

DOI: 10.11913/PSJ.2095-0837.2020.10023

王捷, 魏爱丽, 石瑛, 李艳晖, 韩雨昕, 王中杰. 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 AL, Shi Y, Li YH, Han YX, Wang ZJ. Adaptive evolutionary analysis of *hetR* gene in *Nostoc* [J]. Plant Science Journal, 2020, 38(1): 23~31

# 念珠藻属植物 *hetR* 基因的适应性进化分析

王捷<sup>1</sup>, 魏爱丽<sup>1</sup>, 石瑛<sup>1\*</sup>, 李艳晖<sup>1</sup>, 韩雨昕<sup>2</sup>, 王中杰<sup>3</sup>

(1. 太原师范学院生物系, 山西晋中 030619; 2. 山西大学生命科学学院, 太原 030006;

3. 中国科学院武汉植物园, 中国科学院植物种质创新与特色农业重点实验室, 武汉 430074)

**摘要:** 以念珠藻属(*Nostoc*)及其近缘类群 *hetR* 基因的 51 条序列为研究对象, 对 *hetR* 基因的编码蛋白进行生物信息学分析和系统发育分析, 并使用分支模型、位点模型和分支-位点模型进行该基因位点的适应性进化研究。系统发育分析结果显示, 51 条 *hetR* 基因蛋白序列可分为 4 个大分支。适应性进化分析结果表明, 在 3 种进化模型中, 大多数分支及藻株都没有检测到统计学上具有显著性的正选择位点, 说明检测的位点大多处于负选择压力下。但在普通念珠藻(*Nostoc commune*, CHAB2802)中检测到正选择位点(126T), 提示念珠藻属植物 *hetR* 基因发生了适应性改变。

**关键词:** 念珠藻属; *hetR* 基因; 适应性进化

中图分类号: Q949.2

文献标识码: A

文章编号: 2095-0837(2020)01-0023-09

## Adaptive evolutionary analysis of *hetR* gene in *Nostoc*

Wang Jie<sup>1</sup>, Wei Ai-Li<sup>1</sup>, Shi Ying<sup>1\*</sup>, Li Yan-Hui<sup>1</sup>, Han Yu-Xin<sup>2</sup>, Wang Zhong-Jie<sup>3</sup>

(1. Department of Biology, Taiyuan Normal University, Jinzhong, Shanxi 030619, China; 2. School of Life Science, Shanxi University, Taiyuan 030006, China; 3. Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China)

**Abstract:** Fifty-one sequences of the *hetR* gene from *Nostoc* and related species of cyanobacteria were studied. The coding proteins of *hetR* were analyzed by bioinformatics and phylogenetic analysis. The adaptive evolutionary characteristics of *hetR* were studied using the branch, site, and branch-site models. Phylogenetic analysis showed that the inner group could be divided into four large branches. Based on the adaptive evolutionary analysis, no significant positive selection sites were detected in other branches and algae strains of the three evolutionary models, indicating that most of the sites were under negative selection pressure. However, significant positively selected sites (126T) were detected in *Nostoc commune* (CHAB2802), indicating that the *hetR* gene has undergone adaptive evolution in *Nostoc*.

**Key words:** *Nostoc*; *hetR* gene; Adaptive evolution

念珠藻属(*Nostoc*)是蓝藻门或蓝细菌中的重要类群。该属物种是一类不分枝、具有异形胞且能进行光合作用的丝状固氮蓝藻。由于其属于古老且

多样化的原核生物群, 同时能忍受多种极端环境, 因此在全世界的各种生境中分布广泛<sup>[1, 2]</sup>。作为天然产品的丰富生产者, 念珠藻属具有非常高的经济

收稿日期: 2019-05-24, 退修日期: 2019-08-10。

基金项目: 国家自然科学基金项目(51709197); 山西省“1331 工程”重点创新团队建设计划(TD201718); 山西省应用基础研究计划(201801D221282); 山西省高等学校科技创新项目(201802102)。

This work was supported by grants from the National Natural Science Foundation of China (51709197), Fund for Shanxi “1331 Project” Key Innovative Research Team (TD201718), Applied Basic Research Program of Shanxi (201801D221282), and Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (201802102).

