



蕨类植物叶绿体基因 $ycf94$ 的分子进化研究

李子菲, 苏应娟, 王艇

引用本文:

李子菲, 苏应娟, 王艇. 蕨类植物叶绿体基因 $ycf94$ 的分子进化研究[J]. 热带亚热带植物学报, 2023, 31(4): 541–548.

在线阅读 View online: <https://doi.org/10.11926/jtsb.4780>

您可能感兴趣的其他文章

Articles you may be interested in

橄榄CaICE1基因的克隆和表达分析

Cloning and Expression Analysis of the CaICE1 Gene in *Canarium album*

热带亚热带植物学报. 2018, 26(6): 571–579 <https://doi.org/10.11926/jtsb.3877>

云南木兰科48种野生植物资源的遗传多样性研究

Studies on Genetic Diversity of 48 Wild Species of Magnoliaceae in Yunnan

热带亚热带植物学报. 2020, 28(3): 277–284 <https://doi.org/10.11926/jtsb.4161>

弯枝藻属rbcL基因的适应性进化分析

Adaptive Evolutionary Analysis of the rbc L Gene from *Compsopogon* (Rhodophyta)

热带亚热带植物学报. 2019, 27(1): 36–44 <https://doi.org/10.11926/jtsb.3909>

铁皮石斛DoSMT2基因的克隆与表达分析

Cloning and Expression Analysis of DoSMT2 Gene in *Dendrobium officinale*

热带亚热带植物学报. 2020, 28(6): 591–598 <https://doi.org/10.11926/jtsb.4234>

青花菜BoSCL3基因的克隆和渍水胁迫下的表达特征分析

Cloning BoSCL3 Gene from *Brassica oleracea* var. *italica* and Expression Analysis under Waterlogging Stress

热带亚热带植物学报. 2021, 29(2): 195–200 <https://doi.org/10.11926/jtsb.4263>

向下翻页, 浏览PDF全文

蕨类植物叶绿体基因 *ycf94* 的分子进化研究

李子菲¹, 苏应娟^{2,3*}, 王艇^{1*}

(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, 深圳 518057)

摘要: *ycf94* 基因是近年来在叶绿体基因组中新发现的一个基因, 在蕨类植物中表现高度保守。该研究共选取 94 种蕨类植物, 在系统发育背景下, 对 *ycf94* 基因的结构特征、密码子偏好性、进化速率和适应性进化进行分析。结果表明, *ycf94* 基因的密码子偏好性较弱, 偏好使用以 A/U 结尾的密码子, 且不同物种间的偏好性存在一定差异。密码子偏好性的形成主要受到突变压的影响, 同时也存在其他因素的作用; 基于凤尾蕨科和其他蕨类中 *ycf94* 基因的结构特征存在区别, 对两者的分子替换速率进行了比较, 表明颠换率、非同义替换率和 ω 值间存在显著差异; 仅检测出 1 个正选择位点 74A, 强烈的负选择作用表明 *ycf94* 基因的结构和功能基本趋于稳定。这为蕨类系统发育分析提供了新依据, 并提供了解析 *ycf94* 基因功能的线索。

关键词: 蕨类植物; *ycf94*; 密码子偏好性; 进化速率; 适应性进化

doi: 10.11926/jtsb.4780

Molecular Evolution of Chloroplast Gene *ycf94* in Ferns

LI Zifei¹, SU Yingjuan^{2,3*}, WANG Ting^{1*}

(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 518057, Guangdong, China)

Abstract: As a newly discovered chloroplast gene in recent years, *ycf94* gene, with length of about 200 bp, is conserved among ferns. However, the origin and function of *ycf94* require further investigation. The study was performed on the *ycf94* gene sequence of 94 species of ferns, in the phylogenetic background, analyzing codon usage bias, evolution rate and selection pressure. Results showed that the codon bias was weak in *ycf94* gene which the 3rd position of codon prefers A/U, variously in closely species. The codon bias was mainly produced by gene mutation. In addition, based on the differences in the structural characteristics of *ycf94* between Pteridaceae and other ferns, their evolution rate was compared, suggesting that there were significant difference in transversion rate, nonsynonymous substitution rate and omega. Besides, only one position selection site 74A was detected. The strong negative selection pressure indicated that the function and structure of *ycf94* were mostly stabilized. The results provide a new clue of phylogenetic analysis in ferns and functional studies of *ycf94* gene.

Key words: Fern; *ycf94* gene; Codon usage bias; Evolutionary rate; Adaptive evolution

叶绿体基因组具有长度短、拷贝数高、保守性强的特点, 长期被广泛用于植物系统发育和分子进化研究^[1]。叶绿体基因组作为双链环状闭合 DNA, 大小一般为 110~116 kb, 具有典型的四区结构, 包含了大单拷贝区(large single copy, LSC)、小单拷贝

区(small single copy, SSC)和 2 个反向重复区(inverted repeats, IR)。目前普遍认为陆地植物叶绿体基因组的基因组组成是大致相似的, 大约有 80 个蛋白编码基因(CDS, protein-coding gene)、30 个 tRNA 基因和 4 个 rRNA 基因^[2]。其中也包含一些保守的开放阅读

收稿日期: 2023-02-18 接受日期: 2023-03-13

基金项目: 国家自然科学基金项目(31770587)资助

This work was supported by the National Natural Science Foundation of China (Grant No. 31770587).

