

利用多基因序列探讨稻属药稻复合体二倍体物种的系统发育关系

1,2包颖¹ 葛颂^{1*}

¹(中国科学院植物研究所系统与进化植物学重点实验室 北京 100093)

²(曲阜师范大学生物系 山东曲阜 273165)

Phylogenetic relationships among diploid species of *Oryza officinalis* complex revealed by multiple gene sequences

^{1,2}BAO Ying¹ GE Song^{1*}

¹(Laboratory of Systematic and Evolutionary Botany, Institute of Botany, the Chinese Academy of Sciences, Beijing 100093, China)

²(Qufu Normal University, Qufu 273165, Shandong, China)

Abstract *Oryza officinalis* complex includes five diploid species involving the B, C and E genomes, and is distributed in Asia, Africa and Oceania. These species are very important because *O. australiensis* is the only species with E genome, while *O. punctata* is the single species with B genome. Although there are three species with the C genome (*O. eichingeri*, *O. officinalis* and *O. rhizomatis*), these species might have played different roles in speciation of polyploids due to their disjunction in Asia and Africa. The phylogenetic relationships among these diploids are still unclear, and polyploidy speciation involving these diploids has long been in debate. In this study, we utilized the sequence data of multiple genes, including the chloroplast *matK*, internal transcribed spacer (ITS) fragment of nuclear ribosomal DNA, and two nuclear genes (*Adh1* and *Adh2*), to infer the phylogenetic relationships among the diploids and the related genomes. The separate and combined analyses were implemented by both maximum parsimony and maximum likelihood methods. The results indicate that (1) the C genome showed a closer affinity to the B genome than to the E genome; (2) of the three species with the C genome, *O. officinalis* and *O. rhizomatis* showed a closer relationship to each other than to *O. eichingeri*; (3) notably, in *O. eichingeri*, there exists higher divergence between the samples from Africa and those from Sri Lanka although they formed one monophyletic clade with lower bootstrap support.

Key words phylogenetic relationships, *Oryza officinalis* complex, diploid, multiple gene sequence.

摘要 稻属药稻复合体 *Oryza officinalis* complex 中有 5 个二倍体物种, 涉及 B、C 和 E 3 个染色体组, 分布在亚洲、非洲和大洋洲。其中, *O. australiensis* 是稻属中惟一含 E 染色体组的物种; *O. punctata* 则是惟一含 B 染色体组的物种。虽然 *O. officinalis*、*O. rhizomatis* 和 *O. eichingeri* 都含 C 染色体组, 但由于它们间断分布在亚洲和非洲, 可能在稻属异源多倍体成种中发挥了不同作用。染色体组 B、C 和 E 之间, 以及上述二倍体物种之间的系统发育关系及其在稻属多倍体物种形成中的作用都是稻属系统和进化研究中长期争论的问题。本文利用多基因片段, 包括叶绿体 *matK* 基因、nrDNA ITS 片段、低拷贝核基因 *Adh1* 和

2003-05-30 收稿, 2003-06-24 收修改稿。

基金项目: 国家杰出青年科学基金(30025005); 中国科学院“十五”重要方向项目(KSCXZ-SW-101A)资助 (Supported by the National Natural Science Foundation of China (Grant. No. 30025005) and the Chinese Academy of Sciences (KSCXZ-SW-101A))。

* 通讯作者 (Author for correspondence. E-mail: gesong@ns.ibcas.ac.cn)。

Adh2,探讨了这些二倍体种以及它们所代表的染色体组之间的系统发育关系。独立和合并的分析显示了一致的系统发育关系,即 C 染色体组和 B 染色体组的亲缘关系要比它们和 E 染色体组的更近。在 3 个含 C 染色体组的二倍体中,*O. officinalis* 和 *O. rhizomatis* 表现出较近的亲缘关系。值得注意的是,在 *O. eichingeri* 种内,尽管基于多基因的数据支持来自斯里兰卡的样品和来自非洲的样品为单系分支,但是较低的支持率表明这两个地区的样品间存在着较大的遗传分化,值得进一步利用形态、遗传和地理学证据从居群水平上进行深入的研究。

关键词 系统发育关系;药稻复合体;二倍体;多基因序列

药稻复合体(*Oryza officinalis* complex)是稻属中最大的复合体,包括 9 个种,含 5 种染色体组类型(B, C, BC, CD, E)(Vaughan, 1989, 1994),广泛分布于亚洲、非洲、大洋洲和拉丁美洲。在该复合体中,5 个二倍体种在稻属中具有非常重要的地位。其中,*O. australiensis* Domin. 是稻属中惟一含 E 染色体组的物种;*O. punctata* Kotechy ex Steud. 则是惟一含 B 染色体组的二倍体种。虽然 *O. officinalis* Wall. ex Watt.、*O. rhizomatis* Vaughan 和 *O. eichingeri* A. Peter 均含 C 染色体组,但它们间断分布在亚洲和非洲,在稻属异源多倍体 BBCC 和 CCDD 的物种形成中具有不同的作用。但是,这些二倍体物种之间的系统发育关系以及一些物种的分类地位一直存在争议(Bor, 1960; Sharma & Shastry, 1965; Tateoka, 1962; Vaughan, 1990)。例如,根据 Vaughan (1990) 的研究,斯里兰卡只有 2 个含 C 染色体组的二倍体种,即 *O. rhizomatis* 和 *O. eichingeri*,前者是斯里兰卡特有种,后者则间断分布于亚洲和非洲,在亚洲也只局限于斯里兰卡(Vaughan, 1994)。*O. eichingeri* 曾被定名为 *O. collina* (Sharma & Shastry, 1965),后 Vaughan(1990)将其并入 *O. eichingeri*,但对这些物种的分类仍缺乏统一的认识。Wang 等(1992)利用 RFLP 研究稻属物种的多态性和亲缘关系时,将编号为 103410 和 103421 的 2 份材料定名为 *O. collina*,而 Vaughan 1990 年发表新种 *O. rhizomatis* 时已将这 2 份材料处理为 *O. rhizomatis*。由于不同学者在以往的分类处理时基本依据形态特征和个人经验,因此迫切需要利用新的技术,特别是分子技术探讨这些二倍体的系统发育关系,为物种界定提供进一步的证据。同时,探讨药稻复合体二倍体种间的系统发育关系对于深入理解整个稻属的起源和分化以及物种形成方式具有重要意义。

