



## 保留特殊种质材料的核心库构建方法

胡晋 徐海明 朱军

(浙江大学 种子科学中心, 浙江 杭州 310029)

**摘要:** 本研究提出了能保留特殊种质材料的多次聚类构建种质资源核心库方法. 对种质材料的表现型数据采用合适的遗传模型及混合线性模型统计分析方法无偏预测基因型值, 用基因型值构建核心库. 计算材料间的马氏距离, 用不加权类平均法进行聚类. 根据聚类图选取材料构建核心库时, 优先保留特殊遗传材料. 用方差  $F$  测验、均值  $t$  测验、极差比和变异系数评价核心库代表原有种质资源群体遗传多样性的程度. 以 168 个棉花基因型的 5 个纤维性状构建核心库.

**关键词:** 基因型值; 核心库; 种质; 多次聚类; 棉花

**中图分类号:** S32

**MR 分类号:** 62H25; 92B05

**文献标识码:** A

**文章编号:** 1001-9626(2001)03-0348-05

### 0 引言

目前, 种质资源库内巨大的遗传资源材料成了种质库管理, 材料评价、整理、分发、繁殖和利用的限制因素. 构建种质资源核心库可以有效地克服这一难题. 核心库概念由 Franke(1984)最早提出<sup>[1]</sup>, 核心库由来自一个已存在的种质库中的一批有限遗传材料所组成, 用于代表整个种质库的遗传范围(Brown, 1989)<sup>[2]</sup>. 核心库中入选的遗传材料都是有代表性的, 核心库应包括尽可能多的遗传多样性. 育种工作中可以利用核心库材料的代表性, 筛选需要的性状, 优先与当地品种进行配合力测定, 提高育种效率. 对于昂贵和复杂性状的评价, 核心库可以优先进行, 可为育种家提供更详细和有用的育种材料. 如为了芸苔属抗病虫基因的筛选, Boukema 等(1997)首先将甘蓝分成栽培和野生类型, 再以地理起源分类构建了甘蓝的核心库<sup>[3]</sup>. 目前种质资源核心库的构建研究在国际上处于起步阶段<sup>[4]</sup>, 已有报导利用苜蓿, 野生大豆, 花生, 硬粒小麦, 大麦, 黑麦草, 兵豆, 玉米等为材料进行研究<sup>[4,5,6,7]</sup>. 在构建核心库过程中遇到的一个问题是, 可能丢失一些有用的特殊种质遗传材料<sup>[4]</sup>, 本研究应用多次聚类及合适的取样方法, 使构建的核心库能够保持特殊的遗传材料(或极端类型), 同时提出了评价核心库的标准.

### 1 材料和方法

#### 1.1 材料

以棉花种质资源 168 个基因型, 5 个纤维性状数据为例, 构建核心库. 5 个纤维性状分

收稿日期: 1999-03-04

基金项目: 国家自然科学基金资助项目(39470377)

作者简介: 胡晋(1957-), 男, 浙江永康人, 浙江大学种子科学中心副教授. 博士.

别为 2.5% 跨长 (mm)、整齐度 (%)、强度 (gf/tex)、伸长度 (%)、麦克隆。为方便起见, 分别以 1 至 168 号代表各基因型。

## 1.2 基因型值预测

规模较大的种质资源遗传试验, 可采用不设区组按田间行列编号顺序种植基因型, 以一定间隔种植对照基因型的试验设计方法。用对照基因型控制田间不同位置的差异。对于多年份的遗传试验, 观察值可分解为环境效应、环境内的行效应、环境内的列效应、基因型效应、基因型与环境互作效应, 以及随机机误等分量(朱军, 1994; 1996; 胡晋等, 2000)<sup>[8,9,10]</sup>。可采用朱军(1993)提出的混合线性模型统计分析方法进行统计分析, 用调整无偏预测(AUP)法无偏预测基因型值<sup>[11]</sup>, 用基因型预测值计算遗传距离并用于聚类分析。