作者简介: 李子菲(1998 年生), 女, 硕士研究生, 研究方向为植物系统发育与分子进化。E-mail: fuyjun420@qq.com

* 通讯作者 Corresponding author. E-mail: suyj@mail.sysu.edu.cn; tingwang@scau.edu.cn

框(open reading frame, ORF), 其所编码的蛋白相关功能仍在摸索中, 通常称为假定叶绿体开放阅读框(hypothetical chloroplast open reading frame)基因(*ycf*)。

蕨类作为最早脱离水体适应陆生环境、进化出维管组织的植物类群, 在整个陆地植物的起源和演化历史研究中有着重要的地位^[3]。蕨类植物现存约有 12 000 种, 它在形态上具有丰富的多样性, 广泛分布在热带和亚热带地区。同时, 蕨类的质体基因组有独特的结构, 相比于与种子植物的质体基因组结构基本一致的蕨类祖先类型“莲座蕨型”, “铁线蕨型”分别在跨 IR/LSC 区和 LSC 区中发生了约 20 和 3 kb 的两次重叠倒位, 而“铁线蕨型”占据了蕨类 90% 的种类^[4]。2016 年发表的 PPG I (the pteridophyte phylogeny group I) 系统是一个基本建立在分子系统发育研究之上的现代蕨类植物分类系统^[5], 其中描述的分类较为全面且被接受。目前, 有关蕨类叶绿体基因组和分类系统等各类信息仍在不断完善中。

Song 等^[6]于 2018 年在薄囊蕨类中发现 1 个新的假定 ORF, 并将其命名为 *ycf94*。研究表明, *ycf94* 位于 *matK* 和 *rps16* 之间, 在薄囊蕨类中表现高度保守。同时在蕨类、石松类和苔藓类的质体中都发现其同源性, 而种子植物中不存在该序列。在陆地植物中质体的进化历史是动态的, 其基因组存在基因丢失的现象, 基因会从质体基因组转移到核基因组^[7]。*ycf94* 或许存在于陆生植物的祖先质体中, 随后在种子植物中发生基因丢失。

根据 *ycf94* 序列的保守程度, 以及其编码跨膜蛋白和具有多个 RNA 编辑位点的特征^[6], 可以表明 *ycf94* 是一个具有功能意义的蛋白编码基因^[8], 但其起源和功能目前尚不明确。本研究以 94 种蕨类植物的 *ycf94* 基因为研究对象, 通过分析 *ycf94* 的碱基组成和密码子偏好性, 并在系统发育背景下, 分析其进化速率和选择压力, 来探究 *ycf94* 的基因结构特征和在蕨类植物中的进化样式, 挖掘 *ycf94* 在叶绿体基因组系统发育关系分析中的可用性, 为蕨类植物分类提供新的依据, 为进一步研究 *ycf94* 的起源和功能等提供基础信息。

1 材料和方法

1.1 提取序列数据

根据 *Adiantum shastense* 的 *ycf94* 基因序列作为参考序列, 用 NCBI-BLAST 在线比对(E value<1e-6)

在 GenBank 数据库上选取了 94 种蕨类植物叶绿体全基因组, 涵盖了 3 目 19 科 60 属(附录 1)。采用 Genious prime 软件^[9]提取 *ycf94* 基因, 并使用 MEGA X 软件通过 MUSCLE (密码子)模块进行序列比对并对结果进行手工校正。

1.2 构建系统发育树

进行 94 种蕨类植物基于 *matK+rbcL* 串联数据集的多基因系统发育分析, 构建 3 种系统发育树。PhyloSuite v1.2.2 软件^[10]导入 94 条序列并提取 *matK* 和 *rbcL* 基因, 在 MATTF 插件上序列比对后进行基因串联。用 ModelFinder 插件来选择最佳的核苷酸替代模型 GTR+F+I+G4, 以 6 种合囊蕨目作为外类群, 利用 MrBayes 方法重建贝叶斯(bayesian inference, BI)树。并且用 MEGA 分别构建邻接(neighbor-joining, NJ)树和最大似然(maximum likelihood, ML)树。综合分析 3 种树, 结合 PPGI 系统发育关系进行手工调整, 得出系统发育树。

1.3 密码子偏好性分析

以 94 种蕨类植物的 *ycf94* 基因为对象, 使用在线软件 CUSP 计算密码子第一、第二、第三位的 GC 含量(分别用 GC1、GC2、GC3 表示)和总 GC 含量; CodonW v1.4.2 软件统计同义密码子相对使用度(RSCU, relative synonymous codon usage)和有效密码子(ENC, effective number of codon)。使用软件 SPSS 24 对数据进行 Pearson 相关性分析。

GC3-GC12 分析 密码子的使用受到碱基的突变压和自然选择的双重作用^[11], 通过分析 GC12 和 GC3 的相关性, 研究密码子 3 个位置上的碱基组成是否相似和影响密码子偏好性的主要因素。当 GC12 和 GC3 之间显著相关, 说明密码子 3 个位置上的碱基组成无差异, 密码子的使用主要受突变的影响; 当 GC12 和 GC3 之间相关性不显著, 说明密码子第一、第二和第三位碱基组成不相似, 密码子的使用主要受自然选择的影响。

