

# 居群遗传结构研究中显性标记 数据分析方法初探

钱 韦, 葛 颂

(中国科学院植物研究所系统与进化植物学开放研究实验室, 北京 100093)

**摘要:** 为对比显性标记应用于居群遗传结构研究时不同统计参数的适用性, 利用 RAPD 技术对中国 5 个居群的 100 个疣粒野生稻个体进行了遗传结构分析。在衡量居群遗传多样性水平时, 多态位点比率 (PPB) 会低估遗传变异的量, 其价值不如 Shannon 多样性指数和 Nei 基因多样性指数, 而采用 Nei 指数时不必进行 Lynch-Milligan 矫正。对个体间遗传关系进行分析时, 17 种遗传相似性指数矩阵两两之间的 Mantel 检测都表现出极显著的相关性 ( $r > 0.95, r > r_{0.01}$ ), 且 UPGMA 聚类图的模式基本相似。基于  $\Phi$  统计量 ( $\Phi_{st}$ ) 的分子方差分析 (AMOVA) 和基于 Hardy-Weinberg 平衡假设的 Nei 遗传距离分析结果具有显著相关性, 它们都适合于遗传结构分析, 且应用后者时应通过 Lynch-Milligan 矫正减少显性遗传对变异估计偏低的影响, 增加检测居群间遗传分化的能力。此外, AMOVA 分析、 $G_{st}$  分析和 Shannon 多样性指数巢式分析都得出一致的结果, 表明疣粒野生稻的遗传变异绝大部分存在于地区之间, 而地区内和居群内的多样性程度较低。

**关键词:** 显性标记; 居群遗传结构; 疣粒野生稻; RAPD; Hardy-Weinberg 平衡

**中图分类号:** Q346.1 **文献标识码:** A **文章编号:** 0379-4172(2001)03-0244-12

由于无需鲜活实验材料, 可以直接检测 DNA 水平的变异且对基因组的取样更具代表性, DNA 分子标记正在逐步替代居群遗传学和保护遗传学中常用的等位酶分析方法, 成为衡量居群遗传变异水平的主流技术<sup>[1-3]</sup>。然而, 这些标记中的大部分 (如 RAPD、DAF、AFLP 和 ISSR) 都表现为显性遗传方式, 不能区分显性纯合子与杂合子, 从而无法直接估计居群内的基因频率, 导致不能利用各种成熟的、基于多位点、共显性遗传方式的居群遗传学统计参数<sup>[4]</sup>。虽然从实验方法上可以解决上述问题, 如对 RAPD 产物片段进行限制性酶切或利用标记的 RAPD 探针进行 Southern 杂交等确定扩增条带的基因型。但是, 对基于一定样本数量的居群遗传学和保护遗传学研究而言, 这意味着实验工作量会增加到不现实的程度而在研究中极少应用<sup>[5, 6]</sup>。在此情况下, 简单地认为显性标记不适用于居群遗传学和保护遗传学研究, 或者不加任何处理就利用标准  $F_{st}$  统计来进行居群遗传结构分析

收稿日期: 2000-05-14; 修订日期: 2000-08-07

基金项目: 中国科学院“九·五”重大项目 (No. KZ-951-B1-102) 资助

作者简介: 钱 韦 (1973), 男, 云南省昆明市人, 博士生, 目前的研究方向为野生稻的居群遗传结构及其形成与维持机制

都是不适宜的。事实上,如果利用适宜的数学分析方法,可保证遗传变异水平的估计偏差降低到最小程度,获得极有价值的分析资料<sup>[4,5]</sup>。这对于在等位酶水平上变异极低的珍稀濒危物种的保护研究来说尤为迫切。

疣粒野生稻(*Oryza granulata*)是分布于中国的3种野生稻种之一,也是重要的水稻种质资源。由于它不能和栽培稻杂交产生可育后代,目前疣粒野生稻是稻属中研究最为缺乏,在品种资源保护中也最少受到关注的类群<sup>[7-9]</sup>。我们对中国疣粒野生稻所进行的全面调查表明,对该物种的干扰和破坏已经相当严重,亟需了解其居群遗传特征,从而进行有效的保护。此外,从居群生态学的角度来说,疣粒野生稻具有群落内居群数量多而各居群个体数量少,居群之间分散分布但又能通过种子散布进行基因交流的特点,维持了较为典型的Metapopulation结构<sup>[10,11]</sup>。对其进行居群遗传学研究可以作为濒危物种遗传多样性研究的范例,有较为普遍的意义。

目前,已经出现数种针对显性标记的居群遗传结构统计方法<sup>[4,5]</sup>,但是各方法之间缺乏直接对比,对各种参数在不同类型物种中的适用性了解不足。本研究利用RAPD作为显性标记技术的代表,对分布于中国的疣粒野生稻5个居群进行了遗传结构的研究,其目的在于:(1)对比各种遗传结构分析方法的异同,对它们的适用性进行评判;(2)在上述分析的基础上,讨论将显性标记用于植物居群遗传学和保护遗传学分析中需要重视的问题。

## 1 材料和方法

1.1 材料 疣粒野生稻实验材料来自云南和海南的5个居群(表1),根据居群大小在每个居群内各取20~100株个体的新鲜叶片,分别用变色硅胶进行快速干燥固定后带回实验室,室温下保存。为防止取到同一个无性系的不同分株,居群内个体间的取样间隔在