作者简介: 王捷(1984-), 男, 博士, 讲师, 主要研究方向为藻类植物系统分类(E-mail: nostoc@126.com)。

\* 通讯作者(Author for correspondence, E-mail: lucia0812@163.com)

价值和应用价值。念珠藻属的生物活性和实用性在药物开发和生物技术应用中得到广泛认可，同时还具有其他生物医学方向巨大的利用价值。因此，念珠藻属被视为新药物的潜在开发来源<sup>[3]</sup>。

异形胞(heterocyst)是由一些丝状蓝藻的营养细胞分化出来的一种特殊细胞。在固氮酶的催化下，异形胞可把大气中的N<sub>2</sub>转化为藻种可利用的氮素化合物<sup>[4]</sup>，具有固氮作用。念珠藻属的hetR基因是操纵异形胞分化的主要基因，在异形胞分化阶段发挥着极其重要的功能。同时，该基因对于蓝藻厚壁孢子(akinete)的形成也是必需的<sup>[5]</sup>。因此HetR在蓝藻细胞分化阶段是一个重要的调控因子，而不仅仅局限于异形胞的分化阶段。

物种在进化过程中发挥重要功能的蛋白质可能会受到较强烈的选择压力。遗传水平上适应性进化的剖析及关键功能位点和氨基酸结构的识别，可为探索植物进化的奥秘提供证据，还能更加深刻地了解基因结构和功能的变异<sup>[6]</sup>。在适应性进化研究中，蛋白质水平上的选择压力可以通过核苷酸的非同义替换率(*dN*)与同义替换率(*dS*)的比值( $\omega$ )来度量，并进一步判断选择压力在非同义替换的稳固进程中起遏制或促进作用。一些核苷酸的替代不会引起氨基酸变化，称为同义替换；而更多情况下密码子的替代会引起氨基酸变化，称为非同义替换。当 $\omega = 1$ ，即 *dN* = *dS* 时，表明选择对适合度无影响；当 $\omega < 1$ ，即 *dN* < *dS* 时，表明非同义突变有害；当 $\omega > 1$ ，即 *dN* > *dS* 时，表明非同义突变有利。目前对于基因的适应性进化研究多集中在高等植物中，如森林等<sup>[7]</sup>和周媛等<sup>[8]</sup>对凤尾蕨科(Pteridaceae)植物*rbcL*基因进行了适应性进化分析，并在氨基酸水平上检测到多个正选择位点，其中一些位点对核酮糖-1,5-二磷酸羧化酶/加氧酶(Rubisco)行使正常功能起关键作用。陈晓霞等<sup>[9]</sup>、许可等<sup>[10]</sup>和吴筱婷等<sup>[11]</sup>对细鳞苔科(Lejeuneaceae)*psbA*基因、蕨类植物*psbD*和*psaA*基因进化适应性进行了分析，进一步检验了统计学上正选择位点的显著性，结果显示这些正选择位点的P值大于0.05，说明负选择对这些基因起核心作用。此外，巩超彦等<sup>[12]</sup>和韩雨昕等<sup>[13]</sup>分别对红藻进化中的重要类群，即串珠藻目(Batrachospermales)和弯枝藻属(*Compsopogon*)植物的*rbcL*基因进行了适应性进化研究，仅在串珠藻目植物中

检测到正选择位点。目前，对蓝藻中重要功能基因的适应性进化的研究报道较少。

本研究以念珠藻属植物及其近缘类群的hetR基因序列为对象，采用位点模型、分支模型和分支-位点模型对念珠藻属hetR基因进行适应性进化分析，探讨环境压力下念珠藻属植物 HetR 是否有正选择发生，以期为深入开展念珠藻属植物的分子进化研究奠定基础。

## 1 材料与方法

### 1.1 基因序列测定及系统发育树构建

本研究共测定了31株念珠藻属植物hetR基因的序列(表1)，将所测序列和GenBank数据库(<https://www.ncbi.nlm.nih.gov/>)中的基因序列进行对位排列。采用MEGA 6.06软件<sup>[14]</sup>中的Kimura2-parameter模型分析每条序列的特征，使用