ENC-plot 分析 分析 GC3 和 ENC 的相关性, 研究碱基组成对密码子偏好性的影响。以 GC3 为横坐标、ENC 值为纵坐标绘制散点图, 并绘制 ENC 期望值标准曲线($ENC=2+GC3+29/[GC3^2+(1-GC3)^2]$)进行比较。当基因散点分布在标准曲线附近, 说明密码子偏好性受突变影响较大; 而散点位于曲线较远下方位置时, 密码子偏好性受自然选择影响较大。

最优密码子 ENC 值反映了同义密码子非均衡使用的偏好程度, 取值范围为 20~61。通常基

因表达量越高, 密码子偏好性越强, ENC 值越小^[12]。分析 94 种蕨类植物 *ycf94* 基因主要偏好密码子, 构建以 ENC 值为标准的高低表达库(选择 ENC 值最低和最高的 10 个基因), 计算 Δ RSCU 值(Δ RSCU=RSCU_{High}-RSCU_{Low}), 选择 Δ RSCU>0.08 且 RSCU>1 的作为最优密码子。

1.4 分子进化速率分析

采用 HyPhy v2.2.4 软件^[13]基于核苷酸模型 HKY85 和密码子模型 MG94×HKY85_3×4 计算各物种分支 *ycf94* 基因的转换率(transition rate, trst)、颠换率(transversion rate, trsv)、转换率和颠换率的比值(trst/trsv)、非同义替换率(nonsynonymous substitution rate, dN)、同义替换率(synonymous substitution rate, dS), 以及非同义替换速率和同义替换速率的比值(ω =dN/dS)。通过软件 SPSS 24 使用 Mann-Whitney 秩和检验对数据进行相关性分析。

1.5 适应性进化分析

使用 PAML v4.9 软件^[14]中的 codeml 程序检测 *ycf94* 基因中是否存在氨基酸正选择位点。采用检测谱系之间选择压力的 3 种模型: 分支模型(Branch model)、位点模型(Site model)和分支-位点模型(Branch-site model)。

分支模型中, 单比率模型(one-ratio, M0)假定所有分支的 ω 值相同; 自由比率模型(free-ratio, F)假定所有分支的 ω 值均不相同; 二比率模型(two-ratio, M2)假定前景支与背景支的 ω 值不相同。通过似然比检验进行模型比较, M0 和 F 检验不同分支间 ω 值的差异性, M0 和 M2 检验前景支和背景支间 ω 值的差异性。

位点模型假定不同位点的 ω 值不同, 其中存在 4 对比较模型: M0 (one-ratio)和 M3 (discrete)检测位点间 ω 值是否一致; M1a (neutral)和 M2a (selection), M7 (beta)和 M8 (beta& ω), M8 和 M8a (beta& ω =1) 用于检测是否存在正选择位点。

分支-位点模型假定了分支间和位点间 ω 值都存在差异。对模型 Model A 和零假设进行比较, 检测前景支中的部分位点是否受到正选择作用。

2 结果和分析

2.1 *ycf94* 的基因特征分析

取样的 94 种蕨类植物 *ycf94* 基因长度为 222~156 bp。其中可以大致分为 2 种类型: 以起始密码子为 T(ACG)开头发生 C to U RNA 编辑的序列共 28 条,

其中包括全部的凤尾蕨科 27 条, *ycf94* 大多为 219 bp (共 20 条); 起始密码子为 M(AUG)的序列共 66 条, 包含除凤尾蕨科外的 18 科, *ycf94* 大多为 216 bp (共 50 条), 其中 6 条合囊蕨目的大小均为 177 bp。

2.2 94 种蕨类植物的系统发育树

利用 *matK+rbcL* 串联基因构建 NJ、ML 和 BI3 种系统发育树并对其进行比较, 用蕨类分类系统 PPGI 进行修正。

在 ML 和 BI 中水龙骨亚目和铁角蕨亚目为姐妹群且都是单系, 与 PPGI 系统一致; 而 NJ 中水龙骨亚目属于多系。BI 支持铁角蕨科和肠蕨科为姐妹群, ML 支持铁角蕨科位于铁角蕨亚目的基部位置。BI 与 PPGI 系统最为接近, 所以最后以 BI 为标准。构建出的系统发育树(附录 2)主要为后续研究分析提供系统发育背景。

2.3 密码子偏好性分析

2.3.1 密码子碱基组成分析

附录 3 为 94 种蕨类 *ycf94* 基因的密码子碱基组成, 其中 GC 含量为 23.73%~47.75%, GC1、GC2 和 GC3 分别为 19.72%~47.22%、23.94%~47.22%和 20.34%~56.94%, 说明密码子总体的碱基组成偏向 A/U。ENC 值为 61~34.89, 在近缘物种中 *ycf94* 基因的密码子偏好性存在差异。其中有 19 条序列的 ENC=61 即每个密码子都被均衡使用。ENC 值以 35 作为判断密码子偏好性强弱的标准^[15], 大部分的 *ycf94* 密码子偏好性较弱。

对 *ycf94* 基因的密码子相关参数进行热图统计, 大多数的 GC3 要高于 GC1、GC2 和 GC 值; 合囊蕨目的 GC 各项数值要低于其他蕨类, 凤尾蕨科的 GC1 和 GC2 总体上也低于其他蕨类(图 1)。

由表 1 可见, GC1、GC2、GC3 和 GC 相互间均为极显著相关, 表明 *ycf94* 基因密码子三位的碱基组成相似; 而 ENC 值与各项 GC 值相关性不显著, 碱基组成对 *ycf94* 密码子偏好性的影响较小。

