林业科学研究 2006, 19(2): 165～169
Forest Research

文章编号: 1001-1498(2006)02-0165-05

竹类植物与水稻等其它禾本科作物的系统进化关系及基因序列组成的比较

樊龙江[1], 郭兴益[1], 马乃训[2]

(1. 浙江大学作物科学研究所/生物信息学研究所,浙江 杭州 310029;
2. 中国林业科学研究院亚热带林业研究所,浙江 富阳 311400)

摘要:利用 8个全长 mRNA基因序列比较了竹类植物和水稻等禾本科作物的系统进化和序列组成等。结果表明,在与水稻、玉米和麦类作物的比较中,竹类植物与水稻有着最近的亲缘关系和更相似的基因序列特征(GC含量分布和密码子使用频率),它们甚至比水稻与玉米或水稻与麦类之间的亲缘关系更近。这一结果提示,水稻作为禾本科植物的模式物种,应与竹类植物存在良好的基因组共线性关系,水稻基因组序列信息对竹类植物基因组研究与分析具有重要参考价值。

关键词:竹类植物;水稻;系统进化; GC含量;密码子使用频率
中图分类号: S795 S718.46 **文献标识码**: A

Comparative Study on Phylogenetics and Sequences Composition of Bamboos and Cereals

FAN Long-jiang[1], *GUO Xing-yi*[1], *MA Nai-xun*[2]

(1. Institute of Crop Science/Institute of Bioinformatics, Zhejiang University, Hangzhou 310029, Zhejiang, China;
2. Research Institute of Subtropical Forestry, CAF, Fuyang 311400, Zhejiang, China)

Abstract: Phylogenetics and Sequence composition of bamboos and cereals were analyzed based on current 8 full-length mRNA sequences in the public nucleotides databases. The results indicated that bamboos had more shorter phylogenetic distance and similar sequence composition (GC content and codon usage) with rice than other cereals. The results suggested that rice, as model plant for Poaceae, should have significant genomic synteny with bamboos, and its genomic sequences are important resource for bamboo genome research.

Key words: bamboo; rice; phylogenetics; GC content; codon usage

竹类植物是禾本科(Gramineae)植物中最原始的亚科(竹亚科 Bambusoideae)之一,也是禾本科植物中最具多样化的一个种群,它以木质的秆、复合分枝、发达的根系和很少开花的特点区别于同科的其它植物。禾本科的其他成员还包含水稻(*Oryza sativa* L.)、玉米(*Zea mays* L.)、大麦(*Hordeum vulgare* L.)、小麦(*Triticum aestivum* L.)等主要粮食作物。竹子在地球的纬度分布范围为 46°N～47°S,包括热带和亚热带的广大地区。其生长的海拔可高达 4 000 m,主要分布在喜马拉雅山区、中国的部分地区和南美洲安第山脉等地区。全世界共有竹类植物 107个属,1 300多个种,其中主要为木本竹类植物;

收稿日期: 2005-03-30
基金项目: 国家自然科学基金(30270810)
作者简介: 樊龙江(1965—),男,浙江衢州人,浙江大学作物科学研究所和生物信息学研究所教授,博导,目前主要从事禾本科作物基因组分析及其遗传改良。电话: 0571-86971730; E-mail: fanlj@zju.edu.cn

同时,我国竹类植物资源非常丰富,有许多特有珍贵竹种[1]。竹类植物的遗传和分子生物学研究近年来虽然有了长足发展,但还是相对比较薄弱[2],与禾本科中的水稻等主要作物等的研究有较大的差距。例如,国际公共核苷酸序列数据库[3](GenBank EMBL DDBJ)中,来自竹类植物的序列目前仅有 257条。

