

# 群体遗传学研究中的数据处理方法 I . RAPD 数据的 AMOVA 分析

张富民 葛 颂\*

(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)

**摘要:** 近年来 RAPD 数据和 AMOVA 分析广泛地应用于群体遗传学和保护遗传学研究。然而, 由于 RAPD 标记具显性特点, 加上目前进行 AMOVA 分析所依赖的 RAPDistance 软件不完善, 使得对 RAPD 数据进行 AMOVA 分析时存在许多不足。本文介绍了 AMOVA 分析的基本过程, 同时引入一个新的程序 DCFA 用以替代 RAPDistance, 并详述了将 DCFA 与 WINAMOVA 联用, 对 RAPD 数据进行 AMOVA 分析的具体步骤与注意事项。最后, 以产自中国和巴西 8 个普通野生稻 (*Oryza rufipogon*) 天然群体为例, 演示了对 RAPD 表型数据进行 AMOVA 分析的过程, 讨论了 AMOVA 分析结果在群体遗传结构上的意义。通过对 AMOVA 算法的分析, 同时比较 4 种距离系数所得 AMOVA 结果, 我们认为在进行 AMOVA 分析时选择 NEI-LI 距离和欧氏距离平方较为合适, 而目前国内使用较多的 JACCARD 系数不适合 AMOVA 分析。

**关键词:** 群体遗传结构, RAPDistance, DCFA, 进化距离

中图分类号: Q347

文献标识码: A

文章编号: 1005-0094(2002)04-0438-07

## Data analysis in population genetics. I. analysis of RAPD data with AMOVA

ZHANG Fu-Min, GE Song\*

Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093

**Abstract:** Random Amplified Polymorphic DNA (RAPD) data have been increasingly used in studies on population and conservation genetics. Analysis of molecular variance (AMOVA) has become one of the important methods utilized in analysis of population genetic structure. Currently, WINAMOVA is the most popular software that is used for AMOVA analysis, and often runs together with RAPDistance, a software to calculate genetic distances. However, cautions should be taken when AMOVA analysis is used to process the RAPD data because of the dominant characteristic of RAPD marker and the limitation of RAPDistance. In this paper, we briefly introduce the principle and algorithm of AMOVA analysis and describe a new program, DCFA, that substitutes for RAPDistance. We also illustrate the processes of running the programs DCFA and WINAMOVA. In addition, we analyze eight *Oryza rufipogon* populations as an example to demonstrate how to process RAPD data with AMOVA, and discuss the results in light of population genetic structure. Finally, based on the comparison of four commonly used distance coefficients, we suggest that the coefficients of Euclidean squared distance and NEI-LI rather than JACCARD are suitable for analysis of molecular variance.

**Key words:** population genetic structure, RAPDistance, DCFA, evolutionary distance

群体遗传结构是指遗传变异在物种或群体中的一种非随机分布,即遗传变异在群体内、群体间的分布样式以及在时间上的变化(Hamrick & Loveless, 1989)。这种非随机分布是由不同过程共同作用产生的,包括物种长期的进化历史(分布区改变、生境破碎、群体隔离)、突变、遗传漂变、繁育系统、基因流与选择(Slatkin, 1987;葛颂, 1997;Schaal *et al.*, 1998)。因此,要阐明生物进化的过程、式样和机理,必须首先探讨生物群体遗传变异大小、遗传结构及其变化规律以及影响群体遗传结构的各种因素(Hamrick, 1983;葛颂, 1994, 1997)。群体遗传结构的研究也是遗传资源利用和物种保护的基础(葛颂, 1997;李昂, 葛颂, 2002)。

