

蕨类植物 *psbD* 基因的适应性进化和共进化分析

许可^{1,2}, 王博^{1,2}, 苏应娟³, 高磊¹, 王艇^{1*}

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

2. 中国科学院大学, 北京 100049; 3. 中山大学生命科学学院, 广州 510275)

摘要: D2 蛋白是植物光系统 II 复合体 (PS II) 核心蛋白之一, 由叶绿体 *psbD* 基因编码。为了深入理解核心薄囊蕨类植物在阴生环境下的“辐射”式演化, 我们对 12 种蕨类植物的 *psbD* 基因进行了克隆和测序, 然后联合已公布的其他 8 种蕨类植物的 *psbD* 序列, 基于 ω 值 (非同义替换率 d_N 和同义替换率 d_S 的比值) 探讨了该基因经受的选择压力。发现 D2 蛋白在大多数分支和位点受到强烈的负选择, 但是树蕨类分支的 *psbD* 进化速率低且 ω 值较高。借助多种模型进行的共进化分析显示, 树蕨类 D2 蛋白的 168R、245H 和 272M 两两组成具有共进化关系的氨基酸位点对。

关键词: 核心薄囊蕨; *psbD* 基因; 选择压力; 共进化

中图分类号: Q941

文献标识码: A

文章编号: 2095-0837(2013)05-0429-10

Molecular Evolution of *psbD* Gene in Ferns: Selection Pressure and Co-evolutionary Analysis

XU Ke^{1,2}, WANG Bo^{1,2}, SU Ying-Juan³, GAO Lei¹, WANG Ting^{1*}

(1. Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: The D2 protein of the Photosystem II (PSII) reaction center complex is encoded by the chloroplast gene *psbD*. To gain a better understanding of the adaptive radiation of core leptosporangiate ferns in low-light environments, we sequenced *psbD* in twelve species and collected eight published sequences to represent extant ferns in order level. We then evaluated the selection pressure of D2 protein based on ω values (nonsynonymous synonymous substitution ratio). The results of statistical tests based on different models indicated that most sites and branches were under strong negative selection, but tree ferns exhibited a slower evolutionary rate and higher ω value compared with other ferns. Co-evolution analysis based on different approaches showed that the D2-R168, D2-H245 and D2-M272 of tree ferns were involved in the same co-evolution network.

Key words: Core leptosporangiates; *psbD* gene; Selective pressure; Co-evolution

一般认为, 伴随被子植物在白垩纪的兴盛, 包括蕨类在内的、非种子植物以外的其他维管植物的多样性和丰富度都极度下降^[1]。然而, 近来有学者利用分子和化石资料相结合的方法研究却发现,

随着被子植物的兴起, 部分蕨类植物因对阴生环境的适应, 其物种多样性反而显著增加^[2]。现存的蕨类植物超过 12000 种, 是微管植物中仅次于被子植物的第二大类群^[3], 不仅形态性状变异极其

收稿日期: 2013-04-17, 修回日期: 2013-06-28。

基金项目: 国家自然科学基金资助项目 (31070594); 湿地演化与生态恢复湖北省重点实验室开放课题。

作者简介: 许可 (1987-), 男, 硕士研究生, 主要从事植物进化遗传学研究 (E-mail: xuke09@mails.ucas.ac.cn)。

* 通讯作者 (Author for correspondence. E-mail: tingwang@wbgcas.cn)。

多样, 而且具有陆生、水生、石生和附生等多种类型。考虑到不同生境的环境因子, 特别是光条件的明显差异, 推测蕨类植物在光合作用中起关键作用的一些功能基因, 极有可能伴随物种的“辐射”式分化而发生适应性进化。

光系统 II 反应是植物吸收光能、进行光诱导的电荷分离、产生电子传递并催化水光解的过程^[4], 该过程是在光系统 II (PS II) 复合体中完成^[5]。PS II 复合体的核心是由两个结构相似的 D1 和 D2 蛋白组成的光反应中心, 其中 D2 蛋白由叶绿体基因 *psbD* 编码^[6]。D2 蛋白在维持 PS II 反应中心构像的稳定和电子传递方面起重要作用, 同时还参与调节 D1 蛋白的表达^[7]。由于自然选择的作用, 功能重要的蛋白质在进化过程中通常会受到较强的选择压力。为度量分子水平的选择压力, 可估算蛋白编码基因序列的核苷酸非同义替换率(d_N)和同义替换率(d_S)的比值 ω ($\omega = d_N / d_S$), 作为选择压力的指标^[8]。另一方面, 同一蛋白质的氨基酸残基之间还可能在进化上存在相关性: 重要氨基酸残基在进化过程中发生的替换有可能因位于其他位点的残基的替换而得到补偿, 从而维持蛋白质结构和功能的稳定。这些进化上相关的氨基酸残基, 被称为共进化残基 (co-evolving residues)^[9, 10]。鉴定共进化残基, 有助于深化对适应性进化复杂机制的理解^[11]。

本研究根据 *psbD* 基因侧翼的保守区序列设计引物, 克隆并测定了 12 种蕨类植物的 *psbD* 基因序列, 联合其他已经公布的蕨类植物 *psbD* 基因序列, 进而对 *psbD* 基因及其编码的 D2 蛋白展开了适应性进化和共进化分析。目的是: 第一, 检测 *psbD* 基因在蕨类不同分支所受的选择压力; 第二, 鉴定 D2 蛋白发生适应性进化的氨基酸位点; 第三, 揭示蕨类植物 D2 蛋白氨基酸位点的共进化式样。