## 1.3 多次聚类选取核心库材料

在核心库构建过程中, 不同性状的量纲是不同的, 因此会影响遗传距离的计算。同时性状间的相关也会影响遗传距离的计算。马氏(Mahalanobis)距离处理性状间的相关性, 并可以排除不同性状的量纲影响, 无须将原始数据作适当的变换<sup>[12,13]</sup>。本试验采用马氏距离计算方法计算材料间的遗传距离(胡晋等, 2000)<sup>[10]</sup>。采用系统聚类法中的不加权类平均法(Unweighted pair-group average)聚类。假设类  $G_i$  与  $G_j$  分别有  $n_i$  与  $n_j$  个基因型, 其合并所得的新类为  $G_r$ , 基因型数为  $n_r (= n_i + n_j)$ , 它与其它各类  $G_s$  的类间距离计算公式为<sup>[14]</sup>

$$D_{rs}^2 = \frac{n_i}{n_r} D_{si}^2 + \frac{n_j}{n_r} D_{sj}^2$$

根据遗传变异量和等位基因变异的保持情况, 选取原有种质材料量的 20% ~ 30% ( $p = 0.2 \sim 0.3$ ) 作为核心材料构成核心库<sup>[10,15]</sup>。根据遗传距离进行聚类后, 得到聚类图。根据聚类的原理, 最低分类水平上类(组)内的差异最小。因此从每组二个遗传材料中取样时, 如遇特殊材料(或极端类型), 则优先取该材料进入下一轮聚类分析; 如该组二个材料均为需保存的特殊材料时, 二个材料均进入下一轮聚类; 当该组二个材料均为非特殊材料时, 随机选取一个材料进入下一轮聚类。如组内只有一个材料, 则该材料直接进入下一轮聚类分析; 对所取材料多次聚类, 按同样方法取样, 直到所取材料量达到设定的要求, 即可构成核心库。本研究以保持每性状所有基因型中具最大和最小值材料(极端类型)为例构建核心库。

## 2 结果与分析

首先对表现型值采用调整无偏预测(AUP)法无偏预测基因型效应值<sup>[11]</sup>。然后计算基因型间的马氏距离, 并用不加权类平均法进行系统聚类, 画出聚类图。

在最低分类水平上, 选取材料, 保留每一性状具最大值和最小值的基因型, 进入第二轮聚类的材料有 105 个, 其中 6 个为具最大和最小值的材料。对 105 个材料的基因型值重新计算马氏距离, 进行第二轮聚类, 取样, 得到 69 个材料。第三轮聚类后, 取样得到 44 个材料(图 1), 材料数量符合  $p = 0.2 \sim 0.3$ , 因此决选 44 个材料构成核心库(材料号为: 167, 55, 93, 38, 130, 141, 19, 169, 129, 145, 36, 163, 47, 101, 73, 139, 112, 57, 142, 30, 10, 79, 146, 48, 76, 65, 71, 40, 25, 105, 58, 113, 45, 111, 156, 126, 120, 86, 28, 74, 50, 21, 3, 1)。

以方差、极差、均值和变异系数来衡量构建的核心库是否能够很好地代表原种质资源群体的遗传多样性. 对方差进行  $F$  测验, 对均值进行  $t$  测验. 提出若符合以下条件, 则可认为此资源核心库基本代表了原有资源群体的遗传多样性:

- 条件 (1): 检验  $V_C > V_I$  显著 ( $\alpha = 0.05$ ) 的性状  $\geq 70\%$ ;
- 条件 (2): 检验  $M_C \neq M_I$  显著 ( $\alpha = 0.05$ ) 的性状  $\leq 30\%$ ;
- 条件 (3): 70% 以上性状的  $R_C/R_I$  值  $\geq 0.7$ ;
- 条件 (4): 70% 以上性状的  $CV_C > CV_I$ ;

其中  $V_C, M_C, R_C, CV_C$  分别为资源核心库的方差、均值、极差和变异系数;  $V_I, M_I, R_I, CV_I$  分别为原资源群体的方差、均值、极差和变异系数.

