

# 植物遗传多样性研究中等位基因酶分析的 遗传参数及其计算方法(综述)

葛学军

(中国科学院华南植物研究所, 广州 510650)

## THE GENETIC PARAMETERS AND THEIR CALCULATING METHODS OF ALLOZYME ANALYSIS IN PLANT GENETIC DIVERSITY RESEARCH

Ge Xuejun

(South China Institute of Botany, Academia Sinica, Guangzhou 510650)

**关键词** 等位基因酶; 遗传参数; 计算公式

遗传多样性是生物多样性的重要组成部分。它主要指种内不同居群之间或一个居群内不同个体的遗传变异的总和<sup>[1]</sup>。

遗传多样性可以分为四个水平, 即居群水平、个体水平、细胞水平和分子水平。分子水平的遗传多样性, 即结构基因和调节基因以及基因的产物—蛋白质的多样性是当前研究的热点之一。其中等位基因酶(Allozymes)作为由同一基因位点的不同等位基因编码的酶, 它的酶谱反映了等位基因的组成和变化, 是基因的极好标记。因此在调查生物遗传多样性的工作中, 等位酶分析目前占有最重要的位置<sup>[2]</sup>。

在我国的遗传学研究以及植物系统和进化研究中, 也作过一些等位酶电泳工作, 但大多是个别取样, 而不是居群取样。同时对酶谱只是描述性报道, 不作遗传学解释和分析, 个别甚至直接把酶谱上的带的迁移率的差异作为数量性状进行聚类分析, 这是非常危险的, 会导致完全错误的结论<sup>[2]</sup>。

对于等位基因酶的电泳酶谱带, 必须首先经过遗传分析, 确定出位点和等位基因, 写出每个个体的基因型组成, 然后求得每个位点上每个基因在居群中的频率, 最后以基因频率为基础, 对居群的遗传结构进行分析。

近一二十年来, 随着同工酶电泳方法, 特别是“水平切片淀粉凝胶电泳方法”的发展, 使得植物的分子遗传标记测定不仅变得快速、简便, 而且花费相对低廉。所需的植物材料也仅为叶片或幼苗等。目前国内已有北京植物所等单位开展了植物居群的等位酶分析工作<sup>[3]</sup>。

### 1 居群遗传变异水平的度量

居群的遗传变异一般从等位基因的丰富程度(Richness)和等位基因频率分布的均匀程度

(Evenness) 两方面来进行估算。最常用的估算参数有 5 个。

表 1 光泽石松 Woodridge 居群的遗传参数值<sup>[9]</sup>

Table 1 Genetic variability estimates of Woodridge population in *Lycopodium lucidulum*

Locus	Allelic frequencies	J	He	A	Ae	The amount of genetic variation
PGI-2						
a	0.68	0.5648	0.4352	2	1.77	P=22.2% A=1.22 Ap=2 Ae=1.12 He=0.07 J=0.927
b	0.32					
G <sub>6</sub> PD-1		0.8698	0.1302	2	1.15	
a	0.93					
b	0.07					
G <sub>6</sub> PD-2		1	0	1	1	
a	1.00					
b	0.00					
PGM						
a	0.86	0.7592	0.2408	2	1.32	
b	0.14					
LGGP-1						
a	0.50	0.5	0.5	2	2	
b	0.50					

\* 测定位点总数为 18 个, 另外 13 个位点均是单态位点

A total of 18 loci was scored, another 13 loci were monomorphic.

### 1.1 居群中多态位点的百分率 (P)

多态位点的定义为: 绝大多数等位基因的频率小于或等于 0.99 的位点。即当居群中某位点有两个以上等位基因而且每个等位基因的频率均在 0.01 以上时, 该位点就被称为多态位点, 否则就是单态位点<sup>[9]</sup>。

$$P = \frac{k}{n} \times 100\%$$

式中: k 为等位酶测定的多态位点的数目;

n 为等位酶测定的位点总数。

表 1 中, 18 个位点中有 4 个多态位点, 因此 P=22.2%。

### 1.2 每个位点的等位基因平均数 (A)

$$A = \frac{\sum x}{n}$$

式中,  $\sum x$  为所有测定位点 (包括单态位点) 上等位基因的数目之和;

n 为等位酶测定位点的总数。

表 1 中, 18 个位点上共有 22 个等位基因, 因此 A=1.22。

另外一个相关的参数为: 多态位点上的等位基因平均数 (A<sub>p</sub>)

$$A_p = \frac{\sum x_p}{n_p}$$

表 1 中, 在 4 个多态位点上有 8 个等位基因, 因此 A<sub>p</sub>=2。

### 1.3 每个位点上等位基因的有效数目 (Ae)

一个位点上  $A_e$  的计算方法为:  $A_e = 1/\sum p_i^2$

式中,  $p_i$  为第  $i$  个等位基因的频率。

当测定了多个位点时, 先计算一个位点上的等位基因有效数目, 然后各位点的算术平均值即为一个居群的每个位点上等位基因的有效数目。表 1 中,  $A_e = 1.12$ 。

