

• 药材与资源 •

五加科药用植物鲨烯合酶的适应性进化分析

王 卓^{1,2}, 国红玉², 张妍彤², 朱金丽², 邢朝斌^{2*}

1. 华北理工大学药学院, 河北 唐山 063210

2. 华北理工大学生命科学学院, 河北 唐山 063210

摘 要: 目的 对五加科药用植物的鲨烯合酶(squalene synthase, SS)进行适应性进化分析。方法 利用 PAML 软件的分支模型、位点模型、分支-位点模型以及 MEC 模型、SLAC、FEL 和 REL 等方法对五加科 7 种药用植物的 23 条 SS 基因进行适应性进化分析。结果 在 PAML 与 MEC 模型的分析中发现多数分支与位点处于较强的负选择下, 未发现正选择位点; 通过 SLAC、FEL 和 REL 分析同样显示存在大量的负选择位点, 仅 412P、413N、415K 为正选择位点。结论 负选择对五加科 SS 基因起主导作用, 而发现的 412P、413N、415K 位点可能与 SS 的活性有关。

关键词: 五加科; 鲨烯合酶; 适应性进化; SLAC; FEL; REL

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2019)14-3413-07

DOI: 10.7501/j.issn.0253-2670.2019.14.022

Adaptive evolution analysis of squalene synthase in medicinal plants of Araliaceae

WANG Zhuo^{1,2}, GUO Hong-yu², ZHANG Yan-tong², ZHU Jin-li², XING Zhao-bin²

1. School of Pharmacy, North China University of Science and Technology, Tangshan 063210, China

2. College of Life Sciences, North China University of Science and Technology, Tangshan 063210, China

Abstract: Objective To analyze the adaptive evolution of squalene synthase (SS) in the medicinal plants of Araliaceae. **Methods** The adaptive evolutionary analysis of 23 SS genes of seven medicinal plants of Araliaceae was carried out by using the branch model, site model, branch-site model, MEC model, SLAC, FEL, and REL of PAML software. **Results** In the analysis of PAML and MEC models, most of the branches and loci were found to be under strong negative selection, and no positive selection sites were found. The analysis by SLAC, FEL, and REL also showed that there were a large number of negative selection sites, only 412P, 413N, and 415K were positive selection sites. **Conclusion** This indicates that negative selection plays a leading role in SS gene of Araliaceae. The 412P, 413N, and 415K sites found may be involved in the activity of SS.

Key words: Araliaceae; squalene synthase; adaptive evolution; SLAC; FEL; REL

五加科(Araliaceae)全科约有 80 属 900 余种, 我国有 21 属 150 种^[1], 药用价值丰富, 共 114 种植物可药用, 其中不乏人参 *Panax ginseng* C. A. Mey.、三七 *P. notoginseng* (Burk.) F. H. Chen 和刺五加 *Eleutherococcus senticosus* (Rupr. Maxim.) Harms 等名贵药材。多数具有抗衰老、调血脂、滋补、活血补气、利脾补肾等药理活性^[2]。其中三萜皂苷类化合物是其主要有效活性成分之一, 具有极高的经济与药用价值^[3]。药用植物在长期的生物进化历程中, 因不同环境的选择使其产生与相应环境高度适应的次生代谢产物, 如五加科中的萜类化合物等, 这些

次生代谢产物往往能增强植物自身抵御逆境的能力或能更好地繁育后代^[4]。一般认为包括催化药用植物次生代谢产物生物合成的关键酶在内的重要蛋白质所承受的选择压力也会较强^[5]。在分子水平度量选择压力时, 可通过估算蛋白编码基因序列的核苷酸非同义替换率(d_N)与同义替换率(d_S)的比值(ω)来推测出其进化趋势与所经受的选择压力。当 $d_N = d_S$, 即 $\omega = 1$, 表现为中性选择; 当 $d_N < d_S$, 即 $\omega < 1$, 表明受到负选择; 当 $d_N > d_S$, 即 $\omega > 1$, 表明出现了正选择。若 ω 显著大于 1 可被视作蛋白质发生适应性进化的证据^[6]。同时, 通过对蛋白质进行适应

收稿日期: 2019-02-12

基金项目: 国家自然科学基金项目(31570683); 河北省教育厅资助科研项目(ZD2019075); 华北理工大学培育基金资助(SP201508)

