

DOI:10.11913/PSJ.2095-0837.2020.10001

平晶耀, 祝铭, 苏应娟, 王艇. 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): 1-9Ping JY, Zhu M, Su YJ, Wang T. Molecular evolution of chloroplast gene *rps12* in ferns[J]. *Plant Science Journal*, 2020, 38(1): 1-9

# 蕨类植物叶绿体 *rps12* 基因的分子进化研究

平晶耀<sup>1</sup>, 祝铭<sup>1</sup>, 苏应娟<sup>2,3</sup>, 王艇<sup>1,4\*</sup>

(1. 华南农业大学生命科学学院, 广州 510642; 2. 中山大学生命科学学院, 广州 510275; 3. 中山大学深圳研究院, 广东深圳 518057; 4. 中国科学院华南植物园广东省应用植物学重点实验室, 广州 510650)

**摘要:** 以 68 种蕨类植物和 2 种石松类植物的 *rps12* 基因为对象, 在系统发育背景下, 结合最大似然法, 使用 HyPhy 和 PAML 软件对该基因进行进化速率和适应性进化研究。结果显示: 位于 IR 区的外显子 2~3, 其替换率明显降低, *rps12* 基因编码序列的替换率也随之降低, 且 *rps12* 基因密码子第 3 位的 GC 含量明显升高; 在蕨类植物的进化过程中, 3'-*rps12* 更倾向定位于 IR 区, 以保持较低的替换率; *rps12* 基因编码的 123 个氨基酸位点中, 共检测到 4 个正选择位点和 116 个负选择位点。研究结果表明基因序列进入到 IR 区后, 显示出降低的替换率; 强烈的负选择压力表明 RPS12 蛋白的高度保守性以及 *rps12* 基因的功能和结构已经趋于稳定。

**关键词:** 蕨类植物; 叶绿体; IR 区; *rps12* 基因; 进化速率; GC 含量; 适应性进化

中图分类号: Q949.36

文献标识码: A

文章编号: 2095-0837(2020)01-0001-09

## Molecular evolution of chloroplast gene *rps12* in ferns

Ping Jing-Yao<sup>1</sup>, Zhu Ming<sup>1</sup>, Su Ying-Juan<sup>2,3</sup>, Wang Ting<sup>1,4\*</sup>

(1. College of Life Sciences, South China Agricultural University, Guangzhou 510642, China; 2. School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China; 3. Research Institute of Sun Yat-Sen University in Shenzhen, Shenzhen, Guangdong 518057, China; 4. Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China)

**Abstract:** Plant chloroplast genomes consist of an inverted repeat (IR) region and two single copy (SC) regions; however, the patterns of molecular evolution in the IR and SC regions differ. The *rps12* gene encodes the ribosomal small subunit S12 protein, which is composed of 5'-*rps12* (exon 1) and 3'-*rps12* (exon 2-3), with the 3'-*rps12* protein of different species located in different genomic regions. The *rps12* genes of 68 species of ferns and two species of lycophytes were studied, in the phylogenetic background, combined with the maximum likelihood method, the evolution rate and selection pressure of this gene were analyzed with HyPhy and PAML software. Results showed that exon 2-3 was located in the IR region, and its substitution rate was significantly reduced. The substitution rate of the coding sequence of *rps12* was also reduced, and the GC content in the third position of the codon of *rps12* was significantly increased. During the evolution of ferns, 3'-*rps12* tended to be located in the IR region to maintain a low substitution rate. Among the 123 amino acid sites encoded by *rps12*, four positive selection sites and 116 negative selection sites were detected. Results indicated that the gene translocated into the IR region showed decelerated substitution rates, and the strong negative selection pressure indicated that the RPS12 protein was highly conserved and

收稿日期: 2019-06-23, 退修日期: 2019-07-23。

基金项目: 国家自然科学基金项目 (31770587)。

This work was supported by a grant from the National Natural Science Foundation of China (31770587).

作者简介: 平晶耀 (1993-), 男, 博士研究生, 研究方向为植物系统发育与分子进化 (E-mail: pingjinyao@foxmail.com)。

\* 通讯作者 (Author for correspondence, E-mail: tingwang@scau.edu.cn)。

the function and structure of *rps12* were mostly stabilized.

**Key words:** Ferns; Chloroplast; Inverted repeat region; *rps12* gene; Evolutionary rate; GC content; Adaptive evolution

蕨类植物是维管植物的第二大类群, 约有 12 000 种。典型的叶绿体基因组一般为四区结构, 即由两个序列完全相同、方向相反的反向重复区 (inverted repeats, IRa 和 IRb)、大单拷贝区 (large single copy, LSC) 和小单拷贝区 (small single copy, SSC) 4 个部分组成<sup>[1]</sup>。*rps12* 基因编码叶绿体核糖体 30S 小亚基 S12 蛋白<sup>[2]</sup>。1987 年 Zaita 等<sup>[3]</sup> 鉴定出烟草 (*Nicotiana tabacum* L.) 的 *rps12* 基因及其转录过程中存在的反式剪接作用。该基因由 5'-*rps12* 和 3'-*rps12* 两部分组成, 5'-*rps12* 只含有外显子 1 (exon 1), 3'-*rps12* 含有外显子 2 (exon 2)、内含子 (intron) 和外显子 3 (exon 3), 5'-*rps12* 位于 LSC 区, 3'-*rps12* 位于 IR 区, 且两部分在基因组上相距甚远。研究发现, 在一些物种谱系中, 丢失了内含子或 3'-*rps12* 进入到 LSC 区。