目前对竹类植物进行核苷酸测序主要是用于分子系统发育或进化(Phylogenetics)分析。这类分析中只有少数完全针对竹类植物进行系统发育分析,如中科院云南植物研究所的研究工作[4~6],其他大多数研究主要是针对整个禾本科或更上一层物种进行的[7~13]。竹类植物作为禾本科的成员,它们在系统分类上往往并不一致,不同来源(核、叶绿体和线粒体)的基因序列或同一来源但不同基因获得的系统分类会有所差异。例如,在他们构建的系统进化树中有关竹类与水稻等主要禾本科作物关系的结果:Zhang[9]基于叶绿体 rp116基因获得的系统进化(聚类)关系为竹类与水稻先聚合然后与玉米再聚在一起,简单表示为((竹 +稻) +玉米),Natodt等[8]同样基于叶绿体基因 rps4获得(((竹 +稻) +玉米) +麦),Gaut等[12]基于 adhI基因片段获得((竹 1 +稻) + (竹 2 +玉米) +麦),Mathews等[11]基于细胞色素 B基因片段获得((竹 1 + ((竹 2 +麦) +稻)) +玉米),Mason-Gamer等[7]基于 Wx基因片段获得(竹 + (稻 +麦)) +玉米)等。但一个总的趋势是在主要禾本科作物中,竹类植物与水稻的亲缘关系比较近。同时,最近开展的禾本科植物基因非编码保守序列(CNS, conserved noncoding sequence)分析表明,水稻和玉米等禾本科作物基因内含子序列中的保守区段,在竹类植物中同样也是保守的[14,15]。另外,少量竹类植物基因的全长 mRNA序列被测序,包括毛竹等[16,17]。这些序列是难得的开展竹类植物基因序列分析的基础数据。由于水稻等禾本科作物分子生物学研究较全面,这些竹类全长基因序列在水稻等基因组中的直系同源(orthologous)基因也往往被测序,这样就为基于这些基因序列开展分子进化和序列组成比较分析提供了绝好的机会。

本文通过搜索 GenBank数据库获得最新的竹类植物具有全长核苷酸序列的基因记录及其直系同源基因,利用其中 8个序列数据比较了竹类植物与主要禾本科大田作物(水稻、玉米、大麦、小麦)的系统进化关系以及它们的序列组成特征等。

1 材料与方法

1.1 材料

通过搜索 GenBank数据库获得最新的竹类植物具有全长 mRNA序列的基因记录及其水稻等主要禾本科作物直系同源基因全长 mRNA序列(表 1)。获得的 8条竹类基因序列用于以下系统进化和序列分析。

表 1 竹类植物具有全长 mRNA序列的核基因[3]及其水稻等其它禾本科作物同源 mRNA序列情况

物种	基因	数据库记录号	数量/条	同源序列			
				水稻	玉米	小麦	大麦
绿竹	suerose synthase	AF412037 - AF412039	3	NM_184941	L22296	AJ000153	X69931
绿竹	UDP - glucese pyrophosphorylase(UGP)	AY178448	1	AK119197	AY103595	BT009219	X91347
绿竹	phenylalanine ammonia - lyase(PAL1)	AY450643	1	XM_466846	AY104679		Z49147
绿竹	chitinase	AY453406	1	L40337	AY532766	X76041	L34211
绿竹	Sucrose - phosphate synthase	AY445835	1	XM_481429	AY109435	AF347064	
毛竹	ACC Synthase	AB085172	1	AK071011	AY359571	U42336	
毛竹	1 - aminocyclopropane - 1 - carboxylate oxidase	AB044747	1	AK104933	AY109855		
麻竹	MADS 1 - 18 Protein	AY395714 - AY395715 AY599750 - AY599765	18	AY332478	AJ430641	AY280870	AJ249144

1.2 方法

同义替换率估计:利用 Smith-Waterman算法(EMBOSS软件包下的 water程序)[18]联配翻译的蛋白质序列,根据该联配结果进行核苷酸序列联配。去除缺口(gap)后,利用 PAML(Phylogenetic Analysis by Maximum Likelihood)程序包的 codeml程序[19]和 F3 ×4模型[20]计算同义替换率。系统发育树利用 PHYLIP软件包 PROTDIST和 NEIGHBOR(邻接法)方法构建。

密码子使用频率:利用密码子使用频率数据库(Codon Usage Database) Countcodon 程序[21](version4)对 8条全长基因序列(其中大麦 7条,见表 1)进行频率计算。