作者简介: 王 卓(1993—), 男, 在读硕士研究生, 研究方向为生物信息在药用植物中的应用。

*通信作者 邢朝斌(1975—), 男, 教授, 研究方向为药用植物学。Tel: (0315)8805590 E-mail: xzbheuu@126.com

性进化分析还可为识别重要的氨基酸结构和功能位点提供信息^[7]。

鲨烯合酶(squalene synthase, SS)是催化三萜皂苷生物合成第一步酶促反应的重要调控酶,催化 2 分子的法呢基焦磷酸(farnesyl diphosphate, FPP)产生 1 分子的鲨烯,是引导产物进入三萜皂苷类化合物合成途径的关键酶之一^[8]。随着生物技术的不断成熟,通过对 SS 基因的表达调控提高三萜类皂苷产量的研究逐步成为热点^[9]。但五加科药用植物 SS 的适应性进化特点及重要的氨基酸结构和功能位点等尚未有系统分析。本实验利用(PAML, phylogenetic analysis by maximum likelihood)软件

中的分支、位点、分支-位点模型,同时结合 MEC 模型、随机效应似然法 REL、固定效应似然法对五加科药用植物的 SS 基因序列进行适应性进化分析,以期深入认识五加科药用植物三萜类物质生物合成途径的进化及通过生物技术手段改良 SS 基因,提高三萜皂苷的含量等提供理论和技术基础。

1 材料与方法

1.1 序列数据及系统发育树的构建

本研究所用五加科 SS 基因序列均为 GenBank 数据库中已测序发表的五加科植物的 SS 基因序列。以木榄 *Bruguiera gymnorhiza* (L.) Poir. 的 SS 基因序列为外类群,共计 24 条序列(表 1)。

表 1 所用植物材料名称及其 SS 基因 GenBank 登录号
Table 1 Plant materials and their SS gene Gankbank accession number

学名	GenBank 登录号
辽东槲木 <i>Aralia elata</i>	GU354 313.1 ^[10]
竹节参 <i>Panax japonicus</i>	KP890 782.1
西洋参 <i>Panax quinquefolius</i>	GU997 681.1
人参 <i>Panax ginseng</i>	GU183 406.1、GQ468 527.2、AB115 496.1 ^[11] 、EU502 717.1、KP689 323.1 ^[12] 、KP689 322.1 ^[12] 、KP689 321.1 ^[12] 、KP689 320.1 ^[12] 、KP689 314.1 ^[12]
<i>Panax sokpayensis</i>	KT936 528.1
刺五加 <i>Eleutherococcus senticosus</i>	JN714 465.1、HQ456 922.1、HQ456 921.1、HQ456 920.1、HQ456 919.1、HQ456 918.1、HQ456 917.1
三七 <i>Panax notoginseng</i>	KC953 032.1 ^[13] 、KT123 897.1、DQ186 630.1
木榄 <i>Bruguiera gymnorhiza</i>	GU478 981.1

使用 ClustalW 软件对五加科植物的 SS 基因序列进行比对,使用 DAMBE v5.3.48 检测比对序列的置换饱和度,通过 Jmodeltest 2.1.7 计算获得最佳氨基酸序列进化模型,进而作为构建发育树的参数设定。利用软件 PhyML3.0 构建^[14]五加科药用植物 SS 基因序列的系统发育树。

1.2 进化分析

根据上述得到的系统发育树,利用基于机理式模型的 PAML 软件中 Comdelc 程序^[15]的分支模型、位点模型、分支位点模型对五加科的 SS 基因在进化过程中所受到的选择压力进行分析。

分支模型^[16]允许不同的分支存在不同的 ω 值。采用单一比率(one ratio)模型假定进化树上所有分支的 ω 比值都一致,即为 ω_0 ;自由比率(free ratio)模型作为最全面的模型则允许不同分支存在不同的 ω 比值,同时还通过二比率(two ratio)模型假设特定前景支 ω 值与背景支不同,通过两模型进行

LRT 检验,得到两者是否存在差异。位点模型^[17]假设不同位点的 ω 值不同,而树上的各分支 ω 值并无差异。此次研究采用 3 对模型,即 M1a(近中性)对 M2a(选择),M0(单一比值)对 M3(离散),M7(beta)对 M8(beta& ω)^[16,18]。对 3 对模型进行 LRT 检验确定模型之间是否存在差异。其中 M0 对 M3 的比较是为了验证不同位点的 ω 值是否不同,而并非为了检测证选择位点。而 M7(beta)对 M8(beta& ω)易出现假阳性结果,所以本实验组以 M1a(近中性)对 M2a(选择)的检验结果为可能的正选择位点。分支-位点模型^[19]为了检测特定分支是否有存在正选择位点,将发育树分为前景支与背景支。为了减少假阳性结果出现的可能,采用分支-位点模型 A 的 TEXT 2 鉴别特定前景分支的正选择位点^[11]。