1 材料和方法

1.1 植物材料

实验所用的植物材料采自中国科学院武汉植物园 (WBG-CAS), 中国科学院华南植物园 (SCBG-CAS) 和深圳市中国科学院仙湖植物园 (FBG-

SCAS), 另外由 GenBank 数据库 (<http://www.ncbi.nlm.nih.gov/>) 获得了 8 种蕨类植物以及 1 种外类群的 *psbD* 基因序列 (表 1)。取样的原则是根据 Smith 等^[12] 和刘红梅等^[13] 对蕨类植物的分类,

表 1 材料来源和 GenBank 登录号
Table 1 Source of materials and GenBank accession numbers

种名 Species	采集地点 ** Location	GenBank 登录号 GenBank accession No.
鹿角蕨 <i>Platyserium bifurcatum</i> (Cav.) C.Chr.	WBG-CAS	KC831598
巢蕨 <i>Asplenium australasicum</i> (J.Sm.) Hook.	WBG-CAS	KC831602
欧洲蕨 * <i>Pteridium aquilinum</i> (L.) Kuhn		HM535629
碎米蕨 * <i>Cheilanthes lindheimeri</i> Hook.		HM778032
铁线蕨 * <i>Adiantum capillus-veneris</i> L.		NC_004766
瘤足蕨 <i>Plagiogyria japonica</i> Nakai	FBG-CAS	KC831599
桫欏 * <i>Alsophila spinulosa</i> (Hook.) R. M. Tryon		NC_012818
笔筒树 <i>Sphaeropteris lepifera</i> (Hook.) R. M. Tryon	WBG-CAS	KC831594
苹 <i>Marsilea quadrifolia</i> L.	WBG-CAS	KC831603
蜂巢槐叶蕨 <i>Salvinia molesta</i> Mitchell	WBG-CAS	KC831595
满江红 <i>Azolla caroliniana</i> Willd.	WBG-CAS	KC831600
海金沙 * <i>Lygodium japonicum</i> (Thunb.) Sw.		EU328240
芒萁 <i>Dicranopteris linearis</i> (Burm.) Underw.	FBG-CAS	KC831601
漏斗瓶蕨 <i>Vandenboschia radicans</i> (Sw.) Cop.	WBG-CAS	KC831597
华南紫萁 <i>Osmunda vachellii</i> Hook.	SCBG-CAS	KC831605
观音座莲 * <i>Angiopteris evecta</i> (J. R. Forst.) Hoffmann		NC_008829
松叶蕨 * <i>Psilotum nudum</i> (L.) Beauv.		NC_003386
劲直阴地蕨 <i>Botrychium strictum</i> Underw.	WBG-CAS	KC831596
七指蕨 <i>Helminthostachys zeylanica</i> (L.) Hook.	SCBG-CAS	KC831604
问荆 * <i>Equisetum arvense</i> L.		NC_014699
水韭 * <i>Isoetes flaccida</i> Shuttlew. ex A. Braun		NC_014675

注: *, 表示由 GenBank 获得的数据; **, WBG-CAS—中国科学院武汉植物园, SCBG-CAS—中国科学院华南植物园, FBG-CAS—深圳市中国科学院仙湖植物园。
Notes: * Sequences data were downloaded from GenBank; ** WBG-CAS—Wuhan Botanical Garden, Chinese Academy of Sciences, SCBG-CAS—South China Botanical Garden, Chinese Academy of Sciences, FBG-CAS—Fairylake Botanical Garden, Shenzhen and Chinese Academy of Sciences.

使得材料样本覆盖到蕨类植物的目级水平; 物种丰富度较高核心薄囊蕨类植物的水龙骨科(Polypodiales)、树蕨目(Cyatheaales)和水生异型孢子蕨目(Salviniales), 我们分别选取了5种、3种和3种代表植物。

1.2 总DNA提取、PCR、克隆和测序

采用植物基因组DNA抽提试剂盒(TIANGEN, 北京)对材料进行总DNA的提取, 实验步骤按照试剂盒说明书进行。根据*psbD*基因侧翼的保守区设计两对PCR引物: F1(5'-TCCYTGTTCATACATAGGCTTYTC-3')和R1(5'-ATGGTAAGGCGTAAGTCGTC-3')、F2(5'-CGTGGGCACCAAGTAATTTACCAGA-3')和R2(5'-ATGGTAAGGCGTAAGTCGTC-3')。引物由上海英骏生物技术有限公司(上海)合成。每50 μ L PCR反应体系包括: 2.5 ng DNA模板、正反向引物各2 μ L(10 μ mol/L)、1.5 U LA *Taq* DNA聚合酶(TAKARA, 大连)和10倍PCR缓冲液5 μ L、8 μ L dNTPs。PCR反应程序为: ①预变性94 $^{\circ}$ C 7 min, ②变性94 $^{\circ}$ C 30 s, ③退火和延伸62 $^{\circ}$ C 3 min, ④后延伸72 $^{\circ}$ C 10 min; 其中②~③重复35个循环。将获得的PCR产物纯化后与PCR 2.1载体连接, 热激转化至大肠杆菌DH5 α , 然后挑选阳性克隆进行PCR鉴定。随机挑选3个阳性克隆送华大生物技术有限公司(武汉)进行测序。

1.3 系统发育树的构建

利用MUSCLE软件^[14]对序列进行比对, 采用Modeltest 3.7软件^[15]选取最适核苷酸进化模型。运行MrBayes 3.2软件^[16]经贝叶斯途径推测系统发育关系, 后验概率依Metropolis-Hastings-Green算法通过马尔可夫链(Markov Chain Monte Carlo, MCMC)运行1000000代估计。MCMC分析以随机树起始, 每100代保存一棵树。使用Tracer v1.5软件^[17]检验收敛程度; 当所有参数的有效取样大小(Efficient Sampling Size, ESS)值均大于500时, 认为迭代运算达到收敛。摒弃老化样本(Burnin = 2500), 由剩余样本构建一致树。