表 1 *ycf94* 密码子各参数的相关性分析

Table 1 Correlation analysis between the parameters of *ycf94* gene

	GC1	GC2	GC3	GC	ENC
GC1					
GC2	0.589**				
GC3	0.585**	0.720**			
GC	0.796**	0.889**	0.913**		
ENC	0.047	-0.041	0.062	0.089	

*: $P<0.05$; **: $P<0.01$

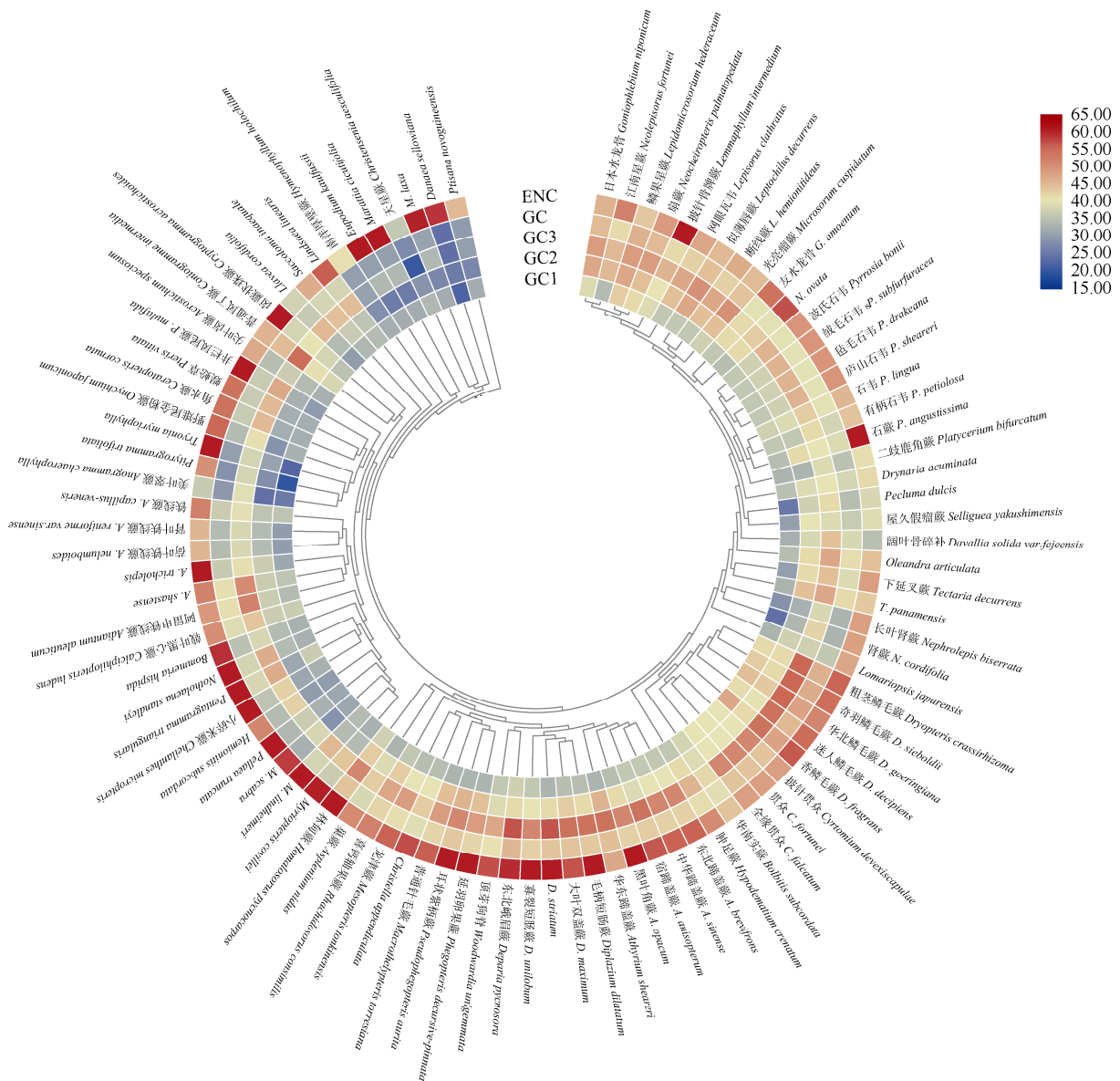


图 1 *ycf94* 基因的密码子相关参数热图统计

Fig. 1 Heat map of correlation parameter of *ycf94* codon

2.3.2 GC3-GC12 分析

由图 2 可见，大部分散点位于对角线之下，除天星蕨(*Christensenia aesculifolia*)外，大部分物种的 *ycf94* 基因的 GC3 值大于 GC12。相关性分析结果表明，双尾检验 $P=0.00$ (<0.01)，相关系数为 0.739，GC3 和 GC12 相关性极显著，说明密码子偏好性主要受到突变的影响。

2.3.3 ENC-plot 分析

由图 3 可见，散点的分布较为分散，且一部分落在距离曲线较远的位置，则该部分基因密码子偏好性受选择影响较大。有 40 条序列的 ENC 比值为 $-0.1\sim 0.1$ ，表明 ENC 值与期望值相差较小，该部分