表1 疣粒野生稻居群取样地点

Table 1 Populations of *Oryza granulata* from China for RAPD analysis

居群 Population	地理位置 Location	生境情况 Enviroments	居群大小 Population size	样本数 Sample size
M16	沧源县南曩 Nanja Township, Cangyuan County	云南 Yunnan 720m alt., NW20, gradient 15, under bamboo forest.	60	20
M22	盈江县那邦 Nabang Township, Yingjiang County	云南 Yunnan 430m alt., SE 10, gradient 10, under a gap in a mountain rainforest.	25	20
M27	思茅市竹林 Zhulin Township, Simao City	云南 Yunnan 780 m alt., SE 45, gradient 25, under a second successive forest.	3000~5000	20
M5	乐东县冲坡 Chongpo Town, Ledong County	海南 Hainan 50 m alt., S, gradient 0, under shrubs at the edge of a road.	150	20
M14	白沙县金波 Jinbo Township, Baisha County	海南 Hainan 220 m alt., SE 20, gradient 20, under a bamboo-wood forest.	300~500	20

5m 以上,但在破坏严重,个体数较小的居群(如 M22)则采取了完全取样的方法。

## 1.2 方法

1.2.1 疣粒野生稻总 DNA 提取 按照 Xie<sup>[12]</sup> 的方法提取总 DNA,并根据疣粒野生稻的特点略作改动。用于正式实验的样品总 DNA 用 RNase A 进行纯化。上述 5 个居群各取 20 株,共 100 株个体用于 RAPD 分析。

1.2.2 RAPD 产物及检测 利用 4 份来自云南和海南的材料进行预备实验,从 79 个 RAPD 引物(Operon 公司)中筛选出 20 个能获得清晰条带,反应稳定的引物(B-6、B-7、B-12、K-3、K-6、K-8、K-9、K-11、K-12、K-13、K-15、Y-1、Y-2、Y-9、Y-14、Y-15、Y-18、Y-19、Z-3 和 Z-4)进行居群遗传结构分析。每个 RAPD PCR 反应体积为 10 $\mu$ l/管,其中包括 50mmol/L Tris-HCl, pH8.3, 500 $\mu$ g/ml BSA, 10% Ficoll, 1mmol/L Tartrazine, 2mmol/L MgCl<sub>2</sub>, 200 $\mu$ mol/L dNTP, 1 $\mu$ mol/L 引物, 5ng DNA 模板和 0.5U Taq 酶。反应在 1818 型毛细管型快速热循环仪(Idaho Tech.)上进行。首先 94 $^{\circ}$ C 变性 1min, 35 $^{\circ}$ C 10s, 72 $^{\circ}$ C 20s, 2 个循环,然后进行以下 40 个循环: 94 $^{\circ}$ C 变性 0s, 35 $^{\circ}$ C 0s, 72 $^{\circ}$ C 1min。最后于 72 $^{\circ}$ C 延伸 7min。反应产物在含有 EB 的 1.5% 琼脂糖凝胶中电泳检测。紫外灯下观察拍照,用 100bp 的 DNA ladder 作为分子量标记。

1.2.3 数据分析 RAPD 是显性标记,同一引物扩增产物中电泳迁移率一致的条带被认为具有同源性,按扩增阳性(1)和扩增阴性(0)记录电泳带谱,形成原始数据矩阵。采用两类方法对该原始数据进行解释。其一是将 RAPD 标记视为表征性状,直接利用原始二元数据矩阵(表征矩阵)进行计算;其二是将每一个 RAPD 扩增条带看作具有 2 个等位基因(1 和 0)的复等位基因位点,然后假设居群内基因频率处于 Hardy-Weinberg 平衡,通过隐性基因型(扩增阴性)的频率( $q^2$ ),利用平衡定律  $p^2 + 2pq + q^2 = 1$  ( $p =$  显性基因频率,  $q =$  隐性基因频率)计算各基因的频率(基因频率矩阵)后用于下一步统计。居群遗传结构分析包括居群内遗传变异水平的检测、个体之间遗传关系的测度以及居群间遗传分化程度的评价 3 个方面。首先,采用以下两类方法分析居群内遗传变异水平:(1)根据表征矩阵,统计 RAPD 扩增产物的条带总数和多态条带数,计算多态条带比率(PPB)和单位引物多态条带数,并计算基于各引物条带类型的 Shannon 多样性指数。(2)根据基因频率矩阵,按照 Lynch-Milligan 矫正小样本 RAPD 数据的方法,剔除  $q^2 < 3/N$  ( $N$  为取样大小)的条带后<sup>[5]</sup>,用 POPGEN 32 软件<sup>[13]</sup>计算观测等位基因数( $A$ )、有效等位基因数( $A_e$ )和基因多样性指数( $H_e$ )来反映不同结构层次居群内的遗传多样性水平。同时,在不进行 Lynch-Milligan 矫正的情况下,重新计算上述居群遗传学参数,对比两者的差异;其次,为了研究个体间的遗传关系,利用 RAPDistance<sup>[14]</sup> 软件计算 17 种基于表征矩阵的所有被研究个体间的相似性系数,使用 NTSYS 软件<sup>[15]</sup> 中的 Mantel 检测分析这 17 个相似性系数矩阵之间的相关关系,然后筛选出适用者进行下一步分析;最后,采用以下 3 种方法分析居群分化的程度。(1)根据基因频率矩阵,在采用和不采用 Lynch-Milligan 矫正两种情况下,利用 POPGENE 32 软件计算总居群基因多样性( $H_t$ ),各居群基因多样性( $H_s$ )和基因分化系数( $G_{st}$ ),计算居群间的 Nei 无偏遗传距离并用 UPGMA 法进行聚类分析。(2)根据个体间的遗传相似性系数,用 WINAMOVA<sup>[16]</sup> 进行分子方差分析(AMOVA),计算居群内、居群间和地区之间的变异方差分布和  $\Phi$  统计量

