

DOI:10.11913/PSJ.2095-0837.2017.20177

吴筱娉, 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): 177-185Wu 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

蕨类植物 *psaA* 基因的分子进化研究

吴筱娉¹, 森林^{1,2*}, 陈楠¹, 张潇¹, 马朝霞¹, 张钦宇¹

(1. 湖北中医药大学药学院, 武汉 430065; 2. 中药资源与中药复方教育部重点实验室, 武汉 430065)

摘要: 采用“放松分子钟”模型、氨基酸位点正选择模型和分子内共进化网络估算方法, 对蕨类植物光合系统 I 核心蛋白 PSAA 编码基因 *psaA* 的进化趋势进行了研究。结果显示, 叶绿体基因 *psaA* 编码区全序列具备成为蕨类植物系统发育关系重建位点的潜力, 与 *rbcL* 基因联合后能构建高后验概率的系统发育树; 蕨类植物的 PSAA 蛋白中存在一些曾经经历正选择的氨基酸位点, 其中 29 个位点聚合成为 16 个共进化组, 通过共进化网络的方式协同影响光合系统 I 的内部调整, 提升其在被子植物兴起后光合环境下的适应能力。本文对蕨类植物进化潜能与分子机理的研究结果为揭示蕨类植物适应新生境提供了科学依据, 也为植物系统分类学研究提供了分子依据。

关键词: 蕨类植物; 薄囊蕨; 光合系统 I; 适应性进化; 共进化

中图分类号: Q346+.4; Q949.36

文献标识码: A

文章编号: 2095-0837(2017)02-0177-09

Study on the molecular evolution of the *psaA* gene from ferns

Wu Xiao-Ping¹, Sen Lin^{1,2*}, Chen Nan¹, Zhang Xiao¹, Ma Zhao-Xia¹, Zhang Qin-Yu¹

(1. College of Pharmacy, Hubei University of Chinese Medicine, Wuhan 430065, China; 2. Key Laboratory of Education Ministry on Traditional Chinese Medicine Resources and Compound Prescription, Wuhan 430065, China)

Abstract: The *psaA* gene is a fundamental protein-coding gene of photosystem I. Using the relaxed molecular clock model, positive selection model, and co-evolutionary analysis, we attempted to unravel its evolutionary pattern. Results indicated that the full-length coding sequences of *psaA* from fern species could be novel sites for the reconstruction of phylogenetic trees, and could exhibit high posterior probability when combined with the full-length coding sequences of *rbcL*. The current research also illustrated that the PSAA protein had 29 positively selected amino-acid sites aggregated into 16 co-evolution groups, indicating that ferns had internally modified their photosystem I to enhance their ability to adapt to the photosynthetic environment after the rise of angiosperms.

Key words: Monilophytes; Leptosporangiate; photosystem I; Adaptive evolution; Co-evolution

新近植物分类学对原归于蕨类植物门(Pteridophyta)的物种开展了重排研究^[1], 并以石松类植物(Lycophytes)与蕨类植物(Monilophytes, 即

ferns)为名将物种进行重新区分, 从而对传统意义上的蕨类植物分类认识进行了修订^[2]。新分类系统的蕨类植物包括: 松叶蕨(Whisk ferns)、木贼

收稿日期: 2016-12-05, 退修日期: 2017-1-20。

基金项目: 国家自然科学基金青年项目(31500260); 湖北中医药大学教育部重点实验室科研启动经费(5114000914); 湖北中医药大学“青苗计划”(2016ZZX015)资助。

This work was supported by grants from the National Natural Science Foundation for Young Scientists of China (31500260), Key Laboratory of the Education Ministry at the Hubei University of Chinese Medicine (5114000914), and Young Talent Project of the Hubei University of Chinese Medicine (2016ZZX015).

作者简介: 吴筱娉(1996-), 女, 本科生, 中药学专业(E-mail: wuxiaoping_hbtcm@126.com)。

* 通讯作者(Author for correspondence. E-mail: senlin_hbtcm@126.com)。

(Horsetails)、厚囊蕨 (Eusporangiate)、薄囊蕨 (Leptosporangiate) 4 大类群, 其中厚囊蕨并不是一个单系类群^[3]。有研究认为自从地球上被子植物兴起后, 薄囊蕨出现了物种辐射, 以应对高大木本植物冠盖带来的光合作用环境改变^[4]。然而, 其他蕨类植物的进化潜力及薄囊蕨物种辐射现象所蕴藏的分子机理尚未完全厘清。

为达到适应新光合环境的目标, 陆生植物的基因会经历分子水平的正选择^[5,6]。Ran 等^[7]对陆生植物气孔研究时发现被子植物和裸子植物 *SPCH* 基因的进化趋势各异; 周媛和森林等^[8,9]对凤尾蕨科 (Pteridaceae) 植物的研究表明, 水生分支、附生分支和旱生分支物种的 Rubisco 酶大亚基编码 *rbcL* 基因呈现各自不同的进化趋势; Kapralov 等^[10]针对陆生植物 *rbcL* 基因的大规模扫描研究表明, 该基因曾伴随 CO₂ 含量的变化而发生微调; Sen 等^[11]对裸子植物罗汉松科 (Podocarpaceae)、红豆杉科 (Taxaceae) 及三尖杉科 (Cephalotaxaceae) 的研究也印证了上述结论; Sen 等^[12,13]对裸子植物的研究表明, 裸子植物光合系统 II 核心蛋白编码 *psbA* 基因和蕨类植物单拷贝 *psbA* 基因也在正选择作用下发生了改变。物种类群的更替是环境改变的外因与遗传历程的内因共同作用的结果。对遗传内因的探究不仅有利于理解物种进化潜力及进化机理, 还有助于物种资源的人工利用。