2　结果与分析

2.1　GenBank数据库中有关分类植物核苷酸序列记录情况

GenBank数据库[3]是国际公共核苷酸序列数据库(GenBank/EMBL/DDBJ)之一,是世界最主要的分子生物学数据库。所有研究论文在公开发表前,均被要求将测定的相关核苷酸序列递交该数据库。所以从该数据库中的竹类植物的核苷酸序列递交情况,可以清楚地了解到目前竹类植物分子生物学,特别是基于核苷酸序列的研究现状。

截止 2005年 3月 12日, GenBank数据库中来自竹亚科的核苷酸序列总计 295条,其中 257条来自竹族。该数量是非常少的,与同科的大田作物(如水稻 150多万条记录)相差甚远。在竹族的 257条序列中,主要来自一些重要的属,如箭竹属(*Fargesia* Franch.)、刺竹属(*Bambusa* Schreb.)、青篱竹属(*Arundinaria* Michaux)、丘斯夸竹属(Chusquea)、牡竹属(*Dendrocalamus* Nees)和刚竹属(*Phyllostachys* Sieb & Zucc.)。这些序列中主要为细胞器(叶绿体和线粒体)基因和核糖体 RNA序列,它们主要在一些与竹类植物、禾本科或更宽泛的植物物种系统进化研究中被测序;同时,一些重要基因片段(如淀粉合成酶、细胞色素等)也在竹类植物内被测序用于系统进化分析。另外,毛竹(*Phyllostachys heterocycla* var. *pubescens* (Mazel) Ohwi)和绿竹(*Dendrocalamopsis oldhami* (Munro) Keng f.)一些全长 mRNA被日本和台湾学者测序(表 1);同时还有一些竹类转座子、微卫星 DNA序列等被提交。

2.2　竹类植物与水稻等其他禾本科作物的系统进化关系

利用竹类植物现有的 8条全长 mRNA序列进行的系统进化分析表明,竹类与水稻间的所有 8个基因同义替换率(Ks)均为最小,而竹类与玉米和竹类与麦类作物间的同义替换率则互有大小(数据未列出)。它们之间的平均同义替换率见表 2,自然地,竹类与水稻的平均值最小。根据平均同义替换率值和分子钟假设,竹类与水稻的物种分化时间约在 60万年前,而它们与玉米和麦类则在 75～85万年前后分开(表 2)。同时,从水稻与其他禾本科作物的平均同义替换率值来看,它们之间的亲缘关系均远于竹类与水稻。

同样地,利用这些基因序列均可以分别进行聚类分析,并构建竹类植物与水稻等禾本科作物的系统进化树。在所建成的进化树中,竹类与水稻均首先聚类在一起(例见图 1)。

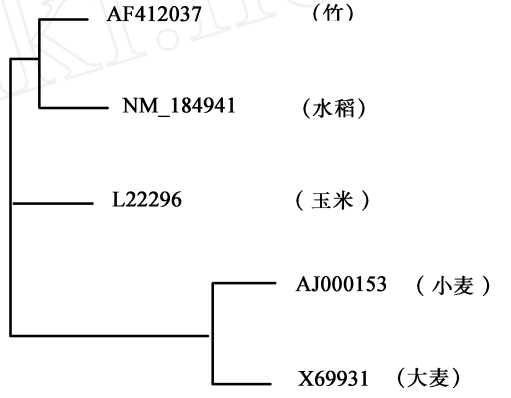

图 1　竹类植物与水稻等其他禾本科作物的系统进化关系
(该系统发育树是基于蔗糖合成酶基因构建,详见表 1)

表 2　竹类植物与其他禾本科作物基因同义替换率比较及其分化时间估计

项目	水稻	玉米	小麦	大麦
竹类				
同义替换率 *Ks*	0.383 ±0.150	0.499 ±0.139	0.557 ±0.227	O.528 ±0.160
分化时间/百万年前	58.9	76.7	85.7	81.2
水稻				
同义替换率 *Ks*	/	0.487 5 ±0.120	0.554 ±0.179	0.576 ±0.166
分化时间/百万年前	/	75.0	85.3	88.6