叶绿体 DNA(cpDNA)和细胞核核糖体 DNA(nrDNA)是系统学研究中最常用的分子标记。尽管大多数叶绿体基因进化速率较慢,从而主要用于探讨较高级分类等级的系统发育关系,但一些进化稍快的叶绿体基因(如 *matK*)或基因间隔区(如 *trnL-F*)等,被有效地用于较低等级的系统发育研究(Johnson & Soltis, 1994; Sang et al., 1997; Bakker et al., 1999; Kajita et al., 1998; Ge et al., 1999; Wang et al., 1999)。由于 nrDNA 的转录间隔区(ITS)是经历了致同进化(concerted evolution)的多拷贝片段,进化速率快且易于 PCR 扩增和测序,从而被广泛地用于低等级的系统发育研究中,对揭示近缘属和属内种间的系统发育关系提供了重要的依据(Baldwin, 1992; Sang et al., 1997; Baldwin et al., 1995; Soltis & Soltis, 1998)。近年来,一些低拷贝核基因由于包含多个内含子和外显子,含有丰富的系统发育信息,被成功地应用于各个不同等级的系统发育重建,并显示了巨大的潜力(Soltis & Soltis, 1998; Ge et al., 1999; Sang, 2002)。但是,单基因片段提供的系统发育信息有限,而且采用单基因片段构建的系统树可能反映的仅仅是该片段的基因树,在许多情况下

基因树并不等同于物种树(Soltis et al., 1996; Wendel & Doyle, 1998; Sang, 2002)。因此, 当前分子系统发育重建的一个主要趋势是通过多个独立片段构建系统发育树并进行比较分析, 最后推断物种的系统发育关系(Hillis, 1995; Wendel & Doyle, 1998; Soltis & Soltis, 1998; Sang, 2002)。本研究选取了叶绿体 *matK* 基因、nrDNA 的 ITS 片段和低拷贝核基因乙醇脱氢酶基因 *Adh1* 和 *Adh2*。通过这些 DNA 片段的独立和合并分析, 探讨稻属药稻复合体中二倍体物种间的系统发育关系, 为该复合体以及稻属的分类提供新的资料, 同时为进一步阐明 B、C 和 E 染色体组之间的关系以及揭示该复合体中多倍体物种形成方式和机制奠定基础。

1 材料和方法

1.1 材料

本研究所用的材料代表药稻复合体的 5 个二倍体物种。每个二倍体种原则上选取 2 份样品。由于 *Oryza eichingeri* 间断分布于亚洲和非洲, 故选取了 4 份材料, 2 份来自非洲, 2 份来自斯里兰卡; 仅分布于澳大利亚的 *O. australiensis* 除在 ITS 片段分析时选用了 2 份样品, 其他基因片段均选用 1 份样品。另外, 1 份来自中国海南的 *O. granulata* Nees & Arn. ex Watt 样品被选作外类群, 因为该种被证明是稻属最基部的谱系(Wang et al., 1992; Ge et al., 1999)。上述 12 份样品的详细资料见表 1。

表 1 植物样品材料

Table 1 Plant materials used in the present study

样品 Sample ¹⁾	染色体组 Genome	IRRI 编号 IRRI ²⁾ accession	原产地 Origin	序列接收号 GenBank numbers			
				<i>matK</i>	ITS	<i>Adh1</i>	<i>Adh2</i>
<i>Oryza officinalis</i> -1	CC	105085	菲律宾 Philippines	AF148658	AF479067	AF148579	AF148613
<i>O. officinalis</i> -2	CC	106159	巴布亚新几内亚 Papua New Guinea	-	AF479063	AY169473	AY169483
<i>O. eichingeri</i> -s1	CC	81803	斯里兰卡 Sri Lanka	AY176644	AF479068	AY169475	AY169487
<i>O. eichingeri</i> -s2	CC	105407	斯里兰卡 Sri Lanka	-	AY181994	AY169476	AY169486
<i>O. eichingeri</i> -u1	CC	101422	乌干达 Uganda	AY318858	AF479069	AY169477	AY169485
<i>O. eichingeri</i> -u2	CC	105159	乌干达 Uganda	-	AY181995	AY169478	AY169488
<i>O. rhizomatis</i> -1	CC	105448	斯里兰卡 Sri Lanka	AF148660	AF479066	AF148580	AF148614
<i>O. rhizomatis</i> -2	CC	103410	斯里兰卡 Sri Lanka	-	AF479065	AY169474	AY169484
<i>O. punctata</i> -1	BB	104071	乍得 Chad	AF148661	AF479070	AF148577	AF148611
<i>O. punctata</i> -2	BB	105607	乍得 Chad	-	AF479071	-	-
<i>O. australiensis</i> -1	EE	105263	澳大利亚 Australia	AF148667	AF520777	AF148589	AF148623
<i>O. australiensis</i> -2	EE	101144	澳大利亚 Australia	-	AF520778	-	-
<i>O. granulata</i>	GG	940208	中国海南 Hainan, China	AF148674	AF520779	AF148597	AF148631

1) 种名后数字代表不同样品, s 代表斯里兰卡, u 代表乌干达。2) IRRI, 国际水稻研究所。

1) Numbers after species name indicate different accessions; s indicates Sri Lanka; u indicates Uganda. 2) IRRI, International Rice Research Institute.