在研究群体遗传结构时, Wright (1951, 1965) 提出的  $F$ -statistics 是广泛采用的模型(Weir, 1990; Rousset, 1996)。但 Wright (1951, 1965) 的  $F$ -statistics 在应用中有一些问题(Weir & Cockerham, 1984; Long, 1986) 如该模型是由共显性双等位基因位点推导而来、等级结构较少等。因此不同学者又在 Wright 的  $F$ -statistics 基础之上建立或完善了与之相关的研究群体遗传结构的各种  $F$ -statistics (Cockerham, 1969, 1973; Nei, 1977; Weir & Cockerham, 1984; Long, 1986; Hartl & Clark, 1997)。随着近些年来 DNA 分子数据在群体遗传学研究中日益受到重视,  $F$ -statistics 也被广泛地应用到各种 DNA 分子标记中, 如 SSR、ISSR、RAPD、AFLP 等, 但针对其中的显性标记如 ISSR、RAPD、AFLP,  $F$ -statistics 的应用则需要一定的前提与假设(Lynch & Milligan, 1994; Apostol *et al.*, 1996; 钱韦, 葛颂, 2001)。另外, 在  $F$ -statistics 中, 没有考虑等位基因(单倍型)之间的差异程度, 而这种差异实际上是分子进化的结果。为此, 一些对分子进化提出假设的方法被应用到群体遗传结构的研究中(Takahata & Palumbi, 1985; Lynch & Crease 1990), 但这些假设往往因具体的研究而不同。为此, Excoffier *et al.* (1992) 发展出了一种分子方差分析(Analysis of Molecular Variance, AMOVA)方法, 通过估计单倍型(含等位基因)或基因型之间的进化距离, 进行遗传变异的等级剖分, 并提出了与  $F$ -statistics 类似的  $\Phi$ -statistics 等方法来有效地度量亚群体的分化。AMOVA 方法被提出后, 很快就被用于 RAPD 表型数据的分析(Huff *et al.*, 1993)。由于 AMOVA 方法适用于所有类型的

遗传学数据, 充分考虑单倍型之间的趋异程度, 且可以在不需要假设的情况下直接对显性标记数据进行群体遗传结构的分析, 加上相应分析软件 WINAMOVA (Excoffier, 1993) 的应用, 所以使得各种单倍型和显性标记数据在群体遗传结构研究中得到了广泛的应用(Shim & Jorgensen, 2000; Hapke *et al.*, 2001; Parducci *et al.*, 2001), 尤其是对近年来在遗传多样性和群体遗传结构研究中大量应用的 RAPD、ISSR、AFLP 技术, AMOVA 方法受到广泛的欢迎(Ge *et al.*, 1999; Nebauer *et al.*, 1999; 钱韦, 葛颂, 2001; Li & Ge, 2001; Qian *et al.*, 2001; Roman *et al.*, 2001; Sales *et al.*, 2001)。

虽然使用 AMOVA 分析 RAPD 数据优势明显, 但在具体研究与操作过程中仍然存在如下问题: (1) 通常使用欧氏距离平方、NEI-LI 系数与 JACCARD 系数等来估计 RAPD 表型之间的进化距离, 但 Nybom & Bartish (2000) 根据已发表的文献, 认为使用这 3 种系数进行 AMOVA 分析所得结果相同, 事实上这一看法是有问题的。(2) AMOVA 分析的基础数据是各种距离矩阵, 而目前与软件 WINAMOVA 1.55 (Excoffier, 1993) 联用的 RAPDistance 软件(Armstrong *et al.*, 1996) 最新版本 V1.04 仅能处理 100 个样本、20 个引物、210 条带的 RAPD 数据, 操作格式繁琐(DOS 软件), 并非为 WINAMOVA 分析专门设计。(3) WINAMOVA 最多只能分析 255 个(种) RAPD 表型。虽然软件 Arlequin 2.000 (Schneider *et al.*, 2000) 能够突破这一限制, 处理对象自由度高达 1000, 但由于该软件的工程文件(project file) 编写比较复杂, 所以使用者很少。

针对以上这些问题, 本文从 AMOVA 分析的基本原理与算法入手, 介绍了我们自己开发的用以替代 RAPDistance 的 DCFA1.1 程序, 以及如何将其与软件 WINAMOVA 和 Arlequin 配套使用。同时, 我们用已发表的普通野生稻(*Oryza rufipogon*) RAPD 数据(Ge *et al.*, 1999) 为例, 探讨了使用不同距离系数进行 AMOVA 分析的差异。通过演示对 RAPD 数据进行 AMOVA 分析的过程, 解释了 AMOVA 分析结果的生物学含义, 并对开展此类分析时应注意的事项进行了讨论。

## 1 AMOVA 分析的基本原理与算法

在群体遗传学研究中处理的基本数据通常有两

种类型:一是单倍型(haplotype)数据,二是基因型(genotype)数据(Schneider *et al.*, 2000)。前者反映的是单倍染色体组中的遗传信息,如单个等位基因、cpDNA 与 mtDNA 单倍型等,而基因型数据指在二倍体中由不同等位基因组合出的基因型(如单位点基因型和多位点基因型),通常包括两种情形:一种可还原为单倍型,如杂合体可辨认的共显性标记数据,另一种只能得到对应于基因型的表型,如杂合体不可辨认的显性标记检测数据(如 RAPD、AFLP、ISSR 表型数据等)。AMOVA 分析的优势在于处理单倍型数据的同时,亦可对基因型数据进行有效处理。