距离( $\Phi_{st}$ ),用UPGMA法分析居群间的关系。(3)利用Shannon多样性指数计算基于各引物扩增条带在居群( $H_{pop}$ )和物种水平的表型多样性( $H_{sp}$ ),分别根据 $(H_{pop}/H_{sp})$ 和 $(H_{sp}-H_{pop})/H_{sp}$ 计算居群内与居群间变异所占的比例<sup>[17]</sup>。同时,为了评价上述居群分化程度分析间的一致性程度,利用Mantel检测分析各种遗传距离矩阵之间的相关系数,并做显著性检测。

## 2 结果与分析

2.1 居群内遗传多样性 20个RAPD引物在疣粒野生稻样品中的扩增结果见表2。5个疣粒野生稻居群共获得199条RAPD条带,分子量从220到2000bp,其中有61条(30.65%)为多态条带。根据PPB值,各居群遗传变异由高向低为居群M27 > M5 > M16 > M22 > M14。以地区为单位,云南3个居群的PPB值(20.60%)高于海南2个居群

表2 中国疣粒野生稻居群RAPD分析的遗传多态性

Table 2 RAPD polymorphism of populations of *Oryza granulata* from China

居群 Population	$N_b$	$N_{pb}$	PPB	$N_b/p$	$I$
M16	151	13	6.53	0.65	0.0385
M22	145	9	4.53	0.45	0.0228
M27	150	26	13.06	1.30	0.0741
云南 Yunnan	160	41	20.60	2.05	0.0927
M5	148	25	12.56	1.25	0.0941
M14	156	9	4.52	0.45	0.0290
海南 Hainan	162	31	15.58	1.25	0.1432
中国 China	199	61	30.65	3.05	0.3171

注:  $N_b$ : 条带总数;  $N_{pb}$ : 多态条带总数; PPB: 多态条带比率;  $N_{pb}/p$ : 单位引物多态比率;  $I$ : 香农多样性指数

Note:  $N_b$ : No. of RAPD bands,  $N_{pb}$ : No. of polymorphic bands, PPB: percentage of polymorphic bands,  $N_{pb}/p$ : No. of polymorphic bands per primer,  $I$ : Shannon diversity

(15.58%)。Shannon多样性指数分析与PPB分析不同。遗传多样性高下为居群M5 > M27 > M16 > M14 > M22。并且以地区为单位时,海南两居群的Shannon多样性指数( $I = 0.1432$ )高于云南3个居群( $I = 0.0927$ )。

基于基因频率矩阵,利用平均观察等位基因数( $A$ ),平均有效等位基因数( $A_e$ )和Nei基因多样性指数( $H_e$ )对疣粒野生稻居群内遗传多样性进行评价的结果见表3,同时列出是否进行Lynch-Milligan矫正后的计算结果。在未进行矫正时,5个居群 $H_e$ 的变化趋势与Shannon多样性指数分析的结果完全一致。而在进行Lynch-Milligan矫正后,所获结果与上述两种分析的结果不完全相同,即居群M27 > M5 > M16 > M14 > M22。表3表明矫正后的基因多样性指数 $H_e$ 普遍高于不矫正的结果。以地区为单位时,无论是否进行Lynch-Milligan矫正,海南居群的 $H_e$ 都高于云南居群,与Shannon多样性指数的计算结果一致。

2.2 个体间的遗传关系 Mantel检测分析所获17种相似性系数矩阵(数据未显示)两两之间的相关系数见表4。它们之间都存在极显著的相关性( $t > t_{0.01}$ ),相关系数 $r$ 均高于0.950,17个相应的UPGMA聚类图的形状也非常一致(数据未显示)。因此,这些相似性系数在应用于疣粒野生稻时的区别非常微小,其中,Apostol和Excoffier指数甚至表现出完全的相关。根据上述结果,我们选用了在居群生物学中常用的Jaccard相似性系数进行

表3 疣粒野生稻居群的基因多样性

Table 3 Gene diversity of populations of *Oryza granulata* from China

	Lynch-Milligan修正			未修正		
	$A$	$A_e$	$H_e$	$A$	$A_e$	$H_e$
M16	1.1146	1.0411	0.0363	1.0754	1.0581	0.0253
M22	1.0312	1.0249	0.0117	1.0402	1.0190	0.0152
M27	1.1875	1.0864	0.0708	1.1307	1.1219	0.0503
云南 Yunnan	1.3329	1.0876	0.0891	1.2312	1.1428	0.0575
M5	1.1562	1.1126	0.0651	1.1558	1.1138	0.0646
M14	1.0833	1.0325	0.0316	1.0503	1.0515	0.0196
海南 Hainan	1.3854	1.1655	0.1514	1.2663	1.2741	0.0958
中国 China	1.9896	1.3710	0.3923	1.6131	1.7014	0.2130

注:  $A$ : 等位基因平均数;  $A_e$ : 有效等位基因数;  $H_e$ : 基因多样性

Note:  $A$ : allele number,  $A_e$ : effective allele number,  $H_e$ : gene diversity