表1 本研究所用的念珠藻藻株信息  
Table 1 Strains of *Nostoc* used in this study

编号 Number	藻株 Strain	样品采集地 Sample collection site
CHAB 2801	<i>Nostoc sphaeroides</i> Kütz	湖北省鹤峰市稻田
CHAB 2802	<i>Nostoc commune</i> Vauch	山西省忻州市神池县山坡
CHAB 2803	<i>Nostoc commune</i> Vauch	山西省忻州市偏关县山坡
CHAB 2804	<i>Nostoc</i> sp.	浙江省宁波市东钱湖湖边
CHAB 2805	<i>Nostoc</i> sp.	云南省大理州洱源县海西湖边
CHAB 2806	<i>Nostoc commune</i> Vauch	湖南省永州市
CHAB 2807	<i>Nostoc commune</i> Vauch	日本广岛大学校园土地
CHAB 2813	<i>Nostoc</i> sp.	吉林省白山市靖宇县
CHAB 2814	<i>Nostoc</i> sp.	苏铁共生念珠藻
CHAB 2815	<i>Nostoc flagelliforme</i> Born et Flah	宁夏回族自治区土地
CHAB 2816	<i>Nostoc flagelliforme</i> Born et Flah	内蒙古自治区土地
CHAB 2817	<i>Nostoc flagelliforme</i> Born et Flah	甘肃省
CHAB 2818	<i>Nostoc</i> sp.	中国科学院水生生物研究所院内
CHAB 2820	<i>Nostoc commune</i> Vauch	南京师范大学校园土地
CHAB 2821	<i>Nostoc sphaeroides</i> Kütz	湖北省鹤峰市稻田
CHAB 2822	<i>Nostoc sphaeroides</i> Kütz	湖北省鹤峰市稻田
CHAB 2823	<i>Nostoc sphaeroides</i> Kütz	菲律宾
FACHB 88	<i>Nostoc paludosum</i> Kütz	湖北省武汉市武昌区水稻田
FACHB 261	<i>Nostoc commune</i> Vauch	美国
	<i>Nostoc muscorum</i> C.	
FACHB 316	Agardh ex Bornet & Flahault	美国
FACHB 838	<i>Nostoc flagelliforme</i> Born et Flah	中国科学院植物研究所
XA009-XA019	<i>Nostoc</i> sp.	新疆古尔班通古特沙漠

PhyML 3.0 软件<sup>[15]</sup>构建最大似然树(ML)，最后通过 Modeltest(Version 3.7)软件<sup>[16]</sup>计算获得最适取代模型及相关参数。

## 1.2 适应性进化分析

以构建的最大似然树为本原数据，运行 PAML 4.9 软件包<sup>[17]</sup>中 codeml 模块中的 3 种模型(位点模型、分支模型和分支-位点模型)进行适应性进化分析<sup>[7-11]</sup>。

分支模型中，主要使用单比率模型、自由比率模型和二比率模型进行检测。位点模型中，主要检测 *hetR* 基因的正选择和负选择位点存在与否，如果  $\omega > 1$ ，说明存在正选择位点；如果  $\omega < 1$ ，则不存在正选择位点。采用的 3 对比较模型分别为：M1a 和 M2a、M0 和 M3、M7 和 M8。然后按照  $\chi^2$  分布对这 3 对模型进行 LRT (likelihood ratio test) 检验，通过比较模型间的差异显著性来检验正选择位点。分支-位点模型中，将系统发育树的支系分为前景支和背景支，仅允许前景支中出现正选择位点，并对其进行 LRT 检验。在 test 1 中将 Ma 和 M1a 进行似然比检验，在 test 2 中将 Ma 和无效模型( $\omega$  设置为 1)进行比较。经比较发现选择 test 2 算法更具有准确性。

## 1.3 三维结构建模

从 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>) 中选取代表物种普通念珠藻 (*Nostoc commune*)。以其 *hetR* 基因序列 (GenBank 登录号: AP018326) 为参比序列，翻译成氨基酸序列后提交瑞士生物信息研究所网站 (European bioinformatics institute, <http://www.isb-sib.cn/>)，利用在线软件 Swiss-model，基于同源建模原理来预测 HetR 蛋白的三维结构。

将用于建模的 AP018326 序列和鉴定出正选择位点的 HetR 蛋白对应的氨基酸序列放入 BioEdit 软件<sup>[18, 19]</sup>中进行比对，定位正选择位点。运用 Raswin 软件<sup>[20]</sup>将正选择位点标记在构建的 HetR 蛋白参考三维模型中。

## 2 结果与分析

### 2.1 系统发育分析

本研究用于 *hetR* 基因序列分析的蓝藻共有 51 株，选择集胞藻 (*Synechocystis* sp. PCC 6803) 为外类群。经 ModelTest 软件计算结果，采用念珠

藻 *hetR* 基因序列构建 ML 系统发育树的最优进化模型为 GTR + I + G。系统进化树分析结果显示(图 1)，本研究测定序列的念珠藻和中国科学院水生生物研究所淡水藻种库(以下简称：淡水藻种库)保存的 2 株念珠藻聚为 4 个大的分支 A~D。其中，A 分支包含 25 株本研究所测定的念珠藻(E、F、G 3 个小分支  $\omega$  值均大于 1)；B 分支包含 1 株与苏铁 (*Cycas*) 共生的念珠藻；C 分支包含 2 株淡水藻种库保存的念珠藻；D 分支包含 3 株未能定种的念珠藻。因此，选择含有本研究测定的和淡水藻种库保存的念珠藻所在的分支，并以这 6 个分支进一步进行适应性进化分析。