除外类群为野外采集, 其余的材料均来自国际水稻研究所(IRRI)的种子资源库。野生稻种子首先在 50–55 °C 处理 5 d 以打破休眠, 然后将去壳的种子悬浮在灭菌液中充分冲洗, 在温水(30–35 °C)中萌发。生长 2 周的幼苗被移植至中国科学院植物研究所(北京)的温室中栽培。用 2 个月幼苗的叶片提取 DNA。取样植株均保留至成熟, 以观察其形

态特征。

1.2 总 DNA 的提取、PCR 扩增和 DNA 测序

从个体新鲜叶片中提取总 DNA 均使用改良的 CTAB 法(Doyle & Doyle, 1987)。聚合酶链式反应(PCR)在 PTC-200(PE)热循环仪上完成。扩增叶绿体 *matK* 基因的程序包括 94 °C 预变性 1.5 min; 35 个循环的 94 °C 变性 1.5 min, 55 °C 退火 1.5 min 和 72 °C 延伸 1.5 min; 最后 72 °C 延伸 5 min。ITS 的扩增反应中加入 8% 的二甲基亚砜 (DMSO) 以增加引物特异性 (Baldwin et al., 1995), 扩增程序包括 70 °C 预热 4 min; 2 个循环的 94 °C 变性 1 min, 52 °C 退火 20 s 和 72 °C 延伸 1.5 min; 接 35 个循环的 94 °C 变性 20 s, 52 °C 退火 20 s 和 72 °C 延伸 1.5 min; 最后 72 °C 延伸 10 min。 *Adh1* 或 *Adh2* 基因的扩增程序包括 70 °C 预热 4 min; 2 个循环的 94 °C 变性 1 min, 56 °C 退火 30 s 和 72 °C 延伸 1.5 min; 接 35 个循环的 94 °C 变性 20 s, 55 °C 退火 20 s 和 72 °C 延伸 1.5 min; 最后 72 °C 延伸 10 min。扩增及测序所用的引物见表 2。扩增产物利用 Pharmacia 纯化试剂盒纯化后在 ABI377 自动测序仪上进行测序。为保证序列的精确性,所有序列均对双链进行测序。

表 2 用于扩增和测序的引物序列

Table 2 The primer sequences for PCR amplification and sequencing

引物 Primer	序列 Sequence	参考文献 Reference
<i>matK</i>		Ge et al., 1999
<i>matKF</i>	5'-TAATTAAGAGGATTCACCAG-3'	
<i>matKF2</i>	5'-ATTGCCITTCCTTGATATCG-3'	
<i>matKR2</i>	5'-ACTACTCGAATTGGAATAG-3'	
<i>matKR1</i>	5'-ATGCAACACCCTGTTCTGAC-3'	
ITS		White et al., 1990
ITS1	5'-AGAAGTCGTAACAAGGTTCCGTAGG-3'	
ITS4	5'-TCCTCCGCTTATTGATATGC-3'	
<i>Adh1</i>		Ge et al., 1999
<i>AdhF1</i>	5'-CACACCGACGCTCTACTTCTG-3'	
<i>AdhF2</i>	5'-AGAGTGTTGGAGAGGGTGTGAC-3'	
<i>Adh1R2</i>	5'-ACTCACAGCAAGGCCTACAGC-3'	
<i>Adh1bR</i>	5'-TCAGCAAGTACCCTAAATTATC-3'	
<i>Adh2</i>		Ge et al., 1999
<i>AdhF1</i>	5'-CACACCGACGCTCTACTTCTG-3'	
<i>AdhF2</i>	5'-AGAGTGTTGGAGAGGGTGTGAC-3'	
<i>Adh2R2</i>	5'-ACAGCAAGGCCAACAGCTCC-3'	
<i>Adh2RR</i>	5'-CCACCGTTGCTCATCTCAAT-3'	

1.3 数据分析

序列通过 CLUSTAL W (Thompson et al., 1994) 进行对位排列(alignment)后手工校正; 系统发育分析采用 PAUP 4.0 版 (Swofford, 1998) 中的简约法和似然法进行。最大简约性 (MP) 分析选择分支限界(branch and bound search)法进行,对所有核苷酸位置的碱基性状同等加权 (Fitch, 1971)。对于由于序列长度多态性所造成的空位(gap),在数据统计中不参与运算,即处理为缺失(missing)状态。最大似然性 (ML) 分析依据 HKY 序列进化模式

(Hasegawa et al., 1985), 选择分支限界法进行 ML 搜寻 1000 次。碱基频率设定为数据估算(empirical), 转换和颠换比设定为 2。拓扑结构的可靠性用 1000 次重复的自展(bootstrap)检验来完成(Felsenstein, 1985)。为了检测来自不同 DNA 片段所得到的系统发育树是否存在冲突, 利用 PAUP 4.0 版程序中的 partition-homogeneity test (PHT) 对叶绿体 *matK* 基因、核 ITS 片段、*Adh1* 和 *Adh2* 基因系统树之间的拓扑一致性进行评价(Johnson & Soltis, 1998), 检测进行 1000 次重复。当来自不同 DNA 片段所得到的系统发育树间不存在显著冲突($P > 0.05$)时, 不同片段的数据将进一步合并进行系统发育分析。

2 结果

2.1 序列特点

扩增的叶绿体 *matK* 基因片段长 1487 – 1552 bp。对位排列后的长度为 1552 bp, 包括 34 个变异位点, 其中 15 个含系统发育信息。ITS 片段的长为 578 – 610 bp, 包括 ITS1(188 – 220 bp)、5.8S(164 bp)和 ITS2(228 – 230 bp)3 个区段, 对位排列后的长度为 668 bp, 含 56 个变异位点, 其中 33 个含系统发育信息。尽管 ITS1 的长度短于 ITS2, 但含系统发育信息的位点数(16 个)与 ITS2(16 个)相同, 5.8S 基因序列高度保守, 在 5 个二倍体物种中只发现 2 个变异位点, 其中 1 个含系统发育信息。扩增的 *Adh1* 和 *Adh2* 基因片段包括 5 个内含子和 5 个外显子。扩增的 *Adh1* 基因片段长 1770 – 1829 bp, 对位排列后的长度为 1940 bp, 包含 264 个变异位点, 其中 133 个含系统发育信息; *Adh2* 片段长 1645 – 1668 bp, 对位排列后的长度为 1752 bp, 包含 236 个变异位点, 其中 137 个含有系统发育信息。4 个片段中, 核基因 *Adh1* 和 *Adh2* 提供的系统发育信息最多(信息位点分别占 7% 和 6%); ITS 次之(5%); 叶绿体 *matK* 片段提供的信息最少(1%)。另外, ITS 片段的 GC 含量(72%)要明显高于其他片段(35% – 43%)(表 3)。