注:时间估计按每年每同义位点平均替换率 6.5 ×10^{-9}计算[22]。

2.3　竹类植物与水稻等其他禾本科作物基因序列组成

随着水稻基因组序列的测序,一些水稻,甚至是整个禾本科植物基因特有的现象被逐渐发现,例如水稻基因 GC含量沿转录方向呈负梯度分布[23]、基因组中 TIR～NBS～LRR类抗性基因的缺失[24]等。竹类植物基因看来同样具有基因 GC含量沿转录方向呈负梯度分布的特征(图 2)。图中可见,竹类植

物基因在 5 端编码启始位点附近约 1 000 bp区域,GC含量负梯度分布特征明显,曲线走势与水稻基因极为类似,相对地,与玉米和麦类基因的走势则有较明显的差异。

不同物种的基因编码对密码子都有所偏好。根据竹类植物现有的 8条具有全长 mRNA的基因序列可以大致看出其密码子使用偏好。结果表明,竹类植物使用频率最高的密码子为 GAG(0.54)、AAG(0.45)、GCC(0.35)、GGC(0.34)、CUC(0.33)和 GAC(0.33)(使用频率均大于 0.30),而最不常用的密码子为 CGA(0.02)、UUA(0.03)、UGU(0.03)、CUA(0.04)、GUA(0.05)等,这些趋势与禾本科植物总的密码子使用趋势是一致的。利用竹类植物这 8个基因的直系同源序列,同样地可以获得水稻等基因密码子使用频率。比较竹类与它们的密码子使用偏好可以看出,竹类植物更多地与水稻的密码子使用偏好相近,而与玉米和麦类作物之间的差异略大些(图 3)。

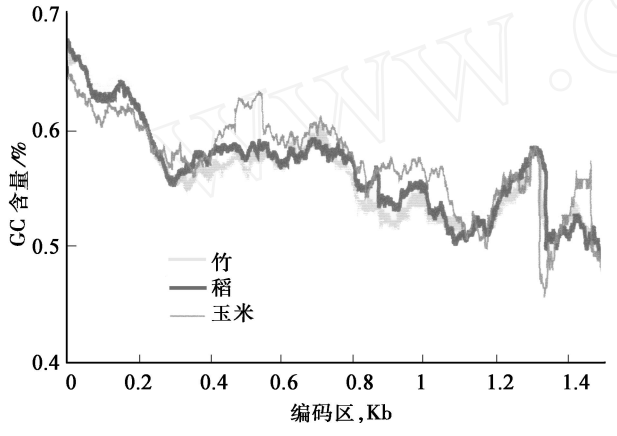

图 2 竹类植物、水稻和玉米基因编码区 GC含量变化趋势
(基因序列按编码启始位点对齐并以 129bp窗口平均)

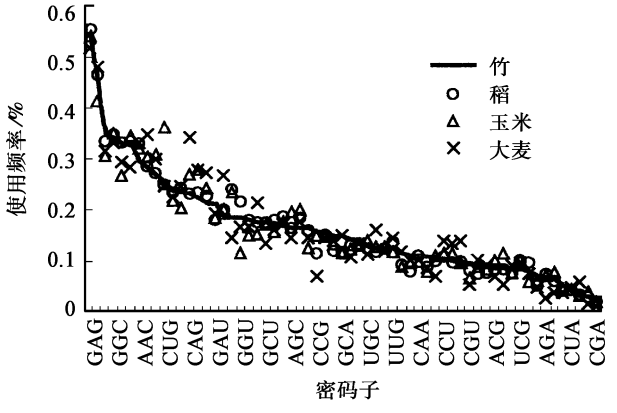

图 3 竹类植物及水稻等禾本科作物密码子使用频率
(横坐标包括了 61个非终止密码子,图中仅标出部分密码子)