光合系统 I 自从被发现后一直是光合作用研究的热点^[14]。作为光合系统 I 跨膜复合物的中心蛋白, PSAA 蛋白与 PSAB 蛋白相互结合, 共同承担着重要的生物学功能^[15]。其中, PSAA 蛋白的编码基因 *psaA* 和 *rbcL* 基因均位于叶绿体基因组的大单拷贝区 (Large single copy, LSC 区), 而且前者编码区长度比后者更长^[16]。有研究报道, *rbcL* 基因作为分子标记在一些较低的分类单位中的辨识度较低^[17], 引发了利用更多位点组成联合数据集参与系统发育关系重建的研究热潮。虽然我国在观赏植物矢竹属 (*Pseudosasa*) 物种^[18]、栽培植物紫菜属 (*Porphyra*) 物种中^[19]已有 *psaA* 基因的克隆测序报道, 但该基因是否具备成为蕨类植物系统发育关系重建位点的潜力, 尚少有研究探讨。

为了明晰蕨类植物对陆生环境响应的分子机制, 我们利用 18 科 24 属 27 种具有代表性的蕨类植物 *psaA* 基因编码区进行系统发育关系重建、适

应性进化分析及共进化关系研究, 以阐明“放松分子钟”模型下, 基于光系统 I 重要编码基因 *psaA* 核苷酸序列信息能否正确描绘出蕨类植物 4 大类群的分化式样, 并比较与 *rbcL* 基因编码区全序列及二基因序列联合数据集间的构树效率; 探讨伴随薄囊蕨类群兴盛, *psaA* 基因内部是否存在适应性进化; 并估算 PSAA 蛋白氨基酸位点间的共进化关系。

1 材料与方法

1.1 序列数据

本研究所选取的蕨类植物 4 大类群 18 科 24 属 27 种植物及其由 GenBank 获得的 *psaA* 基因及 *rbcL* 基因编码区全长序列见表 1。使用 MAFFT 软件进行序列比对^[20-22], 经手工校正后单条 *psaA* 基因序列包括 750 个氨基酸编码密码子和 1 个终止密码子, 单条 *rbcL* 基因包括 475 个氨基酸编码密码子和 1 个终止密码子。为便于比较, 密码子位点的编号参照桫欏 (*Alsophila spinulosa* (Wall. ex Hook.) R. M. Tryon, NC_012818) 的基因序列。构建 3 个数据集: (1) 仅包括 *psaA* 编码区的序列; (2) 仅包括 *rbcL* 编码区的序列; (3) 将来自相同叶绿体基因组的 *psaA* 基因与 *rbcL* 基因的编码区全长逐一首尾相连后由 *psaA* 编码区与 *rbcL* 编码区构成的联合序列。

1.2 统计分析

将上述 3 个数据集的结果导入 jModeltest V2.1.4 软件, 选取该数据集对应的核苷酸进化模型^[23-25]。分别利用 BEAST V1.8.4 软件的 UCLD (Uncorrelated lognormal distributed relaxed clock model) 分子钟模型估算系统树每个节点的分歧时间^[26]。参照 Sen 等^[12]估算的蕨类植物系统发育关系, 用最近共同祖先时间 (t_{MRCA}) 值校正主要分支的分歧时间: 薄囊蕨类植物总体类群 t_{MRCA} 值为 299.9 Mya (百万年前, 下同); 松叶蕨-瓶尔小草 (Psilotaceae-Ophioglossaceae) 支系 t_{MRCA} 值为 135.7 Mya^[12]。根据 jModeltest V2.1.4 软件所选取的 GTR + G + I 模型, 在 BEAUti V1.8.4 软件中设置贝叶斯迭代模型, 迭代计算 2×10^7 代, 每 1000 代保存 1 株样本; 将输出结果导入 Tracer V1.6.0 软件检测各参数后验概率的收敛程度, 所有参数的有效取样大小 (Efficient sampling size, ESS) 值均大于 500 时, 确定运行达到收敛终点。将树结果导入

表 1 本研究所选取的植物及其 GenBank 登录号
Table 1 Sampled plants and GenBank accession numbers

类别 Classification	科名 Family	种名 Species	中文名 Chinese name	GenBank 登录号 GenBank Acc. No.	<i>psaA</i> 基因碱基数目 Bp numbers in <i>psaA</i> (bp)	<i>rbcl</i> 基因碱基数目 Bp numbers in <i>rbcl</i> (bp)
薄囊蕨 Leptosporangiate	Dicksoniaceae	<i>Dicksonia squarrosa</i> (Forst.) Sw.		KJ569698	2253	1428
	Lygodiaceae	<i>Lygodium japonicum</i> (Thunb.) Sw.	海金沙	NC_022136	2253	1428
	Cyatheaceae	<i>Alsophila spinulosa</i> (Wall. ex Hook.)	桫欏	NC_012818	2253	1428
		R. M. Tryon				
	Pteridaceae	<i>Cheilanthes lindheimeri</i> Hook.	碎米蕨	NC_014592	2253	1428
		<i>Adiantum capillus-veneris</i> L.	铁线蕨	NC_004766	2253	1428
		<i>Ceratopteris richardii</i> Brongn.		KM052729	2253	1428
	Woodwardioideae	<i>Woodwardia unigemmata</i> (Makino) Nakai	顶芽狗脊	NC_028543	2253	1428
	Dennstaedtiaceae	<i>Pteridium aquilinum</i> (L.) Kuhn	欧洲蕨	NC_014348	2253	1428
	Dryopteridaceae	<i>Cyrtomium devesicapulae</i> (Koidz.) Ching	披针贯众	NC_028542	2253	1428
		<i>Cyrtomium falcatum</i> (L. f.) C. Presl	全缘贯众	NC_028705	2253	1428
	Marsileaceae	<i>Marsilea crenata</i> C. Presl	南国田字草	NC_022137	2253	1428
	Gleicheniaceae	<i>Diplopterygium glaucum</i> Thunb. ex Houtt.		NC_024158	2253	1428
	Osmundaceae	<i>Osmundastrum cinnamomeum</i> (L.) C. Presl		NC_024157	2253	1428
	Cystopteridaceae	<i>Cystopteris protrusa</i> (Weath.) Blasdell		KP136830	2253	1428
	Polypodiaceae	<i>Polypodium glycyrrhiza</i> D. C. Eaton		KP136832	2253	1428
	Plagiogyriaceae	<i>Plagiogyria glauca</i> (Blume) Mett.		KP136831	2253	1428
	Dipteridaceae	<i>Dipteris conjugata</i> Reinw.	双扇蕨	KP136829	2253	1428
厚囊蕨 Eusporangiate	Ophioglossaceae	<i>Mankyua chejuensis</i> B. Y. Sun, M. H. Kim & C. H. Kim		NC_017006	2253	1434
		<i>Helminthostachys zeylanica</i> (L.) Hook.	七指蕨	KM817788	2253	1431
		<i>Botrychium ternatum</i> (Thunb.) Sw.	阴地蕨	KM817789	2253	1428
		<i>Ophioglossum californicum</i> Prantl		NC_020147	2253	1428
	Marattiaceae	<i>Angiopteris evecta</i> (G. Forst.) Hoffm.	观音座莲	NC_008829	2253	1428
		<i>Angiopteris angustifolia</i> C. Presl		NC_026300	2253	1428
木贼类 Horsetails	Equisetaceae	<i>Equisetum arvense</i> L.	问荆	NC_014699	2253	1428
		<i>Equisetum hyemale</i> L.	木贼	NC_020146	2253	1428
松叶蕨类 Whisk ferns	Psilotaceae	<i>Tmesipteris elongata</i> Danguy		KJ569699	2253	1428
		<i>Psilotum nudum</i> (L.) Beauv.	松叶蕨	NC_003386	2253	1428