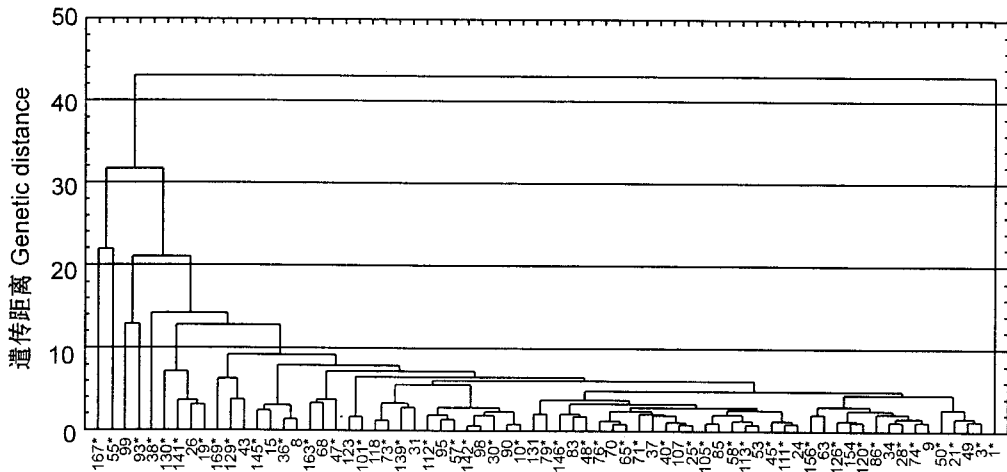


图1 棉花5个纤维性状69个基因型的聚类图(\*为选中的核心库材料)

本方法构建的核心库, 所有性状的方差显著 ( $\alpha = 0.05$ ) 大于原有资源群体的方差 (表1). 核心库所有性状的均值与原有资源群体的均值均无显著差异 ( $\alpha = 0.05$ ), 差异性状为 0%. 所有性状的极差与原有材料的极差相同, 极差比  $R_C/R_I$  值  $\geq 0.7$  的性状为 100%. 所有性状的变异系数均大于原资源群体的变异系数. 可以认为核心库基本代表了原资源群体的遗传多样性.

### 3 讨论

在植物群体里, 遗传多样性不是随机分布的<sup>[16]</sup>, 由于植物的自然进化、栽培的驯化以及现代育种工作的结果, 遗传多样性形成自身的结构, 一般可以通过遗传多样性的系统分类或聚类方法形成遗传多样性树, 然后根据遗传多样性的结构采用合理的方法选取进入核心库的材料. 遗传群体较大的作物种质资源, 可先分成栽培种和野生种, 再按它们的分布、地理起源、生态起源、遗传标记等进行划分大类. 如棉花可先分为栽培种和野生种, 再分为陆地棉、海岛棉、亚洲棉和非洲棉, 然后对每类进行多次聚类构建资源核心材料, 将各大类的核心材料合并为资源核心库. 用于聚类的种质资源个体的各种性状数据, 应尽量采用基因型值, 得到的数据与 DNA 越密切越好, 如可采用 RFLP, RAPD 和同工酶的数据<sup>[17]</sup>.

表1 构建的棉花资源核心库与原资源遗传材料的比较

Table 1 Comparison between the Core Collection and the Initial Accessions of Cotton

性状		方差	均值 **	变异系数	$R_C/R_I$ 值
2.5% 跨长	原资源遗传材料	2.075	26.504	0.054	1.000
	资源核心库	3.467*	26.455	0.070	
整齐度	原资源遗传材料	1.963	50.720	0.028	1.000
	资源核心库	3.115*	50.582	0.035	
强度	原资源遗传材料	0.926	20.011	0.048	1.000
	资源核心库	1.669*	19.821	0.065	
伸长度	原资源遗传材料	0.176	6.064	0.069	1.000
	资源核心库	0.266*	5.994	0.086	
麦克隆	原资源遗传材料	0.087	4.053	0.073	1.000
	资源核心库	0.127*	4.028	0.089	

\* 资源核心库方差显著大于原资源遗传材料方差 ( $\alpha = 0.05$ )

\*\* 资源核心库与原资源遗传材料间均值均无显著差异 ( $\alpha = 0.05$ ),  $R_C$  为资源核心库极差,  $R_I$  为原资源遗传材料极差.

构建核心库时, 如完成了分类(组), 即也确定了不同材料之间的遗传关系, 再从每组进行取样. 取样的方法有按随机, 地理起源<sup>[7]</sup>, 随机结合起源和频率<sup>[18]</sup>等. 但以往的取样尚未考虑到构建核心库时对特殊材料的保持. 构建核心库方法的优劣, 主要看其能否代表原有种质资源群体遗传多样性, Basigalup 等(1995)认为具有最大极差的方法最好, 方差增加的性状愈多, 方法愈好<sup>[6]</sup>, 但评价方差的变化还需结合取样的方法. 采用本研究的取样方法可使所构建的核心库保留原有种质资源的特殊材料(或极端类型), 以较少的材料量最大限度地保存原有群体的遗传变异量, 同时又兼顾了原有种质资源库的遗传多样性结构.