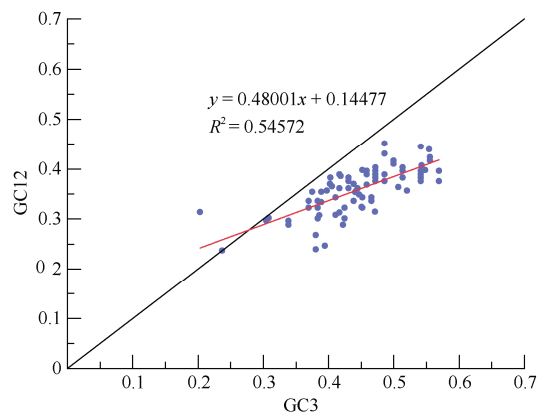


图 2 中性绘图

Fig. 2 Neutrality plot analysis

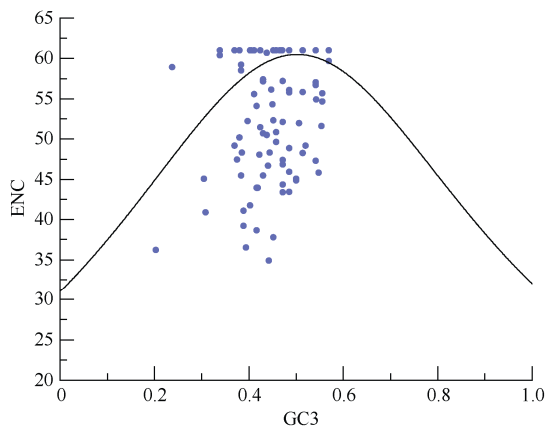


图 3 ENC-plot 绘图
Fig. 3 ENC-plot analysis

基因的密码子偏好性受突变影响较大。

2.3.4 *ycf94* 最优密码子

从表 2 可见,共筛选出高频密码子(RSCU>1) 26 个,高表达密码子(Δ RSCU>0.08) 24 个,而同时满足这两个条件的最优密码子有 14 个: UUU、GUU、GUA、GUG、UCU、CCC、ACC、GCC、UAU、AAA、GAU、AGG、GGC、GGA。其中 8 个以 A/U 结尾,6 个以 G/C 结尾。

2.4 进化速率分析

对凤尾蕨科和其余蕨类的替换速率进行差异显著性分析,数据表明(附录 4),凤尾蕨科除了同义替换率外的进化速率均值都大于其他蕨类(图 4),秩

表 2 *ycf94* 基因的密码子 RSCU 统计以及最优密码子

Table 2 Relative synonymous codon usage of genes and the optimal codons in *ycf94* gene

氨基酸 Amino acid	密码子 Codon	RSCU _{Low}	RSCU _{High}	Δ RSCU	RSCU	氨基酸 Amino acid	密码子 Codon	RSCU _{Low}	RSCU _{High}	Δ RSCU	RSCU	
苯丙氨酸 Phe	UUU*	0.99	1.52	0.53**	1.37***	酪氨酸 Tyr	UAU*	1.18	1.42	0.24**	1.10***	
	UUC	1.01	0.48	-0.53	0.63		UAC	0.82	0.58	-0.24	0.90	
亮氨酸 Leu	UUA	1.28	1.20	-0.08	1.36***	组氨酸 His	CAU	2.00	0.67	-1.33	1.54***	
	UUG	0.88	1.50	0.62**	0.83		CAC	0.00	1.33	1.33**	0.46	
	CUU	1.76	1.80	0.04	1.78***	谷氨酰胺 Gln	CAA	0.40	0.67	0.27**	0.39	
	CUC	1.12	0.90	-0.22	1.02***		CAG	1.60	1.33	-0.27	1.61***	
	异亮氨酸 Ile	CUA	0.88	0.60	-0.28	0.75	天冬酰胺 Asn	AAU	1.67	1.67	0.00	1.74***
CUG		0.08	0.00	-0.08	0.25	AAC		0.33	0.33	0.00	0.26	
AUU		1.17	1.00	-0.17	0.95	赖氨酸 Lys	AAA*	1.37	1.58	0.21**	1.46***	
AUC		1.11	0.95	-0.16	1.10***		AAG	0.63	0.42	-0.21	0.54	
缬氨酸 Val	AUA	0.72	1.05	0.33**	0.96	天冬氨酸 Asp	GAU*	1.18	1.67	0.49**	1.37***	
	GUU*	1.17	1.33	0.16**	1.16***		GAC	0.82	0.33	-0.49	0.63	
	GUC	1.08	0.57	-0.51	0.74	谷氨酸 Glu	GAA	1.47	1.43	-0.04	1.30***	
	GUA*	1.17	1.33	0.16**	1.03***		GAG	0.53	0.57	0.04	0.70	
丝氨酸 Ser	GUG*	0.58	0.76	0.18**	1.08***	半胱氨酸 Cys	UGU	2.00	2.00	0.00	1.00	
	UCU*	1.52	2.62	1.10**	1.72***		UGC	0.00	0.00	0.00	1.00	
	UCC	1.62	0.76	-0.86	1.27***	色氨酸 Trp	UGG	1.00	1.00	0.00	1.00	
	UCA	0.19	0.34	0.15**	0.14		精氨酸 Arg	CGU	0.00	0.10	0.10**	0.03
	UCG	0.00	0.08	0.08	0.09			CGC	0.45	0.10	-0.35	0.26
脯氨酸 Pro	CCU	1.33	0.91	-0.42	0.87	CGA	0.09	0.58	0.49**	0.34		
	CCC*	1.00	1.64	0.64**	1.69***	CGG	0.36	0.48	0.12**	0.58		
	CCA	0.67	0.73	0.06	0.92	丝氨酸 Ser	AGU	1.52	1.52	0.00	1.69***	
	CCG	1.00	0.73	-0.27	0.52		AGC	1.14	0.68	-0.46	1.08***	
苏氨酸 Thr	ACU	0.67	0.31	-0.36	0.52	精氨酸 Arg	AGA	3.85	3.39	-0.46	3.28	
	ACC*	2.33	2.46	0.13**	2.28***		AGG*	1.25	1.35	0.10**	1.51***	
	ACA	0.33	0.31	-0.02	0.39	甘氨酸 Gly	GGU	0.70	0.21	-0.49	0.36	
	ACG	0.67	0.92	0.25**	0.81		GGC*	0.96	1.74	0.78**	1.35***	
丙氨酸 Ala	GCU	1.33	0.71	-0.62	0.92		GGA*	1.22	1.44	0.22**	1.19***	
	GCC*	1.07	2.55	1.48**	1.89***	GGG	1.13	0.62	-0.51	1.10***		
	GCA	1.33	0.36	-0.97	0.84							
	GCG	0.27	0.36	0.09**	0.34							