TreeAnnotator V1.8.4 软件去除最初的 2 000 000 代后，用剩余样本中后验概率大于 95%的拓扑结构重建时间尺度下的一致树。利用 Figtree V1.4.2 软件展开 3 个数据集各自一致树的拓扑结构。

运行 PAMLx V1.3.1 软件提供的正选择模型 (M2a、M5、M8) 与对照模型 (M1a、M7、M8a) 开展位点适应性进化分析^[27]。模型 M5 假设蛋白质序列存在 3 类位点：保守位点($0 < \omega < 1$)、中性位点($\omega = 1$)和正选择位点($\omega > 1$)，满足 γ 分布，估算上述位点的 ω 值和概率 P 值。模型 M7 假设无正选择位点存在，所有位点的 ω 值满足 β 分布，且介于 0~1 之间。模型 M8 在 M7 基础上增加一类正选择位点，其概率由数据估计得出，位点 ω 取值可大

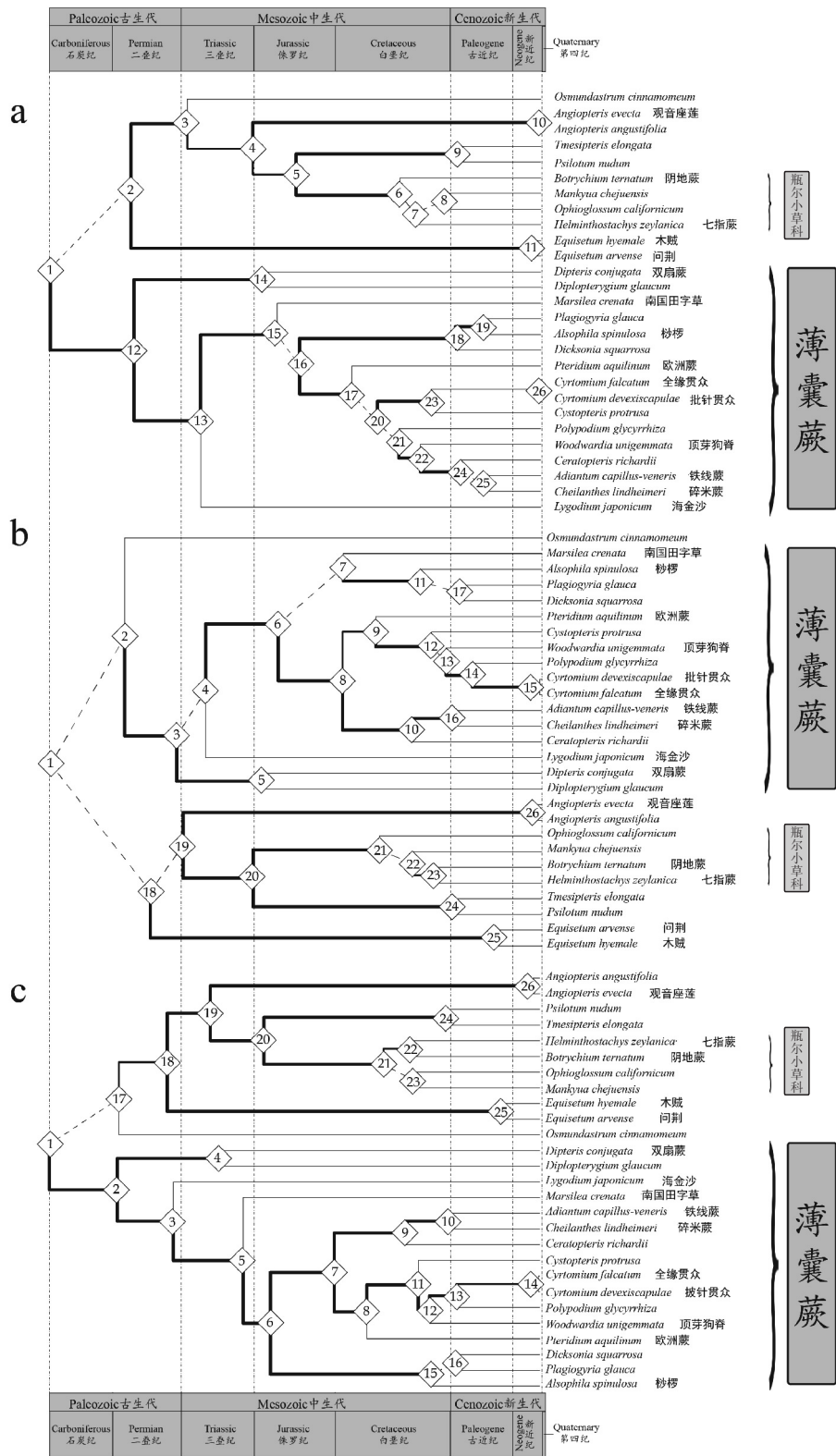
于 1。模型 M8a 同 M8 类似，只是将 ω_1 值固定为 1。最终经过似然率检测 (Likelihood ratio test, LRT)，比较模型 M7 和 M8、模型 M8 和 M8a 即可确定是否存在正选择位点。上述正选择模型中，M2a 最为原始和保守，分别估算 3 类位点的比例与 ω 值，需要比较 M1a 及 M2a 间的模型似然率。

