

被子植物 SQUA 类基因适应性进化的统计检测^{*}

陈永燕^{1,2} 钟扬^{3**} 田波¹ 杨继⁴ 李德铎^{1**}

1. 中国科学院昆明植物研究所生物多样性与生物地理学开放研究实验室, 昆明 650204;

2. 大连大学生物工程学院, 大连 116622;

3. 复旦大学生物多样性与生态工程教育部重点实验室, 上海 200433;

4. 北京大学生命科学学院, 北京 100871

摘要 SQUAMOSA (SQUA) 类基因是 MADS-box 基因家族中参与花被发生发育的重要基因. 测定了麻竹的 SQUA 基因序列并分析了被子植物中 SQUA 类基因的系统发育式样; 采用相对速率检验、同义与非同义置换位点统计以及似然比检验方法, 分析了单子叶和真双子叶植物最近共同祖先中 SQUA 基因发生基因重复后的进化速率与适应机制. 结果表明: 真双子叶植物分支中 SQUA 类基因的相对速率和同义与非同义置换位点均显著高于单子叶植物分支; 同时, d_N/d_S 值表明真双子叶植物分支中可能存在正选择压力.

关键词 被子植物 麻竹 SQUA 类基因 相对速率检验 似然比检验 适应性进化

调控花器官发生发育的 MADS-box 基因家族在整个植物界中高度保守. ABC 模型^[1]认为花器官发生发育的调控过程在进化上是古老的, 不同的植物类群花部式样则随基因的重复和分化逐渐形成. 由于基因重复在进化过程中频繁发生, 人们可以推测细胞核基因的大部分是由一个基因经基因重复所形成的基因家族构成的. 例如, 植物中 MADS-box 基因家族是一类调控花器官形成和分化的基因, 其直系同源成员的进化关系在很大程度上可以反映被子植物花的进化过程, 这对于被子植物花的起源与进化研究具有关键作用.

迄今为止, 人们已从拟南芥 (*Arabidopsis thaliana*)、水稻 (*Oryza sativa*) 和玉米 (*Zea mays*) 等植物中分离到许多 MADS-box 基因家族成员^[2~6], 其中 SQUA 类基因属于 ABC 模型中的 A 类, 单独决定花萼的形成, 也参与决定花瓣的形成. 从形态上看, 单子叶和真双子叶植物的花器官

结构在雄蕊和雌蕊上差别并不大, 主要区别在花被. 单子叶植物花部比较简单, 花被很少出现花萼和花瓣的分化, 花粉单沟, 是一类古草本植物^[7], 而真双子叶植物的花被则大多数明显分化为花萼和花瓣, 且物种间呈现出极大的多样性. 研究结果表明, 在单子叶与真双子叶植物的最近共同祖先中已含有 SQUA 类基因, 但在被子植物的共同祖先中尚未发现该拷贝^[8]. 本文应用多种统计分析方法, 结合从单子叶植物麻竹 (*Dendrocalamus latiflorus*) 中分离到的 SQUA 基因, 对被子植物中 SQUA 类基因的进化速率与适应机制进行检测, 以期被子植物花器官的进化发育研究提供新的证据.

1 材料与方法

来自不同单子叶和双子叶植物的共 32 个 SQUA 类基因的 GenBank 收录号见表 1. 其中麻竹的 SQUA 基因由本实验室自行分离和测序. 在初步

2004-02-10 收稿, 2004-04-05 收修改稿

* 国家自然科学基金 (批准号: 30200015)、国家重点基础研究发展规划 (批准号: 2003CB715904) 和云南省自然科学基金 (批准号: 2002C0056M) 资助项目

** 通讯作者, E-mail: dzl@mail.kib.ac.cn, yangzhong@fudan.edu.cn

分析中, 先对 GenBank 中 225 个 MADS-box 基因序列构建系统树, 再从与单、双子叶 SQUA 基因分支构成姐妹群关系的序列中选取 ZmZMM 28, 作为对 32 个单子叶和双子叶植物序列进行系统发育分析的外类群。