这种方法趋向于降低等位基因作为遗传变异来源的重要性。

如某一位点中只有两个等位基因  $p$  和  $q$ , 如果它们的频率为  $p_1 = 0.5$ ,  $q_1 = 0.5$ , 那么  $A_e = 1/(0.5^2 + 0.5^2) = 2$ ; 如果  $p_2 = 0.95$ ,  $q_2 = 0.05$ , 那么  $A_e = 1/(0.95^2 + 0.05^2) = 1.1$ 。

#### 1.4 平均期望杂合度 ( $H_e$ )

每个位点的纯合度为  $J_k = \sum p_i^2$

杂合度为:  $H_k = 1 - J_k = 1 - \sum p_i^2$

所有测定位点(包括单态位点)上纯合度及杂合度的算术平均, 即为居群的平均纯合度 ( $J$ ) 及平均期望杂合度 ( $H_e$ )。

表 1 中,  $J = 0.927$ ,  $H_e = 0.07$ 。

此概念是通过随机交配居群的分析发展而来的, 是在 Hardy-Weinberg 平衡下, 随机交配居群的期望值。因为 Hardy-Weinberg 平衡是指: 在一个大的、随机交配的居群中, 在没有选择、突变以及迁移的情况下, 该居群基因的相对频率将保持恒定<sup>[4]</sup>。但在非随机交配的居群中, 上述定义的杂合度则与居群内杂合体的频率无关。不过它仍然是一个居群内基因变异的理想测度, 它可以适用于任何生物。因此 Nei 建议把  $H_e$  改称为“基因多样性”<sup>[5]</sup>。

#### 1.5 固定指数 ( $F$ )

$$F = 1 - \frac{H_o}{H_e}$$

式中,  $H_o$  为观察到的实际杂合体比率。  $H_o$  的计算可按下式进行<sup>[6]</sup>:

$$H_o = \frac{1}{n} \sum_{i=1}^n (1 - \sum_{j=1}^{l_i} p_{ij}) = \frac{1}{n} \sum_{i=1}^n h_{oi}$$

式中:  $h_{oi} = 1 - \sum_{j=1}^{l_i} p_{ij}$ , 为第  $i$  个位点上实际杂合体比率;

$p_{ij}$  为第  $i$  个位点上第  $j$  个纯合体的频率;

$l_i$  为第  $i$  个位点上纯合体的种类数。

表 3 中, 4 个等位基因纯合体的频率之和  $\sum p_{ij} = 0.1613 + 0.1684 + 0.0058 + 0.0197 = 0.3552$ , 因此观察到的实际杂合体比率  $H_o = 1 - 0.3552 = 0.6448$ 。

理论期望杂合度  $H_e = 1 - \sum p^2 = 1 - (0.1326 + 0.1681 + 0.0102 + 0.0229) = 0.6662$ 。

因此  $F = 1 - \frac{0.6448}{0.6662} = 0.032$ 。

可以通过固定指数来衡量居群中基因型的实际频率是否偏离 Hardy-Weinberg 理论比例; 当  $F = 0$  时, 基因型频率符合 Hardy-Weinberg 理论比例; 当居群中纯合体过量时,  $F > 0$ ; 反之, 当杂合体过量时,  $F < 0$ 。

表2 美洲印第安人3个居群 MNSs 位点基因 ( $p_i$ ) 和纯合子 ( $P_i$ ) 频率及其  $F_{ISik}$  估计值<sup>[6]</sup>  
Table 2 Gene ( $p_i$ ) and homozygote ( $P_i$ ) frequencies for the MNSs locus in three populations of the Xavante American Indians and estimates of  $F_{ISik}$

Allele	D (79)*			M (287)			L (171)		
	p	P	$F_{ISik}$	p	P	$F_{ISik}$	p	P	$F_{ISik}$
MS	0.33	0.1392	0.1370	0.37	0.1812	0.1900	0.39	0.1637	0.0488
Ms	0.38	0.1646	0.0857	0.36	0.1185	-0.0482	0.48	0.2222	-0.0328
NS	0.09	0.0000	-0.0989	0.15	0.0174	-0.0400	0.01	0.0000	-0.0101
Ns	0.20	0.0380	-0.0125	0.12	0.0035	-0.1032	0.12	0.0176	0.0303
Weighted average			0.0578			0.0244			0.0108

\* Number of individuals examined.

