

聚类分析的理论研究及 其在植物育种中的应用(综述)

李永忠

(吉林省农科院大豆所)

聚类分析是各个学科领域广泛应用的一种多元统计分析方法,是随着近十几年电子计算机技术的发展,逐渐从数值分类中发展而形成的一个分枝。不少学者进行了在植物的遗传育种中的应用研究。

进行聚类分析的统计量有多种,常用的大致为两类:一类是相似系数(包括夹角余弦和相关系数等);一类是距离(包括有欧氏距离和马氏距离等)。由于Pearson的相似系数只适用于性状是独立的情况,马氏广义距离 D^2 适用于相关性状而在生物界得到了广泛应用。

马氏距离 $D_{ij}^2 = (\bar{X}_i - \bar{X}_j)' \Sigma^{-1} (\bar{X}_i - \bar{X}_j)$, 其中 $X_i = (\bar{X}_{i1}, \bar{X}_{i2}, \dots, \bar{X}_{in})'$, $X_j = (\bar{X}_{j1}, \bar{X}_{j2}, \dots, \bar{X}_{jn})'$, 这里 n 为性状数, Σ 为多元方差分析表中误差项期望协方差阵的估计值。当 Σ 为单位矩阵时,就简化为普通的欧氏距离。由于性状间往往有相关存在,所以性状的协差阵不可能为单位矩阵,因而有必要将性状值转变为标准化的新变数,使新变数的方差为1,协方差为0,这样新变数的协方差阵 Σ 为单位矩阵,即

$D_{ij}^2 = (\bar{Y}_i - \bar{Y}_j)' (\bar{Y}_i - \bar{Y}_j)$, 其中 $\bar{Y}_i = (\bar{y}_{i1}, \bar{y}_{i2}, \dots, \bar{y}_{in})'$, $\bar{Y}_j = (\bar{y}_{j1}, \bar{y}_{j2}, \dots, \bar{y}_{jn})'$, 这里 n 为性状数。于是只要计算新变数的欧氏距离即可。在此基础上进行分类,这就是Rao(1952)的中枢压缩法(Pivotal condensation)。国外的研究者基本上都是采用这种方法进行聚类分析,探讨遗传差异。

Chandra⁽⁷⁾, 对马氏距离的方法和Metroglyph技术表示的遗传差异进行了比较,结果二者具有显著的一致性。并建议在分类种质比较多时,用马氏距离进行分类之前,先用Metroglyph技术进行初步分类。

国内刘来福提出了以分类单位的遗传型协方差阵,采用标准化的主成分进行转换。因为主成分是独立的,相互间无关,标准化使其方差为1,这样标准化主成分协方差阵也是单位矩阵。所以,也只要计算每两个品种的标准主成分之间的欧氏距离即可。国内的研究者除了何国浩⁽⁸⁾采用Rao法外,基本为这种方法。

刘垂珏⁽¹⁾进一步研究了多数量性状遗传分析的数据结构指出,利用遗传协方差阵和遗传相关阵会给多元统计方法应用于多数量性状的遗传分析造成带有根本性的困难。可能出现遗传相关系数的绝对值大于1,使相关阵失去正定性,无法再进行分析,更易出现生物学上无法解释的负特征根和关联阵特征根大于1,并建议用品种相关阵代替遗传相关阵。

另外,似乎遗传相关阵也不能很好地反映性状间的相关性。我们知道,遗传相关系数

是标准化的遗传协方差, $r_g = \text{CoVg}_{12} / (\sigma_{g_1} \cdot \sigma_{g_2})$, 而 $\text{CoVg} = \text{CoVp} - \text{CoVe}$, $\sigma_g = h \cdot \sigma_p$, 所以 $r_g = (\text{CoVp} - \text{CoVe}) / (h_1 \cdot h_2 \cdot \sigma_{p_1} \cdot \sigma_{p_2})$ 。由此可见, r_g 的绝对值不但受表型方差和协方差的影响, 而且受两个有关性状的遗传力的强烈影响。Robertson⁽¹⁰⁾ 在假定遗传相关系数的抽样分布为正态的条件下, 从数理统计上推导出了由方差—协方差分析法所求的遗传相关系数的估计标准误公式。他认为, 当遗传相关系数不太大, 表型相关和遗传相关系数很相似时, 遗传相关系数的估计标准误近似于