## 2.2 适应性进化分析

适应性进化分析中各模型选择位点的鉴定结果见表 2 和表 3。分支模型中，二比率模型指定 A~D 4 个分支为前景支，其他是背景支。前景支 C 的  $\omega$  估计值为 16.49504，表明此分支可能存在正选择位点；前景支 A、B、D 的  $\omega$  估计值均小于 1，表明大多数分支均处于负选择压力下。自由比率模型中，检测到大多数分支  $\omega$  值远小于 1，仅有 3 个小分支的  $\omega$  值大于 1(CHAB2802 分支的  $\omega$  值为 1.3454, CHAB2815 和 CHAB2823 分支的  $\omega$  值为 999.0000)，因此对这 3 个分支进行了分支-位点模型检测。对分支模型中二比率模型 A~D 以及自由比率模型 H 进行 LRT 检验(表 2)，结果表明各分支后验概率均不具有可靠性。

位点模型中，模型 M3(离散)、M2a(选择)和 M8( $\beta$  &  $\omega$ )允许  $\omega$  值  $> 1$ ，与其对应的零假设模型为 M1a(近中性)模型、M0(单一比值)模型和 M7( $\beta$ )模型。经 LRT 检验，M3 模型显著优于 M0 零假设模型 ( $P < 0.01$ )，表明各位点间经受的选择压力具有差异性。在位点模型中均没有检测出正选择位点，表明 *hetR* 基因处于强烈负选择压力下。

分支-位点模型中，指定 7 个分支为前景支，其中分支 E 的序列代号为 CHAB2802，分支 F 为 CHAB2815，分支 G 为 CHAB2823。在分支 A 检测出 95 K(后验概率为 88.6%)、分支 B 检测出 43 E(后验概率为 55.7%)、分支 C 检测出 45 Y(后验概率为 64.9%) 和 81 V(后验概率为 74.1%)、分支 D 检测出 40 V(后验概率为 53.4%)、分支 E 检测出 126 T(后验概率为 99.7%)、分支 F 检测出



节点处的数字代表最大似然法步靴值，A~G 代表选定的分支。

Bootstrap values are noted on nodes and selected branches are marked as A–G.

图 1 基于 *hetR* 基因蛋白序列构建的系统发育树  
Fig. 1 Phylogenetic tree based on *hetR* gene sequence

11 E(后验概率为 98.2%)、分支 G 检测出 71 L(后验概率为 98.4%)为正选择位点。但经似然比检验, 仅分支 E 的后验概率小于 0.05, 支持其检测出正选择位点的结论; 其余分支不可靠, 拒绝其存在正选择位点。

### 2.3 正选择位点定位

本研究基于同源建模原理, 对 PDB 数据库

进行 Blast 搜索, 发现普通念珠藻与 *Fischerella* 属藻类(PDBID: 4J00) HetR 蛋白的结构相似度高达 92.36%, 满足同源建模的可靠性条件。基于此来预测 HetR 蛋白的三维结构。经对位排列(图 2), 确定被检测出的正选择位点的相对位置。在 HetR 蛋白三维预测结构中标示出正选择位点(图 3)。

表 2 各模型参数的估计值和对数似然值  
Table 2 Parameter estimates and log-likelihood values for different models