表 1 本文所用 SQUA 类基因植物种名、缩写、基因名及 GenBank 收录号^{a)}

植物物种	缩写	基因名	GenBank 收录号
金鱼草 <i>Antirrhinum majus</i>	Am	DEFH28	AY040247
拟南芥 <i>Arabidopsis thaliana</i>	At	AGL8	U33473
		CAL	L36925
		API	Z16421
茼蒿 <i>Chrysanthemum x morifolium</i>	Cm	CDM41	AY173055
		CDM111	AY173054
		CDM8	AY173056
麻竹 <i>Dendrocalamus latiflorus</i>	Dl	API like	AY395714 ^{a)}
		MI9	AY395715 ^{a)}
蓝桉 <i>Eucalyptus globulus</i>	Eg	API	AF305076
		AP2L	AF306349
向日葵 <i>Helianthus annuus</i>	Ha	M75	AF462152
大麦 <i>Hordeum vulgare</i>	Hv	M5	AJ249144
		M8	AJ249146
黑麦草 <i>Lolium perenne</i>	Lp	MADS2	AY198327
毒麦 <i>Lolium temulentum</i>	Li	MADS1	AF035378
番茄 <i>Lycopersicon esculentum</i>	Le	MADS-MC	AF448521
		TDR4	X60757
烟草 <i>Nicotiana sylvestris</i>	Ns	MADS2	AF068726
水稻 <i>Oryza sativa</i>	Os	FDRMADS6	AF139664
		MADS14	AF058697
		MADS15	AF058698
碧冬茄 <i>Petunia x hybrida</i>	Ph	FBP29	AF335245
		PFG	AF176782
豌豆 <i>Pisum sativum</i>	Ps	M4	AJ279089
大叶蝇子草 <i>Silene latifolia</i>	Sl	M4	X80491
		M5	X80492
白芥 <i>Sinapis alba</i>	Sa	MADS B	U25695
小麦 <i>Triticum aestivum</i>	Ta	MADS11	AB007504
玉米 <i>Zea mays</i>	Zm	ZMM4	AJ430641
		ZMM15	AJ430632
		ZAP1	L46400
		ZMM28	AJ430695

a) 自测序列

应用 DNASStar 软件包中的 MegAlign 程序分别

对核苷酸和氨基酸序列进行对位排列; 根据氨基酸排列后的矩阵, 去除出现空位的氨基酸所对应的核苷酸, 组成不含空位的核苷酸序列矩阵; 以 PAUP 软件构建的 MP 系统树为基础¹⁾, 采用 Li-Bousquet 方法及 RRTree 程序检测不同分支间的相对速率^[9-11]。

采用 Mega2 软件计算各分支的非同义置换位点数, 并进行差异显著性检验^[12]。采用似然比检验方法和 PAML 软件^[13], 分析单子叶和真双子叶植物最近共同祖先中的 SQUA 基因在基因重复后的适应性进化机制; 密码子置换模型为“branch”模型, 即判断谱系间非同义置换与同义置换的比率 $\omega(d_N/d_S)$ 是否存在显著性差异^[13, 14]。

2 结果与分析

2.1 SQUA 基因树与相对速率检验

基于 32 个 SQUA 类基因序列的 MP 系统树见图 1, 其中单子叶和真双子叶的序列分别构成自展支持率为 100% 的两大分支。

以 ZmZMM 28 为参照类群, 对单子叶和真双子叶两个分支间的核苷酸置换速率进行了差异显著性检验。结果表明, 非同义置换位点的 $d(Ka)/SD = -2.9344$, $P = 0.0033$, 即上述两个分支间的相对速率存在极显著差异, 而真双子叶分支中的速率显著高于单子叶分支。

2.2 似然比检验

采用 Mega2 计算获得单子叶和真双子叶两个分支的 d_N/d_S 分别为 1.04 和 0.59。差异显著性检验结果为: $\Delta d_N/SE = 125.14/8.69 = 14.4$, $\Delta d_S/SE = 108.98/3.6 = 5.27$, 表明同义和非同义置换位点数在上述两个分支间均存在显著差异。

采用 PAML 软件, 在基于密码子置换“branch”模型的似然比检验中, 先采用两种不同的模型来分析谱系间的进化式样: 第 1 种为“one-ratio”模型(假定所有谱系都具有相同的 ω 值), 计算获得对数似然率 $\ln L = -10677.76$, ω 估计值为 0.1220; 第 2 种为“free-ratio”模型(假设每一谱系都具有不同的 ω 值), 对数似然率 $\ln L = -10614.04$ 。