3 讨论

本研究利用最新的基因全长序列比较了竹类植物和水稻等禾本科作物的系统进化和序列组成等,结果表明,在水稻、玉米和麦类作物中,竹类植物与水稻有最近的亲缘关系和更相似的基因序列特征,它们比水稻与玉米或水稻与麦类之间的亲缘关系更近。虽然水稻和竹类植物现被分属为不同的亚科内[3,25],但在早期的禾本科分类中,水稻曾被分在竹亚科之下,与竹类植物分属不同的族,而玉米和麦类则被分在其他亚科内[26]。同时,竹类植物有着与水稻一样的染色体基数($X=12$)[27]。这些结果使我们初步认为,水稻作为禾本科植物的模式物种,与竹类植物基因组应该存在明显的共线性,就象与玉米和麦类作物基因组一样,其基因组序列信息对竹类植物基因组研究与分析具有重要价值。大量研究表明,水稻与玉米等其他禾本科作物基因组间存在很好的共线性(synteny)[28,29]。

根据最新的水稻基因组分析结果,水稻等禾本科作物均经历过一次全基因组倍增(Whole-genome duplication),通过序列分析可以在水稻基因组中看到清晰的基因组倍增遗留下来的痕迹;同时水稻还经历了一次近代的染色体片段倍增(Segmental duplication)[30,31]。这些重大的遗传事件是否在竹类植物基因组中留下痕迹,以及这一事件对竹类植物可能引起的遗传效应等有待进一步研究和证实。

在以前的有关竹类系统进化分析中,均是基于核基因片段或细胞器基因[7~11]。本研究使用的竹类植物具有全长序列的基因尚未见其在系统进化和序列分析中被应用(其中 MADS基因是由中科院云南植物所近期提交,应用于竹类植物系统进化或分类分析,但尚未见文章报道)。这些全长基因提供了完整的序列和代表性样本数据,使本研究得以进行。本研究中由于麦类作物的部分同源基因全长 mRNA序列尚未被测序或测序完成,这可能会部分影响竹类与麦类作物的相对结果,但其影响的程度不会明显。

参考文献:

[1] 傅懋毅,杨校生. 我国竹类研究展望与竹林生境利用[J]. 竹类研究汇刊, 2003, 22(2): 1~8

[2] 马乃训. 中国竹类植物生物多样性及其开发利用[A],见:竺肇华. 中国热带地区竹藤发展[M]. 北京:中国林业出版社, 2001: 36~46

[3] Benson D A, Karsch-Mizrachi I, Lipman D J, et al. GenBank: update. Nucleic Acids Res. 2004, 32 (Database issue): D23-6. www [J/OL]. ncbi. nlm. nih. org

[4] Guo Z H, Chen Y Y, Li D Z, et al. Genetic Variation and Evolution of the Alpine Bamboos (Poaceae: Bambusoideae) using DNA Sequence Data[J]. Plant Res, 2001, 114: 315～322

[5] Guo Z H, Chen Y Y, Li D Z. Phylogenetic Studies on the *Thamnocalamus* Group and Its Allies (Gramineae: Bambusoideae) Based on ITS Sequence Data[J]. Mol Phylogenet Evol, 2002, 22 (1): 20～30

[6] Guo Z H, Li D Z. Phylogenetics of the *Thamnocalamus* group and its allies (Gramineae: Bambusoideae): inference from the seqdences of GBSSI gene and ITS spacer[J]. Mol Phylogenet Evol, 2004, 30: 1～12

[7] Mason-Gamer R J, Weil C F, Kellogg E A. Granule-bound starch Synthase: Structure, function, and phylogenetid utility[J]. Mol Biol Evol, 1998, 15 (12): 1658～1673

[8] Nadot S, Bittar G, Carter L, et al. A phylogenetic analysis of monocotyledons based on the chloroplast gene rps4, using parsimony and a new numerical phenetics method[J]. Mol Phylogenet Evol, 1995, 4 (3): 257～282

[9] Zhang W. Phylogeny of the grass family (Poaceae) from rp116 intron Sequence data[J]. Mol Phylogenet Evol, 2000, 15 (1): 135～146