1.4 适应性进化分析

主要利用PAML 4.5软件^[18]完成分析。采用

3种模型: (1)分支模型^[19], 它允许非同义替换和同义替换的比值 ω 在不同支系上有变化, 其中单比率模型(One ratio model)假定所有进化支的 ω 值都是一致的, 而自由比率模型(Free ratio model)假定各支的 ω 值各不同, 还有居于两者之间的二比率模型(Two ratio model)假定前景支与背景支的 ω 值不同, 自由比率模型和二比率模型分别与单比率模型进行似然比检验(Likelihood ratio test, LRT), 检验模型之间是否有显著性差异; (2)位点模型^[8,20]假设不同位点有不同的选择压力, 模型M2a(选择)、M3(离散)及M8(beta和 ω)的假定条件中均允许存在 ω 值大于1的位点, 与各自对应的零假设模型M1a(近中性)、M0(单一比率)及M7(beta)进行LRT分析, 通过比较模型间差异的显著性来检验正选择位点; (3)分支-位点模型^[21]将进化树上的支系分成前景支和背景支两类, 仅仅允许前景支有正选择位点, 采用分支-位点模型A的检验2检验前景支是否存在正选择位点。

此外, 还基于机理式和经验式整合模型(Mechanistic-empirical combination model, MEC)^[22]、固定效应似然模型(Fixed effects likelihood model, FEL)和单一似然祖先计数法(Single likelihood ancestor counting, SLAC)^[23]进行分析。MEC分析利用网络服务器Selecton(<http://selection.tau.ac.il/>)完成, FEL和SLCA分析采用Datamonkey服务器(<http://www.datamonkey.org/>)完成。

1.5 共进化分析

采用Pearson相关系数法^[24]、参数检验法^[25]和互信息法(Mutual Information)^[26]等对氨基酸残基的共进化关系进行了检测。通过网络服务器CAPS(<http://bioinf.gen.tcd.ie/caps/>)^[27]、Spidermonkey^[28](<http://www.datamonkey.org/>)和Intermap3D^[29](<http://www.cbs.dtu.dk/services/InterMap3D/>)完成分析。

2 结果

2.1 系统发育关系树的构建

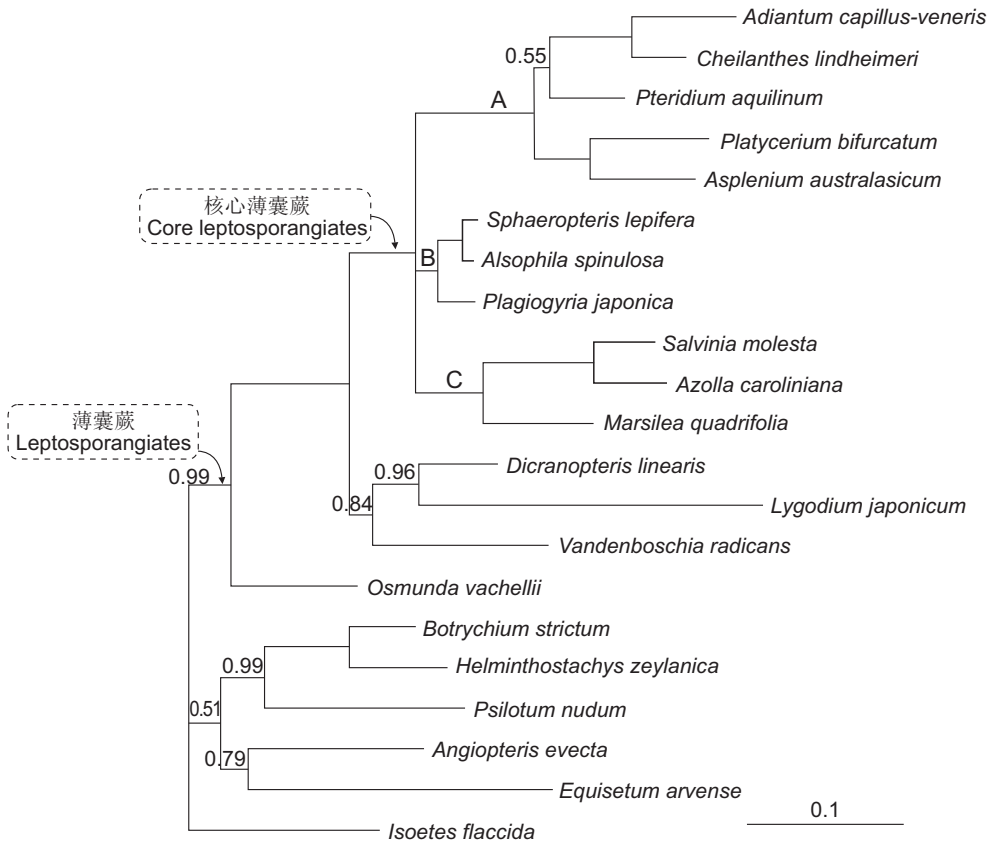
测得的12种蕨类植物的*psbD*基因的ORF

区长度均为 1059 bp，编码 353 个氨基酸。序列提交至 GenBank，登录号为：KC831594-KC831605(表 1)。以石松类植物水韭(*Isoetes flaccida*)为外类群，利用 *psbD* 基因序列基于贝叶斯法重建了 20 种蕨类植物的系统发育树(图 1)，该系统树的拓扑结构与 Smith 等^[12]的报道基本一致。木贼科的问荆(*Equisetum arvense*)、合囊蕨科的观音座莲(*Angiopteris evecta*)、松叶蕨科的松叶蕨(*Psilotum nudum*)以及瓶儿小草科的劲直阴地蕨(*Botrychium strictum*)和七指蕨(*Helminthostachys zeylanica*)都是蕨类中比较原始的基部类群，聚在一起。薄囊蕨类植物的早期成员华南紫萁(*Osmunda vachellii*)独自构成一支，占据薄囊蕨类的最基部位置；随后是漏斗瓶蕨(*Vandenboschia radicans*)与互为姊妹群的芒萁(*Dicranopteris linearis*)和海金沙(*Lygodium*