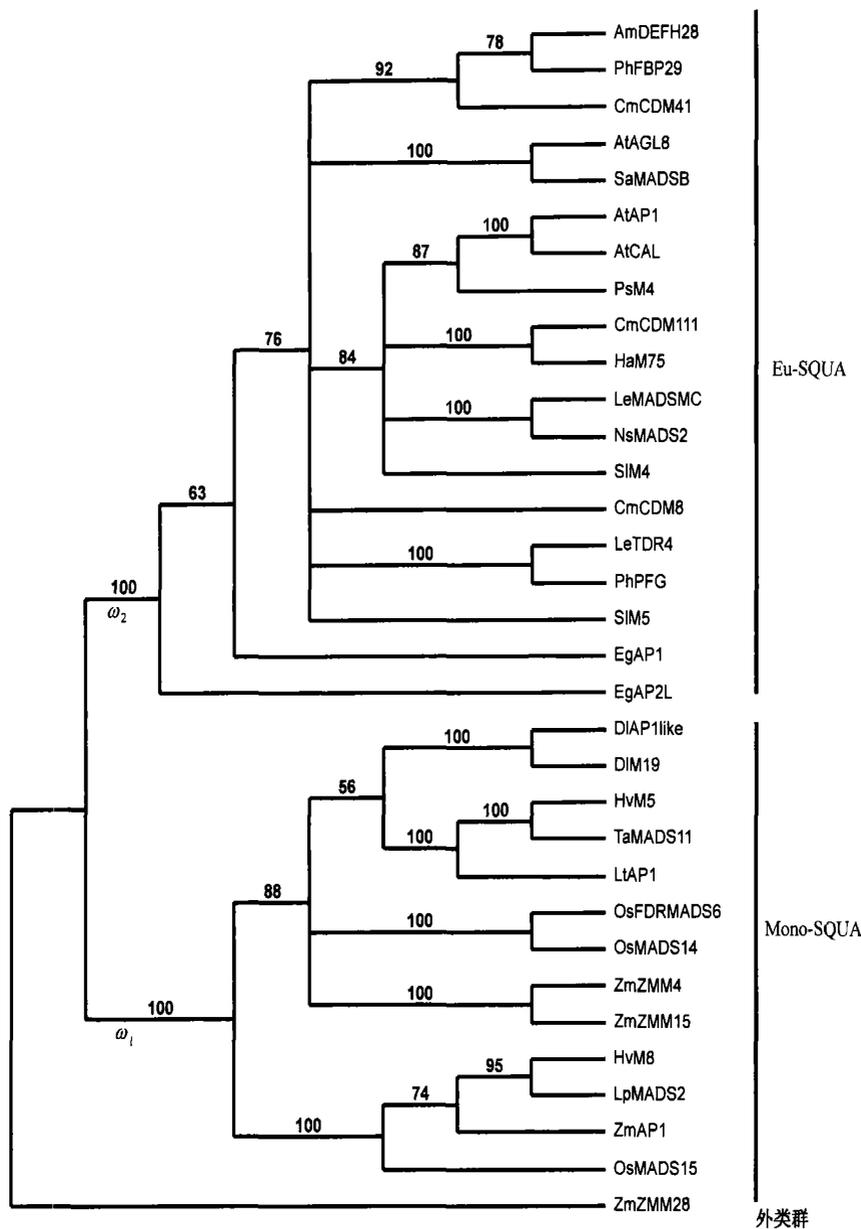


图 1 SQUA 类基因的 MP 系统树

Eu-SQUA 为真双子叶植物分支, Mono-SQUA 为单子叶植物分支. 数字表示该分支的自展支持率(%)

然而, χ^2 检验结果拒绝 “one-ratio” 模型 ($P < 0.01$). 因此, 我们采用第 3 种模型 (“three-ratio” 模型, 假定某些谱系具有和背景 ω_0 不同的 ω 值)^[13], 分别标注单子叶和真双子叶分支为 ω_1 和 ω_2 (图 1), 它们在不同的模型中具有不同的假定 (表 2). 进一步的 χ^2 检验结果拒绝 “free-ratio” 模型 ($P < 0.01$) (表 3), 说明分支内各谱系间的 ω 值差异并不符合

“free-ratio” 模型假定.

表 2 不同模型的似然值以及非同义-同义置换速率比

模 型	$\ln L$	ω_0	ω_1	ω_2
One-ratio; $\omega_0 = \omega_1 = \omega_2$	-10677.76	0.1220	ω_0	ω_0
Free-ratio; $\omega_0, \omega_1, \dots, \omega_{63}$	-10614.04	-	-	-
Three-ratio; $\omega_0, \omega_1, \omega_2$	-10676.07	0.1203	0.1165	0.3092

表 3 不同模型的似然比检验

零假设	已有假设	模型比较	$2 \ln L$
One-ratio	$\omega_0 = \omega_1 = \omega_2$	One-ratio 比 free-ratio	127.44 ^{a)}
Free-ratio	$\omega_0, \omega_1, \omega_2$	Free-ratio 比 three-ratio	123.34 ^{a)}

a) 极显著差异 ($P < 0.01$)