利用 CAPS 软件分析蛋白质 PSAA 内部的共进化氨基酸位点。

2 结果

2.1 3 种数据集估测的蕨类植物系统发育关系

利用贝叶斯法重建的蕨类植物系统发育关系显示(图 1)，3 种数据集估测的蕨类植物系统发育关



a: 基于 *psaA* 基因编码区全序列数据集。b: 基于 *rbcL* 基因编码区全序列数据集。c: 基于 *psaA* + *rbcL* 基因联合数据集。粗实线表示该分支后验概率大于 95%，实线表示该分支后验概率在 90%~95% 之间，虚线表示该分支后验概率低于 90%。Phylogenetic tree in Fig. 1a resulted from the multiple alignment of full-length coding sequences of the *psaA* gene in 27 fern species, whereas Fig. 1b and Fig. 1c resulted from *rbcL* and *psaA* + *rbcL* datasets respectively. Heavily thickened lines, real lines, and dotted lines indicate posterior probability greater than 95%, between 90% and 95%, and less than 90%, respectively.

图 1 基于贝叶斯算法根据 3 种数据集分别构建的蕨类植物系统发育树
Fig. 1 Three different phylogenetic trees based on distinct datasets via Bayesian approaches

系中，瓶尔小草科(*Ophioglossaceae*)的 4 个成员(*Mankyua chejuensis*, *Helminthostachys zeylanica*, *Botrychium ternatum* 和 *Ophioglossum californicum*)均以较高的后验概率聚为一支，其中 *psaA* 基因编码数据集的后验概率为 1.00(图 1: a 节点 6)，分歧时间估测值为 103.47 Mya; *rbcl* 基因编码数据集的后验概率为 1.00(图 1: b 节点 21)，分歧时间估测值为 116.67 Mya; *psaA* + *rbcl* 联合数据集的后验概率为 1.00(图 1: c 节点 21)，分歧时间估测值为 110.99 Mya。这 3 个分歧时间估算值均指向瓶尔小草科在白垩纪早白垩纪世阿尔布期后发生了较大规模的物种分化。

从图 1 可看出，在蕨类植物系统发育关系重建过程中，*psaA* + *rbcl* 联合数据集具有最高的可信度：*rbcl* 基因编码区数据集估测的系统发育关系中有 8 个节点后验概率不足 90%(节点 2、4、7、13、17、18、19、22)；*psaA* 基因编码区数据集估测的

系统发育关系中后验概率不足 90%的节点减少到 7 个(节点 2、7、8、16、20、21、25)；而 *psaA* + *rbcl* 联合数据集则仅有 4 个后验概率不足 90%的节点(节点 12、16、17、23)。此结果表明：(1) *psaA* 基因编码区全长序列具备作为蕨类植物系统发育关系构建位点的潜力，甚至有时比传统的 *rbcl* 基因编码区全长表现更好；(2) 两个位点(*psaA* 和 *rbcl*)的联合能够显著提高系统发育关系构建的可信度。

2.2 蕨类植物 *psaA* 基因的正选择位点的估算

利用图形化的 PAMLx V1.3.1 软件，针对来自蕨类植物 4 大类群 18 科 24 属 27 个物种 *psaA* 编码区全序列数据集，对不同模型对应的对数似然值(lnL)以及参数估测值进行了测算。模型 M2a(选择)、M5(γ)和 M8(β 和 ω)均假定编码区位点存在正选择的情况，并通过估算寻找其位置(表 2)。经过 LRT 检测后发现，M2a 和 M8 均比其对照模型更符合所分析的数据(表 3)，其原假设模型在显

表 2 PSAA 蛋白中的正选择氨基酸位点及其参数估测值
Table 2 Estimated parameters and positively selected sites among protein PSAA

模型 Models	lnL	参数估计值 Estimated value of parameters	正选择位点 Positively selected sites*
M1a: 近中性 Near neutral	-8250.93	$p_0 = 0.7142, \omega_0 = 0;$ $p_1 = 0.2858, \omega_1 = 1$	不允许 Not allowed
M2a: 选择 Positive selection	-7987.89	$p_0 = 0.6525, \omega_0 = 0;$ $p_1 = 0.18195, \omega_1 = 1;$ $p_2 = 0.16555, \omega_2 = 10.00016$	16K、20K、25K、34R、35N、63L、70I、82F、84R、100S、108V、135V、144L、148S、165V、209G、215I、240L、247L、250T、255A、265N、267S、270S、273L、296L、304V、319T、336K、344S、351A、370A、399I、476L、484I、490L、493G、494S、500T、501A、502G、510S、511D、513V、514A、516A、522A、532L、548L、549L、553L、619S、620V、625A、626V、638S、692W、701W、707K、720I
M5: γ	-7769.26	$a = 0.10965, b = 0.0283$	16K、20K、25K、34R、35N、56F、63L、70I、79A、82F、84R、100S、108V、135V、144L、148S、152S、163A、165V、186A、209G、215I、232E、240L、247L、250T、255A、265N、267S、270S、273L、296L、304V、319T、327K、336K、338I、344S、351A、356T、370A、399I、476L、484I、490L、493G、494S、500T、501A、502G、510S、511D、513V、514A、516A、522A、532L、548L、549L、553L、619S、620V、625A、626V、638S、692W、701W、707K、720I
M7: β	-8251.21	$p = 0.005, q = 0.01173$	不允许 Not allowed
M8: β 和 ω	-7988.04	$p_0 = 0.982, p = 0.005;$ $q = 0.01936, p_1 = 0.16816,$ $\omega = 9.83383$	16K、20K、25K、34R、35N、63L、70I、82F、84R、100S、108V、135V、144L、148S、165V、209G、215I、240L、247L、250T、255A、265N、267S、270S、273L、296L、304V、319T、336K、344S、351A、370A、399I、476L、484I、490L、493G、494S、500T、501A、502G、510S、511D、513V、514A、516A、522A、532L、548L、549L、553L、619S、620V、625A、626V、638S、692W、701W、707K、720I