### 1.1 基本原理

假定在一或多组群体中,第  $k$  组第  $j$  个群体的第  $i$  个单倍型(或基因型)的  $x_{ijk}$  频率向量可用一个线型可加模型假定:  $x_{ijk} = x + a_k + b_{jk} + c_{ijk}$ 。  $x$  是在整个研究中  $x_{ijk}$  的未知期望值,  $a$  是组(或地区)的效应,  $b$  是群体的效应,  $c$  是群体内单倍型(或基因型)的效应。这三个效应具可加性、随机性、独立性,且分别具方差  $s_a^2, s_b^2, s_c^2$ 。通过等级剖分(nested analysis of molecular variance)计算组间(Among Groups, 简称 AG)、组内/群体间(Among Populations/Within Groups, 简称 AP/WG)、群体内(Within Population, 简称 WP)方差组分  $\sigma_a^2, \sigma_b^2, \sigma_c^2$  的期望值,分别计算各个等级对总遗传变异  $\sigma^2$  ( $\sigma^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2$ ) 的贡献率。相应的  $\Phi$ -statistics 由以下公式计算:

$$\Phi_{ST} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma^2} \quad \Phi_{CT} = \frac{\sigma_a^2}{\sigma^2} \quad \Phi_{SC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2}$$

$\Phi_{ST}$ 、 $\Phi_{CT}$ 、 $\Phi_{SC}$  分别反映群体间、组间、群体内的遗传分化。

### 1.2 基本算法

#### 1.2.1 单倍型与基因型之间进化距离的估计

AMOVA 分析引入进化距离(evolutionary distance)来度量并计算单倍型(或基因型,下同)间的差方( $\delta^2$ ),十分巧妙地避开了分子数据不便于直接计算离差方的问题。所有种类的单倍型之间的差方组成一个距离矩阵。这是 AMOVA 分析的基础数据,也是 AMOVA 与 ANOVA 的主要区别。

#### 1.2.2 各等级方差计算

$G$  组  $P$  个群体  $N$  个单倍型的总离差方为:

$$SSD(TOTAL) = \frac{1}{2N} \sum_{j=1}^N \sum_{k=1}^N \delta_{jk}^2$$

由于单倍型的种类最多为  $N$ ,通常小于  $N$ ,所以在计算总离差方时有的单倍型种类之间的差方需引用 2 或 2 次以上。

相应可计算出:

$$SSD(WP) = \sum_{g=1}^G \sum_{i=1}^{I_g} \frac{\sum_{j=1}^{N_{ig}} \sum_{k=1}^{N_{ig}} \delta_{jk}^2}{2N_{ig}}$$

$I_g$  为第  $I$  组的群体数,  $N_{ig}$  为第  $g$  组第  $i$  个群体的单倍型数。

$$SSD(AP/WG) = \sum \left\{ \frac{\sum_{i=1}^{I_g} \sum_{j=1}^{N_{ig}} \sum_{k=1}^{N_{ig}} \delta_{jk}^2}{\sum_{i=1}^{I_g} 2N_{ig}} - \frac{\sum_{i=1}^{I_g} \sum_{j=1}^{N_{ig}} \delta_{jk}^2}{\sum_{i=1}^{I_g} 2N_{ig}} \right\}$$

$SSD(AG) = SSD(TOTAL) - (SSD(WP) + SSD(AP/WG))$

各等级方差计算  $s_a^2, s_b^2, s_c^2$ :

由于各等级自由度(degree of freedom, 简称为:  $d.f.$ )分别为:

$$d.f.(AG): G - 1; \quad d.f.(AP/WG): P - G;$$

$$d.f.(WP): N - P$$

所以:

$$s_a^2 = SSD(AG)/(G - 1), \quad s_b^2 = SSD(AP/WG)/(P - G), \quad s_c^2 = SSD(WP)/(N - P)$$