表 3 各 DNA 片段的序列参数
Table 3 Sequence characteristics for each DNA fragment

DNA 片段 DNA fragment	序列排列长度 Aligned length (bp)	变异位点 Variable site	信息位点 Informative site	信息位点(%) Informative site (%)	树长 Tree length	<i>CI</i>	<i>RI</i>	GC (%)
<i>matK</i>	1552	34	15	1	425	0.93	0.90	35
ITS	668	56	33	5	57	0.97	0.91	72
<i>Adh1</i>	1940	264	133	7	110	0.85	0.84	41
<i>Adh2</i>	1752	236	137	6	392	0.93	0.90	43
<i>matK</i> /ITS/ <i>Adh1</i>	4160	354	181	4	590	0.92	0.88	44

2.2 系统发育分析

对各个 DNA 片段分别采用最大简约法(MP)和最大似然法(ML)进行系统发育分析。4 个基因片段各自得到最简约树的数目分别为 1 个(*matK*)、2 个(ITS)、2 个(*Adh1*)和 3 个(*Adh2*), 其严格一致树的树长、一致性指数(*CI*)和保持性指数(*RI*)列于表 3。ML 得到的系统发育树在分支结构上与 MP 的结果一致, 仅在自展支持率上略有不同, 故在此只展示 ML 的结果。图 1 – 3 为根据 ML 分别得到的 4 个 DNA 片段的单基因系统发育树。由图可见, 4 个 ML 树都是按照染色体组 C、B 和 E 分成 3 个主要分支, 含 C 染色体组物种构成的

分支和含 B 染色体组物种构成的分支形成姐妹群关系,然后再与含 E 染色体组物种的分支构成姐妹群。这 4 个基因树的主要不同在于 C 染色体组分支内部。在 *matK* 基因树上, *O. officinalis* 的不同样品形成一单独小分支(支持率 81%),其余两个二倍体以多歧分支(polytomy)的形式出现(图 1)。然而在 ITS 的 ML 树上, *O. officinalis* 和 *O. rhizomatis* 都各自构成单独的小分支,而 *O. eichingeri* 的斯里兰卡样品和乌干达样品构成了另外两个小分支,这两个小分支构成的单系未得到显著的支持 (< 50%)(图 2)。在 *Adh1* 的基因树上, *O. officinalis* 表现出和 *O. rhizomatis* 更近的关系(支持率 81%),但 *O. eichingeri* 的 2 个地区样品也没有构成单系(图 3); *Adh2* 的 ML 树的基本结构和 *Adh1* 的类似,主要冲突在于 *O. rhizomatis* 的两个样品没有构成独立的小分支,其中一个样品先和 *O. officinalis* 小分支聚合,再和另一个样品形成姐妹群(图 3)。另外,这 4 个基因树在分辨率上也存在不同,其中 *matK* 基因树的分辨率最小,ITS 的次之, *Adh* 基因的最大,体现了各 DNA 片段所含信息量的差异。

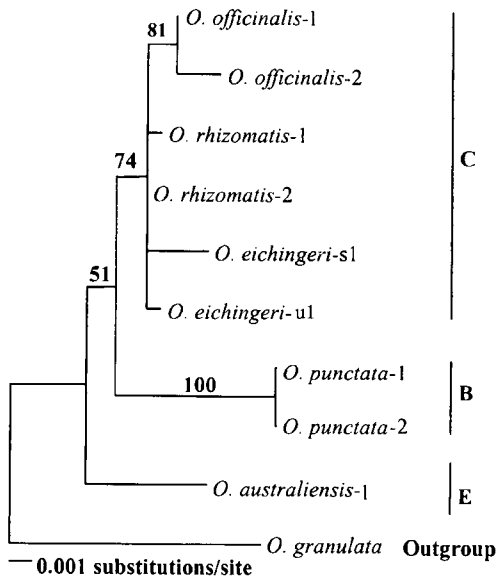


图 1 叶绿体 *matK* 基因的最大似然树(种名后的数字代表不同样品, s 代表斯里兰卡, u 代表乌干达;分支上数字为大于 50% 的自展支持率,字母代表染色体组类型)

Fig. 1. The maximum likelihood tree based on *matK*. Numbers after species name indicate different accessions; s indicates Sri Lanka; u indicates Uganda; numbers above branches indicate bootstrap values above 50%; letters in capital represent the genome types.

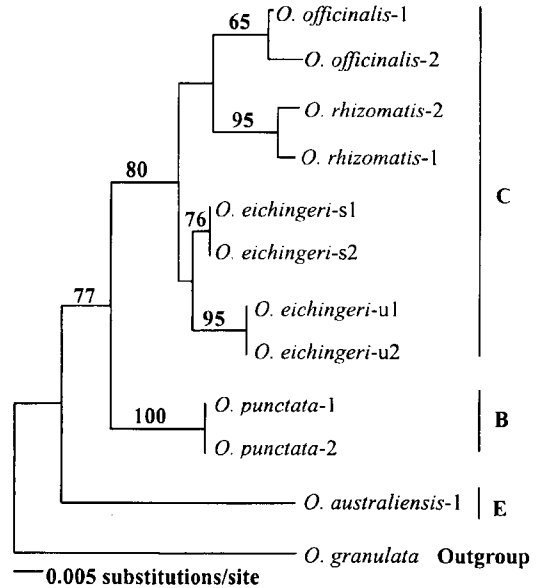


图 2 ITS 的最大似然树(种名后的数字代表不同样品, s 代表斯里兰卡, u 代表乌干达;分支上的数字为大于 50% 的自展支持率,字母代表染色体组类型)

Fig. 2. The maximum likelihood tree based on ITS. Numbers after species name indicate different accessions; s indicates Sri Lanka; u indicates Uganda; numbers above branches indicate bootstrap values above 50%; letters in capital represent the genome types.