此外,将序列数据提交到 Datamonkey (<http://www.datamonkey.org/>)^[20]和 MEC (<http://selecton>。

tau.ac.il/)。通过 Datamonkey Web 服务器中的随机效应似然模型 (random effects likelihood model, REL)、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) 对位点的选择压力进行在线分析, 对于 SLAC, FEL 方法, 处于 $P < 0.1$ 水平的位点即可认为是受正选择的影响, 对于方法 REL 而言应用贝叶斯因子 < 50 即可接受。

1.3 正选择位点相对位置的确定

以 SS(伞形目, 五加科, 登录号为 HQ456922.1) 的氨基酸序列为参考序列, 利用在线软件 TMHMM 预测五加科 SS 蛋白的跨膜区域, 通过 NPSA、NCBI 预测二级结构、保守结构域与功能位点, Genedoc 软件显示结果。

2 结果与分析

2.1 系统发育分析

在进行系统发育和正选择试验之前, 通过置换饱和度的检验, 以测试序列是否适合于系统发育和选择分析, 结果表明替代饱和指数 (I_{ss}) 显著低于临界 I_{ss} ($I_{ss.c}$)。基于得到的五加科 SS 序列, 通过 JModeltest 软件计算得到核苷酸最优替换模型 TrN+I+G, 以木榄的 SS 基因序列为外类群, 建立系统发育树 (图 1)。结果表明, 五加科 SS 基因构建的系统发育树主要分为 2 大支, 其中刺五加的 SS 独自成一支, 并处在最基部的位置, 另一只则是由人参、西洋参 *Panax quinquefolius* L.、竹节参 *Panax japonicus* C. A. Mey.、辽东槲木 *Aralia elata* (Miq.) Seem.、三七等组成。而人参 SS2、SS3 又单独构成一分支, 其余如三七、西洋参、竹节参、辽东槲木等聚为一支。而三七与辽东槲木、刺五加分离株 YET-1 互为姊妹类群处于该支的核心位置。

2.2 选择压力检测

2.2.1 运用 PAML 检测选择压力 为了评估自然选择对五加科 SS 的影响并鉴定受正选择的特定位点, 使用 PAML v4.5 包的 CODEML 程序以 PHYML3.0 得到的系统发育树为基础, 对五加科 SS 基因进行选择压力分析。如表 2 所示, 在分支模型中单比率模型 M0 的似然值 ($\ln L$) = -3 887.584 255, 设计参数 $np=48$; 自由比率模型 Mf 的 $\ln L$ = -3 845.074 970, 设计参数 $np=93$; 通过自由比率与单比率模型的 LRT 检验得到自由比模型优于单比率模型的结果 (表 3), 说明不同的分支所经受的选择压力存在差异。在二比率模型中, 分别以 A、B、C 为前景支,