*: 最优密码子; **: 高表达密码子; ***: 高频密码子。
*: Optimal codon; **: High expression codon; ***: High frequency codon.

和检验中颠换率,非同义替换率和 ω 的 P 值分别是 0.011、0.017 和 0.036 ($P < 0.05$), 说明凤尾蕨科和其他蕨类 *ycf94* 基因的 *trsv*、*dN* 和 ω 存在显著差异, 而 *trst*、*trsv/trst* 和 *dS* 无显著差异。

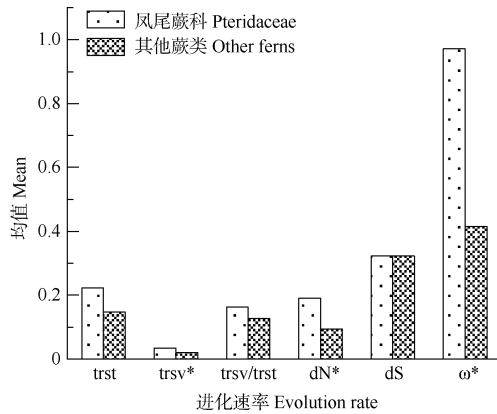


图 4 凤尾蕨科与其他蕨类 *ycf94* 进化速率均值比较及 Mann-Whitney 检验
Fig. 4 Comparison of mean values of *ycf94* evolution rate between Pteridaceae and other ferns and Mann-Whitney U tests

表 3 不同模型下的参数估计值和对数似然值

Table 3 Parameter estimates and log-likelihood values under different models

模型 Model	对数似然值 Likelihood-value	参数 Parameter	正选择位点 Positive selected site
分支模型 Branch model	M0	$\omega=0.416\ 06$	
	F		
	M2	$\omega_1=0.329\ 94, \omega_2=0.578\ 80$	
位点模型 Site model	M1a	$P_0=0.689\ 72, \omega_0=0.261\ 43$ $P_1=0.310\ 28, \omega_1=1.000\ 00$	不允许
	M2a	$P_0=0.675\ 37, \omega_0=0.266\ 54$ $P_1=0.302\ 59, \omega_1=1.000\ 00$ $P_2=0.022\ 04, \omega_2=2.193\ 99$	74 A
	M3	$P_0=0.175\ 58, \omega_0=0.062\ 16$ $P_1=0.573\ 65, \omega_1=0.350\ 45$ $P_2=0.250\ 77, \omega_2=1.015\ 19$	无
	M7	$P=0.960\ 67, q=1.168\ 03$	不允许
	M8	$P_0=0.918\ 31, P=1.204\ 73$ $q=1.879\ 88, P_1=0.081\ 69$ $\omega=1.472\ 03$	74 A**
	M8a	$P_0=0.816\ 16, P=1.219\ 64$ $q=2.737\ 83, P_1=0.183\ 84$ $\omega=1.000\ 00$	无
	分支-位点模型 Branch-site model	Null model	$P_0=0.447\ 51, P_1=0.144\ 96$ $P_{2a}+P_{2b}=0.407\ 53, \omega=1.000\ 00$
Model A		$P_0=0.447\ 51, P_1=0.144\ 96$ $P_{2a}+P_{2b}=0.407\ 53, \omega=1.000\ 00$	

**：后验概率 < 99%

**：Posterior probability < 99%

2.5 适应性进化分析

表 3 和 4 是 PAML 得出的 *ycf94* 基因适应性进化分析的结果。对各模型下计算出的对数似然值进行模型间的似然比检验, 分析显著性选择更为适合的模型。结果表明, 分支模型下 M0 和 F 检验结果显著 ($P=0.024$), 表明不同分支间 ω 值存在显著差异。二比率模型中设定了以凤尾蕨科为前景支, 其他蕨类为背景支, M0 和 M2 检验结果极显著 ($P=6.188 \times 10^{-6}$), 表明凤尾蕨科和其他蕨类的 ω 值存在极显著差异。