2.3 不同 DNA 片段的合并分析

多基因数据的合并分析不但可以增加信息量,而且能更可靠地代表基因组。但在进行合并分析前,必须检验不同基因树间是否存在冲突。通过 PHT(partition-homogeneity test) (Johnson & Soltis, 1998),对 4 组数据的不同组合进行了评估(表 4)。结果表明, *Adh2* 和

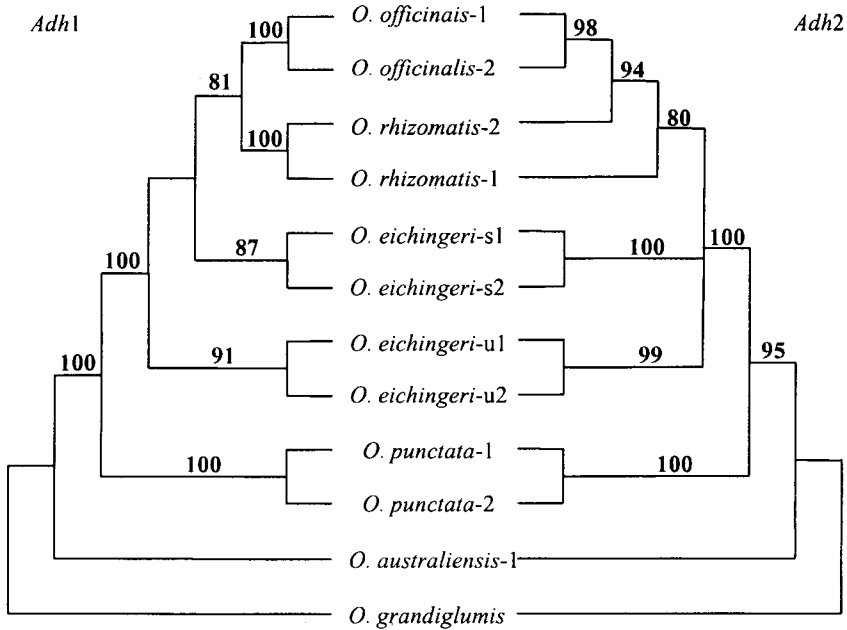


图 3 核 *Adh* 基因的最大似然树 左: *Adh1* 树; 右: *Adh2* 树。种名后的数字代表不同样品, s 代表斯里兰卡, u 代表乌干达; 分支上的数字为大于 50% 的自展支持率。

Fig. 3. The maximum likelihood tree based on *Adh* genes. Left, *Adh1*; right, *Adh2*. Numbers after species name indicate different accessions; s indicates Sri Lanka; u indicates Uganda; numbers above branches indicate bootstrap values above 50%.

表 4 不同数据组合的 PHT 检验结果

Table 4 Results conducted by partition-homogeneity test for various combinations of the data sets

数据矩阵 Data sets	P
matK vs. ITS	1.00
matK vs. <i>Adh1</i>	1.00
matK vs. <i>Adh2</i>	1.00
ITS vs. <i>Adh1</i>	1.00
ITS vs. <i>Adh2</i>	0.24
<i>Adh1</i> vs. <i>Adh2</i>	0.03
matK vs. ITS vs. <i>Adh1</i>	1.00
matK vs. ITS vs. <i>Adh2</i>	0.58
matK vs. <i>Adh1</i> vs. <i>Adh2</i>	0.11
ITS vs. <i>Adh1</i> vs. <i>Adh2</i>	0.15
matK vs. ITS vs. <i>Adh1</i> vs. <i>Adh2</i>	0.25

Adh1、*matK* 和 ITS 这 3 个 DNA 片段。

基于合并数据得到惟一 MP 树(图 4)($CI = 0.92, RI = 0.88$), 其基本的拓扑结构和 *Adh1* 单基因树类似, 但是各分支的分辨率都普遍提高, 3 个主要分支的支持率均达到 100%; *O. officinalis*-*O. rhizomatis* 分支支持率为 96%, 而且 *O. eichingeri* 的全部样品也以 56% 的支持率聚合成单独的分支(图 4)。合并数据的 ML 树(图 4)同样显示了 3 个大的分支, 分别代表了不同的染色体组类型, 含 C 染色体组的全部二倍体物种首先构成一个独

Adh1 的基因树间存在显著的冲突 ($P = 0.03$), *Adh2* 和其他片段的基因树有一些差异, 但冲突不显著 ($P = 0.11 - 0.15$), *Adh1*、*matK* 和 ITS 基因树之间不存在任何冲突 ($P = 1.00$)。由前面的分析及图 2、3 可见, *Adh2* 和其他片段的冲突主要在 *O. rhizomatis* 的系统位置, 为此我们对 *O. rhizomatis* 的样品进行了重复扩增和测序, 系统发育分析的结果仍然不变, 说明这一冲突并非来自取样和实验误差, 而很可能是 *O. rhizomatis* 的 *Adh2* 基因片段经历了某些特殊的进化事件。因此, 在进一步合并 DNA 片段数据进行系统发育分析时, 我们只利用

立的分支(支持率 100%), 然后与含 B 染色体组的 *O. punctata* 构成姐妹群(支持率 100%), 含 E 染色体组的 *O. australiensis* 则与 C、B 染色体组构成的分支形成姐妹群。在 C 染色体组分支内部, *O. officinalis* 和 *O. rhizomatis* 先组合在一起(支持率 96%), 再和 *O. eichingeri* 小分支构成姐妹群。尽管 *O. eichingeri* 的斯里兰卡样品和乌干达样品在合并的基因树上构成了单系小分支, 但支持率较低(65%)。

