 4.176\ 44$ | 255K 1.000 * * |
| 分支-位点模型 | 备择假设 a | 80 | -6 982.621 826 | $P_{2a} = 0.000\ 01, P_{2b} = 0.000\ 00, \omega_{b1} = 0.042\ 35, \omega_{b2} = 1.000\ 00, \omega_{f1} = 1.000\ 00, \omega_{f2} = 1.000\ 00$ | 208I 0.572 |
| | 零假设 a0 | 79 | -6 982.621 564 | $P_{2a} = 0.000\ 00, P_{2b} = 0.000\ 00, \omega_{b1} = 0.042\ 35, \omega_{b2} = 1.000\ 00, \omega_{f1} = 1.000\ 00, \omega_{f2} = 1.000\ 00$ | 无应答 |
| | 备择假设 b | 80 | -6 975.747 374 | $P_{2a} = 0.007\ 15, P_{2b} = 0.000\ 64, \omega_{b1} = 0.041\ 83, \omega_{b2} = 1.000\ 00, \omega_{f1} = 999.000\ 00, \omega_{f2} = 999.000\ 00$ | 75S 1.000 * * 217V 0.839 |
| | 零假设 b0 | 79 | -6 981.110 664 | $P_{2a} = 0.028\ 11, P_{2b} = 0.002\ 50, \omega_{b1} = 0.041\ 89, \omega_{b2} = 1.000\ 00, \omega_{f1} = 1.000\ 00, \omega_{f2} = 1.000\ 00$ | 无应答 |
| | 备择假设 c | 80 | -6 982.621 577 | $P_{2a} = 0.000\ 00, P_{2b} = 0.000\ 00, \omega_{b1} = 0.042\ 35, \omega_{b2} = 1.000\ 00, \omega_{f1} = 1.000\ 00, \omega_{f2} = 1.000\ 00$ | 无 |
| | 零假设 c0 | 79 | -6 982.621 644 | $P_{2a} = 0.000\ 00, P_{2b} = 0.000\ 00, \omega_{b1} = 0.042\ 35, \omega_{b2} = 1.000\ 00, \omega_{f1} = 1.000\ 00, \omega_{f2} = 1.000\ 00$ | 无应答 |
| | 备择假设 d | 80 | -6 982.621 633 | $P_{2a} = 0.000\ 01, P_{2b} = 0.000\ 00, \omega_{b1} = 0.042\ 35, \omega_{b2} = 1.000\ 00, \omega_{f1} = 1.000\ 00, \omega_{f2} = 1.000\ 00$ | 无 |
| | 零假设 d0 | 79 | -6 982.621 589 | $P_{2a} = 0.000\ 00, P_{2b} = 0.000\ 00, \omega_{b1} = 0.042\ 35, \omega_{b2} = 1.000\ 00, \omega_{f1} = 1.000\ 00, \omega_{f2} = 1.000\ 00$ | 无应答 |
| | 备择假设 e | 80 | -6 978.593 911 | $P_{2a} = 0.014\ 13, P_{2b} = 0.001\ 28, \omega_{b1} = 0.041\ 21, \omega_{b2} = 1.000\ 00, \omega_{f1} = 14.417\ 29, \omega_{f2} = 14.417\ 29$ | 29L 0.544 78V 0.568 94S 0.997 * * |
| | 零假设 e0 | 79 | -6 979.891 571 | $P_{2a} = 0.068\ 72, P_{2b} = 0.006\ 23, \omega_{b1} = 0.041\ 00, \omega_{b2} = 1.000\ 00, \omega_{f1} = 1.000\ 00, \omega_{f2} = 1.000\ 00$ | 无应答 |

(续表)

| 模型 | 数量 | 似然值 | 估计值 | 正选择位点 |
|--------|----|----------------|--|---|
| 备择假设 f | 80 | -6 976.107 738 | $P_{2a}=0.038 91, P_{2b}=0.003 47,$ $\omega_{b1}=0.040 89, \omega_{b2}=1.000 00,$ $\omega_{f1}=999.000 00, \omega_{f2}=999.000 00$ | 13S 0.968 * 42V 0.955 * 44F 0.848 75S 0.997 ** 152R 0.986 * 208I 0.827 |
| 零假设 f0 | 79 | -6 978.211 613 | $P_{2a}=0.261 71, P_{2b}=0.023 32, \omega_{b1}=0.040 94,$ $\omega_{b2}=1.000 00, \omega_{f1}=1.000 00, \omega_{f2}=1.000 00$ | 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.

附表 II LRT 检验统计量
Attached tab. II Likelihood ratio statistics

| 模型比较 | 2ΔL | 自由度 | P | |
|---------|-------------|-------------|----|------------|
| 分支模型 | M0 vs. A | 0.004 496 | 1 | 0.946 |
| | M0 vs. B | 0.807 418 | 1 | 0.369 |
| | M0 vs. C | 0.276 106 | 1 | 0.599 |
| | M0 vs. D | 0.005 382 | 1 | 0.942 |
| | M0 vs. E | 0.664 658 | 1 | 0.415 |
| | M0 vs. F | 3.698 052 | 1 | 0.054 |
| | M0 vs. H | 124.449 690 | 74 | 0.000 2 ** |
| 位点模型 | M0 vs. M3 | 429.072 988 | 4 | 0.000 ** |
| | M1a vs. M2a | 13.286 666 | 2 | 0.001 ** |
| | M7 vs. M8 | 30.370 262 | 2 | 0.000 ** |
| 分支-位点模型 | a vs. a0 | 0.000 524 | 1 | 0.982 |
| | b vs. b0 | 10.726 580 | 1 | 0.001 ** |
| | c vs. c0 | 0.000 134 | 1 | 0.991 |
| | d vs. d0 | 0.000 088 | 1 | 0.992 |
| | e vs. e0 | 2.595 320 | 1 | 0.107 |
| | f vs. f0 | 4.207 750 | 1 | 0.040 |

注: ** 表示 $P < 0.01$.