* 仅后验概率大于 99%的正选择位点被列入表 2 中，估算结果以 Naïve Empirical Bayes (NEB) 分析为准。
* Only sites with posterior probabilities greater than 99% are listed into Table 2. All results were based on Naïve Empirical Bayes (NEB) analysis.

表 3 各正选择模型与原假设模型间的似然比值检验统计量(2ΔlnL)

Table 3 Likelihood ratio statistics of positive selection models against their null models (2ΔlnL)

模型比较 Comparison between models	2ΔlnL	d.f.	$\chi^2_{1\%}$
M1a vs. M2a	526.08	2	9.21
M7 vs. M8	526.34	2	9.21
M8a vs. M8	436.8	1	6.635

著水平 $P = 1\%$ 之下被极显著拒绝。

从表 2 可看出，在 99% 水平上，模型 M2a 及 M8 中均发现了一致的 60 个正选择位点，在模型 M5 中发现有另外 9 个正选择位点的后验概率大于 99%，即：56F、79A、152S、163A、186A、232E、327K、338I、356T。经过 M8 模型中的 Naïve Empirical Bayes (NEB) 分析(图 2)及 Bayes Empirical Bayes (BEB)分析(图 3)，各自发现了 129 个和 130 个潜在的正选择位点。其中有 7 个位点(56F、79A、152S、163A、232E、

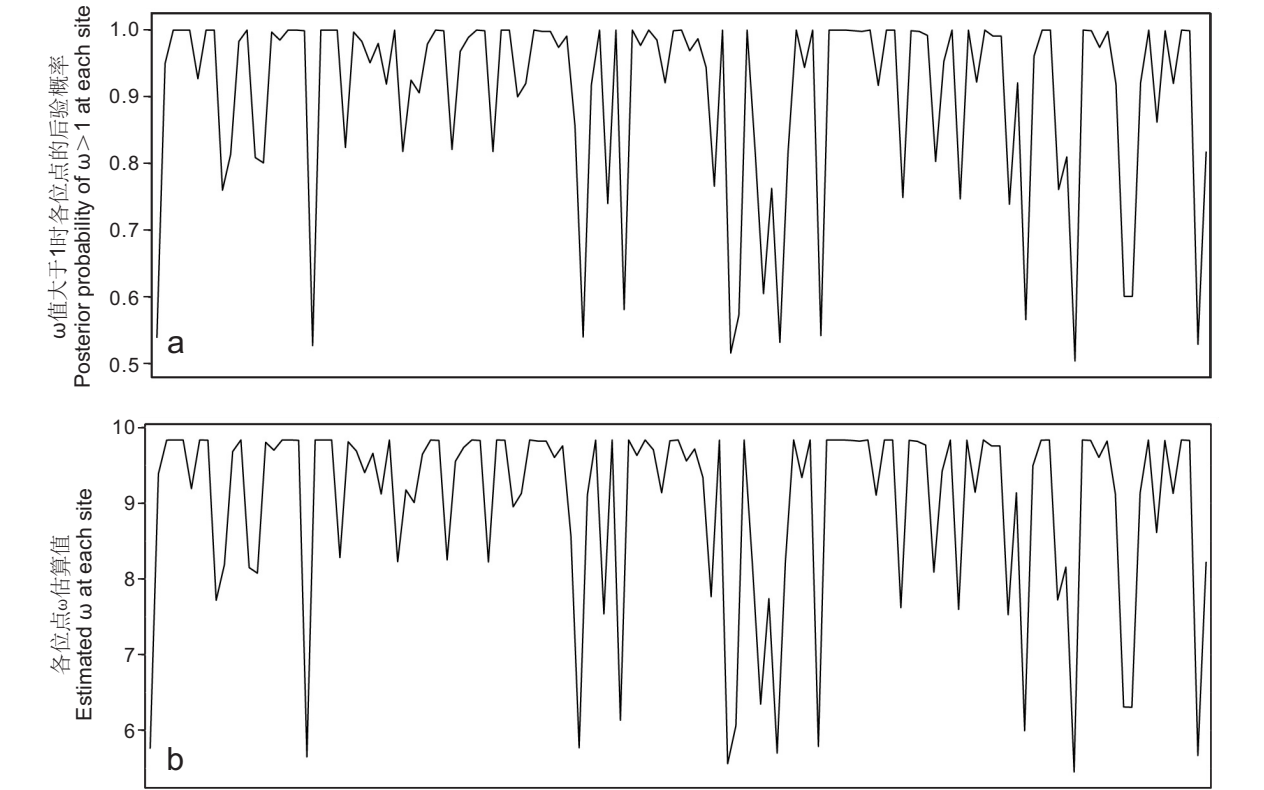
338I 和 356T)，既出现在 M5 模型估算结果中又在模型 M8 的 BEB 分析中的后验概率大于 99%。