IR 区与 SC 区的分子进化模式存在差异, 且 IR 区经常会发生扩张、收缩甚至完全丢失, 使一些基因序列进入或离开 IR 区。在一些被子植物中观察到 IR 区的同义替换率、非同义替换率比 SC 区低<sup>[4, 5]</sup>; 在苏铁 (*Cycas revoluta* Thunb.) 中发现 IR 区具有低的替换率<sup>[6]</sup>; 在一些丢失 IR 区的豆科植物中, 原位于 IR 区的基因显示出与位于 SC 区的基因具有相似的同义替换率<sup>[7]</sup>。位于 SC 区的基因, 其同义替换率是 IR 区基因的 3.7 倍, 基因从 SC 区转移到 IR 区或从 IR 区转移到 SC 区后, 都表现出与转移后所在区域基因一致的替换率<sup>[8]</sup>; 从 SC 区转移进入到 IR 区的基因序列, 其替换率降低, 密码子第 3 位的 GC 含量升高<sup>[9]</sup>。也有一些例外, 比如银杏 (*Ginkgo biloba* L.) 的 *ycf2* 基因从 IR 区转移进入 SC 区后, 其替换率并没有增加<sup>[10]</sup>。这些研究通过比较不同的基因或者不同物种中的相同基因探讨了 IR 区的性质, 但是不同基因的生物功能不同, 经受的选择压力也不同, 其替换率可能存在本质上的差别, 从而影响研究结果。*rps12* 基因提供了一个天然的条件, 可以通过比较该基因两部分的进化速率差异, 探究基因序列位于 IR 区后替换率的变化。

非同义替换率 (nonsynonymous substitution rate,  $dN$ ) 和同义替换率 (synonymous substitution rate,  $dS$ ) 的比值  $\omega$  ( $dN/dS$ ) 可在蛋白质水平上去度量选择压力。若  $dN = dS$  ( $\omega = 1$ ) 即为中性进化; 若  $dN < dS$  ( $\omega < 1$ ) 即受到负选择 (净化选择) 作用; 若  $dN > dS$  ( $\omega > 1$ ) 即受到正选择作用。已有报道对 *rbcL*<sup>[11]</sup>、*matK*<sup>[12]</sup> 及 *rps4*<sup>[13]</sup> 等叶绿体基因进行了适应性进化分析, 但是蕨类植物 *rps12* 基因是否发生适应性进化的相关研究则尚少。本研究以 68 种蕨类植物和 2 种石松类植物的 *rps12* 基因为对象, 在系统发育关系的背景下, 结合最大似然法, 通过分析 *rps12* 基因的进化速率, 探讨当基因序列位于反向重复区后, 其进化速率的变化, 并对 *rps12* 基因的适应性进化进行分析, 以期阐明 *rps12* 基因结构的改变和功能变异研究提供参考。

## 1 材料与方法

### 1.1 序列数据

选取的 70 种植物序列数据来自 NCBI 的 GenBank (<https://www.ncbi.nlm.nih.gov/>) 数据库 (表 1)。其中黑桫欏 (*Alsophila podophylla* Hook.) 的叶绿体全基因组序列已发表<sup>[14]</sup>。从 NCBI 中下载 70 种植物的叶绿体基因组全序列数据, 采用软件 Genious Prime 2019.1.3<sup>[15]</sup> 分别提取 *rbcL*、*matK* 以及 *rps12* 基因序列数据, 并用 MEGA 7.0<sup>[16]</sup> 软件中的 ClustalW 模块进行序列比对并进行手工校正。

### 1.2 分析方法

#### 1.2.1 系统发育树的构建

基于 *rbcL* 基因、*matK* 基因以及两者的串联数据集, 分别采用最大简约法、最大似然法和贝叶斯法构建系统发育树<sup>[17, 18]</sup>, 并结合 PPG (Pteridophyte phylogeny group) 已发表的蕨类系统发育关系<sup>[19]</sup>, 构建 70 种植物的系统树。采用 Modeltest 3.7 软件<sup>[20]</sup> 进行核苷酸替换模型的选择; MEGA 7.0 软件构建最大似然树; PAUP 4.0 软件<sup>[21]</sup> 构建最大简约树; MrBayes 3.2.6 软件<sup>[22]</sup> 构建贝叶斯树。

表 1 植物材料及其叶绿体基因组全序列的 GenBank 登录号  
Table 1 Plant materials and GenBank accession numbers of chloroplast genome sequences