位点模型下, 结果显示位点间 ω 值存在显著差异 ($P=0.000$), 且 3 对模型比较都显示 *ycf94* 基因中存在正选择位点。以 BEB 后验概率大于 95% 为标准, 模型 M2a 和 M8 均鉴定出 1 个正选择位点 74A。

在分支-位点模型中, 同样以凤尾蕨科为前景支, 似然比检测结果显示零假设模型和 Model A 间不存在差异 ($P=1.000$), 故不接受模型 A 的结果。

表 4 似然比检验

Table 4 Likelihood ratio tests

模型 Model	模型比较 Comparison of model	$2\Delta l$	自由度 d.f.	<i>P</i>
分支模型 Branch model	M0-F	224.855 78	185	0.024*
	M0-M2	20.429 00	1	6.188×10^{-6} **
位点模型 Site model	M0-M3	258.779 00	4	0.000**
	M1a-M2a	8.609 06	2	0.014*
	M7-M8	15.463 73	2	4.386×10^{-4} **
	M8-M8a	6.835 55	1	0.009**
分支-位点模型 Branch-site model	Null model-Model A	0	1	1.000

*: $P < 0.05$; **: $P < 0.01$

3 结论和讨论

ycf94 基因在蕨类中表现出高度保守性,但在种子植物中没有发现同源序列这个重要特征,可能成为系统发育分析中的一个新的分子标记。同时,研究发现 *ycf94* 在蕨类中存在 2 种结构样式:起始密码子 ACG 发生 RNA 编辑的、基因长度多为 219 bp 的序列和以起始密码子为 AUG、基因长度多为 216 bp 的序列。其中全部的风尾蕨科 *ycf94* 基因序列都表现为前一种模式,以及三叉蕨科的 *Tectaria panamensis* 也属于同一结构,其余的序列均为后者。因此, *ycf94* 的起始密码子发生 C to U RNA 编辑虽然不属于风尾蕨科的独有结构,但 RNA 编辑作为表观遗传现象之一拥有可遗传性,推测风尾蕨科的共同祖先中 *ycf94* 基因起始密码子存在 RNA 编辑,该基因结构是风尾蕨科的共衍征。

编译氨基酸过程中同义密码子的使用是存在偏好性的,而物种在进化中会形成特有的密码子使用模式,不同物种间的密码子偏好性不同,而近缘物种间的密码子使用模式有可能相似^[6]。这种偏好性的差异对揭示物种间的进化关系,探究基因特征和蛋白质功能有重要意义。本研究中, *ycf94* 基因的密码子偏好性较弱,同义密码子的使用倾向平均,其中有 19 条序列密码子没有使用偏好,说明 *ycf94* 的表达水平可能相对较低,非高表达基因。统计表明碱基组成整体偏向 A/T,密码子第三位也以 A/U 为主,这与叶绿体基因组中密码子大多偏好 A/U 结尾这一特征相符合^[7],且 GC3 大多高于 GC1 和 GC2。研究结果中 94 种蕨类的 *ycf94* 基因 ENC 值从 34 到 61 不等,其密码子偏好性存在一定差异。通常,在没有其他压力作用下,同义密码子的使用概率可能因为基因长度的限制而造成统计上的误差,这种物种间的差异性可能与之相关。同时,GC3 与 GC12 表

现出显著的相关性, *ycf94* 密码子不同位置下碱基组成的差异较小。而 ENC 与 GC 各项数值都不相关,说明碱基组成对 *ycf94* 密码子偏好性的影响较小。通过 GC3-GC12 和 ENC-plot 分析得出结论, *ycf94* 基因的密码子偏好性主要受突变影响,同时也受到选择压力和基因长度等其他因素的作用。

基于凤尾蕨科和其他蕨类中 *ycf94* 基因的结构特征存在区别,对两者的分子替换速率进行比较,研究表明,凤尾蕨科除了同义替换率外的进化速率均值都大于其他蕨类,且颠换率、非同义替换率和 ω 值间两者存在显著差异。凤尾蕨科具有陆生、水生、旱生和附生等多个生态位^[8],有可能是生境的差异使 *ycf94* 在某一生态位上受到较高的选择压力,导致凤尾蕨科的进化速率升高。

在 *ycf94* 基因的适应性进化分析中,分支模型下似然比检验支持的自由比率模型中检测出 15 条 $\omega > 1$ 的分支,说明该部分分支可能有正选择存在,其中有 7 条分支属于凤尾蕨科,其余分支包括了蹄盖蕨科、鳞毛蕨科、水龙骨科和合囊蕨科。同时似然比检验表示二比率模型优于单比率模型,凤尾蕨科和其他蕨类的 ω 值存在极显著差异,这与 HyPhy 的分子进化速率计算结果一致。但是前景支 ω 值 ($\omega = 0.578$) 并不高于 1,尚不能认为它是受正选择作用的,选择压力约束放松可能是较为合理的解释。*ycf94* 基因共编码 80 个氨基酸位点,位点模型 M2a 和 M8 均检测出一个正选择位点 74A,占总序列 1.25%;基于 M8 模型检测负选择位点 63 个 ($P < 0.05$),占总序列 78.75%。当一个基因受到正选择压力时,蛋白功能可能发生改变来适应外界环境变化;而当一个基因受到负选择压力,则说明该基因倾向保持原有的功能^[9]。总的来说 *ycf94* 基因大部分位点受到负选择作用,推断其进化高度保守,在蕨类中 *ycf94* 基因的结构和功能基本趋于稳定。下一步分析氨基酸位