模型 Model	参数数量 Number	对数似然值 lnL	参数估计值 Estimated value	正选择位点 Positive selection site
分支模型 Branch model	单比率模型 M0 One-ratio model M0	101	$\omega = 0.06112$	无应答 Not allowed
	二比率模型 A Two-ratios model A	102	$\omega_0 = 0.05854, \omega_1 = 0.09144$	无应答 Not allowed
	二比率模型 B Two-ratios model B	102	$\omega_0 = 0.06109, \omega_1 = 0.06250$	无应答 Not allowed
	二比率模型 C Two-ratios model C	102	$\omega_0 = 0.06112, \omega_1 = 16.49504$	无应答 Not allowed
	二比率模型 D Two-ratios model D	102	$\omega_0 = 0.06088, \omega_1 = 0.07592$	无应答 Not allowed
	自由比率模型 H Free-ratio model H	199	-3855.852594	无应答 Not allowed
位点模型 Site model	M1a: 近中性 M1a: Near neutral	102	$p_0 = 0.91384, p_1 = 0.08616$ $\omega_0 = 0.03494, \omega_1 = 1.00000$	无应答 Not allowed
	M2a: 选择 M2a: Selection	104	$p_0 = 0.91384, p_1 = 0.04196$ $p_2 = 0.04420, \omega_0 = 0.03494$ $\omega_1 = 1.00000, \omega_2 = 1.00000$	无应答 Not allowed
	M3: 离散 M3: Discrete	105	$p_0 = 0.26165, p_1 = 0.60411$ $p_2 = 0.13424, \omega_0 = 0.00000$ $\omega_1 = 0.03062, \omega_2 = 0.33599$	无 None
	M7 ( $\beta$ )	102	$p = 0.36378, q = 4.72437$	
	M8 ( $\beta$ & $\omega$ )	104	$p_0 = 0.99999, p = 0.36378,$ $q = 4.72443$ $p_1 = 0.00001, \omega = 1.00000$	无 None
	备择假设 a Alternative model a	104	$p_{2a} = 0.03806, p_{2b} = 0.00335$ $\omega_{b1} = 0.03413, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	95 K, 0.886
分支-位点模型 Branch-site model	零假设 a0 Null model a0	103	$p_{2a} = 0.03806, p_{2b} = 0.00335$ $\omega_{b1} = 0.03413, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	无应答 Not allowed
	备择假设 b Alternative model b	104	$p_{2a} = 0.00000, p_{2b} = 0.00000$ $\omega_{b1} = 0.03494, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	43 E, 0.557
	零假设 b0 Null model b0	103	$p_{2a} = 0.00000, p_{2b} = 0.00000$ $\omega_{b1} = 0.03494, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	无应答 Not allowed
	备择假设 c Alternative model c	104	$p_{2a} = 0.01470, p_{2b} = 0.00136$ $\omega_{b1} = 0.03335, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.04156, \omega_{f2} = 1.04156$	45 Y, 0.649; 81 V, 0.747
	零假设 c0 Null model c0	103	$p_{2a} = 0.01502, p_{2b} = 0.00139$ $\omega_{b1} = 0.03334, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	无应答 Not allowed
	备择假设 d Alternative model d	104	$p_{2a} = 0.00155, p_{2b} = 0.00015$ $\omega_{b1} = 0.03488, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	40 V, 0.534

续表 2

模型 Model	参数数量 Number	对数似然值 lnL	参数估计值 Estimated value	正选择位点 Positive selection site
分支-位点模型 Branch-site model	零假设 d0 Null model d0	103 -3885.281658	$p_{2a} = 0.00155, p_{2b} = 0.00015$ $\omega_{b1} = 0.03488, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	无应答 Not allowed
	备择假设 e Alternative model e CHAB2802	104 -3881.030046	$p_{2a} = 0.00791, p_{2b} = 0.00075$ $\omega_{b1} = 0.03411, \omega_{b2} = 1.00000$ $\omega_{f1} = 999.00000, \omega_{f2} = 999.00000$	126 T, 0.997 **
	零假设 e0 Null model e0 CHAB2802	103 -3884.711178	$p_{2a} = 0.10109, p_{2b} = 0.00960$ $\omega_{b1} = 0.03425, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	无应答 Not allowed
	备择假设 f Null model f CHAB2815	104 -3882.537141	$p_{2a} = 0.91318, p_{2b} = 0.08682$ $\omega_{b1} = 0.03404, \omega_{b2} = 1.00000$ $\omega_{f1} = 115.62496, \omega_{f2} = 115.62496$	11 E, 0.982 *
	零假设 f0 Null model f0 CHAB2815	103 -3882.887586	$p_{2a} = 0.91318, p_{2b} = 0.08682$ $\omega_{b1} = 0.03404, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	无应答 Not allowed
	备择假设 g Null model g CHAB2823	104 -3882.546225	$p_{2a} = 0.91317, p_{2b} = 0.08683$ $\omega_{b1} = 0.03404, \omega_{b2} = 1.00000$ $\omega_{f1} = 127.18806, \omega_{f2} = 127.18806$	71 L, 0.984 *
	零假设 g0 Null model g0 CHAB2823	103 -3882.892224	$p_{2a} = 0.91318, p_{2b} = 0.08682$ $\omega_{b1} = 0.03404, \omega_{b2} = 1.00000$ $\omega_{f1} = 1.00000, \omega_{f2} = 1.00000$	无应答 Not allowed

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。

Note: Positively selected sites with posterior probabilities greater than 95% and 99% are marked as “\*” and “\*\*”, respectively.