japonicum)构成一个单系分支(后验概率 0.84)。核心薄囊蕨的水生异型孢子蕨类(以下简称：水生蕨类)、树蕨类和水龙骨科聚在一起构成单系分支(后验概率为 1.00)。

2.2 选择压力检测
2.2.1 运用 PAML 检测选择压力

以用贝叶斯法构建的进化树为树文件，利用 PAML 软件对 *psbD* 基因进行进化分析。分支模型中，单比率模型涉及 40 个参数，似然值 $\ell_0 = -6641.19$ ，估测的 ω 值为 0.0165；自由比率模型包括 77 个参数，似然值 $\ell_1 = -6580.19$ (表 2)。以 $2\Delta\ell = 2(\ell_1 - \ell_0) = 122.00$ ，进行自由度为 37 的 χ^2 的比较，结果显示自由比率模型显著优于单比率模型($p < 0.01$)(表 3)，这提示各支系的 ω 值不同，选择压力在支系间异质。以树蕨类分支(B)为前景支，其他分支的为背景支，运行二比率模型



分支上的数字为后验概率，后验概率为 1.00 的未标示，A、B 和 C 分别代表水龙骨科、树蕨类和水生蕨类。
Posterior probability values are given above branches, the value of 1.00 are not displayed. A, B and C stand for Polypodiales, Cyatheaales and Salviniales.

图 1 基于蕨类 *psbD* 基因序列数据利用贝叶斯法得出的一致树
Fig. 1 Consensus tree resulting from the Bayesian inference of *psbD* gene sequences of ferns

表 2 *psbD* 基因在不同模型下的参数估计值和対数似然值
Table 2 Parameter estimates and log-likelihood values for *psbD* gene under different models

模型 Models	参数个数 p^*	似然值 ℓ	参数估计值 Estimated value of parameters	正选择位点 Positively selected sites
分支模型 Branch model	单一比率 (M0)	40	$\omega = 0.01647$	不允许 Not allowed
	二比率 A (MA)	41	$\omega_0 = 0.01862, \omega_1 = 0.00907$	不允许 Not allowed
	二比率 B (MB)	41	$\omega_0 = 0.014503, \omega_1 = 0.08188$	不允许 Not allowed
	二比率 C (MC)	41	$\omega_0 = 0.01784, \omega_1 = 0.00748$	不允许 Not allowed
	自由比率 F (MF)	77	$\omega_a = 0.0001, \omega_b = 0.2174, \omega_c = 0.0333$	不允许 Not allowed
位点模型 Site model	Model 1a (M1a): 近中性	41	$p_0 = 0.98951, \omega_0 = 0.01402, p_1 = 0.01049, \omega_1 = 1.00000$	不允许 Not allowed
	Model 2a (M2a): 正选择	43	$p_0 = 0.98951, \omega_0 = 0.01402, p_1 = 0.01049, \omega_1 = 1.00000, p_2 = 0.00000, \omega_2 = 48.75004$	无 None
	Model 3 (M3): 离散	44	$p_0 = 0.69673, \omega_0 = 0.00000, p_1 = 0.29046, \omega_1 = 0.04699, p_2 = 0.01281, \omega_2 = 0.34518$	无 None
	Model 7 (M7): beta	41	$p = 0.17182, q = 8.47194$	不允许 Not allowed
	Model 8 (M8): beta 和 $\omega > 1$	43	$p_0 = 0.99999, p = 0.17182, q = 8.47190, p_1 = 0.00001, \omega = 2.60238$	无 None
分支-位点模型 Branch-site model	水龙骨分支 (Ma0): ω_2 固定为 1	42	$p_0 = 0.98951, \omega_0 = 0.01402, p_1 = 0.01049, \omega_1 = 1.00000, p_2 + p_3 = 0.00000, \omega_2 = 1.00000$	不允许 Not allowed
	水龙骨分支 (Ma): ω_2 为估计值	43	$p_0 = 0.98951, \omega_0 = 0.01402, p_1 = 0.01049, \omega_1 = 1.00000, p_2 + p_3 = 0.00000, \omega_2 = 1.00000$	无 None
	树蕨分支 (Mb0): ω_2 固定为 1	42	$p_0 = 0.93391, \omega_0 = 0.01280, p_1 = 0.00924, \omega_1 = 1.00000, p_2 + p_3 = 0.05685, \omega_2 = 1.00000$	不允许 Not allowed
	树蕨分支 (Mb): ω_2 为估计值	43	$p_0 = 0.93391, \omega_0 = 0.01280, p_1 = 0.00924, \omega_1 = 1.00000, p_2 + p_3 = 0.05685, \omega_2 = 1.00000$	168R (50.0%)
	水生蕨分支 (Mc0): ω_2 固定为 1	42	$p_0 = 0.98951, \omega_0 = 0.01402, p_1 = 0.01049, \omega_1 = 1.00000, p_2 + p_3 = 0.00000, \omega_2 = 1.00000$	不允许 Not allowed
	水生蕨分支 (Mc): ω_2 为估计值	43	$p_0 = 0.98951, \omega_0 = 0.01402, p_1 = 0.01049, \omega_1 = 1.00000, p_2 + p_3 = 0.00000, \omega_2 = 1.00000$	无 None

注: p^* 为 ω 分布中参数的个数; ℓ 为最大似然值的対数。
Notes: p^* is number of parameters in ω distribution; ℓ means log maximum likelihood value.