点在蛋白质结构域中的分布,对正选择位点的定位可以探究 *ycf94* 基因功能,进一步了解适应性进化发生的原因和作用。

本研究以 94 种蕨类植物的 *ycf94* 基因为对象,揭示了 *ycf94* 基因的基本结构特征,为叶绿体基因组的基因组成补充信息。在探讨 *ycf94* 的碱基组成与密码子偏好性的研究中, *ycf94* 基因的密码子偏好较弱,偏好使用 A 和 U 结尾的密码子,其密码子偏好性主要受到突变的影响,并且不同物种之间密码子偏好性具有一定差异。同时,通过对 *ycf94* 基因的进化分析,发现该基因表现高度保守,为 *ycf94* 基因的功能研究提供参考。

参考文献

- [1] RAUBESON L A, JANSEN R K. Chloroplast genomes of plants [M]// HENRY R. Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants. Cambridge: CABI, 2005: 45–68. doi: 10.079/9780851999043.0045.
- [2] JANSEN R K, RUHLMAN T A. Plastid genomes of seed plants [M]// Genomics of Chloroplasts and Mitochondria: Advances in Photosynthesis and Respiration. New York: Springer, 2012: 103–126. doi: 10.1007/978-94-007-2920-9_5.
- [3] ZHANG X C, WEI R, LIU H M, et al. Phylogeny and classification of the extant lycophytes and ferns from China [J]. Chin Bull Bot, 2013, 48(2): 119–137. [张宪春, 卫然, 刘红梅, 等. 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. doi: 10.3724/SP.J.1259.2013.00119.]
- [4] PPG I. A community-derived classification for extant lycophytes and ferns [J]. J Syst Evol, 2016, 54(6): 563–603. doi: 10.1111/jse.12229.
- [5] WOLF P G, ROPER J M, DUFFY A M. The evolution of chloroplast genome structure in ferns [J]. Genome, 2010, 53(9): 731–738. doi: 10.1139/G10-061.
- [6] SONG M, KUO L Y, HUIET L, et al. A novel chloroplast gene reported for flagellate plants [J]. Am J Bot, 2018, 105(1): 117–121. doi: 10.1002/ajb2.1010.
- [7] CULLIS C A, VORSTER B J, VAN DER VYVER C, et al. Transfer of genetic material between the chloroplast and nucleus: How is it related to stress in plants? [J]. Ann Bot, 2009, 103(4): 625–633. doi: 10.1093/aob/mcn173.
- [8] HSU P Y, BENFEY P N. Small but mighty: Functional peptides encoded by small ORFs in plants [J]. Proteomics, 2018, 18(10): 1700038. doi: 10.1002/pmic.201700038.
- [9] KEARSE M, MOIR R, WILSON A, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. Bioinformatics, 2012, 28(12): 1647–1649. doi: 10.1093/bioinformatics/bts199.
- [10] ZHANG D, GAO F L, JAKOVLIĆ I, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies [J]. Mol Ecol Resour, 2020, 20(1): 348–355. doi: 10.1111/1755-0998.13096.
- [11] SUEOKA O. Total cross section measurements for positron and electron scattering on benzene molecules [J]. J Phys B: At Mol Opt Phys, 1988, 21(20): L631. doi: 10.1088/0953-4075/21/20/003.
- [12] WRIGHT F. The ‘effective number of codons’ used in a gene [J]. Gene, 1990, 87(1): 23–29. doi: 10.1016/0378-1119(90)90491-9.
- [13] POND S L K, FROST S D W, MUSE S V. HyPhy: Hypothesis testing using phylogenies [J]. Bioinformatics, 2005, 21(5): 676–679. doi: 10.1093/bioinformatics/bti079.
- [14] YANG Z H. PAML 4: Phylogenetic analysis by maximum likelihood [J]. Mol Biol Evol, 2007, 24(8): 1586–1591. doi: 10.1093/molbev/msm088.
- [15] JIANG Y, DENG F, WANG H L, et al. An extensive analysis on the global codon usage pattern of baculoviruses [J]. Arch Virol, 2008, 153(12): 2273–2282. doi: 10.1007/s00705-008-0260-1.
- [16] ROMERO H, ZAVALA A, MUSTO H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces [J]. Nucl Acids Res, 2000, 28(10): 2084–2090. doi: 10.1093/nar/28.10.2084.
- [17] ZHOU M, LONG W, LI X. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants [J]. For Stud China, 2008, 10(4): 235–242. doi: 10.1007/s11632-008-0047-1.
- [18] SCHUETTPELZ E, SCHNEIDER H, HUIET L, et al. A molecular phylogeny of the fern family Pteridaceae: Assessing overall relationships and the affinities of previously unsampled genera [J]. Mol Phylogenet Evol, 2007, 44(3): 1172–1185. doi: 10.1016/j.ympev.2007.04.011.
- [19] PRINCE V E, PICKETT F B. Splitting pairs: The diverging fates of duplicated genes [J]. Nat Rev Genet, 2002, 3(11): 827–837. doi: 10.1038/nrg928.

附录 1 蕨类物种信息

附录 2 94 种蕨类植物的系统发育树

附录 3 *ycf94* 基因的密码子碱基组成与 ENC 值

附录 4 *ycf94* 基因分子进化速率

http://jtsb.ijournals.cn/ajax/common/download_attache_file.aspx?seq_id=20230218234211003&file_no