#### 1.2.3 等级剖分

对不等大小的样本,  $n$  系数代表特定等级样本的平均值,计算方法如下:

$$S_G = \sum_{g=1}^G \sum_{p=1}^p \frac{N_p^2}{N_g}; \quad n = \frac{N - S_G}{P - G}$$

$$S_G = \sum_{p=1}^p \frac{N_p^2}{N}; \quad n' = \frac{N - \sum_{g=1}^G \frac{N_g^2}{N}}{G - 1}; \quad n'' = \frac{N - \sum_{g=1}^G \frac{N_g^2}{N}}{G - 1}$$

$S_G$  为计算  $n$  系数的中间值,  $N_g$  为  $g$  组单倍型的总个数,  $N_p$  为  $p$  群体单倍型总个数。

$$\sigma_a^2 = n''s_a^2 + n's_b^2 + s_c^2 \quad \sigma_b^2 = n's_b^2 + s_c^2 \quad \sigma_c^2 = s_c^2$$

然后计算各个等级对总遗传变异  $\sigma^2$  ( $\sigma^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2$ ) 的贡献率,得出遗传变异在组(地区)间、组内/群体间、群体内的分布。

#### 1.2.4 计算 $\Phi$ -statistics

若为来自二倍体核 DNA 的单倍型(等位基因)数据,则可计算  $\Phi$ -statistics,用来度量亚群体的分化。

#### 1.2.5 统计检验

为了检验群体各等级对遗传变异影响的显著水平,AMOVA 通过以下方法进行统计检验:

通过在组间、群体间交换单倍型,检验  $\sigma_c^2$  与  $\Phi_{ST}$ ;通过在组内群体间交换单倍型,检验  $\sigma_b^2$  与  $\Phi_{SC}$ ;通过在组间交换群体,检验  $\sigma_a^2$  与  $\Phi_{CT}$ 。据此来推断群体各等级(组间、组内/群体间、群体内)对遗传分化影响的显著性。

### 1.2.6 群体间的遗传距离

若为来自二倍体核 DNA 的单倍型(等位基因)数据,还可通过以下两个公式计算两两群体间的  $\Phi_{ST}$ ,作为群体间遗传距离的度量:

$$D = -\log(1 - \Phi_{ST}) \quad (\text{Reynolds } et al., 1983)$$

$$D = \Phi_{ST} / (1 - \Phi_{ST}) \quad (\text{Slatkin}, 1995)$$

目前,人们对等位基因数据的处理通常用 Nei's (1972, 1978) 遗传距离。当得到等位基因的序列数据时,以上两个公式更适用于估计两两群体间的遗传距离。

## 2 AMOVA 分析 RAPD 数据的具体过程

由于 RAPD 数据的显性特点,在研究群体遗传结构时,往往需先将其还原为单倍型,再进行各种  $F$ -statistics 分析,如常用软件 BIOSYS-1 (Swofford & Selander, 1981) 或 POPGENE (Yeh *et al.*, 1997)。AMOVA 可直接对 RAPD 表型数据进行分析 (Huff *et al.*, 1993; Ge *et al.*, 1999; 钱韦,葛颂, 2001), 剖分群体 RAPD 表型变异,不依赖于各种假设条件。下面介绍如何用 DCFA1.1 与 WINAMOVA1.55 来对 RAPD 数据进行 AMOVA 分析。

### 2.1 RAPD 数据的编码和运算

假定每一条 RAPD 电泳谱带代表一个具双等位基因的位点 (Williams *et al.*, 1990)。在此位点上,有带记为 1 (存在),无带记为 0 (不存在)。所有个体在全部位点上的谱带式样 (RAPD 表型) 组成一个 (0 或 1) 矩阵,若有个体在全部位点上谱带式样相同,则按相同 RAPD 表型对待,只保留一个。本文将这样组成的矩阵称为运算矩阵。在得到 RAPD 原始数据后,运行 WINAMOVA 还需完成下列主要步骤 (参见表 1)。

### 2.2 距离矩阵与软件 DCFA1.1

目前进行 WINAMOVA 分析的常用软件是 WINAMOVA1.55 (Excoffier, 1993), 该软件分析是建立在距离文件 (文件扩展名 .dis) 基础上的。由于目前与 WINAMOVA 联用的 RAPDistance 软件 (Armstrong *et al.*, 1996) 操作格式繁琐 (DOS 环境), 更重要的是其运算矩阵不能大于 100 (样本)  $\times$  210 (条带), 这样小的数据容量显然制约了许多群体遗传学数据的分析,在具体使用中很不理想。为此,本文作者用 Visual C++ (MFC) 开发了运行于 MS Windows 操作系统的 DCFA1.1, 可按要求鉴定相同 RAPD 表型,并组成运算矩阵,给出 RAPD 表型的频数,计算 16 种距离系数。该程序界面友好,对运算矩阵没有大小限制,能自动为单倍型或基因型加上标识符,输出的结果文件可直接联用于 WINAMOVA 软件分析。但是,应该说明的是,WINAMOVA1.55 对单倍型或 RAPD 表型有限制,不能多于 255 个 (种)。对运算矩阵用 DCFA1.1 可计算 RAPD 表