### AMOVA 分析.

2.3 居群遗传分化 用 AMOVA 进行的遗传变异(hierarchy structure)巢式方差分析见表 5。中国的疣粒野生稻在云南与海南之间以及各居群之间产生了极显著的( $P < 0.001$ )遗传分化。对三级谱系的 AMOVA 分析表明,有 73.85% 的变异存在于云南与海南之间,有 19.45% 的变异存在于地区内居群之间,而各居群内只包含了 6.70% 的遗传变异。基于 Jaccard 相似性系数,利用 AMOVA 计算出居群间的  $\Phi_{st}$  遗传距离(数据未显示)。5 个疣粒野生稻居群之间的平均遗传距离为 0.846,其中,云南与海南之间的平均遗传距离达 0.934。图 1A 为根据  $\Phi_{st}$  遗传距离的居群间 UPGMA 聚类图。根据基因频率矩阵,利用基因分化系数  $G_{st}$  对 5 个疣粒野生稻居群按二级谱系进行了分析(表 6)。是否经 Lynch-Milligan 修正计算的中国疣粒野生稻  $G_{st}$  值分别为 0.7093 和 0.6579,均表明居群之间产生了比较大的分化。其中,经 Lynch-Milligan 修正后计算得到的基因分化系数  $G_{st}$  都高于未经 Lynch-Milligan 修正的数值。5 个疣粒野生稻居群之间的平均遗传距离为 0.7480,其中,云南与海南居群之间的平均遗传距离为 1.1667。在不进行 Lynch-Milligan 修正的情况下,5 个疣粒野生稻居群之间的平均遗传距离为 0.2781,云南与海南居群之间的平均遗传距离达 0.4227。根据上述两种 Nei 遗传距离(数据未显示)用 UPGMA 进行的居群水平聚类图分别见图 1B 和图 1C。

Mantel 检测表明,有否进行 Lynch-Milligan 修正的两个遗传距离矩阵之间存在极显著的相关性( $r = 0.9986$ ,  $t > t_{0.01}$ ), 3 000 次 permutation 过程中无一次大于 Z 值。而  $\Phi_{st}$  遗传距离与两种 Nei 遗传距离之间存在显著相关性,相关系数分别达到 0.8530 和 0.8413。因此,这 3 个距离矩阵之间存在线性关系。此外,Nei 遗传距离对云南和海南两地区之内和地区之间的分辨程度明显大于  $\Phi_{st}$  遗传距离矩阵。而在两个 Nei 遗传距离矩阵之间,进行了 Lynch-Milligan 修正者体现出更大的检测遗传距离的能力(图 1)。

利用 Shannon 多样性指数分析变异在居群内和居群间的分布情况见表 7。根据( $H_{sp} / H_p$ )和( $H_{sp} / H_{pop} / H_p$ )分析所得结果显示平均有 79.7% 的变异存在于居群之间,这个结论与利用 AMOVA 和  $G_{st}$  分析的结论是比较一致的。此外 Shannon 多样性指数分析还表明不同引物在居

表4 个体间遗传相似性系数的Mantel检测

Table 4 Mantel test among matrix of genetic similarities based on individuals

1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1																	
2	0.9991	1															
3	0.9999	0.9990	1														
4	0.9951	0.9965	0.9949	1													
5	0.9827	0.9797	0.9830	0.9946	1												
6	0.9946	0.9980	0.9942	1.0000	0.9733	1											
7	0.9951	0.9965	0.9949	0.9950	0.9659	0.9945	1										
8	1.0000	0.9991	1.0000	0.9991	0.9829	0.9944	0.9950	1									
9	0.9919	0.9951	0.9915	0.9814	0.9600	0.9958	0.9990	0.9917	1								
10	0.9823	0.9763	0.9828	0.9824	0.9606	0.9632	0.9816	0.9826	0.9724	1							
11	0.9995	0.9972	0.9996	0.9918	0.9844	0.9907	0.9925	1.0000	0.9878	0.9854	1						
12	0.9987	0.9966	0.9988	0.9980	0.9816	0.9902	0.9919	0.9987	0.9874	0.9836	0.9992	1					
13	0.9994	0.9994	0.9993	0.9980	0.9777	0.9959	0.9980	0.9993	0.9956	0.9827	0.9980	0.9974	1				
14	0.9994	0.9995	0.9993	0.9980	0.9777	0.9959	0.9980	0.9993	0.9956	0.9827	0.9980	0.9974	1.0000	1			
15	0.9970	0.9992	0.9967	0.9979	0.9732	0.9991	0.9978	0.9968	0.9979	0.9726	0.9939	0.9934	0.9985	0.9985	1		
16	0.9995	0.9980	0.9995	0.9962	0.9798	0.9923	0.9962	0.9995	0.9923	0.9875	0.9993	0.9983	0.9994	0.9994	0.9960	1	
17	0.9994	0.9994	0.9993	0.9980	0.9777	0.9959	0.9980	0.9993	0.9956	0.9827	0.9980	0.9974	1.0000	1.0000	0.9985	0.9994	1

注(Notes): 相似性系数(Name of similarities): 1: Nei & Li(1979), 2: Jaccard, 3: Kulczynski, 4: Phi, 5: Russell & Rao, 6: Anderberg, 7: Sokal & Sneath, 8: Ochiai, 9: Sokal & Sneath 2, 10: Yule & Kendall, 11: Upholt, 12: Li & Graur, 13: Apostol, 14: Excoffier, 15: Rogers & Tanimoto, 16: Sneath and Sokal 3, 17: Spath

表5 中国的疣粒野生稻居群的AMOVA分析

Table 5 Analysis of molecular variance (AMOVA) for 100 individuals of *Oryza granulata* using RAPD