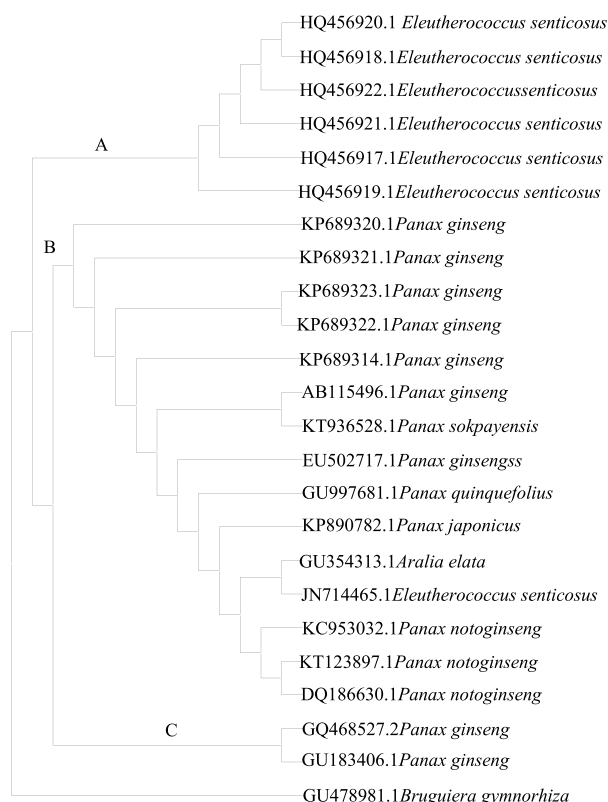


图 1 基于 SS 基因序列构建的系统发育树

Fig. 1 Phylogenetic tree established based on SS gene sequence

其他分支为背景支。经 LRT 检验得到仅二比率模型 A 优于单比率模型 (表 3)。

为了确定在序列的一些特定位点或区域上是否存在多样选择的可能性, 通过比较位点模型下的 3 对模型: M0 (单比率) 对 M3 (离散), M1a (近中性) 对 M2a (选择), M7 (beta) 对 M8 (beta 和 ω) 来检测。根据五加科 SS 基因数据, 得算出各模型相对应的设计参数值与似然值 (表 3)。在 M0 和 M3 模型比较中自由度 (df) = 4、 $2\Delta\ln L = 101.019 5$, 根据 χ^2 分布, 可确认 $P < 0.01$, 其备选假设模型 M3 成立, M3 显著优于 M0, 说明各个位点间所承受的选择压力是存在差异; 由 M3 模型 ω_1 、 ω_2 、 ω_3 的值 0、0.358 01、1.759 41 可知受到较强的纯化选择影响的位点约占 95% 左右, 证明在 SS 基因进化过程中纯化选择占主导地位; 由 M1a 与 M2a 模型 $2\Delta\ln L = 0$ 、 $P = 1$, 可知备选假设模型 M2a 并不成立。在模型 M7 与 M8 的比较中, 在模型 M7 与 M8 比较中 $df = 2$ 、 $2\Delta\ln L = 6.046 784$, 得到 $P = 0.048$, 即概率 95% 的水平上备选模型 M8 成立。根据其 ω 值 $\omega = 6.046 784 < 1$, 可知可能存在正选择

表 2 各模型参数估计值和对数似然值

Table 2 Parameter estimates and log-likelihood values for different models

模型	模型	参数	$\Delta \ln L$	参数估计值	正选择位点
分支模型 1	单比率 M0	48	-3 887.584 255	$\omega=0.166\ 91$	不允许
	二比率 MA	49	-3 871.902 144	$\omega_0=0.276\ 59$ 、 $\omega_1=0.085\ 56$	不允许
	二比率 MB	49	-3 886.128 084	$\omega_0=0.160\ 28$ 、 $\omega_1=0.355\ 00$	不允许
	二比率 MC	49	-3 887.460 514	$\omega_0=0.168\ 94$ 、 $\omega_1=0.127\ 92$	不允许
	自由比率 MF	93	-3 844.945 315	$\omega_A=1.316\ 78$ 、 $\omega_B=0.374\ 26$ $\omega_C=0.093\ 25$	不允许
位点模型	M1a:近中性	49	-3 839.394 674	$P_0=0.823\ 46$ 、 $\omega_0=0.055\ 04$ $P_1=0.176\ 54$ 、 $\omega_1=1.000\ 00$	不允许
	M2a:正选择	51	-3 839.394 674	$P_0=0.823\ 45$ 、 $\omega_0=0.055\ 04$ $P_1=0.004\ 43$ 、 $\omega_1=1.000\ 00$	无
	M3:离散	52	-3 837.074 513	$P_0=0.579\ 89$ 、 $\omega_0=0.000\ 00$ $P_1=0.376\ 81$ 、 $\omega_1=0.358\ 01$ $P_2=0.043\ 30$ 、 $\omega_2=1.759\ 41$	无
	M7:beta	49	-3 840.173 063	$P=0.157\ 60$ 、 $q=0.581\ 60$	不允许
	M8: beta& ω	51	-3 837.149 671	$P_0=0.963\ 30$ 、 $P=0.178\ 84$ $q=0.873\ 86$ 、 $P_1=0.036\ 70$ $\omega=1.850\ 36$	无
分支-位点模型	MA	51	-3 838.870 408	$P_{2a}=0.117\ 08$ 、 $P_{2b}=0.025\ 38$ $\omega_{b1}=0.521\ 00$ 、 $\omega_{b2}=1.000\ 00$ $\omega_{f1}=1.000\ 00$ 、 $\omega_0=1.000\ 00$	无
	MA0	50	-3 838.870 408	$P_{2a}=0.117\ 07$ 、 $P_{2b}=0.025\ 37$ $W_{b1}=0.521\ 00$ 、 $W_{b2}=1.000\ 00$ $W_{f1}=1.000\ 00$ 、 $W_{f2}=1.000\ 00$	不允许
	MB	51	-3 838.456 609	$P_{2a}=0.108\ 08$ 、 $P_{2b}=0.022\ 10$ $W_{b1}=0.543\ 4$ 、 $W_{b2}=1.000\ 00$ $W_{f1}=1.032\ 95$ 、 $W_0=1.032\ 95$	无
	MB0	50	-3 838.456 758	$P_{2a}=0.111\ 48$ 、 $P_{2b}=0.022\ 79$ $W_{b1}=0.543\ 4$ 、 $W_{b2}=1.000\ 00$ $W_{f1}=1.032\ 95$ 、 $W_0=1.032\ 95$	不允许
	MC	51	-3 839.562 010	$P_{2a}=0.069\ 10$ 、 $P_{2b}=0.014\ 78$ $W_{b1}=0.055\ 11$ 、 $W_{b2}=1.000\ 00$ $W_{f1}=3.825\ 66$ 、 $W_0=3.825\ 66$	无
	MC0	50	-3 839.394 675	$P_{2a}=0.000\ 00$ 、 $P_{2b}=0.000\ 00$ $W_{b1}=0.055\ 04$ 、 $W_{b2}=1.000\ 00$ $W_{f1}=1.000\ 00$ 、 $W_0=1.000\ 00$	不允许