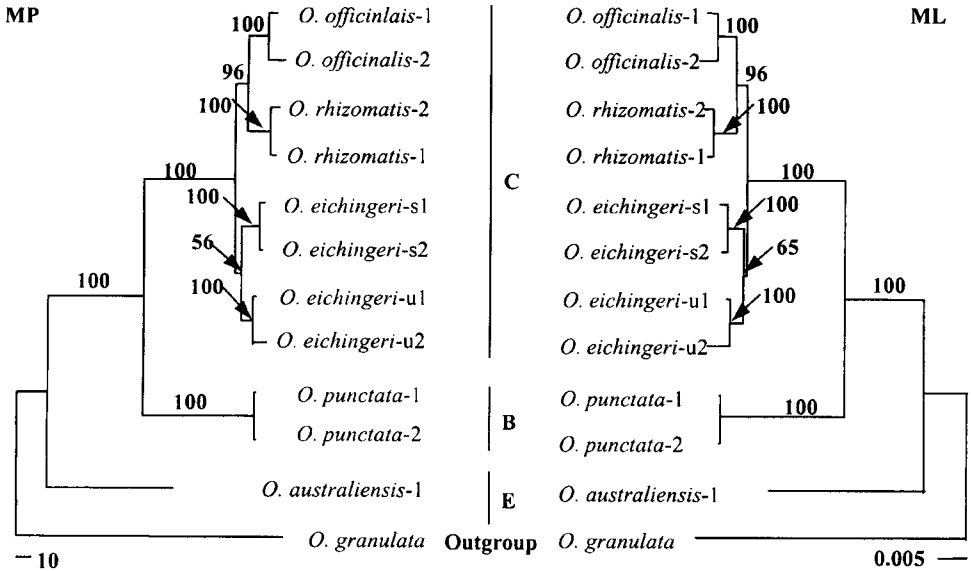


图 4 叶绿体 *matK* 基因、nrDNA ITS 片段和核基因 *Adh1* 数据合并后的基因树 左: 惟一最简约树(树长 = 590, $CI = 0.92$, $RI = 0.88$); 右: 最大似然树。种名后的数字代表不同样品, s 代表斯里兰卡, u 代表乌干达; 分支旁边数字为大于 50% 的自展支持率, 字母代表染色体组类型。

Fig. 4. The gene tree based on combined *matK* gene, ITS region and *Adh1* gene data sets. Left: the single most parsimonious tree (tree length = 590, $CI = 0.92$, $RI = 0.88$); right: the maximum likelihood tree. Numbers after species name indicate different accessions; s indicates Sri Lanka; u indicates Uganda; numbers on the branch indicate the bootstrap value above 50%; the letters C, B, and E represent genome types.

3 讨论

3.1 染色体组 B、C 和 E 的系统发育关系

在稻属中, 染色体组 B、C 和 E 的系统发育关系一直存在争议。Nayar (1973) 在总结稻属染色体组间关系时指出, A 染色体组和 C 染色体组是公认的基本染色体组, B、C 和 D 染色体组彼此间有一定的关系, E、F 染色体组和其他染色体组的关系不清楚。Wang 等 (1992) 采用核 DNA 限制性片段长度多态性 (RFLP) 方法研究了稻属 21 个种的亲缘关系, 在单独探讨二倍体亲缘关系时, 他们认为 C 染色体组和 E 染色体组的关系要近于和 B 染色体组的关系, 但是在他们对全部 21 个物种构建的聚类图上, E 染色体组却明显位于 C 染色体组和 B 染色体组共同的分支之外, 显示 E 染色体组和另两个染色体组间较远的关系。Li 等 (2000) 凭借基因组原位杂交技术 (GISH), 将 *O. punctata* (BB) 和 *O. australiensis* (EE) 的总 DNA 作为探针, 和 *O. minuta* (BBCC) 的染色体组分别进行杂交, 结果显示: 在前一杂交组合中, B 染色体组杂交信号明显, C 染色体组信号微弱; 而在后一杂交组合中, B、C 染色体组上得到了相同的杂交信号, 从而推论 B 和 C 染色体组的分化非常大, 而 E 和

B、C染色体组的关系都很近,C和E染色体组的关系要近于C和B染色体组的关系。但正如Belyayev等(2000)指出的,GISH方法在研究异源多倍体的系统发育关系上存在技术上的局限,它更多的是反映重复序列相互作用的一般技术,而不适用于探讨系统发育关系。在我们目前的研究中,3组DNA片段的独立和合并分析均显示,C染色体组是首先和B染色体组以100%的支持率形成单系,而E染色体组则与含B-C染色体组的分支形成姐妹群,表明B和C染色体组的关系要近于它们和E染色体组的关系(图1-4)。这个结果和Aggarwal等(1999)以及Joshi等(2000)分别利用AFLP和ISSR方法对稻属系统发育关系研究得出的结果一致。

3.2 含C染色体组二倍体物种间的系统发育关系

含C染色体组的二倍体物种包括分布在亚洲的*O. officinalis*和*O. rhizomatis*以及间断分布在亚洲和非洲的*O. eichingeri*。对于这3个物种的界定基本上没有疑义。*O. officinalis*和*O. rhizomatis*在形态上有明显区别,*O. officinalis*小穗较小,内稃顶端稍短,圆锥花序基部节点分枝较多,等长,轮生,小穗着生位置远离基部的初级分枝(Vaughan, 1990)。*O. rhizomatis*则小穗较大,内稃顶端伸长,小穗着生位置常靠近圆锥花序的主轴,在圆锥花序基部节点处缺少轮生分枝等(Vaughan, 1990)。另外由于杂种不育,这两个种在遗传上也存在一定差异(Sharma & Shastry, 1965)。*O. eichingeri*和以上2个种的主要不同在于没有根状茎(Vaughan, 1989, 1994)。本研究中,多基因片段的独立和合并分析均显示*O. officinalis*和*O. rhizomatis*系统发育关系较近,而它们与*O. eichingeri*的关系则稍远(图1-4)。