谱系结构 Source of variance	SSD	VC	百分比 Percentage	P-value
Yunnan vs. Hainan	7.6233	14.29	73.85	<0.001
Among populations/regions	2.2967	3.76	19.45	<0.001
Within populations	1.2319	1.30	6.70	<0.001

注: SSD: 方差总和; VC: 变异组分; 3 000次permutation显著性检测

Note: SSD: sum of squared deviation, VC: variant component; significance tests after 3 000 permutations

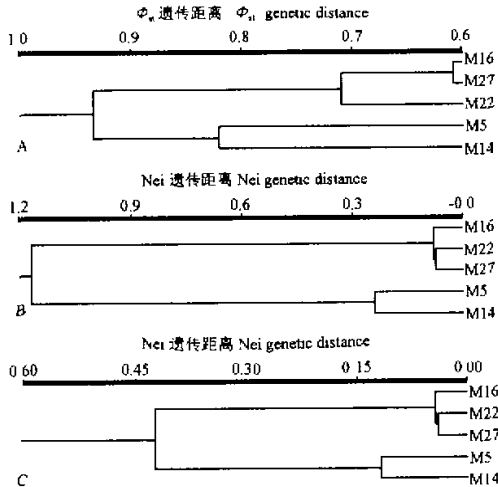


图1 中国疣粒野生稻居群遗传距离

Fig.1 Dendrograms of populations of *Oryza granulata* from China

群内和居群间检测变异的能力有较大差异。例如引物 Y14、Z03、和 Y03 检测居群内变异的能力较强,而 K8、Y14 和 Y18 检测居群间变异的能力较强。

表6 中国的疣粒野生稻居群基因分化 $G_{st}$ 分析Table 6  $G_{st}$  analyses of population genetic differentiation of *Oryza granulata* from China

	Lynch-Milligan修正 Lynch-Milligan 'pruning		未修正 Non pruning			
	$H_s$	$H_t$	$G_{st}$	$H_s$	$H_t$	$G_{st}$
云南	0.0891	0.0396	0.5557	0.0575	0.0303	0.4735
Yunnan	(0.0239)	(0.0056)		(0.0152)	(0.0043)	
海南	0.1514	0.0483	0.6806	0.0958	0.0421	0.5608
Hainan	(0.0429)	(0.0088)		(0.0295)	(0.0084)	
中国	0.3923	0.1140	0.7093	0.2130	0.0729	0.6579
China	(0.0154)	(0.0153)		(0.0422)	(0.0113)	

注:  $H_s$ : 居群内基因多样性;  $H_t$ : 总基因多样性;  $G_{st}$ : 基因分化系数; 括号中为标准差Note:  $H_s$ : gene diversity within populations;  $H_t$ : total gene diversity;  $G_{st}$ : coefficient of gene differentiation; Standard deviations are in parentheses

表7 疣粒野生稻居群遗传多样性的Shannon多样性指数分析

Table 7 Partitioning of the Shannon diversity between and within populations of *Oryza granulata* for 20 primers of RAPD

引物 Primers	$H_{pop}$	$H_T$	$(H_T - H_{pop}) / H_T$	引物 Primers	$H_{pop}$	$H_T$	$(H_T - H_{pop}) / H_T$
K3	0.564	2.898	0.805	Y3	0.880	2.202	0.600
K6	0.125	4.947	0.975	Y8	0.497	2.332	0.787
K8	0.000	1.674	1.000	Y14	1.095	6.034	0.819
K9	0.765	3.883	0.803	Y15	0.489	2.197	0.777
K11	0.720	4.003	0.820	Y18	0.478	6.670	0.928
K12	0.356	2.922	0.878	Y19	0.212	2.993	0.929
K13	0.481	3.348	0.856	Z3	1.180	4.957	0.762
K15	0.693	3.408	0.797	Z4	0.125	3.158	0.960
B6	0.299	0.863	0.654	Mean	0.514	3.155	0.797
B7	0.130	0.215	0.395	平均			
B12	0.699	1.602	0.564	标准差	0.322	1.618	0.149
Y1	0.500	2.799	0.821	St.dev.			

注:  $H_{pop}$ : 居群内 Shannon 多样性;  $H_T$ : 总 Shannon 多样性;  $(H_T - H_{pop}) / H_T$ : Shannon 多样性分化系数

Note:  $H_{pop}$ : Shannon diversity within populations,  $H_T$ : total Shannon diversity,  $(H_T - H_{pop}) / H_T$ : gene differentiation by Shannon diversity

### 3 讨论

**3.1 RAPD 在居群遗传分析中的特点** RAPD 的遗传多态现象比较复杂。Williams 发现约有 5% 的条带表现为共显性遗传方式<sup>[18]</sup>, 此外, Hunt 等和 Heun 等的研究表明 RAPD 的遗传多态还包括条带亮度的差别和异源双链条带上<sup>[19,20]</sup>, 前者表现为显性遗传的方式, 其形成原因尚不清楚, 本研究也观察到大量这种类型的变异, 后者则更多的是一种人为现象。尽管如此, 大多数 RAPD 遗传多态表现为扩增条带的有无, 对二倍体  $F_2$  分离群体的研究证实是为典型的显性遗传方式<sup>[18]</sup>。因此, 在利用 RAPD 进行居群遗传学研究时, 绝大部分条带被当做显性标记来进行处理。其中, PCR 扩增阴性主要是由于引物结合位点发生了突变导致不能形成稳定的引物-模板复合物, 或者引物结合位点之间的 DNA 片段长度变异超过了 PCR 的扩增范围 (2~3kb) 所致<sup>[18]</sup>。从理论上来说, 条带的浓度本可以反映显性纯合子与杂合子之间的差异, 然而由于精确控制 PCR 反应条件极其困难, 事实上无法以此确定扩增阳性条带的基因型是显性纯合还是杂合<sup>[4]</sup>。这使得将 RAPD 这一类显性标记运用于居群遗传学分析中时的数据处理与解释比较复杂, 根据它们得到的结果比起共显性标记如等位酶或微卫星 (microsatellite) 分析要损失一定的信息。