2.3 蕨类植物光合系统 I 反应中心 PSAA 蛋白的重要共进化位点

从表 2 可看出，与 *rbcl* 基因和 *psbA* 基因编码区相比，蕨类植物类群的 *psaA* 基因编码区有较多的正选择位点。这些位点不仅两两之间形成共进化对 (Co-evolution pairs)，而且还多点聚合形成了共进化组 (Co-evolution groups) (表 4)，这种现象表明，这些正选择位点并非单独影响 PSAA 蛋白的适合度，而是通过多点协同共同影响蛋白的空间构象，进而影响蛋白的整体适合度。

3 讨论

本研究中单独的 *psaA* 编码区或者 *rbcl* 编码区构建的蕨类植物系统发育关系都存在较多后验概率低的节点。例如，图 1: b、c 中的节点 2 与图 1: a 中的节点 12 一致，均代表薄囊蕨的祖先分

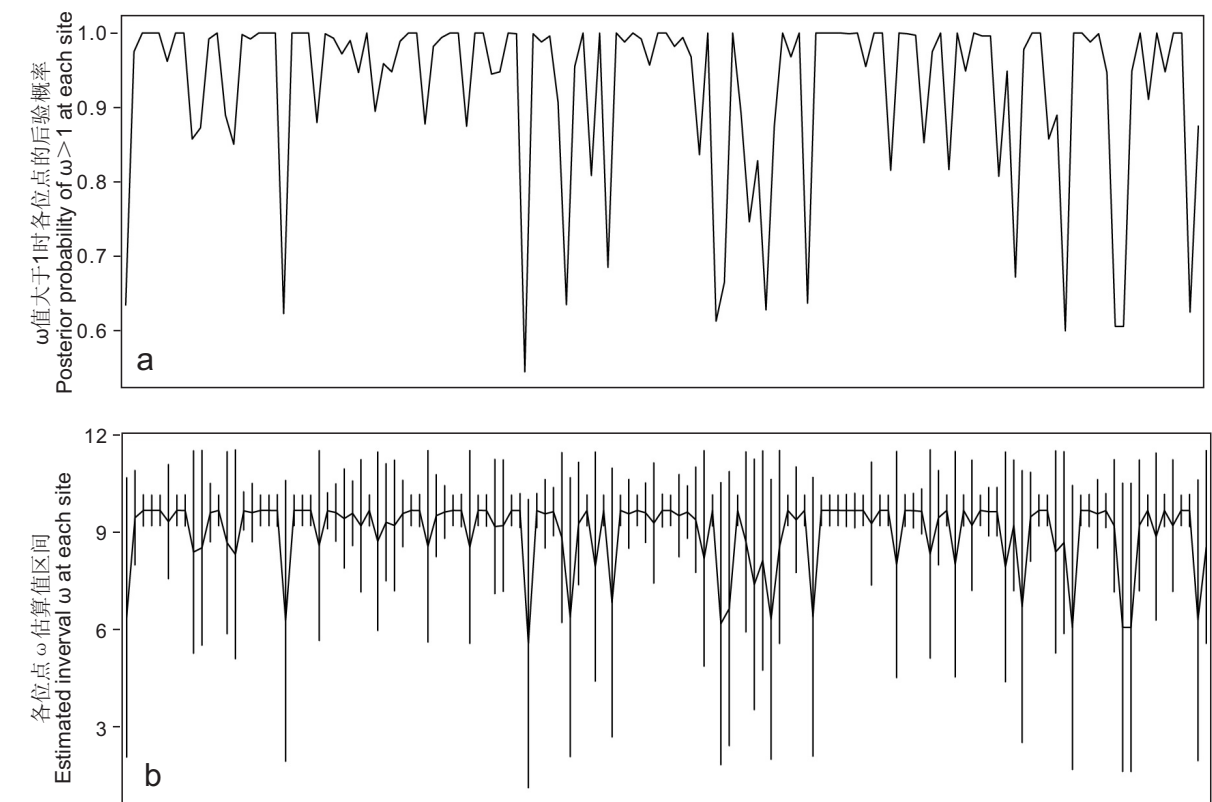


M8模型NEB算法下估算的PSAA蛋白编码区中拥有高 ω 值的129个位点
Candidate sites in PSAA protein after positive selection with high estimated ω in NEB analysis under model M8

a: 129 个位点 ω 值大于 1 时(即正选择)的后验概率; b: 129 个位点的 ω 估测值。
Fig. 2a: Posterior probability of estimated ω values greater than 1 (positively selected); Fig. 2b: Estimated ω values from NEB analysis.

图 2 M8 模型中 Naïve Empirical Bayes (NEB) 分析的参数估测值

Fig. 2 Estimated parameters from Naïve Empirical Bayes (NEB) analysis in model M8



M8模型BEB算法下估算的PSAA蛋白编码区中拥有高 ω 值的130个位点
Candidate sites in PSAA protein after positive selection with high estimated ω in BEB analysis under model M8

a: 130 个位点 ω 值大于 1 时 (即正选择) 的后验概率; b: 130 个位点的 ω 估测值及置信区间。
Fig. 3a: Posterior probability of estimated ω values greater than 1 (positively selected); Fig. 3b: Estimated ω values and their intervals from BEB analysis.

图 3 M8 模型 Bayes Empirical Bayes (BEB) 分析的参数估测值
Fig. 3 Estimated parameters from Bayes Empirical Bayes (BEB) analysis in model M8

表 4 极显著正选择氨基酸位点构建的共进化组
Table 4 Estimated co-evolution groups within positively selected amino acid sites

共进化组 Co-evolution group	组内位点 Positively selected sites in co-evolution groups	共进化组 Co-evolution group	组内位点 Positively selected sites in co-evolution groups
1	20K、108V、494S、510S	9	135V、501A、51D1、619S、707K
2	20K、494S、510S、511D	10	144L、209G、240L、399I
3	20K、510S、626V	11	247L、265N、502G
4	20K、511D、514A	12	247L、501A、502G、510S、619S、707K
5	34R、144L、399I	13	265N、296L、336K、344S、502G
6	100S、399I、484I	14	344S、502G、514A
7	135V、165V、511D、516A、619S、707K	15	484I、493G、638S
8	135V、247L、265N、501A、619S、707K	16	490L、510S、626V

注：表中所有共进化组来自 CAPS2 分析结果，且组内仅包含表中显示的正选择位点。
Notes: All co-evolution groups resulted from CAPS2, with every site in each group positively selected.