进行检测, 结果显示树蕨分支的 ω 值 是其他分支的 5 倍多 ($0.08188/0.014503 \approx 5.65$); 秩和检验结果表明树蕨类分支的非同义替换速率 (d_N) 与其他蕨类植物差异不显著 ($p > 0.05$), 而同义突变速率 (d_S) 极显著降低 ($p < 0.01$), 进而导致树蕨类植物分支的 ω 值较高 (图 2)。

对 D2 蛋白编码序列位点间选择压力检测结果表明, M3 模型显著优于 M0 模型 ($p < 0.01$) (表 3),

表 3 PAML4.5 软件中不同模型的
似然比值检验统计量 ($2\Delta\ell$)

Table 3 Likelihood ratio statistics of different
models in PAML version 4.5 ($2\Delta\ell$)

模型比较 Comparison of models	自由度 <i>d.f.</i>	$2\Delta\ell$	<i>p</i> 值 <i>p</i> value
M0 & MA	1	5.08	<0.05 *
M0 & MB	1	16.90	<0.01 **
M0 & MC	1	4.11	<0.05 *
M0 & MF	37	122.00	<0.01 **
M1a & M2a	2	0.00	1.00
M0 & M3	4	53.03	<0.01 **
M7 & M8	2	0.00	1.00
Ma0 & Ma	1	0.00	1.00
Mb0 & Mb	1	0.00	1.00
Mc0 & Mc	1	0.00	1.00

注: MA 和 Ma, MB 和 Mb, MC 和 Mc 分别代表:以水龙骨科分支、树蕨类分支和水生蕨类分支为前景支的二比率模型和分支位点模型;MF 代表自由比率模型; ** 代表极显著, * 代表显著。

Notes: MA and Ma, MB and Mb, MC and Mc, stands for poly-podiales branches, cyatheales branches and salviniales branches as the foreground branches of two ratio model and branch-site model, respectively; MF means freed-ratio model; ** indicates extremely significant difference, * indicates significant difference.

说明位点间承受的选择压力具有异质性;而 M2a 和 M8 不优于各自对应的零模型 M1a 和 M7 ($p>0.05$) (表 3), 因此这两对模型进行比较无法检验正选择。模型 M3 中 3 类密码子位点的 p_0 、 p_1 和 p_2 分别对应的 ω 值为 0.00000、0.04699 和 0.34518, 说明 98.7%左右的位点处于较强的负选择压力(表 2)。

由于核心薄囊蕨(水龙骨科, 树蕨类和水生蕨类)占蕨类总数 80%以上^[30], 分布范围更加广泛, 生境也更加多样化(包含旱生、水生和附生等^[31,32]), 为了检测核心薄囊蕨的 *psbD* 基因是否经历插曲式自然选择, 分别设置水龙骨科(a)、树蕨类(b)和水生异型蕨类(c)为前景支, 采用分支-位点模型进行正选择位点的检测。分支 a 和分支 c 上没有检测出正选择位点, 分支 b 虽然检测到 168R 一个正选择位点(后验概率为 50.0%), 但似然比检验拒绝存在正选择位点的假设(表 3), 因此分支 b 的检测不能作为可信的正选择证据。