**3.2 居群内遗传变异水平** 在评价居群内遗传多样性的参数中, 虽然多态条带比率 (PPB) 计算简单直观, 能够反映一定的遗传多样性程度, 在研究中得到了广泛的应用<sup>[21,22]</sup>, 但是它不能确定各条带在频率上的均匀程度, 且受到样本大小和条带总数的影响, 它只是衡量居群遗传多态性的一个粗略估计值, 且倾向于低估遗传多态水平。在这种情况下, 基于条带表型频率的 Shannon 多样性指数<sup>[17]</sup>和基于 Hardy-Weinberg 假设的 Nei 基因多样性指数<sup>[23]</sup>可以得到更为可信的衡量指标。在对中国的 5 个疣粒野生稻居群遗传多态性进行评价时, Shannon 多样性指数和未进行 Lynch-Milligan 矫正的 Nei 基因多样性指数获得了完全一致的居群变异趋势, 它们能同时反映多态条带类型和相对频率上的变化, 比 PPB 值



更为客观。从方法上来考虑,显性标记倾向于低估居群内的遗传变异水平,这主要是因为某些扩增阳性的单态条带有可能是杂合的。另外,在进行居群遗传结构分析时,由于对某一条带而言,其频率超过0.5会导致遗传多态估计偏低,而频率低于0.5则又可能对遗传多态估计偏高<sup>[4,5]</sup>,因此在进行数据解释时应该考虑这些问题。

作为显性标记,RAPD位点无法得到实际杂合度  $H_o$  信息,因此居群内的遗传变异程度只能利用基因多样性  $H_e$  来描述。在实际运用时,因为 Lynch-Milligan 矫正删除了所有扩增阴性条带频率低于  $3/N$  的标记,减少了单态位点的数目,而某些单态位点有可能是杂合的,位点删除会导致低估居群内的基因多样性,根据研究我们认为在利用基因多样性指数  $H_e$  评价居群内遗传多样性水平时,不必对 RAPD 原始数据进行 Lynch-Milligan 矫正。

**3.3 个体间遗传关系** Nei 遗传相似性系数是应用最广泛的度量指标,但是它是从基因频率数据估计每个位点的平均密码子差异<sup>[24]</sup>,不把 RAPD 数据假定为双等位基因位点而直接利用 Nei 遗传距离来衡量个体或居群之间的计算方法是错误的。在此情况之下,可行方法之一是将 RAPD 条带的有无视为表型数据,利用表型相似性系数来反映 OTU 之间的遗传相似性。在本研究中对疣粒野生稻 100 个体 17 个相似性系数<sup>[14]</sup> 矩阵之间的 Mantel 检测表明,它们之间存在着极显著的相关性 ( $r > 0.95, t > t_{0.01}$ ),根据它们所得出的 UPGMA 聚类图虽然在分支长度上略有区别,但是基本的模式都比较一致,即将所有个体分为云南和海南两大类,各类中来自同一居群的个体被聚为一群,居群内个体间的遗传相似性较高而地区间的相似性较低。因此,具体利用哪一种可以根据研究的目的和与其他研究的可比性来选择。其中,Jaccard 系数和 Nei & Li<sup>[25]</sup> 系数属于结合系数,在对二元编码的数据矩阵进行相似性检测时,它们测度 OTU 之间的相符程度,所获信息优于测度 OTU 在多维空间中相对位置的距离系数和测度有关对偶向量间比例性的相关系数,特别是 Nei & Li 系数利用样品间共有产物的期望值来衡量它们之间的相互关系,在条带同源性一致的前提下,具有明显的生物学意义<sup>[26]</sup>。此外,鉴于各种相关系数在对不同数据矩阵分析时可能产生相异的结果,在进行研究时,可以用 Mantel 检测或 DIPLOMO<sup>[27]</sup> 分析对比各结果之间的一致程度,然后根据 OTU 的生物学特点来进行选择。所获得的个体间遗传相似性系数可用于 UPGMA 聚类和 AMOVA 分析。

特别重要的是,对于 RAPD 而言,实验中存在非同源但分子量相近的扩增片段在电泳中共迁移到同一位置的可能性,因此绝对不能在不做同源性假设和利用实验加以证实的情况下直接使用分支系统学分析方法(如 parsimony)来研究样品所代表的类群之间(特别是种上等级)的谱系关系<sup>[21]</sup>。在应用于属下和种内关系研究时,虽然研究表明 RAPD 条带存在比较高的同源性<sup>[28]</sup>,但在未证实的情况之下,将其作为同源性状数据处理比作为表型数据处理得出有偏差结论的可能性要高<sup>[29]</sup>。

**3.4 居群遗传结构** 利用显性标记(如 RAPD)对自然居群的遗传结构进行分析一般采用两类方法,第一类把条带看作是含两个复等位基因的位点系统,假设居群处于 Hardy-Weinberg 平衡下估计出基因频率,并利用 Lynch-Milligan 对小样本进行矫正后计算其居群遗传学参数。另一类不进行 Hardy-Weinberg 平衡假设,利用条带的表型来估计遗传多样性,包括 AMOVA 分析中的  $\Phi_s$  统计量分析和 Shannon 多样性指数的阶层分析<sup>[4]</sup>。根据对 5 个中国的疣粒野生稻居群遗传结构的研究,在利用第一类方法时,其缺点在于由于 RAPD