表3 美洲印第安人 MNSs 位点的 F 统计<sup>[6]</sup>  
Table 3 F-statistics for the MNSs locus in the Xavante population of American Indians

Allele	Average gene frequency ( $\bar{p}$ )	$P^*$	$\bar{p}^2$	$\bar{p}^2 \dagger$	$F_{ISk}$	$F_{ITk}$	$F_{STk}$
MS	0.3633	0.1613	0.1320	0.1326	0.1244	0.1266	0.0026
Ms	0.4067	0.1684	0.1654	0.1681	0.0013	0.0124	0.0112
NS	0.0833	0.0058	0.0069	0.0102	-0.0602	-0.0144	0.0432
Ns	0.1467	0.0197	0.0215	0.0229	-0.0258	-0.0144	0.0112
Weighted average					0.0321	0.0436	0.0119

\* Average frequency of homozygotes over the three populations with an equal weight.

† Average of  $p^2$  over the three populations with an equal weight.

## 2 居群遗传分化的度量

不同居群或亚居群间遗传分化格局的研究是植物遗传多样性研究中的一个中心问题。对于等位基因酶资料, 有两种常用的居群分化指标, 即 Wright<sup>[7]</sup> 的  $F_{ST}$  统计量和 Nei<sup>[5]</sup> 基因多样性统计。

### 2.1 F 统计量法

Wright 在研究居群的遗传分化时, 介绍了下面公式<sup>[7]</sup>:  $1 - F_{IT} = (1 - F_{IS})(1 - F_{ST})$ ,  $F_{IS}$  和  $F_{IT}$  是分别测定亚居群和总居群中基因型的实际频率与 Hardy-Weinberg 理论期望比例的偏离程度, 而  $F_{ST}$  是衡量亚居群之间的分化程度。该方法是建立在固定指数 (F) 基础上的。

根据 Nei 发表的论文<sup>[5]</sup>,  $F_{IS}$ 、 $F_{IT}$  和  $F_{ST}$  的计算公式、步骤如下:

对每个亚居群而言  $F_{ISik}$  是第  $i$  亚居群中第  $k$  个等位基因的  $F_{IS}$ 。  $F_{ISi}$  是第  $i$  亚居群中一个位点的所有等位基因  $F_{IS}$  的加权平均值。

$$F_{ISk} = \frac{P_{ik} - p_{ik}^2}{p_{ik}(1 - p_{ik})}$$

$$F_{ISi} = \frac{\sum_k (P_{ik} - p_{ik}^2)}{\sum_k p_{ik}(1 - p_{ik})}$$

其中  $P_{ik}$  为第  $i$  亚居群中  $k$  基因纯合子的频率,  $p_{ik}$  为  $k$  基因的频率。

对总居群中每一个等位基因的  $F_{IS}$ 、 $F_{IT}$ 、 $F_{ST}$  而言,  $F_{ISk}$ 、 $F_{ITk}$ 、 $F_{STk}$  分别表示总居群中  $k$  等

$$\text{位基因的 } F_{IS}, F_{IT}, F_{ST}, F_{ISk} = \frac{P_k - \bar{p}_k^2}{\bar{p}_k - \bar{p}_k^2}, F_{ITk} = \frac{P_k - \bar{p}_k^2}{\bar{p}_k - \bar{p}_k^2}, F_{STk} = \frac{\bar{p}_k^2 - \bar{p}_k^2}{\bar{p}_k - \bar{p}_k^2},$$

这里  $P_k = \sum_i w_i P_{ik}$ ,  $\bar{p}_k = \sum_i w_i p_{ik}$ ,  $\bar{p}_k^2 = \sum_i w_i p_{ik}^2$ ,  $w_i$  是第  $i$  个亚居群的相对大小,  $\sum_i w_i = 1$ . 在实际计算中,  $w_i$  常被平均为  $1/s$ ,  $s$  为亚居群数目, 所以  $P_k$ 、 $\bar{p}_k^2$  都是等量加权.