对于*O. eichingeri*种内分化问题,特别是斯里兰卡样品的分类地位一直不是非常清楚。以往的研究多集中在形态观察和杂交实验以及由此进行的推断。例如,Sharma和Shastry(1965)根据这个类群没有根状茎,秆较细,叶片窄和小穗基部不着色等性状,将其处理为一个新种*O. collina*;Gopalakrishnan和Sampath(1966)则认为这些形态特点在*O. officinalis*变异范围之内;通过观察*O. collina*和*O. officinalis*(印度)以及*O. eichingeri*(乌干达)分别杂交时减数分裂中期的染色体配对行为和杂种的育性,Sampath和Subramanyam(1968)认为*O. collina*是*O. officinalis*和*O. eichingeri*的中间过渡类型,但倾向将其作为*O. eichingeri*的亚种。Vaughan(1990)在全面研究斯里兰卡的野生稻时指出,分类学家同意*O. eichingeri*是斯里兰卡的药稻复合体植物的惟一代表,主要是受Bor(1960)和Tateoka(1962)的影响。但Nayar(1973)指出,Bor(1960)将*O. collina*包括在*O. eichingeri*之中,可能误认为*O. collina*也是四倍体,因为当时他认为*O. eichingeri*是四倍体。而Tateoka(1962)可能相信了Bor(1960)的判断,因为他没有给出任何理由就认为药稻复合体中所有Ceylon(即现在的斯里兰卡)的材料都和*O. eichingeri*相同。同时,Nayar(1973)也提到,Tateoka(1962)把小穗的尺寸作为区别*O. minuta*、*O. officinalis*、*O. eichingeri*和*O. punctata*的最重要性状,那么按此标准,*O. collina*应该包括在Tateoka界定的*O. punctata*之内。本研究结果表明,尽管斯里兰卡和非洲的*O. eichingeri*构成单系分支,显示彼此较近的系统发育关系,但它们在合并的MP和ML树上的支持率(分别为56%和65%)均显著低于*O. officinalis*和*O. rhizomatis*两个亲缘种间的支持率(96%)(图4),说明这两个地区间的*O. eichingeri*发生了较大的遗传分化。Nayar(1973)推测,

O. collina 可能和 *O. officinalis* 关系较近, 它的形态变异可能是由于周缘的隔离造成的, 而 *O. collina* 也可能是亚洲和非洲药稻复合体植物的过渡类群或纽带。但是, 对于这一问题的解决, 还有待进一步对斯里兰卡的稻属植物进行大量包括形态和地理在内的细致考察, 从居群的角度探讨两个大洲间类群的遗传关系, 从而最终进行可靠的分类处理和系统发育关系推断。

3.3 系统发育研究中数据的合并及其意义

虽然分子系统学研究正在飞速发展, 但是重建生命系统发育树仍然是具有挑战性的难题。一个主要的困难在于基因的系统发育关系和生物类群间的系统发育关系并不等同 (Wendel & Doyle, 1998; Sang, 2002)。因此, 利用单基因构建的系统发育树, 更多的是反映单基因的进化历史而不一定代表真正的物种系统发育关系。为此, 当前分子系统发育重建的趋势是通过多个独立基因 (DNA) 片段来构建多基因系统发育树并进行比较分析 (Hillis, 1995; Wendel & Doyle, 1998; Ge et al., 1999)。因为多个相互独立 DNA 片段构建的基因树更能反映物种真正的系统发育关系 (Miyamoto & Fitch, 1995; Sang, 2002)。而且, 将来源不同的 DNA 片段的序列数据进行合并处理可以有效地增加系统发育信息和减少噪音, 提高系统发育重建的准确性 (Donoghue & Sanderson, 1992; Olmstead & Sweere, 1994; Miyamoto & Fitch, 1995)。Olmstead 和 Sweere (1994) 利用来自茄科 17 个种的 2 个叶绿体的基因片段 *ndhF* 和 *rbcL* 序列以及叶绿体全基因组限制性位点基因图谱的资料, 展示了在系统发育研究中合并数据较单个数据具有更强的分辨率和解释系统发育关系的能力。在本研究中, *matK*、ITS 和 *Adh1* 各自单独的基因树虽然不存在任何冲突, 但是在分辨率和支持率上明显不同。*Adh1* 的基因树分辨率最高, 各分支平均支持率为 94.9%; ITS 分辨率次之, 各分支的平均支持率为 84%, *matK* 分辨率最小, 各分支平均支持率为 76.5% (图 1-3)。当将这 3 组数据合并后, 不但分辨率较各个单基因树要明显提高, 而且各分支的平均支持率达到 95.7% (图 4)。通过分析 *Adh1* 树和 *Adh2* 树之间的冲突, 可以发现冲突起因于 *O. rhizomatis*。PHT 分析还表明, 如果将全部数据矩阵进行合并将会掩盖由 *Adh2* 数据造成的不一致。例如, 当分别加入 *matK* 或 ITS 数据时, P 值达到 0.11 或 0.15, 均大于 0.05, 而把 4 组数据全部合并后得到的 P 值 (0.25) 也表现出明显的增大的趋势, 说明 *Adh2* 数据中的噪音在合并的数据矩阵中被真实的系统发育信息所掩盖 (表 4); 但是, 如果将 4 组数据强行合并, 得到的系统发育树和以上 3 个非冲突数据合并树比较, 具有冲突的分支的支持率明显下降。因此, 数据合并之前进行必要的一致性检验, 是发现基因树间冲突并决定数据如何合并的关键。在此基础上的合并树, 不仅提高了分辨率, 而且排除了个别特殊基因 (DNA) 片段的干扰, 从而更能反映真正的系统发育关系。

参 考 文 献

- Aggarwal R K, Brar D S, Nandi S, Huang N, Khush G S. 1999. Phylogenetic relationships among *Oryza* species revealed by AFLP markers. *Theoretical and Applied Genetics* 98: 1320-1328.
- Bakker F T, Culham A, Daugherty L C, Gibby M. 1999. A *trnL-F* based phylogeny for species of *Pelargonium* (Geraniaceae) with small chromosomes. *Plant Systematics and Evolution* 216: 309-324.
- Baldwin B G. 1992. Phylogenetic utility of the internal transcribed spacer of the nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution* 1: 3-16.