No.	物种 Species	GenBank 登陆号 accession no.	No.	物种 Species	GenBank 登陆号 accession no.
1	<i>Athyrium anisopterum</i> Christ	NC_035738	35	<i>Drynaria roosii</i> (Kunze ex Mett.) Hovenkamp & S. Linds.	KY075853
2	<i>Athyrium opacum</i> (D. Don) Copel.	KY427335	36	<i>Hypodematium crenatum</i> (Forssk.) Kuhn	KY427351
3	<i>Anisocampium shearerii</i> (Baker) Ching	KY427330	37	<i>Pteridium aquilinum</i> (L.) Kuhn.	NC_014348
4	<i>Athyrium sinense</i> Rupr.	KY427333	38	<i>Adiantum capillus-veneris</i> (L.) Hook.	NC_004766
5	<i>Diplazium bellum</i> (C. B. Clarke) Bir	KY427343	39	<i>Cheilanthes lindheimeri</i> Hook.	NC_014592
6	<i>Diplazium dilatatum</i> Blume	KY427344	40	<i>Odontosoria chinensis</i> (L.) J. Sm.	MG913608
7	<i>Diplazium dushanense</i> (Ching ex W. M. Chu & Z. R. He) R. Wei & X.C. Zhang X.C.Zhang X.C.Zhang	KY427345	41	<i>Cibotium barometz</i> (L.) J. Sm.	MH105066
8	<i>Diplazium striatum</i> (L.) C. Presl	KY427346	42	<i>Alsophila podophylla</i> Hook.	MG262389
9	<i>Deparia lancea</i> (Thunb.) Fraser-Jenk.	KY427338	43	<i>Alsophila spinulosa</i> (Wall. ex Hook.) R. M. Tryon	NC_012818
10	<i>Deparia pycnosora</i> (H. Christ) M. Kato	KY427339	44	<i>Dicksonia squarrosa</i> (Forst.) Sw.	KJ569698
11	<i>Deparia viridifrons</i> (Makino) M. Kato	KY427340	45	<i>Azolla filiculoides</i> Lam.	MF177094
12	<i>Austroblechnum melanocaulon</i> (Brack.) Gasper & V. A. O. Dittrich	KY427334	46	<i>Azolla mexicana</i> C. Presl	MF177088
13	<i>Woodwardia unigemmata</i> (Makino) Nakai	NC_028543	47	<i>Azolla microphylla</i> Kaulf.	MF177089
14	<i>Matteuccia struthiopteris</i> (L.) Tod.	KY427353	48	<i>Azolla nilotica</i> Decne. ex Mett	MF177090
15	<i>Onoclea sensibilis</i> L.	KY427354	49	<i>Azolla rubra</i> R.Br.	MF177091
16	<i>Woodsia macrochaena</i> Mett. ex Kuhn	KY427358	50	<i>Salvinia cucullata</i> Roxb.	MF177095
17	<i>Woodsia polystichoides</i> D. C. Eaton	KY427359	51	<i>Marsilea crenata</i> L.	NC_022137
18	<i>Ampelopteris prolifera</i> (Retz.) Copel.	KY427329	52	<i>Schizaea elegans</i> (Vahl) Sw.	KX258660
19	<i>Christella appendiculata</i> (C. Presl) Holttum	KY427336	53	<i>Schizaea pectinata</i> (Vahl) Sw.	KX258661
20	<i>Stegnogramma sagittifolia</i> (Ching) L. J. He & X. C. Zhang	KY427357	54	<i>Actinostachys pennula</i> (Sw.) Hook.	KU764518
21	<i>Macrothelypteris torresiana</i> (Gaudich.) Ching	KY427352	55	<i>Lygodium japonicum</i> (Thunb.) Sw.	NC_022136
22	<i>Pseudophegopteris aurita</i> (Hook.) Ching	KY427355	56	<i>Diplopterygium glaucum</i> (Thunb. ex Houtt.) Nakai.	NC_024158
23	<i>Asplenium pekinense</i> Hance	KY427331	57	<i>Hymenophyllum holochilum</i> (Bosch) C. Chr.	MH265124
24	<i>Asplenium prolongatum</i> Hook.	KY427332	58	<i>Osmundastrum cinnamomeum</i> (L.) C. Presl	NC_024157
25	<i>Hymenasplenium unilaterale</i> (Lam.) Hayata	KY427350	59	<i>Angiopteris angustifolia</i> C. Presl	NC_026300
26	<i>Diplaziopsis cavaleriana</i> (Christ) C. Chr.	KY427341	60	<i>Angiopteris evecta</i> (G.Forst.) Hoffm.	NC_008829
27	<i>Diplaziopsis javanica</i> (Blume) C. Chr.	KY427342	61	<i>Botrychium ternatum</i> (Thunb.) Sw.	KM817789
28	<i>Homalosorus pycnocarpos</i> (Spreng.) Pic. Serm	KY427349	62	<i>Helminthostachys zeylanica</i> (L.) Hook.	KM817788
29	<i>Rhachidosorus consimilis</i> Ching	KY427356	63	<i>Mankyua chejuensis</i> B. Y. Sun, M. H. Kim & C. H. Kim	NC_017006
30	<i>Cystopteris chinensis</i> Ching	KY427337	64	<i>Ophioglossum californicum</i> Prantl	NC_020147
31	<i>Cyrtomium devexiscapulae</i> (Koidz.) Koidz. & Ching	NC_028542	65	<i>Psilotum nudum</i> (L.) P. Beauv.	NC_003386
32	<i>Cyrtomium falcatum</i> (L. f.) C. Presl	NC_028705	66	<i>Tmesipteris elongata</i> Danguy	KJ569699
33	<i>Dryopteris decipiens</i> (Hook.) O. Kuntze	KY427348	67	<i>Equisetum arvense</i> L.	NC_014699
34	<i>Lepisorus clathratus</i> (C. B. Clarke) Ching	NC_035739	68	<i>Equisetum hyemale</i> L.	NC_020146
			69	<i>Isoetes flaccida</i> A. Braun	GU191333
			70	<i>Huperzia lucidula</i> (Michx.) Trevis.	AY660566

### 1.2.2 进化速率分析

使用软件 HyPhy 2.2.4<sup>[23]</sup> 计算每个物种分支的转换率 (transition rate, *trst*)、颠换率 (transversionrate, *trsv*)、颠换率与转换率的比值 (*trsv/*

*trst*)、同义替换率 (*dS*)、非同义替换率 (*dN*)、非同义替换率与同义替换率的比值 ( $\omega$ )。调用约束模块进行假设约束。使用 SPSS 22 软件进行相关性检验, 包括威尔科克森符号秩检验 (Wilcoxon

test)、曼-惠特尼秩和检验(Mann-Whitney U test)以及  $t$  检验( $t$ -test)。

### 1.2.3 选择压力分析

使用 HyPhy 软件, 选用密码子模型, 调用类别处理器, 执行后验贝叶斯分析, 设定贝叶斯因子, 检测位点经受的选择压力。

### 1.2.4 氨基酸正选择位点的鉴定

使用软件 PAML 4.9 的 codeml 程序进行 *rps12* 基因的氨基酸正选择位点的鉴定分析<sup>[24]</sup>。包含 3 种模型: (1) 分支模型。它允许非同义替换和同义替换的比值  $\omega$  在不同分支上有变化, 其中单比率模型(one ratio, M0)假定所有进化支的  $\omega$  值都一致, 自由比率模型(Free ratio, F)假定各支的  $\omega$  值各不相同; 还有居于两者之间的二比率模型(two ratio, Model 2)假定前景支与背景支的  $\omega$  值不同。(2) 位点模型。该模型假定不同的位点有不同的  $\omega$  值, 而在系统树的不同分支上没有差异。主要包含 4 对嵌套模型: M1a(近中性)和 M2a(正选择)、M0 和 M3(离散)、M7( $\beta$ )和 M8( $\beta$ 和 $\omega$ )、M8 和 M8a( $\beta$ 和 $\omega = 1$ )。(3) 分支-位点模型<sup>[25]</sup>。允许  $\omega$  同时在分支和位点间存在差异, 可以检测指定分支上的正选择位点。该模型将系统发育分支分为前景支和背景支, 仅允许前景支中出现正选择位点的分支。模型 A 对 MA 和零假设(MA 中  $\omega_2$  值设置为 1)进行比较。