位点。在模型 M8 检测中共得到 11 个位点 (99 R 0.574、196 S 0.503、366 T 0.609、369 E 0.650、372 T 0.556、393 A 0.622、395 I 0.579、408 E 0.603、409 L 0.630、411 P 0.844、413 K 0.868)，但均小于 95%

的标准，故上述位点均不成立。在分支-位点模型，本次研究的以分支 A、B、C 为检验正选择的前景支，其他则为背景支，由上述分支模型结果可知仅分支 A 的模型优于单比率模型 M0，经检验 2 验证，

发现其备选假设并不优于零假设。

2.2.2 基于 MEC 模型分析选择压力 将数据输入 selecton 服务器以 MEC 模型作为备选模型, 于三七 (登录号为 KT123897.1) 的 SS 一级结构上进行标注。结果如图 2 所示, 紫色标注位点占主要部分,

表 3 LRT 检验统计量

Table 3 Likelihood ratio statistics

模型比较	df	2ΔlnL	P 值
M0&MA	1	31.364 20	<0.010
M0&MB	1	2.912 34	0.080
M0&MC	1	0.247 48	0.610
M0&MF	45	85.277 90	<0.010
M0&M3	4	101.019 00	<0.010
M1a&M2a	2	0	1.000
M7&M8	2	6.046 78	0.048
Ma0&Ma	1	0	1.000
Mb0&Mb	1	0.000 29	0.990
Mc0&Mc	1	0.334 67	0.570

MGSLGAILKHPDDFYPLLLKLTAAARHAEKQIPPEPHWAFCSYMLHKVSRSGFGLVIOQLGPF
FGLVIOQLGPF LRDAVCISYLVLRALDTVEDDTSIPEVKVEILMAEHRHIYDKDWHFSCGCTKEYKVLMD
YDKDWHFSCGCTKEYKVLMD TKEKVLMDDEFHVSNAFLEL GNGYQEAIEDITMRMGAGMAKFI
KFIKEVETIDDDYDEYCHYVAGLVGLGLSKLFLHASGAEDLATDSLNSNSMGLFLOKTNIRDYLEDINEIPKSRMFWPQIWSIYVDKLE
LFLQKTNIRIRDYLEDINEIPKSRMFWPQIWSIYVDKLE WSKYVDKLED LKYEENSAKAVRCLNDMVTNALLHVEDCLKYM
VRCLNDMVTNALLHVEDCLKYM SDRGPAIFRCAIPQIMAIGTLALCEDLKYEENSAKAVRCLNDMVTNALLHVEDCLKYM
NTQVFRGVVKLRGLTAKVIDRNTMSDVYGAFFDFSCMLKSKVDNNDPNATKTLSRLFNNTQVFRGVVKLRGLTAKVIDRNTMSDVYGAFFDFSCMLKSKVDNNDPNATKTLSRL
ATKTLSRLEAATKTLSRLEAATKTLSRLEAATKTLSRLEAATKTLSRLEAATKTLSRLEAATKTLSRLEAATKTLSRLEAATKTLSRLEAATKTLSRLEA
YAYLELLQPNKQ

浅黄为经历了正选择的位点, 白色为中性位点, 由白到紫为经历了净化选择的位点, 金黄和深紫则是在统计学中具有显著性的正选择位点与负选择位点

Light yellow is site of positive selection, white is neutral sites, white to purple is site of purification selection, and golden and deep purple are statistically significant positive selection sites and negative selection site