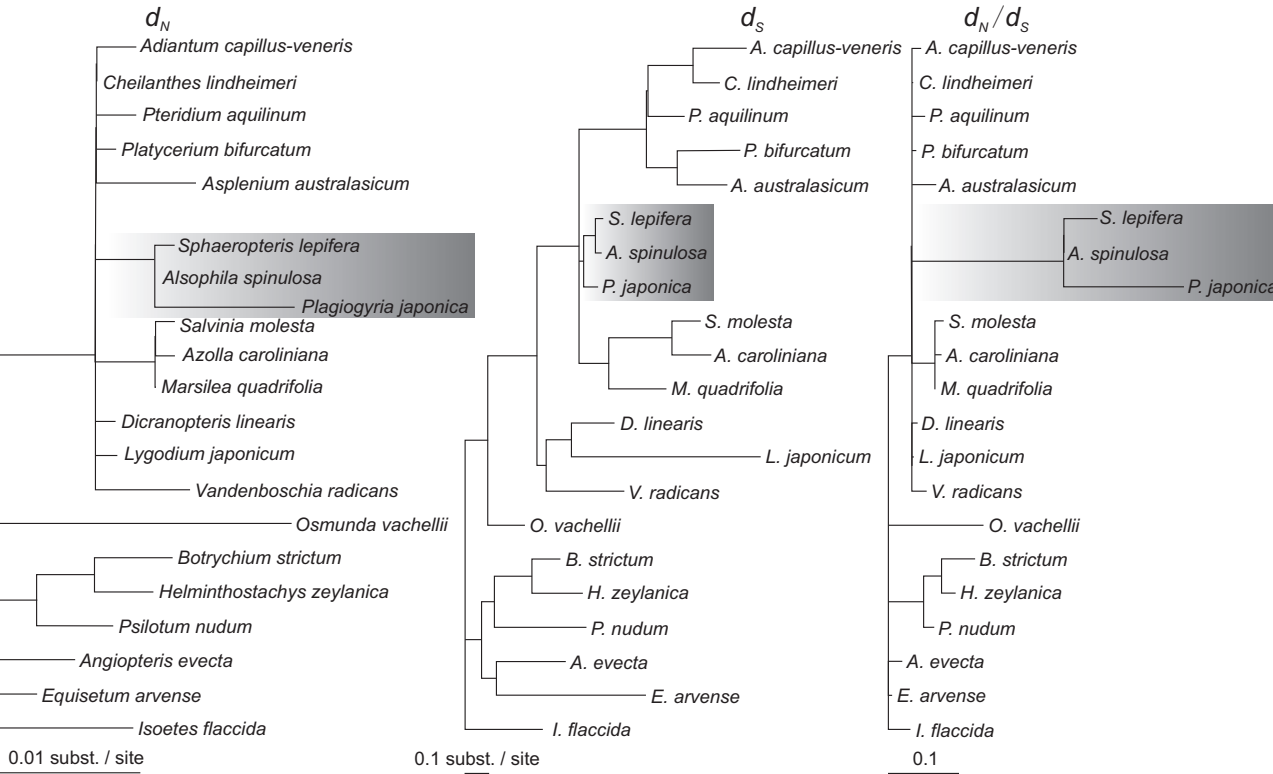


图 2 自由比率模型下的选择压力检测结果, 支长分别代表 d_N 、 d_S 和 ω ($\omega=d_N/d_S$) 值, 阴影区域代表树蕨类

Fig. 2 Selective pressure based on free-ratio model, branch length indicates d_N , d_S and ω ($\omega=d_N/d_S$) values, tree ferns are shown in shaded regions

2.2.2 基于 Selecton 和 Datamonkey 检测选择压力

2.2.2.1 利用 Selecton 的 MEC 模型分析选择压力

根据 MEC 模型计算出的各个位点 $\omega > 1$ 的 BEB 后验概率以及 ω 值绘制成图 3，总长为 353 个氨基酸位点，在后验概率大于 95% 条件下，以 $\omega \leq 0.21$ 为标准的负选择位点有 349 个，占总序列的 98.9%；以 $\omega > 1$ 为标准在后验概率大于 80% 的条件下，没有正选择位点。

2.2.2.2 基于 Datamonkey 检测选择压力

利用 SLAC 模型和 FEL 模型，在 $p = 0.05$ 的标准下，未检测到正选择位点，分别鉴定出 234 个和 268 个负选择位点。这提示负选择起主导作用。

2.3 D2 蛋白序列的共进化分析

采用多种方法在 D2 蛋白氨基酸序列内部共鉴定出 13 对共进化位点对。CAPS 方法鉴定出 1 对共进化位点对，它们的共进化方式与蛋白质的疏水性和分子量均呈极显著相关 ($p < 0.01$)。Spider-monkey 模块共检测 5 对共进化位点对，鉴定过程如下：经贝叶斯图模型 (Bayesian graphical models, BGMs) 进行预处理，在 25 个分支的 57 个位点上发现有 91 个非同义替换存在潜在的共进化关系，在 $p = 0.5$ 的水平，进一步运行 BGMs 模

型的单亲模型 (One parent) 和双亲模型 (Two parent) 共鉴别出 5 对共进化对。最后利用 InterMap 3D 中的行列加权互信息 (RCW-MI)、互信息-熵比 (MI/E) 和依赖性 (Dependency) 3 个模型共鉴定出 10 对共进化位点对 (表 4)。其中有 4 对共进化位点在两种服务器均被检测到。

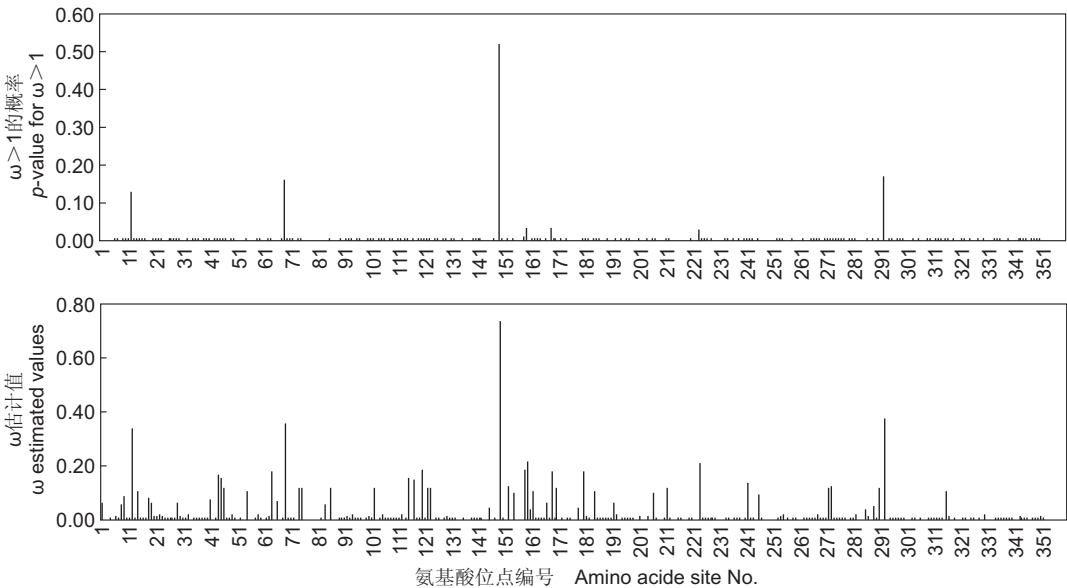
3 讨论

3.1 树蕨类植物分子进化速率降低

利用 *psbD* 基因构建的系统发育树中树蕨类植物分支 (B) 的支长较短 (图 1)，说明树蕨类植物 *psbD* 基因的核苷酸替换速率低于其他蕨类植物，树蕨类进化速率降低不仅存在于叶绿体基因^[33]，也存在于核基因和线粒体基因^[34]，可能因为：(1) 树蕨类的世代周期较长，如瘤足蕨属 (*Plagiogyria*) 的世代周期为 15 年^[35]，因生殖引起的核苷酸替换的几率较小，同样的现象也存在于被子植物中，多年生植物进化速率低于一年生种类^[36]；(2) 树蕨类植物大多是零星式分布，有效种群规模较小^[37]，缺乏有效的基因交流，种群中的遗传变异容易因漂变而丢失。