[10] Clark L G, Zhang W, Wendel J F. A Phylogeny of the Grass Family (Poaceae) Based on ndhF Sequence Data[J]. Syst Bot, 1995, 20 (4): 436～460

[11] Mathews S, Tsai R C, Kellogg E A. Phylogenetic Structure in the grass family (Poaceae): evidene from the nuclear gene phytochrome B[J]. Am J Bot, 2000, 87 (1): 96～107

[12] Gaut B S, Peek A S, Morton B R, et al. Patterns of genetic diversiffication within the Adh gene family in the grasses (Poaceae)[J]. Mol Biol Evol, 1999, 16 (8): 1086～1097

[13] Hilu K W, Alice L A, Liang H. Phylogeny of Poaceae inferred from matK Sequences[J]. Ann Mo Bot Gard, 1999, 86 (4): 835～851

[14] Kaplinsky N J, Braun D M, Penterman J, et al. Utility and distribution of conserved noncoding sequences in the grasses[J]. Proc Natl Acad Sci USA, 2002, 99: 617～651

[15] Baner P, Lubkowitz M, Tyers R, et al. Regulation and a conserved intron sequence of liguleless3/4knox class-I homeobox genes in grasses[J]. Planta, 2004, 219: 359～368

[16] Matsui T, Hatase Y, Ohobayashi K. A wound-induced ACC oxidase gene of Moso Bamboo Shoot[J]. Pak J Biol Sci, 2001, 4: 228～232

[17] Matsui T, Yokozeki Y, Inoue H. A wound-induced ACC Synthase Gene of Moso Bamboo Shoot[J]. Asian J Plant Sci, 2003, 2: 205～211

[18] Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Softwar Suite [J/OL]. Trends in Genetics, 2000, 16 (6): 276～277. http://www.uk.embnet.org/Software/EMBOSS/

[19] Yany Z. Phylogenetic Analysis by Maximum Likelihood (PAML) Version 2[M]. London: University College, 1999

[20] Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences[J]. Mol Biol Evol, 1994, 11: 725～736

[21] Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucl Acids Res [J/OL], 2000, 28: 292. http://www.kazusa.or.jp/codon

[22] Gaut B S, Morton B R, Mccaig B M, et al. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL[J]. Proc Natl Acad Sci USA, 1996, 93: 10274～10279

[23] Wong G K, Wang J, Tao L, et al. Compositional gradents in Gramineae genes[J]. Genome Res, 2002, 12: 851～856

[24] Tian Y, Fan L, Thurau T, et al. The absene of TIR-type resistance gene analogues in sugar beet (*Beta vulgaris* L.) genome[J]. J Mol Evol, 2004, 58 (1): 40～53

[25] GPWG (Grass Phylogeny Working Group). Phylogeny and subfamilial classification of the grasses (Poaceae)[J]. Annals of Missouri Botanical Garden, 2001, 88: 373～457

[26] **查普曼 G P, 皮特 W E. 禾本科植物导论 (包括竹子及禾谷类作物)**[M]. **王彦荣译. 北京: 科学出版社**, 1996: 30～49

[27] Gielis J. Upstream Fundamental Research in Bamboo: Possibilities and Directions[A]. In: Proceedings of Vth International Bamboo Congres, Costa Rica, 1999

[28] Gale M P, Devos K M. Comparative genetics in the grasses[J]. Proc Natl Acad Sci USA, 1998, 95: 1971～1974

[29] The rice chromosome 10 sequencing consortium. In-depth view of structur, activity, and evolution of rice chromosome 10[J]. Science, 2003, 300: 1566～1569

[30] Paterson A H, Bowers J E, Chapman B A. Ancient polyploidization predating divergence of the cereals and its consequences for comparative genomics[J]. Proc Natl Acad Sci USA, 2004, 101: 9903～9908

[31] Zhang Y, Xu G, Guo X, et al. Two ancient rounds of polyploidy in rice genome[J]. J Zhejiang Univ (SCIENCE), 2005, 6B: 87～90