图 2 五加科 SS 基因 MEC 模型的适应性进化结果

Fig. 2 Adaptive evolution results of SS gene MEC model

1 MGSLGAILKHPDDFYPLLLKLTAAARHAEKQIPPEPHWAFCSYMLHKVSRSGFGLVIOQLGPF 60
2 MGSLGAILKHPDDFYPLLLKLTAAARHAEKQIPPEPHWAFCSYMLHKVSRSGFGLVIOQLGPF 60
3 MGSLGAILKHPDDFYPLLLKLTAAARHAEKQIPPEPHWAFCSYMLHKVSRSGFGLVIOQLGPF 60
4 QLRDAVCISYLVLRALDTVEDDTSIPEVKVEILMAEHRHIYDKDWHFSCGCTKEYKVLMD 120
5 QLRDAVCISYLVLRALDTVEDDTSIPEVKVEILMAEHRHIYDKDWHFSCGCTKEYKVLMD 120
6 EFHHVSNFAFLELCSGYKDAIEDITMRMGAGMAKFIKEVETIDDDYDEYCHYVAGLVGLGL 180
7 EFHHVSNFAFLELCSGYKDAIEDITMRMGAGMAKFIKEVETIDDDYDEYCHYVAGLVGLGL 180
8 EFHHVSNFAFLELCSGYKDAIEDITMRMGAGMAKFIKEVETIDDDYDEYCHYVAGLVGLGL 180
9 SKLFLHASGAEDLATDSLNSNSMGLFLOKTNIRDYLEDINEIPKSRMFWPQIWSIYVDKLE 240
10 SKLFLHASGAEDLATDSLNSNSMGLFLOKTNIRDYLEDINEIPKSRMFWPQIWSIYVDKLE 240
11 SKLFLHASGAEDLATDSLNSNSMGLFLOKTNIRDYLEDINEIPKSRMFWPQIWSIYVDKLE 240
12 EDLKYEENSAKAVRCLNDMVTNALLHVEDCLKYM SDRGPAIFRCAIPQIMAIGTLALCEDLKYEENSAKAVRCLNDMVTNALLHVEDCLKYM 300
13 EDLKYEENSAKAVRCLNDMVTNALLHVEDCLKYM SDRGPAIFRCAIPQIMAIGTLALCEDLKYEENSAKAVRCLNDMVTNALLHVEDCLKYM 300
14 YNNIQVFRGVVKLRGLTAKVIDRNTMSDVYGAFFDFSCMLKSKVDNNDPNATKTLSRLFNNTQVFRGVVKLRGLTAKVIDRNTMSDVYGAFFDFSCMLKSKVDNNDPNATKTLSRL 360
15 FNNTQVFRGVVKLRGLTAKVIDRNTMSDVYGAFFDFSCMLKSKVDNNDPNATKTLSRLFNNTQVFRGVVKLRGLTAKVIDRNTMSDVYGAFFDFSCMLKSKVDNNDPNATKTLSRL 360
16 EAIQKTCRSGTLTKRKSYYIIESGSHNSALIAIIFITLAILYAYLSSNLPNSL---- 415
17 EAIQKTCRSGTLTKRKSYYIIESGSHNSALIAIIFITLAILYAYLSSNLPNSL---- 415
18 EAIQKTCRSGTLTKRKSYYIIESGSHNSALIAIIFITLAILYAYLSSNLPNSL---- 415

1-刺五加 HQ456922.1 2-人参 KP689322.1 3-三七 KC953032.1 选取的参考序列, 通过 Gendoc 软件显示了 SS 在 3 种常见药用植物参考序列上的正选择位点; 保守位点用阴影表示, 每个符号的含义在特定的符号中给出: # 为正选择位点, * 为 N-糖基化位点

1-reference sequence selected by *Eleutherococcus senticosus* HQ456922.1 2-ginseng KP689322.1 3-*Panax notoginseng* KC953032.1, shows the positive selection position of SS on the reference sequences of 3 common medicinal plants by Gendoc software; Conserved loci are indicated by shading, and the meaning of each symbol is given in a specific symbol: # is a positive selection site; * is an N-glycosylation site

图 3 五加科 SS 基因氨基酸序列比对图

Fig. 3 Alignment diagram of amino acid sequence of SS gene of Araliaceae

2.3 基于 SS 三维结构的正选择位点的分析

以刺五加分离株 EWLS-1 鲨烯合酶（伞形目五加科，登录号为 HQ456922.1）的氨基酸序列为参考序列，通过 NPSA、NCBI、TMHMM 网站预测 SS 的保守区域以及跨膜区域，并根据上述得到的正选择检测结果，确定正选择位点的相对位置。

结果显示，SS 基因的二级结构中 α 螺旋占 65.78%，无规卷曲占 24.58%，延伸链占 6.75%， β -转角占 2.89%， α 螺旋与无规卷曲为 SS 蛋白结构的主要组成，见图 4。415K 位于糖基化位点见图 3，412P、413K 位于无规卷曲上，415K 位于 α 螺旋上。