- Baldwin B G, Sanderson M J, Wojciechowski M F, Campbell C S, Donoghue M J. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of Missouri Botanical Garden* 82: 247 - 277.
- Belyayev A, Raskina O, Korol A, Nevo E. 2000. Coevolution of A and B genomes in allotetraploid *Triticum dicoccoides*. *Genome* 43: 1021 - 1026.
- Bor N L. 1960. The Grasses of Burma, Ceylon, India and Pakistan. Oxford: Pergamon Press.
- Donoghue M J, Sanderson M J. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. In: Soltis P S, Soltis D E, Doyle J J eds. *Molecular Systematics of Plants*. London: Chapman and Hall. 340 - 368.
- Doyle J J, Doyle J L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11 - 15.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783 - 791.
- Fitch W M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Zoology* 20: 406 - 416.
- Ge S, Sang T, Lu B R, Hong D Y. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proceedings of the National Academy of Sciences, USA* 96: 14400 - 14405.
- Gopalakrishnan R, Sampath S. 1966. The American species of *Oryza*. *Oryza* 3 (1): 35 - 40.
- Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22: 160 - 174.
- Hillis D M. 1995. Approaches for accessing phylogenetic accuracy. *Systematic Biology* 44: 3 - 16.
- Johnson L A, Soltis D E. 1994. *MatK* DNA-sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Systematic Botany* 19: 143 - 156.
- Johnson L A, Soltis D E. 1998. Assessing congruence: Empirical examples from molecular data. In: Soltis D E, Soltis P S, Doyle J J eds. *Molecular Systematics of Plants*. Boston: Kluwer Academic Publications. 1 - 42.
- Joshi S P, Gupta V S, Aggarwal R K, Ranjekar P K, Brar D S. 2000. Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. *Theoretical and Applied Genetics* 100: 1311 - 1320.
- Kajita T, Kamiya K, Nakamura K, Tachida H, Wickneswari R, Tsumura Y, Yoshimaru H, Yamazaki T. 1998. Molecular phylogeny of Dipterocarpaceae in Southeast Asia based on nucleotide sequences of *matK*, *trnL* intron, and *trnL-trnF* intergenic spacer region in chloroplast DNA. *Molecular Phylogenetics and Evolution* 10: 202 - 209.
- Li C-B (李常宝), Zhang D-M (张大明), Ge S (葛颂), Lu B-R (卢宝荣), Hong D-Y (洪德元). 2000. Detection of differentiation among BB, CC and EE genomes in the genus *Oryza* by two-probe genomic *in situ* hybridization (GISH). *Acta Botanica Sinica (植物学报)* 42: 988 - 990.
- Miyamoto M M, Fitch M W. 1995. Testing species phylogenies and phlogenetic methods with congruence. *Systematic Biology* 44: 64 - 76.
- Nayar N M. 1973. Origin and cytogenetics of rice. In: Caspari E W ed. *Advances in Genetics*. Vol. 17. New York, London: Academic Press.
- Olmstead R G, Sweere J A. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* 43: 467 - 481.
- Sampath S, Subramanyam M D. 1968. The taxonomy of a wild *Oryza* species from Ceylon. *Oryza* 5 (1): 75 - 76.
- Sang T, Crawford D J, Stuessy T F. 1997. Chloroplast phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120 - 1136.
- Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121 - 147.
- Sharma S D, Shastry S V S. 1965. Taxonomic studies in the genus *Oryza* L. IV. The Ceylonese *Oryza* spp. aff. *O. officinalis* Wall. ex Watt. *India Journal of Genetics and Plant Breeding* 25: 168 - 172.
- Soltis D E, Johnson L A, Looney C. 1996. Discordance between ITS and chloroplast topologies in the *Boykinia* group (Saxifragaceae). *Systematic Botany* 21: 169 - 185.

- Soltis D E, Soltis P S. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis D E, Soltis P S, Doyle J J eds. *Molecular Systematics of Plants. II. DNA Sequencing*. Boston: Kluwer Academic Publications.
- Swofford D L. 1998. PAUP* 4.0: Phylogenetic Analysis Using Parsimony (*, and other methods). Beta version 4.0b4A. Sunderland: Sinauer Associates.
- Tateoka T. 1962. Taxonomic studies of *Oryza* I. *O. latifolia* complex. *Botanical Magazine (Tokyo)* 75: 418 – 427.
- Thompson J D, Higgins D G, Gibson T J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673 – 4680.
- Wang X R, Tsumura Y, Yoshimaru H, Nagasaka K, Szmidt A E. 1999. Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpL20-rps18* spacer, and *trnV* intron sequences. *American Journal of Botany* 86: 1742 – 1753.
- Wang Z Y, Second G, Tanksley S D. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theoretical and Applied Genetics* 83: 565 – 581.
- Wendel J F, Doyle J J. 1998. Phylogenetic incongruence: window into genomes history and molecular evolution. In: Soltis D E, Soltis P S, Doyle J J eds. *Molecular Systematics of Plants. II. DNA Sequencing*. Boston: Kluwer Academic Publications.
- White T J, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M, Gelfand D, Sninsky J, White T eds. *PCR Protocols: A Guide to Methods and Applications*. San Diego: Academic Press. 315 – 322.
- Vaughan D A. 1989. The genus *Oryza* L. Current Status of Taxonomy. Philippines: International Rice Research Institute.
- Vaughan D A. 1990. A new rhizomatous *Oryza* species (Poaceae) from Sri Lanka. *Botanical Journal of the Linnean Society* 103: 159 – 163.
- Vaughan D A. 1994. *The Wild Relatives of Rice: A Genetic Resources Handbook*. Philippines: International Rice Research Institute.