致 谢 辽宁经作所李瑞祥研究员为本文提供实例研究所需的数据, 特致谢意.

### [参 考 文 献]

- [1] Frankel O H. *Genetic Perspectives of Germplasm Conservation*. In: *Genetic Manipulation: Impact on Man and Society*[M]. (Edited by Arber W et al.). Cambridge: Cambridge University Press, 1984.161-170.
- [2] Brown A H D. Core collections: A practical approach to genetic resources management[J]. *Genome*, 1989, 31(5):818-824.
- [3] Boukema I W, Hintum van T J L, Astley D. Creation and composition of the *Brassica oleracea* core collection[J]. *Plant Genetic Resources Newsletter*, 1997, (111):29-32.
- [4] Brown A H D. *The Core Collection at the Crossroads*. In: *Core Collection of Plant Genetic Resources*[M]. (Edited by Hodgkin T et al.), Chichester: Published by John Wiley & Sons, 1995.3-19.
- [5] Casler M D. Patterns of variation in a collection of perennial ryegrass[J]. *Crop Sci*, 1995, 35(4):1169-1177.
- [6] Basigalup D H, Barnes D K, Stucker R E. Development of a core collection for perennial *Medicago* plant introductions[J]. *Crop Sci*, 1995, 35(4):1163-1168.
- [7] Diwan N, McIntosh M S, Baughan G R. Methods of developing a core collection of annual *Medicago* species[J]. *Theor Appl Genet*, 1995, 90(6):755-761.

- [8] 朱军. 广义遗传模型与数量遗传分析新方法 [J]. 浙江农业大学学报, 1994, 20(6):551-559.
- [9] 朱军. 包括基因型环境互作效应的种子遗传模型及其分析方法 [J]. 遗传学报, 1996, 23(1):56-68.
- [10] 胡晋, 徐海明, 朱军. 基因型值多次聚类法构建作物种质资源核心库 [J]. 生物数学学报, 2000, 15(1):103-109.
- [11] 朱军. 作物杂种后代基因型值和杂种优势的预测方法 [J]. 生物数学学报, 1993, 8(1):32-44.
- [12] 裴鑫德. 多元统计分析及其应用 [M]. 北京: 北京农业大学出版社, 1991. 89-195.
- [13] Mahalanobis P C. On the generalized distance in statistics[J]. *Proc Natl Inst Sci India*. 1936, 2(1):49-55.
- [14] Sokal R R, Michener C D. A statistical method for evaluating systematic relationships[J]. *Univ Kansas Sci Bull*, 1958, 38(5):1409-1438.
- [15] Yonezawa K, Nomura T, Morishima H. *Sampling Strategies for Use in Stratified Germplasm Collections. In Core Collection of Plant Genetic Resources*[M]. (Edited by Hodgkin T et al). Chichester: Published by John Wiley & Sons, 1995, 35-53.
- [16] Hintum van T J L. *Hierarchical Approaches to the Analysis of Genetic Diversity in Crop Plants. In: Core Collections of Plant Genetic Resources*[M]. (Edited by Hodgkin T et al). Chichester: Published by John Wiley & Sons, 1995. 23-34.
- [17] Tanksley S D, McCouch S R. Seed banks and Molecular maps: Unlocking genetic potential from the wild[J]. *Science*, 1997, 277(5329):1063-1066.
- [18] Zeuli S P L, Qualset C O. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat[J]. *Theor Appl Genet*, 1993, 87(3):295-304.

## A Method of Constructing Core Collection Reserving Special Germplasm Materials

HU Jin XU Hai-ming ZHU Jun

(Seed Science and Technology Center, Zhejiang University, Zhejiang Hangzhou 310029 China)

**Abstract:** A method of constructing core collection reserving special germplasm materials was proposed. Phenotype data of genetic accessions were used to predict genotype values by a genetic model controlling systematical errors in fields and by mixed model approaches. Predicted genotype values were used to construct core collection. Genetic distance among accessions was calculated with Mahalanobis distance, then unweighted pair-group average method of hierarchical cluster was used for grouping. According to a dendrogram of clusters the accession with special merit was preferred to be chosen for core collection. *F* test for variances, *t*-test for means, range rate and coefficient of variation were used to evaluate the property of core collection for initial accessions. An worked example of constructing core collection on 168 accessions of cotton by five fiber traits was presented.

**Key words:** Genotype value ; Core collection ; Germplasm ; Multiple cluster ; Cotton