表 1 WINAMOVA1.55 数据文件示例(下划线部分为注释部分,可自由加注)

Table 1 Examples of the files for program WINAMOVA1.55. The underlined words can be edited without limitation

文件类型 File type	距离文件 Distance file	组文件 Group file	群体文件 Population file
文件名 File name	<u>rice</u> .dis	<u>rice</u> .grp	<u>rice1</u> .pop
编写格式 Format	1 2 3 4 5 0.00 2.00 0.00 1.00 3.00 0.00 1.00 1.00 2.00 0.00 1.00 3.00 2.00 2.00 0.00	2 <u>groups</u> 1 2 3 4 5 6 7 8	11 <u>individuals</u> 1 4 3 5 4 2
说明 Interpretation	5 个 RAPD 表型间的距离矩阵 The matrix of distances among 5 RAPD phenotypes	8 个群体分为 2 组,第 1 组含群体 1、2、3 与 4;其余为第 2 组 Eight populations are divided into 2 groups. Group 1 includes populations 1, 2, 3 and 4. Group 2 includes the remaining populations	该群体有第 1、第 3 和第 4 号 RAPD 表型,分别有 4 个、5 个和 2 个个体 This population has RAPD phenotypes 1, 3, 4, and their frequencies are 4, 5 and 2, respectively

型间的距离系数,组成距离系数( $\delta^2$ )矩阵文件,即 WINAMOVA 所要的距离文件(.dis)。实际上,一个位点任意两个 RAPD 表型之间的差异可看做 4 个单倍型的遗传差异的和效应(Stewart & Excoffier, 1996)。另外,还须在距离文件的第一行加上每个基因型的代号,通常按数字顺序排列,DCFA 可按要求自动完成此项操作(表 1)。

### 2.3 组文件和群体文件

WINAMOVA 的组文件(文件扩展名 .grp)是用来把群体按研究目的分成组。通常,群体代号从 1 开始依次用数字表示。在组文件的第一行首先给出组数,其后各行填上每组里的群体代号,注意代号之间用空格隔开(表 1)。有了组文件,还需要编写群体文件(文件扩展名 .pop),所有群体文件名前应保持一致,最好用字母,后用群体数字代号。如:rice1.pop, rice2.pop, .....。每个群体文件的第一行首先给出群体个体数,随后每行给出本群体出现的 RAPD 表型代号与频数。注意 RAPD 表型代号与频数之间用空格分开。

在我们开发的 DCFA1.1 软件可按要求自动为研究者编写组文件与群体文件。

### 2.4 WINAMOVA 软件的运行

如无特殊需要,所有的文件名最好用传统的 8.3 格式,尽量不要用 WINDOWS 的长文件名。尽管文件扩展名不是 .txt,但所有的文件都必须是纯文本格式。在上述工作完成后,即可运行软件 WINAMOVA,在“File”菜单项里选“Select input files”,然后分别填上距离文件名,组文件名,群体文件名。注意群体文件名仅填前边部分,代号部分不填,以代表一组文件。如表 1 例中,填“rice”而不要填“rice1”。然后在“Setting”菜单项里选择要进行各项运算及统计检验的运算次数(Number of permutations)。最后点“Go”菜单项进行运算。由于 WINAMOVA1.55 的准备工作的繁琐,易出错,当“Go”以后,如果提示某一个文件出错,请仔细按本文所述进行核对,确保文件空白处是空格(space)符号等。如果使用 DCFA1.0 自动编写组文件与群体文件,则可最大限度地避免这些问题。

程序 WINAMOVA 最多只能处理 255 个 RAPD 表型,当多于此数时,研究者可以使用软件 Arlequin2.000。但该软件的使用比较复杂,为方便起见,DCFA1.1 被设计成自动计算与编写能用于该软