对总居群中每个位点而言, 加权平均一个位点上所有等位基因  $F_{ISk}$ 、 $F_{ITk}$ 、 $F_{STk}$ , 即得到  $F_{IS}$ 、 $F_{IT}$  和  $F_{ST}$ .

$$F_{IS} = \frac{\sum_k (\bar{p}_k - \bar{p}_k^2) F_{ISk}}{\sum_k (\bar{p}_k - \bar{p}_k^2)}, F_{IT} = \frac{\sum_k (\bar{p}_k - \bar{p}_k^2) F_{ITk}}{\sum_k (\bar{p}_k - \bar{p}_k^2)}, F_{ST} = \frac{\sum_k (\bar{p}_k - \bar{p}_k^2) F_{STk}}{\sum_k (\bar{p}_k - \bar{p}_k^2)}.$$

对于多个位点而言, 最后计算各位点  $F_{IS}$ 、 $F_{IT}$  和  $F_{ST}$  的算术平均值.  $F_{ISk}$ 、 $F_{ISi}$ 、 $F_{ISk}$ 、 $F_{ITk}$ 、 $F_{STk}$  及  $F_{IS}$ 、 $F_{IT}$ 、 $F_{ST}$  具体数据见表 2、3.

如果所有居群都处于平衡状态, 则  $F_{IS} = 0$ ,  $F_{IT} = F_{ST}$ .

Nei 将  $F_{IS}$ 、 $F_{IT}$  和  $F_{ST}$  解释为观察到的和期望的杂合度的函数<sup>10</sup>. 它们相互之间的关系为:  $F_{IS} = (H_S - H_O)/H_S$ ,  $F_{IT} = (H_T - H_O)/H_T$ ,  $F_{ST} = (H_T - H_S)/H_T$ . 式中  $H_T$  为总居群的杂合度,  $H_S$  和  $H_O$  分别为期望的和观察到的杂合度.

这里,  $H_T = 1 - \sum \bar{p}_k^2$ ,  $H_S$  和  $H_O$  是所有亚居群  $H_{Si}$  和  $H_{Oi}$  的加权平均.

$$H_{Si} = 1 - \sum_k p_{ik}^2, H_{Oi} = 1 - \sum_k P_{ik}.$$

表 3 中, 观察到的杂合度  $H_O = 1 - \sum P = 1 - (0.1613 + 0.1684 + 0.0058 + 0.0197) = 0.6448$ , 期望杂合度  $H_S = 1 - \sum \bar{p}_k^2 = 1 - (0.1326 + 0.1681 + 0.0102 + 0.0229) = 0.6662$ , 总居群的杂合度  $H_T = 1 - \sum \bar{p}_k^2 = 1 - (0.1320 + 0.1654 + 0.0069 + 0.0215) = 0.6742$ .

所以  $F_{IS} = (0.6662 - 0.6448)/0.6662 = 0.0321$ ,  $F_{IT} = (0.6742 - 0.6448)/0.6742 = 0.0436$ ,  $F_{ST} = (0.6742 - 0.6662)/0.6742 = 0.0119$ . 这与按上面公式计算的  $F_{ISk}$ 、 $F_{ITk}$ 、 $F_{STk}$  加权平均后的结果相同. 由此可见, Nei 氏后一种方法比较简便.

## 2.2 基因多样性法

这种方法是 Nei<sup>11</sup> 提出的, 它是建立在分解基因多样性基础上的, 即将总居群的基因多样性 ( $H_T$ ) 分解为亚居群内的基因多样性 ( $H_S$ ) 和亚居群间的基因多样性 ( $D_{ST}$ ) 两部分, 从而求出亚居群内和亚居群间基因多样性的相对量. 对任一一位点均有:  $H_T = H_S + D_{ST}$ , 亚居群间基因分化的相对量可用下式求出:  $G_{ST} = D_{ST}/H_T$ , 由于  $G_{ST}$  可以反映居群分化的大小, 故 Nei 将  $G_{ST}$  称为基因分化系数. 该值从 0 到 1 不等. 当亚居群间没有分化时,  $G_{ST} = 0$ . 随着亚居群间分化程度的加大,  $G_{ST}$  值趋向于 1.