分析中的数学假设,不能得到单位位点等位基因数( $A$ )和观测杂合度( $H_o$ )这两个指标。其次相对于共显性标记来说,分析所作的假设会导致在计算居群遗传参数时的可信度有所降低。但是,这类分析仍然能给出了极有价值的信息。例如,利用  $G_{st}$  基因分化系数的分析表明,在疣粒野生稻居群之间存在着较高等度的遗传分化。而 Nei 遗传距离分析表明这些分化主要存在于云南和海南之间,地区内和居群内的遗传一致性则较高。这与 Gao<sup>[30]</sup>等利用等位酶研究同一物种所获得的结论是一致的。这种独特的遗传多样性格局反映了疣粒野生稻的物种起源和散布过程有可能在其居群遗传结构的形成中起着主导作用。此外,根据  $G_{st}$  的计算结果表明疣粒野生稻居群的每代迁移数( $N_m$ 值)较低(数据未显示),因此居群之间缺乏基因交流,不同形式的生殖隔离也是影响其居群遗传结构的重要因素。最后,研究证明在利用这一类方法进行居群遗传格局分析时如果不对样本进行 Lynch-Milligan 矫正,则会导致遗传多样性水平和遗传分化程度的估计偏低,我们的研究证实在进行 Hardy-Weinberg 平衡假设分析其遗传结构的情况下,应该对样本进行 Lynch-Milligan 矫正。与此同时,由于这一矫正减少了可用位点的数目,为了获得与应用共显性标记相近的参数估计,Lynch 等研究建议分析时应扩大居群内样品的采样数量和增加 RAPD 位点的数目<sup>[5]</sup>。

基于 RAPD 条带表型的 AMOVA 分析支持基于 Hardy-Weinberg 假设的分析结果。尽管它们所获得的居群间遗传分化的程度略有差异,且后者比前者能探测到更大程度的居群间分化,但 Mantel 检测表明  $\Phi$  遗传距离和 Nei 遗传距离矩阵之间存在极显著的相关性。它们对中国的疣粒野生稻居群遗传结构格局所得出的结论也比较一致。因此,这两种分析方法都可以比较可信地应用于实际的研究当中。我们倾向于在研究中使用 AMOVA 分析,主要是因为它不必作 Hardy-Weinberg 假设,避免了数学假设带来的系统误差;另外 AMOVA 分析可以利用显著性检测分析遗传多样性的数值结果,因而更为可靠,可作为利用显性标记分析居群遗传结构时优先选择的统计方法<sup>[16,31]</sup>。

Shannon 多样性指数分析提供了第三种区分遗传变异在阶层结构中分布的方法<sup>[17]</sup>。对比其与  $G_{st}$  和 AMOVA 分析的结果可见,Shannon 多样性指数分析与后两者得出的结论基本上是一致的,而且相对来说,它与由 AMOVA 得出的结论更相似,这可能是由于它们都将 RAPD 数据作为表型数据处理的缘故。此外,Shannon 多样性可以直接获得关于特定 RAPD 引物在居群内和居群间检测变异能力的衡量指标,因此,该分析方法可以作为 AMOVA 和  $G_{st}$  分析的有益补充。该指标的缺点在于它是关于体系信息量的指标而与任何遗传物质无关,因而无法确定它的绝对值与遗传变异的确切关系。

值得注意的是,上述几类遗传结构分析方法之间的一致性也与疣粒野生稻的生物学特性有密切的关系。由于疣粒野生稻居群内的遗传多样性水平极低,小居群、无性系和随机遗传漂变增加了个体间近交的机会,因此,居群内个体的纯合比率比较高,遗传相似性分析的结果证实了这一点。所以,当利用 RAPD 这一类显性标记对疣粒野生稻进行研究时,不能检测杂合体对该物种遗传多样性水平评价的数值偏估远小于那些居群内杂合体丰富,遗传多样性水平高的物种。同时,这种情况也有利于基于条带表型的分析方法所作的关于表型关系有可能反映遗传关系的假设。显然这一结论对将显性标记运用于珍稀濒危物种的保育遗传学研究非常有利,因为居群内遗传多样性偏低和存在大量的近交是这

一类物种的基本生物学特征。

致谢: 野外工作得到了本室谢中稳博士和有关县市农业局的大力协助, 数据分析时得到了美国 Colorado State University 的 William C Black 教授的指点, 在此特致感谢。