件分析的距离文件与工程文件。另外,DCFA1.1 还可按要求为研究者计算群体条带频率与不同标准的多态条带比率(PPB),以及编写用于 POPGENE 分析的 RAPD 数据文件。

## 3 实例

以 *Oryza rufipogon* 自然群体的 RAPD 数据(Ge et al., 1999)为例,演示用 WINAMOVA 软件对 RAPD 表型数据进行剖分的过程,同时比较用不同距离系数估计进化距离所得 AMOVA 分析的结果。

所分析的 8 个 *Oryza rufipogon* 自然群体分别来自中国与巴西,各 4 个群体。8 个群体共 80 个个体,20 个 RAPD 引物,共选取稳定可靠(一般取强带,弱带舍去)的 95 个位点(条带),组成 80×95 大小的个体×位点 RAPD 表型矩阵,由于 2 对个体的 RAPD 表型完全一样,各合并为一个 RAPD 表型,即共有 78 个 RAPD 表型,组成 78×95 大小的 RAPD 表型×位点 RAPD 表型矩阵(运算矩阵)。

对运算矩阵分别用 RAPDistance(Armstrong et al., 1996)和 DCFA1.1(张富民, 2001<sup>①</sup>)计算 RAPD 表型间的欧氏距离平方、欧氏距离、JACCARD 系数与 Nei-Li 系数,由此得到的 4 个 78×78 距离系数( $\delta^2$ )矩阵文件,即距离文件(rice.dis),两个软件所得距离系数矩阵完全一致。8 个群体按中国和巴西 2 个地区分成两组,编写组文件(rice.grp)。以 rice1.pop, rice2.pop, rice3.pop, rice4.pop, rice5.pop, rice6.pop, rice7.pop, rice8.pop 为文件名,分别按表 1 编写 8 个群体文件。然后,将距离系数矩阵与软件 WINAMOVA 联合进行等级剖分,统计检验运算 1000 次。结果见表 2。如有必要,可分别组织数据对各地区内及两两群体进行剖分。

从表 2 看出,不论用哪种距离系数来估计 RAPD 表型的遗传差异,各个等级都对遗传差异有极显著影响( $P < 0.001$ )。但不同系数的剖分结果不同,欧氏距离平方与 NEI-LI 系数的剖分结果几乎相等,61.5% 的遗传差异是由中国与巴西两个地区间的差异造成的。在中国与巴西地区内部,群体间平均贡献的遗传差异占 15%,群体内平均贡献占 23.5%。换句话说,*Oryza rufipogon* 在中国和巴西

①张富民, 2001. DEFA1.0, 一个与 AMOVA 联用的计算机距离矩阵的计算机程序。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

表 2 利用 4 个不同距离矩阵对基因型进行等级剖分

Table 2 Partition of genetic variance with four different distance matrices

	欧氏距离平方 Euclidean squared distance				欧氏距离 Euclidean distance		
	P	方差期望	百分比		P	方差期望	百分比
组间( AG )	13.28	61.80%	<0.001	组间( AG )	1.269	38.75%	<0.001
组内/群体间( AP/WG )	3.19	14.86%	<0.001	组内/群体间( AP/WG )	0.489	14.93%	<0.001
群体内( WP )	5.02	23.34%	<0.001	群体内( WP )	1.517	46.32%	<0.001
	NEI-LI 系数 NEI-LI coefficient				JACCARD 系数 JACCARD coefficient		
	P	方差期望	百分比		P	方差期望	百分比
组间( AG )	0.1215	61.56%	<0.001	组间( AG )	0.1519	53.81%	<0.001
组内/群体间( AP/WG )	0.0290	14.74%	<0.001	组内/群体间( AP/WG )	0.0467	16.57%	<0.001
群体内( WP )	0.0467	23.70%	<0.001	群体内( WP )	0.0836	29.61%	<0.001

注:AG: Among Groups, AP/WG: Among populations/Within Groups, WP: Within Population

之间已经出现了较大的分化,而在各地区内群体间的分化不大,小于群体内的遗传变异( Ge *et al.* 1999)。欧氏距离、JACCARD 系数的剖分结果显示,遗传变异在地区间较小,群体内较大,群体间影响不大。