将推测的氨基酸序列提交到瑞士生物信息研究所(Swiss-Model, <https://www.swissmodel.expasy.org/>), 基于同源建模原理预测榲蕨 *Drynaria roosii* RPS12 蛋白的三维结构。

## 2 结果与分析

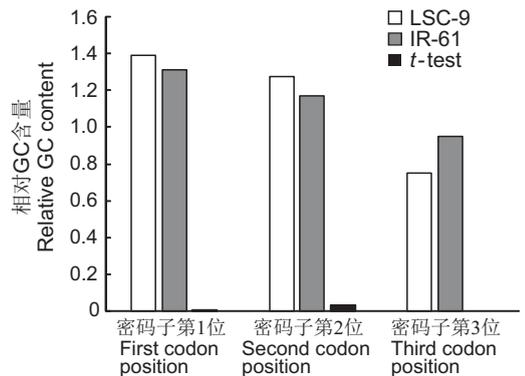
### 2.1 *rps12* 基因特征分析

研究结果显示, 70 种植物的 *rps12* 基因大小为 372~971 bp。编码序列(coding sequence, *rps12*-cds)的大小均为 372 bp, 其中外显子 1 为 114 bp、外显子 2 为 232 bp、外显子 3 为 26 bp。其中 56 种植物含有内含子, 大小为 447~599 bp。根据 *rps12* 在基因组上的定位, 将 70 种植物分成两类, 一类是外显子 2~3 位于 IR 区的 61 种植物(IR-61), 另一类是外显子 2~3 位于 LSC 区的 9 种

植物(LSC-9)。*rps12* 特征分析详见附件 1-1<sup>1)</sup>。

### 2.3 密码子 GC 含量

本研究对 *rps12* 密码子每一位上的 GC 含量(*rps12* 基因密码子每位上的 GC 含量与其对应物种的叶绿体基因组 GC 含量的比值)及相对值进行了计算, 结果见附件 1-2<sup>1)</sup>。进一步分析显示(图 1), LSC-9 密码子第 1 位和第 2 位的 GC 含量高于 IR-61( $P_1 = 0.003$ ,  $P_2 = 0.033$ ); IR-61 密码子第 3 位的 GC 含量显著高于 LSC-9( $P_3 = 0$ )。



IR-61: IR-61 的相对 GC 含量均值; LSC-9: LSC 的相对 GC 含量均值; t-test:  $t$  检验  $P$  值。  
IR-61: Relative GC content mean of IR-61; LSC-9: Relative GC content mean of LSC; t-test:  $P$  of  $t$ -test.

图 1 *rps12* 基因密码子每位的相对 GC 含量均值及  $t$  检验

Fig. 1 Relative GC content of each *rps12* gene codon and  $t$ -tests

### 2.4 系统发育树

对 70 种植物的系统发育树分析结果详见附件 1-3<sup>1)</sup>。研究发现, 70 种植物的外显子 1 位于 LSC, 蕨类植物的进化类群(水龙骨科 Polypodiales、桫欏目 Cyatheales、槐叶苹目 Salviniaceae、莎草蕨目 Schizaeales)与祖先类群石松类(Lycophytes)植物的外显子 2~3 都位于 IR 区; 在蕨类基部类群中, 外显子 2~3 多次进入和离开 IR 区(图 2)。