表 3 LRT 检验统计量  
Table 3 Likelihood ratio statistics

模型比较 Comparison between models	2Δ/	自由度 d. f.	P-value
分支模型 Branch model	M0 vs. A	1.556674	0.212
	M0 vs. B	0.000862	0.977
	M0 vs. C	0	1
	M0 vs. D	0.074922	0.78
	M0 vs. H	118.129482	0.0813
位点模型 Site model	M0 vs. M3	132.172684	0 **
	M1a vs. M2a	0	1
	M7 vs. M8	0.002106	1
分支-位点模型 Branch-site model	a vs. a0	0	1
	b vs. b0	0	1
	c vs. c0	0.001302	0.97
	d vs. d0	0	1
	e vs. e0	7.362264	0.0067 **
	f vs. f0	0.70089	0.4025
	g vs. g0	0.691998	0.4055

\*\*：表示在  $P < 0.01$  水平上差异极显著。

\*\*：Indicates extremely significant difference at  $P < 0.01$ .

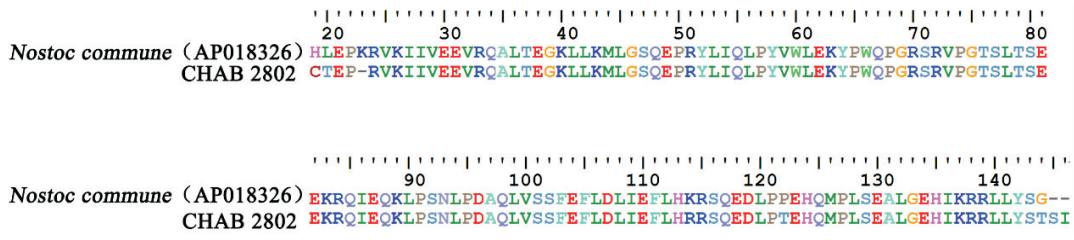
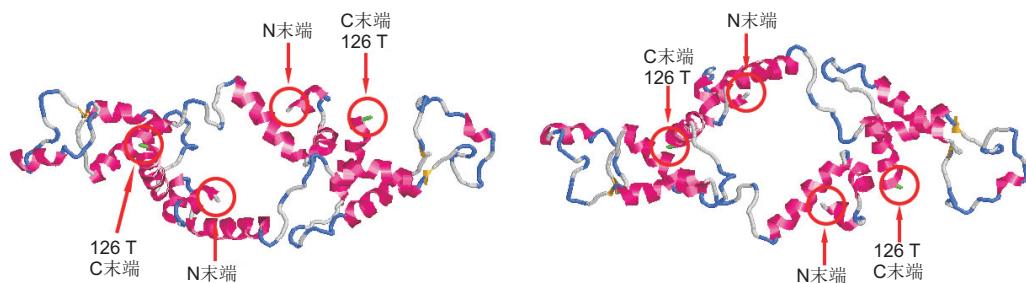


图 2 HetR 氨基酸序列对位排列  
Fig. 2 Alignment of amino acid sequence of HetR



正选择位点用红色圆圈标出。  
Positively selected sites are in red circles.

图 3 普通念珠藻 HetR 正选择位点的空间位置  
Fig. 3 Spatial locations of positively selected sites of HetR in *Nostoc commune*

### 3 讨论

念珠藻属是蓝藻门中极其重要的类群,它可以分化出异形胞固定大气中的氮气。*hetR* 基因是念珠藻异形胞分化和厚壁孢子形成中的关键基因。这提示 *HetR* 在念珠藻细胞分化过程中是重要的调控因子,而不仅局限于异形胞的分化过程。由于各模型中正选择/负选择位点判断的统计方法的进一步改进,对基因序列正选择作用重要位点的研究已成为热点<sup>[6, 21]</sup>。研究正选择位点不仅有助于预测基因的功能区,还可为攻克一些生物技术问题提供新依据。目前,在分子适应性进化研究中,主要度量基于密码子置换模型计算出的非同义替换率(*dN*)和同义替换率(*dS*)的比值( $\omega$ ),通过这项指标来判断自然选择压力<sup>[22, 23]</sup>。研究表明,基于  $\omega$  比值开发的 DNA 编码序列进化模型是检验分子适应的实用方法。当  $dN > dS$ , 即  $\omega > 1$ , 且计算出模型间具有显著性差异,则认为非同义突变有利于选择,即编码序列在相应的位点或分支发生过正选择。这种依托 ML 方法建立模型来检测正选择位点的手段,可更灵敏地识别被较强负选择和中性漂变作用掩盖的正选择位点<sup>[24]</sup>。同时,也可独立分析不同支系和位点来确定氨基酸的正选择作用<sup>[25, 26]</sup>。