对于大量位点而言, 平均基因多样性  $\bar{H}_T$ 、 $\bar{H}_S$ 、 $\bar{D}_{ST}$  是通过对所有位点的  $H_T$ 、 $H_S$  和  $D_{ST}$  算术平均获得的. 计算  $H_S$  时, 应先计算各亚居群中的基因多样性  $He_i$ , 然后平均即得  $H_S$ .

$$He_i = 1 - \sum p_{ik}^2.$$

$H_T = 1 - \sum \bar{p}_k^2$ , 式中  $\bar{p}_k$  为第  $i$  个位点第  $k$  个等位基因在总居群的平均频率.

$D_{ST} = H_T - H_S$ . 将各位点的数据平均就可得到居群在全部位点上总的分化情况.

表 2、3 中,  $H_S$ 、 $H_T$ 、 $G_{ST}$  的计算如下:

居群 D 的  $H_e = 1 - (0.33^2 + 0.38^2 + 0.09^2 + 0.20^2) = 0.6986$ , 居群 M 的  $H_e = 1 - (0.37^2 + 0.36^2 + 0.15^2 + 0.12^2) = 0.6966$ , 居群 L 的  $H_e = 1 - (0.39^2 + 0.48^2 + 0.01^2 + 0.12^2) = 0.603$ . 3 个居群  $H_e$  的平均值即为  $H_s$ ,  $H_s = 0.6661$ .

$$H_T = 1 - \Sigma \bar{p}^2 = 1 - (0.3633^2 + 0.4067^2 + 0.0833^2 + 0.1467^2) = 0.6742.$$

$$D_{ST} = H_T - H_s = 0.6742 - 0.6661 = 0.0081.$$

$$G_{ST} = D_{ST} / H_T = 0.0081 / 0.6742 = 0.0120.$$

$G_{ST}$  与 Wright 的 F 统计中的  $F_{ST}$  相当。当一个位点只有两个等位基因时, Nei 的  $G_{ST}$  就与 Wright 的  $F_{ST}$  完全相等; 当有多个等位基因的情况下,  $G_{ST}$  等于所有等位基因  $F_{ST}$  的加权平均值, 因此  $G_{ST}$  可以看作  $F_{ST}$  的一个扩展。

通过以上参数, 我们就了解了居群内的遗传变异及居群间或亚居群间的遗传分化, 即居群的遗传多样性。最后可以利用居群内等位基因的频率, 计算出不同居群之间相互的遗传相似性系数(I)或遗传距离(D), 并且通过对两个指数的聚类分析, 绘出表明居群间亲缘关系的系统树<sup>[9]</sup>。

以上各公式的应用近十几年来在国际植物学杂志上经常见到, 如对北美鹅掌楸 (*Liriodendron tulipifera*)<sup>[10]</sup> 及喷瓜 (*Ecballium elaterium*)<sup>[11]</sup> 的遗传分析。

#### 参考文献

- 1 陈灵芝. 中国的生物多样性—现状及其保护对策. 北京: 科学出版社, 1993, 99—113
- 2 王中仁. 植物遗传多样性和系统学研究中的等位酶分析. 生物多样性, 1994, 2(1):38—43
- 3 胡志昂, 王洪新. 北京地区野大豆天然群体遗传结构. 植物学报, 1985, 27(6):599—604
- 4 梅里尔 D J. 黄瑞复等译. 生态遗传学. 北京: 科学出版社, 1991, 44—50
- 5 Nei M. Analysis of gene diversity in subdivided populations. Proc Nat Acad Sci USA, 1973, 70(12):3321—3323
- 6 葛颂. 用同工酶定量分析林木群体变异和分化的方法. 西南林学院学报, 1989, 9(1):84—90
- 7 Wright S. The genetical structure of populations. Annals of Eugenics, 1951, 15:323—354
- 8 Nei M. F-statistics and analysis of gene diversity in subdivided populations. Ann Hum Gent Lond, 1977, 41:225—233
- 9 根井正利. 王家玉译. 分子群体遗传学与进化论. 北京: 农业出版社, 1983, 134, 169—172, 191—196
- 10 Parks C R et al. The Significance of allozyme variation and introgression in the *Liriodendron tulipifera* complex (Magnoliaceae). Amer J Bot, 1994, 81(7):878—889
- 11 Costich D E, Meagher T R. Genetic variation in *Ecballium elaterium* (Cucurbitaceae): breeding system and geographic distribution. J Evol Biol, 1992, 5:589—601