### 2.5 进化速率分析

#### 2.5.1 物种分支内外显子 1 与外显子 2~3 的参数均值比较

分析结果显示(图 3; a), IR-61 外显子 1 的参数均值是外显子 2~3 的 2~6 倍, 且存在差异( $\omega$

1) 如需查阅附件内容请登录《植物科学学报》网站(<http://www.plantscience.cn>)查看本期文章附件。

除外)。LSC-9 外显子 2~3 的 *trst* 和 *dS* 均值为外显子 1 的 2~3 倍, 也存在差异(图 3; b)。

### 2.5.2 物种分支间基因序列的参数均值比较

研究发现, *rps12*-cds(图 4; a) 中 LSC-9 的

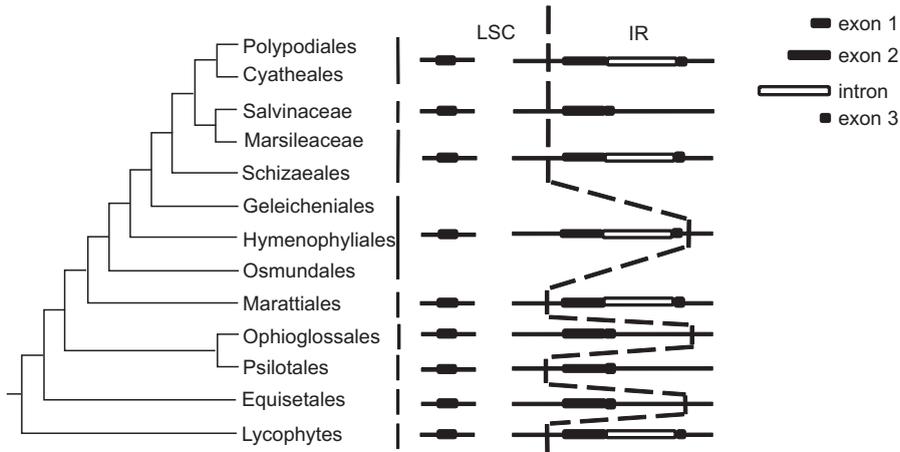
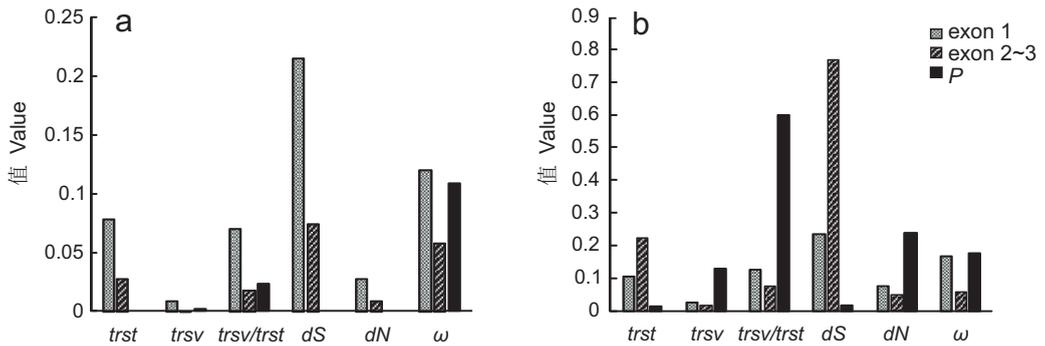


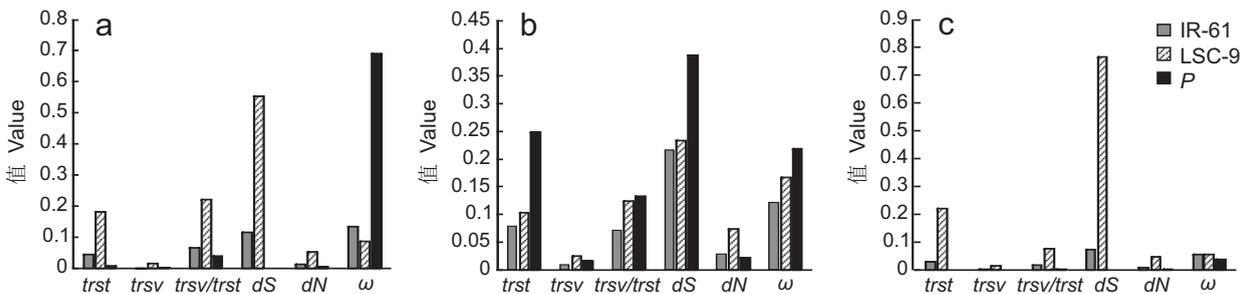
图 2 *rps12* 基因在 70 种植物叶绿体基因组中的定位  
Fig. 2 Localization of *rps12* genes in chloroplast genome of 70 plants



a: IR-61; b: LSC-9. exon 1: 外显子 1 的参数均值; exon 2~3: 外显子 2~3 的参数均值; P: 威尔科克森符号秩检验 P 值。*trst*: 转换率; *trsv*: 颠换率; *trsv/trst*: 颠换率/转换率; *dS*: 同义替换率; *dN*: 非同义替换率;  $\omega$ : 非同义替换率/同义替换率。下同。  
a: IR-61; b: LSC-9. exon 1: Parameter mean of exon 1; exon 2-3: Parameter mean of exon 2-3; P: P-value of Wilcoxon test. *trst*: Transition rate; *trsv*: Transversion rate; *trsv/trst*: Transversion rate/Transition rate; *dS*: Synonymous substitution rate; *dN*: Nonsynonymous substitution rate;  $\omega$ : *dN/dS*. Same below.

### 图 3 外显子 1 与外显子 2~3 的参数均值比较及 Wilcoxon 检验

Fig. 3 Comparison of mean values between exon 1 and exon 2-3 and Wilcoxon tests



a: *rps12*-cds; b: 外显子 1; c: 外显子 2~3. IR-61: 61 个物种的参数均值; LSC-9: 9 个物种的参数均值; P: 曼-惠特尼秩和检验 P 值。  
a: *rps12*-cds; b: exon 1; c: exon 2-3. IR-61: Parameter mean of 61 species; LSC-9: Parameter mean of nine species; P: P-value of Mann-Whitney U test.

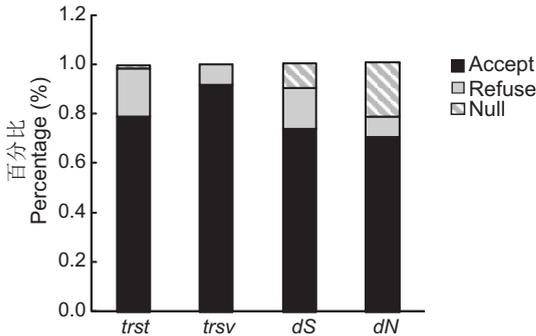
### 图 4 IR-61 与 LSC-9 间参数均值比较及 Mann-Whitney U 检验

Fig. 4 Comparison of mean values between IR-61 and LSC-9 and Mann-Whitney U tests

参数均值是 IR-61 的 3~5 倍, 且存在差异( $\omega$  除外)。外显子 1 中(图 4: b), LSC-9 的 *trsv* 和 *dN* 均值为 IR-61 的 2~3 倍, 且存在差异。外显子 2~3 中(图 4: c), LSC-9 的参数均值为 IR-61 的 4~10 倍, 也存在差异。

### 2.5.3 假设检验

IR-61 的 61 个物种分支中, 分别有 80%、90%、70%和 70%的物种分支接受外显子 1 的 *trst*、*trsv*、*dS*、*dN* 等于外显子 2~3 的假设(图 5)。



Accept:  $P > 0.05$  的物种分支占比; Refuse:  $P < 0.05$  的物种分支占比; Null: 不能进行有效约束假设的物种分支占比。*trst*: 转换率; *trsv*: 颠换率; *dS*: 同义替换率; *dN*: 非同义替换率。  
Accept: Percentage of species branches with  $P > 0.05$ ; Refuse: Percentage of species branches with  $P < 0.05$ ; Null: Percentage of species branches that cannot be effectively constrained by assumptions. *trst*: Transition rate; *trsv*: Transversion rate; *dS*: Synonymous substitution rate; *dN*: Nonsynonymous substitution rate.