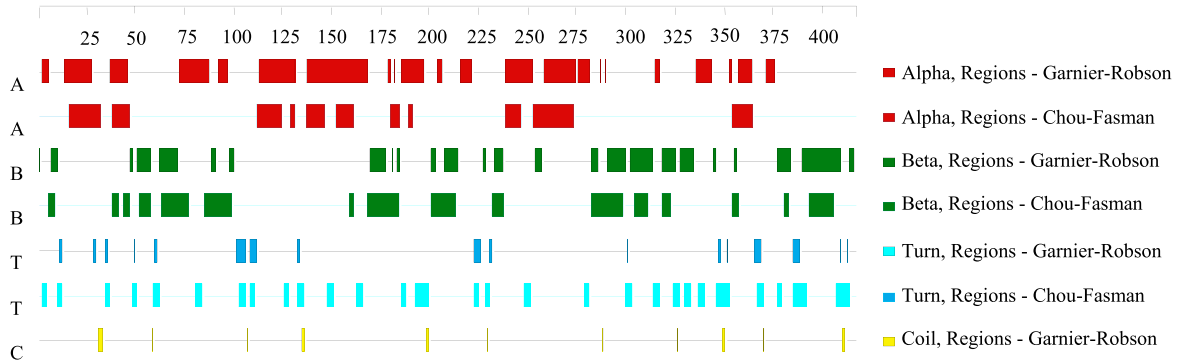


图 4 五加科 SS 蛋白二级结构预测

Fig. 4 Prediction of secondary structure of SS protein in Araliaceae

2.4 SS 的保守结构域分析

以刺五加分离株 EWLS-1 鲨烯合酶（伞形目五加科，登录号为 HQ456922.1）的氨基酸序列为参考序列，通过 NCBI 保守结构数据库 CDD 分析五加科 SS 蛋白的保守区域。发现 SS 蛋白中同时存在有反式异戊二烯二磷酸合成酶（Trans_IPPS_HH）、角鲨烯/八氢番茄红素合成酶（SQS_PSY）、八氢番茄红素/角鲨烯合成酶 [脂质转运和代谢]（ERG9）3 种保守结构。

3 讨论

经 PAML 的研究显示，五加科植物 SS 基因各个分支所承受的选择压力存在差异，其中大多数分支主要承受着净化选择的影响即 $\omega < 1$ 。这同时也说明了五加科 SS 基因在进化过程中较为保守，主要为净化选择起主导作用。在分支模型中检测出分支 A 有存在正选择位点的可能，但分支位点模型并未在 A 支检测出正选择位点，基于此前曾报道过类似的问题^[21]，Yang 等^[22]和 Zhang 等^[23]也针对此类选择压力情况对模型做出了改进，故可以排除是假阳性的可能，推测是由于选择压力的放松所致。位点模型与分支-位点模型并未检测到正选择位点的存在。由于五加科 SS 酶作为三萜类化合物合成途径中关键酶^[9]，需要维持其结构的稳定性以确保其功能，故在进化过程中较为保守。或是参与分析比对的数据所包含的变异较少或基因在较早时期发

生的适应性进化，其进化的信号被普遍的中性或净化选择淹没^[24]。

研究中通过 PAML、MEC 模型、Datamonkey 3 种方式对 23 种五加科植物次生代谢途径中的关键酶 SS 进行了分析。结果显示，仅使用 Datamonkey 得到统计意义显著的正选择位点 412P、413N、415K，进而比对发现 412P、413N 仅在刺五加各分离株上为“P”“N”，而在人参、三七等均不相同，推测位点可能与 SS 基因功能活性有关，但有待进一步探究。而位于糖基化位点的 415K 的变化可能会对该区域的糖链的连接产生影响进而影响 SS 酶活性。

通过 NCBI 进行 BLAST 比对发现 SS 序列中存在有 Trans_IPPS_HH、SQS_PSY、ERG9 的保守结构^[25]，而三者同属于反式异戊二烯基二磷酸合成酶（IPPS）和 I 类萜烯环化酶超家族（cl00210）。其催化位点则是由反平行 α 螺旋为主要结构的大中心腔组成，其腔壁上 2 个富含天冬氨酸的区域，作为催化的关键位点，通过桥接 Mg^{2+} 介导异戊二烯基磷酸酯的结合。以反式异戊二烯二磷酸合成酶（Trans_IPPS_HH）为对照，得到其富含天冬氨酸的区域为点位 50 位至 55 位的 DTLED 与 190 位至 195 位的 DYLED，且均受到不同程度净化选择的影响^[26]，为了保证维持其特定的功能，故在进化过程中保持高度保守。