欧氏距离平方事实上就是两个基因型的差的平方,从方差的定义出发,显然更适合用其计算离差方和。欧氏距离是欧氏距离平方的算术平方根,会减小基因型的差距。NEI-LI 系数将任意两个基因型放在总体的尺度上进行比较,而 JACCARD 系数只在任意两个基因型内部进行比较。显然 JACCARD 会得出较大的相似性,相反则会使基因型的差距变小,除非任意两个基因型具有总体所有的位点(谱带)。因此, JACCARD 系数显然是不适合用来做进化距离的估计,从而不适合做 AMOVA 分析,故 Excoffier (1993)在 WINAMOVA 软件中推荐使用欧氏距离平方。

#### 4 结论

AMOVA 分析从单倍型(基因型)的分化程度出发研究群体的遗传结构,使许多原来不便于研究群体遗传分化的数据,可以很方便地进行此类分析。同时,在估计单倍型(基因型)之间的进化距离时,每个研究者都可从具体的研究实例出发,进行自己的计算和矫正,甚至提出自己合理的分子进化假设,显示了 AMOVA 具有包容性和实用性。实际上, AMOVA 分析更适用于单倍型序列数据,尤其是等位基因序列数据( Excoffier *et al.*, 1992)。只不过目前此类数据的实验成本太高、难以找到合适的标记,因而目前应用较少。而在 RAPD、AFLP、ISSR 等数据的分析中, AMOVA 却得到广泛应用( Ge *et al.* 1999; Nebauer *et al.*, 1999; Shim & Jorgensen, 2000;

Li & Ge, 2001; Qian *et al.*, 2001; Roman *et al.*, 2001; Sales *et al.*, 2001; 钱韦, 葛颂, 2001)。尽管国内已有不少工作利用 RAPD 等分子标记进行遗传多样性和群体遗传结构的研究,但如何充分有效地处理 RAPD 数据,仍是值得注意的问题。我们认为 AMOVA 分析在进行此类分析时十分有效,是一种值得大力提倡的方法。

#### 参考文献

- 葛颂, 1994. 同工酶和植物进化生物学研究. 见: 陈家宽, 杨继(编著), 植物进化生物学. 武汉: 武汉大学出版社, 153~208
- 葛颂, 1997. 植物群体遗传结构研究的回顾和展望. 见: 李承森(主编), 植物科学进展(第一卷). 北京: 高等教育出版社, 1~15
- 李昂, 葛颂, 2002. 植物保护遗传学研究进展. 生物多样性, **10**(1): 61~71
- 钱韦, 葛颂, 2001. 居群遗传结构研究中显性标记数据分析方法初探. 遗传学报, **28**: 244~255
- Apostol B, W C Black IV, P Reiter and B R Miller, 1996. Population genetics with RAPD-PCR markers: the breeding structure of *Aedes aegypti* in Puerto Rico. *Heredity*, **76**: 325~334
- Armstrong J, A Gibbs, R Peakall and G Weiler, 1996. RAPD-Distance programs: version 1.04 for the analysis of patterns of RAPD fragments. Australian National University, Canberra, Australia
- Cockerham C C, 1969. Variance of gene frequencies. *Evolution*, **23**: 72~84
- Cockerham C C, 1973. Analysis of gene frequencies. *Genetics*, **84**: 679~700
- Excoffier L, 1993. Analysis of molecular variance (AMOVA) version 1.55. Genetics and Biometry Laboratory, University of Geneva, Switzerland
- Excoffier L, P E Smouse and J M Quattro, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics*, **131**: 479~491
- Ge S, G C X Oliveira, B A Schaal, L-Z Gao and D-Y Hong, 1999. RAPD variation within and between natural populations of the wild rice *Oryza rufipogon* from China and