图 5 假设检验

Fig. 5 Hypothesis testing

## 2.6 选择压力

在贝叶斯因子大于 1 的条件下, 本研究检测到 3 个正选择位点(17、74、108); 在贝叶斯因子大于 20 的条件下, 检测到 116 个负选择位点(1~16、18~24、26~59、61~73、75~80、82~97、99~107、109~123)。

## 2.7 正选择氨基酸位点的鉴定

### 2.7.1 分支模型

本研究中, 分支模型参数估计值和对数似然值详见附件 1-4<sup>2)</sup>; 似然比检验结果显示(表 2), 对于 *rps12-cds*: M0 与 F 检验结果不存在差异( $P = 0.918$ ); 对于 *rps12-cds*、外显子 1、外显子 2~3: M0 与 Model 2 的检验结果也不存在差异( $P_{rps12-cds} = 0.870$ ,  $P_{\text{外显子1}} = 0.963$ ,  $P_{\text{外显子2~3}} = 0.096$ )。

### 2.7.2 位点模型

位点模型下, 各个模型的参数估计值和对数似然值详见附件 1-4<sup>1)</sup>; 4 对嵌套模型的似然比检验结果见表 2。本研究以模型 M2a 的结果为准, 以后验概率大于 95% 为标准, 检测到一个可能的正选择位点 16G(后验概率为 98%)(附件 1-4<sup>2)</sup>)。

### 2.7.3 分支-位点模型

根据分支模型结果设置两种前景支, 其中, a: 将分支模型中 F 模型检测到的 11 个  $\omega > 1$  的

表 2 似然比值检验统计

Table 2 Likelihood ratio statistics

模型 Model	模型比较 Comparison of model	$2\Delta l$	自由度 <i>d. f.</i>	<i>P</i>
分支模型 Branch model				
<i>rps12</i> 编码序列 ( <i>rps12-cds</i> )	M0-Model 2	0.027	1	0.870
	M0-F	113.675	135	0.918
外显子 1 (exon 1)	M0-Model 2	0.002	1	0.963
外显子 2~3 (exon 2~3)	M0-Model 2	2.766	1	0.096
位点模型 Site model				
	M0-M3	324.748	4	0.000 **
	M1a-M2a	7.644	2	0.022 *
	M7-M8	17.439	2	0.000 **
	M8-M8a	5.351	1	0.021 *
分支-位点模型 Branch-site model				
a	型 A: 零假设—备选假设 a Model A: Null—alternative a	0.751	1	0.386
b	模型 A: 零假设—备选假设 b Model A: Null—alternative b	0.000	1	1.000

注:  $2\Delta l$ : 表示两个模型对数似然值的差值乘以 2。M0: 单比率; F: 自由比率; Model 2: 二比率; M1a: 近中性; M2a: 正选择; M3: 离散; M7:  $\beta$ ; M8:  $\beta$  和  $\omega$ ; M8a:  $\beta$  和  $\omega = 1$ 。\*:  $P < 0.05$ ; \*\*:  $P < 0.01$ 。

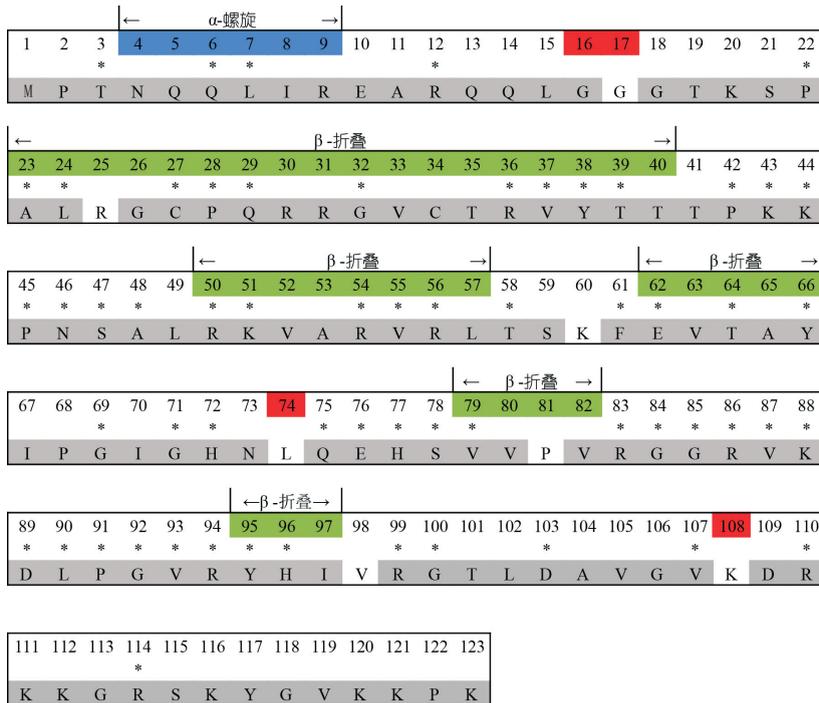
Notes:  $2\Delta l$ : Represents difference between logarithmic likelihood of two models multiplied by 2. M0: One ratio; F: Free ratio; Model 2: Two ratio; M1a: Neutral; M2a: Selection; M3: Discrete; M7:  $\beta$ ; M8:  $\beta$  &  $\omega$ ; M8a:  $\beta$  &  $\omega = 1$ . \*:  $P < 0.05$ ; \*\*:  $P < 0.01$ .