支，图 1: c 中该节点的后验概率具备统计学显著性 ($P = 100\%$)，而图 1: a、b 中的后验概率均低于 70%。将来在这些后验概率较低的节点附近的物种应增加叶绿体基因组测序并扩大取样，这将有助于深入了解其物种辐射的真实历史。本研究中厚

囊蕨类植物出现了多歧分化，印证了之前被认为不是单系类群的研究结果^[3]。

较 *rbcL* 基因而言，*psaA* 编码区数据集虽然显示出较好的可作为系统发育关系位点的潜力，但是 *psaA* + *rbcL* 基因联合数据集的系统发育关系构建

能力更优越。将来利用联合数据集比单独仅依靠某一个数据集更有助于明晰蕨类植物的进化动态。本研究结果表明 *psaA* 基因编码区全长具备作为蕨类植物系统发育关系重建位点的潜力, 建议在以后的研究中应更多地取样, 深入探讨。

早先的研究针对真叶植物(蕨类植物、裸子植物和被子植物)的叶绿体基因组曾经有过大规模的筛选, 确定了存在正选择信号的编码基因^[28], 为本研究提供了帮助。当时选取的对象包括蕨类植物的 7 个物种(桫欏、铁线蕨、欧洲蕨、松叶蕨、观音座莲、问荆和碎米蕨), 它们分别来自薄囊蕨类(桫欏、铁线蕨、欧洲蕨和碎米蕨)、厚囊蕨类(观音座莲)、木贼类(问荆)和松叶蕨类(松叶蕨)。然而, 本研究增加蕨类植物物种数目后(增至 27 种), 发现原本没有正选择信号的 *psaA* 基因中也出现了具备统计学显著性的正选择信号。出现这种状况的原因是: 第一, 正选择信号检测手段可能依赖于研究对象物种数目的多少, 物种过少时会因为数据集中包含的信号过于微弱而出现假阴性结果; 第二, *psaA* 基因中的正选择信号可能来自本研究新增加的物种, 特别是瓶儿小草科的 4 个物种(*Mankyua chejuensis*, *Helminthostachys zeylanica*, *Botrychium ternatum* 及 *Ophioglossum californicum*)。上述情况有望将来进一步对瓶儿小草科其它成员的叶绿体基因测序而得到揭示。

先前的研究表明^[4], 在被子植物冠盖之下, 蕨类植物特别是薄囊蕨植物不仅没有趋于灭亡, 相反出现了大规模的物种辐射。随着对 *rbcL* 基因^[8, 9]和 *psbA* 基因^[12]等多种光合系统重要基因的分子进化研究, 已经发现了蕨类植物出现适应性进化的多种分子证据。本研究结果表明, 蕨类植物类群的 PSAA 蛋白中存在着 60 个曾经受正选择的氨基酸位点分布在蛋白的各个结构中, 占蛋白氨基酸位点总数的 8%; 与 *rbcL* 基因^[9](模型 M8, 约 1.8%)及 *psbA* 基因^[12](模型 M8, 约 1.78%)相比, *psaA* 基因编码产物 PSAA 蛋白中具有更多的氨基酸位点经历了正选择。今后借助已知叶绿体基因组提供的大量基因编码区全长信息, 将可以推断光合系统 I 在蕨类植物获得进化优势动态过程中的重要程度。本研究结果表明通过 *psaA* 基因的适应性进化, 蕨类植物的部分类群对光合系统 I 进行了内部调整, 这是蕨类植物类群为寻求自身物种繁衍

而积极响应环境改变的又一重要分子证据。此结果一方面印证了前人对蕨类植物进化潜力的推测, 另一方面也为探讨上述潜力提供了新思路, 即广泛开展对真叶植物各种光系统的分子进化研究。

仅由诸模型共同认定的正选择位点所组成的共进进化组 (Co-evolution groups) 是 PSAA 蛋白质内部氨基酸位点协同进化的分子证据。如本研究中由 29 个正选择位点构成的 16 个共进进化组(表 4)所代表的“共进进化网络”揭示了蕨类植物类群中的光合系统 I, 为了适应高大种子植物兴起而改变的光合作用环境所呈现的进化动态。今后随着对特定蕨类植物类群, 如由被子植物兴起后幸存的合囊蕨科 (Marattiaceae)、瓶尔小草科阴地蕨属 (*Botrychium*) 等早期蕨类植物的深入研究, 将进一步揭示上述适应性进化的内在规律。

参考文献:

- [1] 刘红梅, 王丽, 张宪春, 曾辉. 石松类和蕨类植物研究进展: 兼论国产类群的科级分类系统[J]. 植物分类学报, 2008, 46(6): 808-829.
Liu HM, Wang L, Zhang XC, Zeng H. Advances in the studies of lycophytes and monilophytes with reference to systematic arrangement of families distributed in China [J]. *Journal of Systematics and Evolution*, 2008, 46(6): 808-829.
- [2] 张宪春, 卫然, 刘红梅, 何丽娟, 王丽, 张钢民. 中国现代石松类和蕨类的系统发育与分类系统[J]. 植物学报, 2013, 48(2): 119-137.
Zhang XC, Wei R, Liu HM, He LJ, Wang L, Zhang GM. Phylogeny and classification of the extant lycophytes and ferns from China[J]. *Chinese Bulletin of Botany*, 2013, 48(2): 119-137.
- [3] Smith AR, Pryer KM, Schuettpelz E, Korall P, Schneider H, Wolf PG. A classification for extant ferns[J]. *Taxon*, 2006, 55(3): 705-731.
- [4] Schuettpelz E, Pryer KM. Evidence for a cenozoic radiation of ferns in an angiosperm-dominated canopy[J]. *Proc Natl Acad Sci USA*, 2009, 106(27): 11200-11205.
- [5] Bock DG, Andrew RL, Rieseberg LH. On the adaptive value of cytoplasmic genomes in plants[J]. *Mol Ecol*, 2014, 23(20): 4899-4911.
- [6] Kok B. On the reversible absorption change at 705 mμ in photosynthetic organisms [J]. *Biochim Biophys Acta*, 1956, 22(2): 399-401.
- [7] Ran JH, Shen TT, Liu WJ, Wang XQ. Evolution of the *bHLH* genes involved in stomatal development: implications for the expansion of developmental complexity of sto-