- Brazil. *Heredity*, **82**: 638 ~ 644
- Hamrick J L and M D Loveless, 1989. Associations between the breeding system and the genetic structure of tropical tree populations. In: J Bock and Y B Linhart (eds.), *Evolutionary Ecology of Plants*. Westview Press, Boulder, 129 ~ 146
- Hamrick J L, 1983. Plant population genetics and evolution. *American Journal of Botany*, **69**: 1685 ~ 1693
- Hapke A, D Zinner and H Zischler, 2001. Mitochondrial DNA variation in Eritrean hamadryas baboons (*Papio hamadryas hamadryas*): life history influences population genetic structure. *Behavioral Ecology and Sociobiology*, **50**: 483 ~ 492
- Hartl D L and A G Clark, 1997. Principles of Population Genetics. Sunderland, Sinauer Associates, Inc.
- Huff D R, R Peakall and P E Smouse, 1993. RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloe dactyloides* (Nutt.)]. *Theoretical and Applied Genetics*, **86**: 689 ~ 696
- Li A and S Ge, 2001. Genetic variation and clonal diversity of *Psammochloa villosa* detected by ISSR markers. *Annals of Botany*, **87**: 585 ~ 590
- Long J C, 1986. The allelic correlation structure of Gainj- and Kalam-speaking people. I. The estimation and interpretation of Wright's *F*-statistics. *Genetics*, **112**: 629 ~ 647
- Lynch M and B G Milligan, 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**: 91 ~ 99
- Lynch M and T J Crease, 1990. The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution*, **7**: 377 ~ 394
- Nebauer S G, L del Castillo-Agudo and J Segura, 1999. RAPD variation within and among natural populations of outcrossing willow-leaved foxglove (*Digitalis obscura* L.). *Theoretical and Applied Genetics*, **98**: 985 ~ 994
- Nei M, 1972. Genetic distance between populations. *American Naturalist*, **6**: 283 ~ 293
- Nei M, 1977. *F*-statistics and the analysis of gene diversity in subdivided populations. *Annals of Human Genetics*, **41**: 225 ~ 233
- Nei M, 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**: 583 ~ 590
- Nybom H and Igor V Bartish, 2000. Effect of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspectives in Plant Ecology, Evolution and Systematics*, **3**: 93 ~ 114
- Parducci L, A E Szmidi, A Madaghiale, M Anzidei and G G Vendramin, 2001. Genetic variation at chloroplast microsatellites (cpSSRs) in *Abies nebrodensis* (Lojac.) Mattei and three neighboring *Abies* species. *Theoretical and Applied Genetics*, **102**: 733 ~ 740
- Qian W, S Ge and D-Y Hong, 2001. Genetic variation within and among populations of a wild rice *Oryza granulata* from China detected by RAPD and ISSR markers. *Theoretical and Applied Genetics*, **102**: 440 ~ 449
- Reynolds J, B S Weir and C C Cockerham, 1983. Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, **105**: 767 ~ 779
- Roman B, D Rubiales, A M Torres, J L Cubero and Z Satovic, 2001. Genetic diversity in *Orobanche crenata* populations from southern Spain. *Theoretical and Applied Genetics*, **103**: 1108 ~ 1114
- Rousset R, 1996. Equilibrium of measure of population subdivision for stepwise mutation process. *Genetics*, **142**: 1357 ~ 1362
- Sales E, S G Nebauer, M Mus and J Segura, 2001. Population genetic study in the Balearic endemic plant species *Digitalis minor* (Scrophulariaceae) using RAPD markers. *American Journal of Botany*, **88**: 1750 ~ 1759
- Schaal B A, D A Hayworth, K M Olsen, J T Rauscher and W A Smith, 1998. Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7**: 465 ~ 474
- Schneider S, D Roessli and L Excoffier, 2000. Arlequin: a software for population genetics data analysis. Ver. 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva
- Shim S I and R B Jorgensen, 2000. Genetic structure in cultivated and wild carrots (*Daucus carota* L.) revealed by AFLP analysis. *Theoretical and Applied Genetics*, **101**: 227 ~ 233
- Slatkin M, 1987. Gene flow and the geographic structure of populations. *Science*, **236**: 787 ~ 792
- Slatkin M, 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**: 457 ~ 462
- Stewart C N and L Excoffier, 1996. Assessing population genetic structure and variability with RAPD data: application to *Vaccinium macrocarpon* (American cranberry). *Journal of Evolutionary Biology*, **9**: 153 ~ 171
- Swofford D L and R B Selander, 1981. BIOSYS-1: a FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *Journal of Heredity*, 1981, **72**: 281 ~ 283
- Takahata N and S R Palumbi, 1985. Extranuclear differentiation and gene flow in the finite island model. *Genetics*, **109**: 441 ~ 457
- Weir B S and C C Cockerham, 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**: 1358 ~ 1370
- Weir B S, 1990. Genetics Data Analysis. Sunderland, Sinauer.
- Williams J G K, A R Kubelik, K J Livak, J A Rafalski and S V Tingley, 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, **18**: 6531 ~ 6535
- Wright S, 1951. The genetical structure of populations. *Annals of Eugenics*, **15**: 323 ~ 334
- Wright S, 1965. The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*, **19**: 395 ~ 420
- Yeh F C, R C Yang, T B J Boyle, Z H Ye and J X Mao, 1997. POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, Alberta, Canada