$$\sigma_{r_g} = (1 - r_g^2) \{ \sigma_{h_1}^2 \cdot \sigma_{h_2}^2 / (2 \cdot h_1 \cdot h_2) \}^{1/2}$$

这里可以看出, 即使遗传相关系数的抽样分布服从正态, 那么在一个固定的试验里, 具有不同遗传力的相同的遗传相关系数, 自由度相等, 其 σ_{r_g} 也不会相同。因而虽绝对值相同, 但所达到的显著水平却不一定相同。经抽样的研究表明, 一般遗传力低于 $2/5$ 时, 遗传相关系数的抽样分布已经不符合正态, 这就造成了遗传相关系数的不可比性。只有遗传力大于 $2/5$, 且两个 r_g 各自所涉及的两个遗传力相等, 自由度也相同时, 其 σ_{r_g} 才唯一地受 r_g 绝对值的影响, 二者方可凭绝对值的大小来比较。显然这种条件在生物界是很难满足的。因此, 遗传相关系数绝对值的差异, 并不能很好地反映性状间相关程度的差异。绝对值差异较大的两个遗传相关系数由于遗传力的差异, 达到的显著水平可能相差无几, 大小相等的两个遗传相关系数也可能分别达到相当不同的显著水平。因而, 标准化的遗传协方差阵并不能很好地反映各个数量性状之间相互关系的本质。以其为基础所进行的聚类分析势必还缺乏理论依据。品种相关系数在一定程度上可以反映基因型间的联系, 又具有简单相关系数的性质, 彼此间具有可比性, 又可以避免可能出现的遗传相关矩阵非正定的异常情况以及生物学上无法解释的负特征根。因而, 用品种相关阵代之, 可能更符合实际情况。

Bhatt⁽⁹⁾ 以小麦为材料, 对自花授粉作物选配杂交亲本的几种方法进行研究, 结果表明, 以多元分析法选择那些遗传差异较大的亲本杂交更有效。Arunachalum 进一步研究植物育种中马氏距离 D^2 的应用指出, 用遗传距离表示的遗传差异可能在遗传育种中得到广泛应用。毛盛贤在冬小麦上所进行的研究也说明, 利用多元分析法测定的材料间的遗传距离, 可以作为选配杂交亲本的一个依据。

Arunachalum⁽¹¹⁾ 在花生上研究了杂种优势和特殊配合力与遗传差异的关系。通过分析将双列杂交的亲本分成四个差异组, 中间组间的材料杂交, 其产量及其构成因素的杂种优势明显高于两边差异组, 说明并非遗传差异越大, 杂种优势也越大。其特殊配合力也是如此。Ghaderi⁽⁸⁾ 在食用豆和菜豆上的研究表明, 食用豆产量因素的每株荚数和每荚粒数的杂种优势与亲本间的遗传距离呈明显的正相关, 而百粒重则无相关; 菜豆只有叶重的杂种优势与之呈显著的正相关, 其余不显著。徐静斐⁽³⁾ 在水稻上的研究也表明, 遗传距离与杂种优势呈明显的正相关, 李成荃、赵安常在梗稻和籼稻上的研究也得到了同样的结果。

Singh⁽¹²⁾, Katiyar, Singh, Maurya, Asthana, Singh, Jrehan, Joshi, Angadi, Kalloo, Amalray, Rao, Sidhu, Chaudhary, Jain, 赖仲铭, 潘佩萱等人