参考文献

- [1] Kim K, Nguyen V B, Dong J, *et al.* Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 *Panax*-related species [J]. *Sci Rep*, 2017, 7(1): 4917-4926.
- [2] 任玉新, 金立弟, 韩 宇. 五加科植物多糖的药理作用研究进展 [J]. 中国药业, 2012, 21(6): 83-86.
- [3] Kim T D, Ham J Y, Hu G H, *et al.* Expression and functional characterization of three squalene synthase genes associated with saponin biosynthesis in *Panax ginseng* [J]. *Plant Cell Physiol*, 2011, 52(1): 125-137.
- [4] Peng S L, Nan P, Zhong Y, *et al.* Terpenoids in higher plants and their roles in ecosystems [J]. *Chin J Ecol*, 2002, 21(3): 33-38.
- [5] 许 可, 王 博, 苏应娟, 等. 蕨类植物 *psbD* 基因的适应性进化和共线性分析 [J]. 植物科学学报, 2013, 31(5): 429-438.
- [6] Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models [J]. *Mol Biol Evol*, 2000, 17(1): 32-43.
- [7] 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.
- [8] Jennings S M, Tsay Y H, Fisch T M, *et al.* Molecular cloning and characterization of the yeast gene for squalene synthetase [J]. *Proc Natl Acad Sci USA*, 1991, 88(14): 6038-6042.
- [9] Kim Y S, Chon J H, Park S, *et al.* Gene regulation patterns in triterpene biosynthetic pathway driven by overexpression of squalene synthase and methyl jasmonate elicitation in *Bupleurum falcatum* [J]. *Planta*, 2011, 233(2): 343-355.
- [10] 吴 颖, 成 慧, 邹宏达, 等. 辽东槲木 *AeSS* 基因的克隆及植物表达载体的构建 [J]. 中药材, 2010, 33(9): 1363-1365.
- [11] Lee M H, Jeong J H, Seo J W, *et al.* Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene [J]. *Plant Cell Physiol*, 2004, 45(8): 976-984.
- [12] Cao H, Nuruzzaman M, Xiu H, *et al.* Transcriptome analysis of methyl jasmonate-elicited *Panax ginseng* adventitious roots to discover putative ginsenoside biosynthesis and transport genes [J]. *Int J Mol Sci*, 2015, 16(2): 3035-3057.
- [13] Niu Y, Luo H, Sun C, *et al.* Expression profiling of the triterpene saponin biosynthesis genes *FPS*, *SS*, *SE*, and *DS* in the medicinal plant *Panax notoginseng* [J]. *Gene*, 2014, 533(1): 295-303.
- [14] Guindon S, Dufayard J F, Lefort V, *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3. 0 [J]. *Systematic Biol*, 2010, 59(3): 307-321.
- [15] Yang Z H. PAML 4: Phylogenetic analysis by maximum likelihood [J]. *Mol Biol Evol*, 2007, 24(8): 1586-1591.
- [16] Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution [J]. *Mol Biol Evol*, 1998, 15(5): 568-573.
- [17] Yang Z H, Swanson W J, Vacquier V D. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites [J]. *Mol Biol Evol*, 2000, 17(10): 1446-1455.
- [18] Yang Z H, Bielawski J P. Statistical methods for detecting molecular adaptation [J]. *Trends Ecol Evol*, 2000, 15(12): 496-503.
- [19] Hong S, Spreitzer R J. Complementing substitutions at the bottom of the barrel influence catalysis and stability of ribulose-bisphosphate carboxylase/oxygenase [J]. *J Biol Chem*, 1997, 272(17): 11114-11117.
- [20] Delpont W, Poon A F, Frost S D, *et al.* Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology [J]. *Bioinformatics*, 2010, 26(19): 2455-2457.
- [21] 巩超彦, 南芳茹, 冯 佳, 等. 串珠藻目植物 *rbcL* 基因的适应性进化分析 [J]. 海洋与湖泊, 2017, 48(3): 527-535.
- [22] Yang Z H, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages [J]. *Mol Biol Evol*, 2002, 19(6): 908-917.
- [23] Zhang J Z, Nielsen R, Yang Z H, *et al.* 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.
- [24] Marchler-Bauer A, Bo Y. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures [J]. *Nucleic Acids Res*, 2017, 45(1): 200-203.
- [25] 王 静, 王 艇, 苏应娟, 等. 裸子植物中光敏色素 *PHY-PAS1* 结构域的适应性进化 [J]. 植物学报, 2009, 44(5): 608-618.
- [26] 刘 镛, 吴耀生, 胡艳玲, 等. 陆生植物鲨烯合酶适应性进化正选择位点分析 [J]. 中国生物化学与分子生物学报, 2013, 29(1): 91-97.