当基因经受正选择时,表明该物种与环境中某关键因素的适应性,提示在该进化时间点可能产生一些蛋白新功能;而当基因处于强烈负选择时,则意味着该物种的蛋白质保持原有的重要功能且趋于稳定<sup>[11]</sup>。本研究在分支模型和位点模型均没有检测出正选择位点,表明 *hetR* 基因处于强烈负选择压力下。可能的原因是,对于念珠藻属植物来说, *hetR* 基因发挥着非常重要的作用,在长期的进化过程中,该基因在结构甚至是功能上已经具有一定的稳定性。也可能是由于 *hetR* 基因发生了较早的适应性进化,而普遍的中性替换位点或者净化选择掩盖了后期的适应性进化信号。巩超彦等<sup>[12]</sup>和韩雨昕等<sup>[13]</sup>分别对藻类植物进化中的重要类群即串珠藻目和弯枝藻属植物的 *rbcL* 基因进行了适应性进化研究,均并未能在位点模型下发现正选择位点。这些结果显示藻类 *hetR* 和 *rbcL* 基因序列数据可能包含的变异太少,适应性信号被淹没于极其普遍的净化选择之中。

本研究分支-位点模型的检验结果表明,分支 A~G 均检测出正选择位点,但经似然比检验后发现仅有分支 E(CHAB2802)的后验概率小于 0.05,支持其检测出正选择位点的结论。其余分支不可靠,拒绝其存在正选择位点。Zhang<sup>[27]</sup> 的研究检

验出分支-位点模型的似然比检验可能存在假阳性，随后对分支-位点模型进行了改进，构建了两个（test 1 和 test 2）LRT 检验。通过大量数据验证 test 2 应用于实际数据分析较为可靠，因此假阳性问题迎刃而解<sup>[28]</sup>。通过大量相关文献的分析，本研究选取的序列数量较少和系统发育树分支较短可能导致了假阳性结果的存在，同时选择压力放松也可能会出现这一结果。

本研究采用分支模型、位点模型以及分支-位点模型对念珠藻属 *hetR* 基因的选择位点进行检测，发现绝大多数藻株未发现正选择位点，说明所检测物种的很多位点处于负选择压力下，仅有 CHAB2802 这株藻在分支-位点模型中检测到一个正选择位点（126T），提示念珠藻属植物 *hetR* 基因发生了适应性变化。同时，此结果表明，*hetR* 基因变异较小而不宜作为种水平的分类研究。而在高等植物研究中发现，很多基因普遍存在多个正选择位点<sup>[7-11]</sup>。可能念珠藻 *hetR* 基因数据较少是造成这一结果的重要原因。通过查阅文献，发现目前对藻类的适应性进化研究较少，今后可加强对藻类其他类群的重要基因适应性进化分析，并推断其在环境条件影响下分子水平的变化。