- mata in land plants [J]. *PLoS One*, 2013, 8 (11): e78997.
- [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.
- 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.
- [9] 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.
- Sen L, Su YJ, Zhang B, Wang T. Adaptive evolution of the *rbcl* gene in Pteridaceous ferns [J]. *Journal of Tropical and Subtropical Botany*, 2010, 18(1): 1–8.
- [10] Kapralov MV, Filatov DA. Widespread positive selection in the photosynthetic Rubisco enzyme [J]. *BMC Evol Biol*, 2007, 7(1): 73.
- [11] Sen L, Fares MA, Liang B, Gao L, Wang B, Wang T, Su YJ. Molecular evolution of *rbcl* in three gymnosperm families: Identifying adaptive and coevolutionary patterns [J]. *Biol Direct*, 2011, 6(1): 1–19.
- [12] Sen L, Fares MA, Su YJ, Wang T. Molecular evolution of *psbA* gene in ferns: unraveling selective pressure and coevolutionary pattern [J]. *BMC Evol Biol*, 2012, 12(1): 145.
- [13] 森林, 余坤, 胡志刚, 汪文杰, 徐雷, 刘合刚, 潘宏林. 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 ZG, Wang WJ, Xu L, Liu HG, Pan HL. Molecular evolutionary patterns of the *psbA* gene in gymnosperms [J]. *Journal of Tropical and Subtropical Botany*, 2016, 24(2): 151–159.
- [14] Strehler BL, Lynch VH. Photosynthetic luminescence and photoinduced absorption spectrum changes in *Chlorella* [J]. *Science*, 1956, 123(3194): 462–463.
- [15] Murray JW, Duncan J, Barber J. CP43-like chlorophyll binding proteins: structural and evolutionary implications [J]. *Trends Plant Sci*, 2006, 11(3): 152–158.
- [16] 石开明, 彭昌操, 彭振坤, 罗正荣. DNA 序列在植物系统进化研究中的应用 [J]. 湖北民族学院学报: 自然科学版, 2002, 20(4): 5–10.
- Shi KM, Peng CC, Peng ZK, Luo ZR. Application of DNA sequences in phylogenetic and evolutionary studies of plant [J]. *Journal of Hubei Institute for Nationalities: Natural Sciences Edition*, 2002, 20(4): 5–10.
- [17] Müller KF, Borsch T, Hilu KW. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trnT-F*, and *rbcl* in basal angiosperms [J]. *Mol Phylogenet Evol*, 2006, 41(1): 99–117.
- [18] 安苗苗, 徐丽丽, 姜可以, 许冰清, 杨海芸, 周明兵, 方伟. 花叶矢竹 *psaA* 基因克隆及功能分析 [J]. 竹子研究汇刊, 2015 34(1): 41–48.
- An MM, Xu LL, Jiang KY, Xu BQ, Yang HY, Zhou MB, Fang W. The *psaA* gene cloning and function analysis of *Pseudosasa japonica* cv. Akebonosuji [J]. *Journal of Bamboo Research*, 2015, 34(1): 41–48.
- [19] 余大为, 张海波, 王月, 侯和胜. 坛紫菜叶绿体 *psaA* 基因片段的序列分析 [J]. 水产科学, 2007, 26(5): 289–291.
- She DW, Zhang HB, Wang Y, Hou HS. Sequence analyses of chloroplastic *psaA* gene fragment from *Porphyra haitanensis* [J]. *Fisheries Science*, 2007, 26(5): 289–291.
- [20] Yamada KD, Tomii K, Katoh K. Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees [J]. *Bioinformatics*, 2016, 32(21): 3246–3251.
- [21] Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment [J]. *Nucleic Acids Res*, 2005, 33(2): 511–518.
- [22] Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform [J]. *Nucleic Acids Res*, 2002, 30(14): 3059–3066.
- [23] Santorum JM, Darriba D, Taboada GL, Posada D. jModelTest.org: selection of nucleotide substitution models on the cloud [J]. *Bioinformatics*, 2014, 30(9): 1310–1311.
- [24] Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing [J]. *Nat Methods*, 2012, 9(8): 772.
- [25] Posada D. jModelTest: phylogenetic model averaging [J]. *Mol Biol Evol*, 2008, 25(7): 1253–1256.
- [26] Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7 [J]. *Mol Biol Evol*, 2012, 29(8): 1969–1973.
- [27] Xu B, Yang Z. PAMLX: a graphical user interface for PAML [J]. *Mol Biol Evol*, 2013, 30(12): 2723–4.
- [28] 王博, 高磊, 苏应娟, 王艇. 基于叶绿体基因组全序列分析真叶植物叶绿体基因的适应性进化 [J]. 中山大学学报: 自然科学版, 2012, 51(3): 108–113.
- Wang B, Gao L, Su YJ, Wang T. Adaptive evolutionary analysis of chloroplast genes in Euphyllophytes based on complete chloroplast genome sequences [J]. *Acta Scientiarum Naturalium Universitatis Sunyatseni: Natural Science Edition*, 2012, 51(3): 108–113.