### 参考文献

- [1] Cruzan M B. Genetic markers in plant evolutionary ecology. *Ecology*, 1998, 79: 400~412.
- [2] Hedrick P W. Conservation genetics and molecular techniques: A perspective. In: Smith J B, Wayne R K. (eds) *Molecular Genetic Approaches in Conservation*. Oxford University Press, London: 1996, 459~477.
- [3] Schaal B A, Leverich W J, Rogstad S H. Comparison of methods for assessing genetic variation in plant conservation biology. In: Falk D A, Holsinger K E (eds) *Genetics and Conservation of Rare Plants*. Oxford University Press, New York: 1991, 123~134.
- [4] Fritsch P, Rieseberg L H. The use of random amplified polymorphic DNA (RAPD) in conservation genetics. In: Smith T B, Wayne R K (eds) *Molecular Genetic Approaches in Conservation*. Oxford University Press, London: 1996, 54~73.
- [5] Lynch M, Milligan B G. Analysis of population genetic structure with RAPD markers. *Mol. Ecol.*, 1994, 3: 91~99.
- [6] Clark A G, Lanigan C M S. Prospects for estimating nucleotide divergence with RAPDs. *Mol. Ecol. Evol.*, 1993, 10: 1096~1111.
- [7] Ge S, Sang T, Lu B R, Hong D Y. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl. Acad. Sci. USA*, 1999, 96: 14400~14405.
- [8] Vaughan D A. *The Wild Relatives of Rice*, IRRI, Manila: 1994.
- [9] Aggarwal R *et al.* Two new genomes in the *Oryza* complex identified on the basis of molecular divergence analysis using total genomic DNA hybridization. *Mol. Gen. Genet.*, 1997, 254: 1~12.
- [10] Hanski I. Metapopulation dynamics. *Nature*, 1998, 369: 41~49.
- [11] Hanski I, Gilpin M. (eds) *Metapopulation Biology: Ecology, Genetics, and Evolution*, Academic Press, London: 1996.
- [12] 谢中稳, 葛 颂, 洪德元. 从普通野生稻硅胶干燥的小量叶片中制备 DNA 用于 RAPD 分析和总 DNA 库的建立. *植物学报*, 1999, 41: 807~812.
- [13] Yeh F C, Yang R. POPGENE v 1.31 <http://www.ualberta.ca/~fyeh/>.
- [14] Armstrong J S, Gibbs A J, Peakall R, Weiller G. The RAPDistance Package. 1994, <ftp://life.anu.edu.au/pub/software/RAPDistance>.
- [15] Rohlf F J. NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System, v. 1.80, Exeter Software, NY, 1994.
- [16] Excoffier L. Analysis of molecular variance (AMOVA) version 1.5. Genetics and Biometry Laboratory, University of Geneva. 1993.
- [17] Lewontin R C. The apportionment of human diversity. *Evol. Biol.*, 1972, 6: 381~398.
- [18] Williams J G K, Kubelik A R, Livak K J, Rafalski J A, Tingey S V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.*, 1990, 18: 6531~6535.
- [19] Hunt G J, Page J r. Patterns of inheritance with RAPD molecular markers reveal novel types of polymorphism in the honey bee. *Theor. Appl. Genet.*, 1992, 85: 15~20.
- [20] Heun M *et al.* Inheritance of RAPDs in  $F_1$  hybrids of corn. *Theor. Appl. Genet.*, 1993, 85: 961~968.
- [21] Ge S, Oliveira G C, Schaal B A, Gao L Z, Hong D Y. RAPD variation within and between natural populations of the wild rice *Oryza rufipogon* from China and Brazil. *Heredity*. 1999, 82: 638~644.
- [22] Fahima T, Sun G L, Beharav A, Krugman T, Beiles A, Nevo E. RAPD polymorphism of wild emmer wheat populations, *Triticum dicoccoides*, in Israel. *Theor. Appl. Genet.*, 1999, 98: 434~447.
- [23] Nei M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA*, 1973, 70: 3321~3323.

- [24] Nei M. Genetic distance between populations. *Amer. Naturalist*, 1972, 6: 283~293.
- [25] Sneath P, Sokal, R. *Numerical Taxonomy*. W. H. Freeman & Company, San Francisco, 1973.
- [26] Nei M, Li W H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA*, 1979, 76: 5269~5273.
- [27] Weiller G F, Gibbs A. DIPLOMO: The tools for a new type of evolutionary analysis. *CABIOS*, 1995, 11: 535~540.
- [28] Rieseberg L H. Homology among RAPD fragments in interspecific comparisons. *Mol. Ecol.*, 1996, 5: 99~105.
- [29] Thormann C E, Ferreira M E *et al.* Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. *Theo. Appl. Genet.*, 1994, 88: 973~980.
- [30] Gao L Z, Ge S, Hong D Y *et al.* A study on population genetic structure of *Oryza meyeriana* (Zoll. et Mor. ex Steud.) Baill. from Yunnan and its *in situ* conservation significance. *Science in China (Series C)*, 1999, 42: 102~108.
- [31] Excoffier L *et al.* Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 1992, 131: 479~491.

## Analyses of Population Genetic Structure by Using Dominant Markers

QIAN Wei, GE Song

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

**Abstract:** Dominant markers tend to under-estimate the amount of genetic diversity relative to codominant systems when applying in population genetics. In order to compare various existing methods for analysis of genetic structure, RAPD markers were used to detect genetic variability of 5 populations of *Oryza granulata* from China. The results indicated that both Shannon index of diversity and Nei gene diversity were superior to percentage of polymorphic bands (PPB) because the latter lacked of ability to describe frequency difference of polymorphic bands. Mantel test showed significant relation ( $r > 0.95$ ,  $t > t_{0.01}$ ) among matrices of 17 different genetic similarities, which indicated that all of them could be used in analysis of genetic relations of individuals of *Oryza granulata*. Both AMOVA analysis based on  $\Phi_{st}$  distance and analysis of Nei's distance showed consistent results in defining relationship among the 5 populations, and Lynch-Milligan pruning should be used to improve the estimation of population parameters. All of AMOVA,  $G_{st}$  and Shannon diversity analyses obtained similar results with majority of genetic variation occurring between Yunnan and Hainan, and low levels of genetic diversity resided within regions and populations.

**Key words:** dominant markers; population genetic structure; *Oryza granulata*; RAPD; Hardy-Weinberg equivalent