物种分支设为前景支；b: F 模型检测出的可能存在正选择的 11 个物种分支均属于薄囊蕨类，故将整个薄囊蕨类设为前景支。模型 A 的检测结果见附件 1-4，模型 A 与零假设模型检验结果不存在差异 ( $P_a = 0.386$ ,  $P_b = 1$ )，模型 A 的结果均不予接受(表 2)。

### 2.8 RPS12 蛋白质三维结构

将 70 种植物的 RPS12 蛋白序列与 *Escherichia coli* FVEC1412 (EFE98771)、地钱(*Marcha-*

*ntia polymorpha* L.) (AZU95199)、苏铁 (YP\_007474599)、烟草 (AMM05616)、*Pisum fulvum* Sibth. & Sm. (AUM82294) 等 75 条 RPS12 蛋白序列进行比对，以槲蕨为参考序列，分析结果显示(图 6)，完全一致的位点有 60 个，且这些一致位点都检测到受负选择作用；位于  $\alpha$ -螺旋区和  $\beta$ -折叠区的 42 个位点都检测到受负选择作用。槲蕨 RPS12 蛋白的三维结构预测结果表明(图 7)，位



数字 1~123 表示氨基酸位点编号；蓝色表示位于  $\alpha$ -螺旋区的位点；绿色表示位于  $\beta$ -折叠区的位点；红色表示 HyPhy 与 PAML 软件中检测到的经受正选择作用的位点；“\*”表示氨基酸序列比对中完全一致的位点；灰色区域表示 HyPhy 软件中检测到的经受负选择作用的位点。

Numbers 1–123 indicate amino acid site numbers; Blue indicates sites located in  $\alpha$ -helix region; Green indicates sites located in  $\beta$ -sheet region; Red indicates sites that have undergone positive selection detected by HyPhy and PAML; “\*” indicates a completely identical site in amino acid sequence alignment; Gray area represents sites subjected to negative selection detected by HyPhy.

图 6 蛋白序列比对(以槲蕨为参考序列)

Fig. 6 Protein sequence alignment (with *Drynaria roosii* as reference sequence)

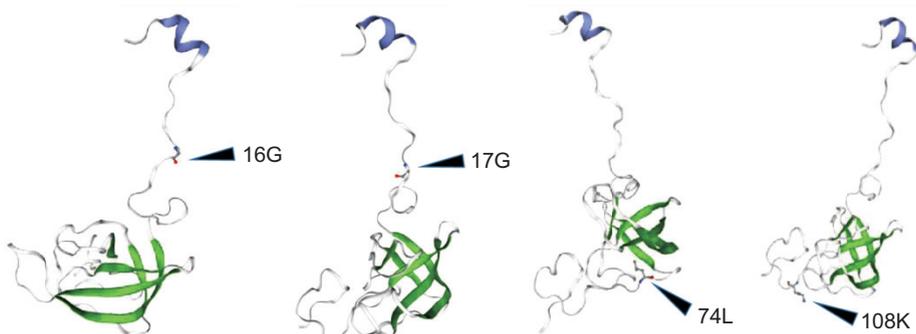


图 7 槲蕨 RPS12 蛋白正选择位点的空间位置

Fig. 7 Spatial locations of positively selected sites in RPS12 protein of *Drynaria roosii*

点 16G、74L 和 98V 位于蛋白质的环上或者不规则区；位点 108K 位于  $\pi$  螺旋区。

### 3 讨论

#### 3.1 反向重复区内基因序列替换率的降低

本研究结果显示，当外显子 2~3 位于 IR 区时，其替换率降低，*rps12* 基因编码序列的替换率和外显子 1 序列的颠换率和非同义替换率也随之降低。同时，*rps12* 基因密码子第 3 位的 GC 含量显著升高，这与前人<sup>[9]</sup> 研究结果一致，说明反向重复区内基因转换存在 GC 偏好。假设检验结果暗示替换率的降低可能与 IR 区有关；分支模型中 M0 与 Model 2 的检验结果显示替换率的降低不是由序列所经受的选择压力引起的，表明替换率的降低是 IR 区的一种特性。在蕨类的进化过程中，3'-*rps12* 多次进入和离开 IR 区，且 3'-*rps12* 更倾向定位于 IR 区，以保持较低的进化速率。在其他植物中也发现了 IR 区这种低的替换率模式<sup>[4-9]</sup>，这可能是质体基因组的一个标志特征。IR 区低的替换率可能是由于其双拷贝相关的修复机制<sup>[7]</sup>。比如基因转换偏向于一个新的突变<sup>[26]</sup>，有偏倚的基因转换可以在整个基因组中发生，IR 区由于其双拷贝的性质，发生基因转换的频率相对更高，同时 IR 区的两个拷贝间也可以进行基因转换<sup>[27]</sup>，最终导致被检测到的替换数目较少，从而表现出降低的替换率。

#### 3.2 适应性进化

本研究共检测到 4 个正选择位点(占 3.25%) 和 116 个负选择位点(占 94.31%)。正选择位点都位于该蛋白的一般结构域中，其中 98V、74L 为疏水性氨基酸残基，对于维持分子的热稳定性有重要作用；108K 为亲水性氨基酸残基，可能参与蛋白质在高盐浓度下的结构稳定；74L 参与氨酰-tRNA 的相互作用，这些位点具体的功能有待进一步验证。与 S8、S17、16S rRNA、23S rRNA、链霉素、氨酰-tRNA 等相互作用的绝大多数位点都检测到受强烈的负选择作用。氨基酸序列比对和蛋白质三维结构的预测结果显示了 RPS12 的高度保守性。某个基因如果受到正选择作用，表明这个群体需要产生新的基因功能去适应外界环境；如果其受到负选择作用则说明该基因保持原有的功能且趋于稳定。RPS12 的大多数主要功能位点受到强烈的负