## 参考文献：

- [ 1 ] McGregor GB. Freshwater cyanobacteria of north-eastern Australia: 3. Nostocales [ J ]. *Phytotaxa*, 2018, 359 ( 1 ): 1.
- [ 2 ] Williams W, Büdel B, Williams S. Wet season cyanobacterial N enrichment highly correlated with species richness and *Nostoc* in the northern Australian savannah [ J ]. *Bio-geosciences*, 2018, 15 ( 7 ): 2149-2159.
- [ 3 ] Nowruzi B, Haghigat S, Fahimi H, Mohammadi E. *Nostoc* cyanobacteria species: a new and rich source of novel bioactive compounds with pharmaceutical potential [ J ]. *Journal of Pharmaceutical Health Services Research*, 2017, 9: 5-12.
- [ 4 ] Muro-Pastor AM, Hess WR. Heterocyst differentiation: from single mutants to global approaches [ J ]. *Trends Microbiol*, 2012, 20 ( 11 ): 548-557.
- [ 5 ] Leganés F, Fernández-Pinas F, Wolk CP. Two mutations that block heterocyst differentiation have different effects on akinete differentiation in *Nostoc ellipsosporum* [ J ]. *Mol Microbiol*, 1994, 12 ( 4 ): 679-684.
- [ 6 ] Nei M, Kumar S. Molecular Evolution and Phylogenetics [ M ]. New York: Oxford University Press, 2000.
- [ 7 ] 森林, 苏应娟, 张冰, 王艇. 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.
- [ 8 ] 周媛, 王博, 高磊, 王艇. 凤尾蕨科旱生蕨类 *rbcL* 基因的适应性进化和共进化分析 [ J ]. 植物科学学报, 2011, 29 ( 4 ): 409-416.
- [ 9 ] Zhou Y, Wang B, Gao L, Wang T. Adaptive evolution and coevolution of the *rbcL* gene in xeric Pteridaceae ferns [ J ]. *Plant Science Journal*, 2011, 29 ( 4 ): 409-416.
- [ 10 ] 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.
- [ 11 ] Chen XX, Su YJ, Wang T. Adaptive evolution analysis of the *psbA* gene in Lejeuneaceae [ J ]. *Acta Botanica Boreali-Occidentalia Sinica*, 2010, 30 ( 8 ): 1534-1544.
- [ 12 ] 许可, 王博, 苏应娟, 高磊, 王艇. 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.
- [ 13 ] Xu K, Wang B, Su YJ, Gao L, Wang T. Molecular evolution of *psbD* gene in ferns: selection pressure and co-evolutionary analysis [ J ]. *Plant Science Journal*, 2013, 31 ( 5 ): 429-438.
- [ 14 ] 吴筱婷, 森林, 陈楠, 张潇, 马朝霞, 张钦宇. 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.
- [ 15 ] Wu XP, Sen L, Chen N, Zhang X, Ma ZX, Zhang QY. Study on the molecular evolution of the *psaA* gene from ferns [ J ]. *Plant Science Journal*, 2017, 35 ( 2 ): 177-185.
- [ 16 ] 巩超彦, 南芳茹, 冯佳, 吕俊平, 刘琪, 谢树莲. 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.
- [ 17 ] Gong CY, Nan FR, Feng J, Lü JP, Liu Q, Xie SL. Adaptive evolutionary analysis on *rbcL* gene of Batrachospermales [ J ]. *Oceanologia et Limnologia Sinica*, 2017, 48 ( 3 ): 527-535.
- [ 18 ] 韩雨昕, 南芳茹, 巩超彦, 冯佳, 吕俊平, 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.
- [ 19 ] Han YX, Nan FR, Gong CY, Feng J, Lü JP, et al. Adaptive evolutionary analysis of the *rbcL* gene from *Compsopogon* (Rhodophyta) [ J ]. *Journal of Tropical and Subtropical Botany*, 2019, 27 ( 1 ): 36-44.
- [ 20 ] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0 [ J ]. *Mol Biol Evol*, 2013, 30 ( 12 ): 2725-2729.
- [ 21 ] Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0 [ J ]. *Syst Biol*, 2010, 59 ( 3 ): 307-321.

- [16] Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests[J]. *Syst Biol*, 2004, 53(5): 793–808.
- [17] Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood[J]. *Mol Biol Evol*, 2007, 24(8): 1586–1591.
- [18] Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. Multiple sequence alignment with the Clustal series of programs [J]. *Nucleic Acids Res*, 2003, 31(13): 3497–3500.
- [19] Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT [J]. *Nucleic Acids Symp Ser*, 1999, 41: 95–98.
- [20] Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all[J]. *Trends Biochem Sci*, 1995, 20(9): 374–376.
- [21] Wang B, Su YJ, Wang T. Molecular cloning of *rbcS* genes in *Selaginella* and the evolution of the *rbcS* gene family [J]. *Arch Biol Sci*, 2015, 67(2): 373–383.
- [22] Yang ZH. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution[J]. *Mol Biol Evol*, 1998, 15(5): 568–573.
- [23] Yang ZH. Computational molecular evolution [M]. New York: Oxford University Press, 2006.
- [24] Yang ZH. Inference of selection from multiple species alignments[J]. *Curr Opin Genet Dev*, 2002, 12(6): 688–694.
- [25] Yang ZH, Bielawski JP. Statistical methods for detecting molecular adaptation [J]. *Trends Ecol Evol*, 2000, 15(12): 496–503.
- [26] Yang ZH, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites[J]. *Genetics*, 2000, 155(1): 431–449.
- [27] Zhang JZ. Frequent false detection of positive selection by the likelihood method with branch-site models[J]. *Mol Biol Evol*, 2004, 21(7): 1332–1339.
- [28] Zhang JZ, Nielsen R, Yang ZH. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level [J]. *Mol Biol Evol*, 2005, 22(12): 2472–2479.

(责任编辑:周媛)