3.2 不同分支上自然选择作用

侏罗纪时代蕨类植物曾在地球上极为繁盛，伴



氨基酸位点编号参照笔筒树 *psbD* 基因序列。
Sites are numbered according to the *psbD* sequence of *Cyathea lepifera*.

图3 MEC 所估计的参数值
Fig. 3 Estimated values of parameters under MEC

表 4 D2 蛋白的共进化对
Table 4 Co-evolution pairs with the D1 protein

服务器 Websites	共进化对 Co-evolutionary pairs	
CAPS	18	19
Spidermonkey	168	245
	180	64
	210	285
	241	178
	288	19
	45	115
InterMap3D	41	122
	168	245
	168	272
	245	272
	19	288
	8	14
	8	84
	64	180
	18	117
	45	115

注：字体加粗的数字表示被一种以上服务器鉴定出的共进化位点。
Note: Numbers in bold indicate that sites have been detected more than once.

随着被子植物在白垩纪的兴起，陆地生态系统开始发生明显改变。水龙骨科祖先和 水生蕨类祖先通过生态型分化(石生、附生和泽生等)，大部分栖息于被子植物难于生存的环境^[3]，而树蕨类植物与部分被子植物生境重叠(如发现桫欏和竹子部分生态位重叠^[37])，争夺光的环境压力更大。在分子水平进行选择压力检测证实，树蕨类植物的 PS II 复合体 D2 蛋白受到选择压力显著不同于其他蕨类。

3.3 具有重要功能的位点进化

本研究采用的位点模型和分支位点模型均未检测到正选择位点。可能是因为 *psbD* 基因序列数据包含的变异太少，适应性信号被淹没于极其普遍的净化选择之中。其次，从进化的时间上考虑，核心薄囊蕨的水龙骨科、树蕨类和水生蕨类在 2.1 ~ 2.2 亿年前已经分化^[2]，如果适应性进化发生在早期并已经被固定下来，鉴定单个或者是少数几个的正选择会更加困难。所以，即使 *psbD* 基因编码的 D2 蛋白在核心薄囊蕨分化的早期发生过正选择，但是由于有可能被掩盖在不影响其功能的中性替换中，因此需要开发新的方法检测功能重要位点的适应进化。

本研究利用多种方法鉴定出蕨类植物 D2 蛋白

分子内 13 对共进化位点，这些位点为理解蕨类植物 D2 进化机制提供了重要信息，如鉴定出的 45S 和 115I 位点，在水生蕨类中共进化，这两个位点对应的密码子均是回复突变，45S 与水韭的一致，115I 与蓝藻的一致。考虑到水韭和蓝藻均为水生植物，这两个位点的回复突变可能与水生生活形态有关。

2011 年，Umena 等报道了蓝藻(*Thermosynechococcus vulcanus*) PSII 复合体分辨率为 1.9Å 的晶体结构，表明 D2-Y244 与 Fe⁺² 结合相关^[38]，该位点对其结构的稳定和功能的实现具有重要意义。然而树蕨类植物与之相应的位点 245，由疏水性酪氨酸(Y)替换成碱性组氨酸(H)，考虑到正电荷与负电荷之间可形成离子键，该位点的突变可能加强了与 Fe⁺² 的结合。共进化分析显示，树蕨类植物 168R、272T 和 245H 构成的共进化网络，一方面说明 245H 位点很重要，该位点的突变需要其他位点进行补偿以维持 PSII 复合体结构的稳定和功能的实现；另一方面说明蛋白质水平分子进化的作用机制复杂，单个具有重要功能的氨基酸位点改变可能引起连锁反应。

参考文献：

[1] Crane P R, Friis E M, Pedersen K R. The origin and early diversification of angiosperms [J]. *Nature*, 1995, 374(6517): 27–33.

[2] Schneider H, Schuettpelz E, Pryer K M, Cranfill R, Magallon S, Lupia R. Ferns diversified in the shadow of angiosperms[J]. *Nature*, 2004, 428(6982): 553–557.

[3] Page C N. Ecological strategies in fern evolution: a neopteridological overview [J]. *Rev Palaeobot Palyno*, 2002, 119(1–2): 1–33.

[4] 孙瑞雪, 杨春虹. 光系统 II 的结构与功能以及光合膜对环境因素的响应机制[J]. *生物物理学报*, 2012, 28(7): 537–548.

[5] Shi L X, Hall M, Funk C, Schroder W P. Photosystem II , a growing complex: updates on newly discovered components and low molecular mass proteins[J]. *Biochim Biophys Acta*, 2012, 1817: 13–25.