3 讨论

植物花发育的 ABC 模型^[1]认为: A 功能基因决定花萼的形成, A 和 B 基因共同决定花瓣的形成, B 和 C 基因共同决定雄蕊的形成, 心皮则由 C 基因单独决定. SQUA 类基因属于 MADS-box 基因家族中的一类重要成员, 为 ABC 模型中的 A 基因. 本文检测了被子植物中 SQUA 类基因的系统发育型式及其在单子叶和真双子叶植物分支中的进化速率与适应机制, 结果具有以下两方面的意义:

(1) 形态比较研究表明, 除兰科植物外, 大部分单子叶植物的花被都无分化^[7], 如禾本科植物的花被为很薄很小的浆片^[15], 而真双子叶植物的花被一般明显分化为花萼和花瓣, 且物种间花萼和花瓣的分化形式极为多样. 因此, 一般认为, 单子叶植物花被的分化程度较真双子叶植物要低^[7]. 目前, 单子叶植物中的 MADS-box 基因研究主要集中于禾本科, 本文所涉及的 SQUA 类基因也主要来自禾本科, 其花被分化程度在单子叶植物中相对较高, 因而有关的相对速率检验、置换位点统计分析和似然比检验结果为解释单子叶植物在花被分化程度上低于真双子叶植物的现象提供了新的进化生物学线索. 诚然, 随着所分离的 MADS-box 基因及相关类群的增加, 应当利用更多的直系同源基因来阐明植物花器官的进化规律.

(2) 单子叶植物构成单系类群的观点已被目前的植物分子系统学研究所证实^[16]. 对不同 MADS-box 基因的系统发育分析表明, 几乎每个亚家族中的直系同源基因都在单子叶、真双子叶两个分支间呈现出明显的分化(数据未显示). 然而, 迄今为止, 仅从 SQUA 类基因中检测出两个分支间在进化速率上存在着显著差异. 同时, 同义与非同义置换位点统计分析和似然比检验结果表明真双子叶分支中还可能存在着正选择压力, 这为进一步结合该类基因的功能结构域分析来研究非同义置换位点在 SQUA 基

因结构域上的分布规律提供了线索. 然而, 有关适应性进化的统计检测方法在实际应用中仍存在一些. 本文使用 Mega 和 PAML 计算所获得的结果也不尽一致, 说明相关方法和程序还需进一步改进.

参 考 文 献

- Weigel D, et al. The ABCs of floral homeotic genes. *Cell*, 1994, 78: 203
- Rounsley S D, et al. Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell*, 1995, 7(8): 1259
- Kang H G, et al. Characterization of two rice MADS box genes that control flowering time. *Mol Cells*, 1997, 7(4): 559
- Kyoizuka J, et al. Ectopic expression of OsMADS3, a rice ortholog of AGAMOUS, caused a homeotic transformation of lodicules to stamens in transgenic rice plants. *Plant Cell Physiol*, 2002, 43(1): 130
- Fischer A, et al. Chromosomal mapping of the MADS-box multi-gene family in *Zea mays* reveals dispersed distribution of allelic genes as well as transposed copies. *Nucleic Acids Res*, 1995, 23(11): 1901
- Cacharron J, et al. Expression of MADS box genes *ZMM8* and *ZMM14* during inflorescence development of *Zea mays* discriminates between the upper and the lower floret of each spikelet. *Dev Genes Evol*, 1999, 209(7): 411
- Taylor, T W, et al. eds. *Flowering Plants: Origin Evolution and Phylogeny*. New York: Chapman and Hall, 1996
- Theissen G, et al. A short history of MADS-box genes in plants. *Plant Mol Biol*, 2000, 42(1): 115
- Li P, et al. Relative-rate test for nucleotide substitutions between two lineages. *Mol Biol Evol*, 1992, (9): 1185
- Robinson M, et al. Sensitivity of the relative-rate test to taxonomic sampling. *Mol Biol Evol*, 1998, 15: 1091
- Robinson-Rechavi M, et al. RRTree: Relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics*, 2000, 16: 296
- Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*, 1997, 13: 555
- Kumar S, et al. MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics*, 2001, 17(12): 1244
- Yang J, et al. Duplication and adaptive evolution of the chalcone synthase genes of *Dendranthema* (Asteraceae). *Mol Biol Evol*, 2002, 19: 1752
- Clifford H T. Spikelet and floral morphology. In: *Grass Systematics and Evolution*. Washington DC: Smithsonian Institution Press, 1987
- APG II. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants; APG II. *Bot J Linn Soc*, 2003, 141: 399