选择作用，由此推断 *rps12* 基因进化保守，无论在低等的原核生物还是高等的陆生植物叶绿体中，它在核糖体小亚基形成过程中都扮演着重要的角色，且已经基本趋于稳定，强烈的负选择压力有助于维持 *rps12* 基因功能与结构的稳定。

本研究以蕨类植物叶绿体 *rps12* 基因为对象，利用其在基因组上的特殊定位，探讨了 IR 区替换率和 GC 含量的变化。研究结果为基因位于 IR 区后具有降低的替换率和升高的 GC 含量提供了有力的证据，同时，通过对 *rps12* 基因的进化分析，发现了该基因的高度保守特征，可为认识 *rps12* 基因结构的改变和功能变异提供参考。通过对 IR 区的研究，可以了解叶绿体基因组的基本特征和基因组内的替换率异质性提供参考。

#### 参考文献：

- [1] 王博. 真叶植物叶绿体基因的分子进化及卷柏属 *rbcS* 基因家族的进化研究[D]. 北京: 中国科学院大学, 2015: 1-67.
- [2] Harris EH, Boynton JE, Gillham NW. Chloroplast ribosomes and protein synthesis[J]. *Microbiol Rev*, 1994, 58(4): 700-754.
- [3] Zaita N, Torazawa K, Shinozaki K, Sugiura M. Trans splicing *in vivo*: joining of transcripts from the 'divided' gene for ribosomal protein S12 in the chloroplasts[J]. *Febs Lett*, 1987, 210(2): 153-156.
- [4] Maier RM, Neckermann K, Igloi GL, Kossel H. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing[J]. *J Mol Biol*, 1995, 251(5): 614-628.
- [5] Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs[J]. *Proc Natl Acad Sci USA*, 1987, 84(24): 9054-9058.
- [6] Wu CS, Chaw SM. Evolutionary stasis in *Cycad* plastomes and the first case of plastome GC-biased gene conversion[J]. *Genome Biol Evol*, 2015, 7(7): 2000-2009.
- [7] Perry AS, Wolfe KH. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat[J]. *J Mol Evol*, 2002, 55(5): 501-508.
- [8] Zhu A, Guo W, Gupta S, Fan WS, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates[J]. *New Phytol*, 2016, 209(4): 1747-1756.
- [9] Li FW, Kuo LY, Pryer KM, Rothfels CJ. Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content[J]. *Genome Biol*

- Evol*, 2016, 8(8): 2452–2458.
- [10] Lin CP, Wu CS, Huang YY, Chaw SM. The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction[J]. *Genome Biol Evol*, 2012, 4(3): 374–381.
- [11] 苏应娟, 王艇. 水龙骨科附生蕨类 Rubisco 大亚基的适应性进化: 正向选择位点的鉴定[J]. 中山大学学报(自然科学版), 2008, 47(5): 74–80.  
Su YJ, Wang T. Adaptive evolution of large subunits of Rubisco solanopteris of the polypodiaceae: identification of positive selection sites[J]. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 2008, 47(5): 74–80.
- [12] Hao DC, Chen SL, Xiao PG. Molecular evolution and positive Darwinian selection of the chloroplast maturase *matK*[J]. *J Plant Res*, 2010, 123(2): 241–247.
- [13] 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 LJ, Chen J, Wang T. Adaptive evolution in the chloroplast gene *rps4* in ferns[J]. *Bulletin of Botanical Research*, 2010, 30(1): 42–50.
- [14] Liu SS, Ping JY, Wang Z, Wang T, Su YJ. Complete chloroplast genome of the tree fern *Alsophila podophylla* (Cyatheaceae) [J]. *Mitochondrial DNA Part B*, 2018, 3(1): 48–49.
- [15] Kearse M, Moir R, Wilson A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. *Bioinformatics*, 2012, 28(12): 1647–1649.
- [16] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets [J]. *Mol Biol Evol*, 2016, 33(7): 1870–1874.
- [17] Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode [J]. *PLoS One*, 2011, 6(5): e19254.
- [18] Li FW, Kuo LY, Rothfels CJ, Ebihara A, Chiou WL, et al. *RbcL* and *matK* earn two thumbs up as the core DNA barcode for ferns [J]. *PLoS One*, 2011, 6(10): e26597.
- [19] PPG I. A community-derived classification for extant lycophytes and ferns [J]. *J Syst Evol*, 2016, 54(6): 563–603.
- [20] Posada D, Crandall KA. MODELTEST: Testing the model of DNA substitution [J]. *Bioinformatics*, 1998, 14(9): 817–818.
- [21] Swofford DL. PAUP\*: Phylogenetic Analysis Using Parsimony (and other methods): Version 4.0b10 [M]. Sunderland: Sinauer Associates, 2002.
- [22] Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees [J]. *Bioinformatics*, 2001, 17(8): 754–755.
- [23] Pond SL, Frost SD, Muse SV. HyPhy: Hypothesis testing using phylogenies [J]. *Bioinformatics*, 2005, 21(5): 676–679.
- [24] Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood [J]. *Mol Biol Evol*, 2007, 24(8): 1586–1591.
- [25] Zhang J, Nielsen R, Yang Z. 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.
- [26] Birky CJ, Walsh JB. Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes [J]. *Genetics*, 1992, 130(3): 677–683.
- [27] Khakhlova O, Bock R. Elimination of deleterious mutations in plastid genomes by gene conversion [J]. *Plant J*, 2006, 46(1): 85–94.

(责任编辑: 周媛)