[6] Marder J B, Chapman D J, Telfer A, Nixon P J, Bar-

- ber J. Identification of *psbA* and *psbD* Gene-Products, D1 and D2, as reaction center proteins of photosystem-2 [J]. *Plant Mol Biol*, 1987, 9(4): 325–333.
- [7] Diner B A, Rappaport F. Structure, dynamics, and energetics of the primary photochemistry of photosystem II of oxygenic photosynthesis [J]. *Annu Rev Plant Biol*, 2002, 53: 551–580.
- [8] Yang Z, Bielawski J P. Statistical methods for detecting molecular adaptation [J]. *Trends Ecol Evol*, 2000, 15(12): 496–503.
- [9] Süel G M, Lockless S W, Wall M A, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins [J]. *Nature Struct Biol*, 2002, 10: 59–69.
- [10] 吕品一, 郑珩, 劳兴珍. 蛋白质共进化分析研究进展 [J]. *生物信息学*, 2010(1): 34–37.
- [11] Fares M A. Computational and statistical methods to explore the various dimensions of protein evolution [J]. *Current Bioinformatics*, 2006, 1(2): 207–217.
- [12] Smith A R, Pryer K M, Schuettpelz E, Korall P, Schneider H, Wolf P G. A classification for extant ferns [J]. *Taxon*, 2006, 55(3): 705–731.
- [13] 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.
- [14] Edgar R C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity [J]. *BMC Bioinformatics*, 2004, 5: 113.
- [15] Posada D, Crandall K. Modeltest: testing the model of DNA substitution [J]. *Bioinformatics*, 1998, 14(9): 817–818.
- [16] Ronquist F, Teslenko M, Van Der Mark P, Ayres D L, Darling A, Höhna S, Larget B, Liu L, Suchard M A, Huelsenbeck J P. MrBayes 3. 2: Efficient bayesian phylogenetic inference and model choice across a large model space [J]. *Syst Biol*, 2012, 61(3): 539–542.
- [17] Rambaut A, Drummond A. Tracer v1.5 [CP]. <http://tree.bio.ed.ac.uk/software/tracer/> [2009-11-30].
- [18] Yang Z H. PAML 4: Phylogenetic analysis by maximum likelihood [J]. *Mol Biol Evol*, 2007, 24(8): 1586–1591.
- [19] Yang Z H. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution [J]. *Mol Biol Evol*, 1998, 15(5): 568–573.
- [20] Nielsen R, Yang Z H. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene [J]. *Genetics*, 1998, 148(3): 929–936.
- [21] Zhang J Z, Nielsen R, Yang Z H. 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.
- [22] Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach [J]. *Nucleic Acids Res*, 2007, 35: 506–511.
- [23] Delpont W, Poon A F Y, Frost S D W, Pond S L K. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology [J]. *Bioinformatics*, 2010, 26(19): 2455–2457.
- [24] Göbel U, Sander C, Schneider R, Valencia A. Correlated mutations and residue contacts in proteins [J]. *Protein: Struct Funct Genet*, 1994, 18(4): 309–317.
- [25] Pollock D D, Taylor W R, Goldman N. Coevolving protein residues: maximum likelihood identification and relationship to structure [J]. *J Mol Biol*, 1999, 287(1): 187–198.
- [26] Martin L, Gloor G, Dunn S, Wahl L. Using information theory to search for co-evolving residues in proteins [J]. *Bioinformatics*, 2005, 21(22): 4116–4124.
- [27] Fares M A, Mcnally D. CAPS: coevolution analysis using protein sequences [J]. *Bioinformatics*, 2006, 22(22): 2821–2822.
- [28] Poon A F, Lewis F I, Frost S D, Pond S L K. Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models [J]. *Bioinformatics*, 2008, 24(17): 1949–1950.
- [29] Gouveia-Oliveira R, Roque F S, Wernersson R, Sicheritz-Ponten T, Sackett P W, Molgaard A, Pedersen A G. InterMap3D: predicting and visualizing

- co-evolving protein residues [J]. *Bioinformatics*, 2009, 25(15): 1963–1965.
- [30] Pryer K M, Schuettpelz E, Wolf P G, Schneider H, Smith A R, Cranfill R. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences [J]. *Am J Bot*, 2004, 91(10): 1582–1598.
- [31] Rothfels C J, Windham M D, Grusz A L, Gastony G J, Pryer K M. Toward a monophyletic *Notholaena* (Pteridaceae): resolving patterns of evolutionary convergence in xeric-adapted ferns [J]. *Taxon*, 2008, 57(3): 712–724.
- [32] Schuettpelz E, Pryer K M. Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy [J]. *Proc Natl Acad Sci USA*, 2009, 106(27): 11200–11205.
- [33] Korall P, Schuettpelz E, Pryer K M. Abrupt deceleration of molecular evolution linked to the origin of arborescence in ferns [J]. *Evolution*, 2010, 64(9): 2786–2792.
- [34] Wikstrom N, Pryer K M. Incongruence between primary sequence data and the distribution of a mitochondrial *atp1* group II intron among ferns and horsetails [J]. *Mol Phylogenet Evol*, 2005, 36(3): 484–493.
- [35] Windisch P G, Pereira-Noronha M. Notes on the ecology and development of *Plagiogyria fialhoi* [J]. *Am Fern J*, 1983: 79–84.
- [36] Smith S A, Donoghue M J. Rates of molecular evolution are linked to life history in flowering plants [J]. *Science*, 2008, 322(5898): 86–89.
- [37] Wang T, Su Y, Li Y. Population genetic variation in the tree fern *Alsophila spinulosa* (Cyatheaceae): effects of reproductive strategy [J]. *Plos One*, 2012, 7(7): e41780.
- [38] Umena Y, Kawakami K, Shen J R, Kamiya N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å [J]. *Nature*, 2011, 473(7345): 55–60.

(责任编辑: 王豫鄂)