分别对珍珠粟、鹰嘴豆、芥菜、水稻、亚麻、大麦、芝麻、豇豆、黄瓜、棉花、红花、马铃薯、串豆、谷子、玉米自交系和小麦等进行了遗传距离测定的聚类分析。结果都得出，地理上的远缘与遗传差异并没有必然的联系。其原因是近代国家和地区之间品种的互换；另一方面由于遗传漂移，不同方向的人工选择下，使相邻地区的材料可能造成很大的遗传差异。但Rao⁽⁸⁾和Bhutani在水稻和马铃薯上的研究得出与之不同的结论。何国浩⁽²⁾对江淮下游地区大豆地方品种的聚类分析也得到同样的结果，并认为主要原因是大豆对光照很敏感，同时由于大豆起源于我国，这些地方品种都是在一定的自然环境条件下，通过长期的自然与人工选择获得的，这就在一定区域内形成相应的生态型。

Abou—Ei—Fittouh, Compbell, Narayan, Singh⁽¹¹⁾, Bainiwal在棉花、小麦、玉米、鹰嘴豆和鸽豆作物上，研究了基因型与环境互作及其适应性问题。他们认为，在不同环境条件下所表现出的遗传差异的不同，主要是基因型与环境互作产生的。而且Singh进一步指出，研究遗传差异应在比较正常的环境条件下（即较肥沃土壤）进行。此外，Broich⁽⁶⁾和Liany借助于此方法分别研究了大豆和高粱的种间关系。

陈洪文和吕景良分别研究了春小麦和春大豆数量性状的遗传距离及其系谱，探讨了遗传距离在杂交育种和种质资源研究上的应用。结果指出，主成分分析是评价品种综合性状优劣，鉴定优良种质的有效方法，遗传距离及其聚类分析的方法对杂交育种和种质资源的研究具有一定的实用价值。

以上研究者基本上是采用品种或纯系进行研究的。Tikka⁽³⁾和Kumar则分别研究了豇豆和珍珠粟分离群体的遗传差异。他们认为，对于遗传差异比较大的组合或亲本，早期世代进行互交，这样更有利于基因的重组和产生较大的杂种优势。

综上所述，多元分析的方法正在作物的遗传育种上进行应用研究。随着育种水平的提高，更需要有突破性的杂交组合出现。因而，有必要研究优良亲本的综合农艺性状，进一步探讨用 D^2 表示的亲本间的遗传差异与育种实践（包括非杂种优势利用的作物）的关系。由于这方面的研究尚少，一些非杂种优势利用的自花授粉作物的杂交育种实践证明，多以优×优组合为宜。因而，尚不能肯定遗传距离理论在作物育种中对亲本选配的指导作用，需进一步研究。

参 考 文 献

- [1] 刘垂环：1984，《安徽农学院学报》，(1)，1—5。
- [2] 何国浩等：1985，《大豆科学》，2(4)：253—265。
- [3] 徐静斐等：1981，《安徽农业科学》，水稻数量遗传论文专辑：55—71。
- [4] Arunachalum, V. : 1984, *Euphytica*, (33): 33—39.
- [5] Bhatt, C. M. : 1973, *Aust. J. Agr. Res.*, 24: 257—264.
- [6] Broich, S. et al: 1980, *Euphytica*, 29: 23—32.
- [7] Chandra, S. : 1977, *Euphytica*, 26: 141—148.
- [8] Ghaderi, A. et al: 1984, *Crop Sci.*, 24(1): 37—42.
- [9] Rao, A. V. et al: 1981, *Indian J. Genet.*, 41(2): 179—185.
- [10] Robertson, A. : 1969, *Biometrics*, 15: 469—485.
- [11] Singh, S. et al: 1979, *Indian J. Genet.*, 39(2): 210—215.
- [12] Singh, Y. P. et al: 1981, *Indian J. Genet.*, 41(2): 186—190.
- [13] Trehan, K. B. et al: 1975, *Indian J. agric. Sci.*, 44(2): 203—212.